

INHIBITION AND MECHANISTIC STUDIES OF FTO

A thesis submitted to the Board of the faculty of Physical Sciences of the University of Oxford in partial fulfilment of degree of Doctor of Philosophy in Organic Chemistry

Shifali Shishodia

Corpus Christi College

University of Oxford

Michaelmas 2017

Abstract

Shifali Shishodia D.Phil.
Corpus Christi College Michaelmas Term, 2017

The Fat Mass and Obesity Associated protein (FTO) is a nucleic acid demethylase that belongs to the 2-oxoglutarate dependent oxygenase superfamily (2OG oxygenases). Like other 2OG oxygenases, FTO uses Fe(II), 2-oxoglutarate (2OG) and molecular oxygen (O_2) to catalyse oxidation of its substrates. Mutation of the gene encoding for FTO have been linked to obesity, type II diabetes, cancer, and Alzheimer's disease. One aspect of the work described in this thesis aimed at developing small molecule inhibitors of FTO. An NMR binding assay was developed using 2OG as a reporter ligand. The inhibitors developed in this thesis exhibit high potency and selectivity for FTO (the lead inhibitor has an IC_{50} value of 1.32 μM).

The work also aimed at the synthesis of N^6-methyladenosine (m^6A) phosphoramidite with the aim of synthesising single stranded RNA (ss-RNA) oligonucleotide containing m^6A. Apart from the literature reported substrates, work also focussed on the investigations of other nucleosides as natural and unnatural substrates of FTO. During the study, new synthetic routes to the m^6A phosphoramidite, ethanoadenosine, ethylethanoadenosine and methylethanoadenosine were developed. The iodouridine phosphoramidite was also synthesised with the future aim of synthesising a ss-RNA substrate crosslinked to FTO.

The reaction of formaldehyde (the FTO co-product) with canonical and non-canonical nucleotides was then investigated using NMR. Various nucleotide/HCHO adducts were identified, including multiple novel species. Among the canonical nucleosides, TMP and UMP reacted the fastest followed by dCMP, dGMP and lastly dAMP. However, dAMP was observed to form the most stable adduct with formaldehyde, while the adducts of TMP and UMP were dynamically favoured. Comparing cytidine, 5-methyldeoxycytidine (5mdC) and 5-hydroxymethyldeoxycytidine (5hmdC) revealed that 5mdC reacts the fastest. Among m^6A, N^1-methyladenosine (m^1A) and adenosine, the initial rate of reaction with formaldehyde was fastest with adenosine.

Finally, the FTO and AlkB catalysed oxidative demethylation of substrates was investigated by NMR using [^{13}C] labelled substrates (e.g., N^6-[^{13}C]-methyladenosine ([^{13}C]-m^6A), N^3-[^{13}C]-methythymidine ([^{13}C]-m^3T) or N^1-methyladenosine ([^{13}C]-m^1A)) as probes. The studies reveal the first direct evidence, using NMR, for the formation of N^6-hydroxymethyladenosine as an intermediate of FTO-catalysed oxidative demethylation of m^6A. The results also confirmed the formation of formaldehyde. The observation of formaldehyde in AlkB-catalysed oxidative demethylation of m^1A led to the invalidation of the previously reported hypothesis that a phosphate group attached to a nucleoside is the minimum requirement for AlkB substrates.

Overall, the work described in this thesis suggests new approaches to the inhibition of FTO, helps to define the range of substrates for FTO, reassign the requirement for AlkB-catalysed oxidative demethylation, and informs on the reaction of the FTO-catalysed reaction co-product, HCHO, with nucleosides.

Dedicated to Hanumanji and my Parents

Acknowledgements

Firstly, I should like to thank my supervisor, Prof. Christopher J. Schofield, for accepting me into his lab and providing me with a lot of scientific freedom to pursue research. I would also like to thank him for his ever encouraging nature and the faith he put in me and in my scientific ideas.

I would take this opportunity to especially express my gratitude to Dr Richard Hopkinson for introducing me to the realm of NMR and formaldehyde, and guiding me throughout the project with patience. I thank Dr Marina Demetriades for providing me with an initial platform for FTO inhibitors, biochemical assays and the guidance during my initial days in lab.

I would like to thank Prof. Timothy Claridge for enriching discussions and guidance on NMR experiments. I would also like to thank my collaborator, Prof. Afaf El-Sagheer, for her excellent guidance on oligonucleotide synthesis.

I would like to thank Dr Michael McDonough for insightful ideas and discussions on FTO.

Thanks to Dong Zhang for always being a sweet buddy and showing faith in me, Dr Rachel Schiller for sharing a bond I would enjoy throughout life, Dr Rashed Chowdhury for the inspirational and motivational talks, every single time I needed them, Dr Kuo-Feng, and all the past and present members of the Schofield group for their day to day guidance and help in experiments and maintaining enjoyable environment in the lab. I would like to thank Dr Adam Hardy and Dr Wendy Sobey for lab administration.

Thesis-wise I would like to thank Dr Richard Hopkinson, Dr Christopher Lohans, Dr Ian Clifton, Dr Jaideep Saha for extensive proof reading.

I am greatly indebted to the donors of "Felix" scholarship for generously funding my tuition fees and living expenses for three years of my D.Phil. I also express my gratitude to the

generous donors of “Vice-Chancellor” Fund for supporting my living expenses for the last leg of my stay in Oxford.

My heartfelt thanks to my lovely parents, Mr. Dinesh Kumar Shishodia and Mrs. Lokesh Shishodia, for the unconditional love, support and encouragement, they have showered on me throughout my life. I would take this opportunity to thank my mausaji, Mr. Dinesh Pratap Singh, for the mentor you have been to me. Though, younger than me, Siddhant Pratap Shishodia, your encouragement always boosted my morals. Kokil Tandon, thanks for everything.

Last, but not the least, I would like to thank Hanumanji, for constantly showering blessings on me and patiently listening to my endless saga of life and work. Without your support, I would not have made so far.

List of Abbreviations

2OG	2-oxoglutarate
A	adenosine
T	thymidine
C	cytidine
G	guanosine
U	uridine
I	inosine
5mC	5-methylcytidine
5hmC	5-hydroxymethylcytidine
5fC	5-formylcytidine
tRNA	transfer ribonucleic acid
rRNA	ribosomal ribonucleic acid
m^3T	3-methylthymidine
m^6A	6-methyladenine
m^3U	3-methyluridine
m^1A	1-methyladenosine
$m^6{}_2A$	N^6,N^6-dimethyladenosine
m^3T	3-methylthymidine
m^4C	4-methylcytidine
m^3C	3-methylcytidine
m^6Am	N^6,O^2-dimethyladenosine
εA	ethanoadenosine
εC	ethanocytidine
m^1G	1-methylguanosine
hm^6A	N^6-hydroxymethyladenosine
f^6A	N^6-formyladenosine
Ψ	pseudouridine
AlkB	alkylation repair enzyme B
ABH(1-8)	AlkB human homologues (1-8)
ABHs	AlkB Human Homologues
ACCO	1-aminocyclopropane-1-carboxylate oxygenase
Ac	acetyl
AFN	ammonium formate
Bn	benzyl
BOP	1-*H*-benzotriazol-1-yloxytris(dimethylamino)phosphoniumhexafluorophosphate
$CDCl_3$-d	deuterated chloroform
CH_2Cl_2/ DCM	dichloromethane
CH_3I	methyl iodide
CPH	procollagen prolyl hydroxylases
Cs_2CO_3	cesium carbonate
$CuSO_4$	copper sulphate
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DMF	dimethylformamide
DNA	deoxyribonucleic acid
DIPEA	diisopropylethylamine
DMSO-d_6	deuterated dimethylsulfoxide

DSBH	double stranded βhelix
D_2O	deuterium oxide
DTBS triflate	Di-*tert*-butylsilyl bis(trifluoromethanesulfonate)
DNA	deoxyribonucleic aacid
DNMTs	DNA methyltransferases
DMTr	dimethoxytrityl
E.Coli	Escherichia coli
ESI-MS	electrospray ionisation mass spectrometry
EtOAc	ethyl acetate
EtOH	ethanol
FTO	fat mass and obesity protein
FGG2216	(1-chloro-4-hydroxyisoquinoline-3-carbonyl)glycine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high performance liquid chromatography
HMBC	heteronuclear multiple bond correlation
HSQC	heteronuclear single quantum correlation
ITC	isothermal titration calorimetry
JMJD2C	Jumonji Domain-Containing Protein 2C
JMJD2A	Jumonji C Domain-Containing Histone Demethylase 2A
JMJD3	Jumonji Domain Containing 3, Histone Lysine Demethylase
K_D	dissociation constant
K_D^{app}	apparent dissociation constant
KI	association constant
KDM4C	lysine (K)-Specific Demethylase 4C
KDM3A	lysine (K)-Specific Demethylase 3A
KDM6B	lysine (K)-Specific Demethylase 6B
$K_2S_2O_8$	potassium peroxodisulfate
HPPE	*S*-hydroxypropylphosphonic acid epoxidase
HRMS	high resolution mass spectrometry
IR	infrared
LCMS	liquid chromatography mass spectrometry
LiHMDS	lithium bis(trimethylsilyl)amide
MMS	methylmethanesulfonate
MNU	*N*-methyl-*N*′-nitrosourea
METTL3/METTL14	methyltransferase Like 3/ Methyltransferase Like 14
MALDI	matrix assisted laser desorption ionisation
MMTr	monomethyltitryl
MeOH	methanol
NOESY	nuclear overhauser effect spectroscopy
mRNA	messenger ribonucleic acid
ssRNA	single-stranded ribonucleic acid
dsRNA	double-stranded ribonucleic acid
T_m	melting shift temperature
mp	melting point
MS	mass spectrometry
mw	microwave irradiation
NAOXs	nucleic acid oxygenases
$NaBH_4$	sodium borohydride
NBS	*N*-bromosuccinimide
NMR	nuclear magnetic resonance

OTf	triflate
PCR	Polymerase chain reaction
PDB	protein databank
PHD2	prolyl hydroxylase domain 2
$POCl_3$	Phosphoryl chloride
PyBOP	(Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
Q-TOF	quadrupole-time of flight
RNA	ribonucleic acid
SAR	structure-activity relation
SAM	*S*-adenosylmethionine
S_N1	unimolecular substitution reaction
S_N2	bimolecular substitution reaction
S_NAr	nucleophilic aromatic substitution
r.t	room temperature
TET	Ten-eleven translocation
Tris	tris(hydroxymethyl)aminomethane
T3P	propylphosphonic anhydride
THF	tertahydrofuran
TFA	trifluoroacetic acid
TBDMSCl	*t*-Butyldimethylsilylchloride
TLC	thin layer chromatography
TOF	time of flight
TrCl	trityl chloride
TBAI	tetrabutylammonium iodide
UV	ultraviolet
WTAP	Wilms' Tumor 1 Associated Protein
YTHD	YTH domain

Table of Contents

Chapter1: Introduction 1

1.1 2OG Oxygenases 1

1.1.1 General Introduction 1

1.2 Structure of 2OG oxygenases 4

1.2.1 The double stranded ß-helix fold (DSBH) 4

1.2.2 The Fe binding motif/ the catalytic site 4

1.2.3 2OG binding pocket 5

1.2.4 Substrate binding pocket 6

1.2.5 Other DSBH enzymes independent of 2OG 7

1.3 Mechanism of 2OG oxygenases 7

1.3.1 Initial Fe(II) complex 8

1.3.2 Binding of 2OG to the Fe(II) centre 8

1.3.3 Binding of substrate and oxygen to the Fe(II) centre 10

1.3.4 Formation of Fe=O species 10

1.3.5 The release of carbon dioxide 11

1.3.7 Release of product and succinate 12

1.4 Nucleic Acid Demethylases 12

1.4.1 AlkB 14

1.4.2 AlkB Human Homologue proteins 17

1.5 RNA epigenetics: 25

1.5.1 Other potential epigenetic modifications 29

1.6 References 31

Chapter 2: Inhibition studies on FTO 41

2.1 Introduction 41

2.1 Previous studies on Inhibitors 41

2.1.1 General 2OG inhibitors 41

2.2 Previous studies on FTO inhibition 43

2.3 The objectives of the work described in this chapter 46

2.4 Results and Discussion 46

2.4.1 Development of FTO Inhibitors 46

2.4.2 Synthesis of Inhibitors 49

2.4.3 Buchwald-Hartwig Coupling 52

2.4.4 Coupling of amines using Hartwig-Buchwald coupling 54

2.4.5 Differential Scanning Fluorimetry 60

2.4.6 IC_{50} values 62

2.4.7 Dissociation constants for the FTO inhibitors 65

2.4.8 Optimisation of the binding assay 67

2.5 Conclusions and Future Work 77

2.6 References 79

Chapter 3: Synthesis of substrates for FTO 85

3.1 Introduction 85

3.2 Summary of Oligonucleotide synthesis 86

3.2.1 Use of protecting groups in oligonucleotide synthesis 87

3.2.2 Summary of Oligonucleotide procedure 87

3.3 Objectives of the work described in this chapter 92

3.4 Results and Discussion 92

3.4.1 Optimisation of the synthetic route for m^6A phosphoramidite 92

3.4.2 Synthesis of protected monomer of 5-iodouridine 108

3.4.3 Synthesis of analogues of m^6A 110

3.5 Conclusions 127

3.6 Acknowledgements 129

3.7 Reference 129

Chapter 4: Studies on the reaction of formaldehyde with nucleotides 133

4.1 Introduction 133

4.1.1 Physical Properties, Structural Information and Industrial Uses 133

4.1.2 Sources of Formaldehyde 133

4.1.3 Crosslinking through formaldehyde 134

4.1.4 Toxicity of formaldehyde 135

4.2 Previous studies on formaldehyde 135

4.3 Objectives of work described in this chapter 139

4.4 Results and Discussion 140

4.4.1 NMR Analysis of formaldehyde solution 140

4.4.2 Optimisation of the reaction conditions of formaldehyde with nucleotides 141

4.4.3 Studies on the reaction of nucleotide/nucleoside with formaldehyde 143

4.4.4 Studies on the reaction of formaldehyde with deoxyadenosinemonophosphate (dAMP)..... 144

4.4.5 Studies on the reaction of formaldehyde with deoxyguanosinemonophosphate (dGMP).....151

4.4.6 Studies on the reaction with formaldehyde with deoxycytidinemonophosphate (dCMP)..... 159

4.4.7 Studies on the reaction of formaldehyde with thymidinemonophosphate (TMP).... 172

4.4.8 Studies on the reaction of formaldehyde with uridinemonophosphate (UMP)..178

4.4.9 Studies on the reactivity of N^6-methyladenosine (m^6A) with formaldehyde.....185

4.4.10 Studies on the reaction of N^1-methyldeoxyadenosine (m^1dA) with formaldehyde... 191

4.4.11 Studies on the reaction of formaldehyde with m^6G 193

4.4.12 Studies on the reaction of formaldehyde with inosine (I) 200

4.4.13 Studies on the reaction of formaldehyde with 5-hydroxymethylcytidine (hm^5C).... 206

4.4.14 Studies on the reaction of formaldehyde with 5-methyldeoxycytidine (m^5dC).....213

4.4.15 Studies on reaction of formaldehyde with pseudouridine (Ψ) 219

4.4.16 Comparison of the reactivity of cytidine (C), hm^5dC and m^5dC 227

4.4.17 Comparison of the reactivities of dAMP, dGMP, dCMP, TMP and UMP with formaldehyde. 231

4.4.18 Effect of temperature on the rates of reaction of formaldehyde with nucleotide..... 236

4.4.19 Effect of change in the sugar moiety on the rates of reaction of formaldehyde with nitrogenous base. 238

4.4.20 Effect of additional phosphate group on the rate of reaction 243

4.4.21 Stability of the *N*-hyroxymethyladduct 247

4.5 Discussion 251

4.6 References 254

Chapter 5: Mechanistic studies on FTO using carbon-13 labelled substrates as probes 259

5.1 Introduction 259

5.2 Objectives of the work described in this chapter 263

5.3 Results 263

5.3.1 Studies on m^6A as a substrate for FTO 263

5.3.2 Optimisation of the FTO assay 264

5.3.3 Studies on hm^6A as a substrate for FTO 281

5.3.4 Studies on N^3-[^{13}C]-methylthymidine ([13C]-m^3T) as a substrate for FTO 282

5.3.5 Studies on m^1A as a substrate for AlkB 285

5.4 Discussion 289

5.5 Acknowledgements 292

5.6 References 292

Chapter 6: Experimental 295

6.1 General experimental 295

6.1.1 Chemical Synthesis 295
6.1.2 Buffer solution 296
6.1.3 Enzymatic assays-stock solutions 296
6.2 General NMR sample handling and processing 296
6.3 Experimental for Chapter 2 297
6.3.1 Experimental details 297
6.3.2 Synthesis and Characterisation of FTO Inhibitors 298
6.4 Binding Experiment 308
6.5 FTO expression and purification 309
6.6 Apo Protein 309
6.7 In Vitro Demethylation Assay 310
6.8 Dynamic Scanning Fluorometry Assays 310
6.9 Experimental for Chapter 3 311
6.10 Experimental for Chapter 4 348
6.10.1 ^{1}H NMR Time-Courses 348
6.10.2 1D HSQC NMR experiment 349
6.10.3 EXSY experiments 349
6.11 Experimental for Chapter 5 349
6.11.1 Optimised assay conditions for FTO 349
6.11.2 Optimised assay conditions for AlkB 349
6.11.3 Synthetic Procedures 350

6.11.4 AlkB expression and purification 355

6.12 References.....357

Chapter1: Introduction

1.1 2OG Oxygenases

1.1.1 General Introduction

The ferrous iron/2-oxoglutarate (2OG) dependent oxygenases, often referred to as 2OG oxygenases, comprise an enzyme superfamily that catalyse a diverse set of reactions, some of which are medicinally important.[1] The 2OG oxygenases use a citric acid cycle intermediate, 2OG,[2] and oxygen as co-substrates and Fe(II) as a cofactor. 2OG oxygenases produce succinate and carbon dioxide as co-products. The first 2OG oxygenases to be reported, by Hutton *et al.,*[3] were the procollagen prolyl hydroxylases (CPHs) which are involved in collagen biosynthesis. The 2OG oxygenases are found in all life forms, except archaea, and comprise the largest known non-haem iron containing protein superfamily.[4] More than 60 human 2OG oxygenases have been defined, with likely more to be identified.[5] The 2OG oxygenases catalyse reactions including hydroxylations,[6–8] demethylations,[9–14] desaturations,[15] ring formations,[16] epimerisations,[15] epoxidations,[17] and halogenations.[18–20] In human metabolism, presently known 2OG oxygenases only catalyse hydroxylation and demethylation (proceeding via hydroxylation) reactions.[21] The reactions carried out by 2OG oxygenases, especially the functionalisation of inert C-H bonds, are often challenging to perform synthetically, thus details of their mechanisms are of chemical interest. Studies on 2OG oxygenases are also pursued because of their importance in many environmental,[22] medicinal and pharmacological fields.[23–26] 2OG oxygenases have been found to be involved in genetic diseases such as Ehlers–Danlos syndrome[27] and adult Refsum disease.[28]

2OG oxygenases catalyse the two electron oxidation of their substrates with the concomitant oxidative decarboxylation of 2OG to produce succinate and carbon dioxide. The biological functions carried out by 2OG oxygenases are among the most diverse of any known enzyme

family.[1,4] Their functions include chromatin regulation involving modifications of nucleic acids (5-methylcytidine)[29–33] and histone tails (*N*-methyllysine),[23,34,35] fatty acid metabolism (biosynthesis of L-carnitine and degradation of phytanic acid),[1] and hypoxia signalling[36,37] to reversing endogenous and exogenous methylation on nucleic acids. [9–12,14,21,30,31,38–45] 2OG oxygenases acting on nucleic acids can protect the genome from malfunctioning, and include the AlkB Human Homologues 1-8 (ABHs1-8) and the Fat Mass and Obesity Associated Protein (FTO). The substrates and reactions of characterised human 2OG oxygenases are summarised in Figure 1.1.

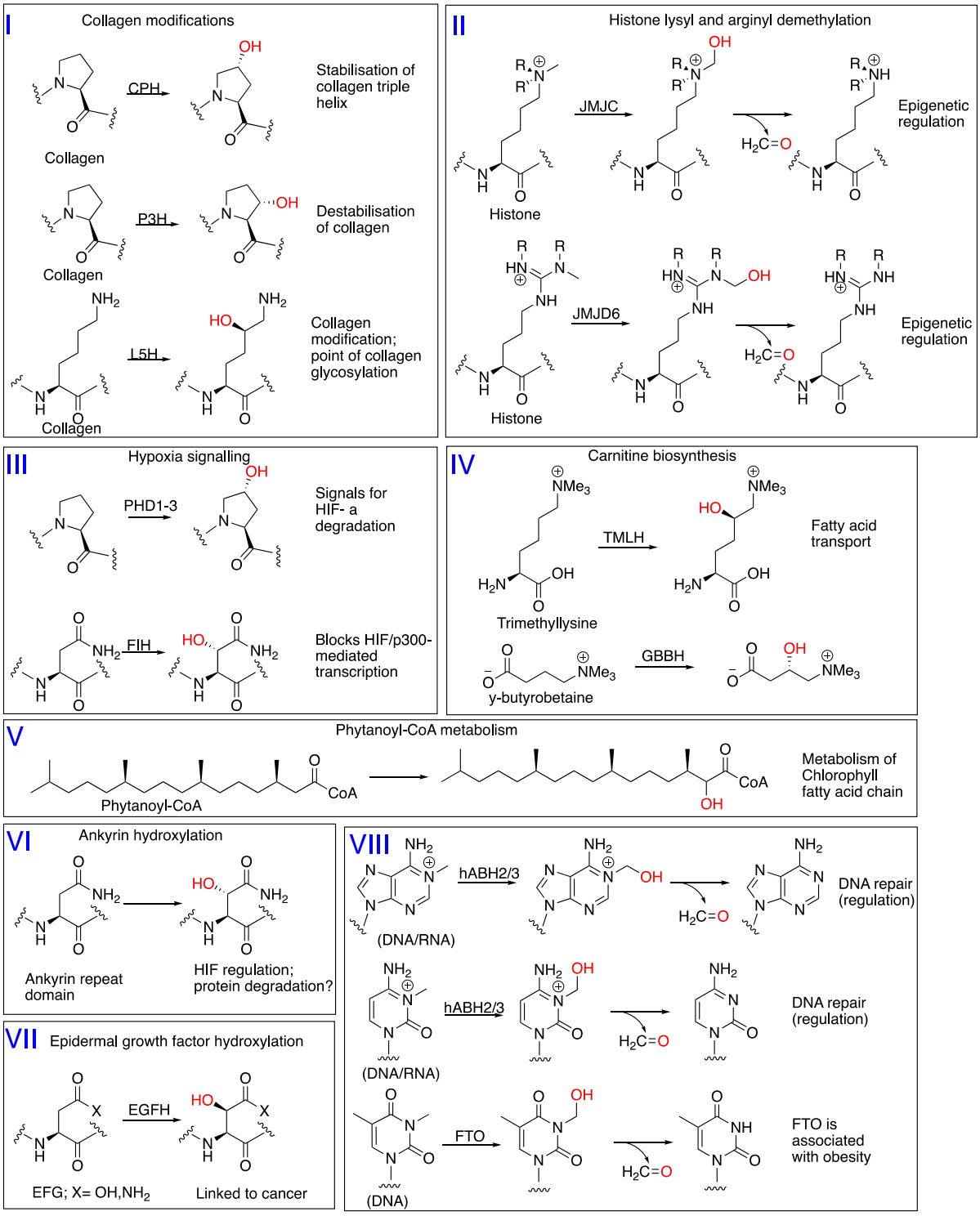


Figure 1.1 Schematic representation of the reactions catalysed by 2OG human oxygenases. (I) Pro-collagen modifications: enzymes involved CPH, P3H and L5H; (II) epigenetic regulation: enzymes involved JMJC, JMJD6; (III) hypoxia sensing: enzymes involved PHD1-3 and FIH; (IV) carnitine biosynthesis: enzymes involved TMLH, GBBH; (V) phytanoyl-CoA-metabolism; (VI) ankyrin hydroxylation; (VII) epidermal growth factor hydroxylation; (VIII) DNA repair and modification.[1]

1.2 Structure of 2OG oxygenases

1.2.1 The double stranded ß-helix fold (DSBH)

Crystallographic studies have revealed the existence of a characteristic distorted double stranded ß-helix fold (DSBH, also known as jelly roll motif or double Greek key motif) in the 2OG oxygenases (Figure 1.2B).[46–51] The first crystallographic structure from the superfamily that was solved was that of isopenicillin N synthase (IPNS); this structure revealed the existence of the core DSBH and the details of the non-haem Fe(II) active site.[51] Apart from the 2OG oxygenases, the DSBH fold has been observed in other protein families, including the plant cupin-fold proteins[52] and the jumonji C (JmjC) chromatin associated proteins.[34,53] The DSBH fold consists of eight antiparallel ß strands that form a ß-sandwich consisting of two four stranded antiparallel ß-sheets, one major ß-sheet consisting of ß-strands I, VIII, III and VI and the other minor ß-sheet consisting of ß strands II, VII, IV, and V (Figure 1.2b).[5,54,55] Adjacent ß-strands are in different sheets with the exception of the fourth and the fifth ß-strands.[54,55] The major ß-sheet is generally extended by parallel or antiparallel ß-strands from its N-terminal region; the minor ß-sheet is generally not extended by more than one ß-strand from the C-terminal region.[54,55] The role of DSBH fold is likely, in part, to provide a rigid scaffold to provide support to Fe(II) and confer selectivity for the binding of the substrate.[54]

1.2.2 The Fe binding motif/ the catalytic site

The non-haem Fe(II) binding site, is highly, but not universally, conserved in 2OG oxygenases. It normally consists of a facial motif His-Xxx-Asp/Glu (generally, 2 His-1-carboxylate facial triad) situated at the more open end of the DBSH core.[6,34,50,56] The presence of Asp as the acidic residue is more common than Glu; Asp accounts for the 80% of

the acidic residues present in the facial triad motif of 2OG oxygenases.[54,55] Variations to the facial triad sequence have been observed with some 2OG oxygenases. For instance, in the 2OG dependent halogenases where the active site, Fe(II) is coordinated by two His residues and a chloride ion; the carboxylate residue is replaced by an Ala.[57,58] Crystallographic studies reveal that the Fe(II) exists in an octahedral complex with three sites occupied by the facial triad and the other three by water molecules. The binding of Fe(II) by 2OG oxygenases is normally weak as compared to Fe(II)-haem oxygenases, and there have been considerable variations observed in the strength of Fe(II)-enzyme binding in different 2OG enzymes. The differences in the binding strength of Fe(II) to 2OG oxygenases may be of functional relevance in some cases.[59,60]

1.2.3 2OG binding pocket

The 2OG binding site is buried within the major and the minor ß-sheets of the DSBH. [5,54,55] The 2OG binding residues are less conserved and are characteristic of the different sub-families within the 2OG oxygenase superfamily. In the enzyme-Fe(II) complex, 2OG chelates to Fe(II) in a bidentate manner, causing two molecules of water to leave the octahedral complex (as supported by crystallographic studies).[5,24,54,55] The non-coordinating oxygen atom of the C-1 carboxylate of 2OG is sometimes involved in electrostatic interactions with an alcohol or a basic residue, for instance, an Arg residue in TauD[61] and in the case of most nucleic acid demethylases.[62–64]

The binding elements of the C-5 carboxylate of 2OG have common features, e.g the presence of a basic residue (Arg/ Lys) and at least one hydroxyl group from either a Ser, Thr or Tyr residue. These residues form electrostatic and hydrogen bonding interactions with the C-5 carboxylate of 2OG. For instance, the C-5 carboxylate of 2OG in the AlkB human homologue 8 (ABH8)[65,66] and the fat mass and obesity associated protein (FTO),[10,67,68]

interacts with an Arg-X-Ser motif. An interesting variation is found in the JmjC subfamily of 2OG oxygenases in which C-5 carboxylate of 2OG is positioned to interact with a Lys, rather than an Arg side chain.[34]

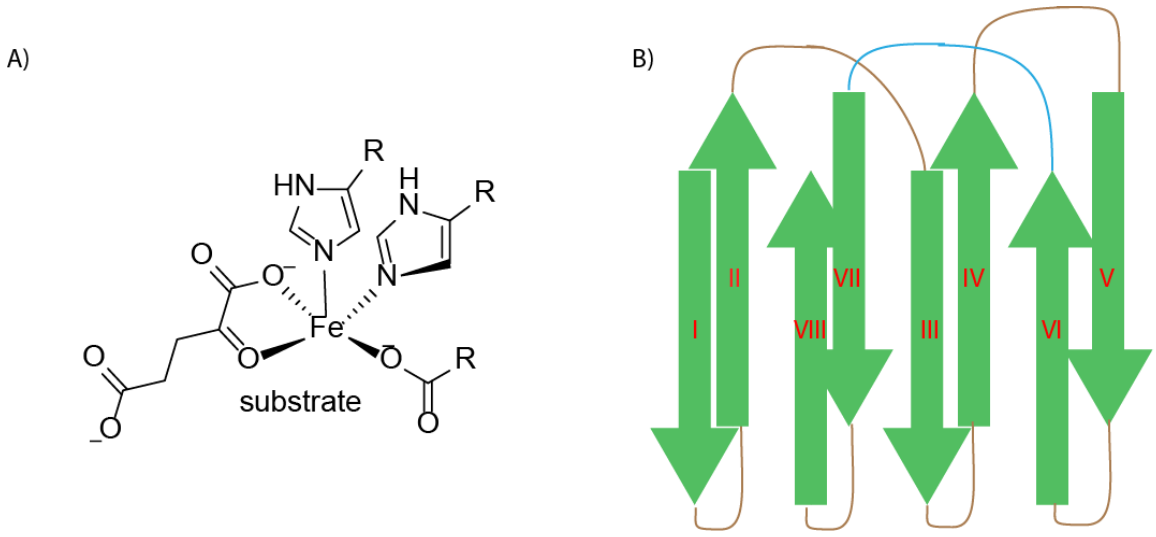


Figure 1.2 A) Schematic representations of the 2-His-1-carboxylate facial triad found in most 2OG oxygenases. It should be noted that 2OG binds in a bidentate manner. B) Topology diagram of DSBH fold of 2OG oxygenases.[5]

1.2.4 Substrate binding pocket

Although the 2OG oxygenases have similar core DSBH folds and have highly, but not universally, conserved Fe(II) and 2OG binding sites, there is much more variation in their substrate binding modes, as may be anticipated by the range of their substrate types.[4,55,69] 2OG oxygenases likely evolved to give different subfamilies with different substrate recognition modes. Their substrate binding often involves mobile elements that can be involved in enabling selectivity.[55] The length and mobility of the loops involved in substrate binding vary. For instance, the residues linking the ß-IV and ß-V strands in PHD2[70] form relatively tight turn compared to the more mobile loops in TauD[48] and related 2OG oxygenases. 2OG oxygenases substrate types include alkyl chains, amino acids (including peptides and proteins) and nucleic acids (including RNA and DNA). Substrate binding can involve hydrogen bonding, hydrophobic and electrostatic interactions; collectively these

interactions must position the substrate carbon-hydrogen bond that is oxidised close to the active site, Fe (II).[54] 2OG oxygenases can be highly selective, e.g. for binding to single stranded-DNA/single stranded- RNA (ssDNA/ss-RNA), double stranded-DNA (dsDNA), or binding to positively charged, neutral or negatively charged substrates. In other cases, the substrates accepted can be promiscuous. In the case of AlkB and the AlkB human homologues, the residues N-terminal to the DSBH fold are involved in binding the nucleic acid backbone (phosphate group) and flipping the methylated nitrogenous base into the active site.[54,62,71]

1.2.5 Other DSBH enzymes independent of 2OG

There are three other known oxygen dependent enzymes that are structurally similar to 2OG oxygenases and require iron as a cofactor, however, which do not utilise 2OG as a cosubstrate. IPNS catalyses the four electron oxidation of a tripeptide to form the penicillin nucleus.[15,72] ACCO (1-aminocyclopropane-1-carboxylate oxygenase) catalyses the formation of ethylene involved in plant signaling.[73] HPPE (*S*-hydroxypropylphosphonic acid epoxidase) catalyses the formation of the epoxide antibiotic, fosfomycin.[74,75]

1.3 Mechanism of 2OG oxygenases

In 1982, an insightful mechanism for the prolyl *4R*-hydroxylases was proposed.[76] This proposal included Fe(II) coordination in a facial triad by protein residues, binding of 2OG in a bidentate manner and the presence of a metal-oxo intermediate.[76] Many features described in the mechanism are still relevant. These proposal were subsequently supported by spectroscopic and crystallographic analyses for several 2OG oxygenases.[5,6,16,17,48,51,55,77–91]

1.3.1 Initial Fe(II) complex

Kinetic, spectroscopic and crystallographic studies have revealed that Fe(II) binds to the active site prior to the substrate or 2OG binding.[50,92,93] Crystallographic studies suggest that the iron exists in an octahedral complex with three sites occupied by the residues from the protein (two His and one carboxylate, either Asp or Glu) and the other three sites occupied by water molecules, prior to binding of 2OG and substrate (**I** in Figure 1.3).[50,92,93]

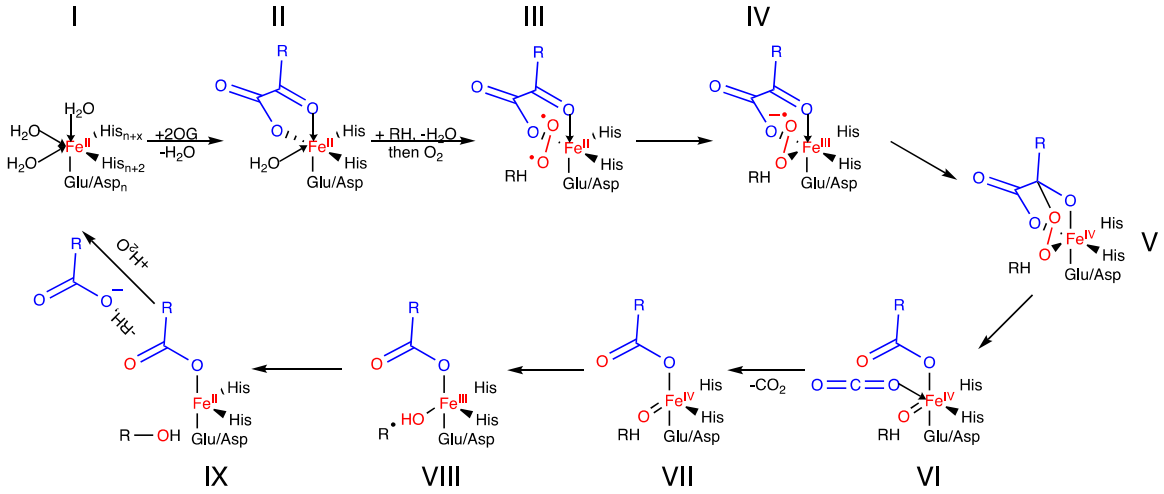


Figure 1.3 Schematic representation of the proposed common mechanism for 2OG oxygenases.[5]

1.3.2 Binding of 2OG to the Fe(II) centre

The first proposed mechanistic step is the binding of 2OG to the iron centre. According to crystallographic studies, 2OG binds to Fe(II) in a bidentate manner via its 2-oxo group and one of its 1-carboxylate oxygens causing loss of coordination of two water molecules (**II** in Figure 1.3).[5,24,54,55] After ligation of 2OG into the octahedral complex, the sixth coordination site is either occupied by a water molecule or remains vacant.[6,16,54,55]

The 2-oxo group of 2OG has been reported to coordinate to Fe(II) *trans* to the carboxylate residue of the facial triad (2-His and 1-carboxylate).[54,55] However, the coordination position of the carboxylate oxygen has been reported to vary; carboxylate oxygen may coordinate

trans to the metal chelating proximal His (His_{n+2}) or trans to the *distal* His ($His_{n+x)}$ of the facial triad.[5,55,94] The variations observed in the binding pattern of 2OG may be of mechanistic relevance.[94]

1.3.2.1 Binding of 1-carboxylate oxygen trans to the proximal or the distal His residue:

When 2OG C-1 carboxylate binds *trans* to the proximal His residue (His_{n+2}), the vacant or water coordinated site (*trans* to the distal His (His_{n+x})) is close to the substrate bond which is to be oxidised. Coordination of oxygen to this site with concomitant formation of a highly oxidising Fe=O intermediate put the substrate appropriately close to the oxidising site.[85,86] In some cases, it has been observed that the 2OG C-1 carboxylate binds *trans* to the distal His (His_{n+x}) residue of the HXD…H facial triad; these observations suggest that binding of oxygen in this position would lead to the formation of a ferryl species projecting away from the substrate bond to be oxidised.[95] Thus, the Fe=O species may have to rearrange in order to be proximate to the substrate C-H bond. The first case where this binding orientation of oxygen was observed was with clavaminic acid synthase (CAS) where nitric oxide was used as an oxygen analogue.[95] A CAS-Fe-2OG-substrate complex was crystallised then exposed to nitric oxide, which coordinated to the metal in a manner projecting away from the substrate.[95] Binding of oxygen in this orientation has the advantage that the reactive species is likely to be buried within the active site and away from the environment. Moreover, the binding of nitric oxide as an oxygen analogue to the CAS-Fe-2OG-substrate complex suggests the possibility that oxygen might pass through the ß-sheets of DSBH in order to coordinate to Fe(II)-complex. It may be that oxygen can bind to either of the possible coordination sites depending upon the Enzyme-Fe(II)-2OG-substrate complex and the availability of the vacant site.

1.3.3 Binding of substrate and oxygen to the Fe(II) centre

Following 2OG binding, the substrate or dioxygen binds to the active site. Spectroscopic and crystallographic studies involving various substrates reveal that substrate binding in proximity to the Fe(II)-2OG complex activates the Fe(II) centre for the binding of dioxygen following displacement of the metal-ligated water molecule.[16,54,55] The late stage binding of oxygen to the Fe(II) complex also minimises the formation of highly oxidising species (Fe(IV)=O) until the substrate is bound, thus minimising the inadvertent oxidation of the active site residues or protein residues which could lead to functional imparity of the enzyme. However, there have been studies which, at least under some conditions, support the binding of oxygen prior to substrate binding (as observed in AlkB and other nucleic acid demethylases), this can result in a substantial uncoupled turnover of 2OG to succinate.[96] Thus, the initial sequence of binding of substrate or oxygen may alter depending upon the enzymes and the environment.

The binding of molecular oxygen to the Fe(II).2OG.substrate complex is proposed to oxidise the iron centre to a Fe(III)-superoxo intermediate (**IV** in Figure 1.3).[79] The nucleophilic distal oxygen of the Fe(III)-superoxo intermediate can then attack the electrophilic carbonyl (C-2) of 2OG, forming peroxide (**V** in Figure 1.3). The formation of intermediate peroxide (**V** in Figure 1.3) is supported by computational studies.[97]

1.3.4 Formation of Fe=O species

The next proposed mechanistic step involves the oxidative decarboxylation of the cyclic peroxide intermediate to give carbon dioxide, succinate and the substrate oxidising Fe(IV)=O intermediate (**VI** in Figure 1.4). The first direct evidence supporting the presence of the highly oxidising species came from the studies on TauD where a Fe(IV)=O complex was observed.[78,80] The presence of a Fe(IV)=O intermediate has been shown with many other, but

not all, 2OG oxygenases characterised by kinetic and spectroscopic techniques.[77–80,83,85] As discussed above, binding of oxygen *trans* to the distal His (His_{n+x}) at the facial triad leads to the formation of Fe(IV)=O projecting away from the substrate bond to be oxidised and needs the ferryl species to flip to close to the substrate in order to oxidise it. The flip could be accompanied by subsequent or concomitant release of carbon dioxide.[95]

1.3.5 The release of carbon dioxide

Following the formation of Fe(IV)=O intermediate, the next proposed mechanistic step is the release of carbon dioxide leading to the formation of a five coordinated Fe(IV)=O species (**VII** in Figure 1.3). Some studies suggest that the presence of an alcohol or a basic residue close to the 2OG C1 carboxylate can promote decarboxylation prior to substrate hydroxylation. For instance, crystallographic studies of TauD[48,87] suggests that the presence of an Arg residue involved in the interaction with the non-coordinating oxygen atom of C-1 of 2OG facilitates the release of CO_2.

1.3.6 Abstraction of Hydrogen from the substrate C-H bond

The highly oxidising Fe=O species abstracts a hydrogen atom from the inert C-H bond leading to the formation of a substrate radical, with concomitant reduction to Fe(III)-OH species (**VIII** in Figure 1.3).[86] The substrate radical then reacts with the hydroxyl radical attached to Fe(III) with concomitant reduction of the iron centre to Fe(II) (I**X** in Figure 1.3).

Of the two oxygen atoms of molecular oxygen, one is incorporated in succinate (at a complete or a very high level) and the other fully or partially incorporated into the prime substrate oxidised product (Figure 1.4).[69] The studies involving labelled oxygen showed that the oxygen incorporated into the substrate is at least partially derived from dioxygen, with partial exchange with water derived oxygen occurring in some cases. The exchange process

may occur at the ferryl stage, with the extent of exchange depending on the accessibility of the ferryl intermediate to water molecules.

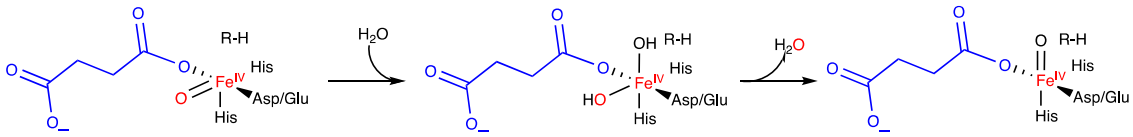


Figure 1.4 Schematic representations of the proposed "ferryl flip" in 2OG oxygenases catalysis.[5] Binding of a water molecule leads to the protonation of the Fe(IV)=O intermediate with subsequent release of a water molecule with one oxygen atom from dioxygen and concomitant formation of Fe(IV)=O intermediate again.

1.3.7 Release of product and succinate

There is some ambiguity regarding the sequence of release of the product and succinate from the Fe(II) complex. Some studies suggest that product leaves prior to succinate; however, some suggest the release of succinate prior to covalent reaction of the substrate.[98,99] The vacated sites are then occupied by water molecules, giving the initial octahedral complex which is ready to enter into a new catalytic cycle.

1.4 Nucleic Acid Demethylases

Genomes are exposed to various alkylating agents from both endogenous and exogenous sources. Endogenous sources include the common S_N2 type methylating agent *S*-adenosylmethionine (SAM). Exogenous sources include methylmethanesulfonate (MMS), which is present in cigarette smoke and *N*-methyl-*N*′-nitrosourea (MNU) (Figure 1.5A).[100] Methylation of DNA can occur at different positions on the nitrogenous bases and backbone depending on availability /susceptibility of the reaction site (Figure 1.5B).

Methylation is involved in crucial roles in biological signaling. Most methyltransferases use SAM as a methyl donor to modify DNA, RNA, and proteins, thus generating distinct biological methylations influencing the regulation of gene expression.[56,101] Methylation in

RNA plays roles in signalling and fine tuning. Though not all methylations are toxic, a few of them can be undesired leading to cell malfunction and cell death (apoptosis). Some of the lesions are toxic because they interfere with Watson-Crick type base pairing leading to instability of DNA.

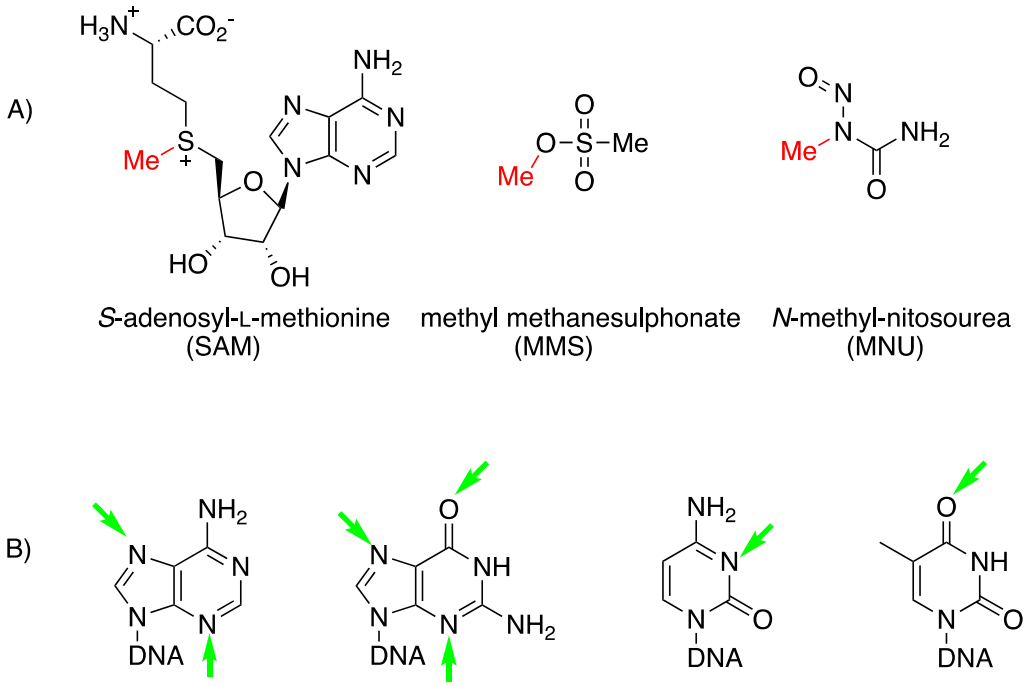


Figure 1.5 A) Chemical structure of S-adenosyl-L-methionine (SAM), methyl methanesulphonate (MMS), N-methyl-nitrosourea (MNU). The methyl groups reacting to alkylate DNA are highlighted in red. B) Common methylation sites of DNA bases.[100]

Organisms have evolved to maintain genome integrity and proper cellular functions by encoding for DNA/RNA-repair proteins that include DNA glycosylases in base-excision repair, O^6 -methylguanine methyltransferases in guanine methylation repair, and AlkB family proteins in direct oxidative repair.[102–105] Expression of the gene encoding for AlkB in *Escherichia coli* is induced as an adaptive response to MMS mild exposure which causes alkylation of DNA.[106] Homologues of AlkB are found in every form of life except archaea.[107–110] Apart from AlkB, the other three enzymes that function in the *E. coli* adaptive response to alkylation damage are Ada, AlkA, and AidB.[111]

The AlkB family of enzymes were first discovered as DNA-repair proteins that couple oxidative decarboxylation of 2OG into succinate and concomitant release of carbon dioxide along with the oxidative reversal of DNA methylation damage.[12,13,30,64,102–104,112] Although the methylation enzymes in DNA, proteins, and RNA were well characterised, studies regarding enzymes catalysing demethylation in mammals remained elusive for decades before the first report on the TET enzymes.[113]

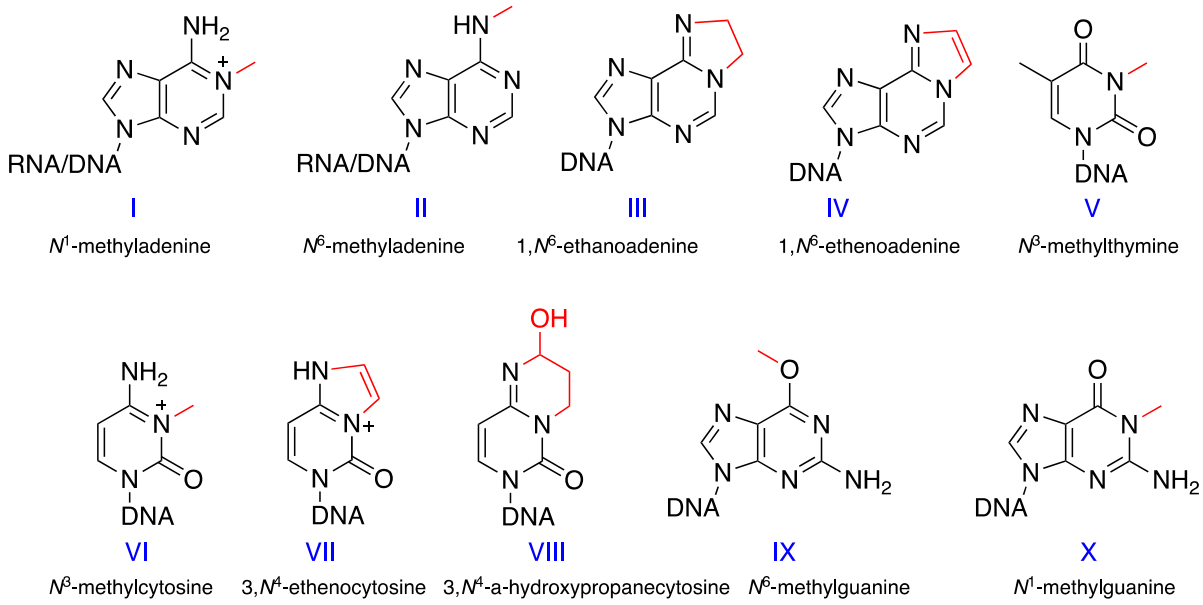


Figure 1.6 Chemical structures of the literature reported substrates of AlkB and the AlkB Human homologues.

1.4.1 AlkB

E. coli AlkB catalysis, like that of other 2OG oxygenases, employs non-haem Fe(II) and oxygen to catalyse the two electrons oxidation of 2OG and its substrates.[114] Like other 2OG dependent oxygenases, AlkB, and related enzymes contain a double-stranded ß-helix fold and a 2-His-1-carboxylate motif (His-131, Asp-133, and His-187) that bind the iron cofactor, which is chelated by 2OG. The 2OG C-5 carboxylate is involved in salt bridging with Arg-204. [115] AlkB has a structurally flexible nucleotide recognition lid consistent with the broad range of substrates accepted by the enzyme.[115] The range of substrates oxidised by AlkB

includes N^1-methyladenosine (m^1A, **I** in Figure 1.6),[12,38,116] N^6-methyladenosine (m^6A, **II** in Figure 1.6),[117] , N^3-methylthymine (m^3T,**V** in Figure 1.6),[116] N^6-ethenoadenine (εA, **IV** in Figure 1.6),[118] 3,N^4-ethenocytosine (εC, **VII** in Figure 1.6),[116] 1-methylguanine (m^1G, **X** in Figure 1.6),[116] 1,N^6-ethanoadenosine (**III** in Figure 1.6),[119] 3,N^4-a-hydroxypropanocytosine (**VII** in Figure 1.6).[116] AlkB has a preference for positively charged substrates via interactions with the negatively charged side chain of Asp135.[71] AlkB apparently has an electropositive binding groove with Thr51 to Tyr55, Ser129 and Lys 127, it was suggested that the groove would anchor the phosphodiester backbone or simply a phosphate group. In some studies, it was observed that at least a phosphate group attached to the substrate was necessary to observe any catalytic turnover, as the analogous nucleoside substrate only stimulated 2OG uncoupled turnover.[120] Crystallographic studies suggest that Trp69 and His131 are involved in π−π stacking with the nitrogenous base in the active site.[71] It has been observed that the coordinating oxygen of the C-1 carboxylate of 2OG coordinates to the distal His i.e. His-187 leading to the binding of oxygen away from the substrate.[71] Thus, the resultant ferryl species Fe=O would need to flip to come close to the substrate inert C-H bond to be oxidised.

A modest non-catalytic 2OG turnover to succinate by AlkB has been observed in the absence of substrate.[96] It was observed that the Fe=O species produced in the absence of substrate, hydroxylates the side chain of Trp 178 of AlkB, thus catalysing an irreversible enzyme modification.[96] However, the production of succinate increases significantly upon addition of methylated substrate, suggesting conformational changes on binding of substrate leading the protein to be catalytically efficient.[120]

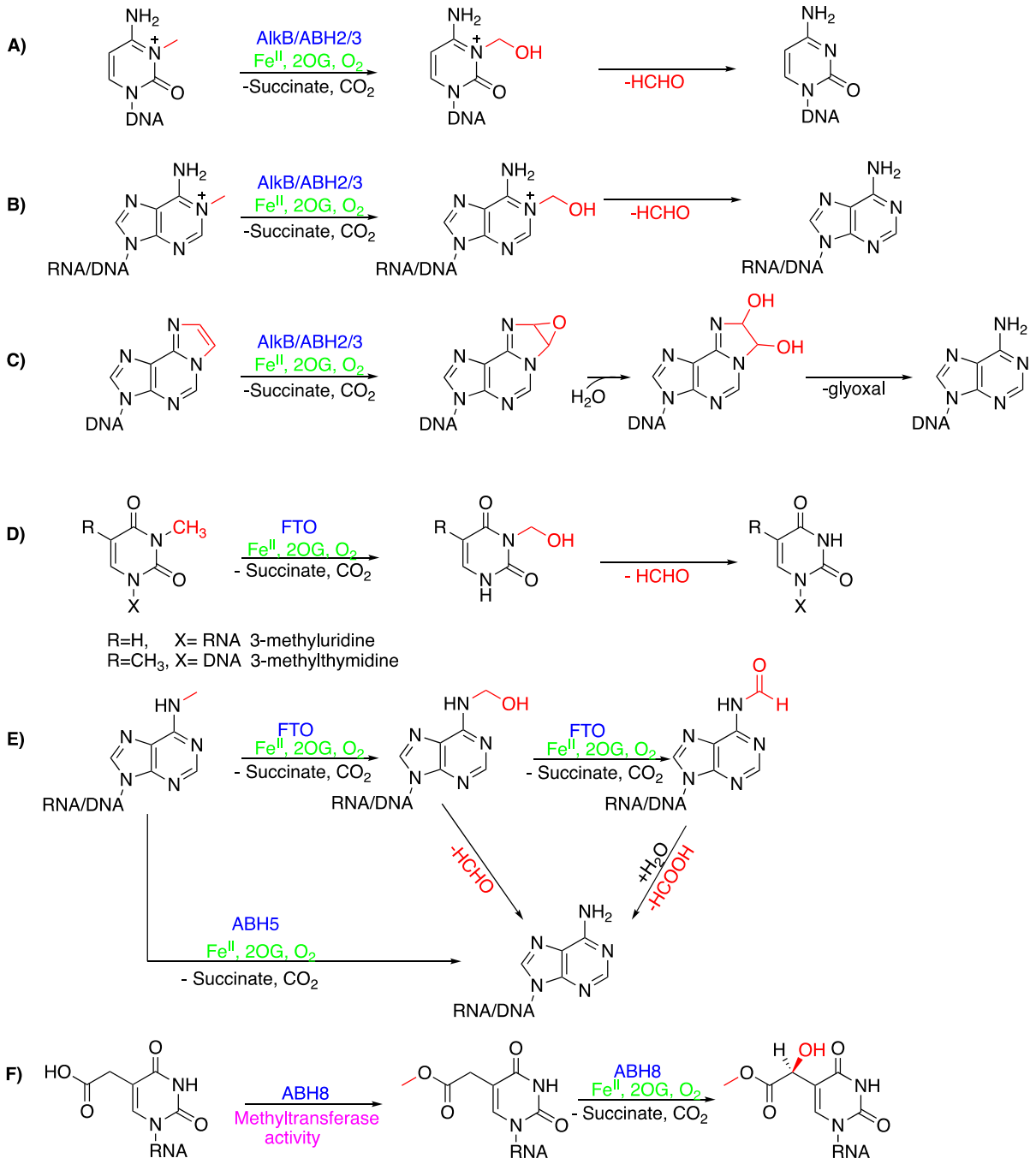


Figure 1.7 Schematic representation of the oxidative demethylation of A) m^3C catalysed by AlkB/ABH2/3; B) m^1A catalysed by AlkB/ABH2/3; C) εA catalysed by AlkB/ABH2/3; D) m^3U/ m^3T catalysed by FTO; E) m^6A catalysed by FTO and ABH5. It should be noted that FTO is involved in the sequential oxidation of m^6A to N^6-hydromethyladenosine (hm^6A) to N^6-formyladenosine (f^6A) to adenosine (A), unlike ABH5 which is supposed to catalyse the degradation of hm^6A to A; F) Schematic representation of the methylation of tRNA to form 5-methoxycarbonylmethyluridine (mcm^5U) and hydroxylation of mcm^5U catalysed by ABH8. The Fe(II) cofactor and the co-substrates are highlighted in green; enzymes catalysing the reaction are highlighted in blue; the groups undergoing oxidation are highlighted in red.

1.4.2 AlkB Human Homologue proteins

To date, nine homologues of AlkB, namely AlkBHs 1-8 (commonly called ABH 1-8) and AlkB 9 (commonly known as FTO), have been found in mammals. Five of these ABHs (ABH1-3, ABH5, and FTO) have shown measurable demethylation activity towards nucleic acids, however, the range of substrates for these enzymes is narrow as compared to AlkB.[9,14,40,62,115,121–123]The active sites of these enzymes, like other 2OG oxygenases, is supported by the characteristic DSBH fold that binds Fe(II) and 2OG using conserved residues.[62,63,65,67,71,115,124,125] Similarly to other 2OG oxygenases, with the AlkB Human homologues cleavage of molecular oxygen generates a highly reactive Fe=O species that abstracts a hydrogen atom from a substrate C-H bond, subsequently leading to the oxidation of substrate in a mechanism proposed to involve radical rebound type process.[80]

1.4.2.1 ABH1

ABH1 is widely expressed in human tissues, but its biological role is still unclear.[40,45] ABH1 has a sequence homology with AlkB,[110] and was initially proposed to behave similarly to AlkB. However, the purified ABH1 gene product did not show any activity towards nucleic acids.[39] Later, ABH1 was reported to have potential demethylation activity towards m^1A and m^3T (**I** and **V** in Figure 1.6); however, the activity was not direct.[40] One *in vitro* study, has suggested that ABH1 has demethylation activity towards 3-methylcytidine (m^3C, **VI** in Figure 1.6). A more recent study suggested ABH1 is a histone dioxygenase exclusively acting on histone H2A, which is predicted to have a functional role in neuron differentiation.[45] Another study detected lyase activity of ABH1 at abasic sites independent of the presence of iron or 2OG, in vitro.[110] However, definitive knowledge of the reaction(s) catalysed by ABH1 is not available, and the biological significance of the ABH1 demethylation activity (if any) is not understood.

1.4.2.2 ABH2

Of the human ABHs, ABH2 and ABH3 most closely resemble the *E*.coli AlKB. ABH2 and ABH3 are the bonafide DNA repair enzymes that preserve genome integrity in mammals (Figure 1.7 A, B and C).[39] ABH2 shows robust demethylation activity towards cytotoxic m^1A and m^3C (**I** and **VI,** respectively, in Figure 1.6) in vitro and vivo and reduced activity towards m^3T (**V** in Figure 1.6) in ds DNA.[30,64] ABH2 also repairs εA (**IV** in Figure 1.6) and εC (**VII** in Figure 1.6).[123] ABH2 acts on both ds DNA and ss-DNA; preferring the ds-DNA compared to ABH3. ABH2 has two extra loops interacting with the opposite strand of the DNA, supporting its substrate preference for dsDNA.[115] Various crystal structures have been solved for ABH2 providing insight into the binding pockets of 2OG, substrate and the Fe(II) site.[115] Five residues of ABH2 interact with the substrate in contrast to the two residues employed by AlkB; ABH2 uses a finger residue, Phe-102, to intercalate into the DNA, unlike AlkB which lacks an analogous finger residue, while repairing the methylated base by flipping the base residue.[115] ABH2 detects the weakening of the base pairing by alkylation of residues as a search tool for its substrates.[124,126] ABH2 is reported to be the primary oxidative DNA demethylase present in mouse cells.[127]

1.4.2.3 ABH3

Similarly to ABH2, ABH3 also functions as a DNA repair enzyme in mammalian cells. ABH3 prefers ss DNA and ssRNA as a substrate unlike ABH2 which prefers ds DNA.[125] ABH3 catalyses the oxidative demethylation of m^1A and m^1C along with m^3T and εA, (**I**, **VI**, **V** and **IV**, respectively, in Figure 1.6) albeit with lower efficiency (Figure 1.7A, B and C).[39,128] However, the active site of ABH3 is more polar as compared to the hydrophobic core of AlkB or ABH2, thus narrowing the substrate specificity.[125] Arg-122 is predicted to

intercalate into DNA in contrast to Phe-102 in ABH2. ABH3 is highly expressed in prostate cancer cells and thus has been proposed to be a therapeutic target for prostate cancer.[129]

1.4.2.4 ABH4:

ABH4 has been proposed to catalyse the oxidative decarboxylation of α-KG; however, to date, ABH4 has failed to show any demethylation activity towards nucleic acids as substrates.[30] ABH4 has been reported to be involved in cell division.[130]

1.4.2.5 ABH5

ABH5 is an RNA demethylase showing robust demethylation activity towards m^6A (marked as **II** in Figure 1.6) (Figure 1.7E).[42] The demethylation activity of ABH5 is dependent on the sequence, with the consensus sequence [GGACU] being demethylated with the highest efficiency.[42] ABH5 prefers to demethylate ss-RNA rather than ds-RNA.[42] The presence of a distinctive loop in the substrate binding pocket of ABH5 is proposed to confer its selectivity towards single stranded RNA over double stranded RNA strands.[93] Studies have suggested the involvement of ABH5 in spermatogenesis in mice.[42] ABH5 knock out mice display impaired male fertility resulting from compromised spermatogenesis, aberrant apoptosis, and altered gene expression in the testis.[42]

1.4.2.6 ABH6

ABH6 is highly expressed in the testis and pancreas and is localised throughout the cell.[131] No function for ABH6 has been reported in the literature to date.

1.4.2.7 ABH7

ABH7 is a mitochondrial protein.[132,133] ABH7 has not shown any observed oxidative repair activity towards nucleic acids as a substrate.[130] However, ABH7 is observed to trigger the collapse of the mitochondrial membrane.[133] ABH7 is also reported to initiate the loss of mitochondria leading to the depletion of energy and cell death.[133] Interestingly, mice in which ABH7 expressed levels are reduced, gain body weight and have elevated body fat suggesting a role of ABH7 in fatty acid metabolism.[134]

1.4.2.8 ABH8

ABH8 is classified as a hydroxylase rather than a demethylases unlike the other members of the AlkB family and its human analogues.[135] ABH8 is the only enzyme identified in the AlkB 2OG oxygenases subfamily to be a tRNA modifying enzyme.[135] Apart from the AlkB domain, ABH8 also contains a RNA recognition motif (RRM) as a N-terminal, and a Trm9-like methyltransferase domain as a C-terminal domain.[65] ALKBH8 is proposed to be involved in bladder cancer progression by increasing the production of reactive oxygen species.[136] The Fe(II)-binding site in ABH8 is proposed to be exposed to solvent leading to the suggestion that uncoupled turnover of 2OG could be more for ABH8 as compared to AlkB. However, lower uncoupled oxidation of 2OG has been observed in ABH8 compared to AlkB, possibly due to an inactive orientation of the 2OG and Arg 334 in the active site in the absence of substrate; changes in conformation upon binding to tRNA may be required in order to be catalytically active.[65] ABH8 catalyses tRNA hydroxylation (Figure 1.7F) which could alter the codon selection during translation.[66] The knockdown of ABH8 leads to a defect in cardiac development in D. melanogaster[137] and can either result in embryonic death or affect sterility in animals surviving to adulthood when an internal deletion in the gene encoding ABH8 is done as observed in C. elegans.[138]

1.4.2.9 Fat Mass and Obesity Associated Protein (FTO)

The "*fto*" gene, first identified in genome-wide association studies (GWAS), encodes for a Fe(II) and 2OG dependent oxygenases called FTO, which belongs to the sub family of nucleic acid demethylases.[10,139] The highest expression levels of FTO in mice have been observed in brain tissues.[10] Studies on mouse models suggested that FTO is involved in energy homeostasis and whole body metabolism.[140] A single mutation in the FTO protein in mouse resulted in a lean phenotype. However, the underlying mechanism relating FTO to body mass is still unknown. FTO has been linked to obesity,[139–142] type 2 diabetes, food intake,[143] development,[144] breast cancer,[145] and Alzheimer's disease.[103] Mutation of an amino acid present in the FTO active site impairs its catalytic activity and correlates with human postnatal growth retardation, facial dysmorphism, and certain brain malfunctions.[144]

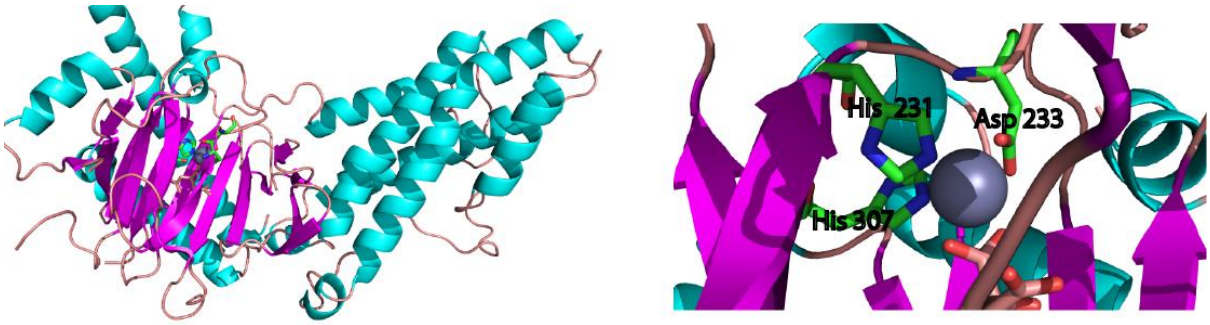


Figure 1.8 A) View from a crystal structure of FTO.Zn(II) NOG(PDB ID 41E7). B). The active site (with Fe replaced by zinc in the crystal structure) of FTO.

FTO and ABH5 were the first RNA demethylases to be identified.[10,42] FTO also shows relatively weak demethylation activity towards m^3T in ssDNA and m^3U in ssRNA.[9] Subsequent biochemical studies revealed that FTO has a higher activity on ssRNA than on ssDNA.[67] In 2011, FTO was reported to catalyse relatively efficient demethylation of m^6A in ss RNA.[121] It was also observed that the overexpression of FTO in cells reduces m^6A levels in mRNA. m^6A is found at a low-level in genomic DNA of higher eukaryotes;[146] however,

m^6A is a prevalent modification present in mRNA. Later, it was observed that FTO oxidises m^6A to N^6-hydroxymethyladenosine (hm^6A) which can be further oxidised to N^6-formyladenosine (f^6A).[44] Molecular dynamics studies suggest that FTO binds to hm^6A in a similar manner to m^6A. Recently, FTO has been reported to catalyse demethylation with a preference for N^6,O^2-dimethyladenosine (m^6Am) over m^6A.[147]

A crystal structure of FTO reveals that FTO has the characteristic AlkB family N-terminal fold consisting of a distorted DSBH fold and the C-terminal with a novel fold comprising mostly α-helices.[67] It was observed that the C-terminal region makes extensive interaction with the N-terminal; the interaction was found to be necessary for FTO to be catalytically active. The FTO active site contains the conserved Fe(II) binding residues (2-His-1-carboxylate) and 2OG binding residues; Fe(II) is coordinated to His 231, Asp 233 and His 307 and the C5 carboxylate of the 2OG forms salt bonds with Arg 316 and Arg 322.[67] The crystals structure of FTO also reveals the existence of an extra loop, surrounding one side of the jelly roll motif, which competes with the unmethylated strand of the ds-DNA and ds-RNA for binding to FTO, thus, supporting the observed FTO preference for ss-DNA and ss-RNA as a substrate.[67]

1.4.2.10 TET enzymes:

The Ten Eleven Translocation 1-3 enzymes (TET1-3) were the first mammalian proteins to be identified as DNA oxygenases acting on 5-hydroxymethylcytidine (5hmC).[113] The TET1-3 proteins belong to the AlkB-like Fe/2OG-dependent oxygenases subfamily and catalyses the sequential oxidation of 5-methylCytidine (5mC) to generate 5-hydroxymethylcytidine (5hmC), 5-formylcytidine (5fC) and 5-carboxylcytidine (5caC).[32,113] The three human TET enzymes are conserved in all jawed vertebrates.[148] Similarly to other 2OG oxygenases, TET1-3 contain a catalytic domain consisting of a DSBH fold, a conserved Fe(II) binding

motif (2-His-1-carboxylate) and 2OG binding residues.[148] Additionally, TET proteins contain a characteristic Cys rich domain adjacent to their DSBH fold which is observed to be involved in substrate recognition and binding.[148,149] The Cys rich domain has been found essential for the catalytic activity of the TET proteins.[149] The TET enzymes are reported to play roles in preimplantation development,[150] gene transcription,[151] suppressing breast cancer,[152] and colon cancer growth.[153] A relatively lowered level of 5hmC has been linked to different types of cancers including skin,[154] breasts, myeloid,[155] lung[156] and prostate cancer.[157]

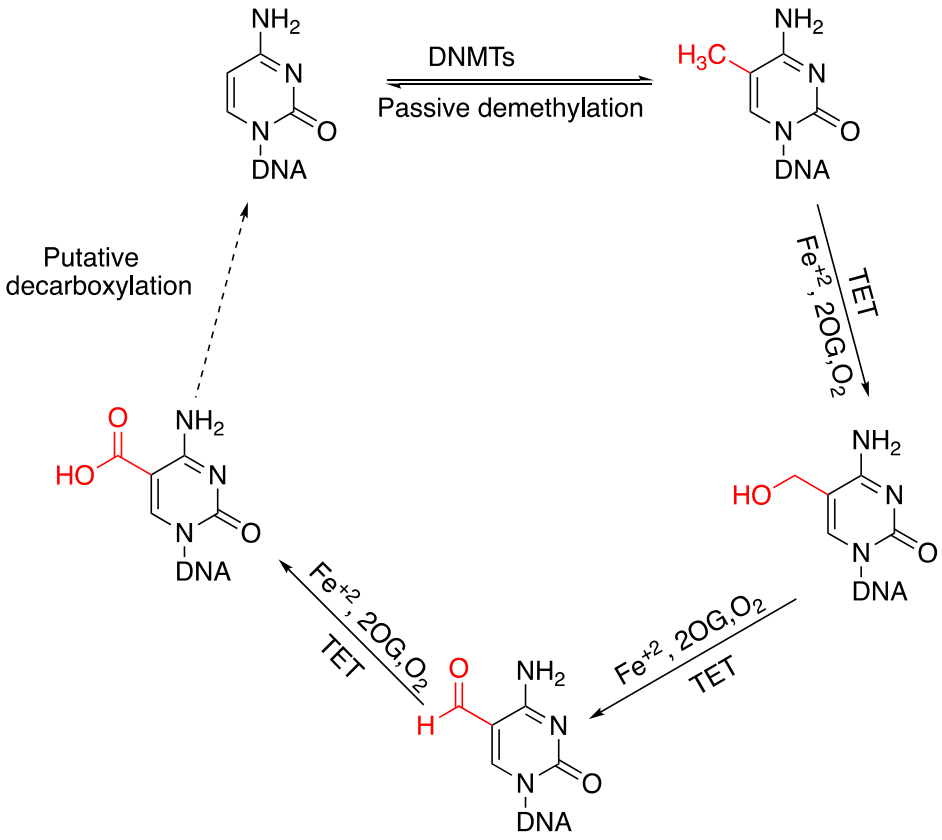


Figure 1.9 Schematic representation of the oxidative demethylation of cytosine in DNA. The modifications are highlighted in red.

Table 1.1 Table showing the reported substrates of AlkB and the AlkB human homologues. Molecules highlighted in green are apparently poor substrates.

Enzymes	Substrate	Functions	Location
AlkB	m^1A; m^6A; 1,N^6-ethenoadenine, 3,N^3-ethenocytosine, m^6G, m^1G, 1,N^6-ethanoadenosine, 3,N^4-a-hydroxypropanocytosine, N^{2-}alkylguanine	DNA repair	ssDNA
ABH1	me^3C, H2A	Unknown roles for demethylases and AP lyase, histone dioxygenases	dsDNA
ABH2	m^1A, m^3T, ethenoA, ethenoC	DNA repair	
ABH3	m^1A, m^3C, m^3T, 1, N^6-ethenoadenine		ssDNA, rna
ABH4	Actin	Cell division	
ABH5	m^1A, m^6A,	RNA demethylation	RNA
ABH6	Unknown	Unknown	
ABH7	Unknown	Obesity	
ABH8	Cm^5U	Translation	tRNA
FTO (ABH9)	m^3T; m^3C; m^6A; m^6Am	RNA demethylation	RNA, ssDNA
TET1-TET3	hm^5C; m^5C; f^5C; ca^5C	Gene regulation	DNA

1.5 RNA epigenetics:

According to the central dogma of molecular biology, DNA holds the key place in information storage; consequently, compared to RNA, DNA has perhaps garnered more scientific attention. However, in retrospect, the discovery of pseudouridine as a major component of transfer RNA about 60 years ago opened up the new era for RNA post transcriptional modifications.[158] The discovery of the catalytic RNAs has also helped to shift the spotlight to RNA research which is now a fast growing research field.

More than 100 structurally different modifications to various forms of RNA, including ribosomal RNA (rRNA), transfer RNA (tRNA), messenger RNA (mRNA), long non-coding RNA (lncRNA) have been reported.[159] Many modifications involve methylation of bases and 2'-hydroxyl group of RNA nucleotides;[159] others are complex modifications being carried out by enzymes but the functional and biological consequences of most of these modifications are still unknown.

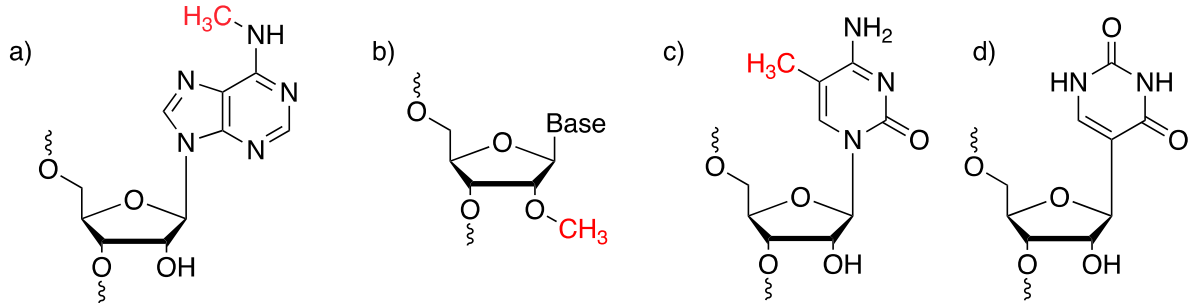


Figure.1.10 Chemical structure of general modifications found in internal messenger/ long noncoding/ ribosomal RNA. a) N^6-methyladenosine b) O^2-methylbase c) C^5-methylcytosine d) pseudouridine. Modifications are highlighted in red, except for pseudouridine.

Although the modifications in DNA and histones have been extensively studied,[12,23,29,32,33,35,38,39,56,104,111,112,118,149,154,160–163] modifications to RNA (particularly, mRNA) were not well studied until recently.[10,30,31,164–174] The most prevalent messenger RNA modification is m^6A which is reported to be present at almost 3 sites per mRNA molecule.[175–

[177] m^6A is installed by multi-component methyltransferase [METTL3/METTL14][178] with the help of the regulatory protein WTAP (Wilms' tumor 1-associating protein) in mammalian cells.[179] Both METTL3 and METTL14 are reported to transfer a methyl group from the cofactor *S*-adenosyl methionine (SAM) to GGAC(U)U sequences, however, METTL14 catalyses the transfer with less efficiency than that of the complex.[178,179]

In mRNA, m^6A modified bases are often found near stop codons[180] and the noted conserved RRACH (R = A, G; H = A, C, U) sequence motif which suggests m^6A be a marker in RNA biology. The biological role of m^6A is significant as small inhibitory RNA (siRNA) knockdown of METTL3 leads to cell death.[181] m^6A is proposed to play roles in mRNA splicing, transport, stability, and immune tolerance.[169,182] The discovery of m^6A as a substrate for FTO[121] and ABH5,[42] increased the 'lost' attention to the RNA modifications, helping to open up the 'new' field of RNA epigenetics.

Proteins involved in epigenetic regulation, are often characterised as "writers" (catalysing a specific type of modification), "readers" (recognising the modified sites), and "erasers" (removing the specific type of modification). For m^6A, the three kinds of proteins have been identified, but for most other RNA modifications, all of the three categories of proteins are yet to be identified. In the case of m^6A, METTL3, METTL14 are writers,[178,179] FTO and ABH5 are erasers,[10,42] and YTHDF1 and YTHDF2 are readers.[183] The YTH family of proteins were identified as readers by m^6A affinity chromatography using methylated RNA as a bait.[184] Mammals encode five YTH proteins namely YTH Domain family(YTHDF) 1, 2, 3 and YTH Domain Containing proteins 1 and 2.[183] As thousands of m^6A sites have been identified, there is a strong possibility that other readers and erasers specific to m^6A exist.

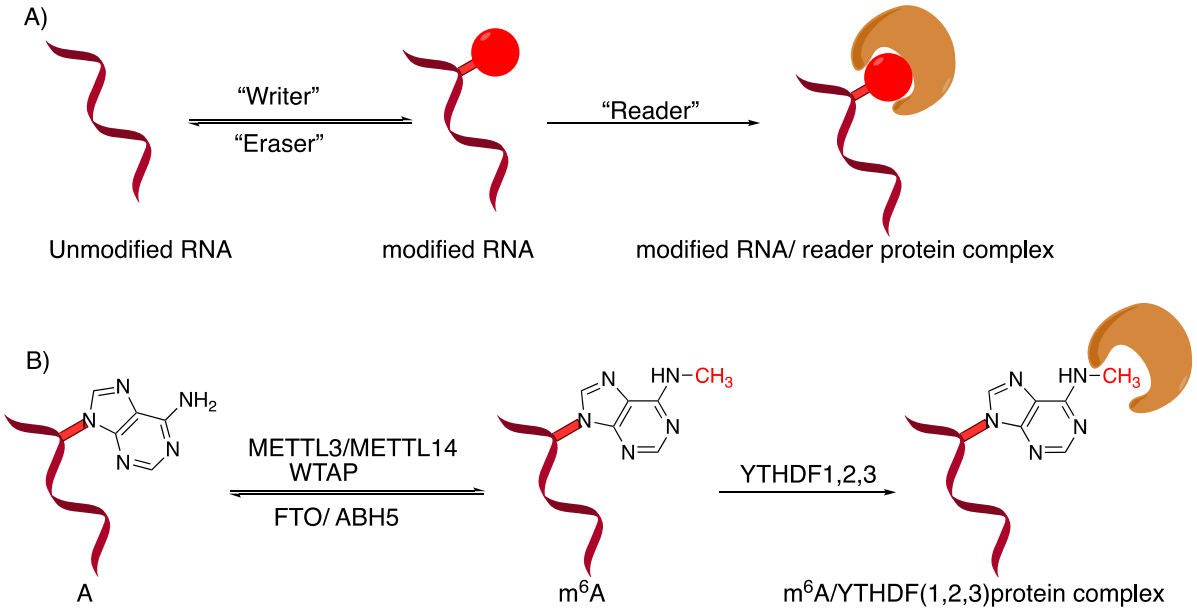


Figure 1.11 A) Schematic plot for general RNA modification. Writer (catalysing modification); Eraser (removing modification); Reader (identifying modification) B) Writer (METTL3/METTL14), Eraser (FTO/ABH5), Reader (YTHDF 1,2,3) proteins for m^{6}A

m^{6}A is involved in a variety of biological process of high importance and has been linked to many diseases (Table 1.2 and references therein). The "writer" proteins of m^{6}A are involved in cell development and changes in their behaviour have been associated with cell death.[181,185] The regulatory protein WTAP plays roles in RNA splicing and is associated with diseases such as Wilms' tumour, kidney cancer, and others.[186,187] FTO has been associated with obesity,[10,139,142] type II diabetes,[188] Alzheimer's[189] and other diseases. ABH5 loss or mutation has been found to affect spermatogenesis.[42] Some of the diseases related to m^{6}A levels are listed in Table 1.2.

Table 1.2 Table showing the proteins associated with the regulation of m^{6}A and associated diseases.

Gene	Function	Phenotype	Associated disease
METTL3	Methyltransferase	Apoptosis,[185] development[190]	Prostatitis, Aicardi syndrome
METTL14	Methyltransferase	Apoptosis,[185] development[190]	Alcohol dependence, alcoholism
WTAP	Scaffolding or localisation[170]	mRNA splicing[187]	Wilms' tumour, hypospadias,[186] sarcoma, malignant mesothelioma, synovial sarcoma
FTO	Nucleic acid demethylase[121]	Cellular energy homeostasis[10]	Obesity[10], diabetes,[188] polycystic ovary syndrome, heart attack, cancer,[191] alcoholism, mental disorders,[143] cataracts, hepatitis
ALKBH5	Nucleic acid demethylase[42]	Male fertility[42]	Hypoxia,[14] Smith-Magenis syndrome
YTHDF1	m^{6}A Binding[184]	mRNA stability[192]	Pancreatic cancer, [193] pancreatitides,[193] dermatomyositis
YTHDF2	m^{6}A Binding[184]	mRNA stability[192]	Leukaemia, renal cell carcinoma, breast cancer[193]
YTHDF3	m^{6}A Binding[184]	n/a	n/a

Most RNA modifications, in principle, are reversible. Apart from m^6A, other candidates, very likely for reversal modifications include m^1A, m^5C, and m^1G.

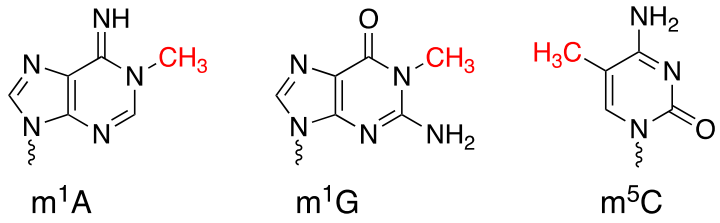


Figure 1.12 Potential other epigenetic markers in RNA. The methylated modifications are highlighted in red.

1.5.1 Other potential epigenetic modifications

Several other nucleobase modifications are proposed to be involved in epigenetic regulation. m^1A is present in almost all tRNAs and is involved in stabilising tRNA structure.[172] In particular, the presence of m^1A is conserved at A58 in most eukaryotic tRNA molecules.[172] m^1A can be oxidatively demethylated by AlkB like enzymes in humans (Figure 1.7 B).[12,116] m^1G also confers tRNA stability.[194] m^5C is proposed as an epigenetic marker in DNA. As described above, the methyl group of m^5C in DNA can be oxidised sequentially to hydroxymethyl, formyl, and carboxylate by the TET enzymes (Figure 1.9);[32] however, the activity of the TET enzymes is not well studied in the context of RNA. m^5C is present in rRNA, tRNA, and mRNA and is more prevalent in nontranslated regions and relatively scarce in coding sequences.[159,195] Pseudouridine is generated by isomerisation of uridine by pseudouridine synthase enzymes.[196] Studies on methylating enzymes as potential "writer" proteins for this modification are of interest. Pseudouridine is present in tRNA, rRNA, and mRNA.[159,195] Its biological function is still unknown but it could play role in altering secondary structure due to additional H-bonding enabled by the rearrangement (Figure 1.13).[196]

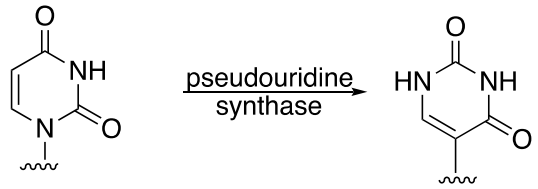


Figure 1.13 Schematic representation of the rearrangement of uridine to pseudouridine as catalysed by pseudouridine synthase.

1.6 References

1. Loenarz, C. & Schofield, C. J. Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat. Chem. Biol.* **4,** 152–156 (2008).
2. Kelly, D. J. & Hughes, N. J. *The Citric Acid Cycle and Fatty Acid Biosynthesis*. *Helicobacter pylori: Physiology and Genetics* (ASM Press, 2001).
3. Hutton, J. J., Kaplan, A. & Udenfriend, S. Conversion of the amino acid sequence gly-pro-pro in protein to gly-pro-hyp by collagen proline hydroxylase. *Arch. Biochem. Biophys.* **121,** 384–91 (1967).
4. Flashman, E. & Schofield, C. J. The most versatile of all reactive intermediates? *Nat. Chem. Biol.* **3,** 86–87 (2007).
5. McDonough, M. A., Loenarz, C., Chowdhury, R., Clifton, I. J. & Schofield, C. J. Structural studies on human 2-oxoglutarate dependent oxygenases. *Curr. Opin. Struct. Biol.* **20,** 659–672 (2010).
6. Schofield, C. J. & Zhang, Z. Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr. Opin. Struct. Biol.* **9,** 722–731 (1999).
7. Valegård, K., Terwisscha Van Scheltinga, A. C., Dubus, A., Ranghino, G.,Öster, L. M., Hajdu, J., Andersson, I. The structural basis of cephalosporin formation in a mononuclear ferrous enzyme. *Nat. Struct. Mol. Biol.* **11,** 95–101 (2004).
8. Kershaw, N. J., Caines, M. E. C., Sleeman, M. C. & Schofield, C. J. The enzymology of clavam and carbapenem biosynthesis. *Chem. Commun.* **256,** 4251 (2005).
9. Jia, G. Yang, C. G.,Yang, S., Jian, X., Yi, C., Zhou, Z., He, C. Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Lett.* **582,** 3313–3319 (2008).
10. Gerken, T., Girard, C. A., Tung, Y. C. L.,Webby, C. J., Saudek, V., Hewitson, K. S., Yeo, G. S.H., McDonough, M. A., Cunliffe, S., McNeill, L. A., Galvanovskis, J., Rorsman, P., Robins, P., Prieur, X., Coll, A. P., Ma, M., Jovanovic, Z., Farooqi, I. S., Sedgwick, B., Barroso, I., Lindahl, T., Ponting, C. P., Ashcroft, F. M., O'Rahilly, S., Schofield, C. J. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science (80).* **318,** 1469–1472 (2007).
11. Yi, C., Yang, C.-G. & He, C. A non-heme iron-mediated chemical demethylation in DNA and RNA. *Acc. Chem. Res.* **42,** 519–29 (2009).
12. Falnes, P. Ø., Johansen, R. F. & Seeberg, E. AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature* **419,** 178–182 (2002).
13. Aravind, L. & Koonin, E. V. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate-and iron-dependent dioxygenases. *Genome Biol.* **2,** 7–1 (2001).
14. Thalhammer, A., Bencokova, Z.,Poole, R., Loenarz, C., Adam, J., O'Flaherty, L., Schödel, J., Mole, D., Giaslakiotis, K., Schofield, C. J., Hammond, E. M., Ratcliffe, P. J.,Pollard, P. J. Human AlkB homologue 5 is a nuclear 2-oxoglutarate dependent oxygenase and a direct target of hypoxia-inducible factor 1α (HIF-1α). *PLoS One* **6,** (2011).
15. Hamed, R. B., Gomez-Castellanos, J. R., Henry, L.,Ducho, C., McDonough, M. A., Schofield, C.J. The enzymes of β-lactam biosynthesis. *Nat. Prod. Rep.* **30,** 21–107 (2013).
16. Zhou, J., Kelly, W. L., Bachmann, B. O., Gunsior, M., Townsend, C. A., Solomon, E. I. Spectroscopic studies of substrate interactions with clavaminate synthase 2, a multifunctional α-KG-dependent non-heme iron enzyme: Correlation with mechanisms and reactivities. *J. Am. Chem. Soc.* **123,** 7388–7398 (2001).
17. Hollenhorst, M. A., Bumpus, S. B., Matthews, M. L., Bollinger, J. M., Kelleher, N. L.,

Walsh, C.T. The nonribosomal peptide synthetase enzyme DdaD tethers N ??-fumaramoyl-l -2,3-diaminopropionate for Fe(II)/ α- ketoglutarate-dependent epoxidation by DdaC during dapdiamide antibiotic biosynthesis. *J. Am. Chem. Soc.* **132,** 15773–15781 (2010).

18. Frédéric H. Vaillancourt, , Ellen Yeh , David A. Vosburg, Sylvie Garneau-Tsodikova, and & Christopher T. Walsh. Nature's Inventory of Halogenation Catalysts: Oxidative Strategies Predominate. (2006).
19. Blasiak, L. C. & Drennan, C. L. Structural Perspective on Enzymatic Halogenation. *Acc. Chem. Res.* **42,** 147–155 (2009).
20. Butler, A. & Sandy, M. Mechanistic considerations of halogenating enzymes. *Nature* **460,** 848–854 (2009).
21. Loenarz, C. & Schofield, C. J. Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem. Sci.* **36,** 7–18 (2011).
22. Johnson, B. M., Stier, B. A. & Caltabiano, S. Effect of food and gemfibrozil on the pharmacokinetics of the novel prolyl hydroxylase inhibitor GSK1278863. *Clin. Pharmacol. Drug Dev.* **3,** 109–117 (2014).
23. Thinnes, C. C.,England, K. S., Kawamura, A., Chowdhury, R., Schofield, C.J., Hopkinson, R. J. Targeting histone lysine demethylases — Progress, challenges, and he future. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839,** 1416–1432 (2014).
24. Rose, N. R., McDonough, M. A., King, O. N. F., Kawamura, A. & Schofield, C. J. Inhibition of 2-oxoglutarate dependent oxygenases. *Chem. Soc. Rev.* **40,** 4364 (2011).
25. Yan, L., Colandrea, V. J. & Hale, J. J. Prolyl hydroxylase domain-containing protein inhibitors as stabilizers of hypoxia-inducible factor: small molecule-based therapeutics for anemia. *Expert Opin. Ther. Pat.* **20,** 1219–1245 (2010).
26. Thalhammer, A., Aik, W., Bagg, E. A. L. & Schofield, C. J. The potential of 2-oxoglutarate oxygenases acting on nucleic acids as therapeutic targets. *Drug Discov. Today Ther. Strateg.* **9,** 91–100 (2012).
27. Malfait, F. & de Paepe, A. Molecular genetics inclassic Ehlers-Danlos syndrome. *Am. J. Med. Genet. Part C Semin. Med. Genet.* **139C,** 17–23 (2005).
28. McDonough, M. A., Kavanagh, K. L., Butler, D., Searls, T., Oppermann, U., Schofield, C. J.Structure of Human Phytanoyl-CoA 2-Hydroxylase Identifies Molecular Mechanisms of Refsum Disease. *J. Biol. Chem.* **280,** 41101–41110 (2005).
29. Branco, M. R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13,** 7 (2011).
30. Lee, D., Dong H. J., Seung G, Cai, S., Chen, Y., Pfeifer, G. P., O'Connor, T. R. Repair of methylation damage in DNA and RNA by mammalian AlkB homologues. *J. Biol. Chem.* **280,** 39448–39459 (2005).
31. Grosjean, H. *Fine-Tuning of RNA Functions by Modification and Editing*.
32. Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A.,He, C., Zhang, Y. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science (80).* **333,** 1300–1303 (2011).
33. Freudenberg, J. M. Ghosh, S., Lackford, B. L., Yellaboina, S., Zheng, X., Li, R., Cuddapah, S., Wade, P. A., Hu, G., Jothi, R. Acute depletion of Tet1-dependent 5-hydroxymethylcytosine levels impairs LIF/Stat3 signaling and results in loss of embryonic stem cell identity. *Nucleic Acids Res.* **40,** 3364–3377 (2012).
34. Ng, S. S., Kavanagh, K. L., McDonough, M. A., Butler, D., Pilka, E. S., Lienard, B.M. R., Bray, J. E., Savitsky, P., Gileadi, O., Von Delft, F., Rose, N. R.,Offer, J., Scheinost, J. C. Borowski, T.,Sundstrom, M., Schofield, C. J.,Oppermann, U. Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **448,** 87–91

(2007).
35. Cloos, P. A. C., Christensen, J., Agger, K. & Helin, K. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. doi:10.1101/gad.1652908
36. Lando, D., Peet, D. J., Whelan, D. A., Gorman, J. J. & Whitelaw, M. L. Asparagine Hydroxylation of the HIF Transactivation Domain: A Hypoxic Switch. *Science (80-.).* **295,** 858–861 (2002).
37. Ivan, M. HIFalpha Targeted for VHL-Mediated Destruction by Proline Hydroxylation: Implications for O2 Sensing. *Science (80).* **292,** 464–468 (2001).
38. Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature* **419,** 174–178 (2002).
39. Duncan, T. T., Sarah C., Koivisto, P., Bates, P., Lindahl, T.,Sedgwick, B. Reversal of DNA alkylation damage by two human dioxygenases. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 16660–16665 (2002).
40. Westbye, M. P. Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *J. Biol. Chem.* **283,** 25046–25056 (2008).
41. Li, M.-M. Nilsen, A.,Shi, Y., Fusser, M., Ding, Y. H., Fu, Y., Liu, B., Niu,Y.,Wu, Yong S., Huang, C. M., Olofsson, M., Jin, K. X., Lv, Y., Xu, X. Z., He, C., Dong, M. Q., Rendtlew Danielsen, J. M., Klungland, A. Yang, Y.G. ALKBH4-dependent demethylation of actin regulates actomyosin dynamics. *Nat. Commun.* **4,** (2013).
42. Zheng, G. ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Mol. Cell* **49,** (2013).
43. Fu, D., Jordan, J. J. & Samson, L. D. Human ALKBH7 is required for alkylation and oxidation-induced programmed necrosis. *Genes Dev.* **27,** (2013).
44. Fu, Y. Jia, Guifang, Pang, X., Wang, R. N., Wang, X., Li, C. J., Smemo, S., Dai, Q., Bailey, K. A., Nobrega, M. A., Han, K. L., Cui, Q., He, C. FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nat. Commun.* **4,** 1798 (2013).
45. Ougland, R. Ougland, R., Lando, D., Jonson, I., Dahl, J. A., Moen, M. N., Nordstrand, L. M., Rognes, T., Lee, J. T., Klungland, A., Kouzarides, T., Larsen, E. ALKBH1 is a histone H2A dioxygenase involved in neural differentiation. *Stem Cells* **30,** 2672–2682 (2012).
46. Hewitson, K. ., Granatino, N., Welford, R. W. ., McDonough, M. . & Schofield, C. . Oxidation by 2-oxoglutarate oxygenases: non-haem iron systems in catalysis and signalling. *Philos. Trans. R. Soc. London A Math. Phys. Eng. Sci.* **363,** (2005).
47. Clifton, I. J., Hsueh, L.-C., Baldwin, J. E., Harlos, K. & Schofield, C. J. Structure of proline 3-hydroxylase. *Eur. J. Biochem.* **268,** 6625–6636 (2001).
48. Jonathan M. Elkins. X-ray Crystal Structure of Escherichia coli Taurine/α-Ketoglutarate Dioxygenase Complexed to Ferrous Iron and Substrates. (2002).
49. Harlos, K. Structural origins of the selectivity of the trifunctional oxygenaseclavaminic acid synthase. *Nat. Struct. Biol.* **7,** 127–133 (2000).
50. Schofield, C. J., Valegård, K., Van Terwisscha Scheltinga, A. C., Lloyd, M. D., Hara, T., Ramaswamy, S., Perrakis, A.,Thompson, A.,Lee, H. J.,Baldwin, Jack E., Schofield, Christopher J. Hajdu, J., Andersson, Inger Structure of a cephalosporin synthase. *Nature* **394,** 805–809 (1998).
51. Baldwin, J. E. Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **387,** 827–830 (1997).
52. Dunwell, J. M., Culham, A., Carter, C. E., Sosa-Aguirre, C. R. & Goodenough, P. W. Evolution of functional diversity in the cupin superfamily. *Trends Biochem. Sci.* **26,**

740–746 (2001).
53. Clissold, P. M. & Ponting, C. P. JmjC: cupin metalloenzyme-like domains in jumonji, hairless and phospholipase A2beta. *Trends Biochem. Sci.* **26,** 7–9 (2001).
54. Aik, W., McDonough, M. A., Thalhammer, A., Chowdhury, R. & Schofield, C. J. Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr. Opin. Struct. Biol.* **22,** 691–700 (2012).
55. Clifton, I. J., McDonough, M. A., Ehrismann, D., Kershaw, N. J., Granatino, N., Schofield, C. J. Structural studies on 2-oxoglutarate oxygenases and related double-stranded β-helix fold proteins. *J. Inorg. Biochem.* **100,** 644–669 (2006).
56. Walport, L. J., Hopkinson, R. J. & Schofield, C. J. Mechanisms of human histone and nucleic acid demethylases. *Curr. Opin. Chem. Biol.* **16,** 525–534 (2012).
57. Wong, C., Fujimori, D. G., Walsh, C. T. & Drennan, C. L. Structural Analysis of an Open Active Site Conformation of Nonheme Iron Halogenase CytC3. *J. Am. Chem. Soc.* **131,** 4872–4879 (2009).
58. Blasiak, L. C., Vaillancourt, F. H., Walsh, C. T. & Drennan, C. L. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **440,** 368–371 (2006).
59. Hopkinson, R. J. Tumber, A., Yapp, C., Chowdhury, R., Aik, W., Che, Ka H., Li, Xuan S., Kristensen, Jan B. L., King, Oliver N. F., Chan, Mun C., Yeoh, Kar K., Choi, H., Walport, Louise J., Thinnes, Cyrille C., Bush, Jacob T., Lejeune, C., Rydzik, A. M.,Rose, N. R., Bagg, Eleanor A. McDonough, M. A., Krojer, Tobias J., Yue, Wyatt W., Ng, Stanley S., Olsen, L., Brennan, P. E., Oppermann, U., Müller, S., Klose, Robert J., Ratcliffe, Peter J. Schofield, C. J., Kawamura, A. 5-Carboxy-8 hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem. Sci.* **4,** 3110 (2013).
60. Kruidenier, L. A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature* **488,** 404–408 (2012).
61. Elkins, J. M. Structure of Factor-inhibiting Hypoxia-inducible Factor (HIF) Reveals Mechanism of Oxidative Modification of HIF-1α. *J. Biol. Chem.* **278,** 1802–1806 (2003).
62. Yang, C.-G.. Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452,** 961–965 (2008).
63. Yi, C.. Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature* **468,** 330–333 (2010).
64. Aas, P. A. Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature, Publ. online 20 Febr. 2003; | doi10.1038/nature01363* **421,** 859 (2003).
65. Pastore, C. Crystal Structure and RNA Binding Properties of the RNA Recognition Motif (RRM) and AlkB Domains in Human AlkB Homolog 8 (ABH8), an Enzyme Catalyzing tRNA Hypermodification. *J. Biol. Chem.* **287,** 2130–2143 (2012).
66. van den Born, E.ALKBH8-mediated formation of a novel diastereomeric pair of wobble nucleosides in mammalian tRNA. *Nat. Commun.* **2,** 172 (2011).
67. Han, Z. Crystal structure of the FTO protein reveals basis for its substrate specificity. *Nature* **464,** 1205–1209 (2010).
68. Aik, W. Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO). *J. Med. Chem.* **56,** 3680–3688 (2013).
69. Hausinger, R. P. Fe(II)/α-Ketoglutarate-Dependent Hydroxylases and Related Enzymes. *Crit. Rev. Biochem. Mol. Biol.* **39,** 21–68 (2004).
70. Chowdhury, R. Structural Basis for Binding of Hypoxia-Inducible Factor to the Oxygen-Sensing Prolyl Hydroxylases. *Structure* **17,** 981–989 (2009).

71. Yu, B. Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* **439,** 879–884 (2006).
72. Baldwin, J. E. Structure of isopenicillinN synthase complexed with substrateand the mechanism ofpenicillin formation. *Nature* **387,** 827–830 (1997).
73. Brisson, L. 1-Aminocyclopropane-1-carboxylic acid oxidase: insight into cofactor binding from experimental and theoretical studies. *JBIC J. Biol. Inorg. Chem.* **17,** 939–949 (2012).
74. Nair, S. K. & van der Donk, W. A. Structure and mechanism of enzymes involved in biosynthesis and breakdown of the phosphonates fosfomycin, dehydrophos, and phosphinothricin. *Arch. Biochem. Biophys.* **505,** 13–21 (2011).
75. Chang, W.. Mechanistic studies of an unprecedented enzyme-catalysed 1,2-phosphono-migration reaction. *Nature* **496,** 114–118 (2013).
76. Hanauske-Abel, H. M. & Günzler, V. A stereochemical concept for the catalytic mechanism of prolylhydroxylase: applicability to classification and design of inhibitors. *J. Theor. Biol.* **94,** 421–55 (1982).
77. Grzyska, P. K., Appelman, E. H., Hausinger, R. P. & Proshlyakov, D. A. Insight into the mechanism of an iron dioxygenase by resolution of steps following the FeIVboxHO species. *Proc. Natl. Acad. Sci.* **107,** 3982–3987 (2010).
78. Price, J. C., Barr, E. W., Hoffart, L. M., Krebs, C. & Bollinger, J. M. Kinetic Dissection of the Catalytic Mechanism of Taurine:α-Ketoglutarate Dioxygenase (TauD) from *Escherichia coli* †. *Biochemistry* **44,** 8138–8147 (2005).
79. Bollinger, J. M. & Krebs, C. Stalking intermediates in oxygen activation by iron enzymes: Motivation and method. *J. Inorg. Biochem.* **100,** 586–605 (2006).
80. Bollinger, J. M., Price, J. C., Hoffart, L. M., Barr, E. W. & Krebs, C. Mechanism of Taurine: α-Ketoglutarate Dioxygenase (TauD) fromEscherichia coli. *Eur. J. Inorg. Chem.* **2005,** 4245–4254 (2005).
81. Koehntop, K. D., Emerson, J. P. & Que, L. The 2-His-1-carboxylate facial triad: a versatile platform for dioxygen activation by mononuclear non-heme iron(II) enzymes. *JBIC J. Biol. Inorg. Chem.* **10,** 87–93 (2005).
82. Que, L. One motif--many different reactions. *Nat. Struct. Biol.* **7,** 182–184 (2000).
83. Krebs, C., Galonić Fujimori, D., Walsh, C. T. & Bollinger, J. M. Non-Heme Fe(IV)–Oxo Intermediates. *Acc. Chem. Res.* **40,** 484–492 (2007).
84. Costas, M., Mehn, M. P., Jensen, M. P. & Que, L. Dioxygen Activation at Mononuclear Nonheme Iron Active Sites: Enzymes, Models, and Intermediates. *Chem. Rev.* **104,** 939–986 (2004).
85. Wong, S. D. Elucidation of the Fe(iv)=O intermediate in the catalytic cycle of the halogenase SyrB2. *Nature* **499,** 320–323 (2013).
86. Hoffart, L. M., Barr, E. W., Guyer, R. B., Bollinger, J. M. & Krebs, C. Direct spectroscopic detection of a C-H-cleaving high-spin Fe(IV) complex in a prolyl-4-hydroxylase. *Proc. Natl. Acad. Sci.* **103,** 14738–14743 (2006).
87. Price, J. C., Barr, E. W., Tirupati, B., Bollinger, J. M. & Krebs, C. The First Direct Characterization of a High-Valent Iron Intermediate in the Reaction of an α-Ketoglutarate-Dependent Dioxygenase: A High-Spin Fe(IV) Complex in Taurine/α-Ketoglutarate Dioxygenase (TauD) from *Escherichia coli* †. *Biochemistry* **42,** 7497–7508 (2003).
88. van der Donk, W. A., Krebs, C. & Bollinger, J. M. Substrate activation by iron superoxo intermediates. *Curr. Opin. Struct. Biol.* **20,** 673–83 (2010).
89. Costas, M., Mehn, M. P., Jensen, M. P. & Que, L. Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. *Chem. Rev.* **104,** 939–86 (2004).

90. Mirica, L. M., McCusker, K. P., Munos, J. W., Liu, H. & Klinman, J. P. 18O kinetic isotope effects in non-heme iron enzymes: probing the nature of Fe/O2 intermediates. *J. Am. Chem. Soc.* **130,** 8122–3 (2008).
91. Hegg, E. L. & Jr, L. Q. The 2-His-1-Carboxylate Facial Triad - An Emerging Structural Motif in Mononuclear Non-Heme Iron(II) Enzymes. *Eur. J. Biochem.* **250,** 625–629 (1997).
92. Jessica R. O'Brien, §, David J. Schuller, Victoria S. Yang, §, Bret D. Dillard, § and & William N. Lanzilotta, Substrate-Induced Conformational Changes in Escherichia coli Taurine/α-Ketoglutarate Dioxygenase and Insight into the Oligomeric Structure‡. (2003). doi:10.1021/BI0341096
93. Aik, W. Structure of human RNA N6-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation. *Nucleic Acids Res.* **42,** 4741–4754 (2014).
94. Hangasky, J. A., Taabazuing, C. Y., Valliere, M. A. & Knapp, M. J. Imposing function down a (cupin)-barrel: secondary structure and metal stereochemistry in the αKG-dependent oxygenases. *Metallomics* **5,** 287 (2013).
95. Zhang, Z. Crystal structure of a clavaminate synthase-Fe(II)-2-oxoglutarate-substrate-NO complex: evidence for metal centered rearrangements. *FEBS Lett.* **517,** 7–12 (2002).
96. Henshaw, T. F., Feig, M. & Hausinger, R. P. Aberrant activity of the DNA repair enzyme AlkB. *J. Inorg. Biochem.* **98,** 856–861 (2004).
97. Borowski, T., Bassan, A. & Siegbahn, P. E. M. Mechanism of Dioxygen Activation in 2-Oxoglutarate-Dependent Enzymes: A Hybrid DFT Study. *Chem. - A Eur. J.* **10,** 1031–1041 (2004).
98. Puistola, U., Turpeenniemi-Hujanen, T. M., Myllylä, R. & Kivirikko, K. I. Studies on the lysyl hydroxylase reaction. II. Inhibition kinetics and the reaction mechanism. *Biochim. Biophys. Acta* **611,** 51–60 (1980).
99. Holme, E. A kinetic study of thymine 7-hydroxylase from neurospora crassa. *Biochemistry* **14,** 4999–5003 (1975).
100. Sedgwick, B. Repairing DNA-methylation damage. *Nat. Rev. Mol. Cell Biol.* **5,** 148–157 (2004).
101. Cantoni, G. L. THE NATURE OF THE ACTIVE METHYL DONOR FORMED ENZYMATICALLY FROM L-METHIONINE AND ADENOSINETRIPHOSPHATE [1,2]. *J. Am. Chem. Soc.* **74,** 2942–2943 (1952).
102. Yi, C. & He, C. DNA repair by reversal of DNA damage. *Cold Spring Harb. Perspect. Biol.* **5,** a012575 (2013).
103. Falnes, P. Ø., Klungland, A. & Alseth, I. Repair of methyl lesions in DNA and RNA by oxidative demethylation. *Neuroscience* **145,** 1222–1232 (2007).
104. Sedwigh, B., Bates, P., Paik, J., Jacobs, S. & Lindahl, T. Repair of alkylated DNA: Recent advances. *DNA Repair (Amst).* **6,** 429–442 (2007).
105. Yang, C.-G., Garcia, K. & He, C. Damage Detection and Base Flipping in Direct DNA Alkylation Repair. *ChemBioChem* **10,** 417–423 (2009).
106. Kondos, H. *J.B.C* Structure and Expression of the alkB Gene of Escherichia coli Related to the Repair of Alkylated DNA*. **261,** 15772–15777 (1986).
107. Mello, L. V. & Rigden, D. J. A new family of bacterial DNA repair proteins annotated by the integration of non-homology, distant homology and structural bioinformatic methods. *FEBS Lett.* **586,** 3908–3913 (2012).
108. van den Born, E. Bioinformatics and functional analysis define four distinct groups of AlkB DNA-dioxygenases in bacteria. *Nucleic Acids Res.* **37,** 7124–7136 (2009).
109. van den Born, E. Viral AlkB proteins repair RNA damage by oxidative demethylation.

Nucleic Acids Res. **36,** 5451–5461 (2008).
110. Kurowski, M. A., Bhagwat, A. S., Papaj, G. & Bujnicki, J. M. Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. *BMC Genomics* **4,** 48 (2003).
111. Mielecki, D. & Grzesiuk, E. Ada response - a strategy for repair of alkylated DNA in bacteria. *FEMS Microbiol. Lett.* **355,** 1–11 (2014).
112. Li, D. Repair of DNA Alkylation Damage by the Escherichia coli Adaptive Response Protein AlkB as Studied by ESI-TOF Mass Spectrometry. *J. Nucleic Acids* **2010,** 369434 (2010).
113. Tahiliani, M. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science (80-.).* **324,** 930–935 (2009).
114. Aravind, L. & Koonin, E. V. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* **2,** RESEARCH0007 (2001).
115. Yang, C.-G. Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452,** 961–965 (2008).
116. Delaney, J. C. & Essigmann, J. M. Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 14051–6 (2004).
117. Li, D. Exocyclic Carbons Adjacent to the N^6 of Adenine are Targets for Oxidation by the Escherichia coli Adaptive Response Protein AlkB. *J. Am. Chem. Soc.* **134,** 8896–8901 (2012).
118. Yukiko Mishina, Cai-Guang Yang, and & He*, C. Direct Repair of the Exocyclic DNA Adduct 1,N6-Ethenoadenine by the DNA Repair AlkB Proteins. (2005). doi:10.1021/JA055957M
119. Maciejewska, A. M. AlkB Dioxygenase Preferentially Repairs Protonated Substrates SPECIFICITY AGAINST EXOCYCLIC ADDUCTS AND MOLECULAR MECHANISM OF ACTION * S. (2012). doi:10.1074/jbc.M112.353342
120. Koivisto, P., Duncan, T., Lindahl, T. & Sedgwick, B. Minimal methylated substrate and extended substrate range of Escherichia coli AlkB protein, a 1-methyladenine-DNA dioxygenase. *J. Biol. Chem.* **278,** 44348–54 (2003).
121. Jia, G. Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Yun G., He, C. N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7,** 885–887 (2011).
122. S., D., F., W., N., M. & M., G. ALKBH3-dependent DNA alkylation repair promotes cancer cell proliferation. *Langenbeck's Archives of Surgery* **398,** 621–622 (2013).
123. Ringvoll, J. Moen, Marivi N., Nordstrand, L. M., Meira, L. B., Pang, B., Bekkelund, A., Dedon, P. C., Bjelland, S., Samson, L. D. AlkB Homologue 2–Mediated Repair of Ethenoadenine Lesions in Mammalian DNA. *Cancer Res.* **68,** (2008).
124. Lu, L., Yi, C., Jian, X., Zheng, G. & He, C. Structure determination of DNA methylation lesions N1-meA and N3-meC in duplex DNA using a cross-linked protein-DNA system. *Nucleic Acids Res.* **38,** 4415–25 (2010).
125. Yi, C., Hou, G., Dai, Q., Zhang, W., Zheng, G., Jian, X., Yang, Cai G., Cui, Q.,He, C. Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature* **468,** 330–333 (2010).
126. Yi, C. Duplex interrogation by a direct DNA repair protein in search of base damage. *Nat. Struct. Mol. Biol.* **19,** (2012).
127. Ringvoll, J. Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1meA and 3meC lesions in DNA. *EMBO J.* **25,** 2189–2198 (2006).

128. Sundheim, O. Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage. *EMBO J.* **25,** 3389–97 (2006).
129. Tasaki, M., Shimada, K., Kimura, H., Tsujikawa, K. & Konishi, N. ALKBH3, a human AlkB homologue, contributes to cell survival in human non-small-cell lung cancer. *Br. J. Cancer* **104,** 700–706 (2011).
130. Li, M.-M. ALKBH4-dependent demethylation of actin regulates actomyosin dynamics. *Nat. Commun.* **4,** 1832 (2013).
131. Tsujikawa, K. Expression and sub-cellular localization of human ABH family molecules. *J. Cell. Mol. Med.* **11,** 1105–1116 (2007).
132. Rhee, H.-W. Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science (80-.).* **339,** 1328–1331 (2013).
133. Fu, D., Jordan, J. J. & Samson, L. D. Human ALKBH7 is required for alkylation and oxidation-induced programmed necrosis. *Genes Dev.* **27,** 1089–100 (2013).
134. Solberg, A. Deletion of mouse Alkbh7 leads to obesity. *J. Mol. Cell Biol.* **5,** 194–203 (2013).
135. Fu, Y. The AlkB Domain of Mammalian ABH8 Catalyzes Hydroxylation of 5-Methoxycarbonylmethyluridine at the Wobble Position of tRNA. *Angew. Chemie Int. Ed.* **49,** 8885–8888 (2010).
136. Shimada, K. A Novel Human AlkB Homologue, ALKBH8, Contributes to Human Bladder Cancer Progression. *Cancer Res.* **69,** 3157–3164 (2009).
137. Kim, Y.-O., Park, S.-J., Balaban, R. S., Nirenberg, M. & Kim, Y. A functional genomic screen for cardiogenic genes using RNA interference in developing Drosophila embryos.
138. Alexandra B Byrne, Matthew T Weirauch, Victoria Wong, Martina Koeva, Scott J Dixon, J. M. S. author and P. J. R. A global analysis of genetic interactions in Caenorhabditis elegans. *J. Biol.* **25,** 547–554 (2007).
139. Church, C. A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. *PLoS Genet.* **5,** (2009).
140. Fischer, J. Inactivation of the Fto gene protects from obesity. *Nature* **458,** 894–898 (2009).
141. Da Cunha, P. A. Interaction between obesity-related genes, FTO and MC4R, associated to an increase of breast cancer risk. *Mol. Biol. Rep.* **40,** 6657–6664 (2013).
142. Dina, C. Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat. Genet.* **39,** 724–726 (2007).
143. Karra, E. A link between FTO, ghrelin, and impaired brain food-cue responsivity. *J. Clin. Invest.* **123,** 3539–51 (2013).
144. Boissel, S. Loss-of-Function Mutation in the Dioxygenase-Encoding FTO Gene Causes Severe Growth Retardation and Multiple Malformations. *Am. J. Hum. Genet.* **85,** 106–111 (2009).
145. Kaklamani, V. The role of the fat mass and obesity associated gene (FTO) in breast cancer risk. *BMC Med. Genet.* **12,** 52 (2011).
146. Ratel, D., Ravanat, J.-L., Berger, F. & Wion, D. N6-methyladenine: the other methylated base of DNA. *BioEssays* **28,** 309–315 (2006).
147. Mauer, JReversible methylation of m6Am in the 5′ cap controls mRNA stability. *Nature* **541,** 371–375 (2016).
148. Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8,** 1698–710 (2009).
149. Zhang, H. TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res.* **20,** 1390–1393 (2010).

150. Iqbal, K., Jin, S.-G., Pfeifer, G. P. & Szabo, P. E. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl. Acad. Sci.* **108,** 3642–3647 (2011).
151. Kaas, G. TET1 controls CNS 5-Methylcytosine Hydroxylation, active DNA demethylation, gene transcription, and memory formation. *Neuron* **79,** 1086–1093 (2013).
152. Sun, M. HMGA2/TET1/HOXA9 signaling pathway regulates breast cancer growth and metastasis. *Proc. Natl. Acad. Sci.* **110,** 9920–9925 (2013).
153. Neri, F. TET1 is a tumour suppressor that inhibits colon cancer growth by derepressing inhibitors of the WNT pathway. *Oncogene* **34,** 4168–4176 (2015).
154. Lian, C. G. Loss of 5-Hydroxymethylcytosine Is an Epigenetic Hallmark of Melanoma. *Cell* **150,** 1135–1146 (2012).
155. Ko, M. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468,** 839–43 (2010).
156. Imielinski, M. Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing. *Cell* **150,** 1107–1120 (2012).
157. Hsu, C.-H. TET1 Suppresses Cancer Invasion by Activating the Tissue Inhibitors of Metalloproteinases. *Cell Rep.* **2,** 568–579 (2012).
158. Cohn, W. E. & Volkin, E. Nucleoside-5′-Phosphates from Ribonucleic Acid. *Nature* **167,** 483–484 (1951).
159. Machnicka, M. A. MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Res.* **41,** D262-7 (2013).
160. Jin, S.-G. 5-Hydroxymethylcytosine Is Strongly Depleted in Human Cancers but Its Levels Do Not Correlate with IDH1 Mutations. *Cancer Res.* **71,** 7360–7365 (2011).
161. Bestor, T. H. Cytosine methylation: the pros and cons of DNA methylation. *Curr. Biol.* **3,** 384–6 (1993).
162. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9,** 465–476 (2008).
163. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **10,** 295–304 (2009).
164. He, C. Grand Challenge Commentary: RNA epigenetics? *Nat. Chem. Biol.* **6,** 863–865 (2010).
165. Roundtree, I. A. & He, C. RNA epigenetics—chemical messages for posttranscriptional gene regulation. *Curr. Opin. Chem. Biol.* **30,** 46–51 (2016).
166. Zheng, G. ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Mol. Cell* **49,** 18–29 (2013).
167. Mauer, J. Reversible methylation of m6Am in the 5′ cap controls mRNA stability. *Nature* **541,** 371–375 (2016).
168. Fustin, J.-M. RNA-Methylation-Dependent RNA Processing Controls the Speed of the Circadian Clock. *Cell* **155,** 793–806 (2013).
169. Nishikura, K. Functions and Regulation of RNA Editing by ADAR Deaminases. *Annu. Rev. Biochem.* **79,** 321–349 (2010).
170. Zhong, S. MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* **20,** 1278–88 (2008).
171. Holoch, D. & Moazed, D. RNA-mediated epigenetic regulation of gene expression. *Nat. Publ. Gr.* **16,** (2015).
172. Saikia, M., Fu, Y., Pavon-Eternod, M., He, C. & Pan, T. Genome-wide analysis of N1-methyl-adenosine modification in human tRNAs. *RNA* **16,** 1317–27 (2010).
173. Lee, M., Kim, B. & Kim, V. N. Emerging Roles of RNA Modification: m 6 A and U-Tail. (2014). doi:10.1016/j.cell.2014.08.005

174. Kirchner, S. & Ignatova, Z. Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. (2015). doi:10.1038/nrg3861
175. Browne, M. J., Turnbull, J. F., McKay, E. L., Adams, R. L. & Burdon, R. H. The sequence specificity of a mammalian DNA methylase. *Nucleic Acids Res.* **4,** 1039–45 (1977).
176. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11,** 1475–1489 (1983).
177. WEI, C.-M., GERSHOWITZ, A. & Moss, B. N6, O2′-dimethyladenosine a novel methylated ribonucleoside next to the 5′ terminal of animal cell and virus mRNAs. *Nature* **257,** 251–253 (1975).
178. Liu, J. A METTL3–METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* **10,** 93–95 (2013).
179. Ping, X.-L. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* **24,** 177–189 (2014).
180. Meyer, K. D. Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and near Stop Codons. *Cell* **149,** 1635–1646 (2012).
181. Wang, Y. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* **16,** 191–198 (2014).
182. Karikó, K., Buckstein, M., Ni, H. & Weissman, D. Suppression of RNA Recognition by Toll-like Receptors: The Impact of Nucleoside Modification and the Evolutionary Origin of RNA. *Immunity* **23,** 165–175 (2005).
183. Zhang, Z. The YTH domain is a novel RNA binding domain. *J. Biol. Chem.* **285,** 14701–14710 (2010).
184. Dominissini, D. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485,** 201–206 (2012).
185. Niu, Y. N6-methyl-adenosine (m6A) in RNA: An Old Modification with A Novel Epigenetic Function. *Genomics. Proteomics Bioinformatics* **11,** 8–17 (2013).
186. Utsch, B. Exclusion of WTAP and HOXA13 as candidate genes for isolated hypospadias. *Scand. J. Urol. Nephrol.* **37,** 498–501 (2003).
187. Ortega, A. Biochemical function of female-lethal (2)D/Wilms' tumor suppressor-1-associated proteins in alternative pre-mRNA splicing. *J. Biol. Chem.* **278,** 3040–7 (2003).
188. Zeggini, E. Replication of Genome-Wide Association Signals in UK Samples Reveals Risk Loci for Type 2 Diabetes. *Science (80-.).* **316,** (2007).
189. Keller, L. The obesity related gene, FTO, interacts with APOE, and is associated with Alzheimer's disease risk: a prospective cohort study. *J. Alzheimers. Dis.* **23,** 461–9 (2011).
190. Wang, Y. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* **16,** 191–198 (2014).
191. Ferroni, P. Determinants of homocysteine levels in colorectal and breast cancer patients. *Anticancer Res.* **29,** 4131–8 (2009).
192. Wang, X.. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505,** 117–120 (2013).
193. Heiliger, K. J. Novel candidate genes of thyroid tumourigenesis identified in Trk-T1 transgenic mice. *Endocr. Relat. Cancer* **19,** 409–421 (2012).
194. Bjork, G., Wikstrom, P. & Bystrom, A. Prevention of translational frameshifting by the modified nucleoside 1-methylguanosine. *Science (80-.).* **244,** (1989).
195. Limbach, P. A., Crain, P. F. & Mccloskey, J. A. Summary: the modified nucleosides of RNA. *Nucleic Acids Res.* **22,** 2183–2196 (1994).

Chapter 2: Inhibition studies on FTO

2.1 Introduction

Obesity is an epidemic in the Western World and is rapidly spreading its wings into developing countries. Mutations to the fat mass and obesity associated gene (*FTO*) have been linked to an increased obesity *via* genome-wide association studies;[1] FTO is a 2-oxoglutarate (2OG, marked as **I** in Figure 2.1A) and iron-dependent oxygenase, which belongs to the AlkB subfamily of nucleic acid oxygenases (NAOXs).[2] Mice possessing an inactive FTO variant (resulting from a single mutation to the wild-type protein) display a lean phenotype.[3] Furthermore, FTO has been associated with diseases such as type II diabetes,[4] Alzheimer's disease,[5] and breast cancer.[6,7] Therefore, although the exact biological role of FTO has, as yet, not been defined, the inhibition of FTO by small molecules is being pursued as a potential treatment for obesity and related diseases including cardiovascular disease.

2.1 Previous studies on Inhibitors

2.1.1 General 2OG inhibitors

As discussed in detail in Chapter 1, 2OG oxygenases catalyse many reactions of high biological importance,[8] piquing interest in their inhibition using small molecules.[9] Most, but not all, 2OG oxygenases, use Fe(II) as a cofactor, and 2OG and O_2 as co-substrates, and have a core double-stranded ß-helix, DSBH protein fold.[10–12] These common characteristics enable the development of generic templates for the inhibition of 2OG oxygenases.[13] Thus, metal chelators and 2OG mimetics have been shown to act as broad spectrum inhibitors of 2OG oxygenases.[9,14–25] Fe(II) binds relatively weakly at the active site of some 2OG oxygenases;[11,12] consequently, inhibition by divalent transition metals like Mn(II), Ni(II),

Co(II), Cu(II), Zn(II) has been reported.[14–17] 2OG mimetics include *N*-oxalyl amino acids (**II** and **III** in Figure 2.1A), substituted hydroxamic acids (**IV** and **V** in Figure 2.1A), and 2,x-pyridinedicarboxylates (**VI** and **VII** in Figure 2.1A). Pyridine carboxylates are cyclic analogues of 2OG. Hydroxamic acids are well established transition-metal chelators, e.g. of Zn-containing enzymes;[18] hydroxamic acid inhibitors have been reported to compete with 2OG binding in a few 2OG oxygenases.[9,18–20]

2OG oxygenases are reported to be involved in the biosynthesis of many natural products;[21] and some natural products are reported as inhibitors of 2OG oxygenases.[22–24] One such natural product family is the flavonoids, which include catechol and its derivatives (flavonoid related compounds; **VIII-XIII** in Figure 2.1B), which are reported to inhibit 2OG oxygenases.[23,25–27] Some endogenous compounds also inhibit 2OG oxygenases, including the TCA cycle intermediates succinate and fumarate, and others.[21]

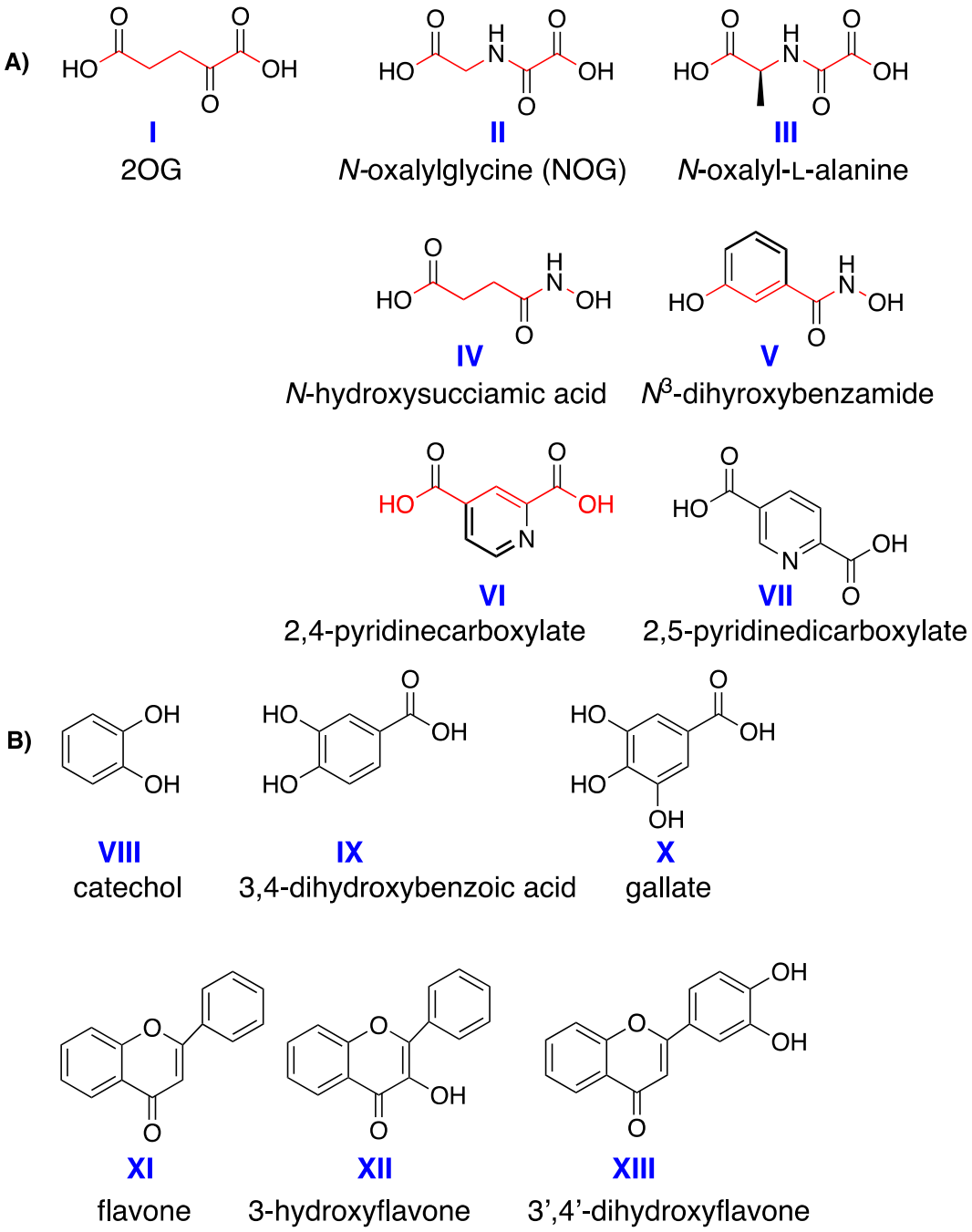
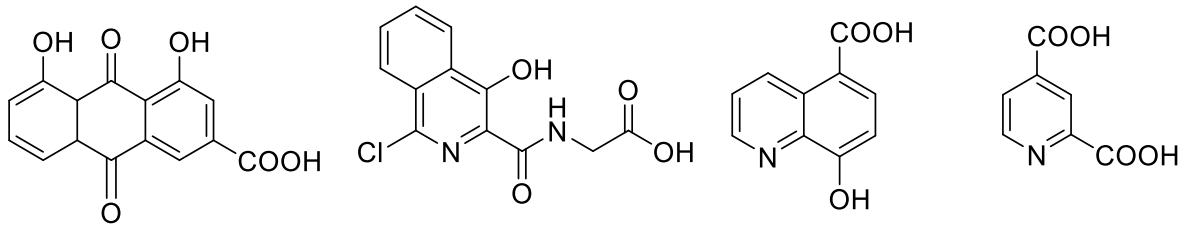


Figure 2.1 A) Some generic 2OG oxygenases inhibitors. (Red colouring highlights the 2OG backbone). B) Examples of a natural product family of 2OG inhibitors.

2.2 Previous studies on FTO inhibition

In 2010, a crystal structure of FTO was reported.[28] To date, reported inhibitors of FTO include rhein, 2OG analogues, meclofenamic acid and compounds that occupy the 2OG and substrate binding site simultaneously.[29–36] The first reported inhibitor of FTO was a natural

compound called rhein (marked as **1** in Figure 2.2).[29] Later, crystallographic studies confirmed that rhein was in fact a substrate competitor.[30] However, rhein showed little selectivity for FTO over other AlkB related oxygenases.[29,35] In studies initiated in Oxford (Dr Wei Shen and Dr Marina Demetriades), other inhibitors of FTO were developed based on crystal structures.[30,37] One of them (**FG2216**, marked as **2** in Figure 2.2) is a potent PHD2 inhibitor in clinical trials. Although the four compounds investigated in our lab (**1-4** in Figure 2.2) are not sufficiently potent, and in the case of the latter three compounds, not sufficiently selective for use as chemical probes of FTO, the development of 2OG and substrate-competitive compounds based on these scaffolds may provide potent and selective FTO inhibitors.



IC_{50}(μM)	**1**	**2**	**3**	**4**
FTO	9	3	3.3	8.3
PHD2	-	0.3	14.3	6
JMJD2C	-	>20	0.6	2.6

Figure 2.2 Reported FTO inhibitors with IC_{50} values against FTO, PHD2, and JMJD2C (KDM4C).

In a 2014 study ,[31] meclofenamic acid (marked as **XIV** in Figure 2.3) was reported as an inhibitor of FTO with an IC_{50} value of 8 μM. Meclofenamic was reported to be exclusively selective for FTO over ABH5.[31] Another inhibitor of FTO was reported in 2014 having anticonvulsant activity.[33] Interestingly, the compounds were designed to inhibit PHD2 rather than FTO;[33] this study also reported that mimics of ascorbate act as inhibitors of FTO with IC_{50} values of 4.9 μM and 8.7 μM (**XV** and **XVI** in Figure 2.3). In 2015,[36] a novel binding site for FTO inhibition was reported; the derivatives of benzene 1,3 diol (**XVII** in Figure 2.3) were tested for FTO inhibition with the most potent compound having an IC_{50} of 4.95 μM. A

study published by the same group in 2016,[32] reported a new inhibitor, 4-chloro-6-(6′-chloro-7′- hydroxy-2′,4′,4′-trimethyl-chroman-2′-yl)benzene-1,3-diol (CHTB, **XVIII** in Figure 2.3), with an IC_{50} of 39.24 µM. In a detailed study published in 2015,[34] inhibitors for the AlkB subfamily were reported. The most potent inhibitor of FTO, to date, is a fumarate derivate (**XIX** in Figure 2.3) with an IC_{50} of 0.81µM and which is reported to have 100-fold selectivity for FTO over ABH5, a very rare selectivity to be achieved within a 2OG oxygenase subfamily. However, **XIX** has been observed to be much more active against JMJD1A (KDM3A; IC_{50} 0.4 µM) over FTO in Oxford (IC_{50}, preliminary results). It should be noted that the work (synthesis of inhibitors) described in this chapter was completed in 2014, before studies by other groups on FTO inhibitors were reported.

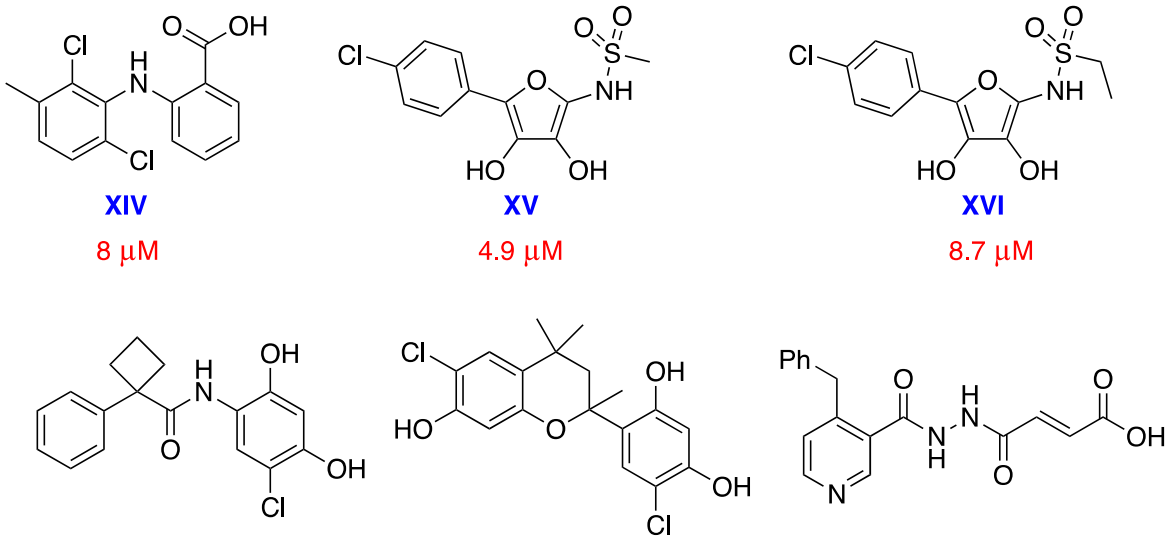


Figure 2.3 Literature reported inhibitors of FTO. Reported IC_{50} values are indicated in red.

2.3 The objectives of the work described in this chapter

The specific objectives of the work described in this chapter are as follows:

- To synthesise potent and selective inhibitors of FTO building on prior studies in Oxford;
- To obtain T_m shift values of the synthesised inhibitors;
- To obtain the IC_{50} values of the inhibitors;
- To obtain the dissociation constants (K_D) for inhibitors.

2.4 Results and Discussion

2.4.1 Development of FTO Inhibitors

Initially, views from crystal structures of the inhibitors, rhein and **FG2216**, (marked as **1** and **2** in Figure 2.2) and, rhein and 2-substituted pyridine **5** (Figure 2.5), were overlaid and analysed (Figure 2.4). It was noted that the close proximity of the pyridyl ring of **2** and **5**, in the 2OG binding site, to **1**, in the substrate binding site suggesting that compounds that occupy both sites, the substrate and the 2OG binding pocket, of FTO, might be potent FTO inhibitors (Figure 2.4). Since the 2OG binding modes of 2OG oxygenases are more conserved than their substrate binding modes, it is more likely that 2OG analogues are likely to bind and inhibit multiple 2OG oxygenases. However, the substrate binding sites of 2OG oxygenases are often less similar, and therefore, offer a more promising target for the development of selective binders. From the analysis of the crystal structures, compound **5** was identified as a useful scaffold for the new generation of inhibitors; it was noted that both positions, 5- and 6-, of the ring, have the potential to be extended to the substrate binding site.

Both 5-substituted compounds **6** and **7** (Figure 2.5), which are potent PHD2 inhibitors, displayed increased IC_{50} values for FTO compared to **5**. In contrast, the 6-bromo substituted pyridyl analogue, **8**, showed an improvement in the potency for FTO. Thus, the 6-position of the pyridyl ring was chosen for substitution.

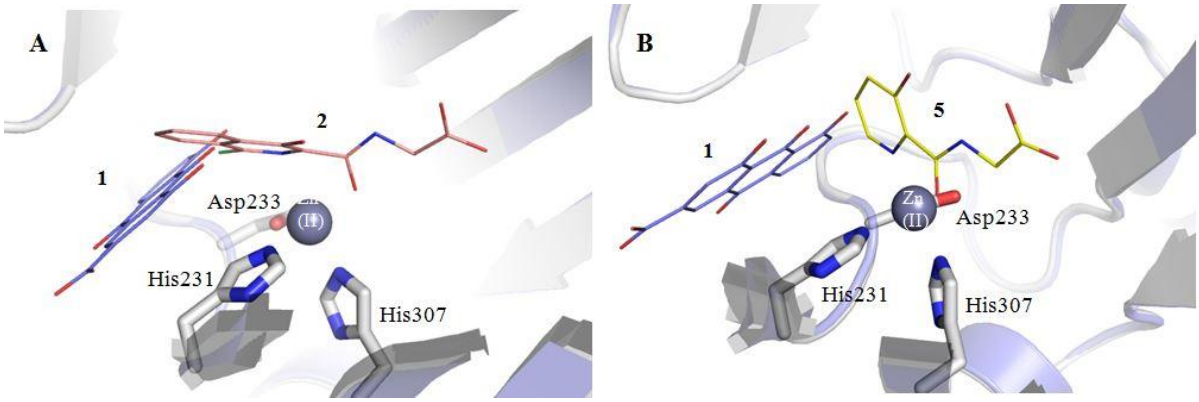
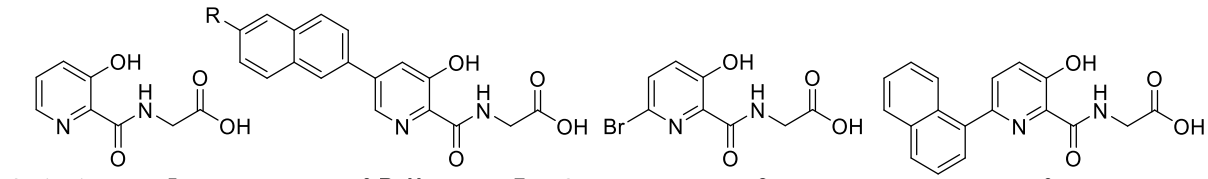


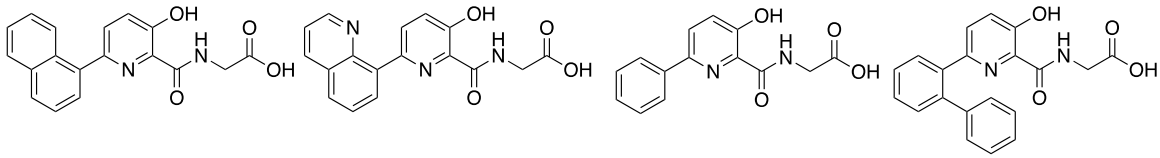
Figure 2.4 Alignment of the crystal structures of A) FTO·1 (PDB ID: 4IE7) and FTO·2 (PDB ID: 4IE6) and B) FTO·1 (PDB ID: 4IE7) and FTO·5 (PDB ID: 4IE5).



IC_{50}(μM)	**5**	**6 R=H**	**7** R=OH	**8**	**9**
FTO	15	43	96	9	4
PHD2	400	0.017	0.004	-	-
JMJD2C/E	>>20(E)	~20(E)	>20(C)	-	>100(C)

Figure 2.5 IC_{50} values for pyridyl inhibitors against FTO, PHD2, and JMJD2E (KDM4E).

Initially, aryl substituted compounds were synthesised by Suzuki coupling (**9-9c**, Figure 2.6). All compounds (**9-9c**, Figure 2.6) synthesised using this method showed weaker inhibitory activity against the 2OG dependent histone demethylases (JMJD2C; also known as KDM4C) relative to compounds **2-4** (IC_{50} >20 μM, Figure 2.2). Compound **9** was the most active FTO inhibitor (IC_{50} 4 μM) and was 7-fold more selective for FTO over the tested demethylases compared to **3**.



IC_{50} (μM)	**9**	**9a**	**9b**	**9c**
FTO	4±1	25±1	35±1	15±1
JMJD2C	>100	>100	>100	27

Figure 2.6 IC_{50} values for pyridyl inhibitors against FTO and JMJD2C (KDM4C).

An FTO crystal structure in complex with **9** was obtained by Dr Wei Shen Aik and revealed that the compound binds to both the 2OG and substrate binding sites supporting the structure-based design process (Figure 2.7).

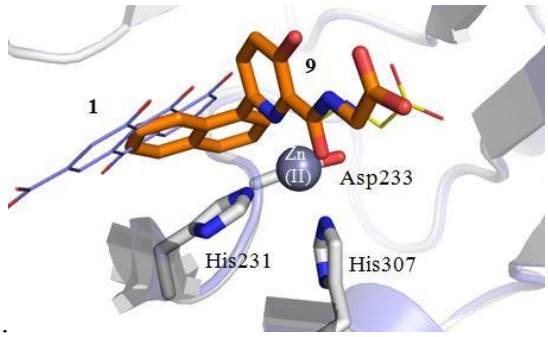


Figure 2.7 View from a crystal structure of FTO complexed with 9 and zinc substituting for iron (PDB ID: 4QHO)

The synthesis of a new generation of analogues containing novel substitutions at the 6-position of the pyridyl ring of **5** (Figure 2.5) was then attempted. Analyses of crystal structure of FTO in complex with **9** (PDB ID: 4QHO) suggested that the formation of a carbon-heteroatom bond between pyridyl ring of **5** and the coupled aryl rings could provide more favourable hydrogen bonding interactions with the residues of FTO over a carbon-carbon bond as in **9** (Figure 2.8).

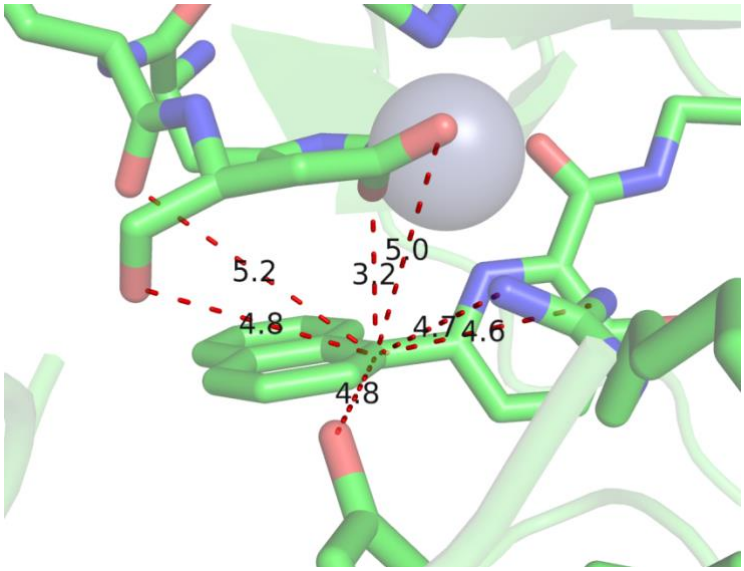


Figure 2.8 View from a crystal structure of FTO.9 (PDB ID: 4QHO) showing the possible H-bonding interactions, if the carbon-carbon bond is replaced by a carbon-heteroatom bond.

*It should be noted that the alignment of crystal structures was done by Dr Marina Demetriades (Figure 2.4). The initial Suzuki coupled biaryl compounds were synthesised by Dr Marina Demetriades and Caitlin O'Connor (**9-9c**) and all the crystal structures were obtained by Dr Wei Shen Aik (Figure 2.4 and Figure 2.7).*

2.4.2 Synthesis of Inhibitors

3-Hydroxypicolinonitrile (**10**) was chosen as the starting material for the synthesis of analogues of **9**. The first step was envisaged as the halogenation at the C-6 position of the pyridine ring. The electrophilic substitution of bromine at the C-6 position of 3-hydroxypicolinonitrile (**10**) was previously done using *N*-bromosuccinimide (NBS) as a bromine source. However, dibromination of **10**, to give 4,6-dibromo-3-hydroxypicolinonitrile (**12**), was observed as a major product; separation of the products through column chromatography was not successful. Bromine in water was then investigated as an alternative for NBS (Figure 2.9). However, dibromination of **10** to give **12** as a major product was observed again. Finally, portion-wise addition of bromine over a period of 3 days enabled the formation of the desired product, 6-bromo-3-hydroxypicolinonitrile (**11**) in a reasonable 53% yield (Figure 2.9). Serendipitously, the dibromo product (**12**) precipitated out of the solution,

leaving behind the reasonably pure mono-brominated product (**11**) which was used in the next step without further purification.

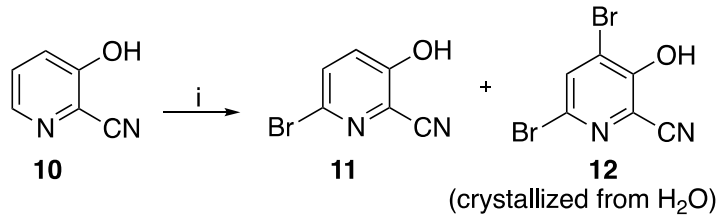


Figure 2.9 Reaction condition: i) Br_2 (1.1 eq), H_2O, 3 days.

Following the bromination at the C-6 position of 3-hydroxypicolinonitrile (**10**) to give **11**, the second step was the basic hydrolysis of the cyanide group of **11** to give 6-bromo-3-hydoxypicolinic acid (**13**; Figure 2.10). Hydrolysis was achieved under harsh conditions (30% NaOH in methanol, 12 hr reflux) which resulted in an apparent yield of 100% after acidification of the aqueous layer (Figure 2.10). Following the hydrolysis, the resultant acid (**13**) was coupled with the protected glycine amide (glycine *tert*-butyl ester hydrochloride or glycine benzyl ester hydrochloride) in reasonable yield using *N*-propanephosphonic acid anhydride (T3P) (Figure 2.10).

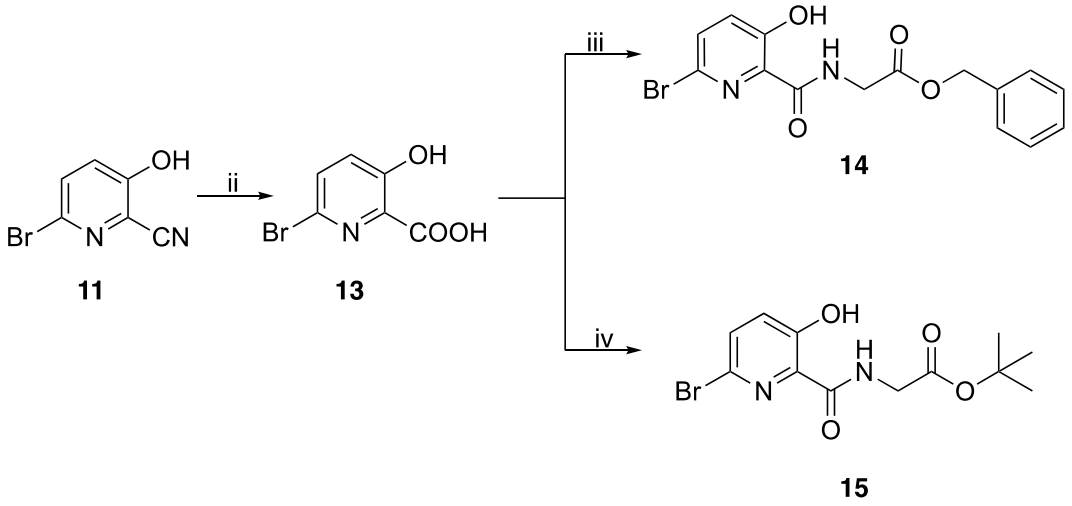


Figure 2.10 Reaction conditions: ii) NaOH(30%) in MeOH, reflux, 12 hr; iii) GlyOCH$_2$Ph.HCl (1.5eq), T3P (1.5eq), DIPEA(4 eq),mw, 120°C, 4 hr; iv) GlyOt-Bu.HCl (1.5eq),DIPEA(4 eq),mw, 120°C, 4 hr.

In the previously used synthetic scheme (Figure 2.11), carbon-carbon bond formation at the C-6 position of **14** or **15** had been achieved by using Suzuki coupling; however, the yield of the reaction was very low and chromatographic purification of the crude material was tedious.

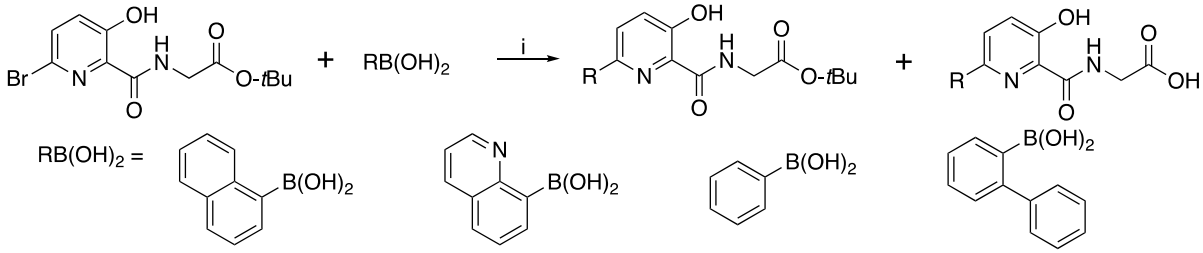


Figure 2.11 Reaction Conditions: i) $RB(OH)_2$ (1.3 eq), K_2CO_3 (4 eq), $Pd(PPh_3)_4$ (10 eq), mw, 130°C, 4hr.

As explained earlier (Figure 2.8), the addition of a heteroatom between the pyridyl and the coupled aromatic rings was considered to be favourable. Thus, instead of using boronic acids in the Suzuki coupling, attempts were made to couple aromatic amines with **14** or **15**. Initially, X-Phos (**18** in Figure 2.12) and tris(dibenzylideneacetone)dipalladium(0) (Pd_2dba_3, **17** in Figure 2.12), were used in an attempt to couple anthraquinone structure based amine (**16,** Figure 2.12) and glycine amide (**14**); sodium bicarbonate ($NaHCO_3$) was used as the base, acetonitrile as the solvent, and the mixture was heated in microwave at 120°C for 4 hrs. However, conversion to the desired product (**19**) was not observed by LCMS, under tested conditions. Subsequently, Hartwig-Buchwald coupling was thoroughly investigated– the initial results were promising for coupling of the aromatic amine (**16**) with haloarenes (**14** or **15**).

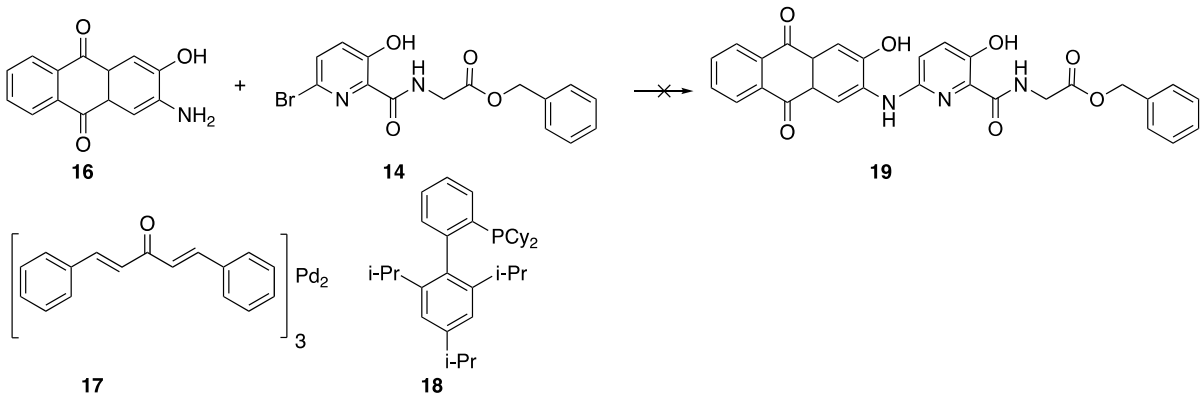


Figure 2.12 Reaction conditions: $NaHCO_3$ (1.2 eq), Pd_2dba_3(5 eq), XPhos (10 eq), CH_3CN, 120°C, mw, 4hr.

2.4.3 Buchwald-Hartwig Coupling

Pd-catalysed cross-coupling reactions are arguably one of the most important synthetic transformations in modern chemistry; the Nobel Prize was awarded in 2010 to the pioneers of Pd-catalysed C-C bond formation reactions (Prof. Richard F. Heck; Prof. Akira Suzuki; Prof. Ei-ichi Negishi) signifying the prominent position these reactions hold in organic synthesis. A general catalytic cycle involved for Pd-based coupling is given in Figure 2.13.

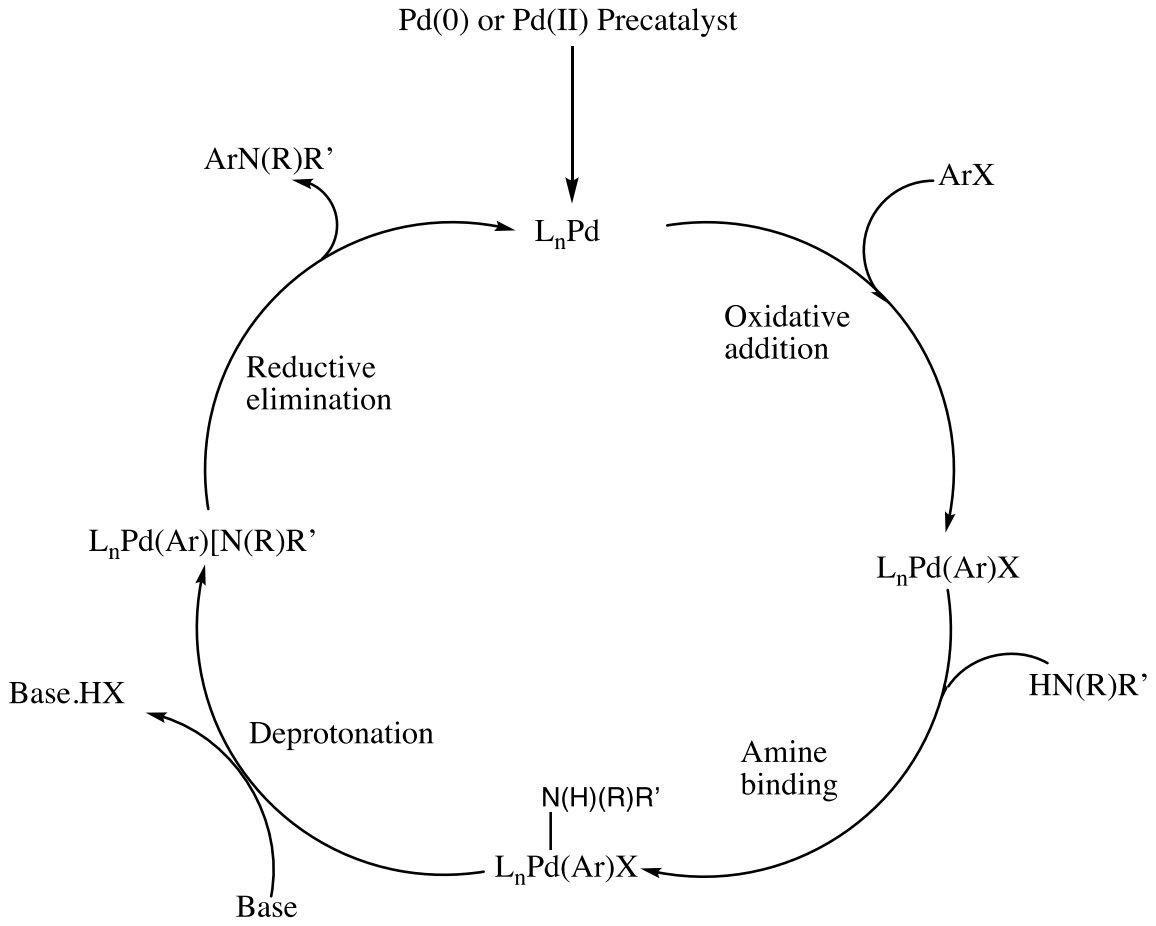
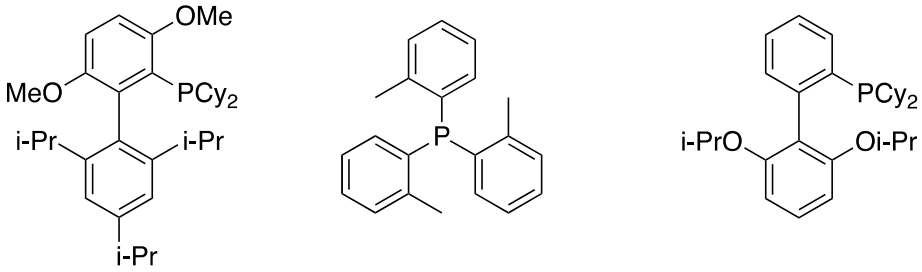


Figure.2.13 Schematic representation of the catalytic cycle of Pd-catalysed amination cross-coupling reaction.[40]

Pd-catalysed amination is believed to involve four intermediates in the catalytic cycle. Except for the first intermediate, which is a ligand-palladium (0) complex, all three intermediates are Pd (II) complexes.[40] The structure of the ligands used for binding to palladium in cross-coupling reactions, has also evolved from simple acetate (OAc; as in $Pd(OAc)_2$) to phosphine ligands e.g. tri(o-tolyl)phosphine ($P(o\text{-}tolyl)_3$; **21a** in Figure 2.14), bidentate ligands, and subsequently bialkylbiaryl and ferrocene-based ligands.[38–42] The large array of ligands has been in part developed to address the variations in the electronic and the steric properties of nucleophiles and electrophiles coupled using Pd based reagents. To date, the most important

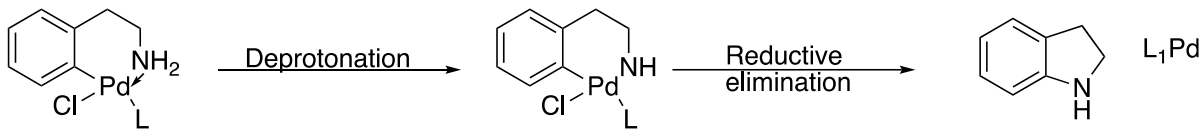
ligands for the coupling of the primary amine and the secondary amines, are arguably BrettPhos and RuPhos (**20** and **21b**, respectively, in Figure 2.14).[40]



20 **21a** **21b**

Figure 2.14 Few phosphine ligands used in Pd-catalysed amination.

In recent times, an amine-ligated oxidative complex of Pd and ligand has been developed which can be activated by a strong base.[43] This complex undergoes reductive elimination producing indoline and monoligated Pd(0) complex (Figure 2.15).[43] This procedure eliminates the requirement for the addition of an extra reducing agent that is otherwise required to form $Pd(0)L_1$ complex to initiate the Pd-based coupling.[40,43]



L=BrettPhos, **22**

Figure 2.15 Schematic representation of the activation of intramolecular coordinated amine oxidative addition pre-catalyst.[43]

2.4.4 Coupling of amines using Hartwig-Buchwald coupling

After unsuccessful attempts to couple anthraquinone based amine (**16** in Figure 2.12) and haloarene (**14**) using Pd_2dba_3 (**17** in Figure 2.12) as a catalyst and X-Phos (**18** in Figure 2.12) as a phosphine ligand, the coupling of **16** and **14** was then attempted using BrettPhos (**20** in Figure 2.14) as a ligand; potassium phosphate (K_3PO_4) as base; *t*-BuOH as solvent, and BrettPhos Pd G1 Methyl-*t*-butyl ester as a Pd source (**22** in Figure 2.15). An equal ratio of

Ligand (BrettPhos; **20** in Figure 2.14) and the Pd source (**22** in Figure 2.15), each 3 equivalents relative to the substrate, was used as a starting point (Figure 2.16).

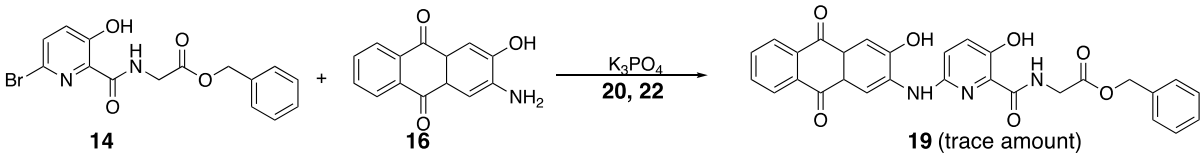


Figure 2.16 Reaction conditions: ***16*** *(1.1eq),* K_3PO_4*(1.2 eq),* ***20****(3 eq),* ***22****(3 eq), mw,120°C,4hr.*

After heating the reaction mixture for 4hr at 120°C in a microwave oven, a trace amount of product (**19**) was observed, as analysed by LCMS. It was envisaged that the presence of additional functional groups apart from the amine in **16**, may have interfered with the reaction yield leading to the formation of undesired products. Hence, simpler amines were then tested as nucleophiles in the coupling (Figure 2.17). **23** was used as a model (Figure 2.17). As evident from the structure of **23**, the aromatic amine is not electron rich, thus optimisation of the reaction conditions was required. Using 1.2 equivalents of K_3PO_4 as a base, 3 equivalents of BrettPhos (**20**) and Pd-source (**22**), the reaction mixture was heated at 120°C for 4hr; 20% conversion to the product (28; Figure 2.18A/B) was observed by LCMS. Lithium bis(trimethylsilyl) amide (LiHMDS) was then tested as a base for the coupling reaction as suggested in the literature;[44] however, no product conversion was observed in this case (though the LiHMDS was not freshly prepared). The attention then turned back to K_3PO_4 as the base for the coupling.

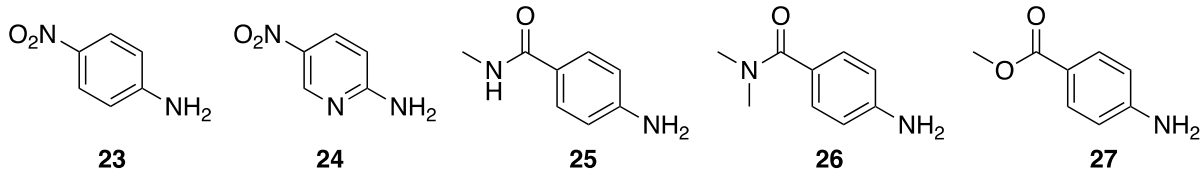


Figure 2.17 List of aromatic amines used in the Hartwig-Buchwald coupling.

It was envisaged that increasing the amount of ligand (**20**) in the reaction mixture would help to ensure the availability of L_nPd throughout the reaction time (Figure 2.13). Thus, the number of equivalents of the ligand, BrettPhos (**20**), was increased to 4.5; the yield of the reaction increased to 50%. It was envisaged that the *t*-butyl ester group might hydrolyse to the corresponding acid under basic conditions at high temperature, therefore Cs_2CO_3, which is a weaker base compared to K_3PO_4, was used with **15** as an electrophile which resulted in a reaction yield of 30%. For further investigations, K_3PO_4 was selected as the base. Increasing the equivalents of BrettPhos (**20**) to 9 and Pd source (**22**) to 6 with the equivalents of other reagents remained unchanged, increased the reaction yield to 60%. Subsequently, the equivalents of BrettPhos (**20**) and Pd source (**22**), relative to the substrate (**14**), were increased to 13.5 and 9, respectively, resulting in an apparent 100 % conversion to the product as observed by LCMS. The optimised conditions were then used with other amines (listed in Figure 2.17) to form novel inhibitors candidates (Figure 2.18).

Table 2.1 Optimisation summary of the Hartwig-Buchwald reaction used in the synthesis of inhibitors. [a]Yields are given after 8 hr at 120°C under mw irradiation as observed by LCMS.

14/15	**Amine**	**Catalyst (eq)**	**L (eq)**	**Base (1.2 eq)**	**Yield (%)[a]**
14	**16**	**17**(5)	**18**(10)	$NaHCO_3$	0
14	**16**	**22** (3)	**20** (3)	K_3PO_4	trace
14	**23**	**22** (3)	**20** (3)	K_3PO_4	20
14	**23**	**22** (3)	**20** (3)	LiHMDS	0
14	**23**	**22** (3)	**20** (4.5)	K_3PO_4	50
15	**23**	**22** (3)	**20** (4.5)	Cs_2CO_3	30
14	**23**	**22** (6)	**20** (9)	K_3PO_4	60
14	**23**	**22** (9)	**20** (13.5)	K_3PO_4	100

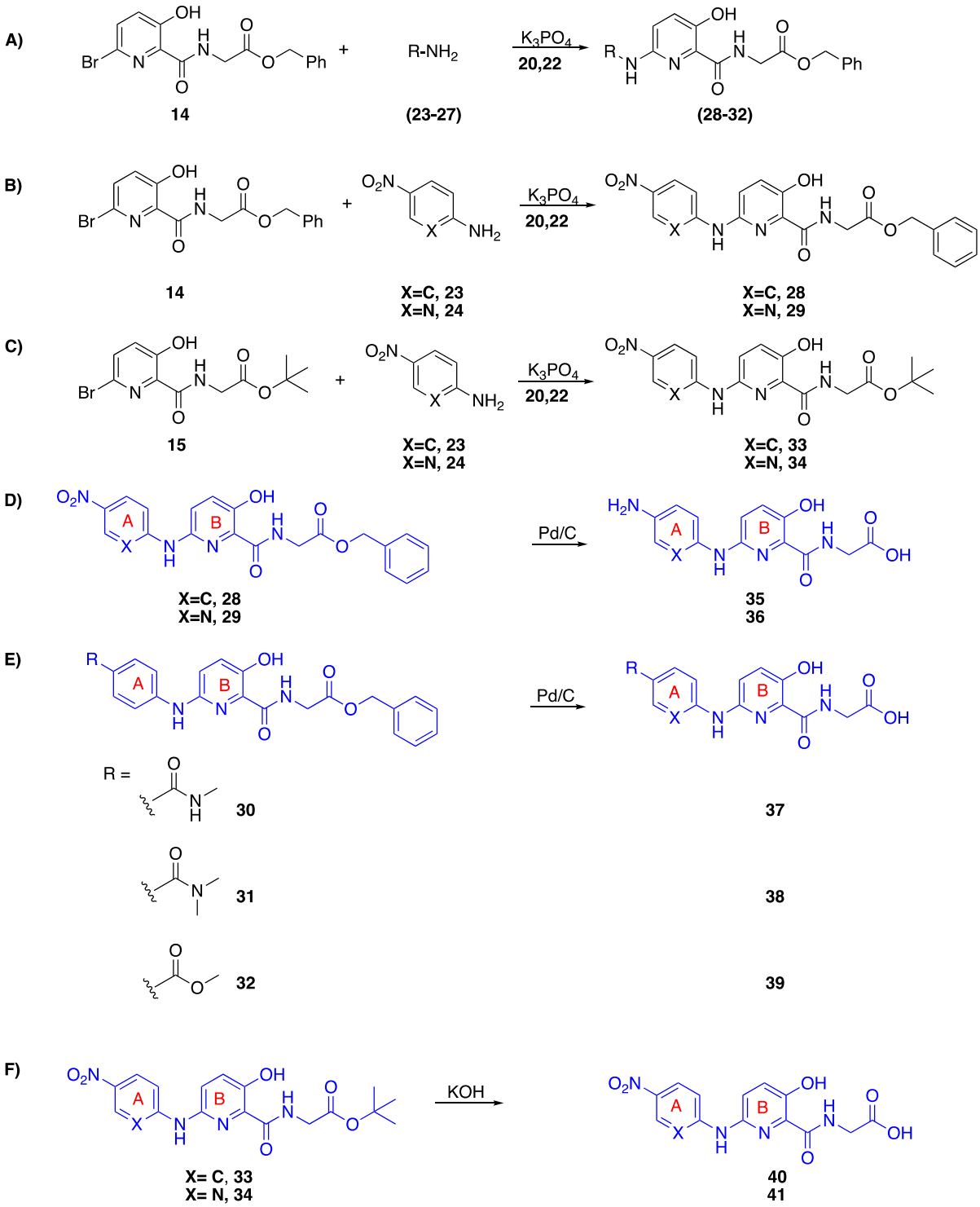


*Figure 2.18 Reaction Conditions: A) R-NH_2(1.1 eq), K_3PO_4(1.2 eq), 20(13.5 eq), **22**(9 eq),mw, 120°C, 8hr; B) **23** or **24**(1.1 eq), K_3PO_4(1.2 eq), **20**(13.5 eq), **22**(9 eq),mw, 120°C, 8hr; C) **23** or **24**(1.1 eq), K_3PO_4 or Cs_2CO_3 (1.2 eq), **20**(13.5 eq), **22**(9 eq),mw, 120°C, 8hr; D) Pd/C, H_2, THF, rt, 24hr; E) KOH(excess), THF, rt, 2days.*

Both benzyl or *t*-butyl ester derivatives of the glycinamide (**14** or **15**) were synthesised. The rationale behind the synthesis of two glycinamides was to investigate two different functional groups (-NO_2 and NH_2) at the C-4' position of ring A (Figure 2.18). It was envisaged that the reduction of **28** or **29** using palladium on charcoal (Pd/C) to the corresponding acid (**35** or **36**) would also reduce the nitro (-NO_2) group to the amine (Figure 2.18 D); however, the hydrolysis of the *t*-butyl group of **33** or **34** using KOH, would leave the –NO_2 group intact, to give **40** or **41** (Figure 2.18F). Consequently, both (–NO_2 and –NH_2) group, at the C-4' position of ring A (indicated in red in Figure 2.18D and 2.18E) to give **35** or **36**, and **40** or **41**, were investigated. The effect of changes in the aromatic ring (ring A is a phenyl ring in **35** or **40**, and pyridine in **36** or **41**) was also investigated. The C-4' position of the ring **A** was investigated via substitution with amide and ester (**37-39**). The role of steric hindrance at position C-4' of ring **A** was also investigated via substitution of a secondary or a tertiary amide at its C-4' position (**37** or **38**).

After the coupling of the protected glycinamides (**14** or **15**) with the amines (**23-27**), the final step was the deprotection of the resulting coupled esters (**28-34**). Pd/C under H_2 was used to deprotect the benzyl protected ester which resulted in the desired free acid (**35-39**). Initially, the deprotection of *t*-butyl ester was investigated in an acidic medium using trifluoroacetic acid (TFA) in dichloromethane (CH_2Cl_2). Different reaction conditions were investigated ranging from an excess of acid to longer reaction time. However, no deprotection of the *t*-butyl group was observed. Subsequently, basic hydrolysis of the ester (**33** or **34**) was investigated using potassium hydroxide (KOH) which resulted in successful conversion to the acid (**40** or **41**) after acidification of the inorganic layer. The possible mechanisms of the unusual base catalysed hydrolysis are as follows: the amide nitrogen (**42**) is deprotonated by base leading to the formation of oxide (**43**), which then attacks the electrophilic carbon of the

carbonyl group leading to the formation of oxazolone (**44**) which on hydrolysis lead to the formation of acid (**45**; Figure 3.19A). Alternatively, Elimination Unimolecular Conjugate Base (E1cB) elimination of *t*-butoxide via deprotonation of the glycine methylene may occur (Figure 3.19B). The final carboxylic acids (**35-41**) were then investigated for their activity against FTO, PHD2, histone demethylases (KDMs) and other 2OG oxygenases to investigate their selectivity for FTO.

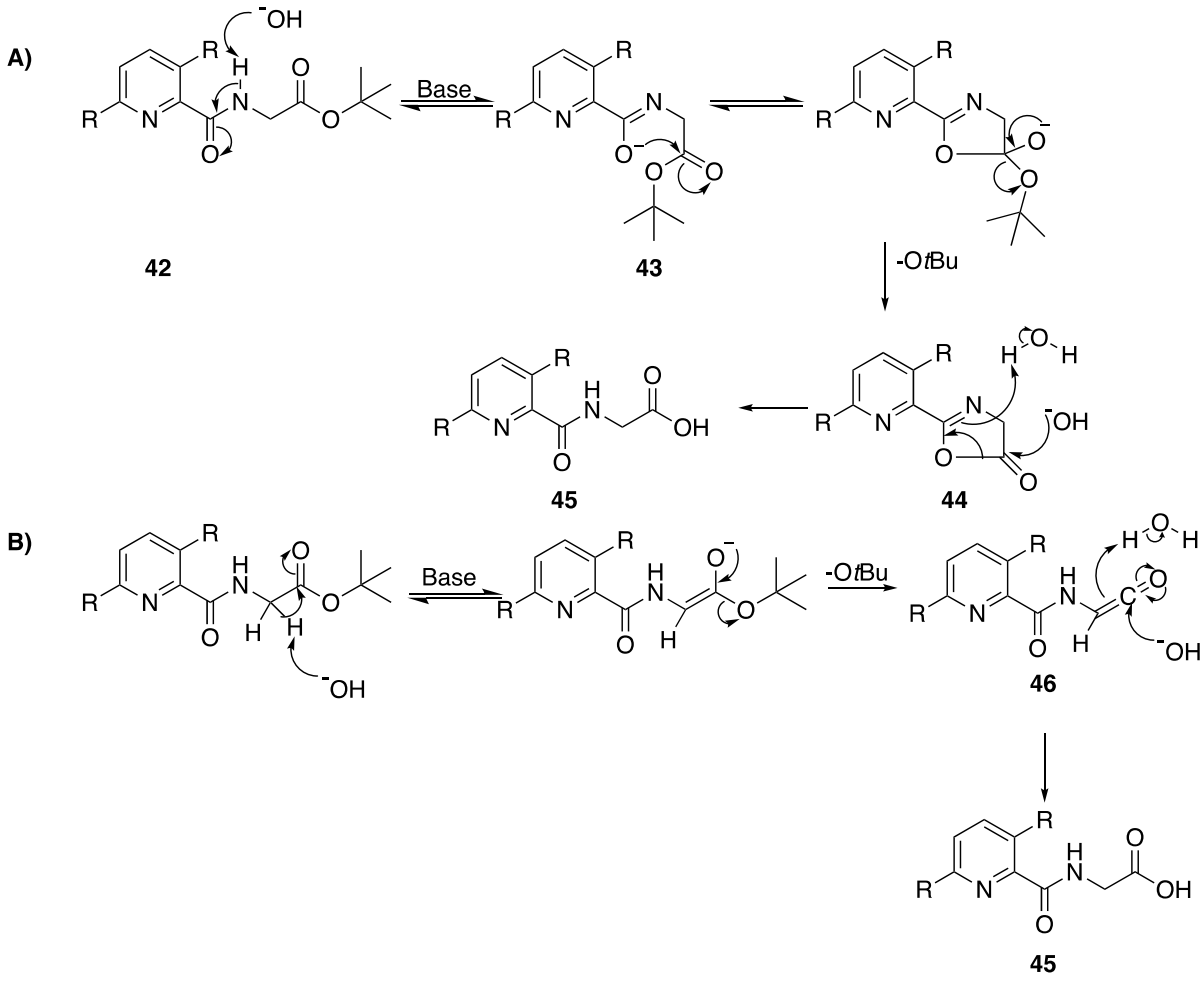


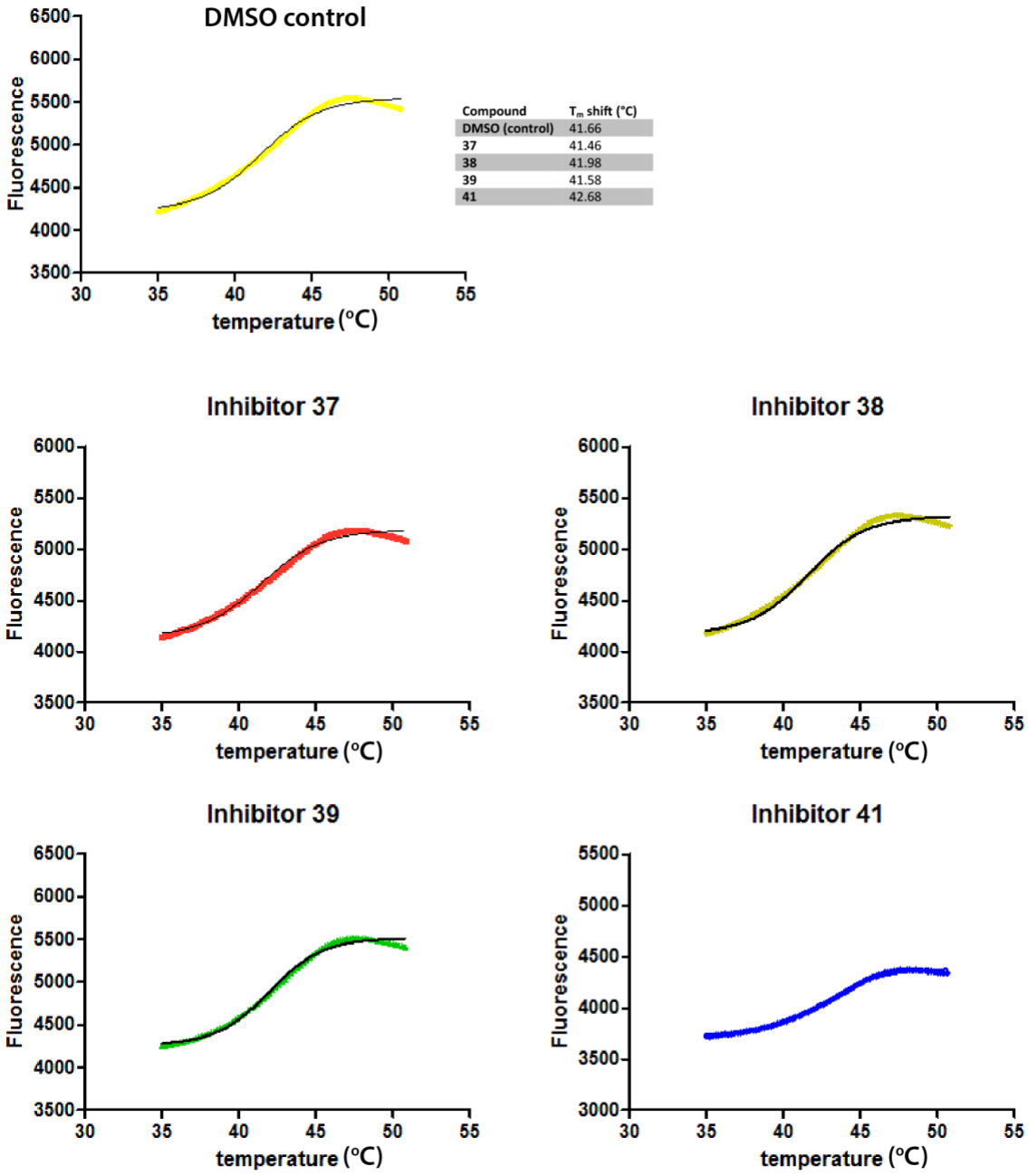
Figure 2.19 Proposed mechanism for hydrolysis of the t-butyl ester; A) involving a 5-exo-trig mechanism; B) involving an E1cB mechanism.

2.4.5 Differential Scanning Fluorimetry

Differential scanning fluorimetry (DSF) is an efficient method for screening low-molecular weight ligands that bind to and, thus stabilise, a protein.[45–47] DSF is based on the measurement of the fluorescence of the dye with strong affinity to hydrophobic regions of the protein.[45] On increasing temperature, the protein unfolds which led to the exposure of the hydrophobic core of the protein to the environment; the dye has a relatively strong affinity for the exposed hydrophobic region, thereby, increasing the fluorescence of the protein-dye complex relative to the non-complexed protein. It is generally, but not always, observed that the stronger the binding, the higher the melting temperature (T_m) shift. [45–47] However, the fluorescence measurement for T_m shift should not be the only criteria for a better ligand binding as some ligands have intrinsic fluorescence properties leading to the false positives. Further, for allosteric binders, an increase in the stability should not be the only criteria for a good inhibitor. Nevertheless, DSF can be a useful general method to investigate initial scaffolds for the inhibitor development.

In the present study, the DSF-assay was conducted by slowly increasing the temperature of the protein solution using MiniOpticon™ Real-time PCR Detection System (Bio-Rad), and the fluorescence reading was noted every 1°C in the range of 25-95°C. The excitation and emission wavelengths were 470 and 570 nm, respectively. The concentrations of the reagents used in the assay were as follows:[30,48] 2 µM FTO, 6× SYPRO orange dye (Invitrogen) (from commercial stock that is 5000×), 50 µM diammonium iron(II) sulfate, 200 µM inhibitors (**35-41**). All stock solutions were made in HEPES buffer, except for the Fe(II) stock solution which was prepared in 20 mM HCl. The T_m values were calculated by fitting the Boltzmann equation to the sigmoidal curves obtained using GraphPad Prism. However, the T_m shift of the FTO-dye-inhibitor (**35-41;** varies from 41.46°C- 42.78°C) tested was observed to be

similar to the FTO-dye-DMSO (41.66°C) control indicating that the binding of inhibitors may not significantly stabilise the enzyme.



Compound	T_m shift (°C)
DMSO (control)	41.66
37	41.46
38	41.98
39	41.58
41	42.68

*Figure 2.20 Some of the T_m shift plots for FTO-x-dye complex, where x = A) DMSO; B) inhibitor **37**; C) inhibitor **38**; D) inhibitor **39**; E) inhibitor **41**.*
It should be noted that the T_m shift assay was carried out under the guidance of Dong Zhang.

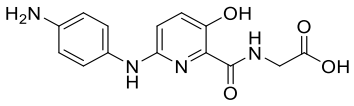
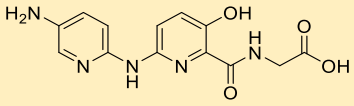
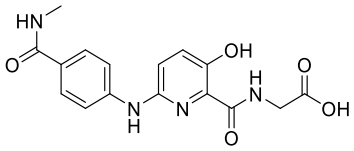
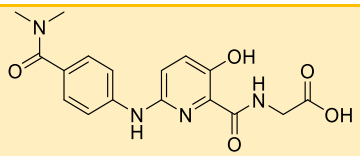
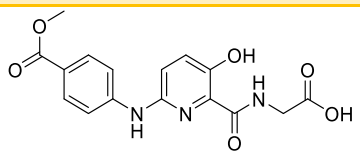
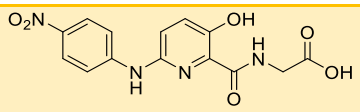
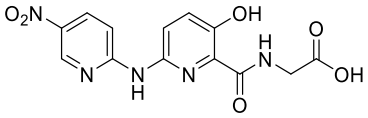
2.4.6 IC_{50} values

The IC_{50} value is defined as the concentration of inhibitor required to inhibit 50% of the enzymatic reaction at a specific substrate concentration. A resolved liquid chromatography-mass spectrometry (LCMS) based demethylation assay was used in my study, as reported.[30] The final concentrations of the reagents in the reaction mixture (25 µl) were as follows: 1.2 µM FTO; 22 µM 5-mer oligonucleotide [GG(m^6A)CU; synthesised in the lab]; 10 µM 2OG; 500 µM L-ascorbate; 10 µM diammonium iron(II) sulphate; and 25 mM Tris-HCl, pH 7.5, and the reaction mixture was incubated for 30 min at rt. The reaction mixture was quenched with 25 µl of methanol. The concentrations of the inhibitors were varied from 0 µM to 10 mM to determine IC_{50} values.

All the synthesised compounds (**35-41**) were tested against representative human 2OG oxygenases including PHD2, JMJD2C (KDM4C), JMJD3 (KDM6B) and were found to be at least 3 fold more selective for FTO over KDMs, as judged by the IC_{50} values comparison. The novel inhibitors (**35-41**) were found to be selective for FTO over PHD2, compared to **FG2216** and previously reported FTO inhibitors[30] (Figure 2.5) which lack selectivity. For most of the histone demethylases (KDMs) tested, the IC_{50} values were >100 µM. The most potent of all the tested inhibitors was **40**, which has a $–NO_2$ group attached to the C-4' position of ring **A** (ring A is a phenyl ring; Figure 2.18F). **40** could be a good candidate from which the optimisation of the cell-penetrating analogues, potentially with blood-brain barrier crossing ability, could be initiated. It should be noted that the difference between the inhibitors, **40** and **41,** is the presence of pyridyl ring in **41** (ring **A** is pyridine; Figure 2.18F) compared to the phenyl ring in **40**; however, the IC_{50} of **41** (7.8 µM) is 5-fold higher compared to **40** (1.3 µM). Thus, the pyridine analogue (**41**) with a $–NO_2$ group at the C-4' position of ring A was found to be less potent inhibitor for FTO compared to the phenyl

analogue (**40**). Crystal structures of **40** and **41**, in complex with FTO, are expected to provide insight into the mode of binding and the observed preference for selectivity. For inhibitors, **35** (29.3 μM) vs **36** (5.3 μM), a reverse mode of preference for FTO inhibition was observed, compared to **40 (**1.3 μM) vs **41** (8.0 μM); **40** is similar to **35** in structure (ring A is a phenyl ring), and **41** is similar to **36** in structure (ring A is pyridine). Thus, the pyridine analogue (**36**), with an amine at the C-4' position of ring A (Figure 2.18D), was observed to have a lower IC_{50} values towards FTO compared to the phenyl analogue (**35**; Figure 2.18E). The IC_{50} values of inhibitors, **36** (5.3 μM) and **39** (5.7 μM), towards FTO were found to be similar, despite differences in their structure; **36** (ring A is pyridine and amine group at the C-4' position of ring A; Figure 2.18D) and **39** (ring A is a phenyl with an ester group at the C-4' position of ring A; Figure 2.18E). The increase in steric hindrance at the C-4' position of the ring **A**, as in **37** (secondary amide; IC_{50} = 6.3 μM) compared to **38** (tertiary amide; IC_{50} = 37.6 μM), increases the IC_{50} of the inhibitors towards FTO by 6 fold. The inhibitor with an ester group at the C-4' position of ring A (**39**, Figure 2.18E; IC_{50} = 5.7 μM) compared to a secondary amide (**37**, Figure 2.18E; IC_{50} = 6.3 μM) showed increased potency towards FTO.

Table 2.2 Table showing the IC_{50} values of the inhibitors against JMJD2C, JMJD3, and FTO. For PHD2, the inhibition value is given as percentage inhibition at a 20μM concentration of inhibitor. The most potent inhibitor IC_{50} values are highlighted in red.

Inhibitor	Structure	IC_{50} (μM)			PHD2(% Inhibition at 20μM)
		JMJD2C (KDM4C)	JMJD3 (KDM6B)	FTO	
35		>100	89.7	29.3	25
36		63.9	>100	5.3	40
37		>100	47.8	6.3	48
38		>100	>100	37.6	53
39		>100	89.7	5.7	42
40		>100	19.5	1.3	43
41		44.4	38.4	8.0	44

It should be noted that the IC_{50} values for FTO were obtained by Dr. Marina Demetriades; JMJD2C and JMJD3 by Dr Anthony Tumber.

The rationale behind the observed preference and selectivity would be more apparent with the information from crystal structures of the inhibitors (**35-41**) in complex with FTO. The results reveal that **40** (IC_{50} of 1.3 μM), is the most potent inhibitor synthesised in the study. It should be noted that the observed IC_{50} is likely to be much higher than the K_I, *in vitro*, as the partially optimised assay was employed with the enzyme concentration of 1.2 μM.

2.4.7 Dissociation constants for the FTO inhibitors

Although the DSF assay was repeated twice, the T_m shift values did not provide supportive data towards the efficiency of binding of inhibitors (**35-41**) to the protein. Thus, it was considered desirable to obtain K_D (dissociation constant) values of the inhibitors to compare their relative strength of binding. K_D is defined as the concentration of the ligand at which half of the protein/ receptor sites are occupied with the ligand. Many binding assays have been used to measure the K_D values for 2OG oxygenases, including non-denaturing electron mass spectrometry (ESI-MS),[49–53] isothermal titration calorimetry (ITC),[54] and NMR.[54–56]

In this study, NMR was used as a tool to measure the binding constant of inhibitors of FTO. There are two general approaches to calculate K_D via NMR, namely, the ligand-based and the protein-based approaches.[55,57–64] The ligand-based approach is based on observing changes in the signals of the ligand [57,58,65] and, the protein-based approach is based on observing the changes in the signals of protein (often involving the use of isotopically labelled proteins).[63] The main drawback of the protein-based approach for FTO is a requirement for labelled protein. Thus, ligand-based direct binding was investigated in these studies. Ligand binding approaches to investigate binding can be categorised into two types as described below.

2.4.7.1 Direct Ligand Binding

When a ligand binds to a protein/ receptor, the environment around the ligand changes (Figure 2.21). These changes can be observed/measured in terms of changes in the chemical shift, line width, peak intensity, and longitudinal (R1) and transverse (R2) relaxation rates,[60,61] and can be monitored via various NMR active nuclei i.e. ^{1}H, ^{13}C, ^{19}F, and ^{31}P.[62,66–69]

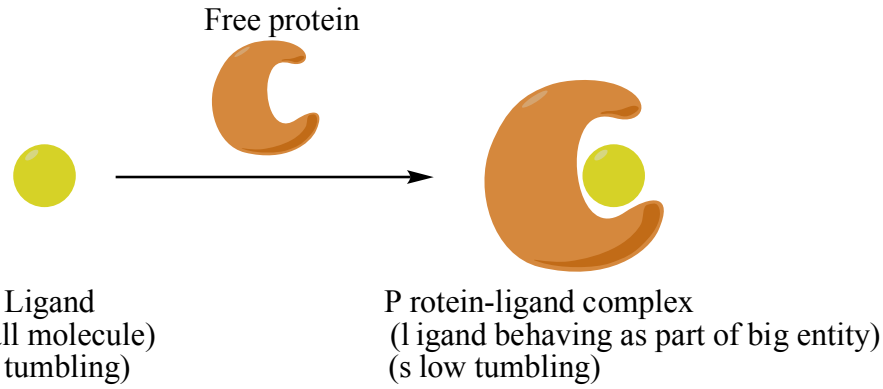


Figure 2.21 Schematic representation of protein-ligand interactions.

2.4.7.2 Reporter Ligand Binding

The Reporter Ligand Binding based approach can be used when the binding site in an enzyme and an established reporter ligand are known (Figure 2.22). It was postulated that the relative K_D value of the inhibitor/ competing ligand could be measured by observing/ monitoring the changes in the intensity peak of the reporter ligand or any other changes in the NMR parameter (chemical shift, relaxation rate).

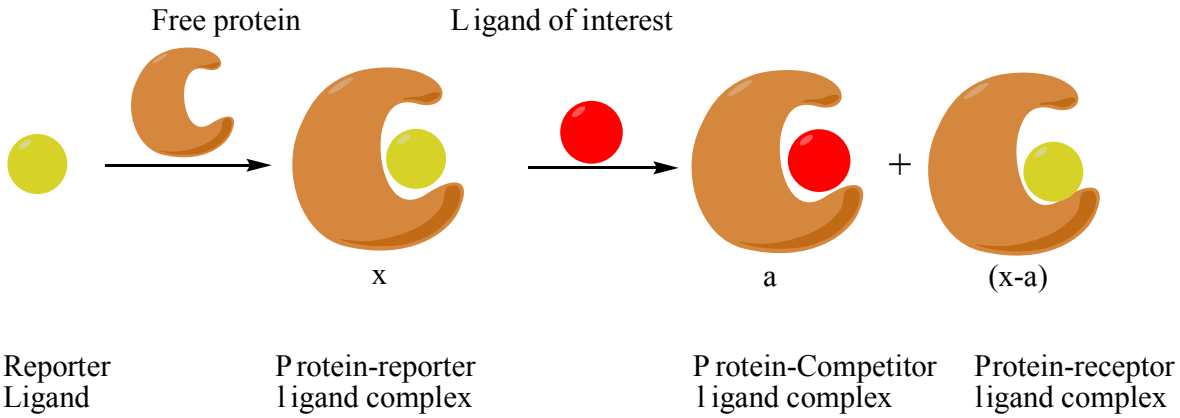


Figure 2.22 Schematic representation of the reporter ligand-based screening method.

In the case of 2OG oxygenases, it was envisaged that 2OG could be investigated as a reporter ligand for competition binding experiments.[8] Moreover, the inhibitors under investigation were designed to compete with the 2OG binding site. For the experiments, apo-FTO was produced (without Fe (II) in the binding site) to avoid non-catalytic turnover of 2OG to succinate and to help ensure the analyses were carried out with a single type of metal ion at the active site; Zn (II) was used as a divalent ion substitute for Fe (II), to inhibit the activity of the enzyme.[70]

2.4.8 Optimisation of the binding assay

The rationale behind the approach is that the loss/gain of the signal intensity of 2OG would report of the binding of a ligand (inhibitor) to FTO. It was envisaged that binding of 2OG to the protein would decrease the intensity of the signals of 2OG (δ_H 2.3 ppm and δ_H 2.8 ppm) in the ^{1}H NMR as some of the 2OG would be a part of the large molecule (protein+2OG) and tumble slowly relative to the NMR timescale. In the presence of a strong competitive binder, 2OG will be displaced into the solution (free state) leading to the regain/sharpening of the intensity of resonances at δ_H 2.3 ppm and δ_H 2.8 ppm.

By monitoring the intensity of the signals of displaced 2OG at different concentrations of inhibitors, the relative K_D of the inhibitors can be calculated using titration provided the dissociation constant of the reporter (2OG, K_D^{rep}) is known.[71–73] It is also possible to compare the relative binding strength of the inhibitors to FTO by correlating to the percentage of 2OG displaced at a specified concentration of inhibitors (**35-41**).

The percentage of 2OG displacement can be calculated as:

$$\% \text{ 2OG Displacement} = \frac{I_{2OG} - I_{2OG(0)}}{I_{2OG(blank)} - I_{2OG(0)}} \text{ X } 100\% \qquad \ldots 2.1$$

where,

I_{2OG} = Integral of 2OG in the presence of both protein and inhibitor;

$I_{2OG(0)}$ = Integral of 2OG in the presence of protein (FTO) and absence of inhibitor;

$I_{2OG(blank)}$ = Integral of 2OG in the absence of both protein (FTO) and inhibitor.

Initially, the K_D of 2OG (K_D^{rep}) was obtained by Dr Ivanhoe Leung for the Zn(II) complex of FTO, using the water relaxation experiment[74] and found to be 1.31 μM. After obtaining the K_D value for 2OG, attempts were then made to obtain the apparent K_D (K_D^{app}) values of the synthesised inhibitors (**35-41**) for FTO by using the methods of Morton. *et al.,*[75] and Dalvit. *et al.*,[64] involving equation 2.2.

$$K_D^{app} = \frac{[ET]\times[L] - [ET]\times[EL] + [EL]\times[EL] - [L]\times[EL]}{[EL]} \qquad \ldots 2.2$$

where,

[ET] = the total enzyme concentration;

[L] = the concentration of the inhibitor;

[EL] = the concentration of the enzyme−inhibitor complex.

From K_D^{app}, the K_D of an inhibitor can be obtained provided the dissociation constant of the reporter ligand (K_D^{rep}) is known.[71–73]

Initially, a ^{1}H NMR spectrum of a reaction mixture containing 2OG (10 μM) and Zn (II) (100 μM) in deuterated Tris (dTris, hereafter) buffer was acquired (Figure 2.23A). A ^{1}H NMR spectrum of a reaction mixture containing 2OG (10 μM), Zn (II) (100 μM), and FTO (10 μM) was acquired (Figure 2.23B).

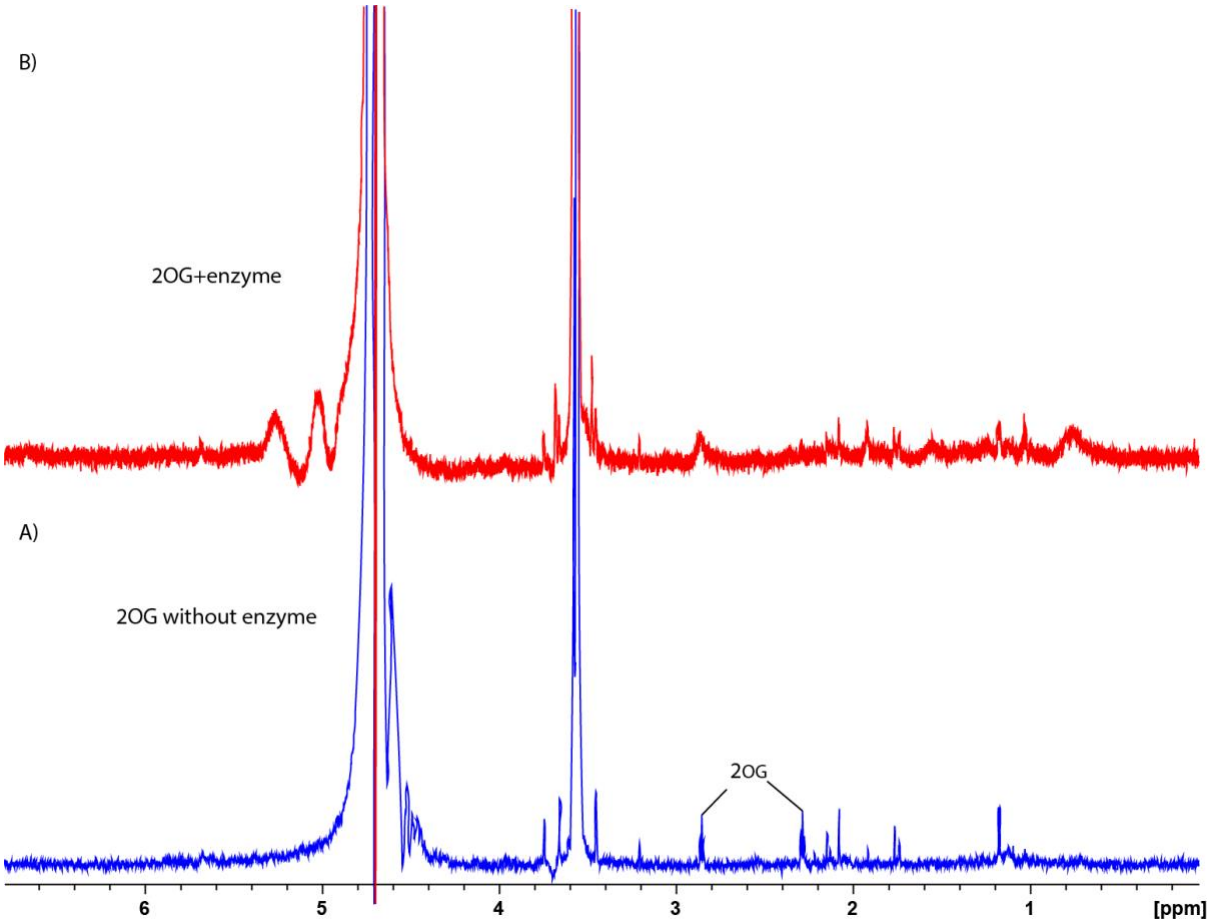


Figure 2.23 ^{1}H NMR (700 MHz) spectrum of a reaction mixture containing A) 2OG (10 μM) and Zn(II)(100 μM) at 27°C, pD-7.5 in 50mM dTris; B) 2OG (10 μM) and Zn(II)(100 μM) and FTO(10 μM) at 27°C, pD-7.5 in 50mM dTris. 2OG signals are indicated.

In the ^{1}H NMR spectrum of a reaction mixture containing 2OG and Z(II) (Figure 2.23A), 2OG was observed to resonate at δ_H 2.3 ppm and δ_H 2.8 ppm as two sharp triplets (Figure 2.23). As is evident from Figure 2.23B, in the ^{1}H NMR spectrum of a reaction mixture

containing 2OG, Zn(II) and FTO, signals corresponding to free 2OG disappear indicating that FTO forms a stable complex with 2OG that saturates at an ~1:1 ratio. It should be noted that the broadening of lines and interference of the signals from the protein was observed in the ^{1}H NMR indicating 2OG as a strong binder, thus, precluding the quantification of the 2OG signals (Figure 2.23B). Therefore, the Carl-Purcell-Meiboon-Gill, often known as CPMG, sequence was employed to attenuate the broad signals from the spectra; its modified version CPMG-PROJECT was employed for the analysis.[76–79]

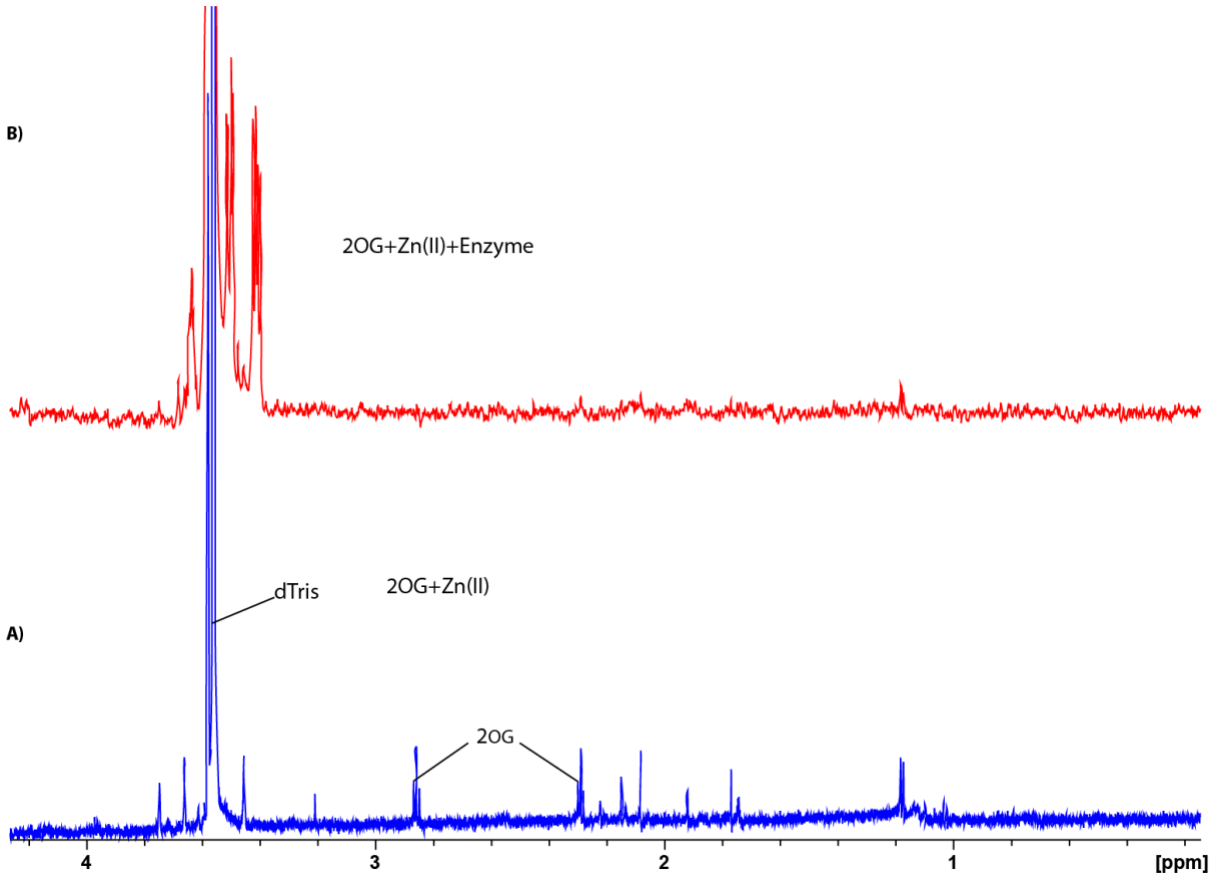


Figure 2.24 ^{1}H NMR (700 MHz) spectrum (CPMG sequence) of a reaction mixture containing A) 2OG (10 µM) and Zn(II) (100 µM) at 27°C, pD-7.5 in 50 mM dTris; B) 2OG (10 µM) and Zn(II)(100 µM) and FTO(10 µM) at 27°C, pD-7.5 in 50mM dTris. 2OG, dTris signals are indicated.

It should be noted that using the CPMG sequence attenuates the line broadening in the ^{1}H NMR spectrum of the FTO: 2OG complex (Figure 2.24B) compared to Figure 2.23B. After selecting the pulse sequence, work focussed on determining the apparent K_D of inhibitors

(**35-41**). Initially, all the inhibitors at 400 µM were screened against the solution containing 2OG (10 µM); FTO (10 µM); Zn(II) (100 µM) to eliminate weak binders, by comparing the amount of 2OG displaced into the solution from the protein:2OG complex, at 400 µM of inhibitors (Figure 2.25).

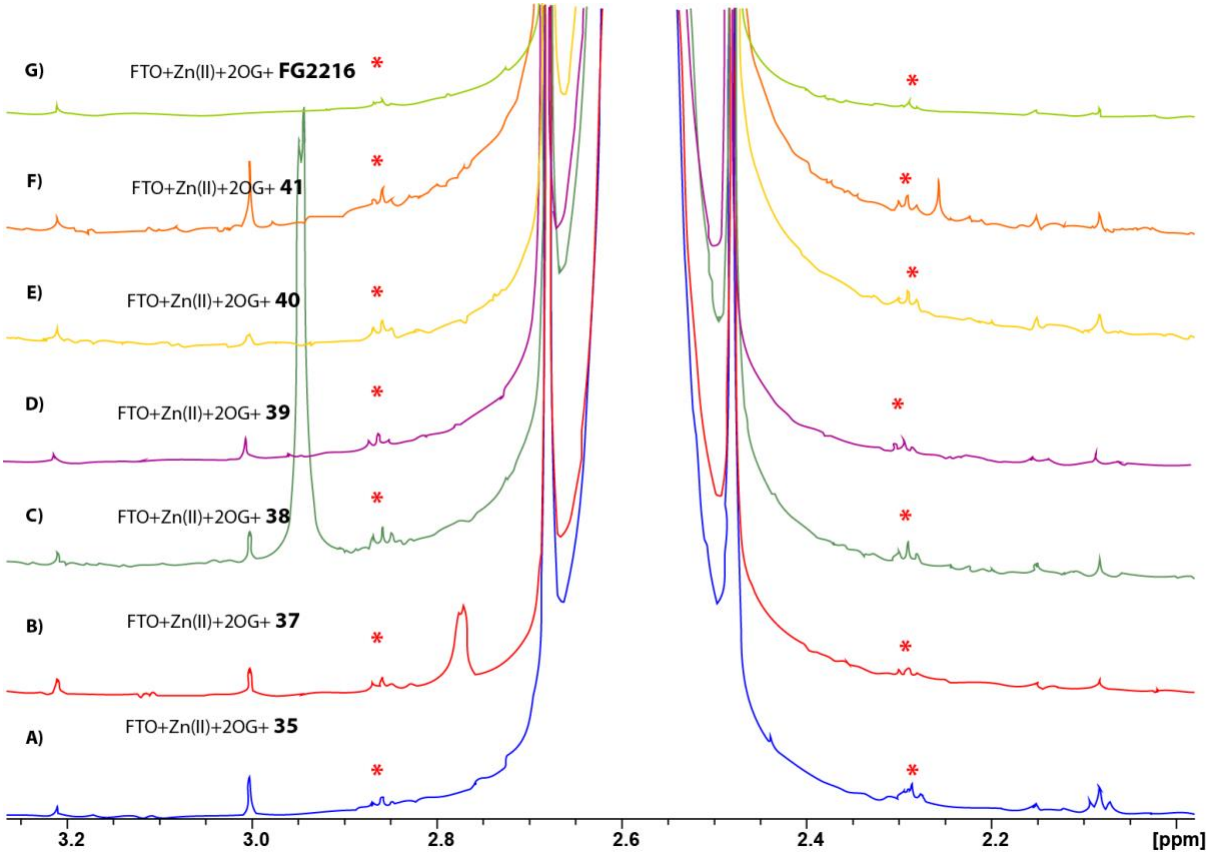


Figure 2.25 *^{1}H NMR (700 MHz) spectrum (CPMG sequence) of a reaction mixture containing 2OG (10 µM), Zn(II) (100 µM), FTO (10 µM) and 400 µM of A) **35**; B) **37**; C) **38**; D) **39**; E) **40**; F) **41**; G) **FG2216**; at 27°C, pD-7.5 in 50mM dTris. 2OG signals are indicated with asterisk (*) in red, at 27°C, pD-7.5 in 50mM dTris.*

From the ^{1}H NMR spectra of the inhibitors, it should be noted that, at 400 µM concentration, the synthesised inhibitors (**35-41**) displaced more 2OG into the solution compared to **FG2216**. Hence, under experimental conditions, inhibitors, **35-41,** are better likely binders to FTO as compared to **FG2216**. Next, **FG2216** (marked as **2** in Figure 2.2) was investigated for binding to FTO. For determining the apparent K_D value, the signals of displaced 2OG

were monitored at seven different concentrations of **FG2216** i.e. at 20 µM, 40 µM, 70 µM, 110 µM, 190 µM, and 350 µM.

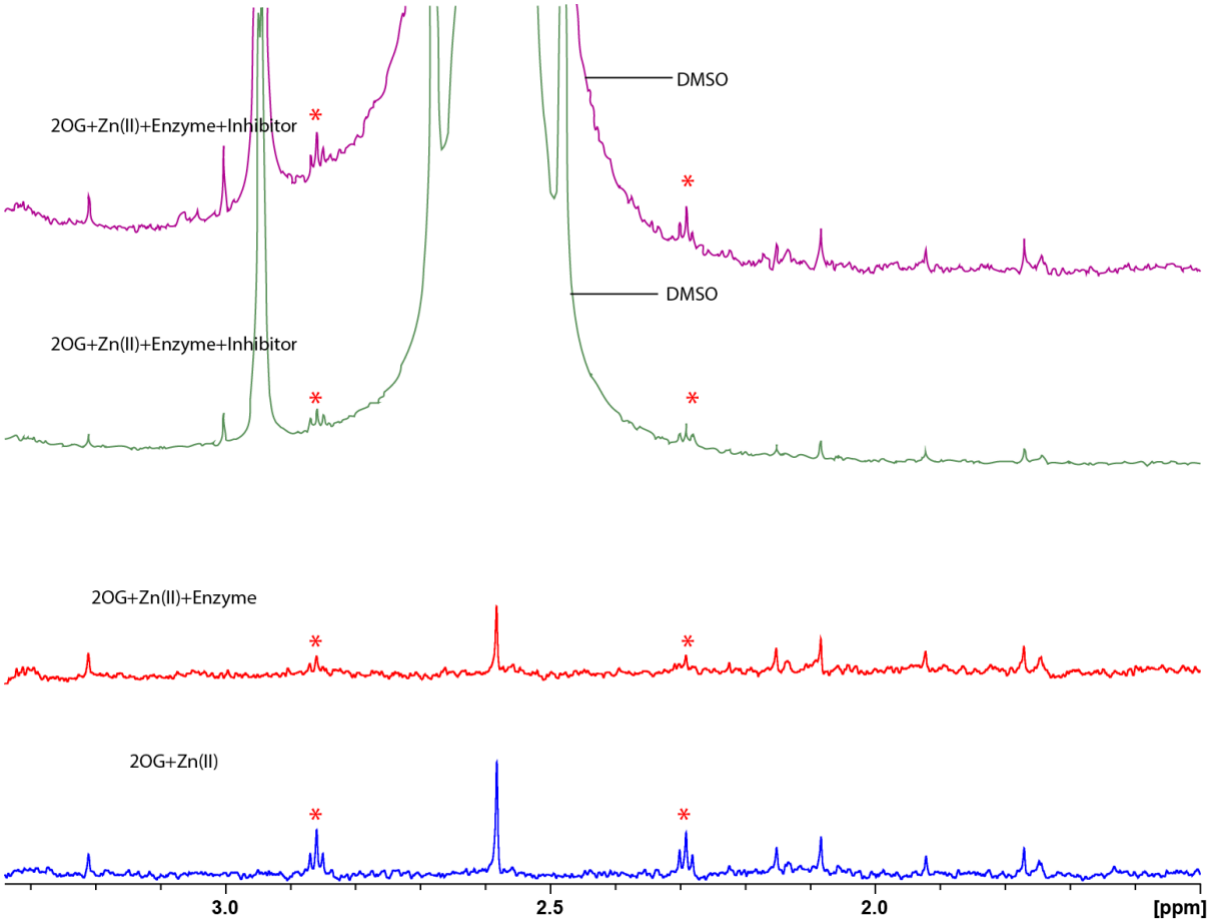


*Figure 2.26 ^{1}H NMR (700 MHz) spectrum (CPMG sequence) of a reaction mixture containing A) 2OG (10 µM) and Zn(II)(100 µM) at 27°C, pD-7.5 in 50mM dTris; B) 2OG (10 µM) and Zn(II)(100 µM) and FTO(10 µM) at 27°C, pD-7.5 in 50mM dTris; C) 2OG (10 µM), Zn(II) (100 µM, FTO(10 µM), **FG2216**(20 µM) at 27°C, pD-7.5 in 50mM dTris; D) 2OG (10 µM), Zn(II)(100 µM), FTO(10 µM) and **FG2216**(40 µM) at 27°C, pD-7.5 in 50mM dTris. 2OG signals are indicated by asterisk (*) in red. DMSO signals from inhibitors stock are also indicated.*

The resonance of 2OG at δ_H 2.8 ppm was observed to overlap with the resonances of the protein (FTO). Thus, the resonance at of 2OG at δ_H 2.3 ppm was selected for monitoring the concentration of 2OG in free state. The initial inhibitor stock was stored in DMSO. The close proximity of the large DMSO signals with the signal of interest (δ_H 2.3 ppm), which are relatively very small, precluded the accurate quantification of the signal (Figure 2.26C and

2.26D). It was envisaged that the inhibitors stock should be dissolved in minimum amount of deuterated DMSO and diluted further with MiliQ water to reduce the signals from DMSO. Therefore, the initial DMSO stock of inhibitors was heated and dried under N_2 overnight.

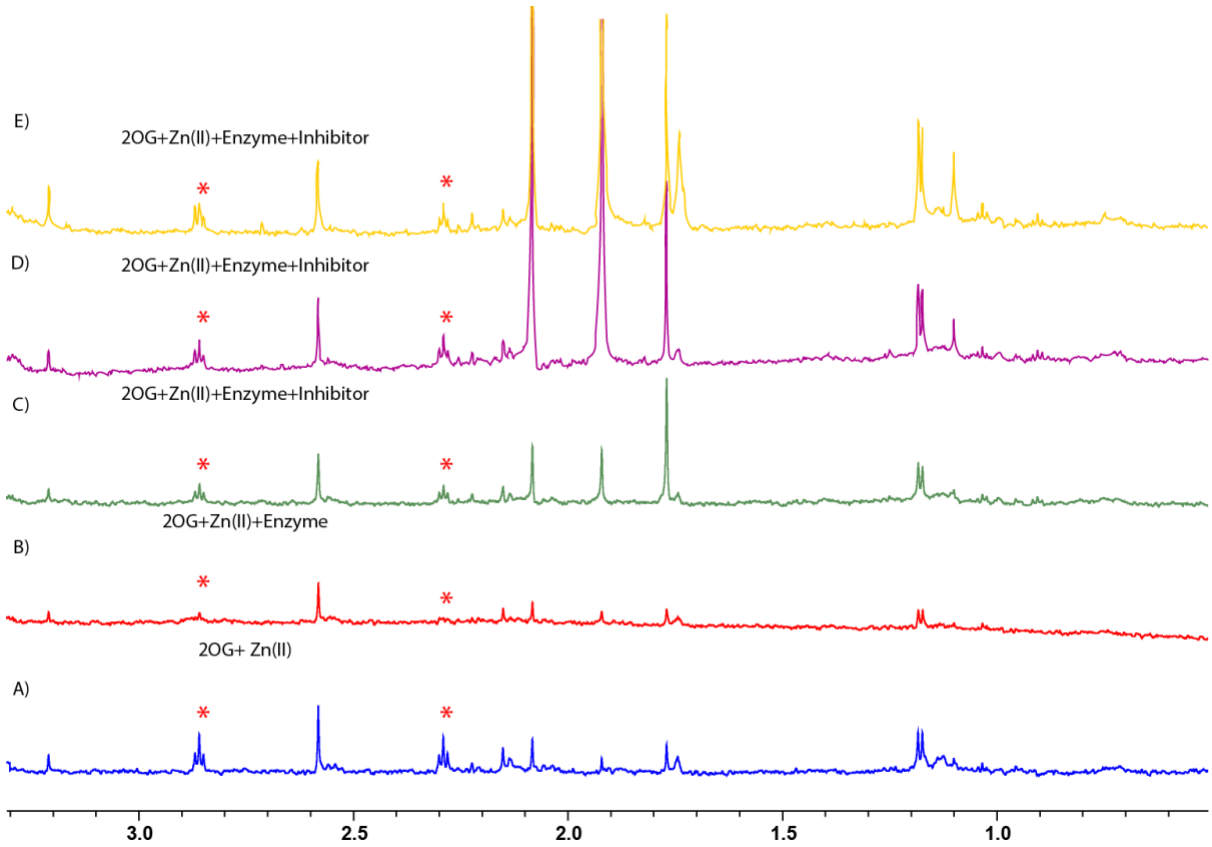


*Figure 2.27 ^{1}H NMR (700 MHz) spectrum (CPMG sequence) of a reaction mixture containing A) 2OG (10 µM) and Zn(II)(100 µM) at 27°C, pD-7.5 in 50mM dTris; B) 2OG (10 µM) and Zn(II)(100 µM) and FTO(10 µM) at 27°C, pD-7.5 in 50mM dTris; C) 2OG (10 µM), Zn(II) (100 µM), FTO(10 µM), **FG2216**(20 µM) at 27°C, pD-7.5 in 50mM dTris; D) 2OG (10 µM), Zn(II)(100 µM), FTO(10 µM) and **FG2216**(40 µM) at 27°C, pD-7.5 in 50mM dTris. E) 2OG (10 µM), Zn(II)(100 µM), FTO(10 µM) and **FG2216**(60 µM) at 27°C, pD-7.5 in 50mM dTris. 2OG signals are marked as *, highlighted in red. The reduction in the signal intensity of the deuterated DMSO signals from inhibitor stock should be noted.*

After optimising the DMSO concentrations, the ^{1}H NMR spectrum of the reaction mixture containing 2OG and Zn(II) was acquired to obtain the $I_{2OG(blank)}$ for the Equation 2.1. FTO was added to the reaction mixture and the ^{1}H NMR spectrum of the reaction mixture was acquired; the signal intensity of 2OG was recorded to obtain $I_{2OG(0)}$. Subsequently, varied amounts of inhibitor were titrated into the reaction mixture and the ^{1}H NMR spectra of the

reaction mixtures were acquired. Thus, I_{2OG} was integrated in the ^{1}H NMR at seven different concentrations of each inhibitor (**35-41**) and the percentage displacement of 2OG was calculated at respective concentrations of inhibitors. The 2OG displacement values were then fitted into a nonlinear fit using Origin 8.5 against the concentrations of inhibitor to obtain the percentage displacement of 2OG vs inhibitor concentration graph. The concentration at which half of the 2OG from the FTO: 2OG complex was displaced by the inhibitors was noted. The K_D^{app} of the inhibitor was obtained by nonlinear curve fitting using equation 2.2. The inhibitor **40** was then investigated. Analogous assay conditions were used and K_D^{app} for the inhibitors were calculated. Later, it was envisaged that increasing the concentration of FTO and 2OG to 20 µM each, would decrease the signal to noise ratio for 2OG resonances and also reduce the acquisition time for the ^{1}H NMR spectrum. Therefore, the studies were then conducted at increased concentrations of FTO and 2OG. The concentration of inhibitors under investigation and adjusted accordingly.

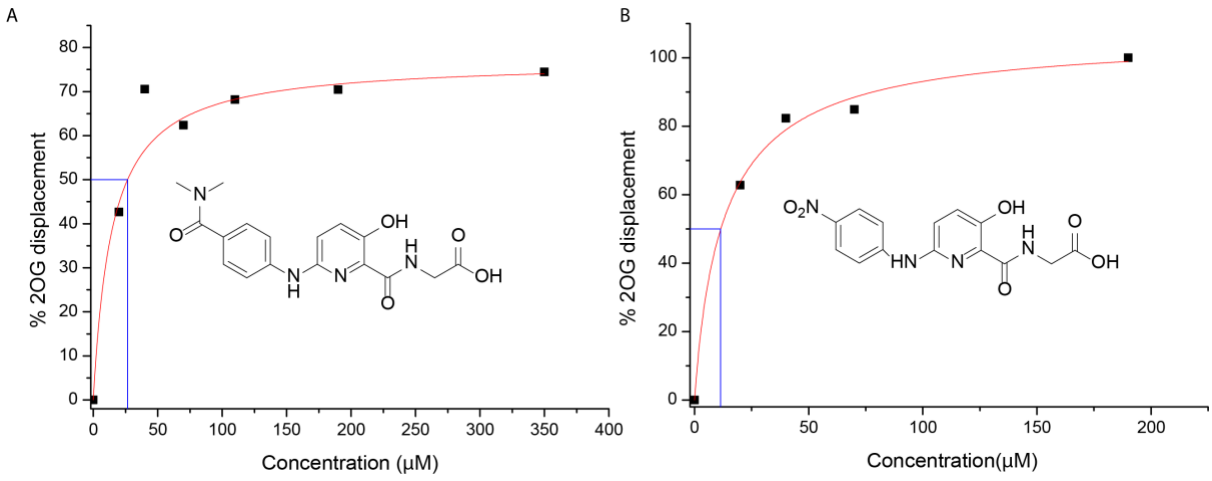


*Figure2.28 Plot showing single data point titration curve of Inhibitors A) **38**; B) **40** against FTO.*

It was observed while analysing the single data point graphs (Figure 2.28) that most of the 2OG from the protein-2OG complex was displaced at very low concentrations of the inhibitors (**38** and **40**, in this case) and reached saturation after the second concentration point

(40 µM) i.e. the concentration of 2OG displaced into the solution remained constant after the second concentration point (40 µM). It was then envisaged that the studies should be conducted at lower concentrations of inhibitors to get more data points before the graph reaches saturation/plateau region. Hence, the stock solution of the inhibitor was diluted to 3 µM compared to previous 7.5 µM. Inhibitors were titrated to the final concentration of 10 µM, 20 µM, 30 µM, 50 µM, 70 µM, and 90 µM into the reaction mixture containing FTO (20 µM), 2OG (20 µM) and Zn(II) (100 µM) (single point data of **37** and **39** is shown in Figure 2.29).

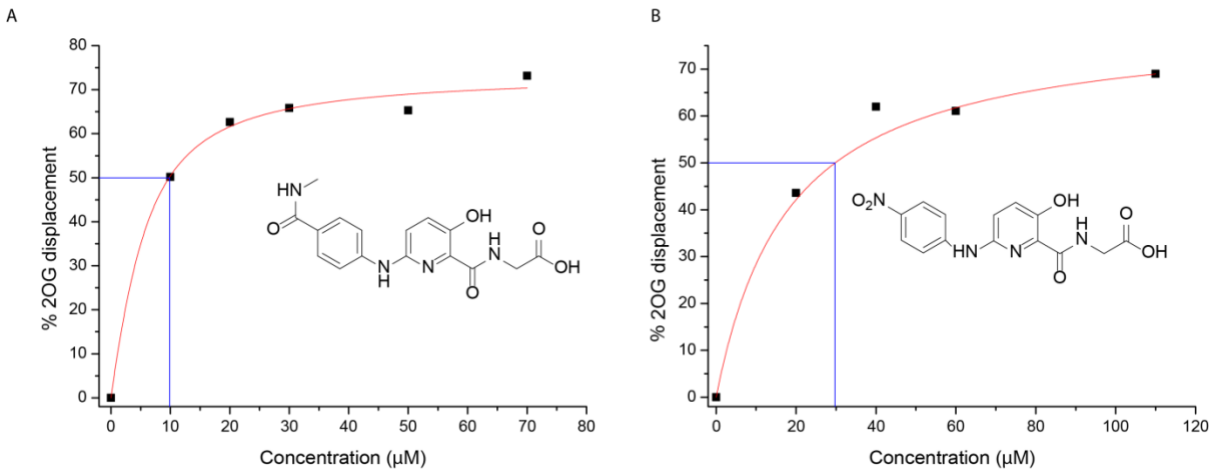


*Figure 2.29 Plot showing single data point titration curve of Inhibitors A) **37**; B) **39** against FTO*

Finally, the assay was conducted in triplicate for the four inhibitors, namely **FG2216**, **37**, **38** and **40** (Figure 2.30).

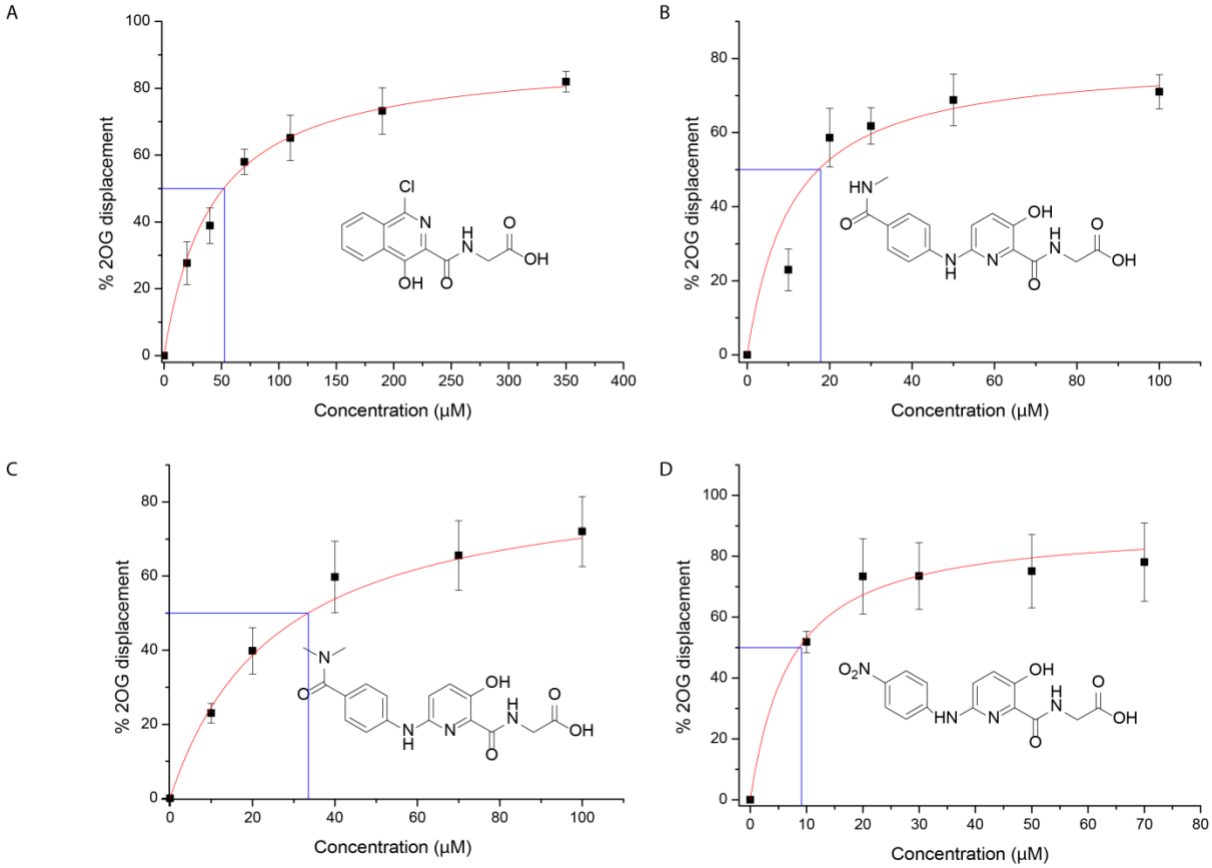


Figure2.30 Plot showing titration curve of Inhibitors A) ***FG2216****; B)* ***37****; C)* ***38****; D)* ***40*** *against FTO*

It should be noted that the attempts to obtain titration curve for the inhibitor **35** were unsuccessful owing to the presence of an impurity at δ_H 2.3 ppm in the stock solution, which overlapped with the resonance of 2OG at δ_H 2.3 ppm, thus, eluding the quantification of the displaced 2OG resonances.

The K_D^{app} values calculated from the graphs were found to be approximately 52 μM, 18 μM, 34 μM, 30 μM, and 9 μM, respectively for inhibitors **FG2216**, **37**, **38**, **39** and **40**. The actual K_D for the inhibitors can be calculated by using the equations described[64,71,72,75] and used in the literature.[80]

Table 2.3 Table showing the binding constants (K_D) for the Catalytic domain of FTO, obtained by using the procedure described in the literature.[80] The errors shown are the standard deviation from three separate measurements.

Compound	**K_D values (nM)**
FG2216	7000±2000
37	45±25
38	600±200
39	50 (single point)
40	54±30

2.5 Conclusions and Future Work

The work described in this chapter aimed at the synthesis of the first generation FTO inhibitors based on the structure-based relation (SAR) studies conducted previously in Oxford. The work also focussed on the optimisation of the synthetic scheme for the synthesis of novel compounds. One important objective of the work was to demonstrate the possibility of selectivity in inhibition for FTO over other 2OG oxygenases. Work was also focussed on optimising the binding assay for FTO to obtain the relative K_D values of the inhibitor, apart from the DSF based assay previously developed in the laboratory.

The Hartwig-Buchwald coupling conditions were successfully optimised under microwave conditions. Three different bases namely, Cs_2CO_3, LiHMDS, and K_3PO_4 were investigated and K_3PO_4 was then selected as the base for the studies. Based on the SAR studies and crystal structure analysis, a potent inhibitor for FTO, **40,** with an IC_{50} of 1.3 µM and a 15-fold selectivity over PHD2, was synthesised. The K_D^{app} value of **40** via NMR binding assay, was found to be 9 µM (actual K_D 54±30 nM). The ester analogue of **40** was found to be non-toxic in cells. The next most potent inhibitors were found to be **39** and **37** with IC_{50} values of

5.7 μM and 6.3 μM, respectively, and reasonably good K_D^{app} values of 34 μM and 30 μM, respectively. The increase in steric hindrance at the C-4' position of the ring **A** apparently resulted in the increase in the IC_{50} values of the inhibitor, observed via substitution of secondary amide vs tertiary amide (**37**, IC_{50} 6.3 μM; **38**, IC_{50} 37.6 μM). The IC_{50} value of a compound is dependent on the amount of substrate used in the assay; the concentration of substrate used in the assay was 22 μM; it is possible that the actual K_I values may be lower than the IC_{50} values observed in the assay used.

To conclude, SAR studies and modelling have been used to design and then synthesise compounds as potential FTO inhibitors. In the key synthetic step, the Buchwald-Hartwig reaction was optimised for the electron deficient amines used in this study and seven inhibitors were synthesised. The IC_{50} values (by LCMS) and binding constants (K_D, by NMR), T_m shifts for the inhibitors were acquired. The results direct the design of the second generation of compounds that will incorporate selectivity and better pharmacological properties.

Crystal structures of FTO in complex with inhibitors will also be sought to identify the inhibitor binding modes and to inform further inhibitor development. Once a suitably active and selective compound *in vitro* has been developed, work will then focus on assessing inhibitory activity in cells using immunofluorescence and potentially translation efficiency and qPCR (quantitative polymerase chain reaction) experiments.

2.6 References

1. Fischer, J. Koch, L.,Emmerling, C., Vierkotten, J., Peters, T., Brüning, J. C., Rüther, U. Inactivation of the Fto gene protects from obesity. *Nature* **458,** 894–898 (2009).

2. Gerken, T. Girard, Christophe A., Tung, Y. C. L., Webby, C. J., Saudek, V., Hewitson, K. S.,Yeo, Giles S.H., McDonough, M. A., Cunliffe, S., McNeill, L. A., Galvanovskis, J., Rorsman, P., Robins, P., Prieur, X., Coll, A. P., Ma, M., Jovanovic, Z., Farooqi, I. S., Sedgwick, B., Barroso, I., Lindahl, T., Ponting, C. P., Ashcroft, F. M., O'Rahilly, S., Schofield, C. J. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318,** 1469–72 (2007).

3. Church, C., Lee, S., Bagg, E. A L, McTaggart, J. S., Deacon, R., Gerken, T., Lee, A., Moir, L., Mecinović, J. M., Schofield, C. J., Ashcroft, F. M., Cox, Roger D. A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. *PLoS Genet.* **5,** (2009).

4. Wehr, E., Schweighofer, N., Möller, R., Giuliani, A., Pieber, T. R., Obermayer-Pietsch, B. Association of FTO gene with hyperandrogenemia and metabolic parameters in women with polycystic ovary syndrome. *Metabolism.* **59,** 575–580 (2010).

5. Keller, L., Xu, W., Wang, Hui X., Winblad, B., Fratiglioni, L., Graff, C. The obesity related gene, FTO, interacts with APOE, and is associated with Alzheimer's disease risk: a prospective cohort study. *J. Alzheimers. Dis.* **23,** 461–9 (2011).

6. Da Cunha, P. A., De Carlos Back, L. K., Sereia, A. F. R., Kubelka, C., Ribeiro, M. C. M., Fernandes, B. L., De Souza, I. R. Interaction between obesity-related genes, FTO and MC4R, associated to an increase of breast cancer risk. *Mol. Biol. Rep.* **40,** 6657–6664 (2013).

7. Kaklamani, V., Yi, N., Sadim, M., Siziopikou, K., Zhang, K., Xu, Y., Tofilon, S., Agarwal, S., Pasche, B., Mantzoros, C. The role of the fat mass and obesity associated gene (FTO) in breast cancer risk. *BMC Med. Genet.* **12,** 52 (2011).

8. Loenarz, C. & Schofield, C. J. Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat. Chem. Biol.* **4,** 152–156 (2008).

9. Rose, N. R., Ng, S. S., Mecinović, J., Liénard, B. M R., Bello, S. H. Sun, Z., McDonough, M. A., Oppermann, U., Schofield, C.J. Inhibitor Scaffolds for 2-Oxoglutarate-Dependent Histone Lysine Demethylases . *J. Med. Chem.* **51,** 7053–7056 (2008).

10. Schofield, C. J. & Zhang, Z. Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr. Opin. Struct. Biol.* **9,** 722–731 (1999).

11. Clifton, I. J., McDonough, M. A., Ehrismann, D., Kershaw, N. J., Granatino, N., Schofield, C. J. Structural studies on 2-oxoglutarate oxygenases and related double-stranded β-helix fold proteins. *J. Inorg. Biochem.* **100,** 644–669 (2006).

12. Aik, W., McDonough, M. A., Thalhammer, A., Chowdhury, R. & Schofield, C. J. Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr. Opin. Struct. Biol.* **22,** 691–700 (2012).

13. Rose, N. R., McDonough, M. A., King, O. N. F., Kawamura, A. & Schofield, C. J. Inhibition of 2-oxoglutarate dependent oxygenases. *Chem. Soc. Rev.* **40,** 4364 (2011).

14. Chen, H., Giri, Nitai C., Zhang, R.,Yamane, K., Zhang, Y., Maroney, M., Costa, M. Nickel Ions Inhibit Histone Demethylase JMJD1A and DNA Repair Enzyme ABH2 by Replacing the Ferrous Iron in the Catalytic Centers. *J. Biol. Chem.* **285,** 7374–7383 (2010).

15. Sekirnik, R., Rose, N. R., Mecinovic, J. & Schofield, C. J. 2-Oxoglutarate oxygenases are inhibited by a range of transition metals. doi:10.1039/c004952b

16. Chen, H., Ke, Q., Kluz, T., Yan, Y. & Costa, M. Nickel Ions Increase Histone H3 Lysine 9 Dimethylation and Induce Transgene Silencing. *Mol. Cell. Biol.* **26,** 3728–3737 (2006).

17. Hirsila, M., Koivunen, P., Xu, L., Seeley, T., Kivirikko, K. I., Myllyharju, J. Effect of desferrioxamine and metals on the hydroxylases in the oxygen sensing pathway. *FASEB J.* **19,** 1308–10 (2005).

18. Conejo-Garcia, A. Structural basis for binding of cyclic 2-oxoglutarate analogues to factor-inhibiting hypoxia-inducible factor. *Bioorg. Med. Chem. Lett.* **20,** 6125–6128 (2010).

19. Hamada, S. Design, synthesis, enzyme-inhibitory activity, and effect on human cancer cells of a novel series of jumonji domain-containing protein 2 histone demethylase inhibitors. *J. Med. Chem.* **53,** 5629–38 (2010).

20. Linden, T. The antimycotic ciclopirox olamine induces HIF-1α stability, VEGF expression, and angiogenesis. *FASEB J.* **17,** 761–3 (2003).

21. Mole, D. R. 2-oxoglutarate analogue inhibitors of HIF prolyl hydroxylase. *Bioorg. Med. Chem. Lett.* **13,** 2677–80 (2003).

22. Hopkinson, R., King, O. N. F. Quantitative High-Throughput Screening Identifies 8-Hydroxyquinolines as Cell-Active Histone Demethylase Inhibitors. *PLoS One* **5,** (2010).

23. Sakurai, M. Rose, Nathan R., Schultz, L., Quinn, Amy M., Jadhav, A., Ng, Stanley S., Oppermann, U., Schofield, C. J.,Simeonov, A. A miniaturised screen for inhibitors of Jumonji histone demethylases. *Mol. Biosyst.* **6,** 357–64 (2010).

24. Schlemminger, I. Mole, David R., McNeill, L.A., Dhanda, A., Hewitson, K. S., Tian, Y. M., Ratcliffe, P. J., Pugh, C. W.,Schofield, C.J.. Analogues of dealanylalahopcin are inhibitors of human HIF prolyl hydroxylases. *Bioorg. Med. Chem. Lett.* **13,** 1451–4 (2003).

25. Welford, R. W. D., Schlemminger, I., McNeill, L. A., Hewitson, K. S. & Schofield, C. J. The Selectivity and Inhibition of AlkB. *J. Biol. Chem.* **278,** 10157–10161 (2003).

26. Zhou, Y.-D. Hypoxia-Inducible Factor-1 Activation by (−)-Epicatechin Gallate: Potential Adverse Effects of Cancer Chemoprevention with High-Dose Green Tea Extracts. *J. Nat. Prod.* **67,** 2063–2069 (2004).

27. Jeon, H., Kim, H., Choi, D., Kim, D., Park, S., Kim, Y., Kim, Young M., Jung, Y. Activates an Angiogenic Pathway, Hypoxia Inducible Factor (HIF)-1-Vascular Endothelial Growth Factor, by Inhibiting HIF-Prolyl Hydroxylase: a Structural Analysis of Quercetin for Inhibiting HIF-Prolyl Hydroxylase. *Mol. Pharmacol.* **71,** 1676–1684 (2007).

28. Han, Z., Niu, T., Chang, J., Lei, X., Zhao, M., Wang, Q., Cheng, W., Wang, J., Feng, Y., Chai, J. Crystal structure of the FTO protein reveals basis for its substrate specificity. *Nature* **464,** 1205–1209 (2010).

29. Chen, B., Ye, F., Yu, L., Jia, G., Huang, X., Zhang, X., Peng, S., Chen, K., Wang, M., Gong, S., Zhang, R., Yin, J. L., H.,Yang, Y., Liu, H., Zhang, J., Zhang, H., Zhang, A., Jiang, H., Luo, C., Yang, C. G.Development of cell-active N6-methyladenosine RNA demethylase FTO inhibitor. *J. Am. Chem. Soc.* **134,** 17963–17971 (2012).

30. Aik, W. Demetriades, M., Hamdan, M. K. K., Bagg, E. A L., Yeoh, K K., Lejeune, C., Zhang, Z., McDonough, M. A., Schofield, C. J. Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO). *J. Med. Chem.* **56,** 3680–3688 (2013).

31. Huang, Y., Yan, J., Li, Q., Li, J., Gong, S., Zhou, H., Gan, J., Jiang, H., Jia, Gui F., Luo, C.,Yang, C. G.Meclofenamic acid selectively inhibits FTO demethylation of m 6 A over ALKBH5. *Nucleic Acids Res.* **43,** 373–384 (2015).

32. Qiao, Y. Zhou, B., Zhang, M., Liu, W., Han, Z., Song, C., Yu, W., Yang, Q., Wang, R., Wang, S., Shi, S., Zhao, R., Chai, J., Chang, J. A Novel Inhibitor of the Obesity-Related Protein FTO. *Biochemistry* **55,** 1516–1522 (2016).

33. Zheng, G., Cox, T., Tribbey, L., Wang, G. Z. Iacoban, P., Booher, M., E. Gabriel, G., J. Zhou, L., Bae, N.,Rowles, J., He, C., Olsen, M. J. Synthesis of a FTO Inhibitor with Anticonvulsant Activity. *ACS Chem. Neurosci.* **5,** 658–665 (2014).

34. Gao, Y.-G., Woon, E.C Y, Toh, J. D W, Sun, L., Lau, L.Z M., Tan, J., Low, J. J A., Tang, Colin W Q., Cheong, Eleanor J Y., Tan, Melissa J H., Chen, Y., Hong, W. A strategy based on nucleotide specifi city leads to a subfamily-selective and cell-active inhibitor of N 6 -methyladenosine demethylase FTO A strategy based on nucleotide specificity leads to a subfamily-selective and cell-active inhibitor of N 6 -methyl. *Chem. Sci.* **6,** 1–844 (2015).

35. Li, Q. Rhein inhibits AlkB repair enzymes and sensitizes cells to methylated DNA damage. *J. Biol. Chem.* **291,** 11083–11093 (2016).

36. He, W.. Identification of A Novel Small-Molecule Binding Site of the Fat Mass and

Obesity Associated Protein (FTO). *J. Med. Chem.* **58,** 7341–7348 (2015).

37. McMurray, F. Pharmacological Inhibition of FTO. *PLoS One* **10,** e0121829 (2015).

38. Gildner, P. G. & Colacot, T. J. Reactions of the 21st Century: Two Decades of Innovative Catalyst Design for Palladium-Catalyzed Cross-Couplings. *Organometallics* **34,** 5497–5508 (2015).

39. Martin, R. & Buchwald, S. L. Palladium-Catalyzed Suzuki-Miyaura Cross-coupling Reactions Employing Dialkylbiaryl Phosphine Ligands. doi:10.1021/ar800036s

40. Surry, D. S. Dialkylbiaryl phosphines in Pd-catalyzed amination: a user's guide. *Chem. Sci.* **2,** 27–50 (2011).

41. Hanhan, M. E. Ligand effects in palladium-catalyzed Suzuki and Heck coupling reactions. doi:10.1002/aoc.1389

42. Co LLC, S.-A. Application Guide – Phosphine Ligands Phosphine Ligand Application Guide.

43. Bruno, N. C., Tudge, M. T. & Buchwald, S. L. Design and preparation of new palladium precatalysts for C–C and C–N cross-coupling reactions. *Chem. Sci.* **4,** 916–920 (2013).

44. and, D.-Y. L. & Hartwig, J. F. Zinc Trimethylsilylamide as a Mild Ammonia Equivalent and Base for the Amination of Aryl Halides and Triflates. (2005). doi:10.1021/OL050141B

45. Pantoliano, M. W. High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery. *J. Biomol. Screen.* **6,** 429–440 (2001).

46. Lo, M.-C. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem.* **332,** 153–159 (2004).

47. Holdgate, G. A. & Ward, W. H. J. Measurements of binding thermodynamics in drug discovery. *Drug Discov. Today* **10,** 1543–1550 (2005).

48. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2,** 2212–2221 (2007).

49. Hewitson, K. S.. Structural and Mechanistic Studies on the Inhibition of the Hypoxia-inducible Transcription Factor Hydroxylases by Tricarboxylic Acid Cycle Intermediates. *J. Biol. Chem.* **282,** 3293–3301 (2007).

50. McNeill, L. A. Flashman, E., Buck, M. R. G., Hewitson, K. S., Clifton, I. J., Jeschke, G., Claridge, T. D. W., Ehrismann, D., Oldham, N. J. Schofield, C. J.Hypoxia-inducible factor prolyl hydroxylase 2 has a high affinity for ferrous iron and 2-oxoglutarate. *Mol. Biosyst.* **1,** 321 (2005).

51. Mecinović, J. ESI-MS Studies on Prolyl Hydroxylase Domain 2 Reveal a New Metal Binding Site. *ChemMedChem* **3,** 569–572 (2008).

52. Mecinović, J., Loenarz, C., Chowdhury, R. & Schofield, C. J. 2-Oxoglutarate analogue inhibitors of prolyl hydroxylase domain 2. *Bioorg. Med. Chem. Lett.* **19,** 6192–6195 (2009).

53. Rose, N. R., Schofield, C. J.Selective Inhibitors of the JMJD2 Histone Demethylases: Combined Nondenaturing Mass Spectrometric Screening and Crystallographic Approaches. *J. Med. Chem.* **53,** 1810–1818 (2010).

54. Bleijlevens, B. Dynamic states of the DNA repair enzyme AlkB regulate product release. *EMBO Rep.* **9,** 872–7 (2008).

55. Leung, I. K. H., Flashman, E., Yeoh, K. K., Schofield, C. J. & Claridge, T. D. W. Using NMR Solvent Water Relaxation to Investigate Metalloenzyme−Ligand Binding Interactions. *J. Med. Chem.* **53,** 867–875 (2010).

56. Hardy, A. P., Prokes, I., Kelly, L., Campbell, I. D. & Schofield, C. J. Asparaginyl β-Hydroxylation of Proteins Containing Ankyrin Repeat Domains Influences Their Stability and Function. *J. Mol. Biol.* **392,** 994–1006 (2009).
57. Ludwig, C. & Guenther, U. L. Ligand based NMR methods for drug discovery. *Front. Biosci. (Landmark Ed.* **14,** 4565–74 (2009).
58. Ji, Z., Yao, Z. & Liu, M. Saturation transfer difference nuclear magnetic resonance study on the specific binding of ligand to protein. *Anal. Biochem.* **385,** 380–382 (2009).
59. Mayer, M. & Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chemie Int. Ed.* **38,** 1784–1788 (1999).
60. Fielding, L. NMR methods for the determination of protein-ligand dissociation constants. *Curr. Top. Med. Chem.* **3,** 39–53 (2003).
61. FIELDING, L. NMR methods for the determination of protein–ligand dissociation constants. *Prog. Nucl. Magn. Reson. Spectrosc.* **51,** 219–242 (2007).
62. Claudio Dalvit, Paul E. Fagerness, Daneen T. A. Hadden, §, Ronald W. Sarver, ‡ and & Stockman, B. J. Fluorine-NMR Experiments for High-Throughput Screening: Theoretical Aspects, Practical Considerations, and Range of Applicability. (2003). doi:10.1021/JA034646D
63. Zuiderweg, E. R. P. Mapping Protein-Protein Interactions in Solution by NMR Spectroscopy. doi:10.1021/bi011870b
64. Dalvit, C., Flocco, M., Knapp, S., Mostardini, M., Perego, R., Stockman, B. J., Veronesi, M., Varasi, M. High-throughput NMR-based screening with competition binding experiments. *J. Am. Chem. Soc.* **124,** 7702–7709 (2002).
65. Mayer, M. & Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chemie Int. Ed.* **38,** 1784–1788 (1999).
66. Dalvit, C., Fagerness, P. E., Hadden, D. T. A., Sarver, R. W. & Stockman, B. J. Fluorine-NMR Experiments for High-Throughput Screening: Theoretical Aspects, Practical Considerations, and Range of Applicability. *J. Am. Chem. Soc.* **125,** 7696–7703 (2003).
67. Dalvit, C., Gossert, A. D., Coutant, J. & Piotto, M. Rapid acquisition of 1H and 19F NMR experiments for direct and competition ligand-based screening. *Magn. Reson.*

Chem. **49,** 199–202 (2011).
68. Sastrys, M. V. K., Swamyg, J. & Surolia, A. THE JOURNAL OF BIOLOGICAL CHEMISTRY Analysis of Dynamics and Mechanism of Ligand Binding to Artocarpus integrifolia Agglutinin A 13C AND 19F NMR STUDY*. **263,** 14826–14831 (1988).
69. Tengel, T., Fex, T., Emtenas, H., Almqvist, F., Sethson, I., Kihlberg, J. Use of ^{19}F NMR spectroscopy to screen chemical libraries for ligands that bind to proteins. *Org. Biomol. Chem.* **2,** 725–731 (2004).

70. Mecinović, J., Chowdhury, R., Liénard, B. M R., Flashman, E., Buck, M.R G.,Oldham, N. J., Schofield, C. J. ESI-MS Studies on Prolyl Hydroxylase Domain 2 Reveal a New Metal Binding Site. *ChemMedChem* **3,** 569–572 (2008).

71. Cer, R. Z., Mudunuri, U., Stephens, R. & Lebeda, F. J. IC50-to-Ki: a web-based tool for converting IC50 to Ki values for inhibitors of enzyme activity and ligand binding. *Nucleic Acids Res.* **37,** W441–W445 (2009).

72. Nikolovska-Coleska, Z., Wang, R., Fang, X., Pan, H., Tomita, Y., Li, P., Roller, P. P., Krajewski, K., Saito, N.G., Stuckey, J. A., Wang, S. Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization. *Anal. Biochem.* **332,** 261–273 (2004).

73. Yung-Chi, C. & Prusoff, W. H. Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.* **22,** 3099–3108 (1973).

74. Leung, I. K. H., Flashman, E., Yeoh, K. K., Schofield, C. J. & Claridge, T. D. W. Using NMR Solvent Water Relaxation to Investigate Metalloenzyme−Ligand Binding Interactions. *J. Med. Chem.* **53,** 867–875 (2010).

75. Morton, C. J., Pugh, D. J.R., Brown, E. L.J., Kahmann, J. D., Renzoni, D. A.C., Campbell, I. D. Solution structure and peptide binding of the SH3 domain from human Fyn. *Structure* **4,** 705–714 (1996).

76. Aguilar, J. A., Nilsson, M., Bodenhausen, G. & Morris, G. A. Spin echo NMR spectra without J modulation. *Chem. Commun.* **48,** 811–813 (2012).
77. Hajduk, P. J., Olejniczak, E. T. & Fesik, S. W. One-Dimensional Relaxation- and Diffusion-Edited NMR Methods for Screening Compounds That Bind to Macromolecules. *J. Am. Chem. Soc.* **119,** 12257–12261 (1997).
78. Carr, H. Y. & Purcell, E. M. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys. Rev.* **94,** 630–638 (1954).
79. Meiboom, S. & Gill, D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* **29,** 688–691 (1958).

80. Leung, I. K. H., Demetriades, M., Hardy, A. P., Lejeune, C., Smart, T. J., Szöllössi, A., Kawamura, A., Schofield, C.J., Claridge, T. D.W. Reporter Ligand NMR Screening Method for 2-Oxoglutarate Oxygenase Inhibitors. *J. Med. Chem.* **56,** 547–555 (2013).

Chapter 3: Synthesis of substrates for FTO

3.1 Introduction

The classical central dogma of biology comprises the transcription of DNA to give RNA, followed by the translation of RNA into proteins, as described by Francis Crick (Figure 3.1A).[1] DNA and RNA are polymers made up of monomeric units called nucleotides, while the protein oligomer is made up of polymerised amino acids. A nucleotide has three components; a phosphate group, a sugar ring and a nitrogenous base, which is either a purine or a pyrimidine moiety (Figure 3.1B); the nitrogenous base with a sugar ring attached to it is called a nucleoside. DNA contains deoxyribose units whereas RNA contains ribose units (Figure 3.1B). The four nitrogenous bases commonly present in DNA are adenine (A), guanine (G), thymine (T), and cytosine (C); the same nitrogenous bases are present in RNA, except that thymine is replaced by uracil (U) (Figure 3.1B).

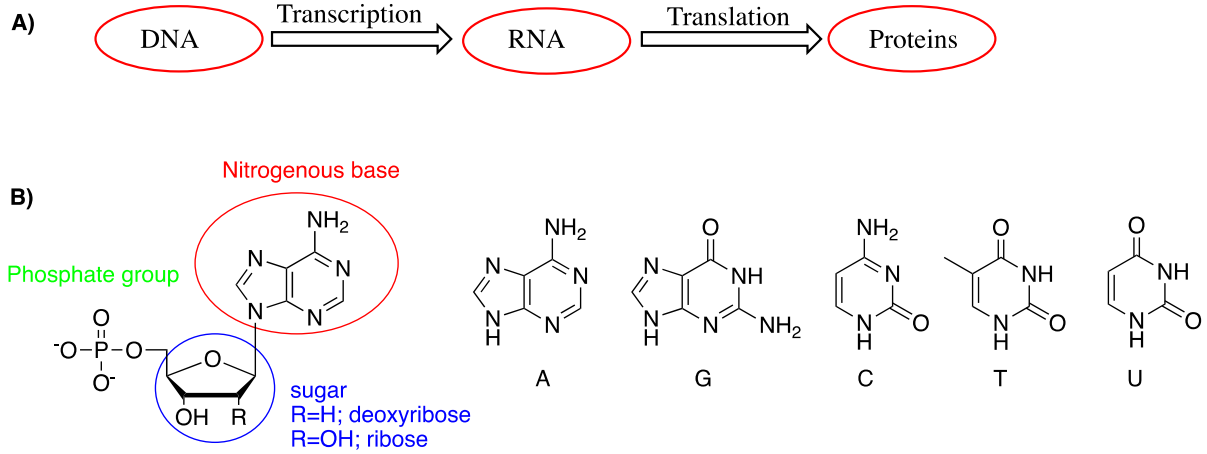


Figure 3.1 A) Schematic representation of the central dogma of biology. B) Schematic representation of nucleotide components (left) and the chemical structures of A, G, C, T, and U (right).

Apart from the canonical nucleotides described above (Figure 3.1B), there are many modified nucleotides present in DNA and RNA.[1] To date, modifications in DNA in eukaryotes are

principally limited to those involving cytidine i.e. 5-methylcytidine and 5-hydroxymethylcytidine.[2–4] By contrast, more than 100 modifications in RNA have been identified, with most in tRNA (~ 80), followed by rRNA (~30), mRNA (12), snRNA (11).[1,5] RNA plays central roles in gene regulation[6–9] and protein synthesis.[10] The malfunctioning of RNA can also cause life threating diseases or death.[11] Consequently, there is a steep growth in research interest in RNA modifications with a piqued interest in the chemical synthesis of modified RNA molecules.

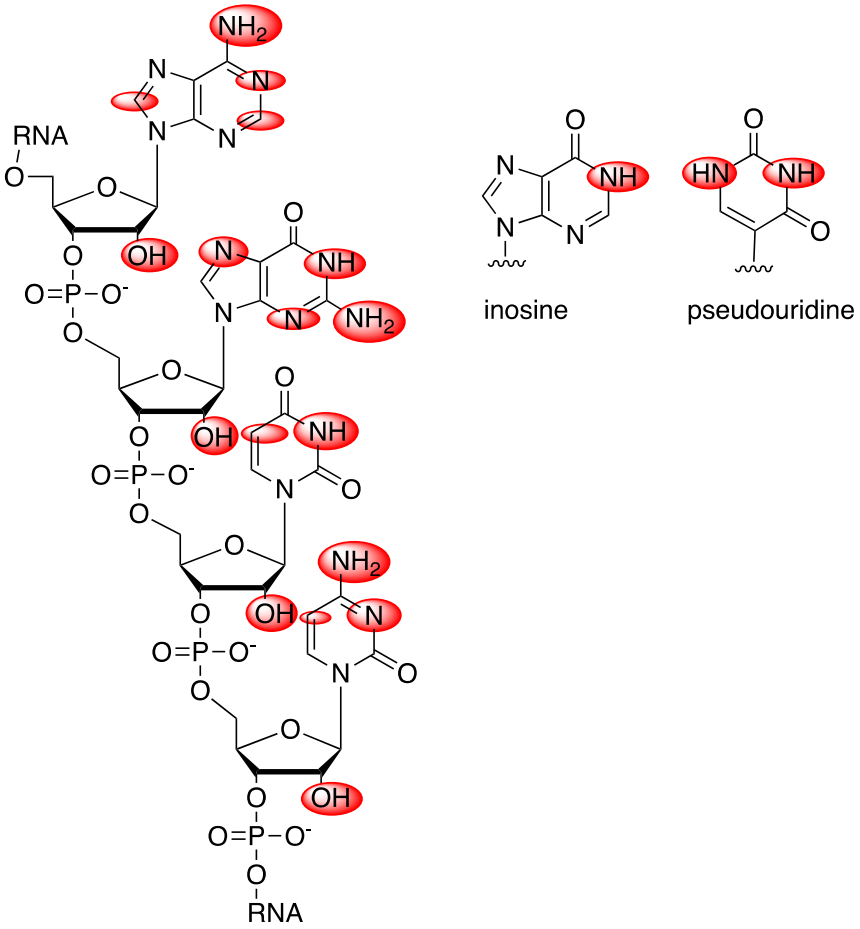


Figure 3.2 Known sites of methylation in RNA.[12]

3.2 Summary of Oligonucleotide synthesis

The first chemical synthesis of oligonucleotides (DNA), reported in 1955, was the thymidine dinucleotide attached via a phosphate group.[13] The chemical synthesis of nucleotides garnered much attention after the synthesis of genes, namely alanine and tyrosine suppressor

tRNAs of yeast and *Escherichia coli*, by Khorana and co-workers.[14] The widely used automated synthesis of DNA oligonucleotides was originally developed by Caruthers and co-workers.[15] The chemical synthesis of RNA is more difficult than DNA synthesis, owing to the presence of extra hydroxyl group at the 2'- position on the sugar ring.

3.2.1 Use of protecting groups in oligonucleotide synthesis

The rationale behind solid phase oligonucleotide synthesis is the step-wise addition of one nucleotide at a time;[15–20] the efficiency of the oligonucleotide synthesis depends on the choice of the protecting groups used for the monomers (nucleosides).[15–20] Protecting groups are used for different functional groups present in the nucleoside, e.g. the exocyclic amine as well as the 5'-,3'-, and 2'-hydroxyl groups.[20,21] The different protecting groups used in oligonucleotide synthesis are described in Section 3.4; Section 3.2.2 which summarises the oligonucleotide procedure.

3.2.2 Summary of Oligonucleotide procedure

Oligonucleotides are generally attached to insoluble particles during their synthesis, known as solid supports; polystyrene and controlled glass pores are generally used as solid supports.[16,17,20,22,23] The oligonucleotide cycle is summarised in Figure 3.3.

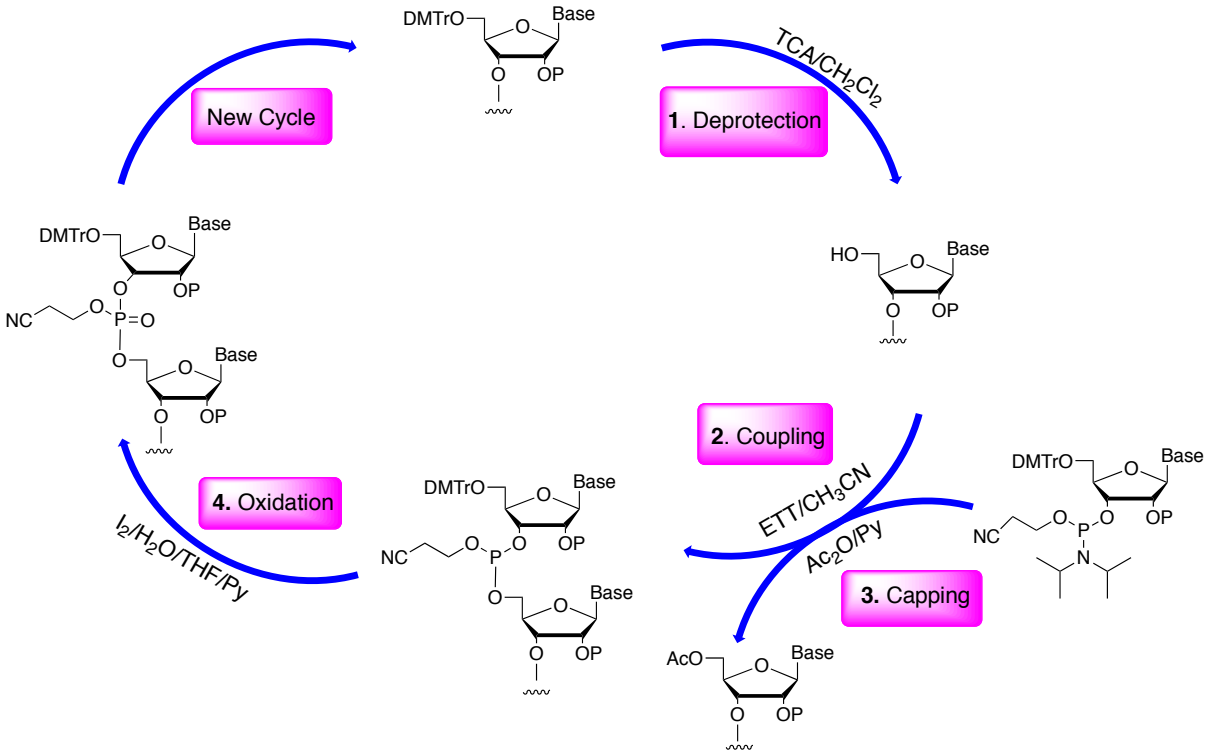


Figure 3.3 Schematic representation of the DNA oligonucleotide synthesis cycle. It should be noted that in the case of RNA, the capping reagent is tert-butylphenoxyacetic anhydride, rather than acetic anhydride.

The cycle generally consists of four steps; the first step being the deprotection of dimethoxytrityl (DMTr) group attached to the 5'-hydroxyl group of the first nucleotide which is generally pre-attached to the resin (solid support) through an amide bond.[16,17,20,22,23] DMTr is deprotected using mild acidic conditions (Figure 3.4).[16,17,20–23]

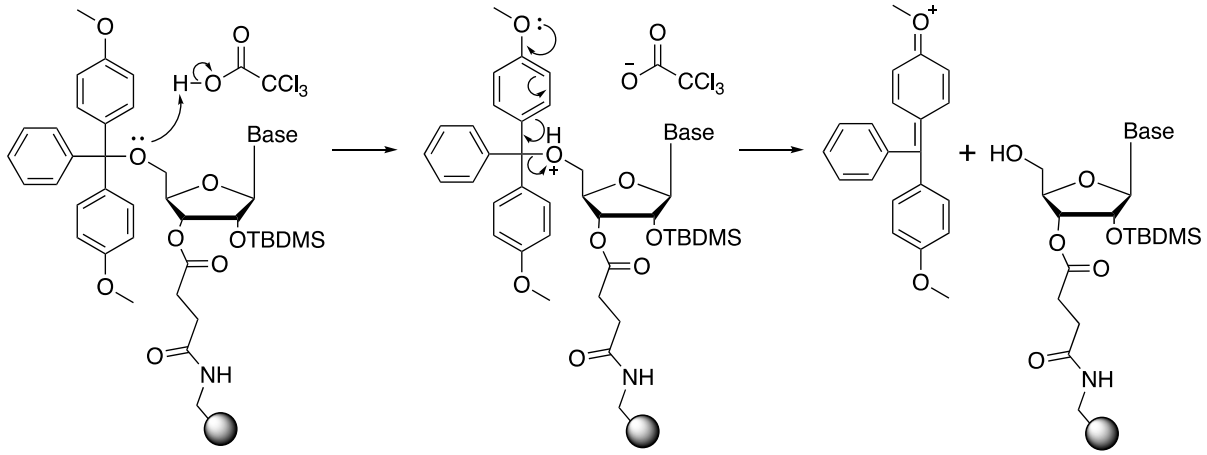


Figure 3.4 Proposed outline mechanism for the deprotection of DMTr group in the oligonucleotide synthesis cycle.[23]

Following the deprotection of the 5'-hydroxyl group, the next step is the coupling of the nucleophilic 5'-hydroxyl group with the freshly added nucleotide in the form of phosphoramidite, along with an activator e.g. tetrazole (Figure 3.5).[16,17,20,22,23] According to the proposed mechanism,[23] tetrazole protonates the diisopropylamine of the phosphoramidite group, making diisopropylamine a good leaving group. It is proposed that the protonated diisopropylamine moiety is then displaced by the negatively charged tetrazole (formed after protonating diisopropylamine), which is subsequently replaced by the 5'-hydroxyl group forming a new phosphorus-oxygen bond (Figure 3.5).[23]

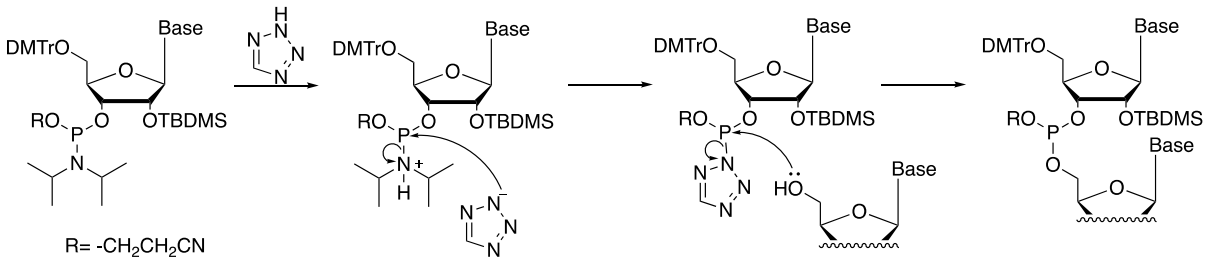


Figure 3.5 Proposed mechanism for coupling during oligonucleotide synthesis.[23]

Following the coupling reaction, the next step is capping of the nucleotides containing unreacted 5'-hydroxyl groups. This step is necessary as it removes any unreacted 5'-hydroxyl groups and also eliminates the possibility of any side reactions. The reagents used for capping are acetic anhydride and *N*-methylimidazole, in case of the DNA oligonucleotides,[20,23] and *tert*-butylphenoxyacetic anhydride in the case of RNA oligoribonucleotides.[22,24]

After the removal of the unreacted starting material, the next step is the oxidation of the trivalent phosphorous [P(III)] to the pentavalent state [P(V)]. The oxidation uses a mixture of iodine, water and pyridine leading to the formation of the phosphate oligonucleotide backbone protected by the cyanoethyl groups (Figure 3.6).[20,23,24]

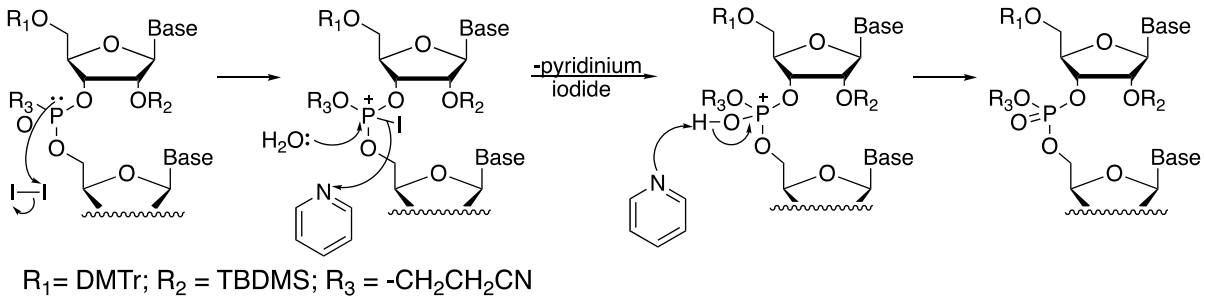


Figure 3.6 Proposed mechanism of the oxidation step in oligonucleotide synthesis.[23]

After the oxidation step, the oligonucleotide enters into a new cycle starting with the removal of the DMTr groups from the 5'-hydroxyl group of the nucleotide; the process is then repeated till the required length of the oligonucleotide is obtained. Subsequently, the oligonucleotide is cleaved and released from the solid support by ester hydrolysis under basic condition, usually using aqueous ammonia (Figure 3.7).[20,22,23]

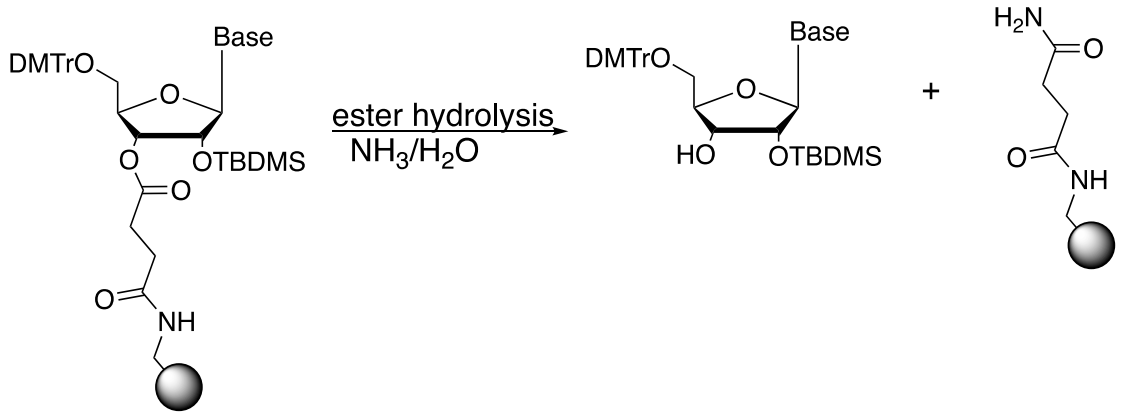


Figure 3.7 Schematic representation of the hydrolysis of the ester bond leading to the release of the oligonucleotide from the solid support.[23]

Following the release of the oligonucleotide from the solid support, the oligonucleotides are heated in concentrated ammonia to deprotect the exocyclic amines and the phosphate group (cyanoethyl) (Figure 3.8).[23]

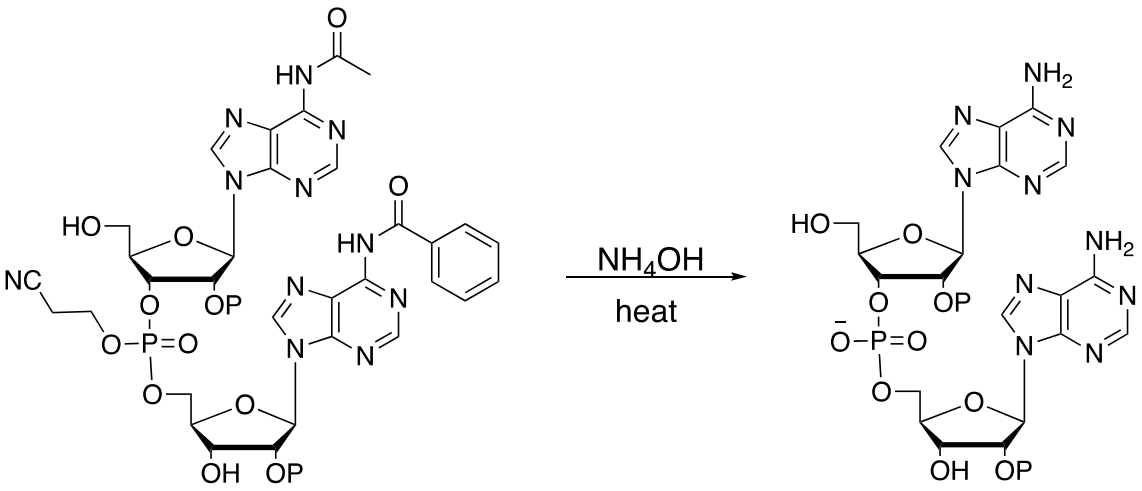


Figure 3.8 Schematic representation of the basic hydrolysis of the amide bond and the removal of cyanoethyl group.[23]

The final step is the deprotection of the fluoride labile TBDMS group which is usually done using *tert*-butylammonium fluoride (TBAF) solution in tetrahydrofuran (THF).[16,20–22,24]

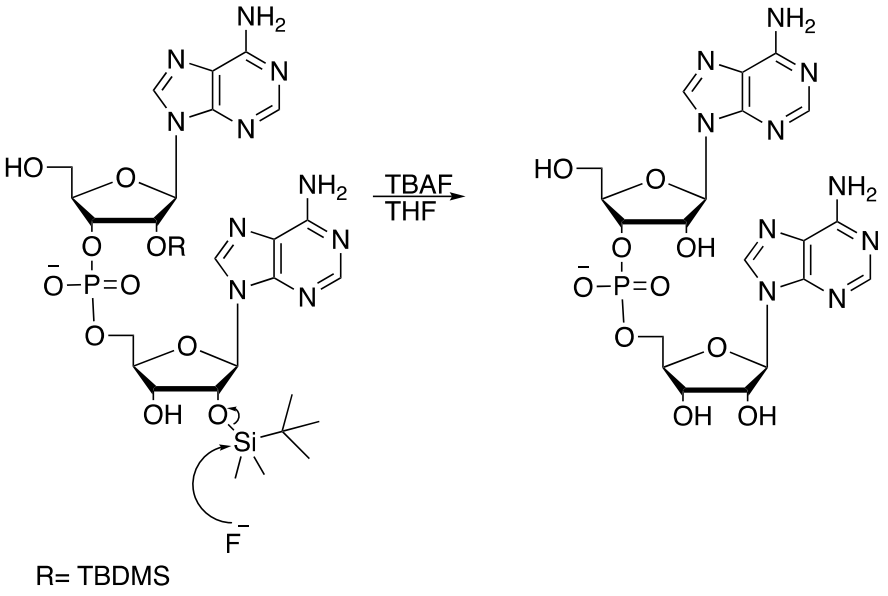


Figure 3.9 Schematic representation of the deprotection of TBDMS group using TBAF.[24]

3.3 Objectives of the work described in this chapter

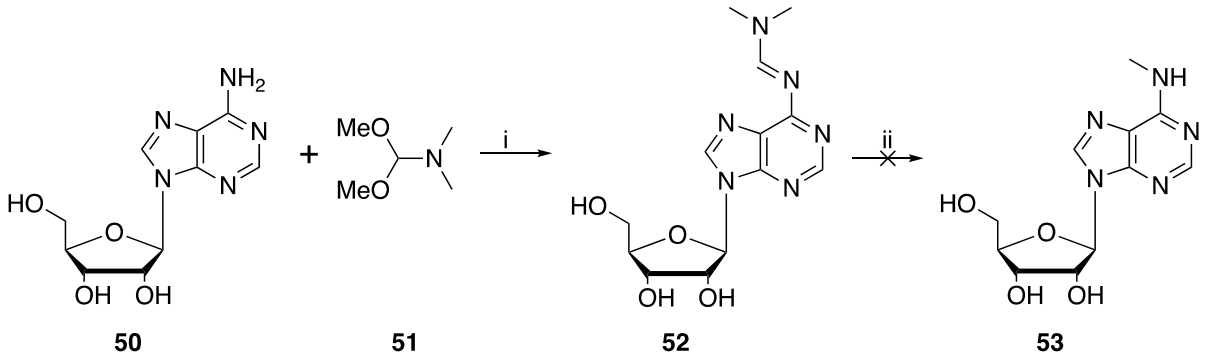
The specific objectives of the work described in this chapter were:

- To synthesise the protected monomer of N^6-methyladenosine (hereafter, m^6A phosphoramidite) for use in the oligonucleotide synthesis;
- To optimise the previously reported synthetic route for the synthesis of m^6A phosphoramidite (or to develop a new route);
- To synthesise the protected monomer of 5-iodouridine (hereafter, iodouridine phosphoramidite) for use in the oligonucleotide synthesis;
- To synthesise analogs of N^6-methyladenosine (m^6A) and investigate them as FTO substrates;
- To synthesise 5-methylcytidine (5mC), 5-hydroxymethylcytidine (5hmC) and 5-formylcytidine (5fC) to study their reaction with formaldehyde.

3.4 Results and Discussion

3.4.1 Optimisation of the synthetic route for m^6A phosphoramidite

Initially, work focussed on the synthesis of m^6A from adenosine. It was envisaged that adenosine could be monoalkylated via an amidine intermediate (**52**; Figure 3.10) as reported previously.[25] Thus, adenosine (**50**; Figure 3.10) was reacted with *N*,*N*-dimethylformamide dimethyl acetal (**51**; Figure 3.10) to give 6-*N*-(1-dimethylamino)methylideneadenosine (**52**; Figure 3.10), and confirmed by mass spectrometric analyses of the reaction mixture. Sodium borohydride ($NaBH_4$) was then added to the crude reaction mixture at room temperature;[25] however, no conversion to m^6A (**53**; Figure 3.10) was observed as suggested by TLC and mass spectrometric analyses. The reaction was then heated at 40°C for 24hr; however, no evidence for the conversion of **52** to m^6A (**53**) was accrued.



50 **51** **52** **53**

*Figure 3.10 Reaction Conditions: i) **51**(5.5 eq), DMF, 40°C, 1hr; ii) $NaBH_4$ (5 eq), DMF, rt, 24hr.*

It was envisaged that the presence of other components in the reaction mixture, apart from amidine (**52**), might be preventing the reduction of amidine (**52**) to m^6A (**53**). Hence, the resultant amidine (**52**) was purified and characterised; attempts to reduce the amidine (**52**) by $NaBH_4$ according to the reported procedure were unsuccessful.[25] Use of some other reducing agents, such as $Na(CN)BH_3$, was also investigated; however, the reduction of amidine (**52**) to m^6A (**53**) was not observed even at elevated temperature (40°C).

Next, search for an alternative method was conducted to access m^6A (**53**). It was envisaged that the direct alkylation of adenosine, via methyl iodide (CH_3I), could result in a mixture of alkylated products (Figure 3.11). Hence, the protection of the three hydroxyl groups present on the sugar ring was deemed necessary.

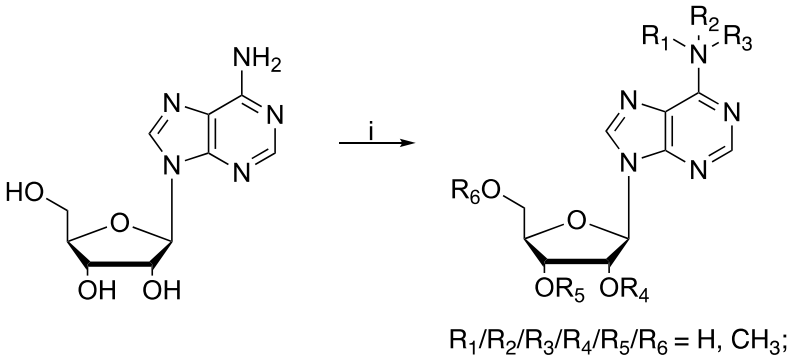


Figure 3.11 Schematic representation of the reaction of adenosine with CH_3I and the resultant combination of possible products.

A wide choice of hydroxyl protecting groups is reported in literature.[26] Di-*tert*-butylsilyl bis(trifluoromethanesulfonate) (hereafter, DTBS ditriflate; **54,** Figure 3.12A) was chosen in the present study for the selective protection of the 3'-and the 5'-hydroxyl groups, intending that the selective deprotection of the DTBS group could be achieved in the presence of appropriate protecting groups on adenosine.[21,26] Some important measures were necessary to choose an appropriate group for the protection of the 2'-hydroxyl group. This group must be stable under mild acidic condition and should not interfere with the removal of the dimethoxytrityl (DMTr) group from the 5'-hydroxyl group in oligonucleotide synthesis, and should be stable under the basic hydrolysis condition used for cleaving the oligonucleotide from the solid support after the completion of oligonucleotide synthesis (Figure 3.3 and Figure 3.7).

Furthermore, given that RNA degrades in the presence of a base (Figure 3.12B); the deprotection of the 2'-hydroxyl group was required to be done under mild conditions. The 2'-hydroxyl protecting group should also not interfere with the coupling reaction between the 3'-phosphoramidite and the 5'-hydroxyl group (Figure 3.3); thus, the use of very sterically hindered protecting groups is preferably avoided considering the close proximity of the 2'-hydroxyl group to the coupling site in the oligonucleotide synthesis. Many protecting groups have been reported for the 2'-hydroxyl group,[21] with examples given in Figure 3.12C. The most widely used reagent for the protection is *tert*-butyldimethylsilyl chloride (TBDMSCl) which was also used in the present study.[21]

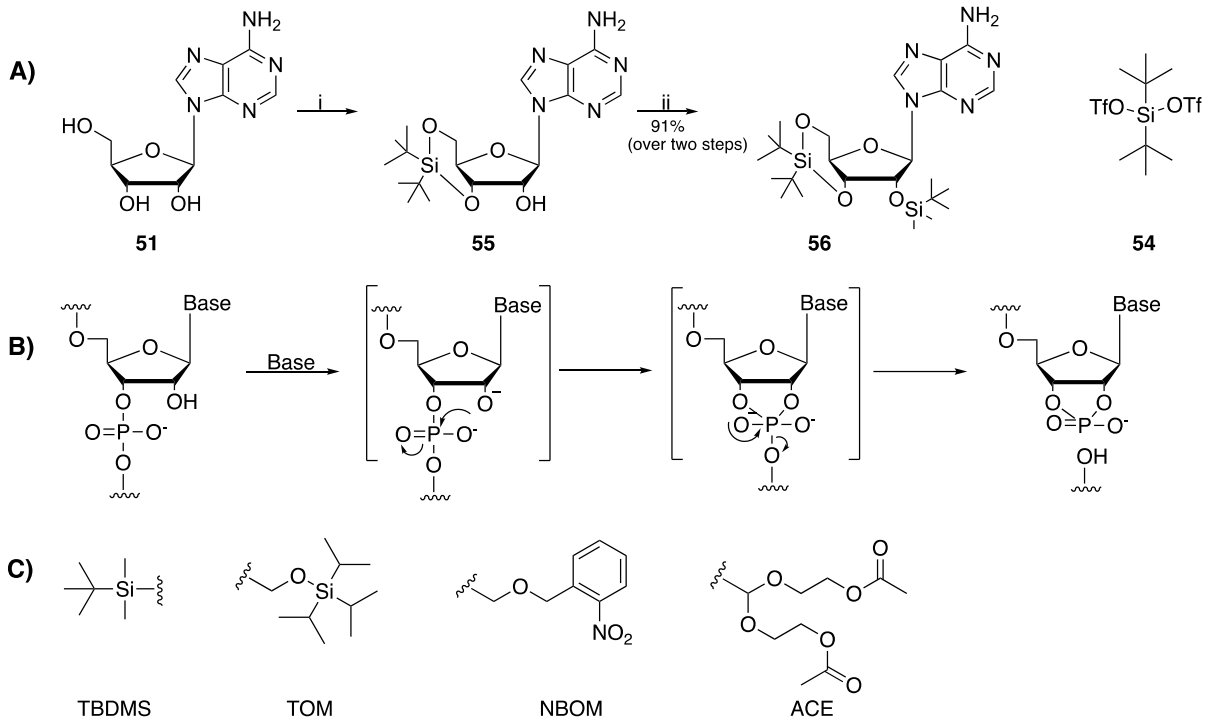


*Figure 3.12 A) Reaction conditions: i) **54**(1.1eq), imidazole (2eq), DMF, 0°C, 1hr; ii) TBDMCl (1.2eq); imidazole (3eq), 60°C, 12hr; B) Outline of the proposed mechanism of RNA degradation under basic conditions; C) Examples of the protecting groups used for the protection of the 2'-hydroxyl group in oligonucleotide synthesis.*

Following the protection of the hydroxyl groups present on the sugar ring of adenosine (**50**) to give 3'5'-*O*-(di-*tert*-butyl)silyl-2'-*O*-dimethyl(*tert*-butyl)silyladenosine (**56**; Figure 3.12A), the next step was the monoalkylation of the exocyclic amine of **56**. It was observed that the direct alkylation of the exocyclic amine using CH_3I resulted in the formation of a inseparable mixture of products likely to be mono-, di-, tri-alkylated amines (Figure 3.13).[27]

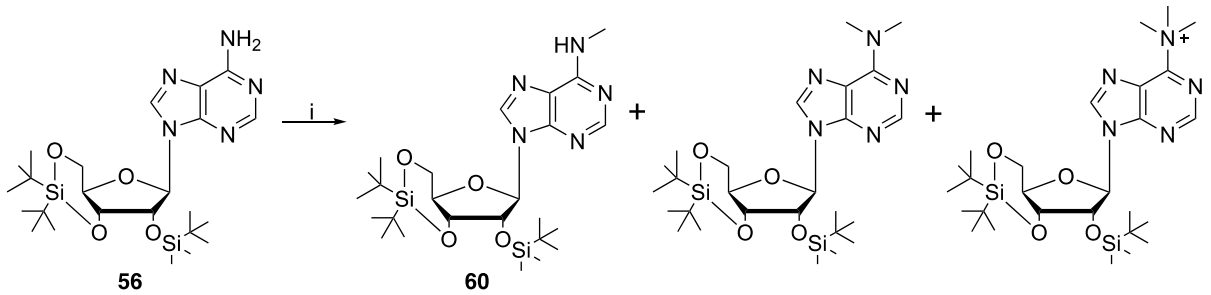


Figure 3.13 Reaction conditions: CH_3I (1.1 eq), sodium hydride (NaH, 1.2 eq), DMF, 0°C, 1hr and possible products.

Hence, an alternative protocol for the protection of the exocyclic amine (**56**) was sought that could avoid any potential dialkylation/trialkylation. Many protecting groups for the exocyclic amine are reported;[21] the most widely used are the acyl-groups, e.g. acetyl (Ac), benzoyl (Bz) or isobutyryl groups (Figure 3.14A).[21] However, the deprotection procedures for the acyl protecting groups require heating for prolonged periods under basic conditions which could lead to the degradation of RNA (Figure 3.12B). Thus, phenoxyacetyl (PAC) or dimethylformamidine groups have been developed as their cleavage requires shorter deprotection times (Figure 3.14A).[21] The benzoyl group was chosen to protect the exocyclic amine (**56**) to give 3'5'-*O*-(di-*tert*-butyl)silyl-2'-*O*-dimethyl(*tert*-butyl)silyl-*N*6-benzoyladenosine (**57**; Figure 3.14B), according to the reported procedure.[28] The addition of benzoyl chloride was done at 0°C, to avoid dibenzoylation.

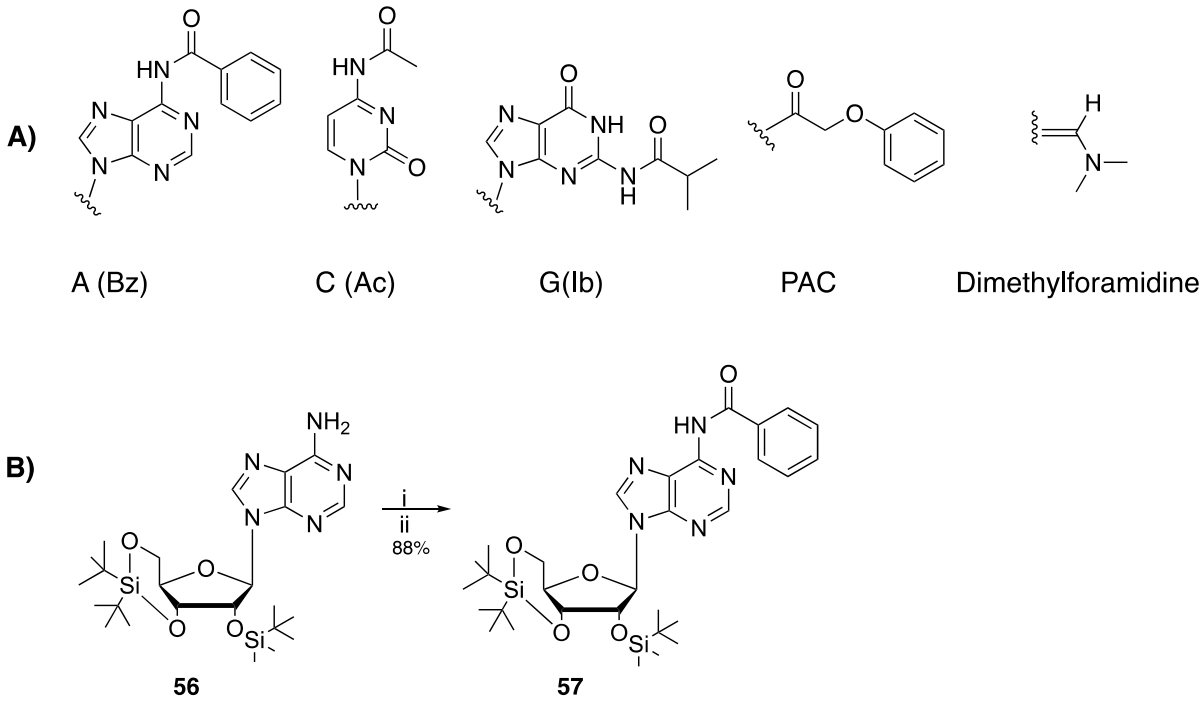


Figure 3.14 A) Protecting groups used for the exocyclic amine in oligonucleotide synthesis. B) Reaction conditions: i) Benzoyl chloride (1.5eq), pyridine, 0-23°C, 12hr; ii) NH_3, MeOH, rt, 2hr.

Following the protection of the exocyclic amine of **56** to give **57**, the next step was monoalkylation of **57** using CH_3I, which was achieved using the reported procedure.[29] The

procedure involves the alkylation of the *N*6 acylated adenosine (**57**) using methyl iodide in the presence of a phase transfer agent, in this case, tetrabutylammoniumbromide (TBAB); thus, **57** was reacted with methyl iodide in the presence of a strong base (in this case, NaOH) and TBAB to give 3'5'-*O*-(di-*tert*-butyl)silyl-2'-*O*-dimethyl(*tert*-butyl)silyl-*N*6,*N*6-benzoylmethyladenosine (**58;** Figure 3.15) in reasonable yield. The use of *tert*-butylammonium iodide (TBAI) as a phase transfer reagent increased the yield of the reaction to 65% compared to 54% with *tert*-butylammoniumbromide. It was observed that the sequential addition of reagents with the nucleoside (**57**), phase transfer reagent and the base first, followed by the addition of CH_3I gave better yields compared to the concurrent addition of all reagents (Figure 3.15).

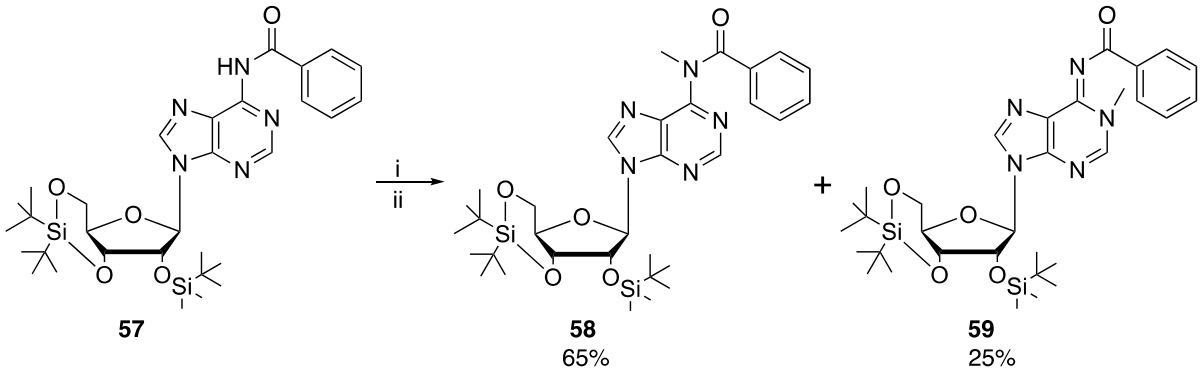


Figure 3.15 Reaction conditions: i) NaOH (10 eq), TBAI (1 eq), dichloromethane (DCM), rt, 30 min; ii) CH_3I (4 eq), rt, 24hr.

The reaction resulted in the formation of two major products as observed by TLC. The presence of the phase transfer agent made the work-up procedure tedious; it was later found that the phase transfer reagent could be easily precipitated from the solution by dissolving the concentrated crude mixture in ethyl acetate. The newly formed two products were observed to have the same mass in the mass spectrometric analyses. Analysis of the ^{1}H NMR spectra of the products indicated the presence of a methyl group in both, resonating at δ_H 3.61 ppm and δ_H 3.64 ppm. It was proposed that one of the products is likely to result from the methylation at the *N*-1 position of adenosine to give 3'5'-*O*-(di-*tert*-butyl)silyl-2'-*O*-dimethyl(*tert*-

butyl)silyl-*N*6,*N*1-benzoylmethyladenosine(**59**; Figure 3.15), in addition to the desired methylation at the *N*-6 position to give **58**. To test this hypothesis, 2D NMR (HSQC, HMBC, and COSY) spectra were recorded for both the products. The methyl group resonating at δ_H 3.64 ppm with higher R_f value (0.5; ethyl acetate:cyclohexane:: 2:3) by TLC, showed correlations to its C-6 of adenosine and the carbonyl of its benzoyl group (2D HMBC, Figure 3.16), thus, confirming the structure to be 5'3'-*O*-Bis(*tert*-butyl)silyl-2'-*O*-dimethyl(*tert*-butyl)silyl-*N*6,*N*6-benzoylmethyladenosine (**58**). The other product with the methyl group resonating at δ_H 3.61 ppm, as predicted, was found to be 5',3'-*O*-Bis(*tert*-butyl)silyl-2'-*O*-dimethyl(*tert*-butyl)silyl-*N*6,*N*1-benzoylmethyladenosine (**59**) with the methyl group showing correlation to the C-2 of its adenosine ring.

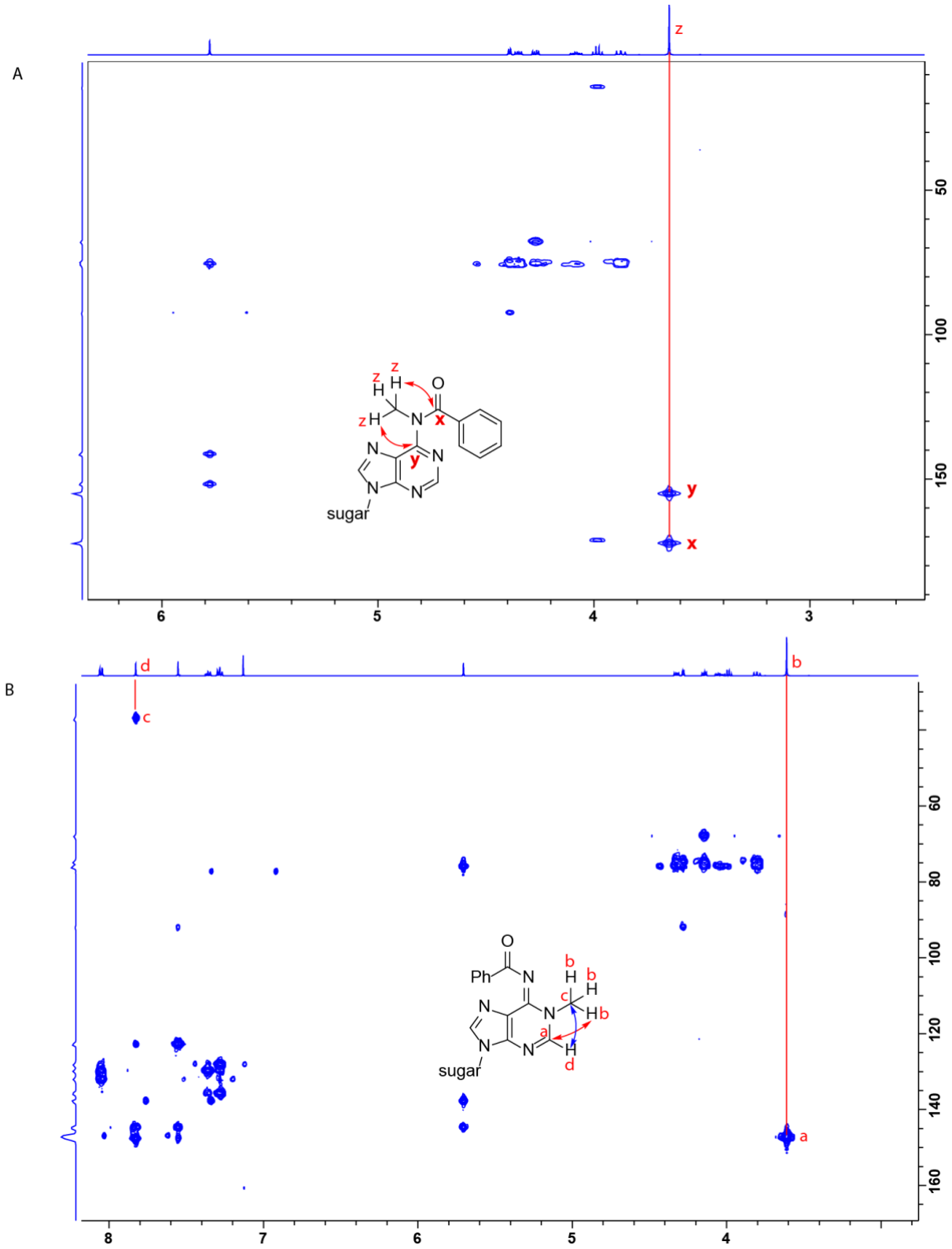


Figure 3.16 HMBC spectra of A) m^6A, showing important correlations of methyl protons (indicated by "z" in red) to the carbons, C-6 (indicated by "y" in red) of aromatic ring and carbonyl group(indicated by "x" in red); B) m^1A showing correlation between methyl protons (indicated by "b" in red) and carbon C-2(indicated by "a" in red) of aromatic ring, and proton (indicated by "d" in red) attached to the carbon C-2 of aromatic ring to the carbon of methyl group (indicated by "c" in red).

Following the synthesis of **58**, the next step was the deprotection of the benzoyl group. As mentioned earlier, the deprotection of the benzoyl group is usually carried out under harsh basic conditions[21,26] and is often followed by time-consuming purification process using Dowex column or adding amberlite resins.[21,30,31] Therefore, it was desirable to cleave the benzoyl group under milder conditions. Use of lithium triethylborohydride ($LiEt_3BH$) in THF for the deprotection of the benzoyl group as reported,[32] gave 3'5'-*O*-(di-*tert*-butyl)silyl-2'-*O*-dimethyl(*tert*-butyl)silyl-*N*6-methyladenosine (**60**; Figure 3.17) in 75% yield.

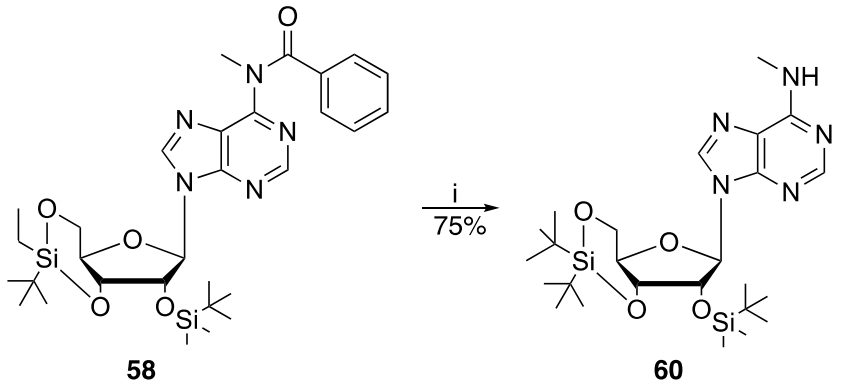


Figure 3.17 Reaction Conditions: i) $LiEt_3BH$ (5 eq), THF, 0°C, 1hr.

Following the deprotection of the benzoyl group to give **60**, the next step was the selective deprotection of the DTBS group without cleaving the TBDMS group present on the 2'-hydroxyl group. As reported in the literature, dilute HF.pyridine mixture was found to selectively cleave the DTBS group with the TBDMS group remaining intact.[28] Accordingly, **60** was subjected to diluted HF.pyridine mixture, at 0°C, to give 2'-*O*-dimethyl(*tert*-butyl)silyl-*N*6-methyladenosine (**61**; Figure 3.18) in 80 % yield.

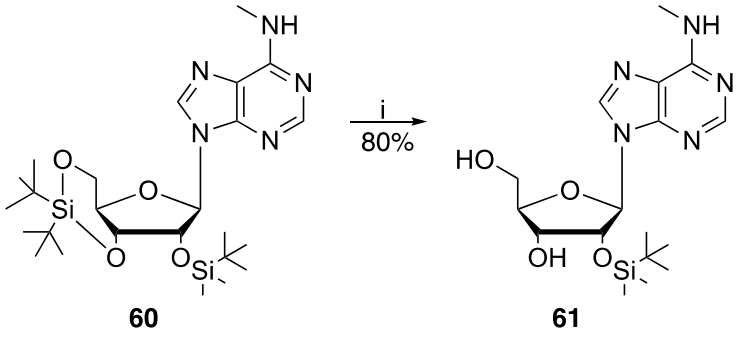


Figure 3.18 Reaction Conditions: HF.pyridine (3 eq), pyridine, 0°C, 5hr.

Following the selective deprotection of the DSBT group of **60** to give **61**, the next step was selective protection of the 5'-hydroxyl group. It was noted that the protecting group on the 5'-hydroxyl group has to be cleaved first to couple the 5'-hydroxyl group with the phosphoramidite; whereas, the 2'-hydroxyl group has to remain intact during oligonucleotide synthesis.[20] Thus, it was envisaged that the deprotection procedure should be compatible with the other functional groups present on the protected nucleoside and should be high yielding and time efficient. Protecting reagents available for the 5'-hydroxyl group include trityl chloride (TrCl), monomethyl trityl chloride (MMTrCl), dimethoxytrityl chloride (DMTrCl), and fluoride labile protecting groups (Figure 3.19A).[21,26] The main advantage of the DMTr and related protecting groups over fluoride labile groups is the release of a colored cation on the cleavage, which could be used for reaction monitoring (Figure 3.19B); the DMTr group can be selectively cleaved using mildly acidic conditions (Figure 3.19C). The 5'-hydroxyl group of **61** was, thus, protected with the DMTr group to give 5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-dimethyl(*tert*-butyl)silyl-*N*6-methyladenosine(**62**) as reported in the literature.[21,28] It was observed that portionwise addition of DMTrCl gave a better yield of the desired product compared to one-time addition of reagent.

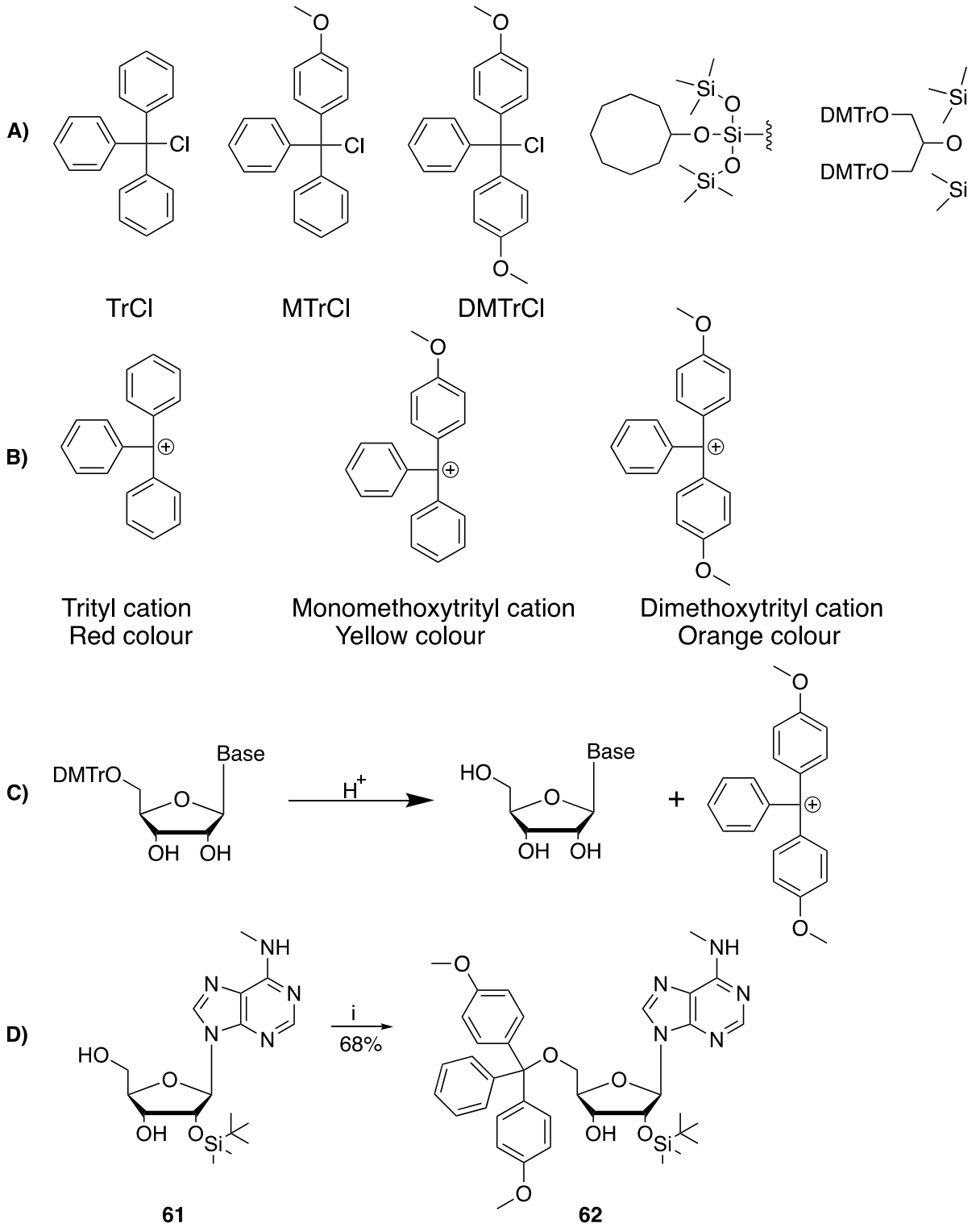


Figure3.19 A) Chemical structures of the protecting groups used for the protection of the 5' hydroxyl group during oligonucleotide synthesis; B) Chemical structure of the trityl cation (red), the monomethyltrityl cation (yellow) and the dimethoxymethyl cation (orange); C) Schematic representation of the deprotection of the DMTr group under mildly acidic conditions. D) Reaction conditions: i) DMTrCl (1.2 eq), pyridine, 0°C, 12hr.

After the successful synthesis of **62**, the next and the final step was the protection of its 3'-hydroxyl group by the phosphoramidite group (**64**; Figure 3.20) to give 5'-*O*-(4,4'-dimethoxytrityl)-(3'-*O*-[(2-cyanoethyl)(*N,N*-diisopropylamino)phosphino]-2'-*O*-dimethyl (*tert*-butyl)silyl-*N*6-methyladenosine (**63**; Figure 3.20), which was carried out as reported.[28] A new spot was observed to develop on TLC with roughly the same R_f as that of the reactant (**62**). The reaction was then analysed by LCMS, which revealed formation of one new peak with the same mass as reactant. No other new spot, except the one with the R_f similar to the reactant, on TLC was observed even after addition of twice the amount of the phosphitylating reagent (**64**) required. The phosphitylating reagent (**64**) was obtained from Sigma-Aldrich, and was found to be impure though LCMS and TLC analysis; the impure reagent probably resulted in the formation of the byproduct.

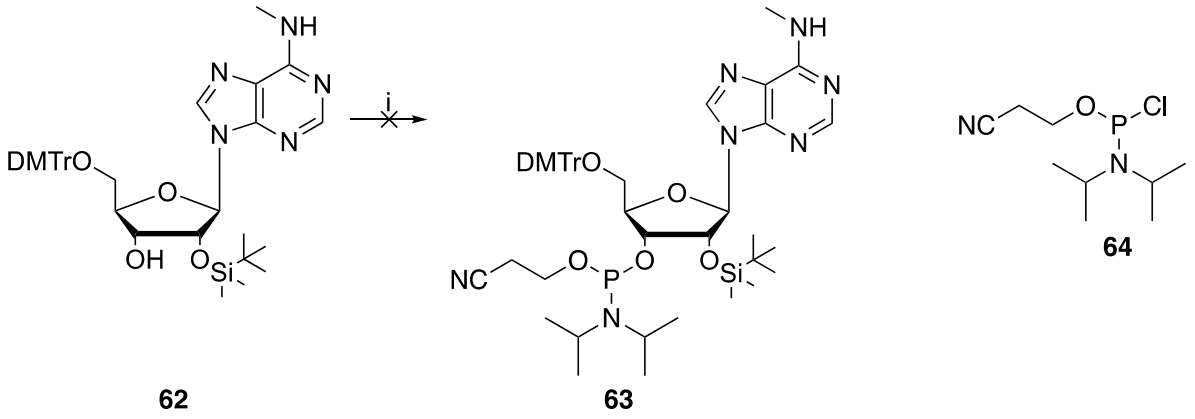
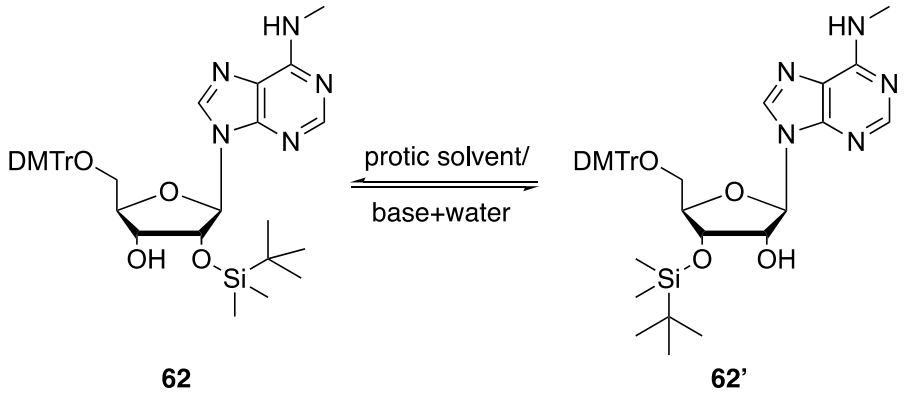


*Figure 3.20 Reaction conditions: **64** (2 eq), diisopropylethylamine (DIPEA, 5eq), THF, 0°C-rt, 6 hr.*

The synthesis was then repeated. Purification of the DMTr protected nucleoside (**62**) using methanol: dichloromethane mixture by column chromatography followed by the subsequent evaporation of the excess solvent under reduced pressure led to the development of another spot on TLC with similar R_f value to that of the desired product (**62**). Attention was then focussed on the precedent literature for the observed behavior of **62**. The TBDMS group

attached to vicinal hydroxyl groups has been reported to migrate to the 3'-hydroxyl from the 2'-hydroxyl group and vice-versa, in protic solvents as reported in literature (Figure 3.21).[33,34] Thus, in methanol at 40°C, which is usually the temperature of the water bath while evaporating the solvent, TBDMS group of **62** migrated to give **62'**. Thus, the other peak observed on TLC corresponded to the isomerised product, **62'** (Figure 3.21).



62 **62'**

Figure 3.21 Schematic representation of the migration of the TBDMS group from the 2'-hydroxyl to the 3'-hydroxyl group in the presence of wet base/ protic solvent.

The consecutive failures in the synthesis of **63** led to an extensive literature search for an alternative route. It was found that inosine (**65**) could be converted to m^6A via conversion of its C-6 carbonyl group to a halogen to give 6-haloinosine that upon substitution with methylamine could afford m^6A (**53**) via nucleophilic aromatic substitution (S_NAr) reaction. The approach seemed promising as it would exclude the formation of undesirable m^1A and the tedious work-up procedure, thus increasing the efficiency of the synthetic route.

Inosine (**65**; Figure 3.22) was used as the starting material; its hydroxyl groups were protected with the acetyl group using acetyl anhydride which resulted in 2'3'5'-*O*-triacetylinosine (**66**; Figure 3.22) in 89% yield. It should be noted that the carbonyl group of **66** is likely to get replaced by a chloride using phosphoryl chloride ($POCl_3$) which would result in the formation of 2'3'5'-*O*-triacetyl-6-chloroinosine (**67**; Figure 3.22), however, the tedious work-up of the reaction and the handling of the toxic chemical, $POCl_3$, precluded its

use, at this stage. *N*-chloromethylene-*N*, *N*-dimethylammoniumchloride (*Vilsmeier* reagent)[35] was then investigated as a chlorinating reagent; the reaction resulted in the formation of **67** in 96% yield.

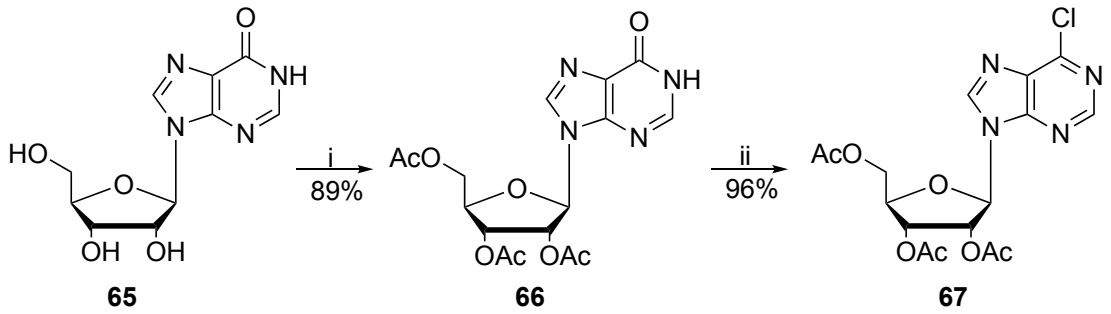


Figure 3.22 Reaction conditions: i) acetyl anhydride (5.7 eq), pyridine, DMF, 75°C, 0.5hr; ii) Vilsmeier reagent (2 eq), DCM, 40°C, 24hr.

Compound **67** was then subjected to S_NAr reaction in the presence of methyl amine ($MeNH_2$) solution in ethanol (EtOH);[35] the reaction was left to stir overnight, leading to subsequent base mediated hydrolysis of the ester protecting groups in the presence of methylamine. Thus, the synthesis of m^6A (**53**) was completed successfully.

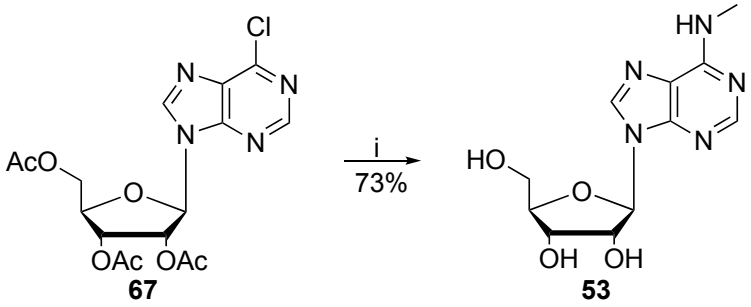
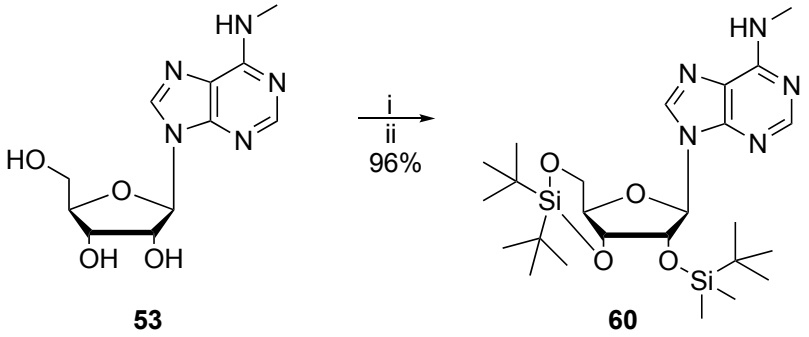


Figure 3.23 Reaction conditions: $MeNH_2$ (33% in ethanol (EtOH), 7eq), EtOH, rt, 18hr.

The next step was appropriate protection of the functional groups present in m^6A (**53**). Earlier attempts to purify the mixture of the 2'- and the 3'-TBDMS isomers of m^6A were unsuccessful (Figure 3.21); it was envisaged that protection of the 3'-and the 5'-hydroxyl groups with the DTBS group would be the best approach to avoid the purification challenges. Thus, similar reaction conditions to those used for the protection of the hydroxyl groups of

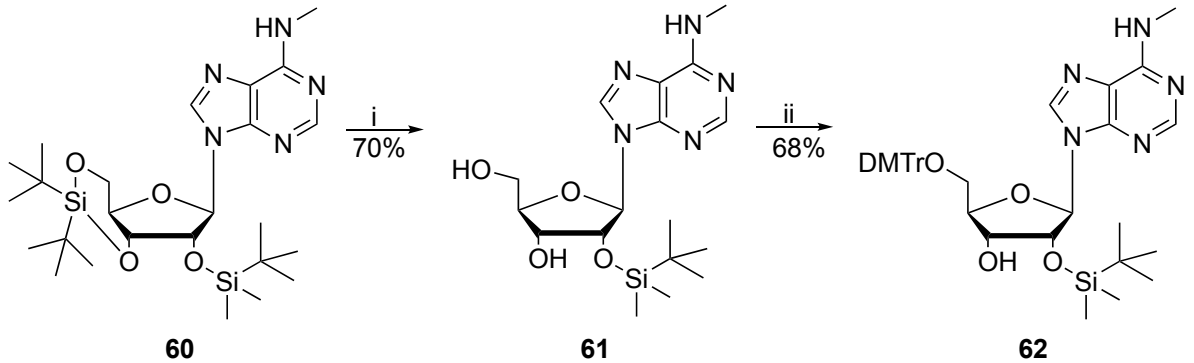
adenosine (Figure 3.12) were used to protect m^6A (**53**). Thus, m^6A (**53**) was protected using the DTBS group for the 3' -and the 5'-hydroxyl groups and TBDMSCl for the 2'- hydroxyl group (Figure 3.24) resulting in the formation of **60**.



53 **60**

*Figure 3.24 Reaction conditions: i) **54** (1.1eq), imidazole (2eq), DMF, 0°C, 1hr; ii) TBDMSCl (1.2eq), imidazole (3eq), 60°C, 12hr.*

The DTBS group of **60** was then selectively cleaved using a dilute solution of HF.pyridine, resulting in **61** (70%). The 5'-hydroxyl group of **61** was protected with the DMTr group to give **62**; purification of the crude mixture by column chromatography was done carefully to avoid isomerisation.



60 **61** **62**

Figure 3.25 Reaction conditions: i) HF.pyridine (3 eq), pyridine, 0°C, 5hr; ii) DMTrCl (1.2 eq), pyridine, 0°C, 12hr.

Following the successful synthesis of the 2'TBDMS isomer (**62**), the next step was the protection of the 3'-hydroxyl group with phosphoramidite (**64**). Initially, the reaction

conditions were optimised; phosphitylation reaction was carried out in the presence of two different bases i.e pyridine and DIPEA. Extra care was taken and the reactions were done under argon with a new batch of phosphitylating reagent (**64**) obtained from Link Technologies. It was observed that the reaction with DIPEA as a base gave better results compared to the reaction with pyridine. Therefore, DIPEA was chosen as the base for the reaction. It was observed that use of 10 equivalents of base and 2.5 equivalents of the phosphitylating reagent (**64**) gave the best yields. After 6 hr of the addition of the phosphitylating reagent, **62** was observed to be consumed and four new spots were observed to develop on TLC. The two upper spots, with similar R_f values, may be the *N*-phosphitilated (**68**; Figure 3.26) by-products; which were not characterised. The two lower spots (R_f of 0.4; ethyl acetate: toluene: pyridine::1:1:0.01) were proposed to be the stereoisomers of **63** because of the presence of a P(III) center in the m^6A phosphoramidite (**63**). The purification of the reaction mixture was done under argon and care was taken to minimise the exposure of column fractions to air and moisture.

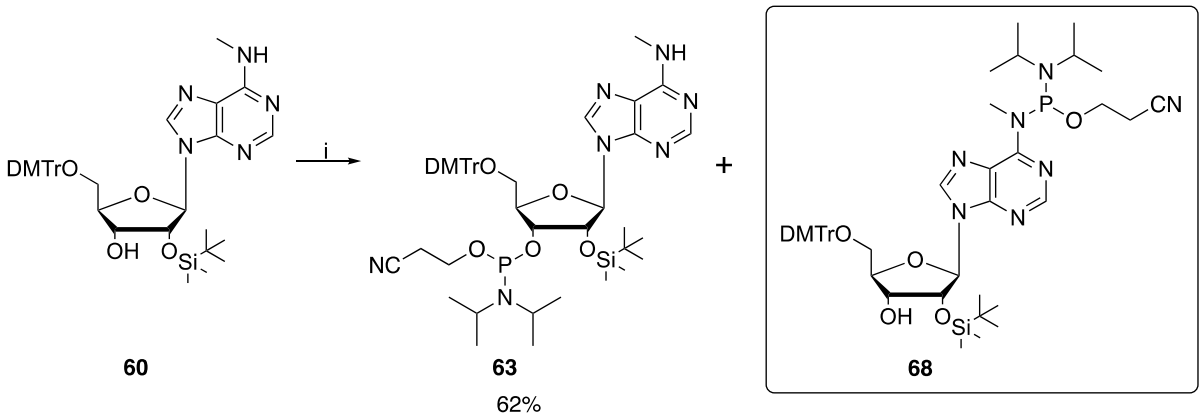


Figure 3.26 Reaction conditions: ***64*** *(2.5eq), DIPEA (10eq),* CH_2Cl_2*, 0°C, 6hr.The possible by-product is highlighted in box.*

Hence, after considerable efforts, the synthesis of the m^6A phosphoramidite (**63**) was completed successfully. Work was then focussed on the synthesis of the 5-iodouridine phosphoramidite (**73**; Figure 3.30).

3.4.2 Synthesis of protected monomer of 5-iodouridine

The rationale behind the synthesis of the iodouridine phosphoramidite (**73**; Figure 3.30) was to synthesise an oligomer containing both 5-iodouridine (**69**; Figure 3.28) and m^{6}A (**53**). It was envisaged that 5-iodouridine (**69**) could be extended via a linker and photo cross-linked to FTO as reported;[36] thus, generating an approach to crystallise FTO cross-linked to its substrate via a diazirine-carbene reaction.

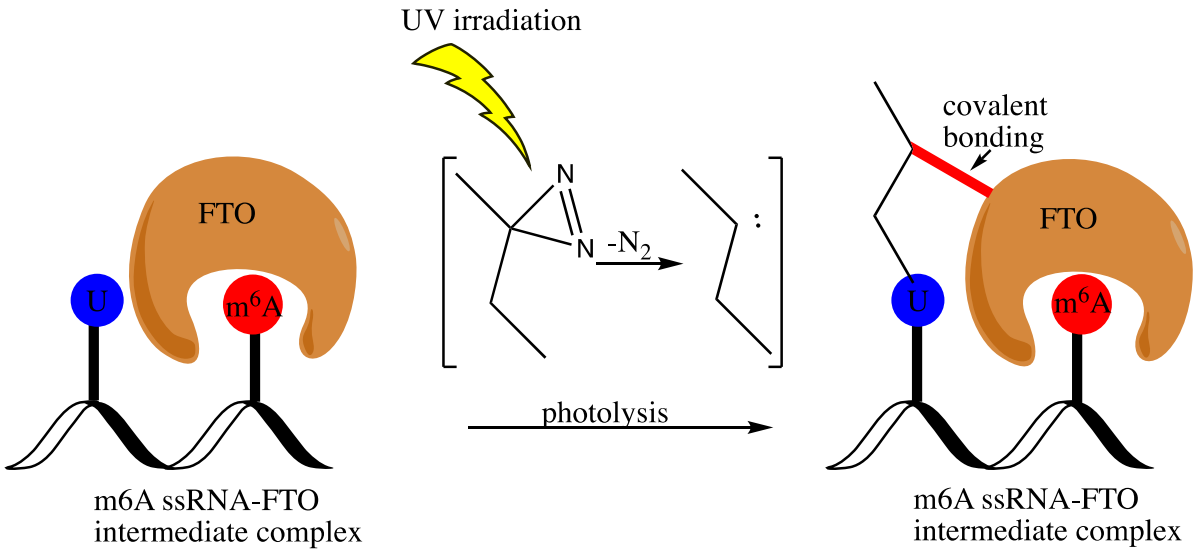
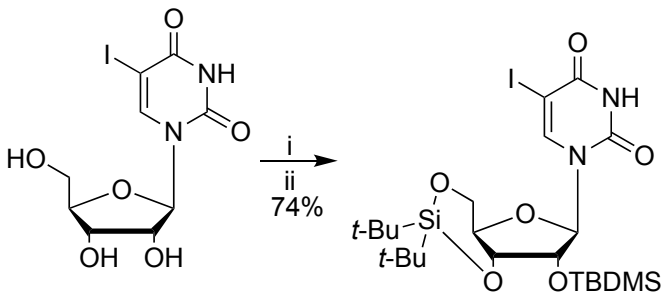


Figure 3.27 Schematic representation of photocrosslinking of single stranded-RNA (ss-RNA) to FTO via diazirine activated by UV irradiation.[36]

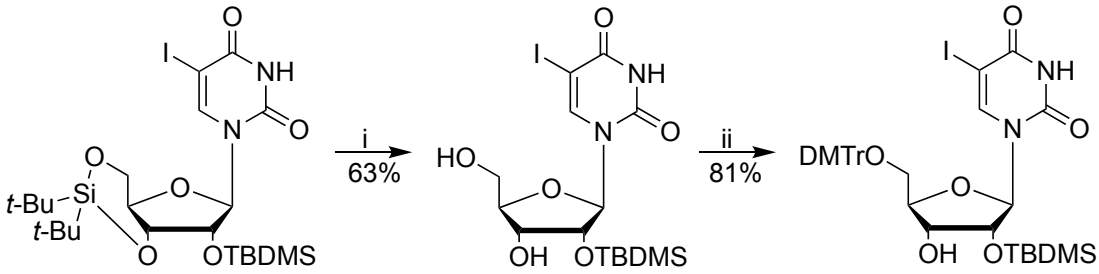
Hence, work focused on the synthesis of the phosphoramidite of 5-iodouridine. An analogous strategy to that of the protection of m^{6}A phosphoramidite (**63**), was applied to the protection of the 5-iodouridine nucleoside. The DTBS group was chosen to selectively protect the 3'- and the 5'-hydroxyl groups of 5-iodouridine, and the 2'-hydroxyl group was then protected with TBDMSCl to give 3'5'-*O*-bis(*tert*-butyl)silyl-2'-*O*-dimethyl(*tert*-butyl)silyl-5-iodouridine (**70**; Figure 3.28) in an apparent quantitative yield.



69 **70**

*Figure 3.28 Reaction conditions: i) **54** (1.1eq), imidazole (2eq), DMF, 0°C, 1hr; ii) TBDMSCl (1.2eq); imidazole (3eq), 60°C, 12hr.*

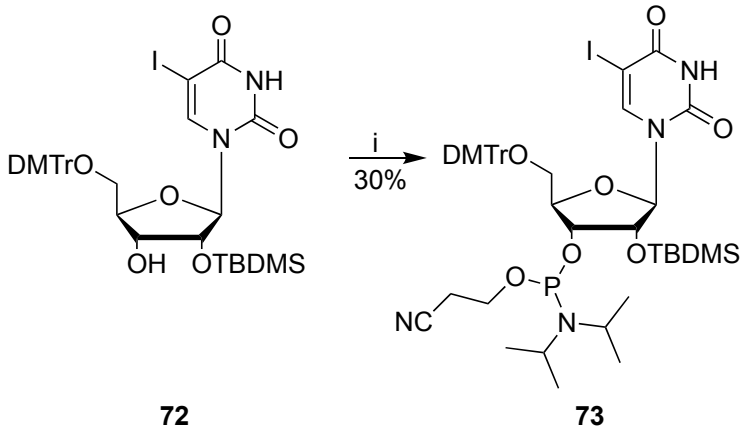
The DTBS group of **70** was then selectively cleaved using dilute HF.pyridine resulting in 2'-*O*-dimethyl(*tert*-butyl)silyl-5-iodouridine (**71**; Figure 3.29) (63%). The 5'- hydroxyl group of **71** was then protected with the DMTr group resulting in 5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-dimethyl(*tert*-butyl)silyl-5-iodouridine (**72**; Figure 3.30) (81 %).



70 **71** **72**

Figure 3.29 Reaction conditions: i) HF.pyridine (3 eq), pyridine, 0°C, 5hr; ii) DMTrCl(1.2 eq), pyridine, 0°C, 12hr.

Protection of the 3'- hydroxyl group of **72** with the phosphitylating reagent (**64**) resulted in 5'-*O*-(4,4'-dimethoxytrityl)-3'-*O*-[(2-cyanoethoxy)(*N*,*N*-diisopropylamino)phosphino]-2'-*O*-dimethyl(*tert*-butyl)silyl-5-iodouridine(**73,** 30 %). Care was taken in chromatographic purification of the crude reaction mixture and the fractions were collected under argon.



72 **73**

Figure 3.30 Reaction conditions: ***64*** *(2.5eq), DIPEA (10 eq), CH_2Cl_2, 0°C, 6hr.*

Hence, the synthesis of the 5-iodouridine phosphoramidite (**73**) was completed successfully for the future work involving synthesis of ss-RNA containing both **73** and the m^6A phosphoramidite (**63**) for use as an FTO substrate.

3.4.3 Synthesis of analogues of m^6A

Following the successful completion of the synthesis of the m^6A and 5-iodouridine phosphoramidite (**63** and **73,** respectively), work focused on the synthesis of analogs of m^6A with the aim of investigating unreported, natural and unnatural, substrates of FTO. Thus, the synthesis of N^6,N^6-dimethyladenosine (m_2^6A, **74**; Figure 3.31), N^6-ethyladenosine (e^6A, **74**; Figure 3.32) and N^6-cyclopropyladenosine (**75**; Figure 3.31), and N^6-(2-hydroxylethyl) adenosine (**76**; Figure 3.31) was attempted using procedures, analogous to those used for the synthesis of m^6A (**53**; Figure 3.23).

3.4.3.1 Synthesis of $m^6{}_2A$, N^6-cyclopropyladenosine, and N^6-(2-hydroxylethyl)adenosine

The hydroxyl groups present on the sugar ring of inosine (**65**) were protected via acylation to form **66**, which was then converted to **67** via chlorination using the *Vilsmeier* reagent (Figure 3.22). Compound **67** was then treated with an excess of dimethylamine for 24 hr to give $m^6{}_2A$

(**74**); treatment of **67** with an excess of cyclopropylamine gave N^6-cyclopropyladenosine (**75**), and treatment of **67** with an excess of ethanolamine gave N^6-(2-hydroxylethyl)adenosine (**76**).

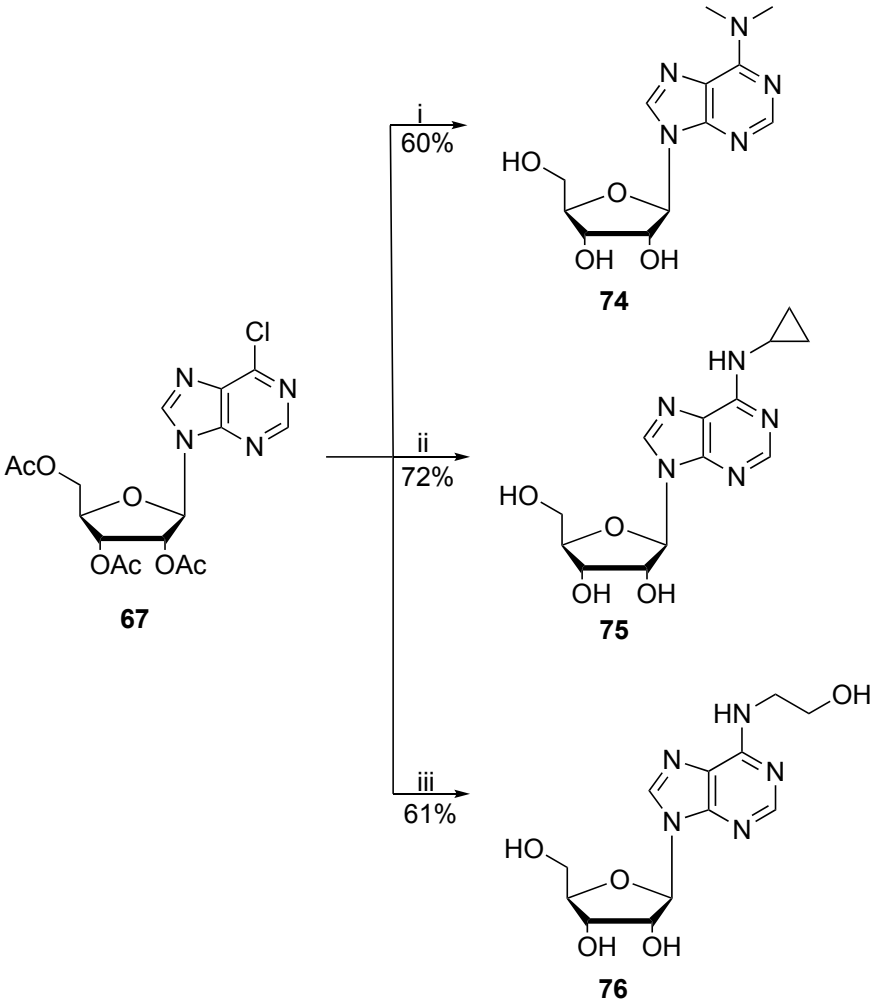
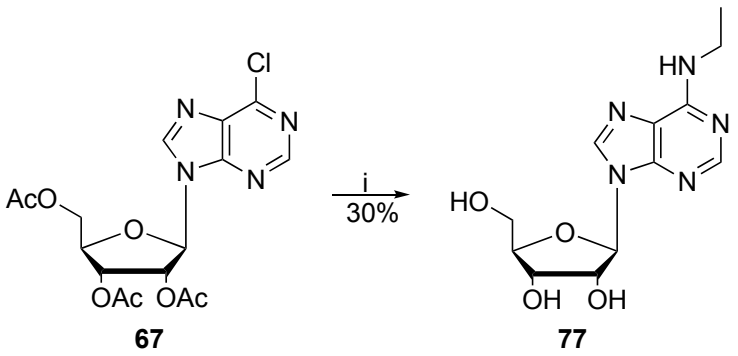


Figure 3.31 Reaction conditions; i) Me_2NH (33% in EtOH, 7 eq), EtOH, rt, 18hr; ii) cyclopropylamine (33% in EtOH, 7 eq), EtOH, rt, 24hr; iii) ethanolamine (5 eq), EtOH, rt, 18hr.

3.4.3.2 Synthesis of N^6-ethyladenosine (e^6A)

Ethylamine solution is not commercially available; hence it was prepared *in situ* by mixing the hydrogen chloride salt of ethylamine and base with subsequent addition of **67** to the reaction mixture gave N^6-ethyladenosine (**77**) (30 %).



67 **77**

Figure 3.32 Reaction condition: $EtNH_2.HCl$ (10 eq), Ag_2O (12 eq), EtOH, rt, 48hr.

The synthesis of **74-77** was carried out, with the aim of investigating the flexibility and selectivity of substrate binding pocket of FTO. Indeed, $m^6{}_2A$ was observed to be oxidised by FTO catalysis (preliminary data not shown, involving mass spectrometric analyses). This result piqued interest in investigating the potential regioselectivity of the site of catalysis by FTO. It was interesting to investigate whether FTO shows any unreported preference in catalysing oxidation of the methyl group syn/anti to the *N*-1 nitrogen of adenosine (Figure 3.33). It was envisaged that rotation of the N^6-methyl group around the bond could be restricted by linking the N^6 and the N^1 nitrogens via a ring. It was also considered that the investigation of methylethanoadenosine (**105**; Figure 3.41) as a substrate for FTO, may provide insights into the regioselectivity of FTO. The syntheses of ethanoadenosine (**100**), methylethanoadenosine (**105**) and ethylethanoadenosine (**106**; Figure 3.41) was thus attempted.

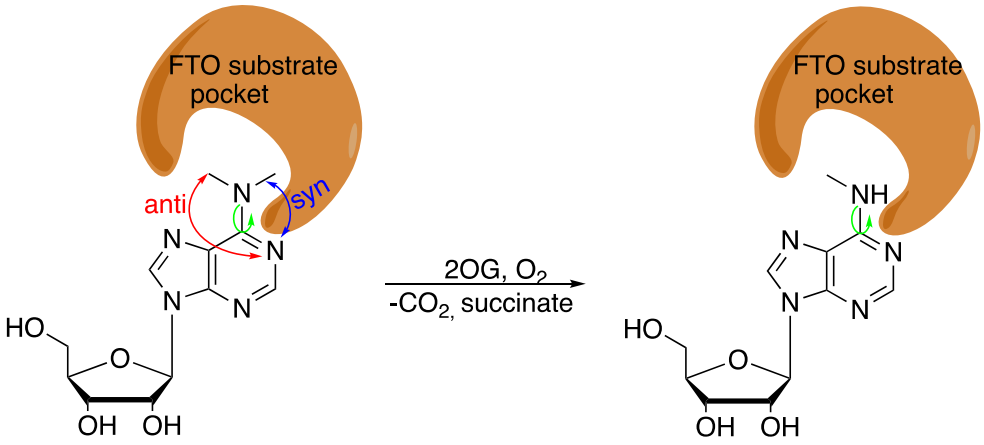


Figure 3.33 Schematic representation of the proposed catalysis of $m^6{}_2A$ by FTO. The methyl group syn and anti to the N-1 of adenosine are highlighted. The green arrow indicates bond rotation.

*3.4.3.3 Synthesis of ethanoadenosine (**100**), methylethanoadenosine (**105**) and ethylethanoadenosine (**106**)*

3.4.3.3.1 Optimisation of the synthesis of ethanoadenosine (**100**)

The hydroxyl groups of inosine (**65**) were protected via acylation to give **66** which was then treated with the *Vilsmeier* reagent to give **67** (Figure 3.22). Compound **67** was then treated with methylethanolamine or ethylethanolamine. However, the presence of amines led to the undesired base mediated hydrolysis of the ester protecting groups, with subsequent conversion to *N*6,*N*6-methyl(2-hydroxyethyl)adenosine (**78**; Figure 3.34) or *N*6,*N*6-ethyl(2-hydroxyethyl)adenosine (**79**; Figure 3.34). The next reaction investigated was cyclisation of the 2-hydroxyethyl group of **78** or **79**; it was envisaged that the conversion of the hydroxyl group of the 2-hydroxyethyl group to a better leaving group and the subsequent nucleophilic attack by the *N*-1 of the adenosine ring, should result in the desired products (**105** or **106,** Figure 3.34). Indeed, it is reported[37] that an analogous hydroxyl group can be converted to an iodo group and subsequent attack by the *N*-1enables the formation of cyclised product. However, the addition of (triphenoxyl)phosphonium-methyliodide[37] to **78** resulted in the formation of a mixture of products, likely due to the presence of other hydroxyl groups on **78**.

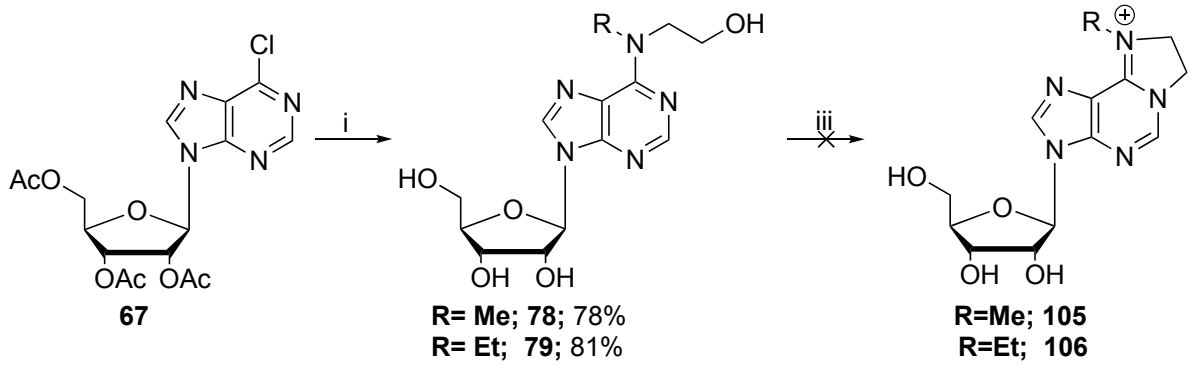


Figure 3.34 Reaction conditions: i) methylethanolamine /ethylethanolamine (5 eq), EtOH, rt, 18hr; ii) (triphenoxy)phosphoniummethyl iodide (1.2 eq), trimethylamine (5eq) DMF, rt, 1 hr.

It was envisaged that hydrolytically stable protection of the hydroxyl groups on the sugar ring of inosine (**65**), prior to S_NAr reaction with the amine (methylethanolamine or ethylethanolamine) may solve the problem. Hence, the hydroxyl groups of the sugar ring of inosine (**65**) were protected using TBDMSCl groups to give 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]inosine (**82**; Figure 3.35), in the presence of imidazole. The resultant **82** was then treated with the *Vilsmeier* reagent intending that the reaction would result in the formation of 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-6-chloroinosine (**83**; Figure 3.35), analogous to the reaction of the *Vilsmeier* reagent with **66** to give **67** (Figure 3.22). However, treatment of **82** with the *Vilsmeier* reagent led to the degradation of the starting material into a mixture and were not characterised.

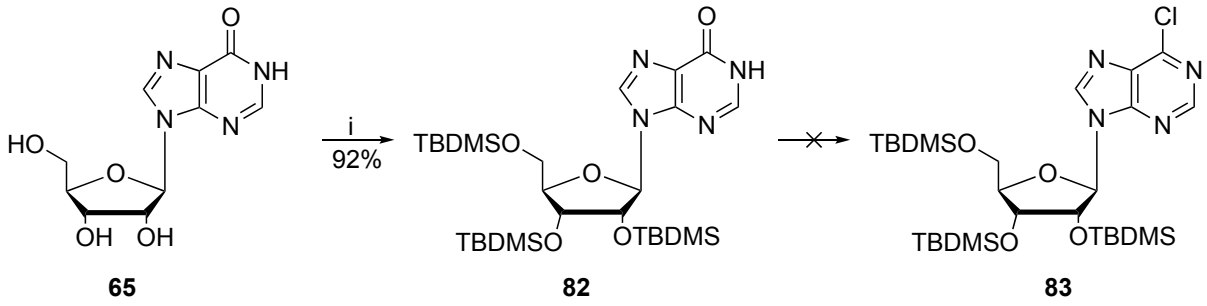


Figure 3.35 Reaction conditions: TBDMSCl (3.3 eq), DMF, 60°C, 12hr; ii) Vilsmeier reagent (2eq), CH_2Cl_2, 40°C, 24hr.

Consistent with my results, it has been reported that use of the *Vilsmeier* reagent for chlorination is not compatible with the TBDMS group, leading to the formation of a formate ester (Figure 3.36). A proposed mechanism for degradation of the TBDMS group in the presence of the *Vilsmeier* reagent is given in Figure 3.36.[38]

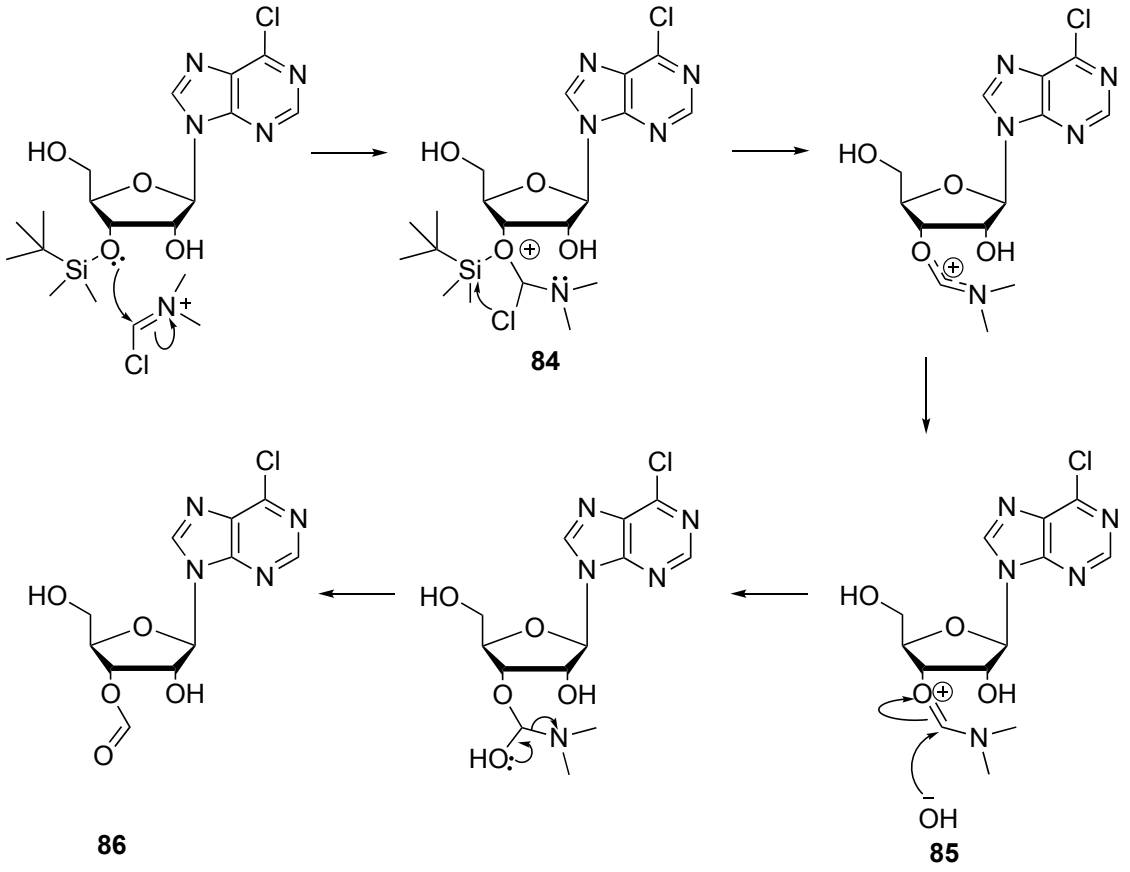


*Figure 3.36 Proposed mechanism of the degradation of TBDMS group in the presence of Vilsmeier reagent and the subsequent formation of the formate group (**86**).*[38]

According to the proposed mechanism,[38] the nucleophilic silyl ether oxygen attacks the *Vilsmeier* reagent leading to the formation of oxonium cation (**84**); the proximity of the chloro group with the silicon atom is likely to favour the nucleophilic attack on the later. The neutral TBDMSCl group leaves leading to the formation of imidate salt (**85**) which on hydrolysis resulted in the formation of a formate ester (**86**).

Following the unsuccessful attempt to synthesise the protected 6-chloroinosine (**83**), it was envisaged that the reaction of **82** with $POCl_3$ with ethanolamine as a base, should be investigated. **82** was treated with excess $POCl_3$ in the presence of a base and 1,2,4-triazole acting as a catalyst. The reaction resulted in the formation of the intermediate 2'3'5'-tris(*tert*-butyldimethylsilyl)-6-(1*H*-1,2,4-triazol-1-yl)-adenosine (**87**; Figure 3.37); however, the conversion of **87** to 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-*N*6-(2-hydroxyethyl)adenosine (**88**; Figure 3.37) was not accrued even after 24 hr reflux.

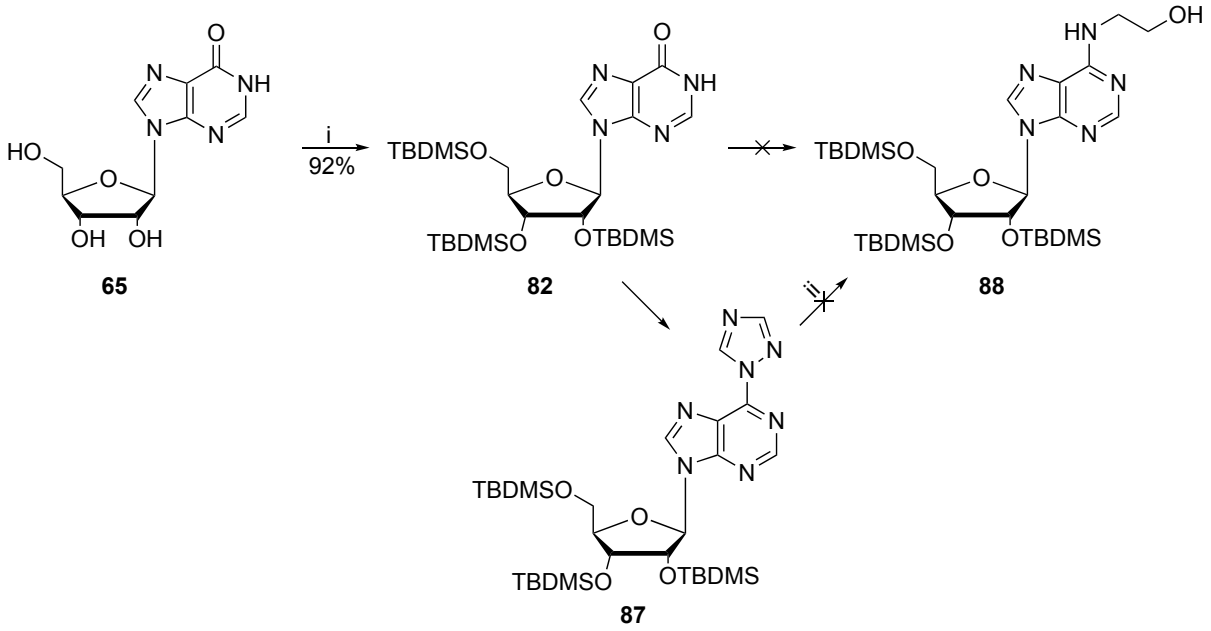


Figure 3.37 Reaction conditions: i) 1,2,4-triazole (18 eq), Et_3N (18 eq), $POCl_3$ (4 eq), CH_3CN, 12hr; ii) Ethanolamine (10 eq), EtOH, rt, 24hr.

Silyl- and the acyl- protecting groups are most commonly used for alcohols in oligonucleotide synthesis.[21,26] Following the failures of both of these groups to provide desired products, work then focussed on finding an alternative pathway for the synthesis of the protected analogs of **78** or **79** (i.e. compound **91** or **93** in Figure 3.38). The synthesis of m^6A analogues using 1-*H*-benzotriazol-1-yloxytris(dimethylamino)phosphoniumhexafluoro-phosphate (BOP, **89**; Figure 3.38) and DIPEA has been reported.[39,40] Thus, the TBDMS protected inosine (**82**) was reacted with methylethanolamine in the presence of PyBOP (**90**;

Figure 3.38) and DIPEA to give 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-*N*6,*N*6-methyl(2-hydroxyethyl)adenosine (**91**; Figure 3.38) (70%). It was noted that the use of BOP (**89**) rather than PyBOP, increased the yield to more than 90%. Subsequently, it was observed that use of 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU), as base, afforded cleaner reaction without side products. Compound **82** was then treated with BOP (**89**) in the presence of DBU as base to give **94** (Figure 3.38) intermediate, at 40°C, in a closed vial; excess ethanolamine or ethylethanolamine was added to the reaction mixture to give 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-*N*6-(2-hydroxyethyl)adenosine (**88**; Figure 3.38) or 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-*N*6,*N*6-ethyl(2-hydroxyethyl)adenosine (**93**; Figure 3.38) in an apparent quantitative yield.

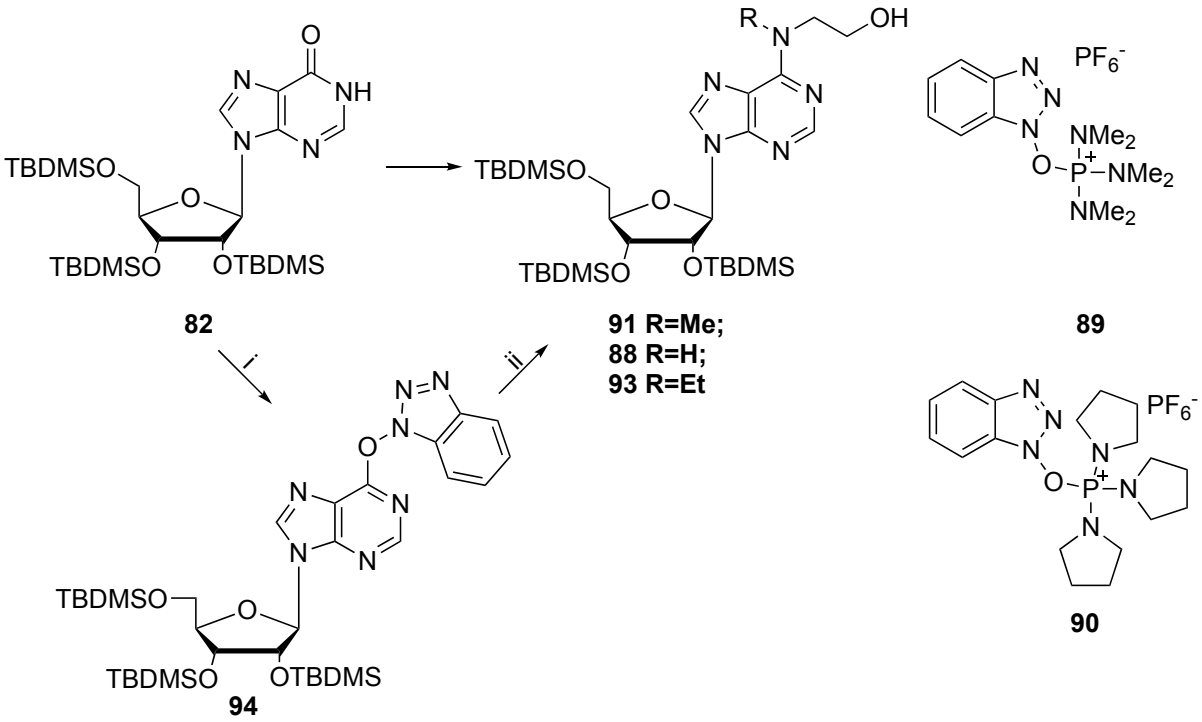


*Figure 3.38 Reaction conditions: i) BOP (**89**;1.2 eq)/ PyBOP (**90**,1.2 eq), DBU (2 eq), EtOH, 40°C, 40 min; ii) methylethanolamine (5eq) or ethanolamine (5 eq) or ethylethanolamine, rt, 4-12hr.*

A likely mechanism for the BOP mediated S_NAr reaction is given in Figure 3.39.[41] According to the proposed mechanism, the base abstracts the proton from the hydroxyl group of the aromatic ring of inosine; the oxide ion, thus formed, attacks the electrophilic phosphorus

centre forming a strong P-O bond, resulting in the formation of (acyloxy)phosphonium salt intermediate (**95**; Figure 3.39); the oxybenzotriazole anion ($^{-}$OBt) attacks the electrophilic centre to form the intermediate (**96**; Figure 3.39) which is subsequently replaced by a nucleophile via S_NAr reaction to form **97** (Figure 3.39).[41]

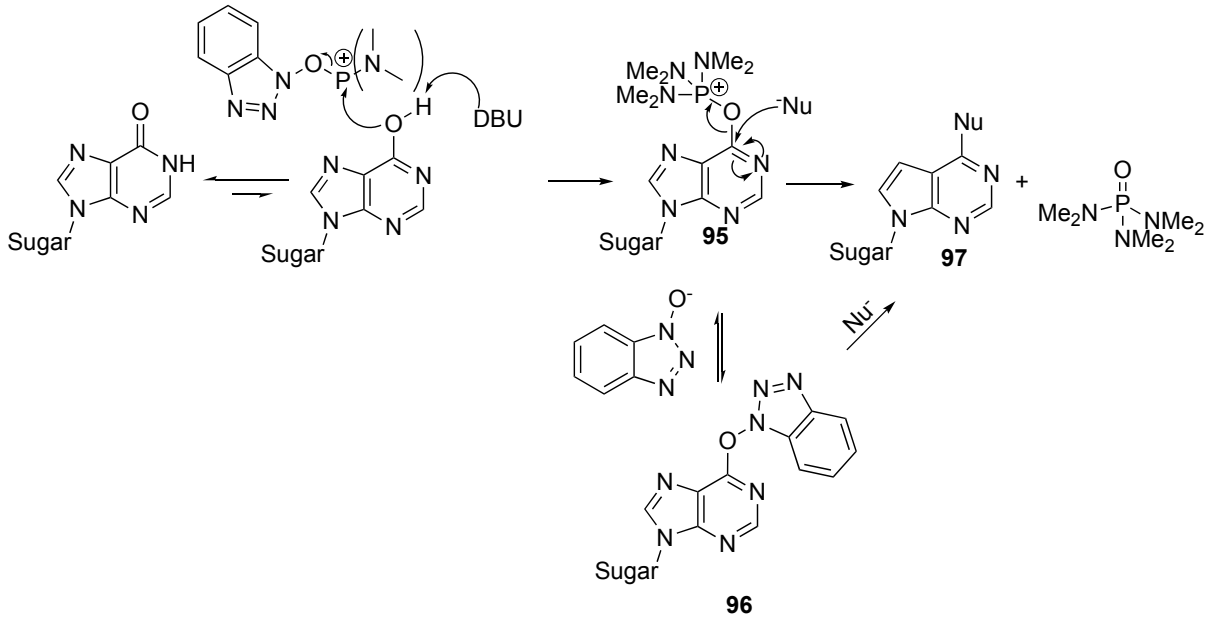


Figure 3.39: Proposed mechanism for the BOP-mediated S_NAr reaction.[41]

Compound **88** (Figure 3.38) was then treated with (triphenoxy)phosphoniummethyl iodide resulting in formation of proposed intermediate 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-*N*6-(2-iodoethyl)adenosine (**98**) which undergoes proposed instantaneous cyclisation to give the desired product 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-*N*6-ethanoadenosine (**99**) (60 %). The final step in the planned synthesis of ethanoadenosine (**100**) was the deprotection of the TBDMSCl group; several deprotection procedures are reported.[21,26] The deprotection of the TBDMS group using TBAF requires a prolonged work-up procedure;[21,29,36,42] however, a chromatographically pure product was not obtained even after running flash column chromatography twice. Hence, 3HF.Et_3N was used as a deprotection reagent.[21,26,43] The deprotection of **99** using 3HF.Et_3N gave ethanoadenosine (**100**; Figure 3.40) (70 %); a purified product was obtained via precipitation with DCM as solvent. It should be noted that

the synthetic route for **100** delineated herein (Figure 3.39 and Figure 3.40), has not been reported in the literature.

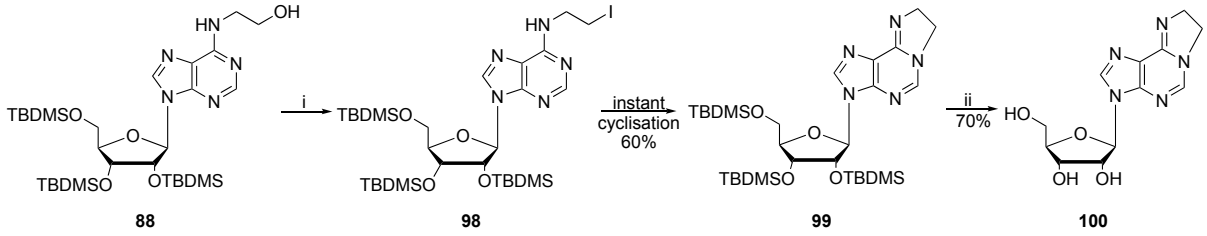


Figure 3.40: Reaction conditions: i) (triphenoxy)phosphoniummethyliodide (1.2 eq), Et_3N (5 eq), DMF, rt, 30min; ii) 3HF.Et_3N (4 eq), DCM, rt, 24hr.

The treatment of **91** or **93** with (triphenoxy)phosphonium-methyl iodide results in the formation of positively charged rings products (**103** or **104**, respectively; Figure 3.41). The purification of crude 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-*N*6,*N*6-methylethanoadenosine **(103)** or 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-*N*6,*N*6-methylethanoadenosine **(104)** by neutral alumina column resulted in **103** and **104**. The final step of the synthesis was the deprotection of the TBDMS group of **103** and **104** which was conveniently achieved by using 3HF.Et_3N solution to give *N*6,*N*6-methylethanoadenosine (**105**; Figure 3.41) or *N*6,*N*6-ethylethanoadenosine (**106**; Figure 3.41) in reasonable yields.

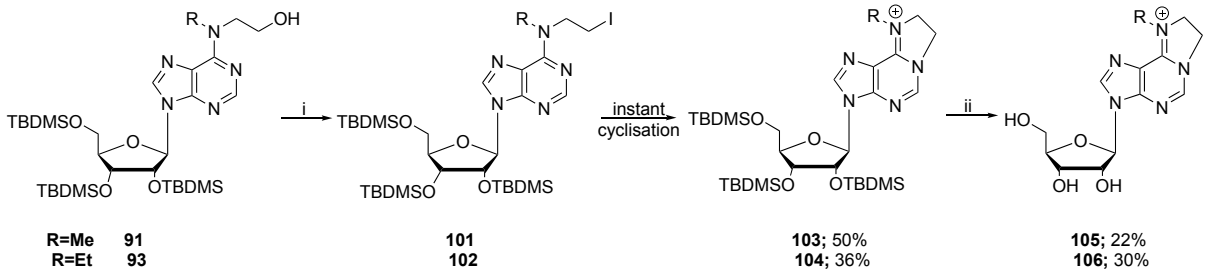


Figure 3.41 Reaction conditions: i) $(OPh)_3$Ph-MeI (1.2 eq), Et_3N (5 eq), DMF, rt, 30 min; ii) 3HF.Et_3N (4 eq), DCM, rt, 24hr.

Following the successful synthesis of **100** (Figure 3.40), **105** (Figure 3.41) and **106** (Figure 3.41), it was envisaged that 6-*O*-methoxyinosine (**112**; Figure 3.43) and 6-*O*-

methoxyguanosine (**113**; Figure 3.43) could be synthesised using an analogous route, with methanol as a nucleophile. Hence, reaction conditions were investigated for the synthesis of 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-6-*O*-methoxyinosine (**107**; Figure 3.42) or 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-6-*O*-methoxyguanosine **(111**; Figure 3.43). Thus, protected inosine (**82**; Figure 3.42) was treated with BOP (**89**) in the presence of DBU which led to the formation of intermediate (**94**; Figure 3.42), as observed by TLC and later characterised. However, attempts for S_NAr reaction at the C-6 of **94** using methanol were not successful, even when an excess of methanol was used. The reaction mixture was heated at 35°C for 12hr; however, no conversion to the desired product **107** was observed via TLC.

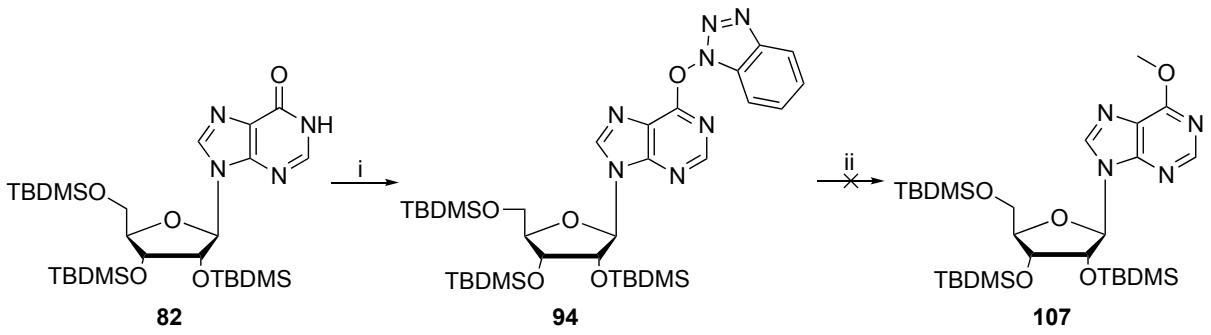


*Figure 3.42 Reaction conditions: i) BOP (**89**;1.2 eq), DBU (2 eq), EtOH, 40°C, 1hr; ii) methanol (10eq), EtOH, rt, 24hr.*

As reported in the literature,[39] it was noticed that evaporating the reaction mixture after the formation of **94** and the subsequent addition of methanol and base led to the formation of the desired product (**107**; Figure 3.43) in an apparent quantitative yield. Analogous reaction conditions were used with the TBDMS protected guanosine 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silylguanosine (**109**; Figure 3.43) which resulted in the formation of intermediate (**110**; Figure 3.43) and subsequent conversion to 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl-*O*6-methylguanosine (**111**; Figure 3.43) in 85% yield. The final step of the synthesis was the deprotection of the TBDMS group which was conveniently achieved using $3HF.Et_3N$,

resulting in the formation of O^6-methylinosine (**112**; Figure 3.43) or O^6-methylguanosine (**113**; Figure 3.43) in reasonable yield.

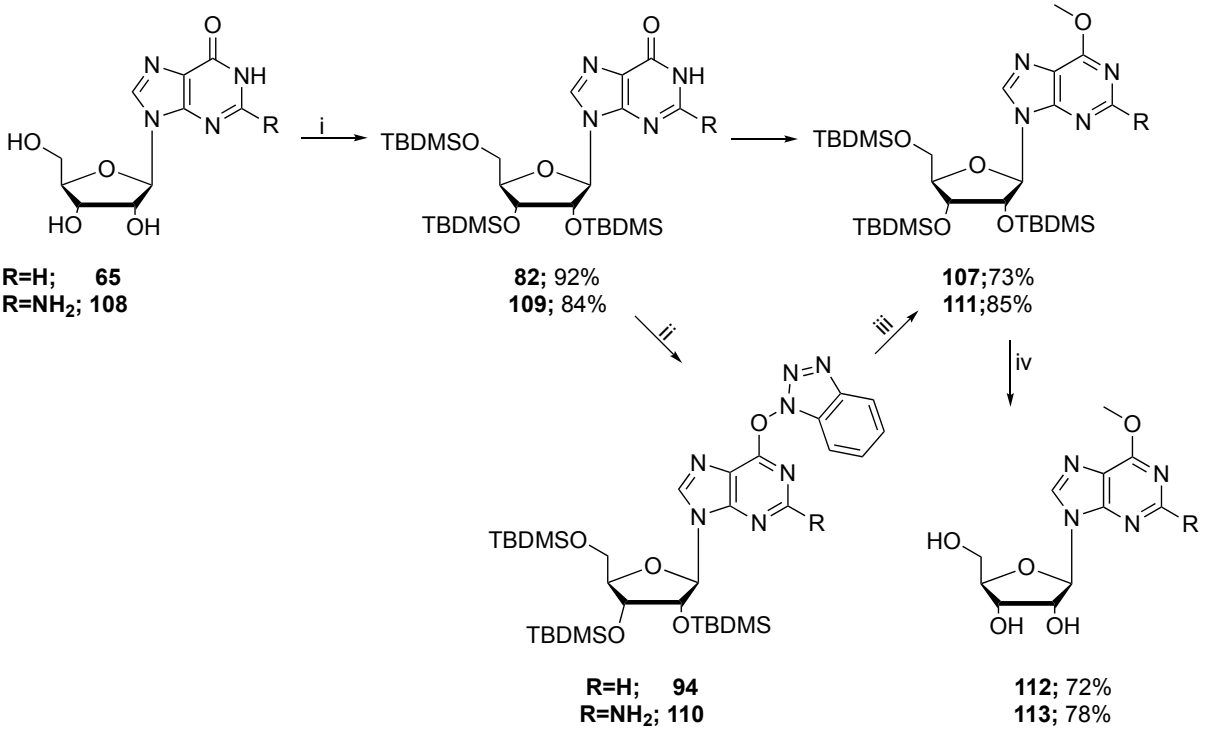


*Figure 3.43 Reaction conditions: i) TBDMSCl (3.3 eq), DMF, 60°C, 12hr; ii) DBU (4 eq), BOP (**89**, 2eq), 40°C, 40 min; iii) DBU (2 eq), MeOH (excess), 4 hr; iv) 3HF.Et_3N (4 eq), DCM, rt, 24hr.*

Following the successful synthesis of compounds **100**, **105**, **106**, **112**, and **113**, it was then envisaged that m^6A (**53**) could be synthesised via S_NAr reaction on inosine (**65**) using methylamine, BOP (**89**) and DBU. Thus, inosine (**65**) was dissolved in ethanol and heated in a closed vial in the presence of BOP (**89**) and DBU to give intermediate **114** (Figure 3.44), as observed by TLC and mass spectrometric analysis; excess methylamine was then added to the reaction mixture at room temperature resulting in the formation of m^6A (**53**) in reasonable yield (53%).

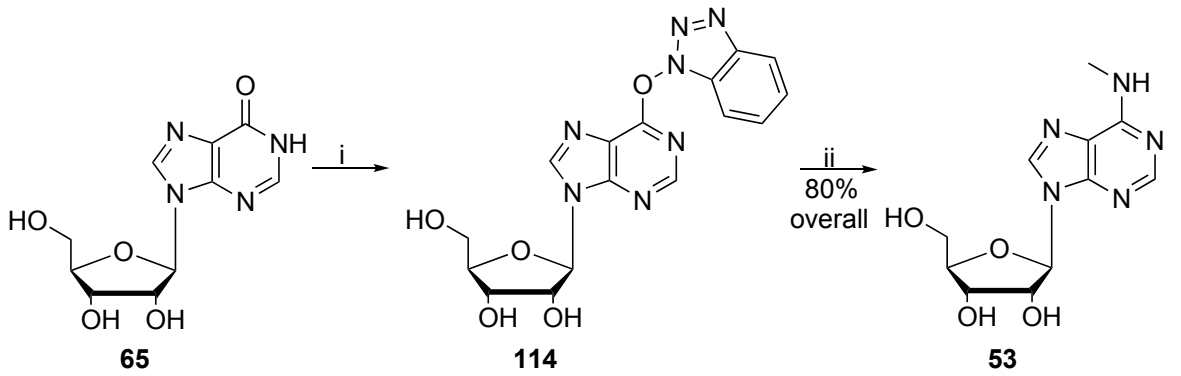


*Figure 3.44 Reaction conditions: i) BOP (**89**;1.2 eq, DBU (2 eq), EtOH, 40°C, 4hr; ii) methylamine (5eq), EtOH, rt, 24hr.*

It was envisaged that the yield of the reaction may be lowered due to insolubility of inosine in ethanol, nevertheless, the synthesis of m^6A (**53**) could be achieved from inosine (**65**) in a single step. The solvent was then changed to DMF, the yield of the reaction increased to approximately 80%. Hence, it was envisaged that the S_NAr reaction employing BOP and DBU could be used to efficiently synthesise the analogs of m^6A in a single step.

It was also envisaged that an analogous approach could be used for the synthesis of the m^6A phosphoramidite (**63**). Thus, the 3'- and the 5'- hydroxyl groups of the sugar ring of inosine (**65**) were selectively protected using DTBS ditriflate and the 2'- hydroxyl group was protected with the TBDMS group in the presence of imidazole resulting in an apparent quantitative yield of 3'5'-*O*-(di-*tert*-butyl)silyl-2'-*O*-dimethyl(*tert*-butyl)silylinosine (**115**; Figure 3.45). **115** was then treated with BOP (**89**) in the presence of DBU giving intermediate **116** (Figure 3.45); excess methylamine was added to the reaction mixture resulting in the formation of **60** (83 %).

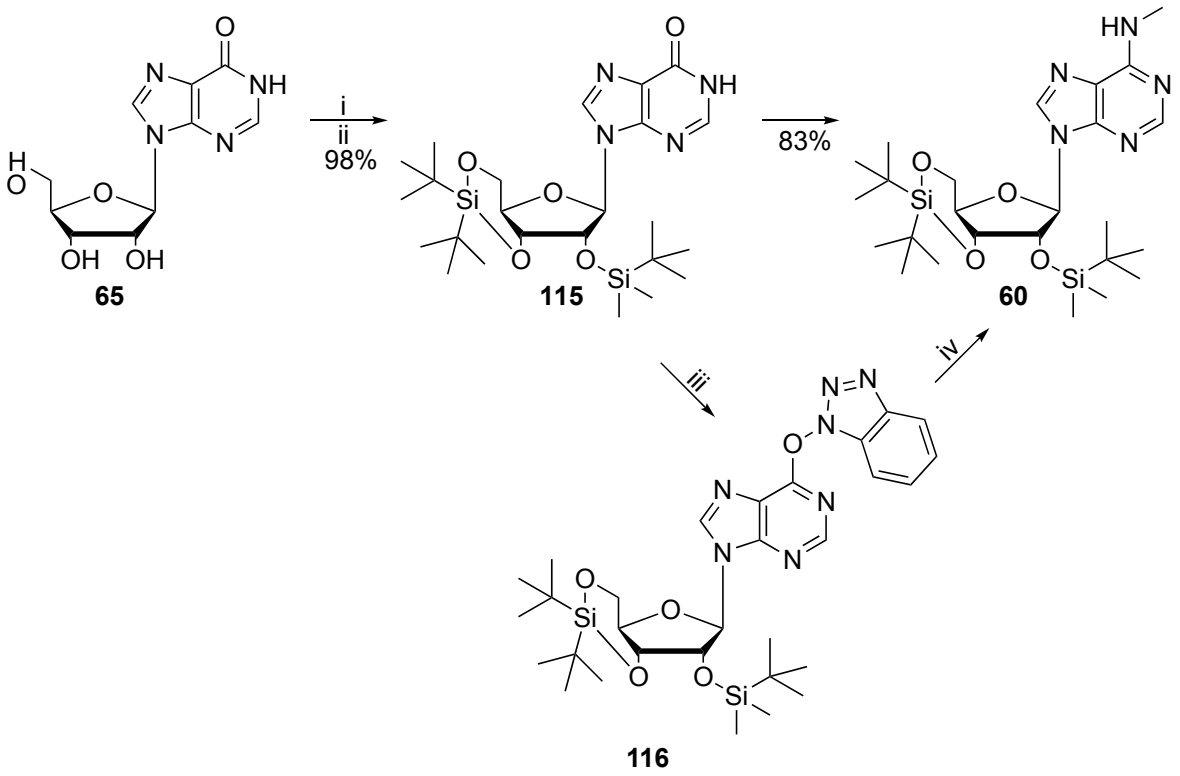


Figure 3.45 ***54*** *(1.1eq), imidazole (2eq), DMF, 0°C, 1hr; ii) TBDMSCl (1.2eq), imidazole (3eq), 60°C, 12hr; iii) DBU (4 eq), BOP (****89****, 2eq), THF, 40°C, 40 min; iv) methylamine (5 eq), 4hr, rt.*

The next step was the selective deprotection of the DTBS group using dilute HF.pyridine solution to give **61** which was subsequently protected at its 5'-hydroxyl group using DMTr group to give **62** (Figure 3.46). Protection of the 3'- hydroxyl group by phosphoramidite (**64**) gave the desired m^6A phosphoramidite (**63**; Figure 3.46). Overall, this route provides m^6A phosphoramidite (**63**) in five steps compared to the seven steps used previously; the shorter synthetic route also increased the yield to approximately 25 % compared to the previous yield of approximately 8 %. The overall synthetic scheme to the m^6A phosphoramidite (**63**) is shown in Figure 3.46.

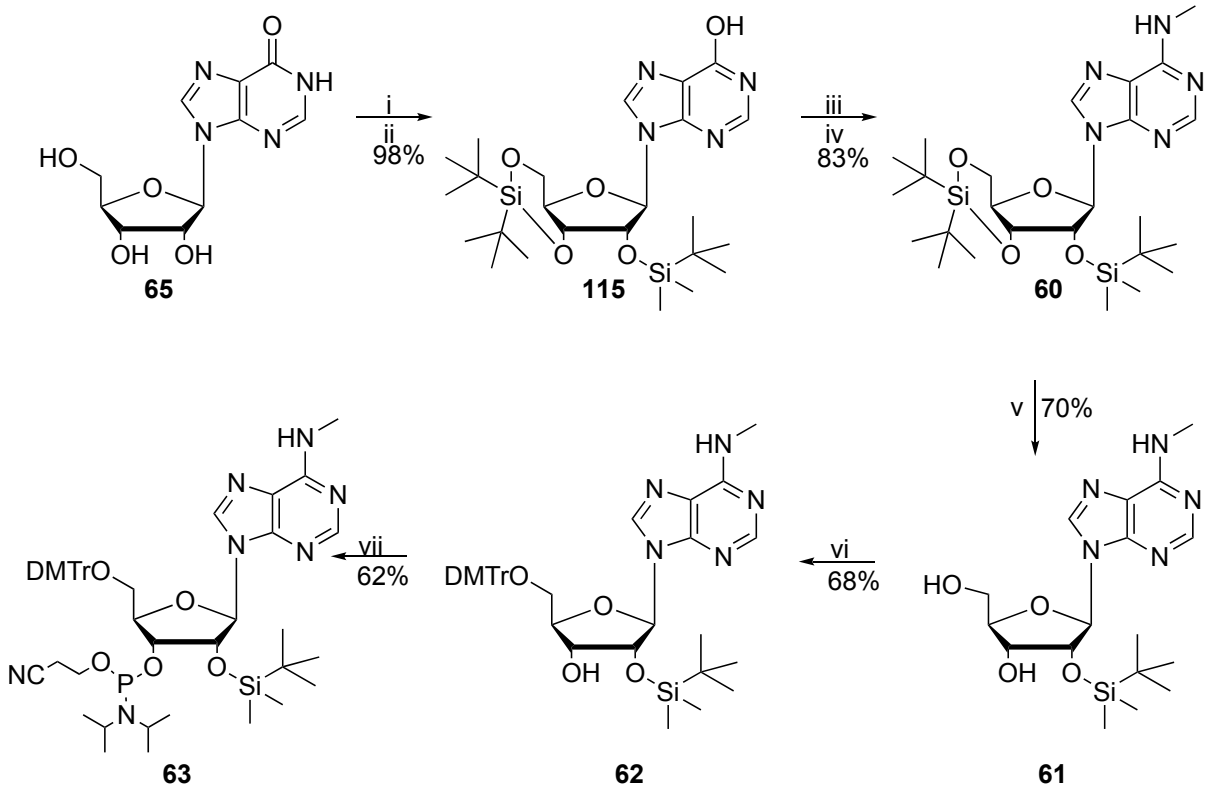


*Figure 3.46 Optimised unreported synthetic route of m^{6}A phosphoramidite. Reaction conditions: i) **54** (1.1eq), imidazole (2eq), DMF, 0°C, 1hr; ii) TBDMS-Cl (1.2eq), imidazole (3eq), 60°C, 12hr; iii) DBU (4 eq), BOP (**89**, 2eq), THF, 40°C, 40 min; iv) methylamine (5 eq), rt, 4hr; v) HF.pyridine (3 eq), pyridine, 0°C, 5hr; vi) DMTrCl (1.2 eq), pyridine, 0°C, 12hr; vii) **64** (2.5eq), DIPEA (10 eq), CH_2Cl_2,0°C, 6hr.*

3.4.3.4 Synthesis of 5-methylcytide (5mC), 5-hydroxymethylcytidine (5hmC) and 5-formylcytidine (5fC)

Following the successful optimisation of the synthetic route of m^{6}A phosphoramidite (**63**), work then focussed on the synthesis of 5mC, 5hmC, and 5fC in order to investigate their reactions with formaldehyde via NMR (Chapter 4). The two hydroxyl groups present on the sugar ring of thymidine (**117**; Figure 3.47) were protected using TBDMSCl group in the presence of imidazole, resulting in an apparent quantitative yield of 3',5'-*O*-bis[dimethyl(*tert*-butyl)silyl]thymidine (**118**; Figure 3.47) which was then treated with BOP (**89**) in the presence of DBU to give the intermediate (**119**, as observed via LCMS; Figure 3.47); excess methanolic ammonia was then added to the reaction mixture, however, no

conversion to the desired product 3',5'-*O*-bis[dimethyl(*tert*-butyl)silyl]cytidine (**120**; Figure 3.47) was observed.

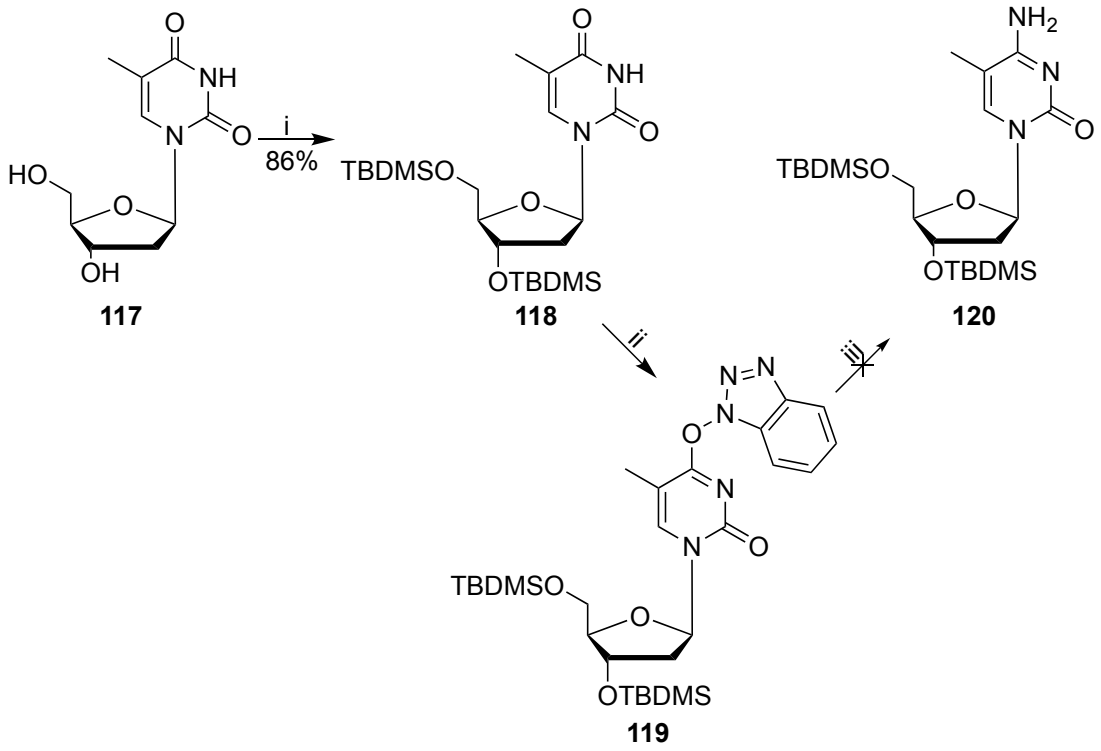


Figure 3.47 Reaction conditions: i) TBDMSCl (2.4 eq), imidazole, DMF, rt, 4hrs; ii) BOP (2 eq), DBU (2 eq), EtOH, 40°C, 40 min; iii) 7M NH_3 in MeOH, 24hrs, rt.

Hence, attention was turned to an alternative synthetic procedure using $POCl_3$, Et_3N, and 1,2,4-triazole (as a catalyst), resulting in the formation of **120** (Figure 3.48)(60 %).[44] Deprotection of **120** using 3HF.Et_3N gave 5-methylcytidine (**121**; Figure 3.48) in reasonable yield.

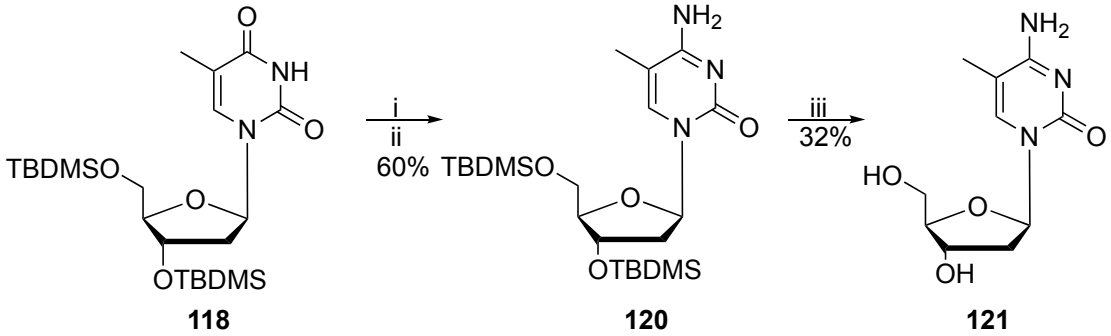
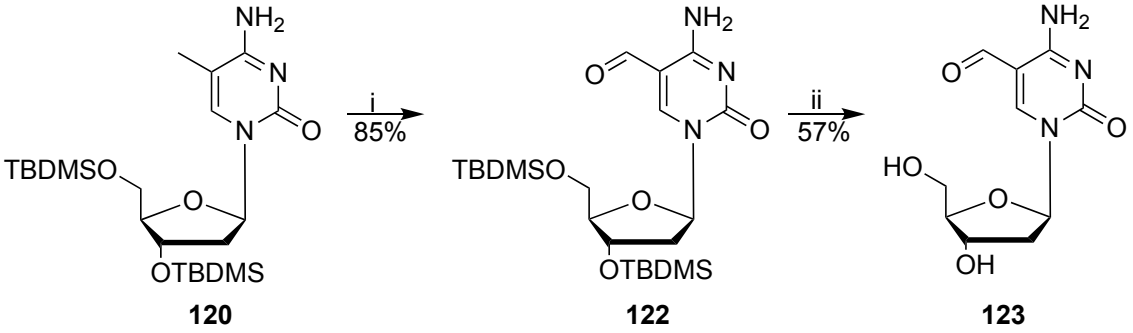


Figure 3.48 Reaction conditions: i) 1,2,4-triazole (18 eq), Et_3N (18 eq), $POCl_3$(4 eq), CH_3CN, 12hr; ii) NH_4OH, dioxane, rt, 4hr; iii) 3HF.Et_3N (2.2eq), DCM, 24hr, rt.

It was envisaged that the controlled oxidation of the 5-methyl group of **120** could result in 5-formylcytidine, and the reduction of the formyl group to hydroxymethyl group could lead to the formation of 5-hydroxymethylcytidine, as reported.[45–47] Compound **120** was thus subjected to oxidation using $K_2S_2O_8$, $CuSO_4.5H_2O$ and 2,6-lutidine in MeOH: H_2O, leading to the formation of the protected 5-formylcytidine, 3',5'-*O*-bis[dimethyl(*tert*-butyl)silyl]-5-formylcytidine (**122**; Figure 3.49), as reported.[46–48] The next step, i.e, the deprotection of **122** was conveniently achieved using $3HF.Et_3N$ giving 5-formylcytidine (**123**; Figure 3.49) in an apparent quantitative yield (crude product).



120 **122** **123**

Figure 3.49 Reaction conditions: i) $K_2S_2O_8$, $CuSO_4.5H_2O$, 2,6-lutidine, MeOH: H_2O, rt, 3hr; ii) $3HF.Et_3N$ (2 eq), DCM, rt, 24hr.

Following the synthesis of 5fC (**123**), the next objective was the reduction of the formyl group to the hydroxymethyl group. Hence, **122** was subjected to reduction using sodium borohydride and $CeCl_3.7H_2O$[45] resulting in formation of 3',5'-*O*-bis[dimethyl(*tert*-butyl)silyl]-5-hydroxymethylcytidine (**124**). Deprotection of the TBDMS group of **124** was conveniently achieved using $3HF.Et_3N$; the purified product 5-hydroxymethylcytidine (**125**) was precipitated using DCM.

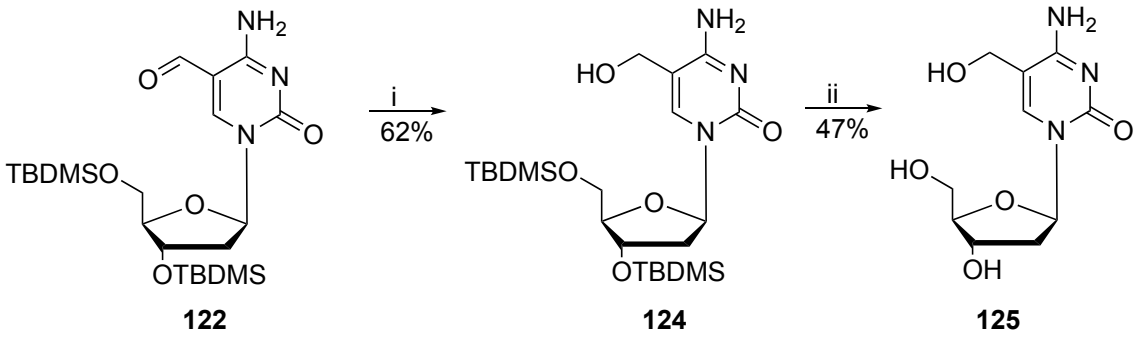


Figure 3.50 Reaction conditions: i) $NaBH_4$, $CeCl_3.7H_2O$, MeOH, rt, 1 hr; ii) $3HF.Et_3N$ (2 eq), DCM, rt, 24hr.

Hence, 5meC (**121**; Figure 3.49), 5fC (**123**; Figure 3.49) and 5hmC (**125**; Figure 3.50) were successfully synthesised.

3.5 Conclusions

The work described in this chapter aimed at the synthesis of m^6A phosphoramidite (**63**). An important initial objective of the work was to optimise the previously reported synthetic procedures for m^6A phosphoramidite (**63**). Work then focused on the synthesis of analogs of m^6A to investigate natural and the unnatural substrates of FTO. The work was also aimed at the synthesis of 5-iodouridine phosphoramidite (**73**) for use in biophysical analysis. A final objective of the synthetic studies was the synthesis of 5mC, 5hmC, and 5fC.

The synthetic route for the synthesis of the m^6A phosphoramidite (**63**) was reduced to five steps compared to the earlier reported procedures which involve at least seven steps.[36,49] The overall yield of the reaction was increased to approximately 25% from approximately 8% as obtained by using the previously reported synthetic route. To the best of my knowledge, this work is the first description of BOP-mediated S_NAr reaction used in phosphoramidite synthesis. BOP-mediated S_NAr has been widely used for other synthetic purposes;[39–41] however, it has not been reported in phosphoramidite synthesis. The work also defines a new synthetic procedure for the synthesis of ethanoadenosine (**100**), methylethanoadenosine (**105**)

and ethylethanoadenosine (**106**). The procedure developed herein brings down the synthesis to four steps starting from inexpensive inosine (5g, £38), rather than using comparatively expensive 6-chloroinosine[50] (5g, £546). The aim of the synthesis of **100**, **105** and **106** was to investigate the existence of regioselectivity in substrate oxidation by FTO. The fact that $m^6{}_2A$ (**74**) is also a substrate for FTO (preliminary data not shown, involving mass spectrometric analyses) suggests further investigations of other potential FTO substrates will be productive. Various analogues of m^6A i.e. $m^6{}_2A$ (**74**), cyclopropyladenosine (**75**), *N*6-(2-hydroxyethyl)adenosine (**76**), ethyladenosine (**77**), *N*6-methyl(2'-hydroxyethyl)adenosine (**78**), *N*6-ethyl(2'hydroxyethyl)adenosine (**79**), *N*6-ethanoadenosine (**100**), *N*6,*N*6-methylethanoadenosine (**105**), *N*6,*N*6-ethylethanoadenosine (**106**), 6-*O*-methylinosine (**107**) and 6-*O*-methylguanosine (**111**) were synthesised to investigate their potential as FTO substrates. The synthesis of 5-iodouridine phosphoramidite (**73**) was completed successfully with the aim of using it in the preparation of oligonucleotides containing m^6A and 5-iodouridine; the 5-iodouridine could be extended via a diazirine linker with a view to enabling photocrosslinking to FTO via UV irradiation, thus, providing a new strategy to crystallise FTO-substrate complexes. The synthesis of 5meC, 5fC, and 5hmC, the reported substrates for the Ten-Eleven Translocation enzymes,[4,51,52] was successfully achieved, for use in investigations on their reaction with formaldehyde in Chapter 4.

In summary, new synthetic routes to m^6A phosphoramidite (**63**), *N*6-ethanoadenosine (**100**), *N*6,*N*6-methyl(2'-hydroxyethyl)adenosine (**105**) and *N*6,*N*6-ethyl(2'hydroxyethyl)adenosine (**106**), were developed. Additionally, m^6A (**53**), m^6A phosphoramidite (**63**), 5-iodophosphoramidite (**73**), *N*6,*N*6-dimethyladenosine (**74**), *N*6-cyclopropyladenosine (**75**), *N*6-(2'-hydroxyethyl)adenosine (**76**), *N*6-ethyladenosine (**77**), *N*6-ethanoadenosine (**100**), *N*6,*N*6-methyl(2'-hydroxyethyl)adenosine (**105**), *N*6,*N*6-ethyl(2'hydroxyethyl)adenosine

(**106**), *O*6-methylinosine (**107**), *O*6-methylguanosine (**111**), 5mC (**121**), 5fC (**123**) and, 5hmC (**125**) were synthesised for use in various biochemical investigations.

Future work could include synthesis of oligonucleotide containing m^6A and 5-iodouridine, the investigation of the synthesised m^6A analogs as FTO substrates, Further, optimise the synthesis of m^6A phosphoramidite to four step route involving the efficient separation and purification of the TBDMS isomers, **62** and **62'** (Figure 3.21).

3.6 Acknowledgements

I thank Dr. Anna Dysko, University of Oxford, for demonstrating the procedure for the phosphoramidite reaction.

3.7 Reference

1. Limbach, P. A., Crain, P. F. & Mccloskey, J. A. Summary: the modified nucleosides of RNA. *Nucleic Acids Res.* **22,** 2183–2196 (1994).
2. Kriaucionis, S. & Heintz, N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science.* **324,** 929–930 (2009).
3. Song, C.-X. & He, C. The hunt for 5-hydroxymethylcytosine: the sixth base. *Epigenomics* **3,** 521–3 (2011).
4. Tahiliani, M. K., Kian P., Shen, Yinghua, Pastor, W. A., Bandukwala, H.,, Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L. Rao, A. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science.* **324,** 930–935 (2009).
5. Sun, W.-J. Li, Jun H., Liu, S., Wu, J., Zhou, H., Qu, Liang H., Yang, Jian H., RMBase: a resource for decoding the landscape of RNA modifications from high-throughput sequencing data. *Nucleic Acids Res.* **44,** (2016).
6. Serganov, A. & Patel, D. J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.* **8,** 776–790 (2007).
7. He, L. & Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5,** 522–531 (2004).
8. Eddy, S. R. Non–coding RNA genes, and the modern RNA world. *Nat. Rev. Genet.* **2,** 919–929 (2001).
9. Motorin, Y. & Helm, M. RNA nucleotide methylation. *Wiley Interdiscip. Rev. RNA* **2,** 611–631 (2011).
10. Lodish H, Berk A, Zipursky SL. Molecular Cell Biology. 4th edition. New York: W. H. Freeman; 2000. Section 4.4, The Three Roles of RNA in Protein Synthesis. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21603/

11. Cooper, T. A., Wan, L. & Dreyfuss, G. RNA, and Disease *Cell*. 2009 Feb 20;136(4):777-93. doi: 10.1016/j.cell.2009.02.011.
12. Motorin, Y. & Helm, M. RNA nucleotide methylation. *Wiley Interdiscip. Rev. RNA* **2,** 611–631 (2011).
13. Michelson, A. M. & Todd, A. R. Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3′: 5′-internucleotidic linkage. *J. Chem. Soc.* **0,** 2632–2638 (1955).
14. Agarwal, K. L., Büchi, H., Caruthers, M. H., Gupta, N., Khorana, H. G., Kleppe, K., Kumar, A., Ohtsuka, E., Rajbhandary, U. L., Van De Sande, J. H., Sgaramella, V., Weber, H.Yamada, T.. Total Synthesis of the Gene for an Alanine Transfer Ribonucleic Acid from Yeast. *Nature* **227,** 27–34 (1970).
15. McBride, L. J. & Caruthers, M. H. An investigation of several deoxynucleoside phosphoramidites useful for synthesizing deoxyoligonucleotides. *Tetrahedron Lett.* **24,** 245–248 (1983).
16. Verma, S. & Eckstein, F. MODIFIED OLIGONUCLEOTIDES: Synthesis and Strategy for Users. *Annu. Rev. Biochem.* **67,** 99–134 (1998).
17. Reese, C. B. The chemical synthesis of oligo- and poly-nucleotides by the phosphotriester approach. *Tetrahedron* **34,** 3143–3179 (1978).
18. Caruthers, M. H. The chemical synthesis of DNA/RNA: our gift to science. *J. Biol. Chem.* **288,** 1420–7 (2013).
19. Ogilvie, K. K., Beaucage, S. L., Schifman, A. L., Theriault, N. Y. & Sadana, K. L. The synthesis of oligoribonucleotides. II. The use of silyl protecting groups in nucleoside and nucleotide chemistry. VII. *Can. J. Chem.* **56,** 2768–2780 (1978).
20. Agrawal, S. *Protocols for oligonucleotides and analogs : synthesis and properties*. (Humana Press, 1993).
21. Sonveaux, E. in 1–71 (Humana Press, 1994). doi:10.1007/978-1-59259-513-6_1
22. Bellon, L. in *Current Protocols in Nucleic Acid Chemistry* 1:3.6.1-3.6.13 (John Wiley & Sons, Inc., 2000). doi:10.1002/0471142700.nc0306s01
23. Brown, Tom and Brown, T. (Jr). Solid-phase oligonucleotide synthesis. Available at: http://www.atdbio.com/content/17/Solid-phase-oligonucleotide-synthesis. (Accessed: 28th August 2017)
24. Brown, Tom and Brown, T. Jr. RNA oligonucleotide synthesis. Available at: http://www.atdbio.com/content/22/RNA-oligonucleotide-synthesis. (Accessed: 28th August 2017)
25. Zhang, J., Chang, H. & Kane, R. Reductive Monoalkylation of Aromatic Amines via Amidine Intermediates. *Synlett* (2001).
26. Wuts, P. G. M. & Greene, T. W. *Greene's Protective Groups in Organic Synthesis*. (John Wiley & Sons, Inc., 2006). doi:10.1002/0470053488
27. lv JONES ANI ROLAND Roi-trxs, J. K. Purine Nucleosides. 111. Methylation Studies of Certain Naturally Occurring Purine Nucleosides1.
28. Ghanty, U., Fostvedt, E., Valenzuela, R., Beal, P. A. & Burrows, C. J. Promiscuous 8‑Alkoxyadenosines in the Guide Strand of a SiRNA: Modulation of Silencing Efficacy and Off-Pathway Protein Binding. *J. Am. Chem. Soc.*, 2012, 134 (42), pp 17643–17652 doi:10.1021/ja307102g
29. Aritomo, K., Wada, T. & Sekine, M. Alkylation of 6-N-acylated adenosine derivatives by the use of phase transfer catalysis. *J. Chem. Soc. Perkin Trans. 1* **0,** 1837 (1995).
30. Bryan E., Ptak, R., Ward, P., Takemoto, J. Y., Bensaci, M., Chang, C. Wei T. Application of the Synthetic Aminosugars for Glycodiversification: Synthesis and Antimicrobial Studies of Pyranmycin. (2004). doi:10.1021/JO035290R
31. Smellie, I. A. S. & Paton, R. M. Synthesis of □-D-ribofuranosylbenzazoles using

nitrile oxide addition chemistry. *Gen. Pap. Ark.* 49–61 (2012).
32. Tanaka, H. & Ogasawara, K. *Utilization of lithium triethylborohydride as a selective N-acyl deprotecting agent*. *Tetrahedron Letters* **43,** (2002).
33. Olgivie, K. K. & Entwistle, D. W. Isomerization of tert-butyldimethylsilyl protecting groups in ribonucleosides. *Carbohydr. Res.* **89,** 203–210 (1981).
34. Jones, S. S. & Reese, C. B. Migration of t-butyldimethylsilyl protecting groups. *J. Chem. Soc. Perkin Trans. 1***0,** 2762 (1979).
35. Hobartner, C., Kreutz, C., Flecker, E., Ottenschläger, E., Pils, W., Grubmayr, K., Micura, R. The Synthesis of 2'-O-[(Triisopropylsilyl)oxy] methyl (TOM) Phosphoramidites of Methylated Ribonucleosides (m1G, m2G, m22G, m1I, m3U, m4C, m6A, m62A) for Use in Automated RNA Solid-Phase Synthesis Claudia. *Monatshefte fur Chemie / Chem. Mon.* **134,** 851–873 (2003).
36. Jeong, H. S., Hayashi, G. & Okamoto, A. Diazirine Photocrosslinking Recruits Activated FTO Demethylase Complexes for Specific N^6 -methyladenosine Recognition. *ACS Chem. Biol.* **10,** 1450–1455 (2015).
37. Murakami, M., Inada, T., Hayashibara, H., Fujii, T Purines. LIV. Intramolecular Cyclization of 9-Ethyl-1-(2-hydroxyethyl)adenine Caused by Nucleophiles: Formation of N6, 1-Ethanoadenine Derivatives. *Chem. Pharm. Bull. (Tokyo).* **41,** 453–457 (1993).
38. Jean-Paul Lellouche, and & Koeller, S. The Particular Sensitivity of Silyl Ethers of d-Glucal toward Two Vilsmeier−Haack Reagents POCl3·DMF and (CF3SO2)2O·DMF. Their Unique and Selective Conversion to the Corresponding C(6)-O-Formates. *J. Org. Chem.*, 2001, 66 (3), pp 693–696 (2001). doi:10.1021/JO005540T
39. and, S. B. & Lakshman, M. K. O6-(Benzotriazol-1-yl)inosine Derivatives: Easily Synthesised, Reactive Nucleosides. *J. Am. Chem. Soc.*, 2007, 129 (4), pp 782–789 (2007). doi:10.1021/JA064682N
40. Wan, Z.-K., Binnun, E., Wilson, D. P. & Lee, J. A Highly Facile and Efficient One-Step Synthesis of N 6 -Adenosine and N6 -2′-Deoxyadenosine Derivatives. *Org Lett.* 2005 Dec 22;7(26):5877-80 doi:10.1021/ol052424+
41. Zhao-Kui W., Wacharasindhu, S., Levins, C. G., Lin, M., Tabei, K., Mansour, T.S.. The Scope and Mechanism of Phosphonium-Mediated SNAr Reactions in Heterocyclic Amides and Ureas. (2007). doi:10.1021/JO7020373
42. Li, H.-Y., Li & Hui-Yin. in *Encyclopedia of Reagents for Organic Synthesis* (John Wiley & Sons, Ltd, 2001). doi:10.1002/047084289X.rt016
43. Pirrung, M. C., Shuey, S. W., Lever, D. C. & Fallon, L. A convenient procedure for the deprotection of silylated nucleosides and nucleotides using triethylamine trihydrofluoride. *Bioorg. Med. Chem. Lett.* **4,** 1345–1346 (1994).
44. Pirrung, M. C., Fallon, L., Lever, D. C. & Shuey, S. W. Inverse phosphotriester DNA synthesis using photochemically-removable dimethoxybenzoin phosphate protecting groups. *J. Org. Chem.* **61,** 2129–2136 (1996).
45. Riml, C., Lusser, A., Ennifar, E. & Micura, R. Synthesis, Thermodynamic Properties, and Crystal Structure of RNA Oligonucleotides Containing 5‑Hydroxymethylcytosine. *J. Org. Chem.*, 2017, 82 (15), pp 7939–7945. doi:10.1021/acs.joc.7b01171
46. Matsuda, A., Inada, M., Nara, H., Ohtsuka, E. & Ono, A. Nucleosides and nucleotides. 126. Incorporation of a mutagenic nucleoside, 5-formyl-2′-deoxyuridine, into an oligodeoxyribonucleotide. *Bioorg. Med. Chem. Lett.* **3,** 2751–2754 (1993).
47. Bhatt, M. V. & Perumal, P. T. Facile conversion of electron rich benzylic hydrocarbons to carbonyl compounds by peroxydisulphate and copper ions. *Tetrahedron Lett.* **22,** 2605–2608 (1981).
48. Krim, J., Grünewald, C., Taourirte, M. & Engels, J. W. Efficient microwave-assisted synthesis, antibacterial activity and high fluorescence of 5 benzimidazolyl-2′-

deoxyuridines. *Bioorg. Med. Chem.* **20,** 480–486 (2012).
49. Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y. G., He, C. N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. (2011). doi:10.1038/nCHeMBIO.687
50. Maruenda, H., Chenna, A., Liem, L.-K. & Singer, B. Synthesis of 1,N 6 -Ethano-2′-deoxyadenosine, a Metabolic Product of 1,3-Bis(2-chloroethyl)nitrosourea, and Its Incorporation into Oligomeric DNA.*J. Org. Chem.*1998, 63, 4385-4389
51. Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., Shi, Y. G. TET1 is a DNA binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res.* **20,** 1390–1393 (2010).
52. Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C. Zhang, Y. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science.* **333,** 1300–1303 (2011).

Chapter 4: Studies on the reaction of formaldehyde with nucleotides

4.1 Introduction

4.1.1 Physical Properties, Structural Information and Industrial Uses

Formaldehyde is a flammable, colourless gas with a molecular mass of 30.03 and a characteristic pungent and unpleasant odour. Formaldehyde is also the simplest and most reactive aldehyde present ubiquitously in nature. The presence of an electrophilic carbon centre makes formaldehyde highly susceptible to attack by nucleophiles (R-NH_2, R-SH, H_2O etc.) resulting in the formation of a C-X bond.

Formaldehyde is used in the manufacture of resins, particle board and plywood, textiles, leather goods, paper, pharmaceuticals and even cosmetics.[1] In the human body formaldehyde is required for the biosynthesis of purines, thymidine and certain amino acids.[2–4]

4.1.2 Sources of Formaldehyde

The sources of biologically relevant formaldehyde can be classified into exogenous and endogenous sources.

4.1.2.1 Exogenous sources

Formaldehyde produced by the oxidation of hydrocarbons (mostly methane) in the troposphere (estimated to be 400 billion kilograms per year), a further 4 billion kilograms produced industrially each year.[5] Formaldehyde is also a component of automobile engine exhaust fumes, tobacco smoke, e-cigarettes and biologically active molecules such as aspartame.[6,7] Cigarette smoke contains 12-106 µg of formaldehyde per cigarette.[8]

4.1.2.2 Endogenous sources

Possibly the most direct formaldehyde source is the direct consumption of methanol, which is oxidised to HCHO within cells. Formaldehyde is produced in animals and other organisms as a consequence of various biological reactions including demethylation reactions (*N*-methylated histone/DNA/ RNA),[9–12] lipid peroxidation,[13] and cytoplasmic reactions, which include P450-catalysed (*N*-/*S*-/*O*-)methyl group demethylations.[13]

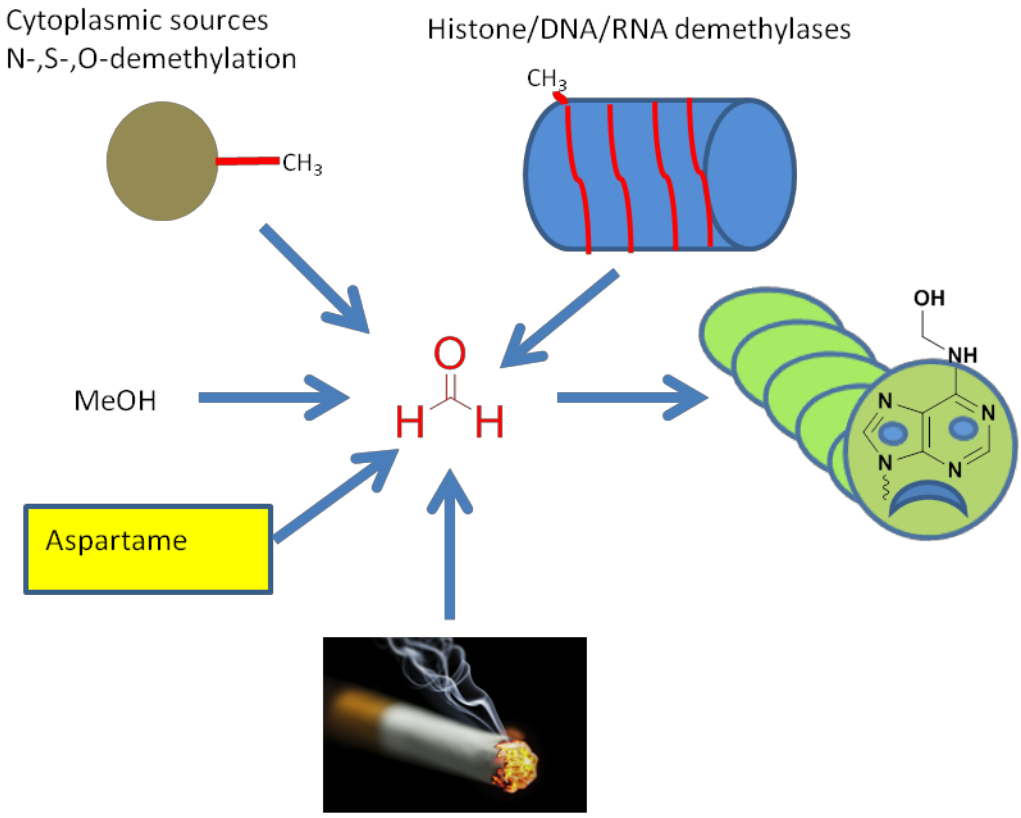


Figure 4.1 Schematic representation of endogenous and exogenous sources of formaldehyde. Formaldehyde is produced as a co-product during demethylation of various N-/S-/O- methylated substrates in cells. The produced formaldehyde can react with various nucleophiles including amino acids residues, nucleotides etc.

4.1.3 Crosslinking through formaldehyde

Since biologically endogenous formaldehyde is produced in close proximity to eukaryotic chromosomes, it has the potential to react with biologically relevant nucleophiles such as

amino groups present in amino acids, nucleotides, thiol groups (e.g.cysteine and glutathione), and alcohols. Thus, formaldehyde has the potential to form adducts with proteins,[14,15] DNA,[15–24] and RNA [25], potentially leading to DNA-protein,[26–28] protein-protein,[8] DNA-DNA,[25,29,30] and DNA-glutathione crosslinks (and many others).[22] However, in most instances formaldehyde is considered to react reversibly with proteins and DNA, unless crosslinks are formed.

4.1.4 Toxicity of formaldehyde

Principally due to its potential to form DNA-protein crosslinks (DPCs), formaldehyde is considered to be genotoxic,[31–33] mutagenic[28,33–35] and carcinogenic[36–38]. However, the precise mode of action of formaldehyde toxicity is still unclear and may involve multiple processes.

The concentration of formaldehyde in normal human blood is reported to about 0.1 mM,[39] while its concentration in healthy human brain is reported to be around 0.2 mM and 0.4 mM in the hippocampus and cortex, respectively.[40] The concentration of formaldehyde in the brain has been reported to be elevated in old age and in patients with neuro-degenerative conditions such as Alzheimer's disease.[40,41]

Endogenous formaldehyde is established as a stem cell genotoxin and metabolic carcinogen.[42] Marked differences are reported in the levels of formaldehyde–DNA adducts in leukocytes of smokers and non-smokers.[43]

Formaldehyde detoxification is conducted by the ADH (Alcohol Dehydrogenase) family of enzymes,[35,44–46] thereby forming a steady state balance of formaldehyde concentration.

4.2 Previous studies on formaldehyde

Formaldehyde reacts with the nucleophilic amino groups on both exocyclic amino and endocyclic imino groups of nitrogenous bases of nucleotides (DNA/RNA) and nucleophilic

hydroxyl, amino and thiol groups in peptides to form adducts of the type –XCH_2OH (X= *S*, *N*, *O*) of varying stability. Several studies on the non-enzymatic reactions of formaldehyde and nucleotides have been reported.[19–21,24,25,27,29,30,47–52]

The first addition is assumed to be rapid and reversible while the formation of a methylene bridge is assumed to be slow but non-reversible.[53]

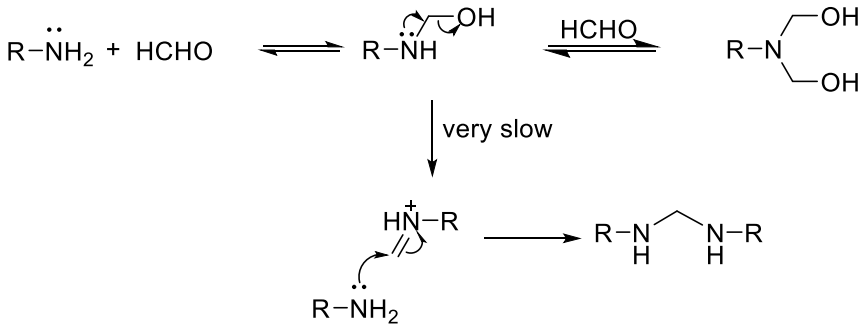


Figure 4.1 Schematic representation of the reactions of formaldehyde with amino groups of peptides/nucleotides. R represents the nucleic acid or amino acid residue. Formaldehyde reacts with amines to form methylol (or hemiaminal) products; the methylol products slowly convert to the corresponding Schiff base which can then react with another nucleophile to form a methylene bridge.

Fraenkel-Conrat[51] reported, using UV-spectrometry, that the reaction of formaldehyde and nucleotides with exocyclic amines results in Schiff base products. In further studies, it was suggested that bases containing only endocyclic imino nitrogens do not react with formaldehyde, as no spectral changes were observed using UV-spectroscopy.

The most detailed study of the non-enzymatic formaldehyde-nucleotide reaction is reported in a series of four papers by Von Hippel and McGhee.[17–20] The first of these studies focused on the reaction of formaldehyde with nucleotides/nucleosides with exocyclic amines and proposed formation of methylol products, rather than the previously reported Schiff bases.[51] Subsequent studies by other research groups support these findings.[20,47,49–51]

In the same study by Hippel and McGhee,[20] the formation of a diadduct with dAMP in the presence of excess formaldehyde, was established for the first time. However, the adduct was not thoroughly characterised. Moreover, no attempts were made to elucidate/validate the

formation of other formaldehyde-induced diadducts with dGMP and dCMP. Furthermore, crosslinks between nucleosides/nucleotides via formaldehyde methylene bridges were not observed in aqueous solution.

In another study by F. A. Beland *et al.*,[54] the presence of the diadduct product with dAMP was not observed, even after using a 67 fold excess of formaldehyde in the reaction. The reaction mixture was analysed using UV-Vis spectroscopy and only two peaks were observed, which were assumed to correspond to the unmodified nucleoside and the mono-hydroxymethyl adduct. Further, they purified the mono-adduct and conducted NMR characterisation. It is possible that the diadduct was not observed in this study due to the overlap of its UV absorbance with that of the mono-adduct; it is plausible that the diadduct degraded during purification.

The non-enzymatic reaction of endocyclic imino groups with formaldehyde has remained controversial.[19,48–50] One possible reason is the smaller spectral difference observed in UV spectroscopy between the endocyclic imine and the corresponding imine-formaldehyde adduct and the other could be a smaller equilibrium constant observed for non-enzymatic reaction of formaldehyde and imino groups as compared to exocyclic amines.[19,20] In the second paper by Hippel and McGheel,[19] studies were conducted on the reaction of formaldehyde and nucleotides with endocyclic imines. They established, for the first time, the formation of endocyclic imine-formaldehyde adducts via UV-spectroscopy and measured the relevant equilibrium constant.

In a study by Chaw, *et al.*[52] calf thymus DNA was incubated with formaldehyde for 40 days and was analysed following enzymatic digestion. The study revealed the formation of five crosslinked adducts i.e dA-dA, dG-dG, dA-dC, dA-dG, dG-dC linked through their exocyclic amine via methylene bridges.

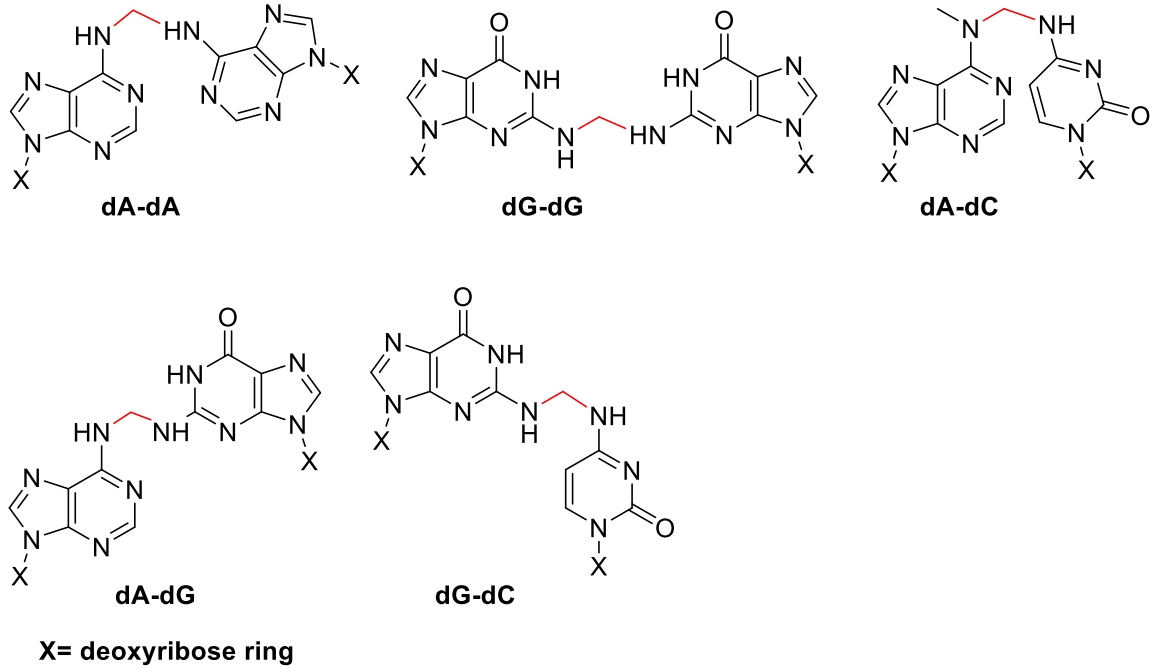


Figure 4.3 Schematic representation of the adducts observed.[25] Methylene bonds derived from formaldehyde are highlighted in red.

Formaldehyde was also shown to crosslink dA to dA in certain AT-rich regions.[29] In this study, DNA duplexes were reacted with 25 mM formaldehyde in 50 mM phosphate buffer for 9 days which resulted in several crosslinks. The ^{1}H-NMR spectrum of the purified dA-methylene-dA adduct in D_2O had a resonance at δ_H 5.4 ppm, which was assigned to the protons of the methylene bridge. Thus, it was concluded that A-T rich regions in DNA are more susceptible to attack by formaldehyde leading to the formation of dA-dA crosslinks. The reason for the observed preference could be the weaker intra-strand H-bonding between A-T as compared to G-C regions, which allowed for local opening of the helix.

In a recent study conducted by Lu, K. *et al.*, formaldehyde-induced crosslinks of lysine (Lys), cysteine (Cys), histidine (His) and tryptophan (Trp) with deoxyguanosine (dG), deoxyadenosine (dA) and deoxycytidine (dC) were reported.[21] However, no crosslinks were observed with thymidine (T), likely because the formation of Schiff bases is not possible with an endocyclic imine base. dG was observed to form a methylene bridge with Lys at the fastest rate, followed by dG and Cys. The Lys-dG adduct could be of interest as Lys residues

are extensively present in the tail of histones and play important roles in transcriptional regulation.

However, with the exception of the work by Lu, K. *et al.*, formaldehyde-induced adducts were mainly characterised by UV-vis and mass spectrometry, lacking thorough and extensive studies using NMR as an analytical tool. Further, the evaluation of adduct formation rates, stabilities, pH and temperature dependence are incomplete. There are also no reports on non-enzymatic reactions with non-canonical bases (e.g. pseudouridine, inosine).

4.3 Objectives of work described in this chapter

The specific objectives of the work described in this chapter were as follows:

- To investigate the non-enzymatic reactions of formaldehyde with nucleotides using NMR;
- To characterise reported formaldehyde-induced nucleotide adducts;
- To investigate the effect of changes in pH on the rates of reaction;
- To investigate the effect of temperature changes on the reaction rates;
- To compare the reaction rates of adenosine mono-, di- and triphosphates with formaldehyde;
- To compare the reactions of nucleotides and their 2'-deoxy-homologues with formaldehyde;
- To define the reactivity of 5-methyldeoxycytidine (m^5dC) and 5-hyroxymethyldeoxycytidine (hm^5dC) with formaldehyde, and compare their reactions with those of deoxycytidine/cytidine (dC/C);
- To define the reactivity of N^6-methyladenosine (m^6A) and N^1-methyladenosine (m^1A) with formaldehyde, and compare to reactions with adenosine (A);

- To compare the reactivity of O^6-methylguanosine (m^6G) and guanosine (G) with formaldehyde;
- To investigate the stability of the formaldehyde-nucleotide adducts by diluting the formaldehyde adducts with water, then monitoring their decomposition.

4.4 Results and Discussion

4.4.1 NMR Analysis of formaldehyde solution

Initially, NMR studies on aqueous formaldehyde were carried out to confirm purity, and to determine the relative prevalence of aldehyde, hydrated and oligomeric forms, which can all exist in aqueous solution. An aqueous solution stock of formaldehyde (800 mM) was prepared by cracking paraformaldehyde in D_2O (Figure 4.4A); after addition of D_2O, the colloidal solution was heated until the solution became clear. The solution was then left for 12 hours at room temperature to reach equilibrium before being further diluted with D_2O to 320 mM and subjected to ^{1}H-NMR analysis (Figure 4.4B).

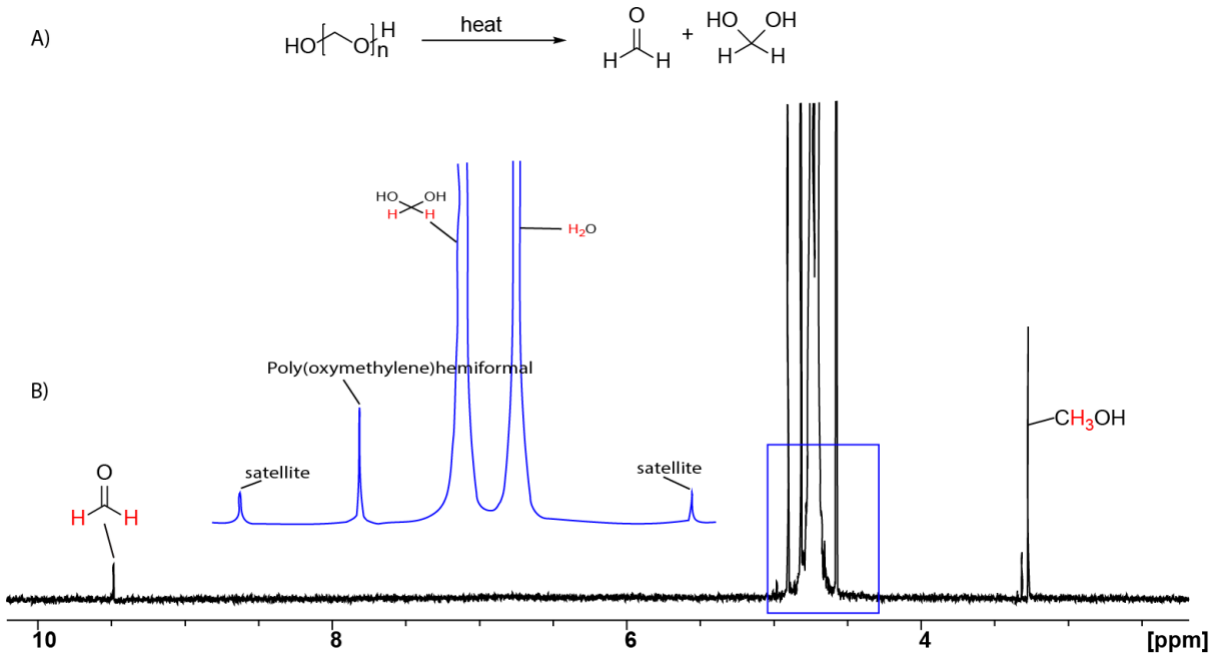


Figure 4.4 A) Schematic representation of cracking of paraformaldehyde to formaldehyde. B) ^{1}H-NMR (600 MHz) spectrum of a 128 mM solution of formaldehyde in D_2O (Probe temperature, 25°C). Protons corresponding to signals are highlighted in red. The region in the blue box is zoomed to 10 times.

Seven major resonances were observed in the ^{1}H NMR spectrum, indicating the presence of species other than formaldehyde/hydrated formaldehyde in the solution. The resonance at δ_H 4.73 ppm corresponds to the hydrated form of formaldehyde. It should be noted that most of the formaldehyde exists in hydrated form in D_2O rather than in the carbonyl form; the methylene protons of the carbonyl resonate at δ_H 9.47 ppm. The resonances at δ_H 8.37 ppm and δ_H 3.27 ppm correspond to the protons attached to the carbonyl of formate and methyl group of methanol respectively, suggesting partial disproportionation (Cannizzaro reaction, Figure 4.5). The concentration of methanol was observed to be higher than that of formate, implying that some methanol was already present in the purchased sample. The resonance at δ_H 4.97 ppm corresponds to oligomeric poly(oxymethylene)hemiformals.[55]

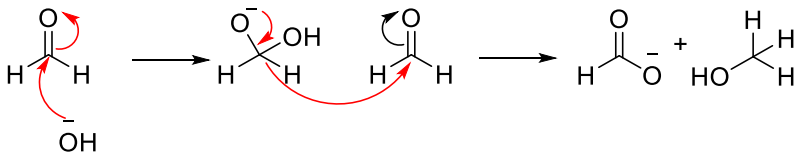
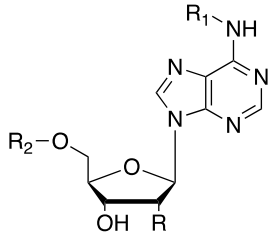


Figure 4.5 Schematic representation of base catalysed Cannizzaro reaction leading to the equimolar formation of methanol and formate.

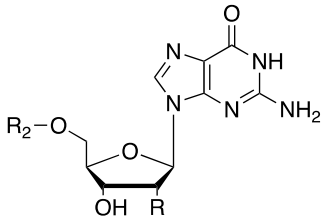
4.4.2 Optimisation of the reaction conditions of formaldehyde with nucleotides

Initially, a 40 mM solution of deoxyadenosine monophosphate (dAMP) in 50 mM sodium phosphate buffer was adjusted to pD 7.5; 16 µl of this solution was then mixed with an 8-fold excess of formaldehyde. However, no reaction was observed, which was likely due to sequestration of formaldehyde by the phosphate buffer. Therefore, reactions were then carried out in D_2O to avoid reactions with buffer, and also to minimise signal overlap between *N*-hydroxymethyl resonances, which likely resonate at roughly δ_H 5.0 ppm, and the residual water resonance (at δ_H 4.7 ppm).

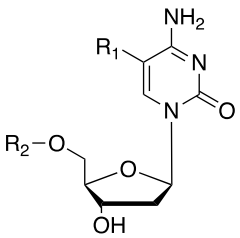
The reaction of formaldehyde was investigated with sixteen different nucleosides/ nucleotides including the canonical bases (adenosine, guanosine, cytidine, thymidine and uridine) and non-canonical bases (N^6-methyladenosine (m^6A), O^6-methylguanosine (m^6G), hm^5dC, m^5dC, m^1A, pseudouridine and inosine) (Figure 4.5.1). For the canonical bases, studies were conducted at three different pD values (pD 6, pD 7.5 and pD 9), while for the non-canonical bases, studies were conducted at pD 7.5 and their reactions with formaldehyde were compared with structurally similar canonical nucleoside/nucleotides.



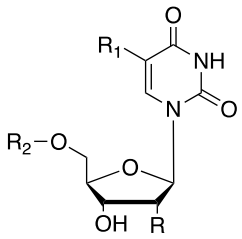
R, R_1, R_2, = H; deoxyadenosine
R_1, R_2 = H, R = OH; adenosine
R_2= PO_3^{2-}, R_1, R, = H; deoxyadenosinemonophosphate
R_2= H, R=OH, R_1= CH_3 ; N^6-methyladenosine



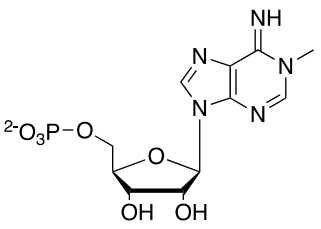
R, R_2 = H, deoxyguanosine
R=H, R_2 = PO_3^{2-} ;deoxyguanosinemonophosphate
R=OH, R_2=H; guanosine



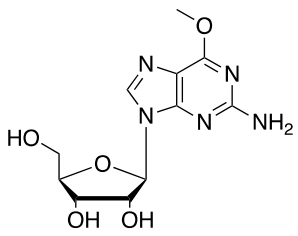
R_1, R_2 =H; deoxycytidine
R_1=H, R_2= PO_3^{2-}; deoxycytidinemonophosphate
R_1= CH_3, R_2=H; 5-methyldeoxycytidine
R_1=CH_2OH, R_2=H; 5-hydroxymethyldeoxycytidine



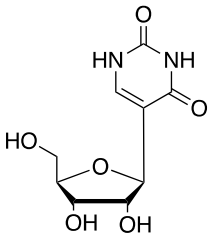
R=H, R_1= CH_3, R_2 = PO_3^{2-} ; Thymidinemonophosphate
R=OH, R_1=H, R_2= PO_3^{2-} ; Uridinemonophosphate



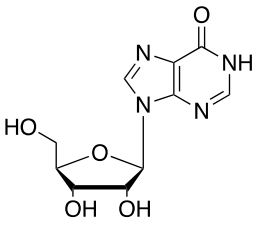
N^1-methyladenosinemonophosphate



O^6-methylguanosine



pseudouridine



inosine

Figure 4.5.1 Chemical structures of the nucleosides/ucleotides investigated in this study.

4.4.3 Studies on the reaction of nucleotide/nucleoside with formaldehyde

It was envisaged that the reaction of formaldehyde with nucleotides/nucleosides would result in the formation of structurally related products. Therefore, resonances corresponding to the ribo-sugar and base of each hydroxymethyl adduct were expected to have similar chemical shifts to those of the corresponding unreacted nucleosides/nucleotides. Therefore, initial studies focussed on assigning the NMR spectra of the unreacted nucleosides/nucleotides so as to simplify/enable assignment of the related adducts.

Attention then shifted to studies on mixtures of nucleotides/nucleosides with formaldehyde. Initially, mixtures were prepared and left for 12 hours to allow significant formation of 'stable' adducts, which were then subjected to NMR characterisation. Each mixture was prepared by adding an excess of formaldehyde (53-fold) to each nucleotide/nucleoside solution in D_2O (final concentration of nucleotide/nucleoside was 2.2 mM); each reaction mixture was then allowed to reach equilibrium/stand overnight in 5mm NMR tubes before analysis using 1D (^1H, ^{13}C) and 2D NMR experiments (HSQC, HMBC).

To investigate the adduct formation rates, ^1H- NMR spectra over the first 8-10 hours after mixing were also recorded. Each nucleoside/nucleotide under study (2.4 mM final concentration) was reacted with a 53-fold excess of formaldehyde and the reaction mixture was immediately transferred to a 3 mm NMR tube. The first ^1H-NMR spectrum was recorded within 5-7 min following initial mixing. This time was required to lock, shim, tune and calibrate the 90-degree pulse for the solvent. For the first hour after mixing, ^1H-NMR spectra were recorded after every 2 min; after this, ^1H NMR spectra were recorded every 4 min for the next two hours and then every 20 min for the rest of the analysis.

The concentrations of each adduct were calculated by integrating their characteristic resonances and normalising the intensity values relative to the internal standard 3-(trimethylsilyl)-2,2', 3,3'-tetradeuteropropionic acid (TSP, δ_H 0.0 ppm) with a molecular weight of 172.27 Da.

For the canonical nucleotides (UMP, TMP, dAMP, dCMP and dGMP), time courses were also carried out at pD 6 and pD 9 to compare the pD dependence of the adduct formation rates.

4.4.4 Studies on the reaction of formaldehyde with Deoxyadenosinemonophosphate (dAMP)

4.4.4.1 ^{1}H-NMR analysis on the reaction of formaldehyde with dAMP

Initially, a ^{1}H NMR spectrum of 40 mM solution of dAMP in D_2O was recorded and assigned. Seven distinct ^{1}H resonances were observed, corresponding to all non-exchangeable protons. The resonances at δ_H 8.41 ppm and δ_H 8.08 ppm correspond to the protons attached to C-8 and C-2 (marked as 5 and 1 in Figure 4.6.A) of the purine ring of dAMP. The resonances at δ_H 6.39 ppm, δ_H 4.6 ppm, δ_H 4.18 ppm, δ_H 3.86 ppm, δ_H 2.76 ppm and δ_H 2.5 ppm correspond to the protons attached to C-1, C-3, C-4, C-5, C-2 and C-2 (marked as 6, 8, 9, 10, 7 and 7, respectively in Figure 4.6.A) of the deoxyribose ring of dAMP. It should be noted that the resonance corresponding to the proton attached to C-3 (marked as 8 in Figure 4.6.A) of the deoxyribose ring resonates close to the residual water signal.

Excess formaldehyde was added to the above solution and ^{1}H NMR spectrum was recorded after 12 hours following initial mixing (Figure 4.6.B). The proton and the carbon assignments of the peaks are given in Table 4.4.4(1-3).

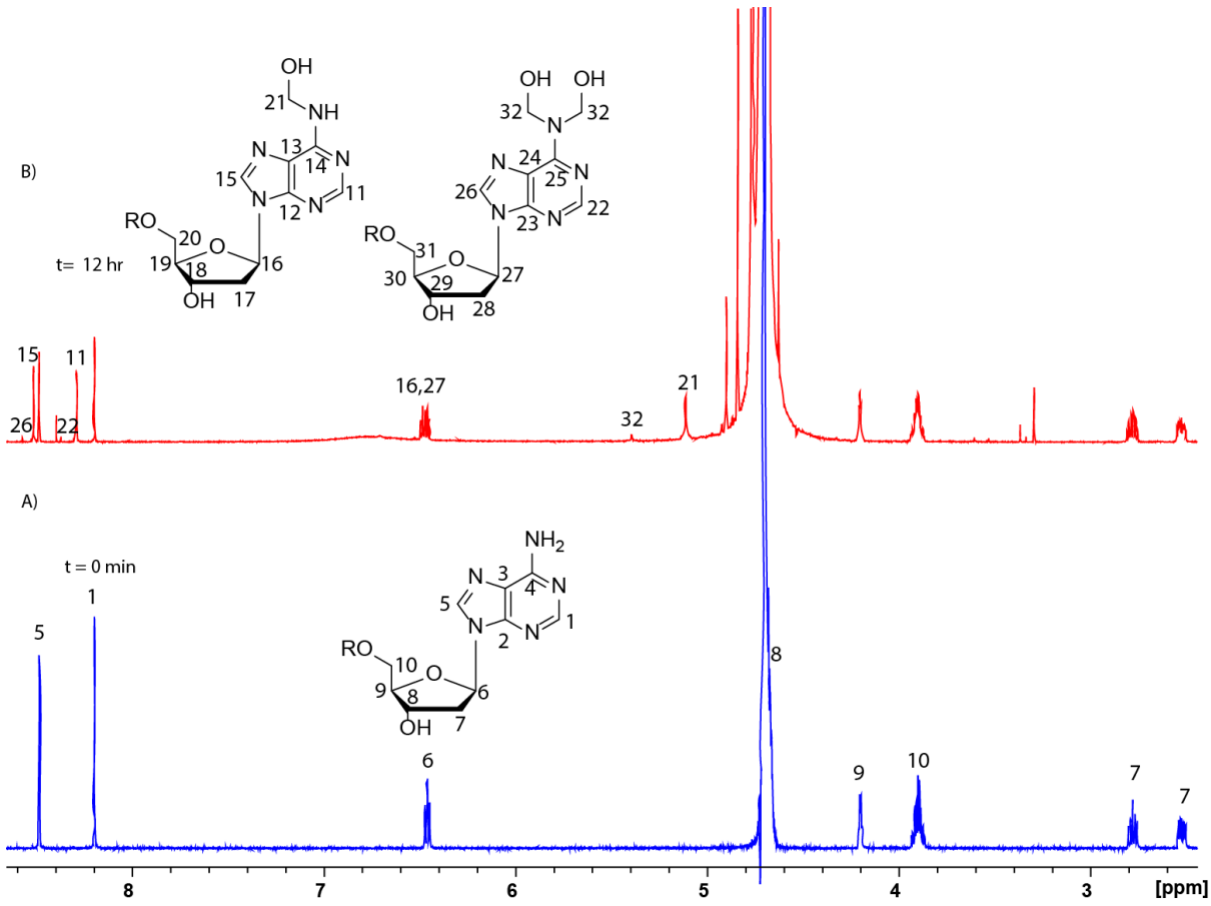


Figure 4.6 A) ^{1}H-NMR (500 MHz) spectrum of a 40 mM dAMP at 27°C, pD 7.5 in D_2O. The ^{1}H assignments of dAMP are indicated. B) ^{1}H-NMR (500MHz) spectrum for the reaction mixture of 40 mM dAMP and excess formaldehyde after 12 hr following initial mixing at 27°C, pD-7.5. ^{1}H assignments of the newly developed resonances are indicated.

New resonances were observed to develop at δ_H 8.57 ppm, δ_H 8.51 ppm, δ_H 8.37 ppm, δ_H 8.29 ppm, δ_H 6.45 ppm, δ_H 6.40 ppm, δ_H 5.40 ppm and δ_H 5.1 ppm. Investigation of the recorded ^{1}H-NMR spectrum revealed the formation of two closely related products. The resonance at δ_H 5.1 ppm (marked as 21 in Figure 4.6.B) correlates with the carbon at δ_C 64.0 ppm in 2D HSQC and shows correlation with the carbon at δ_C 154.3 ppm (marked as 14 in Figure 4.6.B) in 2D HMBC (Figure 4.7). Furthermore, the intensity of the resonance at δ_H 5.1 ppm was found to be approximately twice the intensity of the resonance at δ_H 8.29 ppm (marked as 11 in Figure 4.6.B). Thus, the resonance at δ_H 5.1 ppm (marked as 21 in Figure 4.6.B) was assigned to the methylene protons of the *N*-hydroxymethyl group of N^6-hydroxymethyldeoxyadenosinemonophosphate (hm^{6}dAMP). The resonances at δ_H 8.51 ppm, δ_H 8.29 ppm and δ_H 6.40 ppm correspond to the protons attached to C-8 and C-2 of the purine

ring (marked as 15 and 11 in figure 4.6.B) and C-1 of the deoxyribose ring (marked as 16 in Figure 4.6.B) of hm^{6}dAMP.

The resonance at δ_H 5.4 ppm (marked as 32 in Figure 4.6.B) correlates to the carbon at δ_C 72.4 ppm in the 2D HSQC spectrum and shows correlations with its own carbon (δ_C 72.4 ppm) and the carbon at δ_C 153.6 ppm (marked as 25 in Figure 4.6.B), indicating symmetry in the structure (Figure 4.7). Furthermore, the intensity of the resonance at δ_H 5.4 ppm was found to be approximately four times the intensity of the resonance at δ_H 8.37 ppm (marked as 22 in Figure 4.6.B). Consequently, the resonance at δ_H 5.4 ppm was assigned to the methylene protons attached to two *N*-hydroxymethyl groups of N^6,N^6-di(hydroxymethyl)deoxyadenosinemonophosphate (dhm^{6}dAMP). The resonances at δ_H 8.57 ppm, δ_H 8.37 ppm and δ_H 6.45 ppm correspond to the protons attached to the C-8 and C-2 of the purine ring (marked as 26 and 22 in Figure 4.6.B) and C-1 of the deoxyribose ring (marked as 27 in Figure 4.6.B) of dhm^{6}dAMP. dhm^{6}dAMP was not elucidated completely in previous literature.[20]

No other adducts were observed, at least during the first 12 hours of the reaction. By NMR analysis, the conversion of dAMP to hm^{6}dAMP was ~43% and to dhm^{6}dAMP was ~2% after 12 hours.

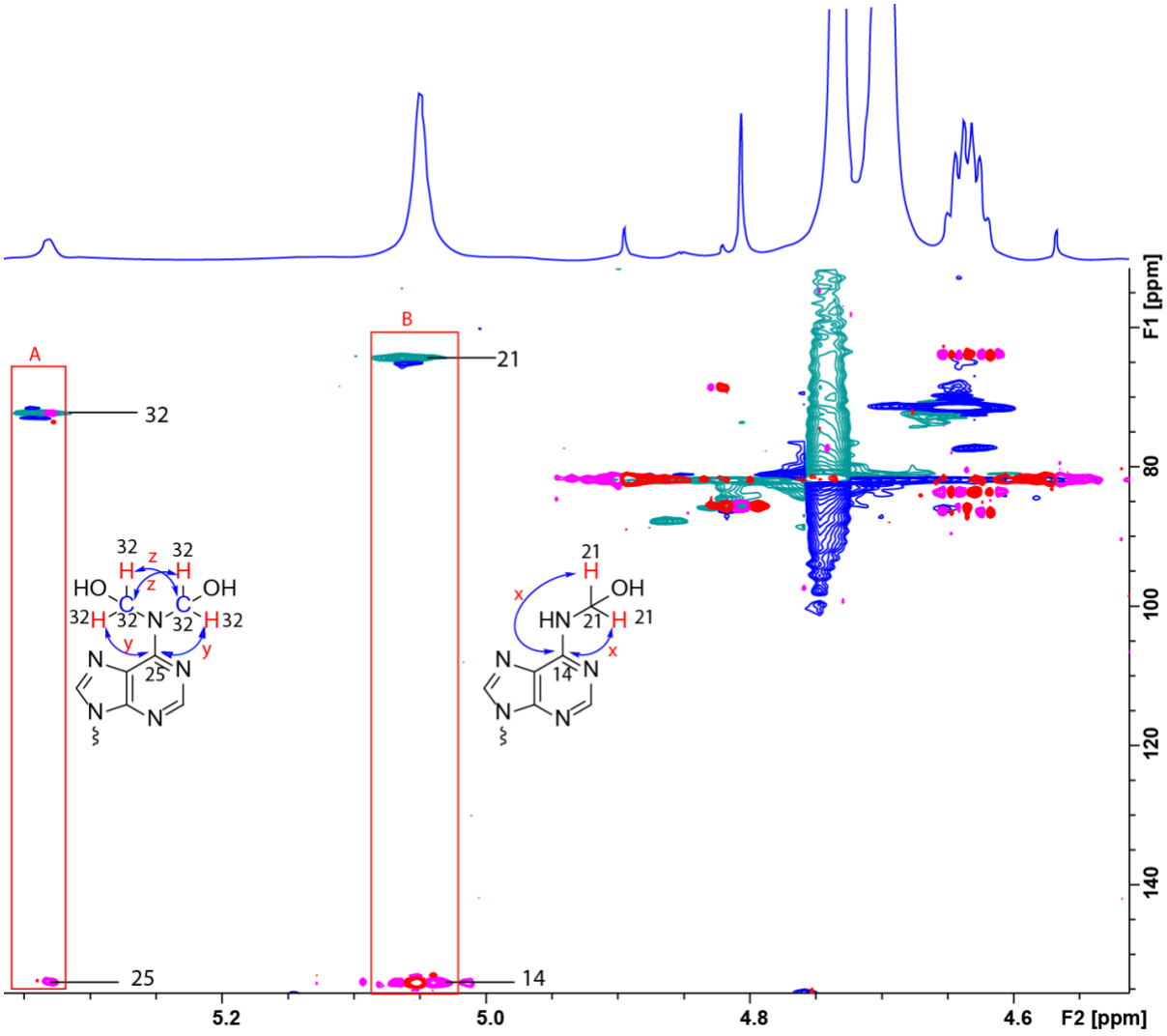


Figure 4.7 Overlaid 2D HSQC (green and blue) and 2D HMBC (red and pink) NMR spectra (500 MHz) of the reaction mixture of 40 mM solution of dAMP and formaldehyde at 25°C. Cross-peaks highlighted in Box A correspond to correlations observed for dhm^{6}dAMP (marked as "y" and "z" in red). Cross-peaks highlighted in Box B correspond to correlations observed for hm^{6}dAMP (marked as "x"in red). Important correlations observed between formaldehyde derived protons (marked in red) and dAMP structure are highlighted with blue curved arrows.

The structures of the adducts were also assessed by analysing the ESI mass spectrum obtained by LC purification. The EI/MS spectrum shows three peaks in the positive ion mode corresponding to m/z: 332.0721 ([dAMP+H]$^+$, calculated: 332.0715), 362.0821 ([hm^{6}dAMP+Na]$^+$, calculated: 362.08600) and 392.0927 ([dhm^{6}dAMP+H]$^+$, calculated: 392.0966).

To investigate the effect of a change in pH on the rates of the reaction, ^{1}H NMR time-courses were recorded at pD (pH+0.4) 6.0, 7.5 and 9.0. The analysis of the time-courses revealed the

formation of hm^{6}dAMP and dhm^{6}dAMP. The resonance at δ_H 5.1 ppm was observed to develop approximately 12 min after initial mixing (i.e. the concentration of hm^{6}dAMP around 12 min was above the detection limit). The resonance at δ_H 5.4 ppm was observed to develop approximately 112 min after initial mixing. The concentration of dhm^{6}dAMP remained consistently low at ~0.4 mM over the time-course. The concentration of hm^{6}dAMP slowly increased with time and reached ~1.1 mM after 600 min following the initial mixing.

^{1}H NMR time-courses were then conducted at pD 6 and pD 9. Analysis of the recorded ^{1}H NMR spectra revealed the formation of hm^{6}dAMP and dhm^{6}dAMP in both cases. The initial rate of the reaction was calculated to be 7.4 x10^{-8} mol dm^{-3} s^{-1}, 6.0 x 10^{-8} mol dm^{-3} s^{-1}, 5.5 x10^{-8} mol dm^{-3} s^{-1} at pD-6.0, 7.5 and 9.0 respectively (Figure 4.9). Thus, the change in pD has a subtle effect on the rates of the reaction. The final concentrations of adducts (hm^{6}dAMP and dhm^{6}dAMP) were observed to be largely independent of the change in pD.

4.4.4.2 Conclusion of the results for the section 4.4.4

In conclusion, ^{1}H-NMR analysis revealed the formation of two closely related adducts when dAMP is reacted with excess formaldehyde. The newly developed resonance at δ_H 5.1 ppm was assigned to the methylene protons of the *N*-hydroxymethyl group of hm^{6}dAMP and the resonance at δ_H 5.4 ppm was assigned to the methylene protons attached to two *N*-hydroxymethyl group of dhm^{6}dAMP (Figure 4.6.B). No cross-linked dAMP product was observed in the reaction mixture during the first 12 hours of reaction, consistent with the previous literature.[20] The NMR results were further supported by MS analysis. The initial rate as calculated from the graph was 6.0 x 10^{-8} mol dm^{-3} s^{-1} at pD 7.5, 27°C. The initial rate of reaction was found to be independent of the change in pD in the range investigated (pD 6-9).

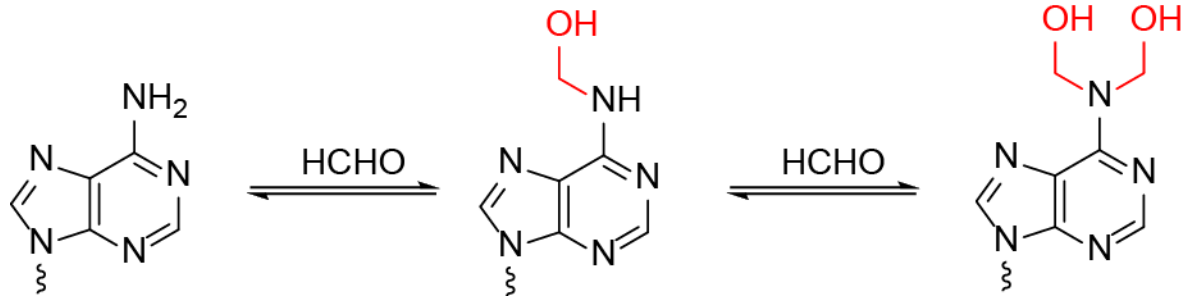


Figure 4.8 Schematic representation of reaction of formaldehyde and dAMP leading to the formation of hm^{6}dAMP and the subsequent formation of dhm^{6}dAMP. The proton/s and carbon/s derived from formaldehyde are highlighted in red.

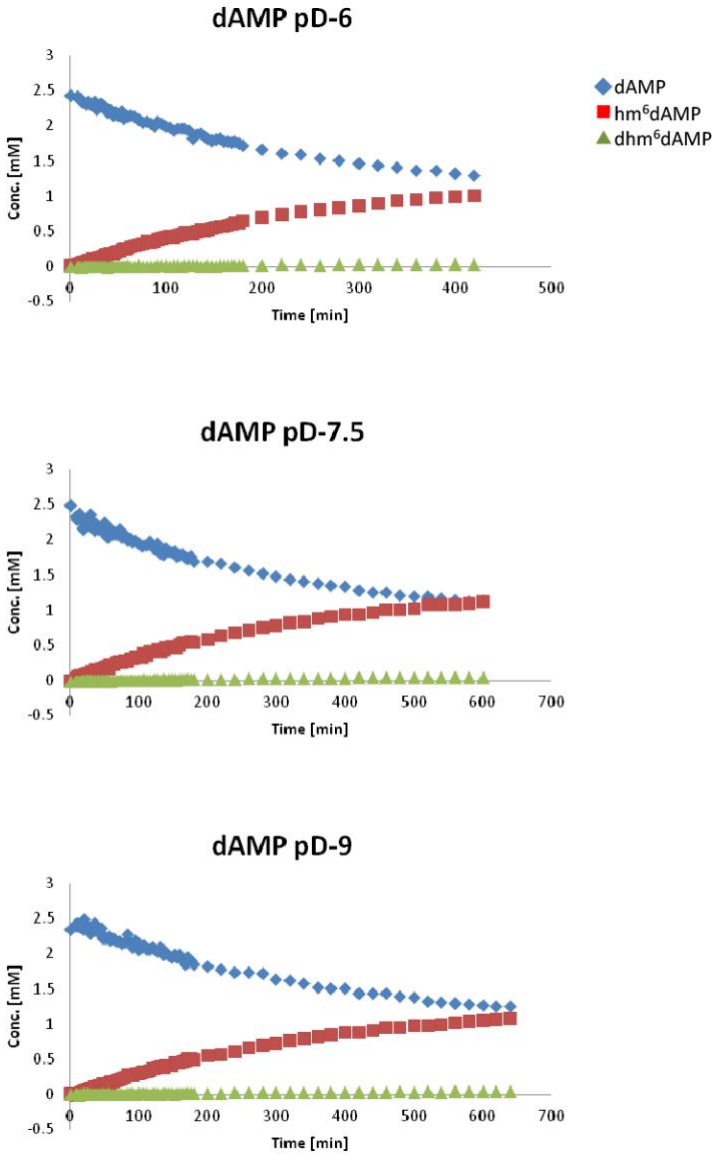


Figure 4.9 Graph showing change in the concentrations of dAMP (blue), hm^{6}dAMP (red) and dhm^{6}dAMP (green) over time at 27°C at pD 6.0 (top); pD 7.5 (middle); pD 9.0 (bottom). The concentrations are obtained by the ^{1}H integration of the peaks corresponding to the resonances of the proton attached to C-2 of the purine ring of dAMP, hm^{6}dAMP and dhm^{6}dAMP (marked as 1, 11 and 22, respectively in Figure 4.6) and normalising the intensity value relative to the internal standard, TSP. ^{1}H-NMR timecourses were recorded for 8-10 hours; every 2 min over the first hour, every 4 min over the next 2 hours and every 20 mins for the next 5-7 hour.

Table 4.4.4.1 Proton and carbon assignments for dAMP

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
1	152.58	8.08	s	-
2	148.5	-	-	-
3	118.5	-	-	-
4	155.3	-	-	-
5	140.02	8.41	s	-
6	83.5	6.39	t	7.0
7	39.0	2.5, 2.76	m, m	-
8	71.3	4.6	dt	6.0, 3.0
9	86.6	4.18	q	4.0
10	63.7	3.86	m	-

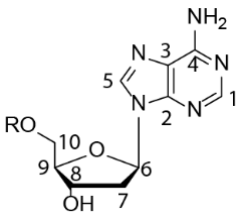


Table 4.4.4.2 Proton and carbon assignment for hm^{6}dAMP

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
11	152.51	8.29	s	-
12	148.5	-	-	-
13	118.9	-	-	-
14	154.3	-	-	-
15	140.33	8.51	s	-
16	83.5	6.40	t	7.0
17	39.0	2.5, 2.76	m, m	-
18	71.2	4.6	dt	6.0, 3.0
19	86.2	4.18	q	4.0
20	63.7	3.86	m	-
21	64.0	5.1	bs	-

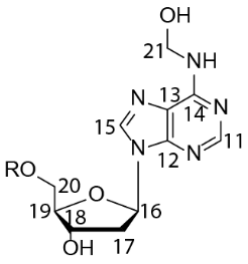
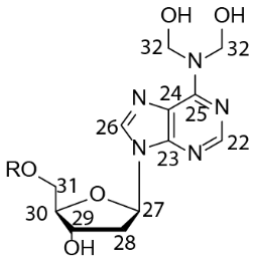


Table 4.4.4.3: Proton and carbon assignment for dhm^{6}dAMP

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
22	152.1	8.37	s	-
23	148.5	-	-	-
24	119.5	-	-	-
25	153.6	-	-	-
26	140.56	8.57	s	-
27	83.5	6.45	t	7.0
28	39.0	2.5, 2.76	m, m	-
29	71.2	4.6	dt	6.0, 3.0
30	86.2	4.18	q	4.0
31	63.7	3.86	m	-
32	72.4	5.40	bs	-



4.4.5 Studies on the reaction of formaldehyde with deoxyguanosinemonophosphate (dGMP)

4.4.5.1 ^{1}H-NMR analysis on the reaction of formaldehyde with dGMP

Initially, the ^{1}H NMR spectrum of a 40 mM solution of dGMP pre-buffered at pD 7.5 was recorded and assigned. Seven distinct ^{1}H resonances were observed. All the non-exchangeable protons were observed. The resonance at δ_H 8.08 ppm corresponds to the proton attached to C-8 (marked as 5 in Figure 4.10.A) of the purine ring of dGMP. The resonances at δ_H 6.25 ppm, δ_H 4.65 ppm, δ_H 4.13 ppm, δ_H 3.83 ppm, δ_H 2.74 ppm and δ_H 2.43 ppm correspond to the protons attached to C-1, C-3, C-4, C-5, C-2 and C-2 (marked as 6, 8, 9, 10, 7 and 7, respectively in Figure 4.10.A) of the deoxyribose ring of dGMP.

To this solution, excess formaldehyde was added and ^{1}H NMR, 2D HSQC, 2D HMBC spectra were recorded on the sample after 12 hr following initial mixing (Figure 4.10.B). The ^{1}H and ^{13}C assignments for dGMP and newly formed adducts are given in Table 4.4.5.(1-3).

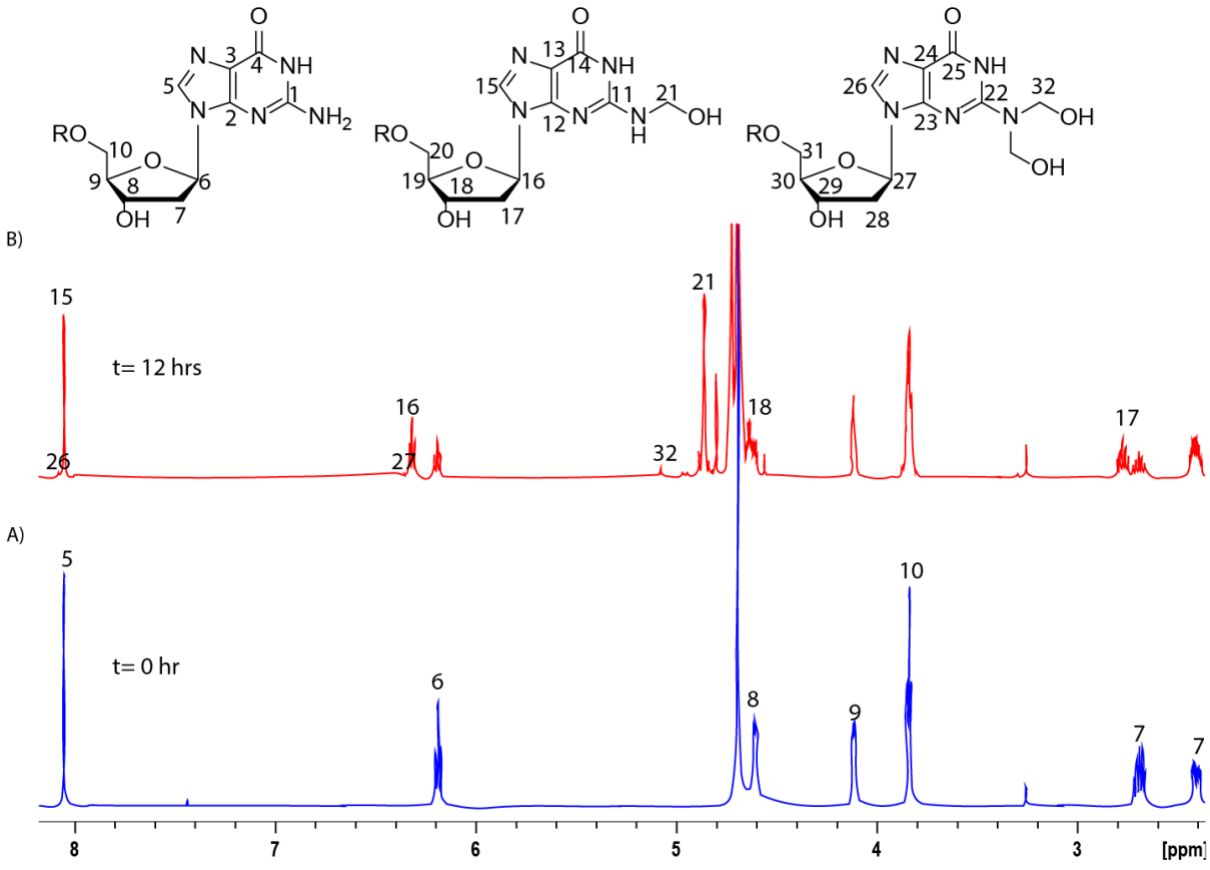


Figure 4.10 A) ^{1}H-NMR (500 MHz) spectrum of a 40 mM dGMP at 27°C, pD 7.5 in D_2O. The ^{1}H assignments of dGMP are indicated. B) ^{1}H-NMR (500MHz) spectrum for the reaction mixture of 40 mM dGMP and excess formaldehyde after 12 hours following initial mixing at 27°C, pD-7.5. ^{1}H assignments of the newly developed resonances are indicated.

New resonances were observed to develop at δ_H 8.14 ppm, δ_H 8.08 ppm, δ_H 6.41 ppm, δ_H 6.38 ppm, δ_H 5.12 ppm, δ_H 4.93 ppm, δ_H 4.68 ppm, δ_H 2.93 ppm and δ_H 2.82 ppm. Investigation of the recorded ^{1}H NMR spectrum revealed the formation of two closely related products.

The resonance at δ_H 4.93 ppm (marked as 21 in Figure 4.10.B) correlates with the carbon at δ_C 64.6 ppm in 2D HSQC and shows correlation with the carbon at δ_C 152.3 ppm (marked as 11 in Figure 4.10.B) in 2D HMBC (Figure 4.11; *wide infra*). Furthermore, the intensity of the resonance at δ_H 4.93 ppm was approximately twice the intensity of the resonance at δ_H 8.08

ppm (marked as 15 in Figure 4.10.B). Thus, the resonance at δ_H 4.93 ppm was assigned to the methylene protons attached to *N*-hydroxymethyl group of N^2-hydroxymethyldeoxyguanosinemonophosphate (hm^2dGMP). The resonances at δ_H 8.08 ppm, δ_H 6.38 ppm, δ_H 4.68 ppm and δ_H 2.82 ppm correspond to the protons attached to C-8 (marked as 15 in Figure 4.10.B) of the purine ring, C-1, C-3 and C-2 of the deoxyribose ring (marked as 16, 18 and 17, respectively in figure 4.10.B) of hm^2dGMP.

The resonance at δ_H 5.12 ppm (marked as 32 in Figure 4.10B) correlates to the carbon at δ_C 72.2 ppm in 2D HSQC and shows correlations with its own carbon and the carbon at δ_C 153.8 ppm (marked as 22 in Figure 4.11) in 2D HMBC, indicating symmetry in the structure (Figure 4.11). Furthermore, the intensity of the resonance at δ_H 5.12 ppm was found to be approximately four times the intensity of the resonance at δ_H 8.14 ppm (marked as 26 in Figure 4.10.B). As a result, the resonance at δ_H 5.12 ppm was assigned to the methylene protons attached to two hydroxymethyl groups of N^2,N^2-di(hydroxymethyl)deoxyguanosinemonophosphate (dhm^2dGMP). dhm^2dGMP has not been previously reported.

The mixture was analysed again after one week, and a new resonance was observed to develop at δ_H 5.0 ppm with carbon correlations at δ_C 47.8 ppm in 2D HSQC and at δ_C 153.7 ppm in 2D HMBC (Figure 4.11). The resonance was assumed to be the methylene protons attached to the carbon forming bridge between two guanidines via *N*-2. However, its concentration was consistently low within 12 hours following the initial mixing, hindering full characterisation. By NMR analysis, the conversion of dGMP to the hm^2dGMP adduct was approximately ~50% and to dhm^2dGMP was approximately ~1%.

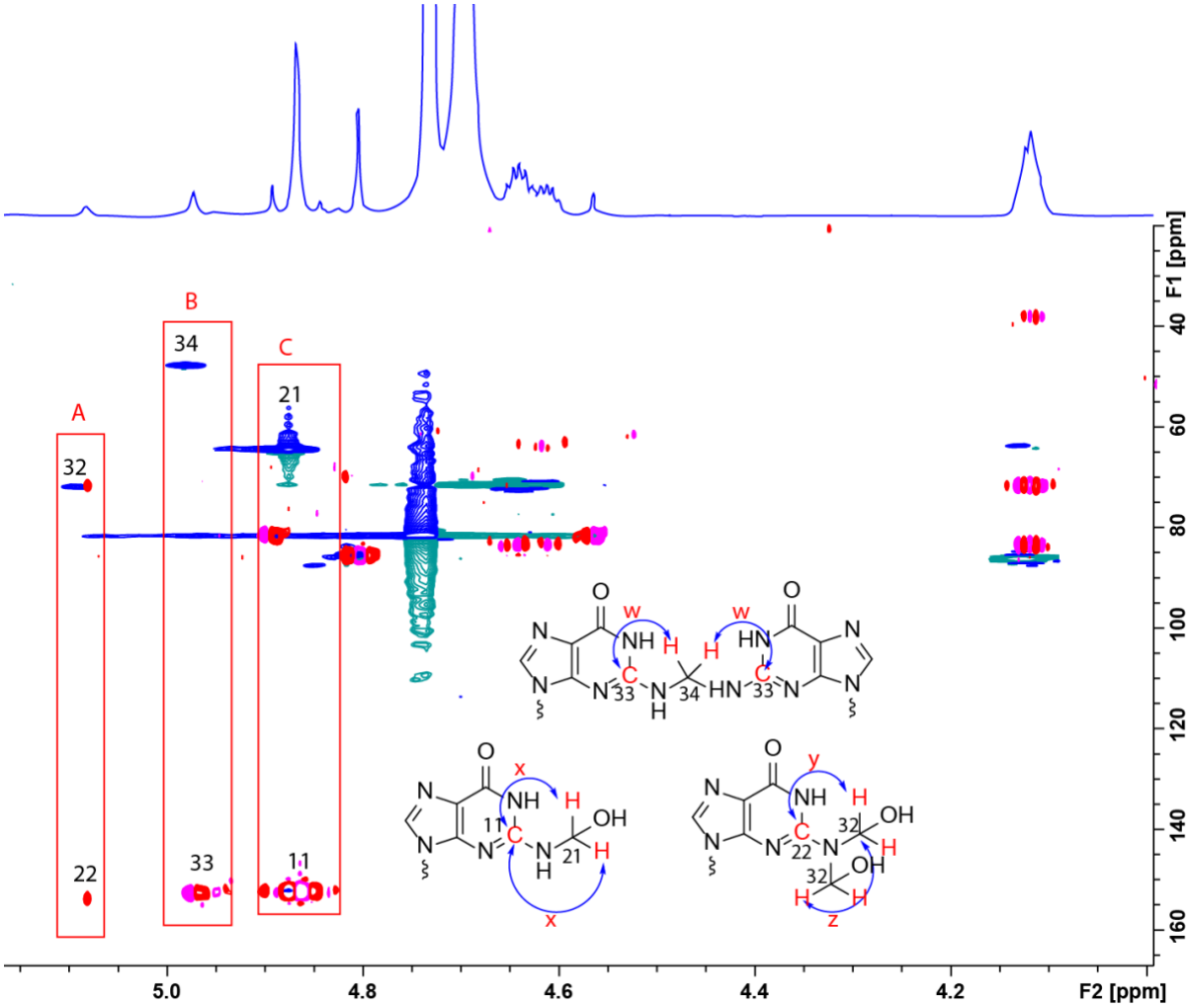


Figure 4.11 Overlaid 2D HSQC (green and blue) and HMBC (red and pink) NMR spectra (500 MHz) of the reaction mixture of a 40 mM dGMP and formaldehyde at 25°C at pD-9.0 after one week. Cross-peaks highlighted in Box A correspond to correlations (marked as "y" and "z" in red) observed for dhm²dGMP. Cross-peaks highlighted in Box B correspond to correlations (marked as "w" in red) observed for dGMP-CH₂-dGMP. Cross-peaks highlighted in Box C corresponded to correlations (marked as "x" in red) observed for hm²dGMP. Important correlations observed between formaldehyde derived protons (marked in red) and dGMP structure is highlighted with blue curved arrows.

The structures of adducts were also assessed by analysing the ESI mass spectrum obtained by LC purification. The EI/MS spectrum showed three peaks in the positive ion mode corresponding to m/z: 348.0699 ($[dGMP+H]^+$, calculated: 348.0704), 400.0628 ($[hm^2dGMP+Na]^+$, calculated: 400.0629), 406.0752 ($[dhm^2dGMP-H]^-$, calculated: 406.0769). The effect of changes in pD on the initial rates of the reaction were then investigated. At pD 7.5, the resonance at δ_H 4.93 ppm was observed to develop approximately 7 min after initial mixing (i.e. the concentration of hm^2dGMP within 7 min was above the NMR detection

limit). The resonance at δ_H 5.12 ppm was observed to develop approximately 74 min after initial mixing. The concentration of dhm^{2}dGMP remained consistently low over the time-course. The concentration of hm^{2}dgmp slowly increased with time. However, after 400 min, no significant change in the concentration of hm^{2}dgmp was observed. The concentration of hm^{2}dgmp reached approximately 1.1 mM after 400 min and remained constant at 1.15 mM after 600 min. The initial rate of the reaction was calculated to be 1.15 x 10^{-7} mol dm^{-3} s^{-1}. Analogous time-courses were then conducted at pDs 6 and 9. The analysis of both these time-courses revealed the formation of hm^{2}dgmp and dhm^{2}dgmp. Furthermore, as evident from the Figure 4.12, the changes in pD have no effect on the initial rates of the reaction (1.19 x 10^{-7} mol dm^{-3} s^{-1} at pD 6, and 1.2 x10^{-7} mol dm^{-3} s^{-1} at pD 9.0).

4.4.5.2 Conclusions from the results of the section 4.4.5

In conclusion, ^{1}H NMR analysis reveals the formation of two closely related adducts when dGMP is reacted with excess of formaldehyde. The newly developed resonance at δ_H 4.93 ppm is assigned to the methylene protons of *N*-hydroxymethyl group of hm^{2}dGMP and the resonance at δ_H 5.1 ppm is assigned to the methylene protons attached to two *N*-hydroxymethyl groups of dhm^{2}dGMP (Figure 4.8). However, ^{1}H NMR analysis after one week after mixing reveals the formation of cross-linked dGMP via *N*-2. No reaction was observed at *N*-1 of dGMP, as proposed in the literature.[20] The NMR results were further supported by MS analysis. The initial rate as calculated from the graph was 1.15 x 10^{-7} mol dm^{-3} s^{-1} at pD 7.5, 27°C. The changes in pD have no effect on the rates of initial reaction.

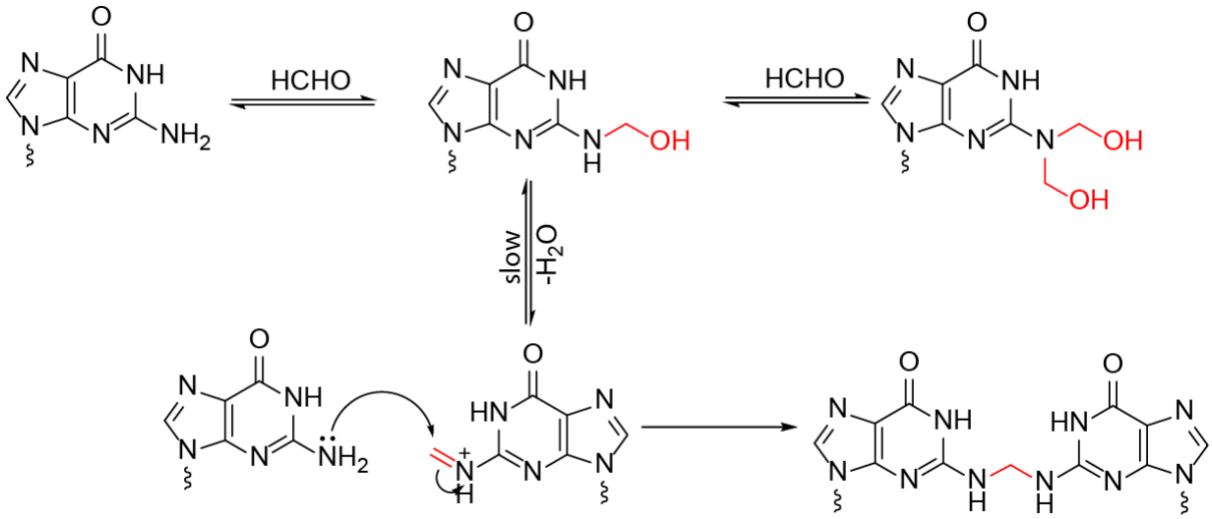


Figure 4.11 Schematic representation of the proposed mechanism for the reaction of dGMP and formaldehyde. The protons and carbon derived from formaldehyde are marked in red.

Table 4.4.5.1: Proton and carbon assignment for dGMP

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
1	153.4	-	-	-
2	150.6	-	-	-
3	115.6	-	-	-
4	158.4	-	-	-
5	137.6	8.08	s	-
6	83.5	6.25	t	7.0
7	38.8	2.43, 2.74	m, m	-
8	71.5	4.65	dt	6.0, 3.0
9	86.1	4.13	q	4.0
10	64.2	3.83	m	-

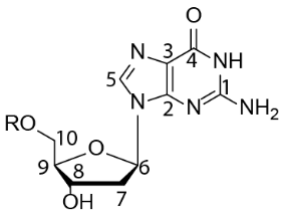


Table 4.4.5.2: Proton and carbon assignments for hm²dGMP.

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
11	152.3	-	-	-
12	150.9	-	s	-
13	116.4	-	-	-
14	159.4	-	-	-
15	138.0	8.08	s	-
16	83.5	6.38	t	7.0
17	38.3	2.45, 2.82	m, m	-
18	71.6	4.68	dt	6.0, 3.0
19	86.1	4.13	q	4.0
20	64.2	3.84	m	-
21	64.6	4.93	bs	-

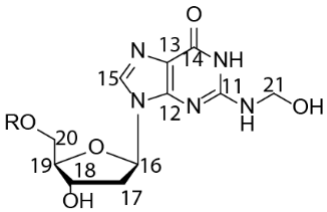
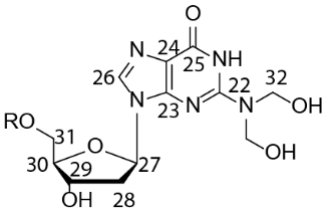


Table 4.4.5.3: Proton and carbon assignments for dhm²dGMP

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
22	153.8	-	-	-
23	150.7	-	s	-
24	116.4	-	-	-
25	158.8	-	-	-
26	138.0	8.14	s	-
27	83.6	6.41	t	7.0
28	38.3	2.45, 2.93	m, m	-
29	85.8	4.13	dt	6.0, 3.0
30	71.6	4.68	q	4.0
31	64.2	3.91	m	-



32	72.2	5.12	bs	-

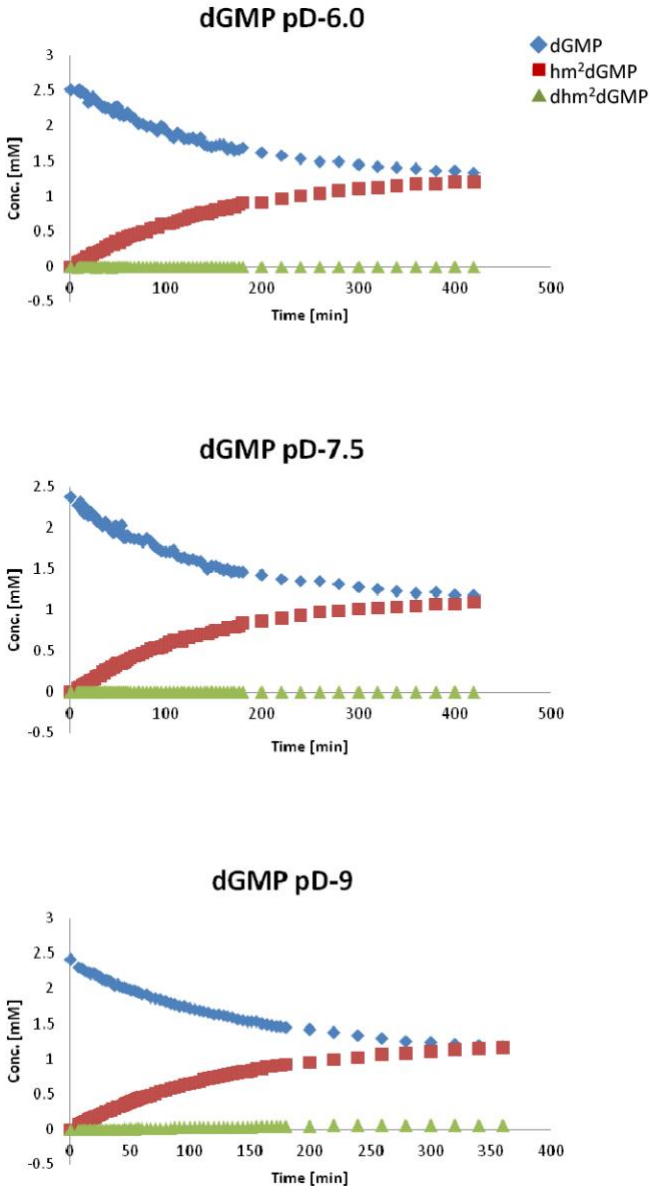


Figure 4.12 Graph showing change in concentrations of dGMP (blue), hm²dGMP (red), dhm²dGMP (green) over time at 27°C at pD 6.0 (top); pD 7.5 (middle); pD 9.0 (bottom). The concentrations are obtained from the integration of the peaks corresponding to resonance of proton attached to C-1 of the deoxyribose ring of dGMP, hm²dGMP and dhm²dGMP (marked as 6, 16 and 27, respectively, in Figure 4.10) and normalising the intensity value relative to internal standard, TSP. ¹H-NMR time-courses were recorded for 8-10 hours; every 2 min over the first hour, every 4 min over the next 2 hours and every 20 mins for the next 5-7 hours.

4.4.6 Studies on the reaction with formaldehyde with Deoxycytidinemonophosphate (dCMP)

4.4.6.1 ^{1}H-NMR analysis on the reaction of formaldehyde with dCMP

Initially, a ^{1}H NMR spectrum of a 40 mM solution of dCMP was recorded and assigned with the help of the correlations observed in recorded 2D NMR (HSQC and HMBC). Seven distinct ^{1}H resonances were observed, including all those correlating to the non-exchangeable protons. The resonances at δ_H 7.97 ppm and δ_H 6.04 ppm were assigned to the protons attached to C-6 and C-5 (marked as 4 and 3 in Fig, 4.13.A) of the pyrimidine ring of dCMP. The resonances at δ_H 6.25 ppm, δ_H 4.48 ppm, δ_H 4.08 ppm, δ_H 3.89 ppm and δ_H 2.36 ppm - 2.18 ppm were assigned to the protons attached to C-1, C-3, C-4, C-5 and C-2 (marked as 5, 7, 8, 9 and 6, respectively, in Figure 4.13.A) of the deoxyribose ring of dCMP.

To this solution, excess formaldehyde was added and ^{1}H-NMR, 2D-HSQC, 2D-HMBC spectra were recorded after 12 hr incubation. The proton and the carbon assignments of dCMP and newly formed adducts are given in Table 4.4.6.(1-3).

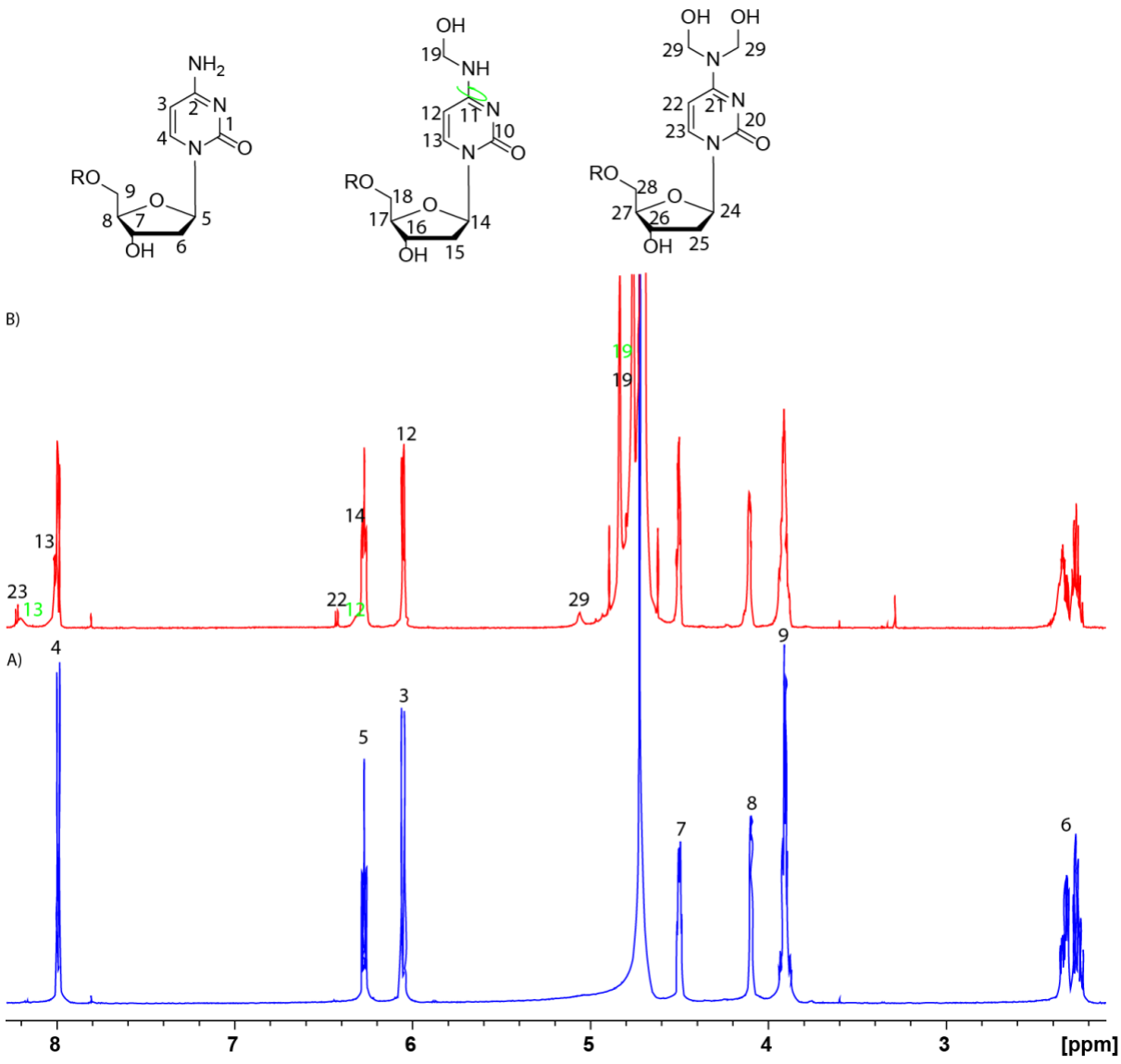


Figure 4.13 A) ^{1}H-NMR (500 MHz) spectrum of a 40 mM dCMP at 27°C, pD 7.5 in D_2O. The ^{1}H assignments of dCMP are indicated. B) ^{1}H-NMR (500MHz) spectrum for the reaction mixture of 40 mM dCMP and excess formaldehyde after 12 hours following initial mixing at 27°C, pD-7.5. ^{1}H assignments of the newly developed resonances are indicated. The assignments in green represent the resonances corresponding to rotamer of hm^{4}dCMP.

New resonances were observed at δ_H 8.22 ppm, δ_H 8.21 ppm, δ_H 7.99 ppm, δ_H 6.4 ppm, δ_H 6.25 ppm, δ_H 6.04 ppm, δ_H 6.31 ppm, δ_H 5.06 ppm and δ_H 4.83 ppm. The new resonances at δ_H 6.31 ppm, δ_H 6.25 ppm, and δ_H 6.04 ppm (marked as 12(green), 14 and 12 in Figure 4.13.B) possessed similar ^{1}H chemical shifts to the protons attached to C-1 of the deoxyribose ring and C-5 of the pyrimidine ring (marked as 5 and 3 in Figure 4.13.A) of dCMP. Further, the new resonance at δ_H 4.83 ppm possessed the same ^{1}H chemical shift as

poly(oxymethylene)hemiformal (Figure 4.4B), which is an impurity arising from formaldehyde oligomerisation. It should be noted that the resonance at δ_H 4.83 ppm (marked as 19 and 19 (green) in Figure 4.13.B) was observed as two resonances with identical ^{1}H chemical shifts but with slightly different ^{13}C chemical shifts (as observed in 2D HSQC, Figure 4.14). The proton resonance at δ_H 4.83 ppm correlates with the carbon resonance at δ_C 63.5 ppm and δ_C 65.3 ppm in 2D HSQC (Figure 4.14) and shows correlation with the carbon resonance at δ_C 164.8 ppm in 2D HMBC (Figure 4.14). The major carbon resonance at δ_C 63.5 ppm in 2D HSQC has a similar chemical shift to resonances corresponding to *N*-hydroxymethyl groups of hm^{6}dAMP and hm^{2}dGMP, and therefore, this resonance was assigned to the *N*-hydroxymethyl group of *N*^{4}-hydroxymethyldeoxycytidinemonophosphate (hm^{4}dCMP). Thus, one of the resonances at δ_H 4.83 ppm is assigned to the methylene protons attached to *N*-hydroxymethyl group of hm^{4}dCMP. The resonance at δ_H 7.99 ppm was assigned to the proton attached to C-6 of the pyrimidine ring of hm^{4}dCMP. The overlapped resonances at δ_H 6.25 ppm and δ_H 6.04 ppm are assigned to protons attached to C-1 of the deoxyribose ring and C-5 of the pyrimidine ring (marked as 14 and 12 in Figure 4.13.B) of hm^{4}dCMP respectively.

The resonance at δ_H 5.06 ppm shows a correlation with the carbon at δ_C 71.6 ppm in 2D HSQC and with the carbon at δ_C 164.8 ppm (Figure 4.14) in 2D HMBC. However, the characteristic correlation with its own carbon at δ_C 71.6 ppm is not observed in 2D HMBC. It was proposed that the fast rotation of the bond could result in broadening of the resonance at δ_H 5.06 ppm, so restricting development of cross-peak/correlation in 2D HMBC. The resonance at δ_H 5.06 ppm is proposed to be the methylene protons attached to the two *N*-hydroxymethyl groups of *N*^{4},*N*^{4}-di(hydroxymethyl)deoxycytidinemonophosphate (dhm^{4}dCMP) based on the similar chemical ^{1}H and ^{13}C chemical shifts observed for previously studied *N*,*N*-dihydroxy nucleoside adducts. The resonances at δ_H 8.22 ppm and δ_H

6.42 ppm are proposed to correspond to the protons attached to C-6 and C-5 of the pyrimidine ring (marked as 23 and 22 in Figure 4.13.B) of dhm^{4}dCMP.

Assigning the resonances at δ_H 8.21 ppm, δ_H 6.31 ppm and δ_H 4.83 ppm (with carbon correlations at δ_C 65.3 ppm in 2D HSQC) posed an analytical challenge. Given their similar chemical shifts to resonances of dhm^{4}dCMP, it was proposed that the resonances correspond to a rotameric form of hm^{4}dCMP. To confirm this hypothesis, experiments were conducted at a lower temperature, which should sharpen the otherwise broaden resonances, and also should shift the solvent and formaldehyde resonances in ^{1}H-NMR, simplifying the analysis (Figure 4.15).

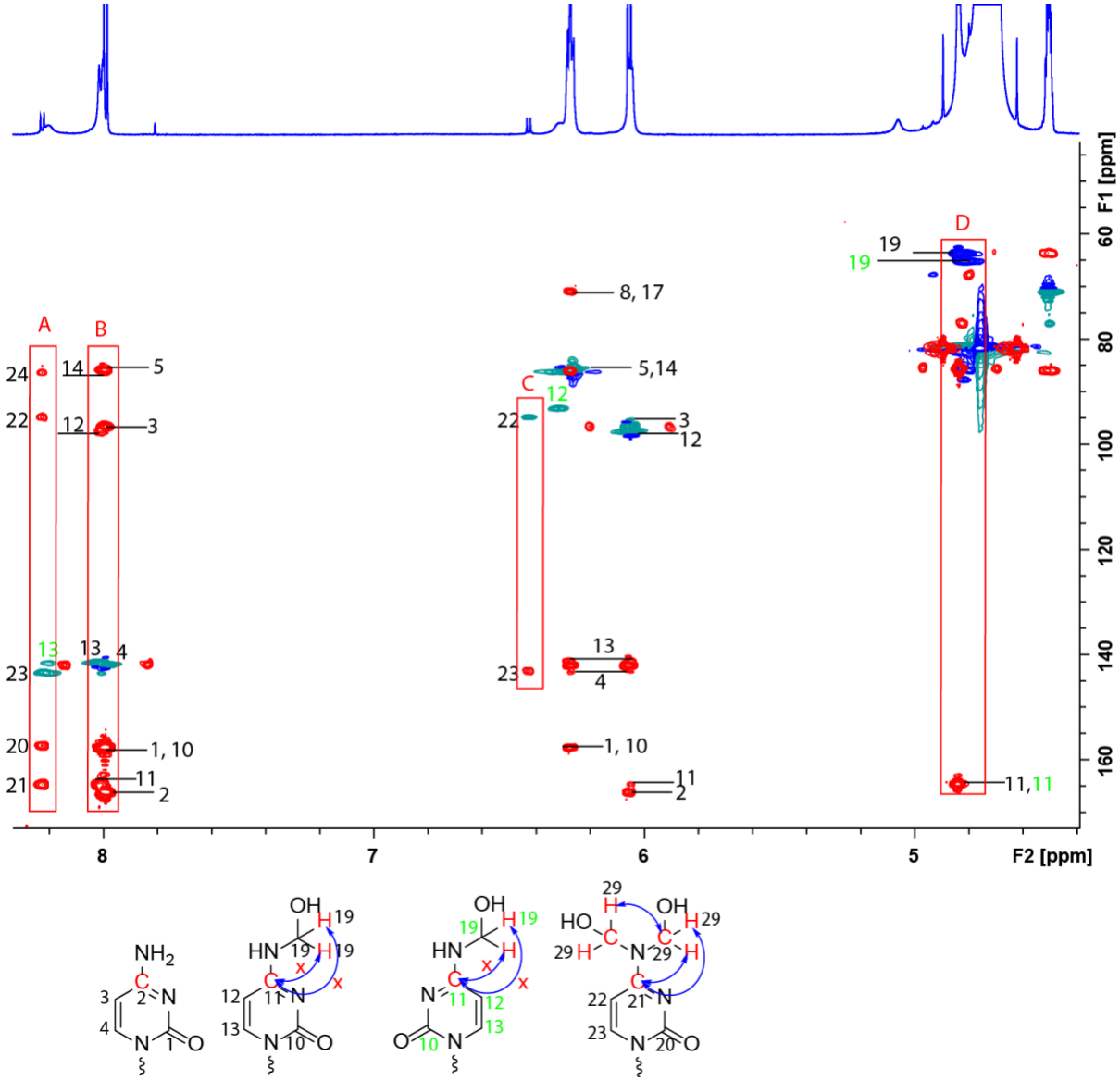


Figure 4.14 Overlaid 2D HSQC (green and blue) and HMBC (red) NMR spectra (500 MHz) of the reaction mixture of 40 mM solution of dCMP and formaldehyde at 25°C. Cross-peaks highlighted in Box A corresponded to correlations observed for dhm^{4}dCMP and rotamer of hm^{4}dCMP. Cross-peaks highlighted in Box B corresponded to correlations observed for hm^{4}dCMP and dCMP. Cross-peaks highlighted in Box C corresponded to correlations observed for dhm^{4}dCMP. Cross-peaks highlighted in Box D corresponded to correlations (marked as "x" in red) observed for both the rotamers of hm^{4}dCMP. Important correlations observed between formaldehyde derived protons (marked in red) and dCMP structure are highlighted with blue curved arrows.

Excess formaldehyde was added to a 40 mM solution of dCMP and the reaction was allowed to stand for 4 days to observe maximum concentrations of products. The temperature of the probe was then lowered down to 15 °C and ^{1}H-NMR and 2D HSQC spectra were recorded on the resultant mixture.

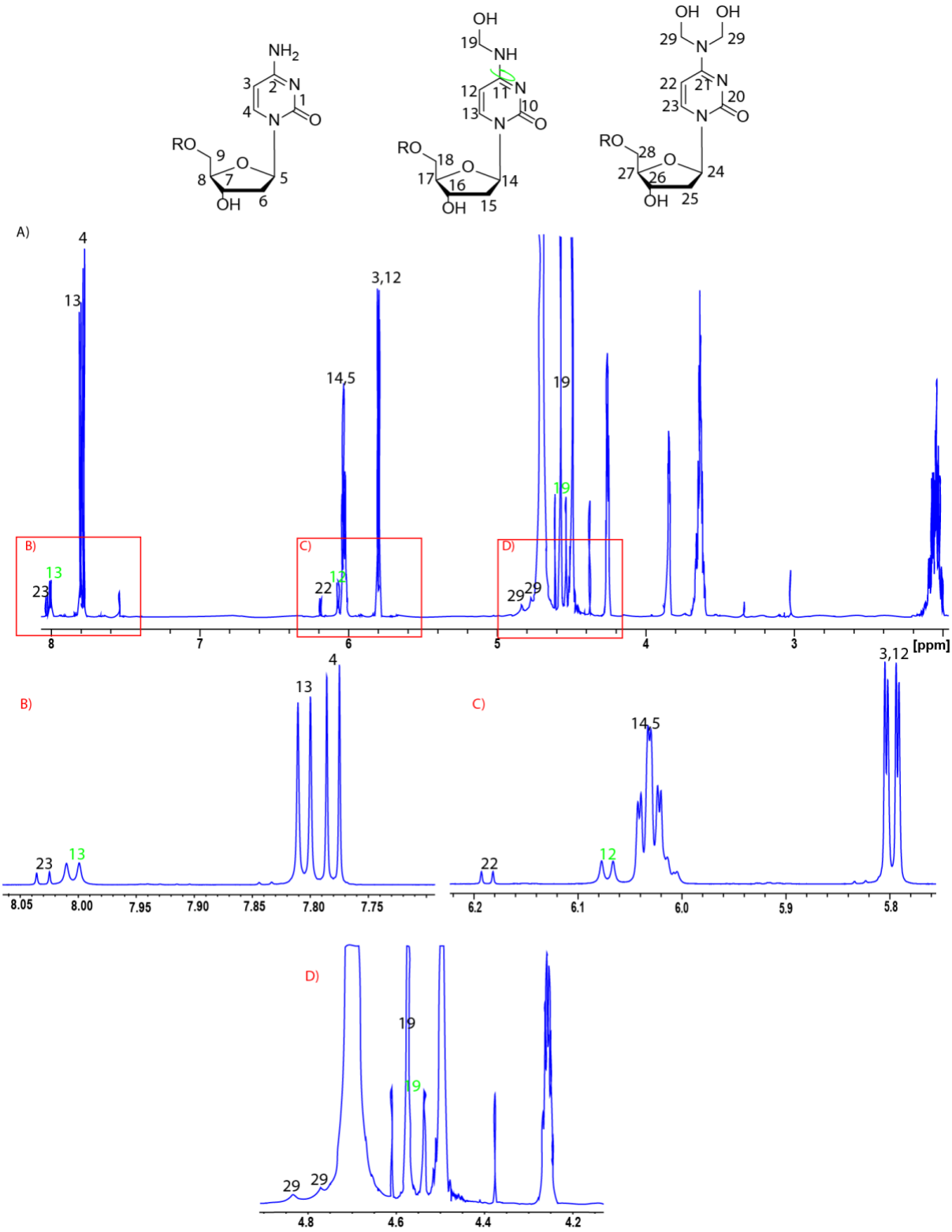


Figure 4.15 ^{1}H-NMR (700 MHz) spectrum for the reaction mixture of a 40mM dCMP and excess formaldehyde after 4 days following initial mixing, at 15°C, pD-7.5(top). The ^{1}H assignments of important resonances are indicated. The assignments in green represent the resonances corresponding to rotameric form of hm^{4}dCMP. B) represents the zoomed ^{1}H-NMR region in Box B (marked red). C) represents the zoomed ^{1}H-NMR region in Box C (marked red). D) represents the zoomed ^{1}H-NMR region in Box D (marked red).

At 15 °C, the resonances corresponding to the protons of formaldehyde had upshifted to δ_H 4.5 ppm from δ_H 4.79 ppm (Figure 4.13.B) and the resonances corresponding to the protons attached to poly(oxymethylene)hemiformal had shifted to δ_H 4.57 ppm from δ_H 4.83 ppm as observed in ^{1}H-NMR at 27°C (Figure 4.13.B). The ^{13}C-satellites of hydrated formaldehyde were observed at δ_H 4.6 ppm and δ_H 4.38 ppm as compared to δ_H 4.78 ppm and δ_H 4.55 ppm at 27 °C.

New resonances were observed at δ_H 8.03 ppm, δ_H 8.0 ppm, δ_H 7.8 ppm, δ_H 6.18 ppm, δ_H 6.07 ppm, δ_H 4.83 ppm, δ_H 4.77 ppm, δ_H 4.574 ppm and δ_H 4.53 ppm. The new resonances at δ_H 6.03 ppm and δ_H 5.8 ppm overlap with the resonances from the starting material. The resonances at δ_H 8.0 ppm and δ_H 6.07 ppm were observed to resolve into sharp doublets; the same resonances were broad at 27°C (Figure 4.13.B). The resonances at δ_H 4.574 ppm and δ_H 4.53 ppm showed correlation with the carbons at δ_C 63.07 ppm and δ_C 64.77 ppm, respectively, in the 2D-HSQC spectrum (Figure 4.16).

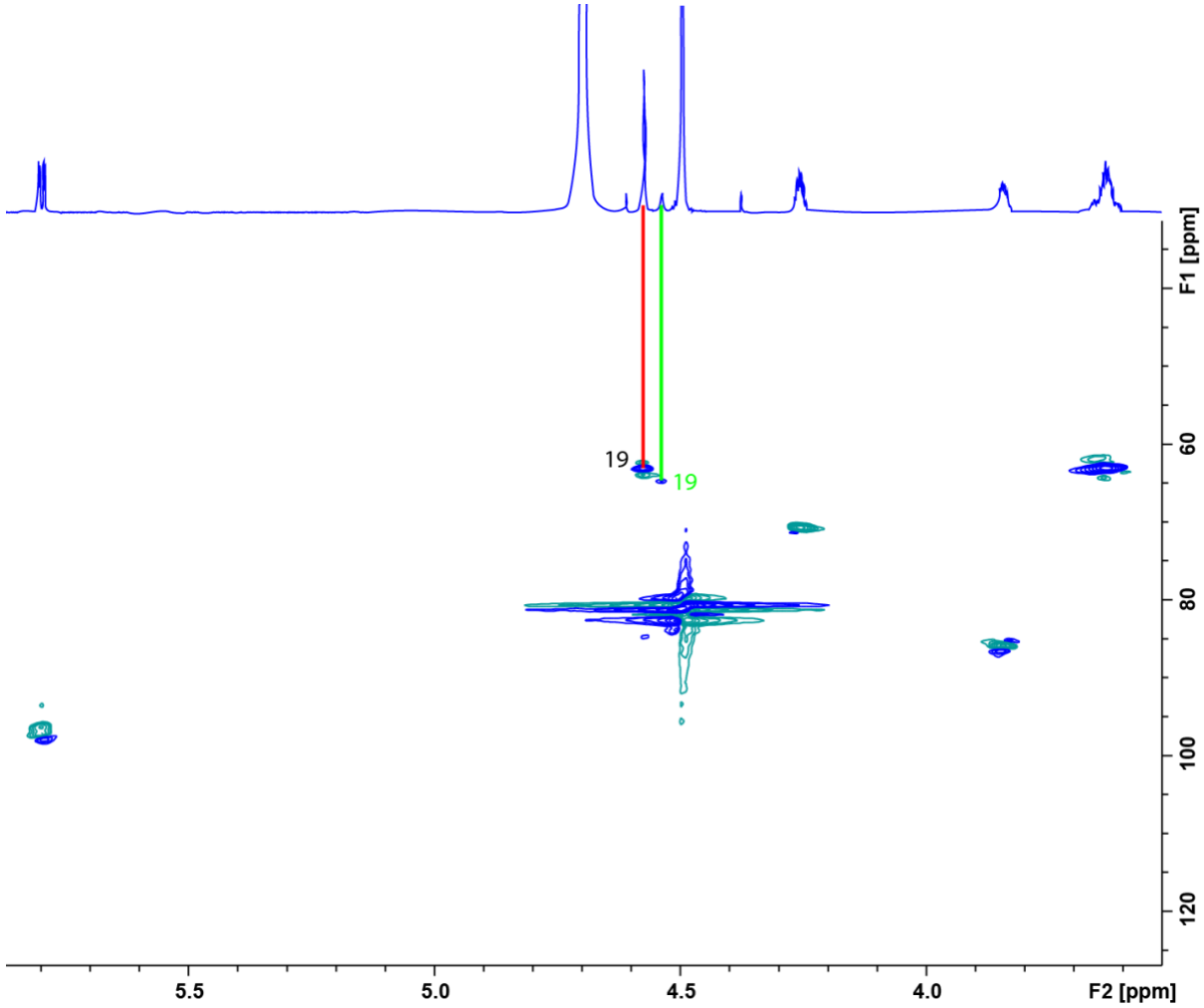


Figure 4.16 2D HSQC (green and blue) (700 MHz) spectrum for the reaction mixture of a 40 mM dCMP and formaldehyde at 15°C. Cross-peaks highlighted by red and green line correspond to the correlations observed for major and minor rotameric form of hm^{4}dCMP, respectively. The cross-peak correlations for methylene carbon of N-hydroxymethyl groups of two rotameric form of hm^{4}dCMP are indicated. Red and Green stick indicate the correlation observed for the major and minor rotamer of hm^{4}dCMP.

The two overlapped resonances at δ_H 4.83 ppm in Figure 4.13.B resolved into two separate resonances at δ_H 4.57 ppm and δ_H 4.53 ppm. The intensity of the resonance at δ_H 4.57 ppm was approximately four times the intensity of resonance at δ_H 4.53 ppm. Thus, the major proposed rotamer was approximately four times the concentration of the minor rotamer. Further, the intensities of the resonances at δ_H 4.57 ppm and δ_H 4.53 ppm were approximately twice those of the resonances at δ_H 7.8 ppm and δ_H 8.0 ppm, respectively. Thus, the resonances at δ_H 4.57 ppm and δ_H 4.53 ppm are assigned to the methylene protons attached to the *N*-hydroxymethyl groups of the two rotameric form of hm^{4}dCMP. The resonances at δ_H

8.0 ppm and δ_H 7.8 ppm were assigned to the protons attached to C-6 (marked as 13 (green) and 13 in Figure 4.15) of the minor and major rotamers of hm^{4}dCMP. To further support the hypothesis that hm^{4}dCMP exists as two rotamers, 1D gradient NOESY experiments were carried out (mixing time = 10-300 µs; Figure 4.17). Following excitation of the resonance at δ_H 7.8 ppm, the intensity of the resonance at δ_H 8.0 ppm was observed to increase with longer mixing times, implying the two resonances correspond to two rotamers of hm^{4}dCMP.

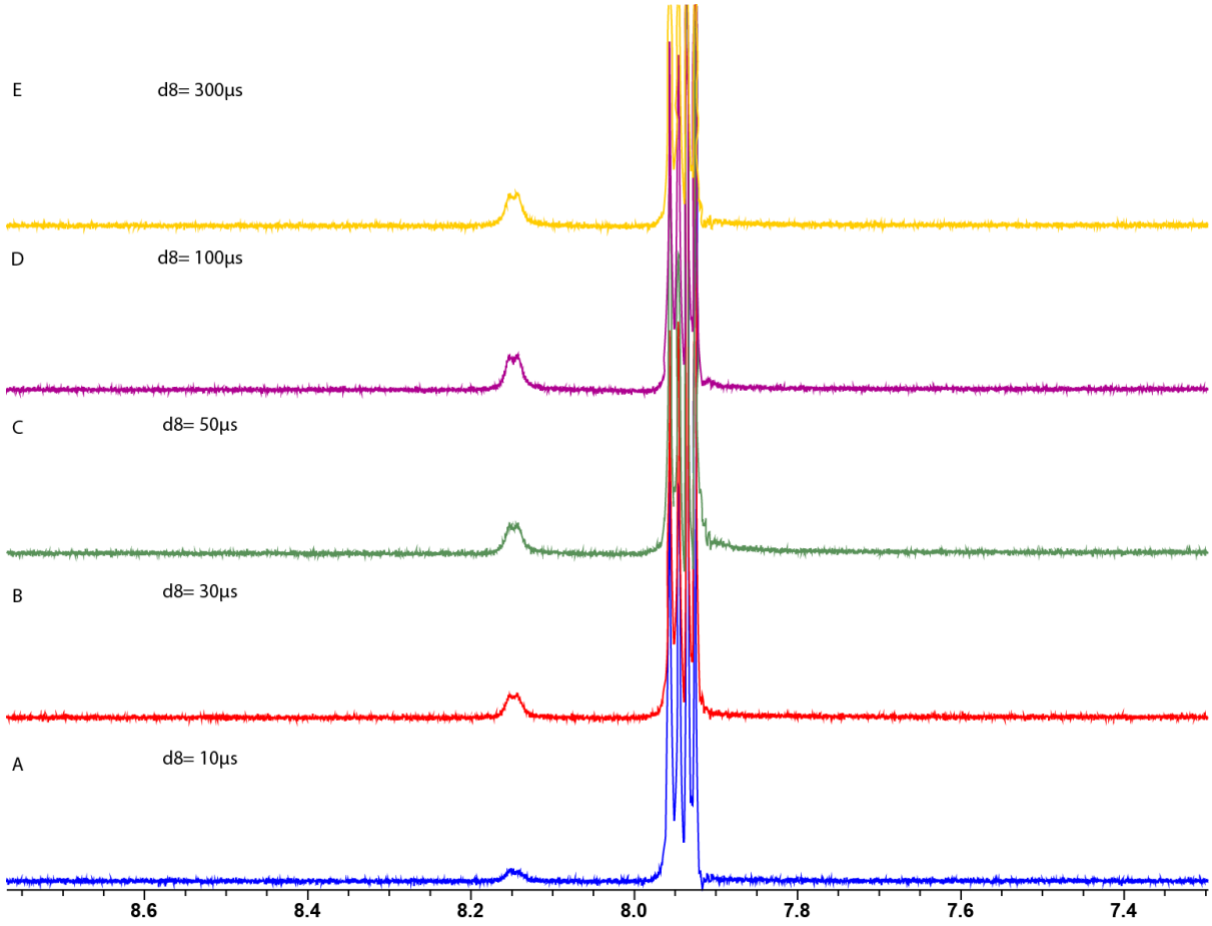


Figure 4.17 Overlaid 1D noesy spectra of a reaction mixture of dCMP and formaldehyde at 15°C. Spectra were obtained by changing the mixing time (d8) from 10µs-300µs.

At 15 °C, the resonances at δ_H 4.83 ppm and δ_H 4.77 ppm were observed as two singlets; at 27 °C, these resonances coalesced as a broad singlet at δ_H 5.06 ppm (marked as 29 in Figures 4.15 and 4.13.B). The intensities of the resonances at δ_H 4.83 ppm and δ_H 4.77 ppm (marked as 29 in Figure 4.15.A) were approximately twice the intensities of resonances at δ_H 8.03 ppm and δ_H 6.18 ppm (marked as 23 and 22 in Figure 4.15.A). Thus, two different observed

resonances, at δ_H 4.83 ppm and δ_H 4.77 ppm, correspond to the methylene protons attached to two *N*-hydroxymethyl groups (marked as 29 in Figure 4.15) of dhm^{4}dCMP. dhm^{4}dCMP has not been previously described. No other adducts were observed, at least during the initial 12 hours of the reaction. By NMR analysis, the conversion of dCMP to hm^{4}dCMP was approximately 65 % and to dhm^{4}dCMP was approximately 3 %.

The structures of adducts were also assessed by analysing the ESI mass spectrum obtained by LC purification. The EI / MS spectrum showed three peaks in the positive ion mode corresponding to m/z: 308.0645 ([dCMP+H]$^+$, calculated: 308.0645), 338.0732 ([hm^{4}dCMP+H]$^+$, calculated: 338.0748) , 366.0718 ([dhm^{4}dCMP-H$^+$], calculated: 366.0708).

Table 4.4.6.1 Proton and carbon assignments for dCMP

Position	δ ^{13}C (ppm)	δ ^{13}H(ppm)	Multiplicity	$J_{H,H}$ (Hz)
1	157.4	-	-	
2	166.4	-	-	
3	96.5	6.04	d	7.5
4	141.6	7.97	d	7.5
5	85.8	6.25	t	7.0
6	39.2	2.36-2.18	m	-
7	71.6	4.48	dt	3.0,3.0,6.0
8	86.1	4.08	m	-
9	63.9	3.89	m	-

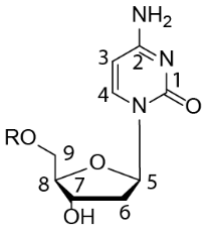


Table 4.4.6.2 Proton and carbon assignments for hm^{4}dCMP

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)

10	157.4	-	-	-
11	164.8	-	-	-
12	93.1,97.5	6.3,6.04	d	7.5
13	141.8,141.8	8.21,7.99	d	7.5
14	85.9	6.25	t	7.0
15	39.2	2.36-2.18	m	-
16	70.6	4.48	dt	3.0,3.0,6.0
17	86.4	4.08	m	-
18	65.2, 63.5	3.89	m	-
19	63.2	4.83	bs	-

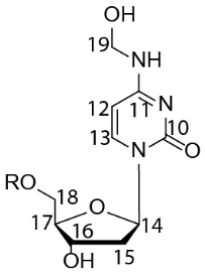
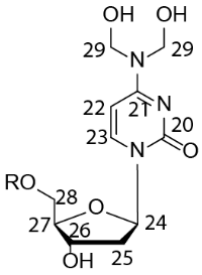


Table 4.4.6.3 Proton and carbon assignments for dhm[4]dCMP

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
20	157.2	-	-	-
21	164.8	-	-	-
22	94.7	6.04	d	7.5
23	143.3	8.22	d	7.5
24	85.9	6.4	t	7.0
25	39.2	2.36-2.18	m	-
26	70.6	4.48	dt	3.0,3.0,6.0
27	86.4	4.08	m	-
28	63.5	3.89	m	-
29	71.6	5.06	bs	-



The effects of changes in pD on the rates of the reaction were then investigated. A ^{1}H-NMR time-course was initially recorded; 2.4mM of dCMP, pre-buffered at pD 7.5, was reacted

with a 53-fold excess of formaldehyde, then immediately transferred to a 3mm NMR tube and analysed directly by ^{1}H-NMR. However, the recorded time-course at either 500 MHz or 600 MHz manifested highly overlapping resonances which compromised accurate analysis. Conducting the experiment on the 700 MHz spectrometer, however, reduced signal overlap, and consequently, all subsequent time-courses with dCMP were recorded at 700 MHz. The resonances corresponding to hm^{4}dCMP were observed to develop approximately 4 min following initial mixing (i.e. the concentration of hm^{4}dCMP after 4 min was above the NMR detection limit). The resonances corresponding to dhm^{4}dCMP were observed to develop after approximately 18 min following initial mixing. The concentration of dhm^{4}dCMP remained consistently low over the time-course. The concentrations of hm^{4}dCMP and its rotamer slowly increased with time. However, after approximately 500 min following initial mixing, significant changes in the concentration of hm^{4}dCMP were not observed. The concentration of hm^{4}dCMP reached approximately 1.5 mM after 500 min and remained constant afterwards. The initial rate of the reaction was calculated to be 2.1 x 10^{-7} mol dm^{-3} s^{-1}.

Analogous ^{1}H-NMR time-courses were conducted at pD-6 and 9. The analysis of these time-courses revealed the formation of hm^{4}dCMP and dhm^{4}dCMP. Furthermore, the changes in pD has no effect on the rates of the reaction (Figure 4.18; 1.5 x10^{-7} mol dm^{-3} s^{-1} at pD 6; 1.7 x10^{-7} mol dm^{-3} s^{-1} at pD 9). The concentrations of adducts effectively did not change with a change in pD.

4.4.6.2 Conclusion for the results described in section 4.4.6

In conclusion, ^{1}H-NMR analysis reveals the formation of three closely related products when dCMP is reacted with excess formaldehyde. The newly developed resonances at δ_H 4.83 ppm have been assigned to the methylene protons attached to the *N*-hydroxymethyl group of two hm^{4}dCMP rotamers; the resonance at δ_H 5.06 ppm was assigned to the methylene protons attached to two *N*-hydroxymethyl groups of dhm^{4}dCMP (Figure 4.13). 1D-NOESY spectra

were recorded to further support the presence of hm^{4}dCMP rotamers in the reaction mixture. The ^{1}H-NMR results are further supported by mass spectrometry analysis. The resonances corresponding to a cross-linked dCMP adduct are not observed. The initial rate as calculated from the graph was 2.1 x 10^{-7} mol dm^{-3} s^{-1} at pD-7.5. dCMP is found to be the most reactive among the nucleotides with exocyclic amines. Changes in pD have no effect on the rates of reaction.

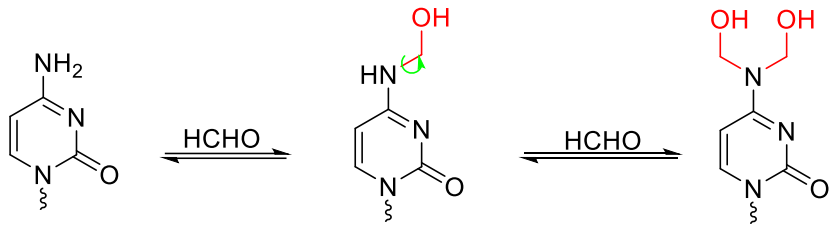


Figure 4.18 Schematic representation of the proposed mechanism for the reaction of dCMP and formaldehyde. The protons and carbon derived from formaldehyde are marked in red.

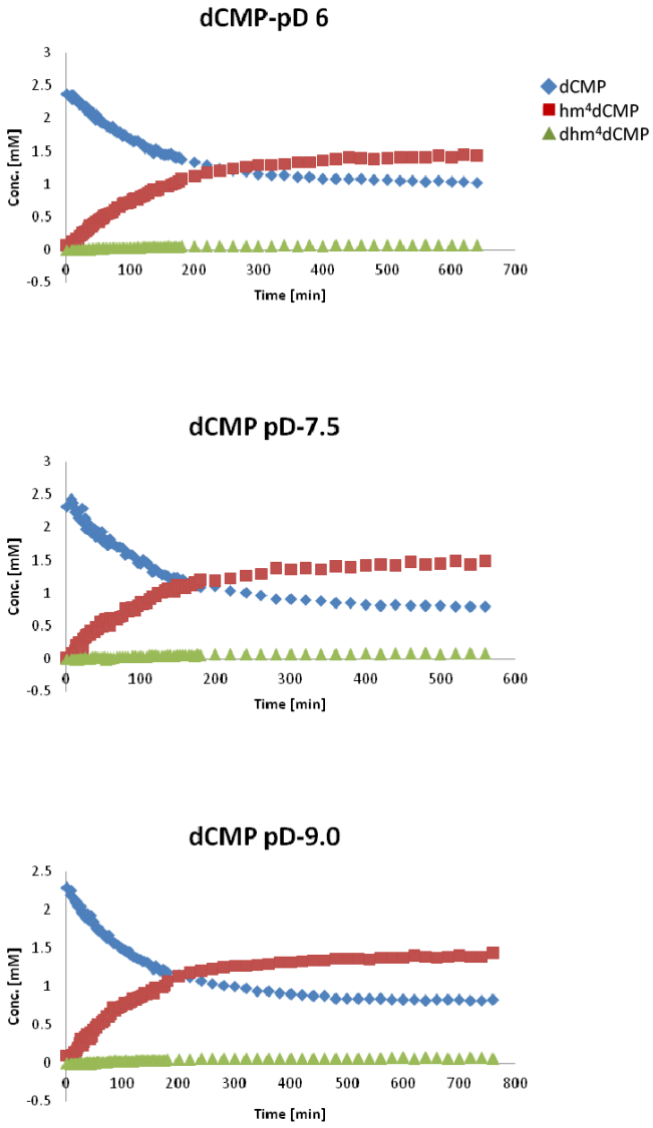


Figure 4.19 Graph showing changes in the concentrations of dCMP (blue), hm^{4}dCMP (red), dhm^{4}dCMP (green) over time at 27°C at pD 6.0 (top); pD 7.5 (middle); C) pD 9.0 (bottom). The concentrations are obtained from the integration of the peaks corresponding to resonance of the proton attached to C-6 of the pyrimidine ring of dCMP, hm^{4}dCMP and dhm^{4}dCMP (marked as 4, 13 and 23, respectively in Figure 4.13) and normalising the intensity value relative to internal standard, TSP. ^{1}H NMR timecourses were recorded for 8-10 hours; every 2 min over the first hour, every 4 min over the next 2 hours and every 20 mins for the next 5-7 hours.

4.4.7 Studies on the reaction of formaldehyde with Thymidinemonophosphate (TMP)

4.4.7.1 ^{1}H-NMR analysis on the reaction of formaldehyde with TMP

A ^{1}H NMR spectrum of a 40 mM solution of TMP in D_2O was initially recorded and assigned with the help of recorded 2D NMR (HSQC and HMBC). Seven distinct ^{1}H

resonances were observed. All the non-exchangeable protons were observed. The resonances at δ_H 7.75 ppm and δ_H 1.86 ppm were assigned to the protons attached to C-6 and methyl group (marked as 4 and 5 in Figure 4.20.A) of the pyrimidine ring of TMP respectively, while the resonances observed at δ_H 6.27 ppm, δ_H 4.51 ppm, δ_H 4.18 ppm, δ_H 4.07 ppm and δ_H 2.36-2.19 ppm were assigned to the protons attached to C-1, C-3, C-4, C-5 and C-2 (marked as 6, 8, 9, 10 and 7, respectively in Figure 4.19.A) respectively.

Excess formaldehyde was then added to a solution of TMP and monitored by ^{1}H-NMR, 2D HSQC, and 2D HMBC (Figure 4.19.B). The proton and the carbon assignments of TMP and a newly formed adduct are given in Tables 4.4.7.(1-2).

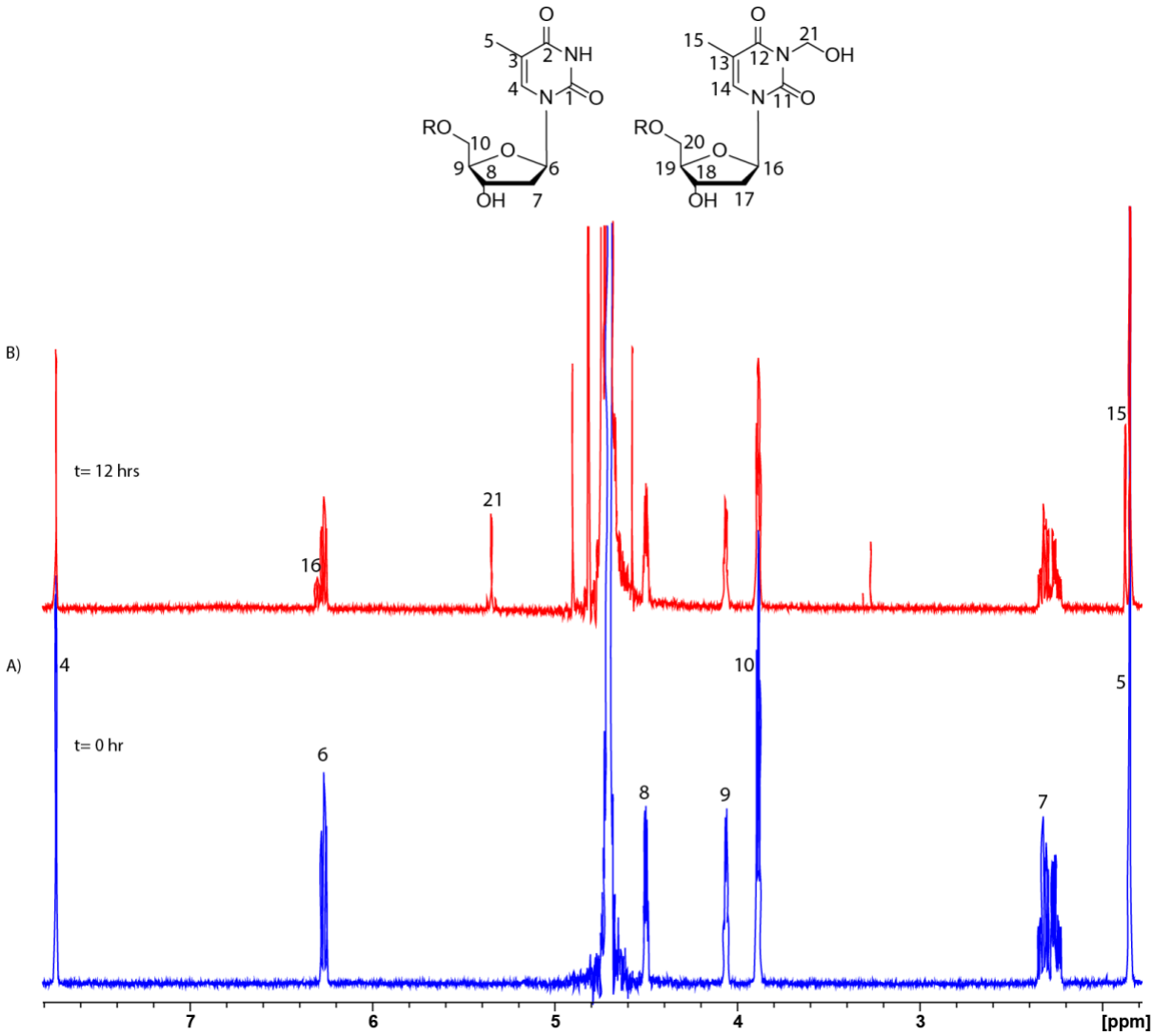


Figure 4.20 ^{1}H-NMR (500 MHz) spectrum of A) a 40 mM TMP at 27°C, pD 7.5 in D_2O; B) the reaction mixture of 40mM TMP and excess formaldehyde after 12 hours following initial mixing at 27°C, pD-7.5. The ^{1}H assignments of TMP and the newly developed resonances are indicated.

New resonances were observed at δ_H 6.3 ppm, δ_H 5.34 ppm and δ_H 1.87 ppm, suggesting formation a TMP-formaldehyde adduct with a similar overall structure to TMP. The resonance at δ_H 5.34 ppm (marked as 21 in Figure 4.20.B) correlates to the carbon at δ_C 64.7 ppm in 2D HSQC and shows correlations with the carbons at δ_C 165.0 ppm and δ_C 151.7 ppm (marked as 12 and 11 in Figure 4.21) in 2D HMBC. Furthermore, the intensity of the resonance at δ_H 5.34 ppm was approximately twice the intensity of the resonance at δ_H 6.3 ppm. Thus, the resonance at δ_H 5.34 ppm was assigned to the methylene protons attached to the *N*-hydroxymethyl group of N^3-hydroxymethylthymidinemonophosphate (hm^3tmp). The resonances at δ_H 6.3 ppm and δ_H 1.87 ppm were assigned to the protons attached to C-1 of the deoxyribose ring and C-5 methyl group of the pyrimidine ring (marked as 16 and 15, respectively in Figure 4.20.B) of hm^3tmp respectively. No other adducts were observed.

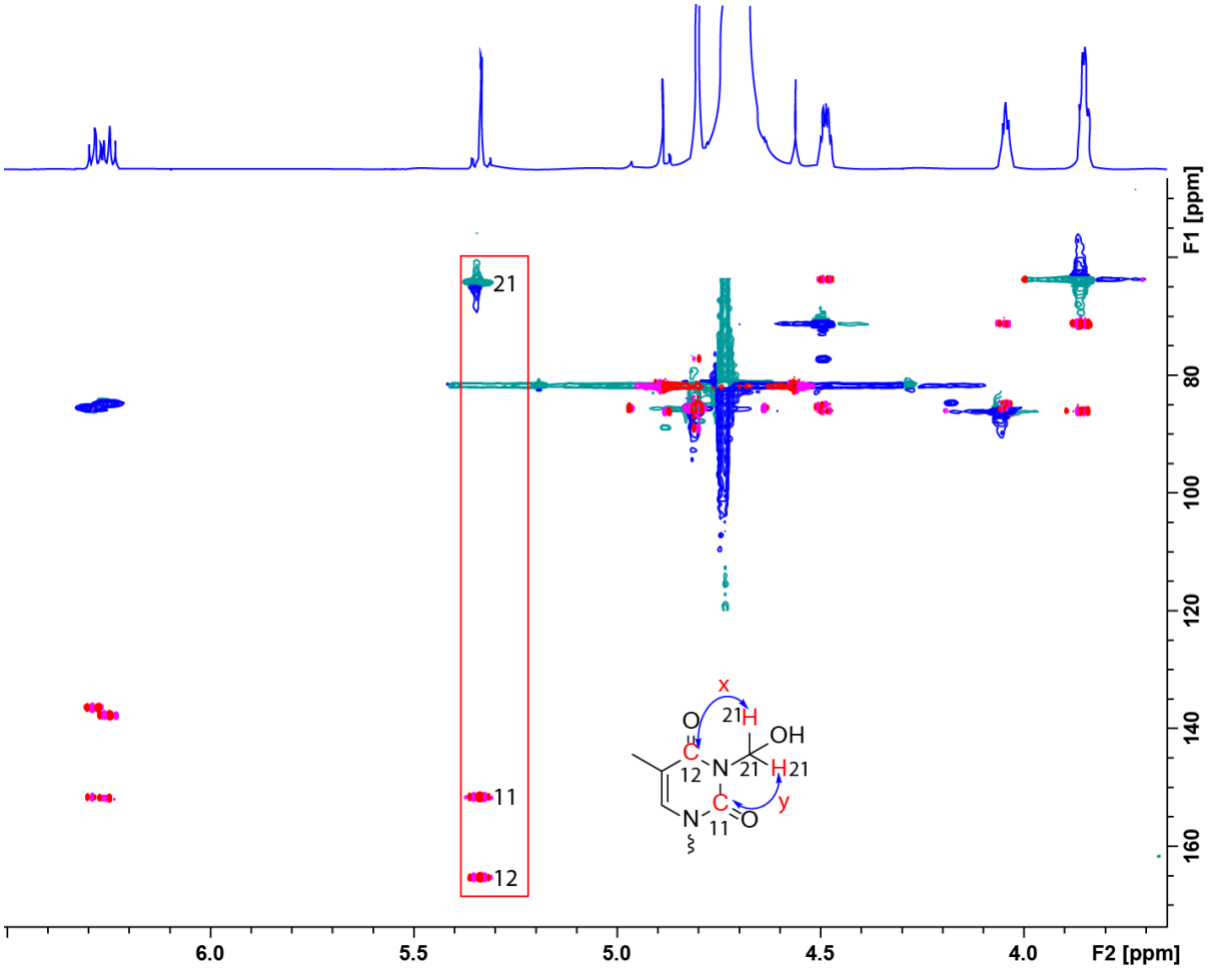


Figure 4.21 Overlaid 2D HSQC (green and blue) and HMBC (red and pink) NMR spectra (500 MHz) for the reaction mixture of 40 mM solution of TMP and formaldehyde at 25°C. Cross-peaks highlighted in Box A correspond to the correlations (marked as "x" and "y" in red) observed for hm^{3}TMP. Important correlations observed between formaldehyde derived protons (marked in red) and TMP structure are highlighted with blue curved arrows.

The conversion of TMP to hm^{3}tmp adduct was ~ 18%, as measured by ^{1}H-NMR. An EI / MS spectrum of the reaction mixture showed two peaks in the negative ion mode corresponding to m/z: 321.0494 ([TMP-H]$^-$, calculated: 321.0493) and 351.0602 ([hm^{3}TMP-H]$^-$, calculated: 351.0599). The effect of changing pD on the rate of the reaction was then investigated. A ^{1}H NMR time-course was initially recorded at pD 7.5; 2.4mM of TMP, prebuffered at pD 7.5, was reacted with a 53-fold excess of formaldehyde and transferred immediately to a 3mm NMR tube and analysed by ^{1}H-NMR. However, the resonance at δ_H 5.34 ppm was observed to develop to maximum intensity within approximately 6 min following initial mixing and the intensity remained constant thereafter. Attempts were then made to reduce the time-lapse by

recording the first ^{1}H-NMR spectrum within 3 minutes following initial mixing. The resonance at δ_H 5.34 ppm was observed to develop to a maximum intensity within 3 min following initial mixing and intensity remained constant thereafter. Thus, it was concluded that the concentration of hm^{3}tmp was observed to reach a maximum (in this case, 0.5 mM) almost immediately after addition of formaldehyde, suggesting a dynamic equilibrium. No other adducts were observed at least during 12 hours following the initial mixing. The initial rate of the reaction was too fast to determine the exact value of the initial formation rates. Analogous time-courses were conducted at pD 6.0 and 9.0. Analysis of these time-courses revealed the formation of hm^{3}tmp. The rate of reaction was found to be slowest at pD 6.0 (Figure 4.22; 8.0 x 10^{-7} mol dm^{-3} s^{-1}). The rates of reaction were comparable at pD 7.5 and 9.0 (Figure 4.20). The final concentrations of hm^{3}tmp effectively did not change upon changing pD.

4.4.7.2 Conclusions of the results described in section 4.4.7

In conclusion, the ^{1}H NMR analyses reveals the formation of one closely related adduct when TMP is reacted with excess formaldehyde. The newly developed resonance at δ_H 5.34 ppm is assigned to the methylene protons attached to the *N*-hydroxymethyl group of hm^{3}tmp. The NMR results were further supported by mass spectrometry analysis. The resonances corresponding to cross-linked TMP were not observed. The reaction of TMP with formaldehyde was found to be near-instantaneous. The reaction was observed to be slowest at pD-6.0 (Figure 4.22).

Table 4.4.7.1 Proton and carbon assignment for TMP

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
1	151.8	-	-	-
2	166.6	-	-	-
3	111.7	-	-	-
4	137.5	7.75	s	-
5	11.6	1.86	s	-
6	84.7	6.27	t	7.0
7	38.0	2.36-2.19	m,m	-
8	70.9	4.51	dt	3.0, 6.0
9	71.5	4.18	m	-
10	63.7	4.07	m	-

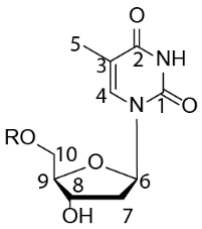
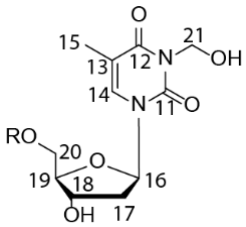


Table 4.4.7.2 Proton and carbon assignment for hm^{3}TMP

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
11	151.7	-	-	-
12	165.0	-	-	-
13	111.1	-	-	-
14	136.1	7.75	s	-
15	12.2	1.87	d	-
16	85.5	6.3	t	7.0
17	38.0	2.36-2.19	m	-
18	70.9	4.51	dt	3.0, 6.0
19	85.71	4.18	m	-
20	63.7	4.18	m	-



21	64.7	5.34	bs	-

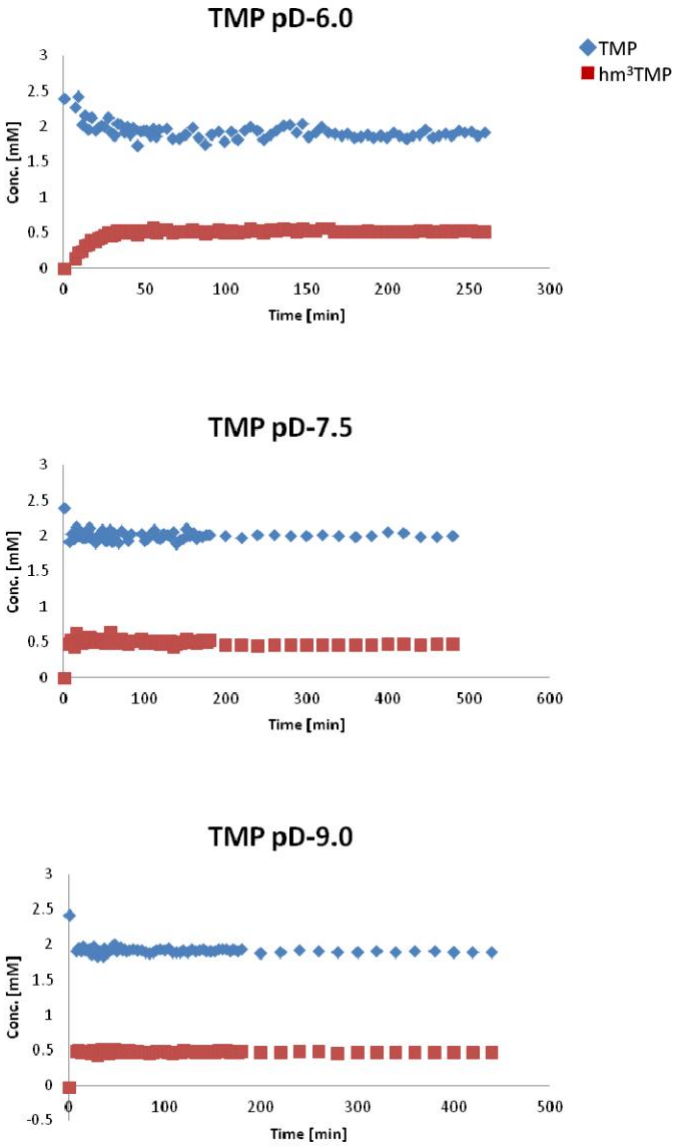


Figure 4.22 Graph showing changes in concentrations of TMP (blue) and hm^{3}TMP (red) over time at 27°C at pD 6.0 (top); pD 7.5 (middle); pD 9.0(bottom). The concentrations are obtained from the integration of the peaks corresponding to resonance of the protons attached to C-1 of the deoxyribose ring of TMP and hm^{3}TMP (marked as 6 and 16, respectively in Figure 4.19) and normalising the intensity value relative to internal standard, TSP. The ^{1}H NMR timecourses were recorded for 8-10hours; every 2 min over the first hour, every 4 min over the next 2 hours and every 20 mins for the next 5-7 hours.

4.4.8 Studies on the reaction of formaldehyde with Uridinemonophosphate (UMP)

4.4.8.1 ^{1}H-NMR analysis on the reaction of formaldehyde with UMP

Initially, a ^{1}H NMR spectrum of a 40 mM solution of UMP was recorded and assigned with the help of 2D NMR (HSQC and HMBC). Seven distinct ^{1}H resonances were observed. All the non-exchangeable protons were observed. The resonances at δ_H 8.05 ppm and δ_H 5.91 ppm were assigned to the protons attached to C-6 and C-5 of the pyrimidine ring (marked as 4 and 3 in Figure 4.23.A) of UMP respectively. The resonances at δ_H 5.92 ppm, δ_H 4.32 ppm, δ_H 4.26 ppm, δ_H 4.16 ppm and δ_H 3.83 ppm-3.96 ppm correspond to the protons attached to C-1, C-2, C-3, C-4 and C-5 (marked as 5, 6, 7, 8 and 9, respectively, in Figure 4.23.A) of the ribose ring of UMP.

A sample containing UMP and excess formaldehyde was then prepared and monitored by ^{1}H-NMR, 2D HSQC, 2D HMBC spectra (Figure 4.23). The proton and the carbon assignments of the UMP and newly formed adduct are given in Tables 4.4.8.(1-2).

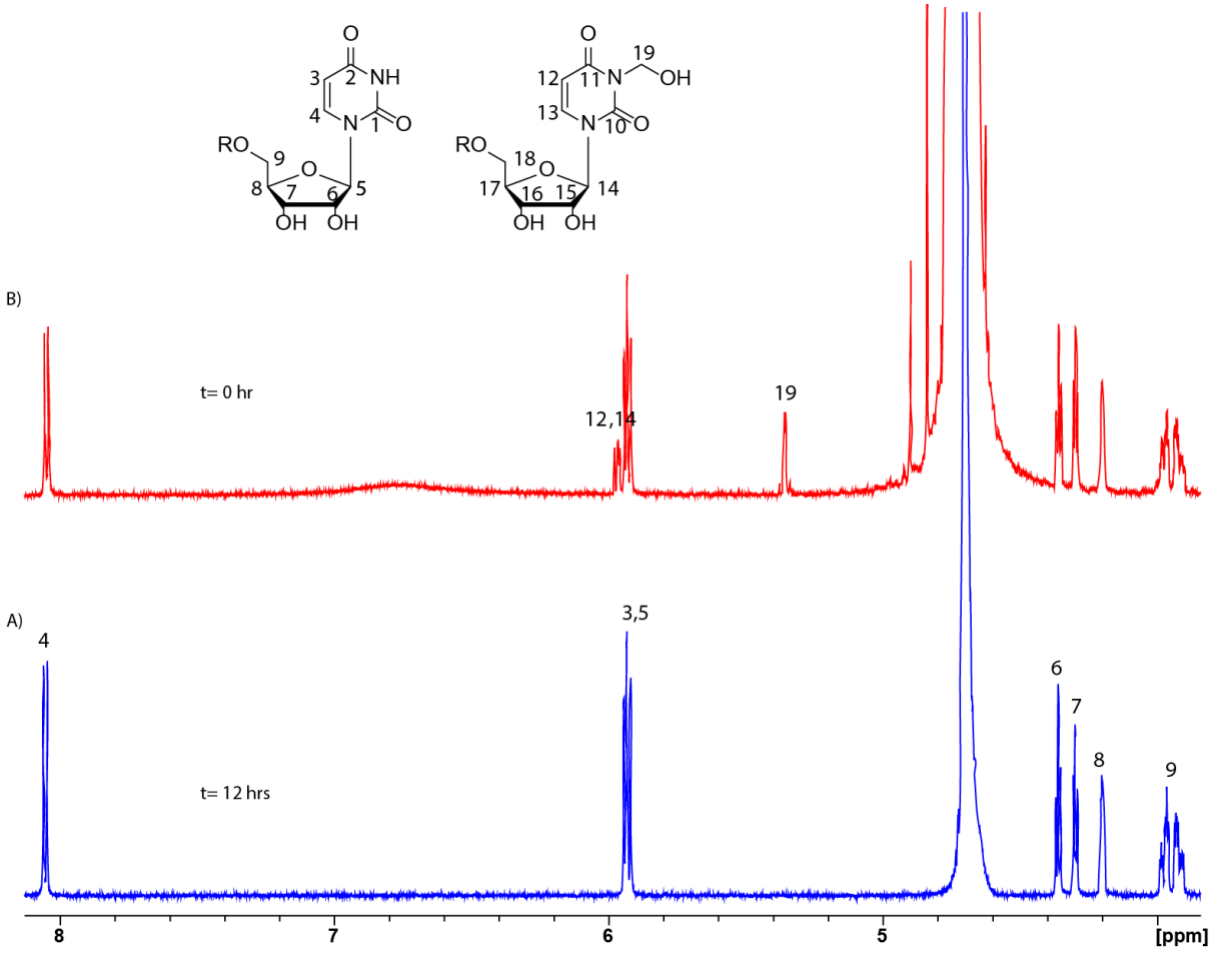


Figure 4.23 ^{1}H-NMR (500 MHz) spectrum of A) 40 mM UMP at 27°C, pD 7.5 in D_2O; B) reaction mixture of 40 mM UMP and excess formaldehyde after 12hours following initial mixing at 27°C, pD-7.5. The ^{1}H assignments of UMP and the newly developed resonances are indicated.

New resonances were observed at δ_H 5.96 ppm and δ_H 5.36 ppm; the resonance at δ_H 5.36 ppm (marked as 19 in Figure 4.23.B) correlates with the carbon at δ_C 64.0 ppm in 2D HSQC and shows correlations with the carbons at δ_C 151.6 ppm and δ_C 164.6 ppm (marked as 10 and 11 in Figure 4.24) in 2D HMBC. Furthermore, the intensity of the resonance at δ_H 5.36 ppm was approximately the same intensity as that of the resonance at δ_H 5.96 ppm (marked as 12 and 14 in Figure 4.23.B). The resonance at δ_H 5.36 ppm was therefore assigned to the methylene protons of the *N*-hydroxymethyl group of N^3-hydroxymethyluridinemonophosphate (hm^3UMP). The resonance at δ_H 5.96 ppm was assigned to the protons attached to C-1 of the deoxyribose ring and C-5 of the pyrimidine ring respectively (marked as 14 and 12 respectively in Figure 4.23.B). No other adducts were

observed during the 12 hours following initial mixing. By NMR analysis, the conversion of UMP to hm^{3}UMP adduct was approximately 24%.

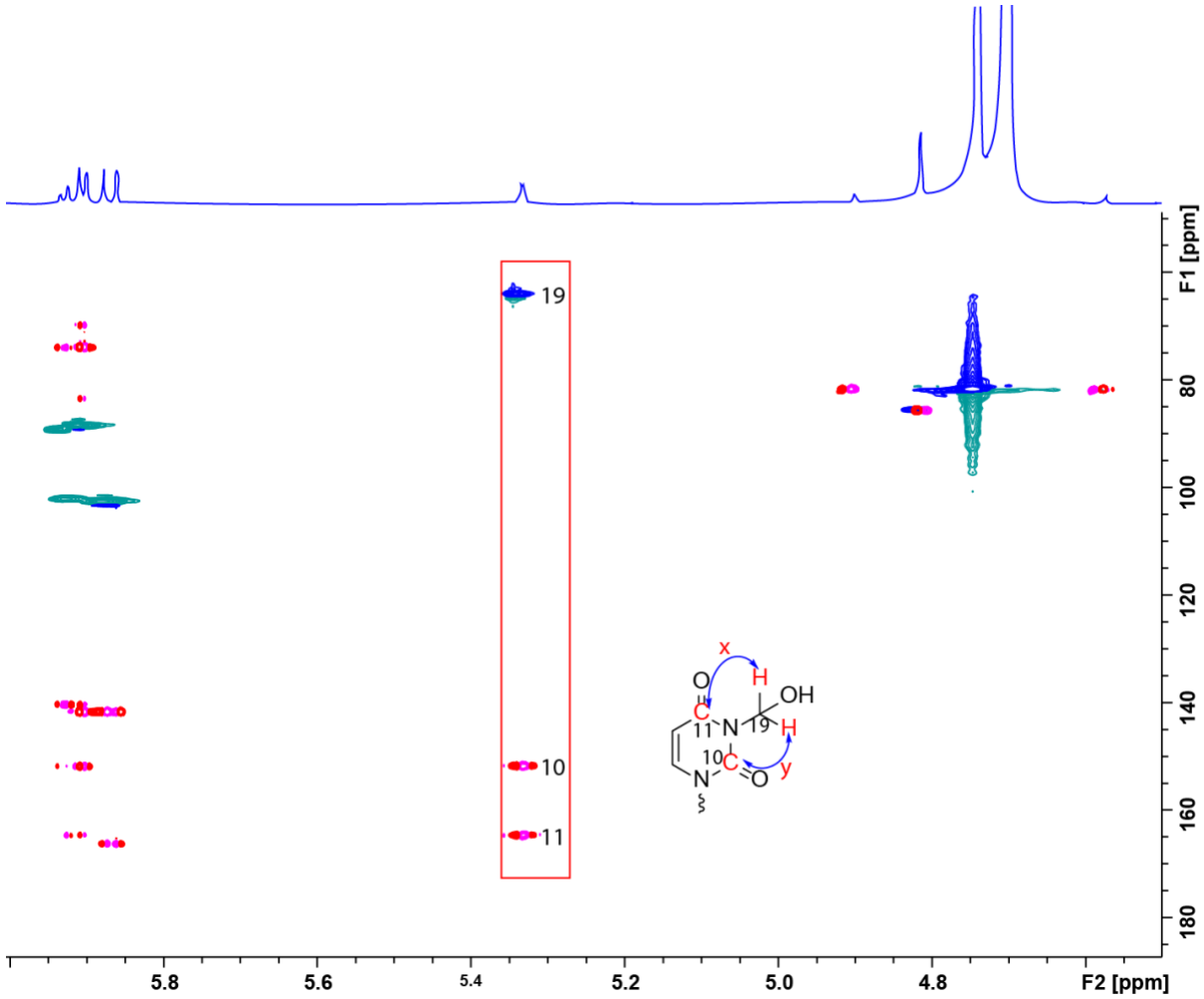


Figure 4.24 Overlaid 2D HSQC (green and blue) and HMBC (red and pink) NMR spectra (500 MHz) for the reaction mixture of a 40 mM UMP and formaldehyde at 25°C. Cross-peaks highlighted in red Box correspond to the correlations (marked as "x" and "y" in red) observed for hm^{3}UMP. Important correlations observed between formaldehyde derived protons (marked in red) and UMP structure are highlighted with blue curved arrows.

The structures of adducts were also assessed by analysing the ESI mass spectrum obtained by LC purification. The EI / MS spectrum showed two peaks in the negative ion mode corresponding to m/z: 323.0288 ([UMP-H]$^-$, calculated: 323.0286) and 353.0390 ([hm^{3}UMP-H]$^-$, calculated: 353.0392).

The effect of changing the pD on the reaction rate was then investigated. A ^{1}H NMR time-course was initially recorded at pD 7.5. 2.4mM of UMP, pre-buffered at pD 7.5, was reacted with a 53-fold excess of formaldehyde, then transferred immediately to a 3mm NMR tube and monitored by ^{1}H-NMR. Surprisingly, the resonance at δ_H 5.36 ppm was observed to develop to a maximum observed intensity within 6 min following initial mixing; the intensity remained constant thereafter. Attempts were made to reduce the time-lapse by recording the ^{1}H-NMR spectrum within 3 minutes following the initial mixing. The resonance at δ_H 5.36 ppm was observed to develop to maximum intensity within 3 min following initial mixing; the intensity remained constant thereafter. The concentration of hm^{3}UMP reached approximately 0.5 mM immediately after addition of formaldehyde.

Analogous time-courses were conducted at pD 6.0 and 9.0. The analysis of these time-courses revealed the formation of hm^{3}UMP. The rate of reaction is slowest at pD 6.0 (Figure 4.26; ~8 x 10^{-7} mol dm^{-3} s^{-1}). Furthermore, the rates of reaction are comparable at pD 7.5 and 9.0 (figure 4.26). The final concentrations of hm^{3}tmp effectively did not change with a change in pD.

4.4.8.2 Conclusion of the results described in section 4.4.8

In conclusion, ^{1}H NMR analyses revealed the formation of a structurally similar adduct when UMP is reacted with excess formaldehyde. The newly developed resonance at δ_H 5.36 ppm was assigned to the methylene protons of the *N*-hydroxymethyl group of hm^{3}UMP. The NMR results were further supported by mass spectrometric analysis. Resonances corresponding to methylene bridge cross-linked UMP were not observed. The reaction of UMP with formaldehyde was found to be near-instantaneous. The reaction was observed to be the slowest at pD-6.0.

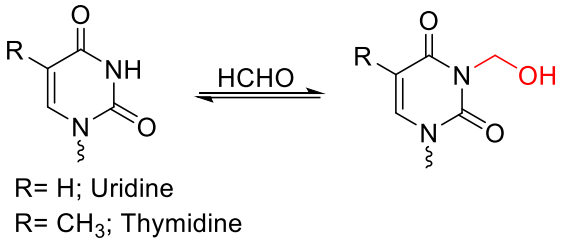


Figure 4.25 Schematic representation of the proposed mechanism for the reaction of UMP/TMP and formaldehyde. The protons and carbon derived from formaldehyde are marked in red.

Table 4.4.8.1 Proton and carbon assignment for UMP

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
1	151.7	-	-	-
2	166.2	-	-	-
3	102.4	5.91	d	8.0
4	141.8	8.05	d	8.0
5	88.2	5.92	d	5.0
6	73.6	4.32	t	5.0
7	69.4	4.26	t	5.0
8	84.1	4.16	m	-
9	60.7	3.83-3.96	m	-

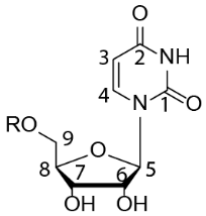
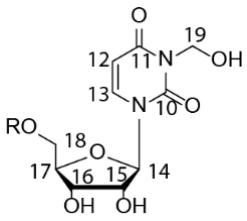


Table 4.4.8.2 Proton and carbon assignment for hm^{3}UMP

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
10	151.6	-	-	-
11	164.6	-	-	-
12	102.1	5.96	d	8.0
13	140.4	8.05	d	8.0
14	88.8	5.96	d	5.0
15	73.7	4.32	t	5.0



16	69.3	4.26	t	5.0
17	84.2	4.16	m	-
18	60.6	3.83-3.96	m	-
19	64.0	5.36	-	-

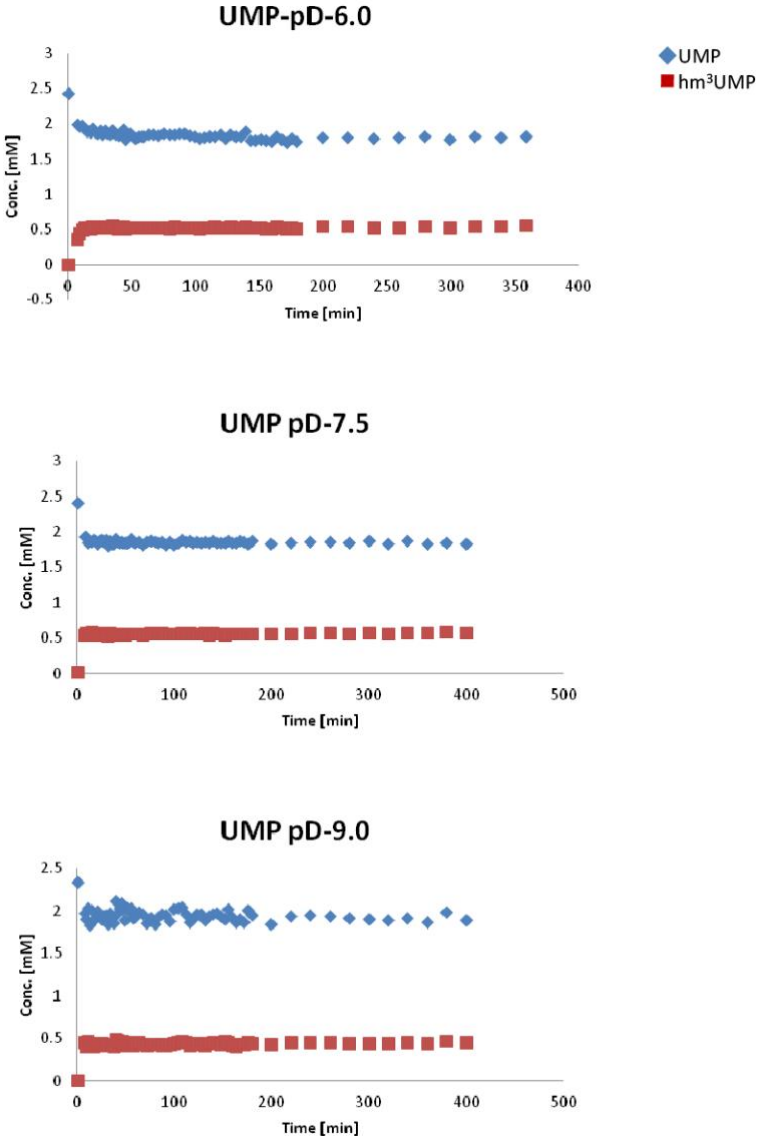


Figure 4.26 Graph showing changes in the concentrations of UMP (blue) and hm^{3}UMP (red), over time at 27°C at pD 6.0 (top); pD 7.5 (middle); pD 9.0 (bottom). The concentrations are obtained from the integration of the peaks corresponding to the resonance of the protons attached to C-6 of the pyrimidine ring and C-1 of the ribose ring of UMP and hm^{3}UMP (marked as 3 and 5, 12 and 14, respectively in Figure 4.23) and normalising the intensity value relative to internal standard, TSP. ^{1}H NMR timecourses were recorded for 8-10hours; every 2 min over the first hour, every 4 min over the next 2 hours and every 20 mins for the next 5-7 hours.

4.4.9 Studies on the reactivity of N^6-methyladenosine (m^6A) with formaldehyde

m^6A in RNA is one of the most prevalent RNA modifications, and was first identified in 1974 (Identification of Methylated Nucleosides in Messenger RNA from Novikoff Hepatoma Cells). m^6A is present in mRNA and long non-coding RNAs. m^6A has emerged as the first example of a reversible modification in RNA, opening a wide field of RNA epigenetics;[56,57] the METTL3/14 and WTAP complexes act as m^6A writers,[58,59] YTHDF1-3 and YTHDC1 act as m^6A readers[60] and FTO and ABH5 act as m^6A erasers,[61–63] respectively. Given that adenosine was observed to react with formaldehyde on the N^6-nitrogen, it was proposed that m^6A might also react with formaldehyde, and also that the additional methyl group might affect reactivity/stability of any adduct(s) formed.

4.4.9.1 ^{1}H-NMR analysis on the reaction of formaldehyde with m^6A

A ^{1}H NMR spectrum of a 20 mM solution of m^6A in D_2O was initially recorded and assigned with the help of recorded 2D-NMR (HSQC, HMBC). Nine distinct ^{1}H resonances were observed. All the non-exchangeable protons were observed. The resonances at δ_H 8.17 ppm and δ_H 8.15 ppm correspond to the protons attached to C-8 and C-2 (marked as 5 and 1, respectively, in Figure 4.27.A) of the purine ring of m^6A. The resonances at δ_H 5.91 ppm, δ_H 4.66 ppm, δ_H 4.32 ppm, δ_H 4.20 ppm, δ_H 3.83 ppm and δ_H 3.74 ppm correspond to the protons attached to C-1, C-2, C-3, C-4, C-5 and C-5 (marked as 7, 8, 9, 10, 11 and 11, respectively, in Figure 4.27.A) of the ribose ring of m^6A. The resonance at δ_H 3.2 ppm corresponds to the protons attached to the N^6-methyl group of m^6A.

A sample containing m^6A and excess formaldehyde was then prepared and monitored by ^{1}H NMR, 2D HSQC and 2D HMBC after 12 hours incubation (Figure 4.27.B). The proton and carbon assignments of the peaks are given in the Tables 4.4.9.(1-2).

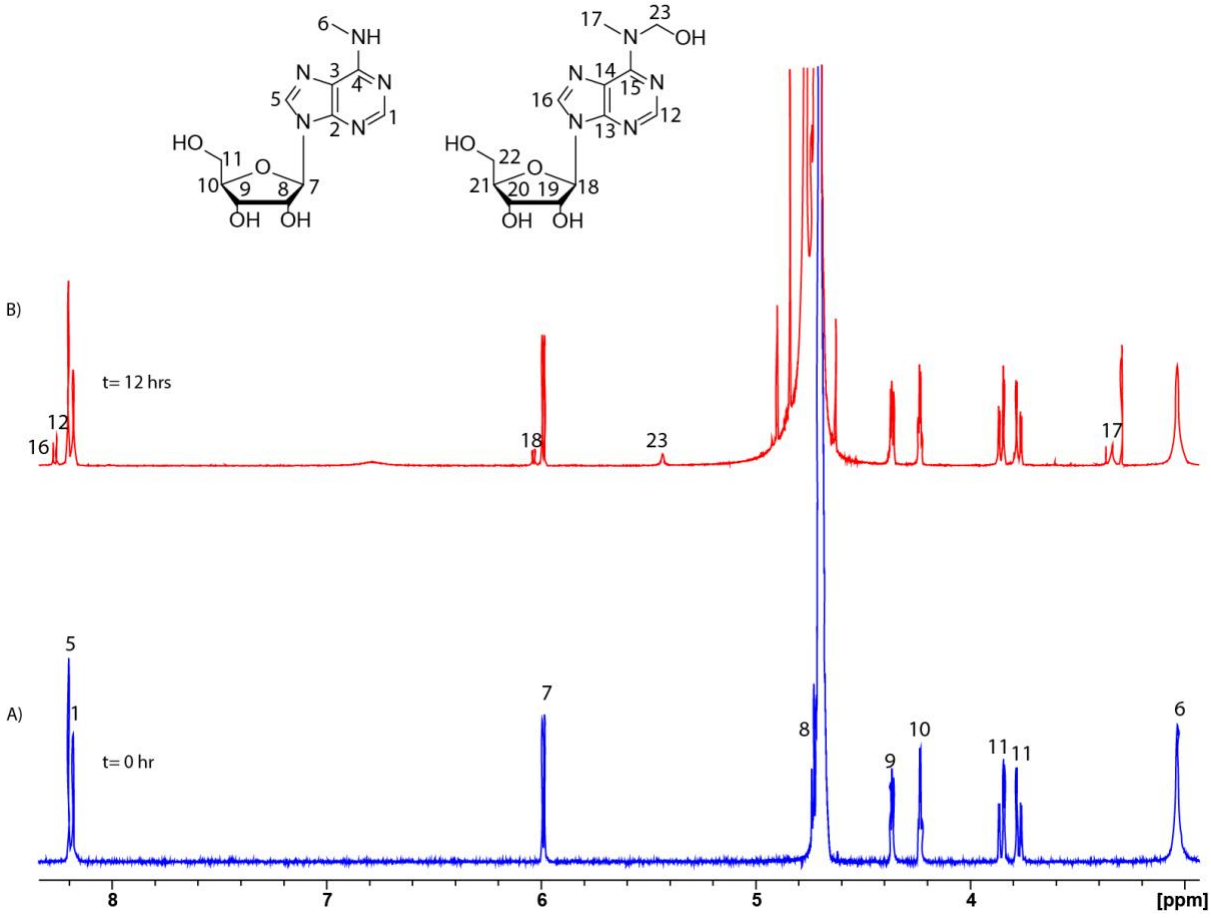


Figure 4.27 ^{1}H-NMR (500 MHz) spectrum of A) 20 mM m^{6}A at 27°C, pD 7.5 in D_2O; B) reaction mixture of 20mM m^{6}A and excess formaldehyde after 12hours following initial mixing at 27°C, pD-7.5. The ^{1}H assignments of m^{6}A and the newly developed resonances are indicated.

New resonances were observed at δ_H 8.27 ppm, δ_H 8.26 ppm, δ_H 6.03 ppm, δ_H 5.44 ppm and δ_H 3.34 ppm. The resonance at δ_H 5.44 ppm (marked as 23 in Figure 4.27.B) correlates to the carbon at δ_C 74.0 ppm in 2D HSQC and shows correlations with the carbons at δ_C 154.5 ppm and δ_C 36.0 ppm (marked as 15 and 17 in Figure 4.26) in 2D HMBC (Figure 4.28). The intensity of the resonance at δ_H 5.44 ppm was found to be approximately twice the intensity of the resonance at δ_H 6.03 ppm (marked as 23 and 18, respectively in Figure 4.27.B). Thus, the resonance at δ_H 5.44 ppm was assigned to the methylene protons of the *N*-hydroxymethyl group of *N*^{6},*N*^{6}-(hydroxymethyl(methyl))adenosine (hm^{6}m^{6}A). The resonances at δ_H 8.27 ppm, δ_H 8.26 ppm, δ_H 6.03 ppm and δ_H 3.34 ppm were assigned to the protons attached to C-8, C-2 of the purine ring, C-1 of the ribose ring and *N*^{6}-methyl group of hm^{6}m^{6}A respectively

(marked as 16, 12, 18 and 17, respectively in figure 4.27.B). No other adducts were observed during the first 12 hours following initial mixing. By NMR analysis, the conversion of m^6A to hm^6m^6A was approximately 12%.

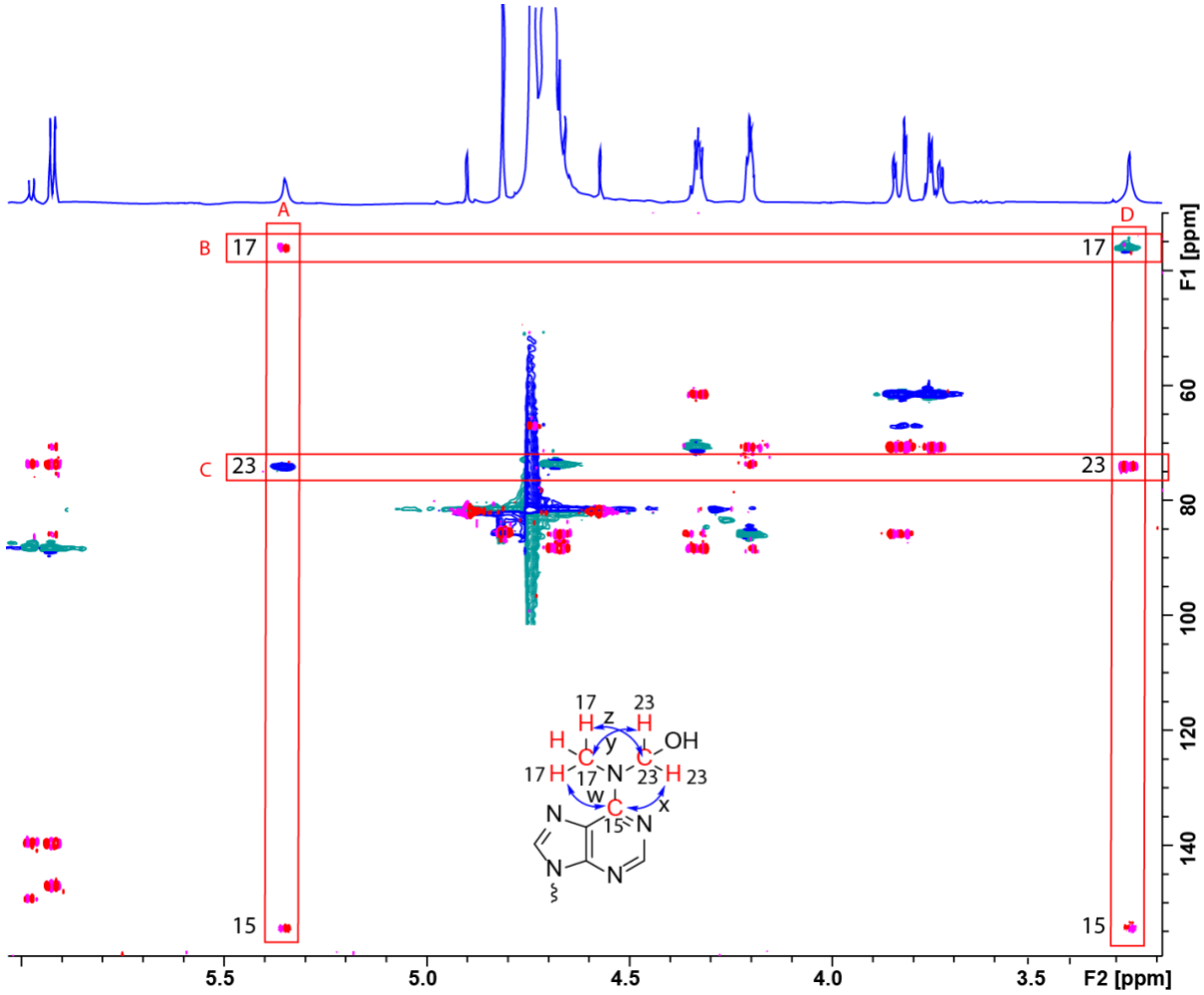


Figure 4.28 Overlaid 2D HSQC (green and blue) and HMBC (red and pink) NMR spectra (500 MHz) for the reaction mixture of a 20 mM m^6A and formaldehyde at 25°C. Cross-peaks highlighted in Box A correspond to correlations (marked as "x" and "y") observed for hm^6m^6A. Cross-peaks highlighted in Box B and Box D correspond to the methyl and methylene carbons (marked as 17 and 23, respectively) of hm^6m^6A. Cross-peaks highlighted in Box D correspond to correlations (marked as "w" and "z") observed for hm^6m^6A. Important correlations observed between formaldehyde derived protons (marked red) and m^6A structure are highlighted with blue curved arrows.

The structure of the adducts was also assessed by analysing the ESI mass spectrum obtained by LC purification. The EI / MS spectrum showed two peaks in the positive ion mode corresponding to the observed m/z: 282.1197 ($[m^6A+H]^+$, calculated: 282.1197) and 312.1302 ($[hm^6m^6A+H]^+$, calculated: 312.1308)

The rate of of hm^{6}m^{6}A formation was then compared to the formation rate of hm^{6}dAMP (formed from reaction of dAMP and formaldehyde). A ^{1}H NMR time-course was initially recorded at pD 7.5; 2.4 mM of m^{6}A, pre-buffered at pD 7.5, was reacted with a 53-fold excess of formaldehyde, then transferred immediately to a 3 mm NMR tube and monitored by ^{1}H-NMR. The resonance at δ_H 5.44 ppm was observed to develop approximately 20 min following initial mixing (i.e. concentration of hm^{6}m^{6}A approximately 20 min following initial mixing was above the NMR detection limit). The concentration of hm^{6}m^{6}A slowly increased with time; however, after 200 min, changes in the concentration of hm^{6}m^{6}A were not observed – the concentration remained constant at approximately 0.3mM thereafter. The initial rate of the reaction, as calculated from graph in Figure 4.27, is 3.0 x 10^{-8} mol dm^{-3} s^{-1} which is much slower than dAMP. Although one might expect the *N*-methyl group to increase the nucleophilicity of the *N*-6 amine, it is possible that the increased steric bulk of the *N*^{6}-methyl group of m^{6}A might result in the observed lower rate of reaction; it is possible that the bulky *N*-hydroxymethyl group forces the methyl group anti to *N*-1 of m^{6}A, leading to a steric clash with *N*-7. Further, addition of a methyl group at *N*-6 subtly changes the tautomeric preference of adenosine, leading to more sp^2 character at *N*-6.

4.4.9.2 Conclusion of the results described in section 4.4.9

In conclusion, ^{1}H NMR analyses have revealed the formation of a novel adduct when m^{6}A is reacted with excess formaldehyde. The newly developed resonance at δ_H 5.44 ppm was assigned to the methylene protons of the *N*-hydroxymethyl group of this new adduct, hm^{6}m^{6}A. The NMR analyses were supported by mass spectrometric analysis. The resonances corresponding to *N*^{6}-methylene bridge cross-linked m^{6}A were not observed. The reaction of m^{6}A with formaldehyde was found to be slower than adenosine/dAMP. The initial rate of the reaction was 3.0 x 10^{-8} mol dm^{-3} s^{-1}. The additional methyl group on *N*-6 of m^{6}A possibly contributes to the steric hindrance on the formaldehyde reacting site. The restriction in

rotation of the exocyclic amine owing to the presence of additional methyl group on m^6A might also contribute to the observed lower initial rate of reaction as compared to the initial rate of reaction of dAMP/adenosine with formaldehyde.

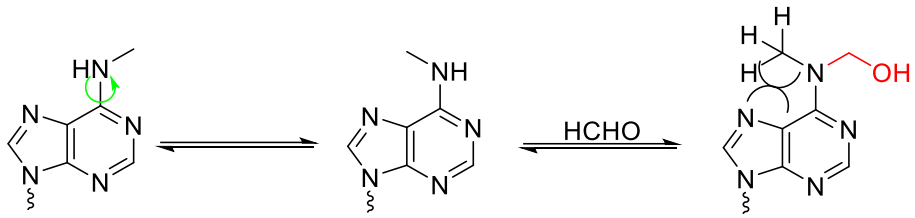


Figure4.29 Schematic representation of the proposed mechanism for the reaction of m^6A and formaldehyde. The protons and carbon derived from formaldehyde are marked in red.

Table 4.4.9.1 Proton and carbon assignments for m^6A

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
1	152.6	8.15	s	-
2	147.0	-	-	-
3	119.8	-	-	-
4	154.9	-	-	-
5	139.84	8.17	s	-
6	27.1	3.2	b.s	-
7	88.21	5.91	d	6.0
8	73.2	4.66	t	5.5
9	70.6	4.32	dd	4.0,5.0
10	85.6	4.20	q	3.0
11	61.2	3.83,3.74	dd,dd	3.0,3.5,13.0
11	27	3.2	bs	-

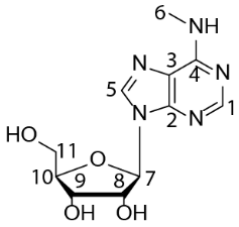
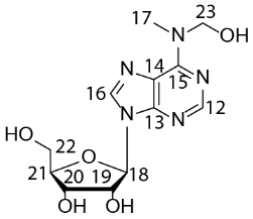


Table 4.4.9.2 Proton and carbon assignments for hm^6m^6A

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
12	152.07	8.26	s	-
13	149.3	-	-	-
14	120.0	-	-	-
15	154.5	-	-	-
16	137.75	8.27	s	-
17	36.0	3.34	b.s	-
18	88.11	6.03	d	6.0
19	73.2	4.66	t	5.5
20	70.6	4.32	dd	4.0,5.0
21	85.6	4.20	d	3.0
22	61.2	3.83,3.74	dd,dd	3.0,3.5,13.0
23	74.0	5.44	bs	-



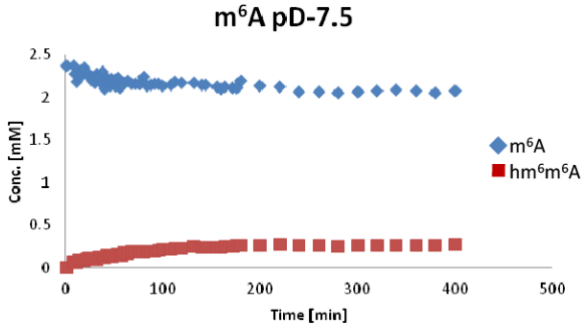


Figure 4.30 Graph showing changes in concentrations of m^6A (blue) and hm^6m^6A (red), over time at 27 °C at pD 7.5. The concentrations are obtained from the integration of the peaks corresponding to the resonance of the proton attached to C-1 of the ribose ring of m^6A and hm^6m^6A (marked as 7 and 18, respectively in Figure 4.27) and normalising the intensity value relative to internal standard, TSP. ^{1}H NMR timecourses were recorded for 8-10hours; every 2 min over the first hour, every 4 min over the next 2 hours and every 20 mins for the next 5-7 hours.

4.4.10 Studies on the reaction of N^1-methyldeoxyadenosine (m^1dA) with formaldehyde

After investigating the reaction of formaldehyde with m^6A, work focussed on investigating the reaction of other modified form of Adenosine, m^1dA, present in RNA with formaldehyde.

4.4.10.1 ^{1}H-NMR analysis on the reaction of formaldehyde with m^1dA

Initially, a ^{1}H-NMR spectrum was recorded on a 20 mM m^1dA and assigned with the help of recorded 2D HSQC, 2D HMBC. Seven distinct ^{1}H resonances were observed to develop. All the non-exchangeable protons were observed. The resonance at δ_H 8.55 ppm and δ_H 8.53 ppm correspond to protons attached to C-8 and C-2 (marked as 5 and 1 in Figure 4.31.A) of the purine ring of m^1dA. The resonances at δ_H 6.16 ppm, δ_H 4.79 ppm, δ_H 4.45 ppm, δ_H 4.28 ppm, δ_H 3.92 ppm and δ_H 3.85 ppm correspond to the protons attached to C-1, C-2, C-3, C-4, C-5 and C-5 (marked as 7, 8, 9, 10, 11 and 11, respectively in Figure 4.31.A) of the deoxyribose ring of m^1dA. The resonance at δ_H 3.94 ppm corresponds to the protons attached to N^1-methyl group of m^1dA.

Excess formaldehyde was added to the above sample and the reaction was allowed to attain equilibrium. ^{1}H-NMR, 2D HSQC, 2D HMBC spectra were recorded on the sample. Proton and carbon assignments of m^1AMP are given in Tables 4.4.10.(1-2).

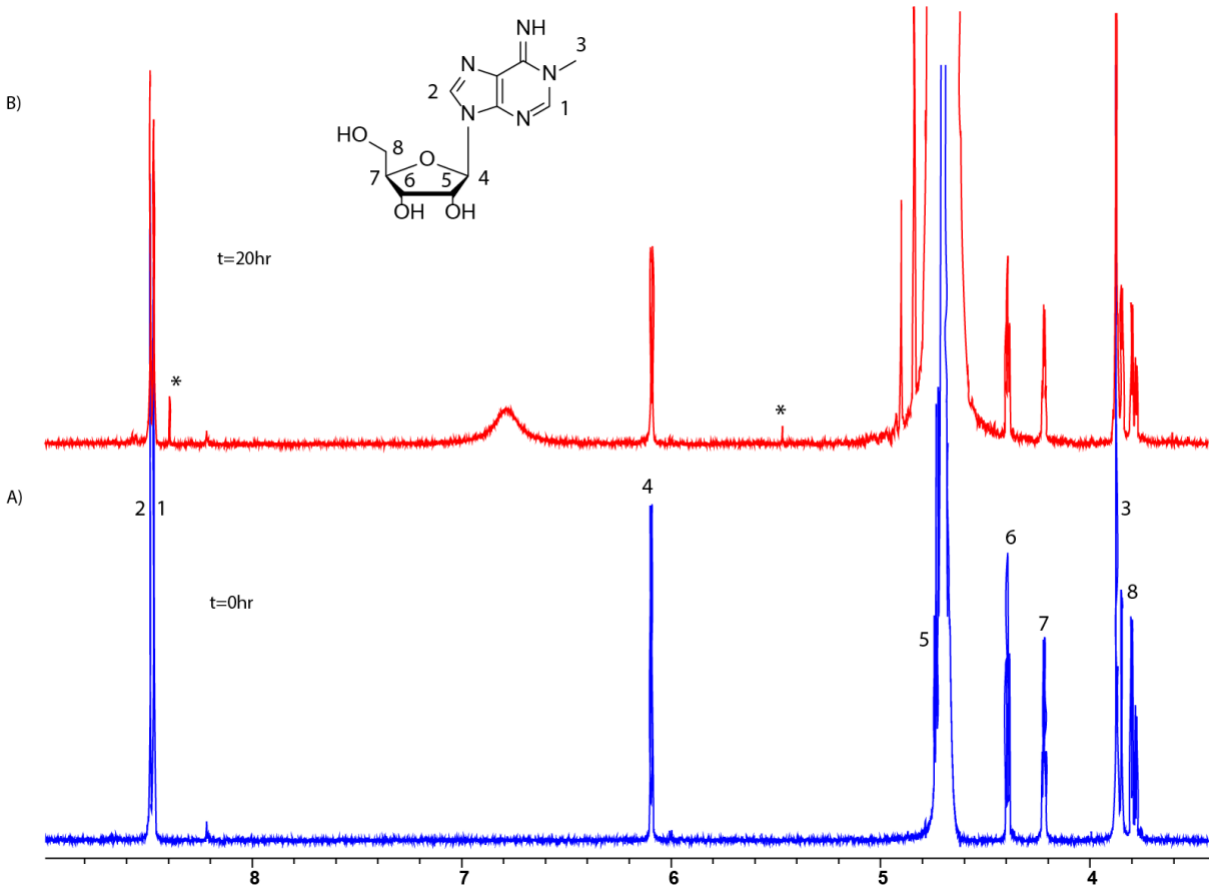
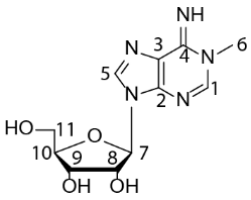


Figure 4.31 ^{1}H-NMR (600 MHz) spectrum of A) 20 mM m^{1}A at 27°C, pD 7.5 in D_2O; B reaction mixture of 20 mM m^{6}A and excess formaldehyde after 20 hours following initial mixing at 27°C, pD-7.5. The ^{1}H assignments of m^{1}A are indicated and newly developed resonances are marked by asterisk "".*

New resonances were observed to develop at δ_H 8.39 ppm, δ_H 5.46 ppm (marked by* in Figure 4.31). However, the data was not conclusive in determining the structure of the adduct, if formed at all. As evident from the figure 4.31, new resonances developed at low level.

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
1	147.8	8.53	s	-
2	146.3	-	-	-
3	119.0	-	-	-
4	150.6	-	-	-
5	143.0	8.55	s	-
6	37.8	3.94	s	-
7	89.4	6.16	d	5.5
8	74.0	4.79	t	5.0
9	69.8	4.45	t	4.5
10	84.0	4.28	q	4.0
11	61.1	3.92, 3.85	dd, dd	3.0,4.5,13.0



4.4.11 Studies on the reaction of formaldehyde with m^{6}G

m^{6}G modification, which is produced by methylation of guanosine, is reported to be the most carcinogenic DNA modification.[64,65] m^{6}G is mutagenic because of its ability to polymerise[66,67] and mis-insert T, instead of C.[68–70] Thus, it was of interest to investigate the reaction of formaldehyde with O^6-methylguanosine (m^{6}G) and compare the reactivity of m^{6}G with formaldehyde to that observed with guanosine/dGMP.

4.4.11.1 ^{1}H-NMR analysis on the reaction of formaldehyde with m^{6}G

Initially, ^{1}H NMR spectrum of a 20 mM m^{6}G was recorded and assigned with the help of 2D HSQC and 2D HMBC. Seven distinct ^{1}H resonances were observed. All the non-exchangeable protons were observed. The resonances at δ_H 8.08 ppm and δ_H 4.09 ppm were assigned to the protons attached to C-8 and the *O*^{6}-methyl group (marked as 5 and 6 in Figure 4.33.A) of the purine ring of m^{6}G respectively. The resonances at δ_H 5.96 ppm, δ_H 4.7 ppm,

δ_H 4.42 ppm, δ_H 4.26 ppm, δ_H 3.8 ppm and δ_H 3.74 ppm were assigned to the protons attached to C-1, C-2, C-3, C-4, C-5 and C-5 (marked as 7, 8, 9, 10, 11 and 11, respectively in Figure 4.33.A) of the ribose ring of m^6G.

The reaction of m^6G and excess formaldehyde was then monitored by ^{1}H-NMR, 2D HSQC, 2D HMBC (Figure 4.33 and 4.34). The proton and carbon assignments of m^6G and the newly formed adducts are given in Tables 4.4.11.(1-3).

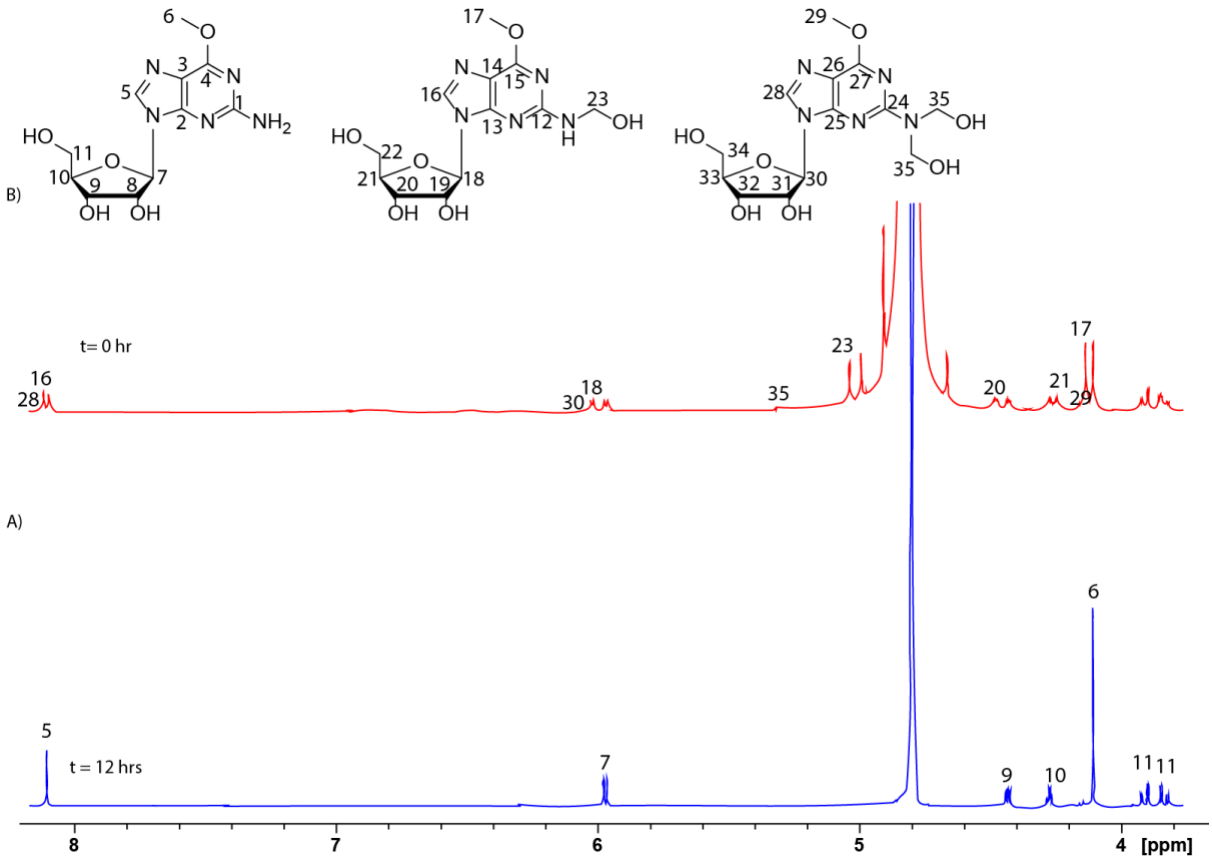


Figure 4.33 A) ^{1}H-NMR (600 MHz) spectrum of a 20 mM m^6G at 27°C, pD 7.5 in D_2O. The ^{1}H assignments of m^6G are indicated. B) ^{1}H-NMR (600MHz) spectrum of the reaction mixture for 20mM m^6G and excess formaldehyde after 12 hours following initial mixing at 27°C, pD-7.5. The ^{1}H assignments of the newly developed resonances are indicated.

New resonances were observed at δ_H 8.16 ppm, δ_H 8.11 ppm, δ_H 6.08 ppm, δ_H 6.01 ppm, δ_H 5.3 ppm, δ_H 5.03 ppm, δ_H 4.47 ppm, δ_H 4.24 ppm, δ_H 4.15 ppm and δ_H 4.12 ppm (Figure 4.33.B). The resonance at δ_H 5.03 ppm (marked as 23 in Figure 4.33.B) correlates with the carbon at δ_C 65.4 ppm in 2D HSQC and shows correlation with the carbon at δ_C 158.5 ppm

(marked as 12 in Figure 4.33.B) in 2D HMBC (Figure 4.34). The intensity of the resonance at δ_H 5.03 ppm was approximately twice the intensity of the resonance at δ_H 8.11 ppm (marked as 16 in Figure 4.33.B). Thus, the resonance at δ_H 5.03 ppm was assigned to the methylene protons of the *N*-hydroxymethyl group of N^2,O^6-hydroxymethyl(methyl)guanosine (hm^2m^6G). The resonances at δ_H 8.11 ppm and δ_H 4.12 ppm were assigned to the protons attached to C-8 and O^6-methyl group of the purine ring (marked as 16 and 17 in Figure 4.33.B) of hm^2m^6G. The resonances at δ_H 6.01 ppm, δ_H 4.47 ppm and δ_H 4.24 ppm were assigned to the protons attached to C-1, C-3 and C-4 of the ribose ring (marked as 18, 20 and 21, respectively, in Figure 4.33.B) of hm^2m^6G.

The resonance at δ_H 5.3 ppm (marked as 35 in Figure 4.33.B) correlates with the carbon at δ_C 71.8 ppm in 2D HSQC and shows correlations with its own carbon and the carbon at δ_C 157.2 ppm (marked as 24 in Figure 4.33.B) in 2D HMBC, indicating symmetry in the structure (Figure 4.34). The intensity of the resonance at δ_H 5.3 ppm was approximately four times the intensity of the resonance at δ_H 8.16 ppm (marked as 28 in Figure 4.33.B). Consequently, the resonance at δ_H 5.3 ppm was assigned to the methylene protons attached to the two hydroxymethyl groups of N^2,N^2,O^6-dihydroxymethyl(methyl)guanosine (dhm^2m^6G). The resonances at δ_H 8.16 ppm, δ_H 6.08 ppm and δ_H 4.15 ppm were assigned to the protons attached to C-8 of the purine ring, C-1 of the ribose ring and O^6-methyl group (marked as 28, 30 and 29, respectively, in Figure 4.33.B) of dhm^2m^6G. This adduct is not reported in literature. No other adducts were observed during the 12 hours following initial mixing. By NMR analysis, the conversion of m^6G to hm^2m^6G adduct was approximately 57 % and to dhm^2m^6G adduct was approximately 5 % after 12 hours incubation. The conversion of m^6G to hm^2m^6G was more prevalent than conversion of G (dGMP) to hm^2G (hm^2dGMP).

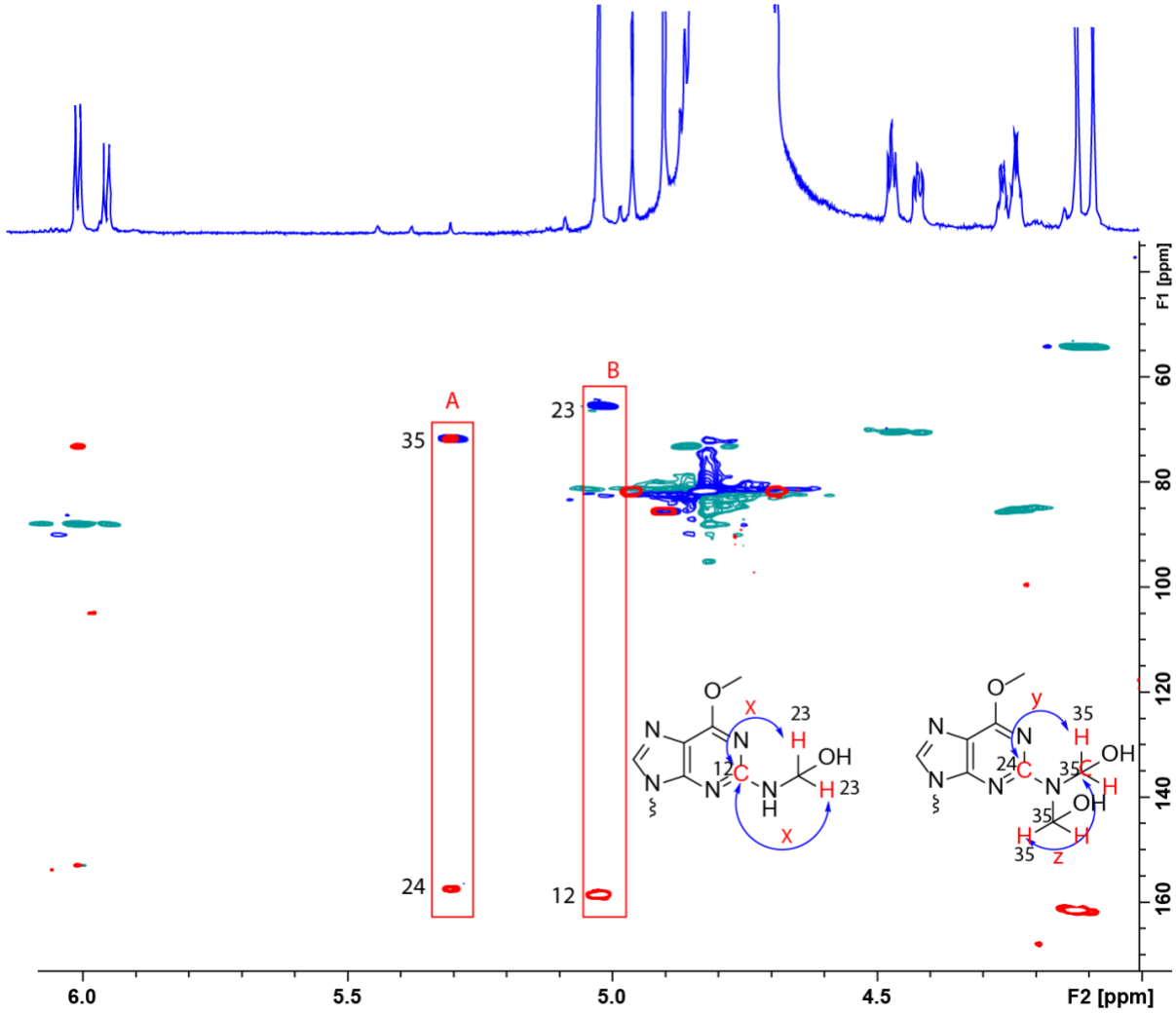


Figure 4.34 Overlaid 2D HSQC (green and blue) and HMBC (red and pink) NMR spectra (500 MHz) for the reaction mixture of 20 mM solution of m^6G and formaldehyde at 25°C. Cross-peaks highlighted in Box A correspond to the correlations (marked as "y" and "z" in red) observed for dhm^2m^6G. Cross-peaks highlighted in Box B correspond to the correlations (marked as "x" in red) observed for hm^2m^6G. Important correlations observed between formaldehyde derived protons (highlighted in red) and m^6G structure are highlighted with blue curved arrows.

Given the cross-linked product of dGMP and formaldehyde (i.e. the symmetric dGMP dimer connected via a methylene bridge) was observed after prolonged incubation (one week), it was hoped that m^6G might also form an analogous cross-linked product with formaldehyde. A ^{1}H-NMR spectrum was recorded on a sample of m^6G and formaldehyde after a one week incubation; however, no cross-linked m^6G aminal was observed. Conversion of m^6G to hm^2m^6G after a week was found to be approximately 72 % and to dhm^2m^6G was

approximately 13 %. It should be noted that this is the highest observed conversion to products of all the nucleosides studied.

The structures of adducts were also assessed by analysing their ESI mass spectra after separation by liquid chromatography. The EI / MS spectrum showed three peaks in the positive ion mode corresponding to m/z: 300.1301 ($[m^6G + H]^+$, calculated:300.1302), 328.1251 ($[hm^2m^6G + H]^+$, calculated: 328.1252) and 358.1357 ($[dhm^2m^6G + H]^+$, calculated: 358.1357).

Table 4.4.11.1 Proton and carbon assignment for m^6G

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
1		-	-	-
2	152.76	-	-	-
3	114.6	-	-	-
4	161.77	-	-	-
5	139.3	8.08	s	-
6	54.41	4.09	s	-
7	88.04	5.96	d	6.0
8	70.3	4.7	t	5.0
9	70.73	4.42	dd	3.0, 3.0,5.0
10	85.81	4.26	q	3.0
11	61.6	3.8, 3.74	dd, dd	3.0, 3.5, 13.0

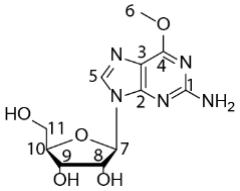


Table 4.4.11.2: Proton and carbon assignment for hm^2m^6G

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
12	158.5	-	-	-
13	153.11	-	-	-
14	114.66	-	-	-
15	161.57	-	-	-
16	139.6	8.11	s	-
17	54.38	4.12	s	-
18	88.08	6.01	d	6.0
19	70.3	4.7	t	5.0
20	70.67	4.47	dd	3.0, 3.0, 5.0
21	85.23	4.24	q	3.0
22	61.6	3.8, 3.74	dd, dd	3.0, 3.5,13.0
23	65.4	5.03	b.s	-

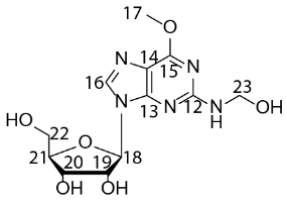
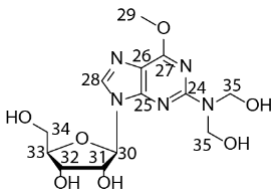


Table 4.4.11.3 Proton and carbon assignment for dhm^2m^6G

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
24	157.2	-	-	-
25	153.11	-	-	-
26	115.4	-	-	-
27	161.6	-	-	-
28	139.68	8.16	s	-
29	54.8	4.15	s	-
30	87.94	6.08	d	6.0
31	70.3	4.7	t	5.0



32	70.07	4.47	dd	3.0, 3.0, 5.0
33	85.72	4.24	q	3.0
34	61.56	3.8, 3.74	dd, dd	3.0, 3.5,13.0
35	71.8	5.3	b.s	-

The effect of the O^6-methyl group on the initial rate of the reaction was then investigated. A ^{1}H NMR time-course was initially recorded at pD-7.5. 2.4 mM of m^{6}G, pre-buffered at pD-7.5, was reacted with a 53-fold excess of formaldehyde and transferred immediately to a 3mm NMR tube. The resonance at δ_H 5.03 ppm was observed to develop approximately 10 min after initial mixing (i.e. the concentration of hm^{2}m^{6}G after 10 min was above the NMR detection limit). The resonance at δ_H 5.3 ppm was observed to develop after approximately 90 min after mixing. The concentration of dhm^{2}m^{6}G remained low during the time-course. The concentration of hm^{2}m^{6}G slowly increased with time and reached approximately 1.2 mM after 800 min. The rate of the reaction was 5.7 x 10^{-8} mol dm^{-3} s^{-1}.

4.4.11.2 Conclusion of the results described in section 4.4.11

In conclusion, ^{1}H-NMR analyses have revealed the formation of two closely related adducts when m^{6}G is reacted with an excess of formaldehyde. The newly developed resonance at δ_H 5.03 ppm is assigned to the methylene protons of the *N*-hydroxymethyl group of hm^2 m^{6}G and the resonance at δ_H 5.3 ppm is assigned to the methylene protons attached to the two hydroxymethyl groups of dhm^{2}dGMP (Figure 4.33.B). The resonances corresponding to methylene bridge cross-linked m^{6}G were not observed. Further, reaction at the *N*-1 position of m^{6}G was not observed. The NMR results were supported by mass spectrometric analyses. The initial reaction rate of formation of dhm^{2}dGMP as calculated from the graph was 5.7 x 10^{-8} mol dm^{-3} s^{-1} at pD-7.5, 27°C (Figure 4.35).

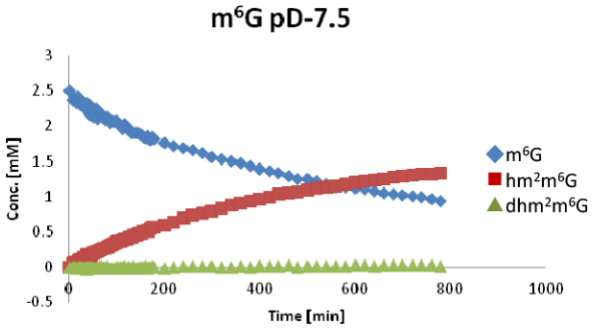


Figure 4.35 Graph showing changes in concentrations of m^6G (blue), hm^2m^6G (red), dhm^2m^6G (green) over time at 27°C at pD 7.5. The concentrations are obtained from the integration of the peaks corresponding to resonance of protons attached to C-1 of the ribose ring of m^6G, hm^2m^6G and dhm^2m^6G (marked as 7, 18 and 30, respectively in Figure 4.33) and normalising the intensity value relative to internal standard, TSP. The ^1H NMR timecourses were recorded for 8-10hours; every 2 min over the first hour, every 4 min over the next 2 hours and every 20 mins for the next 5-7 hours.

4.4.12 Studies on the reaction of formaldehyde with inosine (I)

Deamination of a nucleotide includes hydrolytic loss of exocyclic amines; in the case of adenosine, the product of deamination is inosine. The conversion of the exocyclic amino group to a keto (carbonyl) group changes the H-bonding properties of the nucleotides from hydrogen bond donor to hydrogen bond acceptor. Inosine may also occur in DNA/RNA by deamination of adenosine.[72,73]Given the structural similarity of inosine and guanosine, the former is often erroneously read as guanosine by DNA polymerases which can also lead to incorporation of cytidine in place of thymidine as a consequence of incorrect templating.[71] Although, inosine is mutagenic in DNA, it is naturally present in RNA. In RNA, inosine is an essential modification introduced by specific deaminases.[71] Work was therefore focussed on investigating the reaction(s) of formaldehyde with inosine.

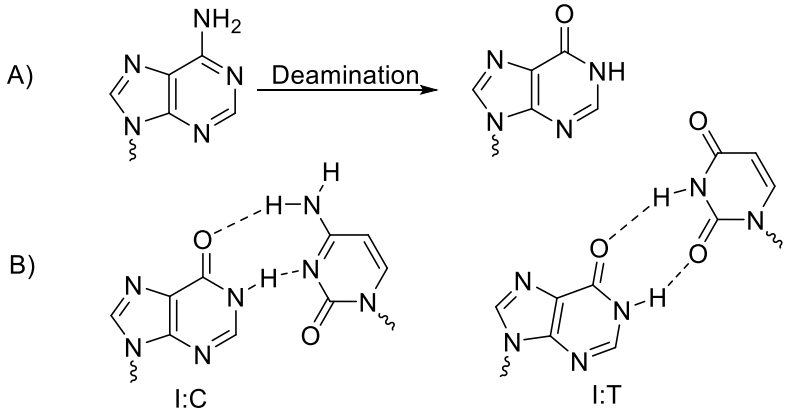


Figure 4.36 A) Schematic representation of the hydrolytic deamination of adenosine to inosine. B) Schematic representation of the base pairing of I with C and T. H-bonding are shown with dotted lines.

4.4.12.1 ^{1}H-NMR analysis on the reaction of formaldehyde with inosine

Initially, a ^{1}H NMR spectrum of a 20 mM inosine solution in D_2O was recorded and assigned with the help of recorded 2D NMR (HSQC and HMBC). Seven distinct ^{1}H resonances were observed. All the non-exchangeable protons were observed. The resonances at δ_H 8.34 ppm and δ_H 8.23 ppm correspond to the protons attached to C-8 and C-2 (marked as 5 and 1 in Figure 4.37.A) of the purine ring of inosine. The resonances at δ_H 6.1 ppm, δ_H 4.78 ppm, δ_H 4.44 ppm, δ_H 4.28 ppm, δ_H 3.9 ppm and δ_H 3.84 ppm correspond to the protons attached to C-1, C-2 C-3, C-4, C-5 and C-5 (marked as 6, 7, 8, 9, 10 and 10, respectively, in Figure 4.37.A) of the inosine ribose ring.

A sample was then prepared containing inosine (2.4 mM) and excess formaldehyde, and was then monitored by ^{1}H-NMR, 2D HSQC and 2D HMBC (Figure 4.37.B). The proton and carbon assignments of inosine and an identified adduct are given in Table 4.4.12.(1-2).

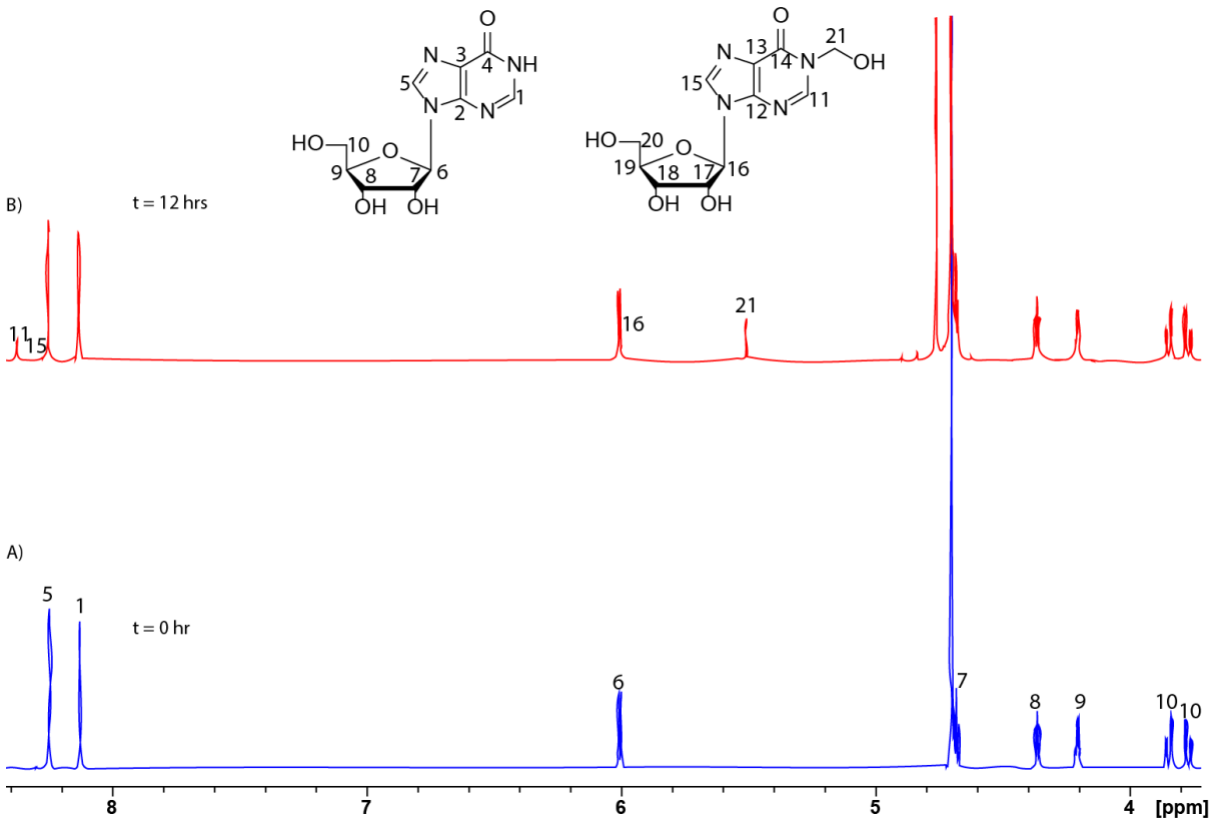


Figure 4.37 A) ^{1}H-NMR (500 MHz) spectrum of a 20 mM inosine at 27°C, pD 7.5 in D_2O. The ^{1}H assignments of inosine are indicated. B) ^{1}H-NMR (500MHz) spectrum for the reaction mixture of 20 mM inosine and excess formaldehyde after 12 hrs following initial mixing at 27°C, pD-7.5. The ^{1}H assignments of the newly developed resonances are indicated.

New resonances were observed to develop at δ_H 8.46 ppm, δ_H 8.37 ppm, δ_H 6.09 ppm and δ_H 5.6 ppm, implying formation of a novel adduct with a similar structure to inosine. The resonance at δ_H 5.6 ppm (marked as 21 in Figure 4.37.B) correlates to the carbon at δ_C 69.4 ppm in 2D HSQC and shows correlation with the carbons at δ_C 157.69 ppm and δ_C 148.4 ppm (marked as 14 and 11 in Figure 4.37.B) in 2D HMBC (Figure 4.38). The intensity of the resonance at δ_H 5.6 ppm was approximately twice the intensity of the resonance at δ_H 8.46 ppm (marked as 11 in Figure 4.37.B). Thus, the resonance at δ_H 5.6 ppm was assigned to the methylene protons of the *N*-hydroxymethyl group of N^1-hydroxymethylinosine (hm^{1}I). The resonances at δ_H 8.46 ppm, δ_H 8.37 ppm and δ_H 6.09 ppm were assigned to the protons attached to C-2, C-8 of the purine ring and C-1 of the ribose ring (marked as 11, 15 and 16,

respectively in Figure 4.37.B) of hm^2I respectively. No other adducts were observed during 12 hours following initial mixing. Conversion of I to hm^1I was approximately 20%.

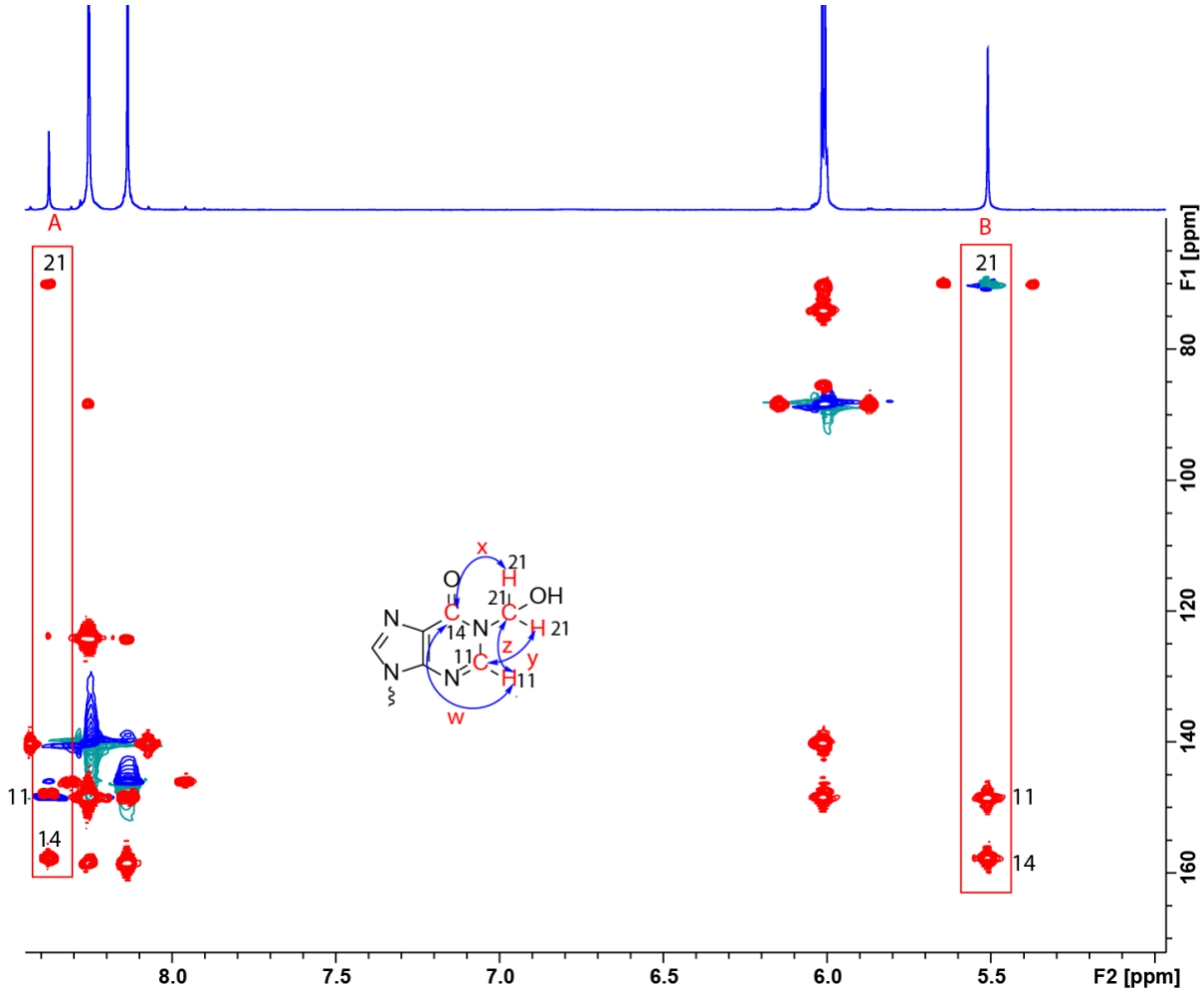


Figure 4.38 Overlaid 2D HSQC (green and blue) and 2D HMBC (red and pink) NMR spectra (600 MHz) for the reaction mixture of a 20 mM inosine and formaldehyde at 25°C. Cross-peaks highlighted in Box A correspond to the correlations (marked as "w" and "z") and Box B correspond to the correlations (marked as "x" and "y") observed for hm^1I. Important correlation observed between formaldehyde derived protons and inosine structure are highlighted with blue curved arrows.

The structure of the adducts were also assessed by analysing the ESI mass spectrum obtained by LC purification. The EI / MS spectrum showed two peaks in the positive ion mode corresponding to m/z: 269.2365 ($[I+H]^+$, calculated: 269.2365) and 299.2595 ($[hm^1I+H]^+$, calculated: 299.2625).

Table 4.4.12.1 Proton and carbon assignments for I

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
1	146.1	8.23	s	-
2	148.4	-	-	-
3	124.22	-	-	-
4	158.55	-	-	-
5	140.3	8.34	s	-
6	88.4	6.1	d	6.0
7	73.9	4.78	t	5.5
8	85.6	4.44	t	4.5
9	61.42	4.28	q	3.5
10	69.4	3.9, 3.84	dd, dd	3.0, 4.0,13.0

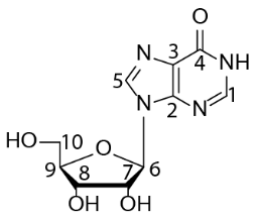
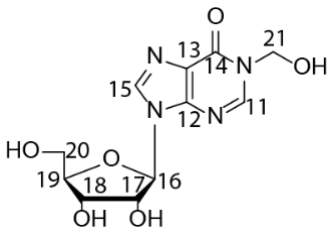


Table 4.4.12.2 Proton and carbon assignment for hm^{1}I

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
11	148.4	8.46	s	-
12	147.89		-	-
13	123.72		-	-
14	157.69		-	-
15	140.1	8.37	s	-
16	87.28	6.09	d	6.0
17	73.9	4.78	t	5.5
18	70.3	4.44	t	4.5
19	85.6	4.28	q	3.5
20	61.42	3.9, 3.84	dd, dd	3.0,4.0,13.0
21	69.4	5.6	bs	-



Analogous to the previous studies, a ^{1}H-NMR time course was recorded at pD-7.5. 2.4 mM of I, pre-buffered at 7.5, was reacted with a 53-fold excess of formaldehyde, then the reaction mixture was transferred immediately to a 3mm NMR match tube. The resonance at δ_H 5.6 ppm was observed to develop to a maximum intensity of 0.5 mM approximately 5 min following initial mixing (i.e. at the first time-point); the concentration of hm^{1}I remained constant, thereafter.

4.4.12.2 Conclusion of the results described in section 4.4.12

In conclusion, the ^{1}H-NMR analyses revealed the formation of a novel adduct when I is allowed to react with an excess of formaldehyde. The newly developed resonance at δ_H 5.6 ppm is assigned to the methylene protons of the *N*-hydroxymethyl group of hm^{1}I. The resonances corresponding to methylene bridge cross-linked product of I were not observed. The NMR results were supported by mass spectrometric analysis.

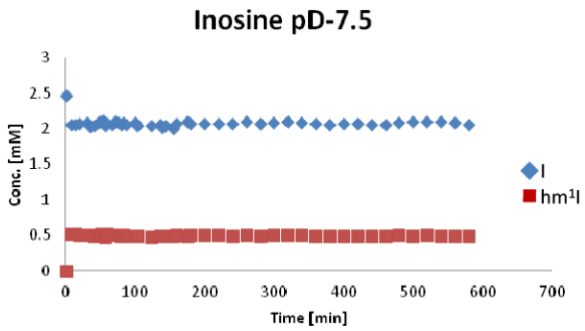


Figure 4.39 Graph showing changes in concentrations of I (blue) and hm^{1}I (red) over time at 27°C at pD 7.5. The concentrations are obtained from the integration of the peaks corresponding to the resonance of the proton attached to C-2 of aromatic ring of I and hm^{1}I (marked as 1 and 11, respectively in Figure 4.37) and normalising the intensity value relative to internal standard, TSP. The ^{1}H NMR time-courses were recorded for 8-10hours; every 2 min over the first hour, every 4 min over the next 2 hours and every 20 mins for the next 5-7 hours.

4.4.13 Studies on the reaction of formaldehyde with 5-hydroxymethylcytidine (hm^5C)

The next non-canonical nucleoside studied was hm^5dC, which is produced by the oxidation of m^5C, as catalysed by the three TET enzymes (TET1-3).[74] hm^5C has been termed the 6th base of DNA; hm^5dC is abundant in embryonic stem cells and in the brain. Reduced hm^5dC levels are reported to impair cell-renewal.[75] hm^5C also plays a role in epigenetic regulation. hm^5C levels in cancer cells are reported to be strongly reduced compared to healthy cells.[76] Among the studied nucleotides containing exocyclic amines, dCMP was found to be the most reactive with formaldehyde. It was therefore interesting to determine whether the additional hydroxymethyl group attached to C-5 might affect this reactivity.

4.4.13.1 ^{1}H-NMR analysis on the reaction of formaldehyde with hm^5dC

Initially, a ^{1}H NMR spectrum of a 20 mM hm^5dC solution in D_2O was recorded and assigned with the help of 2D NMR (HSQC, HMBC). Nine distinct ^{1}H resonances were observed. All the non-exchangeable protons were observed. The resonances at δ_H 7.90 ppm and δ_H 4.46 ppm were assigned to the protons attached to C-6 and the C-5 hydroxymethyl group (marked as 4 and 5, respectively, in Figure 4.40.A) of hm^5dC. The resonances at δ_H 6.27 ppm, δ_H 4.44 ppm, δ_H 4.06 ppm, δ_H 3.86 ppm, δ_H 3.77 ppm, δ_H 2.44 ppm and δ_H 2.32 ppm were assigned to the protons attached to C-1, C-3, C-4, C-5, C-5, C-2 and C-2 (marked as 7, 9, 10, 11, 11, 8 and 8, respectively, in Figure 4.40.A) of the deoxyribose ring of hm^5dC. A sample of hm^5dC (2.4 mM) and excess formaldehyde was prepared and monitored by ^{1}H-NMR, 2D HSQC and 2D HMBC (Figure 4.40). The proton and carbon assignments of the hm^5dC and the newly formed adduct are given in Table 4.4.13.(1-3).

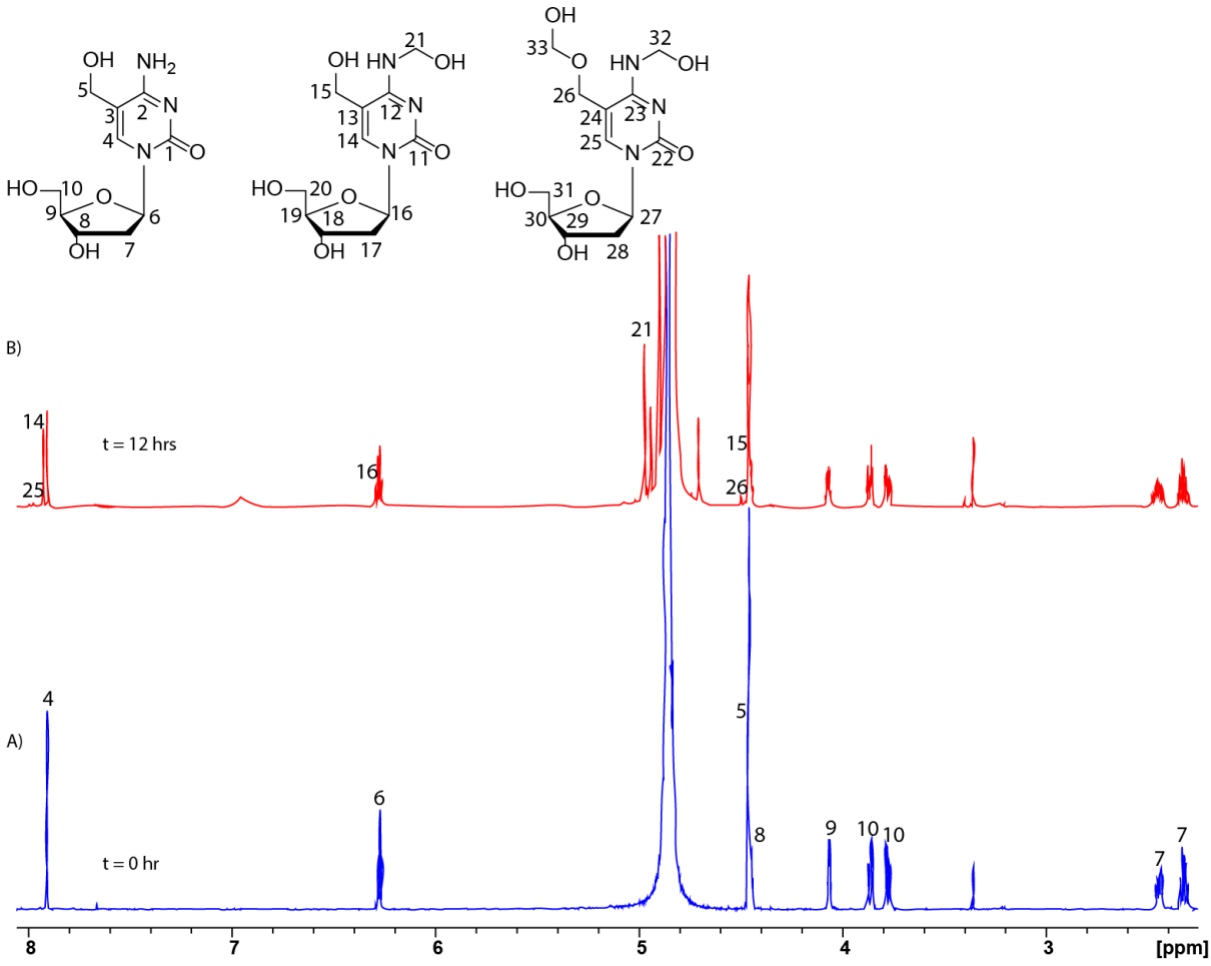


Figure 4.40 ^{1}H-NMR (700 MHz) spectrum of A) 20 mM hm^{5}dC at 27°C, pD 7.5 in D_2O; B) reaction mixture of 20mM hm^{5}dC and excess formaldehyde after 12 hours following initial mixing at 27°C, pD-7.5. The ^{1}H assignments of hm^{5}dC and.the newly developed resonances are indicated.

New resonances were observed at δ_H 7.99 ppm, δ_H 7.92 ppm, δ_H 6.30 ppm, δ_H 6.28 ppm, δ_H 4.97 ppm, δ_H 4.84 ppm, δ_H 4.49 ppm and δ_H 4.465 ppm. The ^{1}H- NMR spectrum revealed the formation of two closely related products. The resonance at δ_H 4.97 ppm (marked as 21 in Figure 4.40.B) correlates with the carbon at δ_C 63.68 ppm in 2D HSQC and shows correlation with the carbon at δ_C 163.36 ppm (marked as 12 in Figure 4.40.B) in 2D HMBC (Figure 4.41). The intensity of the resonance at δ_H 4.97 ppm was found to be approximately twice the intensity of the resonance at δ_H 7.92 ppm (marked as 14 in Figure 4.40.B). Thus, the resonance at δ_H 4.97 ppm was assigned to the methylene protons of the *N*-hydroxymethyl group of *N*^{4},*C*^{5}-di(hydroxymethyl)deoxycytidine(hm^{4}hm^{5}dC). The resonances at δ_H 7.92 ppm and δ_H 4.465 ppm were assigned to the protons attached to the C-6 and C-5 hydroxymethyl

group of the pyrimidine ring (marked as 14 and 15 in Figure 4.40.B) of hm^4hm^5dC. The resonance at δ_H 6.28 ppm was assigned to the protons attached to C-1 (marked as 16 in Figure 4.40.B) of the deoxyribose ring of hm^4hm^5dC. The resonance at δ_H 7.99 ppm (marked as 25 in Figure 4.40.B) shows similar carbon correlations in 2D HMBC with that of the proton attached to the C-6 (marked as 14 in Figure 4.40.B) of the pyrimidine ring of hm^4hm^5dC, indicating formation of similar adduct to that of hm^4hm^5dC. However, the resonance at δ_H 4.84 ppm shows correlation with the carbon at δ_C 87.3 ppm in 2D HSQC (Figure 4.41). This correlation at δ_C 87.3 ppm was found to resonate close to the correlations observed for formaldehyde peak at δ_C 81.5 ppm in 2D HSQC. Further, the peak at δ_H 4.84 ppm overlapped with the carbon shift of poly(oxymethylene)hemiformal at δ_C 85.3 ppm in 2D-HSQC, indicating the presence of a similar $–CH_2-O-$ unit in the newly developed entity (Figure 4.41). Further, the resonance at δ_H 4.84 ppm shows correlation to the protons of the hydroxymethyl group attached to C-5 (marked as 26 in Figure 4.40.A) of the pyrimidine ring, in 2D HMBC (Figure 4.41). Examination of the correlations observed in 2D HSQC and 2D HMBC led to the assignment of the resonance at δ_H 4.84 ppm to the methylene protons of the *O*-hydroxymethyl group attached to C-5 hydroxymethyl group of N^4,C^5,C^5-di(hydroxymethyl)(3'hydroxy-2'-oxomethyl)deoxycytidine. The resonances at δ_H 7.99 ppm, δ_H 6.30 ppm and δ_H 4.49 ppm were assigned to the protons attached to C-6 of the pyrimidine ring, C-1 of the deoxyribose ring and C-5 hydroxymethyl group (marked as 25, 27 and 26 respectively in Figure 4.40.B) of

N^4,C^5-(hydroxymethyl)(hydroxymethyloxymethyl)deoxycytidine (Figure 4.41).

N^4,C^5-(hydroxymethyl)(hydroxymethyloxymethyl)deoxycytidine is a novel compound and has not been previously reported. The conversion of hm^5dC to hm^4hm^5dC was approximately 62% and to N^4,C^5-(hydroxymethyl)(hydroxymethyloxymethyl)deoxycytidine was approximately 1%.

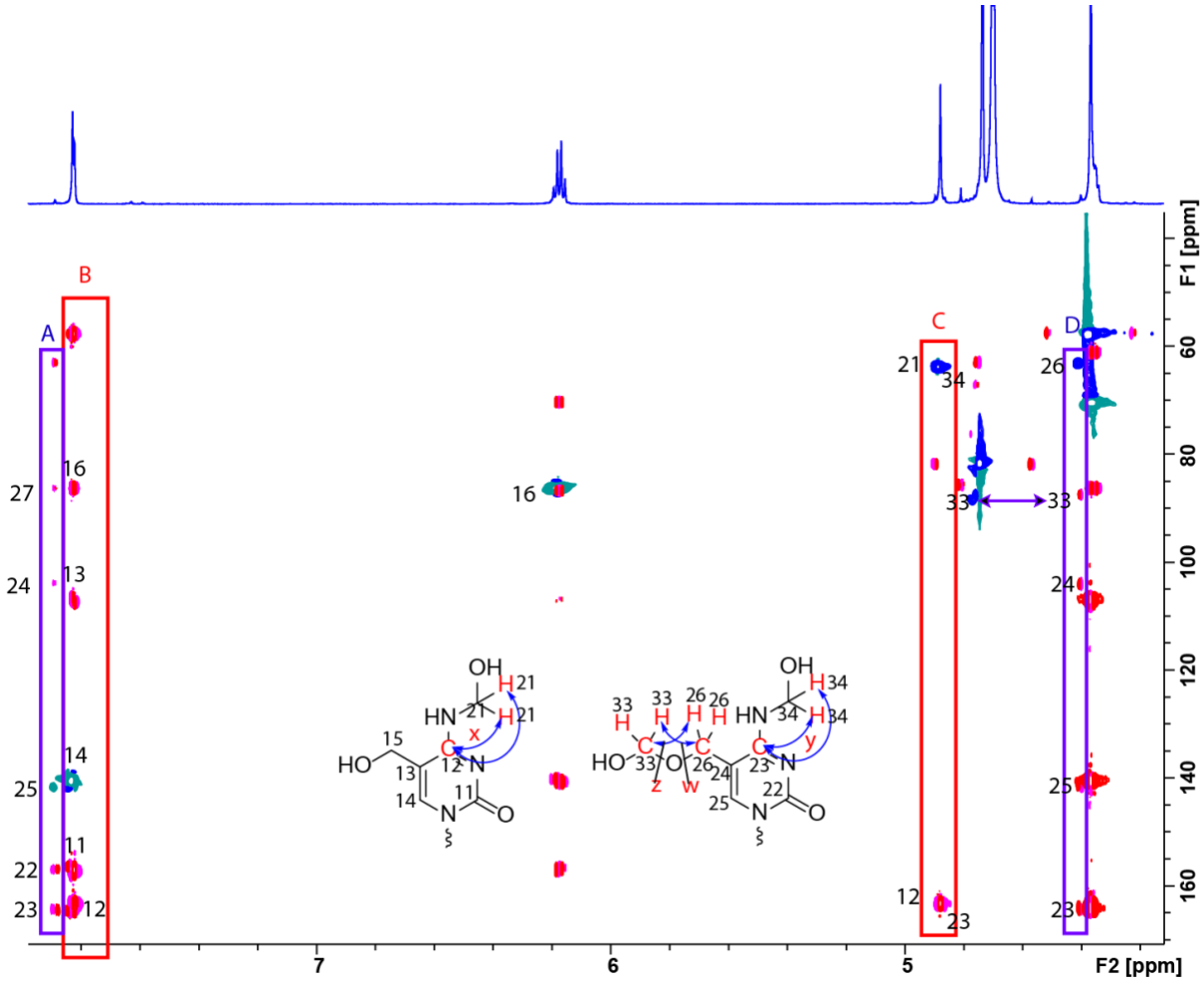


Figure 4.41 Overlaid 2D HSQC (green and blue) and 2D HMBC (red and pink) NMR spectra (500 MHz) for the reaction mixture of a 20 mM solution of hm^{5}dC and formaldehyde at 25°C. Cross-peaks highlighted in Box A (purple box) correspond to the correlations observed for N^4,C^5-(hydroxymethyl)(3'hydroxy-2'-oxoethyl)deoxycytidine. Cross-peaks highlighted in Box B correspond to the correlations observed for hm^{4}hm^{5}dC. Cross-peaks highlighted in Box C correspond to the correlations (marked as "x" in red) observed for hm^{4}hm^{5}dC and correlation (marked as "y" in red) observed for N^4,C^5-(hydroxymethyl)(hydroxymethyloxymethyl)deoxycytidine. Cross-peaks highlighted in Box D (purple box) correspond to the correlations (marked as "z" in red) observed for N^4,C^5-(hydroxymethyl)(3'hydroxy-2'-oxoethyl)deoxycytidine. Important correlations observed between formaldehyde derived protons (marked as red) and hm^{5}dC structure are highlighted with blue curved arrows.

The adduct structures were also assessed by analysing their ESI mass spectra, which were obtained during purification by liquid chromatography. The EI / MS spectra showed three peaks in the positive ion mode corresponding to m/z: 258.1090 ([hm^{5}dC +H]$^+$, calculated:258.1084), 288.1191 ([hm^{4}hm^{5}dC +H]$^+$, calculated: 288.1190) and 340.1117

([N^4,C^5-(hydroxymethyl)(hydroxymethyloxymethyl)deoxycytidine +Na]$^+$, calculated: 340.1115).

Table 4.4.13.1 Proton and carbon assignments for hm^{5}dC

Position	δ ^{13}C (ppm)	δ ^{13}H(ppm)	Multiplicity	$J_{H,H}$ (Hz)
1	156.85	-	-	-
2	163.36	-	-	-
3	107.28	-	-	-
4	141.17	7.90	s	-
5	57.6	4.460	s	-
6	85.69	6.27	t	6.5
7	39.5	2.44, 2.32	m, m	-
8	70.4	4.44	m	-
9	86.8	4.06	q	4.0
10	61.4	3.86,3.77	dd, dd	3.5, 5.0, 13.0

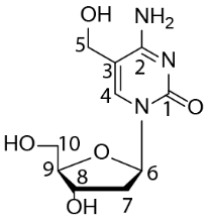
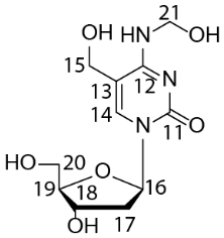


Table 4.4.13.2 Proton and carbon assignments for hm^{4}hm^{5}dC

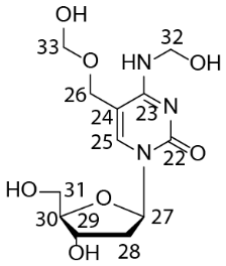
Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
11	156.8	-	-	-
12	163.4	-	-	-
13	107.1	-	-	-
14	141.0	7.92	s	-
15	57.4	4.465	s	-
16	86.6	6.28	t	6.5
17	39.5	2.44, 2.32	m, m	-
18	70.4	4.44	m	-



19	86.8	4.07	q	4.0
20	61.4	3.86,3.77	dd, dd	3.5, 5.0, 13.0
21	63.7	4.97	b.s	-

Table 4.4.13.3 Proton and carbon assignment of N^4,C^5(hydroxymethyl)(hydroxymethyloxymethyl)deoxycytidine.

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
22	156.8	-	-	-
23	164.4	-	-	-
24	103.6	-	-	-
25	141.6	7.99	s	-
26	63.1	4.49	s	-
27	86.6	6.30	t	6.5
28	39.5	2.44, 2.32	m, m	-
29	70.4	4.44	m	-
30	86.8	4.07	q	4.0
31	61.4	3.86,3.77	dd, dd	3.5, 5.0, 13.0
32	63.6	4.90	b.s	-
33	87.3	4.84	s	-



Analogous to the previous studies, a ^{1}H-NMR time-course was recorded at pD-7.5. 2.4 mM of hm^{5}dC, pre-buffered at pD-7.5, was reacted with a 53-fold excess of formaldehyde, then the reaction mixture was transferred immediately to a 3mm NMR match tube and monitored by ^{1}H-NMR. The resonance at δ_H 4.97 ppm (assigned to the methylene protons of hm^{4}hm^{5}dC) was observed to develop approximately 7 min following initial mixing (i.e. at the first NMR time-point). The resonance at δ_H 4.84 ppm (assigned to the *O*-hydroxymethyl group attached to C-5 hydroxymethyl group of

N^4,*C*^{5}-(hydroxymethyl)(hydroxymethyloxymethyl)deoxycytidine) developed after 120 min. The concentration of hm^{5}dC reached to approximately 1.2 mM after 100 min; the concentration remained constant afterwards. However, the resonances corresponding to *N*,*N*-dihydroxyadduct and methylene bridge hm^{5}dC cross-linked product via *N*-4 were not observed. The initial rate of the reaction was 1.6 x10^{-7} mol dm^{-3} s^{-1} at pD-7.5.

4.4.13.2 Conclusion of the results described in section 4.4.13

In conclusion, the ^{1}H-NMR analyses revealed the formation of two closely related products when hm^{5}dC is allowed to react with excess of formaldehyde. The newly developed resonance at δ_H 4.97 ppm was assigned to the methylene protons of the *N*-hydroxymethyl group of hm^{4}hm^{5}dC. Investigation of the 2D HSQC and 2D HMBC revealed the formation of a novel entity, *N*^{4},*C*^{5}-(hydroxymethyl)(hydroxymethyloxymethyl)deoxycytidine, with the methylene protons attached to the *N*-hydroxymethyl group resonating at δ_H 4.90 ppm. The resonances corresponding to a methylene bridged cross-linked hm^{5}dC via *N*-4, or a *N*,*N*-dihydroxymethyl adduct of hm^{5}dC were not observed. The NMR results were supported by mass spectrometric analyses. The initial rates of formation of major product was 1.6 x10^{-7} mol dm^{-3} s^{-1} at pD-7.5, 27°C (Figure 4.42).

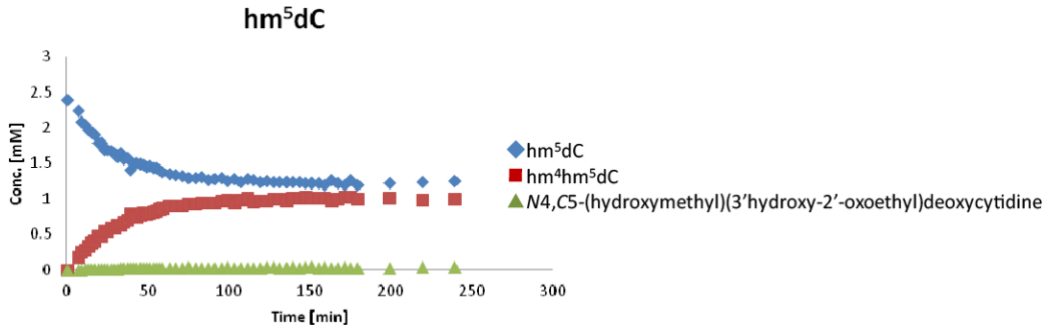


Figure 4.42 Graph showing changes in concentrations of hm^{5}dC (blue), hm^{4}hm^{5}dC (red) and N^4,C^5-(hydroxymethyl)(3'hydroxy-2'-oxoethyl)deoxycytidine (green) over time at 27°C at pD 7.5. The concentrations are obtained from the integration of the peaks corresponding to the resonance of the proton attached to C-6 of the pyrimidine ring of hm^{5}dC, hm^{4}hm^{5}dC and N^4,C^5-(hydroxymethyl)(hydroxymethyloxymethyl)deoxycytidine (marked as 4, 14 and 25, respectively, in Figure 4.40) and normalising the intensity value relative to internal standard, TSP. ^{1}H NMR timecourses were recorded for 8-10 hours; every 2 min over the first hour, every 4 min over the next 2 hourrs and every 20 mins for the next 5-7 hours.

4.4.14 Studies on the reaction of formaldehyde with 5-methyldeoxycytidine (m^5dC)

Methylation at the fifth position of the pyrimidine ring of cytidine (forming m^5dC) is considered as a key epigenetic marker.[77–79] m^5dC is often considered the 5th base of DNA and was the only known modified DNA base before the discovery of hm^5dC. Since, the reaction of formaldehyde with hm^5dC revealed formation of novel adducts, it was deemed interesting to investigate the reaction of m^5dC under similar conditions in order to compare the reactivity of these modified bases with cytidine.

4.4.14.1 ^{1}H-NMR analysis on the reaction of formaldehyde with m^5dC

Initially, a ^{1}H NMR of a 20 mM m^5dC solution in D_2O was recorded and assigned with the help of 2D NMR (HSQC and HMBC). Nine distinct ^{1}H resonances were observed. Each resonance was assigned with the help of 2D NMR (HSQC and HMBC). All the non-exchangeable protons were observed. The resonances at δ_H 7.67 ppm and δ_H 1.98 ppm were assigned to the protons attached to C-6 and C-5 methyl group (marked as 4 and 5, respectively, in Figure 4.43.A) of the pyrimidine ring of m^5dC. The resonances at δ_H 6.30 ppm, δ_H 4.475 ppm, δ_H 4.06 ppm, δ_H 3.86 ppm, δ_H 3.79 ppm, 2.42 ppm and δ_H 2.32 ppm correspond to the protons attached to C-1, C-3, C-4, C-5, C-5, C-2 and C-2 (marked as 7, 9, 10, 11, 11, 8 and 8, respectively, in Figure 4.43.A) of the deoxyribose ring of m^5dC.

A sample containing m^5dC (2.4 mM) was treated with excess formaldehyde and the mixture was monitored by ^{1}H-NMR, 2D-HSQC and 2D-HMBC (Figure 4.43). The proton and carbon assignments of m^5dC and a newly formed adduct are given in Table 4.4.14.(1-2).

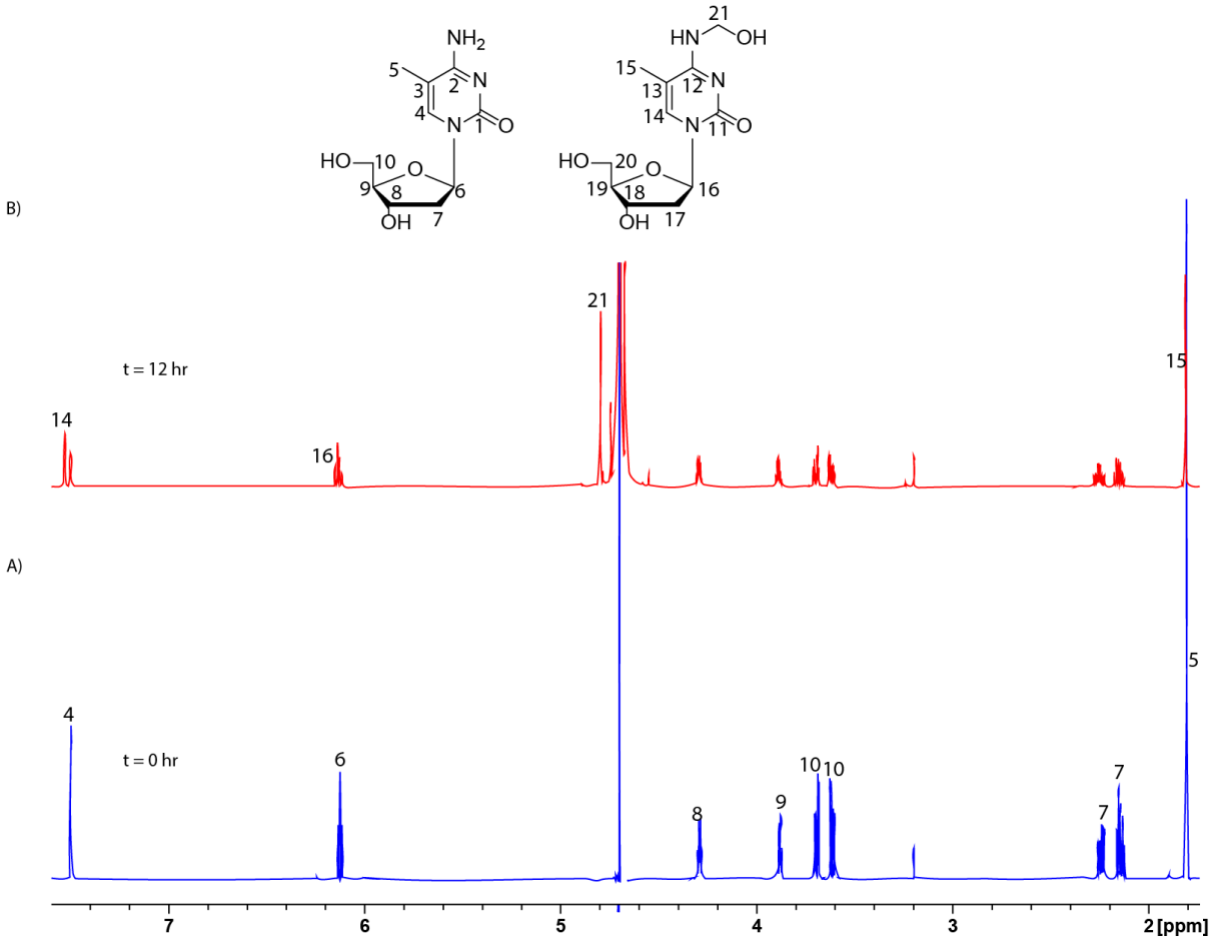


Figure 4.43 ^{1}H-NMR (500 MHz) spectrum of A) 20 mM m^{5}dC at 27°C, pD 7.5 in D_2O; B) reaction mixture of 20 mM m^{5}dC and excess formaldehyde after 12 hours following initial mixing at 27°C, pD-7.5. The ^{1}H assignments of m^{5}dC and the newly developed resonances are indicated.

New resonances were observed at δ_H 7.76 ppm, δ_H 7.70 ppm, δ_H 6.31 ppm, δ_H 4.974 ppm, δ_H 4.89 ppm, δ_H 2.03 ppm and δ_H 1.99 ppm. Investigation of the ^{1}H- NMR spectrum revealed the formation of two closely related products. The resonance at δ_H 4.974 ppm (marked as 21 in Figure 4.43.B) correlates with the carbon at δ_C 63. 8 ppm in 2D HSQC and shows a correlation with the carbon at δ_C 163.97 ppm (marked as 12 in Figure 4.43.B) in 2D HMBC (Figure 4.44). The intensity of the resonance at δ_H 4.974 ppm was approximately twice the intensity of the resonance at δ_H 7.76 ppm (marked as 14 in Figure 4.43.B). Thus, the resonance at δ_H 4.974 ppm was assigned to the methylene protons of the *N*-hydroxymethyl group of *N*^{4},*C*^{5}-(hydroxymethyl)(methyl)deoxycytidine(hm^{4}m^{5}dC). The resonances at δ_H 7.76 ppm and δ_H 1.99 ppm correspond to the protons attached to C-6 and C-5 methyl group of the

pyrimidine ring (marked as 14 and 15 in Figure 4.43.B) of hm^4m^5dC. The resonance at δ_H 6.31 ppm was assigned to the proton attached to C-1 (marked as 16 in Figure 4.43.B) of the deoxyribose ring of hm^4m^5dC. The resonance at δ_H 7.76 ppm (marked as 25 in Figure 4.43.B) shows similar carbon correlations in 2D HMBC with that of the proton attached to the C-6 (marked as 14 in Figure 4.43.B) of the pyrimidine ring of hm^4m^5dC, indicating formation of a similar adduct to hm^4m^5dC (Figure 4.44). However, the resonance at δ_H 4.89 ppm shows correlations with the carbon at δ_C 67.8 ppm in 2D HSQC and the carbon at δ_C 164.3 ppm in 2D HMBC. The intensity of the resonance at δ_H 4.89 ppm was approximately four times the intensity of proton at δ_H 7.76 ppm. The resonance at δ_H 7.76 ppm and δ_H 2.03 ppm were assigned to the protons attached to C-6 and the C-5 methyl group of the pyrimidine ring of *N^4*, *$N^{4,}C^5$*-di(hydroxymethyl)(methyl)deoxycytidine(dhm^4m^5dC). However, the conversion of m^5dC to hm^4m^5dC was approximately 65%, while conversion to dhm^4m^5dC was approximately 0.5 %.

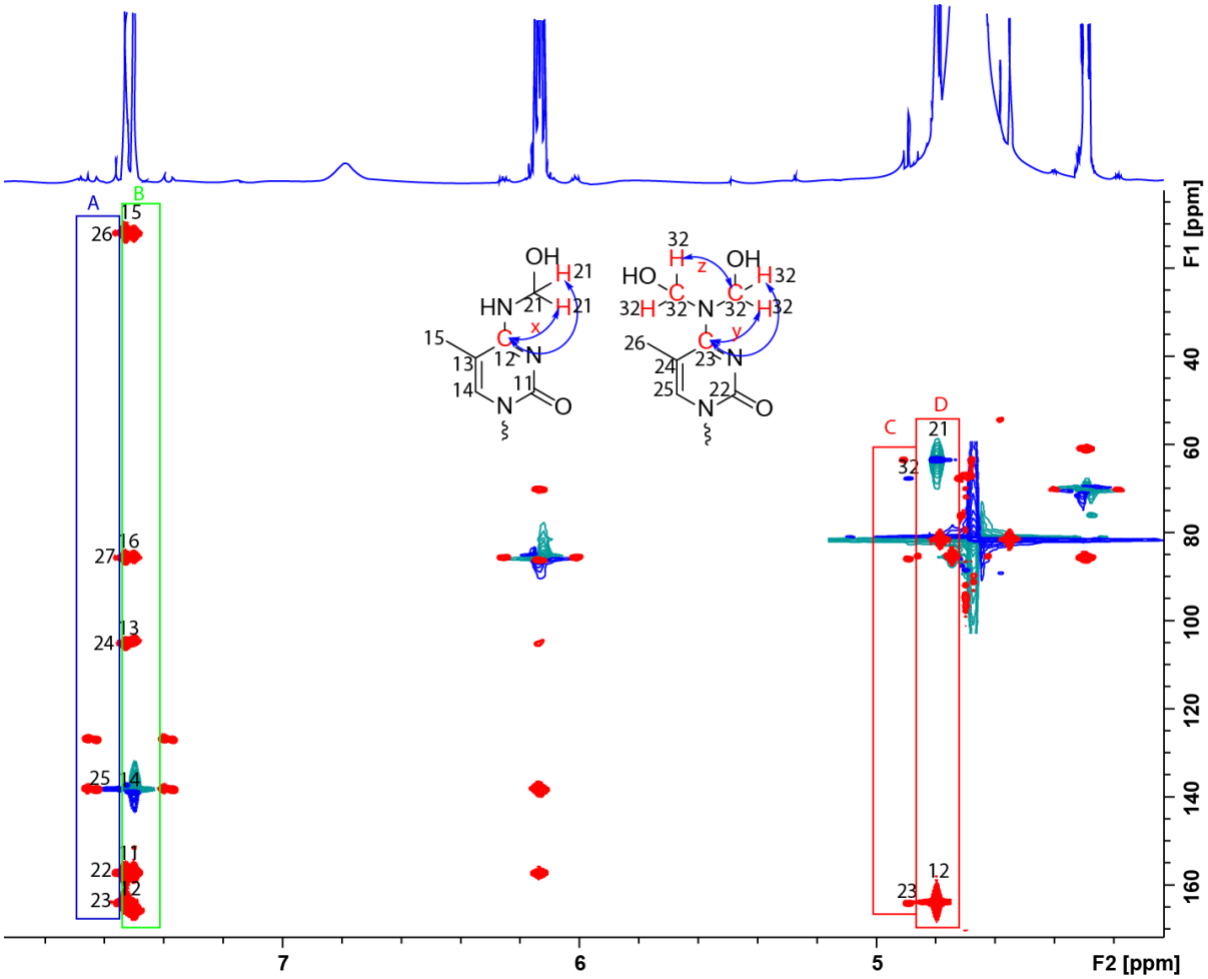


Figure 4.44 Overlaid 2D HSQC (green and blue) and 2D HMBC (red and pink) NMR spectra (700 MHz) for the reaction mixture of a 20 mM m^5dC and formaldehyde at 25°C. Cross-peaks highlighted in Box A (highlighted in purple) correspond to the correlations observed for dhm^4m^5dC. Cross-peaks highlighted in Box B (highlighted in green) correspond to the correlations observed for hm^4m^5dC. Cross-peaks highlighted in Box C (highlighted in red) correspond to the correlation (marked as "y" and "z" in red) observed for dhm^4m^5dC. Cross-peaks highlighted in Box D (highlighted in red) correspond to the correlations (marked as "x" in red) observed for hm^4m^5dC. Important correlations observed between formaldehyde derived protons (marked as red) and hm^5dC structure are highlighted with blue curved arrows.

The structures of the adducts were also assessed by analysing their ESI mass spectra obtained after LC purification. The EI / MS spectrum showed three peaks in the positive ion mode corresponding to m/z: 241.1063 ($[m^5dC +H]^+$, calculated: 241.1063), 271.1133 ($[hm^4m^5dC +H]^+$, calculated: 271.1168) and 291.1117 ($[hm^4m^5dC +H]^+$, calculated: 291.1115).

Table 4.4.14.1 Proton and carbon assignments of m^5dC

Position	δ ^{13}C (ppm)	δ ^{13}H(ppm)	Multiplicity	$J_{H,H}$ (Hz)
1	157.27	-	-	-
2	165.8	-	-	-
3	104.64	-	-	-
4	138.46	7.67	b.s	-
5	12.23	1.98	b.s	-
6	85.62	6.30	t	6.5
7	38.9, 38.9	2.42, 2.32	m, m	-
8	70.3	4.475	dt	-
9	86.43	4.06	q	4.0
10	61.2	3.87, 3.79	dd, dd	3.5, 5.0, 13.0

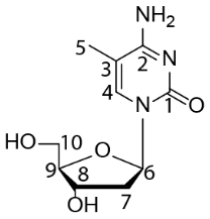


Table 4.4.14.2 Proton and carbon assignment for hm^4m^5dC

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
11	157.41	-	-	-
12	163.97	-	-	-
13	105.29	-	-	-
14	138.2	7.70	b.s	-
15	12.08	1.99	b.s	-
16	85.78	6.31	t	6.5
17	38.9, 38.9	2.42, 2.32	m, m	-
18	70.3	4.475	dt	-
19	86.43	4.06	q	4.0
20	61.2	3.87, 3.79	dd, dd	3.5, 5.0, 13.0
21	63.8	4.974	b.s	-

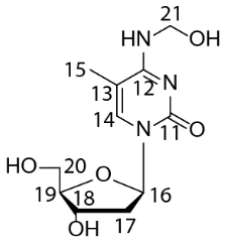
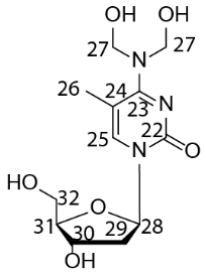


Table 4.4.14.3 Proton and carbon assignments for hm^{4}m^{5}dC

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
22	157.41	-	-	-
23	164.29	-	-	-
24	105.29	-	-	-
25	138.28	7.76	b.s	-
26	12.08	2.03	b.s	-
28	85.78	6.31	t	6.5
29	38.9	2.42, 2.32	m, m	-
30	70.3	4.475	dt	-
31	86.43	4.06	q	4.0
32	61.2	3.87, 3.79	dd, dd	3.5, 5.0, 13.0
27	67.8	4.89	b.s	-



Analogous to the previous studies, a ^{1}H-NMR time course was recorded at pD-7.5. 2.4 mM of m^{5}dC, pre-buffered at pD-7.5, was reacted with a 53 fold excess of formaldehyde, then reaction mixture was transferred immediately to a 3mm NMR match tube and monitored by ^{1}H-NMR. The resonance at δ_H 4.974 ppm (assigned to the *N*-hydroxymethyl group of hm^{4}m^{5}dC) was observed approximately 5 min following initial mixing (i.e. at the first NMR time-point). The resonance at δ_H 4.89 ppm (assigned to the *N*-hydroxymethyl group of dhm^{4}m^{5}dC) was observed after approximately 100 min following initial mixing. The concentration of hm^{4}m^{5}dC reached approximately 1.5 mM after 200 min; the concentration remained constant thereafter. However, resonances corresponding to the methylene bridged cross-linked m^{5}dC adduct were not observed. The initial rate of the formation of major product was 2 x10^{-7} mol dm^{-3} s^{-1} at pD-7.5.

4.4.14.2 Conclusion of the results described in section 4.4.14

In conclusion, the ^{1}H-NMR analyses revealed the formation of two closely related products when m^{5}dC is allowed to react with an excess of formaldehyde. The newly developed resonance at δ_H 4.974 ppm was assigned to the methylene protons of the *N*-hydroxymethyl group of hm^{4}m^{5}dC. Analysis of the 2D HSQC and 2D HMBC spectra revealed the formation of another entity, very similar in structure to hm^{4}m^{5}dC; the methylene protons of the entity attached to the *N*-hydroxymethyl group resonate at δ_H 4.89 ppm. This entity was proposed to be the rotameric form of hm^{4}m^{5}dC. However, its low concentration precluded structural elucidation. The resonances corresponding to the methylene bridged cross-linked hm^{5}dC adduct were not observed. The NMR results were supported by mass spectrometric analyses. The initial rates as calculated from the graph were 2 x10^{-7} mol dm^{-3} s^{-1} at pD-7.5, 27°C (Figure 4.45).

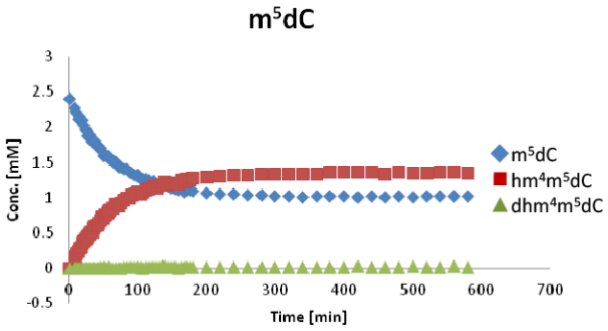


Figure 4.45 Graph showing changes in concentrations of m^{5}dC (blue), hm^{4}m^{5}dC (red), dhm^{4}m^{5}dC (green) over time at 27°C at pD 7.5. The concentrations are obtained from the integration of the peaks corresponding to the resonance of the proton attached to C-6 of the pyrimidine ring of m^{5}dC, hm^{4}m^{5}dC and dhm^{4}m^{5}dC (marked as 4, 14 and 25, respectively in Figure 4.43) and normalising the intensity value relative to internal standard, TSP. ^{1}H NMR timecourses were recorded for 8-10 hours; every 2 min over the first hour, every 4 min over the next 2 hours and every 20 mins for the next 5-7 hours.

4.4.15 Studies on reaction of formaldehyde with Pseudouridine (Ψ)

Pseudouridine was amongst the first post-transcriptional modifications to be identified in t-RNA, r-RNA and m-RNA.[80–82] Ψ is derived from uridine by base specific modification by an

enzyme called pseudouridine synthase.[83] The biological function of Ψ is still largely unknown. However, the presence of extra H-bonding provided by shifted nitrogen could play roles in altering secondary structure. Unlike m^6A modification, the conversion of uridine to Ψ is probably irreversible. By analysing the structure of Ψ, it was envisaged that formaldehyde might react preferentially to one of the base nitrogens.

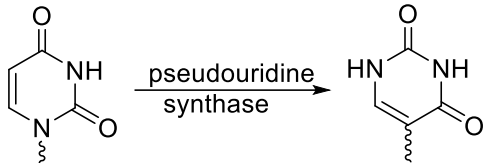


Figure 4.46 Schematic representation of the rearrangement of uridine to pseudouridine as catalysed by pseudouridine synthase.

4.4.15.1 ^{1}H-NMR analysis on the reaction of formaldehyde with Ψ

Initially, the ^{1}H NMR spectrum of a 20 mM Ψ solution in D_2O was acquired and assigned with the help of 2D NMR (HSQC and HMBC). Seven distinct ^{1}H resonances were observed. All the non-exchangeable protons were observed. The resonances at δ_H 7.675 ppm were assigned to the protons attached to C-6 (marked as 4 in Figure 4.47.A) of the aromatic ring of Ψ. The resonances at δ_H 4.68 ppm, δ_H 4.30 ppm, δ_H 4.15 ppm, δ_H 4.02 ppm, δ_H 3.73 ppm and δ_H 3.84 ppm were assigned to the protons attached to C-1, C-2, C-3, C-4, C-5 and C-5 (marked as 5, 6, 7, 8, 9 and 9, respectively, in Figure 4.47.A) of the ribose ring of Ψ. A solution of Ψ (2.4 mM) in D_2O was then treated with excess formaldehyde and monitored by ^{1}H-NMR, 2D HSQC and 2D HMBC (Figure 4.47). The proton and carbon assignments of the m^5dC and newly formed adducts are given in Table 4.4.15 (1-4).

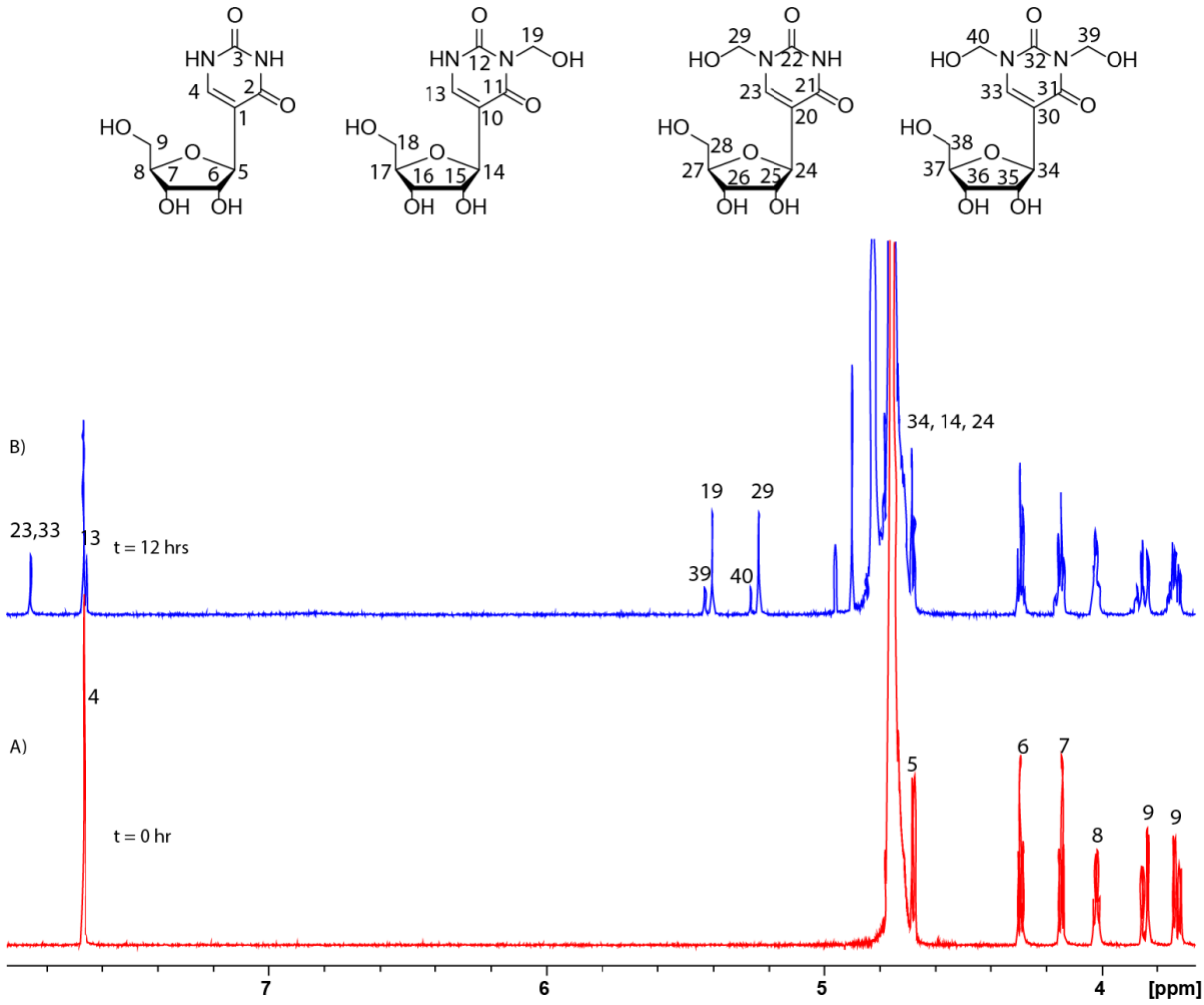


Figure 4.47 ^{1}H-NMR (600 MHz) spectrum of A) 20 mM Ψ at 27°C, pD 7.5 in D_2O; B) reaction mixture of a 20 mM Ψ and excess formaldehyde after 12 hours following initial mixing at 27°C, pD-7.5. The ^{1}H assignments of Ψ and the newly developed resonances are highlighted.

New resonances were observed at δ_H 7.86 ppm, δ_H 7.66 ppm, δ_H 5.43 ppm, δ_H 5.4 ppm, δ_H 5.27 ppm, δ_H 5.24 ppm, δ_H 4.72 ppm and δ_H 4.71 ppm. Investigation of the recorded ^{1}H NMR and 2D-HSQC spectra revealed the formation of three closely related products. It should be noted that the resonance at δ_H 7.86 ppm was assigned to two different carbon shifts in 2D HSQC, implying the presence of two different adducts with the same ^{1}H-chemical shift (Figure 4.48). The resonance at δ_H 5.4 ppm (marked as 19 in Figure 4.47.B) correlates with the carbon at δ_C 63.2 ppm in 2D HSQC and shows correlations with the carbons at δ_C 163.7 ppm, δ_C 152.38 ppm and δ_C 110.42 ppm (marked as 11, 12 and 10 in Figure 4.47.B) in 2D HMBC (Figure 4.48). The intensity of the resonance at δ_H 5.4 ppm was approximately twice the intensity of the resonance at δ_H 7.66 ppm (marked as 13 in Figure 4.47.B). Thus, the

resonance at δ_H 5.4 ppm was assigned to the methylene protons of the *N*-hydroxymethyl group of N^3-hydroxymethylpseudouridine (hm^3 Ψ). The resonances at δ_H 7.66 ppm and δ_H 4.71 ppm correspond to the protons attached to C-6 (marked as 13 in Figure 4.47.B) of the aromatic ring and C-1 of the ribose ring (marked as 14 in Figure 4.47.B) of hm^3 Ψ.

The resonance at δ_H 5.24 ppm (marked as 29 in Figure 4.47.B) correlates to the carbon at δ_C 71.3 ppm in 2D HSQC and shows correlations with the carbons at δ_C 151.7 ppm, δ_C 143.7 ppm and δ_C 111.7 ppm (marked as 22, 23 and 20, respectively, in Figure 4.47.B) in 2D HMBC (Figure 4.48). The resonance at δ_H 7.86 ppm shows correlation to the carbon at δ_C 71.3 ppm (marked as 29 in Figure 4.47.B) in 2D HMBC (Figure 4.48). Hence, the resonances at δ_H 5.24 ppm and δ_H 7.86 ppm appear to belong to the same entity. The intensity of the resonance at δ_H 5.24 ppm was observed to be little less than twice the resonance at δ_H 7.86 ppm. Consequently, the resonance at δ_H 5.24 ppm was assigned to the methylene protons attached to the *N*-hydroxymethyl group of N^5-hydroxymethylpseudouridine (hm^5 Ψ). The resonances at δ_H 7.86 ppm and δ_H 4.72 ppm were assigned to the protons attached to C-6 of the aromatic ring and C-1 of the ribose ring (marked as 23 and 24, respectively, in Figure 4.47.B).

The resonance at δ_H 5.43 ppm correlates with the carbon at δ_C 64.1 ppm in 2D HSQC and shows correlations with the carbons at δ_C 151.6 ppm and δ_C 163.1 ppm (marked as 32 and 31 in Figure 4.47.B) in 2D HMBC (Figure 4.48). It should be noted that the carbon correlations observed for resonance at δ_H 5.43 ppm in 2D HSQC and 2D HMBC were similar to correlations observed for hm^3 Ψ, indicating a similar structure to hm^3 Ψ.

The resonance at δ_H 5.27 ppm correlates to the carbon at δ_C 72.2 ppm in 2D HSQC and shows correlations with the carbons at δ_C 151.6 ppm and δ_C 142.56 ppm in 2D HMBC. It should be noted that the carbon correlations in 2D HSQC and 2D HMBC observed for resonance at δ_H 5.27 ppm were similar to the correlations observed for hm^5 Ψ, indicating structural similarity to hm^5Ψ. Further, the resonances at δ_H 5.43 ppm and δ_H 5.27 ppm

correlate to the same carbon chemical shift at δ_C 151.6 ppm (marked as 32 in Figure 4.47.B) in 2D HMBC, supporting the assignment of the resonances at δ_H 5.43 ppm and δ_H 5.27 ppm to the methylene protons of the *N*-hydroxymethyl groups attached to *N*-3 and *N*-5 of N^3,N^5-di(hydroxymethyl)pseudouridine (hm^3 hm^5 Ψ). No other adducts were observed during the 12 hr following initial mixing. By NMR analysis, the conversion of Ψ to hm^3 Ψ adduct was approximately 14%, while conversion to hm^5 Ψ was approximately 27%; conversion to hm^3 hm^5 Ψ ~4%.

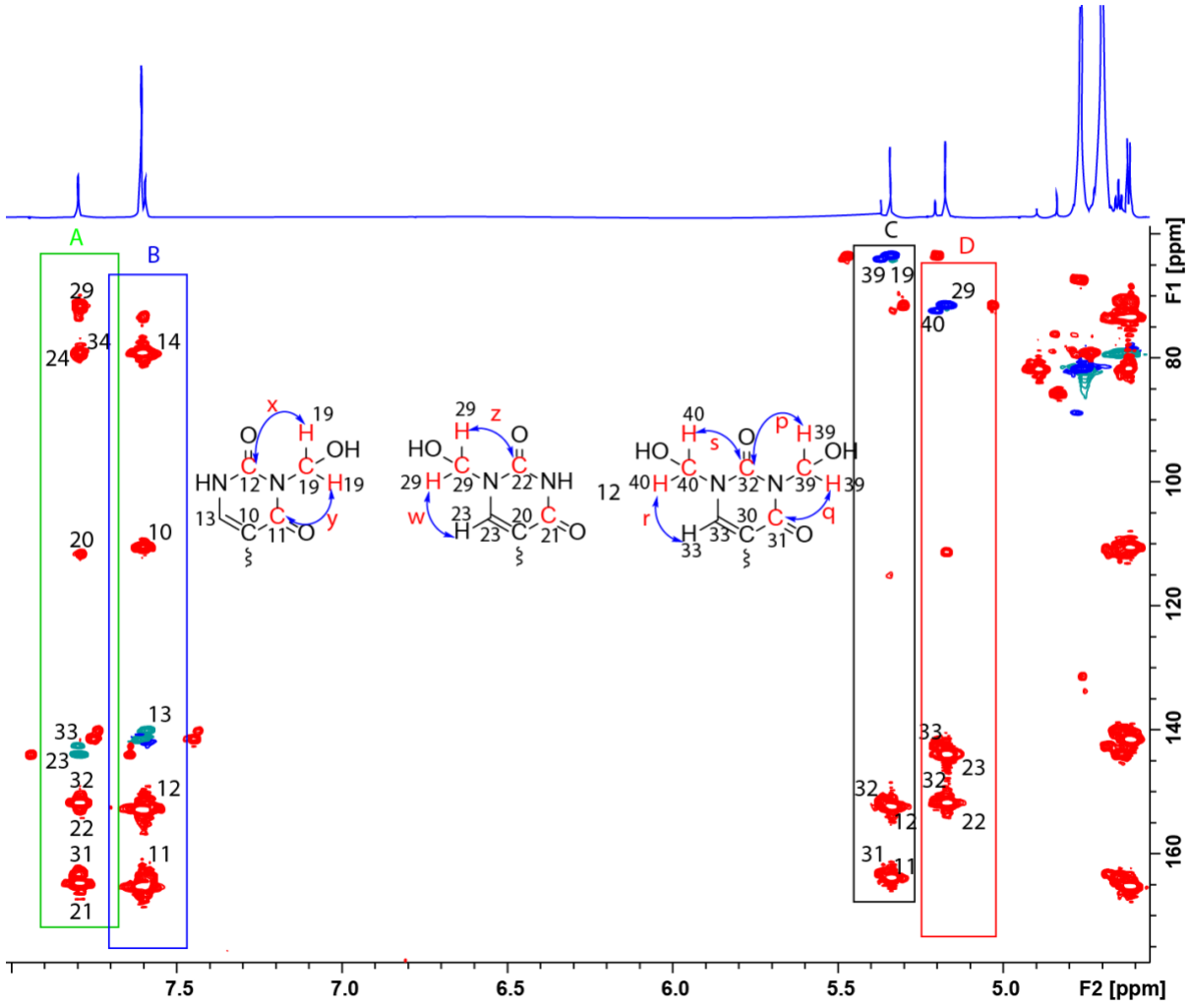


Figure 4.48 Overlaid 2D HSQC (green and blue) and 2D HMBC (red and pink) NMR spectra (600 MHz) for the reaction mixture of a 20 mM Ψ and formaldehyde at 25°C. Cross-peaks highlighted in Box A (highlighted in green) correspond to the correlations observed for hm^5 Ψ and hm^3 hm^5 Ψ. Cross-peaks highlighted in Box B (highlighted in blue) correspond to the correlations observed for hm^3Ψ. Cross-peaks highlighted in Box C (highlighted in black) correspond to the correlations (marked as "p" and "q" in red) observed for hm^3 hm^5 Ψ and the correlations (marked as "x" and "y" in red) observed for hm^3Ψ. Cross-peaks highlighted in Box D (highlighted in red) correspond to the correlations (marked as "s" and "r" in red) observed for hm^3 hm^5 Ψ and the correlations (marked as "w" and "z" in red) observed for hm^5 Ψ. Important correlations observed between formaldehyde derived protons (marked as red) and Ψ structure are highlighted with blue curved arrows.

Table 4.4.15.1 Proton and carbon assignments for Ψ

Position	δ ^{13}C (ppm)	δ ^{13}H(ppm)	Multiplicity	$J_{H,H}$ (Hz)
1	110.42	-	-	-
2	165.3	-	-	-
3	152.8	-	-	-
4	141.5	7.675	b.s	-
5	79.2	4.68	d	5.5
6	73.3	4.30	t	5.5
7	70.9	4.15	t	5.5
8	83.4	4.02	q	4.0
9	61.51	3.73, 3.84	dd, dd	3.0, 5.0, 12.5

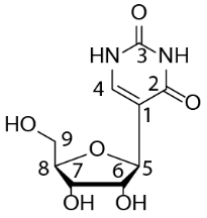


Table 4.4.15.2 Proton and carbon assignments for hm^3 Ψ

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
10	110.42	-	-	-
11	163.7	-	-	-
12	152.38	-	-	-
13	139.98	7.66	b.s	-
14	79.2	4.71	d	5.5
15	73.3	4.29	t	5.5
16	70.9	4.16	t	5.5
17	83.4	4.02	q	4.0
18	61.51	3.87, 3.75	dd, dd	3.0, 5.0, 12.5
19	63.2	5.4	s	-

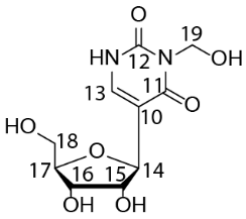


Table 4.4.15.3 Proton and carbon assignments for hm^5 Ψ

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
20	111.7	-	-	-
21	164.7	-	-	-
22	151.7	-	-	-
23	143.7	7.86	b.s	-
24	79.2	4.72	d	5.5
25	73.3	4.29	t	5.5
26	70.9	4.16	t	5.5
27	83.4	4.02	q	4.0
28	61.51	3.87, 3.75	dd, dd	3.0, 5.0, 12.5
29	71.3	5.24	s	-

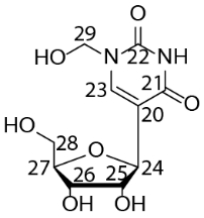
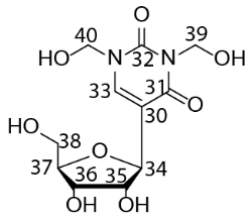


Table 4.4.15.4 Proton and carbon assignments for hm^3 hm^5 Ψ

Position	δ ^{13}C (ppm)	δ ^{13}H (ppm)	Multiplicity	$J_{H,H}$ (Hz)
30	111.7	-	-	-
31	163.1	-	-	-
32	151.6	-	-	-
33	142.56	7.86	b.s	-
34	79.2	4.71	d	5.5
35	73.3	4.29	t	5.5
36	70.9	4.16	t	5.5
37	83.4	4.02	q	4.0
38	61.51	3.87, 3.75	dd, dd	3.0, 5.0, 12.5
39	64.1	5.43	s	-
40	72.2	5.27	s	-



Analogous to the previous studies, a ^{1}H-NMR time course was recorded at pD-7.5. 2.4 mM of Ψ, pre-buffered at pD-7.5, was reacted with a 53-fold excess of formaldehyde, then the reaction mixture was transferred immediately to a 3mm NMR match tube and monitored by ^{1}H-NMR. The resonances at δ_H 5.43 ppm, δ_H 5.4 ppm, δ_H 5.27 ppm and δ_H 5.24 ppm were observed from the first NMR time-point (5 min after mixing) and remained at the same intensity throughout the analysis. The concentration of hm^{3}Ψ reached 0.4 mM, while hm^{5}Ψ reached 0.45 mM; this suggests formaldehyde slightly prefers to react at the *N*-5 nitrogen over the *N-3* nitrogen. The concentration of the dihydroxymethyl adduct remained constant at 0.1 mM throughout the time-course. No other adducts were observed.

4.4.15.2 Conclusion of the results described in section 4.4.14

In conclusion, the ^{1}H-NMR analyses revealed the formation of three closely related products when Ψ is allowed to react with excess of formaldehyde. The newly developed resonances at δ_H 5.4 ppm and δ_H 5.24 ppm were assigned to the methylene protons of the *N*-hydroxymethyl groups of hm^{3}Ψ and hm^{5}Ψ. The resonances at δ_H 5.43 ppm and δ_H 5.27 ppm were assigned to the methylene protons of the *N*-hydroxymethyl groups attached to *N*-3 and *N*-5, respectively of hm^{3}hm^{5}Ψ. Formaldehyde slightly prefers to react at *N*-5 over *N*-3 of the aromatic ring of Ψ. The NMR results were supported by mass spectrometric analysis.

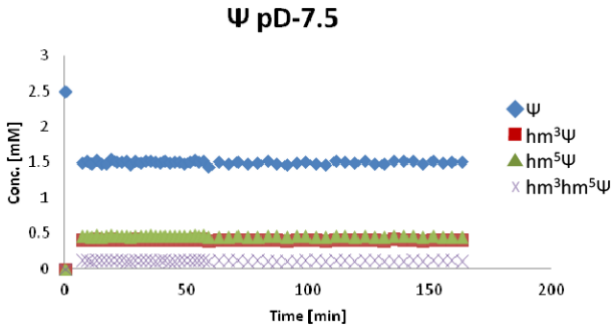


Figure 4.49 Graph showing change in concentrations of Ψ (blue), hm^3 Ψ (red), hm^5 Ψ (green) and hm^3 hm^5 Ψ (purple) over time at 27°C at pD 7.5. The concentrations are obtained from the integral value in the ^1H spectra of the peaks corresponding to resonances of the proton attached to C-6 of the pyrimidine ring of Ψ, methylene protons attached to N-hydroxymethyl group of $hm^3Ψ$ and $hm^5Ψ$ and methylene protons attached to N^3-hydroxymethyl group hm^3 hm^5 Ψ (marked as 4, 19, 29 and 39, respectively in Figure 4.47) and normalising the intensity value relative to internal standard, TSP. ^1H NMR timecourses were recorded for 8-10 hours; every 2 min over the first hour, every 4 min over the next 2 hours and every 20 mins for the next 5-7 hours. It should be noted that the graph is shown for the first 180 min as change in concentration of species were not observed after 5 min.

4.4.16 Comparison of the reactivity of Cytidine (C), hm^5dC and m^5dC

Prior to comparison of the reactivity of C, hm^5dC and m^5dC, it was necessary to complete the ^1H assignments. dCMP has been characterised earlier in the Chapter in section 4.4.6.1 and not cytidine. However, the ^1H-assignments of cytidine were expected to be similar to that of dCMP.

4.4.16.1 ^1H-NMR analysis on the reaction of formaldehyde with C.

A ^1H-NMR spectrum was recorded on a 20 mM cytidine and assigned with the help of recorded 2D HSQC and 2D HMBC. All the non-exchangeable protons were observed. Seven resonances were observed. The resonance at δ_H 8.01 ppm and δ_H 6.04 ppm were assigned to the protons attached to C-6 and C-5 (marked as 4 and 3 in Figure 4.50.A) of the pyrimidine ring of C. The resonances at δ_H 5.91 ppm, δ_H 4.25 ppm, δ_H 4.15 ppm, δ_H 3.97 ppm and δ_H 3.89 ppm were assigned to the protons attached to C-1, C-2 (and C3), C-4, C-5 and C-5 of the ribose ring of C.

Excess formaldehyde was added to an identical cytidine solution and the reaction mixture was allowed to stand for 12 hours to reach equilibrium. ^{1}H-NMR, 2D HSQC and 2D HMBC spectra were then recorded.

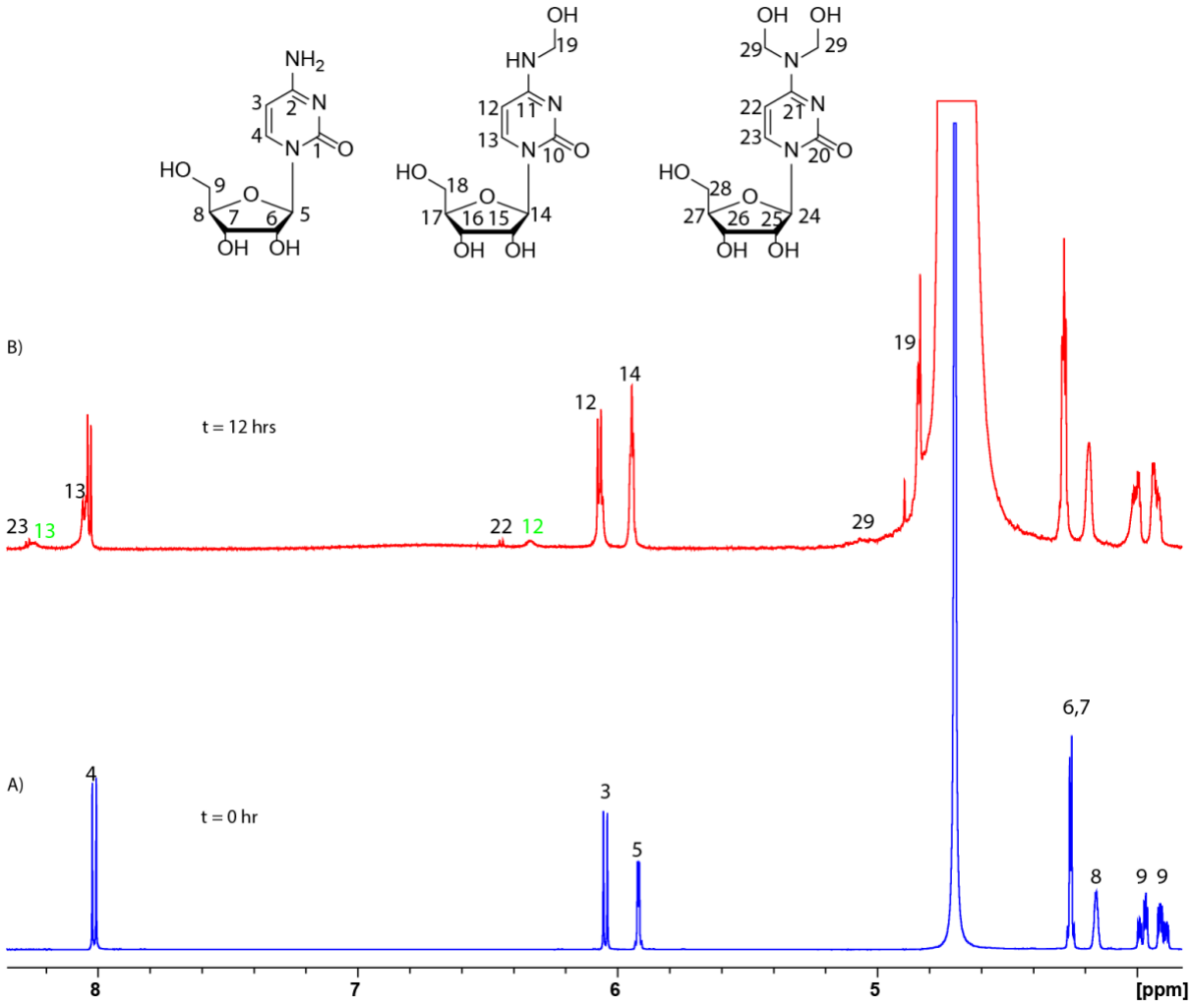


Figure 4.50 ^{1}H-NMR (500 MHz) spectrum of A) 20 mM cytidine at 27°C, pD 7.5 in D_2OB) reaction mixture of a 20 mM cytidine and excess formaldehyde after 12 hours following initial mixing at 27°C, pD-7.5. The ^{1}H assignments of cytidine and the newly developed resonances are indicated.

Similar to dCMP, seven new resonances were observed at δ_H 8.27 ppm, δ_H 8.25 ppm, δ_H 8.05 ppm, δ_H 6.45 ppm, δ_H 6.34 ppm, δ_H 5.06 ppm and δ_H 4.84 ppm. The resonance at δ_H 6.06 ppm and δ_H 5.94 ppm were observed to overlap with the resonances from the starting material. Similar to the ^{1}H NMR spectrum of dCMP and formaldehyde reaction mixture and further supported by correlations observed in 2D HSQC and 2D HMBC, the resonance at δ_H

4.84 ppm was assigned to the methylene protons attached to the *N*-hydroxymethyl group of N^4-hydroxymethylcytidine (hm^4C). Analogous to the observations in the ^{1}H-NMR spectrum of the reaction mixture of dCMP and formaldehyde, the resonance at δ_H 4.84 ppm was assigned to the methylene protons of the *N*-hydroxymethyl group of two rotameric forms of hm^4C. The resonances at δ_H 8.05 ppm, δ_H 6.06 ppm and δ_H 5.94 ppm were assigned to the protons attached to C-6 and C-5 of the pyrimidine ring and C-1 of the ribose ring of major rotamer of hm^4C. The resonances at δ_H 8.25 ppm and δ_H 6.34 ppm were assigned to the protons attached to C-6 and C-5 of the pyrimidine ring of the minor rotamer of hm^4C.

The resonance at δ_H 5.06 ppm corresponds to the methylene protons of the two *N*-hydroxymethyl groups of N^4,N^4-di(hydroxymethyl)cytidine (dhm^4C). No other adducts were observed during 12 hr following initial mixing.

4.4.16.2 ^{1}H-NMR analysis on the reaction of formaldehyde with C, hm^5dC and m^5dC.

After assigning C, hm^5dC and m^5dC by NMR, work then focused on investigating their reactivity with formaldehyde. It should be noted that the reactivity of the nitrogenous bases with formaldehyde are independent of the presence of additional hydroxyl group at the 2'C of the sugar ring (ribose vs deoxyribose) as would be proved later in section 4.4.19.

A ^{1}H-NMR spectrum was recorded on a sample containing 2.4 mM of C, 2.4 mM of hm^5dC, 2.4 mM of m^5dC. All the non-exchangeable protons were observed. It should be noted that the full characterisation of the ^{1}H-NMR spectra of C, hm^5dC and m^5dC had been carried earlier in section 4.4.16.1, 4.4.13.1 and 4.4.14.1, respectively.

The resonances at δ_H 7.9 ppm, δ_H 7.85 ppm and δ_H 7.66 ppm correspond to the protons attached to the C-6 of the pyrimidine rings of hm^5dC, cytidine and m^5dC. The resonances at δ_H 6.29 ppm and δ_H 6.27 ppm correspond to the protons attached to C-1 of the deoxyribose ring of m^5dC and hm^5dC. The resonances at δ_H 6.05 ppm and δ_H 5.9 ppm correspond to the protons attached to C-5 of the pyrimidine ring and C-1 of the ribose ring of C. The

resonances at δ_H 4.43 ppm and δ_H 1.97 ppm correspond to the protons attached to the C-5 hydroxymethyl and C-5 methyl groups of the pyrimidine ring of hm^5dC and m^5dC, respectively. The resonances corresponding to the sugar ring of each nucleotide were observed to possess similar ^1H chemical shifts owing to same chemical structure, leading to a highly overlap region between δ_H 3.7 ppm and δ_H 4.5 ppm.

76.8 mM of formaldehyde was then reacted with a solution containing 2.4 mM each of C, hm^5dC and m^5dC and the mixture was transferred immediately to a 3mm NMR tube. The first ^1H-NMR spectrum was recorded approximately 6 min following initial mixing.

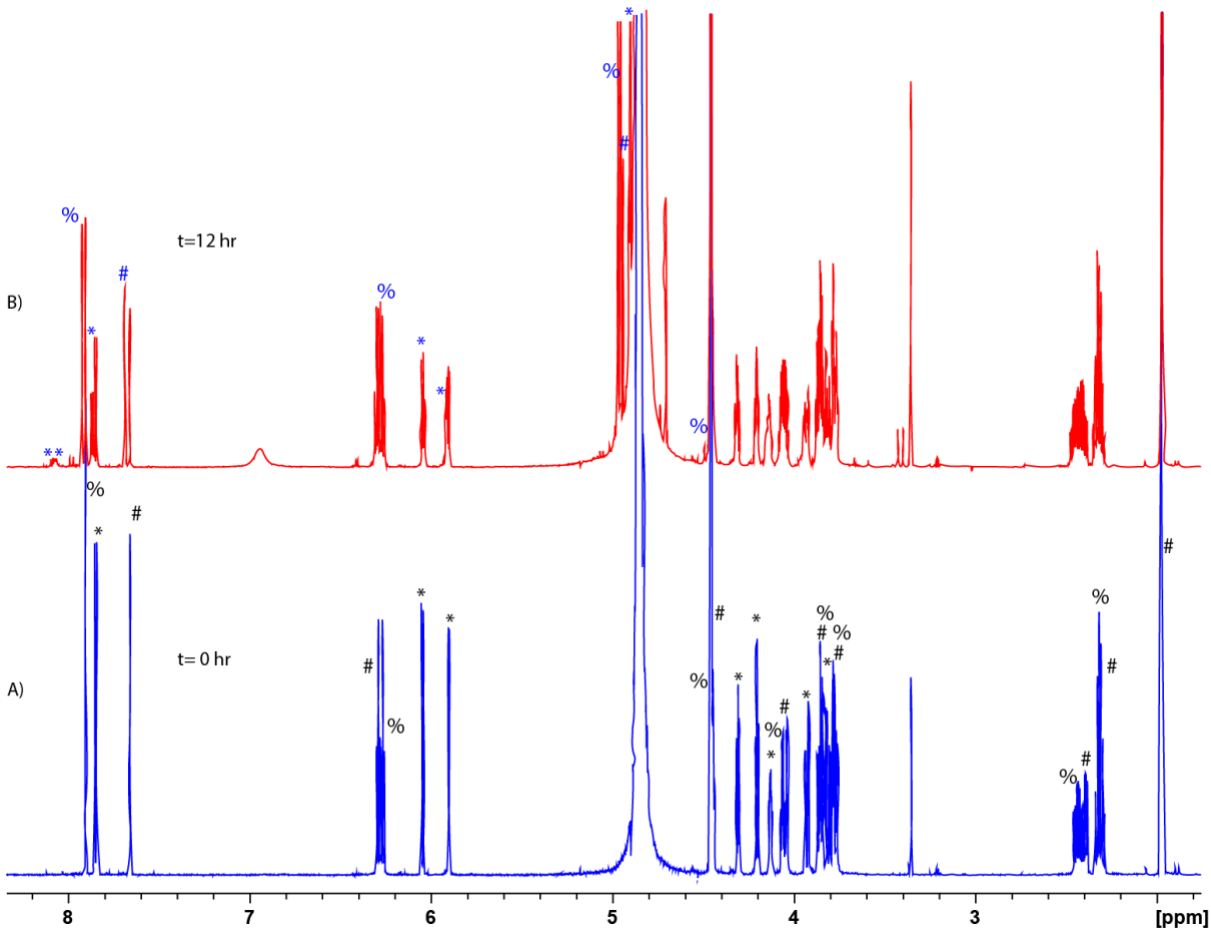


Figure 4.51 ^1H-NMR (700 MHz) spectrum of A) 20 mM each of C, hm^5dC and m^5dC at 27°C, pD 7.5 in D_2O; B) reaction mixture of 20 mM each of C, hm^5dC and m^5dC and excess formaldehyde after 12 hours following initial mixing at 27°C, pD-7.5. The ^1H assignments of C (*), hm^5dC (%) and m^5dC (#) are indicated and the assignments ocorresponding tothe newly developed resonances are indicated in blue.

The resonance at δ_H 7.92 ppm corresponds to the proton attached to C-6 of the pyrimidine ring of hm^{4}hm^{5}dC. The resonance at δ_H 7.99 ppm corresponds to the proton attached to C-6 of the pyrimidine ring of dhm^{4}meC. The resonances at δ_H 8.07 ppm and δ_H 7.87 ppm correspond to the protons attached to C-6 of the pyrimidine rings of hm^{4}C (two rotamers). The resonance at δ_H 8.09 ppm corresponded to the proton attached to C-6 of the pyrimidine ring of dhm^{4}C. The resonance at δ_H 7.68 ppm corresponds to the proton attached to C-6 of the pyrimidine ring of hm^{4}m^{5}dC. The resonance at δ_H 4.97 ppm, δ_H 4.96 ppm and δ_H 4.84 ppm correspond to the methylene protons attached to the *N*-hydroxymethyl groups of hm^{4}hm^{5}dC, hm^{4}m^{5}dC and hm^{4}C, respectively.

4.4.16.3 Comparison of the reactivity of Cytidine (C), hm^{5}dC and m^{5}dC with formaldehyde.

The initial rates of reaction of hm^{5}dC and C were slower than that of m^{5}C. However, the concentrations of hm^{4}hm^{5}dC, hm^{4}C and hm^{4}m^{5}dC were found to be approximately 1.24 mM, 1.4 mM and 1.36 mM after 400 min following initial mixing; the concentration remained constant thereafter. The final concentration of hm^{4}C was higher than hm^{4}m^{5}dC, which might be due to the impurities overlapping with the resonances at δ_H 8.05 ppm. The initial rate of reaction of C with formaldehyde was slower than m^{5}dC, however, hm^{4}C was observed as the most abundant product.

4.4.16.4 Conclusion of the results described in section 4.4.16:

In conclusion, the ^{1}H-NMR analyses of the mixture of C and formaldehyde revealed the formation of two closely related products when C is allowed to react with excess of formaldehyde. The newly developed resonances at δ_H 4.84 ppm is assigned to the methylene protons attached to the *N*-hydroxymethyl groups of the two rotamers of hm^{4}C. The resonance

at δ_H 5.06 ppm corresponds to the methylene protons of the two *N*-hydroxymethyl groups of dhm^{4}C. The resonances corresponding to the crosslinked C via *N*-4 methylene bridge were not observed.

Among C, m^{5}dC and hm^{5}dC, the initial rate of reaction is fastest for m^{5}dC. The initial reaction rates for hm^{5}dC and C were found to be comparable. The conversion to product was highest for C, closely followed by m^{5}dC and hm^{5}dC. hm^{4}hm^{5}dC reaches equilibrium faster than C and m^{5}dC, however, the equilibrium concentration of hm^{5}dC was lowest.

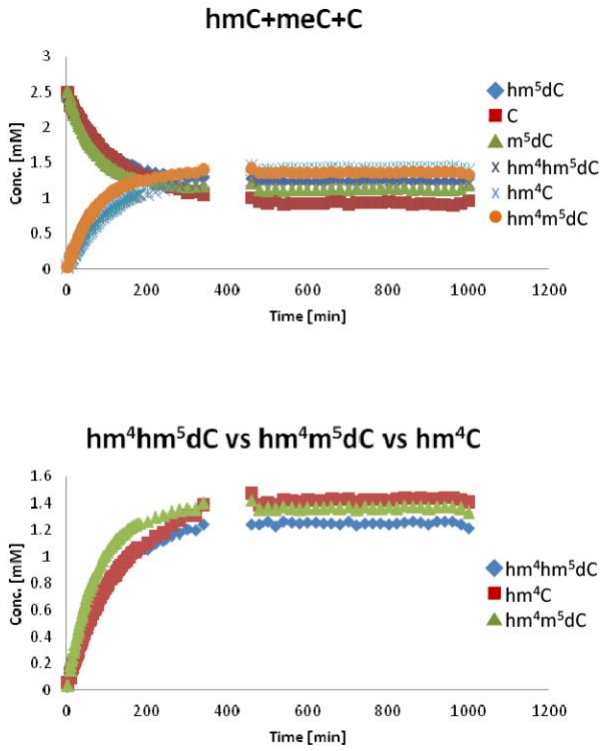


Figure 4.52 Graph showing changes in concentration of hm^{5}dC (blue), C (red), m^{5}dC (green), hm^{4}hm^{5}dC (purple), hm^{4}C (light blue) and hm^{4}m^{5}dC (orange) over time at 27°C at pD 7.5(top); Graph showing changes in concentration of major products i.e. hm^{4}hm^{5}dC (blue), hm^{4}C (red) and hm^{4}m^{5}dC (green) over time at 27°C at pD 7.5(bottom). The concentrations are obtained from the integration of the peaks corresponding to the resonances of the proton attached to C-6 of the pyrimidine ring of hm^{5}dC, C, hm^{5}dC, hm^{4}hm^{5}dC, hm^{4}C and hm^{4}m^{5}dC and normalising the intensity value relative to internal standard, TSP. ^{1}H NMR timecourses were recorded for 8-10 hours; every 2 min over the first hour, every 4 min over the next 2 hours and every 20 mins for the next 15 hours.

4.4.17 Comparison of the reactivities of dAMP, dGMP, dCMP, TMP and UMP with formaldehyde.

4.4.17.1 ^{1}H-NMR analysis on the reaction of formaldehyde with dAMP, dGMP, dCMP, TMP and UMP.

Then, the competition experiments comparing the reactions of canonical nucleosides with formaldehyde was conducted. Initially, a ^{1}H-NMR spectrum of a solution containing 2.2 mM of each nucleotide (dAMP, dGMP, dCMP, TMP and UMP) was recorded and assigned. All the non-exchangeable protons were observed. The resonances corresponding to the deoxyribose rings of each nucleotide possessed similar ^{1}H chemical shifts, leading to highly overlap region between δ_H 2.2 ppm and δ_H 4.7 ppm. The resonances corresponding to the protons attached to the aromatic rings of each nucleotide were observed similar but distinct ^{1}H chemical shifts.

The resonances at δ_H 8.54 ppm and δ_H 8.25 ppm correspond to the protons attached to C-8 and C-2 of the purine ring of dAMP. The resonance at δ_H 8.16 ppm corresponds to the proton attached to C-8 of the purine ring of dGMP. The resonances at δ_H 8.13 ppm and δ_H 6.0 ppm and δ_H 5.99 ppm correspond to the protons attached to C-6 of the pyrimidine ring, C-1 of the ribose ring and C-5 of the pyrimidine ring of UMP. The resonances at δ_H 8.07 ppm and δ_H 6.12 ppm correspond to the protons attached to C-6 and C-5 of the pyrimidine ring of dCMP. The resonances at δ_H 7.84 ppm and δ_H 1.93 ppm correspond to the protons attached to C-6 and C-5 methyl group of the pyrimidine ring of TMP.

Then, 76.8 mM of formaldehyde was reacted with the solution containing 2.2 mM of each nucleotide, pre-buffered at pD-7.5, and transferred immediately to a 3mm NMR tube. The first acquisition was acquired approximately 6 minutes following initial mixing.

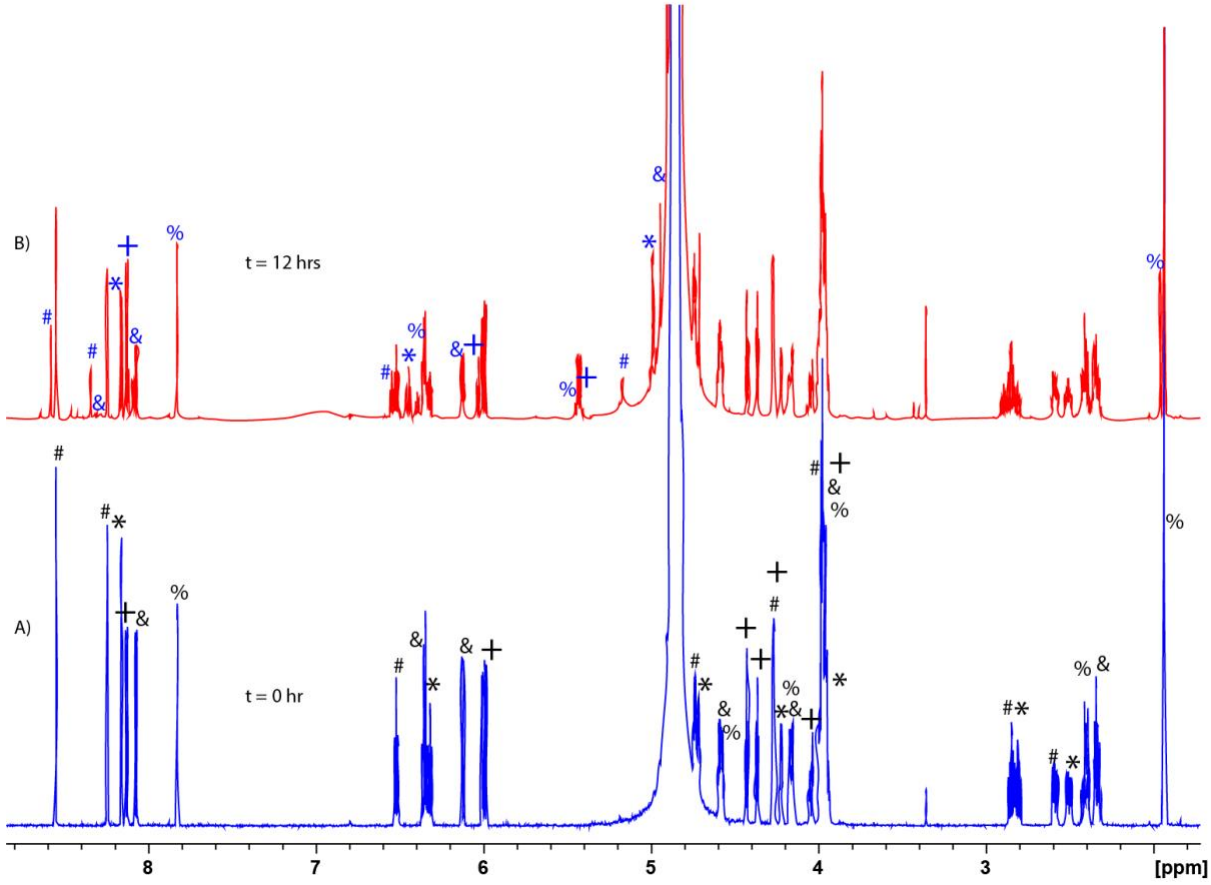


Figure 4.53 A) ^{1}H-NMR (700 MHz) spectrum of a 20 mM each of dAMP, dGMP, dCMP, TMP and UMP at 27°C, pD 7.5 in D_2O. The ^{1}H assignments of dAMP (#), dGMP (*), dCMP (&), TMP (%) and UMP (+) are indicated. B) ^{1}H-NMR (700MHz) spectrum for the reaction mixture of a 20 mM each of dAMP, dGMP, dCMP, TMP and UMP and excess formaldehyde after 12 hours following initial mixing at 27°C, pD-7.5. The ^{1}H assignments of the newly developed resonances are marked in blue symbols.

New resonances were observed to develop. The resonances at δ_H 5.43 ppm and δ_H 5.42 ppm corresponding to the methylene protons of *N*-hydroxymethyl group of hm^{3}TMP and hm^{3}UMP, respectively were observed to develop to a maximum intensity in the first ^{1}H-NMR spectrum recorded. The concentration of hm^{3}TMP and hm^{3}UMP remained constant at approximately 0.6 mM. As expected, the resonances at δ_H 8.64 ppm and δ_H 8.42 ppm, corresponding to the protons attached to the purine ring of hm^{6}dAMP, were observed to develop at a slower rate than the resonances for hm^{3}TMP and hm^{3}UMP. The resonance at δ_H 5.1 ppm corresponds to the methylene protons attached to the *N*-hydroxymethyl group of hm^{6}dAMP. The resonance at δ_H 8.16 ppm, corresponding to the proton attached to C-8 of the

purine ring of hm^{2}dGMP, develop at a faster rate than the resonances corresponding to hm^{6}dAMP but slower than those of hm^{3}TMP and hm^{3}UMP. The resonance at δ_H 4.9 ppm corresponds to the methylene protons attached to the *N*-hydroxymethyl group of hm^{2}dGMP.The resonance at δ_H 8.19 ppm corresponds to the protons attached to C-8 of the purine ring of dhm^{2}dGMP. The resonance at δ_H 8.09 ppm corresponding to the proton attached to the C-6 of the pyrimidine ring of the major rotamer of hm^{4}dCMP was observed to develop at faster rate than hm^{2}dGMP and hm^{6}dAMP. The resonance corresponding to the proton attached to C-6 of the pyrimidine ring of minor rotamer of hm^{4}dCMP was observed to develop at δ_H 8.28 ppm. The resonance at δ_H 4.84 ppm corresponds to the methylene protons attached to the *N*-hydroxymethyl groups of both the rotamers of hm^{4}dCMP. It should be noted that the concentration of hm^{4}dCMP in Figure 4.54 includes the total concentration of hm^{4}dCMP (major and minor rotamer of hm^{4}dCMP). The resonance at δ_H 8.3 ppm corresponds to the proton attached to C-6 of the pyrimidine ring of dhm^{4}dCMP.

The concentration of hm^{6}dAMP, hm^{2}dGMP and hm^{4}dCMP was found to be approximately 0.72 mM, 1.18 mM and 1.36 mM after 800 min following initial mixing, respectively.

4.4.17.2 Conclusion of the results described in section 4.4.11

Although dGMP was previously assumed to be the most reactive nucleotide amongst the exocyclic amines, the present studies suggest that dCMP is the most reactive exocyclic canonical nucleotide. The initial rates of reaction of TMP and UMP are comparable and the fastest amongst the tested nucleotides, followed by dCMP, dGMP and dAMP respectively. However, the conversion to product was found to be highest for dCMP, followed by dGMP and dAMP. The minimum conversion to products were found in TMP and UMP. Hence, the products formed for TMP and UMP are kinetically controlled and the equilibrium favours the reactant side (free nucleotide and formaldehyde) whereas the rate of formation of products with exocyclic amine is slower but are thermodynamically more stable.

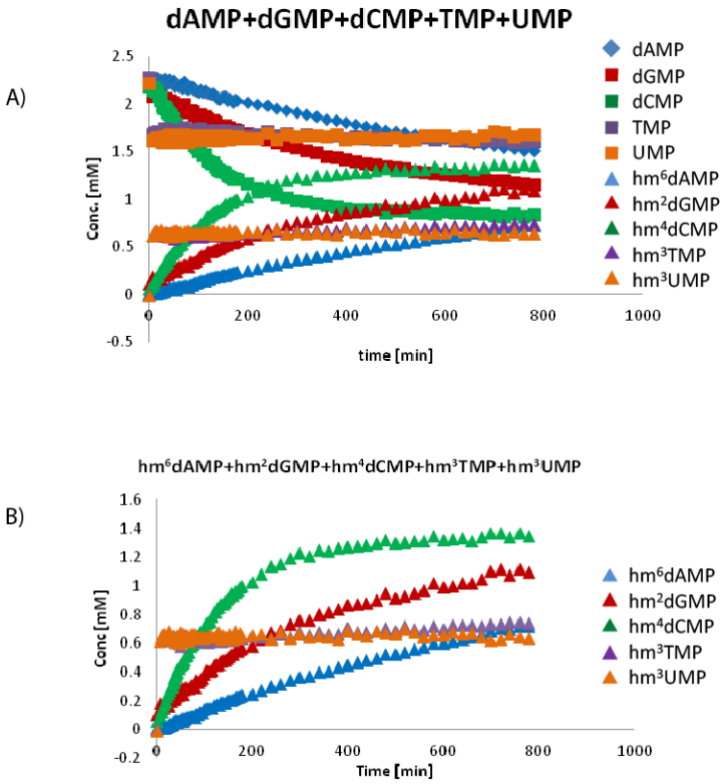


Figure 4.54 Graph showing changes in concentrations of A) dAMP (blue; square), dGMP (red; square), dCMP (green; square), TMP (purple; square), UMP (orange; square), hm^{6}dAMP (blue; triangle), hm^{2}dGMP (red; triangle), hm^{4}dCMP (green; triangle), hm^{3}TMP (purple; triangle), hm^{3}UMP (orange; triangle) over time at 27°C at pD 7.5; B) hm^{6}dAMP (blue; triangle), hm^{2}dGMP (red; triangle), hm^{4}dCMP (green; triangle), hm^{3}TMP (purple; triangle), hm^{3}UMP (orange; triangle) over time at 27°C at pD 7.5. The concentrations are obtained from the integration of the peaks corresponding to the resonances of the proton attached to C-8 of the purine ring of dAMP and hm^{6}dAMP, C-2 of the purine ring of dGMP and hm^{2}dGMP , C-6 of the pyrimidine ring of dCMP and hm^{4}dCMP, C-6 of the pyrimidine ring of TMP and UMP, N-hyroxymethyl group of hm^{4}TMP and hm^{4}UMP and normalising the intensity value relative to internal standard, TSP. ^{1}H NMR timecourses were recorded for 8-10 hours; every 2 min over the first hour, every 4 min over the next 2 hours and every 20 mins for the next 5-7 hours.

4.4.18 Effect of temperature on the rates of reaction of formaldehyde with nucleotide.

The effect of temperature on the initial rates of the reaction of nucleotides with formaldehyde was investigated. The ^{1}H-NMR time-course was recorded at two different temperatures (15 °C and 37 °C) at pD 7.5. The probe of the 600 MHz NMR was either cooled or heated to the desired temperature. The initial mixing was carried out at room temperature and the reaction mixture was transferred immediately to a 3mm NMR tube to minimise time-lapse.

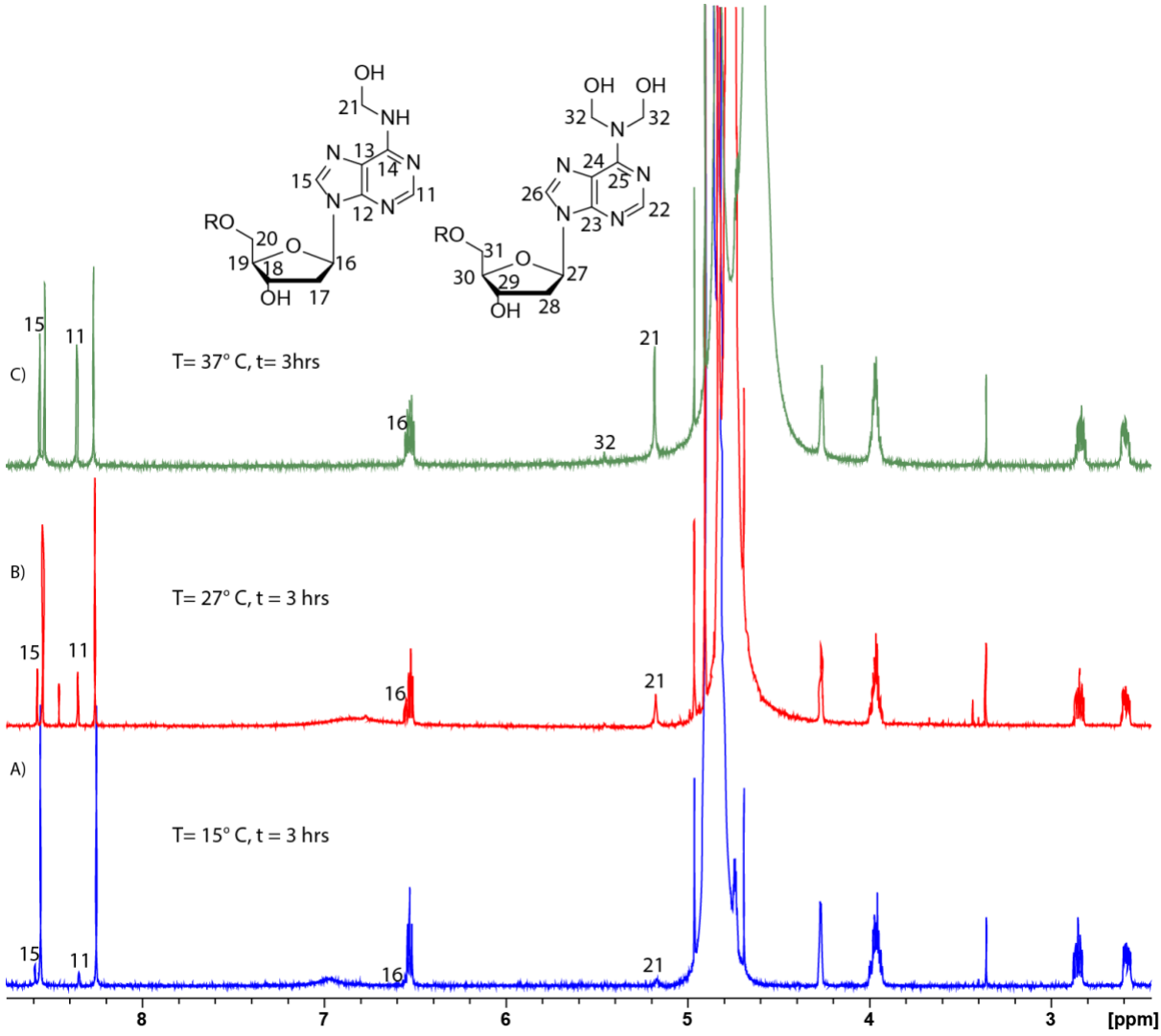


Figure 4.55 ^{1}H-NMR (500 MHz) spectrum for the reaction mixture of 40 mM dAMP and excess formaldehyde at pD 7.5 in D_2O A) 15 °C; B) 27 °C; C) 37 °C. The ^{1}H assignments of hm^{6}dAMP and dhm^{6}dAMP are indicated.

After 3 hours following initial mixing, the concentrations of hm^{6}dAMP were found to be approximately 0.15 mM, 0.54 mM and 1.14 mM at 15°C, 27°C and 37°C, respectively. It should be noted that the concentration of hm^{6}dAMP at 37°C is more than double the concentration at 27°C. The concentration of hm^{6}dAMP at 37°C was observed to halt at 1.16 mM after 200 min following initial mixing, i.e. the reaction reached equilibrium. The concentrations of hm^{6}dAMP were observed to be approximately 1.08 mM (and constant, thereafter) and 0.35 mM at 27 °C and 15 °C, respectively after 500 min following initial mixing. Compared to the rates at 27 °C, the initial rate of reaction increased by 2.5 fold at 37 °C and decreased by 3 fold at 15 °C.

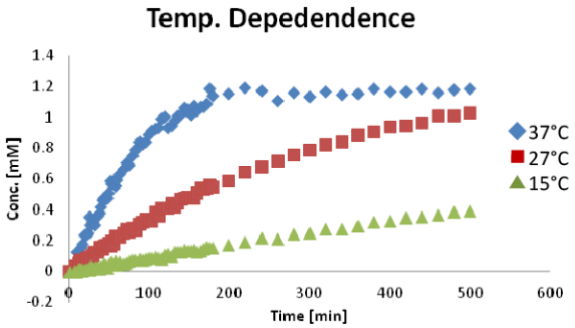


Figure 4.56 Graph showing changes in concentrations of hm^{6}dAMP over time at 37°C (blue), 27°C (red), and 15°C (green). The concentrations are obtained from the integration of the peaks corresponding to the resonance of the proton attached to C-2 of the purine ring of hm^{6}dAMP (marked as 11 in Figure 4.55) and normalising the intensity value relative to internal standard, TSP. ^{1}H NMR time-courses were recorded for 8-10 hours; every 2 min over the first hour, every 4 min over the next 2 hours and every 20 mins for the next 5-7 hours.

4.4.19 Effect of change in the sugar moiety on the rates of reaction of formaldehyde with nitrogenous base.

The work was then focussed on investigating the effect of the additional 2'-hydroxyl group attached to the deoxyribose ring on the reactivity of nitrogenous base (deoxyribose vs ribose). It should be noted that the ^{1}H-NMR assignments of dAMP have been studied in section 4.4.4. The ^{1}H-NMR assignments of AMP would be studied in the following section.

4.4.19.1 ^{1}H-NMR analysis on the reaction of formaldehyde with AMP

Initially, a ^{1}H-NMR spectrum of a 40 mM AMP was recorded and assigned using 2D HSQC and 2D HMBC. All the non-exchangeable protons were observed. The resonances at δ_H 8.53 ppm and δ_H 8.2 ppm correspond to the protons attached to C-8 and C-2 (marked as 15 and 11 in Figure 4.57.B) of the purine ring of AMP. The resonances at δ_H 6.97 ppm, δ_H 4.73 ppm, δ_H 4.45 ppm, δ_H 4.3 ppm and δ_H 3.96 ppm correspond to the protons attached to C-1, C-2, C-3, C-4 and C-5 (marked as 16, 17, 18, 19 and 20, respectively, in Figure 4.57.B) of the ribose ring of AMP.

It should be noted that the proton attached to C-1 (marked as 16 in Figure 4.57.B) of the ribose ring of AMP resonates up field (as a doublet) relative to the corresponding proton

(marked as 6 in Figure 4.57.A) of the deoxyribose ring of dAMP. The proton attached to C-2 (marked as 17 in Figure 4.57.B) of the AMP ribose ring shifts downfield relative to the corresponding protons (marked as 7 in Figure 4.57.A) of the deoxyribose ring of dAMP. The proton attached to C-3 (marked as 18 in Figure 4.57.A) of the AMP ribose ring of shifts up field relative to the corresponding proton (marked as 8 in Figure 4.57.A) of the deoxyribose ring of dAMP.

Excess formaldehyde was then added to the solution and the sample was allowed to stand for 12 hours to attain equilibrium. ^{1}H-NMR, 2D HSQC and 2D HMBC spectra were then recorded on the sample.

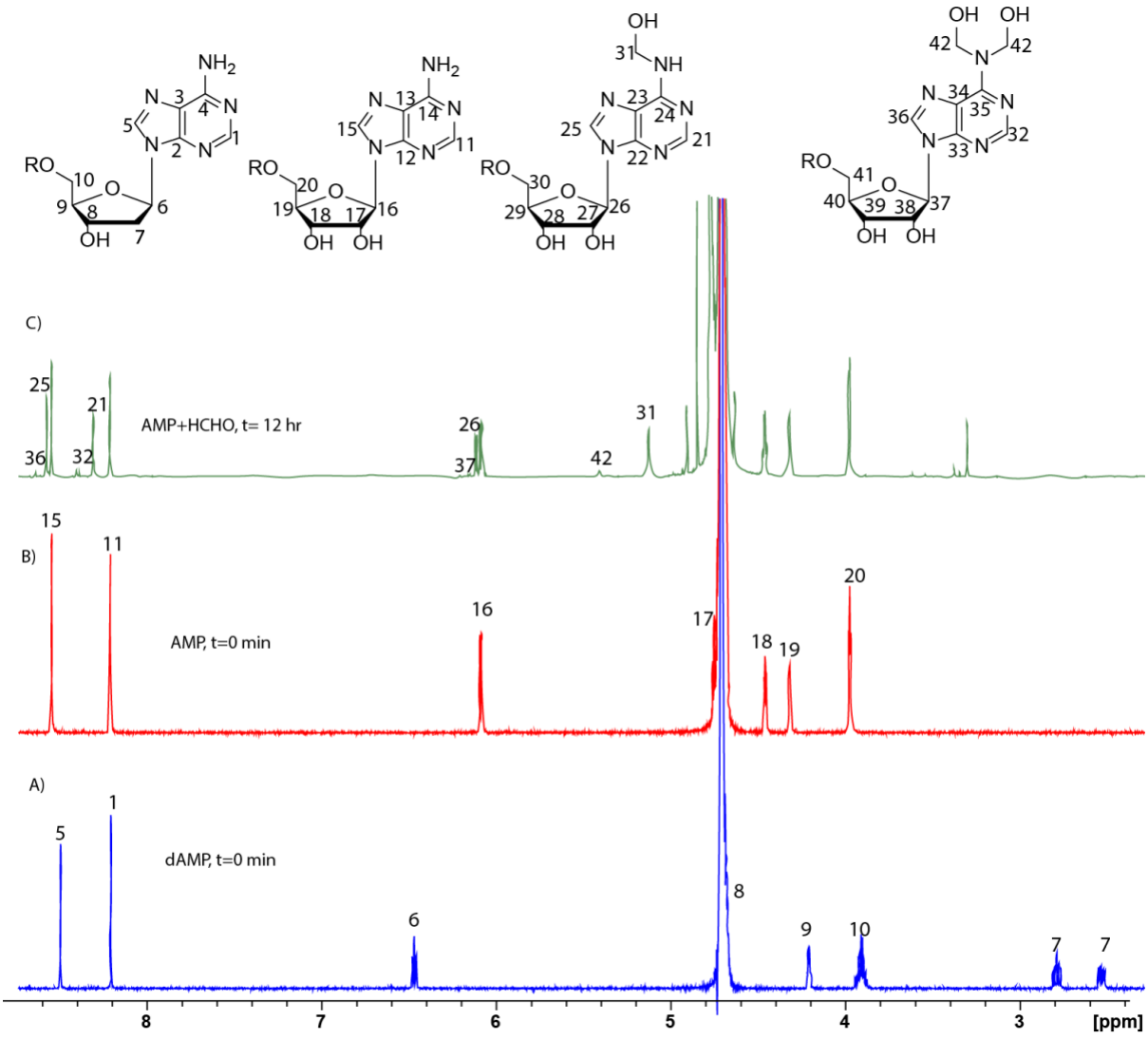


Figure 4.57 ^{1}H-NMR (600 MHz) spectrum of A) 40 mM dAMP at 27°C, pD 7.5 in D_2O. The ^{1}H assignments of dAMP are indicated. B) 40 mM AMP at 27°C, pD 7.5 in D_2O. The ^{1}H assignments of AMP are indicated. C) the reaction mixture of 40 mM AMP and excess formaldehyde after 12 hours following initial mixing at 27°C, pD-7.5. The ^{1}H assignments of the newly developed resonances are indicated.

New resonances were observed at δ_H 8.62 ppm, δ_H 8.56 ppm, δ_H 8.37 ppm, δ_H 8.3 ppm, δ_H 6.14 ppm, δ_H 6.1 ppm, δ_H 5.4 ppm and δ_H 5.1 ppm. Analogous analyses with dAMP, a resonance at δ_H 5.1 ppm was observed and is assigned to the methylene protons attached to the *N*-hydroxymethyl group (marked as 31 in Figure 4.57.C) of hm^{6}AMP. The resonances at δ_H 8.56 ppm, δ_H 8.3 ppm and δ_H 6.1 ppm correspond to the protons attached to C-8 and C-2 of the purine ring and C-1 of the ribose ring of hm^{6}AMP (marked as 25, 21 and 26, respectively in Figure 4.57.C).

The resonance at δ_H 5.4 corresponds to the methylene protons attached to the two *N*-hyroxymethyl groups (marked as 42 in Figure 4.57.C) of dhm^{6}AMP (as assigned by 2D HSQC and HMBC). The resonances at δ_H 8.62 ppm, δ_H 8.37 ppm and ppm $_H$ 6.14 ppm correspond to the protons attached to C-8 and C-2 of the purine ring and C-1 of the ribose ring of dhm^{6}AMP (marked as 36, 32 and 37 in Figure 4.57.C). No other adducts were observed at least during the 12 hours following initial mixing.

4.4.19.2 Effect of change in the sugar ring on the reactivity of nitrogenous base with formaldehyde

AMP (2.4 mM), pre-buffered at pD-7.5, was reacted with a 53 fold excess of formaldehyde and then transferred immediately to a 3mm NMR tube and analysed by ^{1}H-NMR. The first ^{1}H-NMR spectrum was recorded after approximately 5 min following initial mixing. The resonance at δ_H 5.1 ppm was observed to develop after approximately 12 min following initial mixing. The resonance at δ_H 5.4 ppm was observed to develop after approximately 112 min following initial mixing. The initial rate of reaction was calculated to be ~ 6 x10^{-8} mol dm^{-3} s^{-1} (Figure 4.58).

The final concentrations of hm^6(d)AMP and dhm^6(d)AMP after 600 min following initial mixing were approximately ~1.1 mM in both the timecourses. The initial rate of reaction of formation of hm^{6}AMP was almost equal to the formation rates of hm^{6}dAMP. Hence, the presence of additional hydroxyl group at the C-2 of the sugar ring had a subtle/no effect on the reactivity of nucleobase with formaldehyde i.e. the reactivity of a nucleoside with formaldehyde is independent of the presence of additional 2'-hydroxyl group on the sugar ring.

4.4.19.3 Conclusion of the results described in section 4.4.19

In conclusion, the ^{1}H-NMR analyses of reaction mixtures of AMP and formaldehyde reveal the formation of two closely related products. The newly developed resonance at δ_H 5.1 ppm is assigned to the methylene protons attached to the *N*-hydroxymethyl groups of hm^{6}AMP. The resonance at δ_H 5.4 ppm corresponds to the methylene protons of the *N*-hydroxymethyl groups of dhm^{6}AMP. The resonances corresponding to the methylene bridge crosslinked AMP were not observed. The initial rate of reaction of AMP was found to be comparable to that of dAMP, leading to the conclusion that the initial rates of reaction are independent of the presence of 2'- hydroxyl group on the sugar ring, at least for A.

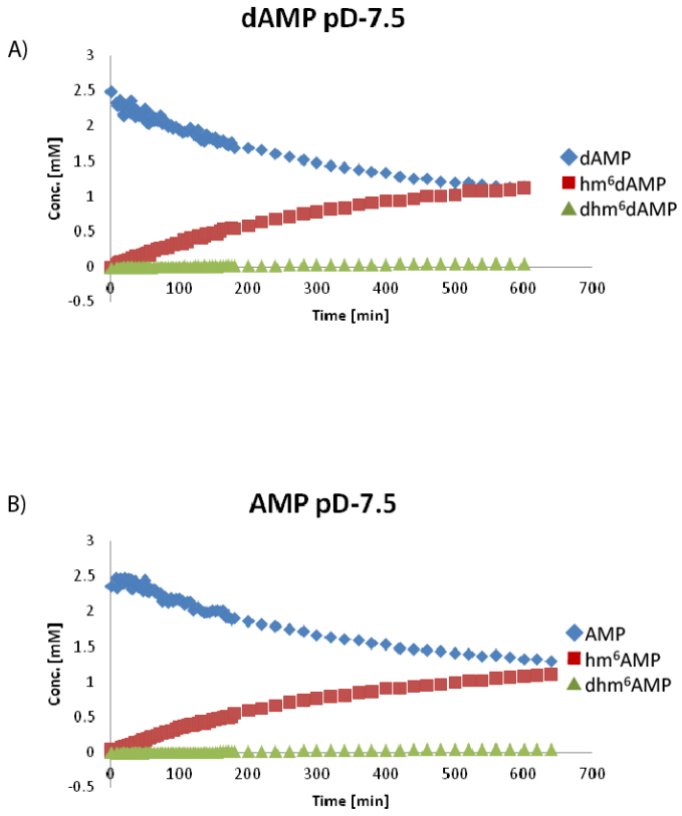


Figure 4.58 Graph showing changes in concentration of A) dAMP (blue), hm^{6}dAMP (red), dhm^{6}dAMP (green) over time at 27°C at pD 7.5. The concentrations are obtained from the integration of the peaks corresponding to the resonance of the proton attached to C-2 of the purine ring of dAMP, hm^{6}dAMP and dhm^{6}dAMP (marked as 1, 11 and 22, respectively, in Figure 4.6); B) AMP (blue), hm^{6}AMP (red), dhm^{6}AMP (green) over time at 27°C at pD 7.5.The concentrations are obtained from the integration of the peaks corresponding to the resonance of the proton attached to C-1 of the ribose ring of AMP, hm^{6}AMP and dhm^{6}AMP (marked as 16, 26 and 37, respectively, in Figure 4.57) and normalising the intensity value relative to internal standard, TSP. ^{1}H NMR timecourses were recorded for 8-10 hours; every 2 min over the first hour, every 4 min over the next 2 hrs and every 20 mins for the next 5-7 hours.

4.4.20 Effect of additional phosphate group on the rate of reaction

The effect of additional phosphate groups attached to A was then investigated. Initially, ^{1}H-NMR spectra of 20 mM A, 40 mM ADP and 40 mM ATP in D_2O were recorded. All the non-exchangeable protons were observed.

4.4.20.1 ^{1}H-NMR analysis of 20 mM adenosine.

The ^{1}H-NMR spectrum of adenosine revealed eight resonances. The resonances at δ_H 8.27 ppm and δ_H 8.18 ppm correspond to the protons attached to C-8 and C-2 (marked as 5 and 1 in Figure 4.59.A) of the purine ring of adenosine. The resonances at δ_H 6.01 ppm, δ_H 4.73 ppm, δ_H 4.36 ppm, δ_H 4.23 ppm, δ_H 3.85 ppm and δ_H 3.77 ppm correspond to the protons attached to C-1, C-2, C-3, C-4, C-5 and C-5 (marked as 6, 7, 8, 9, 10 and 10, respectively, in Figure 4.59.A) of the ribose ring of adenosine.

4.4.20.2 ^{1}H-NMR analysis of 40 mM ADP

The ^{1}H-NMR spectrum of ADP revealed seven resonances. The resonances at δ_H 8.39 ppm and δ_H 8.11 ppm correspond to the protons attached to C-8 and C-2 (marked as 25 and 21 in Figure 4.59.C) of the purine ring of ADP. The resonances at δ_H 5.99 ppm, δ_H 4.6 ppm, δ_H 4.46 ppm, δ_H 4.23 ppm, δ_H 4.11 ppm and δ_H 4.04 ppm correspond to the protons attached to C-1, C-2, C-3, C-4, C-5 and C-5 (marked as 26, 27, 28, 29, 30 and 30, respectively, in Figure 4.59.C) of the ribose ring of ADP.

4.4.20.3 ^{1}H-NMR analysis of 40 mM ATP.

The ^{1}H-NMR spectrum of ATP revealed eight resonances. The resonances at δ_H 8.46 ppm and δ_H 8.17 ppm correspond to the protons attached to C-8 and C-2 (marked as 35 and 31 in Figure 4.59.D) of the purine ring of ATP. The resonances at δ_H 6.05 ppm, δ_H 4.72 ppm, δ_H

4.54 ppm, δ_H 4.34 ppm, δ_H 4.2 ppm and δ_H 4.1 ppm correspond to the protons attached to C-1, C-2, C-3, C-4 and C-5 (marked as 36, 37, 38, 39 and 40, respectively in Figure 4.59.D) of the ribose ring of ADP.

As evident from the Figure 4.4.59, the only observed difference in the ^{1}H-NMR spectra of A, AMP, ADP and ATP was the resonances observed for the protons attached to C-5 of the ribose rings of different nucleoside/ nucleotides; the protons attached to C-5 of the ribose ring of adenosine was observed as dd in ^{1}H-NMR spectrum as compared to multiplets for other nucleotides owing to the coupling of ^{1}H with ^{31}P. The proton attached to C-8 of the purine ring of adenosine was observed to shift downfield with the addition of a phosphate group to the ribose ring. Additional phosphate group was observed to have a subtle effect on the chemical shift of the proton attached to C-8 of the purine ring of AMP.

Excess formaldehyde was added to each solution and ^{1}H-NMR spectra were recorded for the resultant mixtures after 12 hours following initial mixing. The newly developed resonances at δ_H 5.1 ppm and δ_H 5.4 ppm are assigned to the protons attached to the *N*-hydroxymethyl groups of N^6-hydroxymethyladenine and N^6,N^6-di(hydroxymethyl)adenine, respectively and are observed in all the ^{1}H-NMR spectra recorded, independent of the number of phosphate groups attached to ribose ring.

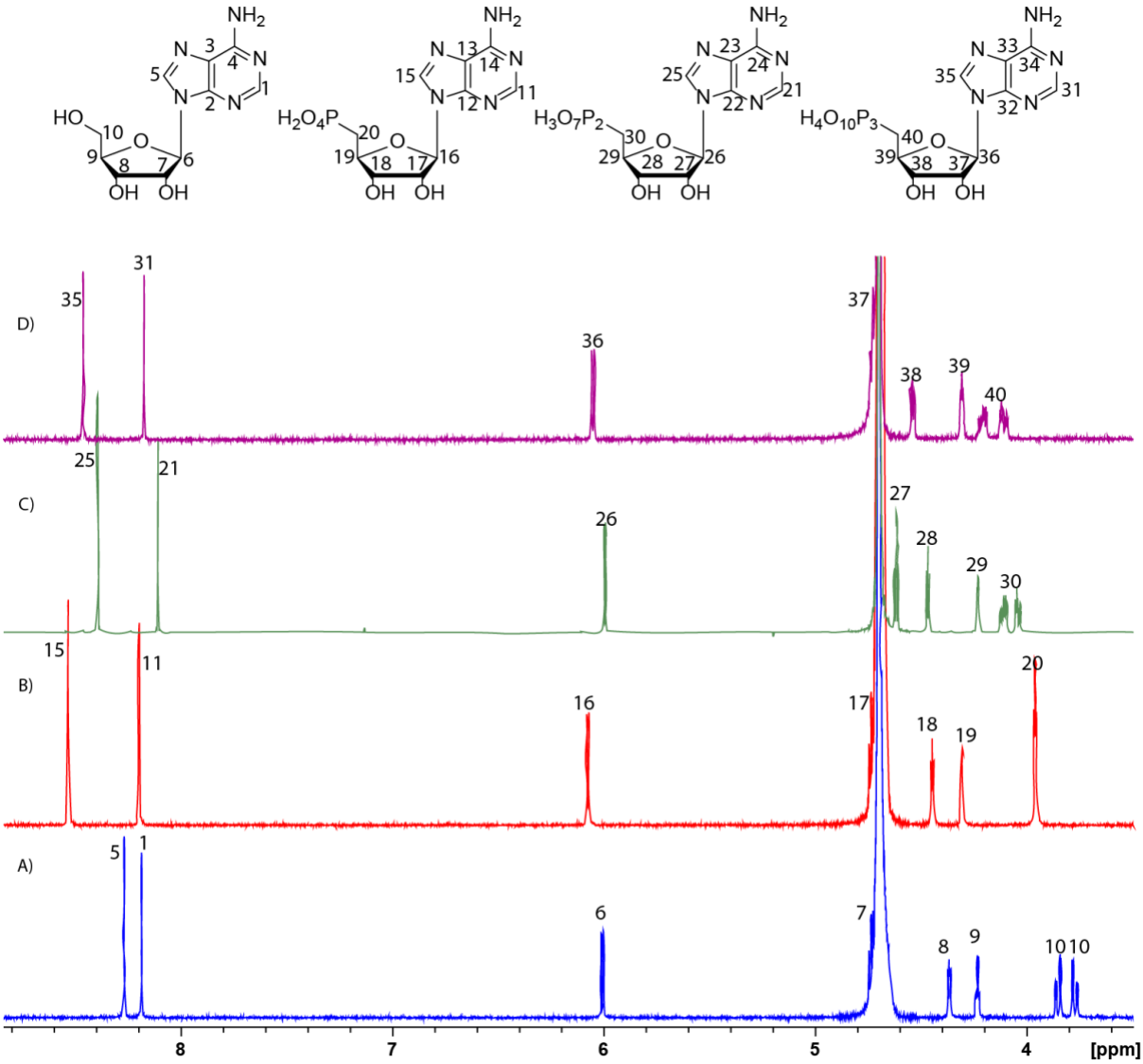


Figure 4.59 A) ^{1}H-NMR (500 MHz) spectrum of a 20 mM adenosine at 27°C, pD 7.5 in D_2O. The ^{1}H assignments of adenosine are indicated. B) ^{1}H-NMR (500 MHz) spectrum of a 40 mM AMP at 27°C, pD 7.5 in D_2O. The ^{1}H assignments of AMP are indicated. C) ^{1}H-NMR (500 MHz) spectrum of a 40 mM ADP at 27°C, pD 7.5 in D_2O. The ^{1}H assignments of ADP are indicated. D) ^{1}H-NMR (500 MHz) spectrum of a 40 mM ATP at 27°C, pD 7.5 in D_2O. The ^{1}H assignments of ATP are indicated.

To study the effect of additional phosphate group on the initial rates of the reaction of formaldehyde with nucleotide, a series of time-course were recorded. 2.4 mM of each of A/ AMP/ ADP/ ATP, pre-buffered at pD-7.5, was reacted with a 53-fold excess of formaldehyde, separately. The mixtures were transferred immediately to a 3mm NMR Match tube to minimise time lapse. The first acquisitions were recorded approximately 5-6 min following initial mixing. The concentrations of hm^{6}AXP and dhm^{6}AXP after 600 min following initial mixing were approximately 0.93 mM and 0.4 mM (hm^{6}A and dhm^{6}A), 0.96

mM and 0.04 mM (hm^6AMP and dhm^6AMP), 0.9 mM and 0.3 mM (hm^6ADP and dhm^6ADP), 0.91 mM and 0.4 mM (hm^6ATP and dhm^6ATP). The initial rates of reaction of A, AMP, ADP and ATP were found to be 7 $x10^{-8}$ mol dm^{-3} s^{-1}, 6 $x10^{-8}$ mol dm^{-3} s^{-1}, 4 $x10^{-8}$ mol dm^{-3} s^{-1} and 2.8 $x10^{-8}$ mol dm^{-3} s^{-1} as calculated from the Figure 4.60. Thus, addition of phosphate groups attached to the ribose ring of adenosine have subtle effects on the initial rate of reaction of formation of major product.

4.4.20.4 Conclusion of the results described in section 4.4.19

In conclusion, the ^{1}H-NMR analysis of mixture of AXP (X=0, M, D, T; when X=0, AXP=A) and formaldehyde revealed the formation of two closely related products when AXP is allowed to react with an excess of formaldehyde. The newly developed resonances at δ_H 5.1 ppm is assigned to the methylene protons attached to the *N*-hydroxymethyl groups of hm^6AXP. The resonance at δ_H 5.4 ppm corresponds to the methylene protons of the *N*-hydroxymethyl groups of dhm^6AXP. The resonances corresponding to the methylene bridge crosslinked AXP were not observed. The initial rates of reaction of AXP were observed to be comparable leading to the conclusion that the initial rates of reaction are almost independent of the presence of additional phosphate group attached to the ribose ring (at least under the studied conditions).

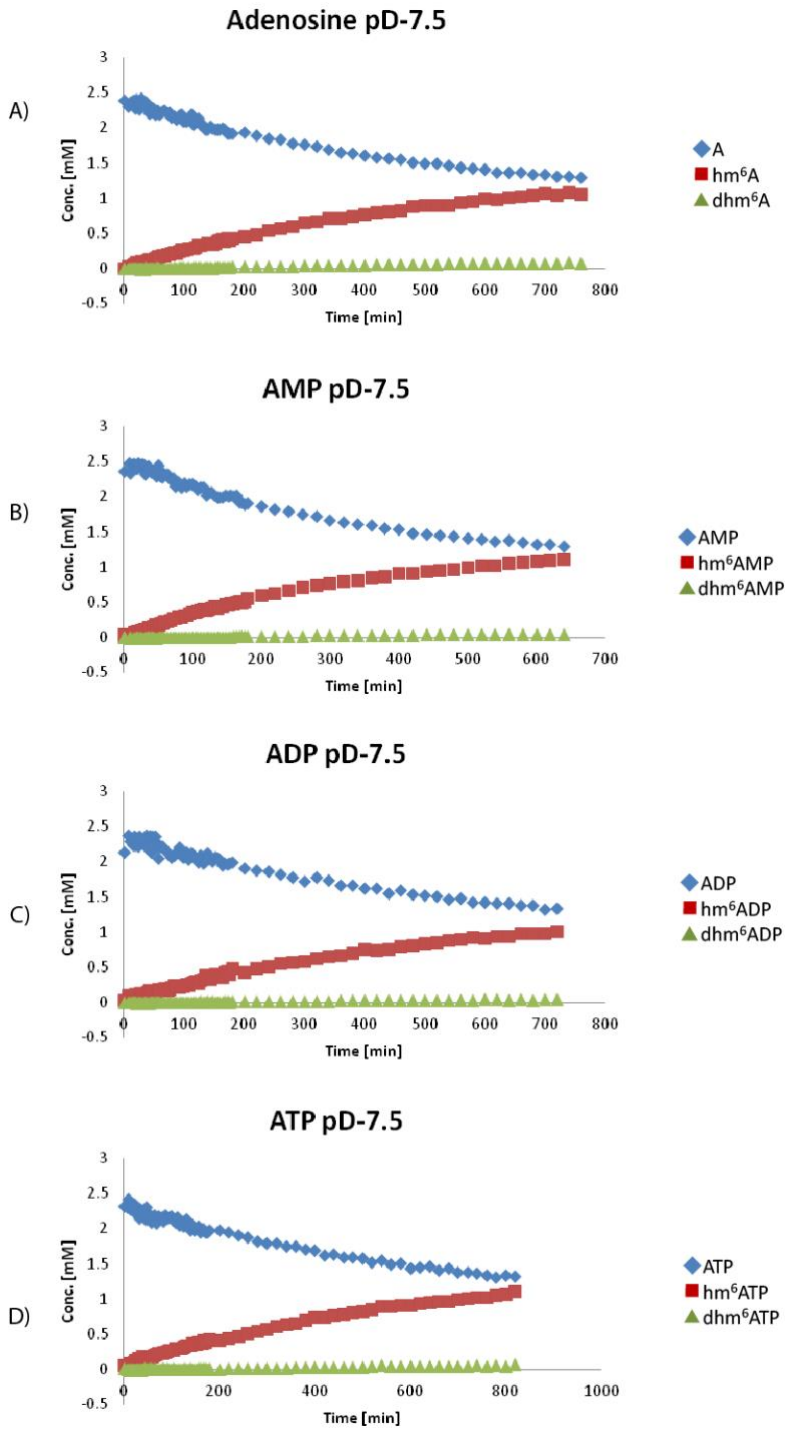


Figure 4.60 Graph showing changes in concentrations (pD 7.5, 27°C) of A) A (blue), hm^{6}A (red), dhm^{6}A (green) over time; B) AMP (blue), hm^{6}AMP (red), dhm^{6}AMP (green) over time; C) ADP (blue), hm^{6}ADP (red), dhm^{6}ADP (green) over time at 27°C at pD 7.5; D) ATP (blue), hm^{6}ATP (red), dhm^{6}ATP (green) over time at 27°C at pD 7.5. The concentrations are obtained from the integration of the peaks corresponding to the resonance of the proton attached to C-1 of the ribose ring of reactant (A/AMP/ADP/ATP), hm^6(A/AMP/ADP/ATP) and dhm^6(A/AMP/ADP/ATP).

4.4.21 Stability of the *N*-hydroxymethyladduct

The stability of the hydroxymethyl adducts formed was then investigated. 16 µl of a 40 mM solution of the nucleotides was allowed to react with an eight-fold excess of formaldehyde. The reaction was allowed to stand for two weeks to give the desired adducts as described. 22.4 µl of a solution was diluted to 24 fold with D_2O and the diluted mixture was transferred immediately to a 5mm NMR tube. The first ^{1}H-NMR spectrum was recorded after approximately 5 min following initial dilution and the time-course was recorded for 8-10 hours following initial dilution.

4.4.21.1 Stability of hm^{6}dAMP with respect to dilution

The concentration of hm^{6}dAMP was observed to decrease from 0.65 mM to 0.60 mM after approximately 500 min following initial dilution and reduced further to 0.55 mM after approximately 700 min following initial dilution. The decrease in the concentration of hm^{6}dAMP was found to be coupled to the increase in the concentration of dAMP. The initial rate of hm^{6}dAMP hydrolysis was to be 2 x10^{-9} mol dm^{-3} s^{-1}.

4.4.21.2 Stability of the hm^{2}dGMP with respect to dilution

The concentration of hm^{2}dGMP decreased from 0.7 mM to 0.2 mM after approximately 500 min following initial dilution. The decrease in the concentration of hm^{2}dGMP was coupled to the increase in the concentration of dGMP. The initial rate of hm^{2}dGMP hydolysis was 7 x10^{-8} mol dm^{-3} s^{-1}. The $t_{1/2}$ was calculated to be approximately 250 min. The rate of hydrolysis of hm^{2}dGMP was faster than hm^{6}dAMP when subjected to dilution.

4.4.21.3 Stability of hm^{4}dCMP with respect to dilution:

The concentration of hm^{4}dCMP decreased from 0.55 mM to 0.2 mM after approximately 800 min following initial dilution. The decrease in the concentration of hm^{4}dAMP was coupled to the increase in the concentration of dCMP. The initial rate of hm^{4}dCMP hydrolysis was found to be 6 x10^{-8} mol dm^{-3} s^{-1}. The $t_{1/2}$ was calculated to be approximately 300 min. It should be noted that the concentration of dCMP and hm^{4}dCMP was constant after 800 min, suggesting the re-establishment of equilibrium at the newly diluted conditions.

4.4.21.4 Stability of hm^{3}TMP with respect to dilution

The resonances corresponding to hm^{3}TMP and hm^{3}UMP were not observed in the first ^{1}H-NMR spectrum recorded approximately within 5 min following initial dilution. Attempts were made to reduce the time-lapse to 3 min, however, ^{1}H resonances corresponding to hm^{3}TMP/ hm^{3}UMP were still not observed in the diluted mixture. Therefore, it was concluded that hm^{3}TMP/hm^{3}UMP are quite unstable to hydrolysis, leading the conversion to TMP/UMP instantaneously on dilution.

4.4.21.5 Conclusion of the section 4.4.21

The most stable *N*-hydroxy adduct was hm^{6}dAMP followed by hm^{4}dCMP and hm^{2}dGMP. As already established, the reaction of formaldehyde and TMP/UMP is kinetically controlled; hm^{3}TMP and hm^{3}UMP instantaneously converted to TMP and UMP, respectively on dilution. Thus, the stability of hm^{6}dAMP compared to other *N*-hydroxy adducts implies the important unreported biological function of hm^{6}dAMP.

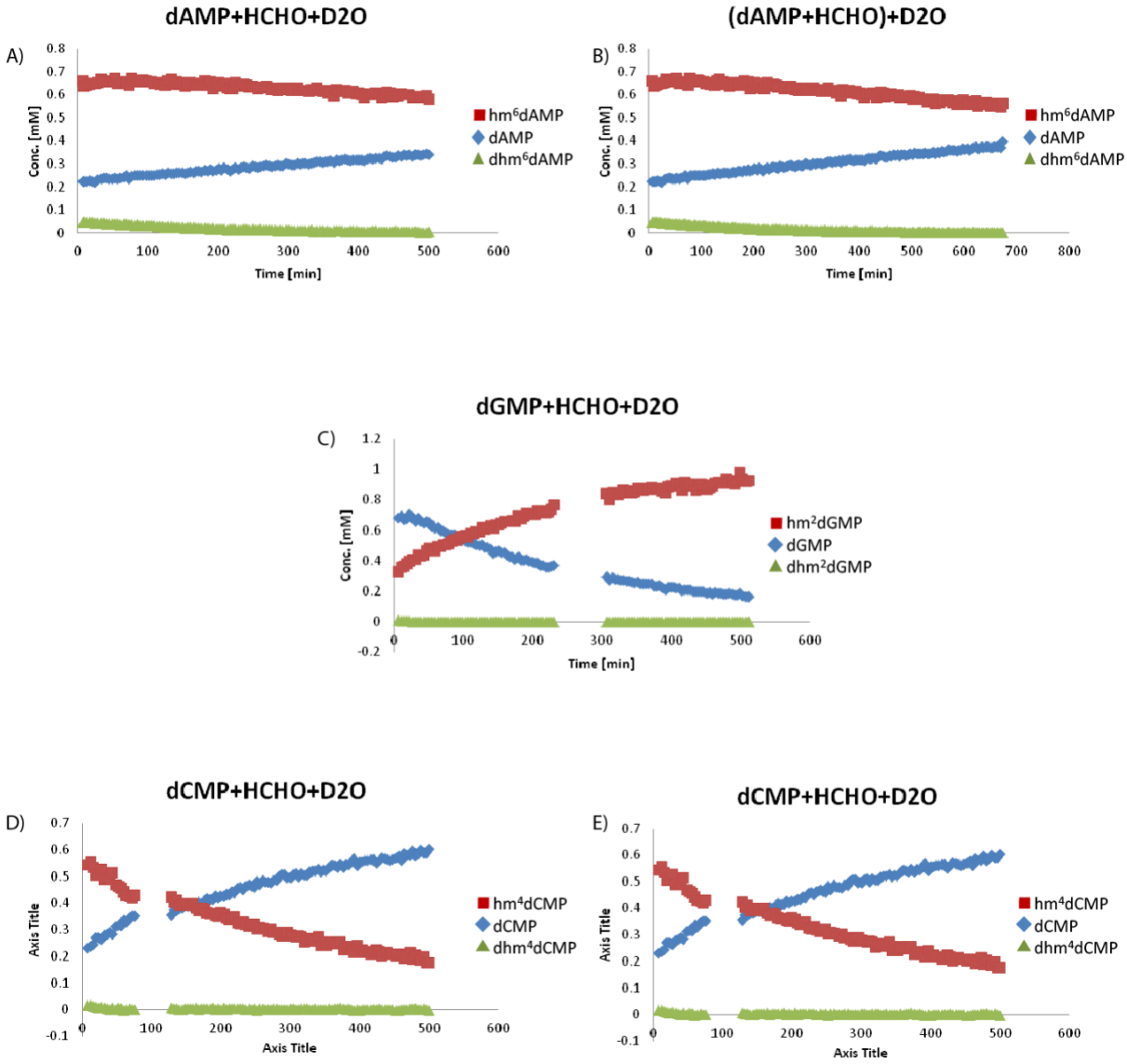


Figure 4.61 Graph showing changes in concentrations (at pD 7.5, 27°C) of A) dAMP, hm^6 dAMP(red) and dhm^6 dAMP (green) over a period of 500 min; B) dAMP, hm^6 dAMP(red) and dhm^6 dAMP (green) over a period of 700 min. C) dGMP, hm^2 dGMP(red) and dhm^2 dGMP (green) over time over a period of 500 min; D) dCMP, hm^4 dCMP(red) and dhm^4 dCMP (green) over time over a period of 500 min; E) dCMP, hm^4 dCMP(red) and dhm^4 dCMP (green) over time over a period of 1300 min. The concentrations are obtained from the integration of the peaks corresponding to the resonances of the proton attached to C-6 of the purine ring of dAMP, hm^{6}dAMP and dhm^6 dAMP; C-1 of the deoxyribose ring of dGMP, hm^2 dGMP and dhm^2 dGMP; C-6 of the pyrimidine ring of dCMP, hm^4 dCMP and dhm^4 dCMP. The intensities of the resonances are normalised relative to internal standard, TSP. ^{1}H NMR timecourses were recorded for 8-10 hours; every 2 min over the first hour, every 4 min over the next 2 hours and every 20 mins for the next 5-7 hours.

4.5 Discussion

The work described in this chapter uses NMR to investigate the reaction of formaldehyde with canonical and non-canonical nucleotides/nucleosides. Important objectives of the work were to investigate the effect of changes in the pH, temperature, phosphate groups attached to the sugar ring, and the 2' alcohol (deoxyribose vs ribose), on the rates of reaction. The reactions of formaldehyde with the modified nucleosides hm^{5}dC, m^{5}dC, inosine, m^{6}G, m^{6}A, m^{1}A and Ψ were also investigated. Competition experiments were carried out to compare the reactivity of hm^{5}dC, m^{5}dC and C with formaldehyde. Competition experiments were also carried out to compare the reactivity of dAMP, dGMP, dCMP, TMP and UMP with formaldehyde. Finally, the stability of the products formed with canonical structure was investigated through dilution which provided insight data into the stability of various adducts. Consistent with the work of Von Hippel and McGhee,[20] two adducts were formed when adenine (dAMP/A/AMP/ADP/ATP) were allowed to react with excess formaldehyde. The monoadduct, hm^{6}A, was the major adduct while the diadduct dhm^{6}A was the minor adduct. No reaction was observed at *N*-1 of adenine, or at the hydroxyl group of the sugar ring, or the phosphate backbone. Further, no methylene bridged crosslinked products were observed. The rate of reaction of formaldehyde appeared largely independent of the pH over the range investigated (pH 5.6-9.0). The rate of the reaction was also independent of the number of additional phosphate groups attached to the sugar ring, at least for adenine, and the presence of additional hydroxyl group on the sugar ring (deoxyribose vs ribose). The initial rate of reaction decreased to almost half when a methyl group was attached to *N*-6 of the adenine ring (i.e m^{6}A). The observed decrease in the reactivity of m^{6}A compared to adenosine might be due to the conformational constraints about *N*-6, as well as sterics. [20] The reaction at *N*-1 of m^{1}A was slowest among all the variants of adenine investigated in the study. The conversion of m^{1}A to product was found to be less than 0.5%; the lowest conversion for all the nucleosides/nucleotides investigated.

Similar to the results for dAMP, two products were observed in the reaction mixture of dGMP and formaldehyde, hm^2dGMP and dhm^2dGMP. No reaction at *N*-1 of Guanosine was observed. However, the crosslinked product dG-CH_2-dG was observed after prolonged incubation. The initial reaction rate of formaldehyde with dGMP to form hm^2dGMP was 1.15 x 10^{-7} mol dm^{-3} s^{-1} at pD 7.5, two fold that of dAMP and formaldehyde to form hm^6dAMP.

Unlike with dAMP and dGMP, three products were initially observed in the reaction mixture of dCMP and formaldehyde; however, two rotameric forms of hm^4dCMP were present, and therefore, only two structurally different products were formed, hm^4dCMP and dhm^4dCMP. hm^4dCMP has not been observed in two rotameric forms in the previous studies. It should be noted that the two *N*-hydroxymethyl groups of dhm^4dCMP are established as magnetically non-equivalent. ^{1}H-NMR resonances corresponding to any methylene bridge crosslinked species were not observed even when the reaction mixture was analysed after 1 week. The initial rate of reaction of dCMP with formaldehyde to form hm^4dCMP was higher than the other exocyclic amine nucleotides studied. The concentration of hm^4dCMP was ~1.3 fold higher than the concentration of hm^2dGMP and almost ~1.5 fold higher than the concentration of hm^6dAMP after approximately 500 min following initial mixing; indicating that dCMP is the most reactive nucleotide with formaldehyde. The reactions of hm^5dC and m^5dC with formaldehyde were then investigated under similar conditions. The formation of monoadducts (hm^4hm^5dC and hm^4m^5dC) and diadducts (dhm^4hm^5dC and dhm^4m^5dC) were observed for both the nucleosides. The formation of rotamers was observed for these nucleosides. However, the concentration of minor rotameric form was approximately 0.5% of the major rotameric forms. Competition experiment among C, hm^5dC and m^5dC revealed that the initial rate of reaction of formaldehyde with m^5dC was the fastest; closely followed by hm^5dC then C. Although, the initial rates of reaction of hm^5dC and C were observed to be roughly equal; the conversion of C to the *N*-hydroxy product was observed to be higher than

hm^5dC and m^5dC conversion. Hence, hm^5dC reaches equilibrium faster than the other two investigated nucleosides.

The reaction of formaldehyde with TMP and UMP resulted in single observed products, hm^3TMP and hm^3UMP, respectively. Crosslinked bridged methylene products were not observed. The rate of reaction of these nucleosides were the fastest among all nucleosides investigated. However, the conversion to their respective product was ~ 20%. The rate of reaction was also pH dependent, with slowest rate at pD-6.0 (pH-5.6). Hence, the results suggest that reactions are base catalysed.

Interestingly, the reaction of Ψ with formaldehyde resulted in three different products, namely $hm^3Ψ$, hm^5 Ψ and hm^3 $hm^5Ψ$. The conversion of Ψ to hm^5 Ψ was slightly higher than that to hm^5 Ψ. The rate of reaction was equivalent to TMP and UMP; no change in the concentration of products were observed after 3 min following initial mixing. Considering the abundance and location of Ψ in t-RNA, the reaction of Ψ with formaldehyde could play important roles in mutations observed in proteins leading to unprecedented role of protein or origin of disease.

In conclusion, the reaction of fourteen different nucleosides / nucleotides with formaldehyde was investigated. The effect of changes in pD, temperature, phosphate groups and 2'-hydroxyl groups was investigated. Reaction with TMP, UMP, I and Ψ were faster than reactons with AMP, GMP, CMP, m^6A, hm^5C, m^5C and m^6G. The initial rates of reaction for m^5dC were the highest amongst the exocyclic amines followed by C (dC/dCMP) > hm^5dC> m^6G> dGMP> dAMP (A/AMP/ ADP /ATP)> m^6A> m^1A. Consistent with the work of McGhee and Von Hippel,[19,20] the rates of reaction of formaldehyde with nucleotides with exocyclic amines were independent of changes in pH in the range of pD 6.0 to pD 9.0, however, the rate of reaction of endocyclic imines (TMP and UMP) were slower at pD- 6.0. The unprecedented stability of hm^6dAMP compared to other *N*-hydroxymethyl adducts strongly supports the hypothesis that hm^6dAMP has an important biological function as it is

formed as an intermediate in FTO enzymatic catalysis. The possibility of formaldehyde forming inter-strand DNA-DNA crosslinks is unlikely in vivo mainly because of steric hindrance.[84] Several studies have to failed to prove DNA-DNA crosslinking upon incubation with formaldehyde.[85,86]

4.6 References

1. IARC Monographs on evaluation of carcinogenic risks to humans. *International Agency for research on Cancer. Vol 62* 217–245 (1995). Available at: http://monographs.iarc.fr/ENG/Monographs/vol62/. (Accessed: 24th March 2017)
2. Byerrium, R. U., Fairley, J. L., Hamill, R. L. & Herrmann, R. L. The synthesis of purines and thymine from formaldehyde in the rat. *Biochim. Biophys. Acta* **21,** 394–5 (1956).
3. Robertst, D., Nichol, C. A. Biosynthesis of Thymine and Formyltetrahydrofolate by Vitamin B,-deficient Cells of Lactobacillus Zeichmannii. *J.Biol.Chem,* **237,** (1962).
4. Fox, S. W. & Windsor, C. R. Synthesis of amino acids by the heating of formaldehyde and ammonia. *Science* **170,** 984–6 (1970).
5. WHO Task Group on Environmental Health Criteria for Formaldehyde., United Nations Environment Programme., International Labour Organisation., World Health Organization. & International Program on Chemical Safety. *Formaldehyde.* (World Health Organization, 1989).
6. Trocho, C., Pardo, R., Rafecas, I., Virgili, J., Remesar, X., Fernández-López, J. A., Alemany, M. Formaldehyde derived from dietary aspartame binds to tissue components in vivo. *Life Sci.* **63,** 337–49 (1998).
7. Jensen, R. P., Luo, W., Pankow, J. F., Strongin, R. M. & Peyton, D. H. Hidden Formaldehyde in E-Cigarette Aerosols. *N. Engl. J. Med.* **372,** 392–394 (2015).
8. IARC Monographs - Monographs available in PDF format. *A review of human carcinogens: chemical agents and related occupations* 401–435 (2012). Available at: http://monographs.iarc.fr/ENG/Monographs/vol100F/. (Accessed: 17th March 2017)
9. Jia, G. H., Timothy F., Hausinger, R.P., Lindahl, T., Sedgwick, B. Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Lett.* **582,** 3313–3319 (2008).
10. Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., Vågbø, C. B., Shi, Y., Wang, W. L., Song, S. H., Lu, Z., Bosmans, R. P.G. Dai, Q., Hao, Ya J., Yang, X., Zhao, Wen M., Tong, Wei M., Wang, Xiu J., Bogdan, F., Furu, K., Fu, Y., Jia, G., Zhao, X.,Liu, J., Krokan, H. E., Klungland, A., He, C. ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Mol. Cell* **49,** 18–29
11. Cloos, P. A. C., Christensen, J., Agger, K. & Helin, K. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. doi:10.1101/gad.1652908
12. Loenarz, C. & Schofield, C. J. Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat. Chem. Biol.* **4,** 152–156 (2008).
13. Hazen, S. L., D'Avignon, A., Anderson, M. M., Hsu, F. F. & Heinecke, J. W. Human neutrophils employ the myeloperoxidase-hydrogen peroxide-chloride system to oxidize -amino acids to a family of reactive aldehydes: Mechanistic studies identifying

labile intermediates along the reaction pathway. *J. Biol. Chem.* **273,** 4997–5005 (1998).
14. Metz, B., Kersten, G. F.A., Baart, G. J.E., De Jong, A., Meiring, H., Ten Hove, J., Van Steenbergen, M. J., Hennink, W. E., Crommelin, D. J.A. Jiskoot, Wim Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. *J. Biol. Chem.* **279,** 6235–43 (2004).
15. Lu, K., Boysen, G., Gao, L., Collins, L. B. & Swenberg, J. A. Formaldehyde-Induced Histone Modifications *in Vitro*. *Chem. Res. Toxicol.* **21,** 1586–1593 (2008).
16. Siomin, Y. A., Simonov, V. V & Poverenny, A. M. The reaction of formaldehyde with deoxynucleotides and DNA in the presence of amino acids and lysine-rich histone. *Biochim. Biophys. Acta* **331,** 27–32 (1973).
17. Holde, V., McGheel, J. D. & von Hippel, P. H. Formaldehyde as a Probe of DNA Structure. 4. Mechanism of the Initial Reaction of Formaldehyde with DNAt. *Biopolym. Biochem. J . Mol. Biol. J . Mol.. Chem. Biol. J . Mol. Biol* **13,** 815–726 (1972).
18. McGhee, J. D. & Von Hippel, P. H. Formaldehyde as a probe of DNA structure. 3. Equilibrium denaturation of DNA and synthetic polynucleotides. *Biochemistry* **16,** 3267–3276 (1977).
19. Mcghee, J. D. & Von Hippel, P. H. Formaldehyde as a Probe of DNA Structure. 11. Reaction with Endocyclic Imino Groups of DNA Basest.
20. McGhee, J. D. & Von Hippel, P. H. Formaldehyde as a probe of DNA structure. I. Reaction with exocyclic amino groups of DNA bases. *Biochemistry* **14,** 1281–1296 (1975).
21. Lu, K. Structural Characterization of Formaldehyde-Induced Cross-Links Between Amino Acids and Deoxynucleosides and Their Oligomers. *J. Am. Chem. Soc.* **132,** 3388–3399 (2010).
22. Lu, K., Ye, W., Gold, A., Ball, L. M. & Swenberg, J. A. Formation of *S*- [1-(N^2 -Deoxyguanosinyl)methyl]glutathione between Glutathione and DNA Induced by Formaldehyde. *J. Am. Chem. Soc.* **131,** 3414–3415 (2009).
23. Cheng, G., Wang, M., Upadhyaya, P., Villalta, P. W. & Hecht, S. S. Formation of Formaldehyde Adducts in the Reactions of DNA and Deoxyribonucleosides with α-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 (2008).
24. Chang, Y.-T. & Loew, G. H. Reaction Mechanisms of Formaldehyde with Endocyclic Imino Groups of Nucleic Acid Bases. *J. Am. Chem. Soc.* **116,** 3548–3555 (1994).
25. Chaw, Y. F. M., Crane, L. E., Lange, P. & Shapiro, R. Isolation and identification of cross-links from formaldehyde-treated nucleic acids. *Biochemistry* **19,** 5525–5531 (1980).
26. Speit, G., Schütz, P. & Merk, O. Induction and repair of formaldehyde-induced DNA-protein crosslinks in repair-deficient human cell lines. *Mutagenesis* **15,** 85–90 (2000).
27. Solomon, M. J. & Varshavsky, A. Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc. Natl. Acad. Sci. U. S. A.* **82,** 6470–4 (1985).
28. Merk, O. & Speit, G. Significance of formaldehyde-induced DNA-protein crosslinks for mutagenesis. *Environ. Mol. Mutagen.* **32,** 260–8 (1998).
29. Huang, H., Solomon, M. S. & Hopkins, P. B. Formaldehyde preferentially interstrand cross-links duplex DNA through deoxyadenosine residues at the sequence 5'-d(AT). *J. Am. Chem. Soc.* **114,** 9240–9241 (1992).
30. Huang, H. & Hopkins, P. B. 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 (1993).
31. Natarajan, A. T., Darroudi, F., Bussman, C. J. & van Kesteren-van Leeuwen, A. C. Evaluation of the mutagenicity of formaldehyde in mammalian cytogenetic assays in vivo and vitro. *Mutat. Res.* **122,** 355–60 (1983).
32. Ragan, D. L. & Boreiko, C. J. Initiation of C3H/10T1/2 cell transformation by formaldehyde. *Cancer Lett.* **13,** 325–31 (1981).
33. Goldmacher, V. S. & Thilly, W. G. Formaldehyde is mutagenic for cultured human cells. *Mutat. Res.* **116,** 417–22 (1983).
34. Auerbach, C., Moutschen-Dahmen, M. & Moutschen, J. Genetic and cytogenetical effects of formaldehyde and related compounds. *Mutat. Res.* **39,** 317–61 (1977).
35. Friedenson, B. A common environmental carcinogen unduly affects carriers of cancer mutations: Carriers of genetic mutations in a specific protective response are more susceptible to an environmental carcinogen. *Med. Hypotheses* **77,** 791–797 (2011).
36. Swenberg, J. A., Kerns, W. D., Mitchell, R. I., Gralla, E. J. & Pavkov, K. L. Induction of squamous cell carcinomas of the rat nasal cavity by inhalation exposure to formaldehyde vapor. *Cancer Res.* **40,** 3398–402 (1980).
37. Monticello, T. M.Correlation of regional and nonlinear formaldehyde-induced nasal cancer with proliferating populations of cells. *Cancer Res.* **56,** 1012–22 (1996).
38. Kerns, W. D., Pavkov, K. L., Donofrio, D. J., Gralla, E. J. & Swenberg, J. A. Carcinogenicity of formaldehyde in rats and mice after long-term inhalation exposure. *Cancer Res.* **43,** 4382–92 (1983).
39. Heck, H. d'A. & Casanova, M. The implausibility of leukemia induction by formaldehyde: a critical review of the biological evidence on distant-site toxicity. *Regul. Toxicol. Pharmacol.* **40,** 92–106 (2004).
40. Tong, Z., Zhang, J., Luo, W., Wang, W., Li, F., Li, H., Luo, H., Lu, J., Zhou, J., Wan, Y., He, R. Urine formaldehyde level is inversely correlated to mini mental state examination scores in senile dementia. *Neurobiol. Aging* **32,** 31–41 (2011).
41. Tong, Z., Han, C., Luo, W., Wang, X., Li, H., Luo, H., Zhou, J., Qi, J., He, R. Accumulated hippocampal formaldehyde induces age-dependent memory decline. doi:10.1007/s11357-012-9388-8
42. Pontel, L. B. Endogenous Formaldehyde Is a Hematopoietic Stem Cell Genotoxin and Metabolic Carcinogen. *Mol. Cell* **60,** 177–188 (2015).
43. Wang, M., Cheng, G., Balbo, S., Carmella, S. G., Villalta, P. W., Hecht, S. S. Clear Differences in Levels of a Formaldehyde-DNA Adduct in Leukocytes of Smokers and Non-smokers. doi:10.1158/0008-5472.CAN-09-1571
44. Thompson, C. M., Ceder, R. & Grafström, R. C. Formaldehyde dehydrogenase: Beyond phase I metabolism. *Toxicol. Lett.* **193,** 1–3 (2010).
45. Staab, C. A., Ålander, J., Morgenstern, R., Grafström, R. C. & Höög, J.-O. The Janus face of alcohol dehydrogenase 3. *Chem. Biol. Interact.* **178,** 29–35 (2009).
46. MacAllister, S. L., Choi, J., Dedina, L. & O'Brien, P. J. Metabolic mechanisms of methanol/formaldehyde in isolated rat hepatocytes: Carbonyl-metabolizing enzymes versus oxidative stress. *Chem. Biol. Interact.* **191,** 308–314 (2011).
47. Boedtker, H. The reaction of ribonucleic acid with formaldehyde. I. Optical absorbance studies. *Biochemistry* **6,** 2718–27 (1967).
48. Eyring, E. J. & Ofengand, J. Reaction of formaldehyde with heteroocyclic imino nitrogen of purine and pyrimidine nucleosides. *Biochemistry* **6,** 2500–6 (1967).
49. GROSSMAN, L., LEVINE, S. S. & ALLISON, W. S. The reaction of formaldehyde with nucleotides and T2 bacteriophage DNA. *J. Mol. Biol.* **3,** 47–60 (1961).
50. Staehelin, M. REACTION OF TOBACCO MOSAIC VIRUS NUCLEIC ACID WITH FORMALDEHYDE*. *Biochim. Biophys. ACTA VOL* **29,** (1958).
51. Fraenkel-Conrat, H. Reaction of nucleic acid with formaldehyde. *Biochim. Biophys.*

Acta **15,** 307–309 (1954).
52. Chaw, Y. F., Crane, L. E., Lange, P. & Shapiro, R. Isolation and identification of cross-links from formaldehyde-treated nucleic acids. *Biochemistry* **19,** 5525–31 (1980).
53. Feldman, M. Y. Reactions of nucleic acids and nucleoproteins with formaldehyde. *Prog. Nucleic Acid Res. Mol. Biol.* **13,** 1–49 (1973).
54. FREDERICK A. BELAND*, N. F. F. and R. H. RAPID ISOLATION, HYDROLYSIS AND CHROMATOGRAPHY OF FORMALDEHYDE-MODIFIED DNA. *J. Chromatogr.* **308,** 121–131 (1984).
55. Rivlin, M., Eliav, U. & Navon, G. NMR studies of the equilibria and reaction rates in aqueous solutions of formaldehyde. *J. Phys. Chem. B* **119,** 4479–4487 (2015).
56. Roundtree, I. A. & He, C. RNA epigenetics—chemical messages for posttranscriptional gene regulation. *Curr. Opin. Chem. Biol.* **30,** 46–51 (2016).
57. He, C. Grand Challenge Commentary: RNA epigenetics? *Nat. Chem. Biol.* **6,** 863–865 (2010).
58. Ping, X.-L. S., Bao F., Wang, L., Xiao, W., Yang, X., Wang, W. J., Adhikari, S., Shi, Y., Lv, Y., Chen, Yu S., Zhao, X., Li, A., Yang, Y., Dahal, U., Lou, Xiao M., Liu, X., Huang, J., Yuan, Wei P., Zhu, Xiao F., Cheng, T., Zhao, Yong L., Wang, X., Danielsen, Jannie M.R., Liu, Feng., Yun G. Mammalian WTAP is a regulatory subunit of the RNA N6 methyladenosine methyltransferase. *Cell Res.* **24,** 177–189 (2014).
59. Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., Dai, Q., Chen, W., He, C. A METTL3–METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* **10,** 93–95 (2013).
60. Zhang, Z. The YTH domain is a novel RNA binding domain. *J. Biol. Chem.* **285,** 14701–14710 (2010).
61. Han, Z., Niu, T., Chang, J., Lei, X., Zhao, M., Wang, Q., Cheng, W., Wang, J., Feng, Y., Chai, J. Crystal structure of the FTO protein reveals basis for its substrate specificity. *Nature* **464,** 1205–1209 (2010).
62. Gerken, T. Girard, Christophe A., Tung, Y. C. L., Webby, C. J., Saudek, V., Hewitson, K. S., Yeo, Giles S.H., McDonough, M. A., Cunliffe, S., McNeill, L. A., Galvanovskis, J., Rorsman, P., Robins, P., Prieur, X., Coll, A. P., Ma, M., Jovanovic, Z., Farooqi, I. S., Sedgwick, B., Barroso, I., Lindahl, T., Ponting, C. P., Ashcroft, Fr. M., O'Rahilly, S., Schofield, C. J. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318,** 1469–72 (2007).
63. Zheng, G. Dahl, John A., Niu, Y., Fedorcsak, P., F., Furu, K., Fu, Y., Jia, G., Zhao, X., Liu, J., Krokan, Hans E., Klungland, A., Yang, Yun-Gui, He, C. ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Mol. Cell* **49,** (2013).
64. Newbold, R. F., Warren, W., Medcalf, A. S. C. & Amos, J. Mutagenicity of carcinogenic methylating agents is associated with a specific DNA modification. *Nature* **283,** 596–599 (1980).
65. LOVELESS, A. Possible Relevance of O–6 Alkylation of Deoxyguanosine to the Mutagenicity and Carcinogenicity of Nitrosamines and Nitrosamides. *Nature* **223,** 206–207 (1969).
66. Green, C. L., Loechler, E. L., Fowlert, K. W. & Essigmann, J. M. Construction and characterization of extrachromosomal probes for mutagenesis by carcinogens: Site-specific incorporation of 06-methylguanine into viral and plasmid genomes (chemical carcinogenesis/mutagenesis/O'-alkylguanine/carcinogen-DNA adducts). **81,** 13–17 (1984).
67. Edward L. Loechler, Calvert L. Green, A. J. M. Essigmann. In vivo mutagenesis by 06-methylguanine built into a unique site in a viral geno'me. *Proc. Natl. Acad. Sci.*

USA **81,** 6271–6275, (1984).
68. and, A. M. W. & Guengerich, F. P. Effect of the O6 Substituent on Misincorporation Kinetics Catalyzed by DNA Polymerases at O6-Methylguanine and O6-Benzylguanine. (2001). doi:10.1021/BI011495N
69. Singer, B., Chavezt, F., Goodmant, M. F., Essigmann, J. M. & Dosanjht, M. K. Effect of 3' flanking neighbors on kinetics of pairing of dCTP or dTTP opposite 06-methylguanine in a defined primed oligonucleotide when Escherichia coli DNA polymerase I is used (fidelity/mutation/base pairing/Drosophila DNA polymerase a). *Biochemistry* **86,** 8271–8274 (1989).
70. Haracska, L., Prakash, S. & Prakash, L. Replication past O(6)-methylguanine by yeast and human DNA polymerase eta. *Mol. Cell. Biol.* **20,** 8001–7 (2000).
71. Yasui, M. Suenaga, E., Koyama, N., Masutani, C., Hanaoka, F., Gruz, P., Shibutani, S., Nohmi, T., Hayashi, M., Honma, M. Miscoding Properties of 2′-Deoxyinosine, a Nitric Oxide-Derived DNA Adduct, during Translesion Synthesis Catalyzed by Human DNA Polymerases. *J. Mol. Biol.* **377,** 1015–1023 (2008).
72. Nishikura, K. Functions and Regulation of RNA Editing by ADAR Deaminases. *Annu. Rev. Biochem.* **79,** 321–349 (2010).
73. Nishikura, K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Publ. Gr.* **17,** (2015).
74. Ito, S. Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., Zhang, Y. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science (80).* **333,** 1300–1303 (2011).
75. Freudenberg, J. M., Ghosh, S., Lackford, B. L., Yellaboina, S., Zheng, X., Li, R., Cuddapah, S., Wade, Paul A., Hu, G., Jothi, R. Acute depletion of Tet1-dependent 5-hydroxymethylcytosine levels impairs LIF/Stat3 signaling and results in loss of embryonic stem cell identity. *Nucleic Acids Res.* **40,** 3364–3377 (2012).
76. Jin, S.-G. Jiang, Y., Qiu, R., Rauch, T. A., Wang, Y., Schackert, G., Krex, D., Lu, Q., Pfeifer, G. P. 5-Hydroxymethylcytosine Is Strongly Depleted in Human Cancers but Its Levels Do Not Correlate with IDH1 Mutations. *Cancer Res.* **71,** 7360–7365 (2011).
77. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **10,** 295–304 (2009).
78. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9,** 465–476 (2008).
79. Bestor, T. H. Coxon, A., Szeto, J., McNoe, L.A., Yun, K., Maw, M.A., Smith, P.J., Reeve, A.E., Miyamoto, M., Tawa, R., Hirose, S., Okada, S. Cytosine methylation: the pros and cons of DNA methylation. *Curr. Biol.* **3,** 384–6 (1993).
80. Lane, B. G. in *Modification and Editing of RNA* 1–20 (American Society of Microbiology, 1998). doi:10.1128/9781555818296.ch1
81. DAVIS, F. F. & ALLEN, F. W. Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* **227,** 907–15 (1957).
82. COHN, W. E. & VOLKIN, E. Nucleoside-5′-Phosphates from Ribonucleic Acid. *Nature* **167,** 483–484 (1951).
83. Hamma, T., Ferre-D-Amare, R. A. Pseudouridine Synthases. *Chem. Biol.* **13,** 1125–1135 (2006).
84. Bedford, P. & Fox, B. W. The role of formaldehyde in methylene dimethanesulphonate-induced DNA cross-links and its relevance to cytotoxicity. *Chem. Biol. Interact.* **38,** 119–126 (1981).
85. Ross, W. E., McMillan, D. R. & Ross, C. F. Comparison of DNA damage by methylmelamines and formaldehyde. *J. Natl. Cancer Inst.* **67,** 217–21 (1981).
86. Ross, W. E. & Shipley, N. Relationship between DNA damage and survival in formaldehyde-treated mouse cells. *Mutat. Res.* **79,** 277–83 (1980).

Chapter 5: Mechanistic studies on FTO using carbon-13 labelled substrates as probes

5.1 Introduction

The *Escherichia coli* enzyme AlkB and the human homologues of AlkB (AlkBHs) make up a subfamily of 2-oxoglutarate (2OG) dependent oxygenases that are *N*-methyl nucleic acid demethylases. Like other 2OG oxygenases, they require non-heme Fe(II) as a cofactor.[1] 2OG oxygenases use 2-oxoglutarate (2OG) and molecular oxygen (O_2) to generate a high-valent iron(IV)-oxo intermediate (as well as the by-products carbon dioxide and succinate), which is used to catalyse reactions of high biological importance.[2]

The fat mass and obesity associated protein (FTO) is a human AlkB homologue, mutations to the gene of which has been linked to obesity.[3–5] The reported substrates for FTO are N^3-methylthymidine (m^3T),[6] N^3-methyluridine (m^3U),[6] N^6-methyladenosine (m^6A),[7] and N^6,O^2-dimethyladenosine (m^6Am)[8] (Figure 5.1C). FTO is also reported to catalyse the conversion of N^6-hydroxymethyladenosine (hm^6A) to N^6-formyladenosine (f^6A), in a two electron second oxidation of m^6A (Figure 5.1.A).[9]

AlkB which is found in many bacteria, catalyses the oxidative demethylation of a broader range of substrates than FTO. Reported AlkB substrates include N^1-methyladenosine (m^1A),[10] m^3T,[11] m^6A,[12] N^1-methylguanosine (m^1G),[11] N^2-methylguanosine (m^2G),[13] N^3-methylcytidine (m^3C),[14] and N^4-methylcytidine (m^4C)[13] (Figure 5.1.C). AlkB is also reported to act on more complex DNA lesions, e.g. 1,N^6-ethenoadenosine (εA),[15] and to have a preference for positively charged substrates at biological pH.[11]

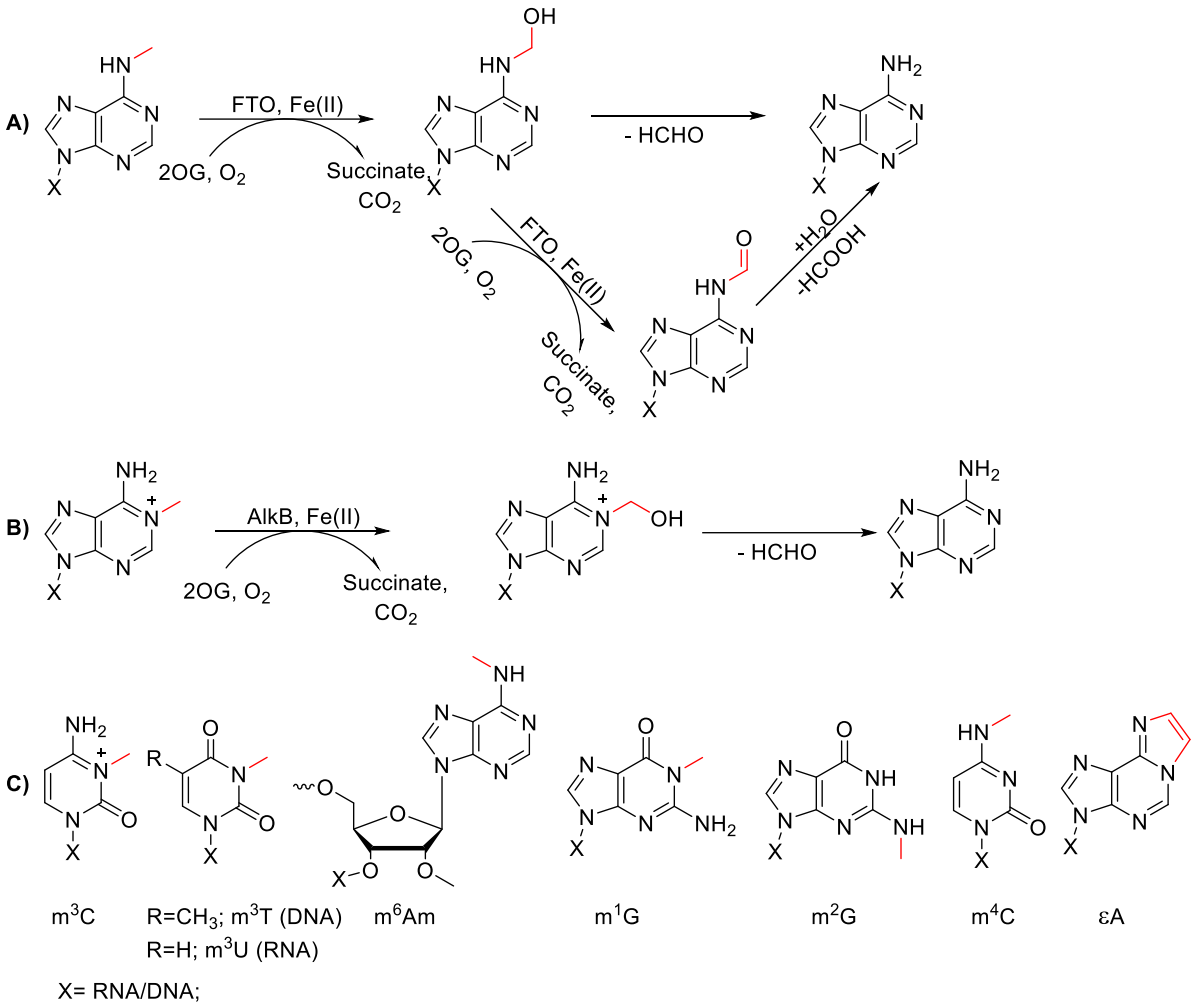


Figure 5.1 Schematic representation of the reaction of A) FTO and B) AlkB. C) Other reported substrates of FTO include m^3U, m^3T, m^6Am, and for AlkB, m^3C, m^3U, m^3T, m^1G, m^2G, m^4C, and εA. The methyl groups undergoing modification are indicated in red.

The consistent role of 2OG, O_2, CO_2, and succinate as co-substrates/co-products in the catalysis of 2OG oxygenases has enabled development of generic assays for these enzymes.[16] Generic 2OG oxygenase assays are mainly based on the measurement of isotopically labelled co-substrates or co-products. Assays based on CO_2 measurement have quantified the amount of $^{14}CO_2$ released during decarboxylation of 2-oxo-[1-^{14}C]glutarate.[17] Assays based on succinate measurement have involved separation of [1-^{14}C]succinate from 2-oxo-[5-^{14}C]glutarate by ion exchange chromatography,[18] determining the concentration of [1-

^{14}C]succinate after precipitation of 2-oxo-[5-^{14}C]glutarate with 2,4-dinitrophenyl hydrazine,[19] or measuring the concentration of succinate by LCMS.[20] Assays based on 2OG depletion have included a fluorescence based assay in which residual 2OG is derivatised to yield a fluorescent product[21] and spectrophotometric assays focussing on the measurement of 2OG remaining in the assay solution.[17] In some of these assays, interference from other components renders them non-robust. In general, radioactivity-based assays are likely more reliable than the fluorescence-based assays, as they give fewer false positives. They are also often less laborious compared with mass spectrometry (MS) based assays, although MS based assays are often considered the most reliable. However, assays based on succinate/2OG measurement suffer from the drawback that they equate the consumption/production of 2OG/succinate to product formation, which is not always the case. The conversion of 2OG to succinate can occur non-enzymatically, and enzymatic conversion of 2OG is not always fully coupled to product formation.[21]

Assays that are specific for 2OG oxygenases generally measure the quantity of released hydroxylated/demethylated product or co-product, often by MS,[22,23] NMR,[24] or antibody-based methods using luminescence/fluorescence-based methodology.[25] In one variant of the formaldehyde release assay, a fluorescence based method is used in which formaldehyde dehydrogenase (FDH) converts formaldehyde to formic acid with concomitant reduction of NAD^+ to NADH; the production of NADH is monitored by fluorescence spectrometry.[26] However, this indirect measurement of formaldehyde can lead to artefacts.

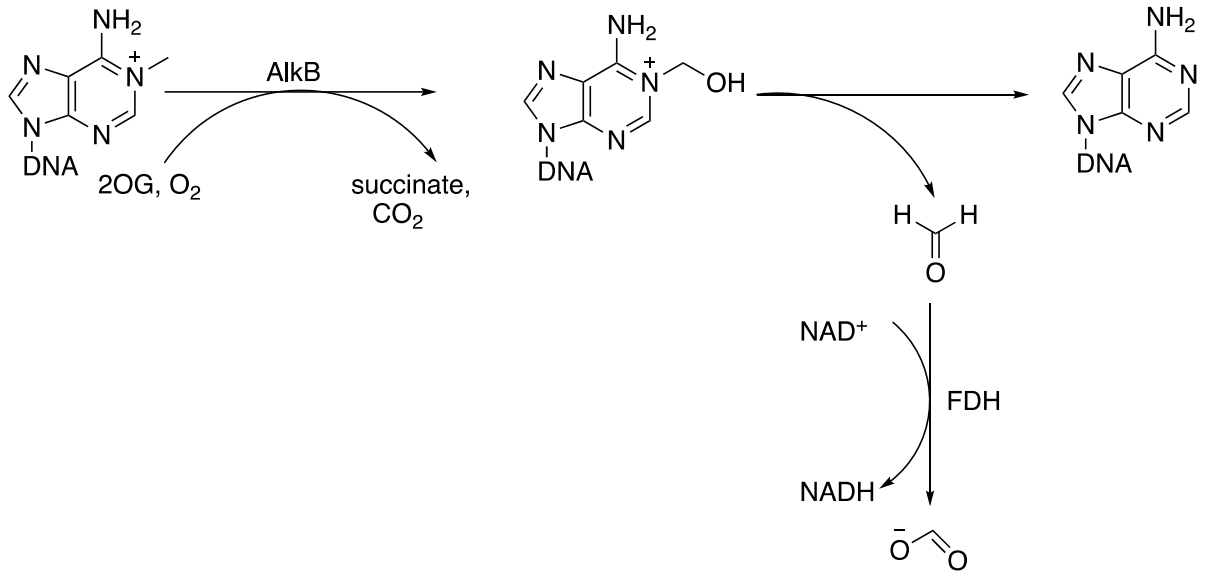


Figure 5.2 Schematic representation of the formaldehyde dehydrogenase (FDH) coupled assay, in which the oxidation of formaldehyde to formate occurs with concomitant reduction of NAD^+ to NADH as catalysed by FDH. The production of NADH is monitored by fluorescence spectroscopy.

Previous assays developed for the AlkB family of enzymes include the gel electrophoresis of labelled DNA,[2] analysis of non-radioactive DNA using mass spectrometry,[27] HPLC,[6,10,15] and fluorescence based methods.[21,26] Apart from these activity-based assays, studies have also been conducted to visualise the predicted enzyme-catalysed intermediates.[6,26–28] Anaerobic crystals of AlkB with bound substrates were exposed to air, allowing the capture of the predicted hemiaminal-nucleotide (hydroxymethyl-nucleotide) intermediate, supporting the proposed mechanism (Figure 5.1B).[28] The intermediates of the enzymatic reactions have also been examined by matrix-assisted laser desorption/ionisation (MALDI) mass spectrometric based assays,[27] fluorescence assays,[26] and UV-Vis assays.[6] However, NMR studies on the intermediates of FTO catalysis have not been reported prior to the work described here. Further, no direct evidence for the release of HCHO as a consequence of AlkB/FTO oxidation has been reported.[28]

5.2 Objectives of the work described in this chapter

The objectives of the work described in this chapter were as follows

- To characterise the intermediates and products formed during the demethylation of m^6A and m^3T by FTO using NMR;
- To characterise the intermediates formed during the demethylation of m^1A by AlkB using NMR;
- To validate the formation of a recently proposed intermediate/product, f^6A, formed during catalytic oxidation of hm^6A by FTO using NMR;
- To quantify the formaldehyde produced during FTO catalysis.

Here, I report what is believed to be the first NMR-based FTO assay (with substrates m^3T and m^6A) and AlkB assay (with substrate m^1A). These results reveal the release of formaldehyde as a direct consequence of the oxidative demethylation of substrates m^3T, m^6A, and m^1A. These results also reveal FTO-catalysed formation of a stable hemiaminal intermediate/product, which gradually decomposes to give formaldehyde and the corresponding demethylated nucleoside.

5.3 Results

5.3.1 Studies on m^6A as a substrate for FTO

m^6A in RNA is reported to be a major substrate for FTO.[7] The proposed intermediate formed during the FTO-catalysed reaction, hm^6A, should have characteristic *N*-hydroxymethyl proton and carbon resonances around δ_H 5.1 ppm and δ_C 64.0 ppm (see Chapter 4). However, the proton signals corresponding to water and formaldehyde are observed at similar chemical

shifts (δ_H 4.7 ppm and δ_H 4.8 ppm, respectively) by ^{1}H NMR. It was envisaged that the detection of the proposed hemiaminal (hm^{6}A) intermediate would be easier with carbon-13 ([^{13}C]) labelling of the *N*-methyl group undergoing oxidation. The [^{13}C] labelling should split the characteristic $^{13}CH_3$ / $^{13}CH_2OH$/ $^{13}CH_2O$ resonances into a doublet with a characteristically large $^1J_{CH}$ (140 Hz, 160 Hz, and 164 Hz, respectively; exact values were measured from recorded ^{1}H NMR during the analysis) value in both ^{1}H and ^{13}C spectra. It was hoped that, although the hm^{6}A hemiaminal proton signals resonate close to the residual water peak in the ^{1}H NMR spectrum, the large $^1J_{CH}$ value (160 Hz) would separate this resonance away from water residual peak. [^{13}C] labelling should also reduce the acquisition time for the detection of the carbon signals of interest (-$^{13}CH_3$ / -$^{13}CH_2OH$ / -$^{13}CH_2O$) in two-dimensional heteronuclear single quantum resonance (2D HSQC) experiments.

5.3.2 Optimisation of the FTO assay

Initially, deuterated Tris.HCl Buffer at pD 7.5 (hereafter, dTris) in Milli-Q purified water was used for the FTO assays. The methylene peaks of Tris buffer resonate at δ_H 3.6 ppm in ^{1}H NMR (Figure 5.3). Therefore, it was envisaged that the use of deuterated Tris buffer (dTris) would minimise the overlap of buffer signals with the resonances of interest. First, samples of 2OG, sodium L-ascorbate, and m^{6}A in dTris buffer were analysed by ^{1}H NMR (Figure 5.3).

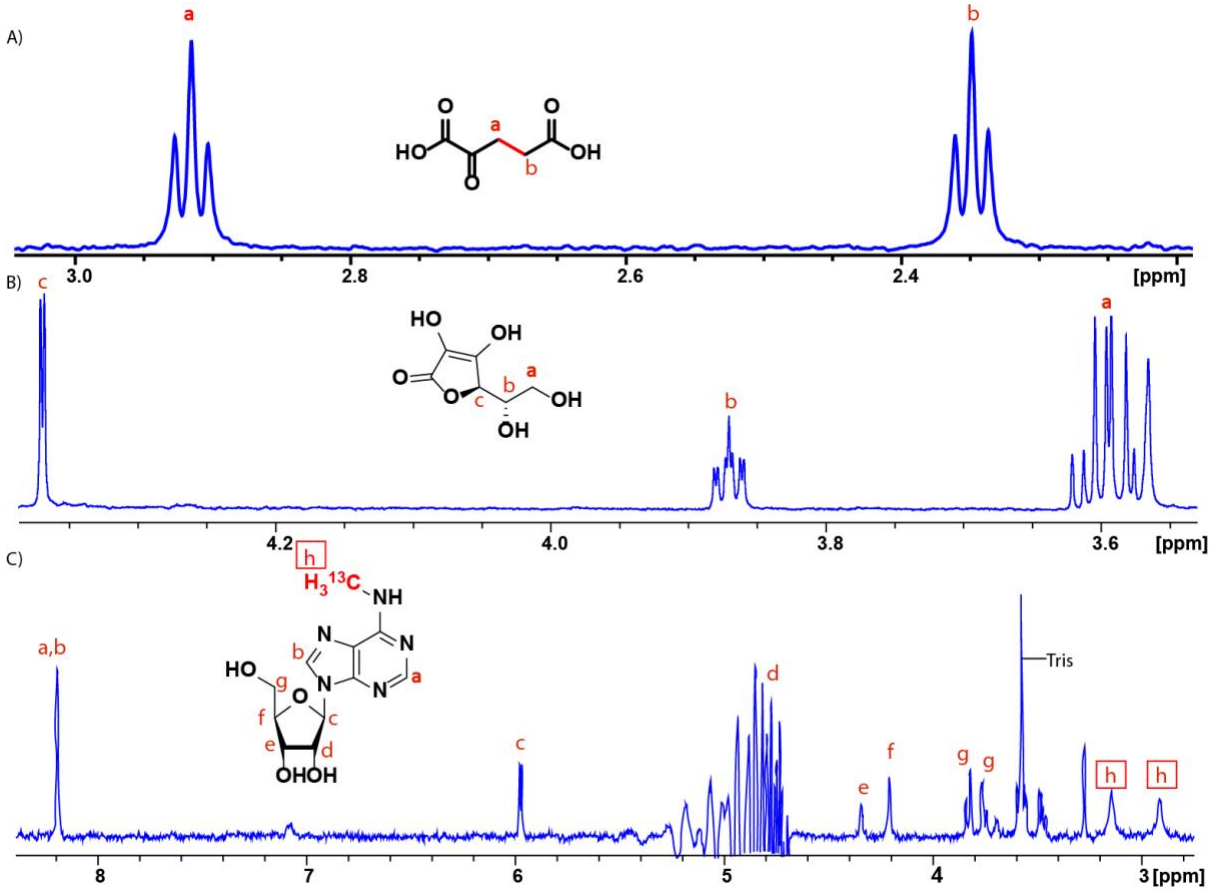


Figure 5.3 ^{1}H NMR spectra (700 MHz) of A) 2OG in dTris buffer, B) sodium L-ascorbate in dTris buffer, and C) m^{6}A in dTris buffer. The ^{1}H assignments are labelled. The proton signal for residual Tris is observed at δ_H 3.6 ppm (probe temperature: 27°C).

In the ^{1}H NMR spectrum of 2OG, the two sets of methylene protons (marked as "***a***" and "***b***", in Figure 5.3) were observed to resonate as triplets at δ_H 2.35 ppm and δ_H 2.95 ppm. The downfield resonance at δ_H 2.95 ppm corresponds to the C-3 methylene protons of 2OG, and the upfield resonance at δ_H 2.35 ppm corresponds to the C-4 methylene protons (Figure 5.3 A). In the ^{1}H NMR spectrum of sodium L-ascorbate, the resonances at δ_H 4.36 (d) ppm, δ_H 3.8 (ddd) ppm and δ_H 3.6 (m) ppm correspond to the protons attached to the dihydrofuran ring, the methylidine proton and terminal methylene protons, respectively (marked as ***c, b, and a,*** respectively, in Figure 5.3B). In the ^{1}H NMR spectrum of m^{6}A (see Section 6.11.3 for the synthesis of [^{13}C]-m^{6}A), the broad singlet at δ_H 8.19 ppm corresponds to the protons attached

to C-2 and C-8 of the purine ring (marked as ***a*** and ***b***, respectively in Figure 5.3C). The doublet at δ_H 5.97 ppm corresponds to the anomeric proton (marked as ***c*** in Figure 5.3C) of the ribose ring. The proton attached to C-2 of the ribose ring overlaps with the residual water peak and resonates at δ_H 4.65 ppm (marked as ***d*** in Figure 5.3C), as confirmed by a correlation with the carbon at δ_C 74.0 ppm in a 2D HSQC spectrum (obtained without using water suppression). It should be noted that a 2D HSQC spectrum of m^6A in D_2O obtained with water suppression masked the correlation for C-2; therefore, a 2D HSQC spectrum of m^6A in D_2O was acquired without water suppression (Figure 5.4). The resonances at δ_H 4.3 (b.s) ppm, δ_H 4.2 (b.s) ppm, δ_H 3.83 (dd) ppm, δ_H 3.74 (dd) ppm correspond to the protons attached to C-3, C-4, C-5, and C-5, respectively, (marked as ***e, f, g, and g,*** respectively, in Figure 5.3C) of the ribose ring of m^6A. The N^6-[^{13}C]-CH_3 signal was observed as a doublet with a $^1J_{CH}$ value of 140 Hz (δ_H 3.14 ppm and δ_H 2.91 ppm, marked as ***h***, Figure 5.3), while the carbon resonance was observed at δ_C 27.43 ppm.

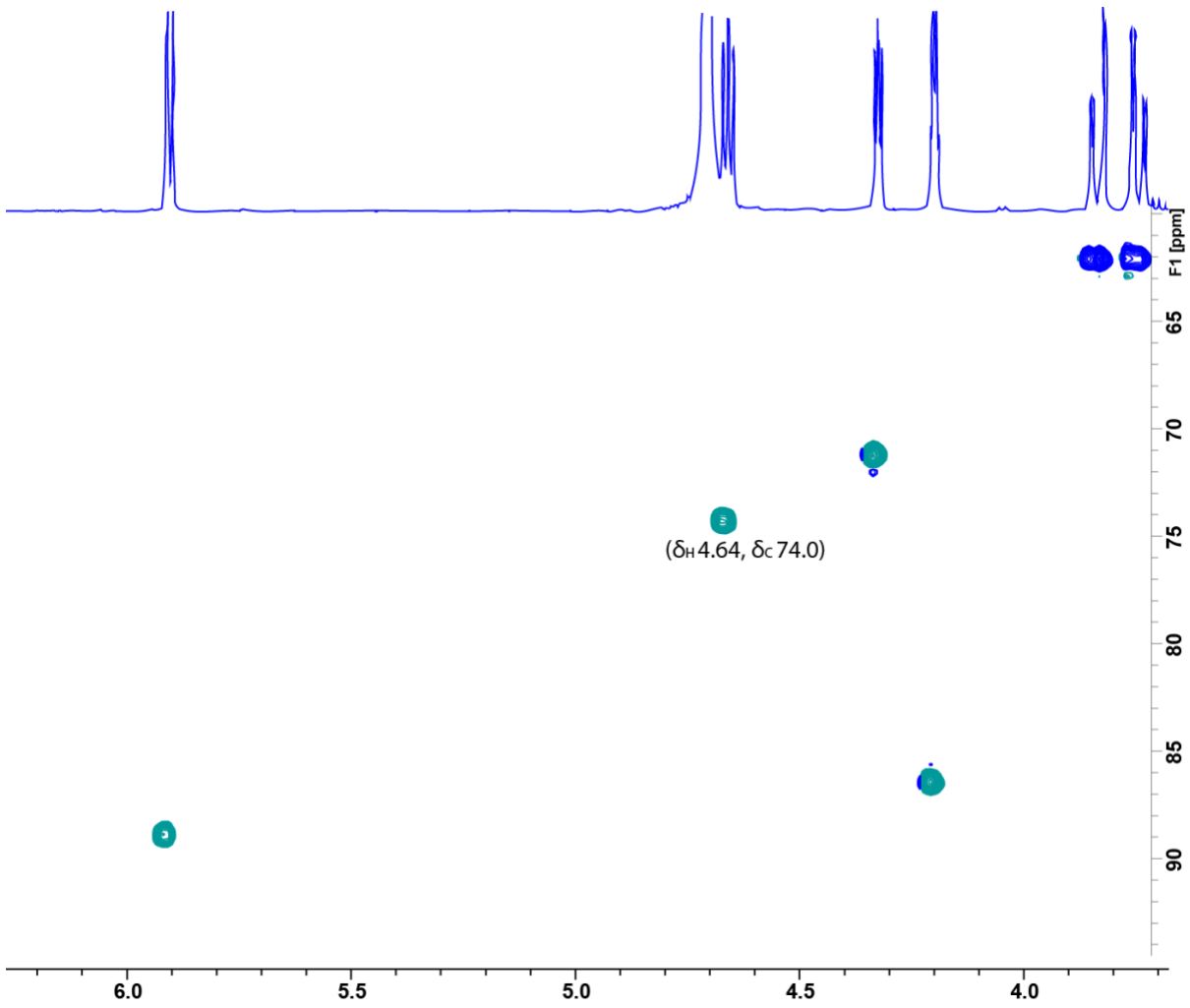


Figure 5.4 2D HSQC (600 MHz) NMR spectrum of m^6A in D_2O, showing a correlation between C-2 and the attached proton. The resonances in green represent the methylidine carbon and those in blue represent methylene carbon. (temperature of the probe: 27°C).

With the proton and carbon chemical shifts of the reagents used in the assay assigned, the FTO assay conditions were optimised such that the Fe(II):enzyme ratio was 1:1, while 2OG and sodium L-ascorbate were present in excess [1 mM each, 50-fold excess relative to Fe(II)]. The addition of ascorbate is reported to increase the catalytic activity of many 2OG-dependent oxygenases *in vitro*.[29] One reason is likely the reducing properties of ascorbate, which may ensure the presence of catalytically active Fe(II) in the reaction mixture by reducing oxidised Fe(III)/Fe(IV) species formed due to the presence of oxygen or hydrogen peroxide in the sample or during enzyme catalysis. Uncoupled 2OG turnover (i.e., turnover

without substrate oxidation) can sometimes be coupled to the oxidation of ascorbate, which apparently substitutes for the normal substrate.[29] Ascorbate might also prevent inadvertent oxidation of some amino acids residues during some catalytic cycles. The concentrations of the reagents used for the partially optimised FTO assay are given in Table 5.1.

Table 5.1 Final concentration of reagents used in FTO enzymatic assays

Reagent	Concentration (μM)
[^{13}C]-m^{6}A	400
diammonium iron(II) sulphate	20
2OG	1000
sodium L-ascorbate	1000
FTO	20

To minimise the time-lapse before NMR analyses, samples were transferred immediately after initial mixing to 3 mm NMR tubes. A pulse sequence employing excitation sculpting with gradient suppression was used for water suppression.[30] In the ^{1}H NMR spectrum of the reaction mixture, the downfield 2OG resonance at δ_H 2.95 ppm overlaps with one of the resonances (δ_H 2.91 ppm) of the N^6-$^{13}CH_3$ group. However, this overlap was considered unlikely to interfere with reaction monitoring. The first spectrum was acquired after approximately 5 min following initial mixing. However, the activity of FTO in these assays was decreased compared to the analogous reaction carried out in an Eppendorf tube. The extent of turnover was approximately 20% in the NMR tube assays compared to the approximately 50% turnover when using an Eppendorf tube. The limited availability of oxygen in the NMR tube may have contributed to the observed decrease in enzyme activity. For simplicity, the numbered positions on the structures of m^{6}A, hm^{6}A and adenosine (A) in

Figure 5.5 correspond to the resonances labelled in Figure 5.6 and in the following descriptions.

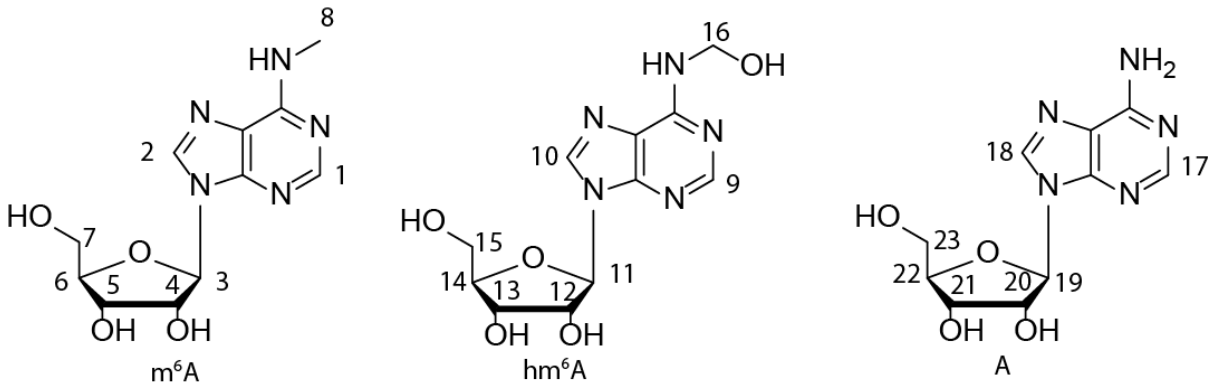


Figure 5.5 Structures of m^{6}A, hm^{6}A and A. The protons monitored by NMR are numbered; the same numbering system is used throughout this chapter.

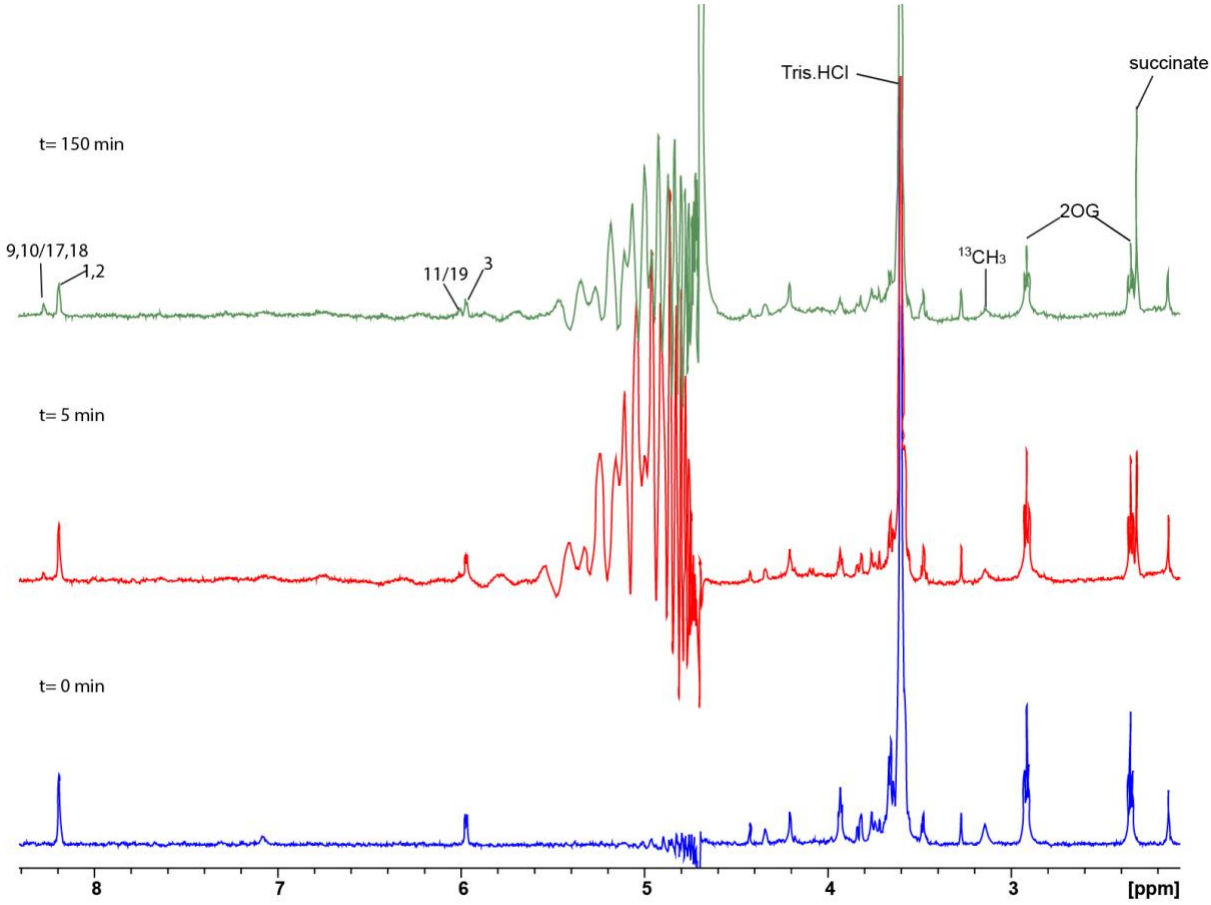


Figure 5.6 ^{1}H NMR (600 MHz) spectra of a time-course monitoring the FTO-catalysed demethylation of m^6A in 50 mM dTris buffer in D_2O. The signals corresponding to the C-2, C-8 and anomeric sugar protons of the reactants, products, dTris, N^6-$^{13}CH_3$ group of m^6A, succinate are indicated. The temperature of the probe was 27°C.

During the time-course, new resonances were observed to develop at δ_H 2.31 ppm (s), δ_H 6.0 ppm (d), and δ_H 8.27 ppm (s) corresponding to the methylene protons of succinate, the sugar anomeric proton (**19** or **11** in Figure 5.5) of the products, and the protons attached to C-2 and C-8 of the purine ring of products (**9** and **10,** or **17** and **18** in Figure 5.5), respectively (Figure 5.6). It was expected that the *N*-hydroxymethyl peak (**16** in Figure 5.5) of hm^6A, if present, would be observed around δ_H 5.1 ppm; the resonance corresponding to formaldehyde should resonate at δ_H 4.8 ppm. However, these signals are likely obscured by the water suppression. As a result, the product structure could not be confirmed based on these ^{1}H-NMR spectra; it was not clear whether m^6A or A were present. Furthermore, there was no signal in the 2D

HSQC spectrum corresponding to the protons (**16** in Figure 5.5; expected at δ_H 5.1 ppm and δc 64 ppm) of the *N*-hydroxymethyl group of hm^{6}A.

The lack of correlations observed in the 2D HSQC spectrum may be due to an insufficient acquisition time. Furthermore, no signals for formaldehyde were observed. It was considered that this may be due to low levels of formation (below the detection limit of NMR), due to overlap of the hydrated HCHO signal (δ_H 4.73 ppm) with that of residual water, or sequestration of formaldehyde by a nucleophile. Formaldehyde is known to react with nucleophiles (in this case, dTris buffer) rendering the quantification of HCHO and hm^{6}A difficult.[31] It should be noted that the conversion of 2OG to succinate exceeded the conversion of reactant to product. The excess production of succinate may be due to non-enzymatic turnover, e.g. by hydrogen peroxide. Alternatively, even in the presence of substrate, uncoupled 2OG turnover can occur owing to a non-productive conformation of the substrate at the active site, or be due to the absence of the substrate at the active site when the ferryl species (Fe=O) is generated. A single nucleoside is likely to be a poor substrate for FTO due to the absence of other binding interactions, potentially resulting in increased uncoupled 2OG turnover.

Therefore, it was confirmed that a new species similar to the reactant was formed in the FTO assay. However, the absence of characteristic correlations observed in the 2D HSQC spectrum and the obstruction of characteristic signals by the water suppression signals in the ^{1}H NMR spectrum led to ambiguity in the identification of this product.

4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) was next investigated as a buffer for the reaction. First, a ^{1}H-NMR spectrum for a sample of 2OG in HEPES buffer was obtained.

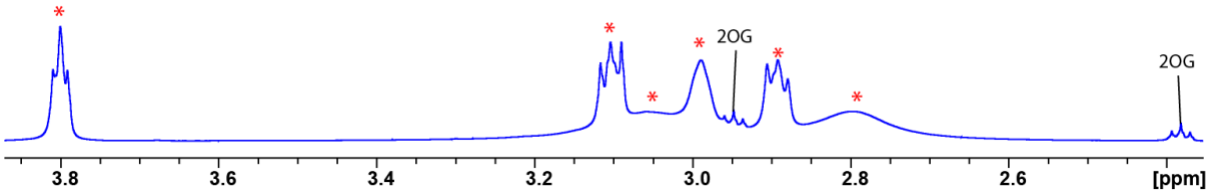


Figure 5.7 ^{1}H NMR (600 MHz) spectrum of 1 mM 2OG in 50 mM HEPES. The HEPES proton signals are marked with red asterisks, and the 2OG signals are labelled. The temperature of the probe was 25°C.

The proton signals of HEPES were observed to the same region as those of 2OG, thus making the quantification of 2OG and succinate signals difficult. Further, in the ^{1}H NMR spectrum of m^{6}A, the resonances corresponding to the *N*^{6}-^{13}CH$_3$ group of m^{6}A (**8** in Figure 5.5) resonate around δ_H 3.0 ppm. Thus, HEPES buffer was not considered suitable for the analysis of the FTO reaction by NMR.

Ammonium formate (AFN) buffer was next investigated as the reaction buffer for NMR analysis. AFN was selected as it was considered unlikely to interfere with observation of the enzyme assay components, as the proton resonance of AFN is observed at δ_H 8.4 ppm by ^{1}H NMR. First, a ^{1}H NMR spectrum of a solution of m^{6}A, 2OG and sodium L-ascorbate in AFN was acquired (Figure 5.8).

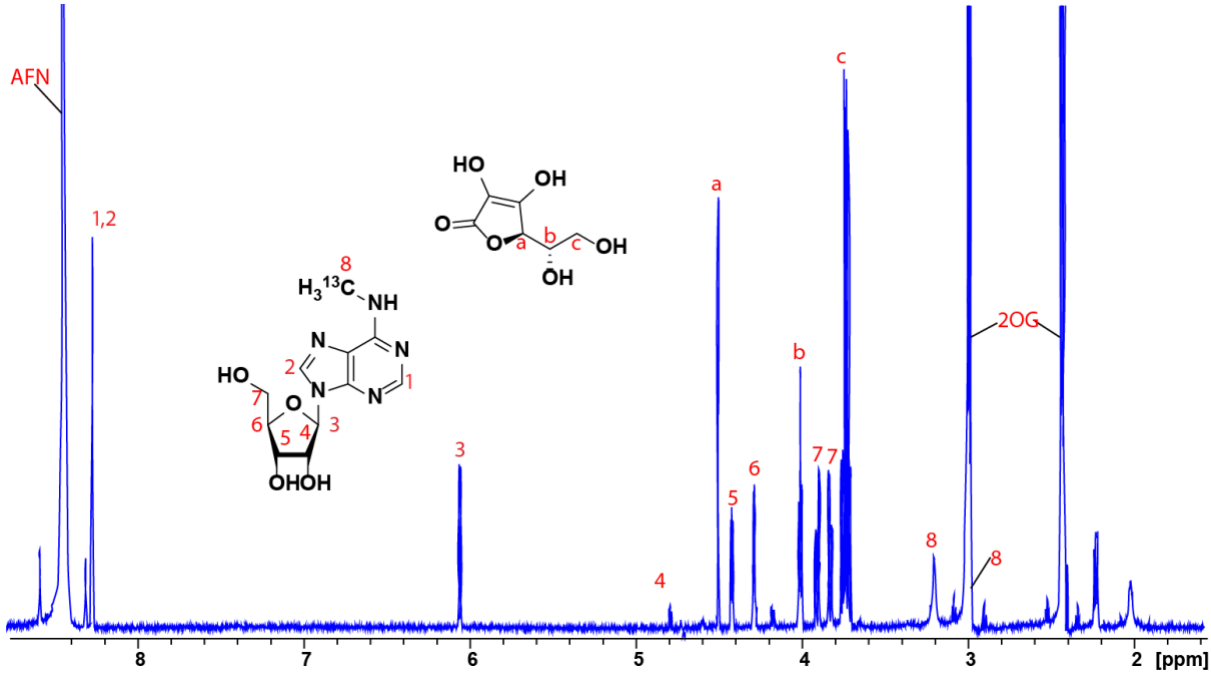


Figure 5.8 ^{1}H NMR (700 MHz) spectrum of [^{13}C]-m^6A, 2OG, and sodium L-ascorbate in 50 mM ammonium formate buffer at pD 7.5. The ^{1}H assignments are labelled as in the chemical structures. The temperature of the probe is 25°C.

The protons (**1** and **2** in Figure 5.8) attached to the purine ring of m^6A were observed near the residual signal of AFN buffer; however, it was proposed that this would not interfere with the analysis of formaldehyde and hm^6A. Therefore, the reaction conditions used were similar to the reaction conditions described above, except AFN was used as the reaction buffer (Table 5.1). After mixing the reaction components, the sample was immediately transferred to a 3 mm NMR tube and the first spectrum was acquired approximately 5 minutes following the initial mixing. Alternating ^{1}H NMR and 2D HSQC spectra were acquired, such that 235 min passed between the consecutive ^{1}H NMR spectra.

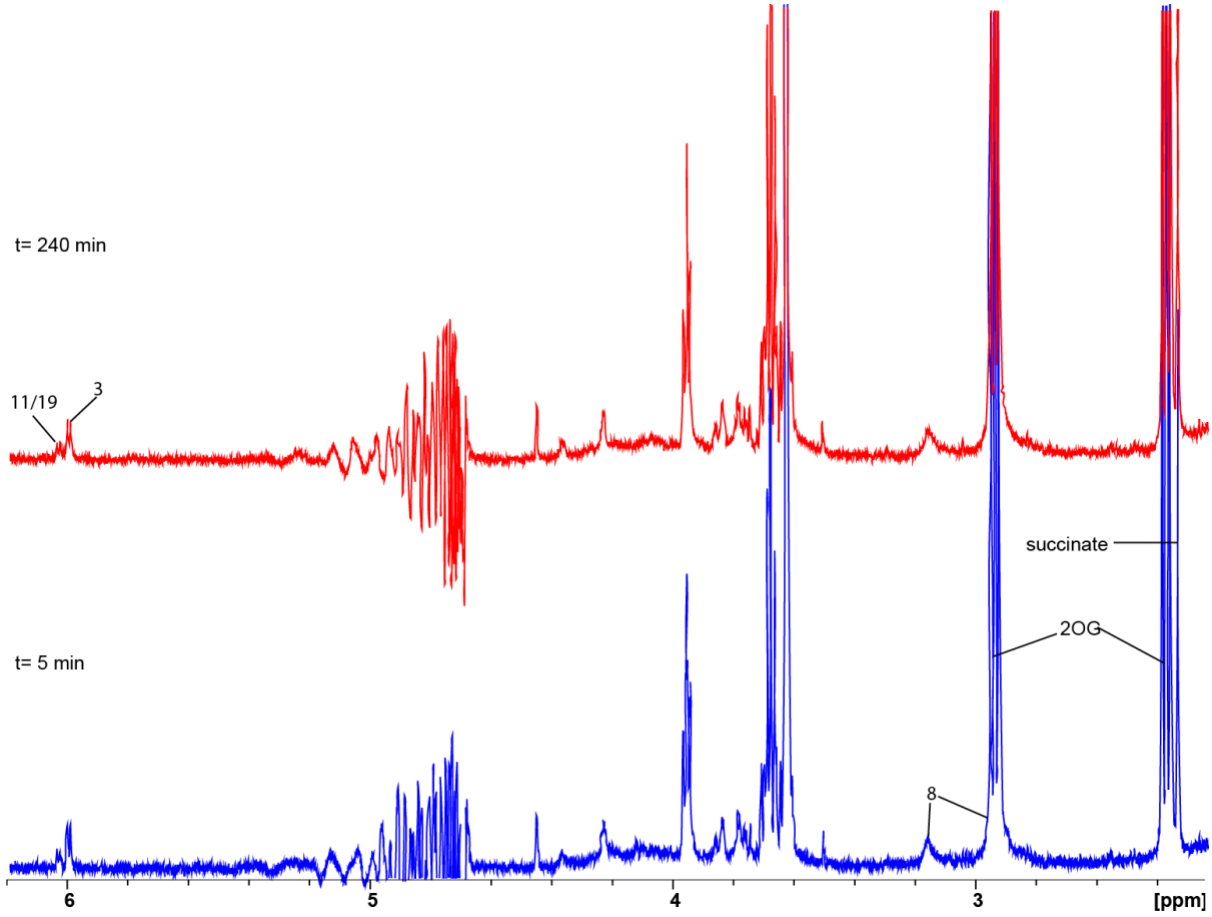


*Figure 5.9 ^{1}H NMR (600 MHz) spectra of a time-course monitoring the FTO-catalysed demethylation of m^6A. Hydroxylation of the N^6-methyl group of m^6A led to the formation of a hemiaminal (hm^6A), which subsequently decomposed to formaldehyde and adenosine (A). The signals corresponding to the anomeric proton of the reactant and the products, N^6-$^{13}CH_3$ (labelled as **8**) of m^6A, 2OG and succinate are labelled. The temperature of the probe was 25°C.*

Similar to what was observed when dTris was used as the reaction buffer (e.g., Figure 5.6), new resonances were observed with AFN as a reaction buffer. These resonances appeared at δ_H 2.33 ppm (s), δ_H 3.6 ppm (s), δ_H 6.04 ppm (d), and δ_H 8.31 ppm (bs) corresponding to succinate methylene protons, residual Tris buffer (FTO was stored in Tris buffer), the anomeric proton (**11** or **19** in Figure 5.5) of the product, and the protons (**9** and **10** or **17** and **18** in Figure 5.5) attached to the aromatic ring of the product (Figure 5.9). Importantly, a correlation was observed between δ_C 64.7 ppm and δ_H 5.13 ppm in the 2D HSQC spectrum (Figure 5.10A), consistent with the formation of N^6-hydroxymethyladenosine (hm^6A). Correlations corresponding to formaldehyde were not observed in the 2D HSQC spectrum,

possibly due to the hydrated formaldehyde signal being obscured by water suppression, or due to low levels of formaldehyde production below the detection limit of the NMR analysis. After incubating the reaction mixture for 8 hrs, a new correlation was observed between δ_C 82.0 ppm and δ_H 4.78 ppm in the 2D HSQC spectrum (Figure 5.10A), indicating the presence of formaldehyde in the reaction mixture. Thus, the absence of HCHO at early time points likely is due to the slow degradation of hm^6A to formaldehyde and adenosine (Figure 5.10B). It should be noted that analysis of ascorbate degradation during the reaction time course is not included in this thesis; these analyses are described in the thesis of Dr Amjad Khan (University of Oxford, 2016).

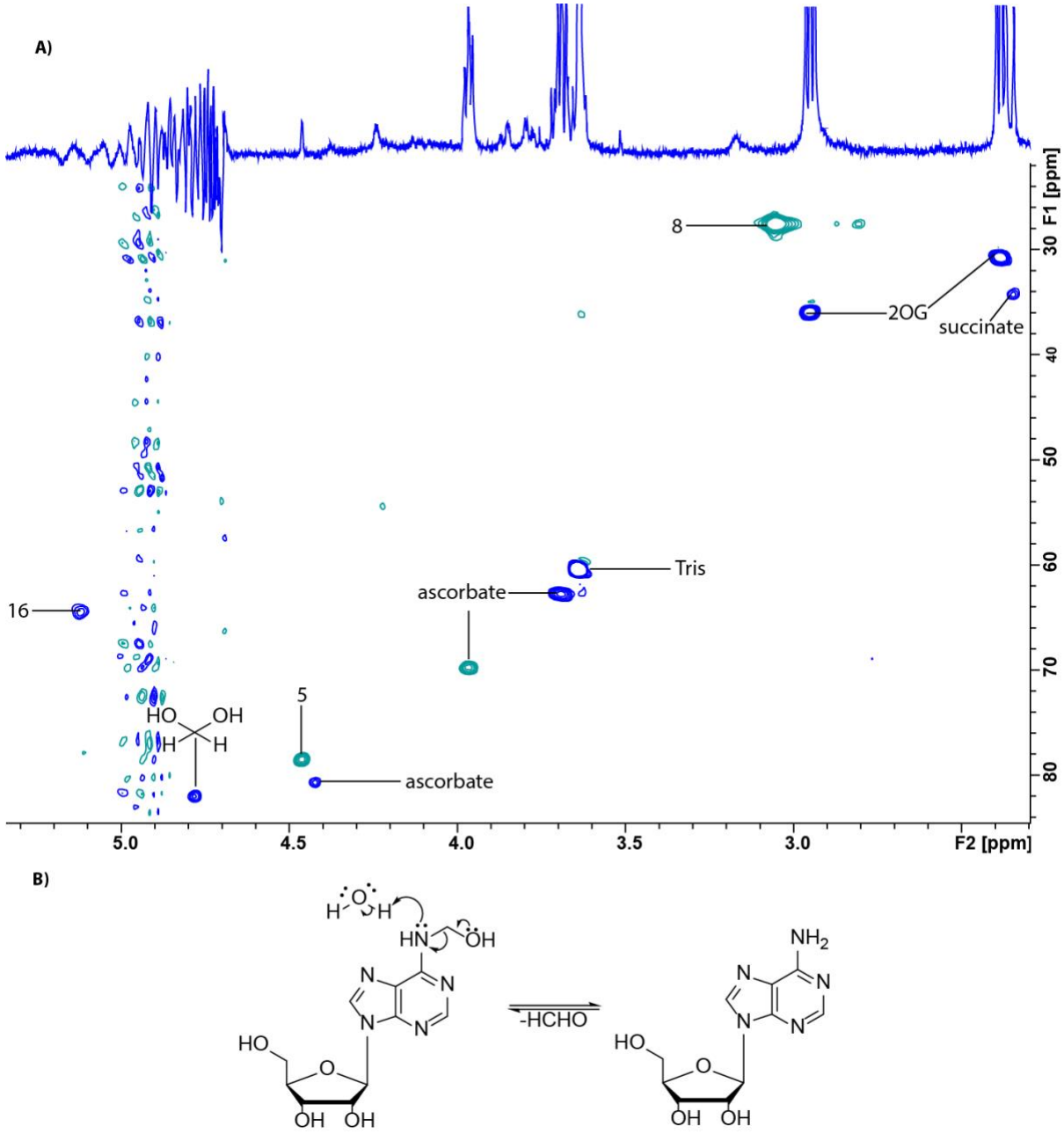


*Figure 5.10 A) 2D HSQC (600 MHz) spectrum showing the correlations between δ_C 64.7 ppm and δ_H 5.13 ppm, and between δ_C 82.0 ppm and δ_H 4.78 ppm, which support the formation of hemiaminal (hm^6A) and formaldehyde, respectively. The correlations marked **8**, **5**, and **16** correspond to the N^6-$^{13}CH_3$ group, the 3' sugar proton and the N^6-[^{13}C]-CH_2OH group, respectively. The correlations for 2OG, succinate, Tris and formaldehyde are indicated. The temperature of the probe was 27°C. B) Schematic representation of hm^6A degradation to adenosine and formaldehyde.*

This assay supported the formation of hm^6A as an intermediate, and formaldehyde as a co-product in the oxidative demethylation of m^6A by FTO. After optimising the assay conditions, work then focussed on better quantifying and characterising intermediate(s), and to investigate the relationship between hm^6A degradation and formaldehyde formation. It was

postulated that 1D heteronuclear single quantum correlation (1D HSQC) experiments would be useful in addressing this analytical challenge, as they would selectively show signals for the [^{13}C]-labelled methyl/hemiaminal groups.

In order to minimise the possible spectral overlap of formaldehyde with residual water, deuterated water was used in combination with milder water suppression based on nuclear Overhauser effect spectroscopy (NOESY).[32] It was envisaged that this suppression scheme would remove the residual water peak, while having less of an impact on peaks close to the residual water signal (e.g., formaldehyde). Further, it was envisaged that increasing the temperature of the NMR probe may increase the rate of enzyme catalysis. As many human enzymes are most active at 37°C (the temperature of the human body), these analyses were performed at 37°C.

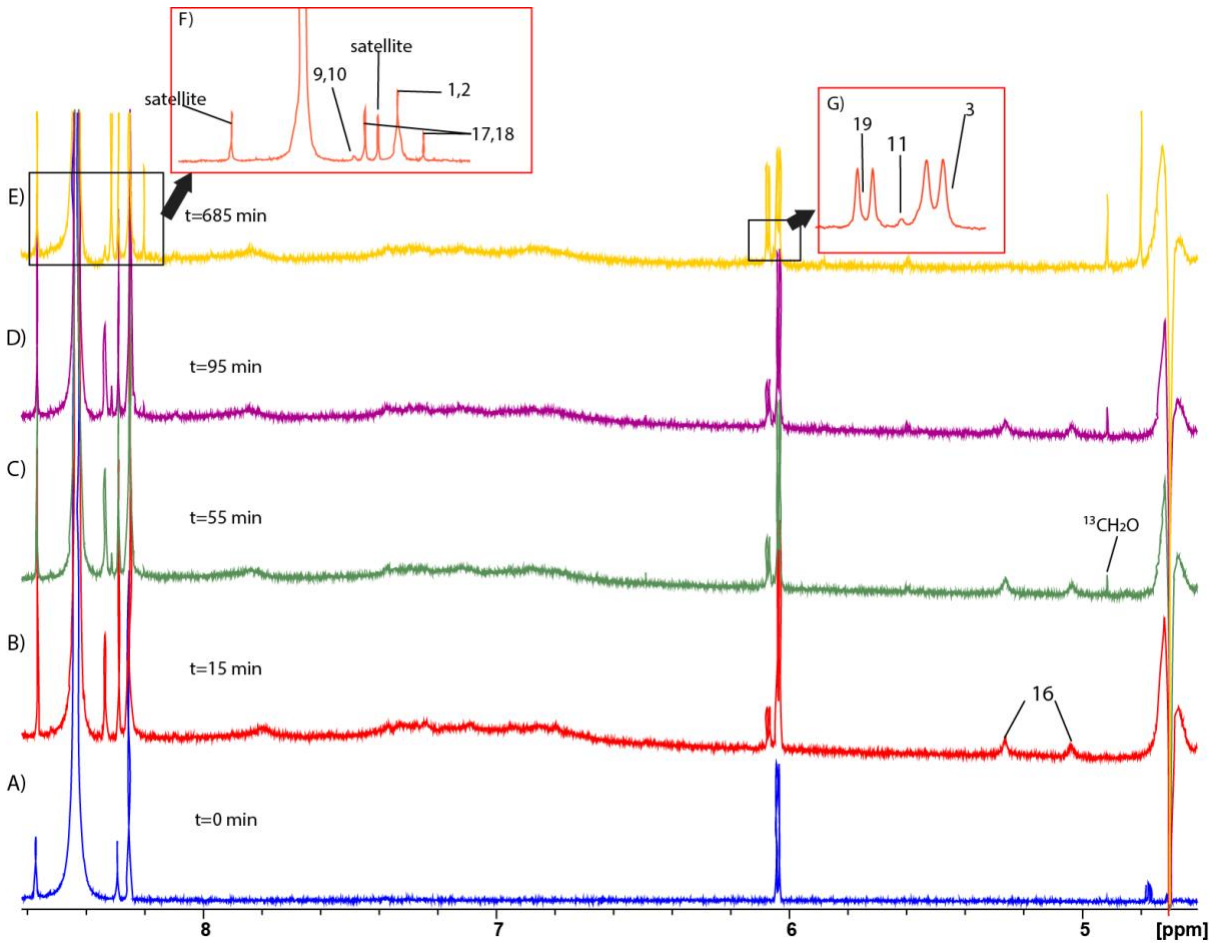


*Figure 5.11 ^{1}H-NMR (700 MHz) time-course monitoring the FTO-catalysed demethylation of m^{6}A, with the formation of hm^{6}A (δ_H 5.26 ppm and δ_H 5.04 ppm correspond to the N^6-$^{13}CH_2OH$ group; labelled as **16**) and formaldehyde (δ_H 4.91 ppm and δ_H 4.67 ppm). The signals corresponding to the aromatic protons and anomeric sugar proton of m^{6}A (marked as **1**,**2**, and **3**), hm^{6}A (marked as **9**,**10** and **11**), A (marked as **17**,**18**, and **19**), N^6-$^{13}CH_2OH$ of hm^{6}A (marked as **16**), and formaldehyde are labelled. The temperature of the probe is 37°C.*

As the reaction was carried out in D_2O with mild water suppression, it was hoped that signals for the N^6-$^{13}CH_2OH$ group (indicating the formation of hm^{6}A in the reaction mixture) and formaldehyde could be observed. Indeed, in the ^{1}H NMR spectrum of the reaction time-course (Figure 5.11), the N^6-$^{13}CH_2OH$ group (labelled as **16** in Figure 5.11B) was observed at δ_H 5.26 ppm and δ_H 5.04 ppm ($^{1}J_{CH}$ 160.0). However, resonances corresponding to formaldehyde were not observed. As a result, it was proposed that formaldehyde was present only at low levels during the first 15 min of the FTO-catalysed reaction. The appearance of the N^6-$^{13}CH_2OH$ resonances (**16** in Figure 5.11B) inversely correlates with the disappearance

of N^6-[^{13}C]-CH_3 resonances (representing m^6A). Approximately 55 min after initial mixing, a peak at δ_H 4.91 ppm was observed, indicating the formation of formaldehyde. An increase in this formaldehyde resonance was observed to correlate inversely with the decrease in the N^6-[^{13}C]-CH_2OH resonances (**16** in Figure 5.11B). New resonances were also observed in the aromatic region. Over the time-course, the resonance at δ_H 8.33 ppm (corresponding to the protons attached to C-2 and C-8 of hm^6A; **9**, **10** in Figure 5.11F) were observed to decrease with a concomitant increase in the resonances at δ_H 8.31 ppm and δ_H 8.20 ppm (corresponding to adenosine; **17**, **18** in Figure 5.11F). Analogous conclusions were drawn after analysing the 1D-HSQC spectra (Figure 5.12).

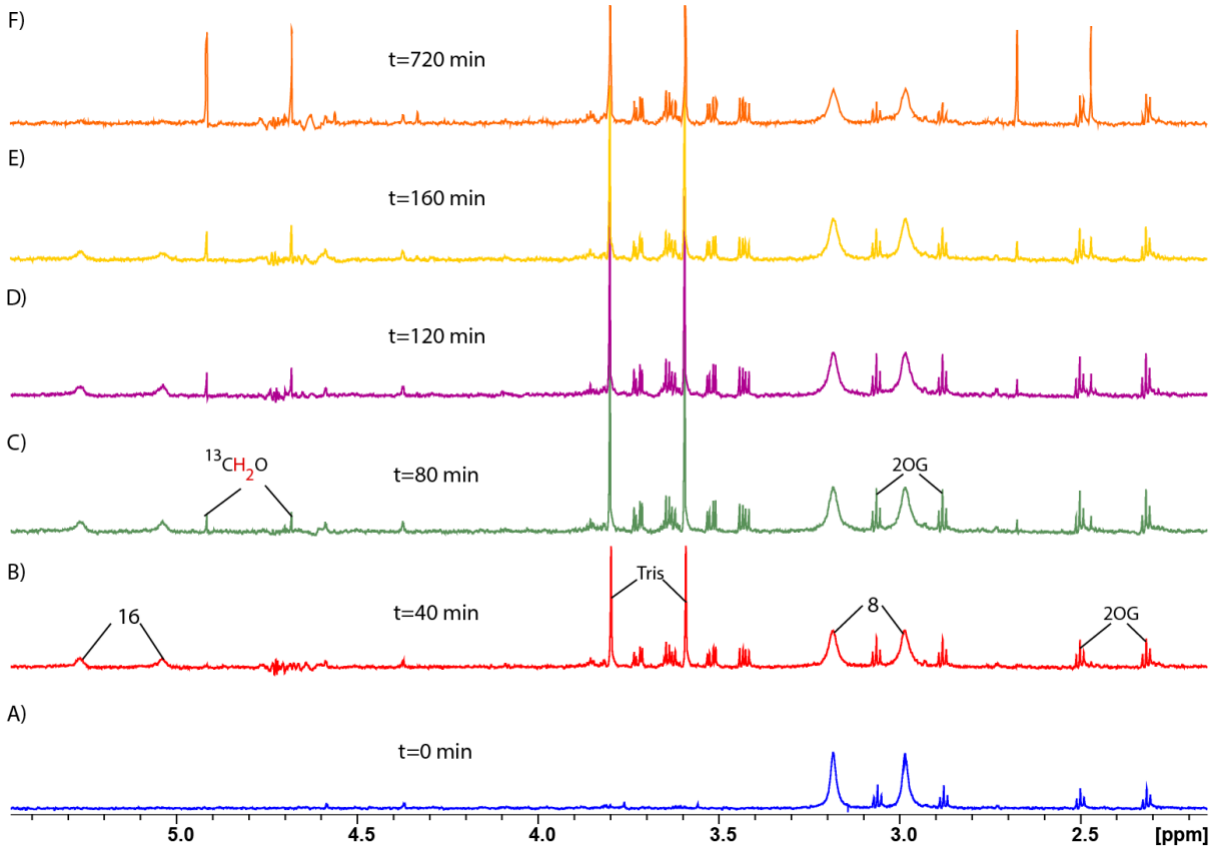


*Figure 5.12 1D HSQC (700 MHz) time-course of the FTO-catalysed demethylation of N^6-[^{13}C]-CH_3-adenosine. The signals corresponding to the methylene protons of the N^6-$^{13}CH_2OH$ group of hm^6A (**16** in panel B), [^{13}C]-CH_2O, Tris buffer, N^6-[^{13}C]-methyl (**8** in panel B) of m^6A and 2OG are labelled. The temperature of the probe is 37 °C.*

The resonances at δ_H 5.26 ppm and δ_H 5.04 ppm were proposed to correspond to the protons attached to the methylene carbon of [^{13}C]-hm^6A (**16** in Figure 5.12B), and the resonances at

δ_H 4.91 ppm and δ_H 4.67 ppm were proposed to correspond to the protons of formaldehyde. A 2D HSQC spectrum of the sample revealed correlations between the resonances at δ_H 5.26 ppm and δ_H 5.04 ppm with the carbon at δ_C 64.5 ppm, and the resonances at δ_H 4.91 ppm and δ_H 4.67 ppm with the carbon at δ_C 81.74 ppm. These resonances are consistent with the resonances observed for authentic samples of N^6-[^{13}C]-hm^{6}A and [^{13}C]-formaldehyde. The [^{13}C]-hm^{6}A sample was prepared by mixing adenosine with an excess of [^{13}C]-formaldehyde (Figure5.13).

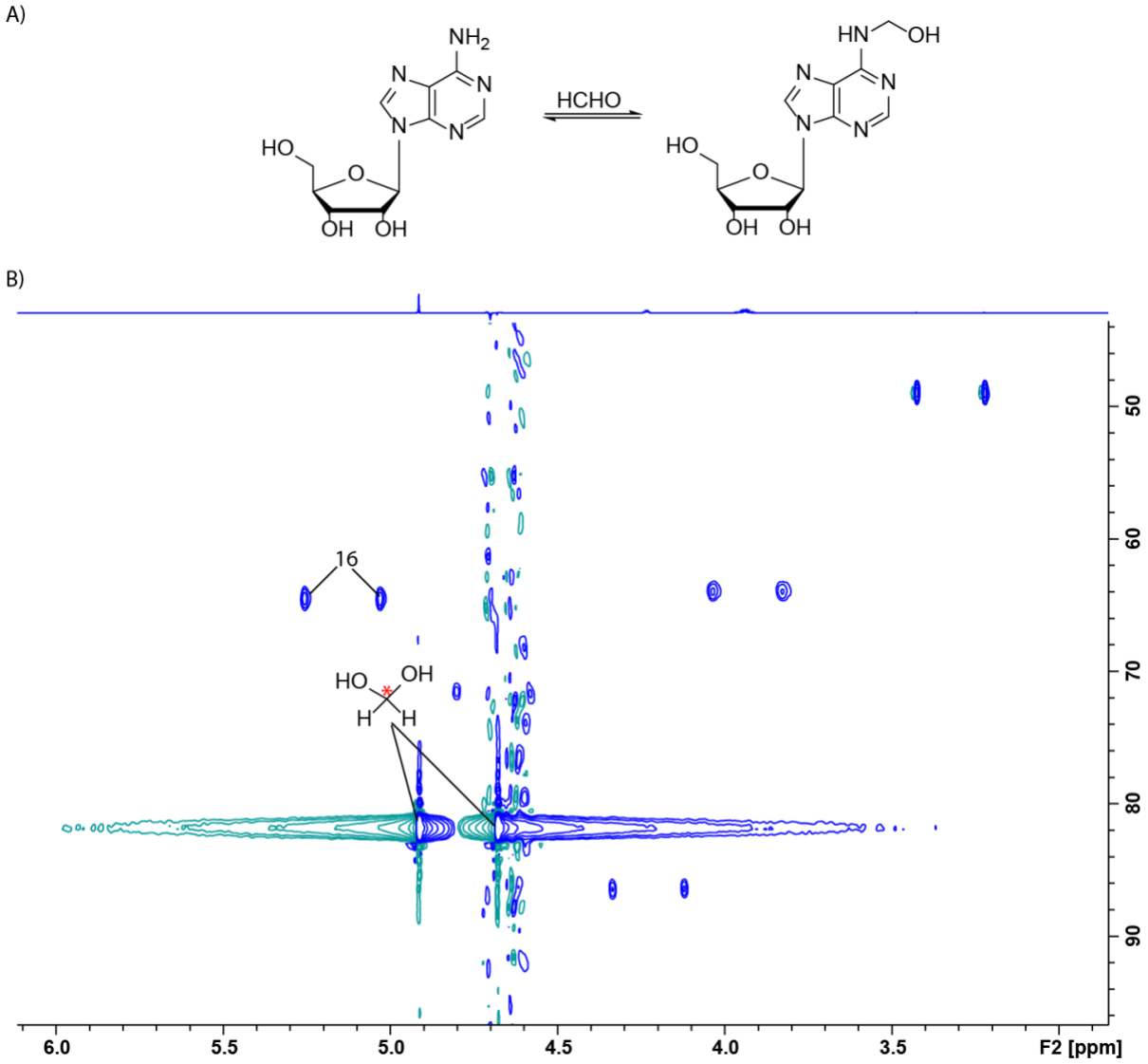


*Figure 5.13 A) Schematic representation of the reaction of adenosine and formaldehyde, leading to the formation of hm^{6}A. B) 2D HSQC (700 MHz) spectrum of a mixture of adenosine and $^{13}CH_2O$, revealing the formation of N^6-[^{13}C]-hm^{6}A. The signals corresponding to the N^6-hydroxymethyl group of hm^{6}A (labelled as **16**) and $^{13}CH_2O$ (^{13}C is labelled by asterisk (*) in red) are labelled. The temperature of the probe was 37°C.*

It was observed that the amount of succinate produced exceeded that of formaldehyde (Figure 5.14). As shown in Figure 5.14A, the formation of succinate was observed to halt approximately 40 min after the reaction was started (possibly due to the limited availability of oxygen). However, the formaldehyde resonances continued to increase even 600 min after initial mixing (Figure 5.14C). This observation indicates that the formation of formaldehyde is, at least partially, a non-catalytic process, likely involving the gradual decomposition of hm^{6}A into formaldehyde and adenosine (Figure 5.10B).

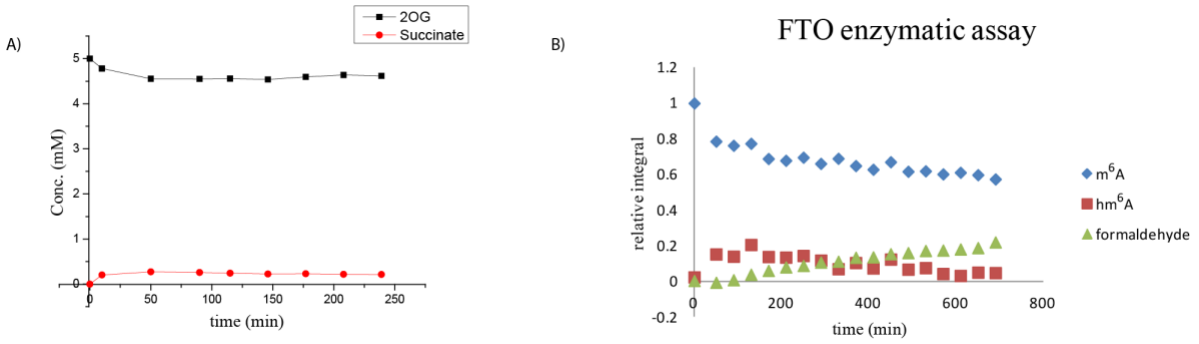


Figure 5.14: Plots demonstrating A) the change in the concentrations of 2OG (mM) and succinate (mM) over time, and B) the relative change in the signal intensities of m^{6}A, hm^{6}A and formaldehyde with time.

5.3.3 Studies on hm^{6}A as a substrate for FTO

To investigate the viability of hm^{6}A as a substrate for FTO, as was reported in the literature,[9] *N*^{6}-[^{13}C]-hydroxymethyladenosine ([^{13}C]-hm^{6}A) was prepared *in situ* by mixing [^{13}C]-CH_2O and adenosine. The final concentration of [^{13}C]-hm^{6}A was found to be 1 mM, based on an internal triisopropylsilane (TSP; 1 mg/mL) standard. The concentrations of the reagents used are given in Table 5.1. A reaction mixture consisting of 5 mM 2OG, 1 mM sodium L-ascorbate, 1 mM hm^{6}A solution, 20 µM diammonium iron(II) sulphate, and 20 µM FTO was prepared in an Eppendorf tube, and then immediately transferred to a 3 mm NMR tube. The first NMR spectrum was recorded approximately 5 minutes after the initial mixing. However,

no new resonances appeared during the time-course, and only signals corresponding to uncoupled 2OG turnover were observed. It is possible that formaldehyde inhibits FTO, or that hm^{6}A is not a substrate for FTO, at least under the NMR compatible conditions tested.

5.3.4 Studies on N^3-[13C]-methylthymidine ([13C]-m^{3}T) as a substrate for FTO

[^{13}C]-m^{3}T was then investigated as a possible substrate for FTO. [^{13}C]-m^{3}T (see Section 6.11.3 for the synthesis of [^{13}C]-m^{3}T) was dissolved in 50 mM AFN buffer and a ^{1}H NMR spectrum was recorded. The protons corresponding to m^{3}T and thymidine (T) are assigned as in Figure 5.15A; the same numbering system is used in further descriptions.

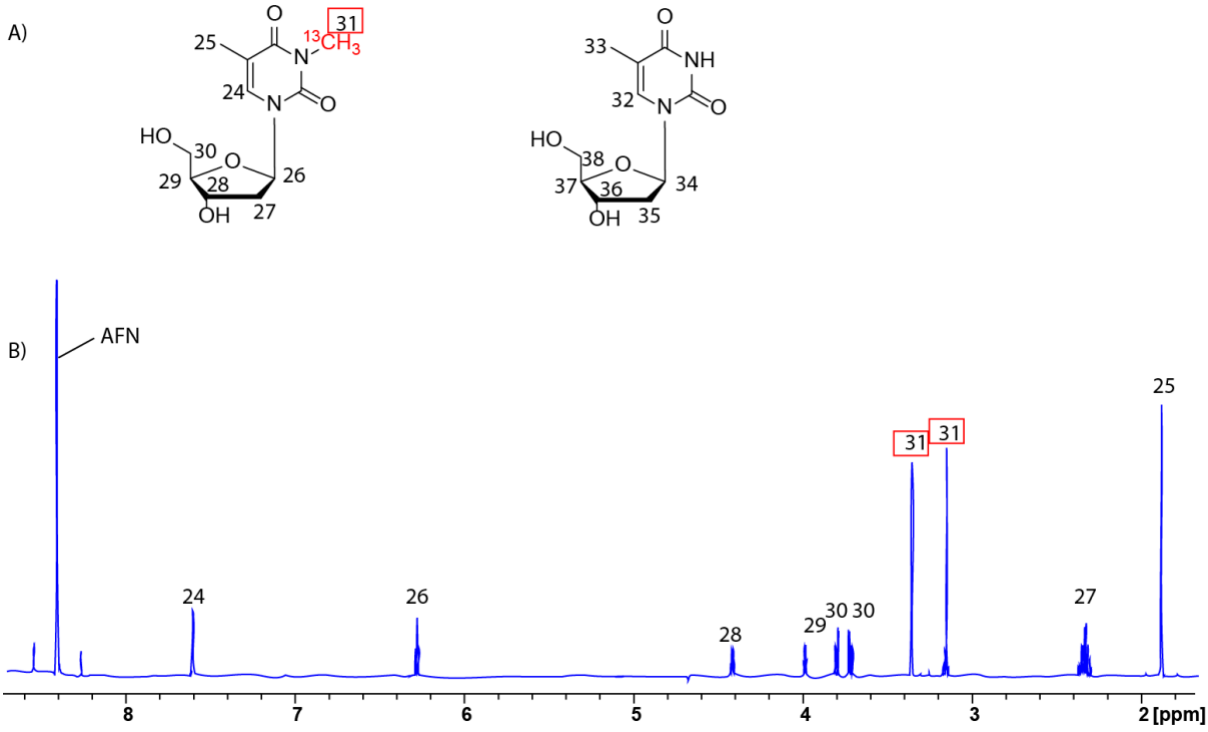


Figure 5.15 A) The structures of [^{13}C]-m^{3}T and T, with protons numbered for NMR analyses. The same numbering system will be used for further descriptions. B) ^{1}H-NMR (700 MHz) spectrum of [^{13}C]- m^{3}T in 50 mM AFN buffer in D_2O at pD 7.5. The temperature of the probe was 37°C.

In the ^{1}H NMR spectrum of [^{13}C]-m^{3}T, the N^3-[^{13}C]-CH$_3$ group was observed as a doublet (δ_H 3.35 ppm and δ_H 3.10 ppm; **31** in Figure 5.15A) with a large $^1J_{CH}$ value (142 Hz). The

resonances at δ_H 7.6 ppm (s), δ_H 6.26 ppm (t), δ_H 4.4 ppm (q), δ_H 3.97 ppm (t), δ_H 3.78 ppm (dd), δ_H 3.70 ppm (dd), and δ_H 2.3 ppm (m) correspond to the proton attached to C-6 (aromatic proton, marked as **24** in Figure 5.15) of the pyrimidine ring and the protons attached to C-1, C-3, C-4, C-5, C-5 and C-2 of the sugar ring (marked as **26**, **28**, **29**, **30**, **30**, and **27**, respectively, in Figure 5.15). The resonance at δ_H 1.86 ppm (s) corresponds to the methyl protons (**25** in Figure 5.15) of the pyrimidine ring of m^3T.

The concentrations of the reagents used are given in Table 5.1. In an Eppendorf tube, a mixture of 5 mM 2OG, 1 mM sodium L-ascorbate, 400 µM [^{13}C]-m^3T, 20 µM diammonium iron(II) sulphate, 20 µM FTO was prepared and immediately transferred to a 3 mm NMR tube. Approximately 15 min after initial mixing, a series of alternating ^{1}H-NMR and 1D-HSQC spectra were acquired.

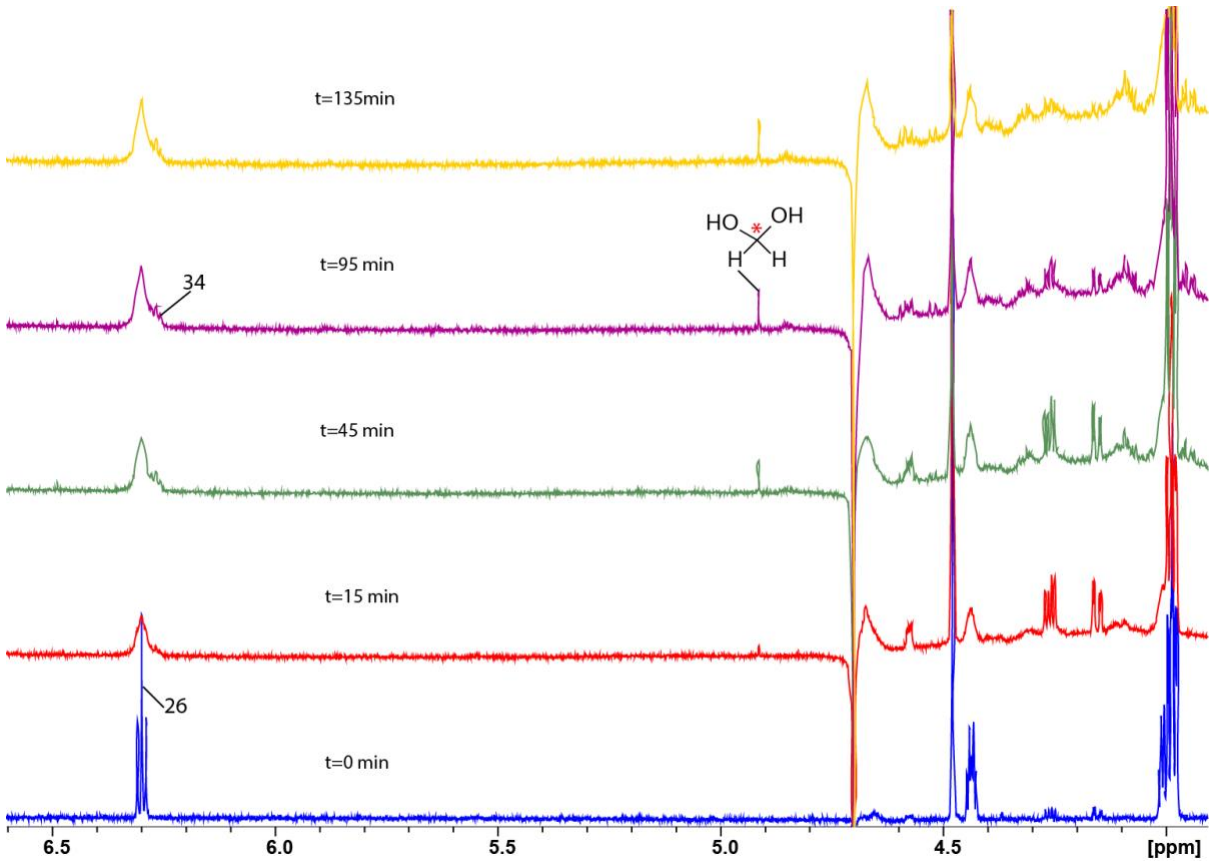


*Figure 5.16 ^{1}H NMR (700 MHz) spectra of a time-course of FTO-catalysed demethylation of N^3-[^{13}C]-CH_3-T. The resonances for formaldehyde, the anomeric proton of the ribose sugar of [^{13}C]-m^3T (**26**) and T (**34**) at δ_H 4.87 ppm, δ_H 6.30 ppm and δ_H 6.28 ppm, respectively, are labelled. As expected, resonances corresponding to N^3-[^{13}C]-hydroxymethylthymidine were not observed. ^{13}C is highlighted by asterisk (*) in red.*

Over the time-course of the reaction, new peaks were observed at δ_H 7.58 ppm (s), δ_H 6.23 ppm (t), δ_H 4.87 ppm (s), δ_H 3.65 ppm (s), δ_H 2.34 ppm (s) and δ_H 1.83 ppm (s) (Figure 5.16). The observed resonances correspond to the protons attached to the C-6 of the pyrimidine ring of T (**32** in Figure 5.15), to the C-1 of the ribose ring of T (**34** in Figure 5.15), formaldehyde, Tris buffer (FTO was stored in Tris), the succinate methylene protons, and the methyl group of the pyrimidine ring of T (**33** in Figure 5.15), respectively (Figure 5.16). Analogous conclusions were drawn by analysing the 1D HSQC spectra of the reaction mixture. Furthermore, in the 1D HSQC spectra, resonances at δ_H 4.87 ppm and δ_H 4.92 ppm were observed, corresponding to [^{13}C]-formaldehyde. Thus, FTO catalyses the oxidative demethylation of m^{3}T to T under the NMR conditions. However, in the ^{1}H NMR spectrum, the resonance corresponding to formaldehyde expected at δ_H 4.82 ppm was not observed, likely due to its close proximity to the region impacted by water suppression (Figure 5.16).

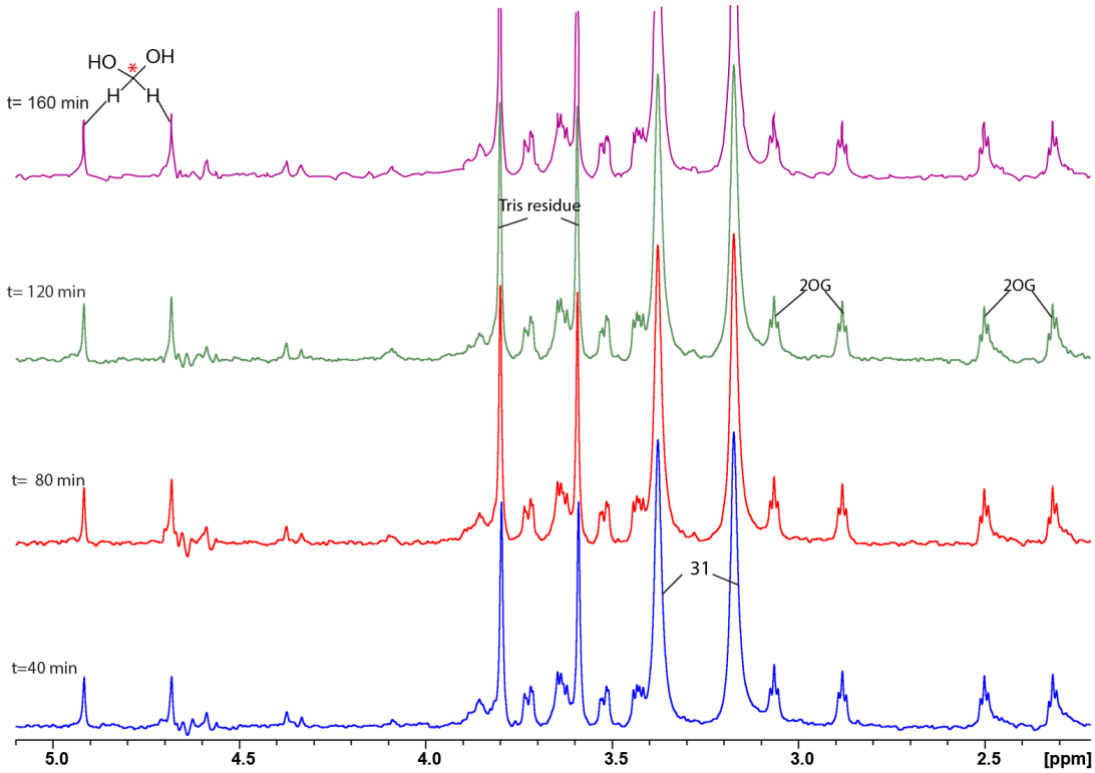


*Figure 5.17 1D HSQC (700 MHz) spectra of a time-course of FTO-catalysed demethylation of [^{13}C]-m^{3}T. The signals corresponding to [^{13}C]-CH_2O, Tris buffer, the N^3-$^{13}CH_3$ group of T (**31**), and 2OG are indicated. ^{13}C is highlighted by asterisk (*) in red. The temperature of the probe was 37°C.*

Resonances corresponding to N^3-[^{13}C]-hydroxymethylthymidine ([^{13}C]-hm^{3}T) were not observed, even while using the sensitive 1D HSQC method (Figure 5.17). This is likely due to the inherent instability of a hydroxymethylated intermediate, as derived from methylation on endocyclic imine (see section 4.4.21, Chapter 4). This would lead to spontaneous dissociation to T and formaldehyde, although the absence of signals could be due to the low levels of [^{13}C]-hm^{3}T present (i.e. below the detection limit of NMR). It should be noted that no changes in the formaldehyde concentration were observed after first time point (40 min, Figure 5.18).

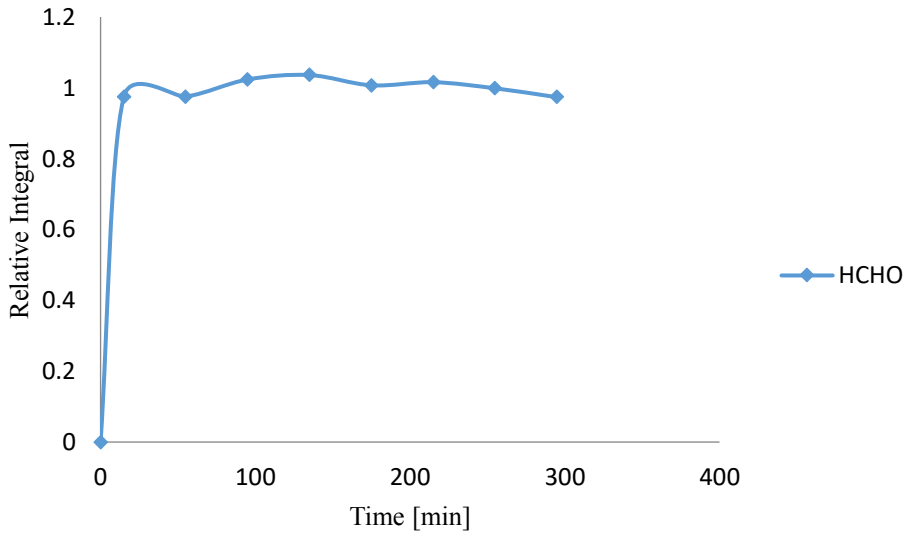


Figure 5.18 Formaldehyde levels over time, as observed during the 1D HSQC (700 MHz) time-course. It should be noted that the concentration of formaldehyde does not change after the first time point.

5.3.5 Studies on m^{1}A as a substrate for AlkB

The potential formation of a hydroxymethyl-nucleoside was also investigated in AlkB catalysis. The substrate used for the investigation of AlkB oxidative activity was N^1-[^{13}C]-methyladenosine ([^{13}C]-m^{1}A; see Section 6.11.3 for the synthesis of [^{13}C]-m^{1}A). However, productive AlkB catalysis requires a a phosphate group attached to the nucleoside.[33] It is proposed that this phosphate group anchors the substrate to the protein, and helps to stabilise the enzyme-substrate complex.[33] Initially, the simple nucleoside (m^{1}A) was investigated as a

substrate for AlkB to validate this hypothesis. First, the ^{1}H-NMR spectrum of [^{13}C]-m^{1}A in 50 mM AFN buffer was recorded the protons of m^{1}A are numbered as in Figure 5.19).

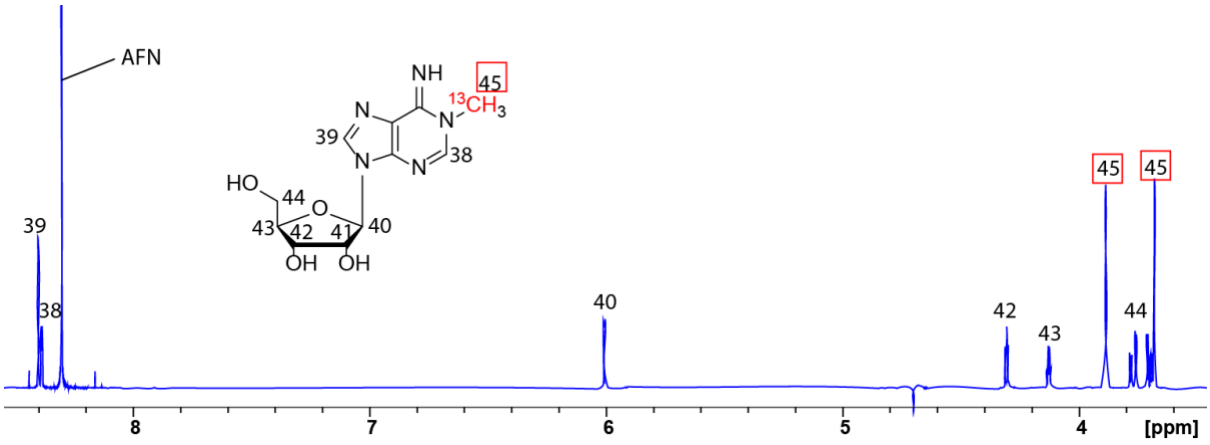


Figure 5.19 ^{1}H-NMR (700 MHz) spectrum of [^{13}C]-m^{1}A in 50 mM AFN buffer in D_2O at pD 7.5. The ^{1}H assignments are numbered; the same numbering system is used elsewhere. The temperature of the probe was 37°C.

The N^{1}-^{13}CH$_3$ peak was observed as a doublet with a large coupling value ($^{1}J_{CH}$ = 144 Hz;δ_H 3.84 ppm and δ_H 4.04 ppm; **45** in Figure 5.19). The resonances at δ_H 8.40 ppm (s) and δ_H 8.38 ppm (d) correspond to the protons attached to C-8 and C-2 (**39** and **38** in Figure 5.19) of the purine ring of [13C]-m^{1}A. The resonances at δ_H 6.00 ppm (d), δ_H 4.30 ppm (t), δ_H 4.12 ppm (q), δ_H 3.76 ppm (dd), and δ_H 3.70 ppm (dd) correspond to the protons attached to C-1, C-3, C-4, C-5 and C-5 (**40**, **42**, **43**, **44**, and **44**, respectively, in Figure 5.19) of the ribose ring of [^{13}C]-m^{1}A.

The reaction conditions used for AlkB assays were based on modifications of literature protocols.[34] A sample containing 1 mM $MgCl_2$, 1 mM sodium L-ascorbate, 5 mM 2OG, 20 µM diammonium iron(II) sulphate, and 20 µM AlkB was prepared and transferred immediately to a 3 mm NMR Match tube. ^{1}H-NMR spectra were recorded approximately 5 min following the initial mixing. While no substrate modification was observed, some uncoupled 2OG turnover was observed. This is likely due to the fact that AlkB accepts its

tryptophan residue, Trp-178, as an autocatalytic substrate in the absence of external substrate or in the presence of a poor substrate.[35]

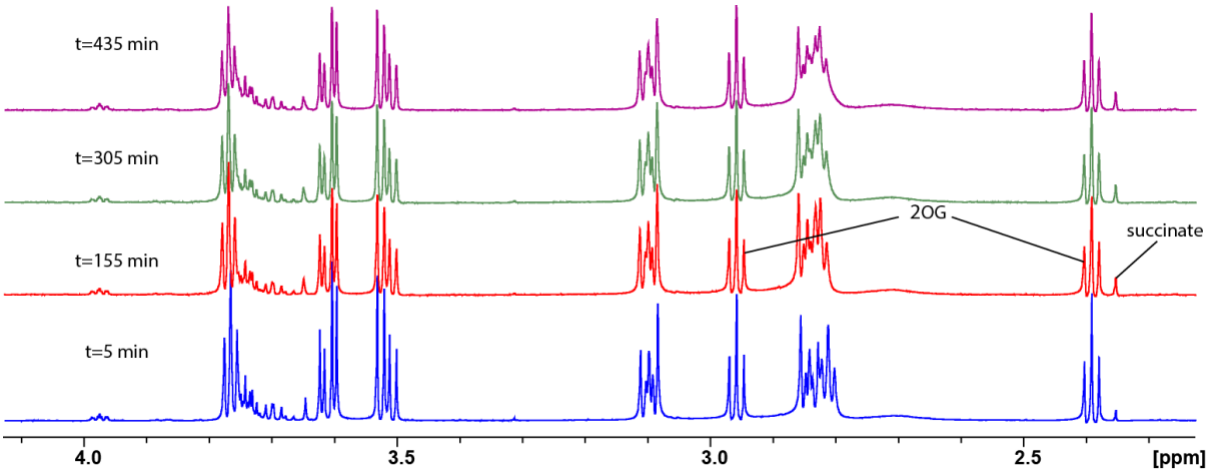


Figure 5.20 ^{1}H-NMR (700 MHz) spectra of a time-course monitoring the potential AlkB-catalysed demethylation of [^{13}C]-m^{1}A. The signals corresponding to 2OG and succinate are labelled. The temperature of the probe was 37°C.

Reaction conditions similar to those used for the FTO assay (Table 5.1) were then tested. In these assay conditions, $MgCl_2$ was not used. It was considered that different enzymes require buffers with different ionic strengths, which can impact hydrogen bonding, and subsequently catalytic activity. The assay solution consisted of 5 mM 2OG, 5 mM sodium L-ascorbate, 2.5 mM [^{13}C]-m^{1}A, 40 µM of diammonium iron(II) sulphate, and 10 µM AlkB. First, a 24 hr reaction was carried out in an Eppendorf tube, in order to ensure sufficient oxygen availability by providing a high surface area (when compared to incubation in an NMR tube). NMR spectra for this sample revealed the development of new resonances with chemical shifts consistent with the formation of Adenosine.

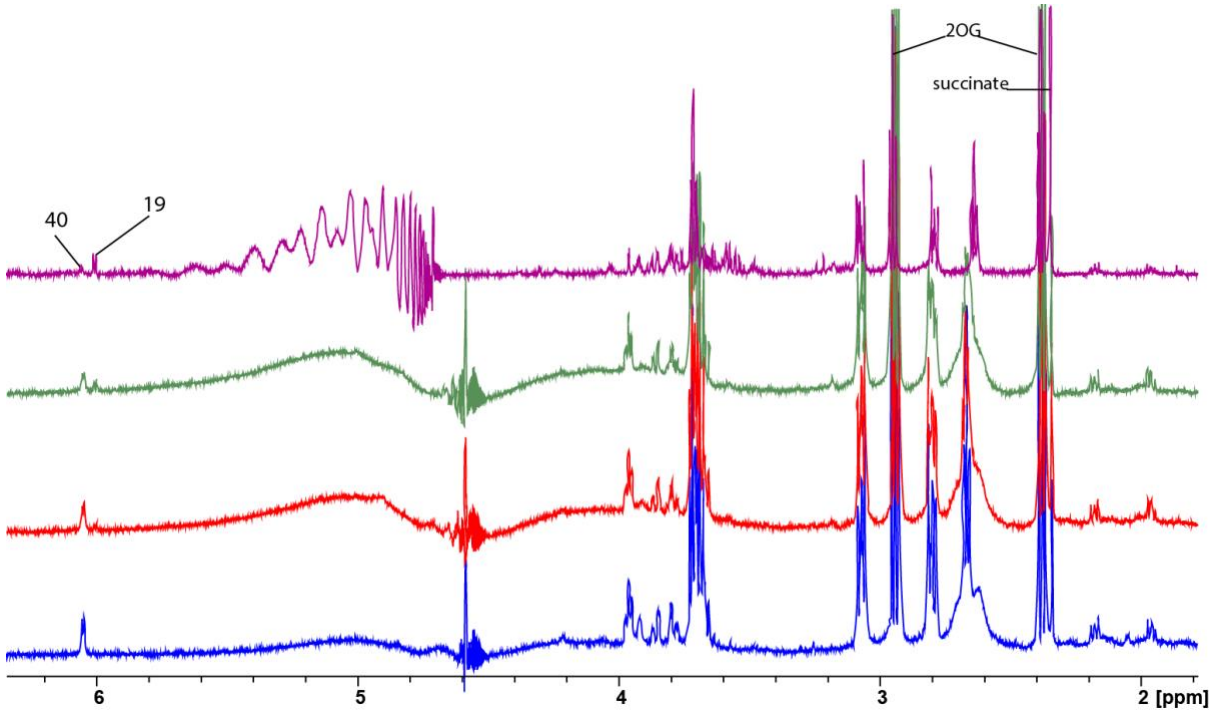


*Figure 5.21 A) ^{1}H-NMR (600 MHz) spectra of a time-course monitoring the AlkB-catalysed reaction of [^{13}C-]m^{1}A. The signals corresponding to the anomeric protons of reactant (**40**) and product (**19**), to 2OG, and to succinate are labelled.*

Another sample was prepared using the same reaction conditions; the reaction mixture was immediately transferred to a 3 mm NMR tube, and a ^{1}H NMR time-course was recorded. New resonances were observed at δ_H 8.13 ppm (s), δ_H 8.12 ppm (s), δ_H 6.00 ppm (d), and δ_H 2.25 ppm (s), corresponding to the protons attached to the pyrimidine ring (C-8 and C-2; **18** and **17** in Figure 5.5) of Adenosine, to C-1 of the ribose ring (**19** in Figure 5.5) of Adenosine and to succinate, respectively (Figure 5.21). Since [^{13}C]-hm^{1}A is very susceptible to hydrolysis (Chapter 4), the new resonances in the spectrum were assigned to adenosine. m^{1}A is positively charged at neutral pH, and so the spontaneous release of formaldehyde from [^{13}C]-hm^{1}A leads to the formation of neutral adenosine. It should be noted that the proximity of the formaldehyde resonances to the water suppression region complicate the detection of formaldehyde. To confirm these results, a 1D HSQC spectrum was recorded on the sample.

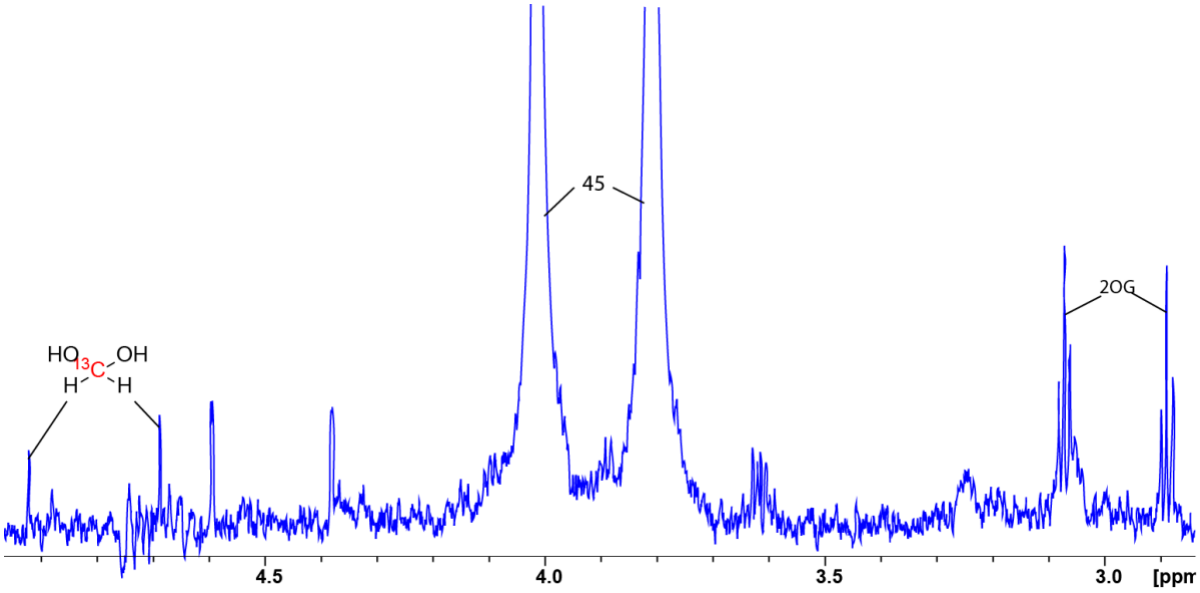


*Figure 5.22 1D HSQC (700 MHz) spectrum of the AlkB-catalysed modification of [^{13}C]-m^{1}A. The signals corresponding to $^{13}CH_2O$, -$^{13}CH_3$ protons (**45**) of [^{13}C]-m^{1}A, and 2OG are labelled.*

Indeed, the 1D HSQC of the reaction mixture revealed the development of resonances at δ_H 4.92 ppm and δ_H 4.69 ppm which corresponded to formaldehyde. These formaldehyde resonances were observed within 15 min following the initial mixing. By contrast, no resonances were observed corresponding to [^{13}C]-hm^{1}A. This is likely because, as described in Chapter 4, hm^{1}A is relatively unstable due to the hydroxymethylation of an endocyclic nitrogen (note that the FTO-catalysed product [^{13}C]-hm^{6}A is relatively more stable to hydrolysis). The positive charge on the *N*-1 nitrogen of hm^{1}A increases the rate of hydroxylation of formed hemiaminal. Moreover, as concluded in Chapter 4, hydroxymethylation on the endocyclic basic nitrogen is a kinetic product, with the equilibrium favouring the reactant side (i.e., nucleoside and formaldehyde formation).

5.4 Discussion

The work described in this chapter aimed to use NMR as a tool to investigate the enzymatic activity of two nucleic acid demethylases, FTO and AlkB, and to characterise the

intermediates/products formed during their catalysis. m^6A, hm^6A and m^3T were investigated as substrates for FTO, and m^1A was investigated as a substrate for AlkB. This work also aimed to characterise the recently reported FTO intermediate, f^6A, formed by further oxidation of hm^6A.[9] One important objective of this work was to correlate these results with those described in Chapter 4, regarding the stability and fragmentation of hemiaminals (hm^6A, hm^1A and m^3T). This work was also focussed on validating the hypothesis that a phosphate group attached to nucleoside is required for AlkB catalysis. NMR assay conditions were optimised for FTO and AlkB, and attempts were made to detect the formation of hydroxylated intermediates (hm^6A, hm^3T, hm^1A) and demethylated products (A and T).

The fact that hm^6A can be observed and characterised by NMR supports the observation that hm^6A is relatively stable, as compared to hm^1A and hm^3T (Chapter 4). Given the half-life of hm^6A is 3 hrs in biological conditions,[9] it is possible that hm^6A might play a regulatory role in RNA stability and function. The inverse correlation observed between hm^6A degradation and formaldehyde production suggests that formation of formaldehyde is, at least partially, not catalysed by FTO. By contrast, hm^3T and hm^1A were not observed by ^1H NMR due to their inherent susceptibility to fragmentation, as described in Chapter 4. This is likely because the positive charge on the endocyclic nitrogen (*N*-1) of hm^1A increases the rate of fragmentation. Moreover, the increased stability of hm^6A relative to hm^1A and hm^3T is possibly due to the difference between the site of hydroxymethylation (i.e., on an exocyclic nitrogen in hm^6A as compared to an endocyclic nitrogen in hm^3T and hm^1A), as supported by the data in Chapter 4. Thus, one important proposal resulting from the present study is that FTO, perhaps, should be considered as a hydroxylase rather than a demethylase. Of the AlkB human homologues, ABH8 is an established hydroxylase,[36] while the other enzymes are either thought to act as demethylases or are not functionally characterised. The classification

of FTO as a hydroxylase would be notable, as it establishes the hydroxylases and the demethylases as two sub-families of the AlkB human homologues.

Although [^{13}C]-hm^{6}A formation could be detected, attempts to observe the formation of [^{13}C]-f^{6}A by NMR were unsuccessful. One possible explanation is the modification of the enzyme active site by formaldehyde, which would likely make the protein inactive. However, even if excess formaldehyde renders the protein inactive, and hm^{6}A is the substrate of FTO, trace amounts of [^{13}C]-f^{6}A should have been observed. However, one possible rationale behind f^{6}A not being observed by NMR was that the production of f^{6}A was below the detection limit of NMR. Alternatively, f^{6}A could be produced by a non-enzymatic oxidation, which did not occur under the incubation conditions used. However, the observation of f^{6}A by mass spectrometry could be questioned, as it could occur as a consequence of the mass spectrometry conditions used, rather than being catalysed by FTO. However, the possibility of FTO-catalysed f^{6}A formation cannot be ruled out.

The substrate requirements for AlkB suggest that at least a single phosphate group must be attached to the nucleoside (e.g., 1-methyl-dAMP(5'), or m^{1}AMP), while nucleosides such as m^{1}A and m^{3}C stimulate 2OG turnover but are not demethylated.[33] However, in the above assay for AlkB, when m^{1}A was tested as a substrate in place of m^{1}AMP, resonances corresponding to formaldehyde were observed by NMR. Therefore, it can be concluded that the minimal substrate requirement for AlkB is a nucleoside, and not a nucleotide as was previously reported.[33]

Overall, optimised conditions for an NMR-based FTO assay were developed. This assay establishes FTO as a hydroxylase rather than a demethylase. These results are also consistent with the unexpected stability of hm^{6}A (as compared to hm^{1}A and hm^{3}T) as concluded in Chapter 4. This stability suggests that hm^{6}A could act as a biologically relevant RNA modification, rather than just an intermediate during the demethylation of RNA.

5.5 Acknowledgements

I thank Prof. Tim Claridge for suggesting 1D-HSQC experiments to be used for the time-course assays, Dr Richard Hopkinson for useful suggestions, and Dong Zhang for FTO and AlkB preparation and purification.

5.6 References

1. Aravind, L. & Koonin, E. V. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate-and iron-dependent dioxygenases. *Genome Biol.* **2,** 7–1 (2001).
2. Falnes, P. Ø., Johansen, R. F. & Seeberg, E. AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature* **419,** 178–182 (2002).
3. Fischer, J., Koch, L., Emmerling, C., Vierkotten, J., Peters, T., Brüning, J. C., Rüther, U. Inactivation of the Fto gene protects from obesity. *Nature* **458,** 894–898 (2009).
4. Dina, C., Meyre, D., Gallina, S., Durand, E., Körner, A., Jacobson, P., Carlsson, L. M. S., Kiess, W., Vatin, V., Lecoeur, C., Cauchi, S., Chèvre, J., Froguel, P. Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat. Genet.* **39,** 724–726 (2007).
5. Church, C., Lee, S., Bagg, E. A. L., McTaggart, J. S., Deacon, R., Gerken, T., Lee, A., Moir, L., Mecinović, J., Quwailid, M. M., Schofield, C. J., Ashcroft, F. M., Cox, R. D. A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. *PLoS Genet.* **5,** 2009 Aug;5(8):e1000599.
6. Jia, G., Yang, C. G., Yang, S., Jian, X., Yi, C., Zhou, Z., He, C. Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Lett.* **582,** 3313–3319 (2008).
7. Jia, G. Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y. G., He, C. N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature Chemical Biology*, 7, 885–887 (2011)
8. Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A. V., Patil, Deepak P., Linder, B., Pickering, B. F. Vasseur, J., Chen, Q., Gross, S.S., Elemento, O., Debart, F., Kiledjian, M., Jaffrey, S. R. Reversible methylation of m6Am in the 5′ cap controls mRNA stability. *Nature* **541,** 371–375 (2016).
9. Fu, Y. Jia, G., Pang, X., Wang, R. N., Wang, X., Li, C. J., Smemo, S., Dai, Q., Bailey, K. A., Nobrega, M. A., Han, K. L., Cui, Q., He, C. FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nat. Commun.* **4,** 1798 (2013).
10. Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature* **419,** 174–178 (2002).
11. Delaney, J. C. & Essigmann, J. M. Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 14051–6 (2004).
12. Li, D., Delaney, J. C., Page, C. M., Yang, X., Chen, A. S., Wong, C., Drennan, C. L.,

Essigmann, J. M. Exocyclic Carbons Adjacent to the N^6 of Adenine are Targets for Oxidation by the Escherichia coli Adaptive Response Protein AlkB. *J. Am. Chem. Soc.* **134,** 8896–8901 (2012).

13. Li, D. Fedeles, B. I., Shrivastav, N., Delaney, J. C., Yang, X., Wong, C., Drennan, C. L., Essigmann, J. M. Removal of N-Alkyl Modifications from N^2 -Alkylguanine and N^4 -Alkylcytosine in DNA by the Adaptive Response Protein AlkB. *Chem. Res. Toxicol.* **26,** 1182–1187 (2013).
14. Begley, T. J. & Samson, L. D. AlkB mystery solved: oxidative demethylation of N1-methyladenine and N3-methylcytosine adducts by a direct reversal mechanism. *Trends Biochem. Sci.* **28,** 2–5 (2003).
15. Yukiko Mishina, Cai-Guang Yang, and & He, C. Direct Repair of the Exocyclic DNA Adduct 1,N6-Ethenoadenine by the DNA Repair AlkB Proteins. *J. Am. Chem. Soc.*, 2005, 127 (42), pp 14594–14595.
16. Martinez, S. & Hausinger, R. P. Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases. *J. Biol. Chem.* **290,** 20702–11 (2015).
17. Rhoads, R. E. & Udenfriend, S. Decarboxylation of alpha-ketoglutarate coupled to collagen proline hydroxylase. *Proc. Natl. Acad. Sci. U. S. A.* **60,** 1473–8 (1968).
18. Cunliffe, C. J., Franklin, T. J. & Gaskell, R. M. Assay of prolyl 4-hydroxylase by the chromatographic determination of [^{14}C]succinic acid on ion-exchange minicolumns. *Biochem. J.* **240,** (1986) Dec 1; 240(2): 617–619.
19. Kaule, G. & Günzler, V. Assay for 2-oxoglutarate decarboxylating enzymes based on the determination of [1-14C]succinate: Application to prolyl 4-hydroxylase. *Anal. Biochem.* **184,** 291–297 (1990).
20. Kaule, G. & Günzler, V. Assay for 2-oxoglutarate decarboxylating enzymes based on the determination of [1-14C]succinate: Application to prolyl 4-hydroxylase. *Anal. Biochem.* **184,** 291–297 (1990).
21. McNeill, L. A., Bethge, L., Hewitson, K. S. & Schofield, C. J. A fluorescence-based assay for 2-oxoglutarate-dependent oxygenases. *Anal. Biochem.* **336,** 125–131 (2005).
22. Rose, N. R., Ng, S. S., Mecinović, J., Liénard, B. M.R., Bello, S. H., Sun, Z., McDonough, M. A., Oppermann, U., Schofield, C. J. Inhibitor Scaffolds for 2-Oxoglutarate-Dependent Histone Lysine Demethylases $^{\perp}$. *J. Med. Chem.* **51,** 7053–7056 (2008).
23. Hewitson, K. S., McNeill, L. A., Schofield, C. J., Coles, C., Mecinovic, J., Edelmann, M., Cook, K. M., Cockman, M. E., Lancaster, D. E., Kessler, B. M. Modulating the hypoxia-inducible factor signaling pathway: applications from cardiovascular disease to cancer. *Curr. Pharm. Des.* **10,** 821–33 (2004).
24. Flashman, E., Hoffart, L. M., Hamed, R. B., Bollinger Jr, J. M., Krebs, C., Schofield, C. J. Evidence for the slow reaction of hypoxia-inducible factor prolyl hydroxylase2 with oxygen. *FEBS J.* **277,** 4089–4099 (2010).
25. Dao, J. H., Kurzeja, R.J.M., Morachis, J. M., Veith, H., Lewis, J., Yu, V., Tegley, C. M.,Tagari, P. Kinetic characterization and identification of a novel inhibitor of hypoxia-inducible factor prolyl hydroxylase 2 using a time-resolved fluorescence resonance energy transfer-based assay technology. *Anal. Biochem.* **384,** 213–223 (2009).
26. Roy, T. W. & Bhagwat, A. S. Kinetic studies of *Escherichia coli* AlkB using a new fluorescence-based assay for DNA demethylation. *Nucleic Acids Res.* (2007);35(21): e147. doi:10.1093/nar/gkm1031
27. Li, D., Delaney, J. C., Page, C. M., Chen, A. S., Wong, C., Drennan, C. L., Essigmann, J. M. Repair of DNA Alkylation Damage by the *Escherichia coli* Adaptive Response Protein AlkB as Studied by ESI-TOF Mass Spectrometry. *J.*

Nucleic Acids **2010,** 369434 (2010).
28. Yi, C., Jia, G., Hou, G., Dai, Q., Zhang, W., Zheng, G., Jian, X., Yang, C. G., Cui, Q., He, C. Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature* **468,** 330–333 (2010).
29. Myllylä, R., Majamaa, K., Günzler, V., Hanauske-Abel, H. M. & Kivirikko, K. I. Ascorbate is consumed stoichiometrically in the uncoupled reactions catalyzed by prolyl 4-hydroxylase and lysyl hydroxylase. *J. Biol. Chem.* **259,** 5403–5 (1984).
30. Hwang, T. L. & Shaka, A. J. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *J. Magn. Reson. Ser. A* \ **112,** 275–279 (1995).
31. Sutherland, B. W., Toews, J. & Kast, J. Utility of formaldehyde cross-linking and mass spectrometry in the study of protein–protein interactions. *J. Mass Spectrom.* **43,** 699–715 (2008).
32. Mckay, R. T. How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: An examination of the pulse sequence components and evolution. *Concepts Magn. Reson. Part A* **38A,** 197–220 (2011).
33. Koivisto, P., Duncan, T., Lindahl, T. & Sedgwick, B. Minimal methylated substrate and extended substrate range of *Escherichia coli* AlkB protein, a 1-methyladenine-DNA dioxygenase. *J. Biol. Chem.* **278,** 44348–54 (2003).
34. Falnes, P. Ø., Bjørås, M., Aas, P. A., Sundheim, O. & Seeberg, E. Substrate specificities of bacterial and human AlkB proteins. *Nucleic Acids Res.* **32,** 3456–3461 (2004).
35. Henshaw, T. F., Feig, M. & Hausinger, R. P. Aberrant activity of the DNA repair enzyme AlkB. *J. Inorg. Biochem.* **98,** 856–861 (2004).
36. Fu, Y. Dai, Q., Zhang, W., Ren, J., Pan, T., He, C., The AlkB Domain of Mammalian ABH8 Catalyzes Hydroxylation of 5-Methoxycarbonylmethyluridine at the Wobble Position of tRNA. *Angew. Chemie Int. Ed.* **49,** 8885–8888 (2010).

Chapter 6: Experimental

6.1 General experimental

6.1.1 Chemical Synthesis

Synthesis reagents were obtained either from Sigma-Aldrich, Alfa Aesar, Cambridge Biotech, Fischer Scientific or Link Technology, unless stated otherwise. Anhydrous solvents used in reactions were either obtained commercially (Alfa Aesar) or freshly collected from distils in house. HPLC grade solvents were employed for work-up and chromatography. For chromatography purification of phosphoramidites, solvents were dried over P_2O_5 prior to use. All reactions involving moisture-sensitive reagents were carried out under argon atmosphere; glassware was oven dried and cooled under nitrogen before use. Reagents were used as supplied (analytical or HPLC grade) without prior purification. Anhydrous $MgSO_4$ was used as a drying agent.

Microwave experiments were carried out using a Biotage Initiator 8.

Thin layer chromatography was performed on aluminium plates coated with 60 F254 silica. Plates were visualised using UV light (254 nm), or 1% aq. $KMnO_4$. Flash column chromatography was performed on Kieselgel 60 silica in a glass column, or on a Biotage SP4 flash column chromatography platform. Retention factors (R_f) are quoted to a precision of 0.05.

Deuterated solvents were from Sigma and Apollo Scientific Ltd. Tris-d_{11} was obtained from Cambridge Isotope Laboratories. ^{1}H NMR and ^{13}CNMR spectra were recorded using Bruker AVIII400, AVII500, AVIII600 and AVIII700 NMR spectrometers. Fields were locked by external referencing to the relevant residual deuterium resonance. Chemical shifts (δ) are reported in ppm; coupling constants (*J*) are measured in Hz and are recorded to the nearest 0.5 Hz; when peak multiplicities are reported, the following abbreviations are used: s =

singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broadened, dd = doublet of doublets, dt = doublet of triplets, td = triplet of doublets. Spectra were recorded at room temperature unless otherwise stated.

Low resolution mass spectra (*m*/*z*) and high resolution mass spectra (HRMS) were recorded on an LCT Premier XE (Waters, MA, USA) or a microTOF (Bruker, MA, USA).

Melting points were recorded on a Gallenkamp Hot Stage apparatus. IR spectra were recorded using a Bruker Tensor 27 FT-IR spectrometer as thin films. Selected characteristic peaks are reported in cm^{-1}.

6.1.2 Buffer solution

NMR experiments were carried out in solution buffered with either 50 mM deuterated tris(hydroxymethyl)aminomethane (Tris-D11; pH 6.6 or 7.5), ammonium formate buffer (AFN; pH 7.5) or simple D_2O (pD 6.0, 7.5, 9.0; where pD = pH meter reading + 0.40)[1]. Buffers were prepared in Milli-Q H_2O or in D_2O. pH adjustments were carried out using a Jenway 3305 pH meter with a Sigma-Aldrich polymer bodied combination pH electrode. pH was adjusted using HCl and NaOH or DCl and NaOD (D_2O-based buffer).

6.1.3 Enzymatic assays-stock solutions

Water was purified using a Millipore Elix® 10 system and then passed through a 0.22 µm filter by a Millipore Milli-Q system. All stock solutions were prepared in assay buffer (50 mM, pH 7.5, 6.5), except Fe(II), which was prepared by dissolving Fe(II) ammonium sulphate in 20 mM hydrochloric acid followed by dilution with Milli-Q water/ D_2O. Disodium salt of 2OG was used, and ascorbate was used as the sodium salt.

6.2 General NMR sample handling and processing

All samples were prepared in Eppendorf tubes before being transferred to 3 mm or 5mm NMR tubes (Wilgenberg) for analysis. The spectrometer conditions were optimised before

each experiment and spectra were processed using automated routines. All spectra were acquired at 25 °C, except the 1D NOESY spectra of the dCMP-formaldehyde reaction mixture, which was conducted at 15°C (Chapter 4); the ^{1}H-NMR and 1D-HSQC analyses for enzymatic assays were conducted at 37 °C (Chapter 5).

6.3 Experimental for Chapter 2

6.3.1 Experimental details

General procedure A Synthesis of protected glycinate Inhibitors

To an overnight dried microwave vial were added aryl halide (40 mg, 0.12 mmol), K_3PO_4 (27.9 mg, 0.13 mmol), BrettPhos (**20**; 7.9 mg, 0.15 mmol), Pd precatalyst (**22**; 7.9 mg, 0.01 mmol). The solid mixture was evacuated and filled with N_2 three times. To this degassed solid mixture was added amine (x mg, 0.13 mmol) and 2.5 ml of *t*-BuOH. The reaction mixture was again evacuated and filled with N_2 three times. The mixture was then subjected to microwave radiation for 8 hours at 140ºC. The solvent was removed under reduced pressure and the residue was diluted with ethyl acetate. The organic layer was washed with 3 X H_2O and crystallised from DCM. Yields varies from 22%-86%.

General Procedure B Deprotection of Benzyl group

To a stirred solution of compound (**28**-**32**) in 2 ml of THF was added Pd/C (10 mol%). The reaction mixture was evacuated and filled with N_2 three times. The reaction was allowed to stir for 24 hours under H_2 atmosphere. The suspension was filtered through celite and the filtrate was concentrated under reduced pressure. The oily residue was washed with methanol to afford the desired acids.

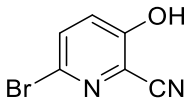
General Procedure C Deprotection of *tert*-Butyl group

To a stirred solution of compound (**33**/**34**) in THF was added KOH and the reaction was allowed to stir for 2 days at room temperature. The crude product mixture was concentrated under reduced pressure and was diluted with ethyl acetate and water. The aqueous layer was

acidified with 1M HCl to pH 3. The resulting precipitate was then filtered and washed with cold water (5 mL).

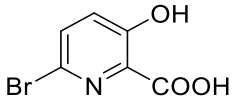
6.3.2 Synthesis and Characterisation of FTO Inhibitors

6-Bromo-3-hydroxypicolinonitrile[3] (11)



To a suspension of 3-hydroxypicolinonitrile (1.20 g, 10 mmol) in H_2O, Br_2 (0.6 mL, 11.5 mmol) was added portion-wise over 3 days at room temperature. The mixture was washed with 10% sodium thiosulfate (50 mL). The product was crystallised from H_2O, resulting in white solid (670 mg, 53%), mp 135 ºC. R_f = 0.20; IR (neat) ν/cm^{-1} 3150 (b, ArO-H), 2290 (-CN); ^{1}H NMR (400 MHz, CD_3CN) δ 7.37 (d, *J*= 9.0, 1H), 7.60 (d, *J*= 9.0, 1H), 11.56 (s,1H); ^{13}C NMR (125 MHz, CD_3CN) δ 127.5 (ArC), 128.2 (ArC)130.4 (ArC),132.9 (ArC) 142.5 (ArC) 156.6 (CN); HRMS (ESI) m/z: Calculated for $C_6H_2{}^{79}BrN_2O$ $[M-H]^-$ 196.9356, Observed 196.9350.

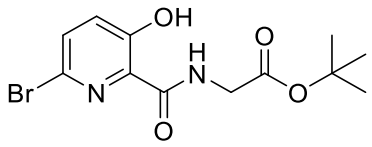
6-Bromo-3-hydroxypicolinic acid[3] (13)



To a stirred suspension of 6-bromo-3-hydroxypicolinonitrile (**11**; 2.8 g, 14.1 mmol) in methanol (50 mL), 30% w/v NaOH (44 mL, 51.8 mmol) was added and the reaction mixture was heated at reflux for 12 hours. The solvent was removed under reduced pressure and the suspension was acidified with concentrated HCl to pH 2. The resulting precipitate was filtrated and washed with cold water (5ml). Filtration afforded product **13** (3.1 g, 100 %) as white needles, mp 122-124 ºC; IR (neat) ν/cm^{-1} 3100 (ArO-H), 2850 (COO-H), 1716

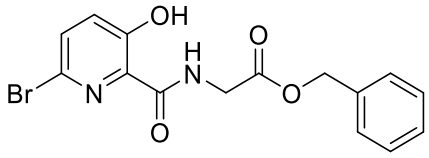
(HOC=O); ^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ 7.05-7.10 (m, 1H), 7.33-7.75 (m, 1H), 11.43 (s,1H); ^{13}C NMR (125MHz, DMSO-*d*$_6$) 131.2 (ArC), 133.7 (ArC), 135.8 (ArC); 137.9 (ArC), 143.3 (ArC), 169.1 (C=O); HRMS (ESI) m/z: Calculated for $C_6H_3{}^{79}BrNO_3$ $[M-H]^-$ 215.9302, Observed: 215.9298.

***tert*-Butyl (6-bromo-3-hydroxypicolinoyl)glycinate[3] (15)**



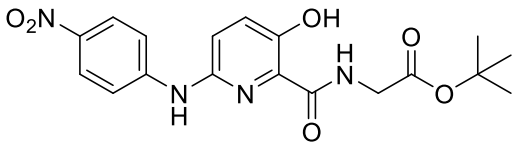
To a mixture of 6-bromo-3-hydroxypicolinic acid (**13**; 0.2 g, 0.92 mmol), DIPEA (0.5 mL, 2.75 mmol) and T3P (0.70 mL, 1.2 mmol) in Ethyl Acetate (2 mL) was added glycine *t*-butyl ester hydrochloride (0.2 g, 1.2 mmol). The reaction was heated under microwave irradiation at 120 ºC for 4 hour. The mixture was diluted with ethyl acetate and washed with water (3x10 mL) and concentrated in vacuum. The residue was purified by column chromatography (9:1 to 6:4 cyclohexane/ ethyl acetate) which resulted in pale yellow solid (0.13 g, 42 %). mp 105-106 ºC; R_f = 0.40; IR (neat) ν/cm^{-1} 3349 (NH), 1748 (*t*-BuOC=O), 1675 (NHC=O); ^{1}H NMR (500 MHz, $CDCl_3$) δ 1.44 (s, 9 H), 4.04 (d, *J*= 5.5, 2 H), 7.13 (d, *J*= 9.0, 1 H), 7.40 (d, *J*= 9.0, 1 H), 8.10 (br. s,1 H), 11.83 (s, 1 H); ^{13}C NMR (125MHz, $CDCl_3$) δ 28.0 (C-(**C**H_3)$_3$); 41.5 (**C**H_2), 82.7 (**C**$(CH_3)_3$), 129.0 (ArC), 129.4 (ArC), 131.3(ArC), 133.2 (ArC), 157.3 (ArC), 167.7 (C=O), 168.0 (C=O); HRMS (ESI) m/z: Calculated for $C_{12}H_{15}BrN_2NaO_4$ $[M+Na]^+$ 353.0107, Observed: 353.0095.

Benzyl-(6-bromo-3-hydroxypicolinoyl)-glycinate[3] (14)



To a mixture of 6-bromo-3-hydroxypicolinic acid (**13**; 0.20 g, 0.92 mmol), DIPEA (0.5 mL, 2.75 mmol) and T3P (0.7 mL, 1.19 mmol) in ethyl acetate (2 mL) was added glycine *t*-butyl ester hydrochloride (0.2 g, 1.20 mmol). The reaction was heated under microwave irradiation at 120 ºC for 4 hours. H_2O (10 mL) was added to the mixture and was washed with ethyl acetate. Organics were combined, dried over MgSO4 and concentrated in vacuum. Chromatography (cyclohexane/ethyl acetate, 9:1 to 6:4) afforded (0.13 g, 42%) as a pale yellow solid. mp 102-104 °C; R_f = 0.45; IR (neat) ν/cm^{-1} 3391(Ar-OH) 3035(N-H) 2949 (COO-H) 1741 (PhOC=O) 1654 (NHC=O); ^{1}H NMR (400 MHz, $CDCl_3$) δ 4.19 (d, *J*=6.0, 2 H) 5.17 (s, 2 H) 7.15 (d, *J*=9.0, 1 H) 7.19 (s, 2 H) 7.31 (m, 4 H) 8.12 (b.s., 1 H) 11.72 (s, 1 H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 40.9 (**C**H_2) 67.5 (**C**H_2) 128.5 (ArC) 128.7 (ArC) 128.7 (ArC) 129.1 (ArC) 129.5 (ArC) 131.2 (ArC) 133.5 (ArC) 135.0 (ArC) 157.4 (ArC) 168.0 (**C**=O) 168.8 (**C**=O); HRMS (ESI) m/z: Calculated for $C_{15}H_{12}BrN_2O_4$ $[M-H]^-$ 362.9996, Observed: 362.9981.

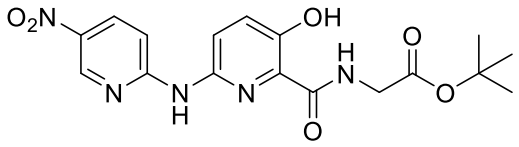
***tert*-Butyl (3-hydroxy-6-((4-nitrophenyl)amino)picolinoyl)glycinate (33)**



Prepared according to general procedure A, aryl **15** and amine **23** afforded **33** (40mg, 86%) as yellow solid. mp 196-200 °C; IR (neat) ν/cm^{-1} 3400 (N-H), 3367 (ArO-H), 1741 (*t*-BuC=O); 1645 (NHC=O), 1502, 1319 (N-O); ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.56 (s, 9 H) 4.15 (d, *J*=5.0, 2 H) 6.82 (br. s., 1 H) 7.07 (d, *J*=9.0, 1 H) 7.36 (d, *J*=9.0, 1 H) 7.45 (d, *J*=1.0,

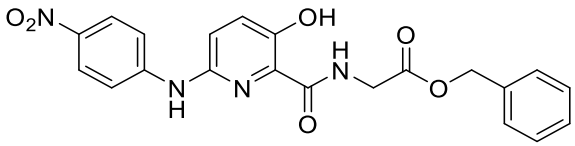
2 H) 8.18 (br. s., 1 H) 8.24 (d, *J*=1.0, 2 H) 11.60 (s, 1 H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 28.1 (C-$(CH_3)_3$) 41.6 (CH_2) 83.0 (C-$(CH_3)_3$) 116.1 (ArC) 118.9 (ArC) 125.8 (ArC) 127.4 (ArC) 129.6 (ArC) 141.2 (ArC) 144.6 (ArC) 146.9 (ArC) 153.5 (ArC) 168.3 (C=O) 168.5 (C=O); HRMS (ESI$^-$) m/z: Calculated for $C_{14}H_{12}N_4O_6$ [M-H]$^-$ 331.0684, Observed: 331.0684.

***tert*-Butyl (3-hydroxy-6-((5-nitropyridin-2-yl)amino)picolinoyl)glycinate (34)**



Prepared according to general procedure A, aryl **15** and amine **24** afforded **34** (34 mg, 73%) as yellow solid. mp 196-200 °C; IR (neat) ν/cm^{-1} 3387 (N-H), 3300 (ArO-H), 1742 (*t*-BuC=O); 1647 (NHC=O), 1504, 1382 (N-O); ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.47 (s, 9 H) 4.07 (d, *J*=5.0, 2 H) 7.32-7.37 (m, 2 H) 7.59 (s, 1 H) 7.75 (d, *J*=9.0, 1 H) 8.10-8.16 (m, 1 H) 8.33 (s, 1 H) 9.07 (t, *J*=5.0, 1 H) 11.56 (s, 1 H); ^{13}C NMR (100 M, $CDCl_3$) δ 28.1 (C-(**C**$H_3)_3$) 41.6 (**C**H_2) 68.3 ((**C**-$(CH_3)_3$) 109.4 (ArC) 120.2 (ArC) 122.3 (ArC) 129.4 (ArC) 133.5 (ArC) 138.1 (ArC) 143.1 (ArC) 145.9 (ArC) 154.2 (ArC) 157.1 (ArC) 168.2 (C=O) 168.7 (C=O); HRMS (ESI) m/z: Calculated for $C_{17}H_{18}N_6O_8$ [M-H]$^-$ 388.1262, Observed: 388.1266.

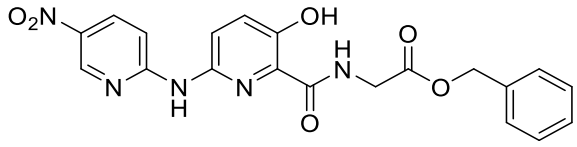
Benzyl (3-hydroxy-6-((4-nitrophenyl)amino)picolinoyl)glycinate (28)



Prepared according to general procedure A, aryl **14** and amine **23** afforded **28** (35 mg, 76%) as orange solid. mp 194-196 °C; IR (neat) ν/cm^{-1} 3340 (N-H), 3289 (ArO-H), 1732 (*t*-BuC=O), 1633 (NHC=O), 1469, 1319 (NO_2); ^{1}H NMR (400 MHz, DMSO-d_6) δ 4.28 (d, *J*=6.0, 2 H) 5.20 (s, 2 H) 7.32 - 7.43 (m, 5 H) 7.46 (d, *J*=9.0, 1 H) 7.58 (d, *J*=9.0, 1 H) 7.99

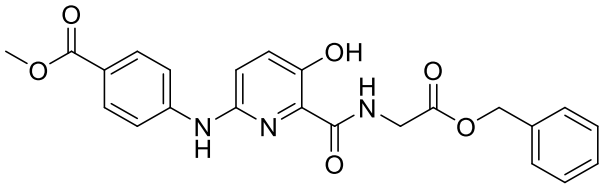
(d, *J*=1.0, 2 H) 8.17 (d, *J*=1.0, 2 H) 8.75 (t, *J*=6.0, 1 H) 10.99 (s, 1 H) 11.99 (s, 1 H); ^{13}C NMR (100 MHz, DMSO-d_6) 49.0 (**C**H_2) 66.6 (**C**H_2) 116.5 (ArC) 126.0 (ArC) 126.4 (ArC) 127.4 (ArC) 128.5 (ArC) 128.9 (ArC) 136.2 (ArC) 139.4 (ArC) 147.0 (ArC) 148.7 (ArC) 149.5 (ArC) 153.0 (ArC) 166.4 (C=O) 168.9 (C=O); HRMS (ESI) m/z: Calculated for $C_{21}H_{16}N_4O_6$ [M-H]$^-$ 421.1148. Observed: 421.1163.

Benzyl (3-hydroxy-6-((5-nitropyridin-2-yl)amino)picolinoyl)glycinate (29)



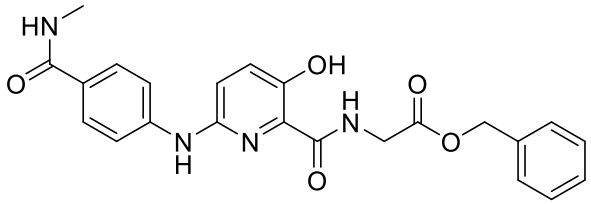
Prepared according to general procedure A, aryl **14** and amine **24** afforded **29** (30 mg, 65%) as orange solid. mp 190-195 °C; IR (neat) ν/cm^{-1} 3352 (N-H), 3340 (Ar-OH), 1740 (MeOC=O), 1644 (BzOC=O), 1605 (NHC=O), 1542, 1324 (N-O); ^{1}H NMR (400 MHz, $CDCl_3$) δ 4.22 (d, *J*=6.0, 2 H) 5.21 (s, 2 H) 7.25 - 7.43 (m, 7 H) 7.57 (br. s., 1 H) 7.69 - 7.78 (m, 1 H) 8.11 (br. s., 1 H) 8.33 (d, *J*=10.0, 1 H) 9.00 - 9.21 (m, 1 H) 11.47 (s, 1 H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 40.5 (CH_2) 67.6 (CH_2) 109.4 (ArC) 116.5 (ArC) 118.5 (ArC) 120.25 (ArC) 122.8 (ArC) 129.4 (ArC) 133.5 (ArC) 138.1 (ArC) 143.1 (ArC) 145.9 (ArC) 154.2 (ArC) 157.1 (ArC) 168.2 (C=O) 168.7(C=O); HRMS (ESI) m/z: Calculated for $C_{20}H_{16}N_5O_6$ [M-H]$^-$ 422.1106, Observed: 422.1110.

Methyl 4-((6-((2-(benzyloxy)-2-oxoethyl)carbamoyl)-5-hydroxypyridin-2-yl)amino) benzoate (32)



Prepared according to general procedure A, aryl **14** and amine **27** afforded **32** (15 mg, 32%) as yellow solid. mp 185-189 °C; IR (neat) ν/cm^{-1} 3408 (N-H), 3352 (Ar-OH), 1740 (MeOC=O), 1690 (BzOC=O), 1651 (NHC=O); ^{1}H NMR (400 MHz, $CDCl_3$) δ 3.83 (s, 3 H) 4.14 - 4.25 (m, 2 H) 5.19 (s, 2 H) 6.59 (s, 1 H) 6.97 (d, *J*=9.0, 1 H) 7.20 - 7.35 (m, 8 H) 7.88 - 7.97 (m, 2 H) 8.13 (t, *J*=6.0, 1 H) 11.33 (s, 1 H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 41.0 (CH_2) 51.9 (CH_3) 67.6 (CH_2) 116.5 (ArC) 118.5 (ArC) 122.8 (ArC) 127.0 (ArC) 128.6 (ArC) 128.7 (ArC) 128.7 (ArC) 129.4 (ArC) 131.3 (ArC) 135.0 (ArC) 145.2 (ArC) 145.6 (ArC) 152.9 (ArC) 166.9 (C=O) 168.8 (C=O) 169.3 (C=O); HRMS (ESI) m/z: Calculated for $C_{23}H_{21}N_3O_6$ $[M-H]^-$ 434.1357, Observed: 434.1359.

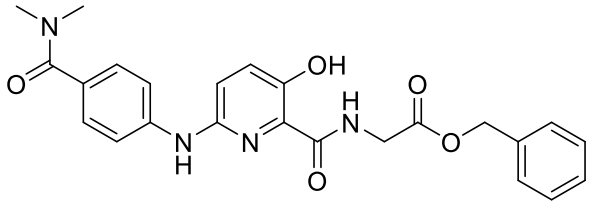
Benzyl (3-hydroxy-6-((4-(methylcarbamoyl)phenyl)amino)picolinoyl)glycinate(30)



Prepared according to general procedure A, aryl **14** and amine **25** afforded **30** (30 mg, 64%) as yellow solid. mp 165-167 °C; IR (neat) ν/cm^{-1} 3402 (NH), 3300 (ArOH), 1741(BzOC=O), 1646 (HNC=O), 1602 (HNC=O); ^{1}H NMR (400 MHz, $CDCl_3$) δ 2.95 (d, *J*=6.0, 3 H) 4.20 (d, *J*=6.0, 2 H) 5.19 (s, 2 H) 5.96 - 6.07 (m, 1 H) 6.48 (s, 1 H) 6.93 (d, *J*=9.0, 1 H) 7.20 - 7.35 (m, 8 H) 7.66 (d, *J*=8.5, 2 H) 8.13 (t, *J*=6.0, 1 H) 11.30 (s, 1 H); ^{13}C NMR (100 MHz, $CDCl_3$) 41.0 (CH_2) 51.9 (CH_3) 67.6 (CH_2) 116.5 (ArC) 118.5 (ArC) 122.8 (ArC) 127.0 (ArC) 128.6

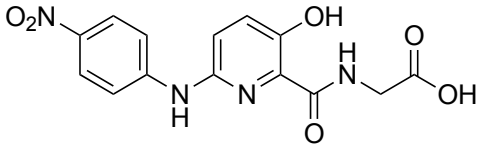
(ArC) 128.7 (ArC) 128.73 (ArC) 129.4 (ArC) 131.3 (ArC) 135.0 (ArC) 145.2 (ArC) 145.6 (ArC) 152.9 (ArC) 166.9 (C=O) 168.8 (C=O) 169.3 (C=O); HRMS (ESI) m/z: Calculated for $C_{23}H_{21}N_4O_5$ $[M-H]^-$ 433.1517, Observed: 433.1523.

Benzyl (3-hydroxy-6-((4-(dimethylcarbamoyl)phenyl)amino)picolinoyl)glycinate (31)



Prepared according to general procedure A, aryl **14** and amine **26** afforded **31** (32 mg, 65%) as yellow solid. mp 201-204 °C; IR (neat) ν/cm^{-1} 3323 (b,NH, ArOH), 1750 (BzOC=O), 1632 (HNC=O), 1601 (Me_2NC=O); ^{1}H NMR (400 MHz, $CDCl_3$) δ 3.02 (b.s., 6 H) 4.20 (d, *J*=6.0, 2 H) 5.18 (s, 2 H) 6.38 (s, 1 H) 6.94 (d, *J*=9.0, 1 H) 7.21 - 7.34 (m, 8 H) 7.34 - 7.39 (m, 2 H) 8.16 (t, *J*=6.0, 1 H) 11.28 (s, 1 H); ^{13}C NMR (100 MHz, CD_3COCD_3-d_6) δ 26.6 (CH_3) 40.8 (CH_2) 66.5 (CH_2) 113.0 (ArC) 116.6 (ArC) 119.5 (ArC) 126.0 (ArC) 128.2 (ArC) 128.5 (ArC) 128.6 (ArC) 129.0 (ArC) 129.2 (ArC) 136.1 (ArC) 143.1 (ArC) 147.5 (ArC) 152.2 (ArC) 169.2 (C=O) 169.5 (C=O) 170.6 (C=O); HRMS (ESI^-) m/z: Calculated for $C_{24}H_{23}N_4O_5$ $[M-H]^-$ 447.1674, Observed: 447.1675.

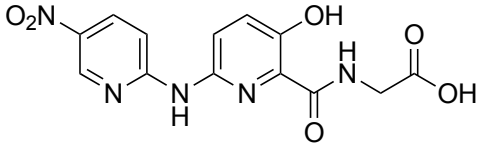
3-Hydroxy-6-((4-nitrophenyl)amino)picolinoyl)glycine (40)



Prepared according to general procedure C, glycinate **33** (20 mg, 0.05 mmol) afforded **40** (10 mg, 66%) as yellow solid. mp >300 °C; IR(neat) ν/cm^{-1} 3387(NH), 3386(ArOH), 2939 (COO-H) 1728(HOC=O), 1647(HNC=O), 1577, 1334(N-O); ^{1}H NMR (400 MHz, DMSO-d_6)

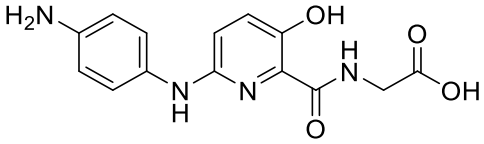
δ 4.09 (d, *J*=5.0, 2 H) 7.22 (d, *J*=9.0, 1 H) 7.47 (d, *J*=9.0, 1 H) 7.81 (d, *J*=1.0, 2 H) 8.18 (d, *J*=1.0, 2 H) 8.83 (b.s.,1H) 9.96 (t, *J*=5.0, 1 H) 11.98 (s, 1 H) 12.96 (b.s., 1 H); ^{13}C NMR (100 MHz, DMSO-*d*$_6$) δ 28.5 (CH_2) 112.4 (ArC) 121.5 (ArC) 128.6 (ArC) 130.3 (ArC) 133.8 (ArC) 137.5 (ArC) 144.6 (ArC) 158.2 (ArC) 169.0 (ArC) 172.7 (C=O) 178.5 (C=O); HRMS (ESI) m/z: Calculated for $C_{16}H_{15}N_4O_5$ $[M-H]^-$ 331.0684, Observed: 331.0685.

3-Hydroxy-6-((5-nitropyridin-2-yl)amino)picolinoyl)glycine (41)



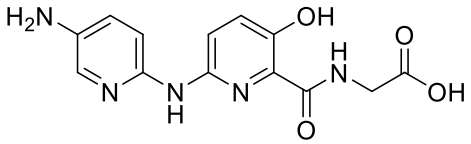
Prepared according to general procedure C, glycinate **34** (20 mg, 0.05 mmol) afforded **41** (14 mg, 78%) as yellow solid. mp >300 °C; IR (neat) ν/cm^{-1} 3400(NH), 3300(ArOH), 2856 (COO-H) 1725 (HOC=O), 1647(HNC=O), 1577, 1334(N-O); ^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ 4.09 (d, *J*=6.0, 2 H) 7.53 (d, *J*=9.0, 1 H) 7.77 (d, *J*=9.0, 1 H) 7.84 (d, *J*=9.0, 1 H) 8.39 (dd, *J*=9.0, 3.0, 1 H) 8.66 (t, *J*=6.0, 1 H) 9.09 (d, *J*=3.0, 1 H) 10.66 (s, 1 H) 11.98 (s, 1 H); ^{13}C NMR (100 MHz, DMSO-*d*$_6$) δ 21.5 (CH_2) 110.4 (ArC) 121.9 (ArC) 127.5 (ArC) 129.5 (ArC) 133.8 (ArC) 137.5 (ArC) 144.6 (ArC) 146.0 (ArC) 153.8 (ArC) 158.2 (ArC) 171.0 (C=O) 172.4 (C=O); HRMS (ESI) m/z: Calculated for $C_{13}H_{10}N_5O_6$ $[M-H]^-$ 332.0636, Observed: 332.0636.

6-((4-Aminophenyl)amino)-3-hydroxypicolinoyl)glycine (35)



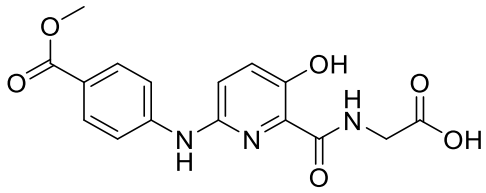
Prepared according to general procedure B, glycinate **28** (20 mg, 0.05 mmol) afforded **35** (6 mg, 43%) as violet solid. mp >300 °C; IR (neat) ν/cm^{-1} 3298 (b, N-H, ArO-H), 2952 (COO-H), 1720 (HOC=O), 1630 (HNC=O); ^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ 4.22 - 4.29 (m, 2 H) 5.21 (br. s., 2 H) 7.30 - 7.46 (m, 6 H) 7.66 (br. s., 1 H) 8.18 (t, *J*=6.0, 1 H) 11.70 (br. s., 1 H); ^{13}C NMR (100 MHz, DMSO-*d*$_6$) δ 49.0(CH_2) 116.5 (ArC) 126.0(ArC) 126.4 (ArC) 128.5 (ArC) 128.6 (ArC) 128.9 (ArC) 129.7 (ArC) 136.2 (ArC) 139.4 (ArC) 153.0 (ArC) 166.4 (C=O) 169.8 (C=O); HRMS (ESI$^-$) m/z: Calculated for $C_{16}H_{14}N_3O_6$ $[M-H]^-$ 302.1015, Observed: 302.1008.

6-((5-Aminopyridin-2-yl)amino)-3-hydroxypicolinoyl)glycine (**36**)



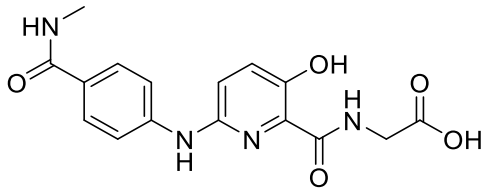
Prepared according to general procedure B, glycinate **34** (20 mg, 0.047mmol) afforded **36** (3mg, 21 %) as violet solid. mp > 300°C; IR (neat) ν/cm^{-1} 3250 (b, N-H, ArO-H), 3000 (COO-H), 1721 (HOC=O), 1680 (HNC=O); ^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ 4.16 (d, *J*=6.0, 2 H) 7.47 (d, *J*=9.0, 1 H) 7.56 - 7.62 (m, 1 H) 7.84 - 7.89 (m, 2 H) 8.14 - 8.17 (m, 2 H) 8.51 (t, *J*=6.0, 1 H) 8.98 (s, 1 H) 9.14 (s, 1 H) 11.49 (br. s., 1 H) 11.67 (s, 1 H); ^{13}C NMR (100 MHz, DMSO-*d*$_6$) 41.6 (**C**H_2) 109.4 (ArC) 120.2 (ArC) 126.3 (ArC) 128.7 (ArC) 129.6 (ArC) 136.2 (ArC) 143.1 (ArC) 145.9 (ArC) 154.2 (ArC) 157.08 (ArC) 165.8 (C=O) 170.2 (C=O); HRMS (ESI) m/z: Calculated for $C_{14}H_{13}N_4O_4$ $[M-H]^-$ 301.0942, Observed: 301.0942.

3-Hydroxy-6-((4-(methoxycarbonyl)phenyl)amino)picolinoyl)glycine (**39**)



Prepared according to general procedure B, glycinate **32** (20 mg, 0.05 mmol) afforded **39** (13 mg, 82%) as white solid. mp 264-267 °C; IR (neat) ν/cm^{-1} 3400 (NH), 3370 (ArOH), 2930 (COO-H) 1710 (MeOC=O), 1682 (HOC=O), 1592 (HNC=O); ^{1}H NMR (400 MHz, DMSO-d_6) δ 3.81 (s, 3 H) 4.09 (d, *J*=5.5, 2 H) 7.16 (d, *J*=9.0, 1 H) 7.41 (d, *J*=9.0, 1 H) 7.73 (d, *J*=1.0, 2 H) 7.91 (d, *J*=1.0, 2 H) 8.49 (t, *J*=5.5, 1 H) 9.55 (s, 1 H) 11.87 (br. s., 1 H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 49.1 (**C**H$_2$) 52.0 (**C**H$_3$) 116.5 (ArC) 120.5 (ArC) 120.9 (ArC) 126.4 (ArC) 129.8 (ArC) 131.1 (ArC) 146.5 (ArC) 147.2 (ArC) 152.3 (ArC) 166.5 (**C**=O) 169.2 (**C**=O) 171.2 (C=O); HRMS (ESI) m/z: Calculated for $C_{16}H_{14}N_3O_6$ [M-H]$^-$ 344.0888, Observed: 334.0888.

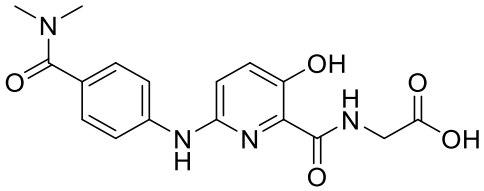
3-Hydroxy-6-((4-(methylcarbamoyl)phenyl)amino)picolinoyl)glycine (**37**)



Prepared according to general procedure B, glycinate **30** (30 mg, 0.07 mmol) afforded **37** (12 mg, 51%) as yellow solid. mp >300 °C; IR(neat) ν/cm^{-1} 3402(NH), 3330 (ArOH), 2946 (COO-H) 1728 (HOC=O), 1645 (HNC=O), 1602 (HNC=O); ^{1}H NMR (400 M, DMSO-d_6) δ 2.77 (d, *J*=4.5, 3 H) 4.08 (d, *J*=5.5, 2 H) 7.13 (d, *J*=9.0, 1 H) 7.39 (d, *J*=9.0, 1 H) 7.65 (d, *J*=1.0, 2 H) 7.80 (d, *J*=1.0, 2 H) 8.18-8.21 (m, 2 H) 8.52 (t, *J*=5.5, 1 H) 9.31 (s, 1 H) 11.82 (s, 1 H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 26.6 (CH$_3$) 49.1 (CH$_2$) 116.4 (ArC) 120.2 (ArC) 126.3 (ArC) 128.7 (ArC) 129.6 (ArC) 144.6 (ArC) 147.6 (ArC) 152.0 (ArC) 166.8 (ArC)

169.4 (ArC) 171.1 (C=O) 180.0 (C=O) 208.8 (C=O); HRMS (ESI) m/z: Calculated for $C_{16}H_{15}N_4O_5$ $[M-H]^-$ 343.1048, Observed: 343.1048.

6-((4-(Dimethylcarbamoyl)phenyl)amino)-3-hydroxypicolinoyl)glycine (38)



Prepared according to general procedure B, glycinate **31** (12 mg, 0.03 mmol) afforded **38** (3 mg, 30%) as white solid. mp >300 °C; IR(neat) ν/cm^{-1} 3449 (NH), 3387 (ArOH), 2863 (COO-H) 1730 (HOC=O), 1647 (HNC=O), 1592 (HNC=O); ^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ 2.98 (s, 6 H) 4.02 (d, *J*=5.5, 2 H) 7.12 (d, *J*=9.0, 1 H) 7.37 (d, *J*=8.0, 3 H) 7.66 (d, *J*=9.0, 2 H) 8.50 (t, *J*=5.5, 1 H) 9.27 (s, 1 H) 11.82 (br. s., 1 H); ^{13}C NMR (100 MHz, DMSO-*d*$_6$) δ 21.5 (CH_2) 49.1 (CH_3) 116.6 (ArC) 112.0(ArC) 126.3 (ArC) 127.9 (ArC) 129.0 (ArC) 129.6 (ArC) 143.2 (ArC) 147.7 (ArC) 151.8 (ArC) 169.2 (C=O) 170.7(C=O) 171.0 (C=O); HRMS (ESI) m/z: Calculated for $C_{14}H_{12}N_4O_6$ $[M-H]^-$ 357.1204, Observed: 357.1205.

6.4 Binding Experiment

The CPMG displacement experiments were carried out using the AV700 instrument. Experiments were conducted at 298 K. The PROJECT-CPMG sequence was used.[2] Experimental parameters were as follows: total echo time 48 ms (τ = 2 ms, n = 6) and number of scans 200-400. Water suppression was achieved by presaturation. For binding experiments, given concentration of inhibitors were titrated into the assay, and the 2OG displacement was measured via reappearance of free 2OG signals.

6.5 FTO expression and purification

The expression and purification of hFTO was carried out as reported.[3,4] The reported procedure is as follows: E. coli BL21 (DE3) transformed with the pET28a_hFTO plasmid (encoding for N-terminally hexahistidine-tagged full-length human FTO without the thrombin cleavage site between the hexahistidine tag and the protein) or pET28a_hFTOΔ31 (encoding for N-terminally hexahistidine-tagged human FTO with residues 2−31 deleted from the N terminus) were grown (37 °C; 180 rpm) to an OD_{600} of 0.6. FTO production was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and growth was continued at 16 °C for 16 h (hFTO). The resultant cell pellets were stored at −80 °C. The cell pellets were thawed and resuspended in 20 mM Tris−HCl, pH 7.5; 500 mM NaCl; and 10 mM imidazole, then lysed by sonication on ice. The lysate was cleared by centrifugation, and the supernatant was loaded onto a 5 mL HisTrap FF column (GE Healthcare) and purified using an AKTA FPLC system. The column was treated with 20 mM Tris−HCl, 500 mM NaCl, and 40 mM imidazole, then eluted with 20 mM Tris−HCl, pH 7.5; 500 mM NaCl; and 500 mM imidazole. hFTO was then buffer-exchanged into 25 mM Tris−HCl, pH 7.5, and concentrated to 20-40 mg/mL for storage.

6.6 Apo Protein

Apo protein was made by treating the purified protein with EDTA as follows: The fraction containing the purified protein were diluted to a final concentration of less than 1 mg/mL in 200 mM EDTA and 15 mM ammonium acetate (pH 7.0 to 8.0). The mixture was incubated overnight at 4°C and then concentrated to 2 mL volume using Amicon Ultra-4 or Amicon Ultra-15 centrifugal filter units.

6.7 In Vitro Demethylation Assay

For catalytic assays, the assay conditions used as reported.[3] The reported procedure is as follows: a 50 µL reaction mixture containing final concentrations of 3 µM FTO; 70 µM 3meT nucleoside; 160 µM 2OG; 500 µM L- ascorbate; 100 µM diammonium iron(II) sulfate complex; 200 µM test compound (2% DMSO in control reaction); and 50 mM 2-(N-morpholino)ethanesulfonic acid (MES), pH 6.3, was incubated at room temperature for 1 h. The reaction was quenched with methanol (50 µL) then centrifuged to remove precipitated protein. The supernatant was then dried using an Eppendorf Speedvac concentrator and reconstituted with (50 µL) water. The product (thymidine) and substrate (3meT) were separated using a Waters Acquity Ultra Performance Liquid Chromatography (UPLC) BEH C18 column (130 Å, 1.7 µm, 2.1 mm × 50 mm) with a gradient of 95% A (H_2O with 1% formic acid) to 80% B (methanol with 0.1% formic acid) over 5 min. The UV detection wavelength was set to 266 nm room temperature (thymidine) = 0.95 min, room temperature (3meT) = 1.48 min. The peaks were identified using a positive ion mode in the ESI-TOF mass spectrometry (Waters LCT Premier XE machine). The UV peaks were integrated using MassLynx software (Agilent), and the percentage conversion of 3meT to thymidine was used to quantify the activity of FTO in the presence of inhibitors. For IC_{50} determinations, assays were carried out in triplicate with a range of compound concentrations (0, 10, 30, 100, and 300 µM and 1, 3, and 10 mM). IC_{50} values were calculated from dose− response curves plotted using GraphPad Prism 5.

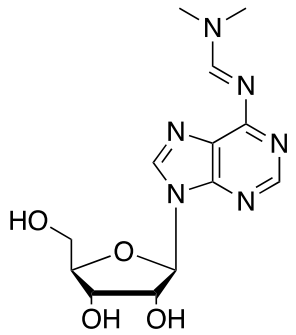
6.8 Dynamic Scanning Fluorometry Assays

The assay has been carried out as reported.[3,5] The procedure reported is as follows: a mixture (50 µL) containing 2 µM full-length FTO protein; 1× SYPRO orange dye (Invitrogen) (from

commercial stock that is 5000×); 50 μM diammonium iron(II) sulfate complex; 200 μM test compound (or 2% DMSO control); and 50 mM HEPES, pH 7.5, was prepared. Assays were performed using a MiniOpticon Real-Time PCR Detection System (Bio-Rad). The sample plate was inserted into the machine, and fluorescence readings were taken between 25 and 90 °C, increasing the temperature linearly over 1 hr 20 min. The excitation and emission wavelengths were 470 and 570 nm, respectively. Melting curves were analysed using GraphPad Prism.

6.9 Experimental for Chapter 3

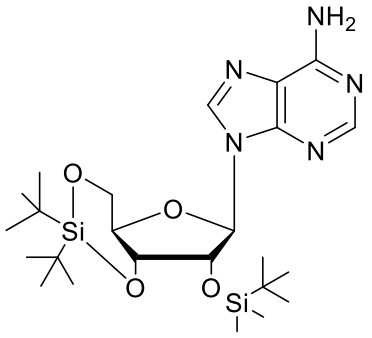
6-*N*-(1-Dimethylamino)methylideneadenosine[6] (52)



To a stirred solution of adenosine (**50**, 270 mg, 1 mmol), dried by co-evaporating with pyridine, in 10 mL anhydrous DMF was added *N*,*N*-dimethylformamide dimethyl acetal (**51**; 1 mL, 5.5 mmol) and the resultant mixture was heated at 40°C for 1 hour under N_2. The reaction mixture was then evaporated to dryness under reduced pressure and the resultant was dissolved in ethyl acetate forming a suspension; the precipitate was collected through suction filtration. The filtrate was discarded, and the white precipitate was washed with ethyl acetate and dried in *vacuo* which resulted in **52** in 30% yield. ^{1}H NMR (700 MHz, DMSO-d_6) δ 3.34 (s, 6H), 3.54 – 3.59 (m, 1H), 3.68 (dd, J = 12.0, 3.5 Hz, 1H), 3.97 (q, J = 3.5 Hz, 1H), 4.15 (dd, J = 5.0, 3.0 Hz, 1H), 4.61 (t, J = 5.5 Hz, 1H), 5.14 – 5.26 (m, 1H), 5.43 (b.s, 2H), 5.88

(d, *J* = 6.0 Hz, 1H), 7.34 (s, 2H), 8.14 (s, 1H), 8.35 (s, 1H); ^{13}C NMR (176 MHz, DMSO-*d*$_6$) δ 40.5, 62.2, 71.1, 73. 9, 86.4, 88.4, 119.8, 140.4, 149.5, 152.8, 156.6; HRMS (ESI) m/z: Calculated for $C_{13}H_{18}O_4N_6$ $[M+H]^+$ 323.1462, Observed:323.1459.

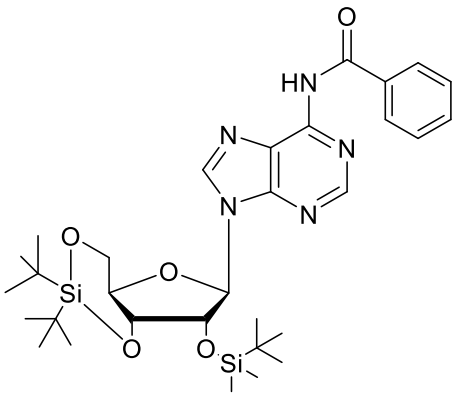
3', 5'-*O*-Bis(*tert*-butylsilyl)-2'-*O*-(*tert*-butyldimethylsilyl)adenosine[7] **(56)**



To a stirred suspension of adenosine (2.10 g, 8 mmol) in 40 mL anhydrous DMF at 0 °C, di-*t*-butylsilyl ditriflate (3.9 mL, 8.8 mmol) was added dropwise under N_2 atmosphere. After full consumption of starting material (TLC, 30 min), the reaction was quenched immediately with imidazole (2.7 g, 40 mmol) at 0°C. After 5 minutes, the reaction was warmed to room temperature. *t*-Butyldimethylsilyl chloride (1.45 g, 9.6 mmol) was added portion wise; the reaction was then refluxed at 60°C for 12 hour. The suspension was cooled down to room temperature, water was added and the precipitate was collected by suction filtration. The filtrate was discarded, and the white precipitate was washed with cold methanol. The methanol layer was evaporated under vacuum and the product was crystallised from CH_2Cl_2 which resulted in a white solid (3.8 g, 91%). mp 191-193.4°C; TLC R_f 0.45 (3:2 cyclohexane/ethyl acetate); ^{1}H NMR (600 MHz, $CDCl_3$) δ 0.17 (s, 3H), 0.19 (s, 3H), 0.95 (s, 9H), 1.07 (s, 9H), 1.11 (s, 9H), 4.06 (dd, *J* = 10.5, 9.0 Hz, 1H), 4.24 (dd, *J* = 10.5, 5.0 Hz, 1H), 4.51 (dd, *J* = 9.0, 5.0 Hz, 1H), 4.57 (dd, *J* = 9.0, 5.0 Hz, 1H), 4.65 (d, *J* = 5.0 Hz, 1H), 5.72 (b.s, 2H), 5.93 (b.s, 1H), 7.84 (s, 1H), 8.35 (s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$) δ -6.3, -6.0, 16.9, 20.4, 22.8, 24.6, 27.0, 27.5, 62.4, 70.5, 74.5, 75.7, 83.0, 126.9, 143.6, 152.3, 154.4,

157.5. HRMS (ESI) m/z: Calculated for $C_{24}H_{44}O_4N_5{}^{28}Si_2$ $[M+H]^+$ 522.2926, Observed: 522.2923.

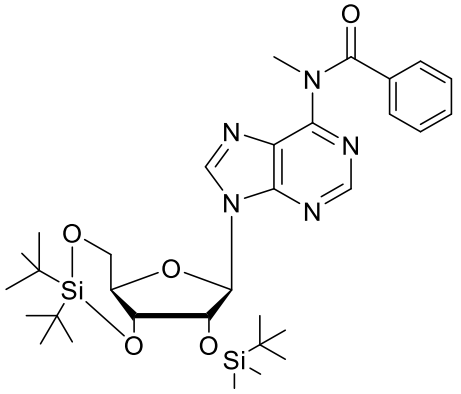
3',5'-*O*-Bis(*tert*-butyl)silyl-2'-*O*-(*tert*-butyldimethyl)silyl-N^6-benzoyladenosine[7] (57)



To a stirred solution of 3',5'-*O*-bis(*t*-butylsilyl)-2'-*O*-(*tert*-butyldimethylsilyl)adenosine (2.1 g, 4 mmol) in 10 mL anhydrous pyridine at -5°C (using a salt-ice bath), benzoyl chloride (BzCl) (0.9 mL, 8 mmol) was added dropwise. The reaction mixture was allowed to warm to room temperature for one hour, then stirred for 18 hours under N_2. The reaction was quenched first with 6 mL methanol at 0 °C and then with excess methanolic ammonia (1 ml; 7 N NH_3) after 1 hour. The crude product mixture was concentrated *in vacuo*, then diluted with ethyl acetate and saturated $NaHCO_3$ solution. The organic layer was washed with water (3 x 10 mL) and dried (anhydrous $MgSO_4$) and again concentrated *in vacuo*. The mixture was purified (6:1 to 3:1 cyclohexane/ethyl acetate) by column chromatography, which resulted in an off-white solid (2.2 g, 88%). mp 127-128.2 °C; TLC R_f 0.35 (2:3 cyclohexane/ethyl acetate); ^{1}H NMR (400 MHz, $CDCl_3$) δ 0.00 (s, 3H), 0.02 (s, 3H), 0.78 (s, 9H), 0.89 (s, 9H), 0.92 (s, 9H), 3.88 (dd, *J* = 10.5, 9.0 Hz, 1H), 4.09 (dd, *J* = 10.5, 5.0 Hz, 1H), 4.25-4.38 (m, 2H), 4.47 (d, *J* = 5.0 Hz, 1H), 5.84 (b.s, 1H), 7.40 – 7.32 (m, 2H), 7.42 – 7.48 (m, 1H), 7.83 – 7.89 (m, 3H), 8.60 (s, 1H), 8.85 (s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$) δ -5.0, -4.2,18.3, 20.4, 21.1, 25.9, 27.0, 27.5, 67.8, 74.8, 75.5, 75.9, 92.5, 123.8, 127.9, 128.9, 132.8, 133.5, 141.1,

149.6, 151.2, 152.7, 164.6; HRMS (ESI) m/z: Calculated for $C_{31}H_{48}O_5N_5{}^{28}Si_2$ $[M+H]^+$ 626.3188, Observed: 626.3185.

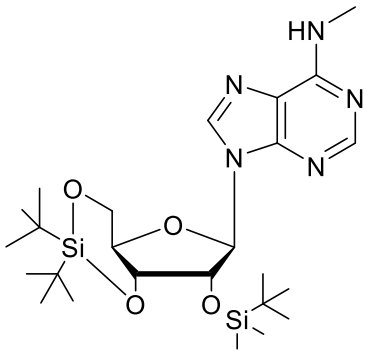
3',5'-*O*-Bis(*tert*-butylsilyl)-2'-*O*-(*tert*-butyldimethylsilyl)-N^6,N^6-benzoylmethyl adenosine[8] (**58**)



To a solution of 3',5'-*O*-Bis(*tert*-butylsilyl)-2'-*O*-(*tert*-butyldimethylsilyl)-N^6-benzoyladenosine (**57**, 1g, 1.6mmol) in 5 ml CH_2Cl_2, tetra-butylammonium bromide (0.9 g, 2.4 mmol) and 1N aq. NaOH (excess) were added. After 20 min, methyl iodide (0.4 mL, 6.4 mmol) was added to the reaction mixture and further stirred for 24 hour. Excess CH_2Cl_2 was added to the residue and washed with water (3x10 mL). The organic phase was dried ($MgSO_4$) and concentrated under vacuum. The residue was purified by column chromatography (4:1 to 3:2 cyclohexane / ethyl acetate) which resulted in an oil (663 mg, 65%). TLC R_f 0.5 (7:3 C_6H_{12} /EtOAc); ^{1}H NMR (400 MHz, $CDCl_3$) δ 0.00 (s, 3 H) 0.02 (s, 3 H) 0.78 (s, 9 H) 0.90 (s, 9 H) 0.94 (s, 9 H) 3.05 (s, 3 H) 3.74 – 3.85 (m, 1H), 3.94 – 4.09 (m, 1H), 4.14 (dd, *J* = 9.0, 5.0 Hz, 1H), 4.27 (d, *J* = 5.0 Hz, 1H), 4.32 (dd, *J* = 9.0, 5.0 Hz, 1H), 5.70 (b.s, 1H), 7.23 – 7.31 (m, 2H), 7.31 – 7.43 (m, 1H), 7.55 (s, 1H), 7.82 (d, *J* = 2.5 Hz, 1H), 8.04 (s, 1H), 8.06 (s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$) δ -5.0, -4.3, 18.3, 20.4, 21.1, 25.9, 27.0, 27.5, 35.8, 67.7, 74.7, 75.5, 75.8, 92.4, 126.7, 127.9, 128.7, 130.7, 136.0, 141.3,

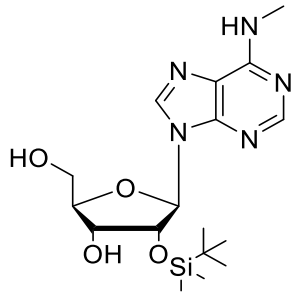
151.7, 152.0, 155.0, 172.2; HRMS (ESI) m/z: Calculated for $C_{32}H_{50}O_5N_5{}^{28}Si_2$ $[M+H]^+$ 640.3345, Observed: 640.3331.

3',5'-*O*-Bis(*tert*-butyl)silyl-2'-*O*-(*tert*-butyldimethyl)silyl-N^6-methyladenosine[9] (60)



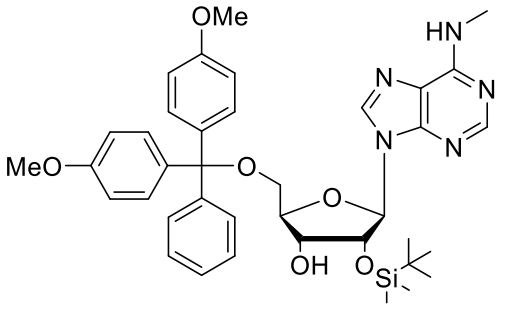
To a stirred solution of the 3',5'-*O*-Bis(*tert*-butylsilyl)-2'-*O*-(*tert*-butyldimethylsilyl)-N^6,N^6-benzoylmethyladenosine (**58**; 663 mg, 1.1 mmol) in THF (5.2 ml) at 0°C lithium triethylborohydride (1 M in THF, 4.15 mL, 4.15 mmol) was added and stirred for 1 hour at room temperature. The reaction was quenched by addition of saturated aqueous NH_4Cl followed by 1N NaOH. The crude product mixture was concentrated under reduced pressure and then diluted with ethyl acetate and was washed with water (3x10 mL). The organic layer was dried ($MgSO_4$) and concentrated under vacuum. The residue was purified by column chromatography (9:1 to 3:2 cyclohexane / ethyl acetate) which resulted in oil (240 mg, 43%). TLC R_f 0.20 (7:3 cyclohexane/ethyl acetate); ^{1}H NMR (400 MHz, $CDCl_3$) δ 0.00 (s, 3 H) 0.02 (s, 3 H) 0.78 (s, 9 H) 0.90 (s, 9 H) 0.94 (s, 9 H) 3.05 (d, *J*=1.0, 3 H) 3.86-3.90 (m, 1 H) 4.02 - 4.10 (m, 1 H) 4.34 (dd, *J*=9.0, 5.0, 1 H) 4.38 - 4.44 (m, 1 H) 4.47 (d, *J*=4.5, 1 H) 5.76 (b.s, 2 H) 7.62 (s, 1 H) 8.22 (s, 1 H); ^{13}C NMR (101 MHz, $CDCl_3$) δ -5.0, -4.3, 18.3, 20.4, 22.8, 25.9, 27.1, 27.5, 27.6, 67.9, 74.6, 75.5, 75.8, 92.4, 127.6, 128.5, 138.0, 153.4, 155.5; HRMS (ESI) m/z: Calculated for $C_{25}H_{46}O_4N_5{}^{28}Si_2$ $[M+H]^+$ 536.3082, Observed: 536.3078.

2'-*O*-(*tert*-Butyldimethyl)silyl-N^6-methyladenosine[9] (61)



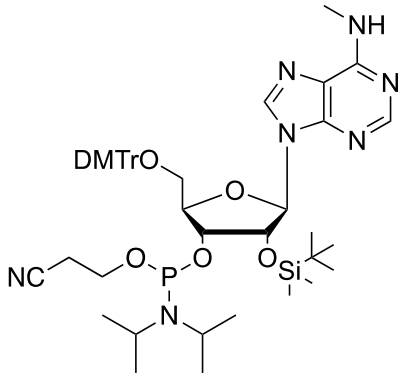
To a stirred solution of 3',5'-*O*-Bis(*tert*-butylsilyl)-2'-*O*-(*tert*-butyldimethylsilyl)-N^6,N^6-benzoylmethyladenosine (**60**; 240 mg, 0.5 mmol) in 4 mL of CH_2Cl_2 at -15°C was added cooled solution of (HF)x**.**py (0.06 mL, 2.3 mmol) diluted with 365 µL pyridine. The reaction temperature was maintained at -15°C and allowed to stir for 12 hours. The reaction was diluted with CH_2Cl_2 and was washed first with sat. aq.$NaHCO_3$ solution then with water (3x10 mL). The organic layer was dried (anhydrous $MgSO_4$) and concentrated under reduced pressure. The residue was purified by column chromatography (9:1 to 3:2 cyclohexane/ethyl acetate) which resulted in oil (93 mg, 53%). TLC R_f 0.15 (2:3 hexane/ethyl acetate); ^{1}H NMR (400 MHz, $CDCl_3$) δ 0.00 (s, 3H), 0.02 (s, 3H), 0.78 (s, 9H), 0.90 (s, 9H), 0.94 (s, 9H), 3.42 (d, *J*=1.0, 3 H), 3.89 (dd, *J* = 10.5, 9.0 Hz, 1H), 4.01 – 4.11 (m, 1H), 4.34 (dd, *J* = 9.0, 5.0 Hz, 1H), 4.41 (dd, *J* = 9.0, 5.0 Hz, 1H), 4.47 (d, *J* = 5.0 Hz, 1H), 5.76 (s, 1H), 7.62 (s, 1H), 8.22 (s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$) δ -5.4, -5.3, 17.92, 25.6, 27.5, 63.5, 73.1, 74.4, 87.8, 91.3, 119.7, 127.6, 140.1, 152.9, 156.0; HRMS (ESI) m/z: Calculated for $C_{17}H_{30}O_4N_5{}^{28}Si$ $[M+H]^+$ 396.2062, Observed: 396.2068.

5'-*O*-(4,4'-Dimethoxytrityl)-2'-*O*-dimethyl(*tert*-butyl)silyl-*N*6-methyladenosine[9] (**62**)



To a stirred solution of 2'-*O*-dimethyl(*tert*-butyl)silyl-*N*6-methyladenosine (**61**) (2.6 g, 6.4 mmol) in 4 mL anhydrous pyridine at 0°C, DMTrCl (2.5 g, 7.6 mmol) was added portion wise at regular intervals for 12 hour. The reaction was quenched by addition of an excess of anhydrous methanol (0.5 mL) at room temperature. After 1hour, the solution was concentrated under vacuum. The crude solid was first fractioned between aqueous $NaHCO_3$ and ethyl acetate, and then the organic layer was washed with water (3x10 mL). The organic layer dried $MgSO_4$ and concentrated under vacuum. The residue was purified by column chromatography (9:1 to 3:2 cyclohexane /ethyl acetate) which resulted in green coloured oil (3 g, 68%). TLC R_f 0.45 (2:3 cyclohexane /ethyl acetate); ^{1}H NMR (400 MHz, $CDCl_3$) δ -0.13 (s, 3 H) 0.00 (s, 3 H) 0.86 (s, 9 H) 2.77 (d, *J*=4.0, 1 H) 3.17 (s, 3 H) 3.36 - 3.43 (m, 1 H) 3.54 (dd, *J*=10.5, 3.5, 1 H) 3.80 (s, 6 H) 4.27 (d, *J*=3.5, 1 H) 4.33 - 4.37 (m, 1 H) 5.02 (t, *J*=5.5, 1 H) 5.85 (d, *J*=4.5, 1 H) 6.04 (b.s, 2 H) 6.83 (d, *J*=9.0, 4 H) 7.18 - 7.28 (m, 3 H) 7.36 (d, *J*=8.0, 4 H) 7.47 (dd, *J*=8.5, 1.3, 2 H) 7.98 (s, 1 H) 8.35 (s, 1 H); ^{13}C NMR (100 MHz, $CDCl_3$) δ -6.3, -6.0, 16.9, 24.6, 27.5, 54.2, 62.4, 70.5, 74.5, 83.0, 85.5, 112.1, 125.9, 126.9, 127.1, 129.1, 134.6, 137.4, 143.6, 152.2, 154.4, 157.5; HRMS (ESI) m/z: Calculated for $C_{38}H_{48}O_6N_5{}^{28}Si$ $[M+H]^+$ 698.3368; Observed: 698.3359.

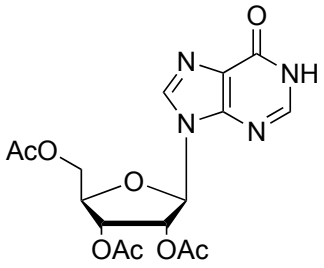
5'-*O*-(4,4'-Dimethoxytrityl)-(3'-*O*-[(2cyanoethyl)(*N*,*N*-diisopropylamino)phosphino]-2'-*O*-dimethyl(*tert*-butyl)silyl-*N*6-methyladenosine[9] (**63**)



To a stirred solution of 5'-*O*-(4,4'-Dimethoxytrityl)-2'-*O*-dimethyl(*tert*-butyl)silyl-*N*6-methyladenosine (**62,** 500 mg, 0.7 mmol) in anhydrous CH_2Cl_2 in an over-dried flask under argon was added DIPEA (1.25 mL, 7.2 mmol) dropwise and the reaction mixture was allowed to stir at 0°C for 10 min. (2-cyanoethyl)-*N,N*-diisopropylchlorophosphoramidite reagent (0.40 mL, 1.8 mmol) was added to the reaction mixture dropwise at 0°C under under argon atmosphere. Reaction was stirred at 0°C for 30 minute and then gradually (about 30 min) warmed up to room temperature. After another five hours under inert atmosphere, the reaction mixture was passed through saturated aq. KCl solution and was evaporated by rotary evaporation. The desired product was separated by silica gel column chromatography (1:1:0.01 hexane:ethyl acetate: pyridine) which resulted in colourless oil (400 mg, 62%) yield. TLC R_f 0.40 (1:1:0.01 hexane:ethyl acetate: pyridine); ^{1}H NMR (700 MHz, CD_2Cl_2-*d*$_2$) δ -0.15 (s, 3H), -0.01 (s, 3H), 0.82 (s, 9H), 1.10 (s, 3H), 1.11 (s, 3H), 1.22 (s, 3H), 1.22 (s, 3H), 1.65 (s, 2H), 2.62 – 2.74 (m, 2H), 3.19 (s, 3H), 3.36 (dd, *J* = 10.5, 4.5 Hz, 1H), 3.54 (dd, *J* = 10.5, 4.0 Hz, 1H), 3.82 (s, 6H), 3.85-3.93 (m, 1H), 3.95-4.10 (m, 1H), 4.41 – 4.49 (m, 1H), 5.12 (dd, *J* = 6.1, 4.4 Hz, 1H), 5.33 – 5.40 (m, 2H), 5.79 (s, 1H), 5.99 (d, *J* = 6.0 Hz, 1H), 6.78 – 6.90 (m, 4H), 7.23 – 7.29 (m, 1H), 7.28 – 7.34 (m, 2H), 7.34 – 7.40 (m, 4H), 7.47 – 7.52 (m, 2H), 7.94 (s, 1H), 8.25 (s, 1H); ^{13}C NMR (176 MHz, CD_2Cl_2) δ -5.4, -5.0, 0.8, 17.8, 20.4, 24.4, 24.4, 25.4, 42.9, 43.0, 55.2, 55.2, 58.9, 63.5, 72.8, 72.9, 74.7, 74.7, 83.5,

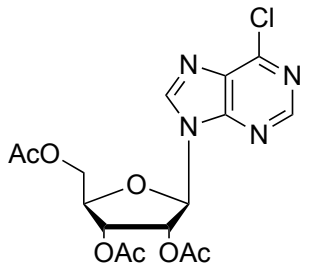
86.5, 88.4, 113.1, 126.8, 127.8, 128.1, 129.0, 130.1, 130.1, 135.7, 139.0, 144.9, 153.0, 155.5, 158.7; ^{31}P NMR (202 MHz, CD_2Cl_2) δ 148.0, 150.8.

2'3'5'-*O*- Triacetylinosine[10] **(66)**



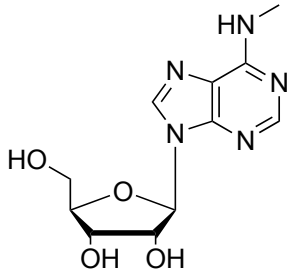
To a stirred solution of inosine (**65**; 10 g, 37.2 mmol) and anhydrous pyridine (10 mL) in 24 mL anhydrous DMF, at room temperature, was added acetic anhydride (20 mL, 212 mmol) dropwise and the reaction mixture was heated to reflux at 75°C for 30 min. The resultant mixture was concentrated under reduced pressure and was diluted with *iso*-propanol. The precipitate was collected through suction filtration and was dried under *vacuo* which resulted in off-white solid (12.3 g, 89%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 2.03 (s, 3H), 2.05 (s, 3H), 2.12 (s, 3H), 4.20 – 4.31 (m, 1H), 4.34 – 4.44 (m, 2H), 5.56 (dd, J = 6.0, 5.0 Hz, 1H), 5.91 (t, J = 5.5 Hz, 1H), 6.20 (d, J = 5.5 Hz, 1H), 8.11 (s, 1H), 8.32 (s, 1H), 12.48 (s, 1H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 20.6, 20.8, 21.0, 63.2, 70.4, 72.7, 80.0, 86.1, 125.3, 139.8, 146.8, 148.4, 156.9, 169.7, 169.9, 170.5; HRMS (ESI) m/z: Calculated for $C_{16}H_{19}O_8N_4$ $[M+H]^+$ 395.1197, Observed: 395.1188.

2'3'5'-*O*- Triacetyl-6-chloropurine riboside[11] **(67)**



To a stirred solution of *Vilsmeier* reagent (7.7 g, 60.9 mmol) in anhydrous DCM in a 50 mL round bottom flask was added 2'3'5'-*O*- triacetylinosine (**66**; 12 g, 30.4 mmol) portion wise, at 0°C, under N_2. The yellow suspension was refluxed at 40°C for 24 hours. The resultant mixture then slowly poured onto saturated aq. The DCM layer was washed with water (3x10 mL) then brine. The organic layer was dried over $MgSO_4$ and concentrated in *vacuo*. The mixture was purified (99:1 to 40:1 DCM/ MeOH) by column chromatography which resulted in yellowish oil (12 g, 96%). TLC R_f 0.75 (40:1 DCM/MeOH); ^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ 2.02 (s, 3H), 2.05 (s, 3H), 2.13 (s, 3H), 4.26 – 4.33 (m, 1H), 4.40 – 4.47 (m, 2H), 5.67 (t, *J* = 5.5 Hz, 1H), 6.04 (t, *J* = 5.0, 1H), 6.38 (d, *J* = 5.0 Hz, 1H), 8.84 (s, 1H), 8.90 (s, 1H); ^{13}C NMR (100 MHz, DMSO-*d*$_6$) δ 20.6, 20.8, 20.9, 63.1, 70.4, 72.6, 80.2, 86.8, 132.1, 146.8, 150.2, 151.7, 152.4, 169.7, 169.9, 170.5. HRMS (ESI$^+$) m/z: Calculated for $C_{16}H_{18}O_7N_4{}^{37}Cl$ $[M+H]^+$ 413.0859, Observed: 413.0869.

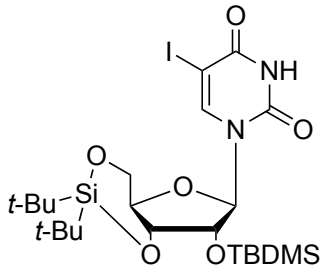
***N*^{6}-Methyladenosine**[12] **(53)**



To a stirred solution of 2'3'5'-*O*- triacetyl-6-chloroinosine (12 g, 28.98 mmol) in 200 ml of ethanol in 500 mL round bottom flask was added methylamine solution (33% in ethanol; 17.2

mL, 196 mmol) slowly and allowed to stir at room temperature under N_2 for 18 hrs. The mixture was concentrated under reduced pressure and the resulted solid was recrystallised twice from *iso*-propanol. The solid was dried under *vacuo* and resulted in white solid (6g, 73%). mp 214.5°C; ^{1}H NMR (500 MHz, D_2O) δ 2.93 (s, 3H), 3.74 (dd, *J* = 13.0, 3.5 Hz, 1H), 3.83 (dd, *J* = 13.0, 2.5 Hz, 1H), 4.20 (q, *J* = 3.0 Hz, 1H), 4.32 (dd, *J* = 5.0, 3.0 Hz, 1H), 4.66 (t, *J* = 5.5 Hz, 1H), 5.90 (d, *J* = 6.0 Hz, 1H), 8.02 (s, 1H), 8.11 (s, 1H); ^{13}C NMR (126 MHz, D_2O) δ 16.8, 61.5, 70.6, 73.7, 85.8, 88.3, 119.2, 139.7, 145.3, 152.3, 154.9. HRMS (ESI) m/z: Calculated for $C_{11}H_{16}O_4N_5$ $[M+H]^+$ 282.1197; Observed: 282.1196.

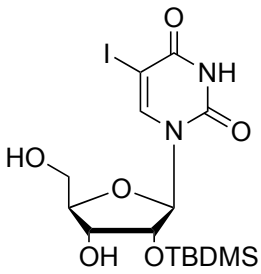
3'5'-*O*-Bis(*tert*-butyl)silyl-2'-*O*-dimethyl(*tert*-butyl)silyl-5-iodouridine[12] (70)



To a stirred suspension of 5-iodouridine (**69**, 1 g, 2.7 mmol) in 40ml anhydrous DMF at 0 °C, di-*t*-butylsilyl ditriflate (1 mL, 3 mmol) was added dropwise under N_2 atmosphere. After consumption of starting material (TLC, 30 min), the reaction was quenched immediately with imidazole (0.9 g, 13.5 mmol) at 0°C. After 5 minutes, the reaction was warmed to room temperature. *tert*-Butyldimethylsilyl chloride (0.5 g, 3.2 mmol) was added portion wise; the reaction was then refluxed at 60°C for 12 hour. The suspension was cooled down to room temperature, water was added and the precipitate was collected by suction filtration. The filtrate was discarded, and the white precipitate was washed with cold methanol. The methanol layer was evaporated under vacuum and the product was crystallised from CH_2C1_2 which resulted in a white solid (1.3 g, 74 %). ^{1}H NMR (600 MHz, DMSO-d_6) δ 0.00 (s, 3H), 0.03 (s, 3H), 0.80 (s, 9H), 0.90 (s, 9H), 0.96 (s, 9H), 3.79 (td, *J* = 10.0, 5.0 Hz, 1H), 3.83 –

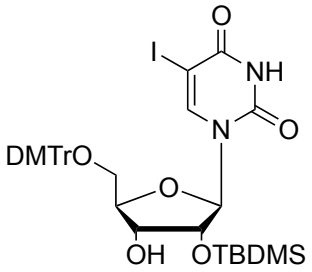
3.90 (m, 1H), 4.04 (dd, *J* = 9.5, 5.0 Hz, 1H), 4.31 (dd, *J* = 9.0, 5.0 Hz, 1H), 4.39 (d, *J* = 5.0 Hz, 1H), 5.51 (b.s, 1H), 7.84 (s, 1H), 11.69 (s, 1H); ^{13}C NMR (151 MHz, DMSO d_6) δ -4.7, -4.0, 18.5, 20.4, 22.7, 26.2, 27.3, 27.8, 67.4, 70.5, 74.4, 74.42, 75.4, 94.2, 145.7, 150.3, 161.0; HRMS (ESI) m/z: Calculated for $C_{23}H_{40}O_6N_2{}^{127}I^{28}Si_2$ $[M+H]^+$ 623.1475, Observed: 623.1471.

2'-*O*-Dimethyl(*tert*-butyl)silyl-5-iodouridine[12] **(71)**



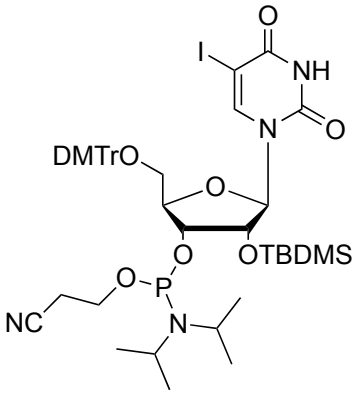
To a stirred solution of 3',5'-*O*-Bis(*tert*-butylsilyl)-2'-*O*-(*tert*-butyldimethyl)silyl-5-iodouridine (**70**; 1 g, 1.63 mmol) in 4mL of CH_2Cl_2 at -15°C was added cooled solution of (HF)x•py (0.2 mL, 8.2 mmol) diluted with 1.2 mL pyridine. The reaction temperature was maintained at -15°C and allowed to stir for 5hrs. The reaction was diluted with CH_2Cl_2 and was washed first with sat. aq. $NaHCO_3$ solution then with water (3x10ml). The organic layer was dried (anhydrous $MgSO_4$) and concentrated under reduced pressure. The residue was purified by column chromatography (9:1 to 3:2 cyclohexane/ethyl acetate) which resulted in oil (680 mg, 63%). TLC R_f 0.2 (2:1 cyclohexane /ethyl acetate); ^{1}H NMR (600 MHz, $CDCl_3$+DMSO-d_6) δ -0.01 (s, 3H), 0.01 (s, 3H), 0.79 (s, 9H), 3.38 (s, 1H), 3.61 (dd, *J* = 11.5, 3.5 Hz, 1H), 3.71 – 3.88 (m, 1H), 3.92 – 4.00 (m, 1H), 4.05 (t, *J* = 5.0 Hz, 1H), 4.13 (t, *J* = 4.0 Hz, 1H), 4.83 (s, 1H), 5.71 (d, *J* = 4.0 Hz, 1H), 8.31 (s, 1H), 8.52 (s, 1H); ^{13}C NMR (151 MHz, $CDCl_3$) δ -0.2, -0.0, 22.8, 30.5, 65.1, 73.2, 74.4, 81.1, 89.6, 94.3, 141.0, 155.2, 165.5; HRMS (ESI) m/z: Calculated for $C_{15}H_{26}O_6N_2{}^{127}I^{28}Si$ $[M+H]^+$ 483.0454, Observed: 483.0449.

5'-*O*-(4,4'-Dimethoxytrityl)-2'-*O*-dimethyl(*tert*-butyl)silyl-5-iodouridine[12] **(72)**



To a stirred solution of 2'-*O*-dimethyl(*tert*-butyl)silyl-5-iodouridine (**71**; 0.6 g, 1.3 mmol) in 4 mL anhydrous pyridine at 0°C, DMTrCl (0.8 g, 1.5 mmol) was added portion wise at regular intervals for 12 hours. The reaction was quenched by addition of an excess of anhydrous methanol (0.5 mL) at room temperature. After 1hour, the solution was concentrated under vacuum. The crude solid was first fractioned between aqueous $NaHCO_3$ and ethyl acetate, and then the organic layer was washed with water (3x10 mL). The organic layer dried $MgSO_4$ and concentrated under vacuum. The residue was purified by column chromatography (9:1 to 1:1 cyclohexane /ethyl acetate) which resulted in yellowish oil (0.7 g, 81%). TLC R_f 0.5 (2:1 cyclohexane /ethyl acetate); ^{1}H NMR (600 MHz, $CDCl_3$) δ 0.00 (s, 3H), 0.01 (s, 3H), 0.79 (s, 9H), 2.56 (d, *J* = 3.5 Hz, 1H), 3.25 (dd, *J* = 11.0, 2.5 Hz, 1H), 3.34 (dd, *J* = 11.0, 2.0 Hz, 1H), 3.67 (s, 6H), 4.05 (q, *J* = 2.5 Hz, 1H), 4.13 (dt, *J* = 5.0, 3.5 Hz, 1H), 4.37 (t, *J* = 5.5 Hz, 1H), 5.87 (d, *J* = 5.5 Hz, 1H), 6.68 – 6.74 (m, 3H), 7.08 – 7.14 (m, 1H), 7.15 – 7.23 (m, 6H), 7.27 – 7.31 (m, 3H), 8.01 (s, 1H), 8.07 (s, 1H); ^{13}C NMR (151 MHz, $CDCl_3$) δ -5.1, -4.8, 18.0, 25.6, 55.3, 63.2, 68.7, 71.5, 76.3, 84.1, 87.3, 88.1, 113.4, 113.5, 127.2, 128.0, 128.2, 130.1, 135.2, 135.3, 144.2, 144.3, 149.6, 158.77, 158.80, 159.4; HRMS (ESI) m/z: Calculated for $C_{36}H_{44}O_8N_2{}^{127}I^{28}Si$ $[M+H]^+$ 785.1761, Observed: 785.1756.

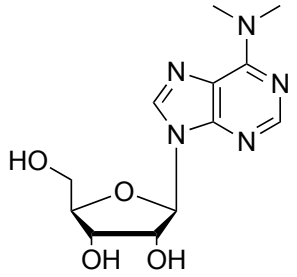
5'-*O*-(4,4'-Dimethoxytrityl)-3'-*O*-[(2-cyanoethoxy)(*N*,*N*-diisopropylamino)phosphino]-2'-*O*-dimethyl(*tert*-butyl)silyl-5-iodouridine[12] (**73**)



To a stirred solution of 5'-*O*-(4,4'-Dimethoxytrityl)-2'-*O*-dimethyl(*tert*-butyl)silyl--5-iodouridine (**72,** 250 mg, 0.32 mmol) in anhydrous CH_2Cl_2 in an over-dried flask under argon was added DIPEA (0.55 mL, 7.17 mmol) dropwise and the reaction mixture was allowed to stir at 0°C for 10 min. (2-cyanoethyl)-*N,N*-diisopropylchlorophosphoramidite reagent (0.2 mL, 0.8mmol) was added to the reaction mixture dropwise at 0°C under under argon atmosphere. Reaction was stirred at 0 °C for 30 minute and then gradually (about 30 minute) warmed up to room temperature. The reaction was monitored by TLC. After another five hours under inert atmosphere, the reaction mixture was passed through saturated aq. KCl solution and was reduced under pressure and further concentrated under *vacuo*. The compound was directly applied for RNA purification.

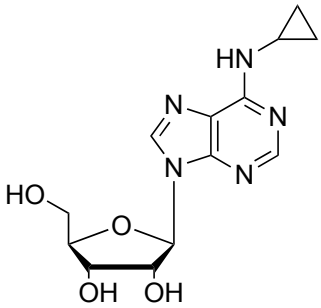
^{31}P NMR (202 MHz, CD_2Cl_2) δ 147.2, 149.2.

***N6,N6*-Dimethyladenosine**[10] (**74**)



To a stirred solution of 2'3'5'-*O*- triacetyl-6-chloroinosine (**67**; 3 g, 7.24 mmol) in 25 mL of ethanol in 100 mL round bottom flask was added dimethylamine solution (33% in ethanol; 25 mL, 50 mmol) slowly and allowed to stir at room temperature under N_2 for 18 hours. The mixture was concentrated under reduced pressure and the resulted solid was recrystallised twice from *iso*-propanol. The solid was dried under *vacuo* and resulted in white solid (1.3 g, 60%). ^{1}H NMR (600 MHz, D_2O) δ 3.06 (s, 6H), 3.75 (dd, *J* = 13.0, 3.5 Hz, 1H), 3.85 (dd, *J* = 13.0, 2.5 Hz, 1H), 4.19 (q, *J* = 3.0 Hz, 1H), 4.29 – 4.33 (t, *J* = 4.5 Hz 1H), 4.59 (t, *J* = 5.5 Hz, 1H), 5.82 (d, *J* = 6.0 Hz, 1H), 7.78 (b.s, 1H), 8.00 (s, 1H); ^{13}C NMR (151 MHz, D_2O) δ 38.7, 61.4, 70.5, 73.7, 85.5, 88.2, 118.9, 138.2, 148.2, 151.3, 153.6; HRMS (ESI) m/z: Calculated for $C_{12}H_{18}O_4N_5$ $[M+H]^+$ 296.1353; Observed: 296.1352.

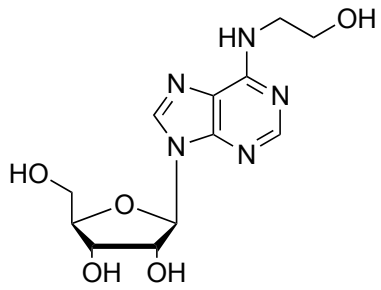
***N6*-Cyclopropyladenosine**[13] (**75**)



To a stirred solution of 2'3'5'-*O*- triacetyl-6-chloroinosine (**67**; 3 g, 7.24 mmol) in 25 mL of ethanol in 100 mL round bottom flask was added cyclopropylamine solution (3.5 mL, 50 mmol) dropwise and allowed to stir at room temperature under N_2 for 18 hours. The mixture

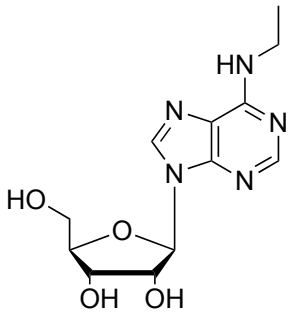
was concentrated under reduced pressure and the resulted solid was recrystallised twice from *iso*-propanol. The solid was dried under *vacuo* and resulted in white solid (1.6 g, 72%). ^{1}H NMR (700 MHz, DMSO-d_6+D_2O) δ 0.62 – 0.71 (m, 2H), 0.88 (dd, *J* = 7.0, 2.0 Hz, 2H), 3.00 (s, 1H), 3.67 (dd, *J* = 12.0, 3.5 Hz, 1H), 3.77 (dd, *J* = 12.0, 3.5 Hz, 1H), 4.21 – 4.28 (m, 1H), 4.67 (t, *J* = 5.5 Hz, 1H), 5.97 (d, *J* = 6.0 Hz, 1H), 8.32 (s, 1H), 8.41 (s, 1H), (1H under solvent); ^{13}C NMR (176 MHz, DMSO-d_6+D_2O) δ 7.0, 61.9, 70.8, 70.9, 73.9, 86.2, 86.3, 88.4, 119.8, 122.7, 140.4, 152.7, 155.9; HRMS (ESI) m/z: Calculated for $C_{13}H_{18}O_4N_5$ $[M+H]^+$ 308.1353, Observed: 308.1351.

N^6-(2-Hydroxyethyl)adenosine[14] (**76**)



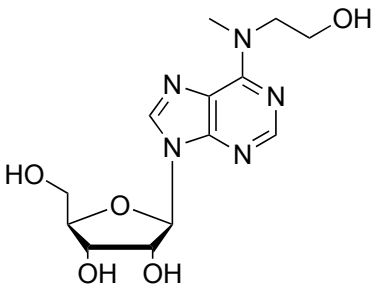
To a stirred solution of 2'3'5'-*O*- triacetyl-6-chloroinosine (**67**; 2.6 g, 6.28 mmol) in 25 ml of ethanol in 100 mL round bottom flask was added cyclopropylamine solution (2.66 mL, 44 mmol) dropwise and allowed to stir at room temperature under N_2 for 18 hrs. The mixture was concentrated under reduced pressure and the resulted solid was recrystallised twice from *iso*-propanol. The solid was dried under *vacuo* and resulted in white solid (1.2 g, 61%). ^{1}H NMR (600 MHz, DMSO-d_6) δ 3.57 (dd, *J* = 12.0, 4.0 Hz, 1H), 3.67 (dd, *J* = 12.0, 3.5 Hz, 1H), 4.13 – 4.21 (m, 1H), 4.51 – 4.64 (m, 3H), 5.96 (d, *J* = 6.0 Hz, 1H), 8.36 (s, 1H), 8.49 (s, 1H), 8.52 (s, 1H) (4 protons under residual water peak); ^{13}C NMR (151 MHz, DMSO-d_6) δ 39.8, 61.6, 63.4, 70.7, 74.0, 86.1, 88.2, 121.5, 142.7, 152.0, 152.2, 160.6; HRMS (ESI) m/z: Calculated for $C_{12}H_{18}O_5N_5$ $[M+H]^+$312.1302, Observed: 312.1297.

N^6-Ethyladenosine[10] **(77)**



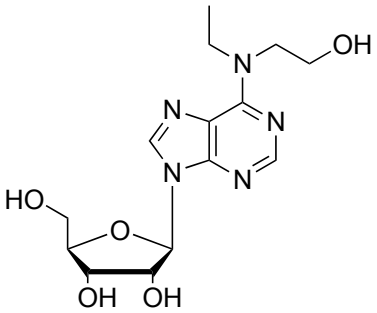
To a stirred solution of ethylamine hydrochloride (1.1g, 13.7mmol) in 10 ml of ethanol in 50 mL round bottom flask was added Ag_2O (3.8g, 16.4 mmol) and the mixture was allowed to stir at room temperature under N_2 for 1 hr. The precipitate is filtered off through suction filtration and 2'3'5'-*O*- triacetyl-6-chloroinosine (**67**; 0.5 g, 1.4 mmol) was added to the filtrate and the reaction mixture was stirred for 24hr. The mixture was concentrated under reduced pressure and the resulted solid was recrystallised twice from *iso*-propanol. The solid was dried under *vacuo* and resulted in yellowish solid (0.12 g, 30%). ^{1}H NMR (600 MHz, D_2O+DMSO-d_6) δ 2.63 (t, J = 2.0 Hz, 3H), 3.53 (b.s, 2H), 3.77 (dd, J = 13.0, 3.5 Hz, 1H), 3.85 (dd, J = 13.0, 3.0 Hz, 1H), 4.23 (q, J = 3.0 Hz, 1H), 4.35 (dd, J = 5.0, 3.0 Hz, 1H), 5.98 (d, J = 6.5 Hz, 1H), 8.18 (s, 1H), 8.24 (s, 1H), (1H under solvent peak); ^{13}C NMR (151 MHz, D_2O) δ 13.9, 38.5, 61.6, 70.8, 73.7, 86.0, 88.3, 117.6, 123.8, 140.1, 152.7, 154.6; HRMS (ESI) m/z: Calculated for $C_{12}H_{18}O_4N_5$ $[M+H]^+$ 296.1308, Observed: 296.1353.

N^6,N^6-Methyl(2-hydroxyethyl)adenosine[15] (**78**)



To a stirred solution of 2'3'5'-*O*- triacetyl-6-chloroinosine (**67**; 1 g, 2.4 mmol) in 25 mL of ethanol in 50 mL round bottom flask was added methylethanolamine solution (1 mL, 12 mmol) dropwise and allowed to stir at room temperature under N_2 for 18 hours. The mixture was concentrated under reduced pressure and the resulted solid was recrystallised twice from *iso*-propanol. The solid was dried under *vacuo* and resulted in white solid (0.56 g, 78%). ^{1}H NMR (700 MHz, DMSO-d_6) δ 3.56 (ddd, J = 12.0, 7.0, 3.5 Hz, 1H), 3.61 – 4.43 (m, 6H), 4.59 (q, J = 6.0 Hz, 1H), 4.75 (t, J = 5.5 Hz, 1H), 5.18 (d, J = 5.0 Hz, 1H), 5.37 (dd, J = 7.0, 4.5 Hz, 1H), 5.45 (d, J = 6.0 Hz, 1H), 5.91 (d, J = 6.0 Hz, 1H), 8.22 (s, 1H), 8.37 (s, 1H) (3 methyl protons and one hydroxyl group under residual water peak in DMSO); ^{13}C NMR (176 MHz, DMSO-d_6) δ 37.2, 52.7, 60.0, 62.0, 71.0, 73.9, 86.2, 88.3, 120.1, 139.2, 150.4, 152.2, 154.6; HRMS (ESI) m/z: Calculated for $C_{13}H_{20}O_5N_5$ $[M+H]^+$ 326.1459, Observed: 326.1459.

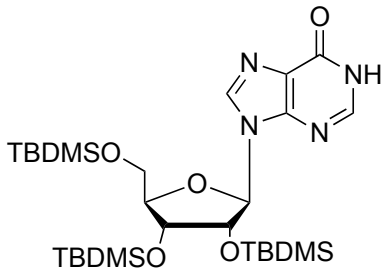
N^6,N^6-Ethyl(2-hydroxyethyl)adenosine[15] (**79**)



To a stirred solution of 2'3'5'-*O*- triacetyl-6-chloroinosine (**67**; 0.15 g, 1.12 mmol) in 25 mL of ethanol in 50 mL round bottom flask was added methylethanolamine solution (0.58 mL, 6

mmol) dropwise and allowed to stir at room temperature under N_2 for 18 hours. The mixture was concentrated under reduced pressure and the resulted solid was recrystallised twice from *iso*-propanol. The solid was dried under *vacuo* and resulted in white solid (0.1 g, 81%). ^{1}H NMR (700 MHz, D_2O) δ 1.18 (t, *J* = 7.0 Hz, 3H), 3.75 (dd, *J* = 13.0, 3.5 Hz, 1H), 3.78 – 4.09 (m, 7H), 4.21 (q, *J* = 3.5 Hz, 1H), 4.34 (dd, *J* = 5.0, 3.5 Hz, 1H), 5.98 (d, *J* = 6.0 Hz, 1H), 8.13 (s, 1H), 8.17 (s, 1H); ^{13}C NMR (176 MHz, D_2O) δ 15.5, 44.6, 47.1, 50.2, 61.5, 70.6, 73.5, 85.8, 88.1, 119.4, 138.7, 149.2, 152.0, 154.2; HRMS (ESI) m/z: Calculated for $C_{14}H_{22}O_5N_5$ $[M+H]^+$ 340.1617, Observed: 340.1615.

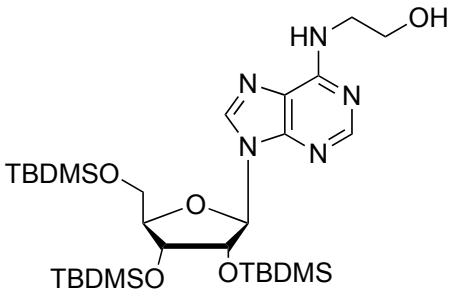
2',3', 5'-*O*-Tris(*tert*-butyldimethyl)silylinosine[16] (**82**)



To a stirred solution of inosine (**65**; 3.75 g, 13.24 mmol) and imidazole (3.6 g, 53.0 mmol) in anhydrous DMF in a 50 mL round bottom flask was added TBDMSCl (6.6 g, 43.7 mmol) portion wise. The reaction was heated at 60°C for 12 hours. The suspension was cooled down to room temperature, water was added and the precipitate was collected by suction filtration. The filtrate was discarded, and the white precipitate was washed with cold methanol. The methanol layer was evaporated under vacuum and the product was crystallised from CH_2C1_2 which resulted in a white solid (7.8 g, 94%). TLC R_f 0.6 (1:9 MeOH/DCM); ^{1}H NMR (400 MHz, $CDCl_3$) δ -0.31 (s, 3H), -0.16 (s, 3H), -0.04 (s, 3H), -0.03 (s, 3H), 0.00 (s, 3H), 0.01 (s, 3H), 0.68 (s, 9H), 0.79 (s, 9H), 0.82 (s, 9H), 3.66 (dd, *J* = 11.5, 2.5 Hz, 1H), 3.86 (dd, *J* = 11.5, 4.0 Hz, 1H), 4.00 (q, *J* = 3.5 Hz, 2H), 4.17 (t, *J* = 4.0 Hz, 1H), 4.38 (t, *J* = 4.5 Hz, 1H), 5.88 (d, *J* = 5.0 Hz, 1H), 7.97 (d, *J* = 5.0 Hz, 1H), 8.09 (s, 1H), 12.83 (s, 1H); ^{13}C NMR (100

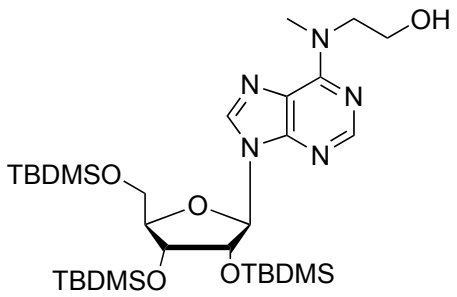
MHz, $CDCl_3$) δ -5.4, -5.0, -4.7, -4.7, -4.4, 17.9, 18.1, 18.6, 25.7, 25.8, 26.1, 62.4, 71.8, 85.5, 88.3, 125.0, 139.1, 144.5, 148.9. 159.2, 159.7; HRMS (ESI) m/z: Calculated for $C_{28}H_{55}O_5N_4{}^{28}Si_3$ $[M+H]^+$ 611.3474; Observed: 611.3468.

2',3',5'-*O*-Tris[dimethyl(*tert*-butyl)silyl]-N^6-(2-hydroxyethyl)adenosine (88)



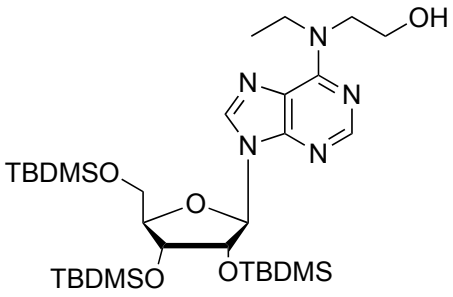
To a stirred solution of 2',3', 5'-*O*-Tris(*tert*-butyldimethyl)silylinosine (**82**; 0.1 g, 0.16 mmol) and PyBOP (0.1 g, 0.2 mmol) in 10 mL of EtOH in a 50 mL round bottom flask was added DIPEA (42 μL, 0.24 mmol) dropwise and the mixture was heated at 40°C. After the consumption of the starting material (40 min, TLC) and the reaction was cooled to room temperature, ethanolamine (0.2 mL, 0.35 mmol) was added dropwise and the reaction was left to stir overnight. The crude product mixture was concentrated under reduced pressure and then diluted with ethyl acetate and was washed with water (3x10 mL). The organic layer was dried ($MgSO_4$) and concentrated under reduced pressure. The residue was purified by column chromatography (99:1 to 94:6 DCM / MeOH) which resulted in oil (85 mg, 80%). ^{1}H NMR (500 MHz, CD_3OD) δ -0.28 (s, 3H), -0.02 (s, 3H), 0.17 (s, 6H), 0.18 (s, 6H), 0.79 (s, 9H), 0.98 (s, 9H), 0.99 (s, 9H), 3.71 – 3.84 (m, 1H), 3.85 (dd, *J* = 11.5, 3.0 Hz, 1H), 4.06 (dd, *J* = 11.0, 4.5 Hz, 1H), 4.15 (dt, *J* = 5.0, 3.0 Hz, 1H), 4.40 (dd, *J* = 4.5, 2.5 Hz, 1H), 4.83 (dd, *J* = 6.0, 4.5 Hz, 1H), 4.89 (s, 5H), 6.07 (d, *J* = 6.0 Hz, 1H), 8.26 (s, 1H), 8.31 (s, 1H); ^{13}C NMR (126 MHz, CD_3OD) δ -6.6, -6.5, -6.3, -5.6, -5.6, -5.5, 17.4, 17.6, 18.0, 24.9, 25.1, 25.2, 45.97, 46.0, 60.4, 62.7, 72.7, 76.0, 86.2, 87.7, 119.5, 139.3, 148.6, 152.5, 155.0; HRMS (ESI) m/z: Calculated for $C_{30}H_{60}O_5N_5{}^{28}Si_3$ $[M+H]^+$ 635.3713; Observed:635.3797.

2',3',5'-*O*-Tris[dimethyl(*tert*-butyl)silyl]-N^6,N^6-methyl(2-hydroxyethyl)adenosine[15] (91)



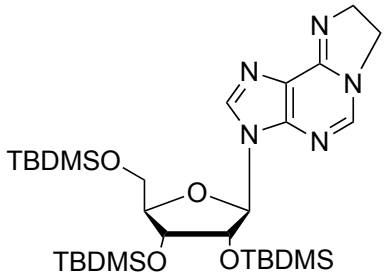
To a stirred solution of 2',3', 5'-*O*-Tris(*tert*-butyldimethyl)silylinosine (**82**; 1 g, 1.64 mmol) and BOP (0.9 g, 1.96 mmol) in 25mL of EtOH in a 50 mL round bottom flask was added DBU (0.3 mL, 1.97 mmol) dropwise and the mixture was heated at 40°C. After the consumption of the starting material (40 min, TLC) and the reaction was cooled to room temperature, methylethanolamine (0.65 mL, 8.2 mmol) was added dropwise and the reaction was left to stir overnight. The crude product mixture was concentrated under reduced pressure and then diluted with ethyl acetate and was washed with water (3x10 mL). The organic layer was dried ($MgSO_4$) and concentrated under reduced pressure. The residue was purified by column chromatography (99:1 to 94:6 DCM / MeOH) which resulted in oil (0.82 g, 80%). TLC R_f 0.6 (6:94 MeOH/DCM). ^{1}H NMR (600 MHz, $CDCl_3$) δ -0.29 (s, 3H), -0.14 (s, 3H), -0.01 (s, 3H), 0.00 (s, 3H), 0.02 (s, 3H), 0.03 (s, 3H), 0.71 (s, 9H), 0.83 (s, 9H), 0.85 (s, 9H), 3.42 (s, 3H), 3.67 (dd, *J* = 11.5, 3.0 Hz, 1H), 3.87 (t, *J* = 5.0 Hz, 2H), 3.92 (dd, *J* = 11.4, 4.0 Hz, 1H), 3.94 – 4.09 (m, 4H), 4.21 (t, *J* = 4.0 Hz, 1H), 4.58 (t, *J* = 4.5 Hz, 1H), 5.92 (d, *J* = 5.0 Hz, 1H), 7.97 (s, 1H), 8.20 (s, 1H); ^{13}C NMR (151 MHz, $CDCl_3$) δ -5.38, -5.36, -5.0, -4.72, -4.70, -4.4, 17.9, 18.1, 18.5, 25.7, 25.9, 26.1, 36.3, 53.8, 61.6, 62.5, 71.9, 75.6, 85.3, 88.3, 120.4, 137.7, 150.5, 152.2, 155.6; HRMS (ESI) m/z: Calculated for $C_{31}H_{62}O_5N_5{}^{28}Si_3$ $[M+H]^+$ 668.4053, Observed: 668.4042.

2',3',5'-*O*-Tris[dimethyl(*tert*-butyl)silyl]-N6,*N*6-ethyl(2-hydroxyethyl)adenosine (93)



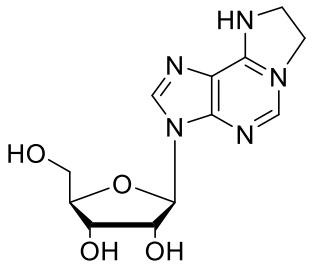
To a stirred solution of 2',3', 5'-*O*-Tris(*tert*-butyldimethyl)silylinosine (**82**; 1 g, 1.64 mmol) and BOP (0.87 g, 1.96 mmol) in 25mL of EtOH in a 50 mL round bottom flask was added DBU(0.3 mL, 1.97 mmol) dropwise and the mixture was heated at 40°C. After the consumption of the starting material (40 min, TLC) and the reaction was cooled to room temperature, ethylethanolamine (0.65 mL, 8.2 mmol) was added dropwise and the reaction was left to stir overnight. The crude product mixture was concentrated under reduced pressure and then diluted with ethyl acetate and was washed with water (3x10 mL). The organic layer was dried ($MgSO_4$) and concentrated under reduced pressure. The residue was purified by column chromatography (99:1 to 94:6 DCM / MeOH) which resulted in oil (0.85 g, 76%). TLC R_f 0.5 (6:94 MeOH/DCM); ^{1}H NMR (600 MHz, $CDCl_3$) δ -0.30 (s, 3H), -0.15 (s, 3H), -0.02 (s, 3H), -0.01 (s, 3H), -0.00 (s, 3H), 0.01 (s, 3H), 0.69 (s, 9H), 0.82 (s, 9H), 0.83 (s, 9H), 1.18 (t, *J* = 7.0 Hz, 3H), 3.66 (dd, *J* = 11.0, 3.0 Hz, 1H), 3.72 – 4.15 (m, 8H), 4.21 (t, *J* = 4.0 Hz, 1H), 4.59 (t, *J* = 5.0 Hz, 1H), 4.67 – 5.17 (m, 1H), 5.89 (d, *J* = 5.0 Hz, 1H), 7.94 (s, 1H), 8.16 (s, 1H); ^{13}C NMR (151 MHz, $CDCl_3$) δ -5.39, -5.38, -5.0, -4.72, -4.70, -4.4, 13.3, 17.9, 18.1, 18.5, 25.7, 25.8, 25.9, 26.1, 44.6, 51.6, 62.5, 63.1, 71.9, 75.5, 85.2, 88.3, 120.1, 137.8, 150.5, 152.2, 155.0. HRMS (ESI) m/z: Calculated for $C_{32}H_{64}O_5N_5{}^{28}Si_3$ $[M+H]^+$ 682.4209, Observed: 682.4201.

2',3',5'-*O*-Tris[dimethyl(*tert*-butyl)silyl]-N^1,N^6-ethanoadenosine[17] (**99**)



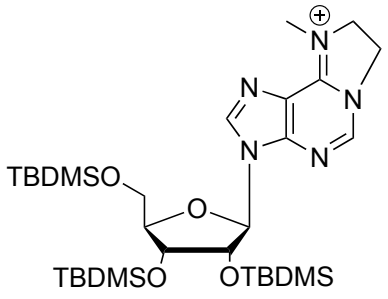
To a stirred solution of 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-*N*1,*N*6-(2-hydroxyethyl)adenosine (**88**; 1.3 g, 2 mmol) and Et_3N (1.4 mL, 10 mmol) in 30 mL of anhydrous DMF in a 50 mL round bottom flask was added methyltriphenoxyphosphonium iodide (1 g, 2.4 mmol) and the mixture was stirred at room temperature for 1 hour. Anhydrous methanol was added and the crude product mixture was concentrated under reduced pressure and then diluted with ethyl acetate and was washed subsequently with $NaHCO_3$ and water (3x10 mL). The organic layer was dried ($MgSO_4$) and concentrated under reduced pressure. The residue was purified by column chromatography (99:1 to 85:15 DCM / MeOH) which resulted in oil (0.75 g, 60%). ^{1}H NMR (600 MHz, $CDCl_3$) δ -0.18 (s, 3H), -0.07 (s, 3H), -0.00 (s, 3H), 0.01 (s, 3H), 0.05 (s, 3H), 0.06 (s, 3H), 0.75 (s, 9H), 0.83 (s, 9H), 0.87 (s, 9H), 3.71 (dd, *J* = 11.5, 2.5 Hz, 1H), 3.93 (dd, *J* = 11.5, 3.0 Hz, 1H), 4.06 (dt, *J* = 5.5, 3.0 Hz, 1H), 4.18 (t, *J* = 4.5 Hz, 1H), 4.26 – 4.40 (m, 3H), 4.90-4.95 (m, 2H), 5.92 (d, *J* = 4.0 Hz, 1H), 8.38 (s, 1H), 8.51 (s, 1H); ^{13}C NMR (151 MHz, $CDCl_3$) δ -5.3, -5.27, -4.8, -4.6, -4.2, 17. 9, 18.0, 18.5, 25.7, 25.8, 26.2, 46.3, 46.4, 62.0, 71.1, 76.7, 85.2, 89.0, 117.6, 142.3, 143.4, 149.0, 151.3; HRMS (ESI) m/z: Calculated for $C_{30}H_{58}O_4N_5{}^{28}Si_3$ $[M+H]^+$ 636.3767, Observed: 636.3791.

***N*^{1},*N*^{6}-Ethanoadenosine**[17] (**100**)



To a stirred solution of 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-*N*^{6}-ethanoadenosine (**99**; 470 mg, 0.74 mmol) in CH_2Cl_2 was added 3HF.Et_3N (361 µL, 2.2 mmol) drop wise. The solution was left to stir for 48 hour. The mixture was reduced under pressure and was purified by column chromatography (99:1 to 9:1 ethyl acetate/methanol) which resulted in a white solid (150 mg, 70%). ^{1}H NMR (700 MHz, DMSO-d_6) δ 3.49 – 3.57 (m, 2H), 3.60 – 3.68 (m, 1H), 3.86 – 3.97 (m, 3H), 4.08 – 4.16 (m, 3H), 4.44 (t, *J* = 5.0 Hz, 1H), 5.11 – 5.21 (m, 2H), 5.46 (s, 1H), 5.78 (d, *J* = 6.0 Hz, 1H), 8.09 (s, 1H), 8.17 (s, 1H); ^{13}C NMR (176 MHz, DMSO-d_6) δ 46.5, 53.3, 61.9, 70.8, 74.5, 86.1, 88.0, 120.0, 138.5, 145.2, 145.4, 150.2; HRMS (ESI) m/z: Calculated for $C_{12}H_{16}O_4N_5$ $[M+H]^+$ 294.1197, Observed: 294.1191.

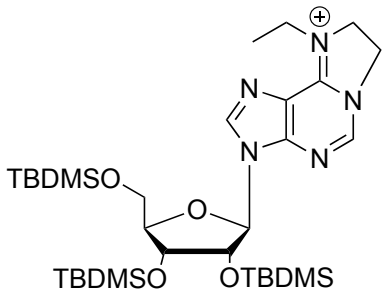
2',3',5'-*O*-Tris[dimethyl(*tert*-butyl)silyl]-*N6,N1,N6*-methylethanoadenosine (**103**)



To a stirred solution of 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-*N*6-(2-hydroxyethyl)adenosine (**91**; 0.5 g, 0.75 mmol) and Et_3N (0.54 mL, 3.75 mmol) in 30 mL of anhydrous DMF in a 50 mL round bottom flask was added methyltriphenoxyphosphonium iodide (0.85 g, 1.9 mmol) and the mixture was stirred at room temperature for 1 hour. Anhydrous methanol was added and the crude product mixture was concentrated under

reduced pressure and then diluted with ethyl acetate and was washed subsequently with $NaHCO_3$ and water (3x10ml). The organic layer was dried ($MgSO_4$) and concentrated under reduced pressure. The residue was purified by alumina column chromatography (99:1 to 90:10 CH_3Cl / MeOH) which resulted in oil. ^1H NMR (600 MHz, $CDCl_3$) δ -0.20 (s, 3H), -0.12 (s, 3H), -0.08 (s, 3H), -0.07 (s, 3H), 0.00 (s, 3H), 0.01 (s, 3H), 0.70 (s, 9H), 0.75 (s, 9H), 0.81 (s, 9H), 3.57 (s, 3H), 3.65 (dd, *J* = 11.5, 2.0 Hz, 1H), 3.88 (dd, *J* = 11.5, 3.0 Hz, 1H), 4.00 (dt, *J* = 5.0, 2.5 Hz, 1H), 4.12 (dd, *J* = 5.5, 4.0 Hz, 1H), 4.17 – 4.26 (m, 2H), 4.35-4.42 (m, 1H), 5.05-5.3 (m, 2H), 5.87 (d, *J* = 3.5 Hz, 1H), 8.43 (s, 1H), 8.50 (s, 1H); ^{13}C NMR (151 MHz, $CDCl_3$) δ -5.4, -5.2, -4.8, -4.6, -4.5, -4.2, 17.9, 18.1, 18.6, 25.7, 25.8, 26.2, 34.6, 45.9, 49.1, 51.6, 61.7, 70.7, 85.0, 89.3, 115.4, 142.8, 143.7, 149.9, 150.1; HRMS (ESI) m/z: Calculated for $C_{31}H_{60}O_4N_5{}^{28}Si_3$ [M] 650.3947, Observed: 650.3924.

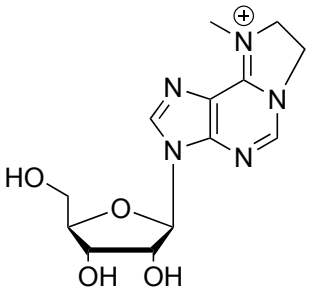
2',3',5'-*O*-Tris[dimethyl(*tert*-butyl)silyl]-*N6,N1,N6*-methylethanoadenosine (104)



To a stirred solution of 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-*N*6-(2-hydroxyethyl)adenosine (**93**; 0.4 g, 0.6 mmol) and Et_3N (0.4 mL, 3 mmol) in 20 mL of anhydrous DMF in a 50 mL round bottom flask was added methyltriphenoxyphosphonium iodide (0.6 g, 1.2 mmol) and the mixture was stirred at room temperature for 1 hour. Anhydrous methanol was added and the crude product mixture was concentrated under reduced pressure and then diluted with ethyl acetate and was washed subsequently with $NaHCO_3$ and water (3x10 mL). The organic layer was dried ($MgSO_4$) and concentrated under reduced pressure. The residue was purified by alumina column chromatography (99:1 to 95: 5

DCM / MeOH) which resulted in solid (0.26 g, 36%). mp 195°C; ^{1}H NMR (600 MHz, $CDCl_3$) δ -0.25 (s, 3H), -0.18 (s, 3H), -0.03 (s, 3H), -0.01 (s, 3H), -0.00 (s, 3H), 0.01 (s, 3H), 0.70 (s, 9H), 0.74 (s, 9H), 0.83 (s, 9H), 1.33 (t, *J*= 7.0, 2H), 3.2 (q, *J*= 7.0, 3H), 3.65 (dd, *J* = 12.0, 2.0 Hz, 1H), 3.88 (dd, *J* = 12.0, 3.0 Hz, 1H), 4.00 (dt, *J* = 5.0, 2.5 Hz, 1H), 4.23 – 4.28 (m, 2H), 4.59 (t, *J* = 5.0 Hz, 1H), 4.69 (t, *J*=5.0, 1H) 5.0-5.18 (m, 2H), 5.83 (d, *J* = 5.0 Hz, 1H), 8.34 (s, 1H), 8.46 (s, 1H); ^{13}C NMR (151 MHz, $CDCl_3$) δ -5.3, -5.0, 5.0, -4.6, -4.6, -4.2, 13.0, 17.9, 18.1, 18.5, 25.7, 25.8, 26.2, 35.6, 48.4, 51.6, 61.7, 70.7, 73.4, 85.0, 89.3, 115.4, 144.8, 149.9, 150.0, 152.0; HRMS (ESI) m/z: Calculated for $C_{31}H_{60}O_4N_5{}^{28}Si_3$ [M] 664.3886, Observed: 664.3896.

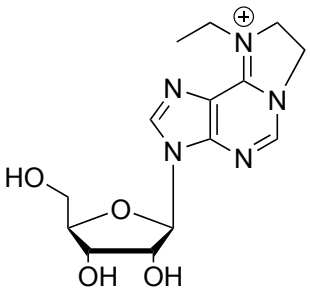
***N*^{6},*N*^{1},*N6*-Methylethanoadenosine (105)**



To a stirred solution of 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-N6,*N6*-methylethano-adenosine (**103**; 500 mg, 0.8 mmol) in CH_2Cl_2 was added 3HF.Et_3N (0.4 mL, 2.3 mmol) drop wise. The solution was left to stir for 48 hour. The mixture was reduced under pressure and was purified by alumina column chromatography (99:1 to 9:1 ethyl acetate/methanol) which resulted in a white solid (54 mg, 22%). mp 202.5°C; ^{1}H NMR (600 MHz, DMSO-d_6) δ 3.01 (s, 3H), 3.29 (m, 1H), 3.38 (dt, *J* = 12.5, 4.5 Hz, 1H), 3.68 (q, *J* = 4.0 Hz, 1H), 3.76-3.84 (m, 2H), 3.87 (b.s, 1H), 4.18 (q, *J* = 5.0 Hz, 1H), 4.43 (t, *J* = 9.5 Hz, 1H), 4.79 (t, *J* = 5.5 Hz, 1H), 5.01 (s, 1H), 5.28 – 5.37 (m, 2H), 5.68 (d, *J* = 5.0 Hz, 1H), 8.50 (s, 1H), 8.52 (s, 1H); ^{13}C NMR (151 MHz, D_2O) δ 36.4, 50.5, 53.4, 63.6, 72.6, 77.1, 88.6, 90.8, 119.0, 145.7, 148.0,

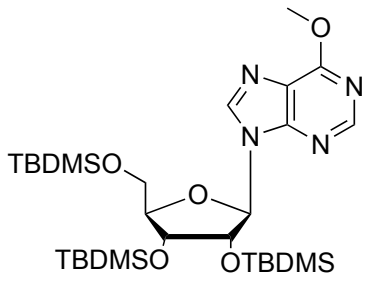
152.2, 152.3; HRMS (ESI) m/z: Calculated for $C_{13}H_{18}O_4N_5$ $[M]^+$ 308.1353, Observed: 308.1353.

***N6,N1,N6*-Ethylethanoadenosine (106)**



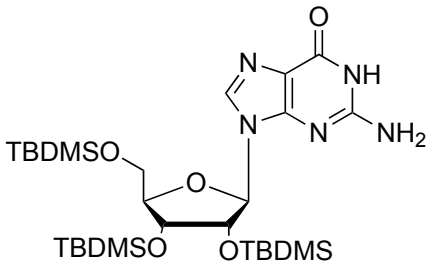
To a stirred solution of 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-N6,*N6*-methylethanoadenosine (**103**; 470 mg, 0.74 mmol) in CH_2Cl_2 was added 3HF.Et_3N (360 µl, 2.2 mmol) drop wise. The solution was left to stir for 48 hours. The mixture was reduced under pressure and was purified by alumina column chromatography (99:1 to 9:1 ethyl acetate/methanol) which resulted in a white solid (65 mg, 30%). mp 210°C; ^{1}H NMR (600 MHz, DMSO-d_6) δ 1.15 (t, J = 7.0 Hz, 3H), 2.88 (q, J = 7.0, 2H), δ 3.49 – 3.57 (m, 2H), 3.36 (dt, J = 12.5, 4.0 Hz, 1H), 3.54 (q, J = 4.0 Hz, 1H), 3.76-3.84 (m, 2H), 3.95 (b.s, 1H), 4.15-4.27 (m, 1H), 4.34-4.43 (m, 1H), 5.1 (b.s, 1H), 5.43 – 5.57 (m, 2H), 6.0 (d, J = 5.0 Hz, 1H), 8.48 (s, 1H), 8.63 (s, 1H); ^{13}C NMR (151 MHz, DMSO-d_6) δ 13.0, 32.4, 48.3, 52.3, 62.4, 70.4, 75.3, 85.4, 90.8, 119.0, 146.4, 148.2, 151.9, 152.0; HRMS (ESI) m/z: Calculated for $C_{13}H_{18}O_4N_5$ $[M]^+$ 322.1518, Observed: 322.1515.

2',3',5'-*O*-Tris[dimethyl(*tert*-butyl)silyl]-*O*6-methylinosine[18] (**107**)



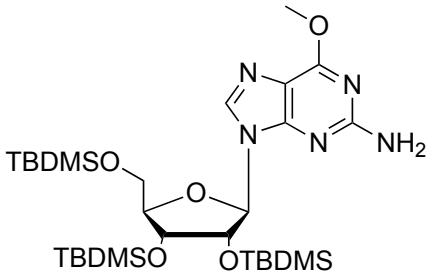
The compound has been synthesised according to reported procedure in literature[7]. To a stirred solution of 2', 3', 5'-*O*-tri(*tert*-butyldimethylsilyl)inosine (1 g, 1.63 mmol) in dry THF was added BOP (1.45 g, 3.28 mmol) and $CsCO_3$ (1.1g, 3.28 mmol). After one hour, the resulting mixture was concentrated under reduced pressure. To this concentrated mixture was added $CsCO_3$ (1.45 g, 3.28 mmol) and MeOH (25ml, excess). After 2 hour, the resulting mixture was reduced under pressure and was diluted with ethyl acetate. The organic layer was washed with water (2x10 mL), brine and dried ($MgSO_4$). The concentrated residue was purified by column chromatography (9:1 to 4:1 cyclohexane / ethyl acetate) which resulted in oil (0.6 g, 73%). TLC R_f 0.8 (9:1 DCM/MeOH); ^{1}H NMR (500 MHz, $CDCl_3$) δ -0.22 (s, 3H), -0.04 (s, 3H), 0.10 (s, 3H), 0.10 (s, 3H), 0.14 (s, 3H), 0.15 (s, 3H), 0.80 (s, 9H), 0.93 (s, 9H), 0.96 (s, 9H), 3.80 (dd, *J* = 11.5, 2.5 Hz, 1H), 4.04 (dd, *J* = 11.5, 4.0 Hz, 1H), 4.14 (q, *J* = 3.5 Hz, 1H), 4.19 (s, 3H), 4.33 (t, *J* = 4.0 Hz, 1H), 4.62 (t, *J* = 4.5 Hz, 1H), 6.09 (d, *J* = 5.0 Hz, 1H), 8.34 (s, 1H), 8.52 (s, 1H); ^{13}C NMR (126 MHz, $CDCl_3$) δ -5.42, -5.38, -5.1, -4.74, -4.70, -4.4, 17.8, 18.1, 18.5, 25.6, 25.8, 26.09, 26. 9, 54.1, 62.4, 71.8, 76.2, 85.4, 88.4, 122.0, 141.2, 151.8, 152.0, 161.0; HRMS (ESI) m/z: Calculated for $C_{29}H_{57}O_5N_5{}^{28}Si_3$ $[M+H]^+$ 625.3631, Observed: 625.3626.

2',3',5'-*O*-Tris[dimethyl(*tert*-butyl)silylguanosine[18] (**109**)



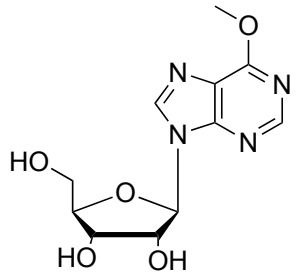
To a stirred solution of guanosine (2 g, 8.3 mmol) and imidazole (3.75g, 13.24 mmol) in dry DMF in a 50 mL round bottom flask was added TBDMSCl (6.58 g, 43.69 mmol) portion wise. The reaction was left to stir overnight. The resulting mixture was concentrated under reduced pressure and was diluted with ethyl acetate. The ethyl acetate layer was washed with water (3x10 mL) and finally with brine. The organic layer was dried over $MgSO_4$ and concentrated in vacuo. Mixture was purified (19:1 to 4:1 cyclohexane/ ethyl acetate) by column chromatography which resulted in off-white solid (7.0g, 84.5%). TLC R_f 0.6 (9:1 DCM /MeOH); mp: 193.5°C; ^{1}H NMR (500 MHz, $CDCl_3$) δ -0.14 (s, 3H), -0.07 (s, 3H), 0.00 (s, 3H), 0.02 (s, 3H), 0.03 (s, 3H), 0.04 (s, 3H), 0.77 (s, 9H), 0.83 (s, 9H), 0.87 (s, 9H), 3.69 (dd, *J* = 11.5, 2.5 Hz, 1H), 3.84 – 3.93 (m, 1H), 4.00 (dt, *J* = 6.0, 2.5 Hz, 1H), 4.18 (b.s, 1H), 4.34 (b.s, 1H), 5.73 (b.s, 1H), 6.11 (b.s, 2H), 7.80 (s, 1H), 11.92 (s, 1H); ^{13}C NMR (126 MHz, $CDCl_3$) δ -5.34, -5.33, -4.8, -4.7, -4.2, 18.0, 18.1, 18.6, 25.8, 25.9, 26.2, 62.1, 71.2, 76.2,84.6, 88.2, 117.5, 136.1, 151.5, 153.5, 159.2; HRMS (ESI) m/z: Calculated for $C_{28}H_{56}O_5N_5{}^{28}Si_3$ [M+H]+ 626.3584, Observed: 626.3576.

2',3',5'-*O*-Tris[dimethyl(*tert*-butyl)silyl-*O*^{6}-guanosine[18] **(111)**



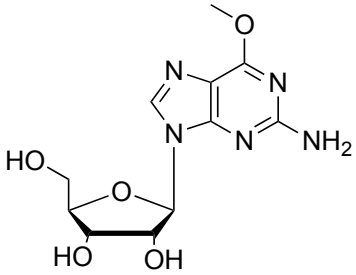
The compound has been synthesised according to reported procedure in literature[7]. To a stirred solution of 2', 3', 5'-*O*-tri(*t*-butyldimethylsilyl)guanosine (0.8 g, 1.3 mmol) in dry THF was added $CsCO_3$ (0.85 g, 2.6 mmol) and BOP (1.2g, 2.7 mmol). After one hour, the resulting mixture was concentrated under reduced pressure. To this concentrated mixture was added $CsCO_3$ (0.85 g, 2.6 mmol) and MeOH (25 mL, excess). After 2 hour, the resulting mixture was reduced under pressure and was diluted with ethyl acetate. The organic layer was washed with water (2x10 mL), brine and dried ($MgSO_4$. The concentrated residue was purified by column chromatography (9:1 to 4:1 cyclohexane / ethyl acetate) which resulted in oil (0.6 g, 73%). TLC R_f 0.8 (9:1 DCM/MeOH); ^{1}H NMR (600 MHz, $CDCl_3$) δ -0.13 (s, 3H), -0.00 (s, 3H), 0.13 (s, 6H), 0.15 (s, 3H), 0.15 (s, 3H), 0.84 (s, 9H), 0.95 (s, 9H), 0.97 (s, 9H), 3.80 (dd, *J* = 11.5, 2.5 Hz, 1H), 3.97 (s, 3H), 4.00 (dd, *J* = 11.5, 4.0 Hz, 1H), 4.12 (q, *J* = 3.5 Hz, 1H), 4.32 (t, *J* = 4.0 Hz, 1H), 4.53 (t, *J* = 4.5 Hz, 1H), 4.95 (s, 2H), 5.93 (d, *J* = 5.0 Hz, 1H), 8.01 (s, 1H); ^{13}C NMR (151 MHz, $CDCl_3$) δ -5.4, -5.0, -4.7, -4.66, -4.3, 17.9, 18.1, 18.6, 25.7, 25.9, 26.1, 53.8, 62.5, 71.9, 76.1, 85.1, 87.9, 115.9, 138.0, 153.6, 159.2, 161.4. HRMS (ESI) m/z: Calculated for $C_{29}H_{58}O_5N_5{}^{28}Si_3$ $[M+H]^+$ 640.3740, Observed: 640.3728.

***O*^{6}-Methylinosine**[18] (**112**)



To a stirred solution of 2', 3', 5'-*O*-tri(*t*-butyldimethylsilyl)-*O*^{6}-methylinosine (0.6 g, 0.94 mmol) in DCM was added 3HF.Et_3N (450 µL, 2.8 mmol) drop wise and the reaction mixture was stirred for 48 hour. The solution was concentrated under vacuo. Mixture was purified (99:1 to 9:1 ethyl acetate/Methanol) by column chromatography which resulted in white solid (0.2g, 72%). TLC R_f 0.3 (9:1 DCM/MeOH); ^{1}H NMR (600 MHz, DMSO-*d*$_6$) δ 3.58 (ddd, *J* = 12.0, 6.0, 4.0 Hz, 1H), 3.69 (dt, *J* = 12.0, 4.5 Hz, 1H), 3.98 (q, *J* = 4.0 Hz, 1H), 4.11 (s, 3H), 4.17 (q, *J* = 4.5 Hz, 1H), 4.60 (q, *J* = 5.5 Hz, 1H), 5.13 (t, *J* = 5.5 Hz, 1H), 5.22 (d, *J* = 5.0 Hz, 1H), 5.50 (d, *J* = 6.0 Hz, 1H), 6.00 (d, *J* = 5.5 Hz, 1H), 8.57 (s, 1H), 8.63 (s, 1H); ^{13}C NMR (151 MHz, DMSO-*d*$_6$) δ 54.5, 61.8, 70.8, 74.2, 86.2, 88.2, 121.6, 142.9, 152.2, 152.24, 160.9; Calculated for $C_{11}H_{13}O_5N_4$ $[M+H]^+$ 283.0934, Observed: 283.0932.

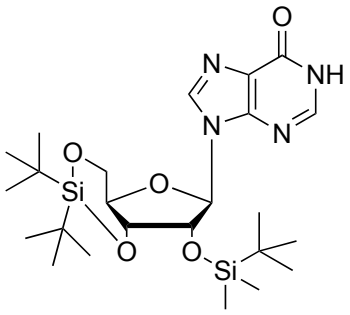
***O*^{6}-Methylguanosine**[18] (**113**)



To a stirred solution of 2', 3', 5'-*O*-tri(*t*-butyldimethylsilyl)-*O*^{6}-methylguanosine (0.6 g, 0.9 mmol) in DCM was added 3HF.Et_3N (450 µL, 2.8 mmol) drop wise and the reaction mixture was stirred for 48 hour. The solution was concentrated under *vacuo*. Mixture was purified (99:1 to 9:1 ethyl acetate/Methanol) by column chromatography which resulted in white solid

(0.24 g, 78%). TLC R_f 0.3 (9:1 Dichloromethane/Methanol); mp 133.7°C; ^{1}H NMR (600 MHz, DMSO-*d*$_6$) δ 3.54 (ddd, *J* = 12.0, 6.0, 4.0 Hz, 1H), 3.64 (ddd, *J* = 12.0, 5.0, 4.0 Hz, 1H), 3.90 (q, *J* = 4.0 Hz, 1H), 3.97 (s, 3H), 4.07 – 4.18 (m, 1H), 4.47 (q, *J* = 6.0 Hz, 1H), 5.08 (t, *J* = 5.5 Hz, 1H), 5.12 (d, *J* = 5.0 Hz, 1H), 5.39 (d, *J* = 6.0 Hz, 1H), 5.79 (d, *J* = 6.0 Hz, 1H), 6.44 (s, 2H), 8.10 (s, 1H); ^{13}C NMR (151 MHz, DMSO-*d*$_6$) δ 53.7, 61.9, 70.9, 73.9, 85.7, 87.0, 114.5, 138.5, 154.6, 160.3, 161.2; HRMS (ESI) m/z: Calculated for $C_{11}H_{16}O_5N_5$ $[M+H]^+$ 298.1146, Observed: 298.1146.

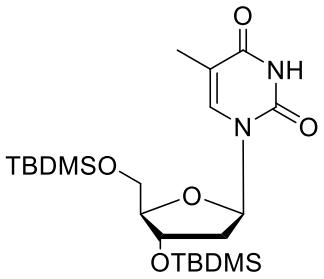
3'5'-*O*-(Di-*tert*-butyl)silyl-2'-*O*-dimethyl(*tert*-butyl)silylinosine[19] (**115**)



To a stirred suspension of inosine (6**5**, 3 g, 11.2 mmol) in 40ml anhydrous DMF at 0 °C, di-*tert*-butylsilyl ditriflate (4 mL, 12.3 mmol) was added dropwise under N_2 atmosphere. After consumption of starting material (TLC, 30min), the reaction was quenched immediately with imidazole (3.8 g, 56 mmol) at 0°C. After 5 minutes, the reaction was warmed to room temperature. *tert*-Butyldimethylsilyl chloride (1.7 g, 24.6 mmol) was added portion wise; the reaction was then refluxed at 60 °C for 12 hr. The suspension was cooled down to room temperature, water was added and the precipitate was collected by suction filtration. The filtrate was discarded, and the white precipitate was washed with cold methanol. The methanol layer was evaporated under vacuum and the product was crystallised from CH_2C1_2 which resulted in a white solid (4.9 g, 84%). ^{1}H NMR (600 MHz, $CDCl_3$) δ 0.17 (s, 3H), 0.18 (s, 3H), 0.96 (s, 9H), 1.07 (s, 9H), 1.10 (s, 9H), 4.02 – 4.09 (m, 1H), 4.25 (td, *J* = 10.0, 5.0 Hz, 1H), 4.38 (dd, *J* = 9.5, 4.5 Hz, 1H), 4.45 – 4.58 (m, 2H), 5.96 (s, 1H), 7.87 (s, 1H), 8.11

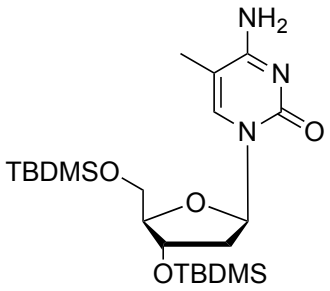
(s, 1H), 12.56 (s, 1H); ^{13}C NMR (151 MHz, $CDCl_3$) δ -5.0, -4.3, 18.3, 20.4, 22.8, 25.9, 27.0, 27.5, 67.8, 74.8, 75.9, 75.9, 92.3, 125.5, 138.3, 144.7, 148.1, 158.9; HRMS (ESI) m/z: Calculated for $C_{24}H_{43}O_4N_5{}^{28}Si_2$ $[M+H]^+$ 523.2767, Observed: 523.2756.

3', 5'-*O*-Bis(*tert*-butyldimethylsilyl)thymidine[20] (**118**)



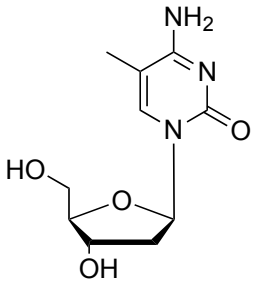
To a stirred solution of thymidine (2 g, 8.3 mmol) and imidazole (2.25 g, 33.03 mmol) in anhydrous DMF in a 50 mL round bottom flask was added TBDMSCl (2.74 g, 18.2 mmol) portion wise. The reaction was left to stir for four hours. The resultant mixture was concentrated under reduced pressure and was diluted with ethyl acetate. The ethyl acetate layer was washed with water (3x10 mL) then brine. The organic layer was dried over $MgSO_4$ and concentrated in *vacuo*. The mixture was purified (19:1 to 4:1 cyclohexane/ ethyl acetate) by column chromatography which resulted in off-white solid (3.3 g, 86%). TLC R_f 0.15 (4:1 cyclohexane/ethyl acetate); mp 150.8°C; ^{1}H NMR (600 MHz, $CDCl_3$) δ 0.10 (s, 3H), 0.11 (s, 3H), 0.14 (s, 3H), 0.14 (s, 3H), 0.92 (s, 9H), 0.96 (s, 9H), 1.94 (d, *J* = 1.0 Hz, 3H), 2.03 (ddd, *J* = 13.0, 8.0, 6.0 Hz, 1H), 2.27 (ddd, *J* = 13.0, 6.0, 2.5 Hz, 1H), 3.79 (dd, *J* = 11.5, 2.5 Hz, 1H), 3.89 (dd, *J* = 11.5, 2.5 Hz, 1H), 3.96 (q, *J* = 2.5 Hz, 1H), 4.43 (dt, *J* = 6.0, 2.5 Hz, 1H), 6.35 (dd, *J* = 8.0, 6.0 Hz, 1H), 7.49 (d, *J* = 1.0 Hz, 1H), 8.15 (s, 1H); ^{13}C NMR (151 MHz, $CDCl_3$) -5.4, -5.35, -4.8, -4.6, 12.5, 18.0, 18.4, 25.8, 25.9, 41.4, 63.0, 72.3, 84.9, 87.9, 110.8, 135.5, 150.1, 163.4; HRMS (ESI) m/z: Calculated for $C_{22}H_{43}O_5N_2{}^{28}Si_2$ $[M+H]^+$ 471.2705, Observed: 471.2697.

3',5'-*O*-Bis[dimethyl(*tert*-butyl)silyl]-5-methylcytidine[20] (**120**)



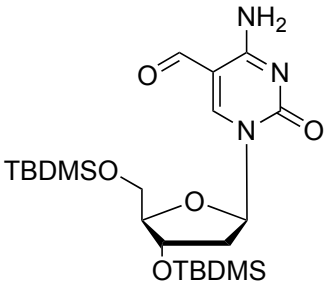
To a stirred solution of 1,2,4-triazole (5.3 g, 0.81 mol, 18eq) and triethylamine (10 mL) in CH_3CN (30 mL) at 0°C, was added phosphorus oxychloride (1.6 mL, 8.5 mmol) dropwise. After stirring for 30 min at 0°C, a solution of 3', 5'-*O*-Bis(*tert*-butyldimethylsilyl)thymidine (**118**; 2 g, 4.2 mmol) in CH_3CN was added dropwise over 30 min and the reaction mixture was allowed to warm to room temperature and left to stir overnight. The reaction mixture was poured into CH_2Cl_2, washed with saturated $NaHCO_3$ and brine, and dried ($NaSO_4$). The solution was concentrated to give the triazolyl intermediate as an orange solid. The orange solid was immediately treated with 1:3 v/v NH_4OH/dioxane (50 mL) and was stirred at room temperature for 4 hrs. The mixture was poured into ether, washed with saturated $NaHCO_3$ and brine, and dried ($MgSO_4$). The organic layer was concentrated under reduced pressure and then purification of the crude by column chromatography (CH_2Cl_2:MeOH 95:5) resulted in a white solid (1.22 g, 60%). ^{1}H NMR (700 MHz, $CDCl_3$) δ 0.06 (s, 3H), 0.07 (s, 3H), 0.11 (s, 3H), 0.11 (s, 3H), 0.89 (s, 9H), 0.92 (s, 9H), 1.91 (s, 3H), 1.96 (dt, *J* = 13.5, 6.5 Hz, 1H), 2.39 (ddd, *J* = 13.5, 6.0, 3.5 Hz, 1H), 3.77 (dd, *J* = 11.5, 2.5 Hz, 1H), 3.88 (dd, *J* = 11.5, 3.0 Hz, 1H), 3.93 (q, *J* = 3.0 Hz, 1H), 4.36 (dt, *J* = 6.5, 3.5 Hz, 1H), 6.32 (t, *J* = 6.5 Hz, 1H), 7.52 – 7.61 (m, 1H); ^{13}C NMR (176 MHz, $CDCl_3$) δ -5.5, -5.4, -4.9, -4.8, 13.4, 18.0, 18.4, 25.8, 25.9, 42.2, 62.8, 71.9, 86.0, 87.7, 101.3, 137.9, 155.9, 165.6; HRMS (ESI) m/z: Calculated for $C_{22}H_{43}O_5N_2{}^{28}Si_2$ $[M+H]^+$ 470.2864, Observed 470.2864.

5-Methylcytidine[21] (**121**)



To a stirred solution of 2',3',5'-*O*-tris[dimethyl(*tert*-butyl)silyl]-N^6,N^6-methylethanoadenosine (**120**; 0.6 g, 1.3 mmol) in CH_2Cl_2 was added 3HF.Et_3N (0.42 mL, 2.6 mmol) drop wise. The solution was left to stir for 48 hours. The mixture was reduced under pressure and was purified by alumina column chromatography (99:1 to 9:1 ethyl acetate/methanol) which resulted in a white solid (100 mg, 32%). ^{1}H NMR (500 MHz, D_2O) δ 1.91 (s, 3H), 2.23 (dt, *J* = 14.0, 6.5 Hz, 1H), 2.34 (ddd, *J* = 14.0, 6.5, 4.0 Hz, 1H), 3.68 (dd, *J* = 12.5, 5.0 Hz, 1H), 3.77 (dd, *J* = 12.5, 3.5 Hz, 1H), 3.96 (q, *J* = 4.0 Hz, 1H), 4.36 (dt, *J* = 6.5, 4.0 Hz, 1H), 6.18 (t, *J* = 6.5 Hz, 1H), 7.68 (s, 1H); ^{13}C NMR (126 MHz, D_2O) δ 12.1, 60.8, 69.6, 70.1, 85.0, 86.9, 103.6, 138.5, 153.1, 162.5; HRMS (ESI) m/z: Calculated for $C_{10}H_{15}O_4N_3$ $[M+H]^+$ 242.1135, Observed: 242.1133.

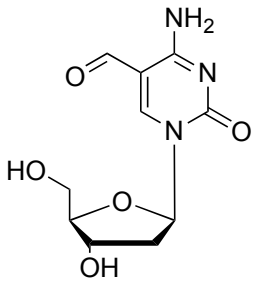
3',5'-*O*-Bis[dimethyl(*tert*-butyl)silyl]-5-formylcytidine[22] (**122**)



To a stirred solution of 3',5'-*O*-Bis[dimethyl(*tert*-butyl)silyl]-5-methylcytidine (**120;** 2.8 g 6 mmol) and 2,6 lutidine (2.4 mL, 20.8 mmol) in CH_3CN:H_2O (1:1), was added a solution containing $K_2S_2O_8$ (3.2 g, 12 mmol) and $CuSO_4.5H_2O$ (0.6 g, 2.4 mmol). The reaction mixture was stirred at 65°C for 3 hours. The reaction mixture was concentrated to half of its

initial volume under reduced pressure and diluted with ethyl acetate and washed with water (3x 50 mL) and subsequently with 10% aq EDTA (pH 4. The combined organic layers were dried over anhydrous Na_2SO_4. The organic layer was concentrated under reduced pressure and the purification by column chromatography on silica gel (CH_2Cl_2:MeOH 95:5) resulted in **122** (5.25 g 85%) as an orange powder. TLC R_f 0.2 (1:1::EtOAc/Cyclohexane). ^{1}H NMR (600 MHz, $CDCl_3$) δ 0.09 (s, 3H), 0.1 (s, 3H), 0.11 (s, 3H), 0.13 (s, 3H), 0.91 (s, 9H), 0.92 (s, 9H), 2.10 (dt, *J* = 13.5, 6.0 Hz, 1H), 2.64 (ddd, *J* = 13.5, 6.0, 4.0 Hz, 1H), 3.67 – 3.92 (m, 1H), 3.98 (dd, *J* = 11.5, 2.5 Hz, 1H), 4.07 (q, *J* = 2.5 Hz, 1H), 4.28 – 4.51 (m, 3H), 6.23 (t, *J* = 6.0 Hz, 1H), 8.57 (s, 1H), 9.51 (s, 1H); ^{13}C NMR (151 MHz, $CDCl_3$) δ -5.3, -5.2, -4.9, -4.5, 18.0, 18.4, 25.7, 26.0, 43.0, 62.6, 71.6, 87.8, 88.7, 105.0, 153.0, 153.3, 162.6, 187.2; HRMS (ESI) m/z: Calculated for $C_{22}H_{41}N_3O_5{}^{28}Si_2$ $[M+H]^+$ 484.2658, Observed: 484.2658.

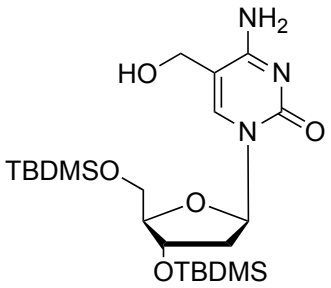
5-Formylcytidine[23] (**123**)



To a stirred solution of 3',5'-*O*-Bis[dimethyl(*tert*-butyl)silyl]-5-formylmethylcytidine (**122**; 0.5 g, 1.03 mmol) in CH_2Cl_2 was added 3HF.Et_3N (0.42 mL, 2.06 mmol) drop wise. The solution was left to stir for 48 hr. The mixture was concentrated under reduced pressure and dissolved in MeOH; the precipitate was collected through suction filtration and dried in *vacuo* which resulted in yellow solid (0.15 mg, 57%). ^{1}H NMR (700 MHz, D_2O) δ 2.30 (dt, *J* = 13.5, 6.0 Hz, 2H), 2.50 (dt, *J* = 13.5, 5.5 Hz, 2H), 3.72 (dd, *J* = 12.5, 5.0 Hz, 1H), 3.84 (dd, *J* = 12.5, 3.5 Hz, 1H), 4.07 (q, *J* = 4.0 Hz, 1H), 4.36 (dt, *J* = 6.5, 4.5 Hz, 1H), 6.13 (t, *J* = 6.0 Hz, 1H), 8.75 (s, 1H), 9.46 (s, 1H); ^{13}C NMR (176 MHz, D_2O) δ 40.1, 60.7, 69.8, 87.2, 87.5,

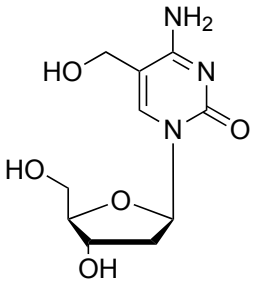
105.9, 154.7, 155.0, 162.5, 190.4; HRMS (ESI) m/z: Calculated for $C_{10}H_{12}N_3O_5$ $[M-H]^-$ 245.0782, Observed: 254.0782.

3',5'-*O*-Bis[dimethyl(*tert*-butyl)silyl]-5-hydroxymethylcytidine[24] **(124)**



To a stirred solution of 3',5'-*O*-Bis[dimethyl(*tert*-butyl)silyl]-5-formylcytidine (**122**, 5.0 g, 3.8 mmol) in MeOH (110 mL) was added $CeCl_3.7H_2O$ (11.55 g, 31 mmol 3.0 eq) under argon, and the mixture was cooled to 0 ^oC with an ice-bath. $NaBH_4$ (392 mg, 10.4 mmol 1.0 equiv) was added portion wise over 15 min and the reaction mixture was warmed to room temperature and stirred for 1 hour. After the consumption of the starting material (TLC), the reaction mixture was concentrated under reduced pressure. The residue was purified by silica gel chromatography (CH_2Cl_2:MeOH 9:1) to give **124** (3.10 g, 62%) as an orange solid. ^1H NMR (600 MHz, $CDCl_3$) δ -0.01 (s, 3H), 0.00 (s, 3H), 0.01 (s, 3H), 0.01 (s, 3H), 0.81 (s, 9H), 0.82 (s, 9H), 2.19 – 2.25 (m, 1H), 2.29 – 2.38 (m, 1H), 3.50 – 3.99 (m, 6H), 4.28 (dq, J = 5.5, 3.0 Hz, 1H), 4.41 (q, J = 13.5 Hz, 2H), 6.10 (t, J = 6.5 Hz, 1H), 7.71 (s, 1H); ^{13}C NMR (151 MHz, $CDCl_3$) δ -5.4, -5.3, -4.8, -4.5, -3.6, 18.0, 18.4, 25.6, 25.8, 25.81, 26.0, 42.1, 58.0, 62.9, 72.0, 86.1, 88.4, 107.1, 140.2, 152.2, 161.7; HRMS (ESI) m/z: Calculated for $C_{22}H_{43}N_3O_5{}^{28}Si_2$ $[M+H]^+$ 486.2814, Observed: 486.2814.

5-Hydroxymethylcytidine[24] (**125**)



To a stirred solution of 3',5'-*O*-Bis[dimethyl(*tert*-butyl)silyl]-5-hydroxymethylcytidine (**124**; 0.2 g, 0.42 mmol) in CH_2Cl_2 was added 3HF.Et_3N (0.15 mL, 0.84 mmol) drop wise. The solution was left to stir for 48 hours. The mixture was concentrated under reduced pressure and dissolved in MeOH; the precipitate was collected through suction filtration and dried in *vacuo* which resulted in yellow solid (0.05 mg, 47%). ^{1}H NMR (500 MHz, D_2O) δ 2.17 – 2.29 (m, 1H), 2.29 – 2.43 (m, 1H), 3.69 (dd, J = 12.5, 5.0 Hz, 1H), 3.78 (dd, J = 12.5, 3.5 Hz, 1H), 3.98 (dt, J = 5.0, 3.5 Hz, 1H), 4.29 – 4.45 (m, 3H), 6.18 (t, J = 6.5 Hz, 1H), 7.82 (s, 1H); ^{13}C NMR (126 MHz, D_2O) δ 39.3, 57.7, 61.0, 70.3, 86.1, 86.6, 106.6, 140.5, 157.0, 164.7; HRMS (ESI) m/z: Calculated for $C_{10}H_{16}O_5N_3$ $[M+H]^+$ 258.1084, Observed: 258.1084.

6.10 Experimental for Chapter 4

6.10.1 ^{1}H NMR Time-Courses

Time course experiments were run according to automated routines. The samples were first prepared in Eppendorf tubes. Immediately after mixing, the samples were transferred to 3 mm Bruker MATCH NMR tubes (160 µL capacity) or 5 mm NMR tubes (400 µL capacity) for NMR analysis. The total lag time between mixing and the first NMR analysis was 5-7 min. For the analysis of formaldehyde-nucleotides adducts formation, the number of scans vary from 32 to 320 (2-20 min). For the first hour, spectra were recorded every 2 min; for the next two hours, spectra were recorded every 4 min; for the rest of the analysis, spectra were recorded every 20 min. For time courses used in enzyme catalysed assays (FTO/AlkB), the

number of scans vary from 64 to 120 (4- 13 min). In some cases, water suppression was achieved by excitation sculpting or noesy pulse program.

6.10.2 1D HSQC NMR experiment

The experiments were carried out using the AV700 instrument. 3 mm MATCH NMR tubes were employed. All experiments were conducted at 310 K. The CLIP-HSQC sequence was used (without ^{13}C decoupling) to detect 13C-HCHO release.[25] The $^{1}J_{CH}$ was set to 145 or 160 Hz. Number of scans vary from 200-400.

6.10.3 EXSY experiments

EXSY experiments utilised 1D pulse sequences using selective refocusing with Gaussian and CHIRP pulses. 1D experiments accumulated 16 scans with a τ_m varying from 10ms -1200s.

6.11 Experimental for Chapter 5

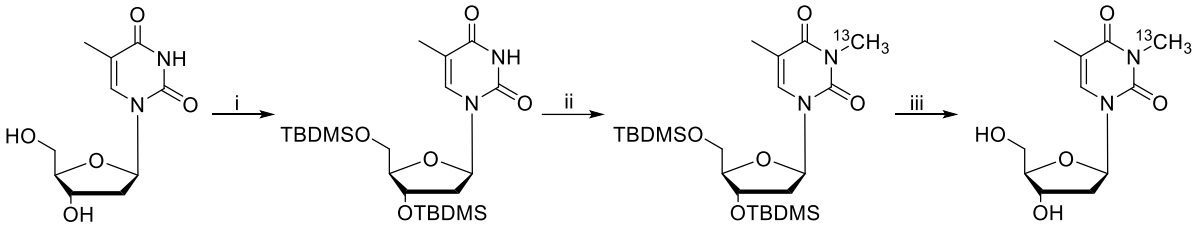
6.11.1 Optimised assay conditions for FTO

For FTO assays, a 160 µL reaction mixture consisted of 400 µM N^6-[^{13}C]-methyladenosine/N^3-[^{13}C]-methylthymidine (from a 1 mM stock in 50 mM AFN buffer in D_2O at pD 7.5), 5 mM of α-ketoglutaric acid sodium salt (from a 50 mM stock solution in 50 mM AFN buffer in D_2O at pD 7.5), 1 mM sodium L-ascorbate (from a 10 mM stock solution in 50 mM AFN buffer in D_2O at pD 7.5), 20 µM diammonium iron(II) sulphate (from a freshly prepared 200 µM stock, from a 100 mM stock solution in 20 mM DCl), and 20 µM FTO. The reagents were added in the specific order indicated above: substrate was added first, followed by 2OG, sodium L-ascorbate, Fe(II), and lastly FTO.

6.11.2 Optimised assay conditions for AlkB

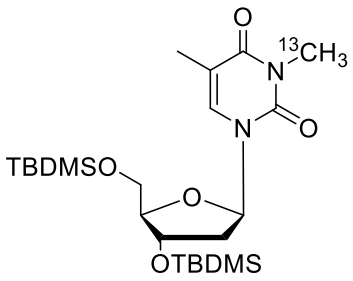
For AlkB assays, a 160 µL reaction mixture consisted of 400 µM of N^1-[^{13}C]-methyladenosine (from a 1 mM stock in 50 mM AFN buffer in D_2O at pD 7.5), 5 mM α-ketoglutaric acid sodium salt (from a 50 mM stock solution in 50 mM AFN buffer in D_2O at pD 7.5), 5 mM sodium L-ascorbate (from a 50 mM stock solution in 50 mM AFN buffer in D_2O at pD 7.5), 40 µM diammonium iron(II) sulphate (from a freshly prepared 200 µM stock, from a 100 mM stock solution in 100 mM DCl), and 10 µM AlkB. The reagents were added in the specific order above: substrate was added first, followed by 2OG, sodium L-ascorbate, Fe(II) and lastly the enzyme, AlkB.

6.11.3 Synthetic Procedures:



i)Imidazole, TBDMSCl, DMF, 4 hr; ii) NaH, DMF, $^{13}CH_3I$; iii) $3HF.Et_3N$, CH_2Cl_2

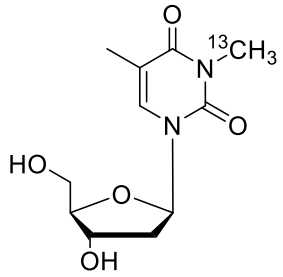
3',5'-*O*-Bis(*t*-butyldimethylsilyl)-N^3-[^{13}C]-2'-deoxymethylthymidine[3]



3',5'-*O*-Bis(*t*-butyldimethylsilyl)-2'-deoxythymidine (0.8 g, 1.7 mmol) was dissolved in anhydrous DMF at 0°C under nitrogen atmosphere. NaH (60% dispersion in mineral oil; 0.07 g, 1.78 mmol) was dissolved in DMF, and the resultant solution was carefully added (due to

H_2 development) to the reaction mixture which was stirred for 1 hr to give a clear solution. [^{13}C]-methyl iodide was added drop wise and the solution was stirred overnight. The solvent was removed *in vacuo*. The concentrated mixture was diluted with ethyl acetate and washed subsequently with water and brine. The organic layer was dried over $MgSO_4$ and concentrated *in vacuo*. The mixture was purified (19:1 to 4:1 cyclohexane/ethyl acetate) by column chromatography which resulted in a white solid (0.2 g, 24 %). TLC R_f 0.2 (4:1 cyclohexane/ethyl acetate); ^{1}H NMR (600 MHz, $CDCl_3$) δ 0.10 (s, 3H), 0.11 (s, 3H), 0.13 (s, 3H), 0.14 (s, 3H), 0.92 (s, 9H), 0.95 (s, 9H), 1.96 (d, *J* = 1.0 Hz, 3H) 2.02 (ddd, *J* = 13.0, 8.0, 6.0 Hz, 1H), 2.29 (ddd, *J* = 13.0, 6.0, 2.5 Hz, 1H), 3.25, 3.48 (d, *J*=142.0, 3H), 3.79 (dd, *J* = 11.5, 2.5 Hz, 1H), 3.89 (dd, *J* = 11.5, 2.5 Hz, 1H), 3.96 (q, *J* = 2.5 Hz, 1H), 4.38 – 4.45 (m, 1H), 6.35 (dd, *J* = 8.0, 6.0 Hz, 1H), 7.49 (d, *J* = 1.0 Hz, 1H); ^{13}C NMR (151 MHz, $CDCl_3$) δ -5.4, -5.4, -4.8, -4.6, 13.3, 18.0, 18.4, 25.8, 25.9, 27.8, 41.5, 62.9, 72.2, 85.6, 87.8, 109.8, 133.3, 151.1, 163.7; HRMS (ESI$^+$) m/z: Calculated for $C_{10}{}^{13}CH_{17}O_5N_2Na$ $[M+Na]^+$ 508.2714, Observed: 508.2708.

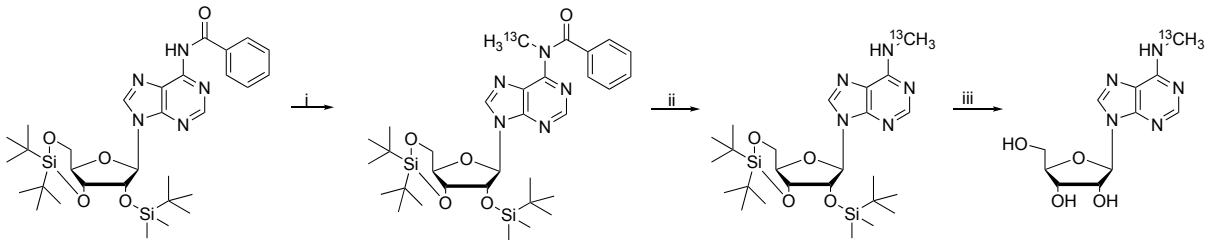
N^3-[^{13}C]-Methylthymidine[25a]



To a stirred solution of 3',5'-*O*-Bis(*t*-butyldimethylsilyl)-N^3-[^{13}C]-2'-deoxymethylthymidine (0.13 g, 0.3 mmol) in CH_2Cl_2 was added 3HF.Et_3N (100 µL, 0.5 mmol) drop wise and the reaction mixture was stirred for 48 h. The solution was concentrated *in vacuo*. The mixture was purified (99:1 to 20:3 ethyl acetate/methanol) by column chromatography which resulted in a white solid (0.2 g, 24%). TLC R_f 0.20 (4:1 ethyl acetate/methanol); ^{1}H NMR (600 MHz,

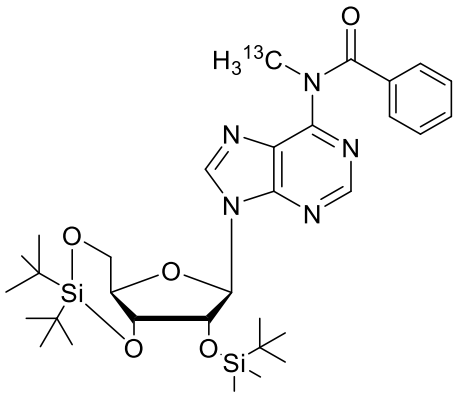
D_2O) 1.87 (s, 3H), 2.26-2.37 (m, 2H), 3.12, 3.36 (d, *J*=142.0, 3H), 3.71 (dd, *J* = 12.5, 3.5 Hz, 1H), 3.79 (dd, *J* = 12.5, 3.5 Hz, 1H), 3.97 (q, *J* = 3.5 Hz, 1H) 4.2-4.46 (m, 1H), 6.27 (dd, *J* = 8.0, 6.0 Hz, 1H), 7.60 (s, 1H); ^{13}C NMR (151 MHz, D_2O) 12.4, 38.2, 27.9, 61.0, 86.6. 70.8, 85.8, 135.4, 110.4, 152.2, 163.7; HRMS (ESI) m/z: Calculated for $C_{10}{}^{13}CH_{17}O_5N_2$ $[M+H]^+$ 258.1166, Observed: 258.1167.

Synthesis scheme for N^6-[^{13}C]-methyladenosine



i)$^{13}CH_3I$, NaOH, CH_2Cl_2, tertabutylammonium Iodide,24 hr ii) $LiEt_3BH$, DMF,1 hr; iii) 3HF. Et_3N, DCM, 48 hrs.

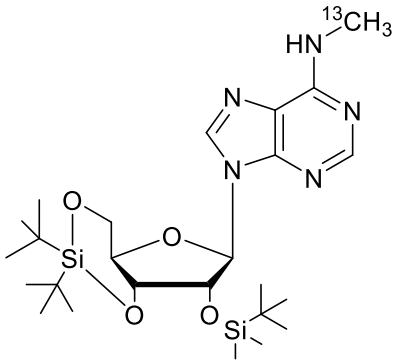
3',5'-*O*-Bis(*t*-butylsilyl)-2'-*O*-(*t*-butyldimethylsilyl)-N^6,N^6-benzoyl-[^{13}C]-methyladenosine[26]



To a solution of 3',5'-*O*-bis(*t*-butylsilyl)-2'-*O*-(*t*-butyldimethylsilyl)-N^6-benzoyladenosine (260 mg, 0.42 mmol) in 5 mL CH_2Cl_2, tetrabutylammonium bromide (0.23 g, 1.6 mmol) and 1 N aq. NaOH (excess) were added. After 20 min, [13C]-methyl iodide (104 µL, 1.7 mmol) was added to the reaction mixture and further stirred for 24 hours. Excess CH_2Cl_2 was added to the residue and washed with water (3 x 10 mL). The organic phase was dried ($MgSO_4$) and

concentrated under vacuum. The residue was purified by column chromatography (4:1 to 3:2 cyclohexane/ethyl acetate) which resulted in an oil (140 mg, 52%). TLC R_f 0.5 (7:3 C_6H_{12}/EtOAc); ^{1}H NMR (400 MHz, $CDCl_3$) δ -0.01 (s, 3H), 0.00 (s, 3H), 0.79 (s, 9H), 0.89 (s, 9H), 0.90 (s, 9H), 3.43, 3.77(d, *J*=140.0, 3H) , 3.74 – 3.85 (m, 1H), 3.94 – 4.09 (m, 1H), 4.14 (dd, *J* = 9.0, 5.0 Hz, 1H), 4.27 (d, *J* = 5.0 Hz, 1H), 4.32 (dd, *J* = 9.0, 5.0 Hz, 1H), 5.70 (b.s, 1H), 7.23 – 7.31 (m, 2H), 7.31 – 7.43 (m, 1H), 7.55 (s, 1H), 7.82 (d, *J* = 2.5 Hz, 1H), 8.04 (s, 1H), 8.06 (s, 1H); ^{13}C NMR (101 MHz, $CDCl_3$) δ -5.0, -4.3, 18.3, 20.4, 22.8, 25.9, 27.0, 27.5, 35.9, 36.9, 67.8, 74.6, 75.8, 75.9, 91.9, 122.7, 128.1, 129.8, 132.0, 135.7, 137.7, 144.7, 146.0, 146.9, 147.6; HRMS (ESI) m/z: Calculated for $C_{31}{}^{13}CH_{50}O_5N_5Si_2$ $[M+H]^+$ 641.3378, Observed: 641.3376.

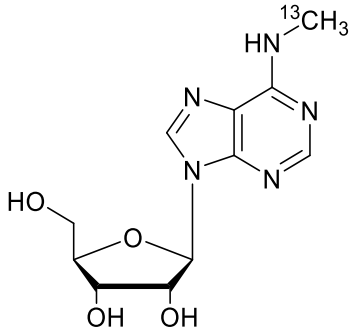
3',5'-*O*-Bis(t-butylsilyl)-2'-*O*-(*t*-butyldimethylsilyl)-N^6- [13C]methyladenosine[26]



To a stirred solution of 3',5'-*O*-Bis(*t*-butylsilyl)-2'-*O*-(*t*-butyldimethylsilyl)-N^6,N^6-benzoyl-[^{13}C]-methyladenosine (140 mg, 0.22 mmol) in THF (2 mL) at 0°C, lithium triethylborohydride (1 M in THF, 1.093 mL, 1.09 mmol) was added and stirred for 1 hour at room temperature. The reaction was quenched by addition of saturated aqueous NH_4Cl followed by 1 N NaOH. The crude product mixture was concentrated *in vacuo*, then diluted with ethyl acetate, then washed with water organic layer was dried ($MgSO_4$) and concentrated under vacuum. The residue was purified by column chromatography (9:1 to 3:2 cyclohexane/ethyl acetate) which resulted in an oil (88 mg, 75%). TLC R_f 0.20 (7:3

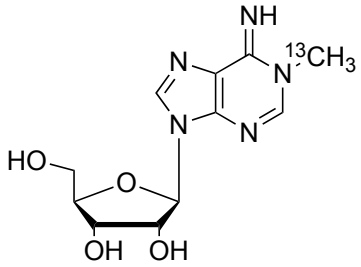
cyclohexane/ethyl acetate); ^{1}H NMR (400 MHz, $CDCl_3$) δ 0.00 (s, 3H), 0.02 (s, 3H), 0.78 (s, 9H), 0.90 (s, 9H), 0.94 (s, 9H), 2.88, 3.23 (d, *J*=140.0, 3H), 3.89 (dd, *J* = 10.5, 9.0 Hz, 1H), 4.01 – 4.11 (m, 1H), 4.34 (dd, *J* = 9.0, 5.0 Hz, 1H), 4.41 (dd, *J* = 9.0, 5.0 Hz, 1H), 4.47 (d, *J* = 5.0 Hz, 1H), 5.76 (b.s, 2H), 7.62 (s, 1H), 8.22 (s, 1H); ^{13}C NMR (101 MHz, $CDCl_3$) δ -4.9, -4.3, 18.3, 20.4, 22.8, 25.9, 27.1, 27.5, 67.9, 74.7, 75.5, 75.8, 92.4, 120.6, 137.9, 153.4, 155.5; HRMS (ESI) m/z: Calculated for $C_{24}{}^{13}CH_{46}O_4N_5Si_2$ $[M+H]^+$ 537.3116, Observed: 537.3117.

N^6- [^{13}C]-Methyladenosine[30]



To a stirred solution of 3',5'-*O*-bis(t-butylsilyl)-2'-*O*-(*t*-butyldimethylsilyl)-N^6-[^{13}C]methyladenosine (100 mg, 0.186 mmol) in CH_2Cl_2 was added $3HF.Et_3N$ (303 μL, 1.86 mmol) drop wise. The solution was left to stir for 48 hours. The mixture was reduced under pressure and was purified by column chromatography (99:1 to 9:1 ethyl acetate/methanol) which resulted in a white solid (40 mg, 76%). R_f 0.4 (9:1 ethyl acetate/methanol); mp 214.5°C; ^{1}H NMR (400 MHz, CD_3OD-d_4) δ 2.95 (s, 2H), 3.22 – 3.49 (m, 5H), 3.77 (dd, *J* = 12.0 Hz, 1H), 3.91 (dd, *J* = 12.0 Hz, 1H), 4.13 – 4.23 (m, 1H), 4.25 – 4.44 (m, 1H), 4.71 – 4.80 (m, 1H), 5.97 (d, *J* = 6.0 Hz, 1H), 8.24 (s, 2H); ^{13}C NMR (100 MHz, CD_3OD) δ 26.3, 62.1, 71.3, 74.1, 86.8, 89.9, 120.2, 140.0, 147.5, 152.1, 155.5; HRMS (ESI) m/z: Calculated for $C_{10}{}^{13}CH_{16}O_4N_5$ $[M+H]^+$ 283.1231, Observed: 283.1232.

N^1- [^{13}C]-Methyladenosine[27]



To stirred suspension of adenosine (2 g, 7.48 mmol) was added 2 mL of [13C]-methyliodide (2 mL, 32 mmol). The reaction was left to stir at 28°C for 18 hours and subsequently filtered through celite. 100 mL of acetone was added to the filtrate. The solution was left to stand at 5°C for 14 hours, then filtered. To the hydroiodide salt of N^1-[13C]-methyladenosine, 10 mL of distilled water was added and the pH was adjusted to pH 8 with ammonia. The solution was then diluted with acetone and the mixture was left to stand for 2 days at 10 °C. The resultant crystals were collected by filtration (0.98 g, 46%). ^{1}H NMR (700 MHz, D_2O) δ 3.68, 3.89 (s, 3H), 3.70 (dd, J = 12.5, 4.5 Hz, 1H), 3.75 – 3.80 (m, 2H), 4.11 – 4.15 (m, 1H), 4.31 (t, J = 12.5, 1H), 6.01 (d, J = 5.0 Hz, 1H), 8.39 (d, J = 2.5 Hz, 1H), 8.40 (s, 1H), (1H missing due to water suppression); ^{13}C NMR (100 MHz, D_2O) δ 37.7, 60.6, 69.8, 74.1, 85.7, 88.6, 119.5, 143.0, 147.1, 147.9, 151.4; HRMS (ESI) m/z: Calculated for $C_{10}{}^{13}CH_{16}O_4N_5$ $[M+H]^+$ 283.1230, Observed: 283.1231.

6.11.4 AlkB expression and purification

AlkB protein was expressed and purified as reported;[28,29] BL21 (DE3) *E. coli* containing pET24a AlkB plasmid was grown at 37 °C and 220 rpm to an OD_{600} of 0.6. Protein expression was induced by addition of 0.2 mM IPTG (Melford Laboratories Ltd). Growth was continued at 28 °C for 4 hours (AlkB) followed by harvesting of cells by centrifugation. The resulting cell pellet was stored at – 80 °C. Cell pellets were resuspended to homogeneity

in 0.1 M MES, pH 5.8, 1 mM $MgCl_2$, 1 x Roche complete EDTA-free protease inhibitor cocktail. Cells were lysed on ice by sonication, and the insoluble particulates removed by centrifugation and filtration. AlkB was purified from the lysate by cation exchange chromatography using a 50 mL S sepharose column, with elution achieved by application of a gradient to 1 M NaCl. Further purification was achieved by gel filtration using a 300 mL Superdex 75 column (Pharmacia) in a buffer of 50 mM HEPES pH 7.5.

6.12 References

1. Glasoe, P. K. & Long, F. A. USE OF GLASS ELECTRODES TO MEASURE ACIDITIES IN DEUTERIUM OXIDE. *J. Phys. Chem.* **64,** 188–190 (1960).

2. Aguilar, J. A., Nilsson, M., Bodenhausen, G. & Morris, G. A. Spin echo NMR spectra without J modulation. *Chem. Commun.* **48,** 811–813 (2012).

3. Aik, W., Demetriades, M., Hamdan, M.K K, Bagg, E. A. L., Yeoh, K. K., Lejeune, C., Zhang, Z., McDonough, M.A., Schofield, C. J. Structural basis for inhibition of the fat mass and obesity associated protein (FTO). *J. Med. Chem.* **56,** 3680–3688 (2013).

4. Gerken, T. G., C. A., Tung, Y. C. L., Webby, C. J., Saudek, V., Hewitson, K. S., Yeo, G. S.H., McDonough, M. A., Cunliffe, S., McNeill, L. A., Galvanovskis, J, Rorsman, P., Robins, P., Prieur, X., O'Rahilly, Stephen Schofield, Christopher J. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science (80-.).* **318,** 1469–1472 (2007).

5. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2,** 2212–2221 (2007).

6. Zhang, J., Chang, H. & Kane, R. Reductive Monoalkylation of Aromatic Amines via Amidine Intermediates. *Synlett* (2001).

7. Serebryany, V. & Beigelman, L. Synthesis of 2′- O -Substituted Ribonucleosides. *Nucleosides, Nucleotides and Nucleic Acids* **22,** 1007–1009 (2003).

8. Aritomo, K., Wada, T. & Sekine, M. Alkylation of 6-N-acylated adenosine derivatives by the use of phase transfer catalysis. *J. Chem. Soc. Perkin Trans. 1* **0,** 1837 (1995).

9. Jia, G. Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T.,Pan, T., Yang, Yun G., He, Chuan N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7,** 885–887 (2011).

10. Bhattarai, S. Freundlieb, M.,Pippel, J., Meyer, A., Abdelrahman, A., Fiene, A., Lee, S., Zimmermann, H., Yegutkin, G. G., Strä, N., El-Tayeb, A., Mü, C.E α,β-Methylene-ADP (AOPCP) Derivatives and Analogues: Development of Potent and Selective ecto-5′-Nucleotidase (CD73) Inhibitors. doi:10.1021/acs.jmedchem.5b00802

11. Höbartner, C. Kreutz, C., Flecker, E., Ottenschläger, E., Pils, W., Grubmayr, K., Micura, R. The Synthesis of 2′-O-[(Triisopropylsilyl)oxy] methyl (TOM) Phosphoramidites of Methylated Ribonucleosides (m1G, m2G, m22G, m1I, m3U, m4C, m6A, m62A) for Use in Automated RNA Solid-Phase Synthesis. *Monatshefte fur Chemie* **134,** 851–873 (2003).

12. Jeong, H. S., Hayashi, G. & Okamoto, A. Diazirine Photocrosslinking Recruits Activated FTO Demethylase Complexes for Specific N^6 -methyladenosine Recognition. *ACS Chem. Biol.* **10,** 1450–1455 (2015).

13. Sauro Vittori, Anna L.,Christina S., Stefano C., Rosaria V., Adriaan P. I., Jakobien K. V.,and Gloria C. †N-Cycloalkyl Derivatives of Adenosine and 1-Deazaadenosine as Agonists and Partial Agonists of the A1 Adenosine Receptor. (1999).

doi:10.1021/JM9911231

14. Ubiali, D. Morelli, C., Rabuffetti, M., Cattaneo, G., Serra, I., Bavaro, T., Albertini, A., Speranza, G. Substrate Specificity of a Purine Nucleoside Phosphorylase from *Aeromonas hydrophila* Toward 6-Substituted Purines and its Use as a Biocatalyst in the Synthesis of the Corresponding Ribonucleosides. *Curr. Org. Chem.* **19,** 2220–2225 (2015).

15. Guengerich, F. P. & Raney, V. D. Formation of etheno adducts of adenosine and cytidine from 1-halooxiranes. Evidence for a mechanism involving initial reaction with the endocyclic nitrogen atoms. *J. Am. Chem. Soc.* **114,** 1074–1080 (1992).

16. Ashton, T. D. & Scammells, P. J. An expedient synthesis of N6-substituted-5′-modified adenosines. *Bioorg. Med. Chem. Lett.* **16,** 4564–4566 (2006).

17. Buckmann, AF; Wray, V; Vanderplas, H. SIMULTANEOUS CONVERSION OF N-(1)-(2-AMINOETHYL)ADENOSINE TO N-6-(2-AMINOETHYL)-ADENOSINE AND TRICYCLIC 1,N-6-ETHANOADENOSINE UNDER MILD AQUEOUS CONDITIONS. *Heterocycles* 1399–1419 (1995).

18. and, S. B. & Lakshman*, M. K. O6-(Benzotriazol-1-yl)inosine Derivatives: Easily Synthesized, Reactive Nucleosides. (2007). doi:10.1021/JA064682N

19. Ching, S. M., Tan, W. J., Chua, K. L. & Lam, Y. Synthesis of cyclic di-nucleotidic acids as potential inhibitors targeting diguanylate cyclase. *Bioorganic Med. Chem.* **18,** 6657–6665 (2010).

20. Pirrung, M. C., Fallon, L., Lever, D. C. & Shuey, S. W. Inverse phosphotriester DNA synthesis using photochemically-removable dimethoxybenzoin phosphate protecting groups. *J. Org. Chem.* **61,** 2129–2136 (1996).

21. Fernández-Lucas, J., Fresco-Taboada, A., Acebal, C., de la Mata, I. & Arroyo, M. Enzymatic synthesis of nucleoside analogues using immobilized 2′-deoxyribosyltransferase from Lactobacillus reuteri. *Appl. Microbiol. Biotechnol.* **91,** 317–327 (2011).

22. Matsuda, A., Inada, M., Nara, H., Ohtsuka, E. & Ono, A. Nucleosides and nucleotides. 126. Incorporation of a mutagenic nucleoside, 5-formyl-2-deoxyuridine, into an oligodeoxyribonucleotide. *Bioorganic Med. Chem. Lett.* **3,** 2751–2754 (1993).

23. Lusic, H. Gustilo, E. M., Vendeix, F.A.P., Kaiser, R., Delaney, M. O., Graham, W. D., Moye, V. A., Cantara, W. A., Agris, P. F., Deiters, A. Synthesis and investigation of the 5-formylcytidine modified, anticodon stem and loop of the human mitochondrial tRNAMet. *Nucleic Acids Res.* **36,** 6548–6557 (2008).

24. Münzel, M. Globisch, D., Brückl, T., Wagner, M., Welzmiller, V., Michalakis, S., Müller, M., Biel, M., Carell, T. Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain. *Angew. Chemie Int. Ed.* **49,** 5375–5377 (2010).

25. a) Mohammad A. and Bergstrom, D. E.,A simple and regioselective synthesis of 13C-methylthymidine, Nucleosides & nucleotides, 1183-1190 (1998). b) Entharta, A; Freudenberger, J; Furrer, J. The CLIP/CLAP-HSQC:Pure absorptive spectra for the measurement of one-bond couplings. *J. Magn. Reson.***192,** 314–322 (2008).

26. Ghanty, U., Fostvedt, E., Valenzuela, R., Beal, P. A. & Burrows, C. J. Promiscuous 8-Alkoxyadenosines in the Guide Strand of an SiRNA: Modulation of Silencing Efficacy and Off-Pathway Protein Binding. doi:10.1021/ja307102g

27. Mikhailov, S. N. Chemical incorporation of 1-methyladenosine into oligonucleotides. *Nucleic Acids Res.* **30,** 1124–31 (2002).

28. Yu, B. Edstrom, W. C., Benach, J., Hamuro, Y., Weber, P. C., Gibney, B. R., Hunt, J. F. Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* **439,** 879–884 (2006).

29. Welford, R. W. D., Schlemminger, I., McNeill, L. A., Hewitson, K. S. & Schofield, C. J. The selectivity and inhibition of AlkB. *J. Biol. Chem.* (2003). doi:10.1074/jbc.M211058200

30. Mechtild H., Montse A., Patricia F., David W., Ramon E., Antonio R. DíazSynthesis of oligodeoxynucleotides containing6-N-([13C]methyl)adenine and2-N-([13C]methyl)guanine. *J. Chem. Soc., Perkin Trans.* 1, 1997,0, 1825-1828