Engineering Methyltransferase-Cofactor Binding Interactions: Towards Therapeutic and Diagnostic Applications

by

STEVEN JOHN COULTHARD

UNIVERSITY OF BIRMINGHAM

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Chemistry

College of Engineering and Physical Sciences

University of Birmingham

November 2023

Declaration of Authorship

I, Steven Coulthard declare that the work that has been presented in this thesis, titled "Engineering Methyltransferase-Cofactor Binding Interactions: Towards Therapeutic and Diagnostic Applications" is my own original work and is a result of my original research.

The compounds that are listed in Chapter 2 are novel with the exception of ((S)-3-amino-3-carboxypropyl)((5-(6-((2-aminoethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(8-(2-((E)-4-((2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)carbamoyl)benzylidene)hydrazineyl)-8-oxooct-2-yn-1-yl)sulfonium, ((S)-3-amino-3-carboxypropyl)(8-(2-((E)-4-((2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)carbamoyl)benzylidene)hydrazineyl)-8-oxooct-2-yn-1-yl)((3,4-dihydroxy-5-(6-((2-hydroxyethyl)amino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)sulfonium, 5'-iodo-6-chloro-5'-deoxyadenosine, 8-hydroxyoct-6-ynoic acid, tert-butyl 2-(8-hydroxyoct-6-ynoyl)hydrazine-1-carboxylate and tert-butyl 2-(8-bromooct-6-ynoyl)hydrazine-1-carboxylate.

The Mutant MTase Enzymes that were created in Chapter 3 are novel.

List of Publications

1) Irving, O. J., Matthews, L., Coulthard, S., Neely, R. K., Grant, M. M., & Albrecht, T. (2023). Sterically enhanced control of enzyme-assisted DNA assembly. https://doi.org/10.1101/2023.02.27.530245
2) Wilkinson, A. A., Jagu, E., Ubych, K., Coulthard, S., Rushton, A. E., Kennefick, J., Su, Q., Neely, R. K., & Fernandez-Trillo, P. (2020). Site-Selective and Rewritable Labeling of DNA through Enzymatic, Reversible, and Click Chemistries. ACS Central Science, 6(4), 525–534. https://doi.org/10.1021/acscentsci.9b01023

Abstract

DNA methylation is an important epigenetic modification that regulates gene expression and maintains cellular function. The main cofactor for this process, S-adenosylmethionine (SAM), provides the methyl group necessary for DNA methyltransferases (DNMTs), enzymes that catalyse the addition of methyl groups to cytosine residues. While SAM is essential for maintaining normal DNA methylation patterns, its limitations—such as poor bioavailability, lack of tissue specificity, and off-target effects—present challenges for therapeutic applications that target DNA methylation in diseases like cancer, neurological disorders, and developmental diseases. To address these challenges, the development of SAM analogues with improved stability, selectivity, and targeted delivery could enhance the precision and efficacy of epigenetic therapies. Furthermore, the exploration of mutant DNMTs, engineered to recognize and modify specific DNA sequences or epigenetic marks, offers a promising strategy to enhance the specificity of DNA methylation-based interventions.

This thesis presents the synthesis of two novel analogues of SAM using an efficient, protecting-group free synthesis to create N6-substitued analogues. The activity of these analogues was tested on DNA with two methyltransferases (M.TaqI and M.MpeI). The M.MpeI active site has been probed, and mutations to the active site were achieved and subsequently tested against multiple cofactors using a new *in vitro* transcription translation method. Mutants of interest were upscaled to discern the required concentrations of cofactor and MTase DNA activity and whether this was comparable to the wild type M.MpeI. This thesis reports identification of a new cofactor and a mutant MTase that are compatible with one another. However, this mutant MTase is incompatible with SAM which makes this mutant a more selective MTase that could be expanded upon for diagnostic and therapeutic purposes.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Robert Neely for his support and guidance throughout my PhD, without his aid this thesis wouldn't have been possible. Whether it was guiding me through an experimental challenge or encouraging me to develop new ideas for the project, he has helped with the countless issues that were encountered during this PhD and helped to develop as a scientist.

I would also like to thank Dr. Krystian Ubych for being there since the beginning of my PhD and being extremely supportive in teaching me the techniques that were used in this thesis. I would also like to thank all the past members of the Neely group for their support and their aid whether that be with solving a problem or providing encouragement.

I would also like to thank the analytical facility in the chemistry department for use of the NMR machines and providing the mass spectrometry service, this has allowed me to collect NMR data for the compounds I have synthesised and obtain their mass spectrum profiles. I would also like to express my deepest thanks to the whole chemistry department for the support and the numerous opportunities that were made available to me during this PhD.

I would also like to thank the UKRI for funding this project as without this funding this project wouldn't have been possible, and we wouldn't have the knowledge we now do.

Finally, I would like to thank my friends and my family for listening to my ramblings on this PhD whether it be about a positive result or a kind ear to vent about my issues whilst continuing to support me.

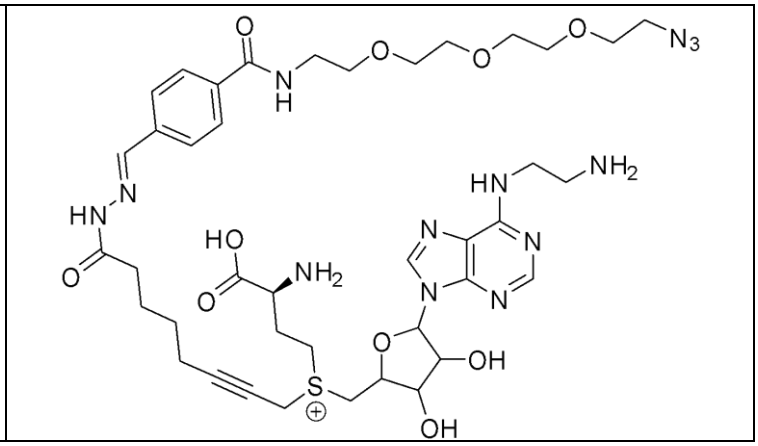
Contents

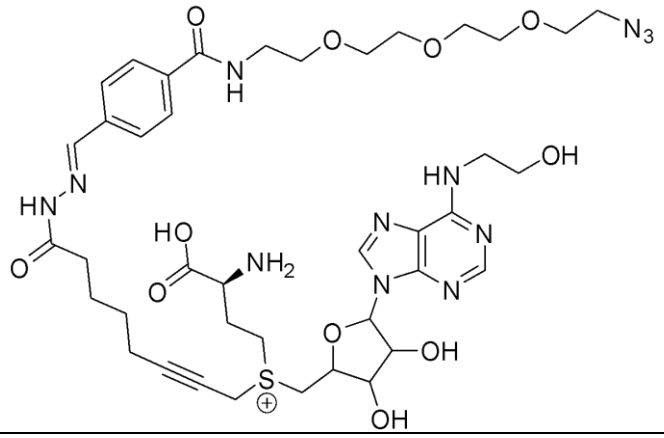
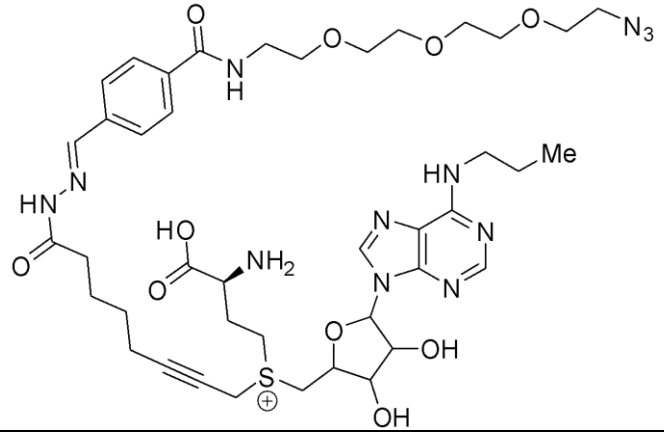
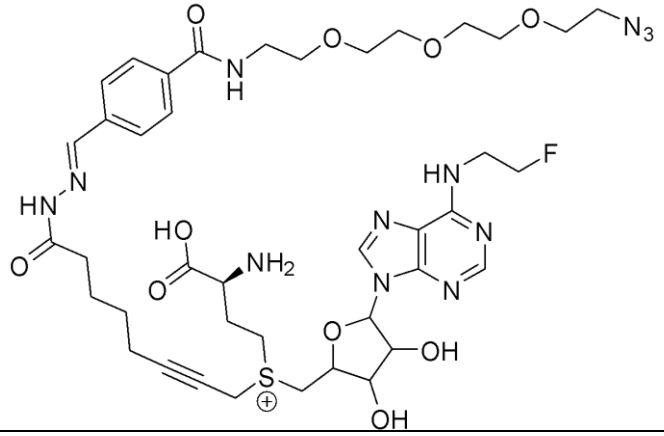
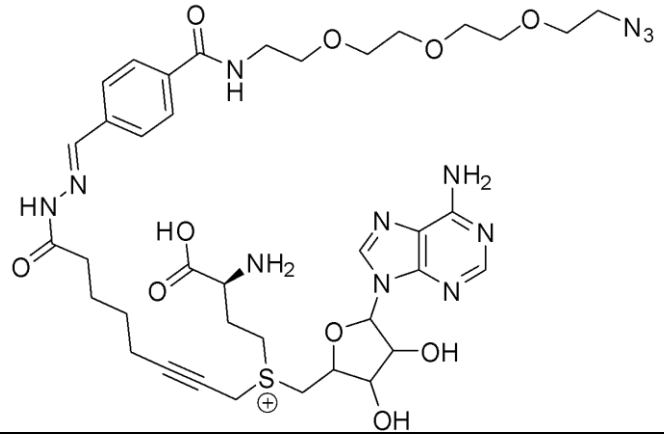
Abbreviations List

Abbreviation	**Meaning**
MTase	Methyltransferase
SAM	*S*-adenosyl-L-methionine
SAH	*S*-adenosyl-L-homocysteine
TET	Ten-eleven translocation
DCM	Dichloromethane
EDA	Ethylenediamine
n-BuLi	*n*-butyllithium
HMPA	Hexamethylphosphoramide

PPh_3	Triphenylphosphine
CBr_4	Carbon tetrabromide
THF	Tetrahydrofuran
NMP	*N*-methyl-2-pyrrolidone
TFA	Trifluoroacetic acid
RT	Room temperature
HPLC	High performance liquid chromatography
SpAAC	Strain-promoted azide alkyne cycloaddition
DNA	Deoxyribose nucleic acid
RNA	Ribose nucleic acid
5mC	5-methylcytosine
5hmC	5-hydroxymethylcytosine
5fC	5-formylcytosine
5caC	5-carboxylcytosine
BER	Base excision repair
AID	Activation-induced cytidine deaminase
APOBEC	Apolipoprotein B mRNA-editing enzyme catalytic polypeptide
TDG	Thymine-DNA glycosylase
DSBH	Double stranded β-helix
AML	Acute myeloid leukaemia
DAC	5-aza-2'-deoxycytidine
siRNA	Small interfering RNAs
miRNA	MicroRNAs
lncRNA	Long non-coding RNAs
mTAG	Methyltransferase-directed transfer of activated groups
HNSCC	Head and neck squamous cell carcinoma
uPA	Urokinase-type plasminogen activator
MMP(s)	Matrix metalloprotein(s)
IVTT	*In vitro* transcription translation
GABA	γ-aminobutyric acid
PBS	Phosphate buffered saline
IPTG	Isopropyl β-d-1-thiogalactopyranoside
1	*((S)-3-amino-3-carboxypropyl)((5-(6-((2-aminoethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(8-(2-((E)-4-((2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)carbamoyl)benzylidene)hydrazineyl)-8-oxooct-2-yn-1-yl)sulfonium*
2	*((S)-3-amino-3-carboxypropyl)(8-(2-((E)-4-((2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)carbamoyl)benzylidene)hydrazineyl)-8-oxooct-2-yn-1-yl)((3,4-dihydroxy-5-(6-((2-hydroxyethyl)amino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)sulfonium*
4	*((S)-3-amino-3-carboxypropyl)(8-(2-((E)-4-((2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)carbamoyl)benzylidene)hydrazineyl)-*

	8-oxooct-2-yn-1-yl)((3,4-dihydroxy-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)sulfonium
3	*((S)-3-amino-3-carboxypropyl)(8-(2-((E)-4-((2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)carbamoyl)benzylidene)hydrazineyl)-8-oxooct-2-yn-1-yl)((5-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)sulfonium*
5	*((S)-3-amino-3-carboxypropyl)((5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(8-(2-((E)-4-((2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)carbamoyl)benzylidene)hydrazineyl)-8-oxooct-2-yn-1-yl)sulfonium*

1	

2	
4	
3	
5	

Chapter 1: Literature Introduction

Many diseases that have developed are caused by mutations in our own DNA. These are known as epigenetic diseases. Genetically inherited differences can lead to serious disease; however, our understanding of the way in which certain alleles, the genetic polymorphisms that express phenotypic traits, are inherited can help in preventing serious diseases being passed down to future generations. Epigenetic diseases that arise from these mutations

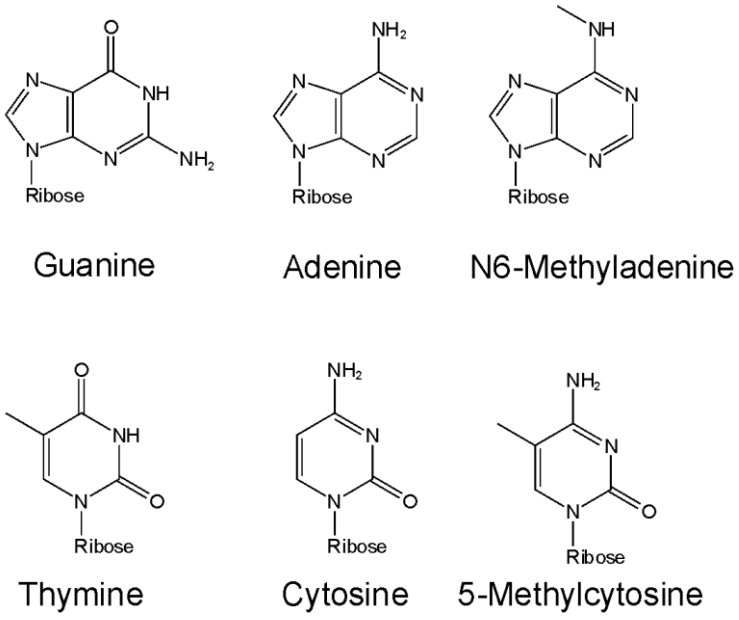
include things like cancers, autoimmune and neurological diseases such as Alzheimer's and schizophrenia. These diseases have issues in both treatment and diagnosis, which makes them very difficult to treat, since when they are diagnosed it can be too late to administer effective treatments. The other hurdle is that treatment can involve using aggressive medication which attacks the mutated DNA. However, a DNA strand with a single mutation does not appear significantly different to a healthy strand. This can then result in the treatment attacking healthy DNA as collateral.

The change to the DNA that causes an epigenetic disease is usually very small, like the addition of a single methyl group to the DNA strand, and this is why diagnosis is difficult. These groups are significantly inert so many chemical tests cannot discern a methylated base from a non-methylated base. Developments have been made to change this group from methyl to a chemically reactive group which would aid with earlier diagnoses, as well as allowing for treatment to be more effective. This would then result in less severe side effects than something like cisplatin would generate as a treatment for cancer. This would allow a new medicine to be used that is more selective to this new group than the methyl. This is also further complicated by the fact that methylated DNA is important for several biological processes. These will be discussed in depth further in this review.

1.1 DNA Introduction

This thesis focuses heavily on DNA and its chemical reactivity with enzymes and their relation to disease. As DNA is a vast topic the focus of this introduction will be on the bases and how the structure of DNA is depicted in the literature. This will help in the understanding of other processes that are discussed in this thesis, such as epigenetics, mainly DNA methylation.

DNA is made up of four bases. These are Adenine (A), Cytosine (C), Guanine (G) and Thymine (T). These four bases align themselves in a way that are complementary to each other. In DNA, the A and T bases are complementary, and the C and G bases are also complementary to each other, due to the ability for the bases to form hydrogen bonds to each other. A and T can form two hydrogen bonds to each other whilst C and G make three hydrogen bonds to each other.[1]



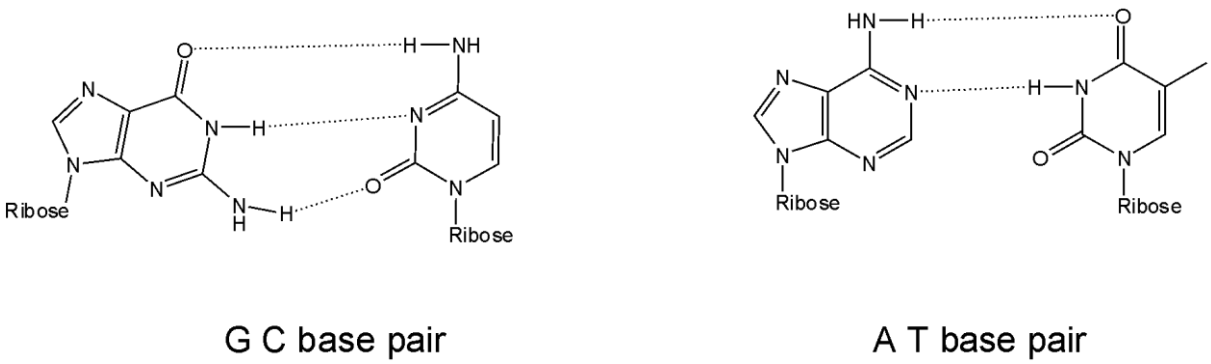


Figure 1: Structures of DNA bases, including methylated forms (above). Complementary base pairs, showing the hydrogen bonds.[2]

When this is repeated with several complementary pairs this forms a new structure, known as the double helix.[2] This double helix arranges in this format and this is advantageous in many ways. The double helix is a stable structure and maintains the stability of the DNA bases and corrects errors. This structure is important in the function of DNA for protein synthesis. This double helix provides a scaffold for correcting errors and removing damaged DNA. This double helix structure also gives rise to two new features, which are known as the minor groove and the major groove.[3] This is the space that is found between the backbones of the helix, the minor grove is where the backbones are close together and the major grove is found where the backbones are far apart.[4] Both grooves have different functions and different properties. Water can be found in the major groove as it is bigger and thus has the space to accommodate a number of water molecules. [5]The major groove will also contain more chemical information that will determine the difference between a CG or GC pair and an AT or TA pair. This is crucial for DNA binding proteins such as DNA polymerase, which is crucial for DNA replication.[6] These proteins will bind to the DNA via the major groove. This is advantageous as this enables the proteins to recognise DNA sequences without having to open the DNA and disrupt the double helix. Non-sequence specific DNA binding proteins such as histones or ribosomes, will bind to the minor groove which can cause the DNA to unwind or bend.[7]

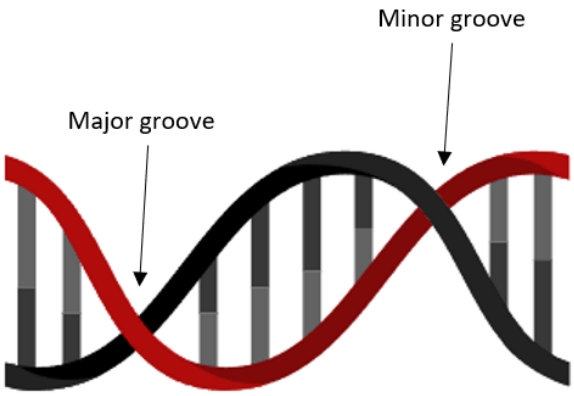


Figure 2: Structure of the DNA double helix, highlighting the major and minor grooves.[7]

If a DNA base is matched incorrectly this can lead to the DNA molecule becoming deformed due to the hydrogen bonds not aligning.[8] This can then lead to a mutation which can cause genetic diseases, such as cancer and neurodegeneration. Mismatched base pairs can occur when a wrong nucleotide is incorporated during DNA synthesis; if the nucleotide is chemically damaged and incorporated into DNA, or if a damaged nucleotide is present in a template strand and an undamaged one is incorporated opposite.[9] These mismatched pairs can be repaired by certain enzymes which will be discussed later in the thesis. If left unchecked, however, this could lead to mutations as the DNA replicates leading to diseases, and potentially, an increase in the rate of spontaneous mutations.[10]

1.1 Epigenetics

Epigenetics is the study of heritable traits that can impact on cellular biology, but that do not alter the genetic sequence.[11] Some diseases can develop due to changes in gene expression that result from a change in the epigenetic status of the genome; these are known as epigenetic diseases.[12] Epigenetic modifications allow the cell to control which genes are expressed at key moments, i.e. changing whether a gene is transcribed or not. Epigenetic modifications are achieved by one of three main processes:

- Non-coding RNA: this is defined as the process when DNA makes an RNA strand that can control gene expression. This is achieved by attaching the non-coding strand to a coding RNA strand alongside proteins that break down the coding RNA, which in turn regulates how much protein is made. [13,14]

- Histone modification: this is defined as the process where a DNA molecule wraps around another protein known as a histone. When DNA is wrapped around the histone, proteins are unable to read the DNA, so this gene is not transcribed. When the gene is required, the DNA is unwrapped from the histone allowing the gene to be transcribed. Chemical groups can be added or removed from the histone protein causing the DNA to pack more tightly or loosely around the protein. This will then affect whether the gene is switched or transcribed.[15]

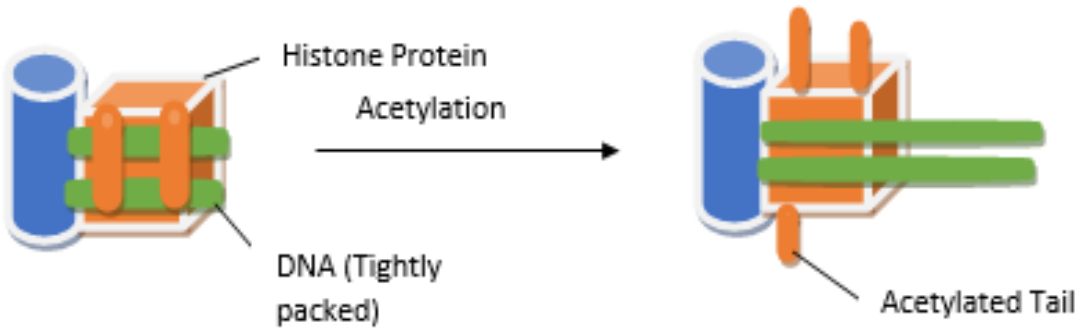


Figure 3: Cartoon illustration showing histone modification and how it blocks DNA transcription, left side shows histone packs DNA tightly and prevents proteins gaining access to the DNA, the right shows post acetylation where tails have released the DNA and now is now loosely packed giving the gene the ability to be transcribed. [15]

-The final method is DNA methylation which is the primary focus of this thesis. DNA methylation is when a DNA base is modified by addition of a methyl group, usually either added to base cytosine or eukaryotes.[16] The addition of a methyl group to the base can change whether a gene is expressed or not depending on the gene that is being added to. When some important genes are methylated they can become inactive. An example of this are the tumour suppressor genes which will be discussed later in this review. One such

disease is liver cancer. This disease is often caused by factors such as another virus (hepatitis C), microbial toxins (aflatoxin) or alcohol consumption, but the risk of this disease also has a genetic component such as hereditary hemochromatosis.

1.2 DNA Methylation

DNA methylation works by addition of a methyl group to a DNA base, either cytosine or adenine. Despite the methyl group being chemically unreactive it still has an effect on the physical properties of the DNA helix. This methyl group is added to the 5 position of cytosine which then adds a hydrophobic group to the major groove of DNA.[17] This can slightly alter the structure of the helix as the methyl group causes the major groove to widen whilst simultaneously causing the minor groove to narrow.[18] The magnitude of the change in shape depends on the sequence environment. The result of this structural change not only affects how DNA-binding proteins interact with the DNA but can also affect how transcription factors bind to DNA. The addition of the methyl group also increases the hydrophobic component of base stacking which can result in the DNA becoming more thermally stable. This methyl group can also block transcription factors from binding to gene promoter sequences.[19]

Methylation of DNA is known to be one of the mechanisms by which people can develop cancer. This is supported by findings in cancer cells; that the genome has regions that have increased DNA methylation (hypermethylation). This causes certain genes to not be transcribed. Key genes that are found to have this methylation in cancer are the tumour suppressor genes.[20,21] Tumour suppressor genes control the cell division rate. If the tumour

suppressor genes get switched off, then damaged cells can continuously divide at rapid rates which can then develop into a tumour. Tumour suppressor genes aid in preventing cancer by being responsible for informing a cell to undergo apoptosis (programmed cell death) when damaged DNA is found within the cell, or when a process fails such as DNA replication.[22] These genes also encode the necessary information that prevents further replication of genetic material through cell division, which can be detrimental to health.[23] If it is switched off then damaged cells can start dividing and this damaged DNA will be passed down to future generations of cells. [24]

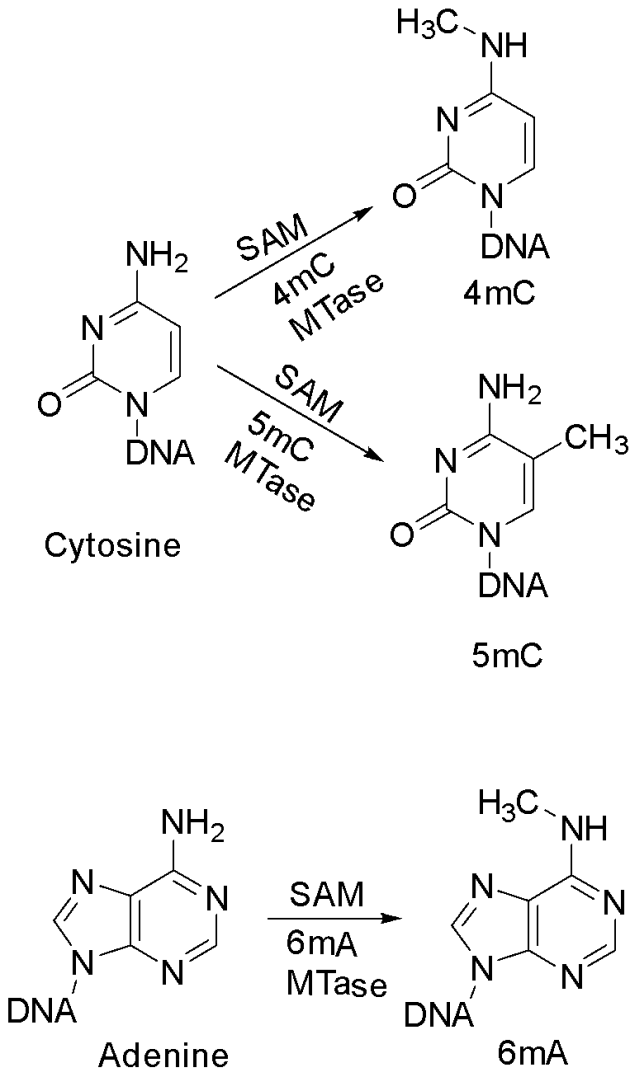
The ability to selectively transcribe genes could be a powerful tool that would aid in treatment of many diseases, if it were able to be controlled. An example would be in a therapeutic sense, if a patient was suffering with an epigenetic disease and we could identify what it was, and the genes that were involved, we would be prevent the genes that are responsible for this disease being transcribed or transcribe the genes that were responsible for preventing the disease from occuring in the first instance. An example of such a use would be the treatment of cancer; the tumour suppressor genes would not be transcribed, allowing for its reverse and thus allowing for a prevention of the development and progression of the cancer. When DNA becomes methylated there is a time when the base needs to be reverted to cytosine. This can be achieved through demethylation, which can be achieved by the use of Ten-Elven Translocation (TET) enzymes.[25]

1.21 The Mechanism of Cytosine C5 Methylation

Methylation is catalysed by MTases. These enzymes cause methylation to a range of different biomolecules such as DNA, RNA, proteins lipids and polysaccharides.[26] This methylation is crucial to many processes within biological systems including, DNA repair, regulation of replication and regulating protein expression. In a bacterial system these

enzymes protect host genomic DNA from being digested by a restriction enzyme; this process allows the system to discern between healthy and damaged DNA.[27]

DNA methylation is catalysed by MTases with the cofactor *S*-Adenosyl-L-methionine (SAM), this supplies the methyl group that is transferred to the DNA base.[28] In mammalian genomes, the DNA base that is predominantly methylated is cytosine.[29] Adenine has also been known to be the target of methylation (at the N6 position) but this is more common in bacteria and plants, and has only recently been seen in mammalian DNA.[30,31] Cytosine methylation is much more widespread in eukaryotes. Cytosine has two positions at which the base can be methylated; either C5 or N4. [28,29] These bases are often abbreviated as mC for methylated cytosine, and mA for methylated adenine.



Scheme 1: Structure of cytosine (top) and the two methylated forms (4mC & 5mC) and adenine (bottom) and its methylated form (6mA). [29] [30]

These new methylated bases can be detected through a method called bisulfite sequencing. This method converts unmethylated cytosine into uracil in the library preparation step and after a PCR step, these converted bases are then identified as thymine from next generation sequencing. The untreated DNA sample can then be sequenced and any cytosine that is present in both can then be labelled as a methylated cytosine base. This method can be advantageous as it can be used to analyse low-density CpG regions, regulatory elements

and incomplete methylated domains. This method is expensive, however, and requires a large amount of DNA. The conditions of the process (Thermal and acidic) can also degrade the DNA sample.

In replication regulation for bacteria, methylation is used to ensure that as soon as replication begins the replication origin becomes hemimethylated[32]. This is achieved by the MTases, which causes the origin to be suppressed for a period. This is vital, as the interaction between the origin and a SeqA protein is essential for timing chromosomal replication initiation.[33] in protein expression these also play a role in certain proteins being transcribed or repressed, an example of this would be DNA adenine methylase in *Aggregatibacter actinomycetemcomitans*, when the gene that transcribes this enzyme is knocked out, the bacterium shows dysregulated levels of the protein, which in turn changes the bacterium's levels of leukotoxin.[34]

This is also important for strand discrimination when undergoing DNA repair. DNA repair timing and location is controlled by the DNA adenine methylase enzyme, which is involved in differentiating between parent and daughter strands of DNA, Heithoff et al[35] concludedthat without the DNA adenine methylase enzyme present, the bacteria has a higher rate of mutation.

MTases have different classifications based on their structural folds, many of these belong to class I which incorporates the Rossmann-fold.[36] Class II includes the MetH reactivation domain, class III comprises of precorrin-4 MTases, the SpoU-TrmD (SPOUT) family of RNA is included in class IV, and class V contains SET domain protein MTases.[37] The Rossmann fold is a protein structure which has repeated units in tandem composed of 4 α-helices intercalated between 6 β-strands. This assortment of structures gives an alternating β – α – β

pattern, allowing the β-strands to be aligned parallel and held together by hydrogen bonds, Figure 4 shows the structure of this fold. [38]

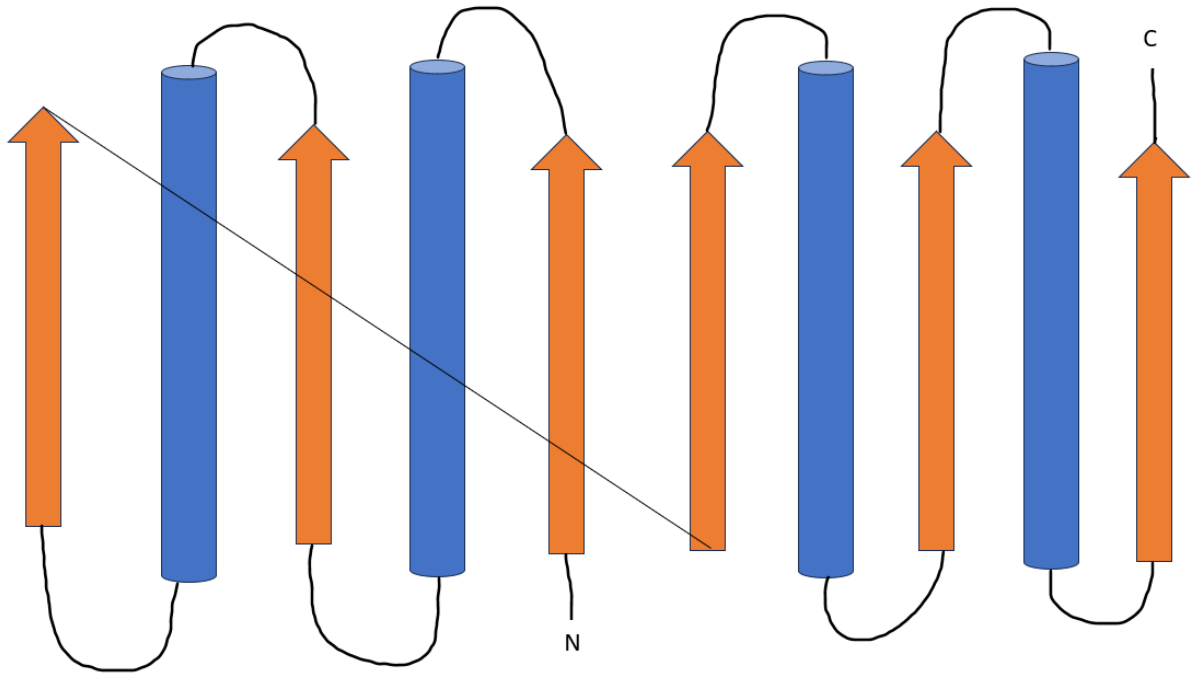


Figure 4 Cartoon schematic of Rossman fold, orange arrows show the β-sheets and what direction they travel in, blue cylinders show α-helices.[29]

As previously mentioned the β-sheets are made up of six strands, but the final sheet can include up to seven strands, which are aligned in the order of 3-2-1-4-5-7-6 instead of the usual 3-2-1-4-5-6. The 7th strand is aligned antiparallel to the others, resulting in a sandwich topology;[38] It has been found that these strands have a shared nucleoside binding motif which has a glycine-rich region found at β-1 loop which is found in between β1 and the α1 helix.[39] This region of the enzyme interacts with the cofactor, and then β2 supplies an acidic residue such as aspartic or glutamic acid to interact with the ribose of the cofactor in a bidentate manner.[40]

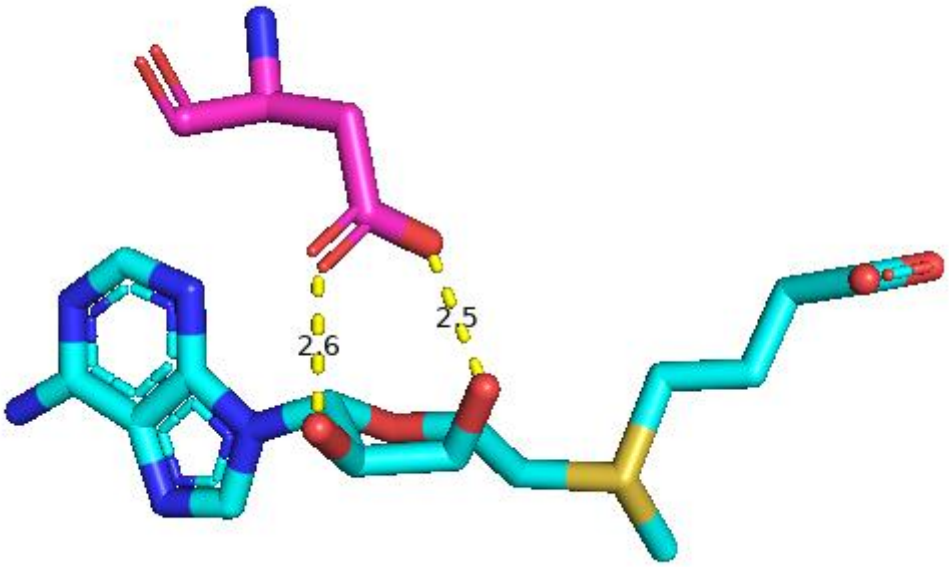
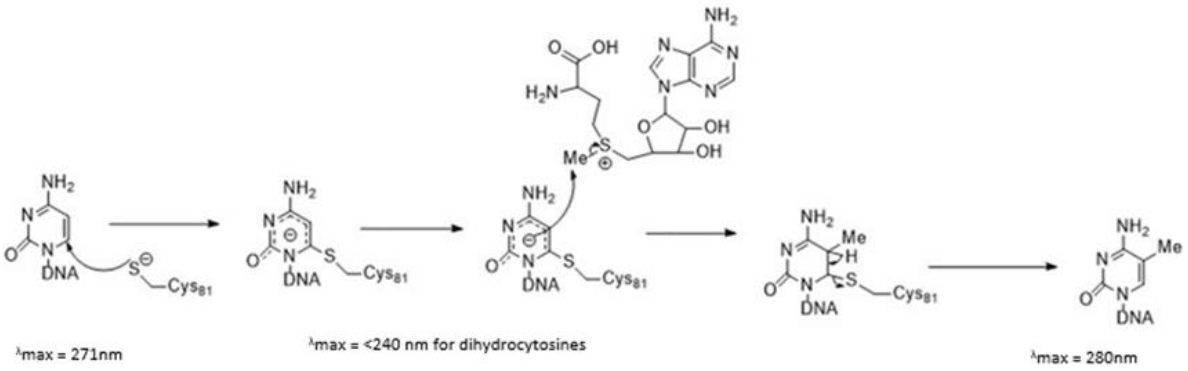


Figure 5: S-Adenosyl-L-methionine (blue) binding to an aspartic acid (pink) residue in bidentate manner shown by yellow hatched lines. [40]

By studying MTases that did not contain a Rossman fold, and enzymes that are not MTases that did contain the Rossman fold, Chouhan et al[41] discovered that the glycine-rich motif was unique for MTase enzymes. This thesis will mainly focus on class I MTases.



Scheme 2: Mechanism of methylation of the cytosine base [41]

Prokaryotes use DNA methylation as a protective mechanism that allow the host bacterium to distinguish between its own DNA from that injected by an invading bacteriophage.[42] They work in tandem with a restriction enzyme that has the same DNA recognition sequence. In the host genome, these sites are methylated and protected from restriction endonuclease activity.This is why, when unmethylated phage DNA is injected into the cell, it is rapidly digested by the restriction enzymes and proliferation of the bacteriophage is prevented. [43]

For eukaryotes, DNA methylation is used in more varied processes, but the main method it is used for, and the focus of this project, is epigenetic control.[44] This involves the genes being switched from transcribed to not-transcribed without changing the genetic sequence.[45] When a gene is silenced or not-transcribed this means the phenotype that it was originally transcribing for will no longer be expressed.[46]

The mechanism of enzymatic cytosine methylation has been elucidated. One key discovery, made by Gerasimait et al. , was the observation that the flipping out of the base occured simultaneously alongside the covalent bond formation. This was observed during the monitoring of the UV absorbance change that occured in hyperchromicity and the loss of the cytosine chromophore.[47] This works in tandem with the SAM cofactor. This cofactor donates its methyl group to the cytosine; when this happens the DNA base that receives the methyl is

flipped out of the DNA helix into the active site of the enzyme. This is achieved by breaking the H-bonds between the bases. When this is “flipped” out into the active site of the enzyme, the base is near the cofactor and so it can receive the methyl group. For the prototypical M.HhaI MTase, Gerasimait et al. hypothesised that base flipping happens simultaneously with the closure of a mobile, catalytic loop of the enzyme. This loop closure serves to stabilise the DNA duplex during flipping, as well as maintaining the extrahelical conformation for the cytosine base.[47]

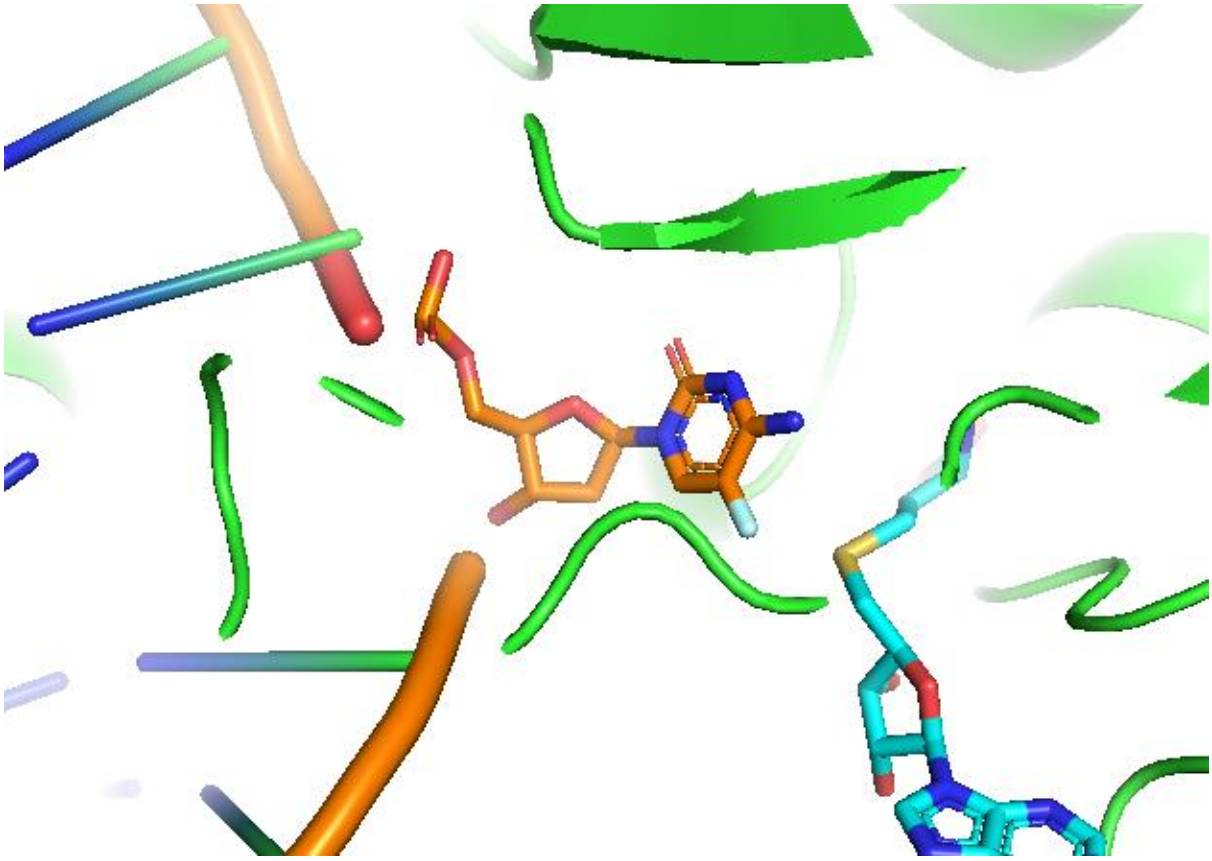


Figure 6 Pymol image of a cytosine base being flipped out of the DNA helix (orange) when bound with M.MpeI (green) with SAM cofactor (blue) PDB code 4DKJ. [47]

This position enables the cytosine to be modified to an intermediate known as a dihydrocytosine which can be observed by UV, as this causes changes in the aromatic system which in turn changes the wavelength that the cytosine absorbs at. The methyl group is then transferred from the SAM cofactor to the cytosine forming the methylcytosine. This new base is then flipped back to reform the DNA helix.

J. Yangs et al. proposed a new explanation of the mechanism which postulates this process being elucidated further.[48] The cytosine base could be flipped into a pocket located on the enzyme. The first stage of such a reaction begins with an amino acid of the MTase reacting with the DNA base. The thiol group of M.HhaI Cys81 is deprotonated, and the charged sulphur attacks the C6 of the cytosine which causes a negative charge to be distributed among the cytosine base, due to breaking of the double bond between the C6 and C5. This negative charge on the DNA base attacks the methyl group of the SAM molecule, quenching the positive charge on the sulphonium ion. This transforms the cofactor from SAM to *S*-adenosyl-L-homocysteine (SAH) which is removed from the active site and can be used to regenerate SAM in other reactions.[49] This now methylated cytosine is removed from the MTase by reforming the double bond between the C5 and C6. β-hydrogen elimination, causing the double bond to reform and the 5mC base to return to its original orientation as the double helix.[50,51]

The free energy profile of this step shows that the intermediate was in a very energetically stable state. The activation energy barrier is 15.8 kcal/mol and the intermediate had a lower free energy when compared to the reactant with a value of -30.4 kcal/mol. This is a specific case for the M.HhaI MTase.[48]

Zhang and Bruice have hypothesized that a base is required for the methylation of cytosine, and adduced that the hydroxide ion as a base comes from a proton wire that is formed from

an excess of water. This is regulated by a solvent water channel which manages the proton interchange to generate the base. This process was determined to cost about 12 kcal/mol, but this was the lowest energy value of the mechanisms that were explored. [52] From simulations J. Yangs et al. were able to determine that the barrier for the reaction had an energy value of 8.7 kcal/mol; combining this with the energy required to generate the hydroxide ion gives an overall energy barrier of 20.7 kcal/mol which makes this step the rate-determining one.

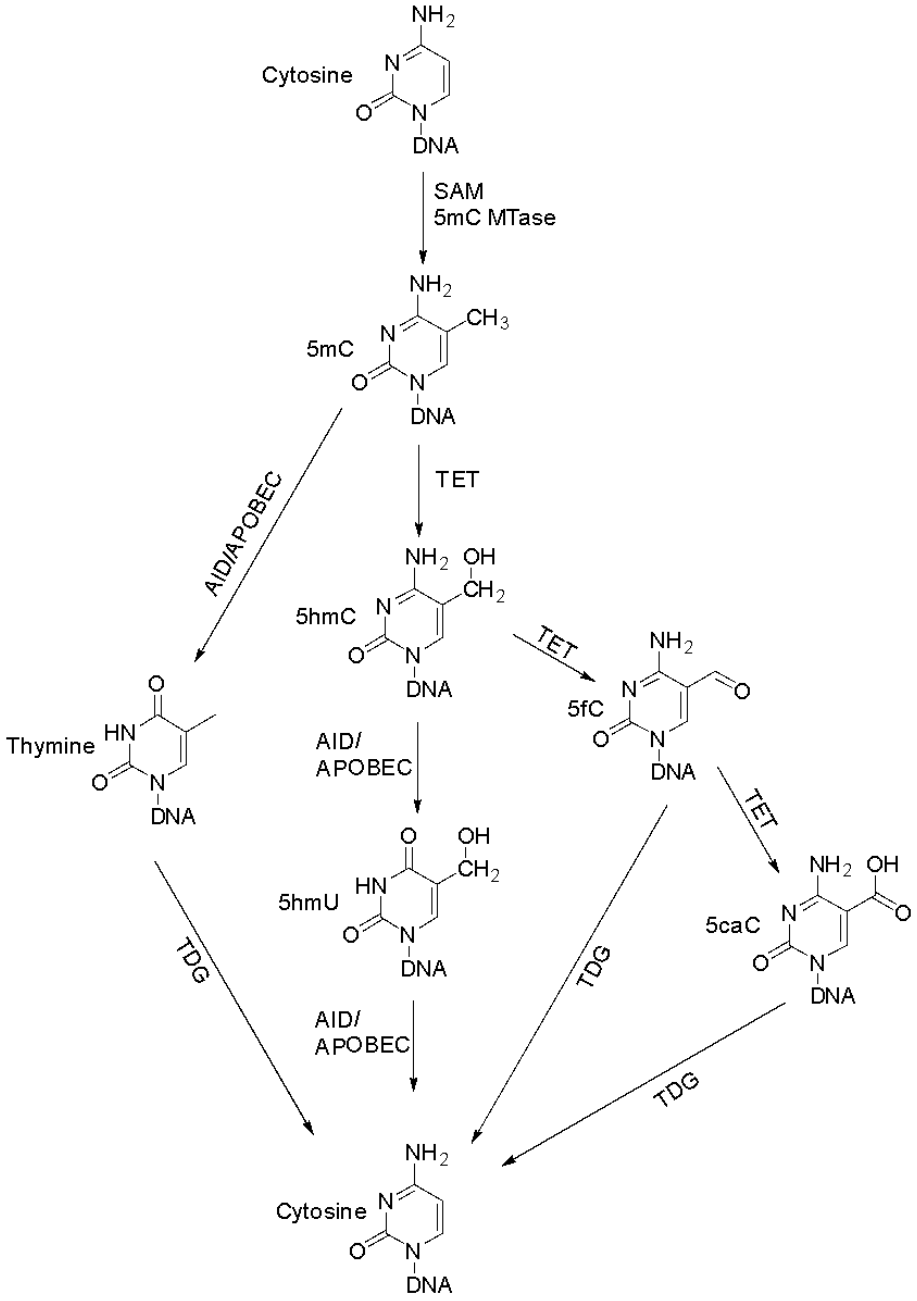
1.3 Life Cycle of 5mC

Methylcytosine (5mC) can be demethylated by two mechanisms; the pathway followed depends on the enzyme that is present.[53] The first pathway; 5mC is removed from the genome through a process that is managed by the base excision repair (BER) enzymes.[54] The first step of this is deamination by activation-induced cytidine deaminase (AID) (a deamination enzyme), whilst apolipoprotein B mRNA-editing enzyme catalytic polypetide (APOBEC) is an enzyme complex. If either of these are present, then the amine is converted into a carboxyl. The resulting base of thisis thymine which can then revert to cytosine via the thymine-DNA glycosylase (TDG) enzyme.[55]

5mC Can be converted into 5-hydroxymethylcytosine (5hmC) through the ten-eleven translocation (TET) family of dioxygenases.[56] Through further research it was observed that these enzymes can oxidise the 5hmC base to products 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) which can then subsequently be removed out of the genome by TDG.[57]

1.31 DNA Demethylation by TET Enzymes

TET enzymes are responsible for converting 5mC back to cytosine. This is achieved by oxidising the methylated cytosine to become 5hmC[58], this species can continue to further oxidise as shown in scheme 3.



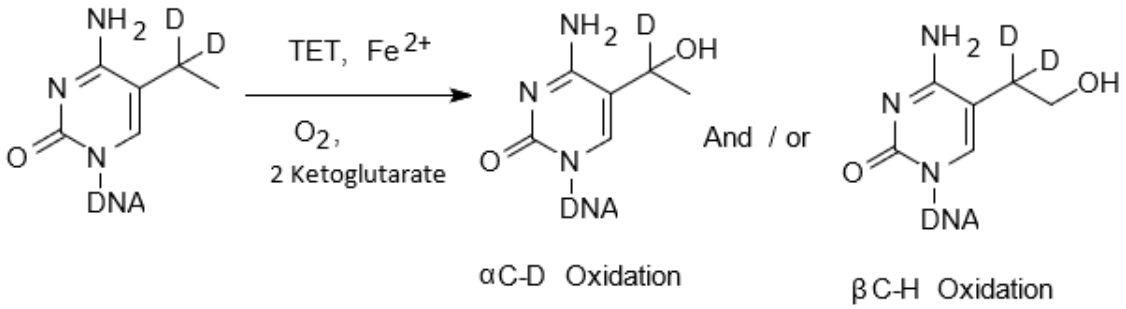
Scheme 3: Cytosine methylation cycle. [58]

The TET family contains three members which are labelled as TET1, TET2 and TET3. All members of this family contain a region for catalysing oxidation at the C-terminus which has

sites that are rich in cysteine and contains double stranded β-helixes (DSBH) as well as cofactor binding sites. These features along with an alpha-ketoglutarate, which can act as a co-substrate, enables oxidation of 5mC to 5hmC. Within the CXXC domain of TET1 and TET3 another feature was identified, which was a zinc finger, the purpose of this finger is mostly unknown. This CXXC region is found in other proteins and are known to bind CpG dinucleotides which are unmethylated, however the Tet1 CXXC domains not only recognise cytosine but also recognise the 5mC and 5hmC products as well. It also shows preferential binding to regions rich in CpG content.

The TET2 enzyme has historically been the most studied.[59] Not only is this enzyme able to demethylate DNA through an oxidative process, it also has been shown to have a cancer suppressing mechanism.[59] This was observed through studies of the TET2 enzyme as it is one of the more commonly mutated genes in haematopoietic malignancies.[59] This mutation can affect its function. Though this was observed with the TET2 enzyme originally, further studies have shown that all three members of the family can become mutated, resulting in all TET enzymes being expressed at a reduced amount. Due to the occurrence of this being a very early event in the process, it was concluded that this must have a role in tumour suppression. A recent study done by Kavoosi et al.[59] used methylated DNA as a control to compare the effect of ethylated, propylated and iso-propylated DNA on the TET enzyme. Their results showed that the iso-propylated DNA base was unable to undergo further oxidation. The propyl labelled DNA base showed that a fraction of propyl-labelled DNA was oxidised to the next product. Ethylated DNA, however, showed it could be oxidised selectively to 5-hydroxethyl cytosine, but it also had a comparable rate to methylated DNA being oxidised, though further oxidation of this product was not observed. The former two analogues (iso-propylated and propylated DNA) of 5mC not oxidising, could be explained by sterics; the binding site of TET2 is unable to accommodate the larger cofactor. Next, regioselectivity was investigated to determine which alkyl group (CH_2 or CH_3) is preferably

oxidised. This was determined by converting the CH_2 to CD_2. When the deuterated base was introduced to the TET2 enzyme, hydroxylation was observed to occur at the heavy methylene site. This was concluded to be due to an increase of fifteen mass units as the OH is exchanged for D. If the other site was hydroxylated an increase of sixteen mass units would have been observed. This observation supports a radical based mechanism with the stability of the intermediate owed to the benzylic radical forming at the CD_2 site. Specific oxidation levels could be explained by the lack of conformational freedom for the ethyl group when it is within the TET2 binding pocket.[59]



Scheme 4: Reaction scheme showing the formation of 5-hydroxyethyl cytosine from 5-ethylcytosine [59]

An alternative mechanism for the removal of 5mC from the genome, is via the demethylation pathway.[59] The TET enzyme converts 5mC to 5hmC, this base will then have two paths it can follow. If AID/APOBEC is used, then the primary amine is removed, which generates 5-hydroxymethyl-uracil which can be reverted to cytosine via thymine DNA glycosylase (TDG)/single-strand selective monofunctional uracil DNA glycosylase.[60] If 5hmC is further treated with TET, 5-formylcytosine (5fC) is made and when exposed to TET further, this produces 5-carboxycytosine (5caC). This final product can regenerate cytosine via the BER pathway to lose the COOH group.[61]

The TET enzymes each contain a catalytic domain found at the core, which has a double-stranded β-helix fold. This fold contains important metal-binding residues, which can be found in the family of Fe(II)/α-ketoglutarate-dependent oxygenases.[62] These enzymes employ O_2 as a substrate to catalyse the decarboxylation of α-ketoglutarate, to allow for the regeneration of a reactive high-valent enzyme-bound FE(IV)-oxo intermediate which then converts the 5mC to 5hmC.[63] However, despite the core being a fraction of the TET enzymes size and where the mechanistic function happens, the rest of the protein could be important in regulatory roles.[64] All isoforms of the TET enzymes contain a cysteine rich domain that precedes the core. This appears to be required for activity.[65]

Since its discovery, 5hmC has been extensively studied. It has been found to be present in several processes, however the reason for its presence has not been deduced yet.[65] One likely explanation for its function is its involvement in epigenetics. This base has been linked to the reduction of the effect of gene silencing that is provided by the 5mC base. This is achieved by the 5hmC base being bigger than 5mC, which prevents methyl-binding proteins from attaching to the base. [65] This has been hypothesised from previous studies on chromosomes which did show that 5hmC is mainly euchromatic.[65]

The removal of a methyl group from most things has a very high energetic barrier, however removing the methyl group from an oxidised base is more realistic.[66] Decarboxylation of 5caC could revert the base to the original cytosine which one study has suggested with the aid of isotopic labelling of 5caC.[67] To date, however, a 5caC decarboxylase has yet to be identified. It has been observed that without a methyl donor present, DNMT enzymes are able to promote the addition or removal of oxidised 5-position substituents, which include reacting with 5hmC in vitro. DNMTs could thus theoretically function in demethylation, but the relevance of this reverse DNMT reaction is unknown as SAM is a general methyl-group donor that can be found in every cell type.[68]

An alternative pathway for the regeneration of cytosine involves the DNA repair enzymes. [69] Though pathways involving nucleotide excision repair have been investigated, the primary focus has been on BER. This pathway results in removal of the whole modified base and then followed up by subsequent repair to replace the residue with an unmodified C.[70] Several key parts of the BER pathway are present at important transition steps of DNA methylation patterns.[71]

The TDG enzyme has been shown to interact with several transcription factors, chromatin modifying enzymes and DNMTs. This raises the possibility of a functional role for TDG in maintaining gene transcription, either through glycosylase activity or as a transcriptional coactivator.[72]

The next big discovery in DNA demethylation was shown that unlike other DNA glycosylases, TDG is compulsory for embryonic development.[73] Studies on TDG-null embryos or using a catalytically inactive mutant infer an epigenetic abnormality. Amongst other alterations the mutants showed a decrease in expression of important developmental transcription factors, such as the *hox* gene family members, and also showed perturbed methylation at their regulatory sequences. However, even though these studies show a possibility of TDG being involved in DNA demethylation, its genomic target is unknown.[74]

TDG has been a focus of many biochemical and structural studies due to its intriguing role as a DNA repair enzyme that is able to remove a normal DNA base.[75] Initial hypotheses focus on the possibility that the TDG enzyme activity would be related to deamination of 5mC or 5hmC since T.G or hmU.G mismatches are known substrates of TDG. If repair occurs, then unmodified C would be regenerated. For this model, AID/APOBEC enzymes were used as the likely candidates that would catalyse deamination, and some of the studies have suggested a role for the deaminases in the reprogramming of stem cells or in embryogenesis.[76,77] These pathways of deamination-mediated demethylation could also involve the DNA damage response protein GADD45 or MBD4 as potential alternative

enzymes for excision of T.G mismatches. However, a challenge is that the deamination of the 5hmC base with AID/APOBEC enzymes is not able to be detected in vitro or within cells.[78] However, the deamination of 5mC by AID/APOBEC enzymes can occur *in vitro* at a rate that is ten times slower than when compared to deamination with unmodified C.[79] This makes the role of deaminases in demethylation very limited since there is a lower concentration of 5mC than unmodified C present in the genome. This is also limited by the enzyme having selectivity for single-stranded DNA and prefers specific sequence contexts, which is backed up by observation of no significant developmental defects which are associated with a deficiency of AID/APOBEC.[80,81]

Though limited, the role of TDG in processing the T.G mismatches that occur from deamination of 5mC by AID/APOBEC has remained as one possibility for its requirement for embryonic development. It was observed previously that some matched base pairs, those containing modified C bases, could be targeted by the TDG enzyme. The C bases with substituents on the 5-position are able to destabilise the *N*-glycosidic bond through electronic effects, these would include the 5fC base which has been shown to be excised efficiently by glycosylase.[82] This result opens up the possibility that TDG could excise TET oxidation products. Though no impactful in vitro base excision activity has been seen with the other oxidised bases, TDG has robust in vitro base excision activity on 5fC and 5caC when it is paired to a G base in duplex DNA. This *in vitro* activity is very important in cells, shown by knockout of the gene that encodes for TDG, which results in increased concentration of 5caC in embryonic stem cells.[83] In addition to this overexpression of TDG and TET within the HEK293 cell line this will lead to a depletion of both TET-related 5caC and 5fC. These studies on TDG and TET mediated oxidation have led to the discovery of the first biological and biochemically validated path for active DNA demethylation.[84]

1.4 The M.MpeI DNA Methyltransferase Enzyme

M.MpeI is an MTase found in the prokaryote *Malacoplasma penetrans HF-2* (which was formally known as *Mycoplasma penetrans*) NCBI accession number is BA000026.2. This enzyme is classified as a transferase enzyme, this enzyme's role is to transfer chemical groups to other biomolecules. This enzyme is an MTase which means to achieve its role it transfers a methyl group to DNA. This enzyme shares the same target recognition sequence (5'-CG-3') as the eukaryotic C5-MTases, which make it a prime candidate for researching epigenetic modifications and DNA methylation in eukaryotes.[85] The M.MpeI is crystalised, so that it's structure could be observed.[86] This structure could aid in understanding the MTases CpG specificity and how tolerant the MTase in question is to sequence insertions. M. Wojciechowski et al. used a combination of 5fC in place of the substrate base. This was expected to interfere with the final step of catalysis which would then regenerate the active enzyme. This would be achieved by stopping the elimination of the cysteine found in the active site. Another replacement that was used was the co-substrate which was SAM in this particular case. M. Wojciechowski et al managed to achieve a crystal from a drop which had mixtures of SAM and the degradation product, 5'-methylthioadenosine (achieved from scission of SAM to 5'-methylthioadenosine and homoserine lactone) as well as SAH. This structure contains what is known as classic features of a C5-MTase-DNA complex, the 5fC is extruded from the DNA which is replaced by an intercalating amino acid (Gln141). The flipped base can be housed in a pocket of the MTase with the Wastson-Crick edge facing Glu184. The residues Arg228 and Arg230 are found in very close proximity to the flipped DNA base which is approximately 3.0 Å and has the ability to donate one or two H-bonds to its carbonyl group.

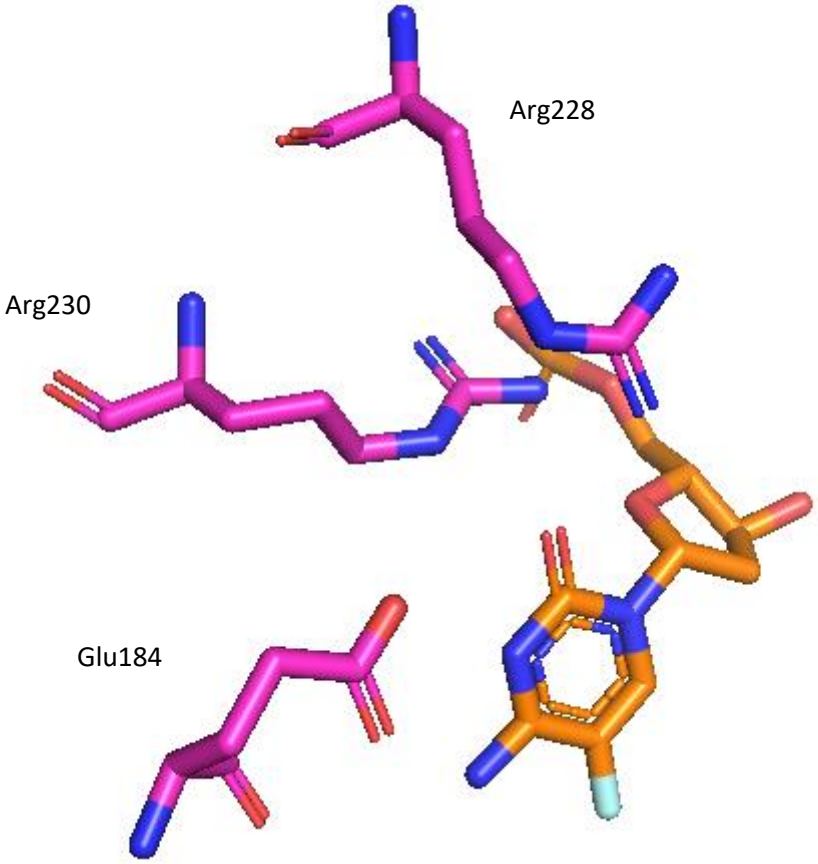


Figure 7: 5fC base (orange) from M.MpeI active site, with the close residues shown (Glu184, Arg228 & Arg230 in purple) PDB code 4DKJ. [86]

Wojciechowski et al. hypothesised that using 5fc in place of the substrate base would interfere with the regeneration of the enzyme, by preventing the elimination of cysteine from the active site. There is no observable evidence for the methyl group transfer to the substrate base, but there is also no methyl group seen on the cofactor. This could be explained by the binding site for the cofactor most likely being occupied by a combination of SAM, the degradation product, as well as SAH; the details are difficult to discern since the density of the base is well defined. The other regions such as the sugar and the sulphonium or thioether regions are not as well defined.[87] The catalytic cysteine residues of the M.MpeI (Cys135) can be found in a good position for nucleophilic attack on the C6 atom. The distance however for Cys-C6 is 2.9 Å which is an intermediate between a covalent bond

which is 1.8 Å and noncovalent interactions which can be found at 3.6 Å or greater. A similar distance has been observed for other complexes of C5 MTases when coupled with SAH which is 2.6 – 2.8 Å. This has been speculated as evidence for the generation of the partial covalent bond. If true, then it would be accompanied by a fractional loss of cytosine, however, the resolution was too poor to be able to determine this detail. [86]

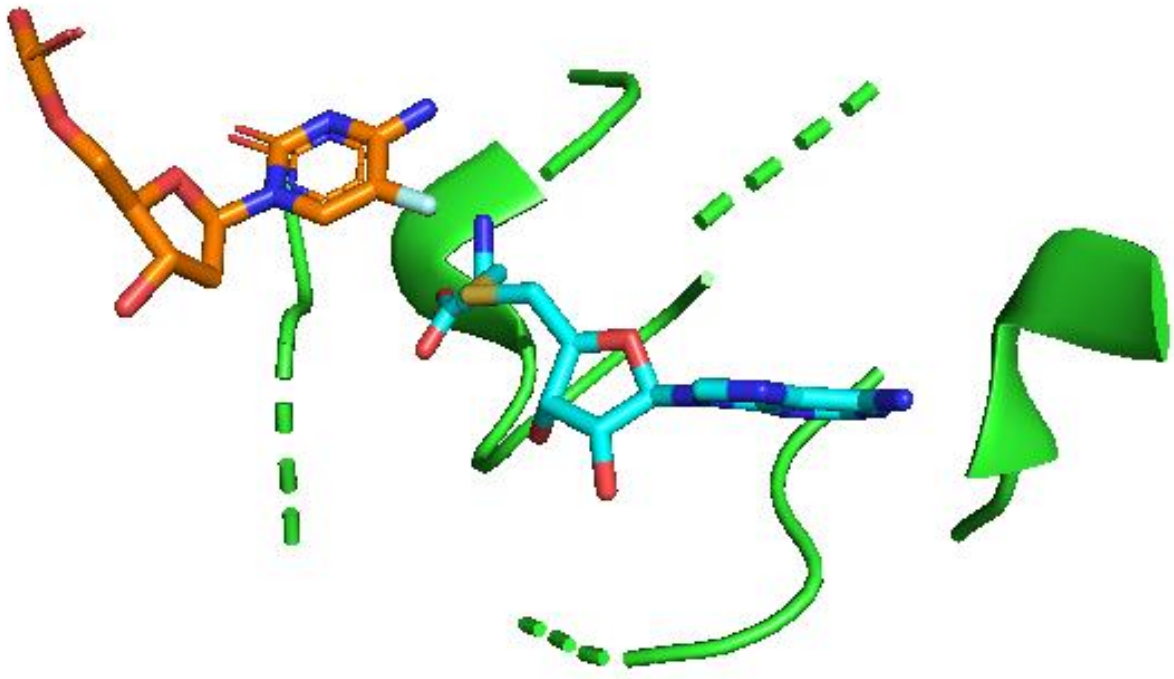


Figure 8: Crystal structure of M.MpeI enzyme active site (green) with DNA base to be methylated (orange) flipped out with SAM bound (blue) PDB code 4DKJ. [86]

Here, the M.MpeI enzyme not only interacts with the base that is flipped out of the double helix, but can also engage in other extensive interactions with three of the other bases in the helix. The guanine opposite the flipped base can partake in some hydrogen bonding between O and N, which are H-bond donors from Ser304, these donate H-bonds to the guanine O6, the N group on Asn303 is able to donate to its N7 atom. In addition to these bonds, Gln141 which displaces the cytosine base can accept a bifurcated H-bond from N1 and N2 atoms

from the guanine. The major feature of the sequence recognition by M.MpeI is insertion of Phe302 from the major groove side, to the 5mCpG step of the non-substrate DNA strand. The height of this step is nearly twice larger than usual, which ultimately leads to a substantial DNA bend or a kink. Despite this massive distortion, the second base-pair of the recognition sequence can retain the Watson-Crick H-bonding. The cytosine of this base can donate a H-bond from the N4 atom to the carboxylate of Glu305. The glutamine is on the minor groove edge and mainly faces solvent but can accept a H-bond from Ala323 N to the O6 atom on guanine.

The recognition of short sequences such as CpG requires multiple interactions. The unexpected one was intercalation of Phe302, which is because 5'-pyrimidine-purine-3' dinucleotides can unstack themselves more easily than other dinucleotides. This means, the intercalation could contribute to sequence specificity. A similar role of methionine intercalation has been reported for CpG sequence recognition by ThaI restriction endonuclease.

This M.MpeI enzyme complex can be compared to Dnmt1 (eukaryotic DNA Mtase) and M.HhaI (prototypical bacterial MTase which can methylate CpG in GCGC context). The catalytic cysteine (Cys135 M.MpeI, Cys1229 DNMT1, Cys81 HhaI) and glutamate (Glu184, Glu1269, and Glu119 respectively) as well as two arginine residues that can flank the pocket for the flipped base are all consistent across the three MTases. For all three protein-DNA complexes, the flipped base is displaced by amino acid side chains being inserted from the minor groove (Gln131 M.MpeI, Met1235 DNMTt1) and major grove (Lys1537 DNMT1, Gln237 M.HhaI). CpG recognition differs between the three complexes, M.MpeI and DNMT1 unstack the CpG step of the non-substrate strand via intercalation of an amino acid side chain from the major groove (Phe302 M.MpeI, Trp1512 DNMT1). This is not seen for HhaI. There are also observable differences between M.MpeI and DNMT1 CpG recognition. For M.MpeI, the estranged guanine interacts with the intercalating gln141, this potential mode of

interaction can make CpG recognition independent. The DNMT1 complex equivalent guanine interacts with another guanine which is located on the 5'-side of the substrate strand CpG in the crystal. However, follow-up experiments prove this interaction is unimportant in solution which means that the recognition by these two MTases could be more similar than suggested by comparison of their crystal structures.[86]

1.41 MTases in Therapeutics

There are 208 different DNA methyltransferases (MTases) in the human body, each responsible for methylating specific DNA sequences.[88] This process is crucial for regulating gene expression, but it can also contribute to disease. When certain genes are methylated, they may be silenced, which can lead to a variety of issues. If the methyl group is not removed, cells with damaged DNA can continue to divide instead of being destroyed, leading to more copies of these damaged cells, which may then divide more rapidly.[89]

Researchers are exploring MTase inhibitors as potential treatments for cancer. A 2021 study by Hu et al. found that combining MTase inhibitors with traditional cancer drugs, like cisplatin, showed higher effectiveness than using chemotherapy alone. However, more research is needed, as the study had a small sample size (<50 patients), and there are few randomized or blind trials.[90]

DNA methylation errors play a role in many diseases, as it's crucial for controlling gene expression. In cancer, DNA methylation is often disrupted, leading to global demethylation and local hypermethylation.[91] For example, in acute myeloid leukaemia (AML), mutations in the DNMT3A gene are common, and mutations in DNMT1 are linked to changes in DNA methylation, especially in colon cancer.[92] Studies have also explored how non-coding RNAs, and synthetic compounds, can regulate DNA methylation. While these studies are promising, more research is needed to determine how best to use them, either alone or in combination with cancer treatments.[93]

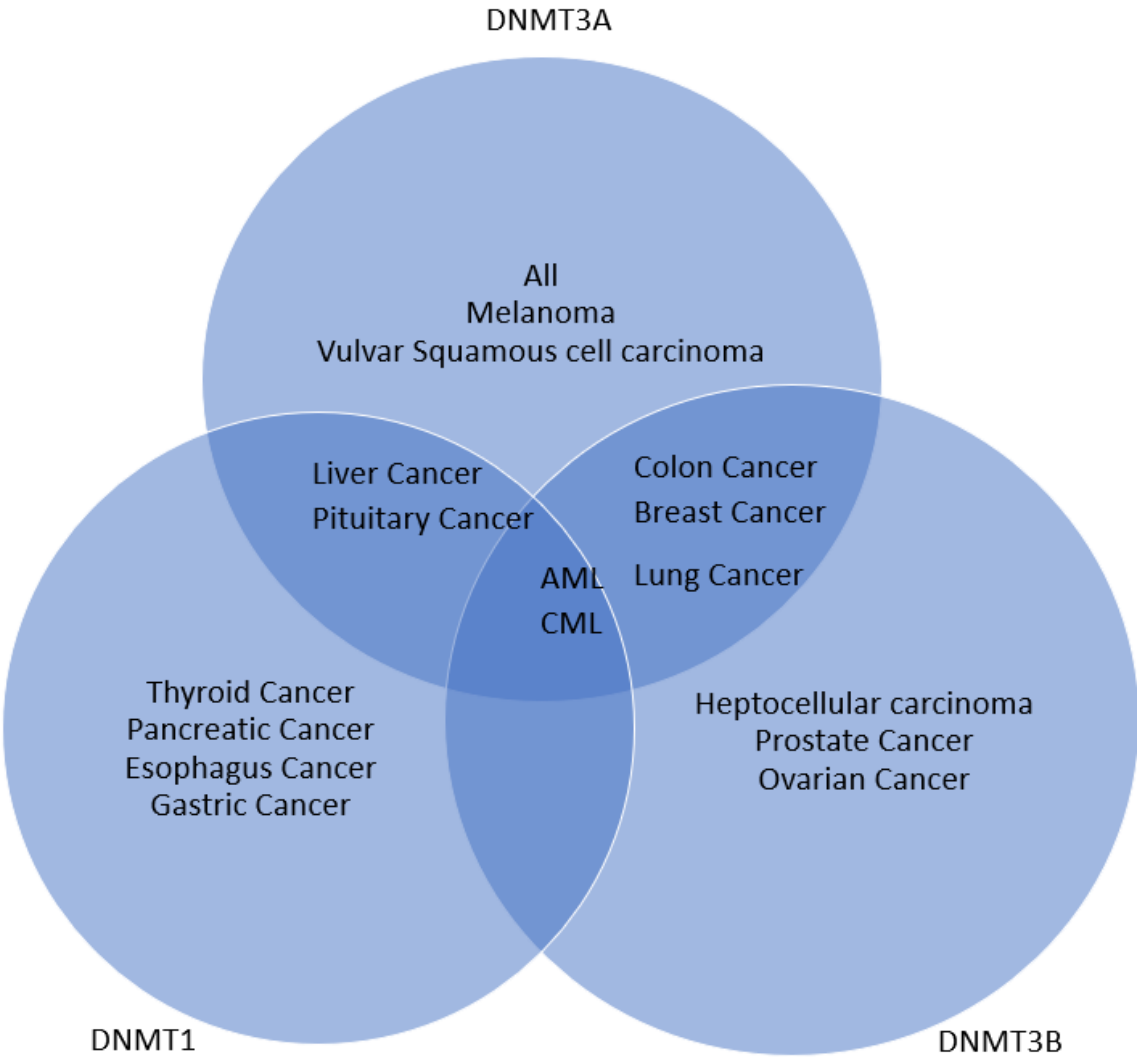


Figure 9: Venn diagram showing which MTases expression levels are altered in different cancers. [93]

There are still very few drugs approved for cancer treatment.[94] One of the main challenges for chemotherapy is improving drug efficiency, stability, and minimizing side effects.[95] Oligonucleotides and drugs that target DNMTs are becoming important tools in cancer therapy.[96] The FDA, however, has approved only a few nucleoside analogues, such as 5-azacytidine and 5-aza-2'-deoxycytidine (DAC), in order to treat AML and myelodysplastic syndrome.[97] Other compounds, like zebularine, which is more stable than the drugs currently used, are currently in preclinical trials.[98] Most recently, second-generation DAC and deoxyguanosine are now in the process of being developed, showing longer half-lives due to their resistance to deamination. [99]

A new series of quinoline-based inhibitors has also been discovered. These belong to non-nucleoside inhibitors, the first being SGI-1027[100]. This inhibitor has developed two further analogues, which are known as MC3343 and MC3353. Novel therapies are not only including DNMT inhibitors but are also including immunotherapy and HDACs inhibitors as well.[101]

Non-coding RNAs (ncRNAs), including small interfering RNAs (siRNAs), microRNAs (miRNAs), and long non-coding RNAs (lncRNAs), have been shown to inhibit the action of DNA methyltransferases (DNMTs).[102] siRNAs (21-23 nucleotides) and miRNAs (19-25 nucleotides) regulate gene expression by targeting mRNA after transcription.[103] LncRNAs, which are longer than 200 nucleotides, influence gene expression both before and after transcription. Many of these ncRNAs have reached early-stage clinical trials, but the results have been mixed. This suggests that further research is needed to determine the optimal doses and combinations for better future outcomes.

miR29b is a small molecule that can control two important enzymes; DNMT3A and DNMT3B, which add chemical marks to DNA. In cancers like lymphoma and pancreatic cancer, miR29b levels are lower, and DNMT3B is more active. Increasing miR29b can reduce cancer cell growth and trigger cell death by affecting DNMT3B. In AML (a type of leukemia), miR29b also lowers the levels of DNMT3A and DNMT3B. [104,105]

Another molecule, miR145, works in ovarian cancer[106] by reducing DNMT3A. This decreases DNA marks at the miR145 gene, making it more active and further lowering DNMT3A in a feedback loop.[107] In bladder cancer, a long molecule called DBCCR1-003 stops DNMT1 from marking DNA in a gene that helps prevent tumors.[108] In AML, when another long molecule (CCDC26) is lost, DNMT1 gets misplaced and causes changes in DNA marks. [109]

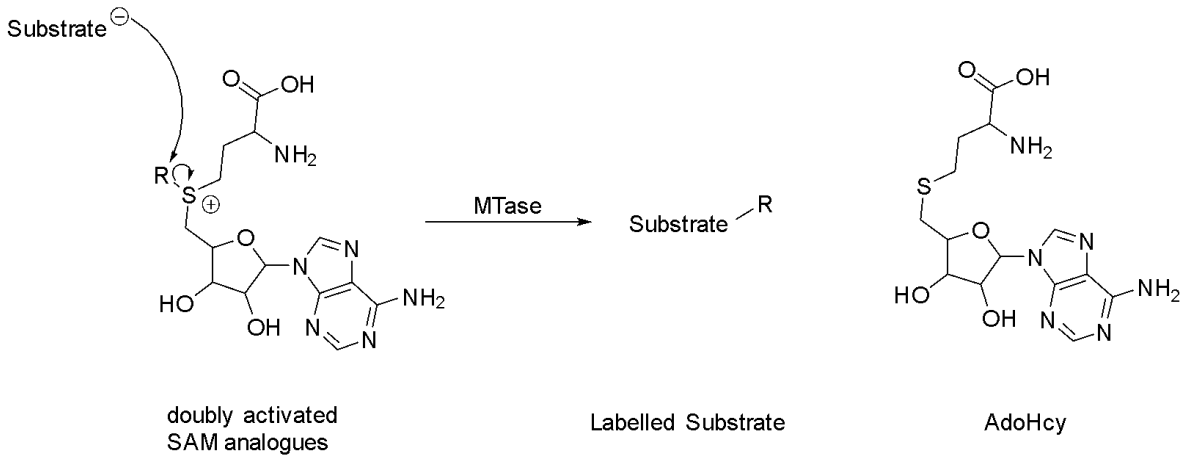
Other long molecules in colon and lung cancers can also affect DNMT1. While these molecules show some promise for cancer treatment, only two (MG98 and miR29b) have been tested so far. [110,111] A challenge with these treatments is that they can accidentally turn on cancer-causing genes, worsening the problem instead of treating it. [112]

Cancer is often caused by changes in how DNA is controlled, and DNMT enzymes play a big role in this.[113] These enzymes are influenced by molecules like non-coding RNAs, and their malfunction has been linked to many cancers.[114] Understanding how DNMTs work and how they change over time can help improve cancer treatment by predicting how mutations might affect cells. By controlling DNMT activity, it could be possible to use drugs and small molecules to treat cancer more effectively. [115]

1.5 SAM Cofactor

The SAM cofactor is used in many biological functions such as methylation, transulphuration and amino propylation.[116] The SAM molecule is produced from ATP, *L*-methionine, and H_2O

to generate SAM amongst other materials.[117] This cofactor has been the target of modification, as it has direct control on what groups are transferred to DNA. For example, by attaching a florescent tag, we can follow and observe what happens to that DNA. Another use of this cofactor is adding a reactive group to capture the DNA at a later point. This is achieved by addition to the cytosine base of a CG in DNA. The groups that are usually transferred to achieve this are a terminated azide or a terminated alkyne. This group can be transferred to the DNA and subsequently be reacted in click-chemistry fashion to capture the DNA. This is known as methyltransferase-directed transfer of activated groups or mTAG for short.[118–122]



Scheme 5: Schematic of reaction between doubly activated SAM analogues with an MTase [122].

One of the changes that has been made to the cofactor, is replacing its centre with sulphonium instead of selenium. This change was made in order to study a cofactor analogue that bears an alkynyl fragment to be transferred. With the sulphur centre, however, these analogues had a half-life of a few minutes when at pH 7.5 and would degrade along the alkyne hydration path.[123] With the exchange to the selenium centre the hydration of the alkyne was no longer observed, and had a different degradation pathway where the

formation of a homoserine lactone is produced.[124] This selenium derivative is more useful than sulpur, as the reactivity of the selenium analogue increases, which then increases the efficiency of group transfer to DNA when in presence of the MTase. This efficency can be explained by the increased level of activation of the selenium centre. This allows for proteomic analysis on the activity of protein MTases in several differing cell lines.[125,126]

Knowing that SAM is the key cofactor for methylation, analogues of SAM were synthesised in which the methyl group was replaced with other molecules. These analogues have allowed further probing into the roles of the enzyme or the DNA base for gene transcription. There are currently two types of SAM analogues that are widely used: these are doubly activated SAM molecules and aziridine based SAM molecules.[127,128] Aziridine SAM analogues have only been generated through synthetic means; these were first reported by the Weinhold group. The difference between these analogues and original SAM is that the methyl group on the sulphonium has been substituted to an aziridine ring. When methyl transfer occurs, this ring, instead of being transferred, opens due to nucleophilic attack. This causes the whole cofactor to be transferred to the DNA base of interest. [129]

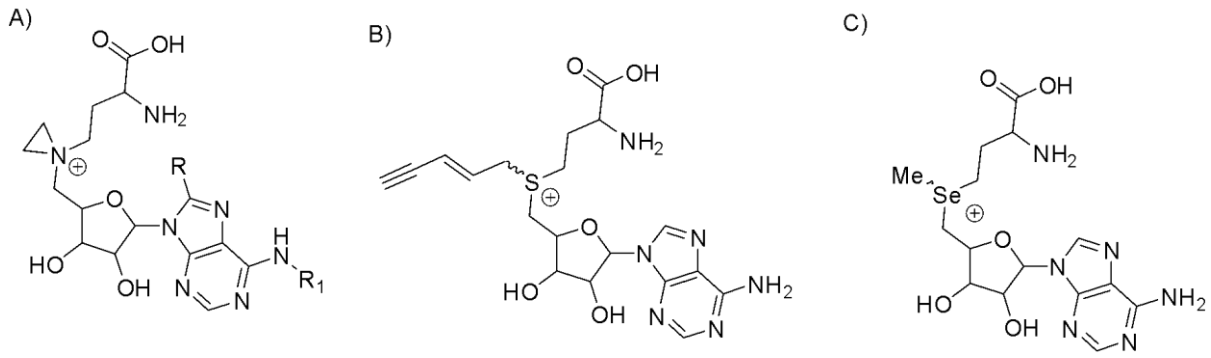


Figure 10: Structures of SAM analogues, A) aziridine SAM analogue, B) doubly activated analogue, C) selenium-based SAM analogue. [127,128 129]

After learning that multiple MTases were able to carry out the same chemistry with this aziridine attached to the sulphonium, the process of involving the addition of reporter groups

such as biotin being added to this SAM analogue was then carried out.[130,131] This brought about a new method of labelling called “sequence-specific methyltransferase-induced labelling”. Studies on the binding pocket of SAM have shown that steric clashes between the enzyme and the cofactor play a role in which analogues are compatible with specific MTases. [132] This elucidated that the 8-carbon position on the adenine group on the cofactor is generally exposed to the solvent. This means that if bulky groups are added at this position, they are more likely to be accepted by the MTase as the steric clash will be minimal between the analogue and the enzyme.[133,134]

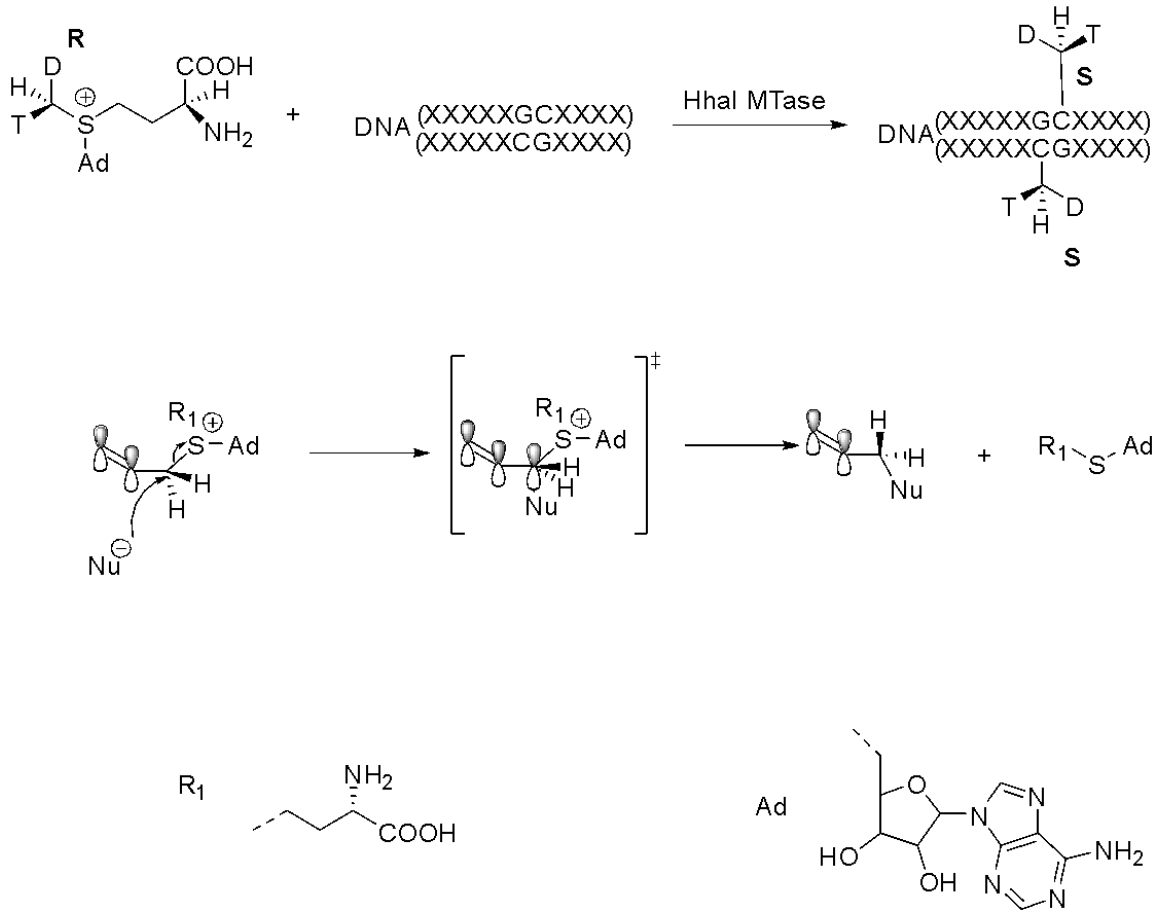
This work also developed the use of 5’-*N*-substitued nitrogen mustard derivatives instead of the aziridine.[135] The advantages of the nitrogen mustard analogues are numerous; they have been shown to have a higher stability than the aziridine analogues and have higher activity with the MTases compared to the alkyne substituted aziridines. Bio-orthognal ligation reactions are also possible if an alkyne or azide is added to the N-mustard. However, the nitrogen mustard analogues have one major disadvantage which was shown by Osborne et al.; the reaction is not a catalytic one and thus requires stoichiometric amounts of the MTase to go to completion. This evidence of this was found by using a protein MTase enzyme 1 that used a nitrogen mustard precursor to transalkylate a peptide. The product of this was shown to be an inhibitor of the enzyme which made the reaction self-limiting.[135]

Furthermore, aziridine analogues are extremely reactive, due to the charged nitrogen as well as the strained three-membered ring. This can make a product very susceptible to fast degradation which leads to an increase in non-specific alkylation without the presence of an enzyme.[135]

The second analogue of the native SAM is the doubly activated SAM analogues. There are two methods to make these, the first is through a total synthesis pathway, whilst the second

is by using an enzymatic method which contains a methionine adenosyl transferase and a synthetic methionine derivative.[135]

Since MTases are biological catalysts, these enzymes can catalyse the transfer of alkyl groups that are bigger than methyl groups. However, as the transfer group size increases (going from methyl to ethyl and further) the rate of transfer of these groups decreases. [136] This can be rationalised by considering an S_N2 mechanism, which necessitates inversion of the alpha carbon adjacent to the sulphur group. As the group that is inverted becomes bulkier, the reaction rate falls because of destabilisation of the transition state intermediate. Dalhoff et al. hypothesised and later demonstrated that adding a double bond to the beta position, relative to the sulphonium, stabilises the transition state intermediate. The π-orbital of this adjacent unsaturated bond actively stabilises the p-orbitals that are inverted on the alpha carbon. A collaboration between the Weinhold and Klimašauskas group were able to prove that this is the case. The addition of a double or triple carbon-carbon bond can stabilise the transition state and ultimately recover the reaction rate. Labelling using these doubly activated analogues has given rise to the term mTAG.[137]



Scheme 6: Schematic of inversion at the alpha carbon (top), schematic showing the p-orbital overlap being used to stabilise the transition state of the S_N2 (bottom) reaction. [137]

These doubly activated analogues offer a significant advantage over aziridine and nitrogen mustard analogues, that is mTAG labelling is a catalytic process so stoichiometric amounts of enzyme are not needed. However, the analogue is a magnitude slower than nature's method of methyl transfer. Despite this, however, this method allyl, alkenyl and alkynyl SAM analogues have been produced successfully and have been used in follow up mTAG experiments. This simple technique offers the ability to tag DNA with a wide range of

complex molecules. This can be achieved with a functional group that can either react or be altered that will produce a more complex molecule. An example is the ability to substitute different amine groups which can react with N-hydroxysuccinimide ester, this will produce an amide bond with a new group attached. Research in this field has been focused on making these analogues contain terminal azides or alkynes.[137] The advantage is being able to react in an azide-alkyne cycloaddition (click reaction). However, the alkyne terminated analogue proved to be very unstable. This was observed by Weinhold and Luo et al. [137] The issue in stability was circumvented by using an analogue known as AdoenyYn which contained a terminal alkyne in conjugation with a double-bond. This enabled the analogue to be used in applications like bio-orthogonal ligations. Ketones are another possibility of transferrable groups which could be useful for reacting hydrazides or hydroxylamine derivatives. This ketone analogue could be transferred to the DNA and coupled with a hydroxylamine species; the product could then be used to couple DNA to a fluorophore.[137] Developing on this idea, an analogue of SAM was generated which had a dye molecule attached which led to a single step labelling reaction, this was shown by Grunwald et al. where a TAMRA dye was used to couple to the SAM analogue.[137]

Other double-activated analogues of this cofactor have been synthesised that don't have a triple bond attached. The MTases Ecm1 and M.TaqI, were shown to be compatible with an anologue that contains the norborane fragment with a p-xylylene linker. The fragment was covalently bound to the sulphur atom. This resulted in the norborane fragment being transferred to the DNA.[137] Bioconjugation can be followed up, with each MTase in absence of Cu (I) salts. This is advantageous for a thiol-rich environment as azides have very low stability in this environment. Further study around this led to the development of SAM analogues which contain very common photocrosslinking fragments. These include: diazirin, benzophenon & arylazide.[137] The crosslinkers could be transferred using MTase and then

reacted via photoactivation with UV radiation, labelled DNA could then be crosslinked with a photocage with the binding protein.[137]

In 2020 the Nealy group made a further development. They made a cofactor analogue that has a reversible linker on it- a hydrazone group. This is a group that contains two nitrogens' bonded together via a single bond, one nitrogen bound to a carbonyl whilst the other is bound to an alkyl group via a double bond. This allows for further modification of the substrates of the MTases. These can be subsequently removed but can also be immediately followed by another modification on the same base with new functionalised groups that could then remain permanently on the substrate or be rewritten again should it be required. To test this, the Nealy group used the M.MpeI DNA MTase as a test. They were able to develop methodology that isolated fluorescent DNA; this DNA was then followed up with rewriting the label with biotin. This allowed the DNA of interest to be captured via streptavidin beads. This rewriteable component arises when the label is treated with H2NOH.HCl at 50 ℃ for one hour which will release the alkyl group from the hydrazine. The hydrazine is then free to react again with another carbonyl to form another hydrazone. This can then either be reacted with hydroxylamine again to release the hydrazine to rewrite the label or it can be left labelled permanently.[137]

1.51 SAM uses in Therapeutics

SAM is the primary source of methyl groups for DNA during methylation. This plays a key role in epigenetic changes. SAM has also been studied for its potential anti-cancer effects, including inhibiting cell growth, preventing metastasis, and promoting cell death in several types of cancer.[137] In head and neck squamous cell carcinoma (HNSCC), the sixth most

common cancer globally, SAM has been shown to reduce the ability of two HNSCC cell lines, Cal-33 and JHU-SCC-011, to migrate and invade. SAM also triggered apoptosis (cell death) in these cells. [137]

SAM was found to slow cell cycle progression, reduce levels of certain cyclins, and lower the expression of the p21 cell cycle inhibitor. It inhibited cell migration in a dose- and time-dependent manner, with effects seen at 24 hours for Cal-33 cells and 48 hours for JHU-SCC-011 cells. SAM also significantly decreased the invasive ability of these cells. Western blot analysis showed that SAM influenced several key signalling pathways involved in migration and invasion. Additionally, combining SAM with cisplatin showed a synergistic effect, enhancing its ability to inhibit HNSCC cell migration. These findings suggest that SAM could be a promising candidate for drug development, particularly for targeting metastatic cancers.[137]

Much research has gone into SAM, and it has been shown that SAM is involved in regulating the genes that are responsible for cell invasion as well as metastasis.

SAM plays a key role in regulating genes that control tumor invasion and metastasis, such as urokinase-type plasminogen activator (uPA) and matrix metalloproteinases (MMPs).

Research has shown that SAM can reduce the expression of these genes in cancer cells, including breast (MDA-231) and prostate (PC-3) cancer cells, which decreases their ability to

invade and spread. SAM also works with other treatments like DAC to block tumor cell invasion by targeting uPA and MMPs. [137]

In colorectal cancer cells (SW-620), SAM reduced MMP2 levels and increased the tissue inhibitor of MMP2. In osteosarcoma cells (LM-7 and MG-63), SAM decreased cell growth and invasion by inhibiting uPA and MMPs, which are key to metastasis.[137] SAM also enhanced the anti-cancer effects of gemcitabine in pancreatic cancer by blocking the JAK2/STAT3 pathway. In cervical cancer cells (HeLa), SAM worked with selenium compounds to affect key signalling pathways (ERK and AKT). This reduces cell growth, migration, and attachment. [137]

SAM has shown promising anti-proliferative and anti-metastatic effects in various cancer cells. Although, its impact on all cancer types is still being studied. In head and neck squamous cell carcinoma (HNSCC). Regulating this cell cycle is crucial for tumor growth. In cancer cells, key cell cycle regulators, like p53, are often mutated or inactive, leading to uncontrolled cell division. In HNSCC, SAM was found to stop the cell cycle at the G2/M phase by reducing cyclin levels, whilst leaving p53 levels unaffected. [137]

SAM also helped improve wound healing when combined with the chemotherapy drug cisplatin (cDDP). This combination of SAM and cDDP created a strong pro-apoptotic effect on cancer cells. Less of the drug is then required, thus reducing side effects. cDDP works by damaging DNA, leading to cell death. This process activates stress pathways, like MAPK, which can influence both apoptosis and drug resistance. [137]

These findings suggest that SAM could be a valuable anticancer treatment, as it inhibits cell growth, migration, and invasion. SAM also works well when combined with other drugs, making it a promising option for treating metastatic cancers. Additionally, SAM is available as

a dietary supplement, and when used in the right doses, it has few side effects and doesn't harm non-cancerous cells.

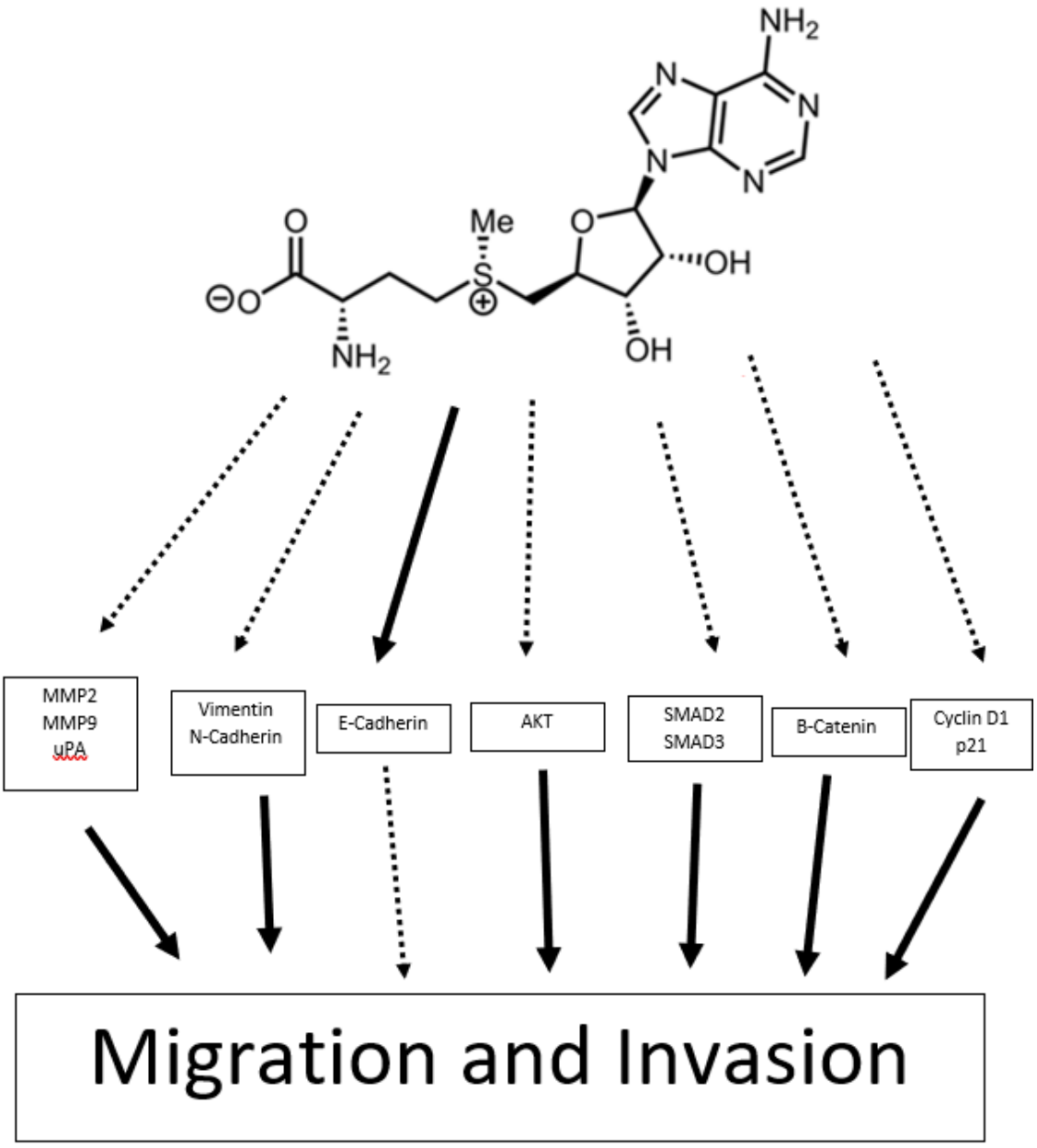


Figure 11: SAM and its effects on the levels of species that are important to cancerous migration and invasion, hatched arrows leads to a decrease in concentration or effect whilst bold arrows mean an increase in concentration or effect e.g., SAM increases the levels of E-cadherin which leads to a decrease in migration/invasion of tumour cells. [137]

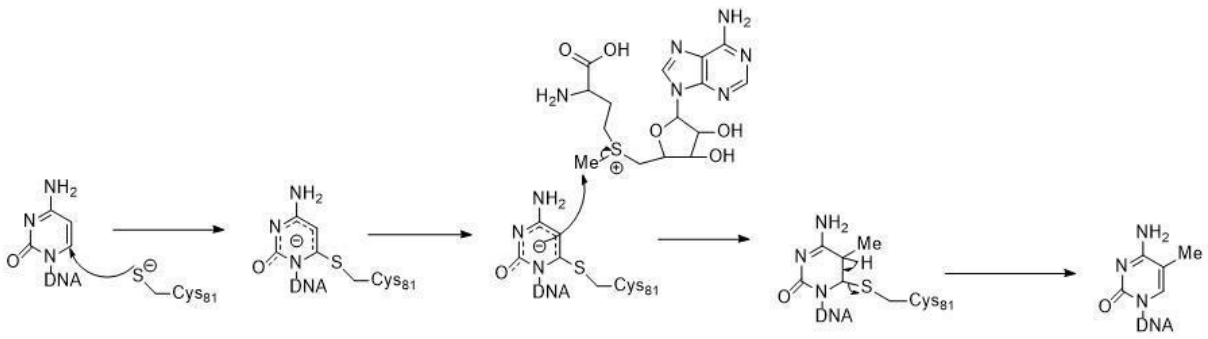
1.6 Thesis Aims

This thesis aims to demonstrate the development of novel cofactor analogues based on the SAM cofactor. It also explores how these new analogues were tested with methyltransferase enzymes, particularly the M.MpeI enzyme. Additionally, the thesis investigates the use of docking studies to examine the enzyme's active site and explain the observed differences in activity between structurally similar analogues. This analysis provides insights into the design of new analogues that, despite being tested, show minimal improvement with the enzyme. Further investigation of the active site reveals how a single amino acid change in the M.MpeI enzyme alters its interaction with cofactors. This has led to the creation of a new cofactor analogue and a mutant methyltransferase combination that enhanced enzyme selectivity.

Chapter 2: Synthetic Cofactor Analogues

2.1 Introduction

Enzymes serve as catalysts for cellular processes. With many having evolved to perform their tasks quickly and efficiently;[14] for example, M.MpeI catalyzes DNA methylation within its active site. This site is occupied by both the substrate (cytosine) and a cofactor (or cofactor analogue). Once methyl transfer is complete, cytosine and the cofactor leave the active site, allowing the enzyme to participate in another cycle of methylation.[24] S-adenosylmethionine (SAM) is the second most ubiquitous cofactor in the body and plays a role in several biological functions. It is involved in the immune system, where methyltransferases (MTases) use the methyl group on SAM to methylate DNA or RNA. This promotes T cell proliferation and differentiation. SAM is also crucial in the production and regulation of brain chemicals like serotonin.[25] This thesis will focus specifically on SAM's role as a methyl donor in DNA methylation.[19] The methyltransferase process converts SAM into S-adenosyl-L-homocysteine (SAH), which dissociates from the enzyme, regenerating the vacant active site. This enables the enzyme to catalyse another cycle of DNA methylation until all required DNA is methylated or the SAM cofactor is depleted. A schematic illustrating the DNA methylation process is shown below:



Scheme 7: Methylation mechanism for DNA with SAM. This starts with attack of the cytosine base from the enzyme (S—Cys81) to generate the intermediate which then can attack the methyl group of the SAM cofactor to form the

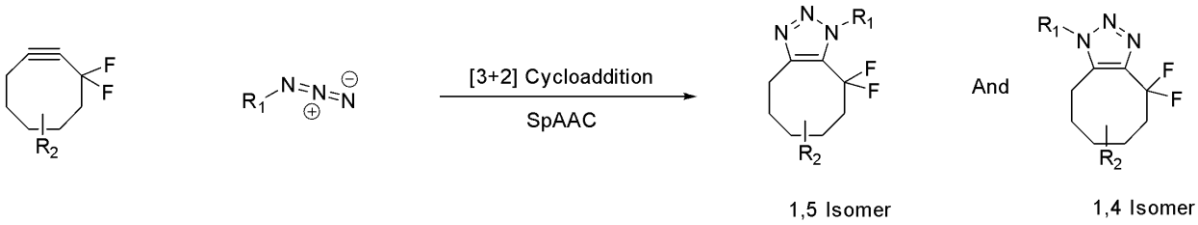
second intermediate which undergoes a β-hydride elimination to remove the now methylated cytosine from the enzyme and regenerate the double bond. [38]

In scheme 7, the sulphur atom from Cys81 of the methyltransferase (M.HhaI) is deprotonated, enabling it to attack the cytosine base as a nucleophile. This forms a bond between the cytosine base and the active site of the methyltransferase. The resulting intermediate makes the base negatively charged. This charge can be stabilized through π-resonance. The cytosine then reacts with the methyl group of the SAM cofactor, which is attached to the sulfonium centre. This generates a second intermediate. This intermediate undergoes β-elimination, where the hydrogen is removed by a Lewis base, reforming the double bond and expelling Cys81, thus releasing the methylated base from the enzyme's active site.

Methylation is a mechanism the body uses to switch a gene from an 'active' to a 'repressed' state. However, this process can sometimes be detrimental when important genes, crucial for preventing disease, become methylated (and thus repressed). If we can understand how this occurs and learn how to control it, it could become a valuable tool for preventing the methylation of key genes, thus ensuring they remain active and functional.

Given the critical role of DNA methylation in regulating gene expression, considerable effort has been dedicated to modifying the SAM cofactor to influence or control the behaviour of MTases. For instance, altering the transferable moiety can impact the types of functional groups attached to the DNA. Modifying SAM itself is also possible and could enhance binding or even inhibit the MTase. This has been shown to be feasible. In past research, the Neely group has altered the moiety transferred to the DNA. This has enabled DNA to be captured on beads through DNA transalkylation, where an extended alkyne group, terminated with an azide group, is transferred using bacterial MTases.[20] Subsequently,

strain-promoted azide-alkyne cycloaddition (SPAAC) allows the attachment of biotin to the DNA, providing a method to selectively capture labelled DNA via the biotin-streptavidin interaction. DNA transalkylation with an azide-terminated group enables the modification of DNA with various functional groups, such as attaching a fluorophore, which could allow for tracking the DNA's movement within the cell. .[21]



Scheme 8: Schematic showing reaction between a fluorocyclooctyne species and azide species in a SpAAC context. Ordinarily, this type of reaction requires the aid of a Cu species however as Cu is known to damage DNA this needs to be avoided, and so the SpAAC reaction has the advantage in that it requires no other species to act as a catalyst. [22]

Another commonly reported modification of SAM occurs around the adenosine ring, specifically at the N6 or C8 positions.[22] C8 analogues, for example, have been used in the past as a method to covalently bind the cofactor to DNA.[138] The C8 carbon can also be modified to influence the binding affinity of SAM to the enzyme, as this alteration can impact the interactions between the enzyme's amino acids and the SAM itself. [25]

This Chapter details the development of new analogues based on the SAM cofactor. The Neely group originally synthesized two versions of a SAM analogue. The first is ((S)-3-amino-3-carboxypropyl)((5-(6-((2-aminoethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(8-(2-((E)-4-((2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)carbamoyl)benzylidene)hydrazineyl)-8-oxooct-2-yn-1-yl)sulfonium (**1**), and the second is ((S)-3-amino-3-carboxypropyl)(8-(2-((E)-4-((2-(2-(2-(2-

azidoethoxy)ethoxy)ethoxy)ethyl)carbamoyl)benzylidene)hydrazineyl)-8-oxooct-2-yn-1-yl)((3,4-dihydroxy-5-(6-((2-hydroxyethyl)amino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)sulfonium (**2**).

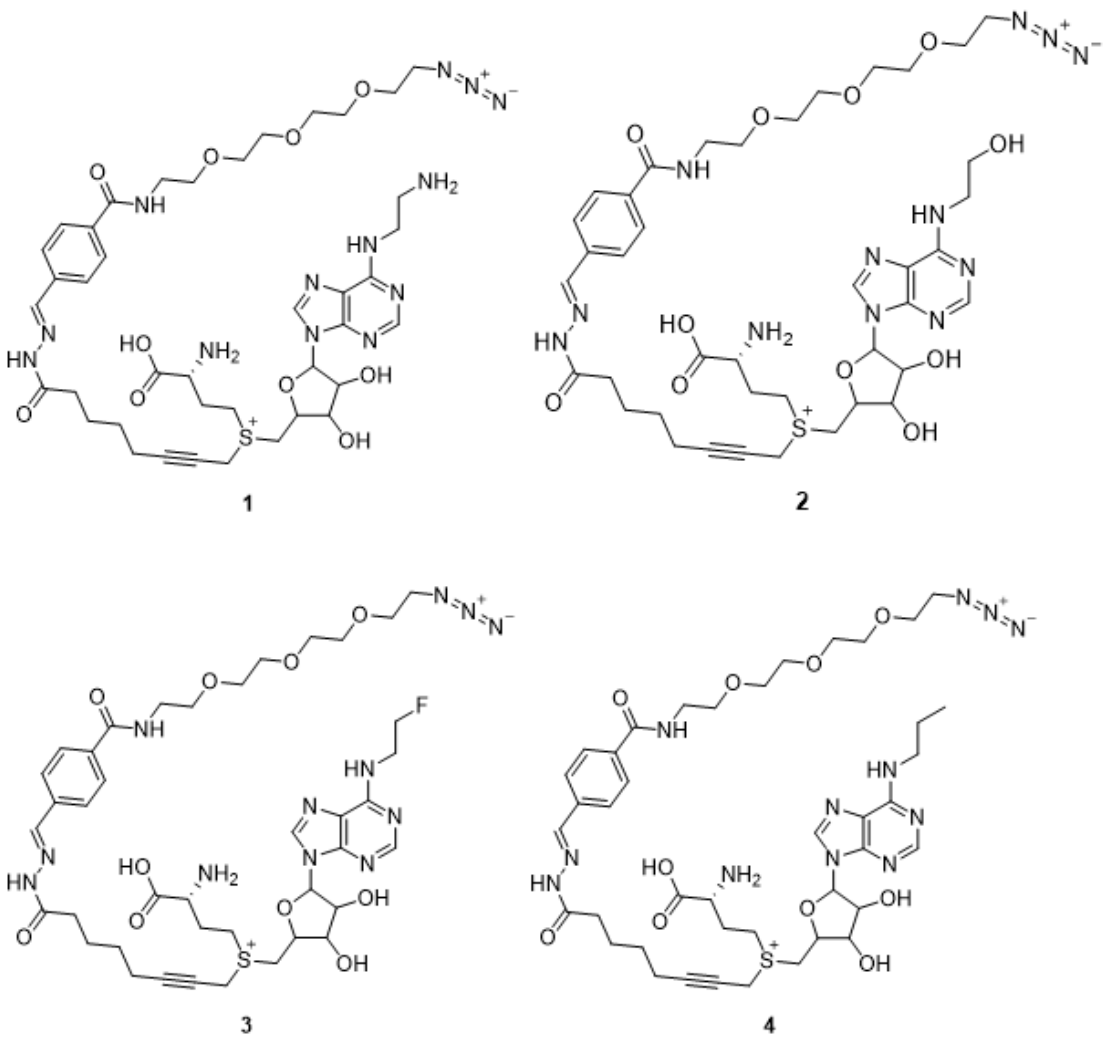


Figure 12: Structures of all cofactor analogues used in this chapter

2.2 Chapter Aims

This Chapter investigates whether or not analogues of the two cofactors (**1** and **2**) will affect the activity of the MTase when acting on DNA. The two analogues designed and synthesized

are ((S)-3-amino-3-carboxypropyl)(8-(2-((E)-4-((2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)carbamoyl)benzylidene)hydrazineyl)-8-oxooct-2-yn-1-yl)((5-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)sulfonium (**3**) and ((S)-3-amino-3-carboxypropyl)(8-(2-((E)-4-((2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)carbamoyl)benzylidene)hydrazineyl)-8-oxooct-2-yn-1-yl)((3,4-dihydroxy-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)sulfonium (**4**).

Throughout this thesis, a modified version of M.MpeI has been used. This follows the work of Lukinavičius et al.[139], in modifying M.HhaI, M.HpaII, and M2.Eco31. Previous work from the Neely group utilized the Q5 Site-Directed Mutagenesis Kit (NEB) to introduce N374A and Q136A mutations into the M.MpeI protein sequence. This improved its efficiency with synthetic cofactors. This modified version is referred to as wild-type M.MpeI. More details on this protein can be found in section 6.41 of the Materials and Methods section.

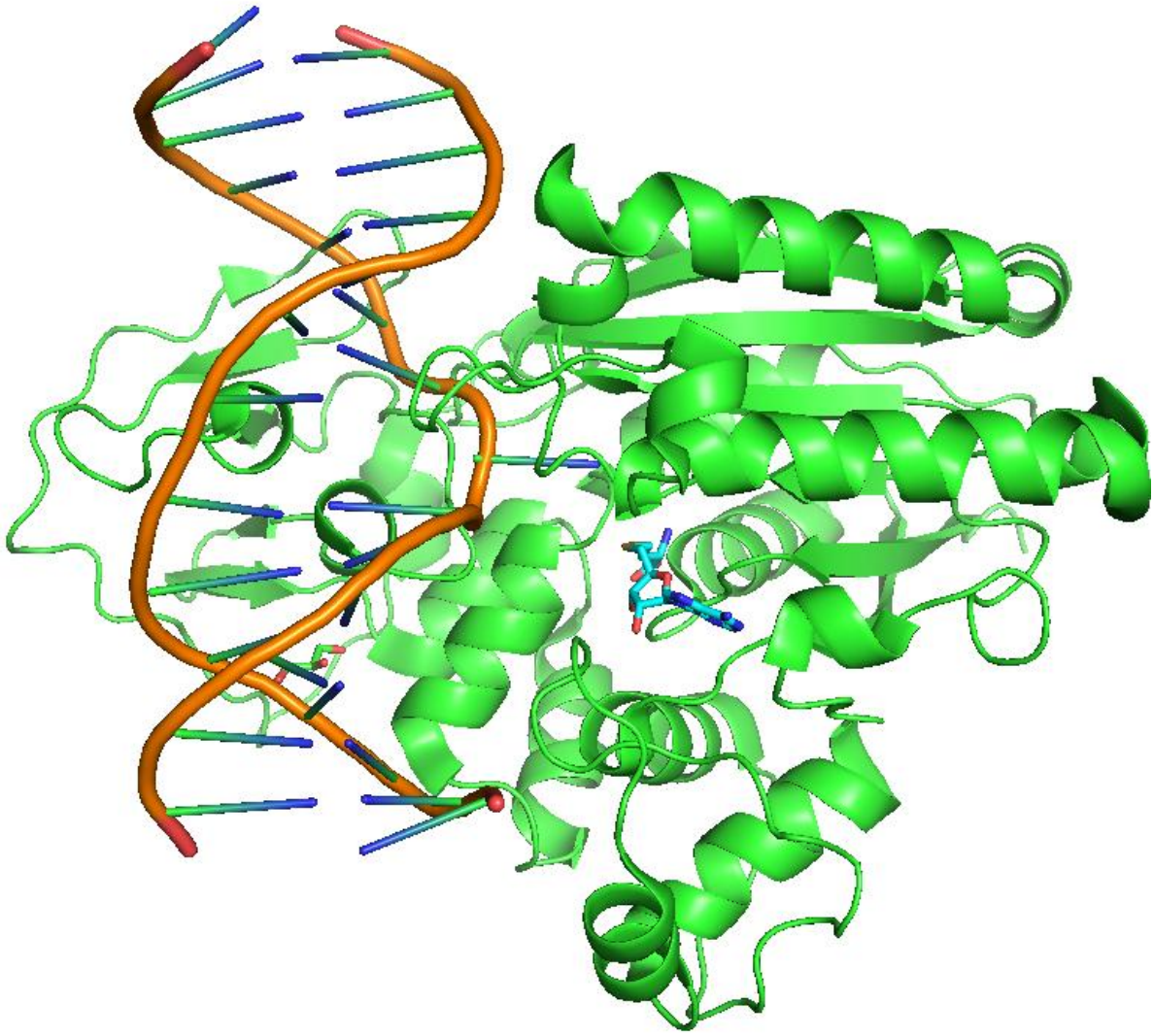


Figure 13: Illustration of crystal structure of M.MpeI (green) with pymol rendering of DNA (orange) using ribbon imagining & SAM (blue) bound PDB code 4DKJ. [139]

2.3 Results

The cofactor analogues 1 and 2 were tested in a protection assay, a technique used throughout this thesis, involving an enzyme, DNA, and cofactor. In this assay, when the enzyme and cofactor react together, the enzyme transfers a moiety from the cofactor to the DNA, thereby protecting the DNA from digestion by later introducing a restriction enzyme. If this enzyme or cofactor is incompatible, the moiety is then not transferred, and the restriction

enzyme will digest the DNA into smaller fragments. Both the digested and undigested DNA strands can be visualized on a gel. Bands that appear further down the gel indicate smaller, digested DNA fragments, whilst bands higher up represent undigested DNA. The standard method for testing DNA MTase activity is an in vitro DNA methylation reaction, followed by digestion with a restriction enzyme, whose activity is then blocked by the MTase of interest. This can be seen in the protection assay. DNA digestion or protection is assessed by measuring DNA fragment lengths using gel electrophoresis. Figure 14 shows a simplified DNA ladder, along with positive and negative controls.

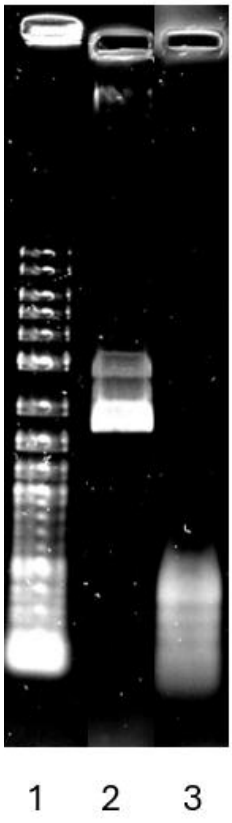
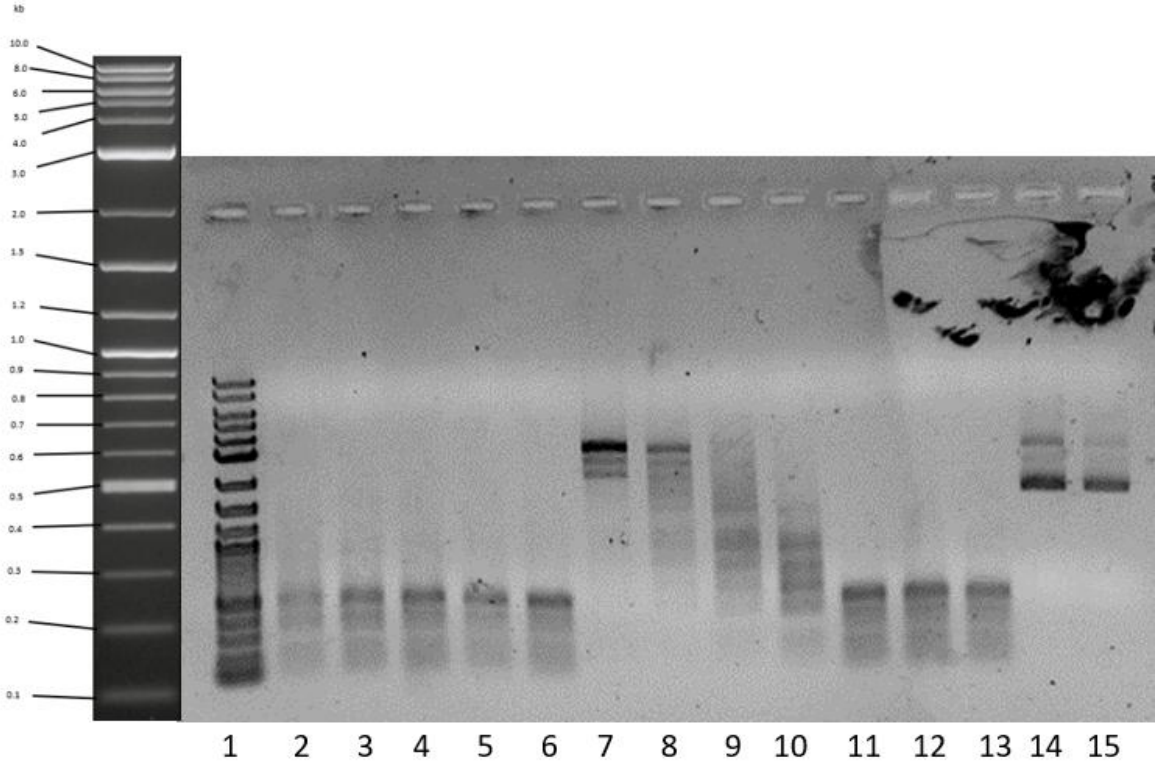


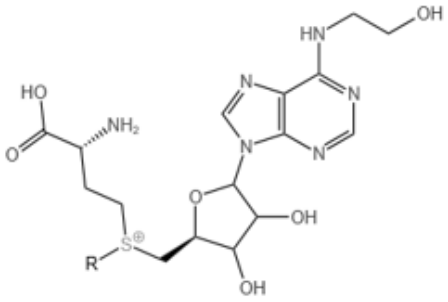
Figure 14 Simplified agarose gel which highlights how the DNA ladder appears (lane 1) in relation to a positive control "protected" (lane 2) and a negative control "unprotected" (lane 3).

Both cofactors were tested with two different MTases. M.TaqI, whose previous work was conducted by Dr. Elodie Jagu, showed that both cofactor analogues enabled the MTase to be active on DNA.[140] The second MTase tested was M.MpeI, the focus of this thesis. When I performed a protection assay using M.MpeI, I observed a significantly different activity profile for the two cofactors. The **2** cofactor analogue enabled the MTase to protect DNA to a much higher degree compared to the ((S)-3-amino-3-carboxypropyl)((5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(8-(2-((E)-4-((2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)carbamoyl)benzylidene)hydrazineyl)-8-oxooct-2-yn-1-yl)sulfonium (**5**) analogue. In contrast, the **1** cofactor analogue showed the opposite effect when added to the same MTase, resulting in very little activity on DNA and minimal protection. This unexpected result was an interesting discovery that warranted further investigation. It suggests that specificity for a given MTase enzyme can be introduced by modifying the SAM scaffold. This finding could open several avenues for improving current diagnostics and therapeutics by targeting the activity of specific MTases in our cells. This investigation was pursued further to understand why M.MpeI behaves so differently in the presence of two structurally similar cofactor analogues.

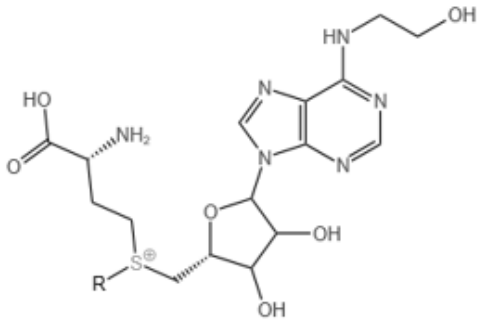
A protection assay was performed, this is the technique used throughout this thesis to test the activity of the MTase on DNA with the cofactor analogue in question. In this assay, the MTase is exposed to the DNA and cofactor analogue, which will then methylate (protect) the DNA if the cofactor is compatible with the MTase. After methylation, the DNA is treated with a restriction enzyme that cuts at sites that have not been methylated. The DNA is then loaded onto a gel and subjected to an electrical current, causing it to travel through the gel. Cut DNA fragments, being smaller, move faster through the gel compared to protected DNA strands. The resulting DNA bands are visualized, allowing observation of protected and

unprotected DNA and providing insight into the compatibility between the cofactor and enzyme.





2 cofactor isomer I

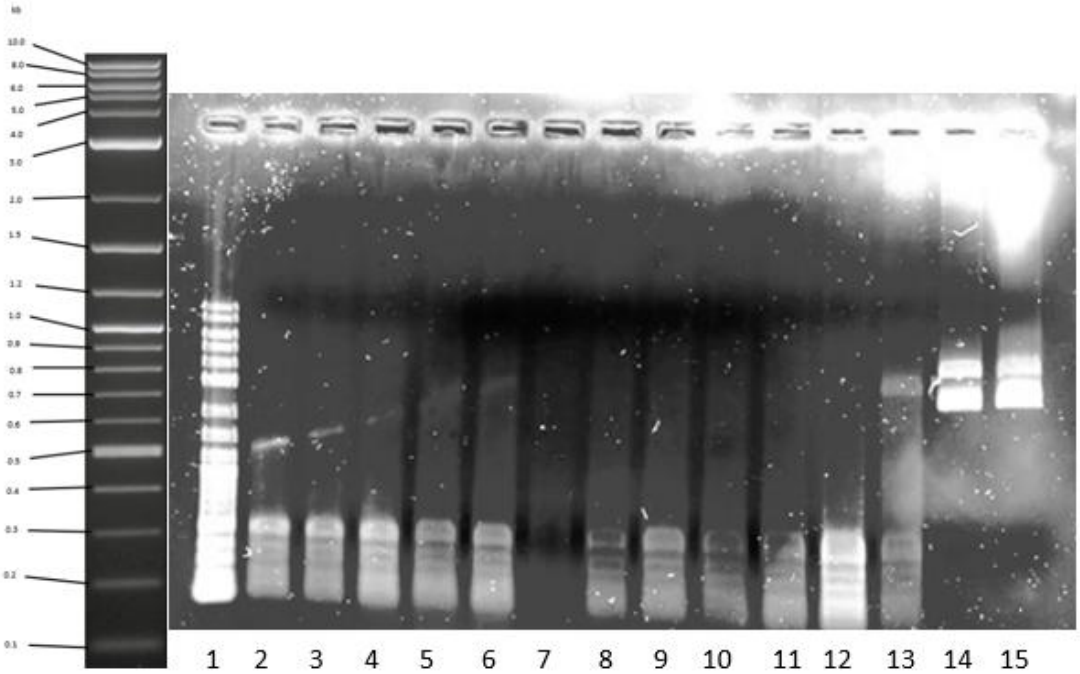


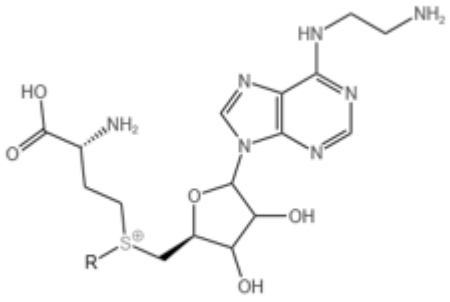
2 cofactor isomer II

*Figure 15: (Above) Gel image of the **2** cofactor in protection assay with M.MpeI, lanes from left to right;*
1: DNA ladder
*2-5: **2** cofactor isomer I serial dilution 500 µM – 62.5 µM*
*6: **2** cofactor isomer I control*
*7-10 **2** cofactor isomer II serial dilution 500 µM – 62.5 µM*
*11: **2** cofactor isomer II cofactor control*
12: Control (no cofactor)
13: Control (no cofactor, no enzyme)
14: SAM control
15: SAM control (no restriction)
EC_{50} value is 125.0614 µM
(below) structures of the different isomers

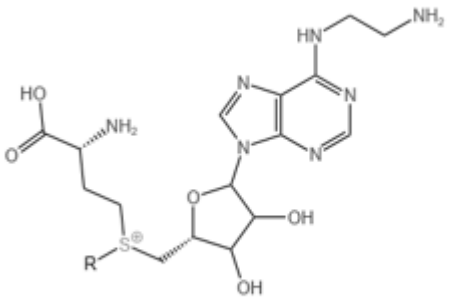
Figure 15 shows the results of the gel from the protection assay using the **2** cofactor. Isomer I of the **2** cofactor (Lanes 2-5) was found to be incompatible with the M.MpeI enzyme, as it showed no protection of the DNA. In contrast, isomer II (Lanes 7-10) demonstrated DNA protection, with the level of protection decreasing as the cofactor was diluted, as seen in Lane 10. The transition from 90% to 60% DNA protection is evident, with Lane 6 showing near full protection (90%) and Lane 10 showing reduced protection (60%). This was determined by comparing the position of each band in the gel to the DNA ladder, along with the positive and negative controls. If the band migrated the same distance as the positive control, this indicates the DNA had been protected and resisted digestion. If it migrated the same distance as the negative control, then this indicates that the DNA was unprotected and digested.

The next test aimed to determine whether cofactor **1** exhibited a similar activity level to cofactor **2** with the M.MpeI enzyme. The controls for both cofactors had the same composition, but in each control, no enzyme was present (Lanes 6, 11, 12, and 13). These controls confirm that the observed protection is due to the interaction between the MTase and the cofactor analogue. The SAM control, where no restriction enzyme is applied, is included to demonstrate that the MTase remains active in the presence of SAM. If this were not the case, the SAM control in Lane 14 would show no protection.





1 cofactor isomer I



1 cofactor isomer II

*Figure 16: (Above) Gel image of **1** cofactor in protection assay with M.MpeI, lanes from left to right;*
1: DNA ladder
*2-5: **1** cofactor isomer I serial dilution 500 µM – 62.5 µM*
*6: **1** cofactor isomer I control*
*7-10 **1** cofactor isomer II serial dilution 500 µM – 62.5 µM*
*11: **1** cofactor isomer II cofactor control*
12: Control (no cofactor)
13: Control (no cofactor, no enzyme)
14: SAM control
15: SAM control (no restriction)
(Below) structures of the different isomers

This protection assay was conducted using the M.MpeI enzyme alongside cofactor **1**. The results were unexpected, as it was initially assumed that the switch from alcohol to amine would not cause such a significant change in activity. Lanes 7-10 were expected to show a similar pattern to those seen in Figure 15 (Lanes 7-10) since both cofactor analogues are similar in size and electronegativity. Therefore, it was anticipated that both the **1** and **2** analogues would behave similarly under the same conditions. The next step was to test the **2** analogue against cofactor **5**.

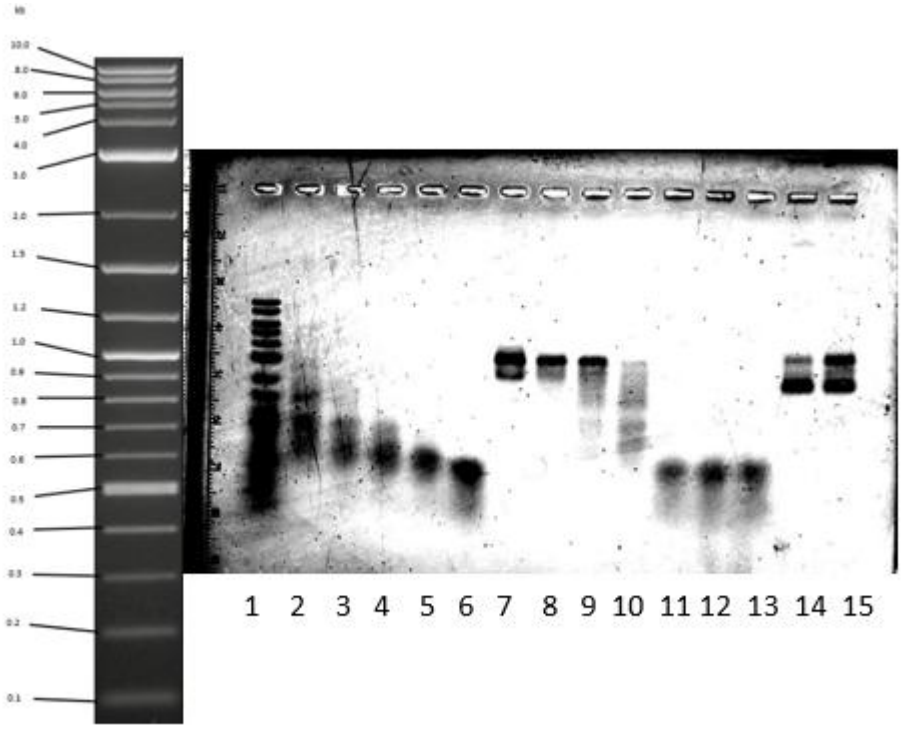


*Figure 17: Gel image of the **5** cofactor with **2** analogue with M.MpeI, enzyme lanes from left to right;*
1: DNA Ladder
*2-5: **5** serial dilution 500 µM – 62.5 µM*
*6: **5** control*
*7-10: **2** cofactor serial dilution 500 µM – 62.5 µM*
*11: **2** cofactor control*
12: Control (no cofactor)
13: Control (no cofactor, no enzyme)
14: SAM control
15: SAM control (no restriction)

This result compares the activities of the **2** cofactor and the **5** cofactor. The findings are interesting, as the **2** cofactor provides better protection of DNA than the **5** cofactor. This can be seen when comparing Lane 2 and Lane 10, where both show similar levels of DNA protection, but the **5** cofactor is at a concentration of 500 µM (Lane 2), while the **2** cofactor is at a concentration of 62.5 µM (Lane 10). When the concentration of the **2** cofactor is increased, the DNA protection improves significantly, as shown in Lane 7 at 500 µM, which demonstrates 90-100% DNA protection. In contrast, the **5** cofactor at 500 µM (Lane 2) provides only about 60% DNA protection.

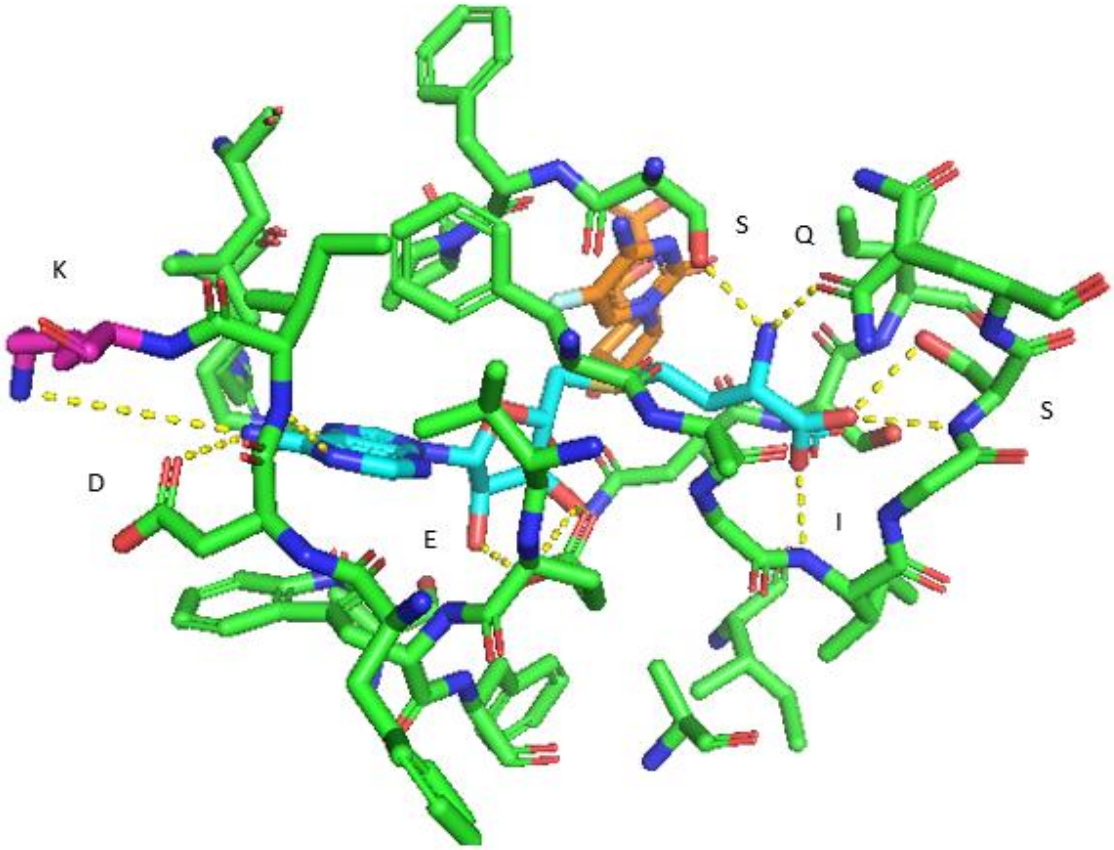


Figure 18: M.MpeI (green) binding site with cytosine (orange) as well as SAM (blue) bound, showing key interactions (yellow) between SAM and other residues, notably the 115-lysine residue (purple) PDB code 4DKJ.

The two cofactor analogues tested are structurally similar, differing by only one group: OH, in the **2** analogue and NH_2 in the **1** analogue. Any differences in MTase activity can be attributed to these groups. Since OH and NH_2 have similar mass, it was expected that their radii would also be similar. Additionally, the electronegativity values of oxygen (3.44) and nitrogen (3.04) are close on the Pauling scale. Both cofactors are likely to engage in a binding interaction between the heteroatom (O or N) and a nearby lysine residue. More specifically, this interaction would involve the lone pair of electrons on either O or N, interacting with a hydrogen bound to a nearby lysine residue.

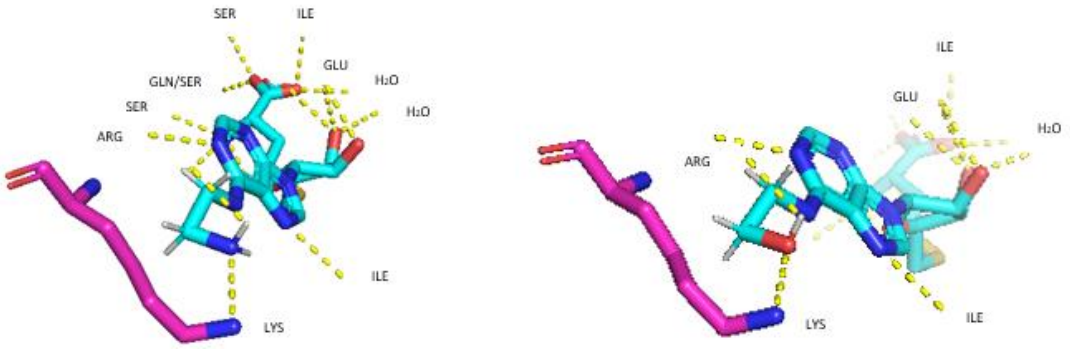


*Figure 19: Interaction between the terminal amino/hydroxy group (left/right(blue)) for **1** and **2** and glutamine K115 (purple) in active site of M.MpeI PDB code 4DKJ.*

As shown in Figure 19, hydrogen bonds were formed between the O or N of the cofactor analogues and the 115-lysine residue. Both analogues exhibited this interaction, as oxygen and nitrogen each have at least one lone pair (with oxygen having two). However, when tested with the M.MpeI enzyme, the **2** cofactor showed methylation of DNA, while the **1** cofactor showed no methylation. This result was unexpected, as both cofactors were anticipated to produce similar outcomes. Further investigation revealed a major difference in how the experiment was performed. The conditions were meant to reflect the body's homeostasis, with pH being the most important factor; maintained around pH 7. At this pH, amine groups become protonated due to the lone pair on the nitrogen atom, which results in a positive charge on the amino group. Meanwhile, any acidic groups are deprotonated, generating a negative charge on the acidic group.

It was hypothesized that pH could be the cause of the observed difference in activity. To address this, new analogues were developed that would not be affected by pH. [24] Specifically, analogues of the OH group were investigated, as this analogue was the one that successfully protected DNA from restriction by M.MpeI.

2.31 Synthetic Results

The two analogues behaved very differently with the same enzyme. This showed that it would be beneficial to test other analogues. However, with many options available, such as methyl, fluoro, chloro, and thiol, and the assay being low throughput, there was a need for a rational method to narrow down the possibilities. Computational docking studies were used to help prioritize which analogues to test experimentally. These studies were advantageous because they allowed for the identification of analogues that could bind similarly to the original cofactor, without the need for extensive synthesis and activity screening.

A docking simulation was performed for each cofactor analogue in the active site of the M.MpeI enzyme. Details of the docking process can be found in the appendix (Docking Method). The results generated a data file that could be visualized with PyMOL, along with a text file containing the energy values for each docked state. These two data sets were then used to compare the analogues to the orientation of the native cofactor, SAH (SAM without the transferrable methyl group), from the M.MpeI-DNA-SAH co-crystal structure. The docked states of the cofactor analogues were analyzed to determine which could orient in a similar manner to SAM, thus ensuring that transfer was possible. The energy value of each state was then compared to that of the SAM cofactor to assess the similarity. A high energy value would indicate improper orientation, while a low value would suggest that the analogue binds too tightly, potentially preventing it from leaving the active site after the reaction, thus inhibiting enzyme turnover.

The docking studies revealed that the **4** and **3** cofactors could orientate in the M.MpeI cofactor binding pocket in a similar position and orientation as the 5 cofactor. The energy for each cofactor in these positions was calculated and compared to that of the 5 cofactor. The docking studies showed that the proposed analogues had energy values similar to the **5**

cofactor binding to the M.MpeI protein (-9.2 kcal/mol). Table 1 shows the energy values of the cofactors when oriented in the position that allows for transfer.

Table 1: Table that includes the docked energy values that were calculated by docking studies when the wild type M.MpeI enzyme was simulated with each cofactor

Enzyme	Cofactor	Docked result (kcal / mol)
Wild_Type	**5**	-9.2
Wild_Type	**1**	-9.0
Wild_Type	**2**	-9.6
Wild_Type	**3**	-9.3
Wild_Type	**4**	-8.9

However, this process has some limitations that could lead to false positives. In all the docking studies, water was removed from the protein, but in practice, the protein would contain water. This could alter the energy values of the docked states. Future work could involve docking the cofactors with the water included in the protein to observe how this affects the energy values. It is possible that the presence of additional hydrogen bond acceptors and donors may provide stabilizing interactions for some cofactor analogues. Additionally, an exhaustiveness value of 8 was used in all studies, which is relatively low. This was chosen as a starting point but could be increased to provide a broader range of energy values by increasing the number of docked states. Although this method allows for rapid screening of many analogues and is less computationally demanding than other binding techniques, it could be complemented by molecular dynamics simulations. While more computationally intensive and time-consuming, molecular dynamics would offer more accurate binding values and could also be used to assess the stability of cofactors in the active site. Free energy perturbation could also be employed to precisely predict the binding potency between the protein and ligand.

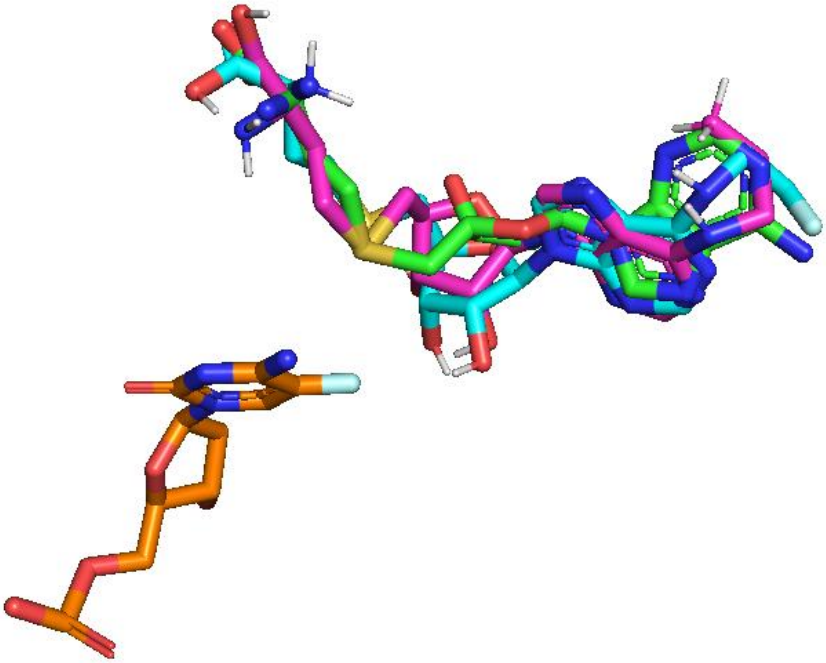
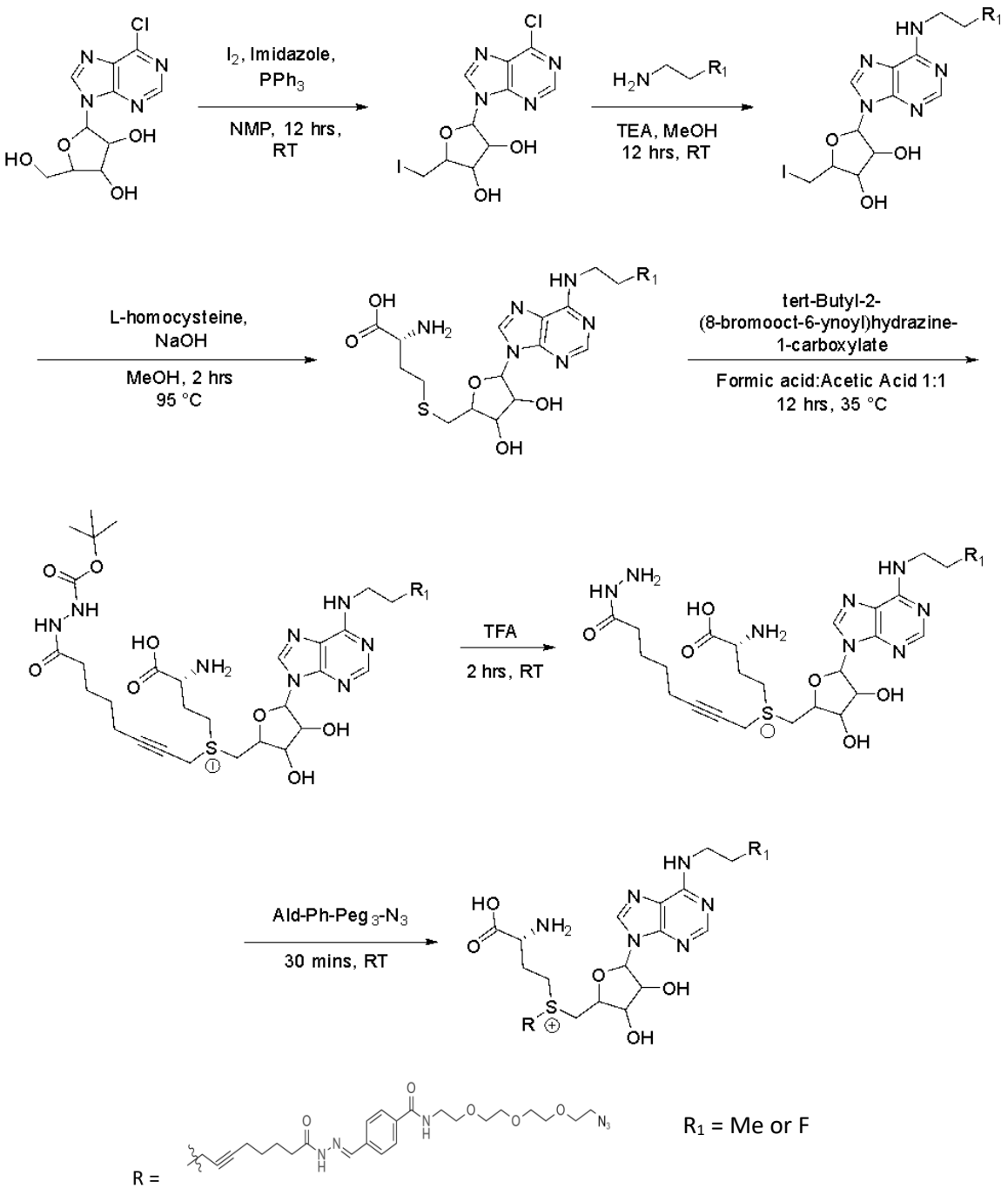


*Figure 20: Overlay of docked cofactors in the M.MpeI active site in proximity to cytosine (orange), SAM (shown in green), **4**(shown in purple) **3** (shown in cyan) PDB code 4DKJ.*

:

2.4 Overview of the Synthetic Approach

The two molecules selected for synthesis and further experimental investigation were the N6-substituted methyl and fluoro SAM analogues. This selection was made in the hope of gaining deeper insight into the significance of hydrogen-bonding, dipole-dipole interactions, and steric effects between the enzyme and cofactor, particularly in relation to the catalytic activity of M.MpeI. The analogues were synthesized following an established procedure, as outlined briefly below. [141]



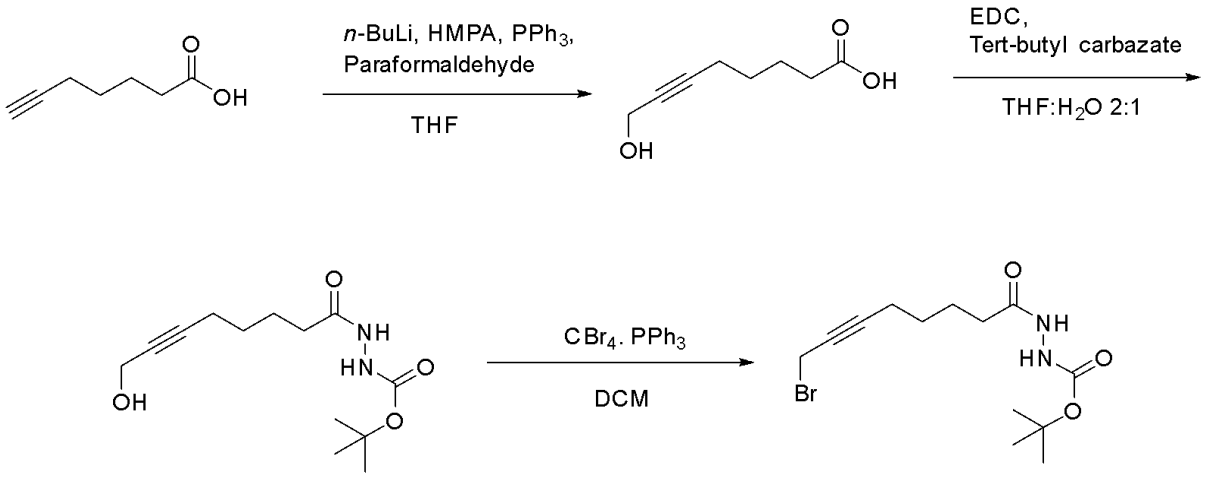
Scheme 27: Synthetic scheme of cofactor synthesis along with reagents. Conditions and other reagents are found in the methods section.

Starting with 6-chloropurine riboside, an Appel reaction was employed to convert the 5'-primary alcohol into a halide. This was followed by the substitution of the C6-chlorine with an amine alkyl chain, which terminated in either a methyl or fluoro group. The propyl/fluoro-I-adenosine was then converted into the SAH analogue by adding L-Homocysteine in the presence of a base. The base used was NaOH.

The 'linker' for the cofactor analogue is covalently transferred to DNA by the MTase enzyme. This is attached to the sulphur of the SAH analogue S-((3,4-dihydroxy-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)-D-homocysteine / S-((5-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-D-homocysteine, forming the complete cofactor analogue. During the methylation process, this group is transferred to the DNA. This linker was chosen due to its ability to reversibly label biomolecules, allowing its chemical functionality to be easily removed.[142] The linker was synthesized by using 6-heptynoic acid and n-BuLi with paraformaldehyde to generate a primary alcohol from a terminal alkyne. The resulting product was then converted into a protected hydrazine using EDA and tert-butylcarbazate, forming the core structure of the linker. This was then reacted with the SAH analogue.

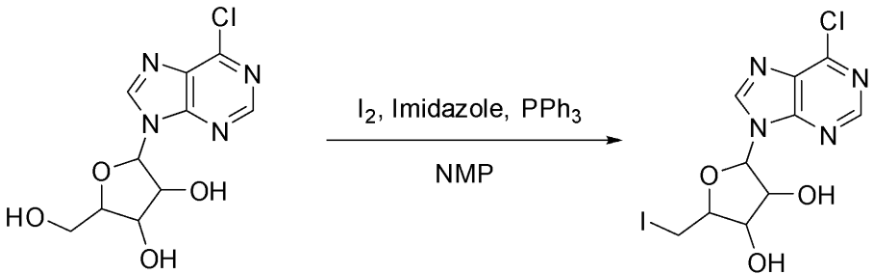
Another Appel reaction was used to convert the primary alcohol to a bromine, which was then added to the SAH analogue in a 1:1 mixture of formic and acetic acid to create the sulphonium center. The compound was purified using HPLC. Lastly, the azide was introduced to the cofactor via the hydrazine. The hydrazine was deprotected with TFA, then followed by reaction with ald-Ph-PEG3-N3 to generate the final product. With all these steps complete, a new cofactor analogue was produced, ready for testing with DNA.

The next step was to test these cofactors with the two MTases in the presence of DNA, starting with M.TaqI, known for its larger active site.



Scheme 9: Synthetic scheme of linker synthesis along with reagents. Conditions and other reagents are found in the methods section. First stage used n-BuLi, HMPA, PPh_3 and paraformaldehyde in THF, yield was 58%. Second stage used EDC, Tert-butyl carbazate in THF:H_2O in a 2:1 ratio, yield was 60%. Third stage used CBr_4, PPh_3 in DCM, yield was 66%.

2.5 Synthesis Results and Discussion

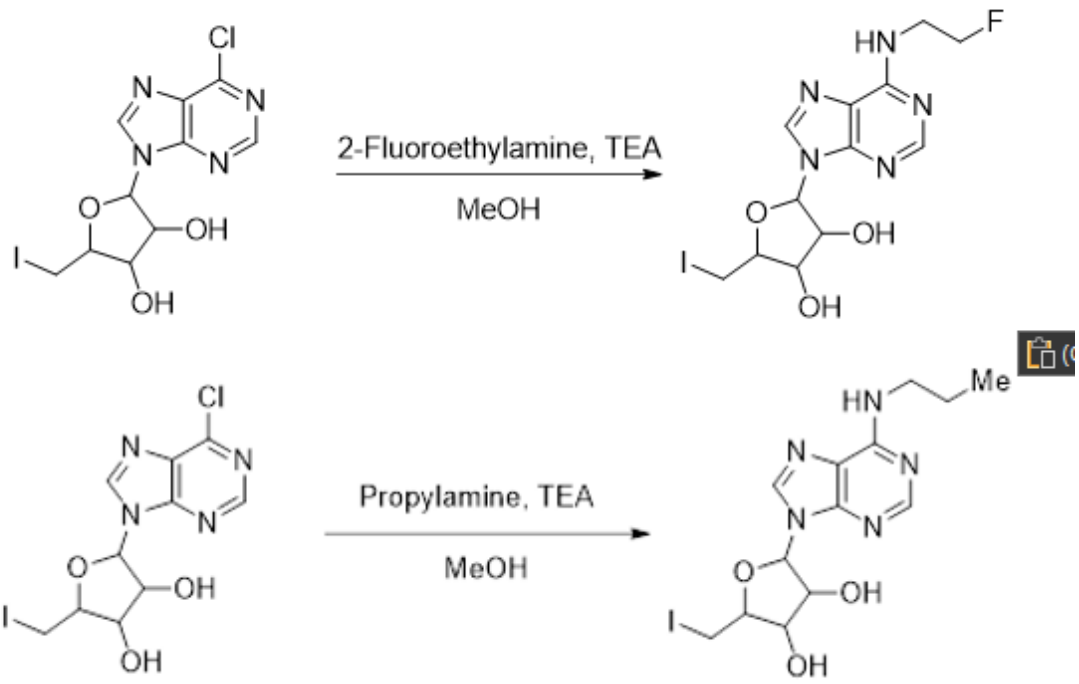


Scheme 10: Reaction scheme of 6-chloropurine riboside converted to 2-(6-chloro-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol.

The reaction was left overnight and was worked up the following day. The crude product was dissolved in the minimum amount of methanol and purified by HPLC. Three peaks were

identified at 6.5, 12, and 22 minutes, with the peak at 22 minutes corresponding to the product. The fractions containing the product were combined, and methanol was removed under reduced pressure. Water was then removed via lyophilization, resulting in a white powder. The final yield of the product was 50%.

Subsequent analysis using NMR spectroscopy and mass spectrometry confirmed the identity of the product. The ^{1}H NMR spectrum confirmed the product by the disappearance of the terminal OH group (NMR spectrum data can be found in Appendix Figure 4). Mass spectrometry was used to determine the molecular weight of the product. Full details of the synthesis process can be found in Section 6.31 of the Materials and Methods.

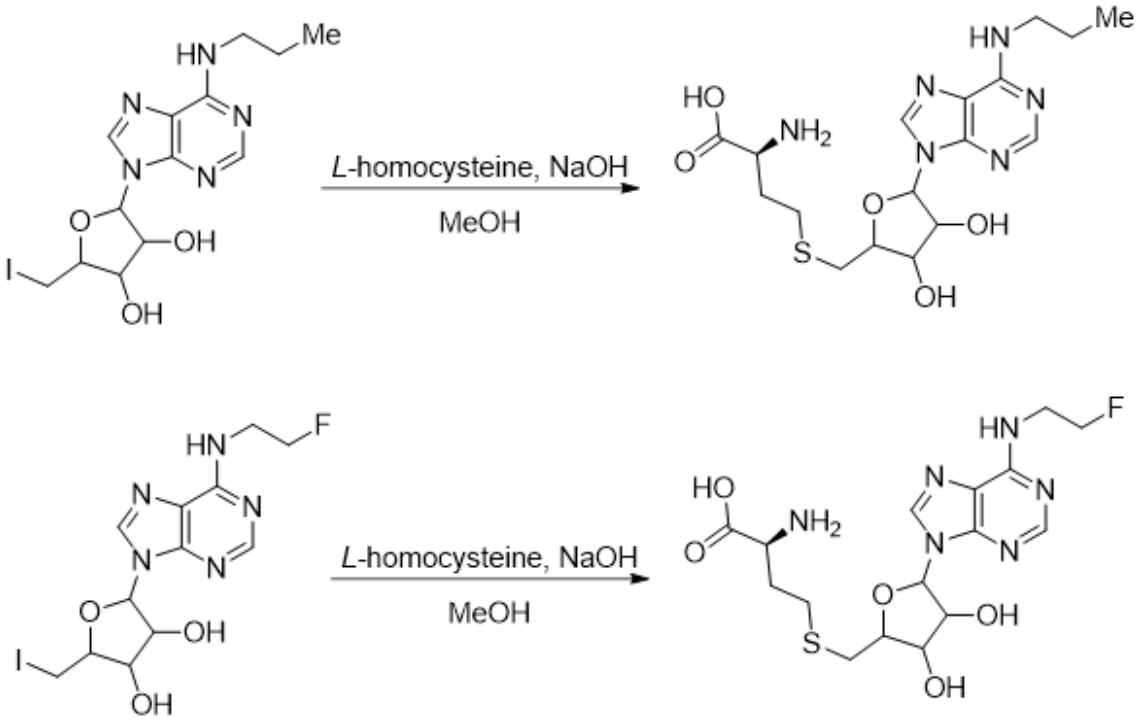


Scheme 11: Reaction scheme of 2-(6-chloro-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diolconverted to 2-(iodomethyl)-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-3,4-diol (above)/ 2-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol (below).

Two different products, leading to the **3** and **4** analogues, were synthesized using the same method, with the only difference being the amine reagent. Propylamine was used to generate the **3** precursor, whilst fluoroethylamine was used to generate the **4** precursor. The amine was stirred with 5'-iodo-6-chloro-5'-deoxyadenosine, along with the base triethylamine. They reacted overnight until completion.

The products were then purified by HPLC, resulting in a single peak for each product, observed at 260 nm. The 2-(iodomethyl)-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-3,4-diol was collected at 45 minutes, and the 2-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol was collected at 34 minutes. The difference in retention times was due to the differing polarities of the two analogues. This was confirmed by ^{13}C NMR spectroscopy, which showed 3 and 2 new peaks, respectively, for each analogue. The mass spectrum further confirmed the molecular weight of the HPLC peaks (see Appendix Figures 5, 7, 13, 14 and 24). Full details of the synthesis procedure can be found in Sections

6.32a and 6.32b of the Materials and Methods.

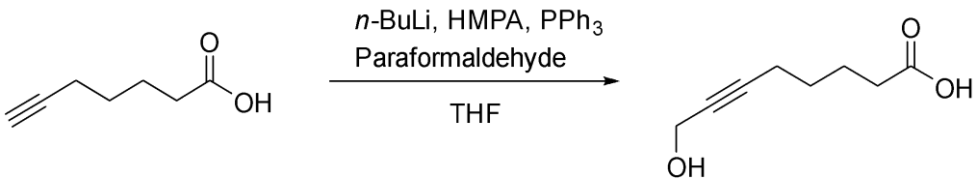


Scheme 12: Reaction scheme of 2-(iodomethyl)-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-3,4-diol converted to S-((3,4-dihydroxy-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)-L-homocysteine (above). 2-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol converted to S-((5-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-L-homocysteine (below).

The next step was to convert the halide to the cofactor skeleton. The alkyl halide was reacted with L-homocysteine in the presence of NaOH, which would then generate two stereoisomers, if homocysteine were used. However, since L-homocysteine was used, only one stereoisomer was produced.

The crude product was subjected to reduced pressure and then diluted in a mixture of 3% MeOH and H_2O, followed by purification via HPLC. The retention times were 34 minutes for the **3** precursor and 44 minutes for the **4** precursor. Further analysis was conducted to

confirm the identity of the products using a combination of ^{1}H NMR spectroscopy, ^{13}C NMR spectroscopy, and mass spectrometry. For the **4** analogue, the presence of an extra carbon in the ^{13}C NMR spectrum confirmed the product. In contrast, the **3** analogue was identified by a shift in the ppm value in the ^{13}C NMR spectrum, which corresponded to the alkyl halide carbon. Mass spectrometry also confirmed the presence of the molecular ion peak, aiding in the identification. (See Appendix Figures 6, 8, 15, 16 and 25 for detailed analysis.) Details of this synthesis can be found in Sections 6.33a and 6.33b of the Materials and Methods.

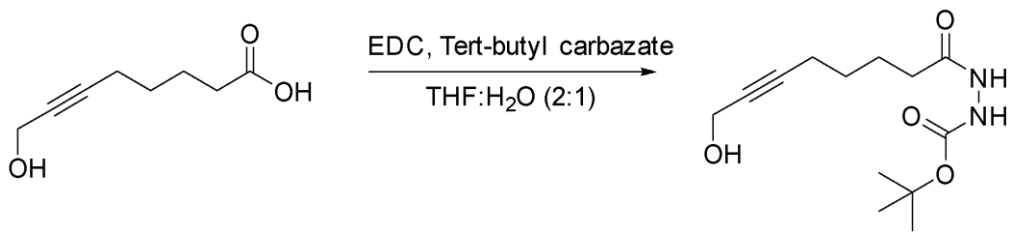


Scheme 13: Reaction scheme of hept-6-ynoic acid converted to 8-hydroxyoct-6-ynoic acid

The linker synthesis begins with the hydromethylation of 6-heptynoic acid, using 6-heptynoic acid, triphenylphosphine, n-BuLi, HMPA, and paraformaldehyde. The HMPA is then added to solvate the lithium ions, improving the selectivity of the reaction by disrupting oligomers of lithium metals. It also accelerates the S_N2 reactions by increasing the concentration of butyl anions, leading to faster deprotonation and a quicker S_N2 step, as more butyl anions are available to react with the paraformaldehyde.

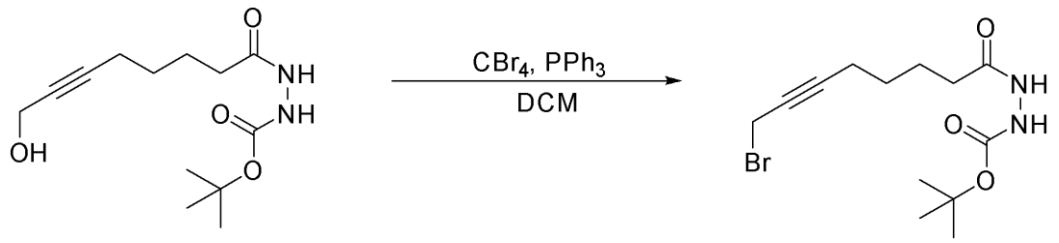
This initial step was purified via column chromatography, where triphenylphosphine oxide was removed using a gradient eluent system of hexane:ethyl acetate (8:2) to (6:4) to increase polarity. The product was confirmed by mass spectrometry, ^{1}H NMR spectroscopy, and ^{13}C NMR spectroscopy. Several peaks were observed that confirmed the formation of the desired product. In the ^{1}H NMR spectrum, a peak disappeared due to the loss of an

alkyne hydrogen, while new peaks appeared corresponding to the alcohol and the hydrogens bound to the new carbon. In the ^{13}C NMR spectrum, a new carbon peak appeared, indicating the presence of a carbon bound to the alkyne carbon and bearing an alcohol group. These observations, along with mass spectrometry data, confirm the successful formation of the compound. Details of this synthesis can be found in Section 6.34 of the Materials and Methods.



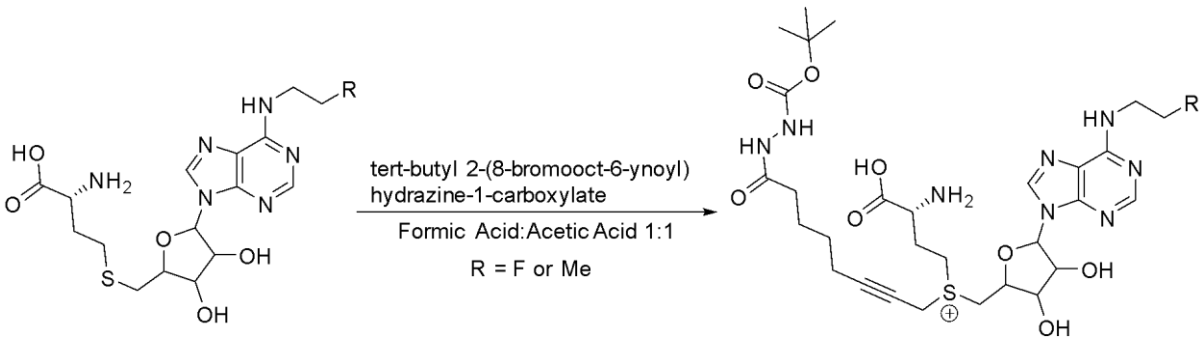
Scheme 14: Reaction scheme of 8-hydroxyoct-6-ynoic acid converted to tert-butyl 2-(8-hydroxyoct-6-ynoyl)hydrazine-1-carboxylate

The next step in the reaction involves converting 8-hydroxyoct-6-ynoic acid into a masked hydrazine molecule. This is achieved by reacting the acid with EDC hydrochloride and tert-butyl carbazate. The product was confirmed using both mass spectrometry and NMR spectroscopy (^{1}H and ^{13}C). In both spectra, new peaks appeared due to the addition of the Boc group. In the ^{1}H NMR spectrum a large singlet was observed which corresponded to 9 hydrogens all in the same environment. This had no other hydrogen atoms to couple with, as the adjacent carbon is a quaternary carbon. In the ^{13}C NMR spectrum, the addition of the Boc group was reflected by a peak in the alkyl region. The mass spectrum showed a large peak corresponding to the fragmentation of the Boc group, along with the molecular ion peak. Details of this synthesis can be found in Section 6.35 of the Materials and Methods.



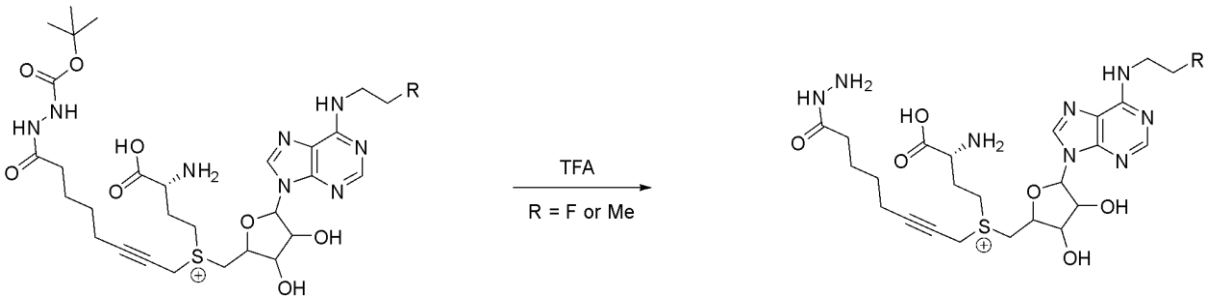
Scheme 15: Reaction scheme of tert-butyl 2-(8-hydroxyoct-6-ynoyl)hydrazine-1-carboxylate converted to tert-butyl 2-(8-bromooct-6-ynoyl)hydrazine-1-carboxylate

The next stage involved converting the alcohol into a bromide via an Appel reaction, using hydrazine, carbon tetrabromide, and triphenylphosphine. The product was purified by column chromatography. This was achieved by dry loading the column, which involved adding silica to the crude product and dissolving the silica in a minimal amount of dichloromethane (DCM). The DCM was then removed under reduced pressure, allowing the crude product to adsorb onto the silica. The purification was carried out using an eluent system of hexane and ethyl acetate in a 7:3 ratio. The product was confirmed using NMR spectroscopy and mass spectrometry. The ^{1}H NMR spectrum confirmed the product; the OH group is exchanged for the bromide, resulting in the loss of a peak corresponding to the OH. Details of this synthesis can be found in Section 6.36 of the Materials and Methods.



Scheme 16: Reaction scheme of S-((3,4-dihydroxy-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)-D-homocysteine / S-((5-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-D-homocysteinereacting with tert-butyl 2-(8-bromooct-6-ynoyl)hydrazine-1-carboxylate to form the masked hydrazine.

The next step involved the generation of the cofactor by reacting the linker with the cofactor skeleton in a 1:1 mixture of formic and acetic acid. Acidic conditions were required to generate the sulphonium ion. After allowing the reaction to proceed overnight at 35°C, the product was immediately used in the next step, which involved the addition of trifluoroacetic acid (TFA) due to the limited stability of the product. Details of this synthesis can be found in Section 6.37 of the Materials and Methods.

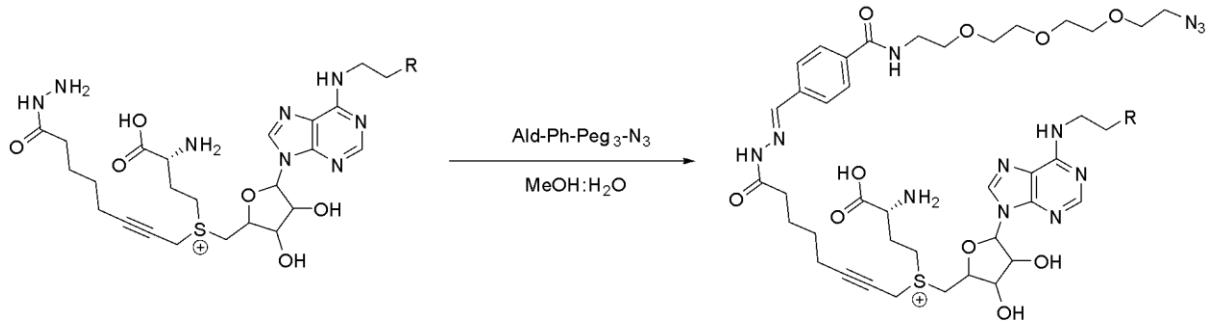


Scheme 17: Reaction scheme of masked hydrazine being converted to the unmasked hydrazine.

The penultimate step of this reaction was the deprotection of the Boc group to unmask the hydrazine. This was achieved by reacting the compound with trifluoroacetic acid (TFA), resulting in the formation of the TFA salt with the free amine.

After the Boc group was deprotected by the TFA, the acid was removed under an argon flow, and the product was purified via HPLC. The product was then dissolved in the starting eluent of the HPLC system, which consisted of 3% MeOH and 97% water. Each product showed different peaks in the HPLC. The **3** cofactor exhibited peaks at 27, 31, and 33 minutes, while

the **4** cofactor had peaks at 41 and 46 minutes. Due to the limited stability of the products, they were immediately used in the next step. Details of this synthesis can be found in Section 6.38 of the Materials and Methods.



*Scheme 18: Reaction scheme of unmasked hydrazine reacting to form either **4** or **3** when R = Me or F respectively.*

The final step in this synthesis was the formation of the final product by reacting the hydrazine with Ald-Ph-P3-Azide, generating an imide linkage between the hydrazine and the aldehyde.

This step was carried out immediately after recovering the product from HPLC, due to the limited stability of the intermediate. After adding the aldehyde to the product, the reaction mixture was rolled for another half hour. The methanol was then removed under reduced pressure, and the water was removed by lyophilization, yielding the product as a white solid for both cofactors. Due to the small amount of solid obtained, NMR spectroscopy could not be used, thua mass spectrometry was employed to confirm the products. Details of this synthesis can be found in Section 6.39 of the Materials and Methods.

Overall, this synthetic route was chosen because it had been previously employed by other groups to create new analogues of the SAM cofactor. This is a relatively straightforward synthesis with a limited number of steps which also uses relatively safe conditions. Aside from HMPA and BuLi, most of the chemicals involved are mild in terms of safety. An alternative would have been to use an enzymatic approach, which would have been safer but more costly due to reagents and would have yielded significantly less material compared to

this synthetic approach. A limitation of this route, however, is the low yield of the final cofactor, as it produces only small quantities (in the µg range), which results in a lower yield for the overall process. Another major critique of this work is the reliance on high-resolution mass spectrometry to confirm the product. Using software like ChemDraw, a predicted molecular weight for the compound was calculated. However, the mass spectrum did not exactly match the predicted mass value, likely due to the presence of isotopes and larger molecules. ChemDraw averages the isotope ratios, which could explain the discrepancy in the mass values. If more time were available, this could be explored further to better understand the differences.

2.51 Screening for Methyltransferase Activity with Novel Cofactor Analogues

After successfully synthesising the two cofactor analogues, the next step was to evaluate their impact on MTase activity with DNA. Both analogues demonstrated a similar effect on enzyme activity and provided protection to DNA when tested with the M.TaqI enzyme. However, an interesting observation was made at higher cofactor concentrations: the enzyme no longer protected the DNA. This suggests that, at elevated concentrations, the cofactor may inhibit the enzyme's activity rather than support its turnover. The gels corresponding to these results are shown in Figure 21.

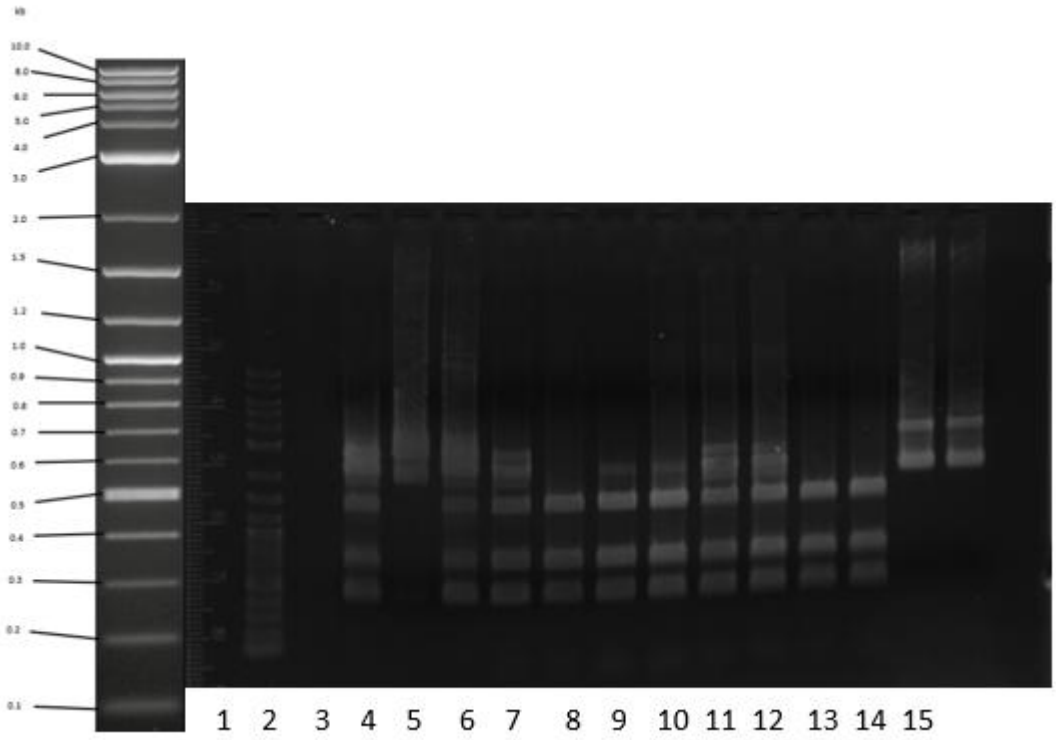


*Figure 21: Gel image of **4** cofactor with M.TaqI enzyme lanes from left to right;*
1: DNA ladder
2: Empty
*3-6: **4** isomer I serial dilution 500 µM – 62.5 µM*
*7: **4** isomer I control*
*8-11: **4** isomer II serial dilution 500 µM – 62.5 µM*
*12: **4** isomer II control*
13: Control (no cofactor, no enzyme)
14: SAM control
15: SAM control (no restriction)

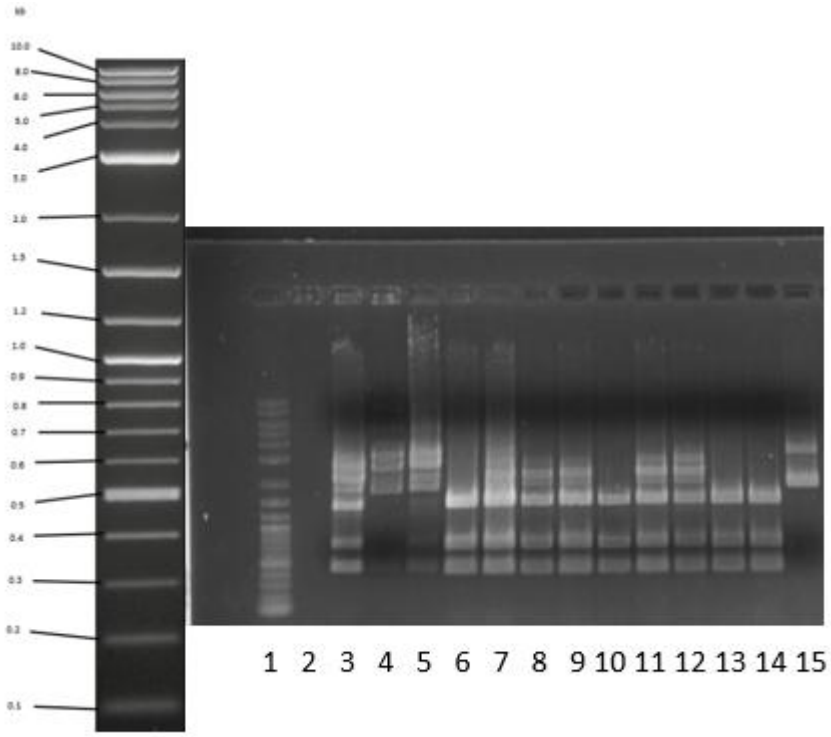


*Figure 22: Gel image of **3** cofactor with M.TaqI enzyme lanes from left to right;*
1: DNA ladder
2: Empty
*3-5: **3** serial dilution isomer I 500 µM – 125 µM*
*6: **3** isomer I control*
*7-9: **3** isomer II serial dilution 500 µM – 125 µM*
*10: **3** isomer II control*
*11-12: **3** isomer II serial dilution 500 µM – 250 µM*
*13: **3** isomer II control*
14: Control (no cofactor, no enzyme)
15: SAM control

The next experiment was designed to test how each of the cofactor analogues would influence the activity of the M.MpeI enzyme on DNA. However, both cofactors showed no protection in the presence of the enzyme. In contrast, Figures 21 and 22 demonstrate that DNA is protected when using the M.TaqI enzyme with cofactor analogues **3** and **4**. On the other hand, Figures 23 and 24 show that DNA is digested when using the M.MpeI enzyme with the same cofactor analogues.

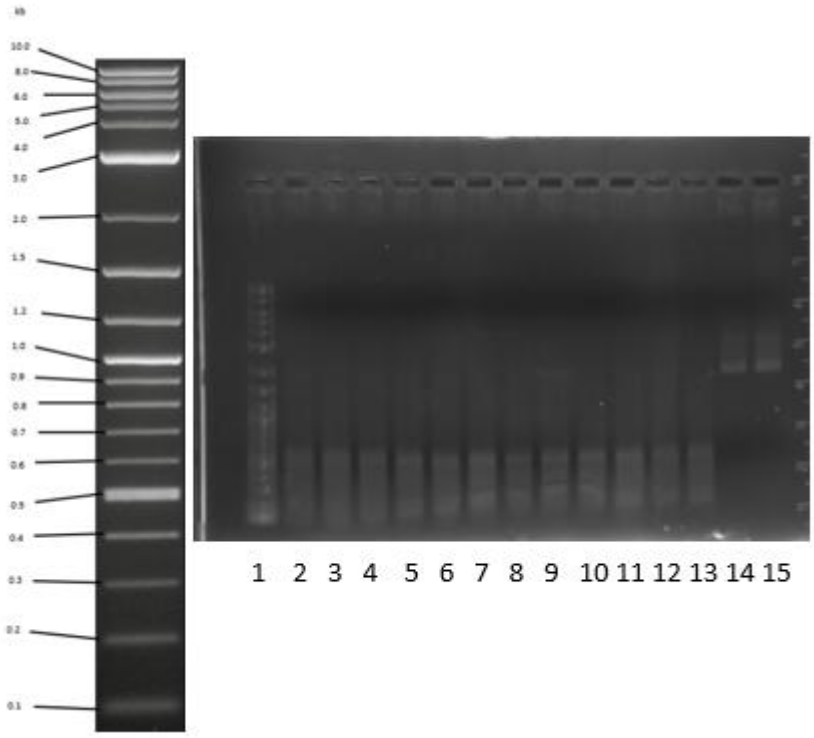


*Figure 23: Gel image of **4** cofactor with M.MpeI enzyme lanes from left to right;*
1: DNA ladder
*2-5: **4** serial dilution 500 µM – 62.5 µM*
*6: **4** isomer I control*
*7-10: **4** isomer II serial dilution 500 µM – 62.5 µM*
*11: **4** isomer II control*
12: Control (no cofactor)
13: Control (no cofactor, no enzyme)
14: SAM control
15: SAM control (no restriction)

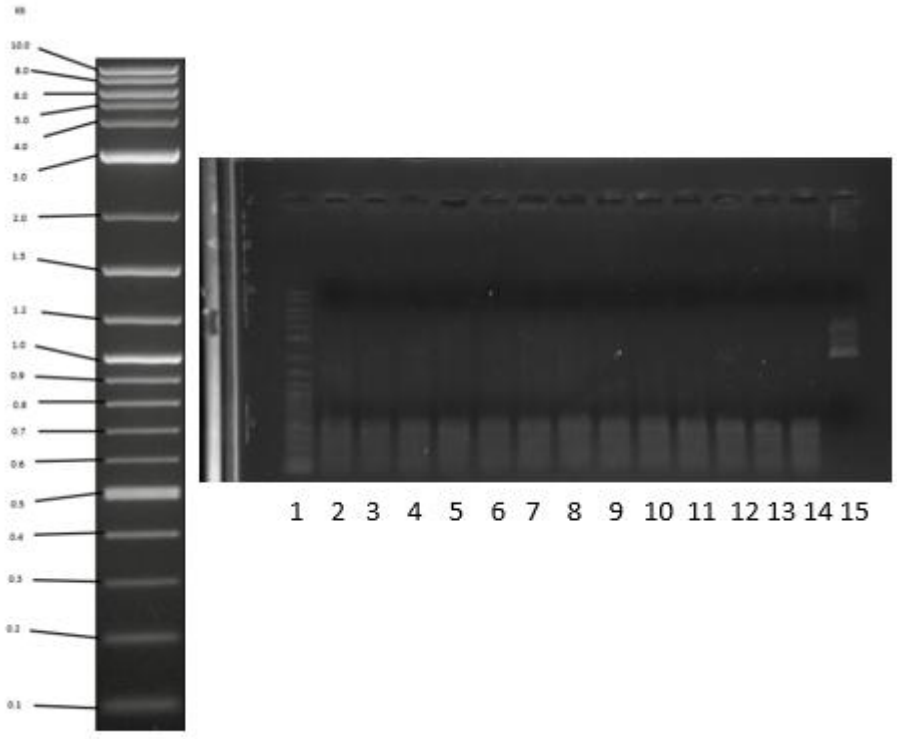


*Figure 24: Gel image of **3** cofactor with M.MpeI Enzyme lanes from left to right;*
1: DNA ladder
*2-6: **3** serial dilution 500 µM – 31.25 µM*
*7: **3** isomer I control*
*8-11: **3** isomer II serial dilution 500 µM – 62.5 µM*
*12: **3** isomer II control*
13: Control (no cofactor)
14: Control (no cofactor, no enzyme)
15: SAM control

This result highlights a discrepancy between the predictions made using docking studies and the actual experimental outcomes. Despite the docking simulations suggesting that these analogues should be suitable for the enzyme, none of them demonstrated protection in practice. This indicates that the loss of a hydrogen bond between the NH2 group and the nearby lysine residue is crucial for the transfer process to occur. This hydrogen bond likely serves as a stabilizing interaction for the transition state, lowering the energy barrier for the intermediate formation. Without this stabilizing interaction, the transition state may have too

high an energy, preventing the formation of the intermediate and, consequently, inhibiting transalkylation.

2.6 Discussion

This prompted the investigation of other cofactor analogues, such as those incorporating a carboxylic acid group. At pH 7, the carboxylic acid moiety would be deprotonated, giving the cofactor a negative charge. This negative charge could help stabilize the transition state by creating a positive coulombic interaction between the cofactor and the MTase. Additionally, the lone pairs on the carbonyl and the acidic oxygen could form several hydrogen bonds, further stabilizing the complex. A similar effect could be achieved with an amide, as the amide would not become protonated at this pH. Another analogue worth exploring would be a branched diol or triol. These structures may be as effective or even better, as they would provide more hydrogen bonding opportunities between the cofactor and the MTase. However, this approach might also pose challenges, as the larger size of the analogue could lead to clashes or interfere with other intermolecular forces, potentially disrupting the orientation necessary for efficient transfer. Additionally, if the enzyme-substrate complex is more stable than the turnover state, it could result in the enzyme becoming "used up" in forming the intermediate, leading to an inactive catalyst and the cofactor acting as an inhibitor instead.

The synthesis process encountered several challenges that slowed down the procedure, including issues with solubility, yield, and purification. For instance, the Appel reaction involving carbon tetrabromide proved difficult to purify due to its poor solubility in the column eluent, leading to low yields. To address this, the crude product was dry-loaded onto the column to improve the purification process.

2.7 Conclusion of Chapter

Analogues of the SAM cofactor have been developed with modifications made around the adenosine ring, leading to the synthesis of the **1** and **2** analogues. Both analogues were tested with two MTases, M.TaqI and M.MpeI, revealing significant differences in enzyme activity. M.MpeI exhibited varying levels of activity with the analogues compared to the **5** analogue. Notably, the **2** analogue showed better activity with M.MpeI compared to the **5** analogue, while the **1** analogue displayed no activity with M.MpeI on DNA. It was intriguing to observe such contrasting responses to the same enzyme for two different cofactor analogues.

Two additional analogues of the **2** cofactor, the methyl and fluoro variants, were also tested. Docking studies predicted that these modifications would be compatible with the MTase. After synthesis, the products were tested against both M.TaqI and M.MpeI. Whilst both analogues showed activity with M.TaqI, no observable activity was seen with M.MpeI in the presence of these analogues. This lack of activity may be due to the loss of critical bonding interactions within the enzyme-substrate complex, potentially leading to an increase in the transition state energy of the intermediate.

Docking studies also highlighted an important amino acid residue in M.MpeI, a lysine, which is located close to the amino group on the adenosine ring. This lysine residue, which carries a positive charge at pH 7, may contribute to repulsion between the cofactor and the enzyme, thus hindering DNA protection. This suggests that the failure of DNA protection could be caused by the electrostatic repulsion between the positively charged lysine and the cofactor's amino group.

Given the varying responses of the two MTases to the same cofactor, it appears to be possible to make M.MpeI active with otherwise incompatible cofactors. Achieving this, however, would require a deeper understanding of the design principles governing both the cofactor and the enzyme. With this knowledge, it may then be possible to manipulate the MTase function more effectively. The next phase of this project aims to explore how to make

these cofactors compatible with M.MpeI. This will be addressed in the following chapter, which delves into the interactions between the cofactor and the enzyme, focusing on the 115-lysine residue discussed earlier. This lysine residue plays a key role in binding interactions, particularly for the **2** cofactor, as it forms additional hydrogen bonds. However, this interaction becomes unfavorable for the **1** cofactor, as both amine groups become positively charged, leading to repulsion. The next chapter will introduce mutations of the lysine residue to explore both steric and electronic effects, testing the mutant MTases with each cofactor in a bacterial transformation experiment.

Chapter 3: *In vitro* Transcription and Translation of M.MpeI Mutant Methyltransferases: A Novel Medium-Throughput Screen

3.1 Introduction

Chapter 3 of this thesis delves deeper into the active site of the M.MpeI enzyme, with a particular focus on the amino acid residues surrounding the cofactor. These residues may influence interactions with cofactor analogues by either stabilizing or destabilizing their binding. This chapter also explores potential modifications to the enzyme to examine how steric and electronic factors affect the stability of the cofactor-enzyme complex. As a result, several new M.MpeI mutants have been designed and synthesized using a novel approach, and their activity was screened against various cofactors in a rapid bacterial assay.

In the previous chapter, it was described how new cofactor analogues containing fluoro and methyl groups were synthesized and tested with the M.MpeI enzyme in a protection assay. This assay aimed to assess how the cofactor influences the MTase activity on DNA. The results showed that these analogues did not enable the MTase to protect DNA. This observation prompted a further investigation of the active site, leading to the development of a new hypothesis. Specifically, the interaction between the MTase and the cofactor. The 115-lysine residue in M.MpeI and the amine group attached to the adenosine, may play a critical role in whether the analogue can enable the MTase to protect DNA. To test this hypothesis, the 115-lysine residue was mutated to enhance the intermolecular interactions between the MTase and the analogue.

3.2 Chapter Aims

The gold-standard protection assay typically requires the production of tens to hundreds of micrograms of purified protein, which would have been prohibitively time-consuming for the current study. To address this challenge, I developed an approach that enabled the

simultaneous production of small amounts of mutant protein using in vitro transcription and translation, whilst also testing enzymatic activity for DNA methylation. The enzymatic activity was then assessed through a bacterial transformation readout. This chapter outlines the production and initial characterization of the catalytic behavior of six M.MpeI K115 mutants in the presence of six cofactor analogues.

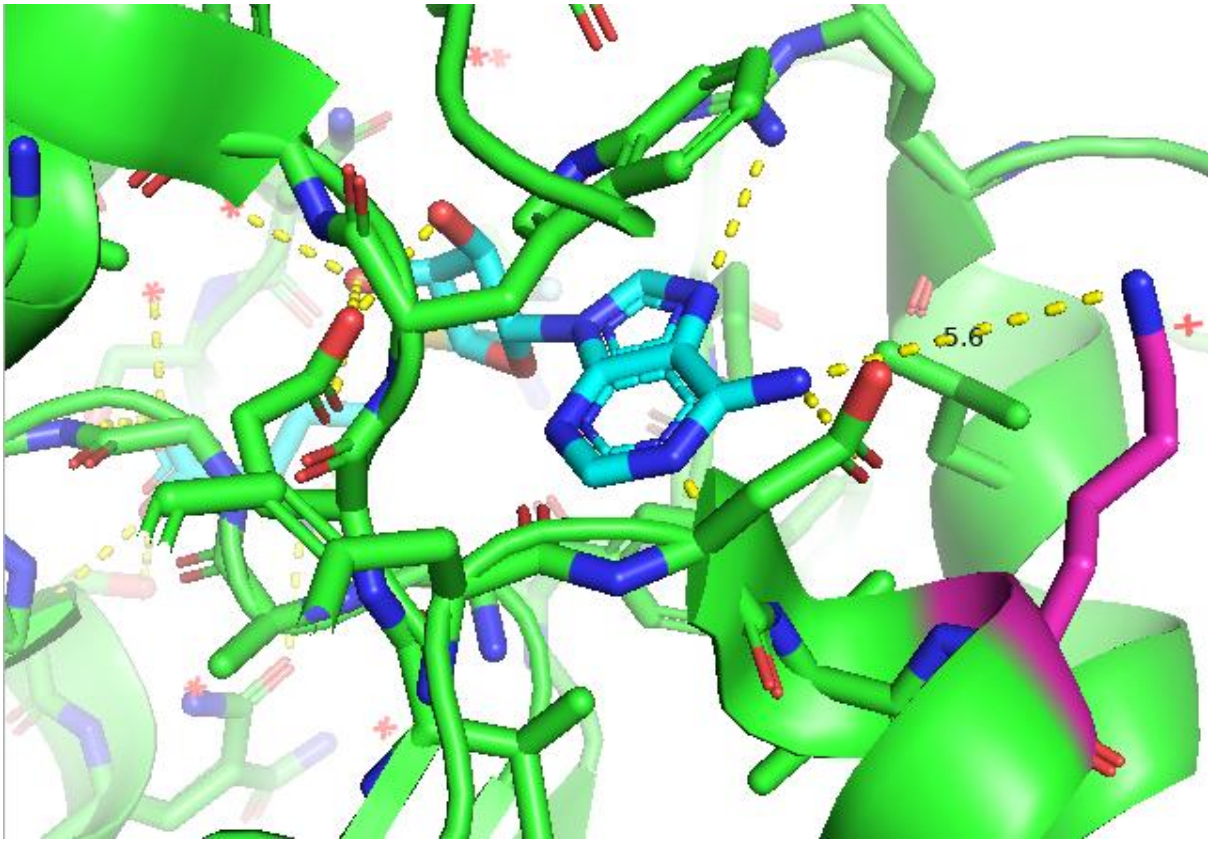


Figure 25: Pymol image of interactions with the cofactor (blue) with 115-lysine residue (magenta) yellow hatched lines show intermolecular interactions namely H-bonds between cofactor and the MTase PDB code 4DKJ.

3.3 Docking Studies

Chapter 2 detailed the discovery of the modifications to the exocyclic amino group of the cofactor analogue. This led to significant changes in the M.MpeI enzyme's ability to catalyze

DNA methylation. I identified that these changes were likely due to specific interactions between the cofactor analogue and the M.MpeI K115 side chain, with the lysine residue potentially forming a hydrogen bond with the amino group on the cofactor analogue. The results described in Chapter 2 further demonstrated that cofactor analogues lacking this hydrogen bonding ability did not function as methyl donors for the M.MpeI enzyme.

3.31 Results: Docking Studies

To then further investigate this, docking studies were conducted to explore the substitution of various amino acids for the 115-lysine residue. As observed in the previous chapter, the M.MpeI enzyme showed no activity with either isostere. The docking studies revealed a residue near the cofactor analogue, suggesting that it could influence whether a cofactor enhances the MTase activity on DNA. This residue was identified as the 115-lysine, due to its free nitrogen. The active site, including this residue, is shown in Figure 26.

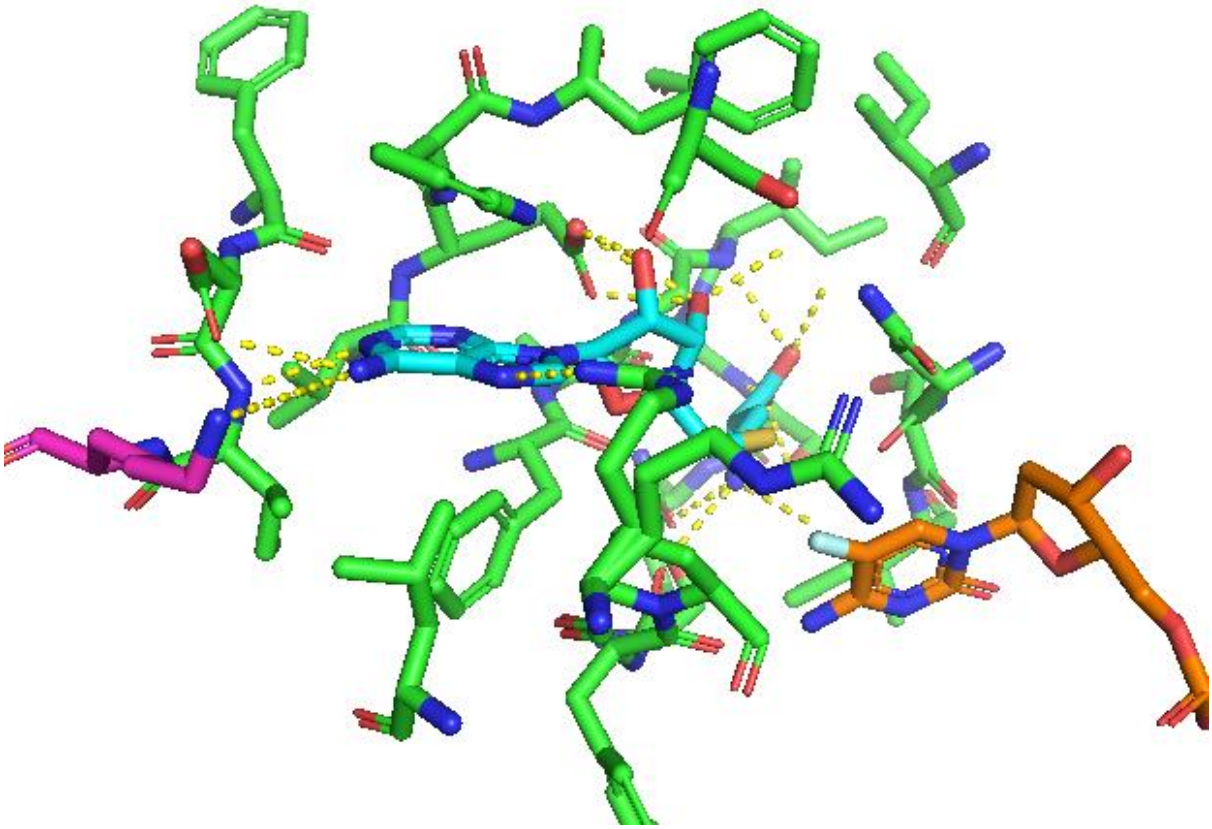


Figure 26: Pymol image of M.MpeI enzyme with SAM (blue) and the K115 residue (magenta) with cytosine bound (orange) PDB code 4DKJ.

The 115-lysine residue features a four-carbon side chain terminating in a primary amine, which could, hypothetically, orient itself in a way that facilitates binding. At pH 7, the free amine group becomes protonated, forming a positive charge on the residue. This ammonium group is close enough to the nitrogen on the adenosine ring of the cofactor analogue to form several hydrogen bonds. However, adding an ethyl chain that terminates in either an amine or alcohol group removes one potential hydrogen bond with the cofactor analogue. For the **2** analogue, the alcohol group introduces additional hydrogen bonds—its lone pairs can form hydrogen bonds, and the hydrogen attached to the alcohol can form another bond. This results in a net increase of two hydrogen bonds for the alcohol. This could then explain why the 2 cofactor performs better than the **5** cofactor, as the additional hydrogen bonds stabilize the transition state of the intermediate. The stabilization lowers the transition state energy, allowing the enzyme to turnover more frequently.

The same should then apply to the **1** cofactor due to the gain of two hydrogen bonds. However, the **1** cofactor is protonated at pH 7, making the amine positive and creating electrostatic repulsion with the lysine residue. This destabilizes the transition state and likely explains why the 1 cofactor shows no activity with M.MpeI. The **3** cofactor and its analogues are also less successful than the **2** cofactor, as these analogues lose a hydrogen bond. The amine on the adenosine ring in the **3** cofactor can only form one hydrogen bond, compared to the two in the **5** cofactor. Another potential cause of transition state destabilization is steric clash between the ethyl chain and the residue.

This leads to the hypothesis that mutating the 115-lysine could allow the enzyme to selectively bind to a specific cofactor. Docking studies did reveal hat the 115-lysine residue could be substituted with a range of different residues to test this hypothesis.

The areas investigated focused on modifying the size of the cofactor's binding pocket and the charge of the residue. To address size, we explored expanding the cofactor binding pocket by substituting smaller residues. We hypothesized that this would create more space for the modified adenosine ring to fit and reduce steric clashes. As a result, alanine, histidine, glutamine, and asparagine were selected for investigation. Figure 27 illustrates how changing the amino acid residue can impact the binding pocket.

In the wild-type MTase, the lysine residue carries a positive charge on its amine group. When the lysine is replaced by alanine, the active site becomes larger, as the lysine arm no longer occupies that space, potentially allowing the cofactor to expand. Furthermore, substituting lysine with glutamic acid flips the polarity of the residue from positive to negative. This change could be crucial in forming more stabilizing interactions between the cofactor and the enzyme, potentially enabling cofactors, such as the **1** cofactor, to facilitate MTase activity on DNA—something that would not have been possible with the original lysine.

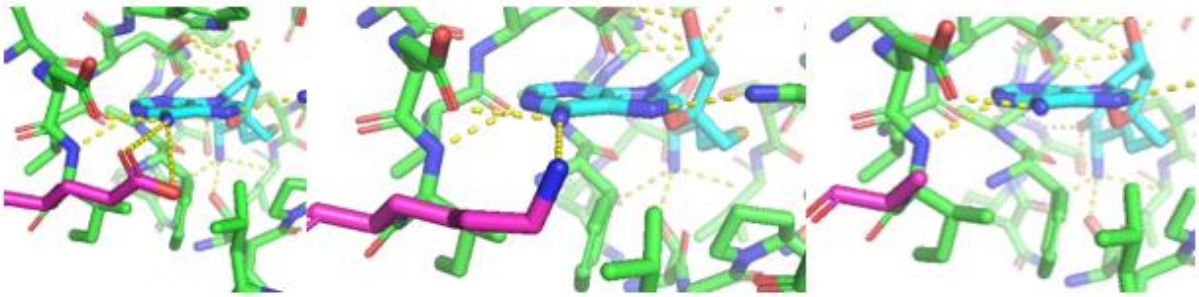


Figure 27: Pymol Images of M.MpeI (green) bound to SAM (blue) with the 115 lysine residue (purple) mutated from left to right: glutamic acid, lysine and alanine PDB code 4DKJ.

Histidine, on the other hand, significantly shortens the carbon chain length by introducing a heterocycle into the enzyme. This substitution would then restrict rotation but also maintain a small size while preserving a positive charge. The final residue, alanine, provides the most space, as its side chain consists of only a methyl group, which takes up minimal space. This would offer the cofactor the greatest flexibility when binding to the enzyme. Both asparagine and glutamine maintain the same size as lysine, but their side chains are neutral rather than positive.

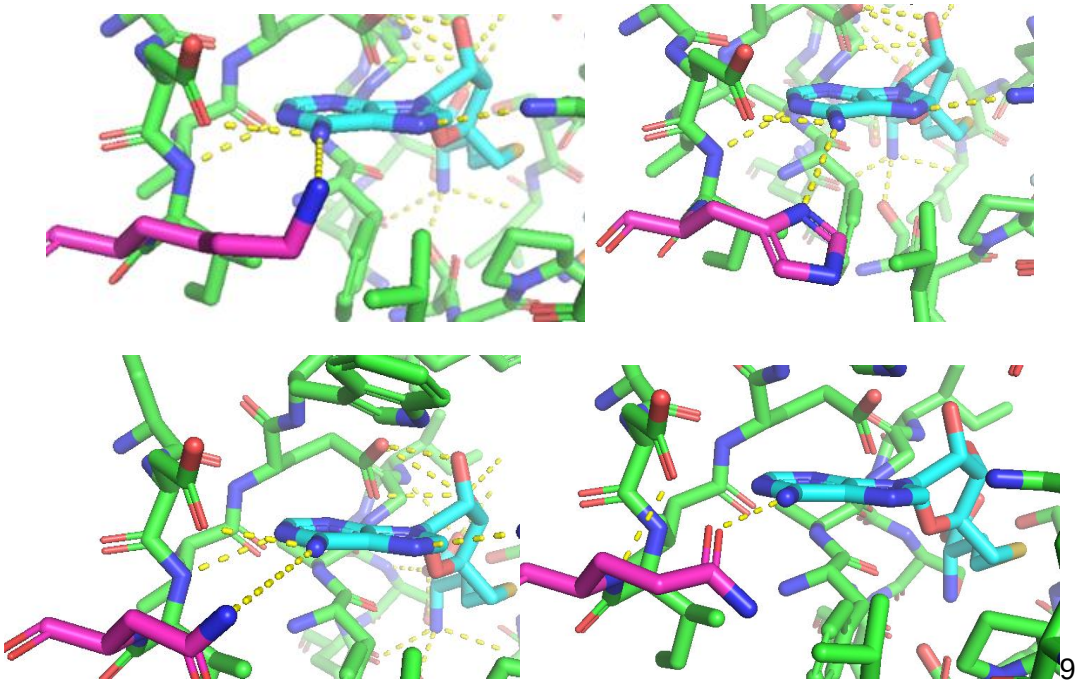


Figure 64: Pymol images of M.MpeI (green) with SAM (blue) bound, with the 115 lysine residue (purple) mutated, in order from left to right, lysine, histidine, asparagine and glutamine PDB code 4DKJ.

Another area of interest was the investigation of the charge of the residue. At pH 7, lysine's free amine group becomes positively charged. To explore how changes in both size and charge affect enzymatic activity, amino acids that alter both properties were tested. Histidine retains a free amine group, so at pH 7, it will also have a positive charge. This then allows any observed changes in enzymatic activity on DNA protection to be attributed to the differences in size between the cofactor analogues used in the assay.

Glutamine and asparagine are slightly shorter than lysine but maintain a flexible chain length, whereas histidine is much shorter and adopts a locked ring conformation. Alanine represents a residue where both the charge and size are altered. With its methyl group side chain, alanine takes up very little space, making it the smallest residue tested. Additionally, the methyl group remains neutral at pH 7.

The final residues tested: glutamic acid and aspartic acid, completely reverse the charge of the enzyme residue from positive to negative. These residues terminate in a carboxylic acid group, which can dissociate at pH 7, as the acid is effective at stabilizing a negative charge.

In this study, docking simulations were conducted on the ternary complex of SAM and each of the proposed M.MpeI mutants with DNA. Binding energy values were assessed for these mutants, as well as for other amino acid substitutions such as K115W and K115Y. Mutant enzymes were excluded from further analysis if they showed either a very high docking energy or if cofactor binding resulted in SAM not orienting correctly (relative to the expected orientation from the crystal structure of the native enzyme-ternary complex). This is important for efficient methyl group transfer. Details of the docking procedure and results can be found in the appendix.

The primary reason for excluding these mutants is that the substituted residues (tryptophan or tyrosine) are bulkier. This increases the space occupied in the active site. This reduction in available space can hinder proper cofactor binding, making the cofactor analogue less favored by the enzyme. Another disadvantage of these residues is their rigid ring structure. Unlike residues with alkyl side chains, which can rotate freely and allow for efficient use of space in the active site, the ring structure of tryptophan and tyrosine restricts rotation. This rigidity can block or limit access to the cofactor binding pocket, further destabilizing the complex and raising the docking energy.

3.32 Discussion: Docking Studies

Docking analysis suggests that several amino acid substitutions at the K115 position will be tolerated, enabling us to manipulate the space and binding interactions within the M.MpeI cofactor binding pocket. This approach is particularly advantageous for the creation of a CpG-targeting enzyme that catalyzes DNA transalkylation using a specific non-natural

cofactor analogue. Such an enzyme, with minimal or low affinity for SAM, could support a range of potential diagnostic or therapeutic applications.

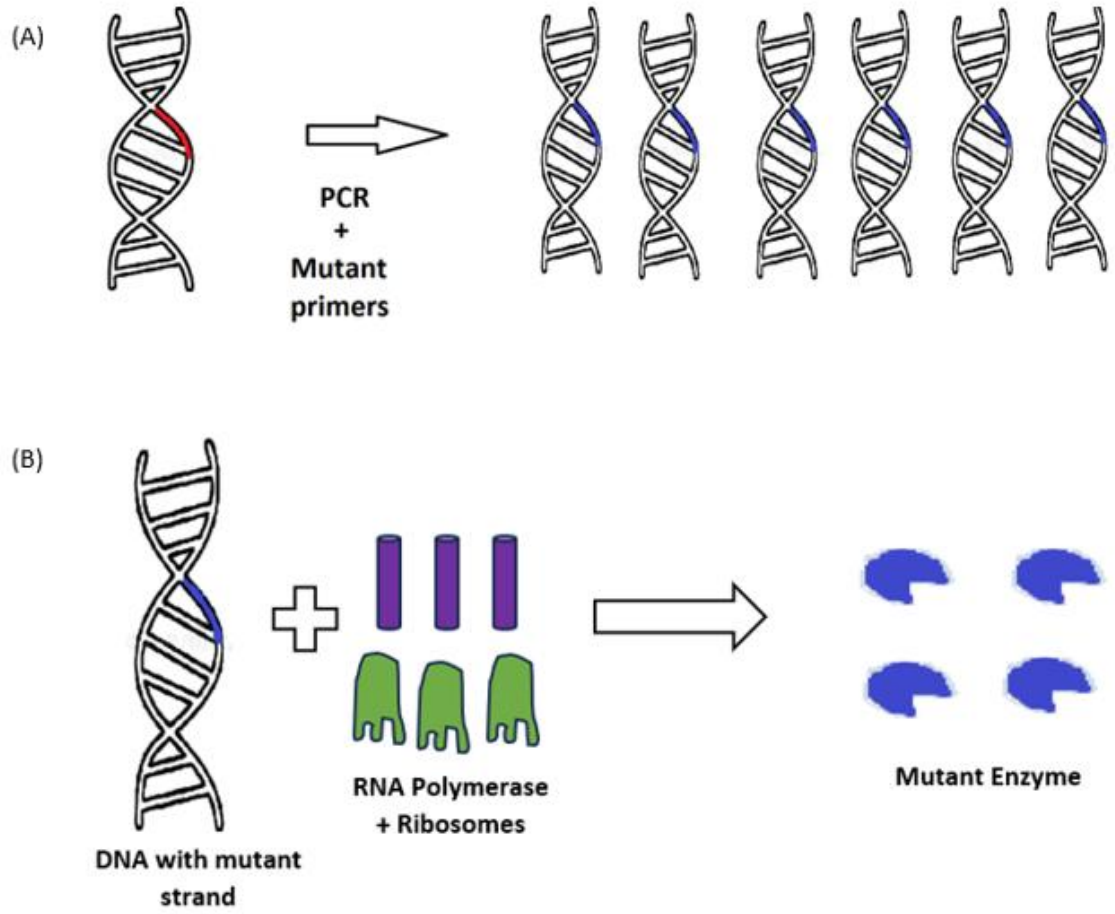
The substitutions made were designed to probe two key factors: the size of the active site and the charge on the residue. The key mutants of interest—K115A, K115D, K115E, K115H, K115N, and K115Q—were created as part of this investigation. A brief description of how these mutants were made is provided below, with full details available in the Biological Methods section (6.44 and 6.46). This method was chosen over more traditional approaches because it allowed for the screening of multiple mutations with different cofactors in a shorter time frame.

This system can be adapted for use with other methyltransferases (MTases), not just M.MpeI. The docking studies could also easily be extended to other enzymes, such as those in the DNMT family. Additionally, this system takes advantage of the speed of docking studies to quickly scan a library of compounds. Once promising candidates are identified, these can then be further analyzed using molecular dynamics, or incorporated into free energy perturbation models to obtain precise energy values for these analogues.

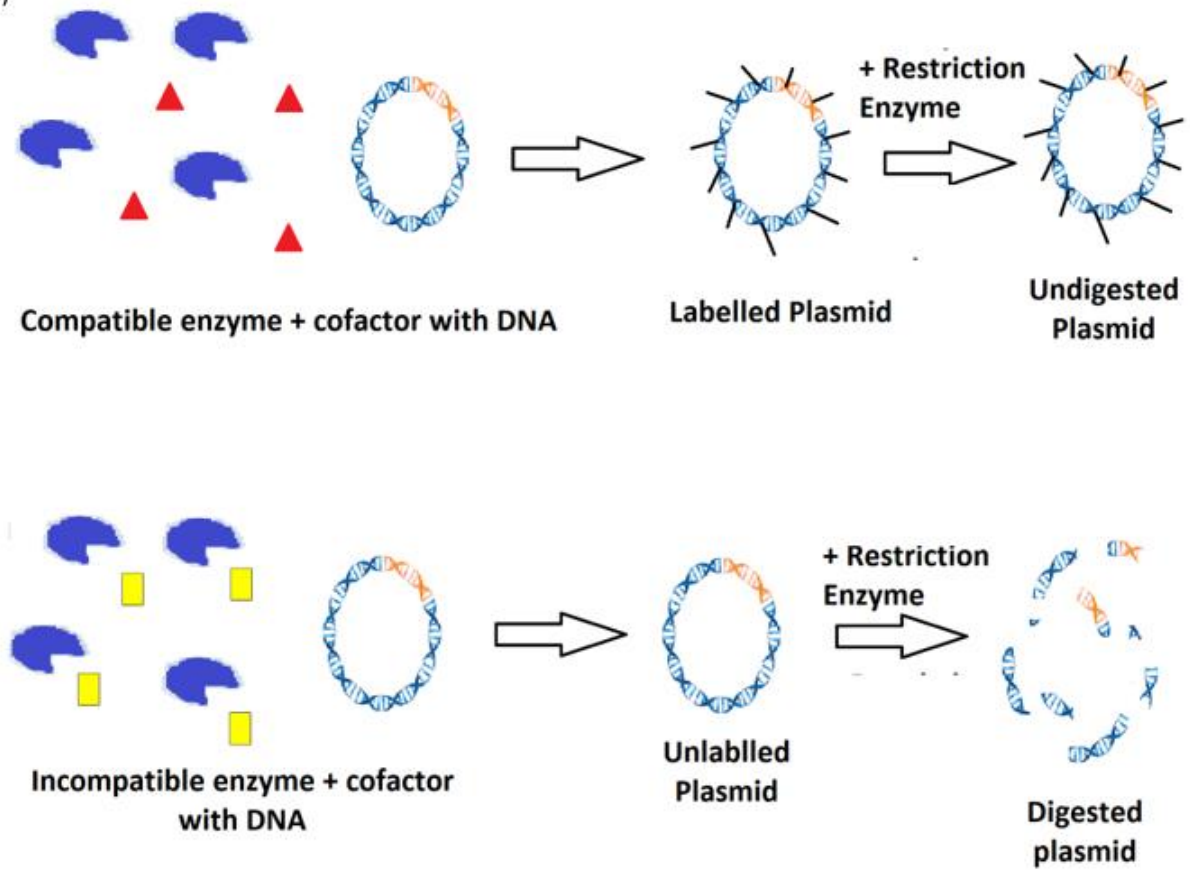
3.33 Development of a Novel Approach for Screening DNA Methyltransferase Activity

To create and test these M.MpeI mutants experimentally, in-vitro transcription and translation (IVTT) was used to produce small quantities of the MTase enzyme in the presence of DNA. This then encoded the enzyme and a selected non-natural cofactor analogue. If the enzyme-cofactor analogue combination was capable of DNA transalkylation, the plasmid encoding the mutant would be modified at available CpG sites and protected from DNA digestion. This would then allow us to identify those MTase-cofactor combinations that demonstrate apparent activity on this scale. This could be prioritized for further study.

This approach eliminates the need for enzyme over-expression, purification, and subsequent screening, thereby significantly accelerating the process of identifying promising enzymes. Figure 28 illustrates how this was achieved.



(C)


+ Restriction Enzyme
Compatible enzyme + cofactor with DNA
Labelled Plasmid
Undigested Plasmid
+ Restriction Enzyme
Incompatible enzyme + cofactor with DNA
Unlablled Plasmid
Digested plasmid

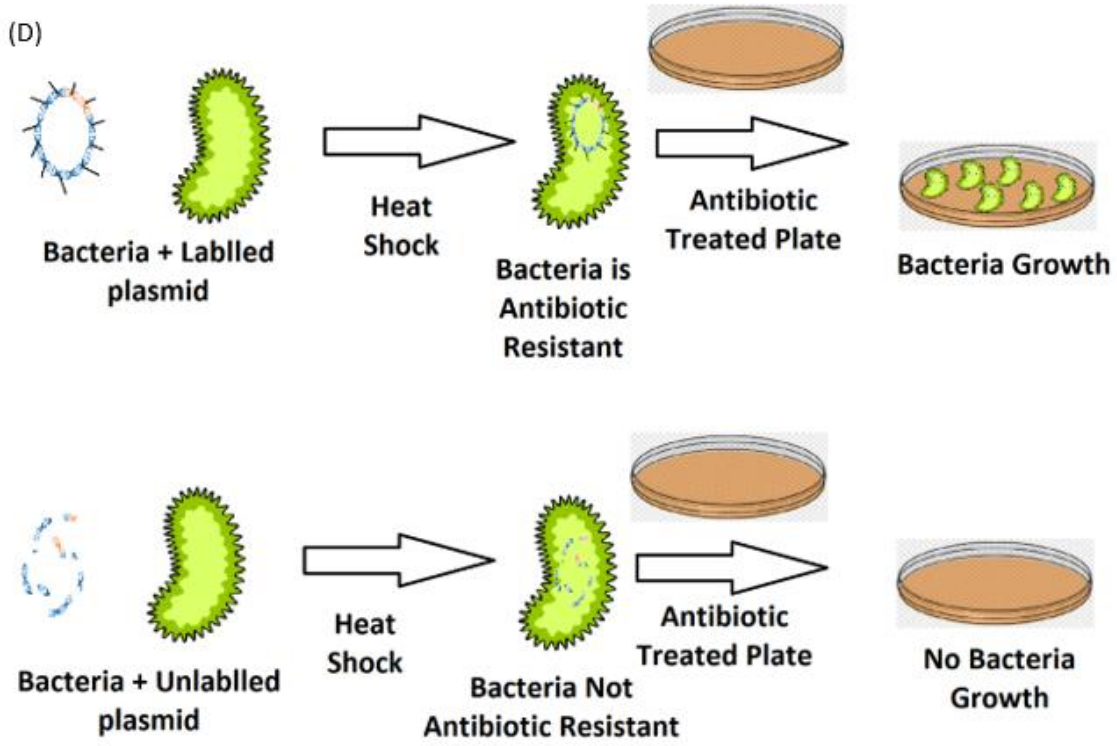


Figure 28: Schematic detailing each major step of transforming the bacteria with the antibiotic resistant plasmid, (A) The original DNA strand that encodes for the 115-lysine residue (red) has PCR performed with mutant primers to change what amino acid is in the 115-position (blue). (B) mutant DNA is then added to in-vitro protein synthesis kit which contains RNA polymerase and ribosomes to generate mutant enzyme. (C) The mutant protein is then mixed with cofactor and DNA if compatible (top). Plasmid is protected from digestion and so is the antibiotic resistant gene (orange) if not compatible (bottom) thus the plasmid is digested when exposed to restriction enzyme. (D) The plasmid is then transferred to bacterium via heat shock and added to the agar plate which has been pre-treated with the antibiotic (top). This shows if the enzyme and cofactor were compatible. The plasmid stays intact and so the bacteria were able to grow on the plate (bottom). This shows that the enzyme and cofactor is not compatible as the bacteria wouldn't grow on the plate.

3.34 Mutation of Plasmid and Bacteria work

The M.MpeI mutants were generated through site-specific mutagenesis of the plasmid encoding the M.MpeI enzyme. For full details, refer to sections 6.44 and 6.46 of the Methods section.

Primers for the site-specific mutagenesis were designed to have a GC content greater than 40%, with a termination on either a C or G, and a melting temperature (Tm) close to that of the reverse primer (52°C).

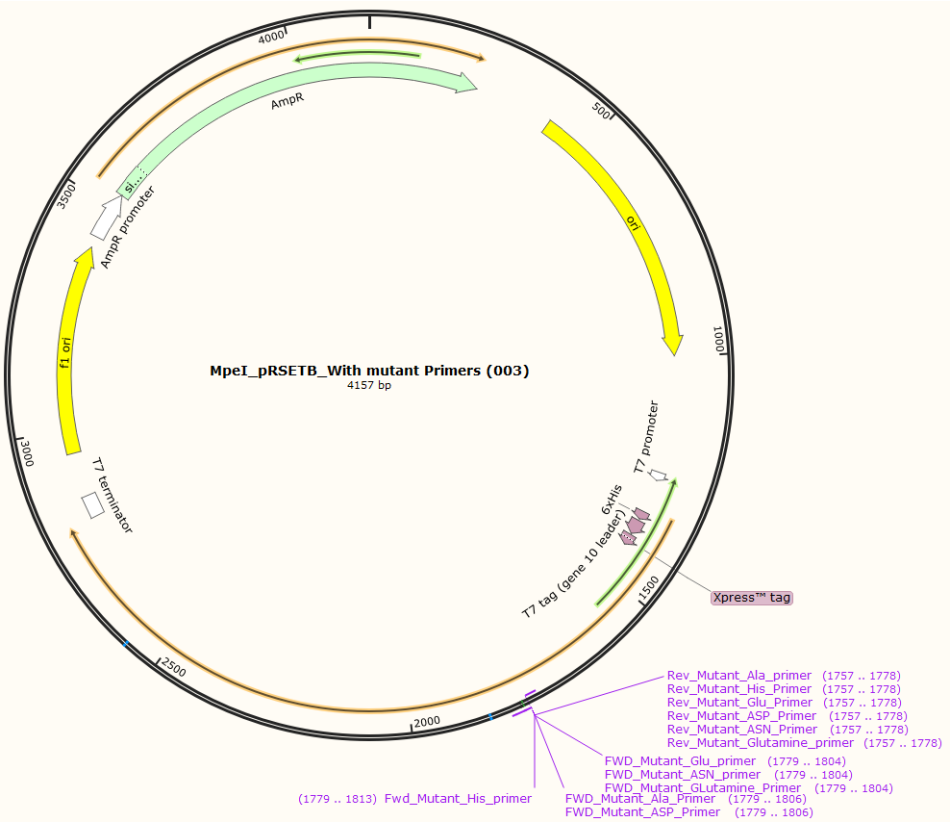


Figure 29: Map of the plasmid that was used to help design the primers that would generate the amplicon with the gene that gives a mutant protein when it undergoes transcription/translation. (top) Shows the sequence where the primers were made to bind to, the green highlight sequence of the top strand shows the codon for the 115-lysine residue above shows how the primer sequence was altered that would give a different codon for another amino acid. (bottom)

To generate a sufficient concentration of plasmid for each mutant MTase, PCR was performed using the primers shown in Figure 29 to produce a high concentration of plasmid containing the specific mutation. The resulting plasmid was purified, and Sanger sequencing was employed to confirm the presence (or absence) of the mutation (see Appendix, Figures 9 to 13).

This plasmid was then used with the NEB PURExpress® In Vitro Protein Synthesis Kit, following the supplier's instructions, to produce the mutant MTase. A detailed description of this method can be found in section 6.46 of the Methods section.

3.4 In Vitro Transcription/Translation of M.MpeI Mutants and Subsequent Screening of their Activity

In a single tube, the plasmid encoded a mutant M.MpeI enzyme, a cofactor analogue, and the in vitro transcription/translation Master Mix were combined. If the mutant enzyme and cofactor analogue form an active complex, the plasmid DNA encoding the mutant enzyme will be protected from restriction digestion at CpG sites.

The plasmid DNA mixtures were then treated with the HpaII restriction enzyme, and the resulting mixture was used to transform the bacteria. The transformed bacteria were plated on agar and incubated at 37°C overnight. The bacteria were subjected to antibiotic selection, with colonies forming only in those containing the intact plasmid. Thus, colony formation on the plate serves as an indicator of an active M.MpeI-cofactor analogue pair.

The number of colonies formed for each mutant enzyme was compared to that of a control plasmid, which contained the wild-type MTase and the SAM cofactor. This comparison allowed for the assessment of how well each mutant MTase-cofactor combination performed.

The table below summarizes the results obtained, with colour coding used as follows:

- **Green** indicates that the mutant enzyme-cofactor combination performed as well as, or better than, the control plasmid, with the same number or more colonies observed compared to the control.
- **Yellow** indicates that colonies grew, but fewer than those observed with the control plasmid.
- **Red** indicates a failed combination, where no colonies grew, or only one or two colonies were present, potentially due to an anomaly.
- **White** indicates the control plasmid, which is not compared to itself.

Representative images of the resulting plates are shown in Figure 30. These images show plates with colonies resulting from the activity of the wild-type M.MpeI with SAM, K115Q with 3, and K115A with 1. Results for all combinations are provided in the Appendix (Appendix Figures 25 to 30).

Table 2: Table of mutant enzymes tested with each cofactor in a bacterial protection assay. The colours represent how well the bacteria grew with each combination. Red means very few or no colonies grew, yellow means a good number of colonies grew compared to the wild-type SAM combination (White). Green means lots of colonies grew, even performing better than comparison combination.

	SAM	γ-aminobutyric acid	**2**	**1**	**3**	**4**
Wild-Type						
A						
D						
E						
H						
N						
Q						

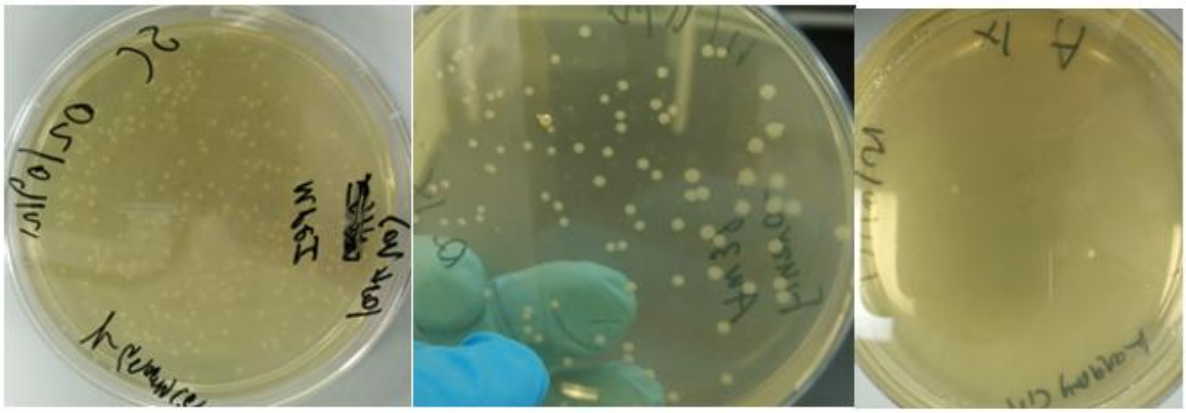


*Figure 30: Photos of some of the results from the in vitro transcription translation, (left) SAM cofactor with wild type M.MpeI, (middle) K115Q M.MpeI with **3**, (right) K115A M.MpeI with **1***

The baseline for these studies was the plate showing the activity of the wild-type M.MpeI enzyme with the SAM cofactor, which yielded around 200 colonies. All mutant and cofactor combinations were compared to this result. No colonies were observed with the wild-type

enzyme and the **1** or **4** cofactors, indicating no activity with these combinations. Wild-type M.MpeI showed activity with both the SAM and the GABA and **2** cofactor analogues.

The K115A mutant showed no colonies with any cofactor, suggesting that it was completely inactive. The K115D and K115N mutants displayed no activity with any cofactor except for **2**. The K115E and K115H mutants showed activity only with SAM and the **2** cofactor. The K115Q mutant showed no activity with cofactors except for **2** and **3**. Interestingly, the K115Q mutant lost activity with the native SAM cofactor, but this activity was recovered with the **3**-cofactor analogue. None of the other mutant MTases showed activity with the **3** cofactor.

K115E and K115H retained activity with the SAM cofactor. K115D and K115N showed no activity with SAM but regained activity with the 2 cofactor analogue, which was also compatible with the wild-type M.MpeI enzyme.

3.5 Discussion

The most significant result was that of the K115Q MTase mutant, which replaced the lysine residue with glutamine. This change resulted in the MTase losing its activity with the SAM cofactor but gaining activity with the **3**-cofactor analogue. This finding demonstrates that an enzyme can become active with a cofactor analogue if certain factors are carefully maintained. Specifically, the size of the residue must remain similar, as glutamine only differs from lysine by one carbon. Significant changes in residue size can lead to the complete loss of enzyme function, as seen with the K115A mutant, which replaces the long side chain with a small methyl group. This alteration was the most substantial in terms of size and caused the MTase to lose all activity, as no bacterial growth was observed with any cofactor.

Electrostatic factors are also important when modifying the residue. The K115Q mutant loses the positive charge of lysine at pH 7 due to the presence of an amide group. This eliminates the ion interactions that lysine could form and prevents electrostatic repulsion, which could

also occur in wild-type M.MpeI with the **1** cofactor analogue. However, K115Q does not benefit from ion interactions between positive and negative charges like K115. Despite this, K115Q retains the ability to act as both a hydrogen-bond donor and acceptor due to the lone pairs on the carbonyl and amino groups, as well as the hydrogens attached to the amino group. This allows for ion-dipole interactions with nearby positive and negative residues without contributing to electrostatic repulsion, as the residue is neutral.

This could explain why the K115Q mutant formed fewer colonies with the **2**-cofactor analogue compared to the K115H mutant. The K115H mutant, though shorter, retained a positive charge, enabling it to form an ion-dipole interaction between the ammonium group on the histidine ring and the OH group of the ethyl chain on the **2** cofactor. In contrast, the K115Q mutant formed weaker dipole-dipole interactions with the **2** cofactor via lone pairs and hydrogen atoms on both the cofactor and the residue. These interactions were less stabilizing than the ion-dipole interactions in K115H.

The size of the residue is critical. This is evidenced by the comparison of K115Q and K115N. K115Q is one carbon longer than K115N and shows activity with the **3** cofactor, whilst K115N does not. This change does not eliminate activity, as K115N is still active with the **2** cofactor, but it needs to make more stabilizing interactions to function effectively. This was achieved by both the cofactor and residue acting as hydrogen-bond donors and acceptors. The **2** cofactor can only act as a hydrogen-bond acceptor, and the absence of an extra carbon in the chain length explains why K115N shows no activity with the **2** analogue.

The other key cofactor analogue that worked was the **2** analogue, which suggests that the electronic properties of both the cofactor and the MTase are more important than size changes in determining the enzyme's function. The **2**-cofactor analogue bears an alcohol group on its ethyl chain, which can act as both a hydrogen-bond acceptor and donor. Additionally, it does not become charged at pH 7, offering an advantage since many of the

mutant MTases become charged at this pH. This neutral charge prevented electrostatic repulsion. This is an issue that could also arise with some other cofactor analogues.

Whilst the electronic properties of the cofactor analogue and the mutant MTase are more important than size alone however, this is only true to a certain extent. This is evident from the K115A residue, which was not viable with the **2**-cofactor analogue. This suggests that the change from lysine to alanine may have disrupted important intramolecular interactions, potentially causing the protein to misfold and resulting in a smaller active site than in the wild-type M.MpeI. Alternatively, it may have lost essential intermolecular interactions between the residue and the cofactor analogue, making the transition state too energetically unstable to form, thus preventing the reaction from taking place.

GABA, a novel cofactor analogue, was also tested with the mutant MTases. This analogue usually terminates in an acidic residue, which deprotonates at pH 7. It was observed to be compatible with wild-type M.MpeI, likely due to the ion-ion interaction between the lysine residue and GABA. However, when lysine was replaced with other residues, the enzyme completely lost activity with this cofactor analogue. Although one might expect the histidine residue to work due to a similar ion-ion interaction, the loss of activity would suggest that the size of the residue is then not the only factor at play. Other residues within the active site may now cause repulsion or steric clashes with GABA, resulting in the enzyme's failure to function.

The K115E and K115H mutants, which remained viable with the native SAM cofactor, further support the idea that maintaining electronic consistency is more important than the size of the residue. The histidine residue, despite being smaller, likely avoided some steric clashes whilst retaining a positive charge. In contrast, K115E replaced the lysine's positive charge with a negative charge. This change may then destabilize the enzyme, as other residues in the active site could have relied on the positive charge of lysine to maintain stability. Whilst some repulsion could be alleviated by other residues forming new positive interactions, the

overall number of interactions may be fewer than those available in the wild-type M.MpeI. This could then explain why K115E with SAM resulted in fewer colonies than K115H with SAM.

For other mutant MTases and cofactor combinations, the enzyme completely lost activity on DNA. This suggests that a different cofactor analogue may be required to restore activity for these mutant MTases.

3.6 Conclusion

To summarize this chapter, six mutant versions of the M.MpeI enzyme were generated and tested against all the cofactor analogues. Many of these mutant enzymes and cofactor analogue combinations were unviable. However, a new mutant MTase, the K115Q mutant, was found to be compatible with one of the cofactor analogues synthesized in Chapter 2, specifically cofactor **3**. This mutant enzyme offers the advantage of not being compatible with the natural SAM cofactor, opening up the potential for selective cofactor use in future applications.

Two hypotheses about the active site were tested by creating the mutant enzymes: one concerning sterics and the other regarding the electronics of the active site. The K115A mutant, which underwent the largest size change compared to the other mutants, showed no activity with any of the tested cofactors. This mutation replaced the long lysine side chain with a small, neutral methyl group. In contrast, the K115N and K115Q mutants, which were also neutral, can act as both hydrogen bond donors and acceptors due to the presence of the amide group at the terminus of their side chains. The inability of the K115A mutant to form these interactions could explain its lack of activity, as the smaller methyl group does not

support hydrogen bonding. This highlights the importance of both sterics and electronic properties in the enzyme's functionality.

The K115(D-Q) mutants demonstrated the ability to perform DNA transalkylation with the **2** cofactor, indicating that the electronic properties of both the MTase and the cofactor analogue play a crucial role in determining the activity of the M.MpeI-cofactor analogue pair. This suggests that while sterics are important, the electronic compatibility between the enzyme and cofactor analogue is the most significant factor that influences enzyme activity.

These results highlight that the enzyme's activity is more dependent on the electronic interactions than on sterics. The K115A mutant, despite being the smallest residue tested, failed to show any activity with the cofactors. This was likely due to the loss of crucial intramolecular interactions that are essential for proper function in the cofactor binding pocket. Whilst histidine is larger than alanine, it is still smaller compared to the other residues tested, and it did show activity with some of the cofactors (SAM and 2, specifically). This success could be attributed to the retention of a positive charge on histidine contributing to the stabilizing of the active site.

The differences between K115E and K115H further emphasize the role of electronic interactions. The K115E mutant introduces a negative charge at the active site, which can then cause destabilization due to repulsion with nearby anionic residues. In contrast, K115H retains a positive charge akin to the wild-type enzyme. This charge interaction likely stabilizes the binding pocket by interacting favourably with the surrounding anionic residues. This electronic compatibility, rather than just the size of the residue, appears to be key in determining the enzymatic activity in these mutants.

It was also found that the majority of the novel M.MpeI mutant-cofactor analogue combinations tested were catalytically inactive under the conditions of the in vitro

transcription translation system. Notably, none of the cofactor analogues developed were used by the wild-type M.MpeI for DNA modification, highlighting the specificity of the enzyme for its natural cofactor, SAM.

The impact of mutating K115 was then also investigated to garner an understanding of how changes to the size and electronics of the residue could affect the enzyme's ability to catalyse DNA transalkylation with various cofactor analogues. The K115 mutants provided a valuable insight into how the size and charge of the residue can influence the compatibility of the enzyme with different cofactors, and ultimately the enzymes ability to carry out the required biochemical reaction. These findings also suggest that both steric and electronic properties of the active site play crucial roles in determining the enzyme's functionality with alternative cofactors.

In conclusion, this chapter has demonstrated that both electronic and steric factors contribute to the viability of a MTase with a cofactor analogue; with a stronger preference for electronic factors. The K115N and K115D M.MpeI mutants serve as examples of mutants that are not compatible with SAM but show compatibility with the **2** cofactor analogue. However, it is also important to note that this cofactor analogue is also compatible with the wild-type M.MpeI, suggesting that it could potentially be utilized by other methyltransferases in the cell as a methyl donor. This broad compatibility could complicate its use in in vivo studies, as it might not be specific enough for targeted modifications. This could potentially lead to off-target effects.

The compatibility of the K115Q mutant with the **3**-cofactor analogue offers exciting potential for targeted therapeutic strategies, particularly for in vivo applications aimed at unmethylated CpG sites in the genome. Since the K115Q mutant is inactive with the natural SAM cofactor, its unique activity with the **3**-cofactor analogue could be harnessed for precise epigenetic interventions, such as targeting specific genes using Cas9-directed labelling or methylation. This approach could pave the way for a novel form of epigenetic therapy.

The principles of modifying the M.MpeI cofactor binding pocket, as outlined in this study, are also applicable to other methyltransferases, providing a foundation for controlling various epigenetic processes within cells. Future research should focus on refining the design principles governing methylase-cofactor interactions. The method described in this chapter enables high-throughput screening, facilitating the identification of key amino acids and the targeted mutation of enzymes to either enhance or eliminate cofactor compatibility. This would be particularly valuable for the development of therapeutic agents that selectively bind and inhibit specific methylases.

By optimising the binding affinity of cofactor analogues, we can potentially improve therapeutic outcomes at lower cofactor concentrations, which would be critical for more effective treatments. Given the structural similarities among methylase family members, the approach established in this study could also be extended to target other enzymes, thus offering a promising route towards the precise treatment of epigenetic diseases.

Chapter 4: Characterisation of M.MpeI K115 Mutant Catalytic Activity

4.1 Introduction

Chapter 4 will explore the next phase of this research, focusing on the scaling up of promising mutant MTases that were identified in the previous chapter. This section will explore the methods used to express these mutant enzymes in large quantities. This is critical for advancing their potential applications. To begin with, this chapter will describe the criteria used for selecting the most promising mutant MTases based on their initial activity and stability profiles, and how these mutants were then scaled up for further study.

The expression of these mutants in a bacterial host, such as *Escherichia coli*, will be explored in detail. The steps taken to optimize enzyme production and increase yield for subsequent testing will be outlined. This chapter will also describe the purification protocols employed to isolate the mutants, whilst ensuring that the enzymes obtained are of sufficient purity for rigorous testing.

Following purification, functional activity assays will be performed, with gel electrophoresis used to determine the minimum concentrations of both the cofactor and MTase necessary to observe effective DNA protection. These experiments should help assess the efficiency of each enzyme and its potential for use in practical applications, such as epigenetic research or therapeutic interventions.

A deeper analysis will be conducted to understand why certain mutants exhibited superior activity compared to others, particularly focusing on the role of the K115 mutation. The outcome of the research detailed in Chapter 3 showed that K115 is a crucial residue in the cofactor binding pocket of M.MpeI, and mutating this position can significantly affect the enzyme's ability to methylate DNA. These specific results from the earlier studies, including the discovery that the K115Q mutant was uniquely compatible with cofactor **3**, will be further

investigated to elucidate how these mutations can influence the interaction between the enzyme and its cofactor, thus impacting its functionality.

By combining experimental data and structural insights, this chapter will provide a comprehensive understanding of the relationship between enzyme structure and function. These findings will be able to guide future efforts to optimize these mutants for therapeutic and research purposes, potentially opening new avenues for the development of selective methylation tools and targeted epigenetic therapies.

4.2 Chapter Aims

Chapter 4 will expand on the findings of Chapter 3 by addressing some of the key unanswered questions regarding the behaviour of the mutant enzymes. Whilst the previous results were binary—indicating whether bacteria grew or not based on DNA protection—this chapter will explore the finer details of enzyme activity, including the rate of transalkylation and the minimum concentrations of cofactor and enzyme required for effective DNA protection. These insights are crucial for understanding the functional properties of the mutants in more depth.

The focus of this chapter will be on the K115X mutants (where X = A, E, H, or Q), which were created to probe the role of the K115 residue in M.MpeI's cofactor binding and DNA methylation activity. This chapter will outline the methods used to express these mutants in sufficient quantities, detailing the optimization of *Escherichia coli* growth conditions to maximize enzyme yield. Following expression, the purification process will be discussed, ensuring that the mutant enzymes were isolated with sufficient purity for subsequent functional assays.

To provide more comprehensive data, this chapter will also describe the specific assays used to measure the enzymatic activity of the K115X mutants. These assays will go beyond simple growth measurements, offering insights into the kinetics of DNA protection and

methylation. Determining the minimum concentration of both enzyme and cofactor needed for activity will also provide critical information about the efficiency of each mutant and help identify the most promising candidates for future applications.

By addressing these questions, this chapter aims to build on the results of Chapter 3 and provide a deeper understanding of the K115X mutant enzymes' potential for targeted methylation, paving the way for more detailed investigations and future therapeutic applications.

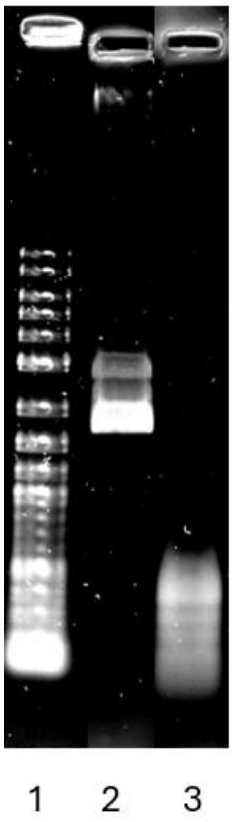


Figure 21: Simplified agarose gel which highlights the DNA ladder appears (lane 1) in relation to a positive control "protected" (lane 2) and a negative control "unprotected" (lane 3).

The focus of Chapter 4 will be on the production and characterization of the K115A, K115E, K115H, and K115Q mutant M.MpeI enzymes. These mutants were selected for synthesis based on the results from Chapter 3, where K115E and K115Q were found to exhibit

interesting properties. K115D and K115N were not included in this chapter due to the structural and electronic similarities between K115E and K115Q.

K115E and K115Q mutants were prioritised because they share similar characteristics in terms of size and electronic properties. The difference between these two mutants lies in the side chain, with glutamine having an additional carbon in its chain compared to asparagine, and glutamic acid being one carbon chain longer than glutamine. Despite this small structural variation, both mutants are expected to exhibit comparable electronic properties since both the glutamine and glutamic acid side chains are similar in terms of functionality and group attachments.

In terms of electronic characteristics, both mutants retain the same functional groups, with glutamic acid carrying a negatively charged oxygen and glutamine remaining neutral, terminating in an amide group. This allows both mutants to form a similar number of intermolecular or intramolecular interactions. Because of these similarities, the K115E and K115Q mutants were expected to behave in a comparable manner in terms of enzyme-cofactor interaction, so the decision was made to focus on the synthesis and testing of just one from this pair (K115E and K115Q), rather than both.

In this chapter, the experiments will explore how these mutations affect the structure and function of the enzyme, including their interaction with different cofactor analogues. Understanding how these mutations influence the enzyme's ability to methylate DNA will provide key insights into the design of more specific, tailored methyltransferase variants for research or therapeutic applications.

In the previous chapter, it was demonstrated that the K115Q mutant MTase was uniquely compatible with the **3** cofactor, a result confirmed by bacterial growth on a kanamycin-infused agar plate. This growth indicated that the plasmid encoding the KanR gene was

protected from restriction digestion, which suggests that the K115Q MTase, in combination with the **3** cofactor, was able to effectively methylate DNA and protect it from restriction enzyme activity. While this finding confirmed the compatibility of the K115Q MTase with the **3** cofactor, it did not provide detailed insights into the enzyme's activity, such as the rate of DNA protection or the minimal concentrations of the MTase or cofactor needed for effective protection.

To address these gaps, this chapter will build on the previous findings by quantitatively assessing the level of activity of the K115Q MTase compared to the wild type M.MpeI enzyme. The primary goal is to determine whether the K115Q mutant is significantly more or less effective than the wild type, providing a clear measure of the enzyme's functionality. This will be achieved through a protection assay, in which varying concentrations of K115Q or wild type M.MpeI MTase and the **3** cofactor will be serially diluted to assess their capacity to protect plasmid DNA from digestion by a restriction enzyme.

By using this serial dilution approach, the concentration thresholds at which protection occurs can be determined, providing a clearer understanding of the enzyme’s efficiency. This assay will also help pinpoint the minimum required concentrations of both the MTase and cofactor for effective protection, which is crucial for understanding the practical application of this enzyme-cofactor pair, whether in vitro or in vivo. Results from this analysis will allow us to make comparisons between the activity of the K115Q mutant and the wild type enzyme, leading to a more comprehensive assessment of the mutant's potential for use in targeted methylation applications, such as epigenetic studies or therapeutic interventions.

4.21 Bacterial Overexpression of M.MpeI Mutants

The first step in this process involved increasing the concentration of the mutant M.MpeI MTase from the plasmid, using the same plasmid that was utilized in Chapter 3. This plasmid, which contains the gene encoding the mutant MTase, was introduced into *Escherichia coli* cells via heat shock treatment. This process allows the bacteria to take up the plasmid and grow in the presence of kanamycin, an antibiotic. Only the bacteria that successfully incorporate the plasmid encoding for the mutant MTase gene will survive and grow on the kanamycin-infused agar plates.

After bacterial growth, a single colony of the transformed bacteria was selected and transferred into a liquid culture medium that also contained kanamycin. The purpose of using kanamycin in the liquid medium was to ensure that only bacteria carrying the correct plasmid could grow. The culture was incubated at 37 °C to allow for the bacteria to proliferate.

Once the bacterial culture reached an optical density (OD) of 0.6, which indicates an appropriate level of bacterial growth, IPTG (Isopropyl β-D-1-thiogalactopyranoside) was introduced. IPTG is commonly used to induce protein expression in bacteria containing plasmids with an inducible promoter. In this case, IPTG acts to trigger the rapid expression of the mutant M.MpeI MTase. The IPTG induction is shown in Figure 31, which likely provides a visual representation of the experimental setup and the timing of IPTG addition to the culture. Following IPTG induction, the bacteria would begin producing the mutant MTase protein, which could then be harvested and purified for further analysis in the subsequent steps of the study.

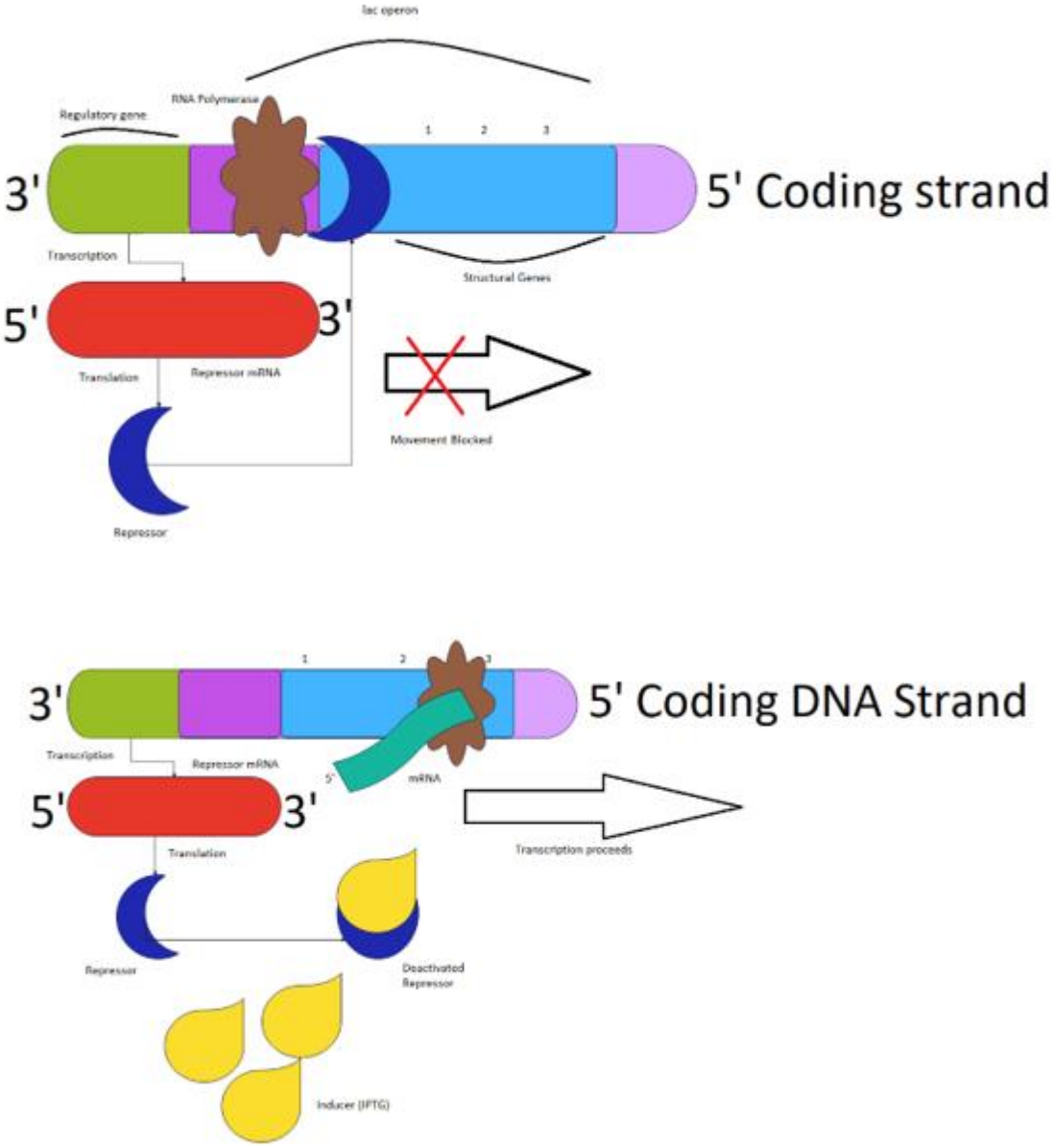


Figure 31: Cartoon illustration of the mechanism of IPTG induction, above diagram shows how the strand is blocked by a repressor which prevents the RNA polymerase producing the mRNA strand that will encode for the mutant protein but when IPTG is added this repressor is removed and transcription can proceed as normal (bottom cartoon)

To purify the expressed mutant M.MpeI MTase, the bacteria were first harvested by centrifugation, forming a pellet containing the expressed proteins. The pellet was washed with cold PBS to remove any remaining media components. Once the wash was complete, the pellet was stored until it was ready for purification.

For the protein purification process, an AKTA system was employed. The AKTA machine is a versatile protein purification device that allows for the purification of proteins using a variety of chromatographic techniques, including affinity and ion-exchange chromatography. In this case, affinity chromatography was used to isolate the mutant M.MpeI MTase proteins.

An affinity column with a nickel His-trap resin was selected for this purification. The His-trap column binded specifically to proteins with a His-tag—six histidine residues added to the C-terminus of the mutant MTase protein. The nickel ions in the column bind to the His-tag, allowing the mutant MTase to be selectively captured while the non-His-tagged proteins pass through the column and are washed away.

Once the mutant MTase was bound to the column, the next step was to elute the protein from the column. This was achieved by introducing imidazole into the column. Imidazole has a higher affinity for nickel than the His-tag, so at sufficiently high concentrations (500 mM), imidazole competes with the His-tag for binding to the nickel ions, causing the MTase to be released from the column. The eluted protein was then collected.

The AKTA system uses UV absorbance to detect protein elution, monitoring the absorption of the solution as the protein passes through the system. One challenge in using UV absorbance to detect the elution of the MTase is that imidazole absorbs in the same region of the UV spectrum as the protein, which can make it difficult to accurately identify when the protein elutes. However, the presence of the MTase in the solution slightly increases the UV absorbance, which results in a small "bump" in the absorbance profile when the MTase is released. This bump can be detected when the concentration of imidazole reaches around 150 mM, indicating the successful elution of the mutant M.MpeI MTase from the column. For a more detailed overview of the purification process and the UV absorbance profile, refer to Appendix Figures 17-20.

The collected MTase fractions were pooled together, and a solvent exchange was performed to remove the excess imidazole in the elution buffer, as it could have interfered with enzymatic activity. Filter columns with a 30 kDa size cutoff were used for this process. The MTase pool was added to the column and subjected to ultracentrifugation, allowing smaller species to pass through and be collected in a separate tube, which was then discarded. The column was then replenished with storage buffer, free of imidazole. This process was repeated at least three times before the MTase was removed from the column. The MTase was eluted by inverting the column into a clean collection tube and centrifuging at low speed, which deposited the MTase into the tube. The MTase was mixed with storage buffer, quantified by UV-vis absorption at 280 nm using a molar extinction coefficient of 86,320 M^{-1} cm^{-1}, and stored in glycerol. The catalytic activity of the mutant MTase was assessed with all cofactors (excluding GABA) used in Chapter 3, and the results were compared to the activity of the wild-type MTase with the SAM cofactor.

4.3 Results

The results of over-producing and purifying each of the M.MpeI mutant enzymes are presented in the Appendix (Appendix Figures 21 to 24). Table 3 summarises the MTases produced along with their corresponding concentrations.

Table 3: Table showing the MTases made and approximate concentrations of each MTase.

Mutant MTase	**Approximate Concentration / M**
K115A	2.38×10^{-4}
K115E	1.38×10^{-4}
K115H	6.34×10^{-6}
K115Q	6.77×10^{-5}

The protection assays conducted involved serial dilution of both the cofactor and MTase to assess the impact of modifications to either the cofactor or MTase on the enzyme's ability to catalyse complete transalkylation (protection) of DNA (50 ng μL^{-1}) over the course of one hour. This approach allowed for the observation of the mutant MTase activity profile, which was then compared to that of the wild-type M.MpeI enzyme with the SAM cofactor, under similar conditions.

4.31 K115A Mutant M.MpeI

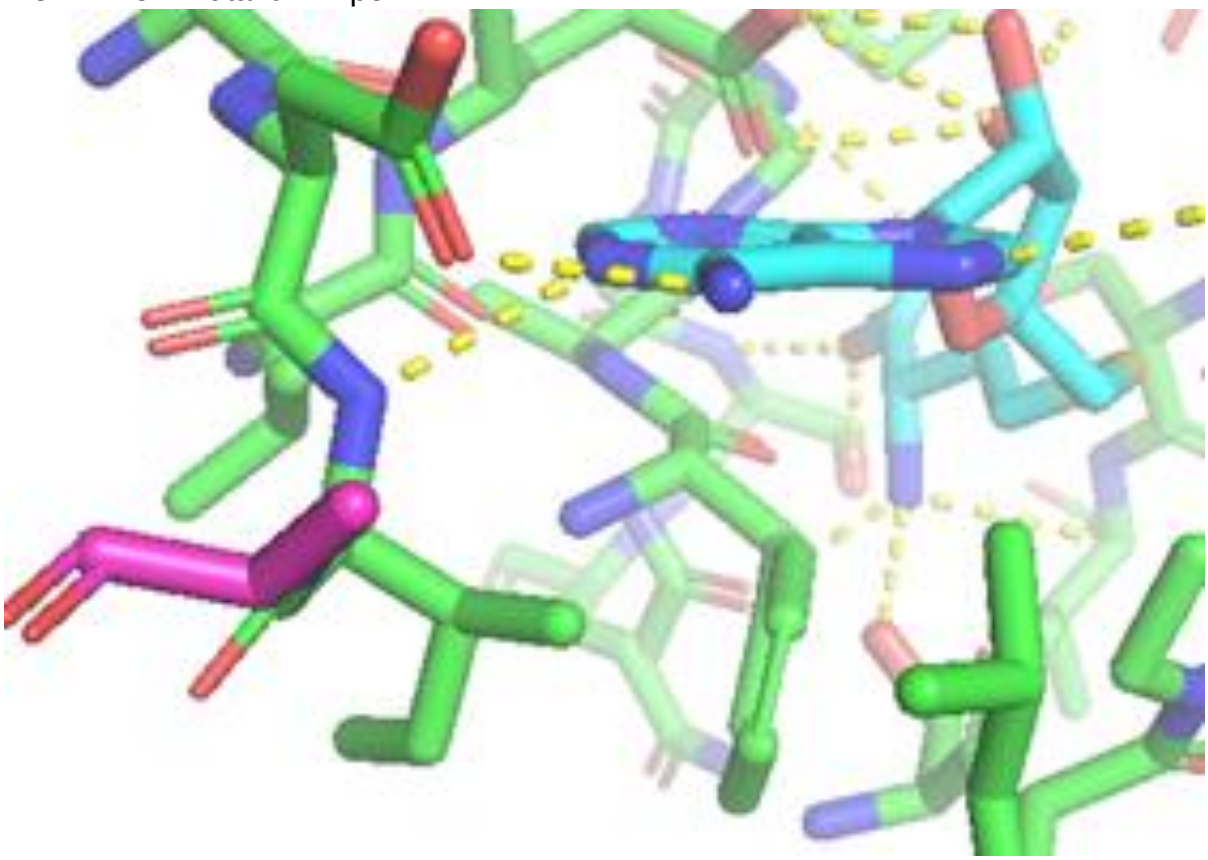


Figure 32: K115A active site, K115A is shown in purple, yellow hatched lines show intermolecular/intramolecular interactions between elements, no intermolecular forces are seen between alanine and cofactor (blue) PDB code 4DKJ.

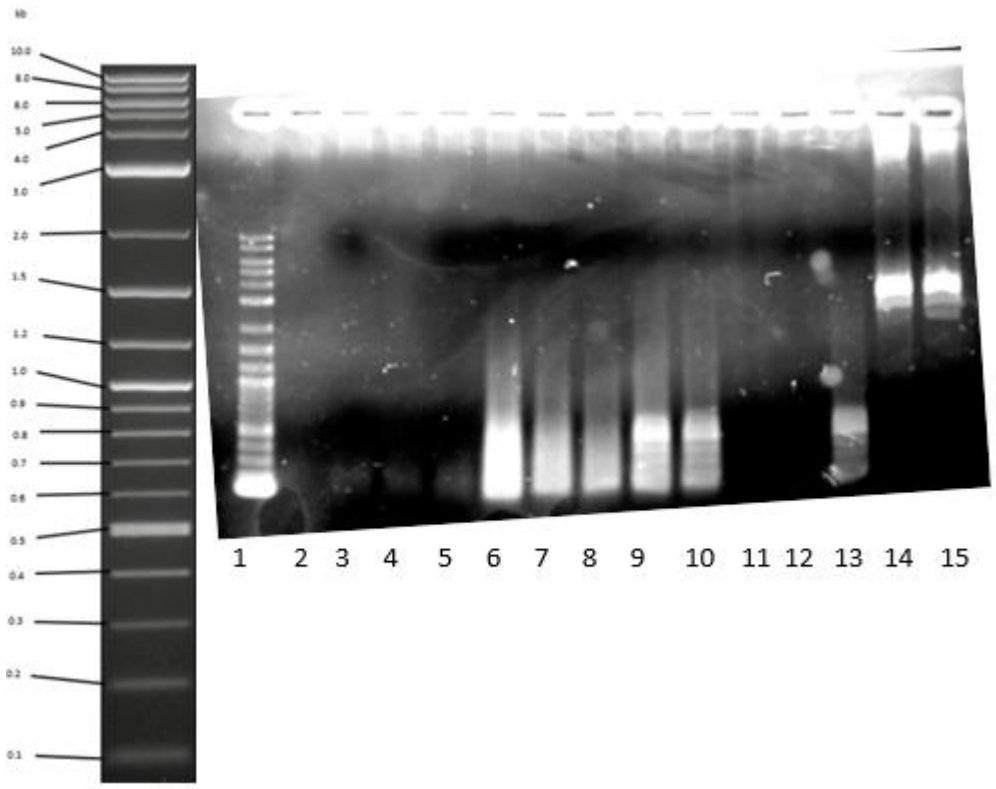


Figure 33: K115A (0.2 mM) gel results with SAM cofactor dilution. (Lane 1: DNA ladder, Lane 2: Empty, Lanes 3-7: Mutant A, 2-fold serial dilution of cofactor (500 µM to 31.25 µM Lane 8: Mutant A control, Lane 9: Control without cofactor Lane 10: Control without cofactor and enzyme Lane 13: Same as Lane 10, Lane 14: SAM control Lane 15: SAM control without restriction enzyme

Figure 33 presents an example of an agarose gel which shows the catalytic activity of the K115A mutant enzyme with the native SAM cofactor. Lanes 3–7 show a two-fold serial dilution of SAM cofactor (500 µM to 31 µM), with no detectable DNA methylation observed under any condition. In lanes 3–5, DNA degradation is observed. This is likely due to protein precipitation, which could cause some DNA to be lost, especially at high cofactor concentrations. However, lanes 6 and 7 demonstrate that no methylation activity is detected with this enzyme. Control experiments (lanes 10 and 13) confirm complete DNA restriction, while lane 14 shows that the SAM cofactor effectively acts as a methyl-donor for the wild-

type enzyme under these conditions. Similar results were obtained for all K115A and cofactor analogue combinations tested (Appendix Figures 32–34).

The K115A mutant was tested with several cofactors used in Chapter 3, including **3**, **1**, and SAM. Consistent with the results from Chapter 3, K115A showed no DNA protection from restriction when using these cofactors. If any protection had occurred, it would have been indicated by a band like those seen in lanes 14/15 in Figure 33. Lanes 3–5 appear to show less plasmid, which could be due to excess cofactor causing precipitation of the enzyme-substrate complex. These results indicate that the K115A MTase has no activity with the tested cofactors, as no protection is observed even at the highest concentration of K115A, and no further enzyme dilution experiments were performed. This outcome may be due to a lack of stabilizing bond interactions between the cofactor and the MTase, or an inability to form necessary intramolecular interactions, leading to structural failure of the MTase. Specifically, an essential H-bond interaction may not be formed, preventing proper folding of the MTase and the formation of an active site capable of accommodating the supplied cofactor. By attempting to enlarge the cofactor binding pocket, it is possible that the mutation inadvertently resulted in a reduction of the active site size. Figure 32 highlights the K115A active site, showing that there are no polar binding interactions between the mutated amino acid and the cofactor. Although van der Waals forces may still be present, they are much weaker than the H-bond interactions that K115 would normally be able to form. Additional HPLC traces and UV-Vis data for the MTase can be found in the appendix (Appendix Figures 17 and 21).

4.32 K115H Mutant M.MpeI

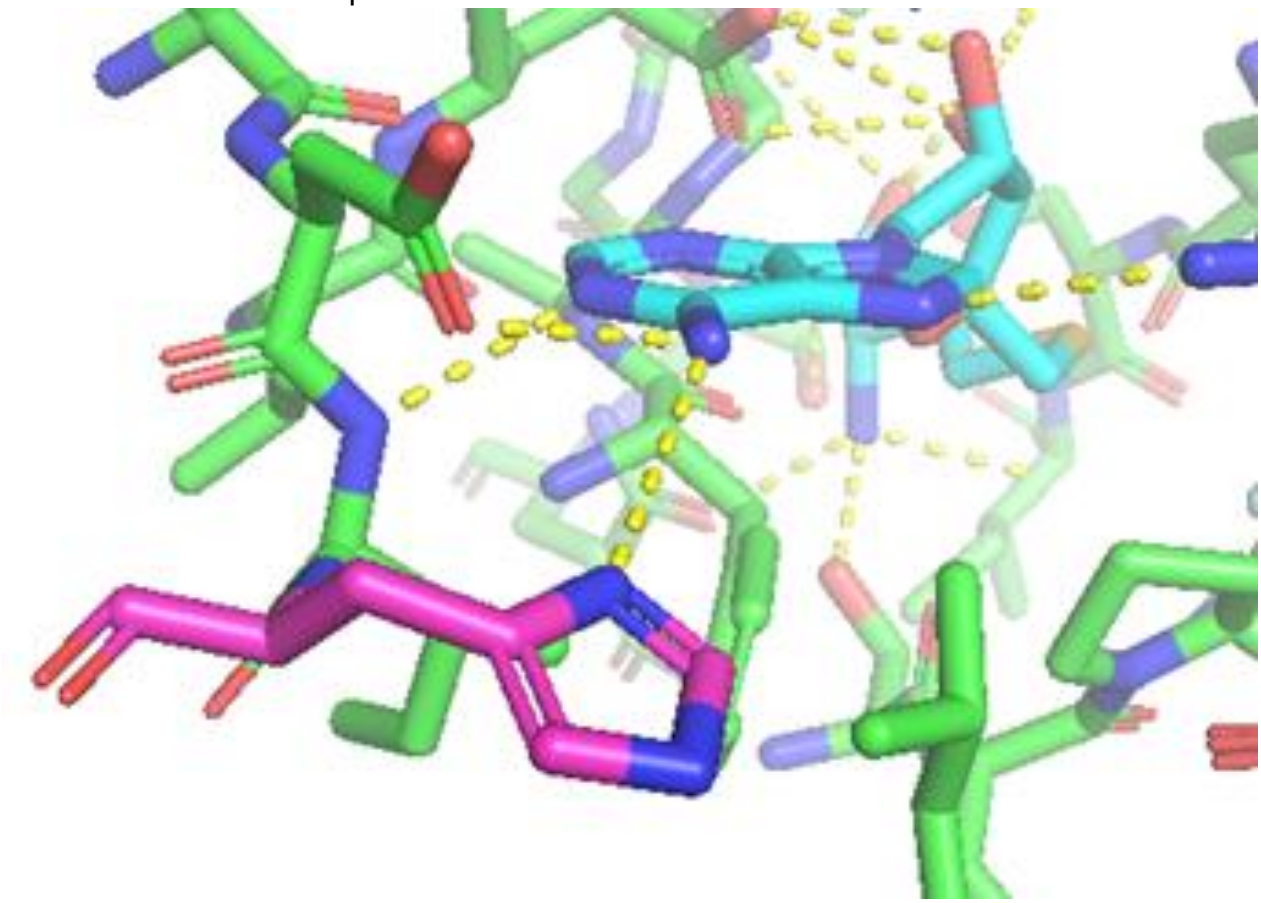


Figure 34: K115H active site, K115H is shown in purple, yellow hatched lines show intermolecular/intramolecular interactions between elements. This shows histidine can form an interaction between the nitrogen in the ring of histidine to form H-bonds with the amine group of SAM (blue) PDB code 4DKJ.

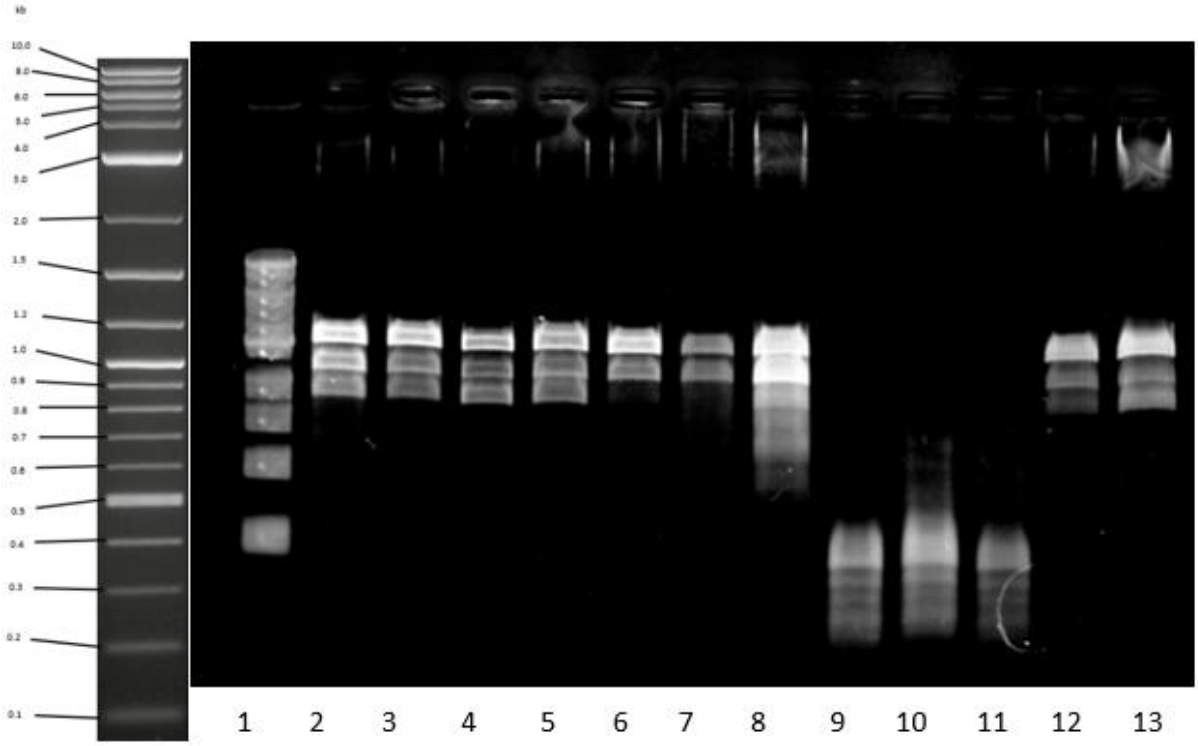


Figure 35: K115H (6.0 µM) gel with SAM cofactor with dilution of SAM (Lane 1: DNA ladder, Lanes 2-9: K11H, serial dilution of SAM (500 µM to 3.9 µM) Lane 10: K115H control (same as lanes 2-9 minus cofactor), Lane 11: Control without cofactor & enzyme, Lane 12: SAM control, Lane 13: SAM control without restriction enzyme. EC_{50} value is 6.42 µM

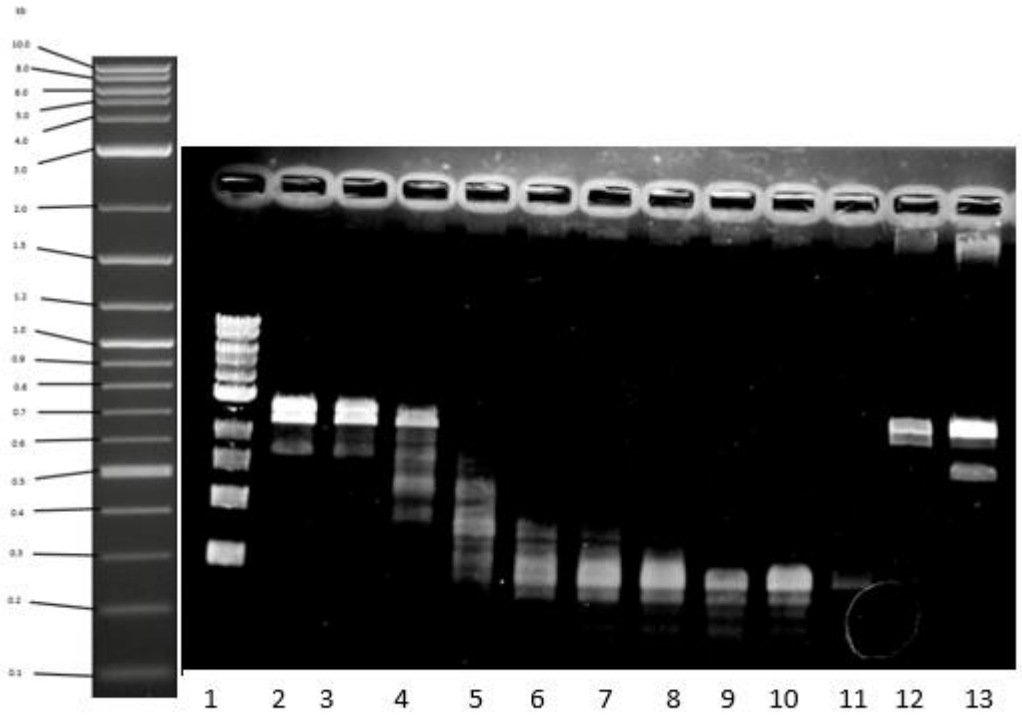


Figure 36: Gel results of K115H with SAM cofactor (500 µM) with enzyme serial dilution (Lane 1: DNA ladder, Lanes 2-8: Mutant H, serial dilution of enzyme (6.0 µM to 0.094 µM) Lane 9: Control without cofactor, Lane 10: Control without cofactor & enzyme, Lane 11: Empty, Lane 12: SAM control, Lane 13: SAM control without restriction enzyme. EC_{50} value is 1.1528 µM

Figures 35 and 36 present example images of agarose gels used to test the catalytic activity of the K115H mutant with the native SAM cofactor. Figure 35 shows the results from the dilution of SAM cofactor, and Figure 36 displays the dilution of the MTase. In Figure 35, lanes 2–8 demonstrate that K115H can protect DNA at low concentrations of SAM (as low as 7.8 µM). In contrast, lanes 4–8 of Figure 36 show that when the MTase concentration is too low (ranging from 1.5 µM to 0.094 µM), meaningful DNA protection is not observed. Control experiments, shown in lanes 10 and 11 of Figure 35, and lanes 9 and 10 of Figure 36, confirm complete restriction of the DNA. Lane 12 in Figures 35 and 36 illustrates that SAM acts as a methyl-donor for the wild-type M.MpeI under these experimental conditions. Further results for K115H and other cofactor analogue combinations can be found in the appendix (Appendix Figures 19 and 23).

This gel demonstrates that at high concentrations, K115H exhibits activity comparable to the wild-type enzyme. As shown in Figure 36, K115H is compatible with the cofactor and successfully protects DNA. However, as the enzyme is diluted to 1.5 µM, its efficiency begins to decline, and by approximately 0.75 µM and lower, it no longer provides effective protection of the DNA.

Histidine has a greater potential to form stabilizing interactions within the active site due to its two nitrogen atoms, one of which carries a positive charge. This allows for the formation of ion-dipole hydrogen bonds. This can be seen in Figure 34, where hatched yellow lines represent this interaction. The histidine ring is also small and rigid, unlike the alkyl chains in other tested mutants that allow more flexibility within the enzyme. This rigidity could introduce unfavorable steric interactions. These results indicate that the K115H mutant, when combined with the SAM cofactor at higher concentrations (>125 µM), protects DNA with comparable activity to the wild-type M.MpeI. However, at lower concentrations, its efficiency decreases. This suggests that the K115H mutant can form a similar number of binding interactions as the wild-type enzyme, unlike the K115A mutant, which fails to protect DNA even at high concentrations of SAM.

The K115H mutant may not form as many stabilizing interactions as the wild-type M.MpeI enzyme. Whilst it shows comparable activity to the wild-type enzyme at high SAM concentrations, as evidenced in Figures 35 and 36, this activity diminishes after two serial dilutions of the MTase and is completely lost after five serial dilutions. One possible explanation of this is that the histidine residue is too short to form long-range stabilizing interactions, which could then result in a reduction in the size of the active site. Additionally, histidine may be unable to form a key stabilizing intramolecular interaction, thus potentially causing the enzyme to misfold and preventing the formation of an active site in the mutant MTase. Figure 33 shows that the histidine residue can form only one H-bond interaction with the SAM cofactor, fewer than the interactions formed by K115. This explains why the

protection efficiency falls off earlier than in the wild-type enzyme. However, histidine offers an advantage over lysine due to its smaller size, which takes up less space in the active site, and its rigid ring structure, which may facilitate potential π-π interactions that could further stabilize the active site.

4.33 K115E Mutant M.MpeI

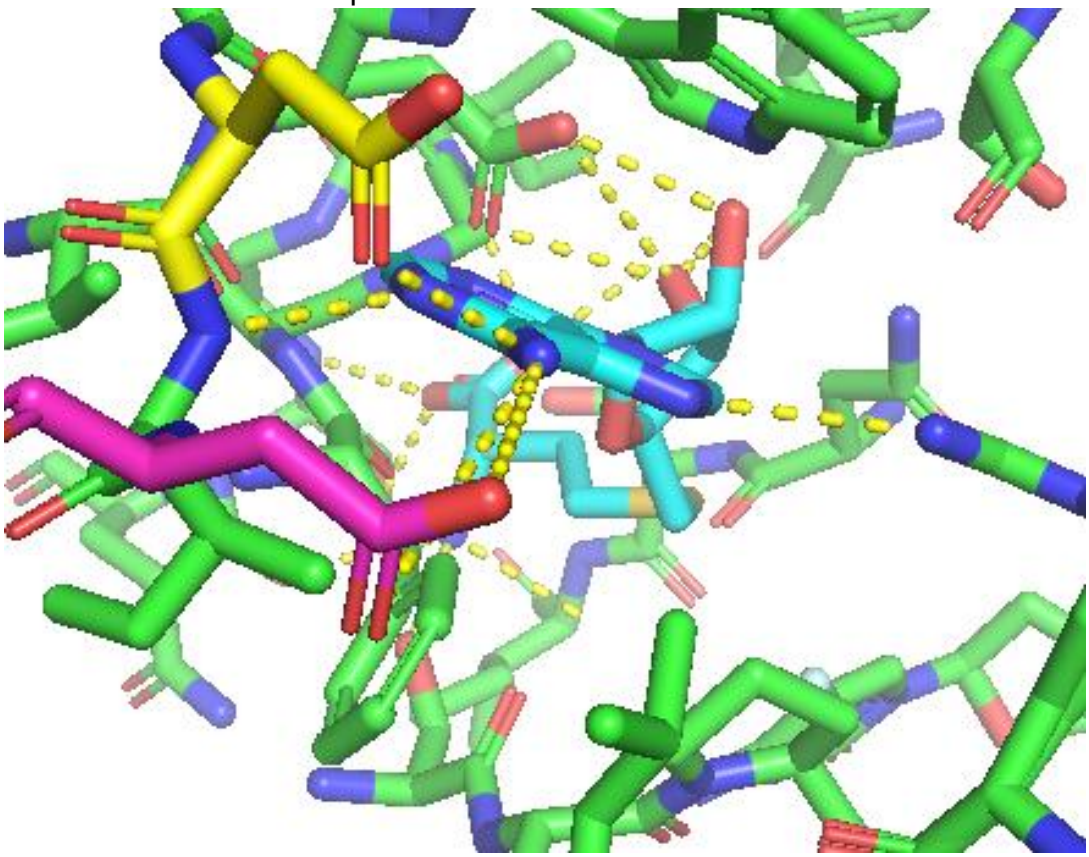


Figure 37: K115E active site, K115E is shown in purple, D113 is shown in yellow, yellow hatched lines show intermolecular/intramolecular interactions between elements. This shows glutamic acid can form two interactions using both oxygens involved in the carboxylic group to form H-bonds with the amine group of SAM (blue) PDB code 4DKJ.

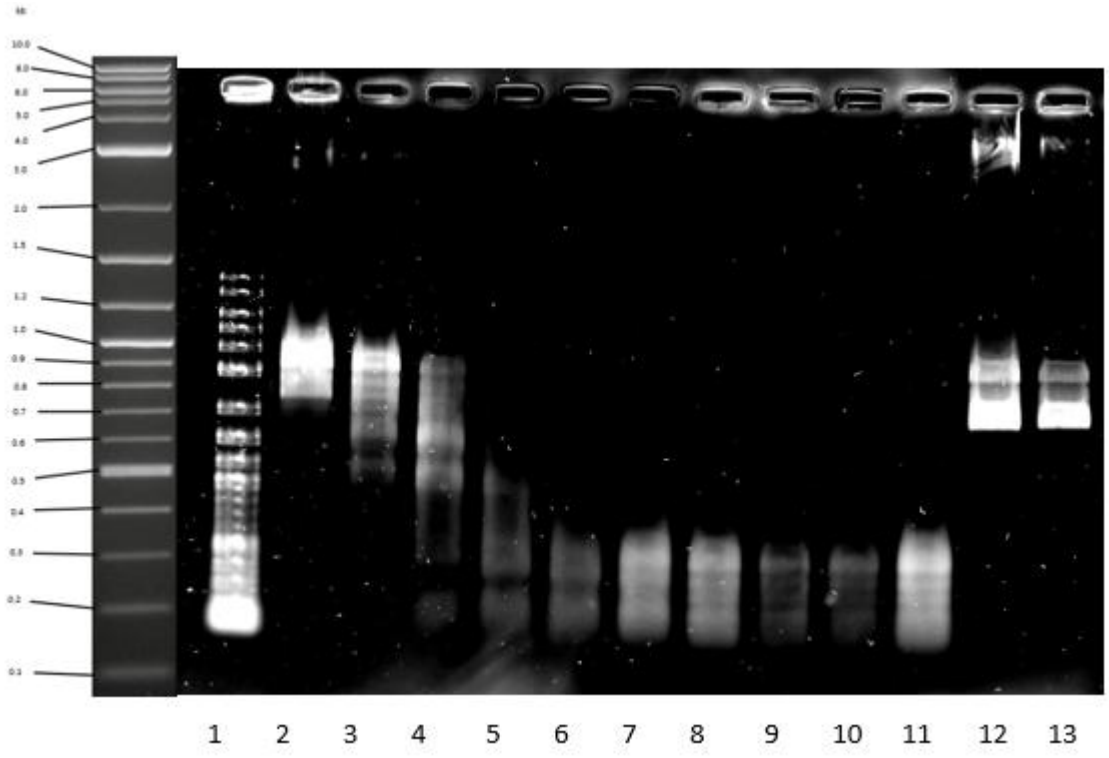


Figure 38: K115E gel (6.0 µM) results with SAM cofactor with 2-fold serial dilution of SAM (Lane 1: DNA ladder, Lanes 2-9: Serial dilution of SAM (500 – 3.9 µM), Lane 10: Control without cofactor, Lane 11: Control without cofactor & enzyme, Lane 12: SAM control, Lane 13: SAM control without restriction enzyme. EC_{50} value is 104.7 µM

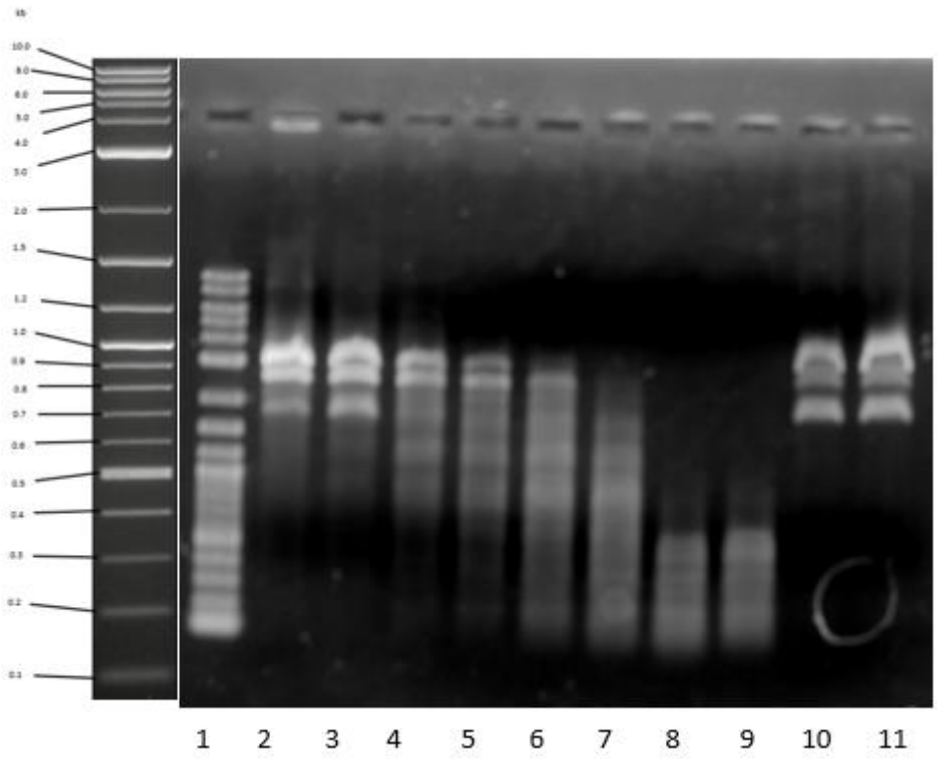


Figure 39: K115E gel results with SAM cofactor with 2-fold serial dilution of enzyme (Lane 1: DNA ladder, Lanes 2-7: Mutant E, serial dilution of enzyme (6.0 µM to 0.19 µM), Lane 8: Control without cofactor, Lane 9: Control without cofactor & enzyme, Lane 10: SAM control, Lane 11: SAM control without restriction enzyme. EC_{50} value is 1.8026 µM

Figures 38 and 39 show examples of agarose gel images used to test the catalytic activity of the K115E mutant with the native SAM cofactor. Figure 38 illustrates the dilution of the SAM cofactor, whilst Figure 39 depicts dilution of the MTase. In Figure 38, lanes 2-9 demonstrate

that the enzyme can protect DNA at SAM concentrations up to 125 µM. In Figure 39, however, shows that at lanes 4-7 at lower MTase concentrations (1.5 µM to 0.19 µM), there is no meaningful DNA protection. Control experiments in lanes 10 and 11 of Figure 38 and lanes 9 and 10 of Figure 39 confirm complete restriction of the DNA. Lane 12 (Figure 38) and lane 10 (Figure 39) show that the SAM cofactor acts as a methyl donor for the wild-type M.MpeI enzyme under these conditions. Additional results for the K115E mutant and other cofactor analogue combinations can be found in the appendix.

The K115E mutant has a side chain that is slightly shorter than the native lysine residue, and the terminal carboxyl group carries a formal negative charge, in contrast to the positive charge on the lysine residue. The ability of this mutant MTase to protect DNA from restriction was tested with all the synthesized cofactor analogues (Appendix Figures 32 to 34). The results showed that the K115E mutant enzyme was unable to protect DNA from restriction, except when used with the SAM cofactor, where it exhibited comparable activity to the wild-type M.MpeI and demonstrated good efficiency up to four serial dilutions. Figure 37 shows the active site of the K115E mutant, which reveals that it can form more stabilizing interactions than K115H. However, the K115E mutant still failed to show detectable catalytic activity when tested with the **4**, **3**, and **1** cofactors. Similar behavior was observed for the K115D mutant enzyme (Chapter 3), suggesting that the relatively short, negatively charged side chain may not be capable of forming key interactions with the cofactor analogues. The carboxyl group may also not be within the 5-angstrom range needed to form stabilizing hydrogen bonds with other residues in the enzyme. This would then lead to the disruption of critical stabilizing intramolecular interactions, preventing the proper formation of the active site required for these cofactors. Additionally, since the carboxyl group is deprotonated and negatively charged at physiological pH, it can only act as a hydrogen bond acceptor, potentially disrupting interactions with specific cofactor analogues. For example, since the **3**

cofactor is exclusively a hydrogen bond acceptor, the lone pairs on the oxygen atoms in the K115E mutant may not pair well with this cofactor.

In principle, the K115E mutant could form a stabilizing interaction with the **1** cofactor, as the amine group is protonated at physiological pH, making it exclusively a hydrogen bond donor. Yet, if the intramolecular interaction in the wild-type M.MpeI enzyme is between lysine and another residue, it is likely that this interaction occurs between the ammonium group of lysine and an acid or a residue with hydrogen bond acceptors. A prime candidate for this interaction is the D113 residue in the wild-type M.MpeI. Figure 37 highlights this D113 residue in yellow, which is near the K115E mutant. This proximity could lead to destabilization due to the presence of two negative charges in close proximity, preventing the active site from forming correctly. In contrast, the K115 residue in the wild-type enzyme would have a positive charge, allowing for stabilization, as it could then form a hydrogen bond with the D113 residue, which would act as a hydrogen bond acceptor.

HPLC traces and UV-Vis absorption spectra of the purified M.MpeI mutant enzymes can be found in the appendix (Appendix Figures 18 and 22).

4.34 K115Q Mutant M.MpeI

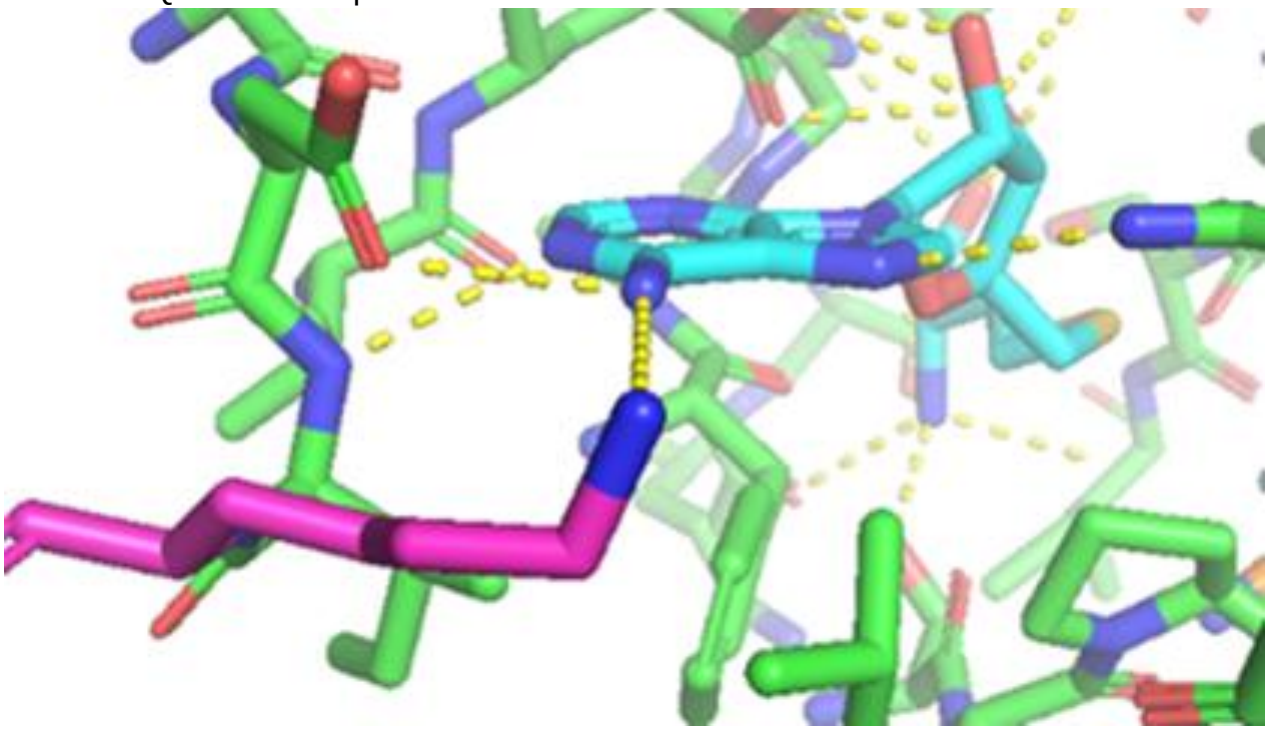


Figure 40: K115Q active site, K115Q is shown in purple, yellow hatched lines show intermolecular/intramolecular interactions between elements. This shows glutamine can form 2 interactions using both the carbonyl and the amide group to form H-bonds with the amine group of SAM (blue) PDB code 4DKJ.

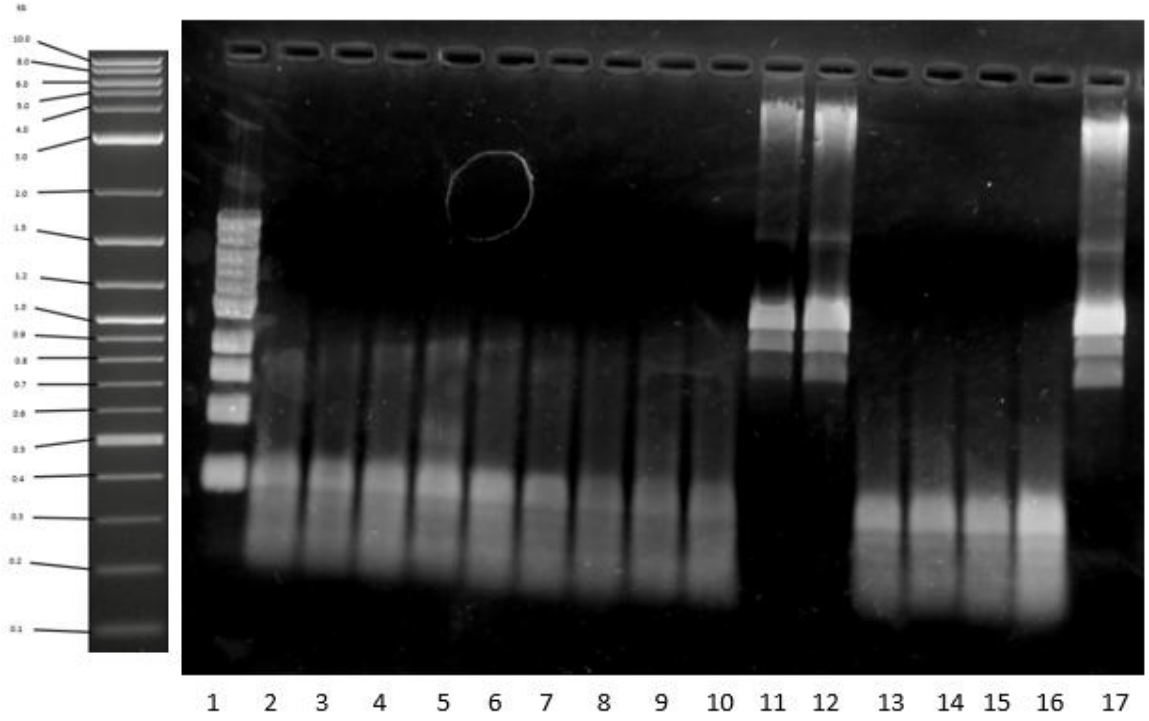


*Figure 41: Gel image of all mutant proteins with the **3** cofactor (Lane 1: DNA ladder, Lanes 2-3: Mutant A, serial 2-fold dilution of cofactor, Lane 4: Mutant A control, Lane 5-6: Mutant E, dilution of cofactor, Lane 7: Mutant E control, Lane 8-9: Mutant H, Serial dilution of cofactor, Lane 10: Mutant H control, Lane 11-13: Mutant Q, serial dilution of cofactor, Lane 14: Mutant Q control, Lane 15: Control without cofactor, Lane 16: control without cofactor or enzyme, Lane 17: SAM Control.*

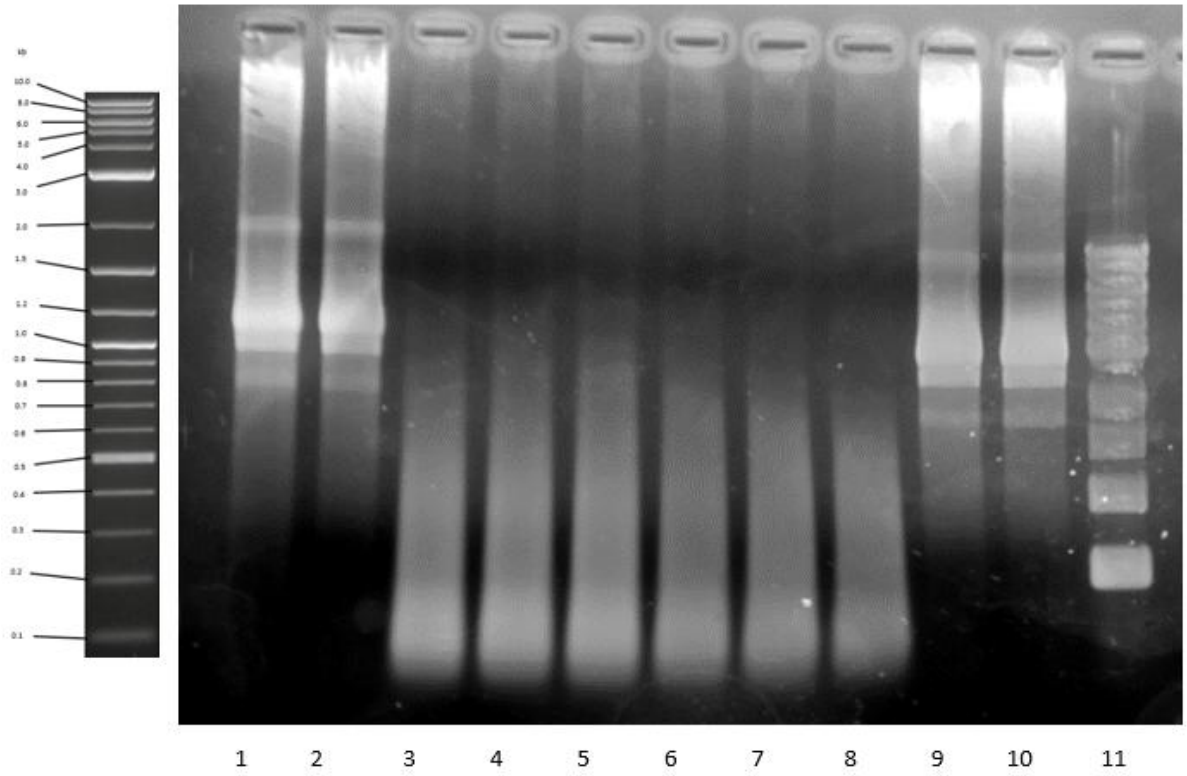


*Figure 42: Gel results of the K115Q mutant (6.0 µM) with **3** with dilution of cofactor Lane 1-5: K115Q, 2-fold serial dilution with **3** (500 µM-31.25 µM), Lane 6: Mutant Q control, Lane 7: Control without cofactor, Lane:8 Control without cofactor or enzyme, Lane 9: SAM Control, Lane 10: SAM control without restriction enzyme, Lane 11: DNA ladder.*

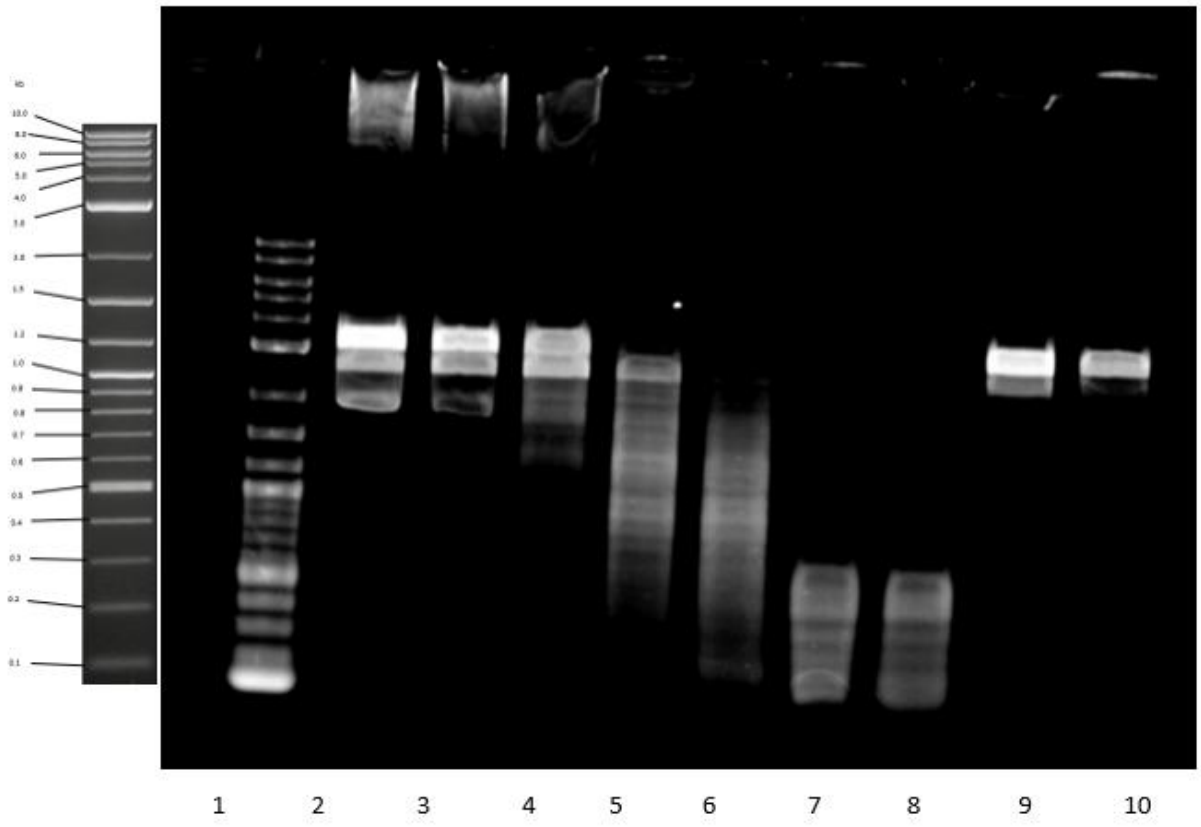


*Figure 43: Gel results of K115Q with **3**, with dilution of enzyme Lane 1: DNA ladder, Lane 2-6: K115Q, 2-fold serial dilution K115Q (6.0 µM to 0.375 µM), Lane 7: K115Q control, Lane 8: Control without cofactor or enzyme, Lane 9: SAM Control, Lane 10: SAM control without restriction enzyme.*

Figure 41 presents an example agarose gel testing the catalytic activity of the K115A, K115E, K115H, and K115Q mutants with the **3**-cofactor analogue. Lanes 2 to 13 show the single dilution of the cofactor paired with each of the K115X mutants. Lanes 2 to 10 demonstrate that K115A, K115E, and K115H do not exhibit activity on DNA with this cofactor. Control experiments in lanes 14 to 16 confirm complete DNA restriction, while lane 17 shows that the SAM cofactor acts as a methyl donor for the wild-type enzyme. Figures 42 and 43 further investigate the catalytic activity of K115Q with the **3** cofactor. Figure 42 shows the dilution of the **3** cofactor, where lanes 3 to 5 reveal no observable activity on DNA at 125

µM cofactor. Lanes 6 to 8 present control experiments confirming complete restriction of DNA, and lane 9 shows that the SAM cofactor acts as a methyl donor for the wild-type M.MpeI under these conditions. Figure 43 focuses on diluting the MTase instead of the cofactor. Lanes 2 to 4 confirm that K115Q is active down to 1.5 µM, but when diluted further to 0.375 µM, the protection efficiency decreases, as shown in lanes 5 to 6. Control experiments in lanes 7 to 8 confirm complete DNA restriction, and lane 9 shows that SAM cofactor acts as a methyl donor for the wild-type enzyme. Additional K115Q and cofactor analogue combinations can be found in the gel images in the appendix (Appendix Figures 32 to 34).

HPLC traces and UV-Vis absorption spectra for all MTases are available in the appendix (Appendix Figures 17 to 24).

4.4 Discussion

The final mutant tested was the K115Q mutant, which has a structure very similar to K115E, with the key difference being that it terminates in an amide group rather than a carboxylic acid. Compared to K115, K115Q is one carbon shorter and, unlike the positively charged lysine, is neutral at physiological pH due to the amide linkage. The amide group's nitrogen typically does not become protonated, as the amide has a lower pKa compared to a standard amine, making it slightly negative. This amide can exist in an equilibrium between a positively charged form, a neutral form, and a negatively charged form. However, at physiological pH, the neutral form predominates due to the nature of the amide's resonance structure. The K115Q mutant, like K115E, was tested with the **3**-cofactor analogue from Chapter 3, and K115Q showed compatibility with the cofactor, successfully protecting DNA from restriction digestion.

The reason for this activity, in contrast to the other mutants, could be that K115Q is one of the longest residues tested, allowing it to potentially form key intramolecular interactions with

other residues. This proximity may facilitate the formation of a more stable active site that can accommodate cofactors. For instance, there may be a key interaction between K115Q and residues such as K116 or D113. Another factor contributing to K115Q's success with the **3** cofactor could be related to the intermolecular interactions. K115E, with its negative charge, could form ion-dipole interactions that might interfere with the fluorine atom on the **3** cofactor, which is highly electronegative. In contrast, K115Q is neutral, avoiding the repulsive interactions with the cofactor, and the neutral amide group may also act as a hydrogen bond donor. The nitrogen's partial positive charge can form hydrogen bonds both with the fluorine on the cofactor and with the lone pairs on the carbonyl and amine components of the amide. This makes K115Q a viable candidate for activity with the **3**-cofactor analogue, as it avoids the potential issues highlighted with the other mutants.

4.5 Conclusion

In summary, this chapter built on the bacterial work presented in Chapter 3, particularly with the **3** cofactor and the K115Q mutant MTase. Several of the mutant MTases tested in Chapter 3 were successfully produced on a larger scale through protein overexpression. These mutant enzymes were then tested with various cofactors in protection assays to compare their ability to protect DNA against restriction digestion, relative to the wild-type M.MpeI enzyme. Most of the cofactors were unsuccessful in aiding the mutant MTases, likely due to an inability to form crucial intramolecular interactions, which may have caused the active site to collapse or shrink, preventing the accommodation of the cofactor. Alternatively, the cofactors might not have been able to form essential intermolecular interactions, such as hydrogen bonds with the MTase. For example, some cofactors, like the **3** analogue, may only act as H-bond acceptors, while certain mutant proteins terminate with H-bond acceptors, thus hindering complementary interactions necessary for successful catalysis.

One successful combination was the **3** cofactor with the K115Q mutant. The K115Q mutant is long enough to potentially form a key intramolecular interaction, and its amide group can both act as an H-bond acceptor via the lone pairs on the carbonyl and amide, and as an H-bond donor via the hydrogens attached to the amide group. This dual functionality allows for favourable interactions with the fluorine on the **3** cofactor. Further testing in serial dilution experiments revealed that K115Q displayed activity comparable to the wild-type M.MpeI enzyme, although it failed after three serial dilutions and showed diminished DNA protection after two serial dilutions.

In conclusion, each mutant MTase was produced at significantly higher concentrations and successfully purified using the protein purification system, allowing for storage until required for protection assays. The results consistently demonstrated that only one cofactor from Chapter 2, the **3** cofactor, was compatible with the K115Q mutant MTase. This compatibility is likely due to a combination of factors, including both intramolecular and intermolecular interactions. Key hydrogen bonds likely formed to ensure the active site is sufficiently large for the cofactor to fit, whilst crucial intermolecular interactions between the cofactor and MTase helped stabilize the enzyme-substrate complex.

From these findings, it can be concluded that the K115 residue is essential for the proper function of M.MpeI. The activity of the enzyme is significantly impacted when this residue is mutated, as seen in the K115A mutation. This results in a complete loss of activity, particularly in protection assays with SAM. On the other hand, the K115H and K115E mutants do maintain some level of activity at high cofactor concentrations, with comparable activity to the wild type at the highest concentrations. The K115A mutant, however, shows complete digestion of DNA even at the highest cofactor concentrations, further emphasizing the functional importance of the K115 residue.

This is likely due to the stabilizing electrostatic interactions that lysine can form with other amino acids and the cofactor. The K115A mutant cannot form meaningful electrostatic interactions because alanine lacks the polar groups necessary for such interactions. w

Whereas K115E and K115H can form some electrostatic interactions, they can only act as hydrogen bond donors or acceptors, not both. In contrast, the K115Q mutant, with its amide and carbonyl groups, can participate in both hydrogen bond donation (via the hydrogens attached to the amide) and acceptance (via the lone pairs on the carbonyl and amide), hence providing a more versatile interaction profile. K115E can only act as a hydrogen bond acceptor due to its lone pairs, whilst K115H, despite having polar hydrogens available for donation, has lone pairs involved in aromaticity, limiting its ability to form hydrogen bonds.

The main areas for improvement in this work are centred around the protein purification system, which presented several limitations during the experiments. One major issue was the interference observed in the UV-Vis absorbance readings from the protein purification system. Both MTase and imidazole absorb at the 260 nm wavelength, making it difficult to accurately determine when the MTase eluted. It would be a significant improvement to first establish a baseline for the imidazole gradient, thus allowing for the absorbance from imidazole to be excluded and making it easier to observe the MTase elution.

Another limitation, particularly evident with the K115A mutant, lies in the uncertainty of whether the cofactors are incompatible with the mutant or if the K115 residue itself is crucial for maintaining the enzyme's structure. Whilst the protein was successfully generated, it's possible that replacing lysine with alanine caused misfolding, which may have resulted in an altered active site, potentially too small to accommodate cofactors. This could explain the failure in protection assays. Future research, such as enzyme crystallization, would help clarify this. Structural analysis would also provide insights into whether K115 is essential for cofactor binding or contributes to the overall structural integrity of the enzyme.

Additionally, another potential cause for the failure to protect DNA could be that while the cofactor may bind to the enzyme, the resulting enzyme-substrate complex could be more stable than the MTase alone. In such a case, the energy required for the enzyme to turnover and protect DNA may be too high, effectively rendering the cofactor as an inhibitor. Crystallizing the MTase in the presence of the cofactor would reveal whether the cofactor binds to the active site. If it does, energy calculations could then be conducted to compare the stability and energy values of the MTase alone versus the enzyme-substrate complex. This would help determine if the cofactor inhibits the enzyme's function.

Such findings could have broader implications, particularly in the context of diseases caused by bacteria or viruses that produce their own enzymes. Introducing synthetic cofactors with higher binding affinity to these disease-causing enzymes could inhibit their function. If the enzyme plays a critical role in the disease progression, its inhibition could reduce the severity or even cure the disease by halting its spread.

Chapter 5: Summary of Results and Future Work

5.1 Results

Chapter **2** focused on investigating the effects of cofactor analogues on the activity of the M.MpeI methyltransferase (MTase) on DNA. This exploration was prompted by the observation that two very similar cofactors, **2** and **1**, led to markedly different enzyme activities on DNA. The only distinction between these cofactors was the terminal group attached to the adenosine: an alcohol group in cofactor **2** and an amine group in cofactor **1**.

To understand the basis for these differences, new analogues of cofactor **2** (**4** and **3**) were synthesized. These analogues were chosen based on previous docking studies, which indicated that their energy profiles were comparable to the natural cofactor when bound to the M.MpeI MTase. The analogues were synthesized using the same method employed by the Neely group for cofactor **2**.

After successful synthesis, the analogues were tested in a protection assay using the M.TaqI enzyme, which had a less restrictive active site compared to M.MpeI. Both cofactor analogues (**4** and **3**) were found to enable MTase activity on DNA when paired with M.TaqI. When tested with M.MpeI, however, neither cofactor analogue demonstrated activity, suggesting that the difference in activity may be related to the specific intermolecular interactions between the cofactor and the enzyme. In particular, the role of the 115-lysine residue in M.MpeI was hypothesized to be a key factor in this differential activity.

Chapter 3 explored the effects of mutating the 115-lysine residue in M.MpeI to different amino acid residues. The primary focus was on residues that could alter the size of the active site as well as the charge of the residue. It was hypothesized that smaller residues would increase the size of the active site, allowing it to accommodate larger cofactors. Additionally,

it was hypothesized that changing the charge of the residue would help mitigate potential clashes between the cofactor and the enzyme. Given that the lysine residue is cationic at pH 7, this could explain why cofactor **1** fails to bind, as both cofactor 1 and the lysine residue carry a positive charge, potentially causing repulsion and destabilizing the enzyme-substrate complex.

To test this hypothesis, a bacterial in-vitro transcription-translation method was used, where plasmids encoding M.MpeI with the 115-lysine residue mutated to other amino acids were generated. These mutant MTases were tested with various cofactors in a protection assay. If bacterial colonies did grow, it would indicate that the MTase and cofactor were compatible.

The results showed that most of the enzyme mutants failed to protect DNA, with no colonies present. However, a few mutants showed promising results. This suggests compatibility with certain cofactors. Notably, the majority of the MTase mutants worked with cofactor **2**, whilst the K115Q mutant exhibited activity with cofactor **3**, but no activity with other cofactors, including the natural SAM.

Chapter 4 built upon the binary results of Chapter 3, where conclusions could only be made about the compatibility of the enzyme and cofactor. However, no information was gathered regarding the minimum concentration of either the cofactor or the MTase needed for activity. The aim of this chapter was to address this gap.

To achieve this, several MTase mutants from Chapter 3 were produced at much higher concentrations than the in-vitro transcription-translation method could generate. A more traditional ion-exchange chromatography method was employed to produce larger quantities of these enzymes. This allowed for the use of protection assays, which required significantly higher enzyme and cofactor concentrations than what was achievable in Chapter 3.

Through these protection assays, both the concentration of cofactor and enzyme could be controlled. This provides insight into the minimum concentrations required for activity. The

results showed that the K115A mutant was completely incompatible with any cofactor, leading to total inactivation of the MTase in the presence of the cofactors. In contrast, K115H and K115E mutants showed improved activity with some SAM cofactors, and their minimum required concentrations for activity could be determined. K115Q exhibited results consistent with Chapter 3, where it was inactive with the SAM cofactor, but showed some protection when paired with cofactor **3**. However, it did not perform well at low concentrations, as seen with K115H and K115E.

Ultimately, a compatible mutant MTase and cofactor analogue was identified, which was incompatible with the natural SAM cofactor. This discovery opens up new avenues for further research, particularly in exploring alternative cofactors that may exhibit distinct interactions with mutant enzymes.

Overall, this thesis has demonstrated the power of combining docking studies, cofactor analogues, and targeted mutation of key residues in the active site to provide a deeper understanding of the importance of specific amino acids. By altering these residues, it is possible to significantly alter the activity profile of the MTase across different cofactors. These insights have the potential to be applied to a wide range of MTases, including human MTases, which could pave the way for advances in diagnostics and therapeutics.

5.2 Future Work

Future work in this field could involve investigating additional functional groups on the cofactor, such as carboxylic acids, thiols, or other chemical groups. By exploring how these modifications affect the interaction between the cofactor and MTase, this could provide further insights into how structural changes impact the enzyme's activity on DNA. Additionally, these studies could uncover novel MTase mutants and cofactor combinations.

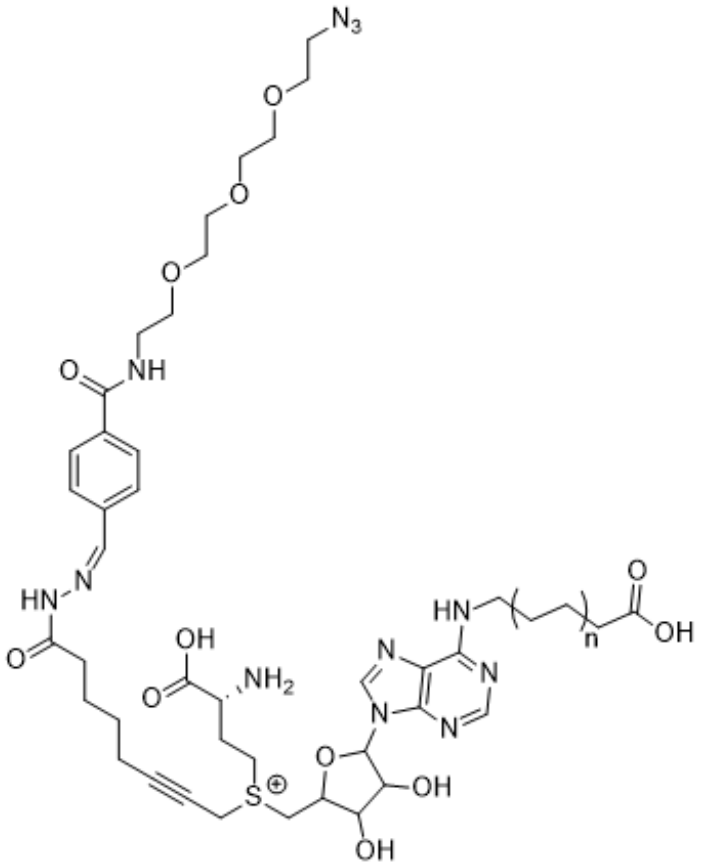


Figure 44: Structure of a potential new cofactor molecule that may have improved binding with K115H where n is between 3 and 10.

The K115H mutant, for example, could be paired with a larger carboxylic acid cofactor which could be promising. The positive charge on the histidine residue could form strong intermolecular interactions with the negative charge on the carboxylic acid, potentially improving activity.

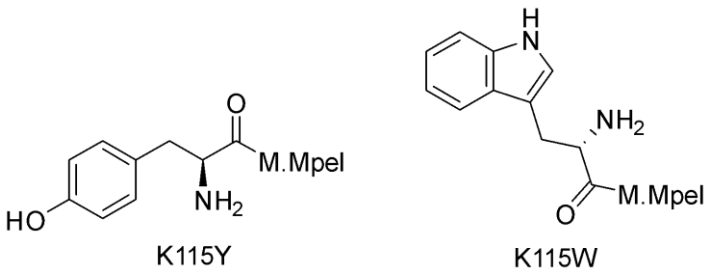


Figure 45: Structures of two other mutations that could be made to K115 to test aromatic interactions.

Other mutations, such as K115W and K115Y, could also be explored. These larger, ringed residues may offer new opportunities for interactions, such as π-π stacking, which could further enhance enzyme-cofactor compatibility and broaden the scope of MTase activity on DNA.

The mutant MTases from Chapter 4 could also be structurally characterised using techniques such as X-ray crystallography or electron microscopy. Determining the structure of these mutants would provide valuable insights into how the cofactor binds to the enzyme, allowing for energy calculations of the enzyme-substrate complex. These calculations could help us determine whether the cofactor is acting as an inhibitor, if it is incompatible with the MTase, or if the mutation has caused any alterations to the active site, potentially distorting its ability to bind the cofactor effectively. With more time and resources, this work could be expanded to explore additional key residues within the active site, further enhancing our understanding of their role in enzyme function.

These findings could also be extended to other MTases, such as DNMTs, or even enzymes involved in human epigenetic diseases. This could aid in diagnosing diseases by confirming enzyme activity or dysfunction. If the therapeutic process involved enzyme-cofactor interactions, this knowledge could then be used to design cofactors or enzymes that enhance therapeutic efficacy. Moreover, for diseases caused by aberrant enzyme-cofactor interactions, engineered mutant enzymes could be created to prevent the natural enzyme from binding its cofactor, potentially halting the disease process.

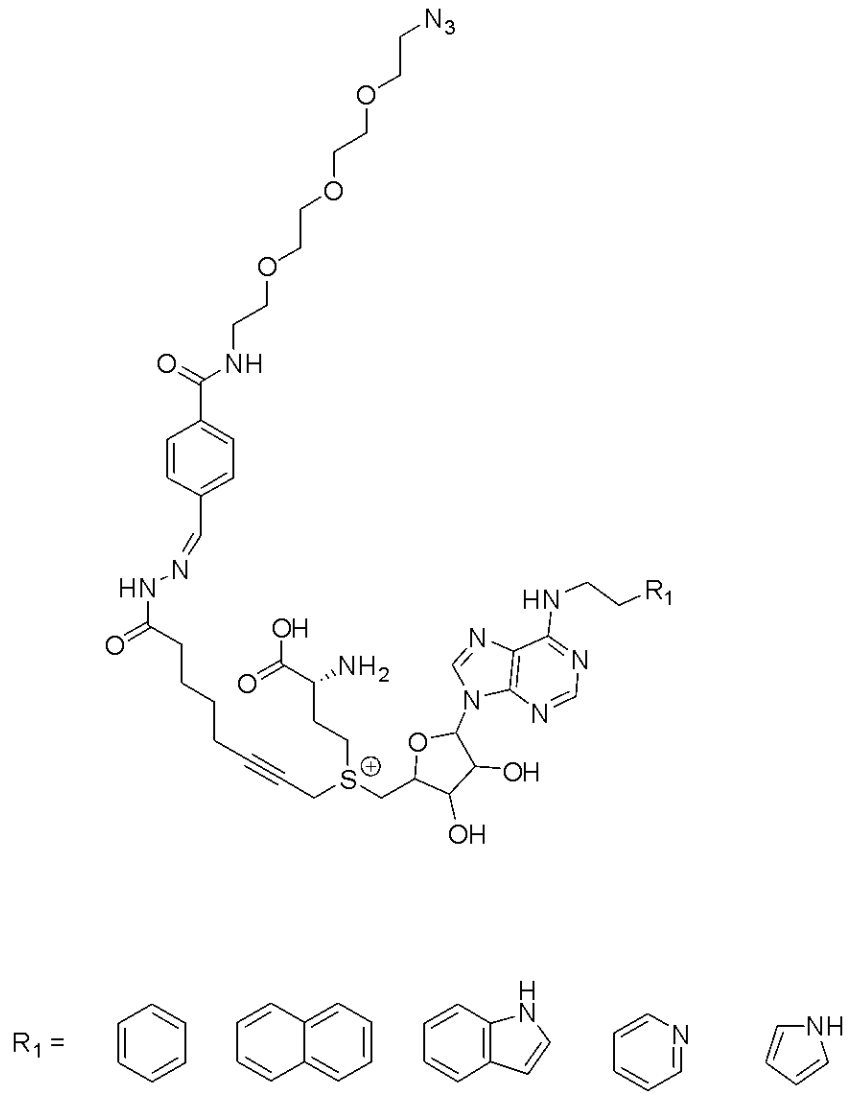


Figure 46: Structure of a new cofactor analogue that would focus on aromatic interactions and what possible groups could be attached.

Another interesting avenue for exploration involves designing cofactor analogues compatible with the tyrosine or tryptophan residues from Figure 45. This could be achieved by creating cofactors that rely on π-stacking interactions rather than traditional H-bonding. For instance, an analogue could be developed with an aromatic structure, such as a benzene ring, naphthalene, indole, or pyridine, at the terminus of the alkyl arm (like cofactors **1** or **2**). These aromatic analogues could benefit from π-stacking interactions with tyrosine or tryptophan

residues positioned above or below the cofactor's adenosine ring. To further enhance binding, additional groups could be added to the aromatic ring (e.g., carbonyl or alcohol groups) to enable H-bond formation alongside π-stacking. However, face-to-face π-stacking may be preferred over edge-to-face stacking, as the latter could occupy too much space in the enzyme's active site, preventing proper cofactor binding. Such approaches could lead to the development of cofactors with a wider range of chemical functionalities, enabling more diverse and potent DNA modifications.

The specificity of these mutant MTases could also be evaluated to determine how selective they are for cofactor analogues, including natural SAM. This could be expanded to investigate their performance within living cells, assessing whether these mutants maintain the same level of activity or if their efficiency decreases in the cellular environment. Additionally, it would be important to explore whether the cell recognizes and rejects these modified enzymes or if they are accepted and function effectively. This research could be further developed to optimize these mutant MTases, thus enhancing their ability to selectively scavenge specific cofactor analogues within cells.

Materials and Methods

6.1 General Experimental:

6.11 Reagents/Materials

Other reagents were bought from Alfa Aesar, Fisher Scientific, Sigma Aldrich and BroadPharm and used without further purification. Bio-reagents such as enzymes and DBCO-biotin analogues were purchased and used accordingly from New England Biolabs (NEB), Jena Bioscience and Thermo Fisher. LB Broth with/without agar was purchased from Sigma-Aldrich and agarose was purchased from sigma and used as is. T7 express cells for the purpose of transformation and protein expression kits and were bought from NEB. Cutsmart buffer has a composition of 50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate, 100 µg/ml BSA, and pH 7.9 at 25 °C. Tango buffer has a composition of 33 mM Tris-acetate, 10 mM magnesium acetate, 66 mM potassium acetate, 0.1 mg/mL BSA and pH 7.9 at 37°C.

All reagents and compounds were handled using appropriate PPE. Additional safety precautions were taken when handling *n*-BuLi. The solution was only dispensed when the reagent had a flow of argon available to the bottle, there was a source of MeOH nearby to quench the needle and syringe after dispensing the reagent. HMPA is a carcinogen and mutagen, extreme caution must be exercised when handling this reagent.

2D NMR spectra, such as HSQC and COSY, were used to assist in assigning the 1D NMR peaks. COSY helped identify which protons are coupled to each other, typically two or three bonds apart. Meanwhile, HSQC was used to correlate proton signals with their corresponding carbon signals, allowing for the complete structure to be determined.

6.12 Purification/Synthetic techniques

Purification was done by performing flash column chromatography using silica gel as the stationary phase. DCM and THF were dried over 3Å molecular sieves. Analytical RP-HPLC was performed on Shimadzu LC-20 Prominence equipped with ACE 5 C18 (250 x 4.6 mm,

flow rate 1 mL/min) with UV channels set to 220, 240, 260 and 280 nm for the wavelength. Preparative RP-HPLC was performed on Agilent Technologies 1260 Infinity equipped with ACE 5 C18 (250 x 21.2 mm, 100 A, flow rate 10 mL/min).

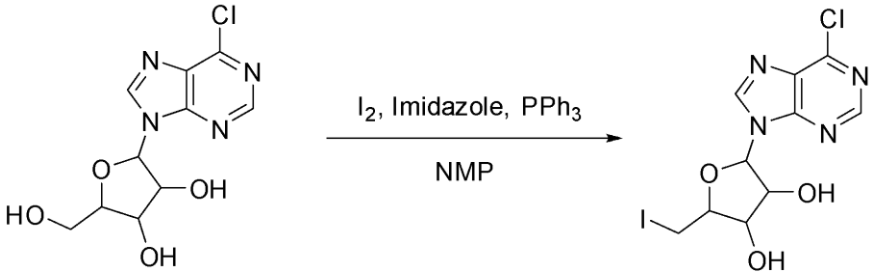
Cooling bath systems that were used were -78 °C (dry ice/acetone) and -196 °C (liquid N_2). Reactions carried out in non-aqueous solvents were done under an argon atmosphere using oven-dried glassware. Solvents were degassed by a needle submerged in the solvent whilst having argon bubble through for 10 mins. Sparging was achieved by blowing argon over the crude mixture for 15 mins.

6.12 Analytical Specifications

^{1}H NMR spectroscopy data was acquired using a Bruker AVII 300 spectrometer. ^{13}C NMR spectroscopy data was acquired using a Bruker AVII 400 spectrometer. Spectra was recorded in deuterochloroform, dimethyl sulfoxide d_6, or deuterated methanol and referenced to $CDCl_3$ (^{1}H, 7.26 ppm) or $CDCl_3$ (^{13}C, 77.2 ppm), $(CD_3)_2SO$ (^{1}H, 2.50 ppm) or $(CD_3)_2SO$ (^{13}C, 39.52 ppm), CD_3OD(^{1}H, 3.31 ppm) or CD_3OD (^{13}C, 49.00 ppm). Chemical shifts were given in ppm and the coupling constants were given in Hz. Abbreviations used for the multiplicity are m- multiplet, s-singlet, d-doublet, t-triplet, q-quartet, and br-broad. Mass spectra were recorded on Waters (LCT) spectrometer and were reported in terms of (m/z (%)).T.L.C. was done on silica plates eluents and RFs were found for each compound where appropriate and UV radiation and $KMnO_4$ was used to visualise compounds.

6.3 Synthetic Techniques

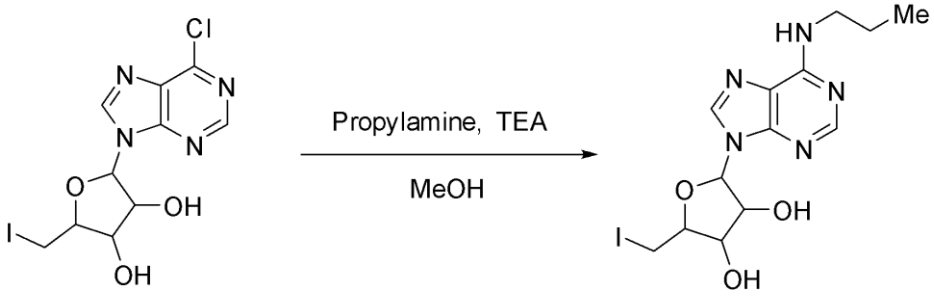
6.31 Synthesis of 2-(6-chloro-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol



A solution of 6-chloropurine riboside (Sigma 500 mg, 1.75 mmol) was made up in *N*-methylpyrrolidine (5 mL) on ice. To this solution imidazole (773.5 mg, 11.4 mmol) and triphenylphosphine (1513 mg, 5.77 mmol) were added to the stirred solution and allowed to dissolve. Iodine (1464 mg, 5.77 mmol) was dissolved in a small amount of *N*-methylpyrrolidine and added to the solution dropwise, waiting for the colour to disperse before adding the next drop. The reaction was then left overnight and worked up with H_2O afterwards. The organic and aqueous layers were both washed with sodium thiosulphate and the organic layer was collected and solvent was removed under reduced pressure. The crude product was then dissolved in a minimum amount of methanol and placed in the fridge overnight. 2-(6-chloro-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol was then purified by prep-HPLC giving retention times of 6.5, 12 and 22 mins. Fractions were combined and had methanol removed via reduced pressure and were freeze dried to give the 2-(6-chloro-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol as a white powder. (198 mg, 50%); Rf = 0.60 (DCM:MeOH 9:1); 1H-NMR (400 MHz, MeO-D) δ 8.26 (s, 1H, aromatic CH), 8.21 (s, 1H, aromatic CH), 6.34 (d, *J* = 12.1 Hz, 1H, OCHN), 4.28 – 4.24 (m, 1H, C*H*CH_2I), 4.05 – 4.00 (m, 1H, C*H*OHCHON), 3.64 – 3.45(m, 1H, CH_2IC*H*CHOHCHOH), 2.34 – 1.97 (m, 2H, IC*H*$_2$CH); HRMS-ESI(+): calculated for $C_{10}H_{11}N_4O_3ICl$: 395.9486, found: 396.9381 [M+H].

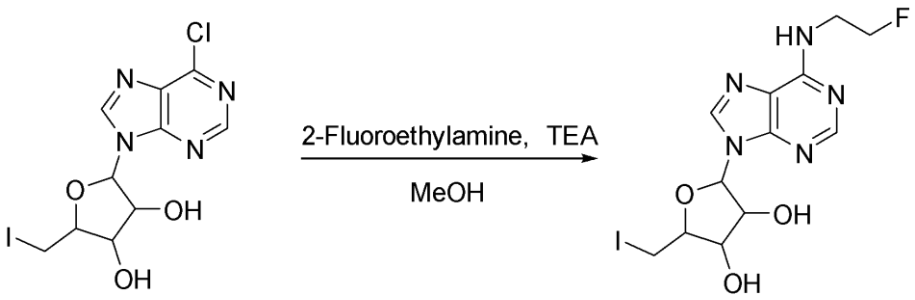
Compound made in patent WO2023047141A, mass spec data in agreement with this patent, NMR spectrum data is not comparable due to use of different deuterated solvent.

6.32a Synthesis of 2-(iodomethyl)-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-3,4-diol



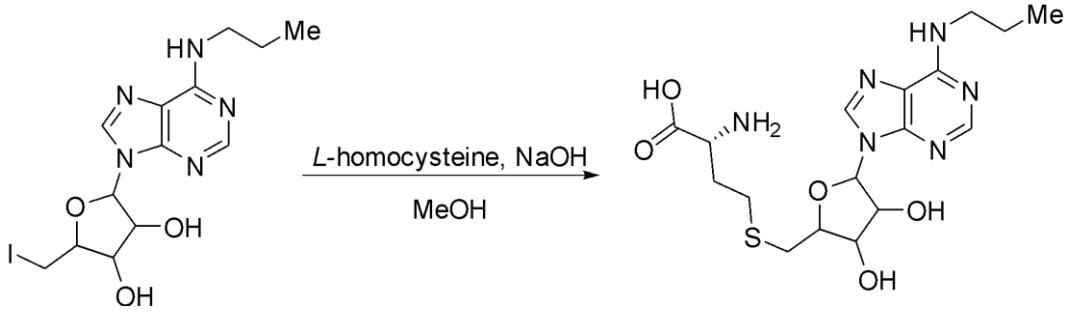
A solution of 2-(6-chloro-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol (50 mg, 0.126 mmol) was made up in MeOH (1 mL) and stirred. To this mixture propylamine (0.09 mL, 0.760 mmol) was added along with TEA (0.105 mL, 0.753 mmol). This was allowed to react overnight at RT. After which the reaction was evaporated and diluted in a mixture of MeOH (3%) and H_2O. This crude product was purified using prep-HPLC, the final product was evaporated, and the remaining solution was lyophilised. This gave the product as a white product. (31.1 mg 52%) Retention times: 44 mins ^{1}H NMR spectroscopy (400 MHz MeO-D) δ 8.26 (s) (1H Aromatic CH), 8.24 (s) (1H Aromatic CH), 6.01 (d) *J*(5.3) (1H OC*H*N), 4.86 (t) *J*(5.5) (1H NCC*H*OH), 4.30 (dd) *J*(5.4, 4.1) (1H $CH_2CHCHOH$) 4.07 (td) *J*(5.8, 4.1) (1H ICH_2CH) 3.63 (dd) *J*(10.7, 5.8) (2H ICH_2) , 3.51 (dd) *J*(10.7, 5.8) (2H $NHCH_2CH_2CH_3$), 1.71 (h) *J*(7.4) (2H $NHCH_2CH_2CH_3$), 1.02 (t) *J*(7.4) (3H CH_3) ^{13}C NMR spectroscopy – 171.7, (Aromatic C) 165.3 (Aromatic C), 164.7, (Aromatic C) 152.29 (Aromatic CH), 139.6 (Aromatic CH) 87.5 (OCHN), 73.2 (NCCHOH), 71.9 (ICH_2CH), 42.1 ($NHCH_2CH_2CH$), 21.2 ($NHCH_2CH_2CH_3$), 9.7 (CH_3), HRMS-ESI(+): calculated for $C_{13}H_{18}N_5O_3I$: 419.0454, found 420.0533 [M + H]

6.32b Synthesis of 2-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol:



A solution of 2-(6-chloro-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol (50 mg, 0.126 mmol) was made up in MeOH (2 mL) and stirred; to this mixture 2-fluoroethylamine (72 mg, 0.726 mmol) was added along with triethylamine (0.202 mL, 0.753 mmol). This was allowed to react overnight at RT to allow reaction to go to completion. After the elapsed time the reaction was evaporated and diluted in a mixture of MeoH (3%) and H_2O. The crude mixture was then purified on HPLC. The resulting product was evaporated and then lypophilised resulting in a white powder. Retention time 34 mins (37 mg 69%) NMR ^{1}H NMR spectroscopy – 8.28 (s) (2H aromatic CH), 6.02 (d) *J*(5.3) (1H OC*H*N) , 4.69 (t) *J*(5.0) (1H NCC*H*OH), 4.57 (m) (1H $NHCH_2CH_2F$), 4.31 (dd) *J*(5.4, 4.1) (1H $CH_2CHCHOH$), 4.07 (td) *J*(5.8, 4.1) (1H ICH_2CH), 3.63 (dd) *J*(10.7, 5.8) ((2H ICH_2), 3.51 (dd) *J*(10.8, 5.8) (2H, $NHCH_2CH_2F$) ^{13}C NMR spectroscopy – 152.50 (Aromatic CH), 139.81 (Aromatic CH), 88.77 (OCHN), 82.59 ($NHCH_2CH_2F$), 80.97 ($ICH_2CHCHOH$), 73.19 (NCCHOH), 47.91($NHCH_2CH_2F$), 28.78 (ICH_2CH), HRMS-ESI(+):calculated for $C_{12}H_{15}N_5O_3IF$: 423.0204, found 424.0282 [M + H]

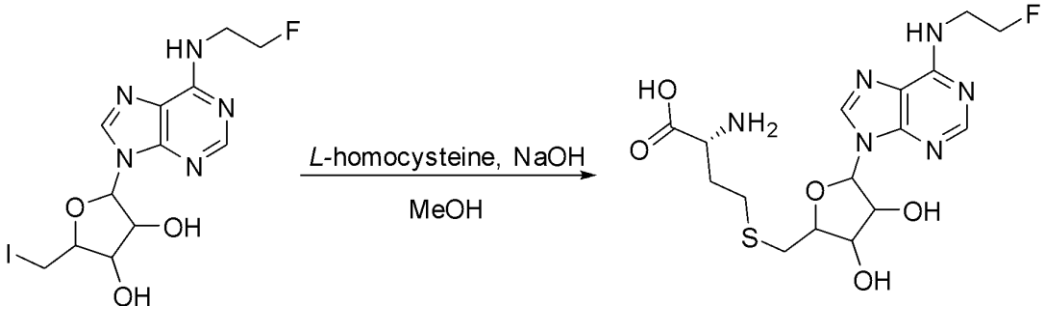
6.33a Synthesis of S-((3,4-dihydroxy-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)-D-homocysteine



A solution of 2-(iodomethyl)-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-3,4-diol (31 mg 0.0742 mmol) was dissolved in MeOH (111 µL) and this was sparged using argon. L-homocysteine (20.1 mg 0.148 mmol) was dissolved in 2M NaOH (111 µL) and this was sparged with argon. These solutions were mixed and then the lost Ar was replaced, the solution was then allowed to mix at 95 °C. The reaction was monitored by analytical HPLC. When the reaction was finished (after about 1-2 hours) the solution was worked up by adding HCl until the solution was at pH 5. The solution was made up in MeOH (3%) and H_2O and the crude mixture was purified on prep-HPLC (retention time 46 mins). The product was then evaporated, and the remaining solution was lyophilised, resulting in a white solid (22 mg, 70%). ^{1}H-NMR - 7.93 (s) (1H Aromatic CH), 7.89 (s) (1H Aromatic CH), 5.58 (d) *J*(4.7) (OC*H*N), 4.29 (t) *J*(5.0) (NCC*H*OH), 3.90 (t) *J*(5.1) (CHC*H*OH), 3.79 (dt) *J*(6.8, 4.9) (OC*H*CH_2), 3.58 (t) *J*(6.4) (C*H*NH_2COOH), 3.00 (dd) *J*(14.4, 7.5) (NHC*H*$_2CH_2$), 2.49 (qd) *J*(14.2, 5.9) (CHC*H*$_2$S), 2.21 (t) *J*(7.5) (SC*H*$_2CH_2$), 1.68 (m), ($NH_2CHCH_2CH_2$), 1.23 (q) *J*(7.3) ($CH_2CH_2CH_3$), 0.45 (t) *J*(7.4) (CH_2CH_3) ^{13}C NMR spectroscopy – 171.74 (aromatic C), 148.13 (aromatic C), 146.24 (aromatic CH), 144.15 (aromatic CH), 88.12 (O*C*HN), 83.14 (O*C*HCH_2), 73.31 (NC*C*HOH), 71.87 (CH*C*HOH), 51.86 (*C*HNH_2COOH), 43.45 (NH*C*H_2CH_2), 32.84

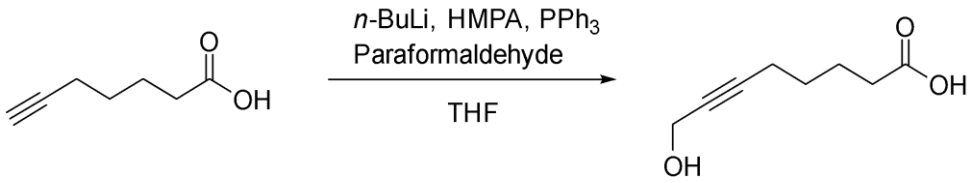
($CHCH_2S$), 29.35 ($NH_2CHCH_2CH_2$), 26.97 (SCH_2CH_2), 20.51 (CH_2CH_3), 9.87 (CH_3); HRMS-ESI(+):calculated for $C_{17}H_{26}N_6O_5S$: 426.1685, found 427.1764 [M + H]

6.33b Synthesis of S-((5-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-D-homocysteine



A solution of 2-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol (26.0 mg 0.0600 mmol) was dissolved in MeOH (91.3 µL), and this was sparged for a period. A solution of L-homocysteine (16.5 mg 0.122 mmol) was dissolved in 2M NaOH (91.3 µL) and this was also sparged for a period. These solutions were then mixed, replacing the lost argon, and the solution was then mixed at 95 °C. The reaction was monitored with analytical HPLC. When the reaction was finished (after 1-2 hours) the solution was worked up by adding HCl until the solution was at pH 5. The solution was made up in MeOH (3%) and H_2O and the crude mixture was purified on prep-HPLC (retention time 23 mins) the product was then evaporated, and the remaining solution was lyophilised. This gave a white solid. (20.0 mg 75%) ^{1}H-NMR 8.34 (s), 8.28 (2H Aromatic CH), 5.98 (d) *J*(4.7) (OC*H*N), 4.68 (t) *J*(5.0) (CH_2F), 4.56 (m) ($CH_2CHCHOH$, $CH_2CHCHOH$), 4.27 (t) *J*(6.8, 4.9) (HOC*H*CHN), 4.17 (q) *J*(6.4) (FCH_2CH_2NH), 3.97 (t) *J*(14.4, 7.5), (CH_2NH_2), 2.87 (qd) *J*(14.2, 5.9) ($CHCH_2SCH_2CH_2$), 2.58 (t) *J*(7.5) ($NH_2CH_2CH_2$), HRMS-ESI(+):calculated for $C_{16}H_{23}N_6O_5FS$: 430.1435, found 431.1513 [M + H]

6.34 Synthesis of 8-hydroxyoct-6-ynoic acid:
A known compound prepared according to literature procedure found in reference [155]

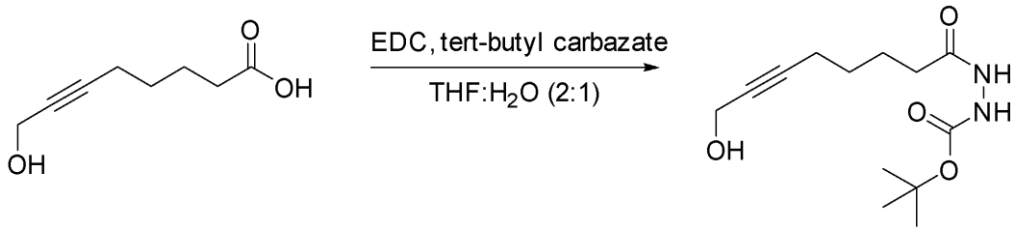


Additional safety note: HMPA is a carcinogen and mutagen; extreme caution must be exercised when handling this reagent.

A solution of 6-heptynoic acid (2.10 mL, 15.9 mmol) was made in dry THF (42 mL) under argon, to this HMPA (6.13 mL, 34.9 mmol) was added and the solution was cooled to -78 °C. To this n-BuLi (21.8 mL, 34.9 mmol) was added dropwise whilst keeping the temperature below -60 °C. The solution was then warmed to -40 °C and stirred for 1 hour. After 1 hour paraformaldehyde (1.47 g, 47.6 mmol) was added through a funnel under an argon flow. The reaction mixture was then warmed to 45 °C for 4 hours. After reaction, the mixture was quenched with 1 M HCl to pH 4-5 and extracted with EtOAc (3 x 20 mL). The organic fraction was then dried over Na_2SO_4 and the EtOAc was removed by rotary evaporation giving the crude product. Purification was completed using flash column chromatography eluent conditions: hexane:EtOAc (8:2) to (6:4) yielding the product as a colourless oil (1.4 g, 58%); Rf = 0.27 (Hex:EtOAc, 6:4); ^{1}H-NMR (400 MHz, DMSO-d_6) δ 5.07 (s, 1H, O*H*), 4.01 (s, 2H, C*H*$_2$OH), 2.22 – 2.18 (m, 4H, C*H*$_2$CO/C), 1.61 – 1.51 (m, 2H, C*H*$_2$CH$_2$COOH), 1.48 –1.37

(m, 2H,CH_2CH$_2$C); ^{13}C-NMR (101 MHz, DMSO) δ 174.6 (*C*OOH), 84.0 (CH$_2$*C*≡CCH$_2$OH), 80.6 (CH$_2$C≡*C*CH$_2$OH), 49.3 (*C*H$_2$OH), 33.3 (*C*H$_2$COOH), 27.8 (C≡CCH$_2$*C*H$_2$), 23.9 (COOHCH$_2$*C*H$_2$), 17.9 (*C*H$_2$C≡CCH$_2$OH); HRMS-ESI(-): calculated for $C_8H_{11}O_3$: 155.0708, found 155.0713 [M-H]

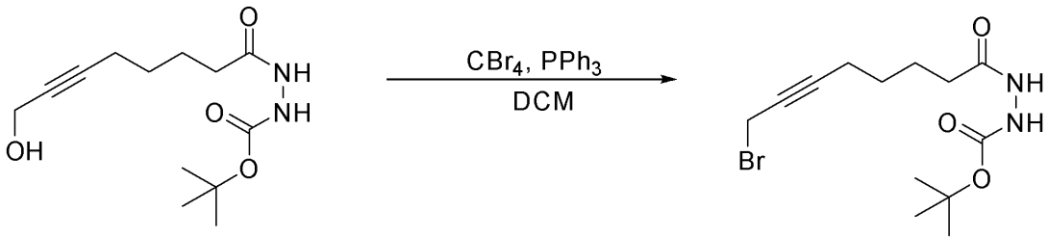
6.35 Synthesis of tert-butyl 2-(8-hydroxyoct-6-ynoyl)hydrazine-1-carboxylate:
A known compound prepared according to literature procedure found in reference [155]



8-Hydroxyoct-6-ynoic acid (1.4 g, 9.21 mmol) and tert-butyl carbazate (1.46 g, 11.1 mmol) were dissolved in 2:1 THF:H$_2$O (13.5:6.75 mL). To this EDC.HCl (1.94 g, 10.1 mmol) was added slowly over 15 minutes. The mixture was then left to stir for 3 hours and then extracted with EtOAc (3 x 20 mL). The organic layer was washed with 0.1 M HCl, water and brine. The organic layer was then collected, dried over Na_2SO_4 and the solvent was removed under reduced pressure yielding the product as a clear colourless oil (1.50 g, 60%); Rf = 0.14 (Hex:EtOAc 1:1); ^{1}H-NMR (400 MHz, $CDCl_3$) δ 8.33 (s, ^{1}H, N*H*COCH$_2$), 7.01 (s, 1H,

N*H*COOC), 4.22 (s, 1H, O*H*), 2.26 (dt, *J* = 5.9, 2.3 Hz, 2H, CH_2OH), 1.78 (tt, *J* = 7.1, 2.3 2H,CCH_2CH$_2$), 1.54 (t, *J* = 7.2 2H, CH$_2$CH_2CO), 1.49 – 1.43 (m, 13H, CH_3/CCH$_2$CH$_2$CH$_2$CH$_2$CO); HRMS-ESI(+): calculated for $C_{13}H_{22}N_2O_4Na$: 293.1477, found: 293.1477 [M+Na].

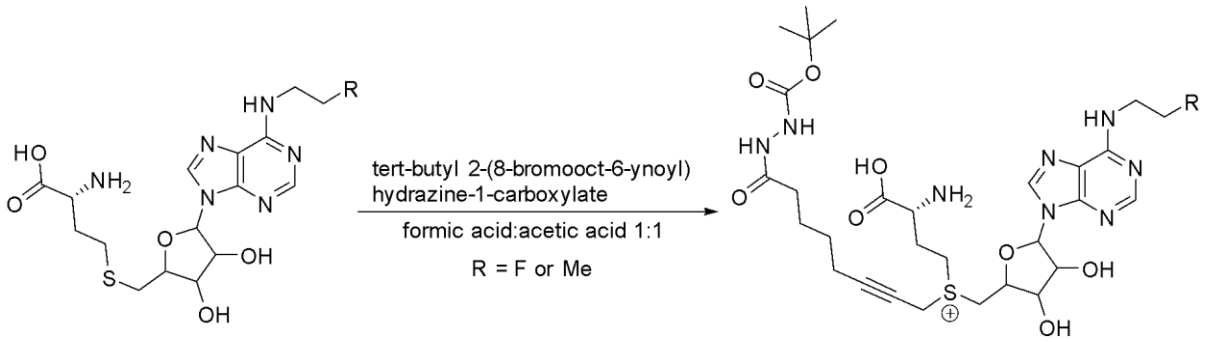
6.36 Synthesis of tert-butyl 2-(8-bromooct-6-ynoyl)hydrazine-1-carboxylate:
A known compound prepared according to literature procedure found in reference [155]



A stirred solution of tert-butyl 2-(8-hydroxyoct-6-ynoyl)hydrazine-1-carboxylate (600 mg, 2.22 mmol) was made in dry DCM (7.00 mL) and cooled on ice. Triphenylphosphine (874 mg, 3.33 mmol) was added and left to dissolve, once dissolved tetrabromomethane (1.10 g, 3.33 mmol) was added slowly. The reaction was then allowed to heat to room temperature and left to stir for 1 hour. After reaction, the solvent was removed under reduced pressure and the crude mixture. Purification was completed using flash column chromatography, eluent conditions: (Hex:EtOAc, 7:3) This was achieved by adding silica to the crude product dissolving in DCM and then removing the solvent under reduced pressure. The resulting silica would have most of the crude adsorbed to it. Crude silica was added on top of the column and was purified by column chromatography. This gave the pure product as an oil. (491.1 mg, 66%); Rf = 0.15 (Hex:EtOAc 7:3); 1H-NMR (400 MHz, $CDCl_3$) δ 8.41 (s, 1H, N*H*COCH$_2$), 6.98 (s, 1H, NHCOOC), 3.84 (t, *J* = 2.3 Hz, 2H, CH_2Br), 2.18 (tt, *J* = 6.9, 3.4 2H,CCH_2CH$_2$), 1.54 (t, *J* = 7.5 2H, CH$_2$CH_2CO), 1.52 – 1.36 (m, 13H,

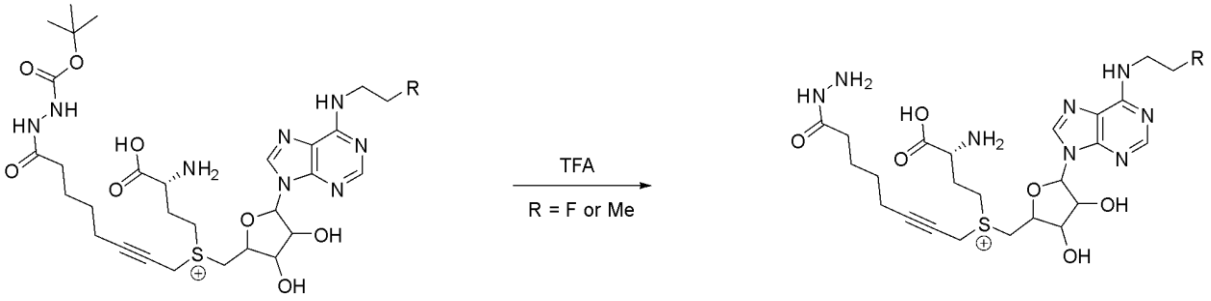
C*H*$_3$/CCH$_2$CH$_2$CH$_2$CH$_2$CO); HRMS-ESI(+): calculated for $C_{13}H_{21}N_2O_3NaBr$: 355.0633, found: 355.0634 [M+Na].

6.37 General coupling procedure



A solution of S-((3,4-dihydroxy-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)-D-homocysteine/S-((5-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-D-homocysteine (10 mg, 0.0234 mmol) was made in a 1:1 mixture of formic and acetic acid (200 µL). Tert-butyl 2-(8-bromooct-6-ynoyl)hydrazine-1-carboxylate (100 mg, 0.300 mmol) was then added dropwise, on ice. The reaction mixture was warmed to 35 °C and left to stir overnight. After this time, the reaction mixture was extracted with diethyl ether and the aqueous layer was collected and dried by lyophilisation:

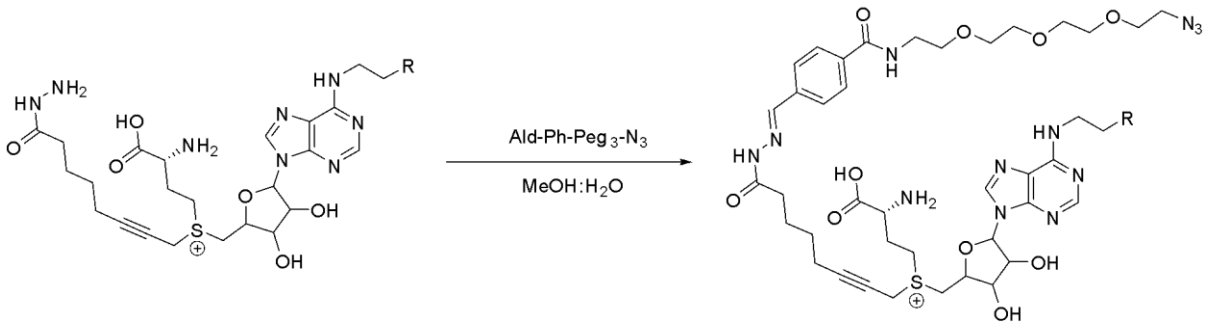
6.38 Cofactor Deprotection



The crude product was dissolved in TFA (400 µl) and left to stir for 2 hours at room temperature. After reaction, the acid was removed under a flow of argon. The crude reaction mixture was then dissolved in water (2 ml). Purification of ((R)-3-amino-3-carboxypropyl)(8-(2-(tert-butoxycarbonyl)hydrazineyl)-8-oxooct-2-yn-1-yl)((3,4-dihydroxy-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)sulfonium / ((R)-3-amino-3-carboxypropyl)(8-(2-

(tert-butoxycarbonyl)hydrazineyl)-8-oxooct-2-yn-1-yl)((5-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)sulfonium was performed by preparative reversed-phase HPLC (ACE 5 C18 25 x 2.12 cm) eluting with 20 mM ammonium acetate pH 5.5 Water (A)/MeOH (B) gradient at a flow rate of 10 ml/min. Gradient system: 30 mins 3-30% B, 30-97% B over 30 mins, hold at 97% B for 5 minutes, stop programme. Retention times (propyl): 37 mins 41 mins, (Fluoro) 27 mins 31 mins 33 mins

6.39 Aldehyde Coupling:



To the collected HPLC fractions Ald-Ph-PEG_3-Azide (1.2 equivs) was added and rolled for 30 mins at room temperature. The fractions were then dried by lyophilsation. Once dry the solids were dissolved in 100 µl 0.1% acetic acid and stored at -20 °C. Concentrations were determined by UV absorption analysis with ε260 = 15.400 dm^{-3} $mol^{-1}cm^{-1}$Acetone needs to be avoided during lyophilisation as exchange with hydrazone to oxime was observed.

6.4 Biological DNA methods

6.41 Wild Type M.MpeI information

M.MpeI Acquisition number: Q8EVR5

From work completed by Lukinavičius et al (2012) to M.HhaI, M.HpaII and M2.Eco311 , previous work in the Neely group used the Q5 Site-Directed Mutagenesis Kit (NEB) to introduce N374A and Q136A mutations in the M.MpeI protein sequence, this improved its

efficiency with synthetic cofactors. The enzyme also had a His-tag (6x histidine) engineered onto the *N*-terminal; this sequence can be found below.

Engineered M.MpeI gene sequence:

GTGGTGGTGG TGGTGGTGCT CGAGTGCGGC CGCAAGCTTT CACGAGGCCC
TTTCGTCTTC AAGCATTCTC ATGTTTGACA GCTTATCACC GATCGCAGCA
CAGCTCGCTG GTCCAGAACT GATTTCTTCG TTATTAACAA ACTCCAGCGT
ATTGAAAATA GCTTCCAGAA TCTTGACCGG AATGGAGGCT CCAGCAATAT
AAATCATCTT GTTCTCGGAA ATGAGATTGG TGCTCTGCAC TTTCTTAAAA
TCGTTGACAT CGAACTGCAT GTATTTGAAA CATTCCAAAG GAGTCAGGTA
GCGCACACCC TGCTGGGTTT CGATCTTGAT GCGGCTATTA GCCCCGGAGG
CGGTAAGTGT AGGGCCGATC CCGTTGATAT TGTAAACGTA ATTTTCAGAG
TTAAAAGTAG TGTAATTCTT AAGGGAACGG CTGATAATAT TGGACTTGGT
TTCGCGAAAA GTGGTTGTTT CATATTTATT GAGATTAAGA TATTTGTAAT
TGCTGCTGTC TACCAGAATG TCCTTGATTT TTTTCGGTGG ATTTTTAACT
TTTTCCAGCT CCTTGAATTT GAAACCAGTT TTCTCCAGGT AGTCATCGCG
GATGGACAGA CAAAAGACAC GCTCACGGTT TTGGCAGTTA TCAAAATTTT
TACTATTTAA GAGATAGGTT TTGCTTTTGT ATCCAAATTT CTCTAATTGT
TTTAACCACG TGTTGTAGTT CTTCTTATTT TTATGGCTTA ACAGATTTTT CACGTTTTCC
ATCAGAAGAT ATTTCGGCAT CTCTTCCTTA CTAAAAGAAT TCTTAATTTC
TTCTAAGATG CGTTCAATCT CCCACAGCAG ACCAGAACGT GTGTTCAGTT
CTTTATCGAT GCCTTTCTGA AGACCTTGCA CCGAGAGATC TGCACAGGGA
AAGCTATACG TGAAGATATC AATGTTTTTC GGAAAATTGT CTTTATTGAC
TTTCTTGATG TCGAACAGGT TGTTGAAATG TTTCTTAGCA TAGTTCAGGT
AGCTCGCCTT AATTGTGTTG TTAATTTTCT TAATGCCGTA TTCGCTGATC
GGCATTTTTG AATCATTGCT AATACTCAGA ATATCTTTAT CCAGCTGTTC
GATCTTAGGG TTGAAGTTCT TTGAGTGAAT CGCCACATAG CTGACGATAG
CGTCTACAAA CCATTCGACC ATACCAGAAT GCTGGATTTC CCAGTTTTTA
GAGCGCGCAA TATTTTTCAA CGCTTTGAAC TGGGAACCAA TCCCCGCAAA
GGCCTCAAAA ACCTTAATCA CTTTAATTTT GTCCTTATTG GAGTTCAT

Engineered M.MpeI amino acid sequence:

MNSNKDKIKVIKVFEAFAGIGSQFKALKNIARSKNWEIQHSG

MVEWFVDAIVSYVAIHSKNFNPKIEQLDKDILSISNDSKMPISE

YGIKKINNTIKASYLNYAKKHFNNLFDIKKVNKDNFPKNIDIFTY

SFPCADLSVQGLQKGIDKELNTRSGLLWEIERILEEIKNSFSKEE

MPKYLLMENVKNLLSHKNKKNYNTWLKQLEKFGYKSKTYLLN

SKNFDNCQNRERVFCLSIRDDYLEKTGFKFKELEKVKNPPKKIK

DILVDSSNYKYLNLNKYETTTFRETKSNIISRSLKNYTTFNSENY

VYNINGIGPTLTASGANSRIKIETQQGVRYLTPLECFKYMQFD
VNDFKKVQSTNLISENKMIYIAGASIPVKILEAIFNTLEFVNNEEI
SSGPASCAAIGDKLSNMRMLEDERASHHHHHH

Molecular weight: 48857 g/mol
Extinction coefficient: 86320 M^{-1} cm^{-1}

6.42 Gel Electrophoresis

First the gel was made up using a gel holder and a comb to make the wells. 0.8 g of agrose was weighed out and then mixed with 80 mL of tris-acetate-EDTA buffer. This mixture was then microwaved for about 1 min until the solution had completely dissolved. The mixture was then poured into the holder and any bubbles were removed. The solution was allowed to cool so the gel was able to set. After this was set each sample had loading dye added to it and loaded onto the gel, alongside the DNA ladder. This was then run at 120 V for 40 mins. After the samples had reached about 2/3 down the gel, the gel was placed into Gel Red, (Sigma-Aldrich) and placed on the shaker for 30 mins to stain the gel. This was then visualised using the UV visualiser with an EtBr filter with UV illumination, λ = 254 nm.

6.43a Protection Assay (M.TaqI)

A Master Mix of 120 µL was made up on ice which contained cutsmart buffer (12 µL, NEB), 1.5 µL of M.TaqI (1.5 µL, 37.5 ng/µL), pUC-19 (6 µL 50 ng/µL), and nuclease free H_2O (100.5 µL) . The Master Mix was split across 9 tubes each with 10 µL of the Master Mix. Each subsequent tube contains 10 times less cofactor making up a concentration of 500 µM to 16.125 µM unless stated otherwise from a stock of 10 mM of cofactor. Tubes 7-11 were controls and tubes 7 and 10 contain no Master Mix and are made up using cutsmart (1 µL) and pUC-19 (5 µL, 50 ng/µL) and either 0.204 µL or 0 µL of cofactor (10 mM) for 7 and 10 respectively. Tubes 8 and 11 contain the Master Mix but also 0.33 µL of 3 mM of SAM. Tube 9 contains only Master Mix. Each of these tubes was incubated at 50 °C for 1 hour. After the elapsed time R.TaqI (0.5 µL, 20k units/mL, NEB,) was added to each tube except 11 and left

to incubate at 50 °C for 1 hour. Proteinase K (0.5 µL, 20 mg /mL, NEB) was added to each tube and left once more to incubate at 50 °C for 1 hour. The samples were analysed using gel electrophoresis.

6.43b Protection Assay (M.MpeI)

For this each tube was prepared individually. In each of the tubes a mixture of nuclease free H_2O (up to 15 µL), cutsmart (2 µL, NEB), Puc19 (50 ng/µL 2 µL, NEB), and then either M.MpeI (6.8 mg/µL 2 µL) and/or cofactor (10 mM, 1 µL). These were left to incubate at 37 °C for 1 hour, after the allotted time proteinase K (4 µL, 20 mg/mL NEB) was added and the samples were left to incubate at 50 °C for 1 hour. After this time, the solutions were purified using Sigma Clean Up Columns or Zymo Clean Up Columns using 37 °C ultrapure water as the eluent. The samples then had Tango buffer (2 µL, NEB) and HPaII (0.3 µL, 10k units/mL, NEB) added except one of the SAM controls and was left to incubate at 37 °C for 1 hour and then proteinase K (0.3 µL, 20 mg/mL) was added, and the solutions were left to incubate at 50 °C for 1 hour. Gel electrophoresis was then performed on all samples. Controls that were prepared were those that didn't contain cofactor, didn't contain enzyme, didn't contain cofactor or enzyme and two SAM controls one of the SAM controls didn't have Tango and HpaII added but the other control did have these added.

6.34 Mutagenesis PCR Amplification

A Master Mix of 400 µL was made up was made up that contained 200 µL of Master Mix Q5 hot start 2x stock, 20 µL of 10 µM K115X forward primer, 20 µL of 10 µM K115X reverse primer, 8 µL of 10 ng μL^{-1} M.MpeI_pRSETB plasmid and 152 µL of H_2O. This Master Mix was then split into 8 PCR tubes each containing 50 µL each and placed into a thermocycler which then was set on an amplification programme: initial denaturation (98 °C) for 30 seconds, then for 30 cycles: 98 °C for 10 seconds, 51 °C for 30 seconds and 72 °C for 2 mins. then held at 4 - 10 °C. The DNA was then purifed using either Sigma PCR Clean Up Columns or Zymo Columns using 37 °C ultrapure water as the eluent. This DNA was sent off for sequencing.

6.45a Purification via sigma columns

Reactions were purified using GenElute PCR Clean-Up Kit (Sigma Aldrich) according to the manufacturer's instructions. Centrifugation steps were done between 12000 –16000 RPM using Eppendorf Centrifuge model 5418 fitted with a rotor of FA-45-18-11.

6.45b Purification via Zymo columns

Before starting, buffers were prepared in accordance with the Zymo DNA Clean & Concentrator manual. The protocol was followed, and any centrifugation steps were performed between 10000 and 16000 RPM using Eppendorf Centrifuge model 5418 fitted with a rotor of FA-45-18-11.

6.46 Mutant MTase IVTT

The K115X PCR plasmid was supplied to NEB PURExpress® In Vitro Protein Synthesis Kit and protocol was followed in accordance with manufacturers protocol. 2 µL of cofactor (10 mM, 1 µL) was added to this mixture and mixed. This was incubated at 37 °C for 1 hour and proteinase K (4 µL, 20 mg/mL) was added and incubated at 50 °C for 1 hour. Mixture was purified using Sigma Clean Up Columns or Zymo Clean Up Columns using 37 °C ultrapure water as the eluent. The sample then had Tango buffer (2 µL, NEB) and HpaII (0.3 µL, 10k units/mL, NEB) added and was left to incubate at 37 °C for 1 hour followed by addition of proteinase K (0.3 µL, 20 mg/mL) and the sample was left to incubate at 50 °C for 1 hour. Mixture was purified using Sigma Clean Up Columns or Zymo Clean Up Columns using 37 °C ultrapure water as the eluent. Bacterial plates were set up using LB Broth with agar. This was weighed out (4% of the volume used) and dissolved in H_2O, which was then autoclaved for an hour and allowed to cool to less than 70 ℃. Kanamycin was added to make a final concentration of 30 µg/mL. The medium was then poured onto the plates under a flame to sterilise the environment and the plates were allowed to set, once set these were warmed 37 ℃ when required. Escherichia coli T7 Express Competent (High Efficiency NEB) were

removed from the –70 ℃ freezer and allowed to thaw on ice for 30 mins. To this was added 5 µL of the purified DNA and the mixture was mixed by flicking the tube and the samples were left on ice for 30 mins. After the time had elapsed the samples were then placed in a water bath at 42 ℃ for 10 seconds. The samples were then placed back on ice for 5 mins. After this the samples had 950 µL of room temperature SOC growth medium added with mixing done by flicking the tube, these were then warmed to 37 ℃ and shaken at 300 RPM for 1 hour. After the elapsed time, dilutions of the samples were made up and the 100 µL of each sample was spread across an agar plate. This was then allowed to incubate at 37 ℃ overnight for colonies to grow.

6.47 MTase Protein Expression and Purification

LB Broth with agar was weighed out (4% of the volume used) and dissolved in H_2O, this was then autoclaved for 1 hour and allowed to cool to less than 70 ℃. Kanamycin was added to make a final concentration of 30 µg/mL. The medium plates were poured under a flame to sterilise the environment allowed to set, once set these were warmed 37 ℃ when required. E. coli T7 Express Competent (High Efficiency NEB) were removed from the –70 ℃ freezer and allowed to thaw on ice for 30 mins. To this was added 5 µL of pET plasmid DNA (500 ng / µL) containing the appropriate mutation, the contents was mixed (by flicking the tube) and the samples were left on ice for 30 mins. The samples were then placed in a water bath at 42 ℃ for 10 seconds. The samples were then placed back on ice for 5 mins. After this the samples had 950 µL of room temperature SOC growth medium added with mixing done by flicking the tube, these were then warmed to 37 ℃ and shaken at 300 RPM for 1 hour. Dilutions of the samples were then made, and each sample (100 µL) was streaked across an agar plate, incubated at 37 ℃ overnight. LB broth (2% of volume used) and autoclaved, after cooling below 70 °C, kanamycin was supplemented to make up a final concentration of 30

µg/mL. A single colony from the plates was selected and removed using a sterile pipette tip and placed inside the broth, lightly sealed and grown at 37 °C. 300 RPM until an OD^{600} was achieved. Temperature was reduced to 28 °C and was supplemented with IPTG to give a final concentration of 500 mM, then incubated overnight at 300 RPM. The culture was split into (50 mL) Falcon tubes and cells were harvested by centrifugation (20 mins at 3993 RPM, 4 °C) using an Eppendorf centrifuge 5430 R fitted with fixed angle rotor F-35-6-30. Pellets were washed with 50 mL of pre-chilled, phosphate buffered saline and re-harvested (30 mins at 7500 RPM, 4 ℃). The cell pellets were resuspended & lysed using BPER (ThermoFisher) as instructed by manufacturers protocol. Cell lysate was cleared using centrifugation (20 mins at 3993 RPM, 4 °C). The supernatant was collected and was passed through a filter (0.45 µM in diameter). A 5mL HisTrap column (Cytiva) was loaded into an protein purification system and equilibrated with 25x column volumes (CV) of cold equilibration buffer (20 mM Tris-HCl, 40 mM imidazole, 0.5 mM DTT, EDTA-free Protease Inhibitor Cocktail & DNAse I). Supernatant was loaded onto the column and washed with 30x CV of Wash buffer (20 mM Tris-HCl, 40 mM imidazole, 500 mM NaCl, 0.5 mM DTT, EDTA-free Protease Inhibitor Cocktail (Thermofisher) & DNAse I (NEB)). MTase mutant was eluted into 20 mM Tris-HCl, 40 mM imidazole, 500 mM NaCl, 0.5 mM DTT, EDTA-free Protease Inhibitor Cocktail & DNAse I, using a 0-500 mM gradient of imidazole and was collected into 2 mL fractions. Fractions containing protein were pooled and concentrated into storage buffer 10 mM Tris-HCl, 0.5 mM TCEP, 1 mM EDTA (pH 7.5) using Amicon Ultra-15 Centrifugal Filters (30 kDa cut off), repeated 3 times as described by manufacture, 50% glycerol was added to the pools for long term storage. All steps that involved exposing bacteria cells or the growth medium to air were done with a flame to sterilise the surrounding environment.

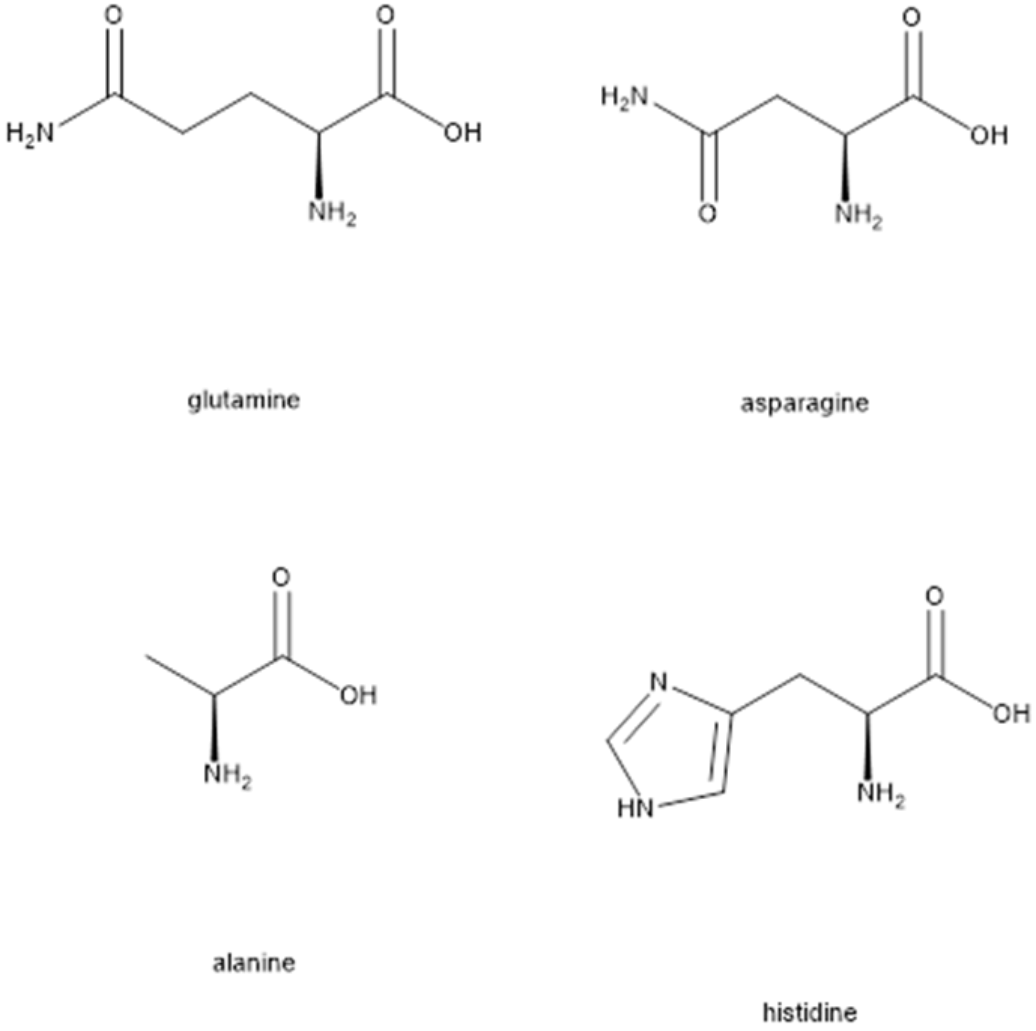
Appendix

7.1 Structure of Amino Acids

Below are the structures for the size and charge related amino acids that the positively charged K115 residue was changed to. Size related residues all decreases the size of the

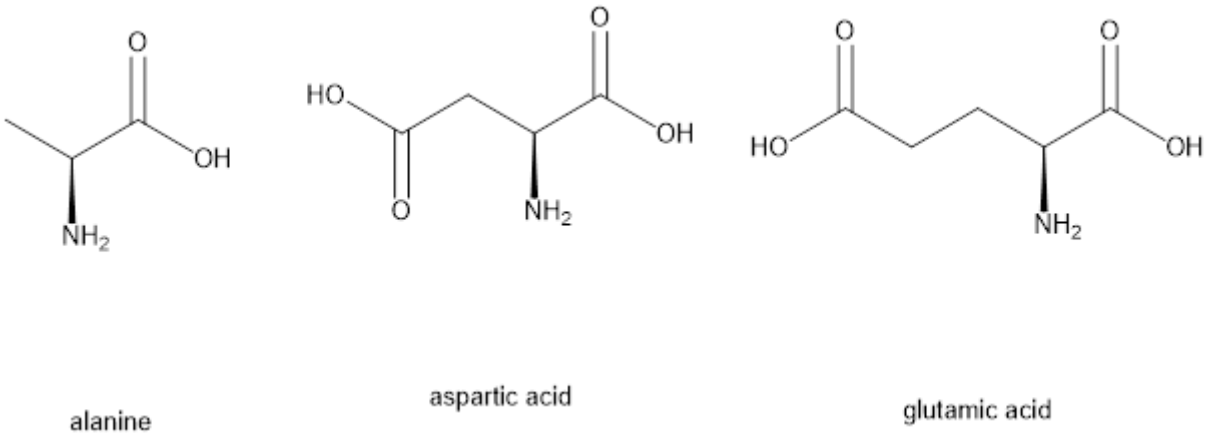
residue but maintained the positive charge, apart from the alanine residue which does change the positive charge from positive to neutral. The charge related amino acids changed the positive charge to either neutral or negative, alanine switched it to neutral whilst the acidic residues changed it to negative.

7.11 Size Related Residues



Appendix Figure 1: Structures of size related amino acids pKa of histidine 6.04

7.12Charge Related Residues



Appendix Figure 2: Structures of charge-related amino acids pKa of aspartic acid and glutamic acid are 3.90 4.07 respectively.

7.2 Docking method

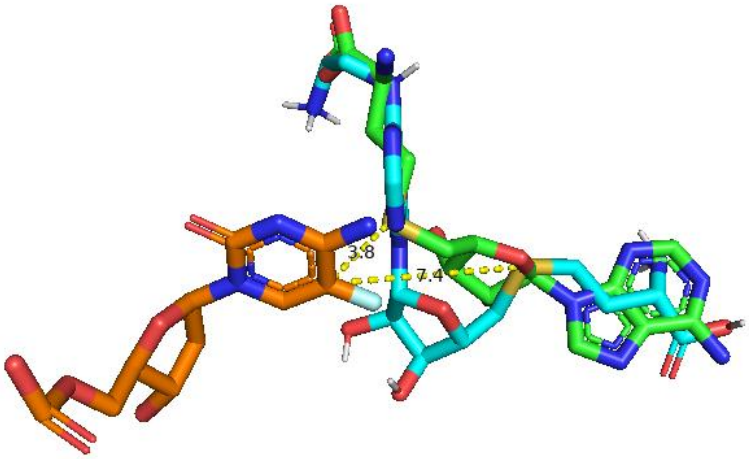
The active site of the enzyme was located. This was achieved by looking through the enzyme sequence to find where the cofactor is located. By using code in pymol the residues around the cofactor within 5 angstroms were then deemed to be the active site. The coordinates of this active site were determined by using pymol coding. This was then used in another program called Autodock Vina, in order to prepare the protein and ligand for docking.

1. The protein was first prepared to receive the ligand: (i) the first step was to remove the water molecules in the protein. If water had been present in the active site, this could interfere with the docking values that were being calculated as it may have bound better when the water is present. (ii) The next stage was to add polar hydrogens to the protein, (iii) finally the charges were added by addition of Kollman charge. Finally the protein was saved in a file format as PDBQT, this prepared the protein for docking. The ligand was then prepared by building in pymol. This was so the ligand was in the PDB format that the Autodock Vina could use. Once this was in the correct format, the ligand was then prepared by selecting input and choosing the

ligand to be prepared for Autodock Vina. Once this step was completed the ligand was ready for docking but needed to be saved as a PDBQT file. Then, both protein and ligand were in the correct format to dock with.

2. The next stage was to prepare the active site using an autobox grid so that the ligand is docked around the active site. By loading both files in Autodock by selecting gridbox a cube is generated in the middle of the protein, then the grid file is generated. Once output to grid file has been selected, this text file can be opened. This needs to be edited as it needs to have the coordinates set around the active site. Once this is complete the docking can go ahead as the protein, ligand files are in the correct format and the grid file informs where the docking can take place.
3. The next step was to make a config text file, as this will contain the information for the command window that will be used to find all the files. The config file needs “receptor”, “ligand”, the centres of the coordinates, and the size of the grid box in each direction which should be 20x20x20. The coordinates of the centre should be 27.8 x 26.6 x 43.11. The energy range must be set: if the default value is 4, then the exhaustiveness is set which has a default value of 8. Once the config file had been generated the docking could proceed using the command window. Through the command window the code was written to start the docking process. First the path directory was given to the command window. This was done by typing “cd” into the command window. By copying the location of the file path, this told the command window where to find all the necessary files for the docking. The executable was then called and this was where the receptor files, ligand file and the config file were specified to the vina executable. The log file is then also produced as well as the output file. Then the docking process had begun; once this had fully finished the output file would be a pdbqt file which could then be loaded into pymol. This could be used to visualise all versions of the docking. The log file should output a text file detailing all the energy values for each dock position. The number of docked

positions should be equal to the number that was supplied in the exhaustiveness config. This can be tweaked for different versions, if required.

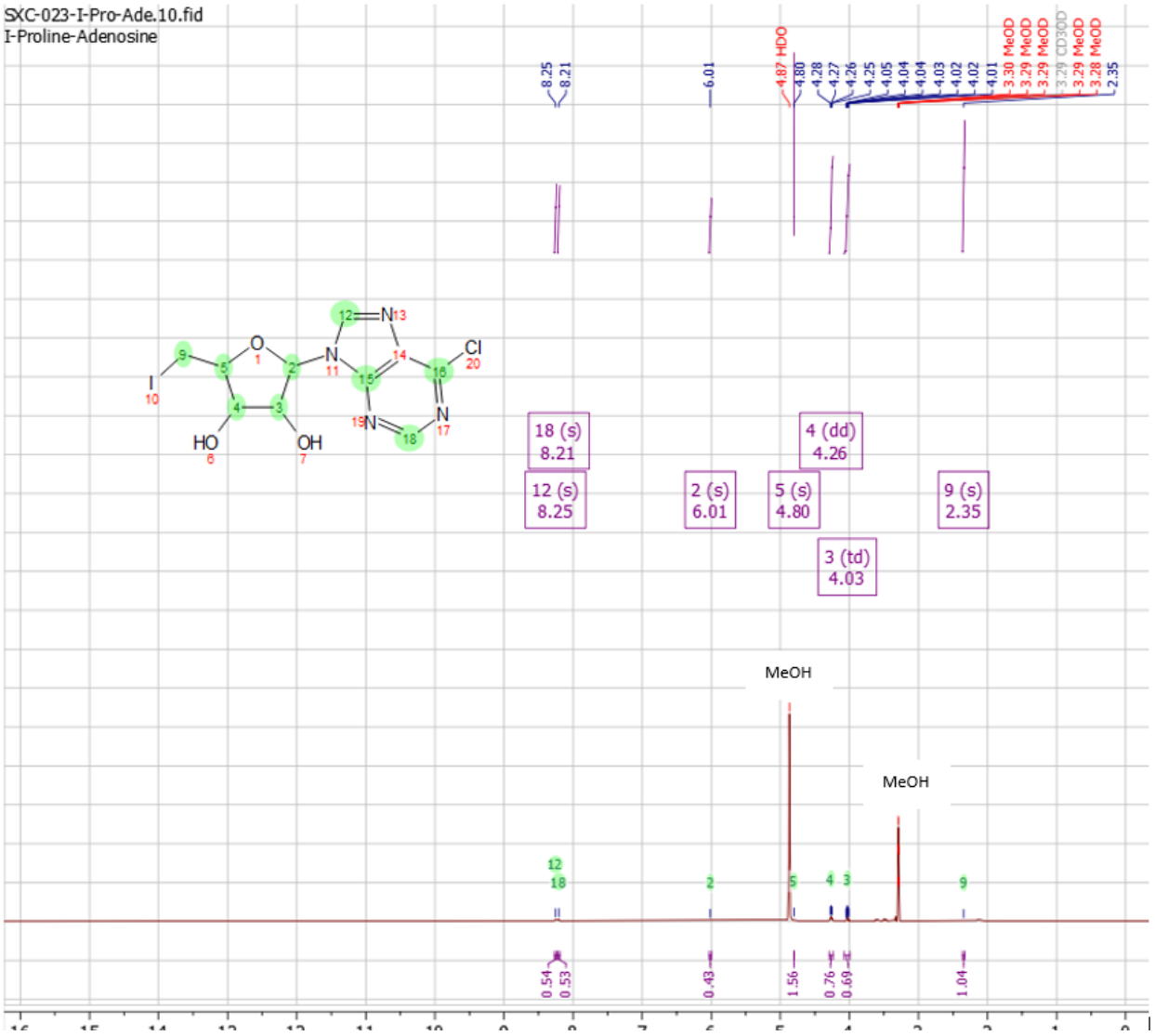


*Appendix Figure 3: Example of when docking between **1** (Blue) and M.MpeI is poor compared to SAH (Green). Adenosine ring is where the Sulphur is located, and Sulphur is out of position for transfer to be viable PDB code 4DKJ.*

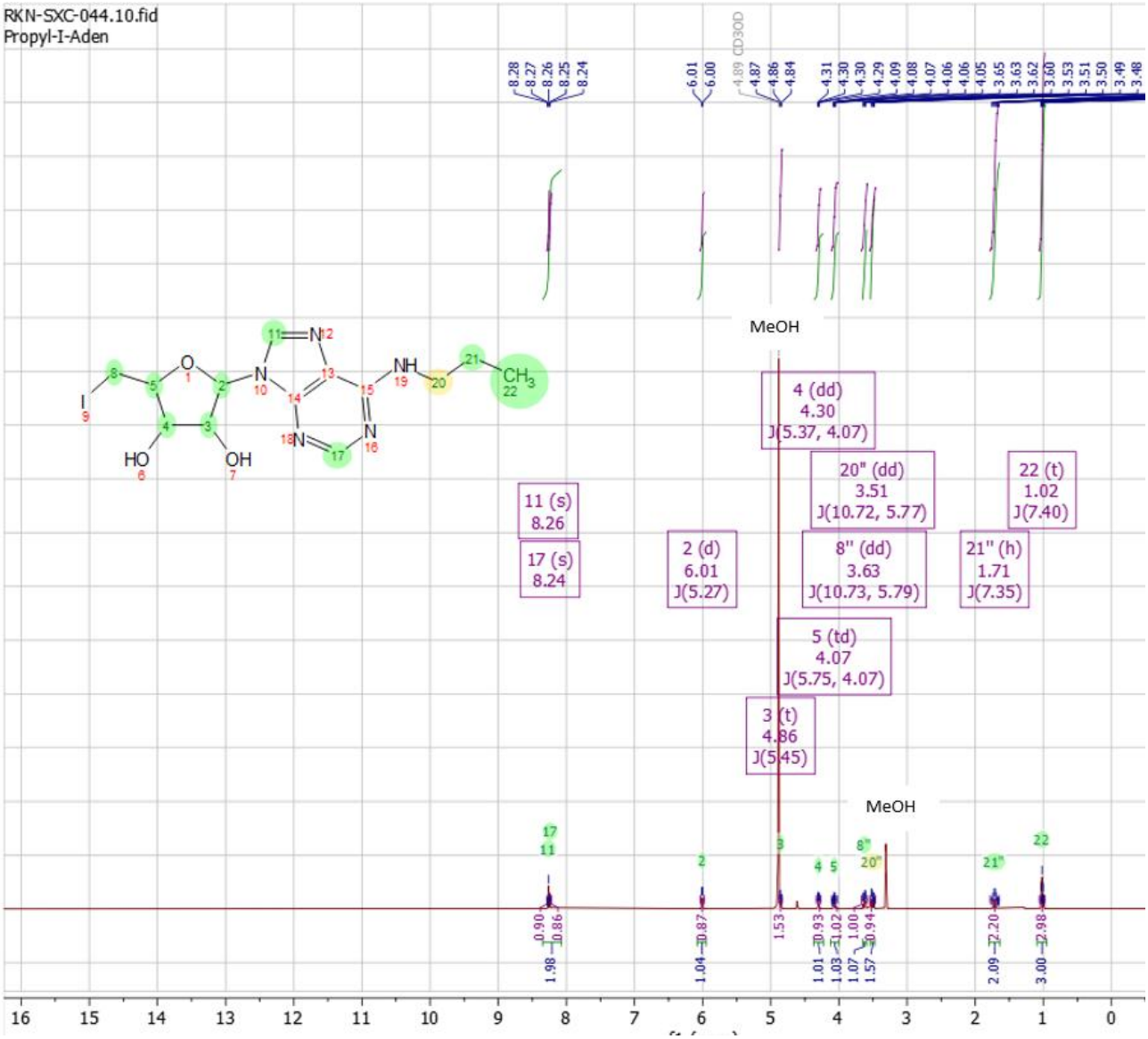
Appendix Table 1: Table detailing energy results for best oriented cofactor and MTase species.

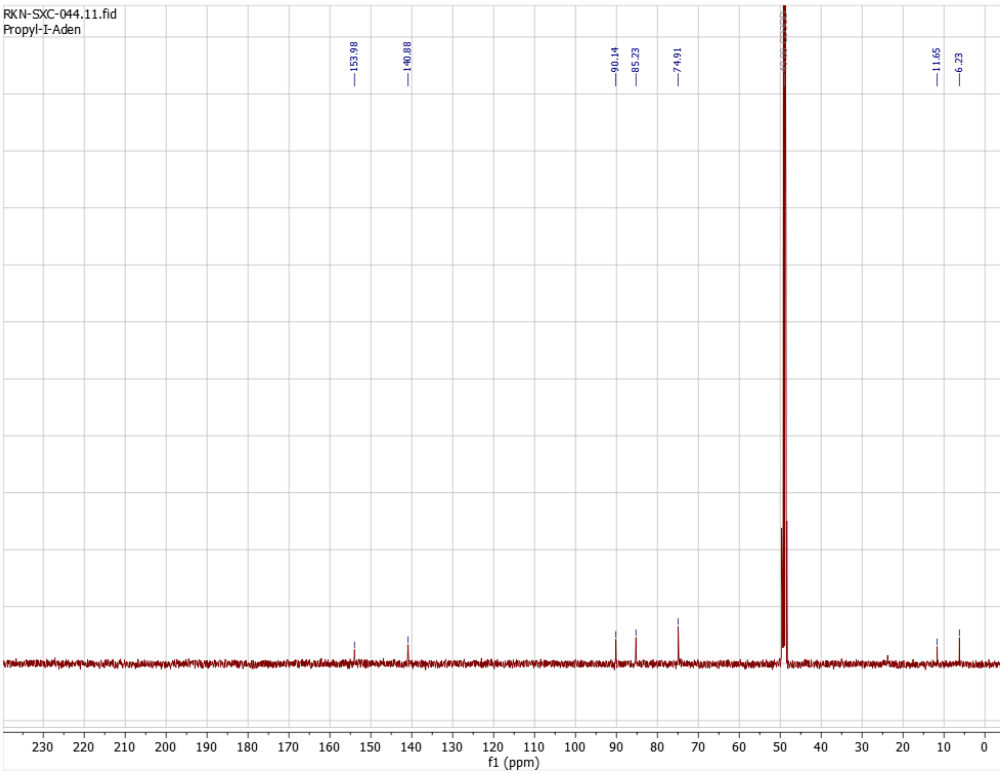
Enzyme	Cofactor	Docked result (kcal / mol)
Wild_Type	SAH	-10.1
Wild_Type	**5**	-9.2
Wild_Type	**1**	-9.0
Wild_Type	**1**(appendix Figure 3)	-8.5
Wild_Type	**2**	-9.6
Wild_Type	**4**	-8.9
Wild_Type	**3**	-9.3
K115A	SAH	-9.1
K115D	SAH	-9.4
K115E	SAH	-9.4
K115H	SAH	-9.5
K115N	SAH	-9.6
K115Q	SAH	-9.6

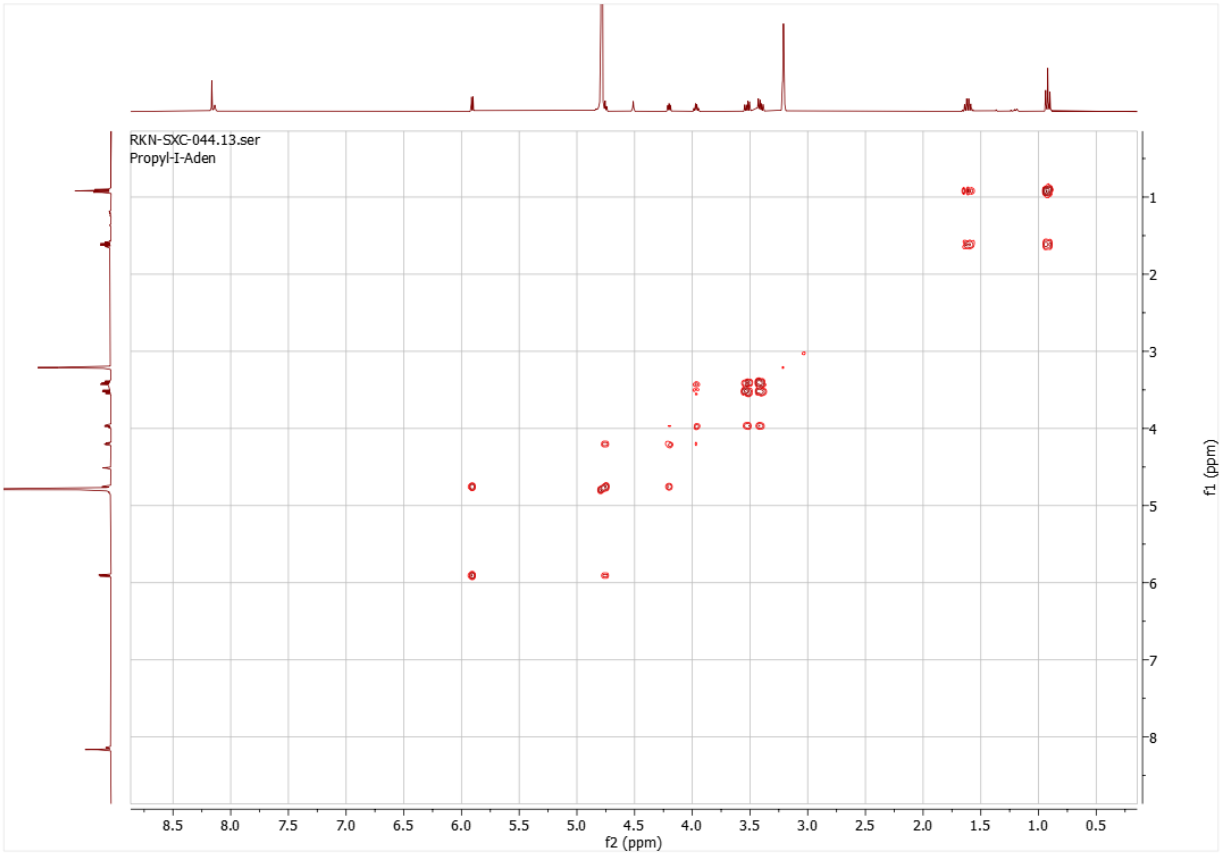
7.3 NMRs

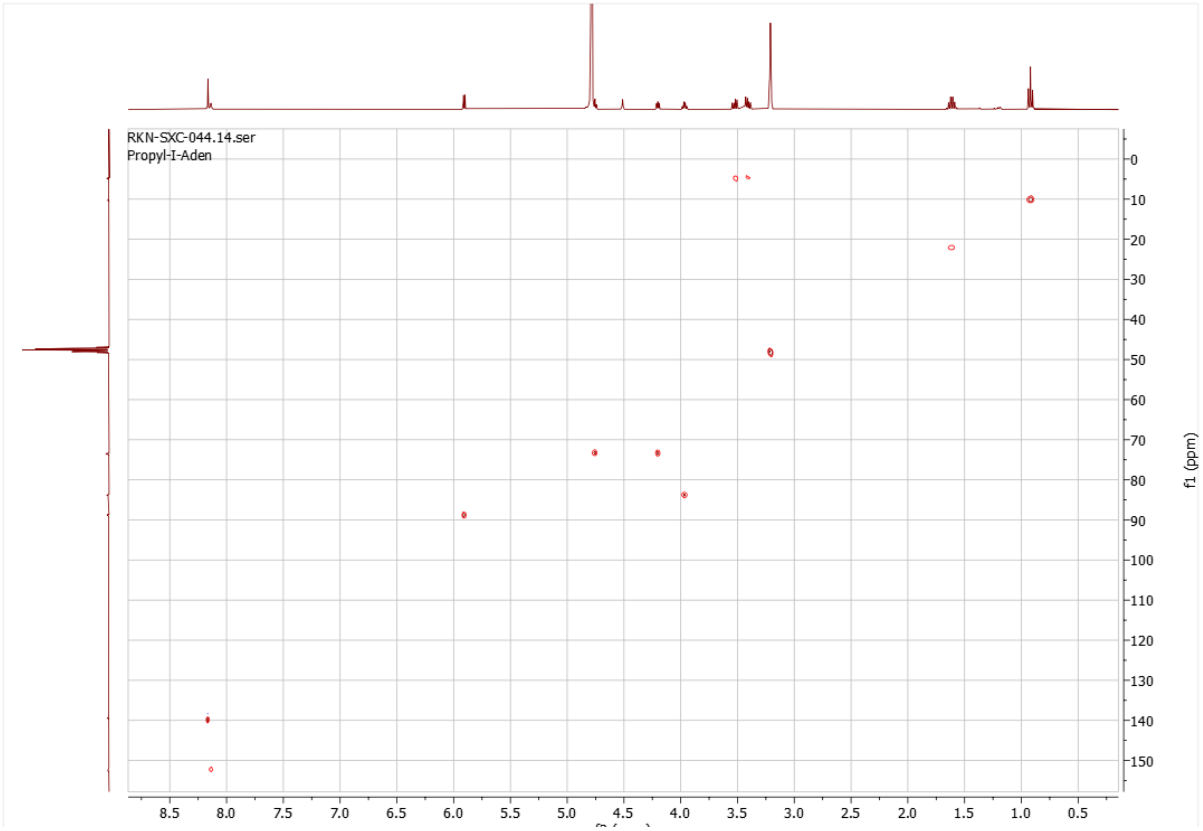


Appendix Figure 4: ^{1}H NMR spectrum of 2-(6-chloro-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol. Acquired using a Bruker AVII 300 and MeOD

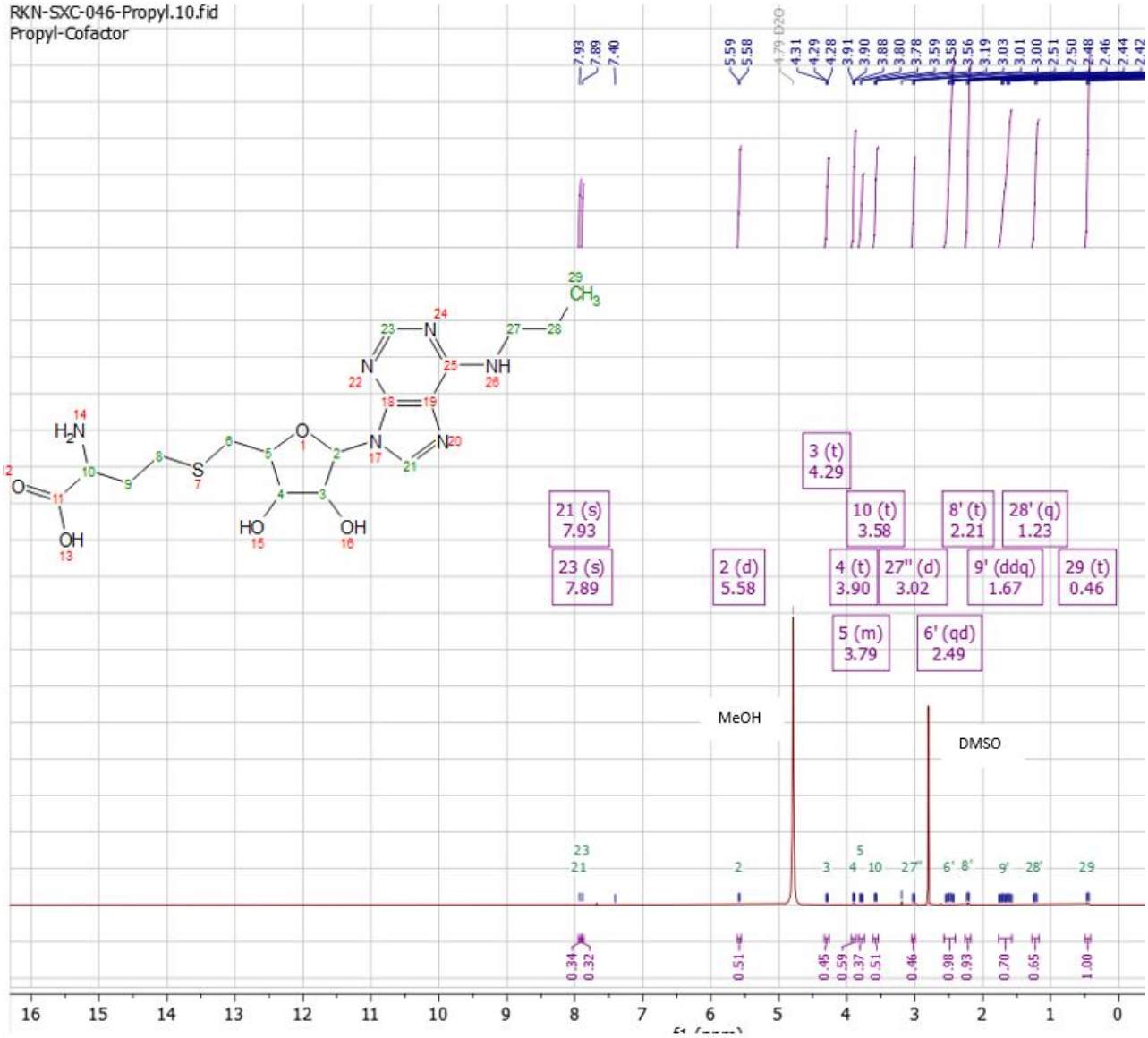

RKN-SXC-044.10.fid
Propyl-I-Aden
CD3OD
MeOH
MeOH
11 (s)
8.26
17 (s)
8.24
2 (d)
6.01
J(5.27)
4 (dd)
4.30
J(5.37, 4.07)
20" (dd)
3.51
J(10.72, 5.77)
8" (dd)
3.63
J(10.73, 5.79)
22 (t)
1.02
J(7.40)
21" (h)
1.71
J(7.35)
5 (td)
4.07
J(5.75, 4.07)
3 (t)
4.86
J(5.45)
HO
OH
I
NH
CH3

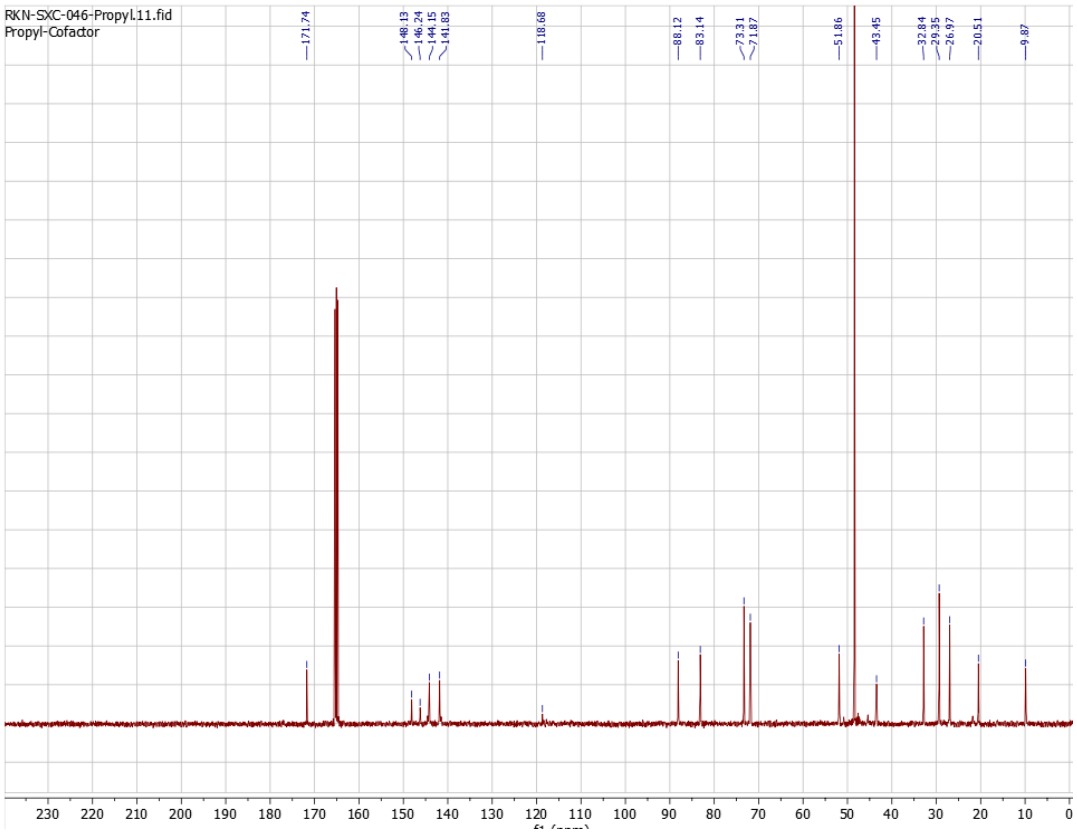

RKN-SXC-044.11.fid
Propyl-I-Aden
153.98
140.88
90.14
85.23
74.91
11.65
6.23
f1 (ppm)

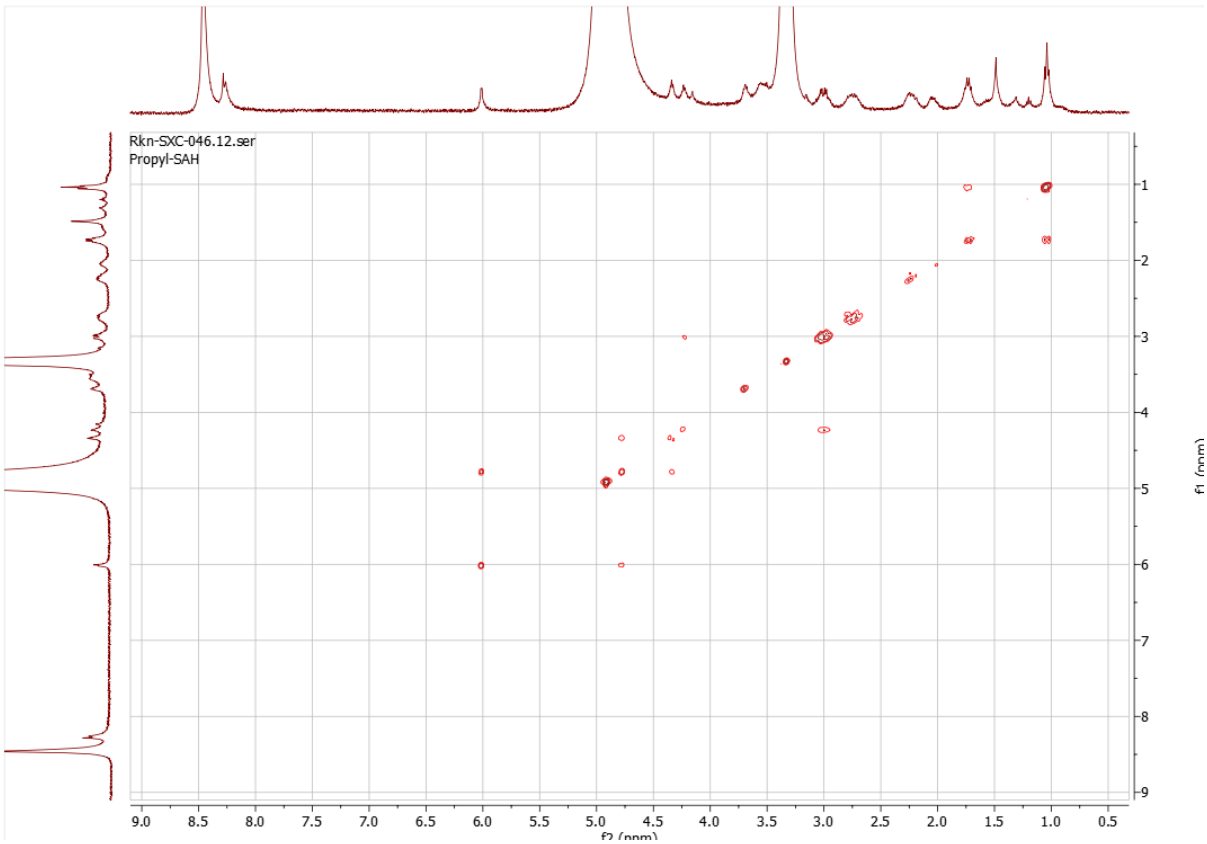

RKN-SXC-044.13.ser
Propyl-I-Aden
f2 (ppm)
f1 (ppm)

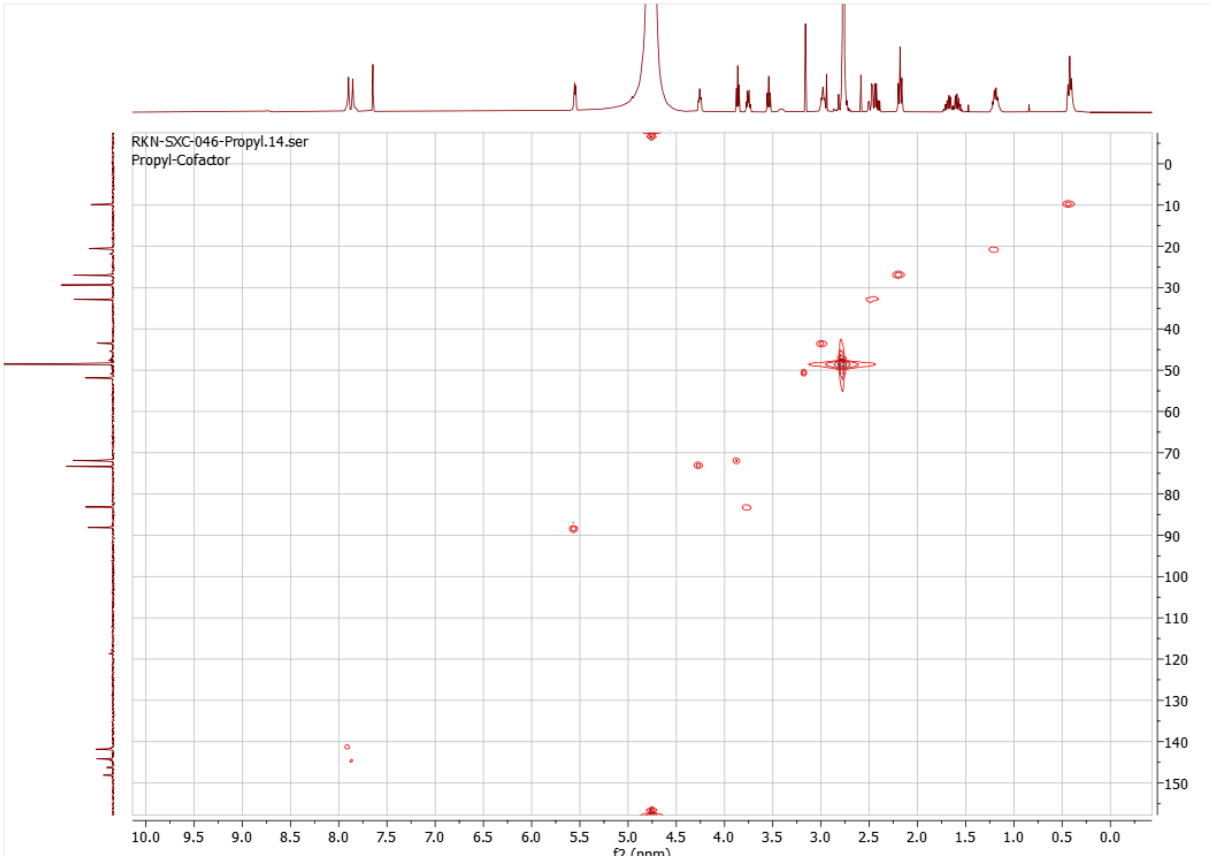


Appendix Figure 5 ^{1}H NMR spectrum of 2-(iodomethyl)-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-3,4-diol (top) ^{13}C NMR spectrum of 2-(iodomethyl)-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-3,4-diol (bottom). Acquired using a Bruker AVII 300 and Bruker AVII 400 and MeOD

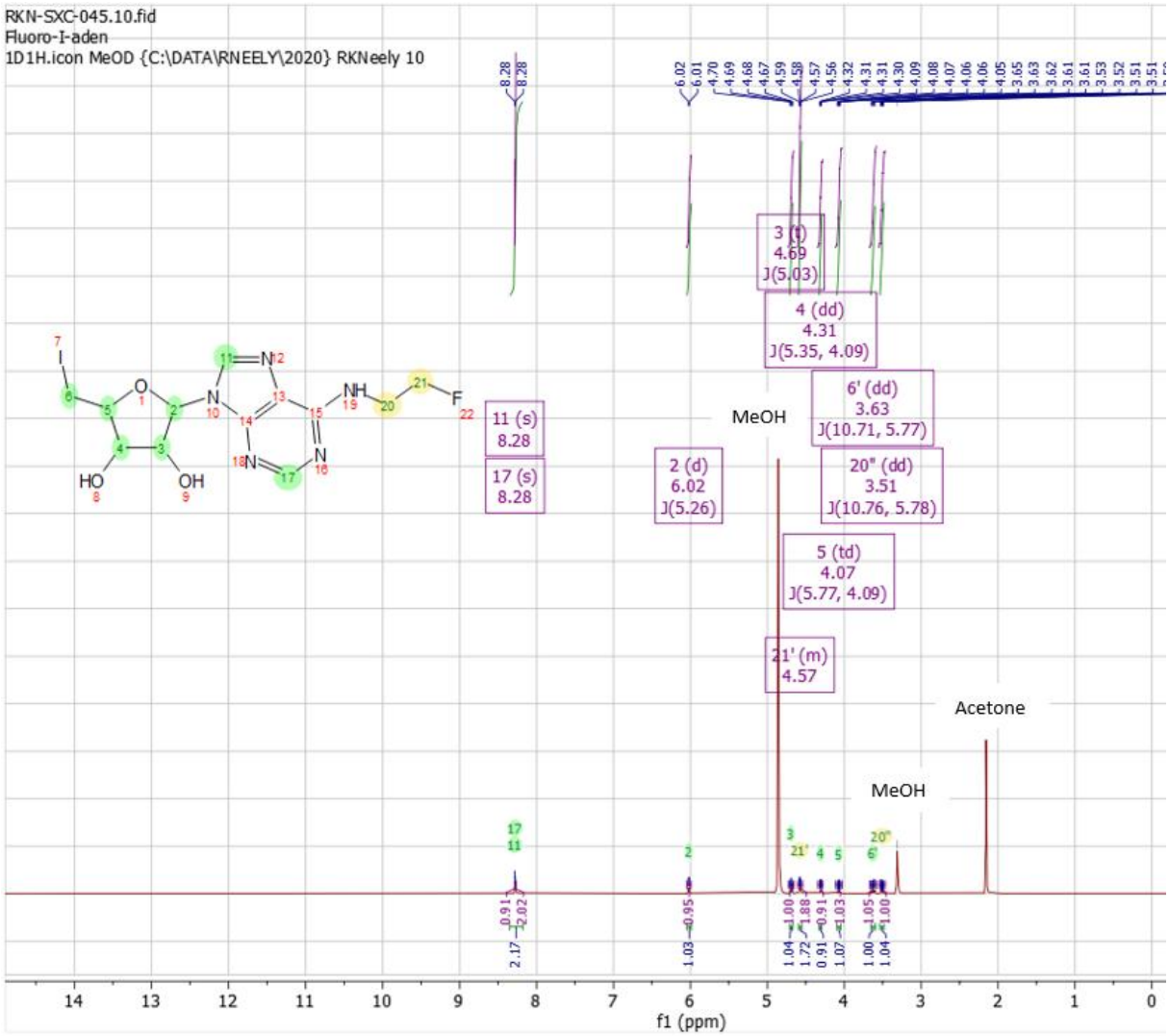

RKN-SXC-046-Propyl.10.fid
Propyl-Cofactor
21 (s) 7.93
23 (s) 7.89
2 (d) 5.58
3 (t) 4.29
4 (t) 3.90
5 (m) 3.79
10 (t) 3.58
27'' (d) 3.02
6' (qd) 2.49
8' (t) 2.21
9' (ddq) 1.67
28' (q) 1.23
29 (t) 0.46
MeOH
DMSO

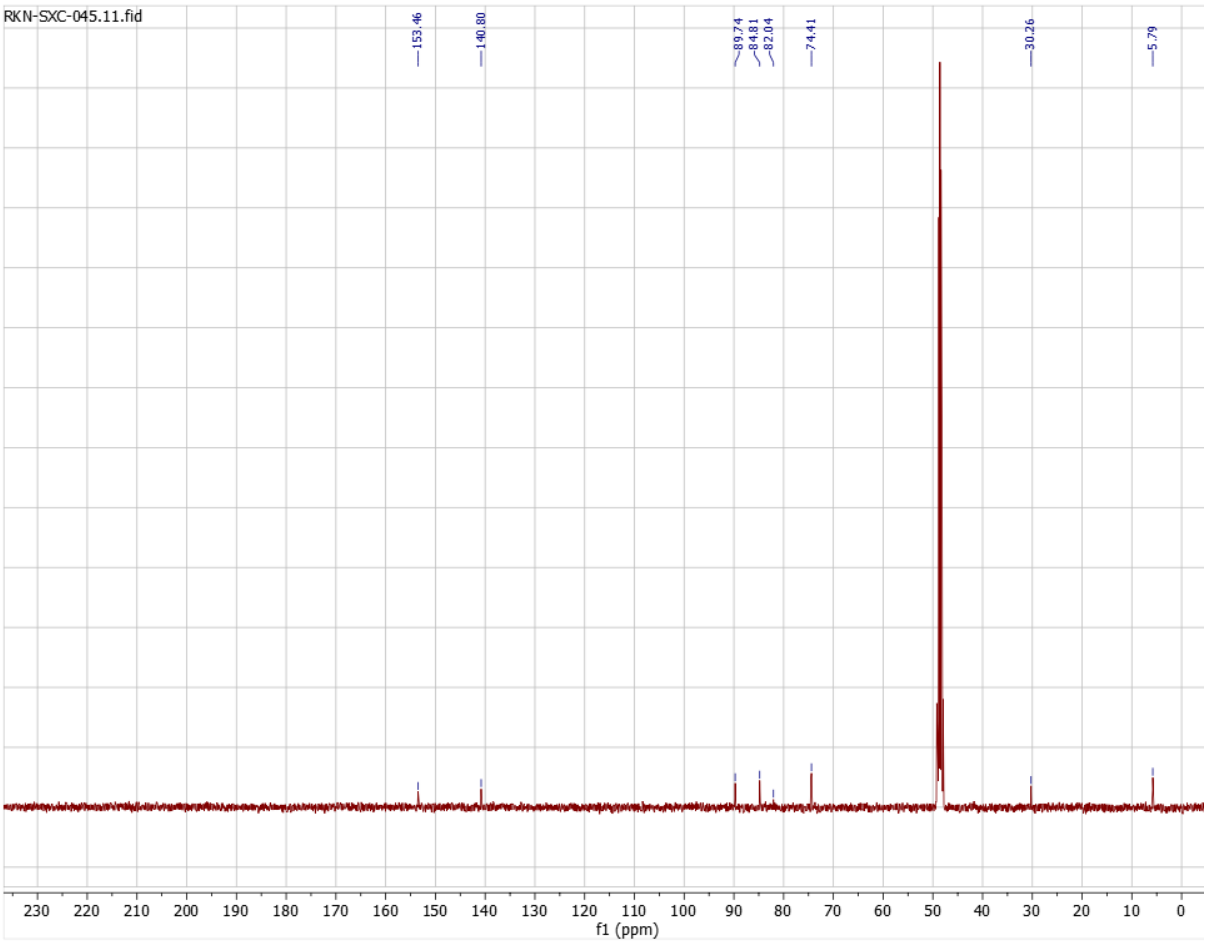

RKN-SXC-046-Propyl.11.fid
Propyl-Cofactor
171.74
148.13
146.24
144.15
141.83
118.68
88.12
83.14
73.31
71.87
51.86
43.45
32.84
29.35
26.97
20.51
9.87
230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0


Rkn-SXC-046.12.ser
Propyl-SAH
1 2 3 4 5 6 7 8 9
f1 (ppm)
9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f2 (ppm)

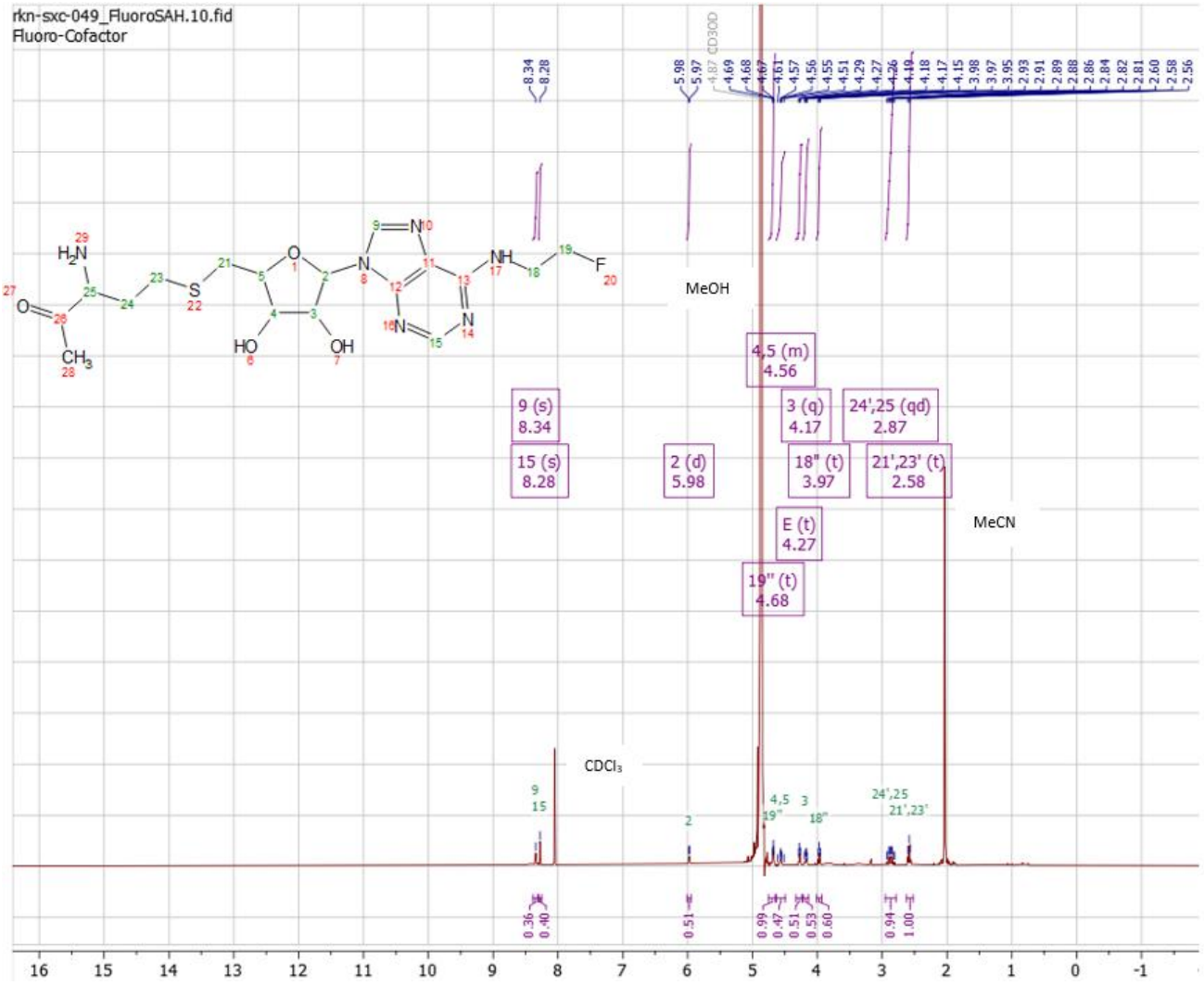


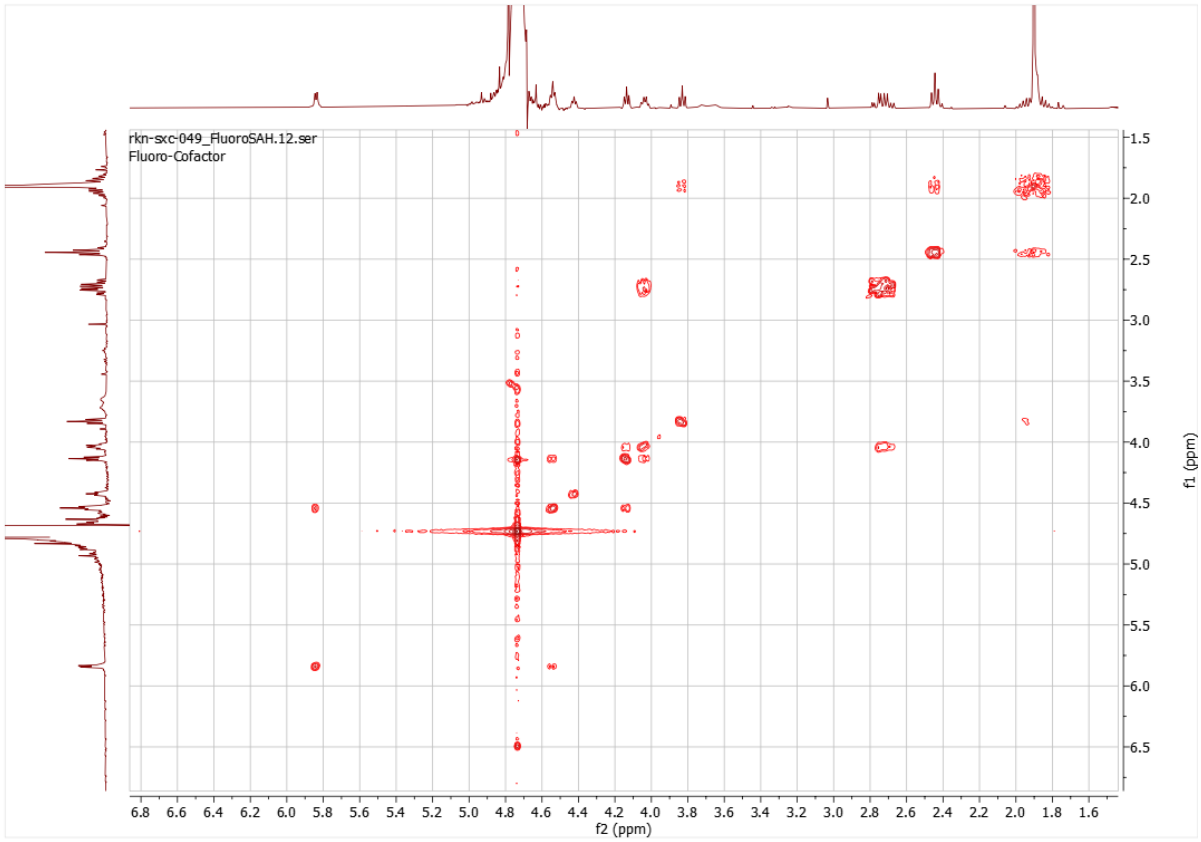
Appendix Figure 6 ^{1}H NMR spectrum (top) and ^{13}C NMR spectrum (bottom) of S-((3,4-dihydroxy-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)-D-homocysteine. Acquired using a Bruker AVII 300 and Bruker AVII 400 and MeOD


RKN-SXC-045.10.fid
Fluoro-I-aden
1D1H.icon MeOD {C:\DATA\RNEELY\2020} RKNeely 10
11 (s)
8.28
17 (s)
8.28
2 (d)
6.02
J(5.26)
3 (t)
4.69
J(5.03)
4 (dd)
4.31
J(5.35, 4.09)
5 (td)
4.07
J(5.77, 4.09)
6' (dd)
3.63
J(10.71, 5.77)
20" (dd)
3.51
J(10.76, 5.78)
21' (m)
4.57
MeOH
MeOH
Acetone
f1 (ppm)



Appendix Figure 7: ^{1}H NMR spectrum of 2-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol (top) and ^{13}C NMR spectrum of 2-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol (bottom). Acquired using a Bruker AVII 300 and Bruker AVII 400 and MeOD

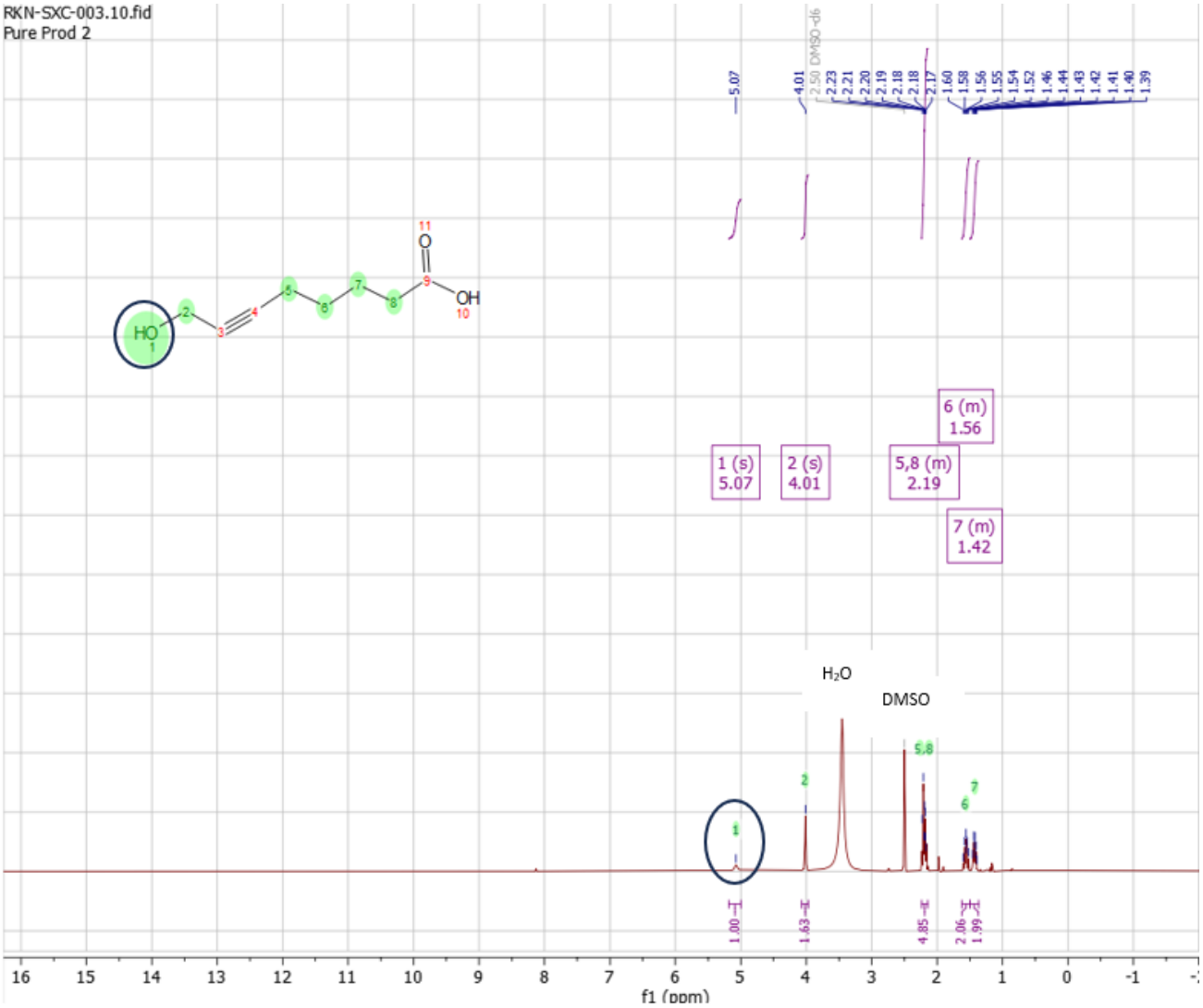

rkn-sxc-049_FluoroSAH.10.fid
Fluoro-Cofactor
MeOH
MeCN
CDCl3
9 (s)
8.34
15 (s)
8.28
2 (d)
5.98
4,5 (m)
4.56
3 (q)
4.17
18" (t)
3.97
24',25 (qd)
2.87
21',23' (t)
2.58
E (t)
4.27
19" (t)
4.68

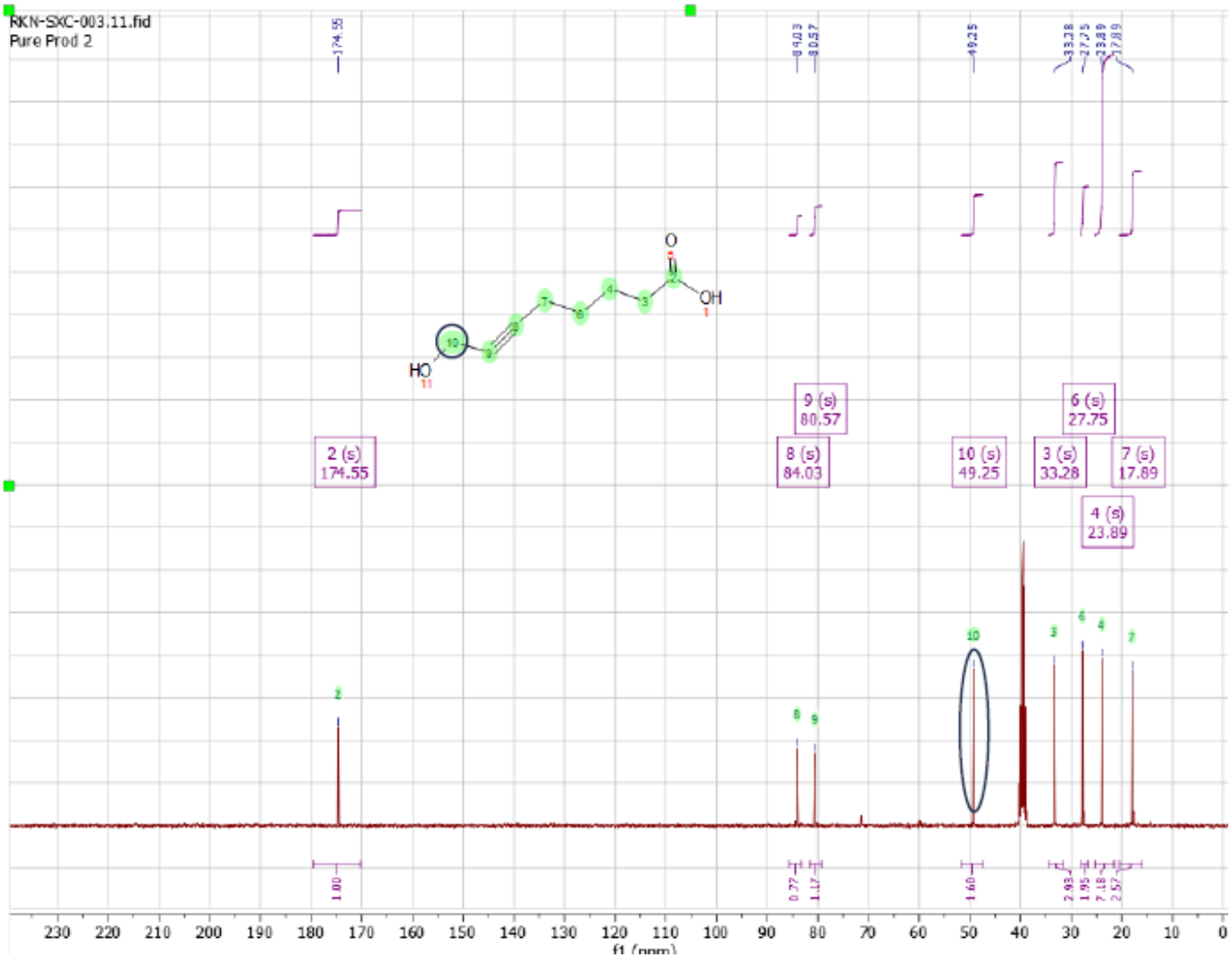


Appendix Figure 8: ^{1}H NMR spectrum of S-((5-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-D-homocysteine. Acquired using a Bruker AVII 300 and MeOD

Appendix Figure 9: ^{1}H NMR spectrum of the brominated Linker with alkyl halide peak highlighted. Acquired using a Bruker AVII 300 and $CDCl_3$.

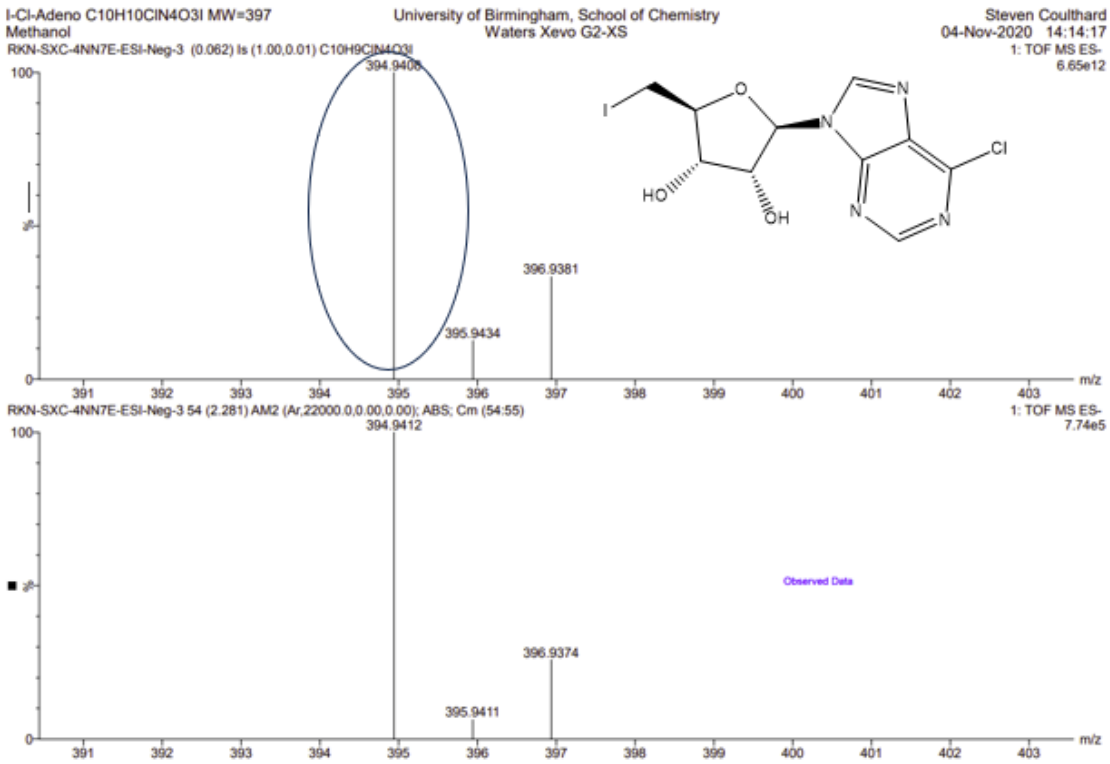
Appendix Figure 10: ^{1}H NMR spectrum of the hydrazine product with the Boc peak highlighted. Acquired using a Bruker AVII 300 and $CDCl_3$.


RKN-SXC-003.10.fid
Pure Prod 2
HO
1
2
3
4
5
6
7
8
9
11
O
OH
10
5.07
4.01
2.50 DMSO-d6
2.23
2.21
2.20
2.19
2.18
2.18
2.17
1.60
1.58
1.56
1.55
1.54
1.52
1.46
1.44
1.43
1.42
1.41
1.40
1.39
1 (s)
5.07
2 (s)
4.01
5,8 (m)
2.19
6 (m)
1.56
7 (m)
1.42
H2O
DMSO
5,8
2
7
6
1
1.00
1.63
4.85
2.06
1.99
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
f1 (ppm)

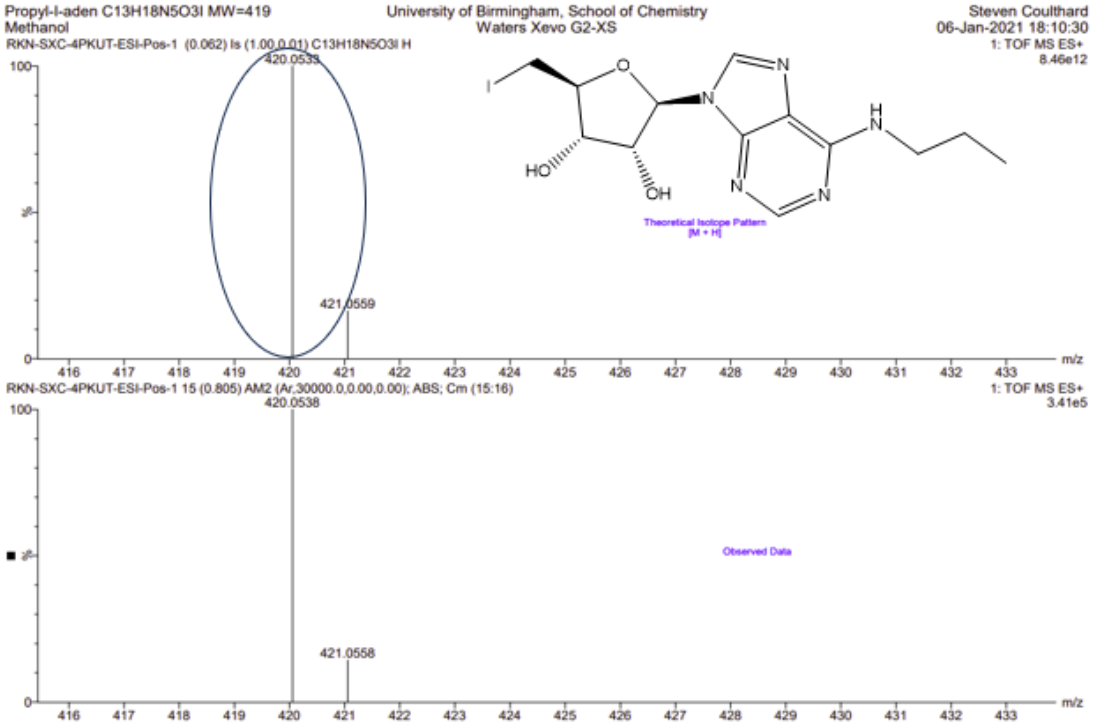


Appendix Figure 11: 13C NMR spectrum of alkynyl alcohol, hydroxyalkyl peak is highlighted which proves the formation of the product. Acquired using a Bruker AVII 300 and Bruker AVII 400 and DMSO.

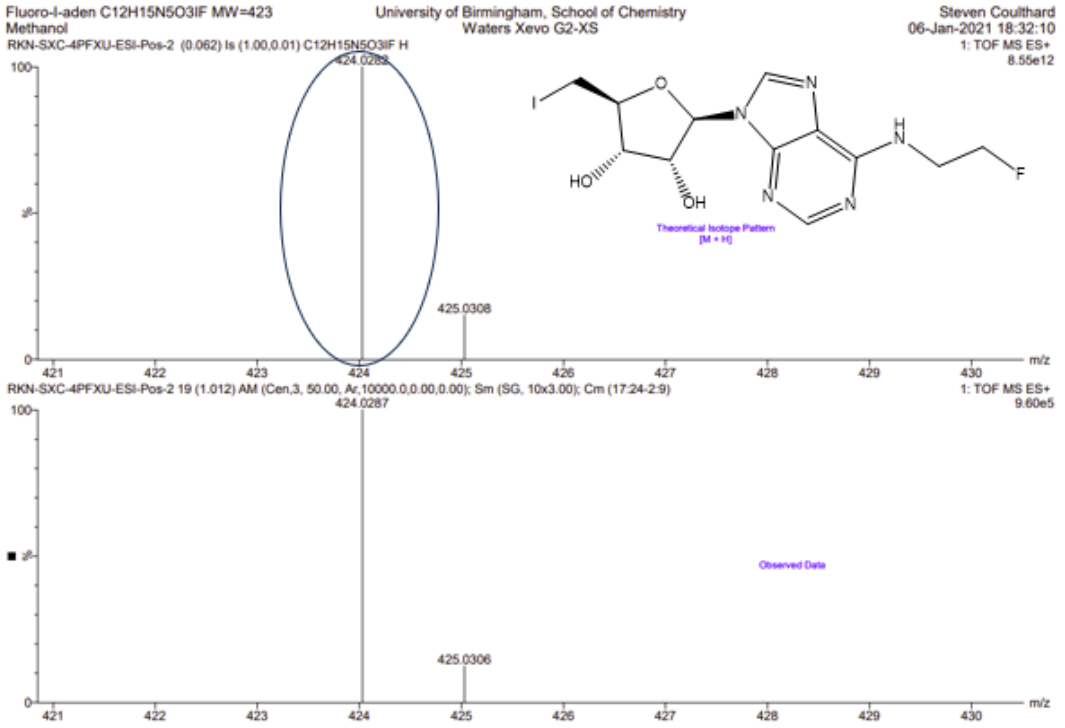
7.4 Mass spectra



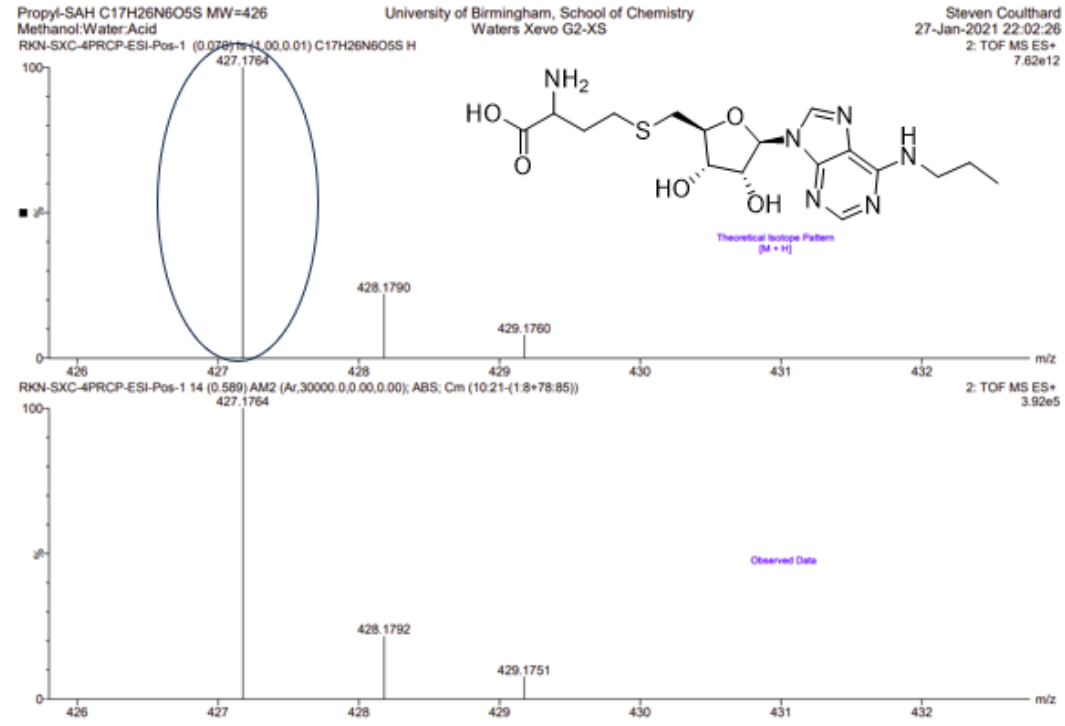
Appendix Figure 12: Mass spectrum of the 2-(6-chloro-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol, the molecular ion peak has been highlighted proving the synthesis of the product.



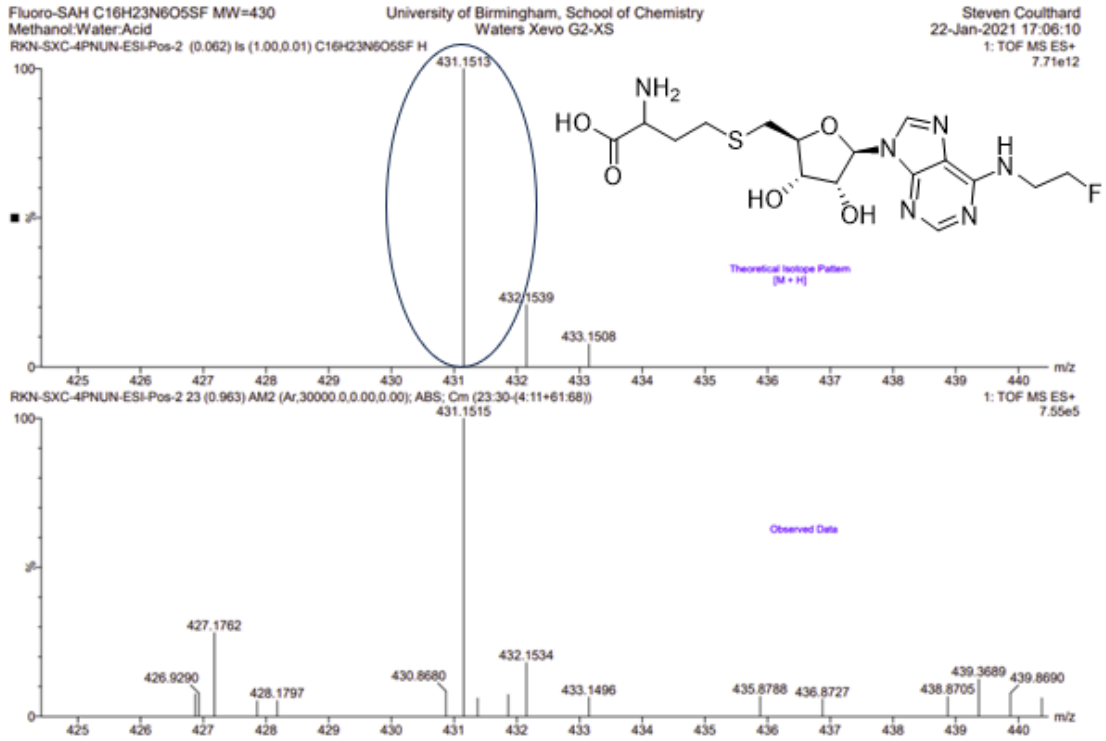
Appendix Figure 13: Mass spectrum of the 2-(iodomethyl)-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-3,4-diol, the molecular ion peak has been highlighted to prove this has been made



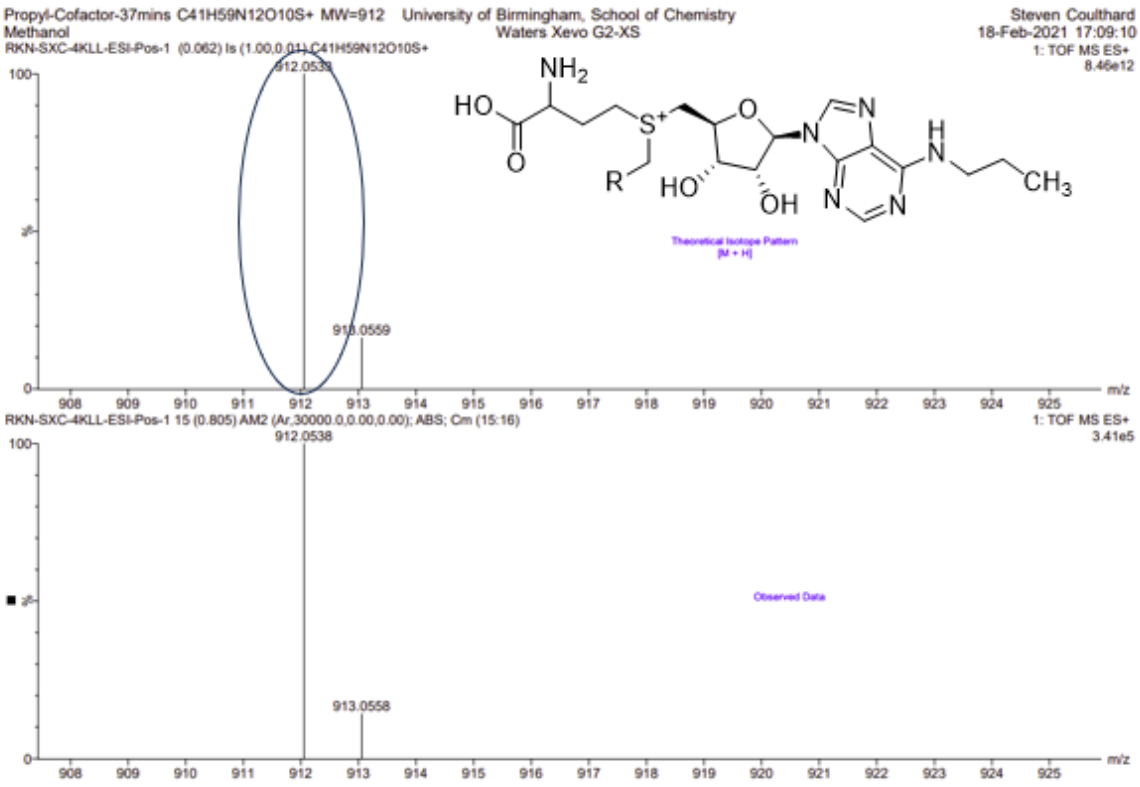
Appendix Figure 14: Mass spectrum of the 2-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol, the molecular ion peak has been highlighted to prove this has been made., the molecular ion peak has been highlighted to prove this has been made.



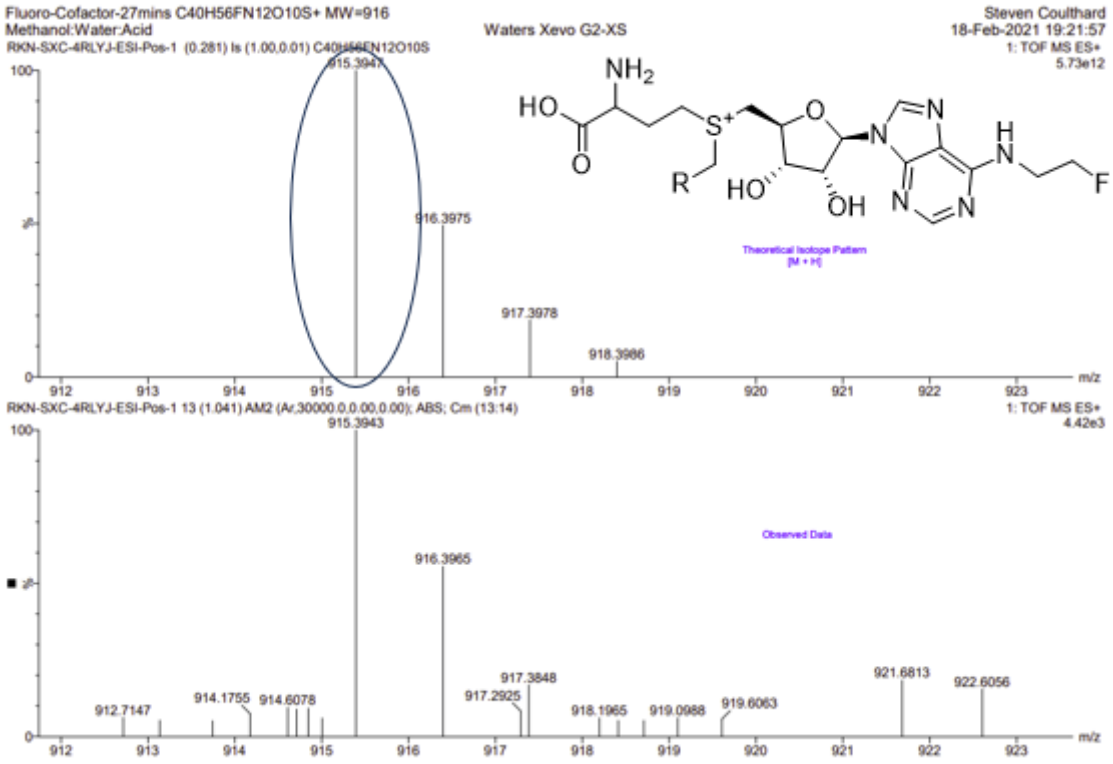
Appendix Figure 15: Mass spectrum of the S-((3,4-dihydroxy-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)-L-homocysteine, the molecular ion peak has been highlighted to prove this has been made.



Appendix Figure 16: Mass Spectrum of the S-((5-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-L-homocysteine, the molecular ion peak has been highlighted to prove this has been made.



*Appendix Figure 17: Mass spectrum of the **4** cofactor with molecular ion peak highlighted.*



*Appendix Figure 18: Mass spectrum of the **3** cofactor with molecular ion peak highlighted.*

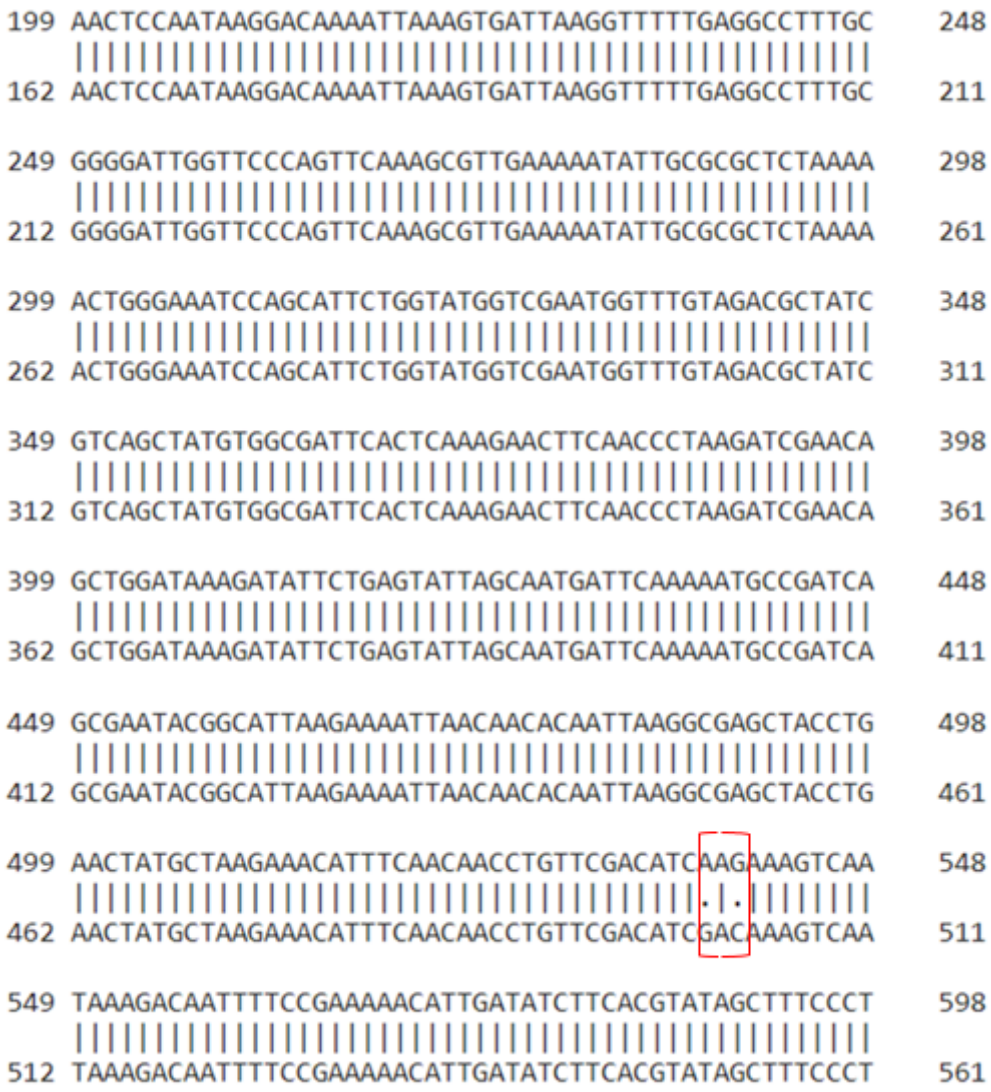
7.5 Sequence Results

Below are the sequencing results for the plasmids that were generated for each of the mutant MTases. The top row represents the wild type M.MpeI sequence and the bottom row represents the mutant plasmid sequence. Lines show where the sequence is the same whereas dots show the differences between the sequences.

7.41 K115A

Appendix Figure 19: Sequence results for K115A.

7.42 K115D



Appendix Figure 20: Sequence results for K115D.

7.43 K115E

```
AAAATATTGCGCGCTCTAAAAACTGGGAAATCCAGCATTCTGGTATGGTCGAATGGTTTG 300
AAAATATTGCGCGCTCTAAAAACTGGGAAATCCAGCATTCTGGTATGGTCGAATGGTTTG 300
************************************************************

TAGACGCTATCGTCAGCTATGTGGCGATTCACTCAAAGAACTTCAACCCTAAGATCGAAC 360
TAGACGCTATCGTCAGCTATGTGGCGATTCACTCAAAGAACTTCAACCCTAAGATCGAAC 360
************************************************************

AGCTGGATAAAGATATTCTGAGTATTAGCAATGATTCAAAAATGCCGATCAGCGAATACG 420
AGCTGGATAAAGATATTCTGAGTATTAGCAATGATTCAAAAATGCCGATCAGCGAATACG 420
************************************************************

GCATTAAGAAAATTAACAACACAATTAAGGCGAGCTACCTGAACTATGCTAAGAAACATT 480
GCATTAAGAAAATTAACAACACAATTAAGGCGAGCTACCTGAACTATGCTAAGAAACATT 480
************************************************************

TCAACAACCTGTTCGACATCAAGAAAGTCAATAAAGACAATTTTCCGAAAAACATTGATA 540
TCAACAACCTGTTCGACATCGAGAAAGTCAATAAAGACAATTTTCCGAAAAACATTGATA 540
********************************************************

TCTTCACGTATAGCTTTCCCTGTGCAGATCTCTCGGTGCAAGGTCTTCAGAAAGGCATCG 600
TCTTCACGTATAGCTTTCCCTGTGCAGATCTCTCGGTGCAAGGTCTTCAGAAAGGCATCG 600
************************************************************

ATAAAGAACTGAACACACGTTCTGGTCTGCTGTGGGAGATTGAACGCATCTTAGAAGAAA 660
ATAAAGAACTGAACACACGTTCTGGTCTGCTGTGGGAGATTGAACGCATCTTAGAAGAAA 660
************************************************************

TTAAGAATTCTTTTAGTAAGGAAGAGATGCCGAAATATCTTCTGATGGAAAACGTGAAAA 720
TTAAGAATTCTTTTAGTAAGGAAGAGATGCCGAAATATCTTCTGATGGAAAACGTGAAAA 720
************************************************************
```

Appendix Figure 21: Sequence results for K115E.

7.44 K115H

Appendix Figure 22: Sequence results for *K115H.*

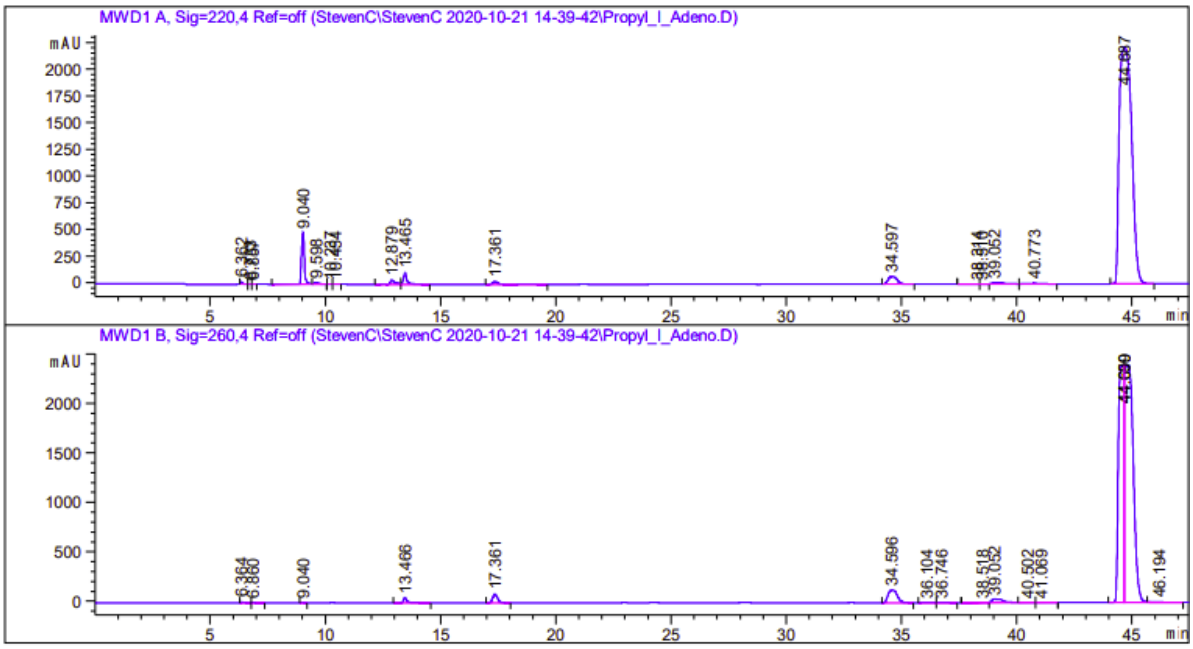
7.45 K115Q

Appendix Figure 23: Sequence results for K115Q.

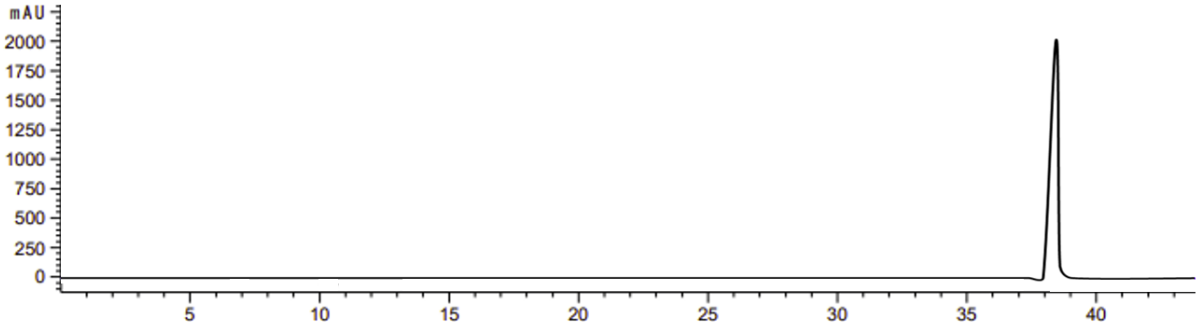
7.6 Traces:

7.61HPLC Traces

7.62 2-(iodomethyl)-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-3,4-diol



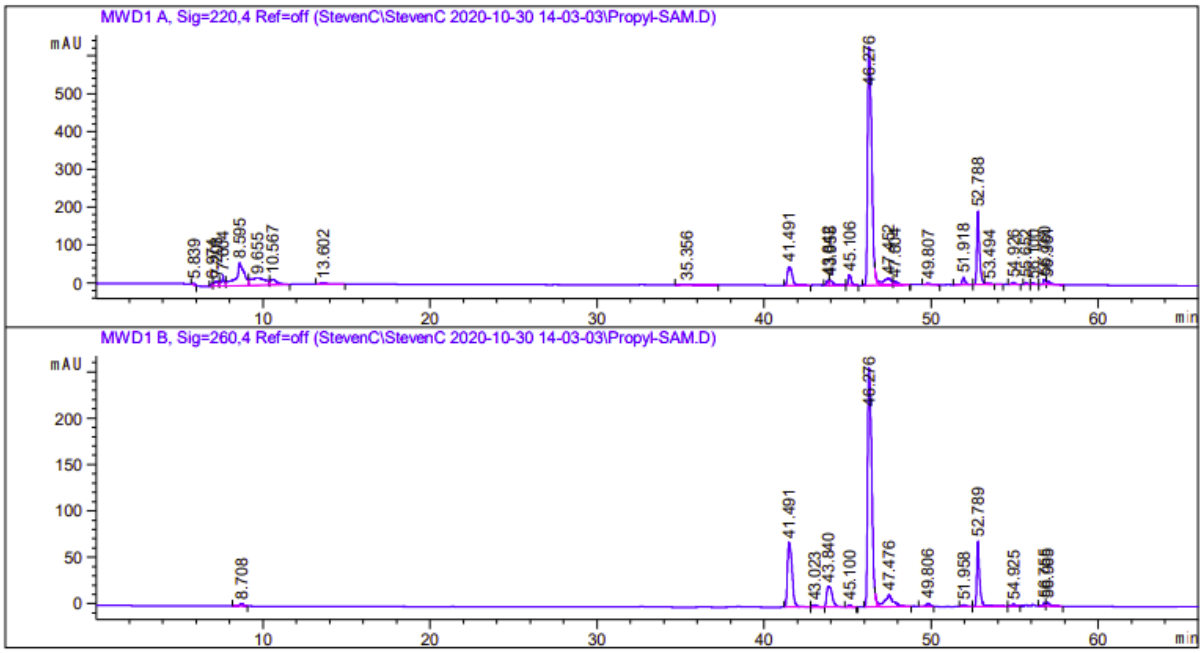
Appendix Figure 24: HPLC trace of 2-(iodomethyl)-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-3,4-diolproduct peak at 45 mins.



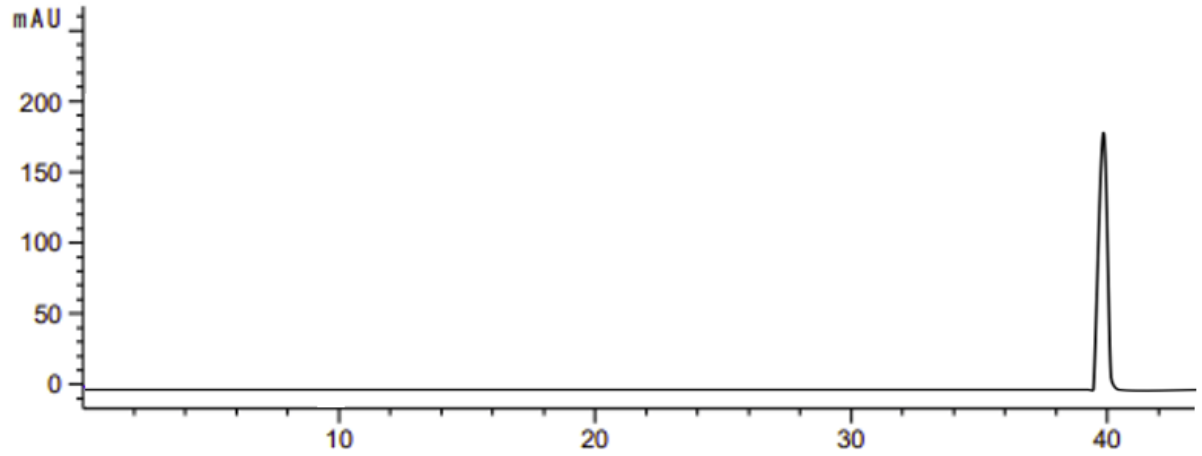
Appendix Figure 25: Analytical HPLC trace of 2-(iodomethyl)-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-3,4-diolproduct peak at 38 mins

.

7.63 S-((3,4-dihydroxy-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)-D-homocysteine

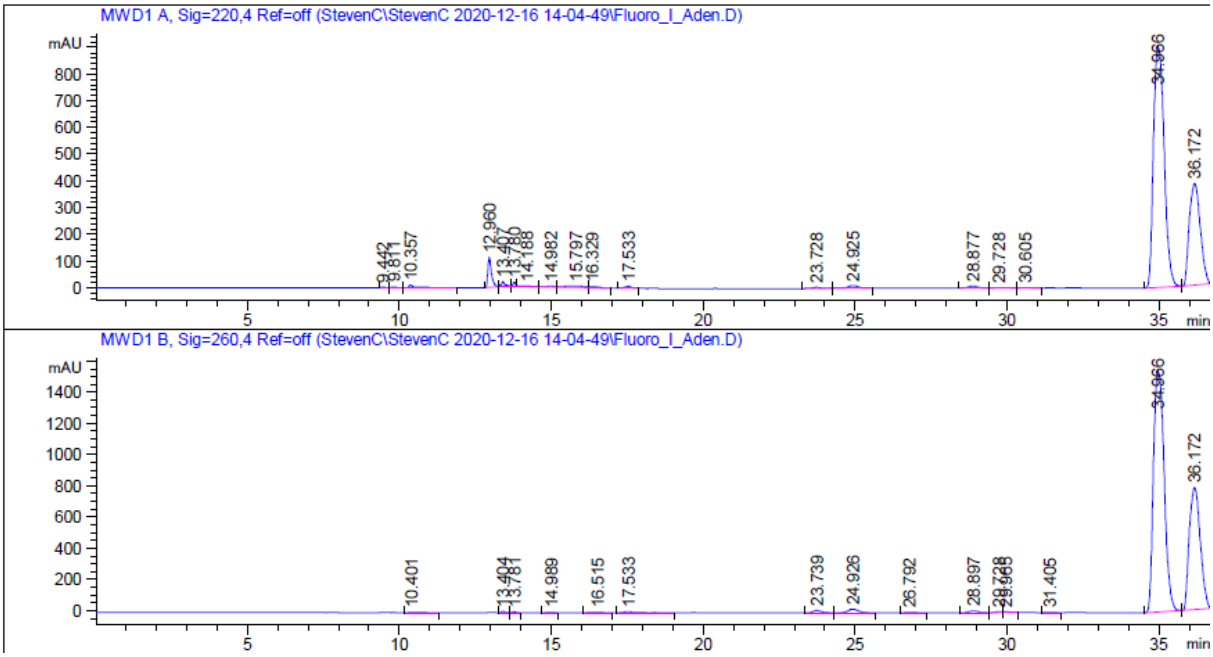


Appendix Figure 26: HPLC trace of S-((3,4-dihydroxy-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)-D-homocysteineproduct peak at 47 mins

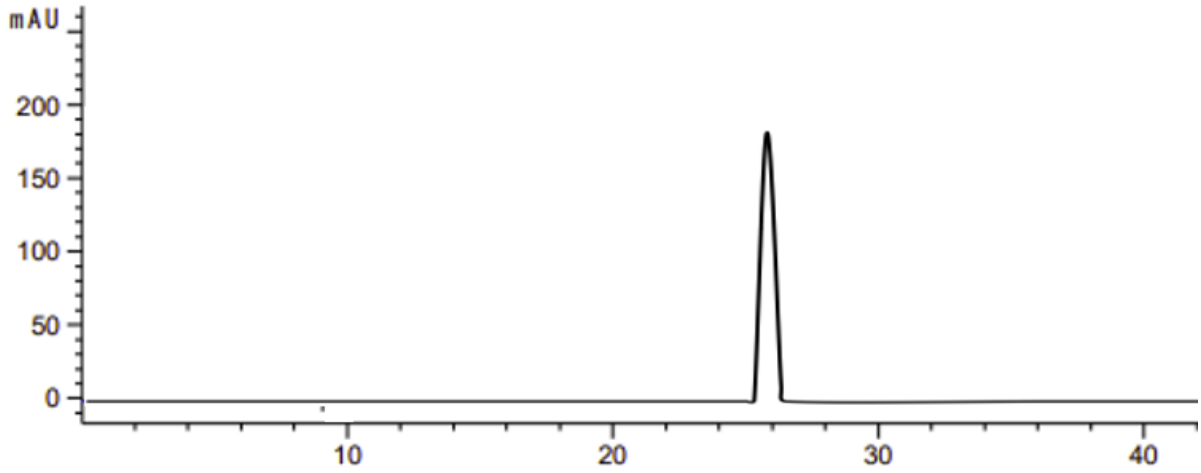


Appendix Figure 27: Analytical HPLC trace of S-((3,4-dihydroxy-5-(6-(propylamino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)-D-homocysteineproduct peak at 40 mins.

7.64 of 2-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol

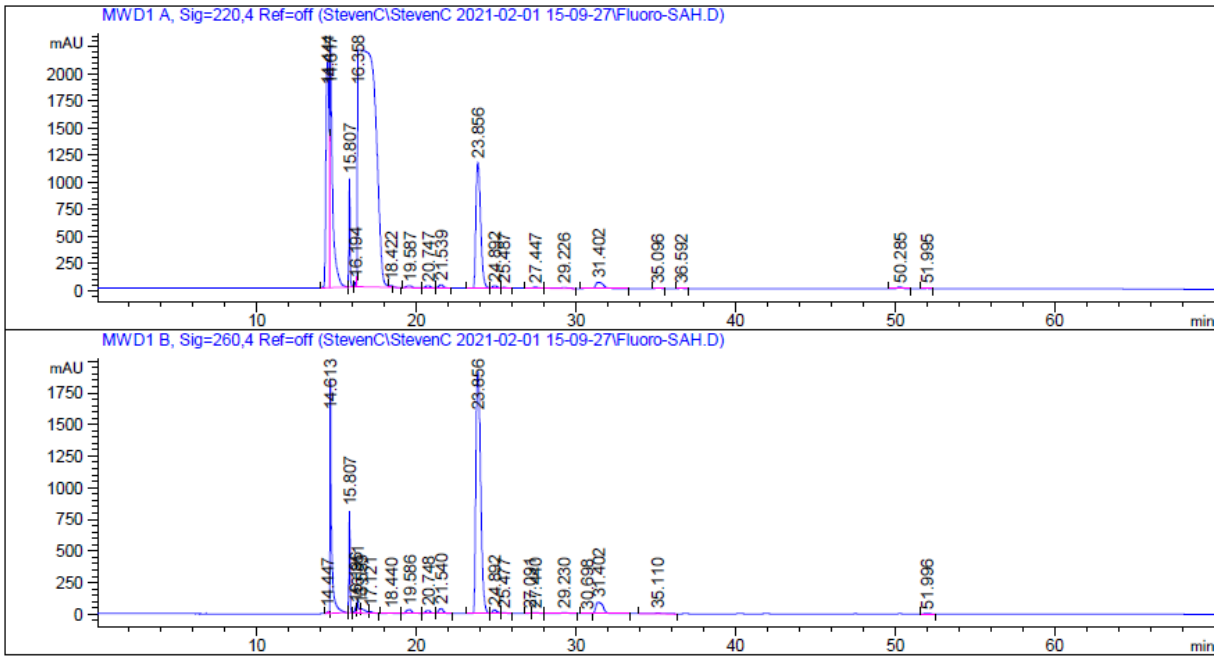


Appendix Figure 28: HPLC trace of 2-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol product peak at 34 mins.

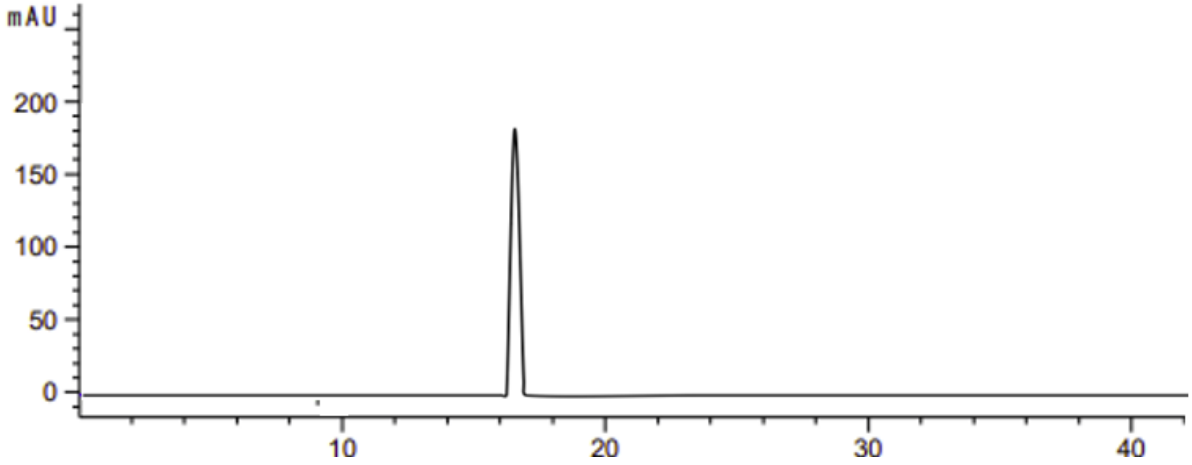


Appendix Figure 29: HPLC trace of 2-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol product peak at 28 mins.

7.65 S-((5-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-D-homocysteine

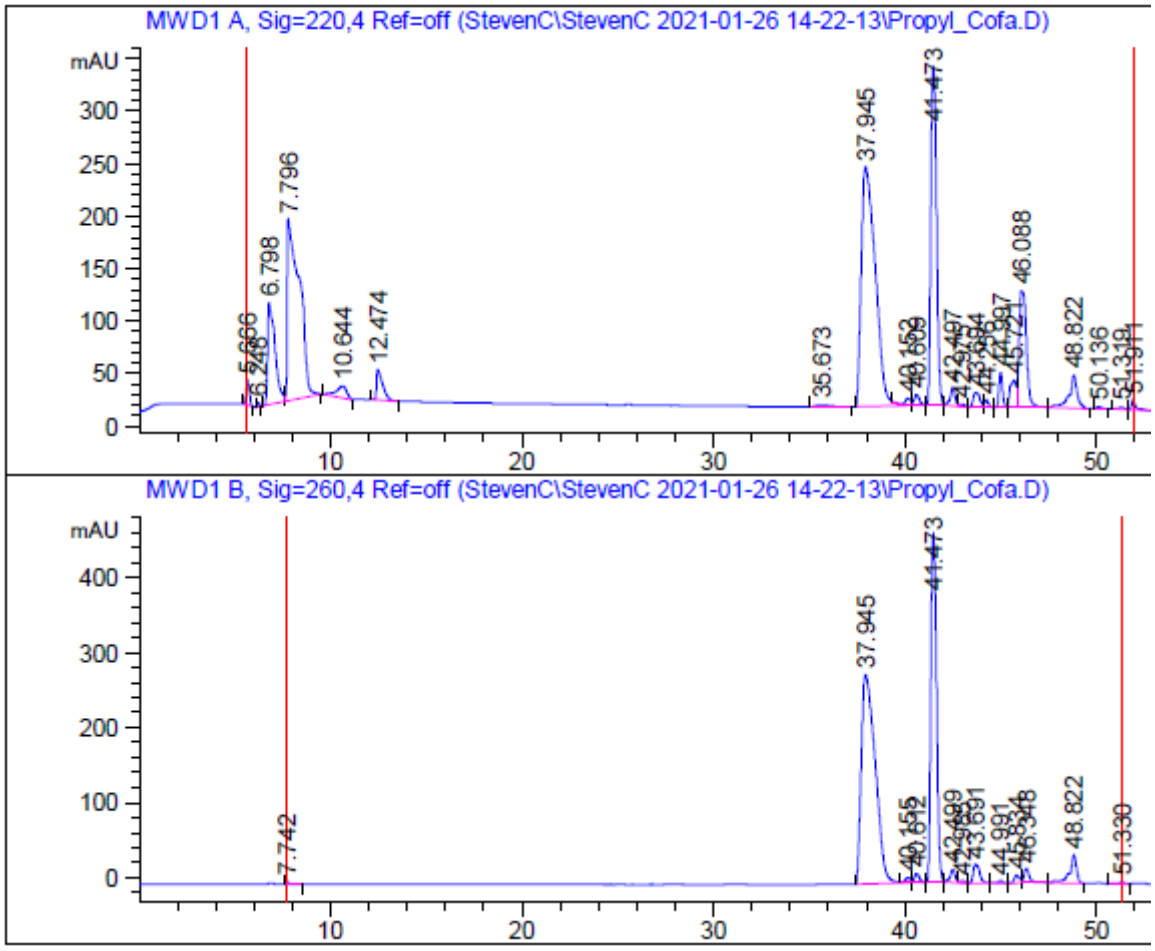


Appendix Figure 28: HPLC trace of S-((5-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-D-homocysteine product peak at 23 mins

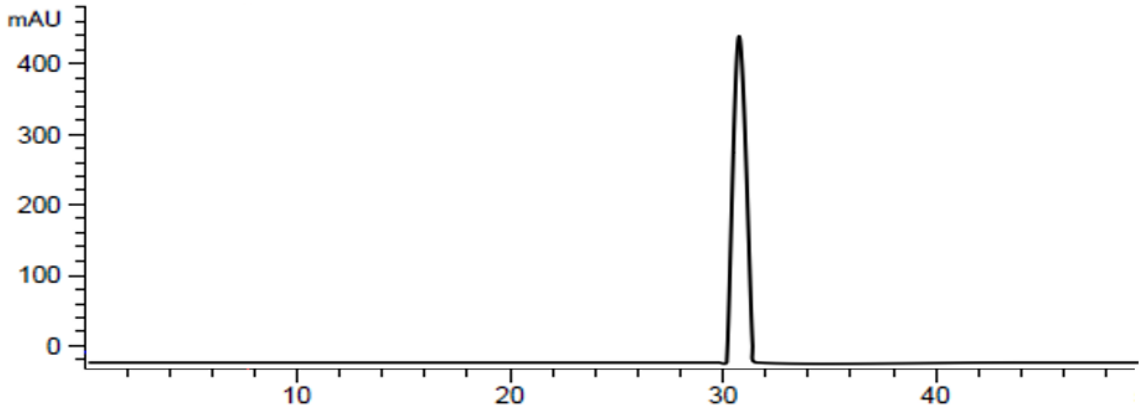


Appendix Figure 29: HPLC trace of S-((5-(6-((2-fluoroethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-D-homocysteine product peak at 23 mins.

7.66 **4**

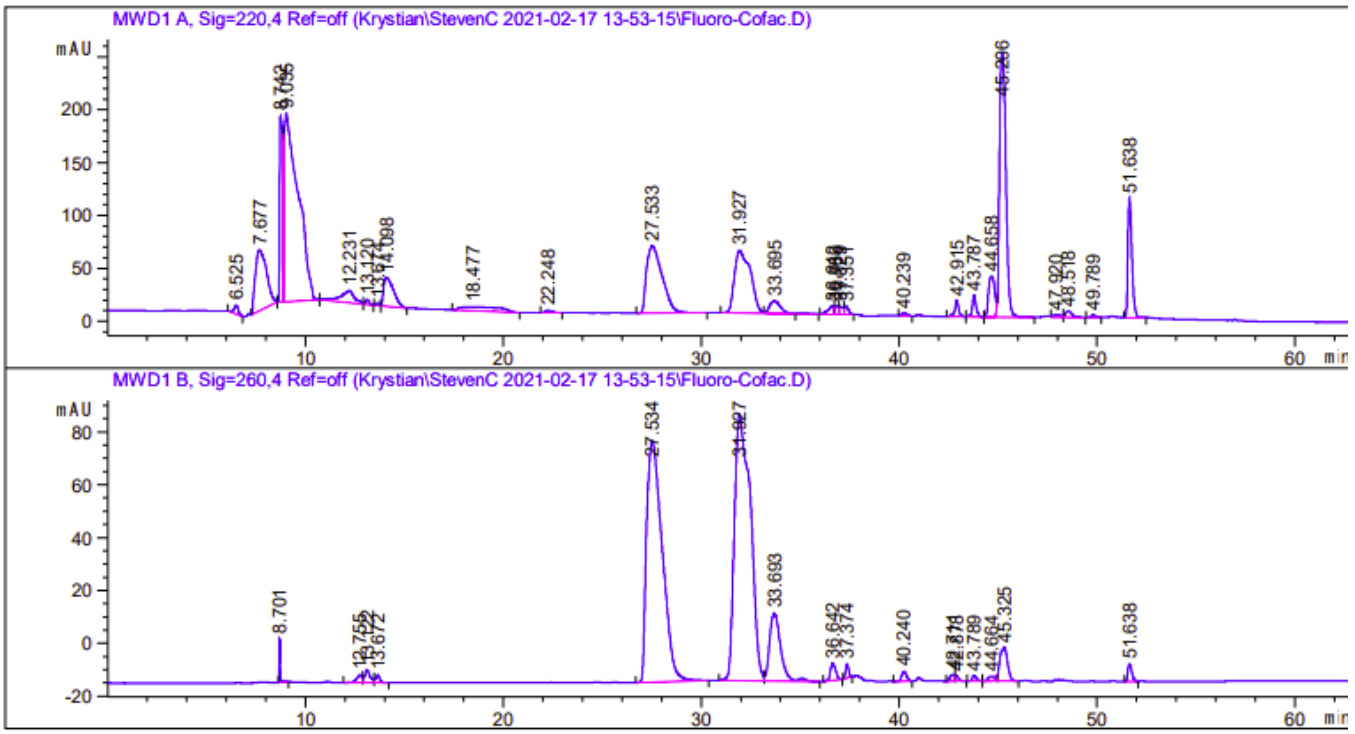


*Appendix Figure 30: HPLC trace of **4** product peaks at 37 and 41 mins*

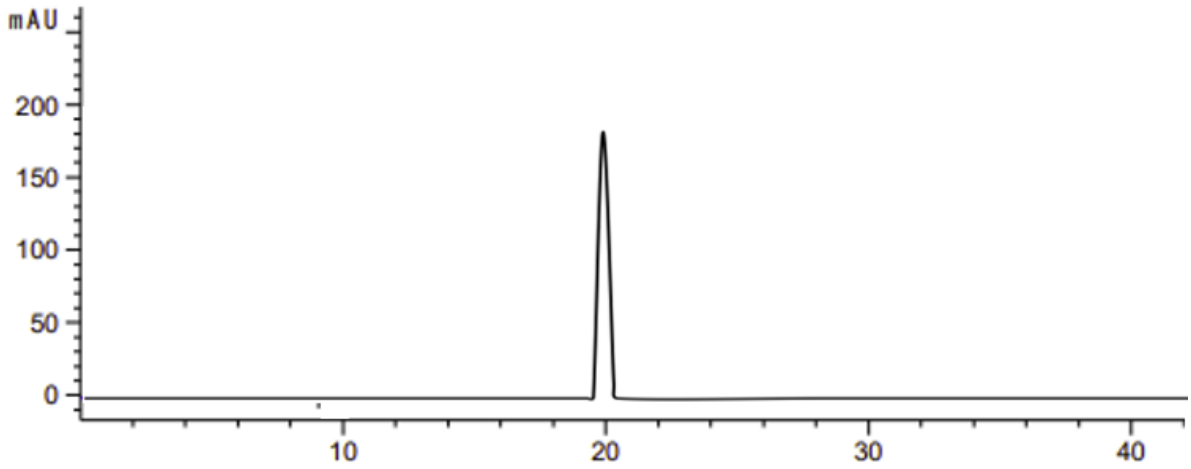


*Appendix Figure 31: Analytical HPLC trace of **4** product peak at 31 mins*

7.67 ***3***



Appendix Figure 32: HPLC trace of ***3*** *product peaks at 28 and 32 mins*

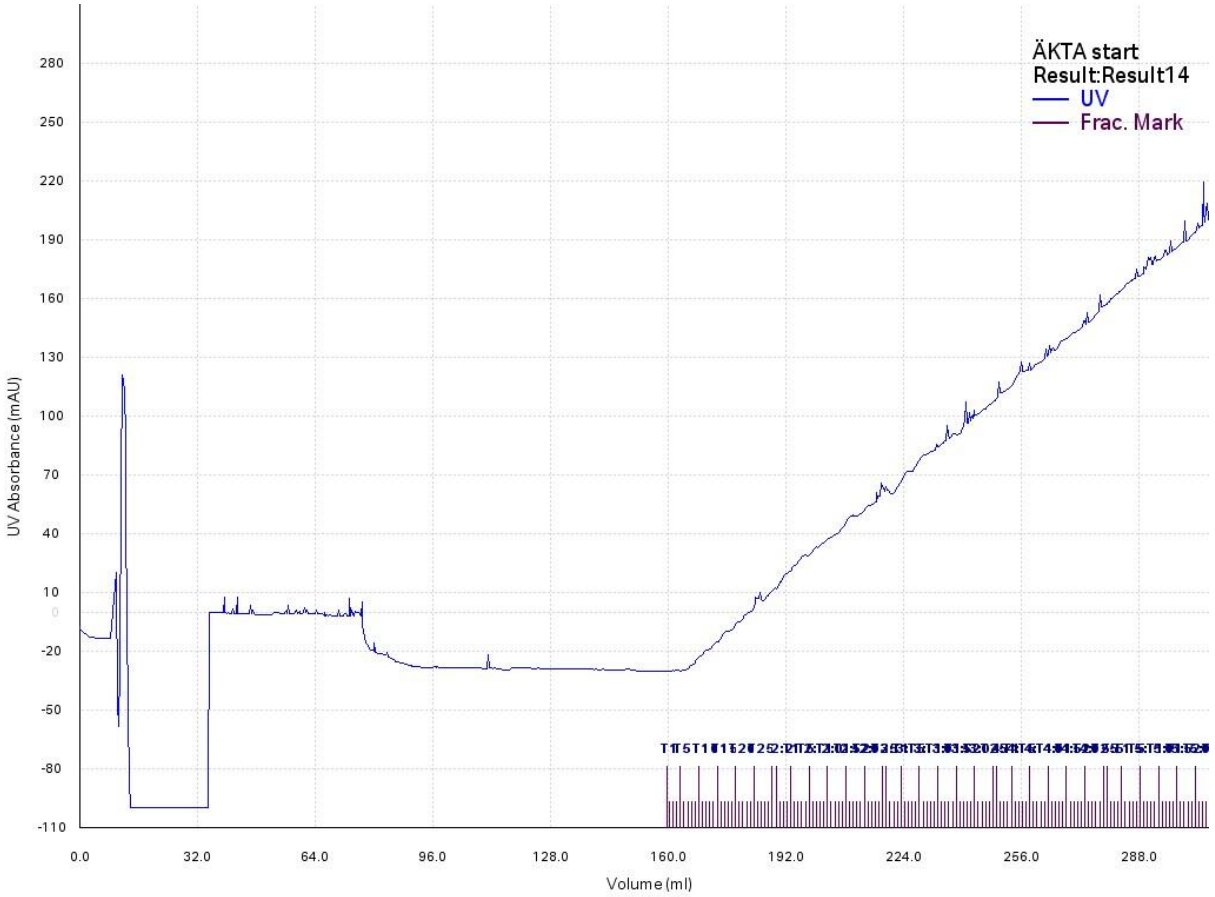


Appendix Figure 33 Analytical HPLC trace of ***3*** *product peak at 20 mins*

7.65 Protein Traces

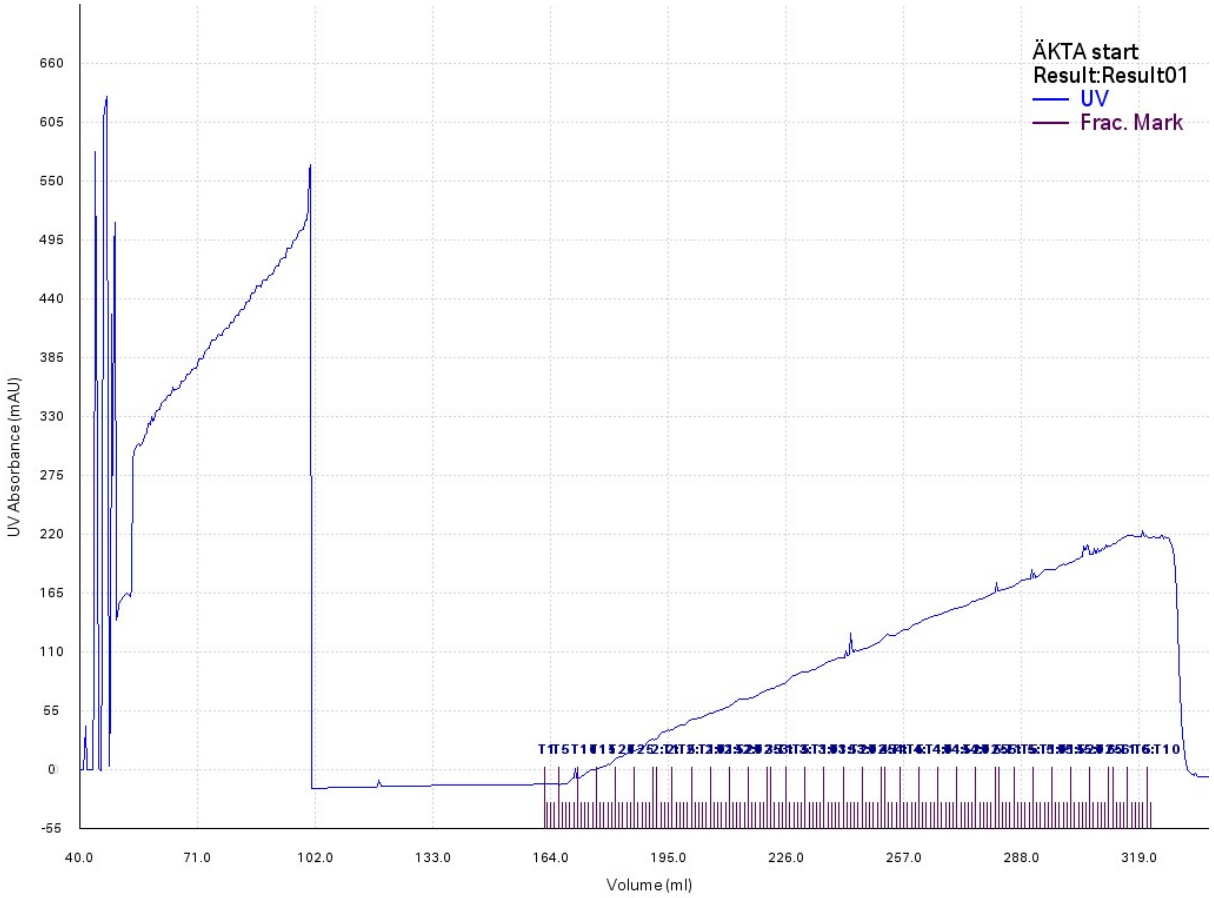
Below are the Cytiva traces from the protein purification system for each of the mutant MTase from the point of injection of the lysed cells.

7.56 K115A



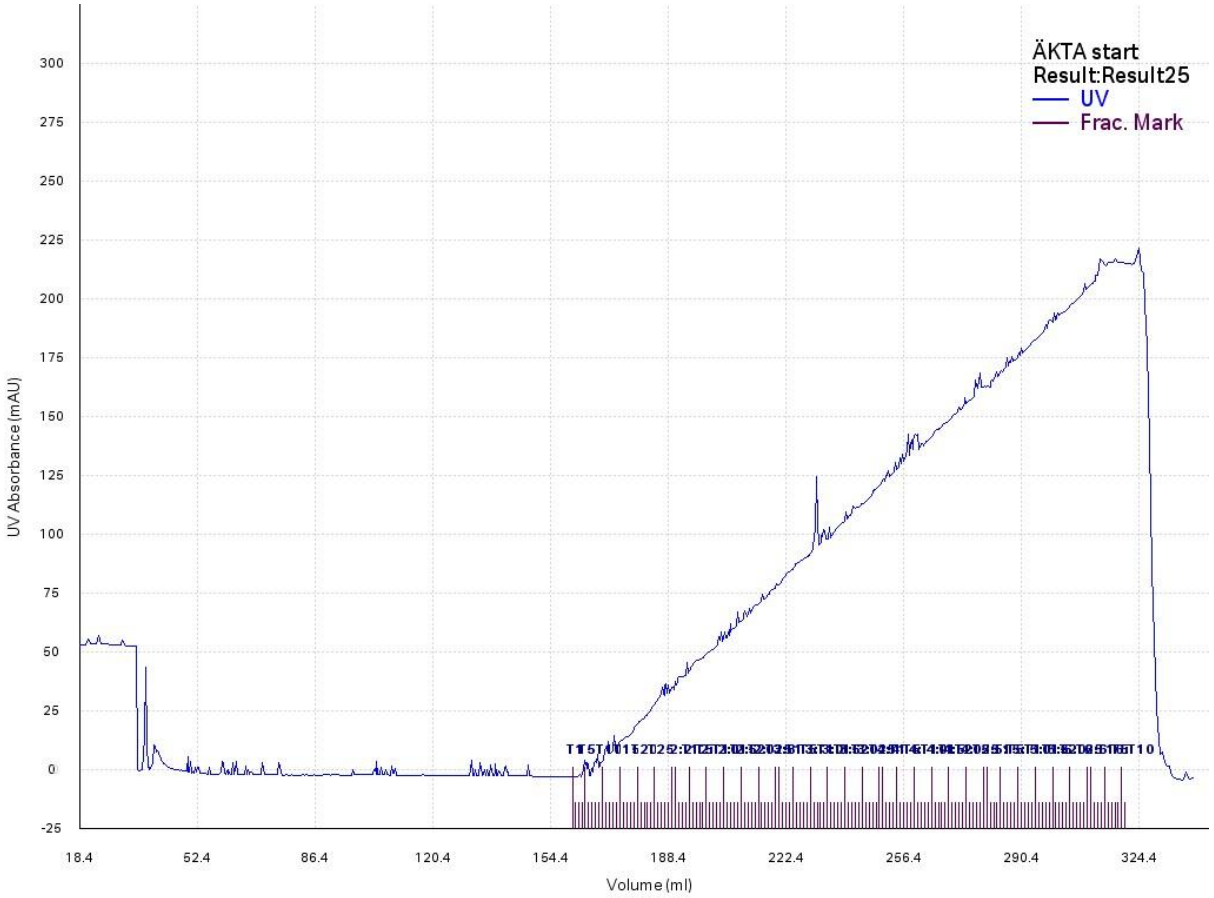
Appendix Figure 34: Cytiva trace of K115A purification.

7.67 K115E



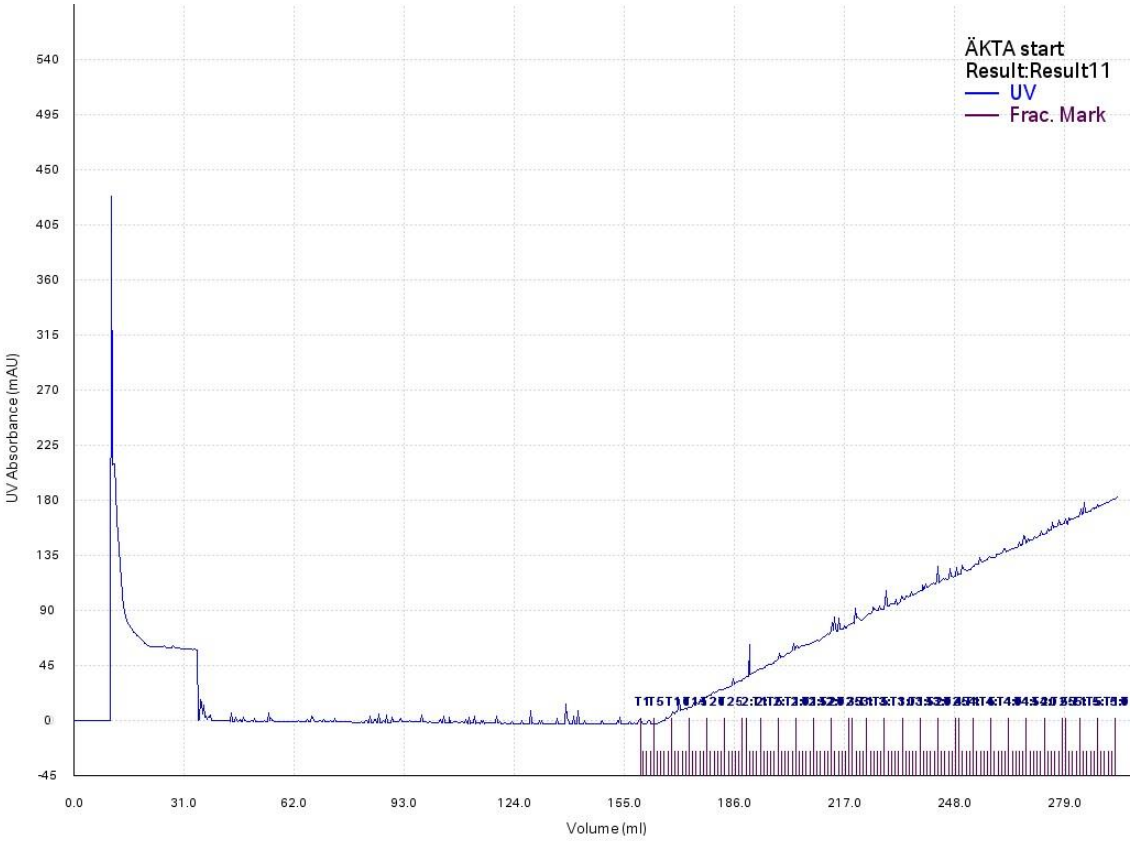
Appendix Figure 35: Cytiva trace of K115E purification.

7.68 K115H



Appendix Figure 36: Cytiva trace of K115H purification.

7.69 K115Q



Appendix Figure 37: Cytiva trace of K115Q purification.

7.7 Protein Concentrations

Below are the protein concentrations that were collected from the protein purification system for each of the mutant MTases. This was calculated by nanodrop using the extinction coefficient as 86320 M^{-1} cm^{-1}.

7.71 K115A

No.	: 21
Sample Name	: MpeI_A_pool_1
Measurement Mode	: Protein Quantitation
Method	: OD280
Label	: Cy3
DateTime	: 22/05/26 12:39:23
User Name	: Steve

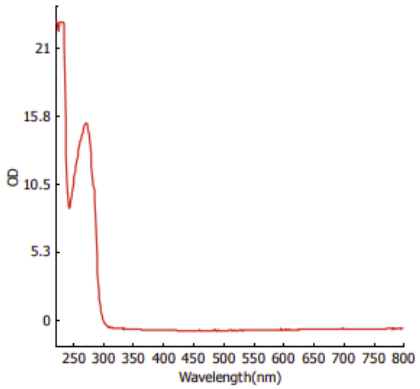
Protein Conc : 2.376E-04 M

Protein Conc : 10952.63 µg/mL

Label Conc : -5.049E-06 M

Labeling Ratio : -0.021

Item	Result
ε280(M-1cm-1)	5.056E+04
MW	46105.9
OD280	11.950
ODLabel	-0.757
Pathlength (mm)	0.694
Dilution	1.000



Appendix Figure 38: K115A protein concentration

7.72 K115E

No.	: 3
Sample Name	: E_130_MINS
Measurement Mode	: Protein Quantitation
Method	: OD280
Label	: Cy3
DateTime	: 22/04/27 13:32:45
User Name	: Steve

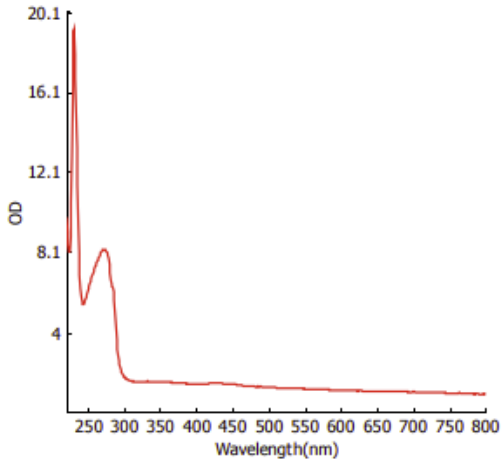
Protein Conc : 1.377E-04 M

Protein Conc : 6348.24 µg/mL

Label Conc : 8.142E-06 M

Labeling Ratio : 0.059

Item	Result
ε280(M-1cm-1)	5.056E+04
MW	46105.9
OD280	7.059
ODLabel	1.221
Pathlength (mm)	0.694
Dilution	1.000



Appendix Figure 39: K115E protein concentration

7.73 K115H

No.	: 39
Sample Name	: MpeI_H_pool1
Measurement Mode	: Protein Quantitation
Method	: OD280
Label	: Cy3
DateTime	: 22/05/26 12:54:29
User Name	: Steve

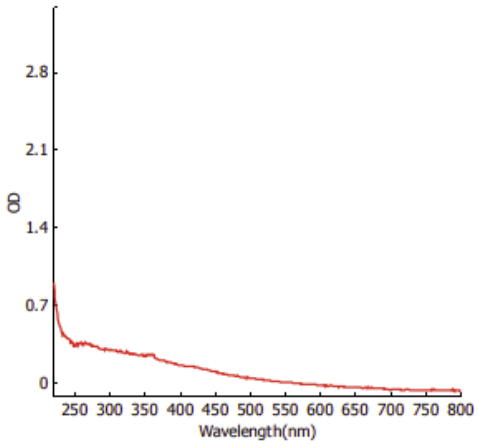
Protein Conc : 6.337E-06 M

Protein Conc : 292.17 µg/mL

Label Conc : 1.609E-08 M

Labeling Ratio : 0.003

Item	Result
ε280(M-1cm-1)	5.056E+04
MW	46105.9
OD280	0.321
ODLabel	0.002
Pathlength (mm)	0.694
Dilution	1.000



Appendix Figure 40: K115H protein concentration

7.74 K115Q

No.	: 33
Sample Name	: Q_pool1
Measurement Mode	: Protein Quantitation
Method	: OD280
Label	: Cy3
DateTime	: 22/05/12 12:45:08
User Name	: Steve

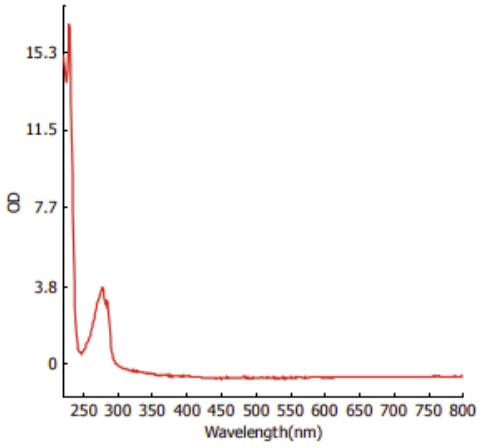
Protein Conc : 6.766E-05 M

Protein Conc : 3119.48 µg/mL

Label Conc : -4.355E-06 M

Labeling Ratio : -0.064

Item	Result
ε280(M-1cm-1)	5.056E+04
MW	46105.9
OD280	3.369
ODLabel	-0.653
Pathlength (mm)	0.694
Dilution	1.000

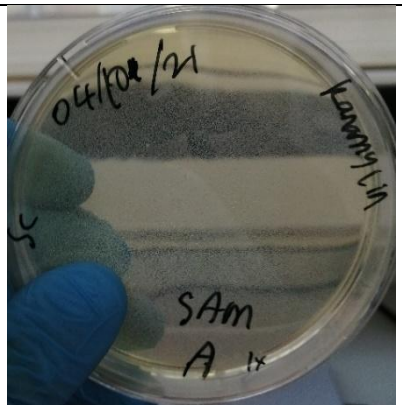
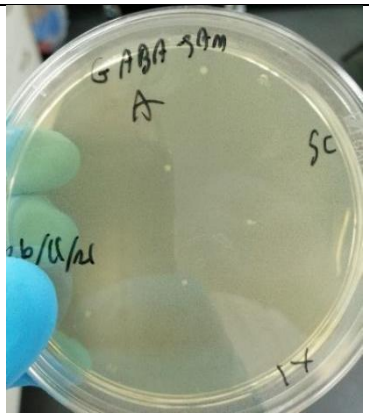
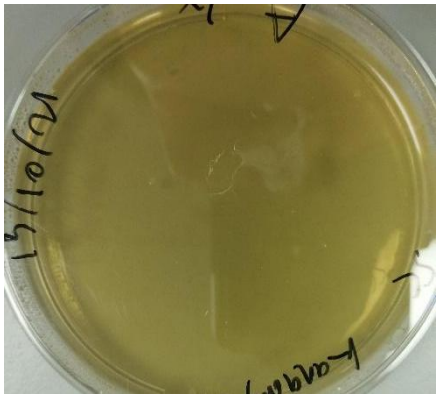
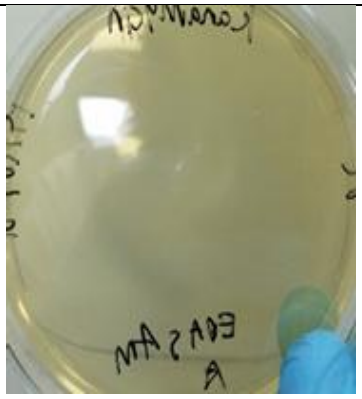
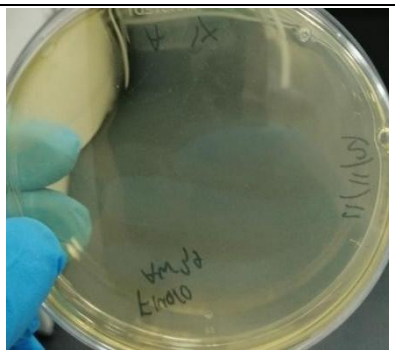
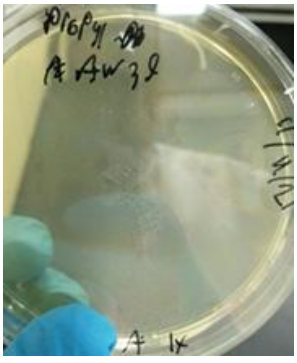


Appendix Figure 41: K115H protein concentration

7.8 Bacterial Images

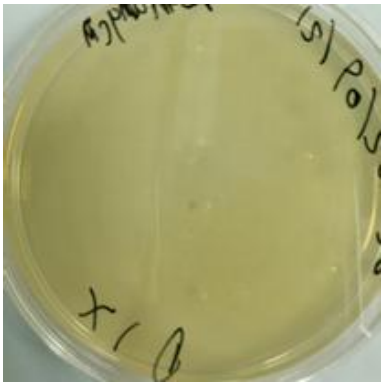
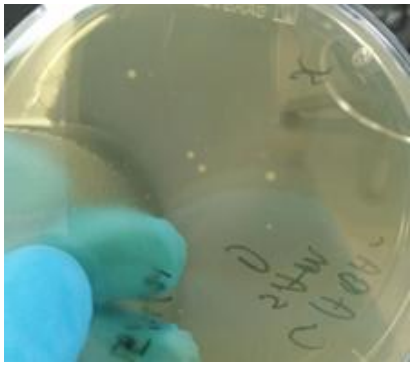
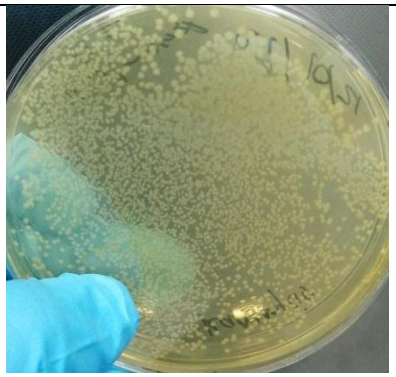
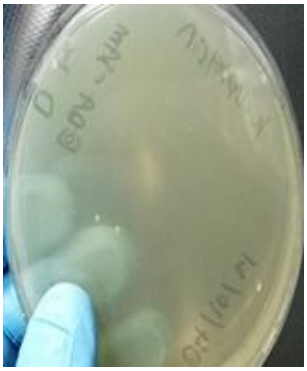
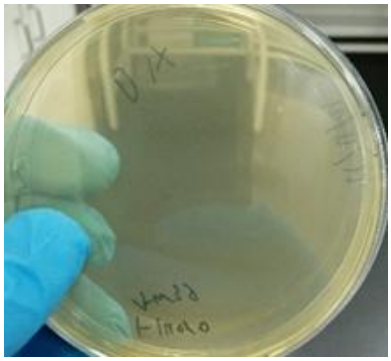
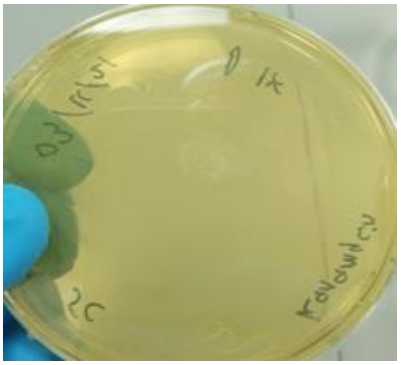
7.81 K115A

Appendix table 2: K115A bacterial results

 SAM	 γ-aminobutyric acid
 **2**	 **1**
 **3**	 **4**

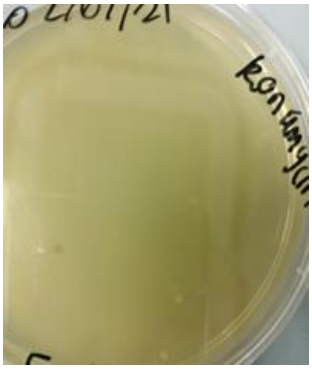
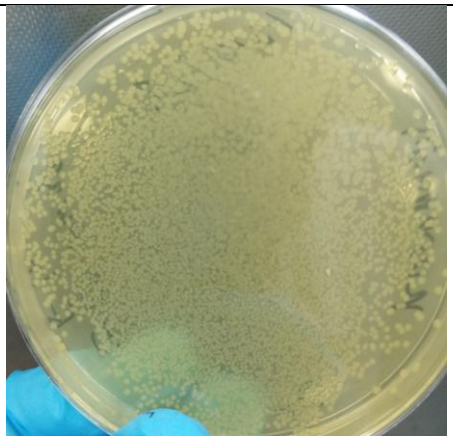
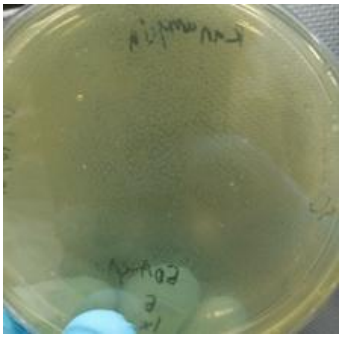
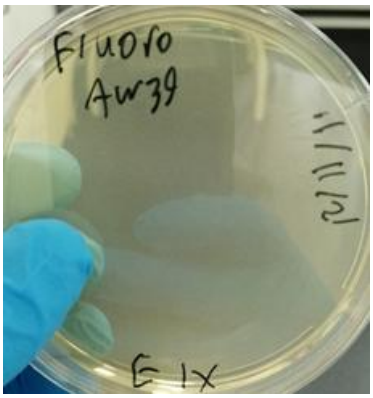
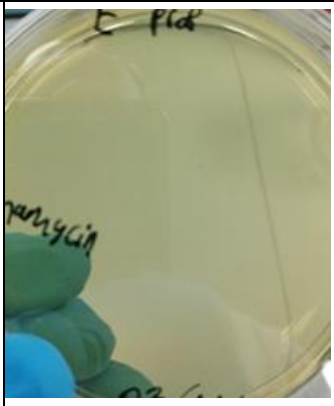
7.82 K115D

Appendix table 3: K115D bacterial results.

 SAM	 γ-aminobutyric acid
 **2**	 **1**
 **3**	 **4**

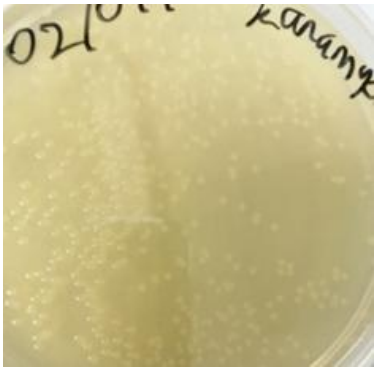
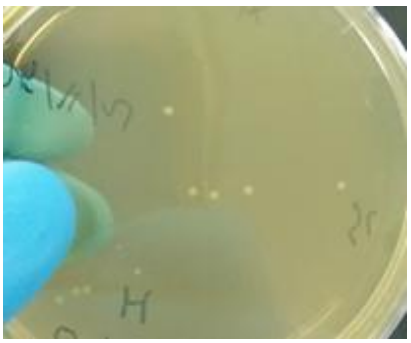
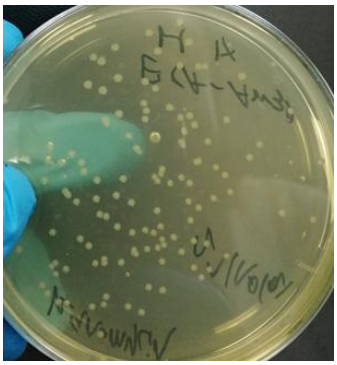
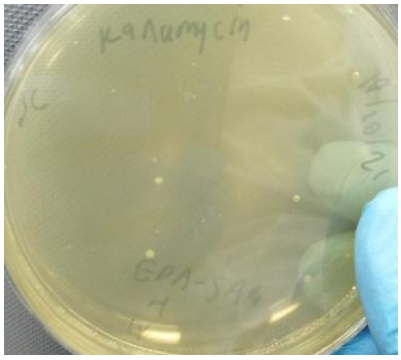
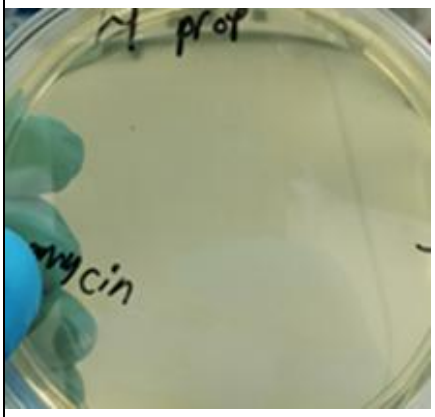
7.83 K115E

Appendix table 3: K115E bacterial results

<table>
<tr>
<td> 
SAM</td>
<td> 
γ-aminobutyric acid</td>
</tr>
<tr>
<td> 
2</td>
<td> 
1</td>
</tr>
<tr>
<td> 
3</td>
<td> 
4</td>
</tr>
</table>

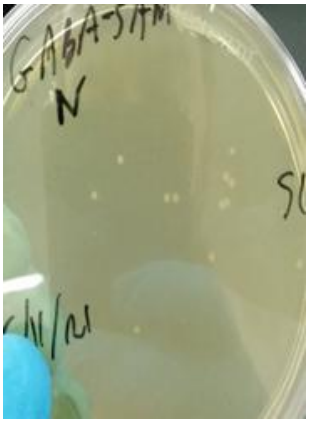
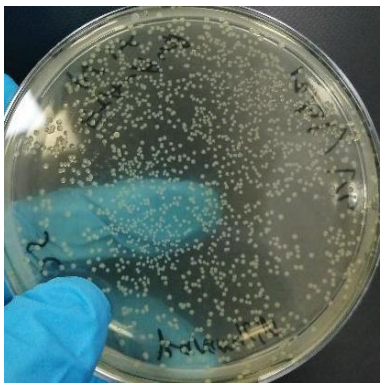
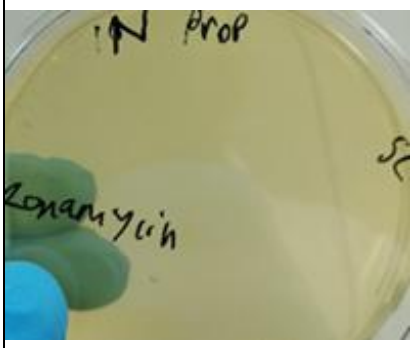
7.84 K115H

Appendix Table 3: K115H bacterial results

 SAM	 γ-aminobutyric acid
 **2**	 **1**
 **3**	 **4**

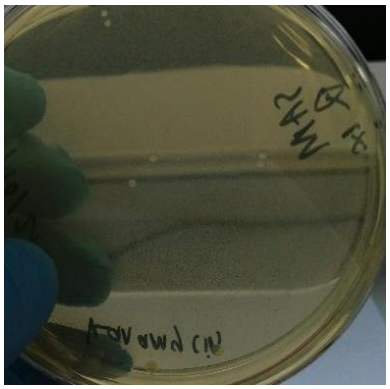
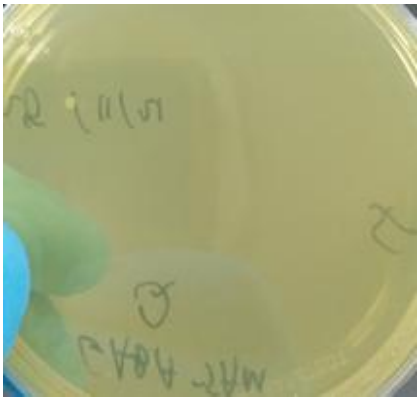
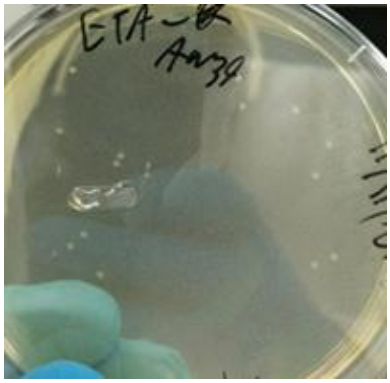
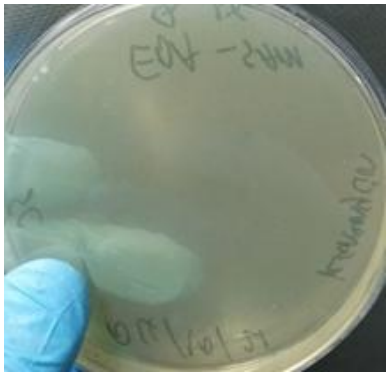
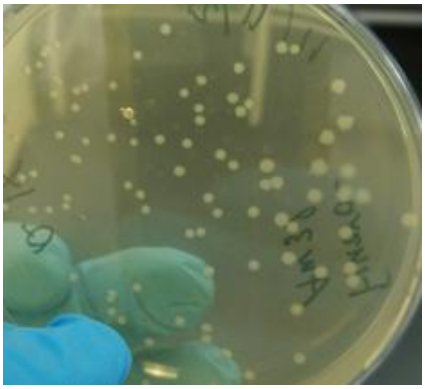
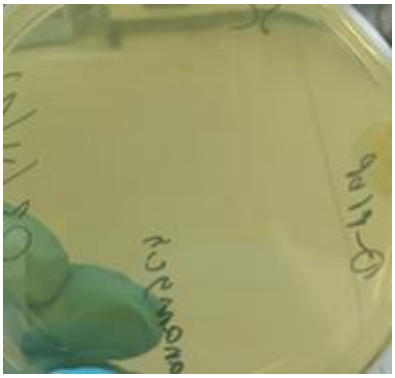
7.85 K115N

Appendix Table 4: K115N bacterial results

 SAM	γ-  aminobutyric acid
 **2**	 **1**
 **3**	 **4**

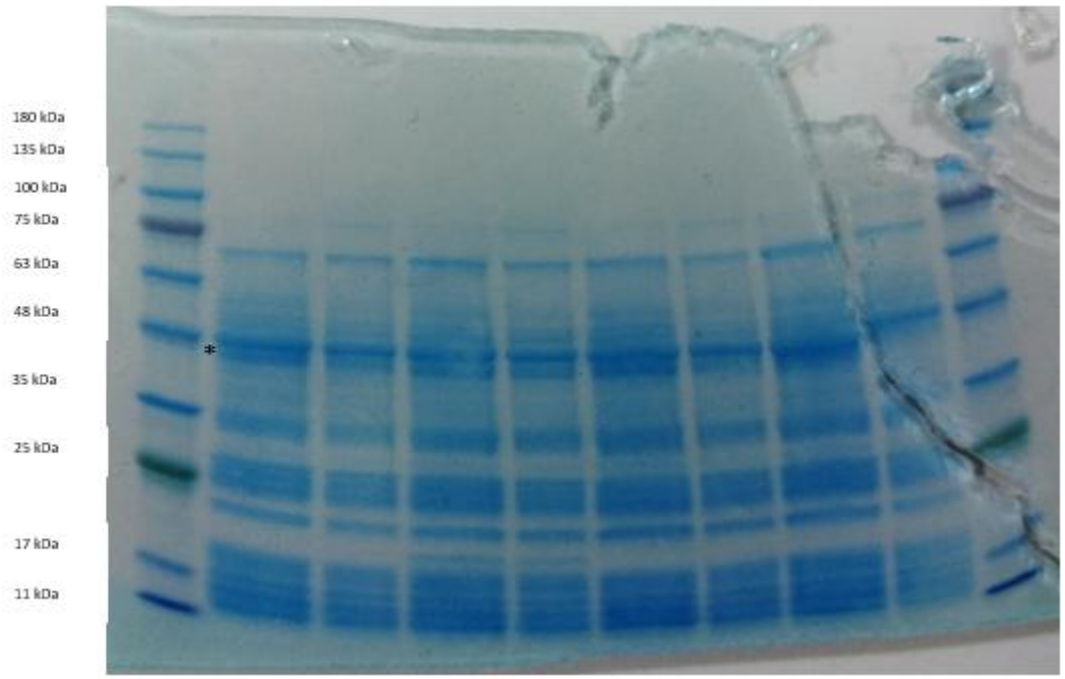
7.86 K115Q

Appendix table 5: K115Q bacterial results

 SAM	 γ-aminobutyric acid
 **2**	 **1**
 **3**	 **4**

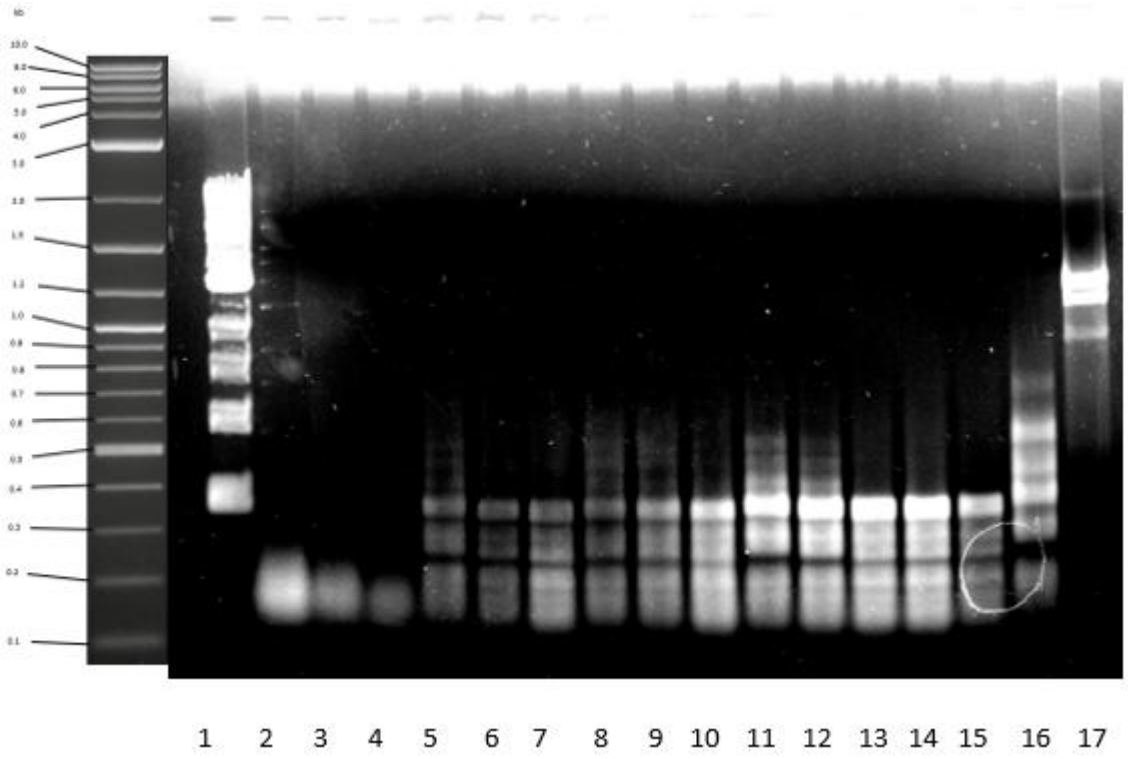
7.9 Gel Images

7.91 SDS PAGE



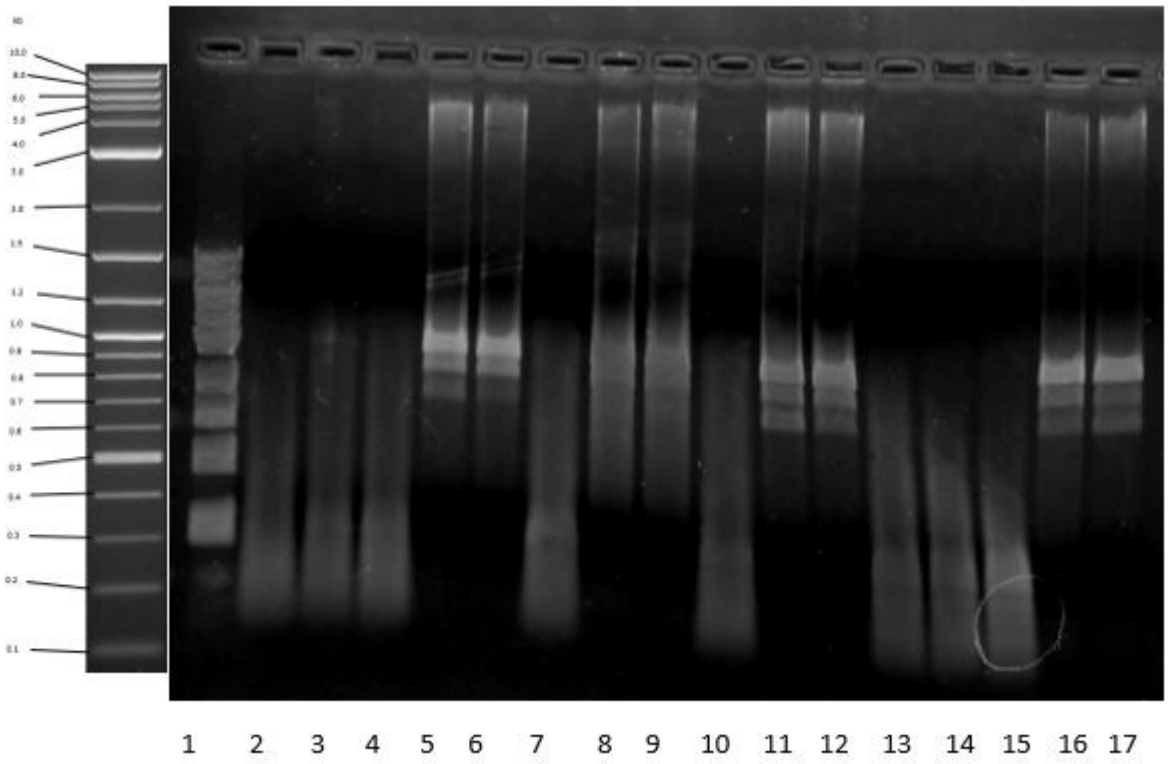
Appendix Figure 42: SDS PAGE of each mutant MTase, lane 1 contains the protein ladder, lanes 2 and 3 contain K115A pools, lanes 4 and 5 contain the K115E pools, lanes 6 and 7 contain K115H pools, lanes 8 and 9 contain K115Q pools and lane 10 contains protein ladder. Asterisk dictates where the band for the M.MpeI protein is found.

7.92 K115X_**1**



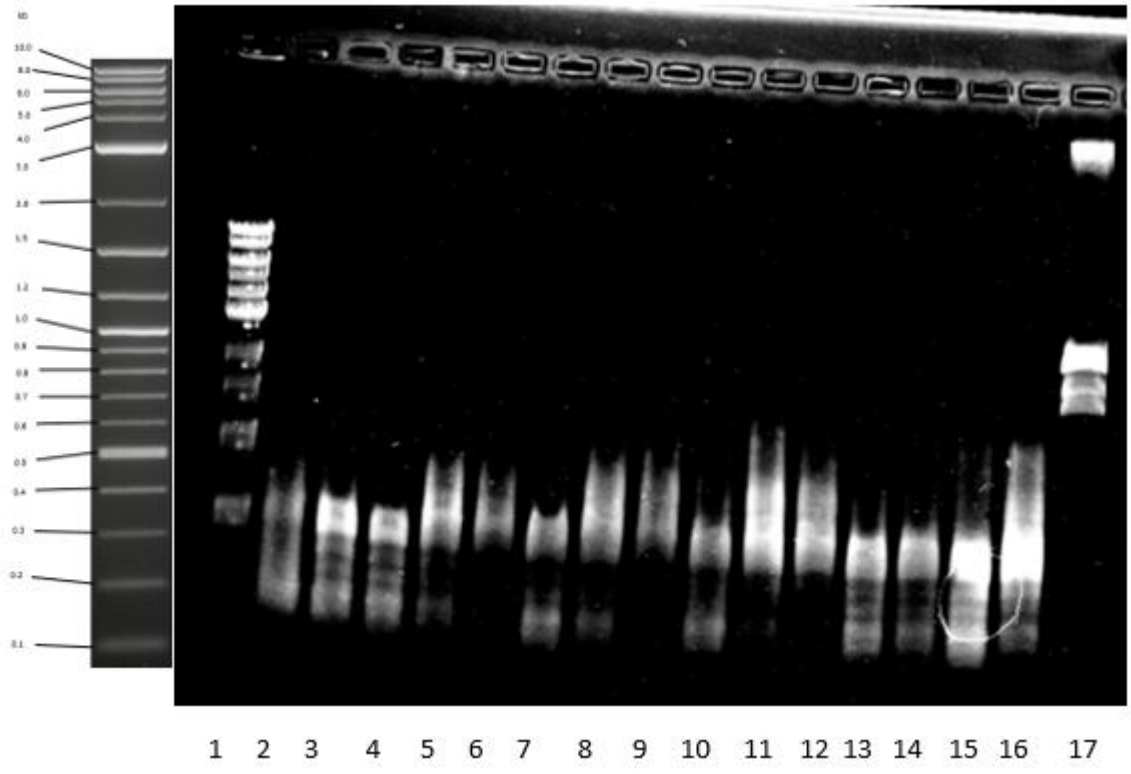
*Appendix Figure 43: Gel image of all mutant proteins with the **1** cofactor (Lane 1: DNA ladder, Lanes 2-3 K115A, 2-fold dilution of cofactor, Lane 4: K115A Control, Lane 5-6 K115E, 2-fold dilution of cofactor, Lane 7: K115E Control, Lane 8-9: K115H, 2-fold dilution of cofactor, Lane 10: K115H Control, Lane 11-13: K115Q, Serial Dilution of Cofactor, Lane 14: K115Q control, Lane 15: Control without cofactor, Lane 16: control without cofactor or enzyme, Lane 17: SAM Control, cofactor concentration is 500 µM to 250 µM Enzyme concentration is 0.2 mM for K115A and 6.0 µM for other mutants.*

7.93 K115X_**2**



*Appendix Figure 44: Gel image of all mutant proteins with the **2** cofactor (Lane 1: DNA ladder, Lanes 2-3 K115A, 2-fold dilution of cofactor, Lane 4: K115A Control, Lane 5-6 K115E, 2-fold dilution of cofactor, Lane 7: K115E Control, Lane 8-9: K115H, 2-fold dilution of cofactor, Lane 10: K115H Control, Lane 11-13: K115Q, Serial Dilution of Cofactor, Lane 14: K115Q control, Lane 15: Control without cofactor, Lane 16: control without cofactor or enzyme, Lane 17: SAM Control, cofactor concentration is 500 µM to 250 µM Enzyme concentration is 0.2 mM for K115A and 6.0 µM for other mutants.*

7.94 K115X_**4**



*Appendix Figure 45: Gel image of all mutant proteins with the **4** cofactor (Lane 1: DNA ladder, Lanes 2-3 K115A, 2-fold dilution of cofactor, Lane 4: K115A Control, Lane 5-6 K115E, 2-fold dilution of cofactor, Lane 7: K115E Control, Lane 8-9: K115H, 2-fold dilution of cofactor, Lane 10: K115H Control, Lane 11-13: K115Q, Serial Dilution of Cofactor, Lane 14: K115Q control, Lane 15: Control without cofactor, Lane 16: control without cofactor or enzyme, Lane 17: SAM Control, cofactor concentration is 500 µM to 250 µM Enzyme concentration is 0.2 mM for K115A and 6.0 µM for other mutants.*

8.0 References

1 C. Bailly, M. J. Waring and A. A. Travers, *J. Mol. Biol*, 1995, **253**, 1–7.

2 J. Watson and F. Crick, *Nature*, 1953, **171**, 737–738.

3 S. Arnott, M. H. F. Wilkins, W. Fuller and R. Langridge~, *J. Mol. Biol*, 1967, **27**, 535–548.

4 B. D. S Arnott, Pr. F. Hutchinson, D. M. Spencer and Pr. M. H F Wilkins, *Nature*, 1966, **211**, 227–232.

5 R. E. Dickerson, H. R. Drew, B. N. Conner, R. M. Wing, A. V Fratini and M. L. Kopka, *Science (1979)*, 1982, **216**, 475–485.

6 S. Arnott, 2006, preprint, DOI: 10.1016/j.tibs.2006.04.004.

7 C. A. Bewley, A. M. Gronenborn and G. M. Clore, *Annu. Rev. Biophys. Biomol. Struct*, 2025, **27**, 105–136.

8 A. C. Vallur, J. A. Feller, C. W. Abner, R. K. Tran and L. B. Bloom, *JBC*, 2002, **277**, 31673–31678.

9 T. R. O'connor, *Nucleic Acids Res*, 1993, **21**, 5561–5569.

10 M. D. Wyatt and L. D. Samson, *Carcinogenesis*, 2000, **21**, 901–908.

11 J. Wu and M. Brown, in *Hematology: Basic Principals and Practice*, 2018, vol. 7, pp. 17–24.

12 D. E. Handy, R. Castro and J. Loscalzo, *Circulation*, 2011, **123**, 2145–2156.

13 Jenuwein T., *Science (1979)*, 2016, **297**, 2215–2218.

14 L. Statello, C. J. Guo, L. L. Chen and M. Huarte, *Nature Research*, 2021, preprint, DOI: 10.1038/s41580-020-00315-9.

15 A. J. Bannister and T. Kouzarides, 2011, preprint, DOI: 10.1038/cr.2011.22.

16 B. Jin, Y. Li and K. D. Robertson, 2011, preprint, DOI: 10.1177/1947601910393957.

17 J. F. Kribelbauer, X. J. Lu, R. Rohs, R. S. Mann and H. J. Bussemaker, *Academic Press*, 2020, preprint, DOI: 10.1016/j.jmb.2019.10.021.

18 A. C. D. Machado, T. Zhou, S. Rao, P. Goel, C. Rastogi, A. Lazarovici, H. J. Bussemaker and R. Rohs, *Brief Funct Genomics*, 2015, **14**, 61–73.

19 S. A. Wajed, P. W. Laird and T. R. Demeester, *Ann Surg*, 2001, **234**, 10–20.

20 N. Datta, S. Chakraborty, M. Basu and M. K. Ghosh, *MDPI*, 2021, preprint, DOI: 10.3390/cells10010046.

21 E. Y. H. P. Lee and W. J. Muller, 2010, preprint, DOI: 10.1101/cshperspect.a003236.

22 B. J. Aubrey, G. L. Kelly, A. Janic, M. J. Herold and A. Strasser, *Nature Publishing Group*, 2018, preprint, DOI: 10.1038/cdd.2017.169.

23 Sherr J. Charles., *Cell*, 2004, **116**, 235–246.

24 E. N. Kontomanolis, A. Koutras, A. Syllaios, D. Schizas, A. Mastoraki, N. Garmpis, M. Diakosavvas, K. Angelou, G. Tsatsaris, A. Pagkalos, T. Ntounis and Z. Fasoulakis, *Anticancer Res*, 2020, **40**, 6009–6015.

25 K. D. Rasmussen and K. Helin, *Genes Dev*, 2016, **30**, 733–750.

26 F. Barras and M. G. Marinus, *Trends in Gentics*, 1989, **5**, 139–143.

27 N. Takahashi, Y. Naito, N. Handa and I. Kobayashi, *J Bacteriol*, 2002, **184**, 6100–6108.

28 L. D. Moore, T. Le and G. Fan, 2013, preprint, DOI: 10.1038/npp.2012.112.

29 C. L. Ulrey, L. Liu, L. G. Andrews and T. O. Tollefsbol, 2005, preprint, DOI: 10.1093/hmg/ddi100.

30 M. Bochtler and H. Fernandes, *John Wiley and Sons Inc*, 2021, preprint, DOI: 10.1002/bies.202000243.

31 C. Shen, K. Wang, X. Deng and J. Chen, *Elsevier Ltd*, 2022, preprint, DOI: 10.1016/j.tig.2021.12.003.

32 E. Boye, *J Bacteriol*, 1991, **173**, 4537–4539.

33 C. Nievera, J. J-C Torgue, J. E. Grimwade and A. C. Leonard, *Mol Cell*, 2006, **24**, 581–592.

34 J. M. Atack, A. Tan, L. O. Bakaletz, M. P. Jennings and K. L. Seib, *Elsevier Ltd*, 2018, preprint, DOI: 10.1016/j.tim.2018.01.008.

35 D. M. Heithoff, E. Y. Enioutina, R. A. Daynes, R. L. Sinsheimer, D. A. Low and M. J. Mahan, *Infect Immun*, 2001, **69**, 6725–6730.

36 K. E. Medvedevid, L. N. Kinch, R. D. Schaeffer and N. V. Grishin, *PLoS Comput Biol*, DOI:10.1371/journal.pcbi.1007569.

37 H. L. Schubert, R. M. Blumenthal and X. Cheng, *Elsevier Ltd*, 2003, preprint, DOI: 10.1016/S0968-0004(03)00090-2.

38 K. E. Medvedevid, L. N. Kinch, R. D. Schaeffer and N. V. Grishin, *PLoS Comput Biol*, DOI:10.1371/journal.pcbi.1007569.

39 B. P. S. Chouhan, S. Maimaiti, M. Gade and P. Laurino, *Biochemistry*, 2019, **58**, 166–170.

40 L. M. Longo, J. Jabtoiiska, P. Vyas, M. Kanade, R. Kolodny, N. Ben-Tal and D. S. Tawfik, *Elife*, 2020, **9**, 1–16.

41 B. P. S. Chouhan, S. Maimaiti, M. Gade and P. Laurino, *Biochemistry*, 2019, **58**, 166–170.

42 H. J. Seong, S. W. Han and W. J. Sul, *Journal of Microbiology*, 2021, **59**, 242–248.

43 J. T. Rostøl and L. Marraffini, *Cell Press*, 2019, preprint, DOI: 10.1016/j.chom.2019.01.009.

44 A. Razin and B. Kantor, *Prog Mol Subcell Biol*, 2005, **38**, 151–67.

45 E. R. Gibney and C. M. Nolan, 2010, preprint, DOI: 10.1038/hdy.2010.54.

46 B. Tycko, *J Clin Invest*, 2000, **105**, 401–407.

47 R. Gerasimait, E. Merkiene and S. Klimašauskas, *Nucleic Acids Res*, 2011, **39**, 3771–3780.

48 J. Yang, L. Lior-Hoffmann, S. Wang, Y. Zhang and S. Broyde, *Biochemistry*, 2013, **52**, 2828–2838.

49 S. Bheemanaik, Y. V. R. Reddy and D. N. Rao, 2006, preprint, DOI: 10.1042/BJ20060854.

50 V. Valinluck, H. H. Tsai, D. K. Rogstad, A. Burdzy, A. Bird and L. C. Sowers, *Nucleic Acids Res*, 2004, **32**, 4100–4108.

51 K. E. Szulwach, X. Li, Y. Li, C. X. Song, J. W. Han, S. S. Kim, S. Namburi, K. Hermetz, J. J. Kim, M. K. Rudd, Y. S. Yoon, B. Ren, C. He and P. Jin, *PLoS Genet*, DOI:10.1371/journal.pgen.1002154.

52 X. Zhang and T. C. Bruice, *PNAS April*, 2006, **18**, 6148–6153.

53 C. Li, *Genet Epigenet*, 2013, **5**, GEG.S12143.

54 A. C. Drohat and C. T. Coey, *American Chemical Society*, 2016, preprint, DOI: 10.1021/acs.chemrev.6b00191.

55 C. S. Nabel, H. Jia, Y. Ye, L. Shen, H. L. Goldschmidt, J. T. Stivers, Y. Zhang and R. M. Kohli, *Nat Chem Biol*, 2012, **8**, 751–758.

56 S. Ito, A. C. Dalessio, O. V. Taranova, K. Hong, L. C. Sowers and Y. Zhang, *Nature*, 2010, **466**, 1129–1133.

57 U. Müller, C. Bauer, M. Siegl, A. Rottach and H. Leonhardt, *Nucleic Acids Res*, 2014, **42**, 8592–8604.

58 G. L. Xu and J. Wong, *Oxford University Press*, 2015, preprint, DOI: 10.1093/nsr/nwv029.

59 M. Bochtler, A. Kolano and G. L. Xu, *BioEssays*, 2017, **39**, 1–13.

60 H. Hashimoto, S. Hong, A. S. Bhagwat, X. Zhang and X. Cheng, *Nucleic Acids Res*, 2012, **40**, 10203–10214.

61 A. Onodera, E. González-Avalos, C. W. J. Lio, R. O. Georges, A. Bellacosa, T. Nakayama and A. Rao, *Genome Biol*, DOI:10.1186/s13059-021-02384-1.

62 L. Scourzic, E. Mouly and O. A. Bernard, *Genome Med*, DOI:10.1186/s13073-015-0134-6.

63 R. M. Kohli and Y. Zhang, 2013, preprint, DOI: 10.1038/nature12750.

64 K. Joshi, S. Liu, P. Breslin S.J and J. Zhang, *Cellular and Molecular Life Sciences*, DOI:10.1007/s00018-022-04396-x.

65 P. Melamed, Y. Yosefzon, C. David, A. Tsukerman and L. Pnueli, *Frontiers Media S.A.*, 2018, preprint, DOI: 10.3389/fcell.2018.00022.

66 C. S. Robb, L. Reisky, U. T. Bornscheuer and J. H. Hehemann, *Biochemical Journal*, 2018, **475**, 3875–3886.

67 Y. Feng, N. Bin Xie, W. B. Tao, J. H. Ding, X. J. You, C. J. Ma, X. Zhang, C. Yi, X. Zhou, B. F. Yuan and Y. Q. Feng, *CCS Chemistry*, 2021, **3**, 994–1008.

68 D. Q. Shi, I. Ali, J. Tang and W. C. Yang, *Frontiers Media S.A.*, 2017, preprint, DOI: 10.3389/fgene.2017.00100.

69 N. Mahmood, D. Cheishvili, A. Arakelian, I. Tanvir, H. A. Khan, A.-S. Pépin, M. Szyf and S. A. Rabbani, *Oncotarget*, 2018, **9**, 5169–5183.

70 Z. Gong and J. K. Zhu, *Cell Res*, 2011, **21**, 1649–1651.

71 A. C. Drohat and C. T. Coey, *American Chemical Society*, 2016, preprint, DOI: 10.1021/acs.chemrev.6b00191.

72 H. E. Krokan and M. Bjørås, *Cold Spring Harb Perspect Biol*, 2013, **5**, 1–22.

73 N. M. Da Costa, A. Hautefeuille, M. P. Cros, M. E. Melendez, T. Waters, P. Swann, P. Hainaut and L. F. R. Pinto, *Cell Cycle*, 2012, **11**, 4570–4578.

74 S. R. Dalton and A. Bellacosa, 2012, preprint, DOI: 10.2217/epi.12.36.

75 T. J. Begley and R. P. Cunningham, *Protein Eng*, 1999, **12**, 333–340.

76 S. U. Siriwardena, K. Chen and A. S. Bhagwat, *American Chemical Society*, 2016, preprint, DOI: 10.1021/acs.chemrev.6b00296.

77 R. Pecori, S. Di Giorgio, J. Paulo Lorenzo and F. Nina Papavasiliou, *Nat Rev Genet*, 2022, **23**, 505–518.

78 E. K. Schutsky, J. E. DeNizio, P. Hu, M. Yun Liu, C. S. Nabel, E. B. Fabyanic, Y. Hwang, F. D. Bushman, H. Wu, R. M. Kohli, H. conceived and N. Biotechnol, *Nat Biotechnol*, 2018, **36**, 1083–1090.

79 P. Wijesinghe and A. S. Bhagwat, *Nucleic Acids Res*, 2012, **40**, 9206–9217.

80 F. Ito, Y. Fu, S. C. A. Kao, H. Yang and X. S. Chen, *J Mol Biol*, 2017, **429**, 1787–1799.

81 M. Jaszczur, J. G. Bertram, P. Pham, M. D. Scharff and M. F. Goodman, 2013, preprint, DOI: 10.1007/s00018-012-1212-1.

82 R. S. Das, M. Samaraweera, M. Morton, J. A. Gascón and A. K. Basu, *Chem Res Toxicol*, 2012, **25**, 2451–2461.

83 Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, N. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He and G. L. Xu, *Science (1979)*, 2011, **333**, 1303–1307.

84 L. A. McGregor, B. Zhu, A. M. Goetz and J. T. Sczepanski, *JBC*, DOI:10.1016/j.jbc.2023.104590.

85 P. Albert, B. Varga, N. Zsibrita and A. Kiss, *PLoS One*, DOI:10.1371/journal.pone.0197232.

86 M. Wojciechowski, H. Czapinska and M. Bochtler, *Proc Natl Acad Sci U S A*, 2013, **110**, 105–110.

87 A. Y. Rudenko, S. S. Mariasina, P. V. Sergiev and V. I. Polshakov, *Pleiades journals*, 2022, preprint, DOI: 10.1134/S002689332202011X.

88 T. C. Petrossian and S. G. Clarke, *Molecular and Cellular Proteomics*, DOI:10.1074/mcp.M110.000976.

89 P. Calabrese and D. Shibata, *BMC Cancer*, DOI:10.1186/1471-2407-10-3.

90 C. Hu, X. Liu, Y. Zeng, J. Liu and F. Wu, *BioMed Central Ltd*, 2021, preprint, DOI: 10.1186/s13148-021-01154-x.

91 J. Zhang, C. Yang, C. Wu, W. Cui and L. Wang, *MDPI AG*, 2020, preprint, DOI: 10.3390/cancers12082123.

92 M. Ehrlich, 2009, preprint, DOI: 10.2217/EPI.09.33.

93 S. Somasundaram, M. E. Forrest, H. Moinova, A. Cohen, V. Varadan, T. Laframboise, S. Markowitz and A. M. Khalil, *Clin Epigenetics*, DOI:10.1186/s13148-018-0555-3.

94 D. T. Debela, S. G. Y. Muzazu, K. D. Heraro, M. T. Ndalama, B. W. Mesele, D. C. Haile, S. K. Kitui and T. Manyazewal, *SAGE Publications Ltd*, 2021, preprint, DOI: 10.1177/20503121211034366.

95 S. Senapati, A. K. Mahanta, S. Kumar and P. Maiti, *Springer Nature*, 2018, preprint, DOI: 10.1038/s41392-017-0004-3.

96 M. G. Uddin and T. E. Fandy, *Adv Cancer Res*, 2021, **152**, 205–223.

97 J. Ma and Z. Ge, *Frontiers Media S.A.*, 2021, preprint, DOI: 10.3389/fphar.2021.701690.

98 M. Billam, M. D. Sobolewski and N. E. Davidson, *Breast Cancer Res Treat*, 2010, **120**, 581–592.

99 C. Hu, X. Liu, Y. Zeng, J. Liu and F. Wu, *BioMed Central Ltd*, 2021, preprint, DOI: 10.1186/s13148-021-01154-x.

100 N. Sun, J. Zhang, C. Zhang, B. Zhao and A. O. Jiao, *Oncol Lett*, 2018, **16**, 5799–5806.

101 C. Zwergel, M. Schnekenburger, F. Sarno, C. Battistelli, M. C. Manara, G. Stazi, R. Mazzone, R. Fioravanti, C. Gros, F. Ausseil, C. Florean, A. Nebbioso, R. Strippoli, T. Ushijima, K. Scotlandi, M. Tripodi, P. B. Arimondo, L. Altucci, M. Diederich, A. Mai and S. Valente, *Clin Epigenetics*, DOI:10.1186/s13148-019-0663-8.

102 S. Kumar, E. A. Gonzalez, P. Rameshwar and J. P. Etchegaray, *MDPI AG*, 2020, preprint, DOI: 10.3390/cancers12123657.

103 W. Filipowicz, L. Jaskiewicz, F. A. Kolb and R. S. Pillai, *Elsevier Ltd*, 2005, preprint, DOI: 10.1016/j.sbi.2005.05.006.

104 L. Wang, N. Mu and N. Qu, *Oncol Rep*, 2021, **45**, 65–72.

105 W. Blum, R. Garzon, R. B. Klisovic, S. Schwind, A. Walker, S. Geyer, S. Liu, V. Havelange, H. Becker, L. Schaaf, J. Mickle, H. Devine, C. Kefauver, S. M. Devine, K. K. Chan, N. A. Heerema, C. D. Bloomfield, M. R. Grever, J. C. Byrd, M. Villalona-Calero, C. M. Croce and G. Marcucci, *Proc Natl Acad Sci U S A*, 2010, **107**, 7473–7478.

106 M. Hua, Y. Qin, M. Sheng, X. Cui, W. Chen, J. Zhong, J. Yan and Y. Chen, *Mol Med Rep*, 2019, **49**, 3575–3583.

107 D. Ye, Z. Shen and S. Zhou, *Dove Medical Press Ltd*, 2019, preprint, DOI: 10.2147/CMAR.S191696.

108 D. Qi, J. Li, B. Que, J. Su, M. Li, C. Zhang, M. Yang, G. Zhou and W. Ji, *Cancer Cell Int*, DOI:10.1186/s12935-016-0356-8.

109 C. Chen, P. Wang, W. Mo, Y. Zhang, W. Zhou, T. Deng, M. Zhou, X. Chen, S. Wang and C. Wang, *Oncol Lett*, 2019, **18**, 2203–2211.

110 S. Somasundaram, M. E. Forrest, H. Moinova, A. Cohen, V. Varadan, T. Laframboise, S. Markowitz and A. M. Khalil, *Clin Epigenetics*, DOI:10.1186/s13148-018-0555-3.

111 X. Guo, Z. Chen, L. Zhao, D. Cheng, W. Song and X. Zhang, *Exp Cell Res*, DOI:10.1016/j.yexcr.2019.06.006.

112 W. Li, J. Zheng, J. Deng, Y. You, H. Wu, N. Li, J. Lu and Y. Zhou, *Gastroenterology*, DOI:10.1053/j.gastro.2014.03.002.

113 Y. Cheng, C. He, M. Wang, X. Ma, F. Mo, S. Yang, J. Han and X. Wei, *Signal Transduct Target Ther*, DOI:10.1038/s41392-019-0095-0.

114 A. Gnyszka, Z. Jastrzebski and S. Flis, *Anticancer Res*, 2013, **33**, 2989–2996.

115 R. B. Mokhtari, T. S. Homayouni, N. Baluch, E. Morgatskaya, S. Kumar, B. Das and H. Yeger, *Oncotarget*, 2017, **8**, 38022–38043.

116 Y.-S. Jung, *Biol. Pharm. Bull*, 2015, **38**, 971–974.

117 W. Niu, S. Cao, M. Yang and L. Xu, *Catalysts*, DOI:10.3390/catal7080238.

118 M. Zhao, Y. S. Wijayasinghe, P. Bhansali, R. E. Viola and R. M. Blumenthal, *Microbiology (United Kingdom)*, 2015, **161**, 674–682.

119 V. Goyvaerts, S. Van Snick, L. D'Huys, R. Vitale, M. Helmer Lauer, S. Wang, V. Leen, W. Dehaen and J. Hofkens, *ChemComm*, 2020, **56**, 3317–3320.

120 A. K. Upadhyay, J. R. Horton, X. Zhang and X. Cheng, 2011, preprint, DOI: 10.1016/j.sbi.2011.08.003.

121 H. M. Cohen, A. D. Griffiths, D. S. Tawfik and D. Loakes, *Org Biomol Chem*, 2005, **3**, 152–161.

122 X. Cheng and R. J. Roberts, *Nucleic Acids Res*, 2001, **29**, 3784–3795.

123 I. R. Bothwell, K. Islam, Y. Chen, W. Zheng, G. Blum, H. Deng and M. Luo, *J Am Chem Soc*, 2012, **134**, 14905–14912.

124 A. Y. Rudenko, S. S. Mariasina, P. V. Sergiev and V. I. Polshakov, *Pleiades journals*, 2022, preprint, DOI: 10.1134/S002689332202011X.

125 L. R. Comstock and S. R. Rajski, *Nucleic Acids Res*, 2005, **33**, 1644–1652.

126 T. Osborne, R. L. W. Roska, S. R. Rajski and P. R. Thompson, *J Am Chem Soc*, 2008, **130**, 4574–4575.

127 J. Zhang and Y. G. Zheng, *ACS Chem Biol*, 2016, **11**, 583–597.

128 M. Pignot, C. Siethoff, M. Linscheid and E. Weinhold, *Angewandte Chemie - International Edition*, 1998, **37**, 2888–2891.

129 A. Plotnikova, A. Osipenko, V. Masevičius, G. Vilkaitis and S. Klimašauskas, *J Am Chem Soc*, 2014, **136**, 13550–13553.

130 G. Pljevaljcic, M. Pignot and E. Weinhold, *J Am Chem Soc*, 2003, **125**, 3486–3492.

131 L. R. Comstock and S. R. Rajski, *J Am Chem Soc*, 2005, **127**, 14136–14137.

132 L. R. Comstock and S. R. Rajski, *Nucleic Acids Res*, 2005, **33**, 1644–1652.

133 R. L. Weller and S. R. Rajski, *Org Lett*, 2005, **7**, 2141–2144.

134 R. L. Weller and S. R. Rajski, *ChemBioChem*, 2006, **7**, 243–245.

135 M. Ramadan, N. K. Bremner-Hay, S. A. Carlson and L. R. Comstock, *Tetrahedron*, 2014, **70**, 5291–5297.

136 C. Dalhoff, G. Lukinavičius, S. Klimašauskas and E. Weinhold, *Nat Chem Biol*, 2006, **2**, 31–32.

137 G. Lukinavičius, V. Lapiene, Z. Staševskij, C. Dalhoff, E. Weinhold and S. Klimašauskas, *J Am Chem Soc*, 2007, **129**, 2758–2759.

138 S. C. Lu and J. M. Mato, *J Gastroenterol Hepatol*, 2008, **23**, 73–77.

139 G. Lukinavičius, A. Lapinaite, G. Urbanavičiute, R. Gerasimaite and S. Klimašauskas, *Nucleic Acids Res*, 2012, **40**, 11594–11602.

140 E. Jagu, 2020, preprint.

141 K. B. Ubych, 2020.

142 A. A. Wilkinson, E. Jagu, K. Ubych, S. Coulthard, A. E. Rushton, J. Kennefick, Q. Su, R. K. Neely and P. Fernandez-Trillo, *ACS Cent Sci*, 2020, **6**, 525–534.