Structure-Guided Directed Evolution of Alkyltransferase Enzymes

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy (PhD) in the Faculty of Engineering and Physical Sciences

September 2015

Matthew Rawlinson Bennett

School of Chemistry

Contents

Table of Figures

Table of Tables

Table of Schemes

Abstract

Currently alkylation methods are dominated by synthetic procedures. The aim of this study has been to seek a greener, safer and more selective alternative by using directed evolution and cofactor manipulation to alter methyltransferases to selective alkyltransferases. The work focussed on the Catechol-*O*-methyltransferase (COMT), Coclaurine-*N*-methyltransferase (CNMT) and Tetrahydroprotoberberine-cis-*N*-methyltransferase (TNMT) enzymes.

The selectivity of COMT was initially improved by mutagenesis, which produced regiocomplementary mutants in K144A and Y200L. K144E showing some *para* regioselectivity (-27.8% r.e DHBAL substrate), whilst the Y200L mutant displayed nearly total *meta* regioselectivity (90.2% r.e DHBAL substrate). The selectivity of quaternary structures of COMT was also investigated and the dimeric COMT was found to be *meta* regioselective (83.8% r.e DHBAL substrate). Crystallographic studies of both the dimeric COMT and Y200L COMT mutant suggested that the orientation of the E199 residue controlled COMT's regioselectivity.

Two coupled systems of COMT were developed aimed at expanding the cofactor and substrate scope of the COMT enzyme. Initially synthetic analogues of *S*-adenosyl-methionine (SAM) were shown to be active with COMT transferring allyl and benzyl groups. The activity of the allylation reaction was improved by using the hMAT2A enzyme, which produced SAM analogues *in situ* using the *S*-allylhomocysteine cofactor. The second coupled system that was developed was a tyrosinase-COMT system and was used to regioselectively methoxylate a range of phenolic peptides, culminating in labelling of the prostrate cancer drug; goserelin.

Finally, the substrate scope of CNMT and TNMT enzymes were investigated through a high-throughput methyltransferase screen. Improvement of the CNMT substrate scope was investigated through crystallography. The crystal structure, coupled with modelling studies, led to identification of a catalytic diad in CNMT (E207 and H208), which was found to be involved in the methylation event. Key binding residues (Y328, W329 and F332) were also identified through these modelling studies.

Declaration

I declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

I.   The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

II. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

III. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

IV. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy, in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations and in The University's policy on Presentation of Theses.

Acknowledgements

"It was the best of times, it was the worst of times. It was the age of wisdom, it was the age of foolishness."-Charles Dickens *A Tale of Two Cities* 1859

Firstly, my thanks must go to my supervisor, without whom I would not have carried out this work. It has been a good four years of hard work and Jason has made working in his lab a wholly good experience. I would also like to thank Dr Reddy's and CoEBio3 for their funding, particularly Dr Ania Fryszkowska for all her help and guidance. I would also like to thank the entire Micklefield group; old members and new. I have had the pleasure to work with a talented group of people in; Dr Mark "Tommo" Thompson, Dr Anna-Winona Struck, Dr Brian Law and Dr Sarah "Sheppy" Shepherd. These people have provided constant support and advice, and occasionally jovial badgering. In Dr Sheppy and Dr Tommo two better proof readers could not have been found, and Dr Anna has been fantastic with all her "NMRing" and NMR tutorials and just general life advice. Some members, such as Dr KK (famous for his guarantee) have made the MIB a great place to work in with his sunny disposition. Dr Chris "Bobby" Robinson and Dr Michael "Winner" Winn have been excellent partners in our lab quiz team.

Finally, my thanks must be aimed at my parents and my partner of some years. My parents have funded me through four years of university (and the 18 years before that!) and without those qualifications I would never have got this far. They have always supported and helped me, despite for some years not understanding a single word of my work. As to my partner Jenny I say, I couldn't have completed this tome without you. Thank you for putting up with me whilst I was disgruntled by the constant rearranging of Word and the debacle with the compound numbering. It's been a wonderful six years, here's to many more.

Abbreviations

ADH-	Alcohol dehydrogenase
API-	Active pharmaceutical ingredient
ATP-	Adenosine triphosphate
BuLi-	Butyl lithium
CNMT-	Coclaurine-*N*-methyltransferase
COMT-	Catechol-*O*-methyltransferase
DHBA-	3,4-dihydroxybenzoic acid
DHBAL-	3,4-dihydroxybenzaldehyde
DIPEA-	N,N-diisopropylethylamine
DMSO-	Dimethyl sulfoxide
DNC-	3,5-dinitrocatechol
DTT-	Dithiothreitol
FDAS-	5'fluoro-5'-deoxyfluoroadenosine synthase
FPLC-	Fast protein liquid chromatography
GNMT-	Glycine-*N*-methyltransferase
HPLC-	High pressure liquid chromatography
HTS-	High throughput screen

HVA-   Homovanillic acid

L-DOPA-   L-3,4-dihydroxyphenylalanine

LB-   Lysogeny broth

LCMS-   Liquid chromatography mass spectrometer

LDA-   Lithium diisopropylamide

MALDI-   Matrix assisted laser desorption ionisation

MAO-   Monoamine oxidase

MAT-   Methionine adenosyl transferase

MTHF-   Methyltetrahydrofolate

NADPH-   Nicotinamide adenine dinucleotide phosphate

NDCC-   7-nitro-2,3-dihydro-1H-cyclopenta[b]chromen-1-one

NO2CAT-   4-nitrocatechol

NRPS-   Non ribosomal peptide synthetase

NTA-   Nitrilotriaceticacid

PKS-   Polyketide synthase

POB-   Propynl-oxy-2-butene

R.e-	Regioisomeric excess
SAAH-	*S*-adenosyl-allylhomocysteine
SABH-	*S*-adenosyl-benzylhomocysteine
SAE-	*S*-adenosyl-ethionine
SAH-	*S*-adenosyl-homocysteine
SAHH-	SAH hydrolase
SAlH-	*S*-allylhomocysteine
SAM-	*S*-adenosyl-methionine
SAMH-	*S*-adenosyl-methylacetatehomocysteine
SAPH-	*S*-adenosyl-propylhomocysteine
SAPrH-	*S*-adenosyl-propargylhomocysteine
SBH-	*S*-benzylhomocysteine
SDH-	Sarcosine dehydrogenase
SeAM-	*Se*-adenosylselenomethionine
SeAPrH-	*Se*-adenosyl-propargylselenohomocysteine
SOB-	Super optimal broth
SPH-	*S*-propylhomocysteine
TFA-	Trifluoroacetic acid

TNMT-   Tetrahydroprotoberberine-cis-*N*-methyltransferase

VMA-   Vanillyl mandelic acid

WT-   Wild-type

Abbreviations-amino acids

A	Alanine
C	Cysteine
D	Aspartic acid
E	Glutamic acid
F	Phenylalanine
G	Glycine
H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
N	Asparagine
P	Proline
Q	Glutamine
R	Arginine

S	Serine
T	Threonine
U	Selenocysteine
V	Valine
W	Tryptophan
Y	Tyrosine

1 Introduction

Methyltransferases are found across nature, with methylation playing a vital role in a number of key biological process.[1,2] From DNA signalling,[3,4] to the metabolism of neurotransmitters,[5] methyltransferases have been found to be ubiquitous, spanning both prokaryotic and eukaryotic classes.[6] The range of substrates active with methyltransferase has made this class of enzymes an interesting area of study. Methyltransferases have also been developed for biocatalysis with varying degrees of success.[7,8] This study will aim to build on these successes and secure more selective methyltransferases with their associated cofactors.

1.1 Classes and biological activity of methyltransferases

1.1.1 Overall classification

5-methyltetrahydrofolate   *S*-adenosyl-methionine (**1a**)
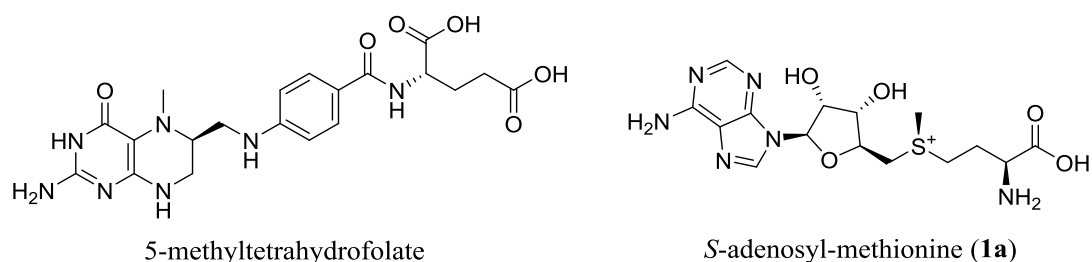
Figure 1: The structures of methyltransferase cofactors tetrahydrofolate and SAM (**1a**).[9]

Methyltransferases are found with two main cofactors; *S*-adenosyl-methionine (SAM) (**1a**) and 5-methyltetrahydrofolate (5-MTHF).[10,11] The structures of the two cofactors can be seen in Figure 1. SAM has a sulfonium centre,[11] which acts as an electrophile to nucleophilic substrates aiding the transfer of the methyl group at the centre sulfonium via an S_N2 mechanism producing *S*-adenosyl-homocysteine (SAH) (**1**) as a byproduct (Scheme 1).[2,4] Whereas 5-MTHF is first generated either through a trimethylsulfonium tetrahydrofolate methyltransferase or through a NADPH dependent methylenetetrahydrofolate reductase.[10,12] The methyl group from 5-MTHF is then transferred to a homocysteine (**2**) substrate to regenerate methionine (**2a**), with the methionine synthase domain (homocysteine methyltransferase).[13]The use of

the 5-MTHF methyltransferases have been shown to be limited by the requirement of two enzymes to complete the methyl transfer.[10,13] The substrate scope was also limited, as there are a limited number of 5-MTHF-dependent methyltransferases.[13-15] Whereas SAM dependent methyltransferases have been shown to have a larger substrate scope, with the cofactor SAM (**1a**) being the second most common cofactor in the cell (after ATP).[16] Therefore, most studies have focussed on SAM-dependent methyltransferases.
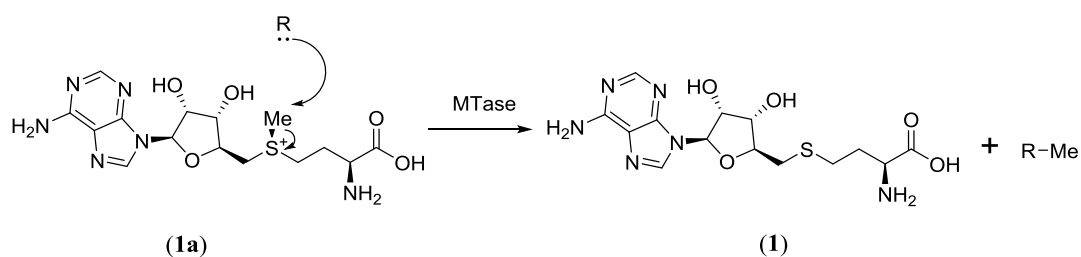
Scheme 1: SAM (**1a**) dependent methyltransferase reaction, R= methyltransferase substrate.[2]

SAM (**1a**)-dependent methyltransferases have been found to be diverse in structure and chemistry. These different chemistries have often been governed by the substrate's nucleophile with; C, N, O, S, P and even As methyltransferases having been discovered.[5,17-21] Categorisation of methyltransferase was needed to develop these enzymes for structural and mutagenesis studies. The best classification strategy for methyltransferases was determined to be by the structure of the enzymes, as structural motifs common to certain substrates have been observed.[22] The classes of methyltransferases have been named as 1-5 and Table 1 shows the classes, their structural motifs and the substrates common to the class.[22]

Table 1 showed how the methyltransferases can be categorised. However, in order to identify methyltransferases suitable for development as biocatalysts, more detail about the substrate scope and structure of each methyltransferase needed to be considered. The structure of each methyltransferase was considered important in understanding how the active sites of methyltransferases assemble. Identifying structural classes, and understanding, which residues of the active sites (of each classification) were important for activity or structural stability was important for further studies into the methyltransferases.

Table 1: The classes of methyltransferase and their common substrates.[22]

Class of methyltransferase	Most common substrates
1. Rossmann fold	A variety of different substrates from small molecules to RNA.
2. Regeneration SAM (**1a**) domain- antiparallel β sheets in a αβ sandwich	SAH (**1**)
3. Porphyrin methyltransferases-2 domain α/β sandwiches	Porphyrins
4. SPOUT domain	RNA
5. SET domain-β clip	Lysine residues of histones, proteins and other targets.

1.1.2   Class I methyltransferases

The Class I methyltransferase enzymes have been determined to be the largest methyltransferase class.[22] They have also been named the Rossmanoids due to the class characterised by their adoption of the Rossman fold, originally found in alcohol dehydrogenase enzymes.[22-24] The Rossman fold comprises of 7 β sheets; 6 parallel sheets and one anti parallel sheet (the last sheet in the sequence), with sandwiching α helices (Figure 2).[1,22] The structural features of the Rossman fold have also enabled researchers to identify key active site residues. Key residues of Rossmanoids have often been noted as being present on either the conjoining loop or C-terminus of the β sheets, which enables easier elucidation of key residues in unknown methyltransferases of this Class.[22] Other conserved motifs preserved across the Class

have also shown researchers about SAM (**1a**) binding, with a GxGxG motif found on C-terminal alpha helices (present in some Class I structures) indicating a nearby acid that binds to the amino group of SAM (**1a**), whilst a conserved acidic residue (or polar residue) at the C terminus in β sheet 2 has often been determined as binding the ribose hydroxyls.[22] Therefore, understanding the structural detail of the methyltransferases through this classification can enable researchers to uncover active site residues quicker.
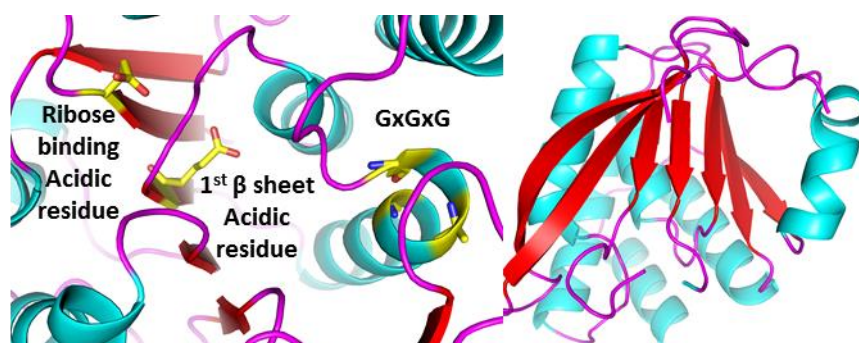
Figure 2: (Left) GxGxG and two acidic residues motif (yellow) from Class I methyltransferase catechol-*O*-methyltransferase (COMT). (Right) overall structure showing α helices (cyan) sandwiching the β sheet core (red). Structures adapted from PDB:1VID.[1]

The conservation and understanding of SAM (**1a**) binding and identification of active site residues through structural classification has enabled detailed understanding of this Class of enzymes. However, the classification rarely determines substrate binding residues due to the wide range of substrates that make up the Class I methyltransferases. Class I methyltransferases, methylate substrates from DNA and RNA to small molecules making determination of a substrate binding motif impossible. Therefore, more focussed studies on individual methyltransferases have elucidated how these substrates bind, as well as determining the activities of the enzymes.

An example of a Class I methyltransferase that has been thoroughly researched, with both structural and activity based studies being conducted; is the catechol-*O*-methyltransferase. COMT is a mammalian enzyme that was first isolated in 1958 by Axelrod and Tomchick[25], which was found to O-methylate catecholic neurotransmitters such as epinephrine.[25] Subsequent work has shown that COMT is a key enzyme in the metabolism of neurotransmitters (Scheme 2),[26,27] and further studies have often focussed on its inhibition due to enzyme's ability to methylate key drugs such as L-DOPA.[28,29] COMT inhibitors have been used clinically in the

treatment of Parkinson's disease in conjunction with L-DOPA, which is currently the best treatment for Parkinson's disease.[30] Their mechanism of action is to inhibit the methylation of L-DOPA, which allows for lower dosages of L-DOPA treatment, reducing L-DOPA side effects.[27] The COMT inhibitor entacapone is FDA approved for Parkinson's disease treatment. The activity studies of the COMT enzyme have shown that study of this Class of methyltransferases can often have key clinical results such as COMT inhibitors' use in Parkinson's disease.
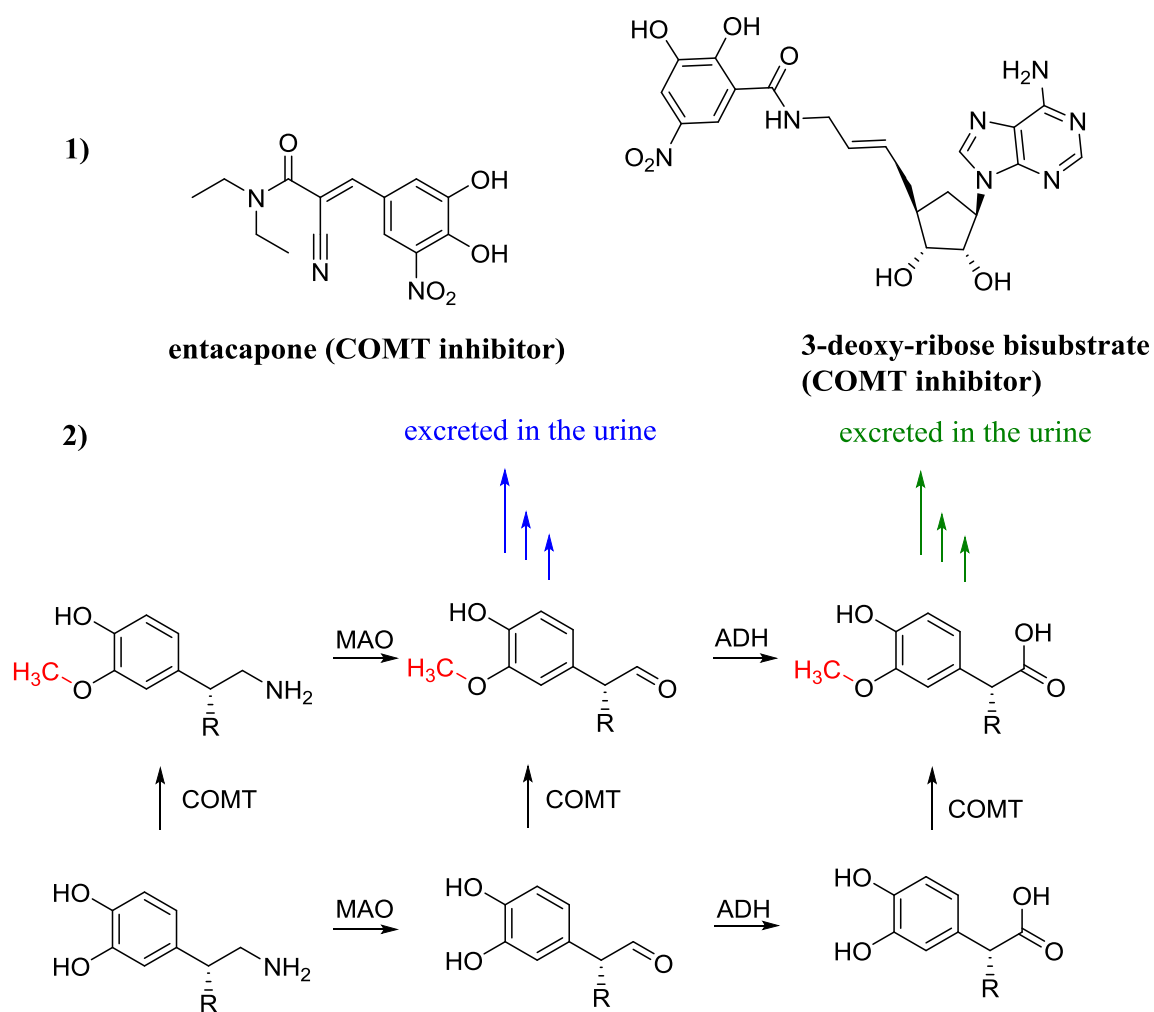
Scheme 2: 1) COMT inhibitor structure entacapone and an example of a 3-deoxyribose bisubstrate inhibitor. 2) COMT metabolism pathway (red) (inhibited by entacapone) of neurotransmitters dopamine (R=H) and noradrenaline (R=OH). Dopamine major metabolite is shown in blue (HVA) and noradrenaline major metabolite is shown in green (HHVA). Enzymes, monoamine oxidase (MAO) and alcohol dehydrogenase (ADH) are also shown.[5,27]

The initial studies into COMT focussed on the enzyme's activity and role in metabolism. Extraction of the cDNA for COMT meant that the enzyme could be fully purified and studies subsequently uncovered that COMT can exist as two

isoforms; soluble COMT (s-COMT) and membrane bound COMT (MB-COMT) coded for by the same gene.[31] S-COMT has since been found to be easier purify and isolate (as well as containing possessing similar properties to MB-COMT) so has been the focus of subsequent studies.[5] Improvement in crystallographic and enzyme purification techniques meant that the crystal structure for the COMT enzyme was solved by Vidgren *et al* in 1994.[1] In that seminal piece of work, Vidgren identified key residues in the COMT enzyme responsible for SAM (**1a**), substrate and the Mg^{2+} binding (Sections 2.1.1 and 2.1.2).[1] This structure was subsequently used to identify COMT 3-deoxyribose-bisubstrate inhibitors with further structural detail for how these inhibitors bind, as well as identifying a genetic polymorphism at the 108 position resulting in either a methionine (**2a**) or valine amino acid.[32,33] The identification of these key residues agreed with the motifs that were subsequently shown for Class I methyltransferases by Kozbial *et al*.[1,22] The structure can be used as a model for the small molecule methyltransferases as well as being used to identify key residues in newly resolved Class I methyltransferase structure.

The value of understanding and developing Class I methyltransferase can be observed by the clinical studies in COMT, especially for the enzyme's role in the treatment of Parkinson's disease. The large size of the substrate scope of Class I methyltransferases, as well as the number of small molecule methyltransferases present in this class, makes the Class I methyltransferases a good target when developing methyltransferases for biocatalysis.

1.1.3 Class II, III, IV, V

Class I methyltransferases has a large range of different substrates as well as having easily identifiable structural motifs.[22] However, despite the structural motifs, there is limited conservation of residues across the Class. Identifying conserved residues would enable the development of substrates and cofactors that could work with the whole class, as well as identifying residues that could fundamentally alter the enzyme's activity with mutagenesis. Examination of the remaining Classes of methyltransferase can be used to explore of the different functions and potential applications of these methyltransferases.

The Class II methyltransferases have been shown to have important biological functions but are limited to just one substrate; cobalamin.[22] Class II methyltransferases form part of the large multifunctional methionine synthase domain (found in mammals and bacteria), with the Class II regeneration domain operating a shuttle mechanism between tetrahydrofolate, two water molecules and SAM (**1a**) to methylate cobalamin (Figure 3).[15] Cobalamin subsequently methylates homocysteine (**2**) regenerating methionine (**2a**) in another domain of the synthase.[15] Whilst Class II methyltransferase could be used for SAM (**1a**) recycling, the large size, different cofactors and complicated machinery of the methionine synthase, have meant that studies have not focussed on further applications of the Class II methyltransferases.
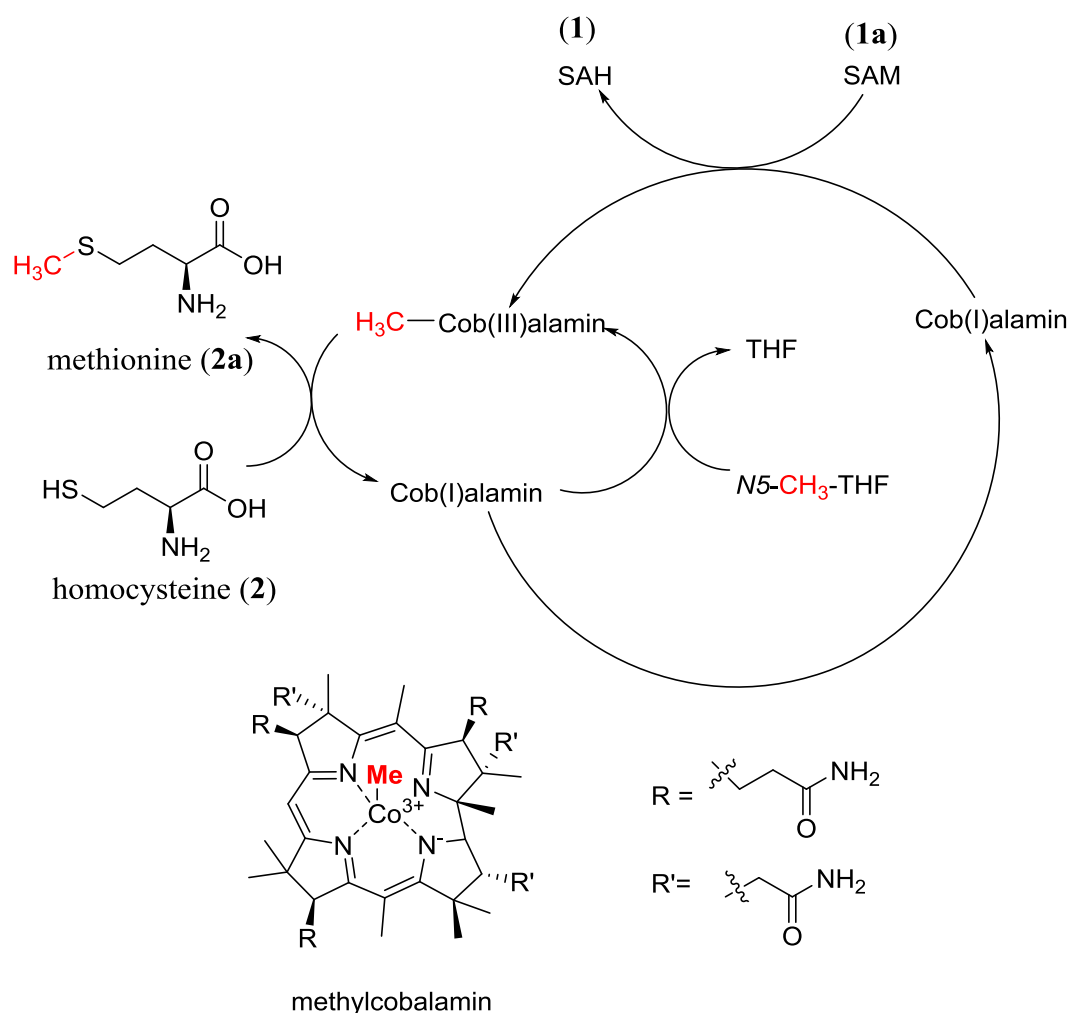
Figure 3: Methionine synthase catalytic cycles.[15]

The Class III methyltransferases have a similarly narrow substrate scope and important biological role. The main function of Class III is methylation of porphyrin substrates at several positions, which leads to the biologically important heme d_1 cofactor for nitrite reduction in plants and bacteria.[22,34] The Class III enzymes have a two domain structure.[22,34] The first domain consists of a structure of five parallel β sheets packed between three α sheets.[22,34] The second domain also consists of five β sheets sandwiched between three α sheets but the β sheets adopt a mixed β sheet structure.[34] Whilst a conserved lysine residue has been suggested to be important for catalytic transfer and several residues have been suggested for SAM (**1a**) binding, the limited number of studies has yet to identify conserved residues in this class.[34] Studies of this Class of enzyme have aimed at understanding the function of the enzyme and how Class III methyltransferases affect photosynthetic mechanisms through chlorophyll metabolism.

The Class IV methyltransferases have a wider substrate range than Classes II-III, although all substrates are either rRNA or tRNAs. Methylation occurs on a broad range of positions on the RNA bases from the N1 atom in guanine to the N3 atom of uridine.[35] The natural function of these methyltransferases is to modify the RNA to improve the folding and the stability of the RNA, which protects the RNA structure from attack by pathogens (RNA methyltransferases are found in organisms from humans to archae).[35] They are known as the SPOUT superfamily (name has been based on the two RNA methyltransferases spoU and trmd) with all structures of these methyltransferase forming a trefoil knot.[22] The trefoil knot is formed through the last alpha helix of the C terminal and the loop being threaded through a hoop that connects the 7 and 8 β sheets.[35] High conservation is found across the Class with arginine and aspartic acid residues all conserved, these residues are all involved in substrate binding (RNA).[35] Conservation is also noted in SAM (**1a**) binding with leucine residues forming a hydrophobic pocket around **1a**, this pocket is conserved across the class.[35] Control of **1a** binding could lead to directed methylation, which would allow control of the RNA's properties and control over cellular machinery. However, the large substrates and range of methylation patterns makes this control difficult to obtain, which has limited the development of this methyltransferase class for further functions and applications.
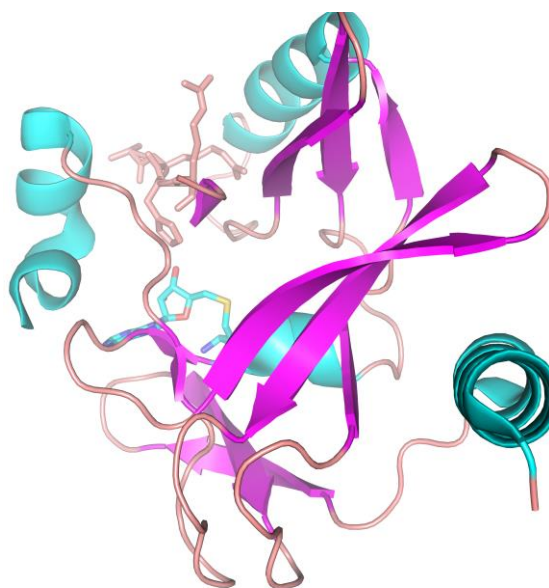
Figure 4: β clip in SET 8 methyltransferases. The knot like clip is indicated by the β sheets (magenta). Structure from Couture *et al* (3F9Y).[36]

The Class V methyltransferases are known as the SET domain methyltransferases.[22] These methyltransferases methylate histones mostly on the nitrogen atom of the side chain amine on lysine residues.[36] Histones are often involved in regulation of transcription and chromatin function.[36] Methylation can either repress or activate these functions.[36] β-sheets in SET8 methyltransferases form a knot like structure and this is the β-clip (Figure 4).[36] The most conserved residue across the class is a histidine involved in binding the adenine ring of SAM (**1a**) through π stacking interactions.[36] However, a tyrosine residue has also been commonly found, involved as a catalytic base, deprotonating the lysine prior to methylation (Figure 5).[37] Conservation is also found in the substrate with a sequence of K-R/S-T-A noted throughout the protein substrates.[36] This could allow for further development of these enzymes by engineering protein substrates to accommodate this conserved sequence. A number of studies have focussed on altering the substrates of these methyltransferases (for example the addition of ubiquitin to the histone H3 leading to enhanced methylation by the lysine methyltransferase Dot1 by R McGinty *et al*), due to their role in histone methylation.[38] Subsequent engineering of the cofactor, could then be used to label the selected proteins allowing for further modifications, indeed Binda *et al* have shown that engineered cofactors work with the SET 8 enzymes.[39]
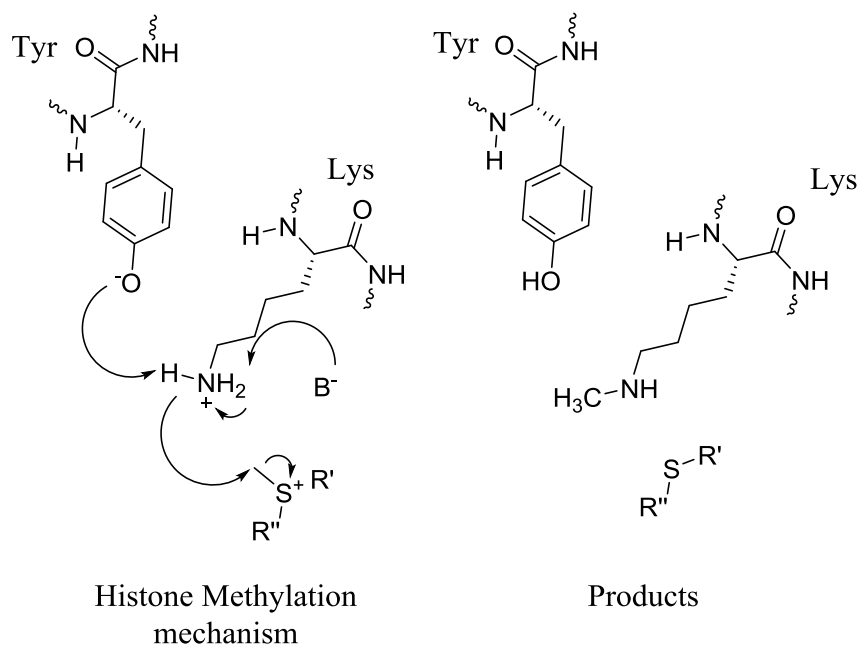
Figure 5: Suggested mechanism of lysine methylation by SET 8 methyltransferases with a catalytic tyrosine (B = tyrosine residue).[37]

Class II-V methyltransferase have all shown a diverse range of structural and functional motifs. These classes also have good conservation of residues across their Classes, making the active sites of unidentified methyltransferases belonging to these Classes easier to find. However, despite Class II methyltransferases being developed for a possible role in SAM (**1a**) recycling, and Class V methyltransferases being developed for possible protein labelling, the potential applications for Classes II-V are limited. Therefore, the focus of this study will be on Class I methyltransferases and the small molecule substrates of this Class.

1.2 Developing alkyltransferases

1.2.1 Existing synthetic means of alkylation

The understanding and classification of methyltransferases has meant that these enzymes' active sites can be understood, and therefore manipulated through mutagenesis to improve activity and selectivity. However, in developing these enzymes for biocatalysis, it is first necessary to consider the existing synthetic

technologies in order to determine whether enzymes can improve upon these existing techniques. A recent review by Schönherr and Cernak indicated that 67% of the top 200 drugs have a C-Me bond.[40] The prevalence of this C-Me bond has been attributable to the "magic methyl effect" where medicinal chemists improve the half-life and affinity of drug candidates by introduction of a methyl group.[2,40] The most prevalent methyltransferases across nature are *C-*, *N-* and *O-* methyltransferases. Therefore, understanding existing synthetic techniques for introducing these "magic methyls" to *C*, *N* and *O* is important when considering the development of methyltransferases for biocatalysis.

The C-C bond forming reaction of methyltransferases are a common tailoring step used in both polyketide synthases (PKSs) and non ribosomal peptide synthetases (NRPSs) to functionalise secondary metabolites.[41] C-C bond forming reaction is also common in synthetic chemistry with a wide range of chemistries developed for different functional groups. The earliest examples of C-C bond forming reactions are the; Grignard and Friedel-Crafts reactions (Scheme 3).[42-44] Grignard reactions alkylate ketones or aldehydes using alkylated magnesium halides in nucleophilic addition reactions.[42] However, the reactions are limited by the lack of regio- (no functional group tolerance for acids, and will react with any ketones on the molecule) and stereo- control.[44] Alternative organo-metals that have been used in alkylation chemistries are the organolithiums (Scheme 3). Organolithiums, such as LDA or BuLi, can be used to initial deprotonate carbons alpha to a ketone, to form an enolate intermediate, which then attacks electrophilic alkylating agents such as organohalides to complete the alkylation.[44,45] The enolates formed by organolithium reagents have also been used in conjugate additions (also known as the Michael addition), where the enolate attacks an enone (or alkene with an electron withdrawing group attached) to form a new C-C bond.[44,46] The Friedel-Crafts reaction alkylates aryl reagents through an electrophilic substitution mechanism.[43] Friedel Crafts reactions were hampered by overalkylation and limited substrate range (only aryl groups could be used).[44]
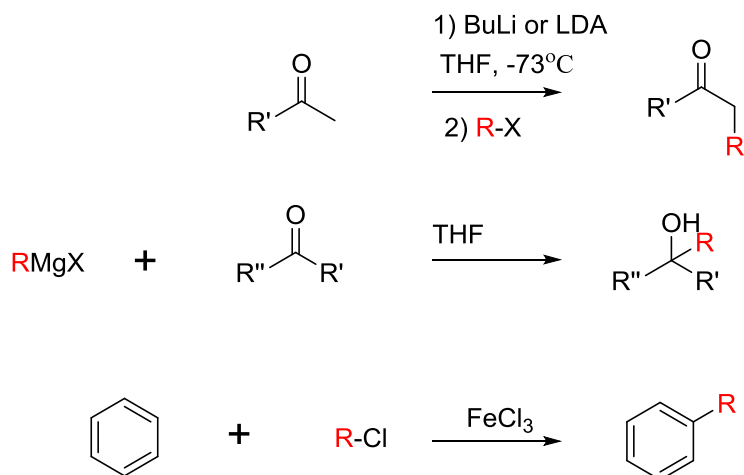
Scheme 3: (Top) Grignard reaction.[42] (Middle) Organolithium alkylation through a lithium intermediate.[45] (Bottom) Friedel Crafts reaction.[43] R=alkyl group, R'=H/alkyl and R"=alkyl chain.

The disadvantages of all three reactions have been improved by catalysts as well as different reagents. The reagent range of the Friedel Crafts reaction has been expanded from aryls to alkene and to even aliphatic groups through an acylation reaction with an aluminium chloride catalyst (although this does require an extra reduction step).[47,48] The acylation reaction, with an acyl chloride, also controls overalkylation.[44] Whereas both organometal reactions have focussed on improving the selectivity of the reactions. The Grignard reaction's regio- and stereo- selectivity has been improved by both palladium (through oxidative addition and reductive elimination pathways) and nickel catalysts (Kumada reaction).[49,50] Whilst, organolithium reagents have been used with chiral auxillaries to perform diastereoselective alkylations,[51,52] as well as stereospecific conjugate additions through carefully controlled conditions and reagents.[53] However, the improvements for synthetic C-C bond forming reactions are often limited in reagent range,[54] require expensive[49] and difficult to assemble catalysts or chiral auxillaries[50,52] in organic solvents.[45,49,50,52,54]

The methylation of oxygen atoms is also a common methylation reaction found in nature with examples ranging from COMT metabolism to PKS methylation.[5,41] Synthetic *O*-methylation has been around for 160 years with the earliest example being the Williamson ether synthesis,[55] the primary method by which most *O*-methylation has been based on (Scheme 4).[56,57] The Williamson ether method has been shown to be an S_N2 alkylation mechanism, with an alkoxide (formed with base such as sodium hydride) attacking an alkyl halide (leading to an inversion of

48

stereochemistry at chiral alkyl halides) to complete the *O*-alkylation.[44,58] Initially these reactions required harsh conditions, with strong bases such as sodium hydride and high temperatures,[56] however, with the addition of phase transfer catalysts (such as tetrabutylammonium bromide)[59] milder conditions (reactions could be performed at room temperature)[59] were developed for these reactions.[57] These improvements have meant that the Williamson ether synthesis (or a derivative thereof) is still the primary method of choice for *O*-alkylation reactions.
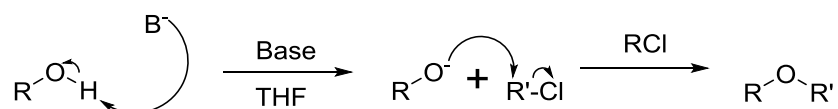
Scheme 4: Example of a Williamson ether synthesis. R=alkyl chain and R'=alkylating agent.[44,57]

The C-N bond is also found across nature, with a vast array of *N*-methyltransferases found with functions ranging from alkaloid biosynthesis to neurotransmitter metabolism.[21,60] There have been a number of *N*-alkylating methods from reductive aminations to the Buchwald Hartwig reaction.[61-63] However, reductive aminations have been used with a variety of reagents and is the main *N*-alkylation technique discussed here. Reductive aminations are performed on ketones where the ketone is attacked by a secondary amine (for alkylation reactions) resulting in an imine, which is reduced by a reducing agent such as sodium borohydride.[44] Originally these reactions were carried out in non-aqueous conditions with flammable catalysts such as Raney-Nickel.[61] One method that avoids these harsh conditions, is the Eschweiler-Clarke reaction (Scheme 5). This method uses formaldehyde as the starting reagent with an amine, which attacks the formaldehyde and the subsequent intermediate is reduced by formic acid yielding the methylated product.[64] Other methods of reductive aminations have involved iridium catalysts, have led to reductive aminations being carried out in milder aqueous conditions with alcohols instead of ketones (through intermediate oxidation by the iridium catalyst).[65]
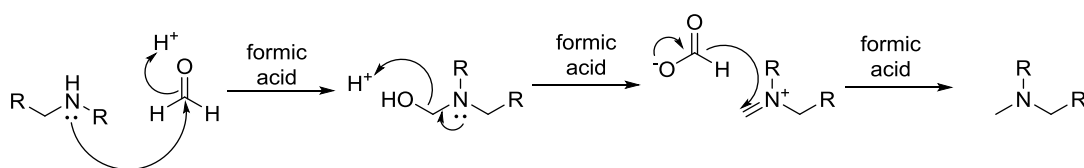
Scheme 5: Eschweiler Clarke reaction.[58]

There are a number of synthetic methods for C-, N- and O- alkylations with different functional group tolerance and selectivities. The alkylations have improved upon existing alkylation reactions (Grignard, organolithium techniques, Friedel Crafts, Williamson Ether and reductive amination reactions) to provide greater selectivity with milder conditions.[47,50,52,57,65] However, barring the Eschweiler Clarke reaction, which is limited by the dialkylation of primary amines, all of the alkylation techniques use alkyl halides, which are genotoxic, as alkyl halides spontaneously alkylate DNA.[58,66] Alternative alkylating agents such as methyl triflate offer no safer alternative, with methyl triflate hydrolysing violently with water.[67] The alkylation synthetic reactions are also limited by their selectivity, which is only improved with expensive catalysts, which are iridium or palladium based.[49,65] Therefore, the development of methyltransferases could offer a more selective alternative, as examples of selective methyltransferases has been observed (for example MppJ).[19]

1.2.2 Methods of developing SAM analogues

The last section clearly indicated a number of drawbacks for existing synthetic alkylation techniques (Section 1.2.1). The lack of selectivity for existing alkylating techniques can be improved by using methyltransferases. However, as the previous Section showed (1.2.1), existing synthetic techniques are alkylations rather than methylations. Whereas wild-type methyltransferases only transfer a methyl group. Therefore, in order to develop alkyltransferases, numerous groups have engineered SAM analogues with longer alkyl chains to determine whether alkyl transfer is possible.[4,7,39,68-71]

S-adenosyl-L-ethionine (**1b**)   S-adenosyl-L-propyl-homocysteine
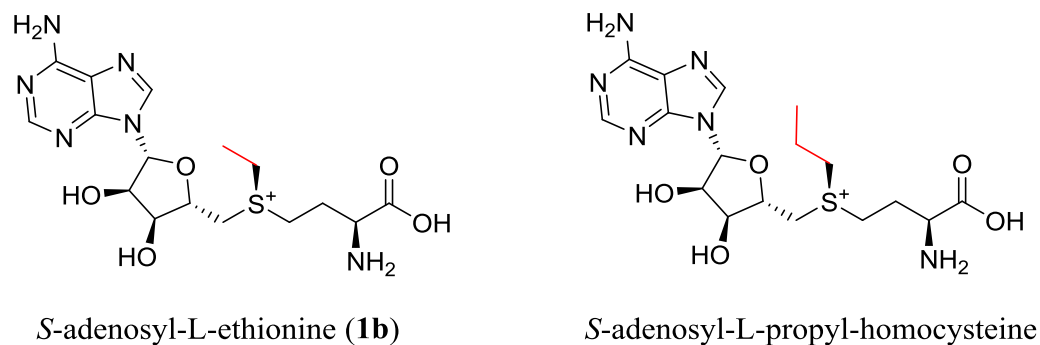
Figure 6: Structures of S-adenosyl-ethionine (SAE) (**1b**) and S-adenosyl-propyl-homocysteine (SAPH).

Development of alkyltransferases has focussed on the SAM (**1a**) sulfonium centre, as alkyl transfer occurs from this point. Early studies focussed on biological methods to alter the sulfonium centre by biological supplements that would affect the metabolic cycle of **1a** (Scheme 6).[68,69] An example of this early biological strategy was developed by Parks, who grew yeast cells on ethionine supplemented media (used to biosynthesise SAE with MAT (Scheme 6), isolating the SAE (**1b**) (Figure 6) and subsequently using SAE (**1b**) in ethylation reactions .[68] This work was then emulated by Schlenk but with S-propyl-homocysteine (SPH) supplements, and subsequently SAPH was isolated and used in propylation reactions.[69] However, subsequent studies with the SAE (**1b**) and SAPH cofactors also found that activity with these cofactors was poor compared to the methylation assays.[68,69,72] This suggested that alkylations with methyltransferases may not be a viable alternative to synthetic alkylation techniques.
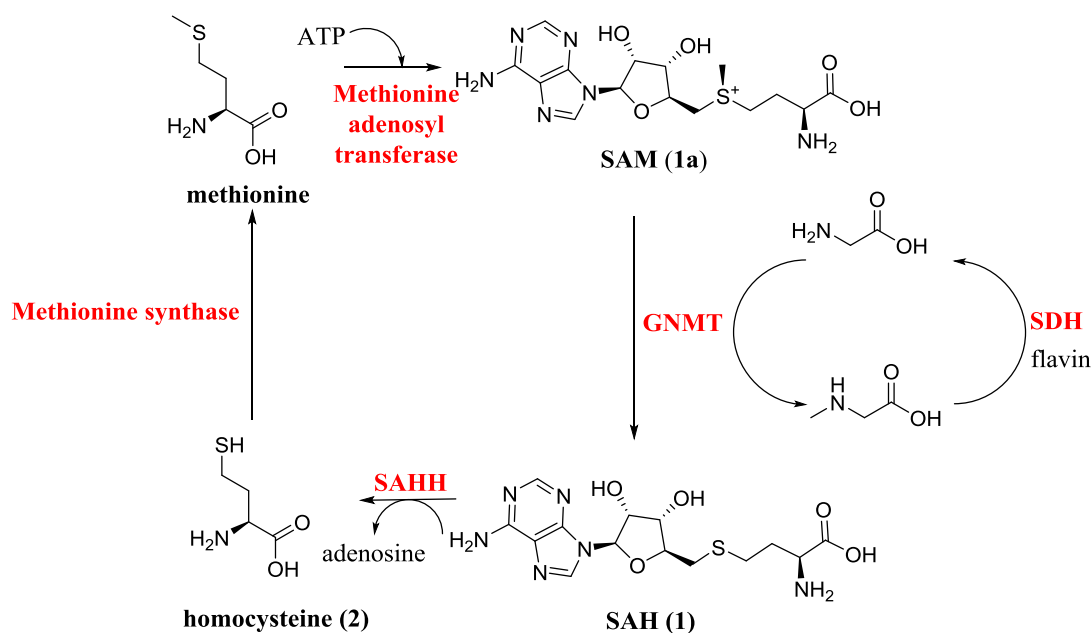
Scheme 6: Simplified metabolic pathway of SAM (**1a**), enzymes are shown in red and key intermediates are labelled in bold.[73] **1a** is the cofactor of the glycine-*N*-methyltransferase (GNMT), which methylates glycine (that is regenerated from sarcosine with sarcosine dehydrogenase (SDH) enzyme) and the SAH (**1**) byproduct is broken down into adenosine and homocysteine (**2**) with SAH hydrolase (SAHH). Parks and Schlenk have adapted the pathway through ethionine and *S*-propyl methionine supplements at the methionine adenosyl transferase (MAT) stage.[68,69]

Initial studies with the saturated SAM analogues of SAE (**1b**) and SAPH had shown poor conversion.[68,69] However, work conducted by Dalhoff *et al* found that introduction of unsaturated groups at the sulfonium centre, adapting a synthetic method from Zorbach *et al*,[74] moderately improved activity of the SAM analogues.[4] The method comprised of an S_N2 alkylation of the sulphur of SAH (**1**), with the corresponding alkyl bromide and acid solvents for protection of the amino and hydroxyl groups.[4,74] The improved activity was shown by the *S*-adenosyl-allylhomocysteine (SAAH) (**1c**) and *S*-adenosyl-propargylhomocysteine (SAPrH) cofactors showing higher activity with DNA methyltransferases M.TaqI, M.Hhal and M.BcnIB (with corresponding DNA substrates) compared to SAE (**1b**) and SAPH cofactors.[4] The generally accepted reason for the stabilisation was proposed by Dalhoff *et al* and was that the π* antibonding orbital of the allyl chain of SAAH stabilises the p orbital, present at the reactive carbon, that was necessary for the S_N2 alkylation reaction with the substrate (Nu) to proceed (Scheme 7).[4] Subsequent studies strongly suggested that this hypothesis was correct, with studies using *S*-adenosyl-benzylhomocysteine (SABH) (**1d**) and various propargylic based

analogues showing modest increases in activity compared to SAE (**1b**) and
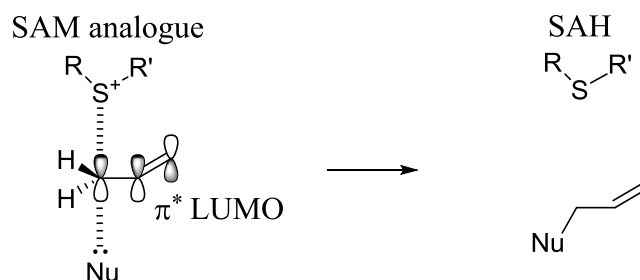
SAPH.[7,39,71]   Nu = Methyltransferase substrate
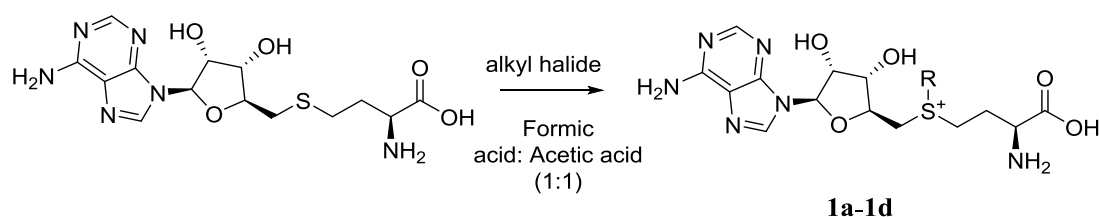
(**1a**) R = CH_3; (**1b**) R = CH_2CH_3; (**1c**) R = $CH_2CH=CH_2$; (**1d**) R = CH_2Ph

Scheme 7:(Top) Suggested stabilisation of the transfer of an allyl chain from the SAM analogue (SAAH) to a methyltransferase substrate (Nu) forming the allylated product (R and R' represent the amino group and adenosyl group of SAM (**1a**) respectively).[4] (Bottom) Dalhoff *et al* method for synthesising SAM analogues (adapted from Zorbach's method).[4,74]

The work of Dalhoff *et al* meant that subsequent studies transferred a rich variety of alkyl chains. This variety was shown by Stecher *et al*, which was the first group to use these SAM analogues with small molecule substrates.[7] Using the enzymes NovO and CouO a range of coumarin structures were alkylated at the C_8 position on the aromatic ring of the coumarin showing desirable regioselectivity, which may be due to the OH substituent at the C_7 position, however, to date no studies have confirmed this (Scheme 8).[7] Stecher *et al* also improved the synthesis developed by Dalhoff *et al*,[4,7] through introduction of the Lewis acid activator silver triflate, with catalysis Stecher *et al* were able to achieve quantitative yields in 1-2 days.[7] Stecher *et al* had thus shown that a range of C-C alkylation reactions could be carried out with two methyltransferases on small molecules,[7] opening the possibility that small molecule methyltransferases could be used for biocatalysis.
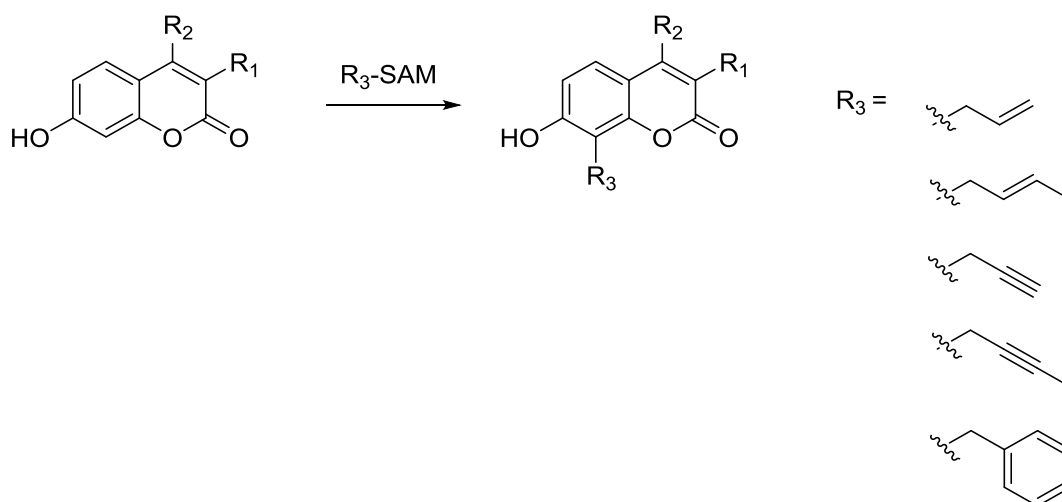
Scheme 8: Stecher *et al* range of alkylation reactions with NovO and CouO methyltransferases.[7]

Stecher *et al* had shown that a facile, quick and high yielding synthesis of SAM analogues with a number of different alkyl groups could be conducted.[7] However, with a number of groups using SAPrH analogues due to the ability to conduct click chemistry based labelling studies, a recurring problem of hydrolysis of the SAM analogue was noted through an allene intermediate (Scheme 9).[59,75,76] Previously selenomethionine had been shown by Mudd and Cantoni to produce the *Se*-adenosyl-selenomethionine (SeAM) analogue, which could be used to methylate glycocyamine with a creatine methyltransferase.[77] Studies have suggested that the due to the weaker C-Se, the SeAM is more active than SAM (**1a**), as the C-Se bond has been shown to be more labile during the S_N2 reaction.[78] The *Se*-adenosyl-propargyl-selenohomocysteine (SeAPrH) analogue was also suggested to effect the neighbouring CH_3 protons, by reducing their acidity, which may make the allene intermediate unable to form (Scheme 9).[78] Based on these hypotheses, Bothwell *et al* synthesised the SeAPrH and did find that the properties of SAPrH were significantly improved by introduction of the selenium with a large reduction in degradation noted (improved of $t_{1/2}$ from <1mins to <60mins).[75] This recent study has shown that improvement of the synthetic SAM analogues can be continued and that selenium analogues should also be considered as well as sulphur based analogues.
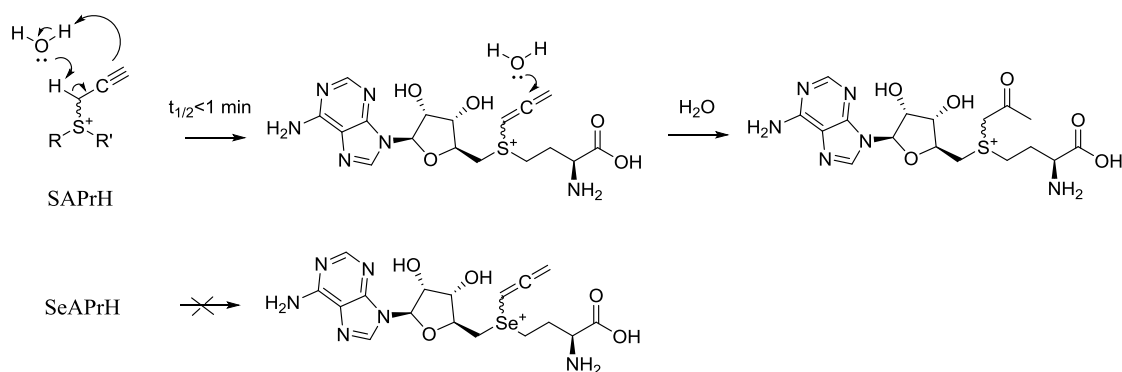
Scheme 9: (Top) Proposed hydrolysis mechanism of SAPrH.[75] (Bottom) Bothwell *et al* SeAPrH does not form the allene as quickly at physiological pH.[75]

Synthetic analogues have shown that alkylation reactions with methyltransferases were possible and that introduction of adjacent sp^2 hybridised groups could improve activity.[4] Initial problems with the SAPrH analogue were also overcome with the introduction of selenium analogues, which improved the yield of these reactions through decreasing degradation by hydrolysis.[75] However, one problem, noted in the synthetic studies,[4,72] was the lack of diastereomeric purity.[4] The S_N2 reaction (as observed in the figures in this section) does not control the direction of alkylation at the sulphur centre and therefore both R and S isomers are made at the sulphur centre. [4]Hoffmann *et al* (amongst others) showed that racemisation at the sulphur centre did occur but the inversion barrier is high enough (105-120 kJmol^{-1})[9] that both diastereomers have to be considered in methyltransferase reactions.[9,79,80] Borchardt *et al* had also shown the (R_s S_c)-SAM (**1a**) was at best inactive and at worst inhibitory.[72] Therefore, in order to solve the diastereomeric impurity of (R_s S_c)-SAM analogues alternative strategies to synthetic SAM analogues were needed.

1.2.3 Enzymatic production of SAM analogues

The problem of diastereomeric impurity was observed in a number of studies of SAM's (**1a**) sulfonium centre.[72,79,80] However, these studies did also determine that enzymatically produced **1a** was a single diastereomer.[79] The result of these studies meant that researchers now consulted an enzymatic solution to the problem of **1a** diastereomers. The two main enzyme classes that have been trialled in the enzymatic production of SAM analogues are; halogenases and MAT (Scheme 6).[8,59,81] These classes of enzymes have produced SAM analogues *in situ* with associated

methionine analogues and coupled the enzymes to downstream methyltransferases with varying degrees of success.[8,81]

There have been two main halogenases used in the *in situ* production of SAM analogues; SalL chlorinase and the fluorinase FDAS enzyme.[82,83] Both enzymes belonged to the halogenase class, with the fluorinase enzyme being the first of its kind to be identified,[83] that halogenate SAM (**1a**) with a fluoride or chloride ion at the C5' on the ribose ring of **1a** (Scheme 10).[82,83] The halogenation of **1a** by SalL and FDAS was reversible, therefore, using methionine analogues (synthesised by a sodium liquid ammonia reduction of L-homocystine and alkylation with alkyl bromides) and 5'-chloro-adenosine Thomsen *et al* were able to form SAM analogues *in situ* and transfer alkyl chains to peptides using the PRMT enzyme.[81] This work was the first diastereoselective *in situ* production of SAM analogues, and the first development of a coupled system for enzymatic alkylation.[81]

(**1a, 2a**) R = CH$_3$; (**1b, 2b**) R=CH$_2$CH$_3$; (**1c, 2c**) R= CH$_2$CH=CH$_2$; (**1d, 2d**) R= CH$_2$Ph
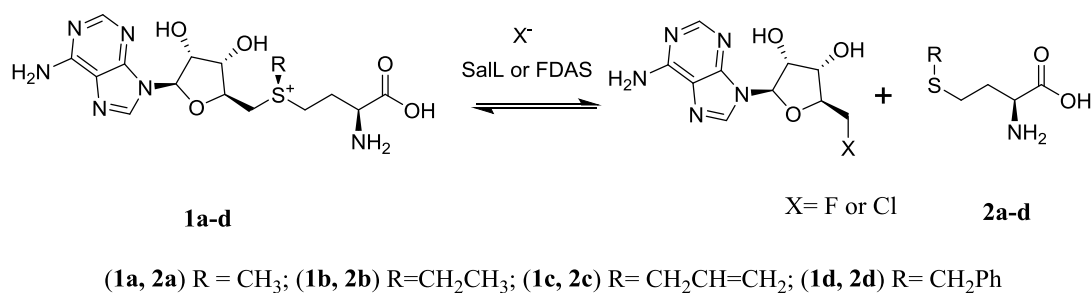
Scheme 10: Reversible reaction of FDAS and SalL enzymes with SAM (**1a**). Introduction of methionine analogues can make SAM analogues *in situ*.[81]

However, despite solving the problem of diastereomeric contaminants, the SalL and FDAS systems had a number of drawbacks. Firstly, the formation of the SAM analogue involved 5'-chloro-5'-deoxy-adenosine,[81] which is expensive and not readily available. Secondly, despite an extensive mutagenesis strategy the activity of SalL and FDAS was much lower with the methionine analogues than with methionine.[81] For example the *S*-allylhomocysteine (SAlH) homocysteine analogue was 4930 times slower with the SalL enzyme than with methionine cofactor.[81] Despite the equilibrium favouring the SAM (**1a**) production,[84] the introduction of the larger alkyl chains was hindered by the presence of two water molecules in SalL's binding pocket for methionine, which intensive mutagenesis did not remove.[81]The drawbacks of the SalL enzyme for producing SAM analogues were substantial and the development of selectivity had sacrificed the yield of SAM analogues.

The other method of producing SAM analogues *in situ* is with the MAT enzyme.[11] The MAT enzyme is found naturally in the metabolic cycle of SAM (**1a**) and is utilised to make **1a** from methionine (**2a**) and ATP (Scheme 7).[11,85] Wang *et al* used this enzyme, again with methionine analogues, to produce SAM analogues *in situ* and then subsequently alkylate histones using the protein arginine methyltransferase (PRMT1).[59] In order to accommodate the larger methionine analogues (such as propynl-oxy-2-butene-SAM (POB SAM) (**1e**)) mutagenesis was conducted on the alkyl binding pocket of *E.coli* MAT (eMAT) (the alkyl binding pocket was noted by Wang as being conserved across the class). The I302V mutant of eMAT was identified as having improved activity with the longer alkyl chain SAM analogues (Figure 7).[59,86] The mutation was thought to maintain the structural integrity of the pocket whilst reducing steric clashes with the longer alkyl chains on the methionine analogues (Figure 7).[59,86] Therefore, Wang *et al* had shown that the MAT enzyme could be used to generate SAM analogues *in situ* with less costly cofactors, and without diastereomeric impurities.
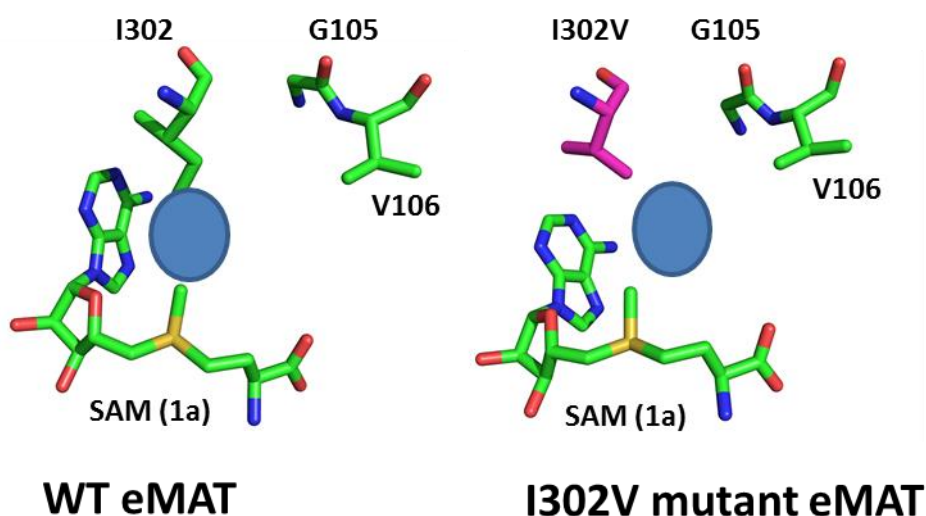
Figure 7: From PDB structure (1RG9), the binding pocket (blue circle) of the alkyl chains of SAM analogues in the eMAT enzyme.[59,86]

The study by Wang *et al* did offer a viable alternative to the SalL cofactor generation.[59,81] However, the initial study on the eMAT only focussed on a few methionine analogues and a few SAM analogues.[59] A more comprehensive study was conducted by Singh *et al*, in which 5 variants of MAT were tested; eMAT (eMAT), human MAT1A (hMAT1A, human MAT2A (hMAT2A), human MAT2

(hMAT2) and *Methanocaldococcus jannaschii* MAT (mMAT) with 24 methionine analogues (and selenomethionine analogues) (Table 2).[8] Singh *et al* showed that (Table 2) the activity of the MAT enzymes with the methionine analogues is far greater than SalL (with SAlH approximately 2500 times faster with hMAT2A than SalL).[8] The investigation subsequently used Rebeccamycin 4-*O*-methyltransferase to alkylate with ethyl, allyl and propargyl groups.[8] Therefore, using MAT to produce SAM analogues overcame the problems of the SalL system (poor cofactor and activity) whilst maintaining the selectivity.[8,81]

Table 2: Methionine analogues and the turnover number with the 5 MAT variants.[8] SePrH=*Se*-propargyl-selenohomocysteine.

Methionine analogue	eMAT	hMAT1A	hMAT2A	hMAT2	mMAT
Methionine (**2a**)	90	95	100	100	100
Ethionine (**2b**)	8	75	90	88	94
SAlH (**2c**)	2	5	50	52	16
SePrH	0	7	44	58	14

SAM analogues have been successfully produced by the two enzymes; SalL and MAT. Both enzymatic transformations produce the SAM analogues as single diastereomers. However, SalL has the main drawback of low activity with the methionine analogues that are used to make the SAM analogues. The problem of low activity is overcome by the MAT enzyme, which displays better activity whilst maintaining the high selectivity.[8] Therefore, future *in situ* production of SAM analogues should be carried out with the more active MAT enzyme.

1.2.4   Applications of SAM (1a) and SAM analogues

The development of SAM analogues by both synthetic and enzymatic means has led to a number of applications. These applications have ranged from labelling studies to selective alkylations of large cyclic polyketides.[59,71,87] This section will focus on how these studies have led to a further understanding of the utilisation of

methyltransferases with SAM analogues. Before concluding on how these developments can be used in future studies of methyltransferases.

The development of SAM analogues has led to a number of studies aimed at using the analogues to label biomolecules; the key examples being DNA and proteins. Previous labelling techniques, have often performed randomly or with large bulky labels that alter the activity or structure of the biomolecule.[2,88,89] Building on the work of Dalhoff *et al*, Lukinavičius *et al* used selective DNA methyltransferases (M.TaqI and M.Hhal), with a propargylic amine SAM analogue, to label sections of DNA, which subsequently reacted with a fluorescent probe through a Huisgen cycloaddition (CLICK reaction) (Scheme 11).[90] The studies also showed that the labelled DNA remained coiled, keeping the DNA's structural integrity, and subsequent characterisation of the nucleotides showed that labelling was specific.[90] This study had shown that previous issues could be overcome by using SAM analogues to specifically label molecules. However, the main drawback of this study was that the SAM analogue was difficult to assemble with a low yield of 3%.[90]
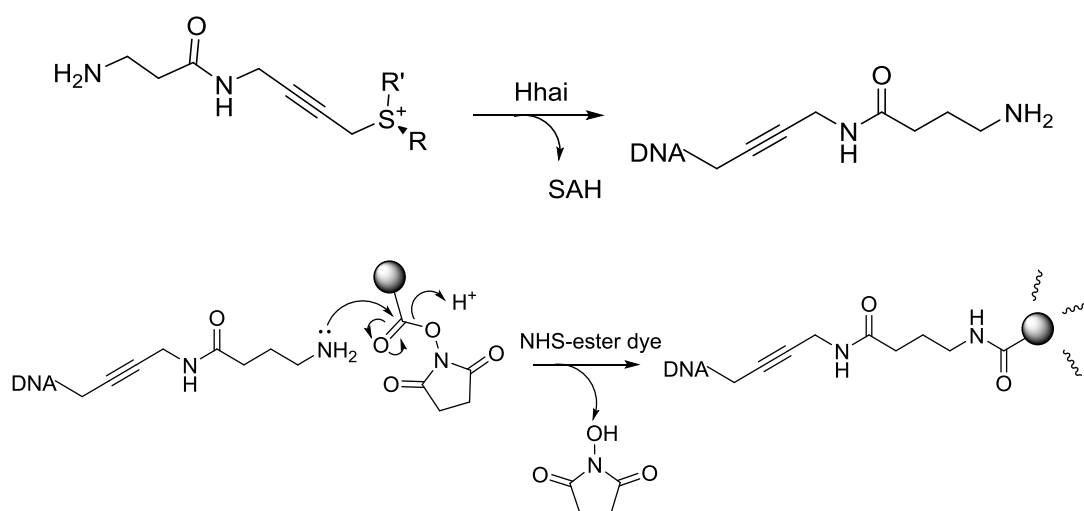
Scheme 11: (Top) DNA methyltransferase HhaI transfers the propargylic chain to segment of DNA. (Bottom) Subsequent reaction with NHS-ester dye leads to fluorescent signal.[90]

The work of Lukinavičius *et al* had shown that labelling could be achieved with SAM analogues and methyltransferases could be used for biomolecular labelling.[90] This work was used to also develop more facile labelling systems. Using PRMTs and SET8 enzymes, protein labelling techniques were developed with the SAPrH.[39] As Scheme 11 shows, Lukinavičius *et al* had built a propargylic group into their chain but this was not terminal, due to aforementioned problems with SAPrH

(Section 1.2.2 and Scheme 9).[75,90] Initially Wang *et al* overcame this problem using the POB SAM (**1e**), which had a chain with an alkene group before a terminal propargyl group.[71] The POB group (**2f**) was transferred by PRMT to arginine residues on histones.[71] The propargyl group was subsequently reacted in an azide-alkyne Huisgen cycloaddition with a rhodamine dye, leading to detection of the alkylated histone by fluorescence.[71] Further improvements were made by Bothwell *et al*'s previously mentioned SeAPrH analogue (Section 1.2.2), leading to similar detection of labelled histones but with a simpler propargyl handle.[75] Finally Wang *et al* progressed this work to labelling histones in living cells, by using the hMAT2A enzyme and cell engineering.[59] This work coupled with the other studies has shown that *in vivo* labelling of targets is possible and SAM analogues can be used to identify targets of unknown methyltransferase.[59]

S-adenosyl-POB-homocysteine
(POB SAM) (**1e**)
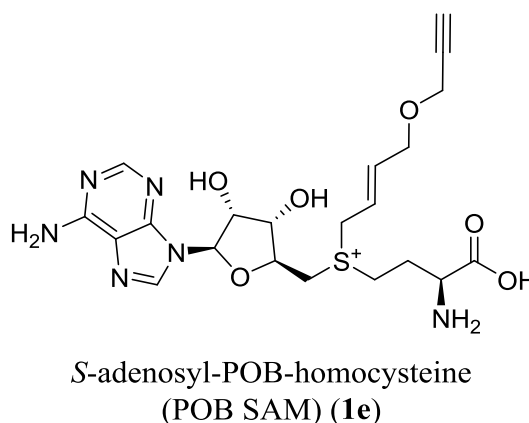
Figure 8: The structure of POB SAM (**1e**).[71]

The advancement of labelling with SAM analogues and methyltransferases have shown that this is a strong area of study as well a good alternative to existing techniques. By comparison the field of methyltransferase biocatalysis is in its infancy. Stecher *et al* have shown that small molecule alkylation is possible, however, there have been few studies since that have developed on this starting point.[2,7] Recently the Micklefield group have shown the advantage of developing methyltransferases for biocatalysis. The study aimed at using the hMAT2A enzyme in a coupled assay with a selective rapamycin *O*-methyltransferase (RapM) and methionine analogues (generating SAM analogues *in situ*).[87] Three *O*-methyltransferase tailoring enzymes are found at the end of the rapamycin PKS pathway and *O*-alkylate in three different positions (Figure 9).[91] Selective alkylation

of the rapamycin molecule was attractive, as the rapamycin has several analogues that are used clinically as immunosuppressants, which often require semi-synthetic methods.[92-94] Selective alkylation was indeed shown with hMAT2A I322V mutant and the RapM enzyme (Figure 9).[87] This work showed that by using selective methyltransferases with hMAT2A I322V and methionine analogues that selective and potentially medically relevant analogues of commercially relevant drugs was possible.

pre-rapamycin   rapamycin analogue
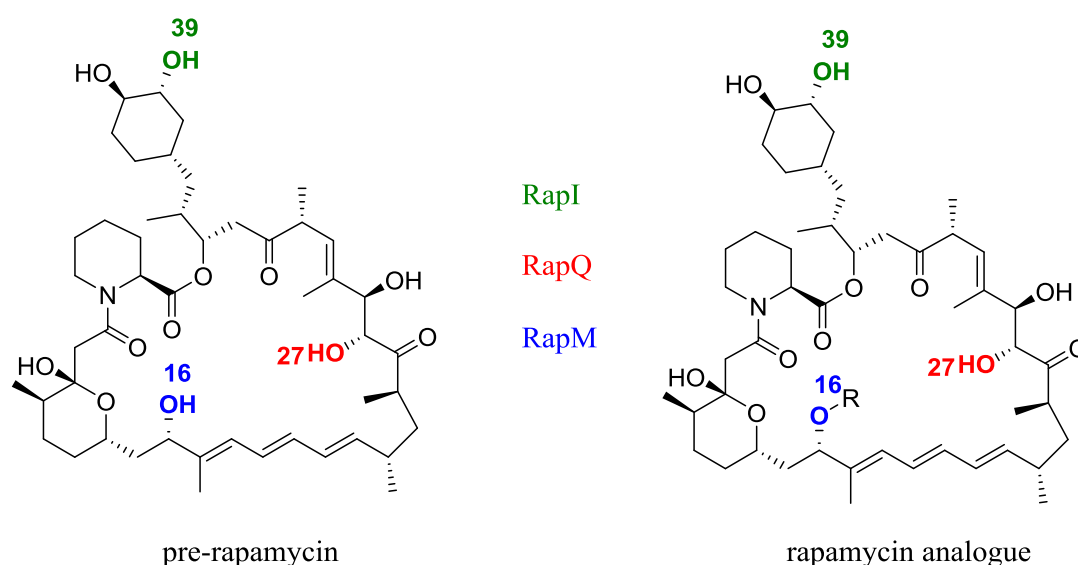
Figure 9: (Left) Pre rapamycin sites of methylation by tailoring O-methyltransferases; RapI, RapQ and RapM. (Right) Alkylation site of RapM R=methyl, allyl and ethyl.[87]

The studies of methyltransferases with alkylating SAM analogue cofactors have two main applications; labelling and biocatalysis. Extensive study with labelling methods have shown that *in vivo* labelling is now possible improving the selectivity and activity from previous labelling methods.[59] Selective alkylations have been shown for biocatalytic applications with the RapM methyltransferases.[87] However, as yet few studies have focussed on mutagenesis or alteration of the methyltransferases themselves. Improving the selectivity or expanding the substrate scope of methyltransferases could lead to a wider range of selective alkylations on a wider range of molecules.

1.3 Developing methyltransferase selectivity

1.3.1   Engineering methyltransferase selectivity

The previous sections have shown that alkyltransferases could be developed from methyltransferases with cofactor engineering. However, few studies have focussed on engineering the methyltransferases for improving the selectivity of these alkylations. Selective biocatalysts have already found their role in industrial processes. For example Lipozyme^{TM} is a regioselective lipase, which is used to esterify fats and amongst other functions; can be used to make margarine.[95,96] When engineering methyltransferases for biocatalysis, the selectivity of methyltransferases will be important. Of the multitude of methyltransferases, this section will focus on the selectivity of C-C keto acid methyltransferases and COMT.

 The enantio- and regio-selectivity of Lipozyme^{TM} has led to a number of useful industrial functions.[95] The advantage of this selectivity has built on the existing selectivity of the natural lipase from *Rhizomucor miehei*.[96] With this in mind the enantioselective *C*-methyltransferases MppJ and GlmT offer an attractive proposal for use with SAM analogues.[19,97] Both of these methyltransferases act on keto acids methylating at the 3 position (Scheme 12).[19,97] The selectivity of the two methyltransferases are different, however, with GlmT methylating to produce the *R* stereoisomer whereas MppJ produces the *S* stereoisomer, as shown in Scheme 12.[19,97] Racemisation will occur at the methylated centres, through formation of an enolate and reprotonation, without either the adenylation domain or a transaminase (TyrB) coupled to the methyltransferases.[19,97] These enzymes offer an attractive route into unnatural amino acids and that the enantioselective methyltransferases can be engineered with coupled enzymes. Therefore, further studies should be conducted with methyltransferases and other enzymes, developing coupled systems, which could improve selectivity.
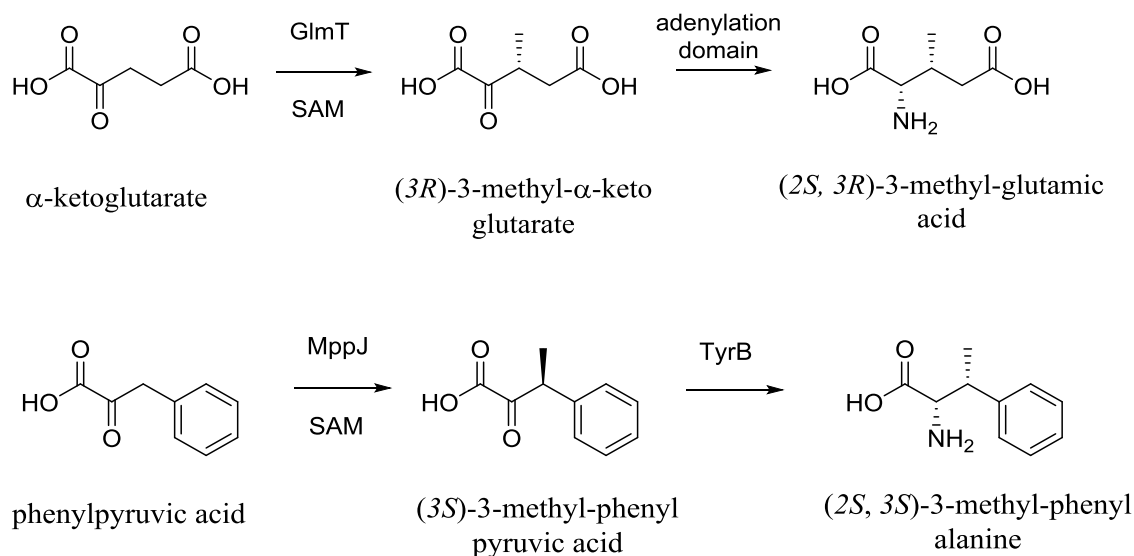
Scheme 12: (Top) GlmT *3R* keto acid methyltransferase with adenylation domain to form unnatural amino acid (*2S, 3R*)-3-methyl-glutamic acid.[97] (Bottom) MppJ *3S* keto acid methyltransferase with a transaminase TyrB to form unnatural amino acid (*2S, 3S*)-3-methyl-phenylalanine.[19]

Previous studies have shown that with coupled transaminases and adenylation domains; MppJ and GlmT methyltransferases are enantioselective.[19,97] Whereas a number of studies have shown that COMT has limited regioselectivity (Scheme 13). Since Creveling *et al* recorded the regioselectivity with a number of substrates (Table 3), COMT has been determined to show some selectivity for methylating the *meta* hydroxyl of catechols.[98] The *meta* selectivity of COMT has since been suggested by Tsao *et al* to be dependent on the substrate and that the enzyme has evolved to methylate the *meta* position due to a lower activation energy for *meta* methylation as opposed to *para* methylation.[99] COMT has a low regioselectivity and the number of useful potential applications of COMT make engineering the regioselectivity of this methyltransferase desirable. Studies will focus on mammalian catechol-O-methyltransferase (COMT) (this study will use the rat COMT) as opposed to the plant methyltransferase caffeic-O-methyltransferase (CaOMT), which has a more limited substrate scope (although also displays *meta* regioselectivity).[100] This work will focus on the catechol-O-methyltransferase, as the well understood active site of COMT, will enable the use of directed mutations to alter the regioselectivity.[1,33]
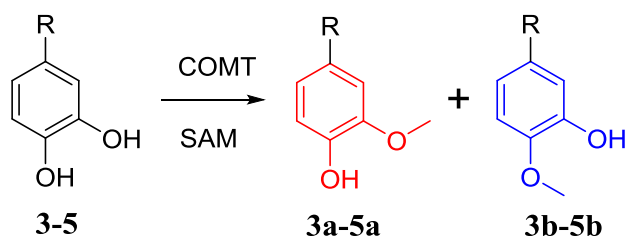
(**3**) R = CHO; (**4**) R=COOH; (**5**) R= NO$_2$

Scheme 13: Methylation by the COMT enzyme. The two regioisomeric products are shown; *meta* (red) and *para* (blue).

Table 3: Substrates, substrate structures and the regioselectivity as determined by Creveling *et al.*[98]

Substrate	Substrate Structure	*meta:para*
Dopamine		10.2
3,4-dihydroxybenzoic acid (DHBA) (**4**)		6.1
4-nitrocatechol (NO2CAT) (**5**)		2.2
3,4-dihydroxybenzaldehyde (DHBAL) (**3**)		2.1
4-methylcatechol		0.85

The low regioselectivity of COMT does lower its value as a biocatalyst. However, the COMT enzyme has been used in metabolic engineering studies to produce vanillin (**3a**). **3a** is a key flavouring, of which only 2% of the world's supply demands are produced by natural sources, thus most is provided by synthetic methods.[101] Increasing the number of vanillin (**3a)** production methods is therefore attractive, as currently **3a** is currently synthesised from guaiacol (*o*-methoxy-phenol), which is extracted from non-renewable petrochemical sources.[102] In order to overcome these issues Li and Frost developed a metabolic engineered process, which used the renewable feedstock glucose.[103] Li and Frost used the enzymes from the shikimate pathway to convert glucose to DHBA (**4**), which COMT methylated before a reductase was used to produce vanillin (**3a**) (Scheme 14).[103] However, this method produced isovanillic acid (**4b**) and isovanillin (**3b**) contaminants.[103] Only a 66% yield of **3a** from vanillic acid (**4a)** was recovered, with the remaining 33% being **3b** and **4b** contaminants.[103] A regioselective COMT would improve the yields of vanillin (**3a**) from the renewable glucose feedstock.
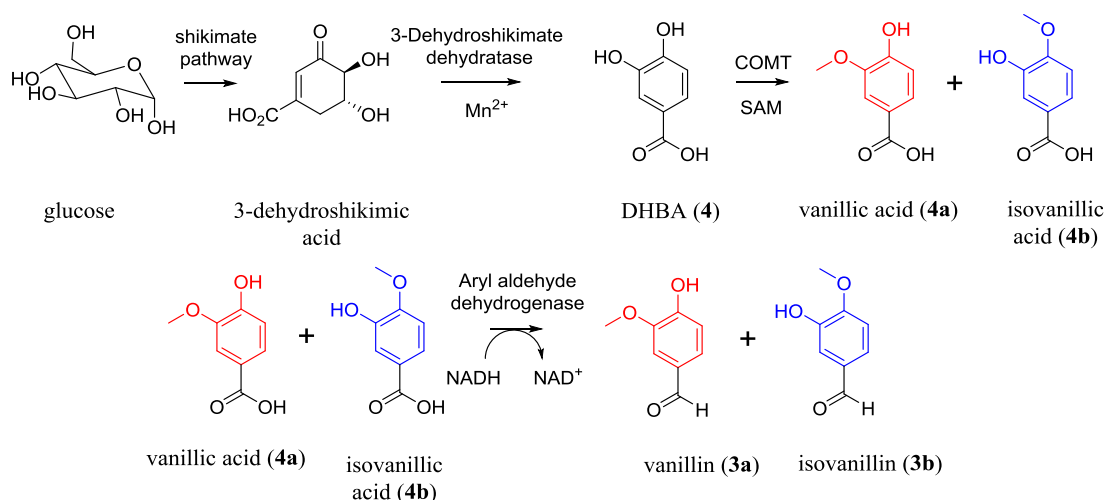
Scheme 14: Li and Frost's pathway for the production of vanillin (**3a**) from glucose *meta* products (red) and *para* (blue) contaminants are shown .[103]

The production of vanillin (**3a**) from renewable feedstocks such as glucose makes development of a regioselective COMT an attractive proposal. **3a** is not the only potential product of COMT, as COMT accepts most side chains. As Figure 10 shows, there are a number of important drug-like molecules that have the catechol motif meaning that COMT could be used to assemble active pharmaceutical ingredients (APIs). The APIs below show that both *meta* and *para* methylated APIs

are necessary. In order to produce *meta* and *para* methylated isomers with good yield a regioselective COMT would be necessary. Therefore, developing regioselective COMT could be used to not only produce *meta* regioisomers (like vanillin (**3a**)) but also important *para* APIs.

This section has shown that existing selectivity is present in methyltransferases, with GlmT and MppJ showing excellent enantioselectivity.[19,97] The enantioselectivity was only maintained with coupled enzymes such as TyrB with the MppJ methyltransferase.[19] Building on the enantioselectivity developed by a coupled system, further coupled systems with methyltransferases should be explored to determine whether more complex transformations are possible with greater selectivity. Whereas the COMT enzyme does not exhibit high regioselectivity, however, improving the regioselectivity of the enzyme could lead to valuable flavourings and APIs. Therefore, coupled systems with methyltransferases and the regioselectivity of the COMT enzyme should be explored.
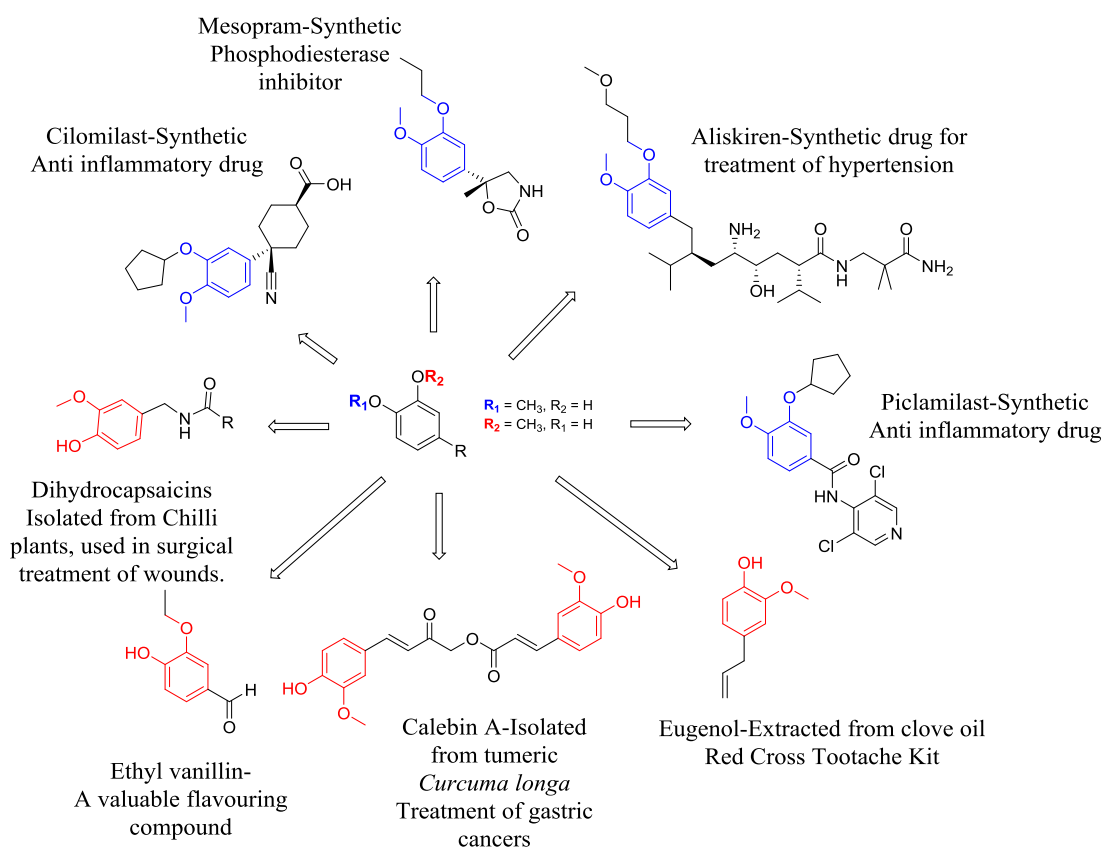
Figure 10: Regioisomeric products, which contain the monomethylated catechol motif. *Meta* (red) and *para* (blue) monomethylated catechols are shown.[102,104-109]

1.3.2   Project Aims

The previous sections have shown the versatility of methyltransferases with SAM analogues (Sections 1.2.2, 1.2.3 and 1.2.4). However, more work is needed to develop methyltransferases into biocatalytic alkyltransferases. Building on the structural studies of Section 1.1 and the examples of Sections 1.2.4 and 1.3.1, the aims for this study can be set out.

The first aim will be to engineer methyltransferase selectivity. The focus of this work will be to understand and develop the regioselectivity of COMT through both structural and mutagenic studies. The study will develop regioselectivity through understanding of the structure of COMT and applying mutagenesis to key residues. Development will aim at regiocomplementary COMT enzymes i.e. a *para* COMT mutant and a *meta* COMT mutant as both regioisomers have important APIs (Figure 10). For this work the DHBAL (**3**) substrate will be primarily used (as well as DHBA (**4**) and NO2CAT (**5**) substrates) (Scheme 15), as the aldehyde functional group offers an R group that can be tailored with a variety of reactions. Through developing **3** as model compound *meta* and *para* COMT methyltranferases will be developed.
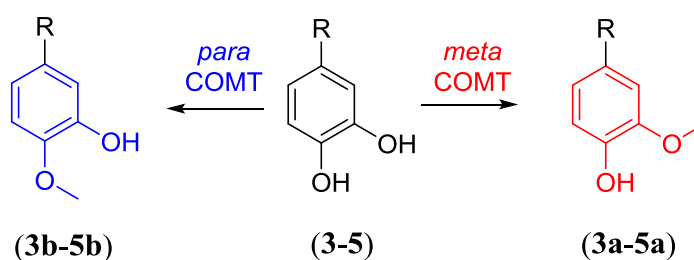
Scheme 15: Hypothetical *para* (blue) and *meta* (red) COMTs. R= -CHO (**3**), -COOH (**4**) and NO2CAT (**5**).

The improvement of COMT selectivity can also be coupled with other enzymes, which could improve the versatility of the COMT enzyme. As seen in Section 1.2.3, coupling COMT with the MAT enzyme will lead to alkylations as well as methylations (Scheme 16). Therefore, coupling the COMT enzyme with MAT enzyme will be used to lead to more alkylations of catechols. The COMT enzyme

could also be trialled with other enzymes to attempt pre or post COMT tailoring, which could lead to improve regioselectivity of the COMT enzyme, as observed for the MppJ and GlmT enzymes.[19,97] Therefore, COMT will be trialled with hMAT to improve versatility of COMT, with the aim to generate an alkyltransferase, and with another enzyme aimed at developing two transformations.

R'=H (**2**), R'=CH$_3$ (**2a**), R'=CH$_2$CH$_3$ (**2b**),
R'=CH$_2$CH=CH$_2$ (**2c**), R'=CH$_2$Ph (**2d**)

R'=H (**3**), R'=CH$_3$ (**3a**), R'=CH$_2$CH$_3$ (**3c**),
R'=CH$_2$CH=CH$_2$ (**3d**), R'=CH$_2$Ph (**3g**)
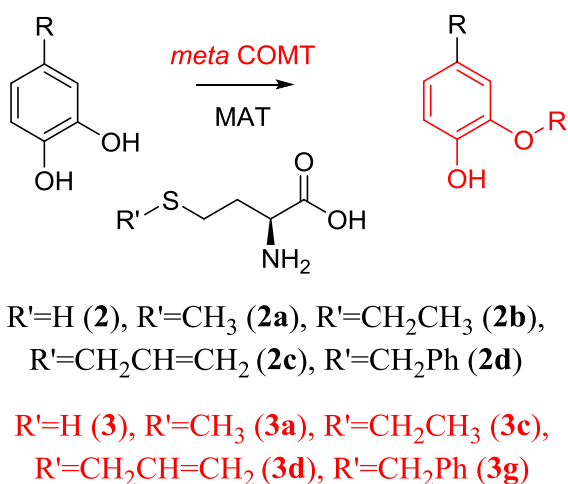
Scheme 16: Hypothetical regioselective *meta* COMT with MAT and methionine analogue.

The final aim of this work will be to apply the results of the COMT experiments to a methyltransferase with limited substrate scope. The results of the structural studies of COMT will aim to develop general structural information about methyltransferases. This information will then be applied to improve the biocatalytic properties of an alternative methyltransferase, such as the plant enzyme methyltransferase coclaurine-*N*-methyltransferase (CNMT). The study of CNMT, which has a limited substrate scope, may require a methyltransferase screen to quickly improve the enzyme's properties. Therefore, the final aim of this study will be to apply the results from COMT development to a methyltransferase (such as CNMT) with the aim of improving the understanding and biocatalytic properties of this methyltransferase.

Therefore, these three aims will be the target of this study with focus on the COMT enzyme and development of COMT as a regioselective alkyltransferase. Coupling systems with COMT will also aim to improve the versatility and selectivity of the enzyme. Finally these studies will be used to widen the substrate scope of another less studied methyltransferase. Accomplishing all of these aims will enable development of the under researched field of methyltransferase biocatalysis.

2 Understanding the effects of mutations and quaternary structure on COMT regioselectivity

2.1 Introduction

2.1.1 COMT active site residues for Mg^{2+} and SAM (1a) binding

The improvement of COMT regioselectivity in either the direction of *para* or *meta* methylation could lead to a number of important APIs and flavouring products (Figure 10, Section 1.3.1). In order to develop the regioselectivity of the enzyme; the residues that are involved in the cofactor SAM (**1a**) or Mg^{2+} ion binding were first necessary to identify, so that any mutagenesis strategy avoided interfering with these interactions. The **1a** amino acid residue contacts should not be mutated (when improving COMT regioselectivity) otherwise loss of **1a** binding and activity could occur. Multiple structural studies have uncovered a number of important residues in the COMT active site and from these studies more can be uncovered about what governs SAM and substrate binding.[1,32,33]

A number of structural studies have been conducted on the COMT enzyme most of which have focussed on the binding of inhibitors such as 3,5-dinitrocatechol (DNC) (Figure 11) and entacapone (as mentioned in Section 1.1.2).[1,30]However, these studies have also given a clear picture of SAM's (**1a**) binding to the COMT methyltransferase. Figure 11 shows that the **1a** binding residues in COMT can be divided into two main groups; adenine binding residues and methionine (**2a**) chain binding residues. The main adenosine binding residues (a multitude of other contacts have also been determined) were shown to be the W143 and E90 residues (found on β sheet 2, following the previously mentioned motif in Section 1.1.2).[1] W143 forms favourable edge to face van der Waals contacts with the adenine group of **1a**, whilst E90 forms strong hydrogen bonds with the hydroxyls of the ribose ring (a motif found across the methyltransferase class).[1,33]
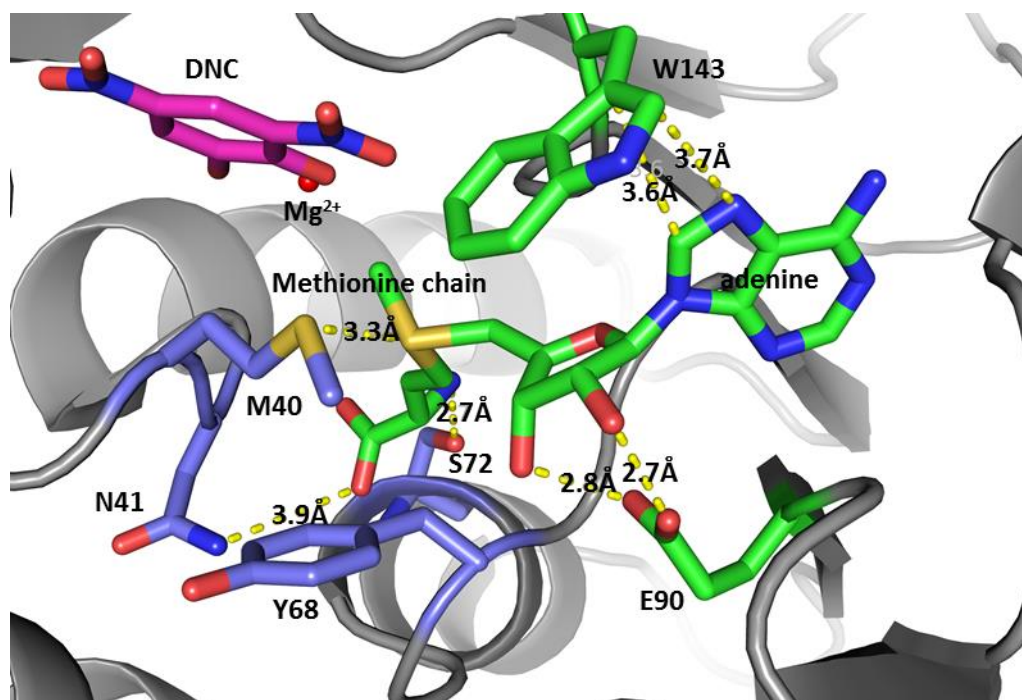
Figure 11: Adapted from PDB (1VID), SAM binding residues. Adenine binding residues shown in green, methionine (**2a**) binding residues shown in blue.[1,33]

The adenosine group has been shown to have two main contacts with the enzyme whereas there were found to be a number of residues involved in the binding of the methionine (**2a**) group of SAM (**1a**); Y68, M40, N41 and S72.[33]Y68 adopts an unfavourable conformation, to accommodate the binding of **1a** and upon binding forms van der Waals contacts with the SAM (**1a**) molecule.[33] M40 has been suggested as orientating the methyl group of **1a** through van der Waals contact with the sulphonium centre of **1a**.[1]N41 formed hydrogen bond contacts with the carboxyl group of the methionine (**2a**) chain whilst S72 forms a hydrogen bond contact with the amine of the methionine (**2a**) chain of SAM.[1,33] The magnesium ion forms an octahedral complex, when in the COMT active site, held by hydrogen bonds from the substrate (represented by the DNC inhibitor), a water molecule and the N170, D169 and D141 residues (Figure 12).[1,33,110]
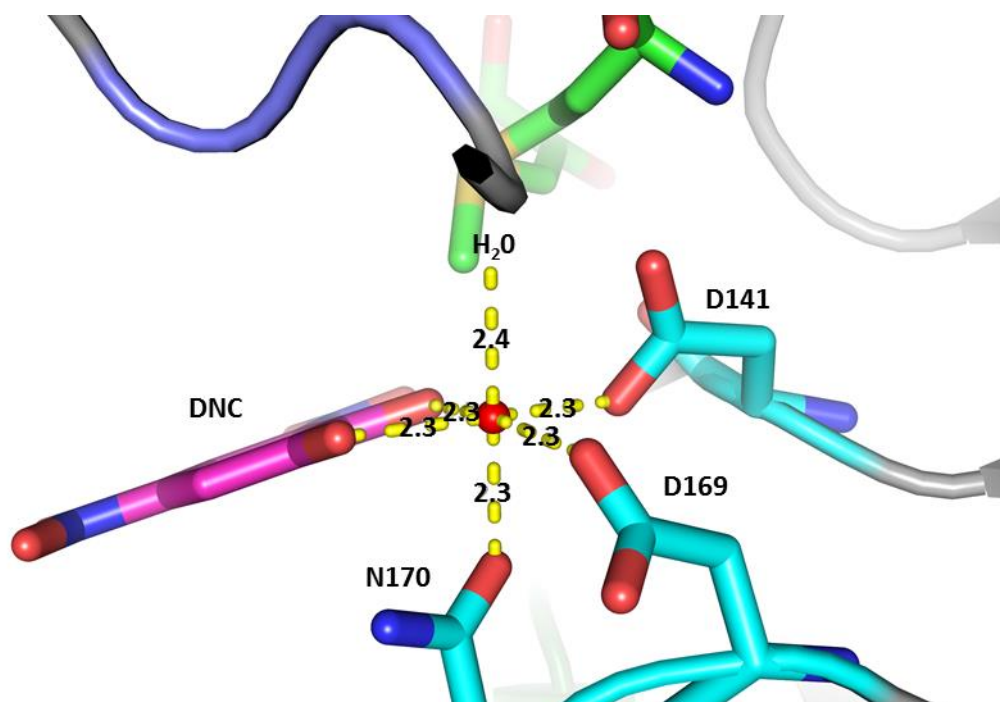
Figure 12: Adapted from PDB (1VID), octahedral binding of Mg^{2+} to DNC, N170, D169, D141 and a water molecule.[1]

Understanding of the SAM (**1a**) and Mg^{2+} binding has shown which residues should be avoided in mutagenesis strategies for improving regioselectivity. From this knowledge the literature was also examined for further information into the substrate binding of COMT. Understanding what residues dictate substrate binding and influence COMT regioselectivity would provide the detail necessary to improve the regioselectivity of the enzyme.

2.1.2 COMT regioselectivity

The residues that are involved in SAM (**1a**) and Mg^{2+} binding have been shown in the previous section (Section 2.1.1). The Vidgren *et al* and Rutherford *et al* studies have also examined substrate binding through monitoring inhibitor interactions with COMT.[1,33] These studies have suggested that COMT's regiocontrol is results from a "hydrophobic wall".[1,99] Examining the residues that make up this "hydrophobic wall" in the COMT active site will lead to a clearer mutagenesis strategy for targeting improvement of COMT regioselectivity.
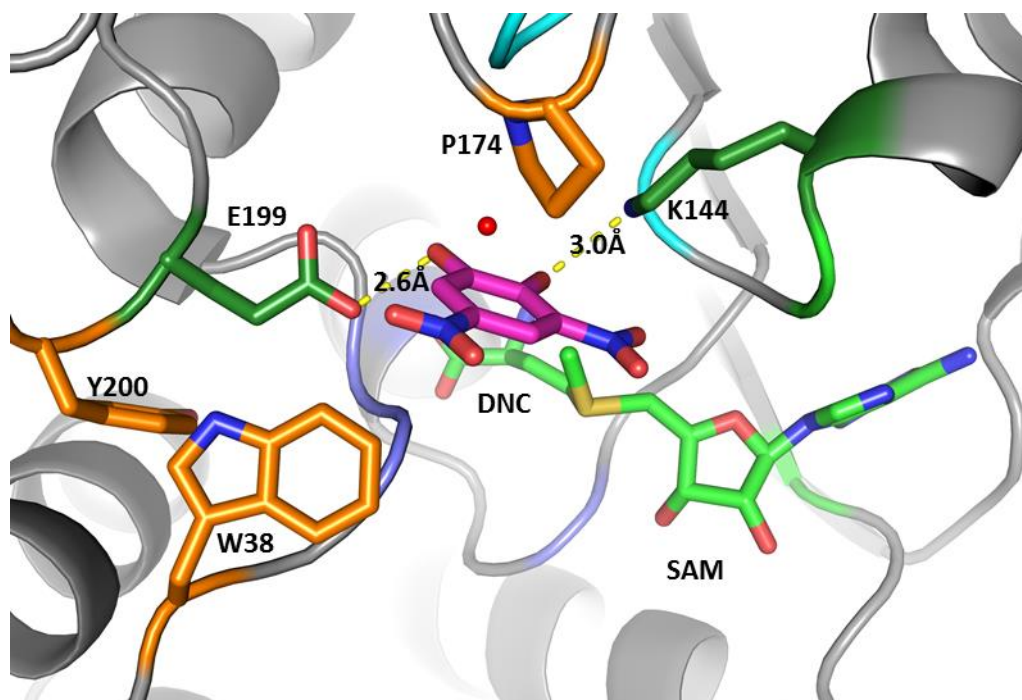
Figure 13: Adapted from PDB (1VID), key residues involved in the substrate binding, regioselectivity and methylation of catechols in COMT.[1]

The structural studies of COMT had shown great detail about the binding of both SAM (**1a**) and the Mg^{2+} ion.[1,33] The inhibitors used were the same in both Vidgren and Rutherford studies, and both found that the key residues involved in substrate binding were; Y200, W38, E199, K144 and P174 (Figure 13).[1,33] W38 and P174 were suggested as the so-called "COMT gatekeeper residues", which were thought to orientate the structure through van der Waals interactions.[1,111,112] A hydrophobic pocket formed from Y200 and W38 was thought to accommodate the side chain of the catechol (for example the amine in dopamine) in the *para* methylating orientation of COMT.[5,99]

Experimental data has shown that catechols with polar side chains have a significantly higher *meta*:*para* products ratio than catechols with nonpolar side chains.[98] The higher *meta*:*para* ratios for polar catechols could be explained if the *para* binding orientation located the side chain of catecholic substrates in the hydrophobic pocket formed by Y200 and W38, as polar groups would be repulsed from this pocket leading to less *para* methylation.[5] Conversely, non-polar side groups (such as 4-methylcatechol) would have a lower *meta*:*para* ratios, as the hydrophobic groups can form hydrophobic contacts with W38 and Y200 residues.[5] More evidence for this hypothesis was provided by Zhang *et al*, who showed that

mutation of the W38 residue (W38A and W38V), could increase in *para* methylation (a decrease of *meta* methylation by 40% r.e. for W38A and a dopamine substrate) suggesting that the W38 was important in controlling the COMT enzyme's regioselectivity.[112] The suggestion of the W38 and Y200 residues were involved in *para* methylation makes these residues interesting targets for the mutagenesis strategy to improve the regioselectivitiy of COMT.[5] Whereas P174 would not be targeted, despite positioning the catecholic substrates, as the P174 could be structurally important involved in a β turn where P174 residue was located.[1]
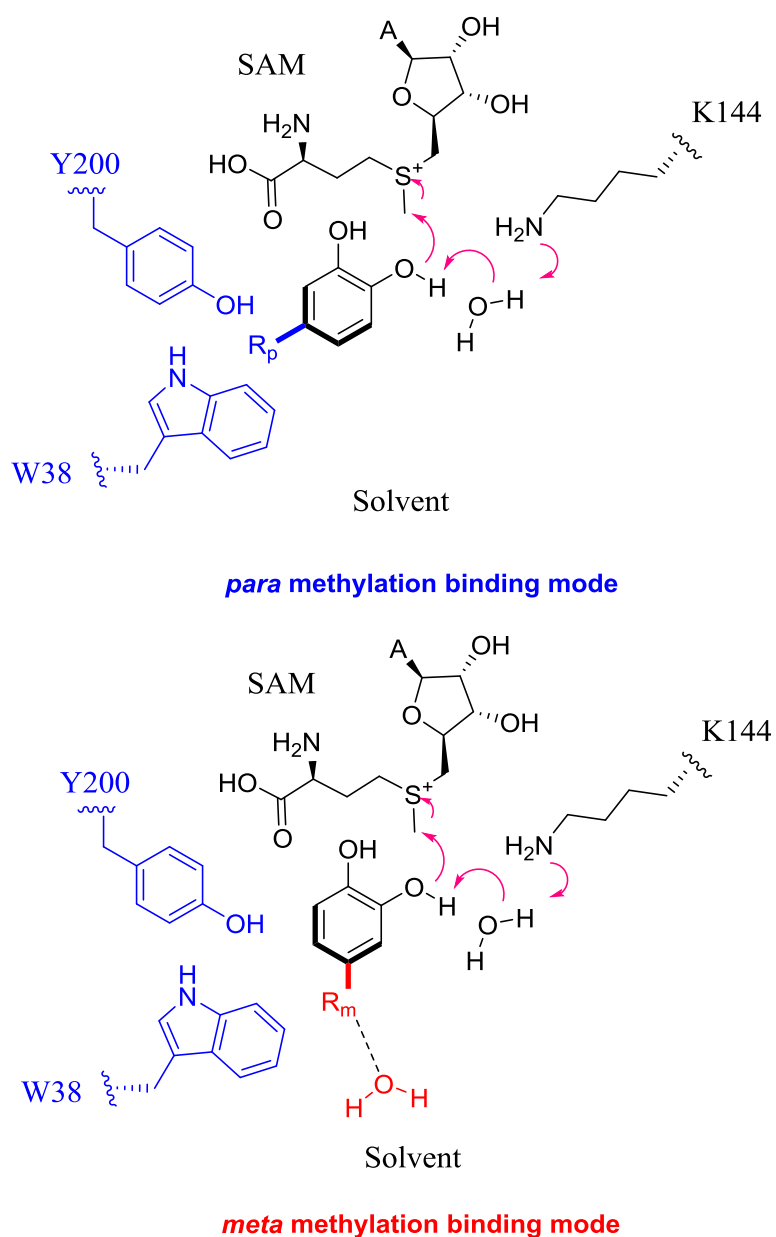
Figure 14: Suggested *para* (blue, top) and *meta* (red, bottom) binding modes for catechol substrate.[5]

The K144 and E199 residues were found to directly interact with the hydroxyls and were noted to be located at C termina of the β sheets (previously noted as important identification points in Section 1.1.2) (Figure 13).[1] K144 has been suggested as the catalytic base for COMT.[1,99] K144 acts as the catalytic base by either directly deprotonating the catechol to form to ionise the hydroxyl, which is then methylated, or the deprotonation occurs through a water molecule, with the K144 deprotonating the water molecule first and the deprotonated molecule in turn deprotonating the hydroxyl.[1] E199 was also postulated to interact with the alternative hydroxyl, through stabilising the adjacent hydroxyl (to the hydroxyl nearest K144) through a hydrogen bond, unlike the K144 residue.[1] Therefore, mutating K144 and E199 may indicate whether these residues were involved in the functions attributed to them.

The literature has shown the key residues involved in; SAM (**1a**) binding, Mg^{2+} binding, substrate coordination and regioselectivity.[1] Identification of these residues meant that a mutagenesis strategy could be drawn up aimed at targeting specific residues to improve the COMT enzyme either for *meta* or *para* methylation. From these studies, mutagenesis to develop COMT's regioselectivity was suggested to plausible and could lead to improved ratios of both *para* and *meta* regioisomeric products of catecholic substrates.

2.2   Results and discussion

2.2.1   COMT regioselectivity-mutagenesis strategy

The value of regioselective methylation of catechols has been clearly demonstrated in Section 1.3.1. In order to develop regioselective COMT enzymes, a library of COMT mutants was created with Dr Brian Law, based on literature studies and structural studies shown in Sections 2.1.1 and 2.1.2.[1,33,112] The primary aim of this library was to develop regiocomplementary mutants i.e. a mutant that solely produced *para* methylation and a mutant that solely produced *meta* methylation. The library of COMT mutants was based around the following positions of the COMT active site; W38, K144 and Y200. The positions were chosen, as Figure 15 shows, due to their interactions with the substrate (as shown in Sections 2.1.1 and 2.1.2).

Therefore, mutating these residues a shift in either *para* or *meta* methylation was expected.
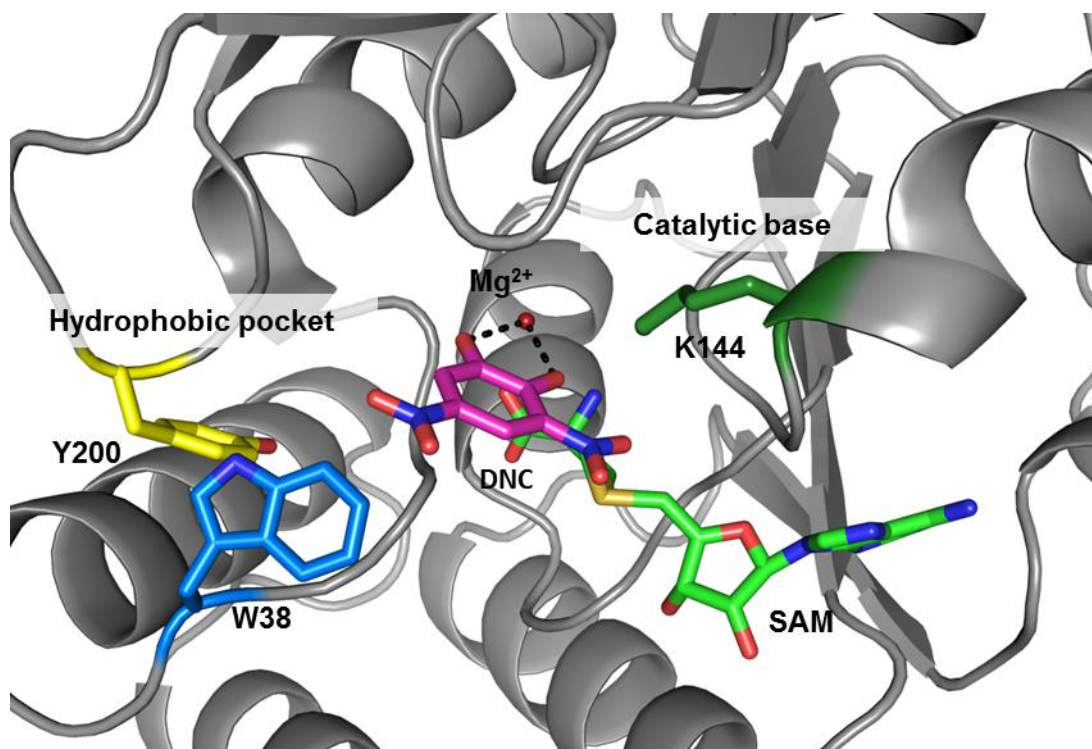
Figure 15: Key residues highlighted in the mutagenesis strategy for COMT. Key residues are; W38 (blue), Y200 (yellow) and K144 (green) from the structure of Vidgren *et al* PDB (1VID).[1]

The positions were decided, but a mutation strategy was now needed. K144 was mutated, initially to an alanine, to determine whether this would knock out activity. The effects of mutating Y200 and W38 were more difficult to determine, as to how the pocket interacts with the side chain had not been extensively studied. Therefore, in the case of Y200 and W38 mutations were chosen to account for different polarities and different sizes of amino acids. The polar mutations that were chosen were; Y200S and W38D. Whilst the mutations based on size were; Y200A, Y200L, Y200W, W38L, W38Y and W38F. The mutagenesis strategy was chosen to target these mutations and primers were then used for site directed saturation mutagenesis (as described in Section 6.2.8) with the pET21b WT COMT (from *Rattus norvegicus*) construct kindly provided by Professor Nigel Scrutton.
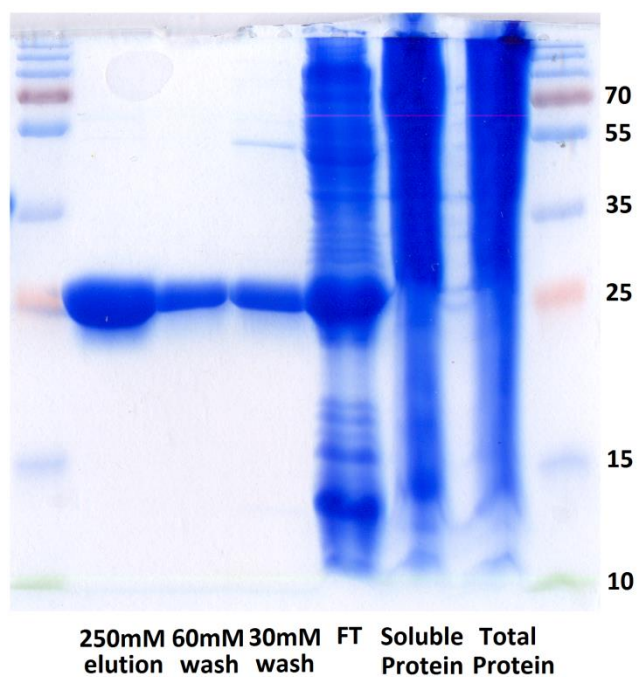
Figure 16: SDS PAGE gel of WT COMT, size consistent with COMT's mass of 25 kDa. (Right to left) Total protein, soluble protein, flow through (FT), 30 mM imidazole wash, 60 mM imidazole wash, 250 mM imidazole elution are shown.

Once the mutants were expressed and purified by the methods laid out in Section 6.2.11 (successful isolation of COMT was determined by SDS PAGE gel (see Figure 16)). COMT was then assayed at 37°C, chosen as COMT was isolated from mammalian systems, which have been shown to be held at 37°C. Initial assay tests were conducted with the DHBAL (**3**) substrate. **3** was chosen, as the *meta* methylated regioisomer (**3a**) (vanillin) has been used as a flavouring, whilst the *para* methylated isomer (isovanillin (**3b**)) has been used as an important API for drugs such as aliskiren.[108] 20 mM Phosphate buffer was used to keep assay conditions at pH 7.4. MgCl$_2$ was added, as all literature studies had shown that the presence of Mg^{2+} was essential for COMT activity (see Section 2.1.1). Dithiothreitol (DTT) was also present, as some COMT studies had suggested that COMT dimerization was possible, and that prevention was necessary.[5]

Using the previously mentioned conditions WT COMT and the mutants were assayed with the DHBAL (**3**) substrate (concentration of **3** was 0.5 mM) for 1 hour before termination with formic acid (final concentration 0.1%). Formic acid was used for termination after being determined to deactivate the enzyme, which was tested through termination assays. The assays were then run on a RP HPLC with the
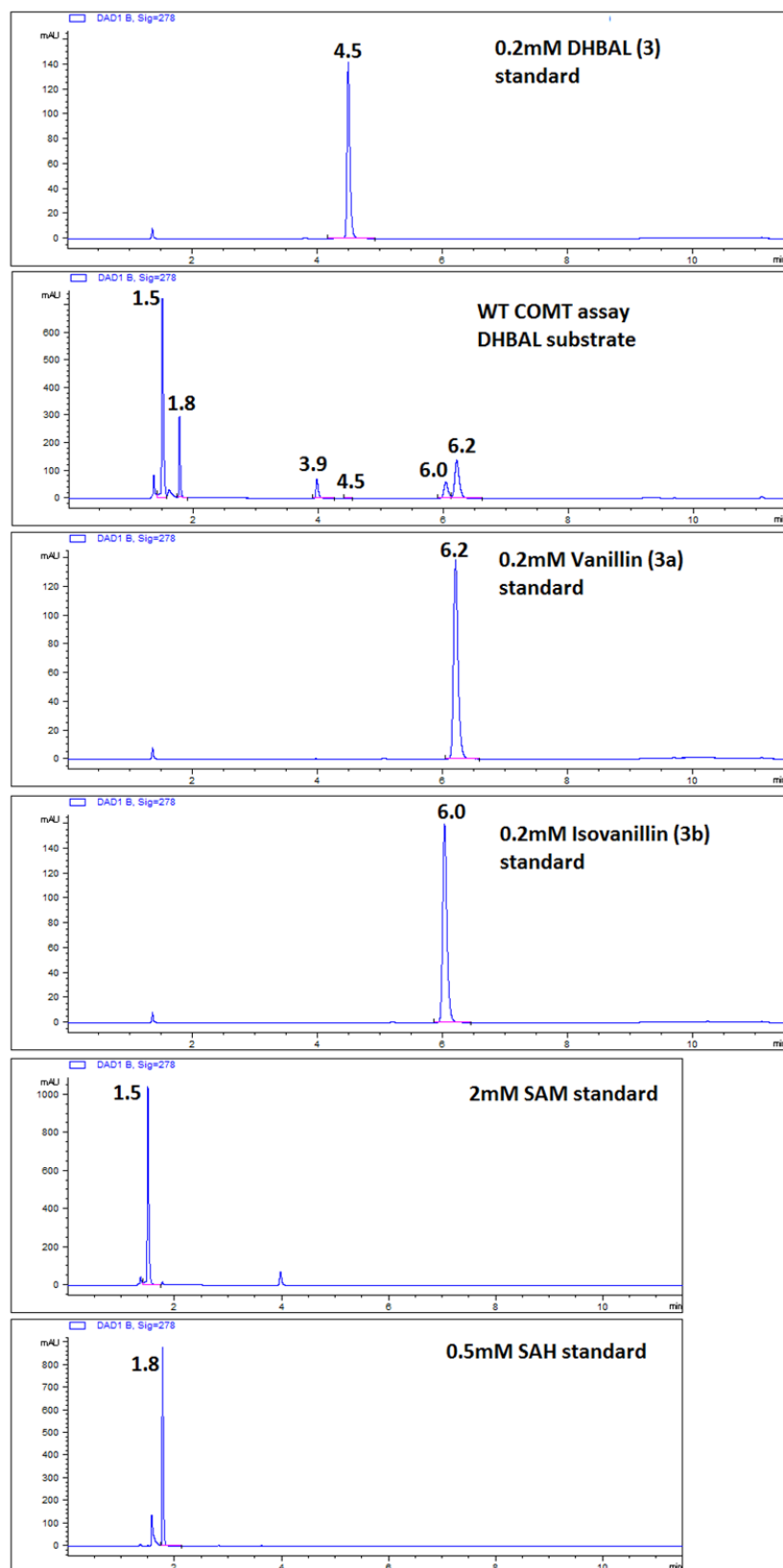
Figure 17: RP HPLC traces for; 0.2 mM DHBAL (**3**) standard, WT COMT **3** substrate assay, 0.2 mM vanillin (**3a**) standard, 0.2 mM isovanillin (**3b**) standard, 2 mM SAM (**1a**) standard and 0.5 mM SAH (**1**) standard using the RP HPLC method shown in Section 6.5.1. The peak at R_t was determined to be 5′-Deoxy-5′-(methylthio)adenosine see Section 7.1.3, Figure 106.

methods shown in Sections 6.5.1 and 6.5.2. Firstly, standards were run on the RP HPLC to determine retention timesfor SAM (**1a**), SAH (**1**), DHBAL (**3**), vanillin (**3a**) and isovanillin (**3b**). Then assays were run with the WT COMT protein initially and compared to the retention times of the two substrates (DHBAL and SAM) and three products (vanillin, isovanillin and SAH). As Figure 17 shows the peaks present in the assay were consistent with the enzyme converting DHBAL to vanillin and isovanillin products (the controls (**-1a** and –COMT) also showed no product peaks).

Once the COMT was confirmed to be active, the mutant and WT COMTs could be analysed for their regioselectivity. Analysing the data from the HPLC, peak areas were determined and concentration of products (vanillin (**3a**) and isovanillin (**3b**)) determined using calibrations (see Section 7.1.4) (that were conducted by Dr Brian Law). Regioisomeric excesses were then calculated (using the equation shown below) for the WT COMT and mutants (Table 4).

$$R.e\ (\%) = Normalised\ meta\ product\ (\%)\ - Normalised\ para\ product\ (\%)$$

Equation 1: Equation for the determination of regioisomeric excess (r.e).

Table 4: Regioisomeric excesses of WT COMT and mutants with DHBAL (**3**) substrate. Residues are colour coded in accordance with the crystal structure in Figure 15 (RP HPLC chromatograms can be seen in Section 7.1.3, Figures; Figure 115, Figure 116, Figure 117, Figure 118, Figure 119, Figure 120 and Figure 121).[1]

Mutant	Conversion to products/ %	Normalised *meta* (3a)/%	Normalised *para* (3b)/%	Regioisomeric excess (r.e) /%
WT COMT	100	69	31	38
Y200L	87	95	5	90
Y200W	68	95	5	90
Y200A	84	94	6	88
Y200S	82	93	8	85
W38D	51	96	4	92

W38Y	85	93	7	86
W38F	93	82	18	64.4
K144A	36	60	40	20

The results of the regioselectivity assays for the COMT mutants with the DHBAL (**3**) substrate showed a range of regioselectivities. Mutations at both W38 and Y200 were mostly *meta* regioselective (producing more vanillin (**3a**) than isovanillin (**3b**)) with W38F showing the lowest *meta* regioselectivity out of the mutants (higher still than WT) and Y200L showing the highest. The increase in *meta* regioselectivity of all the W38 and Y200 mutants suggested that any change in the hydrophobic pocket disturbs *para* regioselectivity. The K144A mutation, however, showed an increase *para* regioselectivity. The change from lysine to alanine, suggested that the basic nature of the lysine may control some of COMT's regioselectivity.

The initial data, collected with Dr Brian Law, showed that the regioselectivity of the COMT enzyme could be altered by a directed mutagenesis strategy. However, the mutations had shown a range of regiochemistries, which from these assays did not explain what caused this change in regioselectivity. Therefore, further work was needed to understand the underlying causes of the regiochemistry in COMT. For this work two positions were targeted, due to their opposite regioselectivity; K144 and Y200.

2.2.2 COMT regioselectivity-explaining the observed regioselectivity

The previous section has shown that mutagenesis of the COMT enzyme could fundamentally alter the regioselectivity of the enzyme. The change in regioisomeric excess had identified two key positions Y200 and K144. These positions were the most interesting; as mutations at Y200 led to the most *meta* regiochemistry being displayed, whilst mutations at the K144 position led to the most *para* regiochemistry being displayed. Therefore, K144A and Y200L were further investigated to understand how the mutations had altered the regiochemistry of the enzyme.

The K144A COMT mutant had shown the most *para* regioselectivity of all the COMT mutants. Previous studies, had suggested that the K144A mutation was the catalytic base, which deprotonates COMT substrates, prior to their methylation.[1,33] The crystal structure produced by Vidgren *et al*, had also shown that K144A was only 3.1 Å away from the *meta* hydroxyl of DHBAL (**3**) (overlaid over the DNC inhibitor) (Figure 18 Right).[1] The small distance between K144 and the DHBAL (**3**) *meta* hydroxyl, suggested that K144 was interacting with the *meta* hydroxyl. The distance also corresponds to a study by Shin *et al* who determined a catalytic serine in a malonamidase also showing a distance of 3.1Å.[113] If K144 was the catalytic base then further mutations and studies could determine this. Studies were focused on the activities and kinetics of the mutants, which aimed at determining the K144 residue's interaction with the substrate.
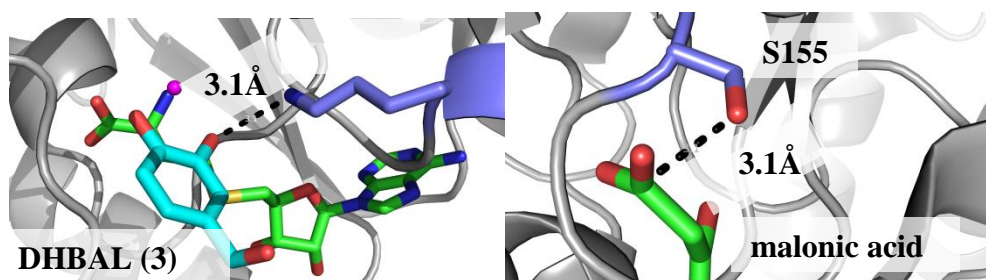
Figure 18: (Left) The distance of K144 (purple), the supposed catalytic base, and DHBAL (**3**) hydroxyl (cyan) from Vidgren *et al* PDB (1VID).[1] (Right) The malonamidase catalytic base serine (purple) and representative substrate malonic acid from Shin *et al* PDB (1OCL) (green).[113]

Activity assays were initially conducted on the K144A COMT mutant to determine the activity compared to WT COMT. The assay was conducted over 1 hour with 0.5 mM DHBAL (**3**) substrate, as WT COMT (5 µM) has been shown to fully methylate **3** at this concentration in this time. The aim of this timing was, therefore, to uncover the maximum regioselectivity and activity of K144A. The concentration of the SAM cofactor (**1a**) was 1.5 mM, in order to keep the cofactor in excess, and ensuring that the activity of the enzyme was the limiting factor (further details of the assay set up can be found in the Section 6.4.1). Finally the assay was conducted in triplicate, in order to uncover any errors in experimental design.

	WTCOMT	K144A
▇ para	29	25
▇ meta	62	37
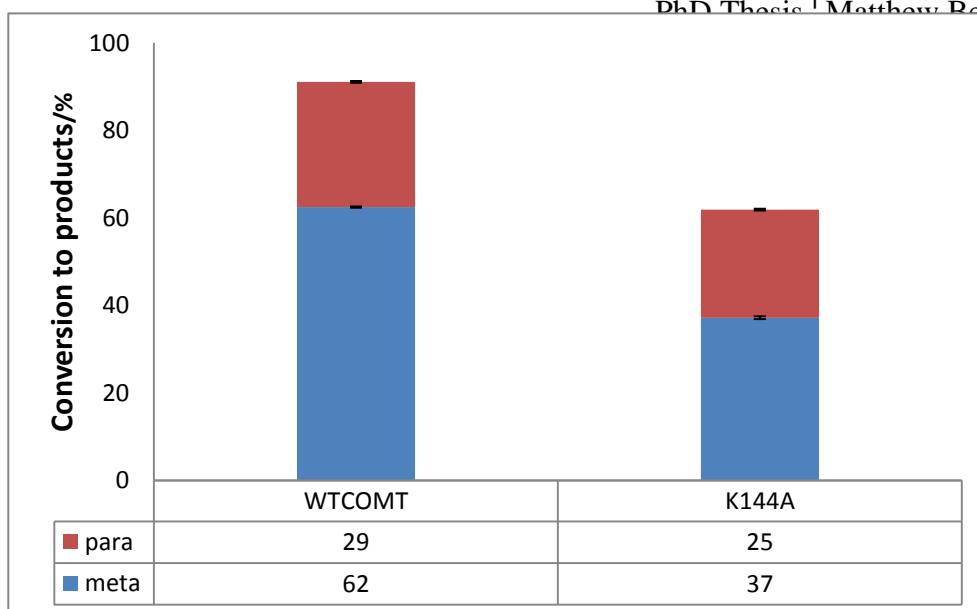
Figure 19: The conversion assays of K144A and WT COMT with DHBAL (**3**) substrate.

Results of the conversion assay (Figure 19) firstly confirmed those shown in the previous section (Section 2.2.1) that K144A leads to a lower regioisomeric excess, which suggested that less of the *meta* regioisomeric product (vanillin) (**3a**) was produced (observed regioisomeric excess in both assays was 20%). The assay also showed a reduction in conversion with the K144A mutant (62%) compared to the WT (91%). However, the removal of the lysine, did not abolish activity suggesting that the K144 residue was not essential for activity.

The change of lysine to alanine did not abolish activity, despite reducing the activity of the COMT enzyme. Analysing the results, the main reduction of activity was due to the lower *meta* methylation. Conversion assays of K144A with the DHBAL substrate (**3**) demonstrated a consistency in *para* methylation with a percentage of 25% opposed to 29% of WT COMT. Whereas *meta* methylation shows a conversion of 37% for K144A opposed to 63% for WT. The lower *meta* methylation suggested that the lysine residue was more important for *meta* methylation than *para* methylation. The *meta* hydroxyl was thought to have a higher pKa than the *meta* hydroxyl due to resonance effects (this can also be observed through theoretical pKa calculators which put the pKa of the *meta* hydroxyl of **3** at 9.1 and the *para* hydroxyl at 7.5). The resonance effects in **3** (Scheme 17) were thought to not stabilise the *meta* anion onto the aldehyde of **3** unlike the *para* anion. Therefore, the *meta* position was thought to have higher pKa, which suggested that the lysine was required to deprotonate the *meta* position, whereas the *para* hydroxyl could be deprotonated at

pH 7.4 by the solvent. The conversion assay results showed that the K144A mutant has lower activity and lower regioselectivity, which could be due to the loss of *meta* deprotonation.
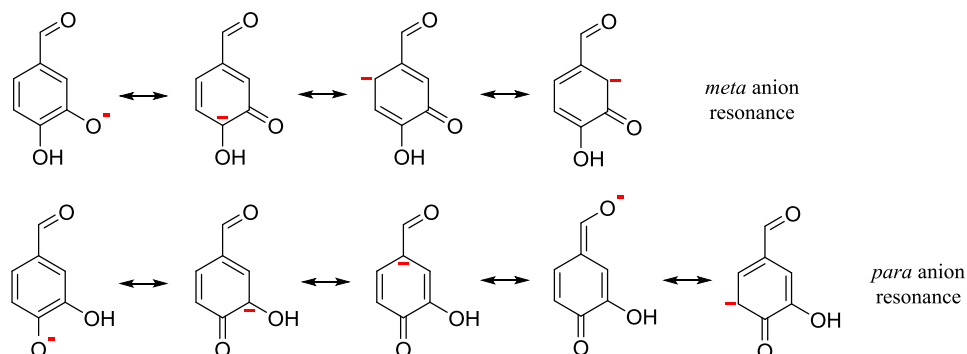
Scheme 17: Resonance stabilisation of *meta* and *para* phenolic anions of DHBAL (**3**) (shown in red).

The conversion assays had suggested that *meta* methylation was reduced through the K144A mutation. In order to understand the loss of activity further the kinetics of the K144A mutant were explored. The conversion assay had shown K144A to be approximately 40% less active than WT COMT. K144A kinetic assays were run with a concentration range of 10-2000 µM with the DHBAL (**3**) substrate. A concentration of 0.5 µM K144A COMT mutant, to prevent full conversion of the lower concentrations of **3**. Time points were measured (2, 3, 4 minutes) in order to determine the initial rates of methylation. Further information into the experimental design of the kinetic K144A studies can be seen in Section 6.4.1.

The kinetic experiments showed a good fit with Michaelis-Menten kinetics. Using the SigmaPlot program, the V_{max} ($8.5 \times 10^{-5} \pm 0.2 \times 10^{-5}$ µmol min^{-1}) and K_m (108 ± 12 µM) was calculated. From the V_{max} a k_{cat} of 1.1 ± 0.1 min^{-1} was calculated. The kinetic experiments did show errors for most concentrations, particularly towards the higher concentrations were saturation was occurring. However, the overall kinetic trend, the K_m and V_{max} could still be calculated.
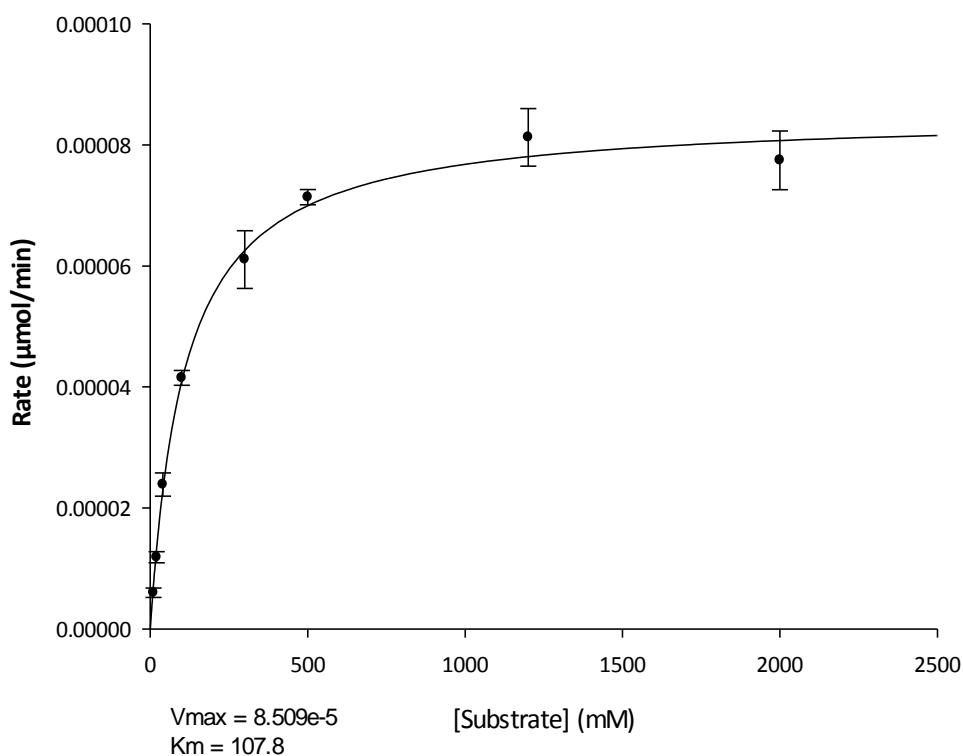
Figure 20: K144A Michaelis-Menten graph.

The kinetic experiments did show the K_m and k_{cat} for the K144A albeit with some errors. The K144A kinetic experiments had shown the binding constants for the mutant, which allowed for a further understanding of how the K144A mutant effects regioselectivity. The K144A COMT mutant (K_m of 108 µM) showed weaker binding compared to the WT COMT (K_m of 3.4 µM) determined by Dr Brian Law. The K144A mutant COMT also showed lower activity displaying a k_{cat} of 1.1 min^{-1} compared to the WT COMT k_{cat} of 4.0 min^{-1} determined by Dr Brian Law. The catalytic efficiency, represented by k_{cat}/K_m, was also significantly lower than the WT; 0.01 min^{-1}µM^{-1} (K144A) compared to 1.2 min^{-1}µM^{-1} (WT COMT). K144A exhibited much lower catalytic efficiency than WT COMT suggested that the K144 was a key catalytic residue. The kinetic results had shown, like the preceding conversion assays, that K144 was an important residue for catalysing the methylation of DHBAL (**3**).

Activity assays and kinetic experiments had suggested that K144 was an important residue, possibly involved primarily in *meta* methylation. In order to provide further

evidence for this result, further mutations were conducted on this residue. The mutations that were selected were K144R and K144E. K144R was selected because arginine can also act as a base in proteins. The K144R mutation was expected to yield WT regioselectivity, with lower activity due to possible steric clashes, from the larger arginine with the DHBAL (**3**) substrate (Figure 21). However, modelling studies using Pymol showed that the K144R may be closer than the lysine, in the WT structure, which suggested that *meta* deprotonation may be faster (2.9Å for the mutant compared to 3.1Å from WT). K144E was expected to push even further towards *para* methylation, as the acidic residue would respulse the hydroxyl rather than deprotonate. However, in modelling with Pymol the K144E residue was 5.8Å (Figure 21), due to the residue's much shorter side chain (Figure 21) so repulsion may not occur. The increased protonation of **3**'s *meta* phenoxide, would lead to less *meta* methylation. Protonation may also occur on the *para* phenoxide, which would lead to a reduced activity.
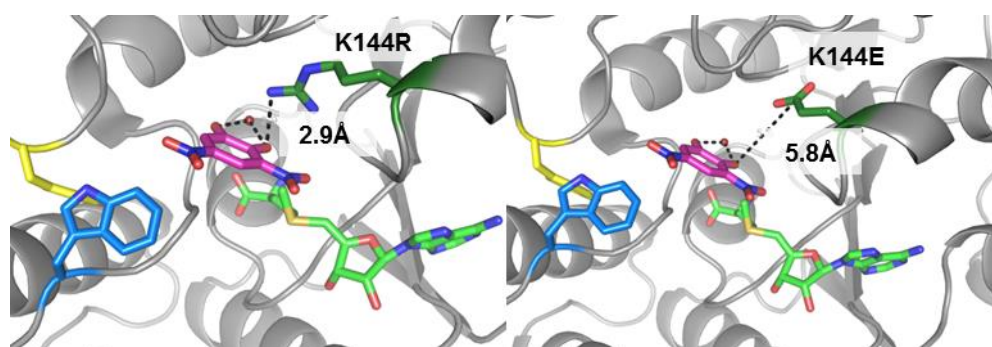
Figure 21: The K144R (Left) and K144E (Right) mutants modelled on the Vidgren *et al* structure PDB (1VID).[1]

K144R and K144E assays were conducted, with the same method as K144A and WT COMT with DHBAL (**3**) substrate. As Table 5 shows, K144E exhibited significantly more *para* than *meta* methylation (the most *para* mutant in this study) whereas K144R displayed WT regioselectivity. These results fitted with the hypotheses previously mentioned. The K144R mutant exhibits WT regioselectivity because the arginine residue can deprotonate the *meta* hydroxyl, like lysine. The K144R mutant displayed lower conversion than WT COMT (Table 5). The lower conversion may be explained by the steric bulk of the arginine mutation leading to steric clashes in the COMT active site, which could slow down the methylation. Unlike K144R, the K144E can either protonate the *meta* phenoxide or repulse the phenoxide, leading to

much less *meta* methylation, explaining the *para* regioselectivity observed. K144E displayed the lowest conversion of the K144 mutants and was most likely attributable to the significant loss of *meta* methylation (Table 5). The K144R and K144E mutations seem to confirm that K144 was the catalytic base and due to the reduction of *meta* methylation, mostly deprotonates the *meta* hydroxyl.

Table 5: Regioselectivities of K144 mutants and WT COMT with the DHBAL (**3**) substrate.

Mutant	Conversion/ %	Regioisomeric excess/%
WT	91	37
K144A	62	20
K144E	29	-29 (*para*)
K144R	35	37

The K144 mutations had shown a decrease in *meta* regioselectivity whereas the Y200 mutations had shown exactly the opposite. Y200L displayed the greatest regioselectivity with a regioisomeric excess of 90% being recorded (Table 5), which was the highest regioselectivity of any of the COMT mutants shown in Section 2.2.1. The high regioselectivity of the Y200L meant that this mutant was an excellent candidate for crystallography trials. The crystal structure would facilitate an understanding of the amino acid interactions in the Y200L COMT active site that lead to the regioselectivity increase.

A crystal structure of Y200L with the DHBAL (**3**) had previously been generated in earlier work Dr Colin Levy and Dr Brian Law; however, the poor binding of **3** meant that the electron density of the structure was uncertain, thus leaving room for further improvement. The purification procedure had also pooled fractions from the gel filtration column, when two peaks had been noted. In order to improve upon the uncertain Y200L structure, the inhibitor DNC and SAM (**1a**) were soaked with Y200L, as DNC (nM binder) and **1a** (μM binder) exhibit better binding than the

SAH (**1**) (μM binder) and DHBAL (**3**) (μM binder).[114] The better binding of DNC and **1a** should give a clearer electron density in any crystal structures obtained.[114]

The purification procedure involved a nickel column, followed by anion exchange and gel filtration chromatography (Section 6.2.12). The extra purification steps were taken to ensure a clean sample of Y200L. The final purification step led to two peaks, which were both separated and trialled for crystallography with 2, 5-dinitrocatechol and SAM (**1a**). The first peak isolated from the FPLC gel filtration step (Section 6.2.12) proved to crystallise the best. The purification was scaled up and the samples were submitted, after soaking with **1a** and DNC, to Dr Colin Levy for crystallography trials.

Y200L was successfully crystallised, yielding a structure of 1.63 Å resolution. The resolution was significantly improved from the previous structure (2 Å resolution) and the ambiguous electron density was cleared up. Y200L was crystallised as a dimer suggesting that the initial elution peak from gel filtration was dimeric Y200L. Figure 22 shows overlaid crystal structures of dimeric Y200L COMT and the Vidgren WT COMT structure.[1] The key difference between residues in the two crystal structures was the E199 residue (Figure 22). The E199 residue's acidic side chain faces out into the solvent in the Y200L mutant structure, whereas in the Vidgren structure the E199 residue was found facing into the active site.[1] The different orientation of the E199 residue in Y200L may cause the shift towards *meta* regioselectivity. E199 may stabilise *para* methylation through a hydrogen bond between E199 and the 4 position of DHBAL (**3**), as shown in Figure 22, and the change in orientation may destabilise *para* methylation, as a result.

The crystal structure of Y200L had suggested that the E199 residue's orientation may control the regioselectivity of COMT. The K144 residue was also shown to be very important in *meta* methylation, reflected in both the kinetic and regioselectivity of the K144 mutants. The understanding shown by these mutations, have shown that regiocomplementary mutations can be conducted on COMT. Controlling the regioselectivity of the enzyme in this manner will increase the attractiveness of COMT, as a biocatalyst, and second generation mutations could build on these initial studies to develop both *para* and *meta* selective COMTs.
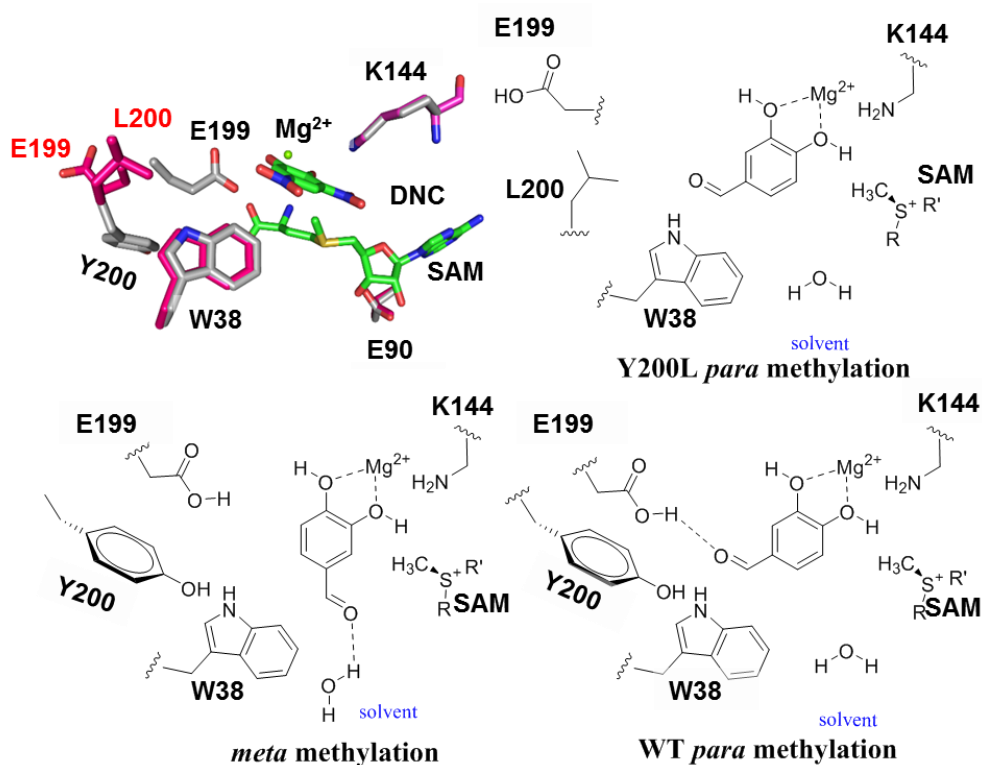
Figure 22: (Top left) Overlaid crystal structure of Y200L and WT COMT, adapted from PDB (1VID).[1] (Top Right and Bottom) Suggested methylation orientations for Y200L and WT COMT, *para* destabilisation is indicated in the top right.

2.2.3 Separation and regioselectivity of oligomeric WT COMT

The previous sections showed that the Y200L COMT mutant had excellent regioselectivity in the direction of *meta* methylation with the DHBAL (**3**) substrate. The crystal structure was also elucidated and the structure suggested that E199's absence from the active site in the Y200L was the reason for this *meta* regioselectivity. During the purification process for the Y200L mutant two peaks were noted on the gel filtration chromatogram. The first peak was used in the crystallisation process. The structure collected showed, in good agreement with the literature, that this peak was dimeric Y200L and that these two peaks from gel filtration were suggested to be oligomeric forms of the COMT enzyme.[115] This

section will examine these oligomers aiming at determining whether dimerization affects the regioselectivity of the COMT enzyme.

The gel filtration chromatogram (see below) showed two peaks, which were also observed in Y200L purification for the crystal structure. The identity of the two peaks was suggested to be monomeric and dimeric structural forms of the COMT enzyme based on the previous literature.[32] The Y200L structure had also shown that this first peak was dimeric Y200L. In order to provide further evidence that these two peaks were the COMT monomer and dimer, the FPLC size exclusion column was calibrated with standard proteins. The Y200L structure was purified on a gel filtration column designed for proteins between 10 kDa and 75 kDa. Based on this a calibration kit was ordered from Sigma Aldrich and the protein's chosen for the calibration were; Cytochrome C from horse heart, Carbonic Anhydrase from Bovine Erythrocytes and Albumin from Bovine Serum. Initially Blue Dextran was run to determine the column's void volume (v_0), which was determined to be 9 minutes. The elution times of the proteins were then recorded (v_e). From this a graph was plotted with the x axis displayed the size of the proteins whilst the y axis displayed the (v_e/v_0) (Figure 23).
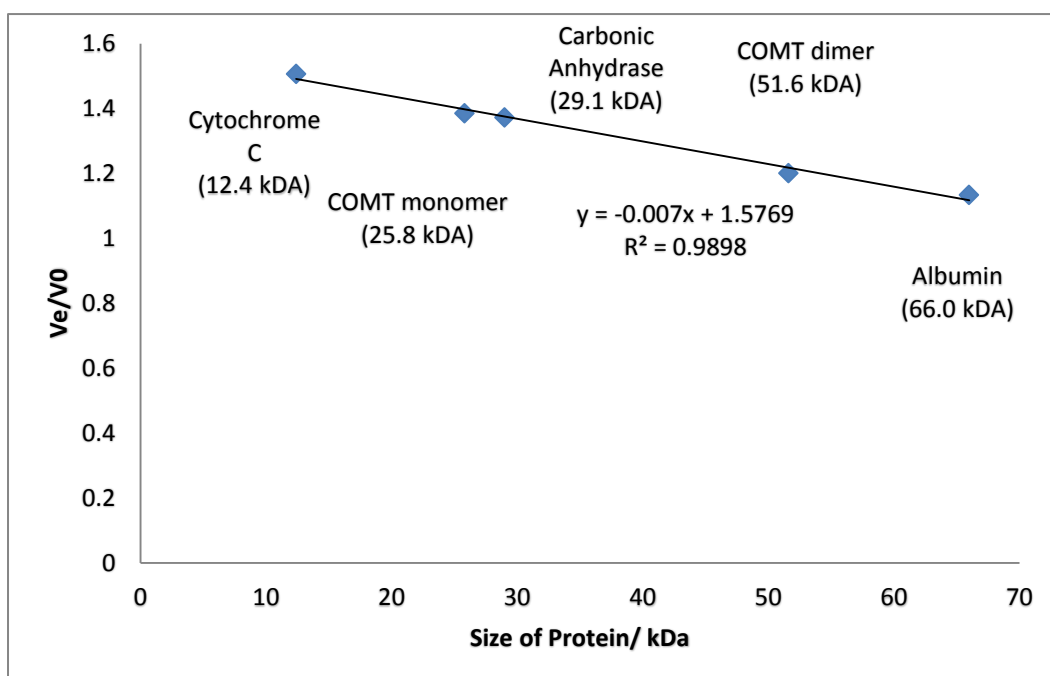
Figure 23: The Gel filtration FPLC calibration graph. Proteins used can be seen as data labels, COMT monomer and dimer can also be seen plotted on the graph.

Using the calibration graph, the mass of the first peak to elute was determined as 55 kDa whilst the second peak was determined to have a mass of 27 kDa. The masses of these peaks were larger than expected for monomeric (expected mass 25.8 kDa) and dimeric COMT (expected mass 51.6 kDa). However, the relationship between the two peaks' sizes was consistent with the first peak being dimeric COMT and the second peak being monomeric COMT. Therefore, the first peak, from the COMT injection, to elute was determined to be dimeric COMT and the second peak, from the COMT injection, to elute was determined to be monomeric COMT.
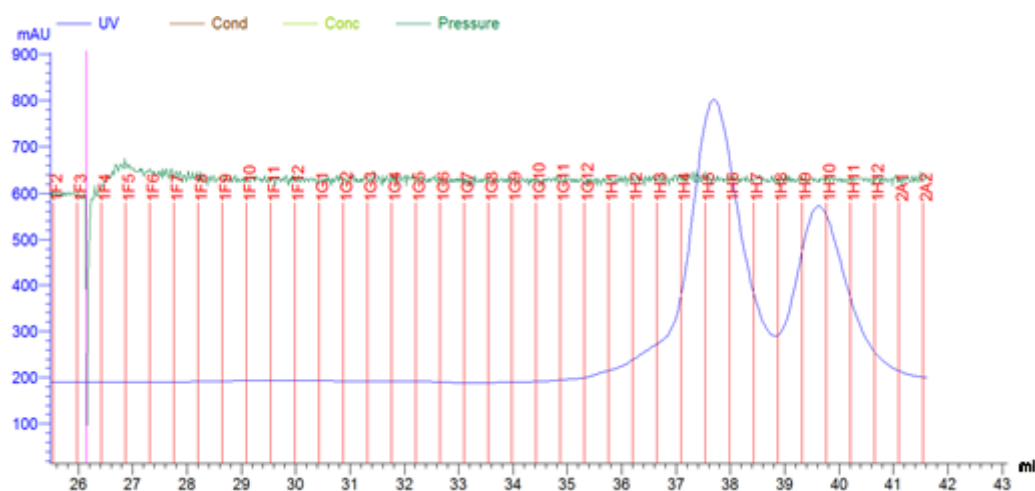
Figure 24: Gel filtration FPLC purification of WT COMT.

The confirmation of the identity of the two peaks seen in Y200L purification meant that further studies could be conducted into the quarternary structure of WT COMT. Purification of the WT enzyme by this process was necessary to determine whether the dimerization was an artefact of the mutation or stabilised by the Y200L mutation. The WT COMT was purified by the same method as the Y200L mutant (mentioned in Section 2.2.2) and again showed the two peaks of monomeric and dimeric COMT. The presence of two peaks showed that the dimeric structure also occurs in the WT enzyme, indicating that dimerization was not an artefact caused by the Y200L mutation.

Having observed the effects of mutation on the regioselectivity of WT COMT the next logical step was to determine the effect of quaternary structure on COMT's regioselectivity. The monomer and dimer peaks were collected (fractions G11-H5 for dimer and H10-2A2 for monomer Figure 24) from the FPLC using the AKTA fraction collector. However, this time a 50 mM potassium phosphate buffer pH 7.4

with 300 mM NaCl was used to elute the COMT. This buffer was chosen to keep consistency with the regioselectivity studies conducted with COMT mutants. The buffer showed no observable change in ratio of monomeric and dimeric COMT. Therefore, the oligomers were separated using these updated conditions.

Monomer and dimer WT COMT were then assayed with the DHBAL (**3**) substrate, a good starting point as this substrate showed the highest range of regioselectivity across the COMT mutants. Assays were set up with 5 µM of enzyme for both monomer (25.8 kDa) and dimer (51.6 kDa), 1 mM SAM (**1a**), 1 mM DTT and 3 mM MgCl$_2$ with a substrate concentration of 0.5 mM **3**. The assay was run for 1 hour at 37˚C with shaking at 800rpm. These conditions were chosen based on the optimal conditions for COMT (further information in Section 6.4.1).The results from the assay showed that dimeric WT COMT produced significantly more of the *meta* methylated product (vanillin (**3a**)) than monomeric WT COMT. The regioisomeric excess for dimeric WT COMT was 84% (in favour of *meta* methylation) whereas monomeric WT COMT displayed a r.e of 31% (Figure 25) (still in favour of the *meta* regioisomeric product). The results showed that the dimeric WT COMT was much more regioselective than the monomeric COMT.

The high regioisomeric excess observed for the dimeric WT COMT was an unexpected result. Previously all literature and data collected had shown that WT COMT produces both the *meta* and *para* methylated regioisomers with 37% regioisomeric excess. However, the previous data collected in this work had not previously separated monomeric and dimeric components. Therefore, it was no coincidence that the 37% regioisomeric excess observed for the WT COMT mix was between the regioisomeric excess values of monomeric and dimeric COMT. However, the regioisomeric excess of the oligomeric WT COMT mix was still lower than if the two values were averaged, suggesting that perhaps monomeric WT COMT dominated in the oligomeric mix at 37˚C. This explains the regioselectivity observed in Section 2.2.1 for WT COMT, however, not for previous literature. As previously mentioned, Creveling *et al* have quoted the *meta* to *para* ratio as 2.1 (36% regioisomeric excess) for the DHBAL (**3**) substrate.[98] The regioisomeric excess value (36%) quoted by Creveling *et al*,[98] was close to that of monomeric WT COMT (31%) but has a large discrepancy from the dimeric WT COMT (84%).

Dimeric COMT has been noted in the literature,[32,115] however, the regioselectivity of the dimeric COMT was not analysed in these studies. The high *meta* regioselectivity observed for the dimeric WT COMT should be further studied in order to yield a great understanding of what governs regioselectivity in both oligomeric forms of the COMT enzyme.

The results of the initial regioselectivity assay also displayed different conversions. The WT COMT dimer (62% conversion to *meta* and *para* methylated products) appeared showed more conversion of the DHBAL substrate to products than the monomeric WT COMT (49% conversion to products). However, the dimeric WT COMT has double the number of active sites (as a value of 51.6 kDa was used for calculating WT COMT dimer's concentration). Assuming both of the dimeric COMT's active sites were active then the conversion of substrate per active site was halved to 31%. This adjustment suggested that the dimeric WT COMT's active sites converted less DHBAL substrate to products than the monomeric enzyme. The lower conversion of the dimeric WT COMT could be the reason behind the WT COMT's 50% regioisomeric excess, observed in purifications of WT COMT without a gel filtration purification step, as the value was closer to the monomeric WT COMT's regioisomeric excess (31%). Thus the conversion of the two oligomeric forms of COMT were compared and also have led to a keener understanding as to what governs the regioselectivity of the COMT enzyme.

In order to test whether the dimeric WT COMT was less active than the monomeric WT COMT conversion assays were set up to establish whether this trend holds over other substrates as well. As well as DHBAL (**3**), NO2CAT (**5**) and DHBA (**4**) were tested as substrates, as **4** and **5** are similar in size and polarity to **3**. Activity assays were set up with the same conditions as those for the mutants (5 μM enzyme, 1 mM SAM (**1a**), 1 mM DTT and 3 mM $MgCl_2$ with a substrate (**3-5**) concentration of 0.5 mM). Unlike the K144E assays, the assays were terminated after 10 minutes to prevent full conversion, as WT COMT has been shown to be significantly more active (this would allow relationships to be drawn between substrates (**3-5**)). After termination, the conversion of NO2CAT (**5**) to *meta* (**5a**) and *para* (**5b**) methylated regioisomers was determined by the HPLC method laid out in Section 6.5.2. Whereas the DHBA (**4**) substrate's conversion to the *meta* (**4a**) and *para* (**4b**)

methylated regioisomers was determined by the same HPLC method as DHBAL (**3**) (Section 6.5.1). These conditions meant that the assay could be trialled to determine activities of the two oligomeric structures respectively.

	DHBAL	DHBA	NO2CAT	DHBAL	DHBA	NO2CAT
■ para	5	9	7	17	16	16
■ meta	57	33	93	32	48	33
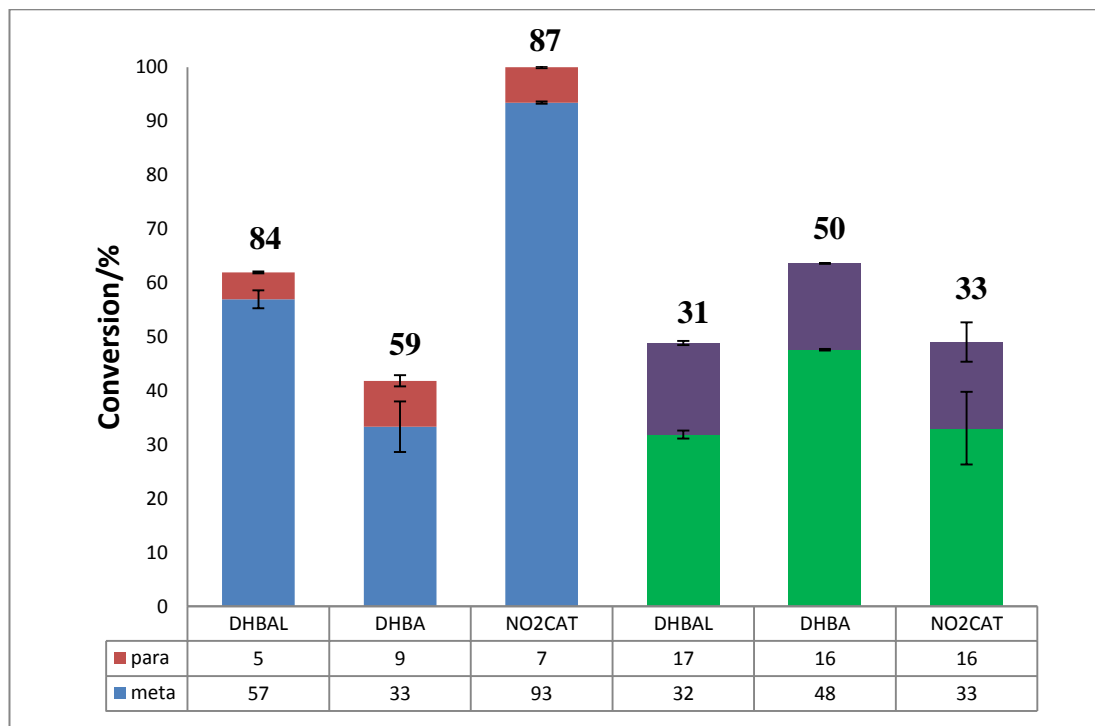
Figure 25: Conversion and regioselectivity assays of monomeric and dimeric COMT for; DHBAL (**3**), DHBA (**4**) and NO2CAT (**5**) substrates. Dimeric COMT activities are shown in red (*para* product (**3b**, **4b**, **5b**) and blue (*meta* product (**3a**, **4a**, **5a**)), monomeric COMT activities are shown in purple (*para* (**3b**, **4b**, **5b**) product) and green (*meta* ((**3a**, **4a**, **5a**) product). R.e. can be seen above each bar.

The assays showed significantly different chemistry for monomeric and dimeric COMT with regards to regioselectivity and conversion. The high *meta* regioselectivity observed for dimeric WT COMT with the DHBAL (**3**) substrate (regioisomeric excess 84%) was also observed with the NO2CAT (**5**) substrate (r.e 87%) (Figure 8). Whereas monomeric WT COMT showed lower r.es for substrates; **3** (r.e 31%) and **5** (r.e 33%). Interestingly, the DHBA (**4**) substrate did not show the large regioisomeric differences between dimeric and monomeric WT COMT with only a 9% r.e difference between the two values. The assay data suggested that the substrates most affected by the differences in the oligomeric forms of the WT COMT enzyme were NO2CAT (**5**) and DHBAL (**3**) whilst DHBA (**4**) regioselectivity was not as affected.

The regioselectivity observed by the conversion assays, suggested that NO2CAT (**5**) and DHBAL (**3**) substrates interact differently with monomeric and dimeric forms of

COMT unlike the DHBA (**4**) substrate. However, the conversion of the **4** substrate to the products; vanillic acid (**4a**) and isovanillic acid (**4b**), with the monomeric or dimeric form of COMT was shown to be very different (Table 6). Table 6 showed that for the **3** and **4** substrates the dimeric WT COMT showed less conversion than the monomeric WT COMT. The conversion of **4** by the dimeric WT COMT (21%) to products was much lower than the conversion of the monomer (64%). This suggested that in the case of **4** either; only one active site of the dimer was active with the DHBA (**4**) substrate or methylation of **4** was somehow hampered by the dimeric structure. Whereas Table 6 showed that the substrate **5** was the substrate, which was most converted to it's products (4-nitroguaiacol (**5a**) and 2-methoxy-5-nitro-phenol (**5b**)) with dimeric WT COMT showing full conversion to the *meta* (**5a**) and *para* (**5b**) methylated regioisomers. The 50% activity per active site was only approximate, as 100% conversion could have occurred before termination of the assay. The high conversion of NO2CAT (**5**) meant that unlike the other two substrates, **5** exhibited greater conversion with the dimeric WT COMT (50% at least) as opposed to the monomeric WT COMT (49%). This data suggested that **5** was a better substrate with the dimeric WT COMT than monomeric WT COMT.

Table 6: Conversions of monomeric and dimeric WT COMT. Conversions for dimeric WT COMT are displayed as both observed (overall activity) and conversion per active site.

Substrate	Monomeric conversion /%	Dimeric observed conversion/%	Conversion per active site (dimer)/%
DHBAL (**3**)	49	62	31
DHBA (**4**)	64	42	21
NO2CAT (**5**)	49	100	≥50

The overall trend suggested that conversions with the dimer of the three substrates was NO2CAT (**5**)>DHBAL (**3**)>>DHBA (**4**) whereas with monomeric COMT **4**>**3**≈**5**. The regiochemistry displayed for the dimer and monomer also showed the greatest difference with the **5** substrate (dimer 87% monomer 33%) and the smallest difference with **4** substrate (dimer 59% monomer 50%). These results suggested that

the DHBA (**4**) interacted differently with the oligomeric states of COMT compared to the NO2CAT (**5**) and DHBAL (**3**) substrates. In order to determine what were the underlying causes of these differences further investigation of the quarternary structures of COMT was needed.

2.2.4   Stability and kinetic assays of oligomeric COMT

The conversion assays had shown fundamental differences in dimeric and monomeric COMT's regiochemistry and activity. In order to understand the causes of these differences the cause of COMT dimerisation was further investigated. Dimerisation often occurs in proteins through disulfide linkages.[116] Dithiol linkages are prevented using DTT, which reduces the disulfide linkages and breaks up protein dimers that are formed through these bonds. However, previously DTT has been used in purifications of COMT, as shown with the gel filtration buffer (Section 6.2.3). Despite this, the gel filtration buffer showed equal ratios of monomeric and dimeric COMT, as buffers that did not have DTT (assessed by FPLC). The consistent amount of dimeric COMT in these buffers with and without DTT, suggested that the dimerisation was not formed through disulfide linkages, as addition of DTT would have prevented this. Based on this data, initial experiments suggested that disulfide linkages are not stabilising COMT dimerisation, as the presence of DTT does not seem to hinder dimer formation.

The hypothesis that dimerisation was not caused by disulfide linkages was further tested by conducting 24 hours experiments on the stability of dimeric COMT. COMT was separated, as previously into monomeric and dimeric samples. The separated monomeric and dimeric COMT were then re-injected onto the FPLC to give a t = 0 time point. Monomeric (2 mg/mL) and dimeric COMT (3 mg/mL) were then stood for 24 hours at either 22˚C or 4˚C with one extra 4˚C sample with 10 mM DTT added. The assays were then rerun on the FPLC with the areas of monomer and dimer being measured (Figure 26 and Figure 27). The result of this experiment again suggested that dimeric WT COMT was not stabilised by dithiol interactions, as the DTT assay shows little change in dimeric to monomeric COMT in ratio compared to the stock without DTT. Whereas the dimeric WT COMT kept at 22 ˚C overnight shows a much greater shift to the monomer (from 24:1 to 2:1 dimer to monomer)

compared to the stock kept at 4 ˚C (24:1 to 11:1 dimer to monomer). The monomeric WT COMT stocks showed the opposite results with monomeric WT COMT shifting towards dimeric WT COMT overnight at 4 ˚C and little change in the sample kept at 22 ˚C. However, the change in monomeric WT COMT to dimeric WT COMT was not as dramatic as dimeric WT COMT.

	4C	22C	4C	22C
	t=0 hours		t=24 hours	
■ Monomer	323	323	309	324
■ Dimer	12	12	25	10
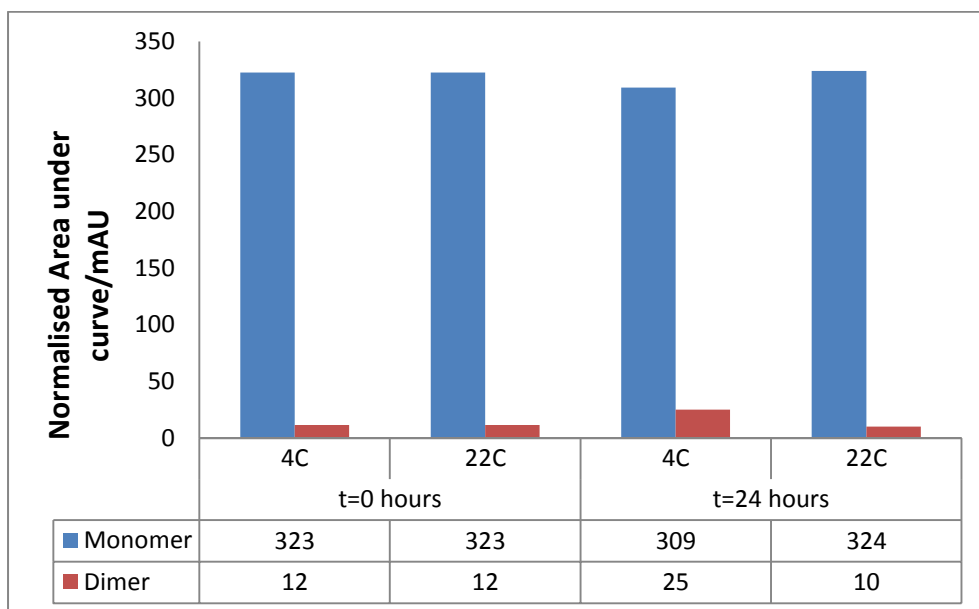
Figure 26: Graph indicating the change from monomeric COMT to dimeric COMT at different temperatures overnight.

	4C	22C	DTT	4C	22C	DTT
	t=0			t=24hours		
■ Monomer	14	14	14	29	98	23
■ Dimer	331	331	331	316	205	330
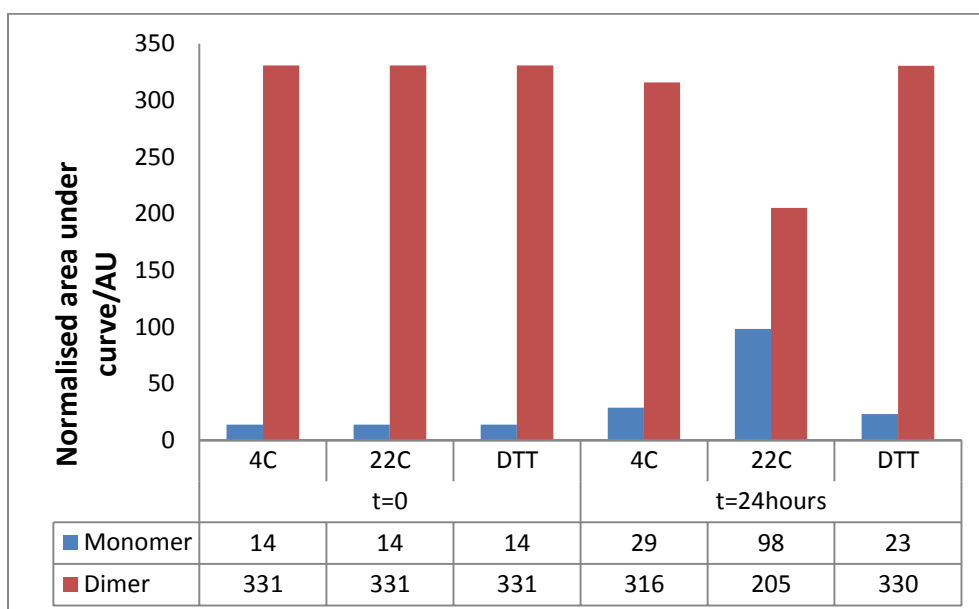
Figure 27: Graph indicating the change from dimeric COMT to monomeric COMT at different temperatures and concentrations of DTT, overnight.

95

The results of the overnight stocks suggested that dimerisation of COMT was not stabilised by disulfide linkages and that the dimer was in equilibrium with the monomer. Disulfide bonds have been shown to be fairly strong with a dissociation energy of 250 kJmol^{-1}, which was a similar strength to C-C peptide backbone bonds (350-355 kJmol^{-1}).[116] However, with dimeric COMT the equilibrium was also shown to be altered by thermal fluctuations, suggesting that the stabilisation was weaker than disulfide linkages. At higher temperatures WT COMT monomer would be the direction in which the equilibrium would lie, whereas at lower temperatures dimer formation was slightly favoured. The equilibrium lying in the direction of WT COMT monomer (at higher temperatures) could therefore account for the regioselectivity of the oligomeric WT COMT mix (r.e. 37 % DHBAL (**3**)) being observed as closer to the regioselectivity of monomeric WT COMT (r.e. 31% for **3**) rather than the dimeric WT COMT (r.e. 84% for **3**). The observation that monomer formation occurs from the dimer at the higher temperatures, has also been observed by Ehler *et al* whilst this study was in progress.[115] Therefore, the overnight experiments suggested that at higher temperatures, monomer formation was more preferred, whilst at lower temperatures the dimeric WT COMT was more stable.

The overnight studies had shown some more information about the two oligomeric forms of the COMT; suggesting that the dimer was not formed by disulphide bonds. However, this detail, whilst useful from a structural perspective did not provide any information on regioselectivity or the methylation of substrates. The understanding of how monomeric and dimeric COMT interact with substrates (**3**-**5**) was best answered by determining the kinetics for monomeric and dimeric WT COMT. The COMT enzyme has been shown to fit Michaelis-Menten kinetics with the K144A mutant shown in Section 2.2.3. Therefore, a similar kinetic experiment can be devised for monomeric and dimeric WT COMT. Understanding the K_m and k_{cat} of the monomer and dimer of WT COMT will yield further information about the binding and relative activities of the monomeric and dimeric WT COMT.

The decision to investigate WT COMT monomer and dimer kinetics meant that a suitable substrate should be chosen. Previous studies of K144A kinetics had shown that kinetics for WT COMT could be measured by HPLC. However, these studies had involved an oligomeric mix of the monomer and dimer. The success of the

K144A kinetics with the DHBAL (**3**) substrate meant that **3** was seen as a good starting point. The kinetic assay set up was initially the same as the K144A kinetic approach (see Section 6.4.1), however, the substrate saturation concentration was observed much earlier for both monomer and dimer. Based on these saturation studies, concentrations from 2 μM- 300 μM for dimeric COMT (full list Section 6.4.1) and 2 μM – 100 μM for monomeric COMT were chosen (full list Section 6.4.1). These initial studies also found that the monomer and dimer were faster than the K144A enzyme so time points of 1, 2 and 3 minutes were used, as opposed to 2,3,4 minutes. The final factor that was altered from the K144A kinetic experiment, based on the saturation experiments, was the concentration of the enzyme with 0.1 μM for dimeric COMT and 0.2 μM for monomeric COMT being used.

The saturation studies showed that Michaelis-Menten kinetic parameters could be applied to monomeric and dimeric COMT. The conditions (as mentioned previously) were used to run kinetic experiments for both monomeric and dimeric WT COMT. The Michaelis-Menten curves (see Figure 28) for monomeric and dimeric COMT recorded low errors and were considered to have a good level of precision. Using these curves the kinetic parameters were calculated (see Section 6.4.1 for calculation process). The results showed that monomeric COMT recorded a K_m of 1.3 ± 0.2 μM and a k_{cat} of 2.3 ± 0.1 min^{-1}. Dimeric COMT recorded a K_m of 3.7 ± 0.3 μM and a k_{cat} of 1.6 ± 0.1 min^{-1} per active site ($k_{cat overall}$ = 3.18 ± 0.03min^{-1}). These values were considered consistent for WT COMT with good agreement with the conversion assays, which showed monomeric WT COMT (49%) to be the more active than dimeric WT COMT (31% corrected) when the data was accounted for per active site. Existing data also showed fairly good agreement with these experiments, with a study conducted by Dr Brian Law recording a K_m of 3.4 ± 0.3 μM and a k_{cat} of 4.1 ± 0.1 min^{-1} for the WT COMT oligomeric mix with the DHBAL (**3**) substrate (the differences observed could be due to different interactions in the oligomeric mix.
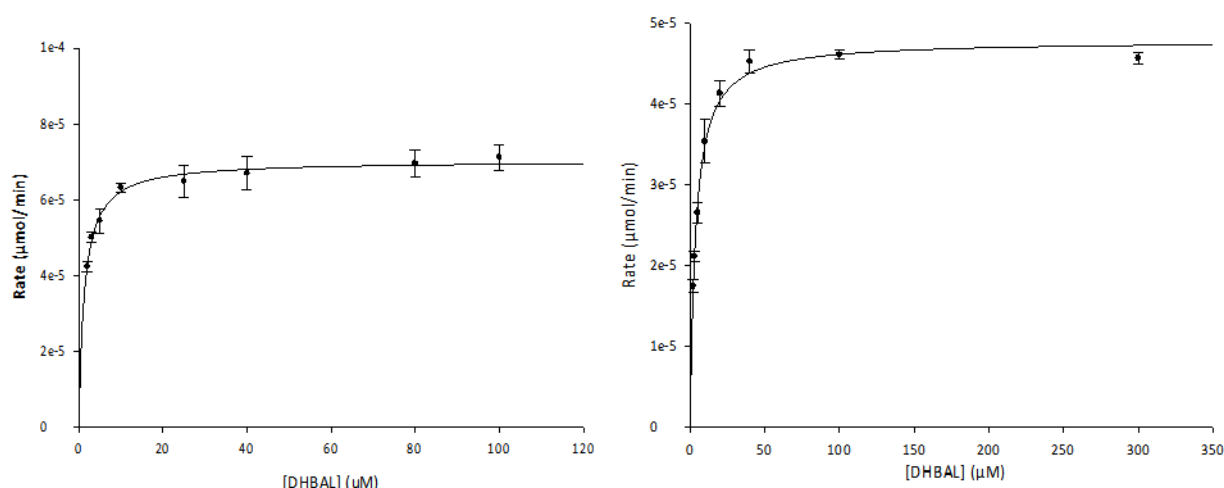
Figure 28: (Left) Monomeric WT COMT DHBAL (**3**) substrate kinetics. (Right) Dimeric WT COMT DHBAL (**3**) substrate kinetics.

The monomeric WT COMT's higher k_{cat} (2.3 ± 0.1 min^{-1}) indicates that the monomeric enzyme methylates faster than the dimeric WT COMT with a lower k_{cat} (1.6 ± 0.1 min^{-1}). The lower K_m (1.3 ± 0.2 µM) of the monomeric WT COMT suggested that the monomer binds the DHBAL (**3**) more tightly than the dimeric WT COMT that has a higher K_m (3.7 ± 0.3 µM). This data again proposed that the dimeric WT COMT has different interactions with **3** compared to the monomeric WT COMT structure. Since the dimeric COMT has been shown to have a mass consistent with being a dimer (see Section 2.2.3) these differing interactions are most likely due to the different arrangement of active site amino acids between the monomeric and dimeric WT COMT rather than truncation or a random mutation in the dimeric WT COMT. Understanding the different arrangement of the active site of the dimeric WT COMT, or indeed the structural differences, between the WT COMT oligomers would enable a further understanding of the observed differences in COMT regioselectivity between the two oligomers.

The data so far observed has shown that both monomeric and dimeric COMT display very different regioselectivity, as well as kinetics between the two oligomers. The differences between monomeric and dimeric COMT have been attributed to the structural differences. In order to determine what these structural changes were a crystal structure of dimeric WT COMT was sought. Whilst dimeric WT COMT structures have been collected,[115] most have aimed at examining inhibitors and few have crystallised WT COMT with SAM (**1a**) or SAH (**1**). Therefore, in order to

understand the regioselectivity observed for dimeric WT COMT, crystallisation of dimeric WT COMT with a **1a** or **1** cofactor and a substrate or appropriate inhibitor was the goal of this work. In order to determine the structural differences between monomeric and dimeric WT COMT, a crystal structure was sought for WT COMT dimer.

2.2.5 Crystal Structure of dimeric WT COMT

The success of the Y200L crystal structure meant that similar conditions were adopted for preparation of the dimeric WT COMT sample. COMT was expressed and purified as before, with a nickel column, anion exchange column and finally the gel filtration step. The gel filtration buffer used, did still include DTT, as this had been shown to have little effect on the WT COMT dimer (Figure 27). The sample purity at each step was verified by a SDS PAGE gel. The gels confirmed the Y200L purification procedure, as only COMT was observed in the gel filtration step.

However, deviation from the Y200L crystallographic procedure was chosen for the gel filtration step. Fractions containing the dimer were selected with extreme prejudice (Figure 29 right) with only fractions 1A, 1B and 1C selected for crystallisation. The strict cut off was chosen as the monomeric and dimeric peaks show slight overlap, therefore, to avoid any chance of monomer crystallisation only 1A, 1B and 1C were chosen. The selected samples (1A-C) were pooled and then dimeric WT COMT (1 μM) was soaked with DHBAL (**3**) (10 μM) and SAH (**1**) (10 μM) in gel filtration buffer (50 mL), overnight at 4°C (to slow down conversion to monomer) before concentration. After concentration the samples were then submitted again to Dr Colin Levy for crystallographic trials.

The crystallisation was successful and yielded a crystal structure with resolution of 2.2 Å. However, the structure that was collected was an *apo* structure, with only the SAH (**1**) bound (Figure 29 left). The collected WT COMT dimer structure shows some structural differences when compared with the Vidgren structure (Figure 29 right).[1] The structure confirmed that the first Peak isolated from the FPLC was indeed dimeric WT COMT and under the crystallisation conditions could be observed as a dimer of dimers (Figure 29). The Rossmann fold could again be

observed, as expected for the Class I methyltransferase COMT structure. However, the C-terminus beta strands of the Rossmann fold were noted to have been domain swapped (see Figure 29 left). The domain swapping can be seen in Figure 29, as the orange and green strands can be seen as swapping to the alternative monomer. In the Vidgren structure (Figure 29 right) the C-terminus contains the β6 and β7 sheets, however, in the dimer structure these sheets forms a flexible strand.[1] The flexible strand associates with the other monomer of the dimer (see in orange). The observed domain swapping, may be lead to the higher regioselectivity observed for the dimeric WT COMT.
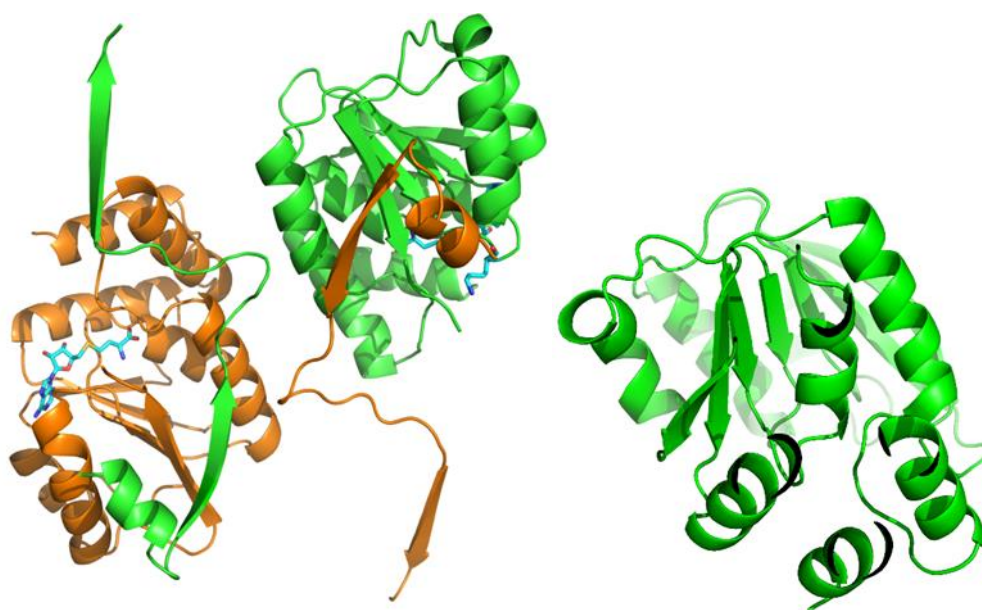
Figure 29: (Left) The apo structure of dimeric WT COMT, one monomeric unit is in green, whilst the other is orange. (Right) Adapted from PDB (1VID), Vidgren's monomeric WT COMT structure.[1]

There was a precedent for domain swapping occurring in dimeric WT COMT structure, firstly Ellerman *et al* reported observing domain swapping (see supporting information of Ellerman *et al*),[32] and whilst this study was being conducted domain swapping was also observed by Ehler *et al*.[115] The domain swapping may be an artefact of the crystallisation procedure, and may not be biologically relevant as Ehler states.[115] However, the dimeric WT COMT has been shown already in this study to have distinct biological activity from the monomeric structure. The distinct biological activity suggested that the dimeric WT COMT was structurally distinct from monomeric WT COMT. Therefore, the dimeric WT COMT's differing

regiochemical properties may be explained by the structural differences caused by domain swapping.

The *apo* structure of dimeric WT COMT has also shown that there were no disulfide linkages in the dimeric WT COMT. The lack of disulfide linkages suggested that the dimer WT COMT needs alternative stabilisation. The stabilisation of the dimeric WT COMT may be generated by the dimeric WT COMT adopting a lower energy state. The lower energy state may be what was observed in the *apo* dimeric WT COMT structure i.e. the domain swapping. Therefore, this suggested that the domain swapping may stabilise the dimerisation of COMT. The lower stabilisation energy provided by dimerisation through domain swapping compared to disulfide linkages (which are likely to be stronger), may account for the equilibrium shifting toward the monomeric WT COMT with thermal fluctuations, as was shown in the overnight experiment earlier (Figure). The *apo* structure and the previous overnight investigation suggested that domain swapping was a dynamic process, which stabilised the dimeric WT COMT.

The *apo* structure showed key structural features, present in the dimeric WT COMT that was not observed in the monomeric WT COMT structure. However, the structure collected may have not been biologically relevant, as the substrate was not bound. The *apo* structure has also shown that the active site residues of Y200 and E199 (yellow Figure 30) are not in the correct closed position. The *apo* structure also presented SAH (**1**) in two differing conformations of which only one would be active (Figure 30 right). Only one site containing an active conformation of **1** suggested that there was half site conversion with the dimer, a phenomenon only accounted for by DHBA (**4**)'s conversion assay results (Figure 30). The inactive site could be providing some allosteric modulation, which could account for the dimer's high regioselectivity. Therefore, in order to determine whether the structural differences between monomeric WT COMT and dimeric WT COMT (as well as determining the correct active site residue and SAH (**1**) orientations) determined from the *apo* structure were biologically relevant a *holo* structure was needed.
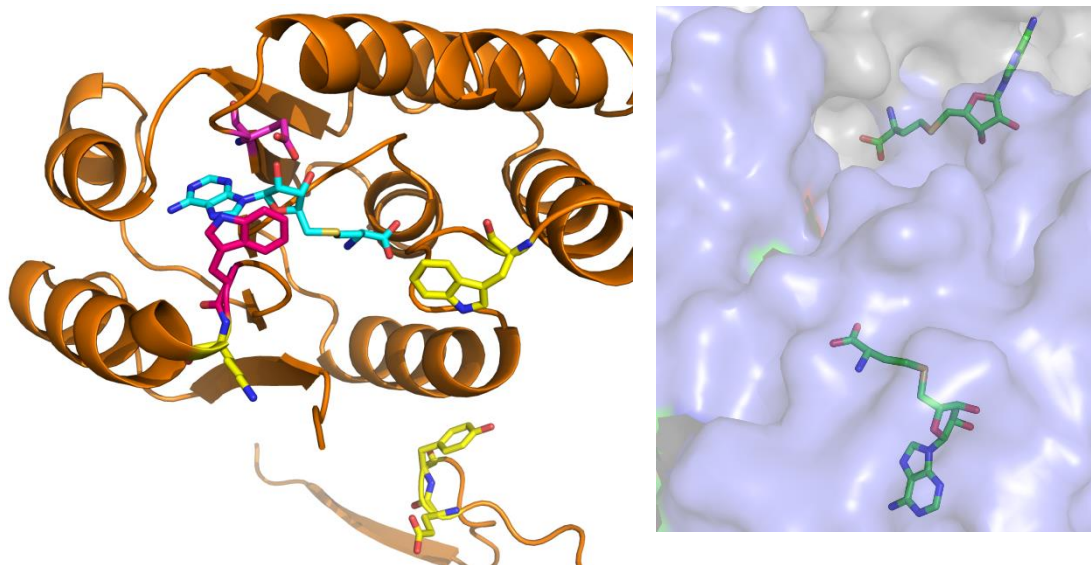
Figure 30: (Left) The dimeric structure with the movement of key residues highlighted in yellow (K144, Y200, W38 and E199). (Right) The two different orientations of SAH (**1**) in the apo dimeric WT COMT structure.

In order to yield a holo crystal structure for the dimeric WT COMT DNC was chosen instead of DHBAL (**3**). As with the Y200L structure, DNC was chosen because DNC has been shown to be an excellent competitive inhibitor of COMT with an IC$_{50}$ of 12 nM.[117] Whereas with **3** binding was much weaker with kinetics showing a K_m of 3.7 μM. DNC was also the inhibitor used by Vidgren *et al* for the original crystal structure of COMT.[1] The use of DNC would, therefore, also allow easy comparison to that of the monomeric Vidgren structure.[1] DNC would also allow SAM cofactor (**1a**) to be used in the crystallisation study due to DNC being a very poor substrate. These conditions would improve the chances of crystals including both substrate and cofactor. The crystallisation conditions and the choice of substrate would also improve the chances of observing the *holo* form of dimeric WT COMT, rather than the *apo* form with just SAH (**1**) bound. Using these ligands the samples were again submitted to Colin Levy for crystallographic trials.
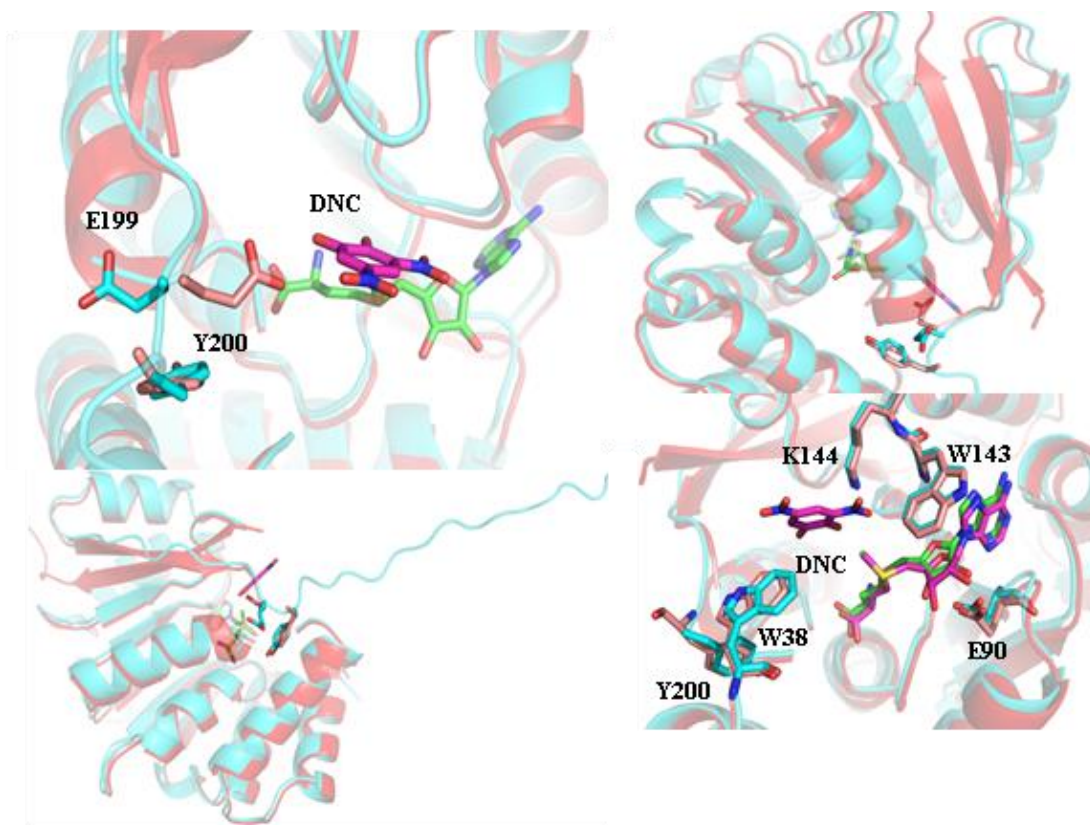
Figure 31: (Top left) The orientations of E199 and Y200 in dimeric WT COMT (cyan) and monomeric WT COMT, adapted from PDB (1VID), (pink).[1] (Top right) The absence of two beta sheets can be observed overlaying the monomeric WT COMT structure (red) and the dimeric WT COMT structure (cyan). (Bottom left) The domain swapped beta sheet can be observed as a flexible strand. (Bottom right) An indication of the residue positions of other key residues in the dimeric and monomeric WT COMT structures.[1]

The crystallisation with the DNC inhibitor was shown to be successful, yielding a *holo* structure with a resolution of 1.6 Å (Figure 31). The cofactor SAM (**1a**) and the inhibitor DNC were both observed bound to the dimeric WT COMT (Figure 31 top left). Overlaying the monomeric WT COMT structure from Vidgren *et al*,[1] again shows the differences between monomeric and dimeric WT COMT seem to be due to domain swapping. The fact that domain swapping was again observed in the *holo* structure suggested that domain swapping was probably not due to the crystallisation process. The monomeric WT COMT has two, previously mentioned, *beta* sheets (β6 and β7) that did not overlay with the dimeric WT COMT (here to amplify the differences only a monomeric unit of the dimeric WT COMT was shown, Figure 31 top right). The domain swapping was clearly observed by the C-terminal of the dimeric WT COMT structure, which again, like the *apo* structure, extended out as a flexible strand with no overlay with the Vidgren structure (Figure 31 bottom right).[1]

The *holo* structure showed the key residues were observed in the active site one example being the K144 residue facing the hydroxyls of the DNC inhibitor. The K144 residue in the *apo* structure faced out into the solvent. Other key residues were also in the correct closed conformation, with W143 and E90 both binding to SAM (**1a**) as in the monomeric WT COMT structure (Figure 31), as well as W38 and Y200 adopting similar positions to the monomeric WT COMT structure.[1] The cofactor was also observed in one conformation, suggesting the two conformations observed in the *apo* structure were artefactual. Therefore, the replacement of DHBAL (**3**) with the inhibitor DNC yielded a *holo* structure with most of the residues adopting similar positions, as observed in the WT closed conformation.

The *holo* dimeric WT COMT structure did show that most residues adopted the same conformation, as they did in the Vidgren structure with one notable difference; E199.[1] The Y200L structure (Section 2.2.2) had shown that the E199 residue could flip out of the active site and it was suggested that this was the cause of Y200L's *meta* regioselectivity. The *holo* dimeric WT COMT structure again shows the E199 flipped out the active site, as has been seen in Figure 31 with the overlay of monomeric WT COMT with dimeric WT COMT. The absence of E199 from the active site suggested, like the structure for Y200L, that this was the cause of dimeric WT COMT's regioselectivity, due to the lack of stabilisation for *para* methylation in the case of DHBAL (**3**) (Figure 31).

The absence of the E199 residue from the active site may also explain the previously observed results for the conversion assays. The dimeric WT COMT had shown strong *meta* regioselectivity for NO2CAT (**5**) (87% r.e) and DHBAL (**3**) (84% r.e) substrates but less *meta* regioselectivity was noted for the DHBA (**4**) (59% r.e) substrate. The E199 residue would not form as strong a hydrogen bond with **4** compared to **3**, which may account for the little regioselectivity difference between the **4** substrate with monomeric and dimeric COMT. The electron density on the acidic group of **4** was shared across two oxygens. The lower electron density on the oxygens of DHBA (**4**) compared with DHBAL (**3**) could have led to a weaker hydrogen bond between **4** and E199 compared with **3** and E199.
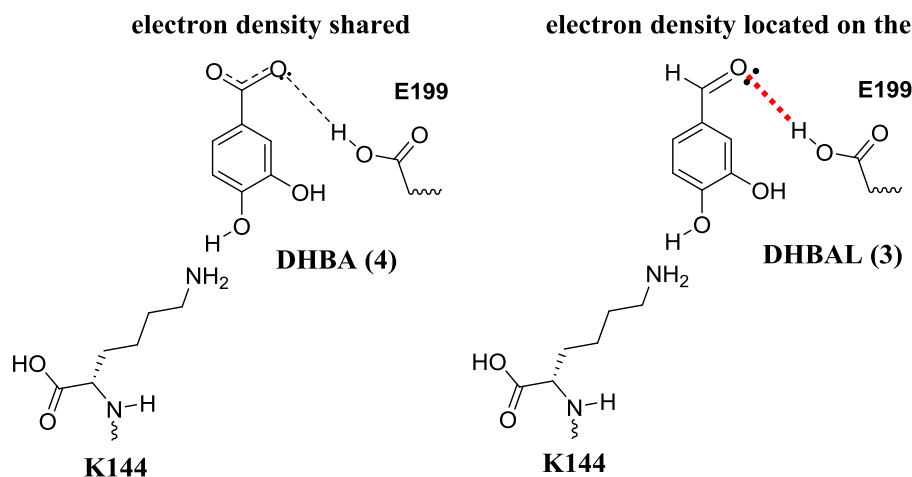
Figure 32: Representation of the suggested electronic interactions between substrates DHBA (**4**) and DHBAL (**3**) with the E199 residue.

The weaker hydrogen bond between DHBA (**4**) and E199 means that **4** was not as affected by E199's orientation, which may be why **4** does not have as large a regiochemical difference between monomeric and dimeric COMT. The repulsive nature of the acidic groups between two molecules of DHBA (**4**) may account for the suggested lack of dual site conversion with **4**. The close proximity of the two monomeric active sites in the WT dimer may cause the repulsion between the DHBA (**4**) molecules. The hydrogen bonding with E199 and the general dimeric structure, therefore, explains the observed results from the conversion assays.

The NO2CAT (**5**) substrate was also capable of resonance, which suggested that the hydrogen bond may be weakened. However, the four valent positively charged nitrogen means that more electron density was delocalised onto the oxygen atoms from the aromatic ring in **5**. The collection of charge at the oxygen atoms of NO2CAT (**5**) could have been substantially more than the acid or aldehyde of DHBA (**4**) and DHBAL (**3**) respectively. Therefore, the hydrogen bond between E199 and the NO2 group could have been greater than or equal to that of the aldehyde. Therefore, the regiochemistry of the **5** substrate would have been expected to be altered in a way similar to **3**. Therefore, the presence (monomeric WT COMT) or absence (dimeric WT COMT) of the E199 residue, could be suggested to affect the regioselectivity of the NO2CAT (**5**) substrate (like **3**) through interference of the *para* stabilising hydrogen bond between **5** and E199 (Figure 32).

105

However, hydrogen bonding of the E199 residue to the 4-position was only one hypothesis for the determined regioselectivity of the two oligomeric states of COMT. The alternative explanation could be based on the predicted pKa's of the substrates, as mentioned previously in Section 2.2.2. A hydrogen bond could have also formed between the acidic side chain of E199 and the hydrogen of the hydroxyl of the catecholic moiety itself, which had been suggested by several literary sources (Figure 33).[1,5] In this case the E199 could have held the catechol in either orientation suitable for *para* methylation or *meta* methylation. However, when the E199 residue was no longer facing into the active site the most stabilised methylation pathway could have occured. *Meta* methylation has been shown to be the quicker pathway for dopamine,[99] and could also have been the quicker pathway for DHBAL (**3**) and NO2CAT (**5**) than *para* methylation (as the deprotonated *meta* position was the most unstable, due to the lack of resonance stabilising effects). Whereas, the the pKas of the two DHBA (**4**) hydroxyls were thought to be nearly equal, so neither anion was destabilised. Therefore, as the results show, an increase in *meta* methylation would be observed if the E199 was absent from the active site (as was the case with dimeric WT COMT) for DHBAL (**3**) and NO2CAT (**5**) substrates, but DHBA (**4**) would not have been as affected.
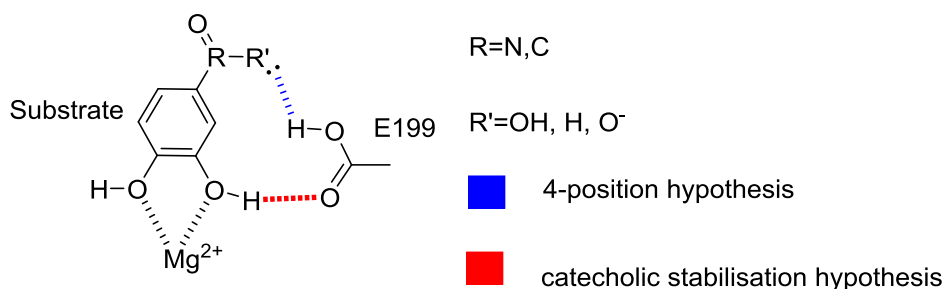
Figure 33: The two suggested hypotheses for E199's effects on COMT regioselectivity for COMT substrate (**3**-**5**).

The dimeric WT COMT structure and the oligomeric WT COMT conversion assay results have shown that monomeric and dimeric COMT exhibit very different regiochemistry. From the dimeric WT COMT structure, this has been explained, like Y200L, as being due to the E199 residue's absence from the active site. The absence of E199 from the active site of the dimeric WT COMT accounts for the regiochemistry exhibited for the three tested substrates. DHBA (**4**) methylation showed the lowest regioselectivity whilst NO2CAT (**5**) shows the highest due the

substrate's electronic interactions with the E199 residue. However, how the absence of E199 leads to more *meta* methylation has led to two possible hypotheses. The discovery of different regiochemistries for the monomeric and dimeric WT COMT also suggested that future studies should investigate whether this phenomenon occurs in all constructs and species of COMT. Investigating these oligomeric regioselectivities in futher studies was determined to be particularly important, in view of developing inhibitors, as E199 has been shown to be very important in substrate binding.

2.2.6   Oligomeric forms of COMT mutants

The previous section had shown that once again E199 appeared to be an important residue in dictating the regiochemistry of the COMT enzyme. Therefore, mutation of the E199 residue into amino acids incapable of hydrogen bonding was conducted. E199 mutants were also aimed at determining, whether position or polarity of the E199 residue was important. This section will also investigate oligomeric forms of the key E199 and Y200 mutants. The aim of investigation of the oligomeric forms of the COMT mutants was to determine what role the mutations play; whether they are stabilising dimerisation or whether the regioselectivity was dictated by the mutation. Investigation of these two areas will allow a more detailed conclusion to be drawn on the factors that affect COMT regioselectivity.

The discovery that E199 may have been effecting regioselectivity via a hydrogen bond to the substrates (DHBA (**4**), DHBAL (**3**) and NO2CAT (**5**)) required further investigation by mutagenesis of the E199 residue. The mutation at E199 was chosen, firstly based on an approximately similar size to glutamic acid and secondly, and most importantly, for the residue's hydrophobicity: E199L. The mutation to E199L was expected, if the current hypothesis was correct (Figure 34), to remove the stabilising hydrogen bond for *para* methylation then E199L should increase *meta* methylation.

Regioselectivity assays with the same conditions as those for the initial screen (with a DHBAL (**3**) substrate) mentioned in Section 6.4.1, were performed for the E199L mutant without separation into monomer dimer initially. No oligomeric separation

was initially carried out, as without the possibility of a hydrogen bond the two oligomers should behave similarly. The result of these assays showed a regioisomeric excess of 85% in favour of *meta* methylation with the DHBAL (**3**) substrate. The 85% value was between the values for dimeric WT COMT (83.8%) and oligomeric Y200L (89%) with **3**. The result of improved *meta* regioselectivity of E199L mutant showed that both hypotheses could still explain the increased production of the *meta* regioisomer of DHBAL (**3a**), as the two suggested hydrogen bonds (Figure 34) would both be removed by the mutation to leucine. The initial data from the E199L mutation, suggested that the E199 residue was important in dictating the regioselectivity of the COMT enzyme.
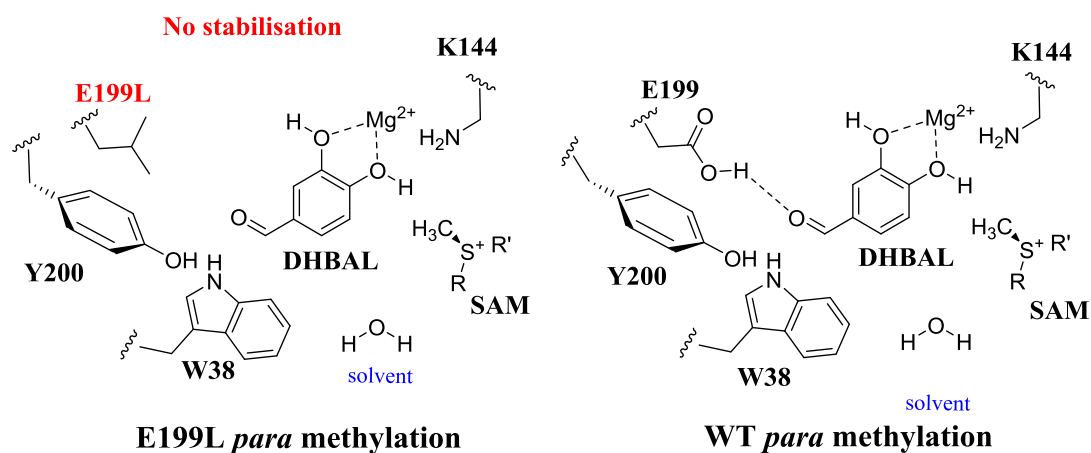
Figure 34: (Left) E199L mutation lacks the stabilisation of the WT COMT structure (right) for *para* methylation. Residues that change highlighted in red.

The E199L mutation regioselectivity assays with DHBAL (**3**) suggested that an acid group was important for *para* methylation. In order to test whether glutamic acid itself was important or whether any other acidic amino acid residue could be used the E199D mutation was introduced. The resulting assays showed similar regioselectivity as the E199L (85%) with E199D recording a regioisomeric excess of 85% (E199D mutagenesis and assays were carried out by Dr Brian Law of the Micklefield group). The *meta* favoured regioisomeric excess of E199D suggested that aspartic acid was unable to stabilise *para* methylation at the E199 position. The lack of stabilisation of *para* methylation by E199D mutant suggested that the position of acidic residue was also important. The importance of the E199 position adds further weight to the hypothesis that the E199 forms the long hydrogen bond

(between the oxygen atom of the aldehyde, acid or N=O for NO2CAT (**5**)), rather than the shorter contact between the hydroxyl and E199L, to stabilise *para* methylation. The E199D was 7.5 Å away from W38, which was already at the extreme end for a hydrogen bond to form in the hydrophobic pocket. Whereas the distance between E199D and the hydroxyl of the DHBAL (**3**) was significantly less at 5.6 Å. Therefore, the E199D mutation suggested that E199 does stabilise *para* methylation through a hydrogen bond to the lone pair of the oxygen of the substrates' (DHBAL (**3**), DHBA (**4**) and NO2CAT (**5**)) 4 position and that this hydrogen bond dictates the regioselectivity of the active site of COMT.
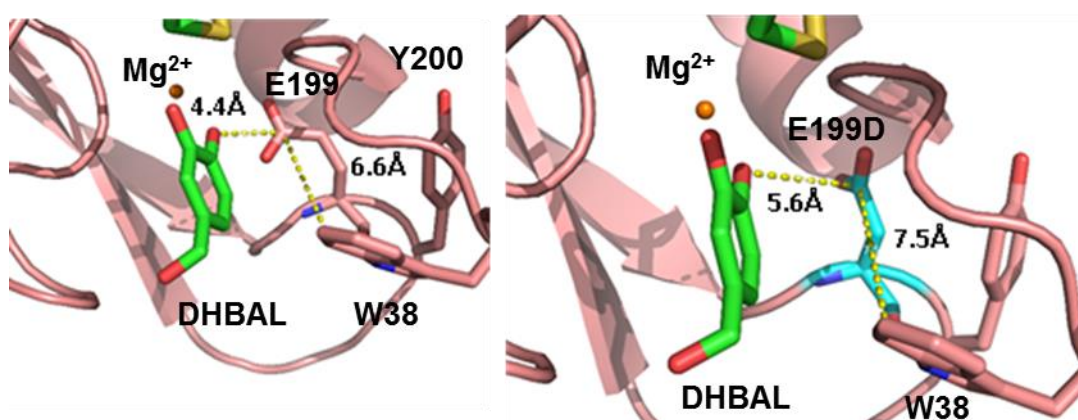
Figure 35: (Left) Adapted from PDB (1VID), the distance between E199 and the DHBAL (**3**) substrate and W38 residue. (Right) Adapted from PDB (1VID), the distance between the modelled E199D and the DHBAL (**3**) substrate and W38 residue.[1]

E199 has been shown to be the pivotal residue in dictating *para* methylation. Oligomeric separation was conducted for the E199 mutant and regioselectivity assays were run as before, with dimeric and monomeric E199L showing very similar regioselectivities. Dimeric E199L showed a regioisomeric excess of 88% whereas monomeric E199L showed a regioisomeric excess of 83%. The result again fitted with the previous result of E199L, which showed a regioisomeric excess of 85% for the mix of the two oligomers. The result shows that the two oligomers of E199L both exhibit *meta* regioselectivity. The regioisomeric excesses are also close with only 5% separating dimeric E199L from monomeric E199L. If E199 controls the *para* methylation, the observed small difference between dimeric E199L and monomeric E199L was to be expected, as the leucine cannot form contacts with the aldehyde whether the leucine faces in or out of the active site. The small difference between E199L monomer and dimer, therefore, did suggest that dimer

regioselectivity was dependent on the E199 residue and the position the residue faces in the COMT active site.

E199 mutants had strongly suggested that the position of the E199 and the 4 position hydrogen bond to E199 controlled regioselectivity of the COMT active site. The Y200L structure overlaid with the dimeric WT COMT also shows that the E199 was flipped out (Figure 31). The two structures were both determined as dimers, therefore, to determine whether Y200L was regioselective, or whether the mutation stabilised the dimer, Y200L was separated into monomeric and dimeric forms. Conversion assays were again conducted for Y200L monomer dimer with all three substrates; DHBAL (**3**), DHBA (**4**) and NO2CAT (**5**). The method of the conversion assays (see Section 6.4.1) were also conducted under the same conditions as those performed for WT COMT. Three substrates were chosen to see if the trends observed across WT COMT were the same for Y200L. Using these conditions, more should be understood about the effect of mutations on the regioselectivity of WT COMT.

	DHBAL	DHBA	NO2CAT	DHBAL	DHBA	NO2CAT
para	5	2	5	5	1	5
#REF!	93	13	94	67	6	61
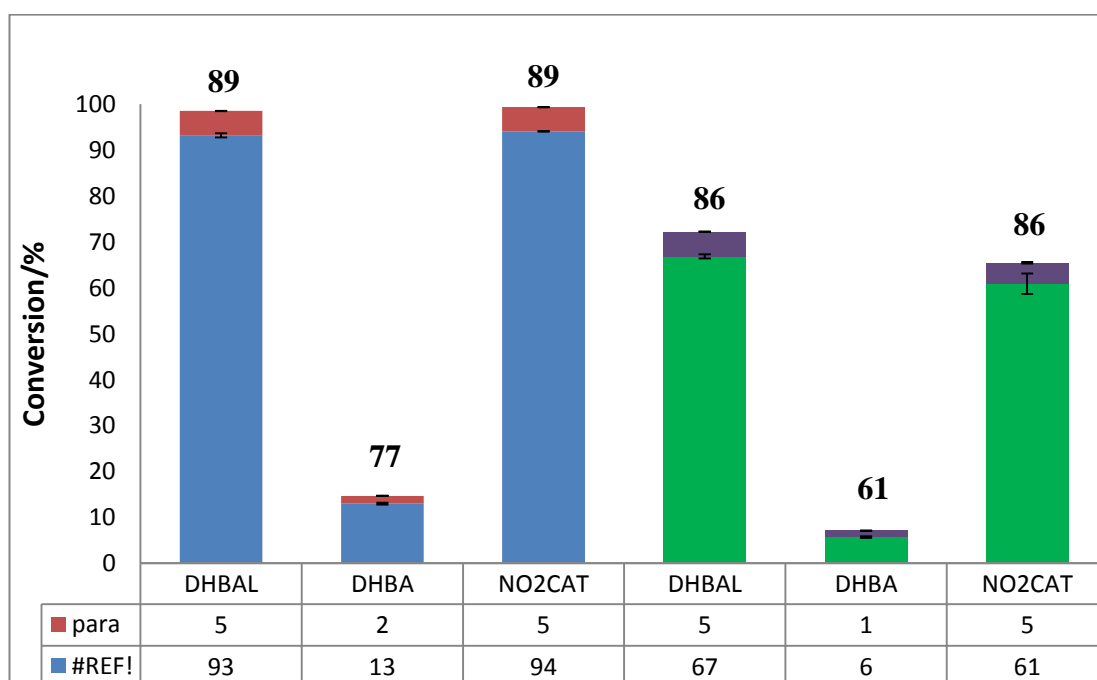
Figure 36: Conversion and regioselectivity assays of monomeric and dimeric Y200L COMT for; DHBAL (**3**), DHBA (**4**) and NO2CAT (**5**) substrates. Dimeric Y200L COMT activities are shown in red (*para* product (**3b**, **4b**, **5b**)) and blue (*meta* product (**3a**, **4a**, **5a**)), monomeric Y200L COMT activities are shown in purple (*para* product (**3b**, **4b**, **5b**)) and green (*meta* product (**3a**, **4a**, **5a**)). Regioisomeric excesses can be seen above the bars.

Results for the Y200L mutant showed that both the monomeric and dimeric Y200L mutant exhibited *meta* regioselectivity (Figure 36). The conversion assays with dimeric Y200L COMT with substrate NO2CAT (**5**) showed the most *meta* methylation. Whereas monomeric Y200L with substrate DHBA (**4**) showed the least *meta* methylation. **4** was also the least active with both dimeric and monomeric Y200L. The low conversion of **4** with the Y200L mutant was also observed in the oligomeric mix by Dr Brian Law. The low conversion of the DHBA (**4**) substrate to vanillic (**4a**) and isovanillic (**4b**) with the Y200L mutant suggested that Y200 has an interaction with **4**. The dimeric Y200L does also seem to show dual site conversion with the DHBA (**4**) substrate, as Table 7 shows.

Table 7: Conversions of monomeric and dimeric Y200L COMT with the DHBAL (**3**), DHBA (**4**) and NO2CAT (**5**) substrates. Conversion for dimeric Y200L COMT are displayed as both observed (overall activity) and conversion per active site.

Substrate	Monomeric Y200L conversion/%	Dimeric Y200L observed conversion/%	Conversion per active site (Y200L dimer)/%
DHBAL (**3**)	72	99	49
DHBA (**4**)	7	15	7
NO2CAT (**5**)	66	99	49

The regioisomeric excesses did show that the monomeric Y200L COMT mutant was *meta* regioselective. The regioselectivity of the monomeric Y200L COMT strongly suggested that the mutation does affect *para* methylation independently of dimerisation. The difference in regioisomeric excesses between monomeric and dimeric Y200L were observed to be small; 3 % DHBAL (**3**), 16 % DHBA (**4**) and 3 % NO2CAT (**5**). The observed small differences between monomeric and dimeric Y200L for **3** and **5** suggested that the Y200L mutation was most likely the cause of the regioselectivity. The lack of difference between monomeric and dimeric Y200L (for **3** and **5**) also indicates that the high regioselectivity observed for Y200L and dimeric WT COMT, was most likely due to the same cause i.e. the E199 residue facing out of the active site. Particularly as both dimeric WT COMT (84 % r.e.) and

Y200L monomer (86% r.e.) exhibit similar regioselectivity. The observation that dimeric WT COMT and Y200L leads to E199 facing out of the active site was also observed in both crystal structures (Figure 22 and Figure 31).

The small differences between monomeric and dimeric Y200L COMT, for the DHBAL (**3**) and NO2CAT (**5**), suggested that the mutation did cause the E199 to face out of the active site without dimerization (which suggested that both Y200L and dimeric COMT force E199 independently). However, once again DHBA (**4**) exhibited different behaviour when compared to the NO2CAT (**5**) and DHBAL (**3**). **4** does present more *meta* methylation for monomeric Y200L (61%) than observed for the most regioselective dimeric WT COMT (59% r.e.) similar to **3** and **5** substrates. The *meta* methylation of DHBA (**4**) did suggest that the Y200L mutation led to less interaction between E199L and **4**, which in turn led to less *para* methylation.

The initial studies of E199 mutations strongly suggested that this residue stabilises *para* methylation. The polarity and the position of the E199 residue was also shown to be important by mutants; E199D and E199L. Separation of the E199L mutant into monomeric and dimeric forms seemed to confirm that the E199 residue's position was also the cause of dimeric regioselectivity. Finally, Y200L oligomeric separation, suggested that the Y200L mutation, was the direct cause of the high *meta* regioselectivity observed for this mutant, rather than simply stabilising the dimeric version of the COMT enzyme.

2.3 Conclusion and Future work

The investigation into COMT regioselectivity produced a number of different regioselectivities, and gave an overall insight into the key residues involved in COMT regioselectivity. The mutation strategy developed novel methods of producing *meta* and *para* methylated regioisomers for catechols. The range of regioselectivity provided by the COMT mutants, coupled with the further understanding of COMT regioselectivity developed in this study, could be used in future for developing industrially relevant compounds such as vanillin (**3a**) and aliskiren. Whilst *para* selectivity has not been achieved above 20%, the development

of the K144 mutants has shown that this strategy was possible. Whereas the *meta* methylation has been very successful, suggesting that near complete *meta* methylation with mutant COMT could be possible. Therefore, the aim of developing regiocomplementary mutants was partially achieved, although, further work was necessary to improve *para* selectivity.

Understanding the *meta* regioselectivity achieved by the mutations at Y200, have identified E199 as key control of COMT regioselectivity. E199 residue facing out of the active site has been attributed to the increased *meta* methylation observed for the Y200L and dimeric WT COMT enzymes, through destabilisation of *para* methylation. The destabilisation was thought to occur through disruption of a hydrogen bond between E199 and the 4-position of polar substrates. The E199 mutations, in turn confirmed this, with the introduction of hydrophobic groups (L) confirming the polarity needed for the hydrogen bond (by showing increase *meta* methylation) and the shorter side group mutation E199D confirming the length of the hydrogen bond. Finally, K144 mutations unequivocally showed that the K144 was the catalytic base, without which, less *meta* methylation occurs for DHBAL (**3**) (K144A and K144E), due to the suspected higher pKa value for the *meta* hydroxyl of **3**.
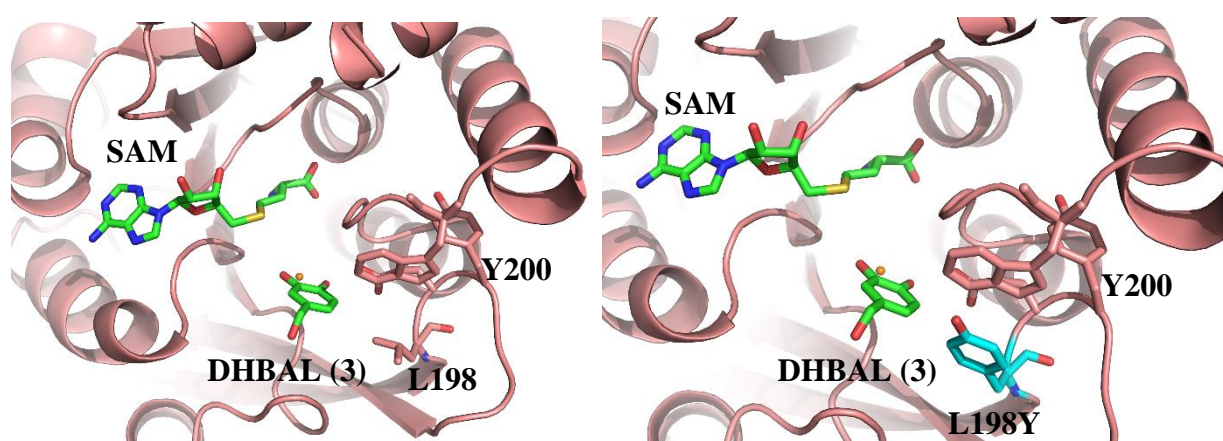
Figure 37: (Right) Adapted from PDB (1VID), WT COMT structure and orientation of W38, Y200 and L198. (Left) Hypothetical mutation of L198Y.[1]

Understanding Y200 regioselectivity through crystal structures has so far indicated the importance of E199 in the regioselectivity of COMT. The crystal structure has

also highlighted other residues that may be able to increase *meta* methylation, when mutated. The best example of these residues is L197, which as shown in Figure 37, is the other side of the E199 residue and is located under the *para* binding hydrophobic pocket. Introducing a bulkier group at this position may prevent substrates binding in the hydrophobic pocket, due to steric clashes. The steric clashes in the active site would, therefore, destabilise the *para* orientation necessary for *para* methylation, increasing the amount of *meta* methylation.

Future mutations could improve the regioselectivity of the COMT enzyme, particularly in favour of *meta* methylation, which as mentioned by Tsao *et al* is most favoured methylation pathway.[99] The data collected so far has suggested that this *meta* methylation is from the position of E199 residue. However, the data does not confirm what the underlying cause of the E199 facing out of the active site. The data has suggested that the Y200L mutation leads to the E199 residue facing out of the active site, observed by the Y200L crystal structure as well as the monomeric Y200L exhibiting *meta* regioselectivity. The E199 facing out may be due to the loss of the rigid aromatic side chain. The leucine would be more flexible, leading to E199 not being locked into the active site. The E199 then may flip out due to more favourable interactions with the solvent. Alternative mutations around the E199 residue may determine whether more flexibility around this loop is the reason for E199's change in orientation. Future investigations should determine how the E199 adopts the "in" or "out" orientation.

Dimeric WT COMT has also been suggested to cause the E199 to face out the active site, through the dimeric WT COMT's repositioning of E199 by domain swapping. The underlying effects of what cause this domain swapping have not been fully explored in this study. There have been many examples of quaternary structure being essential for activity.[118] One of the most famous examples being haemoglobin, which without haemoglobin's quaternary structure significantly loses activity, as shown in investigations with the enzyme's monomeric units.[118] However, there are few examples, that I am aware of, where both activity and regioselectivity can be effected by quaternary structure in this way.

The formation of dimeric COMT could be explained by the presence of the His-Tag at the N terminus. The presence of the His-Tag may make the N-terminus more

flexible, which enables domain swapping. The domain swapping in turn leads to dimerisation. Previous literature,[1,5,119] has not mentioned dimerisation of COMT until Rutherford *et al*.[33,32]The distinct difference between Rutherford *et al* and the previous earlier studies can be attributed to the introduction of vectors containing a His-Tag.[1,33] In order to examine this hypothesis the removal of the His-Tag is necessary. In this study removal of the His-Tag was attempted through initial cloning of COMT into the pET28a vector (which has a thrombin cleavage site before the His-Tag on the C-terminus). Whilst gels indicated the removal of the His-Tag, subsequent scaling issues rendered further studies inconclusive, as to whether the Tag was responsible for dimerization (Figure 38). Future studies should focus on the scaling of this thrombin cleavage step enabling study of whether the His-Tag affects dimerisation. After thrombin cleavage the COMT can be examined by gel filtration to determine whether; firstly COMT dimerisation still occurs and secondly whether domain swapping is still present.
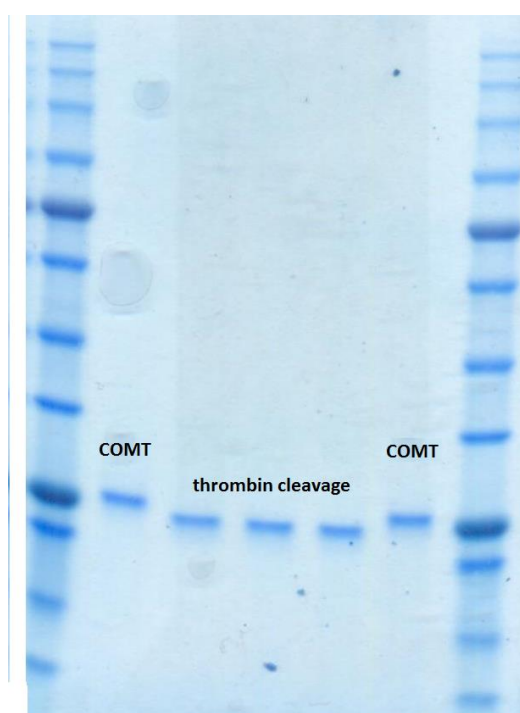
Figure 38: SDS-PAGE gel showing the thrombin cleavage of pET28a COMT. The lanes were shown as follows (from left to right); protein ladder, uncleaved pET28a COMT, cleaved pET28a COMT (1 hr), cleaved pET28a COMT (2 hr), cleaved pET28a COMT (3 hr), uncleaved pET28a COMT and protein ladder.

This investigation has led to a greater understanding of COMT regioselectivity and those residues that are involved; Y200, E199 and K144. The study has also uncovered some novel regiochemistry, for the dimeric WT COMT. The investigation has also determined that the E199 residue was important for the underlying regioselectivity of both the Y200L COMT mutant and dimeric WT COMT. The position that E199 faces was also shown to be important in dictating the regioselectivity of COMT. Dimeric COMT has a greater regioselectivity than the monomeric variant, and through FPLC was shown to be easily separable from the monomer species, meaning a more regioselective COMT biocatalyst can be generated with mutagenesis. Additionally, based on our structures (of dimeric WT COMT and Y200L) and the conversion data shown here, it may be possible to engineer COMT mutants that lead to stabilisation of the dimeric COMT structure.

3 Applications of COMT regioselectivity

3.1 Introduction

3.1.1 Natural incorporation of DOPA into peptides and proteins

The improvement of COMT regioselectivity has shown that the COMT enzyme can be manipulated through mutagenesis to exhibit different behaviour from the WT methyltransferase activity. An alternative field of study with this enzyme would be the expansion of the COMT substrate scope. As mentioned in the Section 1.1.2, COMT has been widely studied for the role the enzyme plays in L-DOPA metabolism.[28,29] L-DOPA, as well as being a key treatment in Parkinson's disease,[28,29] has recently been the object of several studies focussing on the occurrence of the residue in peptides.[120-122]Therefore, as L-DOPA has been shown to be active with the COMT enzyme, investigation of COMT's activity with L-DOPA peptides could be of interest, particularly from a labelling standpoint.

Recent studies have shown that L-DOPA can be translationally incorporated into proteins instead of tyrosine leading to L-DOPA based proteins.[120,122] These L-DOPA proteins are harmful for the body, as they can lead to cross linking, as well as protein misfolding, which could lead to disorders such as atherosclerosis and cataractogenesis.[121,122] Therefore, in order to tackle these problems, which are often prevalent in elderly patients, the quantity and locus of L-DOPA incorporation in proteins must be understood.

There are three main pathways that lead to substitution of tyrosine residues for L-DOPA species; free radical mechanisms, biosynthesis of melanin and unnatural translation incorporation of the L-DOPA by a tyrosyl-tRNA synthetase (Scheme 18).[120] The free radical mechanism has also been termed as the oxidative stress mechanism.[121] The oxidative stress mechanism refers to an imbalance of reactive oxygen species, which caused oxidative reactions leading to the production of L-DOPA residues in the place of tyrosine residues.[120-122] The reactive oxygen species

were thought to occur from a number of sources; leakage from mitochondria, flavoproteins and cytochrome P450 enzymes being just a few.[121] These mechanisms often occur in ageing cells and can lead to the previously mentioned age related diseases atherosclerosis and cataractogenesis.
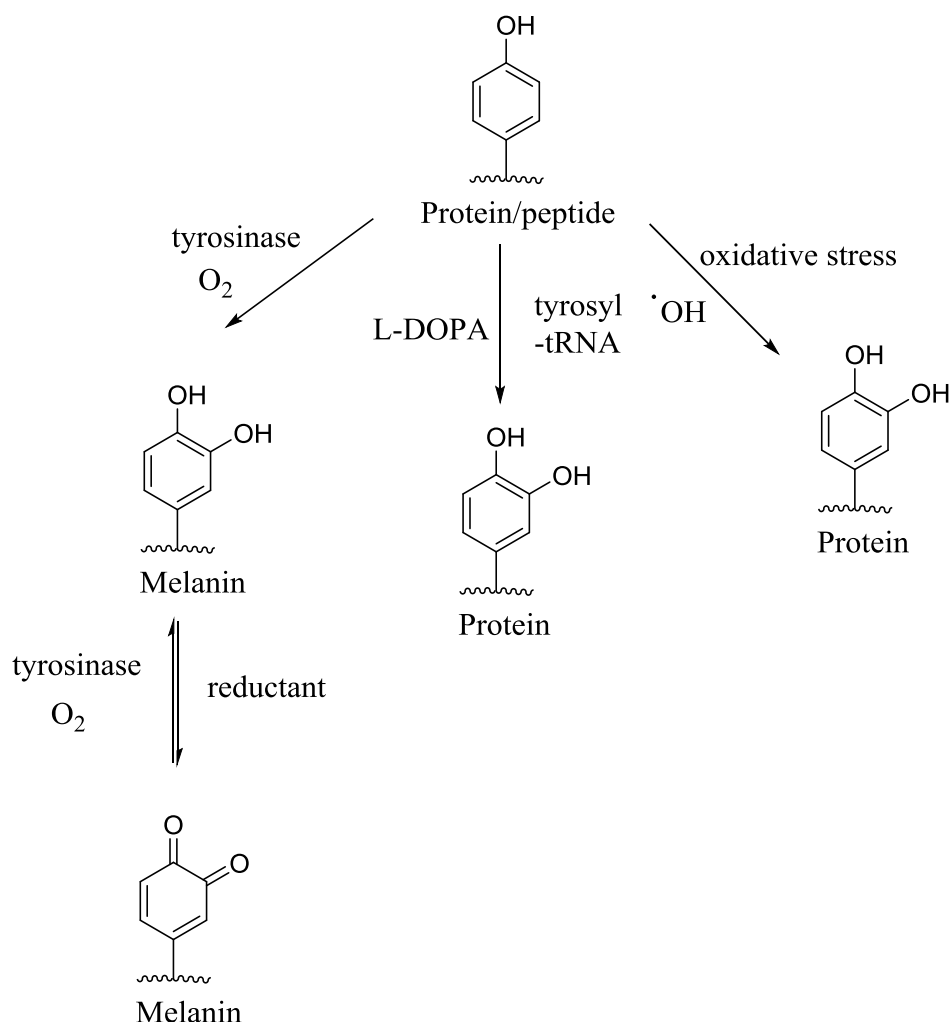
Scheme 18: Various routes for the conversion of tyrosine residues on proteins to DOPA residues on proteins.[121]

Another source that was determined to produce L-DOPA peptides naturally was the tyrosinase enzyme. In mammalian systems tyrosinase was found in melanocytes and has been found to be involved in the production of melanin, a long complex biopolymer, which is used in skin pigmentation.[121] The mechanism by which the tyrosinase hydroxylates phenolic residues, and in turn polymerises them into melanin has often been debated. Currently, L-DOPA production by the tyrosinase enzyme has been suggested to either be formed by direct oxidation of free tyrosine or by reduction of the tyrosinase quinone product.[123,124] Either way the outcome of the

enzyme was found to be the production of L-DOPA residues that are incorporated into melanin. The tyrosinase system has also been found to incorporate L-DOPA residues into proteins.[121] Therefore, there are two natural pathways in mammalian systems by which DOPA bound proteins are generated.

3.1.2   Methods of incorporating DOPA into peptides and proteins

The study of DOPA peptides has increased due their relation to age related disorders. Therefore, as well as the natural methods of production of DOPA peptides mentioned in the previous section, there have been recently a number of ways to produce DOPA peptides through unnatural and *in vitro* means.

DOPA residues have been introduced into proteins through translational incorporation of DOPA into proteins (Scheme 18).[122] Ozawa *et al* have shown that L-DOPA can be incorporated into proteins by using their *E.coli* cell free transcription/translation system.[122] This incorporation was postulated as to have occurred through the tyrosyl-tRNA accepting L-DOPA residues, as well as tyrosine residues.[122] However, the K_m for L-DOPA binding was roughly 200 times higher (1.4 mM) than tyrosine (6 μM) for binding with the tyrosyl-tRNA.[122] The significantly higher K_m for L-DOPA binding suggested that for L-DOPA, formed from free tyrosine residues in the body by oxidative stress mechanisms, L-DOPA incorporation via tyrosyl-tRNA would be low. This would be due to the L-DOPA concentration (formed from the oxidative stress mechanisms) may not be sufficient to be incorporated by the tyrosyl-tRNA. However, for Parkinson's sufferers the larger concentrations of L-DOPA may overcome the high K_m thus leading to translational incorporation. Therefore, the translational incorporation of L-DOPA into proteins may be low under natural conditions, but with L-DOPA supplements incorporation of L-DOPA by tyrosyl-tRNA may lead to L-DOPA residue formation in peptides and proteins.

Methodology has also been developed to transform already incorporated tyrosine residues to L-DOPA residues using the tyrosinase enzyme (mentioned in Section 3.1.1).[125] Marumo and Waite studied the transformation of tyrosine residues, with a mushroom tyrosinase enzyme and an ascorbic acid inhibitor, to L-DOPA residues in

a number of tyrosine containing peptides.[125] The ascorbic acid was introduced to reduce any quinones produced back to the L-DOPA form (Scheme 19).[125] The method was shown to work with a variety of peptides with sequences ranging from dipeptides to decapeptides.[125] Ito *et al* demonstrated that the mushroom tyrosinase could also hydroxylate, and in turn form quinone residues which crosslinked with surface cysteines, three proteins; bovine insulin, bovine serum albumin and alcohol dehydrogenase.[126] Therefore, *in vitro* hydroxylation was shown to be possible by two groups on both peptide and protein substrates.[125,126]
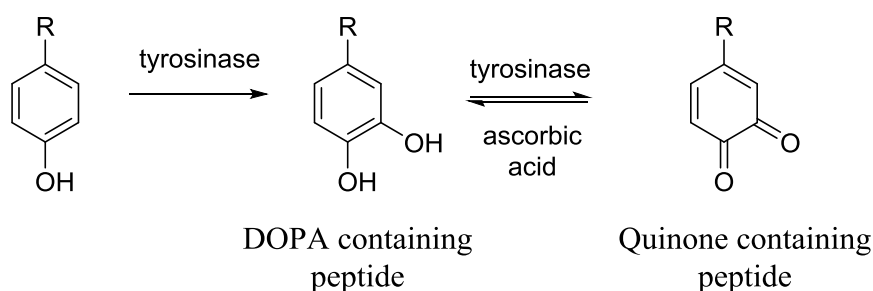
Scheme 19: Hydroxylation of tyrosine containing peptides and the reduction of the quinone with ascorbic acid.

The incorporation of L-DOPA residues has been shown to occur both in both *in vivo* and *in vitro* settings. However, despite *in vitro* hydroxylation being shown to be possible,[125] as yet no studies *in vitro* alkoxylations have been conducted. Coupling an O-alkyltranferase such as COMT with the tyrosinase system could result in a labelling system for L-DOPA-peptides. Therefore, based on the existing studies, mentioned in this section, an alkoxylation system should be devised with COMT as the trial enzyme.

3.2   Results and discussion

3.2.1   WT COMT with synthetic SAM analogues

The expansion of COMT's substrate scope would also be improved by the expansion of COMT's cofactors. The main aim would be to determine how far COMT could be described as an "alkyltransferase" rather than a methyltransferase. The value of alkylation over methylation was clearly apparent. Alkylation could be used as a

"chemical handle" enabling further transformations. Alternatively, alkylation offered different chemistry to methylation and could be used to alter properties of APIs and pharmaceuticals. Developing an alkylation system with the COMT enzyme will lead to a diversity of catecholic products. In order to develop these catecholic products, SAM analogues would first be derived synthetically.

S-adenosyl-allylhomocysteine (**1c**)

S-adenosyl-benzylhomocysteine (**1d**)

S-adenosyl-methylacetatehomocysteine (**1f**)
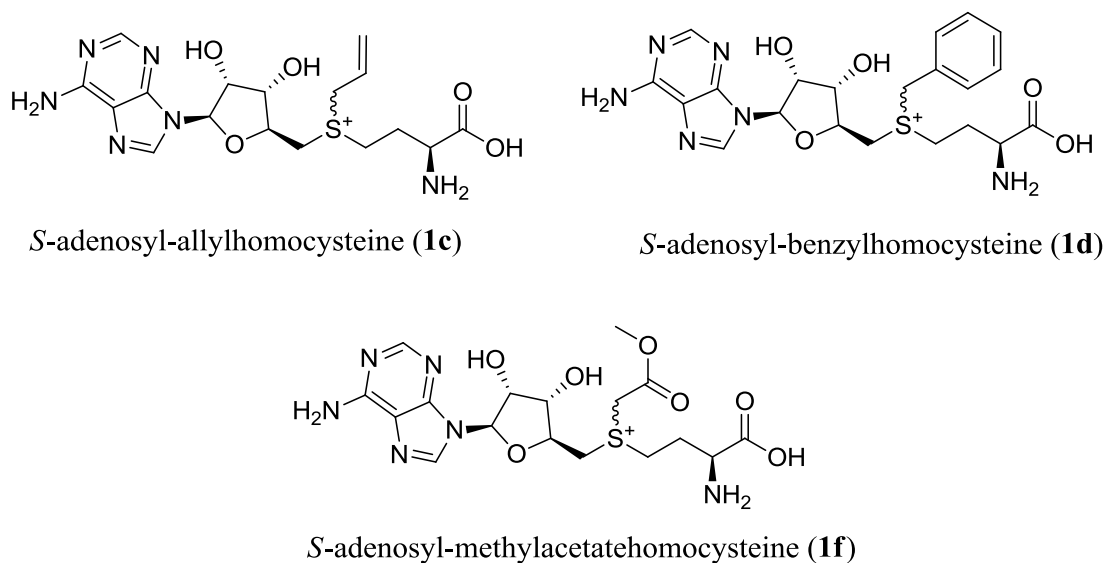
Figure 39: SAM analogues; SAAH (**1c**), SABH (**1d**) and SAMH (**1f**).

SAM analogues were first synthesised by using a similar approach to Dalhoff *et al* (Section 6.7) three analogues were synthesised; SAAH (**1c**), *S*-adenosyl-methylacetatehomocysteine (SAMH) (**1f**) and SABH (Figure 39) (**1d**).[4] These analogues were firstly chosen due to a sp^2 hybridised system connected to the methylene attached to the sulfonium ion, to stabilise the S_N2 alkyl transfer (see Scheme 7). Secondly, they were chosen as the analogues showed a range of sizes and polarities. **1f** was interesting as transfer of the methyl acetate moiety would be an umpolung reaction, as the latent polarity of the methylene on the methyl acetate moiety should be negative. Whereas the **1d** analogue may offer an interesting route into regioselective protection.

Table 8: Yields and activity with WT COMT of the three synthetic analogues; SAMH (**1f**), SAAH (**1c**) and SABH (**1d**).

SAM analogue	Yield/%	Active with COMT
SAMH (**1f**)	5	No
SAAH (**1c**)	91	Yes
SABH (**1d**)	89	Yes

The analogues were isolated in good yields for SABH (**1d**) and SAAH (**1c**) but not SAMH (**1f**) (Table 8). Analogues were then characterised by NMR, HPLC and MALDI (Section 6.7). All analogues were tested with the DHBAL (**3**), as this substrate was the most well studied of the COMT substrates (Section 6.4.6). The SAM analogues were used over 2.5 times the concentration of **3**. The large excess of SAM analogue, was due to only 50% of the synthetic analogue being active with the COMT. Only 50% of the synthetic sample was thought to be active, as the (S_c R_s) SAM (**1a**) diastereomer has been shown to be not active with methyltransferases,[72,79] and the synthetic alkylation of SAH (**1**) was not carried out with diastereomeric control. The resulting samples were then run on the HPLC with controls without enzyme and cofactors. **1f** was tested with COMT but to date, no transfer of the methyl acetate has been noted. **1d** and **1c** both showed two new peaks, one of which was identified, as SAH (**1**) (Figure 40). However, under the initial HPLC conditions no regioisomers were observed.
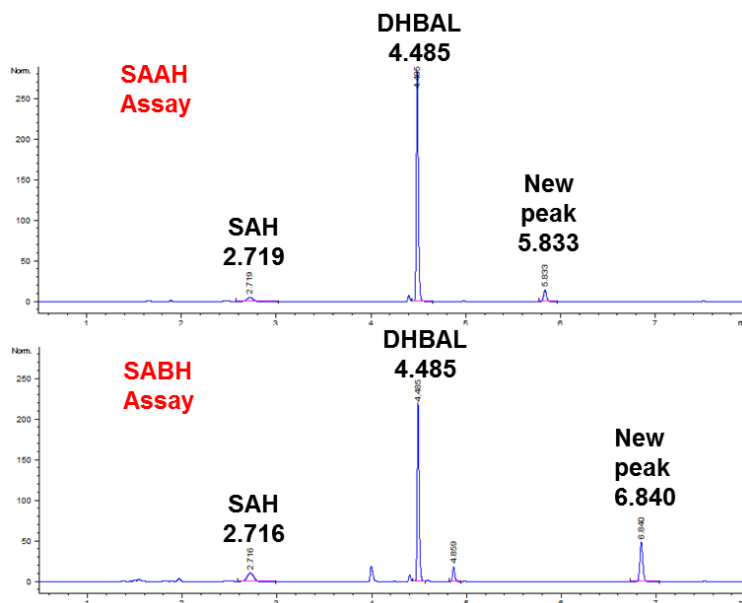
Figure 40: (Top) HPLC trace of WT COMT assay with SAAH (**1c**) analogue. (Bottom) HPLC trace of WT COMT assay with SABH (**1d**) analogue.

In order to confirm that the analogues were active with COMT and to determine whether both regioisomers were present with the SAM analogues; standards were synthesised. Adapting a method from Reitz *et al*, both regioisomers of the allylated DHBAL (3-alloxy-4-hydroxybenzaldehyde (**3e**) and 4-alloxy-3-hydroxybenzaldehyde (**3f**)) were prepared (Scheme 20).[127,128] The synthetic approach showed the value of developing regioselective alkylation, as both methods to the two respective regioisomers led to considerable side products. Scheme 20 shows that to synthesise the **3f**, a weak base was used in potassium carbonate, as the *para* hydroxyl was thought to have lower pKa and would lead to more deprotonation at this position than the *meta* position. The **3f** regioisomer was isolated in 28% yield with **3e** and di allylated side products (22% combined).
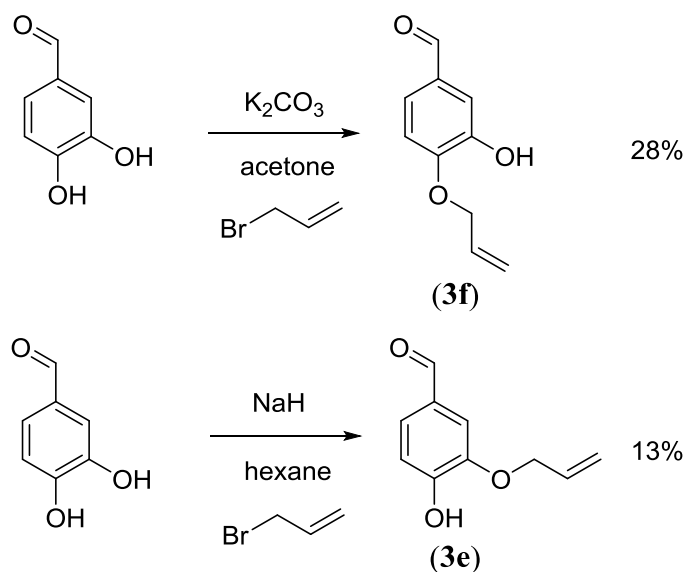
Scheme 20: The two synthetic strategies for *meta* (**3e**) and *para* (**3f**) allylation.[127,128]

The *meta* allylation (**3e**) was carried out with a much stronger sodium hydride base, to remove both protons, which would give an even bigger mix of the three allylated products. Using this method, even less of the *meta* allylated DHBAL (**3e**) regioisomer was isolated with a yield of 13%. The methods were adapted for benzylated DHBAL for both regioisomers (*para* 4-benzyloxy-3-hydroxybenzaldehyde (**3h**) and *meta* (3-benzyloxy-4-hydroxybenzaldehyde (**3g**)) (Section 6.9).The yields of the benzylation reaction were even lower with **3h** isolated in 7% yield whilst **3g** was isolated in 1% yield. Both methods also required long columns and a large amount of solvent to complete separation.

The synthesis of the four regioisomers for the benzylation and allylation reactions meant that they could now be run on the HPLC, for comparison with the assays. As Figure 41 shows, the *para* allyl standard (**3f**) showed the same retention time as the new peak. Figure 42 shows, the same for the benzyl standard (**3h**). Spiking the SAAH (**1c**) assay (Figure 41) with the **3f** standard, showed an increase in the new peak (area changed by 10% proportional to the amount of **3f** added). The SABH (**1d**) assay was also spiked in the same way, with a similar outcome (Figure 42). The growth of the peak, through spiking both assays, and the same retention times for the synthetic standards as the new peaks in the assay suggested that the COMT was active with the SAM analogues.

The assays were then repeated in triplicate with SAM (**1a**), SABH (**1d**) and SAAH (**1c**) cofactors, to determine the relative activities of the three analogues. Assays were terminated at 25 minutes to prevent full conversion of the **1a** assay (time determined from k_{cat}=4.00 µmol min^{-1} by Dr Brian Law). The result unsurprisingly showed that COMT was the most active with the **1a** cofactor. However, the most active cofactor was **1d**. Table 9 shows that use of the **1d** cofactor results in significantly higher conversion of DHBAL (**3**) to *meta* (**3g**) and *para* (**3h**) benzylated regioisomers (26%) compared to; **1c** conversion of DHBAL to *meta* (**3e**) and *para* (**3f**) allylated regioisomers (7%). These preliminary results suggested that the **1d** analogue was more active with COMT than the **1c** analogue.
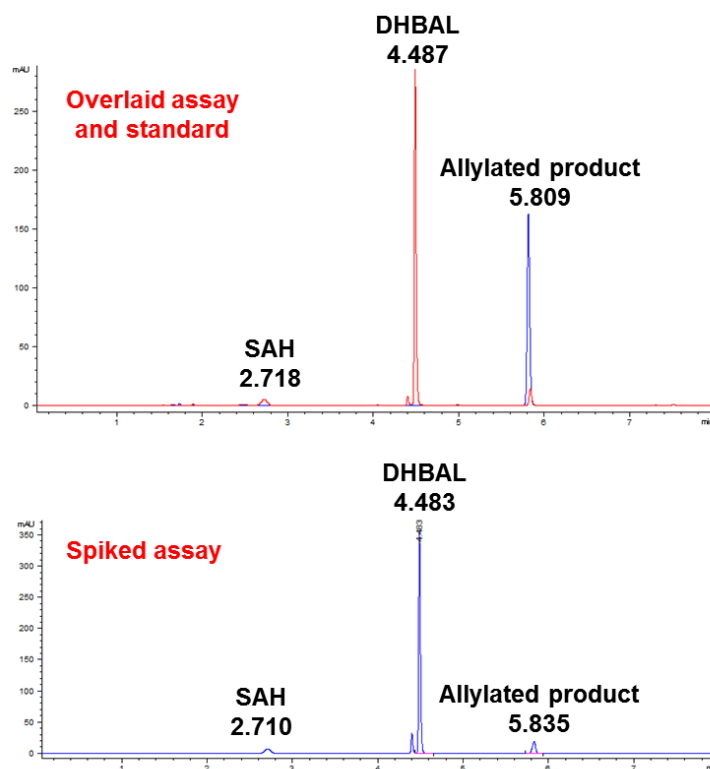
Figure 41: (Top) Overlay of HPLC traces for the SAAH (**1c**) assay ((red) and the *para* allylated DHBAL standard (**3f**). (Bottom) SABH assay spiked with **3f**.
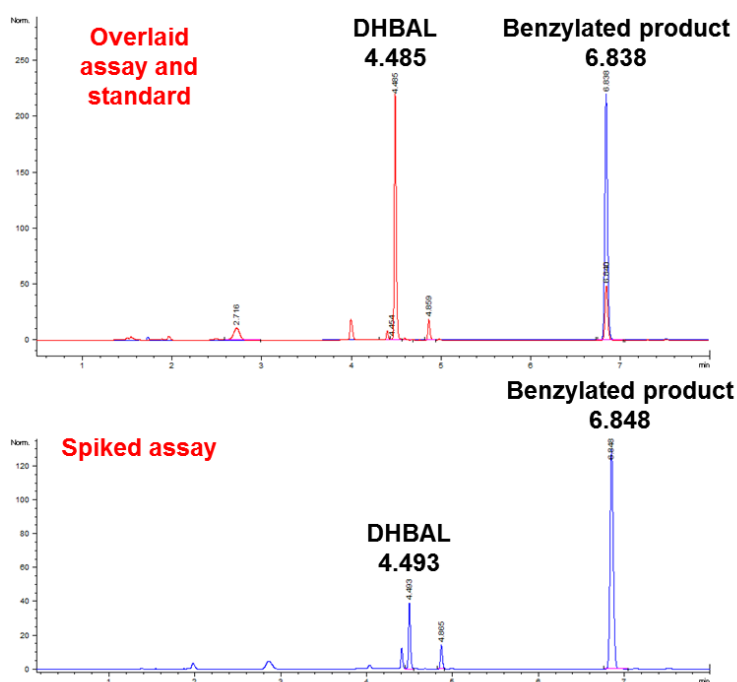
Figure 42: (Top) Overlay of HPLC traces for the SABH assay (**1d**) (red) and the *para* benzylated DHBAL standard (**3h**). (Bottom) **1d** assay spiked with **3h**.

The higher conversion of DHBAL(**3**) to products 3-benzyloxy-4-hydroxybenzaldehyde (**3g**) and 4-benzyloxy-3-hydroxybenzaldehyde (**3h**) observed for the SABH analogue (**1d**) with COMT enzyme (Table 9) was most likely attributable to the benzyl group offering more stabilisation of the S_N2 transistion state during alkylation. As previously stated in Section 1.2.2 an adjacent sp^2 to the carbon attached to the sulphonium centre stabilises the p orbital, which was suggested to be the orbital by which a lone pair attacks.[4] A benzyl group has six sp^2 hybridised carbon atoms to stabilise the centre whilst the allyl only has two sp^2 hybridised carbon atoms. Therefore, the higher number of sp^2 hybridised carbon atoms present in the benzyl group of **1d** may stabilise alkyl transfer more than the SAAH (**1c**) analogue, which in turn may be attributable the **1d** analogue having a higher conversion to products with COMT.

However, despite the initial determination of COMT conversion of DHBAL (**3**) with the SABH (**1d**) and SAAH (**1c**) analogues, there was no separation of alkylated regioisomers. The lack of separation of alkylated regioisomers proposed that under current HPLC conditions, both sets of regioisomers had the same retention time. This accounted for the observation that only one new peak was noted for the

alkylation of DHBAL (**3**) (disregarding SAH (**1**)). The HPLC had shown that the **1c** and **1d** analogue assays were successful with the COMT enzyme, however, a HPLC method was needed to separate the alkylated analogues (**3e-h**).

Table 9: Activity of the two SAM analogues; SABH (**1d**) and SAAH (**1c**) with COMT compared to SAM (**1a**).

Analogue	Conversion/%
SAM (**1a**)	77
SABH (**1d**)	26
SAAH (**1c**)	7

Separation of the regioisomers of benzylated and allylated DHBAL (**3e-h**) needed to find a chemical distinction between *meta* and *para* alkylated substrates. Previously, it has been noted that the *meta* and *para* hydroxyls were thought to have different pKa values, this was also observed during the synthesis of the alkylated standards. Based on this knowledge, the best way to separate the alkylated DHBAL would be to use a buffered system that approaches the pKa of the *para* hydroxyl, as an anionic phenoxide should have a distinct polarity from a hydroxyl group. The buffered system was trialled with 10 mM potassium phosphate buffer pH 7.2. Figure 43 shows that, after method optimisation (final conditions Section 6.5.3), *meta* and *para* regioisomers of the benzylation reactions could be separated (similar separations were also observed for allylated regioisomers see Figure 131).

Successful separation of the alkylated regioisomers meant that the regioselectivity and conversion of the COMT system could now be tested. Initially the WT enzyme was assayed under the conditions set out in Section 6.4.6. The results showed the SABH (**1d**) analogue displayed the lowest regioselectivity with WT COMT (33% r.e.) compared to SAAH (**1c**) (45% r.e.). The assays showed significant lower regioselectivity with the **1d** analogue than that observed for the WT with SAM (**1a**). Whereas the SAAH analogue showed higher but similar regioselectivity to SAM

(**1a**). The assays confirmed that WT COMT with **1c** and **1d** analogues are not regioselective.
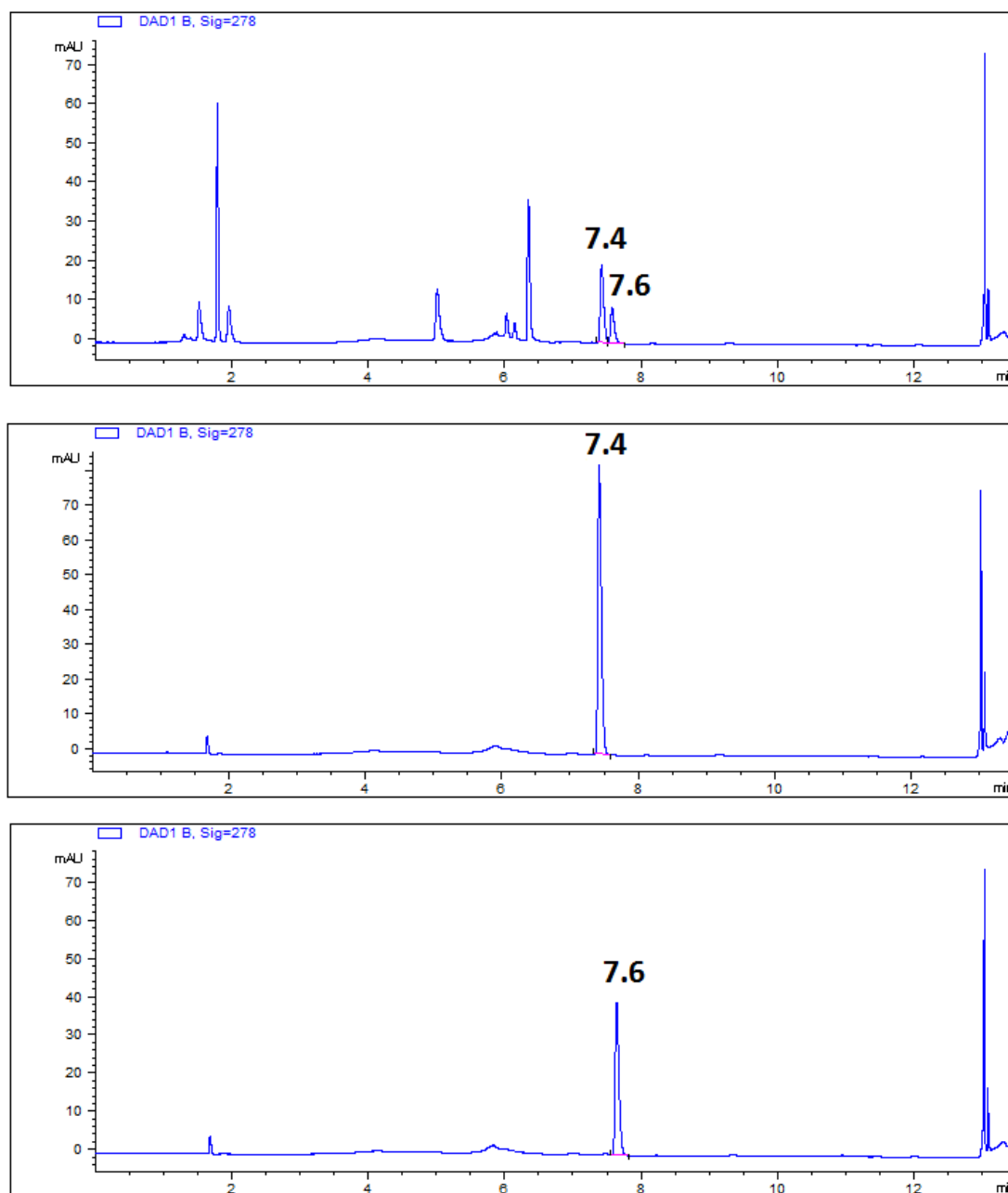
Figure 43: RP HPLC separation of benzylated DHBAL regioisomers; (Top) WT COMT (45 µM) assay with SABH analogue (5 hours at 37˚C), (Middle) 1 mM 3-benzyloxy-4-hydroxybenzaldehyde standard and (Bottom) 1 mM 4-benzyloxy-3-hydroxybenzaldehyde standard.

3.2.2   Regioselective alkylation of DHBAL with COMT mutants

The advantages of regioselective alkylation, as opposed to regioselective methylation, could include regioselective protection methods. Isolation of the different alkylated regioisomers could offer different chemistries, as well as different routes to APIs. In order to investigate the regioselectivity of COMT with the SAAH (**1c**) and SABH (**1d**) analogues, the following COMT mutants were chosen; K144E and Y200L to be assayed with the analogues. K144E and Y200L were chosen, as these mutants showed the extremes of the regioselectivity in the study with methylation. Therefore, these mutants were used to determine whether regioselective alkylation was possible with the SABH and SAAH cofactors.

The very low conversion of DHBAL (**3**) to vanillin (**3a**) and isovanillin (**3b**) products with K144E COMT mutant andSAM (**1a**) (as demonstrated in Section 2.2.2), meant that in order to compare the conversion and regioselectivity of the mutants with the SAAH (**1c**) and SABH (**1d**), high concentrations of COMT enzyme and a long time period was needed. After initial investigations with the K144E COMT mutant, 3 hours was set upon with a concentration of 45 μM for the COMT enzyme (both WT and mutants). **1c** and **1d** were again used at double the concentration of the substrate due to 50% of the cofactor was thought to be an inactive diastereomer (full conditions Section 6.4.6).

Table 10: Conversions and regioselectivities of mutants with WT COMT and the COMT mutants; Y200L and K144E.

Mutant	SABH (1d) conversion	SABH (1d) R.E. %	SAAH (1c) conversion	SAAH (1c) R.E. %
WT COMT	100	33	94	45
Y200L	100	87	89	87
K144E	3	-36	1	-30

Table 10 displays the conversions and regioselectivities of the mutants with the SAAH (**1c**) and SABH (**1d**) cofactors. Y200L again displayed excellent *meta* regioselectivity with very similar r.e.s for the two cofactors. The results of these high r.e.s for both **1c** and **1d** with the Y200L mutant meant that regioselective alkylation was successful and could be used for further applications (such as regioselective protection of catechols). The trend of regioselectivity with the *meta* alkylating mutants was determined to be SAM (**1a**)>SAAH (**1c**)>SABH (**1d**), which suggested that the *para* alkylating orientation may have stabilised the bulkier analogues more than the *meta* alkylating orientation. The suggestion that the bulkier analogues were more stabilised by the *para* alkylating orientation was also reflected by the K144E mutation which showed **1d**>**1c**>**1a** for *para* regioselectivity. The *para* alkylating orientation stabilising the bulkier analogues may account for why **1d** displays lower regioselectivity than **1a** and **1c**.

The conversions displayed by the mutants and the SAAH (**1a**) and SABH (**1d**) cofactor followed a similar trend, as with the SAM cofactor WT>Y200L>>K144E. **1d** conversion of DHBAL (**3**) substrate was again greater than **1a** conversion with **3** and the COMT mutants, again suggesting that benzylation occurs more readily than allylation in the COMT enzyme. However, conversion with both cofactors (SAAH (**1c**) and **1d**) was much lower than with **1a** (reflected in the higher amount of COMT needed and the longer time period) suggesting that inhibition may be occurring with the analogues, or the alkylations are much slower than methylations.

Previous studies have suggested that the (S_cR_s) diastereomer of SAM (**1a**) leads to inhibition of the COMT enzyme. The synthesis of the SAM analogues led to a 50:50 diastereomeric mix, as neither diastereomer was favoured by chemical synthesis, as both angles of attack are as statistically likely to occur in the S_N2 mechanism can occur from the sulphur of SAH (**1**) (this was also observed experimentally .[79] In order to determine whether only one diastereomer was taken by the COMT enzyme, the SAAH (**1c**) analogue was first run on a HPLC method that separates diastereomers and then rerun on the same method after an assay.[79] As Figure 40 shows, only one peak reduces in size after the assay (Peak 2), which suggested that the peak that does not reduce in size was the (S_cR_s) **1c** diastereomer (Peak 1). Secondly, the HPLC that was reduced by the COMT assay (Peak 2) was the only

peak noted in the MAT assay with SAlH, which has been shown in numerous studies to only produce the (S_cS_s) diastereomer (Figure 44).[79] Therefore, the HPLC diastereomeric separation suggested that the (S_cR_s) was not used by the COMT enzyme and maybe inhibition does with occur with this isomer.

The monitoring of synthetic SAM analogues with DHBAL (**3**) substrate had suggested that only one diastereomer had been taken. Therefore, although regioselective alkylation has been shown with synthetic SAM analogues, only 50% of the synthetic SAM analogue has been taken. Lack of conversion of the synthetic SAM analogues suggested that the unnatural diastereomer (S_cR_s) may inhibit the COMT enzyme. Therefore, further work was needed to overcome the diastereomeric contaminants.
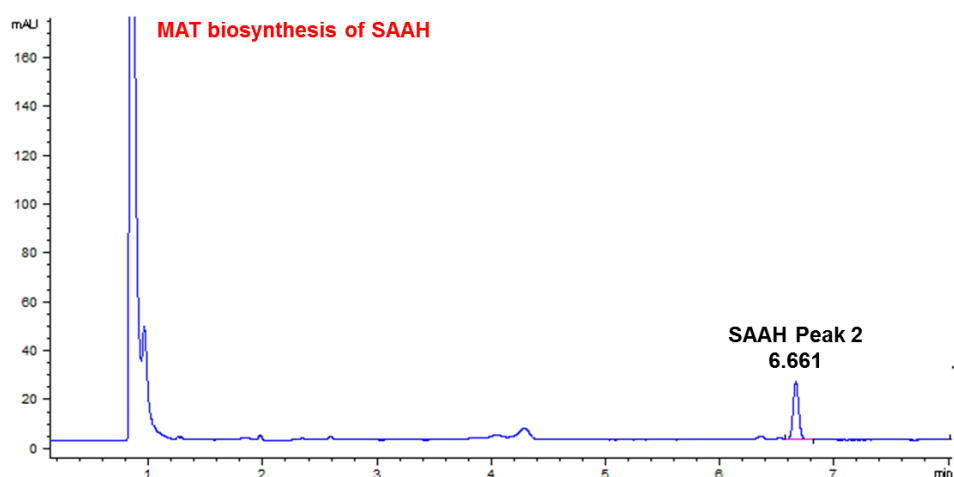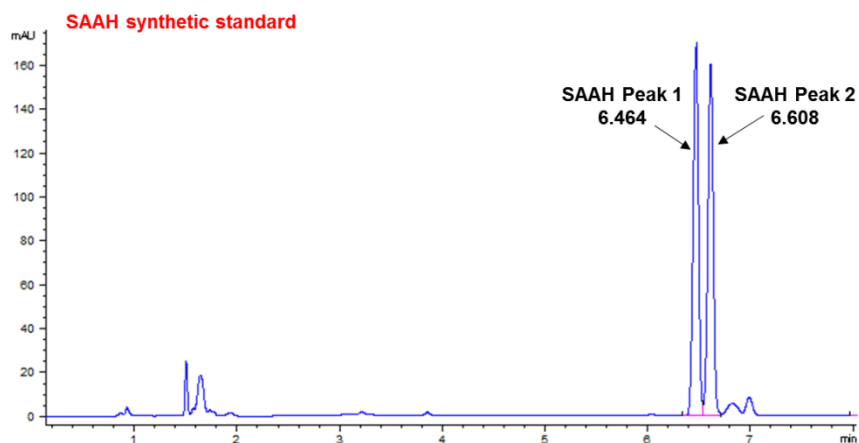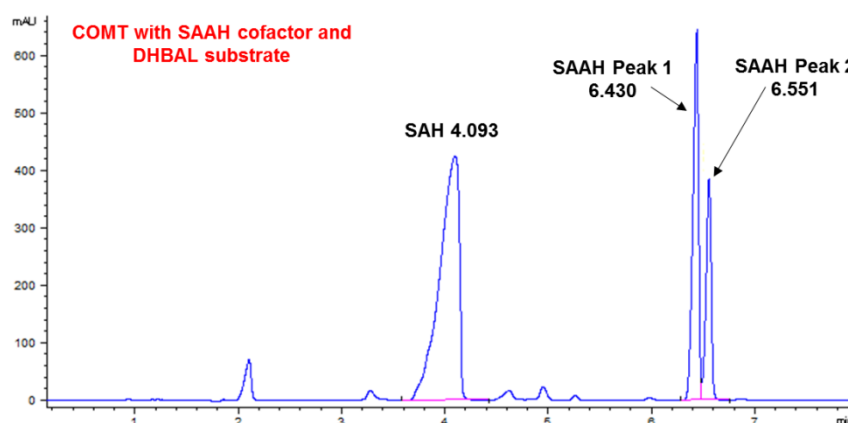
Figure 44: (Top) Synthetic sample of SAAH (**1c**). (Middle) MAT biosynthesis of **1c** with ATP and SAlH (**2c**) (ATP also visible at 1min). (Bottom) COMT assay with **1c** cofactor clear reduction of Peak 2 and not Peak 1 was noted. Peak 1 was thus suggested to be the (S_cR_s) **1c** diastereomer, whilst Peak 2 was suggested to be the (S_cS_s) **1c** diastereomer.

3.2.3   *In situ* preparation of SAM analogues

The synthesis of the synthetic analogues SAAH (**1c**) and SABH (**1d**) had shown that regioselective alkylation was possible with the COMT mutants. However, the results and literature suggested that there was a flaw in using synthetic analogues in that they are not diastereomerically pure. Literature studies, have shown that diastereomerically pure SAM analogues can be generated *in situ* using the methionine adenosyl transferase (MAT) enzyme and the appropriate homocysteine analogue (**2a-d**), followed by a suitable methyltransferase (see Section 1.2.3 and Scheme 6).[8,59] Therefore, to solve the problem of diastereomeric impurities with COMT, the MAT enzyme will be used in coupled assay approach with COMT.

In order to synthesise the SABH (**1d**) and SAAH (**1c**) *in situ* with the MAT enzyme, firstly methionine analogues were needed to be synthesised. Scheme 21 shows that using the literature methods both *S*-benzylhomocysteine (SBH) (**2d**) and *S*-allylhomocysteine (SAlH) (**2c**) were synthesised in 58% and 45% yields respectively (see Section 6.8).[81,129] The successful synthesis of the methionine analogues meant that **2c** and **2d** could be trialled with the MAT enzyme. The hMAT2A enzyme was chosen as literature studies had shown that this enzyme was promiscuous with methionine analogues.[8] The hMAT2A was then mutated to the I322V mutant using a plasmid donated by Dr Professor Udo Oppermann, as this had been shown to be the most active MAT2A mutant. I322V hMAT2A was cloned, expressed and purified by Dr Brian Law (SDS PAGE gel confirming isolation can be observed in Section

7.1.2). The methionine analogues were subsequently tested with the hMAT2A I332V mutant with the conditions set out in Section 6.4.2. Initially SAM analogue formation was tested, using HILIC HPLC and MALDI techniques. **2c** was determined to be active (in line with literature studies) showing the expected 425 mass on MALDI whilst the **2d** analogue did not show the mass expected of 475. This suggested that the **2c** analogue was active with the MAT enzyme but the **2d** analogue was not, which has also been observed in literature.[8]

S-allyl-homocysteine (**2c**)
45% yield

S-benzyl-homocysteine (**2d**)
58% yield
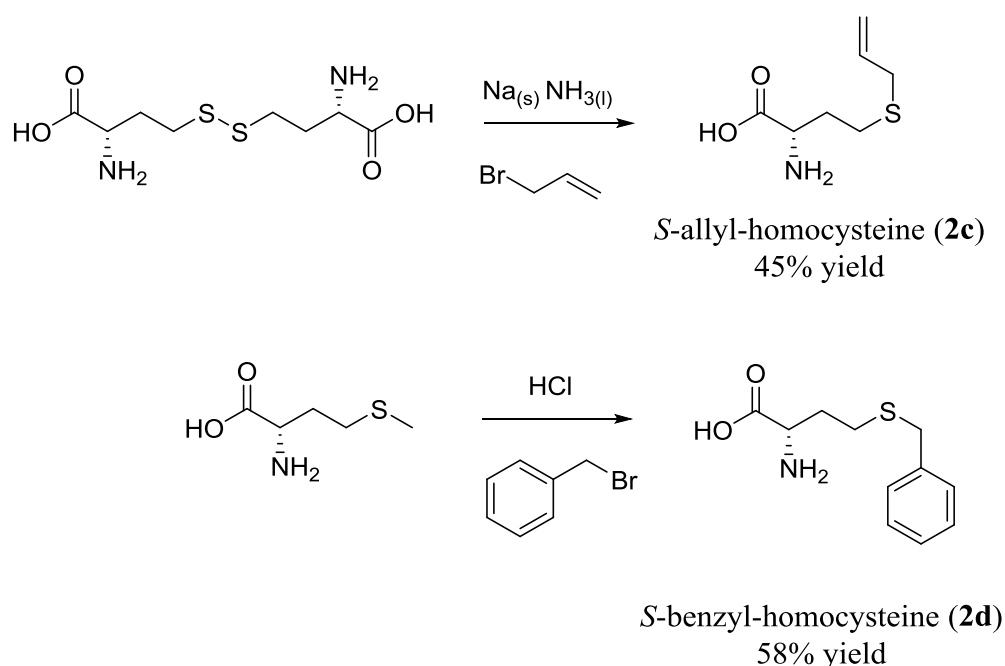
Scheme 21: Synthesis of methionine analogues; SAlH (**2c**) and SBH (**2d**).[81,129]

The conversion of SAlH with hMAT2A I322V suggested that a coupled alkylation was possible. The success of the homocysteine analogue, also suggested that commercial methionine analogues could be trialled. The commercial analogue selected was ethionine. This analogue was chosen to test the theory suggested by Dalhoff that a lack of an adjacent sp^2 system leads to less alkylation. In order to construct a coupled system, an understanding of conversion with the hMAT2A with the methionine analogues was necessary.

The conversion of methionine analogues (**2b** and **2c**) with hMAT2A I322V to SAM analogues (**1b** and **1c**) where first necessary to determine before the coupled hMAT2A-COMT system could trialled. The assays were conducted in duplicate for the three selected methionine analogues (SAlH (**2c**), ethionine (**2b**) and methionine

(**2a**)) at 1 mM concentration with 15 μM hMAT2A I322V over a period of 1 and 2 hours. The assays were terminated at these two time-points with methanol and product was determined using a HILIC column (overlaying synthetic SAAH (**1c**) for **2c** and commercial SAM (**1a**) for **2a**). Calibrations were performed with SAM (**1a**), to equate the area of these peaks,[87] to a concentration (after the determination that SAEH (**1b**) and SAAH (**1c**) analogues do not absorb too differently at 260nm). Figure 45 shows that, as expected, the natural substrate **2a** has the largest conversion with 53% being reached after 2 hours. The next highest conversion was **2b** (45% after 2 hours) and a substantially lower conversion was noted for **2c** (19% after 2 hours). The assays did not show much conversion difference between 1 hour and 2 hours, suggesting that after 1 hour, maximum conversion was almost achieved. The results were also in good agreement with Singh *et al* (Table 2).[8]

	Met (2a)	Eth (2b)	SAlH (2c)	Met (2a)	Eth (2b)	SAlH (2c)
Concentration	0.44	0.36	0.17	0.53	0.45	0.19
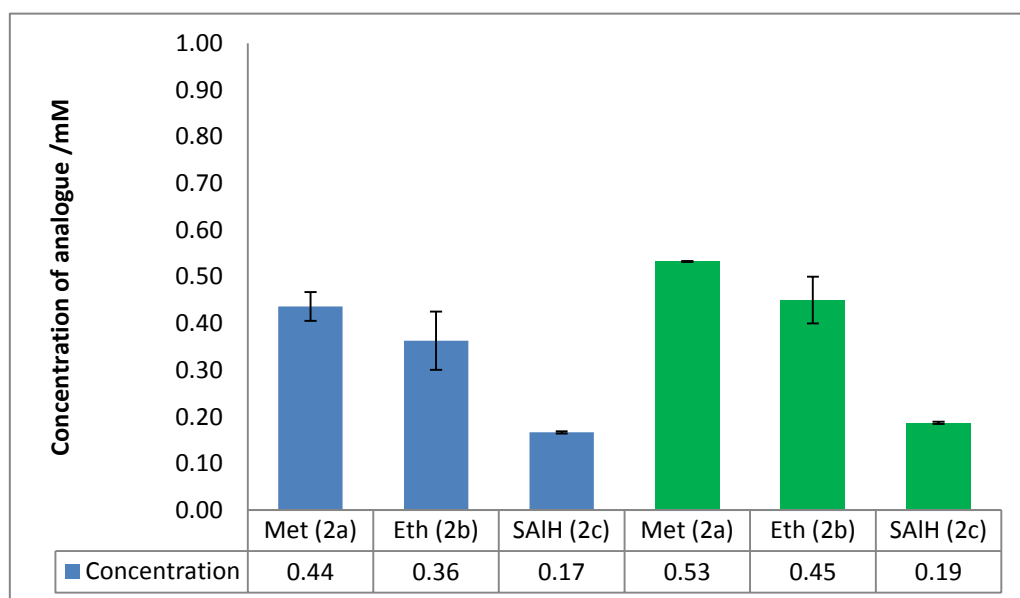
Figure 45: Graph showing the conversions of the three methionine analogues (**2a**-**c**) to SAM analogues (**1a**-**c**) at 1 hour (blue) and 2 hours (green) with hMAT2A I322V.

The hMAT2A I322V conversion assays had shown that maximum conversion could nearly be reached after 1 hour. In order to determine the diastereomerically pure SAM analogues activity with COMT, a conversion assay was set up. In short, the hMAT2A I322V assay was run, after 1 hour the assay would be passed through a centricon to separate the hMAT2A from the homocysteine (**2a**-**c**) analogue and the SAM analogue product. The separation was necessary to determine the exact concentration of the analogue, otherwise hMAT2A would continue to produce more

analogue. A sample of the mix was then run on the HILIC column to determine concentration.

Once the concentration of the diastereomerically pure analogue (SAAH (**1c**) , SAEH (**1b**) or SAM (**1a**)) was determined, then the conditions displayed in Section 6.4.1 were used to determine COMT conversion of the DHBAL (**3**) substrate to the alkylated products (**3a-f**), with assays terminated after 1 hour to prevent methylation (SAM cofactor) going to completion. Figure 46 shows the conversions of diastereomerically pure **1a**, **1b** and **1c**, as well as the synthetic analogues **1d** and **1c**, with COMT. Once again, the diastereomerically pure SAM shows the highest conversion (equal to the value shown earlier). Surprisingly the analogue that led to the most conversion was still the synthetic **1d** (26% conversion) followed by; enzymatic **1c** (21% conversion), synthetic **1c** (7% conversion) and enzymatic **1b** (1%). The assay results did show that that diastereomerically pure analogue (SAAH (**1c**) 21%) showed more conversion to products (**3e** and **3f**) than the synthetically produced mix (SAAH (**1c**) 7%). The higher conversion of the diastereomerically pure analogue suggested that COMT was inhibited by the (S_cR_s) **1c**. The diastereomer (S_cR_s) **1c** was present in the synthetic sample and the inhibition by the (S_cR_s) **1c** diastereomer could have accounted for the synthetic SAAH's lower conversion of DHBAL (**3**) to allylated products. The high conversions of DHBAL (**3**) to alkylated products (**3e-h**) was further shown by high isolated yields of **3e** (72%) and **3g** (63%) after scale up of the COMT concentration and an increase of assay time (see Section 6.4.4 and 6.4.5).

	SAM (1a)	SABH (1d)	SAAH (1c)	SAAH (1c)	SAEH (1b)
■ %conversion	77	26	20	7	1
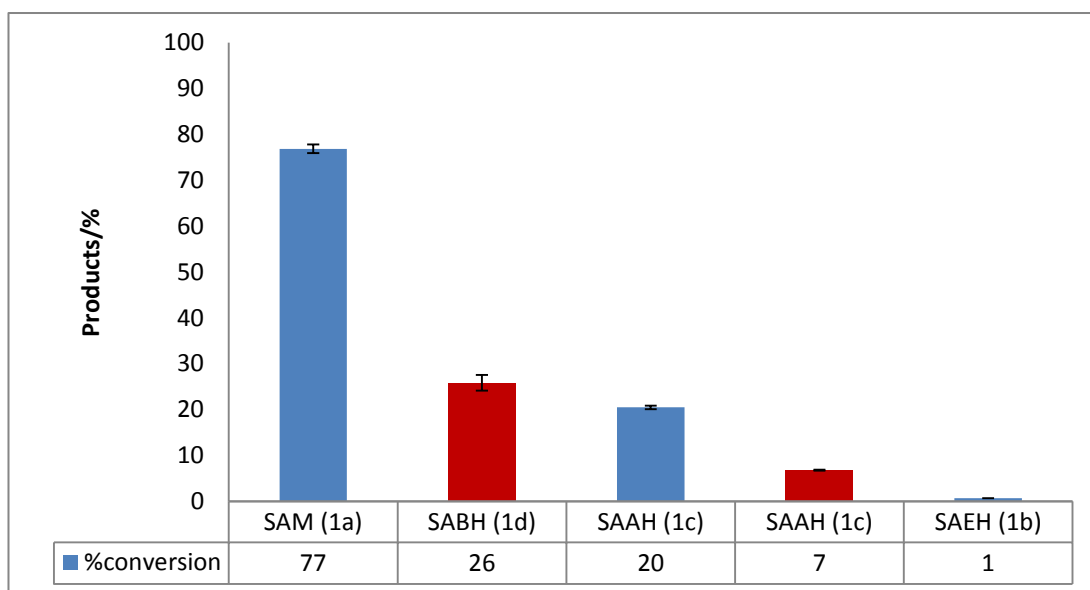
Figure 46: Graph showing the conversions of diastereomerically MAT derived analogues (blue) and synthetic analogues (Red) with COMT and a DHBAL (**3**) substrate. All assays correct to 3% error.

3.2.4 Optimisation of DHBAL (3) ethylation

The conversion assays also showed the low conversion of DHBAL (**3**) with COMT and the SAEH analogue (1%), which agreed with Dalhoff *et al* suggestion, that an adjacent sp^2 system increases activity.[4] The regioselective ethylation of DHBAL (**3**), as mentioned in Section 1.3.2, offers a valuable flavouring agent; ethyl vanillin (**3c**). Therefore, improvement of the hMAT2A-COMT system was necessary to achieve the desirable ethylation. These advantageous conditions were determined with the Y200L mutant, as the desired product was the *meta* ethylated DHBAL (ethyl vanillin) (**3c**) so the most *meta* mutant was used.

Regioselective ethylation has been shown to be useful with COMT and the DHBAL (**3**) substrate. The conversion assays have shown that diastereomerically pure analogues exhibit better activity than the synthetic SAM analogues. Initial tests were focussed on optimisation of the MAT-COMT system with 0.5 mM DHBAL (**3**). The initial data suggested that the bottleneck in the ethyl vanillin (**3c**) production was the COMT enzyme and that poor binding as well as slow turnover of the SAEH (**1b**) cofactor could be the reason for the low conversion. The suggestion of the COMT

bottleneck, was also suggested by the good conversion of the hMAT2A for the ethionine cofactor shown in Figure 47.

In order to further determine whether the COMT would be a bottleneck in the coupled system three concentrations of Y200L COMT were chosen; 5 μM, 25 μM and 50 μM (hMAT2A I322V was used at 25 μM for the three concentrations). The assays were conducted with hMAT2A I322V for 1 hour and then the Y200L was added for a further hour. The same concentrations were also trialled with hMAT2A I322V enzyme (Y200L COMT was used at 25 μM for the three concentrations).
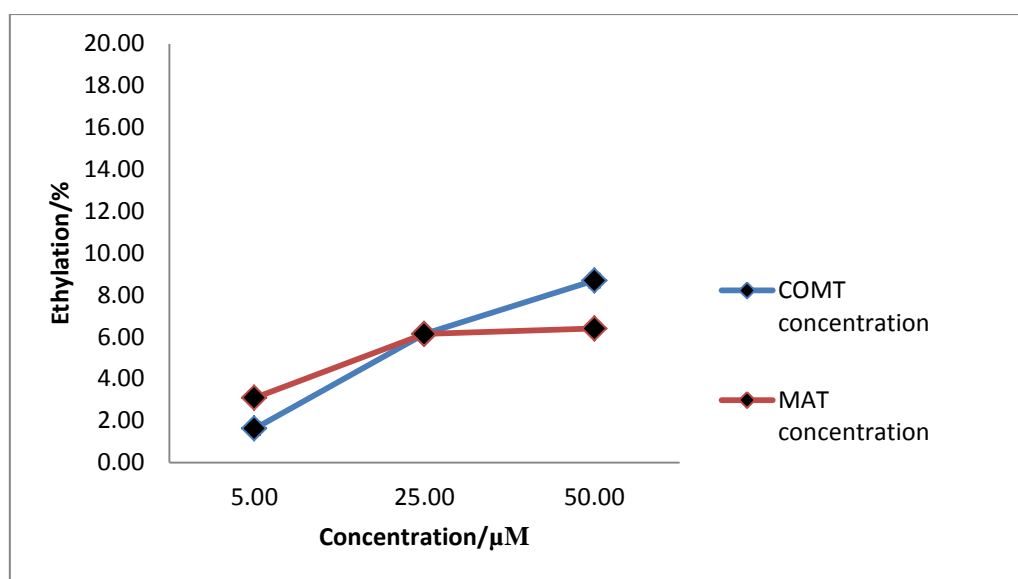
Figure 47: Graph showing the effects on ethyl vanillin (**3c**) production through alteration of the MAT and COMT concentration.

The results of the assays (Figure 47) agreed with the suggestion that the bottleneck was the Y200L COMT enzyme. The highest conversion to ethyl vanillin (**3c**) was achieved with the highest concentration (50 μM) of COMT (9%) whilst at the highest concentration of hMAT2A I322V (50 μM) a plateau was reached (6%), as the middle concentration recorded a similar conversion (6%). Ethylation increased with COMT concentration, suggesting that the K_m of SAEH (**1b**) was a factor in the low amounts of ethylation. Despite the high COMT concentrations, only 9% conversion to ethylated products (ethyl vanillin (**3c**) and ethyl isovanillin (**3d**)) was achieved in 2 hours suggesting that, as well as a high K_m with the SAEH analogue COMT has a low k_{cat}.

The successful conversion to detectable ethylation of the DHBAL (**3**) substrate meant that the regioselectivity of the ethyl analogue with Y200L and **3** could now be confirmed. In order to confirm the regioselectivity, first a standard of ethyl isovanillin (**3d**) was needed. **3d** was synthesised with the potassium carbonate base (in the same manner as the 4-alloxy-3-hydroxy-benzaldehyde (**3f**) and 4-benzyloxy-3-hydroxy-benzaldehyde (**3h**)) and was isolated in a low yield of 19% (see Section 6.9.3). **3d** was commercially available (Sigma) and, with **3c** and **3d** regioisomers now available, a HPLC method was devised (Section 6.5.3) for the separation of **3c** and **3d** regioisomers. Therefore, a HPLC method was developed for separation of the two regioisomeric products of DHBAL (**3**) ethylation (**3c** and **3d**), which meant the regioselectivity of Y200L with the SAEH (**1b**) analogue could now be determined.

Previously a 1% conversion of ethylation would have made determination of the *para* regiosiomer very difficult due to the low peak areas. However, with the higher conversion of ethylation and the HPLC separation of the regioisomers, the r.e. of Y200L with SAEH (**1b**) could be determined. Unsurprisingly, the results showed, that like the SABH (**1d**) and SAAH (**1c**) analogues, **1b** was *meta* regioselective with the Y200L and the DHBAL (**3**) substrate displaying a r.e. 89 %. Interestingly, the r.e. of SAEH showed a trend of **1a**>**1b**>**1c**>**1d** in terms of regioselectivity, further suggesting that the *para* alkylating orientation stabilises bulkier analogues, and thus the regioselectivity will decrease in the *meta* direction for bulkier analogues. Therefore, the regioselectivity of Y200L was again shown to be highly *meta* selective with the **1b**, which was important as the ethyl vanillin (**3c**) was the preferred regioisomer.

The regioselectivity was confirmed for the Y200L COMT enzyme with ethylation of DHBAL (**3**), and suggested that optimisation of the conditions for regioselective ethylation could be achieved. A high concentration of COMT (170 µM) was chosen whilst MAT was elected for use at 25 µM. The high concentration of COMT was used to overcome the weak binding of SAEH (**1b**) and any inhibition of ethyl vanillin (**3c**), to drive towards higher conversion. In order to overcome the suggested low k_{cat} a longer assay time was chosen. Initial tests at 4 hours showed 45% conversion of DHBAL (**3**) to ethyl products (**3c** and **3d**), whilst a good improvement, further improvement was necessary. Therefore, an even longer assay time of 18

hours was chosen to ensure full conversion. The longer assay time led to a conversion of 55%. The improvement from 9% to 55% suggested that the ethylation could be optimised, however, in order to achieve high isolated yields further optimisation was necessary.

The longer reaction assay time did lead to an improvement in yields, however, this was not significant from 4 hours to 18 hours (45% to 55%). The modest increase in yield suggested that inhibition may be occurring between the two time-points. The inhibition of the ethylation could either be due to the larger amounts of ethylated products (**3c** and **3d**) or SAH (**1**) (as well as possible SAM degradation products). In order to overcome possible SAH (**1**) inhibition the SAH homocysteinase enzyme could be used. As Scheme 22 shows, SAH homocysteinase (SAHH) breaks down SAH (**1**) into homocysteine (**2**) and adenosine. Homocysteine (**2**) and adenosine, do not bind as well as SAH (**1**) does to the COMT enzyme, and therefore, **1** inhibition will be decreased by the SAHH. Using the optimal conditions of the MAT-COMT coupled assay, 5 μM of SAH Hydrolase (from *Rattus norvegicus*) (cloned, expressed and purified by Dr Mark Thompson) was added and an improvement of 55% to 77% conversion was achieved for the ethylation of DHBAL (**3**).
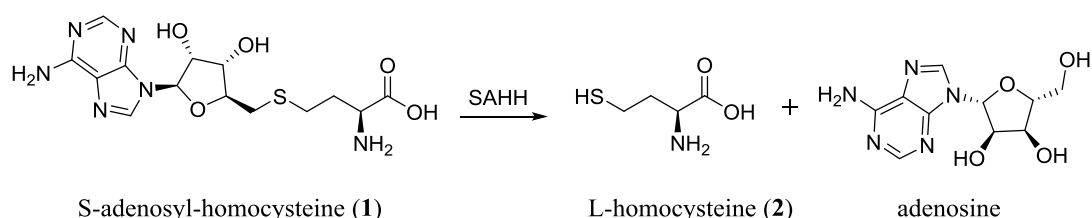
Scheme 22: The breakdown of SAH (**1**) by the SAH Hydrolase enzyme.

Optimisation of the ethylation through coupling of MAT-COMT enzymes meant that optimisation of purification was needed. Using an Agilent Bond Elut C18 column, as a separation technique, a purification method was devised using 5% and 25% washes to remove SAEH (**1b**), DTT, SAH (**1**), DHBAL (**3**) and ethionine (**2b**). The final step was purification of ethyl vanillin (**3c**) by HPLC from vanillin (**3a**) and isovanillin (**3b**), which again proved possible using the gradient in Section 6.6.2. Final separation of the ethyl regioisomers (**3c** and **3d**) was not possible using the HPLC. Therefore, the isolation of the ethyl vanillin (**3c**) will be approximated, with

calculation of integrals in the NMR to compare the relative percentage yield of ethyl isovanillin (**3d**).

Using the optimised assay and purification conditions, the MAT-COMT system was scaled up to ethylate 2.5mg DHBAL (**3**). Using the conditions laid out in Section 6.4.3, scale up was successful with an observed conversion of 75%. After purification 2 mg of ethylated products (**3c** and **3d**) were successfully isolated, the isolated yield of ethylated products was 56%. The percentage of ethyl vanillin (**3c**) of this sample was calculated as 96% from the relative integrals of the aldehyde peaks of **3c** and **3d** in the ^{1}H spectrum. Therefore, a yield of 54% was achieved for ethyl vanillin (**3c**) from 2.5mg of the DHBAL (**3**) substrate, in future optimisation of the purification methods should improve this value.

The successful coupling of MAT-COMT enzymes meant that an industrially relevant molecule; ethyl vanillin (**3c**) could be isolated in modest yields. However, despite expanding the cofactors acceptable with COMT the substrate scope has not been increased. In order to develop the utility of the MAT-COMT system the substrate scope of the COMT enzyme could be improved to diversify the range of substrates that COMT can alkylate. Improvement of COMT's substrate scope will mean that regioselective alkylations can be conducted to produce novel products with novel chemistries.

3.2.5   Methylation of DOPA containing peptides

The regioselective alkylation of the DHBAL (**3**) substrate by SABH (**1d**), SAAH (**1c**) and SAEH (**1b**) analogues had shown that COMT could be used to produce industrially relevant product in ethyl vanillin (**3c**). The success of the alternative cofactors suggested that there was space in the COMT active site to accommodate larger cofactors and substrates. The larger space in the COMT active site meant that the medically interesting DOPA peptides mentioned in Section 3.1.1 could be trialled, as substrates for the COMT enzyme.

Alkylation of DOPA peptides, could be interesting from two points of view. Firstly regioselective alkylation with COMT would provide a method for selective labelling, and secondly, regioselective alkylation or methylation may affect the pharmacokinetics or pharmacodynamics of therapeutic peptides. In order to test the

strategy two sequences were selected containing a DOPA residue; KNFLDDOPA (**6b**) and KETDOPASK (**8b**). The sequence KNFLDDOPA (**6b**) contains a C-terminal L-DOPA residue, and was derived from a peptide sequence present in peptidyl-propyl isomerases (PpiB). [122]Previous studies have shown that L-Dopa can be incorporated into PpiB instead of tyrosine, if an *Escherichia coli* cell-free transcription/translation system was supplemented with L-DOPA.[122] Therefore, the incorporation of L-DOPA into this sequence suggested that a selective alkylation of this residue could be used to label PpiB.[122] The second sequence chosen **8b** has an internal L-DOPA residue, which can be derived from an oxidation of the KETYSK sequence by the Mushroom tyrosinase.[125] Therefore, the two DOPA peptide sequences were chosen based on the ability of the sequences to derive a DOPA residue with either a tyrosinase or through incorporation.

Table 11: Percentage yields and efficiencies of the peptide synthesis for KNFLDDOPA (**6b**) and KETDOPASK (**8b**).

Sequence	Percentage Yield/%	Efficiency/%
KNFLDDOPA (**6b**)	83	97
KETDOPASK (**8b**)	84	97

The two peptide sequences (KNFLDDOPA (**6b**) and KETDOPASK (**8b**)) were synthesised using a Rink amide resin and the Merrifield synthesis, with a HBTU activating agent and DIPEA base (Further details Section 6.10). Both peptides were then purified with HPLC and isolated in good yields showing high efficiency for the method used (Table 11), with sequence confirmed with NMR, MALDI and MS-MS (Section 6.10 Table 18 and Table 19) (all NMR spectra in this section were on the 800MHz spectrometer by either Dr Matthew Cliff or Dr Anna-Winona Struck). The success of the synthesis of the two peptides meant that conditions for methylation could be trialled. Both peptides (**6b** and **8b**) were trialled at 0.5 mM since any higher than this may result in inhibition of COMT, whilst at lower concentrations, there may not be sufficient binding to COMT (Section 6.4.8 for full conditions). Figure 48

and Figure 49 show the MALDI of the assay mix after 1 hour and that peptides **6b** and **8b** were nearly fully converted to the methylated peptides.
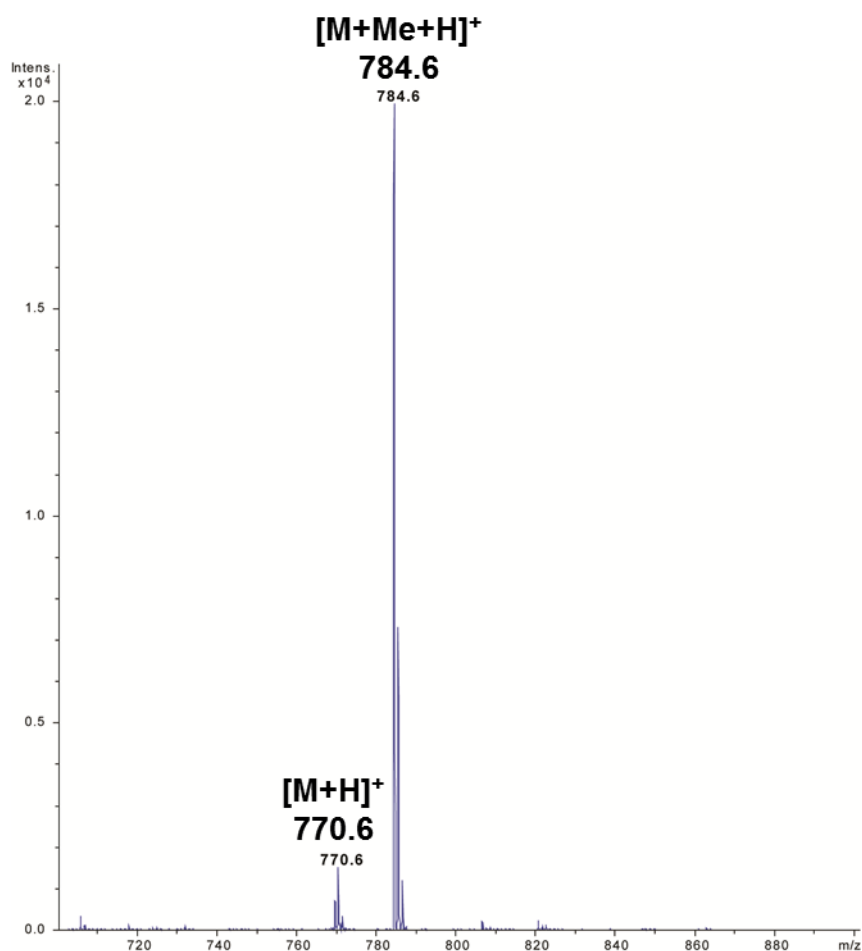
Figure 48: MALDI of the methylation assay of KETDOPASK (**8b**). The [M+H] $^+$ peak of **8b** at 770.6 m/z (calculated 770.4043 m/z) and the [M+Me+H] $^+$-peak of the product KETDOPA-MeSK (**8c**) at 784.5 m/z (calculated 784.4199 m/z) was shown.

The assays were scaled up to isolate the two methylated peptides, and were conducted with WT COMT dimer, to yield regioselective methylation. The methylated peptide was purified by HPLC (Section 6.6.5), which indicated good conversion for both peptides to their methylated products (KNFLDDOPA-Me (**6c**) and KETDOPA-MeSK (**8c**)) (>95%). However, a modest isolated yield of 40.9% was recorded for **6c**. Whereas isolation of the **8c** peptide did not result in an accurate yield, as HPLC purification was difficult and not all of the methylated peptide could be separated from the starting material KETDOPASK (**8b**). MS-MS was again used to confirm the sequence as well as confirming methylation on the tyrosine residue for **6c** (Section 6.11 Table 20 and Table 21) (whereas **8c** was confirmed by ^{1}H NMR

due to **8c** being too salty for MS-MS). The KNFLDDOPA-Me (**6c**) NMR, as can be seen in Figure 50, shows only one methyl peak at 3.7 ppm consistent with a methoxy peak (which was also not present in the starting KNFLDDOPA (**8b**) peptide) suggesting that regioselective methylation had occurred.
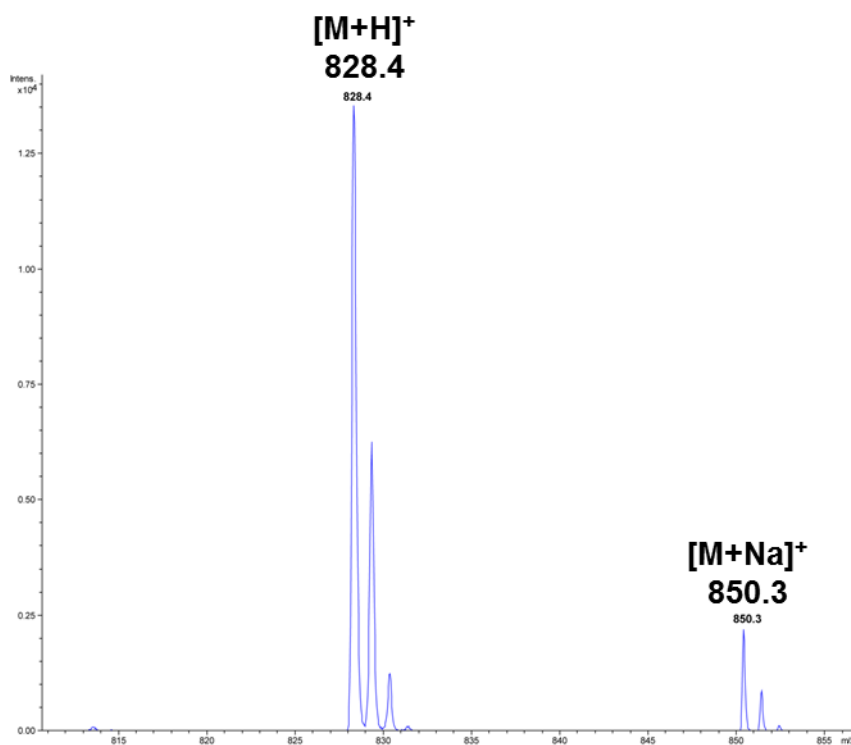
Figure 49: MALDI of the methylation assay of KNFLDDOPA (**6b**). The [M+H]$^+$-peak of the product KNFLDDOPA-Me (**6c**) at 828.4 m/z (calculated 828.4250 m/z), the sodium peak of the **6c** product at 850.3 m/z (calculated 850.4070).

The successful isolation of KNFLDDOPA-Me (**6c**) meant that a HPLC method could be developed to determine the conversion of KNFLDDOPA (**6b**) with the COMT enzyme. The method showed that **6b** could be resolved from **6c** (method was shown in Section 6.5.5). Activity assays were then conducted with **6b** and the COMT (5 μM) for an hour showing a conversion of 94% to the **6c** product. Unfortunately, resolution of KETDOPASK (**8b**) and KETDOPA-MeSK (**8c**) on analytical HPLC was not achieved, however, from the MALDI spectrum (Figure 48) a similar conversion as **6b** was expected.
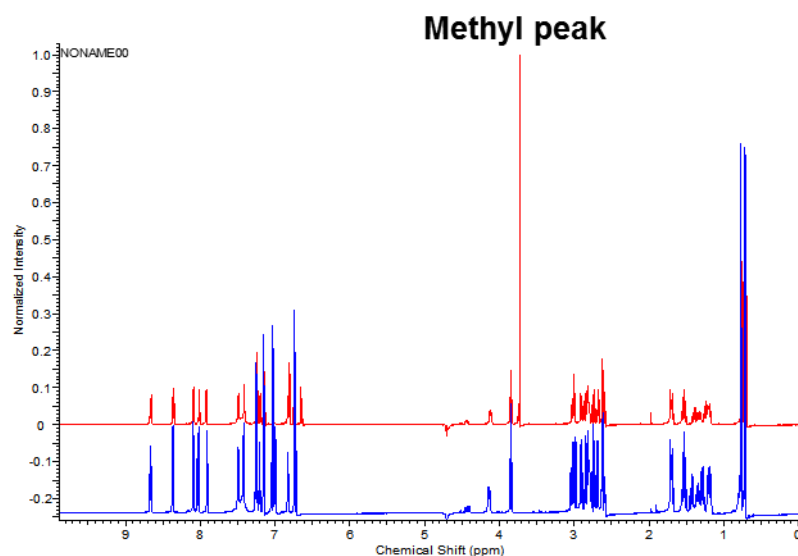
Figure 50: Spectra of KNFLDDOPA-Me (**6c**) (Red) overlaid with KNFLDY (**6a**) (blue) reference standard.

The synthetic peptides showed that COMT could methylate DOPA peptides, indicated that the substrate scope for COMT was larger than initially thought. Methylation of the terminal (KNFLDDOPA (**6b**)) and mid-chain (KETDOPASK (**8b**)) DOPA residues suggested that COMT can methylate most catechols with the CH_2-catechol motif. The methylation also suggested alkylation of these peptides would be possible. However, the need for a DOPA residue does limit the possible scope of this system. Therefore, in order to increase the value of the method, an *in situ* generation of a DOPA residue was necessary.

3.2.6   Methoxylation of phenolic peptides

The *in situ* generation of a DOPA peptide needed to fulfil two criteria; firstly, the generation of the DOPA peptide did not affect other residues, and secondly, that generation was regioselective without formation of side products. Based on these two criteria, oxidation of tyrosine residues was decided as the most desirable method. Chemical oxidation methods would not have been selective for tyrosine residues over other residues. For enzymatic hydroxylation of tyrosine residue, the two most probable enzymes to have chosen from were peroxidases and tyrosinases.[125] Due to the lack of selectivity offered by the peroxidase enzyme and the success of the mushroom tyrosinase with linear peptides, as shown by Marumo and Waite, the mushroom tyrosinase wash chosen to couple with the COMT enzyme.[125]

144

Previously Marumo and Waite had shown that hydroxylation could be achieved with a number of linear peptides.[125] The tyrosine containing peptides; KNFLDY (**6a**) and KNYLDF (**7a**) were synthesised and characterised as before (**7a** was chosen due to the better HPLC compared to KETYSK (**8a**)). Use of peptides **7a** and **6a** would allow a comparison between internal as opposed to terminal tyrosine, as substrates for the tyrosinase. Therefore, an conversion assay with KNYLDF (**7a**) and KNFLDY (**6a**) with the tyrosinase could determine the selectivity of the enzyme.
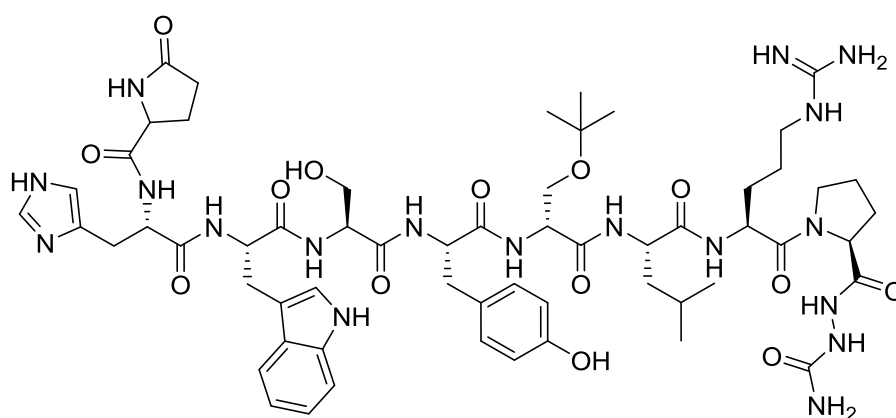
Goserelin (**9a**)
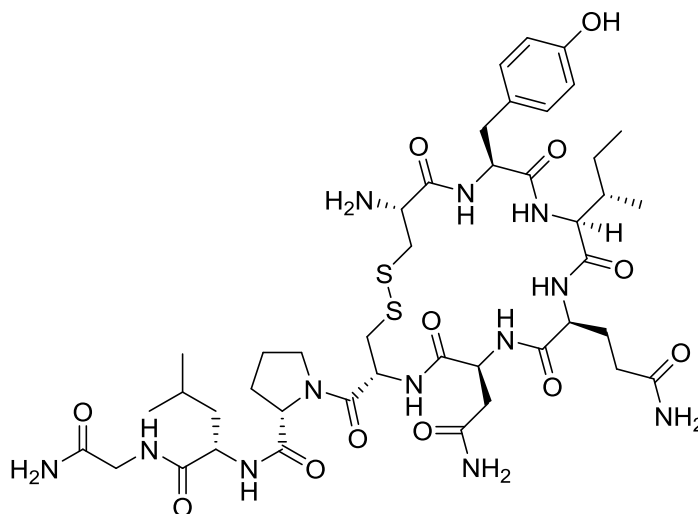
Oxytocin (**10a**)

Figure 51: Goserelin (**9a**) and oxytocin (**10a**) peptides.

In order to see whether the tyrosinase has a larger scope than linear peptides, commercial peptides were also trialled along with KNFLDY (**6a**) and KNYLDF (**7a**). The commercial peptides chosen were goserelin (**9a**) and oxytocin (**10a**) (Figure 51). Goserelin (**9a**) has been used in the clinic as a prostate cancer

treatment, whilst oxytocin (**10a**), a human hormone, has been listed as an essential drug; used in childbirth.[130,131] The therapeutic nature of the tyrosine containing peptides makes both **9a** and **10a** attractive targets. **10a** offers a different structure from the previous targets both in literature and this study, being cyclic in structure and goserelin also contains some unnatural amino acid residues such as *tert*-butyl serine.[132,133] Therefore, the goserelin (**9a**) and oxytocin (**10a**) substrates could be determined to test the scope of the tyrosinase as well as possibly providing therapeutically useful products.

Using conditions from Marumo and Waite, assays were set up at 25°C, 1 mM peptide (**6a-10a**) and 25 mM ascorbic acid to prevent quinone formation with a commercial mushroom tyrosinase (0.7 μM).[125] MALDI showed successful hydroxylation of the two synthetic peptides (KNYLDF (**7a**) and KNFLDY (**6a**)) as well as the commercial peptides (oxytocin (**9a**) and goserelin (**10a**)). Therefore, hydroxylation was shown for all the targets both commercial and synthetic, with two novel substrates determined for the tyrosinase enzyme in **9a** and **10a**.

Based on the success of the tyrosinase enzyme with the oxytocin (**10a**) and goserelin (**9a**) substrates, the tyrosinase assays with the **10a** and **9a** were scaled up to isolate hydroxylated standards for HPLC. The scale up and isolation of the hydroxyoxytocin (**10b**) and hydroxygoserelin (**9b**) were successful, with confirmation of hydroxylation determined by MS for **10b** and **9b**. Hydroxygoserelin (**9b**) was isolated 59% yield whereas hydroxyoxytocin (**10b**) was isolated in 25% yield (also shown by Figure 52). The respective isolated yields were reflected in the conversions, with the oxytocin recording a much lower conversion with the tyrosinase enzyme than the goserelin substrate.

Isolation of the hydroxylated peptides; hydroxygoserelin (**9b**) and hydroxyoxytocin (**10b**), as well as the previously synthesised KNFLDDOPA (**6b**) and KNDOPALDF (**7b**), meant that activity of the four phenolic peptides (KNFLDY (**6a**), KNYLDF (**7a**), goserelin (**9a**) and oxytocin (**10a**)) could be compared with the tyrosinase enzyme. Previous assays had shown that tyrosinase conversion occurred quickly, therefore, the assay was conducted for 30mins with 1 mM substrate and 25 mM ascorbic acid. Termination was, as before, with 1% FA. Assays were run in triplicate and the data displayed in Figure 52.
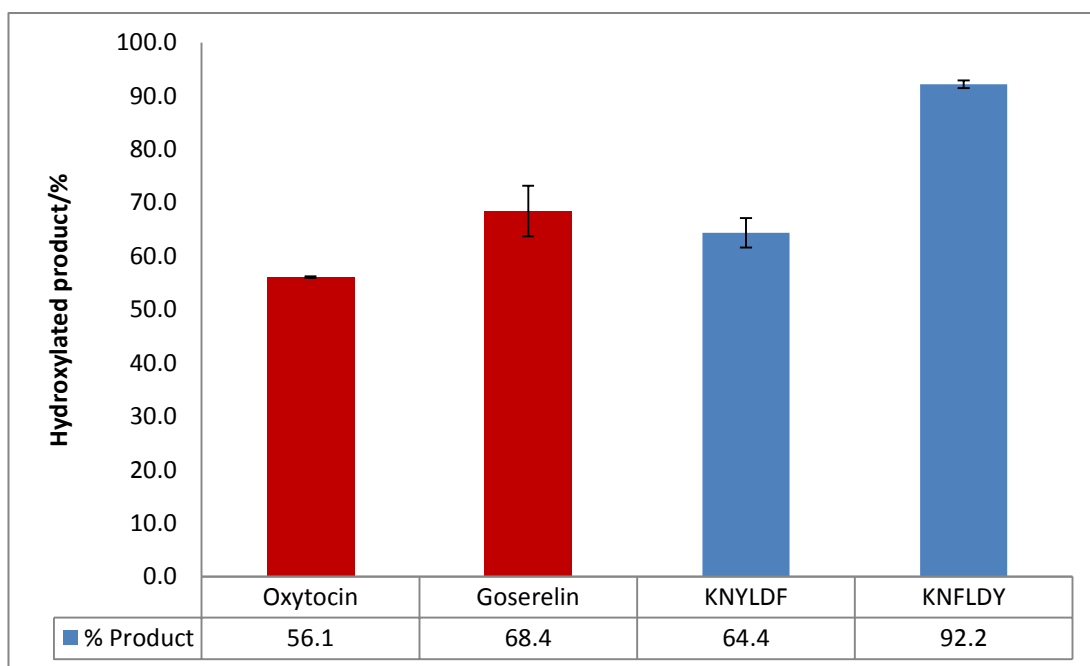
Figure 52: Tyrosinase conversion assays with synthetic peptides (**6a** and **7a**) (blue) and commercial therapeutic peptides (**9a** and **10a**) (red).

The results showed that all the peptides were good substrates with the tyrosinase enzyme with a range of 56-92% hydroxylation noted across the substrates in just 15 minutes. Oxytocin (**10a**) was the worse substrate and this could be due to the large cyclic nature of the molecule, which could weaken binding to the enzyme. The results of the remaining three peptides did suggest that hydroxylation could be selective, with comparison of the peptides KNFLDY (**6a**) and KNYLDF (**7a**) showing higher conversion for the terminal tyrosine containing peptide **6a**. Meanwhile the mid chained tyrosine in goserelin (**9a**) also shows less conversion than the terminal **6a**. The results of the tyrosinase conversion assay suggested that selective hydroxylation of tyrosine residue, at least in KNFLDY (**6a**) and KNYLDF (**7a**) sequences, could be possible.

The high conversion of all substrates by the tyrosinase enzyme meant that a coupled methoxylation with the COMT and tyrosinase enzymes could be attempted. However, Blaschke *et al* had shown ascorbic acid was a substrate for COMT with a K_i of 14 mM.[134] Based on 50% inhibition by ascorbic acid at 14 mM, a lower ascorbic acid concentration of 12.5 mM was used to prevent quinone formation. The DTT, which had been shown to be redundant in monomeric and dimeric experiments, was also discarded. DTT inhibits tyrosinase activity by firstly reducing

the copper centre of the tyrosinase enzyme, the DTT has also been observed to provide catecholic adducts.[135] The coupled assay conditions were also set at 30°C, as initial assays confirmed that both COMT and tyrosinase still retain a good degree of their activities at this temperature. These considerations developed a suitable method for testing methoxylation with the coupled tyrosinase and COMT enzymes.

The conditions were used with 1 mM substrate concentration with the KNYLDF (**7a**), KNFLDY (**6a**) and goserelin (**9a**) peptides with methylation confirmed by MALDI and MS-MS for the peptides (as well as HPLC compared to the methylated standards). However, for oxytocin a lower concentration of 0.1 mM was used, as above this concentration inhibition was thought to occur. 0.7 μM tyrosinase was used, as before. However, a higher concentration of dimeric WT COMT (15 μM) was used in the coupled system than the concentration used for DOPA peptides. A higher concentration of dimeric WT COMT was used because the competitive inhibition from ascorbic acid could decrease binding of the DOPA peptide. The optimised conditions had shown that *in situ* hydroxylation was possible leading to methoxylated products of all the substrates tested.

Successful methoxylation of goserelin (**9a**) and oxytocin (**10a**) meant that both were scaled up again for isolation. However, isolation of methoxyoxytocin (**10c**) was not successful due to the low conversion of the **10a** substrate. Therefore, **10c** was isolated in 7% after two separate steps; firstly hydroxylation to hydroxyoxytocin (isolated yield 25%) and secondly methylation (29%). Whereas goserelin (**9a**) was a better substrate than **10a** for the methoxylation and an isolated yield of 37% was recorded for methoxygoserelin (**9c**). The isolated **9c** and **10c** were then characterised by MS-MS, NMR and HPLC techniques (Sections 6.11.6 and 6.11.9). The NMR characterisation of the methoxygoserelin (**9c**) was also able to determine that *meta* methoxylated regioisomer of **9c** was the product, noted by a cross peak between 5 position of the tyrosine and the methyl peak in the NOESY spectrum (Figure 53). This confirmed the previous regioselectivity noted with dimeric COMT (see Section 2.2.3) that the methylated regioisomers of the DOPA peptides were of the *meta* orientation. The isolation of the methoxylated standards (methoxygoserelin (**9c**) and methoxyoxytocin (**10c**)) meant that conversions of the methoxylations of goserelin (**9a**), oxytocin (**10a**), KNYLDF (**7a**) and KNFLDY (**6a**) could now be determined.
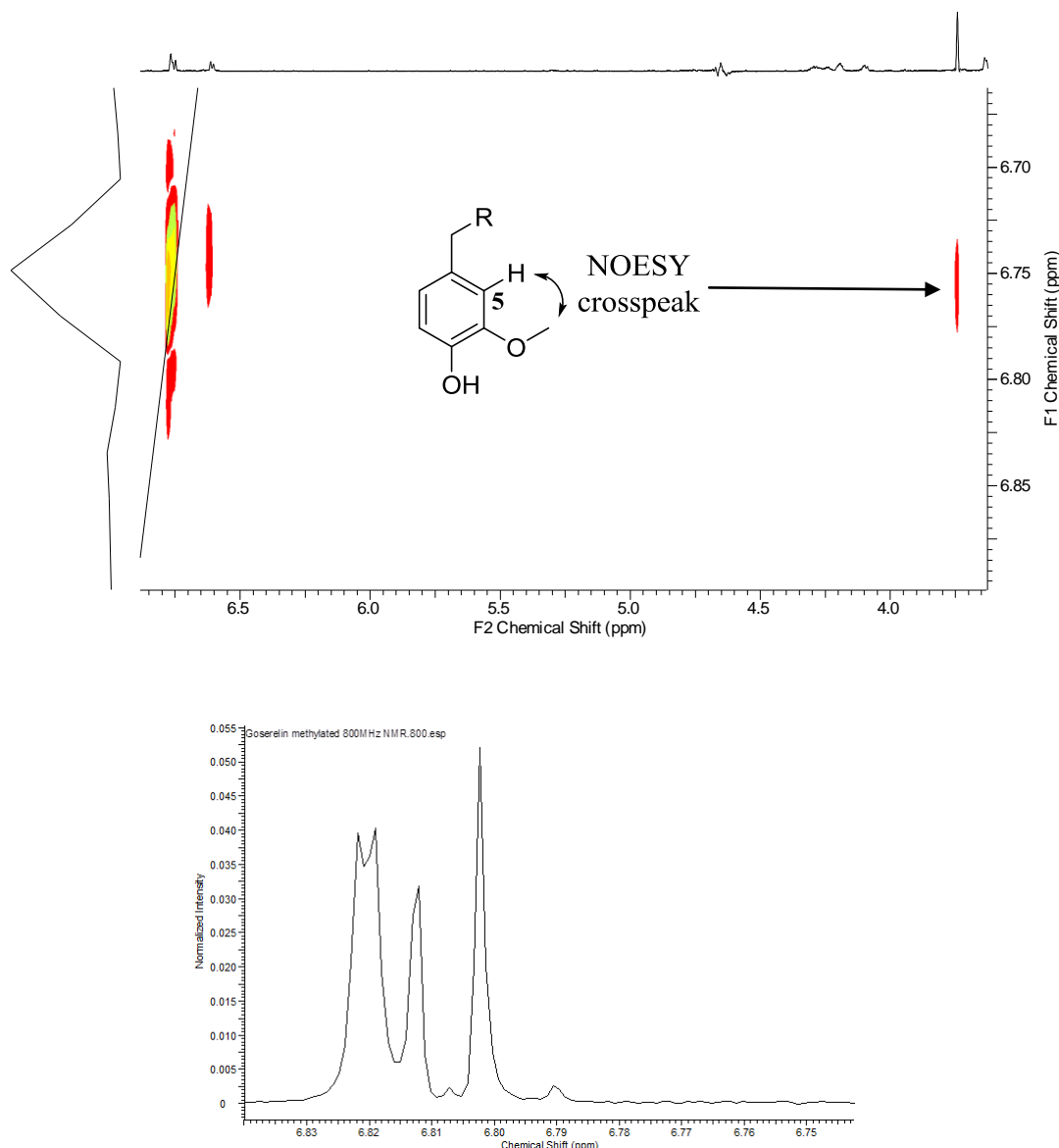
Figure 53: (Top) The NOESY crosspeak between methoxylated DOPA residue's position 5 proton (3.82) (6.83ppm) and the methyl protons of the methoxylated DOPA residue of methoxygoserelin (**9c**). R was the rest of the peptide chain of methoxygoserelin (**9c**). (Bottom) (Zoom) 1H methoxygoserelin, shows that 6.83ppm is a singlet consistent with the 5 position.

The initial assays had suggested that COMT was the slower of the two enzymes with methylation of DOPA peptides taken longer than hydroxylation of tyrosine containing peptides. This was also suggested by the kinetics of the two enzymes, as tyrosinase has higher k_{cat} values than COMT, suggesting the tyrosinase was the faster enzyme.[5,136] Based on these values methoxylations were carried out over 3 hours with a high COMT concentration (15 μM) (remaining conditions can be found in Section 6.4.10). The results of the assay (Figure 54) showed a similar trend to the hydroxylation conversion assays; KNFLDY (**6a**)>KNYLDF (**7a**)>goserelin

(**9a**)>>>oxytocin (**10a**). The much lower conversion of **10a** (compared to the other substrates and the previous hydroxylation assays) was most likely attributable to poor binding of hydroxyoxytocin (**10b**) to COMT, as COMT has a much smaller active site than the tyrosinase. The poor binding of hydroxyoxytocin (**10b**) to COMT could be attributable bulkier structure of **10b**, which will have a worse fit in smaller active sites compared with the other substrates. Therefore, the low methoxylation of oxytocin was most likely due to the methylating enzyme (COMT) rather than the tyrosinase.

	Oxytocin	Goserelin	KNYLDF	KNFLDY
■ % Product	9	54	71	77
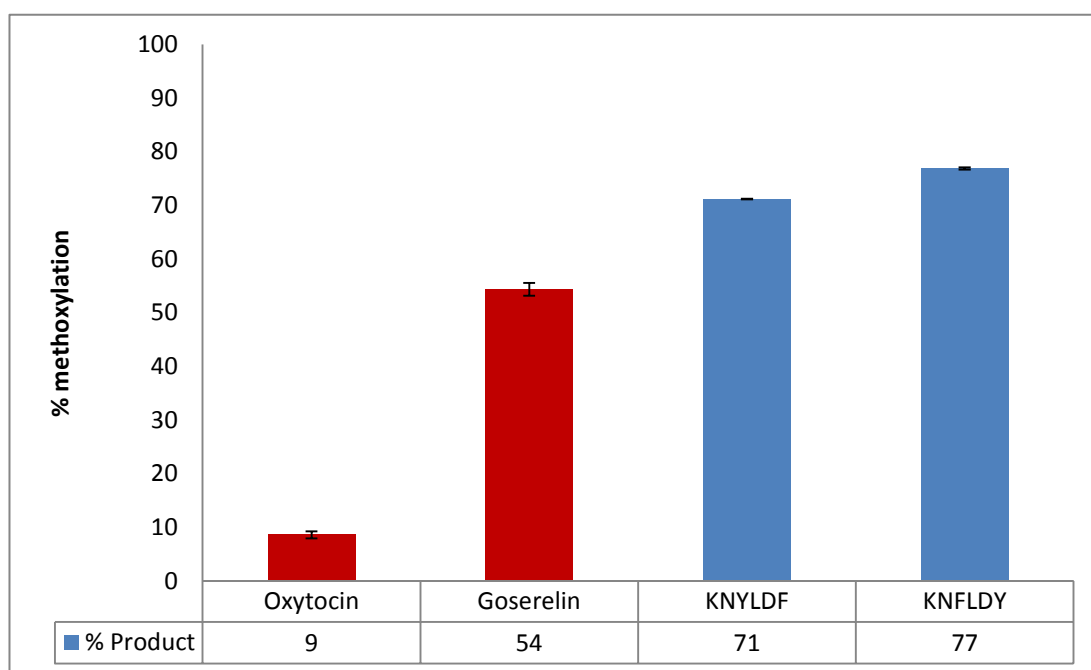
Figure 54: Methoxylation conversion assays with synthetic peptides (**6a** and **7a**) (blue) and commercial therapeutic peptides (**9a** and **10a**) (red).

The assays showed that selective methoxylation of the terminal KNFLDY (**6a**), over KNYLDF (**7a**) was possible. However, the values showed only a ~5% difference in conversion. In future studies selective methoxylation could be achieved by lowering the assay time, which could lead to a bigger discrepancy between terminal and mid chained tyrosines. The two substrates containing mid chain tyrosines, were goserelin (**9a**) and **7a**, which showed similar conversions. However, **9a** was lower in conversion compared to **7a**. The lower conversion of **9a** compared to **7a** may be due to the bulkier residues on the hydroxygoserelin (**9b**) and **9a** substrates such as tert-butyl serine and tryptophan, which like oxytocin (**10a**) may not be accommodated well in the smaller COMT active site.

The successful methylation and methoxylation of the commercial (goserelin (**9a**) and oxytocin (**10a**)) and synthetic peptides (KNFLDY (**6a**) and KNYLDF (**7a**)) meant that another substrate could be considered. Tyrocidine A (**11a**) (Figure 55) was considered as a larger more strained substrate, similar in structure to **10a**, which was aimed to test whether methoxylation was limited by the larger substrates. An early antibiotic, tyrocidines were constructed by a NRPS from *Bacillus brevis*.[137] Tyrocidines have limited clinical use due to tyrocidines' toxicity to red blood cells, but have been found in lozenges, as well as topical creams.[138] Lin *et al* had shown that manipulation of the tyrosine residue with glycosides could improve the selective toxicity of tyrocidine against bacterial cells over human cells.[139] Therefore, selective methoxylation of **11a** would not only test the limits of the coupled tyrosinase-COMT system, but could possibly uncover tyrocidine analogues that may exhibit a better therapeutic profile.
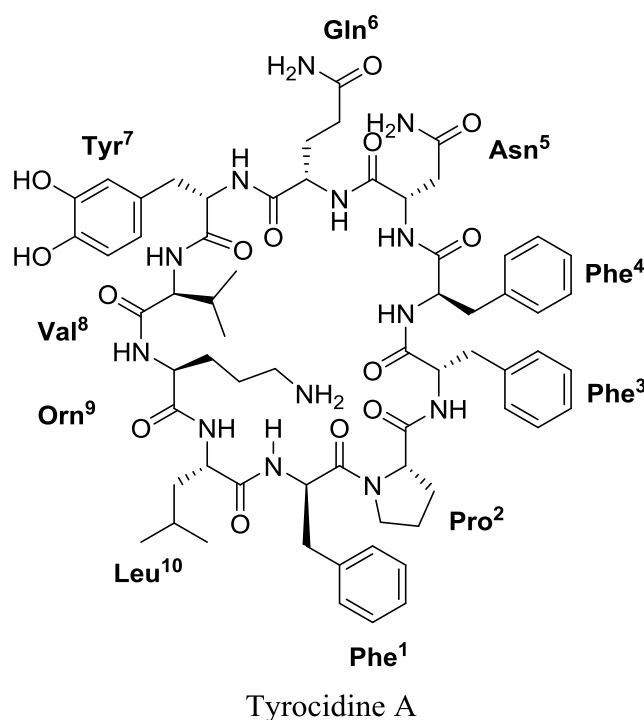
Tyrocidine A

Figure 55: Tyrocidine A (**11a**) structure and amino acid composition.

The tyrocidine A (**11a**) structure being similar to oxytocin (**10a**), meant that **11a** was run at a low concentration of 50 μM with the tyrosinase-COMT coupled assay. Assays were also run overnight, as conversion with the linear peptides (KNFLDY (**6a**), KNYLDF (**7a**) and goserelin (**9a**)) was not complete in 4 hours, and **11a** was thought to be a worse substrate than the linear peptides. Conversion to the

151

methoxylated product was confirmed by the appearance of the 1300.7 mass, which was not present in the controls (Table 12). However, conversion of the substrate was not complete (even with overnight runs) suggesting that hydroxy-tyrocidine (**11b**) may be the limit of what COMT can accommodate in terms of steric bulk.

The suggestion that tyrocidine A (**11a**) was not a good substrate for the methoxylation by the coupled tyrosinase-COMT needed to be determined by conversion assays. However, conversions could not be accurately calculated, as **11a** was not a good candidate for HPLC with a variety of columns trialled, poor retention of the tyrocidine (**11a**) meant that accurate peak areas could not be determined. Conversions were approximated from MALDI and showed under tyrosinase conversion assay conditions an approximate conversion of 34% to hydroxy-tyrocidine A (**11b**) (30 minutes) whereas methoxylation showed an approximate conversion of 50% to methoxy-tyrocidine A (**11c**) (16 hours). These approximate conversions did suggest that methoxylation was limited by larger bulkier substrates.

The successful hydroxylation and methoxylation of tyrocidine A (**11b-c**) was needed to be confirmed by the MS-MS results of the hydroxy-tyrocidine A (**11b**) and methoxy-tyrocidine A (**11c**). Ring opening of the tyrocidine A (**11a**) was determined to occur between the phenylalanine[1] and proline[2] residues (Figure 56) from these fragments were calculated and compared to the observed results (Table 12). The observed results were consistent with the fragments found by Vosloo *et al* (Figure 56 and Table 12).[140]
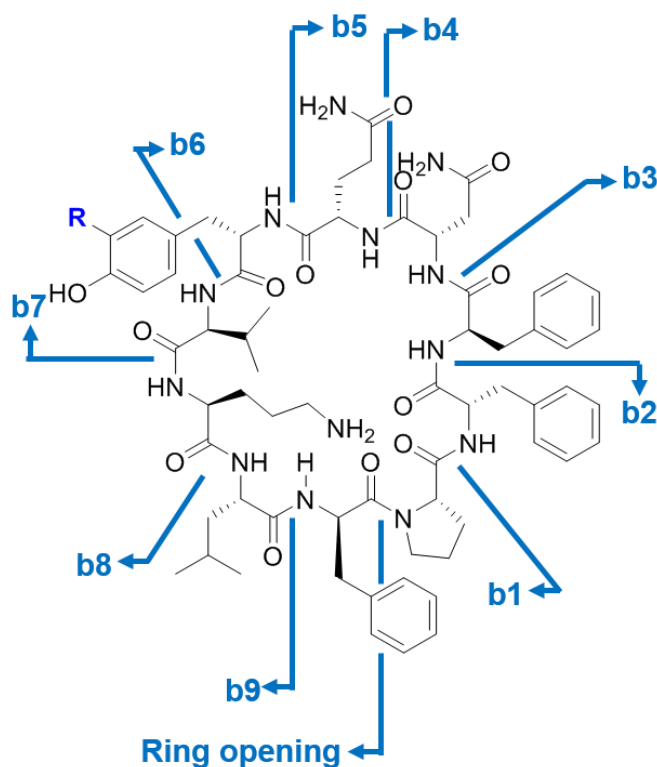
Figure 56: Proposed ring opening for fragmentation of tyrocidine A (**11a**) by MS-MS.[140]

Table 12: MS-MS fragments for tyrocidine (**11a**), hydroxytyrocidine (**11b**) and methoxytyrocidine (**11c**). Fragments were consistent with Visloo *et al.*[140] Calculated (calc.) and observed (obs.) values are shown (an error of 0.7ppm was recorded for all samples, due to poor calibration).

proposed ring opening		Tyrocidine A (11a)		Hydroxytyrocidine A (11b)		Methoxytyrocidine A (11c)	
		calc.	obs.	calc.	obs.	calc.	obs.
none	[M+H]$^+$	1270.7	1271.0	1286.7	1287.0	1300.7	1300.9
Phe1-Pro2	b9	1123.6	1122.8	1139.6	1140.0	1153.6	1153.9
Phe1-Pro2	b8	1010.5	1009.7	1026.5	1025.8	1040.5	1039.7
Phe1-Pro2	b7	896.4	895.7	912.4	911.8	926.4	925.7
Phe1-Pro2	b6	797.4	796.9	813.4	812.8	827.4	826.6
Phe1-Pro2	b5	634.3	633.8	634.3	634.7	634.3	633.5
Phe1-Pro2	b4	506.2	505.7	506.2	505.7	506.2	505.5
Phe1-Pro2	b3	392.2	391.6	392.2	391.6	392.2	391.6
Phe1-Pro2	b2	245.1	244.7	245.1	244.8	245.1	244.6
Phe1-Pro2	b1	115.1	114.8	115.1	114.8	115.1	114.8

The tyrosinase-COMT coupled assay had shown that a range of peptides with different structures and chemistries could be methoxylated. The methoxylation has

been successful with three therapeutic peptides; goserelin, oxytocin and tyrocidine A showing the value of this methodology for selective manipulation of tyrosine containing therapeutic peptides. However, methoxylation may only slightly alter the chemistries of these peptides. In order to expand the scope of the coupled assay, alkylation was also developed for use with the tyrosinase-COMT system.

3.2.7   Alkylation of DOPA peptides

The success of the methoxylation assay had shown that a coupled tyrosinase-COMT system would work on phenolic peptides. The methoxylation was successful with commercial therapeutic peptides, which suggested pharmaceutically relevant analogues could be prepared by this approach. However, this only fulfilled one aim of this assay, the other aim of labelling selected lengths of peptide in possible proteins was left unfulfilled. In order to achieve this secondary aim potential labelling SAM analogues were, firstly, needed to be synthesised and secondly used in conjunction with the tyrosinase-COMT system.

The idea of labelling a peptide sequence required a label that was; water soluble (on the SAM moiety), easily detectable and easy to synthesise. The 4-propargyloxy-but-2-ene (POB) (**2f**) moiety (Scheme) fitted all of these categories. As mentioned in Section 1.2.4, previously Wang *et al* had shown with a POB (**2f**) label that POB SAM (**1e**) could be synthesised and the **2f** chain subsequently transferred by an arginine methyltransferase (PRMT1) to a human histone target.[71] The **2f** label was shown to then undergo further transformations that could be used in a labelling context.[71] Therefore, **2f** was selected as the linker that would be firstly synthesised and then trialled with COMT to see if transfer was possible.

The synthesis of POB SAM (**1e**) was conducted over two steps as Scheme 23 shows. Using methods from Wang *et al* the POB SAM (**1e**) was synthesised with an overall yield of 9%.[71] The POB SAM (**1e**) was characterised by MALDI and 1H NMR both gave the correct spectra associated with the literature (Section 6.7.5).[59,71] Despite being low yielding the synthesis produced enough of **1e** for the purposes of this study. The successful synthesis of the SAM analogue meant that the POB SAM (**1e**) cofactor could now be trialled with both the commercial and synthetic peptides.
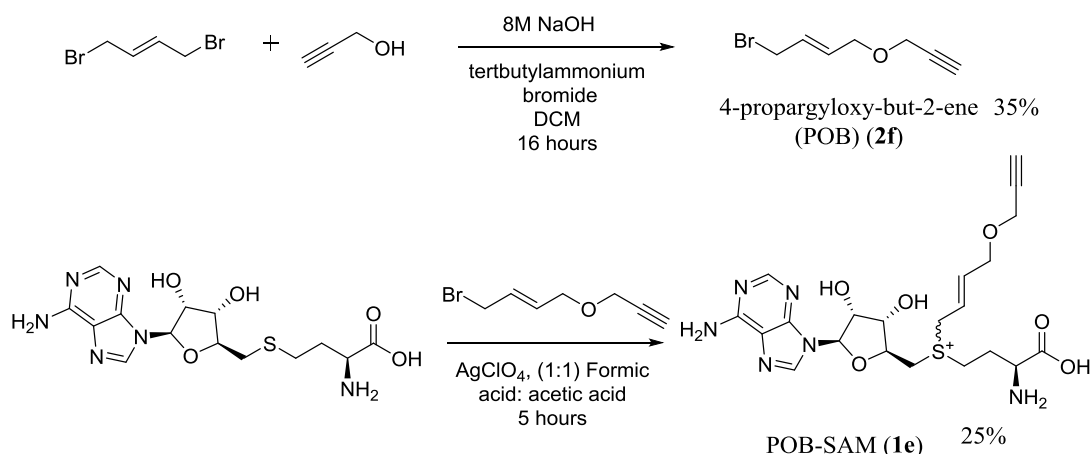
Scheme 23: The two step synthesis of POB SAM (**1e**) adapted from Wang *et al.*[71]

Section 2.1.3 showed that the SAM analogues (**1b-d**) were much less active with COMT compared to SAM. The highest activities of COMT were with the SABH (**1d**) and SAAH (**1c**) analogues. These analogues activities relative to SAM (**1a**) were; 34% (**1d**) and 27% (**1c**). Based on the lower activities of the SAM analogues, the POBylation was first trialled with the COMT enzyme and the hydroxylated peptides (hydroxygoserelin (**9b**), hydroxyoxytocin (**10b**), KNFLDDOPA (**6b**) and KETDOPASK (**8b**)). POBylation was shown to be successful with hydroxygoserelin (**9b**), KNFLDDOPA (**6a**) and KETDOPASK (**8a**) substrates (see Sections 6.11.7, 6.11.2, and 7.1.8 respectively), with the appropriate masses observed for the POBylated products of all three peptides (POBoxy-goserelin (**9d**), KNFLDDOPA-POB (**6d**) and KETDOPA-POBSK (**8d**)) whilst MS-MS confirmed the location on the tyrosine residue for POB-KNFLDDOPA. Finally, the POBylation reaction was scaled up, for the hydroxygoserelin substrate (as the most interesting substrate), and the POBoxy-goserelin was isolated and characterised by ^{1}H NMR and MS. Analysis of the ^{1}H and MS spectra confirmed the transfer of the POB chain to the hydroxygoserelin substrate (see Section 6.11.7 and Figure 178).

The success of the POBylation assays meant that the POBoxylation could be trialled. However, no POBoxylation conversion for any of the peptide substrates was observed, even with a multitude of different conditions trialled. The lack of POBoxylation was thought to be attributable to the inhibition of the COMT enzyme by the ascorbic acid. In order to develop this system for further use, particularly for a labelling function, the problem of ascorbic acid inhibition would have to be overcome.

The development of two coupled systems (MAT-COMT and tyrosinase-COMT) has enabled regioselective transfer of a range of alkyl groups to a diverse range of substrates. DOPA peptides have been identified as novel substrates for COMT, with both synthetic and commercial enzymes methoxylated by the coupled tyrosinase-COMT system. Although, alkoxylation was not shown to be possible with the POB SAM, under the current conditions, the development of the tyrosinase-COMT system has shown the versatility of methyltransferase coupled systems. Further development should make alkoxylations possible with the coupled tyrosinase-COMT system enabling the development a peptide and even protein labelling system.

3.3 Conclusion and Future work

The success of regioselective alkylation with both synthetic and enzymatic SAM analogues cofactors showed that the COMT enzyme could be used as a selective alkyltransferase with a variety of substrates. The enzymatic coupled MAT-COMT system showed that the problem of diastereomeric contamination could be overcome. The tyrosinase-COMT coupled system had shown that methoxylation could be conducted on the novel substrate phenolic peptides, resolving the problem of introducing the DOPA residues. However, in order to achieve some of the aims set out in Section 2.1.1 further development of these coupled systems is necessary.

The regioselective alkylation of catechols has been suggested as a useful method of regioselective protection. As can be seen in Section 1.3.2, there are a number of *para* and *meta* alkylated catechols, a regioselective protection of either hydroxyl of the catechol, is therefore desirable to make these compounds. The benzylation reaction would be the most desirable, as the benzyl group is stable to a variety of conditions, meaning a variety of chemistries could be conducted on the protected molecule. However, under existing conditions the benzylation reaction by the COMT enzyme is held back by diastereomeric contaminants, as only the synthetic SABH (**1d**) analogue is possible. Future work to improve the COMT enzyme for regioselective protection, should thus focus on mutation of the hMAT2A enzyme for activity with SBH. Previous work conducted by Wang *et al*, has shown improved activity with the

hMAT2A I322V mutant and methionine analogues by creating more space in the methionine binding pocket (the mutant used in this study).[59]
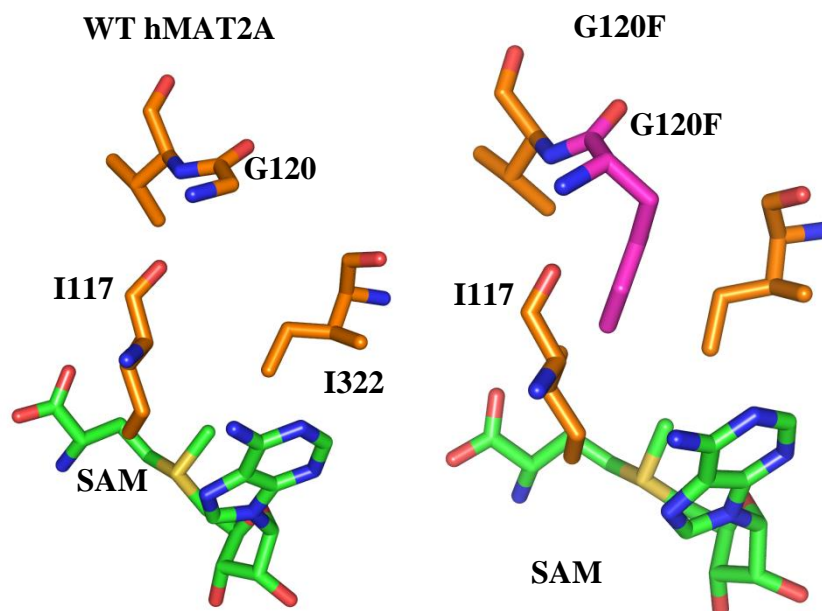
Figure 57: (Left) WT hMAT2A key residues (Right) suggested G120F mutation. Structure adapted from J Komoto *et al* (PDB code 1RG9).[86]

Consulting the crystal structure of MAT (Figure 57) possible mutation sites have been identified and could be used to activate SBH (**2d**) towards MAT. The activity with **2d** could be encouraged through the G120F mutation (or any aromatic amino acid, see right Figure 57), which may encourage binding of **2d** through stabilising π-π stacking interactions. Therefore, through encouragement of **2d** binding the SABH (**1d**) could be made *in situ* and used with the MAT-COMT system to improve the regioselective protection of catechols.

Improvement of the MAT-COMT system could allow for a greater variety of alkylation to occur. Coupling the MAT-COMT system with the tyrosinase-COMT system would also improve the substrate scope of the two systems. The triple enzyme (MAT-tyrosinase-COMT) was attempted in this study, but was unsuccessful. This was most likely due to the ascorbic acid, whilst a useful reductant of the quinone formed by the tyrosinase enzyme, destabilises the hMAT2A enzyme through lowering of the pH and competitively inhibits the COMT enzyme. To overcome these problems an alternative reductant of the quinone should be sought. There are a variety of different tyrosinase inhibitors from flavones to long chain

lipids.[141] However, the tyrosinase inhibitor requires two properties; reduction of the quinone intermediate and no inhibition of COMT or hMAT2A. The first property can be tested through tyrosinase colorimetric screening (if quinone formation occurs the solution colours).[142] The second requirement would need a colorimetric screen of the transfer of an alkyl chain from COMT to a chosen substrate (for example DHBAL (**3**)) to determine whether inhibition occurs. There are a number of methyltransferase screens,[143-145] however, further development may be necessary to improve the versatility of them.

Further developments on the coupled enzyme systems (MAT-COMT and tyrosinase-COMT) produced in this study would enable more selective alkylations of more substrates. Improvements in screening methods could enable the production of a variety of alkylated therapeutic peptides, a library of which could be developed to improve the pharmaceutical properties of the peptide (for example goserelin or oxytocin). Discovery of an improved inhibitor for the tyrosinase enzyme and optimisation of the hMAT2A enzyme would be the key improvements needed for future development of regioselective alkylation.

4 Understanding the active site of N-methyltransferases through optimisation of a methyltransferase screen

4.1 Introduction

4.1.1 Existing methyltransferase screens

The previous section (Section 3.3) had suggested that a methyltransferase screen would improve the ability to identify successful alkylations in a more facile and quicker way than the current HPLC methods employed in this study. Investigation was conducted of the existing literature to determine what the advantages of existing methyltransferase high throughput screens (HTSs). Once the advantages of existing screens were examined, the decision of whether development of a novel screen was necessary to be developed.

Recent literature has shown that there have been developments of methyltransferase HTSs, some of which have been commercialised.[143-145] Current HTSs focus on the side product of the methyltransferase activity SAH (**1**). **1** was selected for these screens, as this molecule is universally present in every methyl and even alkyl transferase reaction. The HTSs then screen the specific breakdown of **1** with the introduction of different enzymes and probes. The two main enzymes used in methyltransferase HTSs, for the breakdown of **1**, were SAH nucleosidase and SAHH (previously mentioned in Section 3.2.4).[146,147] SAH nucleosidase breaks **1** down into adenine and *S*-ribosyl-homocysteine (Scheme 22).[146] Whereas SAHH, as shown in Scheme 22, breaks **1** down into L-homocysteine (**2**) and adenosine. [147]Using these two enzymes a number of methyltransferase screens have been developed.
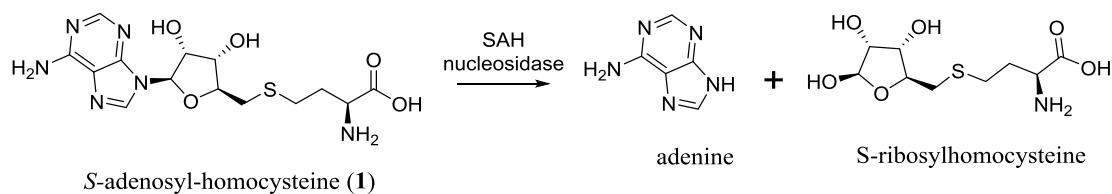
Scheme 24: SAH nucleosidase breakdown of SAH (**1**) to adenine and S-ribosylhomocysteine.[146]

In a study conducted by Biastoff *et al* of putrescine *N*-methyltransferases,[143] the SAH was broken down with SAH nucleosidase and the subsequent *S*-ribosyl-homocysteine was converted to L-homocysteine (**2**) with the *S*-ribosyl-homocysteine lyase enzyme (LuxS). **2** was then detected by the thiol probe 5, 5'dithiobis-2-nitro-benzoic acid (Scheme 25).[143] Whereas, using the SAH hydrolase enzyme (Scheme 22), Collazo *et al* measured the **2** concentration directly with the ThioGlo 1 probe.[148] This reaction was enabled determination of accurate kinetic parameters for the histone methyltransferase CLR4.[148] This work in turn was built on by Palmer *et al* for development with caffeic acid O-methyltransferases.[149] Therefore, both SAH nucleosidase and SAHH have given rise to a number of different screens with different methyltransferases.

The SAHH screens offer the most versatility of the current methyltransferase screens with the ability to immediately detect the L-homocysteine (**2**) molecule with associated probes.[148,149] Of these studies Collazo *et al*, was the most promising with direct reaction of **2** with Thio Glo 1 and the need for only one enzyme.[148] However, Thio Glo 1 was not found to be readily commercially available and the aim of this study was to produce a facile HTS, therefore an alternative probe was sought to use with the SAHH enzyme.
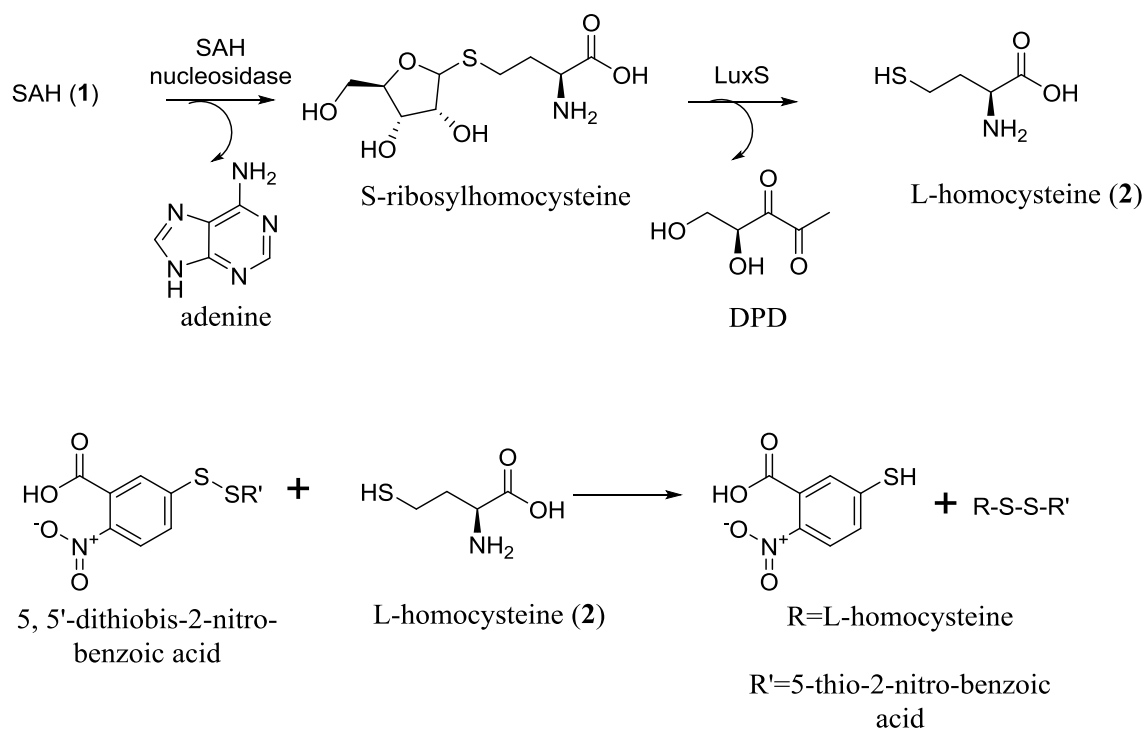
Scheme 25: Biastoff *et al* HTS screen with the SAH nucleosidase and LuxS enzymes with a 5,5'-dithiobis-2-nitro-benzoic acid probe to detect L-homocysteine (**2**).[143]

4.1.2 CNMT and TNMT two closely related methyltransferases

Development of a methyltransferase HTS will enable the understanding and improvement of methyltransferase both in substrate and cofactor scope. The successful understanding of the COMT active site meant that the regioselectivity of the COMT enzyme could be improved. The coclaurine-*N*-methyltransferase (CNMT) and tetrahydroprotoberberine-cis-*N*-methyltransferase (TNMT) enzymes have not been as well studied as the COMT enzyme. The investigation of the CNMT and TNMT enzymes using a facile HTS could lead to an understanding of both enzymes, which previously have not been well characterised.

The cDNA of both TNMT and CNMT enzymes were isolated, transformed, expressed and purified in the early 2000s.[21,150] The TNMT and CNMT enzymes were identified as Class I methyltransferases and were found to be very similar in protein sequence (amino acid residue similarity of 71%) (Figure 58).[150] CNMT was determined to be involved in the biosynthesis of isoquinoline alkaloids and has been shown to methylate the natural substrate coclaurine (Figure 59).[21] TNMT was also

found on isoquinoline alkaloid pathways, but unlike CNMT was later on the pathways and methylated the *S*-stylopine (Figure 59).[150]

```
TNMT    6   EVKKESAGETLGRLLKGEIKDEELKKLIKFQFEKRLQWGYKSSHQEQLSFNLDFIKSLKK   65
            + KK +  E L +L  G +  +++K+LI +   +RLQWGYK +++EQ++    +    SL++
CNMT    7   QTKKAAIVELLKQLELGLVPYDDIKQLIRRELARRLQWGYKPTYEEQIAEIQNLTHSLRQ   66


TNMT   66   MEMSGEIETMNKETYELPSEFLEAVFGKTVKQSMCYFTHESATIDEAEEAAHELYCERAQ  125
            M+++  E+ET++ +  YE+P EFL+ + G   +K S CYF  +S T+DEAE A  +LYCERAQ
CNMT   67   MKIATEVETLDSQLYEIPIEFLKIMNGSNLKGSCCYFKEDSTTLDEAEIAMLDLYCERAQ  126


TNMT  126   IKDGQTVLDIGCGQGGLVLYIAQKYKNCHVTGLTNSKAQVNYLLKQAEKLGLTNVDAILA  185
            I+DGQ+VLD+GCGQG L L++AQKYKNC VT +TNS +Q  Y+ +++ +  L NV+   LA
CNMT  127   IQDGQSVLDLGCGQGALTLHVAQKYKNCRVTAVTNSVSQKEYIEEESRRRNLLNVEVKLA  186


TNMT  186   DVTQYESDKTYDRLLMIEAIEHMKNLQLFMKKLSTWMTKESLLFVDHVCHKTFAHFFEAV  245
            D+T +E   +TYDR+L+IE   EHMKN +L ++K+S W++K+ LLF++H+CHKTFA+ +E +
CNMT  187   DITTHEMAETYDRILVIELFEHMKNYELLLRKISEWISKDGLLFLEHICHKTFAYHYEPL  246


TNMT  246   DEDDWYSGFIFPPGCATILAANSLLYFQDDVSVVDHWVVNGMHMARSVDIWRKALDKNME  305
            D+DDW++ ++FP G   I +A+   LYFQDDVSVV+HW ++G H +R+ +  W K LD N++
CNMT  247   DDDDWFTEYVFPAGTMIIPSASFFLYFQDDVSVVNHWTLSGKHFSRTNEEWLKRLDANLD  306


TNMT  306   AAKEILLPGLGGSHETVNGVVTHIRTFCMGGYEQFSMNNGDEWMVAQLLFKKK   358
              K +    +G   E V  ++ + R FC+ G E F  NNG+EWM + +LFKKK
CNMT  307   VIKPMFETLMGNEEEAVK-LINYWRGFCLSGMEMFGYNNGEEWMASHVLFKKK   358
```

Figure 58: BLAST results of CNMT and TNMT sequences. BLAST showed amino acid similarity of 71%.

The natural substrate of both enzymes has been uncovered (Figure 59) and some studies have built on this work to determine the substrate scope of these enzymes. CNMT has been shown to accept a number of different isoquinoline structures (Figure 60),[151] whilst the substrate scope of TNMT has not been as thoroughly investigated (Figure 60).[150] Despite the limited number of TNMT investigations, the existing substrates of TNMT have been shown either to be racemic or to have the *S* stereoisomer chemistry at position 1 of the tetrahydroisoquinoline substrates.[150] The methylation of only the *S* isomer could have suggested that the TNMT enzyme was enantiospecific.[150] The selectivity of TNMT compared to CNMT could be described as unsurprising as TNMT was found later on alkaloid pathways and would therefore, have less evolutionary need to be multifunctional unlike CNMT. However, the sequences were determined to be similar for both TNMT and CNMT enzymes. In

order to understand what governs the selectivity of the enzymes, further studies are needed on the active sites of both enzymes.
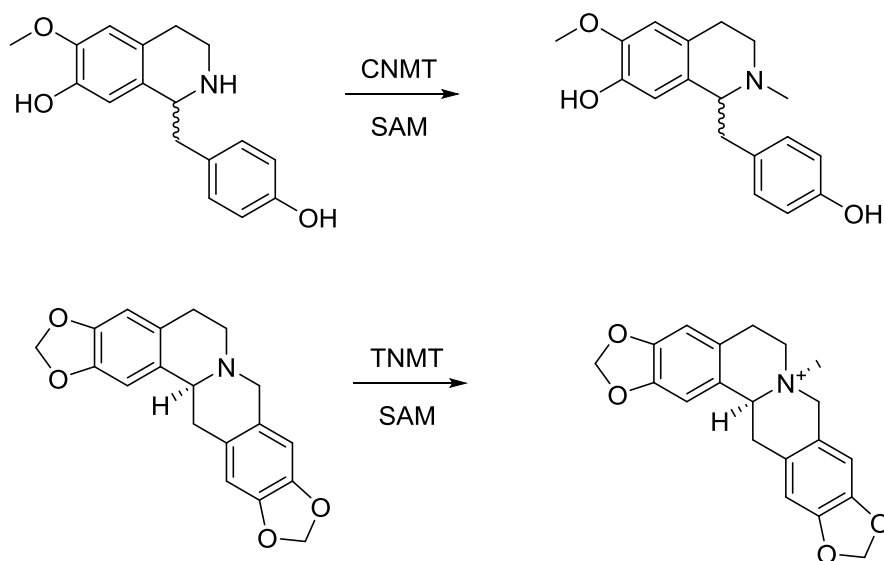
Figure 59: (Top) CNMT methylation of coclaurine. (Bottom) TNMT methylation of stylopine.[21,150]

(*R,S*)-coclaurine

(*R,S*)-norreticulene

(*R,S*)-norlaudanosoline

CNMT Substrates

heliamine (**13**)

6-*O*-methyl-norlaudanosoline

1-methyl-heliamine (**14**)

(*R,S*)-stylopine

(*R,S*)-canadine

(*R,S*)-tetrahydropalmatine
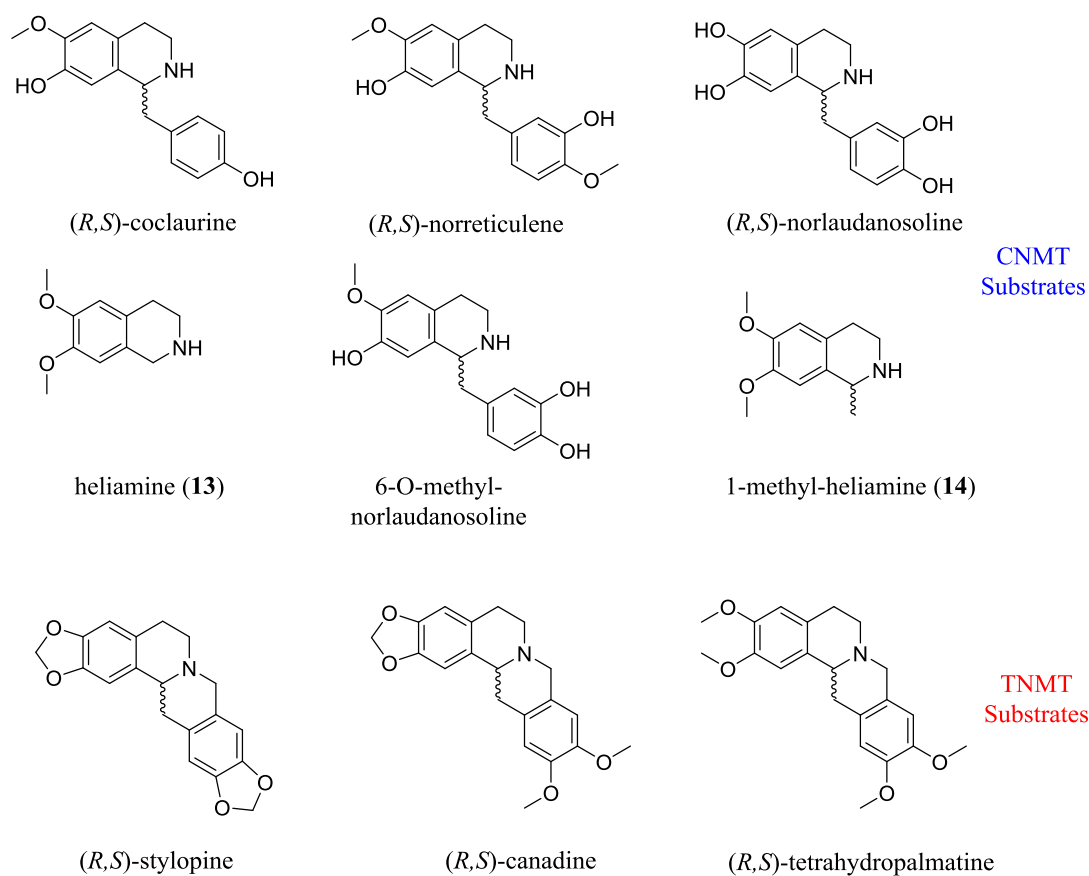
TNMT Substrates

Figure 60: Substrate scope of CNMT (blue) and TNMT (red) enzymes. [150,151]

Previous studies have uncovered some detail on the substrate scope of CNMT and TNMT, as well as suggesting an enantiospecificity for TNMT. However, as yet, no studies have focussed on the enzymes themselves with little to no information about the active site of both enzymes. TNMT and CNMT have both been shown to be active with drug like molecules suggesting that the enzymes could be used to develop APIs. Therefore, understanding of their activity, through active site mutagenesis and structural studies, and what influences the interactions between the enzymes and the substrate will be important to further development of these enzymes.

4.2   Results and discussion

4.2.1   Development of a colorimetric methyltransferase screen

The development of a colorimetric screen for methyltransferase has been the aim of a number of investigations. However, current screens require two enzymes or probes that require certain instrumentation to quantify the results.[143,148] In order to investigate whether a more simple and facile methyltransferase screen could be developed a one enzyme and an associated colorimetric probe will be sought. Once a screen has been developed, the screen can be used to determine activity of the *N*-methyltransferases; TNMT and CNMT with a range of substrates.

Previously Huo *et al* have uncovered the molecule 7-nitro-2, 3-dihydro-1*H*-cyclopenta[*b*]chromen-1-one (NDCC) (**12**) that could selectively react with the thiol containing amino acids (L-homocysteine (**2**) and L-cysteine).[152] The reaction with thiols and NDCC (**12**), leads to a ring opening of **12** that produced a yellow colour that could be measured at 405nm (Scheme 26).[152] Alternatively Ellman's reagent could be used (Scheme 26), however, **12** was shown to be easily synthesised and detectable, so was trialled.
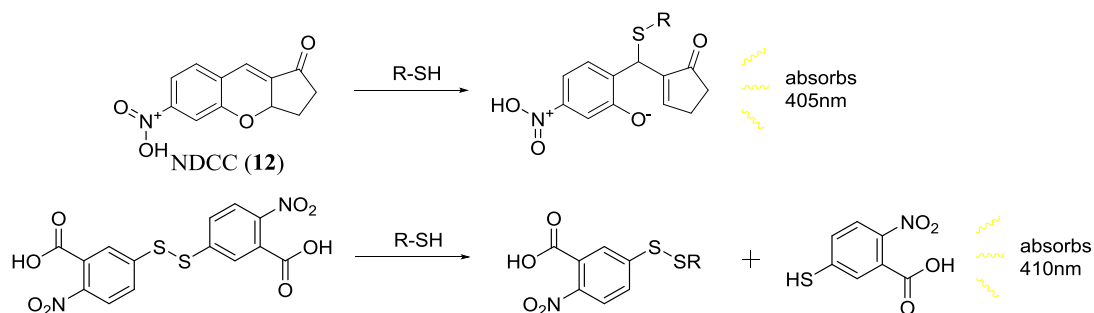
Scheme 26: (Top) NDCC (**12**) reaction with thiol.[152] (Bottom) Reaction of Ellman's reagent with thiol.
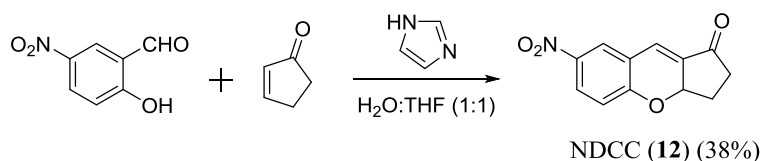
Scheme 27: Synthesis of NDCC (**12**).[152]

The synthesis of the NDCC (**12**) molecule was a facile Baylis-Hillman reaction and **12** was isolated in a modest yield (38%), after the reaction in aqueous ambient conditions (Scheme 27).[152] After purification and isolation of NDCC (**12**), **12** was tested with L-homocysteine (**2**), SAH (**1**) and SAM (**1a**). Addition of **2** to NDCC (**12**) in phosphate buffer immediately led to a colour change of colourless to yellow, whilst with **1** and **1a** no observable change was noted after 4 hours with **12**. The lack of colour change for **1** and **1a** was a positive start, as no background signal from these molecules would be detected and therefore, absorbance at 405nm would be mostly due to **2** rather than background noise.

SAM (**1a**) and SAH (**1**) had shown no reaction with NDCC (**12**) meaning that assay conditions could be sought for developing a colorimetric assay with **12** and the SAHH enzyme. Initially, the reaction of L-homocysteine (**2**) and NDCC (**12**) (25 µM) were measured on the Synergy HT Microplate Reader (plate reader) with a range of L-homocysteine (**2**) concentrations (2 µM-20 µM) over 25 minutes. The results of the experiment showed that **12** reacted quickly with **2**, with maximum absorption observed after a minute. Figure 61 shows that absorbance did increase with increasing concentrations of **2**. However, the increase of absorbance with **2** may not be linear, as shown by the R^2 for the linear trend line, suggesting that the assay may not be quantitative.
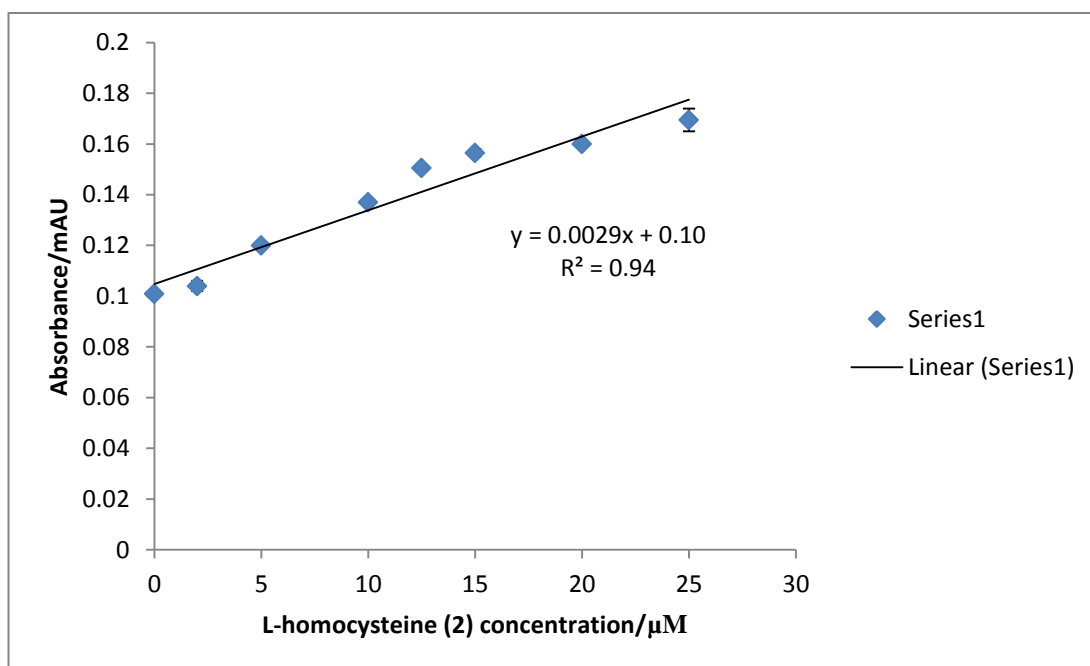
Figure 61: Reaction of NDCC (**12**) with L-homocysteine (**2**) (2-20 µM) over 25 minutes.

SAHH has previously been mentioned in optimisation of ethyl vanillin (**3c**) production (Section 3.2.4) and the gene was cloned, expressed and purified by Dr Mark Thompson. The successful purification of SAHH meant that the enzyme could be tested as a HTS screen for methyltransferases. Three enzymes were chosen to test the methyltransferase screen; COMT, CNMT (from *Coptis Japonic*) and TNMT (from *Papaver somniferum*) with their associated substrates (Figure 62). These three enzymes were chosen for two main reasons. Firstly, the three enzymes were chosen due to them being small molecule methyltransferases and secondly COMT has been well studied and can be used as a model system whereas CNMT and TNMT's substrate scope and activity has not been thoroughly investigated (Figure 59).

1-phenyl-heliamine (**15**)
1-phenyl-6,7-dimethoxy-1,2,3,4-tetrahydroisoquinoline
TNMT

heliamine (**13**)
6,7-dimethoxy-1,2,3,4-tetrahydroisoquinoline
CNMT

DHBAL (**3**)
COMT
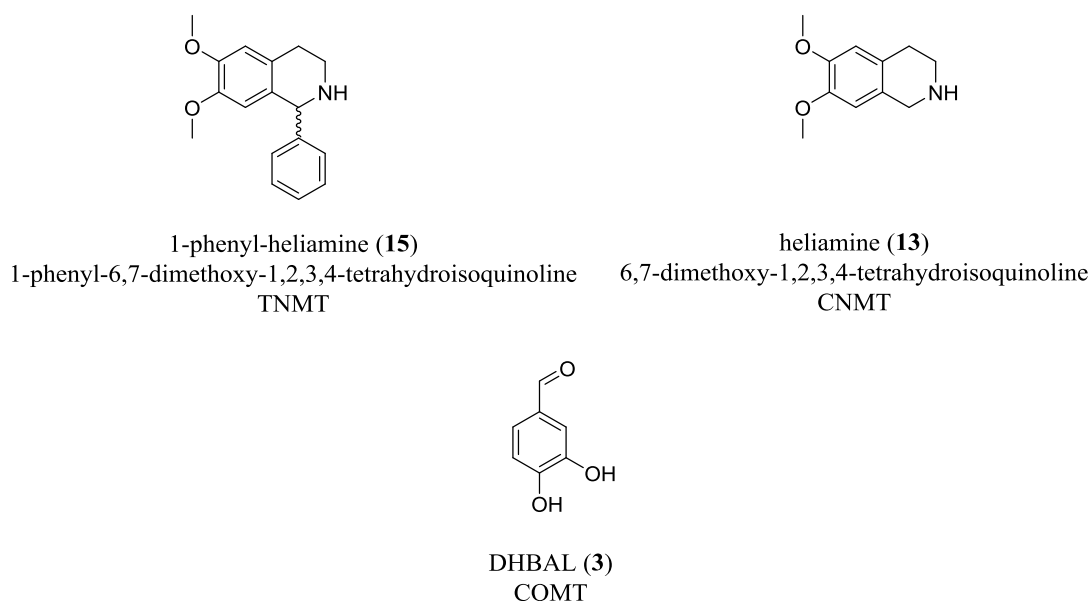
Figure 62: Substrates and their associated enzymes tested for the NDCC (**12**) and SAHH colorimetric assay.

Assays were run of the three methyltransferases (CNMT, COMT and TNMT at a concentration of 1 μM), CNMT and TNMT were cloned, expressed and purified by Dr Mark Thompson, and were assayed with both known and unknown substrates for the enzymes (Figure 62) at 25 μM and the SAM (**1a**) cofactor (25 μM). The COMT assay was run without DTT to limit thiols in the assay. The methyltransferases were separated from the assay mix with a centricon at 4°C. The methyltransferases were not precipitated out with heat or methanol, as heat could lead to SAM (**1a**) and SAH (**1**) degradation, and methanol would lead to deactivation of SAHH. After separation the filtrate from the centricon was collected and divided, one set was run on plate reader with NDCC (**12**) (25 μM) and SAHH (1 μM) with a no enzyme control (SAHH+**12**), and the other was run on the HPLC for comparison. Table 13 shows the results of the plate reader assay. TNMT was active with conversion calculated, using the L-homocysteine (**2**) calibration, showing a rough conversion of 42%. 1-phenyl-heliamine's (**15**) activity was novel and increased the substrate scope of TNMT, the result was confirmed by LCMS of the assay, which showed the characteristic mass of 284.2 for 1-phenyl-*N*-methylheliamine (**15a**) (expected mass 284.2).

Table 13: Initial colorimetric assay results of NDCC (**12**) and SAHH with COMT, CNMT and TNMT enzymes. HCY=L-homocysteine (**2**). Control adjusted concentration was determined through subtraction of the background control signal (which had been determined as 5.2 µM for this experiment).

	Control	COMT	CNMT	TNMT
Average Absorbance/AU	0.12	0.12	0.12	0.15
Homocysteine (**2**)/µM	5.2	4.2	5.6	15.8
Control adjusted Homocysteine concentration (**2**)/µM	-	-	0.4	10.6
Conversion/%	-	-	2	42

However, as Table 13 shows no activity to minimal activity (CNMT) was detected for CNMT and COMT. After reviewing the HPLC traces for each assay this was not consistent with the conversion to products for COMT (98%) and CNMT (39%). However, the traces also revealed that low amounts of substrates, SAM (**1a**) and SAH (**1**) were recovered from the centricon step. The low recovery of the cofactor and byproduct (**1a** and **1**) suggested that they were tightly bound to the enzyme still and were not separated in the centricon. The data also suggested that CNMT and COMT were tighter binders of **1** than TNMT, as the CNMT and COMT enzymes showed no activity with the flow through from the centricon suggesting that **1** was bound more strongly. Therefore, an alternative separation strategy was sought.

Previously with the COMT studies, precipitation of the enzyme had been shown to work most effectively with methanol or heating. However, methanol added to the assays would lead to inactivation of the SAHH enzyme in the following colorimetric assay. Heating was thought to lead to SAH (**1**) degradation, in order to determine the extent of SAH (**1**) degradation two samples of **1** at 2 mM concentration were prepared in D$_2$O. Both samples were heated to 37ºC for 25 minutes to emulate methyltransferase conditions and then one sample was heated to 95 ºC for 5 minutes, whilst the other was kept at 0ºC to prevent any degradation after the assay. Both

samples were then submitted to NMR to determine whether any new peaks were present, and whether they corresponded to degradation products. The results of these assays showed little to no degradation of the unheated sample, whereas **1** heated to 95 ºC did show new peaks at about 5% of the **1** peaks in the NMR sample. The new peaks were unable to be assigned due to their low signal, but were mostly likely due to be products of **1**'s degradation. The low amount of SAH (**1**) degradation, suggested, however, that the heating may be a viable alternative to using centricons to separate protein from the **1**.

Based on the results of the SAH degradation experiments COMT and CNMT were re run with heating to terminate the assays before running with the NDCC (**12**) and SAHH in the colorimetric assay. The results of the assays did show activity for COMT and CNMT. However, the results of the colorimetric assay (Table 13) did not match up with the conversions shown by the HPLC for the COMT or CNMT assays (Table 13). The results of the colorimetric assay; overestimated the conversion of heliamine (**13**) to 2-*N*-methylheliamine (**13a**) with the CNMT enzyme, whilst under estimating the conversion of DHBAL (**3**) for COMT. The errors associated with the assay were most likely attributable to correction factor needed for the large signal displayed in the control. The colorimetric assay activity in the controls suggested that the **12** reacted with the SAHH surface thiols. Therefore, whilst the initial development of this colorimetric assay was shown to be qualitative (shown by the larger signals observed for the active sample) the assay was not quantitative (under current conditions).

Table 14: Results of the colorimetric assay with COMT and CNMT enzyme's after heating termination. HCY=L-homocysteine (**2**).

	Control	COMT	CNMT	HPLC COMT	HPLC CNMT
Average Absorbance/AU	0.12	0.18	0.17	-	-
Calibrated HCY (**2**)/µM	5.6	24.2	20.8	-	-
Control adjusted HCY (**2**)/µM	-	18.6	15.2	-	-
Conversion/%	-	74	60	98	39

The lack of quantification suggested that the NDCC (**12**) assay could not be used to determine activities and kinetic constants of methyltransferases. However, as Figure 63 shows, with a COMT enzyme and DHBAL (**3**) substrate (and excess SAM conditions) the assay can be developed to be used a qualitative screen and can test methyltransferases' substrate scope. The screening can be achieved by running enzyme controls in parallel with the tested substrates to determine activity. Controls were observed to lead to an absorbance equivalent to 5µM, suggesting that assays could detect activity at >10% conversion. Therefore, although the NDCC (**12**) assay was shown to be limited, by the NDCC molecules' indirect activity with the SAHH surface thiols, the assay could be used as a facile qualitative test of methyltransferase activity.
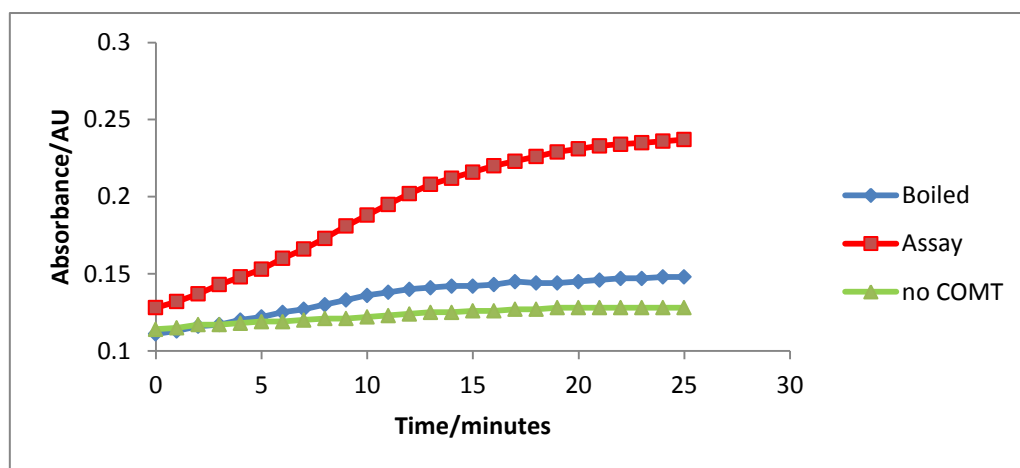
Figure 63: Graph of the NDCC (**12**) reaction across 25 minutes with the COMT assay (red), boiled COMT control (blue) and no COMT control (green).

4.2.2 Understanding CNMT and TNMT's substrate scope

The successful development of a methyltransferase screen meant that more potential substrates could be screened for the TNMT and CNMT enzymes. Section 4.1.2 showed that the potential methylated products of tetrahydroisoquinolines with the TNMT and CNMT enzymes are valuable pharmaceutical scaffolds. Investigation of what governs the selectivity of the CNMT and TNMT enzymes through mutagenesis was important in developing these enzymes to produce important industrially relevant products. Therefore, using the methyltransferase screen new substrates will be screened as well as screening of CNMT and TNMT mutants to understand more about the enzyme's active sites.

Using the methyltransferase screen, 24 substrates were screened for the CNMT and TNMT enzymes along with controls of boiled enzyme. TNMT was only additionally active with 1-phenylheliamine (**15**), whereas CNMT was active with a wide range of the substrates screened (Figure 64). The "hits" from the colorimetric assay were subsequently tested by LCMS to determine methylation (Section 7.1.10). From the substrate screen the CNMT enzyme requires; the central tetrahydroisquinoline core, an electron withdrawing group at the 6 position (MeO- or Cl) and if substituents are present on the tetrahydroisoquinoline core only in positions 6, 7 and 1. Whereas the

structural motif of TNMT substrates appears to need an aromatic substituent at the 1 position as well as methoxy substituents at the 6 and 7 positions.
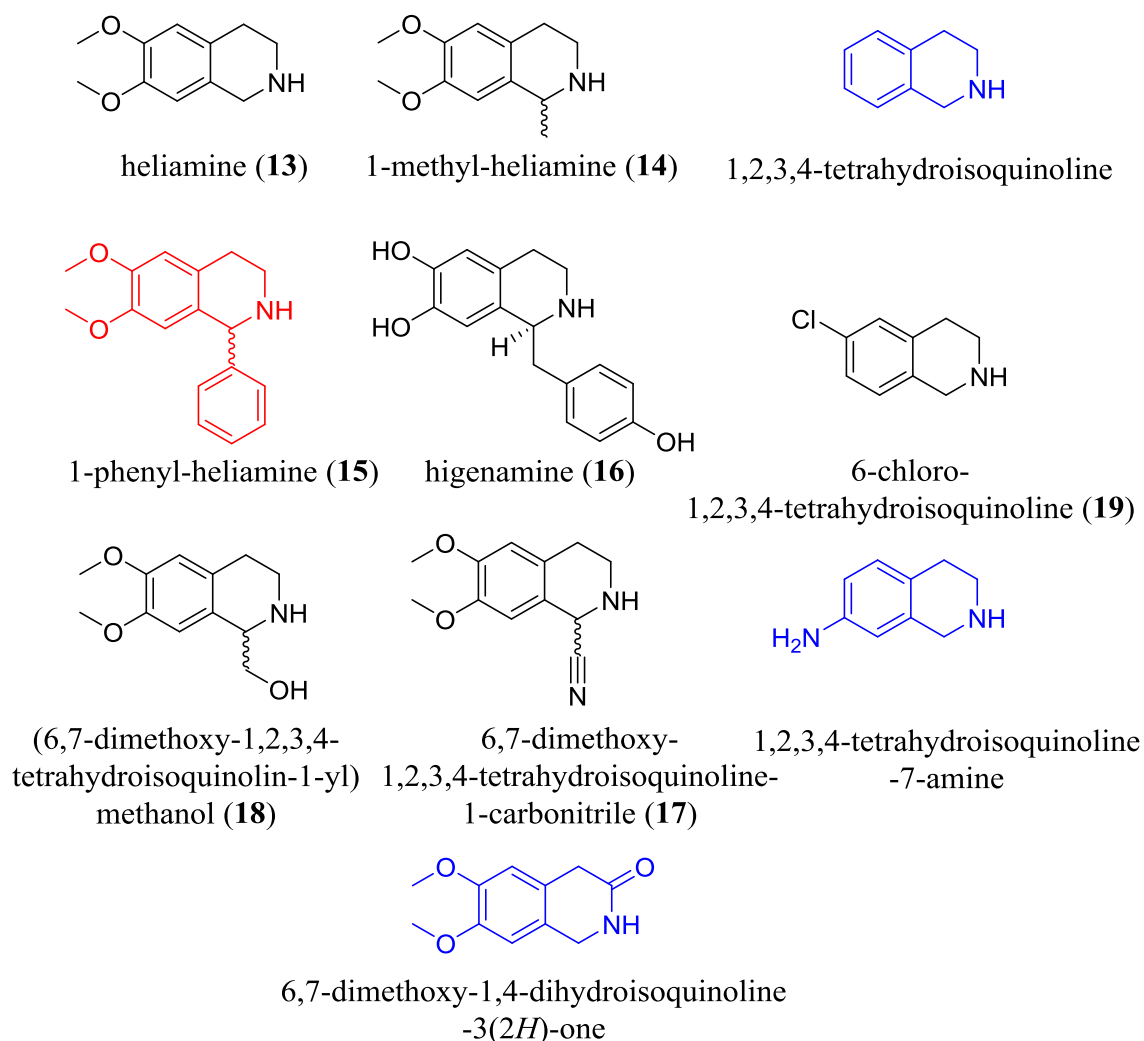
Figure 64: Substrates (**13-19**) active with CNMT, TNMT was only additionally active with 1-phenylheliamine (red). LCMS for all active substrates can be found in Section 7.1.10. Structures shown in blue were inactive.

The screen successfully identified active substrates of CNMT and suggested a motif of CNMT substrates. In order to understand further what substrates CNMT favours 4 substrates were chosen; higenamine (**16**) (close in structure to coclaurine the natural substrate of CNMT), heliamine (**13**), 1-methylheliamine (**14**) and 1-phenylheliamine (**15**) to develop conversion assays. Despite acting as a good initial screen the colorimetric assay will not be used, as the assay was determined to be a better qualitative assay than quantitative based on the existing results. Therefore, HPLC was decided on as the best analytical technique to determine activities of the four

substrates (**13**-**16**) with the CNMT enzyme (**15** was also used with TNMT) (see Section 7.1.3).

In order to set up conversion assays of CNMT, HPLC methods were required to separate methylated products from substrates. Commercial standards of the methylated products of heliamine (**13**) and 1-methylheliamine (**13a**) were available. However, commercial standards were not readily available for the methylated products of 1-phenylheliamine (**15a**) and higenamine (**16a**). *N*-methylheliamine (**15a**) was synthesised in a good yield 75% using an Eschweiler Clarke synthesis (Figure 65).[153] As **15a** was thought to be similar in polarity to **16a** and quantities of higenamine (**16**) for synthesis of methylated products was not available, HPLC methods were developed for 1-phenylheliamine (**15**) and 2-*N*-methyl-1-phenylheliamine (**15a**) first and applied to **16**. Separation of heliamine (**13**), 1-methylheliamine (**14**), 1-phenyl-heliamine (**15**) and higenamine (**16**) substrates from their methyl product standards (**13a**-**16a**) proved successful (Sections 6.5.7 and 6.5.8).
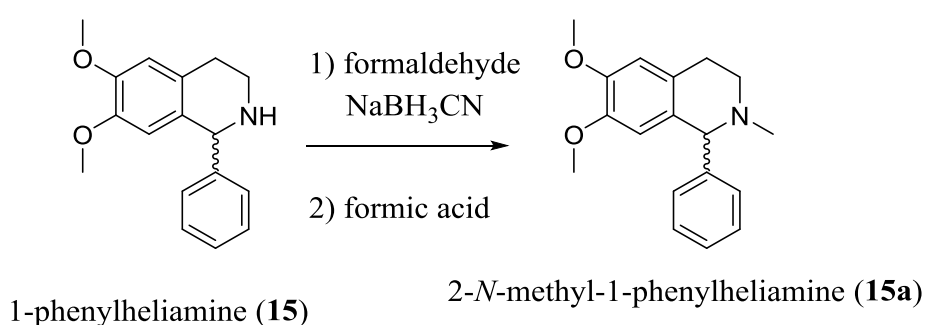
Figure 65: Synthesis of 2-*N*-methyl-1-phenylheliamine (**15a**).[153]

Successful development of HPLC methods for the 4 substrates meant that conversion assays could be developed. After some initial development the conversion assays were fixed for 30 minutes with termination with an assay volume of methanol (further details Section 6.4.11). Figure 66 shows that CNMT displayed good activity with all four substrates. Surprisingly of the four substrates tested higenamine (**16**) (closest to coclaurine in structure) showed the lowest conversion to 2-*N*-methylhigenamine (**16a**) of 62.9%, whereas 1-methylheliamine (**14**) showed the highest conversion to 2-*N*-methyl-1-methylheliamine (**14a**) of 83.2%. The high activity of **14** suggested that the CNMT substrate motif should have a substituent at position 1 (also reflected by the high conversion of 1-phenylheliamine (**15**) to 2-*N*-

methyl-1-phenylheliamine (**15a**)) but that the substituent on **16** may too large (or too polar).

The activity had shown that CNMT worked well with all four substrates, with good conversion. However, with TNMT and the 1-phenylheliamine (**15**) substrate very low conversion to 2-*N*-methyl-1-phenylheliamine (**15a**) (2.6%) was noted in 30 minutes (Figure 66). The low conversion was most likely attributable to **15** having a significantly different structure from TNMT's natural substrate (Figure 59). Alternatively, TNMT has been suggested in literature studies to be enantiospecific,[150] only active with the *S*-enantiomer.[150] The enantiospecificity of TNMT would, therefore, only accept 50% of the racemic **15** and could be inhibited by the other enantiomer. The enantiospecificity of TNMT was also suggested by the results of the colorimetric assay, which calculated an approximate 42% conversion of TNMT with the **15** substrate. Therefore, TNMT showed lower activity with the 1-phenylheliamine (**15**) substrate compared to CNMT, which could be attributable to enantiospecificity of TNMT or **15** acting as a poor substrate for TNMT.

	heliamine (13)	1-methylheliamine (14)	1-phenylheliamine (15)	higenamine (16)	1-phenylheliamine TNMT (15)
■%Products	76	83	83	63	3
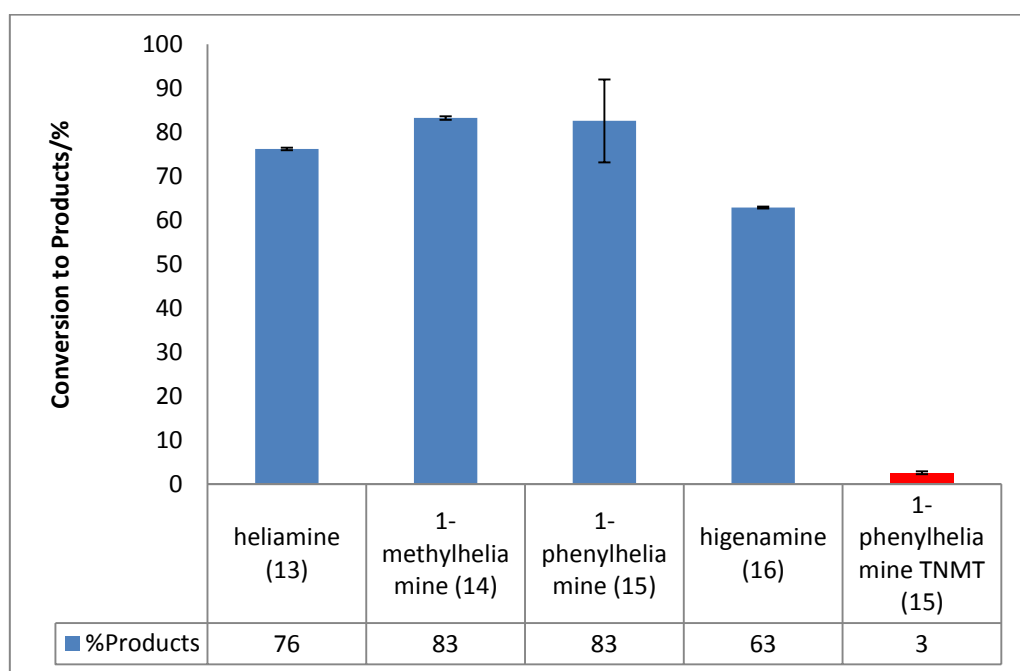
Figure 66: Activity of CNMT (blue) and TNMT (red) with substrate; 1-phenyl-helamine (**15**). CNMT was also assayed with substrates; heliamine (**13**), 1-methyl-heliamine (**14**), and higenamine (**16**). Example chromatograms can be see for all four substrates in Section 7.1.3, Figures; Figure 133, Figure 136, Figure 139 and Figure 142.

The colorimetric assay had revealed the substrate scope of the TNMT and CNMT enzymes, whilst conversion assays had suggested more about the structural motif for

CNMT substrates. Activity assays have also shown that TNMT has low activity with the 1-phenylheliamine (**15**). However, the studies conducted did not determine any structural details about the two enzymes. In order to understand the differences in these two enzymes further structural detail of the active sites were required.

4.2.3   Understanding key residues in CNMT's active site

The colorimetric assay had uncovered key information about CNMT's substrate scope and had uncovered a novel substrate for the TNMT enzyme. In order to understand more about the active site interactions with these substrates structural studies were needed. TNMT was not chosen for structural studies, due to the enzyme's low substrate scope and activity. A crystal structure of CNMT was sought to uncover key residues in the active site and from this a homology model of TNMT could be produced (due to the similar sequences).

An initial *apo* CNMT crystal structure was successfully solved by Dr Mark Dunstan (2.4Å resolution), with protein purification carried out by Dr Mark Thompson. The structure showed the binding of SAH (**1**) but no substrate was bound. The structure indicated that CNMT was a Class I Rossmann folded methyltransferase and the hydroxyl binding residue of adenosine Q163 **1** binding residues were indicated (Figure 67). However, the structure showed no further information about the binding of substrates to the CNMT structure. Therefore, the *apo* structure determined the structural features of the CNMT enzyme but did not uncover key active site-substrate interactions.
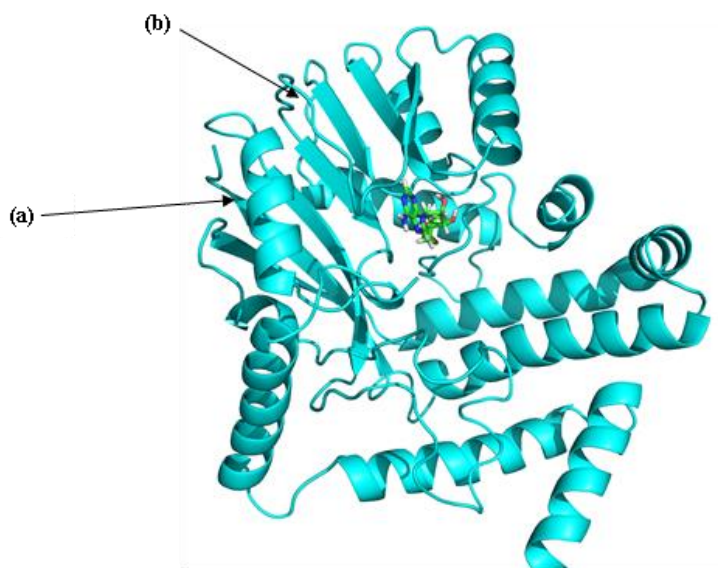
Figure 67: *Apo* crystal structure of CNMT. Key structural features of a Class I Rossmann fold methyltransferase a) antiparallel β sheet b) 6 parallel β sheets alternating β sheet α helix.

In order to overcome the shortcomings of the *apo* structure a modelling experiment was set up with the AutoDock Vina software with the heliamine (**13**) substrate and CNMT *apo* structure.[154] Using the modelling software the most stable orientation (facing the SAH (**1**) molecule) of the substrate was taken (-22.1 kJ mol^{-1}) and loaded into Pymol software (Figure 69) to produce the substrate orientation shown below.[154] The structure clearly showed the nitrogen atom of the heliamine (**13**) substrate 2.7Å away from the H208 of the CNMT enzyme. The E207 was located underneath the H208 residue. The positions of E207 and H208 relative to the substrate suggested that the CNMT structure may show "catalytic triad" behaviour (although in this case these residues would be a catalytic diad). The mechanism of the "catalytic diad" would start with E207 removing the proton from H208, which in turn removes a proton from the substrate for methylation by SAM (**1a**) (Figure 68). Catalytic triad behaviour has been noted in plant *O*-methyltransferases before by Brandt *et al*,[155] therefore, H208 and E207 may be involved in the methylation in the manner suggested.
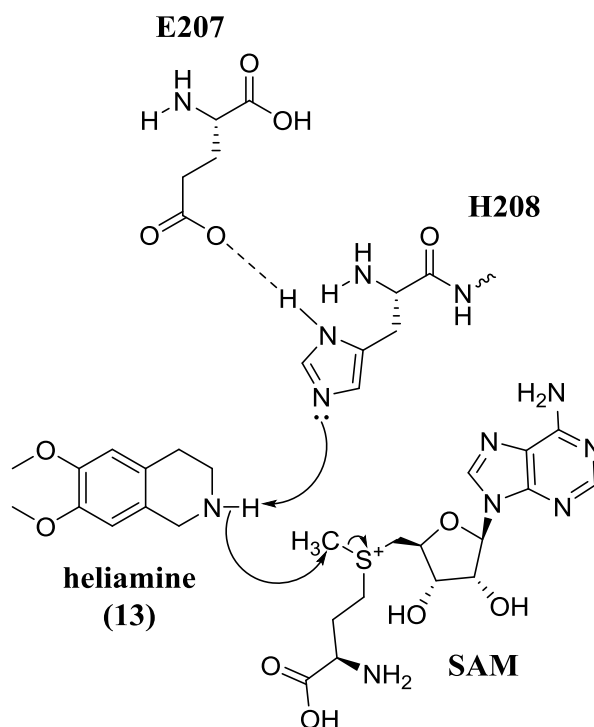
Figure 68: Suggested catalytic cycle of residues E207 and H208, with the heliamine (**13**) substrate and SAM (**1a**) cofactor.
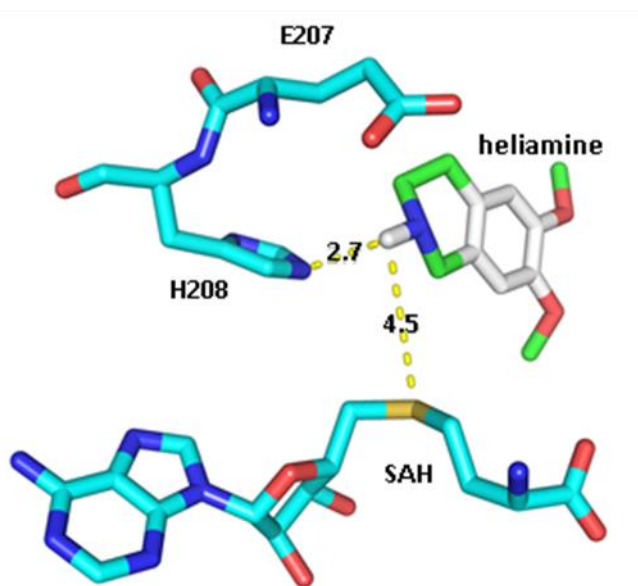
Figure 69: Modelled in heliamine (**13**) substrate into the *apo* CNMT active site. The predicted binding orientation suggested H208 and E207 are the nearest residues.

In order to test the hypothesis laid out by the modelling software the E207 and H208 residues were mutated to E207A, H208A, E207Q and H208F respectively. The mutants were chosen to remove the polarity of the E and H residues, as well as testing whether the size of the residues mattered. Activity assays were conducted on the CNMT mutants with higenamine and heliamine to determine whether activity was reduced or eliminated. The assays showed that with E207Q, H208F and H208A mutants CNMT activity was almost eliminated with both heliamine and higenamine (

Figure 70). The reduction of activity with the H208F mutation suggested that the histidine's aromaticity was not important to methylation (as H208F maintains the aromaticity but still a drop in activity was noted), suggesting that histidine's basicity was more important. These results suggested that the acidic E207 and basic H208 residues were essential for CNMT's methylation.

	WT	H208A	H208F*	E207Q	E207A
higenamine	62.9	3.3	0.0	0.0	57.1
heliamine	76.2	0.9	6.0	0.0	47.4
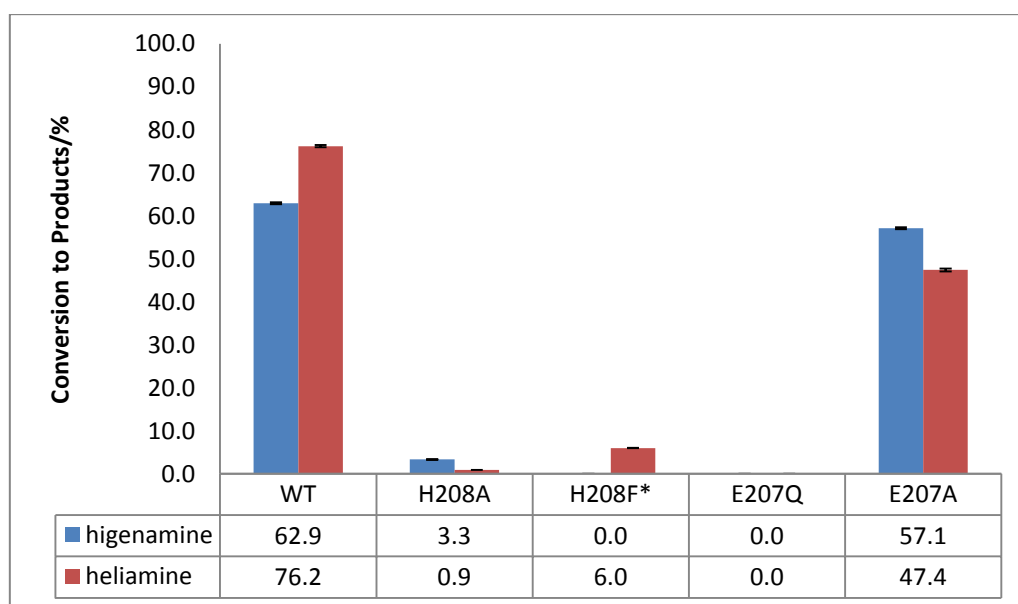
Figure 70: Higenamine (**16**) (blue) and heliamine (**13**) (red) substrates with CNMT WT and mutants; WT, H208A, H208F, E207Q and E207A.*No assay was run for higenamine H208F mutant. No activity was observed with the E207Q mutant for both substrates.

The E207Q mutation had eliminated activity whereas E207A led to a modest reduction in activity. The modest reduction in activity of E207A suggested that the size of the E207 was also important for the interaction with a water molecule. Figure 71 suggests that the E207Q elimination may be due to the size of the Q residue. The Q residue does not offer the stabilising hydrogen bonds of E207 and sits in the space of a water molecule that could be involved in the deprotonation of H208. In the

E207A mutation, the space was not occupied by the alanine residue, which allowed the water molecule to still deprotonate the histidine residue. Therefore, the observation of lower activity for E207Q did suggest that the E207 residue was involved in deprotonation of the histidine through interaction with a water molecule.
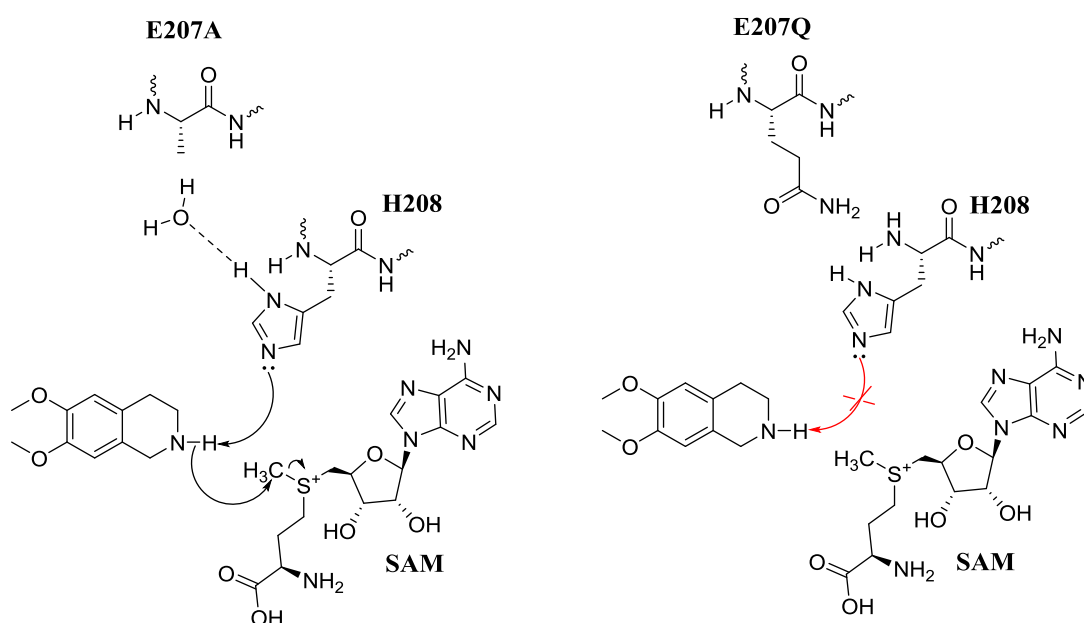
Figure 71: (Left) E207A mutant proposed methyltransferase activity. (Right) E207Q the glutamine residue blocks the water molecule deprotonating H208.

The successful identification of the catalytic diad in the CNMT structure had shown that mutagenesis coupled with modelling could be used to identify key residues in the CNMT active site. In order to gain an understanding of the substrate scope of CNMT and understand the suggested CNMT motif the modelled structure was re-examined. Observing the structure and comparing TNMT and CNMT structures there were two differences in the sequence at the potential 1 position binding site. These residues were W329 and G331 in CNMT, I329 and T331 in TNMT. A double mutant was produced for CNMT (W329I G331T) and TNMT (I329W T331G) with both mutants compared with activities with the substrate 1-phenylheliamine (**15**). The CNMT mutant was also assayed with heliamine (**13**), 1-methylheliamine (**14**) and higenamine (**16**).

	Substrate 13	Substrate 14	Substrate 15	Substrate 16	TNMT Substrate 15
WT	76	83	83	63	3
Mutant	5	3	39	3	4
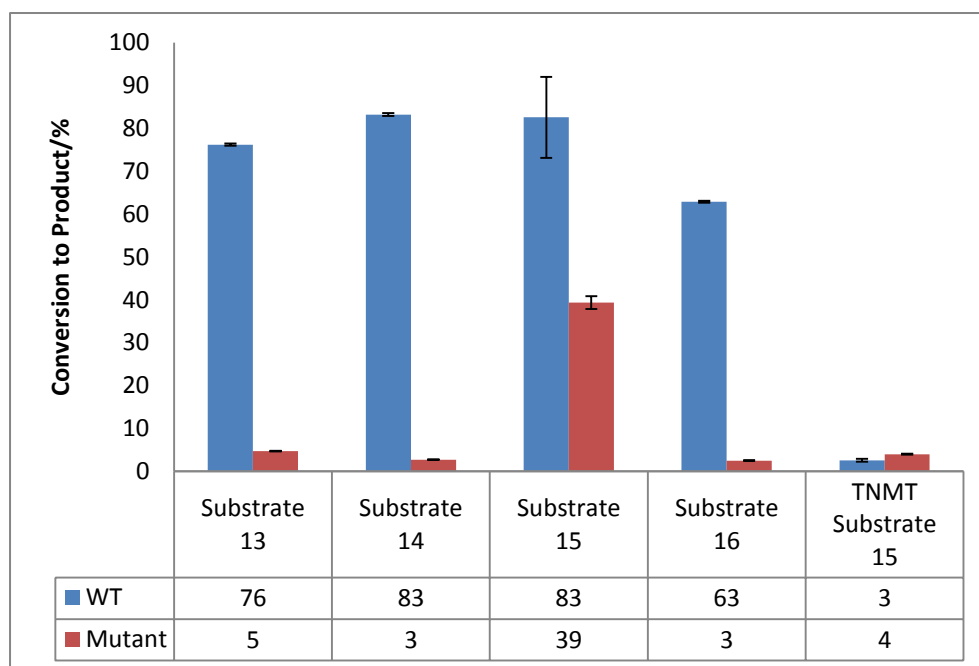
Figure 72: CNMT mutant W329I G331T conversion assay with; heliamine (**13**), 1-methylheliamine (**14**), 1-phenylheliamine (**15**) and higenamine (**16**). TNMT mutant I329W T331G activity with 1-phenyl-heliamine (**15**).

The conversion assays of the TNMT and CNMT mutant showed that mutation of the residues in CNMT led to a reduction in activity with all the substrates whereas TNMT showed an increase in activity for 1-phenylheliamine (**15**) (Figure 72). The conversion of the **15** substrate by the W329I G331T double mutant of CNMT showed the least decrease of 44% compared to conversions of the other substrates (**13**, **14** and **16**), which displayed reductions in their conversions of at least 60%. The decrease in activity for CNMT W329I G331T mutant suggested that these residues were may be involved in binding the tetrahydroisoquinoline aromatic ring. The smaller decrease for **15** may be due to the mutation decreasing steric clashes between the 1 position phenyl ring and the W329 position, which compensates for the loss of interaction with the G331 and W329 residues.

The decreased activity with the CNMT mutant W329I and G331T suggested that the W329 and G331 residues might be important for binding the substrates of CNMT. The increased activity for the TNMT I329W T331G mutant with the 1-phenylheliamine was also observed. The activity of the TNMT mutant with 1-phenylheliamine substrate suggested that CNMT activity could be engineered for the TNMT enzyme. Therefore, the other substrates (1-methylheliamine (**14**), heliamine (**13**) and higenamine (**16**)) were also tested with the TNMT mutant but activity was

still not observed for these substrates. The lack of activity with the other substrates suggested that further understanding of the TNMT active site was also needed to engineer CNMT substrate scope.

Identification of W329 and G331 as possible key binding residues for CNMT substrates had added further details about the CNMT active site. However, the double mutant did not show, which residue had the stronger interactions to the tetrahydroisoquinoline substrate, as well as not determining where on the substrate the residues bound. Further single point mutations were conducted on residues in the vicinity of W329 and G331 introducing an alanine to each position; Y328A, W329A, R330A, G331A and F332A (Figure 73). Alanine was chosen for all mutations, as the alanine residue was a small size and neutral charge, which would show whether the charge or size of the original position was important.
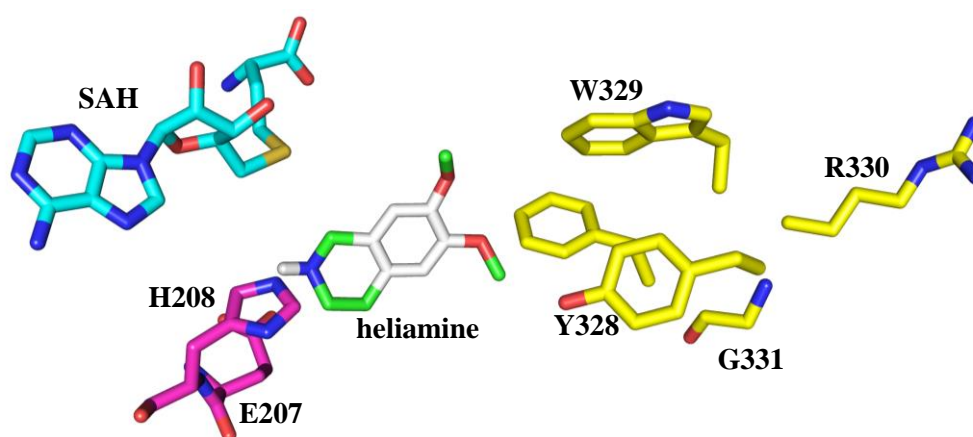
Figure 73: CNMT structure with modelled in heliamine (**13**), catalytic diad is shown in purple (H208 and E207). Suspected substrate binding wall shown in yellow (W329, Y328, F332, G331 and R330).

With the aim of determining which residues were important for binding the tetraisoquinoline substrates; each mutant was tested under conversion assay conditions (full conditions recorded in Section 6.4.11) with the heliamine (**13**) substrate. **13** was chosen for the absence of a substituent, to determine which residues were important for binding the tetrahydroisoquinoline rings of the substrate. The mutants all showed a reduction in activity, however, there were only two mutants that led to a significant reduction in activity; W329A (29% reduction) and F332A (53% reduction), as shown in Figure 74. The significant reduction in activity of CNMT and these two residues suggested that W329 and F332 binds the ring of the tetrahydroisoquinoline substrate through π stacking interactions.

	Y328A	W329A	R330A	G331A	F332A	H208A	WT
■ Series1	60	31	65	61	17	1	70
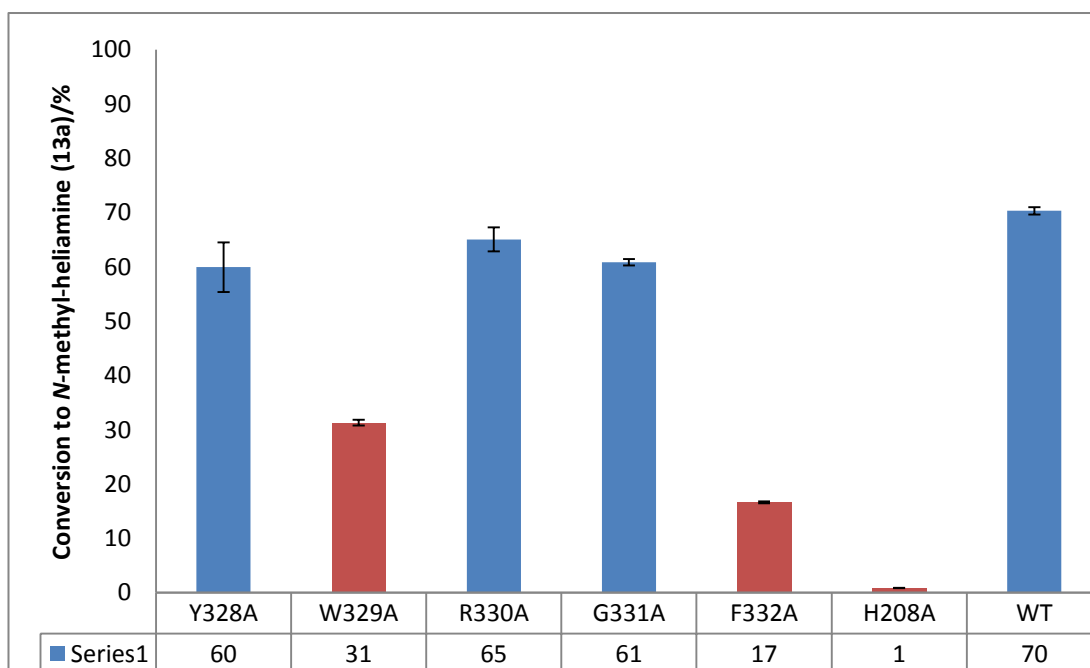
Figure 74: Conversion assay of CNMT mutants with heliamine (**13**).

The results of the heliamine (**13**) conversion assay with the alanine mutants (Y328A, W329A, R330A, G331A and F332A) had suggested that W329 and F332 residues may bind the aromatic ring of the tetrahydroisoquinoline substrates. The higenamine (**16**) substrate was also assayed with the mutants to determine whether the phenolic substituent at the 1 position and free hydroxyls on the aromatic ring would have different interactions with the; Y328, W329, R330, G331 and F332 positions (compared to the **13** substrate).

The results of the conversion assay with higenamine (**16**) showed reductions for Y328A, W329A and F332A (Figure 75). Reductions were also noted for R330A and G331A but these reductions were not significant, and these residues were determined to not be significantly involved in substrate binding. The reduction of activity of F332A was similar to the results observed with heliamine (**13**) suggesting that this residue was primarily involved in binding the aromatic ring of the tetrahydroisoquinoline. Whereas Y328A mutant showed a reduction in activity with the **16** substrate that was significant and larger than the observed reduction for the **13** substrate. W329A showed even more reduction in activity for **16** (57%) than W329A had shown with the **13** substrate (29%). Re-examination of the CNMT active site with the higenamine (**16**) substrate was needed to determine how W329 and Y328 interact with **16**.

	Y328A	W329A	R330A	G331A	F332A	WT
Activity /%	38	6	51	58	9	63
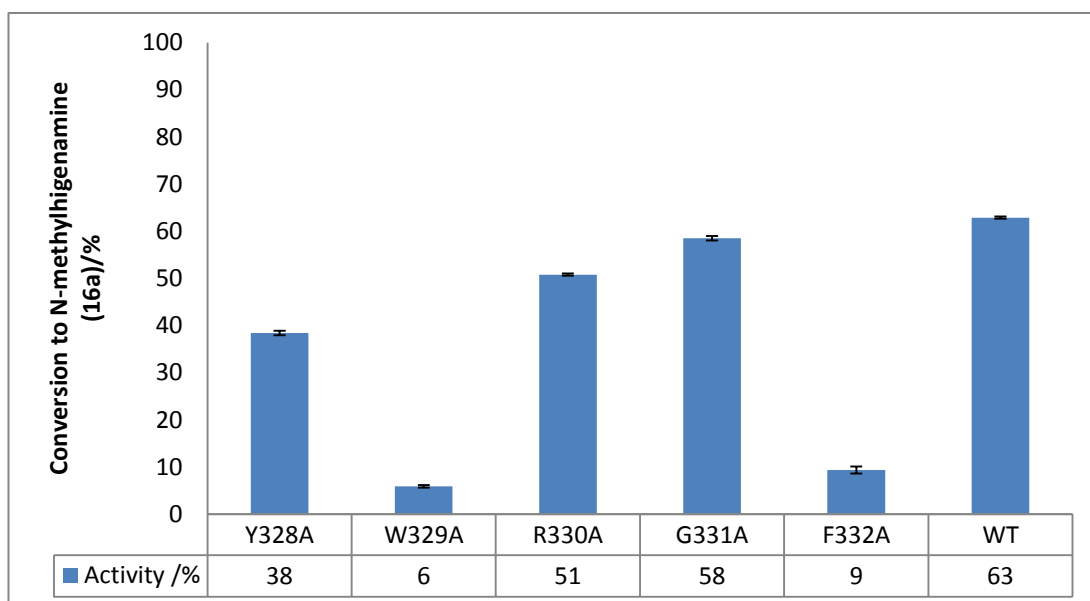
Figure 75: Activity of CNMT mutants with higenamine (**16**).

Modelling studies were conducted with the higenamine (**16**) substrate to determine what role the W329 and Y328 residues played in **16** binding that was distinct from heliamine (**13**). The modelling experiment was set up as mentioned at the beginning of this section and the resulting binding mode showed an -25.9 kJ mol^{-1} binding energy.[154] The higher binding energy calculated by the program might be due to higenamine being closer in structure to coclaurine (CNMT's natural substrate).

The binding model of higenamine (**16**) showed a different binding orientation compared to heliamine (**13**) (Figure 76). The model suggested that the higenamine substrate was much closer to the residues W329, Y328 and F332. The structure suggested that a hydrogen bond forms between Y328 and the free hydroxyls of the higenamine substrate. The loss of the hydrogen bond with the Y328A mutant suggested there was weaker binding with the higenamine (**16**) substrate and could have accounted for the larger reduction in activity with the **16** substrate compared to the heliamine (**13**) substrate. Whereas, W329 was thought to form π stacking interaction between indole ring of W329 and the aromatic ring of the tetrahydroisoquinoline substrate. The interaction between W329 and substrate would be stronger with the **16** substrate compared to the **13** substrate due to the smaller distance (suggested by the model) that occurs with **16** binding (compared to **13**). Therefore, Y328A and W329A mutants were thought to have a bigger reduction in the activity of CNMT with the higenamine (**16**) substrate, due to the suggested

alternative orientation shown by modelling studies (when compared with heliamine (**13**) (Figure 76).
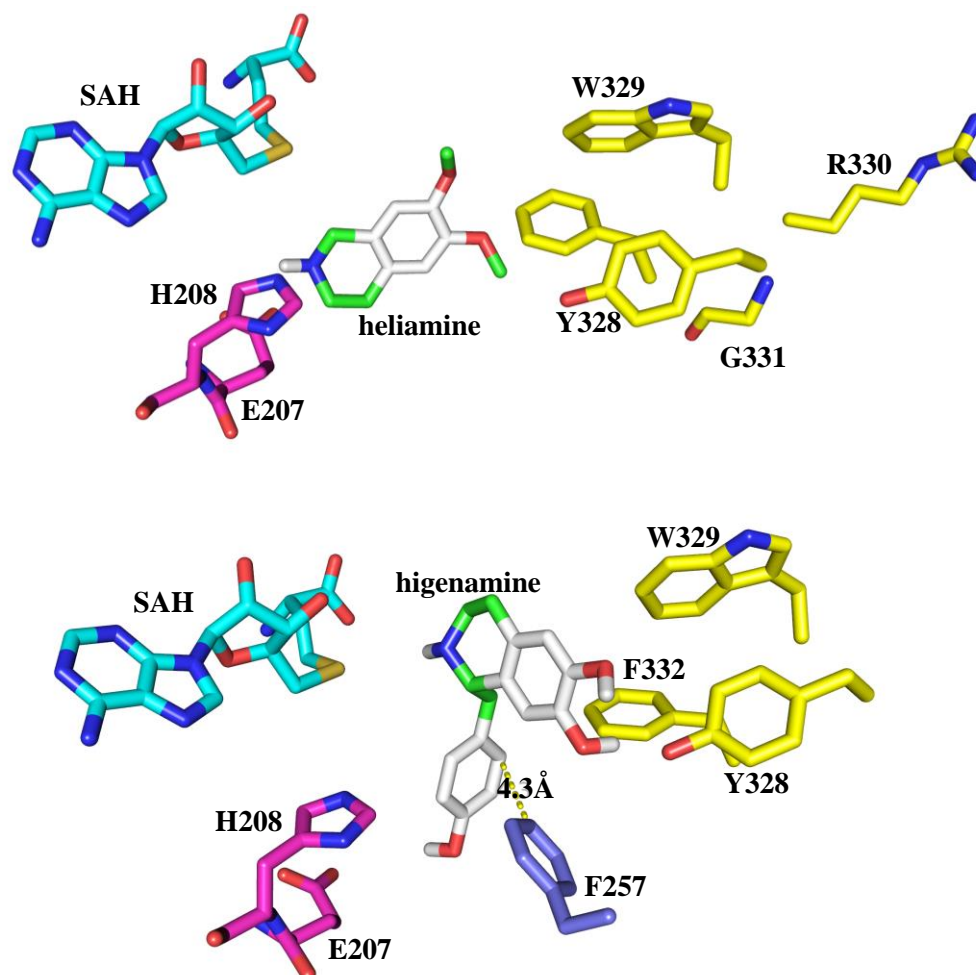
Figure 76: (Top) The heliamine (**13**) substrate docking model. (Bottom) The higenamine (**16**) substrate (-25.9 kJ mol^{-1}) modelled into the *apo* CNMT structure. Substrate binding residues are shown in yellow (W329, F332 and Y328), the catalytic diad is shown in purple (H208 and E207). Proposed 1 position binding residue shown in dark blue (F257). No clashes were noted in either model.

However, the higenamine (**16**) conversion assays with these mutants did not account for any 1-position binding. The model suggested that future studies should focus on F257, which was only 4.3Å away from the 1-position phenolic group, as shown in Figure 76. Understanding the 1-position binding residues was thought to be important in determining the differences between TNMT and CNMT, as the substrate scopes showed that the selectivity of each enzyme was governed by the substituents at this position.

The colorimetric assay had shown that the substrate scope of a methyltransferases could be investigated and even broadened. The screen had improved the understanding of the scope of CNMT and TNMT enzymes. From the understanding of the substrate scope of the CNMT enzyme, an understanding of the active site interactions between substrate and the enzyme were investigated. Modelling and conversion assay studies successfully determined that catalytic diad (E207 and H208) catalyse the methyl transfer and a "π-wall" (W329, F332 and Y328) was responsible for binding the aromatic ring of the tetrahydroisoquinoline substrates.

4.3   Conclusion and future work

The colorimetric assay investigated the substrate scopes for the TNMT and CNMT enzymes. However, the stumbling block for this HTS proved to be the lack of quantification. The qualitative data provided by the screen enabled further study of the CNMT enzyme. Despite the understanding of the CNMT active site developed by this study, no improvement in selectivity or activity was developed for the CNMT enzyme. Future studies are needed to improve the problem of a lack of quantification for the colorimetric screen and the improvement of selectivity or activity for CNMT.

The lack of the quantification shown in Section 4.2.1 suggested that the NDCC screen requires further optimisation. Optimisation of the initial conditions, could allow SAHH to run for the set 25 minutes before termination of the assay with methanol, then addition of the NDCC (**12**) probe for the reaction with the L-homocysteine (**2**) produced. This would slow the speed of the reaction but may improve quantification, by removal of the SAHH through precipitation in methanol there would be less background reactions.

Alternatively the selectivity of the thiol probe could be improved by trying to develop probes for selectivity of L-homocysteine (**2**) over glutathione. The larger nature of the L-glutathione makes this a good thiol model of the surface thiols of the SAHH. Using the NDCC (**12**) or Ellman's reagents as scaffolds varying substituents could be trialled at the 8 position to block the larger thiol from the reaction necessary to release the nitrophenolate (Figure 77). Therefore, improvement of the selectivity of the thiol probes, through functionalization of the probe or optimisation of the

assay conditions would be a useful step towards improvement of the quantification of the colorimetric assay.
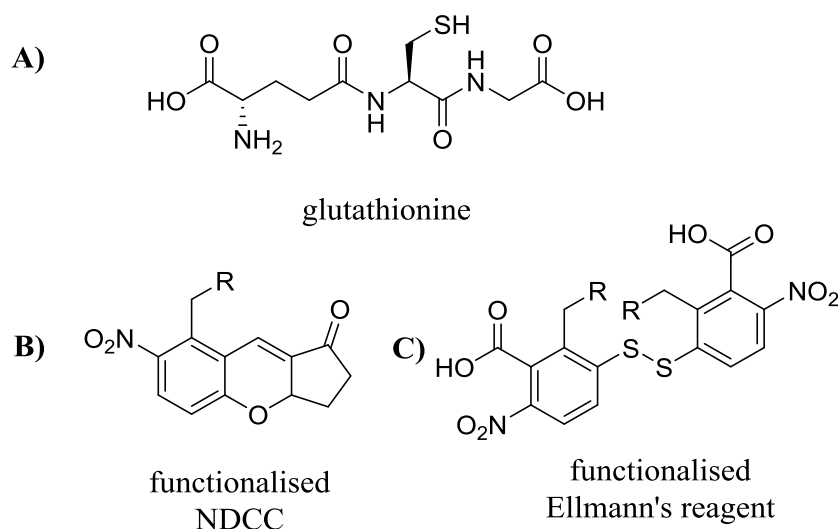
Figure 77: A) Glutathionine model for protein thiols. B) Functionalised NDCC (**12**) and C) Ellman's reagent. R represents a bulky blocking group.

Improvement of the HTS screen would lead to a quicker way of accurately determining methyltransferase activity. For example the activity of the mutants of CNMT and TNMT could be quickly determined by this method. The mutations on the CNMT enzyme have shown some key interactions between the substrate and the CNMT enzyme. However, activity or selectivity of the enzyme has not been improved by the mutations. In order to improve the selectivity of the CNMT enzyme through second generation mutants, mutations at the F257 site could first be contemplated. If the F257 does effect the 1 position binding, then restriction of the substrate scope could be achieved by mutation of this residue to a larger amino acid such as tryptophan. Alternatively the smaller residues such as alanine and glycine could be introduced at this position to accommodate even larger substrates. Finally selectivity for the CNMT enzyme could be improved through further structural studies of TNMT to determine what governs the enzyme's small substrate scope and suspected enantiospecificity. Understanding what controls TNMT's selectivity could then be applied to CNMT, which would be useful as CNMT has the far larger substrate scope. Applying enantiospecificity to a possible larger range of substrate would expand the utility of this selectivity yielding a range of enantiospecifically alkylated tetrahydroisoquinolines.

5   Conclusions

5.1   Regioselective alkylations

5.1.1   Applications of COMT regioselectivity

This study has uncovered a range of methyltransferases and determined novel bio-alkylations on a range of substrates. The study has shown with the COMT enzyme that regioselectivity can be engineered and improved on by fundamental studies of the active site of COMT. The understanding of the COMT active site has led to a range of regioselectivities with the regiocomplementary COMT mutants K144E and Y200L displayed *para* (28.7%) and *meta* selectivity (90.4%) respectively, fulfilling the first aim described in Section 1.3.2. In order to develop the regiocomplimentarity displayed in this study for industrial applications further transformations of the COMT enzyme both *in vivo* and *in vitro* systems would be necessary.

The regioselectivity of COMT could be a valuable tool for synthesis requiring regioselective alkylations. As Section 3.2.1 showed synthesis of regioselective alkylated derivatives of catechols produces low yields with dialkylated and monoalkylated side products. However, in order to introduce COMT as a regioselective alkylating agent, the enzyme would require optimisation for organic solvents or alternatively synthetic reactions would have to be under aqueous conditions. Engineering the COMT enzyme would be less labour intensive than changing pre-established synthetic routes. In order to optimise COMT for organic solvent surface mutations or immobilisation of the enzyme could be trialled. Whilst these transformations are not trivial, regioselective alkylations on catechols could act as a useful protection strategy in the synthesis of APIs. Therefore, development of Y200L COMT with these improvements could provide a useful regioselective alternative to the synthesis of mono and di alkylated catecholic APIs.

The further manipulation of COMT for production of regioselective alkylation protection strategies, whilst offering a viable alternative to existing strategies, may prove too difficult. However, the COMT enzyme can also be optimised for the *in vivo* production. As mentioned in Section 1.3.1 several groups have shown COMT can be included in metabolically engineered pathways in an *E.coli* host using a DHBA substrate for the production of vanillin (**3a**).[103,156] However, the unwanted isovanillin (**3b**) side product was also isolated from fermentations and separation was necessary to remove it.[103] Introduction of the regioselective mutant Y200L would drastically decrease the amount of isovanillin (**3b**), which would improve the yield and viability of the process. Therefore, insertion of the Y200L coding gene into the *E.coli* vanillin (**3a**) producing strain would lead to a better production of **3a**.

The two possible applications mentioned has shown that COMT regioselectivity could be developed and utilised for industrial applications. However, in terms of regioselective alkylation for protection strategies, the development of COMT would be frustrated by the need for the synthesis of unstable and diastereomerically impure analogues. In order to overcome the problem of synthetic SAM analogues further work would be needed to develop the coupled MAT-COMT system, which could produce the analogues *in situ*. Therefore, to engineer COMT for regioselective protection further work would also be needed on the coupled systems developed in this study.

5.1.2   Applications of coupled methyltransferase systems

The regioselectivity of the COMT enzyme had been drastically improved especially in the direction of *meta* alkylation. The COMT enzyme had also been used in coupled systems to produce alkoxylations and alkylations, fulfilling the secondary aim of developing an alkyltransferase by coupling COMT with the MAT enzyme. The success of these experiments greatly expanded the repertoire of the enzyme. Expanding the cofactor and substrate scope of the COMT enzyme had shown that therapeutic agents such as oxytocin (**10a**) and goserelin (**9a**) can be selectively alkylated. The industrially relevant ethyl vanillin (**3c**) was also selectively produced and shown to be isolatable in modest yields. Further work should now aim to build

on these successes with the coupled MAT-COMT and tyrosinase-COMT systems, moving the work onwards towards better production and a diverse range of products.

The regioselective *in vivo* production of vanillin (**3a**) has already been mentioned as an important application that could be developed for COMT regioselectivity. The *in vivo* production of ethyl vanillin (**3c**) would also be useful system to explore. Using the same approach, as mentioned in Section 5.1.1, the hMAT2A I322V and Y200L COMT could be co-expressed in same *E.coli* host.[103] However, unlike previous studies, an alternative *E.coli* strain would have to be sought to avoid SAM (**1a**) present in the cells converting the DHBA (**4**) substrate into vanillic (**4a**) and isovanillic (**4b**) products. The solution to this would be to use a methionine (**2a**) auxotrope strain, which has the **2a** biosynthetic pathway knocked out. The cells could be grown with **2a** supplements initially to enable the development of the protein. Then the cells could be expressed in minimal media, before addition of **4** and ethionine (**2b**) for the whole cell biotransformation. Alternatively, the methionine (**2a**) biosynthetic pathway could be inhibited by excluding **2a** from any media and fixed amounts of other amino acids (known to inhibit **2a** biosynthesis), as laid out by G. Van Duyne *et al*.[157] Using either of these conditions could produce the valuable **3c** flavouring product under fermentation conditions, without unwanted side products.
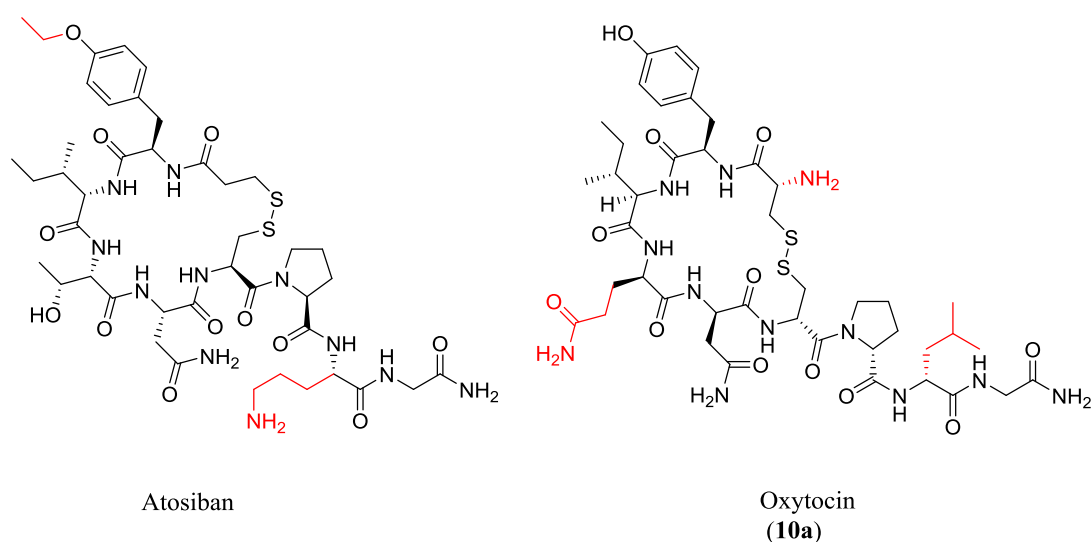
Figure 78: Atosiban and Oxytocin (**10a**). Differences in structure are highlighted in red.

Conducting ethylations on a larger scale could prove to be very useful alternative to existing productions of ethyl vanillin (**3c**). Using the MAT-COMT system on this

189

larger scale would also enable larger production of possibly other alkylated hydroxyls. Alkylated products of the goserelin (**9a**) and oxytocin (**10a**) substrates for the tyrosinase-COMT system could also offer easy access to a range of alkoxylated analogues for these therapeutic peptides and could improve their therapeutic properties. For example atosiban (Figure 78) has been used as a labour suppressant,[158] similar in structure to **10a** but ethylated at the phenolic position.[158] Alkylation of this phenolic position has shown an alternative drug can be developed. Based on this, hydroxy-goserelin (**9b**) and hydroxy-oxytocin (**10b**) could be screened with COMT and a range of analogues (using the colorimetric screen to determine activity) to develop alternative alkoxylated analogues of the **9a** and **10a** drugs (Figure 79). Alternatively, the screen could start from the **9b** and **10b** and use the tyrosinase enzyme, with associated analogues. Using this approach novel analogues of goserelin (**9a**) and oxytocin (**10a**) could be developed and could offer alternative or improved pharmaceutical activities.

The isolation of better quantities and even novel analogues of therapeutic peptides would be a useful route to proceed with the advancement of the coupled MAT-COMT and tyrosinase-COMT systems. The development of the regioselective COMT underpins these systems and utilising this regioselectivity could be a very important tool in expansion of the applications of the COMT enzyme. Through co-development of the COMT coupled systems as well as a methyltransferase HTS, the understanding learnt from the COMT system could at least be used to model further methyltransferase development.
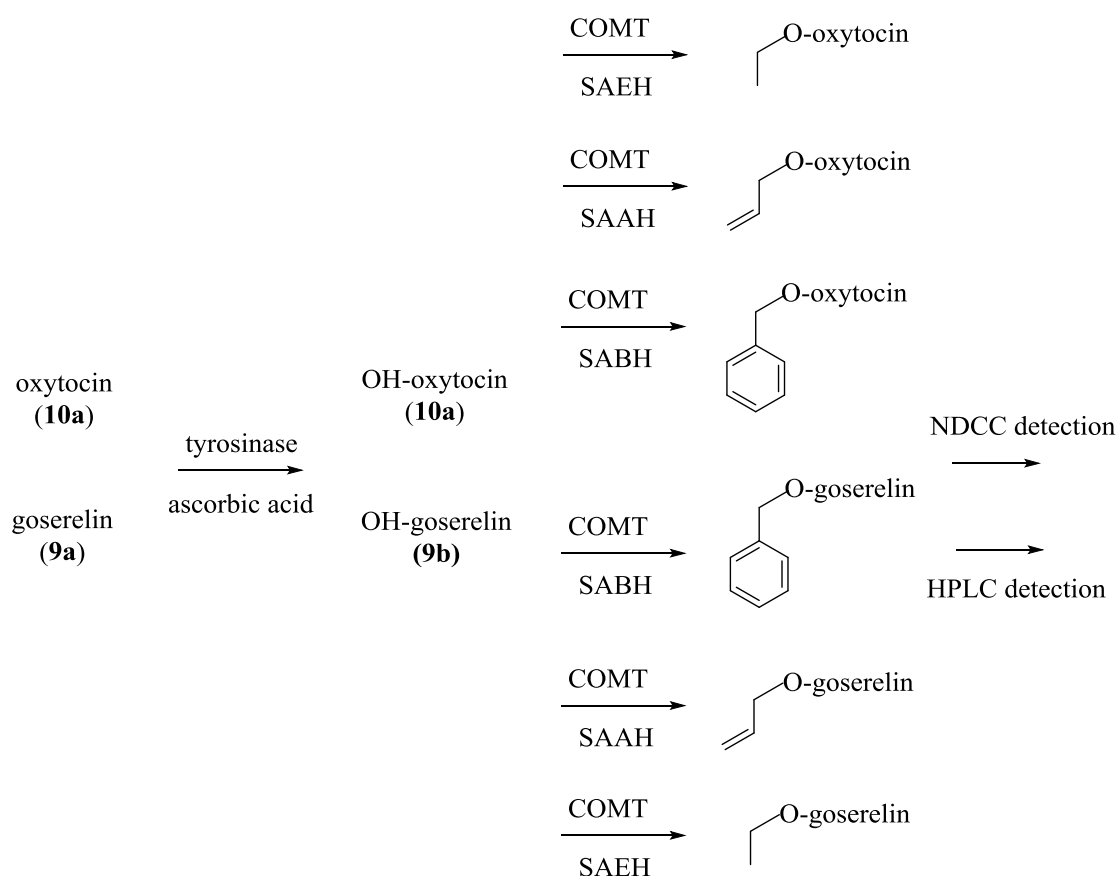
Figure 79: Suggested screen to produce oxytocin (**10a**) and goserelin (**9a**) analogues.

5.2 Expanding the screening of unknown methyltransferases

5.2.1 Methyltransferase screening

The development of a methyltransferase screen that would only take one step and could provide qualitative and quantitative results was only partially achieved in this study. Namely the NDCC (**12**) screen was qualitative but not quantitative. However, the development of the screen was simpler than existing screens and did enable further understanding of the substrate scope of CNMT and TNMT, fulfilling the final aim laid out in Section 1.3.2. The NDCC (**12**) screen could be used for further applications than determination of a substrate scope and this section will examine these applications in more detail.

The previous Section, (Section 5.1.2) had shown how the NDCC (**12**) screen could be used to screen for oxytocin (**10a**) and goserelin (**9a**) analogues. The screening to determine successful production of analogues of **10a** and **9a** therapeutic peptides would focus on the transfer of methyl and alkyl chains. Development of alkyltransferases using the screen could be a very useful application. By using the screen to monitor alkyl transfer the cofactor scope of methyltransferases could be observed, as well as the cofactor scope. In order for this to be achieved on large scale the NDCC (**12**) screen should be tested with the MAT-COMT system, to determine whether initially allylations and ethylations could be monitored. If the monitoring of alkylations were possible then a screening with multiple different methionine analogues and substrates could be envisioned. Expansion of the NDCC monitoring capabilities could thus be used to determine multiple alkylation reactions with multiple substrates.

The NDCC probe (**12**) could be used to determine alkyl chain transfer to a range of substrates. The screen could also be used to develop other properties of methyltransferases. One example would be enzyme stability. As was shown by Poor *et al*,[159] the RebH halogenase enzyme thermal stability was dramatically improved with an optimum temperature change of 5°C, and greater activity achieved at higher temperatures.[159] The study was conducted through the screening of random mutants of the RebH enzyme with a HPLC screen.[159] Conducting random mutagenesis on a selected methyltransferase, and screening with an improved, quicker and more facile NDCC (**12**) screen for activity (compared to HPLC), could look at improving the properties the selected methyltransferases (such as thermo stability) in a similar manner to RebH.

5.2.2   Determination and utilisation of TNMT enantiospecificity

Utilising the NDCC probe (**12**) could be a powerful tool in developing methyltransferases in an industrial setting for both alkylation and improvement of enzyme properties. The NDCC (**12**) screen could also be used to increase understanding of specific methyltransferases and the residues controlling selectivity could be determined by screening mutants. Understanding the selectivity of

individual methyltransferases would enable these discoveries to be transferred on to similar or related methyltranferases.

Section 4.1.2 showed that the sequence identity between TNMT and CNMT was high with a sequence similarity of 71% (Figure 58). The phylogenetic tree analysis also shows that the enzymes are closely related, which was unsurprising seeing that they are both on the berberine pathway. However, the substrate scope of TNMT, as has been shown, was much more limited than CNMT. Unfortunately, TNMT does not yet have a resolved crystal structure, which could be used to determine what the cause of the limited substrate scope was. Therefore, future studies will rely on homology models based on the CNMT structure and mutagenesis on suspected active site residues to determine, which residues are important. The understanding of the active site of TNMT was imperative for determining the enzyme's selectivity for the 1-phenylheliamine (**15**) substrate. Therefore, using the NDCC (**12**) screen a wide number of mutations can be conducted on suspected TNMT active site residues to determine, which residues dictate this selectivity.

Understanding the selectivity of TNMT would also extend to the enzyme's stereospecificity. Determination of the TNMT enzyme's stereospecifictivity could be achieved by using chiral HPLC. The first step would be to separate the enantiomers of racemic 1-phenylheliamine. Then monitoring the conversion of **15** to 2-*N*-methyl-1-phenylheliamine (**15a**) to determining whether only one enantiomer of **15** is utilised by the enzyme. As well as offering an understanding of the TNMT enzyme, determination of the enantiospecificity could be used for further transformations. For example an application of the TNMT could be dynamic resolution. TNMT would be expected to be active with the norlaudanosine (similar to the natural substrate shown in Figure 59).[150] Norlaudanosine is an API for the anaesthetic enantiopure drug cisatracinium besilate (Scheme 28), which has almost replaced the non-enantiopure, first generation atracinium drug in clinic.[160] Therefore, using TNMT with the API norlaudanosine as a substrate, the TNMT enzyme could be used to exclusively alkylate the *S*-enantiomer (expected active isomer from literature studies) allowing easier separation of the API into the two enantiomers.[150] The easier separation of the two isomers could then allow for further transformations necessary to synthesise the cisatracinium drug.

S-Norlaudanosine   *R*-Norlaudanosine   *S*-2-*N*-methyl-Norlaudanosine   *R*-2-*N*-methyl-Norlaudanosine

R-SAM

TNMT

Separated from unmethylated enantiomer

cisatracinium besilate
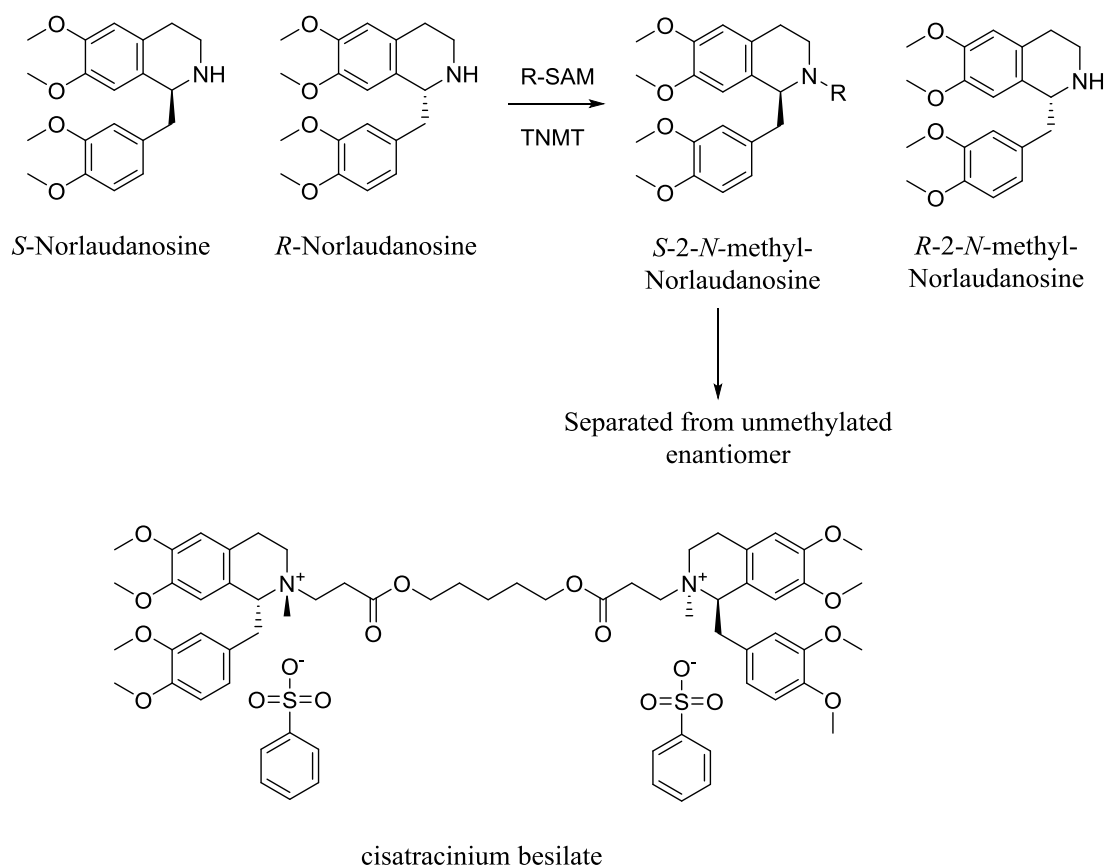
Scheme 28: (Left) suggested dynamic resolution of TNMT. (Right) Cisatracinium besilate an anaesthetic shown have a better therapeutic profile than the atracinium stereoisomeric mix.[150,160]

This study has shown that methyltransferases can be developed for alkylation and regioselectivity. The improvement of methyltransferases meant that these enzymes could then be used in conjunction with tyrosinase and hMAT2A enzymes. Successful alkylation and methoxylations has opened up the substrate scope of the COMT enzyme, as well as the cofactor scope. The introduction of a one-step methyltransferase HTS was also used to develop an understanding of the CNMT active site, as well as understanding the substrate scope of CNMT and TNMT enzymes. The continuing evolution of methyltransferases will rely on improved HTSs allowing further studies to continue from this work to develop substrate and cofactor selectivity for Class I methyltransferases.

6   Materials and methods

6.1   Chemical and reagents

Unless otherwise stated, all reagents were purchased from Sigma Aldrich UK or Fisher Scientific. Solvents for HPLC systems were purchased from Sigma Aldrich and of HPLC-grade quality. Primers were purchased from Sigma Aldrich, (synthesised at 0.05 μmol scale and purified by desalting). Synthetic genes for TNMT and CNMT were purchased by Dr Mark Thompson from the GeneArt® Gene Synthesis service (Life Technologies, Invitrogen™).

6.2   General techniques

6.2.1   Media

Media was sterilised by autoclaving at 121 °C for 10 min and then allowed to cool, before use.

LB broth: The following were added to make the media; tryptone (10 g), yeast extract (5 g), MgCl2 (0.95 g), NaCl (5 g), distilled H_2O (1L).

LB agar: The following were added to make the media; tryptone (10 g), yeast extract (5 g), $MgCl_2$ (0.95 g), NaCl (5 g), agar (15g) distilled H_2O (1L). For LB agar plates antibiotics were added for selection of transformant colonies: kanamycin (50 μg/mL), ampicillin (50 μg/mL) and chloramphenicol (34 μg/mL).

SOB medium: The following were added to make the media; Tryptone (20 g), yeast extract (5 g), NaCl (0.5 g), KCl (0.186 g), distilled H_2O (1L).

Inoue solution: $MnCl_2$ (10.9 g), $CaCl_2$ (2.2 g), KCl (18.7 g), PIPES solution (20 mL), dH_2O (1L). The resulting solution was then filter sterilised.

TAE (Tris-acetate-EDTA) buffer (10x): The following were added to make the buffer; Tris base (242 g), glacial acetic acid (57 mL), 0.5M EDTA pH 8 (100 mL), distilled H_2O (5L).

TE (Tris-EDTA) buffer: 10 mM Tris-HCl pH 8, 1 mM EDTA.

6.2.2   COMT Ni-NTA purification buffers

COMT lysis buffer: 50 mM KH_2PO_4/K_2HPO_4 pH 7.4, 300 mM NaCl

COMT wash 1 buffer: 50 mM KH_2PO_4/K_2HPO_4 pH 7.4, 300 mM NaCl, 30 mM imidazole.

COMT wash 2 buffer: 50 mM KH_2PO_4/K_2HPO_4 pH 7.4, 300 mM NaCl, 60 mM imidazole.

COMT elution buffer: 50 mM KH_2PO_4/K_2HPO_4 pH 7.4, 300 mM NaCl, 250 mM imidazole.

COMT storage buffer: 50 mM KH_2PO_4/K_2HPO_4 pH 7.4, 300 mM NaCl, 10% glycerol (v/v)

6.2.3   COMT crystallography buffers

COMT Tris buffer- 20 mM Tris HCl pH7.4, 20 mM NaCl

Anion exchange buffer 1: 20 mM Tris HCl pH8, 20 mM NaCl

Anion exchange buffer 2: 20 mM Tris HCl pH8, 50 mM NaCl

Anion exchange elution buffer: 20 mM Tris HCl pH8, 1 M NaCl

COMT gel filtration buffer: 20 mM potassium phosphate pH7.4, 300 mM NaCl, 10 mM DTT.

6.2.4   hMAT2A I322V buffers

hMAT2A lysis buffer: 50 mM Tris HCl pH 8.5 500 mM NaCl, 5% glycerol (v/v)

hMAT2A wash 1 buffer: 50 mM Tris HCl pH 8.5, 500 mM NaCl, 5% glycerol (v/v), 20 mM imidazole.

hMAT2A wash 2 buffer: 50 mM Tris HCl pH 8.5, 500 mM NaCl, 5% glycerol (v/v), 60 mM imidazole.

hMAT2A elution buffer: 50 mM Tris HCl pH 8.5, 500 mM NaCl, 5% glycerol (v/v), 250 mM imidazole.

hMAT2A storage buffer: 50 mM Tris HCl pH 8.5, 500 mM NaCl, 10% glycerol (v/v).

6.2.5   CNMT and TNMT Ni-NTA purification buffers

CNMT/TNMT lysis buffer: 20 mM KH_2PO_4/K_2HPO_4 pH 7.4, 200 mM NaCl, 20 mM imidazole.

CNMT/TNMT wash 1 buffer: 20 mM KH_2PO_4/K_2HPO_4 pH 7.4, 200 mM NaCl, 60 mM imidazole.

CNMT/TNMT elution buffer: 20 mM KH_2PO_4/K_2HPO_4 pH 7.4, 300 mM NaCl, 500 mM imidazole.

CNMT/TNMT storage buffer: 100 mM KH_2PO_4/K_2HPO_4 pH 7, 200 mM NaCl, 10% glycerol (v/v)

6.2.6   CNMT crystallography buffers

CNMT Tris buffer- 20 mM Tris HCl pH 8, 20 mM NaCl

Anion exchange buffer 1: 20 mM Tris HCl pH 8, 50 mM NaCl

Anion exchange elution buffer: 20 mM Tris HCl pH 8, 1 M NaCl

CNMT gel filtration buffer: 25 mM Tris HCl pH 8.4, 150 mM NaCl

6.2.7   Chemically competent cells

An overnight seeder culture was prepared with SOB media (5 mL) (Section 6.2.1) inoculated with a stock of competent cells (BL21 or DH$_{5\alpha}$ or Rossetta^{TM} 2) and grown at 37°C overnight with 200 rpm agitation. The seeder culture was diluted (50-fold) into fresh SOB media (400 mL) and grown at 18 °C with 200 rpm agitation until an OD$_{600}$ of 0.45 was reached. The culture was transferred into falcon tubes (8 x 50 mL) and cooled on an ice-water bath for 10 mins. The cells were collected as a cell pellet by centrifugation (2,500 *x g*, 10 min, 4 °C). The cells were washed and suspended in Inoue solution (10 mL per tube) (Section 6.2.1) before pooling together into four tubes. The cell suspensions were again centrifuged (2,500 *x g*, 10 min, 4 °C) and the cell pellets were collected and resuspended in 4 mL Inoue solution per tube (16 mL total). The tubes were then combined into a single tube. Pre-warmed DMSO (37°C) was added at a final concentration of 7% (v/v) and the cells incubated on ice for 10 mins. The competent cells were aliquoted into pre-chilled 1.5 mL Eppendorf tubes and were flash frozen in liquid nitrogen, before storing at -80°C.

6.2.8   Site-directed mutagenesis

COMT mutagenesis- the expression construct pET21b-COMT, which carried an *E. coli* C-terminal His$_6$-tag, was kindly provided by Professor Nigel Scrutton. Mutagenic primers were designed for positions; Trp38, Lys144, Val173, Glu199 and Tyr200 (Section 6.3.1, Table 15). The WT COMT sequence was amplified using the

mutagenic primers by PCR with the following mix (total volume 50 µL); 5x Phusion^{TM} HF buffer (10 µL), 20 mM dNTP mix (2 µL), 10 µM forward primer (2 µL), 10 µM reverse primer (2 µL), 100-50 ng/µL pET21b COMT plasmid (2 µL), DMSO (1 µL), Phusion^{TM} HF polymerase (0.5 µL) and distilled H_2O. The protocol for the G-storm thermocycler (used to amplify the DNA) was as follows; initial denaturation 98°C (2 mins), denaturation 98°C (20 s), annealing 50-70°C (20 s),extension 72°C (185 s), final extension 72 °C (10 minutes) and samples were then kept at 4°C (denaturation, annealing and extension steps were completed in 16 cycles before the final extension).The parental wild-type DNA was then digested with the addition of *Dpn*I restriction enzyme. After digestion, the mutant DNA was transformed into *E. coli* DH5α chemically competent cells for plasmid amplification and verification. The mutant DNA was extracted using a Qiagen miniprep spin kit and mutations were confirmed by nucleotide sequencing (GATC Biotech).

For COMT mutants with one long primer, the steps were the same except an alteration in the sample mix; 5x Phusion^{TM} HF buffer (2.4 µL), 20 mM dNTP mix (0.5 µL), 10 µM primer (0.5 µL), 50-100 ng/µL pET21b COMT plasmid (0.5 µL), DMSO (0.5 µL), Phusion^{TM} HF polymerase (0.5 µL) and distilled H_2O.

CNMT and TNMT mutagenesis-Cloning of the synthetic genes for TNMT and CNMT into pET28a vector was carried out by Dr Mark Thompson. Mutagenesis was conducted using the same method shown for *COMT mutagenesis* with the substitution of pET28a TNMT/ pET28a CNMT and mutagenic primers, with a longer extension step in thermocycling of 300 s due to the longer constructs for CNMT and TNMT (Section 6.3.2 and Table 16).

6.2.9   Plasmid purification

Plasmids were purified using a Qiagen MiniPrep Kit. Transformed $DH_{5\alpha}$ cells were grown overnight in LB media and the appropriate antibiotic (ampicillin, kanamycin or kanamycin and chloramphenicol) at 37°C with 200 rpm agitation. The resulting

solution was centrifuged (4,500 x g, 10 minutes, 4°C) and the supernatant discarded. The cell pellets were resuspended in Qiagen Buffer P1 containing RNAse (250 µL) and Qiagen Buffer P2 (250 µL) was subsequently added. The solution was inverted in the tubes 5-7 times, to ensure the cells were lysed before addition of Qiagen Buffer N_3 (750 µL) to precipitate the cells fats and proteins. The resulting mixture was centrifuged (10,000 x g, 10mins) and the supernatant transferred to a Qiagen miniprep column. The column was centrifuged (10,000 x g, 30 s) and the flow through discarded. The Qiagen Buffer PB (500µL) was then added centrifuged (10,000 x g, 30 s) and the flow through discarded. The same step was repeated with Qiagen Buffer PE (750 µL) with an additional centrifugation step on the empty tube to remove ethanol from the Qiagen PE buffer (10,000 x g, 2 mins). The spin column was transferred to a clean eppendorf and Qiagen EB (50 µL) was added stood for 90s. The column was then centrifuged and the DNA (50 µL) collected, measured by a NanoDrop 2000 UV/Vis spectrophotometer at 260 nm, and confirmed by sequence reading company GATC.

6.2.10 Transformations

The transformations were conducted under sterile conditions. Briefly, DNA (1 µL) was added to thawed competent cells (BL21, Rosetta^{TM}2 or DH$_{5\alpha}$) (100 µL) and kept on ice for 30 minutes. The cells were then heat shocked at 42°C for 45s and placed back on ice for 2 minutes. After 2 minutes LB media (400 µL) was added and the cells were grown at 37°C for 1 hour. After 1 hour the cells (150 µL) were transferred to an agar plate with the associated antibiotic (ampicillin, kanamycin or kanamycin and chloramphenicol) and colonies were incubated overnight at 37°C. The resulting colonies on the plate were then stored at 4°C. Colonies could be transferred to seeder cultures (LB media and appropriate antibiotic). The subsequent seeder cultures were then used to prepare glycerol stocks through addition of 50% glycerol (10% glycerol final concentration), which could be stored at -80°C.

6.2.11 Expression and Purification of COMT

The BL21 cells were first grown overnight, in seeder cultures (40 mL) with LB media and ampicillin (50 µg/mL) at 37°C. The seeder cultures were then diluted 100 fold into fresh LB media and ampicillin (50 µg/mL). The diluted cultures were then grown at 37°C with agitation at a 190 rpm, until an OD of 0.6-1 was reached. The protein was then expressed with IPTG (1 mM) for 3 hours at 37°C. The cells were collected with centrifugation (4,000 x g, 10 min, 4°C) and disrupted in lysis buffer by sonication. The solution was then centrifuged (17,000 x g, 30 min, 4°C) and the supernatant was isolated. WT COMT (and mutants) were then purified by passing the supernatant through a Nickel NTA column with two wash steps before elution; 7 column volumes of Wash 1 Buffer (Section 6.2.2), 5 column volumes of wash 2 buffer (Section 6.2.2) and 7 column volumes of elution buffer (Section 6.2.2). Fractions were analysed by SDS-PAGE gels, then COMT was then dialysed (into storage buffer Section 6.2.2) and concentrated in a 10,000 MWCO Vivaspin 20 centricon (Sartorius Stedim Biotech). Concentration of the enzyme was determined by Nanodrop (as were all subsequent proteins), COMT was stored at -80°C.

6.2.12 Purification of COMT for crystallography

For purification of dimeric WT COMT and Y200L for crystallography; after purification by the Nickel NTA column the eluted COMT was dialysed into the COMT Tris buffer (Section 6.2.3) before purification with an ÄKTA prime FPLC (amersham pharmacia biotech) and HiTrap Q-Hp anion exchange column. The sample was loaded onto the HiTrap Q-Hp anion exchange column, which was then washed with; anion exchange buffer 1 and 2 (Section 6.2.3), and eluted with anion exchange elution buffer (Section 6.2.3). After elution the COMT was dialysed into gel filtration buffer (Section 6.2.3), purified by gel filtration with an ÄKTApurifier FPLC (GE Healthcare) and a Superdex 200, 10/300 column and eluting as two peaks. The first peak was shown to be dimeric WT COMT (this peak was also used for Y200L) and this peak was collected. COMT samples (1 µM) were then soaked

overnight with SAM (**1a**) (10 µM) and DNC (10 µM) samples were then concentrated to 10 mg/mL and submitted for crystallographic trials with Dr Colin Levy.

6.2.13 Expression and Purification of hMAT2A I322V

The Rosetta^{TM} 2 cells (containing pNIC28-Bsa4-hMAT2A plasmid kindly donated by Professor Udo Opperman) were first grown overnight, in seeder cultures (40 mL) with LB media, kanamycin (50 µg/mL) and chloramphenicol (50 µg/mL) at 37°C. The seeder cultures were then diluted 100 fold into fresh LB media and kanamycin (50 µg/mL) and chloramphenicol (50 µg/mL). The diluted cultures were then grown at 37°C with agitation at a 190 rpm, until an OD of 0.6-1 was reached. The protein was then expressed with IPTG (1 mM) for 4 hours at 30°C. The cells were collected with centrifugation (4,000 x g, 10 min, 4°C) and disrupted in lysis buffer (Section 6.2.4) by sonication. The solution was then centrifuged (17,000 x g, 30 min, 4°C) and the supernatant isolated. hMAT2A I322V was then purified by Nickel NTA column with two wash steps before elution; 7 column volumes of wash 1 buffer (Section 6.2.4), 5 column volumes of wash 2 buffer (Section 6.2.4) and 7 column volumes of elution buffer (Section 6.2.4). Fractions were analysed by SDS-PAGE gels and hMAT2A I322V was then dialysed (into storage buffer Section 6.2.4) and concentrated in a 10,000 MWCO Vivaspin 20 centricon (Sartorius Stedim Biotech). Concentration of the enzyme was determined by Nanodrop, hMAT2A I322V was stored at -80°C.

6.2.14 Expression and Purification of CNMT and TNMT

CNMT/TNMT expression and purification-The BL21 cells (containing pET28a-CNMT/TNMT plasmid) were first grown overnight, in seeder cultures (40 mL) with LB media and kanamycin (50 µg/mL) at 37°C. The seeder cultures were then diluted 100 fold into fresh LB media and kanamycin (50 µg/mL). The diluted cultures were then grown at 37°C with agitation at a 190 rpm, until an OD of 0.6-1 was reached.

The protein was then expressed with IPTG (0.5 mM) overnight at 20°C. The cells were collected with centrifugation (4,000 x *g*, 10 min, 4°C) and disrupted in lysis buffer (Section 6.2.5) by sonication. The solution was then centrifuged (17,000 x *g*, 30 min,4°C) and the supernatant isolated. CNMT/TNMT (and mutants) were then purified by Nickel NTA column with two wash steps before elution; 7 column volumes of lysis buffer, 5 column volumes of wash 1 buffer (Section 6.2.5) and 7 column volumes of elution buffer (Section 6.2.5). Fractions were analysed by SDS-PAGE gels, then CNMT/TNMT was then dialysed into storage buffer (Section 6.2.5) and concentrated in a 10,000 MWCO Vivaspin 20 centricon (Sartorius Stedim Biotech). Concentration of the enzyme was determined by Nanodrop, COMT was stored at -80°C.

CNMT purification for crystallography- After purification by the Nickel NTA column (shown above) the eluted CNMT was dialysed into a Tris buffer (Section 6.2.6) before purification with an ÄKTA prime FPLC (amersham pharmacia biotech) and a HiTrap Q-Hp anion exchange column. The sample was loaded onto the HiTrap Q-Hp anion exchange column, which was then washed with anion exchange buffer 1 (Section 6.2.6) before elution of the CNMT in anion exchange elution buffer (Section 6.2.6). After elution the CNMT was purified by gel filtration with an ÄKTApurifier FPLC (GE Healthcare) and a Superdex 200, 10/300 column and a gel filtration buffer (Section 6.2.6), elution was noted as one peak. CNMT samples were then concentrated to 10 mg/mL and were submitted with SAH (**1**) and substrates for crystallographic trials with Dr Mark Dunstan.

6.3   Primer Tables

6.3.1   COMT primer table

Table 15: Primer table for COMT mutants. Where bp stands for base pairs. Mutation site (red).

Primer	Sequence 5′-3′	bp
W38F F	TACTGCACCCAGAAAGAATTTGCCATGAAC	30

W38F R	ATCACCAACGTTCATGGCAAATTCTTTCTG	30
W38D F	TACTGCACCCAGAAAGAAGATGCCATGAAC	30
W38D R	ATCACCAACGTTCATGGCATCTTCTTTCTG	30
W38Y F	TACTGCACCCAGAAAGAATATGCCATGAAC	30
W38Y R	ATCACCAACGTTCATGGCATATTCTTTCTG	30
K144A	GTCTTCCTGGACCACTGGGCAGATCGTTAC	30
K144E	GGATATGGTCTTCCTGGACCACTGGGAAGA TCGTTACCTGCCAGATACCCTGC	51
K144R	GGATATGGTCTTCCTGGACCACTGGCGTGAT CGTTACCTGCCAGATACCCTGC	51
E199L	GTGTACCCACTACTCCAGCTACCTGCTGTA CATGAAAGTGGTCGACGGTCTGG	51
E199D F	CACTACTCCAGCTACCTGGATTACATGAAAG	31
E199D R	GTCGACCACTTTCATGTAATCCAGGTAGC	29
Y200A F	ACTACTCCAGCTACCTGGAAGCGATGAAAGTG	32
Y200A R	ACCGTCGACCACTTTCATCGCTTCCAG	27
Y200L F	ACTACTCCAGCTACCTGGAACTGATGAAAGTG	32
Y200L R	ACCGTCGACCACTTTCATCAGTTCCAG	27
Y200S F	ACTACTCCAGCTACCTGGAATCTATGAAAGTG	32
Y200S R	ACCGTCGACCACTTTCATAGATTCCAG	27
Y200W F	ACTACTCCAGCTACCTGGAATGGATGAAAGTG	32
Y200W R	ACCGTCGACCACTTTCATCCATTCCAG	27

6.3.2 CNMT and TNMT primer table

Table 16: Primer table for CNMT and TNMT mutants. Where bp stands for base pairs. Mutation site (red).

Primer	Sequence 5′-3′	bp
CNMT W329I G331T F	CTGATCAATTAT**ATT**CGT**ACT**TTTTGTCTGAGTGGCATGG	40
CNMT W329I G331T R	CAGACAAAA**AGT**ACG**AAT**ATAATTGATCAGTTTAACCGC	39
CNMT H208A F	CGAACTGTTTGAA**GCT**ATGAAAAACTACGAACTGCTGCTG	40
CNMT H208A R	AGTTTTTCAT**AGC**TTCAAACAGTTCGATCACCAGAATACG	40
CNMT H208F F	CGAACTGTTTGAA**TTT**ATGAAAAACTACGAACTGCTGCTG	40
CNMT H208F R	GTTTTTCAT**AAA**TTCAAACAGTTCGATCACCAGAATACG	39
CNMT E207Q F	CGAACTGTTT**CAA**CATATGAAAAACTACGAACTGCTGCTG	40
CNMT E207Q R	GTTTTTCATATG**TTG**AAACAGTTCGATCACCAGAATACG	39
CNMT E207A F	CGAACTGTTT**GCA**CATATGAAAAACTACGAACTGCTGCTG	40
CNMT E207A R	GTTTTTCATATG**TGC**AAACAGTTCGATCACCAGAATACG	39
CNMT Y328A F	CTGATCAATGCTTG**GCG**TGGTTTTTGTCTGAGTGGCATGG	40
CNMT Y328A	CACTCAGACAAAAACCA**CGC**CAAGCATTGATCAGTTTAACC	41

R		
CNMT W329A F	CTGATCAATTATGCGCGTGGTTTTTGTCTGAGTGGCATGG	40
CNMT W329A R	CACTCAGACAAAAACCACGCGCATAATTGATCAGTTTAACC	41
CNMT R330A F	CTGATCAATTATTGGGCTGGTTTTTGTCTGAGTGGCATGG	40
CNMT R330A R	CACTCAGACAAAAACCAGCCCAATAATTGATCAGTTTAACC	41
CNMT G331A F	CTGATCAATTATTGGCGTGCTTTTTGTCTGAGTGGCATGG	40
CNMT G331A R	CACTCAGACAAAAAGCACGCCAATAATTGATCAGTTTAACC	41
CNMT F332A F	CTGATCAATTATTGGCGTGGTGCTTGTCTGAGTGGCATGG	40
CNMT F332A R	CACTCAGACAAGCACCACGCCAATAATTGATCAGTTTAACC	41
TNMT I329W T331G F	GTTGTTACCCATTGGCGTGGCTTTTGCATGGGTGGTTATG	40
TNMT I329W T331G R	CATGCAAAAGCCACGCCAATGGGTAACAACACCATTAACG	40

6.4 Assays

All assays were conducted in an Eppendorf Thermomixer attached with either a 1.5 mL eppendorf rack or a MTP thermoblock.

6.4.1 Assay conditions for COMT activity and kinetic assays

K144A and WT COMT conversion assay conditions- K144A and WT COMT enzyme assays were prepared with the substrate DHBAL (**3**) as follows: DTT (1 mM), MgCl$_2$ (3 mM), SAM (**1a**) (2 mM), COMT (5 µM), substrate (1 mM), in 20 mM potassium phosphate buffer pH 7.4 (total volume 150 µL). The reactions were incubated at 37 °C, 800 rpm agitation for 1 hour and subsequently quenched with an equal volume of methanol. Protein was then removed through centrifugation (13,000 x *g*, room temperature, 10 minutes) and the assays were analysed by C$_{18}$ RP-HPLC using an Agilent 1260 Infinity HPLC system and the method shown in Sections 6.5.1. Determination of areas of products (**3a**, **3b**) and substrate (**3**) was conducted with the Agilent 1260 Infinity system software.

Monomer and dimeric WT COMT conversion assays with NO2CAT (5), DHBAL (3) and DHBA (4) substrates- Monomeric and dimeric WT COMT (and oligomeric COMT mutants) enzyme assays were prepared with the substrates DHBAL (**3**), DHBA (**4**) and NO2CAT (**5**) as follows: DTT (1 mM), MgCl2 (3 mM), SAM (**1a**) (1 mM), COMT (5 µM), substrate (0.5 mM), in 20 mM potassium phosphate buffer pH 7.4 (total volume 150 µL). The reactions were incubated at 37 °C, 800 rpm agitation for 30 minutes and subsequently quenched with an equal volume of methanol. Protein was then removed through centrifugation (13,000 x *g*, room temperature, 10 minutes) and the assays were analysed by C$_{18}$ RP-HPLC using an Agilent 1260 Infinity HPLC system and the two methods shown in Sections 6.5.1 and 6.5.2. Determination of areas of products (**3a**, **3b, 4a, 4b, 5a, 5b**) and substrates (**3-5**) was conducted with the Agilent 1260 Infinity system software.

Monomeric kinetic assay- The initial reaction rates of monomeric WT COMT with the DHBAL (**3**) substrate were determined for concentrations of DHBAL (**3**); 2 µM, 3 µM, 5 µM, 10 µM, 20 µM, 40 µM, 80 µM, 100 µM. The rates were determined under the following conditions; DTT (1 mM), MgCl$_2$ (3 mM), SAM (**1a**) (250 µM), monomeric WT COMT (0.2 µM) and DHBAL (**3**) (2 µM-100 µM) in 20 mM potassium phosphate buffer, pH 7.4 (total volume 150 µL). Assays were terminated with 1% FA (0.1% FA final concentration) at 1, 2 and 3 minutes. Protein was then removed through centrifugation (13,000 x *g*, room temperature, and 10 minutes).

Assays were then analysed by C_{18} RP-HPLC using an Agilent 1260 Infinity HPLC system (using the method described in Section 6.5.1). Using Michaelis-Menten kinetics; conversions from the HPLC were used to determine Determination of areas of products (**3a**, **3b**) and substrate (**3**) was conducted with the Agilent 1260 Infinity system software. The initial rates for each concentrated were calculated from the areas and submitted to the SigmaPlot 12.0 program, which was used to determine V_{max} and k_{cat}.

Dimeric WT COMT kinetic assays- The conditions were as for the kinetic assay for monomeric WT COMT with the following exceptions; dimeric WT COMT (0.1 µM) and DHBAL (**3**) substrate; 2 µM, 3 µM, 5 µM, 10 µM, 20 µM, 40 µM, 100 µM, 300 µM. Assays were terminated as before, analysed; determination of areas of products (**3a**, **3b**) and substrate (**3**) was conducted with the Agilent 1260 Infinity system software. The initial rates for each concentrated were calculated from the areas and submitted to the SigmaPlot 12.0 program (determining V_{max} and k_{cat} as before).

K144A kinetic assay- The conditions were as for the kinetic assay for monomeric and dimeric WT COMT with the following exceptions; K144A (0.5 µM) and DHBAL (**3**) substrate 10 µM, 20 µM, 40 µM, 100 µM, 300 µM, 500 µM, 1200 µM, 2000 µM. Assays were terminated with 1% FA (0.1% FA final concentration) but at 2, 3 and 4 minutes. Assays were then analysed; determination of areas of products (**3a**, **3b**) and substrate (**3**) was conducted with the Agilent 1260 Infinity system software. The initial rates for each concentrated were calculated from the areas and submitted to the SigmaPlot 12.0 program (determining V_{max} and k_{cat} as before).

Activity assay COMT and SAM analogue cofactors- COMT assays were prepared as follows with both synthetic and pre-prepared biosynthetic analogues; the hMAT2A assays were initially run with the conditions shown in Section 6.4.2 (total volume 300µL).. The hMAT2A enzyme was removed through Amicon® Ultra-0.5 centrifugal filter devices (3,000 MWCO), and the *in situ* SAM analogue (**1a-d**) concentration was measured through Shimadzu Prominence UFLC XR HPLC system using a Phenomenex Luna HILIC column. The assay mix was then used with the COMT enzyme as follows; *in situ* SAM analogue (**1a-d**) (80 µM), COMT (2 µM) and DHBAL (**3**) (80 µM) in 20mM potassium phosphate buffer pH 7.4 (total volume 100 µL). Alternatively with synthetic samples SAM analogue (**1a-d**) (160

μM), COMT (2 μM) and DHBAL (**3**) (80 μM) (due to presence of inactive SAM (**1a**) diastereomer). After 25 minutes the assays were then terminated with equal volume of methanol, protein removed through centrifugation (13,000 x *g*, room temperature, 10 minutes) and analysed on an Agilent 1260 Infinity HPLC with the HPLC conditions shown in Section 6.5.1. Determination of areas of products (**3a-h**) and substrate (**3**) was conducted with the Agilent 1260 Infinity system software.

6.4.2 Assay conditions for MAT conversion assay conditions

The hMAT2A assays were prepared with SAlH (**2c**), methionine (**2a**) and ethionine (**2b**) as follows: DTT (1 mM), MgCl$_2$ (3 mM), ATP (1.5 mM), methionine analogue (**2a-2c**) (1.5 mM), hMAT2A I322V (15 μM), in 20 mM phosphate buffer pH 7.4 (total volume 150 μL). The reactions were incubated at 37 °C, 800 rpm agitation for 1 hour and subsequently quenched with an equal volume of methanol. Protein was removed through centrifugation (13,000 x *g*, room temperature, 10 minutes). The assays were analysed by Shimadzu Prominence UFLC XR HPLC system using a Phenomenex Luna HILIC column (5 μm, 4.6x150 mm column) and the SAM analogues (**1a-d**) verified by ES$^+$ LC-MS.

6.4.3 Assay conditions for scale up of 3-ethoxy-4-hydroxybenzaldehyde (3c) production

The scale up of hMAT2A-COMT system was prepared as follows; the total volume of assays was 36 mL with 4 assays carried out in 9 mL vessels. DTT (1 mM), MgCl$_2$ (3 mM) ATP (3 mM), ethionine (**2b**) (3 mM), hMAT2A I322V (25 μM), Y200L COMT (176 μM), SAHH (20 μM), DHBAL (**3**) (0.5 mM), in 20 mM potassium phosphate buffer, pH 7.4. Assays were run overnight after terminated with an equal volume of methanol. The 4 assays were centrifuged (13,000 x *g*, room temperature, and 10 minutes) and purification proceeded as laid out in Sections 6.6.2 and 6.9.6.

6.4.4 Assay conditions for scale up of 3-allyloxy-4-hydroxybenzaldehyde (3e) production

Allyl vanillin (3-allyloxy-4-hydroxybenzaldehyde) 3e was prepared using a coupled enzyme reaction with hMAT2A (I322V) and COMT (Y200L) as follows: the reaction mixture containing 1 mM DTT, 2 mM S-allyl-homocysteine, 3 mM $MgCl_2$, 0.5 mM DHBAL **1a**, 45 μM COMT (Y200L), 5 μM hMAT2A (I322V) in 20 mM phosphate buffer pH 7.4, was incubated at 37 °C with 800 rpm agitation for 16 hours. The protein was precipitated and the product solubilised through addition of ethyl acetate (5 mL) and purification proceeded as laid out in Sections 6.6.3 and 6.9.7.

6.4.5 Assay conditions for scale up of 3-benzyloxy-4-hydroxybenzaldehyde (3g) production

Benzyl vanillin (3-benzyloxy-4-hydroxybenzaldehyde) **3g** was prepared using an enzyme reaction with COMT (Y200L) and the synthetic S-adenosyl-benzyl-homocysteine cofactor as follows: the reaction mixture containing 1 mM DTT, 2 mM S-adenosyl-benzyl-homocysteine, 3 mM $MgCl_2$, 0.5 mM DHBAL **1a**, 170 μM COMT (Y200L) in 20 mM phosphate buffer pH 7.4, was incubated at 37 °C with 800 rpm agitation for 16 hours. The protein was precipitated and the product solubilised through addition of ethyl acetate (5 mL) and purification proceeded as laid out in Sections 6.6.4 and 6.9.8.

6.4.6 Assay conditions for COMT mutants with SABH (1d) and SAAH (1c) cofactors

Assays were prepared for wild-type and mutant COMT enzymes (Y200L and K144E) with the substrate DHBAL (**3**) as follows: DTT (1 mM), $MgCl_2$ (3 mM), SAM analogue (**1a-d**) (0.25 mM), COMT (45 μM), DHBAL (**3**) (0.1 mM), in 20

mM phosphate buffer, pH 7.4. The reactions were incubated at 37 °C, 800 rpm agitation for 3 hours and subsequently quenched with an equal volume of methanol. The protein precipitate was removed by centrifugation (13,000 x g, room temperature, 10 minutes) and the assays were analysed by C_{18} RP-HPLC using an Agilent 1260 Infinity HPLC system using the method shown in Section 6.5.3. Determination of areas of products (**3a-h**) and substrate (**3**) was conducted with the Agilent 1260 Infinity system software.

6.4.7   Assay conditions for COMT and POB SAM cofactor

Assays were prepared as follows; $MgCl_2$ (3mM), POB-SAM (**1e**) (2mM), COMT (50uM), hydroxylated peptide substrate (**6b**, **8b** and **9b**) (0.5 mM), in a 20 mM potassium phosphate buffer pH 7.4 (total volume 150 µL). The assay mixtures were incubated at 37°C with 800 rpm agitation, overnight. Termination and analysis proceeded as before (for *Monomeric WT COMT kinetic assay*), assays were analysed by C_{18} RP-HPLC using an Agilent 1260 Infinity HPLC system and a MALDI-TOF-MS on a Bruker Ultraflex in reflector mode using DHB as matrix.

6.4.8   Assay conditions for the COMT DOPA peptide assays

Assays were prepared as follows; $MgCl_2$ (3 mM), SAM (**1a**) (2 mM), COMT (5 µM), hydroxylated peptide substrate (**6b-8b**) (0.5 mM), in a 20 mM potassium phosphate buffer pH 7.4 (total volume 150 µL). The assay mixtures were incubated at 37°C with 800 rpm agitation, for 1 hour. Termination and analysis proceeded as before (for *Monomeric WT COMT kinetic assay*), assays were analysed by C_{18} RP-HPLC using an Agilent 1260 Infinity HPLC system and a MALDI-TOF-MS on a Bruker Ultraflex in reflector mode using DHB as matrix. Determination of areas of products (**6c**, **7c**, **8c**) and substrates (**6b**, **7c**, **8c**) was conducted with the Agilent 1260 Infinity system software.

6.4.9 Assay conditions for the tyrosinase conversion assay

Assays were prepared as follows; MgCl$_2$ (3 mM), ascorbic acid (12.5 mM), mushroom tyrosinase (0.2 µM), peptide substrate (**6a**, **7a**, **9a**, **10a**) (0.5 mM), in a 20 mM potassium phosphate buffer pH 7.4 (total volume 150 µL). The assay mixtures were incubated at 25°C with 800 rpm agitation, for 15 minutes. Termination was achieved with 1% hydrochloric acid (0.1% final concentration) and the samples were then centrifuged to remove proteins (13,000 x g, room temperature, 10 minutes).All assays were analysed by Agilent HPLC (methods see Sections 6.5.5 and 6.5.4). Determination of areas of products (**6b**, **7b**, **9b**, **10b**) and substrates (**6a**, **7a**, **9a**, **10a**) was conducted with the Agilent 1260 Infinity system software. MALDI-TOF-MS on a Bruker Ultraflex in reflector mode using DHB as matrix was also used to confirm successful assays. Selected products were further characterised by MALDI-TOF-MS/MS on a Bruker Ultraflex in LIFT mode using DHB as matrix.

6.4.10 Assay conditions for the COMT-tyrosinase conversion assay

Excepting the tyrocidin substrate, assays were prepared as follows; MgCl$_2$ (3 mM), ascorbic acid (12.5 mM), SAM (**1a**) (2 mM), COMT (15 µM), mushroom tyrosinase (0.2 µM), peptide substrate (**6a**, **7a**, **9a**, **10a**) (0.5 mM), in a 20mM potassium phosphate buffer pH 7.4 (total volume 150 µL). The assay mixtures were incubated at 30°C with 800rpm agitation, for 5 hours. Termination was achieved with 1% Formic acid (0.1% final concentration) and the samples were then centrifuged to remove proteins (13,000 x g, room temperature, 10 minutes).All assays were analysed by Agilent HPLC (methods see Section 6.5.5 and 6.5.4). Determination of areas of products (**6c**, **7c**, **9c**, **10c**) and substrates (**6a**, **7a**, **9a**, **10a**) was conducted with the Agilent 1260 Infinity system software. MALDI-TOF-MS on a Bruker Ultraflex in reflector mode using DHB as matrix. Selected products were further characterised by MALDI-TOF-MS/MS on a Bruker Ultraflex in LIFT mode using DHB as matrix.

Tyrocidin COMT-Tyrosinase assay- Conditions were as previously stated with a few amendments; tyrocidin (Cambridge Research Biochemicals) (50 µM) and COMT (45 µM) were run with the tyrosinase system overnight. Tyrocidin products were further characterised by MALDI-TOF-MS/MS on a Bruker Ultraflex in LIFT mode using DHB as matrix identifying both hydroxylated and methoxylated products.

6.4.11 Assay conditions for TNMT and CNMT assays

Assays were prepared as follows; SAM (**1a**) (3 mM), CNMT/TNMT (mutants and WT) (5 µM), substrate (**13-16**) (higenamine (**16**) sourced from Sequoia Research Products) (1 mM), in a 20 mM potassium phosphate buffer pH 7.0 (total volume 150 µL). The reactions were incubated at 30 °C, 800 rpm agitation for 30 minutes and subsequently quenched with an equal volume of methanol. The protein precipitate was removed by centrifugation and the assays were analysed by C_{18} RP-HPLC using an Agilent 1260 Infinity HPLC system using the method shown in Sections 6.5.7 and 6.5.8. Determination of areas of products (**13a**-**16a**) and substrates (**13-16**) was conducted with the Agilent 1260 Infinity system software.

6.4.12 Assay conditions for colorimetric assay

Assays were prepared as follows; the selected methyltransferase (1 µM) was assayed with SAM (**1a**) (25 µM) and the selected substrate (25 µM) in 20 mM potassium phosphate buffer pH 7.4, for 25 minutes (total volume 300 µL). After 25 minutes, the assay was terminated by heating at 95°C for 5 minutes. The assay was cooled before addition of SAHH (1 µM) and NDCC (25 µM) (total volume 150 µL). Assays were then measured on the Synergy HT Microplate Reader over 25 minutes with detection at 405 nm.

6.5 Analytical HPLC methods

6.5.1 HPLC method for COMT assays with DHBAL (3)/DHBA (4) substrates

Using an Agilent 1260 Infinity HPLC and Phenomenex Kinetex C_{18} 5 μm 4.6 x 150 mm column, flow rate 1 mL/min, wavelength 278 nm, gradient: 0-2 min 5-20% B, 2-7 min 20-23% B, 7-7.1 min 23-95% B, 7.1-9 min 95% B, 9-9.1 min 95-5% B, 9.1-11 min 5% B, mobile phase A: H_2O 0.05% TFA, mobile phase B: acetonitrile 0.05% TFA.

6.5.2 HPLC method for COMT assays with NO2CAT (5) substrate

Using an Agilent 1260 Infinity HPLC and Phenomenex Kinetex C_{18} 5 μm 4.6 x 150 mm column, flow rate 1 mL/min, wavelength 340 nm, gradient: 0-2 min 5% B, 2-2.1 min 5-25% B, 2.1-8 min 25-75% B, 8-8.1 min 75-95% B, 8.1-10 min 95% B, 10-10.1 min 95-5% B, 10.1-12 min 5% B, mobile phase A: H_2O 0.05% TFA, mobile phase B: acetonitrile 0.05% TFA.

6.5.3 HPLC method for COMT assays with SAM analogue (1b-d) cofactors

Using an Agilent 1260 Infinity HPLC and Phenomenex Kinetex C_{18} 5 μm 4.6 x 150 mm column, flow rate 1 mL/min, wavelength 278 nm, gradient: 0-2 min 5% B, 2-3 min 5-20% B, 3-4 min 20-50% B, 4-9 min 50-70% B, B, 9-11 min 70% B, 11-11.1 70-5% B, 11.1-13 min 5% B, mobile phase A: 5 mM potassium phosphate buffer, mobile phase B: acetonitrile.

6.5.4 HPLC method for COMT/tyrosinase-COMT assays with Goserelin (9a) and Oxytocin (10a) substrates

Using an Agilent 1260 Infinity HPLC and Phenomenex Kinetex C_{18} 5 μm 4.6 x 150 mm column, flow rate 1 mL/min, wavelength 280 nm, gradient: 0-2 min 5% B, 2-9 min 5-65% B, 9-10.1 min 65-95% B, 10.1-11 min 95% B, 11-11.1 min 95-5% B, 11.1-14 min 5% B, mobile phase A: H_2O 0.1% FA, mobile phase B: acetonitrile 0.1% FA.

6.5.5 HPLC method for COMT/tyrosinase-COMT assays with synthetic peptides KNYLDF (7a) and KNFLDY (6a)

Using an Agilent 1260 Infinity HPLC and Phenomenex Kinetex C_{18} 5 μm 4.6 x 150 mm column, flow rate 1 mL/min, wavelength 280 nm, gradient: 0-2 min 5% B, 2-7 min 5-75% B, 7-7.1 min 75-95% B, 7.1-9 min 95% B, 9-9.1 min 95-5% B, 9.1-12 min 5% B, mobile phase A: H_2O 0.1% FA, mobile phase B: acetonitrile 0.1% FA.

6.5.6 HPLC method for measurement of SAM analogue (1a-d) concentrations *in situ*

Using Shimadzu Prominence UFLC XR HPLC system and a Phenomenex Luna HILIC column (5μm, 4.6x150 mm column) flow rate 1 mL/min, wavelength 260 nm, gradient: 0-1 min 10% B, 1-3 min 10-35% B, 3-8 min 35-60% B, 8-10 min 60% B, 10-10.2 min 60-10% B, 10.2-11.6 min 10% B, mobile phase A: acetonitrile, mobile phase B: 5mM ammonium formate, pH3.3.

6.5.7 HPLC method for CNMT assays with heliamine (13) and 1-methylheliamine (14) substrates

Using an Agilent 1260 Infinity HPLC and Phenomenex Kinetex C_{18} 5 μm 4.6 x 150 mm column, flow rate 1 mL/min, wavelength 283 nm, gradient: 0-1 min 5% B, 1-2 min 5-10% B, 2-8 min 10-33% B, 8-8.1 min 33-95% B, 8.1-10.0 min 95% B, 10.0-10.1 min 95-5% B, 10.1-11.5 min 5%, mobile phase A: H_2O 0.05% TFA, mobile phase B: acetonitrile 0.05% TFA.

6.5.8 HPLC method for CNMT and TNMT assays with higenamine (16), 1-phenylheliamine (15) substrates

Using an Agilent 1260 Infinity HPLC and Phenomenex Kinetex C_{18} 5 μm 4.6 x 150 mm column, flow rate 1 mL/min, wavelength 283 nm, gradient: 0-1 min 5% B, 1-2 min 5-10% B, 2-8 min 10-75% B, 8-8.1 min 75-95% B, 8.1-10.0 min 95% B, 10.0-10.1 min 95-5% B, 10.1-11.5 min 5%, mobile phase A: H_2O 0.05% TFA, mobile phase B: acetonitrile 0.05% TFA.

6.6 Preparative HPLC Methods

6.6.1 HPLC method for preparative purification of SAM analogues

SAAH and SAMH purification (1c, 1f)- Using a Varian ProStar HPLC and Varian Pursuit C_{18} 10 μm 212 x 250 mm column, flow rate 10 mL/min, wavelength 260 nm, gradient: 0-15 min 3-7% B, 15-20 min 7-70% B, 20-22 min 70-97% B, 22-25 min 97% B, 25-25.1 min 97-3% B, 25.1-28 min 3% B, mobile phase A: H_2O 0.05% TFA, mobile phase B: acetonitrile 0.05% TFA.

SABH purification (1d)- Using a Varian ProStar HPLC and Varian Pursuit C_{18} 10 μm 212 x 250 mm column, flow rate 10 mL/min, wavelength 260 nm, gradient: 0-5

min 5-7% B, 5-10 min 7-22% B, 10-25 min 22-30% B, 25-25.1 min 30-95% B, 25.1-30.0 min 95% B, 30-30.1 min 95-5% B, 30.1-35 min 5% B mobile phase A: H_2O 0.05% TFA, mobile phase B: acetonitrile 0.05% TFA.

POB SAM (1e) purification- Using a Varian ProStar HPLC and Phenomenex Kinetex C_{18} 5 μm 10 x 150 mm column, flow rate 5 mL/min, wavelength 260 nm, gradient: 0-6 min 5-25% B, 6-7 min 25-95% B, 7-9 min 95% B, 9-9.1 min 95-5% B, 9.1-15 min 5% B, mobile phase A: H_2O 0.05% TFA, mobile phase B: acetonitrile 0.05% TFA.

6.6.2 HPLC method for preparative purification of 3-ethoxy-4-hydroxybenzaldehyde (3c)

Using a Varian ProStar HPLC and Phenomenex Kinetex C_{18} 5 μm 10 x 150 mm column, flow rate 5 mL/min, wavelength 278 nm, gradient: 0-1 min 5% B, 1-2 min 5-35% B, 2-8 min 35-70% B, 8-8.1 min 70-95% B, 8.1-10 min 95% B, 10-10.1 min 95-5% B, 10.1-11 min 5% B, mobile phase A: H_2O 0.05% TFA, mobile phase B: acetonitrile 0.05% TFA.

6.6.3 HPLC method for preparative purification of 3-allyloxy-4-hydroxybenzaldehyde (3e)

The solution was then purified by RP-HPLC using a Varian Pro Star HPLC system: (Phenomenex Kinetex C_{18} 5 μm 10 x 150 mm column, flow rate 5 mL/min, wavelength 278 nm), gradient 0-1 min 5% B, 1-2 min 5% B-35% B, 2-8 min 35% B-70% B, 8-10 min 95% B, 10.0-12 min 5% B. Mobile phase A consisted of H_2O + 0.05% TFA, mobile phase B consisted of acetonitrile + 0.05% TFA.

6.6.4 HPLC method for preparative purification of 3-benzyloxy-4-hydroxybenzaldehyde (3g)

The solution was then purified by RP-HPLC using a Varian Pro Star HPLC system: (Phenomenex Kinetex C_{18} 5 μm 10 x 150 mm column, flow rate 5 mL/min, wavelength 278 nm), gradient 0-1 min 5% B, 1-2 min 5% B-35% B, 2-8 min 35% B-70% B, 8-10 min 95% B, 10-12 min 5% B. Mobile phase A consisted of H_2O + 0.05% TFA, mobile phase B consisted of acetonitrile + 0.05% TFA.

6.6.5 HPLC method for preparative purification of synthetic peptides and synthetic peptide products (6a-c and 7a-c)

Using a Varian ProStar HPLC and Phenomenex Gemini C_{18} 5 μm 10 x 250 mm column, flow rate 5 mL/min, wavelength 280 nm, gradient: 0-3 min 5% B, 3-30 min 5-75% B, 30-31 min 75-95% B, 31-34 min 95% B, 34-37 min 95-5% B, 37-40 min 5% B, mobile phase A: H_2O 0.1% FA, mobile phase B: acetonitrile 0.1% FA.

6.6.6 HPLC method for preparative purification of goserelin hydroxylated/methoxylated/POBoxylated (9a-d) synthetic peptide products

Using a Varian ProStar HPLC and Phenomenex Gemini C_{18} 5 μm 10 x 250 mm column, flow rate 5 mL/min, wavelength 280 nm, gradient: 0-3 min 5% B, 3-25 min 5-50% B, 25-26 min 50-95% B, 26-29 min 95% B, 29-30 min 95-5% B, 30-33 min 5% B, mobile phase A: H_2O 0.1% FA, mobile phase B: acetonitrile 0.1% FA.

6.6.7 HPLC method for preparative purification of oxytocin hydroxylated/methoxylated (10a-c) synthetic peptide products

Using a Varian ProStar HPLC and Phenomenex Gemini C_{18} 5 μm 10 x 250 mm column, flow rate 5 mL/min, wavelength 280 nm, gradient: 0-3 min 5% B, 3-35 min 5-65% B, 35-38 min 65-95% B, 38-43 min 95% B, 43-43.1 min 95-5% B, 43-48 min 5% B, mobile phase A: H_2O 0.1% FA, mobile phase B: acetonitrile 0.1% FA.

6.7 Synthetic methods for SAM analogues

6.7.1 General synthesis of SAM analogues (*S*-adenosyl-allylhomocysteine) (SAAH) (1c-1e)
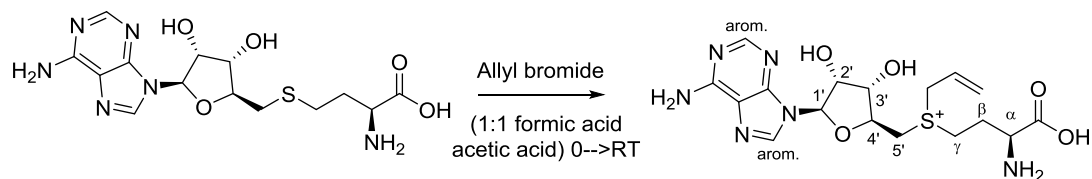
Figure 80: Synthesis of SAAH (**1c**) and general SAM analogue synthesis conditions.

SAAH (**1c**) was synthesised using the method from Dalhoff *et al.*[4] Briefly, a 1:1 mix of acetic acid and formic acid (3 ml) was added to SAH (**1**) (20 mg, 52 μmoL). The mixture was stirred on ice at 0°C until dissolved. To this solution allyl bromide (180 μL, 60 equivalents, 3.1 mmol) was slowly added by drop wise addition. The ice bath was removed and the solution was left stirring for four days. Ice cold distilled water (30 mL) was then added and the resultant solution was washed with diethyl ether (3 x 5 mL). The aqueous layer was then lyphophilized and purified by preparative RP-HPLC (Method Section 6.6.1) yielding **1c** (20 mg, 47 μmol, 90.5%) ^{1}H (400 MHz, D_2O) δ 8.37 (s, 2H, H arom) 6.16 (d, *J* = 3.7, 1H, H1′), 5.87 (m, 1H, H2″), 5.69 (d, *J* = 10.7 Hz, 1H, H3B″), 5.62 (d, *J* = 10.7 Hz, 1H, H3a″), 4.86 (m, 1H, H2′), 4.59 (m, 1H, H3′), 4.50 (ddd, *J* = 6.8, 5.0 Hz, 1H, H4′), 4.14 (d, *J* = 7.9 Hz, 2H H1″), 3.90 (d, *J* = 6.8 Hz, 1H, H5b′), 3.87 (d, *J* = 6.8 Hz, 1H, H5a′), 3.78 (t, *J* = 6.6 Hz, 1H, Hα),

3.56 (m, 1H, Hγ_b), 3.46 (m, 1H, Hγ_a), 2.33 (q, J = 6.6 Hz, 2H, H$_\beta$). MALDI m/z (relative intensity): 425.2 (100) [M]$^+$, (expected mass C$_{17}$H$_{25}$N$_6$O$_5$S$^+$: 425.2).

## 6.7.2 Synthesis of *S*-adenosyl-benzylhomocysteine (SABH) (1d)
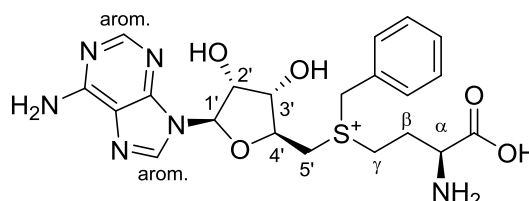
Figure 81: Structure of SABH (**1d**).

SABH (**1d**) was synthesised adapting the method from Dalhoff *et al.*[4] The method was the same as Section 6.7.1 but with a benzyl bromide alkylating agent (185 μL, 60 equivalents, 3.12 mmol). The product was isolated and purified (using Method Section 6.6.1) to give **1d** (22 mg, 46 μmol, 89.3%).^{1}H (400 MHz, D$_2$O) δ 8.15 (s,1H, H arom.), 8.10 (s,1H, H arom.), 7.20 (m, 5H, Ph), 5.94 (d, J = 3.5 Hz, 1H, H1$^{'}$), 4.63 (m, 1H, H2$^{'}$), 4.54 (m,1H,H$_3$'), 4.39 (m, 1H, H4$^{'}$), 3.85 (d, J = 6.8 Hz, 1H, H$_{5b}$$^{'}$) 3.77(d, J = 6.8 Hz, 1H, H$_{5a}$$^{'}$), 3.71 (t, J = 7.9 Hz, 1H, H$_\alpha$), 3.59 (m,1H, Hγ_b), 3.41 (m, 1H, Hγ_a), 2.64 (s, S$^+$-CH$_2$-Ph), 2.24 (q, J = 7.9Hz, 2H, H$_\beta$). MALDI m/z (relative intensity): 475.2 (100) [M]$^+$ (expected mass for C$_{21}$H$_{27}$N$_6$O$_5$S$^+$: 475.2).

6.7.3 Synthesis of *S*-adenosyl-methylacetatehomocysteine (SAMH) (1f) analogue
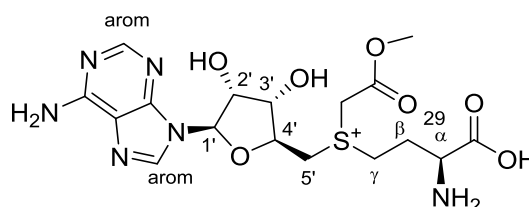
Figure 82: Structure of SAMH (**1f**)

SAMH (**1f**) was synthesised adapting the method from Dalhoff *et al.*[4] The method was the same as Section 6.7.1 but with a bromomethylacetate alkylating agent (310

μL, 60 equivalents, 3.12 mmol). The product was isolated and purified (using Method Section 6.6.1) to give **1f** (1 mg, 3 μmol, 5%).^{1}H (400 MHz, D$_2$O) δ 8.42 (s, 1H, H arom.), 8.35 (s, 1H, H arom.), 6.08 (m, 1H, H1$^{'}$), 4.81 (m, 1H, H2$^{'}$), 4.53 (m, 1H, H$_3$'), 4.35 (m, 1H, H4$^{'}$), 3.57 (s, 3H, Ac-OMe) 2.71 (q, J = 7.9Hz, 2H, H$_\beta$). MALDI m/z (relative intensity): 457.2 (100) [M]$^+$ (C$_{17}$H$_{25}$N$_6$O$_7$S$^+$ expected mass 457.2).

6.7.4   Synthesis of POB Linker (2f)
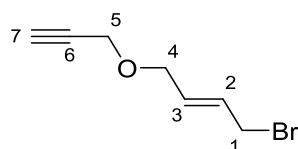
Figure 83: Structure of the POB linker (**2f**).

The POB linker (**2f**) was synthesised using the method from Wang *et al.*[59] Propargyl alcohol (0.6 mL, 10 mmol) and *trans*-1,4-dibromo-2-butene (2.6 g, 12 mmol) were dissolved with stirring in DCM (10 mL). Aqueous sodium hydroxide solution (8 M, 6 mL) and tetrabutylammonium bromide (40 mg) were subsequently added to the solution. Addition of the sodium hydroxide and tetrabutylammonium bromide resulted in a biphasic system, which was vigorously stirred at room temperature (22°C) overnight. After completion distilled water (30 mL) and dichloromethane (30 ml) were added to the reaction mixture, and the organic fraction separated from the aqueous layer. The aqueous layer was then extracted with dichloromethane (2 × 30 mL).The combined organic phases were again washed with distilled water and then brine. The organic layers were then dried over magnesium sulphate. The solvent was removed under reduced pressure and the **2f** was purified by flash chromatography (15% dichloromethane/hexane) yielding the **2f** (650 mg, 3 mmol, 34% yield). ^{1}H NMR (400 MHz, CDCl$_3$) δ 5.97 (dtt, J = 15.0, 7.5, 1.2 Hz, 1 H, H2), 5.83 (dtt, J = 15.0, 5.6, 1.0 Hz, 1 H,H3), 4.14 (d, J = 2.5 Hz, 2 H, H5), 4.08 (dq, J = 5.6, 1.2 Hz, 2 H, H4), 3.95 (dq, J = 7.5, 1.0 Hz, 2 H, H1), 2.45 (t, J = 2.5 Hz, 1 H, H7). ^{13}C (100 MHz, CDCl$_3$) δ 130.7 (C3), 129.4 (C2), 79.33 (C6) 74.6 (C7), 68.6 (C4), 57.3 (C5),

31.7 (C1). IR (neat) $\tilde{\upsilon}$ = 3291, 2851, 2100, 1354cm^{-1}. ES$^+$MS m/z (%) 189.9 (100) [M+H]$^+$ (Mass expected for $C_7H_{10}BrO^+$ 189.0).

6.7.5 Synthesis of *S*-adenosylPOBhomocysteine (POB SAM) (1e)
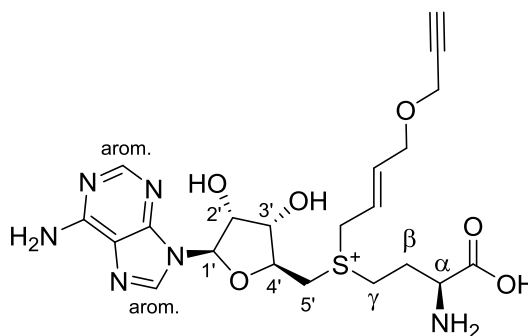
Figure 84: Structure of POB SAM (**1e**).

POB SAM (**1e**) was synthesised using the method from Wang *et al.*[71] In brief, SAH (**1**) (20 mg, 0.052 mmol, 1 equivalent) was dissolved in a mixture of formic and acetic acid (1:1, 3 mL) and stirred at 0 °C. To this mixture was firstly added the POB (**2f**) (300 mg, 1.5 mmol, 30 equivalents) and silver chlorate (24 mg, 0.10 mmol, 2 equivalents). The mixture was then stirred at room temperature for 2.5 hours. After this time an additional portion of each the **2f** (300 mg, 1.5 mmol, 30 equivalents) and silver perchlorate (24 mg, 0.10 mmol, 2 equivalents) and the reaction was left for a further 2.5 hours. At completion, the collected silver bromide precipitate was spun off via centrifugation (4°C, 12000 rpm, 10 minutes) and supernatant added to 30 mL iced water. The resultant cloudy solution was extracted with diethyl ether and the aqueous phase was lyophilised overnight (after freezing in dry ice). The lyphophilized crude was then purified by preparative RP-HPLC (Method Section 6.6.1) yielding **1e** (6 mg, 13 μmol, 25.1%). ^{1}H (400 MHz, D$_2$O) δ 8.42 (m,1H, H arom.), 8.41 (m, 1H, H arom.), 6.12 (s, 1H, H1$^{'}$), 6.07 (m, 0.5H), 5.92 (dt, 0.5 Hz, J = 14.7, 5.6Hz), 5.75 (m, 0.5H), 5.64 (m, 0.5H), 4.75 (m, 1H, H2$^{'}$), 4.62 (m, 1H, H3'), 4.50 (m, 1H, H4$^{'}$), 4.16 (s, 2H, H-C≡C-CH$_2$-O), 4.10 (d, J = 2Hz, 1H), 4.01 (d, J = 5.6 Hz, 1H) 3.85 (m, 2H, H$_5^{'}$) 3.80 (m, 1H, H$_\alpha$), 3.47 (m, 2H, Hγ), 2.85 (s, 1H, R-C≡CH), 2.29 (q, 2H, J = 7.9 Hz, H$_\beta$). MALDI m/z (relative intensity): 493.0 (100) [M]$^+$ ($C_{21}H_{29}N_6O_6S^+$ expected mass 493.2).

6.8 Synthesis of methionine analogues

6.8.1 Synthesis of *S*-allyl-L-homocysteine (SAlH) (2c)
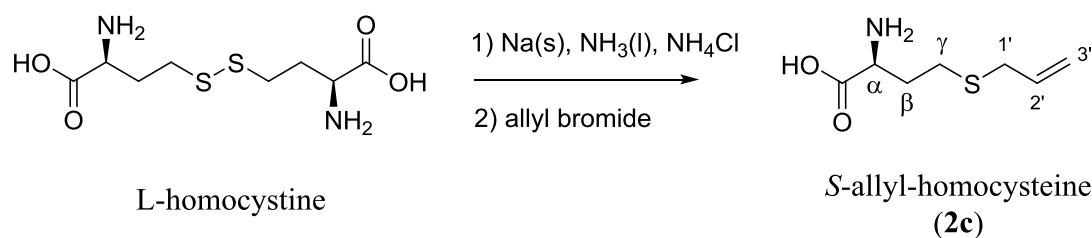
Figure 85: Synthesis of SAlH (**2c**).

SAlH (**2c**) was synthesised using the method from Thomsen *et al.*[81] Briefly, L, L-homocystine (270 mg, 1 mmol) was dissolved in liquid ammonia (30 mL) at -78°C in a dry ice acetone bath. Sodium metal (100 mg) was added until a dark blue colour persisted for 15 minutes. Ammonium chloride (20 mg) was added, to quench the reduction, until the dark blue colour of the mixture faded. After the blue colour had faded, allyl bromide (260 mg, 186 μL, 2.15 mmol) was added and the resulting mixture was stirred at -78°C for 2 hrs. The cooling bath was removed, in order to allow NH_3 to evaporate. The resulting white solid from the reaction was dissolved in water (30 mL) and washed with diethyl ether (2 x 10 mL). The pH of the aqueous phase was adjusted to pH 6 with aqueous HCl (4 M) and aqueous NaOH (1 M) solutions. The volume was then reduced *in vacuo*. The solution was then left to crystallise overnight yielding **2c** (120 mg, 0.69 mmol, 45 %) as a white solid. ^{1}H (400 MHz, D_2O/NaOD) δ 5.62-5.51 (m, 1H, H2'), 4.95- 4.90 (m, 2H, H3'), 3.97 (t, *J* = 6.3 Hz, 1H, H_α), 2.97 (d, *J* = 7.3 Hz, 2H, H1'), 2.42 (m, 2H, H_γ), 2.05-1.90 (m, 2H, H_β); ^{13}C (100 MHz, D_2O) δ 171.38 (COOH), 133.64 (C2'), 117.27 (C3'), 51.6 (C_α), 33.21 (C1'), 29.17 (C_β), 24.90 (C_γ). IR (neat) $\tilde{v}$ = 3297, 2871, 1667, 1522, 1372, 1178, 1023, 941, 845, 796, 786, 612, 546 cm^{-1}. ES$^+$MS *m/z* (%) 176.0 (100) $[M+H]^+$($C_7H_{13}NO_2S^+$ expected mass 176.2).

6.8.2 Synthesis of *S*-benzylhomocysteine (SBH) (2d)
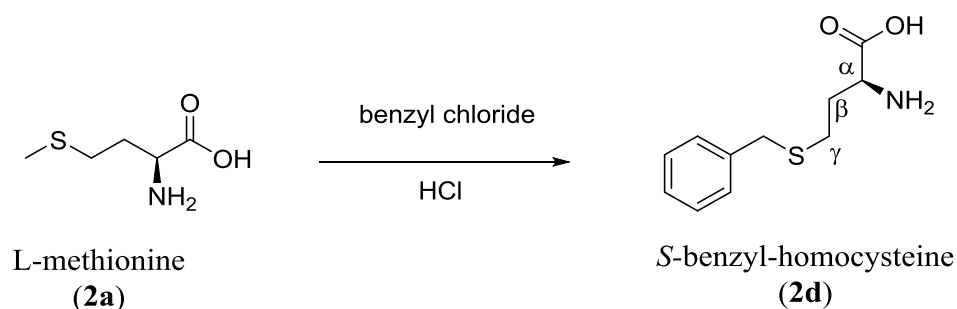
Figure 86: Synthesis of *S*-benzylhomocysteine.

SBH (**2d**) was synthesised using the method from Dekker *et al.*[129] Briefly, L-methionine (**2a**) (10.6 g, 71 mmol) was dissolved in concentrated hydrochloric acid (150 mL). (Dekker *et al*) To the stirring solution, benzyl chloride (8.2 mL, 71 mmol) was added dropwise. After dropwise addition the solution was refluxed overnight. The reaction mixtures was cooled and washed with diethylether. The aqueous phase was evaporated to dryness and was dissolved in hot water. The water was then neutralized to pH 4.7 with NH_4OH. The resulting precipitate was collected by filtration. The precipitate was recrystallized from boiling water yielding **2d** (9.3 g, 41 mmol, 58%). ^{1}H (400 MHz, D_2O) δ 7.26-7.38 (m, 5H, Ph), 3.75 (s, 2H, <u>CH$_2$</u>-Ph), 3.23 (dd, *J* = 7.3, 5.6 Hz, 1H, CH$_\alpha$), 2.46 (m, 2H, CH$_\gamma$), 1.66-1.91 (m,2H, CH$_\beta$), ES$^+$MS *m/z* (%) 226.3 (100) [M+H]$^+$, HRMS ES$^+$-MS *m/z*, mass calculated for $C_{11}H_{16}NO_2S^+$: 226.0896; found 226.0901.

6.9 Synthesis of alkylated catechols

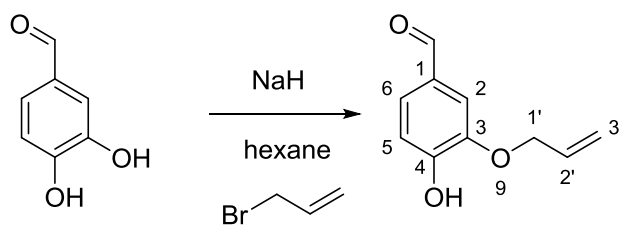### 6.9.1 Synthesis of 4-allyloxy-3-hydroxybenzaldehyde (3f)
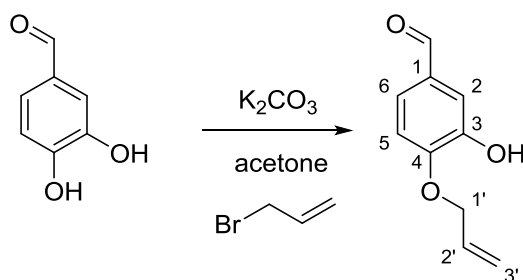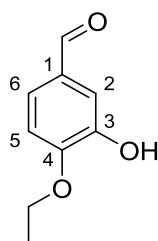

Figure 87: Synthesis of 4-alloxy-3-hydroxybenzaldehyde (**3f**).

4-allyloxy-3-hydroxybenzaldehyde (**3f**) was synthesised using the method from Reitz *et al.*[127] Briefly, a suspension of 3, 4-dihydroxybenzaldehyde (DHBAL) (**3**) (3.1 g, 23 mmol), acetone (100 mL), K_2CO_3 (3.1 g, 23 mmol) and allyl bromide (2.8 g, 2 mL, 23 mmol) was stirred for 5 hours at 60˚C. After 5 hours the suspension was filtered and the acetone removed under reduced pressure. The resulting residue was re-dissolved in diethyl ether (50 mL) and H_2O. The two layers were mixed and acidified to pH 3 with 50% aqueous sulphuric acid. The layers were then separated and the aqueous layer was extracted with diethyl ether (2 x 50 mL). The organic fractions were then combined and dried over magnesium sulphate. The solvent was removed under reduced pressure and the product was purified through flash chromatography using hexane : ethyl acetate : acetic acid (80:20:1) as the eluant to give **3f** (1.1 g, 28%). ^{1}H (400 MHz, CDCl$_3$) δ 9.84 (s, 1H, CHO), 7.42 (dd, J = 8.3, 2.0 Hz, 1H, H6), 7.47 (d, J = 2.0 Hz, 1H, H2), 6.98 (d, J = 8.3 Hz, 1H, H5), 6.29 (s, 1H, OH), 6.07 (ddt, J = 17.3, 10.6, 5.6 Hz, 1H, H2'), 5.45 (ddd, J = 17.3, 2.8, 1.5 Hz, 1H, H3b'), 5.38 (ddd, J = 10.5, 2.4, 1.2 Hz, 1H, H3a'), 4.71 (dt, J = 5.5Hz, 1.3Hz, 2H, H1'), ^{13}C (100 MHz, CDCl$_3$) δ 191.3 (CHO), 151.1 (C4), 146.3 (C3), 131.9 (C2'), 130.6 (C1), 124. (C6), 119.2 (C3'), 114.4 (C2), 111.5 (C5), 70.0 (C1'). IR (neat) $\tilde{v}$ = 3195, 1672, 1576, 1508, 1422, 1343, 1239 cm^{-1}. ES$^-$ MS (%) 177 (60) [M-H]$^-$, 136 (100) [M-C$_3$H$_5$]$^-$, ES$^+$MS *m/z* (%) 179 (100) [M+H]$^+$, HRMS ES$^+$MS *m/z*, mass calculated for C$_{10}$H$_{11}$O$_3$: 179.0708; found 179.0714.

6.9.2 Synthesis of 3-allyloxy-4-hydroxybenzaldehyde (3e)

Figure 88: Synthesis of 3-allyloxy-4-hydroxybenzaldehyde (**3e**).

3-allyloxy-4-hydroxybenzaldehyde (**3e**) was synthesised using the method from Reitz *et al.*[127] Briefly, 60% Sodium hydride (3.8 g, 81 mmol) was washed free of oil with four portions of dry hexane (30 mL) under nitrogen. Dry DMSO (60 mL) was then added. The solution was then cooled to 0˚C and stirred. To the stirring solution,

a solution of 3, 4-dihydroxybenzaldehyde (5.5 g, 40 mmol) in dry DMSO (20 mL) was then added dropwise. The solution was left stirring until no solid NaH was visible, whereupon a solution of allyl bromide (4.8 g, 3.5 mL, 40 mmol) in dry DMSO (20 mL) was added dropwise. The ice bath was then removed and the solution was left to reach room temperature whilst stirring overnight. The solution was then added to ice cool water (100 mL) and acidified with 1M HCl. The product was then extracted with ethyl acetate (3 x 100 mL). The organic layers were combined, washed with brine and dried over $MgSO_4$. The ethyl acetate was then removed under reduced pressure. The product was then purified through flash chromatography using hexane : ethyl acetate : acetic acid (80:20:1) to give **3e** (920 mg, 13%). ^{1}H (400 MHz, $CDCl_3$) δ 9.81 (s, 1H, CHO), 7.45 − 7.38 (m, 2H, H2&H6), 7.05 (d, J = 8.0 Hz, 1H, H5), 6.36 (s, 1H, OH), 6.06 (ddt, J = 17.2, 10.8, 5.6 Hz, 1H, H2'), 5.43 (ddd, J = 17.2, 2.8, 1.3 Hz, 1H, H3b'), 5.35 (ddd, J = 10.5, 2.8, 1.3 Hz, 1H, H3a'), 4.68 (dt, J = 5.6, 1.3 Hz, 2H, H1'), ^{13}C (100 MHz, $CDCl_3$) δ 191.3 (CHO), 153.3 (C4), 146.5 (C3), 132.4 (C2'), 130.2 (C1), 128.0 (C6), 119.6 (C3'), 115.1 (C5), 110.6 (C2), 70.4 (C1'). ES$^-$ MS m/z (%) 177 (60) [M-H]$^-$, 136 (100) [M-C$_3$H$_5$]$^-$, ES$^+$MS (%) 179 (100) [M+H]$^+$.

### 6.9.3   Synthesis of 4-ethoxy-3-hydroxybenzaldehyde (3d)

Figure 89: Structure of 4-ethoxy-3-hydroxybenzaldehyde (**3d**).

4-ethoxy-3-hydroxybenzaldehyde (**3d**) was synthesised adapting the method from Reitz *et al.*[127] Same method as Section 6.9.1 with ethyl iodide (3.6 g, 1.8 mL, 23 mmol) alkylating agent to give **3d** (730 mg, 19.2%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.84 (s, 1H, CHO), 7.45 (d, J = 1.9 Hz, 1H, H2), 7.42 (dd, J = 1.8, 8.2 Hz, 1H, H6), 6.95 (d, J = 8.1 Hz,1H, H5), 5.84 (s, 1H, OH), 4.22 (q, J = 7.1 Hz, 2H, CH$_2$), 1.52 (t, J = 7.1 Hz, 3H, CH$_3$), ^{13}C NMR (100 MHz, $CDCl_3$) δ 191.1 (CHO), 150.7 (C4),

146.3 (C3), 130.2 (C1), 124.7 (C6), 114.2 (C2), 111.0 (C5), 65.0 (CH_2), 14.8 (CH_3). IR (neat) $\tilde{v}$ = 3182, 2848, 1664, 1582, 1506, 1457, 1393, 1273 cm^{-1}.   ES$^-$ MS m/z (%) 165.0 (40) [M-H]$^-$, 137.0 (100) [M-C_2H_5]$^-$ (C_7H_6O expected mass 165.0557), ES$^+$MS (%) 167.0 (100) [M+H]$^+$, HRMS ES$^+$MS m/z, mass calculated for $C_9H_{11}O_3^+$: 167.0703; found 167.0703.

6.9.4   Synthesis of 4-benzyloxy-3-hydroxybenzaldehyde (3h)
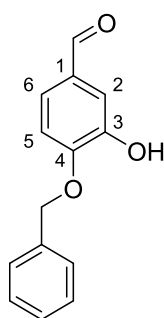
Figure 90: Structure of 4-benzyloxy-3-hydroxybenzaldehyde (**3h**).

4-benzyloxy-3-hydroxy-benzaldehyde (**3h**) was synthesised adapting the method from Reitz *et al*.[127] Same method as Section 6.9.1 with benzyl bromide (3.9 g, 2.7 mL, 23 mmol) alkylating agent yield **3h** (370 mg, 7%). ^{1}H (400 MHz, CDCl$_3$) δ 9.80 (s, 1H, CHO), 7.46 (dd, J = 1.8, 8.2 Hz, 1H, H6), 7.45-7.37 (m, 6H, Ph&H2), 7.04 (d, J = 8.3 Hz, 1H, H5), 6.12 (s, 1H, OH), 5.20 (s, 2H, CH$_2$-Ph), ^{13}C (100 MHz, CDCl$_3$) δ 191.2 (CHO), 151.0 (C4), 146.3 (C3), 135.17 (Ph), 130.6 (C1), 128.76 (Ph), 128.66 (Ph), 127.81 (Ph) 124.3 (C6) 114.0 (C2), 111.1 (C5), 71.2 (CH$_2$-Ph), IR (neat) $\tilde{v}$ = 3178, 2869, 1670, 1576, 1510, 1453, 1388, 1280 cm^{-1}. ES$^-$ MS (%) 227 (60) [M-H]$^-$, 136 (100) [M-benzyl]$^-$, ES$^+$MS m/z (%) 229 (100) [M+H]$^+$, HRMS ES$^+$-MS m/z, mass calculated for $C_{14}H_{13}O_3^+$: 229.0859; found 229.0854.

6.9.5   Synthesis of 3-benzyloxy-4-hydroxybenzaldehyde (3g)
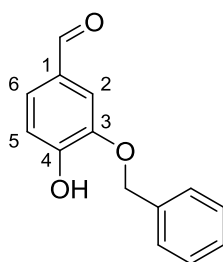
Figure 91: Structure of 3-benzyloxy-4-hydroxybenzaldehyde (**3g**).

3-benzyloxy-4-hydroxybenzaldehyde (**3g**) was synthesised adapting the method from Reitz *et al.*[127] Same method as Section 6.9.1 with benzyl bromide (6.80 g, 4.8 mL, 40 mmol) alkylating agent yield **3g** (91 mg , 1 %). ^{1}H (400 MHz, CDCl$_3$) δ 9.78 (s, 1H, CHO), 7.51 (dd, *J* = 1.8, 8.0 Hz, 1H, H6), 7.47-7.35 (m, 6H, Ph&H2), 7.07 (d, *J* = 8.1 Hz, 1H, H5), 6.56 (s, 1H, OH), 5.15 (s, 2H, H1'), ^{13}C (100 MHz, CDCl$_3$) δ 190.8 (CHO), 151.9 (C4), 146.3 (C3), 135.4 (Ph), 129.6 (C1), 128.7 (Ph), 128.6 (Ph), 128.2 (Ph), 127.6 (C6), 114.7 (C5), 110.2 (C2), 71.2(<u>CH$_2$</u>-Ph), IR (neat): $\tilde{v}$ = 3242, 2855, 1665, 1587, 1511, 1437, 1386, 1274 cm^{-1}. ES$^+$MS *m/z* (%) 229 (100) [M+H]$^+$, HRMS ES$^+$MS *m/z*, mass calculated for C$_{14}$H$_{13}$O$_3^+$: 229.0859; found 229.0861.

6.9.6   Purification and extraction of 3-ethoxy-4-hydroxybenzaldehyde (3c)
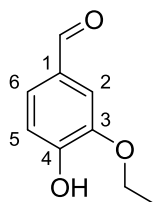
Figure 92: Structure of ethyl vanillin (**3c**).

The assay solution from Section 6.4.3 was first cleaned using an Agilent BondElut with 5% and 25% ACN (in water) washes. Elution of 3-ethoxy-4-hydroxybenzaldehyde (**3c**), 4-ethoxy-3-hydroxybenzaldehyde (**3d**), vanillin (**3a**) and isovanillin (**3b**) was achieved with 50% ACN (in water). The solution was lyphophilised, redissolved in acidic water (0.05% TFA) and purified by HPLC (Section 6.6.2) yielding **3c** (2 mg, 54.0%).^{1}H NMR (400 MHz, CDCl$_3$) δ 9.83 (s, 1H, CHO), 7.43 (d, *J* = 8.3, 1H, H6), 7.41 (s, 1H, H2), 7.06 (d, *J* = 8.3, 1H, H5), 6.21 (s,

1H, H4), 4.22 (q, J = 7.1, 2H, CH$_2$), 1.50 (t, J = 7.1, 3H, CH$_3$). ES$^+$MS m/z (%) 167 (100) [M+H]$^+$, HRMS ES$^+$MS m/z, mass calculated for C$_9$H$_{11}$O$_3^+$: 167.0703; found 167.0705.

6.9.7 Purification and extraction of 3-allyloxy-4-hydroxybenzaldehyde (3e)
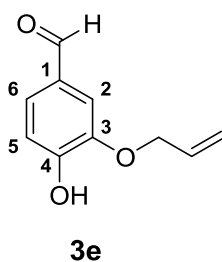
Figure 93: Structure of 3-allyloxy-4-hydroxybenzaldehyde (**3e**).

The mixture was centrifuged (4500 x g, 20 min) and the ethyl acetate layer transferred to a new vessel. The remaining aqueous was further extracted (2 x 5 mL ethyl acetate) as described previously. The organic fractions were combined and dried under a stream of nitrogen. The resulting residue was dissolved in 60% aqueous acetonitrile. 3-allyloxy-4-hydroxybenzaldehyde **3e** (14 mg, 72%) was isolated and shown to be identical to an authentic sample by ^{1}H, HPLC and mass spectrometry. ^{1}H (400MHz, CDCl$_3$) δ 9.82 (s, 1H, CHO), 7.45 − 7.42 (m, 2H, H2+H6), 7.06 (d, J = 8.0 Hz, 1H, H5), 6.30 (s, 1H, OH), 6.07 (ddt, J = 17.2, 10.3, 5.6 Hz, 1H, CH$_2$<u>CH</u>=CH$_2$), 5.45 (ddt, J = 10.4, 1.4, 1.4 Hz, 1H, CH$_2$CH=<u>CH$_2$</u>'), 5.37 (ddt, J = 10.4, 1.3, 1.3 Hz, 1H, CH$_2$CH=<u>CH$_2$</u>), 4.69 (ddd, J = 5.6, 1.4, 1.4 Hz, 2H, <u>CH$_2$</u>CH=CH$_2$), HRMS ES$^+$MS: mass calculated for C$_9$H$_{11}$O$_3^+$: 179.0703; found 179.0705.

6.9.8 Purification and extraction of 3-benzyloxy-4-hydroxybenzaldehyde (3g)
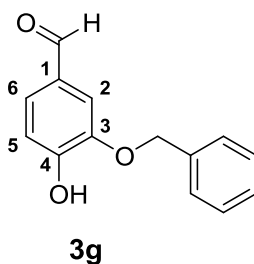
Figure 94: Structure of 3-benzyloxy-4-hydroxybenzaldehyde (**3g**).

The mixture was centrifuged (4500 x g, 20 min) and the ethyl acetate layer transferred to a new vessel. The remaining aqueous was further extracted (2 x 5 mL ethyl acetate) as described previously. The organic fractions were combined and dried under a stream of nitrogen. The resulting residue was dissolved in 60% aqueous acetonitrile. 3-benzyloxy-4-hydroxybenzaldehyde **3g** (3 mg, 63%) was isolated and shown to be identical to an authentic sample by ^{1}H, HPLC and mass spectrometry. ^{1}H (400MHz, CD$_3$CN) δ 9.77 (s, 1H, CHO), 7.52-7.35 (m, 7H, Ph, H6 & H5), 7.01 (d, J = 8.0 Hz, 1H, H5), 5.23 (s, 2H, H1'), HRMS ES^{+}MS, mass calculated for C$_{14}$H$_{13}$O$_3{}^{+}$: 229.0859; found 229.0861.

6.10 General synthesis of peptides

All synthetic peptides (**6a-b**, **7a-b**, **8a-b**) were synthesized on solid phase using the solid phase peptide synthesis strategy on a Rink amide resin (100 μmol). The resin was deprotected with 20 % piperidine in DMF. After deprotection, the resin was washed with DMF (5 x 2 mL). The amino acid at the C terminal (400 μmol) was dissolved in DMF (2 mL) with a HBTU activator (180 mg) and DIPEA (88 μL), and was coupled to the resin. The peptide was capped after every coupling step with 10% acetic anhydride in DMF. After capping, the resin was again washed in DMF (5 x 2 mL) before the cycle started again with the coupling step using the adjacent amino acid to the N terminal sequence. After the final deprotection, the resin was washed with DMF (5 x 2 mL), DCM (2 mL) and cold isopropanol (2 mL). Peptides (**6a-b**, **7a-b**, **8a-b**) were then cleaved off the resin using trifluoroacetic acid (TFA),

Triisopropylsilane and H_2O (95:5:2, v/v). All peptides were then purified by the methods shown in Section 6.6. All peptides were characterized by MALDI-TOF-MS on a Bruker Ultraflex in reflector mode using dihydroxybenzoic acid (DHB) or α-cyano-4-hydroxycinnamic acid (CCA) as matrix. Selected peptides were further characterized by 1D- and 2D-NMR spectroscopy; see Sections 6.10.2 and 6.10.3.

6.10.1 Characterisation of KNFLDY (6a)

Table 17: 1H characterisation of KNFLDY (**6a**) based on 1H, COSY, TOCSY and NOESY data.

	NH	CH_α	CH_β	CH_β'	Aromatics and others
Lys1	-	3.84	1.69	1.69	1.19 (γ) 1.51 (δ), 2.79 (ε), 7.44 (ζ)
Asn2	8.66	4.63	2.69	2.62	-
Phe3	8.36	4.44	3.02	2.91	7.25, 7.20, 7.14 (aromatics)
Leu4	8.03	4.52	1.43	1.43	1.27 (γ) 0.76, 0.71 (δ)
Asp5	8.09	4.52	2.74	2.61	-
Tyr6	7.90	4.40	2.98	2.83	7.03, 6.72 (aromatics), 7.48, 6.83 (CO-NH_2)

6.10.2 Characterisation of KETYSK (8a)

Table 18: 1H characterisation of KETYSK (**8a**) based on 1H, COSY, TOCSY and NOESY data. The n.d. notation refers to not determined.

	NH	CH_α	CH_β	CH_β'	Aromatics and others
Lys1	-	3.81	1.75	1.75	1.32 (γ) 1.56 (δ), 2.90 (ε), n.d. (ζ)
Glu2	n.d.	4.28	1.81	1.81	2.12 (γ)
Thr3	8.29	4.16	4.00	4.00	1.01 (γ)
Tyr4	n.d.	4.52	2.92	2.92	7.01, 6.70 (aromatics 2H/6H&3H/5H)
Ser5	8.17	4.28	3.69	3.69	-
Lys6	8.21	4.15	1.75	1.75	1.33 (γ) 1.60 (δ), 2.89 (ε), n.d. (ζ), 7.47, 7.04 (CO-NH_2)

6.10.3 Characterisation of KNYLDF (7a)

Table 19: 1H characterisation of KNYLDF (**7a**) based on 1H, COSY, TOCSY and NOESY data.

	NH	CH$_\alpha$	CH$_\beta$	CH$_\beta$'	Aromatics and others
Lys1	-	3.84	1.65	1.65	1.17 (γ), 1.51 (δ), 2.80 (ε), 7.42 (ζ)
Asn2	8.65	4.66	2.66	2.66	-
Tyr3	8.35	4.38	2.95	2.81	7.02, 6.72 (aromatics)
Leu4	7.97	4.12	1.43	1.43	1.28 (γ) 0.74 (δ)
Asp5	8.06	4.52	2.74	2.56	-
Phe6	7.97	4.47	3.07	2.90	7.25, 7.21, 7.16 (aromatics), 7.49, 6.82 (CO-NH$_2$)

6.11 Characterisation of isolated peptides

6.11.1 Characterisation of KNFLDDOPA-Me (6c)
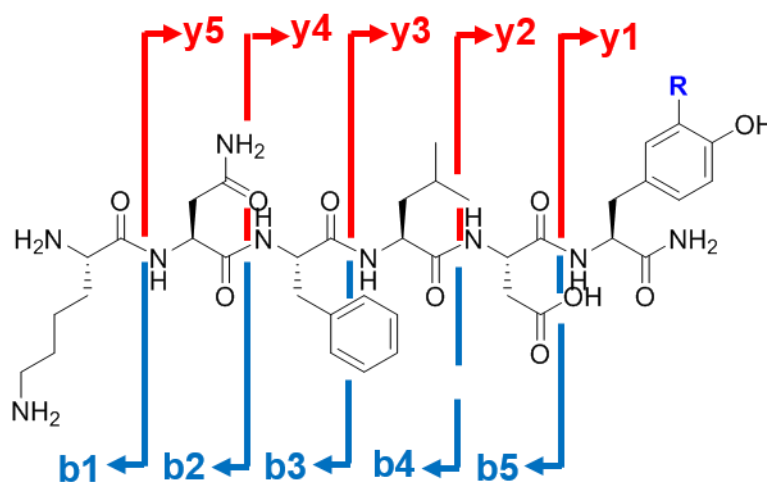
Figure 95: KNFLDDOPA-Me (**6c**) fragmentation pattern to produce b (blue) and y (red) ions. R=OMe.

b ions	Mass Expected	Mass observed	y ions	Mass Expected	Mass observed
1	129.1022	128.9	5	700.3301	700.3
2	243.1452	242.9	4	586.2871	586.2
3	390.2136	390.0	3	439.2187	439.0
4	503.2976	503.2	2	326.1347	-
5	618.3246	618.3	1	211.1077	210.9

Table 20: 1H characterisation of KNFLDDOPA-Me (**6c**) based on 1H, COSY, TOCSY and NOESY data.

	NH	CH_α	CH_β	CH_β'	Aromatics and others
Lys1	-	3.84	1.70	1.70	1.19 (γ) 1.54 (δ), 2.80 (ϵ), 7.43 (ζ)
Asn2	8.66	4.62	2.70	2.60	-
Phe3	8.35	4.43	3.01	2.90	7.24, 7.20, 7.13 (aromatics)
Leu4	8.00	4.49	1.39	1.31	1.27 (γ) 0.75, 0.70 (δ)
Asp5	8.08	4.51	2.74	2.61	-
DOPA-Me6	7.92	4.42	2.99	2.83	6.81, 6.74, 6.69 (aromatics), 7.49, 6.82 ($CO-NH_2$), 3.73 (DOPA-Me)

6.11.2 Characterisation of KNFLDDOPA-POB (6d)
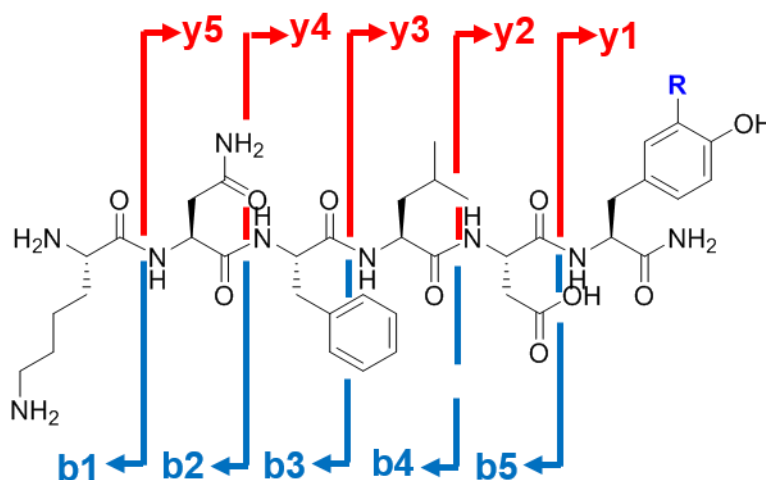
Figure 96: KNFLDDOPA-Me (6c) fragmentation pattern to produce b (blue) and y (red) ions. R=OPOB.

b ions	Mass Expected	Mass observed	y ions	Mass Expected	Mass observed
1	129.1022	129.1	5	794.3719	794.6
2	243.1452	243.1	4	680.3290	680.6
3	390.2136	390.2	3	533.2606	533.4
4	503.2976	503.4	2	420.1765	420.2
5	618.3246	618.5	1	305.1496	305.2

6.11.3 Characterisation of KETDOPA-MeSK (8c)

Table 21: 1H characterisation of KETDOPA-MeSK (**8c**) based on 1H, COSY, TOCSY and NOESY data.

	NH	CH_α	CH_β	CH_β'	Aromatics and others
Lys1	-	3.98	1.78	1.78	1.32 (γ+γ') 1.58 (δ+δ'), 2.88 (ϵ+ ϵ'), 7.47 (ζ)
Glu2	8.70	4.33	1.84	1.84	2.26 (γ+γ')
Thr3	8.27	4.16	4.00	4.00	1.03 (γ)
DOPA-Me4	8.24	4.55	2.94	2.86	6.77, 6.71, 6.63 (aromatics 2H&5H&6H), 3.73 (DOPA-Me)
Ser5	8.11	4.30	3.70	3.70	-
Lys6	8.19	4.12	1.75	1.75	1.32 (γ&γ') 1.62 (δ&δ'), 2.89 (ϵ& ϵ'), 7.47 (ζ), 7.46, 7.03 ($CO-NH_2$)

6.11.4 Characterisation of KNDOPA-MeLDF (7c)
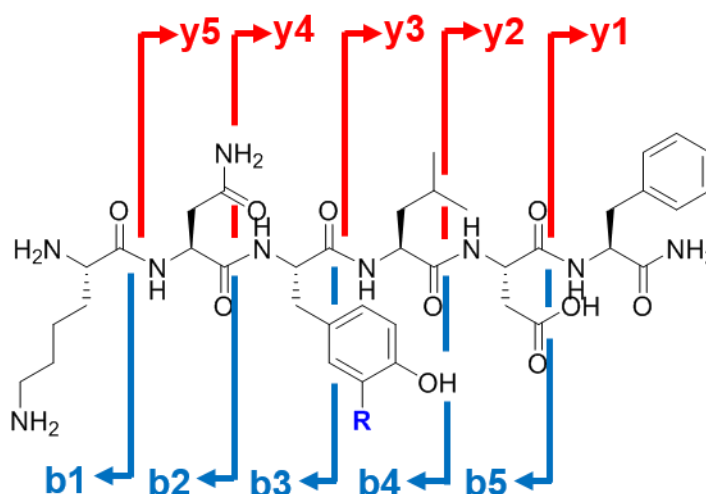
Figure 97: KNDOPA-MeLDF (**7c**) fragmentation pattern to produce b (blue) and y (red) ions. R=OMe.

Table 22: MS-MS fragments expected and observed for KNDOPA-MeLDF.(**7c**)

b ions	Mass Expected	Mass observed	y ions	Mass Expected	Mass observed
1	129.1022	128.8	5	700.3301	699.8
2	243.1452	242.6	4	586.2871	585.8
3	436.2191	435.7	3	393.2132	-
4	549.303	548.8	2	280.1292	279.5
5	664.3301	663.9	1	165.1022	165.7

Table 23: 1H characterisation of KNDOPA-MeLDF (**7c**) based on 1H, COSY, TOCSY and NOESY data. The n.d. notation refers to not determined.

	NH	CH_α	CH_β	CH_β'	Aromatics and others
Lys1	-	n.d.	n.d.	n.d.	n.d. (γ), n.d (δ), n.d. (ϵ), n.d. (ζ)
Asn2	8.65	4.44	2.65	2.65	-
DOPA-Me3	8.40	4.40	2.97	2.81	6.82, 6.73, 6.65 (aromatics), 3.74 (DOPA-Me)
Leu4	7.95	4.09	1.47	1.47	1.21 (γ) 0.70, 0.77 (δ)
Asp5	8.05	4.52	2.72	2.58	-
Phe6	7.97	4.63	3.06	2.89	7.25, 7.21, 7.16 (aromatics), 7.49, 6.80 ($CO-NH_2$)

6.11.5 Characterisation of goserelin (9a)
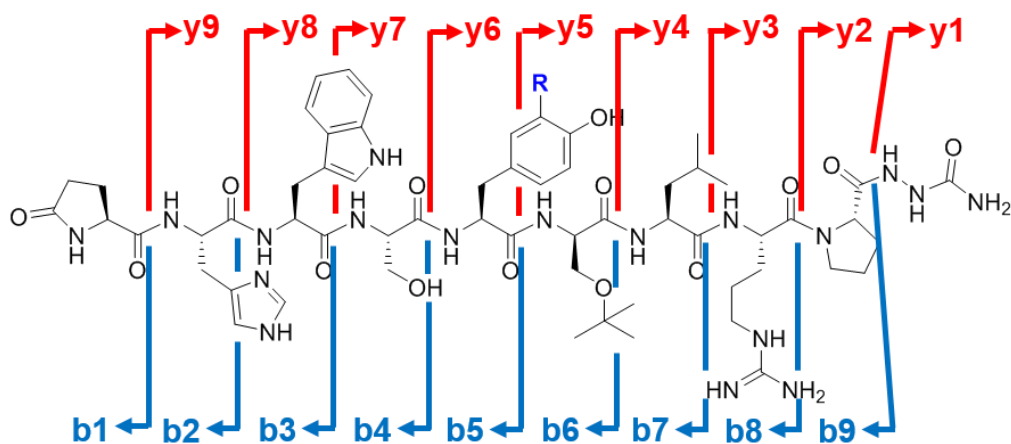
Figure 98: Goserelin (**9a**) fragmentation pattern to produce b (blue) and y (red) ions. R=H.

Table 24: MS-MS fragments expected and observed for goserelin (**9a**).

b ions	Mass Expected	Mass observed	y ions	Mass Expected	Mass observed
1	112.0339	111.9	9	1158.6167	-
2	249.0982	248.8	8	1021.5578	1021.5
3	435.1775	434.9	7	835.4785	835.3
4	522.2096	521.9	6	748.4465	748.2
5	685.2729	685.0	5	585.3831	-
6	828.3675	-	4	442.2885	-
7	941.4516	-	3	329.2044	-
8	1097.5527	1097.4	2	173.1033	173.0
9	1194.6055	-	1	76.0505	-

6.11.6 Characterisation of methoxygoserelin (9c)
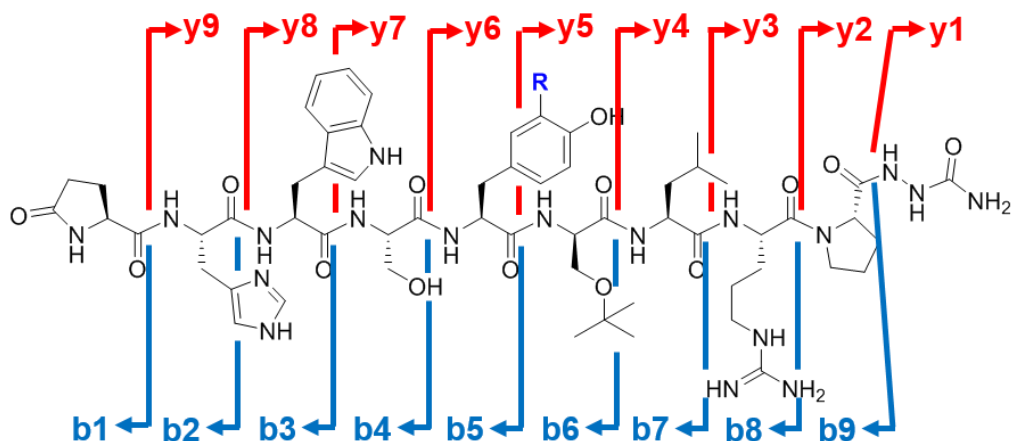
Figure 99: Methoxygoserelin (**9c**) fragmentation pattern to produce b (blue) and y (red) ions. R=OMe.

Table 25: MS-MS fragments expected and observed for methoxygoserelin (**9c**).

b ions	Mass Expected	Mass observed	y ions	Mass Expected	Mass observed
1	112.0339	-	9	1188.6273	-
2	249.0982	248.8	8	1051.5683	-
3	435.1775	434.8	7	865.489	-
4	522.2096	521.9	6	778.457	779.0
5	715.2835	-	5	585.3831	-
6	858.3781	-	4	442.2885	-
7	971.4621	-	3	329.2044	329.0
8	1127.5633	1127.3	2	173.1033	-
9	1224.616	-	1	76.0505	-

Table 26: 1H characterisation of methoxygoserelin (**9c**) based on 1H, COSY, TOCSY and NOESY data.

	NH	CH$_\alpha$	CH$_\beta$	CH$_\beta$'	Aromatics and others
Pyr1	-	4.09	2.25	2.25	2.17 (γ)
His2	8.33	4.55	3.06	2.97	8.44, 7.08 (aromatic 2H+4H)
Trp3	8.15	4.61	3.13	3.06	7.50, 7.39, 7.14, 7.02 (aromatics 4H&7H&5H&6H)
Ser4	8.06	4.24	3.63	3.63	-
DOPA-Me5	7.94	4.29	2.89	2.77	6.76, 6.75, 6.61 (aromatics 2H&5H&6H), 3.74 (DOPA-Me)
^{t}But-Ser6	7.70	4.19	3.37	3.04	0.98 (tert-Butyl)
Leu7	7.99	4.27	1.52	1.52	1.44 (γ), 0.79 (CH$_3\delta$), 0.74 (CH$_3\delta$')
Arg8	8.21	4.29	1.61	1.51	1.40 (γ), 2.97 (δ), 6.98 (NHε)
Pro9	-	4.10	2.25	2.25	1.52 (γ) 3.71 (δ), 6.07 (CO-NH$_2$)

6.11.7 Characterisation of POBoxygoserelin (9d)

Table 27: 1H characterisation of POBoxygoserelin (**9d**) based on 1H, COSY, TOCSY and NOESY data. The n.d. notation refers to not determined.

	NH	CH$_\alpha$	CH$_\beta$	CH$_\beta$'	Aromatics and others
Pyr1	-	4.02	2.80	2.80	n.d. (γ)
His2	8.29	4.56	3.01	3.01	8.45, 7.06 (aromatic 2H&4H)
Trp3	8.13	4.60	3.02	3.02	7.49, 7.35, 7.14, 7.03 (aromatics 4H&7H&5H&6H)
Ser4	8.13	4.24	3.65	3.65	-
DOPA-OPOB5	7.86	4.32	2.82	2.82	6.81, 6.67, 6.61 (aromatics 2H&5H&6H), 4.33, 5.75, 5.68, 3.95 (CH2CH=CHCH2,POB), 4.05, 2.80 (CH2C≡CH, POB)
^{t}But-Ser6	7.76	4.24	3.44	3.17	0.99 (tert-Butyl)
Leu7	8.04	4.29	1.48	1.48	1.48 (γ), 0.81 (CH$_3\delta$), 0.76 (CH$_3\delta$')
Arg8	8.21	4.32	1.60	1.60	1.42 (γ), 2.98 (δ), 7.00 (NHε)
Pro9	-	4.08	2.24	2.24	1.52 (γ) n.d. (δ), n.d. (CO-NH$_2$)

6.11.8 Characterisation of oxytocin (10a)

Table 28: 1H characterisation of commercial oxytocin (**10a**). Assignments in agreement with Ohno *et al.*[161]

	NH	CH$_\alpha$	CH$_\beta$	CH$_\beta$'	Aromatics and others
Cys1	-	4.18	3.38	3.21	-
Tyr2	8.93	n.d.	3.11	2.94	7.14 (H2&H6), 6.80 (H3&H5)
Ile3	7.89	3.99	1.87	-	1.15 (CH$_\gamma$) 0.92 (CH$_\gamma$'), 0.81 (CH$_\delta$)
Gln4	8.18	4.03	1.98	-	2.32 (CH$_\gamma$) 7.50, 6.80(CONH)
Asn5	8.29	4.66	2.77	-	7.55, 6.87 (CONH2)
Cys6	8.15	4.78	3.15, 2.89	-	-
Pro7	-	4.38	2.23	1.85	3.69(CH$_\delta$) 3.63(CH$_\delta$), 1.95 (CH$_\gamma$)
Leu8	8.43	4.24	1.62	1.55	1.80 (CH$_\gamma$) 0.87 (CH$_\delta$) 0.82 (CH$_\delta$)
Gly9	8.34	3.86, 3.82	-	-	7.35, 7.02 (CONH$_2$)

6.11.9 Characterisation of methoxyoxytocin (10c)

Table 29: 1H characterisation of isolated methoxyoxytocin (**10c**).

	NH	CH$_\alpha$	CH$_\beta$	CH$_\beta$'	Aromatics and others
Cys1	-	4.18	3.37	3.20	-
DOPA-OMe2	8.90	4.65	2.95	2.86	6.90, 6.74, 6.75(H2&H5&H6) aromatics, 3.74 (DOPA-OMe)
Ile3	7.85	3.90	1.79	-	1.05 (CH$_\gamma$) 0.75 (CH$_\gamma$'), 0.73 (CH$_\delta$)
Gln4	8.14	4.01	1.95	-	2.29 (CH$_\gamma$) 7.46, 6.77(CONH)
Asn5	8.26	4.59	2.73	-	7.52, 6.84 (CONH2)
Cys6	8.11	4.76	3.11	2.84	-
Pro7	-	4.34	2.18	1.83	3.64(CH$_\delta$) 3.60(CH$_\delta$), 1.93 (CH$_\gamma$)
Leu8	8.40	4.19	1.57	1.51	1.80 (CH$_\gamma$) 0.79 (CH$_\delta$) 0.75 (CH$_\delta$)
Gly9	8.30	3.8, 3.78	-	-	7.31, 6.98 (CONH$_2$)

6.12 Synthesis of 7-nitro-3,3a-dihydrocyclopenta[b]chromen-1(2H)-one (NDCC) (12)

7-nitro-3,3a-dihydrocyclopenta[b]chromen-1(2H)-one

NDCC (**12**)
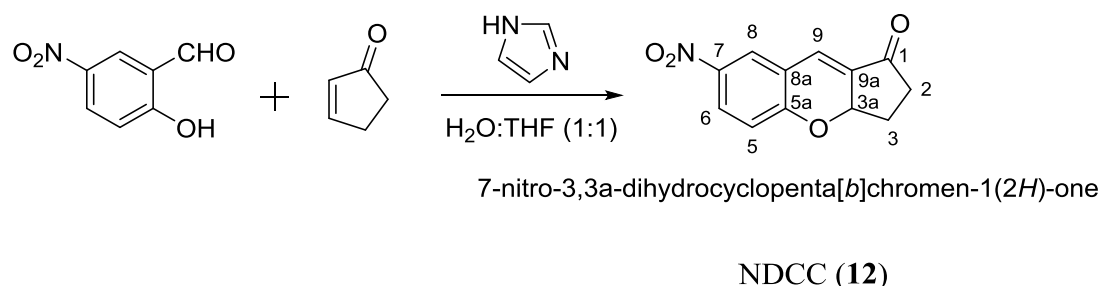
Figure 100: Synthesis of 7-nitro-3, 3a-dihydrocyclopenta[b]chromen-1(2H)-one (**12**).

7-nitro-3,3a-dihydrocyclopenta[b]chromen-1(2H)-one (**12**) was synthesised using the method from Huo *et al.*[152] 5-nitrosalicylaldehyde (180 mg, 1 mmol), 2-Cyclopenten-1-one (160 mg, 170 µL, 2 mmol) and imidazole (60 mg, 1 mmol) were dissolved in a mix of THF and deionised water (1:1). The solution was stirred at ambient temperature overnight. The solution was then diluted with deionised water (10 ml) and extracted with ethyl acetate (3 × 15 ml). The organic fractions were pooled and the solvent was removed under reduced pressure. NDCC was then purified by flash chromatography purified by flash chromatography (4:1; petroleum: ethyl acetate) yielding **12** (88 mg, 38%). ^{1}H NMR (400 MHz, CDCl$_3$) 8.07 (m,2H, H8 and H6), 7.13 (d, 1H, J = 2.5Hz, H9), 6.94 (dd, 1H, J = 6.5, 3.5Hz, H5), 5.35 (m, 1H, 3a), 2.72 (dt, 1H, J = 8.1, 11.9 Hz, H3), 2.58 (m, 1H, H2), 2.33 (m, 1H, H2), 2.15 (m, 1H, H3). ^{13}C NMR (100 MHz, CDCl$_3$) 200.3 (C1), 160.1 (C5a), 142.5 (C7), 133.1 (C8a), 127.6 (C6), 125.7 (C8), 125.4 (C9), 121.8 (C9a), 117.1 (C5), 76.6 (C3a), 36.8 (C2), 28.1 (C3). IR (neat) $\tilde{v}$ = 3060, 1709, 1647, 1605, 1576, 1510, 1339 cm^{-1}. HRMS ES$^+$MS *m/z*, mass calculated for C$_{12}$H$_9$NO$_4^+$: 232.0604; found 232.0602.

6.13 Synthesis 2-*N*-methyl-1-phenylheliamine (15a)
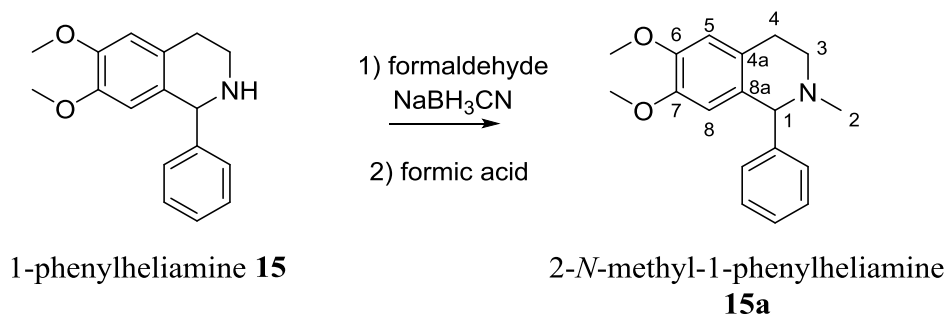
Figure 101: The Eschweiler Clarke synthesis of 2-*N*-methyl-1-phenyl-heliamine (**15a**).

2-*N*-methyl-1-phenylheliamine (**15a**) was synthesised using the method from Valpuesta *et al.*[153] Briefly, 1-phenylheliamine hydrochloride (**15**) (39 mg, 0.15 mmol) was first desalted through a partition of dichloromethane (10 mL) and sodium hydroxide (2.5 M, 10 mL). The dichloromethane layer was dried over magnesium sulphate and the solvent removed under reduced pressure to yield the solid 1-phenyl-heliamine in free amine form. Desalted **15** (39 mg, 0.15 mmol) was dissolved in MeCN (5 mL) followed by a 37% aqueous formaldehyde (130 µL, 1.7 mmol) and NaBH$_3$CN (10 mg, 1.7 mmol). After 30 min, formic acid (130 µL or until neutral) was added and the mixture was stirred overnight. The solvent was removed under reduced pressure and the crude residue was dissolved aqueous sodium hydroxide (2 M, 15 mL). The resulting suspension was extracted with chloroform (3 × 15 mL). The organic layers were combined, washed with brine and dried over anhydrous MgSO4. The solvent was evaporated under reduced pressure to give a yellow solid **15a** (31 mg, 75.5%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.30-7.21 (m, 5H, 1-Ph), 6.58, 6.08 (2 x s, 2H, H5, H8), 4.16 (s, 1H, H1), 3.82, 3.54 (2 x s, 3H, OMe-6, OMe-7), 3.14 (m, 1H, H3), 3.07 (m, 1H, H4), 2.72 (m, 1H, H3'), 2.59 (m, 1H, H4') 2.21 (s, 3H, H2). ^{13}C NMR (100 MHz, CDCl$_3$) 147.4, 147.0 (C6, C7), 143.7 (C1-Ph), 130.3, 126.5 (C4a, C8a), 129.6, 128.3, 127.2 (C1-Ph), 111.4, 110.6 (C5, C8), 71.0 (C1), 55.8, 55.7 (2 x C, OMe-6, OMe-7), 52.1 (C3), 44.2 (*N*-C2), 28.9 (C4). IR (neat) $\tilde{v}$ = 2940, 2837, 1667, 1516 cm^{-1}. HRMS ES$^+$MS *m/z*, mass calculated for C$_{18}$H$_{22}$NO$_2^+$: 284.1645; found 284.1649.

7   Supplementary information

7.1.1   Plasmid construct for pET21b-WT COMT
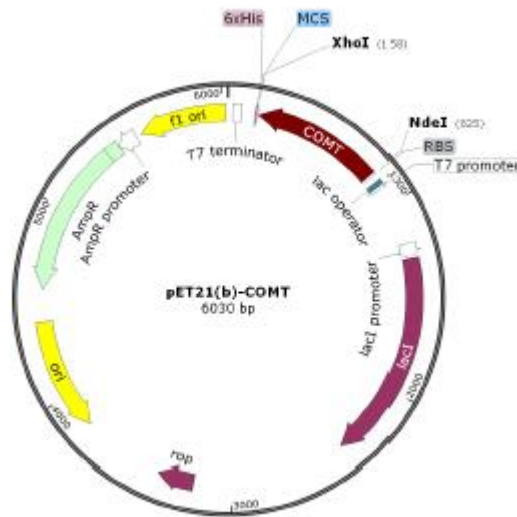
Figure 102: Plasmid construct designed for pET21b-WT COMT with C-terminal His Tag kindly donated by Professor Nigel Scrutton.

7.1.2   SDS PAGE gels
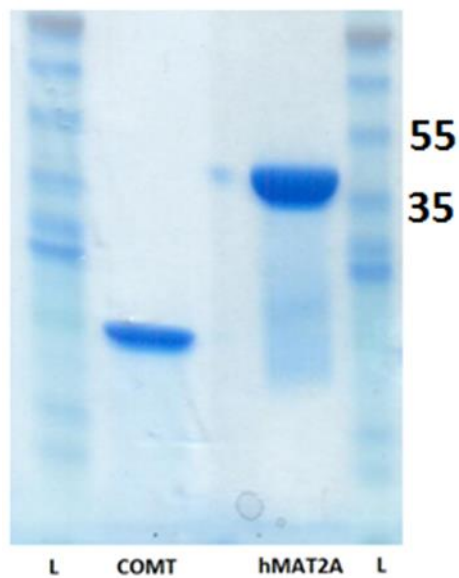
Figure 103: SDS PAGE gel of pure WT COMT and pure hMAT2A, size consistent with COMT's mass of 25 kDa and hMAT2A mass of 44 kDa.
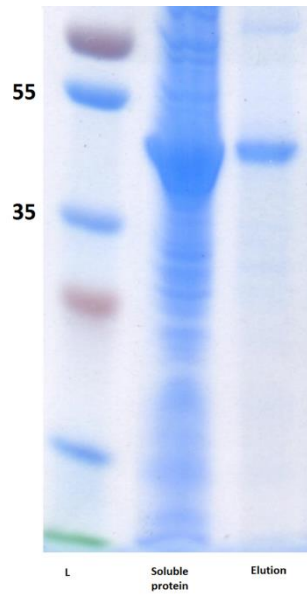
Figure 104: SDS PAGE gel of CNMT purification (from left to right); ladder, soluble protein and elution of CNMT. The size was consistent with CNMT's mass of 42 kDa.
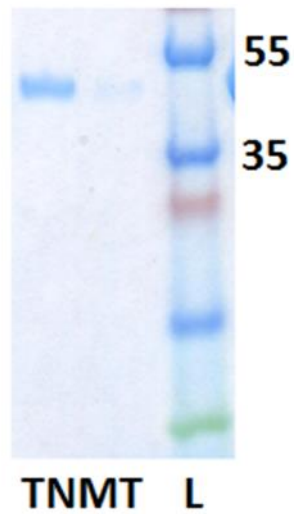
Figure 105: SDS PAGE gel of pure TNMT, size consistent with TNMT's mass of 41 kDa.

7.1.3 HPLC Chromatograms
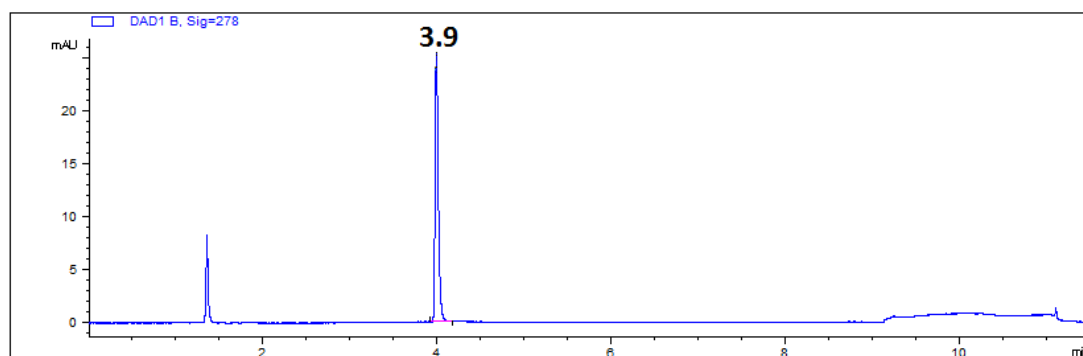
Figure 106: 0.2mM 5′-Deoxy-5′-(methylthio)adenosine standard, run on the RP HPLC method laid out in Section 6.5.1.
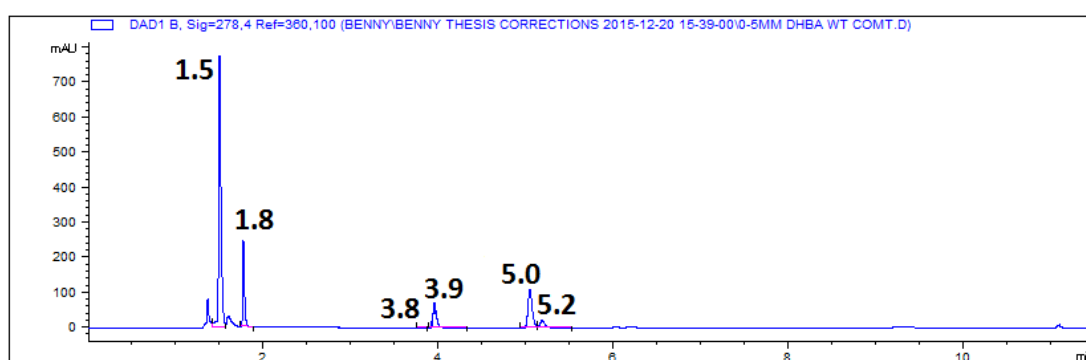
Figure 107: WT COMT assay (5 μM COMT) with DHBA (**4**) (0.5 mM) substrate for 1 hour at 37˚C. Standards can be observed in Figures; Figure 106, Figure 108, Figure 109 and Figure 110. Sample was run by the RP HPLC method laid out in Section 6.5.1.
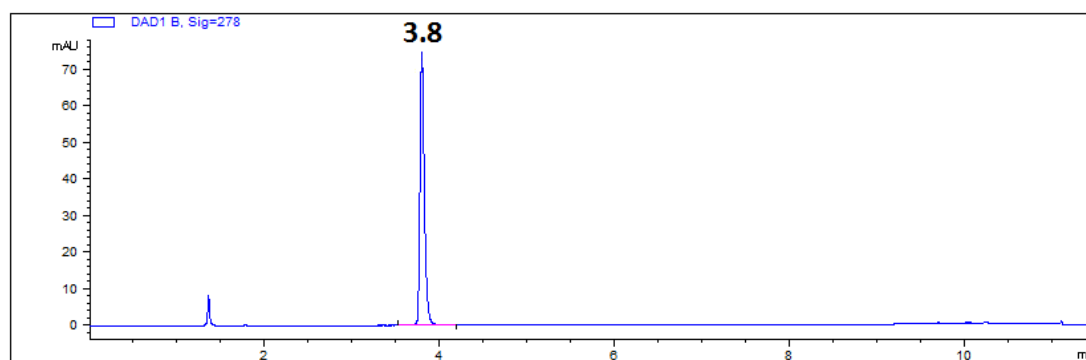
Figure 108: 0.2 mM DHBA (**4**) standard. Sample was run by the RP HPLC method laid out in Section 6.5.1.
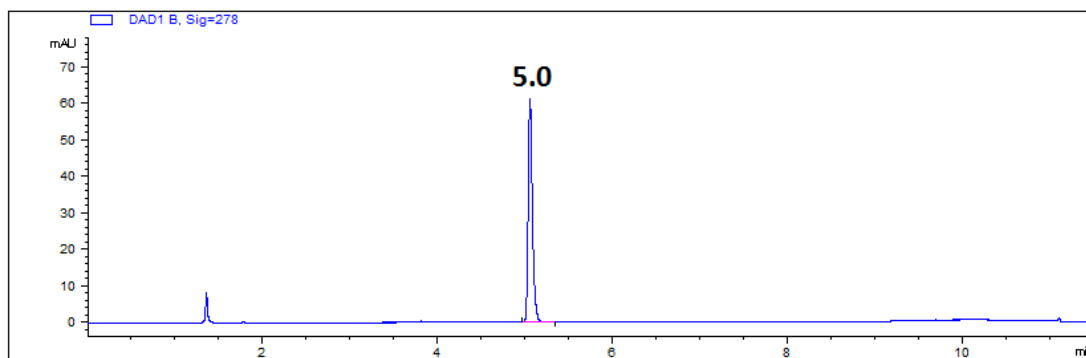
Figure 109: 0.2 mM vanillic acid (**4a**) standard (*meta* methylated DHBA). Sample was run by the RP HPLC method laid out in Section 6.5.1.
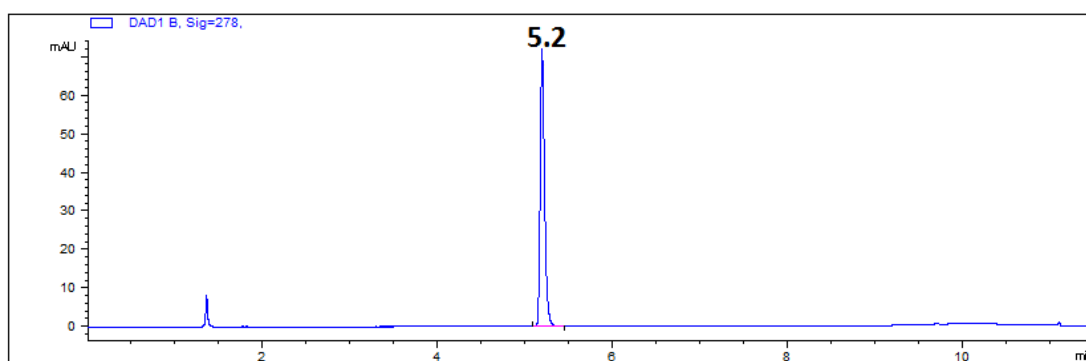
Figure 110: 0.2 mM isovanillic acid (**4b**) standard (*para* methylated DHBA). Sample was run by the RP HPLC method laid out in Section 6.5.1.
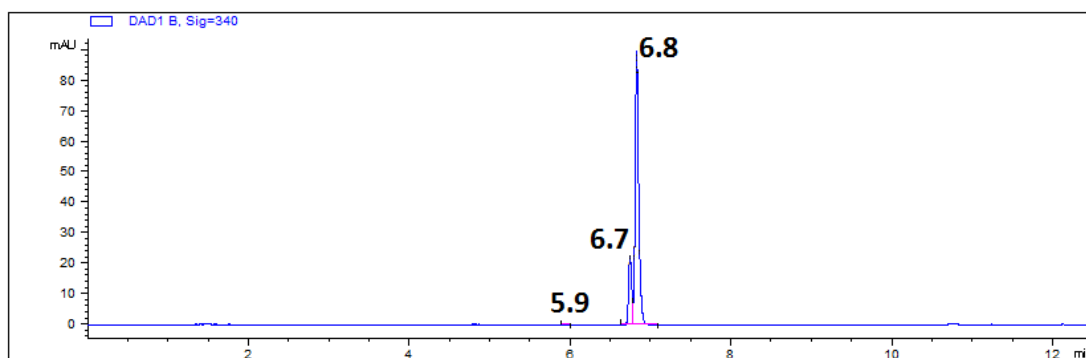
Figure 111: WT COMT assay (5 μM COMT) with NO2CAT (**5**) (0.5 mM) substrate for 1 hour at 37°C. Standards can be observed in Figures;Figure 112, Figure 113 and Figure 114. Sample was run by the RP HPLC method laid out in Section 6.5.2.
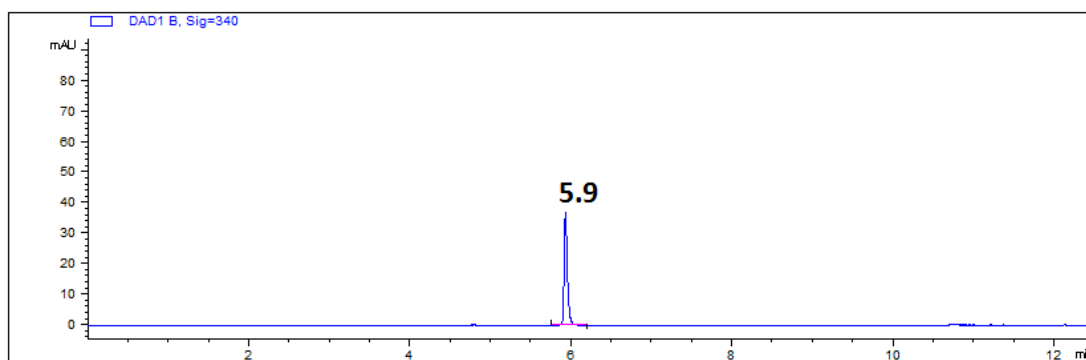
**Figure 112:** 0.2 mM NO2CAT (**5**) standard. Samples was run by the RP HPLC method laid out in Section 6.5.2.
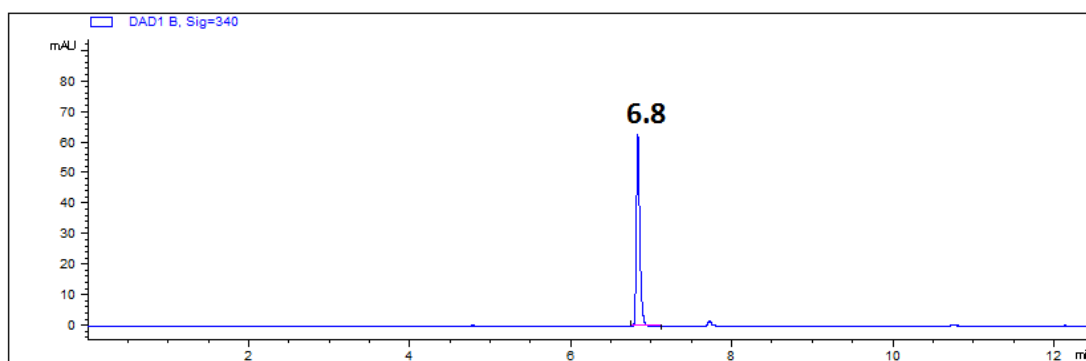
Figure 113: 0.2 mM nitroguaiacol (**5a**) standard (*meta* methylated NO2CAT). Sample was run by the RP HPLC method laid out in Section 6.5.2.
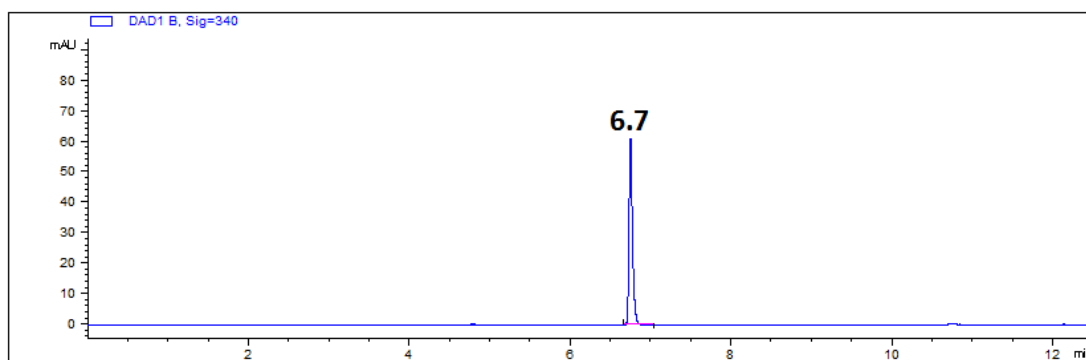
Figure 114: 0.2 mM 2-methoxy-5-nitrophenol (**5b**) standard (*para* methylated NO2CAT). Sample was run by the RP HPLC method laid out in Section 6.5.2.
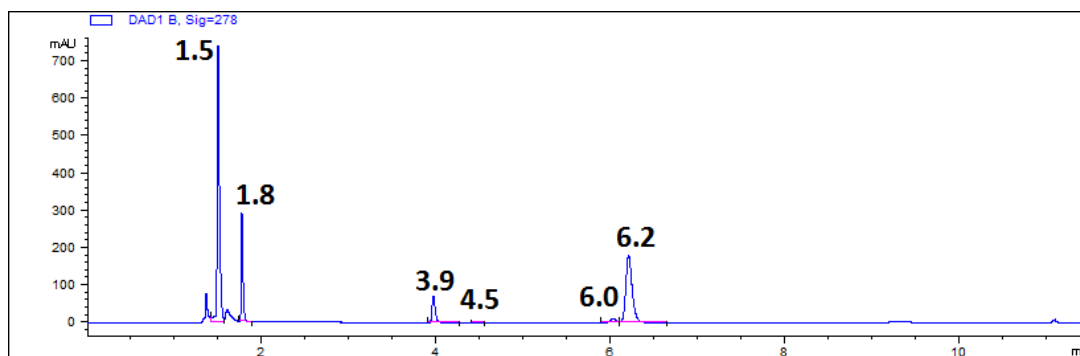
Figure 115: Y200L COMT assay (5 µM Y200L COMT) with DHBAL (**3**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.1.
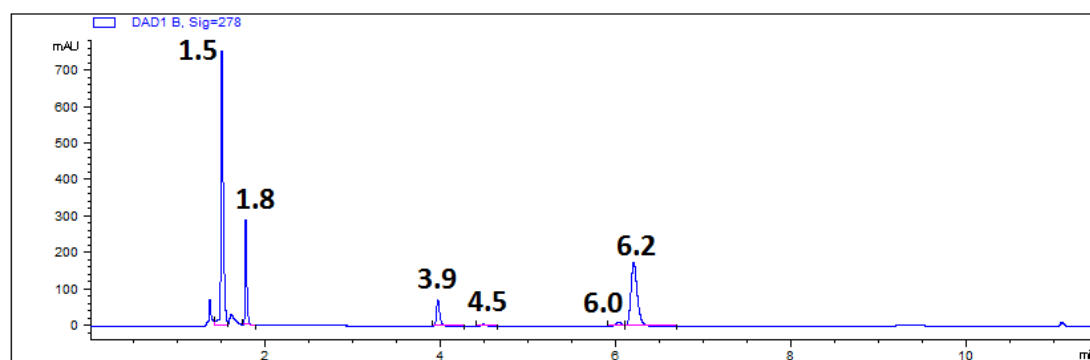
Figure 116: Y200W COMT assay (5 µM Y200W COMT) with DHBAL (**3**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.1.
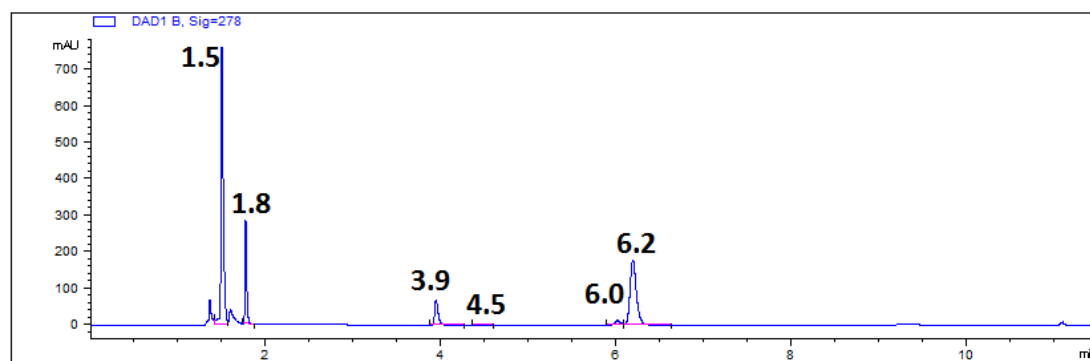
Figure 117: Y200S COMT assay (5 µM Y200S COMT) with DHBAL (**3**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.1.
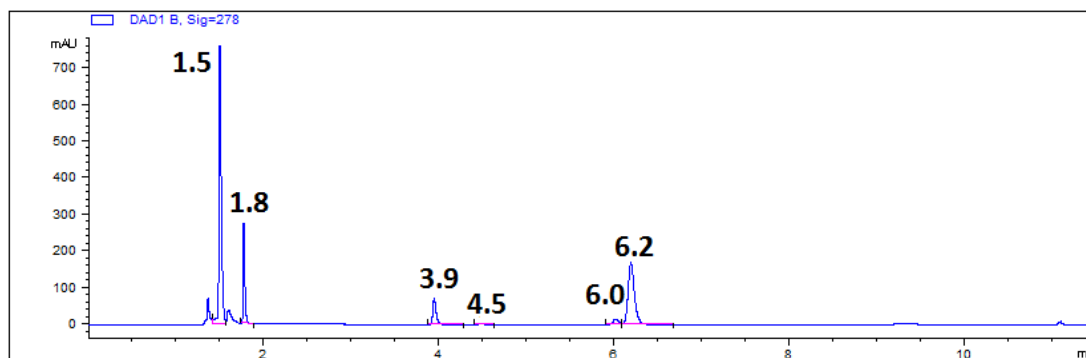
Figure 118: W38Y COMT assay (5 µM W38Y COMT) with DHBAL (**3**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.1.
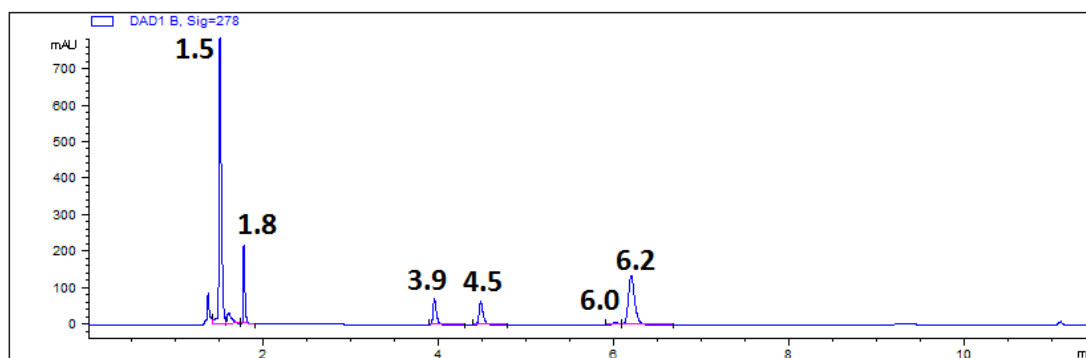
Figure 119: W38D COMT assay (5 µM W38D COMT) with DHBAL (**3**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.1.
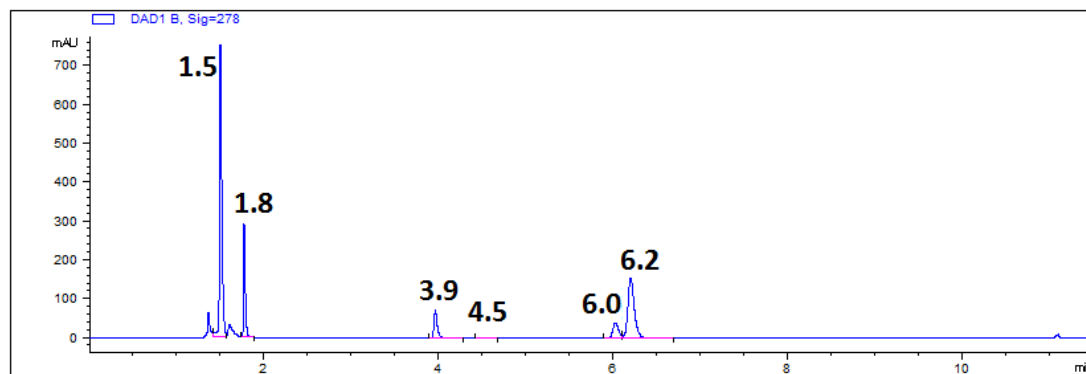
Figure 120: W38F COMT assay (5 µM W38F COMT) with DHBAL (**3**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.1.
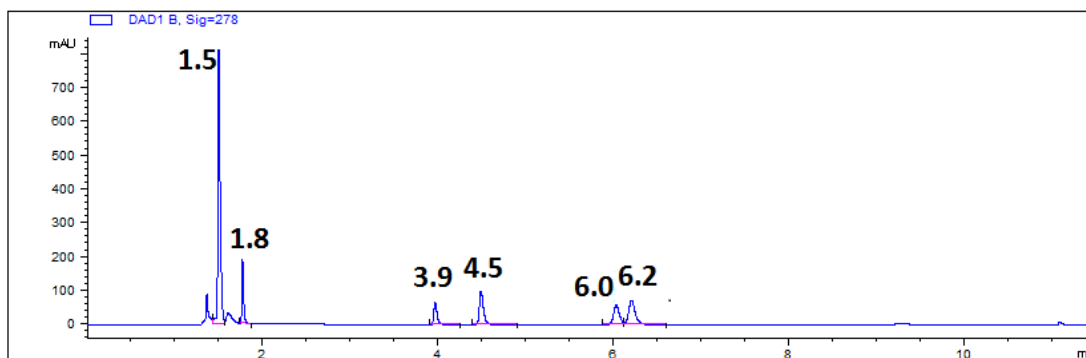
Figure 121: K144A COMT assay (5 µM K144A COMT) with DHBAL (**3**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.1.
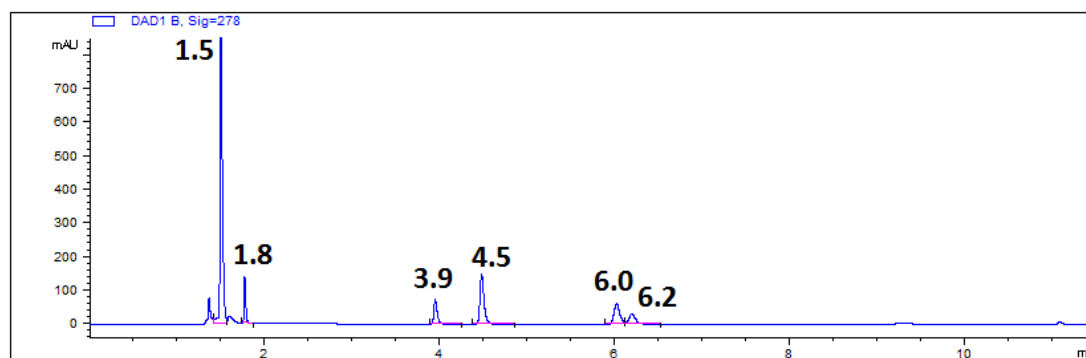
Figure 122: K144E COMT assay (5 µM K144E COMT) with DHBAL (**3**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.1.
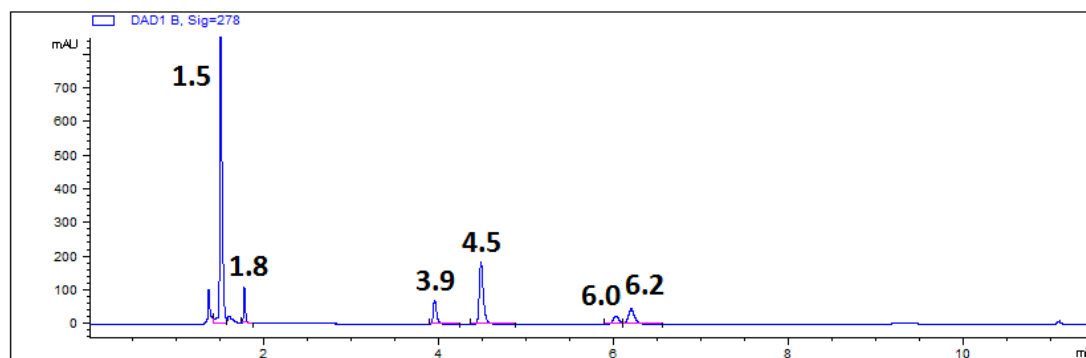
Figure 123: K144R COMT assay (5 µM K144R COMT) with DHBAL (**3**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.1.
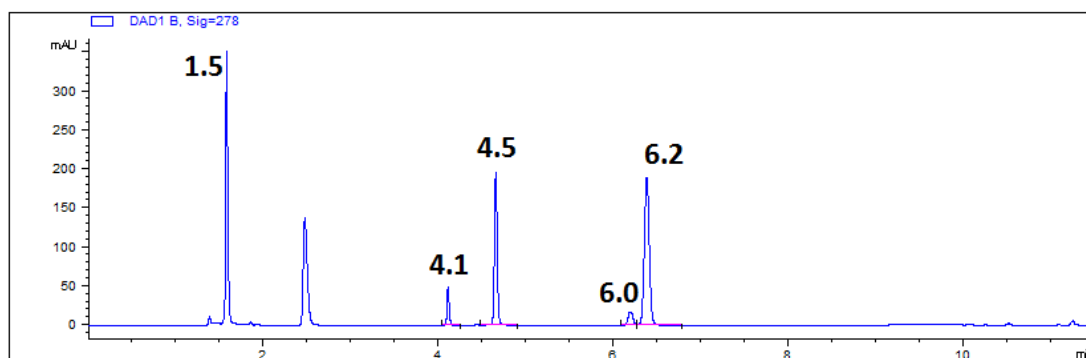
Figure 124: E199L COMT assay (5 µM E199L COMT) with DHBAL (**3**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.1.
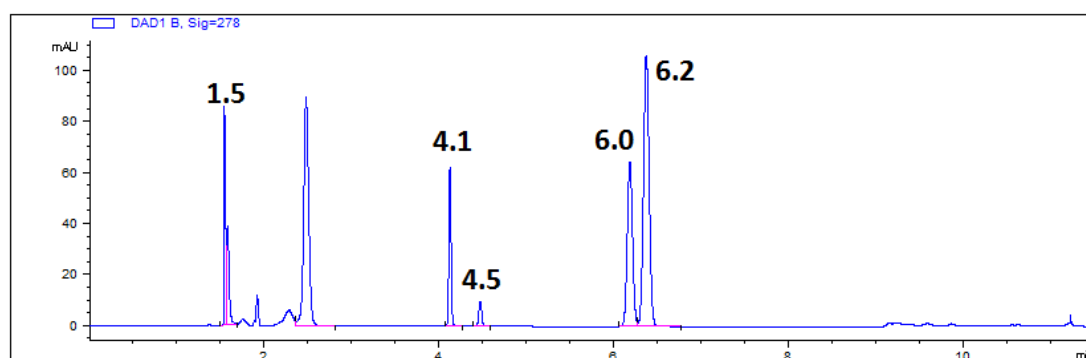
Figure 125: : WT COMT monomer assay (5 µM WT COMT monomer) with DHBAL (**3**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.1.
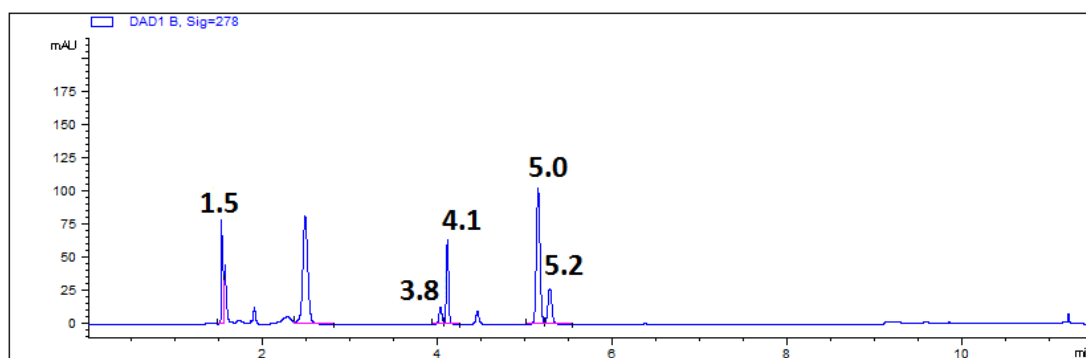
Figure 126: WT COMT monomer assay (5 µM WT COMT monomer) with DHBA (**4**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.1.
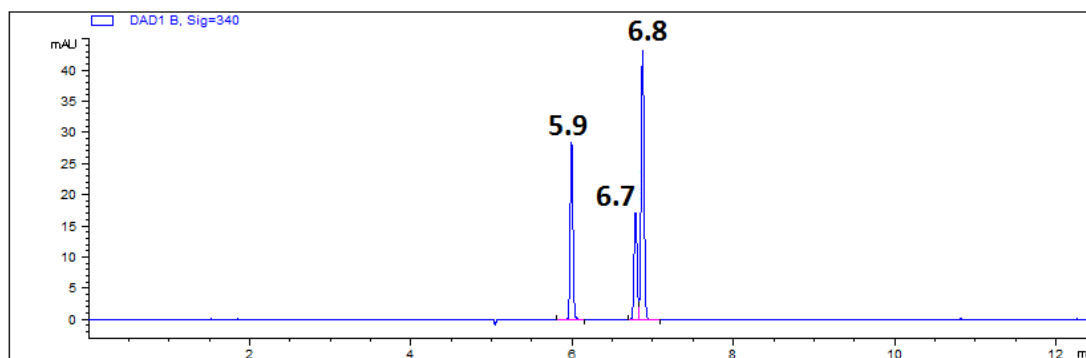
Figure 127: WT COMT monomer assay (5 μM WT COMT monomer) with NO2CAT (**5**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.2.
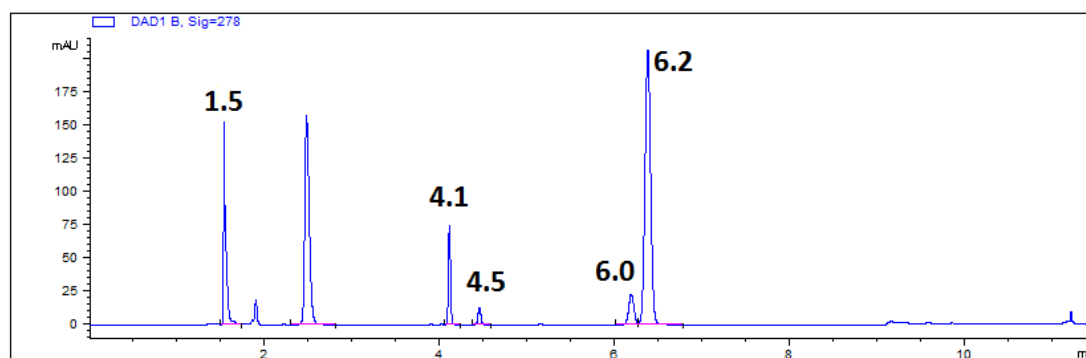
Figure 128: WT COMT dimer assay (5 μM WT COMT dimer) with DHBAL (**3**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.1.
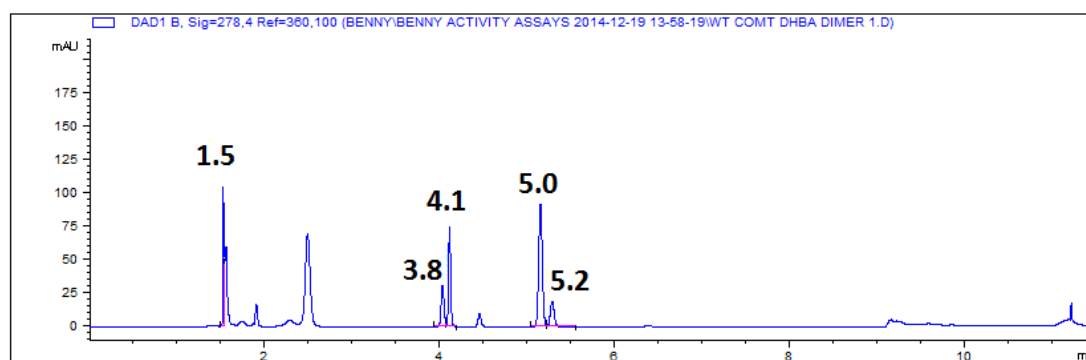
Figure 129: WT COMT dimer assay (5 μM WT COMT dimer) with DHBA (**4**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.1.
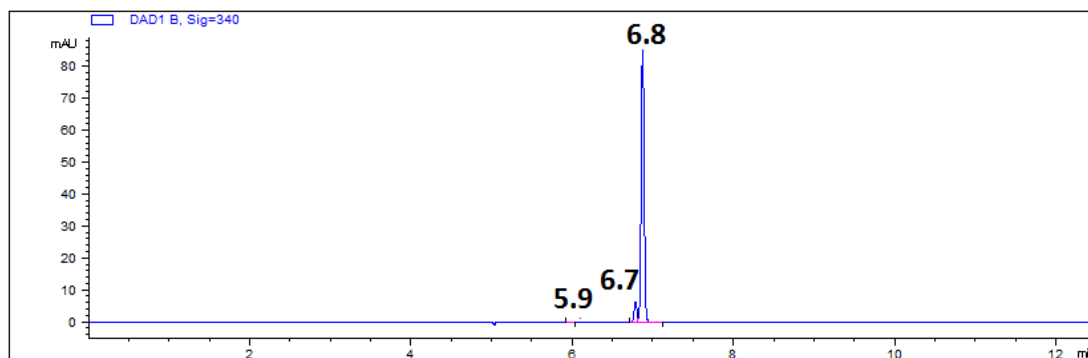
Figure 130: WT COMT dimer assay (5 μM WT COMT dimer) with NO2CAT (**5**) (0.5 mM) substrate for 25 minutes at 37˚C. Sample was run by the RP HPLC method laid out in Section 6.5.2.
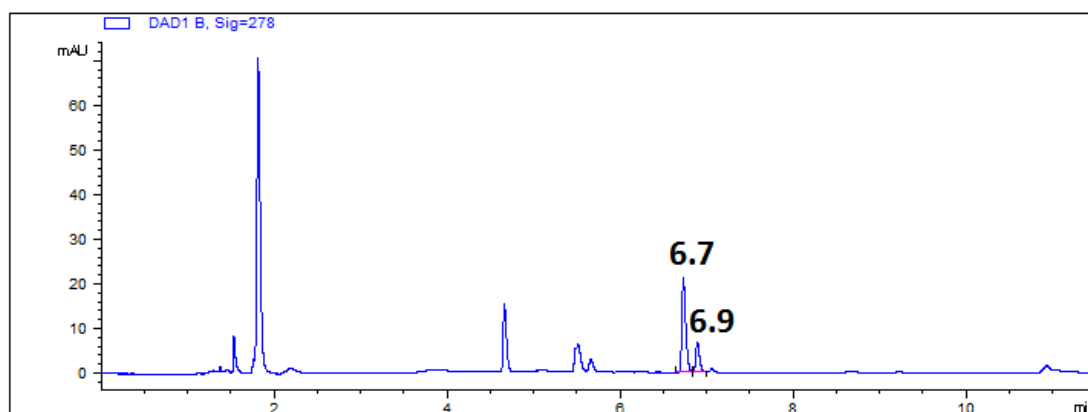
Figure 131: WT COMT (45 μM) with SAAH (**1c**) analogue (250 μM) and DHBAL (**3**) substrate (100 μM), assay was run at 37˚C for 5 hours. Sample was run by the RP HPLC method laid out in Section 6.5.3.
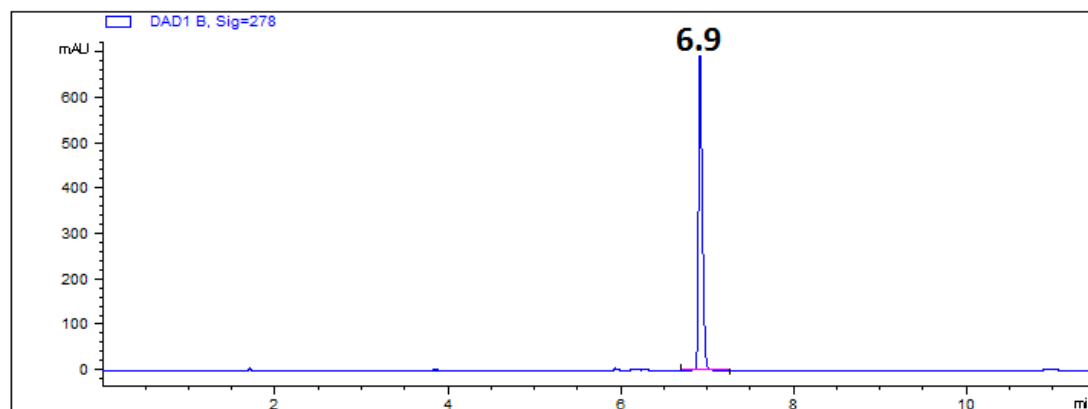
Figure 132: 1 mM 4-allyloxy-3-hydroxy-benzaldehyde (**3f**) (*para* allylated DHBAL). Sample was run by the RP HPLC method laid out in Section 6.5.3.
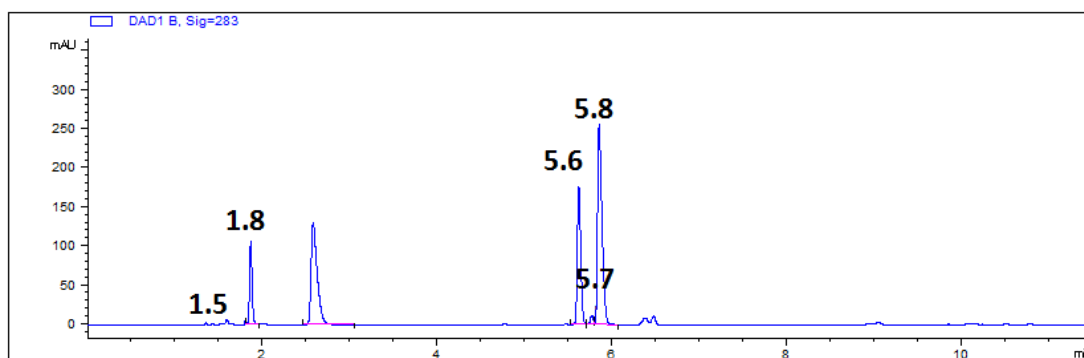
Figure 133: WT CNMT (5 µM) with SAM (**1a**) (2mM) and heliamine (**13**) (0.5 mM) 30°C for 25 minutes. Standards can be observed in Figures; Figure 134 and Figure 135. Sample was run by the RP HPLC method laid out in Section 6.5.7.
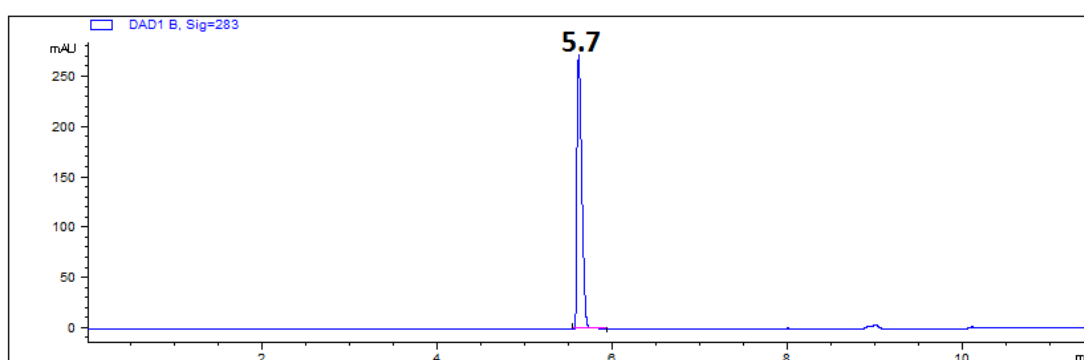
Figure 134: 1 mM heliamine (**13**) substrate standard. Sample was run by the RP HPLC method laid out in Section 6.5.7.
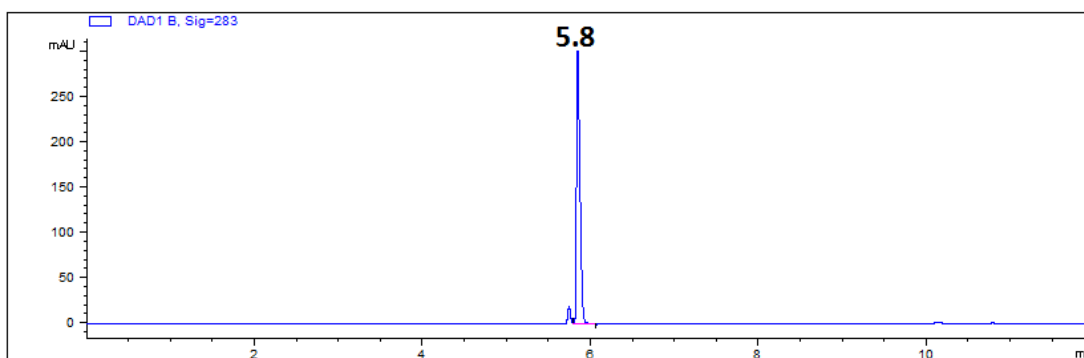
Figure 135: 1 mM 2-*N*-methyl-heliamine (**13a**) product standard. Sample was run by the RP HPLC method laid out in Section 6.5.7.
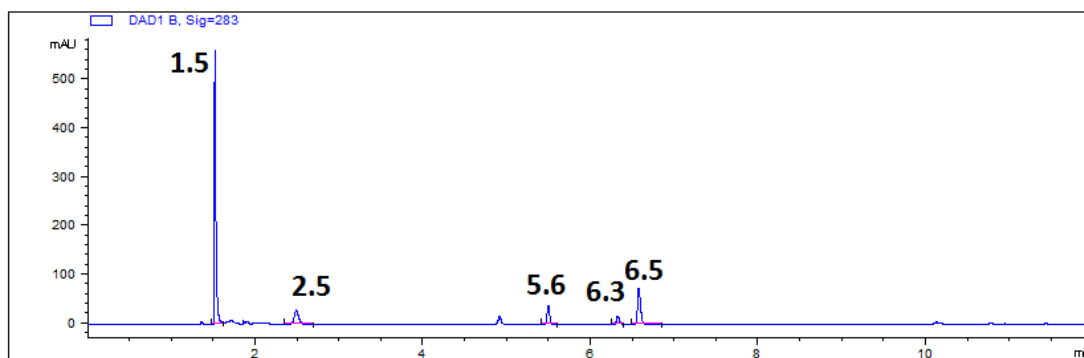
Figure 136: WT CNMT (5 µM) with SAM (**1a**) (2 mM) and 1-methyl-heliamine (**14**) (0.5 mM) 30˚C for 25 minutes. Standards can be observed in Figures; Figure 137 and Figure 138. Sample was run by the RP HPLC method laid out in Section 6.5.7.
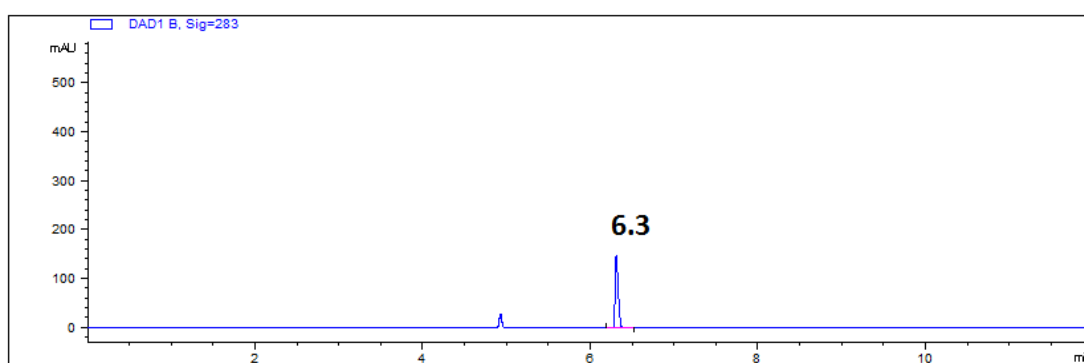
Figure 137: 0.5 mM 1-methylheliamine (**14**) substrate standard. Sample was run by the RP HPLC method laid out in Section 6.5.7.
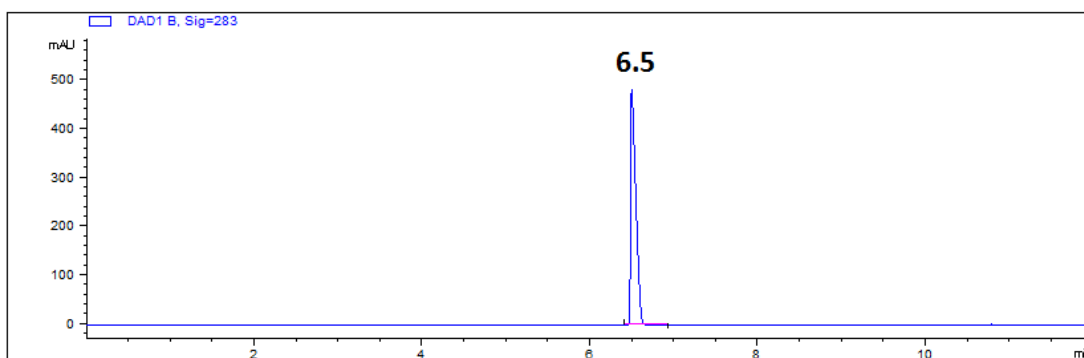
Figure 138: 1 mM 2-*N*-1-methylheliamine (**14a**) product standard. Sample was run by the RP HPLC method laid out in Section 6.5.7.
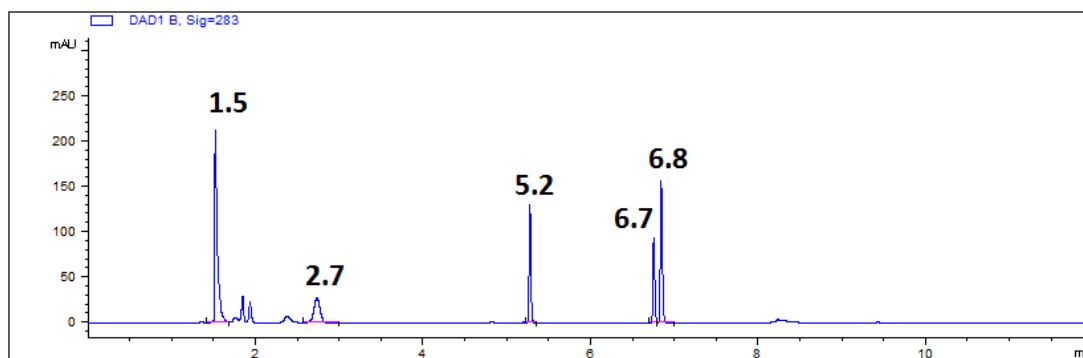
Figure 139: WT CNMT (5 μM) with SAM (**1a**) (2 mM) and 1-phenyl-heliamine (**15**) (0.5 mM) 30°C for 25 minutes. Standards can be observed in Figures; Figure 140 and Figure 141. Sample was run by the RP HPLC method laid out in Section 6.5.8.
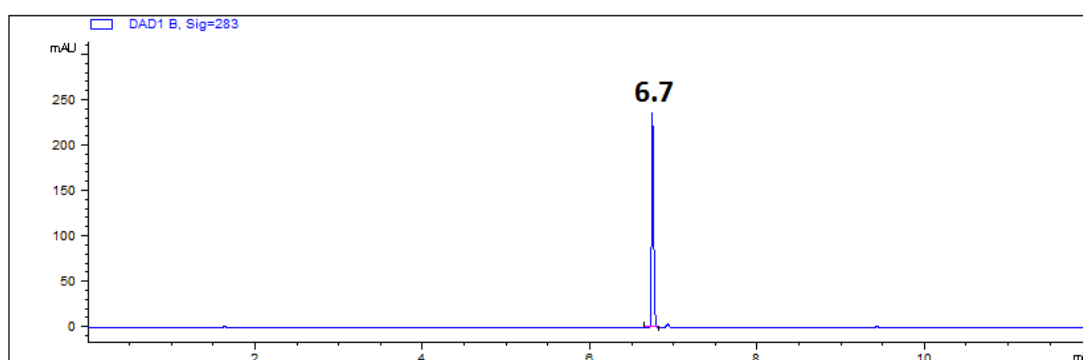
Figure 140: 0.5 mM 1-phenylheliamine (**15**) substrate standard. Sample was run by the RP HPLC method laid out in Section 6.5.8.
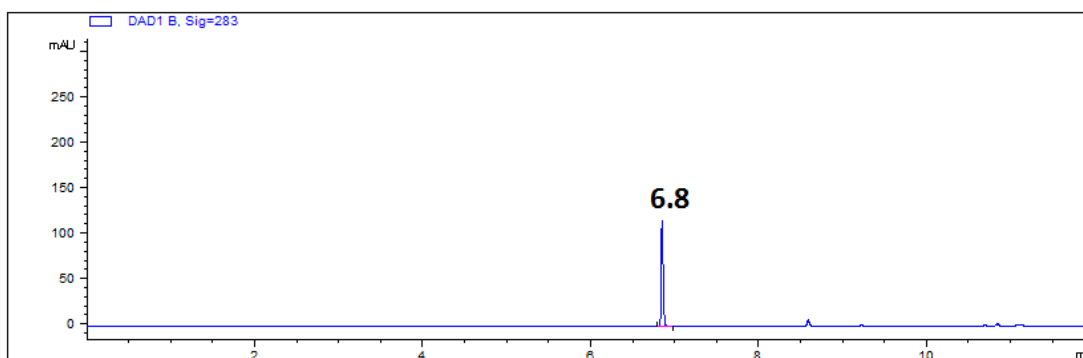
Figure 141: 0.25 mM (**15a**) 2-*N*-methyl-1-phenylheliamine product standard. Sample was run by the RP HPLC method laid out in Section 6.5.8.
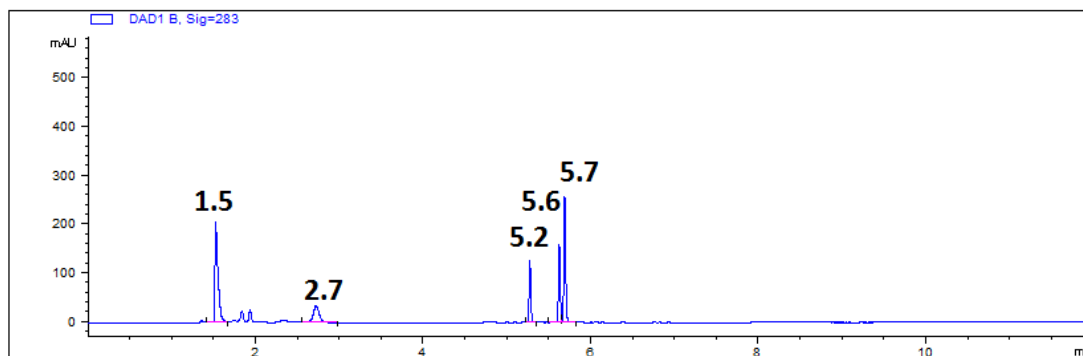
Figure 142: WT CNMT (5 µM) with SAM (**1a**) (2 mM) and higenamine (**16**) (0.5 mM) 30˚C for 25 minutes. Standard can be observed in Figure 143. Sample was run by the RP HPLC method laid out in Section 6.5.8.
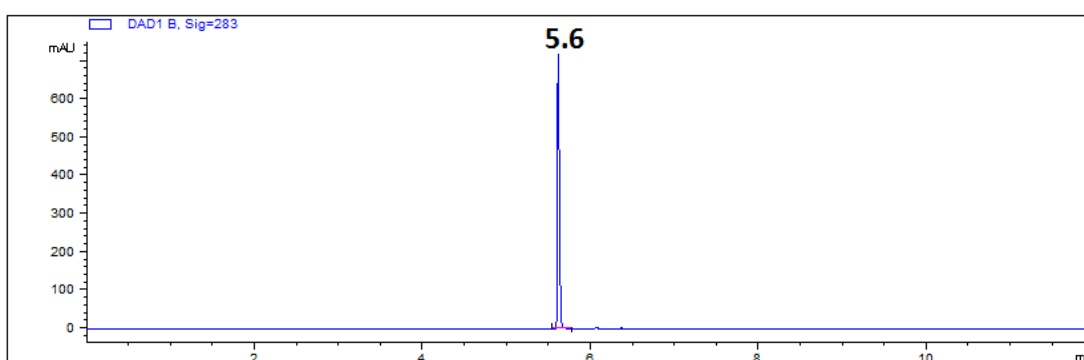
Figure 143: 1 mM higenamine (**16**) substrate standard. Sample was run by the RP HPLC method laid out in Section 6.5.8.

7.1.4 Calibration graphs

$$y = 2.3482x$$
$$R^2 = 1$$
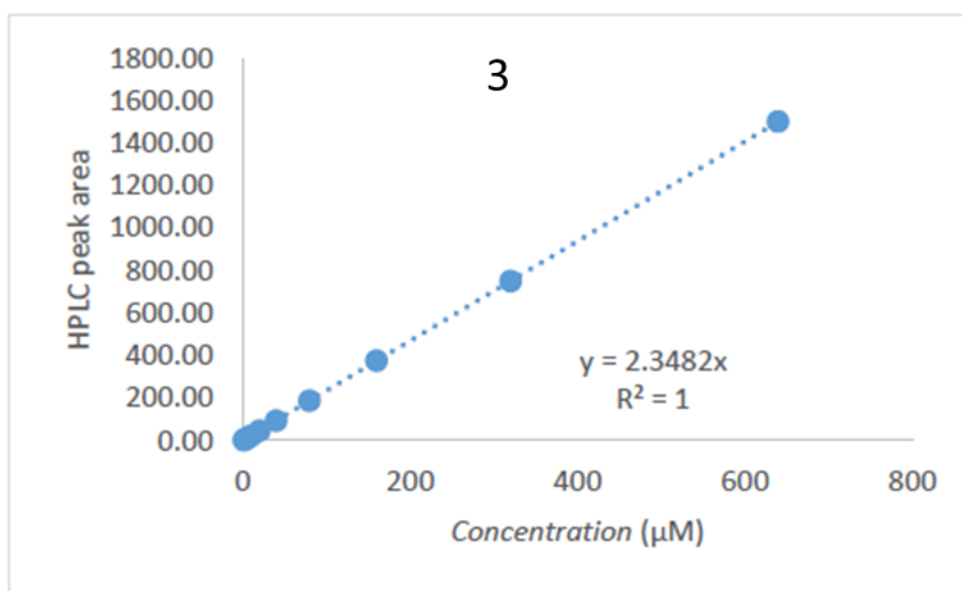
Figure 144: DHBAL (**3**) calibration for HPLC.
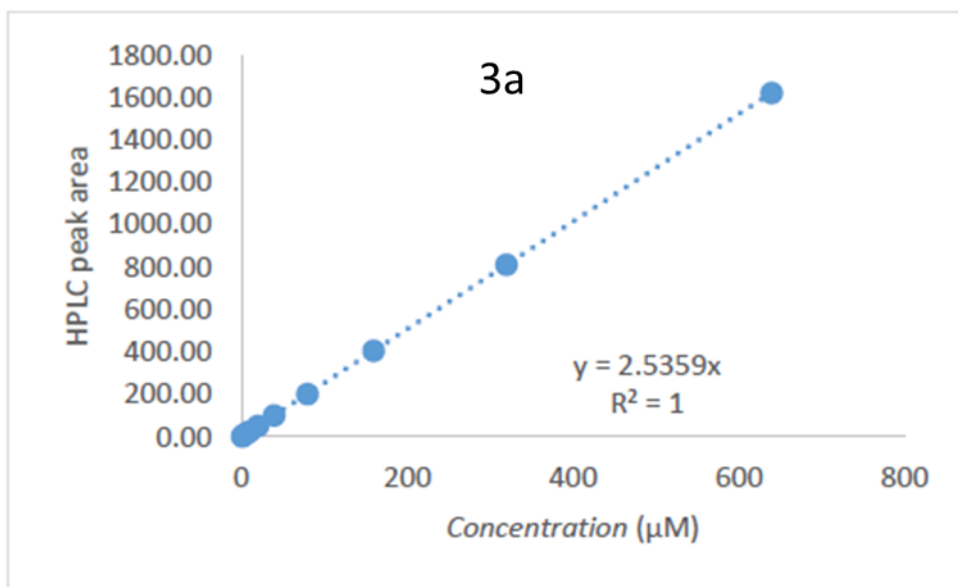
Figure 145: Vanillin (**3a**) calibration for HPLC.
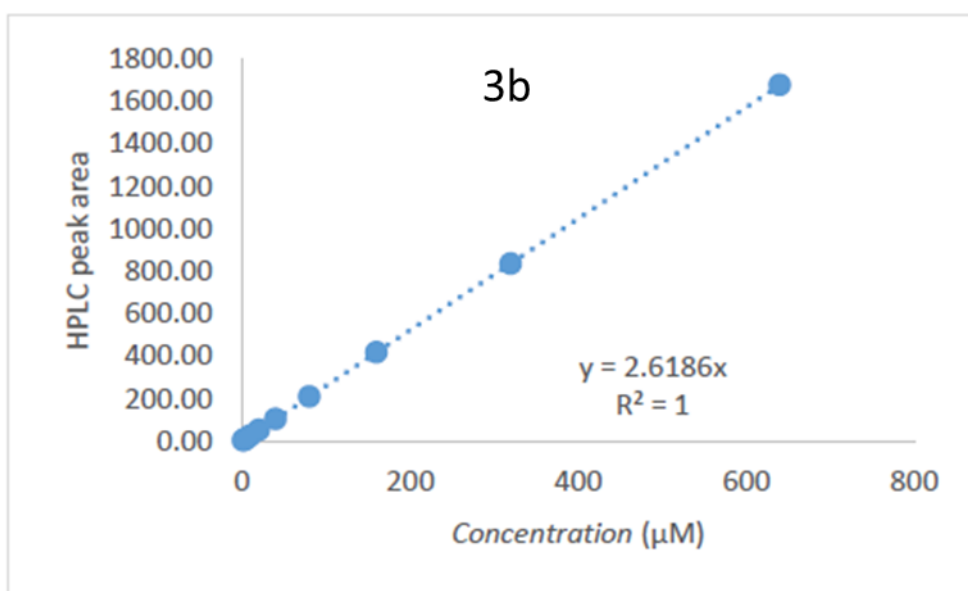
Figure 146: Isovanillin (**3b**) calibration for HPLC.
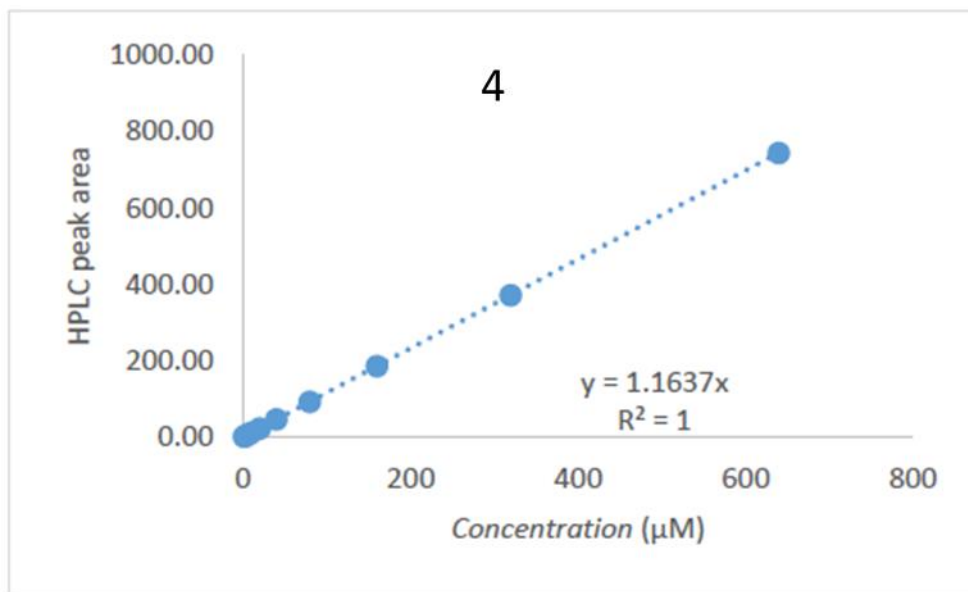
Figure 147: DHBA(**4**) calibration for HPLC.
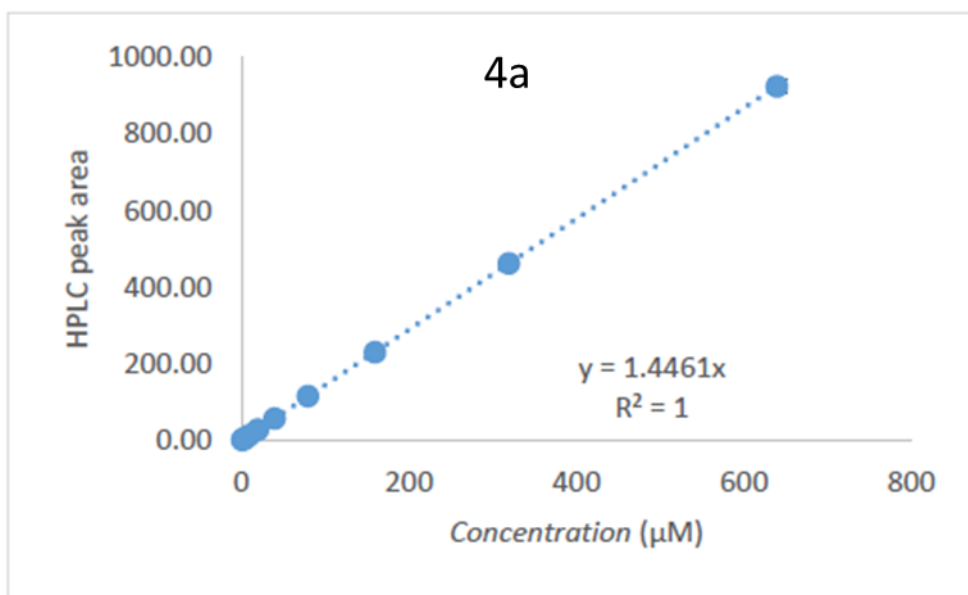
Figure 148: Vanillic acid (**4a**) calibration for HPLC.
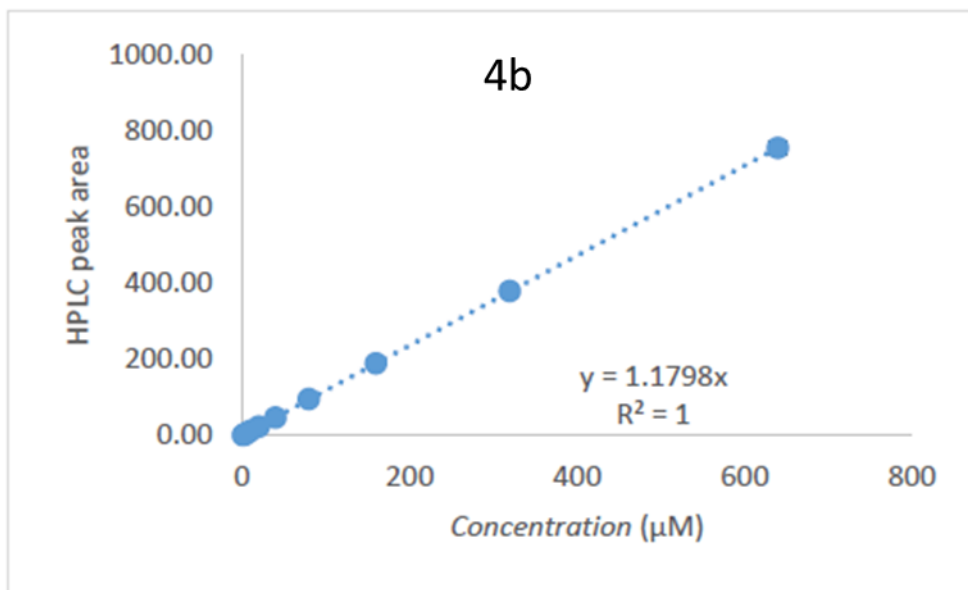
Figure 149: Isovanillic acid (**4b**) calibration for HPLC.
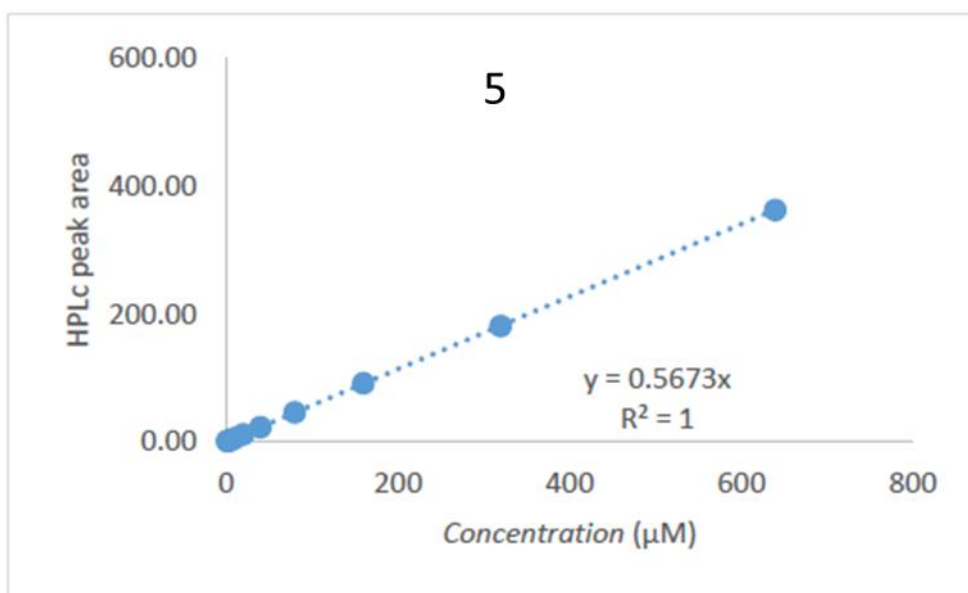
Figure 150: NO2CAT (**5**) calibration for HPLC.
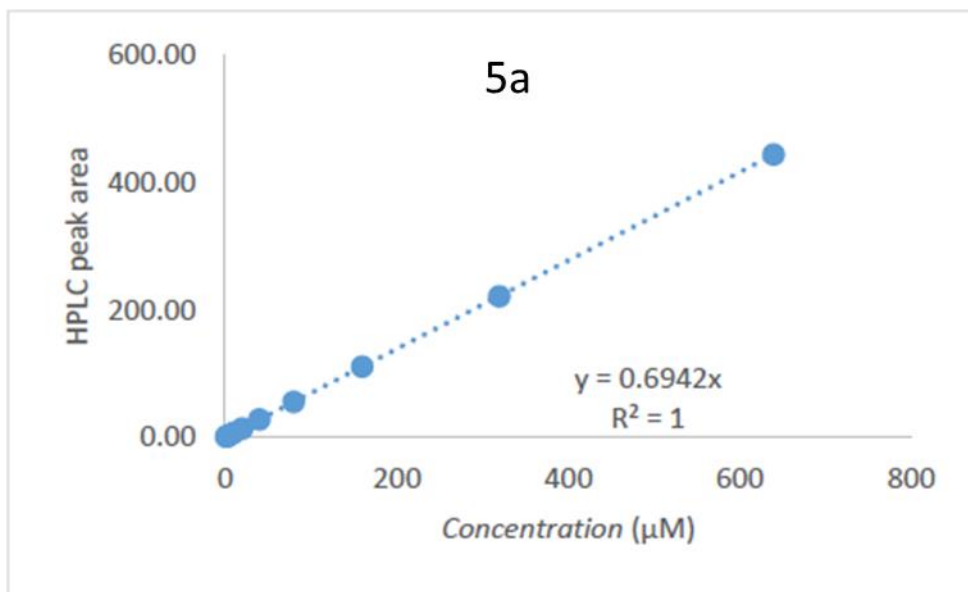
Figure 151: 4-nitroguaiacol (**5a**) calibration for HPLC.
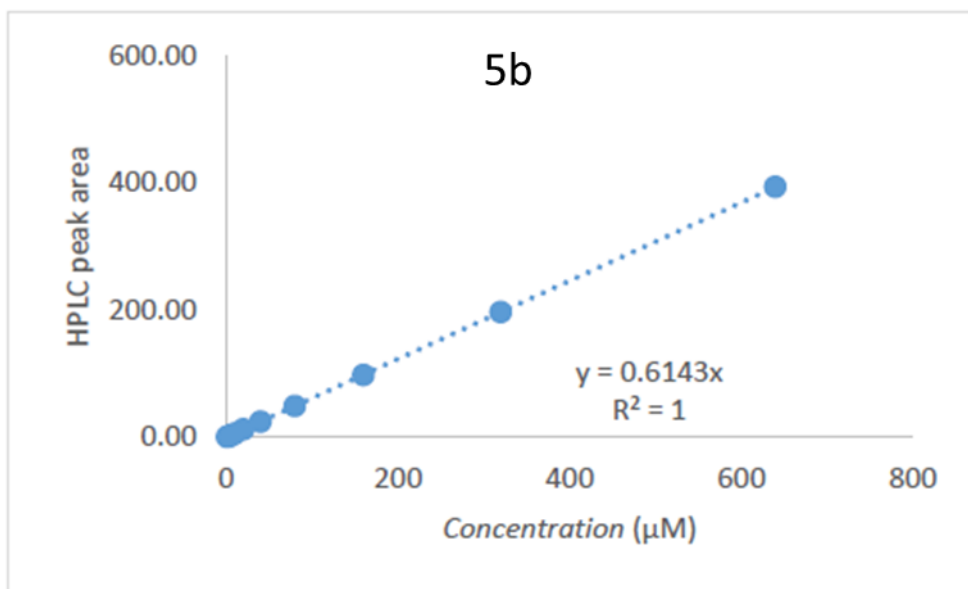
Figure 152: 2-methoxy-5-nitrophenol (**5b**) calibration for HPLC.

7.1.5   Raw data for the characterisation of SAM analogues
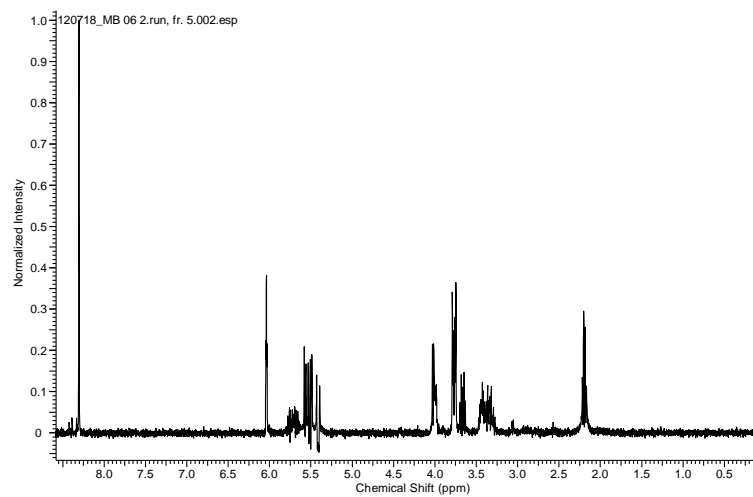
SAAH (1c)

Figure 153: [1]H spectrum of SAAH (**1c**)
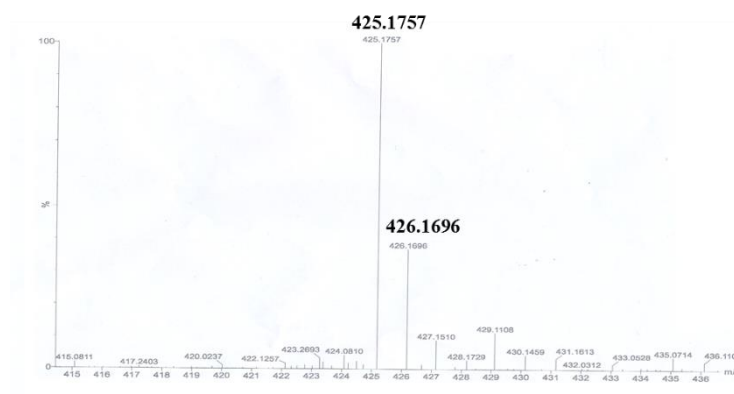
Figure 154: Mass spectrum of SAAH (**1c**).

*SABH (**1d**)*
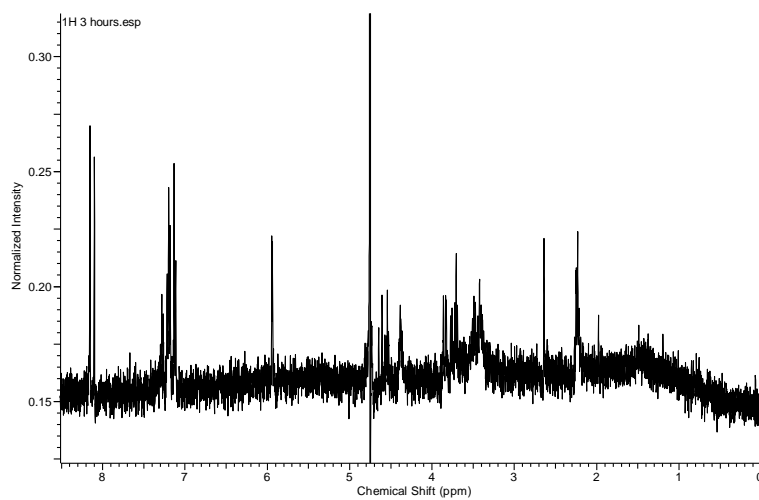
Figure 155: ^{1}H spectrum SABH (**1d**).
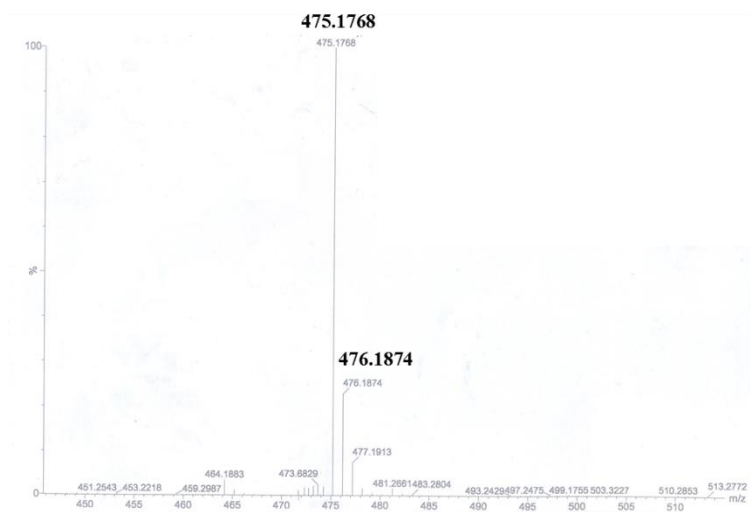
Figure 156: Mass spectrum of SABH (**1d**).

*SAMH (**1f**)*
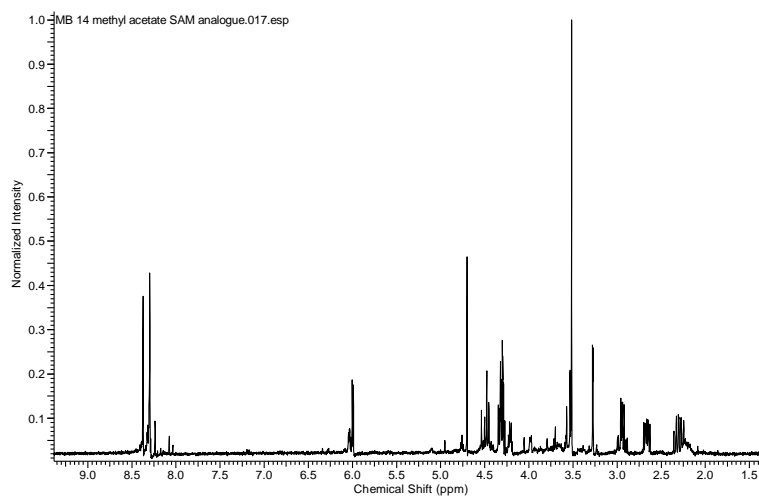
Figure 157: [^1]H spectrum of SAMH (**1f**)
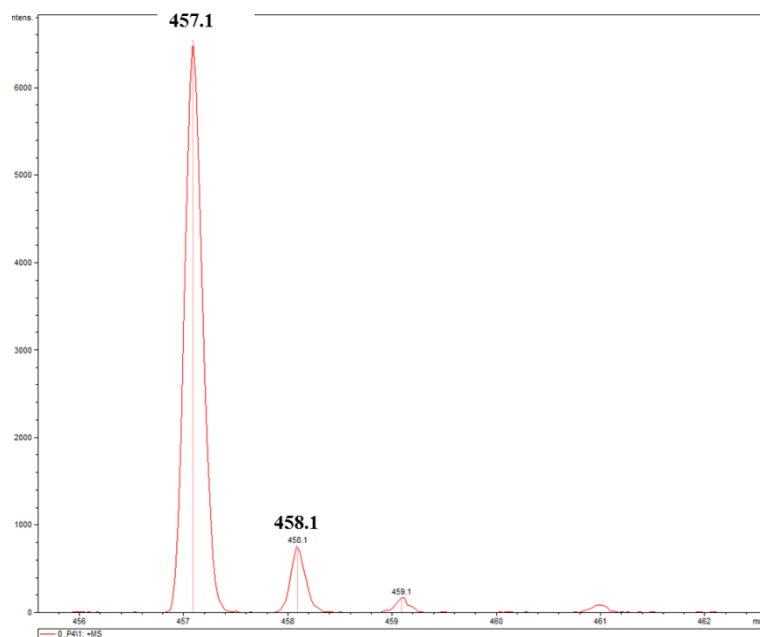
Figure 158: MALDI spectrum of SAMH (**1f**).

7.1.6 Raw data for the characterisation of Alkylated catechols

*4-allyloxy-3-hydroxybenzaldehyde (**3f**)*
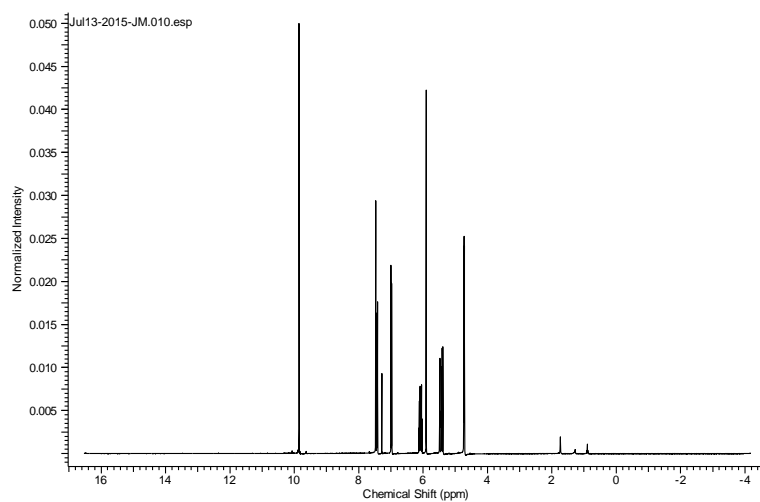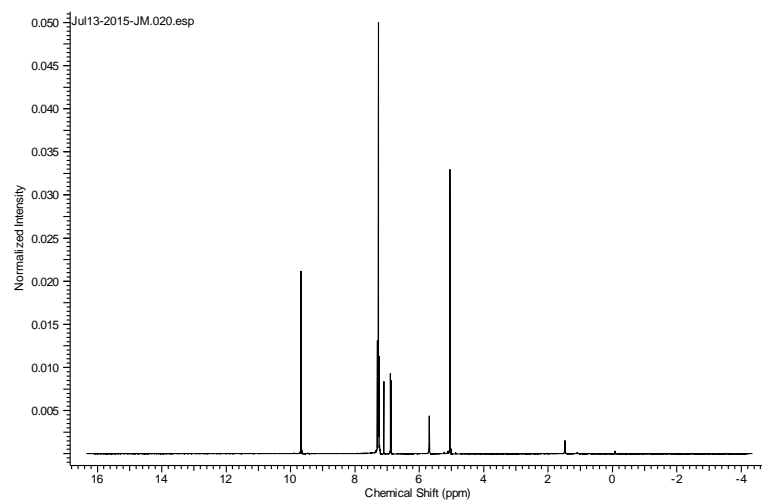

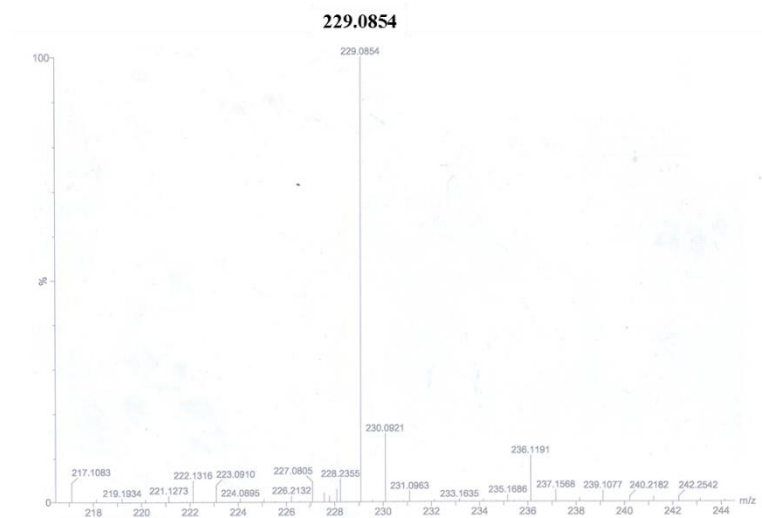**Figure 159:** [1]H spectrum of 4-allyloxy-3-hydroxybenzaldehyde (**3f**).
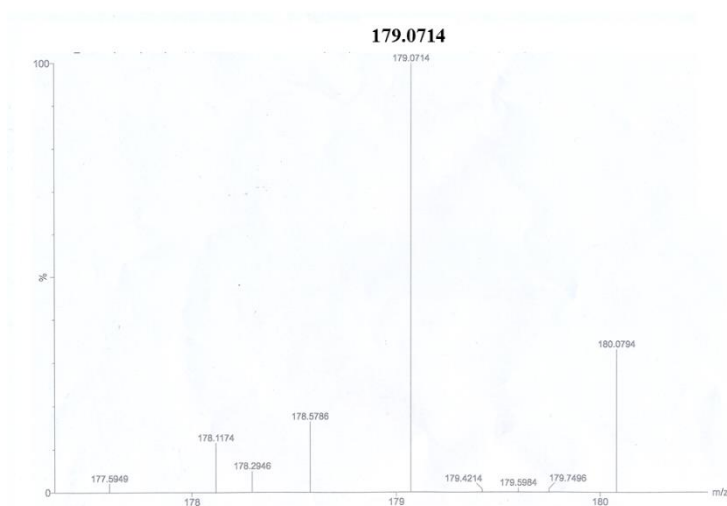
Figure 160: Accurate mass of 4-allyloxy-3-hydroxybenzaldehyde (**3f**).

3-allyloxy-4-hydroxybenzaldehyde (3e)
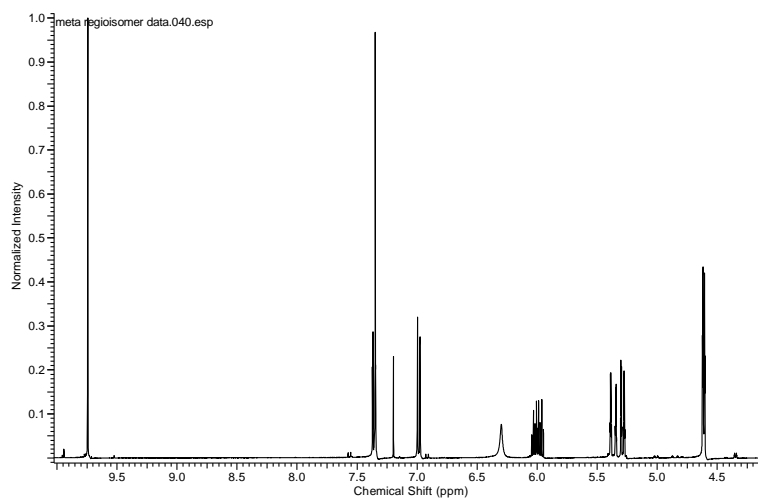
Figure 161: ^{1}H spectrum of 3-allyloxy-4-hydroxybenzaldehyde (**3e**).
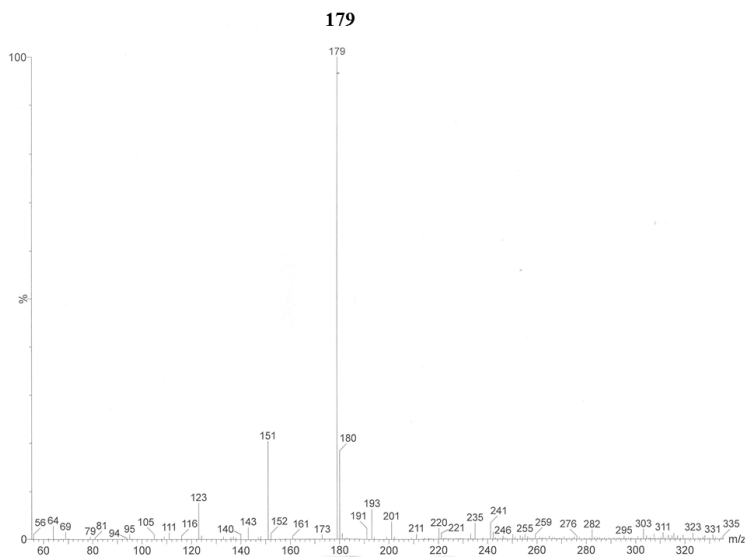
Figure 162: ES$^-$ spectrum of 3-allyloxy-4-hydroxybenzaldehyde (**3e**).

*4-benzyloxy-3-hydroxybenzaldehyde (**3h**)*

Figure 163: [1]H spectrum of 4-benzyloxy-3-hydroxybenzaldehyde (**3h**).

Figure 164: Accurate mass of 4-benzyloxy-3-hydroxybenzaldehyde (**3h**).

3-benzyloxy-4-hydroxybenzaldehyde (3g)
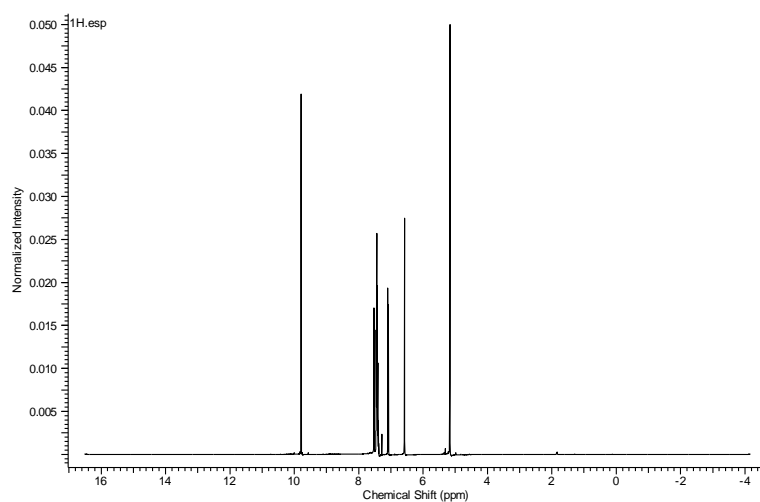
Figure 165: [1]H spectrum of 3-benzyloxy-4-hydroxybenzaldehyde (**3g**).
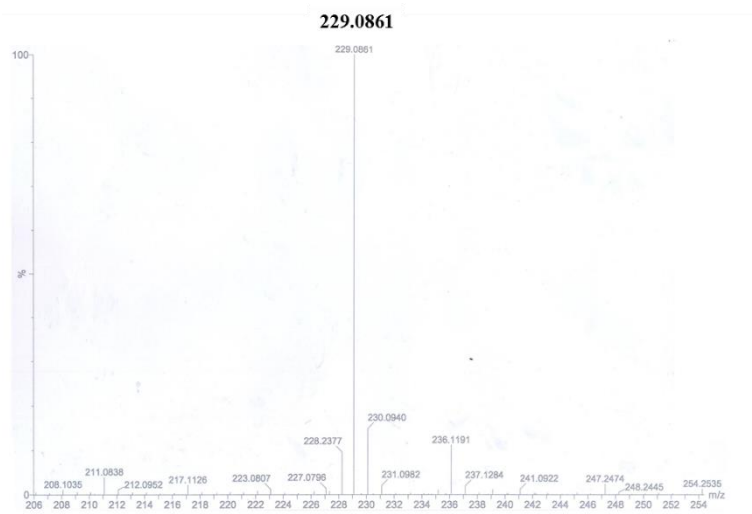
Figure 166: Accurate mass of 3-benzyloxy-4-hydroxybenzaldehyde (**3g**).
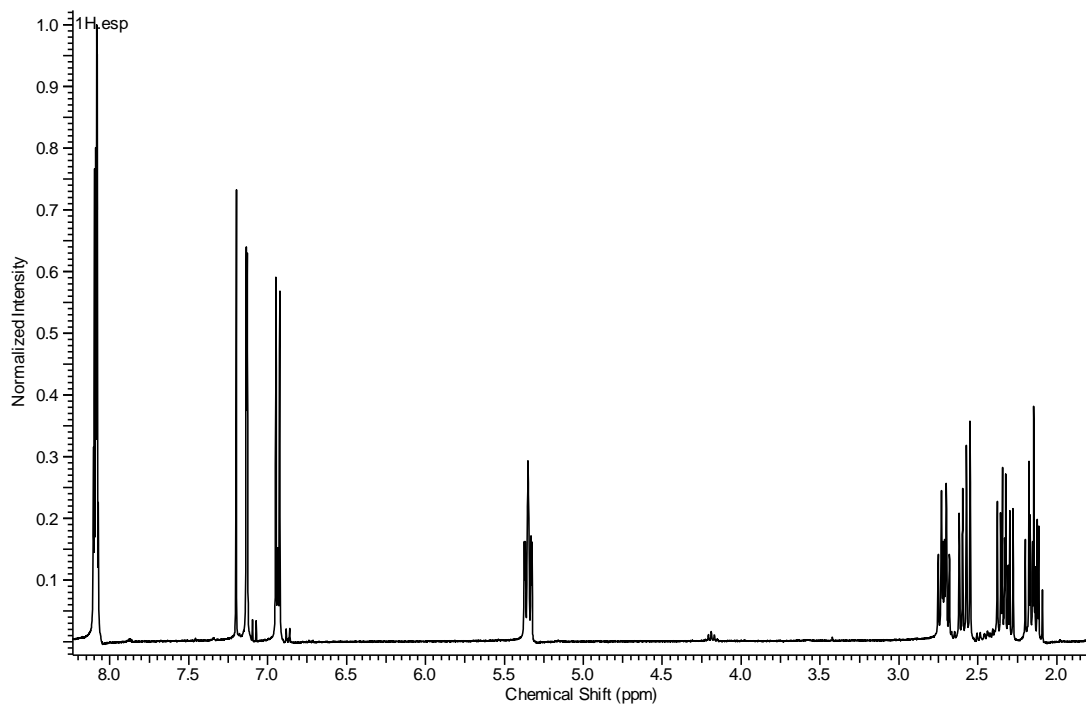
7.1.7 Raw data for the characterisation of NDCC (12)
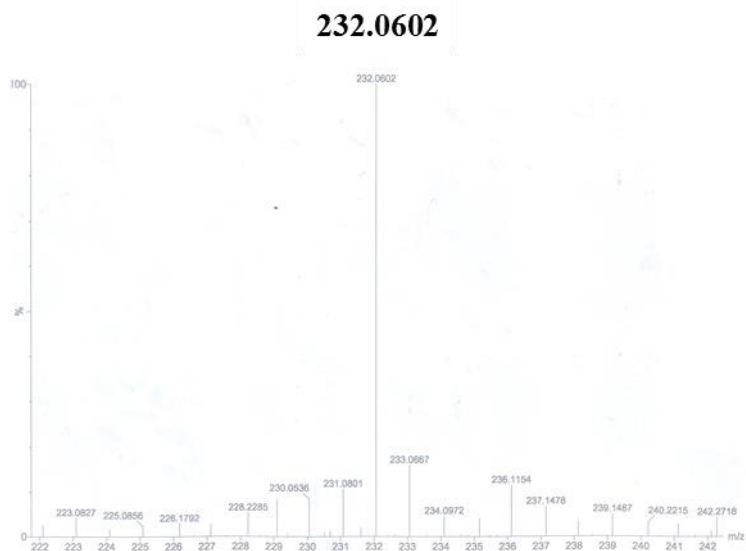

Figure 167: ^{1}H spectrum of NDCC (**12**).

232.0602

Figure 168: Accurate mass of NDCC (**12**).

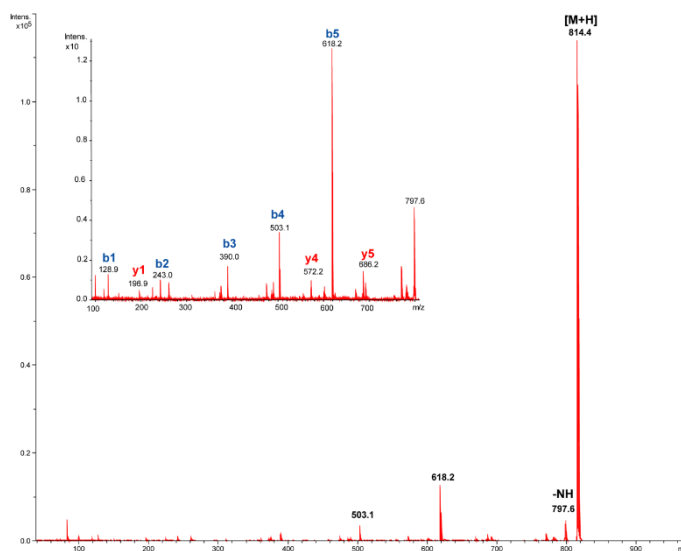## 7.1.8 Raw data for the characterisation of alkylated and hydroxylated peptides

KNFLDDOPA (6b)

Figure 169: MS-MS spectrum of KNFLDDOPA (**6b**) [M+H]$^+$ of 814.4 is clearly shown.
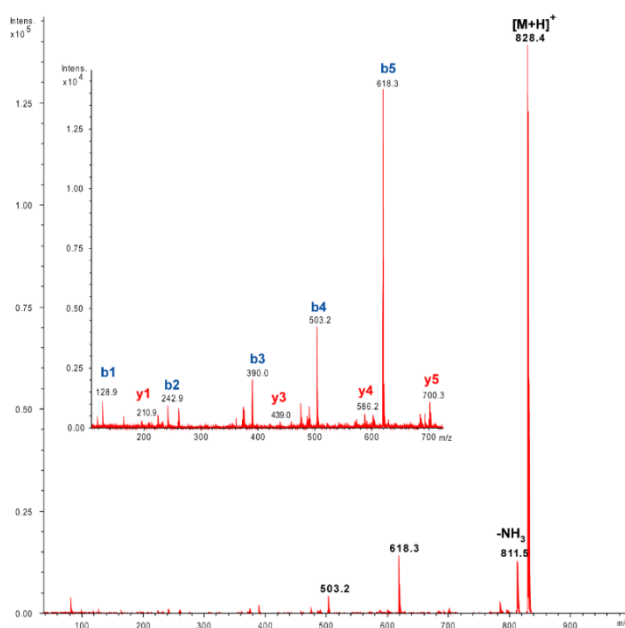
KNFLDDOPA-Me (6c)

Figure 170: MS-MS spectrum of KNFLDDOPA-Me (**6c**) [M+H]$^+$ of 828.4 is clearly shown.
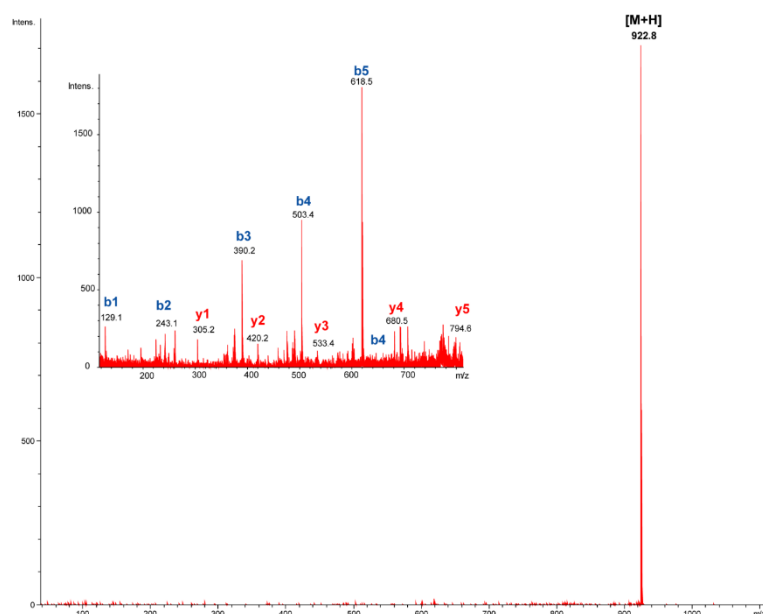
*KNFLDDOPA-POB (**6d**)*

Figure 171: MS-MS KNFLDDOPA-POB (**6d**) [M+H]$^+$ of 922.8 is clearly shown.
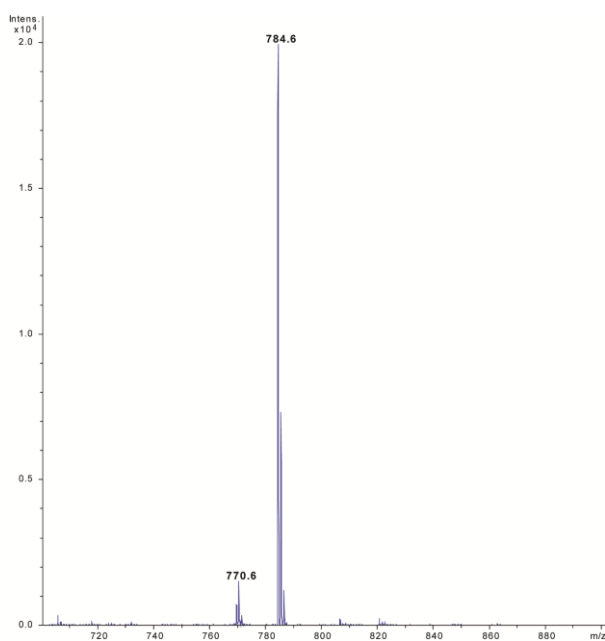
*KETDOPA-MeSK (**8c**)*

Figure 172: MALDI spectrum of methylation assay of KETDOPASK (**8b**) ([M+H]$^+$ of 770.6).
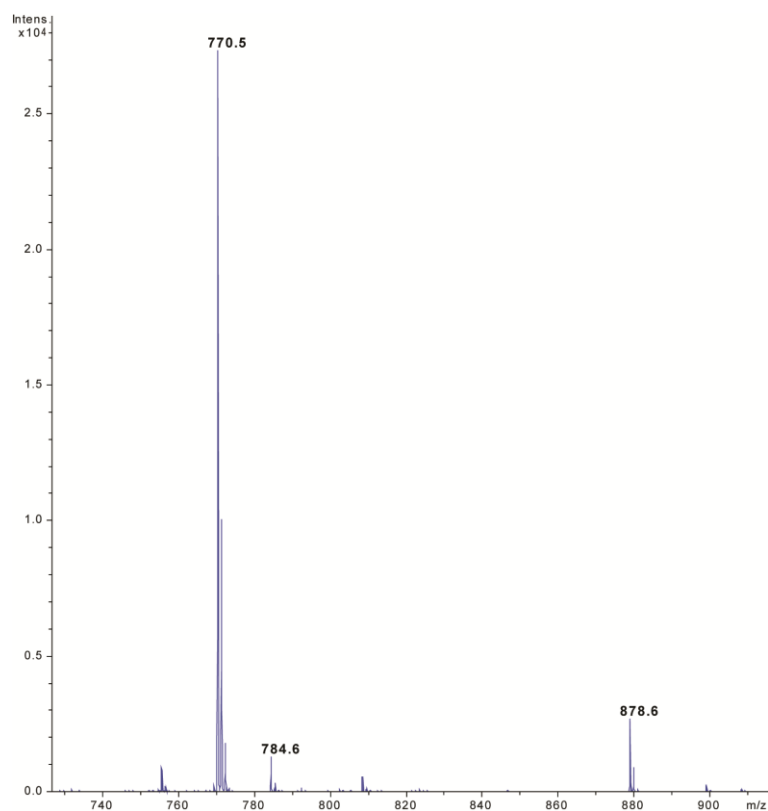KETDOPA-MeSK (**8c**) [M+H]$^+$ of 784.6 is clearly shown.

*KETDOPA-POBSK (**8d**)*

Figure 173: MALDI spectrum of POBylation assay of KETDOPASK (**8b**) ([M+H]$^+$ of 770.5). KETDOPA-POBSK (**8d**) [M+H]$^+$ of 878.6 is clearly shown.

*KNDOPALDF (**7b**)*

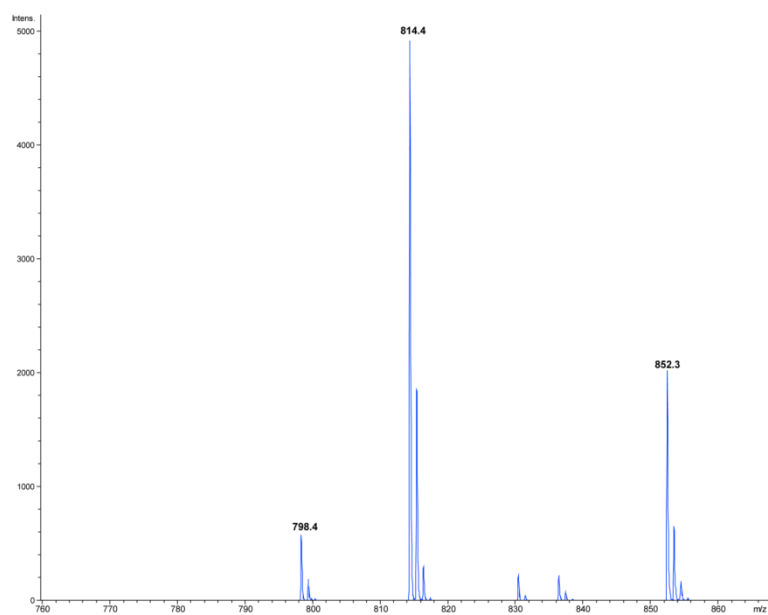**Figure 174:** MALDI spectrum of hydroxylation assay of KNYLDF (**7a**). KNDOPALDF (**7b**) [M+H]$^+$ of 814.4 and [M+K]$^+$ of 852.3 are clearly shown.

*KNDOPA-MeLDF (**7c**)*
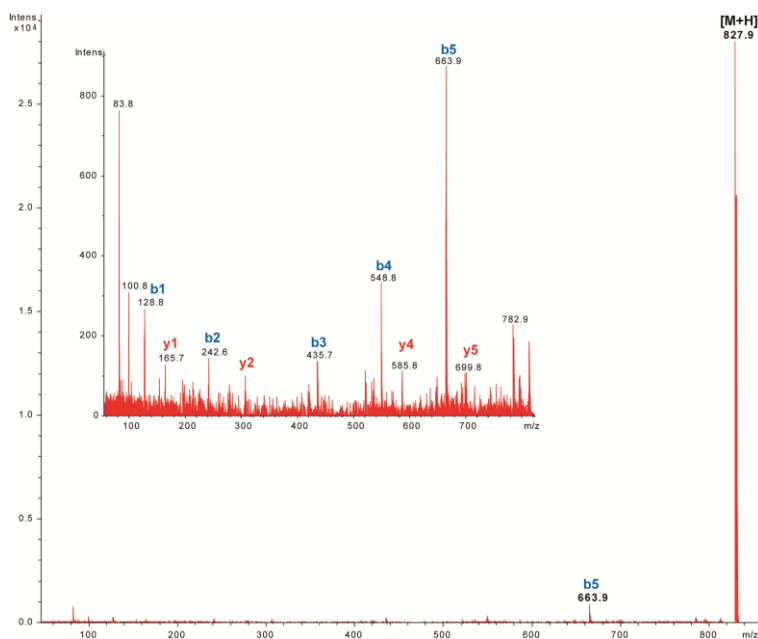
Figure 175: MS-MS spectrum of KNDOPA-MeLDF (**7c**) [M+H]$^+$ of 827.9 is clearly shown.

*Hydroxygoserelin (**9b**)*
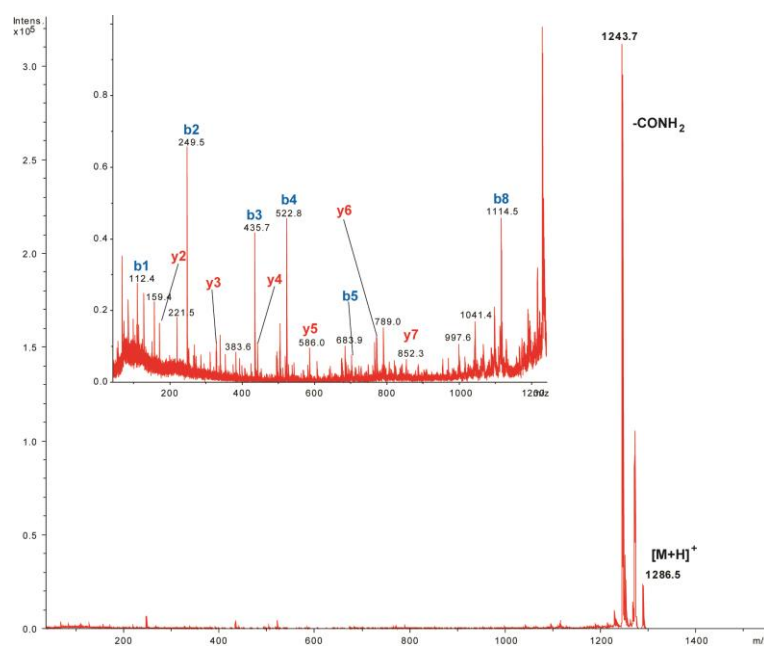
Figure 176: MS-MS spectrum of hydroxygoserelin (**9b**) [M+H]$^+$ of 1286.5 is clearly shown.

*Methoxygoserelin (**9c**)*
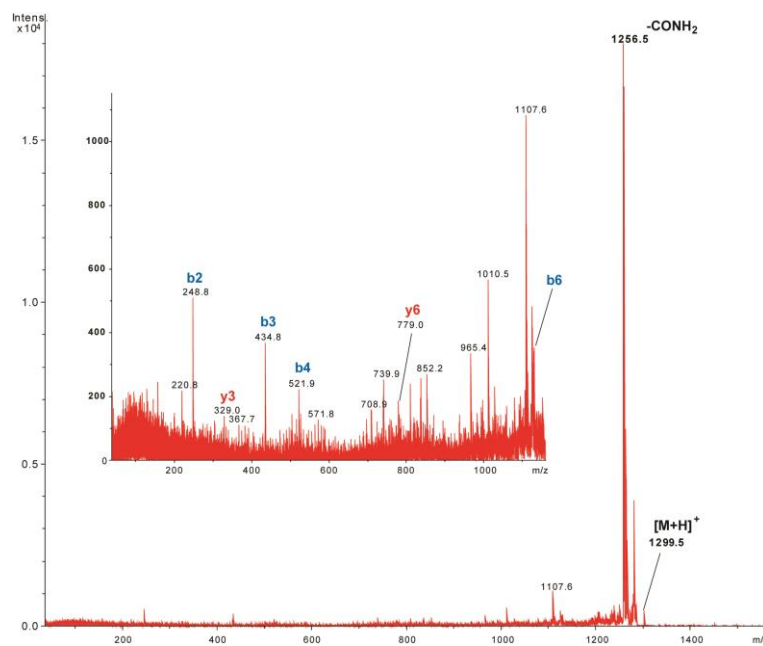
Figure 177: MS-MS spectrum of methoxygoserelin (**9c**) [M+H]$^+$ of 1299.5 is clearly shown.

*POBoxygoserelin (**9d**)*
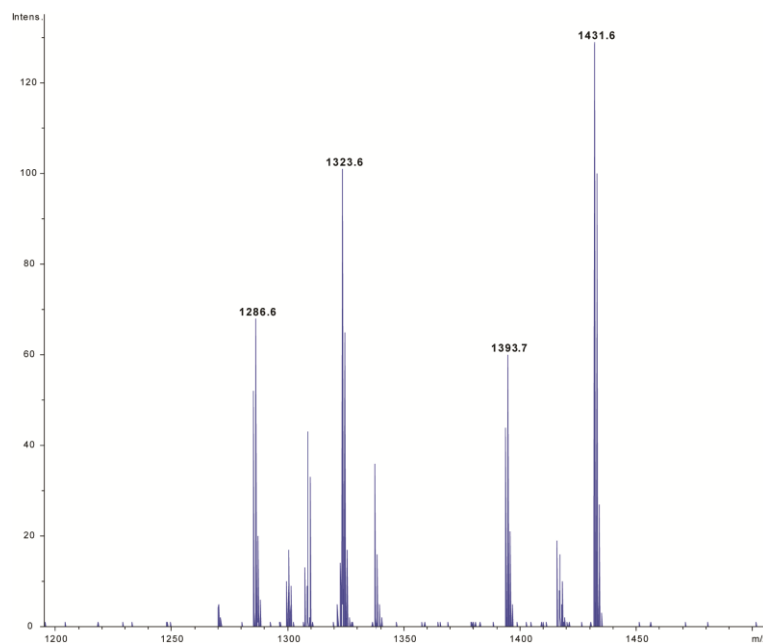
Figure 178: MALDI spectrum of POBylation assay of hydroxygoserelin (**9b**). POBoxygoserelin (**9d**) [M+H]$^+$ of 1393.7 and [M+K]$^+$ 1431.6 are clearly shown.

*Hydroxyoxytocin (**10b**)*
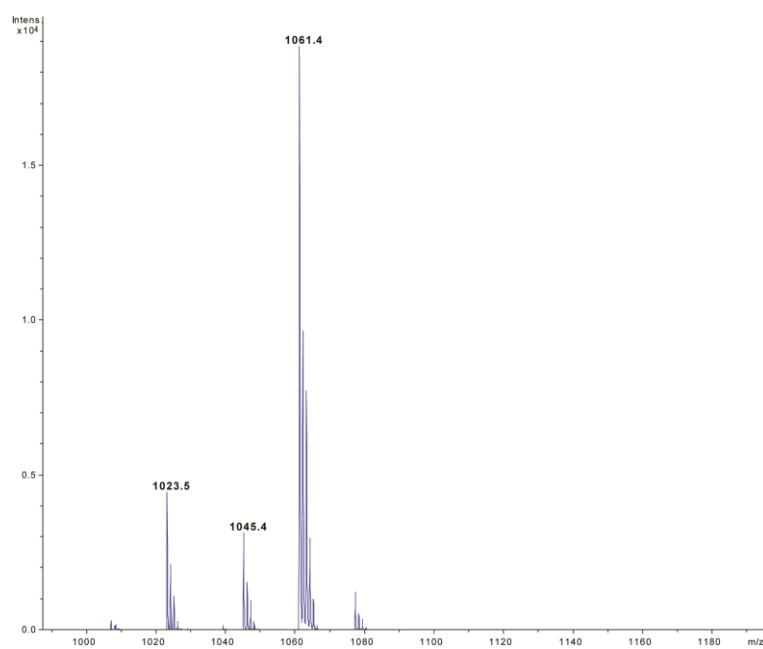
Figure 179: MALDI spectrum of hydroxylation assay of Oxytocin (**10a**). Hydroxyoxytocin (**10b**) [M+H]$^+$ of 1023.5, [M+Na]$^+$ of 1045.4 and [M+K]$^+$ 1061.4 are clearly shown.

*Methoxyoxytocin (**10c**)*
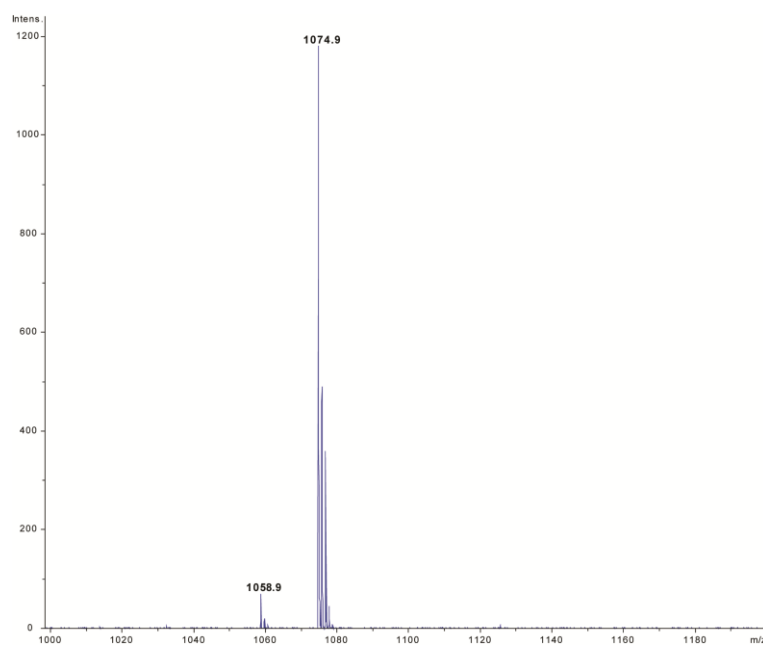
Figure 180: MALDI spectrum of methoxylation assay of oxytocin (**10a**). Methoxytocin (**10c**) [M+Na]$^+$ of 1058.9 and [M+K]$^+$ 1074.9 peaks are clearly shown.

*Hydroxytyrocidine (**11b**)*
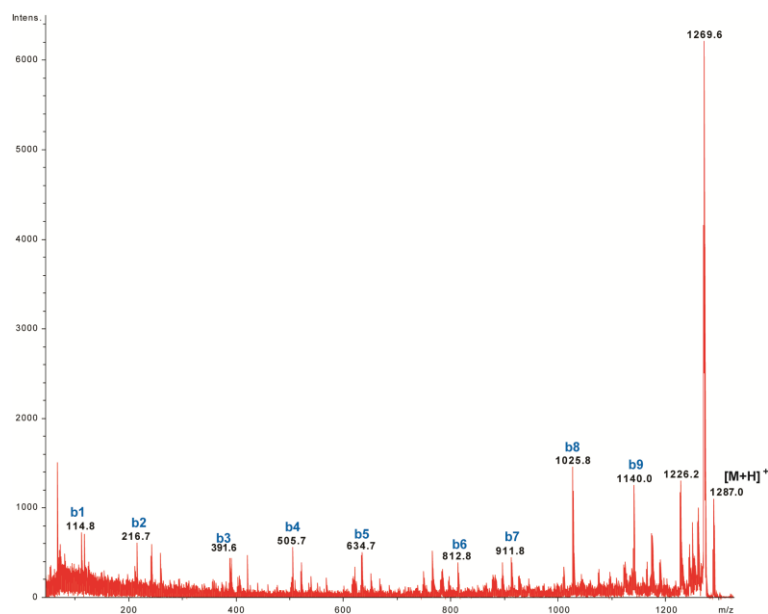
Figure 181: MS-MS spectrum of hydroxytyrocidine (**11b**) [M+H]$^+$ of 1287.0 is clearly shown.

*Methoxytyrocidine (**11c**)*
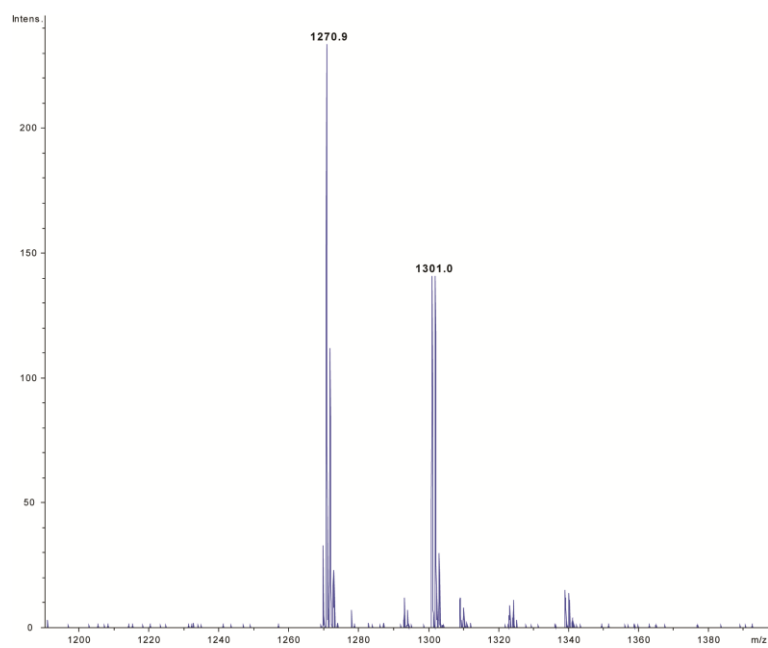
Figure 182: MALDI spectrum of methoxylation assay of methoxytyrocidine (**11c**) [M+H]$^+$ of 1301.0.

7.1.9 Raw data for the characterisation of 2-N-methyl-1-phenylheliamine (15a)
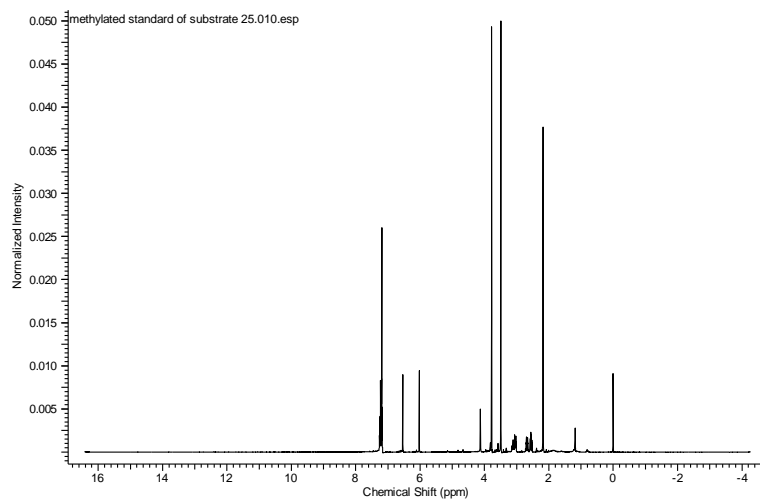
Figure 183: ^{1}H spectrum of synthetic standard of 2-*N*-methyl-1-phenylheliamine (**15a**).
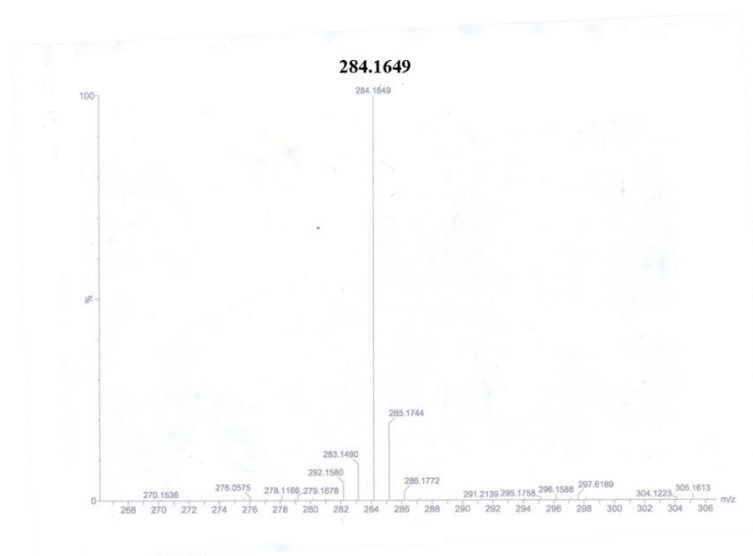
Figure 184: Accurate mass of 2-*N*-methyl-1-phenylheliamine.

7.1.10 LCMS traces of methylated CNMT and TNMT substrates

*2-**N**-methylheliamine (**13a**)*
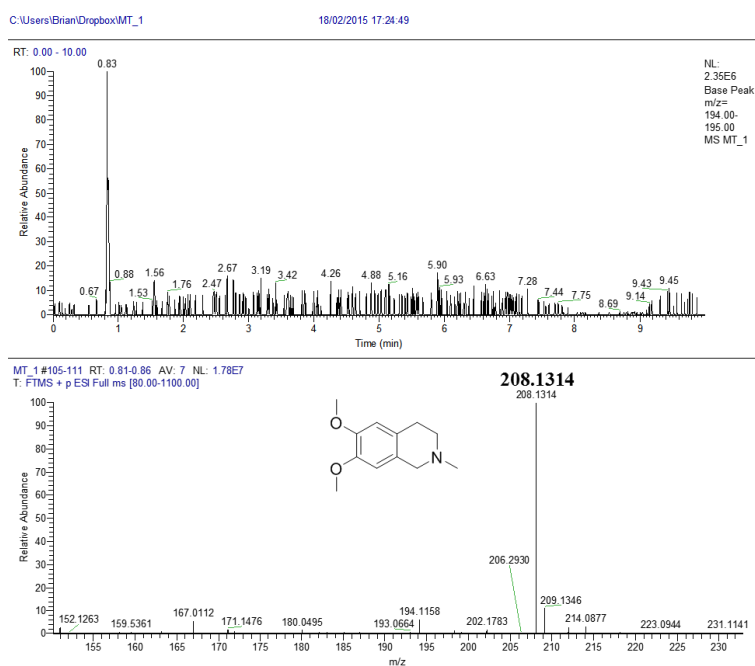
Figure 185: (Top) LC chromatogram, (Bottom) Mass spectrum of peak R_t = 0.8 min consistent with 2-*N*-methylheliamine (**13a**).

*2-**N**-methyl-1-methylheliamine (**14a**)*
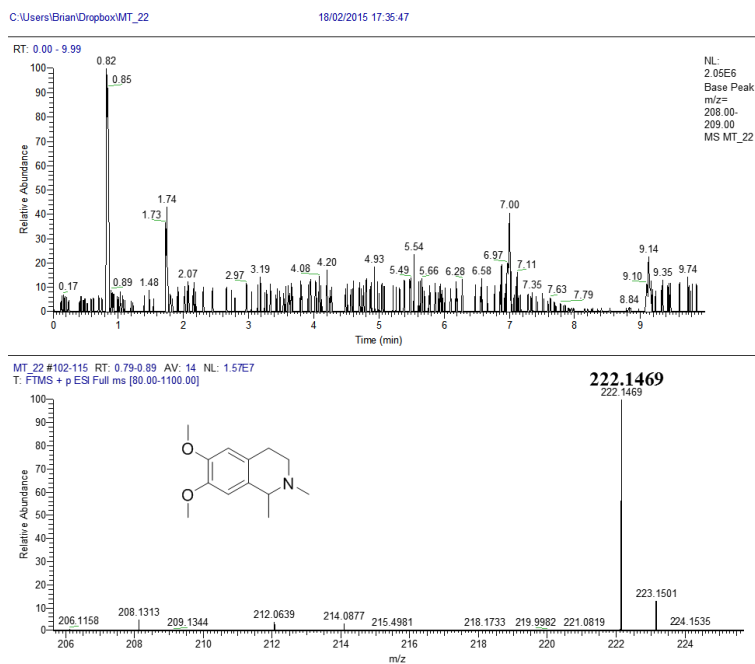
Figure 186: (Top) LC chromatogram, (Bottom) Mass spectrum of peak R_t = 0.8 min consistent with 2-*N*-methyl-1-methylheliamine (**14a**).

2-*N*-methyl-1-phenylheliamine (*15a*)
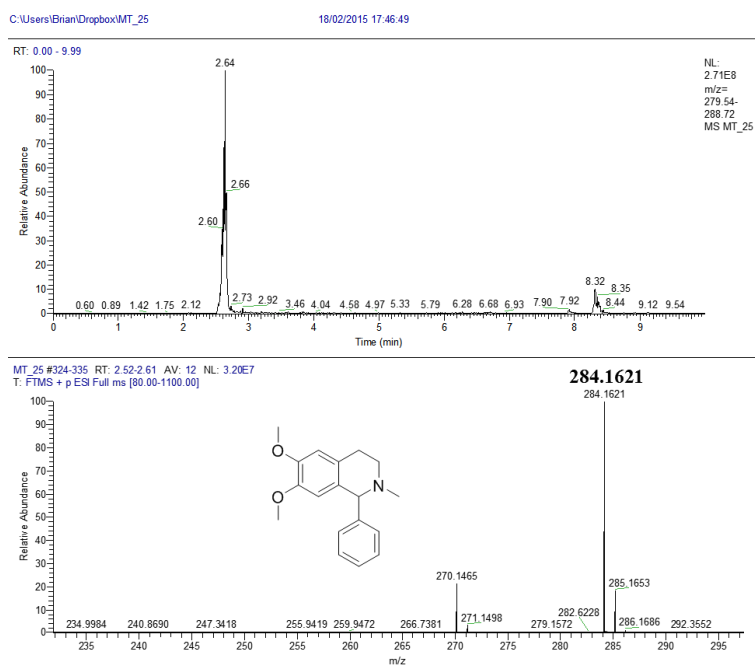
Figure 187: (Top) LC chromatogram, (Bottom) Mass spectrum of peak R_t = 2.7 min consistent with 2-*N*-methyl-1-phenylheliamine (**15a**).

2-*N*-methylhigenamine (*16a*)
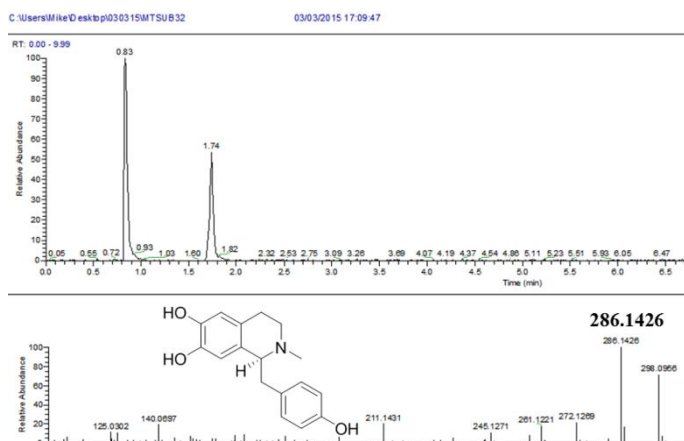
Figure 188: (Top) LC chromatogram, (Bottom) Mass spectrum of peak R_t = 0.8 min consistent with 2-*N*-methylhigenamine (**16a**).

*2-**N**-methyl-6,7-dimethoxy-1,2,3,4-tetrahydroisoquinoline-1-carbonitrile (**17a**)*
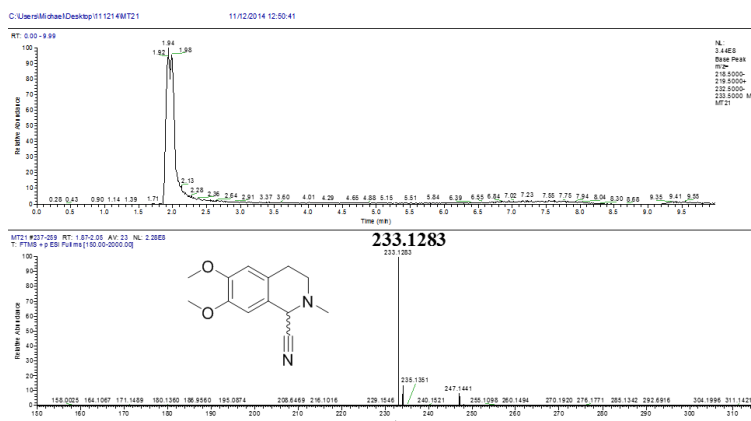
Figure 189: (Top) LC chromatogram, (Bottom) Mass spectrum of peak R_t = 1.9 min consistent with
2-*N*-methyl-6,7-dimethoxy-1,2,3,4-tetrahydroisoquinoline-1-carbonitrile (**17a**).

*2-**N**-methyl-(6,7-dimethoxy-1,2,3,4-tetrahydroisoquinolin-1-yl)methanol (**18a**)*
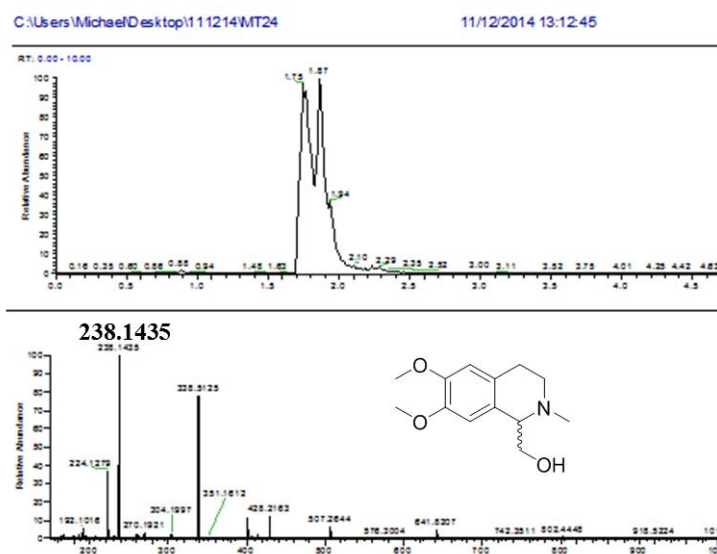
Figure 190: (Top) LC chromatogram, (Bottom) Mass spectrum of peak R_t = 1.9 min consistent with
2-*N*-methyl-(6,7-dimethoxy-1,2,3,4-tetrahydroisoquinolin-1-yl)methanol (**18a**).

*6-chloro-2-methyl-1,2,3,4-tetrahydroisoquinoline (**19a**)*
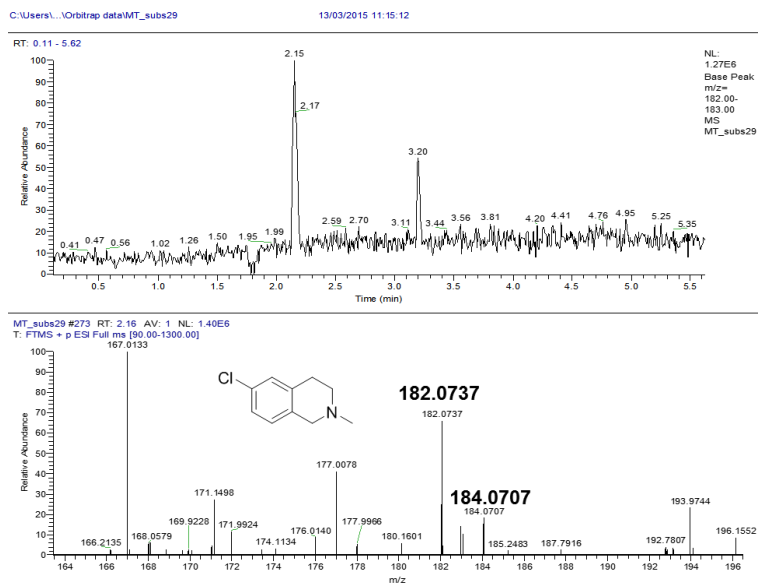
Figure 191: (Top) LC chromatogram, (Bottom) Mass spectrum of peak R_t = 2.2 min consistent with 6-chloro-2-methyl-1,2,3,4-tetrahydroisoquinoline (**19a**).

8 References

(1)   Vidgren, J.; Svensson, L. A.; Liljas, A., *Nature*, **1994**, *368*, 354

(2)   Struck, A. W.; Thompson, M. L.; Wong, L. S.; Micklefield, J. *ChemBioChem* **2012**, *13*, 2642.

(3)   Cheng, X.; Blumenthal, R. *S-adenosylmethionine-dependent methyltransferases: structures and functions*; Singapore, World Scientific, 1999.

(4)   Christian Dalhoff, G. *Nat. Chem. Biol.* **2005**, *2*, 31.

(5)   Männistö, P. T.; Kaakkola, S. *Pharmacol. Rev.***1999**, *51*, 593.

(6)   Gromova, E.; Khoroshaev, A. *Mol. Biol.* **2003**, *37*, 260.

(7)   Stecher, H.; Tengg, M.; Ueberbacher, B. J.; Remler, P.; Schwab, H.; Griengl, H.; Gruber-Khadjawi, M. *Angew. Chemie* **2009**, *121*, 9710.

(8)   Singh, S.; Zhang, J.; Huber, T. D.; Sunkara, M.; Hurley, K.; Goff, R. D.; Wang, G.; Zhang, W.; Liu, C.; Rohr, J. *Angew. Chemie Int. Ed.* **2014**, *53*, 3965.

(9)   Bentley, R. *Chem. Soc. Rev.* **2005**, *34*, 609.

(10)   Wagner, C.; Lusty, S. M.; Kung, H.-F.; Rogers, N. L. *J. Biol. Chem.* **1967**, *242*, 1287.

(11)   Cantoni, G. *J. Am. Chem. Soc.* **1952**, *74*, 2942.

(12)   Goyette, P.; Summer, J.; Milos, R. *Nat. Genet.* **1994**, *7*, 195.

(13)   Banerjee, R. V.; Matthews, R. G. *The FASEB journal* **1990**, *4*, 1450.

(14)   Nishimasu, H.; Ishitani, R.; Yamashita, K.; Iwashita, C.; Hirata, A.; Hori, H.; Nureki, O. *PNAS* **2009**, *106*, 8180.

(15)   Koutmos, M.; Datta, S.; Pattridge, K. A.; Smith, J. L.; Matthews, R. G. *PNAS* **2009**, *106*, 18527.

(16)   Mercer, A. C.; Burkart, M. D. *Nat. Chem. Biol.* **2006**, *2*, 8.

(17)   Zakharyan, R.; Wu, Y.; Bogdan, G. M.; Aposhian, H. V. *Chem. Res. Toxicol.* **1995**, *8*, 1029.

(18)   Lee, D.; Szumlanski, C.; Houtman, J.; Honchel, R.; Rojas, K.; Overhauser, J.; Wieben, E. D.; Weinshilboum, R. M. *Drug Metab. Dispos.* **1995**, *23*, 398.

(19)   Huang, Y. T.; Lyu, S. Y.; Chuang, P. H.; Hsu, N. S.; Li, Y. S.; Chan, H. C.; Huang, C. J.; Liu, Y. C.; Wu, C. J.; Yang, W. B. *ChemBioChem* **2009**, *10*, 2480.

(20)   Werner, W. J.; Allen, K. D.; Hu, K.; Helms, G. L.; Chen, B. S.; Wang, S. C. *Biochemistry* **2011**, *50*, 8986.

(21)   Choi, K.-B.; Morishige, T.; Sato, F. *Phytochemistry* **2001**, *56*, 649.

(22)   Kozbial, P. Z.; Mushegian, A. R. *BMC Struct. Biol.* **2005**, *5*, 19.

(23)   Rao, S. T.; Rossmann, M. G. *J. Mol. Biol.* **1973**, *76*, 241.

(24)   Eklund, H.; Samma, J.; Wallén, L.; Brändén, C. I.; Akeson, A.; Jones, T. A. *J. Mol. Biol.* **1981**, *146*, 561.

(25)   Axelrod, J.; Tomchick, R. *J. Biol. Chem.* **1958**, *233*, 702.

(26)   Whitby, L.; Axelrod, J.; Weil-Malherbe, H. *J. Pharmacol. Th.* **1961**, *132*, 193.

(27)   Guldberg *Pharmacol. Rev.* **1975**, *27*, 135.

(28)   Acquas, E.; Carboni, E.; Ree, R.; Prada, M.; Chiara, G. *J. Neurochem.* **1992**, *59*, 326.

(29)   Ericsson, A. D. *J. Neurol. Sciences* **1971**, *14*, 193.

(30)   Ruottinen, H.; Rinne, U. *Clin. Neuropharmacol.* **1996**, *19*, 283.

(31)   Lundstrom, K.; Salminen, M.; Jalanko, A.; Savolainen, R.; Ulmanen, I. *DNA Cell Biol.* **1991**, *10*, 181.

(32)   Ellermann, M.; Lerner, C.; Burgy, G.; Ehler, A.; Bissantz, C.; Jakob-Roetne, R.; Paulini, R.; Allemann, O.; Tissot, H.; Grunstein, D. *Acta Crystallograph. D.* **2012**, *68*, 253.

(33)   Rutherford, K.; Le Trong, I.; Stenkamp, R.; Parson, W. *J. Mol. Biol.* **2008**, *380*, 120.

(34)   Storbeck, S.; Saha, S.; Krausze, J.; Klink, B. U.; Heinz, D. W.; Layer, G. *J. Biol. Chem.* **2011**, *286*, 26754.

(35)   Chen, H. Y.; Yuan, Y. A. *J. Mol. Cell Biol.* **2010**, *2*, 366.

(36) Couture, J. F.; Hauk, G.; Thompson, M. J.; Blackburn, G. M.; Trievel, R. C. *J. Biol. Chem.* **2006**, *281*, 19280.

(37) Trievel, R. C.; Beach, B. M.; Dirk, L. M.; Houtz, R. L.; Hurley, J. H. *Cell* **2002**, *111*, 91.

(38) McGinty, R. K.; Kim, J.; Chatterjee, C.; Roeder, R. G.; Muir, T. W. *Nature* **2008**, *453*, 812.

(39) Binda, O.; Boyce, M.; Rush, J. S.; Palaniappan, K. K.; Bertozzi, C. R.; Gozani, O. *ChemBioChem* **2011**, *12*, 330.

(40) Schönherr, H.; Cernak, T. *Angew. Chemie Int. Ed.* **2013**, *52*, 12256.

(41) Rix, U.; Fischer, C.; Remsing, L. L.; Rohr, J. *Nat. Prod. Rep.* **2002**, *19*, 542.

(42) Grignard, V. *CR Acad. Sci. Paris* **1900**, *130*, 1322.

(43) Friedel, C.; Crafts, J. M. *Compt. rend.* **1877**, *84*, 1392.

(44) Clayden, J.; Warren, S.; Greeves, N.; Wothers, P., *Org. Chem.*, United Kingdom, Oxford University Press, 2001.

(45) House, H. O.; Gall, M.; Olmstead, H. D. *J. Org. Chem.* **1971**, *36*, 2361.

(46) Corey, E. J.; Hegedus, L. *J. Am. Chem. Soc.* **1969**, *91*, 4926.

(47) Groves, J. *Chem. Soc. Rev.* **1972**, *1*, 73.

(48) Mayr, H.; Striepe, W. *J. Org. Chem.* **1983**, *48*, 1159.

(49) Tamao, K.; Sumitani, K.; Kumada, M. *J. Am. Chem. Soc.* **1972**, *94*, 4374.

(50) Lou, S.; Fu, G. C. *J. Am. Chem. Soc.* **2010**, *132*, 1264.

(51) Wu, G.; Huang, M. *Chemical reviews* **2006**, *106*, 2596.

(52) Holla, E. W.; Napierski, B.; Rebenstock, H.-P. *Synlett* **1994**, 333.

(53) Plunian, B.; Mortier, J.; Vaultier, M.; Toupet, L. *J. Org. Chem.* **1996**, *61*, 5206.

(54) Terada, M.; Sorimachi, K. *J. Am. Chem. Soc.* **2007**, *129*, 292.

(55) Williamson, A. W. *J. Chem. Soc. London* **1852**, *4*, 229.

(56) Olson, W. T.; Hipsher, H. F.; Buess, C. M.; Goodman, I. A.; Hart, I.; Lamneck Jr, J. H.; Gibbons, L. C. *J. Am. Chem. Soc.* **1947**, *69*, 2451.

(57) Freedman, H.; Dubois, R. *Tetrahedron lett.***1975**, *16*, 3251.

(58) Kurti, L.; Czakó, B. *Strategic applications of named reactions in organic synthesis*; United Kingdom, Elsevier, 2005.

(59) Wang, R.; Islam, K.; Liu, Y.; Zheng, W.; Tang, H.; Lailler, N.; Blum, G.; Deng, H.; Luo, M. *J. Am. Chem. Soc.* **2012**.

(60) Caine, J. M.; Macreadie, I. G.; Grunewald, G. L.; McLeish, M. J. *Protein expres. purif.* **1996**, *8*, 160.

(61) Winans, C. F.; Adkins, H. *J. Am. Chem. Soc.* **1932**, *54*, 306.

(62) Guram, A. S.; Rennels, R. A.; Buchwald, S. L. *Angew. Chemie Int. Ed.* **1995**, *34*, 1348.

(63) Louie, J.; Hartwig, J. F. *Tetrahedron lett.* **1995**, *36*, 3609.

(64) Farkas, E.; Sunman, C. J. *The J. Org. Chem.* **1985**, *50*, 1110.

(65) Wetzel, A.; Wöckel, S.; Schelwies, M.; Brinks, M. K.; Rominger, F.; Hofmann, P.; Limbach, M. *Org. lett.***2013**, *15*, 266.

(66) Sobol, Z.; Engel, M.; Rubitski, E.; Ku, W.; Aubrecht, J.; Schiestl, R. *Mutat. Res.-Gen. Tox. En.* **2007**, *633*, 80.

(67) Alder, R. W.; Phillips, J. G. E.; Huang, L.; Huang, X. In *Encyclopedia of Reagents for Organic Synthesis*; John Wiley & Sons, Ltd: 2001.

(68) Parks, L. *J. Biol. Chem.* **1958**, *232*, 169.

(69) Schlenk, F.; Dainko, J. *BBA-Gen. Subjects* **1975**, *385*, 312.

(70) Lee, B. W.; Sun, H. G.; Zang, T.; Kim, B. J.; Alfaro, J. F.; Zhou, Z. S. *J. Am. Chem. Soc.* **2010**, *132*, 3642.

(71) Wang, R.; Zheng, W.; Yu, H.; Deng, H.; Luo, M. *J. Am. Chem. Soc.* **2011**.

(72) Borchardt, R.; Wu, Y. S.; Huber, J.; Wycpalek, A. *J. Med. Chem.* **1976**, *19*, 1104.

(73) Varela-Rey, M.; Martínez-López, N.; Fernández-Ramos, D.; Embade, N.; Calvisi, D. F.; Woodhoo, A.; Rodríguez, J.; Fraga, M. F.; Julve, J.; Rodríguez-Millán, E. *Hepatology* **2010**, *52*, 105.

(74)   Zorbach, W. W.; Tipson, R. S. *Synthetic procedures in nucleic acid chemistry*; Interscience Publishers; New York, Interscience Publishers, 1968.

(75)   Bothwell, I. R.; Islam, K.; Chen, Y.; Zheng, W.; Blum, G.; Deng, H.; Luo, M. *J. Am. Chem. Soc.* **2012**, *134*, 14905.

(76)   Peters, W.; Willnow, S.; Duisken, M.; Kleine, H.; Macherey, T.; Duncan, K. E.; Litchfield, D. W.; Lüscher, B.; Weinhold, E. *Angew. Chemie Int. Ed.* **2010**, *49*, 5170.

(77)   Mudd, S. H.; Cantoni, G., *Nature*, **1957,** *180*, 1052

(78)   Iwig, D. F.; Booker, S. J. *Biochemistry* **2004**, *43*, 13496.

(79)   Hoffman, J. L. *Biochemistry* **1986**, *25*, 4444.

(80)   Wu, S. E.; Huskey, W. P.; Borchardt, R. T.; Schowen, R. L. *Biochemistry* **1983**, *22*, 2828.

(81)   Thomsen, M.; Vogensen, S. B.; Buchardt, J.; Burkart, M. D.; Clausen, R. P. *Org. Biomol. Chem.* **2013**, *11*, 7606.

(82)   Eustáquio, A. S.; Pojer, F.; Noel, J. P.; Moore, B. S. *Nat. Chem. Biol.* **2008**, *4*, 69.

(83)   O'Hagan, D.; Schaffrath, C.; Cobb, S. L.; Hamilton, J. T.; Murphy, C. D. *Nature* **2002**, *416*, 279.

(84)   Deng, H.; Cobb, S. L.; McEwan, A. R.; McGlinchey, R. P.; Naismith, J. H.; O'Hagan, D.; Robinson, D. A.; Spencer, J. B. *Angew. Chemie Int. Ed.* **2006**, *45*, 759.

(85)   Horikawa, S.; Sasuga, J.; Shimizu, K.; Ozasa, H.; Tsukada, K. *J. Biol. Chem.* **1990**, *265*, 13683.

(86)   Komoto, J.; Yamada, T.; Takata, Y.; Markham, G. D.; Takusagawa, F. *Biochemistry* **2004**, *43*, 1821.

(87)   Law, B.; Winona-Struck, A.; Bennett, M. R.; Wilkinson, B.; Micklefield, J. *Chem. Sci.* **2015**, *6*, 2885.

(88)   Klimašauskas, S.; Weinhold, E. *Trends Biotechnol.* **2007**, *25*, 99.

(89)   Klose, J.; Wendt, N.; Kubald, S.; Krause, E.; Fechner, K.; Beyermann, M.; Bienert, M.; Rudolph, R.; Rothemund, S. *Protein Sci.* **2004**, *13*, 2470.

(90)   Lukinavičius, G.; Lapiene, V.; Staševskij, Z.; Dalhoff, C.; Weinhold, E.; Klimašauskas, S. *J. Am. Chem. Soc.* **2007**, *129*, 2758.

(91)   Gregory, M. A.; Hong, H.; Lill, R. E.; Gaisser, S.; Petkovic, H.; Low, L.; Sheehan, L. S.; Carletti, I.; Ready, S. J.; Ward, M. J. *Org. Biomol. Chem.* **2006**, *4*, 3565.

(92)   Saunders, R. N.; Metcalfe, M. S.; Nicholson, M. L. *Kidney Int.* **2001**, *59*, 3.

(93)   Chen, Y.-W.; Smith, M. L.; Sheets, M.; Ballaron, S.; Trevillyan, J. M.; Burke, S. E.; Rosenberg, T.; Henry, C.; Wagner, R.; Bauch, J. *J. Cardiovasc. Pharm.* **2007**, *49*, 228.

(94)   Nicolaou, K.; Chakraborty, T.; Piscopio, A.; Minowa, N.; Bertinato, P. *J. Am. Chem. Soc.* **1993**, *115*, 4419.

(95)   Rodrigues, R. C.; Fernandez-Lafuente, R. *J. Mol. Catal. B: Enzym.* **2010**, *64*, 1.

(96)   Inprakhon, P.; Lalot, T. *J. Biotechnol.* **2007**, *131*, 418.

(97)   Mahlert, C.; Kopp, F.; Thirlway, J.; Micklefield, J.; Marahiel, M. A. *J. Am. Chem. Soc.* **2007**, *129*, 12011.

(98)   Creveling, C.; Morris, N.; Shimizu, H.; Ong, H.; Daly, J. *Mol. Pharmacol.* **1972**, *8*, 398.

(99)   Tsao, D.; Liu, S.; Dokholyan, N. V. *Chem. Phys. Lett.* **2011**.

(100)   Inoue, K.; Sewalt, V. J.; Ballance, G. M.; Ni, W.; Stürzer, C.; Dixon, R. A. *Plant physiol.* **1998**, *117*, 761.

(101)   Clark, G. S. *Perfumer & flavorist* **1991**, *16*, 27.

(102)   Hocking, M. B. *J. Chem. Educ.* **1997**, *74*, 1055.

(103)   Li, K.; Frost, J. *J. Am. Chem. Soc.* **1998**, *120*, 10545.

(104)   Giembycz, M. A. *Expert opin. investigat. drugs* **2001**, *10*, 1361.

(105)   Li, Y.; Li, S.; Han, Y.; Liu, J.; Zhang, J.; Li, F.; Wang, Y.; Liu, X.; Yao, L. *Eur. J. Pharmacol.* **2008**, *591*, 252.

(106)   R. Burch, T. Anderson, J. Lazar, *Capsaicinoid Gel Formulation and Uses Thereof.*, **2006**, US20060148903 A1.

(107)   Dinter, H.; Tse, J.; Halks-Miller, M.; Asarnow, D.; Onuffer, J.; Faulds, D.; Mitrovic, B.; Kirsch, G.; Laurent, H.; Esperling, P. *J. Neuroimmunol.* **2000**, *108*, 136.

(108) Hanessian, S.; Guesné, S.; Chénard, E. *Org. lett.* **2010**, *12*, 1816.

(109) Beeh, K. M.; Beier, J.; Lerch, C.; Schulz, A. K.; Buhl, R. *Lung* **2004**, *182*, 369.

(110) Sparta, M.; Alexandrova, A. N. *PloS one* **2012**, *7*, e47172.

(111) Piedrafita, F.; Elorriaga, C.; Fernández-Alvarez, E.; Nieto, O. *J. Enzym. Inhib. Med. Chem.* **1990**, *4*, 43.

(112) Zhang, J.; Klinman, J. P. *J. Am. Chem. Soc.* **2011**.

(113) Shin, S.; Lee, T. H.; Ha, N. C.; Koo, H. M.; Kim, S. y.; Lee, H. S.; Kim, Y. S.; Oh, B. H. *The EMBO journal* **2002**, *21*, 2509.

(114) Vieira-Coelho, M.; Soares-da-Silva, P. *Brain res.* **1999**, *821*, 69.

(115) Ehler, A.; Benz, J.; Schlatter, D.; Rudolph, M. G. *Acta Crystallograph. D* **2014**, *70*, 2163.

(116) Li, X.-Q.; Zhang, T.; Donnelly, D. *PloS one* **2011**, *6*, e18615.

(117) Nissinen, E.; Lindén, I.-B.; Schultz, E.; Kaakkola, S.; Männistö, P. T.; Pohto, P. *Eur. J. Pharmacol.* **1988**, *153*, 263.

(118) Ferraiolo, B. L.; Onady, G. M.; Mieyal, J. J. *Biochemistry* **1984**, *23*, 5528.

(119) Tsuji, E.; Okazaki, K.; Isaji, M.; Takeda, K. *J. Struct. Biol.* **2009**, *165*, 133.

(120) Moor, N.; Klipcan, L.; Safro, M. G. *Chem. Biol.* **2011**, *18*, 1221.

(121) Rodgers, K. J.; Dean, R. T. *Int. J. Bioch. Cell B.* **2000**, *32*, 945.

(122) Ozawa, K.; Headlam, M. J.; Mouradov, D.; Watt, S. J.; Beck, J. L.; Rodgers, K. J.; Dean, R. T.; Huber, T.; Otting, G.; Dixon, N. E. *Febs Journal* **2005**, *272*, 3162.

(123) Ozeki, H.; Wakamatsu, K.; Ito, S.; Ishiguro, I. *Anal. Biochem.* **1997**, *248*, 149.

(124) Riley, P. *Cell. Mol. Biol.* **1999**, *45*, 951.

(125) Marumo, K.; Waite, J. *BBA Protein Struct. M.* **1986**, *872*, 98.

(126) Ito, S.; Kato, T.; Shinpo, K.; Fujita, K. *Biochem. J.* **1984**, *222*, 407.

(127) Reitz, A.; Avery, M. A.; Verlander, M. S.; Goodman, M. *J. Org. Chem.* **1981**, *46*, 4859.

(128) Kilenyi, S.; Mahaux, J.; Van Durme, E. *J. Org. Chem.* **1991**, *56*, 2591.

(129) Dekker, C. A.; Fruton, J. S. *J. Biol. Chem.* **1948**, *173*, 471.

(130) Peeling, W. B. *Urology* **1989**, *33*, 45.

(131) Sanu, O.; Lamont, R. F. *Ther. Clin. Risk Manag.* **2010**, *6*, 191

(132) Lee, H.-J.; Macbeth, A. H.; Pagani, J.; Young, W. S. *Prog. Neurobiol.* **2009**, *88*, 127–151.

(133) Maynard, P.; Nicholson, R. *Brit. J. Cancer* **1979**, *39*, 274.

(134) Blaschke, E.; Hertting, G. *Biochem. Pharmacol.* **1971**, *20*, 1363.

(135) Naish-Byfield, S.; Cooksey, C. J.; Riley, P. A. *Biochem. J.* **1994**, *304*, 155.

(136) Selinheimo, E.; Gasparetti, C.; Mattinen, M.-L.; Steffensen, C. L.; Buchert, J.; Kruus, K. *Enzyme Microb. Tech.* **2009**, *44*, 1.

(137) Lipmann, F.; Roskoski, R.; Gevers, W.; Kleinkauf, H. *Biochemistry* **1970**, *9*, 4839.

(138) Rammelkamp, C. H.; Weinstein, L. *J. Infect. Dis.* **1942**, 166.

(139) Lin, H.; Thayer, D. A.; Wong, C.-H.; Walsh, C. T. *Chem. Biol.* **2004**, *11*, 1635.

(140) Vosloo, J. A.; Stander, M. A.; Leussa, A. N.-N.; Spathelf, B. M.; Rautenbach, M. *Microbiology* **2013**, *159*, 2200.

(141) Chang, T.-S. *Int. J. Mol. Sciences* **2009**, *10*, 2440.

(142) Spencer, R. P.; Valentine, R. J.; Field, J. B. *Acta pharmacologica et toxicologica* **1956**, *12*, 196.

(143) Biastoff, S.; Teuber, M.; Zhou, Z. S.; Dräger, B. *Planta medica* **2006**, *72*, 1136.

(144) Hendricks, C. L.; Ross, J. R.; Pichersky, E.; Noel, J. P.; Zhou, Z. S. *Analytical biochemistry* **2004**, *326*, 100.

(145) Dorgan, K. M.; Wooderchak, W. L.; Wynn, D. P.; Karschner, E. L.; Alfaro, J. F.; Cui, Y.; Zhou, Z. S.; Hevel, J. M. *Analytical biochemistry* **2006**, *350*, 249.

(146) Lee, J. E.; Cornell, K. A.; Riscoe, M. K.; Howell, P. L. *J. Biol. Chem.* **2003**, *278*, 8761.

(147) Gomi, T.; Date, T.; Ogawa, H.; Fujioka, M.; Aksamit, R.; Backlund, P.; Cantoni, G. *J. Biol. Chem.* **1989**, *264*, 16138.

(148) Collazo, E.; Couture, J.-F.; Bulfer, S.; Trievel, R. C. *Anal. Biochem.* **2005**, *342*, 86.

(149) Palmer, N. A.; Sattler, S. E.; Saathoff, A. J.; Sarath, G. *J. Agr. Food Chem.* **2010**, *58*, 5220.

(150) Liscombe, D. K.; Facchini, P. J. *J. Biol. Chem.* **2007**, *282*, 14741.

(151) Choi, K.-B.; Morishige, T.; Shitan, N.; Yazaki, K.; Sato, F. *J. Biol. Chem.* **2002**, *277*, 830.

(152) Huo, F.-J.; Sun, Y.-Q.; Su, J.; Chao, J.-B.; Zhi, H.-J.; Yin, C.-X. *Org. lett.* **2009**, *11*, 4918.

(153) Valpuesta, M.; Ariza, M.; Díaz, A.; Suau, R. *European Journal of Org. Chem.* **2010**, *2010*, 4393.

(154) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455.

(155) Brandt, W.; Manke, K.; Vogt, T. *Phytochemistry* **2015**, *113*, 130.

(156) Hansen, E. H.; Møller, B. L.; Kock, G. R.; Bünner, C. M.; Kristensen, C.; Jensen, O. R.; Okkels, F. T.; Olsen, C. E.; Motawia, M. S.; Hansen, J. *Appl. Environ. Microb.* **2009**, *75*, 2765.

(157) Van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J. *J. Mol. Biol.* **1993**, *229*, 105.

(158) Åkerlund, M.; Carlsson, A.; Melin, P.; Trojnar, J. *Acta Obstet. Gyn. Scand.* **1985**, *64*, 499.

(159) Poor, C. B.; Andorfer, M. C.; Lewis, J. C. *ChemBioChem* **2014**, *15*, 1286.

(160) Meretoja, O.; Taivainen, T.; Wirtavuori, K. *Brit. J. Anaesth.* **1995**, *74*, 6.

(161) Ohno, A.; Kawasaki, N.; Fukuhara, K.; Okuda, H.; Yamaguchi, T. *Magn. Res. Chem.* **2010**, *48*, 168.