

PROFILING THE METHYLOME TARGETS OF HISTONE LYSINE METHYLTRANSFERASES

A thesis submitted by:

Katie Addison

In partial fulfilment of the requirements for the degree of:

Doctor of Philosophy of Imperial College

Department of Chemistry

Imperial College London

South Kensington

London

SW7 2AZ

September 2018

Declaration of Originality

I hereby confirm that the work presented within this thesis is entirely my own unless otherwise stated. All work performed by others has been acknowledged within the text and referenced where appropriate.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

It is widely understood that dysregulation of epigenetic events such as post-translational modifications (PTMs) of histone proteins, are associated with a plethora of diseases, such as cancer. A prevalent and regularly studied PTM is the methylation of lysine residues on histones, mediated by the family of enzymes – histone lysine methyltransferases (HKMTs). However, to date there are considerable challenges in defining unambiguously the methylation targets of a single HKMT. The major reasons for this are: 1) there is considerable redundancy in histone methylation and multiple HKMTs can ‘write’ the same target methylation modification, 2) HKMTs also often have many non-histone substrates.

The majority of HKMTs contain the SET (Su(var)3-9, Enhancer-of-zeste and Trithorax) binding domain which binds the co-factor S-adenosyl methionine (SAM) enabling the transfer of the methyl donor group to the intended substrate. Through computational studies, the group has located an endogenous ‘hole’ in the SET binding domain of SETD7 making it structurally different to other HKMTs (e.g. G9a). The work presented here shows the synthesis of a small library of SAM analogues to be utilised by SETD7. The analogues have various groups appended at the C-2 position of the adenine ring to compliment the endogenous ‘hole’ in SETD7. Additionally, these cofactors include an isotopically labelled CD₃ reporter group so newly deposited ‘heavy’ methyl marks can be readily distinguished from endogenous methylation using LC-MS/MS analysis. Initial *in vitro* and *in vivo* studies have been performed to demonstrate the validity of this approach for histone targets with the intention of extending to a library of non-histone targets.

Acknowledgements

First of all, I would like to thank both of my supervisors Matthew Fuchter and Peter DiMaggio for the opportunity to be a part of this project. Secondly, I would like to thank all members of the Fuchter and DiMaggio groups for being a great bunch of people in and out of the lab.

I would especially like to thank Harry, Aaron, Thomas, Saulius and Nahide in the DiMaggio group who took their time to help teach me in various biological and mass spectrometry techniques and always there for discussions about anything that I was unsure on. Thank you to Diego who performed the initial *in vitro* assays. Also, I want to thank Melis, Luiza, Hannah and Alex of the Fuchter group for making the past 4 years much more enjoyable and keeping me sane throughout the whole process. I hope we will all be friends for life. On top of that, without the founding of Kelis who knows if I would have made it through these PhD years. Thank you Melis! I also want to thank all members of the Imperial CDT of Chemical Biology of Health & Disease who I have made great friends with and joined me to drown sorrows after many failed experiments in HBar – you know who you are.

Thank you also to everyone who proof read this document: Melis, Luiza, Harry, Alex and John.

Last, but not least, I want to give the biggest thank you to John who gave me continuous support and help throughout the PhD years and generally made the time so much more fun! Also, want to say thank you to all my family who have always been there for me during the hard moments.

Abbreviations

Abbreviation	Full name
Ac	acetyl
ADP	adenosine diphosphate
ADI1	aci-reductone dioxygenase
AEBSF	4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride
AgOTf	silver triflate
APIP	5'-methylthioribulose-1-phosphate dehydratase
aq.	aqueous
Ar	argon
Arg	arginine
ATP	adenosine triphosphate
BSA	bis(trimethylsilyl)acetamide
ChIP	chromatin immunoprecipitation
CID	collision induced dissociation
COSY	correlation spectroscopy
Da	Dalton
dc-SAM	S-adenosylmethionineamine
DMEM	Dulbecco's Modified Eagle's Medium
DMF	dimethylformamide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
ENOPH1	2,3-dioxomethiopentane-1-phosphate enolase/phosphatase
ES	electrospray
eq.	equivalent
FA	formic acid
FBS	faetal bovine serum
FOXO1	forkhead box protein O1
FOXO3	forkhead box protein O3
G ₁ phase	gap 1 phase
G ₂ phase	gap 2 phase

h	hours
H1, H2A, H2B, H3 or H4	histone 1, 2A, 2B, 3 or 4 protein
HMBC	heteronuclear multiple bond correlation
HCY	homocysteine
HDAC	histone deacetylase
HEK-293T	human embryonic kidney cells 293T
HKMT	histone lysine methyltransferase
HMPA	hexamethylphosphoramide
HMT	histone methyltransferase
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
HSQC	heteronuclear single quantum coherence spectroscopy
Hz	hertz
IP	immunoprecipitation
KMT buffer	lysine methyltransferase buffer
KMTB	α -keto- γ -methylthiobutyric acid
LC-MS	liquid chromatography mass spectrometry
Lys	lysine
MAT	methionine adenosyltransferase
MCF7	Michigan Cancer Foundation 7
MET	methionine
MeOTf	methyl triflate
MOB	4-methylthio-2-oxybutyrate
m.p.	melting point
MPip	methylpiperazine
MRI1	5'-methylthioribose-1-phosphate isomerase
MS	mass spectrometry
MSP	methionine salvage pathway
MTA	5'-methylthioadenosine
MTAP	5'-methylthioadenosine phosphorylase
MTR-1P	methylthioribose-1-phosphate
m/z	mass-to-charge ratio
NA buffer	nuclear acetylation buffer
NAD	nicotinamide adenine dinucleotide

NIB	nuclear isolation buffer
NM buffer	nuclear methylation buffer
NMR	nuclear magnetic resonance
NP-40	Nonidet-40
NSD2	nuclear receptor binding SET domain protein 2
ODC	ornithine decarboxylase
o/n	overnight
PARP1	poly [ADP-ribose] polymerase 1
Pip	piperidine
PBS	phosphate buffered saline
Phe	phenyl
PRMT	protein arginine methyltransferase
PMT	protein methyltransferase
PTM	post-translational modification
<i>p</i> -TsOH	<i>p</i> -toluenesulfonic acid
py	pyridine
R _f	retention factor
RP-HPLC	reverse phase-high performance liquid chromatography
rt	room temperature
s	seconds
SAH	S-adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
SAMDC	adenosylmethionine decarboxylase
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax
SETD7	SET domain containing lysine methyltransferase 7
SILAC	stable isotope labelling with amino acids in cell culture
siRNA	small interfering ribonucleic acid
SPh	thiophenyl
S phase	synthesis phase
SPhos	2-dicyclohexylphosphino-2',6'-dimethoxybiphenyl
STAT1	signal transducer and activator of transcription 1
Suv39h1	suppressor of variegation 3-9 homolog 1
TBN	<i>tert</i> -butyl nitrite

TBS	tris-buffered saline
TCA	trichloroacetic acid
TCEP	tris(2-carboxyethyl)phosphine
TFA	trifluoroacetic acid
TLC	thin layer chromatography
TOF	time of flight
Tris	tris(hydroxymethyl)aminomethane
UHPLC	ultra-high-performance chromatography
UV	ultraviolet
wt	wildtype

Table of Contents

Abstract.....	3
Acknowledgements.....	4
Abbreviations	5
1. Introduction.....	12
1.1. Epigenetics	12
1.1.1. Post-translational modifications (PTMs).....	13
1.2. Histone methylation	15
1.2.1. Histone methyltransferases (HMTs)	15
1.2.2. Histone demethylases	18
1.2.3. Histone code and combinatorial effects	18
1.3. SETD7, its targets and their effects on transcription	20
1.3.1. SETD7's methylation targets	20
1.3.2. SETD7 and its structure.....	25
1.4. Techniques to profile histone lysine methylation	27
1.4.1. Conventional techniques	28
1.4.2. Chromatin immunoprecipitation (ChIP) techniques	31
1.4.3. Mass spectrometry techniques	33
1.4.4. Bump-and-hole approach	36
1.5. S-adenosylmethionine (SAM)	41
1.5.1. Biosynthesis of SAM and the methionine cycle.....	41
1.5.2. Extraction and previous syntheses of SAM	43
1.6. Project design.....	45
2. Synthesis of a small library of S-adenosyl methionine analogues specific to SETD7	49
2.1. Synthesis of the SAM analogues	49
2.1.1. Synthesis of the deprotected adenosine intermediate.....	50
2.1.2. Chlorination at the 5' position	52

2.1.3. Addition of the homocysteine side chain.....	53
2.1.4. Introducing nucleophiles at the C-2 position	60
2.1.5. Palladium cross-coupling to install a 2-phenyl group	63
2.1.6. Methylation reactions	67
3. Biological assays for identifying HKMT targets using SAM analogues	
3.1. Mass spectrometry proteomics workflow for analysing histone PTMs	70
3.2. <i>In vitro</i> assays using nucleosome substrates	71
3.3. Cellular assays	76
3.3.1. Assessment of cell permeability of SAM and SAM analogues	76
3.3.2. Analysis of transfection-based methods to improve SAM analogue permeability	81
3.3.3. Testing other cell lines.....	86
3.4. Degradation pathways of SAM	88
3.4.1. Background – Physicochemical properties of SAM.....	88
3.4.2. Degradation studies of CD ₃ -SAM and SAM analogues	90
3.5. <i>In nucleo</i> assays	94
3.6. Investigation into non-histone targets of SETD7	99
3.6.1. Background – Non-histone targets of SETD7.....	99
3.6.2. <i>In vitro</i> assays on non-histone substrates.....	104
3.7. Lysate experiments utilising 2-piperidine SAM.....	112
4. MTA and the methionine salvage pathway	113
4.1. Background – The methionine salvage pathway (MSP)	113
4.2. Utilising CD ₃ -MTA in cellular assays	117
4.3. Investigating two different methionine pathways in tandem using isotopic labels	121
5. Conclusions and future work	129
6. Methods	132
7. Experimental.....	141
8. References.....	164

1. Introduction

1.1. Epigenetics

Epigenetics is defined as a change of gene expression that is not explained by modification of the DNA sequence. C. H. Waddington first devised the term “epigenetics” in 1942 to explain aspects of developmental biology that had little mechanistic understanding.¹ This was an attempt to bring together the two separate areas at the time – developmental biology and genetics.² Epigenetic mechanisms work in addition to the DNA template to regulate gene expression systems, and thus the combination of the two results in different cell-type identities.

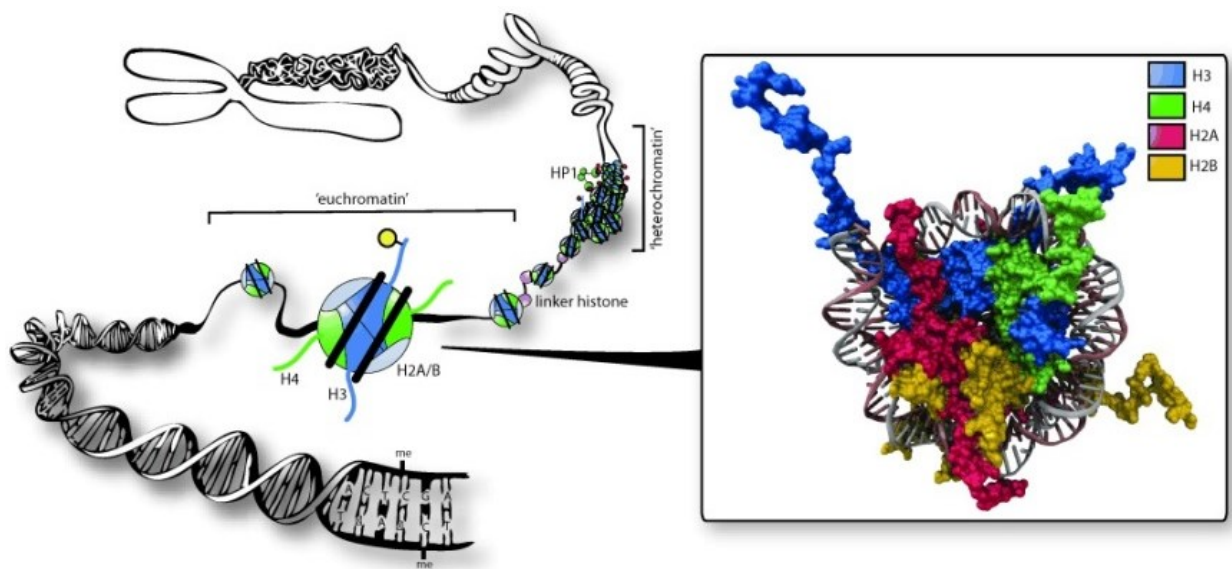


Figure 1 – Schematic representation of how histones, complexed with DNA, are folded into higher order chromatin structures, ultimately leading to chromosomes. The right hand box shows the structure of a nucleosome, the fundamental repeating unit of chromatin. A nucleosome is composed of a octamer of histone proteins – H2A (pink), H2B (yellow), H3 (blue) and H4 (green) surrounded by approximately 147 bases of DNA. Histone tails are shown to protrude from the surface of the nucleosome. Image reproduced with permission of the rights holder, John Wiley and sons.³

The majority of epigenetic information is stored in chromatin modifications, including the exchange of histone primary sequence variants, alteration of the chromatin structure and covalent modifications on histone proteins.³ Chromatin is the form that DNA is packaged into, and nucleosomes are the fundamental subunits of chromatin. The structure of the nucleosome can be seen in Figure 1 and the basic unit of a nucleosome consists of a protein octamer with approximately 147 bases of DNA wrapped around the octamer. The protein octamer consists of two molecules of each of the core histone proteins – H2A, H2B, H3 and H4. In addition to the core histones, there is

also a histone linker protein, H1, which contacts the exit/entry strand of the DNA between nucleosomes. Including linker DNA, this repeating nucleosomal unit produces a “beads-on-a-string” structure that can subsequently lead to higher-order chromatin arrangements and ultimately chromosomes (shown in Figure 1). The core histones are small, highly basic and mainly globular apart from their unstructured *N*-terminal ‘tails’ which protrude from the surface of the nucleosome (Figure 1). Histone proteins and especially their tails are subject to a wide range of histone modifications and these modifications are explained in more detail in the next section.

1.1.1. Post-translational modifications (PTMs)

This project focusses on covalent modifications on the histone proteins, otherwise known as post-translational modifications (PTMs). There are at least eight histone modifications that have been well characterised to date: methylation, acetylation, phosphorylation, ubiquitination, ADP-ribosylation, sumoylation, deamination and proline isomerisation.⁴ On top of these, other modifications have been revealed via mass spectrometry including propionylation, succinylation, butyrylation, formylation, malonylation and crotonylation.⁵ Furthermore, there are over 60 different histone residues that have been identified to be modified using specific antibodies and mass spectrometry techniques but this could be a huge underestimation of the true number. The large number of possible modifications that can occur has the potential to lead to a vast array of biological responses. In fact, abnormal modifications lead to many pathological disorders, including cancer and mental health disorders.^{6,7} These epigenetic modifications are highly dynamic and thus often reversible. This highlights the potential to apply therapeutics to reverse abnormal modifications and subsequently lead to a direct change in the accompanying disease state.

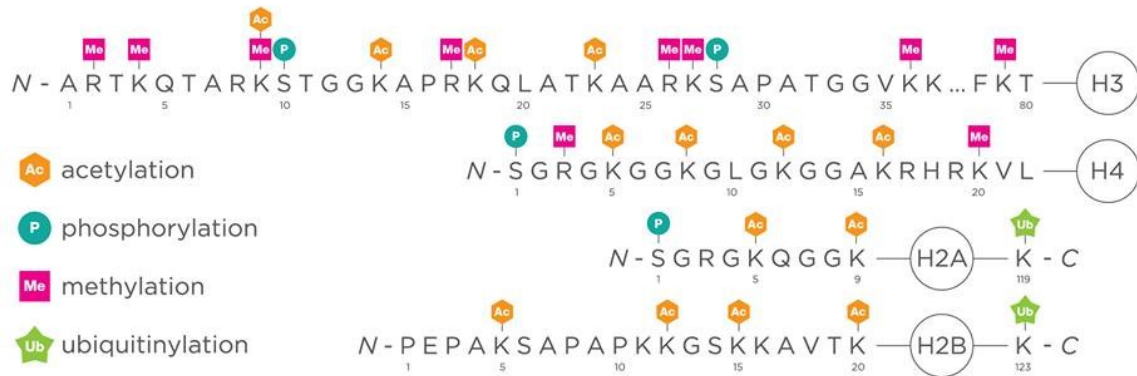


Figure 2 – Modifications that are installed on histone proteins H3, H4, H2A and H2B. Diagram adapted from M. Luo.⁸

Histone tails are extensively post-translationally modified and the most common PTMs and the histone amino acid residues that they are installed on are shown in Figure 2. PTMs can lead to altered gene expression by altering chromatin structure or via the recruitment of effector proteins. Chromatin is highly dynamic and exists in two main forms – heterochromatin and euchromatin (a schematic representation of the two forms can be seen in Figure 3). Heterochromatin is the highly condensed and predominantly silenced form of chromatin, thus the genes associated with heterochromatin are seldom expressed. Euchromatin is the more expanded form, meaning it is more accessible for transcription and the genes that are associated with euchromatin lead to transcriptional activation.

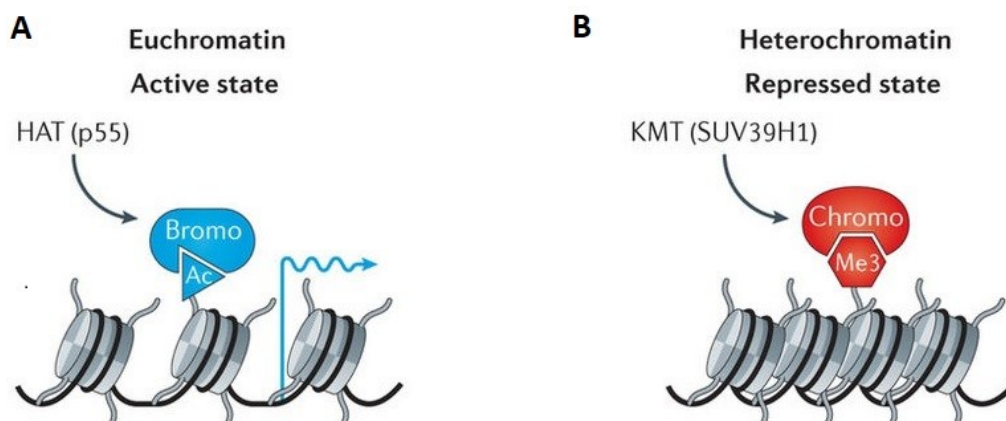


Figure 3 – Diagrams showing enzymatic examples of how PTMs can lead to altered gene expression. (A) – acetylation by p53 leads to an active transcriptional state; (B) – trimethylation by SUV39H1 leads to a repressed transcriptional state. Figure adapted from Allis and Jenuwein.¹

A couple of examples of how PTMs can lead to different transcriptional states via altered chromatin

structures and the recruitment of effector proteins are shown in Figure 3. The first example (**A**, Figure 3) involves the acetylation (Ac) on the H3 histone tail at lysine position 14 (H3K14ac) catalysed by the nuclear histone acetyltransferase (HAT) or p55.¹ The acetylation at this position provides a docking point for bromodomain (Bromo)-containing effector proteins to bind which subsequently leads to promotion of nucleosome accessibility and transcriptional activation. Acetylation can be reversed by histone deacetylases (HDACs) which have been known to cause transcriptional repression. The opposite is true for the second example (**B**, Figure 3) which involves the trimethylation at lysine 9 of histone 3 (H3K9me3) catalysed by human HKMT SUV39H1.¹ The trimethylation mark is recognised by chromodomain (Chromo)-containing effector protein HP1 that induces impaired nucleosome accessibility and leads to gene repression. In recent years, demethylases have been characterised that can reverse methylation marks presumably reversing the effects of these PTMs.

1.2. Histone methylation

In this project, the post-translational modification being investigated is methylation, particular on the lysine residues of histone proteins. It has been over 50 years since methylation was first reported on calf thymus histones, but the mechanisms of methylation on histones and its subsequent effect on biological processes is still being extensively investigated.⁹

1.2.1. Histone methyltransferases (HMTs)

Histone methyltransferases (HMTs) control methylation on specific residues of histone proteins and more recently discovered non-histone proteins. Histone methylation can be split into two major families: lysine methylation and arginine methylation.¹⁰ Methylation of arginine residues is catalysed by a group of enzymes called protein arginine *N*-methyltransferases (PRMTs), and arginines can be monomethylated, symmetrically dimethylated or asymmetrically dimethylated. Methylation of lysine residues is catalysed by a group of enzymes called histone lysine methyltransferases (HKMTs) and lysines can be mono-, di- or tri-methylated. Histone lysine methylation occurs when the methyl group of the cofactor *S*-adenosyl-L-methionine (SAM) is transferred to the lysine target of the histone protein, resulting in the by-product *S*-adenosyl-L-homocysteine (SAH), and the mechanism of

this reaction can be seen in Figure 4. Methylation of histidine residues has also been observed but this appears to be rare and therefore has not been fully characterised.¹⁰

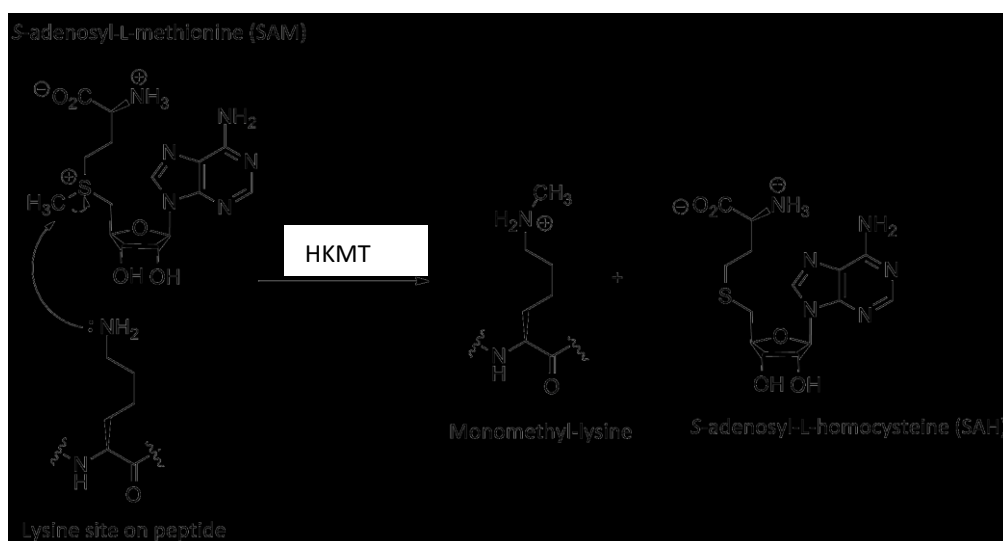


Figure 4 – The general mechanism of the methylation reaction mediated by HKMTs.

The most common methylation marks that are installed on lysine residues of histone proteins are on H3 at lysine positions 4, 9, 27, 36 and 79 and on H4 at lysine position 20 (see Figure 2). Multiple enzymes can methylate each of these lysine sites on histone proteins resulting in a redundancy of activity. For example, the methylation mark on histone H3 lysine 9 (H3K9me) can be installed by at least 8 methyltransferases.^{11,12} The HKMTs known to methylate at each histone lysine site are given in Table 1. Arginine methylation marks are commonly installed on H3 tails at arginine positions R2, R17, R26 and on H4 at arginine R3.¹³

Table 1 – Known HKMTs and the specific histone lysine targets that they methylate.⁴

Site of lysine methylation	Enzymes that can methylate at this site
H3K4	MLL, MLL2, MLL3, MLL4, MLL5, hSET1A, hSET1B, ASH2, SMYD1, SMYD2, SMYD3, SETD7
H3K9	G9a, GLP, ESET/SETDB1, SETDB2, SUV39H1, SUV39H2, PRDM2/RIZ1, CLR4
H3K27	EZH1, EZH2, NSD3, G9a
H3K36	NSD1-3, SET2, SMYD2
H3K79	DOT1L
H4K20	NSD1/2, SETD8, SUV420H1, SUB420H2

Lysine residues can either be mono-, di- or trimethylated, but each individual HKMT only transfers a certain number of methyl groups depending on their product specificity.¹⁴ Each individual methylation state will have different effects on chromatin and transcriptional activity. Methylation can lead either to transcriptional repression or to transcriptional activation, dependent on the site and/or the degree of methylation. An example was discussed earlier in Section 1.1.1 for the SUV39H1-mediated trimethylation on the histone lysine site H3K9, leading to transcriptional repression.¹ Both SUV39H1 and the transcriptional repressor protein HP1 are found to co-localise at transcriptionally silent heterochromatin sites.¹⁵ On the other hand, methylation at H3K4 is known to lead to active transcriptional states. Large stretches of silent heterochromatin found in fission yeast was found to contain histones methylated at lysine 9, but methylation on lysine 4 was absent.¹⁶

The majority of HKMTs are SET-domain containing proteins and the SET-domain stands for **S**uppressor of variegation (SUV), **E**nhancer of zeste (EZ), **T**ritorax and consists of ~130 amino acids of a conserved sequence which acts as the binding site for the cofactor SAM.¹⁷ The only known exception of this is DOT1L, which is a non-SET-domain containing HKMT. There are seven main SET-domain containing methyltransferase families known which are the SUV39, SET1, SET2, EZ, RIZ, SMYD and SUV4-20 families.¹⁷ In addition, there are a couple of orphan HKMTs such as SETD7 and SET8 (or PR-SET7).¹⁸ These families are grouped by similarity of SET domain and comparable sequence motifs surrounding the SET domain.

SET-domain HKMTs bind the cofactor SAM and the substrate peptide at opposite sides of the enzyme and are connected by a hydrophobic channel to allow the transfer of the methyl donor group. The SET-domain contains a specific residue that hydrogen bonds to the two hydroxyl groups of the ribose moiety of SAM. In addition, the SET-domain contains either an arginine or lysine residue that can form a salt bridge with the carboxylate present on the methionine chain of SAM. The combination of these two interactions results in optimum positioning of the methylsulfonium group in the hydrophobic channel and thus allows an efficient transfer of the methyl donor group to the relevant substrate. There are also additional residues that help activate the methyl donor by using C-H → O hydrogen bonds. The function of these hydrogen bonds is to allow an arrangement where the sulphur of SAM, the carbon of the methyl group and the ε-nitrogen on the lysine substrate are all collinear. This collinearity allows a S_N2 nucleophilic attack to occur at the ε-amine for incorporation of the methyl group on the protein substrate.¹⁴ Further discussion on these interactions related to the HKMT of interest, SETD7, can be found in Section 1.3.

1.2.2. Histone demethylases

Like acetylation, methylation is reported to be a dynamic process, meaning the methylation marks installed by HKMTs can be reversible. Methylation marks are removed by two known families of histone demethylases: the amine oxidase KDM1 family and the Jumonji C (JmjC) domain-containing iron-dependent dioxygenases.¹⁰ The majority of the known demethylases to date contain the JmjC domain and were identified after the initial discovery of JHMD1 (JmjC domain-containing histone demethylase 1), as many demethylases contain this kind of domain.¹⁹ However, the first histone demethylase discovered was an amine oxidase homologue LSD1 (lysine-specific demethylase 1) which demonstrated that these methylation marks are indeed reversible.²⁰ LSD1 is a flavin adenine dinucleotide (FAD)-dependent amine oxidase which specifically removes methyl groups from mono- and di-methylated H3K4.^{20–23} It is important in this project to keep H3K4 demethylation in mind; especially since demethylation of dimethylated H3K4 could also regulate monomethylation levels at K4.

1.2.3. Histone code and combinatorial effects

The most common methylation marks that are installed on histone 3 (H3) and the effector proteins that are associated with them are shown graphically in Figure 5. The distinct domains that have been found to recognise specifically histone lysine methylation marks include plant homeodomain (PHD) finger domains and the Tudor ‘royal’ family of domains that include chromodomains, Tudor, Pro-Trp-Trp-Pro (PWWP) and Malignant Brain Tumour (MBT) domains (Figure 5).²⁴ Effector proteins can also exist as multimeric proteins that bind numerous modifications at the same time, whether that is on one nucleosome or adjacent nucleosomes.²⁵ This provides evidence for a ‘histone code’. The ‘histone code’ states that one or multiple histone modifications can act as a code for effector proteins to bind and result in specific biological outcomes.²⁶ An example emphasising this hypothesis was explained in Section 1.1.1 where the binding of the trimethylation mark on H3K9 by the chromodomain on HP1 leads to transcriptional activation.

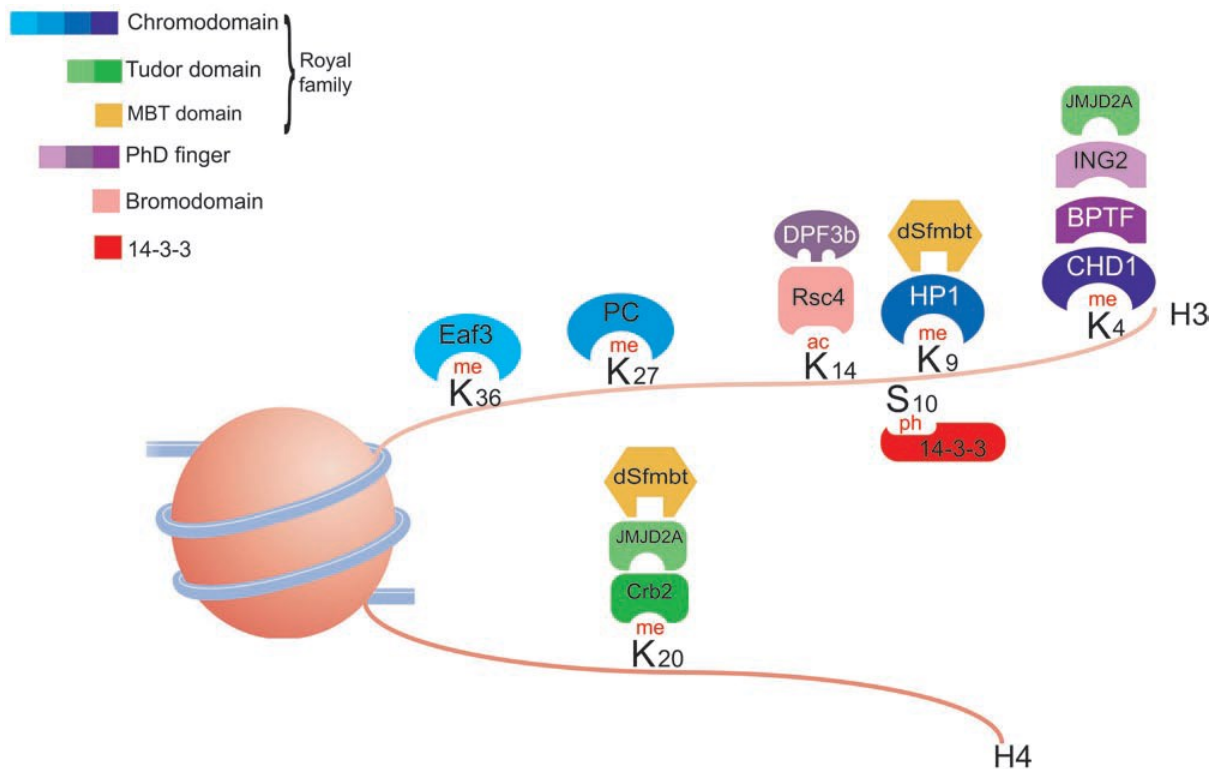


Figure 5 – A diagram showing the structure of a nucleosome with different PTMs installed on the histone tails, and also the effector proteins that can bind to the PTMs. Image reproduced with permission of the rights holder, Springer Nature.²⁴

Methylation marks are not present individually and the same histone protein can be modified by various different modifications simultaneously. Additionally, modifications are not always mutually exclusive meaning that one modification can influence the outcome of another. This ‘cross-talk’ between the abundance of modifications seen on H3 is shown in Figure 6. There are a number of ways that histone modifications communicate with each other, resulting in a positive or negative influence over each other. When modifications have a positive influence over the other it is shown with an arrow (Figure 6), such as recognising its substrate more effectively due to the adjacent modification. An example of this is GCN5, an acetyltransferase, which may recognise H3 more effectively when H3S10 is phosphorylated.⁴ Conversely, negative effects are shown with a dish-line (Figure 6), such as the binding of a protein being disrupted by an adjacent modification. For example, if the serine 10 residue on H3 is phosphorylated then it will block methylation occurring on K9 by SUV39H1, and vice versa (Figure 6).¹³ Moreover, the catalytic activity of an enzyme can be hindered by other modifications installed at its substrate recognition site. For example, isomerisation of H3P38 inhibits the ability of Set2 to methylate H3K36 (Figure 6).⁴ On top of this, communication can

Nishioka *et al.*, used Edman degradation of H3 histones labelled with ^3H -SAM to confirm that the methylation target of SETD7 is H3K4.³⁸ Mori *et al.* designed and synthesised a small library of SETD7 inhibitors that consisted of Aza-SAM substituted with various alkylamino groups (structures shown in Figure 7). When these SETD7 inhibitors were tested in *in vitro* assays against a H3 peptide substrate, the results showed a potent reduction of monomethylation on H3K4.³⁹ Therefore, the results gave evidence that SETD7 monomethylates H3K4. Further evidence of this came from crystal structure studies of SETD7 bound to histone H3 and SAM which indicated that SETD7 transfers a single methyl group to H3K4.⁴⁰

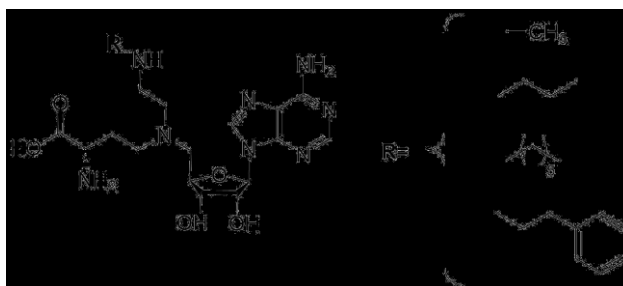


Figure 7 – The structure of Aza-SAM and the alkylamino groups that were appended to produce a library of potent SETD7 inhibitors. Adapted from Mori *et al.*³⁹

The transfer of one methyl group provides potential for subsequent modification but several studies have discounted SETD7 having the ability to di- or tri-methylate its substrates.^{35,41,42} Evidence that SETD7 is a H3K4 monomethylase is further demonstrated by peptide array analysis performed by Dhayalan *et al.*, which showed that H3K4me1, H3K4me2, H3K4me3, and H3K4 acetylation marks blocked SETD7's methylation activity on H3K4.⁴² The specificity profile of SETD7 was investigated using a stepwise iterative approach by methylating a series of peptide arrays. The peptide arrays consisted of a histone H3 tail (1–21) peptide with each amino acid residue in the tail exchanged against all other 20 natural amino acids.⁴² *Ab initio* quantum mechanical/molecular mechanical (QM/MM) molecular dynamics simulations also reported, via analysis of SETD7's crystal structure, that SETD7 is a monomethyltransferase.⁴³ The simulation showed that when the methylated lysine substrate is bound in the active site of SETD7, the SAM binding channel is opened, allowing solvent water molecules to enter the active site. This subsequently disrupts SETD7's catalytic activity and prevents further methylation (di- or tri-methylation) occurring. Further discussion on SETD7's crystal structure and the interactions that explain its catalytic activity is presented in Section 1.3.2.

However, there have been questions raised about the validity of H3K4 being a target for SETD7 in a physiological environment. Even though Wang's study provided evidence that SETD7 methylates H3K4 *in vitro*, they found marked reduction in activity when nucleosomal substrates were used.⁴⁸

This was further evidenced by Nishioka *et al.* who also observed a reduction in SETD7's activity when using nucleosomes as substrates.⁴¹ These studies provide doubt that SETD7 has a significant physiological role in the methylation at H3K4, since nucleosomal histones have a direct association with chromatin. In agreement to these *in vitro* results, RNAi-knockdown of SETD7 experiments had no effect on global methylation levels of H3K4me1 in histones extracted from LNCaP cells⁴⁴ or in MEF cell lysates.⁴⁵ This was further evidenced in a study utilising somatic SETD7 knockdown in MEF cells that resulted in no change in H3K4me1 global levels.⁴⁶

However, these findings are not consistent with SETD7 knockdown in rat mesangial cells which resulted in global H3K4me1 depletion with no change in H3K4me2 or H3K4me3.⁴⁷ In addition to cell-type differences, this disparity could be a result of the method of histone preparation differing to the previous study.²⁹ This study agrees with various studies in both human^{37,44,48,49} and animal^{47,50,51} cells types that describe SETD7-mediated transcriptional regulation by enrichment of the H3K4me1 methylation mark, identified by chromatin immunoprecipitation (ChIP). An example of one of these studies by Okabe *et al.*, proposed that gene regulation mediated by glucose is dependent on SETD7 with distinct H3K4me1 marks on target genes.⁴⁹ Additional studies by Brassachio *et al.* and El-Osta *et al.* showed that sustained upregulation of the NFκB-p65 (RelA) gene as a result of hyperglycaemia, was accompanied with increased levels of H3K4me1 but not H3K4me2 or H3K4me3 marks (Table 2).

Despite this, there are several studies that discount SETD7 as a regulator of H3K4me1 and even suggest that an alternative HKMT is maintaining nucleosomal H3K4me1 patterns. A study by Herz *et al.* showed an association with the HKMT Mll3 with H3K4me1 patterns in embryonic fibroblasts.⁵² They showed that knockdown of Mll3 resulted in significant reduction of H3K4me1 globally.⁵² Nevertheless, H3K4 methylation enrichment resulting from Mll3 was not definitively proven and secondary effects resulting from SETD7 may be contributing to the H3K4me1 pattern that was observed.²⁹ Moreover, there is the possibility that H3K4me1 is regulated by specific demethylases of H3K4me2 and me3, such as the nuclear amine oxidase homolog LSD1.²⁰

Methylation at H3K4 has been associated with euchromatin and thus linked to transcriptional activation.^{53–56} This is further evidenced by SETD7 localising to the promoter genes of collagenase type 1 gene (MMP1)⁵⁷ and the proximal promoter element of the insulin gene (Table 2).⁵⁸ This localisation temporally corresponds with the recruitment of transcription factors, chromatin remodelling factors and other components of the transcriptional machinery. These temporal links imply that SETD7 functions as a general transcription co-activator. SETD7 has been linked in several studies to hyperglycaemia and appears to regulate glucose-induced chromatin changes (Table 2).^{37,49,59} Table 2 summarises the transcriptional regulation that SETD7 has been linked to, as well as

the target promoters it associates with. However, the functions of both SETD7 and monomethylation at H3K4 are still widely unknown.^{60,61}

As well as H3K4, SETD7 has also been linked to other histone targets. SETD7 was shown to methylate recombinant full-length H2A and H2B peptides at a similar level to H3 peptides *in vivo*.⁴² H2A was found to be methylated at multiple sites but H2B was exclusively methylated at the lysine position K14.⁴² However, the role of lysine methylation on these histones is still unknown. Recently, it was reported that PARP1-mediated ADP-ribosylation of H3 histones promoted SETD7-dependent methylation of linker histone variant H1.4.⁶² On top of this, it was shown that there was negative crosstalk between ADP-ribosylation at both H3 and H1.4 and SETD7 methylation of these sites. This crosstalk suggests that ADP-ribosylation has a role in regulating SETD7's histone target preference. Four individual sites were found to be methylated on H1.4 – K121, K129, K159, K171, K177 and K192 – but this has yet to be established *in vivo* and therefore the physiological significance of this target has not been confirmed.⁶²

Table 2 – Studies in cells that demonstrate transcriptional regulation by SETD7-mediated chromatin modification. Table adapted from Keating *et al.*²⁹

Cell type	Experimental conditions	Target promoter	Reference
Human T98G glioblastoma cells	Serum stimulation	MMP1	57
LNCaP human prostate cancer cells	Androgen stimulation	PSA	63
		PSA, NKX3–1	44
Human THP-1 monocytes	TNF- α stimulation	MCP-1, IL8, TNF	64
Human and bovine endothelial cells	Transient hyperglycaemia	RelA	37,59
Human microvascular endothelial cells	Hyperglycemia	IL8	49
Murine β TC3 β -cells	Basal growth	INS1, INS2, SLC2A2	50,65
Murine C2C12 myoblasts	Serum-induced differentiation	MYOD1, MYOG, MCK, MHC	51
Rat mesangial cells	TGF- β stimulation	COL1A1, CTGF, PAI-1	47

In addition to histone targets, SETD7 has been reported to have wide range of non-histone targets. The various non-histone targets are discussed in further detail in Section 3.6 but several will be considered here in relation to SETD7's sequence specificity. The fact SETD7 has a diverse range of targets has led to investigation of the mechanism with which SETD7 recognises specific lysines.

Couture *et al.* compared the crystal structure of SETD7 bound to TAF10 peptide with ones bound to H3 and p53.³³ This study revealed a structurally conserved binding site with a short consensus motif of [K/R][π]K[X]. π represents a small residue such as Ser or Thr, K the lysine methylation site and X any residue. This consensus motif was found to be consistent with additional studies involving oestrogen receptor α ,⁶⁶ DNMT1,⁶⁷ HIF-1 α ⁶⁸ and Tat,⁶⁹ which are also reported to be SETD7s targets. In the case of the study involving DNMT1, both π and X residues were expanded upon to be [K/R][T/L/S]K[Q/S/A]. Further expansion on this motif was achieved by Dhayalan *et al.*, by applying the peptide array methylation mentioned earlier, to discover an optimised target sequence.⁴² This led to the consensus motif reported to be [GRHKPST][K>R][S>KYARTPN]K[QN][AQGMSPTYV], which was also agreed upon with studies performed on ARTD1 (or PARP1).⁷⁰ All these motifs contain both similar and differing elements and therefore they are all worth considering when investigating novel targets of SETD7. This peptide array study resulted in the identification of nine new non-histone targets – AKA6, CENPC1, MeCP2, MINT, PPARBP, ZDH8, Cullin1, IRF1, and [weakly] TTK. Table 3 summarises the most relevant SETD7 targets for this project with their target motifs highlighted, including the non-histone targets that will be investigated further in Section 3.6.

Table 3 – SETD7 substrates and target motifs characterised in *in vitro* and *in vivo* studies. Target motif is underlined with the target lysine in bold. Table adapted from Keating *et al.*²⁹

Substrate	Methylation site		Experiment
Histones	H3	K4 A <u>R</u> <u>T</u> <u>K</u> Q T A R K S	<i>in vitro</i> ²⁷ , cellular ^{37,44,47,50,51,57,59,63–65}
	H2A	Unknown	<i>in vitro</i> peptide assay ⁴²
	H2B	K15 A P K K <u>G</u> <u>S</u> <u>K</u> K A V T K A	<i>in vitro</i> peptide assay ⁴²
	H1.4	K121 E A K P <u>K</u> <u>A</u> <u>K</u> K A G A A K	<i>in vitro</i> ⁶²
		K129 A G A A <u>K</u> <u>A</u> <u>K</u> K P A G A A	
		K159 K T P K <u>K</u> <u>A</u> <u>K</u> K P A A A A	
		K171 A G A K <u>K</u> <u>A</u> <u>K</u> S P K K A K	
		K177 K S P K <u>K</u> <u>A</u> <u>K</u> A A K P K K	
		K192 K S P A <u>K</u> <u>A</u> <u>K</u> A V K P K A	
	FOXO3	K270, K271 G R A A <u>K</u> <u>K</u> <u>K</u> A A L Q T A	<i>in vitro</i> , cellular ^{71,72}
Transcription factors	FOXO1	K274 R S R A A <u>K</u> <u>K</u> <u>K</u> A S L Q S	<i>in vitro</i> ⁷¹
	RelA	K37 G M R F R Y <u>K</u> C E G R S A	<i>in vitro</i> , cellular ⁴⁵
		K314, K315 T F K S I <u>M</u> <u>K</u> <u>K</u> S P F S G	<i>in vitro</i> , cellular ⁷³
Enzyme	PARP1	K508 A A L S K <u>K</u> <u>S</u> <u>K</u> G Q V K E	<i>in vitro</i> , cellular ⁷⁰

1.3.2. SETD7 and its structure

In the same way as all other SET-domain containing HKMTs, SETD7 binds the cofactor SAM and the substrate peptide on opposite faces of the enzyme that are connected by a narrow hydrophobic channel.¹⁴ This is shown in Figure 8 for the tertiary complex of human SETD7.³¹

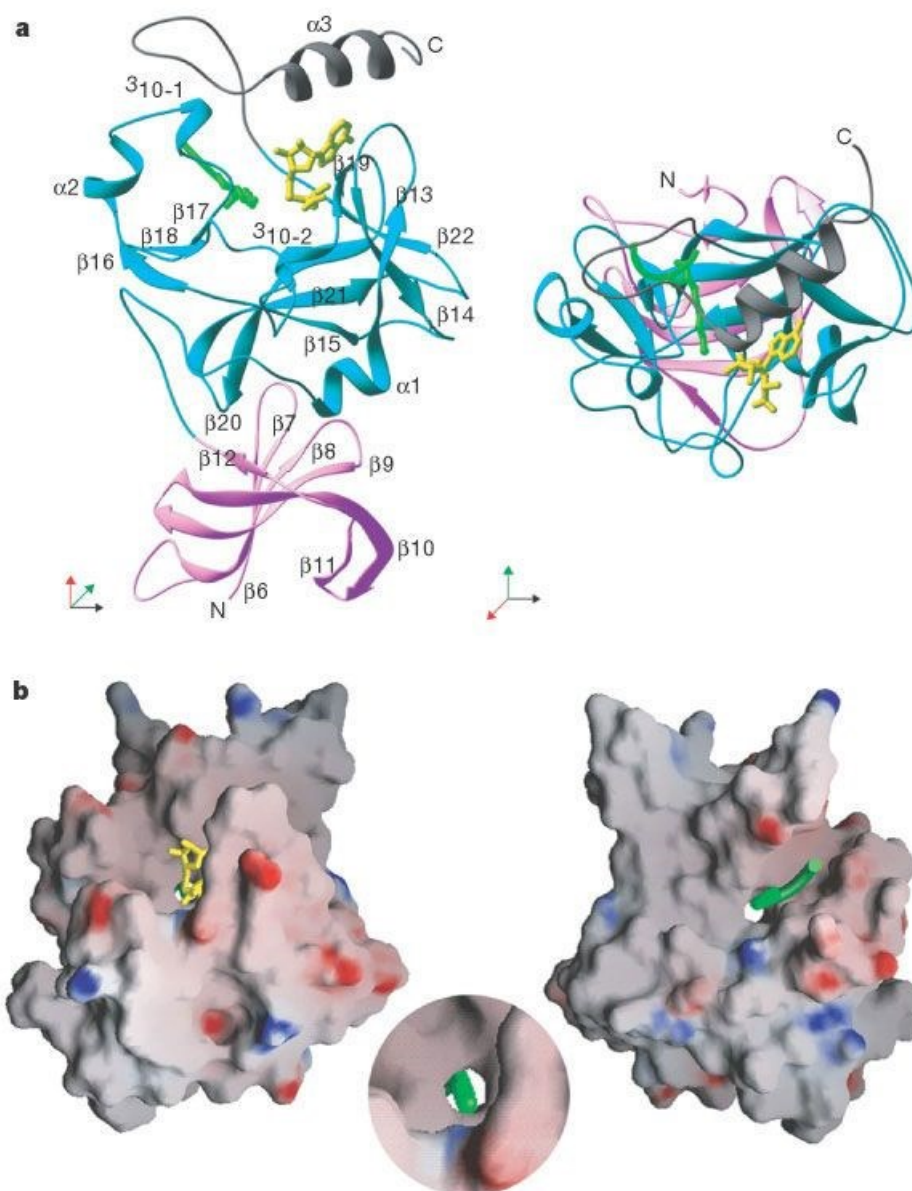


Figure 8 – Structure of the tertiary structure of SETD7 bound to SAM and H3 peptide. a) Tertiary complex in ribbons representation. The C-terminal domain is grey, the N-terminal domain is pink, the SET domain in blue, SAM in yellow and the H3 peptide in green. b) Two views of the SETD7 SET domain in surface representation (the views are related by a 2-fold rotation about a vertical axis). The inset panel shows the lysine hydrophobic channel containing the methyl lysine side chain shown from the SAM binding site. Image reproduced with permission of the rights holder, Springer Nature.³¹

In Section 1.2, it was discussed how hydrogen bonds helped activate the methyl donor by ensuring the sulphur of SAM, the carbon of the methyl group and the ϵ -nitrogen on the lysine substrate are all collinear. For SETD7, the residues involved in these interactions are Tyr-335, Gly-264 and His-293.⁷⁴ The diagram in Figure 9 shows the collinearity that is formed. Tyr-335 is a conserved residue and has been theorised to act as a general base in the methylation reaction.⁷⁵

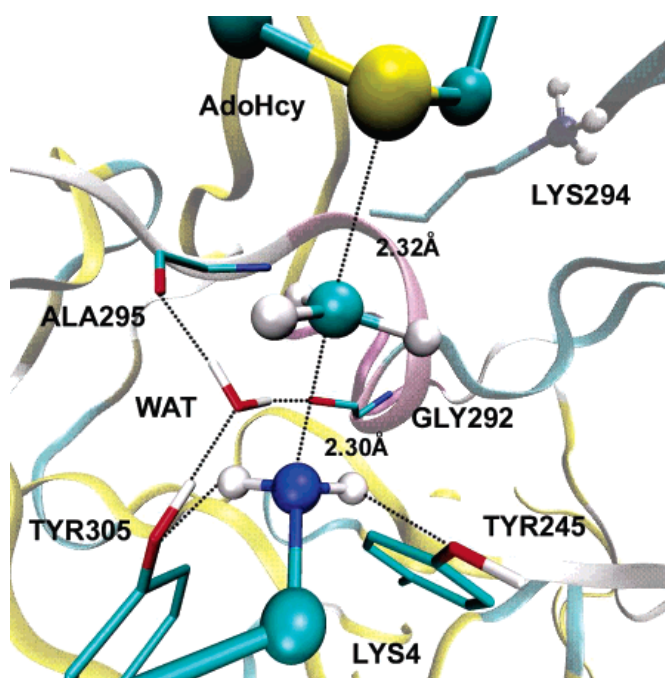


Figure 9 – The active site of HKMT SETD7 at its transition state highlighting the collinearity between the sulphur on SAM (yellow), the carbon of the methyl group (light blue) and the ϵ - nitrogen (dark blue) on the lysine substrate. Reprinted with permission from Hu and Zhang. Catalytic Mechanism and Product Specificity of the Histone Lysine Methyltransferase SET7/9: An ab Initio QM/MM-FE Study with Multiple Initial Structures. *J Am Chem Soc.* 2006; 128:1272–8. Copyright 2006 American Chemical Society.¹⁴

In addition, there are two additional interactions that are important for SETD7 to stabilise the active conformation for monomethylation. The interactions are between O₁ of Tyr-245 and H₁ of the ϵ -amino group and between O₂ of Tyr-305 and H₂ of the ϵ -amino group (also shown in Figure 9).⁷⁵ These two interactions facilitate the correct orientation of the lone pair of electrons towards the methyl group of SAM. Mutating the residue Tyr-245 for either Ala or Phe was found in multiple studies to significantly reduce the activity on a H3K4 peptide.^{31,34,40} Also, a single base mutation from a conserved histidine found in the SET domain to an alanine residue was found to impair SETD7's activity.^{27,28}

In terms of evidence that SETD7 is specifically a monomethyltransferase, free-energy simulations by Guo and Guo showed that the free-energy barrier for the second methyl transfer from SAM to H3K4me is ~5 kcal/mol higher than the barrier for monomethylation in wild-type SETD7.⁷⁵ Mutagenesis studies were performed on Tyr305 and Tyr245 to further explain the product specificity of SETD7. These residues were chosen as there are key residues residing in the lysine binding channel of SETD7 and therefore could sterically exclude the target lysine side chain with methyl groups.³⁶ Also, these residues are conserved between the SET family. Tyr305 was mutated to a Phe residue, which resulted in a lower free-energy barrier to a second methyl transfer to SAM. This mutation also interestingly turned SETD7 into a dimethyltransferase. This is logical due to the wild-type Y305 hydroxyl group being in close proximity to the terminal amine group of the target lysine. Mutation of Tyr245 to an Ala residue resulted in little activity on unmodified lysine; but allowed SETD7 to catalyse mono- and dimethylated lysine peptides to the corresponding di- and tri-methylated lysine products.³¹

1.4. Techniques to profile histone lysine methylation

In an attempt to understand these epigenetic processes further, there have been various techniques developed to profile methylation catalysed by HMTs. However, profiling the specific targets of single HKMTs is a difficult and complex task. One reason for this is the redundancy of activity between the enzymes. This redundancy means that every potential site of modification has multiple candidate enzymes that can install this particular mark. Thus, detecting the relative increase of a specific histone methylation mark only gives information on cumulative and not individual HKMT activity. Additionally, every HKMT, with the exception of DOT1L, exhibits a catalytic SET domain that binds the cofactor SAM with highly conserved hydrogen bonds, making it difficult to differentiate the individual structures of an HKMT by probing with small molecules.

Extra complexity arises from HKMTs working in a context-dependent manner. Therefore, it is important to study these enzymes in their natural chromatin environment to ensure biologically consistent results. This is emphasised by HKMTs activity being dependent on particular substrate states, for example NSD2 is a robust H3K36 dimethyltransferase on nucleosomes but has minor activity on free histones.⁷⁶ Also, HKMTs only exhibit activity in certain cellular conditions, for example – methylation of Reptin by G9a only occurs under hypoxic conditions.⁷⁷ Furthermore, these enzymes commonly form multimeric complexes in the cell to target them to a specific site and/or stabilise them, for example a multimeric complex is formed of subsets of each of the four main H3K9

HKMTs, G9a, GLP, SETDB1 and Suv39h1.⁷⁸ Lysine methylation also occurs on non-histone proteins, which have been widely understudied and the functional consequence of such methylation remaining generally unknown.

For the reasons discussed, to date, the targets of any specific HKMT have not been unambiguously defined in a biologically relevant manner. In this section, we will review the previous techniques employed for profiling targets of HKMTs, including conventional techniques, mass spectrometry techniques and the bump-and-hole approach.

1.4.1. Conventional techniques

The majority of conventional techniques to profile methylation employ activity-based assays. The two main types of activity-based assays use radiometric or antibody-based techniques, though both can be used in tandem. Activity-based assays measure the products that result from the catalysed methylation reaction between the protein substrate and the SAM cofactor – which are the methylated lysine substrate and the by-product SAH. Assays measuring the by-product SAH are performed directly^{79,80} or by converting SAH into different products that can then be measured.^{81–86} The latter approach is preferred because SAH is an inhibitor of HKMTs, and therefore measuring this directly can provide inaccurate results. In this project, the methylated lysine will be measured as the readout and therefore the focus in this section will be on approaches using this readout.

One way of detecting the methylated substrate is to use a radioactively labelled methyl donor group such as [Me-³H] or [Me-¹⁴C] to measure the enzyme activity. The signal from the donor can then be quantified using autoradiography or a scintillation method, typically after a separation step to get rid of the unreacted labelled SAM.⁸ This separation step can be laborious and time-consuming. Therefore, a more efficient assay was developed – a scintillation proximity assay (SPA) that does not require separation steps and thus is readily adaptable for high-throughput screening (HTS). This assay uses an immobilised biotinylated peptide as the substrate and [Me-³H]-labelled SAM as the cofactor. The biotinylated peptide is immobilised by a strong binding to streptavidin-conjugated SPA beads. The biotin-streptavidin binding allows close proximity between the β -particles that are produced from the radiolabelled peptide and the SPA-coated scintillation fluid on the surface of the bead, which subsequently produces a scintillation signal that can be measured (an illustrative representation is given in Figure 10). This approach has previously been used to measure the activity of the PMTs Dim5, G9a and PRMT1.^{87–89} A HTS assay that used this approach was performed by Dhayalan and co-workers to study the activity of the HKMT G9a on its non-histone substrate WIZ.

From this assay, Dhayalan *et al.* were able to determine the IC₅₀ value of SAH and the G9a inhibitor BIX-01294 for methylation of WIZ. The HTS assay was shown to have excellent reproducibility.⁹⁰

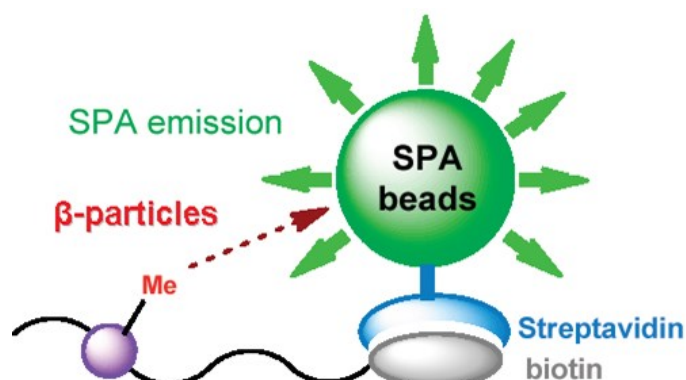


Figure 10 – Function of the SPA-based assay highlighting how the β -particles are received to produce a signal. Figure adapted from Luo *et al.*⁸

Radioactive assays are attractive techniques due to their straightforward method and the ease of access to reagents and instruments. However, these techniques are not environmentally friendly and they cannot report on the degree of methylation that has occurred.⁸ As mentioned previously, the other main type of activity-based assays are antibody-based assays. These assays use specific monoclonal or polyclonal *anti*-methyllysine antibodies to detect methylation on the target substrate. Antibodies are used to recognise protein methylation in Western Blot (WB) analysis and ChIP analysis; the latter will be discussed further in relation to mass spectrometry (MS) techniques.⁹¹

Antibodies can also be utilised in conjunction with technologies such as AlphaScreen/AlphaLISA and Lance *ultra*, shown in Figure 11. Similar to the SPA assay, methylation at the target substrate results in the acceptor and donor components becoming in close proximity with each other. The main difference here is that the donor or acceptor is conjugated to the antibody that recognises the methylated substrate, instead of the methylated substrate producing the donor signal. In the AlphaScreen/AlphaLISA assays, the donor beads are streptavidin-coated to allow binding of the biotin-tagged, methylated peptide product. The *anti*-methyllysine antibody is conjugated to acceptor beads so once the methylated product is recognised by the antibody, the acceptor beads are in close proximity to the donor beads. Excitation of the donor beads at 680 nm leads to the emission of singlet oxygen (¹O₂), which subsequently excites the acceptor beads. Excitation of the acceptor beads leads to emission at 520–620 nm (AlphaScreen) or 615nm (AlphaLISA), which can then be measured.

For the Lance *ultra* assay, the acceptor beads are conjugated to streptavidin and the *anti*-methyllysine antibody is labelled with europium. Once the antibody recognises the methylated substrate it will be in close proximity with the acceptor beads. Excitation of the europium donor at 320 or 340 nm leads to the Förster resonance energy transfer (FRET)-mediated light emission at 665 nm from the acceptor.

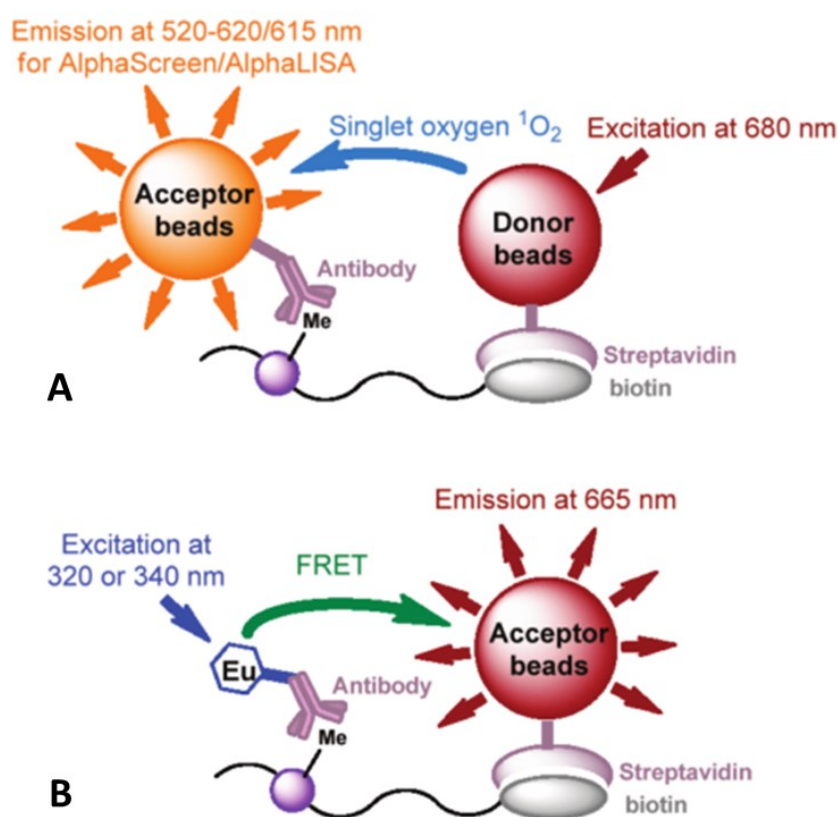


Figure 11 – Examples of homogenous antibody assays. (A) represents the AlphaScreen/AlphaLISA immunoassay using singlet oxygen transfer and (B) shows the LANCE *ultra* assay using FRET. Figure adapted from Luo *et al.*⁸

Both the AlphaLISA and LANCE *ultra* approaches have been developed and optimised to monitor methylation and demethylation at the histone lysine site of H3K4 utilising SETD7 and LSD1 respectively.⁹² These assays were designed as signal increase assays using biotinylated peptides derived from the *N*-terminus of H3. In the case of SETD7, this was also optimised for the full-length H3 protein as the substrate. These optimised assays were high-throughput and robust (Z' factors ≥ 0.7) and sensitive (only requiring nM concentrations of enzyme and substrate). The effect of SAM concentration on the methyltransferase activity from the Lance *ultra* and AlphaLISA gave similar apparent K_m concentrations for SAM of 75 μ M and 79 μ M respectively.

Recently, methods have been developed to enrich the methylation marks on proteins for proteomic analysis with high-affinity “pan-methyl” antibodies and other methyl-binding proteins that are now available.⁹³ Therefore, enrichment provides an alternative way antibodies can be utilised to effectively study methylation events. This enrichment technique would be useful to detect novel methylation events over the complex background that would be produced from cellular or lysate samples.

The activity-based assays described above are useful for determining the activity of certain HKMTs with easy readouts. However, these techniques cannot determine the degree of methylation at particular sites. Also, the antibody assay requires that you know the type of modification you want to study and its position on the substrate. This limits the use of these assays to cases where the substrate is known. Furthermore, these techniques alone cannot profile the specific methylation targets catalysed by a single HKMT. These limitations mean that these techniques alone are unsuitable for this project.

1.4.2. Chromatin immunoprecipitation (ChIP) techniques

There are a variety of ChIP immunoprecipitation techniques that have been developed including ChIP-on-chip and ChIP-Seq as well as the standard ChIP technique. The basic principle of ChIP is to selectively enrich a specific part of the chromatin (e.g. the modified histone) using an immunoprecipitation (IP) step. The modified histone is selectively enriched using specific antibodies for that modification site. This can then be coupled with gene microarrays resulting in the ChIP-on-chip technique or the ChIP sample can be directly sequenced, termed the ChIP-Seq technique. The use of methylation specific antibodies in chromatin immunoprecipitations, coupled to gene-array technology (ChIP-on-chip), produced significant advances in the ability to monitor the global occurrence of histone modifications (see Figure 12).⁴

Conventional ChIP-Seq normalisation and analysis methods are not quantitative, and fail to directly compare samples that are derived from different cell types or between cells that have experienced a global perturbation of methylation levels.⁹⁴ In recent years, there have been methods developed that have allowed direct comparison between two or more ChIP-Seq samples. One method, published by Orlando *et al.*, called ChIP-with reference exogenous genome (ChIP-Rx), used reference or ‘spike-in’ chromatin to enable comparison between samples.⁹⁴ This approach successfully measured genome-wide changes in H3K79me2 levels when human cells were introduced to a small molecule inhibitor of DOT1L, the only known HKMT to specifically methylate H3K79.⁹⁴ Another

similar strategy, that introduced *Drosophila melanogaster* chromatin and a *D. melanogaster*-specific antibody into standard CHIP reactions, successfully observed global reductions in H3K27me3 levels from samples treated with an inhibitor for the HKMT EZH2.⁹⁵

Despite chromatin immunoprecipitation presenting a powerful method to analyse global modification levels within chromatin, there are still disadvantages associated with it. This technique can detect the modification status over a single nucleosome, but it cannot detect the modification status of different histones within that nucleosome.⁴ Therefore, it is not possible to identify if there is a distinct pattern of modifications between both copies of each histone or if they are modified in an identical way.⁴ Furthermore, there are problems associated with antibodies, especially their low specificity. On top of this, adjacent modifications may disturb the binding of the antibody to its recognition site and thus lead to false readings.

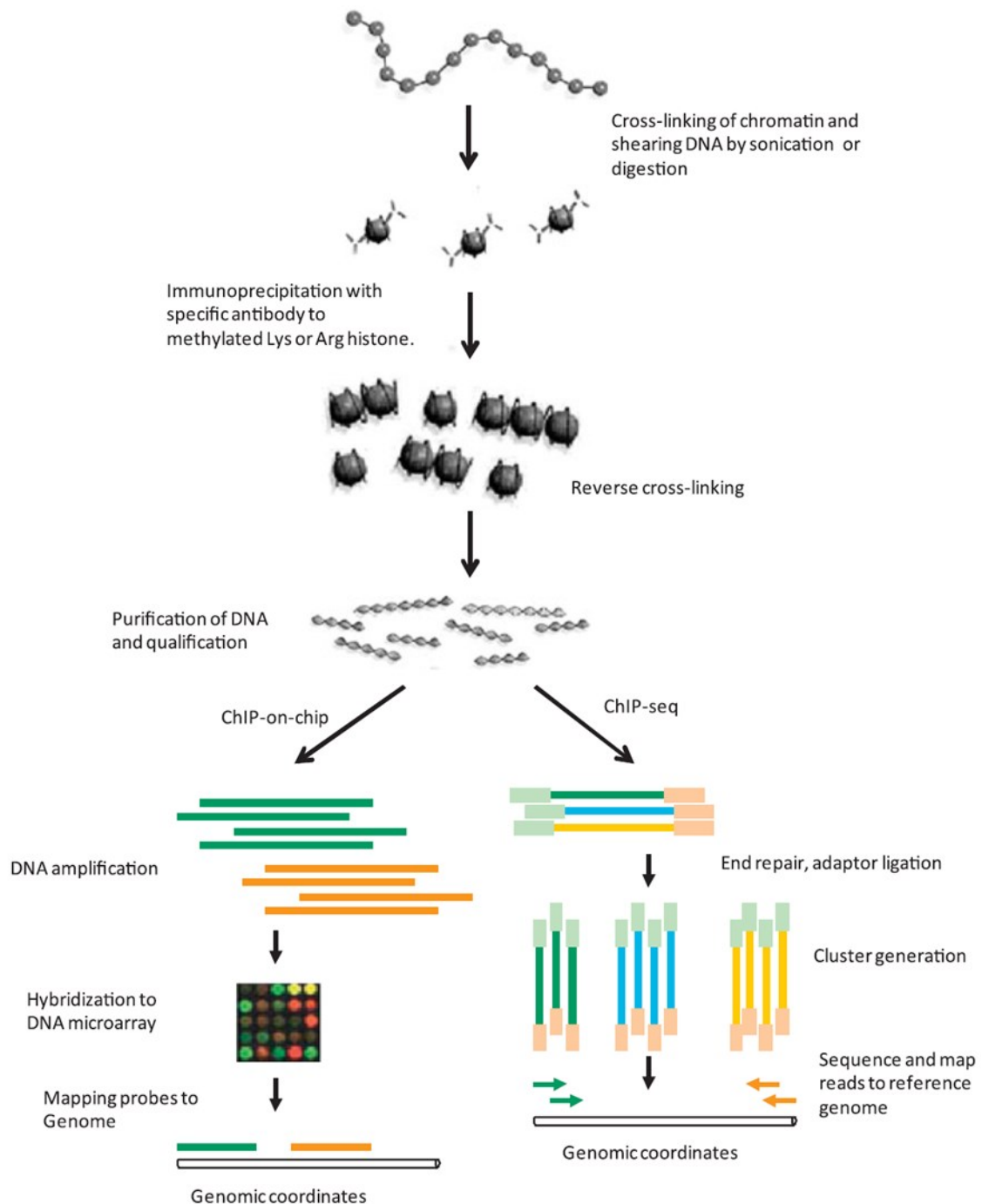


Figure 12 – A schematic diagram showing the process of ChIP, ChIP-on-chip and ChIP-seq techniques. Image reproduced with permission of the rights holder, John Wiley and Sons.⁹³

1.4.3. Mass spectrometry techniques

Mass spectrometry (MS) is a powerful technique for mapping protein methylation sites, including the degree of methylation that is installed. MS can accurately record the respective mass shift that occurs by mono- (+14 Da), di- (+28 Da) or trimethylation (+42 Da). Mapping the specific methylation

site via MS is much quicker than radiometric/antibody techniques, which need to use laborious techniques that utilise truncated peptides or site mutagenesis. Furthermore, methylation does not significantly change the size or charge of the target substrate, so the peak ratio from the MS spectra can be used for direct quantification of methylation.

The two classic proteomic MS workflows use the 'top-down' and 'bottom-up' approaches. 'Top-down' MS analyses intact proteins and is generally used for smaller, high-purity proteins like histones. This is due to complex protein samples not resolving to a sufficient degree. However, the advantage of using intact proteins is that you can unambiguously detect the methylation sites alongside other PTMs on the same protein. A high-throughput characterisation of hyper-modified combinatorial histone codes has been achieved previously by the Garcia lab.⁹⁶

On the other hand, 'bottom-up' MS uses digested proteins and separates the resulting peptides, before MS analysis, using chromatographic techniques like high-performance liquid chromatography (HPLC). The peptides are analysed using MALDI or ESI in combination with tandem mass spectrometer techniques like collision-induced dissociation (CID), electron transfer dissociation (ETD) or electron capture dissociation (ECD). This approach is generally more useful for interpreting histone modifications because it does not have the limitations associated with using highly purified, small proteins. An example of how 'bottom-up' was used to map specific methylation sites is for the tumour suppressor protein pRb. Carr *et al.* found that SETD7 specifically methylated Lys810 of pRb using this technique.⁹⁷

One successful approach for studying histone modifications has been to combine MS proteomic analysis with Stable Isotope Labelling of Amino acids in Cell culture (SILAC) experiments. This approach introduces 'heavy' methyl groups to cell media for incorporation of these methyl groups during normal cellular growth. The isotope label allows an easily observable distinction in the MS spectra between the newly deposited methylation over endogenous methylation. The labelled and unlabelled peptides that result from the SILAC method chromatographically co-elute and possess an identical charge state enabling straightforward analysis. An example MS spectrum that would result from the SILAC method is shown in Figure 13. Traditional SILAC approaches incorporate an isotopically labelled version of methionine ($^{13}\text{CD}_3\text{-MET}$) into the cell media as the source of the 'heavy' methyl group. The 'heavy' methionine will then be synthesised into SAM intracellularly. A similar approach is used in this project, but instead of methionine, an isotopically labelled SAM cofactor was utilised instead.

The Garcia lab have utilised 'top-down' MS combined with heavy methyl SILAC, to analyse *in vivo* dynamics of multiple histone lysine and arginine methylations. In general, the results from this study showed that active-gene-associated histones are methylated faster than silent-gene-associated histones. This is demonstrated by the histone marks H3K27me3 and H4K20me3, that are normally associated with silencing epigenomic regions, having among the slowest rates of turnover.⁹⁸ This is further evidenced by another study investigating global turnover of histone PTMs and variants in human cells, using SILAC in conjunction with high-resolution MS.⁹⁹ Interestingly, Zee *et al.* found that the presence of both silent and active marks on the same peptide resulted in a slower turnover rate than when either mark alone was present.⁹⁹

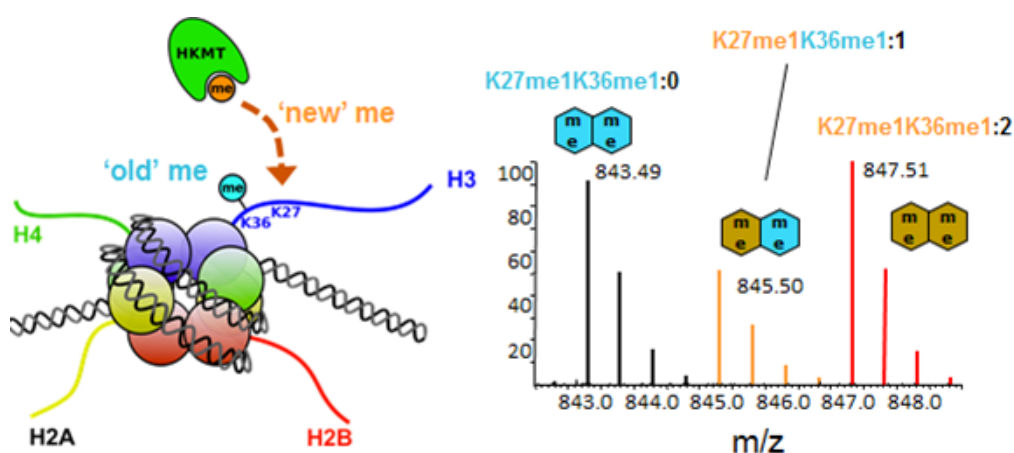


Figure 13 – Schematic representation of the SILAC method. 'New' me in orange represents the isotopically labelled methyl group and the 'old' me represents endogenous methylation.
Picture courtesy of P. DiMaggio, Imperial College London.

Mass spectrometry is an extremely powerful technique that can identify the respective mass shift that occurs by mono-, di- or trimethylation unambiguously and also locate the exact sequence on the protein that modification has been added to. One disadvantage is that peptide coverage is not always equivalent due to the different ionisation properties these modified peptides exhibit. This means that there could be a lack of sensitivity in some regions and thus might result in an incomplete picture of the genome-wide state. However, with the advanced techniques and high-resolution mass spectrometers available, this should not prove detrimental in studying histone modifications. MS will be utilised in this project to map SETD7s methylation sites by monitoring the 'heavy' methyl donor group of our SAM cofactor analogues.

1.4.4. Bump-and-hole approach

In recent years, the bump-and-hole approach has been utilised to profile the targets of both protein kinases and to some extent protein methyltransferases.^{100,101} The principle behind this approach is that a modified 'bumped' cofactor or inhibitor is synthesised that specifically recognises a mutant form of the enzyme of interest over any endogenous enzymes that are present in the system. The mutant enzyme is engineered typically with a point mutation in the active site to produce a 'hole' that can accommodate the 'bump' or substituent on the modified cofactor. In this way, only the engineered enzyme should recognise the modified cofactor and not any endogenous enzymes. In combination with a way to track the products of the enzymatic reaction between the 'bumped' cofactor and the mutant enzyme, the targets of that specific enzyme could be elucidated. A schematic representation of this is shown in Figure 14. The engineered enzyme should exhibit high catalytic efficiency for the bumped cofactor. Furthermore, on top of no cross reactivity between the 'bumped' cofactor and endogenous enzymes, ideally, the engineered enzyme would preferentially bind to the 'bumped' cofactor over the wild-type with a higher catalytic efficiency.

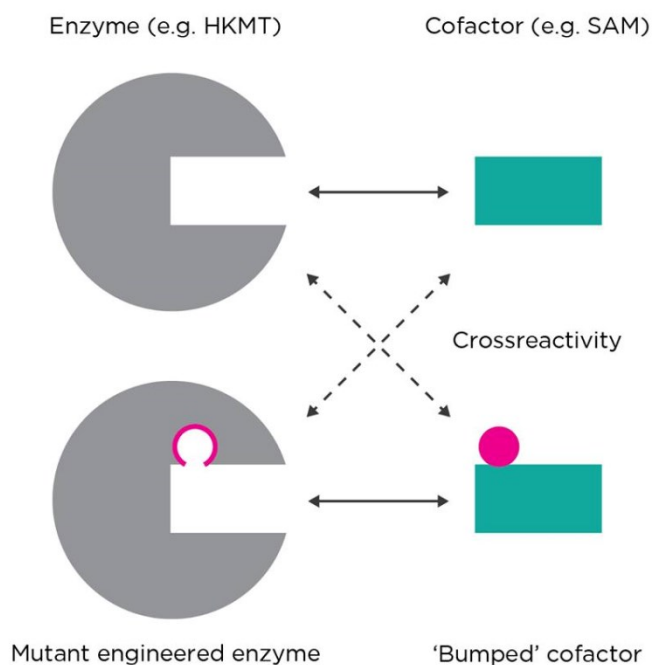


Figure 14 – Schematic representation of the bump-and-hole approach.

This approach has been successfully utilised for protein kinases using modified ‘bumped’ ATP cofactors and engineered protein kinases to profile the targets of a specific protein kinase. Protein kinases utilise ATP as the cofactor and transfer the γ -phosphate group from ATP to various amino acid residues on proteins. From crystal structures, it has been shown that protein kinases bind the adenine moiety of ATP with three hydrogen bonds via a hinge region.¹⁰² Therefore, this is where the enzymes are mutated for the bump-and-hole approach while not disrupting the essential hydrogen bonding. The first successful bump-and-hole approach for protein kinases was achieved over 15 years ago by Shokat and co-workers for the protein tyrosine kinase v-Src.¹⁰³ To furnish the engineered enzyme, a point mutation was performed at the ‘gatekeeper’ residue of the hinge domain. This residue is chosen because mutating this position will not affect the hinge domain hydrogen bonding interactions that are essential for binding SAM (shown in Figure 15). The ‘gatekeeper’ residue is mutated to a smaller residue allowing the accommodation of a bulky group that is attached to the N⁶-position of the adenine group of ATP (see Figure 15). The ATP analogue modified at the N⁶-position will not be able to bind to endogenous kinases due to the ‘gatekeeper’ residue sterically blocking this bulky group.

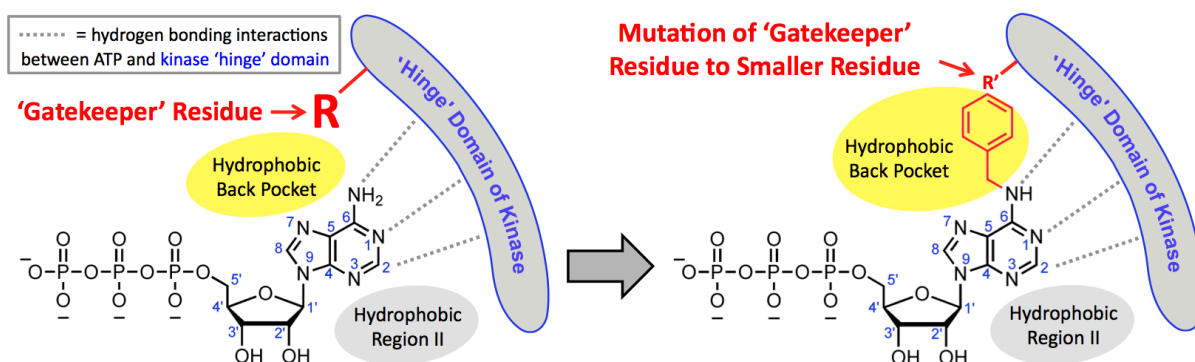
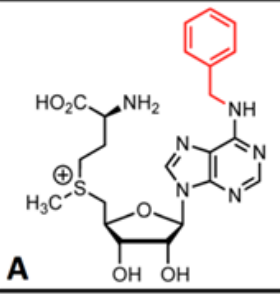
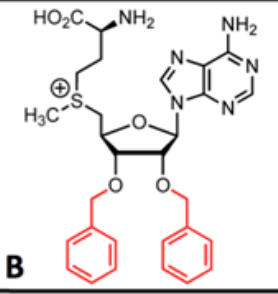
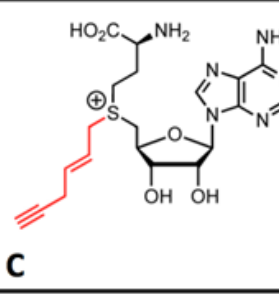


Figure 15 – A diagram highlighting the point mutation introduced to protein kinases and the complimentary 'bump' introduced at the N⁶-position of the adenine ring of ATP. Picture courtesy of P. DiMaggio, Imperial College London.

This approach was successfully utilised to yield the targets of the protein tyrosine kinase v-Src. Specifically, a N⁶-(cyclopentyl)-ATP analogue was synthesised and the ATP binding site residues V323 and I338 were mutated to alanine residues. The mutant enzyme exhibited the same target specificity as the endogenous enzyme showing that the mutation did not affect its activity.¹⁰³ This was extended for another tyrosine kinase c-Src by Witucki *et al.* and later the approach was successfully translated in more complex environment on whole-cell extracts to elucidate novel candidate substrates.^{100,104}

The bump-and-hole approach has also been applied to protein methyltransferases but to a lesser extent than protein kinases. The most successful approaches are summarised in Table 4 and will be briefly discussed here. The first approach by Lin and co-workers translates the approach from protein kinases and relates it to the arginine methyltransferase Rmt1.¹⁰⁵ The same N⁶-position on the adenine ring of SAM is modified to produce various SAM analogues with various groups incorporated, ranging from alkyl to arylamines. A complimentary point mutation is introduced at residue E117G on the HKMT Rmt1. The various SAM analogues were then screened using a MS-based assay to find which one was the most selective for the mutant Rmt1 over wild-type Rmt1 by measuring the catalytic activities for each one. This was elucidated to be the N⁶-benzyl SAM analogue (**A**, Table 4).

Table 4 – Previous bump-and-hole approaches for HKMTs. The substituent or ‘bump’ that was installed to SAM is highlighted in red, with the mutation on which HKMT that compliments this substituent reported underneath. Picture courtesy of P. DiMaggio, Imperial College London.

SAM Analogue			
SAM Modification	N ⁶ -benzyl SAM	2',3'-dibenzyl SAM	Sulfonium-alkyl SAM
PMT Mutation	Rmt1 E117G	vSET L116A	G9a Y1154A
PMT Substrate	Arginine Residues	Histone H3K27	Histone H3K9/H3K27
Substrate Target Assay	Synthetic Prmt1 Peptide	H3 Peptide (13-33)	Histone Peptides, Cells

The second study shown in Table 4 from Li *et al.* uses the SET protein vSET.¹⁰⁶ Their approach originally focussed on modifying SAM at the N⁶-position of adenine as shown before but this led to no significant methylation occurring when incubated with the corresponding mutant vSET. Therefore, this position was not suitable for modification and this was probably because the amino acid that was mutated is responsible for hydrogen bonding to the adenine ring of SAM. This led to them investigating modifications on the 2'-OH and 3'-OH groups of the ribose moiety of SAM. The most promising results came from appending benzyl groups to both the 2' and 3'-OH positions paired with a complimentary ‘hole’ formed from the modification of the Leu116 residue of vSET (**B**,

Table 4). This led to the successful methylation on lysine 27 of histone 3, vSETs target, in an *in vitro* assay.

The most common and successful approach for studying protein methyltransferases uses a 'clickable' terminal alkyne introduced at the sulphonium centre of SAM instead of the methyl group donor. The sulphonium- β sp/sp² carbon in these analogues has been shown to maintain the catalytic activity of the methylation reaction.¹⁰⁷ The alkyne that is installed from this reaction can then be modified using copper catalysed azide-alkyne cycloaddition (CuAAC) to a detectable entity. Analysis of where the detectable entity was installed will provide the specific targets of the enzyme being studied. Islam *et al.* synthesised a Hey-SAM (2c, Figure 16) cofactor where the methyl group has been replaced with a (*E*)-hex-2-en-5-ynyl group in combination with a G9a mutant Y1154A. Mutating Y1154 to a smaller amino acid residue extends the binding pocket of G9a in order to accommodate the bulky alkyl group of Hey-SAM. Islam *et al.* showed that native G9a was inactive toward Hey-SAM whilst the G9a mutant Y1154A led to the efficient labelling of known G9a histone peptide substrates with the Hey-SAM cofactor analogue (**C**, Table 4).¹⁰⁷ On top of this, endogenous SAM was only used to a minor extent by G9a Y1154A mutant, highlighting the selectivity of this mutant pair.

Additionally, this approach has been introduced to an *in vivo* system for genome-wide examination of substrate specificity using this Hey-SAM and G9a Y1154A combination.¹⁰¹ To elucidate the specific methylation targets of G9a, Wang *et al.* used the G9a Y1154A mutant and an engineered methionine adenosyltransferase (MAT) within living cells to process the alkyne methionine analogues (**1b** and **1c**, Figure 16) into the alkyne SAM analogues (**2b** and **2c** in Figure 16) intracellularly.¹⁰¹ This bump-and-hole approach overcomes SAM's limited cell permeability by introducing the cell-permeable methionine analogues. The methionine analogues have the 'clickable' terminal alkyne introduced at the sulfonium centre, which then is converted to SAM inside the cell. Subsequently, the SAM analogue will only bind to the engineered G9a and therefore methylate only G9a's specific targets. Following the methylation assay, Wang *et al.* used a novel approach similar to the ChIP techniques called clickable chromatin enrichment with parallel DNA sequencing (CliEn-seq).¹⁰¹ This uses the CuAAC reaction discussed before which binds to the alkyne moieties that will now be present on the histone proteins, followed by affinity pull-down with streptavidin beads.

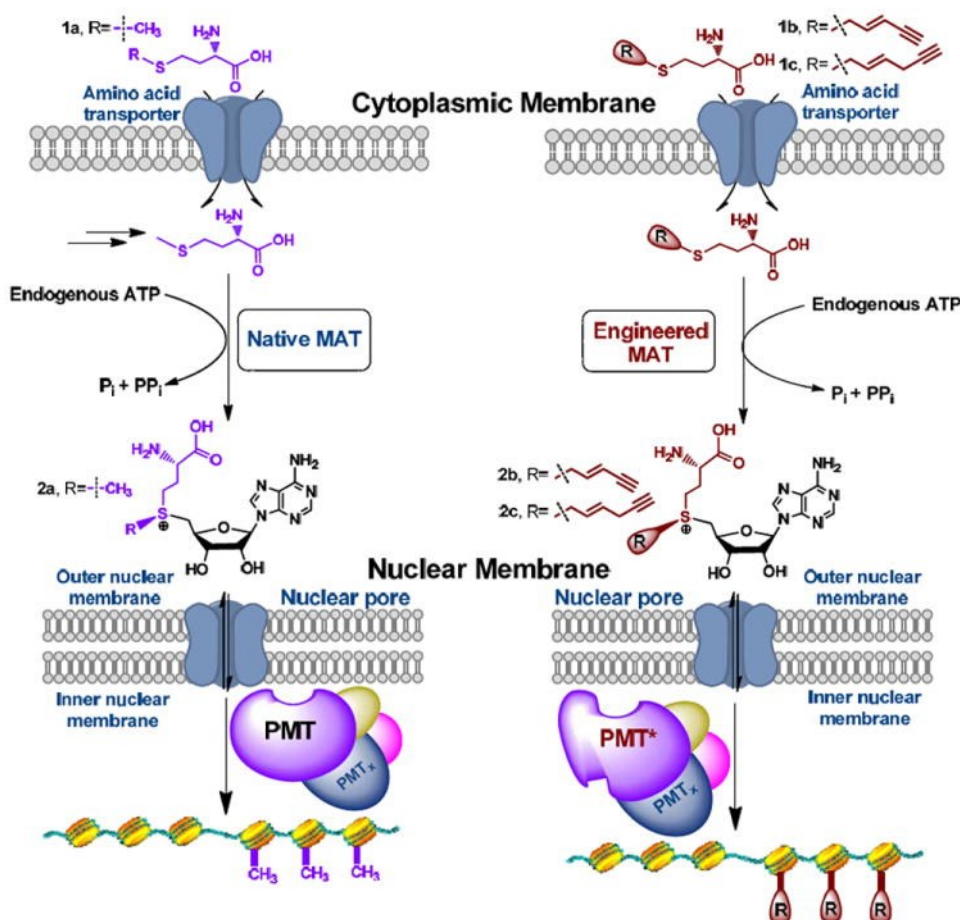


Figure 16 – Bump-and-hole approach using engineered MAT (methionine adenosyltransferase) to synthesise SAM analogues intracellularly. In an endogenous context, methionine (1a) and adenosine triphosphate (ATP) will be processed by MAT into SAM (2a). SAM can then function as a methyl donor cofactor for PMTs. In contrast, a methionine analogue (e.g. Hey-MET, 1c) can pass through the cell membrane and subsequently processed by an engineered MAT to generate the corresponding SAM analogue *in situ* (e.g. Hey-SAM, 2c). The SAM analogue is a cofactor substitute for engineered PMTs (PMT*) for biorthogonal labelling of substrates. Reprinted (adapted) with permission from Wang *et al. J Am Chem Soc.* 2013;135:1048–56. Copyright 2013 American Chemical Society.¹⁰¹

Despite the three approaches discussed showing promising results, all three have significant limitations associated with them. The first two (A and B, Table 4) have not been validated using complex substrates (e.g. chromatin or whole proteins) or in complex proteomes (e.g. lysates or cellular studies). Additionally, these two approaches had no method employed to track the newly deposited methyl mark. The last approach (C) was validated in cells; however, the alkyl donor group has no resemblance to the endogenous methyl donor group. The unnatural alkyl group will not be recognised by effector proteins and therefore the downstream effects cannot be elucidated from these studies. Additionally, the alkyl group also cannot be recognised by demethylases which could lead to long term perturbations in the system and eventually cell death.

Furthermore, all of these approaches require a mutated enzyme to obtain specificity for that enzyme over any others. Utilising mutated enzymes is likely to modify the substrate specificity of that particular enzyme and lead to false hits, especially the mutation to accommodate the bulky alkyl group in **C**. The mutation in this case is a point mutation of the Y1154 amino acid residue in G9a's active site. The Y1154 residue is a catalytically important residue for the methylation activity of G9a and therefore mutating this residue could perturb the methylation profile of G9a.¹⁰⁸

1.5. S-adenosylmethionine (SAM)

1.5.1. Biosynthesis of SAM and the methionine cycle

In order to investigate the specific methylation targets of SETD7, we will be synthesising a small library of SAM analogues that bind specifically to SETD7 and not to any other HKMTs. S-adenosyl-L-methionine (SAM) is the key cofactor substrate for all transmethylation reactions, including the reactions catalysed by HKMTs.¹⁰⁹ SAM is biologically synthesised from the essential amino acid methionine.¹¹⁰ This synthesis is catalysed by the enzyme methionine adenosyltransferase (MAT, EC2.5.1.6) using ATP as the co-substrate.^{101,111} This is the first step of the methionine cycle, which is shown in Figure 17. All mammalian cells produce SAM, but the majority of SAM is generated in the liver due to 50% of dietary methionine being metabolised here and 85% of transmethylation reactions taking place in the liver.¹⁰⁹

The *MAT* gene is conserved throughout evolution and MAT is present in cells of nearly all organisms.¹⁰⁹ There are three identified isozymes of MAT in mammals encoded by three *MAT* genes: *MAT1α* (encodes the catalytic subunit α1), *MAT2α* (encodes the catalytic subunit α2) and *MAT2β* (encodes the β subunit).¹¹² The MAT-I isozyme is a tetramer of α1 and MAT-III is a dimer of α1, whereas MAT-II is a heterotetramer formed by the α2 and β subunits. The three different isozymes of MAT also have different kinetic and regulatory properties. In respect to the metabolism of methionine, MAT-III clears methionine after a high-methionine meal whilst MAT-I and II keep basal SAM levels maintained under fasting conditions as required by the liver.¹¹² Additionally, MAT-II is inhibited by physiological concentrations of SAM.¹¹³

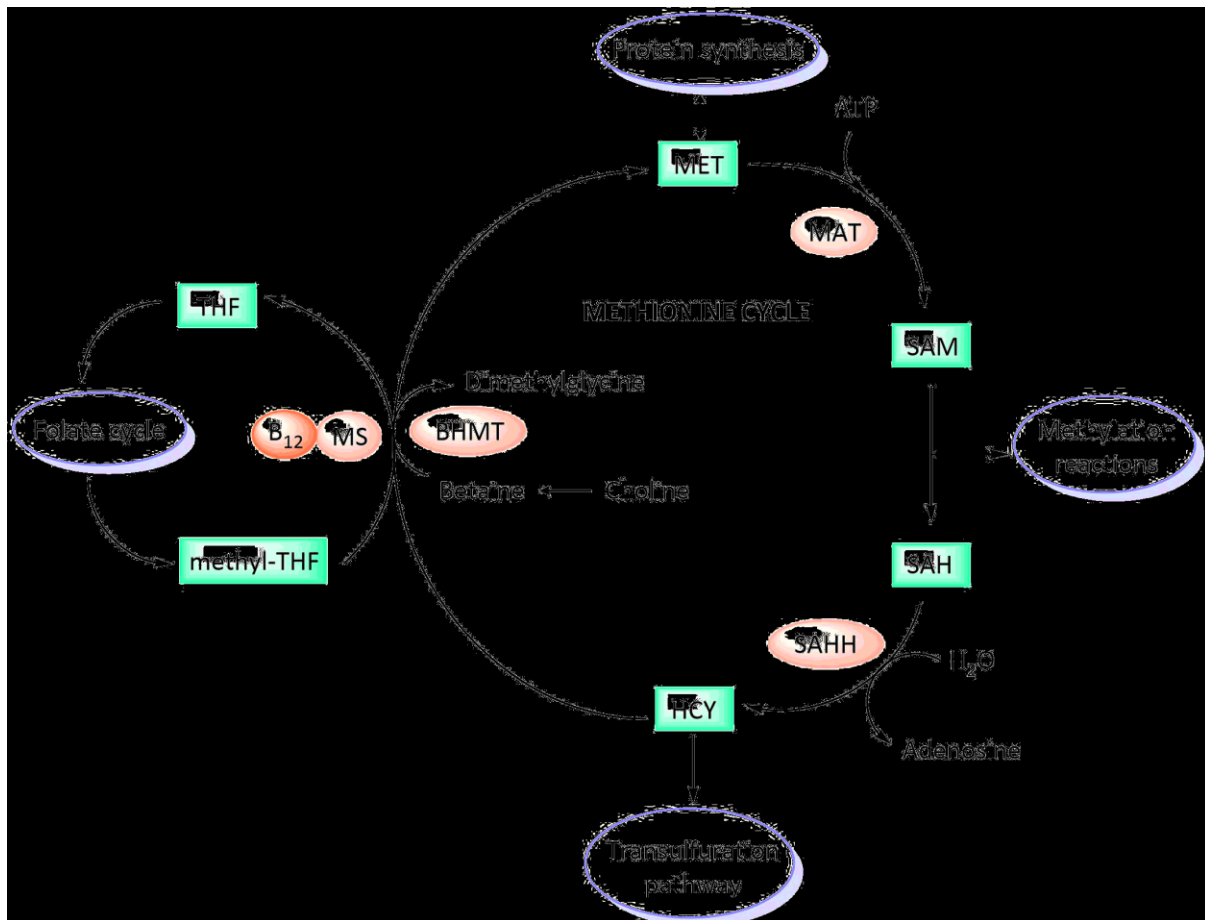


Figure 17 – The methionine cycle. MET (methionine), MAT (methionine adenosyltransferase), SAM (S-adenosylmethionine), SAH (S-adenosylhomocysteine), SAHH (S-adenosylhomocysteine hydrolase), HCY (homocysteine), BHMT (betaine-homocysteine S-methyltransferase), MS (methionine synthase).

SAM is one of the most important metabolites, along with ATP, and is only one of the few sulphur-containing active substances and the sole methyl donor *in vivo*. Once methionine is converted to SAM by MAT, SAM's methyl group can then be transferred in a wide range of methylation reactions catalysed by specific methyltransferases (MTs). The by-product of this reaction is S-adenosylhomocysteine (SAH) which is an inhibitor of the majority of PMTs. Therefore, SAH is hydrolysed to homocysteine via a reaction catalysed by S-adenosylhomocysteine hydrolase to yield adenosine as a by-product.

Homocysteine can then participate in two different processes. It can either enter the transsulphuration pathway to be converted to cysteine and α -ketobutyrate, or it can be methylated and be converted back to methionine completing the methionine cycle. Methionine can be regenerated from homocysteine in two ways: 1) betaine (produced via the conversion of choline) can be used as the methyl donor utilising the enzyme BHMT (betaine-homocysteine S-methyltransferase) or 2) methyl-THF can be used as the donor utilising methionine synthase as the

enzyme which also requires the vitamin B₁₂ as a cofactor. In the latter case, THF is produced as a by-product, which can enter the folate cycle, and subsequently methyl-THF is then regenerated from this folate cycle (Figure 17).

SAM can also participate in another key pathway – the methionine salvage pathway. This will be discussed in more detail in Chapter 4.

1.5.2. Extraction and previous syntheses of SAM

As discussed, SAM is biologically synthesised from the reaction of L-methionine with ATP, catalysed by the enzyme MAT (or SAM synthetase). The MAT enzyme supplies the diastereoselective control to the reaction and thus provides the *S*-configuration needed at the sulphur stereocentre.¹¹⁴ To obtain pure, biologically active samples of SAM, Mudd and Cantoni used extraction techniques from yeast¹¹⁵ or rabbit liver.¹¹⁶ These extraction techniques tend to result in a small yield of final product and thus are not feasible to obtain SAM on a large scale. Semi-synthetic approaches have been performed using recombinant strains from *E.coli* of the MAT enzyme to achieve up to full conversion of methionine.¹¹⁷ This semi-synthetic approach, shown in Figure 18, suffers from product inhibition but this is overcome by adding various additives; the most successful being 400 mM *p*-toluenesulfonate (*p*-TsONa), 8% β-mercaptoethanol (β-ME), 20% acetonitrile, or 2M urea.¹¹⁷

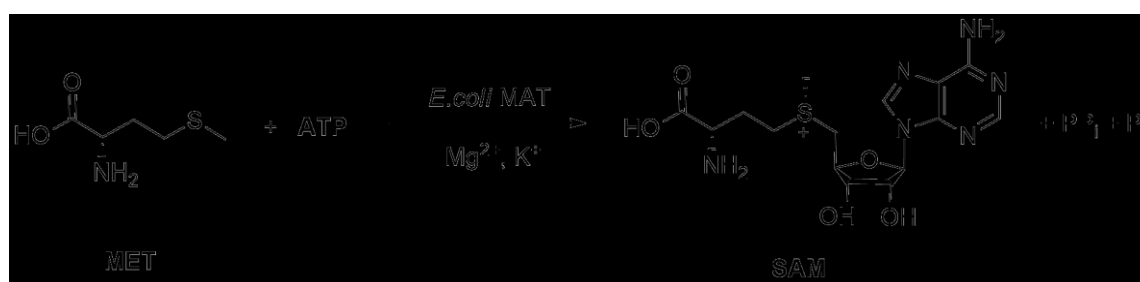


Figure 18 – Enzymatic synthesis of SAM utilising an *E.coli* strain of the enzyme MAT to provide a diastereoselective product. The maximum concentration of methionine was 10 mM and various additives were tested to optimise the reaction. Adapted from Park *et al.*¹¹⁷

Although the biological synthesis gives a highly stereoselective SAM product, it does not allow any functionality to be added to the structure of SAM, which is one of the key aims of this project. Also, the chemical synthesis of SAM from L-MET and ATP would require a high energy reaction. For these reasons, other starting materials are used in published syntheses of SAM. The bond disconnections (a and b, Figure 19) commonly employed in the chemical synthesis of SAM were investigated by

Baddiley and Jamieson in the original total synthesis of SAM.¹¹⁸ It was hypothesised at the time that 5'-haloadenosines were too unstable for these reactions leading to cyclonucleoside salts and thus were not tested for Route **b**.¹¹⁹ Instead, a tosylate group was tested (**4**, Figure 19) but ultimately Route **a** was chosen due to MTA being a more stable starting material (**3**, Figure 19). This led to a successful synthesis of SAM which confirmed the structure proposed by Cantoni *et al*.

An alternative route to SAM is to synthesise SAH and then methylate the SAH product to produce SAM. This is the preferred route in modern syntheses of SAH and SAM analogues and was chosen for this project. Despite 5'-haloadenosines previously reported to be too unstable, these have been used successfully in SAM analogue synthesis and were also utilised in this project.¹²⁰ More detail on the synthesis is presented in Chapter 2.

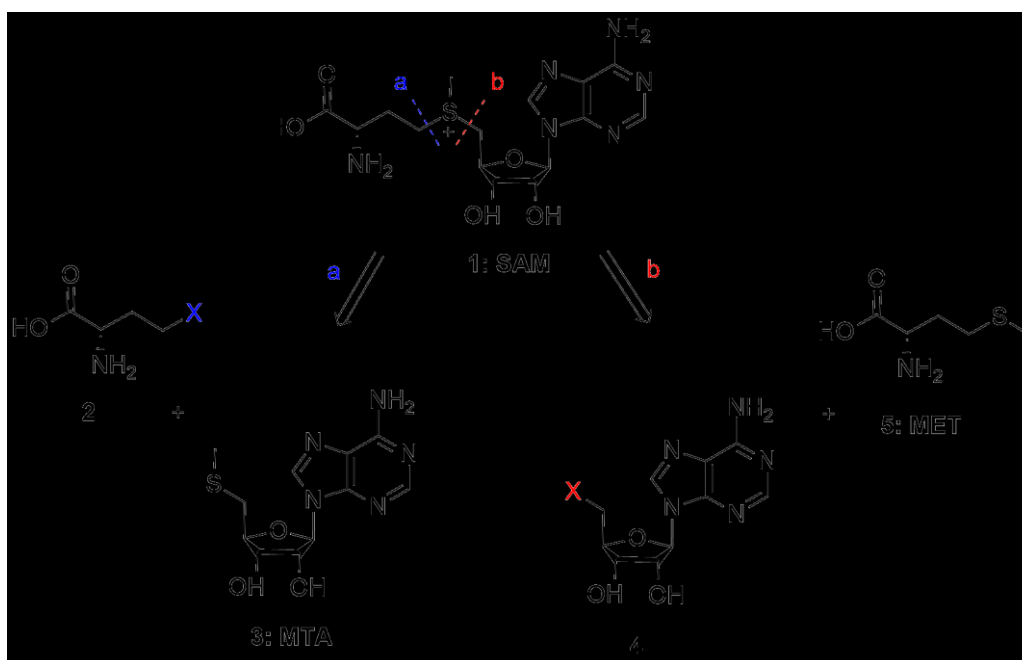


Figure 19 – Bond disconnections investigated by Baddiley and Jamieson for the chemical synthesis of SAM.
For Route **a**, X = Br, Cl or I and for Route **b**, X = OTs.¹¹⁸

1.6. Project design

The overall aim of this project is to profile the specific methylation targets of one HKMT of interest, which in this case is SETD7. In order to achieve this, a small library of SAM analogues will be synthesised that are selectively utilised by the HKMT of interest (SETD7) over other HKMTs. In this way, only lysine methylation produced from SETD7 should be detected, thus identifying the specific lysine targets of that HKMT. The label used in this project is an isotopically labelled 'heavy' methyl group (CD_3) so the targets can be identified and quantified by mass spectrometry. This method is similar to the SILAC method where the CD_3 -labelled substrates are distinguishable over endogenous methylation by the mass difference the isotope provides.

To achieve selectivity for one HKMT over others with SAM analogues, we are using an approach in this project that is conceptually similar to the bump-and-hole approach. However, our approach uses the wild-type protein instead of mutating the active site of the protein. Bump-and-hole approaches use single point mutations to obtain selectivity but using mutations could change the substrate selectivity of the enzyme (e.g. HKMT) and/or change the affinity of the active site for the cofactor (e.g. SAM). Using the wild-type protein in this approach will circumvent these problems and ensure that any targets detected are physiologically relevant.

To utilise the wild-type protein selectively, a natural pocket in the active site needs to be identified first that is structurally different to other HKMTs. In collaboration with Diego Alonso-Martinez (PhD, Chemical Engineering) an in-house program was used for this purpose based on the ESBTL library. The program takes aligned HKMT-cofactor structures as input and identifies pockets that are unique to one of the HKMTs under analysis. The program identified a natural pocket or 'hole' in the SET domains of HKMTs lacking a post-SET domain (e.g. SETD7 and SETD8) which does not exist in post-SET containing HKMTs such as G9a. The natural hole identified in the crystal structure of SETD7 is shown in Figure 20 alongside the crystal structure of G9a which does not exhibit this natural pocket.

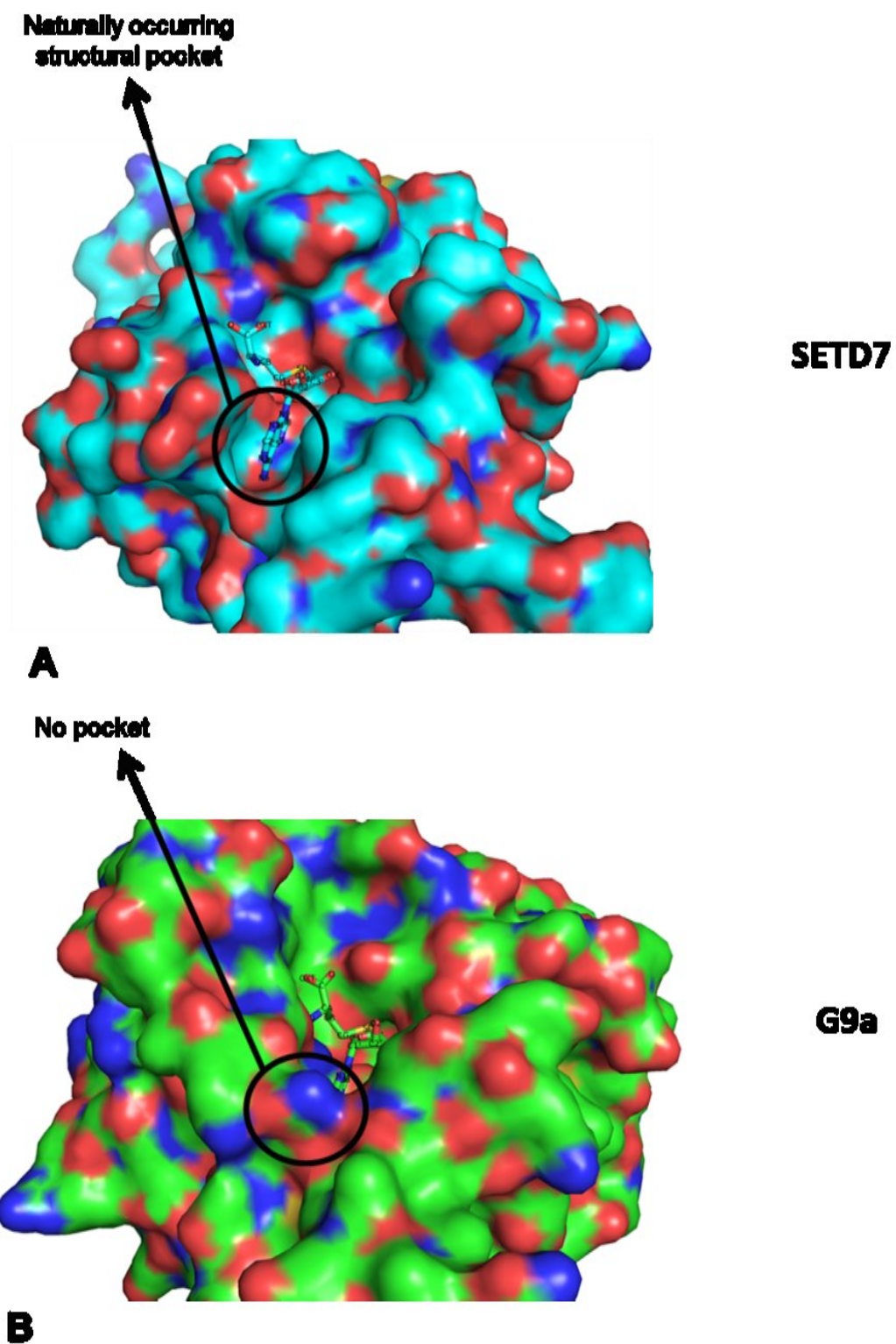


Figure 20 – Crystal structures of (A) SETD7 and (B) G9a bound to SAM. The phenyl group is added computationally and is circled to show it projecting into the naturally occurring structural pocket of SETD7 (A). G9a does not have a structural pocket so the phenyl group added clashes with the area circled (B).

The output of the program was also used to elucidate design points for incorporation of synthetic ‘bumps’ on the SAM analogue to match the identified ‘hole’ in SETD7. SAM with substituents added computationally that successfully projected into SETD7s natural pocket were identified to be introduced at the C-2 position of the adenine ring of SAM. However, for the post-SET containing HKMT, G9a, the same substituents added computationally to the C-2 position were shown to clash with the post-SET domain. Thus, C-2 modified SAM analogues should be selective for SETD7 over other HKMTs such as G9a.

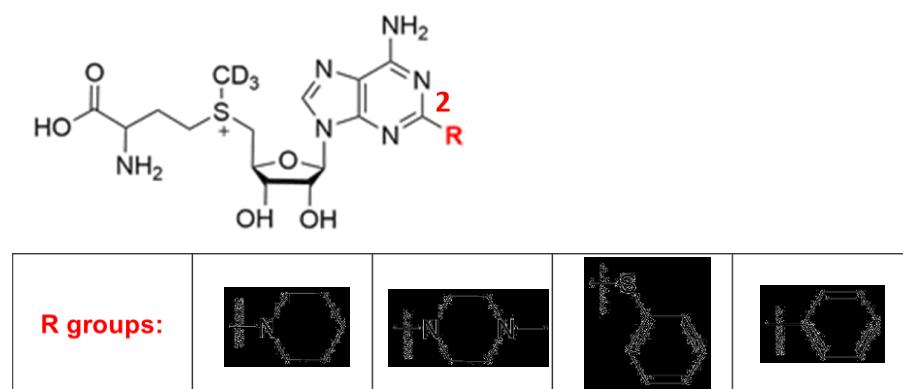


Figure 21 – The structure of the SAM analogues synthesised modified at the C-2 position. R groups incorporated at this position are shown in the table.

A small library of SAM analogues modified at the C-2 position was proposed based on the steric properties of the SETD7 binding pocket. The four that were synthesised and tested in biological assays can be seen in Figure 21. An isotopically ‘heavy’ labelled methyl (CD₃) group is also added to distinguish from endogenous methylation by LC-MS/MS, whilst still being biologically consistent. This means that any methyl-binding proteins and demethylases in the biological system studied will still be able to bind to the CD₃ group. This is important for cellular studies because if these methyl-binding proteins cannot bind to the group added, then this will significantly disrupt epigenetic programming and ultimately lead to cell death. Furthermore, if methyl-binding proteins cannot bind to this methylation event then the downstream events related to the HKMT of interest cannot be elucidated. Cellular assays are the ultimate aim for this project to validate any potential targets found for SETD7 to ensure they are physiologically relevant.

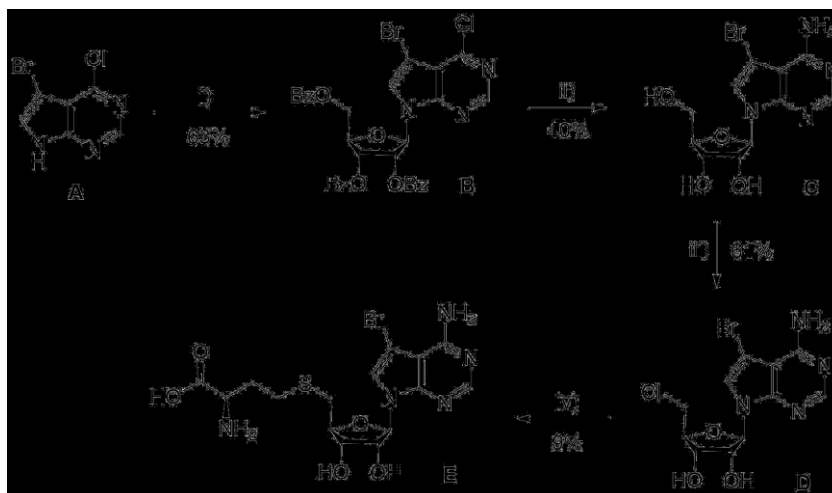
Alongside the synthesis of the C-2 modified analogues, a non-selective control was also synthesised consisting of wild-type SAM with a CD₃ ‘heavy’ label incorporated. The use of CD₃-SAM is important for the optimisation of any biological assay used and as a control to compare the activity of the SAM analogues with. The optimisation and synthesis of these analogues is described in Chapter 2. The biological assays testing the SAM analogues in respect to SETD7’s methylation activity and selectivity

are described in Chapter 3. First, the SAM analogues were validated using simple *in vitro* assays and subsequent MS analysis to quantify any methylation obtained compared against CD₃-SAM. Then, cellular and *in nucleo* assays were explored to obtain a better selectivity profile for the analogues. Cellular and *in nucleo* assays provide a more complete chromatin environment and also provides the full complement of HKMTs. In Chapter 4, we investigate the degradation product of SAM – methylthioadenosine (MTA) – and how this contributes to methylation in a cellular system.

2. Synthesis of a small library of S-adenosyl methionine (SAM) analogues specific to SETD7

2.1. Synthesis of the SAM analogues

The published synthesis of bromo-deaza-SAH from Yu *et al.* was used as inspiration for the synthetic plan of the proposed SAM analogues modified at the C-2 position (proposed analogues are shown in the Project Plan in Section 1.6).¹²⁰ There are no published examples of the complete synthesis to either SAH or SAM modified at the C-2 position of the adenine ring so the synthesis shown in Scheme 1 was used as a template to modify and expand upon.



Scheme 1 – Published synthesis of bromo-deaza-SAH.¹²⁰ i) bis-trimethylsilylacetylamide (BSA), MeCN, rt, 10 min then 1-O-acetyl-2,3,5-tri-O-benzoyl-β-D-ribofuranose, 80 °C, 1 h; ii) NH₃/MeOH, 120 °C, autoclave, 2 h; iii) SOCl₂, HMPA, rt, 18 h; iv) L-homocysteine, 1 N NaOH/H₂O, KI, 80 °C, 2 h.

Our original retrosynthetic scheme to yield SAM analogues modified at the C-2 position from adenine **1** is shown below (Figure 22). The first modification from the published synthesis was to use a different starting material **1** to allow modification at the C-2 position. The additional disconnections that have been introduced, expanding upon the bromo-deaza-SAH published synthesis, are shown with red wavy lines in Figure 22. One of the additional synthetic steps is to introduce a bromine group instead of an amine at the C-2 position, to allow, in turn the incorporation of the various R groups at the C-2 position utilising palladium cross-coupling. The very last step is methylation at the sulphonium centre to furnish the SAM analogues from SAH. In this case, a ‘heavy’ labelled CD₃ group will be used to distinguish newly deposited methylation over

endogenous methylation from wild-type SAM. All synthetic steps and further modifications made are discussed in this Chapter.

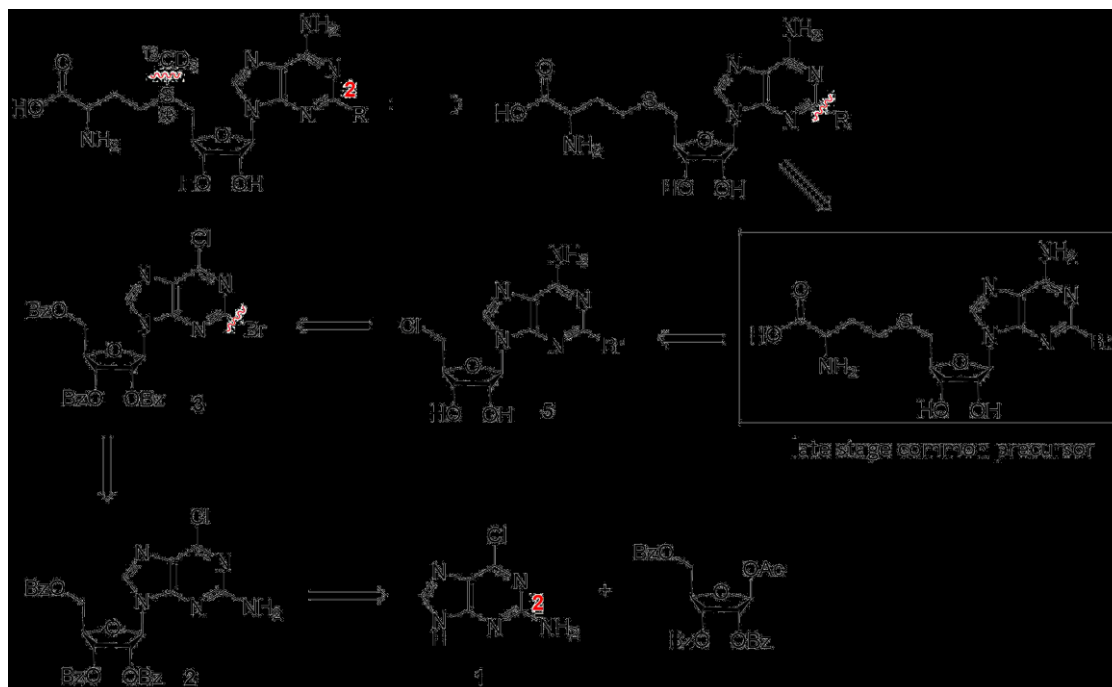
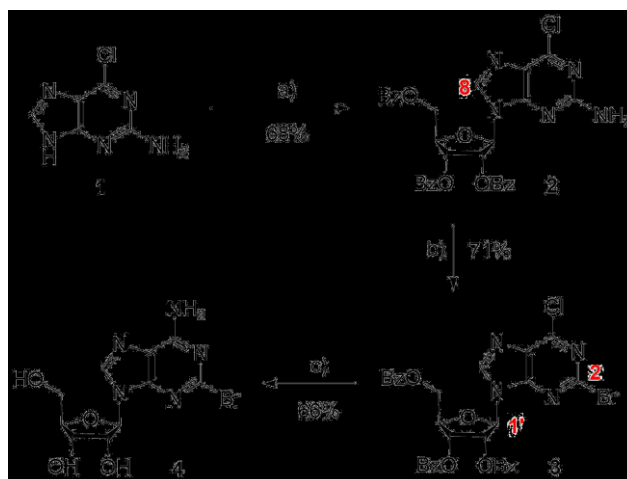


Figure 22 – Original retrosynthetic route to 2-modified SAM analogues. The C-2 position is marked in red.

The original plan was to introduce the substituents to the C-2 position at a late stage common precursor (see Figure 22) to allow an efficient way to produce a range of different analogues. Difficulties in the synthesis led to a modified route and the substituents were introduced slightly earlier, which will be discussed in this chapter.

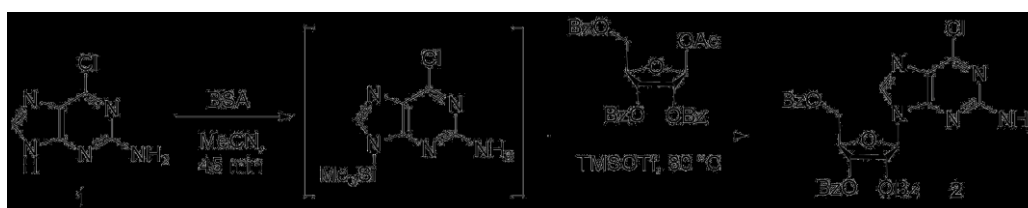
2.1.1. Synthesis of the deprotected adenosine intermediate

The first reaction of the synthesis is a glycosylation reaction between 2-amino-6-chloropurine (**1**) and the protected ribofuranose moiety using silyl-Hilbert-Johnson conditions.¹²¹ The reaction proceeds via the TMS-purine intermediate upon reaction of **1** with bis(trimethylsilyl)acetamide (BSA) (see Scheme 3). This intermediate subsequently attacks the anomeric position of the ribofuranose in a nucleophilic displacement reaction to give adenosine product **2**.



Scheme 2 – Optimised synthesis of triol 4 from 2-amino-6-chloropurine. a) (i) bis-trimethylsilylacetamide (BSA), MeCN, rt, 45 min (ii) 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl- β -D-ribofuranose, TMSOTf, MeCN, 80 °C, 1 h; b) TMS-Br, CH₂Br₂, tert-butyl nitrite (TBN), rt, 1 h; c) NH₃/ MeOH (7 N), μ W, 120 °C, 1 h.

Apart from the different purine (**1**) used as the starting material the only change made from the published synthesis by Yu *et al.* was the first step with BSA was allowed to react longer (45 min instead of 15 min).¹²⁰ The longer reaction time allowed the purine to undergo complete conversion due to the starting materials not fully dissolving in acetonitrile after 15 min. After leaving the first step to react for 45 mins, a homogenous solution was observed and subsequently the sugar was added and left to react for an additional hour. The adenosine product **2** was confirmed by NMR with the benzoyl hydrogens being observed between 7.38 and 8.04 ppm and the 8-H hydrogen (Scheme 2, C-8 position shown in red) seen at 7.92 ppm. Furthermore, the product was confirmed by MS with the correct mass detected at 614.1 *m/z*.



Scheme 3 – TMS-purine intermediate in the glycosylation reaction.

The second step is a halo-dediazotization reaction to introduce the bromine at the C-2 position of the adenine ring (Scheme 2, C-2 position highlighted in red). The procedure followed was reported by Francom and Robins and the reaction is believed to produce NOBr *in situ* and proceeds via a diazonium intermediate.¹²² The published procedure uses recrystallisation to purify the product

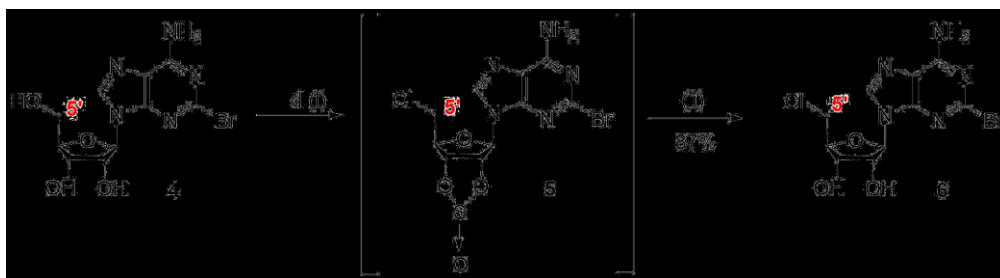
which was unsuccessful here due to different protecting groups being used – benzoyl instead of acyl used in the literature. Therefore, column chromatography was used to furnish a pure product with 71% yield. The adenosine product **3** was confirmed by mass spectrometry with the correct isotopic peaks for ^{79}Br and ^{81}Br and m/z value of 677.0. It was also confirmed by ^1H NMR with no NH_2 hydrogens present and the chemical shift of the 1'-H shifting from 6.26 to 6.47 ppm (C-1' position highlighted in red in Scheme 2).

The next step is a global deprotection of the benzoyl groups along with an aromatic substitution to introduce the amine at the C-6 position of the adenine ring (Scheme 2). The procedure by Yu *et al.* used an autoclave to provide the suitable pressure for this reaction with NH_3/MeOH (7 N). However, microwave irradiation was chosen here due to it being a safer and more convenient method. After various testing of this method, it was found the maximum concentration to be 500 mg in 20 mL in order to keep this microwave-assisted approach safe and not over pressurised. This procedure led to a 66% yield after combination of four batches with the major side product being the 6-amino benzoyl protected product.

2.1.2. Chlorination at the 5' position

To install the homocysteine side chain, first a suitable leaving group was installed at the C-5' position. For this purpose, a chlorine atom was introduced converting 5'-OH to 5'-Cl using thionyl chloride conditions (Scheme 4, C-5' position is highlighted in red). Yu *et al.* used HMPA as the solvent for the chlorination conditions but from previous work in the group this was found to not be easily scalable due to residual HMPA still being present at the end, despite various work-up conditions utilised. Additionally, HMPA is highly toxic and so an alternative solvent system was required for this reaction.

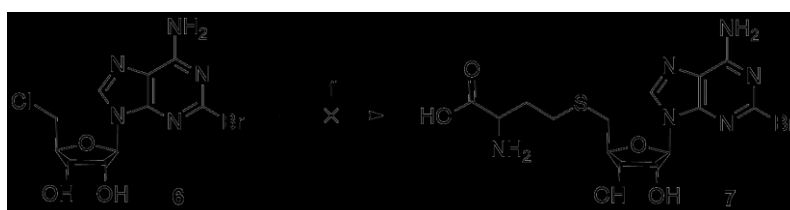
Therefore, conditions reported by Kanai *et al.* were used in which **4** underwent reaction with SOCl_2 and pyridine in MeCN to produce the intermediate sulphite ester **5**.¹²³ The sulphinate **5** was originally isolated for characterisation purposes and the formation was confirmed using MS giving a mass at 407.0 m/z . The non-sulphinated product was also observed at a lesser degree at 396.0 m/z . In all subsequent chlorination reactions, the sulphite ester was not isolated and hydrolysed fully straight after chlorination using $\text{MeOH}/\text{H}_2\text{O}$ (5:1) and NH_4OH to furnish 5'-chlorinated adenosine **6** in 87% yield over 2 steps. The chlorinated adenosine serves as a suitable intermediate to introduce the homocysteine side chain to the C-5' position via displacement of the chloride leaving group.



Scheme 4 – Chlorination at the C-5' position. d) (i) SOCl₂, py., MeCN, (ii) H₂O/MeOH (1:5), NH₄OH.

2.1.3. Addition of the homocysteine side chain

The literature procedure for the addition of the homocysteine side chain by Yu *et al.* produced a substantially low yield (9%) for this reaction using L-homocysteine as the reagent.¹²⁰ In light of this, a different procedure was found using the less expensive thiolactone version of L-homocysteine which reported higher yields (61%) after HPLC.¹⁰⁶ The procedure uses base to ring-open the cyclic thiolactone that subsequently does an S_N2 displacement at the C-5' position to create the sulphide bond. Prior to the S_N2 displacement, it also utilises potassium iodide to introduce the more reactive iodide atom at the C-5' position via an *in situ* Finkelstein reaction. Despite the literature precedence, this reaction yielded no product with mostly starting material present after work-up (Scheme 5).



Scheme 5 – Unsuccessful addition of the homocysteine side chain. f) L-homocysteine thiolactone, NaOH, KI, H₂O, 100 °C.

One reason why the above reaction did not result in any product formation may be due to insufficient ring-opening of the thiolactone ring to the reactive open chain form. Therefore, NMR studies tracking the opening of the thiolactone ring were performed to see how this reaction could be optimised which is described in Section 2.1.3.2.

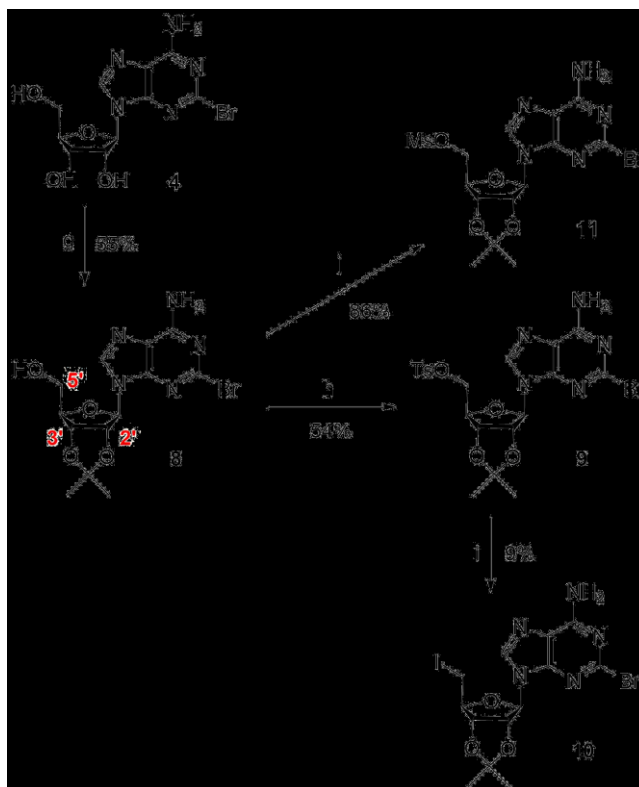
In parallel to the above reaction, other parameters were considered in order to optimise this key step in the synthesis and to find a different route to introduce the homocysteine side chain. One of the potential solutions investigated included incorporation of protecting groups to help the purification and subsequently to introduce more reactive leaving groups at the C-5' position. These protecting group strategies are investigated and discussed in Section 2.1.3.1.

2.1.3.1. Protecting group strategies

In general, the purification of SAH analogues was found to be challenging and taking into consideration that the addition of the homocysteine side chain reaction was initially unsuccessful, protecting group strategies were investigated alongside the optimisation of the previous approach (see Section 2.1.3.2.). With the C-2' and C-3'-hydroxyl groups protected, various leaving groups could be introduced at the C-5' position to both increase the reactivity of this position and improve the purification of these compounds by normal phase chromatography.

Acetonide protecting groups are widely used in the literature as a convenient way to selectively protect *cis*-diols. Therefore, the C-2' and C-3' hydroxyl groups were first protected as an acetonide by utilising the well-known conditions of *p*-TsOH in acetone (Scheme 6, C-2' and C-3' positions highlighted in red).¹²⁴ The *p*-TsOH available was in the monohydrate form so triethylorthoformate was added as a drying agent to stop any potential hydrolysis of the acetonide group. The product was confirmed by the two readily distinguishable CH₃ groups of the acetonide seen in the ¹H NMR at 1.33 and 1.53 ppm.

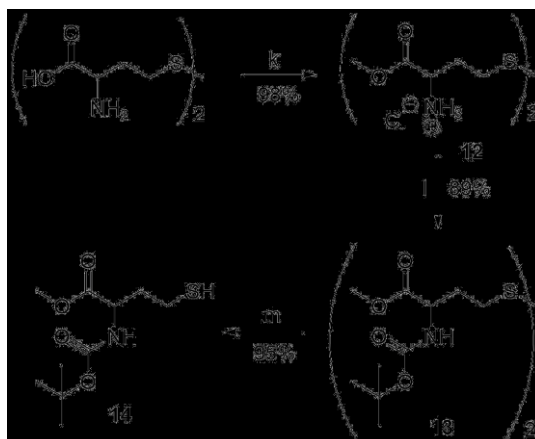
With the C-2' and C-3' hydroxyl groups protected, three separate leaving groups were introduced at the C-5' position that exhibit higher reactivity than the original chlorine atom: a tosylate group, a mesylate group and an iodide atom. Firstly, a tosylate group was introduced using tosyl chloride in pyridine to furnish **9** (Scheme 6) in modest yield (54%). The iodide group could then be introduced from the tosylated product by using conditions sodium iodide in acetone under reflux overnight. However, this strategy afforded product **10** in a very low yield of 9% and consequently, not enough material was formed to proceed with this product. The main product formed from this reaction resulted from hydrolysis of tosylate **9** to produce the 5'-hydroxyl acetonide **8**. In addition, the mesylate product (**11**, Scheme 6) was synthesised successfully utilising mesyl chloride in CH₂Cl₂ in 86% yield.



Scheme 6 – Synthesis of various different leaving groups. g) *p*-TsOH monohydrate, triethylorthoformate, acetone, rt, 18 h; h) TsCl, pyr., rt, overnight; i) NaI, acetone, reflux, o/n; j) MsCl, CH₂Cl₂, NEt₃, rt, 30 min.

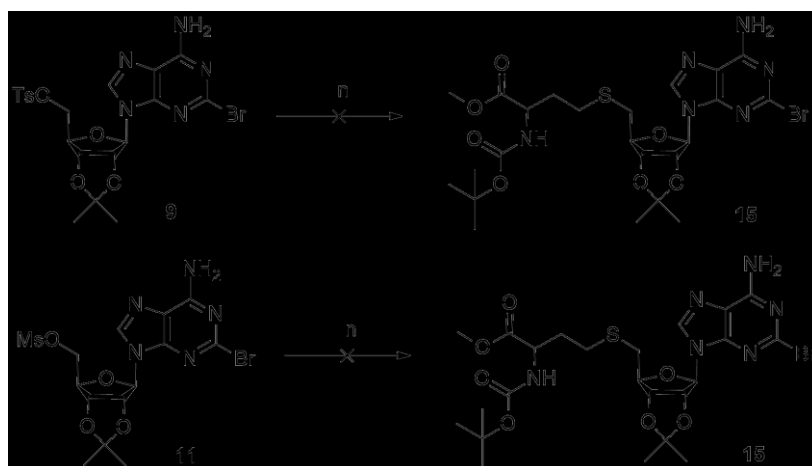
To complement acetonide protection on the adenosine ring, a protected homocysteine side chain was synthesised in an attempt to further improve the purification of these analogues. To install the planned protecting groups, L-homocysteine was used offering a more stable and less expensive alternative to L-homocysteine. The disulphide bond can later be cleaved prior to the addition to the adenosine moiety. The reported procedure from Lherbot and Keillor was followed for the protection of L-homocysteine (Scheme 7).¹²⁵

Firstly, the carboxylic acid was protected as a methyl ester in nearly quantitative yields (98%), followed by Boc protection of the amine leading to **13** in 89% yield (Scheme 7). The next step was the reduction of the disulphide bond before further reaction with the adenosine ring. Two methods were investigated: in the first method, triphenylphosphine was used as the reducing agent to afford the product **14** in 92% yield. However, despite the reaction working efficiently the triphenylphosphine oxide by-product was difficult to separate from the product by column chromatography. Thus, an alternative phosphine TCEP (tris 2-chloroethyl phosphine) was utilised providing a cleaner reaction in 95% yield.



Scheme 7 – Synthesis of protected L-homocysteine starting from L-homocystine. k) SOCl_2 (6 eq.), MeOH, 0 °C; then rt, overnight; l) $(\text{Boc})_2\text{O}$, NEt_3 , MeOH, 0 °C; then rt, overnight; m) TCEP, DMF/ H_2O (10:1), rt, 24 h.

After synthesising the two separate protected moieties **9** and **14**, they were reacted together under the conditions previously used to try to introduce the homocysteine side chain (Scheme 8). Unfortunately, this reaction was unsuccessful yielding mostly starting material and the dimer of **14** that was oxidised either in the reaction or during column chromatography. The same result was observed when utilising the C-5' mesylate-protected adenosine **11** (Scheme 8). It is hard to rationalise why this reaction did not work as the conditions vary to the harsher aqueous conditions used in Scheme 5. The solvent has been changed from H_2O to DMF and the temperature changed from 100 °C to rt. The conditions here could be too mild and has another potential issue of utilising a now very bulky nucleophile (**14**).



Scheme 8 – Attempted coupling of protected L-homocysteine with tosylated and mesylated adenosine, 9 and 11 respectively. n) **14**, K_2CO_3 , KI, DMF, rt, overnight.

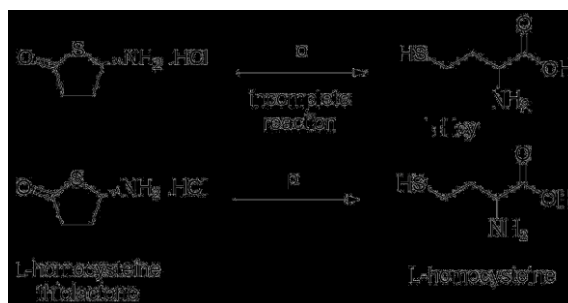
Further investigation could focus on either increasing the temperature or using different leaving groups at the C-5' position of the adenosine moiety. However, with time constraints in this project and in light of more promising results that were obtained in parallel, the optimisation of the ring opening of the thiolactone ring was carried forward to introduce the homocysteine side chain (Section 2.1.3.2).

2.1.3.2. Double substitution

The first step of the planned addition reaction when using the reagent L-homocysteine thiolactone hydrochloride (shown in Scheme 5) uses NaOH to open the cyclic thiolactone. It was hypothesised that there may have not been enough equivalents of base used in the reported literature procedure to fully open the ring, explaining why the reaction did not work previously with the 5'-chloroadenosine derivative. To test this hypothesis, NMR studies were performed where L-homocysteine thiolactone was reacted with both 2.5 eq. (the original amount) and 3 eq. of NaOH in D₂O. It is rational to use 3 eq. of base in this reaction mechanistically as:

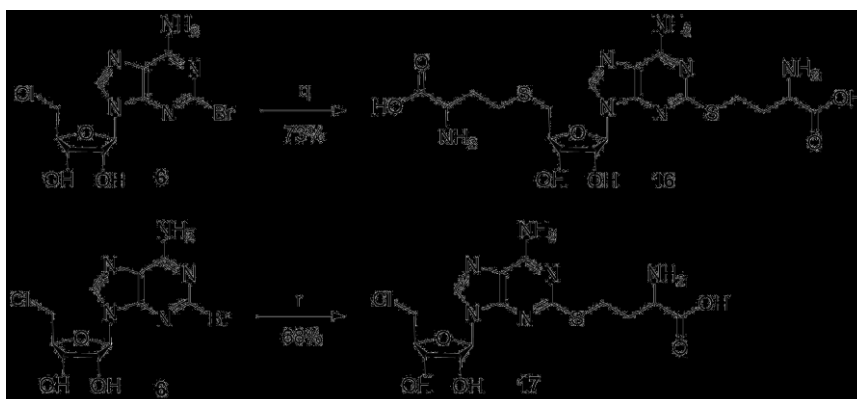
- 1 eq. will be needed to neutralise the hydrochloride salt of the commercial L-homocysteine thiolactone reagent
- 1 eq. will be needed to open the thiolactone ring
- 1 eq. will react with the formed carboxylic acid

With the use of 3 equivalents of base the ring-opening was shown to be successful, whereas only partial formation of the open chain form occurred when 2.5 equivalents was used (Scheme 9). These results confirm that higher equivalents of base are required for this reaction.



Scheme 9 – Opening of the thiolactone with varying concentrations of base. o) 2.5 eq. NaOH, D₂O; p) 3 eq. NaOH, D₂O.

Consequently, these revised conditions were applied to the reaction between the 5'-chloroadenosine derivative and the thiolactone. Although less starting material was recovered after this reaction, the desired product was still not present. From MS and NMR analysis, it was deciphered that double substitution was occurring at both the C-5' and the C-2 positions, substituting both the chlorine and bromine atoms respectively to furnish **16** (Scheme 10).



Scheme 10 – (q) Double substitution occurring at the C-5' and C-2 positions; L-homocysteine thiolactone hydrochloride (2 eq.), NaOH (6 eq.), KI, H₂O, 100 °C (r) Single substitution occurring just at the C-2 position; L-homocysteine thiolactone hydrochloride (1 eq.), NaOH (3 eq.), KI, H₂O, 100 °C.

The equivalents of thiolactone relative to eq. of NaOH were reduced from 2 eq. to 1 eq. to prevent double substitution from occurring. The reduction in equivalents succeeded in preventing double substitution, but substitution now occurred exclusively at the C-2 position instead of the desired C-5' position. Preferential substitution at the C-2 position was confirmed by MS analysis. The chlorine isotope splitting was evident and the correct mass was clearly observed with no bromine isotopes observed. The MS spectra, shown in Figure 23, shows the correct mass for **17** at 419.10 *m/z* with the correct 3:1 ratio of the two (³⁵Cl and ³⁷Cl) chlorine isotopes. This evidence highlights that under the reaction conditions employed the C-2 position is the most reactive position for nucleophilic substitution.

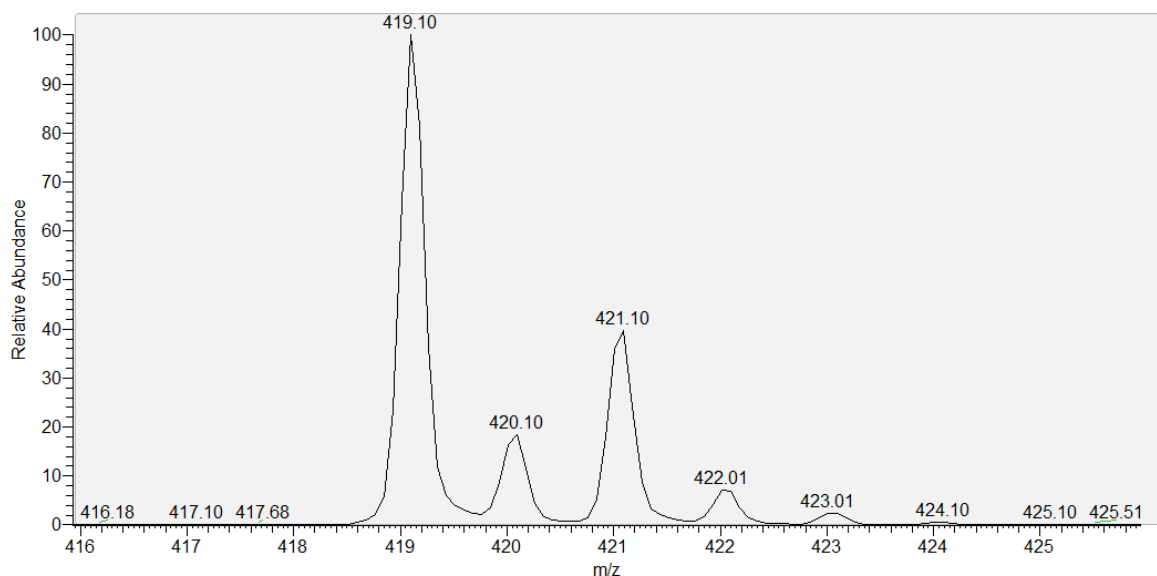
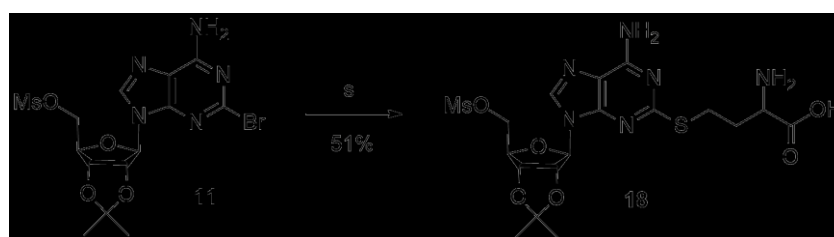


Figure 23 – MS spectra showing the preferential substitution of the homocysteine side chain at the C-2 position of the adenine ring.

One reason that substitution occurs preferentially at the C-2 position could be the bromine atom compared to the chlorine at the C-5' position being a better leaving group. Therefore, the reaction was repeated with the acetonide-protected adenosine with the mesylate leaving group installed at the C-5' position (**11**). Utilising the more reactive leaving group at the C-5' position may alter the preference for the substitution reaction towards the desired product. The acetonide protecting group has to be present here to ensure correct regioselectivity for the introduction of the mesylate group.



Scheme 11 – Testing the reactivity of the nucleophilic substitution reaction on mesylated adenosine **11.
s) L-homocysteine thiolactone hydrochloride (1 eq.), NaOH (3 eq.), KI, H₂O, 100 °C.**

When **11** was reacted with 1 eq. of thiolactone, nucleophilic substitution still occurred at the C-2 position preferentially over the C-5' position (Scheme 11). However, considering the mesylated adenosine **11** contains an acetonide protecting group it is hard to elucidate whether the C-2 position is still the most reactive or if the acetonide is hindering the side chain from attacking the C-5'

position. The acetonide could hinder the reaction by providing steric hindrance or by imposing conformational rigidity to the sugar ring and thus not allowing the attack of the nucleophile.

Despite these problems, the selectivity of this chemistry can be exploited to introduce a variety of bumped substituents to the C-2 position via nucleophilic substitution reactions. This provides an alternative route to a range of 2-substituted SAM analogues, despite the groups not being introduced as late stage as planned initially.

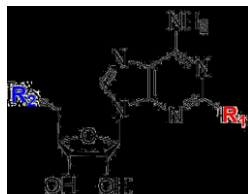


Figure 24 – Example highlighting where different R groups can be introduced sequentially.

Additionally, this chemistry could be exploited in a novel way to quickly introduce different substituents to both the C-5' and the C-2 positions sequentially to synthesise other compounds exhibiting this scaffold (Figure 24). Previous approaches have used various protecting group strategies to enable the incorporation of different substituents at both the C-2 and C-5' positions. An example of how this protecting group strategy has been applied is for the synthesis of 2,N⁶,5'-substituted adenosine derivatives that have been synthesised to exhibit potent activity against the A2B adenosine receptor.¹²⁶ The chemistry outlined here would give a more convenient and quicker route to a wide range of 2,N⁶,5'-substituted adenosine compounds and other compounds exhibiting this scaffold.

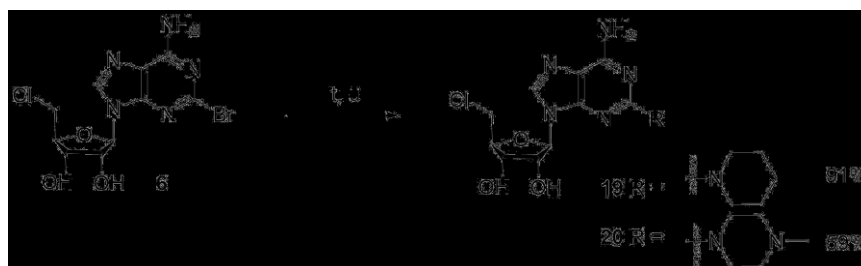
2.1.4. Introducing nucleophiles at the C-2 position

2.1.4.1 Nitrogen nucleophiles

Nitrogen/amine-based compounds are good nucleophiles so substituents incorporating an amine group were utilised first to introduce substituents at the C-2 position of the adenine ring. In previous successful bump-and-hole approaches used for kinases, researchers have incorporated a phenyl group to the adenine ring.¹⁰³ Therefore, in this approach, we wanted to keep the size of the substituents similar and thus the nitrogen substituents piperidine and 1-methylpiperazine were chosen.

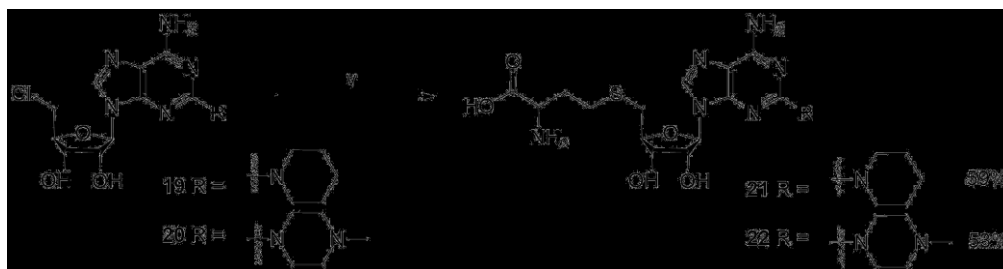
To install the piperidine substituent, **6** was reacted with 2 eq. of piperidine in H₂O at 100 °C using similar conditions to the previous thiolactone reactions to yield the 2-piperidine adenosine **19** in 91% yield (Scheme 12). Two equivalents of piperidine were used since one equivalent is required to act as a base in addition to the other equivalent being introduced at the C-2 position. Higher equivalents were not used to prevent double substitution from occurring.

To confirm the success of this route and to increase the library of analogues another nitrogen nucleophile was employed. Consequently, the nucleophilic substitution reaction was repeated utilising 1-methylpiperazine as the substrate, which gave the resulting product **20** in 59% yield (Scheme 12). These results confirm that nitrogen nucleophiles are suitable to yield adenosine analogues modified at the C-2 position without reacting at the C-5' position. The incorporation of both groups was confirmed by NMR and MS with the MS data revealing no bromine splitting in the product providing evidence that substitution has successfully occurred.



Scheme 12 – Introducing nitrogen substituents at the C-2 position. t) piperidine (2 eq.), H₂O, 100 °C; k) 1-methylpiperazine (2 eq.), H₂O, 100 °C.

Intermediate **19** was then treated with L-homocysteine thiolactone to introduce the homocysteine side chain. The original conditions using 2 eq. of thiolactone and 6 eq. of NaOH were carried out with the piperidine adenosine **19** yielding the piperidine-SAH analogue **21** in 59% yield (Scheme 13). It was decided not to purify this compound by reverse phase chromatography because no major impurities were observed from this reaction. Instead, the product was neutralized with 2M HCl that resulted in a fine precipitate of product.

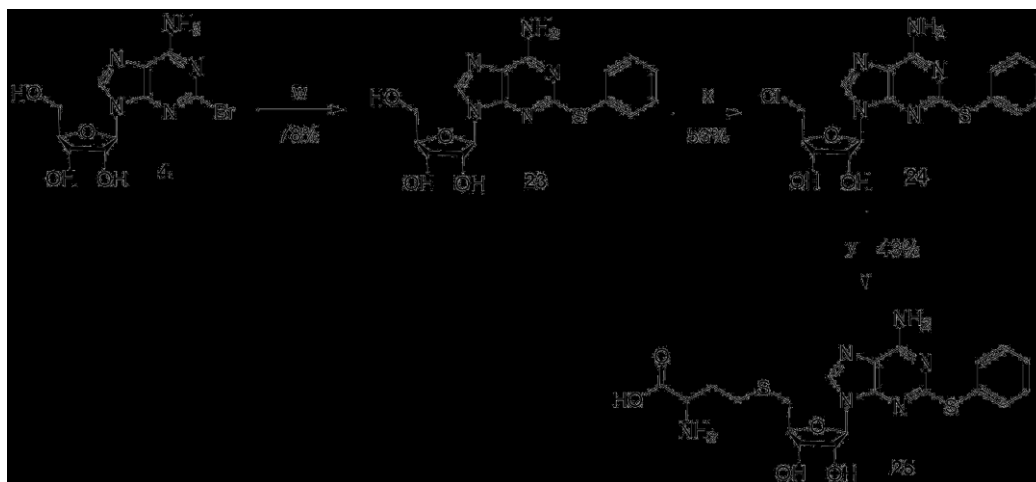


Scheme 13 – Synthesis of first SAH analogues modified with nitrogen substituents at the C-2 position.
v) L-homocysteine thiolactone hydrochloride (2 eq.), NaOH (6 eq.), KI, H₂O, 100 °C.

The same reaction was repeated with the 1-methylpiperazine adenosine derivate **20** to give the second SAH analogue **22** in 53% yield (Scheme 13). The last step to produce the desired SAM analogues is the methylation of the sulphur atom, which is discussed in Section 2.1.6.

2.1.4.2 Sulphur nucleophiles

Other commonly used nucleophiles are sulphur-based compounds which have already been proven viable by the preferential substitution at the C-2 position by the homocysteine side chain. Therefore, attempts were made to install a thiophenyl substituent by reacting sodium thiophenolate (NaSPh) with the chlorinated adenosine **6**. Unlike the addition of the nitrogen substituents, double substitution occurred both at the C-2 and C-5' positions as seen with the addition of the homocysteine side chain. To overcome this, the addition was achieved by reaction of sodium thiophenolate with the triol substrate **4** to afford the substituted product **23** (Scheme 14). After the substitution, the C-5' position was then chlorinated using the same conditions described previously. The homocysteine side chain was then added to yield the third SAH analogue **25** in 43% yield (Scheme 14).

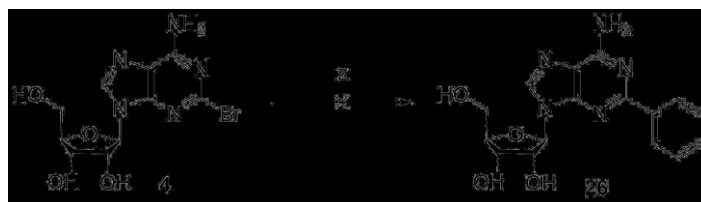


Scheme 14 – Synthesis of thiophenyl SAH analogue. w) NaSPh, H₂O, 100 °C, o/n; x) SOCl₂, py., MeCN, 0 °C to rt, 16 h; y) L-homocysteine thiolactone, NaOH, KI, H₂O, 100 °C, o/n.

2.1.5. Palladium cross-coupling to install a 2-phenyl group

Phenyl substituents have previously been installed successfully in bump-and-hole techniques used for protein kinases and thus they are a good choice here to introduce at the C-2 position of SAM. It is more challenging to utilise carbon nucleophiles to achieve formation of C-C bonds but Suzuki-Miyaura cross-coupling techniques are well established for this role. Conditions are well established for Suzuki-Miyaura cross-coupling reactions to successfully form C-C bonds and commonly use a Buchwald phosphate ligand (e.g. SPhos) in water/organic biphasic solvent systems for similar substrates used here.¹²⁷ A potential problem here is that the C-2 position of the adenine ring has lower reactivity to palladium cross-coupling reactions compared to other positions (e.g. C-6 and C-8) due to the electron-poor nature of the pyrimidine ring. Nonetheless, Buchwald conditions were employed to the triol compound **4** in an attempt to install a phenyl group (see Scheme 15). This reaction resulted in no detectable product formed by NMR or MS.

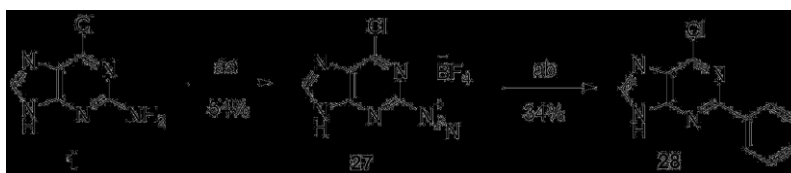
Another phosphate ligand was investigated that is reported to be more efficient than SPhos in aqueous solvent systems – TXPTS (Tris(2,4-dimethyl-5-sulfophenyl)phosphine).¹²⁸ Unfortunately, this reaction also resulted in no conversion to product.



Scheme 15 – Failed reaction employing aqueous Suzuki-Miyaura palladium cross-coupling conditions.
z) Pd(OAc)₂, Ph-B(OH)₂, SPhos, K₂CO₃, H₂O/MeCN (1:1), 80 °C, 24 h.

The lack of reaction observed could potentially arise from the lower reactivity of the C-2 position and may benefit by introducing a more reactive group at this position. Therefore, alongside the reactions in Scheme 15), conditions were found for coupling diazonium salts with boronic acids using baseless and ligandless conditions.¹²⁹ It was reported that adding phosphine ligands to these reactions lead to extensive decomposition of the diazonium salt. Furthermore, adding external base, which are usually key components to Suzuki cross-coupling, is actually detrimental. To circumvent other potential problems of having free polar hydroxyl and amine groups this coupling was attempted using the 2-amino-6-chloro-purine starting material.

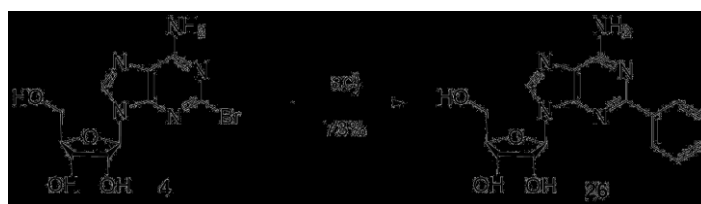
The amine group was converted to a diazonium salt which was achieved by reacting **1** with sodium nitrite and tetrafluoroboric acid to furnish the diazonium tetrafluoroborate salt **27** in 54% yield. Diazonium salts are normally unstable at room temperature but the tetrafluoroborate allows the diazonium salt to be isolated. Once isolated, **27** was subjected to the palladium cross-coupling conditions of Pd(OAc)₂ and phenylboronic acid in degassed solvent to successfully furnish 2-phenyl-6-chloroadenine (Scheme 16, **28**) in 34% yield. The choice of solvent used in this reaction is of great importance and in previous approaches either alcoholic solvents (e.g. MeOH or EtOH) or 1,4-dioxane gave the best results.^{129,130} Following the literature procedure by Sengupta *et al.*, methanol solvent was utilised first as the solvent but the starting material **27** was not soluble and thus the solvent was changed to 1,4-dioxane.



Scheme 16 – Cross-coupling on the diazonium salt of 6-chloroadenine. aa) NaNO₂, H₂O, conc. HCl, HBF₄, 0 °C, 15 min; ab) Pd(OAc)₂, Ph-B(OH)₂, 1,4-dioxane, rt, 8 h.

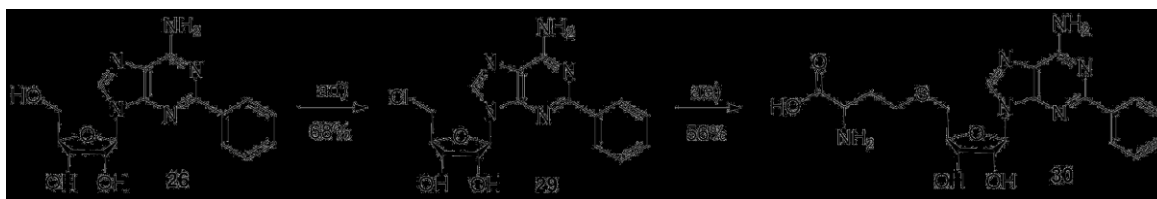
Although the cross-coupling reaction is modestly successful in this case, it will require a lot of material to bring through to the final 2-phenyl SAM analogue and thus it is an inefficient route to this analogue. It is noted that the solvent used for this reaction is not aqueous due to literature precedence of water leading to extensive decomposition in similar reactions.¹²⁹ Furthermore, the Wagner group that use similar chemistry for NAD analogues have reported that aqueous conditions are suitable for coupling with the C-6 and C-8 positions of the adenine ring but not the C-2 position.¹³¹ They did not theorise why this is the case, but, due to the nucleophilic nature of the C-2 position, it is likely that competing hydrolysis is occurring at the C-2 position.

Given these results, the reaction was revisited for the triol adenosine substrate **4** but instead using degassed 1,4-dioxane as solvent. The rest of the reagents were kept consistent using SPhos as the ligand and the reaction successfully produced the 2-phenyl cross-coupled product **26** in 73% yield (Scheme 17).



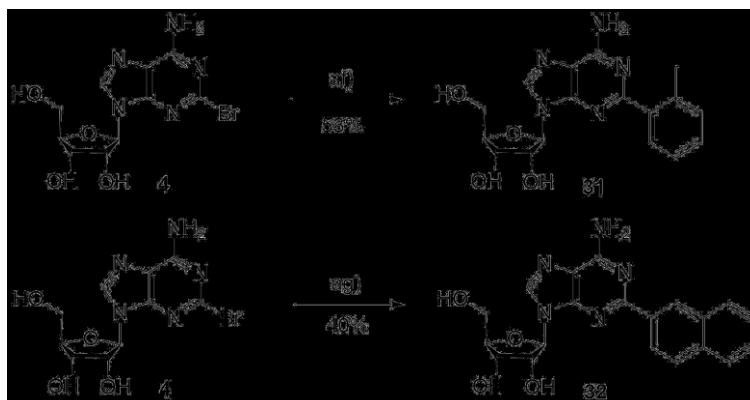
Scheme 17 – Successful addition of a phenyl group at the C-2 position using Suzuki cross-coupling conditions. ac) Pd(OAc)₂, Ph-B(OH)₂, SPhos, K₂CO₃, 1-4-dioxane, 80 °C, 48 h.

The R_f values for the starting material and product are very similar in value and thus during purification by column chromatography they were hard to separate from each other. The mixture was approximately 50:50 starting material:product by NMR, hence the reaction was pushed further to completion to overcome the separation problem. An additional equivalent of phenylboronic acid was added to the mixture and the reaction was left a further 24 hours. After 48 hours, the reaction was near completion allowing the product to be carried forward to the next step after purification by column chromatography.



Scheme 18 – Synthesis of 2-phenyl SAH. ad) SOCl_2 , py., MeCN, 0 °C to rt, 16 h; ae) L-homocysteine thiolactone, NaOH, KI, $\text{H}_2\text{O}:\text{DMF}$ (2:1), 100 °C, o/n.

At this stage, the C-5' position was chlorinated as described previously using thionyl chloride to furnish the chlorinated product **29** (68% yield, Scheme 18). The final step was the addition of the homocysteine side chain. Previously, the addition was achieved using H_2O as the solvent. However, in this case, the 2-phenyl derivative **29** was not soluble even at 100 °C under these conditions. To overcome the solubility problem, DMF was introduced at a ratio of 2:1 ($\text{H}_2\text{O}:\text{DMF}$). All the reagents successfully dissolved in this solvent system and the 2-phenyl SAH analogue **30** was obtained in 56% yield (Scheme 18).



Scheme 19 – Suzuki cross-coupling reactions to introduce *ortho*-tolyl and 2-naphthalene groups to the C-2 position of adenine. af) $\text{Pd}(\text{OAc})_2$, σ -tolyl- $\text{B}(\text{OH})_2$, SPhos, K_2CO_3 , 1-4-dioxane, 80 °C, 48 h; ag) $\text{Pd}(\text{OAc})_2$, 2-naphthyl- $\text{B}(\text{OH})_2$, SPhos, K_2CO_3 , 1-4-dioxane, 80 °C, 48 h.

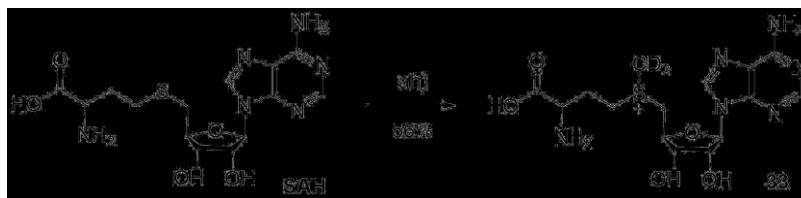
Palladium cross-coupling was also used to introduce an *ortho*-tolyl group and a 2-naphthalene group to the C-2 position with the aim to introduce more rigidity and bulk to the ‘bump’ respectively (Scheme 19). With more rigidity the group may lie more efficiently in the natural pocket of SETD7. Moreover, the substituent may need to be more bulky to get better selectivity for SETD7 over the other HKMTs. Due to time constraints and the need to test the analogues already synthesised these intermediates were not taken forward to final products. However, these analogues can be revisited in the future if necessary and they highlight the power of palladium cross-coupling to introduce a variety of groups at the C-2 position.

2.1.6. Methylation reactions

2.1.6.1. Methylation of SAH

The final step to generate the planned SAM analogues was to incorporate an isotopically labelled methyl group (CD_3 group) on the sulphur atom. To develop methylation conditions before applying it to the synthesised SAH analogues, commercial SAH was utilised to produce 'heavy' labelled 'wild-type' SAM. This reaction also produced a 'heavy' CD_3 -labelled control molecule to use in biological assays as a comparison to the analogues.

Li *et al.* have reported the successful methylation of SAH using methyl iodide (MeI) as the methylation agent.¹⁰⁶ However, the conditions have long reaction times (4 days) and produce low yields of product, likely due to methyl iodide being a weak methylating agent. Therefore, 'heavy' methyl triflate (MeOTf-d_3) was investigated as a stronger methylating agent. To produce the strong methylating agent, 'heavy' methyl triflate was formed *in situ* by the reaction of iodomethane- d_3 and silver triflate (AgOTf). The reaction is observed easily by the precipitation of bright yellow silver iodide salts that drives the reaction forward. After an hour, SAH was added to the mixture and left for 12 h at rt (Scheme 20).

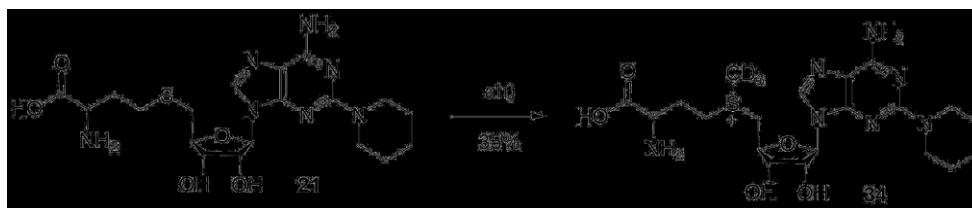


Scheme 20 – CD_3 methylation of SAH. ah) AgOTf , MeI-d_3 , HCOOH , rt, 12 h.

In the first iteration, the crude product was purified on an Amberlite XAD-4 resin column. TLC proved not to be a useful analytical technique due to the polarity of the product. Therefore, every fraction collected from the column was identified by NMR leading to an overall yield of 37% for the product **33**. The product CD_3 -SAM **33** was confirmed by LC-MS/MS as a pure compound of m/z value of 402.17. This method of purification is time consuming and resulted in a low yield of product. Therefore, in subsequent reactions the crude product is purified by reverse phase HPLC to generate CD_3 -SAM in an improved yield of 56% (Scheme 20). The synthesised CD_3 -SAM is epimeric at the sulphur atom and therefore the product is approximately a 1:1 mixture of diastereoisomers.

2.1.6.2. Methylation of SAH analogues

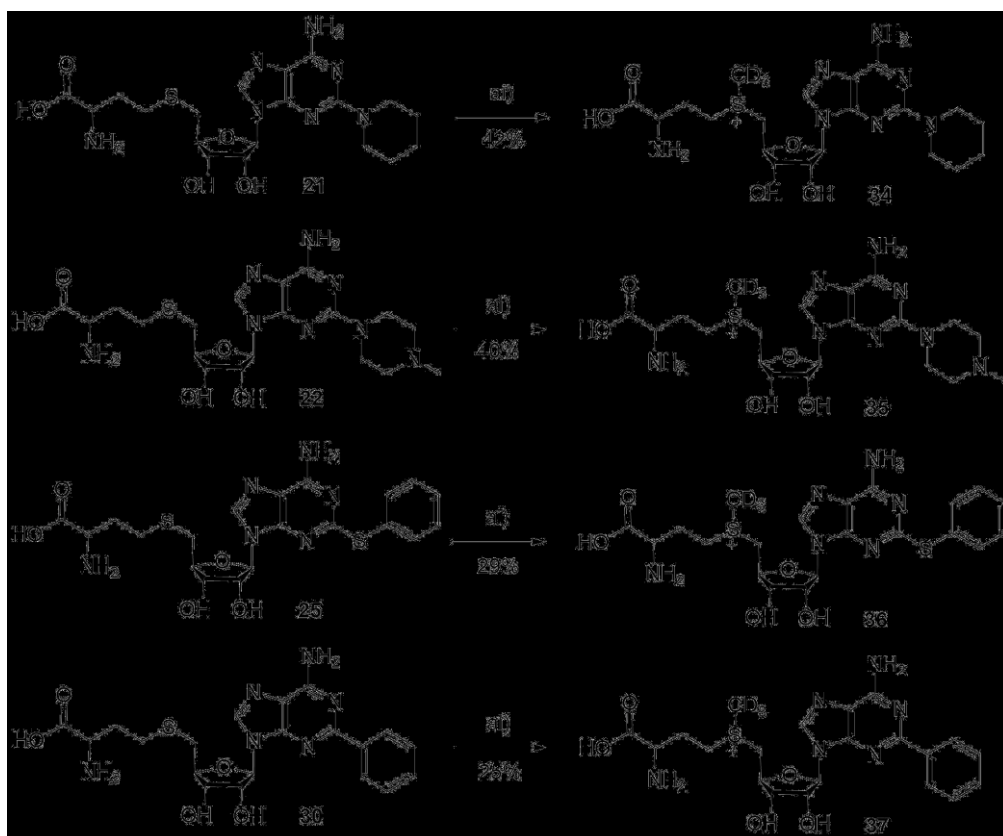
The same conditions described in Scheme 20 were then applied to the four SAH analogues (synthesis described in Section 2.1.4.). After an hour of *in situ* formation of MeOTf-d₃, piperidine-SAH **21** was added to the mixture and left for 12 h. The crude product was then purified by HPLC to generate the first SAM analogue, piperidine SAM **23** in 35% yield (Scheme 21). The product was confirmed by LC-MS as a pure compound of *m/z* value of 485.26.



Scheme 21 – Methylation of piperidine SAH (30). ah) MeI-d₃, AgOTf, HCOOH 1 h then 30, 12 h.

This reaction was repeated with the 1-methylpiperazine-SAH **22** using the same methylation conditions. LC-MS confirmed the successful synthesis of the second ‘heavy’ labelled SAM analogue **24** with *m/z* value of 500.44. Unfortunately, alongside this there was significant degradation thought to be from the silver iodide salts generated in the *in situ* formation of MeOTf-d₃ or from unreacted AgOTf. Therefore, another procedure by Kuethe was employed that uses CCl₄ as the solvent for the *in situ* formation of MeOTf-d₃. This allows filtering out of the silver salts using Celite for further reaction with the substrate (e.g. SAH analogue in this case) in HCOOH.¹³² For the small scale reactions utilised here, syringe pads were employed instead of a Celite pad. This alternative procedure significantly decreased the degradation formed in this reaction and thus these conditions were employed for all subsequent methylation reactions on the SAH analogues.

The successful methylation reactions using the revised conditions are shown for all the four final CD₃ labelled SAM analogues in Scheme 22. The crude products were all purified by reverse phase HPLC to ≥95% purity for use in biological assays. The use of the commercially purchased reagent MeOTf-d₃ instead of the *in situ* formation of the reagent could lead to a cleaner reaction and thus higher yield in the above methylation reactions. However, due to the toxicity and cost of this reagent this procedure was not pursued. The yields are satisfactory to yield the SAM analogues in high purity and so testing was commenced in various biological assays which are discussed in Chapter 3.



Scheme 22 – Methylation of all SAH analogues to furnish the four CD₃-labelled SAM analogues that will be tested in biological assays. ai) MeI-d₃, AgOTf, CCl₄, 1h then SAH analogue (21, 22, 25, 30), HCOOH, 12 h.

3. Biological assays for identifying HKMT targets using SAM analogues

3.1. Mass spectrometry proteomics workflow for analysing histone PTMs

The next step after the small library of 'heavy' labelled analogues had been synthesised was to test them in a biological setting. The performed biological assays vary in complexity from *in vitro* assays to cellular assays. However, the majority of these assays examine modified histone peptides via LC-MS/MS. The workflow employed has been previously reported and optimised for histone analysis and is briefly described here (see Figure 25).¹³³

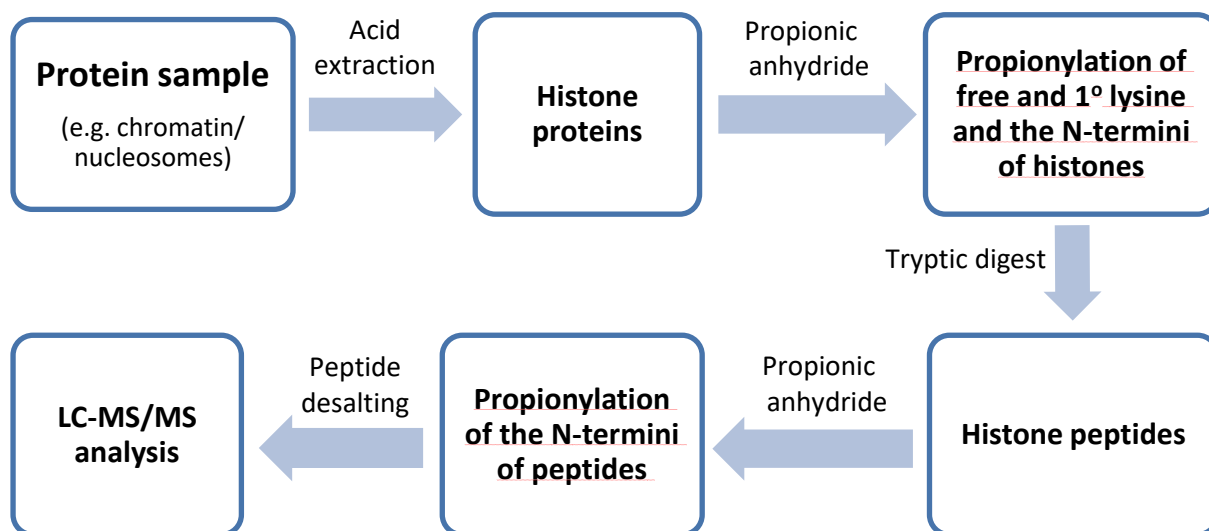


Figure 25 – Standard histone workflow for analysis of histone peptides by LC-MS/MS.

First, the histone proteins are extracted from the sample using acid (histones are solubilised in 0.6 N H₂SO₄ and then precipitated out using 100% trichloroacetic acid (TCA)). Then, the lysine residues of the protein sample are chemically modified with propionic anhydride and followed by digestion with trypsin. The final step after digestion is to propionylate the *N*-termini of the peptides. This method results in highly reproducible peptide fragments allowing for the comparative analysis of histone modifications in different samples tested over a variety of different conditions. Quantifications are normalised per peptide so comparison between two samples are done on a peptide-by-peptide basis. Chemical modification by propionylation has two purposes:

- The lysine residues modified by the propionyl group are blocked hence enzymatic cleavage only occurs on the C-termini of arginine residues. Arg-C is another enzyme that works in the same way, however this method benefits from the efficiency and robustness of trypsin.
- The second round of propionylation neutralises the charge of the N-termini of the peptide fragments. This has two benefits: i) the peptides will be less hydrophilic and therefore are resolved easier by RP-HPLC; ii) the peptides will produce doubly and triply charge fragments in electrospray ionisation allowing straightforward analysis of MS/MS spectra.

The last step is to desalt, in preparation for LC-MS/MS analysis, using stage tips (process described in the Methods section).

3.2. *In vitro* assays using nucleosome substrates

The first step to validate the synthesised SAM analogues (synthesis discussed in Section 2) was to test them in a simple *in vitro* assay using chromatin substrates (e.g. mononucleosomes extracted from HeLa cells). Mononucleosomes are used over more simple substrates like isolated histones in order to represent the endogenous environment more accurately. Furthermore, the specificity of the enzyme being studied is dependent on the substrate used. For example, NSD2 is a robust H3K36 dimethyltransferase on nucleosomes but has minor activity on free histones.¹³⁴ Diego Alonso-Martinez, a PhD student in Chemical Engineering, performed the assays described in this section using the analogues I synthesised. Diego Alonso-Martinez produced all the biological data described in this section with the data figures produced by myself from the resulting data.

To ensure the 'heavy' methyl SAM cofactors are compatible with the HKMT of interest, SETD7, initial *in vitro* assays were performed with CD₃-SAM as a control. Mass spectrometry (MS) was used as the primary readout for validating catalytic activity after performing the standard histone workflow described in Section 3.1. A variety of different concentrations of CD₃-SAM and SETD7 were investigated for assay optimisation. The optimised conditions were found to be 200 µM CD₃-SAM and 500 nM of SETD7. This led to successful labelling on H3K4, the reported histone target of SETD7.^{29–33} The MS spectra in Figure 26 shows selective 'heavy' labelling of H3K4 (left spectra) and no labelling of other H3 or H4 residues were seen (example on right given for the neighbouring methylation mark H3K9me2K14ac). The blue peaks highlight the peptide with the endogenous methylation already installed on the lysine site of interest and the red peaks highlight the installation of the CD₃ 'heavy' methylation onto the lysine site. For example, the blue MS peak at 830.460

(Figure 26, left spectra) is the peptide TKQTAR with endogenous methylation on K4 and the red peak at 833.480 is the same peptide but with the installation of the CD₃ group on K4. Comparison to the methylation site K9 is chosen because if SETD7 was promiscuous then it would likely label on proximal lysine sites, which K9 is. This observed selective labelling confirmed that the assay is working and producing the expected labelling on H3K4.

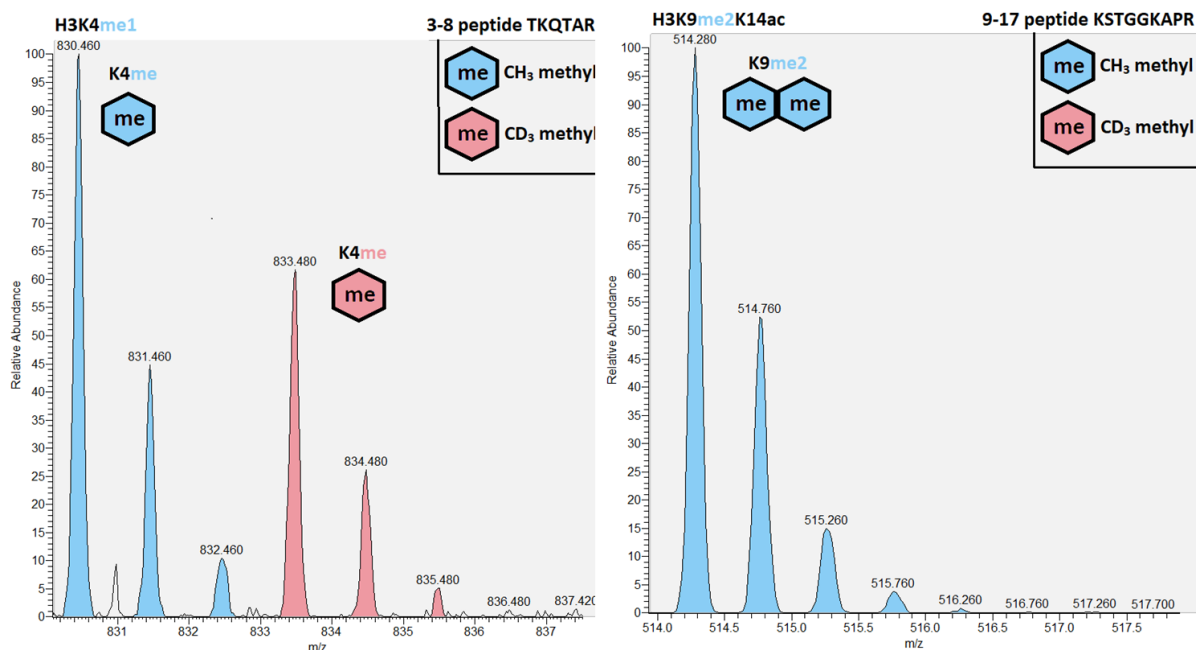


Figure 26 – MS spectra showing CD₃ incorporation (red peaks) on the methylation mark H3K4me1 (left spectrum) and no CD₃ incorporation on the methylation mark H3K9me2K14ac (right spectrum) following incubation of mononucleosomes with 200 μ M CD₃-SAM and 500 nM commercial SETD7 for 1 h.

The *in vitro* assays were repeated with the small library of synthesised SAM analogues to observe whether SETD7 can also utilise the ‘bumped’ analogues. With all the C-2 modified analogues, ‘heavy’ CD₃ labelling was detected on H3K4. The relative ratio of CD₃ incorporation on H3K4 with respect to the endogenous methylation is shown for all analogues in Figure 27 (i.e. the percentage of the red peaks to the blue peaks for the methylation mark H3K4 shown in Figure 26). The 2-piperidine and 2-methylpiperazine analogues show similar incorporation of the CD₃ group to the ‘endogenous’ CD₃-SAM, suggesting that they have similar turnover rates to the isotopically labelled endogenous cofactor under the assay conditions. In contrast, the 2-thiophenyl and the 2-phenyl analogues only show approximately 50% of the CD₃ incorporation compared to CD₃-SAM. These results confirm that the ‘bumped’ analogues can be utilised by SETD7.

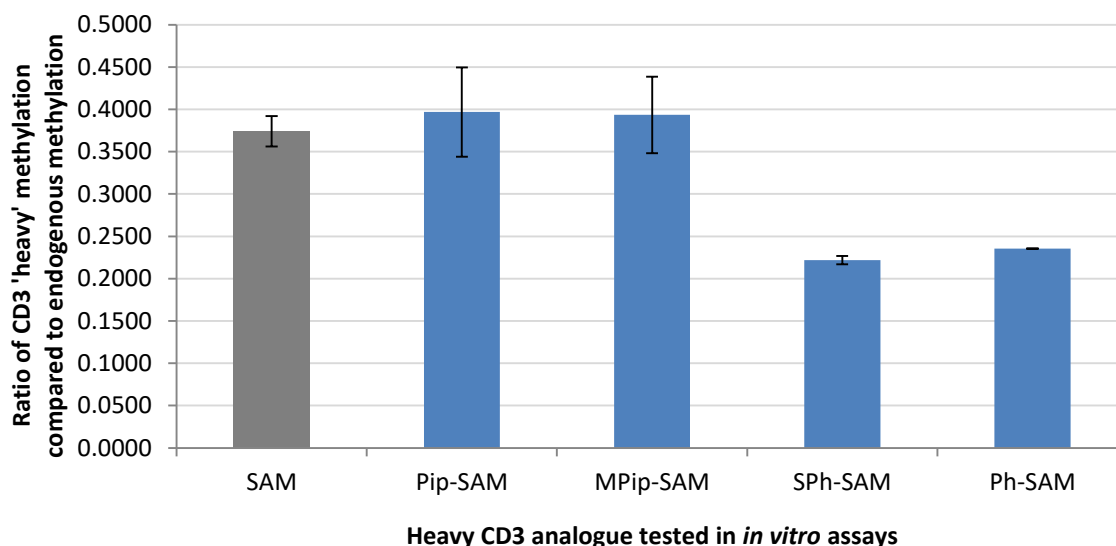


Figure 27 – The ratio of CD₃ monomethylation of H3K4 catalysed by SETD7 on mononucleosomes with the small library of 'heavy' cofactor analogues. The CD₃ methylation present from the experiments with 200 μ M CD₃-SAM is shown on the far left for comparison with the 2-modified analogues. The error bars show one standard deviation (SD). (Pip-SAM = 2-piperidine-SAM, MPip-SAM = 2-methylpiperazine-SAM, TPhe-SAM = 2-thiophenyl-SAM, Phe-SAM = 2-phenyl-SAM).

After confirming that SETD7 can utilise both endogenous SAM and the SAM analogues, the analogues selectivity was tested with G9a. G9a was the HKMT of choice to compare the selectivity of the analogues with SETD7 because it is a well characterised HKMT with a well-established histone target of H3K9.^{108,135,136} Moreover, as discussed in Section 1.6, G9a does not contain a structural pocket in its SAM binding pocket that SETD7 exhibits. The absence of this structural pocket in G9a should hinder the binding of the SAM analogues modified at the C-2 position and thus show no labelling in the *in vitro* assays utilising them. Therefore, G9a is a suitable comparator to SETD7 for selectivity studies of the small library of analogues synthesised. Initially, *in vitro* assays were performed on G9a with CD₃-SAM to validate the assay and subsequently the C-2 modified analogues. If the analogues are selective for SETD7, we should not observe any CD₃ incorporation on the target site of G9a (H3K9).

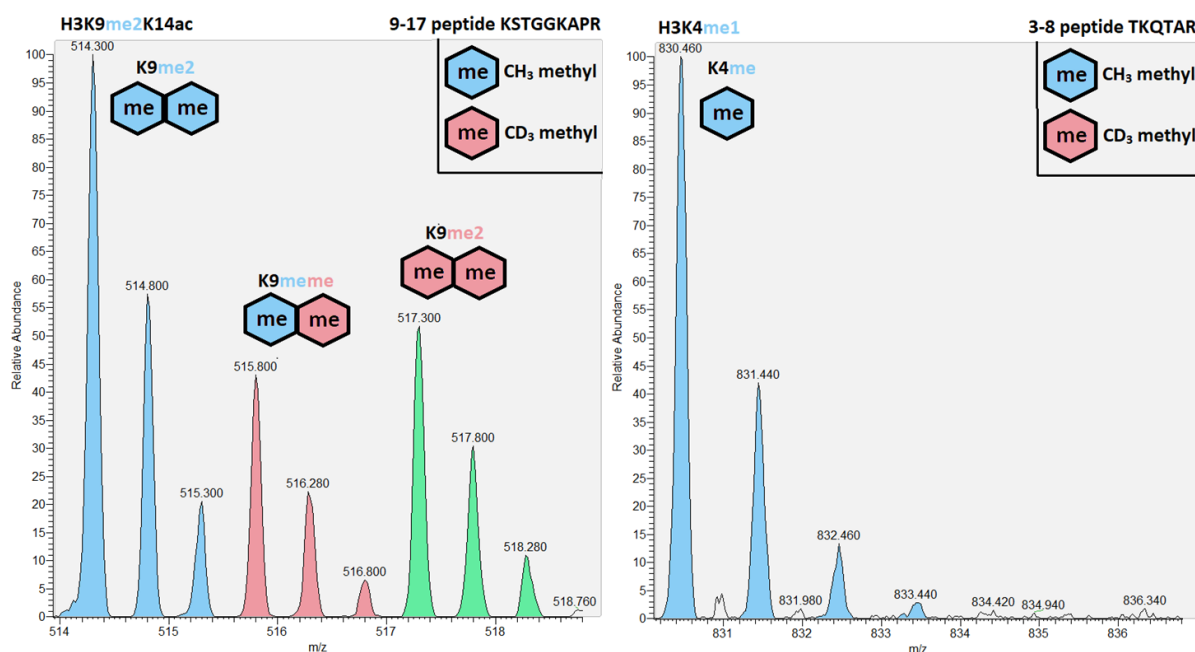


Figure 28 – MS spectra showing CD₃ incorporation (red and green peaks) for the methylation mark H3K9me2K14ac (left spectrum) and no CD₃ incorporation on the methylation mark H3K4me1 (right spectrum) following incubation of mononucleosomes with 1 mM CD₃-SAM and 1 μ M commercial G9a for 1 h.

Following the same methods as for the *in vitro* assays with SETD7, various concentrations of SAM and G9a were tested to optimise the assay for labelling on H3K9. The optimised conditions for the G9a assay were 1 mM of SAM and 1 μ M of G9a which resulted in successful CD₃ incorporation on G9a's target of H3K9. No labelling was observed for the rest of the H3/H4 lysine sites. The spectra in Figure 28 shows an example of this successful CD₃ incorporation for the methylation mark H3K9me2K14ac and an example of no incorporation on the methylation mark H3K4 as a comparison.

With an optimised G9a assay, the next step was to repeat the experiments with the C-2 modified SAM analogues to observe if they are selective for SETD7 over G9a. The conditions were kept identical to the CD₃-SAM assays and the results are shown in Figure 29. The graph shows the ratio of methylated peptides for the SETD7 assays (blue bars) and the G9a assays (pink bars) for each analogue compared to CD₃-SAM. The data used for the SETD7 assays is identical to the data shown in Figure 27 and used here as a comparison to the G9a assays. G9a can both mono- and di-methylate H3K9 and therefore both of these methylation marks are used as the readout of the assay. G9a can trimethylate H3K9 *in vitro*, but this methylation state has not been evidenced *in vivo* and thus will not be suitable when translating these results into cellular environments.¹³⁵ As it can be seen in Figure 29, there is still methylation on H3K9 when the analogues are utilised, particularly for the 2-thiophenyl analogue. However, for the 2-piperidine and 2-methylpiperazine analogues there is

significant selectivity seen for SETD7 over G9a. The selectivity is shown by a significant drop in methylation observed for G9a's target compared to SETD7's and therefore provides evidence for selectivity of these C-2 modified analogues for SETD7. The most promising analogue from these assays is the 2-piperidine analogue, which exhibits the best selectivity. Therefore, this analogue was further analysed.

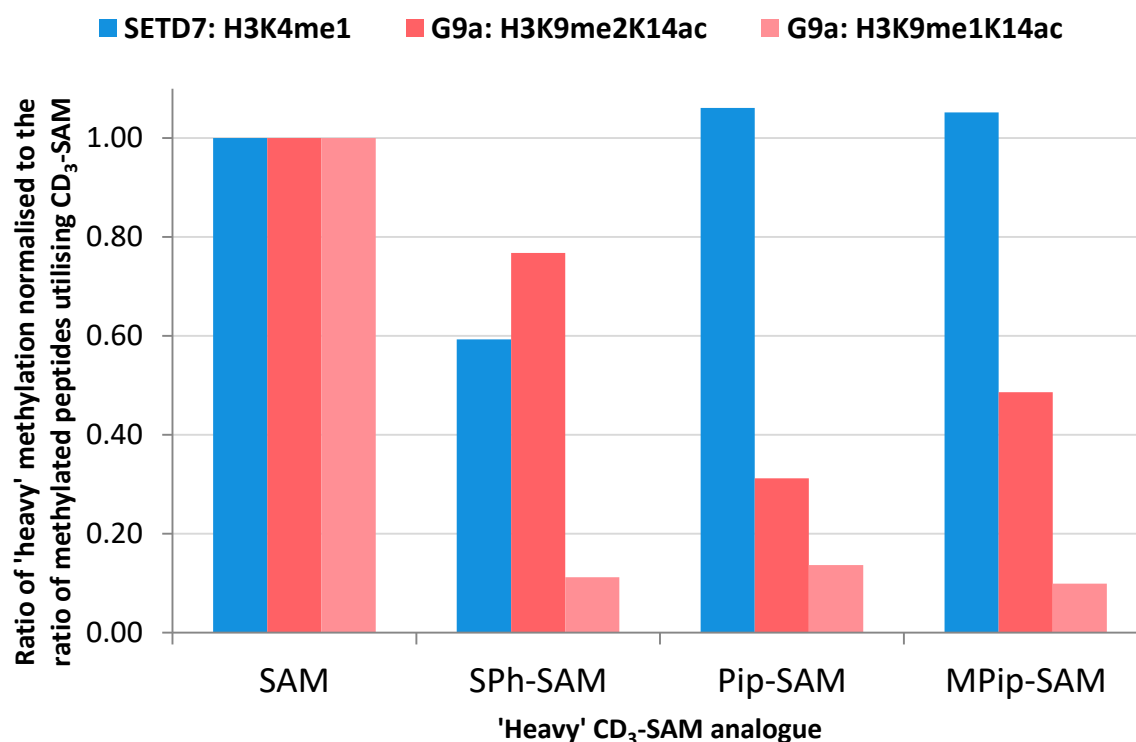


Figure 29 – Comparison of the CD₃ incorporation obtained when mononucleosomes were incubated with the C-2 modified SAM analogues and SETD7 (blue bars) vs G9a (pink bars). The CD₃ methylation present from the experiments with CD₃-SAM is shown on the far left for comparison with the 2-modified analogues. The SETD7 experiments were repeated in triplicate and the bars in this figure represent the mean average across the three repeats. However, only one experiment is shown that yielded usable results for G9a. This was due to repeat experiments showing lack of activity for the obtained commercial G9a. (SPh-SAM = 2-thiophenyl-SAM, Pip-SAM = 2-piperidine-SAM, MPip-SAM = 2-methylpiperazine-SAM).

Unfortunately, the G9a experiments have only been successfully performed once due to the lack of activity from repeat stocks of the commercial G9a provided from various suppliers. Therefore, these experiments will be repeated in future work to ensure the same selectivity is observed.

3.3. Cellular assays

Alongside testing the C-2 modified SAM analogues in the *in vitro* assays (Section 3.2), cellular assays were performed. Henceforth, I have generated all the data that is described, unless otherwise stated. Initially, the cellular assays were performed by incubating different cell lines with CD₃-SAM to optimise the assays and then the C-2 modified SAM analogues synthesised in Section 2 were tested.

In vitro assays are a quick and simple method to confirm that the enzyme of interest can utilise the cofactor analogues; however, they lack the complexity of a cellular environment. The specificity of HKMTs is dependent on the cellular environment, meaning that even though a target could be discovered *in vitro*, this may not translate to *in vivo* or vice versa. HKMTs also work in a context dependent manner and only under certain cellular conditions will methylation occur – e.g. methylation of Reptin by G9a only occurs under hypoxic conditions.⁷⁷ Furthermore, in the cell these enzymes commonly exist in multimeric complexes, which help target the HKMT to a specific site and/or stabilise or increase the activity of them.¹³⁷ Therefore, these endogenous components need to be present to get an accurate analysis of the HKMTs lysine targets. An example of this is the evidence of the multimeric H3K9 methyltransferase complex that is formed of the subsets of each of the four main H3K9 HKMTs (G9a, GLP, SETDB1 and Suv39h1).⁷⁸ In Suv39 or G9a null cells, the other HKMTs in the complex (GLP and SETDB1) are destabilised, highlighting the importance to have all the components of the multimeric complex present. On top of histone targets, HKMTs also commonly methylate non-histone proteins, thus a cellular environment is essential to investigate these proteins. Investigation into non-histone targets of SETD7 is discussed further in Chapters 3.6 and 3.7.

Endogenous SAM is reported to have low cellular permeability due to poor passive transport through the cellular membrane.¹⁰¹ Despite this, in initial cell assays, the incubation of HEK-293T cell line with ‘heavy’ CD₃-SAM resulted in labelling observed across all H3 lysine residues, suggesting reasonable permeability. The optimisation of this approach is therefore discussed herein.

3.3.1. Assessment of cell permeability of SAM and SAM analogues

To validate that the ‘heavy’ labelled SAM analogue (CD₃-SAM) can pass through the cellular membrane, HEK-293T cells were incubated with CD₃-SAM for 20 h at an initial concentration of 200 µM. The concentration of 200 µM was chosen firstly because it corresponds with the concentration of methionine that is included in the media. Also, intracellular SAM concentrations would be 50 – 100 µM and it has been shown in a previous study by Bontemps *et al.* that concentrations have to be

200 μM or over to induce SAM uptake by the cell.¹³⁸ Moreover, whilst the concentration of SAM differs between different cell types and stage of the cell cycle, 200 μM is consistent with the *in vitro* studies performed (Section 3.2). After incubation for 20 h and subjecting the resulting cell pellet to the standard histone workflow, the results showed successful incorporation of the CD_3 label on all lysine sites of H3. An example of this incorporation can be seen in Figure 30 which shows CD_3 monomethylation on K4 (left) and dimethylation on K9 of H3 (right) (the target sites of SETD7 and G9a respectively). This data indicates that CD_3 -SAM can be utilised effectively to label lysine sites on histones in a cellular environment. For these initial assays and for all subsequent assays, the SAM analogues were added while the cells were still dividing. This timeframe will result in the greatest incorporation of the CD_3 label because the epigenetic marks will be incorporated onto new histones during mitosis.

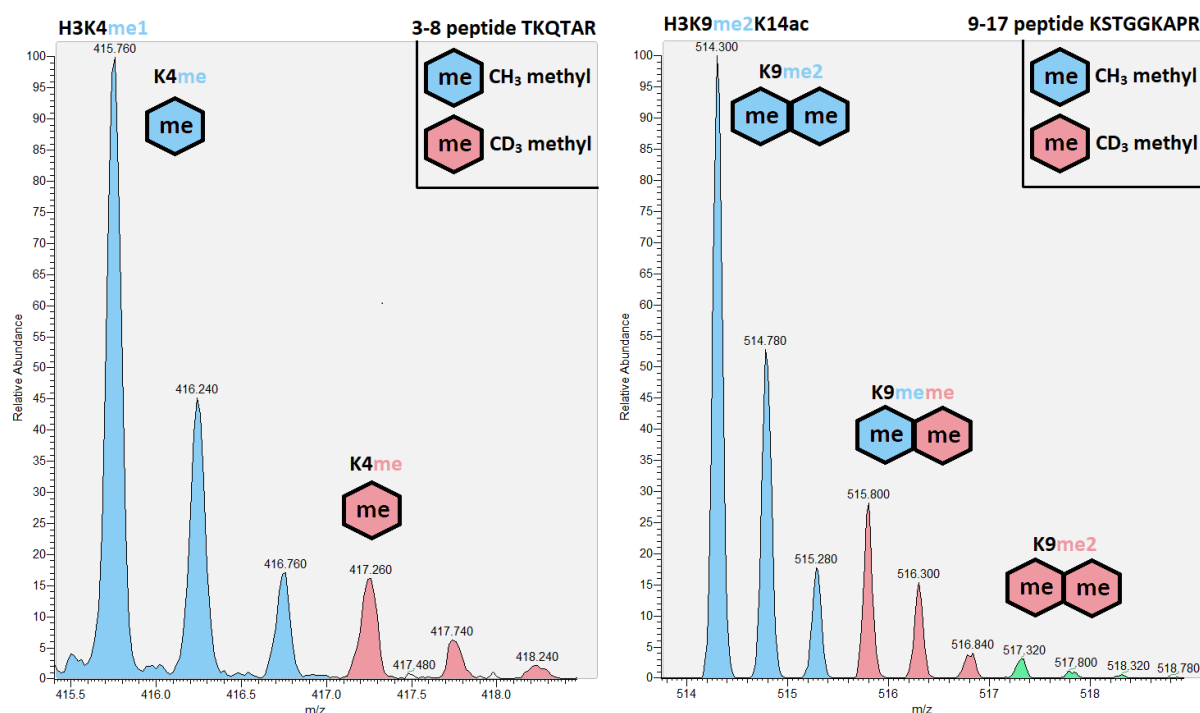


Figure 30 – MS spectra showing CD_3 incorporation (red and green peaks) for the methylation marks H3K4me1 (left spectrum) and H3K9me3K14ac (right spectrum) following incubation of HEK-293T cells with 200 μM CD_3 -SAM for 20 h.

The next step was to employ a range of different concentrations of CD_3 -SAM keeping the time length of 20 h constant. These assays were performed to investigate the optimum concentration required to get sufficient labelling in cells without affecting cell viability. Figure 31 shows the percentage ‘heavy’ labelling observed at various methylation marks and how this percentage changes with

increasing concentration. The percentage reported is the amount of CD₃ incorporation relative to the amount of endogenous CH₃ methylation. As expected, the percentage of CD₃ methylation increases with increased concentration. However, at 500 µM there was a significant increase in cells that became detached from the plate over time. Therefore, for future experiments 200 µM was chosen to circumvent this problem.

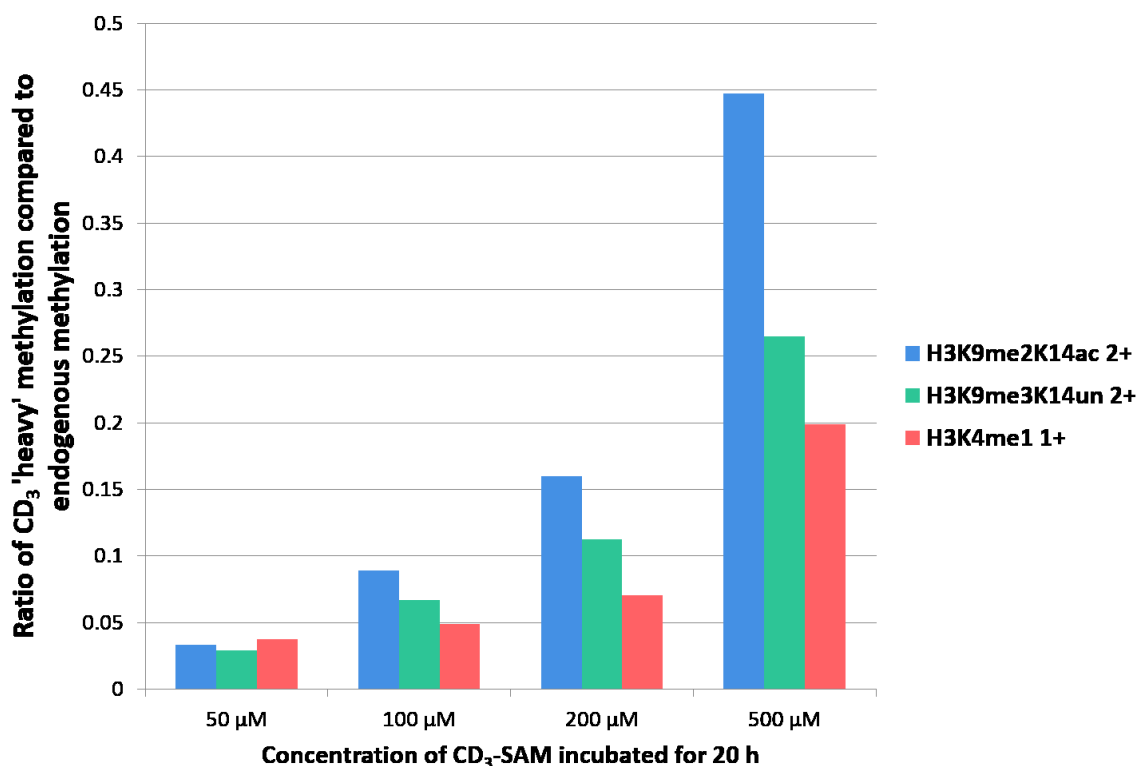


Figure 31 – The ratio of CD₃ methylation obtained on K9 and K4 methylation sites following incubation of HEK-293T cells assays with varying concentrations of CD₃-SAM for 20 h. NB. n=1 (proof of concept).

Next, the change of labelling over time was studied. Five time points were chosen – 2, 6, 12, 20 and 48 h. For the 48 h time point, to prevent the pool of available SAM from being limiting, fresh CD₃-SAM and media were supplied at 24 h. The graph showing the ratio of ‘heavy’ labelling at each time point is shown in Figure 32. At 2 and 6 h time points there is very little ‘heavy’ labelling observed on the H3 lysine sites and thus are not suitable time points for these particular cell assays. At 12 and 20 h, the labelling becomes observable and at 20 h the labelling ranges from 10–20%. The percentage at 20 h is still low but it is sufficient for initial assays. Even though at 48 h there is a higher percentage of labelling observed, 20 h was selected as a suitable time length for subsequent assays in order to keep any potential degradation of the SAM analogues in the assay to a minimum.

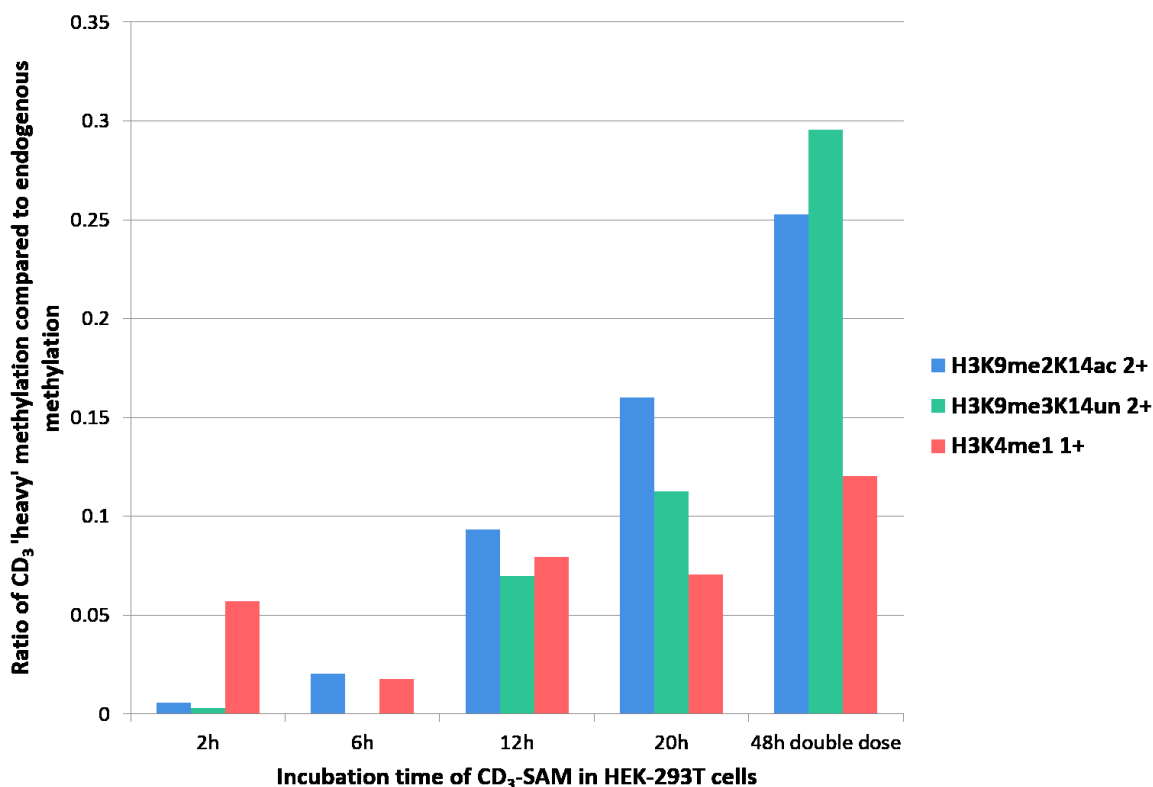


Figure 32 – The ratio of CD₃ methylation obtained on K9 and K4 methylation sites following incubation of HEK-293T cells assays with 200 μ M CD₃-SAM for varying time lengths. NB. n=1 (proof of concept).

The optimal assay conditions (20 h, 200 μ M) were repeated with the four SAM analogues modified at the C-2 position. Figure 33 shows the ratio of 'heavy' methylation at various methylation marks for all four SAM analogues. The labelling observed for CD₃-SAM is included for comparison. This data shows that for all of the four analogues, very little or no labelling is apparent on any of the H3 lysine sites compared to the 10–20% shown for the endogenous 'heavy' labelled CD₃-SAM. Should there be selectivity with SETD7, we would expect to see preferential labelling on H3K4, as was observed in the *in vitro* assays discussed in Section 3.1. However, the data suggests that the SAM analogues are not being utilised in the cellular assays for any of the histone lysine methyltransferases that methylate H3. There appears to be selective labelling for H3K4 monomethylation over the other methylation marks when phenyl-CD₃-SAM is utilised, although if you take the error (standard deviation) into account this result is not significant. A reduced level of 'heavy' labelling is expected to occur on the methylation mark H3K4 when utilising the analogues versus CD₃-SAM. There are multiple HKMTs (including SETD7) that can methylate at H3K4, which can all utilise the CD₃-SAM cofactor. However, only SETD7 should be able to utilise the 'heavy' labelled SAM analogues and therefore the 'heavy' labelling at H3K4 resulting from these assays will inherently be reduced. However, the level of

labelling obtained in these assays is insufficient to draw any conclusions around whether these analogues are specific to SETD7.

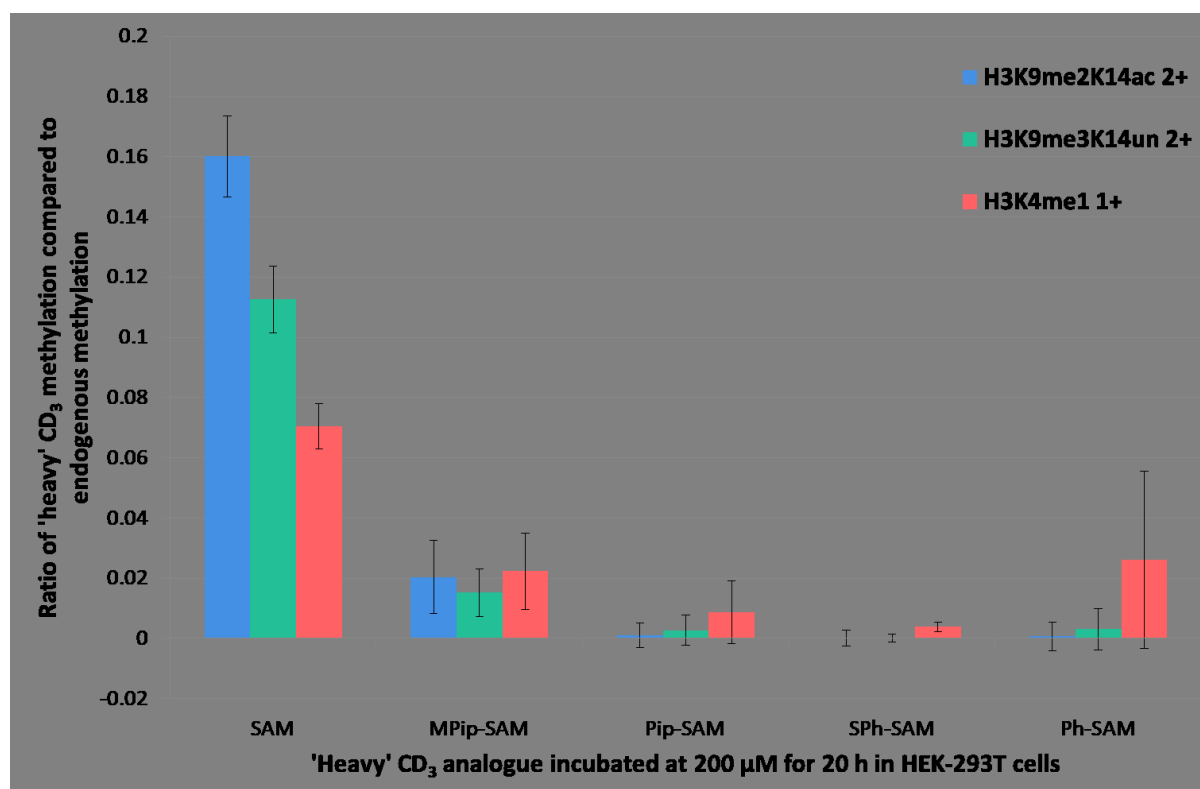


Figure 33 – The ratio of CD₃ methylation obtained on K9 and K4 methylation sites following incubation of HEK-293T cells assays with 200 μ M CD₃-SAM for 20 h for all the 'heavy' labelled SAM analogues. The error bars show one standard deviation (SD). (Pip-SAM = 2-piperidine-SAM, MPip-SAM = 2-methylpiperazine-SAM, SPh-SAM = 2-thiophenyl-SAM, Phe-SAM = 2-phenyl-SAM).

The most likely explanations for the results from these initial cell assays are as follows:

- The SAM analogues have limited cell permeability and therefore cannot be utilised intracellularly.
- The cell line being used (HEK-293T) has a low expression of SETD7 and therefore the labelling associated with this HKMT is not seen.
- SAM analogues are significantly degrading over time.
- H3K4 is not a valid substrate for SETD7 *in vivo* and is only valid in *in vitro* conditions.

These options were considered as points for optimisation and they are discussed in further detail in the following Sections.

3.3.2. Analysis of transfection-based methods to improve SAM analogue permeability

The labelling shown in the cellular assays for the SAM analogues modified at the C-2 position was almost negligible and even the labelling from CD₃-SAM was low. From reference to the Human Protein Atlas, SETD7 is shown to not be expressed in the kidney where HEK-293T cells are derived from. As a result, transfection techniques were pursued to improve the labelling observed in these assays through improved cellular delivery of SAM and/or expression of a SETD7 plasmid. A range of different transfection conditions were tested and a summary of these is shown in Table 5.

Table 5 – A summary of all the transfection conditions that were employed in the HEK-293T cellular assays. For all assays, the SAM analogue concentration was 200 µM and incubation time of analogue was 20 h.

Assay	Transfection method	SAM analogue	Plasmid	Confluency at time of transfection	Time after transfection was performed, before the SAM analogue was added to the media
1	Calcium phosphate	CD ₃ -SAM	None	70%	24 h
2	Calcium phosphate	CD ₃ -SAM	5 µg SETD7	70%	24 h
3	Calcium phosphate	MPip-CD ₃ -SAM	5 µg SETD7	70%	24 h
4	Lipofectamine 2000	CD ₃ -SAM	1:1 SETD7:VP35	70%	24 h
5	Lipofectamine 2000	CD ₃ -SAM	3:2 SETD7:VP35	70%	24 h
6	Lipofectamine 2000	CD ₃ -SAM	3:2 SETD7:VP35	70%	10 h
7	Lipofectamine 2000	CD ₃ -SAM	3:2 SETD7:VP35	50%	24 h
8	Lipofectamine 2000	MPip-CD ₃ -SAM	3:2 SETD7:VP35	50%	24 h

Firstly, traditional calcium phosphate transfection conditions were employed. This transfection works by the introduction of plasmid DNA into cells via a co-precipitation of plasmid DNA-calcium

phosphate.¹³⁹ This co-precipitates then adheres to the cell surface and is taken up by the cell thus allowing the DNA to be introduced and the protein of interest to be expressed. The cells likely have low expression levels of SETD7 and so this technique could be used to overexpress SETD7 in HEK-293T cells and therefore improve the level of CD₃ labelling observed on H3K4. HEK-293T cells were used for these transfection experiments as they are a robust cell line known to produce high transfection efficiencies. Overexpressing SETD7 in this cell line will result in a disruption to the cellular environment but it will provide evidence whether SETD7 can label H3K4 *in vivo*.

The SETD7 plasmid DNA, containing full-length-SETD7, was prepared from a bacterial source using a Maxiprep kit (pcDNA SETD7 wt was a gift from Danny Reinberg (Addgene plasmid #24084)).³⁸ A control experiment was performed using calcium phosphate and CD₃-SAM without plasmid in an attempt to overcome the potential cell permeability issues with SAM analogues (Table 5, assay 1). If this technique improved the uptake of CD₃-SAM into cells, it would then improve the intracellular level of CD₃-SAM and thus improve the observed CD₃ labelling on the histones. Unfortunately, the CD₃ labelling did not increase with the addition of calcium phosphate. Next, 5 µg of SETD7 plasmid was added to the calcium phosphate transfection, in an attempt to increase the percentage of 'heavy' labelling observed at K4 site of H3 through overexpression of SETD7 protein. Again, this did not improve the labelling on the histones, either with CD₃-SAM or the 2-methylpiperazine SAM analogue (Table 5, assays 2 and 3).

Lipofectamine 2000 contains lipid subunits that form cationic liposomes in aqueous media that can capture DNA plasmids and subsequently fuse with the negatively charged plasma membrane of cells, thus allowing internalisation.¹⁴⁰ In this case, the SETD7 plasmid was used alongside an empty vector plasmid VP35 which was shown by Hugh Brady's group to improve transfection efficiencies.¹⁴¹ The VP35 plasmid was gifted from the Brady group for use in these experiments. Two different ratios were tested to get maximum efficacy: 1:1 and 3:2 SETD7:VP35 (Table 5, assays 4 and 5). Unfortunately, instead of increasing 'heavy' monomethylation on K4 as desired, there was decreased labelling on all lysine sites of H3. Nonetheless, the ratio of 3:2 SETD7:VP35 showed slightly better 'heavy' labelling of histones than 1:1 so this ratio was used for subsequent experiments (Figure 34).

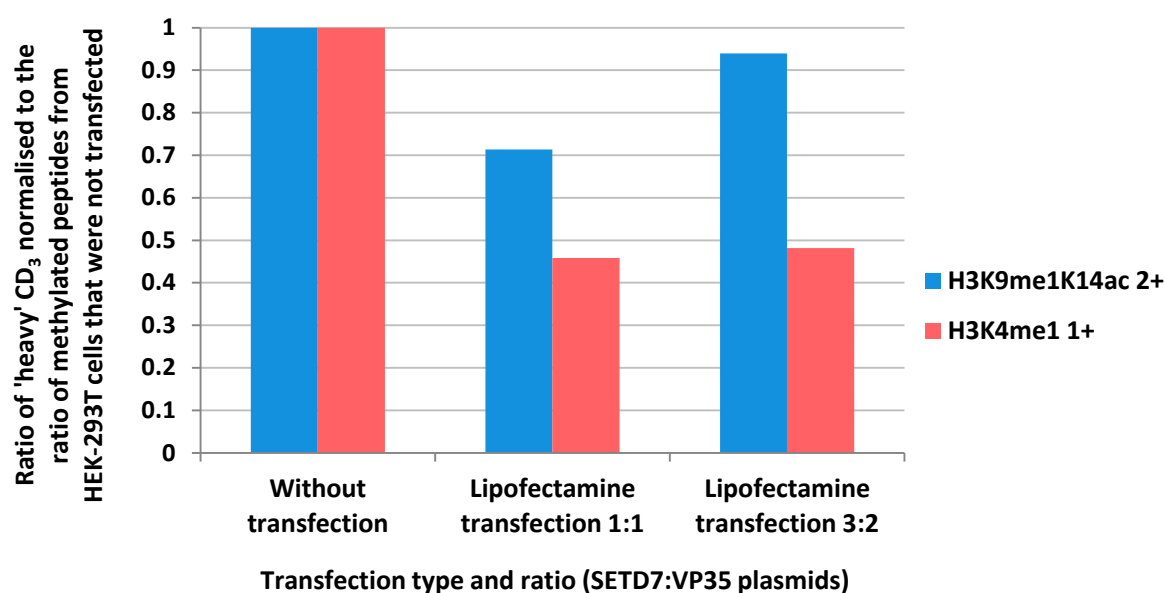


Figure 25 – The ratio of CD₃ methylation obtained on K9 and K4 methylation sites following incubation of HEK-293T cells assays with 200 μ M CD₃-SAM for 20 h for with different transfection methods. The left hand shows incubation with CD₃-SAM without transfection as a comparison. The ratio stated is the ratio of SETD7 and VP35 plasmids used for the Lipofectamine 2000 transfection. NB. n=1 (proof of concept).

The decreased labelling on H3 that was observed in these initial experiments could be due to unsuccessful Lipofectamine transfection of plasmid SETD7. Therefore, the success of this technique was investigated before any more experiments were performed. The plasmid SETD7 acquired from AddGene has a FLAG tag incorporated at the C-terminus. In order to confirm if SETD7 was being successfully expressed in the lipofectamine transfection, a FLAG immunoprecipitation was performed using ANTI-FLAG M2 Affinity Gel (Sigma Aldrich). Protein samples taken from different stages of the protocol were denatured and then resolved on an 8% SDS-PAGE gel (see Figure 35).

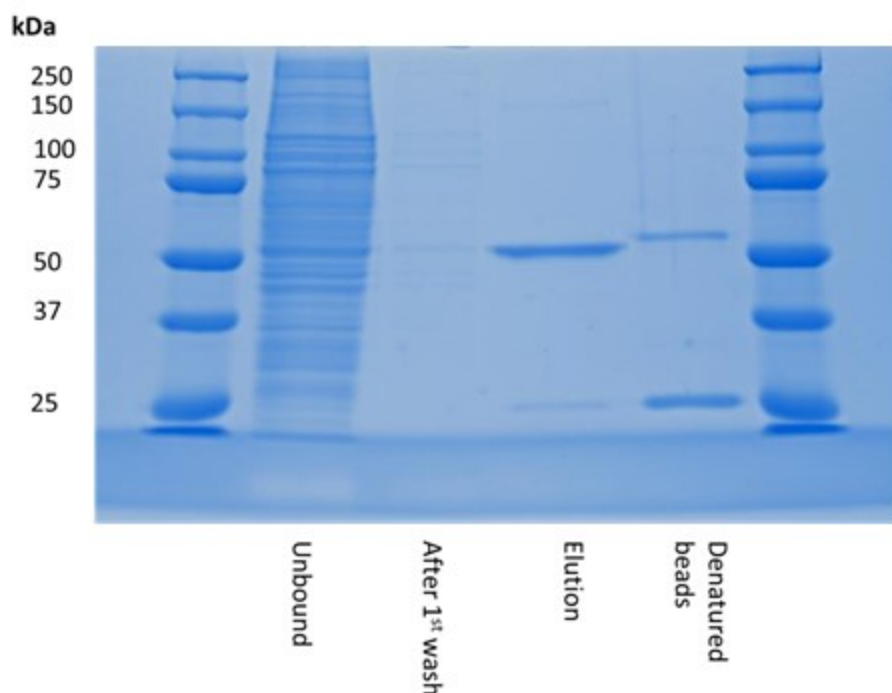


Figure 35 – SDS-PAGE of the various stages throughout the FLAG IP protocol. The SDS-PAGE was run against Precision Plus Protein™ Standard, Dual Colour (BioRad).

There is only one dominant band for the elution sample (lane 4, Figure 35) at approximately 50 kDa in the SDS-PAGE. This is the molecular weight we would expect for SETD7. There is also a faint band at approximately 25 kDa which also appears on the denatured beads at a higher intensity. These two bands were cut out and in-gel digestion was performed on both bands for identification by LC-MS/MS. An aliquot of the elution sample was also subjected to in-solution digestion for identification by LC-MS/MS. All LC-MS/MS data collected was run through SEQUEST to identify the protein relating to the band/sample collected. Both in-gel and in-solution digests for the elution sample confirms the band at 50 kDa is indeed SETD7 and therefore the transfection was successful. The in-gel digestion of the 25 kDa band was also identified as SETD7. The lower 25 kDa band corresponds to the N-terminal fragment of SETD7.

Another FLAG IP was performed to confirm that the expressed FLAG-SETD7 is active. To achieve this, the elution buffer was exchanged using a 30 kDa Amicon centrifugal filter to KMT buffer (the buffer used in the *in vitro* assays). Mononucleosomes were then added to the re-buffered sample and 400 μ M CD₃-SAM was added to start the reaction. The reaction was incubated for 1 h and the same workflow for histone sample preparation was performed as described previously in Section 3.1. The resulting LC-MS/MS data from this experiment showed successful incorporation of 'heavy' labelled monomethylation on H3K4 by CD₃-SAM, which is shown in Figure 26. This experiment confirms that

the SETD7 being expressed in the transfected HEK-293T cells is indeed active. However, overexpressing SETD7 in HEK-293T cells did not result in an increase in labelling on H3K4 in a cellular environment. Therefore, this experiment provides evidence that SETD7 is catalytically active on mononucleosomes but not on chromatin substrates in a cellular environment.

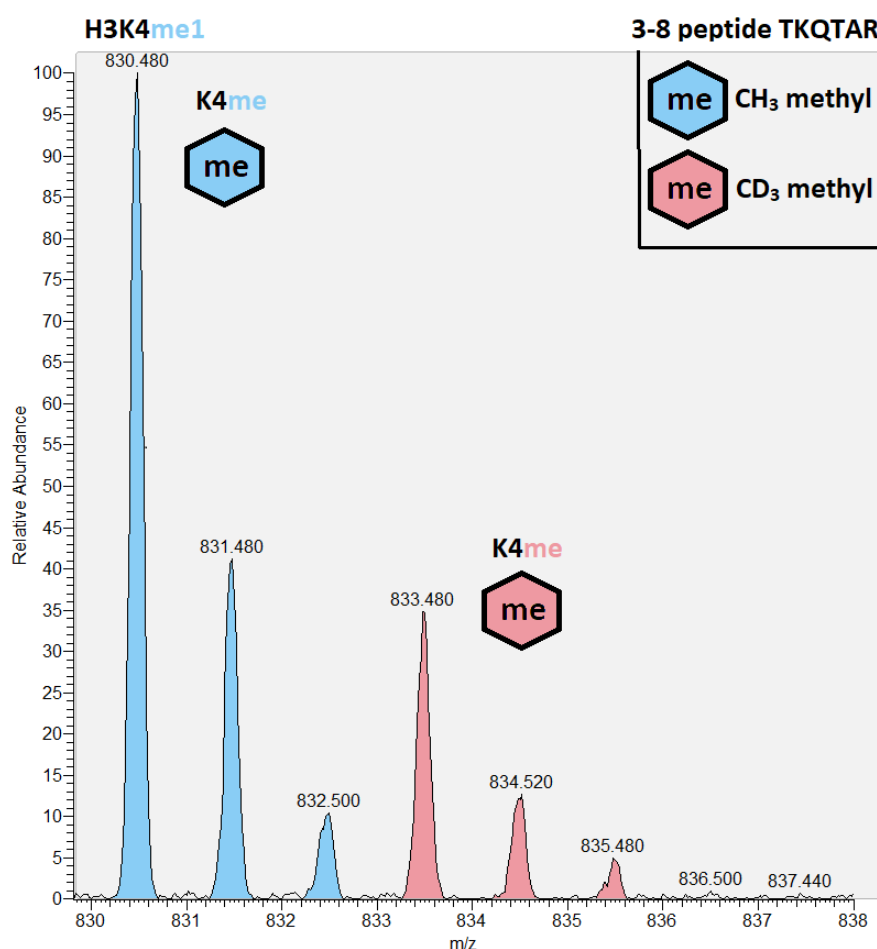


Figure 26 – MS spectra showing CD₃ incorporation (pink peaks) for SETD7's methylation mark H3K4me1 following incubation of mononucleosomes with FLAG-tagged SETD7 from transfected HEK-293T cells and 400 μ M CD₃-SAM for 1 h.

The transfection protocol was revisited to ensure there was nothing hindering the methylation by the overexpressed SETD7. The standard protocol for Lipofectamine transfection is to transfect the cells at 70–90% confluency. Even at 70% confluency, leaving the cells to transfect for 24 hours before adding the probe meant the cells were already close to confluency at the time of probe addition. It is thought that the cells being confluent affect the incorporation of the CD₃ label, hence the decreased labelling in assays 4 and 5. This was verified with a test experiment where confluent

HEK-293T cells were incubated with CD₃-SAM for 20 h. This experiment resulted in approximately 50% decrease of CD₃ labelling seen for all the lysine targets of H3, confirming that confluency does affect the incorporation of the CD₃ label. Considering this confluency issue, Lipofectamine transfection was performed as before but instead the probe was added after 10 hours of transfection instead of 24 hours (Table 5, assay 6). This transfection protocol did not improve labelling significantly compared to assays 4 and 5. Alongside this, transfection was also performed on cells that were 50% confluent instead of 70% incubating both with CD₃-SAM and the methylpiperazine SAM analogue (Table 5, assays 7 and 8). Although a decrease in the percentage of CD₃ labelling was not seen in comparison to the experiments performed without transfection (as seen previously in assays 4 and 5), the percentage did not improve from those experiments. At 50% confluency, the majority of the cells will still be able to divide and thus a greater incorporation of CD₃ marks would be expected compared to cells at 70%, regardless of the transfection.

Despite various changes to the transfection protocol, no method improved the labelling observed on any H3 or H4 lysine sites. Therefore, other experiments were explored to confirm whether or not H3K4 is a valid target by SETD7 in a cellular environment. These studies are discussed in the following sections.

3.3.3. Testing other cell lines

One of the reasons the initial cellular assays showed such low levels of labelling is the low expression levels of SETD7 in HEK-293T cells. In parallel to investigating transfection techniques, other cell lines were investigated that have been reported by the literature and by the Human Protein Atlas to have higher endogenous expression levels of SETD7. The cell lines chosen for this reason were the breast cancer cell lines – MDA-MB-231 and MCF7.¹⁴²

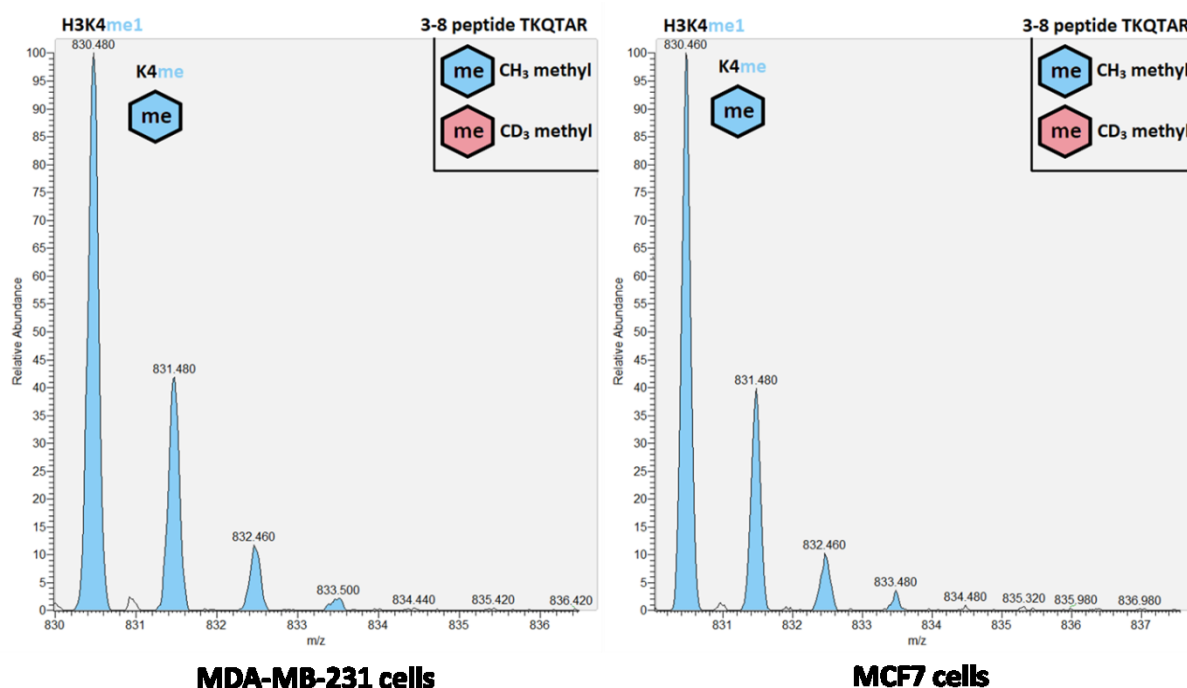


Figure 27 – MS spectra showing no successful CD₃ incorporation on the methylation mark H3K4me1 when incubating MDA-MB-231 cells (left spectrum) and MCF7 cells (right spectrum) with 200 μ M CD₃-SAM for 20 h.

The optimised cellular assay conditions (200 μ M CD₃-SAM for both 20 h and 48 h) were repeated for both MDA-MB-231 and MCF7 cells. The results of the 20 h experiments resulted in no ‘heavy’ labelling observed on any of the H3 lysine sites (Figure 27). Both cell lines from the 48 h experiments showed reduced viability after 24 h and detached from the plate. This loss of cell viability over time seemed unusual so a fresh batch of CD₃-SAM was synthesised to confirm the probe was not the issue. Again, no labelling was observed, though the 48 h experiments did show increased cell viability compared to the previous experiments. Therefore, it was concluded that using HEK-293T cells provides ‘heavy’ labelling on histones with CD₃-SAM but not MDA-MB-231 or MCF7 cell lines.

As previously mentioned, SAM analogues degrade slowly over time and so are always kept as frozen stocks. One issue that could result in the loss of viability after 24 h is the formation of degradation products. Therefore, a study was performed to identify the degradation products of the SAM analogues over time under similar conditions to the cell assays. This would also provide insight into the reason behind the disparity of labelling observed between the two types of cell lines. This study is discussed in the following section (Section 3.4).

3.4. Degradation pathways of SAM

3.4.1. Background – Physicochemical properties of SAM

The SAM analogues, including CD₃-SAM, are synthesised and purified shortly before use in any of the biological assays performed. This is due to the chemical instability of these compounds making the synthesis and purification difficult. The purified stocks are always kept at -20 °C or -80 °C to slow down the degradation. However, before the stocks have been stored for a couple of months, new stock is necessary to ensure the purity of the analogues being tested are sufficiently high for biological purposes. This is because up to 10% of the stock will include degradation products after 2 months, depending on the way the stock was handled and stored.

There are several different ways SAM can degrade over time and the two dominant degradation pathways are shown in Figure 38. The first degradation pathway (left to right in Figure 38) is hydrolysis of the C-S bond in SAM to give the degradation products 5'-methylthioadenosine (MTA) and homoserine lactone. The homoserine lactone can then react with any water molecule that is present to produce its open form, amino acid homoserine.¹⁴³ The other main degradation pathway is cleavage of the glycosidic bond to leave S-ribosyl methionine and adenine as the degradation products.

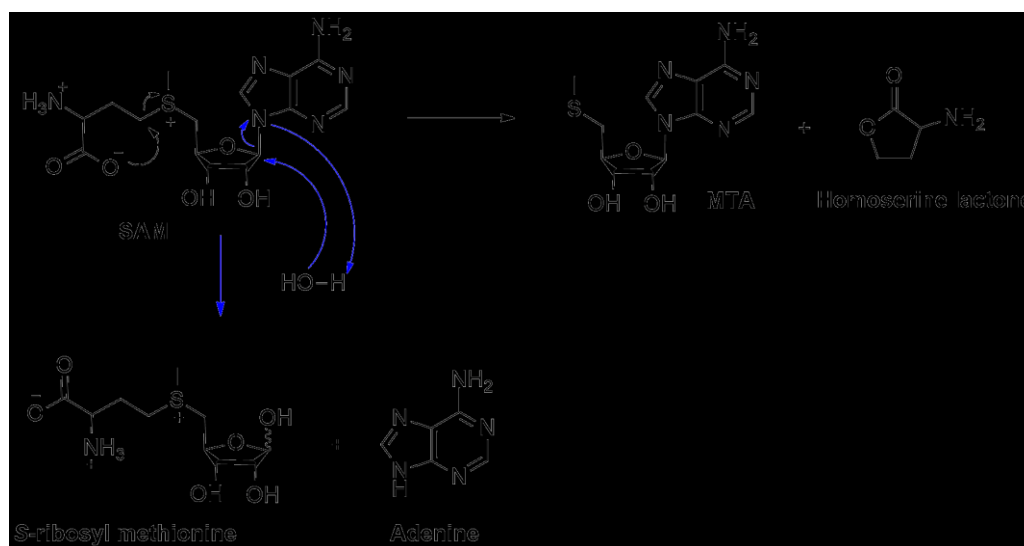


Figure 38 – The dominant degradation pathways of SAM at physiological conditions (37 °C, pH 7.0).

Another mechanism that can occur is the racemisation at the sulphur centre of SAM. In biological systems, only the active (*S,S*)-SAM is synthesised via the enzyme MAT. However, even commercial sources of SAM that have been isolated from yeast have included 20% of the biologically inactive

(*R,S*) diastereoisomer (confirmed by NMR studies conducted by Stolowitz and Minch).¹⁴⁴ More evidence for the racemisation was produced by Wu *et al.* where repeated incubation of freshly synthesised SAM from MAT produced the inactive (*R,S*)-SAM that co-eluted with the active form by reverse phase HPLC. They calculated the constant for racemisation to be $8 \times 10^{-6} \text{ s}^{-1}$ at 37 °C at pH 7.5 (similar to the conditions used in the *in vitro* and cellular assays).¹⁴⁴ Synthesis of the SAM analogues discussed in Chapter 2 yield a mixture of both the active and inactive diastereoisomer of SAM. It has been reported in previous studies that enzymes involved in the utilisation of SAM bind only to the active (*S,S*) configuration.^{114,145} Therefore, the inactive (*R,S*) diastereoisomer will likely be removed via degradation to MTA and adenine to allow for further intracellular metabolism.

To explain the mechanism of SAM degrading to MTA, we can look at the pH profile and the major species of SAM between pH's 0 and 14 (shown in Figure 39). The major species from pH 8.8 until a pH of around 1.5 is a zwitterionic species where the carboxylate oxygen is deprotonated and the amine group on the methionine is protonated. Lower than 1.5, the carboxylic acid is now protonated in the major species of SAM. Cleavage of SAM to MTA occurs via the intramolecular attack of the negatively charged carboxylic acid group to the C-4 of the homocysteine side chain (shown in Figure 38). This mechanism will be unfavourable when the carboxylic acid is protonated thus explaining why the cleavage to MTA diminishes at pH 1.5 and lower.¹⁴⁴ The racemisation at the sulphur centre is the only significant mechanism occurring at pH 1.5 or lower. Therefore, the inherent instability of SAM can be improved by lowering the pH.¹⁴⁴ However, biochemical assays or cellular assays are incompatible at these extreme pH values.

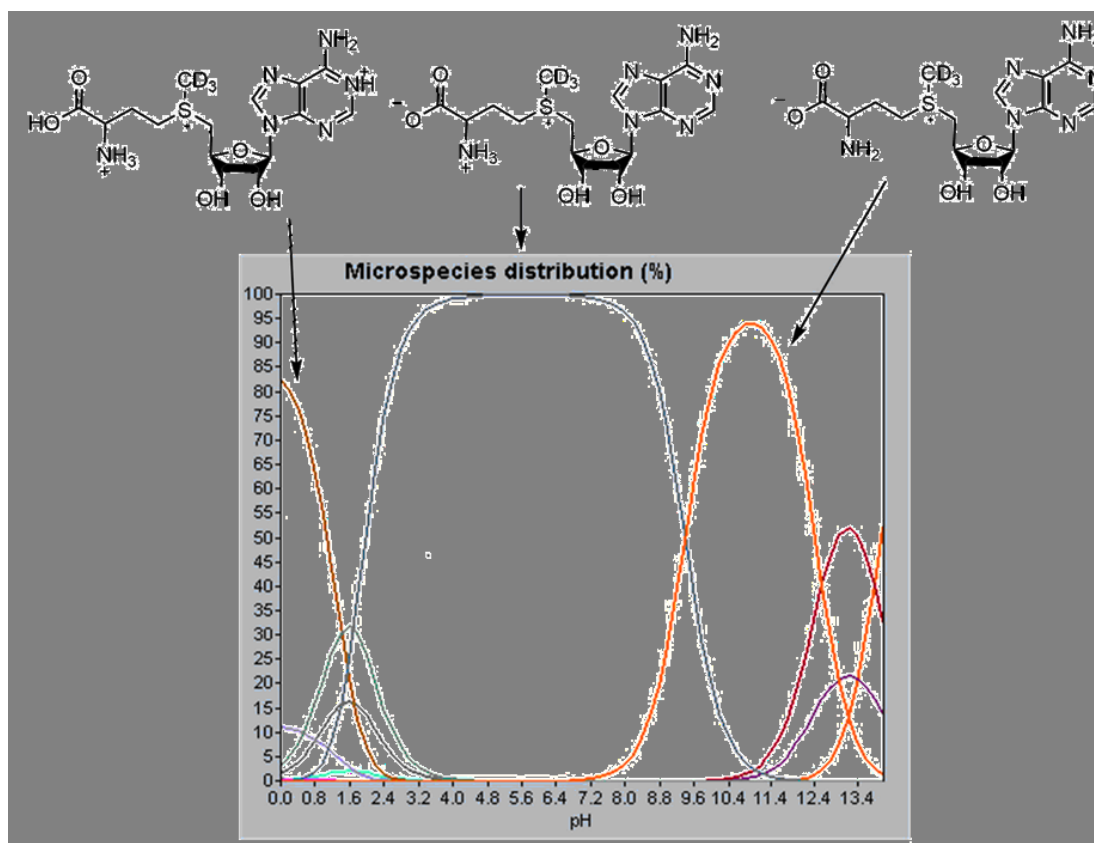


Figure 39 – (A) The pH profile of CD₃-SAM calculated with MarvinSketch and labelled to show the % distribution of each species as pH changes in the range of 0 to 14. (B) Degradation mechanism resulting in cleavage of the C-S bond of SAM to produce MTA and homoserine lactone.

3.4.2. Degradation studies of CD₃-SAM and SAM analogues

To observe whether our analogues are degrading over time, four aliquots were taken of each analogue at 200 μ M. Each aliquot was incubated at 37 $^{\circ}$ C in pH 7.4 Tris-HCl buffer to keep temperature and pH at similar level to the media in the cell assays. Samples were taken at 0, 2, 24 and 48 hours to measure the level of degradation over time. A 2-hour time length was chosen as this is the maximum time length that has been used for *in vitro* assays. At each time point, the sample was analysed by reverse phase HPLC and the peaks were identified by LC-MS/MS analysis.

Firstly, the degradation study was performed with CD₃-SAM and the chromatogram resulting from each time point is shown in Figure 40. At the start of the time course, the first peak at retention time 3.68 min is CD₃-SAM confirmed by LC-MS/MS and is 99.8% pure. Another peak appears at a later retention time of 10.53 min. By LC-MS/MS this peak was elucidated to be the CD₃ analogue of

methylthioadenosine (MTA) with a mass of m/z 301.27 (Figure 41). After 2 hours, there is only 1.9% of this degradation product but after 48 hours there is a significant percentage (>20%) of CD₃-MTA.

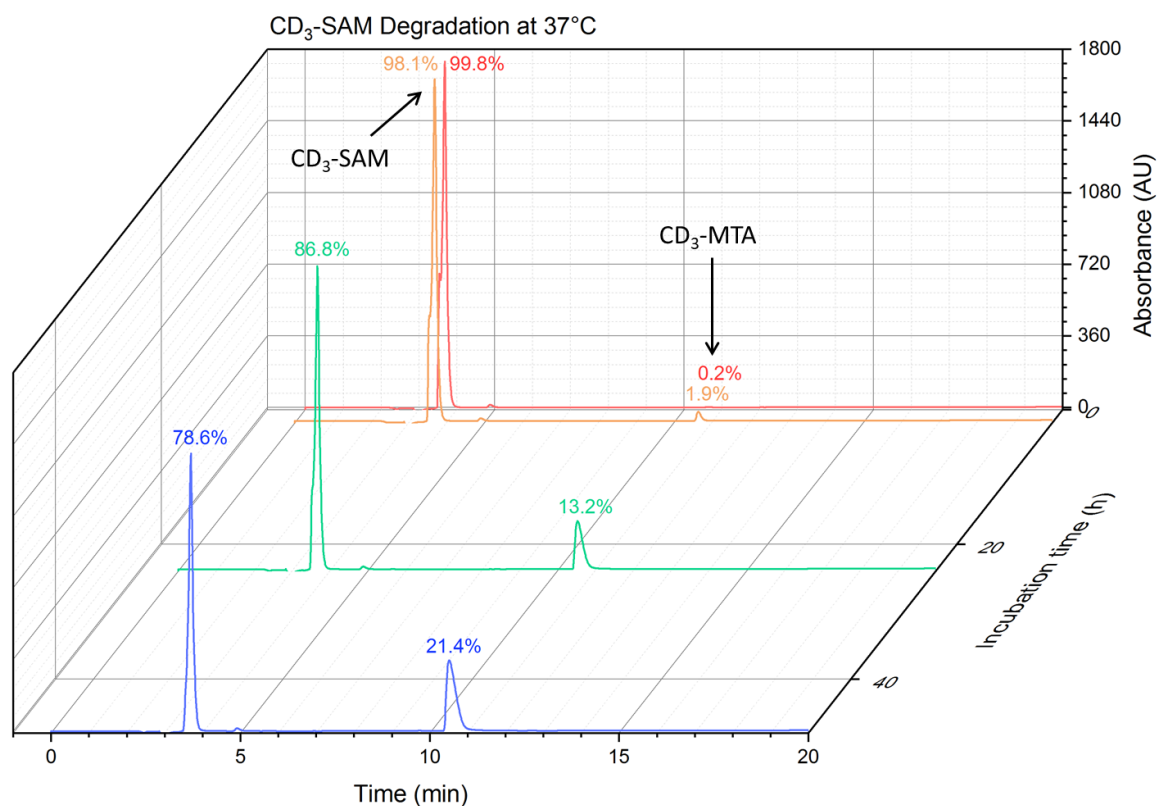


Figure 40 – HPLC chromatogram showing the composition of CD₃-SAM over time at 200 μ M incubated at 37 °C in pH 7.4 Tris-HCl buffer.

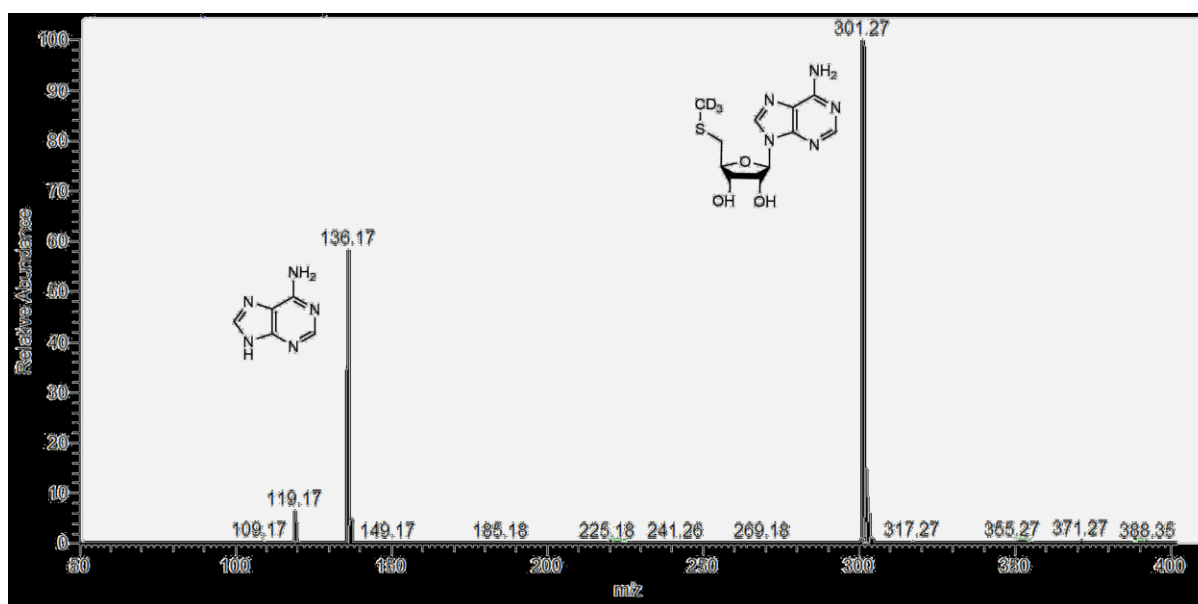


Figure 41 – MS/MS spectrum of CD₃-methylthioadenosine (CD₃-MTA).

The pH in cell culture media is close to neutral at approximately 7.4, thus allowing MTA to form in cellular assays once the cells are incubated with SAM. Therefore, we need to understand the metabolic pathways involving MTA, which will be discussed in Section 4.1.2.

The degradation study was repeated for all the SAM analogues modified at the C-2 position. The study revealed the same trend in degradation to their respective MTA analogue modified at the C-2 position, over time. At 48 hours, the degradation corresponding to the MTA analogue ranged from 14% for the 2-phenyl-CD₃-SAM analogue to 29% for the 2-thiophenyl-CD₃-SAM analogue. The only difference observed was for the 2-thiophenyl SAM analogue, which has two additional peaks at 48 hours – denoted A and B in Figure 42. These peaks (A and B) only contribute a very small amount to the mass balance at 1.7% and 1.8% for A and B, respectively. These two peaks were identified as two products from further degradation of MTA via the hydrolysis of the glycosidic bond – 5-methylthioribose and the 2-thiophenyl adenine analogue (Figure 42).

MTA is a common degradation product of SAM and the biological significance of this metabolite is discussed in further detail in Chapter 4. The degradation of SAM to MTA could provide an explanation for the histone labelling that was observed when the ‘heavy’ labelled SAM analogues were utilised in the cellular assays (Section 3.3.1). The labelling observed in the initial cellular assays was assumed to come from CD₃-SAM but if CD₃-SAM is not cell permeable than it is possible the labelling arises from MTA. Cellular assays utilising CD₃-MTA is investigated and discussed further in Section 4. It is therefore evident from the initial investigation into cellular assays that cell permeability continues to be the prevalent issue. The following sections in this chapter will investigate alternative methods to overcome cell permeability.

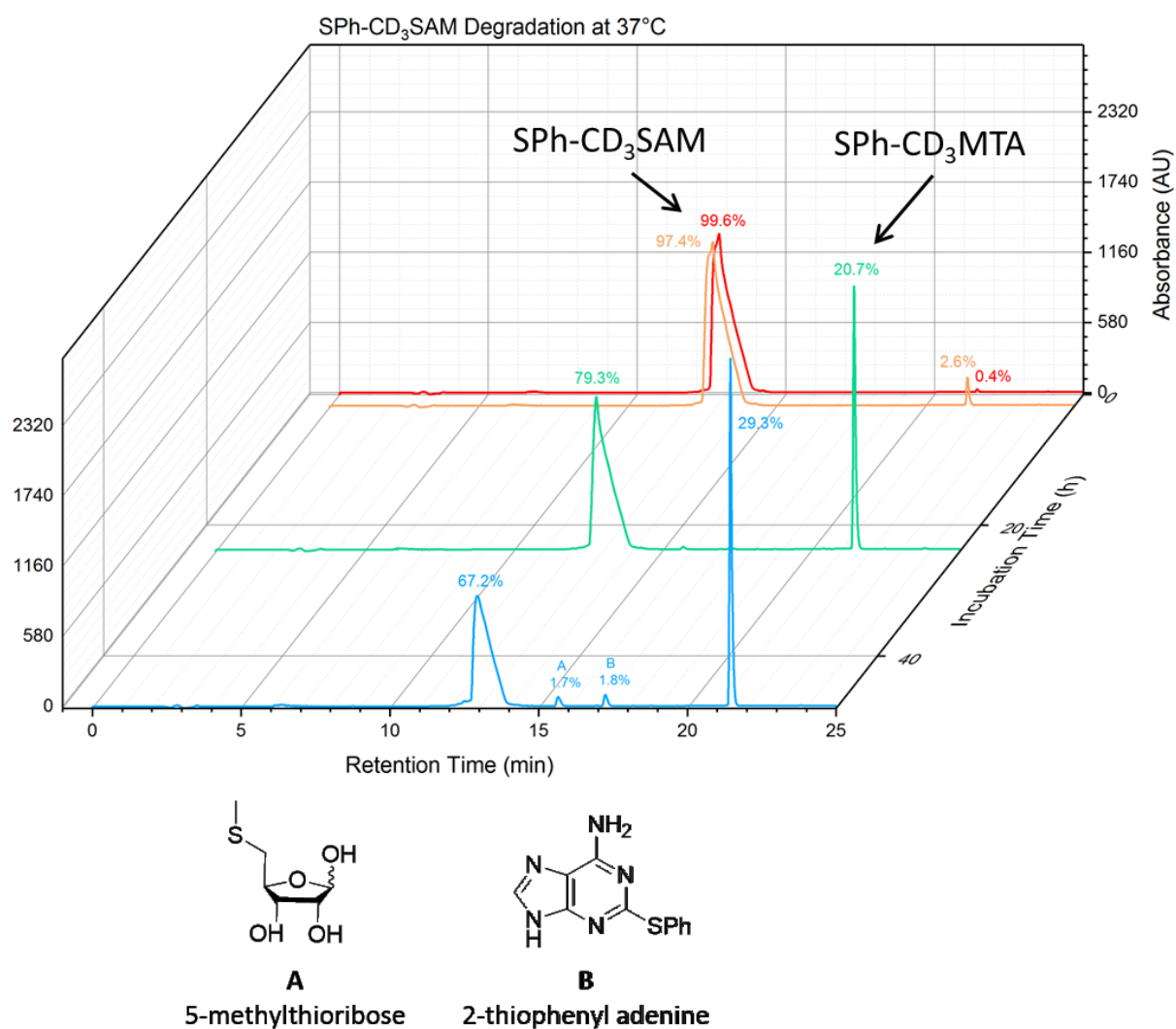


Figure 42 – HPLC chromatogram showing the composition of 2-thiophenyl CD3-SAM over time at 200 μ M incubated at 37 °C in pH 7.4 Tris-HCl buffer.

3.5. *In nucleo* assays

In an attempt to overcome the apparent cell permeability issues, a different assay was optimised utilising only the nuclei of cells. These *in nucleo* assays provide an easy and fast way of analysing histone modifications while still keeping the chromatin environment intact. Although SAM has low cell permeability, SAM and SAM analogues should easily pass through the nuclear membrane because the nuclear pore complex allows free diffusion of proteins under 60 kDa.¹⁴⁶ A large number of non-histone proteins are localised in the nucleus in addition to histones, thus leading to the potential of investigating nuclear non-histone substrates.

The procedure for the isolation of nuclei from HEK-293T cells and subsequent incubation with the CD₃-SAM analogues is described in the Methods section and is modified from the procedure reported by Fischle.¹⁴⁷ The general workflow for these experiments is shown in Figure 43. HEK-293T cells were cultured until 70–80% confluency and then harvested ensuring the cell population is still actively growing and dividing. After the cell pellet was washed, the pellet was incubated with 0.1% NP-40 detergent to break down the cell membrane allowing isolation of the nuclei. Only a fraction of NP-40 is used to ensure the proteins in the cells are not completely unfolded and thus rendering the assay inactive. The nuclei were then incubated with CD₃-SAM at 37 °C to incorporate ‘heavy’ methyl marks on histone proteins. The samples were subjected to the standard histone workflow described previously (Section 3.1) before LC-MS/MS analysis. The main limitation to utilising nuclei is the time course of the labelling step. The time of the assay is limited to a maximum of two hours because nuclei only survive for a couple of hours before becoming inactive.

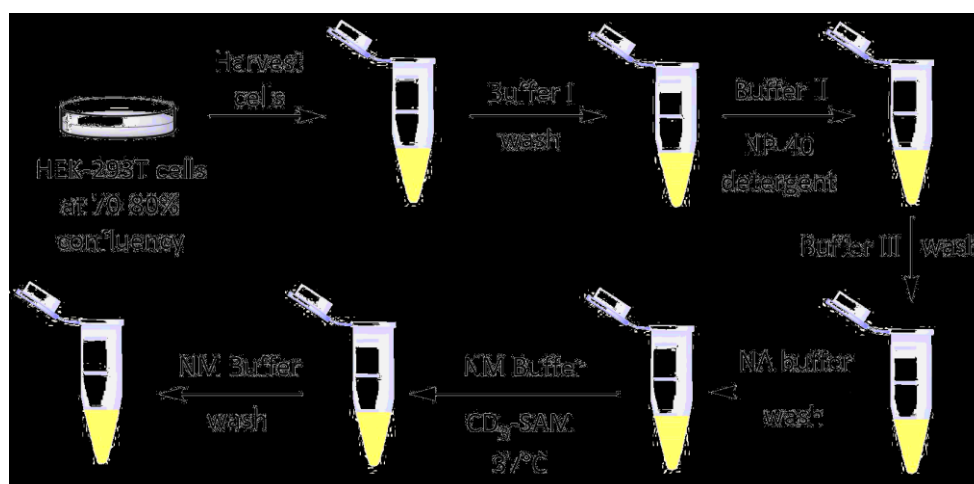


Figure 43 – The general workflow employed for the *in nucleo* assays utilising the nuclei of HEK-293T cells. For details of buffer solutions, refer to the Methods section.

The method was then optimised for CD₃-SAM by varying the conditions listed in Table 6.

Table 6 – The conditions that were varied for optimisation of the *in nucleo* assays and the lysine sites that CD₃ methylation was observed at. The inhibitors used were protease, phosphatase and histone deacetylase (HDAC) inhibitors: microcystin, 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF) and sodium butyrate.

Assay	Time/ h	Douncing	Inhibitors used in methylation incubation step	Conc ⁿ of CD ₃ -SAM used/ μ M	ATP added to NA buffer (1 mM)	Lysine labelling observed
1	1	None	Yes	1	No	None
2	1	After NP-40	Yes	1	No	H3K4 (small degree)
3	1	Before NP-40	Yes	1	No	H3K4 (small degree)
4	1	After NP-40	No	1	No	H3K4
5	1.5	After NP-40	No	10	No	H3K4, K9 (small degree)
6+7	1.5	After NP-40	No	100	No	H3K4, K9, K27, K36, H4K20
8+9	1.5	After NP-40	No	200	No	H3K4, K9, K27, K36, H4K20
10+11	1.5	After NP-40	No	200	Yes	H3K4, K9, K27, K36, H4K20

The first experiment (assay 1, Table 6) followed the Fischle protocol exactly.¹⁴⁷ After analysis of the histone peptides by LC-MS/MS, no conclusive CD₃ methylation was detected. From this result, it was surmised that the NP-40 alone is insufficient to fully lyse the cell membrane and allow CD₃-SAM to enter the nucleus. Consequently, an extra douncing step was introduced either before the addition of NP-40 or after (assays 2 and 3, Table 6). From these two experiments, a small degree of labelling was detected but not at a sufficient level to carry on in-depth studies with. However, from the two assays it was shown that better results were observed when douncing was introduced after NP-40 addition.

To increase the ‘heavy’ labelling observed, the inhibitors used in the protocol were removed from the buffers in the CD₃-SAM incubation step (assay 4, Table 6). The inhibitors used in these assays were microcystin, AEBSF and sodium butyrate which are phosphatase, protease and histone

deacetylase inhibitors respectively. It was proposed that these inhibitors were interfering with the methylation reaction. This assay resulted in easily observable peaks in the MS1 spectra that display CD₃ 'heavy' methylation on H3K4 with the corresponding +3 mass shift. However, 'heavy' labelling of some lysine methylation marks was still low under these conditions, including K9 dimethylation (assay 4, Table 6).

To further improve the assay conditions, the concentration of CD₃-SAM was steadily increased from 1 μ M to 200 μ M to try and improve the labelling on all histone lysine methylation sites (assays 5–9, Table 6). The MS results from these experiments did show increased levels of 'heavy' methylation and K9, K27, K36 and H4K20 methylation was also observed at a sufficient amount. The only site that did not exhibit detectable levels of methylation on MS1 was H3K79. Two examples of the labelling observed on the lysine sites H3K4 and H3K9 are shown in Figure 44.

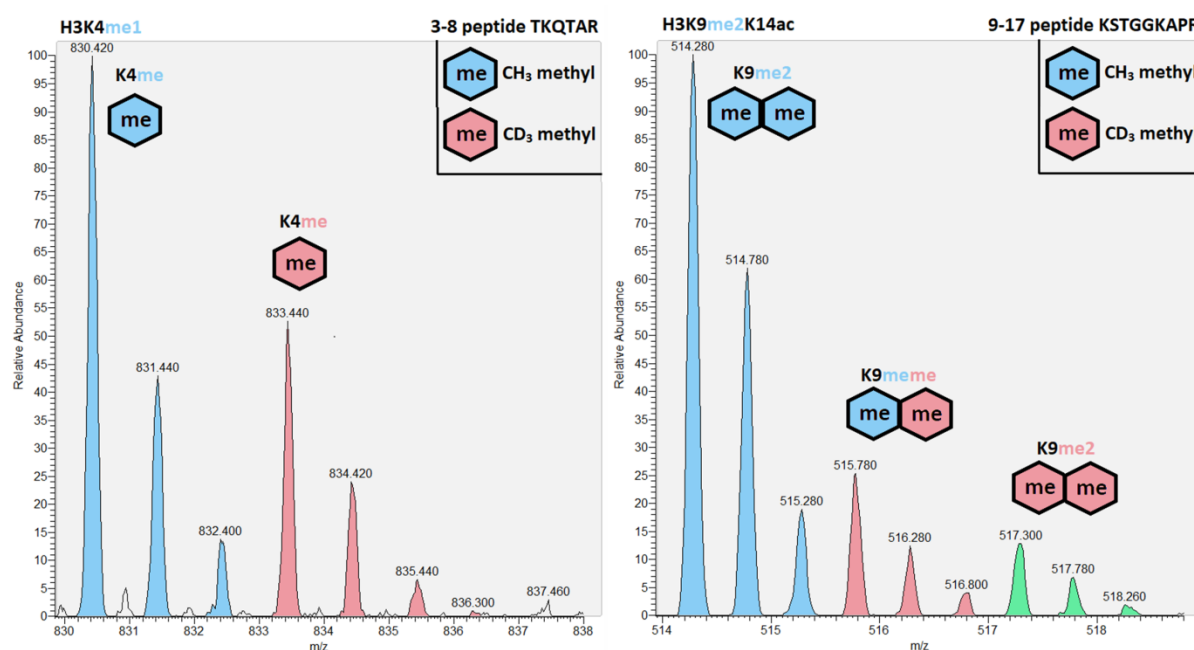


Figure 44 – MS spectra showing CD₃ incorporation (red and green peaks) for the methylation marks H3K4me1 (left spectrum) and H3K9me2K14ac (right spectrum) when incubating HEK-293T nuclei with 200 μ M CD₃-SAM for 1.5 h.

Another experiment was performed which introduced ATP (1 mM) together with CD₃-SAM (200 μ M) at the incubation step. ATP was added in an effort to activate ATP-dependent chromatin remodelling complexes that reside in the nuclei. Activating these complexes should open the chromatin structure thus giving HKMT enzymes access to the more compact marks associated with heterochromatin.

Despite this theory, this assay resulted in lower methylation observed than the previous assays at the same concentration.

With assay conditions optimised for the *in nucleo* assays (Assays 8+9, Table 6), the assays were repeated for the four C-2 modified SAM analogues. The ratio of 'heavy' labelling observed with the 2-modified analogues compared to CD₃-SAM can be seen in Figure 45 for the dimethylation (blue bars) and trimethylation (green bars) marks on H3K9 and monomethylation (pink bars) on K4. The expected result, in comparison with the *in vitro* data in Figure 26, would be that labelling would be observed on H3K4 (SETD7 target) but not on any of the other histone lysine residues. However, the results in Figure 45 show that no labelling is observed on any of the histone methylation sites for the 2-piperidine-, 2-methylpiperazine- or 2-phenyl-CD₃-SAM analogue. In contrast, the thiophenyl analogue did show labelling on both K9 and K4 histone lysine residues. The latter result correlates with the *in vitro* assays performed by Diego Alonso where the 2-thiophenyl analogue showed the least selectivity for SETD7s target over G9as than the other analogues (Section 3.2). The complete lack of labelling that is presented for the other three analogues tested (MPip-CD₃SAM, Pip-CD₃SAM and Ph-CD₃SAM) is unexpected. There should not be any permeability issues here; however, the lack of labelling could be due to another problem that was referred to earlier – for example, whether H3K4 is a valid target of SETD7 in a biologically relevant system. These results combined with the results from the cellular assays do not show enough evidence for SETD7 labelling on H3K4 and thus leads to the next stage of investigation pursued in the next Section – investigating non-histone targets. In light of this issue, the *in nucleo* assays were not progressed until more relevant targets for SETD7 could be established and validated.

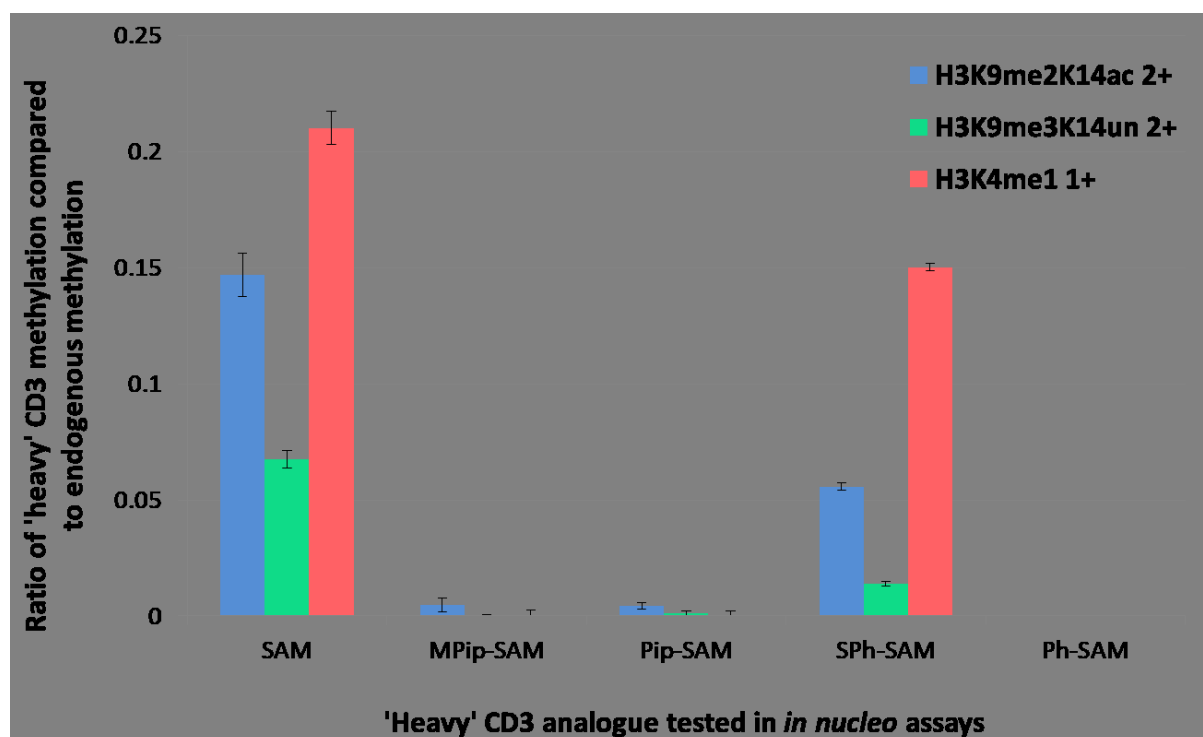


Figure 45 – Comparison of the ratio of CD₃ methylation observed when HEK-293T cell nuclei were incubated with 200 μ M CD₃-SAM (far left) vs the small library of C-2 modified CD₃-SAM analogues. The error bars show one standard deviation (SD). (MPip-SAM = 2-methylpiperazine-SAM, SPh-SAM = 2-thiophenyl-SAM, Pip-SAM = 2-piperidine-SAM, Ph-SAM = 2-phenyl SAM).

3.6. Investigation into non-histone targets of SETD7

3.6.1. Background – Non-histone targets of SETD7

In recent years, lysine methylation has emerged as a prevalent PTM in many non-histone proteins that has many important roles such as regulation in cellular processes.¹⁴⁸ The number of identified lysine methylation sites on non-histone proteins is rapidly increasing but it is suspected that many are still widely uncharacterised. This gap is due to lack of a suitable technology for measuring the methylome. SETD7 is known to have a large range of non-histone targets, exceeding the known number for any other HKMT. Figure 46 highlights the specific proteins that have been observed to be methylated by SETD7 in the literature.

The SETD7 lysine methylation sites of each of these non-histone protein targets are provided in Table 7 along with the techniques that were used to confirm the methylation and/or lysine residue that is methylated.

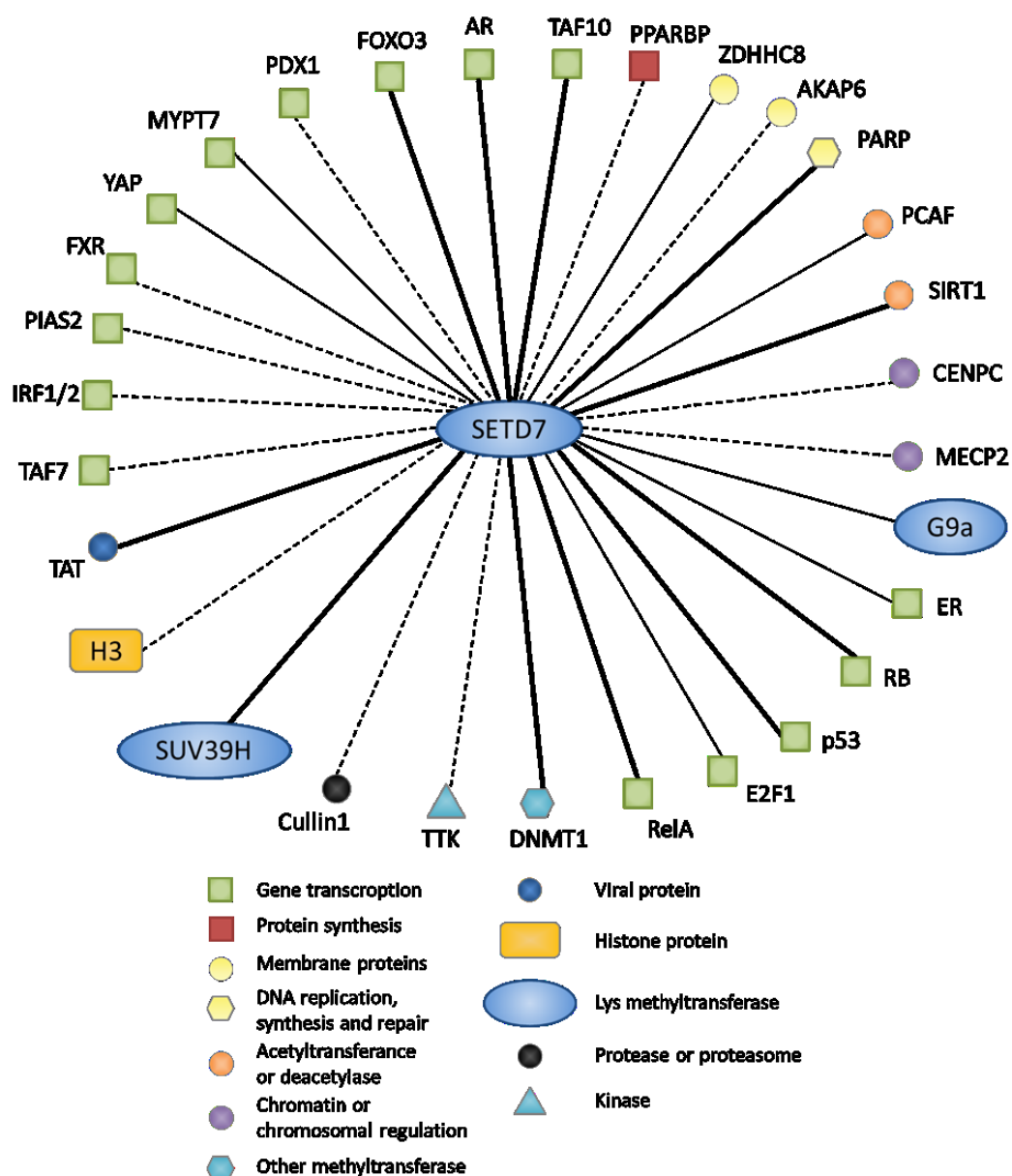


Figure 28 – The known non-histone methylation targets for the HKMT SETD7. Bold lines highlight the targets with more data supporting that they are a valid target and dashed lines show targets with limited data supporting their validity. Figure adapted from Li and Biggar.¹⁴⁸

Table 7 – The non-histone protein targets of SETD7 (including UNIPROT ID) and the lysine residue reported to be methylated by SETD7, the state of methylation and the techniques used to confirm methylation.

PROTEIN	UNIPROT ID	RESIDUES	METHYLATION STATE	ASSAYS/TECHNIQUES USED TO VERIFY METHYLATION		REF.
				<i>In vitro</i>	Cellular assays	
RelA	Q04206	37, 314, 315	1Me	Autoradiography, immunoblotting and MS experiments using full-length and peptide fragments of RelA	Immunoprecipitation and siRNA experiments using cells with endogenous RelA and cells overexpressing RelA and SETD7	45,73
E2F1	Q01094	185	1Me	Autoradiography experiments using full-length and peptide fragments of recombinant E2F1	Immunoprecipitation experiments using cells overexpressing E2F1 and cells with endogenous E2F1	149
p53	P04637	372	1Me	Autoradiography and crystal structure studies using full-length and peptide fragments of p53	IP, Western blots, mutagenesis and siRNA experiments using cells overexpressing p53 and cells with endogenous p53	150–155
RB	P06400	810, 873	1Me	Autoradiography and MS experiments using full-length and peptide fragments of recombinant RB	Immunoprecipitation experiments using cells overexpressing RB and SETD7	97,156
ER	P03372	302	1Me	MS experiments using peptide fragments of ER	Immunoprecipitation experiments using cells overexpressing ER	66
MECP2	P51608	347	1Me	Autoradiography and MS (data not shown) experiments using the target protein domain of MECP1	In gel digest then MS experiment using cells expressing the protein domain of MECP2	42
CENPC	Q03188	414	1Me	Autoradiography and MS (data not shown) experiments using the target protein domain of CENPC	In gel digest then MS experiment using cells overexpressing the protein domain of CENPC	42
SIRT1	Q96EB6	233, 235, 236, 238	1Me?	Autoradiography experiments using full-length and peptide fragments of SIRT1	Immunoprecipitation, autoradiography and MS experiments using cells overexpressing SIRT1 and SETD7	150
PCAF	Q92831	78, 89	1Me	Autoradiography experiments using full-length and peptide fragments of PCAF	Immunoprecipitation experiments using cells overexpressing PCAF and cells with endogenous PCAF	157
AKAP6	Q13023	604	?	Autoradiography and MS experiments	In gel digest then MS (could not detect the	42

				(data not shown) using the target protein domain of AKAP6	peptide containing lysine) experiment using cells expressing the protein domain of CENPC	
PARP1	P09874	508	1Me	Autoradiography, mutagenesis and MS experiments on full-length protein and peptides fragments of PARP1	Immunoprecipitation and mutagenesis experiments on cells with endogenous PARP1 and cells overexpressing SETD7	70
ZDHHC8	Q9ULC8	300	1Me	Autoradiography and MS (data not shown) experiments using the target protein domain of ZDHHC8	In gel digest then MS experiment using cells overexpressing the protein domain of ZDHHC8	42
PPARBP	Q15648	1006	?	Autoradiography and MS (data not shown) experiments using the target protein domain of PPARBP	In gel digest then MS (could not detect the peptide containing lysine) experiment using cells expressing the protein domain of PPARBP	42
TAF10	Q12962	189	1Me	Autoradiography, kinetic assays, crystal structure with TAF10 peptide and MS (data not shown) experiments using full-length protein and GST fusions of TAF10	Immunoprecipitation experiments using nuclear extracts from cells with endogenous TAF10	33
AR	P10275	630, 632	1Me	Autoradiography, mutagenesis, MS using full-length protein and truncated peptide fragments of AR	<i>In nucleo</i> assays, immunoprecipitation, western blots and siRNA experiments using cells with endogenous AR	44,63
FOXO3	O43524	271	1Me	Autoradiography, mutagenesis and MS experiments using full-length protein and peptide fragments of FOXO3	Immunoprecipitation, mutagenesis, siRNA and immunofluorescence to detect localisation using cells overexpressing FOXO3 and cells with endogenous FOXO3	71,72
PDX1	P52945	?	?	Autoradiography – shows methylation of full-length PDX1 in presence of SETD7	–	65
MYPT1	O14974	442	1Me	Autoradiography, mutagenesis and MS experiments using full-length MYPT1	Immunoprecipitation and siRNA experiments using cells overexpressing MYPT1	158
YAP	P46938	494	1Me	–	Immunoprecipitation then MS, siRNA and mutagenesis experiments using cells overexpressing Flag-YAP	159
FXR	P51114	206	?	Autoradiography experiments using full-length FXR	Immunoprecipitation (no mention of the methylation state) experiments using cells overexpressing FXR	160

PIAS2	Q96T58	2076	2Me	Autoradiography and MS (data not shown) experiments using the target protein domain of PIAS2	In gel digest then MS experiment using cells overexpressing the protein domain of PIAS2	⁴²
IRF1/2	P10914	126, 134	1Me, 2Me	Autoradiography and MS (data not shown) experiments using the target protein domain of IRF1/2	In gel digest then MS experiment using cells overexpressing the protein domain of IRF1/2	⁴²
TAF7	Q15545	5	1Me	Autoradiography and mutagenesis experiments using full-length TAF7	–	³³
STAT3	P40763	140	2Me	Autoradiography and Western blot experiments using full-length STAT3	Immunoprecipitation, siRNA and MS experiments of immunoprecipitated STAT3 (but not with incubation with SETD7/labelled methyl) using cells overexpressing STAT3	¹⁶¹
TAT	P04610	51	1Me	Autoradiography, MS and mutagenesis experiments using full-length TAT protein	Immunoprecipitation, Western blot and mutagenesis experiments using cells overexpressing TAT protein	⁶⁹
SUV39H1	O43463	105, 123	1Me	Autoradiography, mutagenesis and MS experiments using full-length and fragment peptides of SUV39H1 peptide	Western blot and siRNA experiments using cells with endogenously expressed SUV39H1	¹⁶²
Cullin1	Q13616	73	?	Autoradiography and MS (data not shown) experiments using the target protein domain of Cullin1	In gel digest then MS (could not detect the peptide containing lysine) experiments using cells overexpressing the protein domain of Cullin1	⁴²
TTK	P33981	708, 710?	?	Autoradiography and MS (data not shown) experiments using the target protein domain of TTK	In gel digest then MS (could not detect the peptide containing lysine) experiment using cells overexpressing the protein domain of TTK	⁴²
DNMT1	P26358, P13864	142, 1094	1Me	Autoradiography, kinetic assays, MS and mutagenesis experiments using full-length and fragment peptides of DNMT1	Western blots, siRNA, immunoprecipitation then autoradiography experiments using cells endogenously expressing DNMT1	^{163,67}

Lysine 4 of histone 3 (H3K4) is a relevant target for SETD7 in the *in vitro* assays described in Section 3.2. However, as discussed in Section 1.3.1, it is debated in the literature whether this is a valid target *in vivo*.^{29,73,164} This is a potential problem for the cellular assays because if histone targets are not relevant substrates, then we will not observe any labelling of H3K4 using our 'heavy' SAM analogues. This is what was observed in the cellular assays described in this chapter, providing evidence that H3K4 is not a physiological target of SETD7. No increased labelling was observed on H3K4 derived from HEK-293T cells, even when using cells that were successfully overexpressing catalytically active, full-length SETD7. Therefore, select non-histone substrates (that have been reported in the literature) were tested in *in vitro* assays to confirm whether they are valid targets for SETD7. Validation is needed before more complex assays were investigated (e.g. cellular and *in nucleio* assays). Success could lead to a better readout for SETD7 activity in cellular systems with CD₃-labelled endogenous SAM and the C-2 modified analogues. Generating annotated MS/MS spectra showing successful methylation of non-histone proteins by SETD7 will allow for the development of a targeted MS approach to validate these targets in cells. Targeted MS approaches will be vital in discovering these methylation signals in complex backgrounds.

The non-histone substrates selected for the assays were chosen based on:

- How many different references confirmed the methylation activity on the protein substrate.
- Whether MS data was included to confirm the methylation activity/residue methylated.
- The commercial availability of the recombinant protein.

This led to a small selection of non-histone substrates consisting of FOXO3, FOXO1 and PARP1. The testing of these proteins will be presented in the next section.

3.6.2. *In vitro* assays on non-histone substrates

All four of the chosen recombinant proteins (FOXO3, FOXO1 and PARP1) were incubated with 1 mM CD₃-SAM and recombinant SETD7 in KMT buffer for 1 h at 37 °C. These conditions were selected in order to be consistent with the optimised *in vitro* assays performed on histone substrates (Section 3.2). The samples were then subjected to in solution digestion (protocol in Section 6.6.1) before LC-MS/MS analysis.

3.6.2.1. FOXO3 and FOXO1 targets

FOXO3 is reported to be monomethylated at lysine residue K271. After in solution digestion using trypsin following the assay, the K271 residue (shown in **bold**) is present in the peptide: KAALQTAPESADDSPSQLSK.⁷¹ The MS data from the FOXO3 *in vitro* experiment was analysed by SEQUEST and this peptide was found as a hit for CD₃ monomethylation on K271. This hit was confirmed and annotated using in-house software, which gave the tandem MS (MS/MS) spectrum shown in Figure 47. This hit also showed 'heavy' dimethylation on lysine residue 290. The dimethylation on K290 was reported previously by Calnan *et al.* but was thought to be an artefact of mass spectrometry due to SETD7 being reported only to monomethylate its targets.⁷² However, it is more likely to be a different modification or protein isoform and thus was investigated further.

A control experiment was performed with no CD₃-SAM present in the *in vitro* assay but all other conditions kept identical. This experiment did not yield any CD₃ methylation on K271 as expected but did show the same 'heavy' CD₃ dimethylation on K290 as before. Since there was no CD₃ feedstock in this control experiment, this could not be CD₃-dimethylation. This led to the investigation of other modifications and changes to the sequence that could compensate for the apparent 'CD₃-dimethylation' that would result in a mass difference of +34 at the C-terminus. It is also noted that there is only a good coverage of b- and y-ions up to the amino acid leucine at residue 288. This indicates that the difference could be present at the leucine residue 288.

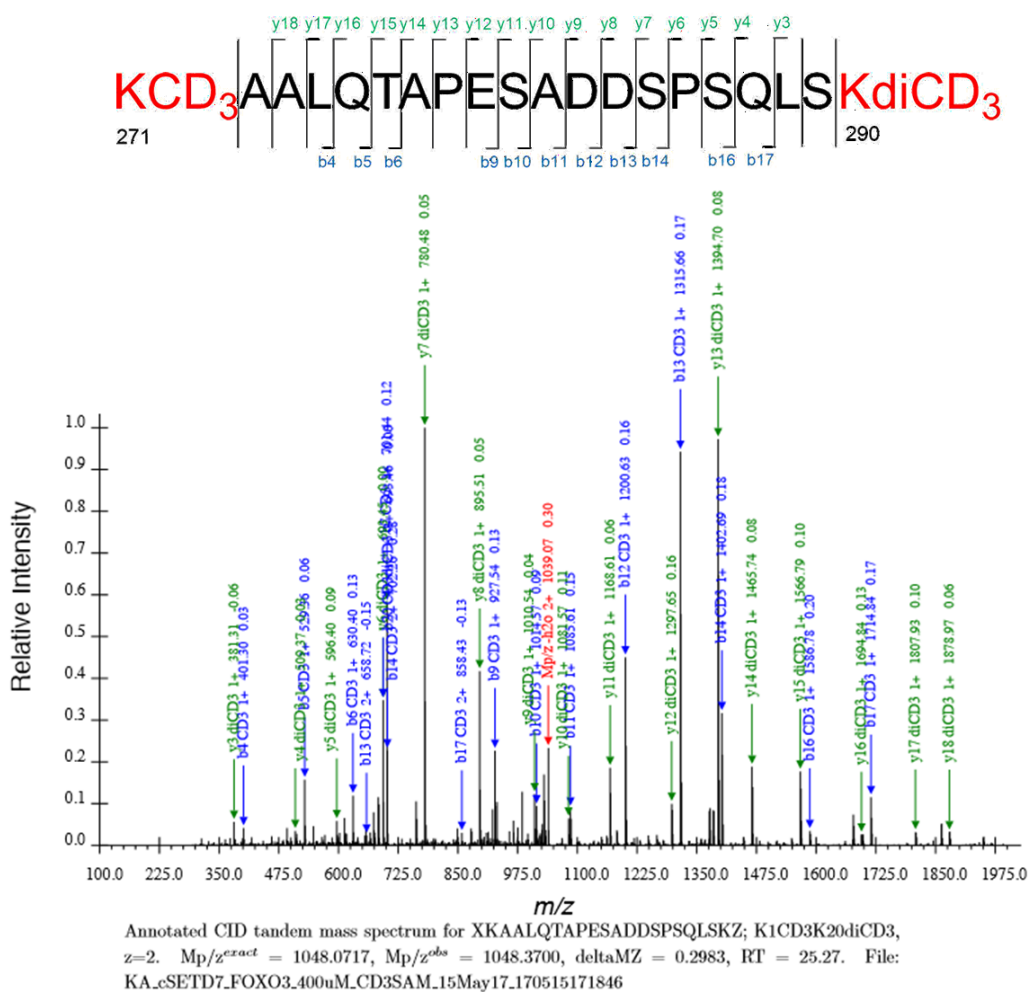


Figure 47 – MS/MS spectra of the peptide containing CD₃ monomethylation on K271 and apparent CD₃ dimethylation on K290 of FOXO3. The coverage of the b- and y-ions is shown above in blue and green, respectively, and the labelled residue is labelled in red.

The only way to compensate for this mass (+34) and not be CD₃ dimethylation would be an amino acid mutation from leucine (L) to phenylalanine (F) at position 288. This modified sequence was then analysed by the in-house software to give the MS spectrum shown in Figure 48. Good b-ion coverage confirmed that it was indeed an F mutation. Following this, the company who sent the recombinant FOXO3 was contacted to confirm whether there was a mutation in the construct they were using to express the recombinant protein. After they re-sequenced the construct, it was confirmed that position 288 is indeed a phenylalanine residue instead of a leucine residue. To conclude, even though there is a single nucleotide polymorphism (SNP) at position 288, these experiments do confirm that SETD7 can monomethylate FOXO3 at lysine 271.

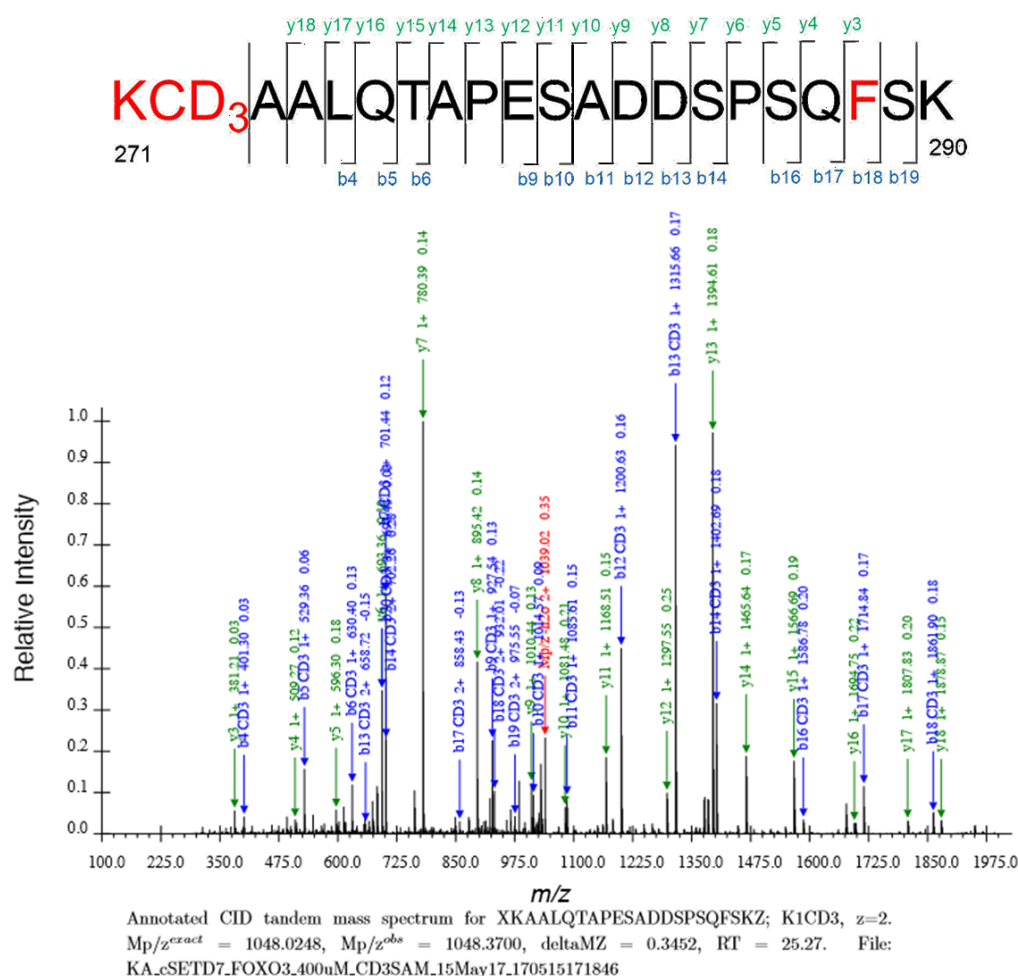


Figure 29 – MS/MS spectra of the peptide containing CD₃ monomethylation on K271 and F mutation at position 288 of FOXO3. The coverage of the b- and y-ions is shown above in blue and green, respectively, and the residues of interest labelled in red.

The *in vitro* assay was then repeated for FOXO1. It has been reported that the methylation site is conserved between the FOXO family members and therefore the lysine residue found to be methylated was K274 for FOXO1.⁷¹ After in solution digestion using trypsin following the assay, the K274 residue (shown in bold) is present in the peptide: **KASLQSGQEGAGDSPGSQFSK**. The MS data from the FOXO1 *in vitro* experiment was analysed by SEQUEST and this peptide was found as a hit for CD₃ monomethylation on K274. This hit was confirmed and annotated using in-house software, which gave the tandem MS spectrum shown in Figure 30. The MS/MS spectrum has good coverage of b- and y-ions giving strong evidence this is a valid methylation hit. To confirm this hit further, the same peptide was run through the in-house software, but without a search for methylation at K274, to provide a comparison spectrum. If the two spectra share a similar peak profile it gives further evidence that the MS/MS spectrum in Figure 30 is valid and K274 is a valid lysine target for SETD7. The MS/MS spectrum for the peptide without any labelling can be seen in Figure 31 and does show a

similar profile to the CD₃-labelled peptide. These experiments therefore confirm that SETD7 is able to monomethylate FOXO1 at lysine 274.

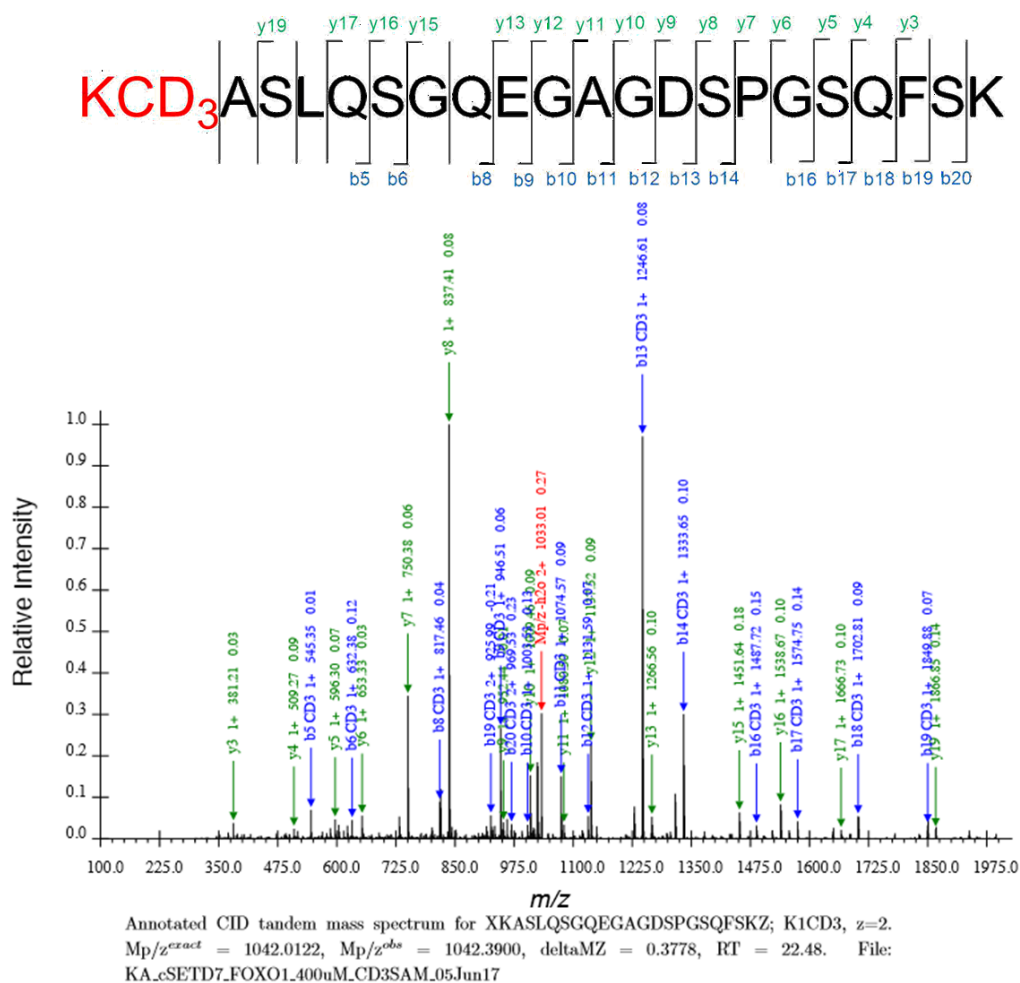


Figure 30 – MS/MS spectra of the peptide containing CD₃ monomethylation on K274 on FOXO1. The coverage of the b- and y-ions is shown above in blue and green, respectively, and the residues of interest labelled in red.

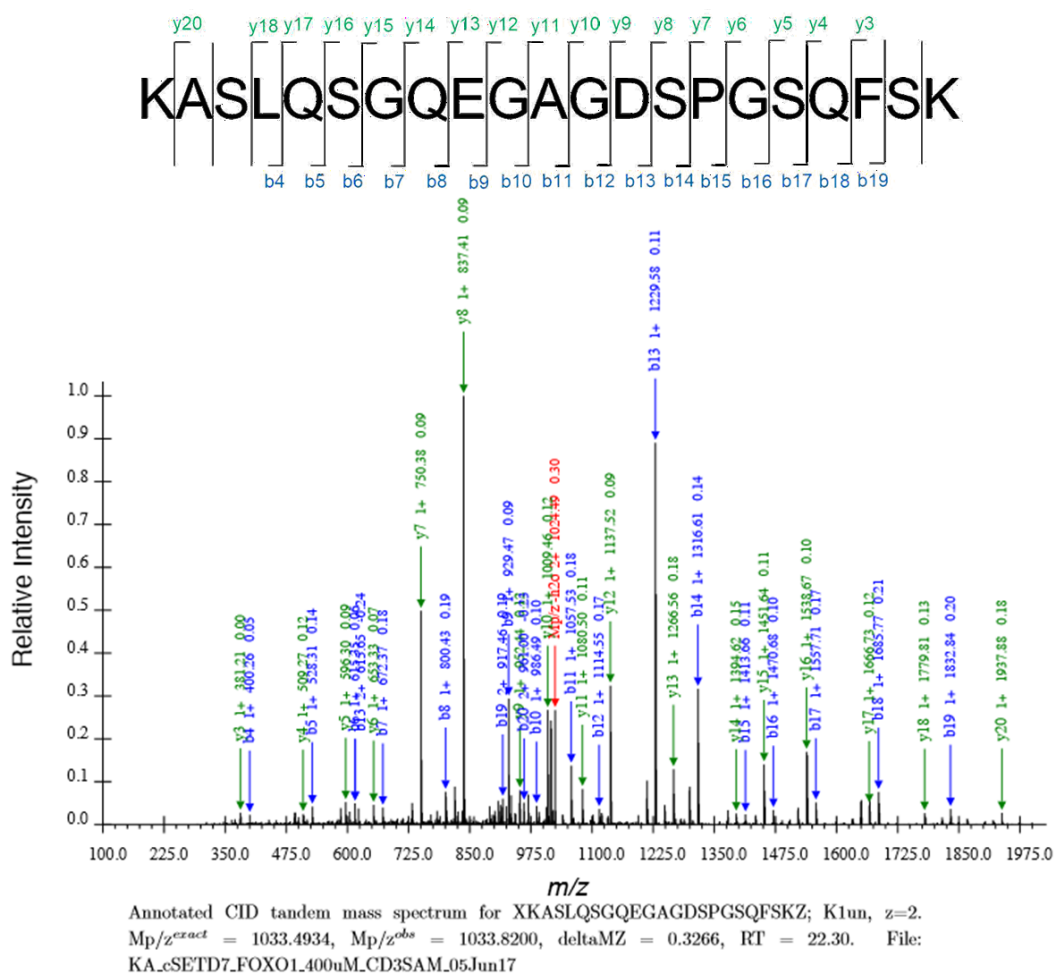


Figure 31 – MS/MS spectra of the peptide containing no CD_3 labelling on K274 of FOXO1. The coverage of the b- and y-ions is shown above in blue and green, respectively.

3.6.2.2. PARP1

PARP1 is reported to be monomethylated at the lysine residue K508.⁷⁰ After in solution digestion using trypsin following the assay, the K508 residue (shown in **bold**) is present in the peptide: **KGQV**. This peptide sequence is very short and thus difficult to confirm by mass spectrometry due to its low mass. Therefore, it was decided that an additional propionylation step should be included in the protocol in an attempt to increase the size of the peptide analyte. Assuming complete propionylation on all lysine sites before trypsin digestion, the resulting peptide would be: GKSGAALSKK**SKGQV**KEEGINKSEKR. The MS data from the PARP1 *in vitro* experiment was analysed by SEQUEST and a truncated version of this peptide (SKGQVKEEGINK) was identified as a hit for CD_3 monomethylation on K508. A truncated version of the expected peptide could have occurred from insufficient propionylation on all available lysine sites. This peptide was then analysed by the in-house software, which gave the tandem MS spectrum shown in Figure 32 for the unmodified

peptide that is propionylated on K2 and K5 in the sequence. The software also produced spectra with good b- and y-ion coverage for this same peptide but with only K2 or K5 propionylated instead of both. However, there was no strong evidence seen for CD₃ monomethylation on K508 (K2 in the peptide sequence).

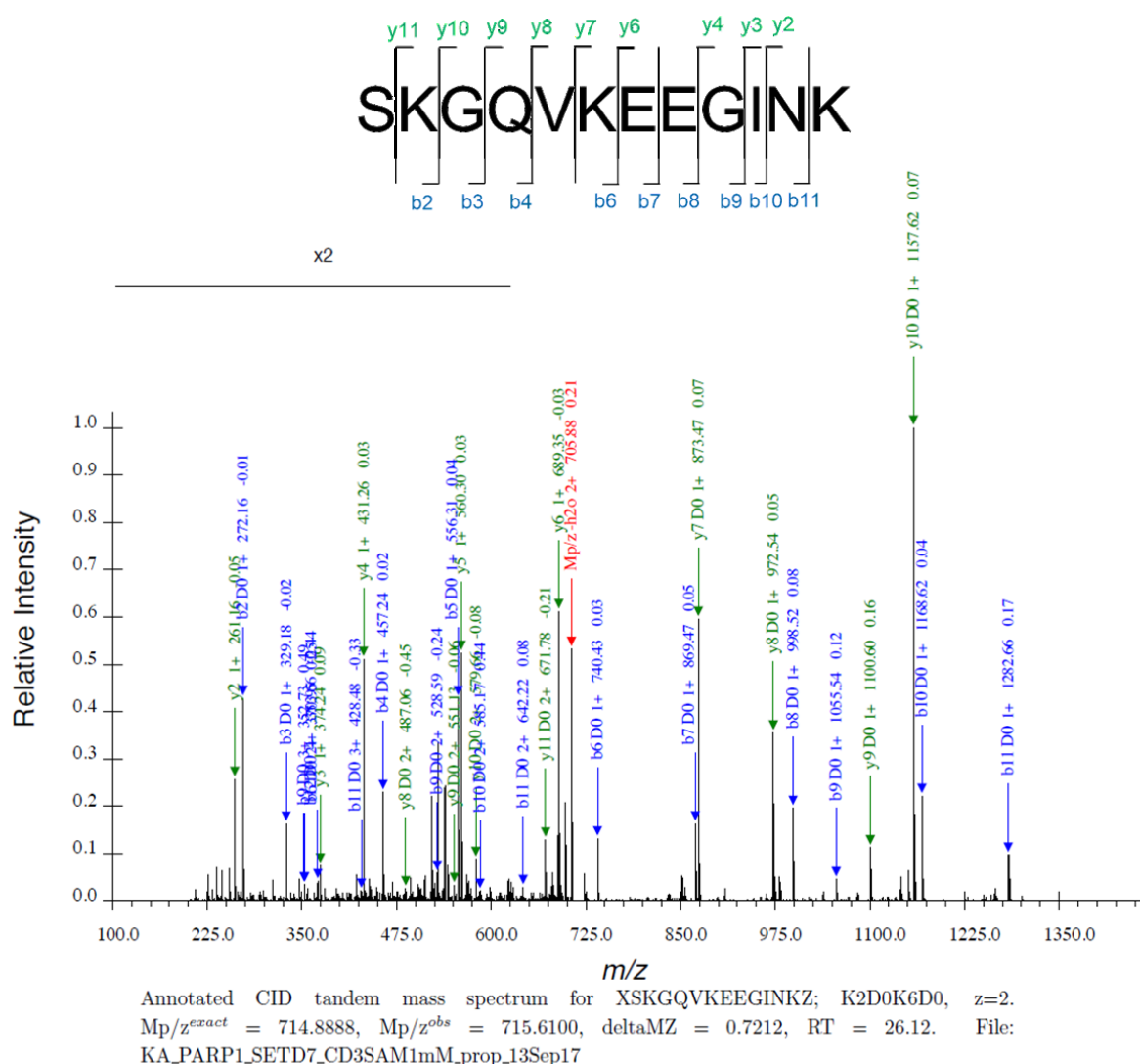


Figure 32 – MS/MS spectra of the endogenous propionylated peptide that included K508 of PARP1.
 The coverage of the b- and y-ions is shown above in blue and green, respectively.

There are a couple of reasons why labelling might not be observed here, whilst still assuming that the K508 residue of PARP1 is a valid target for SETD7. Firstly, it is reported that PARP1 auto-modifies with ADP-ribosylation at various lysine sites around K508, which subsequently inhibits methylation by SETD7 at K508 (Figure 33).^{165,70} Additionally, acetylation at K508 by p300 will stop methylation occurring at this position.¹⁶⁶ Further investigation into this target will need to be performed to

ascertain whether acetylation is blocking this position and/or whether ADP-ribosylation is installed in close proximity (Figure 33). If this is the case, deacetylases can be utilised to see if this allows methylation to occur at K508.

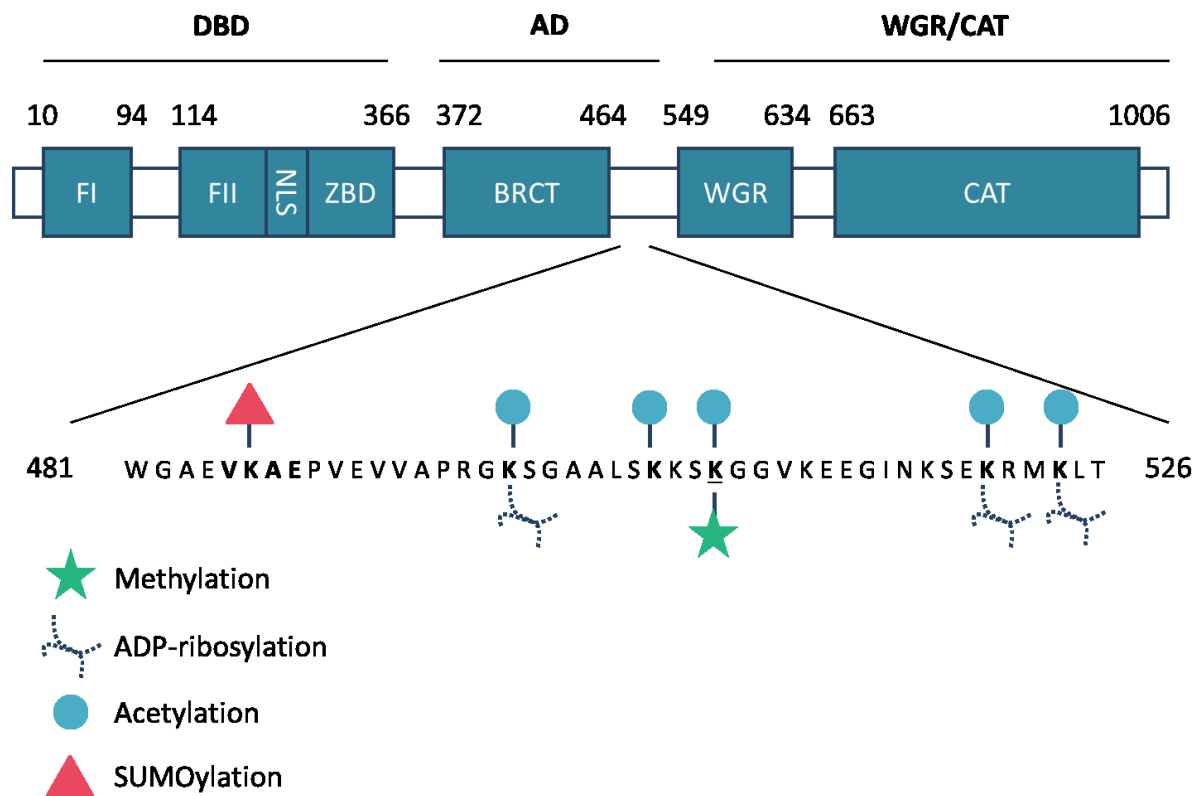


Figure 33 – Schematic representation of the structure of the PARP1 domain and the important post-translational modifications that are installed around the methylation site K508.¹⁶⁶

3.7. Lysate experiments utilising 2-piperidine SAM

In addition to testing individual non-histone targets in *in vitro* assays, another method was used to investigate non-histone targets of SETD7. Cell lysates from HEK-293T cells were incubated with a CD₃-labelled SAM analogue for subsequent analysis by LC-MS/MS. From the *in vitro* assays performed by Diego Alonso-Martinez, the 2-piperidine SAM analogue was shown to exhibit the most promising selectivity for SETD7. In theory, if cell lysates were incubated with the 2-piperidine SAM analogue, only SETD7s lysine targets should be preferentially 'heavy' labelled. Therefore, these experiments were performed on lysates taken from HEK-293T cells that were over-expressing SETD7. After lysates were prepared from cells, they were concentrated and buffer exchanged in 10 kDa Vivaspin® centrifugal filters. Various parameters were altered or introduced in each experiment for purposes of control readings and for optimisation of the assay. The four lysate experiments that were performed included the following reagents:

1. 1 mM 2-piperidine SAM
2. 1 mM 2-piperidine SAM, 2 µg mononucleosomes
3. 1 mM 2-piperidine SAM, 2 µg mononucleosomes, 200 nM recombinant SETD7
4. 1 mM 2-piperidine SAM, 10 µg mononucleosomes, 200 nM recombinant SETD7

Mononucleosomes were added in three of the experiments to facilitate a background reading and to confirm that the methylation reaction worked. Mononucleosomes have shown successful labelling with SETD7 in previous experiments and thus provided a suitable control. The cells that were used to produce the lysates were overexpressing SETD7, but as a control, recombinant SETD7 was added in two of the experiments to ensure signal could be observed.

The difficulty with analysing these samples via LC-MS/MS is the high degree of complexity and background in the sample. To overcome this, the methylated peptides must be enriched to provide adequate signal of the desired 'heavy' methylation over the background noise. A co-worker in the group is in the process of optimising an enrichment technique to achieve this goal. At this present stage of the project, these samples have not undergone analysis and will be processed by the co-worker in future work following the development of the enrichment method.

4. MTA and the methionine salvage pathway

4.1. Background – The methionine salvage pathway (MSP)

S-adenosyl methionine (SAM) is bio-synthesised from methionine and ATP through a reaction catalysed by the enzyme methyl adenosyltransferase (MAT). SAM has many important functions in the cell including the transmethylation of proteins as discussed in the previous sections. SAM also plays a role in the polyamine synthesis of spermine and spermidine, which are both essential for cell growth.¹⁶⁷ The synthesis of these polyamines results in the formation of the critical metabolic by-product 5'-methylthioadenosine (MTA). This pathway, and how it relates to transmethylation, is shown in Figure 34. Polyamine synthesis from SAM occurs in two enzymatic steps: the first step is a decarboxylation of SAM to S-adenosylmethionineamine (dc-SAM), catalysed by adenosylmethionine decarboxylase (SAMDC, step **A** in Figure 34). Subsequently, dc-SAM transfers its *n*-propylamine group to first putrescine, and then to spermidine, resulting in MTA as the by-product (steps **B** and **C**, Figure 34).¹⁶⁸ Another key enzyme in the polyamine synthesis is ornithine decarboxylase (ODC). ODC decarboxylates the amino acid ornithine to yield putrescine, which is the substrate for dc-SAM to transfer its *n*-propylamine group onto to produce spermidine (step **D**, Figure 34).¹⁶⁹

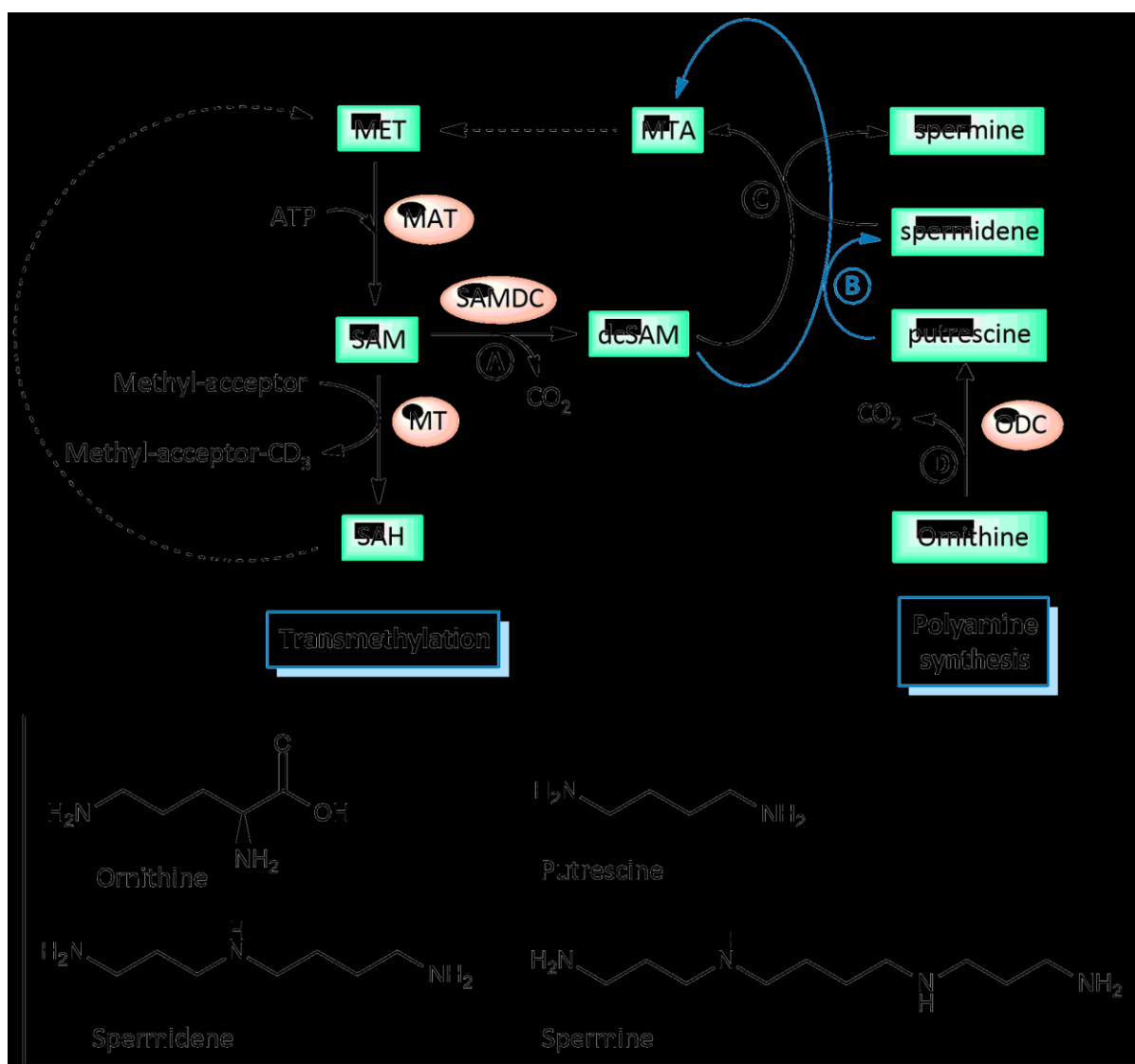


Figure 34 – The polyamine synthesis pathway and its relation to the transmethylation pathway. dcSAM, decarboxylated S-adenosylmethionine; MAT, methionine adenosyltransferase; MTA, methylthioadenosine; SAH, S-adenosylhomocysteine; SAMDC, S-adenosylmethionine decarboxylase; ODC, ornithine decarboxylase; MT, methyltransferase; MET, methionine.¹⁷⁰ Note: SAH can be converted back to MET via two steps (dashed arrow).

MTA can also be formed via the degradation of SAM, which was discussed in Section 3.4. Without a biological mechanism to metabolise MTA further intracellularly, there would be a reduction in the available sulphur pool and this increase in intracellular levels of MTA would result in cytotoxicity.¹⁷¹ The assimilation of sulphur is energetically costly¹⁷² and consequently cells have developed the universal methionine salvage pathway (MSP) to recycle the sulphur in MTA back to L-methionine (see Figure 35). Furthermore, it is essential to maintain the intracellular pool of methionine since methionine is an essential amino acid provided through diet alone.

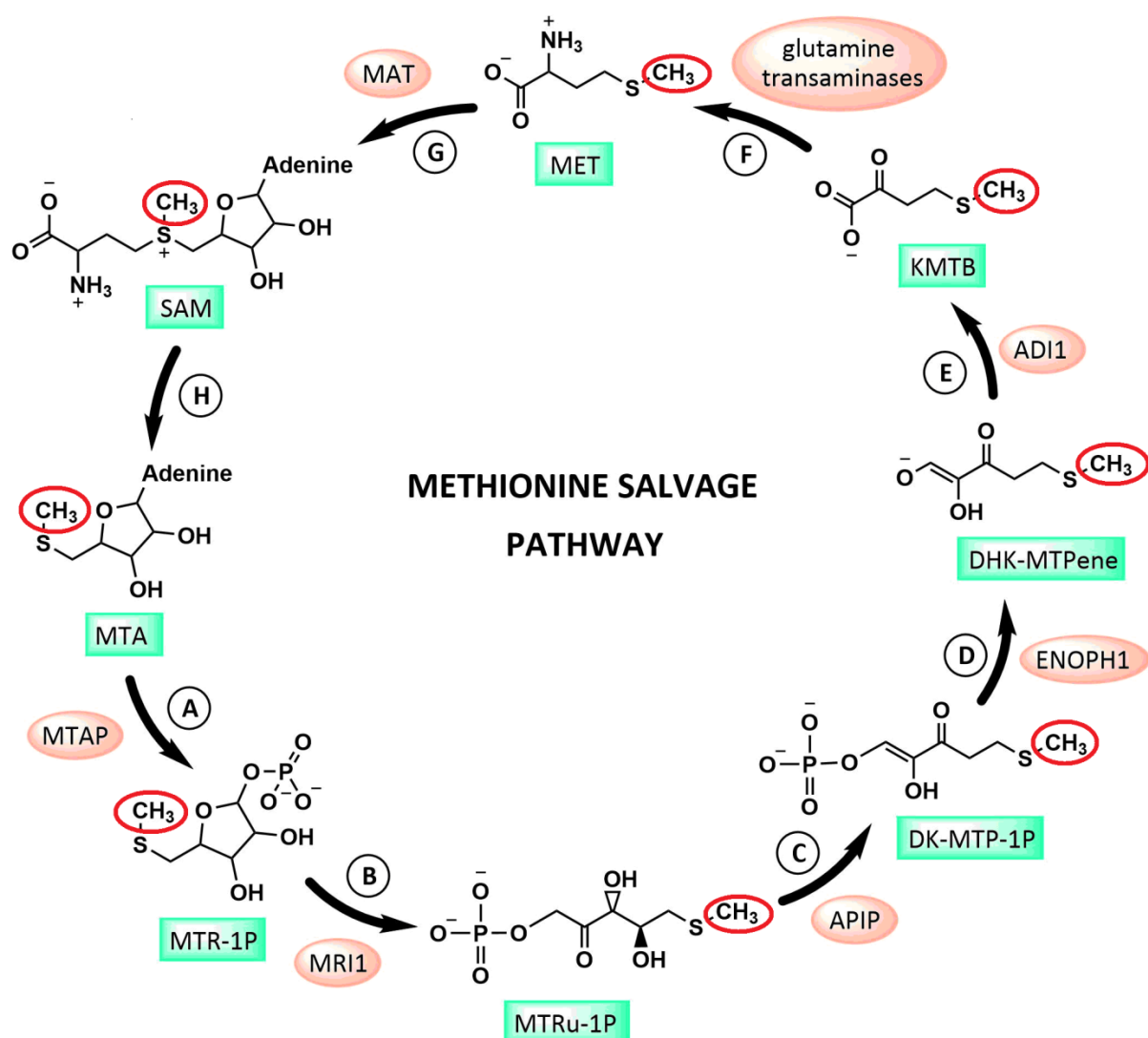


Figure 35 – The universal MSP with the methylthio-group circled in red to highlight the conservation of this group throughout the MSP. Figure adapted from North *et al.*¹⁷¹ Human enzymes have been added in pink that have been reported by Mary *et al.* and the references therein.¹⁷³ Note: Arrow H includes two steps of the pathway, going through the intermediate dc-SAM and these steps are shown in more detail in Figure 34.

As discussed previously, the degradation product we observe for SAM is MTA and this can be utilised in the MSP. Thus, in this chapter we studied the contributions of the two different sources of methionine (from the media or from the MSP) contributing towards histone methylation and whether specific histone methylation marks are correlated with one specific source. Furthermore, it is important to study the function of this salvage pathway in order to fully understand what is taking place in the cell assays performed in Section 3. The first enzymatic step in the methionine salvage pathway is rate limiting and utilises the enzyme MTA phosphorylase (MTAP) to catalyse the phosphorylation of MTA to yield methylthioribose-1-phosphate (MTR-1P) and adenine (step A, Figure 35).¹⁷⁴ Therefore, MTAP is important in the salvage of both methionine and adenine.¹⁷⁵

Following the phosphorylation of MTA by the enzyme MTAP, there are five additional enzymatic steps to yield methionine (steps **B–F**, Figure 35). These steps are catalysed by the following consecutive enzymes: 5'-methylthioribose-1-phosphate isomerase (MRI1), 5'-methylthioribulose-1-phosphate dehydratase (APIP), 2,3-dioxomethiopentane-1-phosphate enolase/phosphatase (ENOPH1), aci-reductone dioxygenase (ADI1) and lastly 4-methylthio-2-oxybutyrate (MOB) transaminase (glutamine transaminases) involved in the final step to form methionine. The specific human enzymes utilised in each step are in brackets and are also shown in pink in Figure 35. Links between the expression levels of enzymes in the MSP and different disease states have been reported and thus deregulation of this pathway would be interesting to investigate.^{176,177}

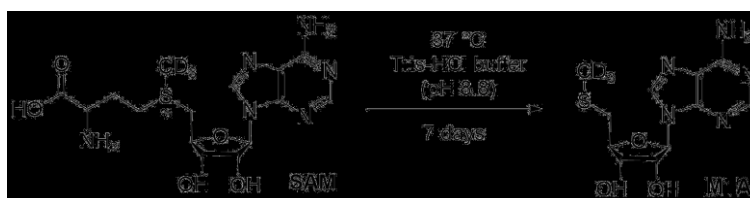
All normal cell lines ubiquitously express the enzyme MTAP which keeps physiological intracellular levels of MTA very low in the nM range.¹⁷⁸ This function is important because MTA is a known inhibitor of both polyamine synthesis and protein methylation, including arginine methylation of STAT1.¹⁷⁹ The methionine salvage pathway has recently gained interest since the discovery that many human tumour cells, including the breast cancer cells lines MDA-MB-231 and MCF7, are deficient in the MTAP enzyme.¹⁸⁰ This deficiency commonly occurs because of the homozygous co-deletion of the MTAP gene with the CDKN2A gene that encodes the tumour suppressor proteins p16 and p14ARF.¹⁷⁵ These genes are co-deleted because of their close proximity on the chromosome region 9p21, which are often deleted in human malignancies. Additionally, loss of MTAP activity in certain types of carcinomas can arise from DNA hyper-methylation of its gene promoter.¹⁸¹

Since various human tumour cells are deficient in MTAP, the methionine salvage pathway cannot be utilised in these systems. The lack of a functioning MSP is highlighted when breast cancer cell lines growing in methionine depleted media are shown to have reduced viability, and thus these cells are known to be 'methionine dependent'.¹⁸² Methionine depleted cells can be rescued with the addition of KMTB (the last intermediate before forming methionine, shown in Figure 35) or by overexpression of the MTAP enzyme. There has been increasing research into administering MTA to treat tumour cells.¹⁸³ The accumulation of MTA leads to cytotoxicity and since these cells cannot remove MTA; this toxicity could lead to potential cancer therapeutics. Therapeutics using a combination therapy of both adenine analogues and MTA have already been developed by Lubin and Lubin to selectively kill MTAP-negative lung tumour cells while keeping MTAP-positive human fibroblasts protected.¹⁷⁵ Another study by Iraburu *et al.* has shown that hepatocarcinoma cells undergo apoptosis when treated with MTA, whereas normal hepatocytes and human fibroblasts are unaffected and continue to be viable.¹⁸⁴ Furthermore, MTA administered to healthy mice and rats was found to not be toxic even at high doses and even demonstrated protective effects against liver damage.¹⁸⁴

The methylthio-group and carbons C-2,3,4 and 5 of the ribose moiety of MTA are all conserved throughout the MSP and therefore they will all be present in the methionine formed from this pathway.¹⁷¹ This conservation means that if 'heavy' labelled CD₃-MTA was utilised in the MSP, the MTA would form CD₃-labelled methionine via this pathway which would subsequently be converted to CD₃-SAM intracellularly. In contrast, breast cancer cell lines like MDA-MB-231 cells should not be able to utilise the metabolite CD₃-MTA due to the absence of the MTAP enzyme. To investigate whether the methionine salvage pathway can be utilised for 'heavy' labelling in cellular assays with CD₃-MTA, cell assays were performed on both HEK-293T cells that do express MTAP and on MDA-MB-231 cells that do not. The results of these experiments are discussed in Section 4.2.

4.2. Utilising CD₃-MTA in cellular assays

To investigate if any 'heavy' methylation on histones can be obtained via the MSP, CD₃-MTA instead of CD₃-SAM was introduced to the cell culture media during the course of the cellular assays. To obtain a sufficient amount of CD₃-MTA to use in cellular assays, batches of CD₃-SAM were incubated at 37 °C in pH 8.8 Tris-HCl buffer for 7 days to induce degradation (Scheme 23). After 7 days of degradation, the samples were purified by HPLC to furnish CD₃-MTA in ≥98% purity. Any remaining CD₃-SAM from the HPLC runs was collected and degraded further to obtain larger quantities of CD₃-MTA.



Scheme 23 – Incubation of CD₃-SAM at 37 °C in Tris-HCl buffer (pH 8.8) for 7 days to induce degradation in order to obtain CD₃-MTA.

Once a pure batch of CD₃-MTA had been obtained, it was introduced to the culture media of HEK-293T cells. The cells were incubated with 200 μM CD₃-MTA for 20 and 48 hours before being subjected to the standard histone workflow and analysed by LC-MS/MS. For both time points, the experiments resulted in successful CD₃ labelling on all lysine methylation sites on H3. Figure 36 is an example of the MS data showing the 'heavy' CD₃ labelling seen for the monomethylation mark on H3K4 for both time points. The left spectrum is the labelling observed at 20 hours and the right spectrum is for the 48 hour time point.

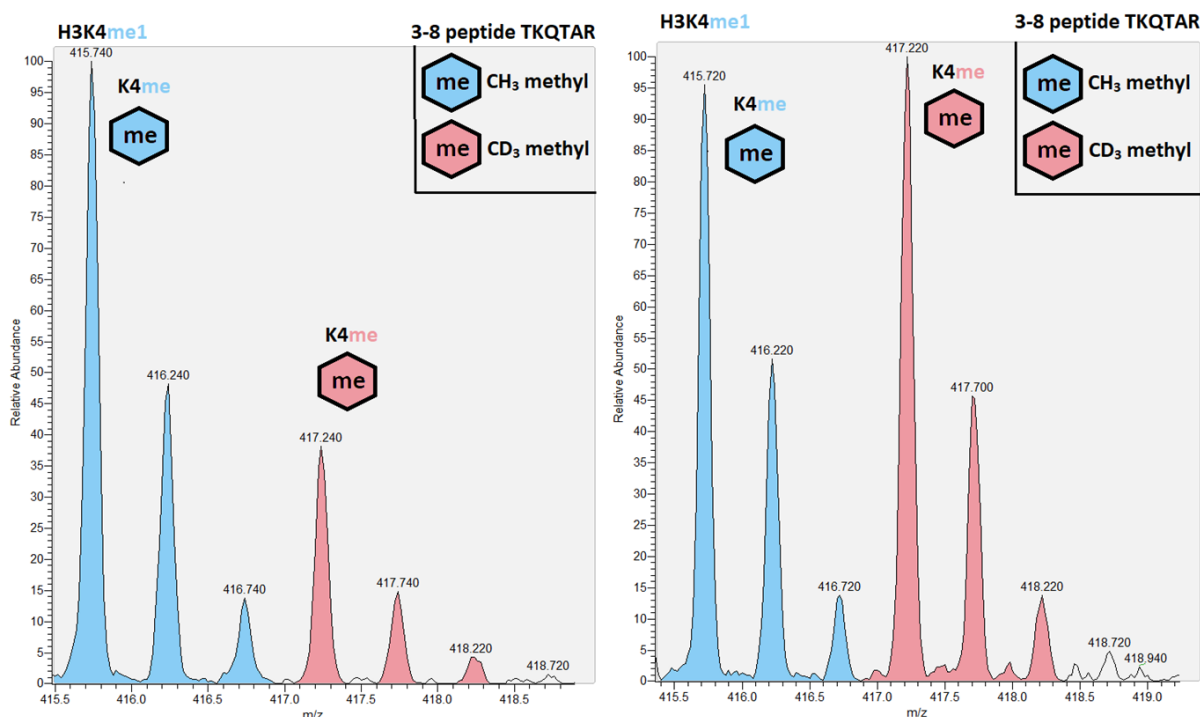


Figure 36 – MS spectra showing CD₃ incorporation on the monomethylation mark H3K4 at 20 h (left spectrum) and 48 h (right spectrum) when incubating HEK-293T cells with 200 μ M CD₃-MTA. The peaks in blue show the original endogenous methylation and the peaks in red show incorporation of a CD₃ group.

If we compare these results to the cell experiments utilising CD₃-SAM, there is a much higher extent of labelling when cells are incubated with CD₃-MTA. Figure 37 shows the MS spectra of CD₃ labelling on the monomethylation mark of H3K4 at 48 hours observed for both CD₃-SAM (left) and CD₃-MTA (right) incubation. The data shown in Figures 55 and 56 indicates that CD₃-MTA can be utilised in the methionine salvage pathway to produce CD₃-SAM via CD₃-methionine. The CD₃-SAM produced via the pathway can then be utilised in intracellular methylation reactions. Importantly this result could explain the 'heavy' labelling obtained when HEK-293T cells were incubated with CD₃-SAM (these experiments are detailed in Section 3.3). When CD₃-SAM degrades to CD₃-MTA over time it can enter the cells and label via the intracellular production of CD₃-SAM produced from the methionine salvage pathway. CD₃-SAM only degrades to 21.4% CD₃-MTA over 48 hours as evidenced from the degradation study in Section 3.4. Therefore, this degradation could explain the lower extent of labelling seen for the CD₃-SAM experiments at the same concentration and time points.

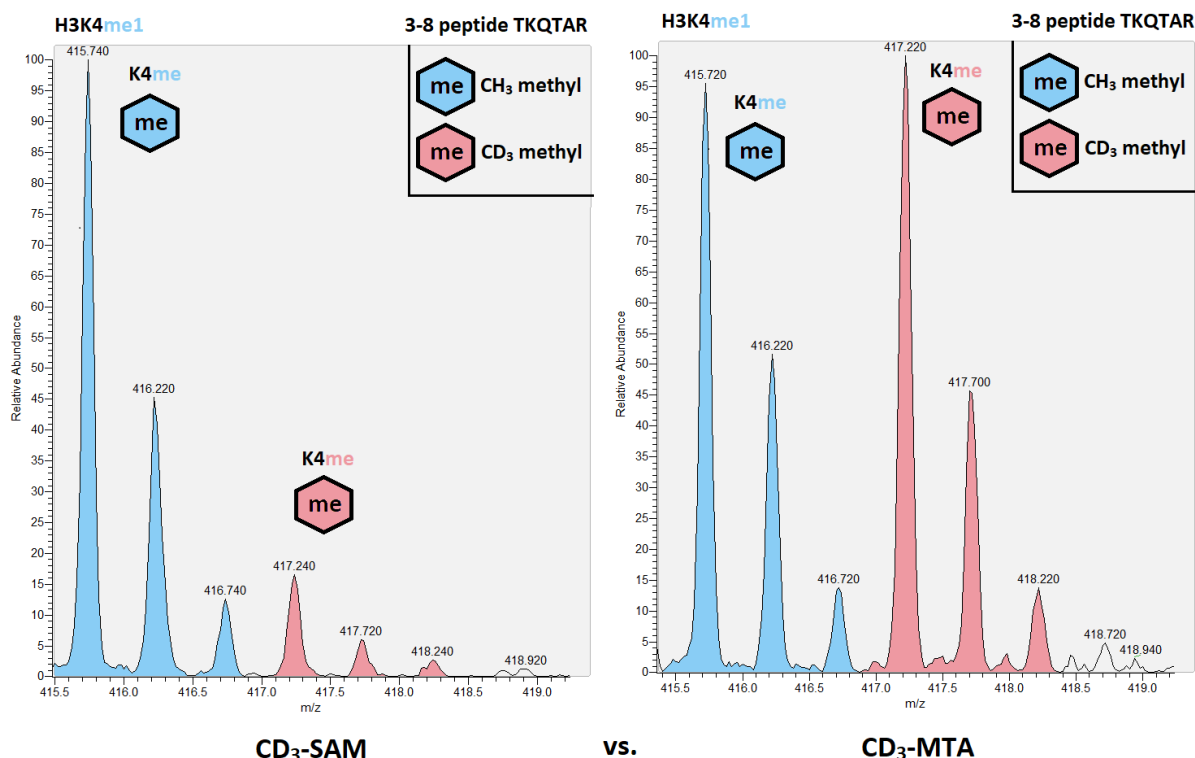


Figure 37 – MS spectra showing the CD₃ incorporation on the monomethylation mark H3K4 when incubating HEK-293T cells with 200 μ M CD₃-SAM (left spectrum) vs. 200 μ M CD₃-MTA (right spectrum) for 48 h.

The next step was to confirm if MDA-MB-231 cells could be labelled with CD₃-MTA. MDA-MB-231 cells are deficient in the MTAP gene and therefore the methionine salvage pathway should not function as it does in HEK-293T cells. If this is the case no labelling should be detected when incubating MDA cells with CD₃-MTA (these experiments are detailed in Section 3.3). As expected, no labelling was seen on any of the H3 lysine methylation sites, which is shown in Figure 38 for the MS spectra of monomethylation mark on H3K4. These results also likely explain the absence of labelling seen when MDA-MB-231 cells were incubated with CD₃-SAM. If CD₃-SAM is not cell permeable and the CD₃-MTA produced from degradation cannot be utilised then no labelling will be observed in these assays.

Another experiment was performed in which the media of MDA-MB-231 cells was changed to methionine-depleted media and then incubated with 200 μ M CD₃-MTA for 20 and 48 hours. There is no exogenous methionine in this assay and if the methionine salvage pathway cannot utilise the CD₃-MTA added, CD₃-MTA will accumulate over time and no methionine will be produced. Methionine is essential for cell viability and it has been suggested that human cells require greater than 25 μ M to maintain cell growth.¹⁸⁵ Also, as discussed earlier, MTA is toxic to cancer cells at high concentrations (μ M) and interferes with cell proliferation, lymphocyte activation, tumour development,

invasiveness and the regulation of apoptosis.¹⁷⁸ The combined effect of the depletion of methionine and the toxicity of MTA was observed when the MDA-MB-231 cancer cell line became non-viable and detached from the cell culture plate in both experiments.

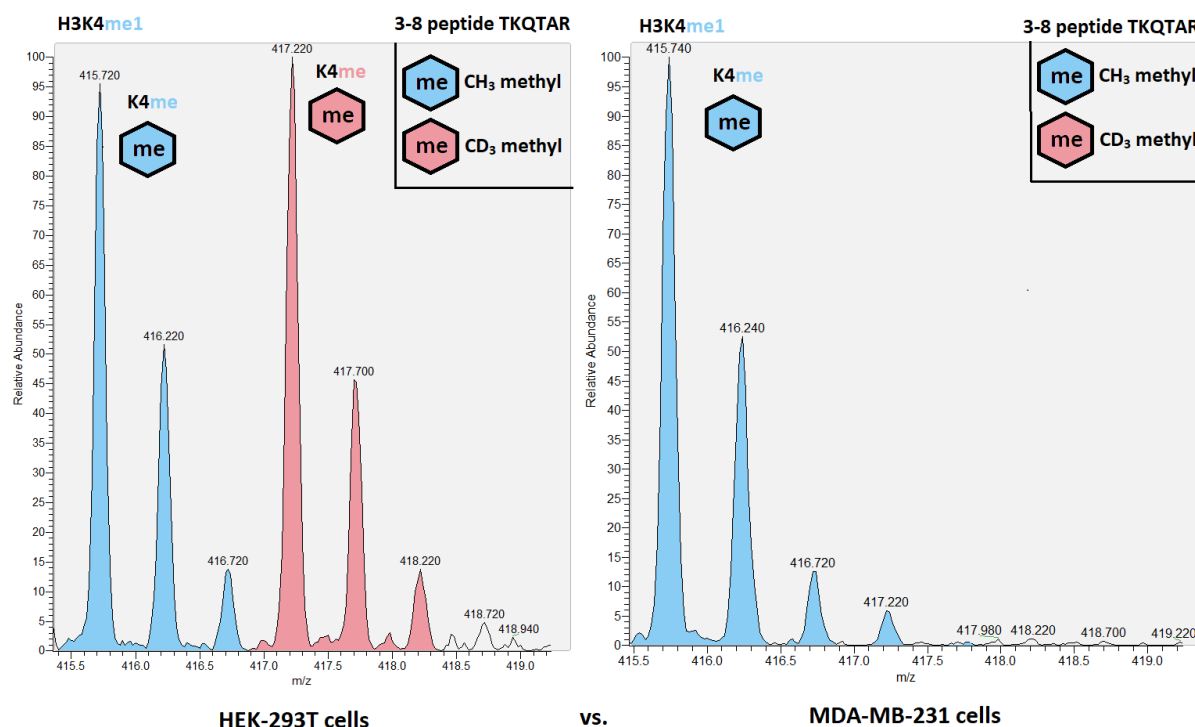


Figure 38 – MS spectra showing CD₃ incorporation on the methylation mark H3K4me1 when incubating HEK-293T cells with 200 μ M CD₃-MTA (left spectrum) vs. no CD₃ labelling observed when incubating MDA-MB-231 cells with 200 μ M CD₃-MTA for 48 h.

The MDA-MB-231 cells were incubated with ¹³CD₃-methionine (¹³CD₃-MET) to ensure that labelling can occur on histones using a ‘heavy’ label in this system and to additionally confirm that the absence of labelling from CD₃-SAM and CD₃-MTA is due to the reasons discussed. Methionine is cell permeable and it does not rely on the methionine salvage pathway to label histones, thus making it a good control for this purpose. After incubation of MDA-MB-231 cells with ¹³CD₃-MET for 20 and 48 hours, labelling was observed with the ¹³CD₃ label on all lysine sites of H3. Two examples of the ‘heavy’ labelling observed can be seen in the MS spectra given in Figure 39 for the 20 h timescale. Consequently, there is good evidence that the problems associated with the cell assays in Section 3 are due to low cell permeability of the SAM analogues and the degradation of SAM to MTA.

One possible reason that no sufficient labelling was observed with the SAM analogues modified at the C-2 position is due to their low cell permeability. The C-2 modified analogues degrade to their

respective MTA analogues; however, the C-2 modified MTA analogues may not be able to be utilised by the endogenous MTAP enzyme. If MTAP cannot utilise the C-2 modified MTA in the first catalytic step of the methionine salvage pathway, then CD₃-SAM cannot be generated. Due to time limitations, a sufficient amount of C-2 modified CD₃-labelled MTA analogue was unable to be produced. Therefore, it would be beneficial to investigate whether any of the MTA analogues can be utilised by endogenous MTAP and thus label histones in cells via CD₃-SAM as future work. The bumped functionality at the C-2 position would be lost in this process; however, it could explain the lack of specificity observed in the cellular assays if the MTA analogue was being utilised instead to provide CD₃-SAM *in vivo*.

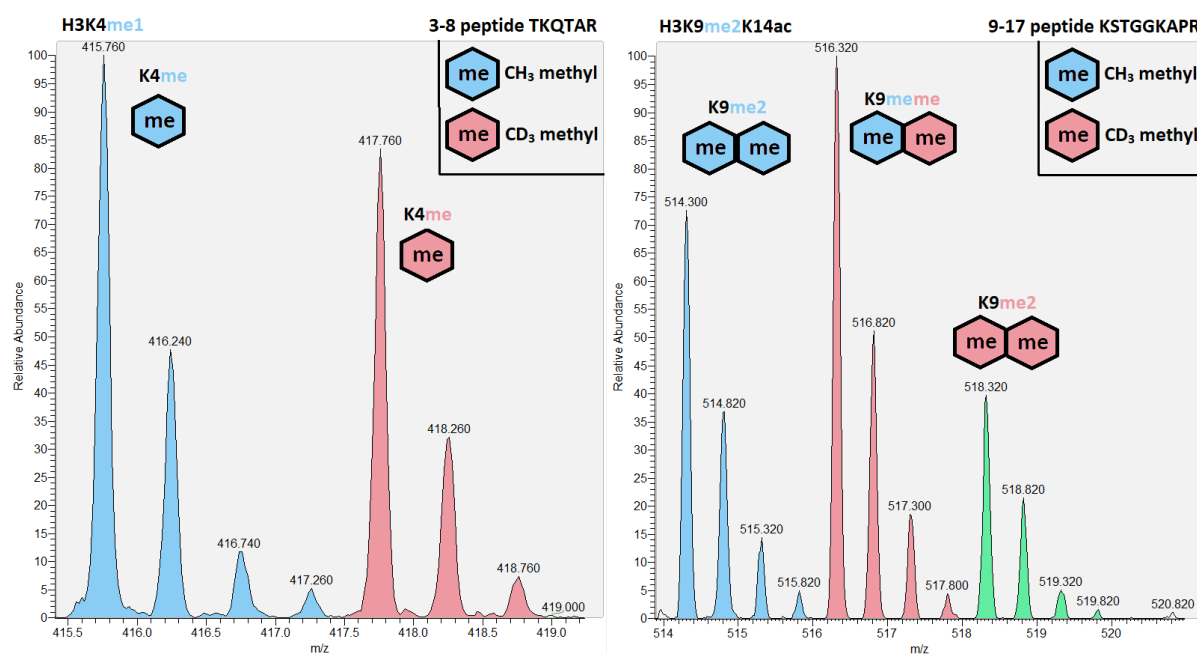


Figure 39 – MS spectra showing ¹³CD₃ incorporation on the methylation marks H3K4me1 (left) and H3K9me2K14ac (right) for the cell assays incubating MDA-MB-231 cells with ¹³CD₃-MET for 20 h.

4.3. Investigating two different methionine pathways in tandem using isotopic labels

It has been confirmed in Section 4.2 that histones can be labelled in cellular assays using both MTA and methionine in separate experiments via two different methionine sources (media or from the MSP). MTA utilises the methionine salvage pathway to produce methionine and subsequently SAM and methionine can be converted to SAM in just one step. To investigate the two sources of

methionine and how they contribute to histone methylation, it was theorised that by incubating cells with differentially labelled metabolites we could observe the separate contributions over time. To track the turnover of methylation on lysine residues of histones in unsynchronised growing cells, HEK-293T cells were transferred into media containing exclusively CD₃-MTA and ¹³CD₃-MET as the two methionine sources. CD₃-MTA was synthesised by degradation of CD₃-SAM as discussed previously and ¹³CD₃-MET was purchased from commercial sources. The cells were then harvested at five different time points – 2, 6, 12, 20 and 48 hours. The samples were analysed by a high resolution Orbitrap mass spectrometer to ensure suitable resolution between the two isotopic labels. As an initial proof-of-concept experiment, a concentration of 100 µM was chosen for each metabolite (CD₃-MTA and ¹³CD₃-MET). In this way, the two isotopic labels can be directly compared to each other and the combined concentration will equal the concentration of L-methionine used in normal Dulbecco's Modified Eagle's Medium (DMEM) media.

An example of the MS data produced from the assays incubating HEK-293T cells with 1:1 CD₃-MTA:¹³CD₃-MET is shown in Figure 40. The MS spectrum in Figure 40 was generated from the 20-hour experiment and it corresponds to the monomethylation mark on H3K4. The peak at *m/z* 830.48 relates to the existing endogenous methylation at the K4 site with incorporation of the CD₃ and ¹³CD₃ isotopic labels, labelled in blue and red respectively (*m/z* 833.50 and 834.50). The incorporation of both isotopic labels is observed across all the H3 lysine methylation marks and on the histone 4 mark H4K20. The incorporation was seen for all methylation states (mono-, di- and tri-) across all the H3 methylation marks with the spectra getting more complicated the higher the methylation state.

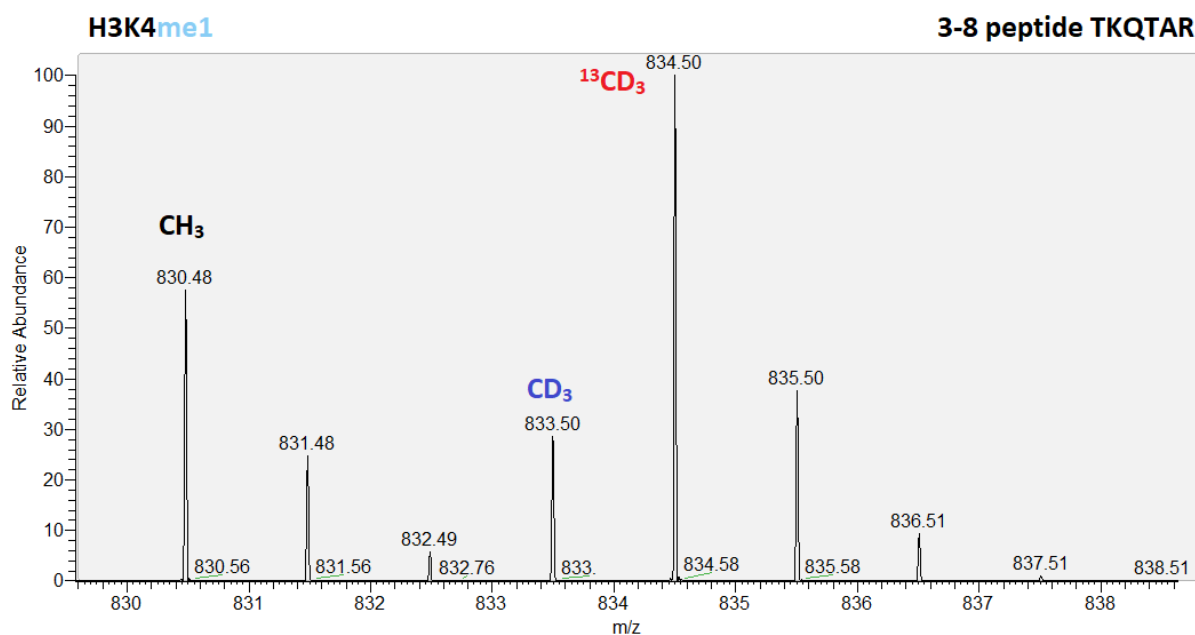


Figure 40 – MS spectrum showing CD₃ incorporation on the monomethylation mark H3K4 for the assay incubating HEK-293T cells with 1:1 CD₃-MTA:¹³CD₃-MET for 20 h. The CD₃ incorporation is labelled in blue and the ¹³CD₃ incorporation is labelled in red.

An example highlighting this complexity is given in Figure 41 for the peptide sequence with dimethylation on both lysine residues K27 and K36. As you can see from this spectrum, there is the incorporation of various degrees of each individual isotopic label but there are also combinations of the two isotopes (CD₃ and ¹³CD₃) together. These combinations make the analysis much more difficult to elucidate the ratio of the different isotopic labels because of the significant overlap of the two contributions. An algorithm could be implemented here to decouple the two contribution areas for the various methylation states. However, for these initial experiments only the lower methylation states were analysed to look at the general trend over time for the two methionine sources.

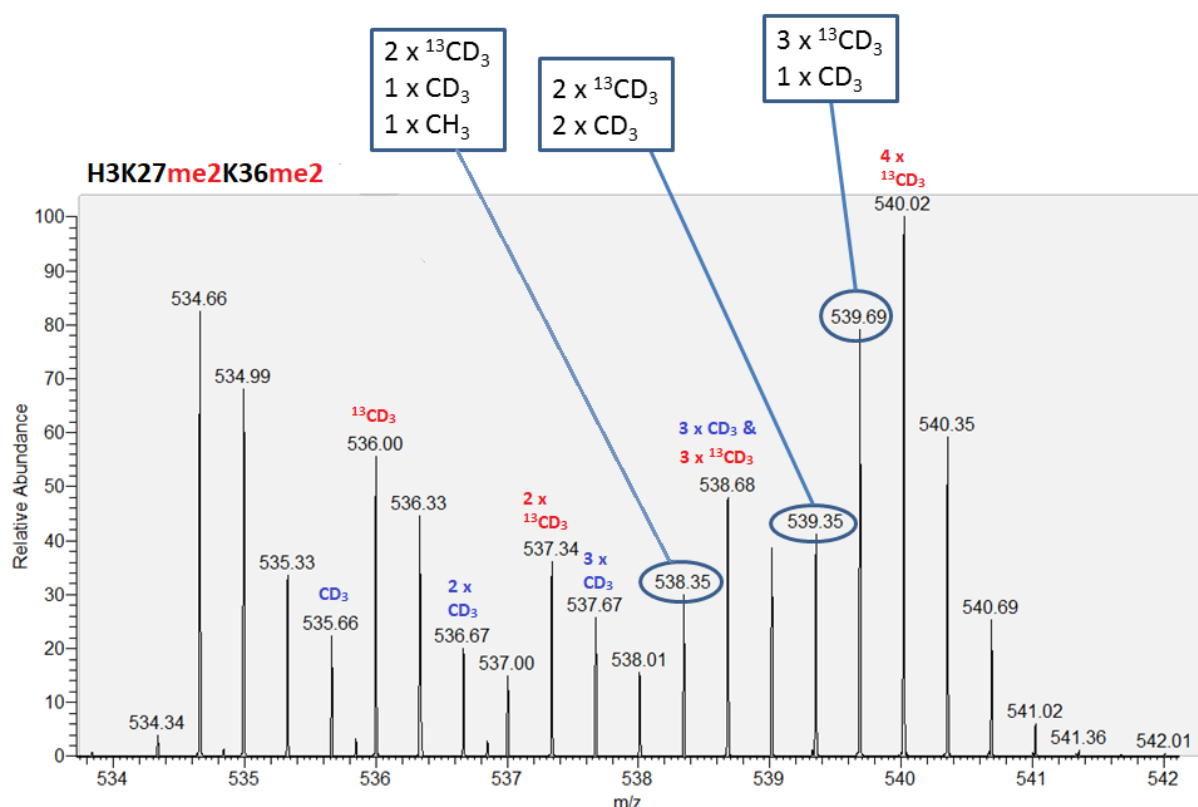


Figure 41 – MS spectrum showing CD₃ incorporation on both H3K27me2 and H3K36me2 methylation marks for the assay incubating HEK-293T cells with 1:1 CD₃-MTA:¹³CD₃-MET for 20 h. The CD₃ incorporation is labelled in blue and the ¹³CD₃ incorporation is labelled in red. Some of the various isotopic combinations are highlighted in the boxes.

To observe if there is any preference for any specific methylation mark/state using one methionine source over the other, the ratios of incorporation of the two isotopes were plotted over time. The ratio of the CD₃ and ¹³CD₃ incorporation on the dimethylation marks for K9, K27, K36 and K79 are shown in Figure 42. After incubation of HEK-293T cells with both isotopically labelled metabolites at a ratio of 1:1 for 48 hours, the ratio of CD₃ incorporation ranged from ~10–20% with the ratio of ¹³CD₃ incorporation ranging from ~80–90% (see Figure 42). These ratios were expected because the CD₃ label that is incorporated via MTA and the MSP pathway includes six additional enzymatic steps to yield isotopically labelled SAM (versus ¹³CD₃-MET) and therefore will have a smaller contribution to histone methylation. From this data, it seems there may be a higher contribution of methylation at K79 via the MSP by around 10%. One possible reason for this higher contribution could be linked to the slower turnover rate of K79 thus allowing for the production of CD₃-SAM via the MSP.¹⁸⁶ At this early stage it is difficult to give a definite answer and therefore more investigation will have to be completed to validate if there is a link between the MSP and H3K79 methylation. It is important to note that a very similar trend is also observed for the monomethylation marks on these H3 lysine sites as well as the dimethylation marks.

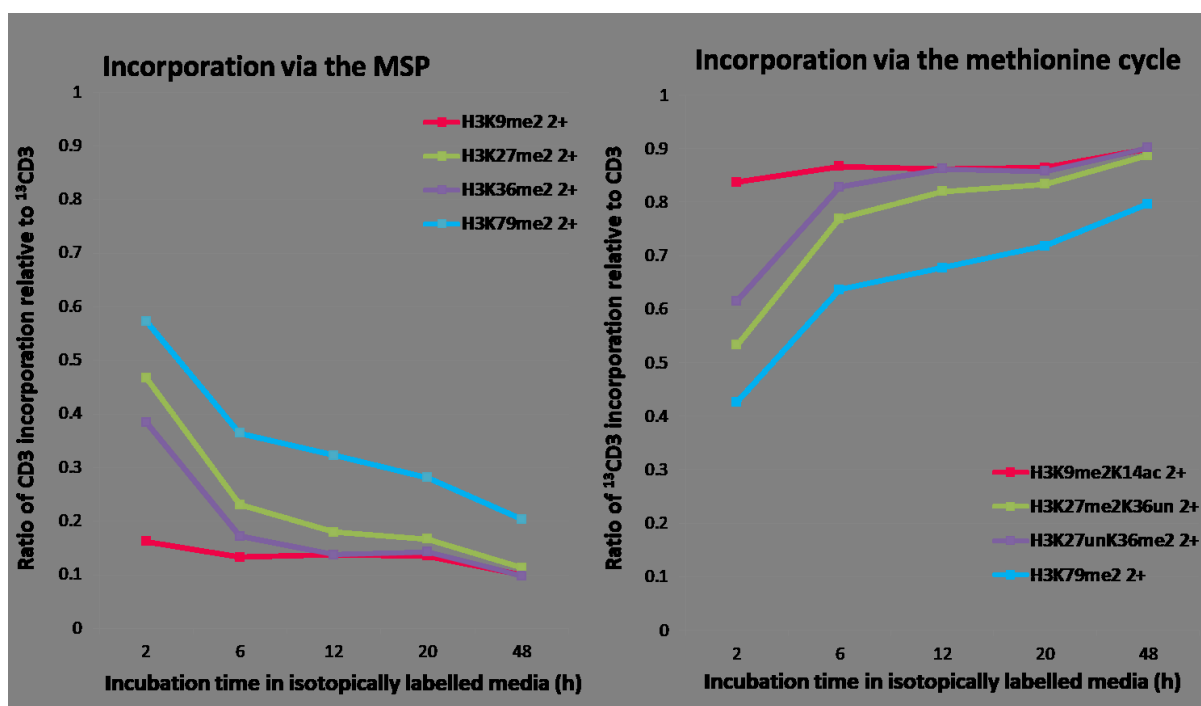


Figure 42 – Incorporation of the two isotopic labels (1:1, 100 μ M:100 μ M, CD_3 : $^{13}\text{CD}_3$) over time in HEK-293T cells. N.B. Trend lines have been added for clarity and do not signify a mathematical estimate of the trend course. NB. n=1 (proof of concept).

Following on from these assays, the ratio of the isotopically labelled metabolites was altered to 3:1 CD_3 -MTA: $^{13}\text{CD}_3$ -MET in an attempt to boost the contribution of the MSP to histone methylation. To boost the MSP contribution further, the overall concentration was lowered for each metabolite to 75 μ M:25 μ M. By lowering the overall pool of methionine available to the cells, it should push the cells to utilise MTA more in histone methylation reactions. The rest of the cell assay conditions were kept identical and the ratios for each isotope over time are shown in Figure 43. The time course experiments show that the contribution of histone methylation via the MSP has indeed increased at the higher 3:1 ratio. At 48 hours, the contribution via MSP was 20–30% compared to 10–20% obtained from the 1:1 ratio experiments. In agreement with the 1:1 experiments, methylation on H3K79 had a higher contribution via the MSP than the rest of the H3 lysine targets. Another interesting trend that appeared out of these experiments is the spike in contribution via the MSP at 20 hours and a corresponding reduction in contribution via methionine at the same time point.

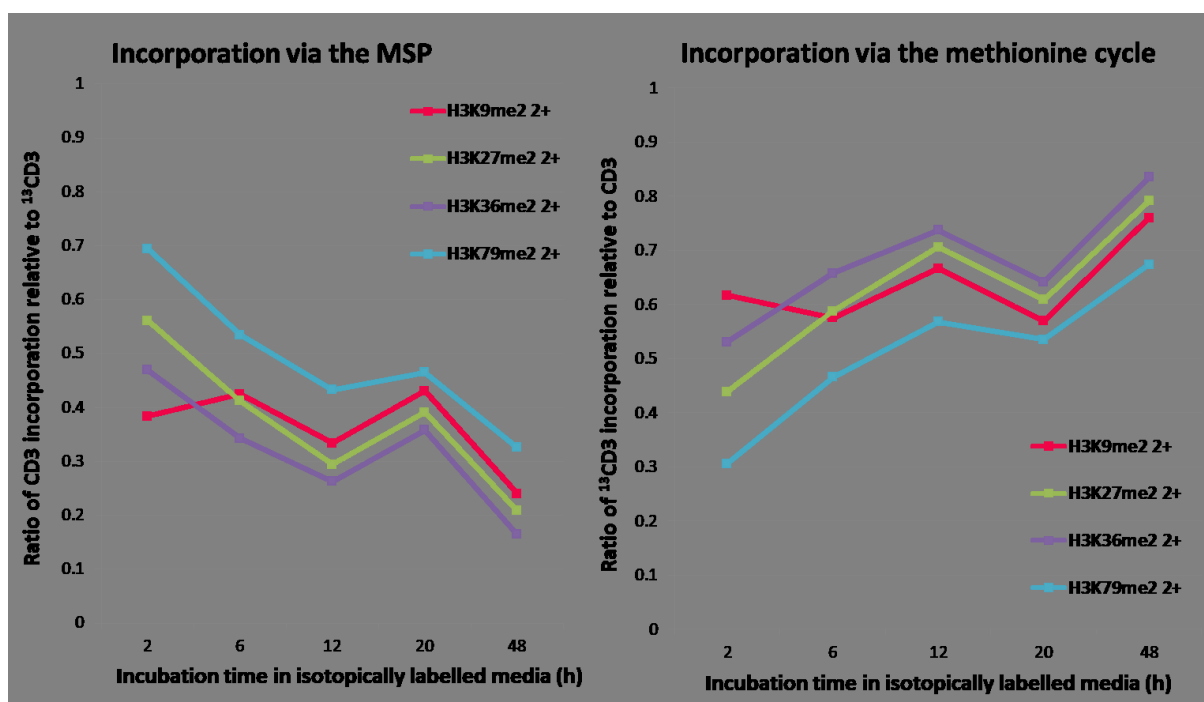


Figure 43 – Incorporation of the two isotopic labels (3:1, 75 μ M:25 μ M CD_3 : $^{13}\text{CD}_3$) over time in HEK-293T cells. N.B. Trend lines have been added for clarity and do not signify a mathematical estimate of the trend course. NB. n=1 (proof of concept).

The time course experiment was repeated at the 3:1 ratio to see if a similar trend can be observed showing a reduction in the methionine contribution at 20 hours. The results from this time course experiment are shown in Figure 44, and comprises the ratio of each isotope (CD_3 and $^{13}\text{CD}_3$) incorporated at the methylation marks utilised in all previous figures (H3K9, K27, K36 and K79). In agreement with the initial 3:1 CD_3 -MTA: $^{13}\text{CD}_3$ -MET experiment, the contribution via the MSP was at 20–30% and methylation on H3K79 has a higher contribution via the MSP than the rest of the H3 lysine targets. Unlike the initial experiment (Figure 43), the reduction in the incorporation via the methionine cycle and increase in incorporation via the MSP at 20 hours was less pronounced; however, it is still present.

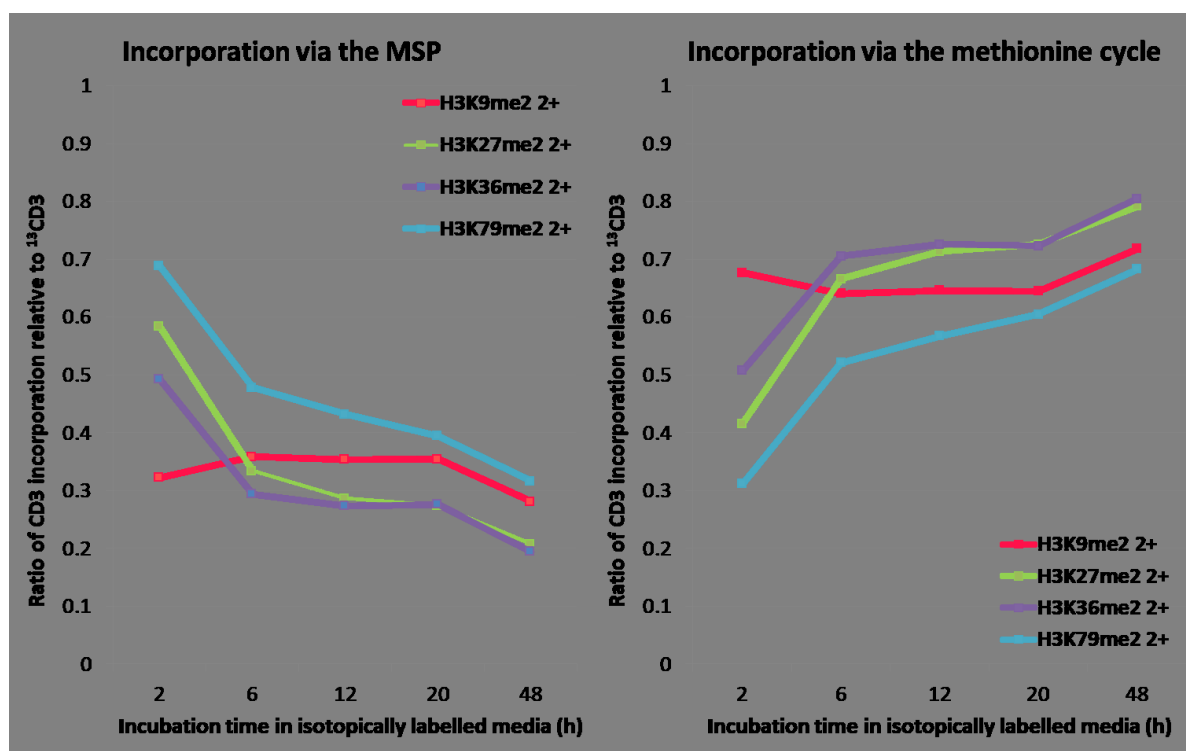


Figure 44 – Repeat experiment of the incorporation of the two isotopic labels (3:1, 75 μ M:25 μ M CD_3 : $^{13}\text{CD}_3$) over time in HEK-293T cells. N.B. Trend lines have been added for clarity and do not signify a mathematical estimate of the trend course. NB. n=1 (proof of concept).

These are just initial experiments and therefore a definite reason of why this trend occurs at this time point cannot be given. However, one could speculate that the trend is related to the stage of the cell cycle the cells are in at the time of the assay. If the cells were in S phase, then the cells will need methionine primarily for DNA synthesis.¹⁸⁷ During S-phase, more histones are synthesised to accommodate the new DNA and hence more histone methylation is required; especially K9 and K27 dimethylation since the majority of chromatin is heterochromatin.¹⁸⁸ Furthermore, rates of protein production during S-phase are very low and therefore a significant amount of methionine will be utilised directly for the production of SAM. However, during G_1 and G_2 , there is a large amount of protein production and thus requires a significant amount of methionine. The heterochromatin marks will not change much during this period, but there may be changes in histone methylation arising from transcription. Since methionine is being heavily used in translation, one could hypothesise that the MSP is utilised more to produce SAM and ensure histone methylation processes are still functioning at their 'usual' rate. Therefore, these histone marks will see an increase in methylation produced from SAM originating from MTA.

The HEK-293T cells used in all these experiments were unsynchronised and therefore it is unknown what stage of the cell cycle they were in when harvested. It is important to note that the cells in this

repeat experiment were less confluent than the initial experiment and therefore when they were harvested at 20 hours the trend may have been missed due to the cells being in a different stage of the cycle. In all these experiments we were only observing snapshots in time and not a continuous time course and this could be why the trend was missed in this repeat experiment. Further investigation of this trend could involve performing experiments on synchronised cells at various stages of the cell cycle to see what effect the cell cycle has on the contributions from both methionine sources.

5. Conclusions and future work

The synthesis of a small library of 'heavy' CD₃-labelled SAM cofactors modified at the C-2 position was successfully completed at ≥95% purity. The discovery of preferential substitution of the homocysteine side chain occurring at the C-2 position of chlorinated adenosine (**6**) led to the utilisation of a nucleophilic substitution reaction to install substituents at the C-2 position. This resulted in the successful introduction of the first three substituents to the C-2 position of the adenine ring – piperidine, methylpiperazine and thiophenyl. After introduction of the homocysteine side chain, a 'heavy' methyl group (CD₃) was introduced to the sulphur position through optimisation of a protocol for the *in situ* generation of the strong methylation reagent, methyl triflate (MeOTf). This provided a readily detectable group via LC-MS/MS analysis in subsequent biological assays. For the introduction of the phenyl group, a palladium cross-coupling reaction was optimised to provide the fourth SAM analogue in the series. This methodology could be considered for future synthesis of SAM analogues.

The activities of the small library of cofactor probes were then validated by a simple *in vitro* assay utilising mononucleosomes. Once the SETD7 and G9a assays were optimised with CD₃-SAM, these assays were then repeated for the four SAM analogues. LC-MS/MS analysis demonstrated that all four SAM analogues methylated at H3K4 via SETD7, but afforded little methylation on H3K9 when G9a was utilised. This significant drop in methylation observed for G9a's target compared to SETD7's provides evidence for the SAM analogues selectivity for SETD7. Future work will entail testing a wider range of HKMTs to establish an improved selectivity profile for the probes synthesised. From the *in vitro* assays performed, the piperidine analogue seems to have the best selectivity profile and thus should be continued further for future biological assays.

Cellular assays were then optimised with CD₃-SAM, which resulted in successful labelling on all histone lysine sites processed from HEK-293T cells; albeit at low quantities. However, no labelling was observed for any of the four SAM analogues under the same conditions in this cell line. Even with overexpression of active SETD7 in HEK-293T cells, this did not increase labelling at H3K4 with either CD₃-SAM or the SAM analogues. These results provide evidence that H3K4 is not a valid target for SETD7 in a physiological environment despite being catalytically active against H3K4 on mononucleosomes. Moreover, investigation into breast cancer cells lines (MDA-MB-231 and MCF7), which are reported to have endogenous expression of SETD7, did not show any 'heavy' labelling. In contrast to HEK-293T cells, no labelling was seen with both breast cancer cell lines even with incubation of CD₃-SAM. This then prompted an investigation into the degradation of CD₃-SAM to

investigate this disparity. The degradation study of CD₃-SAM showed that CD₃-MTA was formed in significant quantities under equivalent conditions to the cellular assays. This also occurred with the SAM analogues modified at the C-2 position, which all degraded to their MTA equivalent. Since MTA is involved in the methionine salvage pathway and is more cell permeable than SAM, this was likely the source of SAM used for methylation in the cell assays. Since MDA-MB-231 and MCF7 cell lines do not have a functioning methionine salvage pathway, this explains why no labelling was observed with these systems.

In light of the permeability issues of SAM, further investigation into the *in nucleio* assays should be performed to confirm that SETD7 cannot utilise chromatin for methylation at H3K4. The *in nucleio* assays do not have the issues of cell permeability and have been optimised for 'heavy' labelling with CD₃-SAM. Experiments should be performed with nuclei derived from HEK-293T cells transfected with SETD7 to see if an increase of labelling at H3K4 can be observed. Also, *in nucleio* assays could be performed using nuclei from the breast cancer lines (MDA-MB-231 or MCF7 cells) which have been reported to have endogenous expression of SETD7.

It is debated in the literature that H3K4 may not be a physiologically valid target for SETD7 and no evidence for methylation at this position with SETD7 was observed in cellular assays overexpressing SETD7. However, there has been substantial evidence that SETD7 methylates at a wide range of non-histone proteins and therefore a small selection were tested for validation. This resulted in two successful methylation hits being discovered for FOXO1 and FOXO3. These non-histone targets should be further analysed in a more complex environment to confirm they are valid targets for SETD7. Moreover, the third non-histone target RelA should be investigated further to see if acetylation at the reported methylation lysine position (K508) or alternative PTM crosstalk is blocking successful 'heavy' labelling occurring in the *in vitro* assays performed.

Another way to discover more novel targets of SETD7 was to utilise the most selective 2-piperidine SAM analogue in lysate experiments; which also provides a more complex environment. In order to get relevant data from the lysate studies that were performed, an optimised enrichment method protocol must be developed. This method would ensure that any novel targets could be elucidated over background data. A post-doc in the group is optimising this protocol and once this is established, lysate experiments will be processed and analysed for any successful methylation hits by SETD7. Any successful hits arising from these experiments can then be further validated to result in a physiological profile of SETD7's targets.

A separate investigation was instigated into how the two confirmed sources of methionine in both assays (MTA and methionine) contribute to histone methylation, which resulted in some interesting trends. At 20 h, there was a decrease in contribution from methionine which corresponded with an increase in contribution from MTA and the methionine salvage pathway. One possible explanation for this is that the contributions from both sources are related to the stages of the cell cycle. One way this research should be continued would be to observe how these contributions change at different stages of the cell cycle. One way to synchronise the cells would be to serum starve them, causing the arrest of the cells in G1 phase. The incorporation of a FUCCI stain would provide a method to easily determine which stage of the cell cycle the cells were in prior to their analysis. However, this method also involves genetic modification of the cell line which might perturb endogenous expression.

6. Methods

6.1. General materials and instrumentation

All cell lines (HEK-293T, MDA-MB-231 and MCF7) were purchased from ATCC and grown in a Nuaire incubator at 37 °C with 5% CO₂. The cells were handled in a Nuaire LabGuard Class II biological safety cabinet and visualised using a CETI microscope (Medline Scientific). The cells were cultured in Dulbeccos Modified Eagle Medium (DMEM + GlutaMAX; Gibco, Life Technologies) with 10% fetal bovine serum and 1% Penicillin-Streptomycin (Gibco, Life Technologies). The cells were split with 0.5% Trypsin-EDTA solution (Gibco, Life Technologies).

For lipofectamine transfection, Lipofectamine 2000 (Invitrogen, Thermo Fisher Scientific) was used and OptiMEM media (Gibco, Life Technologies). The FLAG-SETD7 plasmid was from AddGene plasmid #24084 (pcDNA Set9 wt was a gift from Danny Reinberg).³⁸ ANTI-FLAG M2 Affinity Gel for the FLAG immunoprecipitation was from Sigma Aldrich. The VP35 plasmid was obtained with thanks from the Hugh Brady group at Imperial College London.

All buffer salts and additives were purchased from Sigma Aldrich apart from EDTA (Promega) and NP-40 alternative (CalBioChem) and all buffers prepared with distilled, deionised water (Siemens Ultra Clear water purifier (Triple Red)). The pH was adjusted with an Accumet AB200 pH meter (Fisher Scientific) and then the buffers were stored at 4 °C until use. The inhibitors used when mentioned are all from Sigma Aldrich and include: sodium butyrate, microcystin LR, 4-benzenesulfonyl fluoride hydrochloride (AEBSF) or alternatively HALT protease and phosphatase inhibitor cocktail (100X, Thermo Fisher Scientific).

All centrifugation was performed using a Haraeus MegaFuge 40R centrifuge (Thermo Scientific) for 15–50 mL Falcon Tubes (VWR) or on a 5424R microcentrifuge (Eppendorf) for microcentrifuge tubes (Eppendorf). When heating of eppendorfs was necessary, a Dri-Block DB-2D heating block (Technique, Bibby Scientific) was used. Douncing in the *in nucleio* assays was performed with a 7 mL glass dounce homogenizer (Fisher Brand). All sonication was performed using a FB15055 sonicator (Fisher Scientific).

For the *in vitro* assays, the recombinant proteins were from the following companies: SETD7 (Abcam), FOXO1 and FOXO3 and PARP1 (Cambridge Bioscience, Human HeLa mononucleosomes (AMS Biotechnology)).

For all biological experiments reported in this thesis, three technical repeats were performed for each experiment. Moreover, experiments were repeated in triplicate unless otherwise specified.

6.2. *In nucleo* assays

6.2.1. Cell culture

HEK-293T cells were grown in DMEM + GlutaMAX medium supplemented with 10% fetal bovine serum and 1% Penicillin-Streptomycin. The cells were grown to 70% confluency in 50 cm³ plates and then subsequently harvested using cold (4 °C) PBS and cell scrapers.

6.2.2 *In nucleo* assay

The protocol followed a previous procedure.¹⁴⁷ The protease, phosphatase and histone deacetylase (HDAC) inhibitors used in the buffer solutions were microcystin, AEBSF and sodium butyrate. The harvested cells are washed twice more with cold PBS (4 °C). The cells were then washed twice with 1 mL buffer I (10 mM Hepes-NaOH (pH 7.9), 50 mM KCl, 1.5 mM MgCl₂, 0.5 mM DTT, inhibitors) and centrifuged at 4 °C, 600g, 5 min. Swollen cells were resuspended in 200 µL buffer II (buffer I, 0.1% (v/v) NP-40) and incubated on ice for 10 min before douncing (5 strokes). The nuclei were then collected by centrifugation (4 °C, 600g, 5 min). The nuclei were then washed twice in 200 µL of buffer III (20 mM Hepes-NaOH (pH 7.9), 50 mM KCl, 1.5 mM MgCl₂, 0.5 mM DTT, 0.2 mM EDTA, 20% (v/v) sucrose, inhibitors). Nuclei were resuspended in 200 µL cold NA buffer (nuclear acetylation buffer) supplemented with 1 mM ATP (50 mM Tris-HCl (pH 8.0), 50 mM NaCl, 1 mM EDTA, 0.5 mM DTT, 10% (w/v) sucrose) and then transferred to rt for 5 min. Then 50 µL of 50 µM S-adenosyl-L-(methyl-d₃)methionine or analogue in NM buffer (nuclear methylation buffer; 50 mM Tris-HCl (pH 8.0), 50 mM NaCl, 10 mM MgCl₂, 0.5 mM DTT, 10% (w/v) sucrose) was added to commence the reaction. Nuclei were incubated at 37 °C for 90 min and gently agitated intermittently to keep nuclei in suspension. The nuclei were then washed once with NM buffer with inhibitors and kept at -80 °C ready for histone extraction (see Section 6.8.2).

6.3. Cell assays

6.3.1. Cell culture

HEK-293T, MDA-MB-231 or MCF7 cells were grown in DMEM + GlutaMAX medium supplemented with 10% fetal bovine serum and 1% Penicillin-Streptomycin. The cells were grown to 50% confluency in 50 cm³ plates or 6-well plates. One plate or one well of cells was used for each experiment. The cells were incubated for the indicated time and concentration of CD₃-SAM, CD₃-SAM analogue, CD₃-MTA or ¹³CD₃-L-methionine as described in the results. Cells are harvested at 90–100% confluency by washing with cold (4 °C) PBS and subsequent lifting of the cells with a cell scraper. The harvested cells were washed twice more with cold PBS (4 °C) and then taken forward straightaway to the chromatin isolation step (see Section 6.3.2).

6.3.2. Chromatin isolation

Cell pellets were resuspended in 1 mL NIB buffer (15 mM Tris-HCl (pH 7.5), 60 mM KCl, 15 mM NaCl, 5 mM MgCl₂, 1 mM CaCl₂, 250 mM sucrose) plus NP-40 [0.3% (v/v) added fresh] and incubated on ice for 5 min. The chromatin was then collected by centrifugation (4 °C, 600g, 5 min). The chromatin was then washed twice in 1 mL NIB buffer without NP-40 and the pellet was either taken straightaway to the histone extraction step or kept at -80 °C ready for histone extraction (Section 6.8.2.).

6.4. Transfections

6.4.1. Calcium phosphate transfection

HEK-293T cells were grown to 50% confluency in DMEM + GlutaMAX medium supplemented with 10% fetal bovine serum and 1% Penicillin-Streptomycin in 50cm³ plates. One plate of cells was used for each experiment. 50, 200 or 500 µM of CD₃-SAM or 15 µg of SETD7 plasmid was mixed with 0.5 mL 0.25 M CaCl₂ solution. Equal volume of HBS solution (50 mM HEPES, 280 mM NaCl, 1.5 mM Na₂HPO₄, 12 mM dextrose (pH 7) was added to the mixture and the mixture was thoroughly bubbled with air using a pipette. The mixture was further mixed by shaking vigorously until frothy and then let sit for 15 min. 1 mL of each mixture with the varying concentrations of CD₃-SAM was added to

each plate of cells. The cells were incubated for 24 h and then the contents of these plates were harvested with cold (4 °C) PBS. The harvested cells were washed twice more with cold PBS (4 °C) and then taken forward straightaway to the chromatin isolation step (see Section 6.3.2.).

6.4.2. Lipofectamine transfection

HEK-293T cells were grown to 60–70% confluency in DMEM + GlutaMAX medium supplemented with 10% fetal bovine serum and 1% Penicillin-Streptomycin in 50 cm³ plates. For each 50 cm³ plate to transfect, a total of 24 µg of plasmid was added (14:10 µg SETD7:VP35 plasmids) to 1.5 mL of OptiMEM media. In a separate tube, 60 µL of Lipofectamine 2000 was added to 1.5 mL OptiMEM media and both tubes were incubated for 5 min at rt. After this time, the tubes were combined and mixed and incubated for a further 20 min at rt. During this time the media was changed to 7 mL DMEM + GlutaMAX + FBS w/o Pen-Strep. After 20 min, 3mL of transfection media was added to each plate and after 5–6 h the media was changed to normal DMEM plus supplements. The cells were then harvested after 48 h of transfection with cold (4°C) PBS.

6.5. FLAG immunoprecipitation

6.5.1. Preparation of lysates

Each 50 cm³ plate of HEK-293T cells harvested from lipofectamine transfection (see Section 6.4.2.) were resuspended in 1 mL two-step lysis buffer (20 mM Tris.HCl (pH 7.4), 10 mM MgCl₂, 10 mM KCl, 2 mM EDTA, 10% glycerol, 1% Triton X-100, HALT protease inhibitors (added fresh)) and incubated on ice for 10 min. The cells were then sonicated for 30 sec and NaCl was added to a final concentration of 420 mM. After incubation for 60 min rotating at 4 °C, the cell lysate was then sonicated for a further 30 sec. The lysate was then centrifuged at 10,000 g for 30 min at 4 °C and the supernatant was recovered for FLAG immunoprecipitation. The lysate was stored at -80°C until purification.

6.5.2. FLAG pull-down

1 mL of HEK-293T lysate (for preparation of the lysates see Section 6.5.1.) was diluted with 1:1 low salt TBS (50 mM Tris-HCl (pH 7.4), 100 mM NaCl, 0.01% (v/v) Triton X-100) and split into two

samples. During this time, anti-FLAG M2 affinity gel resin was washed 3 x 0.5 mL TBS (50 mM Tris-HCl (pH 7.4), 150 mM NaCl). Each sample was then mixed with 20 µL of pre-washed anti-FLAG M2 affinity gel beads and left to rotate overnight at 4°C. After centrifugation at 5,000 g for 30 sec, at 4°C, the supernatant was removed and retained as unbound material (for subsequent SDS-PAGE analysis). The samples were then washed 3 x low salt TBS (first wash retained for SDS-PAGE) and then 3 x high salt TBS (50 mM Tris-HCl (pH 7.4), 300 mM NaCl, 0.01% (v/v) Triton X-100). 150 µL of elution buffer (150 ng/µL FLAG peptide in high salt TBS) was added and incubated at 4°C for 1 h. Resin was spun down (5,000 g, 30 sec) and elution fraction collected. Two more elution fractions were collected by adding 150 µL fresh elution buffer for overnight, rotating at 4°C and subsequently for 6 h, rotating at 4 °C. Elution fractions were combined and concentrated using Amicon Ultra 0.5 mL centrifuge filters with 30 kDa cut off (Millipore). Resin was denatured to check elution by adding 150 µL of 2x sample buffer (125 mM Tris-HCl (pH 6.8), 2% (w/v) SDS, 25% (v/v) glycerol, 0.005% (w/v) bromophenol blue) was added to resin and incubated for 5–10 min. Resin was spun down and supernatant collected. 5% β-mercaptoethanol (β-ME) was added and boiled at 95 °C for 2 min. All other samples (unbound, first wash and elution) were mixed 1:1 with 2x sample buffer with 5% β-ME and boiled at 95 °C for 2 min. All samples were kept at -20°C until gel was run.

6.5.3. SDS-PAGE

Protein samples were treated with mercaptoethanol (Sigma) and Laemmli 2 × Sample Buffer (Sigma) at 95 °C for 3 min, and then resolved on an 8% SDS-PAGE polyacrylamide gel (100 V for 60 min). The protein standard used was Precision Plus Protein™ Dual Color (Bio-Rad). To visualise whole protein content, gels were stained with imperial protein stain (Thermo Scientific) for 1 h using manufacturer's instructions. The gel was then destained with H₂O (2 x 30 min).

6.5.4. In-gel digest

In-gel digestion was carried out following a previous protocol.¹⁸⁹ The band of interest was cut from the gel using a scalpel, diced into 1mm³ pieces and transferred to a 1.5 mL Eppendorf. 100 µL of 100 mM ammonium bicarbonate/acetonitrile (1:1, v/v) was added to the sample and shaken vigorously for 30 min. 500 µL of neat acetonitrile was added at rt and shook until the gel pieces become white and shrink and then acetonitrile was removed. Enough trypsin buffer (13 ng µL⁻¹ trypsin in 10 mM ammonium bicarbonate containing 10% v/v acetonitrile) was added to cover the gel pieces (approximately 50 µL) and incubated on ice. After 30 min, the sample was checked to see if the

solution had been fully absorbed and sufficient trypsin buffer was added to ensure gel pieces were covered. Once the gel pieces were incubated for 120 min, 10–20 μL of 100 mM ammonium bicarbonate was added. Samples were then incubated for 6 h at 37 °C. Once the sample was cooled to rt and spun down at 5,000 g for 30 sec, a 1–1.5 μL aliquot was taken for reference. 100 μL of extraction buffer (1:2, 5% v/v formic acid/acetonitrile) was added to the sample and incubated for 15 min at 37 °C. The supernatant was collected and dried down ready for MS/MS analysis.

6.6. *In vitro* assay

To confirm whether they are non-histone targets of SETD7, 10 μg of FOXO3, FOXO1 or PARP1 were incubated with 200 nM SETD7 and 1 mM $\text{CD}_3\text{-SAM}$ in KMT buffer (50 mM Tris-HCl (pH 8.8), 10 mM MgCl_2 , 4mM DTT (added fresh)) for 1 h at 37 °C. The total assay volume was 200 μL . The reaction was stopped by the addition of cold SAH (1 mM, dissolved in 1% FA (formic acid)).

The reaction was then dried down and either directly subjected to in solution digest (Section 6.6.1) or subjected to two rounds of propionylation (see details in Section 6.8.3) and then in solution digest directly afterwards.

6.6.1. In-solution digest

Protein sample was dried down *in vacuo* and then redissolved in 50 μL 100 mM ammonium bicarbonate and checked pH was ~ 8 . 10 μL of 30 mM DTT was added to the sample, vortexed, span down and left at 51 °C for 1 h. After cooling down, 10 μL of 100 mM iodoacetamide was added, pH was checked to be ~ 7.8 , and left at rt in dark for 45 min. Reaction was quenched with 2 μL of 30 mM DTT. Trypsin digest of the proteins was achieved by adding 1.5 μg of trypsin for 6 h at 37°C and subsequently quenched with 2 μL of glacial acetic acid.

6.7. Lysate assay

HEK-293T cells were transfected with SETD7 plasmid at 70% confluency using the Lipofectamine method described in Section 6.4.2. After 48 h of transfection, the cells were harvested, washed twice with PBS and incubated on ice in 2-step lysis buffer (20 mM Tris.HCl (pH 7.4), 10 mM MgCl_2 , 10 mM KCl, 2 mM EDTA, 10% glycerol, 1% Triton X-100, HALT protease inhibitors (added fresh)) for 15 min.

The sample was then sonicated, NaCl was added to a final concentration of 420 mM and incubated for a further 60 min rotating at 4 °C. The sample was sonicated again and centrifuged at 10,000 g for 30 min. The resultant supernatant was concentrated in 10 kDa centrifugal filters (Vivaspin®, GE Healthcare Life Sciences) and then KMT buffer (50 mM Tris-HCl (pH 8.8), 10 mM MgCl₂, 4mM DTT (added fresh)) was added and the solution concentrated again. The sample was resuspended in KMT buffer and 1 mM Pip-CD₃-SAM was added to start the reaction and immediately incubated for 1 h at 37 °C. 10 µg or 2 µg mononucleosomes and/or commercial SETD7 (200 nM) was added in additional control experiments. After 1 h, cold SAH (1 mM) was added to stop the reaction. The samples were stored at -80 °C for further processing.

6.8. Mass spectrometry work-flow

6.8.1. General instrumentation and materials

For LC-MS/MS analysis, samples were loaded onto a LTQ Velos Pro linear ion trap LC-MS system (Thermo Scientific) with an ESI primary ionisation method (200–2000 m/z) and secondary CID fragmentation (collision energy of 35%). The Xcalibur (Thermo Scientific) software was used to create methods that could perform zoom and targeted scans for peaks of interest with known mass and approximate retention time. Xcalibur was also used to visualise and analyse the MS data. Proteome Discovery 1.3 (Thermo Scientific) software was used for SEQUEST database searches to search for methylation events on non-histone proteins. For high resolution LC-MS/MS analysis, samples were loaded onto a LTQ Velos Pro orbitrap LC-MS system (Thermo Scientific) using a UHPLC nanoViper column (20 µM × 550 mm, Thermo Scientific).

A histone standard was run before any histone samples, which was prepared from HEK-293T cells without any 'heavy' CD₃-labelled incubation, to estimate the timings of the zoom scans. Tryptic digest was performed using sequencing grade modified trypsin (Promega). For stage tip clean-up, C-18 extraction discs (Empire, Bioanalytical Technologies) were used to prepare the handmade tips. All solvents were LC-MS grade (Sigma Aldrich or Fisher Scientific) and all other chemicals and materials were purchased from Sigma Aldrich or VWR unless otherwise specified.

6.8.2. Histone extraction

The frozen nuclei were thawed on ice (~ 4 °C) and 0.4 N H₂SO₄ solution (500 µL) was then added slowly, vortexed and rotated in a cold room for 90 min. Extracted histones were then centrifuged at

21,000 g for 5 min and the supernatant containing the histones was transferred to a 1.5 mL Eppendorf. The histones were precipitated by adding 125 μ L 100% TCA and kept on ice for 1 h. Histones were pelleted by centrifugation (21,000 g, 5 min) and washed once with 500 μ L acetone + 0.1% HCl (-20 °C). The histones were washed once more with 500 μ L acetone (-20 °C) and air dried after centrifugation (21,000 g, 5 min).

6.8.3. Propionylation and trypsin digest

This protocol is derived from previous work by Garcia *et al.*¹⁹⁰ The dry histone pellet (see histone extraction in Section 6.8.2.) was then resuspended in 10–100 μ L 100 mM ammonium bicarbonate (NH_4HCO_3) and then 10 μ L was sonicated for 10 min and brought through to propionylation. The pH of the sample was tested and adjusted to pH 7–9 using ammonium hydroxide (NH_4OH) and then 5 μ L (or half volume) of propionic anhydride reagent was added (1:3, propionic anhydride:isopropanol). Immediately after, 3–5 μ L NH_4OH was added to the sample to make sure pH is $\sim$ 8 and then incubated for 15 min at 37 °C. After incubation, the sample was dried to <5 μ L and the whole propionylation procedure was repeated.

Once dried, the sample was made up to 50 μ L in 100 mM NH_4HCO_3 (pH 8.5) and trypsin was added in a 1:20 trypsin:histone ratio and left at 37 °C for 6 h for digestion. The reaction was stopped by freezing at -80 °C. The sample was then dried to <5 μ L and made up to 10 μ L of 100 mM NH_4HCO_3 ready for a second round of propionylation.

Propionylation was repeated twice more and dried down ready for stage tip clean up (Section 6.8.4.). If peptide standards were added these were added just before the second round of propionylation steps and after digestion.

6.8.4 Sample desalting using stage tips

Stage tips were made by punching out one and a half discs of C18 extraction material using a P1000 enlarged tip. This material was then loaded into the bottom of a P200 tip and the tip inserted into the top of 1.5 mL microcentrifuge tubes. The C18 putty was washed and activated by eluting through 200 μ L MeOH by centrifugation (800 g, 5 min). The putty was then washed through with 200 μ L 0.1% AcOH solution (800 g, 5 min) and the two elutions were discarded. The histone peptide sample was then reconstituted in 200 μ L 0.1% AcOH solution (pH $\sim$ 3–4) and loaded on to the putty (800 g, 5 min). The peptides were then eluted from the putty into a fresh 1.5 mL Lo-bind microcentrifuge tube

with 200 μ L elution buffer (75% MeCN, 20% H₂O, 5% FA, 800 g, 5 min). The sample was then dried ready for MS analysis.

7. Experimental

7.1 General Experimental

All reagents were obtained from commercial sources and were used as purchased from the manufacturers. Anhydrous solvents were purchased from Sigma-Aldrich, UK, stored over 4 Å molecular sieves and under an Ar atmosphere or were freshly distilled from calcium hydride (dichloromethane, toluene, triethylamine, pyridine) and sodium benzophenone (tetrahydrofuran). Reactions were monitored by TLC using pre-coated aluminium-backed ICN silica plated (60A F254) and visualised by UV detection (at 254 nm) and staining with potassium permanganate or ninhydrin solution and then heated with a heat gun. Column chromatography was performed on Fluka 60 (40 – 60 µm mesh) silica gel.

All reactions were conducted under normal conditions and at ambient temperature (20 to 25 °C) unless specified otherwise, with magnetic stirring. Evaporation and concentration under reduced pressure was performed using a Büchi rotary evaporator at 40 °C. Residual solvent was removed under high vacuum (1 mbar).

Infrared spectra were recorded neat on a Perkin Elmer Spectrum 100 fitted with a universal ATR accessory. The intensity of each band is described as s (strong), m (medium) or w (weak), and with the prefix v (very) and suffix br (broad) where appropriate. ¹H NMR spectra were recorded at 400 MHz, using a Bruker AV 400 spectrometer. ¹³C NMR spectra were recorded at 100 MHz, using a Bruker AV 400 spectrometer. Chemical shifts are reported as δ values (ppm) referenced to the following solvent signals: **CDCl₃** δ_H (CHCl₃) = 7.26, δ_C = 77.0; **CD₃OD** δ_H (CHD₂OD) = 3.34, δ_C = 49.0; **CD₃SOCD₃** δ_H (CD₃SOCHD₂) = 2.50; δ_C = 77.0; **D₂O** δ_H (DHO) = 4.80. The term 'stack' is used to describe a region where resonances arising from non-equivalent nuclei are coincident, and multiplet, m, to describe a region where a resonance arises from a single nucleus (or equivalent nuclei) in which coupling constants cannot be readily measured. Coupling constants are measured in Hertz (Hz). Connectivities were deduced from COSY90, HSQC and HMBC experiments.

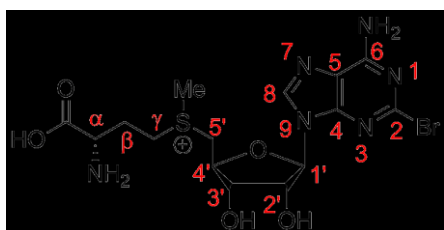
High and low resolution mass spectra were recorded by the Mass Spectrometry Service of Imperial College London, Department of Chemistry using Micromass Platform II and Micromass AutoSpec-Q or on a LTQ Velos Pro linear ion trap spectrometer (DiMaggio group). The samples were eluted over a linear gradient of 100% solvent A (99.9% H₂O, 0.1% FA) to 5% A and 95% B (95% MeCN, 0.4% H₂O,

0.1% FA) over 10 min, at which point the gradient remained isocratic for 2 min. Melting points were recorded using a Reichert-Thermovar melting point apparatus.

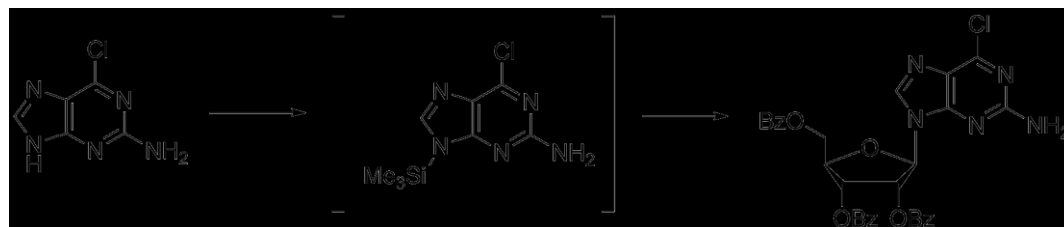
All crude final SAM analogues were purified by RP-HPLC. The compounds were injected onto a Supelco Ascentis Express C18 (2.7 μm particle size, 150 mm $\times$ 4.6 mm, Sigma Aldrich) and were resolved and fractionated by UV detection (215 nm) with a Dionex Ultimate 3000 UHPLC system (Thermo Scientific). The flow rate was 0.5 mL/min and a multi-step gradient was used starting at 0.5% B (95% MeCN, 4.8% H₂O and 0.2% TFA) and 99.5% A (5% MeCN, 94.8% H₂O and 0.2% TFA) for two minutes, then linearly increased to 20% B over 15 minutes, followed by linear increase to 95% B over a further 7 minutes. All solvents used were HPLC grade and purchased from either Sigma Aldrich or Thermo Fisher).

7.2 Experimental Procedures

Numbering sample for NMR data (2-bromo-SAM shown here)



(2*R*,3*R*,4*R*,5*R*)-2-(2-amino-6-chloro-9*H*-purin-9-yl)-5-((benzoyloxy)methyl)tetrahydro-furan-3,4-diyl dibenzoate (**2**)



N,O-Bis(trimethylsilyl)acetamide (BSA, 6.34 mL, 31.2 mmol) was added to a stirred suspension of 2-amino-6-chloro-purine (**1**, 4.80 g, 28.3 mmol) in dry acetonitrile (150 mL). After stirring at rt for 45 min, 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl- β -D-ribofuranose (16.00 g, 31.7 mmol) and trimethylsilyl trifluoromethanesulfonate (TMSOTf, 6.80 mL, 30.6 mmol) were added sequentially. The reaction mixture was stirred at rt for 15 min, after which the flask was transferred to a preheated 80 $^{\circ}\text{C}$ oil

bath and stirred for 1 h. Then the mixture was cooled to rt and diluted with EtOAc (100 mL). The organic phase was sequentially washed with aqueous saturated NaHCO₃ solution and brine, dried with Na₂SO₄, filtered and concentrated under reduced pressure to afford the crude nucleoside. Purification by column chromatography (33–100% EtOAc/pentane gradient) afforded protected adenosine **2** as a white foam (11.83 g, 68%): *R*_f = 0.33 (pentane/ethyl acetate 2:1); m.p. 97 – 101 °C; $\nu_{\max}$ (neat)/cm⁻¹ = 3483w (N-H), 3373w, 3204w (=C-H), 2919w (C-H), 1721s (C=O), 1611s (N-H bending), 1562s, 1512s, 1452s, 1262s, 1091s, 1069s; ¹H NMR (400 MHz, CDCl₃): δ = 4.70 (dd, *J* = 12.2, 4.4 Hz, 1 H, 5'-CH₂), 4.83 – 4.86 (m, 1 H, 4'-CH), 4.94 (dd, *J* = 12.2, 3.7 Hz, 1 H, 5'-CH₂), 5.17 (br s, 2 H, 2-NH₂), 6.26 (d, *J* = 4.5 Hz, 1 H, 1'-CH), 6.39 (dd, *J* = 5.4, 3.2 Hz, 1 H, 3'-CH), 6.44 – 6.47 (m, 1 H, 2'-CH), 7.40 – 7.47 (m, 6 H, aryl CH), 7.57 – 7.62 (m, 3 H, aryl CH), 7.91 (s, 1 H, 8-CH), 7.91 – 8.05 (m, 6 H, aryl CH); ¹³C NMR (400 MHz, CDCl₃): δ = 63.2 (CH), 71.3 (CH), 73.5 (CH), 80.3 (CH), 87.3 (CH), 125.8, [128.5, 128.6, 128.7 (CH, CH-aryl, resonance overlap)], 129.2, [129.7, 129.8, 129.9, 130.0 (CH, CH-aryl, resonance overlap)], [133.5, 133.7, 133.9 (C, C-aryl, resonance overlap)], 141.1 (C), 151.9 (C), 153.1 (C), 159.1 (C), 165.1 (C, C=O), 165.3 (C); MS (TOF ES+) *m/z* 614.1 ([M+H]⁺, 100%); HRMS (TOF ES+) calcd for C₃₁H₂₅N₅O₇Cl [M+H]⁺ 614.1443, found 614.1442.

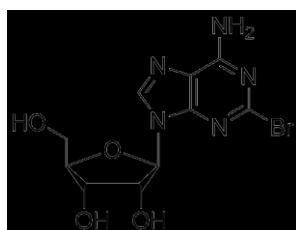
(2*R*,3*R*,4*R*,5*R*)-2-(2-bromo-6-chloro-9*H*-purin-9-yl)-5-((benzoyloxy)methyl)tetrahydro-furan-3,4-diyl dibenzoate (**3**)¹⁹¹



TMS-Br (26.50 mL, 173.3 mmol, 9 eq.) was added dropwise to a stirred solution of adenosine **2** (11.82 g, 19.3 mmol, 1 eq.) in CH₂Br₂ (150 mL) under N₂. TBN (39.70 mL, 385.0 mmol, 20 eq.) was then added dropwise and stirred at ambient temperature for 1 h (TLC). The mixture was cooled for 15 min and then added dropwise to a cold, vigorously stirred mixture of NaHCO₃/H₂O (200 mL)/CH₂Cl₂ (200 mL). The aqueous layer was extracted and the combined organic phase was washed (H₂O) and dried (Na₂SO₄). Solvent was evaporated *in vacuo* and the orange oil was purified by column chromatography (pentane/EtOAc, 2:1) to afford adenosine **1** as a white solid (8.82 g, 68%): *R*_f = 0.64 (pentane/ethyl acetate, 2:1); m.p. 89 – 93 °C; $\nu_{\max}$ (neat)/cm⁻¹ = 3079w (=C-H), 1722s (C=O), 1587s, 1553s, 1352s, 1261s, 1091s, 1069s, 1025s; ¹H NMR (400 MHz, CDCl₃): δ = 4.71 (dd, *J* = 12.1, 4.1 Hz, 1 H, 5'-CH₂), 4.84 – 4.87 (m, 1 H, 4'-CH), 4.91 (dd, *J* = 12.1, 3.3 Hz, 1 H, 5'-CH₂), 6.10 – 6.15

(stack, 2 H, 3'-CH, 2'-CH), 6.47 (d, J = 4.9 Hz, 1 H, 1'-CH), 7.34 – 7.46 (m, 6 H, Ar-H), 7.52 – 7.61 (m, 3 H, Ar-H), 7.90 – 7.93 (m, 2 H, Ar-H), 8.00 – 8.06 (m, 4 H, Ar-H), 8.24 (s, 1 H, 8-CH); ^{13}C NMR (400 MHz, CDCl_3): δ = 63.5 (CH_2), 71.6 (CH), 74.4 (CH), 81.6 (CH), 87.1 (CH), 128.1 (CH), [128.6, 1287, 129.1 (CH, resonance overlap)], [129.6, 129.9, 130.0 (CH, resonance overlap)], 131.7 (C), [133.7, 133.9, 134.0 (CH, resonance overlap)], [143.6, 143.6 (CH, resonance overlap)] 152.0 (C), 152.5 (C), 165.2 (C=O), 165.3 (C=O), 166.1 (C=O); MS (TOF ES+) m/z 677.0 ($[\text{M}+\text{H}]^+$, 48%); HRMS (TOF ES+) calcd for $\text{C}_{31}\text{H}_{23}\text{N}_5\text{O}_7\text{ClBr}$ $[\text{M}+\text{H}]^+$ 677.0439, found 677.0447.

(2*R*,3*R*,4*R*,5*R*)-2-(2-bromo-6-amino-9*H*-purin-9-yl)-5-(hydroxymethyl)tetrahydro-furan-3,4-diol (**4**)¹⁹²



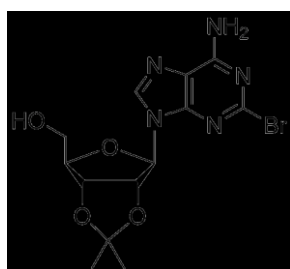
A suspension of **3** (2.00 g, 2.954 mmol) in NH_3/MeOH (ca. 7 *N*, 20 mL) was prepared in a microwave vial and stirred at rt for 10 mins. It was then transferred to a microwave reactor where it was heated at 120 °C and stirred for 1 h. The solvent was evaporated under reduced pressure to afford the crude nucleoside as an orange residue. Purification by column chromatography with sequential elutions of CH_2Cl_2 , $\text{CH}_2\text{Cl}_2/\text{MeOH}$ [7 *M* NH_3] (93:7) and $\text{CH}_2\text{Cl}_2/\text{MeOH}$ [7 *M* NH_3] (9:1) provided **4** as a pale orange amorphous solid (674 mg, 66%): R_f = 0.52 ($\text{CH}_2\text{Cl}_2/\text{MeOH}$ 85:15); m.p. 131 – 135 °C; ν_{max} (neat)/ cm^{-1} = br 3129s (O-H), 1643s, 1594s, 1396s, 1351s, 1296s, 1281s, 1083s, 1043s; ^1H NMR (400 MHz, $\text{DMSO}-d_6$): δ = 3.56 – 3.67 (m, 2 H, 5'- CH_2), 3.92 – 3.96 (m, 1 H, 4'-CH), 4.09 – 4.14 (m, 1 H, 3'-CH), 4.49 – 4.53 (m, 1 H, 2'-CH), 5.05 (t, J = 5.6 Hz, 1 H, 5'-OH), 5.21 (d, J = 4.9 Hz, 1 H, 3'-OH), 5.48 (d, J = 6.2 Hz, 1 H, 2'-OH) 5.81 (d, J = 6.0 Hz, 1 H, 1'-CH), 7.86 (br s, 2 H, 6- NH_2) 8.36 (s, 1 H, 8-CH); ^{13}C NMR (400 MHz, $\text{DMSO}-d_6$): δ = 61.9 (CH_2 , 5'- CH_2), 71.0 (CH, 3'-CH), 74.1 (CH, 2'-CH), 86.2 (CH, 4'-CH), 87.8 (CH, 1'-CH), 119.0 (C, 5-C), 140.3 (CH, 8-CH), 144.7 (C, 4-C), 150.7 (C, 2-C), 157.1 (C, 6-C); MS (TOF ES+) m/z 346.0 ($[\text{M}+\text{H}]^+$, 100%); HRMS (TOF ES+) calcd for $\text{C}_{10}\text{H}_{12}\text{N}_5\text{O}_4\text{Br}$ $[\text{M}+\text{H}]^+$ 346.0151, found 346.0163.

(2*R*,3*R*,4*R*,5*R*)-2-(2-bromo-6-amino-9*H*-purin-9-yl)-5-(chloromethyl)tetrahydro-furan-3,4-diol (**6**)



A suspension of **4** (100 mg, 0.289 mmol, 1 eq.) in pyridine (92 μ L, 0.580 mmol, 2 eq.) and acetonitrile (3 mL) was cooled to 0 °C. Thionyl chloride (155 μ L, 1.300 mmol, 5 eq.) was added dropwise with stirring and stirring was continued for a further 4 h. Then the solution was warmed to rt and stirred a further 16 h. The reaction mixture was evaporated *in vacuo* and the resulting suspension was re-dissolved in MeOH, H₂O (5.5 mL, 5:1) and NH₄OH (1 mL) and stirred for 0.5 h at rt. The reaction mixture was then concentrated under reduced pressure to afford nucleoside **6** as a light brown solid (370 mg, 87%): *R*_f = 0.79 (CH₂Cl₂/MeOH 6:1); m.p. 189 – 191 °C; ¹H NMR (400 MHz, DMSO-*d*₆): δ = 3.83 – 3.96 (m, 2 H, 5'-CH₂), 4.08 – 4.13 (m, 1 H, 4'-CH), 4.16 – 4.19 (m, 1 H, 3'-CH), 4.65 – 4.69 (m, 1 H, 2'-CH), 5.50 (d, *J* = 5.1 Hz, 1 H, 3'-OH), 5.65 (d, *J* = 6.5 Hz, 1 H, 2'-OH), 5.85 (d, *J* = 6.1 Hz, 1'-CH), 7.92 (br s, 2 H, 6-NH₂), 8.36 (s, 1 H, 8-CH); ¹³C NMR (400 MHz, DMSO-*d*₆): δ = 43.8 (5'-CH₂), 71.3 (3'-CH), 73.5 (2'-CH), 84.1 (4'-CH), 88.8 (1'-CH), 98.0 (5-C), 128.3 (8-CH), 144.6 (4-C), 156.5 (2-C), 157.1 (6-C); MS (TOF ES+) *m/z* 365.0 ([M+H]⁺, 100%), 364.0, 368.0; HRMS (TOF ES+) calcd for C₁₀H₁₁N₅O₃BrCl [M+H]⁺ 364.9789, found 364.9834.

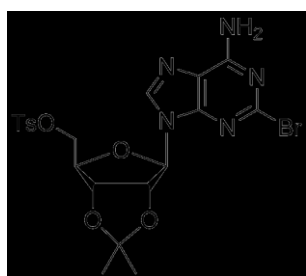
((3*aR*,4*R*,6*R*,6*aR*)-6-(2-bromo-6-amino-9*H*-purin-9-yl)-2,2-dimethyl-tetrahydro-furo[3,4-*d*][1,3]dioxol-4-yl)methanol (**8**)



Adenosine **4** (0.560 g, 1.623 mmol) was suspended in acetone (22 mL) containing *p*-TsOH monohydrate (1.018 g, 5.357 mmol) and then triethylorthoformate (5.6 mL) was added over a period of 10–20 min at rt with vigorous mechanical stirring. After 18 h, water (19 mL) and concentrated ammonium hydroxide (0.43 mL) were added to adjust pH to ~7–8. Volatiles were

evaporated and purification by column chromatography ($\text{CH}_2\text{Cl}_2/\text{MeOH}$, 95:5) afforded acetone **8** as an orange hygroscopic solid (340 mg, 55%): $R_f = 0.78$ ($\text{CH}_2\text{Cl}_2/\text{MeOH}$, 85:15); m.p. 186 – 191 °C; ν_{max} (neat)/ cm^{-1} = br 3322, br 3179, 2937s (C-H), 1638s, 1592s, 1458s, 1374s, 1350s, 1295s, 1206s, 1072s; ^1H NMR (400 MHz, $\text{DMSO}-d_6$): δ = 1.33 (s, 3 H, acetone CH_3), 1.54 (s, 3 H, acetone CH_3), 3.52 – 3.55 (m, 2 H, 5'- CH_2), 4.20 (td, J = 4.8, 2.6 Hz, 1 H, 4'-CH), 4.93 (dd, J = 6.0, 2.6 Hz, 1 H, 3'-CH), 5.08 (t, J = 5.8 Hz, 1 H, 5'-OH), 5.27 (dd, J = 6.0, 2.8 Hz, 1 H, 2'-CH), 6.05 (d, J = 2.8 Hz, 1 H, 1'-CH), 7.88 (br s, 2 H, 6- NH_2), 8.34 (s, 1 H, 8-CH); ^{13}C NMR (400 MHz, $\text{DMSO}-d_6$): δ = 25.7 (CH_3 , acetone CH_3), 27.5 (CH_3 , acetone CH_3), 62.0 (CH_2 , 5'- CH_2), 82.3 (CH, 3'-CH), 84.0 (CH, 2'-CH), 87.2 (CH, 4'-CH), 89.8 (CH, 1'-CH), 113.6 (C, C-(CH_3)₂), 118.5 (C), 140.2 (CH, 8-CH), 156.7 (C), 163.8 (C); MS (TOF ES+) m/z 386.0 ($[\text{M}+\text{H}]^+$, 100%); HRMS (TOF ES+) calcd for $\text{C}_{13}\text{H}_{16}\text{N}_5\text{O}_4\text{Br}$ $[\text{M}+\text{H}]^+$ 386.0464, found 386.0482.

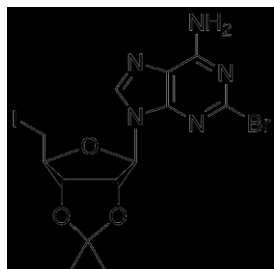
((3aR,4R,6R,6aR)-6-(2-bromo-6-amino-9H-purin-9-yl)-2,2-dimethyl-tetrahydro-furo[3,4-d][1,3]dioxol-4-yl)methyl 4-methylbenzenesulfonate (9)



To a solution of acetone **8** (340 mg, 0.883 mmol) in pyridine (12 mL) was added TsCl (337 mg, 1.766 mmol) and the solution was stirred at rt overnight. The reaction was quenched MeOH, evaporated and worked up. Purification by column chromatography (0–5 % MeOH/ CH_2Cl_2 gradient) afforded tosylate **9** as an orange solid (256 mg, 54%): $R_f = 0.42$ ($\text{CH}_2\text{Cl}_2/\text{MeOH}$, 99:1); m.p. 95 – 100 °C; ν_{max} (neat)/ cm^{-1} = br 3372w, br 3203w, 2927s (C-H), 1721s, 1603s, 1541s, 1443s, 1332s, 1351s, 1298s, 1213s, 1052s, 1023s; ^1H NMR (400 MHz, CDCl_3): δ = 1.36 (s, 3 H, acetone CH_3), 1.57 (s, 3 H, acetone CH_3), 2.38 (s, 3 H, tosyl- CH_3), 4.23 – 4.36 (m, 2 H, 5'- CH_2), 4.43 – 4.47 (m, 1 H, 4'-CH), 4.99 (dd, J = 6.0, 3.4 Hz, 1 H, 3'-CH), 5.22 (dd, J = 6.0, 1.6 Hz, 1 H, 2'-CH), 6.12 (d, J = 1.6 Hz, 1 H, 1'-CH), 7.35 – 7.37 (br m, 2 H, 6- NH_2), 7.55 – 7.57 (m, 2 H, 2 $\times$ aryl-CH), 7.79 – 7.81 (m, 2 H, 2 $\times$ aryl-CH), 8.07 (s, 1 H, 8-CH); ^{13}C NMR (400 MHz, CDCl_3): δ = 21.7 (CH_3 , tosyl- CH_3), 25.4 (CH_3 , acetone CH_3), 27.1 (CH_3 , acetone CH_3), 69.3 (CH_2 , 5'- CH_2), 81.4 (CH, 4'-CH), 84.4 (CH, 3'-CH), 85.1 (CH, 2'-CH), 90.6 (CH, 1'-CH), 114.8 (C, acetone C- CH_3), 118.6 (C, C- NH_2), 126.4 (CH), 127.9 (CH), 129.6 (C), 132.2 (C),

139.8 (C), 144.8 (C), 145.1 (C), 156.0 (C); MS (TOF ES+) m/z 540.1 ($[M+H]^+$, 100%); HRMS (TOF ES+) calcd for $C_{20}H_{23}N_5O_6SBr$ $[M+H]^+$ 540.0552, found 540.0525.

((3aR,4R,6S,6aS)-6-(2-bromo-6-amino-9H-purin-9-yl)-2,2-dimethyl-tetrahydro-furo[3,4-d][1,3]dioxol-4-yl)iodo-methyl (10)



A solution of tosylate **9** (137 mg, 0.253 mmol, 1 eq.) in acetone (5 mL) containing sodium iodide (95 mg, 0.633 mmol, 2.5 eq.) was heated at reflux overnight. The solution was evaporated, dissolved in EtOAc and washed (2 M $Na_2S_2O_3$) and evaporated. The residue was purified by column chromatography (0 – 5% MeOH/ CH_2Cl_2 gradient) which afforded iodo **9** as a light yellow solid (25 mg, 9%): R_f = 0.13 (CH_2Cl_2 /MeOH, 99:1); 1H NMR (400 MHz, $CDCl_3$): δ = 1.37 (s, 3 H, acetonide CH_3), 1.62 (s, 3 H, acetonide CH_3), 3.96 – 4.00 (m, 2 H, 5'- CH_2), 4.50 – 4.51 (m, 1 H, 4'- CH), 5.09 – 5.11 (m, 1 H, 3'- CH), 5.15 – 5.18 (m, 1 H, 2'- CH), 5.79 (d, J = 6.3 Hz, 1 H, 1'- CH), 6.04 (br s, 2 H, 6- NH_2), 7.79 (s, 1 H, 8- CH); ^{13}C NMR (400 MHz, $CDCl_3$): δ = 13.1 (CH_2 , 5'- CH_2), 25.1 (CH_3 , acetonide CH_3), 27.2 (CH_3 , acetonide CH_3), 81.7 (CH), 83.4 (CH), 86.6 (CH), 89.1 (CH), 113.4 (C), 118.2 (C), 139.2 (CH), 156.6 (C), 163.4 (C); MS (TOF ES+) m/z 495.1 ($[M+H]^+$, 50%), 497.1, 498.1; HRMS (TOF ES+) calcd for $C_{13}H_{15}N_5O_3BrI$ $[M+H]^+$ 494.9493, found 495.1230.

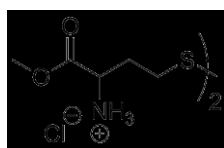
((3aR,4R,6R,6aR)-6-(2-bromo-6-amino-9H-purin-9-yl)-2,2-dimethyl-tetrahydro-furo[3,4-d][1,3]dioxol-4-yl)methyl 4-methanesulfonate (11)



Acetonide **8** (250 mg, 0.649 mmol, 1 eq.) in CH_2Cl_2 (12 mL) was flushed with N_2 for 1 h before methanesulfonyl chloride (60 μ L, 0.779 mmol, 1.2 eq.) and NEt_3 (220 μ L, 1.558 mmol, 2.4 eq.) were added and was stirred at rt for 30 min. Then 10 mL of sat. NH_4Cl -solution was added and the layers

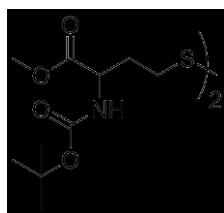
separated. The organic layer was washed two times with brine and the combined organic layers were dried over MgSO_4 and evaporated. Purification by column chromatography (0 – 5% $\text{MeOH}/\text{CH}_2\text{Cl}_2$ gradient) afforded mesylate **11** as a yellow solid (258 mg, 86%): $R_f = 0.45$ ($\text{CH}_2\text{Cl}_2/\text{MeOH}$, 99:1); m.p. 84 – 90 °C; ν_{max} (neat)/ cm^{-1} = br 3327w, br 3175w, 2979s (C-H), 1722s, 1602s, 1558s, 1452s, 1316s, 1264s, 1091s, 1070s, 1025s; ^1H NMR (400 MHz, CDCl_3): δ = 1.38 (s, 3 H, acetonide CH_3), 1.60 (s, 3 H, acetonide CH_3), 2.97 (s, 3 H, mesylate- CH_3), 4.44 – 4.51 (m, 3 H, 4'-CH, 5'- CH_2), 5.09 (dd, 1 H, 3'-CH), 5.30 – 5.32 (m, 1 H, 2'-CH), 5.95 (br s, 2 H, 6- NH_2), 6.06 (d, 1 H, 1'-CH), 7.82 (s, 1 H, 8-CH); ^{13}C NMR (400 MHz, CDCl_3): δ = 25.4 (CH_3 , acetonide CH_3), 27.2 (CH_3 , acetonide CH_3), 37.6 (CH_3 , mesylate CH_3), 68.5 (CH_2 , 5'- CH_2), 81.4 (CH, 3'-CH), 84.1 (CH, 2'-CH), 84.9 (CH, 4'-CH), 90.7 (CH, 1'-CH), 115.0 (C), 119.6 (C), 140.1 (CH, 8-CH), 145.0 (C), 156.0 (C); MS (TOF ES+) m/z 464.0 ($[\text{M}+\text{H}]^+$, 95%), 466.0 ($[\text{M}+\text{H}]^+$, 100%); HRMS (TOF ES+) calcd for $\text{C}_{14}\text{H}_{19}\text{N}_5\text{O}_6\text{SBr}$ $[\text{M}]^+$ 464.0239, found 464.0223.

L-Homocystine methyl ester hydrochloride (**12**)¹²⁵



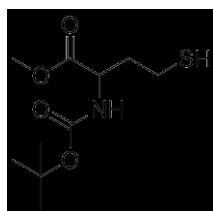
To a solution of *L*-homocystine (0.200 mg, 0.746 mmol, 1 eq.) in methanol (6 mL) at 0 °C was added thionyl chloride (0.54 mL, 4.478 mmol, 6 eq.). The solution was stirred overnight at rt and then removed under reduced pressure. To remove the traces of SOCl_2 the resulting white oil was diluted in CHCl_3 and concentrated again to afford methyl ester **12** as a hygroscopic white amorphous solid (215 mg, 98%): $R_f = 0.63$ ($\text{CH}_2\text{Cl}_2/\text{MeOH}$, 17:3); $[\alpha]^{20}_{\text{D}} = +32.8$ (c 1.00, MeOH); ν_{max} (neat)/ cm^{-1} = br 3330s (N-H), 3075s (C-H), 2936s (C-H), 2853s (C-H), 1723s (C=O), 1590s, 1561s, 1452s, 1369s, 1315s, 1263s, 1092s, 1026s, 707s (C-Br); ^1H NMR (400 MHz, MeOD): δ = 2.36 (m, 4 H, 2 $\times$ γ - CH_2), 2.90 (m, 4 H, 2 $\times$ β - CH_2), 3.88 (s, 6 H, 2 $\times$ CH_3), 4.22 (t, 2 H, 2 $\times$ α -CH), 7.92 (s, 3 H, NH_3); ^{13}C NMR (400 MHz, MeOD): δ = 29.6 (CH_2 , γ - CH_2), 32.2 (CH_2 , β - CH_2), 51.3 (CH_3 , terminal CH_3), 52.5 (CH, α -CH), 169.2 (C, C=O); MS (TOF ES+) m/z 297.1 ($[\text{M}+\text{H}]^+$, 100%).

tert-Butoxycarbonyl *L*-homocystine methyl ester (**13**)¹²⁵



To a solution of protected homocystine **12** (175 mg, 0.591 mmol, 1 eq.) in methanol (10 mL) at 0 °C was added (Boc)₂O (387 mg, 1.774 mmol, 3 eq.) and NEt₃ (0.24 mL, 2.365 mmol, 4 eq.). The solution was stirred overnight at rt. The MeOH was removed under reduced pressure and the residue was diluted in CHCl₃ and washed with aq. HCl (0.1 M). The organic layer was dried (MgSO₄) and removed under reduced pressure. Purification by column chromatography (pentane/EtOAc, 7:3) afforded protected homocystine **13** as an amorphous white solid (376 mg, 89%); R_f = 0.52 (pentane/EtOAc, 7:3); [α]_D²⁰ = +22.8 (c 1.00, CHCl₃); ¹H NMR (400 MHz, CDCl₃): δ = 1.42 (s, 18 H, 2 × C(CH₃)₃), 1.93 – 2.02 (m, 2 H, β-CH₂), 2.19 – 2.26 (m, 2 H, β-CH₂), 2.69 (dd, *J* = 8.2, 6.6 Hz, 4 H, 2 × γ-CH₂), 3.74 (s, 6 H, 2 × CH₃), 4.36 – 4.41 (m, 2 H, 2 × α-CH), 5.09 – 5.11 (br m, 2 H, 2 × NH); ¹³C NMR (400 MHz, CDCl₃): δ = 28.3 (CH₃, C-(CH₃)₃), 32.6 (CH₂, β-C), 34.5 (CH₂, γ-C), 52.5 (CH₃, O-CH₃), 80.2 (C, C-(CH₃)₂), 155.3 (C, C=O), 172.6 (C, C=O); MS (TOF ES+) *m/z* 497.2 ([M+H]⁺, 100%), 498.2, 499.2.

tert-Butoxy-*L*-homocysteine methyl ester (**14**)¹²⁵



Method a)

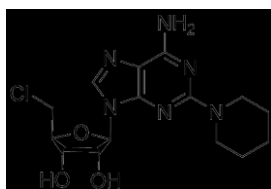
Protected homocystine **13** (50 mg, 0.098 mmol, 1 eq.), triphenylphosphine (27 mg, 0.103 mmol, 1.05 eq.) and sodium acetate (3.3 mg, 0.040 mmol, 0.4 eq.) were suspended in a mixture of MeOH (375 μL), H₂O (200 μL) and glacial acetic acid (3.25 μL) and heated under reflux for 30 min. The mixture was diluted with 5 mL CH₂Cl₂, washed with 10 mL H₂O, 5 mL brine, dried over MgSO₄ and concentrated under reduced pressure to afford clear oil. This was purified by column chromatography (EtOAc/hexane, 3:7) to furnish colourless oil (45 mg, 92%).

Method b)

To a solution of protected homocystine **13** (150 mg, 0.300 mmol) in DMF/H₂O (10:1, 2.2 mL) was added TCEP (105 mg, 0.375 mmol) and the solution was stirred at rt for 24 h. The mixture was

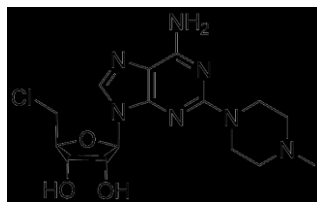
diluted with 10 mL CH₂Cl₂, washed with 20 mL H₂O, 10 mL brine, dried over MgSO₄ and concentrated under reduced pressure to afford protected homocysteine **14** as a colourless oil (140 mg, 95%): *R*_f = 0.90 (pentane/EtOAc, 7:3); *v*_{max} (neat)/cm⁻¹ = br 3363s (N-H), br 2979s (C-H), 1743 (C=O), 1692 (C=O), 1513s, 1440s, 1365s, 1345s, 1305s, 1161s, 1025s; ¹H NMR (400 MHz, CDCl₃): δ = 1.46 (s, 9 H, C-(CH₃)₃), 1.58 (t, *J* = 7.9 Hz, 1 H, SH), 1.92 – 1.99 (m, 1 H, β-CH₂), 2.09 – 2.15 (m, 1 H, β-CH₂), 2.57 – 2.63 (m, 2 H, γ-CH₂), 3.77 (s, 3 H, O-CH₃), 4.38 – 4.50 (m, 1 H, α-CH), 5.10 – 5.13 (br m, 1 H, NH); ¹³C NMR (400 MHz, CDCl₃): δ = 20.7 (CH₂, γ-C), 28.3 (CH₃, ^tBu), 37.2 (CH₂, β-C), 52.2 (CH, α-C), 52.5 (CH, α-C), 80.1, (C, C-(CH₃)₃), 155.4 (C, C=O), 172.8 (C, C=O); MS (TOF ES+) *m/z* 150.1 ([Boc fragmentation +H]⁺, 100%).

(2*R*,3*R*,4*S*,5*S*)-2-(6-amino-2-(piperidin-1-yl)-9*H*-purin-9-yl)-5-(chloromethyl)tetrahydrofuran-3,4-diol (**19**)



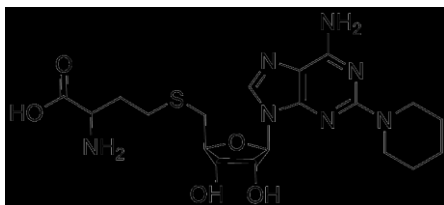
Piperidine (136 μL, 1.374 mmol, 2 eq.) was added to chlorinated adenosine **6** (250 mg, 0.687 mmol, 1 eq.) in H₂O (10 mL) and stirred at 100 °C overnight. The solvent was evaporated and then triturated with H₂O to afford the 2-piperidine adenosine compound **19** as a pale yellow amorphous solid (231 mg, 91%): *R*_f = 0.75 (CH₂Cl₂/ MeOH 6:1); ¹H NMR (400 MHz, DMSO-*d*₆): δ = 1.53 – 1.60 (m, 6 H, piperidine 2 × CH₂), 3.67 – 3.69 (m, 4 H, piperidine CH₂), 3.79 (m, 5'-CH₂), 3.94 (m, 1 H, 5'-CH₂), 4.04 – 4.07 (m, 1 H, 4'-CH), 4.22 (m, 1 H, 3'-CH), 4.77 (m, 1 H, 2'-CH), 5.78 (d, *J* = 6.2 Hz, 1 H, 1'-CH), 6.80 (br s, 2 H, NH₂), 7.93 (s, 1 H, 8-CH); ¹³C NMR (400 MHz, DMSO-*d*₆): δ = 25.0, 25.8, 45.3, 71.8, 72.6, 83.8, 87.7, 137.5, 152.1, 156.2, 159.1, 207.0; MS (TOF ES+) *m/z* 369.1 ([M+H]⁺, 100%).

(2*R*,3*R*,4*S*,5*S*)-2-(6-amino-2-(4-methylpiperazin-1-yl)-9*H*-purin-9-yl)-5-(chloromethyl)tetrahydrofuran-3,4-diol (**20**)



1-methylpiperazine (152 μ L, 1.374 mmol, 2 eq.) was added to chlorinated adenosine **6** (250 mg, 0.687 mmol, 1 eq.) in H₂O (10 mL) and stirred at 100 °C overnight. The solvent was evaporated and then triturated with H₂O to afford the 2-methylpiperazine adenosine **20** as an orange amorphous solid (156 mg, 59%); ¹H NMR (400 MHz, DMSO-*d*₆): δ = 2.34 (s, 3 H, piperazine CH₃), 2.46 – 2.49 (m, 4 H, piperazine, 2 $\times$ CH₂), 3.22 – 3.27 (m, 4 H, piperazine, 2 $\times$ CH₂), 3.77 (m, 1 H, 5'-CH₂), 3.93 (m, 1 H, 5'-CH₂), 4.01 – 4.03 (m, 1 H, 4'-CH), 4.24 (m, 1 H, 3'-CH), 4.78 (m, 1 H, 2'-CH), 5.81 (d, *J* = 6.0 Hz, 1 H, 1'-CH), 6.82 (br s, 2 H, NH₂), 7.96 (s, 1 H, 8-CH); ¹³C NMR (400 MHz, MeOH-*d*₄): 43.8 (methylpiperazine CH₂), 43.9 (methylpiperazine CH₂), 44.4 (methylpiperazine CH₃), 44.9 (methylpiperazine CH₂), 45.2 (methylpiperazine CH₂), 54.5 (5'-CH₂), 71.4 (4'-CH), 72.9 (3'-CH), 83.4 (2'-CH), 88.7 (1'-CH), 113.1 (5-C), 137.4 (8-CH), 151.6 (4-C), 155.8 (2-C), 159.2 (6-C); MS (TOF ES+) *m/z* 384.1 ([*M*+H]⁺, 100%).

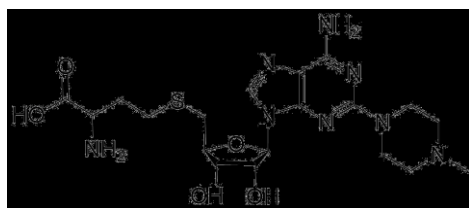
S-(((2*S*,3*S*,4*R*,5*R*)-5-(6-amino-2-(piperidin-1-yl)-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)homocysteine (**21**)



L-homocysteine thiolactone (209 mg, 1.359 mmol, 2 eq.) and NaOH (163 mg, 4.076 mmol, 6 eq.) in H₂O (12 mL) were flushed with N₂ for 10 min at rt. Then, 2-piperidine adenosine **19** (250 mg, 0.679 mmol, 1 eq.) and KI (34 mg, 0.204 mmol, 0.3 eq.) were added and the solution was heated at 100 °C overnight. The solvent was evaporated, dissolved in 0.5 mL H₂O and neutralised to pH 7–8. After trituration, the homocysteine product **21** was afforded as a pale yellow amorphous solid (187 mg, 59%); ¹H NMR (400 MHz, DMSO-*d*₆): δ = 1.47 – 1.53 (m, 6 H, piperidine 3 $\times$ CH₂), 1.85 – 1.89 (m, 2 H, β -CH₂), 2.45 – 2.52 (m, 2 H, γ -CH₂), 2.67 – 2.71 (m, 1 H, α -CH), 2.83 – 2.96 (m, 2 H, 5'-CH₂), 3.54 – 3.59 (m, 4 H, piperidine CH₂), 3.92 – 3.96 (m, 1 H, 4'-CH), 4.13 – 4.16 (m, 1 H, 3'-CH), 4.72 (t, *J* = 4.9

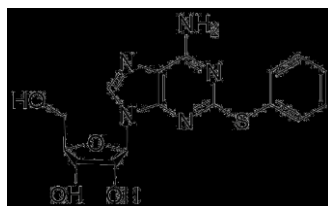
H₂, 1 H, 2'-CH), 5.83 (d, *J* = 5.1 Hz, 1 H, 1'-CH), 7.85 (s, 1 H, 8-CH); ¹³C NMR (400 MHz, DMSO-d₆): 24.3 (β-CH₂), 25.4 (piperidine CH₂), 28.7 (piperidine CH₂), [34.0, 34.5, 34.9 (resonance overlap, piperidine 3 × CH₂)], 46.0 (γ-CH₂), 55.3 (5'-CH₂), 73.4 (3'-CH), 81.5 (2'-CH), 83.9 (4'-CH), 88.9 (1'-CH), 137.6 (5-C), 138.9 (8-CH), 151.9 (4-C), 152.9 (2-C), 159.5 (6-C), 191.33 (C=O); MS (TOF ES+) *m/z* 467.0 ([M+H]⁺, 100%).

S-(((2*S*,3*S*,4*R*,5*R*)-5-(6-amino-2-(4-methylpiperazin-1-yl)-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)homocysteine (**22**)



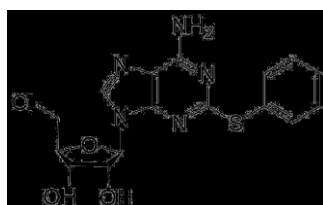
L-homocysteine thiolactone (201 mg, 1.302 mmol, 2 eq.) and NaOH (156 mg, 3.906 mmol, 6 eq.) in H₂O (10 mL) were flushed with N₂ for 10 min at rt. Then, adenosine **20** (250 mg, 0.651 mmol, 1 eq.) and KI (32 mg, 0.195 mmol, 0.3 eq.) were added and the solution was heated at 100 °C overnight. The solvent was evaporated, dissolved in 0.5 mL H₂O and neutralised to pH 7–8. After trituration, the homocysteine product **22** was afforded as a pale yellow amorphous solid (168 mg, 53%); ¹H NMR (400 MHz, D₂O): δ = 1.93 – 2.01 (m, 2 H, β-CH₂), 2.43 (s, 3 H, piperazine CH₃), 2.54 – 2.59 (m, 2 H, γ-CH₂), 2.71 – 2.76 (m, 4 H, piperazine 2 × CH₂), 2.85 – 2.98 (m, 2 H, 5'-CH₂), 3.09 – 3.15 (m, 4 H, piperazine 2 × CH₂), 3.62 – 3.65 (m, 1 H, α-CH), 4.19 (dt, *J* = 7.7, 4.9 Hz, 1 H, 4'-CH), 4.44 (m, 1 H, 3'-CH), 4.99 (m, 1 H, 2'-CH), 5.87 (d, *J* = 5.1 Hz, 1 H, 1'-CH), 7.92 (s, 1 H, 8-CH); ¹³C NMR (400 MHz, DMSO-d₆): 25.2 (β-CH₂), 25.6 (methylpiperazine CH₃), 32.5 (5-CH₂), 38.7 (methylpiperazine CH₂), 45.4 (γ-CH₂), 49.1 (5'-CH₂), 51.3 (methylpiperazine CH₂), 73.2 (3'-CH), 79.7 (2'-CH), 83.8 (4'-CH), 88.2 (1'-CH), 136.2 (5-C), 137.6 (8-CH), 152.1 (4-C), 154.2 (2-C), 157.0 (6-C), 171.2 (C=O); MS (TOF ES+) *m/z* 482.6 ([M+H]⁺, 100%).

(2*R*,3*R*,4*S*,5*R*)-2-(6-amino-2-(thiophenyl)-9*H*-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol
(**23**)



Sodium thiophenolate (306 mg, 2.319 mmol, 2 eq.) and adenosine **4** (400 mg, 1.159 mmol, 1 eq.) were stirred in H₂O (5 mL) at 100 °C overnight. After leaving overnight, another equivalent of sodium thiophenolate was added and the mixture left for a further 24 hours. The mixture was then cooled down, filtered and washed with cold H₂O to yield adenosine **23** as a pale brown solid (341 mg, 78%); *R*_f = 0.23 (CH₂Cl₂/MeOH 19:1); ¹H NMR (400 MHz, MeOH-*d*₄): δ = 3.47 – 3.56 (m, 2 H, 5'-CH₂), 4.02 – 4.05 (stack, 2 H, 3'-CH and 4'-CH), 4.76 (t, *J* = 5.4 Hz, 2'-CH), 5.80 (d, *J* = 6.0 Hz, 1'-CH), 7.43 – 7.49 (m, 3 H, phenyl 3 × CH), 7.60 – 7.63 (m, 2 H, phenyl 2 × CH), 8.14 (s, 1 H, 8-CH); ¹³C NMR (400 MHz, MeOH-*d*₄): δ = 62.1 (5'-CH₂), 70.8 (4'-CH), 72.9 (3'-CH), 86.1 (2'-CH), 89.6 (1'-CH), 128.7 (phenyl CH), 128.8 (phenyl 2 × CH), 130.6 (5-C), 134.9 (phenyl 2 × CH), 140.2 (8-CH), 149.8 (4-C), 155.7 (2-C), 165.3 (6-C); MS (TOF ES+) *m/z* 376.1 ([*M*]⁺, 100%).

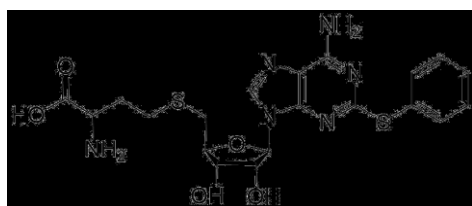
(2*R*,3*R*,4*S*,5*S*)-2-(6-amino-2-(thiophenyl)-9*H*-purin-9-yl)-5-(chloromethyl)tetrahydrofuran-3,4-diol (**24**)



A suspension of adenosine **23** (340 mg, 0.9070 mmol, 1 eq.) in pyridine (146 μL, 1.8130 mmol, 2 eq.) and acetonitrile (8 mL) was cooled to 0 °C. Thionyl chloride (332 μL, 4.5330 mmol, 5 eq.) was added dropwise with stirring and stirring was continued for a further 4 h at 5 °C. Then the solution was warmed to rt and stirred for a further 16 h. The reaction mixture was evaporated *in vacuo* and the resulting suspension was re-dissolved in MeOH, H₂O (12 mL, 5:1) and NH₄OH (2 mL) and stirred for 0.5 h at rt. The reaction mixture was then concentrated under reduced pressure to afford adenosine **24** as a light brown solid (190 mg, 53%); *R*_f = 0.84 (CH₂Cl₂/MeOH, 7:1); ¹H NMR (400 MHz, MeOH-*d*₄): δ = 3.08 (dd, *J* = 11.6, 6.5 Hz, 2 H, 5'-CH₂), 3.89 – 3.91 (m, 1 H, 4'-CH), 3.99 – 4.03 (m, 1 H, 3'-CH), 4.94 – 4.91 (m, 1 H, 2'-CH), 5.75 (d, *J* = 6.3 Hz, 1 H, 1'-CH), 7.31 – 7.39 (m, 3 H, phenyl 3 × CH), 7.62 – 7.64

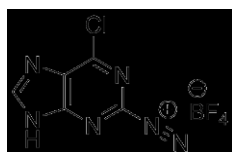
(m, 2 H, phenyl 2 × CH), 8.05 (s, 1 H, 8-CH); ^{13}C NMR (400 MHz, MeOH- d_4): δ = 47.8 (5'-CH₂), 67.9 (4'-CH), 71.2 (3'-CH), 85.6 (2'-CH), 89.3 (1'-CH), 127.9 (phenyl CH), 128.1 (phenyl 2 × CH), 130.3 (8-CH), 134.5 (phenyl 2 × CH), 138.6 (5-C), 149.7 (4-C), 158.5 (2-C), 165.1 (6-C); MS (TOF ES+) m/z 394.1 ([M+H]⁺, 100%); HRMS (TOF ES+) calcd for C₁₆H₁₇N₅O₃SCl [M+H]⁺ 394.0741, found 394.0744.

S-(((2*S*,3*S*,4*R*,5*R*)-5-(6-amino-2-(thiophenyl)-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)homocysteine (**25**)



L-homocysteine thiolactone (39 mg, 0.2545 mmol, 2 eq.) and NaOH (31 mg, 0.7634 mmol, 6 eq.) in H₂O (1 mL) were flushed with N₂ for 10 min at rt. Then, adenosine **24** (50 mg, 0.1272 mmol, 1 eq.) and KI (6 mg, 0.0382 mmol, 0.3 eq.) were added and the solution was heated at 100 °C overnight. The solvent was evaporated, dissolved in 0.5 mL H₂O and neutralised to pH 7–8. After trituration the homocysteine product **25** was afforded as a pale brown solid (24 mg, 43%); ^1H NMR (400 MHz, D₂O): δ = 1.90 – 1.99 (m, 2 H, β -CH₂), 2.21 (dd, J = 14.4, 4.1 Hz, 2 H, 5'-CH₂), 2.48 (t, J = 7.6 Hz, 2 H, γ -CH₂), 3.63 (dd, J = 5.4, 3.4 Hz, 1 H, α -CH), 3.92 – 3.98 (m, 2 H, 3'-CH and 4'-CH), 4.79 – 4.83 (m, 1 H, 2'-CH), 5.73 (d, J = 6.0 Hz, 1 H, 1'-CH), 7.49 – 7.53 (m, 3 H, phenyl 3 × CH), 7.61 – 7.64 (m, 2 H, phenyl 2 × CH), 8.19 (s, 1 H, 8-CH); MS (TOF ES+) m/z 492.1 ([M+H]⁺, 100%).

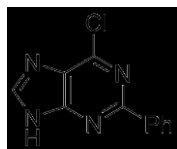
6-chloro-9H-purine-2-diazonium tetrafluoroborate (**27**)



A solution of sodium nitrite (179 mg, 2.599 mmol, 1.1 eq.) in H₂O (9 mL) was added slowly to a stirred solution of the purine **1** (400mg, 2.352 mmol, 1 eq.) in 6 *N* hydrochloric acid (4 mL) at 0 °C and the reaction mixture was stirred at 0 °C for 15 min. Then aqueous fluoroboric acid (47%, 5 mL) was added and the precipitate was collected by filtration and washed thoroughly with cold (0 °C) H₂O, MeOH and diethyl ether to provide purine **27** as white crystals (340 mg, 54%): ^1H NMR (400

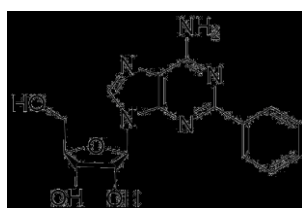
MHz, DMSO- d_6): δ = 7.97 (s, 1 H, 8-CH), 10.86 (s, 1 H, NH); ^{13}C NMR (400 MHz, DMSO- d_6): δ = 128.3 (5-C), 144.1 (8-CH), 150.9 (4-C), 152.2 (2-C), 153.5 (6-C).

*6-chloro-2-phenyl-9H-purine (28)*⁴



The purine diazonium tetrafluoroborate **27** (100 mg, 0.373 mmol, 1 eq.), phenylboronic acid (55 mg, 0.448 mmol, 1.2 eq.) and 5 mol % palladium acetate catalyst (5 mg, 0.019 mmol) were placed in a flask under N_2 atmosphere with protection from light. Degassed 1,4-dioxane (4 mL) was added at rt and stirred for 8 h. The reaction mixture was then diluted with diethyl ether and washed three times with brine. The organic phase was dried (MgSO_4), filtered and solvent removed *in vacuo*. The crude product was purified by column chromatography ($\text{CH}_2\text{Cl}_2/\text{MeOH}$ 17:3) to provide purine **28** as a white solid (30 mg, 34%): R_f = 0.56 ($\text{CH}_2\text{Cl}_2/\text{MeOH}$ 17:3); m.p. 207 – 212 °C; ν_{max} (neat)/ cm^{-1} = 3083s (=C-H), 3051s (=C-H), 3057s (=C-H), 1603s (C-C), 1440s, 1332s, 1304s, 1179s, 1087s, 1025s, 685 (C-Cl); ^1H NMR (400 MHz, DMSO- d_6): δ = 7.31 – 7.40 (m, 3 H, 3 $\times$ phenyl CH), 7.78 – 7.80 (m, 2 H, 3 $\times$ phenyl CH), 8.03 (s, 1 H, 8-CH); ^{13}C NMR (400 MHz, DMSO- d_6): δ = 124.8 (5-C), [128.5, 128.8 (CH, resonance overlap, phenyl CH)], [132.5, 132.8, 133.1 (CH, resonance overlap, phenyl CH)], 138.5 (phenyl C), 144.4 (8-CH), 151.9 (4-C), 153.2 (2-C), 159.4 (6-C); MS (CI+) m/z 234 ($\text{M}+\text{H}$)⁺, 156 [$\text{M}+\text{H}-\text{Ph}$]⁺, 76 [Ph]⁺.

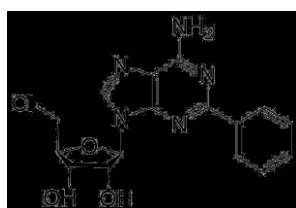
(2R,3R,4S,5R)-2-(6-amino-2-phenyl-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol (26)



Adenosine **4** (200 mg, 0.5797 mmol, 1 eq.), $\text{Pd}(\text{OAc})_2$ (6.5 mg, 0.0290 mmol, 5 mol%), SPhos (28.6 mg, 0.0696 mmol, 12 mol %), phenylboronic acid (120 mg, 0.9855 mmol, 1.7 eq.) and K_2CO_3 (120 mg, 0.8696 mmol, 1.5 eq.) were added to a flask and was purged with N_2 for 30 min. Degassed 1,4-dioxane (5 mL) was added by syringe to the reaction vessel and the reaction was stirred at 80 °C for 12 h. After 12 h, another equivalent of phenylboronic acid was added and the mixture stirred for

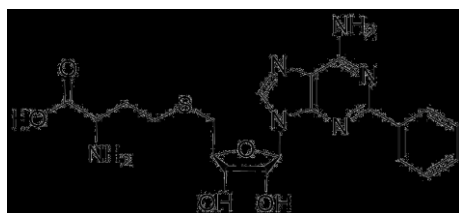
another 24 h. When the reaction was complete, the mixture was evaporated *in vacuo* and the crude purified by column chromatography with sequential elutions of CH₂Cl₂, CH₂Cl₂/MeOH [7 M NH₃] (93:7) and CH₂Cl₂/MeOH [7 M NH₃] (9:1) provided adenosine **26** as a pale yellow amorphous solid (146 mg, 73%): R_f = 0.42 (CH₂Cl₂/MeOH 17:3); ¹H NMR (400 MHz, MeOH-d₄): δ = 3.79 – 3.93 (m, 2 H, 5'-CH₂), 4.17 – 4.20 (m, 1 H, 4'-CH), 4.45 – 4.47 (m, 1 H, 3'-CH), 4.89 – 4.91 (m, 1 H, 2'-CH), 6.11 (t, J = 5.5 Hz, 1 H, 1'-CH), 7.43 – 7.49 (m, 3 H, phenyl 3 × CH), 8.28 – 8.35 (m, 2 H, phenyl 2 × CH), 8.34 (s, 1 H, 8-CH); ¹³C NMR (400 MHz, MeOH-d₄): δ = 61.8 (5'-CH₂), 70.9 (3'-CH), 74.0 (2'-CH), 85.8 (4'-CH), 89.2 (1'-CH), 118.1 (5-C), 126.8 (phenyl CH), 127.8 (phenyl CH), 128.0 (phenyl CH), 129.6 (phenyl CH), 133.5 (phenyl C), 140.7 (8-CH), 150.3 (4-C), 155.9 (2-C), 160.0 (6-C); MS (TOF ES+) m/z 344.1 ([M+H]⁺, 100%); HRMS (TOF ES+) calcd for C₁₆H₁₈N₅O₄ [M+H]⁺ 344.1359, found 344.1354.

(2*R*,3*R*,4*S*,5*S*)-2-(6-amino-2-phenyl-9*H*-purin-9-yl)-5-(chloromethyl)tetrahydrofuran-3,4-diol (**29**)



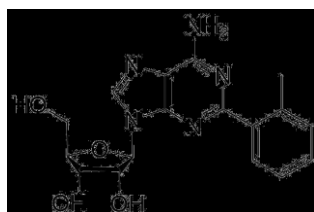
A suspension of adenosine **26** (145 mg, 0.4215 mmol, 1 eq.) in pyridine (68 μL, 0.8619 mmol, 2 eq.) and acetonitrile (4 mL) was cooled to 0 °C. Thionyl chloride (154 μL, 1.2851 mmol, 5 eq.) was added dropwise with stirring and stirring was continued for a further 4 h at 5 °C. Then the solution was warmed to rt and stirred for a further 16 h. The reaction mixture was evaporated *in vacuo* and the resulting suspension was re-dissolved in MeOH, H₂O (6 mL, 5:1) and NH₄OH (1 mL) and stirred for 0.5 h at rt. The reaction mixture was then concentrated under reduced pressure to afford chlorinated nucleoside **29** as a light yellow/orange solid (103 mg, 68%): R_f = 0.66 (CH₂Cl₂/MeOH 17:3); ¹H NMR (400 MHz, DMSO-d₆): δ = 3.89 – 4.02 (m, 2 H, 5'-CH₂), 4.21 – 4.24 (m, 1 H, 4'-CH), 4.50 – 4.53 (m, 1 H, 3'-CH), 4.91 – 4.93 (m, 1 H, 2'-CH), 5.51 – 5.54 (d, J = 5.4 Hz, 1 H, 3-OH), 5.64 – 5.67 (d, J = 5.8 Hz, 1 H, 2-OH), 6.17 (d, J = 5.5 Hz, 1 H, 1'-CH), 7.45 – 7.51 (m, 3 H, phenyl 3 × CH), 8.32 – 8.35 (m, 2 H, phenyl 2 × CH), 8.35 (s, 1 H, 8-CH); ¹³C NMR (500 MHz, DMSO-d₆): δ = 44.8 (5'-CH₂), 71.3 (3'-CH), 72.8 (2'-CH), 83.7 (4'-CH), 87.5 (1'-CH), 118.3 (5-C), 127.6 (phenyl CH), 128.2 (phenyl CH), 129.7 (phenyl CH), 138.2 (phenyl C), 140.5 (8-CH), 150.4 (4-C), 155.8 (2-C), 157.8 (6-C); MS (TOF ES+) m/z 361.0 ([M+H]⁺, 100%).

S-(((2*S*,3*S*,4*R*,5*R*)-5-(6-amino-2-phenyl-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)homocysteine (**30**)



L-homocysteine thiolactone (17 mg, 0.1106 mmol, 2 eq.) and NaOH (13 mg, 0.3317 mmol, 6 eq.) in 2:1 H₂O:DMF (1.6:0.8 mL) were flushed with N₂ for 10 min at rt. Then, adenosine **29** (20 mg, 0.0528 mmol, 1 eq.) and KI (1 mg, 0.0166 mmol, 0.3 eq.) were added and the solution was heated at 100 °C overnight. The solvent was evaporated, dissolved in 0.5 mL H₂O and neutralised to pH 7–8. After trituration the homocysteine product **30** was afforded as a pale brown solid (13.6 mg, 56%); ¹H NMR (400 MHz, D₂O): δ = 1.87 – 1.90 (m, 2 H, β-CH₂), 2.31 – 2.40 (m, 2 H, 5'-CH₂), 2.52 – 2.54 (m, 2 H, γ-CH₂), 3.57 – 3.62 (m, 1 H, α-CH), 3.99 – 4.01 (m, 1 H, 4'-CH), 4.13 – 4.17 (m, 1 H, 3'-CH), 4.79 – 4.83 (m, 1 H, 2'-CH), 5.73 (d, *J* = 5.6 Hz, 1 H, 1'-CH), 7.33 – 7.38 (m, 3 H, phenyl 3 × CH), 7.58 – 7.61 (m, 2 H, 2 × phenyl CH), 8.11 (s, 1 H, 8-CH); MS (TOF ES+) *m/z* 461.36 ([M+H]⁺, 100%).

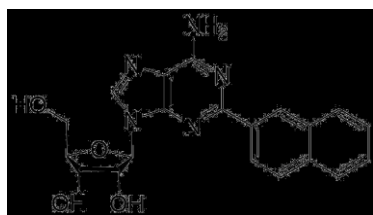
(2*R*,3*R*,4*S*,5*R*)-2-(6-amino-2-(*o*-tolyl)-9*H*-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol (**31**)



Adenosine **4** (100 mg, 0.2899 mmol, 1 eq.), Pd(OAc)₂ (3 mg, 0.0145 mmol, 5 mol %), SPhos (14 mg, 0.0348 mmol, 12 mol %), *o*-tolylboronic acid (67 mg, 0.4928 mmol, 1.7 eq.) and K₂CO₃ (60 mg, 0.4348 mmol, 1.5 eq.) were added to a flask and was purged with N₂ for 30 min. Degassed 1,4-dioxane (2.5 mL) was added by syringe to the reaction vessel and the reaction was stirred at 80 °C for 12 h. After 12 h, another equivalent of *o*-tolylboronic acid was added and the mixture stirred until completion. When the reaction was complete, the mixture was evaporated *in vacuo* and the crude mixture was purified by column chromatography with sequential elutions of CH₂Cl₂, CH₂Cl₂/MeOH [7 M NH₃] (93:7) and CH₂Cl₂/MeOH [7 M NH₃] (9:1) which furnished adenosine **31** as a pale yellow amorphous solid (55 mg, 53%); R_f = 0.38 (CH₂Cl₂/MeOH 17:3); ¹H NMR (400 MHz, MeOH-

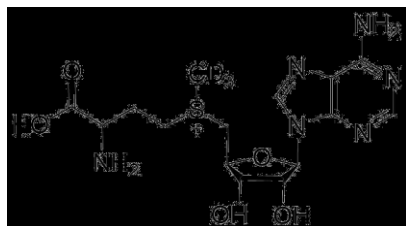
d₄): δ = 2.33 (s, 3 H, σ -CH₃), 3.82 – 3.92 (m, 2 H, 5'-CH₂), 4.15 – 4.17 (m, 1 H, 4'-CH), 4.31 – 4.36 (m, 1 H, 3'-CH), 4.78– 4.80 (m, 1 H, 2'-CH), 6.01 – 6.03 (d, J = 5.2 Hz, 1 H, 1'-CH), 7.11 – 7.19 (m, 1 H, aromatic CH), 7.23 – 7.35 (m, 2 H, aromatic 2 $\times$ CH), 7.51 – 7.53 (m, 1 H, aromatic CH), 8.38 (s, 1 H, 8-CH); ¹³C NMR (400 MHz, MeOH-d₄): 19.1, 60.5, 70.7, 72.9, 84.2, 88.3, 106.7, 117.2, 121.2, 124.2, 126.1, 127.6, 129.6, 130.1, 135.2, 144.3, 156.6; LC-MS (ESI) m/z 357.25 ([M+H]⁺, 100%).

(2*R*,3*R*,4*S*,5*R*)-2-(6-amino-2-(naphthalen-2-yl)-9*H*-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol (**32**)



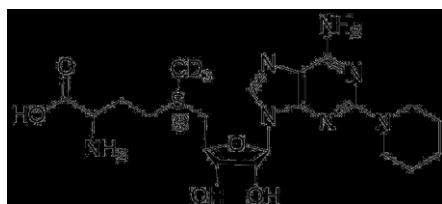
Adenosine **4** (100 mg, 0.2899 mmol, 1 eq.), Pd(OAc)₂ (3 mg, 0.0145 mmol, 5 mol %), SPhos (14 mg, 0.0348 mmol, 12 mol %), 2-naphthylboronic acid (85 mg, 0.4928 mmol, 1.7 eq.) and K₂CO₃ (60 mg, 0.4348 mmol, 1.5 eq.) were added to a flask and was purged with N₂ for 30 min. Degassed 1,4-dioxane (2.5 mL) was added by syringe to the reaction vessel and the reaction was stirred at 80 °C for 12 h. After 12 h, another equivalent of 2-naphthylboronic acid was added and the mixture stirred until completion. When the reaction was complete, the mixture was evaporated *in vacuo* and the crude mixture was purified by column chromatography with sequential elutions of CH₂Cl₂, CH₂Cl₂/MeOH [7 M NH₃] (93:7) and CH₂Cl₂/MeOH [7 M NH₃] (9:1) which furnished adenosine **32** as a pale yellow amorphous solid (46 mg, 40%): R_f = 0.33 (CH₂Cl₂/MeOH 17:3); ¹H NMR (400 MHz, MeOH-d₄): δ = 3.72 – 3.90 (m, 2 H, 5'-CH₂), 4.12 – 4.15 (m, 1 H, 4'-CH), 4.29 – 4.33 (m, 1 H, 3'-CH), 4.66 (t, J = 5.6 Hz, 1 H, 2'-H), 5.91 (d, J = 5.9 Hz, 1 H, 1'-CH), 7.02 – 7.04 (m, 1 H, naphthalene CH), 7.05 – 7.07 (m, 1 H, naphthalene CH), 7.07 – 7.10 (m, 1 H, naphthalene CH), 7.21 (t, J = 7.5 Hz, 1 H, naphthalene CH), 7.30 – 7.36 (m, 1 H, naphthalene CH), 7.60 (d, J = 8.2 Hz, 1 H, naphthalene CH), 7.66 – 7.71 (m, 1 H, naphthalene CH); ¹³C NMR (400 MHz, MeOH-d₄): 61.8, 71.0, 74.1, 86.5, 89.6, 108.5, 117.8, 122.5, 125.7, 125.8, 126.5, 127.2, 127.9, 128.5, 129.0, 135.2, 140.6, 144.3, 154.5, 156.6; LC-MS (ESI) m/z 393.14 ([M+H]⁺, 100%).

S-adenosyl-L-methionine- d_3 (**33**)



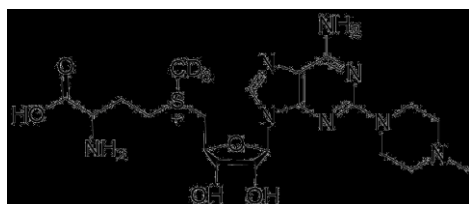
AgOTf (200 mg, 0.780 mmol, 6 eq.) and CD_3I (538 μ L, 7.804 mmol, 60 eq.) were added to $HCOOH$ (5 mL) and stirred for 1 h which resulted in the formation of a yellow precipitate. Then *S*-adenosyl-L-homocysteine (50 mg, 0.130 mmol, 1 eq.) was added and the solution was stirred for 24 h. After this time the solution was diluted with H_2O (5 mL) and the insoluble AgI was filtered over a syringe filter and the solvent was evaporated. The residue was purified using an Amberlite XAD-4 column prepared with H_2O (+0.1% FA). The product was eluted using the following solvents sequentially: i) H_2O (+0.1% FA), ii) H_2O (+0.1% FA)/MeCN (+0.1% FA) 9:1, iii) H_2O (+0.1% FA)/MeCN (+0.1% FA) 8:2, iv) water (+0.1% FA)/MeCN (+0.1% FA) 7:3. Each fraction was concentrated individually and the nature and purity was determined using NMR. The desired CD_3 -labelled SAM product **33** was found to be a colourless residue (24 mg, 46%) which was $\geq 95\%$ pure by LC-MS analysis: 1H NMR (400 MHz, D_2O): δ = 2.39 – 2.45 (m, 2 H, β - CH_2), 3.46 – 3.54 (m, 1 H, γ_a -CH), 3.68 – 3.75 (m, 1 H, γ_b -CH), 3.62 (t, J = 7.7 Hz, 1 H, α -CH), 3.93 – 4.02 (m, 2 H, $2 \times 5'$ - CH_2), 4.57 – 4.60 (m, 1 H, $4'$ -CH), 4.64 (dd, J = 10.3, 4.5 Hz, 1 H, $3'$ -CH), 4.87 (dd, J = 10.3, 4.2 Hz, 1 H, $2'$ -CH), 6.19 (d, J = 4.2 Hz, 1 H, $1'$ -CH), 8.24 (s, 1 H, 2-CH), 8.29 (s, 1 H, 8-CH); ^{13}C NMR (400 MHz, D_2O): δ = 24.6 (β - CH_2), 38.3 ($5'$ - CH_2), 43.8 (γ - CH_2), 50.8 (α -CH), 72.6 ($3'$ -CH), 73.1 ($2'$ -CH), 90.1 ($1'$ -CH), 121.2 (5-C), 143.5 (8-CH), 144.4 (4-C), 148.0 (2-CH), 150.0 (6-C), 169.9 (C=O); LC-MS (ESI) m/z 402.2 ($[M]^+$, 100%); HRMS (TOF ES+) calcd for $C_{31}H_{25}N_5O_7Cl$ $[M]^+$ 402.1663, found 402.1739.

S-(((2*S*,3*S*,4*R*,5*R*)-5-(6-amino-2-(piperidin-1-yl)-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)methionine-*d*₃ (**34**)



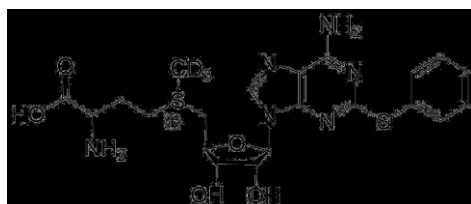
A solution of homocysteine **21** (20 mg, 0.0428 mmol, 1 eq.) was prepared in HCOOH (2 mL) and stirred for 1 h. In tandem, AgOTf (66 mg, 0.257 mmol, 6 eq.) and CD₃I (167 μ L, 2.568 mmol, 60 eq.) were added to CCl₄ (1 mL) resulting in the formation of a yellow precipitate. After 1 h, the second mixture was filtered over a syringe filter into the first solution and subsequently stirred at high speed for a further 24 h. After this time, the solution was diluted with H₂O (3 mL) and the solvent evaporated *in vacuo*. The crude residue was purified by HPLC and the desired CD₃-labelled product **34** was found to be a yellow resin which was $\geq 95\%$ pure by LC-MS analysis (8.5 mg, 42%): ¹H NMR (500 MHz, D₂O): δ = 1.60 – 1.66 (m, 6 H, piperidine 3 $\times$ CH₂), 2.23 – 2.28 (m, 2 H, β -CH₂), 3.37 – 3.42 (m, 2 H, 5'-CH₂), 3.46 – 3.50 (m, 1 H, γ_a -CH), 3.59 – 3.62 (m, 1 H, α -CH), 3.63 – 3.65 (m, 4 H, piperidine 2 $\times$ CH₂), 3.68 – 3.73 (m, 1 H, γ_b -CH), 4.39 – 4.42 (m, 1 H, 4'-CH), 4.49 – 4.53 (m, 1 H, 3'-CH), 4.78 – 4.82 (m, 1 H, 2'-CH), 5.92 (d, J = 3.7 Hz, 1 H, 1'-CH), 8.01 (s, 1 H, 8-CH); ¹³C NMR (500 MHz, D₂O): δ = 23.4 (β -CH₂), 24.9 (piperidine CH₂), 38.6 (5'-CH₂), 40.7 (piperidine CH₂), 43.1 (γ -CH₂), 46.9 (piperidine CH₂), 49.9 (α -CH), 52.5 (piperidine C), 72.2 (3'-CH), 72.7 (2'-CH), 77.9 (4'-CH), 89.2 (1'-CH), 115.1 (5-C), 141.1 (8-CH), 150.3 (4-C), 150.6 (2-C), 151.9 (6-C), 181.9 (C=O); MS (TOF ES+) m/z 485.44 ([M]⁺, 100%).

((2*S*,3*S*,4*R*,5*R*)-5-(6-amino-2-(4-methylpiperazin-1-yl)-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)methionine-*d*₃ (**35**)



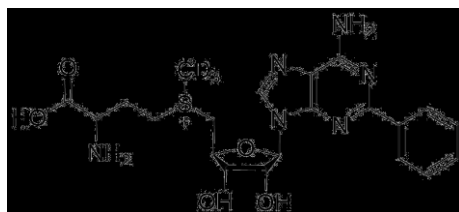
A solution of homocysteine **22** (20 mg, 0.0413 mmol, 1 eq.) was prepared in HCOOH (2 mL) and stirred for 1 h. In tandem, AgOTf (64 mg, 0.2480 mmol, 6 eq.) and CD₃I (161 μL, 2.4792 mmol, 60 eq.) were added to CCl₄ (1 mL) resulting in the formation of a yellow precipitate. After 1 h, the second mixture was filtered over a syringe filter into the first solution and subsequently stirred at high speed for a further 24 h. After this time, the solution was diluted with H₂O (3 mL) and the solvent evaporated *in vacuo*. The crude residue was purified by HPLC and the desired CD₃-labelled product **35** was found to be a clear resin which was ≥95% pure by LC-MS analysis (8.2 mg, 40%): ¹H NMR (400 MHz, D₂O): δ = 2.12 – 2.19 (m, 2 H, β-CH₂), 2.48 (s, 3 H, methylpiperazine CH₃), 2.76 – 2.81 (m, 4 H, methylpiperazine 2 × CH₂), 2.94 – 2.98 (m, 2 H, γ-CH₂), 3.09 – 3.16 (m, 2 H, 5'-CH₂), 3.14 – 3.18 (m, 4 H, methylpiperazine 2 × CH₂), 3.64 – 3.66 (m, 1 H, α-CH), 4.39 (m, 1 H, 4'-CH), 4.50 (m, 1 H, 3'-CH), 4.88 (m, 1 H, 2'-CH), 5.91 (d, *J* = 5.5 Hz, 1 H, 1'-CH), 8.09 (s, 1 H, 8-CH); ¹³C NMR (400 MHz, D₂O): 24.8 (β-CH₂), 25.0 (methylpiperazine CH₃), 32.3 (5'-CH₂), [38.6, 38.8 (resonance overlap, methylpiperazine 2 × CH₂)], [42.4, 42.9 (resonance overlap, γ'-CH₂)], 44.1 (α-CH), [52.0, 52.1 (resonance overlap, methylpiperazine 2 × CH₂)], 72.5 (3'-CH), 72.7 (2'-CH), 77.8 (4'-CH), 89.1 (1'-CH), 114.8 (5-C), 139.4 (8-CH), 150.7 (4-C), 152.5 (2-C), 154.6 (6-C), 171.5 (C=O); MS (TOF ES+) *m/z* 500.17 ([M]⁺, 100%).

S-(((2*S*,3*S*,4*R*,5*R*)-5-(6-amino-2-(thiophenyl)-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)methionine-*d*₃ (**36**)



A solution of homocysteine **25** (20 mg, 0.0406 mmol, 1 eq.) was prepared in HCOOH (2 mL) and stirred for 1 h. In tandem, AgOTf (63 mg, 0.2434 mmol, 6 eq.) and CD₃I (151 μL, 1.0417 mmol, 60 eq.) were added to CCl₄ (1 mL) resulting in the formation of a yellow precipitate. After 1 h, the second mixture was filtered over a syringe filter into the first solution and subsequently stirred at high speed for a further 24 h. After this time, the solution was diluted with H₂O (3 mL) and the solvent evaporated *in vacuo*. The residue was purified by HPLC and the desired CD₃-labelled product **36** was found to be a yellow resin which was ≥95% pure by LC-MS analysis (5.1 mg, 29%): ¹H NMR (500 MHz, D₂O): δ = 2.05 – 2.11 (m, 2 H, β-CH₂), 2.61 (dd, *J* = 14.4, 4.1 Hz, 2 H, 5'-CH₂), 2.69 (t, *J* = 7.6 Hz, 2 H, γ-CH₂), 3.78 (dd, *J* = 5.4, 3.4 Hz, 1 H, α-CH), 3.81 – 3.86 (m, 1 H, 4'-CH), 4.33 (m, 1 H, 3'-CH), 4.90 (t, *J* = 5.8 Hz, 1 H, 2'-CH), 5.77 (d, *J* = 6.0 Hz, 1 H, 1'-CH), 7.59 – 7.63 (m, 3 H, phenyl 3 × CH), 7.67 – 7.70 (m, 2 H, phenyl 2 × CH), 8.31 (s, 1 H, 8-CH); ¹³C NMR (500 MHz, D₂O): δ = 24.8 (β-CH₂), 38.6 (5'-CH₂), 43.3 (γ-CH₂), 51.6 (α-CH), 71.2 (3'-CH), 72.6 (2'-CH), 79.3 (4'-CH), 90.4 (1'-CH), 115.1 (5-C), 128.3 (phenyl C), 130.0 (phenyl 2 × CH), 130.6 (phenyl CH), 135.5 (phenyl 2 × CH), 142.7 (8-CH), 148.7 (4-C), 153.8 (2-C), 163.1 (6-C), 171.0 (C=O); MS (TOF ES+) *m/z* 510.36 ([M]⁺, 100%).

(((2*S*,3*S*,4*R*,5*R*)-5-(6-amino-2-phenyl-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(3-amino-3-carboxypropyl)(methyl-*d*₃)sulfonium (**37**)

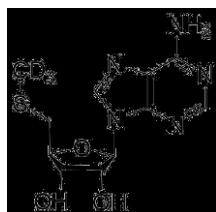


A solution of homocysteine **30** (10 mg, 0.0217 mmol, 1 eq.) was prepared in HCOOH (1 mL) and stirred for 1 h. In tandem, AgOTf (33.5 mg, 0.1304 mmol, 6 eq.) and CD₃I (85 μL, 0.5849 mmol, 60

eq.) were added to CCl₄ (0.5 mL) resulting in the formation of a yellow precipitate. After 1 h, the second mixture was filtered over a syringe filter into the first solution and subsequently stirred at high speed for a further 24 h. After this time, the solution was diluted with H₂O (1.5 mL) and the solvent evaporated *in vacuo*. The crude residue was purified by HPLC and the desired CD₃-labelled product **37** was found to be a yellow resin (2.6 mg, 25%) which was ≥95% pure by LC-MS analysis: ¹H NMR (500 MHz, D₂O): δ = 2.13 – 2.26 (m, 2 H, β-CH₂), 3.20 – 3.28 (m, 1 H, γ_a-CH), 3.37 – 3.46 (m, 1 H, α-CH), 3.60 – 3.65 (m, 1 H, γ_b-CH), 3.90 – 4.00 (m, 2 H, 5'-CH₂), 4.58 – 4.65 (m, 1 H, 4'-CH), 4.71 – 4.75 (m, 1 H, 3'-CH), 5.07 – 5.12 (m, 1 H, 2'-CH), 6.23 (d, *J* = 4.0 Hz, 1 H, 1'-CH), 7.85 – 7.61 (m, 3 H, phenyl 3 × CH), 8.14 – 8.06 (m, 2 H, phenyl 2 × CH), 8.47 (s, 1 H, 8-CH); LC-MS (ESI) *m/z* 478.43 ([M]⁺, 100%).

*¹³C NMR could not be obtained

5'-methylthioadenosine-CD₃ (**CD₃-MTA**)



CD₃-SAM (10 mg, 0.025mmol) was dissolved in 1 mL of pH 8.8 Tris-HCl buffer and heated at 37 °C for 7 days. The resultant mixture was then separated by RP-HPLC to furnish CD₃-MTA as a white residue (3.2 mg, 40%) which was ≥95% pure by LC-MS analysis. Any remaining CD₃-SAM in the mixture was collected from HPLC and left to degrade further to obtain a higher yield: ¹H NMR (500 MHz, D₂O): δ = 3.07 – 2.91 (m, 2H, 5'-CH), 4.38 (dt, *J* = 7.0, 4.9 Hz, 1 H, 4'-CH), 4.46 (t, *J* = 5.0 Hz, 1 H, 3'-CH), 4.91 (t, *J* = 5.2 Hz, 1 H, 2'-CH), 6.17 (d, *J* = 5.0 Hz, 1 H, 1'-CH), 8.46 (s, 1 H, 2-CH), 8.56 (s, 1 H, 8-CH); ¹³C NMR (500 MHz, D₂O): δ = 35.6 (5'-CH₂), 72.4 (3'-CH), 73.6 (2'-CH), 83.7 (4'-CH), 88.4 (1'-CH), 114.4 (5-C), 117.5 (8-CH), 121.1 (4-C), 142.9 (2-CH), 144.7 (6-C); LC-MS (ESI) *m/z* 301.35 ([M+H]⁺, 100%).

8. References

- 1 C. D. Allis and T. Jenuwein, *Nat. Rev. Genet.*, 2016, **17**, 487–500.
- 2 G. Felsenfeld, *Epigenetics*, 2007, 15–22.
- 3 C. D. Allis and T. W. Muir, *ChemBioChem*, 2011, **12**, 264–279.
- 4 T. Kouzarides, *Cell*, 2007, **128**, 693–705.
- 5 T. Bartke, J. Borgel and P. a. DiMaggio, *Brief. Funct. Genomics*, 2013, **12**, 205–218.
- 6 A. Yoshimi and M. Kurokawa, *J. Cell. Biochem.*, 2011, **112**, 415–424.
- 7 B. Weinhold, *Environ. Heal. Perspect.*, 2006, **114**, 160–167.
- 8 M. Luo, *ACS Chem. Biol.*, 2012, **7**, 443–463.
- 9 K. Murray, *Biochemistry*, 1964, **3**, 10–15.
- 10 E. L. Greer and Y. Shi, *Nat. Rev. Genet.*, 2012, **13**, 343–357.
- 11 R. Wang and M. Luo, *Curr. Opin. Chem. Biol.*, 2013, **17**, 729–737.
- 12 R. Wang, G. Ibáñez, K. Islam, W. Zheng, G. Blum, C. Sengelaub and M. Luo, *Mol. Biosyst.*, 2011, **7**, 2970.
- 13 M. Biel, V. Wascholowski and A. Giannis, *Angew. Chemie - Int. Ed.*, 2005, **44**, 3186–3216.
- 14 P. Hu and Y. Zhang, *J. Am. Chem. Soc.*, 2006, **128**, 1272–1278.
- 15 T. Kouzarides, *Curr. Opin. Genet. Dev.*, 2002, **12**, 198–209.
- 16 T. Kouzarides, *Curr. Opin. Genet. Dev.*, 2002, **12**, 198–209.
- 17 F. L. Cherblanc, R. W. M. Davidson, P. Di Fruscia, N. Srimongkolpithak and M. J. Fuchter, *Nat. Prod. Rep.*, 2013, **30**, 605–24.
- 18 S. C. Dillon, X. Zhang, R. C. Trievel and X. Cheng, *Genome Biol.*, 2005, **6**, 227.
- 19 Y. Tsukada, J. Fang, H. Erdjument-Bromage, M. E. Warren, C. H. Borchers, P. Tempst and Y. Zhang, *Nature*, 2006, **439**, 811–816.
- 20 Y. Shi, F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole, R. A. Casero and Y. Shi, *Cell*, 2004, **119**, 941–953.
- 21 M. Yang, C. B. Gocke, X. Luo, D. Borek, D. R. Tomchick, M. Machius, Z. Otwinowski and H. Yu, *Mol. Cell*, 2006, **23**, 377–387.
- 22 C. Marabelli, B. Marrocco and A. Mattevi, *Curr. Opin. Struct. Biol.*, 2016, **41**, 135–144.
- 23 D. Hayward and P. A. Cole, in *Anal Chem.*, 2016, vol. 25, pp. 261–278.
- 24 A. J. Bannister and T. Kouzarides, *Cell Res.*, 2011, **21**, 381–395.
- 25 B. Fierz and T. W. Muir, *Nat. Chem. Biol.*, 2012, **8**, 417–427.

- 26 Y. Zhang and D. Reinberg, *Genes Dev.*, 2001, **15**, 2343–2360.
- 27 H. Wang, R. Cao, L. Xia, H. Erdjument-Bromage, C. Borchers, P. Tempst and Y. Zhang, *Mol. Cell*, 2001, **8**, 1207–1217.
- 28 C. S. Veerappan, Z. Avramova and E. N. Moriyama, *BMC Evol. Biol.*, 2008, **8**, 1–20.
- 29 S. T. Keating and A. El-Osta, *Epigenetics*, 2013, **8**, 361–372.
- 30 R. J. Sims and D. Reinberg, *Genes Dev.*, 2006, **20**, 2779–2786.
- 31 B. Xiao, C. Jing, J. R. Wilson, P. A. Walker, N. Vasisht, G. Kelly, S. Howell, I. A. Taylor, G. M. Blackburn and S. J. Gamblin, *Nature*, 2003, **421**, 652–656.
- 32 A. Kouskouti, E. Scheer, A. Staub, L. Tora and I. Talianidis, *Mol. Cell*, 2004, **14**, 175–182.
- 33 J.-F. Couture, E. Collazo, G. Hauk and R. C. Trievel, *Nat. Struct. Mol. Biol.*, 2006, **13**, 140–146.
- 34 B. Xiao, J. R. Wilson and S. J. Gamblin, *Curr. Opin. Struct. Biol.*, 2003, **13**, 699–705.
- 35 J. R. Wilson, C. Jing, P. A. Walker, S. R. Martin, S. A. Howell, G. M. Blackburn, S. J. Gamblin and B. Xiao, *Cell*, 2002, **111**, 105–115.
- 36 X. Zhang, Z. Yang, S. I. Khan, J. R. Horton, H. Tamaru, E. U. Selker and X. Cheng, *Mol. Cell*, 2003, **12**, 177–185.
- 37 D. Brasacchio, J. Okabe, C. Tikellis, A. Balcerczyk, P. George, E. K. Baker, A. C. Calkin, M. Brownlee, M. E. Cooper and A. El-osta, *Diabetes*, 2009, **58**, 1229–1236.
- 38 K. Nishioka, S. Chuikov, K. Sarma, H. Erdjument-Bromage, C. D. Allis, P. Tempst and D. Reinberg, *Genes Dev.*, 2002, **16**, 479–489.
- 39 S. Mori, K. Iwase, N. Iwanami, Y. Tanaka, H. Kagechika and T. Hirano, *Bioorg. Med. Chem.*, 2010, **18**, 8158–8166.
- 40 T. Kwon, *EMBO J.*, 2003, **22**, 292–303.
- 41 K. Nishioka, S. Chuikov, K. Sarma, H. Erdjument-Bromage, C. D. Allis, P. Tempst and D. Reinberg, *Genes Dev.*, 2002, **16**, 479–489.
- 42 A. Dhayalan, S. Kudithipudi, P. Rathert and A. Jeltsch, *Chem. Biol.*, 2011, **18**, 111–120.
- 43 P. Hu, S. Wang and Y. Zhang, *J. Am. Chem. Soc.*, 2008, **130**, 3806–3813.
- 44 L. Gaughan, J. Stockley, N. Wang, S. R. C. McCracken, A. Treumann, K. Armstrong, F. Shaheen, K. Watt, I. J. McEwan, C. Wang, R. G. Pestell and C. N. Robson, *Nucleic Acids Res.*, 2011, **39**, 1266–1279.
- 45 C. Ea and D. Baltimore, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 18972–18977.
- 46 B. Lehnertz, J. C. Rogalski, F. M. Schulze, L. Yi, S. Lin, J. Kast and F. M. V Rossi, *Mol. Cell*, 2011, **43**, 673–680.
- 47 G. Sun, M. A. Reddy, H. Yuan, L. Lanting, M. Kato and R. Natarajan, *J. Am. Soc. Nephrol.*, 2010, **21**, 2069–2080.
- 48 A. El-Osta, D. Brasacchio, D. Yao, A. Pocai, P. L. Jones, R. G. Roeder, M. E. Cooper and M.

- Brownlee, *J. Exp. Med.*, 2008, **205**, 2409–2417.
- 49 J. Okabe, C. Orłowski, A. Balcerczyk, C. Tikellis, M. C. Thomas, M. E. Cooper and A. El-Osta, *Circ. Res.*, 2012, **110**, 1067–1076.
 - 50 T. G. Deering, T. Ogiwara, A. P. Trace, B. Maier and R. G. Mirmira, *Diabetes*, 2009, **58**, 185–193.
 - 51 Y. Tao, R. L. Nepl, Z. P. Huang, J. Chen, R. H. Tang, R. Cao, Y. Zhang, S. W. Jin and D. Z. Wang, *J. Cell Biol.*, 2011, **194**, 551–565.
 - 52 H. M. Herz, M. Mohan, A. S. Garruss, K. Liang, Y. hei Takahashi, K. Mickey, O. Voets, C. P. Verrijzer and A. Shilatifard, *Genes Dev.*, 2012, **26**, 2604–2620.
 - 53 C. Martin and Y. Zhang, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 838–849.
 - 54 P.-M. Dehé and V. Géli, *Biochem. Cell Biol.*, 2006, **84**, 536–548.
 - 55 G. S. I. S. Noma Ken-ichi, Allis C. David, *Science (80-.)*, 2001, **293**, 1150–1156.
 - 56 B. D. Strahl, R. Ohba, R. G. Cook and C. D. Allis, *Proc. Natl. Acad. Sci.*, 1999, **96**, 14967–14972.
 - 57 J. H. Martens, M. Verlaan, E. Kalkhoven and A. Zantema. *Mol Cell Biol.* 2003, **23**, 1808–1816.
 - 58 S. K. Chakrabarti, J. Francis, S. M. Ziesmann, J. C. Garmey and R. G. Mirmira, *J. Biol. Chem.*, 2003, **278**, 23617–23623.
 - 59 A. El-Osta, D. Brasacchio, D. Yao, A. Pocai, P. L. Jones, R. G. Roeder, M. E. Cooper and M. Brownlee, *J. Exp. Med.*, 2008, **205**, 2409–2417.
 - 60 C. D. Allis, S. L. Berger, J. Cote, S. Dent, T. Jenuwien, T. Kouzarides, L. Pillus, D. Reinberg, Y. Shi, R. Shiekhata, A. Shilatifard, J. Workman and Y. Zhang, *Cell*, 2007, **131**, 633–636.
 - 61 J. Cheng, R. Blum, C. Bowman, D. Hu, A. Shilatifard, S. Shen and B. D. Dynlacht, *Mol. Cell*, 2014, **53**, 979–992.
 - 62 I. Kassner, M. Barandun, M. Fey, F. Rosenthal and M. O. Hottiger, *Epigenetics and Chromatin*, 2013, **6**, 1–9.
 - 63 S. Ko, J. Ahn, C. S. Song, S. Kim, K. Knapczyk-Stwora and B. Chatterjee, *Mol. Endocrinol.*, 2011, **25**, 433–44.
 - 64 Y. Li, M. A. Reddy, F. Miao, N. Shanmugam, J. K. Yee, D. Hawkins, B. Ren and R. Natarajan, *J. Biol. Chem.*, 2008, **283**, 26771–26781.
 - 65 J. Francis, S. K. Chakrabarti, J. C. Garmey and R. G. Mirmira, *J. Biol. Chem.*, 2005, **280**, 36244–36253.
 - 66 K. Subramanian, D. Jia, P. Kapoor-Vazirani, D. R. Powell, R. E. Collins, D. Sharma, J. Peng, X. Cheng and P. M. Vertino, *Mol. Cell*, 2008, **30**, 336–347.
 - 67 P.-O. Esteve, H. G. Chin, J. Benner, G. R. Feehery, M. Samaranayake, G. A. Horwitz, S. E. Jacobsen, S. Pradhan, P.-O. Estève, H. G. Chin, J. Benner, G. R. Feehery, M. Samaranayake, G. A. Horwitz, S. E. Jacobsen and S. Pradhan, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 5076–5081.
 - 68 X. Liu, Z. Chen, C. Xu, X. Leng, H. Cao, G. Ouyang and W. Xiao, *Nucleic Acids Res.*, 2015, **43**,

- 5081–5098.
- 69 S. Pagans, S. E. Kauder, K. Kaehlcke, N. Sakane, S. Schroeder, W. Dormeyer, R. C. Trievel, E. Verdin, M. Schnolzer and M. Ott, *Cell Host Microbe*, 2010, **7**, 234–244.
 - 70 I. Kassner, A. Andersson, M. Fey, M. Tomas, E. Ferrando-May and M. O. Hottiger, *Open Biol.*, 2013, **3**, 120173.
 - 71 Q. Xie, Y. Hao, L. Tao, S. Peng, C. Rao, H. Chen, H. You, M. Dong and Z. Yuan, *EMBO Rep.*, 2012, **13**, 371–377.
 - 72 R. Daniel, A. E. Webb, J. L. White, R. Timothy, T. Goswami, X. Shi, A. Espejo, D. R. Calnan, A. E. Webb, J. L. White, T. R. Stowe, X. Shi, A. Espejo, M. T. Bedford, O. Gozani and S. P. Gygi, 2016.
 - 73 X.-D. Yang, B. Huang, M. Li, A. Lamb, N. L. Kelleher and L.-F. Chen, *EMBO J.*, 2009, **28**, 1055–66.
 - 74 B. C. Smith and J. M. Denu, *Biochim. Biophys. Acta - Gene Regul. Mech.*, 2009, **1789**, 45–57.
 - 75 H.-B. Guo and H. Guo, *Proc. Natl. Acad. Sci.*, 2007, **104**, 8797–8802.
 - 76 K. E. Moore and O. Gozani, *Biochim. Biophys. Acta - Gene Regul. Mech.*, 2014, **1839**, 1395–1403.
 - 77 J. S. Lee, Y. Kim, I. S. Kim, B. Kim, H. J. Choi, J. M. Lee, H. J. R. Shin, J. H. Kim, J. Y. Kim, S. B. Seo, H. Lee, O. Binda, O. Gozani, G. L. Semenza, M. Kim, K. Il Kim, D. Hwang and S. H. Baek, *Mol. Cell*, 2010, **39**, 71–85.
 - 78 L. Fritsch, P. Robin, J. R. R. Mathieu, M. Souidi, H. Hinaux, C. Rougeulle, A. Harel-Bellan, M. Ameyar-Zazoua and S. Ait-Si-Ali, *Mol. Cell*, 2010, **37**, 46–56.
 - 79 T. L. Graves, Y. Zhang and J. E. Scott, *Anal. Biochem.*, 2008, **373**, 296–306.
 - 80 T. M. Lakowski and A. Frankel, *Anal. Biochem.*, 2010, **396**, 158–160.
 - 81 E. Collazo, J. F. Couture, S. Bulfer and R. C. Trievel, *Anal. Biochem.*, 2005, **342**, 86–92.
 - 82 C. Wang, S. Leffler, D. H. Thompson and C. a. Hrycyna, *Biochem. Biophys. Res. Commun.*, 2005, **331**, 351–356.
 - 83 T. a. Klink, M. Staeben, K. Twesten, a. L. Kopp, M. Kumar, R. S. Dunn, C. a. Pinchard, K. M. Kleman-Leyer, M. Klumpp and R. G. Lowery, *J. Biomol. Screen.*, 2012, **17**, 59–70.
 - 84 G. Ibáñez, J. L. McBean, Y. M. Astudillo and M. Luo, *Anal. Biochem.*, 2010, **401**, 203–210.
 - 85 C. L. Hendricks, J. R. Ross, E. Pichersky, J. P. Noel and Z. S. Zhou, *Anal. Biochem.*, 2004, **326**, 100–105.
 - 86 K. M. Dorgan, W. L. Wooderchak, D. P. Wynn, E. L. Karschner, J. F. Alfaro, Y. Cui, Z. S. Zhou and J. M. Hevel, *Anal. Biochem.*, 2006, **350**, 249–255.
 - 87 J. Wu, N. Xie, Y. Feng and Y. G. Zheng, *J. Biomol. Screen.*, 2012, **17**, 237–244.
 - 88 P. Rathert, X. Cheng and A. Jeltsch, *Biotechniques*, 2007, **43**, 602–608.
 - 89 H. Gowher, X. Zhang, X. Cheng and A. Jeltsch, *Anal. Biochem.*, 2005, **342**, 287–291.

- 90 A. Dhayalan, E. Dimitrova, P. Rathert and A. Jeltsch, *J. Biomol. Screen.*, 2009, **14**, 1129–1133.
- 91 A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev and K. Zhao, *Cell*, 2007, **129**, 823–837.
- 92 N. Gauthier, M. Caron, L. Pedro, M. Arcand, J. Blouin, a. Labonte, C. Normand, V. Paquet, a. Rodenbrock, M. Roy, N. Rouleau, L. Beaudet, J. Padros and R. Rodriguez-Suarez, *J. Biomol. Screen.*, 2012, **17**, 49–58.
- 93 K. K. Li, C. Luo, D. Wang, H. Jiang and Y. G. Zheng, *Med. Res. Rev.*, 2010, n/a-n/a.
- 94 D. A. Orlando, M. W. Chen, V. E. Brown, S. Solanki, Y. J. Choi, E. R. Olson, C. C. Fritz, J. E. Bradner and M. G. Guenther, *Cell Rep.*, 2014, **9**, 1163–1170.
- 95 B. Egan, C. C. Yuan, M. L. Craske, P. Labhart, G. D. Guler, D. Arnott, T. M. Maile, J. Busby, C. Henry, T. K. Kelly, C. A. Tindell, S. Jhunhunwala, F. Zhao, C. Hatton, B. M. Bryant, M. Classon and P. Trojer, *PLoS One*, 2016, **11**, 1–20.
- 96 N. L. Young, P. a DiMaggio, M. D. Plazas-Mayorca, R. C. Baliban, C. a Floudas and B. a Garcia, *Mol. Cell. Proteomics*, 2009, **8**, 2266–2284.
- 97 S. M. Carr, S. Munro, B. Kessler, U. Oppermann and N. B. La Thangue, *EMBO J.*, 2010, **30**, 317–327.
- 98 B. M. Zee, R. S. Levin, B. Xu, G. LeRoy, N. S. Wingreen and B. a. Garcia, *J. Biol. Chem.*, 2010, **285**, 3341–3350.
- 99 B. M. Zee, R. S. Levin, P. a Dimaggio and B. a Garcia, *Epigenetics Chromatin*, 2010, **3**, 22.
- 100 J. J. Allen, M. Li, C. S. Brinkworth, J. L. Paulson, D. Wang, A. Hübner, W.-H. Chou, R. J. Davis, A. L. Burlingame, R. O. Messing, C. D. Katayama, S. M. Hedrick and K. M. Shokat, *Nat. Methods*, 2007, **4**, 511–516.
- 101 R. Wang, K. Islam, Y. Liu, W. Zheng, H. Tang, N. Lailler, G. Blum, H. Deng and M. Luo, *J. Am. Chem. Soc.*, 2013, **135**, 1048–1056.
- 102 R. Urich, G. Wishart, M. Kiczun, A. Richters, N. Tidten-Luksch, D. Rauh, B. Sherborne, P. G. Wyatt and R. Brenk, *ACS Chem. Biol.*, 2013, **8**, 1044–1052.
- 103 K. Shah, Y. Liu, C. Deirmengian and K. M. Shokat, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 3565–3570.
- 104 L. a. Witucki, X. Huang, K. Shah, Y. Liu, S. Kyin, M. J. Eck and K. M. Shokat, *Chem. Biol.*, 2002, **9**, 25–33.
- 105 Q. Lin, F. Jiang, P. G. Schultz and N. S. Gray, *J. Am. Chem. Soc.*, 2001, **123**, 11608–11613.
- 106 J. Li, H. Wei and M. M. Zhou, *J. Med. Chem.*, 2011, **54**, 7734–7738.
- 107 K. Islam, W. Zheng, H. Yu, H. Deng and M. Luo, *ACS Chem. Biol.*, 2011, **6**, 679–684.
- 108 H. Wu, J. Min, V. V. Lunin, T. Antoshenko, L. Dombrovski, H. Zeng, A. Allali-Hassani, V. Campagna-Slater, M. Vedadi, C. H. Arrowsmith, A. N. Plotnikov and M. Schapira, *PLoS One*, 2010, **5**.
- 109 J. M. Mato, M. L. Martínez-Chantar and S. C. Lu, *Ann. Hepatol.*, 2013, **12**, 183–189.

- 110 J. D. Finkelstein, *J. Nutr. Biochem.*, 1990, **1**, 228–237.
- 111 J. Mato, L. Alvarez, P. Ortiz and M. a Pajares, *Pharmacol. Ther.*, 1997, **73**, 265–280.
- 112 I. A. Hao, X; Zhou, M; Li, H; Angres, *J. Hepatol. Gastrointest. Disord.*, 2017, **03**, 1114–1125.
- 113 I. A. Hao, X; Zhou, M; Li, H; Angres, *J. Hepatol. Gastrointest. Disord.*, 2017, **03**, 1114–1125.
- 114 G. de La Haba, G. A. Jamieson, S. H. Mudd and H. H. Richards, *J. Am. Chem. Soc.*, 1959, **81**, 3975–3980.
- 115 S. H. MUDD and G. L. CANTONI, *J. Biol. Chem.*, 1958, **231**, 481–92.
- 116 G. L. CANTONI and J. DURELL, *J. Biol. Chem.*, 1957, **225**, 1033–48.
- 117 J. Park, J. Tai, C. A. Roessner and A. I. Scott, *Bioorganic Med. Chem.*, 1996, **4**, 2179–2185.
- 118 J. Baddiley and G. A. Jamieson, *J. Chem. Soc.*, 1954, 4280–4284.
- 119 V. M. Clark, A. R. Todd and J. Zussman, *J. Chem. Soc.*, 1951, 2952.
- 120 W. Yu, D. Smil, F. Li, W. Tempel, O. Fedorov, K. T. Nguyen, Y. Bolshan, R. Al-Awar, S. Knapp, C. H. Arrowsmith, M. Vedadi, P. J. Brown and M. Schapira, *Bioorganic Med. Chem.*, 2013, **21**, 1787–1794.
- 121 F. Seela and X. Peng, *J. Org. Chem.*, 2006, **71**, 81–90.
- 122 P. Francom, Z. Janeba, S. Shibuya and M. J. Robins, *J. Org. Chem.*, 2002, **67**, 6788–6796.
- 123 F. Wnuk, J. Robins, F. Wnuk, J. Robins and F. Wnuk, *Can. J. Chem.*, 1991, **69**, 1468–1474.
- 124 M. J. Thompson, A. Mekhalfia, D. P. Hornby and G. M. Blackburn, *J. Org. Chem.*, 1999, **64**, 7467–7473.
- 125 C. Lherbet and J. W. Keillor, *Org. Biomol. Chem.*, 2004, **2**, 238–245.
- 126 H. Adachi, K. K. Palaniappan, A. a. Ivanov, N. Bergman, Z.-G. Gao and K. a. Jacobson, *J. Med. Chem.*, 2007, **50**, 1810–1827.
- 127 K. L. Billingsley, K. W. Anderson and S. L. Buchwald, *Angew. Chemie - Int. Ed.*, 2006, **45**, 3484–3488.
- 128 K. H. Shaughnessy, *European J. Org. Chem.*, 2006, 1827–1835.
- 129 S. Sengupta and S. Bhattacharyya, *J. Org. Chem.*, 1997, **62**, 3405–3406.
- 130 A. Roglans, A. Pla-Quintana and M. Moreno-Mañas, *Chem. Rev.*, 2006, **106**, 4622–4643.
- 131 G. Pergolizzi, J. N. Butt, R. P. Bowater and G. K. Wagner, *Chem. Commun.*, 2011, **47**, 12655.
- 132 J. T. Kuethe, *J. Label. Compd. Radiopharm.*, 2012, **55**, 180–185.
- 133 B. a Garcia, S. Mollah, B. M. Ueberheide, S. a Busby, T. L. Muratore, J. Shabanowitz and D. F. Hunt, *Nat. Protoc.*, 2007, **2**, 933–938.
- 134 Y. Li, P. Trojer, C.-F. Xu, P. Cheung, A. Kuo, W. J. Drury, Q. Qiao, T. A. Neubert, R.-M. Xu, O. Gozani and D. Reinberg, *J. Biol. Chem.*, 2009, **284**, 34283–95.

- 135 J. C. Rice, S. D. Briggs, B. Ueberheide, C. M. Barber, J. Shabanowitz, D. F. Hunt, Y. Shinkai and C. D. Allis, *Mol. Cell*, 2003, **12**, 1591–1598.
- 136 M. Tachibana, K. Sugimoto, M. Nozaki, J. Ueda, T. Ohta, M. Ohki, M. Fukuda, N. Takeda, H. Niida, H. Kato and Y. Shinkai, *Genes Dev.*, 2002, **16**, 1779–1791.
- 137 X. Cheng, *Cold Spring Harb. Perspect. Biol.*, 2014, **6**, 18747–18748.
- 138 F. Bontemps and G. Van Den Berghe, *Biochem. J.*, 1997, **327** (Pt 2), 383–9.
- 139 *Nat. Methods*, 2005, **2**, 319–320.
- 140 B. Dalby, S. Cates, A. Harris, E. C. Ohki, M. L. Tilkins, P. J. Price and V. C. Ciccarone, *Methods*, 2004, **33**, 95–103.
- 141 H. J. M. Kostrzewski, T.; Borg, A. J.; Meng, Y.; Filipovic, I.; Male, V.; Wack, A.; DiMaggio, P. A.; Brady, *J Immunol*, 2018, **200**, 1370–1381.
- 142 R. Huang, X. Li, Y. Yu, L. Ma, S. Liu and X. Zong, *Oncotarget*, 2017, **Advance Pu**, 1–11.
- 143 T. Lankau, T. N. Kuo and C.-H. Yu, *J. Phys. Chem. A*, 2017, **121**, 505–514.
- 144 J. L. Hoffman, *Biochemistry*, 1986, **25**, 4444–4449.
- 145 R. T. Borchardt and Y. S. Wu, *J. Med. Chem.*, 1976, **19**, 1099–1103.
- 146 R. Wang and M. G. Brattain, *FEBS Lett.*, 2007, **581**, 3164–3170.
- 147 W. Fischle, *Methods*, 2005, **36**, 362–367.
- 148 K. K. Biggar and S. S.-C. Li, *Nat. Rev. Mol. Cell Biol.*, 2014, **16**, 5–17.
- 149 H. Kontaki and I. Talianidis, *Mol. Cell*, 2010, **39**, 152–160.
- 150 X. Liu, D. Wang, Y. Zhao, B. Tu, Z. Zheng, L. Wang, H. Wang, W. Gu, R. G. Roeder and W.-G. Zhu, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 1925–1930.
- 151 G. S. Ivanov, T. Ivanova, J. Kurash, A. Ivanov, S. Chuikov, F. Gizatullin, E. M. Herrera-Medina, F. Rauscher, D. Reinberg and N. A. Barlev, *Mol. Cell. Biol.*, 2007, **27**, 6756–6769.
- 152 J. Huang, R. Sengupta, A. B. Espejo, M. G. Lee, J. A. Dorsey, M. Richter, S. Opravil, R. Shiekhhattar, M. T. Bedford, T. Jenuwein and S. L. Berger, *Nature*, 2007, **449**, 105–108.
- 153 S. Campaner, F. Spreafico, T. Burgold, M. Doni, U. Rosato, B. Amati and G. Testa, *Mol. Cell*, 2011, **43**, 681–688.
- 154 S. Chuikov, J. K. Kurash, J. R. Wilson, B. Xiao, N. Justin, G. S. Ivanov, K. Mckinney, P. Tempst, C. Prives, S. J. Gamblin, N. A. Barlev and D. Reinberg, *Nature*, 2004, **432**, 353–360.
- 155 C. Hsu, K. Peng, H. Jhang, C. Lin, S. Wu, C. Chiang, S. Lee, W. C. Y. Yu and L. Juan, *Oncogene*, 2012, **31**, 2335–2349.
- 156 S. Munro, N. Khaire, A. Inche, S. Carr and N. B. La Thangue, *Oncogene*, 2010, **29**, 2357–2367.
- 157 T. Masatsugu and K. Yamamoto, *Biochem. Biophys. Res. Commun.*, 2009, **381**, 22–26.
- 158 H. S. Cho, T. Suzuki, N. Dohmae, S. Hayami, M. Unoki, M. Yoshimatsu, G. Toyokawa, M.

- Takawa, T. Chen, J. K. Kurash, H. I. Field, B. A. J. Ponder, Y. Nakamura and R. Hamamoto, *Cancer Res.*, 2011, **71**, 655–660.
- 159 M. J. Oudhoff, S. A. Freeman, A. L. Couzens, F. Antignano, E. Kuznetsova, P. H. Min, J. P. Northrop, B. Lehnertz, D. Barsyte-lovejoy, M. Vedadi, C. H. Arrowsmith, H. Nishina, M. R. Gold, F. M. V Rossi, A. Gingras and C. Zaph, *Dev. Cell*, 2013, **26**, 188–194.
 - 160 N. Balasubramanian, M. Ananthanarayanan and F. J. Suchy, *Am. J. Physiol. Liver Physiol.*, 2012, **302**, G937–47.
 - 161 J. Yang, J. Huang, M. Dasgupta, N. Sears, M. Miyagi, B. Wang, M. R. Chance, X. Chen, Y. Du, Y. Wang, L. An, Q. Wang, T. Lu, X. Zhang, Z. Wang and G. R. Stark, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 21499–504.
 - 162 D. Wang, J. Zhou, X. Liu, D. Lu, C. Shen, Y. Du, F.-Z. Wei, B. Song, X. Lu, Y. Yu, L. Wang, Y. Zhao, H. Wang, Y. Yang, Y. Akiyama, H. Zhang and W.-G. Zhu, *Proc. Natl. Acad. Sci.*, 2013, **110**, 5516–5521.
 - 163 J. Wang, S. Hevi, J. K. Kurash, H. Lei, F. Gay, J. Bajko, H. Su, W. Sun, H. Chang, G. Xu, F. Gaudet, E. Li and T. Chen, *Nat Genet*, 2009, **41**, 125–129.
 - 164 J. Castaño, C. Morera, B. Sesé, S. Boue, C. B. Costa, M. Martí, A. Roque, A. Jordan and M. J. Barrero, *PLoS One*, 2016, **11**, 1–21.
 - 165 J. D. Chapman, J. P. Gagné, G. G. Poirier and D. R. Goodlett, *J. Proteome Res.*, 2013, **12**, 1868–1880.
 - 166 I. Kassner, A. Andersson, M. Fey, M. Tomas, E. Ferrando-May and M. O. Hottiger, *Open Biol.*, 2013, **3**, 120173–120173.
 - 167 I. Pirkov, J. Norbeck, L. Gustafsson and E. Albers, *FEBS J.*, 2008, **275**, 4111–4120.
 - 168 A. P. Stevens, K. Dettmer, G. Kirovski, K. Samejima, C. Hellerbrand, A. K. Bosserhoff and P. J. Oefner, *J. Chromatogr. A*, 2010, **1217**, 3282–3288.
 - 169 G. Bistulfi, H. C. Affronti, B. A. Foster, E. Karasik, B. Gillard, C. Morrison, J. Mohler, J. G. Phillips and D. J. Smiraglia, *Oncotarget*, 2016, **7**, 14380–14393.
 - 170 C. Bauchart-Thevret, B. Stoll and D. G. Burrin, *Nutr. Res. Rev.*, 2009, **22**, 175–187.
 - 171 J. A. North, J. Sriram, K. Chourey, C. D. Ecker, R. Sharma, J. A. Wildenthal, R. L. Hettich and F. R. Tabita, *MBio*, 2016, **7**, e00855-16.
 - 172 K. Hoffmann, A. Grünberger, F. Lausberg, M. Bott and L. Eggeling, *Appl. Environ. Microbiol.*, 2013, **79**, 6730–6736.
 - 173 C. Mary, P. Duek, L. Salleron, P. Tienzi, D. Bumann, A. Bairoch and L. Lane, *PLoS One*, 2012, **7**.
 - 174 G. Kirovski, A. P. Stevens, B. Czech, K. Dettmer, T. S. Weiss, P. Wild, A. Hartmann, A. K. Bosserhoff, P. J. Oefner and C. Hellerbrand, *Am. J. Pathol.*, 2011, **178**, 1145–1152.
 - 175 M. Lubin and A. Lubin, *PLoS One*, 2009, **4**.
 - 176 M. A. Avila, E. R. García-Trevijano, S. C. Lu, F. J. Corrales and J. M. Mato, *Int. J. Biochem. Cell Biol.*, 2004, **36**, 2125–2130.

- 177 K. J. Mavrikis, 3. rd. McDonald ER, M. R. Schlabach, E. Billy, G. R. Hoffman, A. deWeck, D. A. Ruddy, K. Venkatesan, J. Yu, G. McAllister, M. Stump, R. deBeaumont, S. Ho, Y. Yue, Y. Liu, Y. Yan-Neale, G. Yang, F. Lin, H. Yin and Gao, *Science (80-.)*, 2016, **5944**, 1–11.
- 178 P. Andreu-Pérez, J. Hernandez-Losa, T. Moliné, R. Gil, J. Grueso, A. Pujol, J. Cortés, M. A. Avila and J. A. Recio, *BMC Cancer*, 2010, **10**, 265.
- 179 K. A. Mowen, J. Tang, W. Zhu, B. T. Schurter, K. Shuai, H. R. Herschman and M. David, *Cell*, 2001, **104**, 731–741.
- 180 S. F. V. De Oliveira, M. Ganzinelli, R. Chilà, L. Serino, M. E. Maciel, C. De Andrade Urban, R. S. De Lima, I. J. Cavalli, D. Generali, M. Broggin, G. Damia and E. M. De Souza Fonseca Ribeiro, *PLoS One*, 2016, **11**, 1–13.
- 181 K. Limm, K. Dettmer, J. Reinders, P. J. Oefner and A. K. Bosserhoff, *PLoS One*, 2016, **11**, 1–13.
- 182 L. Christa, J. Kersual, J. Augé and J. L. Pérignon, *Biochem. J.*, 1988, **255**, 145–52.
- 183 A. P. Stevens, B. Spangler, S. Wallner, M. Kreutz, K. Dettmer, P. J. Oefner and A. K. Bosserhoff, *J. Cell. Biochem.*, 2009, **106**, 210–219.
- 184 M. Iraburu, E. R. García-Trevijano, E. Ansorena, M. A. Avila, J. M. Mato, M. L. Martínez-Chantar, S. C. Lu, L. Chen and Z. Z. Huang, *Hepatology*, 2002, **35**, 274–280.
- 185 N. Shiraki, Y. Shiraki, T. Tsuyama, F. Obata, M. Miura, G. Nagae, H. Aburatani, K. Kume, F. Endo and S. Kume, *Cell Metab.*, 2014, **19**, 780–794.
- 186 B. M. Zee, R. S. Levin, P. A. Dimaggio and B. A. Garcia, *Epigenetics Chromatin*, 2010, **3**, 22.
- 187 R. M. Hoffman, H. Herrera, A. Groce and R. M. Hoffman, *Cancer Res.*, 1993, **53**, 2479–2483.
- 188 Y. Ma, K. Kanakousaki and L. Buttitta, *Front. Genet.*, 2015, **5**, 1–18.
- 189 A. Shevchenko, H. Tomas, J. Havlis, J. V Olsen and M. Mann, *Nat. Protoc.*, 2006, **1**, 2856–2860.
- 190 M. D. Plazas-Mayorca, B. M. Zee, N. L. Young, I. M. Fingerman, G. LeRoy, S. D. Briggs and B. A. Garcia, *J. Proteome Res.*, 2009, **8**, 5367–5374.
- 191 J. Lee, S. E. Kim, J. Y. Lee, S. Y. Kim, S. U. Kang, S. H. Seo, M. W. Chun, T. Kang, S. Y. Choi and H. O. Kim, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 1087–1092.
- 192 E. Halbfinger, D. T. Major, M. Ritzmann, J. Ubl, G. Reiser, J. L. Boyer, K. T. Harden and B. Fischer, *J. Med. Chem.*, 1999, **42**, 5325–5337.