University College London (UCL)

Development of biocatalysts for the alkylation of complex molecules

Benjamin Thair

Thesis submitted to University College London for the degree of Doctor of

Philosophy in Chemical Biology

Declaration

Plasmids containing *Mj*MAT, *Tk*MAT and *Hs*MAT2A were provided by Prof. Jennifer Andexer. The plasmid containing *Uu*MAT was provided by Dr. Shimon Bershtein. *E. coli* strains containing *Rn*COMT, *Ec*MAT, *Ec*MTAN, *Tf*NCS and *Cv*TAm were provided as glycerol stocks by Dr. Fabiana Subrizi. An *E. coli* strain containing HLADH was provided as a glycerol stock by Dr. Daniel Méndez-Sánchez. The method used for HPAE-IPAD analysis was developed by Dr. Gorjan Stojanovski.

A version of Section 1.3 has been published: E. Abdelraheem, B. Thair, R. Fernández Varela, E. Jockmann, D. Popadić, H. C. Hailes, J. M. Ward, A. M. Iribarren, E. S. Lewkowicz, J. N. Andexer, P.-L. Hagedoorn, U. Hanefeld, *Chembiochem*, 2022, **23**, e202200212.

I, Benjamin Thair, confirm that this thesis presents work that is my own and where information has been derived from other sources, this has been indicated.

Abstract

The medicinal effects of compounds can be dramatically improved by even the smallest alkylations of their structures. Making those additions selectively, however, can be challenging for traditional synthesis, especially when molecules have many reactive sites. This has limited the ability to access and study derivatives of complex molecules, which may be hampering the discovery of new bioactives.

Methyltransferases (MTs) are a vast, structurally divergent class of enzymes, responsible for catalysing nearly all alkylation reactions in cells. In many cases, the transferral of a methyl group by an MT is highly selective, and able to significantly alter the interactions of the target. MTs thus have the potential to access the sought-after alkylated derivatives of medicinally promising structures. In doing so, they may allow the study and manufacture of molecules that would otherwise be difficult to produce.

This thesis documents work towards the goal of realising that potential. A three-enzyme cascade was used to generate a reactive cofactor *in situ*, use that cofactor in MT reactions and break down the inhibitory side product. With this system, selective alkylation of compounds featuring the privileged tetrahydroisoquinoline scaffold were explored through small- and preparative-scale assays, with primary analysis by HPLC. The first efforts were towards methylation, comprising investigations using the more established biocatalyst catechol-*O*-MT alongside a search for novel MTs. The work then evolved into the development of capacity for other alkylations. Analogues of the alkyl donor methionine, sourced commercially and later through syntheses, were integrated into the cascades. The challenges encountered during this effort prompted computational investigations, directed mutagenesis to improve enzyme performance and ultimately a high-throughput random mutagenesis screen to search for improved enzyme variants.

Impact Statement

The work presented in this thesis has the potential for impacts both within academia and in wider industry. As part of this research, a mutagenesis workflow, able to generate and assay large numbers of variants with relatively basic equipment, was designed and tested. While more refinement is needed, the method has general applicability to future enzyme engineering efforts both within the group and beyond. Indeed, steps of the workflow have already been used by other researchers in the group looking to screen many enzymes simultaneously. The effort to discover new methyltransferase enzymes was also greatly informative, and a number of recommendations were generated as a result which may expedite future efforts.

On a broader scale, this work may serve to advance the case for biocatalysis in industry. The investigation in this thesis demonstrated preparative-scale reactions from economical starting materials, with reactions under benign conditions. It also showed the applicability and flexibility of the cofactor supply system. All this, alongside parallel efforts by many other groups, may serve to increase the feasibility of biocatalyst use from the perspective of chemical manufacturers. This is especially relevant for methyltransferases, which have enormous potential, but have typically been considered too expensive and specific to be useful. The more this enzyme class can be shown as viable biocatalysts, through work such as this, the greater the chance that they will be developed at larger scales. This has the potential to improve drug discovery by increasing the ease with which diversified compounds can be accessed, in turn increasing the likelihood that molecules with desired characteristics are found. This in turn, would have wider implications for public health. Furthermore, the replacement of high-energy or wasteful steps in chemical manufacture processes with greener, enzyme-based alternatives can contribute to the wider effort of reducing the carbon emissions.

Acknowledgements

My first thank you goes to Professor Helen Hailes, for her guidance, support, and indefatigable optimism. When I was sure my project wasn't worth the agar it was spread on, she helped me find the thread of inquiry that could still be pulled, and the motivation to tug it for just a little while longer. I would also like to thank my former secondary supervisor, Professor John Ward, for being a calm, sage presence throughout my project, who brought patient curiosity to challenges that had become infested with frustration. I'm also grateful to Dr. Jack Jeffries, my current secondary supervisor, upon whom I was foisted at a chaotic moment in his own career, but who has ever been generous with his time.

All members of the Hailes-Tabor group deserve my thanks, for together nurturing a positive work environment. In particular, I want to thank Dr. Fabiana Subrizi for her patience, and for laying the foundation of our group's work with methyltransferases, Dr. Liam Martin for teaching me the little chemistry I've grasped, and Dr. Rebecca Roddan for taking me under her wing when I was but a fledgling researcher. I am grateful for every second others have leant me that I might learn, experience, grow and be reassured.

I would like to thank my funding body, the Wellcome Trust, for their generous investment in my training. I am also grateful to the staff of the Institute of Structural and Molecular Biology for running this doctoral programme. In particular, I would like to thank Professor Finn Werner, whose delight at the progress of our cohort was always touching. The whispered praise he gave me as I left my interview marked an unforgettable beginning to this journey.

The last three people are difficult to thank, because words can't enclose what they have meant to me. Billy, Eve and Esther: know that I am grateful, simply and fully, that you are who you are, and that we were together in the time that we were. You brought light and laughter to moments that were devoid of both, like a magician pulling a rabbit out of an open casket. That I've come to know you feels as much an achievement as anything else I've done here. Everything of merit in these pages stands as testament to your friendship.

Contents

Symbols and Abbreviations

Å	Ångström
ADH	Alcohol Dehydrogenase
aq	Aqueous
AU	Absorbance Units
CH_2Cl_2	Dichloromethane
$CHCl_3$	Chloroform
δ	Chemical Shift
CMT	Carbon methyltransferase
DCM	Dichloromethane
dr	diastereomeric ratio
diH_2O	Deionised water
DMC	Dimethyl carbonate
DMF	Dimethylformamide
DNA	Deoxyribonucleic Acid
Dp	Decomposition point
ee	enantiomeric excess
eqv.	Equivalents
Et_2O	Diethyl ether
EtOAc	Ethyl acetate
g	Graviational Force
h	Hours
H_2O	Water
HCl	Hydrochloric Acid
HCTL	Homocysteine thiolactone
HMBC	Heteronuclear Multiple Bond Correlation
HPLC	High-performance liquid chromatography
Hz	Hertz
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
J	Coupling Constant

kcal	kilocalorie
L-eth.	L-ethionine
L-met.	L-methionine
LB	Lysogeny broth
LC-MS	Liquid Chromatography-Mass Spectrometry
LiOH	Lithium hydroxide
MAT	Methionine adenosyl transferase
μg	Microgram
μL	Microlitre
μM	Micromolar
min	Minutes
M	Molar
MeOH	Methanol
mg	Milligram
mL	Millilitre
mM	Millimolar
mmol	Millimole
Mp	melting point
MT	methyltransferase
MTAN	Methylthioadenosine nucleosidase
NCS	Norcoclaurine synthase
NAD+	Nicotinamide Adenine Dinucleotide
NaOH	Sodium Hydroxide
Na_2SO_4	Sodium Sulphate
NMR	Nuclear Magnetic Resonance
NMT	Nitrogen methyltransferase
NOESY	Nuclear Overhauser Effect Spectroscopy
OD_{600}	Optical Density (at a given wavelength (nm))
OMT	Oxygen methyltransferase
PAGE	Polyacrylamide Gel Electrophoresis

RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal RNA
r.t.	Room temperature
RT	Retention Time
SAH	*S*-adenosyl homocysteine
SAE	*S*-adenosyl ethionine
SAM	*S*-adenosyl methionine
SDS	Sodium Dodecyl Sulfate
sec	second
SOC	Super Optimal Broth with Catabolite Repression
SMT	Sulfur methyltransferase
TFA	Trifluroacetic acid
THF	Tetrahydrofuran
THIQ	Tetrahydroisoquinoline
TLC	Thin Layer Chromatography
tRNA	Transfer RNA
w/v	Weight per Volume
V	Volts
v/v	Volume per Volume

List of Figures

List of Schemes

Chapter 1

Introduction

1.1 Alkyl groups in medicine

1.1.1 From small beginnings

Small structural changes can dramatically impact the function of a drug. Even a change as subtle as methylation can increase potency,[1,2] binding affinity,[3] solubility in an aqueous solvent[4] or a number of other features.[5] Examples of profound changes caused by methylation are so widespread, the term 'magic methyl effect' has arisen to describe them.

The impact of methylation is evident in natural products. The antibiotic peptide cypemycin, produced by *Streptomyces sp.* OH-4156, is completely inactive without an *N*-terminal dimethylation.[6] Similarly, the antibiotic plantazolicin, a powerful agent against anthrax-causing *Bacillus anthracis*, has no activity without dimethylation of its terminal arginine.[7] Celesticetin and lincomycin are natural, small molecule antibiotics, both of which are activated by methylations.[8]

The magic methyl effect often stems from protein-ligand interactions and solubility in aqueous versus hydrophobic environments. A common expectation is that a methyl group will decrease the overall polarity of a molecule. This favours solvation in the hydrophobic interiors of proteins and membranes, promoting passage into the cell and binding to the target.[9] Analyses of simulations and inhibition data for crystal structures have estimated the benefit to hexadecane solvation, per methyl group added, to be 0.7-0.8 kcalmol^{-1}.[9,10] However, the actual benefit to binding affinity is dependant on the context. A survey of

2145 inhibitors found that replacement of a hydrogen atom with a methyl group was at least as likely to harm binding affinity as help it, and cases of substantial improvement were rare.[9]

The benefit of methylation to a drug is thus heavily dependant on the properties of the molecule and the details of its interactions with the target. Shape complementarity can play a large role, with methyl groups increasing affinity when they project into hydrophobic pockets, or decreasing it when they clash with active site residues.[11] Theophylline **1** differs from caffeine **2** by a methyl group at 7-N, but the two have nearly exclusive effects (Figure 1.1).[12] The increased lipophilicty of **1** could change its distribution in the body, and alter its affinity for the 11 distinct isoforms of the compounds' target receptor.

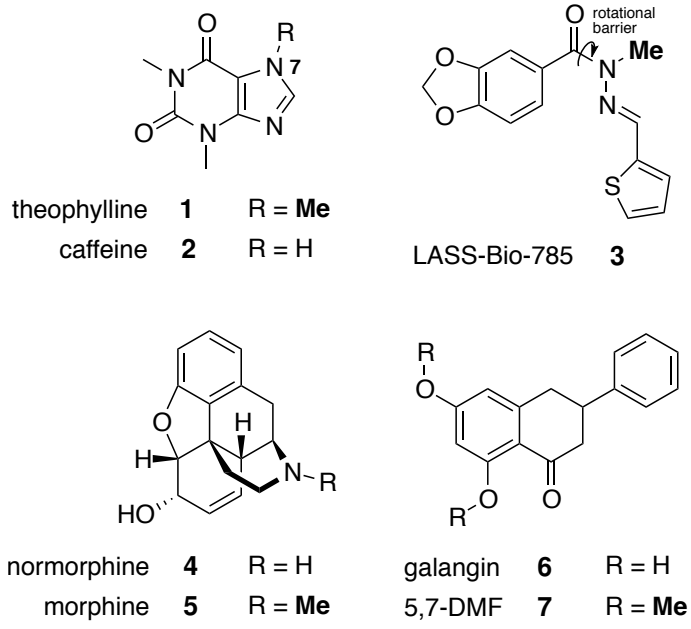


Figure 1.1: Compounds demonstrating magic methyl effects. Relevant methyl groups shown in bold.

Complementarity effects can also be subtle. *N*-methylation of LASS-Bio-785 **3** made it both more selective and seven times more potent (Figure 1.1). Furthermore, the methyl group created a 17 kcalmol^{-1} rotational barrier around the N-C bond. Research has shown that barriers such as these can force flexible molecules to adopt topolgies in solution that more closely mimic the bound conformations.[11,13,14] This, in turn, reduces the entropic penalty of binding and increases affinity.

As a weakly electron-donating species, the methyl group can also influence the elec-

tronic arrangement of a molecule. For example, normorphine **4** has approximately six times less analgesic activity than morphine **5** (Figure 1.1). The difference cannot be explained by active site interactions, most of which are electrostatic. Instead, the lack of a methyl group in **4** polarises the nitrogen, inhibiting its passage through the blood brain barrier to the site of action.[12]

Aside from direct target effects, addition of methyl groups to metabolically vulnerable sites can improve a compound's pharmacokinetic properties. Dimethylation of the flavone galangin **6** to give 5,7-dimethoxyflavone **7** has been found to protect against glucoronidation and sulfation in liver cell preparations (Figure 1.1). These processes lower the oral bioavailability of flavones, so methylation may open new routes of administration for these medicines.[15]

Alkylations of complex molecules thus have the potential to grant new properties or enhance existing ones. With the rate of new medicine discovery is continuing to decline,[16] strategies for the methylation of existing drugs could be invaluable in revealing new and improved aspects of the current arsenal.

1.1.2 The limits of synthetic methylation

In traditional organic synthesis, there are two main methods for incorporating a methyl group into a structure. The first is to use starting materials with the group already installed. This is only feasible if the molecule can be readily synthesised, but many desirable structures are difficult to generate *de novo*.[17] Such a case was noted with a series of mGluR5 antagonists, developed by GlaxoSmithKline. In that study, accessing more potent, methylated variants of a lead compound required recreating them from methylated starting materials.[18] Diversifying an entire combinatorial library by this method would require new synthetic sequences for potentially thousands of compounds.

If starting from an existing structure, the second option is to use an alkylating agent, such as methyl iodide. The reaction involves addition of the nucleophile target to methyl iodide via an S_N2 substitution.

Adding groups directly to the carbon skeleton, however, has typically required deprotonation of sufficient acidic C-H bonds, such as those which are adjacent to carbonyl groups, and reaction of the resulting nucleophilic carbanion.[19] Recent research

has sought to improve on this, with the development of a number of C-H methylation strategies including transition metal catalysis and radical-based reactions.[19] However, despite these successes, the problem of regioselectivity complicates the use of methylating agents, and only grows as the number of reactive sites increases. This, in turn, may require the use of protecting groups, increasing the number of synthetic steps involved.[20] Therefore, even with advances in synthesis, demand persists for methods which enable selective methylation of complex molecules.

1.2 Alkylation in cells

As with many reactions that are challenging for synthesis, selective methylation is performed with relative ease in nature. The enzymes responsible are methyltransferases (MTs), of which there are hundreds of types with roles throughout all cellular systems.

1.2.1 Functions of methyltransferases

On the largest molecular scale, MTs assist in managing entire genomes. In eukaryotes, DNA cytosine residues in the sequence CpG can be methylated to 5-methylcytosine. These CpG sequences cluster together as 'islands' containing hundreds of dinucleotide repeats.[21] The classical understanding is that when these islands are methylated, expression of genes whose transcription start sites are nearby is repressed. However, further research has shown that the effect of CpG methylation is context-dependant, and even the route by which it influences expression is obscure.[21] Studies have further linked CpG methylation to transcription stimulation, splicing, genome stability and the control of transposable elements.[22]

Cytosine C-5 methylation is overseen by three enzymes: DNMT1, DNMT3a and DNMT3b. DNMT1 typically has a maintenance role, copying patterns of CpG methylation from an old strand to the new strand after DNA replication.[23] This ensures that epigenetic signals in a parent cell are inherited by daughters. DNMT3a and 3b are more associated with *de novo* methylation occurring as a response to environmental or internal influences.[23]

In prokaryotes, other forms of methylation are more prevalent, including at adenine

C-6 and cytosine C-4.[24] Prokaryotic DNA methylation often forms part of restriction-modification systems, wherein the mark allows DNA-cleaving endonucleases to distinguish between host DNA, which is left intact, and foreign DNA, which is destroyed. However, the modification has also been implicated in regulation of genes when it occurs in the respective regulatory regions.[24] Thus, DNA MTs across the domains of life act as custodians of genetic information.

Methylation is also a common post-translational modification of proteins. Lysine and arginine residues on histones are methylated to create docking sites for effector proteins on chromatin.[25] Combinations of these modifications forms a complex 'epigenetic code', with consequences for the expression of nearby genes.[25] Reversible protein carboxymethylation also has roles in cell signalling, analogous to the more widespread phosphorylation, but with a more subtle effect on structural dynamics.[26] The function of protein methylation in bacterial memory gives a particularly clear example.[27]

On the smallest scale, MTs can target individual inorganic ions. Arsenic is an environmental toxin with two dominant oxidation states: arsenite (As(III)) and arsenate (As(V)).[28] As(III) is the more mobile and toxic of the two, so many organisms employ As(III) MTs to seqeuntially methylate arsenite. The resulting methylarsenicals readily oxidise to their less toxic pentavalent forms, which helps to manage low levels of the toxic metal in food and water.[28] Mercury(II) is also methylated by some anaerobic microbes, producing the more toxic methyl-mercury which is then exported for as yet unknown reasons.[29]

Somewhere between these limits, MTs have important roles in natural product biosynthesis.[30] Methylated natural products include the mycolic acid of *M. tuberculosis* cell walls,[31] the cofactor cobalamin,[32] insect-repellent nicotine[33] and the airborne signalling compound methyl jasmonate.[34] The natural antibiotics cypemycin, platazolicin, celesticetin and lincomycin were described above, and are also products of MT reactions.[6–8] Regioselective methylations of those compounds are responsible for their transformations into bioactive molecules.

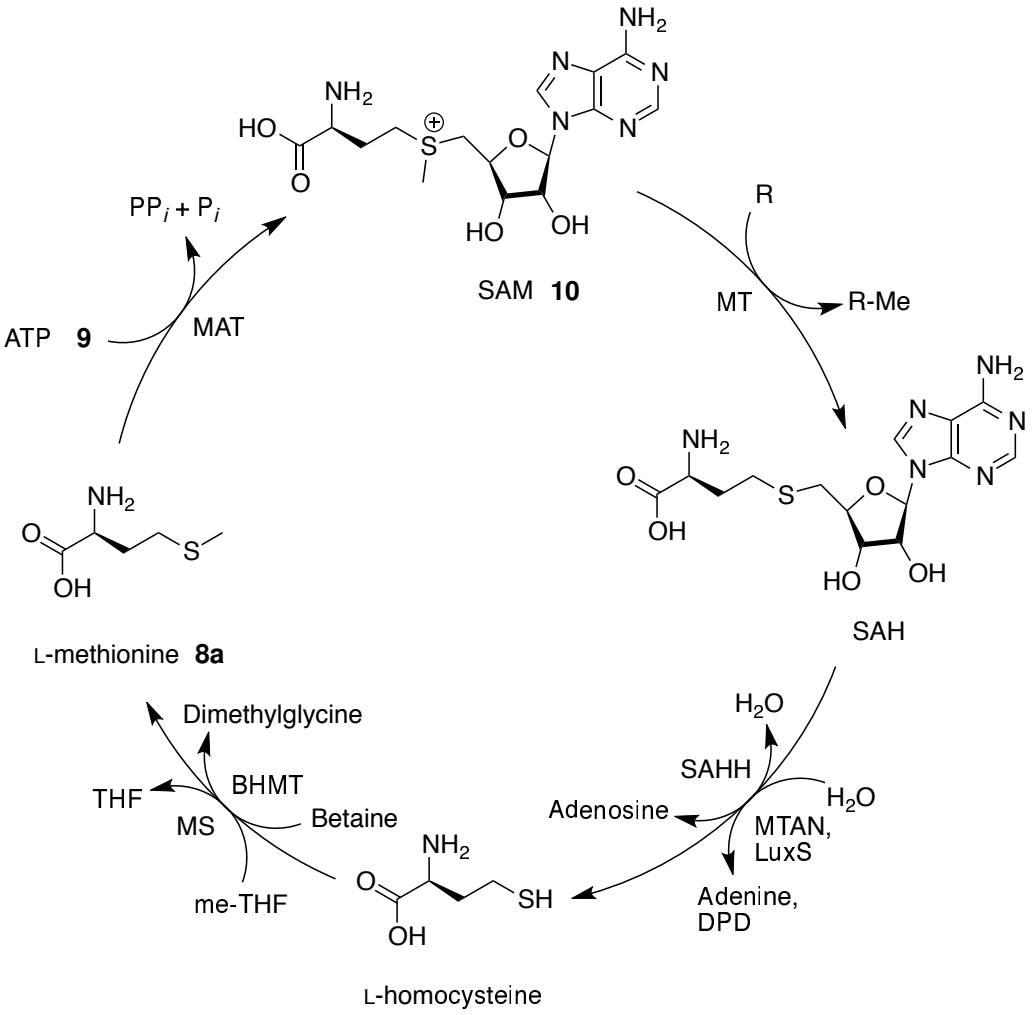
1.2.2 Cellular sources of activated methyl groups

Despite the enormous range of functions performed by MTs, almost all draw from the same source of activated methyl groups: *S*-adenosyl methionine (SAM). Behind ATP,

SAM is the second most-used cofactor in living cells.[35] Although referred to as a cofactor, SAM is consumed by each methylation reaction and must dissociate from the MT to be remade.

Regeneration of SAM occurs by an enzymatic cycle (Scheme 1.1). After methylation of the substrate, the spent cofactor *S*-adenosyl homocysteine (SAH) is cleaved by one of two routes. The first involves reversible hydrolysis to L-homocysteine and adenine, catalysed by SAH hydrolase (SAHH).[36] The second begins with hydrolysis, this time to *S*-ribosyl homocysteine and adenine by methylthioadenosine nucleosidase (MTAN), followed by further lysis of the former by LuxS to give L-homocysteine and 4,5-dihydroxy-2,3-pentanedione.[37] Whether the first, second or both paths are followed depends on the species. In either case, the resulting L-homocysteine is next methylated to L-methionine. This is catalysed either by methionine synthase using N^5-methyl-tetrahydrofolate as a methyl donor,[38] or by betaine homocysteine MT using betaine.[39]

The final step of the cycle consumes ATP to generate SAM from L-methionine. This reaction, and the methionine adenosyltransferase (MAT) that catalyses it, is well-conserved across all domains of life.[40] The MAT *E. coli* (*Ec*) has been well-characterised and the mechanism elucidated. *Ec*MAT exists as a dimer in solution, with two active sites at the interface of the subunits (Figure 1.2).[41] The two substrates, ATP **9** and L-methionine **8a**, bind at these sites. In the first state of the catalytic mechanism, the oxygen which bridges the ribose moiety of ATP to the polyphosphate chain is coordinated by K165 and one of the imidazole nitrogens of H14 (Scheme 1.2). The other nitrogen is itself coordinated by the backbone nitrogens of D16 and K17. This polarizes the ring, triggering a chain of electron transfers beginning from the C-O bonding electrons in ATP, travelling through the O-N hydrogen bond and terminating on the stabilised imidazole nitrogen. Meanwhile, the ribose carbon is attacked by the lone pair of the methionine sulfur, cleaving the C-O bond. The electron transfer chain then reverses, sending electrons back through the imidazole to the phosphoryl oxygen and completing the reaction.[41] In reality, cleavage of the ribose-O bond and formation of the S-ribose bond likely happen in a concerted, S_N2 fashion. The final products are SAM **10**, in which the conformation of the ribose carbon has been inverted, and tripolyphosphate **11** with an additional negatively charged oxygen on the γ phosphate (Figure 1.2, Scheme 1.2).



Scheme 1.1: Overview of the SAM cycle. R respresents any substrate, SAHH = SAH hydrolase, MTAN = methythioadenosine nucleosidase, DPD = 4,5-dihydroxy-2,3-pentanedione, MS = methionine synthase, THF = tetrahydrofolate, BHMT = Betaine homocysteine MT, MAT = methionine adenosyl transferase.

Energy for the next reaction is provided by subsequent cleavage of the polyphosphate inside the active site, giving phosphate and pyrophosphate.[41] The mechanism is assisted by two Mg^{2+} ions. One binds all three phosphoryl moeities, while the other binds only the α and γ groups, bending the polyphosphate chain into a pair of ring systems (Figure 1.2). Quantum mechanical calculations have suggested these ions dissipate the negative charge that would otherwise accumulate on the oxygen as the C-O bond is cleaved.[42] The active site also contains a K^+ ion, which is non-essential for catalysis but improves the reaction rate substantially. It is thought that coordination of this ion by surrounding residues holds the active site in an optimal shape.[42]

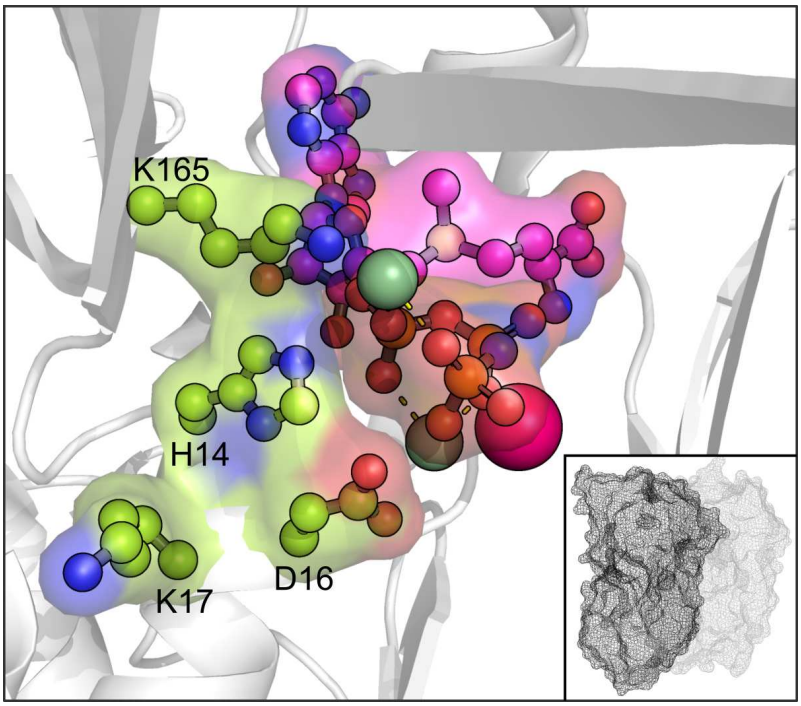
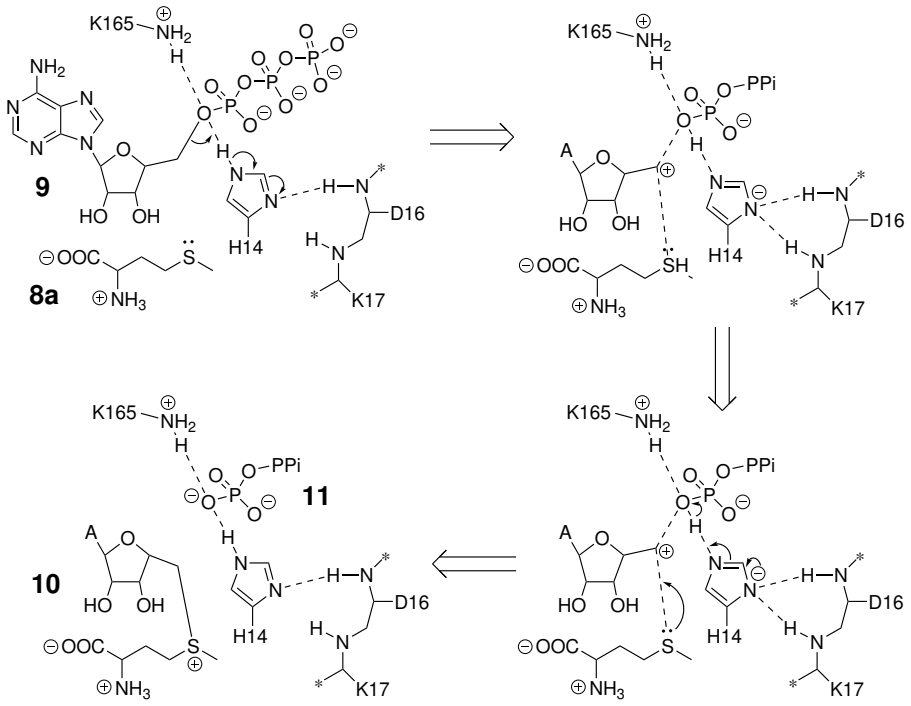


Figure 1.2: Active site and catalysis products of *Ec*MAT. Protein residues shown in green, cofactor shown in magenta, phosphate atoms shown in orange, magnesium and potassium ions shown as pale green and pink spheres, respectively. Dimer topology shown in bottom right corner. The position of tripolyphosphate is represented in the crystal structure by an nitrogen-substituted analogue that is believed to adopt the same conformation (PBD: 1RG9).

Together, this cycle maintains a source of activated methyl groups for use by almost all MTs. Two other methyl donors were described above for converting L-homocysteine to L-methionine: betaine, and N^5-methyl-tetrahydrofolate. The latter has other applications, such as in the synthesis of thymine nucleotides[43] and direct methylation of uracil residues in *t*RNA.[44] Some methanogenic organisms also contain an array of MTs which use various small compounds, such as methylated amines and methanol, as methyl donors. Methyl groups from these molecules are transferred onto the cobalamin cofactors of the specialised MTs, then collected by another set of MTs and transferred to coenzyme M.[45] A number of other niche methyl sources have been found. However, most MTs across the myriad cellular processes with which they are involved have all adapted to use SAM. This unity, especially given their many disparate functions, begs the question of what structural or phylogenetic features might be shared across the family. The answer, however, is not straightforward.



Scheme 1.2: Mechanism of SAM formation from ATP and L-methionine in *Ec*MAT. Adapted from Figure 7 in Komoto *et al.*.[41] While presented as discrete steps for clarity, the real mechanism is most likely a concerted S_N2 reaction.

1.3 Evolution and structures of methyltransferases

Enzymes which use SAM as a methyl donor have been present since the last universal common ancestor. Those primordial MTs have multiplied and diversified over millions of years, while convergence has pulled together unrelated families from the several independent re-discoveries of this enzyme.[46] The present evolutionary picture is therefore complex and, to date, incomplete. Nonetheless, the bioinformatic work of Schubert *et al.* showed that the structures of SAM-dependent MTs coalesce into five fold classes.[35] These classes share little in terms of sequence, mode of SAM binding, tertiary structure and quaternary behaviour. However, the recurrence of structural motifs within the groups has illuminated their key features.

1.3.1 The classical five fold classes

The first MT structure to be solved was of HhaI.[47] Its topology defines class I, which has remained the largest group, and is otherwise referred to as the Rossman-like struc-

ture. Class I MTs feature an alternating α/β sequence, which folds into a 7-strand sheet sandwiched by helices on either side (Figure 1.3A).[47] SAM binds in an elongated conformation at the C-terminal ends of β1 and β2. Many characteristic features of the class are located in this region. They include a loosely conserved GxGxG motif between β1 and αA to accommodate the adenine moiety, a hydrogen bonding residue between β2 and αB to coordinate the ribose moiety, and an acidic residue within β1, which may make water-assisted hydrogen bonds with the methionine.[35,46] However, the neighbouring substrate-binding regions vary heavily. This reflects the diverse functions of the class, which comprise methylations of small molecules,[48–50] cofactors,[51] antibiotics[7] and biomacromolecules.[47,52,53] Substrate selectivity is sometimes enforced by additional domains, which may be appended to the Rossman fold[48,50] or embedded within it.[47,49]

In contrast to class I, class II MTs have only one known function: to reactivate oxidised cobalamin.[54] They form one domain of the modular methionine synthase, which features in the SAM cycle described above and uses cobalamin as a cofactor. If the intermediate Co(I) ion is oxidised, a conformational change presents the cobalamin to the MT, leading to reductive methylation of the ion and reactivation of the cofactor.[38] The central feature of the protein is a long antiparallel β-sheet with a sharp kink near one end (Figure 1.3B). Helices surround the sheet above and below and extend beyond both ends, forming a horseshoe-like topology. As with class I, SAM binds in an elongated conformation, in a groove by the inner edge of the central β-sheet.[38]

Class III methyltransferases are dimeric, with the active site situated between the two domains (Figure 1.3C).[55] The lobes are similar in composition, both featuring central β-sheets flanked by α-helices. However, their topologies are distinct (32415 for the N-terminal sheet, 12534 for the C-terminal), and the C-terminal domain has a β-turn between strands 4 and 5 to give a mostly antiparallel sheet.[55] In solution, the proteins dimerise across the bridge between the two domains, such that the sheets are continuous across the monomers.[55] SAM binds in a deep pocket at the bridge, near the N-terminal end of αE. Unlike in previous classes, the cofactor takes on a tightly folded conformation which exposes the activated methyl group to the exterior.[35] The main targets of class III MTs are tetrapyrroles; the prototypical example, cobalt-precorrin-4 MT, is involved in anaeorobic vitamin B12 biosynthesis.[32] The enzyme accommodates its substrate in a

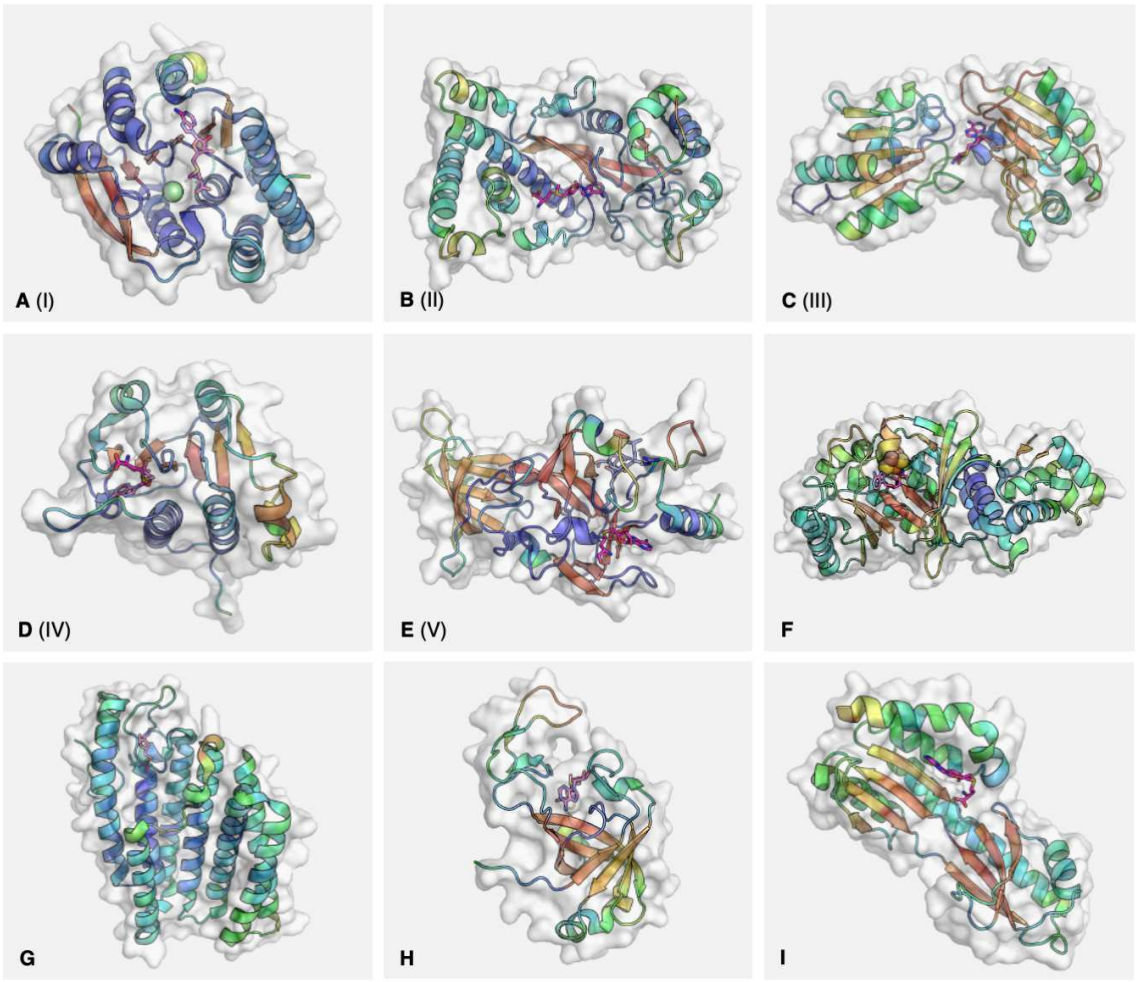


Figure 1.3: Examples of the nine structural classes of methyltransferases. Backbones shown as cartoons with translucent surfaces. Cofactor shown as magenta sticks. PDB codes: A - 6lfe (catechol-*O*-methyl-transferase); B - 1msk (methionine synthase, reactivation domain); C - 1cbf (cobalt precorrin-4-methyltransferase); D - 1mxi (tRNA (cytidine(34)-2'-*O*)-methyl-transferase); E - 1o9s (SET7/9); F - 3rfa (RlmN); G – 5vg9 (isoprenylcysteine carboxyl MT); H – 2nv4 (AF0241); I – 1tlj (Taw3). For E (1o9s), histone peptide represented as blue sticks. For F (3rfa), iron-sulphur cluster represented as spheres.

groove along the N-terminal domain. The site runs across the C-terminal ends of the parallel β-sheet, and is walled by loops emanating from those strands.[55] Other examples with similar structures include SAM uroporphyrinogen III methyltransferase (SUMT)[56] and sirohaem synthase,[57] both of which operate in the same biosynthetic pathway, as well as diphthine synthase which modifies a histidine residue on elongation factor 2.[52]

The class IV MTs are encompassed by the SPOUT enzyme superfamily, named after the bacterial tRNA MTs SpoU (now TrmH) and TrmD.[58] They have roles in post-

transcriptional modification of t- and r-RNAs, and are the second-most populous group after class I.[58] The class IV structure is analogous to class I, with a 6-strand, parallel β-sheet flanked by α-helices, although five of those helices sit on one side of the sheet and two on the other (Figure 1.3D). The crystal structure of TrmL shows SAH in a bent conformation, in a pocket formed by the ends of β3, β4 and β5 and their associated loops.[59] Part of the cofactor binding site is a rare 'knot' topology, formed by the loop between β6 and αG passing through the loop between β4 and αH.[59] This knot is the most characteristic feature of the class.[60] Regions outside the motif, however, have low conservation, even between members,[58] and may facilitate interactions with substrates or other proteins.[60] The catalytic mechanism of SPOUT MTs is in some cases unclear, but is thought to be assisted by a basic residue for both *N*- and *O*-methylation.[61] Dimerization was originally thought necessary for the catalysis. However, recent work has also demonstrated the activity of class IV monomers.[62]

The final numbered class, class V, are the SET domains (*S*uppressor of variegation, *E*nhancer of zeste and *T*rithorax).[63] They methylate the lysine side-chains of histones and other proteins important for transcriptional regulation,[64–69] as well as Rubisco.[70] The superfamily seems to have evolved through domain duplication in eukaryotes, before laterally transferring to bacteria.[46] The typical structure is relatively complex, comprising four α-helices embedded in a sequence of twisted β-sheets (Figure 1.3E). A pseudo-knot is formed by the loop prior to the C-terminal helix, though with a different topology to that in class IV.[63,64] In a crystal structure of SET7/9, SAM binds near this knot, in a channel at the surface of the protein. The histone binds on the opposite face, with the target lysine reaching through a narrow channel between the faces to approach the SAM methyl group from behind.[64] The cofactor is thought to be held in its tightly folded conformation by hydrogen bonding of the methionine amine to a conserved asparagine, with the methyl transfer catalysed by a tyrosine residue.[63] As with other classes, the SET domain itself is often flanked by poorly-conserved but essential pre- and post-SET regions that direct its specificity.[35]

1.3.2 Further fold classes

In the years since Schubert *et al.*'s classification, a number of SAM-dependant MT families with structures outside these groups have been characterised. The first, radical SAM methyltransferases, feature a catalytic domain with an α/β TIM barrel structure (Figure 1.3F).[71] Many enzymes use the 5'-deoxyadenosine radical intermediate generated from SAM, but only a small number use it to methylate their substrate.[71] The mechanisms of methylation have remained obscure. All rely on a [4Fe-4S] cluster, coordinated in the active site at the C-terminal end of the barrel by a characteristic CxxxCxxC motif.[72] RlmN and Cfr both transfer methylene groups to RNA substrates, which are then converted to methyl groups on-target.[73] TsrM, meanwhile, methylates C-2' of the tryptophan indole group, but seems to use methylcobalamin alongside SAM in a way that avoids homolytic cleavage of the latter.[74]

Another new class are the transmembrane MTs, of which isoprenylcysteine carboxyl methyltransferase (ICMT) is a well-characterised example. ICMT performs the final step in prenylcysteine modification of proteins. In eukaryotes, its six to eight α-helices span the endoplasmic reticulum membrane, with the sheltered binding site for SAM protruding into the cytoplasm (Figure 1.3G).[75,76] The substrate binds in a nearby cleft which begins in the cytoplasm and dives down the protein surface into the membrane.[76] A prokaryotic analogue MaICMT (or MaMTase) has a similar structure, with five transmembrane helices and a cytosolic cofactor binding site, but currently has no identified substrates.[77]

A third additional group are the β-barrel MTs, exemplified by the tRNA methyltransferase TrmO.[78] The crystal structure of a homologue from *Archaeoglobus fulgidus* shows a tight, β-barrel formed of six antiparallel strands (Figure 1.3H).[79] At the N-terminal end of the barrel, according to the direction of the β1, an extended helix crosses the circular face to connect β2 and β3. At the C-terminal end, three loops together form a large, lopsided V-shape, with SAM binding in a tight pocket within the point. The enzyme forms a dimer in solution, and the key residues in each monomer's cofactor pocket include M57 and L133 to sandwich the adenine, Q22 and R82 to hydrogen bond the methionine amine and K122, reaching across from the other subunit's β6, to hydrogen bond the carboxyl group.[79] Though the MT activity of the enzyme towards tRNA substrates has been con-

firmed, the binding site of the substrate has not.[78]

The fourth new class has been named Sso0622-like, and includes the *S. cerevisiae* protein TWY3 and its *Sulfolobus solfataricus* homologue Taw3.[80,81] A crystal structure of Taw3 shows two 4-strand, antiparallel β-sheets, connected end-to-end by disordered loops (Figure 1.3I). Both sheets lie against and parallel to the long spine of αE and are flanked by further helices.[81]The protein is actually an amalgamation and extension of two recognised folds. The sheet formed by β2, β3, β4 and β8, as well as αC beside it and part of that spine, constitute a RAGNYA domain, which typically mediates interactions with other macromolecules.[82] Strands β5, β6 and β7 and nearby αD form an SHS2 domain, which can take on a variety of functions.[83] The authors of the study were unable to demonstrate activity of the purified protein *in vitro*, nor were they able to obtain a structure co-crystallised with the cofactor. However, computational docking of the cofactor implicated a groove at the surface of the N-terminal extension, formed by αA, αB and β1, as the binding site.[81] Coordinates for the docked cofactor were not provided with the paper, but docking was able to recreate the binding site presented in the publication (Figure 1.3I).

These nine classes represent our current best picture of the variety of protein structures which can facilitate methylation. However, further characterisation of suspected MT genes, alongside ever-broadening metagenomic analysis, may lead to the identification of yet more classes with distinct traits.

1.3.3 Atom specificity and its relationship to structure

Due to the high electrophilicity of SAM, transfer by SAM-dependent MTs always occurs to a nucleophilic acceptor, regardless of structural class. The targets of all known MTs (according to the IUBMB database) are given in Figure 1.4. More than three-quarters target a heteroatom (O, N, S) while approximately 20% target carbon. Other, far rarer targets include arsenic,[28] tellurium,[84] phosphorus,[85] halides[86] and selenium.[87]

Atom specificities are not strictly divided between the structural classes. Class I, for example, contains MTs which transfer to all of the four most common target atoms.[47,88–90] However, in cases where the class has evolved niche functions, the acceptor atoms can be likewise limited. Class V MTs (SET domains) methylate histone lysine residues, so

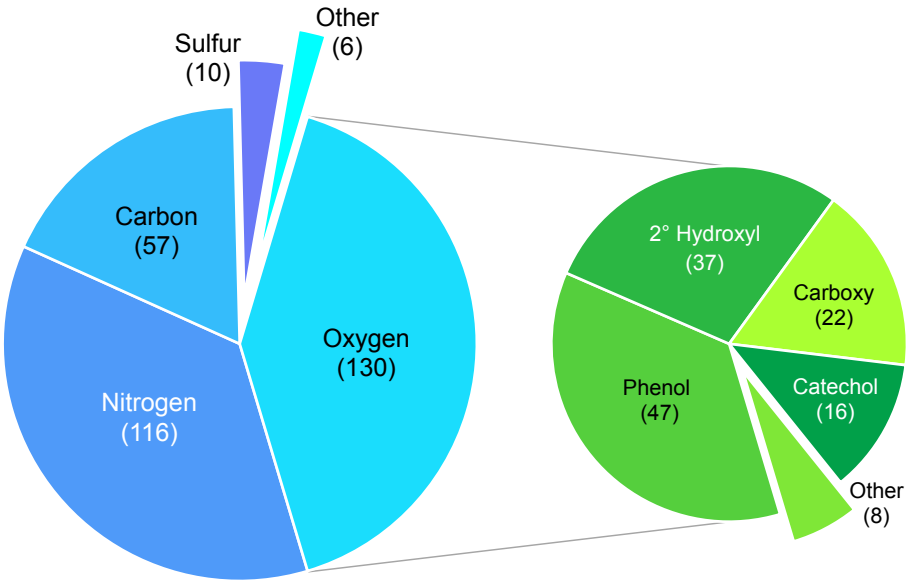


Figure 1.4: Methyltransferases categorised by targets. Left: target atoms of SAM-dependant methyltransferases, determined from the IUBMB database (EC 2.1.1). Number of enzymes shown in brackets. 'Other' category encompasses arsenic, tellurium, phosphorus, halides, cobalt and selenium. Right: specific target groups of OMTs. 'Other' category comprises oximes, primary and tertiary hydroxyls and phosphates.

are all nitrogen-MTs (NMTs).[63] The chemically aggressive mechanism of radical SAM MTs is only necessary with a stubborn acceptor like carbon, so all MTs of this class so far discovered are carbon-MTs (CMTs). However, *C*-methylation is also possible without this mechanism, as demonstrated by MppJ, which utilises an Fe(III) ion instead,[91] and PsmD, which uses a Y-E-H catalytic triad.[92] Therefore, despite interplay between class and targets, there is evidence of enzymes evolving different targets while maintaining their overall structures.

MTs targetting oxygen and nitrogen are by far the most common, perhaps due to the strong nucleophilicity of those atoms and the relative ease of the reaction. Of the two, OMTs are the most populous group of all. Phenolics of various configurations are their most common targets, followed by secondary hydroxyls most often found on sugar moieties (Figure 1.4). As with the atom selectivities, there are a small number of rare targets, which include phosphates, oximes and primary hydroxyls.

Due to the range of targets and structures, it is not possible to give a generic mechanism for methyltransferral, although a specific example will be discussed later in the

chapter. The understanding of MT form and function laid out in this section indicates something important, however: many different protein structures are able to methylate many different targets. It has been suggested that so many MT classes have evolved because the de-methylation of SAM is highly energetically favourable, meaning other parts of the protein can diverge as long as the essential residues are in place.[35] This, in turn, suggests the the core MT architecture may be sufficiently plastic to take on more than its already wide collective range of natural substrates. Specifically, MTs may be able to fill the need identified earlier in this chapter: for an agent which is able to specifically alkylate molecules of pharmacological interest.

1.4 Methyltransferases as biocatalysts

Biocatalysis, the use of enzymes to assist chemical reactions, is already well established, with a number of prominent success stories.[93] Attention on the field is only growing, however, as society looks for ways to reduce the environmental impact of chemical production. Given their selectivity, structural flexibility and potential advantages over synthetic methods, MTs are promising candidates for development into biocatalysts. However, there are specific challenges associated with this enzyme class that must be addressed before it can be used more widely and at scale.

1.4.1 Biocatalysis and sustainable chemistry

Biocatalysis research is situated within the wider sphere of green chemistry, the goal of which is to reduce the waste, energy intensity and environmental harm of chemical processes. Enzymes can be regio-, stereo- and substrate-specific, and their enveloping 3D structure allows them to direct reactions towards one product, all of which reduces waste. Most enzymes also work optimally under the conditions they were evolved for, meaning near-neutral aqueous solutions at mild temperatures.[20] Replacing steps in industrial pipelines with biocatalytic transformations thus provides an opportunity to improve the economy of the process, while reducing the associated hazards and environmental footprint. Many examples of biocatalytic steps which have been implemented in industrial production are given in an excellent review by Wu *et al.*[93] Conceptually, applications of

biocatalysis fall into two fields: use of enzymes in reactions vessels, i.e. *in vitro*, or in living cells, i.e. *in vivo.*

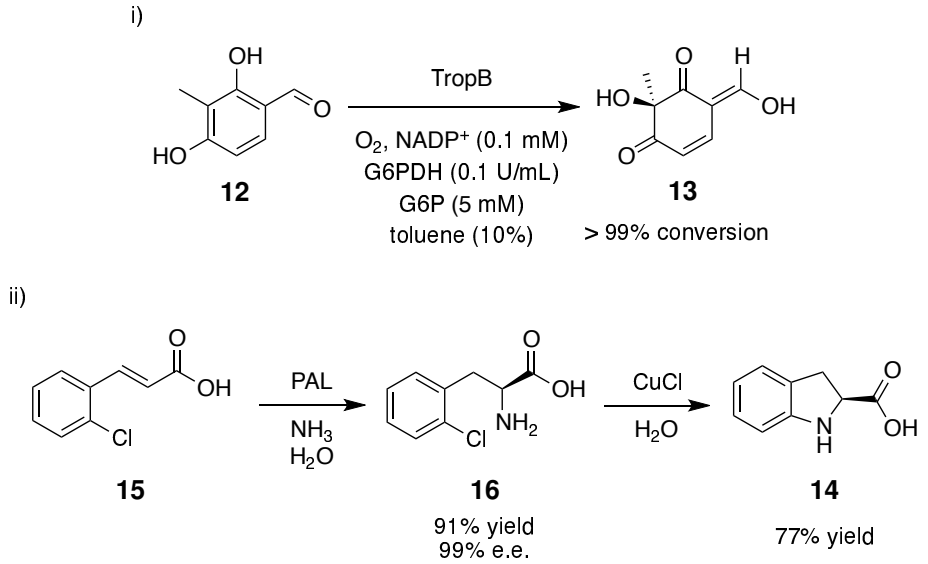
For use *in vitro*, the DNA or protein sequence of an enzyme is determined, and the genomic or synthetic gene is expressed by a suitable host. How the enzyme is then prepared depends on the needs of the application. The least labour-intensive preparation is as whole cells. The host cells, which may be lyophilised for longevity, are added directly to the reaction.[94] This requires that the reactants are able to reach the enzyme, either by crossing cell membrane or because the cells are disrupted as a result of the reaction conditions. Preparation as whole cells also gives limited control over the concentration of enzyme, and side-reactions may occur with the other contents of the cytoplasm. However, if expression can be standardised and if side-reactions are not a concern, whole-cell preparation requires very little time, labor and additional equipment.[94] This has made it a favourite of industry, where efficiency at scale is a chief concern.[95,96]

Alternatively, the host cells can be lysed before the reaction, and the resulting cell lysate containing the enzyme used as the reagent. A number of lysis methods are available, and the one chosen must take account of the downstream application and the resilience of the enzyme.[97] Homogenisation and bead milling both use mechanical force to break open cells, which is effective but can generate unwanted heat. Cells in suspension can also be ruptured by ultrasonication, where microscopic cavitation causes high local pressure changes. This also generates heat, however, and the break down of unwanted cell components may be so thorough that the aqueous lysate is difficult to clarify, i.e. to separate from cell debris by filtration or centrifugation.[97] Cell walls and membranes can also be disrupted by chemical agents such as detergents, or digested by specific enzymes to release the contents. However, care must be taken to ensure that the agents do not interfere with the activity of the enzyme. Use of cell lysate preparations still risks side-reactions with compounds in the host cytosol. However, it obviates the need for the substrate and/or product to cross the membrane of whole cells.

Finally, enzymes might be purified before being added to the reaction as a solution, lyophilised solid or precipitate (with ammonium sulfate). This affords full control over the concentration and eliminates any side-reactions from the host cell cytosol. However, the time and labour required makes purified proteins essentially unviable in many large-

scale applications, and their use is essentially reserved for investigative work, or for fine chemical production where purity is of utmost importance.[94] The stripping away of the cellular environment can also reduce the enzyme's longevity, further reducing the catalyst efficiency. An exception to this are immobilised enzymes, of which *Candida antarctica* Lipase B is an often-cited example.[98] These proteins are purified from the host cells, then covalently attached to the surface of beads. This can both help stabilise them, and allow them to be recovered and reused several times. However, not every enzyme is compatible with this process.

An example of *in vitro* biocatalysis which illustrates the difference between enzyme preparations was published by Baker Dockrey *et al.*.[99] In their study, whole-cells containing the monooxygenase TropB were used to convert phenol **12** to chiral dienone **13**, with full conversion at 1 g scale (Scheme 1.3i). They noted that purification of the enzyme resulted in approximately 10x lower yield of usable catalyst, potentially due to loss during purification, while crude cell lysates gave lower conversion than whole cells (although a specific value was not reported). Furthermore, avoiding purification saved six additional hours of preparation.



Scheme 1.3: Examples of *in vitro* biocatalysis. PAL = phenylalanine ammonia lyase.

This reaction also used toluene as a co-solvent to improve the conversion, which points to another useful property of *in vitro* biocatalysis: the tolerance for solvent conditions that might harm *in vivo* systems, which in turn enables higher loading of substrates

with poor aqueous solubility. However, it should be noted that ionic liquids and deep eutectic solvents, both compatible with living cells, show promise in bridging that gap between the two methodologies.[100]

With optimisation, *in vitro* biocatalysis can be implemented at the highest scale with considerable improvement over the equivalent synthetic process. In a study aiming to produce (*S*)-2-indolinecarboxylic acid **14**, de Lange *et al.* used phenylalanine ammonia lyase (PAL) to catalyse a stereospecific addition, transforming cinnamate **15** to amino acid **16** (Scheme 1.3ii).[95] By manipulating the equilibrium with excess ammonia, and finding a PAL variants that could tolerate such conditions, 91% yield and 99% *ee* were achieved for that step. The process was subsequently adopted by DSM Pharma Chemicals and implemented at ton scale. A life-cycle analysis showed a more than 50% reduction in the carbon footprint of this enzymatic process compared to the reactions it replaced, as well as greatly reducing the need for organic solvents due to the use of water for both steps.[95]

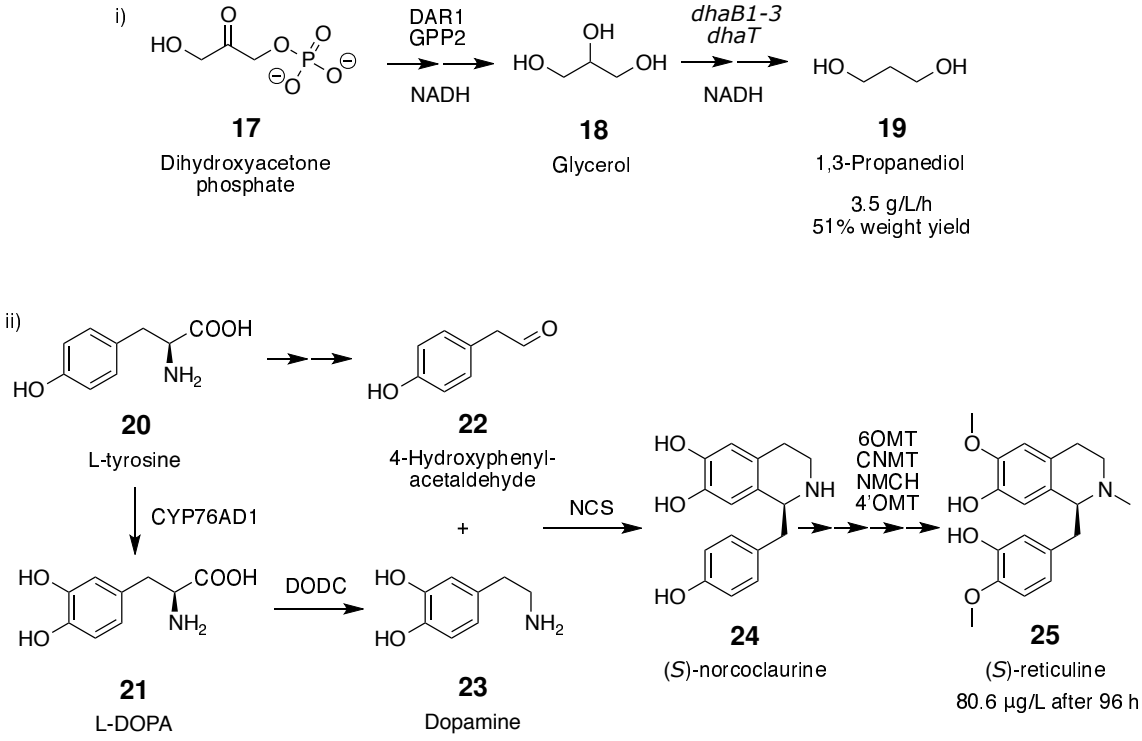
Whichever preparation is chosen, *in vitro* reactions offer control over the biocatalytic process. Relative proportions of the reactants can be tuned to optimise yields, and the isolated enzymes or dead cells might be able to tolerate conditions (e.g. high temperatures or substrate concentrations) that the living cellular system could not. However, any cofactors the enzyme needs will have to be supplied as well, at potentially great expense. Furthermore, the efficiency of the enzyme might drop once it is removed from a cellular environment to which it may have been specifically localised, such as the periplasm or within lysosomes.

The alternative, *in vivo* reactions, are based inside a host which is often genetically engineered. The organism expresses the relevant enzyme(s), while the starting materials may be added to the culture or may come from the organism's own metabolism.[101] The product is generated within the cell, where it may stay or diffuse or be exported into the culture medium. The benefits of *in vivo* biocatalysis directly counter the disadvantages of *in vitro*. The platform can supply any native cofactors, and the enzymes can operate in a more familiar environment, while taking advantage of cellular systems that mend and replace proteins. However, if the product is not membrane-soluble, it will have to be extracted from a cell lysate, which itself will be rich in unwanted soluble compounds. The concentration of enzymes in the pathway also cannot be controlled with any preci-

sion. Moreover, if the system relies on the host's own metabolism, the metabolic flux of starting materials may need to be engineered towards the pathway.[102] This introduces competition between the needs of life and the biocatalytic process, which necessarily limits the yields of these reactions.[101] Nonetheless, there have been successes. An *in vivo* system has been developed in which dihydroxyacetone phosphate **17** (DHAP) is transformed into glycerol **18** and subsequently the bulk chemical 1,3-propandiol **19** with 3.5 g/L/h space-time yield (Scheme 1.4i).[103] Platforms have also been built to produce fine chemicals, such as medicinal alkaloids (Scheme 1.4ii).[104] In one example, L-tyrosine **20** was converted into both L-dopa **21** and 4-hydroxyphenylacetaldehyde (4-HPAA) **22**. Compound **21** was subsequently converted into dopamine **23**, which reacted with **22** to give (*S*)-norcoclaurine **24**. A series of further oxidations and methylations then produced (*S*)-reticuline **25**. Altogether, *in vivo* biocatalysis offers a near-natural conditions for enzymes with specific needs, at the cost of having to consider many other factors.

Both strategies have limitations, which are joined by the more general challenges associated with enzymes. Establishing facilities for biocatalysis can be expensive, and the business case for replacing an established, profitable chemical process with a unfamiliar one can appear weak to manufacturers.[93] Proteins are also often unstable and difficult to recover from the reaction mixture. While enzyme immobilisation can address some problems,[98] that too can contribute to the costs. Low catalytic efficiency also remains a barrier to expanding biocatalysis, especially when the natural product of the enzyme only occurs natively in very small amounts.

Protein engineering has thus been a necessary companion to biocatalysis. Engineering methods can be broadly categorised into rational design and directed evolution. Rational design uses information from the structure of a protein to inform sequence changes that will give a desired effect. This approach has been bolstered by computational analyses, which allow the mechanics of reactions in the active site to be dissected,[105] and which give insight into how mutations might affect substrate binding.[106] Directed evolution, on the other hand, pairs random mutation with a 'selection pressure' in the form of an assay for the desired characteristic. This allows researchers to find beneficial mutations in the absence of structural information.[107] Directed evolution is not as immediately accessible as rational design, however, because the number of mutants which must be



Scheme 1.4: Examples of *in vivo* biocatalysis. i) Production of 1,3-propandiol **19**. DAR1 = glycerol 3-phosphate dehydrogenase, GPP2 = glycerol 3-phosphate phosphatase, *dhaB1-3* = glycerol dehydratase, *dhaT* = 1,3-propanediol oxidoreductase. ii) Production of (*S*)-reticuline **25**. DODC = DOPA decarboxylase, NCS = norcoclaurine synthase, 6OMT = 6-*O*-MT, CNMT = coclaurine-*N*-MT, NMCH = *N*-methylcoclaurine hydroxylase, 4'OMT = 4'-*O*-MT.

screened increases exponentially with the number of positions explored. Furthermore, the process can generate trade-offs, wherein the primary desired characteristic, such as activity towards one substrate, comes at the cost of a secondary characteristic, such as stability or broader substrate tolerance. Studies have used successive rounds of evolution to correct these unwanted consequences,[108] but this adds to the length and expense of the process.

Fortunately, the declining cost of synthetic genes has increased access to both techniques. With that, the number and variety of biocatalysts which can be developed is increasing. This has included, relatively recently, methyltransferases.[91,109] However, the dependence of MTs on SAM introduces a further hurdle, which must be confronted before members of the enzyme class can be used more widely as biocatalysts.

1.4.2 Supply of cofactors with MAT enzymes

The use of SAM-dependant MTs as biocatalysts requires a source of that cofactor. In *in vivo* platforms, SAM is generated within the cells by the cycle given above. For *in vitro* reactions, however, the cofactor must be supplied from an outside source, which is impracticaly expensive for large reactions.[20] For reference, at the time of writing, the price of SAM was nearly £ 500/g. Therefore, methods have been developed to generate SAM *in situ* from cheaper components.

As detailed in Section 1.2.2, SAM is produced in cells by a methionine adenosyltransferase (MAT). This enzyme uses the chemical energy stored in ATP to create the reactive methyl group, which is then attacked by the substrate. The two substrates, ATP and L-methionine, are more stable and less costly than SAM. Therefore, MATs have also been explored as biocatalysts to supply cofactor to MTs as part of enzyme cascades.

A consequence of MAT being essential for life is that almost every species has a gene for one. The only known exceptions are parasites, in which the gene has been lost or degraded, perhaps due to SAM supply from the host.[110] Unlike enzymes from niche secondary metabolism pathways, there is thus a cornucopia of existing variants to search for promising biocatalysts. MATs of thermophilic archaea have received particular attention, because enzymes from those organisms tend to be more stable at room temperature than mesophilic counterparts. Archaeal MATs are also sequence-divergent from

their eukaryotic and prokaryotic homologues, potentially indidicating unique features.[111]

The first archaeal MAT to be characterised was from *Sulfolobus solfataricus* (*Ss*), a thermoacidophilic archaeum with an optimal growth temperature of 87 °C.[112] The enzyme from this organism was notable, because it was hypothesised that the instability of SAM at such temperatures might necessitate an MAT with unusually high turnover.[112] Two isoforms, A and B, were partially purified and characterised. Both were highly thermophilic (i.e. high optimum temperature), but not thermostable, although stability was increased by pre-incubation with ATP.[112] Crystal structures of *Ss*MAT were published later.[111] The overall topology was comparable with *Ec*MAT, as were some key active site residues. However, *Ss*MAT does not need potassium ions to stimulate activity, with the side chain of a lysine residue appearing to perform this role instead.[111]

The MAT from another thermophilic archaeon, *Methanoccocus janaschii* (*Mj*), was the second to be identified and characterised. Like *Ss*MAT, its sequence and optimum temperature differed greatly from *Ec*MAT.[113] A later study on *Mj*MAT highlighted that the subunit association found with many MATs may be the cause of their stability.[114] Two further thermophilic archaeal MATs have since been characterised, from *Thermococcus kodakarensis* (*Tk*)[115] and *Pyrococcus furiosus* (*Pf*).[116] Both show similar properties, with high heat tolerance and a dependence on Mg^{2+} for activity. The crystal structure of *Tk*MAT has also been solved, indicating a typical three-domain architecture with two active sites in the functional dimer.[115] *Pf*MAT is homologous to the other archeael MATs, and like *Ss*MAT is not affected by the presence of K^+ ions. It also shows a lower K_M for methionine and ATP, and a higher k_{cat}, making it one of the most active MAT enzymes at its optimum temperature.

Although not adapted to extreme high temperatures (yet), the MAT from *Homo sapiens* (*Hs*) has also been the subject of detailed study. This is mostly owing to the connection between SAM and a number of health conditions including liver disease and cancer.[117] As a result, its structure has been characterised for research and the development of small-molecule inhibitors.[118,119] There are two catalytic MATs in *H. sapiens*, 1A and 2A, with the former expressed in the liver and the latter expressed everywhere else.[120] *Hs*MAT2A shares the same key structural features with the archaeal and bacterial MATs discussed so far, being a dimer of approximately triangular subunits with two active sites

at the interface.[120] The enzyme is also dependant on Mg^{2+} and K^{+} ions for full activity.

The phylogenetic tree in Figure 1.5 shows the relationship between the six MAT enzymes mentioned so far, and illustrates the archaeal homologues sit separate from both the bacterial and eukaryotic variants.

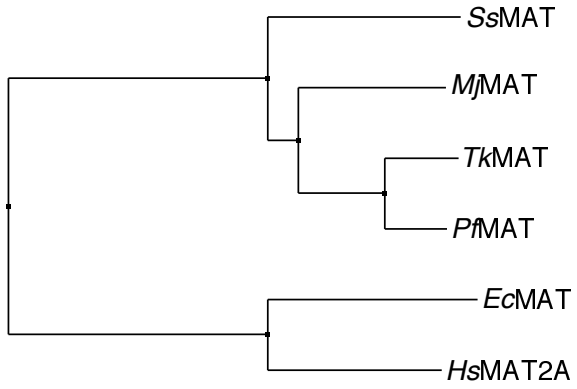


Figure 1.5: Phylogenetic relationship between a number of characterised MAT enzymes. Produced using Jalview 2.[121]

1.4.3 Beyond methylation: alternative alkylations with enzymes

Following characterisation, a number of these MATs have been used experimentally to generate SAM, either alone or as part of cascades with MTs. What is especially noteworthy, however, is that many have shown a tolerance for analogues of methionine. By accepting these as substrates, the MATs can generate the cognate cofactor analogues, which bear the new alkyl group in place of the reactive methyl. This opens the opportunity for MTs to transfer not just methyl groups, but conceivably any group from any accepted cofactor analogue. A figure summarising publications that have used MATs to generate SAM analogues is given in Figure 1.6.

Wang *et al.* demonstrated the acceptance of 31 methionine analogues by *Ss*MAT. These included selenomethionine analogues, which were shown to give higher yields due to the relative weakness of the Se-C bond. Furthermore, they were able to determine a crystal structure of the enzyme bound to the *S*-ethyl cofactor analogue.[111] With this structure, they linked the greater substrate tolerance of *Ss*MAT, relative to *Ec*MAT, to an isoleucine near the alkyl group in place of a leucine.[111]

During the first characterisation of *Mj*MAT, Lu *et al.* showed that the enzyme was

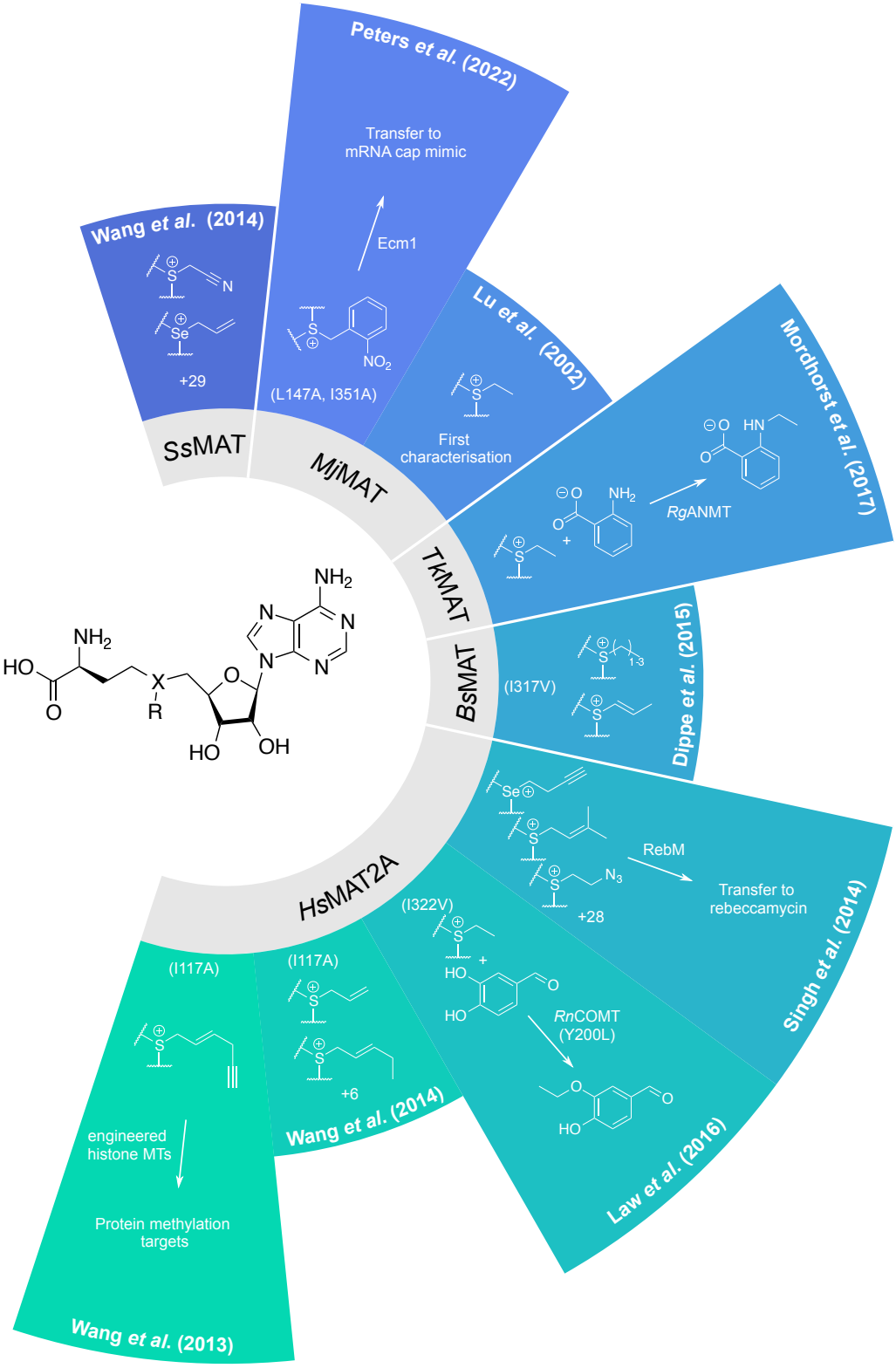


Figure 1.6: Examples of SAM analogue generation via MATs in the literature. On the central structure, 'X' represents S or Se, while 'R' represents any alkyl group.

able to accept the methionine analogue L-ethionine.[113] More recently, Peters *et al.* used a double-mutant to create a cofactor analogue bearing a photolabile group, which was subsequently transferred to a mimic of the mRNA 5' cap by the MT Ecm1.[122] This could be applied in studies of living cells to change the caps from inaccessible to accessible with a flash of light, allowing the consequences to be studied in real time.[122] Meanwhile, as part of a wider investigation into MAT-MT systems, Mordhorst *et al.* used *Tk*MAT in cascade with *N*-methyltransferase *Rg*ANMT to ethylate the 2-aminobenzoic acid using L-ethionine.[123]

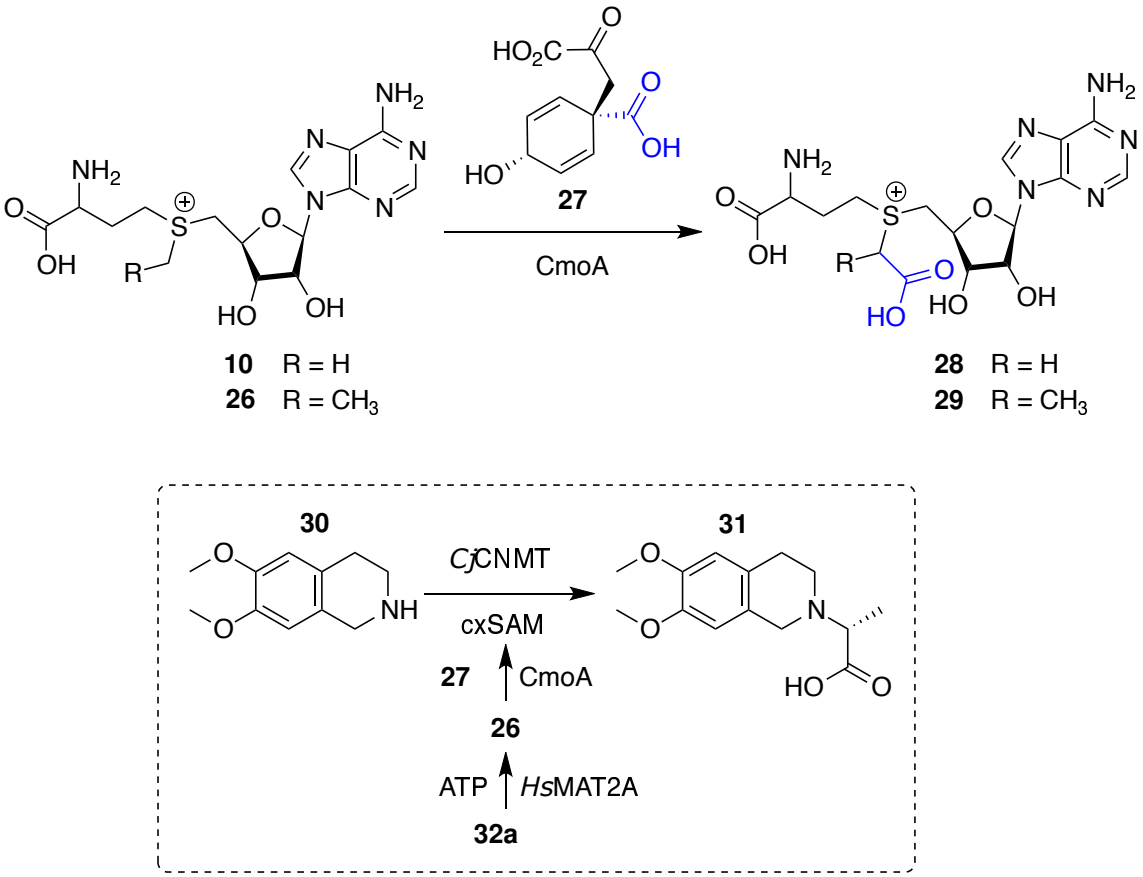
The MAT from a mesophilic organism, *Bacillus subtilis*, has also been investigated by Dippe *et al.*[124] The wild-type enzyme was shown to have low acceptance for methionine analogues with longer alkyl chains. However, the substitution I317V was found to be permissive for *S*-ethyl, -propyl, -butyl and -prop-1-enyl chains.

Singh *et al.* showed that the human MAT2A can also accept a huge variety of methionine analogues.[125] As with the investigation of Wang *et al.* with *Ss*MAT, selenomethionine analogues gave greater conversions.[111] Four pairs of methionine/selenomethionine analogues were also integrated into cascades for alkylation of rebeccamycin by MT RebM. This study also documented more extensively what had been indicated before, that *Ec*MAT is intolerant of nearly all methionine analogues. Later, the I332V variant of *Hs*MAT2A was show by Law *et al.* to accept several methionine analogues. The *S*-ethyl cofactor was generated in cascade with a regioselective variant of the catechol-*O*-MT from *Rattus norvegicus* (*Rn*COMT) to give ethyl-vanillin, a high-value flavour compound.[126] A different variant, I117A, was created by Wang *et al.* with the aim of enlargening the hydrophobic pocket around the cofactor *S*-alkyl group. This variant was shown to accept eight methionine analogues carrying long unsaturated chains. In a separate study, they demonstrated transfer of an alkyne-bearing group by engineered histone MTs to their targets, allowing sections of DNA proximal to specific histone methylations to be pulled down and sequenced.[127,128]

This list of reports is expected to grow as protein engineering allows the tailoring of cofactor supply systems towards a desired application. Though not relevant to alkylation, other work has been done towards generating SAM analogues which vary at a other locations. These include engineering of *Ec*MAT[129] and *Mj*MAT[130] to accept ATP analogues

with alternate purine moieties, and of *Hs*MAT2A to accept bioorthogonal amino-alcohol methionine analogues.[131]

MATs are, however, not the only enzymes capable of creating SAM analogues. In work by Herbert *et al.*, the enzyme CmoA was used to generate carboxy-SAM **28** and carboxy-SAE **29** from SAM **10** and SAE **26**, respectively, using the cofactor prephenate **27** (Scheme 1.5).[132] The latter of these was integrated into a three-step cascade, wherein *Hs*MAT2A(I332V) converted L-ethionine **32a** into SAE **26** which CmoA transformed into carboxy-SAE **29**, whereupon coclaurine-*N*-methyltransferase from *Coptis japonica* (*Cj*CNMT) transferred the carboxy-ethyl group to THIQ substrate **30** (giving **31**). The conversion of the last step was not reported, but this case demonstrates that the diversity of alkylations is not even limited by the tolerance of MATs, and other enzymes may be able to modify cofactor analogues yet further.



Scheme 1.5: Generation and use of cxSAM and cxSAE, utilising the enzyme CmoA alongisde *Hs*MAT2A and *Cj*CNMT.

Another alternative, SalL, was first isolated from the marine bacterium *Salinispora*

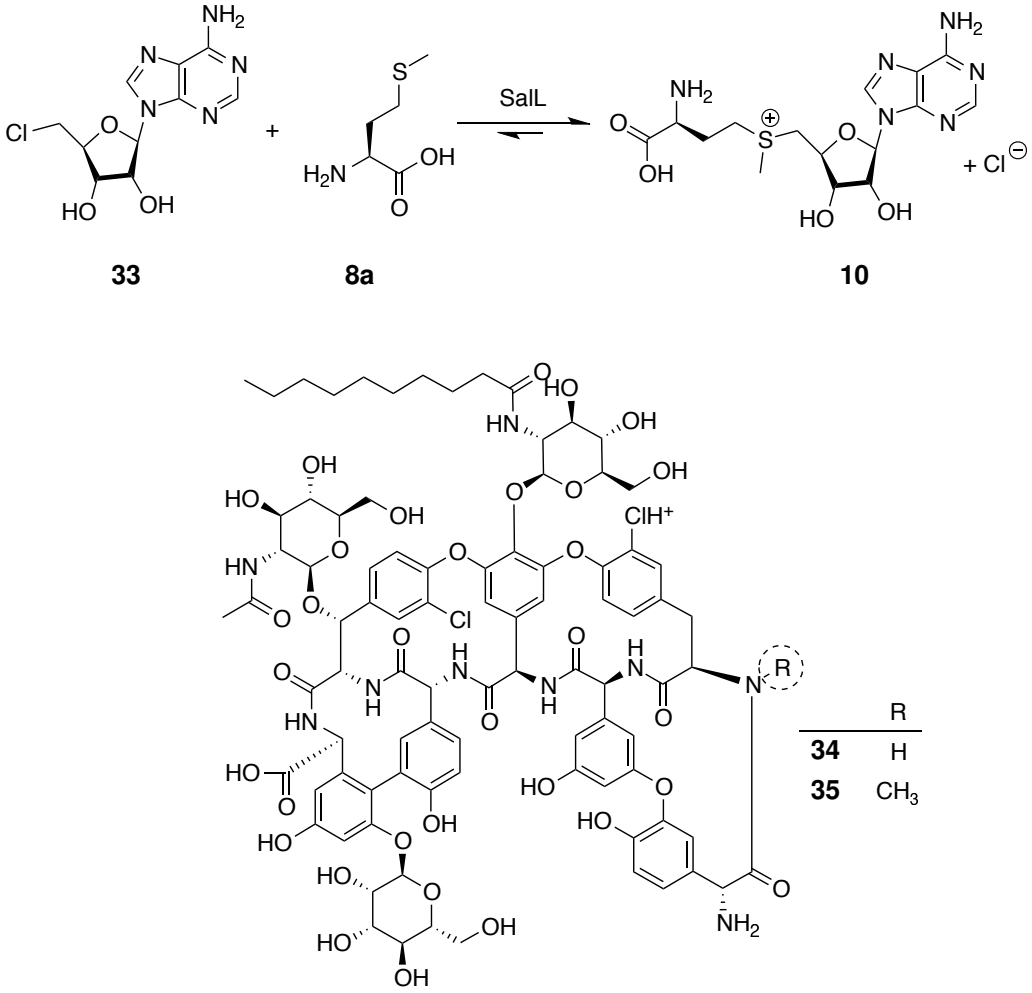
tropica, where it catalyses replacement of the SAM methionine moiety with a halogen.[133] In the original organism, a chain of other enzymes deconstruct this product to give numerous halogenated carbon skeleton fragments, which are then incorporated into natural products. However, Lipson *et al.* demonstrated that, outside of the chloride-rich environment of the ocean, SalL catalyses the opposite reaction, reacting 5'-chloro-5'-deoxyadenosine (ClDA) **33** and an excess of L-methionine **8a** to give SAM **10** (Scheme 1.6).[134] They went on to couple this reaction to methylations, first of DNA using HhaI, then of the natural product teicoplanin **34** with the *N*-MT MtfA. The latter of these cascades gave 74% conversion of **34** to methyl-teicoplanin **35**. The group also conducted a protein engineering study to broaden the substrate scope of SalL. Even the wild-type enzyme was shown to accept ethyl, propyl, butyl, allyl and benzyl methionine analogues, and some mutants increased activity with the L-ethionine by more than 50%.[135,136] Given that **33** can be easily produced in a reaction between adenosine and $SOCl_2$, this method of SAM supply may be at least as viable as MAT-based methods.

What was therefore initially a problem for MT biocatalysis, the supply of expensive cofactor, has given rise to a body of work showing that an impressive range of cofactor analogues can be generated by both mesophilic and thermophilic MATs. However, in general, the larger the alkyl group, the less efficient the reaction seems to become with wild-type enzymes. Yet the common architecture of these proteins suggests that what has been learned from engineering of one homologue may be translatable to others, hinting at a future possibility were a hybrid MAT can combine all advantages.

1.4.4 Alleviation of negative feedback with MTAN enzymes

There is one final obstacle to efficient MT biocatalysis. *S*-adenosyl homocysteine (SAH) is the immediate by-product of methylation, and competitively inhibits MTs via the SAM binding site.[137] As shown above in Scheme 1.1, SAH is hydrolysed in cells either by SAHH or MTAN. MTAN has a dual specificity, able to cleave both SAH **36** and *5'*-methylthioadenosine **37**, with an approximately 2.5 fold rate for the latter (Scheme 1.7).[138]

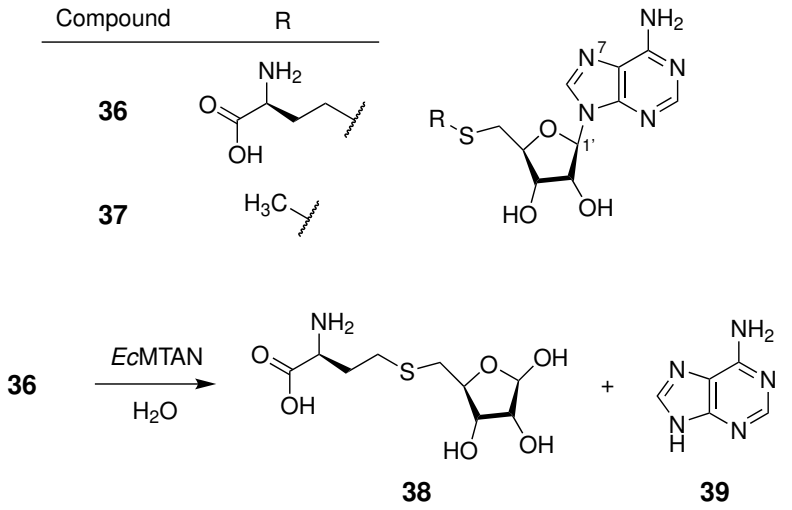
Both enzymes could relieve the negative inhibition of MTs *in vitro*, but the reaction of SAHH is reversible, and the equilibrium in native conditions lies heavily towards SAH. Efficient removal therefore relies on the product of cleavage being drawn into subsequent



Scheme 1.6: SalL-catalysed generation of SAM **10**, coupled to methylation of teicoplanin **34**.

metabolic pathways. The hydrolysis of MTAN, however, is irreversible, so for simple disposal of the unwanted side-product would seem to be preferable. Nonetheless, both have been used at the terminus of MT cascades.[123,125]

The first investigation into the *Ec*MTAN catalytic mechanism was an inhibition study, which deduced a 1' oxocarbenium transition state aided by protonation of the adenine N7.[139] When the crystal structure was later solved, the enzyme was found to feature a twisted α/β topology with a hydrophobic surface for dimerisation.[140] The study also highlighted the role of structural water molecules in catalysis, and implicated D198 as the proton donor to the adenine N7 (Figure 1.7). The presence of a nucleophilic water attacking position 1' of the ribose was later confirmed, first by mass spectrometry[141] and then by a second crystal structure with a transition state analogue[142] (Figure 1.7). Although the mechanism has not been published in detail, the key step is therefore thought



Scheme 1.7: Natural substrates of MTAN and general reaction scheme of hydrolysis.

to involve cleavage of the ribose-adenine bond by water, giving, in the case of SAH, *S*-ribosyl-L-homocysteine **38** and adenine **39**.

The full enzyme cascade to enable scalable MT biocatalysis therefore consists, so far, of three core components: an enzyme to supply the cofactor from relatively low-cost starting materials, an MT to transfer the reactive alkyl group to a target and an MTAN to remove the inhibitory side product. The choice of exactly which MT to use is informed by the desired alkylation substrate. For this project, that in turn was linked to the existing biocatalytic work of the research group.

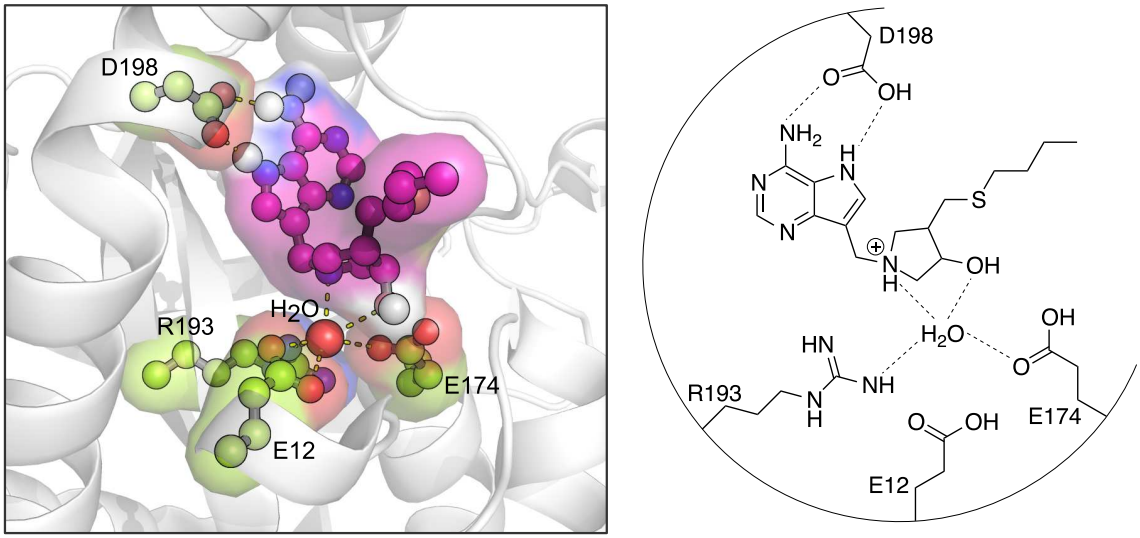


Figure 1.7: Active site of methylthioadenosine nucleosidase (MTAN) from *E. coli*. In 3D representation, SAH mimic shown in magenta, important residues highlighted in green, structural water molecule shown as red sphere, polar contacts represented as dashed lines. The side chains of R193, E12 and E174 are all in range to have polar contacts with the structural water - only two are shown in order to not exceed the H-bonding capacity of the molecule.

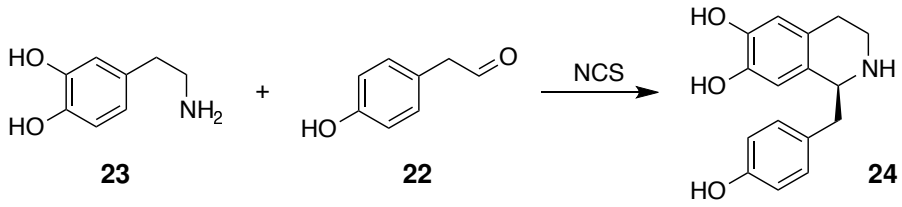
1.5 Research within the group

1.5.1 Uses and creation of THIQs

A focus of prior work in the group had been finding enzymatic methods to generate benzylisoquinolines and their derivatives, particularly tetrahydroisoquinolines (THIQs). Alongside their current applications as analgesics[143] and antitussives,[144] both natural and non-natural THIQs show promise against diabetes,[145,146] cancer[145,147,148] and neurological conditions.[149] However, synthesising these molecules with the desired chirality is difficult.

In nature, specifically some plants, the THIQ scaffold is created by the enzyme norcoclaurine synthase (NCS). NCS catalyses a stereospecific, Picter-Spengler condensation between dopamine **23** and 4-HPAA **22** to give (*S*)-norcoclaurine **24** (Scheme 1.8i).

The protein's structure features a seven-strand, antiparallel β-sheet forming a saddle that encloses two main alpha helices (Figure 1.8A). The active site rests between the sheet and the helices, with the major catalytic residues being Y108, E110, K122 and D141 (Figure 1.8B).[150] The catalytic mechanism begins with binding of dopamine to K122



Scheme 1.8: Reaction natively catalysed by NCS.

by a catechol hydroxyl, and to D141 and E110 by the charged nitrogen (Scheme 1.9). The aldehyde subsequently binds via Y108 and E110 and is attached by the dopamine nitrogen. This results in an intermediate aminol from which water is eliminated. Cyclisation of the resulting iminium is catalysed primarily by K122, and is the step at which the C-1 chirality of the product is decided. Deprotonation by E110, and reversal of the electron transfer chain back through to K122, gives the final THIQ product.[151]

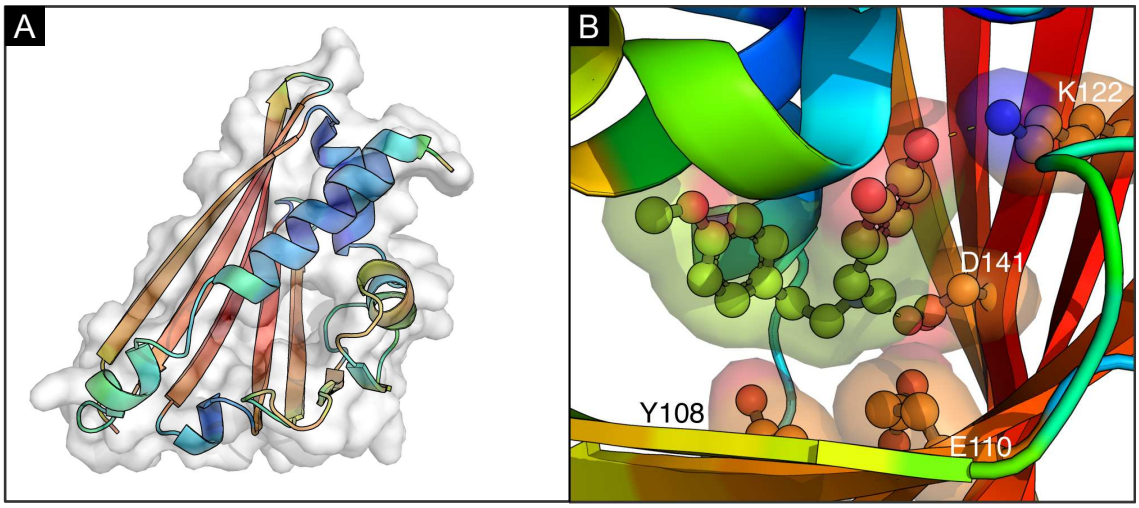
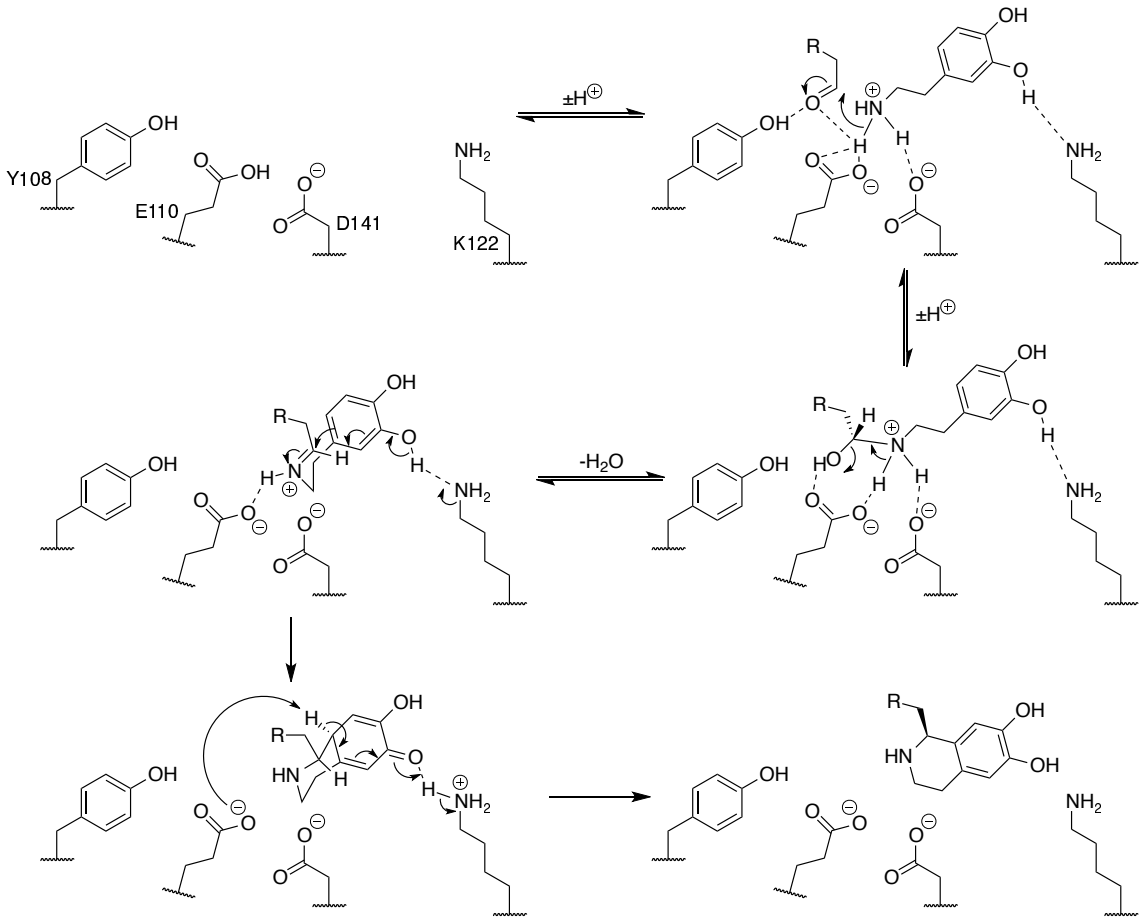


Figure 1.8: Structure and active site of *Tf*NCS. A) Monomer topology. B) Active site with a reaction intermediate mimic. Polar contacts shown as dashed lines. PDB: 5NON.

Research in the group has successfully built enzyme cascades using NCS from *Thalictrum flavum* (*Tf*), generating natural and novel structures with controlled C-1 stereochemistry (Scheme 1.10). Lichman *et al.* used the transaminase *Cv*TAm and pyruvate **40** to generate aldehyde **41** from dopamine **23** *in situ*. The amine and aldehyde could then be combined by *Tf*NCS to give (*S*)-norlaudanosoline **42** (Scheme 1.10ii).[151] They later showed that *Tf*NCS can also accept ketone substrates, e.g. **43**, to give 1,1-disubstituted THIQs, including the spiro compound **44** (Scheme 1.10iii).[152] Zhao *et al.* demonstrated a two-step synthesis, wherein *Tf*NCS-catalysed condensation of **23** and **45**, followed

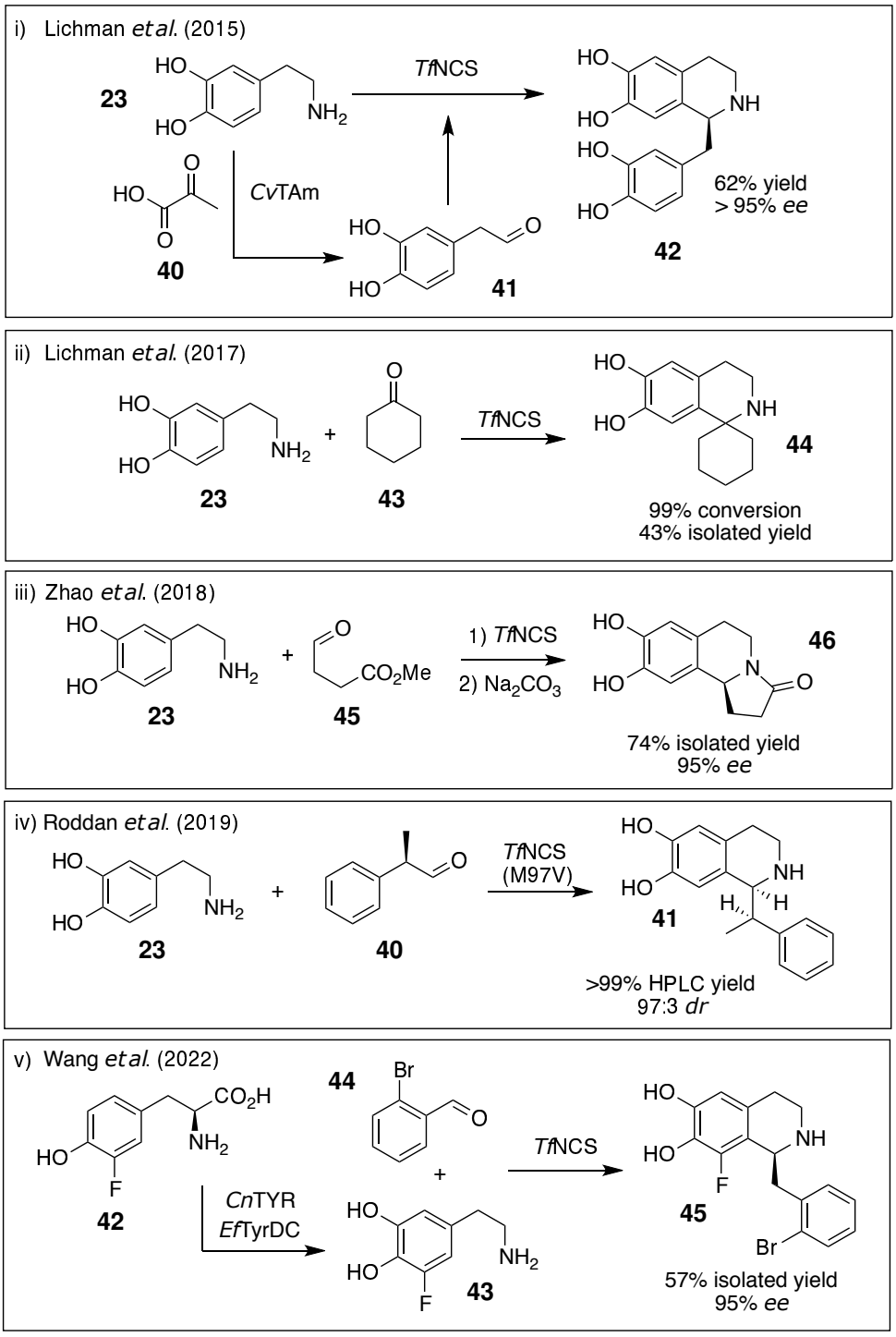


Scheme 1.9: Catalytic mechanism of norcoclaurine synthase. Adapted from Lichman *et al.* (2015), Figure 3.[151]

by treatment with base, generated trolline **46** (Scheme 1.10iv).[153] Roddan *et al.* demonstrated activity of NCS, both wild-type and mutated, towards α-substituted aldehydes (e.g. **47**), to give novel THIQs such as **48** with two defined stereocentres (Scheme 1.10vi).[154] Most recently, Wang *et al.* integrated tyrosinase (*Cn*TYR) and tyrosine decarboxylase (*Ef*TyrDC) enzymes into the cascades. With tyrosine analogues such as **49** as starting materials, they were able to generate the dopamine analogue **50**, which when combined with aldehyde **71** gave the fluorinated THIQ **52** (Scheme 1.10v).[155]

1.5.2 Methylation of THIQs

As explained at the beginning of this chapter, selective alkylation of complex molecules allows exploration of structures that might otherwise be difficult to access. The over-representation of the THIQ moiety in medicines suggests that the region of chemical



Scheme 1.10: Examples of THIQs generated in previous work.

space it occupies may be rich in bioactive compounds. Therefore, enzymatic alkylation of these structures could be a route to finding new and promising medicines.

The THIQ core made by the NCS cascades above contains at least three nucleophilic sites that could be MT targets: the 6' and 7' hydroxy groups, together forming a catechol,

and the secondary amine at position *N*-2. At the commencement of this project, several MTs with catechols as natural substrates were being investigated by the group for their activity towards THIQs. Of these, the generically-named catechol-*O*-methyltransferase (COMT) was particularly notable, as mutants of this MT had been used in reported alkylation reactions (Figure 1.6).

Catechol *O*-MT (COMT) is one of many enzymes in the phenolic OMT family. It is an outlier among MTs in general, however, because of its role in catabolism, as opposed to anabolism or signalling. COMT is found in the brains of mammals, where it deactivates catecholamines such as dopamine, adrenaline and noradrenaline.[48] Histamine *N*-methyltransferase provides another example of this type, being a catabolic MT whose substrate also has a neurotransmitter-like function[49] Because of its role, COMT might thus be expected to have a wider substrate scope than MTs where tight adherence to one target is important for function.

The version of the COMT gene that has been used as a biocatalyst is derived from *Rattus norvegicus*. The reason for this is historical: the enzyme was first isolated over 60 years ago from the brains of rats, as part of studies into neural biochemistry. It has therefore been well-characterised, and conveniently expresses well in bacterial hosts.

A crystal structure of *Rn*COMT has been published, illuminating its topology and catalytic mechanism.[48] It has a class I structure, and the active site sits between αC and β4 and 5 (Figure 1.9A). There are no substrate-specifying residues in the binding region beyond those which interact with the catechol. The active site is near the surface, allowing the groups which extend away from the target group to either interact with the solvent, or lie in one of several grooves around the opening. SAM binds in a characteristic cofactor binding site involving the N-termini of β1 and 2 and the loop emanating from αD. In the active site, the non-methylating hydroxyl forms a hydrogen bond with E199, while both hydroxyls coordinate the buried Mg^{2+} ion (Figure 1.9B). These bonds orient and stabilise the catechol for the S_N2 reaction. The lone pair of K144 attacks the proton of the target hydroxyl group, while the substrate hydroxyl attacks the electrophilic carbon of the SAM methyl group (Scheme 1.11). These components are the basis of many *O*-methylations: residues and/or metal ions abstract electron density while a catalytic base encourages attack by the receiving atom. An MT involved in chalcone biosynthesis uses histidine

as the catalytic base via a structural water molecule,[105] while some bacterial MTs use a bicarbonate ion for deprotonation.[31] However, the overall mechanism demonstrated by *Rn*COMT is recognisable in other MTs.

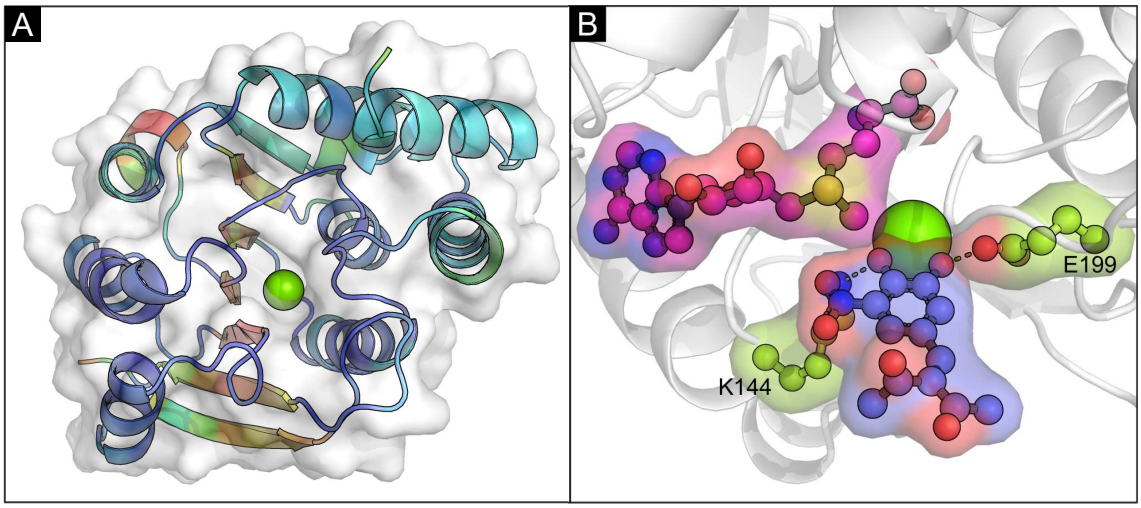
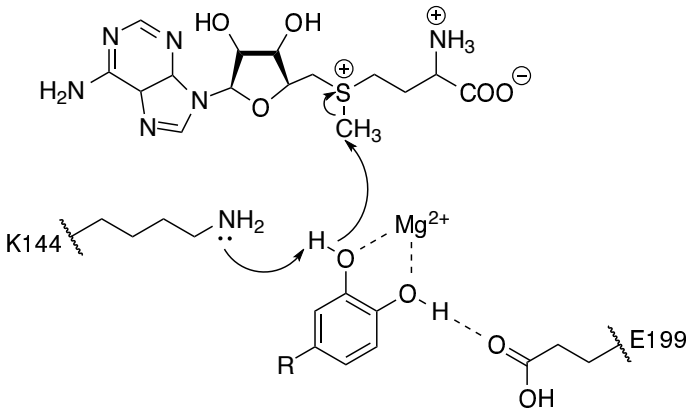


Figure 1.9: Structure and active site of COMT from *Rattus norvegicus*. A) Overall structure. Backbone represented in cartoon under a translucent surface. Mg^{2+} ion shown as green sphere. C) Active site containing an inhibitor. Polar contacts shown as dotted lines.



Scheme 1.11: Mechanism of methyl-transferral in the *Rn*COMT active site with a generic catechol substrate.

The lack of apparent substrate specificity implicated *Rn*COMT as a good biocatalyst for methylating THIQs, whose catechol groups could effectively slot into the narrow 'letterbox' of the active site. An aim of the group was therefore to link NCS cascades, and all the variety they could produce, to methylation cascades consisting of an MAT, catechol MTs and an MTAN. This project was thus seeded from the idea of bringing two biocat-

alytic routes together to generate a rich variety of products from simple starting materials in sustainable conditions.

1.6 Research question and hypothesis

The central objective of this project was thus to explore and expand the capacity of MTs to edit medicinally relevant scaffolds, with scalability being a key consideration. The hypothesis underlying all work towards this objective was that the usefulness of MTs as biocatalysts is a function of their structure. By finding MTs with favourable structures, or creating those structures by design, the characteristics of stability, turnover, substrate scope, cofactor analogue tolerance and selectivity could be improved.

The first aim of this project was characterise the selectivity of an established MT system, based on *Rn*COMT, towards a panel of THIQ substrates. As their most attractive advantage over organic synthesis, selectivity is a integral competency of enzymes. Having detailed knowledge of structure-selectivity relationships was therefore vital for promoting this MT system as a viable biocatalytic strategy for a given substrate group. This work was collaborative, involving a number of members of the group and resulted in publication of our findings.[156]

The central objective then became to explore MT capabilities in two directions. The first was an attempt to develop novel MTs with useful substrate specificities. Candidate enzymes were identified through database searching, and molecular biology techniques used to manipulate the genes and optimise their expression. Bespoke assays were then determine the viability of the candidates as biocatalysts. The focus of this search was on *O*-MTs, specifically, including those targetting phenols, aliphatic hydroxyls and hexoses.

The second direction involved focusing on the more established enzyme *Rn*COMT, but developing a capacity for alternative alkylation with respect to THIQs. As described above, it has been demonstrated that MTs can transfer other chemical groups onto their targets if A) supplied with the cognate cofactor, and B) they have tolerance for that cofactor. This work aimed to meet both these requirements.

Synthetic strategies towards analogues of L-methionine were also sought, with the goal being a generic, low-cost synthesis to produce a variety of analogues. Those which

were the most useful, informative or chemically favourable for the transfer process could then be generated and tested at a larger scale.

Meanwhile, the natural tolerance of the three-enzyme system for such analogues would be tested through assays, analysed predominantly by analytical high-performance liquid chromatography (HPLC). Observations of this natural tolerance would be rationalised by computational techniques, which could generate hypotheses around the engineering of the enzymes for more favourable characteristics. Site-directed mutagenesis could then be used to implement these changes, followed by random mutagenesis if no productive variants were generated. For the latter, a high-throughput assay would need to be developed to screen large numbers of mutants in relatively little time.

Ultimately, the aim was to conduct preparative-scale biocatalytic reactions. These would allow thorough characterisation of the products, and demonstrate the practicality of the enzyme system at a larger scale.

Chapter 2

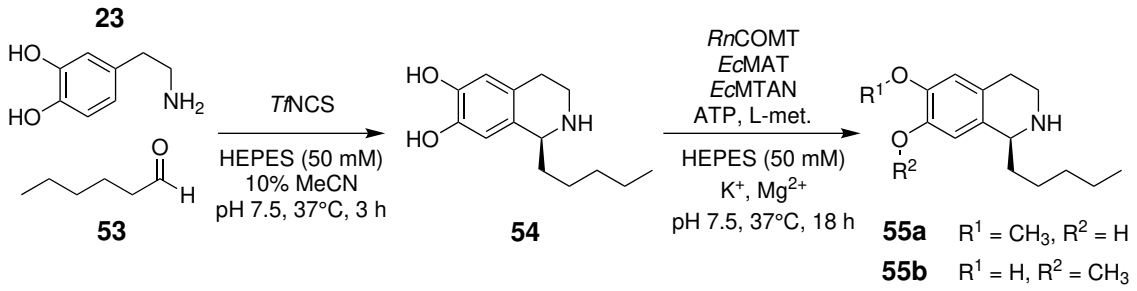
Methylation

2.1 Regioselectivity of *Rn*COMT towards THIQ compound (*S*)-54

At the outset of the project, an aim within the group was to understand the activity of several catechol MTs towards THIQ substrates. It was observed by Dr. F. Subrizi that changing the substituent on the THIQ core affected both the amount and regioselectivity of methylation. Hypotheses were formed around patterns in these relationships, and addressing them required the methylation products of a panel of substrates to be characterised. To develop expertise in the handling of enzymes, this project first explored reactions involving two THIQ substrates on that panel.

For the first substrate, a two-step, one-pot reaction was planned. In the first step, *Tf*NCS would catalyse the condensation of dopamine **23** and hexanal **53**, producing (*S*)-1-pentyl-1,2,3,4-tetrahydroisoquinoline-6,7-diol **(*S*)-54** (Scheme 2.1). The selective creation of (*S*)-THIQs by NCS is well established,[151,157] so this was considered the dominant enantiomer product. A three-enzyme system as described above, using *Rn*COMT, *Ec*MAT and *Ec*MTAN, would then be added to methylate **(*S*)-54**, giving **(*S*)-55** (Scheme 2.1).

Both reaction steps were first tried on a 0.025 mmol scale. The first step used purified *Tf*NCS at 0.5 mg/mL (Assay I. Full list of conditions in 8.3). HPLC analysis of the product showed disappearance of the peak corresponding to dopamine (RT 3.1 min) and appearance of a new peak (RT 7.8 min) (Figure 2.1). The retention time of this new peak matched that of synthetic standard, **(*RS*)-54**, which was synthesised by a phosphate-



Scheme 2.1: Enzyme cascade towards the methylation of (*S*)-1-pentyl-1,2,3,4-tetrahydroisoquinoline-6,7-diol **(*S*)-54**.

catalysed Pictet-Spengler reaction between **23** and **53** (Section 8.8). Therefore, this new peak was taken to be the product, **(*S*)-54**.

In the second step, methylation of this product used clarified lysates of *Rn*COMT (10% *v/v*), *Ec*MAT (10% *v/v*) and *Ec*MTAN (2% *v/v*) (Assay II). Analysis showed quantitative conversion to another new peak (RT 8.4 min) (Figure 2.1). Previous work with *Rn*COMT had indicated 6-OH selectivity, suggesting this peak represented the product **(*S*)-55a**. However, the peak featured a subtle leading shoulder, which might be explained by the presence of a minor 7'-methoxy regioisomer, **(*S*)-55b**. The product was purified by preparative HPLC and analysed by ^{1}H NMR spectroscopy to determine the location(s) of the methyl group. NOESY analysis showed correlation between the methyl ether signal and 5-H, but not 8-H, indicating only the 6-OH had been methylated (Supplementary figure E.1).

To address the disagreement between the NMR and HPLC observations, the cascade was repeated. This time, HPLC analysis of the methylation product showed no shoulder. The product was again purified (44% isolated yield) and analysed by ^{1}H NMR spectroscopy, which confirmed that only **(*S*)-55a** was present. The minor regioisomer **(*S*)-55b** could have been present at the first point of analysis but lost during purification, or deterioration of the HPLC column may have caused double-peaking that appeared like an isomer. The cascade was repeated a final time at 0.1 mmol scale, and analysis of the product corroborated the second cascade's results, with no shoulder appearing alongside the dominant methylation product. *Rn*COMT had therefore demonstrated excellent acceptance of **(*S*)-54** as a substrate, and seemingly full selectivity for the 6-OH position.

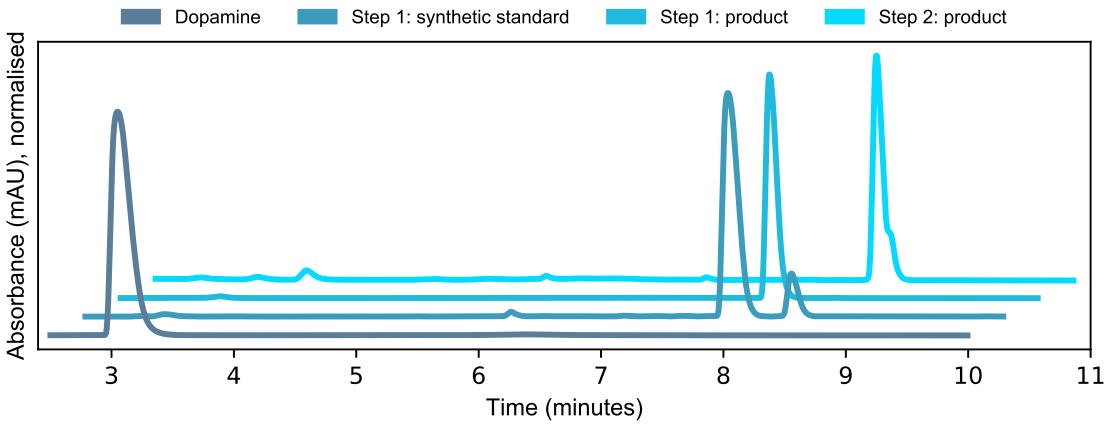
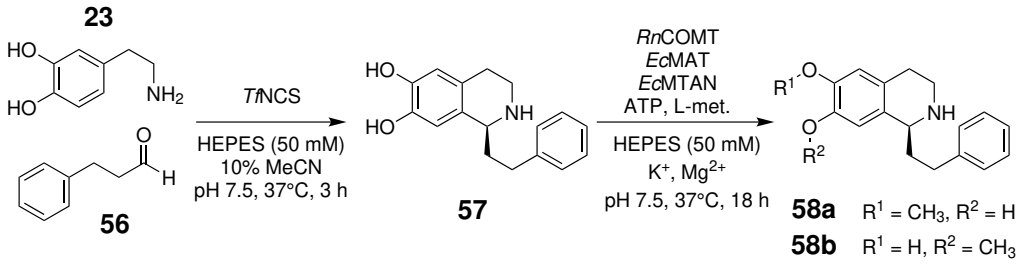


Figure 2.1: HPLC analysis of NCS-methylation cascade, demonstrating methylation of **(*S*)-54** to give **(*S*)-55a**. HPLC method A (see Section 8.4). Assay I and II for NCS and MT stages, respectively (see Section 8.3)

2.2 Regioselectivity of *Rn*COMT towards THIQ compound (*S*)-57

For the second THIQ substrate, a *Tf*NCS-catalysed step between **23** and **56** was planned to give (*S*)-1-phenethyl-1,2,3,4-tetrahydroiso-quinoline-6,7-diol **(*S*)-57**, which could subsequently be methylated to give **(*S*)-58** (Scheme 2.2).



Scheme 2.2: Enzyme cascade towards the methylation of (*S*)-1-phenethyl-1,2,3,4-tetrahydroisoquinoline-6,7-diol **(*S*)-57**.

However, while a small-scale trial of the first step showed formation of a new peak (RT 8.6 min, putatively **(*S*)-57**), substantial dopamine **23** remained (RT 3.4 min) (Figure 2.2). Note that the retention time of **23** had shifted since the traces shown in Figure 2.1 due to adjustment of the HPLC column. It was hypothesised that the incompleteness of the reaction was due to oxidation of **56**. Pulsed addition of the aldehyde over 2 hours was

attempted to ensure a fresh supply. Ultimately, however, simply doubling the equivalents of **56** with respect to **23** was sufficient to give complete consumption of starting material to a new peak (RT 8.5 min, Figure 2.2). The product of this reaction was fed into a methylation reaction with the same conditions as for **(*S*)-54**, which gave another new peak (RT 9.1 min) (Figure 2.2). This signal was taken to represent the methylated product, **(*S*)-58**.

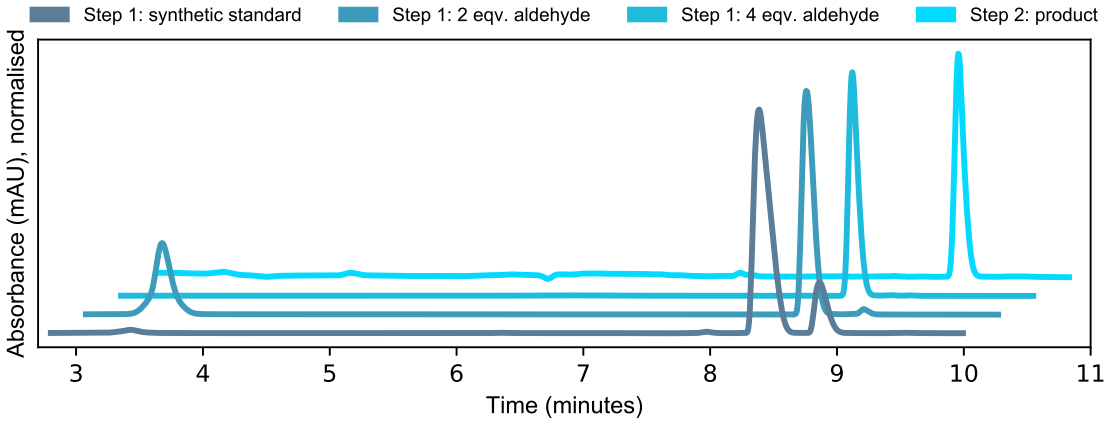


Figure 2.2: HPLC analysis of NCS-methylation cascade, demonstrating methylation of **(*S*)-57** to give **(*S*)-58a** (major) and **(*S*)-58b** (minor). Assay I and II conditions for NCS and MT stages, respectively. HPLC method A.

The two-step cascade to **(*S*)-58** then was attempted at a 0.025 mmol scale. HPLC analysis at both stages indicated full conversion from the starting materials. The product was isolated by preparative HPLC (33% isolated yield) and analysed by ^{1}H NMR spectroscopy. Two peaks were detected for the methyl ether signal, one showing a NOESY correlation to 5-H and the other to 8-H (Supplementary Figure E.2). This suggested there were in fact two regioisomers in the product, each methylated on one of the hydroxyls. The selectivity of *Rn*COMT was estimated by the ratio between the signals' integrals, giving an 86% *meta*-selectivity for this substrate.

Interactions between the C-1 substituents of **(*S*)-57** and **(*S*)-54** and the surface of *Rn*COMT might explain these regioselectivities. Two main grooves lead away from the active site of *Rn*COMT. One is narrower and more enclosed from the solvent, while the other is wider and more exposed (Figure 2.3). The less rigid and bulky C-1 substituent of **(*S*)-54** may fit better into the narrower groove, which would position the *meta*-hydroxyl

towards the SAM methyl group. Conversely, the bulkier substituent of **(*S*)-57** may more often lie in the wider channel, which would position it for *para*-methylation. More advanced techniques, such as protein NMR or X-ray crystallography, would be needed to investigate this hypothesis further.

The work with these two THIQ substrates went on to become part of a larger publication, in which cascades featuring the 3-enzyme MT system were able to produce a variety of natural and non-natural products.[156] This demonstrated the scalability of the system, but only with one target moiety, i.e. catechols, and only with transfer of the methyl group. The focus of this project therefore became to take this foundational biocatalytic method and extend it to new substrates and capabilities.

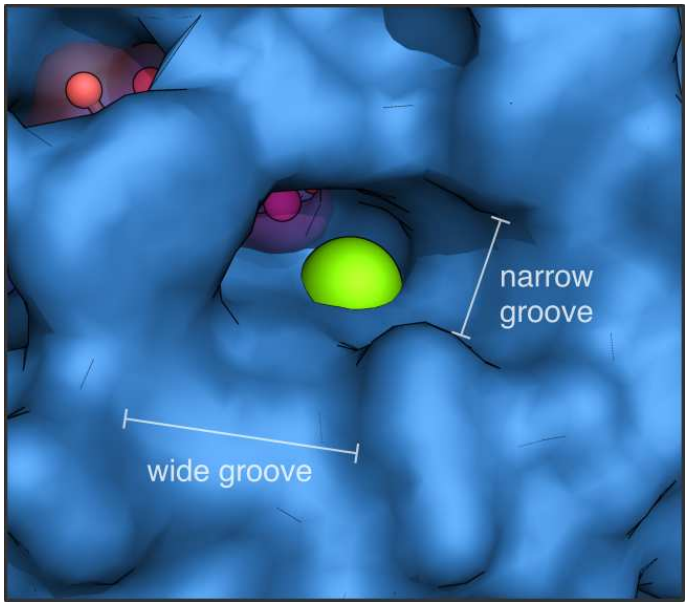


Figure 2.3: Surface model of *Rn*COMT active site. Magnesium ion represented by green sphere. SAM cofactor represented as magenta balls-and-sticks with a translucent surface.

Chapter 3

Exploration of new methyltransferases

3.1 Introduction

*Rn*COMT proved useful as a biocatalyst, combining a wide substrate tolerance with selectivity and efficient turnover in benign conditions. However, catechols comprise only one of the eight chemical groups targetted by OMTs. While enzymes for the other seven have been studied, relatively few have been explored as biocatalysts. Therefore, one branch of this project involved expressing and testing new MT candidates, with the aim of expanding the applications of MTs to other target groups. As there are over a hundred types of *O*-MT, (Figure 1.4), the selection was informed by two factors: whether the target group is of pharmacologial interest, and whether there has already been work to characterise the protein. Prior characterisation saved time by demonstrating the enzyme can be expressed, elucidating its cofactor or ion dependancies and, in some cases, providing a structure to help rationalise results.

Seven MTs matching this criteria were selected. Four targetted aliphatic hydroxyls, of which three had sugar substrates. In mammals, carbohydrates are not methylated outside of DNA. However, the modification exists in bacteria, fungi, plants, worms and molluscs, where it decorates both oligosaccharides and individual monomers. Glycans form components of a number of drug molecules,[158] so the capacity to methylate specific hydroxyls on those rings has value to drug development.

Mtf1 specifically methylates the 3-OH of rhamnose residues (Figure 3.1A). It was first identified in *Mycobacterium smegmatis*, where it functions in the biosynthesis of glycopeptidolipids in the cell wall.[159] Demethylmacrocin OMT (DMOMT) and macrocin OMT (MOMT) operate in the biosynthesis of tylosin, methylating the 2-OH and 3-OH positions, respectively, of a 6-deoxy-D-allose adduct (Figure 3.1B).[160] When first characterised, these enzymes were assumed to act in sequence, with MOMT dependant on the 2-OMe for substrate binding. However, more recent work has demonstrated that they can act in either order, hinting at a broader substrate specificity.[161]

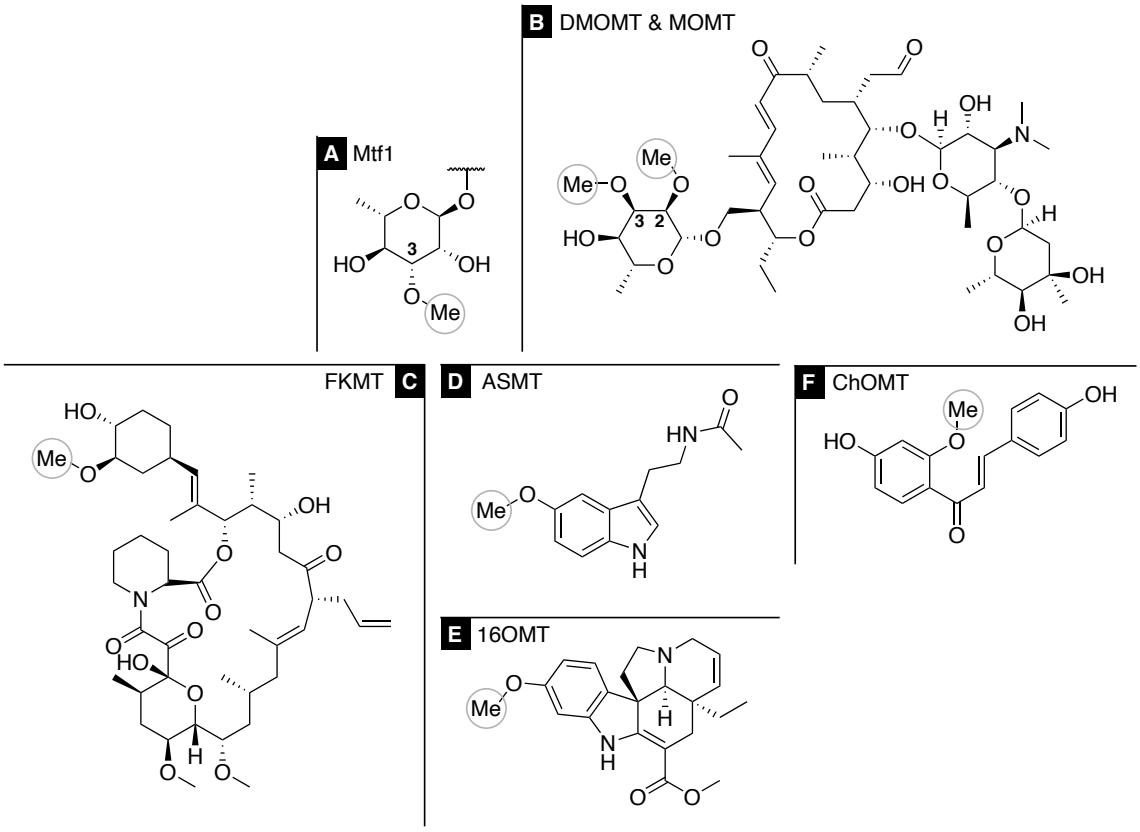


Figure 3.1: Natural substrates of the seven novel candidate MTs. 'Me' denotes site of preferential methylation.

FkbM (FKMT) also targets aliphatic alcohols, specifically the 3-OH of a 3,4-dihydroxycyclohexane adduct (Figure 3.1C). It is native to *Streptomyces tsukubaensis*, where it operates in the biosynthesis of the natural macrolide FK506, a compound with powerful immunosuppresive properties.[162] Once again, recent investigations have shown the nonlinearity of this pathway, with FkbM able to accept a number of related substrates.[163]

The remaining candidates targetted phenols. *N*-acetylseratonin MT (ASMT) methylates a phenol hydroxyl in the final step of melatonin biosynthesis (Figure 3.1D).[88] 16-hydroxytabersnonine-16-OMT (16OMT), meanwhile, performs one step in the biosynthesis of the alkaloid vindoline, a precursor to the anticancer agents vinblastine and vincristine (Figure 3.1E).[164] As such, the pathway has been subject to bio-engineering efforts with the aim of increasing production *in vivo*.[165] However, as yet, the enzyme has not been explored as a biocatalyst in its own right. The final candidate enzyme was Chalcone OMT (ChOMT), which targets the 2'-OH of isoliquirtigenin in a mechanism that has been studied at quantum-mechanical depth (Figure 3.1F).[105] The presence of an unaffected phenol group on that same substrate is indicative of the specificity towards the diphenol. Furthermore, ChOMT demonstrates how methyl groups can have a profound impact on biosynthetic pathways despite their small size. By methylating that hydroxyl, the enzyme prevents cyclisation of isoliquirtigenin into a flavonone, thus regulating the production of a host of secondary metabolites.[105]

Synthetic genes were designed by reverse-transcription of the protein sequences, with codon optimisation for *E. coli*. Minimal extensions were made to the 5' and 3' ends to create restriction sites for *NdeI* and *XhoI*, respectively (See Section 8.2.2 for details). The genes were synthesised by Eurofins Scientific and delivered in a carrier vector. After amplification of that vector in *E. coli*, classical restriction cloning was used to transplant the synthetic genes into the expression vector pET-28a(+). This process was ultimately successful for ASMT, FKMT, DMOMT, MOMT and Mtf1, as confirmed by sequencing of the resulting plasmids. Unfortunately, despite repeated attempts, ligation of digested 16OMT and ChOMT into the vector failed repeatedly. Therefore, these two candidates were set aside. Vectors containing the five successfully cloned enzymes were transformed into *E. coli* BL21(DE3) for expression.

3.2 Optimisation of expression conditions for novel methyltransferases

As a first step, optimal expression conditions were sought for each of the cloned candidates. In the original publications, the enzymes were either extracted from the native

organism, expressed in a system for which the facilities were not available (i.e. mammalian cells) or otherwise no details were given on the expression procedure. The first step, then, was to establish the optimal media for culturing in *E. coli*. Each expression strain was grown in four common bacterial broths: Lysogeny broth (LB), Terrific Broth (TB), 2xYT and Super Optimal Broth (SOB). In each case, mild expression conditions were used (20℃, 0.1 mM IPTG). All strains, except for that containing the Mtf1 gene, showed evidence of expressing their respective proteins (Figure 3.2A).

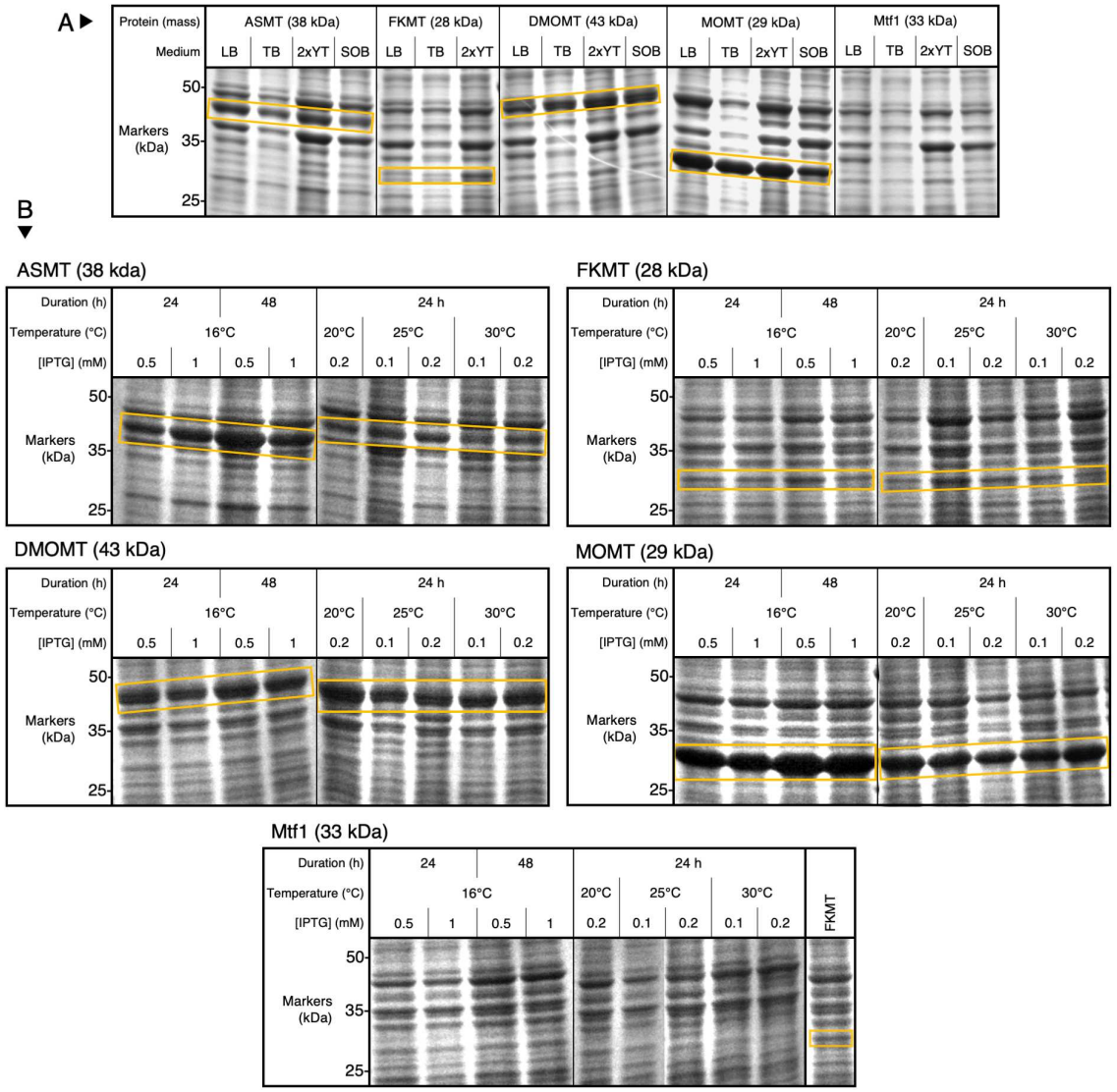


Figure 3.2: SDS-PAGE analysis of expression conditions for ASMT, FKMT, DMOMT, MOMT and Mtf1. Adjacent gel inages aligned by their size markers. Bands of interest highlighted in orange boxes. A) Optimisation of expression medium. FKMT expression in SOB failed to grow, so is not shown. B) Parallel optimisation of temperature, IPTG concentration and time.

In all cases, 2xYT appeared to give the greatest expression, as indicated by the size and intensity of the bands at the expected molecular weights of each protein. The overall abundance of protein across the 2xYT lanes on the SDS-PAGE gel indicated this was a general effect on protein yield, not a specific effect on the protein of interest. Nonetheless, 2xYT was chosen as the optimal broth.

Temperature and IPTG concentration were then optimised in concert. Testing a full range of combinations was not practical. Therefore, pairs of conditions were tested which were most likely to be successful, by seeking to balance the metabolic load of induction with that of temperature. As such, low IPTG concentrations (0.1 and 0.2 mM) were combined with higher temperatures (20, 25 and 30 ℃) while higher IPTG concentrations (0.5 and 1 mM) were tested at 16 ℃. To account for slower metabolism at 16 ℃, these conditions were also tested with 48 h expression periods alongside the standard 24 h.

SDS-PAGE analysis of these trials is shown in Figure 3.2B. ASMT showed reasonable expression across all conditions, with optimal expression at 16 ℃, 0.5 mM and 48 h. FKMT was also visible in all conditions, albeit at low levels. DMOMT expressed well with optimal conditions of 20 ℃, 0.1-0.2 mM IPTG and 24 h. However, the mass of DMOMT overlaps with that of an endogenous protein, so the true extent of the expression was difficult to gauge. MOMT showed the strongest expression of any candidate, with particularly high production at 16 ℃ and 0.5-1 mM IPTG with 48 h expression. Finally, Mtf1 again showed no evidence of expression. A lane from the FKMT SDS-PAGE is given alongside the Mtf1 panel in Figure 3.2B to provide a negative control comparison. No new bands are visible in the region corresponding with the molecular mass of this protein. The growth of strains containing the four detectable proteins, ASMT, FKMT, DMOMT and MOMT, was scaled up to 500 mL for purification, each at their optimal expression conditions.

3.3 Purification and assaying of ASMT

ASMT was the most attractive candidate to develop first because of the detailed structural information available.[50] In the event that engineering was needed, the X-ray structure of *Hs*ASMT would allow mutations to be rationalised and directed.

The first step was affinity purification, in order to conduct an assay without the cell

lysate background and establish the activity of the enzyme. However, during this process, it was discovered that the protein was insoluble. The band in Figure 3.2 was due to the SDS in the lysis buffer solubilising the precipitated protein. A publication by Ben-Abdallah *et al.* noted a similar problem with expressing *Hs*ASMT in *E. coli*. In that case, the authors instead expressed the protein in the eukaryotic host *Leishmania tarentolae*.[166] Copying this method was thus considered as an option. However, Byeon *et al.* later demonstrated expression of ASMT from *Arabidopsis thaliana* in *E. coli* BL21(DE3) strain, showing at least one homologue could be grown in a simpler host.

It was decided to first attempt expression of a small group of ASMT homologues likely to express in *E. coli*, as there was more expertise with this host in our group. This included the *A. thaliana* variety, as well as two more homologues from the BRENDA database: one from *Chromobacterium violaceum*, which had previously produced a robust transaminase,[151] and one from *Cupriavidus necator*, a versatile industrial bioengineering platform.[167–169]

A multiple sequence alignment of these three alongside *Hs*ASMT showed little conservation (Figure 3.3). Referring to the *Hs*ASMT crystal structure, some conserved residues have clear roles in substrate binding (G187 makes a backbone contact with the SAM methionine moiety, D210 coordinates the SAM ribose hydroxyls, F156 and M303 create a hydrophobic environment around the substrate indole ring) or catalysis (H255 and D256). Other conserved positions have uncertain roles, and some residues which seem important in the *Hs*ASMT structure are not conserved at all, such as R252, which coordinates the SAM methionine carboxyl. This lack of conservation was unexpected but potentially useful. It indicated that a high degree of sequence drift had occurred between the homologues, meaning there would be a greater chance of a variant with desireable properties, such as greater substrate tolerance from a differently shaped active site.

Genes for the three enzymes, codon-optimised for *E. coli*, were commercially synthesised and cloned into pET-29a(+) prior to delivery. The open reading frames were designed to include the C-terminal His-tag encoded in that vector. The plasmids were transformed into the *E. coli* BL21(DE3) strain, and all three were expressed using the conditions identified for *Hs*ASMT (Figure 3.2). Samples of the cultures were lysed by sonication and samples from the soluble fraction analysed by SDS-PAGE.

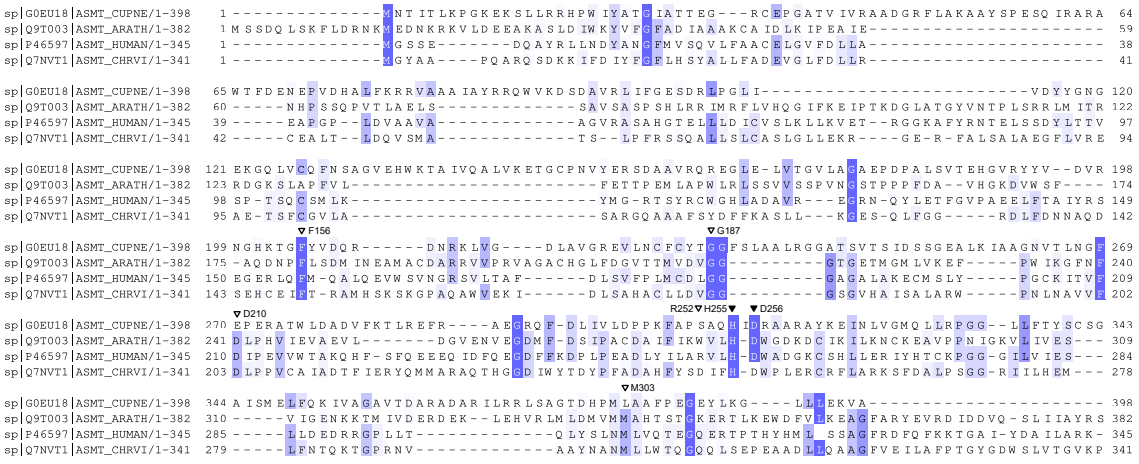


Figure 3.3: Multiple sequence alignment of *N*-acetylseratonin *O*-methyltransferase genes from *H. sapiens* (ASMT_HUMAN), *A. thaliana* (ASMT_ARATH), *C. necator* (ASMT_CUPNE) and *C. violaceum* (ASMT_CHRVI). Identical amino acids highlighted in blue, with stronger colour for more conserved residues. Residues with substrate/cofactor binding and catalytic roles in *Hs*ASMT are indicated by empty and filled triangles, respectively. Sequences aligned by Clustal Omega, alignment visualised with Jalview.

Both the *A. thaliana* and *C. violaceum* homologues showed overexpression bands at the approximately the expected molecular weight (Figure 3.4A). The *C. necator* homologue showed no visible expression. However, due to the saturation of the gel lane, it was possible the enzyme was present but in undetectable amounts.

These three lysates were assayed directly to confirm their activity against the natural substrate, *N*-acetylseratonin (NAS) **59**. The concentration of **59** was 1 mM, with 6 mM of SAM. Cell lysates were used at 50% (*v/v*) concentration: an excessive amount relative to normal conditions, meant to detect even slight levels of enzyme activity. No other enzymes were added, but buffer conditions were kept the same as previous MT assays (Assay III). Formation of the methylated product, melatonin **60**, was analysed by analytical reverse-phase HPLC using a nonspecific 5-95% acetonitrile gradient (HPLC method C).

Surprisingly, no peaks corresponding to **60** were detected in any of the traces (data not shown). Peak area of the starting material was consistent across all conditions, including the empty vector negative control, ruling out precipitation or otherwise degradation of the product.

Solubility of the enzymes which did express had already been determined. The start-

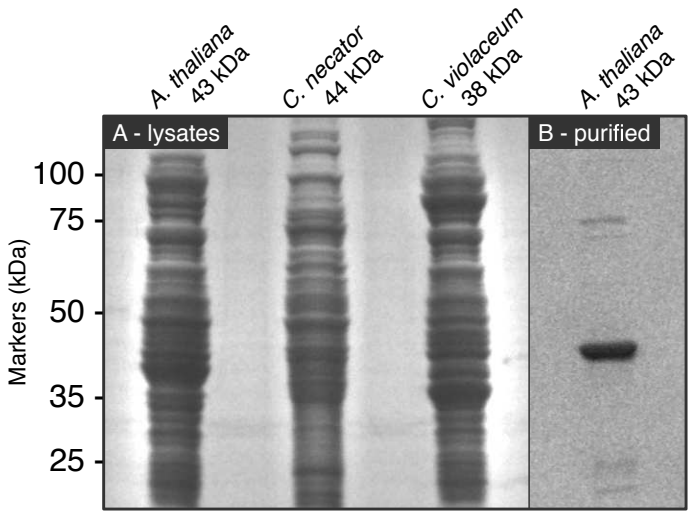
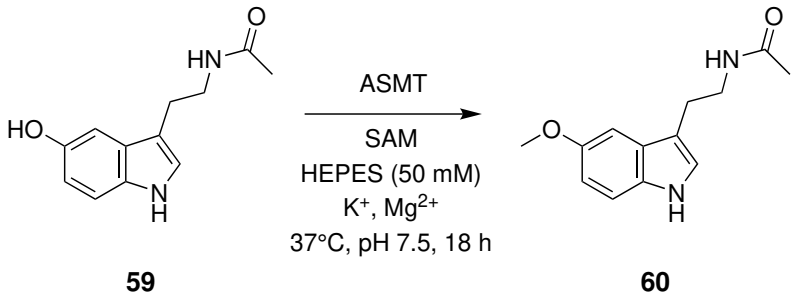


Figure 3.4: SDS-PAGE analysis of the expression of three ASMT homologues. Origin species given at top. A) Clarified lysates of *E. coli* cells expressing each homologue, lysed by sonication. B) Purified ASMT from *A. thaliana*. Gel images aligned by their markers.



59 **60**

Scheme 3.1: Enzyme assay of ASMT homologues, transforming *N*-acetylseratonin **59** into melatonin **60**.

ing material was subjected to NMR spectroscopy, which confirmed its identity and purity. ASMT is not known to have metal ion or cofactor dependencies beyond SAM, and the pH and temperature of these assays was similar or exactly the same as those reported.[88] Therefore, the reason for the total non-functionality of any of the three cell lysates was not clear.

It was decided to repeat the assay with *At*ASMT while following the assay conditions reported by Byeon *et al.* as closely as possible. This comprised: 100 mM phosphate buffer, pH 7.8, 1 mM NAS **59**, 0.5 mM SAM, at 37 ℃ for 1 hour. The authors of that study also used purified enzyme, but did not specify a concentration.

To match these conditions, *At*ASMT was purified by affinity chromatography. SDS-

PAGE analysis of the purified protein showed one dominant band at approximately the expected molecular mass (Figure 3.4B). However, the protein began to precipitate late into the process, as indicated by increasing cloudiness of the purified solution. It was frozen at -80℃ in this state. Upon thawing for the assay, the precipitate was spun down and the concentration of protein in the supernatant determined immediately before use. It was ultimately decided to use the purified *At*ASMT at a concentration of 0.2 mg/mL, although the concentrations used in the literature had not been reported (Assay IV).

Unfortunately, even when the assay was repeated using the literature conditions, no new peaks were observed by analytical HPLC (data not shown). At this stage, it was not clear what further avenues could be explored with the ASMT homologues.

Despite efforts to replicate the reported assay conditions for *At*ASMT, there remained a difference between the DNA constructs used here and those given in the work of Byeon *et al.*. In that study, the gene encoding *At*ASMT was amplified from cDNAs, then recombined first into a gateway donor vector and finally into the destination vector pET300. This would put the His-tag of the expressed protein on the N-terminus, whereas pET-29a(+) features a His-tag at the C-terminus. Despite being a small part of the protein, it is well-documented that the presence and location of the His-tag can affect protein stability and function.[170–173] This could potentially explain why no activity was observed in these assays. A counter might be that, given the wide sequence diversity between the three new homologues (Figure 3.3), it seems unlikely that the position of six additional residues would have such a total knock-down effect on all of them. Still, without testing, it would be impossible to be certain. Addressing this could first involve sub-cloning the *At*ASMT gene into pET-28a(+), which contains an N-terminal His-tag, then again purifying and assaying the proteins using literature conditions. Unfortunately, there was not time in this project to pursue further solutions.

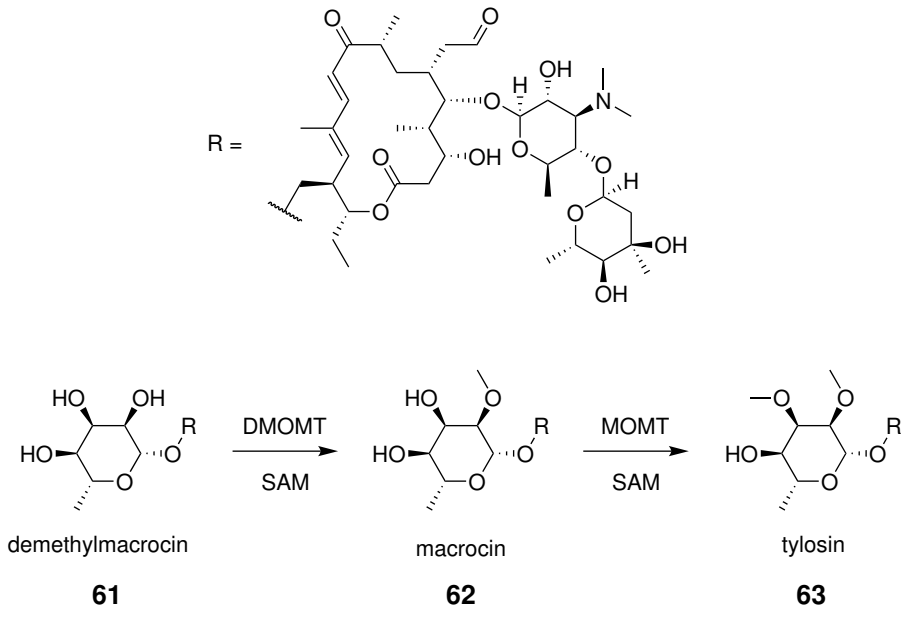
3.4 Purification and assaying of MOMT and DMOMT

As the MT with the greatest expression during the initial screens (Figure 3.2), MOMT was another promising candidate. Purification was first attempted with this enzyme, with a plan to establish a working assay protocol before conducting the same process with the

weaker-expressing sister enzyme, DMOMT.

MOMT remained in the soluble fraction after cell lysis by sonication. Unfortunately, most of the enzyme precipitated immediately after purification. This was found to be true under both pump-assisted and gravity flow methods, in both HEPES- and Tris-based buffers and with and without supplementation with 50 mM L-arginine. However, SDS-PAGE analysis showed that some enzyme remained in solution, and therefore could be assayed for activity.

Unlike ASMT, the natural substrate of MOMT is not easily accessible. The native biosynthetic pathway involves methylation of demethylmacrocin **61** by DMOMT at the 4-OH of a 6-deoxy-D-allose adduct, giving macrocin **62**. Macrocin is then methylated again at the 3-OH by MOMT to give tylosin **63**. This final product is available as a veterinary antibiotic, but the two intermediates are not. Neither is the 6-deoxy-D-allose itself.

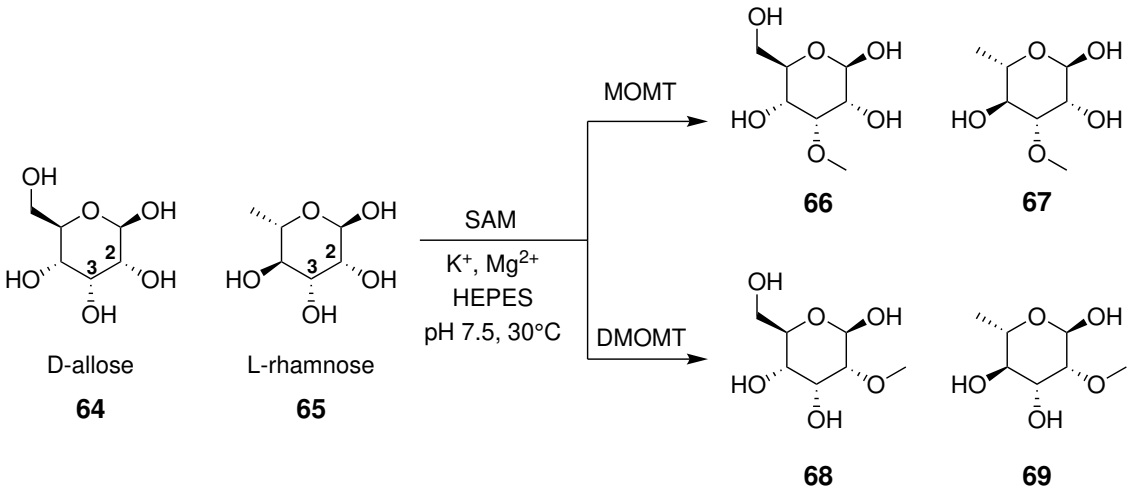


Scheme 3.2: DMOMT- and MOMT-catalysed steps in tylosin biosynthesis

Two substrates were therefore chosen which were structurally similar to the natural substrate's sugar adduct (Scheme 3.3). D-allose **64** is the same, but with a hydroxyl on the position 6 of the ring, while L-rhamnose **65** is the same but with the chirality of positions 4 and 5 reversed.

The published assays of DMOMT and MOMT were conducted largely *in vivo*, in the host *S. venezuelae*, using a reconstructed tylosin biosynthetic pathway. For the *in vitro*

assays, cell-free extracts of that mutant host were prepared by glass-bead homogenisation, mixed with a reaction buffer and supplied with purified macrolide intermediates before incubation, extraction and analysis. The procedures in that work were therefore not applicable to this project, which sought to determine the substrate scope of the enzymes on their own, having been expressed in *E. coli*. However, details on the buffer composition, pH and temperature of those reactions were all informative, and indicated that conducting the assay at 30 ℃ in the standard reaction buffer used so far would be sufficient. Assays were thus run with purified MOMT using the conditions given in Scheme 3.3. The two expected products from **64** and **65** were **66** and **67**, respectively.



Scheme 3.3: Assay for the activity of DMOMT and MOMT against sugar substrates.

HPLC analysis of previous MT assays had used diode-array detection (DAD) relying on the UV absorbance of the substrates' aromatic rings. As **64** and **65** lack chromophores, electrochemical detection (ED) was instead used to analyse the assays. Briefly, ED involves using electrodes, in contact with an eluent stream, to detect eluted compounds by their redox reactions.[174] The specific form of ED used in this experiment was high-performance anion exchange chromatography with integrated pulsed amperometric detection (HPAEC-IPA). For this, a voltage is held constant between a working and a reference electrode. As a chemical species is reduced or oxidised at the working electrode, the resulting current/charge is measured, and gives a peak similar to the absorbance peaks of chromaphores with DAD.[174] As with HPLC, retention on the solid phase prior to detection is characteristic for each species, so peaks can be assigned to compounds and

the appearance/disappearance of peaks can be used to track reactions.

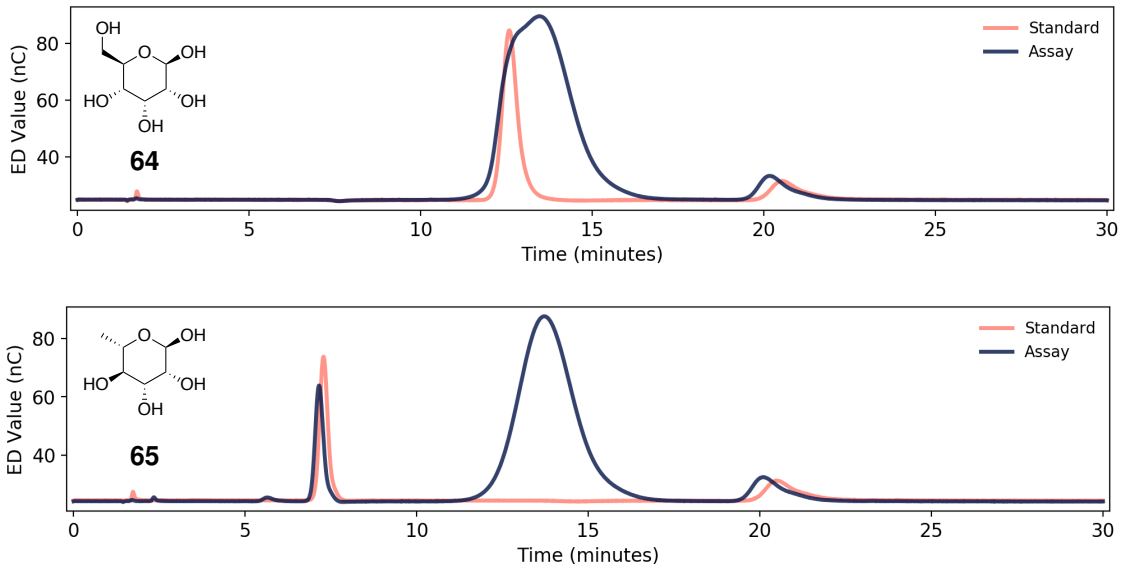


Figure 3.5: HPAE-IPA analysis of MOMT assays with D-allose and L-rhamnose. Assay trace represents crude product of the reactions given in Scheme 3.3. Standard trace represents the same reaction mixtures without enzyme. Assay V conditions. See Section 8.4 for analysis parameters.

For D-allose, this analysis showed two major peaks (Figure 3.5). The larger of the two, at 13.5 min, presented a shoulder. By comparing with the starting material standard trace, this was explained as an overlap between the peak corresponding to **64** and an unidentified later peak. The small peak at 20.2 min appeared in both the assay and standard traces, so was taken to represent an impurity or element of the buffer not indicative of reaction progress. For L-rhamnose, the large peak was separate from the starting material peak at 7.2 min.

From these data, it was concluded was that no methylated product had been observed for either substrate. The peak appearing at 13.5 min was ascribed to glycerol, a cryoprotectant in the protein solution. Otherwise, there were no new peaks in the assay traces compared to the standard traces, indicating no product had formed.

Several causes were considered, including that the enzyme simply did not accept these substrates. However, given the apparent instability of pure MOMT, it was decided to first rule out that the enzyme was precipitating before it could catalyse the reaction. Precipitation had not been observed in the clarified lysate prior to purification. A new assay was thus conducted using 20% (*v/v*) lysate, with the hypothesis that the enzyme

would be more stable in this form. The same substrates were used at 5 mM, and a cell lysate of DMOMT was also included. If productive, this MT was expected to give products **68** and **69** (Scheme 3.3). Otherwise, the same conditions were used (Assay VI).

The products of the second series of assays were analysed by LC-MS, utilising its greater sensitivity to identify even small amounts of methylated product. Unfortunately, no mass peaks corresponding to any possible ionisation state of the methylated sugars was detected (data not shown). In light of these results, it was increasingly likely that the barrier to methylation was acceptance of the sugar substrates. Not only were they different to the 6-deoxy-D-allose adduct of the natural substrate, they also lacked the rest of the macrolide ring and other substituents thereof. Either or both of these differences could impact binding, and thus explain the lack of observed methylation with either substrate.

Using SWISS-MODEL, homology models of MOMT and DMOMT were built to help understand the lack of substrate tolerance. Of the two, the model of MOMT had a better QMEAN score and a recognised cofactor binding site, so was focussed on. The template for this model was mycinnamycin III 3'-*O*-methyltransferase (PDB:4X7U), an enzyme also involved in the biosynthesis of a macrolide antibiotic. The aim was to computationally dock macrocin into this homology model. Ideally, the binding pose thus discovered would illuminate the substrate-enzyme interactions that are important for binding and catalysis, which could help rationalise why **64** and **65** alone were not accepted. Unfortunately, docking was unable to find a pose in which the sugar adduct was located in the active site in an orientation consistent with the known reaction, i.e. with the 3-hydroxyl positioned for attack of the SAM methyl group. It was therefore not possible to draw specific conclusions on why the substrates had not been accepted. Broadly, however, it was hypothesised that, being rich in hydrogen bonding atoms, the sugar itself does not contribute strongly to binding affinity due to the penalty of de-solvation. Therefore, the rest of the macrocin molecule may play a substantial role in binding, and in positioning the sugar adduct in the active site. Without it, binding may fail and catalysis not take place.

3.5 Conclusions

The final enzyme which showed expression, FKMT, could not be investigated in the time available. This branch of the project, then, failed to yield any new MT biocatalysts. On reflection, the approach taken here meant that efforts were spread thin, and the incentive for solving problems with each of the four groups was lower than that to pursue other, more fruitful aspects of the project. The gamble of searching in many places at once for a novel MT may have been encouraged by the success with *Rn*COMT, which performed as an excellent biocatalyst with little optimisation. However, because of the evolution of MTs towards extreme selectivity, cases like these are likely rare. A better strategy might therefore have been to focus on one target group. This would have allowed more deliberacy in candidate selection, potentially employing metagenomics to find varied homologues of known MTs. A larger screening set could then have been optimised and tested together. Future attempts to extend the MT repertoire might therefore benefit from a tighter focus to increase the odds of success.

Chapter 4

Ethylation

4.1 Introduction

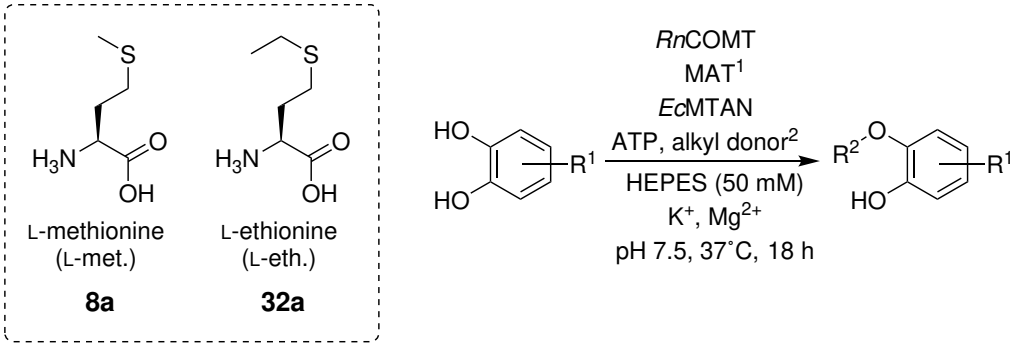
The work described in Chapter 2 showed the scalability of the MT enzyme cascade with THIQ substrates. The goal then was to access methylated derivatives, the potential benefits of which were explained in Chapter 1. However, another direction of investigation in this project involved determining if that system could be expanded to transfer other alkyl groups. The number of derivatives that can be generated is, naturally, increased for every new group that can be transferred, multiplying the volume of chemical space that can be searched for useful molecules. It was therefore desireable to determine if and how the existing system could be adapted to tolerate these new alkylations. Moreover, it was important that the system continue to be scaleable and compatible with THIQs.

4.2 Modifying the MT enzyme cascade

Alternative alkylations have already been demonstrated in the literature for a number of MTs and substrates. Details on these studies are given in Chapter 1. However, for all applications based on the MT enzyme cascade, there are two common considerations: the methionine analogue and the MAT homologue.

The *S*-methyl group of methionine is the chemical fragment that is ultimately transferred to the substrate by the MT. Therefore, alternative alkylations which employ an MAT to generate the cofactor require methionine analogues which carry a different group on the sulfur. The analogue chosen to be investigated first was L-ethionine (L-eth.) **32a**

(Scheme 4.1), which on incorporation into the *Rn*COMT enzyme cascade, would result in ethylation of the target catechol. There were two main reasons for this choice: L-ethionine is commercially available at reasonable cost for small-scale assays, and it was also predicted that the similarity between it and L-methionine would increase the chances of tolerance by the enzymes. This analogue was therefore considered a useful test-case in which any obstacles to alternative alkylation might be first encountered.



Scheme 4.1: Ethylation assays with small and THIQ substrates. [1]MAT was either *Tk*MAT or *Mj*MAT. [2]alkyl donor was either **8a** or **32a**.

For the MAT component, the system used in Chapter 2 employed *Ec*MAT for SAM generation. However, it was reported that this homologue has a very poor tolerance for methionine analogues (with Ottink *et al.* publishing a notable exception[129]).[125] Variants with broader acceptance for analogues include those from *Homo sapiens*,[125,127,128,131,175] *Methanococcus janaschii*,[113] *Thermococcus kodakarensis*[115,123] and *Sulfolobus solfataricus*.[116,127] Our collaborators in the research group of Prof. Jennifer Andexer (University of Freiburg) kindly provided expression plasmids encoding *Hs*MAT2A, *Tk*MAT and *Mj*MAT. These were transformed into *E. coli* BL21(DE3) cells and expression attempted.

SDS-PAGE analysis of the clarified lysates showed clear expression of both *Mj*MAT and *Tk*MAT, though the former more than the latter (Figure 4.1). Our collaborators communicated to us that they, too had observed low expression of *Tk*MAT despite trying a range of host strains. Unfortunately, *Hs*MAT2A did not express. The plasmid was re-transformed several times, but no amount of expression was ever observed under standard conditions. The key components of the first alternative alkylation were therefore established: commercially sourced L-ethionine as the analogue, and *Mj*- and *Tk*MAT to generate the cofactor analogues.

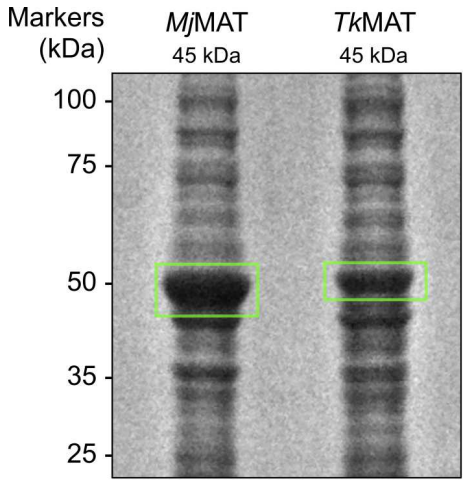


Figure 4.1: SDS-PAGE analysis of *Mj*MAT and *Tk*MAT expression. Bands suspected to represent proteins of interest highlighted with green boxes.

4.3 Initial findings from ethylation assays

Assays of the ethylation system were conducted using the same conditions as had been developed for methylation, with *Rn*COMT (10% *v/v*) as the MT and *Ec*MTAN (2% *v/v*) for SAH degradation (Scheme 4.1). *Tk*MAT and *Mj*MAT were likewise used as clarified lysates at 10% (*v/v*) (Assay II). Four catechol substrates were tested: 3,4-dihydroxybenzaldehyde (DHB) **70**, dopamine **23** and the racemic THIQs **(*RS*)-54** and **(*RS*)-57**. DHB **70** has already been reported to be ethylated by a similar cascade system,[126] while dopamine **23** is a natural substrate of *Rn*COMT so was expected to be well accepted. The two THIQ substrates, meanwhile, would give an indication of the system's tolerance for these more complex compounds. Each substrate was assayed with both **8a** and **32a**.

The products of these assays were analysed by reverse-phase HPLC, which showed the formation of new peaks for **70** (RT 9.4 min), **(*RS*)-54** (RT 9.5 min) and **(*RS*)-57** (RT 9.8 min) (Appendix D.1). The retention times of these peaks were consistent between repeats, eluted later than the methylated products (indicating higher hydrophobicity) and only appeared in the ethylation conditions. It was therefore believed that they represented ethylated products. Conversions to these new products are given in (Figure 4.2).

Conversions were estimated by dividing the peak area of the product peak by the total peak areas of the substrate and all products (see Section 8.4 for diagram). Other methods of calculating conversion, such as measuring starting material consumption or

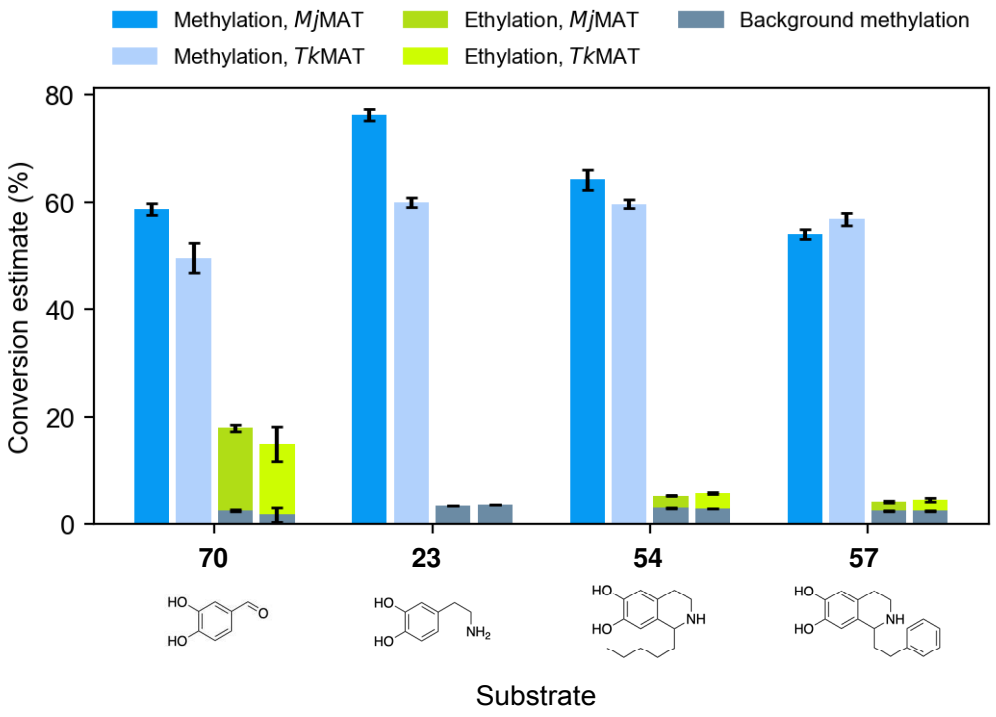


Figure 4.2: Analysis of ethylation assays utilising *Rn*COMT, *Ec*MTAN and either *Mj*MAT or *Tk*MAT. Conversions calculated from HPLC peak areas of product and remaining starting material. Error bars represent one standard deviation above and below the mean of three replicates. Methylation assays supplied with L-methionine, ethylation assays supplied with L-ethionine. Grey bars represent background methylation detected in ethylation experiments. Assay II conditions. HPLC method A.

product formation using calibration curves, were considered but ultimately not used. This is because it was found that reaction samples varied in concentration at the point of analysis, due to evaporation and errors introduced during sample preparation. This meant that the absolute peak areas sometimes varied considerably between repeats of the same reaction, which would make calibration curves of either the starting material or product unreliable. However, ratio between the peak areas was not affected by concentration, so was far more consistent in those cases.

This could only be considered an estimate, however, as it was based on the assumption that the absorbance of the chromophore (i.e. the aryl ring, at 283 nm) was not modified in the product. To test this assumption, a calibration curve was constructed for **70** and a commercially available standard of the expected product, ethyl vanillin (3-ethoxy-4-hydroxybenzaldehyde). These curves showed only a marginal difference between the

absorbance of those two compounds (Appendix C.1). On balance, then, this method of measuring the outcome of alkylations was adopted.

For all substrates, the degree of ethylation was far lower than methylation. DHB **70** gave the highest conversion, while dopamine **23** gave no new peaks at all. This was a surprise, given that dopamine is a natural substrate of the MT. Ethylation of the two THIQ substrates was detectable, but minimal. Background methylation was also observed for all substrates. The was likely due to biogenic L-methionine and SAM in the cell lysates, both of which could be readily taken up by the enzyme cascade.

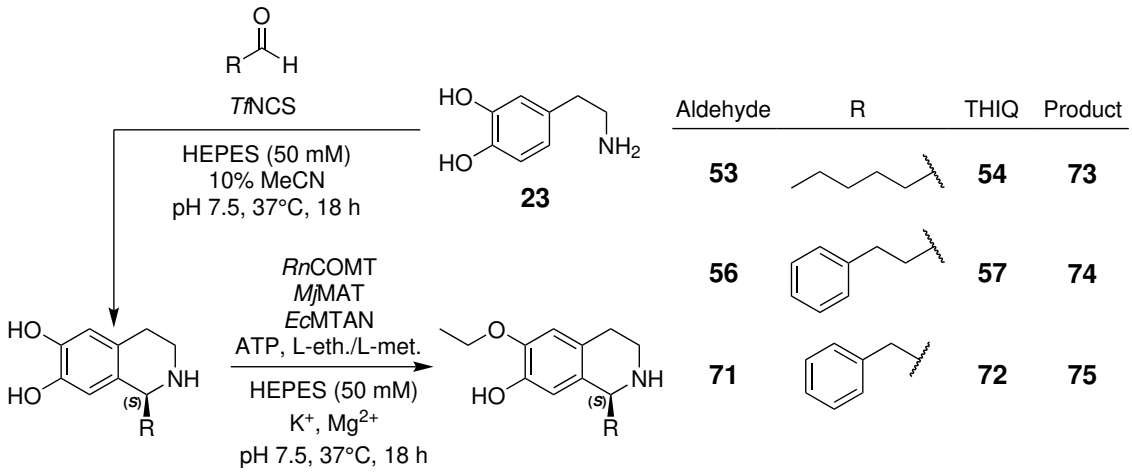
Methylation, though higher than ethylation, was also lower than expected. Substrates **23**, **(*RS*)-54** and **(*RS*)-57** had all been been methylated quantitatively by the system based on *Ec*MAT under the same conditions. The result could be due to *Tk*MAT and *Mj*MAT having lower turnover in these conditions than *Ec*MAT. It had been reported that, as a general trend, enzyme-catalysed reactions in thermophilic organisms (such as *Methanococcus janaschii* and *Thermococcus kodakarensis*) have higher activation energies due to the rigidity of those proteins.[113] However, a lack of like-for-like comparisons of the three enzymes' kinetic parameters makes definitive conclusions difficult. Furthermore, due to variations in enzyme expression between cultures, reactions relying on cell lysates would be subject to batch effects. These had not been noticed during the methylation work, perhaps because that reaction almost always proceeded to completion. However, in a case where some or all of the enzymes involved were operating outside their natural conditions, the MT system might be more sensitive to variations in the content of the lysates.

Is most reactions, *Mj*MAT gave either slightly higher or approximately the same conversion as *Tk*MAT. This was unexpected, as SDS-PAGE indicated that *Mj*MAT was substantially better expressed, so was expected to give commensurately higher conversions. That it did not may indicate higher activity of *Tk*MAT in this context. However, as neither of the two was clearly better, *Mj*MAT was used for subsequent assays due to its more robust expression.

4.4 NCS-ethylation cascades

Following the initial assays, tests were conducted of the full NCS-ethylation cascades towards ethylated (*S*)-THIQs. There were intended as models of reactions that would eventually be scaled up, and as opportunities to learn if the chirality of the C-1 position in THIQs impacted the conversions.

Three reactions were conducted, beginning with **23** and one of three aldehydes: **53**, **56** and **71** (Scheme 4.2). Unlike the experiments in Chapter 2, 20% (*v/v*) desalted clarified lysate of *Tf*NCS was used instead of 0.5 mg/mL purified enzyme. This preparation was less labor-intensive, and the desalting process removed the majority of phosphate that could catalyse a background Pictet-Spengler condensation. SDS-PAGE analysis indicated that the desalting process diluted the lysate as a whole, but did not disproportionately remove *Tf*NCS. Therefore, the desalted lysate were taken to be an acceptable substitute for purified enzyme.



Scheme 4.2: Two-step NCS-ethylation cascades, for generating and ethylating three THIQ substrates.

Dopamine was used at 10 mM, with 2 equivalents of **53** and 4 eqv. of **56** and **71** to compensate for oxidation of the latter two (Assay VII). The components of these reactions were incubated at 37 ℃ overnight, then samples of the crude reactions analysed by HPLC (Figure 4.3: step 1). In all cases, peaks corresponding to the THIQs had appeared, although the reaction with **72** appeared to contain a number of impurities in the 8-11 min region. Precipitates were removed from the crude reactions by centrifugation, then

enzymes and substrates for ethylation were added (Assay II). Control reactions containing the same enzymes but with L-methionine as an alkyl donor were run in parallel. After further overnight incubation at 37℃, samples were once again taken and analysed. In the control reactions, methylation had proceeded to completion or near-completion in all cases (Figure 4.3: Step 2, methylation). This was useful to confirm the activity of the enzymes involved, and to provide a reference peak for the methylated product. Note that the peaks for step 2 are approximately half as intense as those for step 1 because addition of the methylation components diluted the reaction. In the ethylation reactions, the starting material remained as the major peak, with traces of background methylation as seen in the initial assays. For **(*S*)-54** and **(*S*)-57**, small, later peaks were observed (RT 9.5 min and 9.7 min, respectively), which were not present in the methylation controls and which were analogous to the new peaks observed in the assays with **(*RS*)-54** and **(*RS*)-57**. Estimated conversions for the (*S*)-THIQs were also comparable to the (*RS*)-THIQs (6-7%). The trace of the **72**, however, was too crowded to distinguish any new peaks.

To support the assertion that these new peaks represented ethylated products, the reactions were analysed by LC-MS. This found peaks corresponding to M+1 for products **73** (*m/z* = 264), **74** (*m/z* = 298) and even **75** (*m/z* = 284) in their respective samples, confirming that ethylation had occurred (Appendix F). Yet despite demonstration that the NCS and ethylation reactions could be coupled, the low conversions were concerning. Clearly, major obstacles to alternative alkylation remained, and would need to be overcome before it was feasible to scale the reactions up.

The low tolerance of ethionine and SAE could due to the extra methylene clashing with residues in the active sites of either the MAT, *Rn*COMT, or both. Cofactor generation and use are both irreversible reactions, so equilibria do not need to be manipulated, and a lower binding affinity due to sterics should not be terminal to success. However, the enzymes would still be expected to have a half-life on the order of hours. If the reaction rates were reduced so much that the proteins denatured before they could complete, the lower binding affinity would still present a barrier. A second, non-mutually exclusive explanation could center around reactivity. The methylation reaction proceeds via an S_N2 mechanism, with attack of the substrate's nucleophilic oxygen on the electrophilic methyl group of SAM. The carbon is made electrophilic by the positive charge of the adjacent

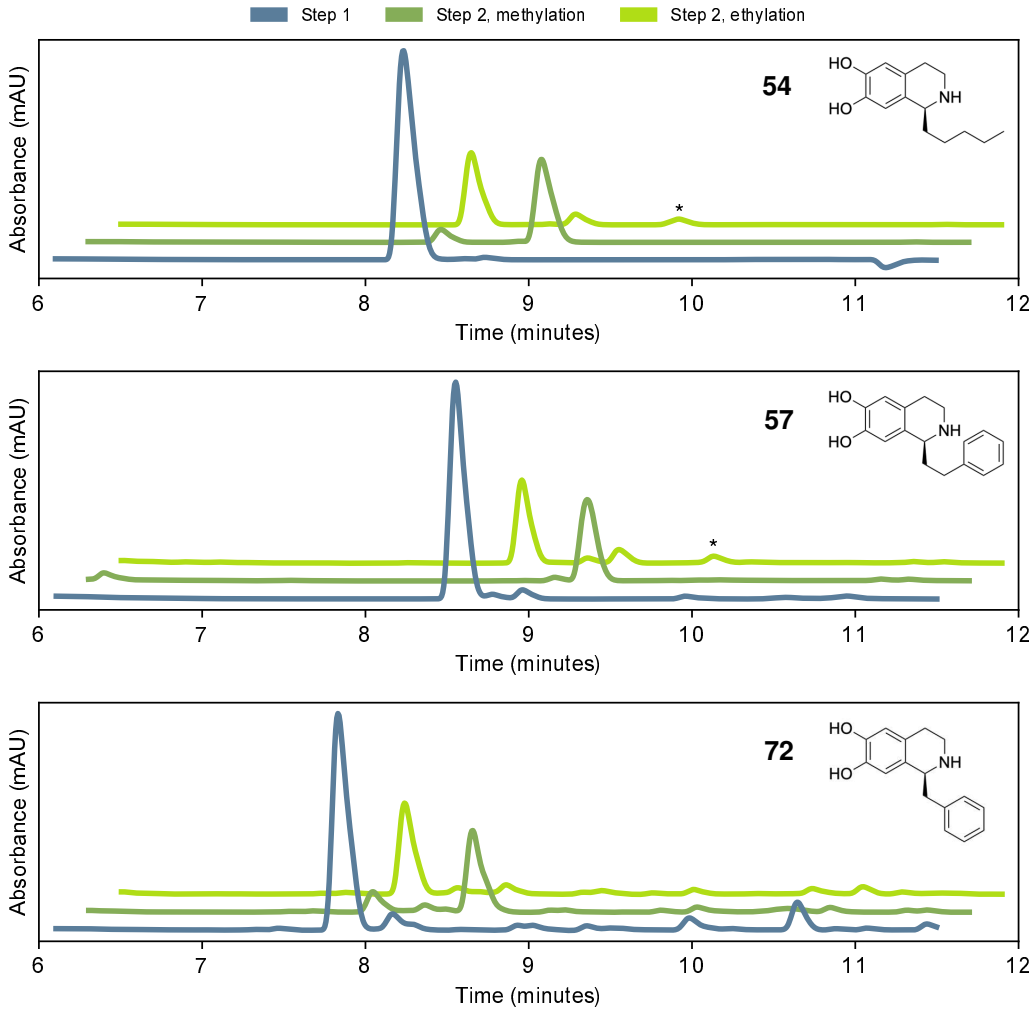


Figure 4.3: HPLC analysis of the NCS-ethylation cascades given in Scheme 4.2. Where identifiable, peaks likely to be ethylated product are marked with '*'. Traces are offset diagonally for clarity. Assay VII conditions for NCS step, II for MT step. HPLC method A.

sulfonium ion. Additional methylene groups would stabilise this positive charge, lessening the electrophilicty of the carbon and decreasing the energetic favourability of the transfer.

The first of these challenges - steric hindrance - could be amenable to enzyme engineering, in which the active site residue that clashed with the ethyl group could be mutated to improve the fit. The acceptance of L-ethionine by *Tk*MAT and *Mj*MAT had been confirmed previously.[113,123] However, no detailed investigations had been described into how the resulting cofactor interacts with the active site of *Rn*COMT. Therefore, computational docking methods were employed to better understand this complex.

4.5 Modelling

4.5.1 Introduction

Molecular docking is a computational technique which aims to find the optimal binding site of a small molecule on a target, often a protein.[176] Simulating both, in solution, with all degrees of freedom, for long enough to observe binding would be too computationally expensive to be practical. Docking programs therefore use algorithms, assumptions and simplifications to strike a balance between accuracy and speed.[176]

AutoDock is a popular choice of software for this purpose. Several algorithms and search strategies are available in the program, but many rely on the same concept: the ligand is placed on a simplified model of the protein surface, and 'wanders' with random changes to its position and torsion angles. Each pose is evaluated with an energy-like quantity, and only realistic conformations are allowed. As the simulation proceeds, the temperature of the system is lowered, making ever more poses unacceptable until the ligand falls into an energy minimum from which it cannot move. It is assumed that the binding site of a small molcule will be the lowest energy conformation in the defined search area. Therefore, by running the simulation a number of times, the user can get a consensus of the likely binding site and the ligand pose adopted therein.[176] Genetic algorithms help to avoid the ligand being stuck in non-optimal local minima by combining acceptable poses and introducing an element of random mutation.[177,178]

One of the inaccuracies in docking comes from solvent effects, which AutoDock estimates empirically using a model of desolvation energy.[106] Autodock Vina was released in 2010 with a number of improvements that dramatically increased the speed and supported compound screening.[106,179] While Vina performs better on benchmark tests,[179] the authors of AutoDock recommend not using Vina for binding sites with metal ions.[106] AutoDock 4.2's scoring function is partly physics-based and contains a screened Coulomb potential for assessing charge, whereas that of Vina is purely empirical, so can model these cases less accurately.[179]

In this project, docking was a means to visualise the difference in substrate binding between SAM-bound and SAE-bound *Rn*COMT. This might highlight residues that could be changed to increase the enzyme's tolerance for alternative alkylation. It also allowed

comparisons in between poses to be quantified by energy-like values, and thus the impact of any mutations to be modelled and assessed prior to experimental work.

4.5.2 Docking Results

Understanding how ethylation might occur in the *Rn*COMT active site required a model of the enzyme bound to SAE. No crystal structures of this complex have been reported, however, so one needed to be built by docking SAE into an 'empty' apo structure of COMT. As mentioned above, docking programs are generally more accurate when the target structure has been determined in complex with a similar substrate to the one being docked. The crystal structure published by Vidgren *et al.*[48] (PDB: 6LFE) featured *Rn*COMT in complex with both SAM and an inhibtor, so was chosen as the starting structure for this model.

The cofactor occupies a narrow pocket within *Rn*COMT, where it is held by well defined interactions. To confirm that docking would be able to find a realistic orientation of SAE within this space, SAM was first re-docked into its binding pocket, and the best result compared to the x-ray crystal structure. Docking was able to 're-discover' the crystal orientation, albeit with some variation. Most of the differences were localised to the carboxyl end of the methionine residue, which does not vary between SAM and SAE, is distant from the active site and is not directly relevant to the reaction mechanism.

A structure for SAE was built *de novo* and docked in the same manner. The expected binding orientation was again the best scoring result. Figure 4.4 provides a comparison between the orientation of SAM in the crystal structure (A) and the docked conformation of SAE (B). Differences persisted at the methionine end of the cofactor, but alignment was close between the decisive atoms adjacent to the sulfur. Therefore, this conformation of SAE was taken as a realistic model.

The final consideration was the Mg^{2+} ion. When AutoDock assigns Gasteiger charges, metal atoms are given a value of zero. As the ion electrostatics are vital for substrate binding, this would not make for an accurate model. There is no official guidance from the software authors on how to account for this. However, after consulting user forums, the consensus advice was to manually add a discrete '2' charge to the atom's line in the *.pdbqt files. This gave two final models for docking: *Rn*COMT in complex with SAM,

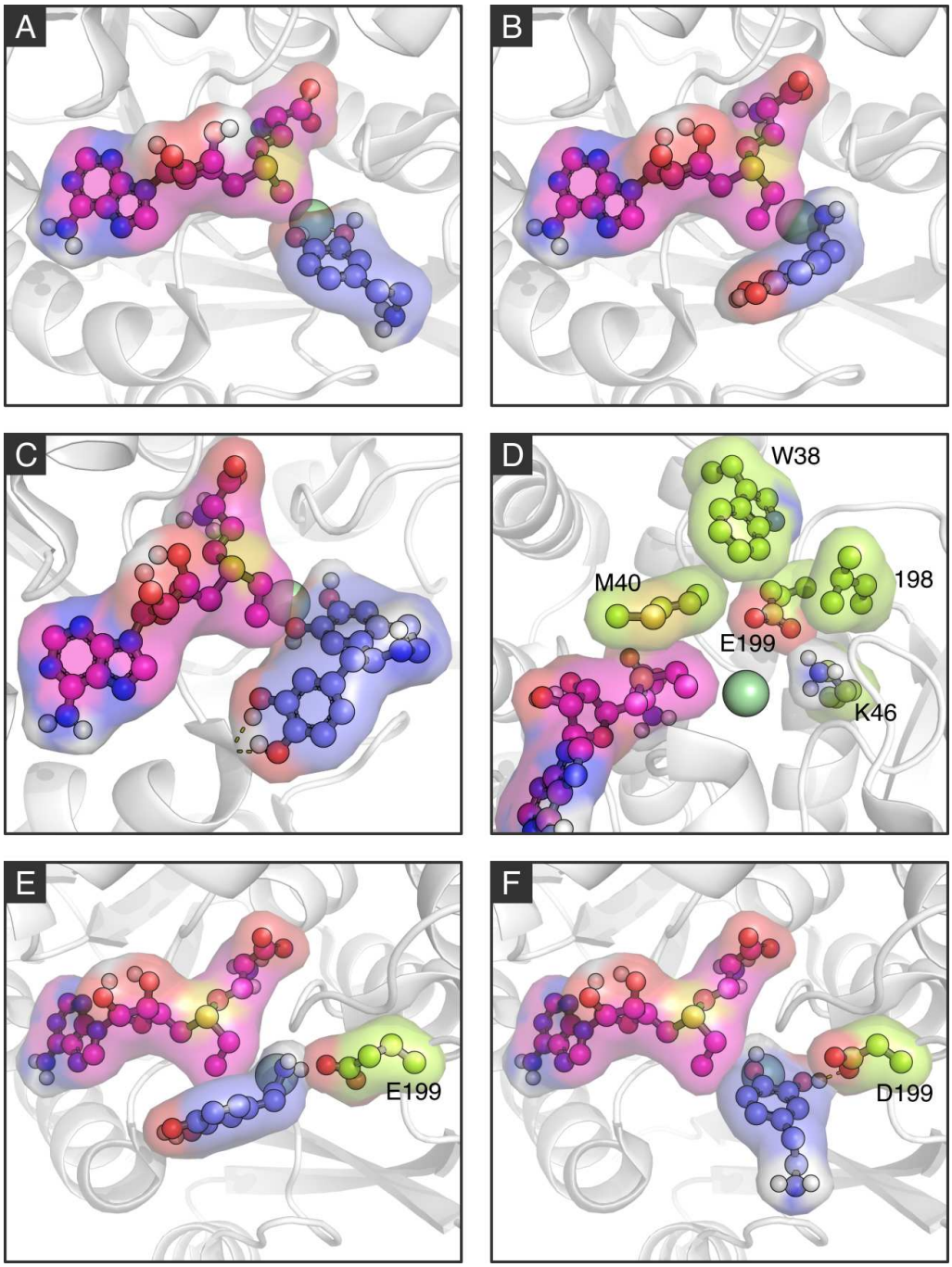
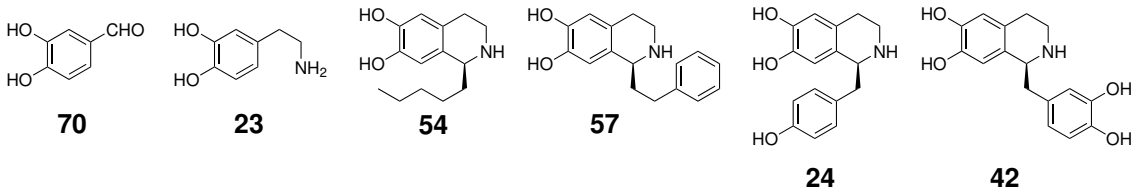


Figure 4.4: Results of docking studies into the effect of the cofactor SAE on substrate binding to *Rn*COMT, and investigation of engineering options. Cofactors shown in magenta, substrates shown in blue, residues shown in green. Polar contacts shown as dashed lines. Magnesium(II) ion shown as a green sphere. A) *Rn*COMT with SAM conformation from crystal structure, docked with dopamine. B) *Rn*COMT with SAE conformation from docking, docked with dopamine. C) *Rn*COMT:SAE docked with **(*S*)-42**. D) Residues considered for engineering. E) & F) the result of docking dopamine into E199 and D199 variants of *Rn*COMT:SAE, respectively.

from the crystal coordinates, and in complex with SAE, from docking.

Six substrates were docked into both models (Scheme 4.3). There was already experimental data for the ethylation of **70** and **23**, and **(*S*)-54** and **(*S*)-57**, suggesting that **70** would have some correct binding to the SAE complex, but the other three would have little to none. The docking largely agreed with these data.



Scheme 4.3: Substrates docked into *Rn*COMT:SAM and :SAE complexes.

Almost none of the substrates docked into the SAE complex with 'productive' conformations, defined by coordination of the catechol hydroxyls with the Mg^{2+} ion and proximity of an oxygen to the target carbon atom of the *S*-ethyl chain. Most instead bound the active site non-productively, i.e. with the catechol rotated away from the cofactor. In this position, the substrates may occasionally sample a conformation that brings the catechol in reaction range, but this conformation may be rare due to the new binding environment created by SAE. Compound **23** gave a different result, as in even the best-scoring result, the catechol was entirely removed from the active site and replaced by the ethylamine substituent (Figure 4.4B). In this position, the amine could be bound by the interactions that were meant for the catechol. It is worth noting a study which has highlighted the tendency of force fields, including those on which AutoDock is based, to over-estimate the energy of such electrostatic interactions.[180] However, the observations here appeared to corroborate the experimental results. Compounds **70**, **(*S*)-54** and **(*S*)-57** showed some ethylation in assays, potentially because of their aforementioned proximity to the cofactor, while **23** showed absolutely none (Figure 4.2). Moreover, docking the same substrates into the *Rn*COMT:SAM complex gave productive conformations for all four as the best-scoring results. This indicated that the electrostatic bias was not so strong as to exclude a productive conformation if it was favourable.

There was one notable exception, however: **(*S*)-42** was able to dock into the SAE complex with a productive conformation. Looking at the interactions with the rest of the

molecule, this was suggested to be a result of polar contacts between the catechol of the C-1 substituent and the amide nitrogen of K144 (Figure 4.4C). This interaction may help anchor the substrate in place. However, **(*S*)-24** should also have hydrogen-bonding capability in its substituent, but the same effect was not observed. The ethylation of these substrates had not yet been tested experimentally, but this observation suggested that conversion may be higher with **(*S*)-42** than any other substrate in the panel.

The inability of other substrates to dock productively in the SAE complex seemed to be due to steric hindrance, caused by the additional methylene in the ethyl chain. Space-filling representations indicated this atom invades the active site and blocks productive binding. It was hypothesised that if amino acid substitutions could relieve this hindrance, the ethyl-transfer may be improved. Two engineering strategies were explored. The first involved amino acid substitutions around the cofactor binding pocket, with the aim of allowing the additional carbon to rotate out of the active site. M40 forms a 'lid' over that part of the cofactor (Figure 4.4D), so was mutated *in silico* to serine, in order to create space while maintaining some of the character of that residue. However, an apo form of this mutant was unable to dock SAM or SAE in the correct conformation. This indicated that mutation of this residue would counter-productively disrupt cofactor binding.

The second strategy was to create more space on the opposite side of the active site, allowing the substrate to shift away from the cofactor while remaining coordinated to the Mg^{2+} ion. That area of the protein is composed of several residues which could be targets, including W38, K46, L198 and E199 (Figure 4.4D). Of these, E199 was the most promising. The carboxyl group of this residue's side chain hydrogen bonds the non-reacting substrate hydroxyl during catalysis. Previous studies have noted the effect of mutating the neighbouring residue, Y200, on the position of this side chain.[126] However, the authors concluded that the E199 side chain is flipped out of the active site by the mutation, as it would be during domain crossover with another monomer. It was decided to investigate the effect of changing E199 directly, in order to 'pull' the substrate away from the cofactor.

E199 was mutated to aspartate *in silico* and the resulting structure energy-minimised to resolve clashes. Examination of the model indicated no interactions were lost as a result of the mutation, so the tertiary structure would be expected to remain largely the

same. It was also assumed that cofactor binding would be unaffected by the change, so the E199D model was completed by transplanting the coordinates for SAE and SAM, in the conformations used thus far, into the mutant model. All six substrates from the original panel (Scheme 4.3) were docked into these new complexes. In a promising result, all were found to dock into the *Rn*COMT (E199D):SAE complex with productive conformations. Dopamine **23** provided an especially striking example. Despite being completely unable to bind with wild-type *Rn*COMT:SAE (Figure 4.4E), a productive conformation was restored in the E199D variant (Figure 4.4F), with the mutation appearing to almost reverse its disrupted binding. A concern at this stage, however, was that the reacting oxygen would be drawn away from SAE in a way which may hinder the reaction. Though quantum-mechanical techniques are able to predict the likelihood of reaction from a model, these were outside the scope of the project. The best validation for the hypothesis therefore remained experiments.

Alongside the improved conformations, comparisons of the binding energy score between the wild type and E199D showed that in all cases, binding was more favourable with the latter (Figure 4.5). This could be a combined effect of resolving the steric clash with SAE as discussed, and bringing the substrates deeper into the hydrophobic interior of the protein. Taken together, these findings provided a strong rationale to obtain and assay *Rn*COMT (E199D) for comparison with the wild-type.

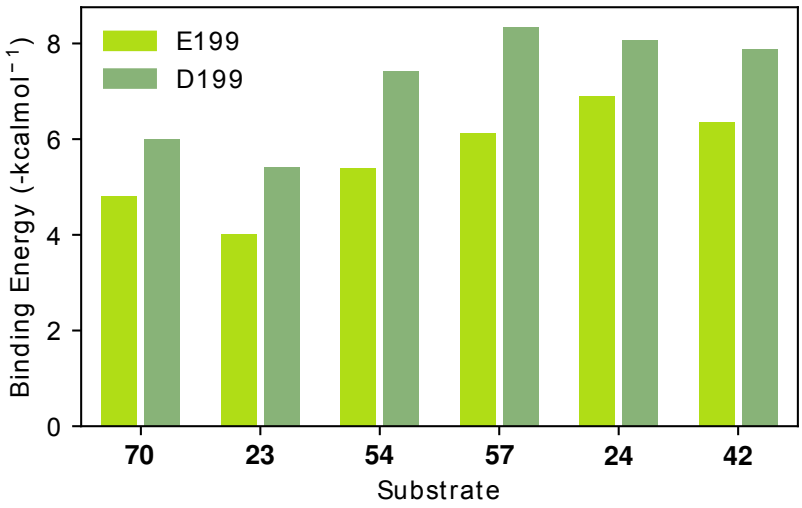
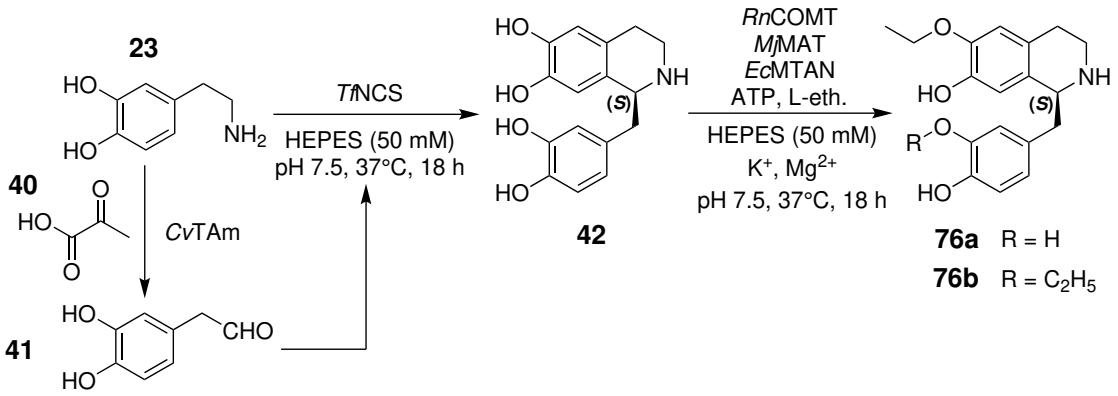


Figure 4.5: Comparison between binding energy scores for wild-type and E199D *Rn*COMT in complex with SAE. Energies shown are the scores of the highest-ranking docking solutions for each substrate.

4.6 Extended NCS-ethylation cascades

One of the predictions made by the docking studies was that (*S*)-norlaudanosoline **(*S*)-42** might have additional interactions with the *Rn*COMT:SAE complex that could improve binding and ultimately turnover. It also indicated that, despite carrying a hydrogen-bonding group in the same location, (*S*)-norcoclaurine **(*S*)-24** would not experience the same benefit. Therefore, assays were conducted to test these predictions.

For the first substrate, **(*S*)-42**, the cascade was split into two steps. The first step employed a method established by Lichman *et al.* to generate the THIQ in a one-pot reaction.[151] The transaminase *Cv*TAm converted dopamine **23** into 3,4-dihydroxyphenylacetaldehyde **41** *in situ*, with pyruvate **40** as the amine acceptor. Compound **41** then reacted with **23** via *Tf*NCS to give **(*S*)-42** (Scheme 4.4). Double the normal concentration of **23** was used (20 mM), along with 10 mM **40**. Both *Cv*TAm and *Tf*NCS were used as clarified lysates, at 20% and 10% (*v/v*) respectively (Assay VIII). The crude product of this reaction was centrifuged to remove precipitates and a sample of the supernatant tested by analytical HPLC. Analysis confirmed the formation of a new peak assumed to be the product (RT 6.7 min, Figure 4.6) and the reduction of the dopamine peak, although conversion was not complete.



Scheme 4.4: Two-step cascade to generate and ethylate norlaudanosoline **42**.

For the second step, as for previous assays, the crude product of step one was centrifuged and aliquots of the supernatant added to the enzymes and reagents for ethylation (Scheme 4.4) (Assay II). Unlike other substrates, **(*S*)-42** contains two catechols that could be alkylated by *Rn*COMT. Previous work in the group had established that for methyla-

tion, these alkylations are sequential, with the product dependant on the amount of SAM provided. To see if the same was true with ethylation, experiments were run with both 1.5 and 3 eqv. of ATP and L-ethionine, which would be expected to give **(*S*)-76a** and **(*S*)-76b** if the pattern held.

However, analysis by HPLC indicated that ethylation had not occurred to any great extent (Figure 4.6). The negative, ATP-only controls showed the peak assigned to **(*S*)-42** was maintained (RT 6.7 min), with small peaks immediately following that could represent products of background methylation reactions. The positive methylation controls, meanwhile, showed stepwise conversion of **(*S*)-42** to a new peak (RT 7.3 min). As the starting material was not fully consumed under either condition and the methylation is known to occur in stages, this peak was ascribed to the mono-methylated product. Finally, in the ethylation reactions, the only large peak was the remaining starting material. A new small peak was observed at 8.0 min in both traces. However, it was uncertain if this represented **(*S*)-76**, as there were a number of unknown peaks in this region, and it did not have any greater area in the reaction with 3 eqv. of ATP and L-ethionine than the reaction with 1.5 eqv.

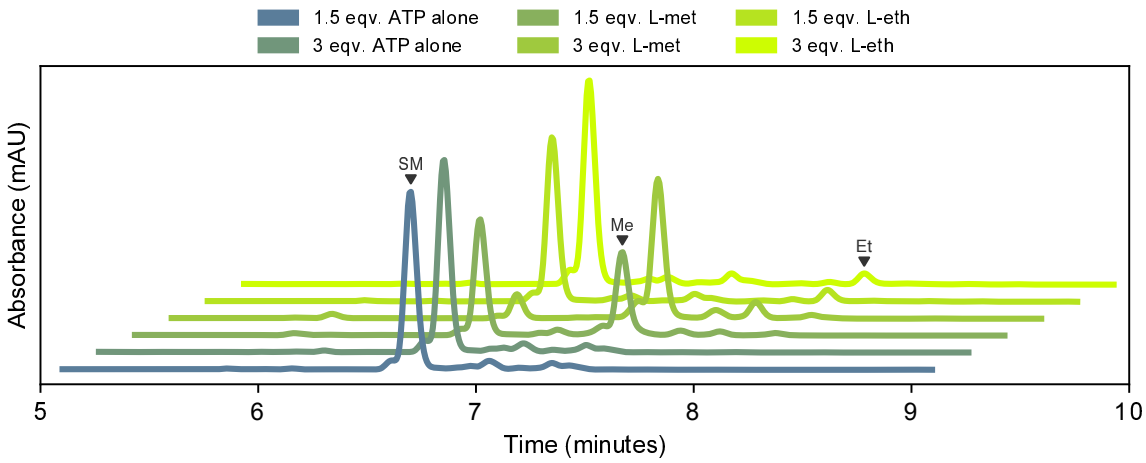
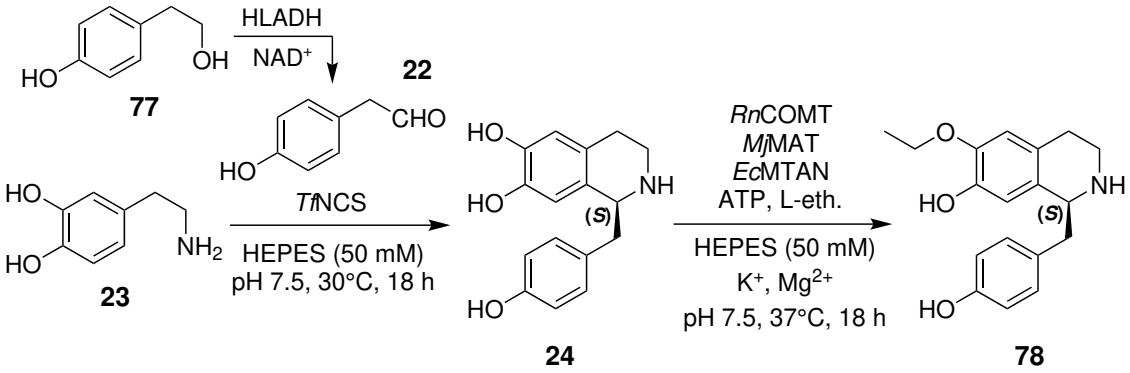


Figure 4.6: HPLC analysis of NCS-ethylation cascade towards **42** (Scheme 4.4). SM, Me and Et indicate the peaks representing starting material, methylated product and suspected ethylated product, respectively. Assay VIII conditions for NCS step, II for MT step. HPLC method A.

To clarify the outcome, the reaction containing 3 eqv. of L-ethionine was analysed by LC-MS. A peak corresponding to M+1 for the mono-ethylated product **76a** (*m/z* = 316) was detected (Appendix F). This confirmed the substrate was partially accepted. How-

ever, regardless which HPLC peak represented that product, the conversion was at no greater level than that seen for the other NCS-ethylation assays (Figure 4.3). Therefore, the hypothesis that **(*S*)-42** would have much higher ethylation conversion was not supported by these experiments. This may be a consequence of the limitation mentioned above, that docking algorithms can over-estimate the contribution of electrostatic effects, such has hydrogen bonds, to binding energy.

Ethylation assays of the other THIQ substrate, **(*S*)-24**, were conducted in parallel. The first step also involved additional enzymes to generate the aldehyde, this time Horse liver alcohol dehydrogenase (HLADH). Using nicotinamide adenine dinucleotide (NAD^+) as a cofactor, HLADH (lysate, 20% (*v/v*)) oxidised tyrosol **77** to 4-HPAA **22** *in situ*. This reacted with **23** via *Tf*NCS to give the THIQ product (Scheme 4.5). Previous work by by Dr. Méndez-Sánchez showed that the NAD^+ present in the enzyme lysates behaves catalytically, as the NADH produced by the reaction reduces oxidised dopamine, regenerating the original cofactor. Therefore, no additional NAD^+ was provided (Assay IX).



Scheme 4.5: Two-step cascade to generate and ethylate (*S*)-norcoclaurine **(*S*)-24**.

HPLC analysis of the first step showed relatively little conversion to **(*S*)-24**, with substantial amounts of **23** (RT 3.4 min) and **77** (RT 6.8 min) remaining (Figure 4.7). However, a new peak at 7.0 min indicated formation of the THIQ. As it was the second step of the assay that was the most important, this crude product was nonetheless taken forward to the second step.

The ATP-only negative control experiments showed maintenance of the starting material peaks, alongside a peak at 7.7 min indicative of background methylation. The methylation positive control corroborated this, showing disappearance of the peak cor-

responding to **(*S*)-24** and a new peak, again at 7.7 min. The ethylation reaction showed substantial remaining THIQ, an amount of background methylation, and a slight suggestion of a new peak at 8.3 min. To confirm that this represented the ethylated product **78**, the reaction was analysed by LC-MS, which detected a peak corresponding to M+1 for that product (*m/z* = 264, Appendix F). However, as before, the conversion was minimal. It was thus concluded that the hydrogen-bonding group on the substituent of this substrate was likewise having no dramatic effect on substrate tolerance.

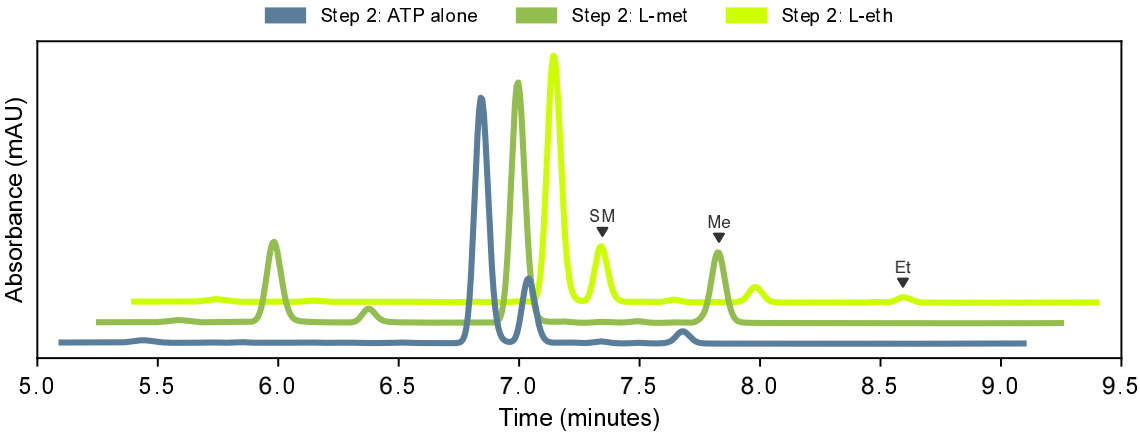


Figure 4.7: HPLC analysis of NCS-ethylation cascade towards **24**. Peaks are labelled according with the compounds they are hypothesised to represent, based on standards and previous similar products. SM, Me and Et indicate the peaks representing starting material, methylated product and suspected ethylated product, respectively. Assay IX conditions for NCS step, II for MT step. HPLC method A.

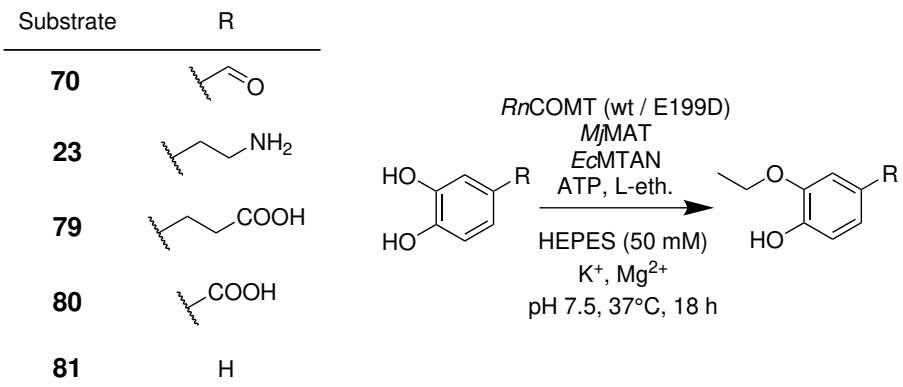
4.7 Cloning and evaluating *Rn*COMT E199D Mutant

The second prediction from the docking studies was that an E199D mutation of *Rn*COMT would facilitate ethylation. A synthetic gene encoding that mutant was thus designed, ordered and subcloned into pET-28a(+), by the same procedures used for cloning the novel MTs in Chapter 3. The resulting plasmids were subsequently verified by sequencing and transformed into *E. coli* BL21(DE3).

In order to directly compare the wild-type and mutant enzymes, it was decided to use purified enzymes, which would enable precise control of the final assay concentrations. Therefore, wild-type and E199D *Rn*COMT, along with *Mj*MAT and *Ec*MTAN, were ex-

pressed in *E. coli* and purified by nickel ion affinity chromatography. Comparable work by others had used *Rn*COMT, *Tk*MAT and *Ec*MTAN concentrations of 0.5, 0.5 and 0.01 mg/mL respectively.[123] The enzymes were therefore used at these concentrations, with ATP and L-ethionine at 10 mM and the substrates at 5 mM (Assay X).

Five simple catechol substrates were screened for ethylation by both the wild-type and mutant enzymes (Scheme 4.6). This was intended to detect any expansion in the substrate scope caused by the mutation, noting that in the docking study, the change was able to completely rescue binding of dopamine **23**. However, after analysis of these reactions by HPLC, only two substrates showed any acceptance by either enzyme. Furthermore, in both cases, the E199D mutant showed far worse conversion than wild-type *Rn*COMT (Figure 4.8).



Scheme 4.6: Assays to compare the activities of wild-type and E199D *Rn*COMT for the ethylation of a panel of simple catechol substrates.

There are several potential explanations for this finding. The mutation may have had a detrimental effect on active site binding not indicated by the docking study, or the greater distance between the target hydroxyl and the reacting carbon of the cofactor may have inhibited the reaction. The mutation could also have had unforeseen effects on protein stability.

Although other residues around the active site were considered during the docking study, none were as promising candidates for mutation as E199, and due to being more distant from the point of binding, their effect on the reaction would be even harder to predict. Around this time, another group reported that an M40A mutation (M40S was discarded during the modelling investigations because of its apparent detriment to cofactor

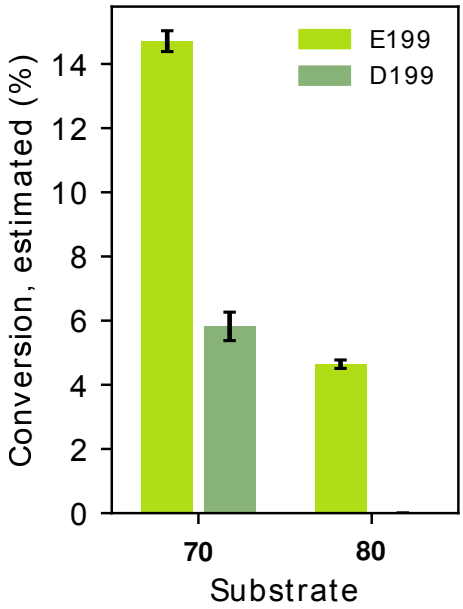


Figure 4.8: Analysis of ethylation assays comparing wild-type and E199D *Rn*COMT (Scheme 4.6). Error bars represent one standard deviation above and below the mean of three replicates. Substrates not shown gave 0% conversion for both varieties of *Rn*COMT. Assay X conditions (with ATP and **32a** at 10 mM). HPLC method A.

binding) is in fact beneficial to the binding of carboxy-SAM (Scheme 1.5.[132] However, the gains in conversion over wild-type COMT were modest. Creating more space in the active site, while maintaining substrate binding, may require a combination mutations which would take further experimentation to uncover. At this point, however, enzymatic ethylation of even simple catechol substrates remained limited.

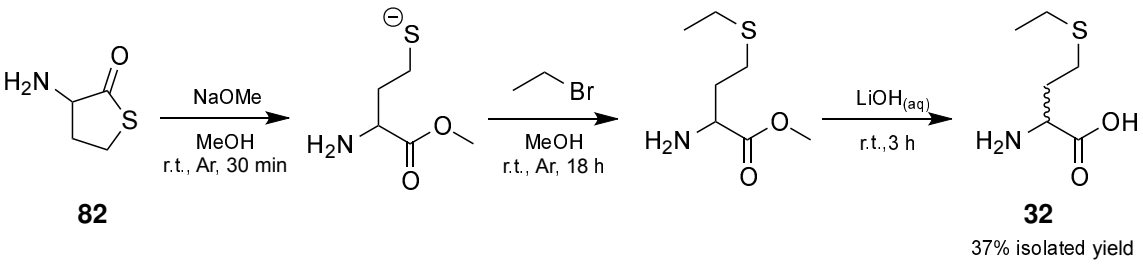
4.8 Synthesis of ethionine

L-ethionine is one of the few *S*-analogues of methionine available commercially in useful quantities, which made it attractive as a place to begin investigations. However, many other analogues which might be of interest are not available. Furthermore, the cost of L-ethionine, while not prohibitive for assays, would be substantial at scale. A number of synthetic methods to create analogues have been developed. Many, though, rely on expensive starting materials or difficult procedures, which are useful for initial experimentation, but impractical for the scaling-up this project aimed to achieve.[125,128,181,182] Therefore, in parallel with the enzyme investigations, a cost-efficient synthesis of ethio-

nine was sought. After validating the product against a commercial standard, this could then be generalised to more analogues.

4.8.1 Synthesis and evaluation of DL-ethionine as an ethyl donor

The first strategy investigated was reported by Dippe *et al.*[124] It begins with a thiolactone form **82** of homocysteine, and features a series of straightforward reactions via two intermediates (not isolated) to generate the racemic analogue **32** (Scheme 4.7). As a variety of alkyl halides are available, this reaction is theoretically versatile, and the low cost of **82** makes it economical to scale-up.



Scheme 4.7: Synthesis of DL-ethionine **32**.

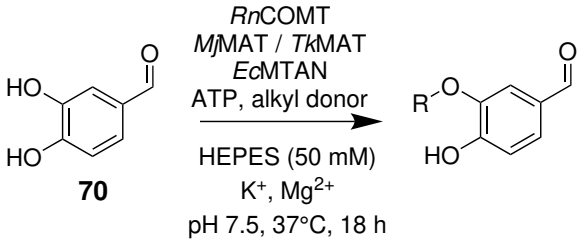
The synthesis of **32** was conducted according to the literature procedure.[124] Isolation of the product from the hydrolysis reaction involved addition of Dowex H^+ resin until the mixture was acidified, then washing away unbound salts with diH_2O and elution of the product with 10% $NH_{3(aq)}$. This gave **32** in 37% yield.

However, D-amino acids can be poorly tolerated by MATs.[113] A concern was that the D-isomer in the racemate could inhibit the MAT. The subsequent methylation by-product, *S*-adenosyl-D-homocysteine, is also cleaved approximately 40% slower by *Ec*MTAN than the L-form, potentially leading to accumulation of this inhibitor.[138]

To address this concern, assays were conducted with the substrate **70**, in which the alkyl donor was L-ethionine **32a**, synthetic DL-ethionine **32** or commercially sourced D-ethionine **32b**, with L-methionine **8a** as a control (Scheme 4.8, Assay II). Both *Mj*MAT and *Tk*MAT were tested, to determine if their tolerance for the D-enantiomer differed.

The control with **8a** indicated that both MAT enzymes were working, although the conversion by *Tk*MAT was approximately 20% lower than that seen in Figure 4.2 (Figure 4.9). For the ethylations, **32a** gave the highest conversion. The racemate **32** gave

Substrate	Enantiomer	R
8a	L	CH_3
32a	L	C_2H_5
32	DL	C_2H_5
32b	D	C_2H_5



Scheme 4.8: Ethylation assays to compare the tolerance of the enzyme cascade for enantiomers of ethionine.

only marginally lower conversion than that, despite the D-enantiomer **32b** giving the lowest conversion. This could be explained by the fact that two eqv. of each alkyl donor were added to the reaction, and the conversions were lower than 20%. This would mean that in the racemate, only 20% of the favoured L-enantiomer had been consumed by the end of the reaction, so the supply of **32a** in this mixture would not yet be limiting. However, the results with **32b** alone indicated that a racemate was nonetheless a sub-optimal form of the analogue. Therefore, routes to the pure L-enantiomer were also explored.

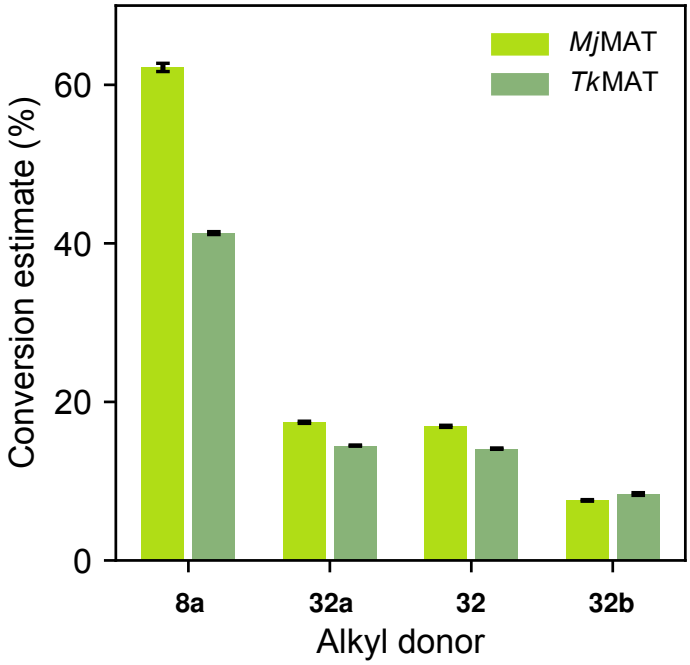
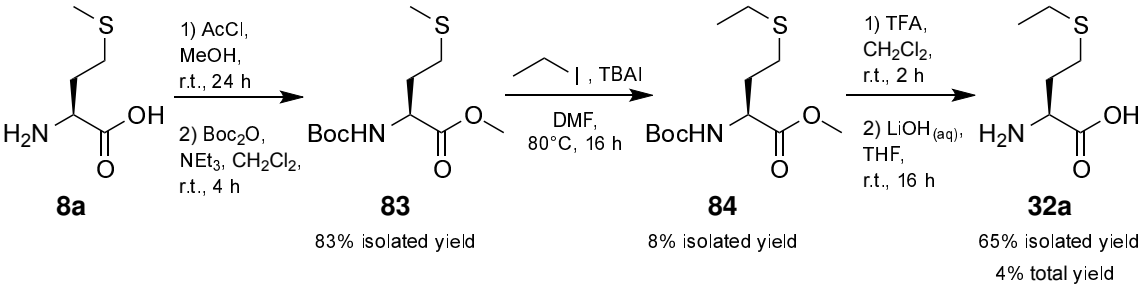


Figure 4.9: Comparison of product formation in ethylations of **70** with ethionine enantiomers. Error bars represent one standard deviation above and below the mean of three replicates. Assay II conditions. HPLC method A.

4.8.2 Attempted synthesis of L-ethionine

The second synthesis explored was reported in the work of Bhushan *et al.*.[183] This strategy centred around an unusual step, in which the methyl group of protected methionine **83** is replaced with the alkyl group of a halide (Scheme 4.9).



Scheme 4.9: Synthesis of isomers of L-ethionine **32a**.

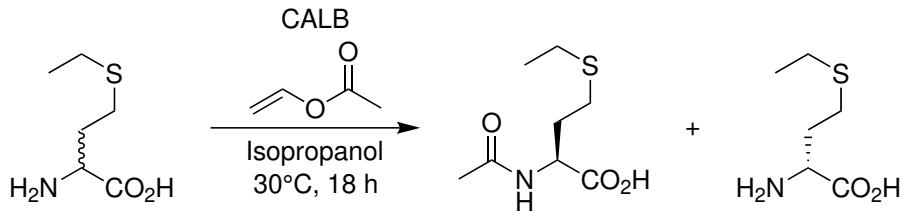
In the original synthesis, the halide in step two was allyl iodide. Replacing this with ethyl iodide, in an attempt to produce **84**, resulted in a sharp drop in the yield compared to that which was reported for the ally group. Following the rest of the synthesis gave **32a** with a 4% total yield. Fortunately, L-methionine **8a**, is available at low cost, and the unreacted starting material **83** could be recovered during the purification. Several attempts were made to improve the central step, such as by adding up to 16 eqv. of ethyl iodide over 48 hours, changing solvents, and purification and re-synthesis of the starting materials. None of these efforts were successful, and the obscurity of the mechanism remained an obstacle to optimisation. The comparative success of the reported procedure was proposed to be owed to the activation of the halide-adjacent carbon in the allyl group, which was not nearly as pronounced in the ethyl group. Therefore, this, too was set aside as a viable route towards the methionine analogue of interest.

4.8.3 Resolution of analogue racemates as a route to chiral products

The ongoing desire for a route to L-methionine analogues prompted exploration of chiral resolution, by which the desired enantiomer could be extracted from the more easily accessible racemate.

One strategy with potential employed the esterase *Candida antarctica* Lipase B (CALB). Melgar *et al.* reported a method wherein organic solvent reverses the enzyme's natural

action, turning it into a chirally-selective acyltransferase.[184] This strategy could be employed to acylate the free amine of **32** to give **85**, which could then be separated from **32b**. A solubility screen with **32**, found that the least polar solvent into which the compound would dissolve was propan-2-ol. Otherwise, the literature method reported by Melgar *et al.* followed as closely as possible, substituting orbital shaking for magnetic stirring for practical reasons (Scheme 4.10).



32 **85** **32b**

Scheme 4.10: Attempted resolution of DL-ethionine **32** using CALB.

The crude reaction was monitored by TLC and ^{1}H NMR spectroscopy. After 18 h, there was no indication that the product **85** had formed. The difference between the substrate in the publication and DL-ethionine was substantial, including a different acceptor atom, which may have been the root of the failure.

At this stage, then, no synthesis of ethionine satisfied all criteria. The route which provided racemic **32** was the best so far, as it was a one-pot procedure which gave a reasonable yield. However, there were ongoing concerns about the impact of the D-enantiomer on the already modest conversions. Therefore, all enzymatic ethylations in this chapter ultimately relied on commercially sourced **32a**.

4.9 Conclusions

Although ethylation had been intended as a simple model of alternative alkylation, the work in this chapter showed that the the existing MT system did not readily tolerate L-ethionine, possibly due to both sterics and reactivity. Furthermore, aside from the general benefit of accessing more diverse products, ethylation would not provide any specific product or functionality that would demonstrate the value of the system at a larger scale. These investigations had, however, given a reliable synthesis of DL-ethionine that could

be generalised to other methionine analogues. It had also given information on the limits of the system, and ideas on how enzyme engineering might be able to overcome those limits, even if the strategies employed so far may have been unsuccessful. From here, the focus of the project therefore shifted to understanding how the MT system might be improved for alternative alkylation, and how it might be used to generate molecules with specific functional value.

Chapter 5

Propargylation

5.1 Introduction

The work with enzymatic ethylation gave some important insights into the MT enzyme cascade. It provided data on the substrate range, some notion of the challenges involved in alkylations beyond methylation and, through modelling, mechanistic hypotheses on the bases of these obstacles. For the next phase of the project, however, attention shifted to a different MT-transferrable alkyl fragment: the propargyl group. This fragment was noteworthy for two reasons. Firstly, a propargyl-bearing SAM analogue would benefit from activation of the sulfonium-adjacent carbon, which could encourage the reaction relative to the ethyl group. Secondly, the terminal alkyne would open the possibility of click chemistry with the target compound.

Click chemistry describes a group of reactions which allow easy linking of modular chemical units.[185] The philosophy of the authors who coined the term was that the focus of synthetic chemistry on challenging, carbonyl-based reactions was slowing the pace of drug discovery. They instead proposed a set of reliable, high-yielding reactions that could generate the greatest diversity of compounds with the least difficulty.[185] Many click reactions involve either alkenes or alkynes, due to their accessibility and reactivity.[186] The reactions' selectivity affords great flexibility in the components being linked, and thus makes them applicable to everything from small-molecule synthesis to macromolecular conjugation.[186]

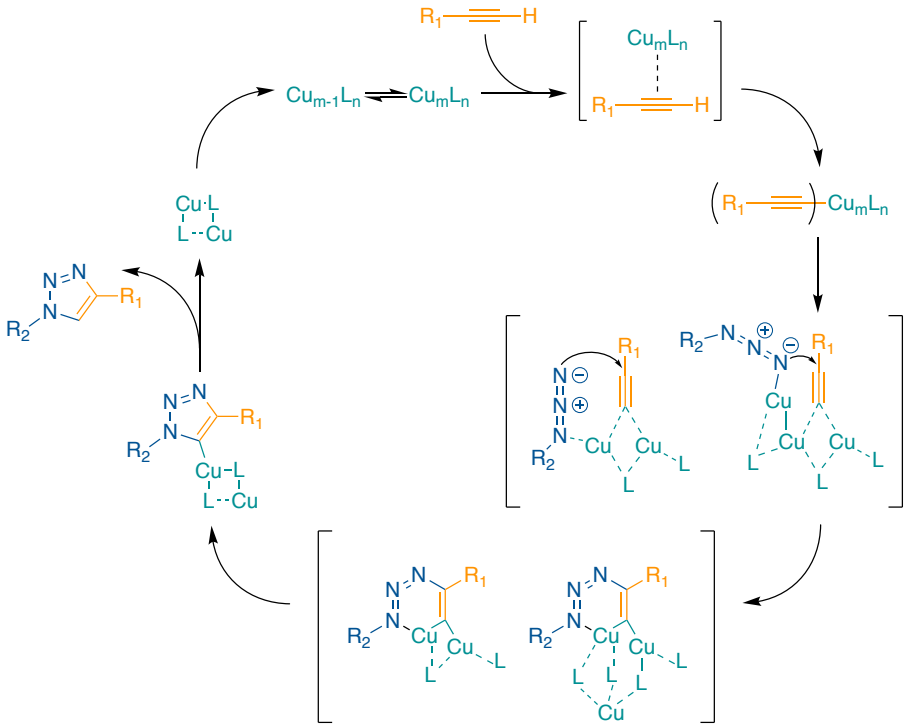
Among this toolkit, the copper(I)-assisted azide/alkyne cycloaddition (CuAAC) reac-

tion has risen to special prominence. CuAAC is highly thermodynamically favourable (-20 kcal mol^{-1}), proceeding rapidly at room temperature in aqueous or organic solvents. As the two component groups, azides and alkynes, are bioorthoganol, reactions can occur selectively even within complex biochemical environments.[187] However, there are some limitations: oxidation of the catalytic copper(I) to copper(II) terminates the cycle, while free thiols and triphenylphosphanes reduce the azides.[187] Strain-promoted azide/alkyne cycloaddition (SPAAC) is a related alternative, invented for situations where copper is undesireable (e.g. in living cells).[188] Instead of a catalyst, the reaction is driven by situation of the alkyne bond in a large aliphatic ring, which strains the *sp*-hybridised bonds.

A mechanistic overview of CuAAC is given in Scheme 5.1. First, copper(I) forms a π-complex with the alkyne. This lowers the pKa of the alkyne, allowing copper(I) to replace the terminal hydrogen. The azide then joins the complex, which may be di- or tri- nuclear with respect to copper. This positions the azide for nucleophilic attack of the alkyne, giving six-membered ring intermediates. A second attack by the azide gives the 1,4-substituded-1,2,3-triazole product with a copper(II)-ligand adduct, which is then ejected. A reducing agent, such as sodium ascorbate, regenerates copper(I) and the cycle repeats.[187]

The triazole produced by this cycle is itself considered a privileged scaffold, appearing in several current and prospective drugs.[189] CuAAC has thus been applied in drug discovery to generate arrays of triazole-containing analogues which can then be screened for activity.[189]

In some applications, the 1,2,3-triazole group can replace similar (bioisosteric) groups in compounds which are promising, but difficult to access.[190] The group is notably similar to the amide bond, but resistant to peptidase activity. This makes it attractive for inclusion into peptide or peptide-mimics. In a striking example, Grob *et al.* generated a series of analogues of the tumour-labelling ligand MG11, in which amide bonds had been replaced with 1,2,3-triazoles.[191] One analogue was found to have not only higher resistance to degradation, but also 2.6-fold higher tumor uptake, by virtue of a 10-fold higher receptor affinity. However, crystallographic and computational studies by Dorian *et al.* have noted differences in R-group angles and hydrogen bonding capacity between amide and 1,2,3-triazole groups, suggesting limits to their interchangeability.[192]



Scheme 5.1: Proposed catalytic cycle of CuAAC reaction. L = ligand. Adapted from Neumann *et al.* (2020)[187]

Aside from its own properties, the triazole product of CuAAC reactions can be a useful rigid linker between pharmacophores. Maurya *et al.* created a diverse library of macrocyles using monosaccharides with allyl and either propargyl or azide functionalisation. CuAAC reactions in the first phase created linear molecules tethered by 1,2,3-triazole groups, permitting subsequent intramolecular metathesis of the olefin moieties to give cyclic products.[193] The carbohydrate structures on either side of the ring could thus be varied to easily give an array of macrocycles.

As the scale of the linking partners increases, the relevance of the 1,2,3-triazole group itself receeds, and the value of the reaction lies in what it can bring together. A common theme of research has been the linking of effectors, such as trackers for medical imaging or drugs, with targetters that guide them to a specific location. This has included partners as large as cells. In one example, Takayama *et al.* exploited the fact that mammalian cells will uptake an azide-labelled sugar and incorporate it into their metabolism, resulting in the labelling of cell-surface glycans. A mouse cancer model was supplied with 'protected' analogues of this sugar, which would only be made metabolically avail-

able by the enzymes overexpressed in tumor cells. Subsequent intravenous injection with strained alkyne-labelled, dye-containing nanoparticles resulted in an *in vivo* SPAAC reaction between the azide-tagged tumor cells and the nanoparticles. The tumor could then be visualised by the fluorescence of the cargo dye.[194]

Click handles have also been attached to free biomolecules themselves. Honacharenko *et al.* developed a technique that installed alkyne linkers on oligonucleotides as a part of the automated solid-phase synthesis workflow.[195] Through CuAAC, these could subsequently be furnished with fluorescent tags for medical imaging, or functional peptides for targetting therapeutic oligonucleotides to their target sequences. CuAAC also holds promise in antibody-drug conjugation.[196] Typically, conjugation involves reacting cysteine residues on the antibody to maleimide groups attached to the drug molecules. However, this reaction can produce unwanted side-products and, even when successful, is slowly reversible. Therefore, click reactions are being explored as an alternative.[197] Recent tests have shown that a 'catch-and-release' method, employing reversible attachment of the antibody to a resin, was remarkably efficient at producing a model antibody-drug conjugate. In this strategy, CuAAC was used to join alkyne-tagged cysteine residues on the antibody to an azide-tagged drug payload.[197]

There is undeniable value, then, in having methods to attach alkyne groups to molecules and materials. The MT enzyme cascade used in this project and beyond could be a means to achieve that. Installing 'click-able' handles with enzymes would have the added benefit of selectivity for A) the substrate and B) the target atom on that substrate. MTs would be able to operate in complex biological environments, for example in culture media, or as a step in an *in vivo* biocatalytic pathway. An application relevant to this group's work could be transferring handles to THIQ compounds, which, given the medicinal potential of that family, might eventually form a step in drug conjugation to antibodies or other carriers.

Publications by other groups had already given precedent for this idea. The seminal work by Singh *et al.*[125] showed that a propargyl-bearing analogue of L-methionine (*S*-propargyl-L homocysteine, **86a**) was accepted by *Hs*MAT2A. The respective L-selenomethionine analogue, **87a** was also tolerated. In both cases, the resulting cofactors were transferred by the MT RebM to a substrate, although conversion was far higher for the

selenomethionine analogue (55%) than the methionine analogue (5%) (Figure 5.1).[125]

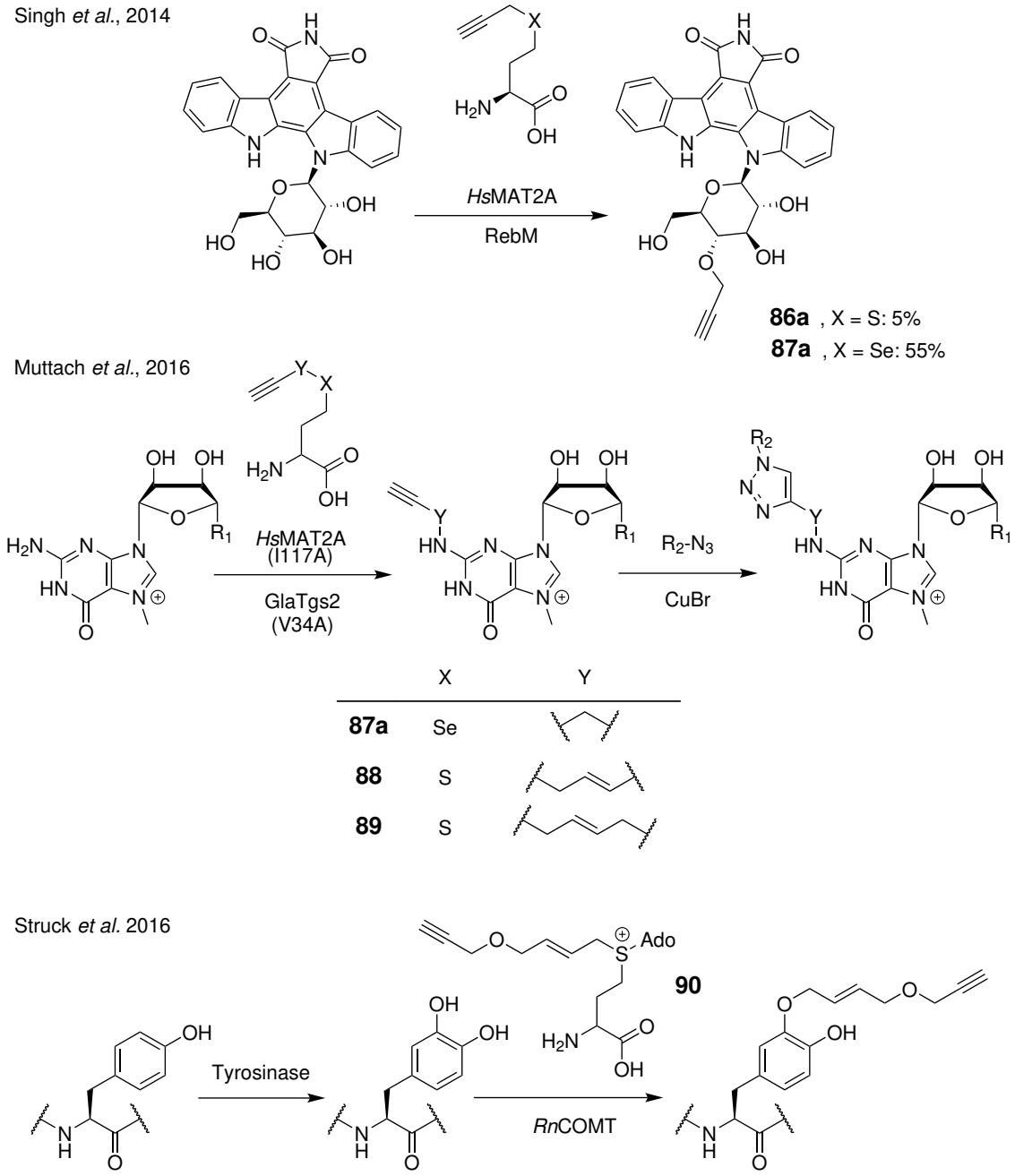


Figure 5.1: Examples of enzymatic cascades to install click functionality into small and macromolecules.

Muttach *et al.* applied the same concept to labelling the 5' cap of RNA.[198] Alkyne-bearing analogues **87a**, **88** and **89** were transformed into their cognate SAM cofactors by a mutated *Hs*MAT2A. The variant (I117A) had been previously developed by another group performing comparable work, and was shown to improve acceptance of

bulky methionine analogues.[128] The tags were thus transferred to the 7-methyl guanosine residue of an RNA cap by an enginnered MT, GlaTgs2 (V34A). This work went a step further, though, in demonstrating a subsequent CuAAC reaction between the alkyne-labelled RNA and azide-tagged biotin and dyes (Figure 5.1).[198] The most similar work to the prospect for this project was undertaken by Struck *et al.*, who employed a tyrosinase-COMT cascade with the alkyne-bearing SAM analogue **90**.[199] The tyrosinase was able to convert tyrosine residues on peptides into L-DOPA, to which the alkyne-bearing propargyloxybut-2-enyl (POB) group of **90** could be transferred by *Rn*COMT (Figure 5.1). Peptides and proteins labelled in this manner could theoretically then be tagged or linked by CuAAC as in the previous examples.

These examples illustrated that the concept of using enzyme cascades to attach clickable handles to a variety of targets was well-established. However, there was a limitation to these studies. All relied on expensive starting materials, and often complex or hazardous syntheses, to generate their methionine or SAM analogues. The synthesis of Singh *et al.* relied on L-homocysteine, which at the time of writing cost £3080/g. That of Muttach *et al.* began with the more affordable racemic isomer, DL-homocysteine (£61/g), but still required hazardous liquid ammonia and sodium metal as reagents. Furthermore, the selenium-containing analogues were derived from L-selenomethionine (£2560/g). Struck *et al.* created the SAM analogue directly from SAH (£5840/g).

For research purposes, and sensitive labelling techniques performed with small amounts of material, these were likely acceptable costs. These syntheses also provided selenomethionine analogues, which have been consistently shown to give better conversions than methionine counterparts.[125,198] However, the vision of this project was towards scalability. In a hypothetical industrial or drug-discovery pipeline, expense of this magnitude would make the enzymatic process unfeasible. A gap was therefore perceived in the literature for a variant of the above cascades, that would be economical and scalable enough to produce larger amounts of click handle-tagged compounds.

The aims for this section of the project were therefore as follows:

1. To develop a simple, low-cost synthetic route to alkyne-bearing methionine analogues, beginning with *S*-propargyl homocysteine.

2. To build an MT cascade that can transfer this alkyne group to substrates. Initially this would be small, simple substrates, but the ultimate aim would be to transfer click handles to THIQs.

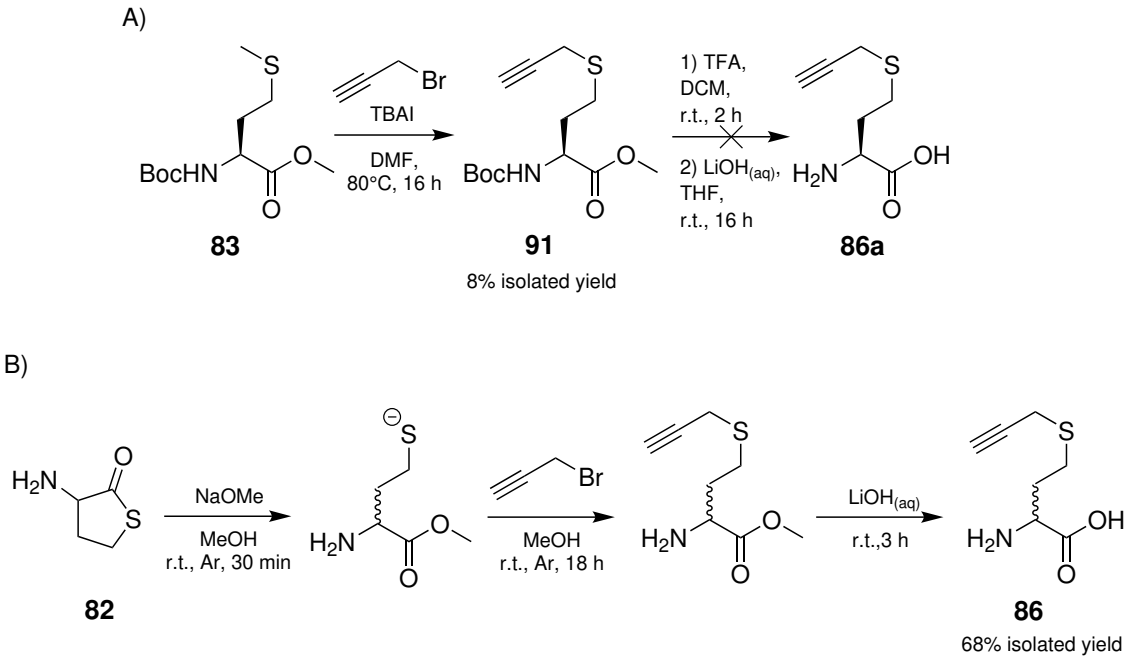
3. To demonstrate a CuAAC reaction with the product of said alkylations.

5.2 Synthesis of *S*-propargyl homocysteine

The lesson learned from the attempted synthesis of ethionine was that both racemic (DL) and chiral (L) forms are accessible. But while the latter is preferable in terms of enzyme acceptance, the yield of that synthesis was low. However, it was hypothesised that the greater chemical activation of the methylene carbon in propargyl bromide, as opposed to that in ethyl bromide, might improve this yield. Therefore, the synthesis leading to *S*-propargyl-L-homocysteine **86a** was attempted (Scheme 5.2A).

The route began with **83**, which had been made previously for the synthesis of L-ethionine **32a**, and proceeded to **91** via an adaptation of the published synthesis. Unfortunately, the yield of this step was as low as it had been in the synthesis of L-ethionine. Furthermore, repeated attempts to complete the subsequent deprotection step failed, and **86a** was never isolated. Possible reasons for this include loss of the relatively small amount of material or deterioration of **91** in the reaction conditions. This route to L-methionine derivatives thus not explored further.

The route based on the work of Dippe *et al.* was therefore revisited.[124] This synthesis had been successful for DL-ethionine **32**, but there were concerns over the effectiveness of the racemate as a substrate, due to poor tolerance of the D-enantiomer **32b** by the MT system (Figure 4.9). These concerns were exacerbated by the already poor acceptance of that analogue in general. However, the acceptance of *S*-propargyl-DL-homocysteine, the propargyl-bearing equivalent, was unknown. If high enough, it could outweigh the disadvantage of using a racemate. Therefore, the synthesis of this analogue was attempted, and gave the final product **86** in reasonable yield (Scheme 5.2B).



Scheme 5.2: Synthetic routes to A) *S*-propargyl-L-homocysteine **86a** and B) *S*-propargyl-DL-homocysteine **86**.

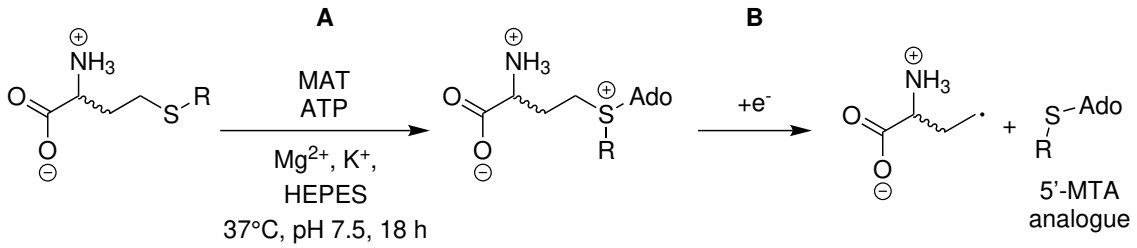
5.3 Accessing the *S*-propargyl cofactor

After synthesis of **86**, a small screen was conducted to establish which, if any, of the MAT enzymes available would accept it. Of the three used previously, *Mj*MAT and *Tk*MAT were included, while *Ec*MAT was excluded as its poor tolerance for methionine analogues is well-documented.[125] A plasmid encoding the MAT from *Ureaplasma urealiticum* had also recently been shared by our collaborators.[200] This enzyme was not suspected to have any particularly desireable properties: its acceptance of methionine analogues had not been studied before, and it was not from a thermophilic organism. However, it was found to express well in *E. coli* BL21(DE3), so was included in the screen.

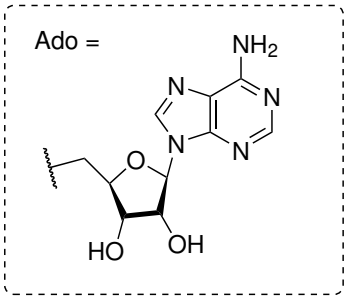
Four methionine analogues were assayed: *S*-propargyl-DL-homocysteine **86**, L-ethionine **32a**, synthetic DL-ethionine **32** and D-ethionine **32b** (Scheme 5.3). The ethionine isomers were included to gain a better understanding of this specific step of the ethylation cascade, which had not been examined by itself so far in the project.

The reactions were performed in similar conditions as the MT assays thus far. However, enzymes were purified to ensure exact concentrations (0.5 mg/mL final), and to

prevent interference from endogenous *Ec*MAT in the cell lysates (Assay XI). Analysis was performed using reverse-phase ion-pair HPLC. In this technique, a reverse-phase column (such as had been used for analysis of previous assays) is equilibrated with a buffer containing an ionic detergent, which modulates binding of compounds to the matrix and improves resolution of polar substances compared with standard RP-HPLC. The gradient and buffer composition used were as reported by Singh *et al.*.[125]



Met. analogue	Chirality	R
32a	L	CH_2CH_3
32	DL	CH_2CH_3
32b	D	CH_2CH_3
86	DL	CH_2CCH



Scheme 5.3: Assay of cofactor analogue production by MAT homologues. First step (A) is the desired reaction, second step (B) is the passive breakdown of product to homoalanine radical and the respective 5'-MTA analogue.

Peaks in the HPLC traces were identified with reference to standards (ATP and SAM) and data reported in the aforementioned publication.[125] Peaks suspected to represent cofactors produced from **32a**, **32** and **32b** (RT 13.2 min) presented with preceding shoulders (Figure 5.2A). These were likely caused by overlap with cofactor degradation products analagous to 5'-methylthioadenosine (Scheme 5.3). However, as the breakdown products could only come from the cofactors, and rate of breakdown was likely similar across reactions, it was decided that for the purpose of comparison between MATs, both the desired product and breakdown product peaks could be considered 'product peaks'. The productivity of the reactions were therefore expressed as the sum of these two peak areas.

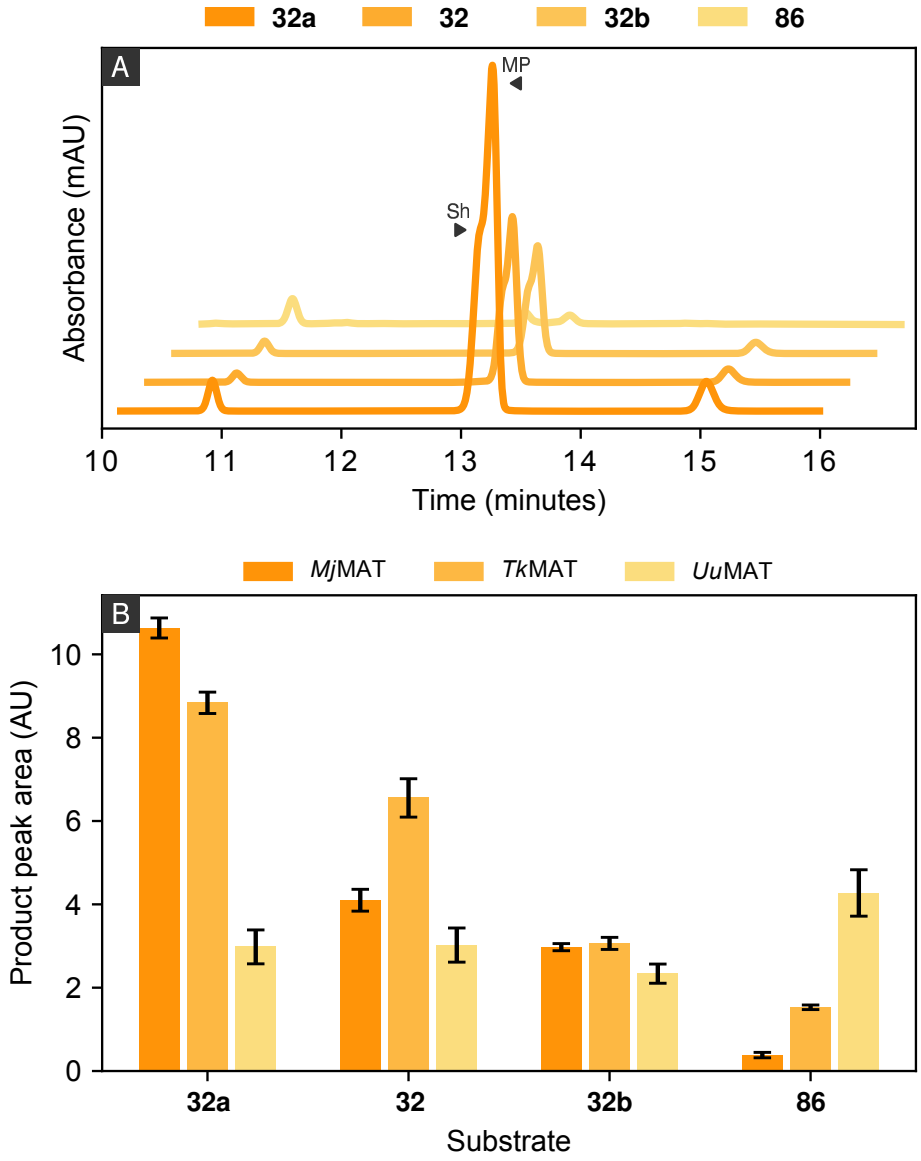


Figure 5.2: Acceptance of selected methionine analogues by a panel of MAT homologues. Note that 'product' here represents both the cofactor analogue and the resulting 5'-MTA analogue due to peak overlap. A) Illustrative HPLC traces, showing product peak (13.2 min) for each analogue in *Mj*MAT assays. MP and Sh indicated main peak and shoulder, respectively. B) Product peak area generated by each MAT homologue across the set of methionine analogues. Error bars represent one standard deviation above and below the mean of three replicates. Assay XI conditions. HPLC method B.

The result of the screen indicated stark differences in methionine analogue acceptance for each of the three MATs (Figure 5.2B). The patterns shown by *Mj*MAT and *Tk*MAT were largely consistent with previous results: **32a** was accepted best, **32b** the worst, and **32** intermediately between the two. For *Mj*MAT, however, the decline from **32a** to **32** was more dramatic than for *Tk*MAT (-61% vs. -26%), despite both having the same productivity with **32b**. This could indicate that for *Mj*MAT, **32b** is a more effective competitive inhibitor of **32a**. Both homologues accepted **86** least of all the analogues, though *Tk*MAT was more tolerant.

*Uu*MAT showed a very different acceptance profile. Compounds **32a**, **32** and **32b** were converted less by this homologue than the other two MATs, but the differences between their acceptance were small to non-existent. In other words, while the conversions were low, *Uu*MAT did not appear to be nearly as enantioselective as *Tk*MAT or *Mj*MAT. Furthermore, *Uu*MAT produced more of the propargyl-bearing cofactor from **86** than either of the other enzymes.

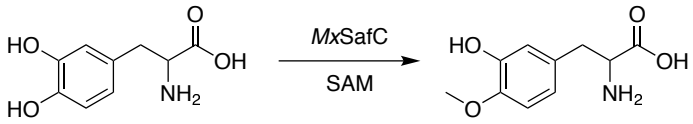
This result gave two promising conclusions. Firstly that there was an MAT in our repertoire which could accept an alkyne-bearing analogue, and secondly that this MAT appeared to be unselective towards the chirality of the alkyl donor. This latter was vitally important, because it removed the need to supply methionine analogues as the L-isomer. Instead, the DL-form **86** would be sufficient.

The first aim of this strand of work was therefore complete. The synthetic route towards **86** was quick (< 24 h), straightforward (three steps, two of which in one pot), low-cost (main starting material **82**: £1.06/g) and scalable (up to 3 g scale attempted). It was recognised that even though *Uu*MAT was the best of the options available for this purpose, the production of cofactor analogue observed here was low. Nonetheless, it provided a foundation to embark on the next aim of the project: building an MT cascade to transfer the propargyl group onto a substrate.

5.4 Incorporating the propargyl cofactor into MT reactions

As a first step, a propargylation assay was designed based on the conditions already established for ethylation. *Rn*COMT was tested, alongside SafC from *Myxococcus xanthus*.

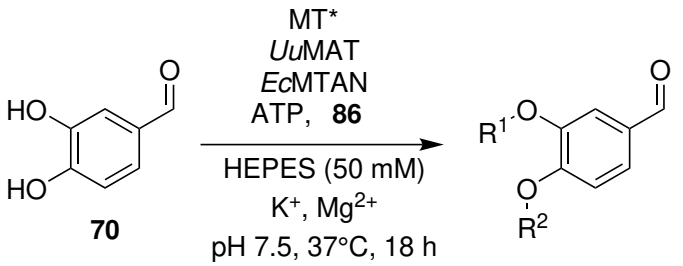
*Mx*SafC is another catechol-*O*-MT, and one of the enzymes studied in the MT collaboration noted in Chapter 2. It performs early step in the biosynthesis of the antitumor agent Saframycin MX1, where it methylates the 4-OH of L-DOPA **21** to give **92** (Scheme 5.4).[201] This is in contrast to the 3-OH selectivity of *Rn*COMT, potentially giving access to the alternate regioisomer if these selectivities were maintained.

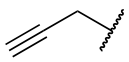
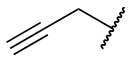


21 **92**

Scheme 5.4: Natural reaction of *Mx*SafC.

Purified enzymes were employed in order to reduce the batch effects and background reactions associated with cell lysates. However, the aim remained to ultimately transition back to using lysates once baseline activity had been established. Both MTs, *Uu*MAT and *Ec*MTAN were used at 0.5, 0.5 and 0.01 mg/mL concentrations as before. The substrate DHB **70**, and the ATP / **86** mix, were used at 5 and 10 mM, respectively (Assay X). The expected products from this assay was therefore **93a** for the *Rn*COMT reactions and **93b** for those of *Mx*SafC (Scheme 5.5).



	R^1	R^2
93a		H
93b	H	

Scheme 5.5: Assays of catechol-*O*-MTs for propargylation of a model substrate, **70**. *MT is either *Rn*COMT or *Mx*SafC. L-methionine **8a** was used in place of **86** for control reactions.

For analysis of these reactions, a new HPLC method (Method D) was developed which could take advantage of equipment now available to the group. This gave a shorter

overall method time and a higher pressure throughout, but otherwise the gradient used up until now (i.e. as in Method A) was maintained, and standards were used to confirm the new retention times of known compounds.

This analysis suggested that propargylated products had been generated by both MTs (Figure 5.3). A pair of new, overlapping peaks appeared in the traces for the reactions with **86** as the alkyl donor (RT 4.6, 4.7 min). These were present in neither the L-methionine nor the no-enzyme controls. The conversion to these new products was greater for *Rn*COMT than *Mx*SafC: 29% vs 20%. Furthermore, the ratio between the peaks, while difficult to quantify due to the overlap, was visibly different between the two MTs. Given this observation it was hypothesised that each peak represented one of the regioisomers **93a** and **93b**. However, it was not possible at this stage to identify which peak was which product.

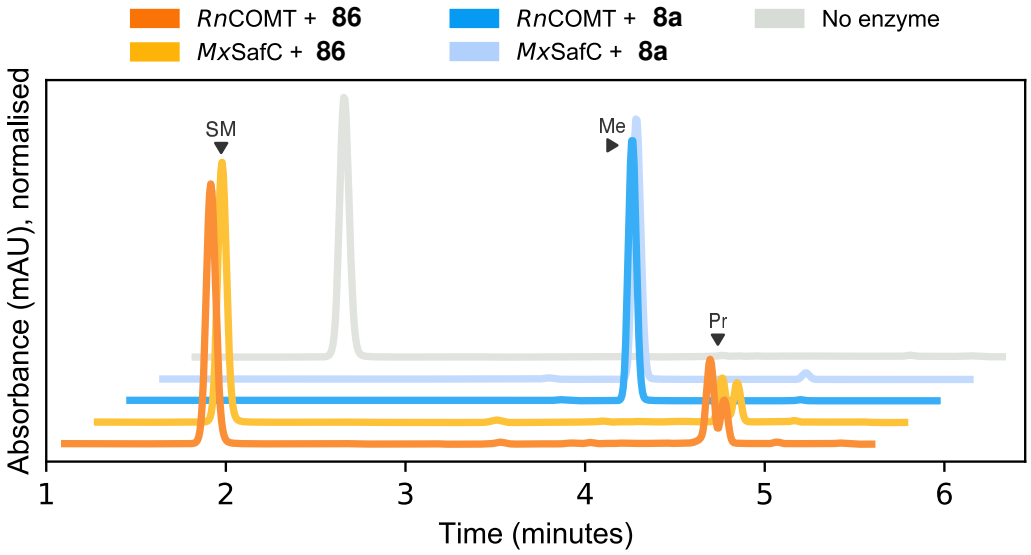


Figure 5.3: Analysis of enzymatic propargylation by *Rn*COMT and *Mx*SafC. Traces normalised to the highest peaks in the viewed region. SM, Me and Pr indicate peaks corresponding to starting material, methylated product and propargylated product, respectively. Assay X conditions (10 mM ATP & **86**). HPLC method D.

The conversions to these putatively propargylated products were comparable to those for ethylation with a similar enzyme system, if not higher. However, they remained substantially lower than the conversions for methylation. Another experiment was thus conducted to understand which variables might be limiting the propargylation reaction.

The setup was the same as the assay described in Scheme 5.5 (Assay X), except only *Rn*COMT was used as the MT, due to its higher conversion and apparently higher selectivity. Across the eight conditions of this experiment, the concentrations of MT, MAT or alkyl donor and ATP were doubled, either individually, in pairs or all together. This meant an increase to 1 mg/mL for the enzymes and an increase to 20 mM for ATP and **86**. The products of these reactions were analysed by HPLC, and the conversions calculated using the product peak area and remaining starting material peak area, as explained in Chapter 4.

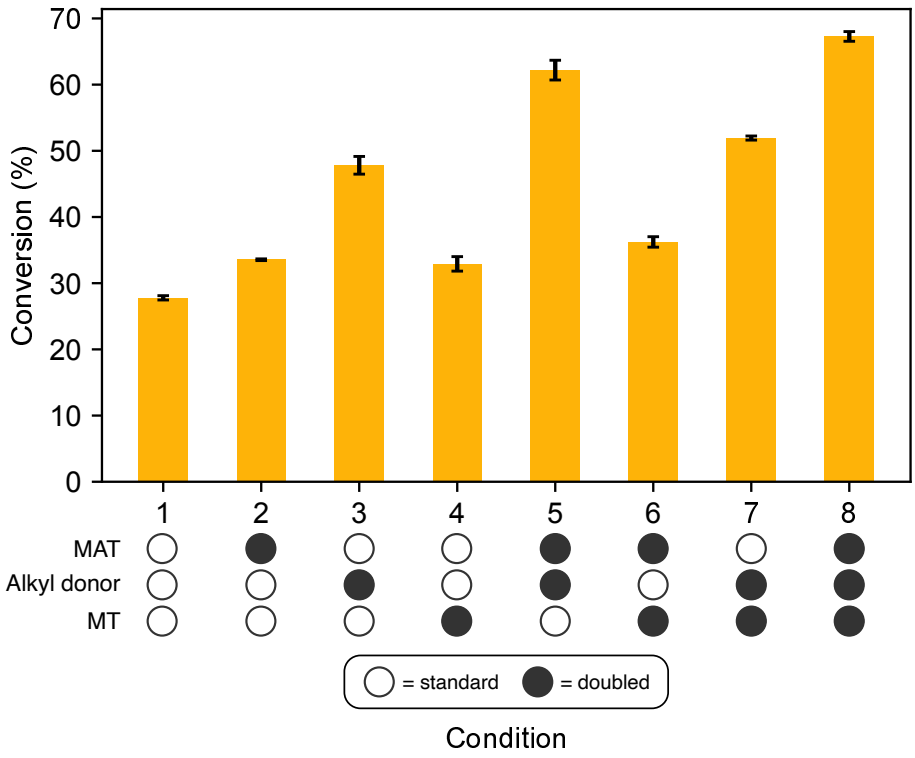


Figure 5.4: Analysis of limiting-factor experiment for the propargylation of **70**, as given in Scheme 5.5. Conversions calculated from HPLC peak areas of product and remaining starting material. Error bars represent one standard deviation above and below the mean of three replicates. Assay X conditions, with varing concentrations of *Rn*COMT, *Uu*MAT and ATP & **86**. HPLC method D.

To avoid confusion between absolute and relative increases in conversions, comparisons between the conditions are given as multipliers instead of percentages. Doubling the concentration of alkyl donor (i.e. **86** and ATP) produced the strongest effect, increasing the conversion by, on average, a factor of 1.75 (1.75x). However, this depended on

the other variables. If the quantity of MAT was already doubled, doubling the alkyl donor increased the conversion by 1.85-1.86x (condition 2 to 5 and condition 6 to 8) whereas if the MT was already doubled, the increase was only 1.58x (4 to 7). Nonetheless, this information was useful, as **86** was now readily available and the concentration could be increased in future assays.

Increasing *Uu*MAT concentration had a weaker effect, but followed the same pattern. From the baseline, doubling the MAT increased the conversion by 1.21x (1 to 2), but this went up to 1.30x if the alkyl donor was already doubled (3 to 5, 7 to 8). Increasing the MAT when *Rn*COMT was already doubled only produced a benefit of only 1.10x, however.

Doubling the concentration of COMT itself produced the weakest effect. From baseline, the increase in conversion was 1.18x (1 to 4). However, when either or both of the other variables were already doubled, this effect shrank to 1.08-1.09x (2 to 6, 3 to 7, 5 to 8).

The relationship between the first two variables was straightforward to explain. Increasing either would lead to more propargyl-cofactor, and if both were increased at the same time, there would be both a greater supply of the alkyl donor and a greater number of MAT enzymes to utilise it, so the combined effect would be greater than the sum of its parts. The pattern of increases with *Rn*COMT, however, was less transparent. The results indicated that it was the least limiting factor, as doubling its concentration produced the smallest effect on conversion. However, this effect was further reduced if either of the other variables were also increased. This defied the expectation that it would also act cooperatively, i.e. having both increased supply of cofactor, and increased MT to use that cofactor, should increase the conversion further than doubling the MT concentration alone.

The negative-feedback effect of the alkylation side product SAH was considered as an explanation. This would be present in all reactions, but the rate at which it was produced would be ultimately controlled by *Rn*COMT. Increasing the MT concentration when there is relatively less cofactor being produced through may lead to a modest increase in both the amount of **93** and SAH. However, increasing the MT when either the MAT or the alkyl donor has already been doubled, so there is a higher output of cofactor, may lead to significantly more SAH being produced than can be degraded by *Ec*MTAN, resulting in

a build-up of SAH which could limit the reaction. The next question, then, is why this does not seem to affect increases in the other variables in the same way. Furthermore, why does this effect not get 'worse' as the absolute conversion increases, i.e. why are the increases from condition 4 to condition 6, and from 5 to 8, both 1.08x, if the system is becoming 'overloaded' with inhibitory side-product, which a higher conversion should exacerbate.

At this stage, and with this limited information, there was not a clear answer. Regardless, the experiment had produced two important insights: that conversions can be improved by doubling the concentration of alkyl donor and ATP, and that at 0.5 mg/mL, the concentration of *Rn*COMT is not limiting.

It was apparent that deeper insight into the enzyme system was needed. Another assay was therefore conducted with similar conditions to the previous two, but with 20 mM ATP and **86** to increase the conversions (Assay X). However, instead of measuring the reaction after overnight incubation, samples were taken at one hour timepoints for six hours. Four parallel reactions were conducted across a tenfold range of *Rn*COMT concentrations, with the intention to observe the overall profile of the reaction over that time and how the concentration of MT affects it. This experiment was intended to give an indication of time to completion and not to provide kinetic parameters, due to the fact that the cofactor concentration could not be controlled.

HPLC method D was amended for this analysis, with the 5-70% B section of the gradient truncated to 5-40% to give method E. This expedited analysis further, but the peaks observed so far retained their approximate retention times.

The clearest observation from these data was that, at all concentrations tested, the propargylation reaction was slow. For comparison, published assays using *Rn*COMT and *Ec*MAT with THIQ substrates showed completion within 90 min,[156] while these reactions had not reached their endpoint after 6 h. At such long time scales, the stability of the enzymes could be an issue. Ways to promote this stability, perhaps by engineering, could be considered as a means to increase conversions overall.

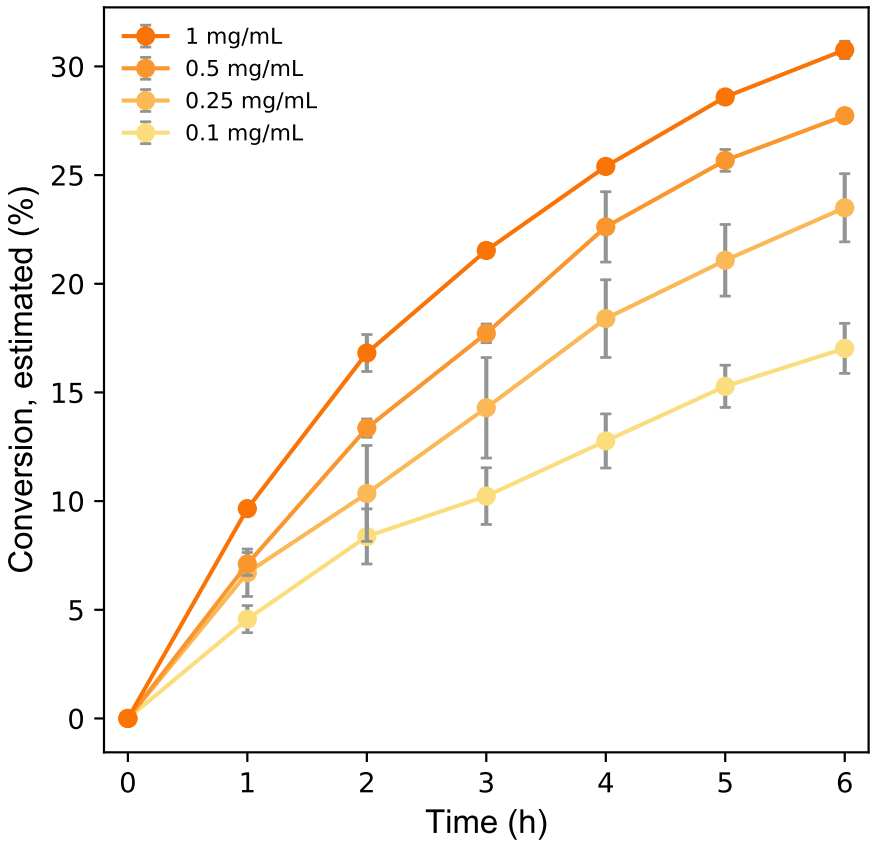


Figure 5.5: Analysis of propargylation assays with varying concentrations of *Rn*COMT, with reaction progress measured every hour for six hours. Conversions calculated from HPLC peak areas of product and remaining starting material. Error bars represent one standard deviation above and below the mean of three replicates. Assay X conditions (20 mM ATP & **86**). HPLC method E.

5.5 Purification of propargylation reaction products

To confirm the product of enzymatic propargylation, the reaction was scaled-up to 0.2 mmol with the aim of isolating **93**. Four eqv. (20 mM) of ATP and **86** were used with respect to substrate **70** (5 mM). While increasing the *Uu*MAT and *Rn*COMT concentrations beyond 0.5 mg/mL would also increase conversion, limited amounts of purified enzymes were available, and the aim was only to obtain sufficient product for analysis. The other conditions remained the same as represented in Scheme 5.5 (Assay X).

Following the reaction, the crude reaction product was quenched with methanol and centrifuged to remove precipitate. Analysis of the supernatant by HPLC showed 29% conversion (Figure 5.6A). The supernatant was then decanted and concentrated under vac-

uum to give a residue, which was resuspended in a minimal volume of 50% $CH_3CN_{(aq)}$.

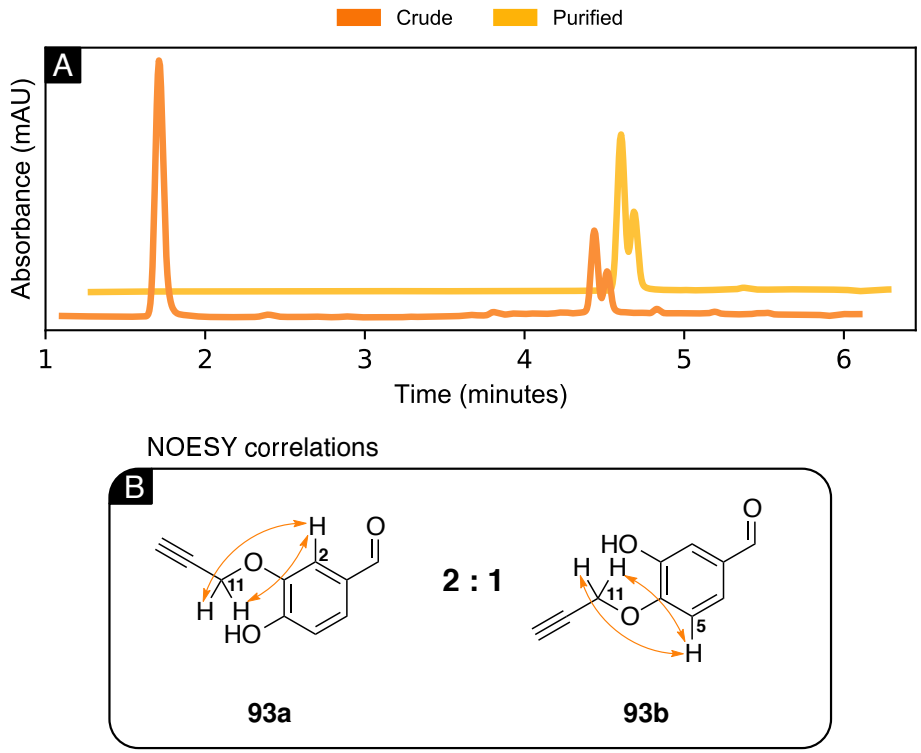


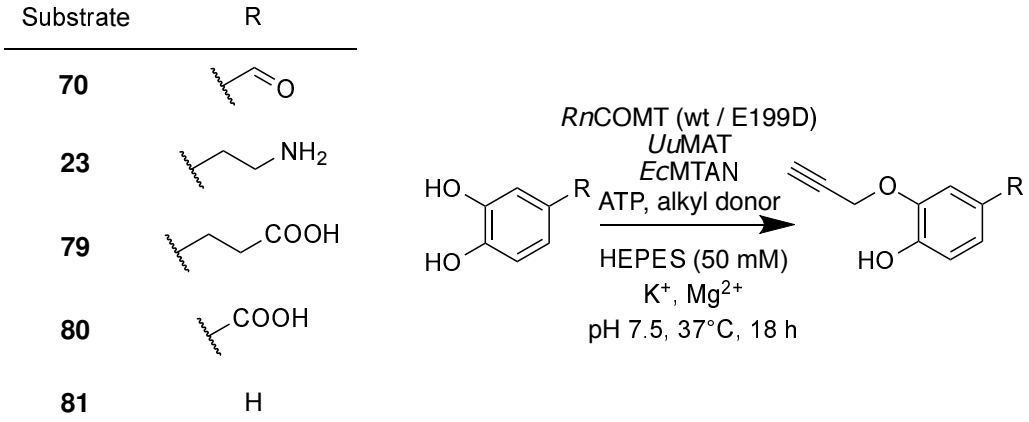
Figure 5.6: Analysis of products from an upscaled propargylation reaction. A) Analytical HPLC traces for crude and purified products. HPLC method E. B) Deduction of the product regioisomeric ratio from ^{1}H NMR NOESY experiments. Assay X conditions (20 mM ATP & **86**).

Preparative HPLC was used to purify the product, and analytical HPLC confirmed that the intended peak had been isolated (Figure 5.6A). For further characterisation, samples of the product were subjected to high-resolution mass spectrometry and ^{1}H NMR spectroscopy. Both confirmed that **70** had been propargylated at the catechol hydroxyls. Furthermore, in the latter, the signals assigned to 2-H, 5-H and 11-H were split, with each appearing as pairs of separate doublets. The integrals of these pairs were all in a ratio of 2:1. Additionally, NOE correlations were observed between the larger peak corresponding to 11-H and the larger peak for 2-H, with another weak correlation observed to the smaller peak for 5-H (Figure 5.6B, Appendix E.3). From these data, it was concluded that two regioisomers were present in the product, with *Rn*COMT producing a major meta-alkylated product **93a** and a minor para-alkylated product **93b** (Figure 5.6B). This confirmed what had been suspected from the HPLC traces, that the two peaks represented regioisomeric products.

Selectivity is one of the major advantages biocatalysis has over synthetic chemistry, allowing it to avoid waste and purification steps. Therefore, the low regioselectivity observed from these data presented another obstacle to developing the system as a useful biocatalytic process, and thus another potential target of enzyme engineering.

5.6 Revisiting the *Rn*COMT (E199D) mutant for propargylation.

The *Rn*COMT (E199D) variant was designed based on the docking studies described in Sectiion 4.5, with the intention of improving substrate binding to the *Rn*COMT:SAE complex. However, *in vitro* assays ultimately showed that the mutation had negative affects on the efficiency and substrate scope of ethylation (Figure 4.8). There was a possibility, however, that the outcome would be different with propargylation. The screen of wild-type and E199D *Rn*COMT against five catechol substrates was thus repeated using *Uu*MAT and **86**, with methylation reactions as controls (Scheme 5.6, Assay X).



Scheme 5.6: Assays to compare the activities of wild-type and E199D *Rn*COMT for the propargylation of simple catechol substrates. Alkyl donor was either **86** or **8a** (control).

Upon analysis of these reactions by HPLC, a very similar result to the ethylation screen was observed (Figure 5.7). Only two of the five substrates were accepted to any degree for propargylation: **70**, as expected, and **80**. In both cases, the E199D mutant was less productive than the wild type, although the decrease in conversion between the two was less dramatic here than for ethylation.

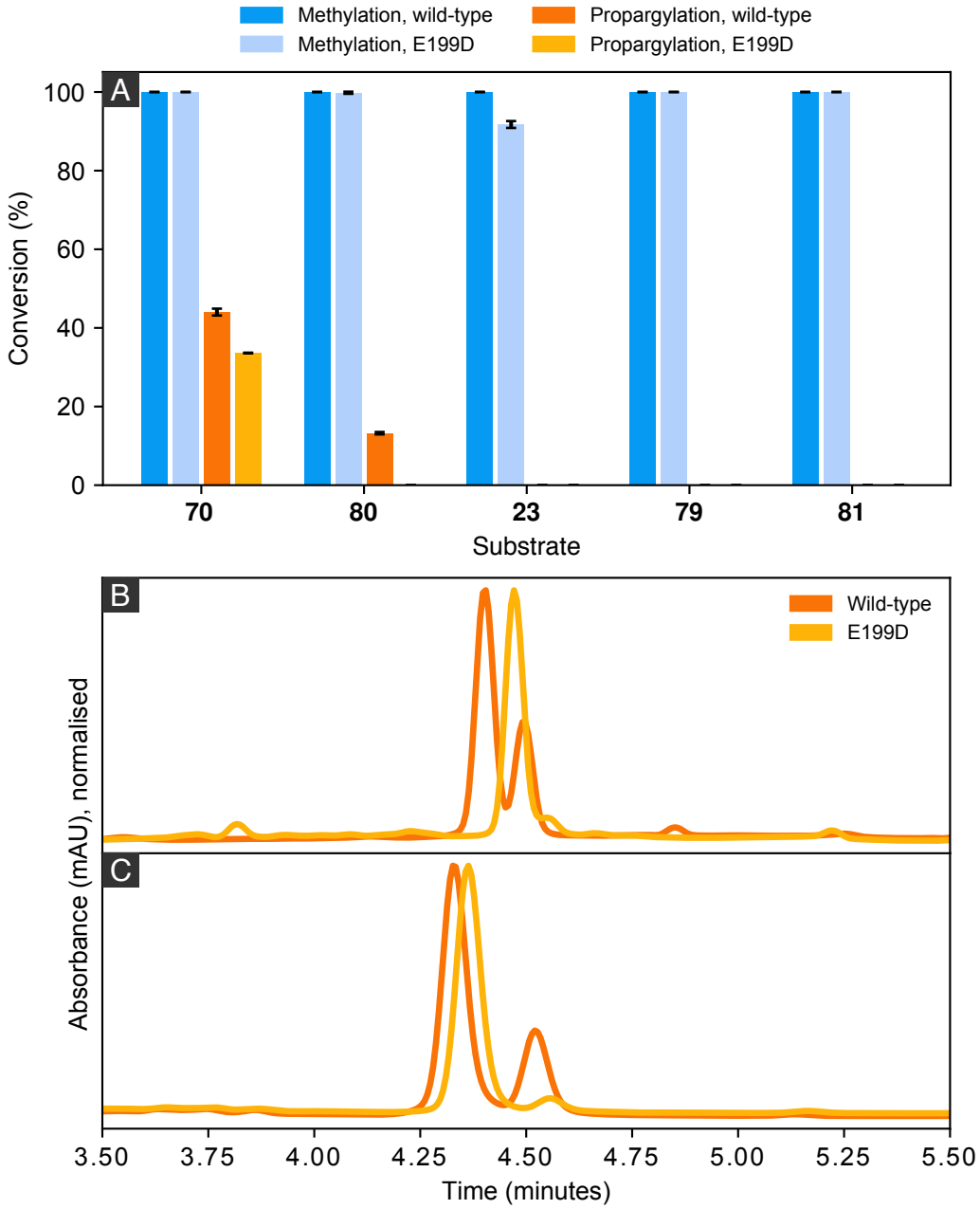


Figure 5.7: HPLC analysis of propargylation screen to compare substrate scope of wild-type and E199D *Rn*COMT (Scheme 5.6). A) Conversions of methylation and propargylation reactions with wild-type and E199D *Rn*COMT, calculated from HPLC peak areas of product and remaining starting material. Error bars represent one standard deviation above and below the mean of three replicates. B) & C) Analytical HPLC traces from the propargylation assay with **70**, using methods E and F, respectively. Assay X conditions (20 mM ATP & alkyl donor).

However, the HPLC traces themselves indicated that, for **70**, the ratio of regioisomer products differed between the two variants. The wild-type traces showed the typical major-minor peaks, while the E199D traces showed only one peak with a small tailing shoulder (Figure 5.7B). Misalignment of the peaks between these traces made it difficult to confidently state which, if either, of the regioisomers was not present. Therefore, a new HPLC method (method F) was developed to separate the regioisomer peaks, and samples from this assay were re-analysed (Figure 5.7C). Although the separation was not completely to the baseline, both peaks were now clearly distinguishable, showing the minor regioisomer **93b** was considerably reduced in the reactions with E199D. From the product peak areas, the ratio of **93a**:**93b** was estimated to be 65:35 for the wild type (in agreement with observations from ^{1}H NMR spectroscopic studies) and 92:8 for the mutant. Although conversions with the mutant were lower, this represented an encouraging result, demonstrating that the regioselectivity could be recovered through changes to the active site.

5.7 Translating the propargylation cascade to clarified lysate preparations.

Having established the products and improved conditions for propargylation with purified enzymes, the next stage involved using clarified lysates. Cell lysates are less labour-intensive to prepare, which makes reactions that use them more applicable to scale-up, especially for industrial applications. However, expression can vary between cultures of the same strain. Furthermore, lysates can contain compounds from the host organism's endogenous metabolism, which might interfere with the reaction. This was observed in the ethylation assays described in Chapter 3, where methylated products were generated due to L-methionine and SAM in the *E. coli* cytosol.

To understand how these factors might affect the reaction, an assay was conducted that attempted to mirror previous propargylation experiments with pure enzymes. The work in Chapters 2 and 3 used fixed volumes of lysates (10, 10 and 2% (*v/v*) for MT, MAT and MTAN, respectively), which were the standard conditions developed by others in the group prior to this project. These had been adequate for methylation, but from here

onwards, an effort was made to estimate the concentration of desired enzyme in each culture batch and adjust the volume of lysate added accordingly. A detailed explanation of the method can be found in Chapter 8. However, in brief, the total protein concentration of a lysate was measured with a Bradford assay. Following SDS-PAGE analysis, the proportion of this total represented by the enzyme was calculated as the ratio between the intensity of its band on the gel, and the intensity of all other bands (i.e. all other proteins in the lysate). The accuracy of this method was limited by the facts that bands can overlap on the gel, the borders of bands can be difficult to pinpoint and the background staining across the gels were not constant. However, it provided better mitigation of batch effects than using fixed lysate volumes.

An assay was therefore conducted using lysates of the three enzymes. The concentrations of *Uu*MAT and *Ec*MTAN were ~0.5 and ~0.01 mg/mL. For *Rn*COMT, a range of concentrations were tested between 0 and ~1 mg/mL. To isolate the effects of the cell lysate itself, a reaction using lysate from an expression host containing empty pET-28a(+) (empty vector, EV) was used as a negative control. The volume of EV lysate in this reaction was matched to the volume of the *Rn*COMT lysate used in the 1.0 mg/mL reaction. As before, 5 mM **70** was used as the substrate, alongside 20 mM ATP and **86** (Assay XII).

Analysis of these reactions by HPLC gave a surprising result: the more *Rn*COMT lysate was added, the lower the conversion observed (Figure 5.8A). In terms of enzyme concentration, this constituted a reversal of the pattern seen in Figure 5.5. However, in the HPLC traces (Figure 5.8B) it was noticed that as the lysate volume increased, the propargylation product peaks decreased (RT 4.5 min) but the methylation side-product peaks increased (RT 3.7 min). This made the effect difficult to explain in terms of a general inhibition of *Rn*COMT, as this would affect both alkylations.

It was hypothesised, then, that background methylation was in fact causing the decrease in propargylation. The MAT and MT would be expected to have higher affinities for the L-methionine **8a** and SAM in the lysate than any analogues. They might therefore be competitively inhibiting propargylation by being taken up into the enzyme cascade (Scheme 5.7). For the purposes of this project, this idea was termed 'methyl poisoning' of the cascade.

An equivalent experiment was then conducted with *Uu*MAT. Conditions were main-

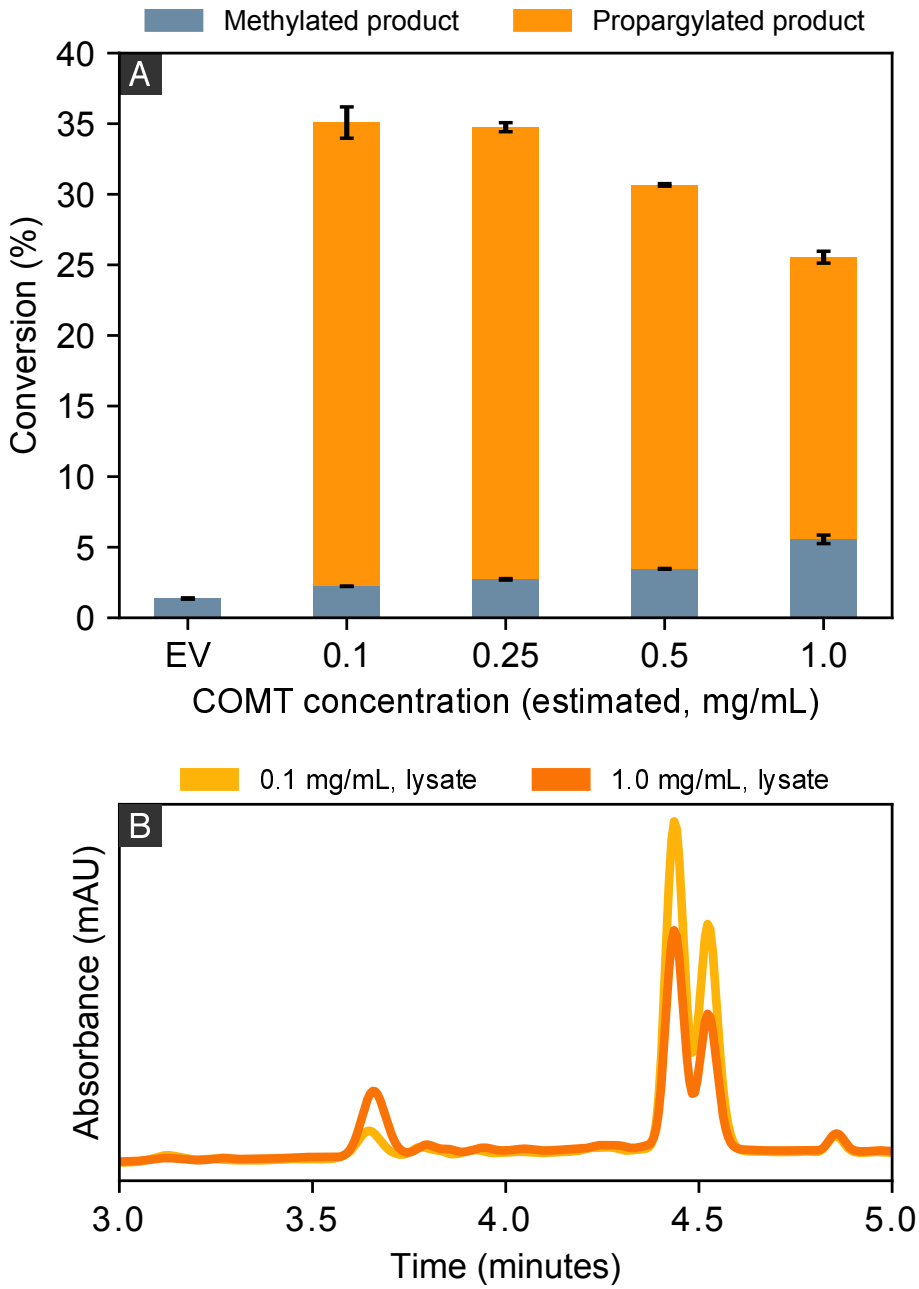
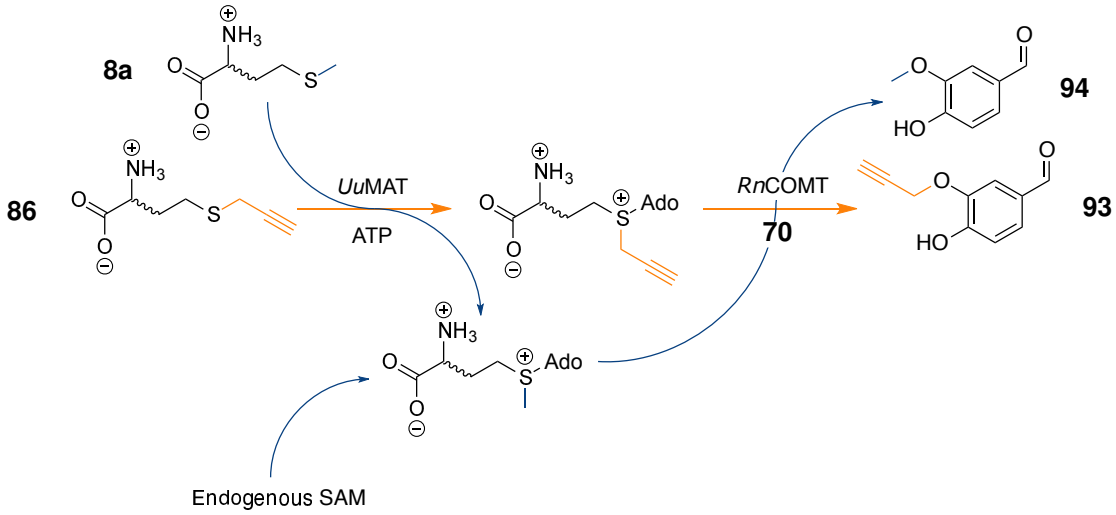


Figure 5.8: HPLC analysis of enzymatic propargylation reactions using clarified lysates, with *Rn*COMT at varying concentrations. A) Conversions to propargylated and methylated products, calculated from HPLC peak areas of product and remaining starting material. Error bars represent one standard deviation above and below the mean of three replicates. Bars are stacked. EV = empty vector. B) Traces for the reactions with the highest and lowest *Rn*COMT lysate concentration to illustrate the simultaneous increase of the methylated product (RT 3.7 min) and decrease of the propargylated product (RT 4.5 min). Assay XII conditions, with varying concentration of *Rn*COMT. HPLC method E.



Scheme 5.7: Illustration of how endogenous L-methionine **8a** and SAM from cell lysates enter the enzyme cascade and give the side-product **94**. Orange represents propargylation pathway, navy blue represents methylation pathway.

tained (Assay XII), but *Rn*COMT was used at ~0.1 mg/mL, as this was the best condition identified in the prior assay. The concentration of *Uu*MAT was varied between 0 and ~1 mg/mL, with an empty vector control that matched the 1 mg/mL condition in terms of volume of lysate added.

Analysis showed that between 0.1 and 0.5 mg/mL, increasing the *Uu*MAT concentration lead to increases in the formation of both methylated and propargylated products, but with diminishing returns for the latter (Figure 5.9).

However, at 1 mg/mL, conversion to the propargylated product declined slightly, while that to the methylated product continued to rise. This suggested that methyl poisoning was also occurring with the MAT. At lower MAT concentrations, poisoning may still be occurring, but would be out-competed by the availability of more enzyme. However, as the concentration increased, the benefit to propargylation would lessen while the poisoning became more pronounced, until the latter overtook the former and propargylation conversions began to decline. This inflection point likely occurred at higher concentrations for *Uu*MAT than *Rn*COMT because, as established in Figure 5.4, the ideal concentration of *Rn*COMT is lower, so the tapering off of benefits from increased concentration would happen earlier.

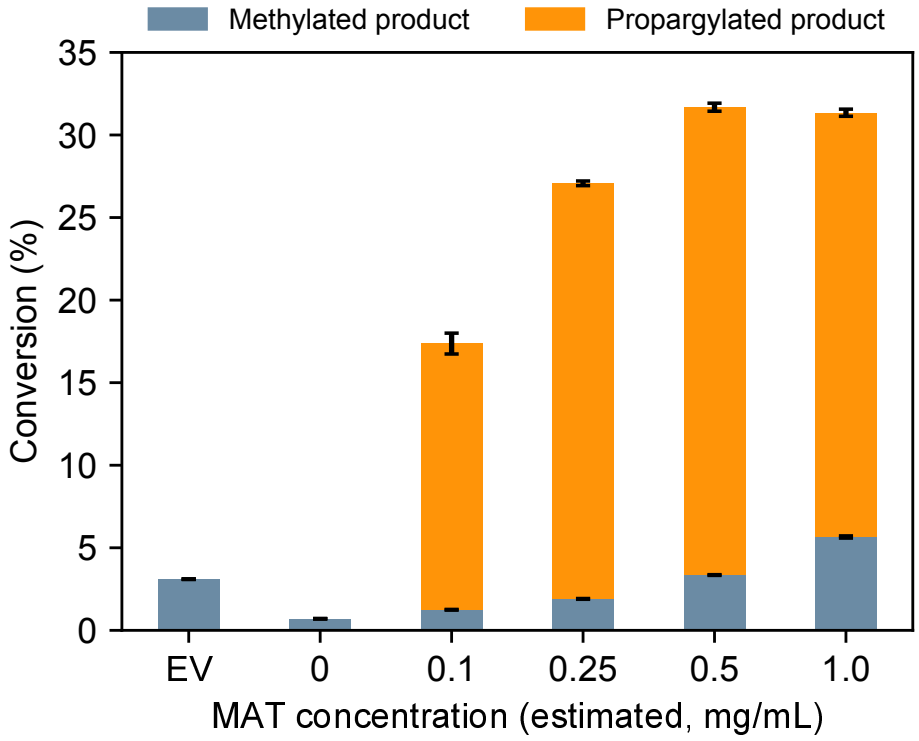
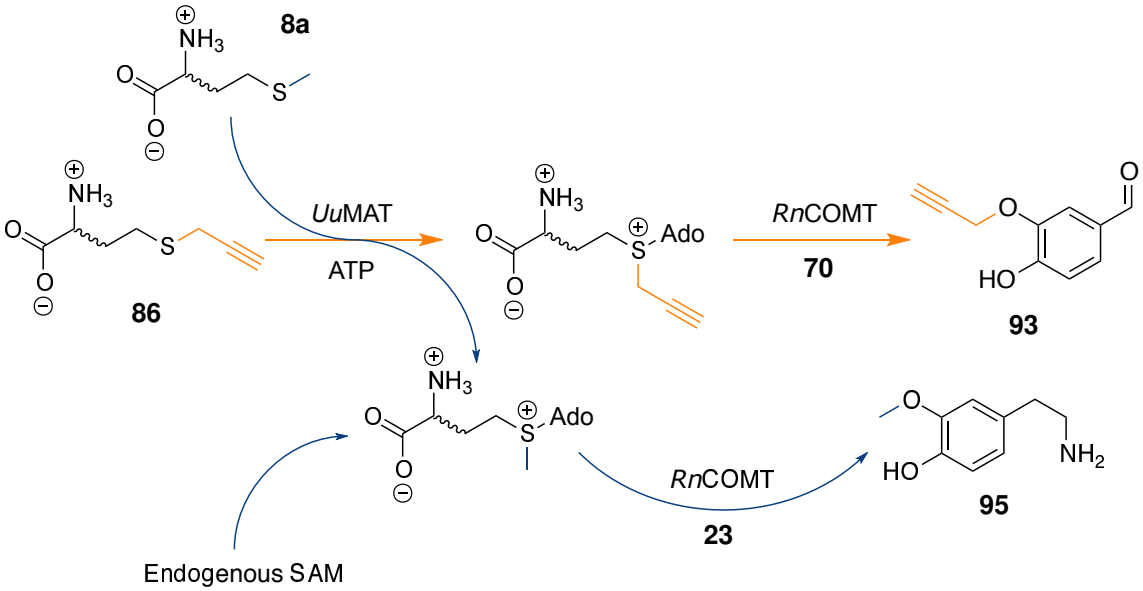


Figure 5.9: HPLC analysis of enzymatic propargylation reactions using clarified lysates, with *Uu*MAT at varying concentrations. Conversions to propargylated and methylated products calculated from HPLC peak areas of products and remaining starting material. Error bars represent one standard deviation above and below the mean of three replicates. Bars are stacked. EV = empty vector. Assay XII conditions, with *Rn*COMT at 0.1 mg/mL (estimated) and varying *Uu*MAT concentration. HPLC method E.

5.8 Investigating and addressing methyl poisoning

Cell lysates remained the ideal preparation of enzymes, especially for larger-scale reactions. Therefore the issue of side-reactions with SAM and L-methionine needed to be addressed.

The first idea involved using dopamine as a sacrificial substrate to consume the L-methionine and SAM in the reaction mixture. Substrate screens with propargylation, and those before with ethylation, had revealed that **23** is not accepted for either reaction. The docking studies in Section 4.5 suggested that this was due to the compound's ethylamine substituent binding in the active site. It was therefore hypothesised that if both substrates were in the reaction, **23** would readily be transformed into **95**, reducing the methylation of **70** (Scheme 5.8).



Scheme 5.8: Illustration of the planned role for dopamine **23** in reducing the amounts of methylated side-product in propargylation reactions containing cell lysates. Orange represents propargylation pathway, navy blue represents methylation pathway.

To test this theory, a modified assay was conducted. Conditions were kept largely the same as prior experiments, with **70** used as the main substrate (Assay XII). However, *Rn*COMT was used at 1 mg/mL in order to deliberately create pronounced methyl poisoning. Dopamine **23** hydrochloride was added to reactions from the beginning, in a small range of concentrations from 0 to 5 mM. If the sacrificial substrate method was successful, it was expected that the conversion to propargylated product would increase, and methylated product would decrease, as more **23** was added.

The results of these experiments were, however, mixed. Conversions in general were lower than had been observed previously, even with exactly the same conditions (compare Figure 5.10: '0 mM' to Figure 5.8A: '1.0 mg/mL'). Addition of **23** did appear to reduce the formation of methylated product, but higher concentrations did not lead to greater reduction. Furthermore, despite lower levels of methylated product, increasing concentrations of **23** also caused a steady decline in the formation of propargylated product. This indicated that the sacrificial substrate was directly or indirectly inhibiting the propargylation instead of helping it.

To avoid the sacrificial substrate interfering with the genuine substrate, it might therefore be necessary to have two orthogonal MT reactions: *Rn*COMT alkylating its cate-

chol substrate, and then a different MT methylating, but not otherwise alkylating, a non-catechol substrate. This would be a more complex system, and would require screening non-catechol MTs and substrates for a lack of tolerance towards alternative alkylations. Another solution might be to engineer a variant of MTAN which is able to degrade SAM (instead of SAH), but not the cofactor analogues. Neither of these ideas were explored further in this project, but could be considered in future work if this poisoning effect remains an obstacle.

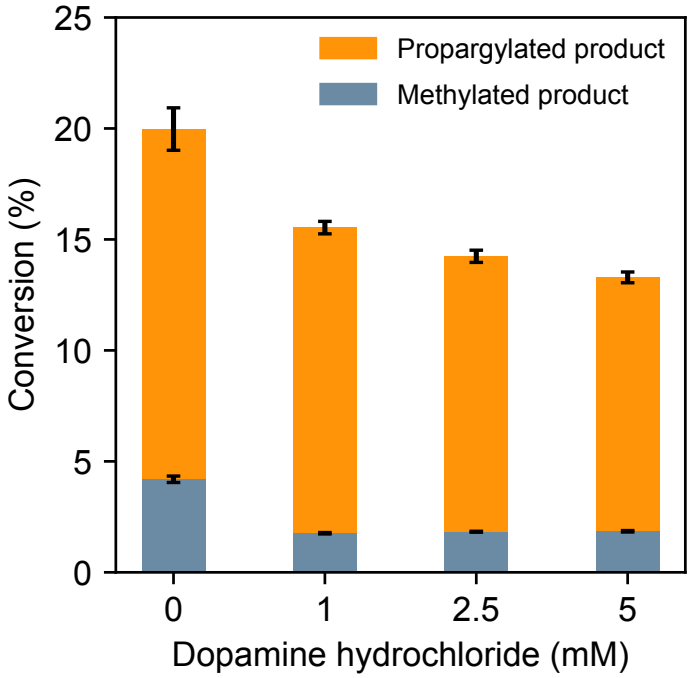


Figure 5.10: HPLC analysis of the impact of sacrificial dopamine hydrochloride, at varying concentrations, on the conversion of propargylation reactions with clarified cell lysates. Conversions calculated from HPLC peak areas of products and remaining starting material. Error bars represent one standard deviation above and below the mean of three replicates. Assay XII conditions, with *Rn*COMT at ~1 mg/mL. HPLC method E.

The next strategy attempted was to remove the SAM and L-methionine from the cell lysates directly via a desalting resin. This was expected to reduce conversion to the methylated product and, if the methionine poisoning hypothesis was correct, to increase conversion to the propargylated product. Desalting would not be selective for these methylating contaminants, though, and would remove all molecules below ~1000 Da in mass. This would introduce the question of whether some other component of the cell lysate was having the observed effect. However, the dual effect of raising one conver-

sion and lowering the other would support the idea that the poisoning effect is related to specific interactions of the cofactors with *Rn*COMT, and not a general effect on the enzyme.

Lysates of *Rn*COMT, *Uu*MAT and *Ec*MTAN were generated as normal and the total protein concentration determined via a Bradford assay. Portions of the lysates were then passed through a PD-10 desalting column, and the concentration tested again. The desalting process was found to have diluted the lysates by approximately 2x. To maintain the enzyme concentration between the reactions, the volume of desalted lysate therefore had to be increased, but there was a limit on how much more lysate could be added without exceeding the volume of the assay, and thus diluting the other components (substrate, **86**, ATP). The estimated concentration of *Rn*COMT in both reactions therefore had to be limited to 0.4 mg/mL. Otherwise, the conditions previously developed were used (Assay XII).

Analysis showed that in the reactions with desalted lysates, propargylated product was increased (+52%) and methylated product was decreased (-62%) relative to reactions with only clarified lysates. This was in agreement with the predictions made by the methyl poisoning hypothesis. It also showed indicated that desalting could provide a way to quickly improve the conversions and product purity of any alkylations affected by the poisoning. However, dilution of the lysates would also limit the concentration of enzymes that could be used, which may itself hamper conversion.

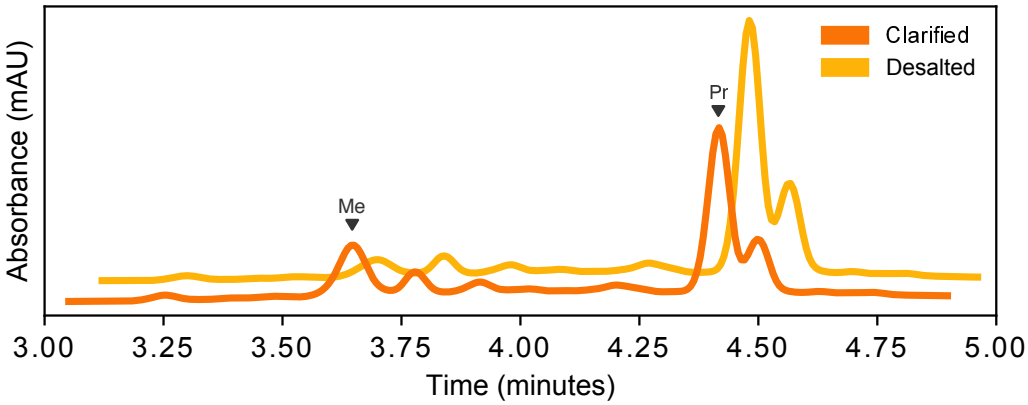
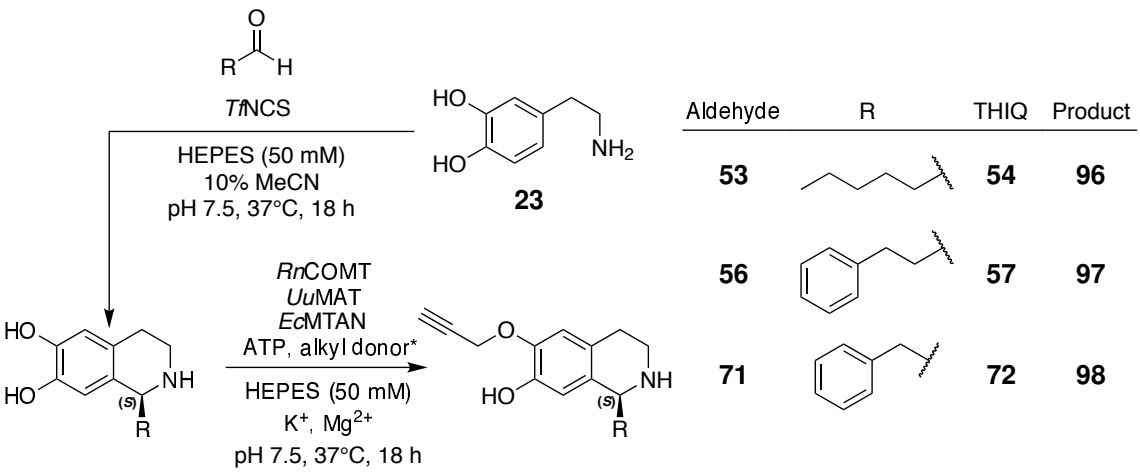


Figure 5.11: HPLC analysis of propargylation assays using clarified or clarified and desalted cell lysates. Traces offset diagonally for clarity. 'Me' and 'Pr' indicate methylated and propargylated products, respectively. Assay XII conditions, with *Rn*COMT at ~0.4 mg/mL. HPLC method E.

5.9 NCS-propargylation cascades

The ultimate aim of this work was to show the transfer of alkyne functionality to a THIQ substrate. Work with the model catechol **70** had so far revealed that, despite some useful optimisations, the MT enzyme cascade was still not demonstrating high levels of conversion. Nonetheless, it was decided to at least establish the acceptance of THIQ substrates for propargylation as a baseline that could be improved later.

The assays conducted for this purpose used the reaction conditions and substrates described in Section 4.4. Three THIQ substrates were produced from **23** and three aldehydes, as shown in Scheme 5.9, using desalted lysates of *Tf*NCS (Assay VII). These reactions were centrifuged to remove precipitate, and the supernatants aliquoted into the propargylation reactions containing ATP, **86** and lysates of *Rn*COMT, *Uu*MAT and *Ec*MTAN (Assay XII). Methylation control reactions were conducted in parallel, substituting **86** for **8a**. The reactions were analysed by HPLC after both the NCS step (step 1) and the alkylation step (step 2). To aid identification of peaks, Method A was used for analysis, to match the earlier work with ethylation of the same compounds



Scheme 5.9: Two-step NCS-propargylation cascades, for generating and propargylating three THIQ substrates. *Alkyl donor was either **8a** (controls) or **86**.

As before, THIQ compounds **54**, **57** and **72** were all successfully generated in the first step (Figure 5.12). The methylation control reactions all showed complete conversion to the methylated product that had been observed previously.

For **(*S*)-54**, analysis of the propargylation reaction showed both a known peak cor-

responding to background methylation (RT 8.9 min), and a pair of new peaks (RT 9.2, 9.4 min). As expected, most of the starting material remained. For **(*S*)-57**, methylated product was also present in the propargylation trace (RT 9.2 min), but that peak seemed to overlap with a slightly later peak of similar height (RT 9.3 min) that was not present in the methylation control. Furthermore, there was another, even later new peak (9.6 min). Lastly, in the propargylation reaction with **(*S*)-72**, there was again evidence of background methylation (RT 8.5 min) and two small new peaks (RT 9.0, 9.3 min).

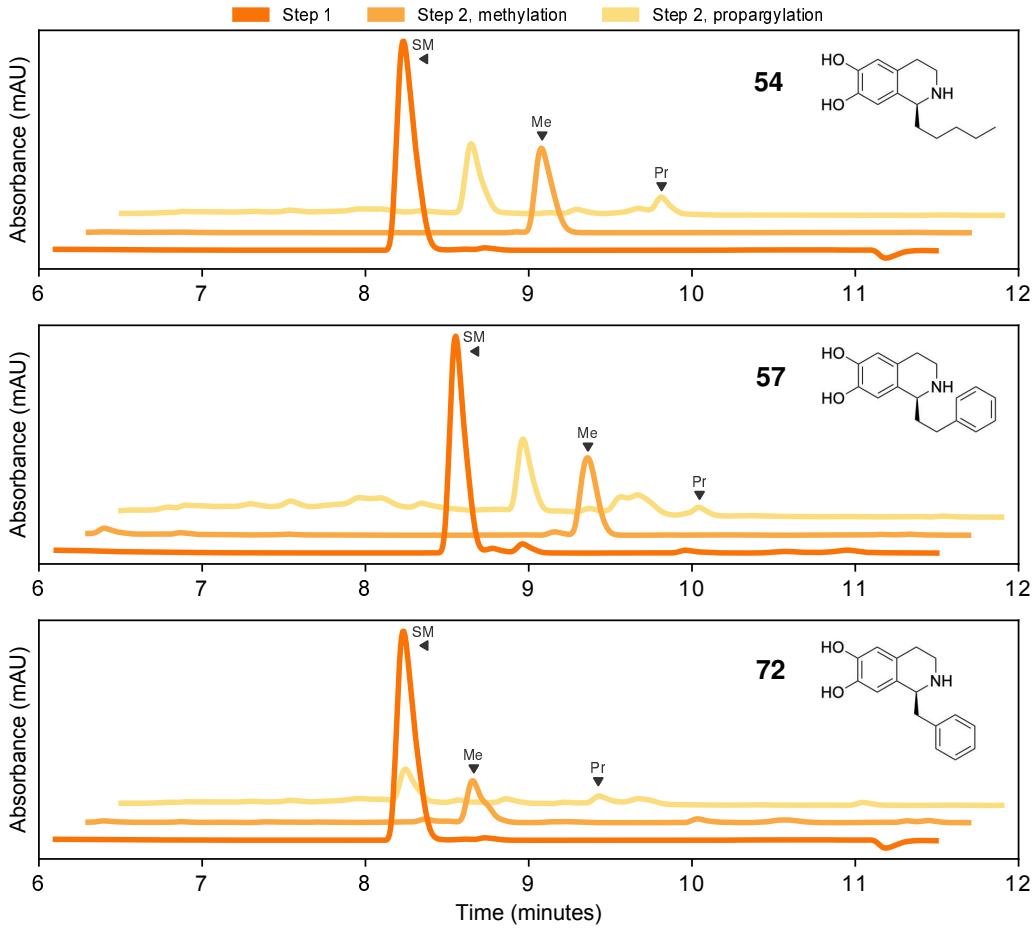
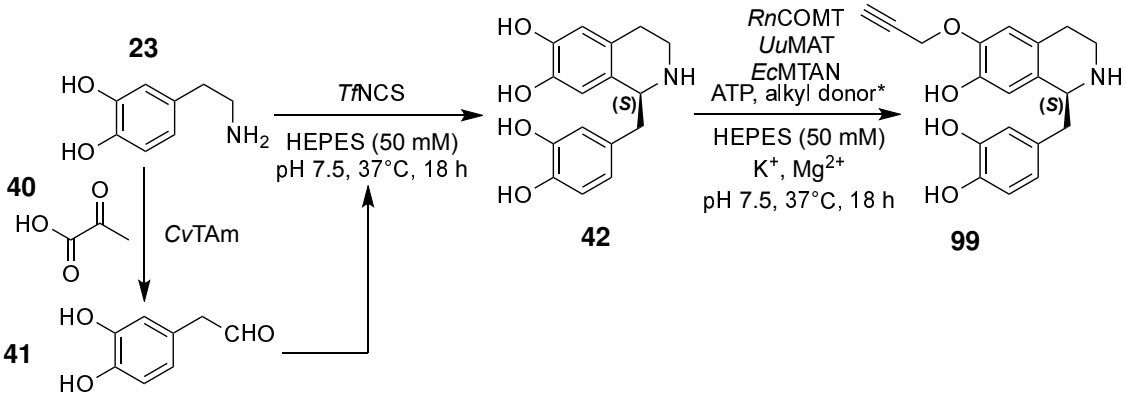


Figure 5.12: Representative traces from HPLC analysis of NCS-propargylation cascades. Step 1 is product from NCS reaction, Step 2 is product from alkylation reaction. Step 2 traces have half the signal of Step 1 traces due to dilution of the THIQ when the alkylation reagents are added. Peaks corresponding to THIQ, methylated THIQ and (suspected) propargylated THIQ are indicated by 'SM', 'Me' and 'Pr', respectively. Assay VII conditions for NCS reaction, XII for propargylation reaction (with *Rn*COMT at ~0.1 mg/mL). HPLC method A.

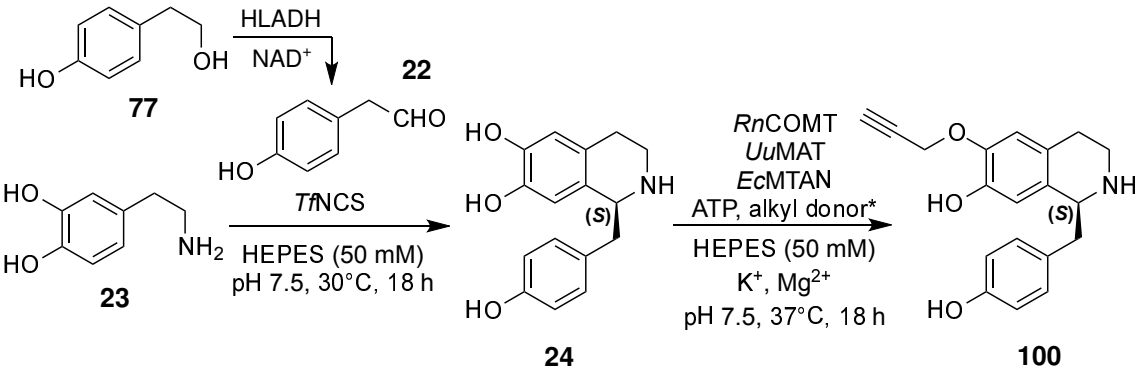
To confirm the suspicion that the new peaks for **(*S*)-54**, **(*S*)-57**, **(*S*)-72** represented

propargylated products, LC-MS analysis was performed on the reactions. Peaks corresponding to M+1 for products **(*S*)-96** (*m/z* = 274), **(*S*)-97** (*m/z* = 308) and **(*S*)-98** (*m/z* = 294) were detected in their respective samples (Appendix F). As had been observed for ethylation, the conversions were low, but established that a level of THIQ propargylation was possible.

The study of THIQ propargylation was extended to **(*S*)-42** and **(*S*)-24**, which had been generated and ethylated in Section 4.6. The same strategies for the NCS step were employed as before, using *Cv*TAm and pruvate **40** to generate aldehyde **41** for **(*S*)-42** (Scheme 5.10, Assay VIII), and HLADH and NAD^+ to generate **22** for **(*S*)-24** (Scheme 5.11, Assay IX).



Scheme 5.10: Two-step cascade to generate and propargylate (*S*)-norlaudanosoline **(*S*)-42**. Alkyl donor was **86** for the reaction shown, and **8a** for the control reactions.



Scheme 5.11: Two-step cascade to generate and propargylate (*S*)-norcoclaurine **(*S*)-24**. Alkyl donor was **86** for the reaction shown, and **8a** for the control reactions.

HPLC analysis of the alkylation reactions following generation of **(*S*)-42** is shown in Figure 5.13. As before, with the knowledge that this THIQ has two target sites for

alkylation, assays were conducted varying amounts of alkyl donor and ATP. However, based on previous results, it was not expected that the di-propargylated product would be observed. In the ATP-only conditions, the peak corresponding to **(*S*)-42** remained (RT 6.7 min), with small later peaks potentially corresponding to background Pictet-Spengler reaction in the NCS step and background methylation in the alkylation step. The assay containing 1.5 eqv. of L-methionine **8a** showed disappearance of the starting material, appearance of a dominant peak likely corresponding to the mono-methylated product (RT 7.3 min), and a second, later peak (RT 7.8). Meanwhile, the reaction containing 3 eqv. of **8a** showed almost complete conversion to that later peak. This peak was not seen in the ethylation reactions, but based on known patterns of methylation with this substrate, it was believed to be the di-methylated product. The fact that this had not been observed during the ethylation assays may be down to higher production of SAM by *Uu*MAT, given that for these assays the volume of MAT lysate was raised to give 0.5 mg/mL of enzyme, whereas in those previous reactions a fixed percentage of 10% (*v/v*) was used regardless of concentration. In the propargylation reactions, a single new peak was observed which was larger when 3 eqv. of **86** (DL-SPHC in Figure 5.13) were added (RT 7.9 min). To confirm that this peak represented the desired product, a sample of the 3 eqv. reaction was submitted for LC-MS, which detected an *m/z* peak at 326 corresponding to M+1 for the mono-propargylated product **(*S*)-99** (Appendix F).

Analysis of the cascade towards **(*S*)-100**, via **(*S*)-24**, showed a similar outcome (Figure 5.14). Only a small amount of the THIQ (RT 6.8 min) was generated in the first step, as had been observed before with ethylation. This was ascribed to a low concentration of HLADH in the cell lysate. However, efforts to improve expression, including supplementation with sorbitol, $ZnCl_2$, NAD^+ and increasing the induction period to up to 72 h were not meaningfully beneficial. If the conversion of the NCS step were to become critical, further optimisation or purification of the enzyme would be the next steps. However, for this exploratory assay, it was sufficient to have a detectable amount of **(*S*)-24**.

The methylation control reactions showed complete conversion to the respective product (RT 7.7 min). The peaks at 5.8 and 6.2 min were likely due to methylation of dopamine, which remained in the reaction mixture due to the low conversion of the NCS step. The propargylation reactions, meanwhile, showed a level of background methylation and

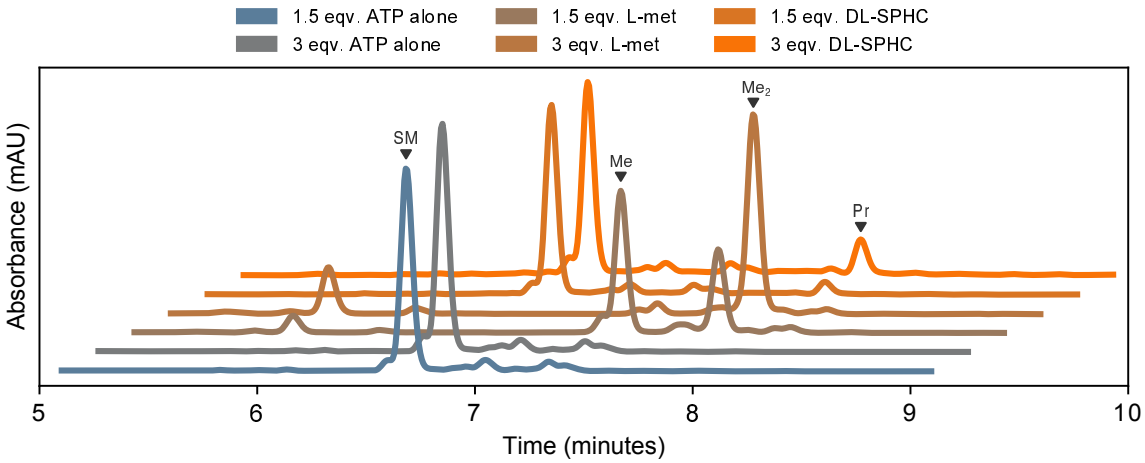


Figure 5.13: HPLC analysis of NCS-propargylation cascade involving **(*S*)-42**. Key shows alkyl donor added. DL-SPHC = *S*-propargyl-DL-homocysteine, **86**. Peaks corresponding to starting material **(*S*)-42**), mono-methylated product, di-methylated product and suspected propargylated product marked as SM, Me, Me_2 and Pr, repsectively. Assay VIII conditions for NCS step, XII for alkylation step (with varying eqv. of ATP and alkyl donor and ~0.1 mg/mL *Rn*COMT). HPLC method A.

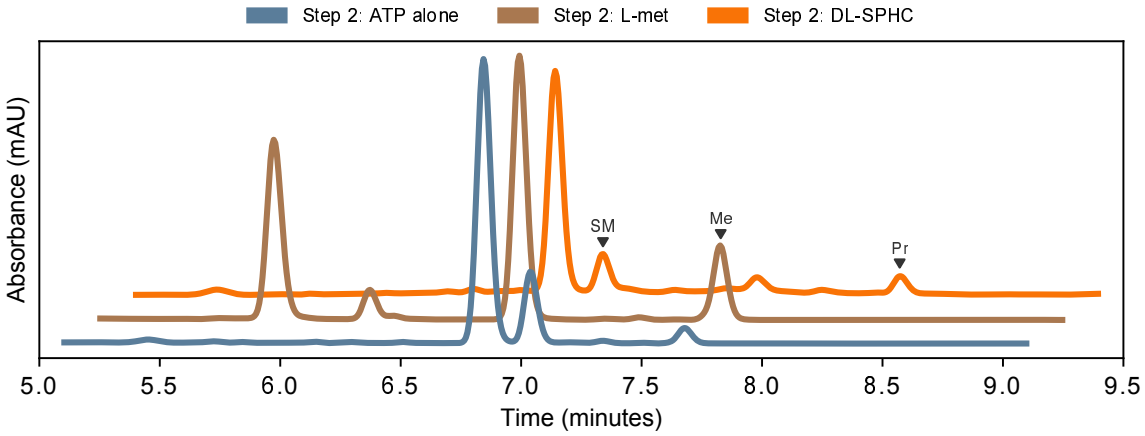


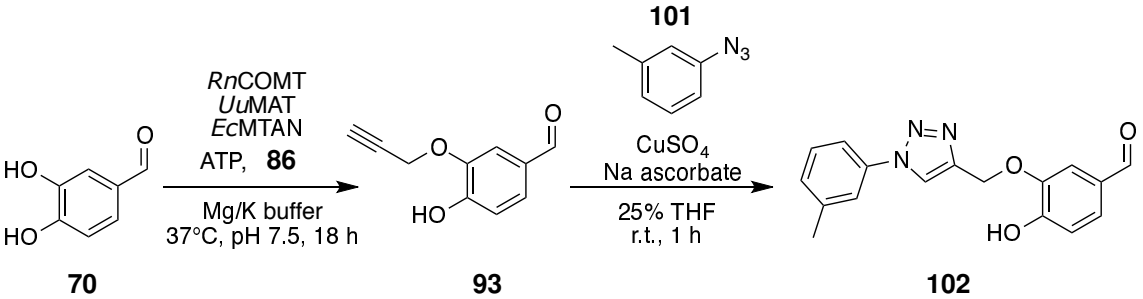
Figure 5.14: HPLC analysis of NCS-propargylation cascade involving **(*S*)-24**. Key shows alkyl donor added. DL-SPHC = *S*-propargyl-DL-homocysteine, **86**. Peaks corresponding to starting material **(*S*)-24**), methylated product, and suspected propargylated product marked as SM, Me and Pr, repsectively. Assay IX for NCS step, XII for alkylation step (with *Rn*COMT at ~0.1 mg/mL). HPLC method A.

the appearance of a small new peak (RT 8.3 min). Samples of this reaction were analysed by LC-MS, which found an *m/z* peak at 310 corresponding to M+1 for the propargylated product **(*S*)-100** (Appendix F).

There was evidence, then that all five NCS-propargylation cascades had produced modest but detectable amounts of propargylated THIQs. With this, the second objective set out at the beginning of this section had been tentatively completed: an MT cascade had been developed which was capable of transferring an alkyne group onto both simple and more complex substrates. Conversions remained low with clarified lysates, and recovering some of the lost regioselectivity with the existing E199D mutant meant reducing productivity even further. Efforts to improve the enzymes themselves had already begun, and will be described in the next chapter. However, it was decided to use the information gathered so far to attempt linking the enzymatic propargylation to a CuAAC click reaction.

5.10 Attempting a chemoenzymatic alkylation-click reaction

A simple two-step reaction was designed, in order to show that the alkyne handle installed by the enzymatic reaction can be used in a subsequent click reaction. Compound **70** would be propargylated, giving **93**: a mixture of *meta* and *para* regioisomers (Scheme 5.12). This would then be coupled to a commercially-available azide 1-azido-4-methylbenzene **101** via CuAAC.



Scheme 5.12: Attempted two-step chemoenzymatic synthesis of **102**. Only major *meta*-regioisomers are shown here for simplicity, but minor *para*-regioisomers also present.

The propargylation step was conducted with established conditions using cell lysates (Assay XII). Upon HPLC analysis peaks corresponding to the expected product **93** were observed (RT 4.4 min, Figure 5.15). If the azide and copper were added to the crude

reaction, it was likely that the alkynes on both **93** and the methionine analogue **86** would react. This could theoretically be overcome with an excess of the azide (i.e. at least 4 eqv.). However, in this case, it was decided to isolate the product by extraction into ethyl acetate. The solvent of the organic fraction was evaporated and the residue re-dissolved in 50% $MeCN_{(aq)}$, then analysed again by HPLC. The trace indicated that the most polar components of the reaction were gone, leaving **70** and **93** (Figure 5.15). The peak corresponding to **70** appeared different in this trace, but not the traces before or after, for unknown reasons. The solution containing the organic products was then mixed with azide **101** (2 eqv.) in THF and a solution of $CuSO_4 \bullet 5H_2O$ (0.15 eqv.) and sodium ascorbate (0.45 eqv.), then shaken at room temperature for 1 h (Scheme 5.12).

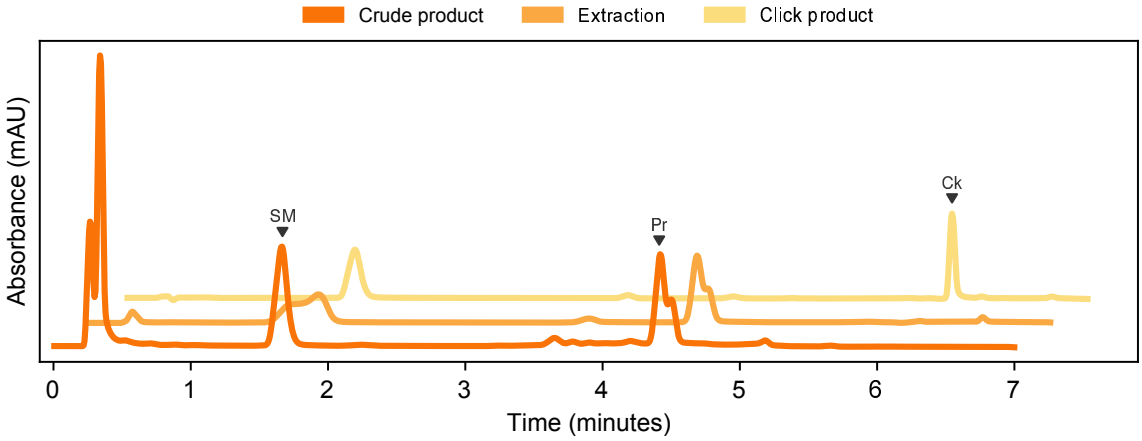
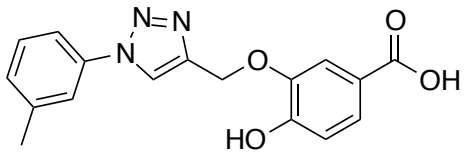


Figure 5.15: HPLC analysis of each stage of chemoenzymatic click cascade. SM, Pr and Ck indicated starting material, propargylated product and suspected click reaction product, respectively. Assay XII conditions for propargylation, with *Rn*COMT at ~0.1 mg/mL. Conditions for click given in text. HPLC method E.

The resulting solution was centrifuged to remove precipitate, and a sample of the supernatant again analysed by HPLC. In this trace, the original starting material **70** was still evident, but the peak corresponding to **93** had dissappeared. Instead, a new peak had appeared at 6.0 min. A standard of **101** was also injected, and was found to have negligible absorbance at 283 nm. This new peak was therefore suspected to be the click product **102**, to which there had apparently been complete conversion. To confirm this, a sample of the reaction was analysed by mass spectrometry.

However, no *m/z* peak corresponding to **102** was identified. A dominant peak at 324 was initially hypothesised to represent an oxidised carboxylic acid, i.e. **103** (Scheme 5.13),

but closer examination showed the to be 1 *m/z* unit less than the hypothetical mass of this product. There was no clear explanation for this peak. The reaction was subsequently repeated with the same conditions for the enzymatic step, but several extraction procedures were tried before and after the click step. These all gave the same result, including the *m/z* peak at 324. It could therefore not be concluded that the observed HPLC peak corresponded to formation of the click product.



103

Scheme 5.13: Click reaction product species initially though to be detected by mass spectrometry.

5.11 Conclusions

The work in this chapter thus achieved the first two aims: finding an economical synthetic route to an alkyne-bearing methionine analogue, and the integration of that analogue into enzyme cascades involving THIQs. The click reaction procedure appeared to need further refinement, which there was unfortunately not enough time for in this project.

There were also still considerable limitations affecting the enzyme-catalysed reactions, which no single solution or optimisation seemed able to overcome. Recovering regioselectivity involved sacrificing conversion, remedying methyl poisoning involved reducing enzyme concentrations, and the conversions of THIQ substrates were still low. These represented fundamental problems with acceptance of the methionine and cofactor analogues by the enzymes of the system. An attempt had been made to address this in Chapter 4 with the development of the E199D mutant. However, with the limits and potential of propargylation now established, the focus once again turned to a concerted engineering effort with the aim of improving *Rn*COMT and *Uu*MAT.

Chapter 6

Mutagenesis

6.1 Introduction

The work described in previous chapters had established that alternative alkylation of THIQ substrates was possible. Furthermore, Chapter 5 had shown that methionine analogues can be readily synthesised. However, there remained a need to improve the acceptance of methionine and cofactor analogues by the enzymes of the cascade.

Of the two archetypal methods of protein engineering, directed and random mutagenesis, only directed mutagenesis had been attempted, resulting in *Rn*COMT (E199D). However, no further mutations had been tested. This was partly due to the limited number of active site residues that are not conserved for catalysis, which gave few options for structural changes that would not be detrimental to the reaction. The MAT from *Ureaplasma urealyticum*, whose acceptance of *S*-propargyl-DL-homocysteine **86** enabled the work in Chapter 5, also had a reported crystal structure.[200] However, to the best of our knowledge, it had not been explored for beneficial mutations. Directed mutagenesis was therefore revisited for both of these proteins in an attempt to make the most of the structural information and published work available.

The second method, random mutagenesis, had not been employed to date. Unlike rationalised substitutions, a random technique would require construction of a mutated DNA library, and development of a high-throughput assay with enough sensitivity to detect beneficial mutations. It would also be dictated by chance - there was no guarantee that a beneficial mutation would occur within the population of variants screened. However,

random mutagenesis could discover advantageous changes that would not be predictable from the X-ray crystal structures. These could include mutations that alter the dynamics of the active site or improve the overall stability of the enzyme. Either could be beneficial to the acceptance of methionine and cofactor analogues by *Uu*MAT and *Rn*COMT. A second branch of work was thus undertaken to build and execute a random mutagenesis workflow.

6.2 Revisiting rational design

In the docking study described in Chapter 4, an M40S mutation was considered as a way to relieve steric clashes in the active site of *Rn*COMT. This residue lies between the substrate and cofactor binding sites and forms a 'lid' over the cofactor sulfonium ion. It was therefore considered whether resecting this residue would allow the terminal methyl group of SAE to rotate out of the active site. However, when this mutation was made *in silico*, neither SAM nor SAE were found to dock in the cofactor binding site of the variant. This suggested that making changes to M40 might harm cofactor binding, so the idea was not pursued.

However, a publication by Herbert *et al.* later showed that an M40A substitution was beneficial for the acceptance of the cofactor analogue carboxy-SAM (Scheme 1.5).[132] The alkyl group of that analogue was bulkier than either the ethyl or the propargyl groups, and had a substantially different physiochemical nature. Nonetheless, it was decided to investigate if the benefits of this mutation could translate to propargylation.

Synthetic genes encoding *Rn*COMT (M40A), as well as (M40A, E199D) to see how the two mutations might interact, were designed and codon-optimised for *E. coli*. The genes were synthesised and cloned into the pET-29a(+) vector by the supplier before delivery. The wild-type and E199D forms of *Rn*COMT used previously were both in pET-28a(+), however, so care was taken during design to ensure that the expression products of all variants would be exactly the same apart from the desired mutations.

By contrast, there were no prior indications which residues in *Uu*MAT might be amenable to engineering. However, MAT enzymes across all species tend to share similar structural characteristics, and mutations had been reported for other homologues which

increased tolerance of methionine analogues. In particular, three publications had indicated that the I117A variant of *Hs*MAT2A increased substrate scope.[127,128,198] The structure of *Uu*MAT was examined, and it was found that that residue I97 is analogous to I117 in *Hs*MAT2A. It was hypothesised, then, that mutation of I97 to alanine would have a likewise beneficial effect on the acceptance of **86**.

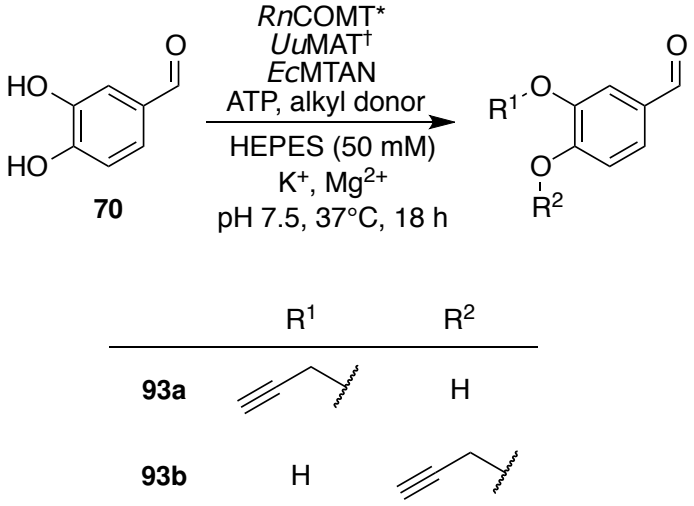
A synthetic gene encoding this mutant enzyme was therefore designed and codon-optimised for *E. coli*. The wild-type gene was held in pET-24a(+), which has a similar sequence to pET-29(+), so no further considerations were needed to keep the new variant otherwise equivalent to the original.

All synthetic genes were transformed into *E. coli* BL21(DE3), and the enzymes produced using standard expression conditions. No significant variations in the levels of expression between the wild-type and mutant enzymes were observed.

To compare the variants, an assay was conducted with **70** as a model substrate. The three variants of *Rn*COMT and two variants of *Uu*MAT were tested in every combination, with both **8a** and **86** as alkyl donors. From prior experiments, propargylation was expected to generate two regioisomers, **93a** and **93b** (Scheme 6.1). Otherwise, the conditions developed in previous chapter for clarified lysate alkylations were used (Assay XII). Use of purified enzymes would have given a more direct comparison due to better control over concentration. However, because of the number of enzymes being investigated, this was intended as a first-pass test. Any variants which showed improvement over the wild type could then be examined more accurately.

HPLC analysis of the reactions showed that instead of improving conversions, every variant lowered them to some extent (Figure 6.1). Conversions for propargylation were lower than for methylation in every case. However, while wild-type and I97A *Uu*MAT gave comparable conversions for methylation, the mutant had a dramatic negative impact on propargylation. This was surprising, as the mutation was specifically designed to create more space for longer *S*-alkyl chains. However, the change may in fact have simply allowed another residue to shift into the space, or compromised the hydrophobic environment. In hindsight, it may have also been wise to examine other, less dramatic substitutions, such as I97V.

An alternate pattern was observed with the *Rn*COMT mutants. M40A severely re-



Scheme 6.1: Assay of *Rn*COMT and *Uu*MAT variants with a model catechol substrate. *Wild-type, M40A or M40A, E199D *Rn*COMT. †Wild-type or I97A *Uu*MAT. Alkyl donor was either **8a** or **86**.

duced methylation, but only slightly affected propargylation. This may be due to the mutation affecting the binding of SAM, as hypothesised from the docking investigation, but not the propargyl-cofactor. The better fit with this cofactor analogue did not translate into any improvement in conversions, however. Furthermore, in every case, adding the E199D mutation lowered conversions further.

This experiment therefore indicated that none of the new mutants were beneficial to propargylation. However, there was a rationale for making changes to each of these positions, so it could be that the specific substitutions made in this study were not optimal. Future work might therefore focus on site-saturation mutagenesis to probe more changes at those locations.

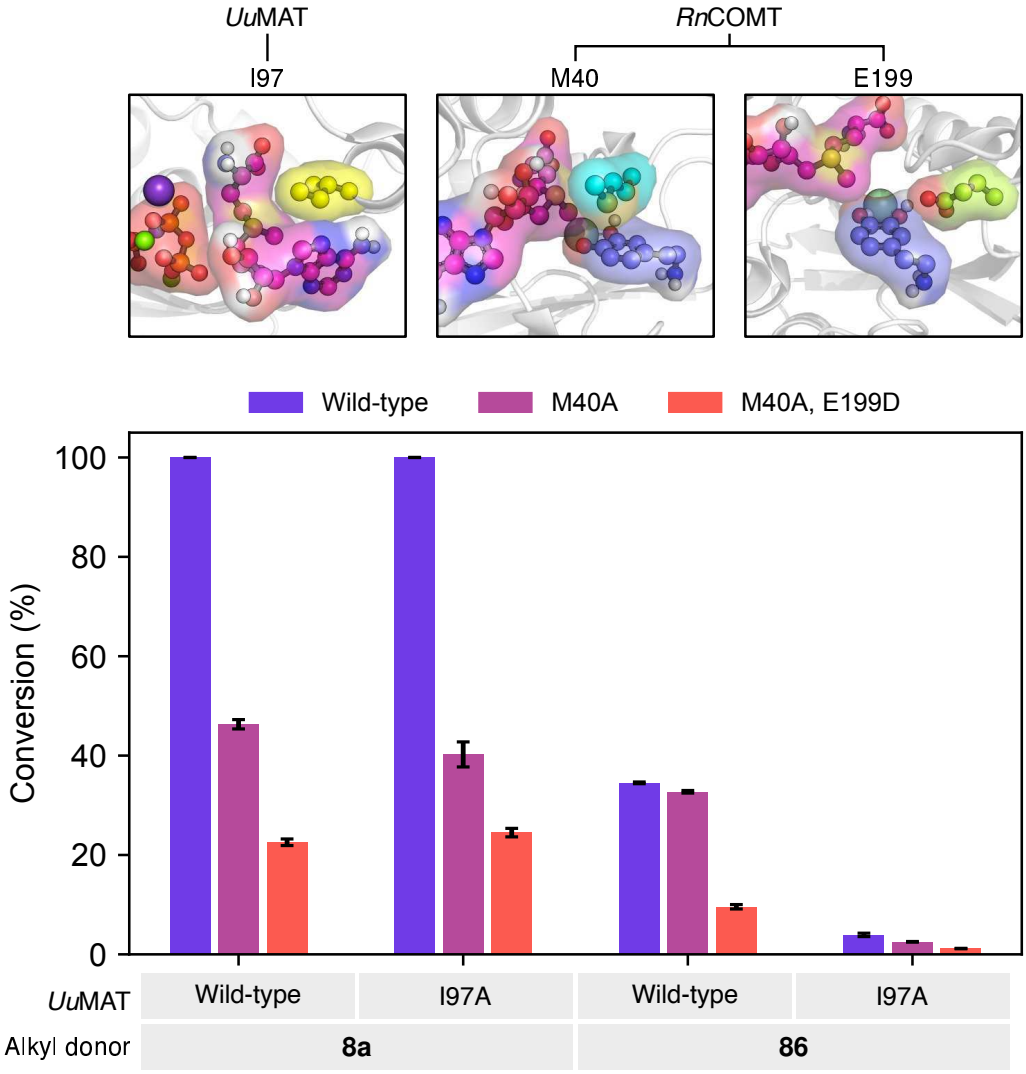


Figure 6.1: Analysis of assays comparing methylation and propargylation of **70** by wild-type and I97A variants of *Uu*MAT, and wild-type, M40A and M40A, E199D variants of *Rn*COMT. Conversions were calculated using HPLC peak areas of product and remaining starting material. Error bars represent one standard deviation above and below the mean of three replicates. Top panels show the location of wild-type residue for each position. Assay XII conditions. HPLC method E.

6.3 Synthesis and testing of *S*-propargyloxybut-2-enyl-DL-homocysteine as an alkyl donor

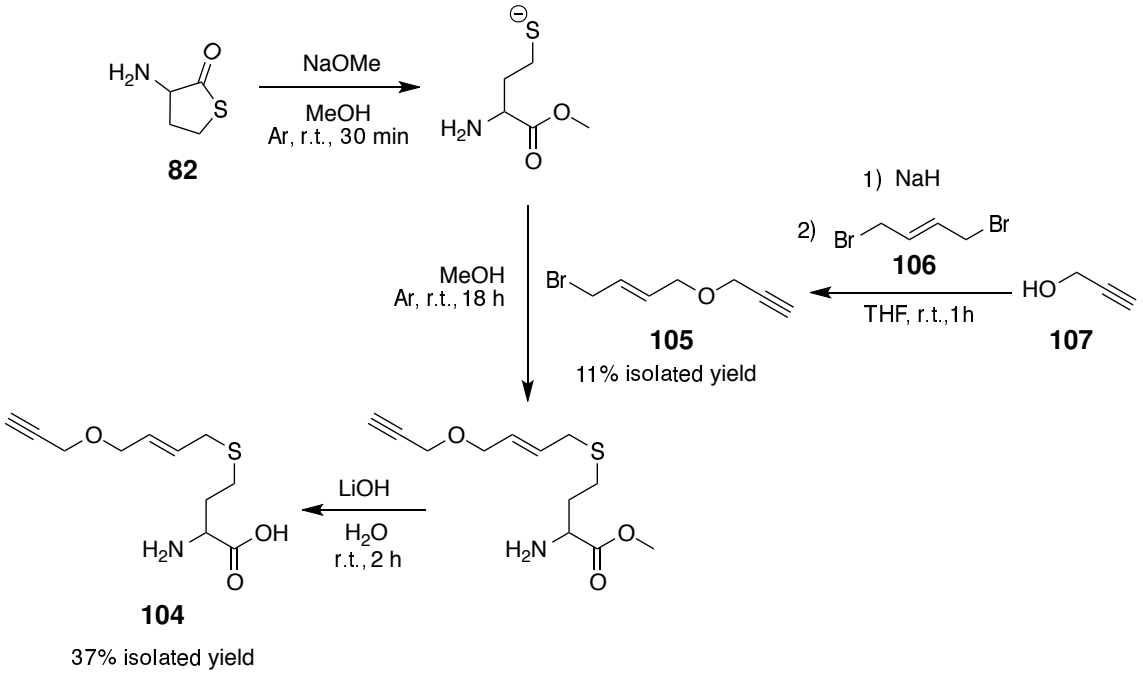
The mutations described above all aimed to relieve steric clashes in the active sites of *Uu*MAT and *Rn*COMT. However, sterics is only one of the likely obstacles to alternative alkylation, the other being activation of the carbon targeted by the S_N2 substitution. This atom is made electrophilic in the cofactor by the adjacent sulfonium ion. However, in-

creasing the size of the alkyl chain, especially with electropositive groups, dilutes this activation and reduces the energetic favourability of the transfer. Some alkyl groups are able to partly restore the activation, notably the allyl group. This has been evidenced by other work in the group, in which a similar MT enzyme cascade using *S*-allyl-L-homocysteine showed higher conversions even with wild-type enzymes.

A new investigation thus sought to determine if the valuable functionality of the alkyne group could be paired with the reactivity of the allyl group in one methionine analogue. In work by Wang *et al.*, engineered variants of *Hs*MAT2A were shown to accept a number of analogues, including *S*-propargyloxybut-2-enyl-L-homocysteine (i.e. the chiral form of **104**, Scheme 6.2).[127] Later work by Struck *et al.* showed how the cofactor analogue with this same *S*-alkyl group, **90**, could be accepted by wild-type *Rn*COMT (Figure 5.1). This analogue had the desired characteristics of a terminal alkyne group and an alkene group adjacent to the reacting carbon. It stood out from other such analogues, however, in that the respective alkyl halide, **105**, could be synthesised from commercially available reagents (Scheme 6.2). Furthermore, this synthesis is modular, in that the component which gives the reactivity, 1,4-dibromobut-2-ene **106**, can be combined with any nucleophile (in this case propargyl alcohol **107**). This would allow a huge range of functionalised methionine analogues to be made without concern over affecting the reactivity of the target carbon.

As acceptance of the resulting cofactor analogue by *Rn*COMT had been reported in the aforementioned study,[199] the only remaining question was whether *Uu*MAT would accept **104**. This MAT was still the principal choice because its chiral non-selectivity made the faster and higher-yeilding synthesis of DL-analogues viable. To test acceptance, racemic **104** was synthesised by combining a literature preparation of **105** with the methionine analogue synthesis used previously in the project (Scheme 6.2).

The synthesis of **105** gave a low isolated yield. However, this was attributed to the small scale of the reaction, and the difficulty in purifying this mono-substituted halide from both the starting material **106** and (what was suspected to be) the di-substituted side product. Nonetheless, once **105** had been isolated and characterised, the second synthesis from **82** was begun. As before, the intermediates were not isolated. However, at the alkylation step, **105** was added. The resulting methyl ether was hydrolysed to give

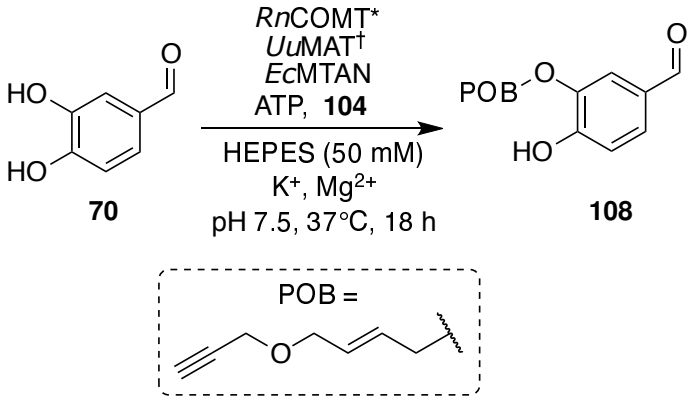


Scheme 6.2: Synthesis of methionine analogue **104** via two converging routes.

104, which was purified via an ion exchange resin. As before, the yield was lower than expected. However, enough material was generated to conduct enzyme assays.

Although they had not shown improvement over wild-type before, it was decided to test the *Rn*COMT and *Uu*MAT variants assayed in the previous section with this new analogue. The hypothesis was that the considerably larger *S*-alkyl group of **104** might benefit more from the additional space created by those mutations. An assay was thus conducted which was almost identical to the previous (Assay XII), albeit with **104** as the alkyl donor and **108** as the expected product (Scheme 6.3). Note that at this stage, it was not certain if both *meta* and *para* regioisomers would form, so the former was taken to be the dominant product.

HPLC analysis of the assays indicated that, if **108** was formed, it was only in very small amounts. In the reactions supplied with **104**, background methylation of **70** was visible at 3.6 min, and a new peak had appeared at 5.7 min (Figure 6.2). Unfortunately, the LC-MS facility that would normally be used to confirm the identity of this peak was not available at this time. However the peak had not been observed before and was not present in the methylation controls, suggesting that it represented a genuine new product. Furthermore, its elution time suggested greater hydrophobicity than the propargylated product, which



Scheme 6.3: Assay for acceptance of **104** and the resulting cofactor by variants of *Uu*MAT and *Rn*COMT.*Wild-type, M40A or M40A, E199D *Rn*COMT. †Wild-type or I97A *Uu*MAT.

was in line with expectations. The new peak was larger in all reactions which contained *Uu*MAT (I97A) compared to the wild type, potentially indicating that the mutation was giving the desired improvement. However, the conversions were nonetheless very low. There were also no noticeable differences in the size of this peak between the three variants of *Rn*COMT.

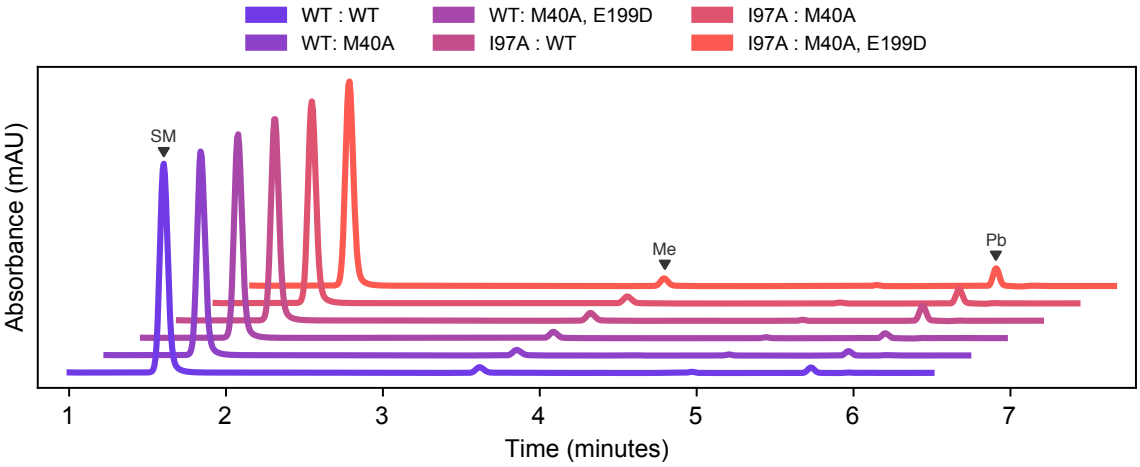


Figure 6.2: Analysis of POB-ylation assays involving wild-type and I97A variants of *Uu*MAT, and wild-type, M40A and M40A, E199D variants of *Rn*COMT. Methylation traces not shown. SM, Me and Pb indicate peaks representing starting material, methylated product and apparent POB-ylated product, respectively. Legend organised as: (*Uu*MAT variant) : (*Rn*COMT variant). Assay XII conditions. HPLC method E.

From this data, it was concluded that there may be some degree of acceptance of **104**, but the anticipated increase in conversion over *S*-propargyl-DL-homocysteine **86**

was not observed. This is likely due to poor acceptance of the much larger alkyl chain by either *Uu*MAT, *Rn*COMT or both. Although the mutant enzymes screened here were all designed to provide that space, more changes may be needed to avoid clashing of the chain with residues around the active site. However, the relative ease with which **104** was synthesised indicates that, if the enzymes of the MT cascade could be improved, this synthesis could provide a gateway to reactive methionine analogues with a variety of functional groups.

6.4 Random Mutagenesis

6.4.1 Construction of mutated DNA libraries

The first step in building the random mutagenesis pipeline was choosing an agent to introduce mutations. Many methods are available, including chemical and physical mutagens and error-prone PCR-based techniques.[202–205] Ultimately, due to expertise within the group, it was decided to use an *E. coli* mutator strain to perform *in vivo* mutagenesis. This method involves transforming the target gene into a bacterial strain which is deficient in one or more DNA repair proteins. The cells are then cultured, and as they grow develop mutations passively from chemical species from their own metabolism, exposure to radiation and errors in DNA replication. Due to the lack of a repair capacity, these mutations accumulate, with some affecting the target gene. After a suitable time, the plasmids carrying the mutated genes are extracted and transformed into a stable strain for screening.[206]

The mutations are acquired non-specifically, and as such the mutator strain itself rapidly becomes less viable. Furthermore, the only way of knowing how mutated and diverse the DNA library has become is by extracting it and screening the resulting enzymes. However, this method is fast and simple to use, and avoids carcinogenic materials. The reliance on spontaneous mutations also reduces the bias of agents which exclusively target one base or cause one type of mutation, allowing a broader spectrum of changes to be represented in the library.

The mutator strain used for this project was *E. coli* XL1-red. The bacteria are commercially sourced and have defects in the genes *mutS*, *mutD* and *mutT*, causing defects

in mismatch repair, DNA polymerase III 3'-5' exonuclease activity and repair of oxidative damage, respectively.[206]

Plasmids containing *Rn*COMT and *Uu*MAT were extracted from the BL21(DE3) expression strains using a Monarch® Plasmid Miniprep Kit (New England Biolabs) according to the manufacturer's instructions. Both DNA samples were commercially sequenced by EuroFins Scientific to confirm that the genes were wild-type. The plasmids were then transformed into 100 μL aliquots of freshly-thawed XL1-red competent cells, again according to the manufacturer's instructions. The transformation mixes were plated on LB-agar + kanamycin (50 μg/mL) and incubated for 24 h at 37 °C. The following day, scrapings from the plates were used to inoculate 10 mL LB + kanamycin (50 μg/mL) media, which were incubated overnight at 37 °C with 180 rpm shaking.

A 5 mL aliquot was taken from each overnight culture, and the plasmid DNA extracted as before. This represented generation 1. According to the manufacturer, this plasmid library should be sufficiently mutated. However, members of the group who had used this method before indicated that more generations may be needed to produce an adequate level of mutation. Therefore, 100 μL samples of XL1-red strains containing *Rn*COMT and *Uu*MAT were used to inoculate 10 mL fresh media, and those cultures were incubated overnight again. The next day, the plasmid DNA was again extracted from 5 mL aliquots of both, and the process repeated one more time with a 72 h incubation, giving three plasmid libraries total from three generations (Figure 6.3). A downside of this approach was that each generation was a bottleneck: only 1% of cells from one generation seeded the next. However, the original cultures could not be grown indefinitely, and it was believed that this would allow the screening to start with generation 2, and then move up (generation 3) or down (generation 1) the scale of mutation if the libraries were found to contain too little or too much diversity.

The first trials of the random mutagenesis workflow would involve screening the *Rn*-COMT mutants. Therefore, each of the three *Rn*COMT mutant plasmid libraries were transformed into *E. coli* BL21(DE3) and incubated overnight at 37 °C on LB agar + kanamycin (50 μg/mL). Individual colonies were then picked and used to inoculate 1 mL volumes of LB + kanamycin (50 μg/mL), which were held in 96-well, 1 mL-deep microplates. One microplate was prepared from colonies from the generation 1 and generation 3 trans-

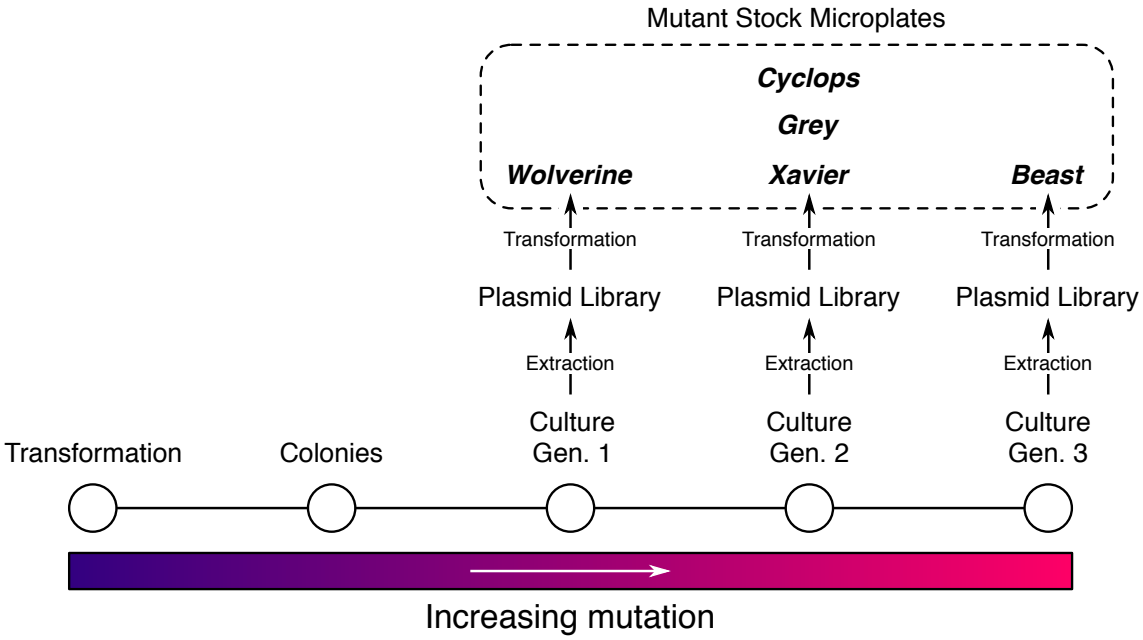


Figure 6.3: Lineage of cultures made from XL1-red strain containing *Rn*COMT gene. 'Transformation' indicates the moment of wild-type plasmid transformation into XL1-red, 'Colonies' represents the moment that the resulting colonies were used to inoculate liquid LB media.

formations, while three microplates prepared from generation 2 because it was anticipated that this library would be the most productive and thus need to be screened to a greater extent.

The microplates, each filled with 96 1 mL media volumes inoculated from 96 different colonies, were covered with a breathable film (Axygen) and incubated overnight at 37℃ with 1000 rpm orbital shaking. The following day, 100 μL aliquots were taken from each well and transferred to a 200 μL-deep microplate. These were mixed with 100 μL 50% aqueous glycerol solution, covered with an aluminum seal and stored at -80℃. These were the mutant stock microplates, with each well from each plate containing the glycerol stock of one mutant that would later be screened.

To distinguish the plates, each was given a unique name. It was decided not to use an alphanumeric naming system because A) the well locations were already alphanumeric, which could cause confusion and B) it was possible that more plates might have to be made from new or branching XL1-red lineages, which would complicate a systematic naming system. Therefore, each plate needed an identifier that was distinct, memorable and would not be reused. Thus, the mutant plate made from the generation 1 plasmid

library was named *Wolverine*, the three plates from generation 2 were named *Xavier*, *Gray*, and *Cyclops*, and the plate from generation 3 was named *Beast* (Figure 6.3).

6.4.2 Exploration of a colourimetric screen

With the libraries in hand, there was now a need to develop a high-throughput screening method. Because of the randomness of this engineering technique, a large volume of mutants would need to be screened to give the best chance of finding one with desired characteristics. At the time, alkylation reactions were analysed by a 14-minute HPLC method, which would be impractical for screening multiple 96-well microplates. To that end, a novel assay was sought which would be able to give fast, low-resolution data on the outcome of reactions, allowing the potential hits to be followed up with more accurate HPLC analysis.

It was reasoned that, if a reagent could be found which changed appearance in the presence of catechols, it would allow alkylation to be determined visually. This could form the basis of a high-throughput screening method. A search was therefore conducted for dyes which would react with that group and produce, or lose, colour as a result. No suitable organic reagents were found. However, the interaction between iron(III) ions and deprotonated catechols was noticed as a potential candidate. When in solution together, the two form reversible complexes that absorb wavelengths within the visible spectrum. The colour of light absorbed depends on the pH of the solution, and so the stoichiometry of the complexes (Figure 6.4).[207]

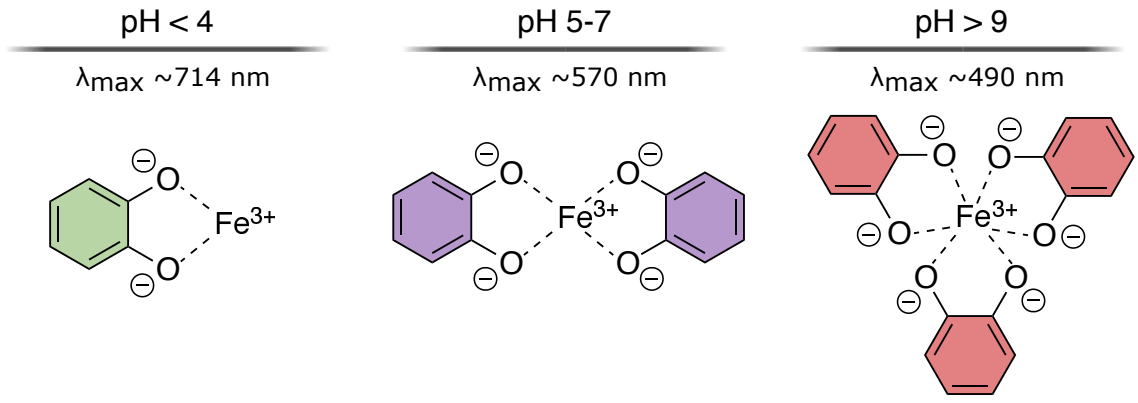
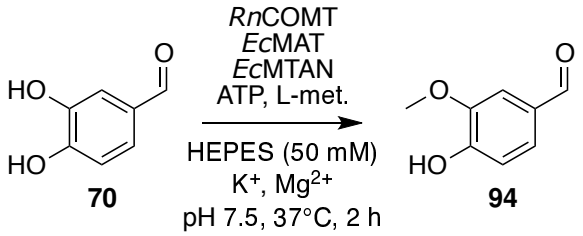


Figure 6.4: The relationship between pH and the stoichiometry and absorbance of iron(III):catechol complexes. Adapted from Bijlsma *et al.*.[207]

In theory, any alkylation of the catechol hydroxyls should interrupt this interaction. To test this, solutions of catechol **70** and an alkylated standard, (3-ethoxy-4-hydroxybenzaldehyde) were made and $FeCl_3$ added. The former turned an intense dark green, while the latter remained almost clear, albeit with a slight yellow tint. Further experimentation indicated that dilution of the catechol was necessary in order for the solution to be translucent, but colour still remained.

To test whether this could be used to track the conversion of an enzymatic assay, three parallel methylation reactions were run with **70** as the model (Scheme 6.4). Each assay in the series was designed to approach Assay XII conditions more closely: the first (A) used purified *Rn*COMT (0.5 mg/mL) and *Ec*MTAN (0.01 mg/mL) with 2 eqv. SAM. The second (B) was the same, but used cell lysates (10% (*v/v*) *Rn*COMT, 2% (*v/v*) *Ec*MTAN). The third (C) used the whole supply system, including *Ec*MAT as a lysate (10% (*v/v*)), ATP and L-methionine **8a** (both 20 mM). Methylation was chosen over the other alkylations due to the relative speed of that reaction, and the expectation that it would go to completion.



Scheme 6.4: Enzymatic methylation reaction to test iron(III)-mediated visualisation of conversion.

Small samples of the reaction were taken at eight timepoints over 2 h and quenched with an equal volume of methanol. Once all samples had been collected, they were diluted 5x with diH_2O, and a solution of $FeCl_3$ added. The final concentration of Fe^{3+} was 2.5 mM, i.e. 0.5 eqv. relative to the starting concentration of the catechol. As the assay was conducted at pH 7.5, this was done to provide the right stoichiometry for the expected Fe^{3+}:catechol complex to form (Figure 6.4).

An image of the resulting coloured samples is shown in Figure 6.5. The colour seemed to decline with time in all three reactions, although the hue also changed from more red at the beginning to more purple later on. The reactions were visually at their endpoint after approximately 45 min, although in reactions A and B, some colour persisted until the end. The amount of colour that developed was also not wholly consistent

between replicates of the same reaction, although that may have been due to errors in collecting and processing the samples. Finally, cell lysates seemed to have an impact on how the colour developed, with samples from reaction C showing reduced colour, even at t=0.

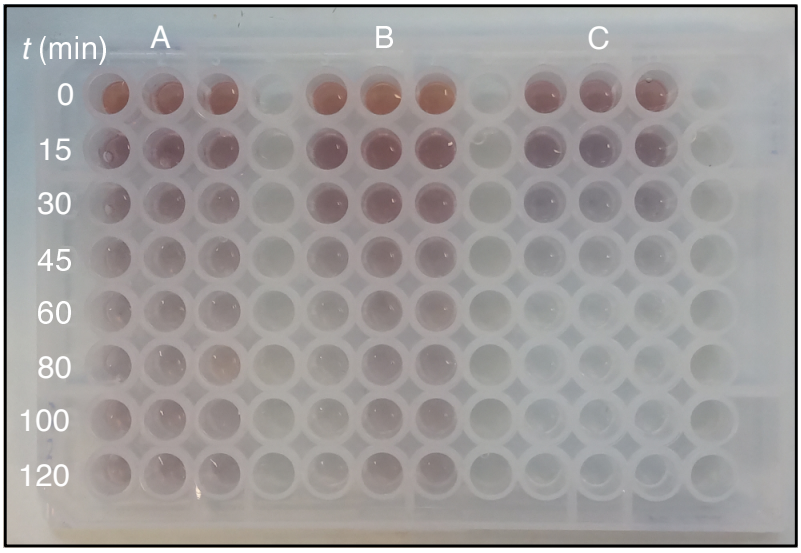


Figure 6.5: Colourimetric analysis of methylation reactions using Fe^{3+} (Scheme 6.4). Each column within the same group is a replicate reaction. A) reactions with purified MT and SAM, B) reactions with cell lysate MT and SAM, C) reactions with cell lysate MT lysate supply system. Full assay conditions described in text.

The indication at this stage was that while the principle of the assay was sound, the method needed refinement. A change in colour over the course of the whole reaction was clear, but there were questions over how great of a difference in alkylation would be needed to make two given reactions visually discernable. Furthermore, there appeared to be some inconsistency in the colour produced, suggesting that more optimisation of the protocols was needed.

To address some of these concerns, Reaction C from above was repeated with some modifications. Samples were instead taken every 10 min for 50 min to focus on the period with the greatest change. Two samples were taken per timepoint: one was analysed by HPLC to calculate conversion by the standard method, while the other was diluted 2x with buffer at pH 7.5 and mixed with 0.5 eqv. $FeCl_3$ (once all samples had been collected). This altered sample processing (dilute 2x with buffer vs. 5x with diH_2O above) was intended to make the catechol more concentrated and so the colour more intense, and to

keep the pH around 7.5 to promote the expected iron-catechol complex. Ethanol (5 μL) was added to each coloured sample to disperse any bubbles on surface. The absorbance at 490 nm was then measured for every well using a CLARIOstar plus microplate reader. The original intention had been to generate the purple (i.e. λ_{max} 570 nm) complex, but after processing, the samples were unambiguously red, so the wavelength corresponding to that complex was measured instead. All absorbance measurements were in reference to a blank sample containing the same reaction components but which had not been treated with $FeCl_3$.

The conversion measured by HPLC and the absorbance at 490 nm were then compared. For the colourimetric analysis to be a good substitute for HPLC, there would need to be a linear correlation between the two values. However, the relationship between the two appeared to become non-linear after ~20% conversion, plateauing such that absorbance did not decrease as quickly as the conversion increased. Furthermore, there was considerable variance in the absorbance of the samples which was not reflected in the conversions.

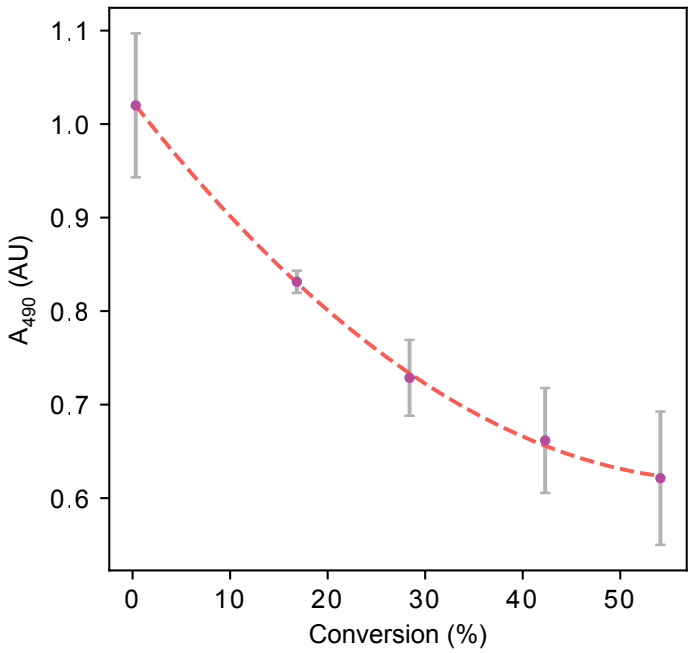


Figure 6.6: Relationship between the conversion of methylation reactions and absorbance at 490 nm following treatment with $FeCl_3$. Conversion calculated from HPLC peak areas of product and remaining starting material. Error bars represent one standard deviation above and below the mean of three replicates, and are obscured in the *x* dimension by the points. HPLC method D.

These results indicated that, unfortunately, the colourimetric analysis would not be a suitable replacement for HPLC. While the general trend of an increase in alkylation correlating with a loss of colour had been upheld, the low sensitivity of the assay, coupled with the nonlinear relationship to conversion, led to concerns that relevant differences between assays might be missed, which could in turn lead to mutant enzymes with desired characteristics being overlooked. The method might be more useful in situations where more dramatic differences in conversion are expected. For example, screening mutants of an enzyme for acceptance of a substrate that the wild-type does not accept at all.

In the time this work was underway, more HPLC equipment became available to the group, along with shorter columns. With this, a method was developed which could separate **70** from its alkylation products in half the time of the earlier method. Relying on HPLC for analysis of large numbers of reactions thus became viable. As such, the work on this colourimetric analysis was set aside, and the project progressed to conducting the first screens of the mutated DNA libraries.

6.4.3 Initial screening of *Rn*COMT mutants

The first attempts to screen mutants began with the mutant stock plate *Xavier*, which contained *Rn*COMT genes that had been mutated in XL1-red cells for two liquid culture generations. The first 16 mutant stocks from that plate (microplate columns 1 & 2) were each used to inoculate 1 mL volumes of MagicMedia™ (ThermoFisher Scientific) + kanamycin (50 μg/mL) in a 96-well, 1 mL-deep microplate. This media was chosen for its autoinduction property, which would help ensure even expression across all mutants. The microplate was covered with a breathable sealing film (Axygen) and incubated overnight at 37℃ with 1000 rpm orbital shaking. The following day, the cultures were centrifuged at 4000 rpm for 5 min to pellet the bacteria, and the supernatant was removed.

The next step, lysis of the cell pellets to release the mutant protein, was subject to extensive experimentation. Sonication by a probe, which was the standard method used to lyse pellets from the larger cultures, was not possible here. Members of the group who had expertise in their own, similar workflows recommended water bath sonication of the microplate containing the pellets. This was tried multiple times, with up to two rounds of 10 min sonications in < 15℃ water . However, this method was never able to lyse the

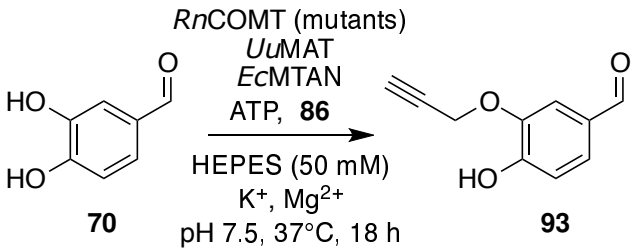
cells adequetely, as indicated by Bradford assays of the resulting clarified lysates.

Next, lysis with a detergent buffer was attempted. Non-denaturing Triton X-100 was chosen to avoid impacting the activity of the enzymes. A lysis buffer was thus composed with Tris HCl (20 mM, pH 8), NaCl (137 mM), Triton X-100 (1% (*v/v*)) and EDTA (2 mM). Tests showed that this buffer was able to lyse the pellets and release some of the protein into the soluble fraction. However, when assays containing this lysate were analysed, it was quickly found that the detergent was deleterious to the performance of the HPLC column.

Sonication was therefore revisited, it being a lysis method which would be fully compatible with the downstream analysis. A new method was tried, in which the cell pellets were first resuspended in 100 μL buffer then transferred to 0.2 mL PCR tubes (Eppendorf). These suspensions were sonicated as before, with strips of 8 tubes on the surface of the chilled water bath for two rounds of 10 min continuous sonication. In principle, separation into tubes would give a larger contact surface with the water and a thinner barrier between the sonic waves and the cell suspensions. This method proved successful: after centrifugation, Bradford assays of the clarified lysates indicated that the levels of protein released were comparable to probe sonication of larger pellets. However, those assays also indicated that lysis was highly variable between tubes. This was despite frequent stirring of the water bath during sonication in an attempt to evenly distribute the strips between areas of high and low sonic intensity. Furthermore, there did not seem to be a link between where on the strip a tube was located and the degree to which the cells inside were lysed. This unpredictability was a concern, as variability in the lysis could mask genuine differences in conversion between mutants. The design of the screen assay was thus adapted to account for total protein concentration, which was taken as directly proportional to overall lysis.

The first sixteen mutants of *Xavier* (columns 1 & 2) were grown as above, and the pellets lysed with PCR-tube water bath sonication. Aliquots of those lysates were added to propargylation assays, which otherwise had the conditions described in the previous chapter: 5 mM substrate, 20 mM ATP and **86**, ~0.5 mg/mL *Uu*MAT and ~0.01 mg/mL *Ec*MTAN, both as lysates (Assay XII). The volume of *Rn*COMT mutant lysate could not practicably be adjusted for the concentration of each. A fixed amount of 20% (*v/v*) was

therefore added (Scheme 6.5).



Scheme 6.5: Assay for screening tolerance for propargylation by *Rn*COMT mutants. **93** represents a mixture of *meta* and *para* propargyloxy regioisomers.

After the reactions were initiated, 2 μL samples of the lysates were aliquoted to a new microplate and mixed with 100 μL Bradford reagent. The colour was allowed to develop for 5 min, then 5 μL ethanol was added to disperse bubbles. The absorbance of every well at 595 nm was then measured with a CLARIOstar plus plate reader. A set of standards with known concentrations of BSA was also measured, and all readings were taken with reference to a blank containing Bradford reagent but no protein. A calibration curve of absorbance vs. protein concentration was then plotted using the standards, and used to calculate the protein concentrations of every lysate.

The completed reactions were analysed by HPLC, and the conversions to propargylated product determined. The ratio of propargylation to background methylation was also calculated from the same traces, in order to detect mutants with increased specificity for propargylation. Conversions, protein concentrations and product ratios were then compared (Figure 6.7A). Conversion and propargylation/methylation ratio were both negatively correlated with protein concentration (i.e. degree of cell lysis), indicating that methyl poisoning was occurring. This also implied that simply identifying mutants associated with the highest conversions would not be sufficient to find hits. Instead, interesting mutants would have to be ones which performed well given their degree of lysis. That is, mutants whose points on the graph in Figure 6.7A lay significantly above the line of best fit.

In these results, *Xavier*-E2 appeared to meet that criteria. However, given the small number of wells sampled for this test, it was not yet known if this result was still within the variance for the whole plate. Nonetheless, these results demonstrated that the random mutagenesis workflow could be completed from start to finish. As such, the assay was

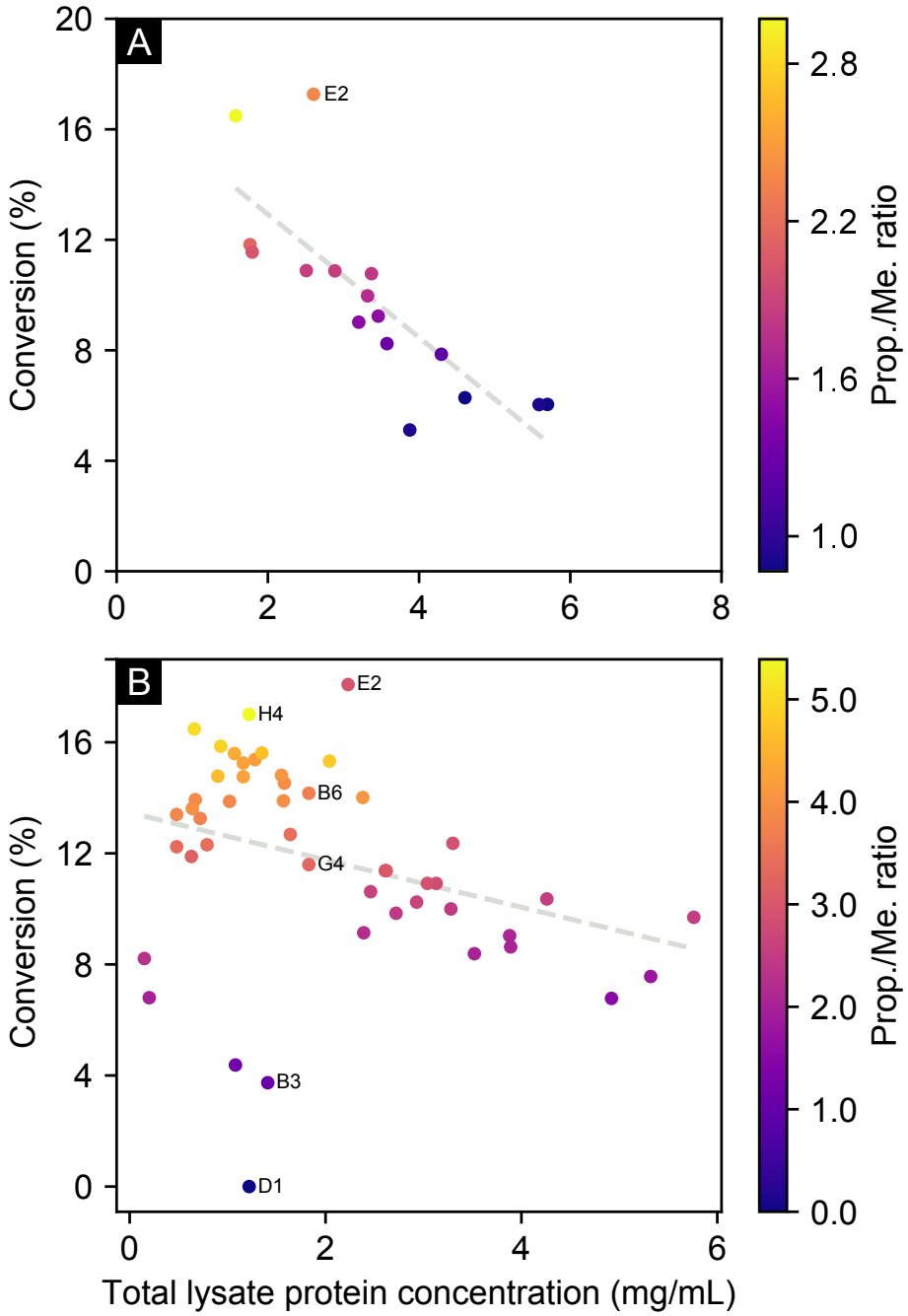


Figure 6.7: Analyses of propargylation assays involving subsets from *Xavier RnCOMT* mutant plate. A) columns 1 & 2. B) columns 1-6. Conversions calculated from HPLC peak areas of product and remaining starting material. Total protein concentrations calculated by Bradford assays. Line of best fit shown in grey. Mutants which were later sequenced are labelled with their microplate well locations. Assay XII conditions, with 20% (*v/v*) *Rn*COMT mutant lysate. HPLC method E.

repeated, this time including the first 48 mutants (columns 1-6) of *Xavier*. All the same conditions and procedures were used.

Upon analysis, the same general trend emerged (Figure 6.7B). A negative relationship between conversion and total protein concentration was apparent, but was not as strong as in Figure 6.7A. The majority of mutant candidates nonetheless aggregated around the line of best fit, suggesting that they were either not mutated (i.e. wild type) or carried mutations that did not substantially effect the activity of the enzyme. The subset did however feature outliers, such as B3 and D1, which showed greatly reduced conversion. Furthermore, E2, which had been identified as a potential hit in the first test screen, again gave the highest conversion of the subset. H4 was also of interest, having a particularly high ratio of propargylated to methylated product.

In order to identify the mutations that might be causing the properties of these outliers, six mutants were selected for sequencing. These six included E2 and H4 as possible hits, B3 and D1 as suspected knock-down mutants, and B6 and G4 to provide controls. Cultures of these six were inoculated from their glycerol stocks on *Xavier*, incubated overnight and the plasmid DNA extracted, then sent for commercial sequencing.

In the meantime, to enable direct comparison with the wild type, *Xavier* (*Xav*)-E2 was cultured and expressed at 500 mL scale using the standard procedure. An assay was then conducted with the same conditions as the screen (Scheme 6.5, Assay XII), but using clarified lysates of both the mutant and wild type, generated by probe sonication as normal. SDS-PAGE analysis and Bradford assays were conducted on the lysates to equalise the concentration of enzyme in the reactions.

Surprisingly, despite it twice showing above-average performance in the screens, *Xav*-E2 gave slightly lower conversion in this assay than wild-type *Rn*COMT (data not shown). This was explained, however, when the sequencing data was returned, and showed that none of the six genes were mutants. All were instead wild-type *Rn*COMT, except for *Xav*-D1, which was wild-type *Uu*MAT.

This had several implications. Firstly, the XL1-red lineage containing *Rn*COMT might have been contaminated with that containing *Uu*MAT. Second, there was a considerable spread of conversions all arising from wild-type enzyme despite similar protein concentrations, implying the assay was not reliable. Lastly, that the generation 2 plasmid library

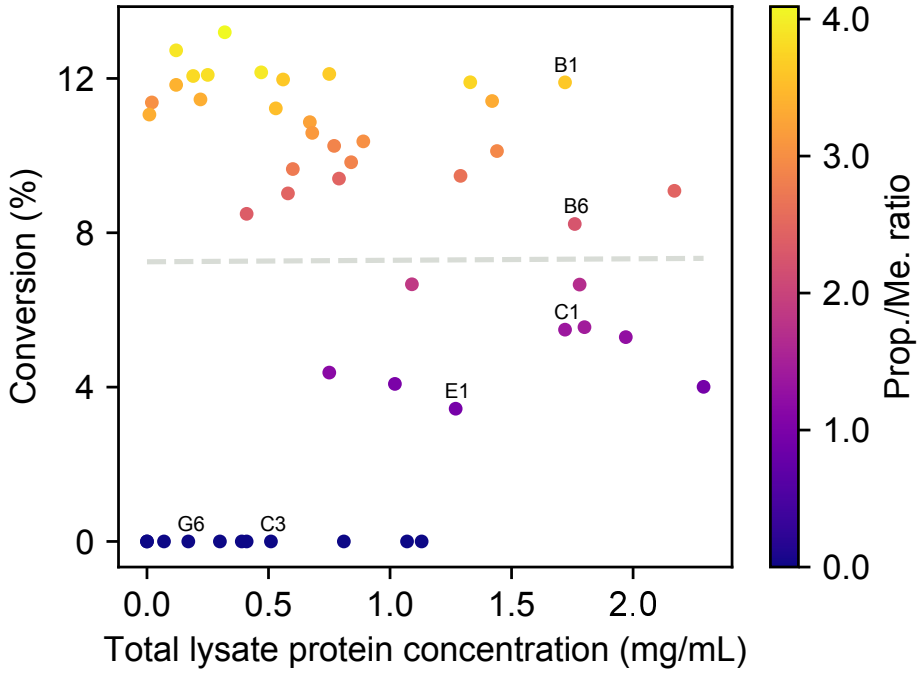


Figure 6.8: Analysis of propargylation assays involving a subset from *Beast* *Rn*COMT mutant plate, columns 1-6. Conversions and product ratio (Prop./Me.) calculated from HPLC peak areas of product and remaining starting material. Total protein concentrations calculated by Bradford assays. Line of best fit shown in grey. Mutants which were later sequenced are labelled with their microplate well locations. Assay XII conditions, with 20% (*v/v*) *Rn*COMT mutant lysate. HPLC method E.

may not contain enough mutants, or any.

The problem of insufficient sequence diversity had been anticipated, and prepared for by collecting the generation 3 plasmid library, which had been incubated in the mutator strain for 72 h longer (Figure 6.3). An exact repeat of the screen above was therefore conducted using columns 1 to 6 from *Beast*, the mutant stock plate made with the generation 3 library.

Analysis of this screen showed that the negative relationship between conversion and protein concentration had almost completely broken down (Figure 6.8). The trend was skewed, however, by the large number of samples of varying protein concentrations giving zero conversion. It was possible that this increased variation in conversion relative to *Xavier* was due to mutation of the enzymes. Therefore, as before, six mutants (*Beast* (*Bea*)-B1, -B6, -C1, -E1, -G6 and C3) were selected from across the spectrum of conversions. The strains containing these genes were cultured, and their plasmid DNA extracted

and commercially sequenced.

Sequencing showed that none of *Bea*-B1, -B6, -C1 or -E1 were mutants. *Bea*-G6 and -C3 were wild-type *Uu*MAT, which suggested that others which had given zero conversion were also contaminants, and that the contamination had grown between generation 2 and generation 3 of the XL1-red culture.

With these results, it was concluded that both the plasmid libraries and the assay approach were flawed. The libraries were less mutated than expected, and the library which should contain the most mutants was also highly contaminated. With respect to the assay, even with the steps to account for lysis, the variation in conversion between even wild-type enzymes meant there was far too much noise. It was not certain if beneficial mutants could even be identified against such a background.

The plasmid libraries for both *Rn*COMT and *Uu*MAT were therefore abandoned. Two new lines of XL1-red cells containing those genes were established by the procedures given in Section 6.4.1. The intention was to sustain these lineages for longer and thus accrue more mutations in the libraries, while taking additional precautions to avoid cross-contamination. Simultaneously, another search was conducted for cell lysis methods that would give less variation in protein concentration.

6.4.4 Testing of CHAPS-based lysis buffer

Lysis buffers based on detergents would, in theory, satisfy the criteria of producing even and predictable lysis across many cell pellets. While the degree of lysis by sonication could vary locally, lysis with a given detergent should only be affected by the concentration of detergent, incubation time, mixing and temperature, all of which could be controlled. An attempt had been made to use Triton X-100 (Figure 6.9) for this purpose. However, it was discovered that this detergent bound tightly to the matrix of the C-18 HPLC column, which further investigation indicated may be a general property of non-ionic detergents.

After sonication had proved too inconsistent to give reliable results, the idea of using detergents was returned to. Further investigation found that while non-ionic detergents bind to HPLC column matrices, ionic detergents do not tend to. Although their hydrophobic parts do interact with the non-polar matrix, the affinity of their charged groups for polar solvents allows them to be removed by washing. An example of this was shown in

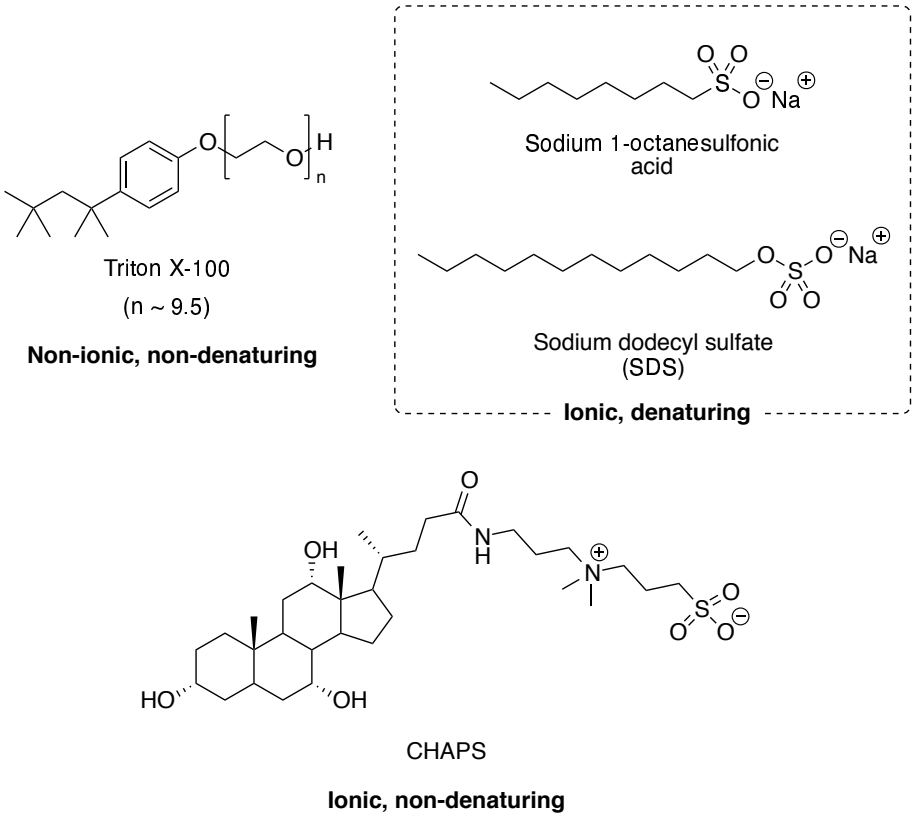


Figure 6.9: Detergent compounds used in this project.

Chapter 5, where a buffer containing ionic 1-octanesulfonic acid (Figure 6.9) was used for ion-pair reverse-phase HPLC. However, the detergent would also need to leave the enzyme intact, and many ionic detergents denature proteins precisely because of their greater polarity. Sodium dodecyl sulfate (SDS, Figure 6.9), the detergent used to denature proteins in SDS-PAGE, exemplifies this.

One detergent described as both ionic and non-denaturing was 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS, Figure 6.9). It is often used in molecular biology to purify delicate membrane proteins,[208] which was a good indication that it would not disrupt the activity of enzymes. To test whether CHAPS was able to lyse cells without compromising enzyme activity, a comparison was conducted with the sonication method. Sixteen 1 mL cultures of *E. coli* containing wild-type *Rn*COMT were grown up and induced in MagicMedia™ as they had been for the mutant screens. Once the cells had been pelleted in the microplate and the supernatant removed, half of the pellets were resuspended in 100 μL MT reaction buffer, transferred to 0.2 mL PCR tubes and lysed by water bath sonication. To the other half, 100 μL of the same buffer plus 1% CHAPS

(*w/v*) was added. Both the buffer and microplate containing the pellets were pre-chilled on ice. The cells were resuspended by agitation of the microplate with a vortex mixer until no pellets were visible, with frequent rests on ice. The microplate was then incubated at 10 °C for 10 min to allow lysis to complete, before both these and the sonicated lysates were centrifuged.

Propargylation reactions, with the same setup as the screening assays (Scheme 6.5, Assay XII), were then conducted using the *Rn*COMT lysates prepared by both methods. The total protein concentrations of the lysates were measured by Bradford assays. After HPLC analysis, the resulting conversions and protein concentrations were compared (Figure 6.10A).

Overall, there were no strong differences between the two lysates. Both gave conversions in a low range, as had been typical for the screening assays, but the lysates made by CHAPS lysis did not perform noticeably worse than those made by sonication. This was encouraging, and indicated that the detergent was indeed not denaturing the components of the assay. However, the range of protein concentrations made by CHAPS lysis was also just as wide as that from sonication. This was concerning, as the principle reason for seeking a new lysis method had been to reduce variance. It was noticed, though, that while the concentrations of sonicated lysates were distributed randomly across the row, those of the CHAPS lysates were arranged in almost perfect order from the beginning to the end (concentrations (mg/mL) were, in order of well: 4.8, 4.9, 6.5, 7.1, 7.2, 6.5, 8.5). This suggested that a bias had affected the degree of lysis. A new experiment was therefore conducted, in which the cultures were grown in wells distributed evenly across the microplate. Pellets from these cultures were lysed by CHAPS in the same manner as above, with particular care taken during resuspension to agitate all areas of the plate equally. When the total protein concentration of these lysates were tested, all were lower than the original experiments, but the variance was much smaller (Figure 6.10B). This indicated that a CHAPS-based lysis buffer could indeed give more even lysis of the cell pellets than sonication. With the first experiments confirming that the detergent would also not impact protein activity, this lysis method was taken forward into the screens of the new plasmid libraries.

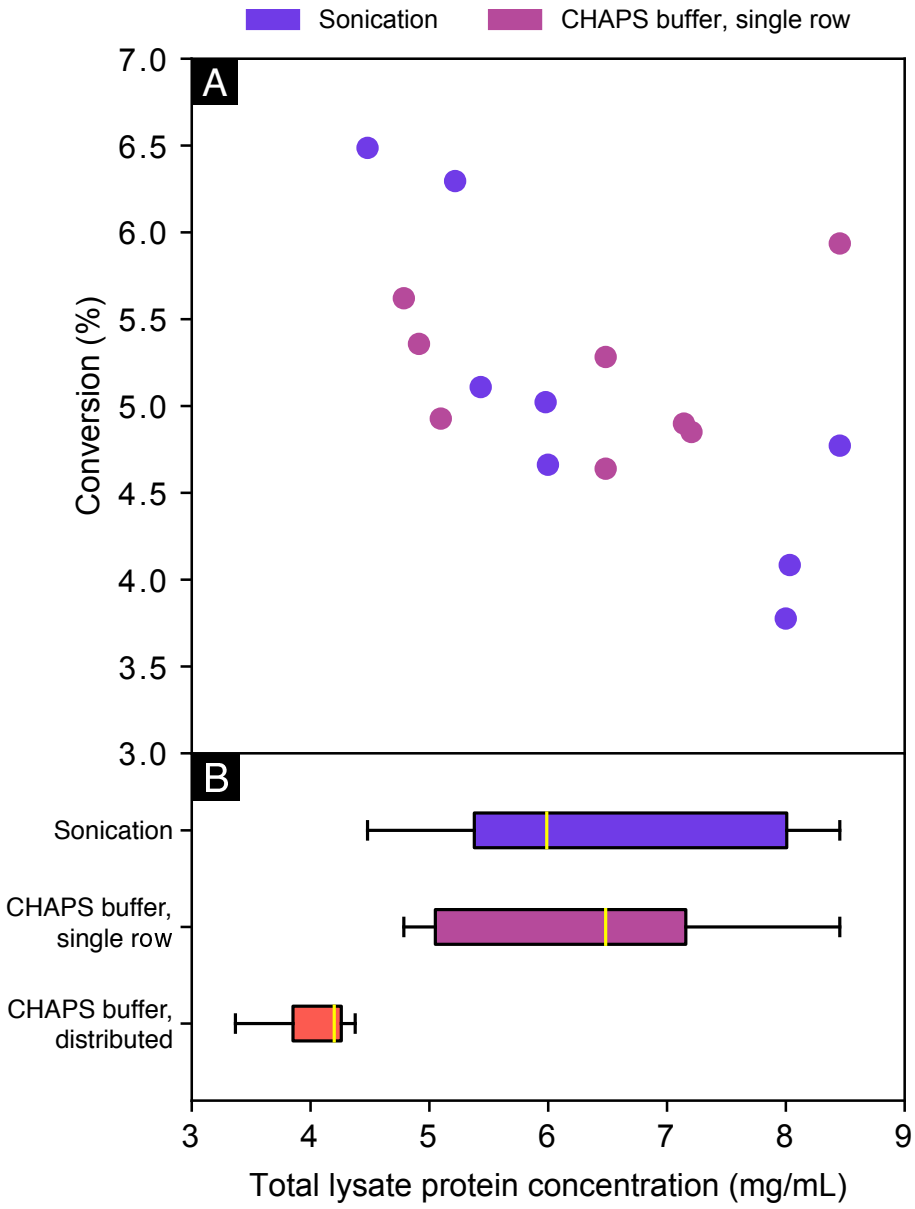


Figure 6.10: Comparison of *Rn*COMT lysates generated by sonication and CHAPS buffer lysis. A) Conversion vs. protein concentration relationships for lysates made by sonication and CHAPS buffer lysis with cell pellets in a single row. Conversion calculated from HPLC traces, using peak areas of product and remaining starting material. Assay XII conditions, with 20% (*v/v*) *Rn*COMT mutant lysate. HPLC method E. B) Protein concentration boxplots for lysates made by sonication, CHAPS buffer lysis with cell pellets in a single row, and CHAPS buffer lysis with pellets distributed across a plate.

6.4.5 Full screening of *Rn*COMT mutant library.

The new XL1-red cell lines were continued for five generations. Plasmid DNA was extracted from the fifth generation of each and sent for commercial sequencing. The resulting chromatograms showed no indication of cross-contamination between the lines carrying *Rn*COMT and *Uu*MAT. Both plasmid libraries were therefore transformed into *E. coli* BL21(DE3). Four mutant stock microplates were made from each set of transformants. The plates of *Rn*COMT mutants were named *Angel*, *Iceman*, *Banshee* and *Havok*. Together, these contained 384 potential mutants.

Each plate was screened by the standard method developed from trials described above. This comprised inoculation of 1 mL volumes of auto-induction MagicMedia™, overnight incubation at 37 ℃, pelleting by centrifugation, cooling on ice and lysis by resuspension and shaking in 100 μL lysis buffer (CHAPS 1% *v/v*). SDS-PAGE analysis and Bradford assays showed that, despite some outliers, enzyme expression and total protein concentration were largely consistent over the wells sampled (Appendix G.1).

The conditions of the initial screens were used again here (Scheme 6.5), although the fixed volume of *Rn*COMT mutant lysate was reduced from 20% to 10% (Assay XII). This was an attempt to strike a balance between limiting methyl poisoning and ensuring consistency, even from cultures with anomalously low expression or poor lysis. Analysis of the completed reactions was done by analytical HPLC. Due to the time taken to analyse each pair of assays (2 x 96 wells, ~26 h), the sample chamber of the instrument was cooled to 15℃ for the duration to limit evaporation and/or deterioration of compounds.

Analyses of the screens are given in Figure 6.11. A common pattern was observed in the distributions of conversions for all four. Each had a subset of poorly performing wells, another of wells which gave slightly higher than average conversions, and then a highly consistent majority between them. This consistent level, which was believed to be the wild-type conversion by consensus, was taken as validation of the new lysis method, which had aimed to reduce variation and thus noise in the interpretation.

Mutants from *Iceman* had a noticeably smaller average conversions than those from the other three plates (Figure 6.11). As this affected the whole plate, it was likely the result of some unintended difference in the culturing and/or processing of these lysates.

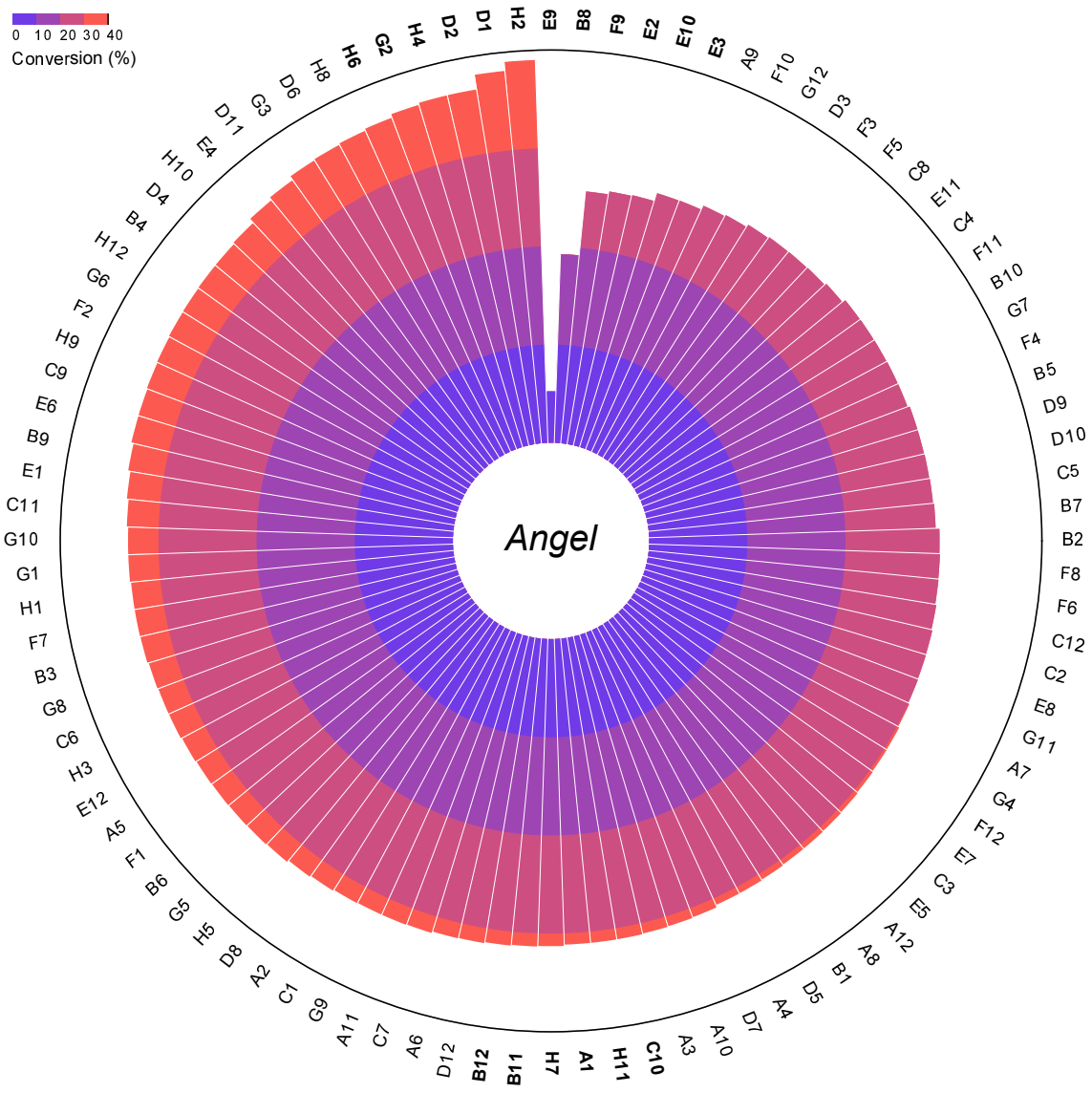


Figure 6.11: *Rn*COMT mutant screens. Well locations given around the outside of the charts, clockwise in ascending order of conversion. Conversions calculated using HPLC peak areas of product and remaining starting material. Locations of mutants that were later assayed again are shown in bold. Assay XII conditions, with 10% (*v/v*) mutant cell lysate. HPLC method E.

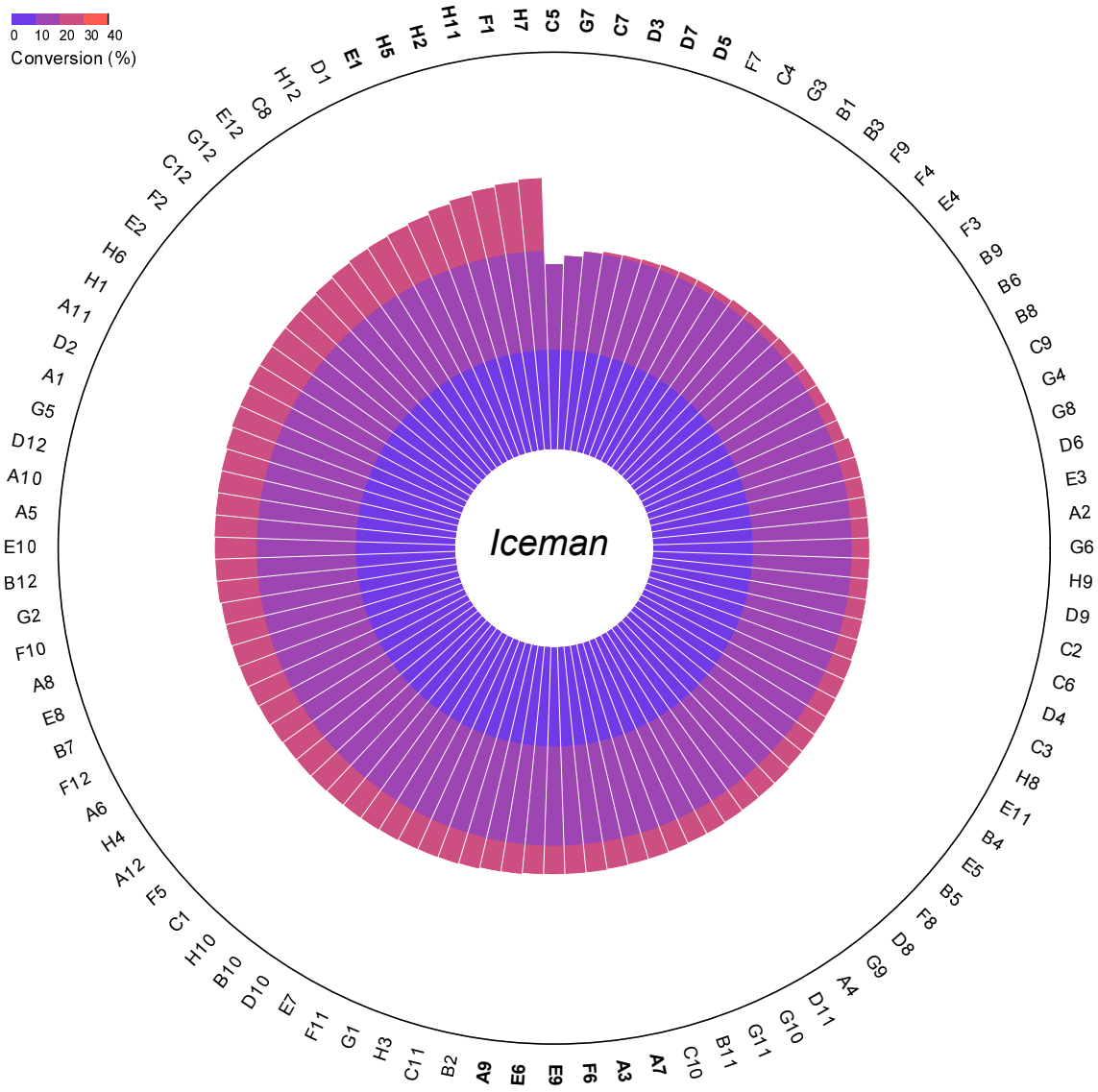


Figure 6.11: *Rn*COMT mutant screens (*cont.*). Well locations given around the outside of the charts, clockwise in ascending order of conversion. Conversions calculated using HPLC peak areas of product and remaining starting material. Locations of mutants that were later assayed again are shown in bold. Assay XII conditions, with 10% (*v/v*) mutant cell lysate. HPLC method E.

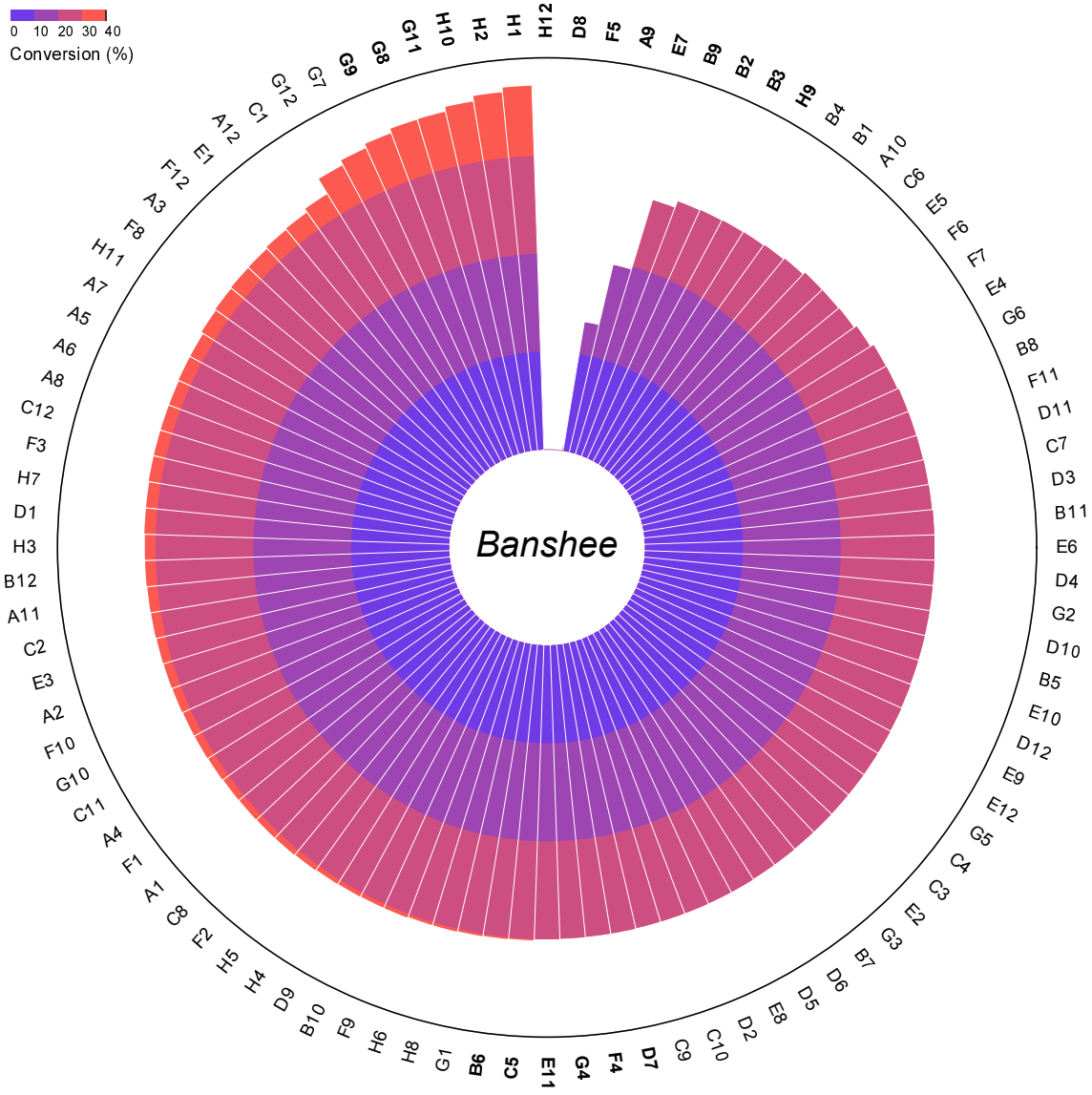


Figure 6.11: *Rn*COMT mutant screens (*cont.*). Well locations given around the outside of the charts, clockwise in ascending order of conversion. Conversions calculated using HPLC peak areas of product and remaining starting material. Locations of mutants that were later assayed again are shown in bold. Assay XII conditions, with 10% (*v/v*) mutant cell lysate. HPLC method E.

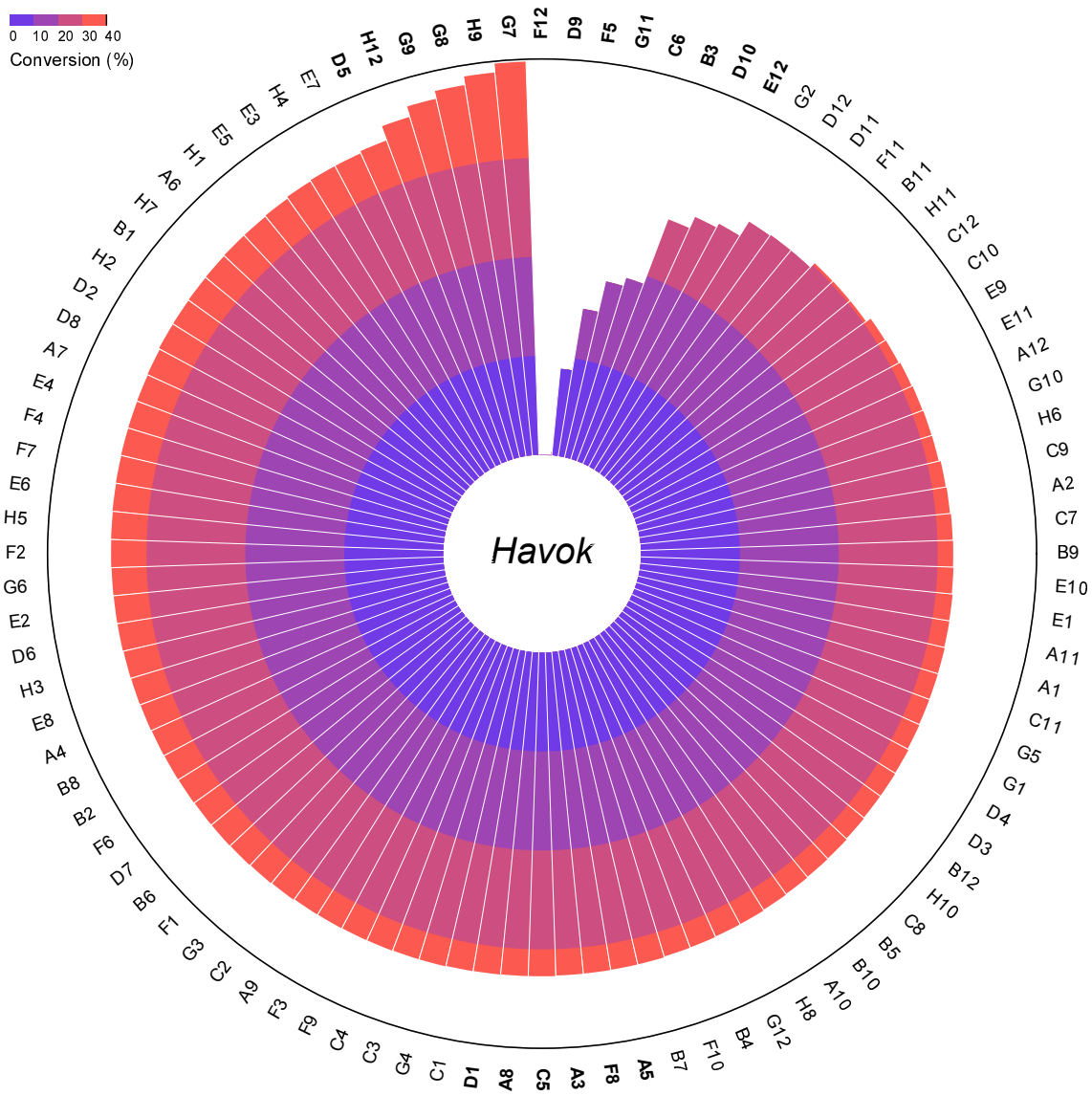


Figure 6.11: *Rn*COMT mutant screens (*cont.*). Well locations given around the outside of the charts, clockwise in ascending order of conversion. Conversions calculated using HPLC peak areas of product and remaining starting material. Locations of mutants that were later assayed again are shown in bold. Assay XII conditions, with 10% (*v/v*) mutant cell lysate. HPLC method E.

However, the data were still sufficient to see the distribution of conversions, so the screen was not repeated.

The mutants which gave the top six, bottom six and middle six conversions of each plate were screened again in duplicate assays (see the bolded well locations around each plate in Figure 6.11). The top six were expected to contain beneficial mutations, while the bottom six were expected to show knock-down mutations. The middle six were included as controls. For *Banshee* and *Havok*, the wells which gave zero conversion were assayed again in addition to the bottom six in order to rule out errors in the experimental setup.

At this stage, it was noticed that there was an over-representation of wells from the final two microplate rows (i.e. G and H) in the highest-converting mutants. No less than four and as many as six of the top six mutants in every plate were from these rows. This raised concerns that there may be a location-bias influencing the results. To visualise this, heatmaps were plotted which displayed the conversions in the row-and-column format of the microplates (Appendix G.2). The maps helped confirm that the bottom rows contained a disproportionate number of the highest-converting wells, but did not indicate any wider trend across the plate. If a location-bias was present, it would likely be introduced during the resuspension of cell pellets in lysis buffer, as this was done manually and involved judging by eye when the pellets had been dispersed. The plates were rotated throughout the resuspension to avoid this, but that may have been insufficient. Furthermore, the differences between the median conversion and highest conversion for all plates were so small, in terms of absolute percentage, that even minor biases might be noticeable. Nonetheless, it was expected that the duplicate assays, which would redistribute the mutants across the plate, would help remove effects based on location.

The mutants of interest were assayed by the same method as the original screen. Media volumes (1 mL) in a 96-well microplate were inoculated from the original mutant stocks. Additionally, eight volumes were inoculated with a strain containing wild-type *Rn*COMT to provide a direct comparison. These mutants were cultured and lysed by the established method, and the lysates assayed as above. Three additional reactions were conducted in parallel using *Rn*COMT lysate generated by probe sonication, in order to confirm, as had been indicated previously (Figure 6.10), that there were no significant differences between lysates generated by those two methods.

Analysis of the second assays showed that the patterns observed in the first screen were largely not upheld (Figure 6.12). The majority of duplicates gave conversions around 30%, with little to no correlation to their original results. This may have been due to the bias described above, wherein the higher conversion of wells in rows G and H was the result of a location-bias, so changing the position of those mutants caused the observed conversions to vanish.

With this result, the hope of finding an *Rn*COMT variant with a beneficial mutation had failed to materialise. Four mutants, however, were found to have notably below-average conversions in both screens: *Hav*-C6, *Hav*-G11, *Ang*-E9 and *Ban*-A9. Cultures of these four were grown and the plasmid DNA extracted and sent for commercial sequencing. The presence of mutations in these sequences would at least confirm that the library was mutated, and thus that the random mutagenesis workflow might still be fruitful with some modifications.

Sequencing showed that of the four plasmids sent, three contained mutations in the *Rn*COMT gene. Despite it giving low conversions in both the original screen and the duplicate assay, no mutations were found in the coding sequence of *Ban*-A9.

Ang-E9 had two substitutions: a transition of G**C**A to G**T**A, turning A118 into valine, and a transition from A**T**C to A**C**C, turning I123 into threonine. The first of these residues faces the solvent, and is part of a loop proximal to the adenine moiety of the cofactor (Appendix G.3). The second faces into the protein core, occupying a space between helices six and seven. Both of these mutations represent dramatic changes to the wild type, but due to their positions it is more likely that they would knock-down activity by destabilising the protein structure as opposed to affecting catalysis directly.

For *Havok*-C6 and -G11, transitions from TG**G** to TG**A** for the former, and from **C**AG to **T**AG for the latter, created stop codons at positions 38 and 7, respectively. This would normally be expected to destroy the protein. Therefore, that either mutant gave any conversion at all may be due to translation re-starting shortly after. However this would require both an in-frame methionine codon (i.e. M40) and for the preceding sequence to be interpreted as an initiation factor binding site, which is not certain. If translation did restart, though, residues 1 to 38 encompass helices one and two, which are exterior to the main Rossman fold and not involved in catalysis, which could explain how some catalytic

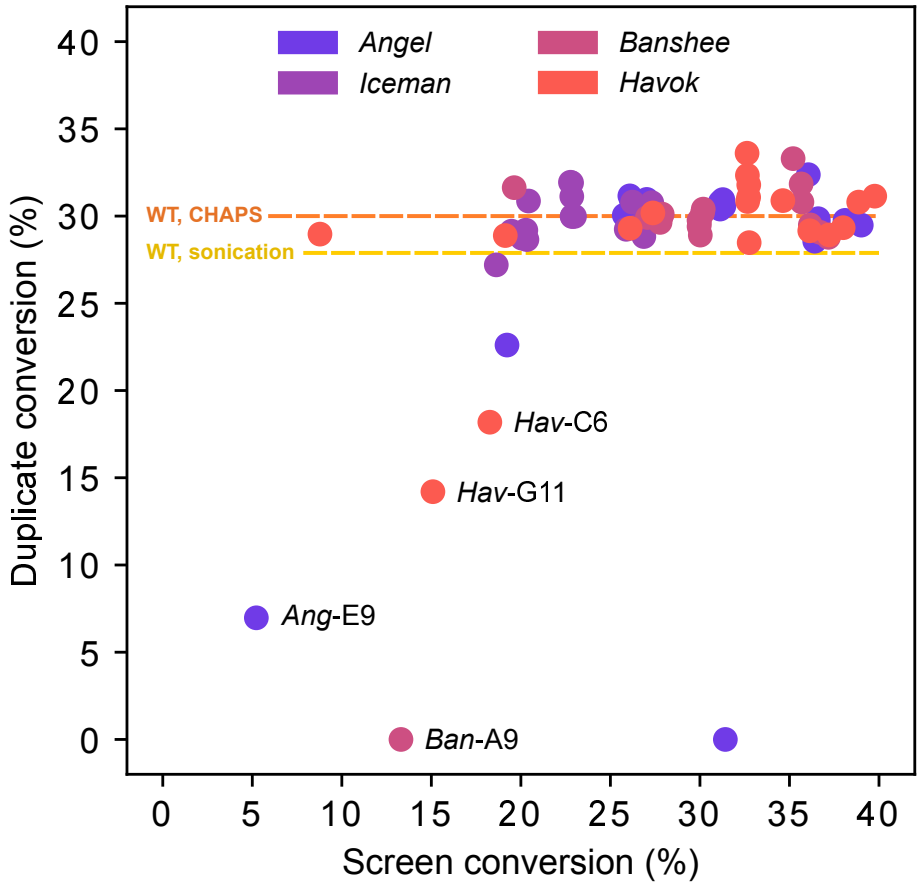


Figure 6.12: Analysis of duplicate assays versus original *Rn*COMT mutagenesis screen. Conversion calculated from HPLC peak areas of product and remaining starting material. Mutants which were later analysed by sequencing are labelled with the first three letters of the plate and their well location. Average conversion of wild-type *Rn*COMT lysate generated by two methods as part of the duplicate assays are shown as dashed lines. Assay XII conditions, with 10% (*v/v*) mutant cell lysate. HPLC method E.

activity survived. This would also represent an approximate 4.5 kDa loss in mass, but SDS-PAGE analysis of cell lysates from *Hav*-C6 and *Hav*-G11 did not show any noticeable decrease in the mass of the band corresponding to the enzyme, relative to wild-type (Supplementary Figure G.4). It was not clear, then, exactly what the consequences of these mutations were for the protein.

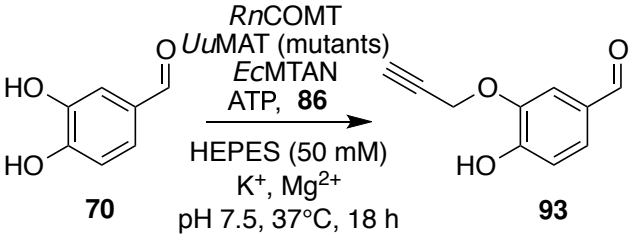
The outcome of this random mutagenesis screen was therefore not as hoped. However, it had confirmed that mutants were present in the library, and could be found by the assay despite some false-positives. In future work, it might be possible to screen the same library for changes in a different enzyme characteristic, such as substrate scope.

However, the effectiveness of this method might still be limited by the lack of diversity in the plasmid library. The choice of mutator strains to create that diversity was informed by expertise within the group and the relative simplicity of the method. However, the degree of mutation is difficult to control in mutator strains, and as explained in Section 6.4.1, inoculating each new generation to keep the strain alive created bottlenecks which also limited diversity. It could therefore be prudent to try other mutation methods. For example, error-prone rolling circle amplification mutagenesis[209] could be implemented into the current workflow with relative ease. This method involves using a DNA polymerase, whose fidelity has been compromised with $MnCl_2$, to amplify a plasmid library while introducing point mutations. This produces long concatemers made up of tandem repeats of the plasmid. When transformed into a host organism, each repeat separates into plasmids by intramolecular homologous recombination, giving the plasmid library with the mutated genes contained therein. As an *in vitro* method, this would be faster than culturing mutator strains and could be repeated as many times as needed to create a diverse library. There was not time in this project to pursue this option, but it could be considered for future attempts.

6.4.6 Screening of the *Uu*MAT mutant library

The plasmid library of mutated *Uu*MAT genes, extracted from the fifth generation of the mutator cell line, was transformed into *E. coli* BL21(DE3). The resulting colonies were used to create four mutant stock microplates, named *Psylocke*, *Domino*, *Gambit* and *Storm*. These plates were screened by the same method as for *Rn*COMT (Assay XII), albeit with a fixed volume of 20% (*v/v*) *Uu*MAT mutant lysate (Scheme 6.6). This value was chosen to match the approximate volume of sonicator probe-lysate normally needed to give 0.5 mg/mL enzyme.

After centrifugation of the mutant cultures, but prior to lysis, it was noticed that some of the cell pellets had a subtle but distinct difference in shade, appearing darker than others (Appendix G.5). *Psylocke*, *Domino*, *Gambit* and *Storm* had 15, 20, 21 and 26 of these wells, respectively. Subsequently, they were found to be more difficult to lyse, tending to take longer to break apart and resuspend than others. When SDS-PAGE analysis and Bradford assays were conducted with a random sample of all lysates, it was found that,



Scheme 6.6: Assay for screening tolerance of propargylation by *Uu*MAT mutants. **93** represents a mixture of *meta* and *para* propargyloxy regioisomers.

while *Uu*MAT expression in lysates from the normal pellets was consistent, the lysates from the dark pellets featured almost no protein at all (Appendix G.5). The reasons for any of this were not clear, as all cultures were grown up in the same conditions and this pattern had not been found with *Rn*COMT. The screen was continued as planned, albeit with the expectation that the lysates from the dark wells might perform poorly.

The results of these screens are given in Figure 6.13. On comparison with the screens of the *Rn*COMT mutants, two differences were immediately apparent. Firstly, conversions overall were lower. This included a larger number of mutants giving zero conversion, a large proportion of which had given dark pellets. A comparison of the distribution of conversions between dark and normal cell pellets is given in Appendix G.5. This was not surprising, given the apparent lack of any protein in these pellets, but the underlying reason for the phenotype remained unclear. Secondly, there was much greater variance in conversions across each plate. Whereas for *Rn*COMT, the variation had been largely confined to the very highest and very lowest ends of the distribution, for this screen the conversions were spread across the ranges. Even despite this variance, however, two hits were evident: *Psylocke*-H3 and *Storm*-D7 (Figure 6.13).

Because of the greater spread of conversions, a different strategy had to be employed to selecting mutants for the duplicate assays. Six groups of three were chosen from each plate, with each group intended to sample a distinct part of the distribution of conversions. The well locations of mutants which were chosen are shown in bold around each chart, and that of *Psylocke* illustrates this strategy. *Psy*-A8, A5 and G7 were chosen to sample those with zero conversion, while E6, F6 and A11 were chosen to sample group giving conversions around 5%. The next group, A9, C12 and C8 were intended to sample the next highest 'plateau', and so on, up to the highest converters (which were always

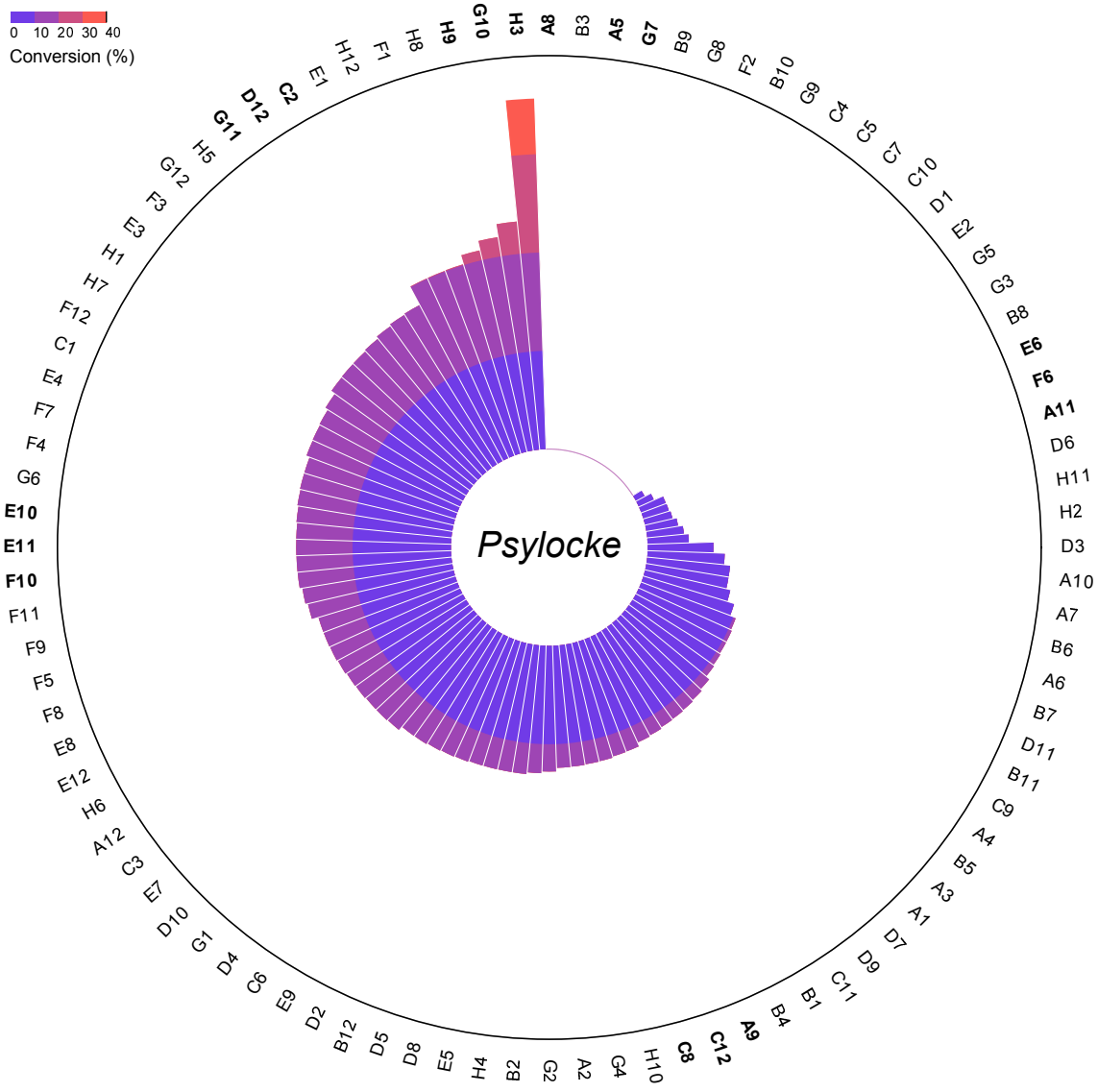


Figure 6.13: *Uu*MAT mutant screens. Well locations given around the outside of the charts, clockwise in ascending order of conversion. Conversions calculated using HPLC peak areas of product and remaining starting material. Locations of mutants which were later assayed again are shown in bold. Assay XII conditions, with 20% (*v/v*) mutant cell lysate. HPLC method E.

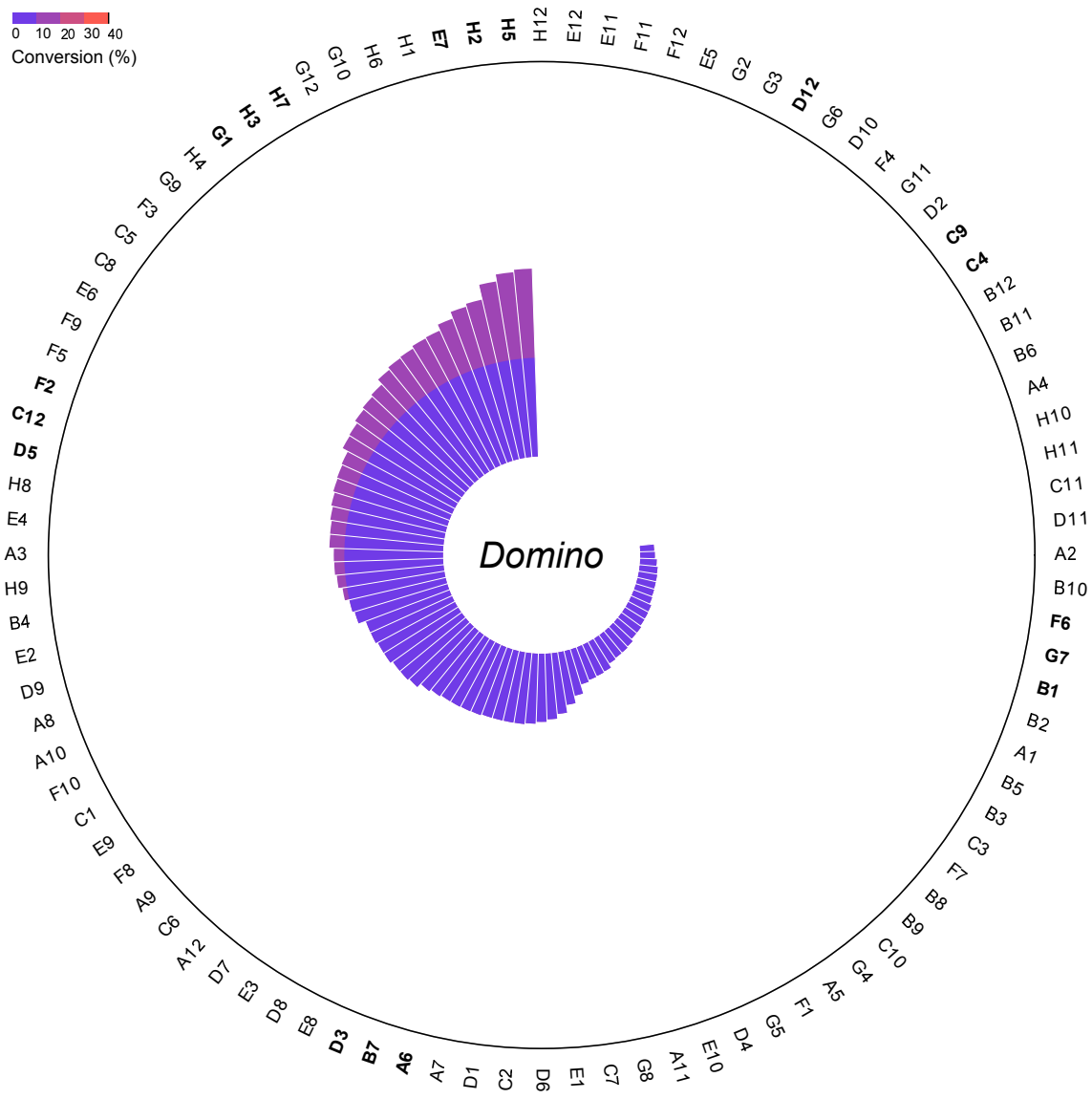


Figure 6.13: *Uu*MAT mutant screens (*cont.*). Well locations given around the outside of the charts, clockwise in ascending order of conversion. Conversions calculated using HPLC peak areas of product and remaining starting material. Locations of mutants which were later assayed again are shown in bold. Assay XII conditions, with 20% (*v/v*) mutant cell lysate. HPLC method E.

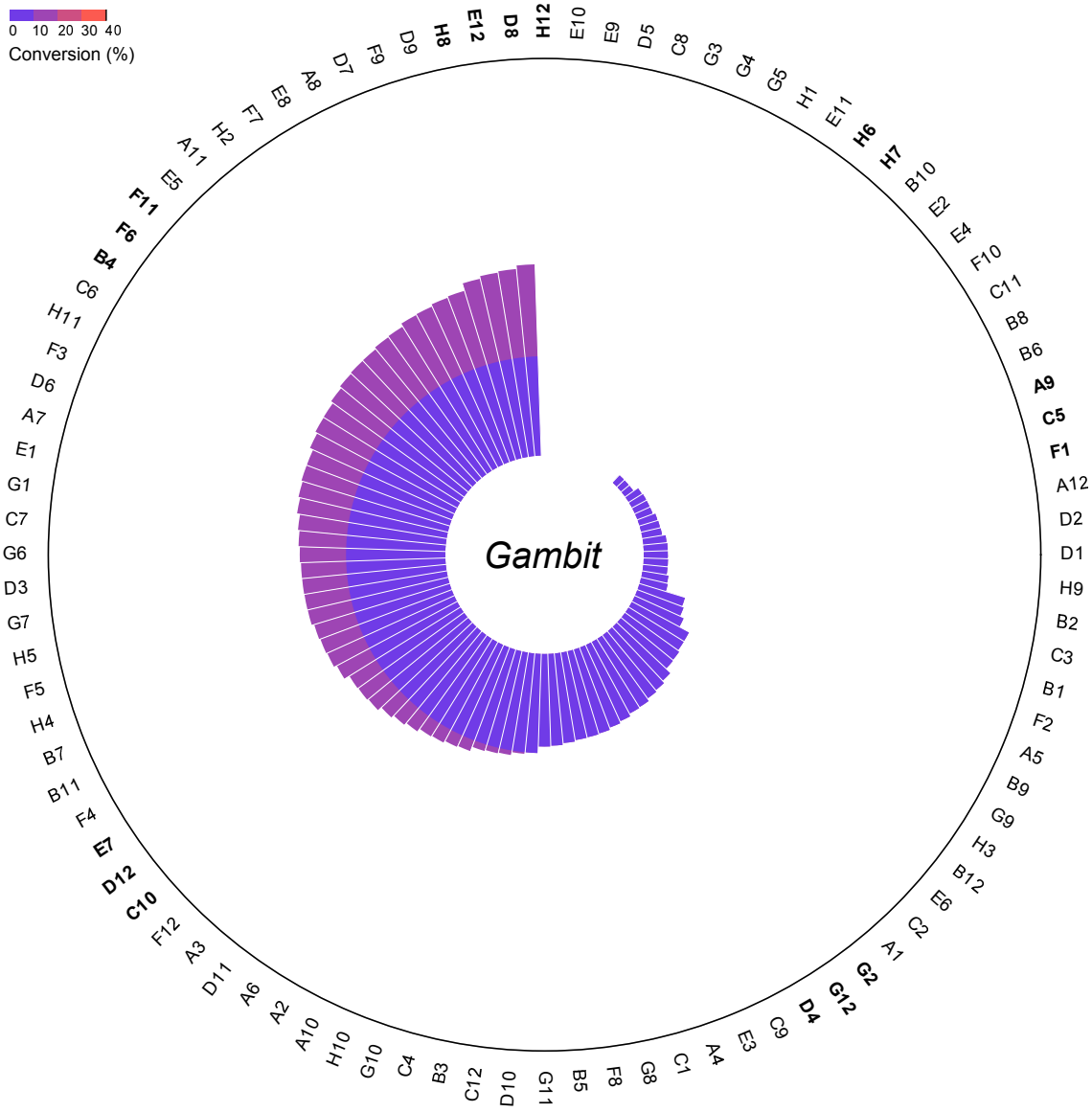


Figure 6.13: *Uu*MAT mutant screens (*cont.*). Well locations given around the outside of the charts, clockwise in ascending order of conversion. Conversions calculated using HPLC peak areas of product and remaining starting material. Locations of mutants which were later assayed again are shown in bold. Assay XII conditions, with 20% (*v/v*) mutant cell lysate. HPLC method E.

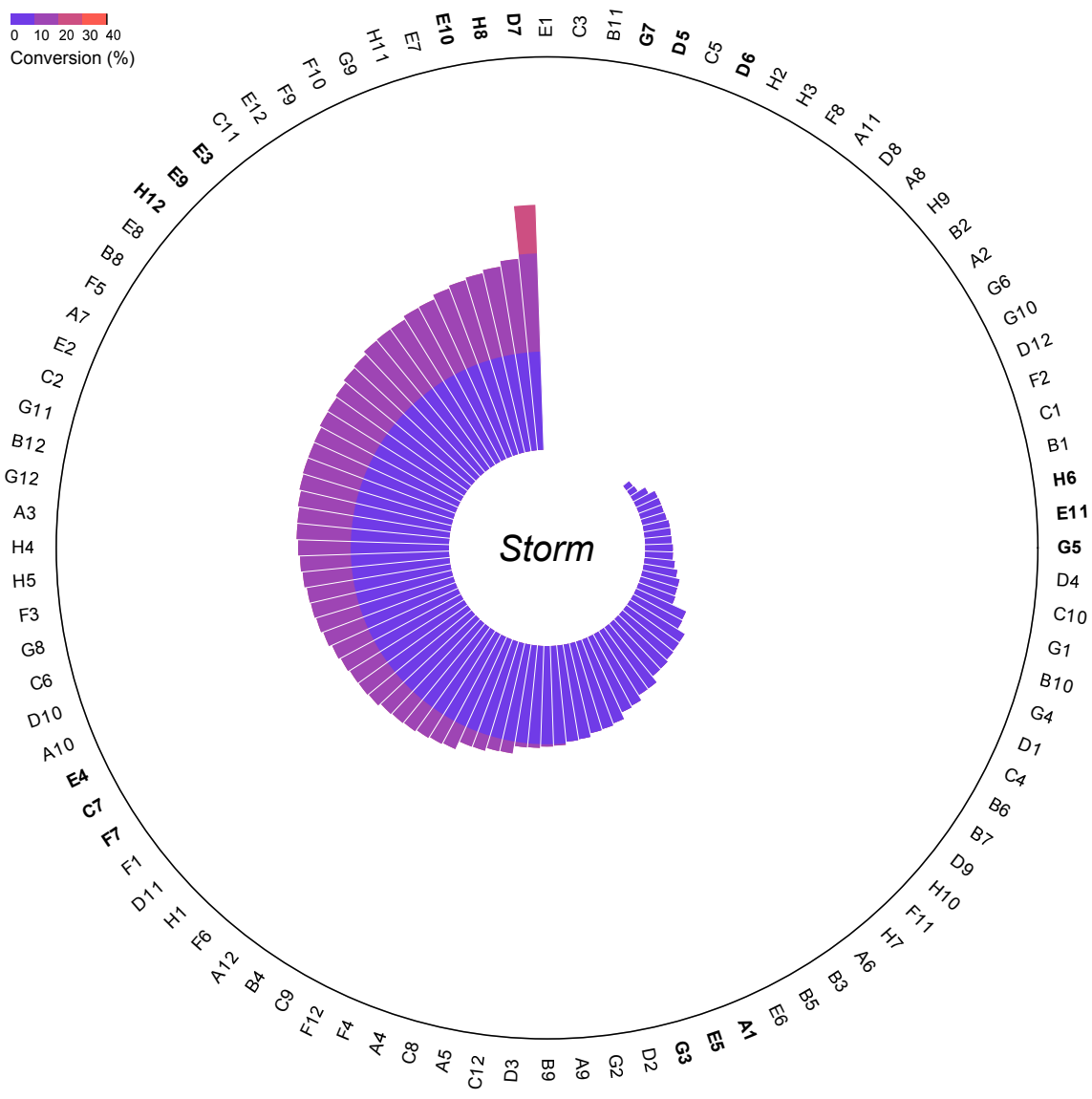


Figure 6.13: *Uu*MAT mutant screens (*cont.*). Well locations given around the outside of the charts, clockwise in ascending order of conversion. Conversions calculated using HPLC peak areas of product and remaining starting material. Locations of mutants which were later assayed again are shown in bold. Assay XII conditions, with 20% (*v/v*) mutant cell lysate. HPLC method E.

assayed again).

The duplicate plates were composed as before, with the original stocks for each chosen mutant used to inoculate 1 mL media volumes. To help confirm the authenticity of the two apparent hits, two additional volumes were inoculated for each at a separate location on the plate. Furthermore, eight wells were inoculated with wild-type *Uu*MAT as a direct control. The duplicate cultures were grown and assayed as per the standard procedure.

The results of the duplicate assays are given in Figure 6.14. There was greater correlation this time with the original screen although still some inconsistency (Figure 6.14A). Importantly, however, *Psy*-H3 and *Sto*-D7 again showed significantly higher conversions than the average for the plate, and the average for the wild type (Figure 6.14B). These higher conversions were consistent with the additional cultures located elsewhere on the plates, further ruling out location-biases as the cause and indicating that the result was due to changes in the MAT sequence. Surprisingly, *Domino*-E7 also showed a very high conversion in the duplicate, despite not doing so in the original screen. Despite the inconsistency, this was considered noteworthy. Therefore, cultures of *Psy*-H3, *Sto*-D7 and *Dom*-E7 were grown, their plasmid DNA extracted and sent for commercial sequencing.

It should also be noted that the cell pellets which were dark in the original screen were dark again in the duplicates. This pointed to the appearance, the lack of protein and the subsequent lack of conversion in the assays as all integral characteristics of those mutants, and not due to circumstantial influences. This phenomenon was not investigated further in this project. However, one hypothesis could be that the plasmids of all these mutants shared a mutation that was deleterious to the *E. coli* cells. Growth did not seem to be impacted, as the dark pellets were as large as the normal ones. Therefore, it is possible that the mutation was in the sequence of *Uu*MAT, so would only have an impact on the health of the cells once the gene was induced. As an MAT, *Uu*MAT would be expected to participate in the metabolism of the cells alongside endogenous *Ec*MAT. Therefore, if a mutation affected its function in some serious way, that could have consequences for the cells themselves. Further characterisation of the dark mutants would be needed to develop this theory.

While the sequencing of the positive hits was awaited, an experiment was conducted to ensure that the differences observed in the screen were upheld when the enzymes

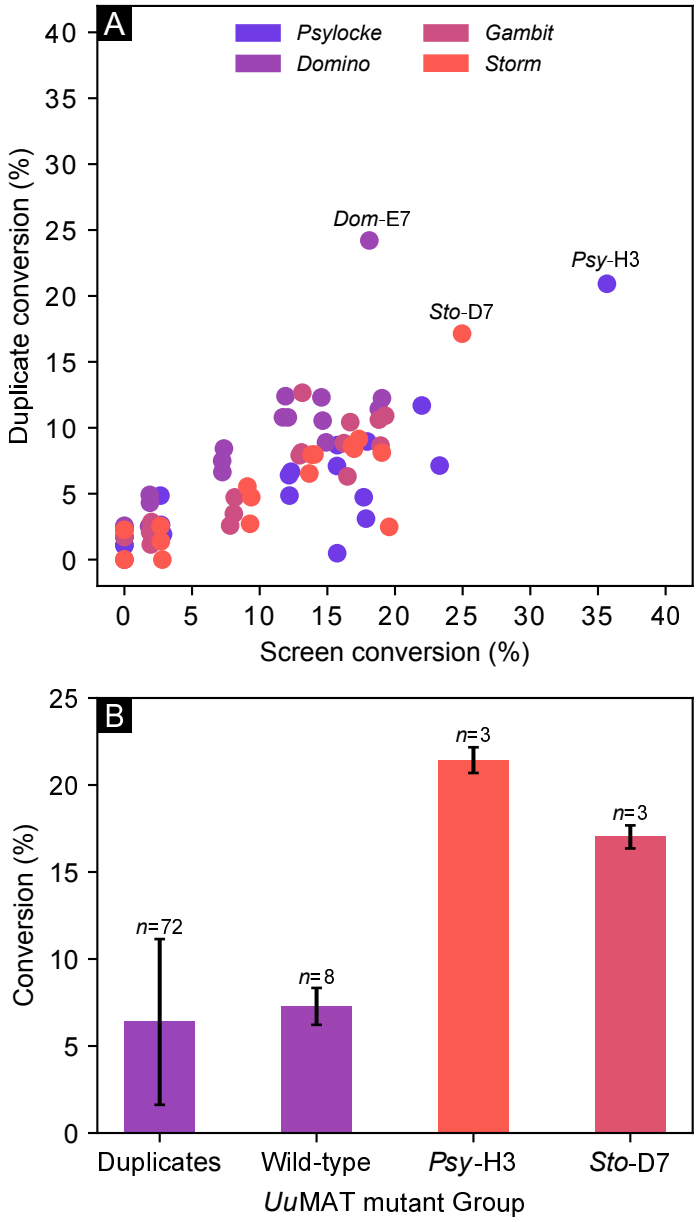
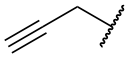
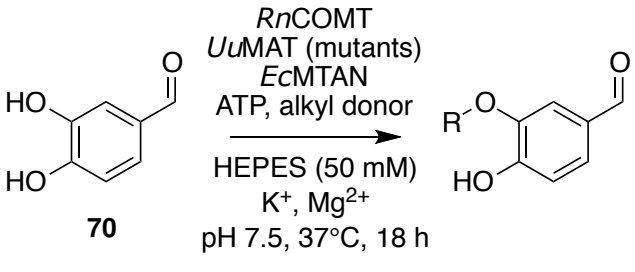


Figure 6.14: Analysis of duplicate assays following *Uu*MAT mutant screen. A) Conversions from duplicates versus original screens. Conversion calculated from HPLC peak areas of product and remaining starting material. Mutants which were later analysed by sequencing are labelled with the first three letters of the plate and their well location. B) Further comparison between average conversions for: the whole duplicate plate; wild-type *Uu*MAT; the two positive hits from the original screen. Error bars show one standard deviation above and below the mean. Number of data points contributing to the averages and standard deviations are given above the error bars. Assay XII conditions, with 20% (*v/v*) mutant cell lysate. HPLC method E.

were cultured on a larger scale. The three *Uu*MAT mutants were grown in 500 mL media volumes, induced by the normal procedure and the resulting cell pellets lysed by probe sonication. Assays were then conducted to compare the variants to the wild-type for the alkylation of **70** with three alkyl donors (Scheme 6.7, Assay XII). This small panel of methionine and two analogues was intended to indicate if the mutations of these hits would have broader impacts on alkyl donor acceptance than the one observed in the screen.

Alkyl donor	R	Product
8a		**94**
32a		**109**
86		**93**



Scheme 6.7: Propargylation assay to compare wild-type *Uu*MAT and hits from the random mutagenesis screen.

As with the rational mutants at the beginning of the chapter, it was acknowledged that purifying the enzymes would allow more accurate control over their concentration and so offer a more direct comparison. However, this assay, with estimated enzyme concentrations from SDS-PAGE and Bradford assays, was considered sufficient to give a first indication that the results of the screen were reliable.

Analysis of these reactions showed that mutants did, in fact, give different results when expressed and lysed this way. The most promising hit, *Psy*-H3, gave the highest conversions for ethylation and propargylation of **70**, between 23% and 28% more than the wild type, respectively (Figure 6.15). This was not nearly as great a difference as the nearly threefold increase seen in the duplicate assays (Figure 6.14). *Dom*-E7 showed nearly equivalent conversion for ethylation compared to wild-type *Uu*MAT, but -30% conversion for propargylation. This, again, was in contrast to the result seen in the duplicate assay. *Sto*-D7 showed a slight improvement over the wild-type for ethylation, but no significant difference for propargylation. None of the three variants had therefore performed as predicted from the screen, although *Psy*-H3 had maintained some benefit.

Unfortunately, the sequencing data gave the final indication that the results from the

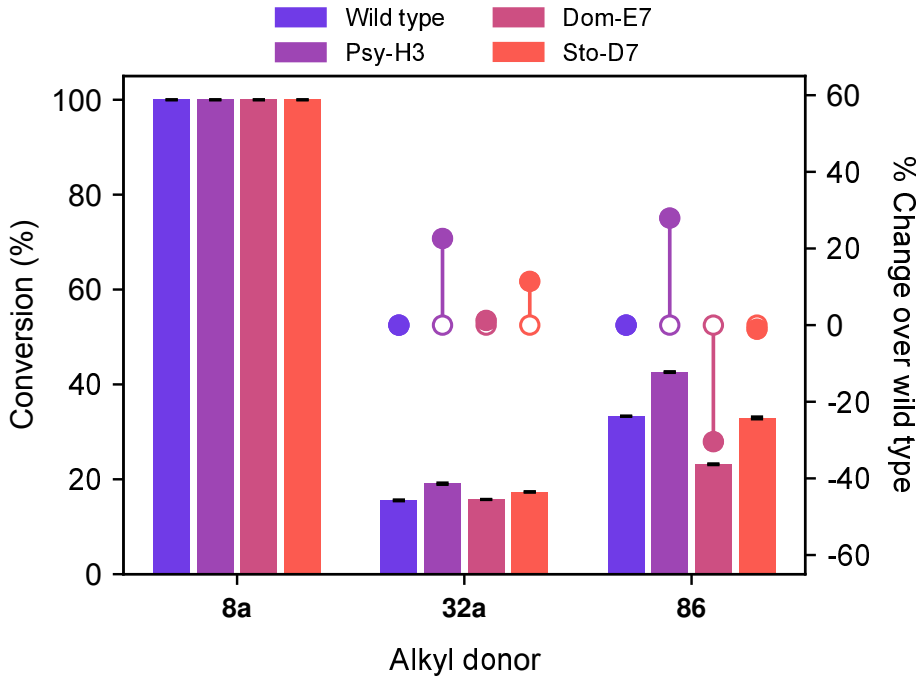


Figure 6.15: Analysis of assay to compare hits from *Uu*MAT random mutagenesis screen with wild-type enzyme in terms of methionine analogue (i.e. alkyl donor) acceptance. Bars show absolute conversion, dots and lines show percentage change over wild-type enzyme for that substrate. Error bars represent one standard deviation above and below the mean of three replicates. All enzymes produced by larger-scale culturing and expression and lysed by probe sonication. Assay XII conditions. HPLC method E.

screen were unreliable. No mutations were detected in any of the three hits. This was especially surprising for *Psy*-H3, given that it had shown higher conversion than wild-type *Uu*MAT on four separate occasions (the original screen, the duplicate assays, the large-scale culture assays, and an initial, failed duplicate assay not shown, in which the uniquely high conversion from this well was the telling sign that there had been a mistake when the reactions were composed). It is possible that there is a difference with this strain, but that it is elsewhere on the plasmid and is somehow influencing expression. This would be difficult to confirm however, and of limited relevance to this investigation. Furthermore, no obvious signs of increased expression were observed by SDS-PAGE analysis.

6.5 Conclusions

Ultimately, the methods described in this chapter were not successful in finding improved enzyme variantsm. The apparent lack of sequence diversity in the mutant libraries, and another mutagenesis method that might be used, were discussed in Section 6.4.5. Future efforts might also benefit from estimating the degree of mutation in a library prior to screening, such as with parallel mutation of an easily-visualised reporter gene. The inaccuracy of the assay method was another concern, given that wild-type *Uu*MAT was consistently shown as a positive hit. However, it may be that, if the plasmid library was more diverse, it would generate genuine positive hits and these false-positives would not be so consequential. Despite these disappointments, a positive outcome of the work in this chapter was the development of a relatively straightforward way to grow, lyse and assay large numbers of potential mutants. Though there are still drawbacks, and though no mutants of real interest were found in this investigation, that core workflow may be adapted and applied to other random mutagenesis efforts in the future that aim to improve aspects of an MT cascade.

Chapter 7

Conclusions and Future Work

The aim of this project was to investigate how MTs might be used to modify THIQ structures in scalable, economic reactions, with the ultimate goal of accessing new compounds with medicinal potential. At its conclusion, several important strides had been made towards this aim, with transfer of several chemical groups onto THIQ structures and extensive study of the challenges confronting these types of reactions. Equally important, however, have been the frontiers for new investigations that have become evident as this work unfolded.

The work presented in Chapter 2 demonstrated a coupled NCS-MT cascade able to convert basic starting materials into methylated THIQs. The use of an MAT enzyme to generate the SAM cofactor *in situ*, and an MTAN to degrade the SAH byproduct, allowed the limitations of expensive SAM supply to be overcome. The enzyme reactions could thus be performed at preparative scale, and the products isolated and characterised to give insight into the regioselectivity of *Rn*COMT.

Chapter 3 documented an effort to develop a set of new *O*-MTs as biocatalysts. The intention was to replicate the success of *Rn*COMT and catechols with other medicinally relevant chemical groups, including phenols, aliphatic hydroxyls and sugars. Ultimately, however, none of these attempts were successful. Some of the novel MTs were unable to be cloned, others did not express adequately in the conditions tested and others still were expressed and even purified, but were not active in assays. As was noted at the end of that Chapter, the approach taken was likely not the optimal one for the given goal. MTs are by nature selective, so finding one that will perform a desired, non-natural reaction

is difficult. Spreading the effort thin over a number of different chemical groups further reduced the likelihood that a suitable enzyme would be found for any given target. To learn from this, a future attempt to find new MT biocatalysts might benefit from picking just one target species and screening as many candidate MTs as possible. These candidates could be sourced both from from biosynthetic pathways described in the literature, as was the case for those explored here, as well as homologues from metagenomic databases in order to maximise the diversity of the pool.

The remaining Chapters described work towards adapting the system for THIQ methylation, as shown in Chapter 2, to accept *S*-alkyl analogues of methionine. Chapter 4 began with L-ethionine, and introduced new MAT enzymes with higher tolerance for these analogues. Conversions for ethylation were far lower than for methylation, however, especially with THIQs. Longer NCS-MT cascades integrating L-ethionine were also attempted, but likewise showed limited conversion. A computational docking study was therefore conducted to probe the reasons for the poor acceptance. Based on the results, it was hypothesised that the additional length of the *S*-ethyl group created major steric clashing in the *Rn*COMT active site. A mutant MT was designed with the aim of reducing that clashing, but when assayed was found to in fact give lower conversion than the wild-type.

The work in Chapter 5 continued to integrate methionine analogues into the casacades, but switched from L-ethionine to *S*-propargyl-DL-homocysteine. This analogue held the promise of being able to install an alkyne click handle onto the THIQs. Furthermore, due the increased chemical activation of this analogue (relative to ethionine), it was hypothesised that conversions would be greater than those seen in Chapter 4. After selection of a new MAT with tolerance of *S*-propargyl-DL-homocysteine, transfer of the alkyne group by *Rn*COMT was demonstrated for both a simple model substrate and, to a limited degree, for a panel of THIQs. These experiments in particular were aided by a published synthetic route towards DL-methionine analogues, which was adapted to give the *S*-propargyl analogue in plentiful quantities at low cost. Obstacles to efficient biocatalysis persisted, however. Regioselectivity with the model substrate was low, and it was found that the use of cell lysate preparations of the enzymes compromised the efficiency of the reaction (hypothesised to be due to cytosolic methionine and SAM). Remedies for these problems were explored, but the fundamental issue of low analogue tolerance at the MAT and MT

stages of the reactions could not be solved by optimisation.

Therefore, in the work described in Chapter 6, both *Rn*COMT and *Uu*MAT were subjected to engineering in order to improve the productivity of the cascades. Rational active site mutants were designed and expressed based on published findings in similar systems, but none were successful at improving conversion. Meanwhile, a random mutagenesis workflow was developed in an attempt to find beneficial but hard to predict substitutions. A large number of potential mutants were screened, with some knockdown mutants identified. Positive hits were also found in the *Uu*MAT screen, but further investigation revealed them to be false positives.

The effort to integrate methionine analogues into the THIQ cascades thus showed some promising first signs, but will need further attention before it is a useful biocatalytic method. In particular, future work could focus on refining or replacing the random mutagenesis method used in this chapter in order to increase the chances of finding positive hits. However, *Rn*COMT itself may be particularly intractable to mutagenesis beyond what has already been published, due to its small active site and thus the limited number of positions that can be manipulated. In that case, it may be prudent to broaden the investigation to homologues of *Rn*COMT, other catechol-*O*-MTs, or *N*-MTs able to target the secondary amine on THIQs. These would still be able to diversify and functionalise the compound family, but may offer more active site plasticity.

Additionally, the 'methyl poisoning' effect observed when cell lysates were used in assays could be addressed in future investigations. If that hypothesis is indeed correct, it would be expected to also impact conversion when whole cells were used, therefore finding a method to reduce it may be important to the wider practicality of MTs as biocatalysts. A partial solution could be the engineering of an MAT to be selective against methionine. Otherwise, an additive or system that is orthogonal to the main enzyme cascade could be used to deplete the methylating molecules in the lysates.

Methyltransferases are the agents of an ancient and foundational cellular process. Through eons of evolution, nature has learned to concentrate the extraordinary energy of ATP onto a tiny chemical fragment, then use that unassuming methyl group to control, regulate, build and activate in hundreds of ways. This project was first inspired by the promise of using that power to reveal and refine latent bioactivity in compounds. Attempts

to realise that promise have been fraught with challenges, from the tight selectivity of methyltransferases to the lability of the cofactor at the heart of the process. However, as this thesis has hopefully shown, the magic of the methyl is only the start of what methyltransferases could offer. With research and engineering, we might yet develop enzymatic tools capable of unlocking a vast garden of chemical opportunity, that we might wander in search of the next great solution.

Bibliography

[1] M. Egbertson, G. B. McGaughey, S. M. Pitzenberger, S. R. Stauffer, C. A. Coburn, S. J. Stachel, W. Yang, J. C. Barrow, L. A. Neilson, M. McWherter, D. Perlow, B. Fahr, S. Munshi, T. J. Allison, K. Holloway, H. G. Selnick, Z. Yang, J. Swestock, A. J. Simon, S. Sankaranarayanan, D. Colussi, K. Tugusheva, M. T. Lai, B. Pietrak, S. Haugabook, L. Jin, I. W. Chen, M. Holahan, M. Stranieri-Michener, J. J. Cook, J. Vacca and S. L. Graham, *Bioorganic and Medicinal Chemistry Letters*, 2015, **25**, 4812–9.

[2] B. P. Fauber, O. René, Y. Deng, J. Devoss, C. Eidenschenk, C. Everett, A. Ganguli, A. Gobbi, J. Hawkins, A. R. Johnson, H. La, J. Lesch, P. Lockey, M. Norman, W. Ouyang, S. Summerhill and H. Wong, *Journal of Medicinal Chemistry*, 2015, **58**, 5308–22.

[3] G. Lunn, B. J. Banks, R. Crook, N. Feeder, A. Pettman and Y. Sabnis, *Bioorganic and Medicinal Chemistry Letters*, 2011, **21**, 4608–11.

[4] M. A. Walker, *Bioorganic and Medicinal Chemistry Letters*, 2017, **27**, 5100–8.

[5] H. Schönherr and T. Cernak, *Angewandte Chemie - International Edition*, 2013, **52**, 12256–12267.

[6] J. Claesen and M. Bibb, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 16297–302.

[7] J. Lee, Y. Hao, P. M. Blair, J. O. Melby, V. Agarwal, B. J. Burkhart, S. K. Nair and D. A. Mitchell, *Proceedings of the National Academy of Sciences of the United States of America*, 2013, **110**, 12954–9.

[8] L. Najmanová, E. Kutejová, J. Kadlec, M. Polan, J. Olšovská, O. Benada, J. Novotná, Z. Kameník, P. Halada, J. Bauer and J. Janata, *ChemBioChem*, 2013, **14**, 2259–62.

[9] C. S. Leung, S. S. Leung, J. Tirado-Rives and W. L. Jorgensen, *Journal of Medicinal Chemistry*, 2012, **55**, 4489–500.

[10] P. R. Andrews, D. J. Craik and J. L. Martin, *Journal of Medicinal Chemistry*, 1984, **27**, 1648–57.

[11] D. J. Kopecky, X. Hao, Y. Chen, J. Fu, X. Y. Jiao, J. C. Jaen, M. G. Cardozo, J. Liu, Z. Wang, N. P. Walker, H. Wesche, S. Li, E. Farrelly, S. H. Xiao and F. Kayser, *Bioorganic and Medicinal Chemistry Letters*, 2008, **18**, 6352–6.

[12] E. J. Barreiro, A. E. Kümmerle and C. A. Fraga, *Chemical Reviews*, 2011, **111**, 5215–46.

[13] R. Angell, N. M. Aston, P. Bamborough, J. B. Buckton, S. Cockerill, S. J. DeBoeck, C. D. Edwards, D. S. Holmes, K. L. Jones, D. I. Laine, S. Patel, P. A. Smee, K. J. Smith, D. O. Somers and A. L. Walker, *Bioorganic and Medicinal Chemistry Letters*, 2008, **18**, 4428–32.

[14] T. Schindler, W. Bornmann, P. Pellicena, W. T. Miller, B. Clarkson and J. Kuriyan, *Science*, 2000, **289**, 1938–41.

[15] X. Wen and T. Walle, *Xenobiotica*, 2006, **36**, 387–397.

[16] W. Wein, *ESMO Open*, 2016, **1**, e000033.

[17] F. Ghirga, M. Mori, A. Bonamore, L. Calisti, A. Boffi, A. Macone, I. D'Acquarica and B. Botta, *International journal of molecular sciences*, 2017, **18**, 2464.

[18] M. Pilla, M. Andreoli, M. Tessari, S. Delle-Fratte, A. Roth, S. Butler, F. Brown, P. Shah, E. Bettini, P. Cavallini, R. Benedetti, D. Minick, P. Smith, B. Tehan, P. D'Alessandro, O. Lorthioir, C. Ball, V. Garzya, C. Goodacre and S. Watson, *Bioorganic and Medicinal Chemistry Letters*, 2010, **20**, 7521–4.

[19] D. Aynetdinova, M. C. Callens, H. B. Hicks, C. Y. Poh, B. D. Shennan, A. M. Boyd, Z. H. Lim, J. A. Leitch and D. J. Dixon, *Chemical Society Reviews*, 2021, **50**, 5517–63.

[20] M. R. Bennett, S. A. Shepherd, V. A. Cronin and J. Micklefield, *Current Opinion in Chemical Biology*, 2017, **37**, 97–106.

[21] P. A. Jones, *Nature Reviews Genetics*, 2012, **13**, 484–492.

[22] W. Ren, L. Gao and J. Song, *Genes*, 2018, **9**, 620.

[23] F. Lyko, *Nature Reviews Genetics*, 2018, **19**, 81–92.

[24] M. A. Sánchez-Romero, I. Cota and J. Casadesús, *Current Opinion in Microbiology*, 2015, **25**, 9–16.

[25] J. Murn and Y. Shi, *Nature Reviews Molecular Cell Biology*, 2017, **18**, 517–27.

[26] Z. Li and J. B. Stock, *Biological Chemistry*, 2009, **390**, 1087–96.

[27] J. Stock, *Annual Review of Biophysics and Biomolecular Structure*, 1991, **20**, 109–36.

[28] K. Huang, Y. Xu, C. Packianathan, F. Gao, C. Chen, J. Zhang, Q. Shen, B. P. Rosen and F. J. Zhao, *Molecular Microbiology*, 2018, **107**, 265–276.

[29] O. Regnell and C. J. Watras, *Environmental Science and Technology*, 2019, **53**, 4–19.

[30] D. K. Liscombe, G. V. Louie and J. P. Noel, *Natural Product Reports*, 2012, **29**, 1238–50.

[31] C. C. Huang, C. V. Smith, M. S. Glickman, W. R. Jacobs and J. C. Sacchettini, *Journal of Biological Chemistry*, 2002, **277**, 11559–69.

[32] E. Raux, H. L. Schubert, S. C. Woodcock, K. S. Wilson and M. J. Warren, *European Journal of Biochemistry*, 1998, **254**, 341–46.

[33] T. Shoji and T. Hashimoto, *Plant Biotechnology*, 2013, **30**, 217–24.

[34] T. Koeduka, H. Suzuki, G. Taguchi and K. Matsui, *Plant Biotechnology*, 2020, **37**, 389–92.

[35] H. L. Schubert, R. M. Blumenthal and X. Cheng, *Trends in Biochemical Sciences*, 2003, **28**, 329–35.

[36] K. Brzezinski, *Biomolecules*, 2020, **10**, 1682.

[37] D. Pei and J. Zhu, *Current Opinion in Chemical Biology*, 2004, **8**, 492–497.

[38] M. M. Dixon, S. Huang, R. G. Matthews and M. Ludwig, *Structure*, 1996, **4**, 1263–75.

[39] M. A. Pajares and D. Pérez-Sala, *Cellular and Molecular Life Sciences*, 2006, **63**, 2792–803.

[40] M. S. McQueney, K. S. Anderson and G. D. Markham, *Biochemistry*, 2000, **39**, 4443–54.

[41] J. Komoto, T. Yamada, Y. Takata, G. D. Markham and F. Takusagawa, *Biochemistry*, 2004, **43**, 1821–31.

[42] G. D. Markham, F. Takusagawa, A. M. DiJulio and C. W. Bock, *Archives of Biochemistry and Biophysics*, 2009, **492**, 82–92.

[43] M. P. Costi, S. Ferrari, A. Venturelli, S. Calo, D. Tondi and D. Barlocco, *Current Medicinal Chemistry*, 2005, **12**, 2241–58.

[44] H. Nishimasu, R. Ishitani, K. Yamashita, C. Iwashita, A. Hirata, H. Hori and O. Nureki, *Proceedings of the National Academy of Sciences of the United States of America*, 2009, **106**, 8180–5.

[45] H.-H. Wu, M. D. Pun, C. E. Wise, B. R. Streit, F. Mus, A. Berim, W. M. Kincannon, A. Islam, S. E. Partovi, D. R. Gang, J. L. DuBois, C. E. Lubner, C. E. Berkman, B. M. Lange and J. W. Peters, *Proceedings of the National Academy of Sciences of the United States of America*, 2022, **119**, e2207190119.

[46] P. Z. Kozbial and A. R. Mushegian, *BMC Structural Biology*, 2005, **5**, 19.

[47] X. Cheng, S. Kumar, J. Posfai, J. W. Pflugrath and R. J. Roberts, *Cell*, 1993, **74**, 299–307.

[48] J. Vidgren, L. A. Svensson and A. Liljas, *Nature*, 1994, **368**, 354–7.

[49] J. R. Horton, K. Sawada, M. Nishibori, X. Zhang and X. Cheng, *Structure*, 2001, **9**, 837–49.

[50] H. G. Botros, P. Legrand, C. Pagan, V. Bondet, P. Weber, M. Ben-Abdallah, N. Lemière, G. Huguet, J. Bellalou, E. Maronde, P. Beguin, A. Haouz, W. Shepard and T. Bourgeron, *Journal of Pineal Research*, 2013, **54**, 46–57.

[51] J. P. Keller, P. M. Smith, J. Benach, D. Christendat, G. T. D. Titta and J. F. Hunt, *Structure*, 2002, **10**, 1475–87.

[52] S. Kishishita, K. Shimizu, K. Murayama, T. Terada, M. Shirouzu, S. Yokoyama and N. Kunishima, *Acta Crystallographica Section D: Biological Crystallography*, 2008, **64**, 397–406.

[53] V. Cura, N. Marechal, N. Troffer-Charlier, J. M. Strub, M. J. van Haren, N. I. Martin, S. Cianférani, L. Bonnefond and J. Cavarelli, *FEBS Journal*, 2017, **284**, 77–96.

[54] J. T. Jarrett, S. Huang and R. G. Matthews, *Biochemistry*, 1998, **37**, 5372–82.

[55] H. L. Schubert, K. S. Wilson, E. Raux, S. C. Woodcock and M. J. Warren, *Nature Structural Biology*, 1998, **5**, 585–92.

[56] J. Vévodová, R. M. Graham, E. Raux, H. L. Schubert, D. I. Roper, A. A. Brindley, A. I. Scott, C. A. Roessner, N. P. J. Stamford, M. E. Stroupe, E. D. Getzoff, M. J. Warren and K. S. Wilson, *Journal of Molecular Biology*, 2004, **344**, 419–33.

[57] M. E. Stroupe, H. K. Leech, D. S. Daniels, M. J. Warren and E. D. Getzoff, *Nature Structural Biology*, 2003, **10**, 1064–73.

[58] A. Krishnamohan and J. E. Jackman, *Biochemistry*, 2019, **58**, 336–45.

[59] K. Lim, H. Zhang, A. Tempczyk, W. Krajewski, N. Bonander, J. Toedt, A. Howard, E. Eisenstein and O. Herzberg, *Proteins: Structure, Function and Genetics*, 2003, **51**, 56–67.

[60] R. J. Liu, M. Zhou, Z. P. Fang, M. Wang, X. L. Zhou and E. D. Wang, *Nucleic Acids Research*, 2013, **41**, 7828–42.

[61] W. E. Swinehart and J. E. Jackman, *RNA Biology*, 2015, **12**, 398–411.

[62] Z. Shao, W. Yan, J. Peng, X. Zuo, Y. Zou, F. Li, D. Gong, R. Ma, J. Wu, Y. Shi, Z. Zhang, M. Teng, X. Li and Q. Gong, *Nucleic Acids Research*, 2014, **4**, 509–25.

[63] C. Qian and M. M. Zhou, *Cellular and Molecular Life Sciences*, 2006, **63**, 2755–63.

[64] B. Xiao, C. Jing, J. R. Wilson, P. A. Walker, N. Vasisht, G. Kelly, S. Howell, I. A. Taylor, M. G. Blackburn and S. J. Gamblin, *Nature*, 2003, **421**, 652–6.

[65] J. A. Q. A. Faria, N. C. R. Corrêa, C. de Andrade, A. C. de Angelis Campos, R. dos Santos Samuel de Almeida, T. S. Rodrigues, A. M. de Goes, D. A. Gomes and F. P. Silva, *Journal of Cancer Science and Therapy*, 2013, **5**, 58–65.

[66] C. Nwasike, S. Ewert, S. Jovanovic, S. Haider and S. Mujtaba, *Annals of the New York Academy of Sciences*, 2016, **1376**, 18–28.

[67] K. Tran and E. M. Green, *Current Genetics*, 2019, **65**, 643–8.

[68] R. M. Woodhouse and A. Ashe, *Epigenetics Insights*, 2019, **12**, 1–4.

[69] O. Dobrovolska, M. Brilkov, N. Madeleine, Øyvind Ødegård Fougner, Øyvind Strømland, S. R. Martin, V. D. Marco, E. Christodoulou, K. Teigen, J. Isaksson, J. Underhaug, N. Reuter, R. B. Aalen, R. Aasland and Øyvind Halskau, *FEBS Journal*, 2020, 15256.

[70] R. C. Trievel, B. M. Beach, L. M. Dirk, R. L. Houtz and J. H. Hurley, *Cell*, 2002, **111**, 91–103.

[71] T. Petrossian and S. Clarke, *Epigenomics*, 2009, **1**, 163–75.

[72] D. G. Fujimori, *Current Opinion in Chemical Biology*, 2013, **17**, 597–604.

[73] T. L. Grove, M. I. Radle, C. Krebs and S. J. Booker, *Journal of the American Chemical Society*, 2011, **133**, 19586–6.

[74] A. J. Blaszczyk, R. X. Wang and S. J. Booker, *Methods in Enzymology*, 2017, **595**, 303–29.

[75] J. D. Romano and S. Michaelis, *Molecular Biology of the Cell*, 2001, **12**, 1957–71.

[76] M. M. Diver, L. Pedi, A. Koide, S. Koide and S. B. Long, *Nature*, 2018, **553**, 526–9.

[77] J. Yang, K. Kulkarni, I. Manolaridis, Z. Zhang, R. B. Dodd, C. Mas-Droux and D. Barford, *Molecular Cell*, 2011, **44**, 997–1004.

[78] S. Kimura, K. Miyauchi, Y. Ikeuchi, P. C. Thiaville, V. D. Crécy-Lagard and T. Suzuki, *Nucleic Acids Research*, 2014, **42**, 9350–65.

[79] F. Forouhar, A. Kuzin, J. Seetharaman, I. Lee, W. Zhou, M. Abashidze, Y. Chen, W. Yong, H. Janjua, Y. Fang, D. Wang, K. Cunningham, R. Xiao, T. B. Acton, E. Pichersky, D. F. Klessig, C. W. Porter, G. T. Montelione and L. Tong, *Journal of Structural and Functional Genomics*, 2007, **8**, 37–44.

[80] T. Wlodarski, J. Kutner, J. Towpik, L. Knizewski, L. Rychlewski, A. Kudlicki, M. Rowicka, A. Dziembowski and K. Ginalski, *PLoS ONE*, 2011, **6**, e23168.

[81] M. A. Currie, G. Brown, A. Wong, T. Ohira, K. Sugiyama, T. Suzuki, A. F. Yakunin and Z. Jia, *RNA*, 2017, **23**, 346–54.

[82] S. Balaji and L. Aravind, *Nucleic Acids Research*, 2007, **35**, 5658–71.

[83] V. Anantharaman and L. Aravind, *Proteins: Structure, Function and Genetics*, 2004, **56**, 795–807.

[84] H. G. Choudhury, A. D. Cameron, S. Iwata and K. Beis, *Biochemical Journal*, 2011, **435**, 85–91.

[85] K. D. Allen and S. C. Wang, *Biochimica et Biophysica Acta - Proteins and Proteomics*, 2014, **1844**, 2135–44.

[86] N. Itoh, H. Toda, M. Matsuda, T. Negishi, T. Taniguchi and N. Ohsawa, *BMC Plant Biology*, 2009, **9**, 116.

[87] S. M. Lyi, L. I. Heller, M. Rutzke, R. M. Welch, L. V. Kochian and L. Li, *Plant Physiology*, 2005, **138**, 409–20.

[88] Y. Byeon, H. J. Lee, H. Y. Lee and K. Back, *Journal of Pineal Research*, 2016, **60**, 65–73.

[89] X. Liu, J. Wu, Y. Sun and W. Xie, *ACS Chemical Biology*, 2017, **12**, 2164–2171.

[90] Y. Peng, Q. Feng, D. Wilk, A. A. Adjei, O. E. Salavaggione, R. M. Weinshilboum and V. C. Yee, *Biochemistry*, 2008, **47**, 6216–25.

[91] X. W. Zou, Y. C. Liu, N. S. Hsu, C. J. Huang, S. Y. Lyu, H. C. Chan, C. Y. Chang, H. W. Yeh, K. H. Lin, C. J. Wu, M. D. Tsai and T. L. Li, *Acta Crystallographica Section D: Biological Crystallography*, 2014, **D70**, 1549–60.

[92] D. A. Amariei, N. Pozhydaieva, B. David, P. Schneider, T. Classen, H. Gohlke, O. H. Weiergräber and J. Pietruszka, *ACS Catalysis*, 2022, **12**, 14130–14139.

[93] S. Wu, R. Snajdrova, J. C. Moore, K. Baldenius and U. T. Bornscheuer, *Angewandte Chemie - International Edition*, 2020, **59**, 2–34.

[94] F. Garzón-Posse, L. Becerra-Figueroa, J. Hernández-Arias and D. Gamba-Sánchez, *Molecules*, 2018, **23**, 1265.

[95] B. de Lange, D. J. Hyett, P. J. D. Maas, D. Mink, F. B. J. van Assema, N. Sereinig, A. H. M. de Vries and J. G. de Vries, *ChemCatChem*, 2011, **3**, 289–292.

[96] H. Yamada and M. Kobayashi, *Bioscience, Biotechnology and Biochemistry*, 1996, **60**, 1391–400.

[97] M. S. Islam, A. Aryasomayajula and P. R. Selvaganapathy, *Micromachines*, 2017, **8**, 83.

[98] J. W. Shen, J. M. Qi, X. J. Zhang, Z. Q. Liu and Y. G. Zheng, *Organic Process Research and Development*, 2019, **23**, 1017–25.

[99] S. A. B. Dockrey, T. J. Doyon, J. C. Perkins and A. R. Narayan, *Chemical Biology and Drug Design*, 2019, **93**, 1207–13.

[100] J. Wachtmeister and D. Rother, *Current Opinion in Biotechnology*, 2016, **42**, 169–77.

[101] J. U. Bowie, S. Sherkhanov, T. P. Korman, M. A. Valliere, P. H. Opgenorth and H. Liu, *Trends in Biotechnology*, 2020, **38**, 766–78.

[102] A. M. Kunjapur, J. C. Hyun and K. L. Prather, *Microbial Cell Factories*, 2016, **15**, 1–17.

[103] C. E. Nakamura and G. M. Whited, *Current Opinion in Biotechnology*, 2003, **14**, 454–9.

[104] W. C. Deloache, Z. N. Russ, L. Narcross, A. M. Gonzales, V. J. Martin and J. E. Dueber, *Nature Chemical Biology*, 2015, **11**, 465–471.

[105] F. C. Cui, X. L. Pan, W. Liu and J. Y. Liu, *Journal of Computational Chemistry*, 2011, **32**, 3068–74.

[106] D. S. Goodsell, M. F. Sanner, A. J. Olson and S. Forli, *Protein Science*, 2020, 1–13.

[107] J. L. Porter, R. A. Rusli and D. L. Ollis, *ChemBioChem*, 2016, **17**, 197–203.

[108] X. M. Gong, Z. Qin, F. L. Li, B. B. Zeng, G. W. Zheng and J. H. Xu, *ACS Catalysis*, 2019, **9**, 147–53.

[109] S. Lee, S. Y. Shin, Y. Lee, Y. Park, B. G. Kim, J. H. Ahn, Y. Chong, Y. H. Lee and Y. Lim, *Bioorganic and Medicinal Chemistry Letters*, 2011, **21**, 3866–70.

[110] G. F. Sánchez-Pérez, J. M. Bautista and M. A. Pajares, *Journal of Molecular Biology*, 2004, **335**, 693–706.

[111] F. Wang, S. Singh, J. Zhang, T. D. Huber, K. E. Helmich, M. Sunkara, K. A. Hurley, R. D. Goff, C. A. Bingman, A. J. Morris, J. S. Thorson and G. N. Phillips, *FEBS Journal*, 2014, **281**, 4224–39.

[112] M. Porcelli, G. Cacciapuoti, M. Cartenì-Farina and A. Gambacorta, *European Journal of Biochemistry*, 1988, **177**, 273–80.

[113] Z. J. Lu and G. D. Markham, *Journal of Biological Chemistry*, 2002, **277**, 16624–31.

[114] F. Garrido, C. Alfonso, J. C. Taylor, G. D. Markham and M. A. Pajares, *Biochimica et Biophysica Acta - Proteins and Proteomics*, 2009, **1794**, 1082–90.

[115] J. Schlesier, J. Siegrist, S. Gerhardt, A. Erb, S. Blaesi, M. Richter, O. Einsle and J. N. Andexer, *BMC Structural Biology*, 2013, **13**, 22.

[116] M. Porcelli, C. P. Ilisso, L. Mosca and G. Cacciapuoti, *Bioengineered*, 2015, **6**, 184–6.

[117] S. C. Lu and J. M. Mato, *Physiological Reviews*, 2012, **92**, 1515–42.

[118] C. L. Quinlan, S. E. Kaiser, B. Bolaños, D. Nowlin, R. Grantner, S. Karlicek-Bryant, J. L. Feng, S. Jenkinson, K. Freeman-Cook, S. G. Dann, X. Wang, P. A. Wells, V. R. Fantin, A. E. Stewart and S. K. Grant, *Nature Chemical Biology*, 2017, **13**, 785–92.

[119] P. Kalev, M. L. Hyer, S. Gross, Z. Konteatis, C. C. Chen, M. Fletcher, M. Lein, E. Aguado-Fraile, V. Frank, A. Barnett, E. Mandley, J. Goldford, Y. Chen, K. Sellers, S. Hayes, K. Lizotte, P. Quang, Y. Tuncay, M. Clasquin, R. Peters, J. Weier, E. Simone, J. Murtie, W. Liu, R. Nagaraja, L. Dang, Z. Sui, S. A. Biller, J. Travins, K. M. Marks and K. Marjon, *Cancer Cell*, 2021, **39**, 209–24.

[120] B. Murray, S. V. Antonyuk, A. Marina, S. C. Lu, J. M. Mato, S. S. Hasnain and A. L. Rojas, *Proceedings of the National Academy of Sciences of the United States of America*, 2016, **113**, 2104–2109.

[121] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp and G. J. Barton, *Bioinformatics*, 2009, **25**, 1189–1191.

[122] A. Peters, E. Herrmann, N. V. Cornelissen, N. Klöcker, D. Kümmel and A. Rentmeister, *ChemBioChem*, 2022, e2021004.

[123] S. Mordhorst, J. Siegrist, M. Müller, M. Richter and J. N. Andexer, *Angewandte Chemie - International Edition*, 2017, **56**, 4037–41.

[124] M. Dippe, W. Brandt, H. Rost, A. Porzel, J. Schmidt and L. A. Wessjohann, *Chemical Communications*, 2015, **51**, 3637–40.

[125] S. Singh, J. Zhang, T. D. Huber, M. Sunkara, K. Hurley, R. D. Goff, G. Wang, W. Zhang, C. Liu, J. Rohr, S. G. V. Lanen, A. J. Morris and J. S. Thorson, *Angewandte Chemie - International Edition*, 2014, **53**, 3965–9.

[126] B. J. Law, M. R. Bennett, M. L. Thompson, C. Levy, S. A. Shepherd, D. Leys and J. Micklefield, *Angewandte Chemie - International Edition*, 2016, **55**, 2683–7.

[127] R. Wang, W. Zheng and M. Luo, *Analytical Biochemistry*, 2014, **450**, 11–9.

[128] R. Wang, K. Islam, Y. Liu, W. Zheng, H. Tang, N. Lailler, G. Blum, H. Deng and M. Luo, *Journal of the American Chemical Society*, 2013, **135**, 1048–56.

[129] O. M. Ottink, F. H. Nelissen, Y. Derks, S. S. Wijmenga and H. A. Heus, *Analytical Biochemistry*, 2010, **396**, 280–3.

[130] M. Erguven, N. V. Cornelissen, A. Peters, E. Karaca and A. Rentmeister, *ChemBioChem*, 2022, e202200511.

[131] T. D. Huber, J. A. Clinger, Y. Liu, W. Xu, M. D. Miller, G. N. Phillips and J. S. Thorson, *ACS Chemical Biology*, 2020, **15**, 695–705.

[132] A. J. Herbert, S. A. Shepherd, V. A. Cronin, M. R. Bennett, R. Sung and J. Micklefield, *Angewandte Chemie - International Edition*, 2020, **59**, 14950–6.

[133] A. S. Eustáquio, F. Pojer, J. P. Noel and B. S. Moore, *Nature Chemical Biology*, 2008, **4**, 69–74.

[134] J. M. Lipson, M. Thomsen, B. S. Moore, R. P. Clausen, J. J. L. Clair and M. D. Burkart, *ChemBioChem*, 2013, **14**, 950–3.

[135] M. Thomsen, S. B. Vogensen, J. Buchardt, M. D. Burkart and R. P. Clausen, *Organic and Biomolecular Chemistry*, 2013, **11**, 7606–10.

[136] T. D. Davis, S. Kunakom, M. D. Burkart and A. S. Eustaquio, *Methods in Enzymology*, 2018, **604**, 367–88.

[137] J. K. Coward and E. P. Slisz, *Journal of Medicinal Chemistry*, 1973, **16**, 460–3.

[138] F. della Ragione, M. Porcelli, M. Carteni-Farina, V. Zappia and A. E. Pegg, *Biochemical Journal*, 1985, **232**, 335–41.

[139] B. Allart, M. Gatel, D. Guillerm and G. Guillerm, *European Journal of Biochemistry*, 1998, **256**, 155–62.

[140] J. E. Lee, K. A. Cornell, M. K. Riscoe and P. L. Howell, *Structure*, 2001, **9**, 941–53.

[141] S. Wang, J. Lim, K. Thomas, F. Yan, R. H. Angeletti and V. L. Schramm, *Journal of the American Chemical Society*, 2012, **134**, 1468–70.

[142] K. Thomas, S. A. Cameron, S. C. Almo, E. S. Burgos, S. A. Gulab and V. L. Schramm, *Biochemistry*, 2015, **54**, 2520–9.

[143] H. Umihara, S. Yokoshima, M. Inoue and T. Fukuyama, *Chemistry - A European Journal*, 2017, **23**, 6993–6995.

[144] L. Chang, J. M. Hagel and P. J. Facchini, *Plant Physiology*, 2015, **169**, 1127–1140.

[145] K. Wang, X. Feng, L. Chai, S. Cao and F. Qiu, *Drug Metabolism Reviews*, 2017, **49**, 139–57.

[146] T. Xu, T. Kuang, H. Du, Q. Li, T. Feng, Y. Zhang and G. Fan, *Pharmacological Research*, 2020, **152**, 104632.

[147] M. A. Altinoz, G. Topcu, A. Hacimuftuoglu, A. Ozpinar, A. Ozpinar, E. Hacker and İlhan Elmaci, *Neurochemical Research*, 2019, **44**, 1796–806.

[148] L. Braconi, G. Bartolucci, M. Contino, N. Chiaramonte, R. Giampietro, D. Manetti, M. G. Perrone, M. N. Romanelli, N. A. Colabufo, C. Riganti, S. Dei and E. Teodori, *Journal of Enzyme Inhibition and Medicinal Chemistry*, 2020, **35**, 974–992.

[149] M. Białoń, M. Żarnowska, L. Antkiewicz-Michaluk and A. Wąsik, *Psychopharmacology*, 2020, **237**, 1577–1593.

[150] B. R. Lichman, A. Sula, T. Pesnot, H. C. Hailes, J. M. Ward and N. H. Keep, *Biochemistry*, 2017, **56**, 5274–7.

[151] B. R. Lichman, E. D. Lamming, T. Pesnot, J. M. Smith, H. C. Hailes and J. M. Ward, *Green Chemistry*, 2015, **17**, 852–5.

[152] B. R. Lichman, J. Zhao, H. C. Hailes and J. M. Ward, *Nature Communications*, 2017, **8**, 14883.

[153] J. Zhao, B. R. Lichman, J. M. Ward and H. C. Hailes, *Chemical Communications*, 2018, **54**, 1323–6.

[154] R. Roddan, G. Gygli, A. Sula, D. Méndez-Sánchez, J. Pleiss, J. M. Ward, N. H. Keep and H. C. Hailes, *ACS Catalysis*, 2019, **9**, 9640–9.

[155] Y. Wang, F. Subrizi, E. M. Carter, T. D. Sheppard, J. M. Ward and H. C. Hailes, *Nature Communications*, 2022, **13**, 5435.

[156] F. Subrizi, Y. Wang, B. Thair, D. Méndez-Sánchez, R. Roddan, M. Cárdenas-Fernández, J. Siegrist, M. Richter, J. N. Andexer, J. M. Ward and H. C. Hailes, *Angewandte Chemie*, 2021, **60**, 18673–9.

[157] M. Nishihachijo, Y. Hirai, S. Kawano, A. Nishiyama, H. Minami, T. Katayama, Y. Yasohara, F. Sato and H. Kumagai, *Bioscience, Biotechnology and Biochemistry*, 2014, **78**, 701–7.

[158] P. Valverde, A. Ardá, N. C. Reichardt, J. Jiménez-Barbero and A. Gimeno, *MedChemComm*, 2019, **10**, 1678–91.

[159] J. H. Patterson, M. J. McConville, R. E. Haites, R. L. Coppel and H. Billman-Jacobe, *Journal of Biological Chemistry*, 2000, **275**, 24900–6.

[160] A. J. Kreuzman, J. R. Turner and W. K. Yeh, *Journal of Biological Chemistry*, 1988, **263**, 15626–33.

[161] E. Kim, M. C. Song, M. S. Kim, J. Y. Beom, E. Y. Lee, D. M. Kim, S. J. Nam and Y. J. Yoon, *Journal of Natural Products*, 2016, **79**, 2014–21.

[162] A. Shafiee, H. Motamedi and T. Chen, *European Journal of Biochemistry*, 1994, **225**, 755–764.

[163] D. Chen, L. Zhang, B. Pang, J. Chen, Z. Xu, I. Abe and W. Liu, *Journal of Bacteriology*, 2013, **195**, 1931–9.

[164] D. Levac, J. Murata, W. S. Kim and V. D. Luca, *Plant Journal*, 2008, **53**, 225–236.

[165] J. Sun, L. Zhao, Z. Shao, J. Shanks and C. A. Peebles, *Biotechnology and Bioengineering*, 2018, **115**, 673–83.

[166] M. Ben-Abdallah, V. Bondet, F. Fauchereau, P. Béguin, H. Goubran-Botros, C. Pagan, T. Bourgeron and J. Bellalou, *Protein Expression and Purification*, 2011, **75**, 114–8.

[167] J. Panich, B. Fong and S. W. Singer, *Trends in Biotechnology*, 2021, **39**, 412–24.

[168] S. Milker and D. Holtmann, *Microbial Cell Factories*, 2021, **20**, 89.

[169] S. Milker, A. Sydow, I. Torres-Monroy, G. Jach, F. Faust, L. Kranz, L. Tkatschuk and D. Holtmann, *Biotechnology and Bioengineering*, 2021, **118**, 2694–702.

[170] W. T. Booth, C. R. Schlachter, S. Pote, N. Ussin, N. J. Mank, V. Klapper, L. R. Offermann, C. Tang, B. K. Hurlburt and M. Chruszcz, *ACS Omega*, 2018, **3**, 760–8.

[171] A. Ferenc-Mrozek, E. Bojarska, J. Stepinski, E. Darzynkiewicz and M. Lukaszewicz, *ACS Omega*, 2020, **5**, 10759–66.

[172] J. M. D. Almeida, V. R. Moure, M. Müller-Santos, E. M. D. Souza, F. O. Pedrosa, D. A. Mitchell and N. Krieger, *Scientific Reports*, 2018, **8**, 10000.

[173] P. D. Parshin, A. A. Pometun, U. A. Martysuk, S. Y. Kleymenov, D. L. Atroshenko, E. V. Pometun, S. S. Savin and V. I. Tishkov, *Biochemistry (Moscow)*, 2020, **85**, 575–82.

[174] K. Robards, P. Haddad and P. Jackson, *High-performance Liquid Chromatography—Instrumentation and Techniques*, Academic Press, 2004, pp. 227–303.

[175] B. J. Law, A. W. Struck, M. R. Bennett, B. Wilkinson and J. Micklefield, *Chemical Science*, 2015, **6**, 2885.

[176] D. S. Goodsell, G. M. Morris and A. J. Olson, *Journal of Molecular Recognition*, 1996, **9**, 1–5.

[177] J. Fuhrmann, A. Rurainski, H. P. Lenhof and D. Neumann, *Journal of Computational Chemistry*, 2010, **31**, 1911–18.

[178] N. S. Pagadala, K. Syed and J. Tuszynski, *Biophysical Reviews*, 2017, **9**, 91–102.

[179] T. Gaillard, *Journal of Chemical Information and Modeling*, 2018, **58**, 1697–706.

[180] M. C. Ahmed, E. Papaleo and K. Lindorff-Larsen, *PeerJ*, 2018, **6**, e4967.

[181] I. R. Bothwell, K. Islam, Y. Chen, W. Zheng, G. Blum, H. Deng and M. Luo, *Journal of the American Chemical Society*, 2012, **134**, 14905–12.

[182] J. Jiracek, M. Collinsova, I. Rosenberg, M. Budesinsky, E. Protivinska, H. Netusilova and T. A. Garrow, *Journal of Medicinal Chemistry*, 2006, **49**, 3982–9.

[183] B. Bhushan, Y. A. Lin, M. Bak, A. Phanumartwiwath, N. Yang, M. K. Bilyard, T. Tanaka, K. L. Hudson, L. Lercher, M. Stegmann, S. Mohammed and B. G. Davis, *Journal of the American Chemical Society*, 2018, **140**, 14599–603.

[184] G. Z. Melgar, E. P. Wendler, A. A. D. Santos and A. L. Porto, *Tetrahedron Asymmetry*, 2010, **21**, 2271–4.

[185] M. G. Finn and K. B. Sharpless, *Angewandte Chemie International Edition*, 2001, **40**, 2004–21.

[186] H. C. Kolb and K. B. Sharpless, *Drug Discovery Today*, 2003, **8**, 1128–37.

[187] S. Neumann, M. Biewend, S. Rana and W. H. Binder, *Macromolecular Rapid Communications*, 2020, **41**, e1900359.

[188] N. J. Agard, J. A. Prescher and C. R. Bertozzi, *Journal of the American Chemical Society*, 2004, **126**, 15046–7.

[189] X. Jiang, X. Hao, L. Jing, G. Wu, D. Kang, X. Liu and P. Zhan, *Expert Opinion on Drug Discovery*, 2019, **14**, 779–89.

[190] J. L. Díaz, U. Christmann, A. Fernández, A. Torrens, A. Port, R. Pascual, I. Álvarez, J. Burgueño, X. Monroy, A. Montero, A. Balada, J. M. Vela and C. Almansa, *Journal of Medicinal Chemistry*, 2015, **58**, 2441–51.

[191] N. M. Grob, D. Häussinger, X. Deupi, R. Schibli, M. Behe and T. L. Mindt, *Journal of Medicinal Chemistry*, 2020, **63**, 4484–95.

[192] J. E. Doiron, C. A. Le, B. K. Ody, J. B. Brace, S. J. Post, N. L. Thacker, H. M. Hill, G. W. Breton, M. J. Mulder, S. Chang, T. M. Bridges, L. Tang, W. Wang, S. M. Rowe, S. G. Aller and M. Turlington, *Chemistry - A European Journal*, 2019, **15**, 1720–30.

[193] S. K. Maurya and R. Rana, *Beilstein Journal of Organic Chemistry*, 2017, **13**, 1106–18.

[194] Y. Takayama, K. Kusamori and M. Nishikawa, *Molecules*, 2019, **24**, 1–20.

[195] D. Honcharenko, K. Druceikaite, M. Honcharenko, M. Bollmark, U. Tedebark and R. Strömberg, *ACS Omega*, 2021, **6**, 579–93.

[196] G. Pander, P. Uhl, N. Kühl, U. Haberkorn, J. Anderl and W. Mier, *Drug Discovery Today*, 2022, **27**, 103311.

[197] E. C. Vatansever, J. Kang, A. Tuley, E. S. Ward and W. R. Liu, *Bioorganic and Medicinal Chemistry*, 2020, **28**, 115808.

[198] F. Muttach and A. Rentmeister, *Angewandte Chemie - International Edition*, 2016, **55**, 1917–20.

[199] A. W. Struck, M. R. Bennett, S. A. Shepherd, B. J. Law, Y. Zhuo, L. S. Wong and J. Micklefield, *Journal of the American Chemical Society*, 2016, **138**, 3038–45.

[200] D. Kleiner, F. Shmulevich, R. Zarivach, A. Shahar, M. Sharon, G. Ben-Nissan and S. Bershtein, *Journal of Molecular Biology*, 2019, **431**, 4796–816.

[201] J. T. Nelson, J. Lee, J. W. Sims and E. W. Schmidt, *Applied and Environmental Microbiology*, 2007, **73**, 3575–80.

[202] E. Grzesiuk and C. Janion, *Mutation Research/Reviews in Genetic Toxicology*, 1993, **297**, 313–21.

[203] M. Hmani, H. Boukedi, S. B. Khedher, A. Elleuch, S. Tounsi and L. Abdelkefi-Mesrati, *Journal of Economic Entomology*, 2018, **111**, 108–11.

[204] Y. Su, W. Shao, A. Zhang and W. Zhang, *FEMS Yeast Research*, 2021, **21**, 1–12.

[205] E. O. McCullum, B. A. Williams, J. Zhang and J. C. Chaput, *Random mutagenesis by error-prone PCR*, Humana Press, 2010, pp. 103–9.

[206] G. Muteeb and R. Sen, *Random mutagenesis using a mutator strain*, Humana Press, 2010, pp. 411–20.

[207] J. Bijlsma, W. J. de Bruijn, J. A. Hageman, P. Goos, K. P. Velikov and J. P. Vincken, *Scientific Reports*, 2020, **10**, 8288.

[208] J. Cladera, J. L. Rigaud, J. Villaverde and M. Dunach, *European Journal of Biochemistry*, 1997, **243**, 798–804.

[209] R. Fujii, M. Kitaoka and K. Hayashi, *Error-Prone Rolling Circle Amplification Greatly Simplifies Random Mutagenesis*, 2014, pp. 23–29.

[210] Wiley, *Current Protocols in Human Genetics*, 2000, **26**, A.2D.1–A.2D.13.

[211] F. Studier and B. A. Moffatt, *Journal of Molecular Biology*, 1986, **189**, 113–130.

[212] N. C. Institute, *Online SMILES Translator*.

[213] N. Guex and M. C. Peitsch, *Electrophoresis*, 1997, **18**, 2714–2723.

[214] T. Pesnot, M. C. Gershater, J. M. Ward and H. C. Hailes, *Chemical Communications*, 2011, **47**, 3242–4.

[215] Q. L. Hu, J. T. Liu, J. Li, Y. Ge, Z. Song, A. S. Chan and X. F. Xiong, *Organic Letters*, 2021, **23**, 8543–8.

[216] Q. Wei, G. K. Seward, P. A. Hill, B. Patton, I. E. Dimitrov, N. N. Kuzma and I. J. Dmochowski, *Journal of the American Chemical Society*, 2006, **128**, 13274–83.

Chapter 8

Experimental Methods

8.1 General Methods & Materials

Unless stated otherwise, all chemicals were supplied commercially. Centrifugations were either performed in an Eppendorf 5415 R benchtop microfuge or an Avanti JXN-26 centrifuge. Buffers were prepared with the ingredients given and corrected to the desired pH with HCl_{aq} or $NaOH_{aq}$. Plasmid purification kits, gel extraction kits, BL21(DE3) competent cells and T4 ligase and buffer were purchased from New England Biolabs (NEB). Restriction enzymes and buffers were purchased from Promega. Synthetic genes were ordered from Eurofins Scientific and arrived lyophilised either in pEX-A128 or pET-28a(+) plasmids. Melting points were recorded with an Electrothermal IA9000 series Melting Point Apparatus. Infrared spectra were recorded with a Bruker Alpha Platinum-ATR.

8.2 Enzyme cloning, expression and purification

8.2.1 Creation of chemically competent cells

10 mL of lysogeny broth media (LB) was inoculated with a frozen glycerol stock of TOP10[i] or NovaBlue[ii] cells. The culture was incubated overnight at 37℃ with 180 rpm shaking. The following day, 2 mL of this culture was used to innoculate 200 mL LB, which was

[i] Full genotype: F$^-$ *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L(StrR) *end*A1 *nup*G

[ii] Full genotype: *endA1 hsdR17* (r_{K12}^- m_{K12}^+) *supE44 thi-1 recA1 gyrA96 relA1 lac* F'[*proA*$^+$*B*$^+$ *lacI*^{q}*ZΔM15::Tn10*] (Tet^R)

incubated at 37℃ with 180 rpm shaking until the OD_{600} was above 0.4. From here, all containers, buffers and equipment were pre-cooled to 4℃ and maintained at such throughout. The culture was divided into 50 mL centrifuge tubes and cooled on ice for 30 min. The aliquots were then centrifuged for 20 min at 5185 x *g*. The supernatant was removed and 20 mL 100 mM $MgCl_2$ added to each of the four pellets. These were cooled on ice for 5 min, then shaken gently and intermittently to resuspend the pellets. The 4x 20 mL suspensions were combined in pairs to give 2x 40 mL suspensions, which were diluted to 50 mL total with 100 mM $MgCl_2$. Those suspensions were centrifuged for 15 min at 5185 x *g*. The supernatants were again removed, and the two pellets resuspended in 20 mL 100 mM $CaCl_2$ each. The cells were resuspended as above, and combined into a single, 40 mL suspension. This was diluted to 50 mL total with 100 mM $CaCl_2$ and again centrifuged for 15 min at 5185 x *g*. The supernatant was removed, and the pellet was resuspended in 2 mL cryoprotectant solution (85 mM 100 mM $CaCl_2$, 15% v/v glycerol). The suspension was pipetted in 50 μL aliquots into pre-cooled (-80℃) 1.5 mL microcentrifuge tubes. The aliquots were stored at -80℃ until use.

8.2.2 Restriction cloning

Protein sequences were obtained from the UniProtKB database, and reverse-translated into *E. coli*-optimised sequences using the respective tool at bioinformatics.org. Restriction sites for *NdeI* and *XhoI* were inserted with the minimal number of new bases at the 5'- and 3'-ends of the sequence, respectively. Mutations, when required, were made to the sequence at this stage. The resulting synthetic genes were ordered from Eurofins Scientific as inserts in pEX-A128 carrier vectors. The lyophilised DNA was dissolved in diH_2O and transformed into TOP10 competent cells by the procedure given below.

Transformant cultures were used to inoculate 10 mL LB + ampicillin (100 μg/mL), which was incubated at 37℃ with 180 rpm shaking overnight. The following day, 5 mL aliquots of those cultures were pelleted by centrifugation, and the plasmids extracted with a Monarch Plasmid DNA Miniprep Kit (New England Biolabs), according to the supplier's instructions.

The digestion mixes constituted the following: 2 μL 10X Tango buffer; 2 μg purified plasmid; 1 μL XhoI; 1 μL NdeI and diH_2O up to 20 μL final volume. The mixes were

incubated at 37℃ overnight, then heat-deactivated at 70℃ for ten minutes. The reaction mixes were combined with 4 μL 6X loading buffer and separated on a gel (1% (w/v) agarose in TAE buffer (Tris-acetate (40 mM), EDTA (1mM), pH 8.3) for 20-30 min at 160 V. Bands corresponding to either the synthetic gene, or in the case of pET-28a(+), the linearised vector, were located with a blue light source and cut from the gel. The DNA was extracted from the gel using a Monarch DNA Gel Extraction Kit, according to the supplier's instructions, with all optional cleaning and concentrating steps observed.

For the ligations reactions, 0.02 pmol of the linearised vector was combined with 0.06 pmol of fragment, along with 1 μL 10X T4 buffer, 0.5 μL T4 ligase and diH_2O up to 10 μL. The reactions were incubated at 4℃ for 24-72 hr, then heat-deactivated at 70℃ for 10 minutes. NovaBlue competent cells were transformed with the ligation products by the following protocol. An aliquot of cells was thawed on ice for 10 min. The full ligation mixtures were added to the thawed cells, mixed by gentle agitation then incubated on ice for 30 min. The cells were then heat-shocked in a water bath at 42℃ for 45 seconds, before being incubated again on ice for 5 minutes. 200 μL room temperature Super Optimal broth with Catabolite repression (SOC)[210] was added to the cells, mixed by gentle agitation and inversion and incubated at 37℃ with 180 rpm shaking for 1 hr. 100 μL aliquots of this culture were spread on selective (LB agar + kanamycin (50 μg/mL)) and control (LB agar) plates, and incubated at 37℃ overnight.

Transformant colonies were used to inoculate 10 mL LB + kanamycin (50 μg/mL) volumes. The cultures were grown and the plasmids extracted using a Monarch DNA Gel Extraction Kit, according to the supplier's instructions, with all optional cleaning and concentrating steps observed. Samples of the purified plasmids were sent to Eurofins Scientific for sequencing with T7 and T7_term primers. Plasmids which sequencing indicated contained the synthetic gene were subsequently transformed into BL21(DE3)[211] competent cells (New England Biolabs) using the protocol above.

8.2.3 Enzyme expression

10 mL LB + kanamycin (50 μg/mL) volumes were inoculated either directly with transformant colonies, or with glycerol stocks of the strains, and incubated overnight at 37℃ with 180 rpm shaking. The following day, the cultures were used to inoculate 250-500 mL LB

+ kanamycin (50 μg/mL) at 1% *v/v*. For instances where the overnight culture was grown directly from a transformant colony, 0.5 mL of the culture was also combined with 0.5 mL 50% *v/v* sterile glycerol solution and stored at -80 ℃ as a stock.

The inoculated expression cultures were incubated at 37 ℃ with 180 rpm shaking until the OD_{600} was above 0.6. The cultures were left at r.t. for 10 min. Isopropyl-β-D-1 thiogalatopyranoside (IPTG) was added to a final concentration of 0.5 mM and the cultures were incubated at 25 ℃ with 180 rpm shaking overnight.

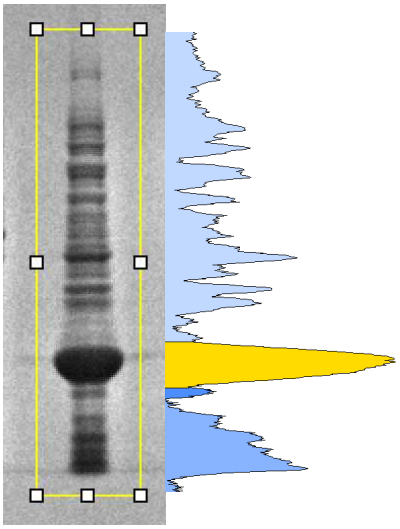
The following day, the cultures were pelleted. For use as lysates, the cultures were pelleted in separate 50 mL volumes, and if not used immediately were stored at -20 ℃. To generate the lysates, the pellets were thawed/cooled on ice and resuspended in the relevant reaction buffer (Section 8.3) by vortexing. The cells in the suspension were kept on ice and disrupted by probe sonication at ~14 W for 5x 10 sec pulses, with 10 sec intervals. This was repeated twice with at least 2 min rest between sonications. The lysates was centrifuged at 6870 x *g* for 30 min to clarify, and the supernatant containing the expressed enzyme decanted into a new container. The total proteins in the lysates were determined with a Quick Start Bradford Assay (Bio-rad) according to the supplier's instructions. Clarified lysates not used immediately were stored at -20 ℃ and thawed only once before being discarded.

For enzyme purification, the cultures were pelleted in 250 mL volumes at 33770 x *g* for 20 min, and the pellets combined. If not used immediately, the pellets were stored at -20 ℃. The pellets were thawed/cooled on ice and resuspended in 35 mL lysis buffer (HEPES 50 mM, NaCl 100 mM, imidazole 20 mM, pH 7.5) by vortexing. The suspensions were kept on ice and disrupted by probe sonication at ~18 W for 10 x 10 sec pulses, with 10 sec intervals. This was repeated twice with at least 2 min rest between sonications. The lysates was centrifuged at 6870 x *g* for 30 min to clarify, and the supernatants containing the expressed enzymes decanted into new containers. The supernatants were passed through 0.45 μM syringe filters, then through HisPur™ Nickel-NTA exchange resins (Thermo Scientific) that had been pre-equilibrated with 5 resin volumes of lysis buffer. Unbound proteins were cleared from the resins with 5 volumes of wash buffer (8% dilution of elution buffer in diH_2O), then eluted as 2.5 mL fractions with elution buffer (HEPES 50 mM, NaCl 100 mM, imidazole 500 mM, pH 7.5) until emerging drops of the

eluate failed to change the colour of Bradford Reagent (Bio-rad). Imidazole was removed from protein-containing fractions by passing through PD-10 desalting resins, which had been pre-equilibrated in reaction buffer. The protein-containing solutions were eluted from the resins with 3.5 mL reaction buffer. The concentration of proteins in the eluates were determined with a Quick Start Bradford Assay (Bio-rad) according to the supplier's instructions, then concentrated as necessary with a PES protein concentrator (10 kDa molecular weight cut off, ThermoFisher Scientific). The final purified protein solution was combined with 20% *v/v* glycerol, divided into 250 μL aliquots and stored at -80℃.

8.2.4 SDS-PAGE Analysis

Protein samples were mixed with an equal volume of 2X concentrate Laemlli buffer and heated to 95℃ for 5 min. Aliquots (5-10 μL) were transferred to the wells of a Tris-Glycine 4-20% gel (Novex), along with 5 μL Broad Range Molecular Weight Markers (Promega), and the gel tank filled with Tris-Glycine SDS running buffer to the appropriate volume. The gel was subjected to 200 V for approximately 50 min, or until the Laemlli buffer stain nearly reached the bottom of the gel. The gel was then submerged in InstantBlue® Coomassie Protein Stain (Abcam) overnight, then washed with diH_2O before being imaged. Estimation of the levels of enzyme in lysates is explained in the following diagram:



34015

15440

858

12931

$$\frac{\mathbf{15440}}{34015 + \mathbf{15440} + 858 + 12931} = 24\%$$

[Total protein] = 7.47 mg/mL

[Enzyme] = 7.47 x 24% = 1.79 mg/mL

Step 1 SDS-PAGE analysis of cell lysate

Step 2 Plot histogram of staining intensity with ImageJ

Step 3 Calculate peak areas for enzyme and other proteins

Step 4 Calculate proportion of enzyme in total protein. Multiply this by the total protein concentration from Bradford assay to give enzyme concentration estimate.

8.2.5 Expression conditions optimisation

10 mL LB + kanamycin (50 μg/mL) volumes were inoculated with glycerol stocks of the strains and incubated overnight at 37℃ with 180 rpm shaking. The following day, 10 mL + kanamycin (50 μg/mL) volumes of the given media were incolulated with 1% (*v/v*) of the overnight cultures. These expression cultures were incubated at 37℃ for 2 h. The cultures were allowed to cool at r.t. for 10 min, before expression was induced by the addition of 0.1, 0.2, 0.5 or 1 mM IPTG. The expression cultures were then incubated at 30, 25, 20 or 16℃ for 24 or 48 h. After this time, the cultures were centrifuged at 6870 x *g* for 3 min and the supernatant removed to leave the pellet. The cells were resuspended in 400 μL lysis buffer and shaken at r.t. for ten min, before being centrifuged at 6870 x *g* for 30 minutes. The clarified lysate supernatant was then extracted and either analysed immediately by SDS-PAGE, or stored at -20℃ for later analysis.

8.3 Enzyme assays

Unless otherwise stated, assays were incubated at 37℃ with 600 rpm orbital shaking for 18 h. Assay volumes were either 200 or 100 μL. Following incubation, proteins were precipitated by addition of 10% (*v/v*) $HCl_{(aq)}$ (1 M) and pelleted by centrifugation. Samples of supernatant were diluted 5-10x with diH_2O prior to analysis. At least three replicates were performed for each reaction. Below are listed the standard assay conditions, which were modified as described in the text.

Table 8.1: **NCS assay condtions**. *Incubated at 30℃

Assay	Buffer	Acetonitrile	Substrate 1	Ascorbate	Substrate 2	NCS	Auxillary
I	HEPES (50 mM, pH 7.5)	10% (*v/v*)	Dopamine (10 mM)	10 mM	Aldehyde 20-40 mM	Purified 0.5 mg/mL	-
VII	HEPES (50 mM, pH 7.5)	10% (*v/v*)	Dopamine (10 mM)	10 mM	Aldehyde 20-40 mM	Cell lysate Desalted, 10% (*v/v*)	-
VIII	HEPES (50 mM, pH 7.5)	10% (*v/v*)	Dopamine (20 mM)	20 mM	Pyruvate (10 mM)	Cell lysate 10% (*v/v*)	*Cv*TAm, lysate 10% (*v/v*)
IX*	HEPES (50 mM, pH 7.5)	10% (*v/v*)	Dopamine (10 mM)	-	Tyrosol (10 mM)	Cell lysate 10% (*v/v*)	HLADH, lysate 10% (*v/v*)

Table 8.2: **Alkylation assay condtions**. MT/MAT buffer: HEPES (50 mM), $MgCl_2$ (200 mM), KCl (20 mM), pH 7.5. *Enzyme concentration estimated from Bradford assay and SDS-PAGE.

Assay	Buffer	Substrate	Cofactor supply	MT	MAT	MTAN
II	MT/MAT buffer	5 mM	Na_2ATP (10 mM) Alkyl donor (10 mM)	Cell lysate 10% (*v/v*)	Cell lysate 10% (*v/v*)	Cell lysate 2% (*v/v*)
III	MT/MAT buffer	1 mM	SAM (6 mM)	Cell lysate 50% (*v/v*)	-	-
IV	Phosphate buffer (100 mM, pH 7.8)	1 mM	SAM (6 mM)	Purified (0.2 mg/mL)	-	-
V	MT/MAT buffer	5 mM	SAM (10 mM)	Purified (0.1 mg/mL)	-	-
VI	MT/MAT buffer	5 mM	SAM (10 mM)	Cell lysate (20% (*v/v*))	-	-
X	MT/MAT buffer	5 mM	Na_2ATP (10-20 mM) Alkyl donor (10-20 mM)	Purified (0.5 mg/mL)	Purified (0.5 mg/mL)	Purified (0.01 mg/mL)
XI	MT/MAT buffer	-	Na_2ATP (10 mM) Alkyl donor (10 mM)	-	Purified (0.5 mg/mL)	-
XII	MT/MAT buffer	5 mM	Na_2ATP (20 mM) Alkyl donor (20 mM)	Cell lysate (0.1-0.5 mg/mL)*	Cell lysate (0.5 mg/mL)*	Cell lysate (0.01 mg/mL)*

8.4 Analysis & Purification

Analytical Reverse-phase high-performance liquid chromatography (RP-HPLC)

Assays were analysed on an Agilent 1260 Infinity II instrument equipped with either an ACE C18-AR column (150 x 4.6 mm, 5 μM pores) (Methods A & B) or an InfinityLab Poroshell 120 EC-C18, (3.0 x 50 mm, 2.5 μM pores) (Methods C-F). Solvent A was H_2O + 0.01% TFA and solvent B was acetonitrile, except for Method B, where solvent A was 5 mM 1-octanesulfonic acid (5 mM), NaH_2PO_4 (10 mM), corrected to pH 3 with phosphoric acid. Flow rate was 1 mL/min and absorbance was recorded at 250 and 283 nm. All absorbance traces shown are from the latter.

Method A

t (min)	B (%)
0	5
2	5
10	70
10.1	95
11	95
11.1	5
13.5	5

Method B

t (min)	B* (%)
0	5
2	5
10	20
12	20
13	5
16	5

Method C

t (min)	B (%)
0	5
1.5	5
8.2	95
9	95
9.1	5
10	5

Method D

t (min)	B (%)
0	5
1.5	5
8	70
8.1	95
9	95
9.1	5
10	5

Method E

t (min)	B (%)
0	5
1.5	5
5	40
5.1	95
6	95
6.1	5
7	5

Method F

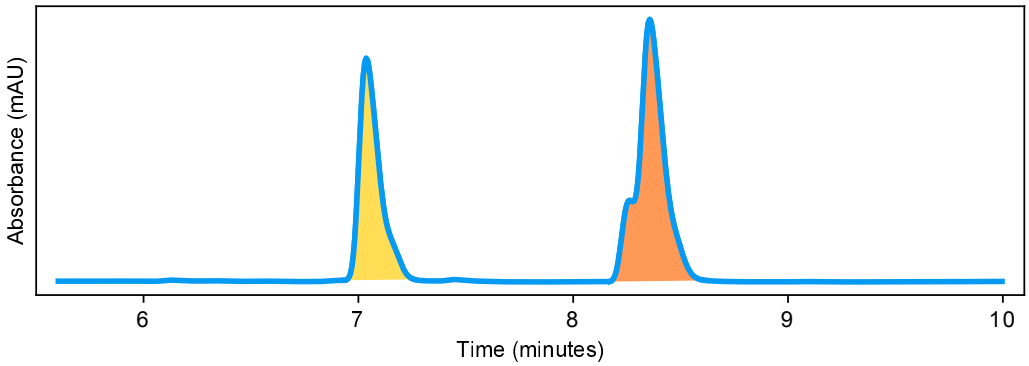
t (min)	B (%)
0	5
1.5	5
2.5	15
6	15
6.1	95
7	95
7.1	5
8	5

* A = 1-octanesulfonic acid (5 mM), NaH_2PO_4 (10 mM), corrected to pH 3 with phosphoric acid.

Conversion was calculated from analytical HPLC traces as the ratio between the peak

areas of the remaining starting material and the product, as shown in the diagram below. This method was predicated on the assumption that the absorbance of the chromophore was not altered between the two species.

HPLC trace from reaction endpoint sample:



Colour	Identity	Area (mAU*s)
	Remaining starting material	959
	Product	1420

$$\text{Conversion} = \frac{\text{Product}}{\text{Product} + \text{Remaining starting material}}$$

$$= \frac{1420}{1420 + 959}$$

$$= \mathbf{60\ \%}$$

Preparative RP-HPLC

Larger scale enzyme cascades were purified on an Agilent 1260 Infinity instrument equipped with a DiscoveryBIO Wide Pore C18 column (10 μM pores, 250 x 21.2 mm). Method: 5% B (3 min), 5-95% B (19 min), 95% B (2 min); where A is H_2O + 0.01% TFA and B is acetonitrile. Flow rate: 8 mL/min. Absorbance was recorded at 250 and 283 nm.

High performance anion exchange chromatography with integrated pulsed amperometric detection (HPAEC-IPAD)

HPAEC-IPAD analysis was performed on an ICS-5000+ Reagent-Free Ion Chromatography System (ICS) (Thermo Scientific). A CarboPac PA1 (4 x 250 mm) anion exchange column was used with a CarboPac PA1 guard column (4 x 50 mm), alongside an elu-

ent generator with a KOH 500 cartridge and a gold electrode electrochemical detector. Method: 1 mL/min flow of isocratic KOH (1 mM) at 30℃ with a 10 μL injection volume.

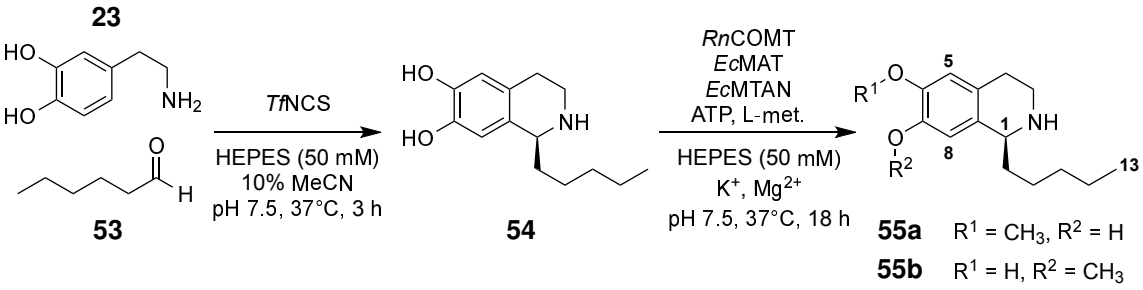
Liquid chromatrography-mass spectrometry (LC-MS)

For LC-MS, assays were prepared as above then analysed on a Walters LCT Premier XE ESI Q-TOF mass spectrometer.

8.5 Docking software and parameters

Coordinates for wild-type *Rn*COMT were downloaded in *.pdb form from the Protein Data Bank entry 6LFE.[48] All other molecules were rendered in ChemDraw Professional (v. 16.0.1.4), converted into SMILES then translated into *.pdb format using the NIH Online SMILES Translator.[212] PyMOL (v. 2.3.0) was used to remove all extraneous molecules including water, solutes and the co-crystallised SAM and nitecapone, but leaving the Mg^{2+} ion in the active site. AutoDock 4.2 was used through the AutoDockTools (ADT, v. 1.5.6) graphic user interface. All molecules were automatically processed into *.pdbqt files by merging non-polar hydrogens and assigning Gasteiger charges. Ligands were allowed the maximum number of torsional degrees of freedom. The grid box was centred on the putative binding site of the molecule being docked, with a 60 x 60 x 60 volume with 0.375 Å spacing between points. AutoGrid was run within ADT to produce the map files. Default settings were used for the docking, with 30 runs of a Lamarckian genetic algorithm and 2.5×10^6 evaluations. The results were subsequently analysed within the ADT interface. Conformations of interest were extracted from the *.dpf file as separate *.pdb files for further analysis in PyMOL. Swiss PDB Viewer[213] was used to make amino acid substitutions to the 6LFE coordinates and perform energy minimisation of the resulting structures.

8.6 Enzyme cascade towards (*S*)-6-methoxy-1-pentyl-1,2,3,4-tetrahydroisoquinolin-7-ol[156] (*S*)-55



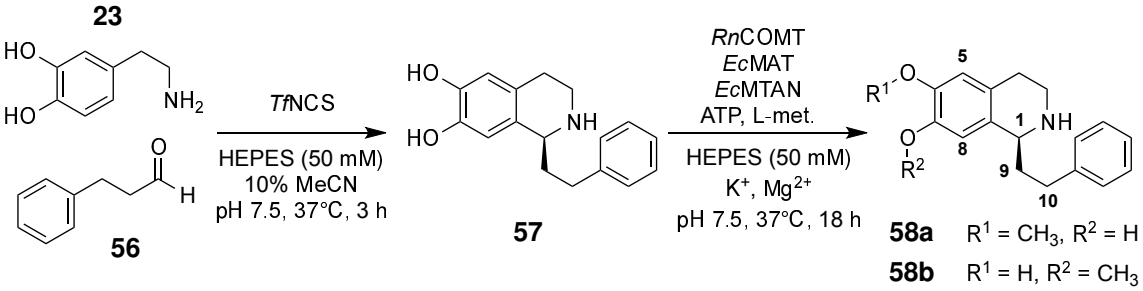
Dopamine **23** hydrochloride (4.75 mg, 0.025 mmol) and sodium ascorbate (4.95 mg, 0.025 mmol, 1 eqv.) were added to buffer (2.5 mL, HEPES 100 mM, pH 7.5) and diH_2O (1.64 mL). Hexanal **53** (12.2 μL, 0.1 mmol, 4 eqv.) in acetonitrile (500 μL) was added, followed by *Tf*NCS (purified as above, to a final concentration of 0.5 mg/mL). The reaction was incubated at 37 ℃ with 200 rpm shaking for 3 h.

After incubation, complete conversion to the product was confirmed by analytical HPLC (Method A). To the crude reaction product was added: 10 X buffer (1 mL, HEPES 500 mM, $MgCl_2$ 200 mM, KCl 2 M, pH 7.5), diH_2O (1.8 mL), *Rn*COMT (1 mL, lysate) *Ec*MAT (1 mL, lysate), *Ec*MTAN (200 μL, lysate), disodium ATP hydrate (28.7 mg, 0.05 mmol, anhydrous basis) and L-methionine (7.5 mg, 0.05 mmol, 2 eqv.). The mixture was incubated at 37 ℃ with 200 rpm shaking for 16 h.

The reaction was quenched with 5% *v/v* 1 M HCl, mixed and centrifuged at 6870 x *g* for 10 min. The supernatant was removed and conversion to the methylated compound confirmed by analytical HPLC (Method A). The crude reaction was then concentrated under vacuum, re-dissolved in 50% aqueous methanol, filtered and purified by preparative HPLC. Fractions containing the product were concentrated under vacuum and re-dissolved in 0.1 M aqueous HCl, before lyophilisation to give the HCl salt of product **(*S*)-55** as a white powder (3.1 mg, 44%). **^{1}H NMR** (700 MHz; D_2O) δ 6.85 (1H, s, 5-H), 6.75 (1H, s, 8-H), 4.40 (1H, dd, *J* = 8.3, 4.8 Hz, 1-H), 3.81 (3H, s, OCH_3), 3.52 (1H, m, 3-*H*H), 3.35-3.29 (1H, m, 3-H*H*), 3.02 (1H, dt, *J* = 17.2, 6.9 Hz, 4-*H*H), 2.95 (1H, dt, *J* = 17.2, 6.2 Hz, 4-H*H*), 2.06-1.97 (1H, m, 9-*H*H), 1.92-1.81 (1H, m, 9-H*H*), 1.46-1.23 (6H, m, 10-H_2, 11-H_2, 12-H_2), 0.83 (3H, t, *J* = 7.6 Hz, 13-H_3); **^{13}C NMR** (700 MHz; D_2O + CD_3OD)

δ 147.7 (C-7), 144.3 (C-6), 125.3 (C-8a), 124.3 (C-4a), 113.6 (C-8), 112.9 (C-5), 57.5 (OCH_3), 55.6 (C-1), 39.9 (C-2), 33.7, (C-9), 31.3, (C-10), 24.8 (C-4), 24.7 (C-11), 22.2 (C-12), 13.7 (C-13); **HRMS** (ESI+) *m/z* Theoretical $[C_{15}H_{25}NO_2]^+$: 250.1802; Measured $[MH]^+$: 250.1802; $[\alpha]_D^{24}$ = +13.2 (H_2O, *c* 0.066); **HPLC** RT (Method A): 8.4 min.

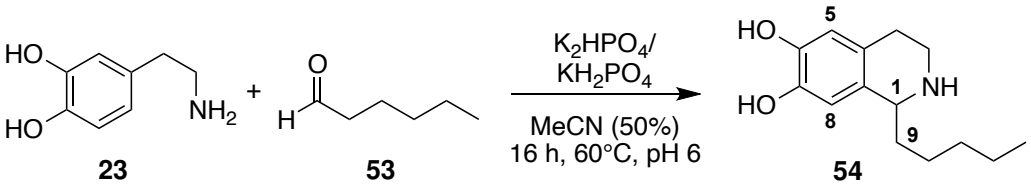
8.7 Enzyme cascade towards (*S*)-6-methoxy-1-phenethyl-1,2,3,4-tetrahydroisoquinolin-7-ol[156] (*S*)-58



Dopamine **23** hydrochloride (4.75 mg, 0.025 mmol) and sodium ascorbate (4.95 mg, 0.025 mmol, 1 eqv.) were added to buffer (2.5 mL, HEPES 100 mM, pH 7.5) and diH_2O (1.64 mL). 3-phenylpropanal **56** (12.2 μL, 0.1 mmol) in acetonitrile (500 μL) was added, followed by *Tf*NCS (purified, to a final concentration of 0.5 mg/mL). The reaction was incubated at 37°C with 200 rpm shaking for 3 h. After incubation, complete conversion to the THIQ was confirmed by analytical HPLC (Method A). To the crude reaction product was added: 10 X buffer (1 mL, HEPES 500 mM, $MgCl_2$ 200 mM, KCl 2 M, pH 7.5), diH_2O (1.8 mL), *Rn*COMT (1 mL, lysate), *Ec*MAT (1 mL, lysate), *Ec*MTAN (200 μL, lysate), disodium ATP hydrate (28.7 mg, 0.05 mmol, anhydrous basis) and L-methionine (7.5 mg, 0.05 mmol, 2 eqv.). The mixture was incubated at 37°C with 200 rpm shaking for 16 h. The reaction was quenched with 5% *v/v* 1 M HCl, mixed and centrifuged at 6870 x *g* for 10 min. The supernatant was removed and conversion to the methylated compound confirmed by analytical HPLC (method A). The crude reaction was then concentrated under vacuum, re-dissolved in 50% aqueous methanol, filtered and purified by preparative HPLC (method above). Fractions containing the product were concentrated under vacuum and re-dissolved in 0.1 M aqueous HCl, before lyophilisation to give the HCl salt of product **(*S*)-58** as a white powder (2.6 mg, 33%). **^{1}H NMR** (700 MHz; D_2O) δ 7.36-7.18

(5H, m, Ph-H) 6.83 (1H, s, 8-H), 6.72-6.66 (1H, m, 5-H), 4.50-4.42 (1H, m,1-H), 3.79 (3H, s, OCH_3), 3.74 (3H, s, OCH_3), 3.57-3.51 (1H, m, 3-*H*H), 3.36-3.31 (1H, m, 3-H*H*), 3.05-2.88 (2H, m, 4-H), 2.68-2.27 (2H, m, 10-H), 2.14-2.34 (2H, m, 9-H). **^{13}C NMR** (700 MHz; D_2O + CD_3OD) δ 147.9 (C-6), 144.3 (C-7), 141.1 (C-11), 129.4 (C-12, C-16), 128.9 (C-13, C-15), 127.1 (C-14), 113.8 (C-5), 112.8 (C-8), 56.5 (OCH_3), 55.3 (C-1), 40.2 (C-3), 35.6 (C-9), 31.3 (C-10), 24.5 (C-4). **HRMS** (ESI+) *m/z* Theoretical $[C_{18}H_{22}NO_2]^+$: 284.1645; Measured $[MH]^+$: 284.1644; **$[\alpha]_D^{24}$** = +6.4 (H_2O, *c* = 0.1); **HPLC** RT (Method A): 9.1 min.

8.8 Synthesis of 1-pentyl-1,2,3,4-tetrahydroisoquinoline-6,7-diol[214] (*RS*)-54

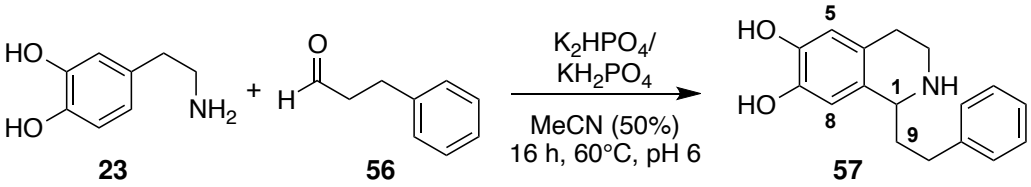


23 **53** **54**

The synthesis was performed according to the procedure reported by Pesnot *et al.*[214] Hexanal **53** (324 μL, 2.64 mmol) was added to a mixture of potassium phosphate buffer (7.5 mL, 0.3 M, pH 6) and acetonitrile (7.5 mL). Dopamine **23** hydrochloride (750 mg, 3.95 mmol) and sodium ascorbate (262 mg, 1.32 mmol) were added. The reaction was heated to 60°C and stirred for 16 h. The pH of the crude product was raised to 7.5 with $NaOH_{(aq)}$ (1 M) and extracted into EtOAc (2x 30 mL). The pH was raised to 8.5 and then 9.5, extracting each time. The combined organic layers (6x 30 mL) were dried over anhydrous Na_2SO_4 then filtered and concentrated under vacuum. The residue was dissolved in a mixture of $HCl_{(aq)}$ (20 mL, 1 M) and dimethyl carbonate (DMC, 20 mL), then the aqueous layer was separated and washed again with DMC (10 mL), before being concentrated under vacuum. The residue was dissolved in diH_2O and lyophilised to give the HCl salt of the product **(*RS*)-54** as an off-white powder (276 mg, 38%). HPLC analysis also indicated the presence of a minor *ortho*-regioisomer. **^{1}H NMR** (700 MHz; CD_3OD) δ 6.65 (1H, s, 8-H), 6.60 (1H, s, 5-H), 4.64-4.31 (1H, m, 1-H), 3.52-3.44 (1H, m, 3-*H*H), 3.33-3.27 (1H, m, 3-H*H*), 3.03-2.86 (2H, m, 4-H), 2.15-1.98 (1H, m, 9-*H*H), 1.98-

1.80 (1H, m, 9-H*H*), 1.61-1.34 (6H, m, 10-H, 11-H, 12-H), 0.97-0.90 (3H, m, 13-H_3); **^{13}C NMR** (700 MHz; CD_3OD) δ 146.5 (C-6), 145.8 (C-7), 124.2 (C-4a), 123.7 (C-8a), 116.0 (C-5), 113.9 (C-8), 56.6 (C-1), 40.9 (C-3), 35.1 (C-9), 32.6 (C-11), 26.0 (C-10), 25.6 (C-4), 23.3 (C-12), 14.3 (C-13). **HRMS** (ESI+) *m/z* Theoretical $[C_{14}H_{22}NO_2]^+$: 236.1645; Measured $[MH]^+$: 236.1642, **HPLC** RT (Method A): 7.8 min.

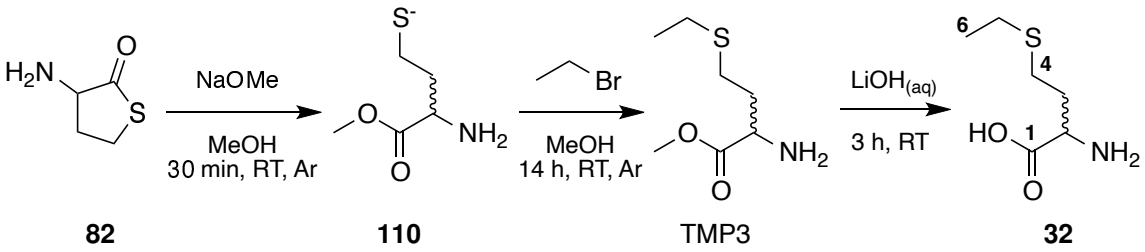
8.9 Synthesis of 1-phenethyl-1,2,3,4-tetrahydroisoquinoline-6,7-diol[214] (*RS*)-57



The synthesis was performed according to the general procedure reported by Pesnot et al.[214] 3-phenylpropanal **56** (351 μL, 2.64 mmol) was added to a mixture of potassium phosphate buffer (7.5 mL, 0.3 M, pH 6) and acetonitrile (7.5 mL). Dopamine **23** hydrochloride (750 mg, 3.95 mmol) and sodium ascorbate (262 mg, 1.32 mmol) were added. The reaction was heated to 60 ℃ and stirred for 16 h. The pH of the crude product was raised to 7.5 with $NaOH_{(aq)}$ (1 M) and extracted into EtOAc (2x 30 mL). The pH was raised to 8.5 and then 9.5, extracting each time. The combined organic layers (6x 30 mL) were dried over anhydrous Na_2SO_4 then filtered and concentrated under vacuum. The residue was dissolved in a mixture of $HCl_{(aq)}$ (20 mL, 1 M) and dimethyl carbonate (DMC, 20 mL). The aqueous layer was separated and washed again with DMC (10 mL), before being concentrated under vacuum. The residue was dissolved in diH_2O and lyophilised to give the HCl salt of the product **(*RS*)-57** as an off-white powder (153 mg, 19%). HPLC analysis indicated the presence of a minor *ortho*-regioisomer.**^{1}H NMR** (700 MHz; CD_3OD) δ 7.73-7.25 (4H, m, 12-H, 13-H, 15-H, 16-H), 7.23-7.16 (1H, m, 14-H), 6.66 (1H, s, 8-H), 6.26 (1H, s, 5-H), 4.71-4.37 (1H, m, 1-H), 3.59-3.51 (1H, m, 3-*H*H), 3.39-3.33 (1H, m, 3-H*H*), 3.05-2.88 (2H, m, 4-H), 2.86-2.76 (2H, m, 9-H), 2.45-2.27 (1H, m, 10-*H*H), 2.24-2.13 (1H, m, 10-H*H*); **^{13}C NMR** (700 MHz; CD_3OD) δ 146.6 (C-7), 146.0 (C-6), 129.6 (C-13, C-15), 129.2 (C-12, C-16), 127.4 (C-14), 123.8 (C-4a), 123.7 (C-8a), 116.2 (C-5), 113.8 (C-8),

56.2 (C-1), 40.8 (C-3), 37.0 (C-10), 32.4 (C-9), 25.5 (C-4); **HRMS** (ESI+) *m/z* Theoretical $[C_{17}H_{20}NO_2]^+$: 270.1489; Measured $[MH]^+$: 270.1489; **HPLC** RT (Method A): 8.5 min.

8.10 Synthesis of DL-ethionine 32[124]

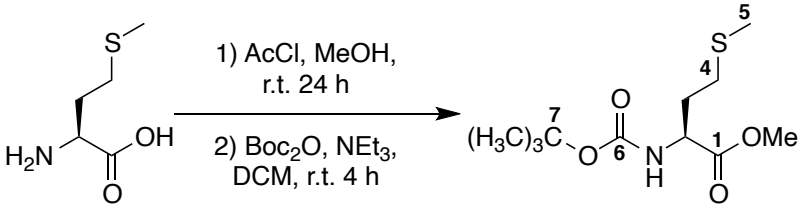


82 **110** TMP3 **32**

The synthesis was performed according to the procedure reported by Dippe *et al.*[124] DL-homocysteine thiolactone hydrochloride **82** (922 mg, 6.00 mmol) was placed under argon and anhydrous methanol (18 mL) was added. Sodium methoxide (30% w/w in methanol, anhydrous, 2.2 mL, 12.0 mmol) was added dropwise and the reaction was stirred at r.t. for 30 min. Ethyl bromide (493 μL, 6.60 mmol) was added dropwise and the reaction left stirring at r.t. overnight. The suspension was filtered under suction, the solids washed with methanol, and filtrate was concentrated under vacuum. The resulting residue was resuspended in $LiOH_{(aq)}$ (1 M, 8 mL) and solution was stirred at r.t. for 4 h.

Dowex™ 50WX8 hydrogen form (200-400 mesh) resin was added to the solution until it was neutralised. The resin suspension was then transferred to a chromatography column and washed by addition of approximately 15 column volumes of diH_2O. The product was then eluted with 30 mL $NH_{3\ (aq)}$ (10% *v/v*). The eluent was concentrated under vacuum, resuspended in a minimal volume of $CH_3CN_{(aq)}$ (5% *v/v*) and lyophilised, giving the ammonium salt of product **32** as a white powder (398 mg, 37%).**^{1}H NMR** (600 MHz; D_2O) δ 3.87-3.84 (1H, m, 2-H), 2.73-2.66 (2H, m, 4-H), 2.62 (2H, q, *J* = 7.5 Hz, 5-H), 2.23-2.09 (2H, m, 3-H), 1.25 (3H, t, *J* = 7.5 Hz, 6-H); **^{13}C NMR** (700 MHz; D_2O + CD_3OD) δ 172.0 (C-1), 52.4 (C-2), 30.0 (C-3), 26.5 (C-4), 25.3 (C-5), 14.3 (C-6); **HRMS** (ESI+) *m/z* Theoretical $[C_6H_{14}NO_2S]^+$: 164.0760; Measured $[MH]^+$: 164.0740; **Mp** 146-152℃ (diH_2O); $\boldsymbol{\nu_{max}}$**/cm^{-1}** (thin film): 3300-2500, 1563, 1410, 1339.

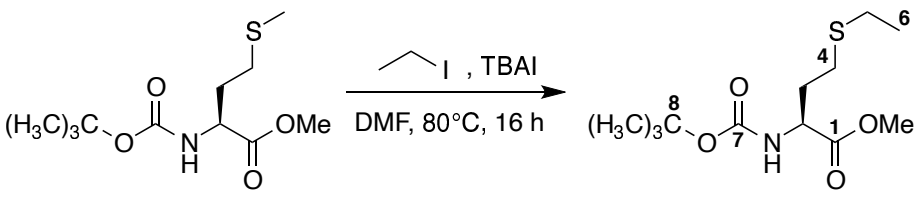
8.11 Synthesis of MeO-Met-NHBoc[183] 83



8a **83**

The synthesis was performed according to the procedure reported by Bhushan *et al.*[183] The reaction was carried out under anhydrous conditions. Acetyl chloride (11.9 mL, 167 mmol) was added dropwise to methanol (30 mL, cooled to 0 °C) over 5 min. The solution was stirred for a further 10 min at 0 °C, then L-methionine **8a** (10.0 g, 67.0 mmol) was added. After stirring for 18 h, the resulting solid was dissolved in methanol, then concentrated under vacuum. The residue was dissolved again in 150 mL dichloromethane and cooled to 0 °C, then triethylamine (18.7 mL, 130 mmol) added in portions, followed by di-tert-butyl decarbonate (17.5 g, 80.2 mmol). The reaction was stirred at r.t. for 4 h, at which time TLC analysis (30% EtOAc in petroleum ether 40-60) showed appearance of the product (R_f = 0.4, $KMnO_4$). The solvent was removed under vacuum to give a pale pink residue, which was purified by silica column chromatography (30% EtOAc in petroleum ether 40-60) to give the product **83** as a yellow oil (14.6 g, 83%). **^{1}H NMR** (400 MHz; $CDCl_3$) δ 4.45-4.2 (1H, m, 2-H), 3.72 (3H, s, OCH_3, 2.50 (2H, t, J = 7.8 Hz, 3-H), 2.13-2.03 (4H, m, 5-H, 4-H), 1.93-1.86 (1H, m, 4-H), 1.42 (9H, s, $C(CH_3)_3$); **^{13}C NMR** (700 MHz; $CDCl_3$) δ 172.8 (C-1), 155.3 (C-6), 80.0 (C-7), 52.7 (C-2), 52.4 (C-11), 32.2 (C-3), 29.9 (C-4), 28.3 ($C(CH_3)_3$), 15.5 (C-5); **HRMS** (ESI+) *m/z* Theoretical $[C_{11}H_{22}NO_4S]^+$: 264.1264; Measured $[MH]^+$: 264.1265.

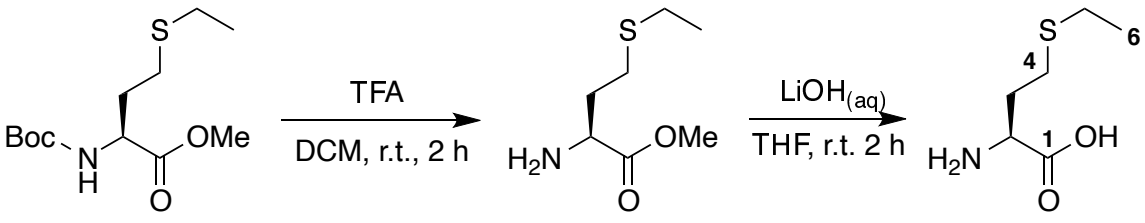
8.12 Synthesis of MeO-Eth-NHBoc 84



83 **84**

MeO-Met-NHBoc **83** (1.0 g, 3.80 mmol) was dissolved in DMF (50 mL). TBAI (7.0 g, 19.0 mmol) was added, followed by ethyl iodide (3.04 mL, 38.0 mmol). The solution was stirred and heated to 80℃. After 16 h, the reaction was cooled to r.t., and half of the DMF was removed under vacuum. DiH_2O (100 mL) was added and extraction performed with Et_2O (3x 50 mL). The organic layers were washed with diH_2O (50 mL) and brine (2x 50 mL) and dried with Na_2SO_4, before being filtered and concentrated under vacuum. The residue was was collected and purified by column chromatography (20% EtOAc in petroleum ether 40-60), giving the product **84** as an orange oil (86.4 mg, 8%).**^{1}H NMR** (600 MHz; $CDCl_3$) δ 4.46-4.21 (1H, m, 2-H), 3.80-3.70 (3H, s, OCH_3), 2.59-2.37 (4H, m, 4-H, 5-H), 2.15-2.07 (1H, m, 3-H), 1.94-1.87 (1H, m, 3-H), 1.45 (9H, s, $C(CH_3)_3$, 1.24 (3H, t, J = 7.4 Hz, 6-H). **^{13}C NMR** (600 MHz; $CDCl_3$) δ 172.9 (C-1), 155.5 (C-7), 80.2 (C-8), 52.9 (C-2), 52.7 (C-12), 32.9 (C-3), 28.0 ($C(CH_3)_3$), 27.4 (C-4), 26.0 (C-5), 14.8 (C-6). **HRMS** (ESI+) *m/z* Theoretical $[C_{12}H_{24}NO_4S]^+$: 278.1264; Measured $[MH]^+$: 278.1265

8.13 Synthesis of L-ethionine[125] 32a

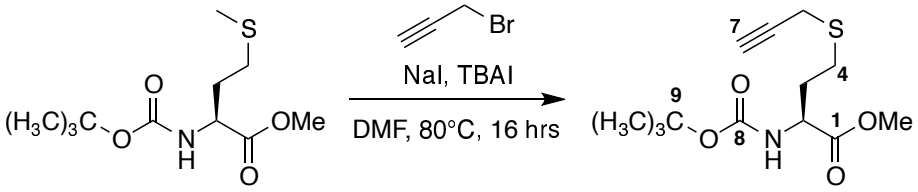


84 **112** **32a**

MeO-Eth-NHBoc **84** (86.4 mg, 0.31 mmol, 1 eqv.) was dissolved in 1 mL DCM. TFA was added (0.7 mL, 9.13 mmol) and the reaction stirred at r.t. for 3 h. After this time, TLC (15% EtOAc in petroleum ether 40-60) indicated formation of the product (Rf = 0.4, KMnO4). The solution was concentrated under vacuum to give a red-brown oil. This was dissolved in 2 mL tetrahydrofuran and cooled to 0℃ for 10 min. $LiOH_{(aq)}$ (1 M, 2 mL) was added dropwise. The reaction was left with stirring at r.t. for 16 h. DiH_2O (4 mL) was added, followed by washed Dowex™ 50WX8 hydrogen form (200-400 mesh) resin (5 cm^3). The suspension was mixed then decanted into a column, and the solution allowed to run through the resin. The resin was washed with diH_2O (15 mL), then the product eluted with $NH_{3(aq)}$ (28% (w/w), 5 mL) under gravity flow. The solution was concentrated

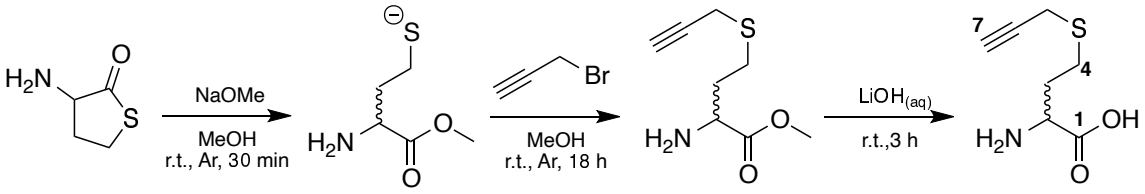
under vacuum to give the ammonium salt of the product **32a** as a yellow solid (23.7 mg, 65%).^{1}H NMR (700 MHz; D_2O) 3.76-3.68 (1H, m, 2-H), 3.04-2.78 (4H, m, 4-H, 5-H), 2.28-2.11 (2H, m, 3-H), 1.33-1.23 (3H, m, 6-H). **^{13}C NMR** (500 MHz; D_2O + CD_3OD) δ 172.0 (C-1), 55.8 (C-2), 32.8 (C-3), 27.9 (C-4), 26.4 (C-5), 15.2 (C-6). **HRMS** (ESI+) *m/z* Theoretical $[C_6H_{14}NO_2S]^+$: 164.0740; Measured $[MH]^+$: 164.0740; **$[\alpha]_D^{24}$** = +19.4 (H_2O, *c* 1.00); **Mp**: 150-158°C (diH_2O), **ν_{max}/cm^{-1}** (thin film): 3500-2500, 1564, 1410, 1339.

8.14 Synthesis of MeO-S-propargyl Hcy-NHBoc[215] 113



83 **113**

To MeO-Met-NHBoc **83** (500 mg, 1.90 mmol) was added DMF (25 mL), TBAI (3.5 g, 9.48 mmol) and NaI (10 mg, 0.067 mmol), followed by propargyl bromide (2.26 mL, 19.0 mmol). The solution was stirred and heated to 80°C for 16 h. The reaction was allowed to cool to r.t., and half of the DMF was removed under vacuum. DiH_2O (50 mL) was added and extraction performed with Et_2O (3x 25 mL). The organic layers were washed with diH_2O (25 mL) and brine (2x 25 mL) and dried with Na_2SO_4, before being filtered and concentrated under vacuum. The residue was collected and purified by column chromatography (20% EtOAc in petroleum ether 40-60), giving the product **113** as an orange oil (43.0 mg, 8% yield). **^{1}H NMR** (600 MHz; $CDCl_3$) δ 7.26 (1H, s, NH), 4.49-4.39 (1H, m, 2-H), 3.79 (3H, s, OCH_3), 3.24 (2H, s, 5-H_2), 2.73 (2H, t, *J* = 7.5 Hz, 4-H_2), 2.24 (1H, s, 7-H), 2.21-2.06 (1H, m, 3-*H*H, 2.00-1.89 (1H, m, 3-H*H*), 1.44 (9H, s, $C(CH_3)_3$); **^{13}C NMR** (600 MHz; $CDCl_3$) δ 173.0 (C-3), 155.5 (C-10), 80.3 (C-13), 79.8 (C-8), 79.2 (C-7), 71.4 (C-9), 52.8 (C-19), 52.7 (C-2), 32.2 (C-4), 28.4 ($C(CH_3)_3$), 19.2 (C-5); **HRMS** (ESI+) *m/z* Theoretical $[C_{13}H_{22}NO_4S]^+$: 288.1264; Measured $[MH]^+$: 288.1260



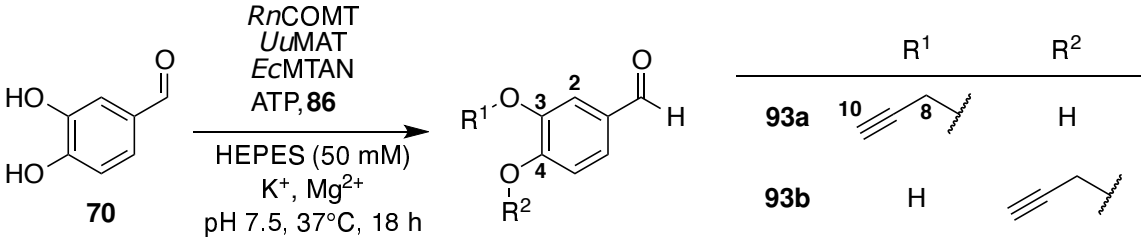
82 **86**

8.15 Synthesis of *S*-propargyl-small DL-homocysteine[198] 86

The synthesis was performed according to the procedure reported by Dippe *et al.*[124] DL-homocysteine thiolactone hydrochloride **82** (3g, 19.5 mmol) was placed under argon, and anhydrous methanol (60 mL) was added. Sodium methoxide (30% w/w in methanol, anhydrous, 7.24 mL, 39.0 mmol) was added dropwise and the reaction was stirred at r.t. for 30 min. Propargyl bromide (80% in toluene, 1.91 mL, 21.5 mmol) was added dropwise and the reaction stirred at r.t. overnight. The suspension was filtered under suction, the solids washed with methanol and the filtrate concentrated under vacuum and resuspended in $LiOH_{(aq)}$ (1 M, 20 mL). The solution was stirred at r.t. for 3 h, then diluted to a total volume of 50 mL with diH_2O. Dowex 50WX8 hydrogen form (200-400 mesh) resin was added gradually and the pH of the solution checked periodically until it had been acidified. The resin suspension was poured into a chromatography column and washed by passing through 100 mL diH_2O. The product was eluted with $NH_{3\,(aq)}$ (10% (w/w), 60 mL) and the eluate concentrated under vacuum until the product barely precipitated. The resulting solution was lyophilised to give the ammonium salt of the product **86** as an orange powder (2.50 g, 68%). **^{1}H NMR** (500 MHz; D_2O) δ 3.87-3.83 (1H, m, 2-H), 3.39 (2H, s, 5-H), 2.89-2.77 (2H, m, 4-H), 2.27-2.1 (2H, m, 3-H); **^{13}C NMR** (700 MHz; D_2O + CD_3OD) δ 176.8 (C-1), 80.5 (C-6), 72.7 (C-7), 54.8 (C-2), 31.7 (C-3), 27.3 (C-4), 18.6 (C-5); **HRMS** (ESI+) *m/z* Theoretical $[C_6H_{14}NO_2S]^+$: 164.0760; Measured $[MH]^+$: 164.0740; **Dp** 158-166℃ (diH_2O); $\boldsymbol{\nu_{max}}$**/cm^{-1}** (thin film): 3276, 3500-2500, 1572, 1407.

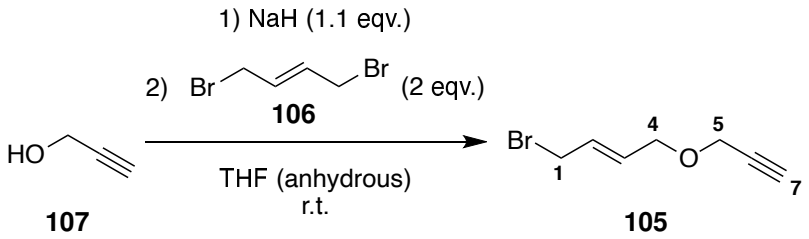
8.16 Enzyme cascade towards 3-hydroxy-4-propargyloxybenzaldehyde[216] 93

Substrate 3,4-dihydroxybenzaldehyde **70** (27.6 mg, 0.2 mmol), disodium ATP hydrate (anhydrous basis, 440.1 mg, 0.8 mmol) and *S*-propargyl-DL-homocysteine **86** (ammonium



salt, 152.2 mg, 0.8 mmol) were dissolved in 30 mL MT buffer (50 mM HEPES, 20 mM $MgCl_2$, 200 mM KCl, pH 7.5). To this was added purified *Rn*COMT, *Uu*MAT and *Ec*MTAN (all in MT buffer) to final concentrations of 0.5, 0.5 and 0.01 mg/mL respectively. The mixture was stirred overnight at 37°C. The proteins were then precipitated by addition of 1 vol. CH_3OH and centrifuged to remove the solids. The supernatant was concentrated under vacuum and the resulting residue redissolved in a minimal volume of $CH_3CN_{(aq)}$ 50% (*v/v*) and passed through a syringe filter (0.2 μm pore). The product was then purified by preparative HPLC, according to the procedure and method given in Section 8.4. Removal of the solvent gave a residue, which was redissolved in a minimal volume of $CH_3CN_{(aq)}$ 50% (*v/v*) and lyophilised. This gave product **93** as a light brown powder (5 mg, 14%).**^{1}H NMR** (500 MHz; CD_3OD) δ 9.81 (0.66H, s, CHO), 9.64 (0.33H, s, CHO), 7.53 (0.66H, d, *J* = 2.0 Hz, 2-H), 7.44 (1H, ddd, *J* = 8.1, 2.0, 1.2 Hz, 6-H), 7.36 (0.33H, d, *J* = 2.0 Hz, 2-H), 7.24 (0.33H, d, *J* = 8.1 Hz, 5-H), 6.87 (0.66H, d, *J* = 8.1 Hz, 5-H), 4.94 (0.66H, d, *J* = 2.4 Hz, 8-H), 4.84 (1.32H, d, *J* = 2.4 Hz, 8-H), 3.06 (0.33H, t, *J* = 2.4 Hz, 10-H), 2.96 (0.66H, t, *J* = 2.4 Hz, 10-H); **HRMS** (ESI+) *m/z* Theoretical $[C_{10}H_9O_3]^+$: 177.0546; Measured $[MH]^+$: 177.0548; **HPLC** RT (Method D): 4.6 min, 4.7 min.

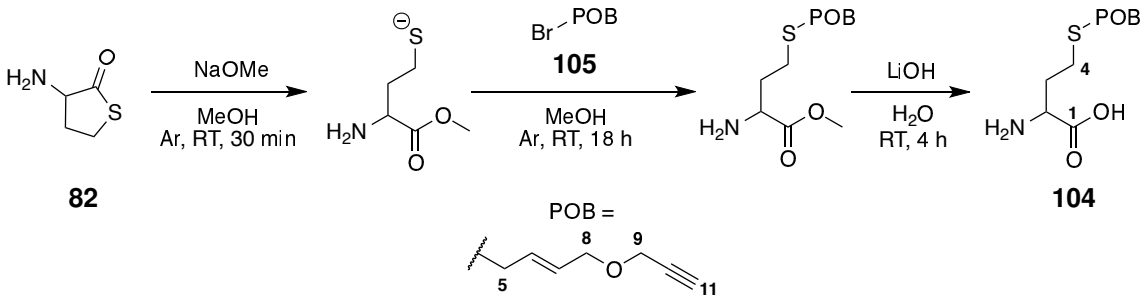
8.17 Synthesis of propargyloxybut-2-ene bromide[128] 105



Sodium hydride (60% in oil, 88 mg, 2.2 mmol) was added to a flask under argon and dissolved in anhydrous THF (2.5 mL) with stirring. Propargyl alcohol **107** (115 μL, 2 mmol) was also dissolved in anhydrous THF (2.5 mL), and this alcohol solution added

dropwise to the sodium hydride suspension. The mixture was stirred at r.t. for 15 min, then added dropwise to a solution of *trans*-1,4-dibromobut-2-ene **106** (856 mg, 4 mmol) in anhydrous THF (5 mL). The reaction was stirred at r.t. overnight, after which TLC analysis (10% EtOAc in petroleum ether) showed the starting material (R_f = 0.2, $KMnO_4$) and two higher spots (R_f = 0.3, 0.5) The products were isolated by silica column chromatography (10% EtOAc in petroleum ether), and upon analysis, the middle spot (R_f = 0.3) was found to represent the product. Fractions containing this compound were combined and the solvent removed under vacuum, giving the final product **105** as a green oil (38 mg, 10% yield). **^{1}H NMR** (500 MHz; $CDCl_3$) δ 6.02-5.94 (1H, m, 2-H), 5.88-5.81 (1H, m, 3-H), 4.15 (2H, d, *J* = 2.4 Hz, 5-H), 4.09, (2H, d, *J* = 5.7 Hz, 4-H), 3.95 (2H, d, *J* = 7.9 Hz, 1-H), 2.44 (1H, t, *J* = 2.4 Hz, 7-H); **^{13}C NMR** (500 MHz; $CDCl_3$) δ 131.0 (C-2), 129.7 (C-3), 79.7 (C-6), 74.8 (C-7), 69.0 (C-4), 57.5 (C-5), 31.8 (C-1). **HRMS** (ESI+) *m/z* Theoretical $[C_7H_9{}^{79}BrO]^+$: 188.9837; Measured $[MH]^+$: 188.9732.

8.18 Synthesis of *S*-propargyloxybut-2-enyl-DL-homocysteine[128] 104



The synthesis was performed according to the procedure reported by Dippe *et al.*[124] DL-homocysteine thiolactone hydrochloride **82** (28 mg, 0.18 mmol) was placed under argon and anhydrous methanol (2 mL) was added. Sodium methoxide (30% w/w in methanol, anhydrous, 67 μL, 0.36 mmol) was added dropwise and stirred at r.t. for 30 min. Compound **105** was added and the reaction was stirred at r.t. overnight. The solvent was removed under vacuum, and the residue resuspended in $LiOH_{(aq)}$ (1 M, 2 mL). This solution was stirred at RT for 4 h. The reaction was then diluted with diH_2O (10 mL), and Dowex™ 50WX8 hydrogen form (200-400 mesh) resin was added until the solution was

neutralised. The resin was transferred to a chromatography column and washed with 50 mL diH_2O. $NH_{3\ (aq)}$ (10% (w/w), 10 mL) was added to elute the product, and the eluate was concentrated under vacuum. The resulting residue was resuspended in a minimal volume of $CH_3CN_{(aq)}$ (5% *v/v*) and lyophilised, giving product **104** as a yellow powder (16.4 mg, 37%). **^{1}H NMR** (500 MHz; D_2O + CD_3OD)) δ 5.85-5.77 (1H, m, 6-H), 5.76-5.69 (1H, m, 7-H), 4.21 (2H, s, 9-H), 4.13 (2H, d, *J* = 6.2 Hz, 8-H), 3.53-3.49 (1H, m, 2-H), 3.23 (2H, d, *J* = 7.0 Hz, 5-H), 2.61-2.48 (2H, m, 4-H), 2.04-1.85 (2H, m, 3-H); **^{13}C NMR** (500 MHz; D_2O + CD_3OD) δ 180.4 (C-1), 132.3 (C-6), 128.7 (C-7), 80.0 (C-10), 73.5 (C-11), 70.5 (C-8), 57.5 (C-9), 55.8 (C-2), 33.7 (C-3), 32.9 (C-5), 27.0 (C-4); **HRMS** (ESI+) *m/z* Theoretical $[C_{11}H_{18}NO_3S]^+$: 244.0929; Measured $[MH]^+$: 244.0996

Appendix A: Sequences

A.1 DNA Sequences

*At*ASMT

CATATGAGCAGCGATCAGCTGAGCAAATTTCTGGATCGCAACAAAATGGAAGAT
AACAAACGCAAAGTGCTGGATGAAGAAGCGAAAGCGAGCCTGGATATTTGGAAAT
ATGTGTTTGGCTTTGCGGATATTGCGGCGGCGAAATGCGCGATTGATCTGAAAATT
CCGGAAGCGATTGAAAACCATCCGAGCAGCCAGCCGGTGACCCTGGCGGAACTG
AGCAGCGCGGTGAGCGCGAGCCCGAGCCATCTGCGCCGCATTATGCGCTTTCTG
GTGCATCAGGGCATTTTTAAAGAAATTCCGACCAAAGATGGCCTGGCGACCGGCT
ATGTGAACACCCCGCTGAGCCGCCGCCTGATGATTACCCGCCGCGATGGCAAAAG
CCTGGCGCCGTTTGTGCTGTTTGAAACCACCCCGGAAATGCTGGCGCCGTGGCTG
CGCCTGAGCAGCGTGGTGAGCAGCCCGGTGAACGGCAGCACCCCGCCGCCGTTT
GATGCGGTGCATGGCAAAGATGTGTGGAGCTTTGCGCAGGATAACCCGTTTCTGA
GCGATATGATTAACGAAGCGATGGCGTGCGATGCGCGCCGCGTGGTGCCGCGCG
TGGCGGGCGCGTGCCATGGCCTGTTTGATGGCGTGACCACCATGGTGGATGTGG
GCGGCGGCACCGGCGAAACCATGGGCATGCTGGTGAAAGAATTTCCGTGGATTAA
AGGCTTTAACTTTGATCTGCCGCATGTGATTGAAGTGGCGGAAGTGCTGGATGGC
GTGGAAAACGTGGAAGGCGATATGTTTGATAGCATTCCGGCGTGCGATGCGATTT
TTATTAAATGGGTGCTGCATGATTGGGGCGATAAAGATTGCATTAAAATTCTGAAAA
ACTGCAAAGAAGCGGTGCCGCCGAACATTGGCAAAGTGCTGATTGTGGAAAGCGT
GATTGGCGAAAACAAAAAAACCATGATTGTGGATGAACGCGATGAAAAACTGGAAC
ATGTGCGCCTGATGCTGGATATGGTGATGATGGCGCATACCAGCACCGGCAAAGA
ACGCACCCTGAAAGAATGGGATTTTGTGCTGAAAGAAGCGGGCTTTGCGCGCTAT
GAAGTGCGCGATATTGATGATGTGCAGAGCCTGATTATTGCGTATCGCAGCCTCG
AG

*Cn*ASMT

CATATGAACACCATTACCCTGAAACCGGGCAAAGAAAAAAGCCTGCTGCGCCG
CCATCCGTGGATTTATGCGACCGGCATTGCGACCACCGAAGGCCGCTGCGAACC
GGGCGCGACCGTGATTGTGCGCGCGGCGGATGGCCGCTTTCTGGCGAAAGCGG
CGTATAGCCCGGAAAGCCAGATTCGCGCGCGCGCGTGGACCTTTGATGAAAACGA
ACCGGTGGATCATGCGCTGTTTAAACGCCGCGTGGCGGCGGCGATTGCGTATCG
CCGCCAGTGGGTGAAAGATAGCGATGCGGTGCGCCTGATTTTTGGCGAAAGCGAT
CGCCTGCCGGGCCTGATTGTGGATTATTATGGCAACGGCGAAAAAGGCCAGCTGG
TGTGCCAGTTTAACAGCGCGGGCGTGGAACATTGGAAAACCGCGATTGTGCAGGC
GCTGGTGAAAGAAACCGGCTGCCCGAACGTGTATGAACGCAGCGATGCGGCGGT
GCGCCAGCGCGAAGGCCTGGAACTGGTGACCGGCGTGCTGGCGGGCGCGGAAC
CGGATCCGGCGCTGAGCGTGACCGAACATGGCGTGCGCTATTATGTGGATGTGC
GCAACGGCCATAAAACCGGCTTTTATGTGGATCAGCGCGATAACCGCAAACTGGT
GGGCGATCTGGCGGTGGGCCGCGAAGTGCTGAACTGCTTTTGCTATACCGGCGG
CTTTAGCCTGGCGGCGCTGCGCGGCGGCGCGACCAGCGTGACCAGCATTGATAG
CAGCGGCGAAGCGCTGAAAATTGCGGCGGGCAACGTGACCCTGAACGGCTTTGA
ACCGGAACGCGCGACCTGGCTGGATGCGGATGTGTTTAAAACCCTGCGCGAATTT
CGCGCGGAAGGCCGCCAGTTTGATCTGATTGTGCTGGATCCGCCGAAATTTGCGC
CGAGCGCGCAGCATATTGATCGCGCGGCGCGCGTATAAAGAAATTAACCTGGT
GGGCATGCAGCTGCTGCGCCCGGGCGGCCTGCTGTTTACCTATAGCTGCAGCGG
CGCGATTAGCATGGAACTGTTTCAGAAAATTGTGGCGGGCGCGGTGACCGATGCG
CGCGCGGATGCGCGCATTCTGCGCCGCCTGAGCGCGGGCACCGATCATCCGATG
CTGGCGGCGTTTCCGGAAGGCGAATATCTGAAAGGCCTGCTGCTGGAAAAAGTGG
CGCTCGAG

*Cv*ASMT

CATATGGGCTATGCGGCGCCGCAGGCGCGCCAGAGCGATAAAAAAATTTTTGA
TATTTATTTTGGCTTTCTGCATAGCTATGCGCTGCTGTTTGCGGATGAAGTGGGCC
TGTTTGATCTGCTGCGCTGCGAAGCGCTGACCCTGGATCAGGTGAGCATGGCGAC
CAGCCTGCCGTTTCGCAGCAGCCAGGCGCTGCTGAGCCTGTGCGCGAGCCTGGG
CCTGCTGGAAAAACGCGGCGAACGCTTTGCGCTGAGCGCGCTGGCGGAAGGCTT
TCTGGTGCGCGAAGCGGAAACCAGCTTTTGCGGCGTGCTGGCGAGCGCGCGCGG
CCAGGCGGCGGCGTTTAGCTATGATTTTTTTAAAGCGAGCCTGCTGAAAGGCGAA

AGCCAGCTGTTTGGCGGCCGCGATCTGTTTGATAACAACGCGCAGGATAGCGAAC
ATTGCGAAATTTTTACCCGCGCGATGCATAGCAAAAGCAAAGGCCCGGCGCAGGC
GTGGGTGGAAAAAATTGATCTGAGCGCGCATGCGTGCCTGCTGGATGTGGGCGG
CGGCAGCGGCGTGCATGCGATTAGCGCGCTGGCGCGCTGGCCGAACCTGAACGC
GGTGGTGTTTGATCTGCCGCCGGTGTGCGCGATTGCGGATACCTTTATTGAACGC
TATCAGATGATGGCGCGCGCGCAGACCCATGGCGGCGATATTTGGTATACCGATT
ATCCGTTTGCGGATGCGCATTTTTATAGCGATATTTTTCATGATTGGCCGCTGGAA
CGCTGCCGCTTTCTGGCGCGCAAAAGCTTTGATGCGCTGCCGAGCGGCGGCCGC
ATTATTCTGCATGAAATGCTGTTTAACACCCAGAAAACCGGCCCGCGCAACGTGGC
GGCGTATAACGCGAACATGCTGCTGTGGACCCAGGGCCAGCAGCTGAGCGAACC
GGAAGCGGCGGATCTGCTGCAGGCGGCGGGCTTTGTGGAAATTCTGGCGTTTCC
GACCGGCTATGGCGATTGGAGCCTGGTGACCGGCGTGAAACCGCTCGAG

*Cv*TAm

ATGCAGAAGCAACGTACGACCAGCCAATGGCGCGAACTGGATGCCGCCCATC
ACCTGCATCCGTTCACCGATACCGCATCGCTGAACCAGGCGGGCGCGCGCGTGA
TGACGCGCGGAGAGGGCGTCTACCTGTGGGATTCGGAAGGCAACAAGATCATCG
ACGGCATGGCCGGACTGTGGTGCGTGAACGTCGGCTACGGCCGCAAGGACTTTG
CCGAAGCGGCGCGCCGGCAGATGGAAGAGCTGCCGTTCTACAACACCTTCTTCAA
GACCACCCATCCGGCGGTGGTCGAGCTGTCCAGCCTGCTGGCTGAAGTGACGCC
GGCCGGTTTCGACCGCGTGTTCTATACCAATTCCGGTTCCGAATCGGTGGACACC
ATGATCCGCATGGTGCGCCGCTACTGGGACGTGCAGGGCAAGCCGGAGAAGAAG
ACGCTGATCGGCCGCTGGAACGGCTATCACGGCTCCACCATCGGCGGCGCCAGC
CTGGGCGGCATGAAGTACATGCACGAGCAGGGCGACTTGCCGATTCCGGGCATG
GCCCACATCGAGCAGCCTTGGTGGTACAAGCACGGCAAGGACATGACGCCGGAC
GAGTTCGGCGTGGTGGCCGCGCGCTGGCTGGAAGAGAAGATTCTGGAAATCGGC
GCCGACAAGGTGGCCGCCTTCGTCGGCGAACCCATCCAGGGCGCCGGCGGCGT
GATCGTCCCGCCGGCCACCTACTGGCCGGAAATCGAGCGCATTTGCCGCAAGTAC
GACGTGCTGCTGGTGGCCGACGAAGTGATCTGCGGCTTCGGGCGTACCGGCGAA
TGGTTCGGCCATCAGCATTTCGGCTTCCAGCCCGACCTGTTCACCGCCGCCAAGG
GCCTGTCCTCCGGCTATCTGCCGATAGGCGCGGTCTTTGTCGGCAAGCGCGTGG
CCGAAGGCCTGATCGCCGGCGGCGACTTCAACCACGGCTTCACCTACTCCGGCC
ACCCGGTCTGCGCCGCCGTCGCCCACGCCAACGTGGCGGCGCTGCGCGACGAG

GGCATCGTCCAGCGCGTCAAGGACGACATCGGCCCGTACATGCAAAAGCGCTGG
CGTGAAACCTTCAGCCGTTTCGAGCATGTGGACGACGTGCGCGGCGTCGGCATG
GTGCAGGCGTTCACCCTGGTGAAGAACAAGGCGAAGCGCGAGCTGTTCCCCGATT
TCGGCGAGATCGGCACGCTGTGCCGCGACATCTTCTTCCGCAACAACCTGATCAT
GCGGGCATGCGGCGACCACATCGTGTCGGCGCCGCCGCTGGTGATGACGCGGG
CGGAAGTGGACGAGATGCTGGCGGTGGCGGAACGCTGTCTGGAGGAATTCGAGC
AGACGCTGAAGGCGCGCGGGCTGGCTTAG

*Ec*MAT

ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAA
ATTGCTGACCAAATCTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAA
AGCACGCGTTGCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGC
GAAATCACCACCAGCGCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTC
GCGAAATTGGCTATGTGCATTCCGACATGGGCTTTGACGCTAACTCCTGTGCAGTT
CTGAGCGCTATCGGCAAACAGTCTCCTGACATCAACCAGGGCGTTGACCGTGCCG
ATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAA
TGAAACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAG
CGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCG
AAAAGCCAGGTGACTTTCCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGT
CGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCG
GTAATGGAAGAGATCATCAAGCCGATTCTGCCCGCTGAATGGCTGACTTCTGCCA
CCAAATTCTTCATCAACCCGACCGGTCGTTTTGTTATCGGTGGCCCGATGGGTGAC
TGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTC
ACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGC
CTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGT
TGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACTTCCATCATGGT
AGAAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACTCTGCTGGTACGTG
AGTTCTTCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCC
GATCTACAAAGAAACCGCAGCATACGGTCACTTTGGTCGTGAACATTTCCCGTGGG
AAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCTGCCGGTCTGAAGTAA

*Ec*MTAN

ATGAAAATCGGCATCATTGGTGCAATGGAAGAAGAAGTTACGCTGCTGCGTGA
CAAAATCGAAAACCGTCAAACTATCAGTCTCGGCGGTTGCGAAATCTATACCGGCC
AACTGAATGGAACCGAGGTTGCGCTTCTGAAATCGGGCATCGGTAAAGTCGCTGC
GGCGCTGGGTGCCACTTTGCTGTTGGAACACTGCAAGCCAGATGTGATTATTAACA
CCGGTTCTGCCGGTGGCCTGGCACCAACGTTGAAAGTGGGCGATATCGTTGTCTC
GGACGAAGCACGTTATCACGACGCGGATGTCACGGCATTTGGTTATGAATACGGT
CAGTTACCAGGCTGTCCGGCAGGCTTTAAAGCTGACGATAAACTGATCGCTGCCG
CTGAGGCCTGCATTGCCGAACTGAATCTTAACGCTGTACGTGGCCTGATTGTTAGC
GGCGACGCTTTCATCAACGGTTCTGTTGGTCTGGCGAAAATCCGCCACAACTTCC
CACAGGCCATTGCTGTAGAGATGGAAGCGACGGCAATCGCCCATGTCTGCCACAA
TTTCAACGTCCCGTTTGTTGTCGTACGCGCCATCTCCGACGTGGCCGATCAACAGT
CTCATCTTAGCTTCGATGAGTTCCTGGCTGTTGCCGCTAAACAGTCCAGCCTGATG
GTTGAGTCACTGGTGCAGAAACTTGCACATGGCTAA

*Mj*MAT (Genomic)

ATGAGAAACATAATTGTAAAAAAATTAGATGTTGAACCAATTGAAGAAAGACCAA
CTGAAATTGTTGAGAGGAAGGGATTGGGGCATCCAGATTCAATTTGTGATGGTATT
GCTGAGAGTGTTAGTAGGGCTTTATGTAAGATGTACATGGAGAAGTTTGGAACTAT
TTTGCACCACAATACAGACCAAGTTGAGCTTGTAGGGGGACATGCATATCCTAAGT
TTGGAGGAGGAGTAATGGTAAGCCCTATTTATATTTTATTATCTGGAAGAGCAACAA
TGGAAATCTTAGATAAGGAGAAAAATGAAGTTATAAAGCTCCCAGTAGGAACAACT
GCTGTTAAAGCTGCTAAAGAATATTTAAAGAAGGTTTTAAGAAATGTTGATGTTGAT
AAAGATGTTATTATTGACTGCAGAATTGGGCAGGGAAGTATGGATTTAGTTGATGT
CTTTGAGAGACAAAAGAATGAAGTTCCTTTAGCTAATGATACATCATTTGGAGTAGG
TTATGCTCCATTATCAACAACAGAGAGGTTAGTTTTAGAAACAGAGAGATTTTTAAA
TAGTGATGAGTTAAAGAATGAGATTCCAGCTGTAGGAGAGGACATAAAGGTTATGG
GATTAAGAGAGGGTAAGAAGATAACTTTAACCATTGCTATGGCTGTTGTTGATAGG
TATGTTAAAAATATTGAGGAATATAAGGAAGTTATTGAAAAGGTTAGAAAGAAGGTT
GAAGATTTAGCTAAGAAGATAGCTGATGGATATGAGGTTGAAATTCATATAAATACA
GCAGATGATTATGAGAGGGAGAGTGTCTATCTAACAGTTACTGGAACATCAGCAGA
GATGGGGGATGATGGTTCAGTTGGGAGAGGAAATAGAGTTAATGGATTGATAACT
CCATTCAGACCTATGAGTATGGAGGCAGCAAGTGGTAAAAACCCAGTAAATCACGT
TGGTAAAATCTACAATATCTTAGCAAACTTAATAGCAAACGATATTGCCAAATTGGA

AGGAGTTAAAGAGTGCTATGTTAGAATATTAAGCCAAATTGGTAAGCCAATCAATGA
GCCAAAGGCTTTAGATATAGAAATTATAACTGAAGATAGCTATGATATAAAGGATAT
TGAACCAAAAGCAAAAGAGATAGCCAATAAATGGTTAGATAACATCATGGAAGTTC
AAAAGATGATTGTTGAAGGAAAAGTAACTACATTCTAA

*Tk*MAT

ATGGCTGGAAAGGTCAGGAACATAGTCGTTGAGGAGCTCGTCAGGACCCCAGT
TGAGATGCAGAAGGTTGAGCTCGTTGAGAGGAAGGGTATCGGTCACCCCGACAGC
ATAGCCGACGGTATAGCCGAGGCCGTCAGCAGGGCGCTCTCCAGGGAGTACGTG
AAGAGGTACGGCATCATTCTCCACCACAACACCGACCAGGTCGAGGTTGTCGGCG
GAAGGGCCTACCCGCAGTTCGGCGGCGGTGAGGTCATCAAGCCGATCTACATCCT
CCTCTCCGGAAGGGCCGTCGAGATGGTTGACAGGGAGTTCTTCCCCGTCCATGAG
ATAGCACTTAAGGCCGCAAAGGACTACCTCAGGAAGGCCGTCAGGCACCTCGACC
TCGAGCACCACGTCATCATAGACTCCCGCATCGGACAGGGAAGCGTTGACCTCGT
TGGAGTCTTCAACAAGGCTAAAAAGAACCCAATCCCGCTCGCCAACGACACCTCCT
TTGGTGTCGGCTACGCCCCGCTCAGCGAGACCGAGAAGATCGTCCTTGAGACCGA
GAAGTACCTCAACAGCGACGAGTTCAAGAAGAAGTACCCCGCCGTCGGTGAGGAC
ATCAAGGTTATGGGTCTCAGGAAGGGAGACGAGATAGACCTCACCATCGCCGCTG
CCATCGTTGACAGCGAGGTTGACAACCCCGACGATTACATGGCCGTGAAGGAGGC
CATCTACGAGGCCGCCAAGGGGATCGTCGAGTCCCACACAGAGAGGCCGACCAA
CATCTACGTGAACACCGCCGATGATCCGAAGGAGGGCATCTACTACATAACCGTC
ACAGGAACTAGCGCCGAGGCCGGCGACGACGGTTCCGTCGGAAGGGGCAACAG
GGTTAACGGCCTCATCACCCCCAACAGGCACATGAGCATGGAGGCAGCGGCGGG
TAAGAACCCGGTGAGCCACGTCGGTAAGATCTACAACATCCTCTCAATGCTCATAG
CGAACGACATAGCCGAGCAGGTCGAGGGCGTTGAGGAGGTCTACGTTAGAATCCT
CAGCCAGATTGGAAAGCCGATAGACGAGCCTCTCGTTGCCAGCGTGCAGATAATC
CCGAAGAAGGGCTACTCAATCGACGTCCTCCAGAAGCCGGCCTACGAGATAGCCG
ACGAGTGGCTGGCCAACATAACGAAGATACAGAAGATGATCCTCGAGGACAAGGT
AAACGTCTTCTGA

*Rn*COMT (Wild-type)

ATGGGTGACACAAAGGAGCAGCGCATCCTGCGCTACGTGCAGCAGAATGCAA

AGCCTGGAGACCCTCAGAGCGTCCTGGAGGCCATCGACACCTACTGCACACAGAA
GGAATGGGCCATGAATGTGGGTGACGCGAAAGGCCAAATCATGGATGCAGTGATT
CGGGAGTACAGCCCCTCCCTGGTGCTGGAACTGGGAGCTTACTGTGGCTACTCAG
CAGTGCGAATGGCTCGCCTGCTGCAGCCTGGAGCCAGGCTTCTCACCATGGAGAT
GAACCCTGACTACGCTGCCATCACCCAGCAAATGCTGAACTTTGCAGGCCTACAG
GACAAAGTCACCATCCTCAATGGGGCATCCCAGGATCTTATCCCCCAGCTGAAGA
AGAAGTACGACGTGGACACACTAGACATGGTCTTTCTTGACCACTGGAAAGACCG
CTACCTTCCAGACACACTTCTCCTGGAGAAATGTGGCCTGCTGCGCAAGGGGACA
GTGCTCCTAGCTGACAACGTCATCGTCCCGGGAACCCCTGACTTCCTGGCGTATG
TGAGAGGGAGCAGCAGCTTCGAGTGCACACACTACAGCTCATACCTGGAGTACAT
GAAAGTTGTAGACGGCTTGGAGAAGGCAATCTACCAGGGTCCAAGTAGCCCTGAC
AAGTCTTGA

*Rn*COMT (E199D)

ATGGGCGATACCAAAGAACAGCGCATTCTGCGCTATGTGCAGCAGAACGCGAA
ACCGGGCGATCCGCAGAGCGTGCTGGAAGCGATTGATACCTATTGCACCCAGAAA
GAATGGGCGATGAACGTGGGCGATGCGAAAGGCCAGATTATGGATGCGGTGATTC
GCGAATATAGCCCGAGCCTGGTGCTGGAACTGGGCGCGTATTGCGGCTATAGCG
CGGTGCGCATGGCGCGCCTGCTGCAGCCGGGCGCGCGCCTGCTGACCATGGAA
ATGAACCCGGATTATGCGGCGATTACCCAGCAGATGCTGAACTTTGCGGGCCTGC
AGGATAAAGTGACCATTCTGAACGGCGCGAGCCAGGATCTGATTCCGCAGCTGAA
AAAAAAATATGATGTGGATACCCTGGATATGGTGTTTCTGGATCATTGGAAAGATC
GCTATCTGCCGGATACCCTGCTGCTGGAAAAATGCGGCCTGCTGCGCAAAGGCAC
CGTGCTGCTGGCGGATAACGTGATTGTGCCGGGCACCCCGGATTTTCTGGCGTAT
GTGCGCGGCAGCAGCAGCTTTGAATGCACCCATTATAGCAGCTATCTGGATTATAT
GAAAGTGGTGGATGGCCTGGAAAAAGCGATTTATCAGGGCCCGAGCAGCCCGGA
TAAAAGCTAA

*Rn*COMT (M40A)

ATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCG
GCAGCCATATGGGCGATACCAAAGAACAGCGCATTCTGCGCTATGTGCAGCAGAA
CGCGAAACCGGGCGATCCGCAGAGCGTGCTGGAAGCGATTGATACCTATTGCACC

CAGAAAGAATGGGCGGCGAACGTGGGCGATGCGAAAGGCCAGATTATGGATGCG
GTGATTCGCGAATATAGCCCGAGCCTGGTGCTGGAACTGGGCGCGTATTGCGGCT
ATAGCGCGGTGCGCATGGCGCGCCTGCTGCAGCCGGGCGCGCGCCTGCTGACC
ATGGAAATGAACCCGGATTATGCGGCGATTACCCAGCAGATGCTGAACTTTGCGG
GCCTGCAGGATAAAGTGACCATTCTGAACGGCGCGAGCCAGGATCTGATTCCGCA
GCTGAAAAAAAAATATGATGTGGATACCCTGGATATGGTGTTTCTGGATCATTGGA
AAGATCGCTATCTGCCGGATACCCTGCTGCTGGAAAAATGCGGCCTGCTGCGCAA
AGGCACCGTGCTGCTGGCGGATAACGTGATTGTGCCGGGCACCCCGGATTTTCTG
GCGTATGTGCGCGGCAGCAGCAGCTTTGAATGCACCCATTATAGCAGCTATCTGG
AATATATGAAAGTGGTGGATGGCCTGGAAAAAGCGATTTATCAGGGCCCGAGCAG
CCCGGATAAAAGCTAACTCGAG

*Rn*COMT (E199D, M40A)

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCG
GCAGCCATATGGGCGATACCAAAGAACAGCGCATTCTGCGCTATGTGCAGCAGAA
CGCGAAACCGGGCGATCCGCAGAGCGTGCTGGAAGCGATTGATACCTATTGCACC
CAGAAAGAATGGGCGGCGAACGTGGGCGATGCGAAAGGCCAGATTATGGATGCG
GTGATTCGCGAATATAGCCCGAGCCTGGTGCTGGAACTGGGCGCGTATTGCGGCT
ATAGCGCGGTGCGCATGGCGCGCCTGCTGCAGCCGGGCGCGCGCCTGCTGACC
ATGGAAATGAACCCGGATTATGCGGCGATTACCCAGCAGATGCTGAACTTTGCGG
GCCTGCAGGATAAAGTGACCATTCTGAACGGCGCGAGCCAGGATCTGATTCCGCA
GCTGAAAAAAAAATATGATGTGGATACCCTGGATATGGTGTTTCTGGATCATTGGA
AAGATCGCTATCTGCCGGATACCCTGCTGCTGGAAAAATGCGGCCTGCTGCGCAA
AGGCACCGTGCTGCTGGCGGATAACGTGATTGTGCCGGGCACCCCGGATTTTCTG
GCGTATGTGCGCGGCAGCAGCAGCTTTGAATGCACCCATTATAGCAGCTATCTGG
ATTATATGAAAGTGGTGGATGGCCTGGAAAAAGCGATTTATCAGGGCCCGAGCAG
CCCGGATAAAAGCTAACTCGAG

Δ29*Tf*NCS

GTTTAACTTTTAGGAGGTAAAACATATGTTGCATCACCAGGGTATCATCAATCAA
GTTAGCACCGTCACGAAAGTAATTCATCACGAGCTGGAAGTTGCGGCATCCGCTG
ACGACATTTGGACCGTGTACAGCTGGCCGGGTCTGGCGAAGCACTTGCCGGATCT
GCTGCCTGGCGCGTTCGAAAAACTGGAGATTATCGGCGATGGCGGTGTTGGTACG
ATTCTGGACATGACCTTTGTCCCGGGTGAATTCCCGCACGAGTATAAAGAGAAATT

CATCCTGGTTGATAACGAACATCGTCTGAAGAAGGTGCAGATGATCGAAGGCGGC
TATCTGGACCTGGGTGTGACGTATTACATGGACACGATTCACGTTGTGCCGACCG
GTAAAGACAGCTGCGTCATCAAGAGCAGCACTGAGTACCACGTCAAGCCGGAGTT
TGTGAAGATTGTTGAGCCGCTGATCACCACCGGTCCACTGGCAGCCATGGCAGAT
GCCATTAGCAAGTTGGTCCTGGAACATAAATCTAAAAGCAACTCCGATGAAATTGA
GGCGGCGATCATCACCGTGCTGGAGCATCACCACCACCATCACTGATAAAAGCTT
CCCC

*Cr*16OMT

ATGGATGTGCAGAGCGAAGAATTTCGCGGCGCGCAGGCGCAGATTTGGAGCC
AGAGCTGCAGCTTTATTACCAGCGCGAGCCTGAAATGCGCGGTGAAACTGGGCAT
TCCGGATACCATTGATAACCATGGCAAACCGATTACCCTGAGCGAACTGACCAACG
CGCTGGTGCCGCCGGTGCATCCGAGCAAAGCGCCGTTTATTTATCGCCTGATGCG
CGTGCTGGCGAAAAACGGCTTTTGCAGCGAAGAACAGCTGGATGGCGAAACCGAA
CCGCTGTATAGCCTGACCCCGAGCAGCCGCATTCTGCTGAAAAAAGAACCGCTGA
ACCTGCGCGGCATTGTGCTGACCATGGCGGATCCGGTGCAGCTGAAAGCGTGGG
AAAGCCTGAGCGATTGGTATCAGAACGAAGATGATAGCAGCACCGCGTTTGAAAC
CGCGCATGGCAAAAACTTTTGGGGCTATAGCAGCGAACATATGGAACATGCGGAA
TTTTTTAACGAAGCGATGGCGAGCGATAGCCAGCTGATTAGCAAACTGCTGATTGG
CGAATATAAATTTCTGTTTGAAGGCCTGGCGAGCCTGGTGGATATTGGCGGCGGC
ACCGGCACCATTGCGAAAGCGATTGCGAAAAACTTTCCGCAGCTGAAATGCACCG
TGTTTGATCTGCCGCATGTGGTGGCGAACCTGGAAAGCAAAGAAAACGTGGAATT
TGTGGCGGGCGATATGTTTGAAAAAATTCCGAGCGCGAACGCGATTTTTCTGAAAT
GGATTCTGCATGATTGGAACGATGAAGATTGCGTGAAAATTCTGAAAAGCTGCAAA
AAAGCGATTCCGGCGAAAGGCGGCAAAGTGATTATTATTGATATGGTGATGTATAG
CGATAAAAAAGATGATCATCTGGTGAAAACCCAGACCAGCATGGATATGGCGATGC
TGGTGAACTTTGCGGCGAAAGAACGCTGCGAAAAAGAATGGGCGTTTCTGTTTAAA
GAAGCGGGCTTTAGCGATTATAAAATTTATCCGAAACTGGATTTTACCCGCAGCCT
GATTGAAGTGTATCCG

*Hs*ASMT

ATGGGCAGCAGCGAAGATCAGGCGTATCGCCTGCTGAACGATTATGCGAACG

GCTTTATGGTGAGCCAGGTGCTGTTTGCGGCGTGCGAACTGGGCGTGTTTGATCT
GCTGGCGGAAGCGCCGGGCCCGCTGGATGTGGCGGCGGTGGCGGCGGGCGTG
CGCGCGAGCGCGCATGGCACCGAACTGCTGCTGGATATTTGCGTGAGCCTGAAA
CTGCTGAAAGTGGAAACCCGCGGCGGCAAAGCGTTTTATCGCAACACCGAACTGA
GCAGCGATTATCTGACCACCGTGAGCCCGACCAGCCAGTGCAGCATGCTGAAATA
TATGGGCCGCACCAGCTATCGCTGCTGGGGCCATCTGGCGGATGCGGTGCGCGA
AGGCCGCAACCAGTATCTGGAAACCTTTGGCGTGCCGGCGGAAGAACTGTTTACC
GCGATTTATCGCAGCGAAGGCGAACGCCTGCAGTTTATGCAGGCGCTGCAGGAAG
TGTGGAGCGTGAACGGCCGCAGCGTGCTGACCGCGTTTGATCTGAGCGTGTTTCC
GCTGATGTGCGATCTGGGCGGCGGCGCGGGCGCGCTGGCGAAAGAATGCATGA
GCCTGTATCCGGGCTGCAAAATTACCGTGTTTGATATTCCGGAAGTGGTGTGGAC
CGCGAAACAGCATTTTAGCTTTCAGGAAGAAGAACAGATTGATTTTCAGGAAGGCG
ATTTTTTTAAAGATCCGCTGCCGGAAGCGGATCTGTATATTCTGGCGCGCGTGCTG
CATGATTGGGCGGATGGCAAATGCAGCCATCTGCTGGAACGCATTTATCATACCTG
CAAACCGGGCGGCGGCATTCTGGTGATTGAAAGCCTGCTGGATGAAGATCGCCG
CGGCCCGCTGCTGACCCAGCTGTATAGCCTGAACATGCTGGTGCAGACCGAAGG
CCAGGAACGCACCCCGACCCATTATCATATGCTGCTGAGCAGCGCGGGCTTTCGC
GATTTTCAGTTTAAAAAAACCGGCGCGATTTATGATGCGATTCTGGCGCGCAAA

*Ms*MTF1

ATGAGCGAAGATGTGCCGACCAGCGAAGATGTGCCGACCCTGCAGAAACGCA
ACGGCCTGTTTGCGACCTGCACCGCGCGCGCGAACGCGATGGTGGGCGCGGTG
GCGAACGGCCTGAGCGTGCGCGCGCGCCTGCTGACCCAGGCGGTGCGCGCGGA
ATATTGGCTGGCGCGCAAACTGCTGCCGGATGTGTATAGCAACGATGCGCTGATT
TGCTTTAACAGCCATGCGTTTATGGATGATCCGGATTTTCAGCGCGCGTATCGCCG
CGGCGCGCGCGCGCTGGGCGATAACGATTGGTATCAGTGGCATTGGCGCGTGCA
TGTGGGCCTGTGGGCGGCGGCGAGCGCGAGCAAAATTGATGGCGCGTTTGTGGA
ATGCGGCGTGAGCTATGGCTTTCTGAGCAGCGCGATTATGGAATATCTGGATTGG
GATAAACTGGGCAAAACCTTTTATCTGCTGGATACCTTTGCGGGCCTGGATCCGCG
CTATGTGACCGAAGCGGAACGCGCGAGCGGCGCGCTGGAACGCAGCGAAGAACA
TCTGCGCAACGGCTTTTATGTGGATAGCGTGGATAGCGTGCGCGCGAACTTTGCG
CAGTGGAAAAACCAGCGCATTATTGTGGGCGCGGTGCCGGAAACCCTGGCGGAA
GTGGATGCGGAAGCGGTGGCGTTTCTGCATATTGATATGAACTGCGCGCCGCCGG

AAGTGGCGACCCTGCGCTATTTTTGGCCGCGCCTGAGCCCGGGCGCGTTTGTGCT
GCTGGATGATTATGCGAACCGCGGCCGCGATGAACAGCGCGTGGCGATGGATGA
AGTGGCGAGCGAACTGGGCGTGCAGATTTGCACCCTGCCGACCGGCCAGGGCCT
GCTGATTAAACCGCCGCTG

*Sf*MOMT

ATGGCGCCGAGCCCGGATCATGCGCGCGATCTGTATATTGAACTGCTGAAAAA
AGTGGTGAGCAACGTGATTTATGAAGATCCGACCCATGTGGCGGGCATGATTACC
GATGCGAGCTTTGATCGCACCAGCCGCGAAAGCGGCGAAGATTATCCGACCGTG
GCGCATACCATGATTGGCCTGAAACGCCTGGATAACCTGCATCGCTGCCTGGCGG
ATGTGGTGGAAGATGGCGTGCCGGGCGATTTTATTGAAACCGGCGTGTGGCGCG
GCGGCGCGTGCATTTTTGCGCGCGGCCTGCTGAACGCGTATGGCCAGGCGGATC
GCACCGTGTGGGTGGCGGATAGCTTTCAGGGCTTTCCGGAACTGACCGGCAGCG
ATCATCCGCTGGATGTGGAAATTGATCTGCATCAGTATAACGAAGCGGTGGATCTG
CCGACCAGCGAAGAAACCGTGCGCGAAAACTTTGCGCGCTATGGCCTGCTGGATG
ATAACGTGCGCTTTCTGGCGGGCTGGTTTAAAGATACCATGCCGGCGGCGCCGGT
GAAACAGCTGGCGGTGATGCGCCTGGATGGCGATAGCTATGGCGCGACCATGGA
TGTGCTGGATAGCCTGTATGAACGCCTGAGCCCGGGCGGCTATGTGATTGTGGAT
GATTATTGCATTCCGGCGTGCCGCGAAGCGGTGCATGATTTTCGCGATCGCCTGG
GCATTCGCGATACCATTCATCGCATTGATCGCCAGGGCGCGTATTGGCGCCATAG
CGGC

*Sf*DMOMT

ATGGCGGTGCAGAAAGAAGCGACCCTGGTGCGCCAGATTATTCGCGCGGCGG
GCGGCCATGCGGCGGATGTGCGCGAACTGGTGGCGGAACATGGCCCGGAAGCG
GTGACCGCGGTGCTGGTGGATGAAATTGTGAGCCGCGCGCCGCATCCGGTGAAC
GATGTGCCGGTGCTGGTGGAACTGGCGGTGCGCAGCGGCGATGCGCTGGTGCC
GCGCCGCCTGGCGGTGGCGCAGGGCGCGCCGGTGCGCCGCGCGGCGCCGGAT
GATGATGGCTTTGTGGCGATGCGCGTGGAATATGAACTGGATGAACTGGTGCGCG
AACTGTTTGGCCCGTGCCGCGAACGCGCGGCGGGCACCCGCGGCACCACCCTGT
TTCCGTATGCGACCAGCGGCACCGGCCATATTGATACCTATTTTCTGGCGGCGCA
GCAGGCGACCGCGACCGTGCTGGCGGGCTGCACCAGCGCGAAACCGGATCTGAA

CGAACTGACCAGCCGCTATCTGACCCCGAAATGGGGCAGCCTGCATTGGTTTACC
CCGCATTATGATCGCCATTTTCGCGAATATCGCAACGAAGAAGTGCGCGTGCTGG
AAATTGGCATTGGCGGCTATCAGCATCCGGAATGGGGCGGCGGCAGCCTGCGCA
TGTGGAAACATTTTTTTCATCGCGGCCTGATTTATGGCCTGGATATTGAAGATAAAA
GCCATGCGGAAGAACAGCGCATTACCACCGTGGTGGGCGATCAGAACGATCCGG
GCTGCCTGACCGAACTGGCGGCGCGCTATGGCCCGTTTGATATTGTGATTGATGA
TGGCAGCCATATTAACGAACATGTGCGCACCAGCTTTCATGCGCTGTTTCCGCATG
TGCGCCCGGGCGGCCTGTATGTGATTGAAGATCTGTGGACCGCGTATTGGCCGG
GCTTTGGCGGCGATAGCGATCCGGGCAAAAGCGATCTGACCAGCCTGGGCCTGG
TGAAAAGCCTGGTGGATAGCCTGCAGCATCAGGAACTGCCGGAAGATAGCGGCC
GCAGCCCGGGCTATGCGGATCGCCATGTGGTGGGCCTGCATGTGTATCATAACCT
GGCGTTTATTGAAAAAGGCGTGAACAGCGAAGGCGGCATTCCGGGCTGGATTCCG
CGCGATTTTGATGCGCTGGTGGCGGCGAGCAGCGGCGGCGCGGCG

*Sm*CHOMT

ATGGGCAACAGCTATATTACCAAAGAAGATAACCAGATTAGCGCGACCAGCGA
ACAGACCGAAGATAGCGCGTGCCTGAGCGCGATGGTGCTGACCACCAACCTGGT
GTATCCGGCGGTGCTGAACGCGGCGATTGATCTGAACCTGTTTGAAATTATTGCG
AAAGCGACCCCGCCGGGCGCGTTTATGAGCCCGAGCGAAATTGCGAGCAAACTG
CCGGCGAGCACCCAGCATAGCGATCTGCCGAACCGCCTGGATCGCATGCTGCGC
CTGCTGGCGAGCTATAGCGTGCTGACCAGCACCACCCGCACCATTGAAGATGGCG
GCGCGGAACGCGTGTATGGCCTGAGCATGGTGGGCAAATATCTGGTGCCGGATG
AAAGCCGCGGCTATCTGGCGAGCTTTACCACCTTTCTGTGCTATCCGGCGCTGCT
GCAGGTGTGGATGAACTTTAAAGAAGCGGTGGTGGATGAAGATATTGATCTGTTTA
AAAACGTGCATGGCGTGACCAAATATGAATTTATGGGCAAAGATAAAAAAATGAAC
CAGATTTTTAACAAAAGCATGGTGGATGTGTGCGCGACCGAAATGAAACGCATGCT
GGAAATTTATACCGGCTTTGAAGGCATTAGCACCCTGGTGGATGTGGGCGGCGGC
AGCGGCCGCAACCTGGAACTGATTATTAGCAAATATCCGCTGATTAAAGGCATTAA
CTTTGATCTGCCGCAGGTGATTGAAAACGCGCCGCCGCTGAGCGGCATTGAACAT
GTGGGCGGCGATATGTTTGCGAGCGTGCCGCAGGGCGATGCGATGATTCTGAAA
GCGGTGTGCCATAACTGGAGCGATGAAAAATGCATTGAATTTCTGAGCAACTGCCA
TAAAGCGCTGAGCCCGAACGGCAAAGTGATTATTGTGGAATTTATTCTGCCGGAAG
AACCGAACACCAGCGAAGAAAGCAAACTGGTGAGCACCCTGGATAACCTGATGTT

TATTACCGTGGGCGGCCGCGAACGCACCGAAAAACAGTATGAAAAACTGAGCAAA
CTGAGCGGCTTTAGCAAATTTCAGGTGGCGTGCCGCGCGTTTAACAGCCTGGGCG
TGATGGAATTTTATAAA

*St*FKMT

ATGAGCGATGTGGTGGAAACCCTGCGCCTGCCGAACGGCGCGACCGTGGCGC
ATGTGAACGCGGGCGAAGCGCAGTTTCTGTATCGCGAAATTTTTACCGATCGCTG
CTATCTGCGCCATGGCGTGGAACTGCGCCCGGGCGATGTGGTGTTTGATGTGGG
CGCGAACATTGGCATGTTTATGCTGTTTGCGCATCTGGAACATCCGGGCGTGACC
GTGCATGCGTTTGAACCGGCGCCGGTGCCGTTTGCGGCGCTGCGCGCGAACGCG
GTGCGCCATCGCGTGGCGGGCCGCGTGGATCAGTGCGCGGTGAGCGATGAAGC
GGGCGTGCGCCGCATGACCTTTTATCCGGATGCGACCCTGATGAGCGGCTTTCAT
CCGGATGCGGCGGCGCGCAAAGAACTGCTGCGCACCCTGGGCCTGAACGGCGG
CTATACCGCGGAAGATGTGGATATGATGCTGGCGCAGCTGCCGGATACCGGCGAA
GAAATTGAAACCAGCGTGGTGCGCCTGAGCGATGTGATTGCGGAACGCGGCATTG
CGGCGATTGGCCTGCTGAAAATTGATGTGGAAAAAAGCGAACGCCGCGTGCTGGC
GGGCGTGGAAGATGCGGATTGGCCGCGCATTCGCCAGGTGGTGGCGGAAGTGCA
TGATGTGGATGGCGCGCTGGGCGAAGTGGTGGCGCTGCTGCGCGGCCATGGCTT
TACCGTGGTGGCGGAACAGGATCCGCTGTTTGCGGGCACCGAAATTCATCAGGTG
GCGGCGCGCCGCACCGCGGGC

*Uu*MAT

ATGCAGTATAAAAAAATTATTACCAGCGAAAGCGTGGGCGCGGGCCATCCGGA
TAAAATTTGCGATCAGATTAGCGATGCGATTCTGGATGAATGCCTGAGCCAGGATC
AGAACAGCCGCGTGGCGTGCGAAGTGCTGGCGTGCAACCGCCTGATTGTGATTG
CGGGCGAAATTACCACCCATGCGTATGTGGATGTGGTGAAAACCGCGTGGGAAAT
TATTAAACCGCTGGGCTATGATGAAAACGATTTTACCATTATTAGCAACGTGAACAA
ACAGAGCGTGGATATTGCGCAGAGCGTGGATAAAACCAACAAAAACCTGATTGGC
GCGGGCGATCAGGGCATTGTGTTTGGCTATGCGTGCGATGAAACCCCGCAGTATA
TGCCGCTGACCAGCGTGCTGGCGCATGAACTGCTGAAAGAAATTGAACGCCAGCG
CCGCAGCAAAGAATTTATTAAAATTCAGGCGGATATGAAAAGCCAGGTGAGCATTG
ATTATAGCAACAGCACCCCGCTGATTGAAACCATGCTGGTGAGCATTCAGCATGAT

GAAGATTATGATGTGGAATATTTTAACAAAAAAGTGAGCGCGATTATGGAACAGATT
GCGAAAAAATATAACCTGAACACCAACTTTAAAAAAATTATTAACAGCAGCGGCCG
CTTTGTGATTGGCGGCCCGATTGGCGATACCGGCCTGACCGGCCGCAAAATTATT
GTGGATACCTATGGCGGCGTGGGCCATCATGGCGGCGGCGCGTTTAGCGGCAAA
GATCCGACCAAAGTGGATCGCAGCGCGAGCTATTTTGCGCGCTGGATTGCGAAAA
ACGTGGTGGCGGCGAAACTGGCGAAACAGTGCGAAATTCAGCTGGCGTTTGCGAT
TGGCCAGCCGCAGCCGGTGGCGATGTATGTGAACACCTTTAACACCAACCTGATT
GATGAAACCAAAATTTTTGAAGCGATTAAAAAAAGCTTTAACTTTGATATTAAAACCT
TTATTAACGATCTGAACCTGTGGACCACCAAATATCTGCCGGTGGCGACCTATGGC
CATTTTGGCCGCGATGATCTGGATCTGAGCTGGGAAAAACTGAACAAAGTGGAAG
ATCTGATTAAAAACAGCAAA

*Uu*MAT (I97A)

ATGCAGTATAAAAAAATTATTACCAGCGAAAGCGTGGGCGCGGGCCATCCGGA
TAAAATTTGCGATCAGATTAGCGATGCGATTCTGGATGAATGCCTGAGCCAGGATC
AGAACAGCCGCGTGGCGTGCGAAGTGCTGGCGTGCAACCGCCTGATTGTGATTG
CGGGCGAAATTACCACCCATGCGTATGTGGATGTGGTGAAAACCGCGTGGGAAAT
TATTAAACCGCTGGGCTATGATGAAAACGATTTTACCATTATTAGCAACGTGAACAA
ACAGAGCGTGGATGCGGCGCAGAGCGTGGATAAAACCAACAAAAACCTGATTGGC
GCGGGCGATCAGGGCATTGTGTTTGGCTATGCGTGCGATGAAACCCCGCAGTATA
TGCCGCTGACCAGCGTGCTGGCGCATGAACTGCTGAAAGAAATTGAACGCCAGCG
CCGCAGCAAAGAATTTATTAAAATTCAGGCGGATATGAAAAGCCAGGTGAGCATTG
ATTATAGCAACAGCACCCCGCTGATTGAAACCATGCTGGTGAGCATTCAGCATGAT
GAAGATTATGATGTGGAATATTTTAACAAAAAAGTGAGCGCGATTATGGAACAGATT
GCGAAAAAATATAACCTGAACACCAACTTTAAAAAAATTATTAACAGCAGCGGCCG
CTTTGTGATTGGCGGCCCGATTGGCGATACCGGCCTGACCGGCCGCAAAATTATT
GTGGATACCTATGGCGGCGTGGGCCATCATGGCGGCGGCGCGTTTAGCGGCAAA
GATCCGACCAAAGTGGATCGCAGCGCGAGCTATTTTGCGCGCTGGATTGCGAAAA
ACGTGGTGGCGGCGAAACTGGCGAAACAGTGCGAAATTCAGCTGGCGTTTGCGAT
TGGCCAGCCGCAGCCGGTGGCGATGTATGTGAACACCTTTAACACCAACCTGATT
GATGAAACCAAAATTTTTGAAGCGATTAAAAAAAGCTTTAACTTTGATATTAAAACCT
TTATTAACGATCTGAACCTGTGGACCACCAAATATCTGCCGGTGGCGACCTATGGC
CATTTTGGCCGCGATGATCTGGATCTGAGCTGGGAAAAACTGAACAAAGTGGAAG
ATCTGATTAAAAACAGCAAA

A.2 Protein Sequences

*At*ASMT

MSSDQLSKFLDRNKMEDNKRKVLDEEAKASLDIWKYVFGFADIAAAKCAIDLKIPEA
IENHPSSQPVTLAELSSAVSASPSHLRRIMRFLVHQGIFKEIPTKDGLATGYVNTPLSRR
LMITRRDGKSLAPFVLFETTPEMLAPWLRLSSVVSSPVNGSTPPPFDAVHGKDVWSFA
QDNPFLSDMINEAMACDARRVVPRVAGACHGLFDGVTTMVDVGGGTGETMGMLVKE
FPWIKGFNFDLPHVIEVAEVLDGVENVEGDMFDSIPACDAIFIKWVLHDWGDKDCIKILK
NCKEAVPPNIGKVLIVESVIGENKKTMIVDERDEKLEHVRLMLDMVMMAHTSTGKERTL
KEWDFVLKEAGFARYEVRDIDDVQSLIIAYRSLEHHHHHH

*Cn*ASMT

MNTITLKPGKEKSLLRRHPWIYATGIATTEGRCEPGATVIVRAADGRFLAKAAYSPE
SQIRARAWTFDENEPVDHALFKRRVAAAIAYRRQWVKDSDAVRLIFGESDRLPGLIVD
YYGNGEKGQLVCQFNSAGVEHWKTAIVQALVKETGCPNVYERSDAAVRQREGLELVT
GVLAGAEPDPALSVTEHGVRYYVDVRNGHKTGFYVDQRDNRKLVGDLAVGREVLNC
FCYTGGFSLAALRGGATSVTSIDSSGEALKIAAGNVTLNGFEPERATWLDADVFKTLRE
FRAEGRQFDLIVLDPPKFAPSAQHIDRAARAYKEINLVGMQLLRPGGLLFTYSCSGAIS
MELFQKIVAGAVTDARADARILRRLSAGTDHPMLAAFPEGEYLKGLLLEKVALEHHHHH
H

*Cv*ASMT

MGYAAPQARQSDKKIFDIYFGFLHSYALLFADEVGLFDLLRCEALTLDQVSMATSLP
FRSSQALLSLCASLGLLEKRGERFALSALAEGFLVREAETSFCGVLASARGQAAAFSY
DFFKASLLKGESQLFGGRDLFDNNAQDSEHCEIFTRAMHSKSKGPAQAWVEKIDLSAH
ACLLDVGGGSGVHAISALARWPNLNAVVFDLPPVCAIADTFIERYQMMARAQTHGGDI
WYTDYPFADAHFYSDIFHDWPLERCRFLARKSFDALPSGGRIILHEMLFNTQKTGPRN
VAAYNANMLLWTQGQQLSEPEAADLLQAAGFVEILAFPTGYGDWSLVTGVKPLEHHH
HHH

*Cv*TAm

MQKQRTTSQWRELDAAHHLHPFTDTASLNQAGARVMTRGEGVYLWDSEGNKIID
GMAGLWCVNVGYGRKDFAEAARRQMEELPFYNTFFKTTHPAVVELSSLLAEVTPAGF
DRVFYTNSGSESVDTMIRMVRRYWDVQGKPEKKTLIGRWNGYHGSTIGGASLGGMK
YMHEQGDLPIPGMAHIEQPWWYKHGKDMTPDEFGVVAARWLEEKILEIGADKVAAFV
GEPIQGAGGVIVPPATYWPEIERICRKYDVLLVADEVICGFGRTGEWFGHQHFGFQPD
LFTAAKGLSSGYLPIGAVFVGKRVAEGLIAGGDFNHGFTYSGHPVCAAVAHANVAALR
DEGIVQRVKDDIGPYMQKRWRETFSRFEHVDDVRGVGMVQAFTLVKNKAKRELFPDF
GEIGTLCRDIFFRNNLIMRACGDHIVSAPPLVMTRAEVDEMLAVAERCLEEFEQTLKAR
GLA

*Ec*MAT

MGSSHHHHHHSSGLVPRGSHMAKHLFTSESVSEGHPDKIADQISDAVLDAILEQD
PKARVACETYVKTGMVLVGGEITTSAWVDIEEITRNTVREIGYVHSDMGFDANSCAVLS
AIGKQSPDINQGVDRADPLEQGAGDQGLMFGYATNETDVLMPAPITYAHRLVQRQAE
VRKNGTLPWLRPDAKSQVTFQYDDGKIVGIDAVVLSTQHSEEIDQKSLQEAVMEEIIKPI
LPAEWLTSATKFFINPTGRFVIGGPMGDCGLTGRKIIVDTYGGMARHGGGAFSGKDPS
KVDRSAAYAARYVAKNIVAAGLADRCEIQVSYAIGVAEPTSIMVETFGTEKVPSEQLTLL
VREFFDLRPYGLIQMLDLLHPIYKETAAYGHFGREHFPWEKTDKAQLLRDAAGLK

*Ec*MTAN

MGSSHHHHHHSSGLVPRGSHMKIGIIGAMEEEVTLLRDKIENRQTISLGGCEIYTGQ
LNGTEVALLKSGIGKVAAALGATLLLEHCKPDVIINTGSAGGLAPTLKVGDIVVSDEARY
HDADVTAFGYEYGQLPGCPAGFKADDKLIAAAEACIAELNLNAVRGLIVSGDAFINGSV
GLAKIRHNFPQAIAVEMEATAIAHVCHNFNVPFVVVRAISDVADQQSHLSFDEFLAVAA
KQSSLMVESLVQKLAHG

HLADH

MGSSHHHHHHSSGLVPRGSHMKAVVVNKNSKANIEVVEKELRPLRSGEALVDVE
YCGVCHTDLHVANHDFGNTDGRILGHEGVGIVTKIADDVNSLKIGDRVSIAWMFQSCG
RCEYCVTGRETFCREVKNAGYSVDGGMAEQCIVTADYAVKVPEGLDPAQASSITCAG

VTTYKAIKVSDIKPGQPIVIYGCGGLGNLAIQYAKNVFGAKVIAVDINDDKLALAKEVGAD
MTINPISQGGPADKIVQEELGGAYAAVVTAVSKVAFNSAVDAVRACGKVVAVGLPVETM
DLNIPRLVLDGIEVVGSLVGTRKDLEEAFMFGAEGKVVPVVQTCSLDKVQNVFEEMEQ
GRIQGRMVIDFKKHNCDCK

*Mj*MAT

MRNIIVKKLDVEPIEERPTEIVERKGLGHPDSICDGIAESVSRALCKMYMEKFGTILH
HNTDQVELVGGHAYPKFGGGVMVSPIYILLSGRATMEILDKEKNEVIKLPVGTTAVKAA
KEYLKKVLRNVDVDKDVIIDCRIGQGSMDLVDVFERQKNEVPLANDTSFGVGYAPLSTT
ERLVLETERFLNSDELKNEIPAVGEDIKVMGLREGKKITLTIAMAVVDRYVKNIEEYKEVI
EKVRKKVEDLAKKIADGYEVEIHINTADDYERESVYLTVTGTSAEMGDDGSVGRGNRV
NGLITPFRPMSMEAASGKNPVNHVGKIYNILANLIANDIAKLEGVKECYVRILSQIGKPIN
EPKALDIEIITEDSYDIKDIEPKAKEIANKWLDNIMEVQKMIVEGKVTTF

*Sf*MOMT

MAPSPDHARDLYIELLKKVVSNVIYEDPTHVAGMITDASFDRTSRESGEDYPTVAH
TMIGLKRLDNLHRCLADVVEDGVPGDFIETGVWRGGACIFARGLLNAYGQADRTVWV
ADSFQGFPELTGSDHPLDVEIDLHQYNEAVDLPTSEETVRENFARYGLLDDNVRFLAG
WFKDTMPAAPVKQLAVMRLDGDSYGATMDVLDSLYERLSPGGYVIVDDYCIPACREA
VHDFRDRLGIRDTIHRIDRQGAYWRHSG

*Sf*DMOMT

MAVQKEATLVRQIIRAAGGHAADVRELVAEHGPEAVTAVLVDEIVSRAPHPVNDVP
VLVELAVRSGDALVPRRLAVAQGAPVRRAAPDDDGFVAMRVEYELDELVRELFGPCR
ERAAGTRGTTLFPYATSGTGHIDTYFLAAQQATATVLAGCTSAKPDLNELTSRYLTPK
WGSLHWFTPHYDRHFREYRNEEVRVLEIGIGGYQHPEWGGGSLRMWKHFFHRGLIY
GLDIEDKSHAEEQRITTVVGDQNDPGCLTELAARYGPFDIVIDDGSHINEHVRTSFHALF
PHVRPGGLYVIEDLWTAYWPGFGGDSDPGKSDLTSLGLVKSLVDSLQHQELPEDSGR
SPGYADRHVVGLHVYHNLAFIEKGVNSEGGIPGWIPRDFDALVAASSGGAA

*Tk*MAT

MGSSHHHHHHSSGLVPRGSHMAGKVRNIVVEELVRTPVEMQKVELVERKGIGHP
DSIADGIAEAVSRALSREYVKRYGIILHHNTDQVEVVGGRAYPQFGGGEVIKPIYILLSG
RAVEMVDREFFPVHEIALKAAKDYLRKAVRHLDLEHHVIIDSRIGQGSVDLVGVFNKAK
KNPIPLANDTSFGVGYAPLSETEKIVLETEKYLNSDEFKKKYPAVGEDIKVMGLRKGDEI
DLTIAAAIVDSEVDNPDDYMAVKEAIYEAAKGIVESHTERPTNIYVNTADDPKEGIYYITV
TGTSAEAGDDGSVGRGNRVNGLITPNRHMSMEAAAGKNPVSHVGKIYNILSMLIANDI
AEQVEGVEEVYVRILSQIGKPIDEPLVASVQIIPKKGYSIDVLQKPAYEIADEWLANITKIQ
KMILEDKVNVF

*Rn*COMT (wild-type)

MGSSHHHHHHSSGLVPRGSHMGDTKEQRILRYVQQNAKPGDPQSVLEAIDTYCT
QKEWAMNVGDAKGQIMDAVIREYSPSLVLELGAYCGYSAVRMARLLQPGARLLTMEM
NPDYAAITQQMLNFAGLQDKVTILNGASQDLIPQLKKKYDVDTLDMVFLDHWKDRYLP
DTLLLEKCGLLRKGTVLLADNVIVPGTPDFLAYVRGSSSFECTHYSSYLEYMKVVDGLE
KAIYQGPSSPDKS

*Rn*COMT E199D

MGSSHHHHHHSSGLVPRGSHMGDTKEQRILRYVQQNAKPGDPQSVLEAIDTYCT
QKEWAMNVGDAKGQIMDAVIREYSPSLVLELGAYCGYSAVRMARLLQPGARLLTMEM
NPDYAAITQQMLNFAGLQDKVTILNGASQDLIPQLKKKYDVDTLDMVFLDHWKDRYLP
DTLLLEKCGLLRKGTVLLADNVIVPGTPDFLAYVRGSSSFECTHYSSYLDYMKVVDGLE
KAIYQGPSSPDKS

*Rn*COMT (M40A)

MGSSHHHHHHSSGLVPRGSHMGDTKEQRILRYVQQNAKPGDPQSVLEAIDTYCT
QKEWAANVGDAKGQIMDAVIREYSPSLVLELGAYCGYSAVRMARLLQPGARLLTMEM
NPDYAAITQQMLNFAGLQDKVTILNGASQDLIPQLKKKYDVDTLDMVFLDHWKDRYLP
DTLLLEKCGLLRKGTVLLADNVIVPGTPDFLAYVRGSSSFECTHYSSYLEYMKVVDGLE
KAIYQGPSSPDKS

*Rn*COMT (E199D, M40A)

MGSSHHHHHHSSGLVPRGSHMGDTKEQRILRYVQQNAKPGDPQSVLEAIDTYCT
QKEWAANVGDAKGQIMDAVIREYSPSLVLELGAYCGYSAVRMARLLQPGARLLTMEM
NPDYAAITQQMLNFAGLQDKVTILNGASQDLIPQLKKKYDVDTLDMVFLDHWKDRYLP
DTLLLEKCGLLRKGTVLLADNVIVPGTPDFLAYVRGSSSFECTHYSSYLDYMKVVDGLE
KAIYQGPSSPDKS

Δ29*Tf*NCS

MLHHQGIINQVSTVTKVIHHELEVAASADDIWTVYSWPGLAKHLPDLLPGAFEKLEII
GDGGVGTILDMTFVPGEFPHEYKEKFILVDNEHRLKKVQMIEGGYLDLGVTYYMDTIHV
VPTGKDSCVIKSSTEYHVKPEFVKIVEPLITTGPLAAMADAISKLVLEHKSKSNSDEIEAA
IITVLEHHHHHH

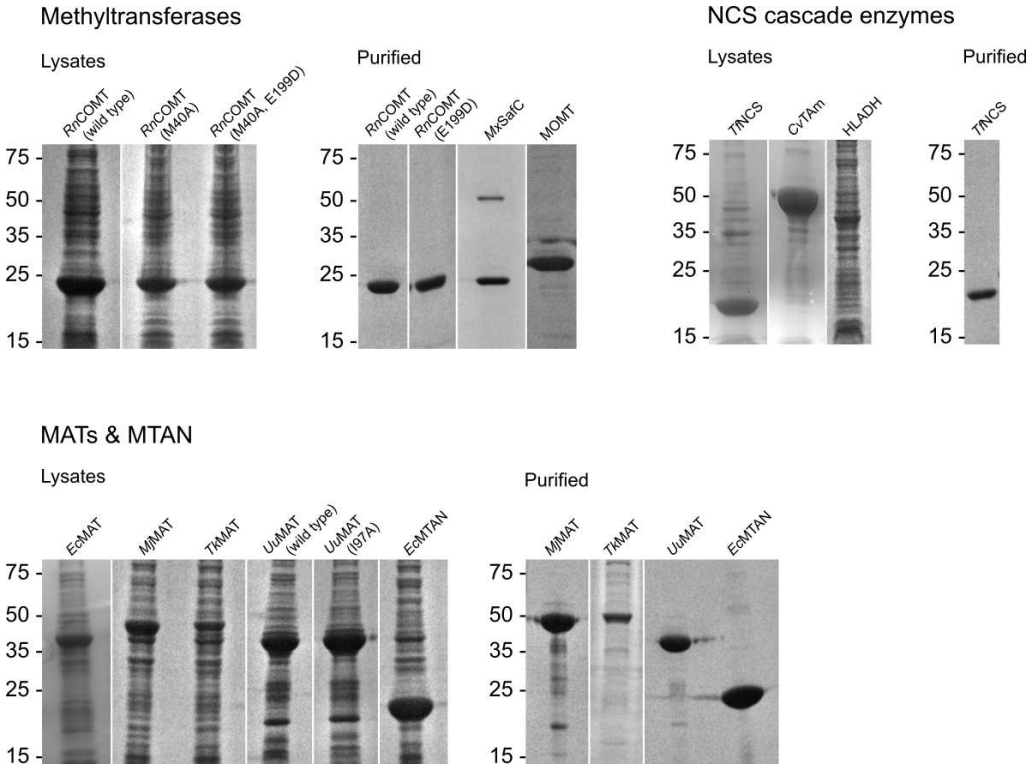
*Uu*MAT

MQYKKIITSESVGAGHPDKICDQISDAILDECLSQDQNSRVACEVLACNRLIVIAGEI
TTHAYVDVVKTAWEIIKPLGYDENDFTIISNVNKQSVDIAQSVDKTNKNLIGAGDQGIVF
GYACDETPQYMPLTSVLAHELLKEIERQRRSKEFIKIQADMKSQVSIDYSNSTPLIETML
VSIQHDEDYDVEYFNKKVSAIMEQIAKKYNLNTNFKKIINSSGRFVIGGPIGDTGLTGRKI
IVDTYGGVGHHGGGAFSGKDPTKVDRSASYFARWIAKNVVAAKLAKQCEIQLAFAIGQ
PQPVAMYVNTFNTNLIDETKIFEAIKKSFNFDIKTFINDLNLWTTKYLPVATYGHFGRDDL
DLSWEKLNKVEDLIKNSK

*Uu*MAT (I97A)

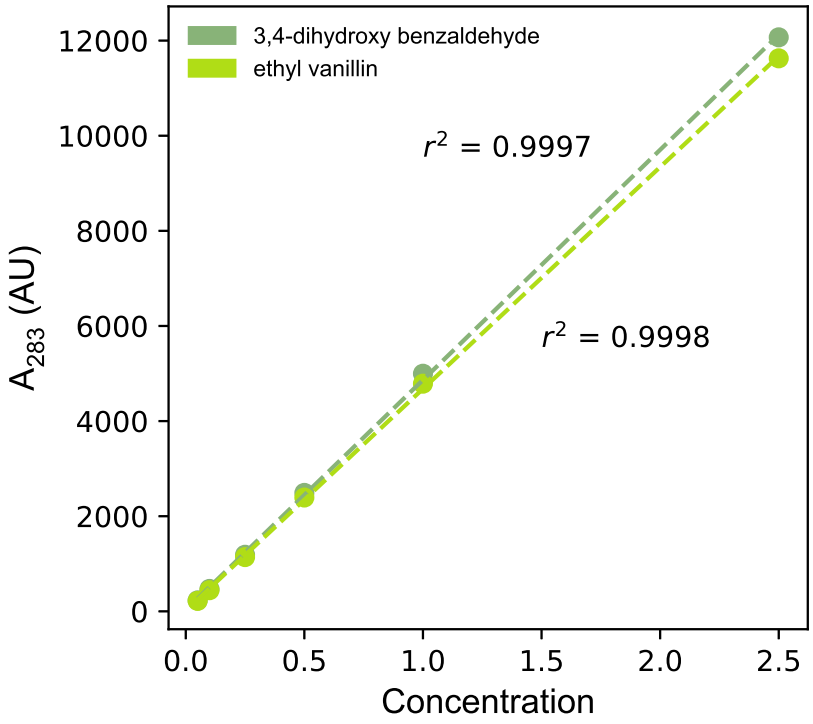
MQYKKIITSESVGAGHPDKICDQISDAILDECLSQDQNSRVACEVLACNRLIVIAGEI
TTHAYVDVVKTAWEIIKPLGYDENDFTIISNVNKQSVDAAQSVDKTNKNLIGAGDQGIVF
GYACDETPQYMPLTSVLAHELLKEIERQRRSKEFIKIQADMKSQVSIDYSNSTPLIETML
VSIQHDEDYDVEYFNKKVSAIMEQIAKKYNLNTNFKKIINSSGRFVIGGPIGDTGLTGRKI
IVDTYGGVGHHGGGAFSGKDPTKVDRSASYFARWIAKNVVAAKLAKQCEIQLAFAIGQ
PQPVAMYVNTFNTNLIDETKIFEAIKKSFNFDIKTFINDLNLWTTKYLPVATYGHFGRDDL
DLSWEKLNKVEDLIKNSK

Appendix B: SDS-PAGE gels



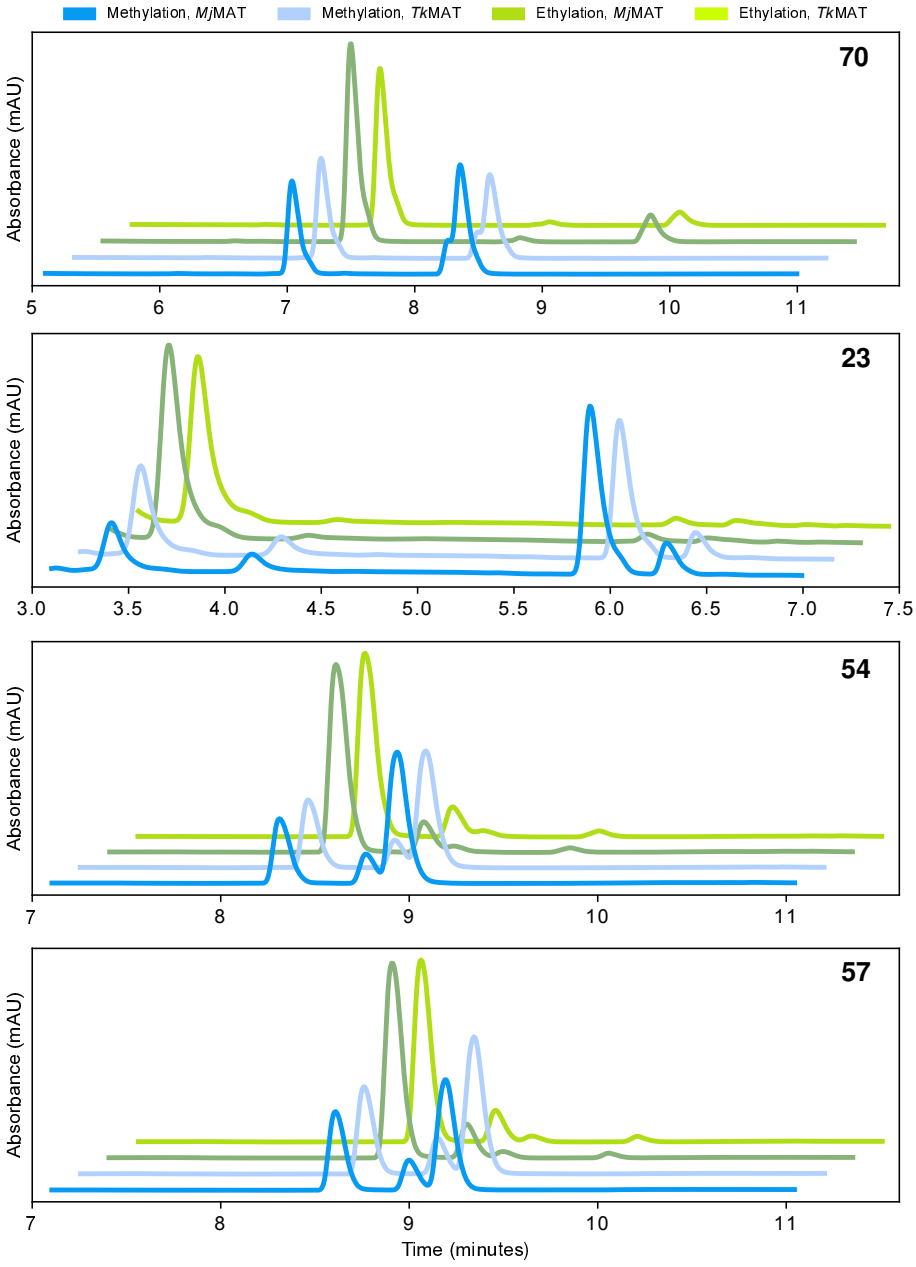
Supplementary Figure B.1: SDS-PAGE of enzymes, visualised with Instant*Blue*™ Coomassie stain according to the manufacturer's instructions. Marker sizes in kDa shown on left side (Promega Broad Range Molecular Weight Markers). Images from different gels are aligned by their markers and demarcated by white lines.

Appendix C: Calibration curves



Supplementary Figure C.1: Absorbance at 283 nm for various concentrations of dihydroxybenzaldehyde **70** and ethyl-vanillin, recorded by analytical HPLC.

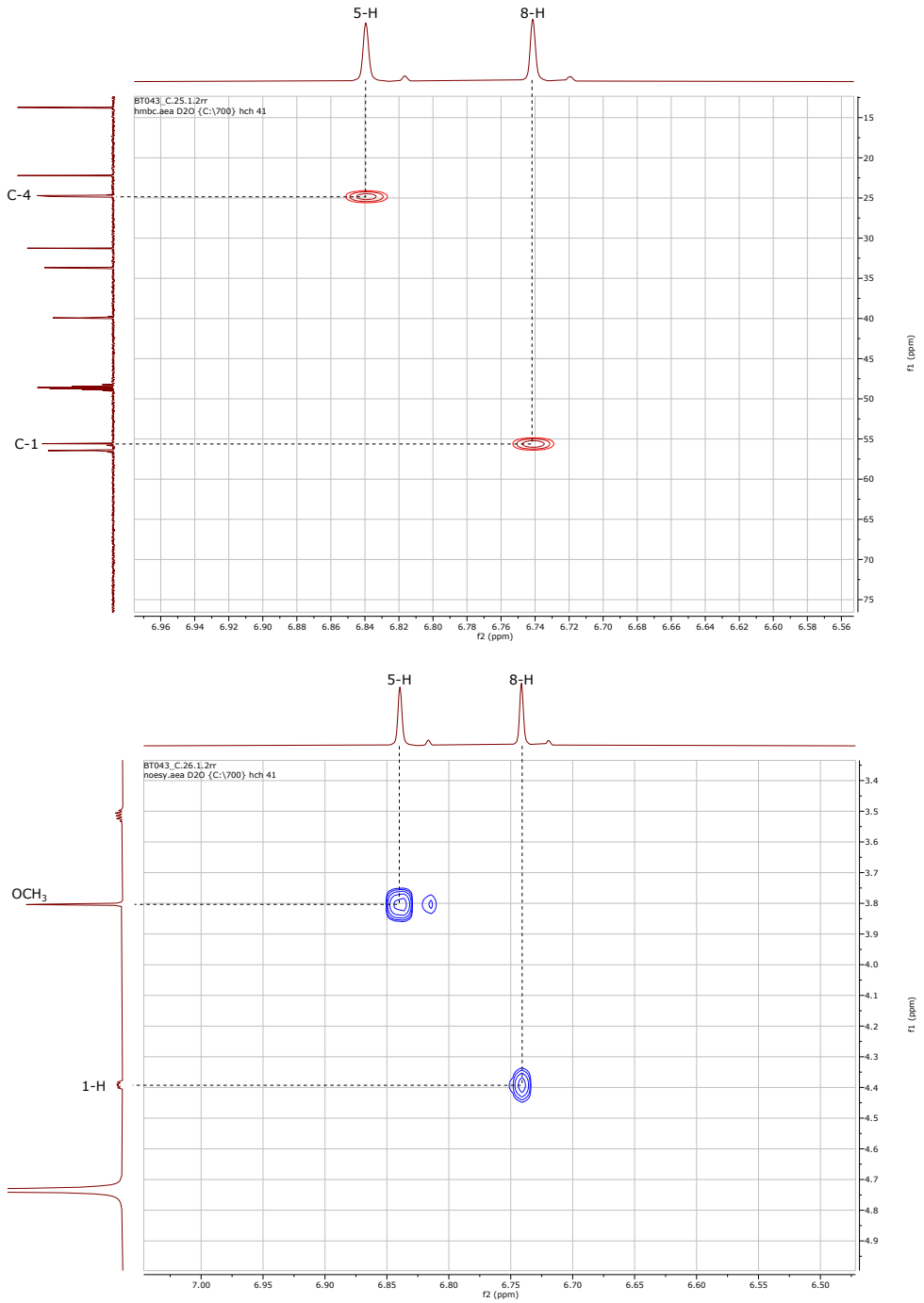
Appendix D: HPLC Traces



Supplementary Figure D.1: Representative analytical HPLC traces for initial ethylation assays. Substrate compound given in top right of each trace.

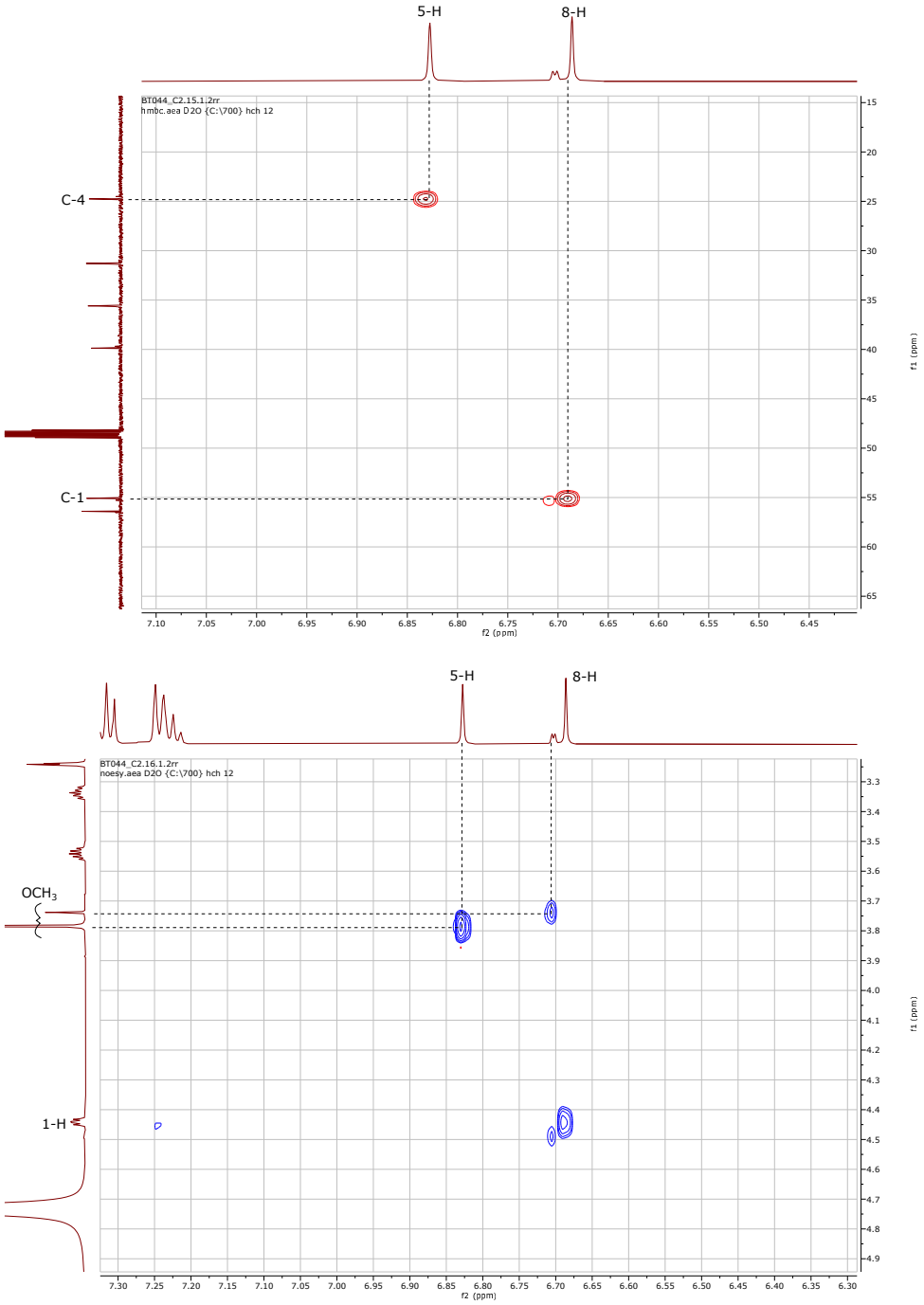
Appendix E: NMR Spectra

E.1 Evidence for selectivity of *Rn*COMT towards (*S*)-54



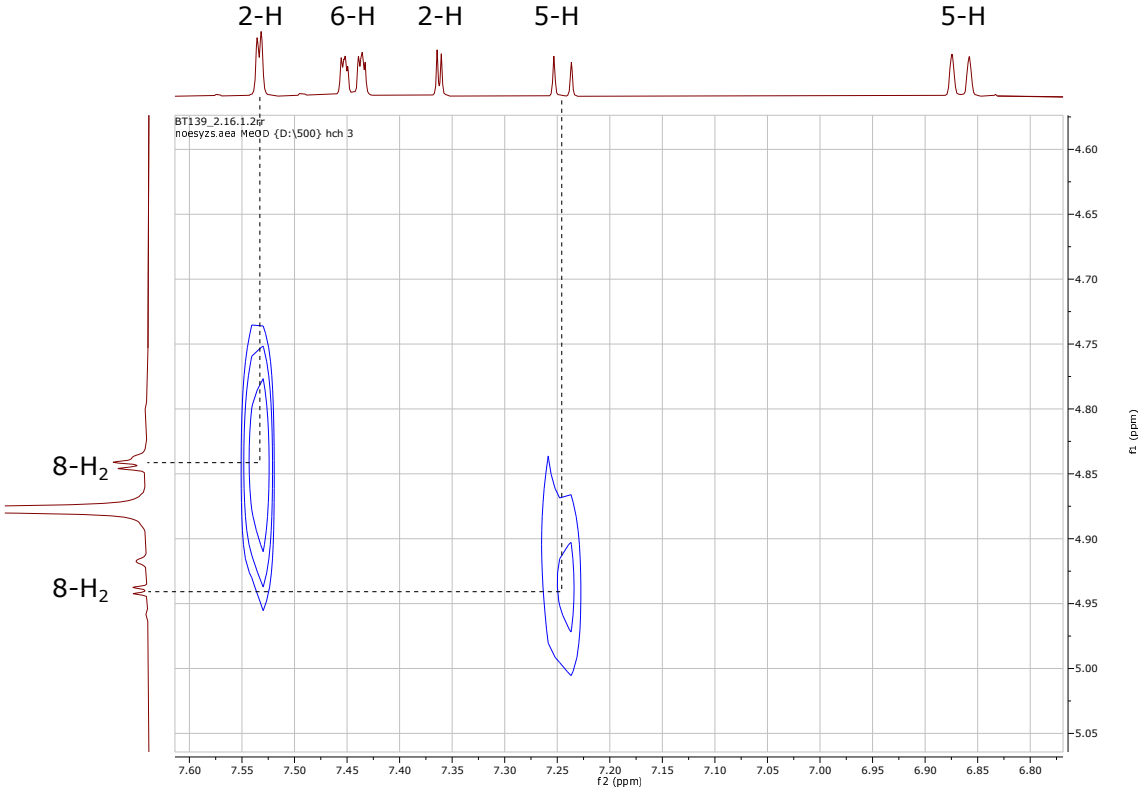
Supplementary Figure E.1: Top: HMBC basis for assignments of aryl proton signals. Bottom: NOESY analysis showing methylation of 6' hydroxyl.

E.2 Evidence for selectivity of *Rn*COMT towards (*S*)-57



Supplementary Figure E.2: Top: HMBC basis for assignments of aryl proton signals. Bottom: NOESY analysis showing methylation of 6' and 7' hydroxyl.

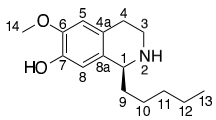
E.3 Evidence for the presence of regioisomers in the enzymatic product 93



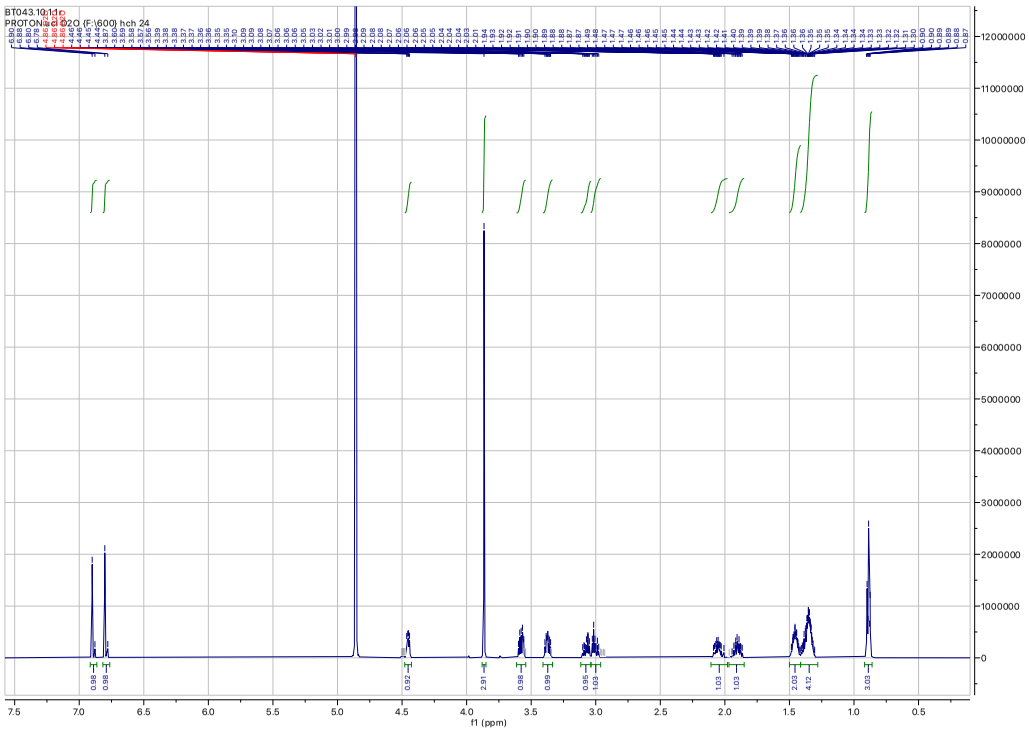
Supplementary Figure E.3: NOESY correlations

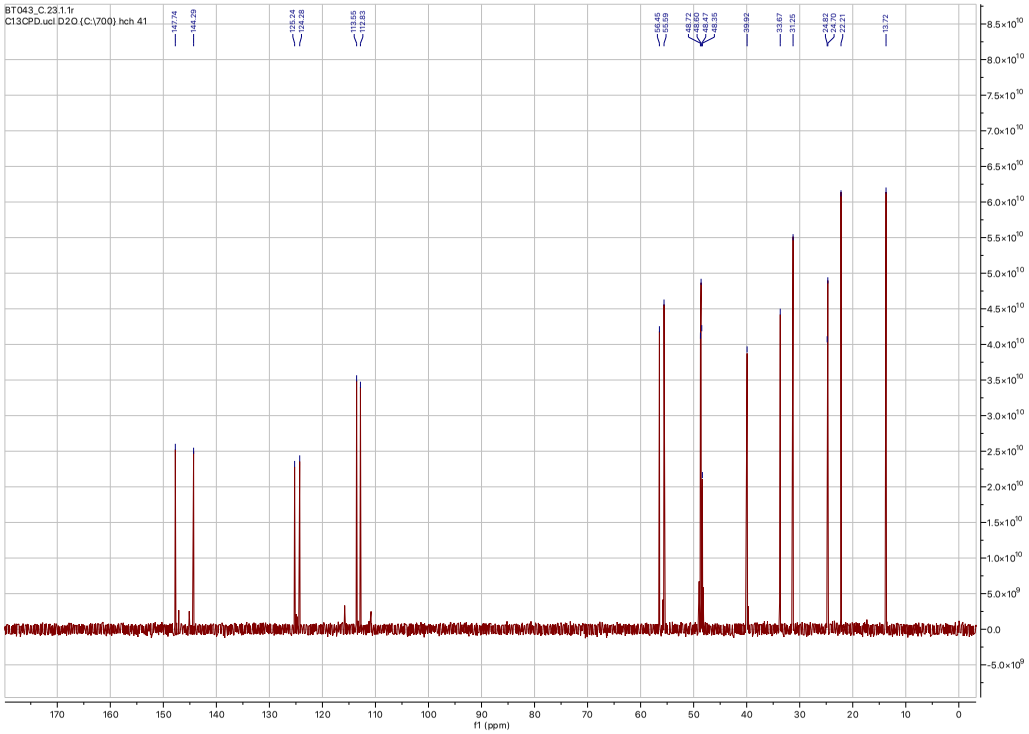
E.4 THIQs

55

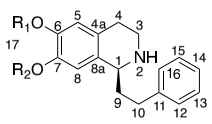


Spectrum	Solvent	Field (MHz)
^{1}H	D_2O	700
^{13}C	$D_2O + CD_3OD$	700



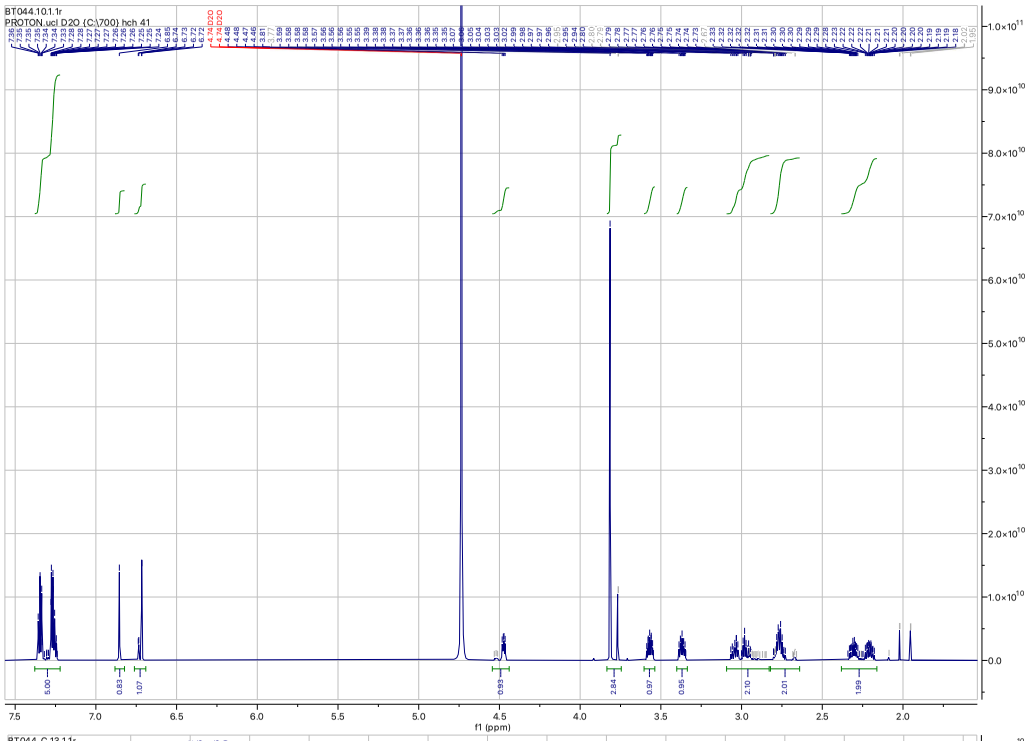


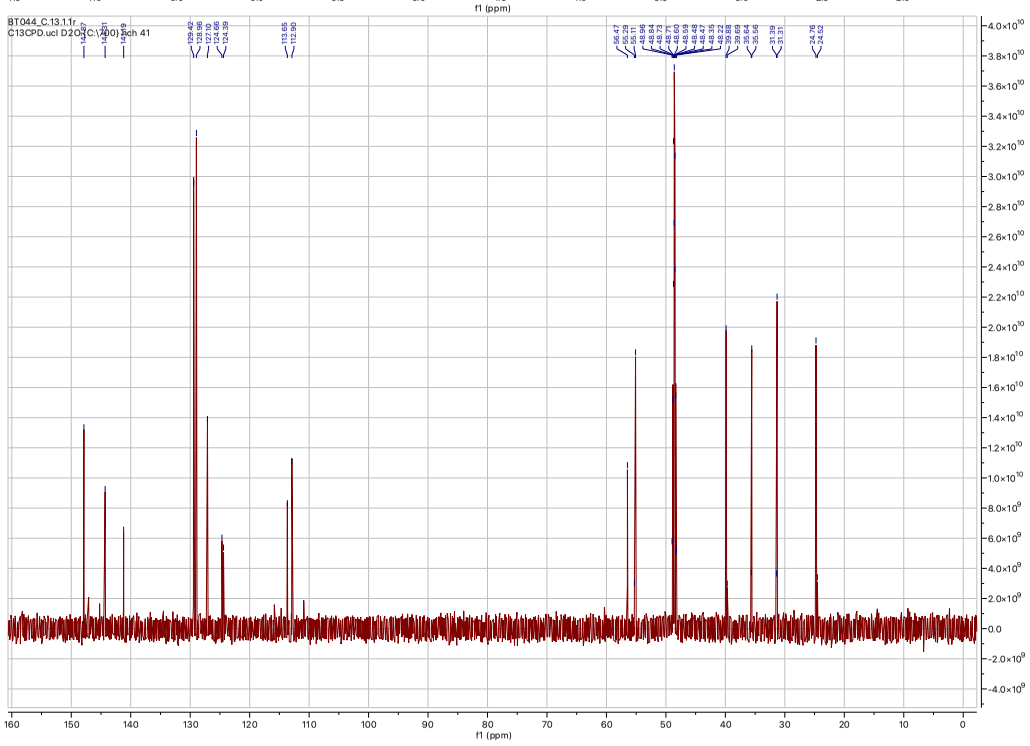
58



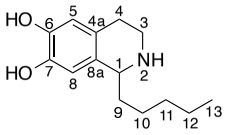
R_1 = CH_3 / H
R_2 = H / CH_3

Spectrum	Solvent	Field (MHz)
^1H	D_2O	700
^{13}C	D_2O + CD_3OD	700

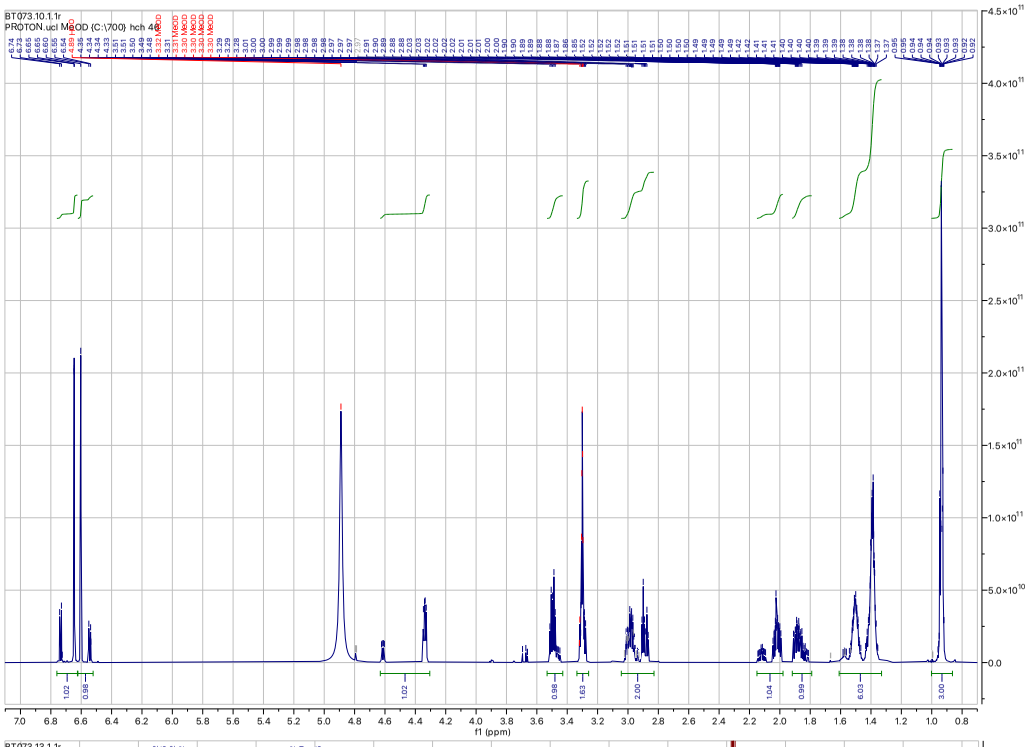


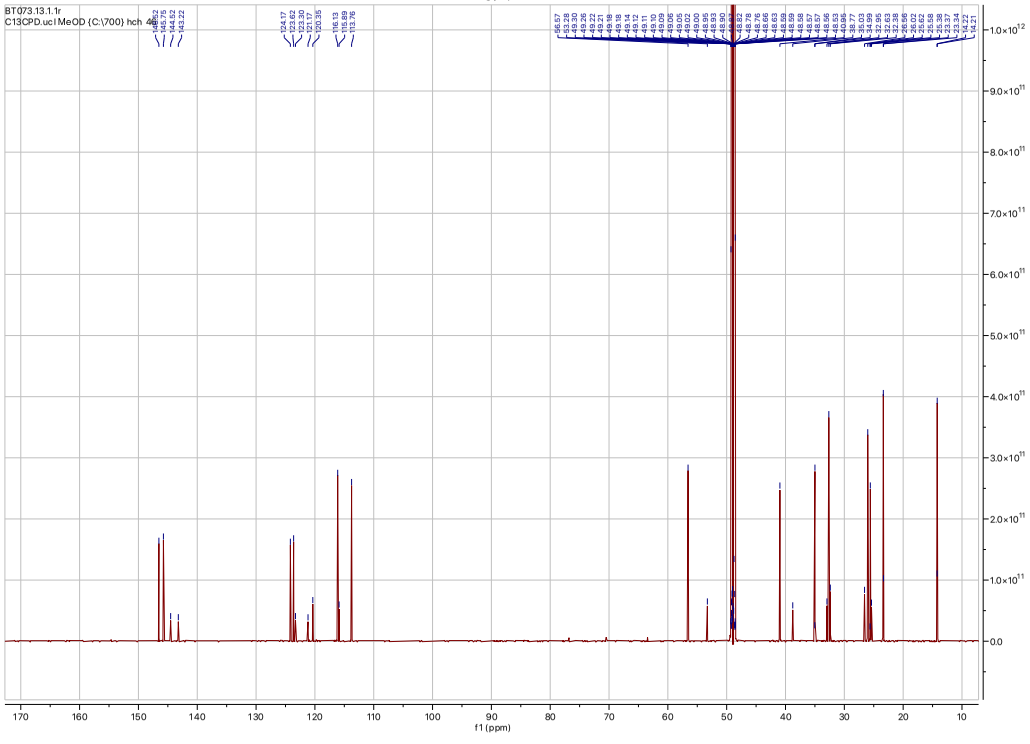


54

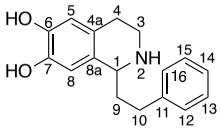


Spectrum	Solvent	Field (MHz)
^{1}H	CD_3OD	700
^{13}C	CD_3OD	700

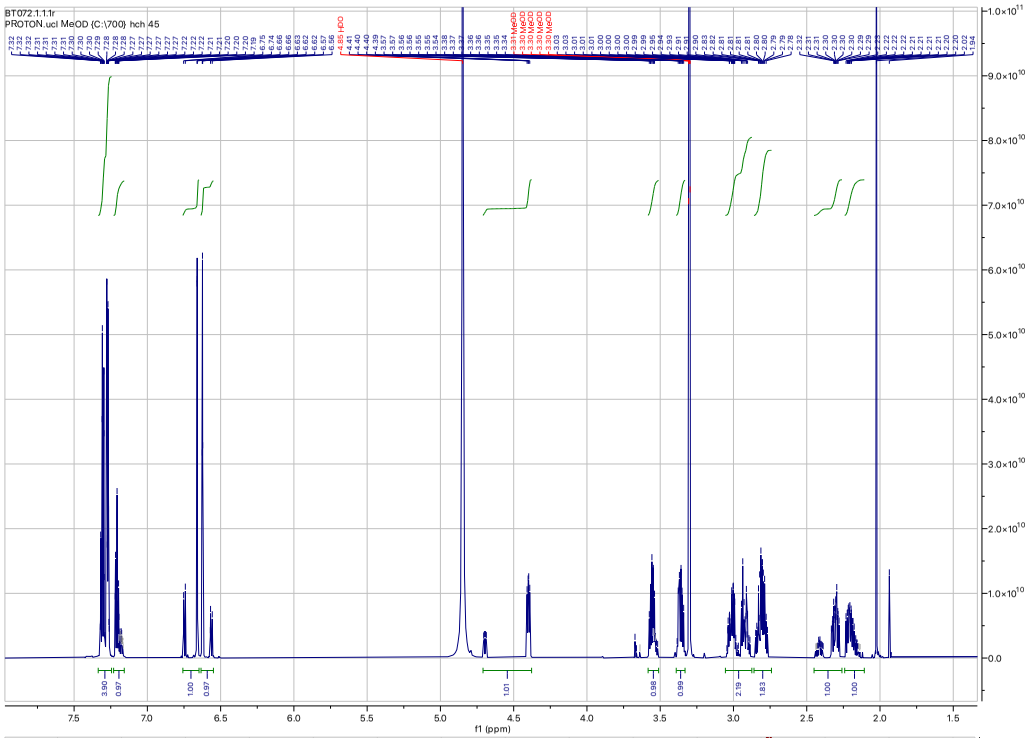


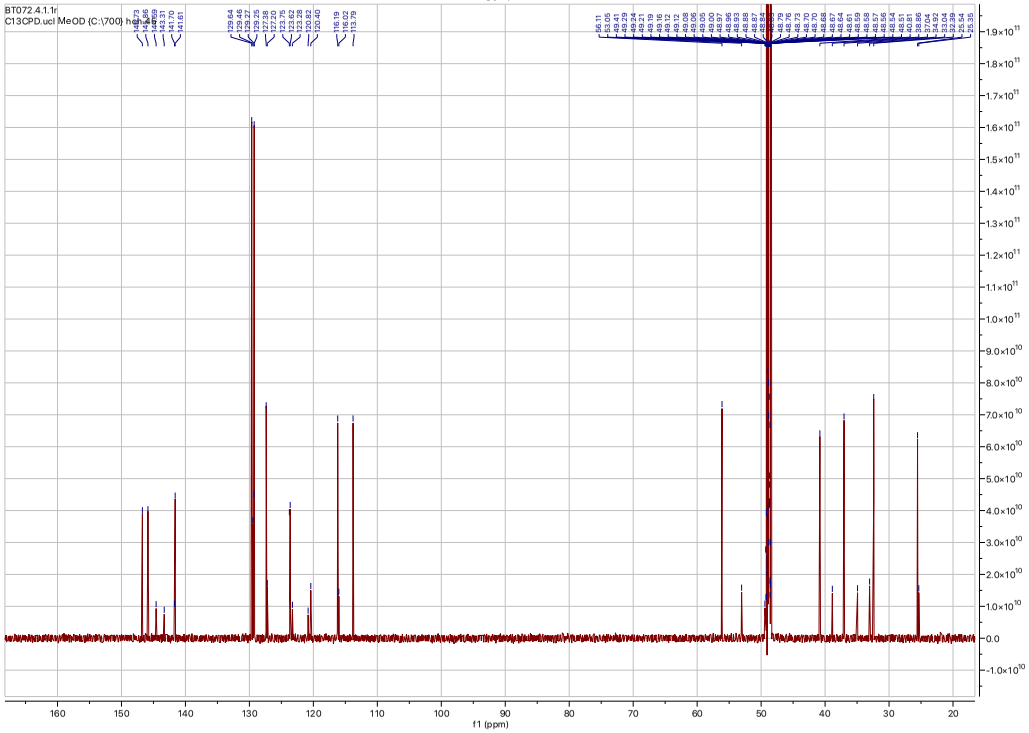


57



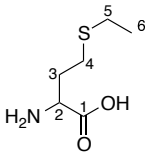
Spectrum	Solvent	Field (MHz)
^{1}H	CD_3OD	700
^{13}C	CD_3OD	700



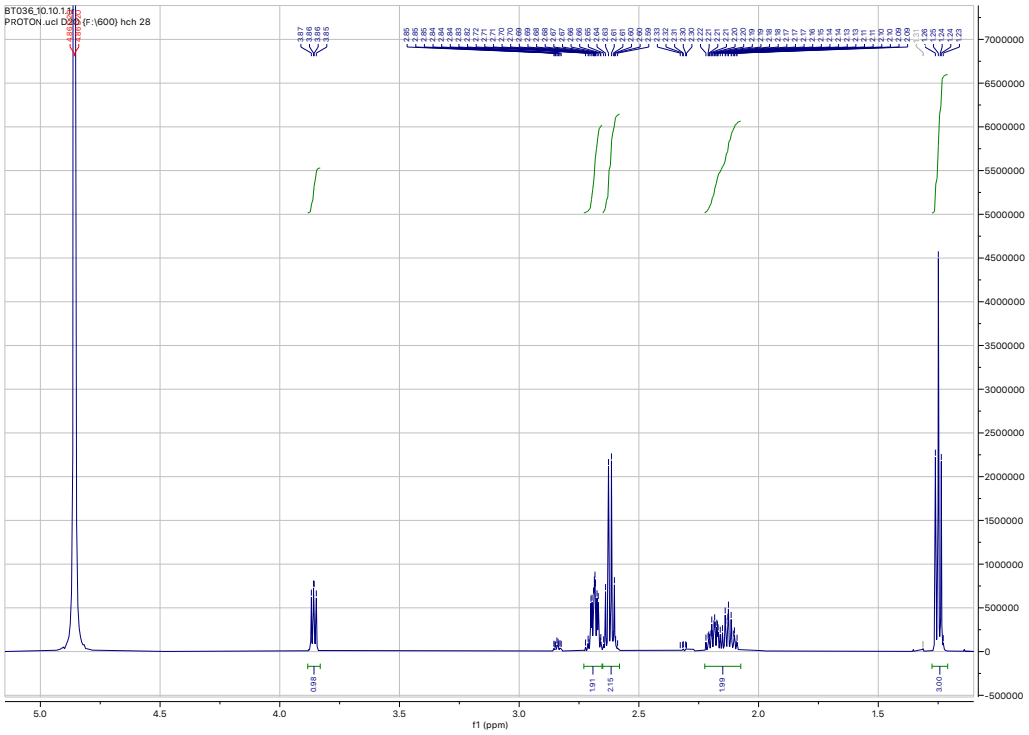


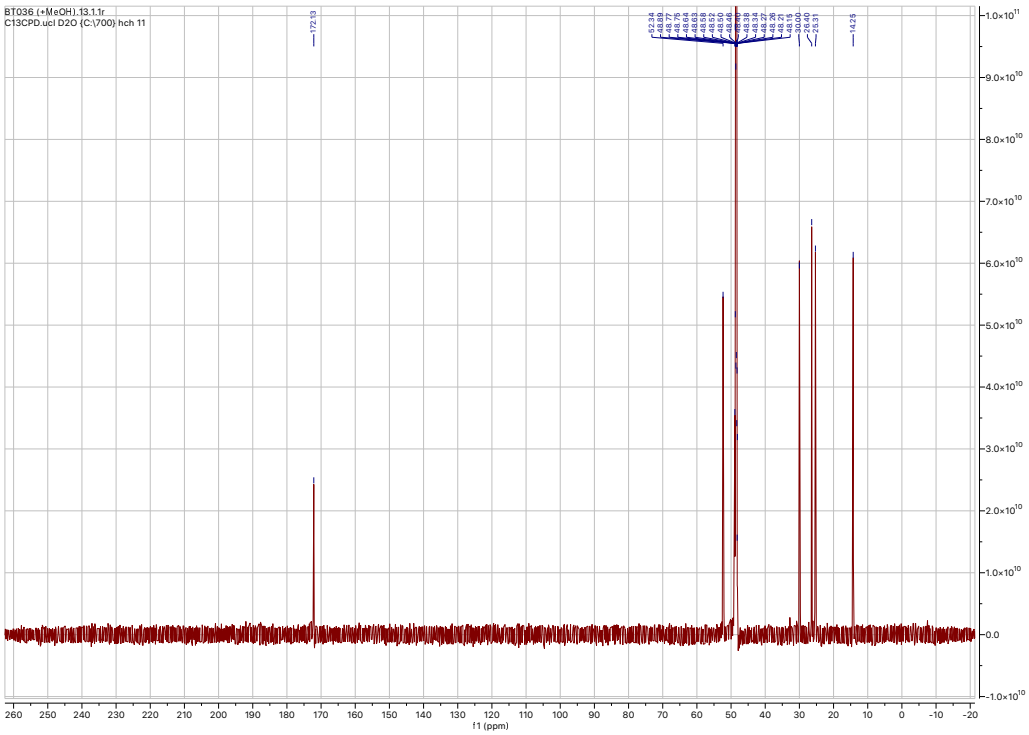
E.5 Methionine Analogues

32

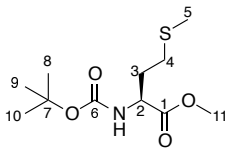


Spectrum	Solvent	Field (MHz)
^{1}H	D_2O	600
^{13}C	$D_2O + CD_3OD$	700

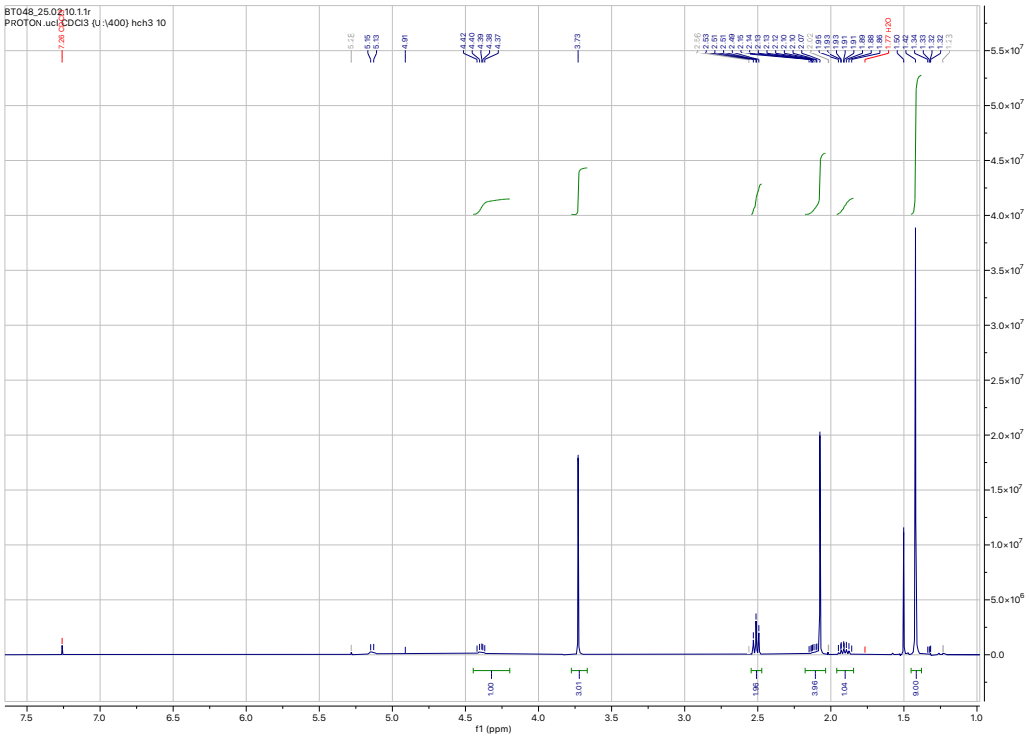


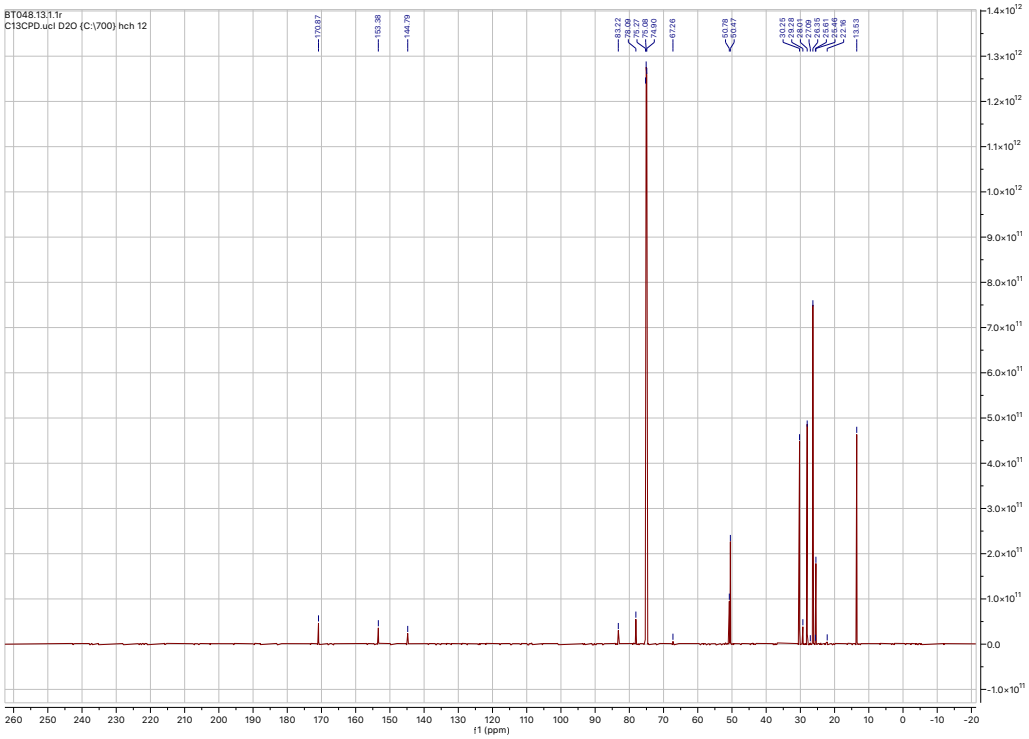


83

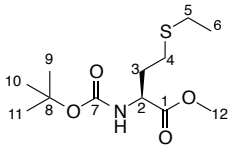


Spectrum	Solvent	Field (MHz)
^{1}H	$CDCl_3$	400
^{13}C	$CDCl_3$	700

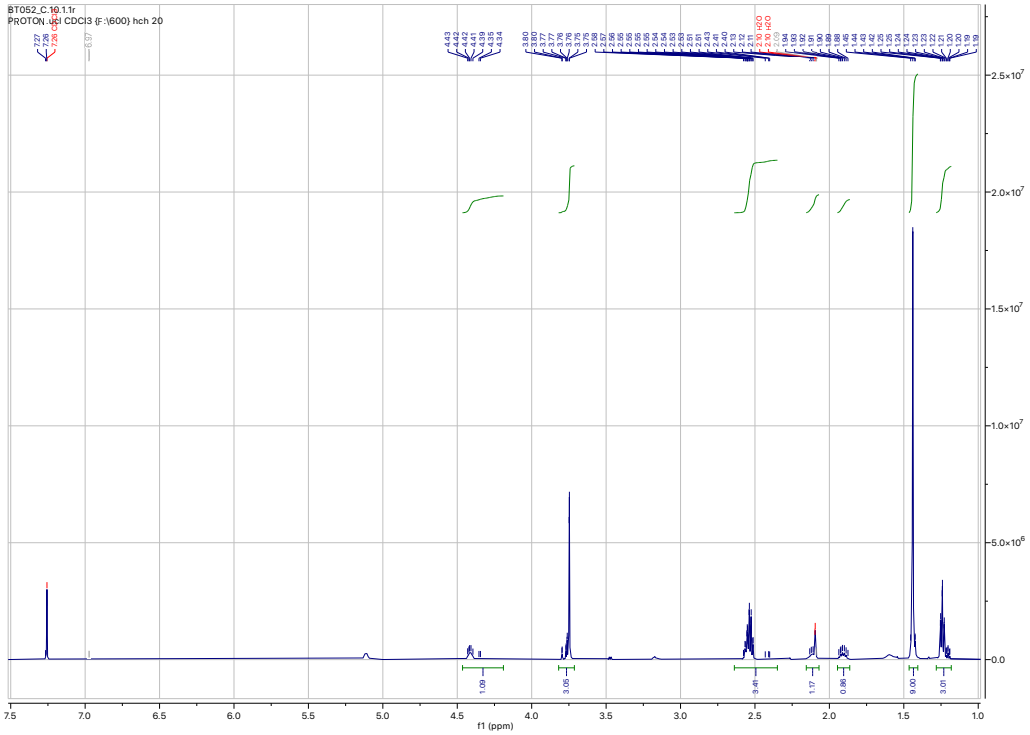


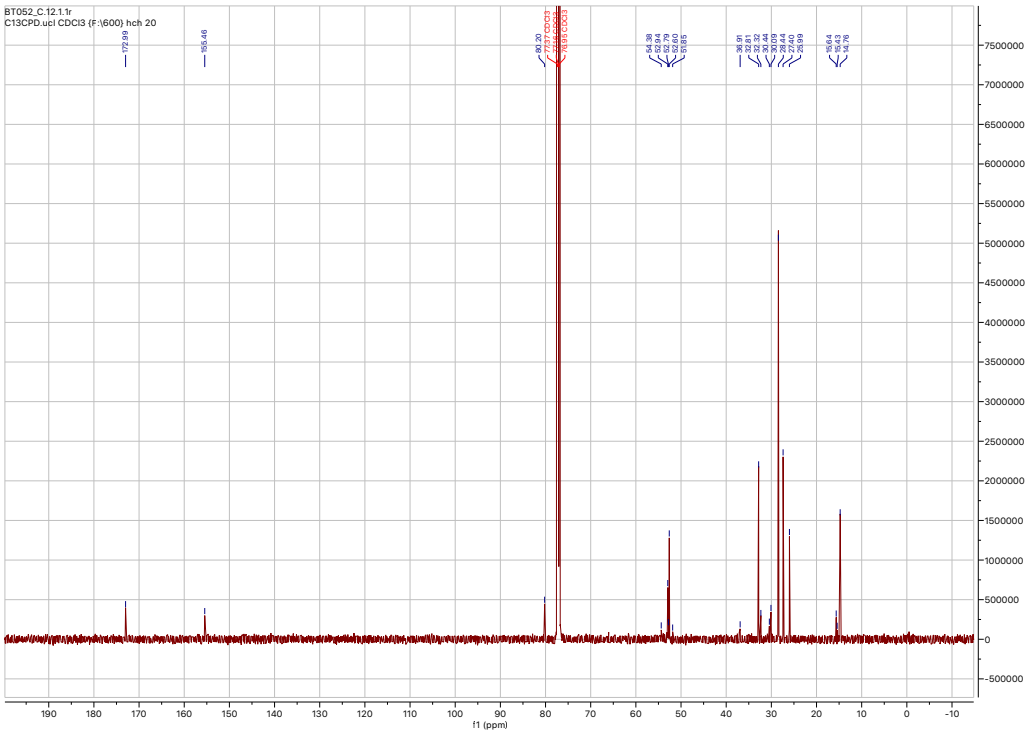


84

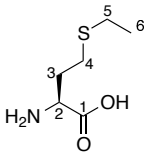


Spectrum	Solvent	Field (MHz)
^{1}H	$CDCl_3$	600
^{13}C	$CDCl_3$	600

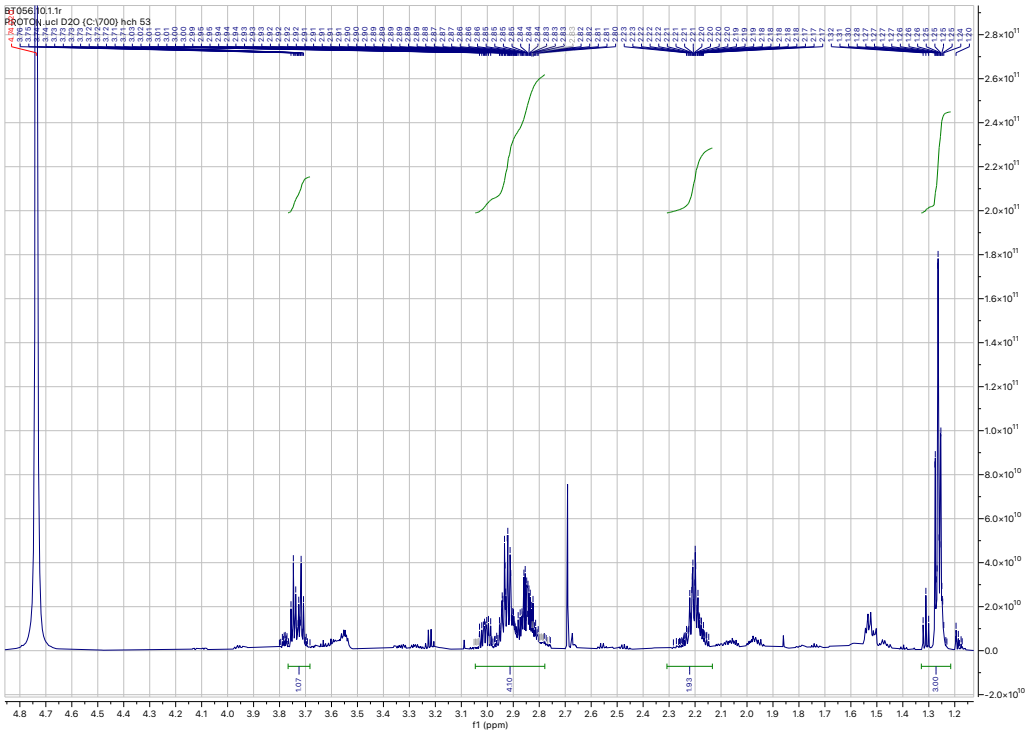


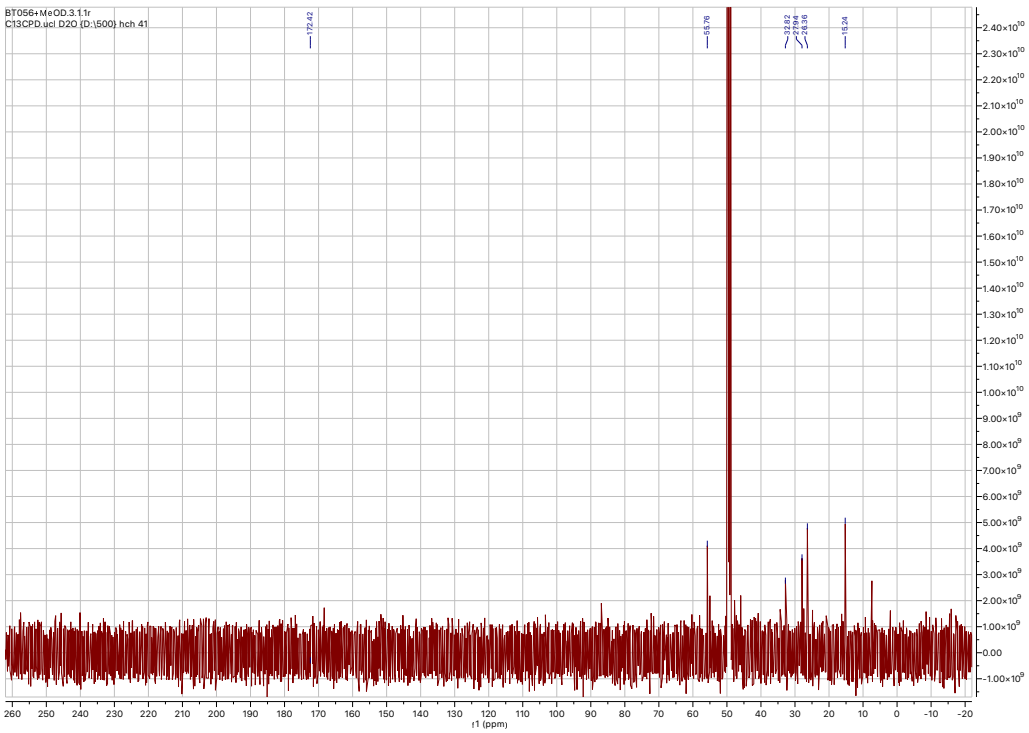


32a

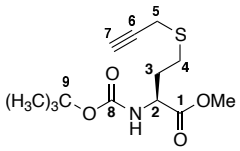


Spectrum	Solvent	Field (MHz)
^{1}H	D_2O	700
^{13}C	D_2O + CD_3OD	500

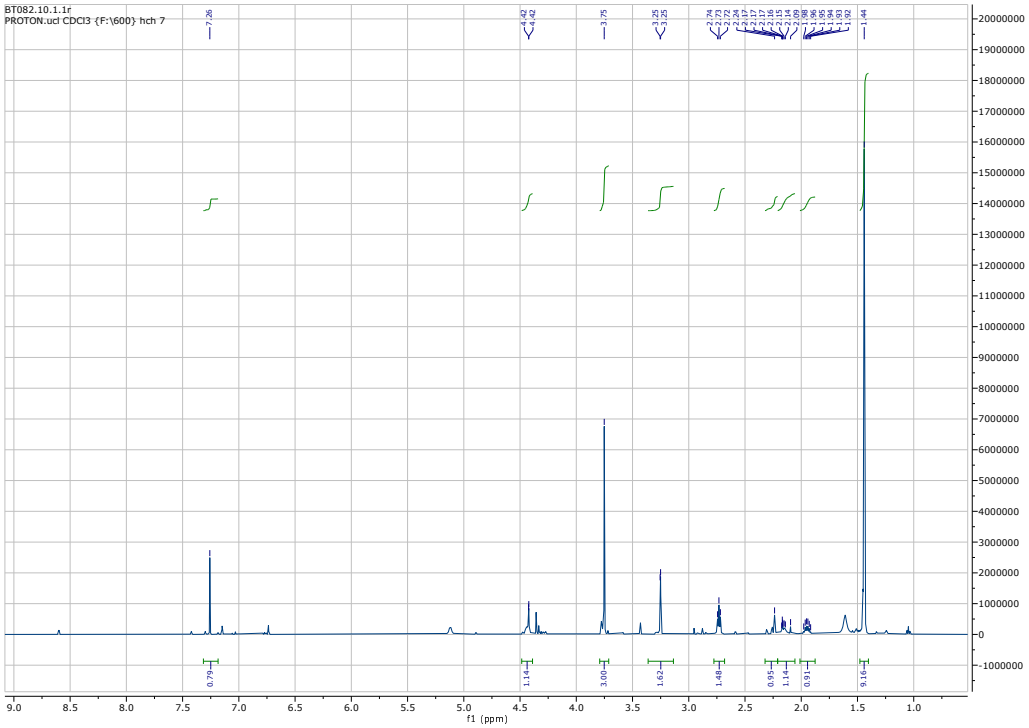


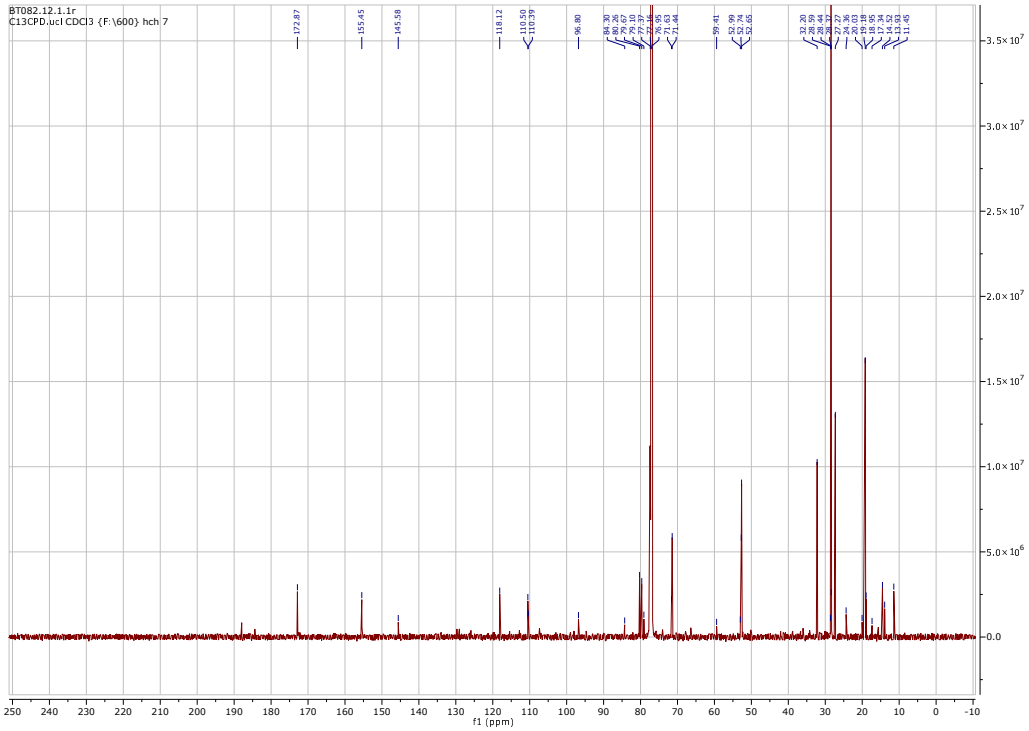


113

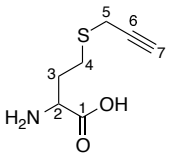


Spectrum	Solvent	Field (MHz)
^{1}H	$CDCl_3$	600
^{13}C	$CDCl_3$	600

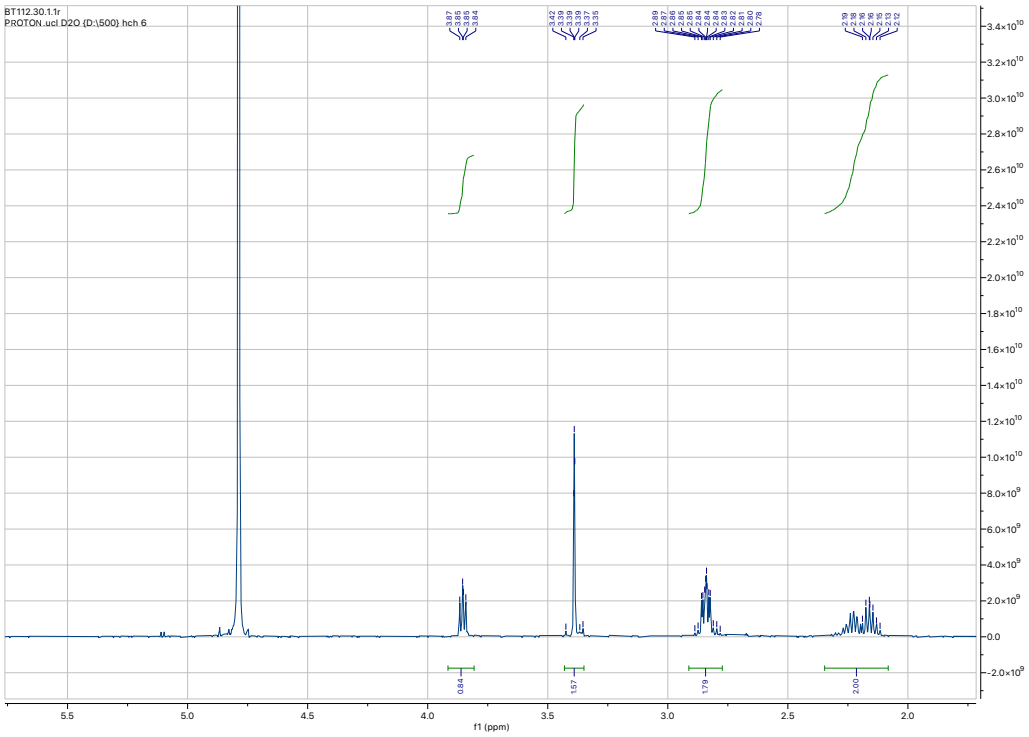


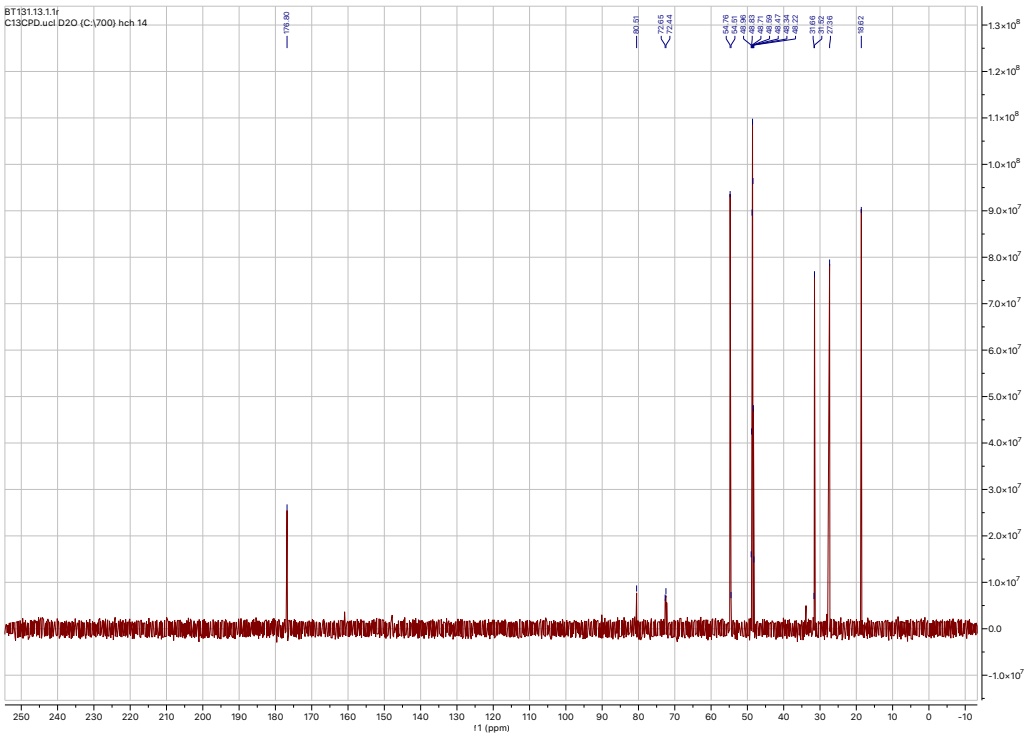


86

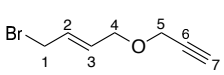


Spectrum	Solvent	Field (MHz)
^{1}H	D_2O	500
^{13}C	D_2O + CD_3OD	700

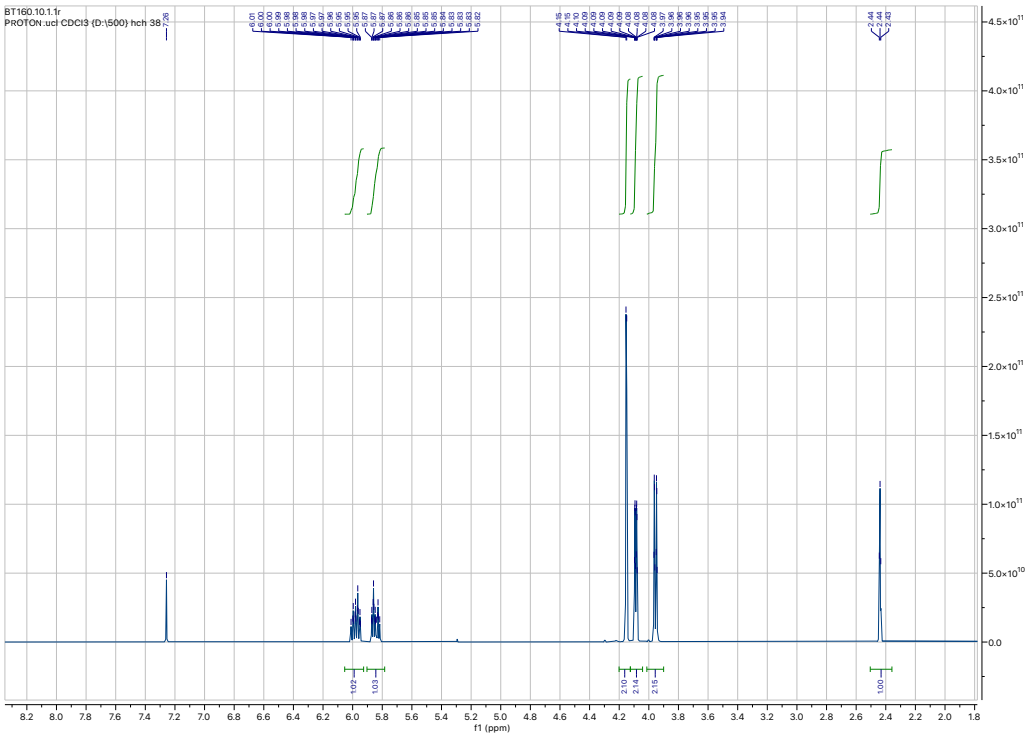


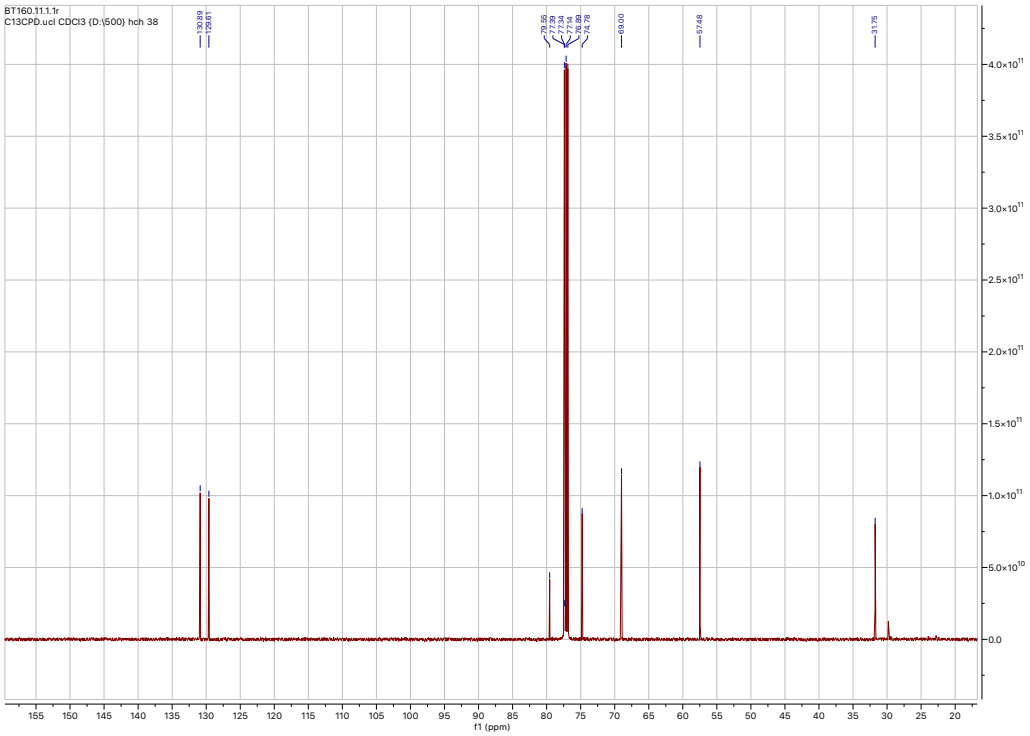


104

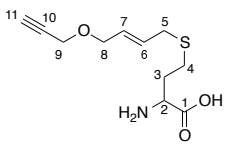


Spectrum	Solvent	Field (MHz)
^{1}H	$CDCl_3$	500
^{13}C	$CDCl_3$	500

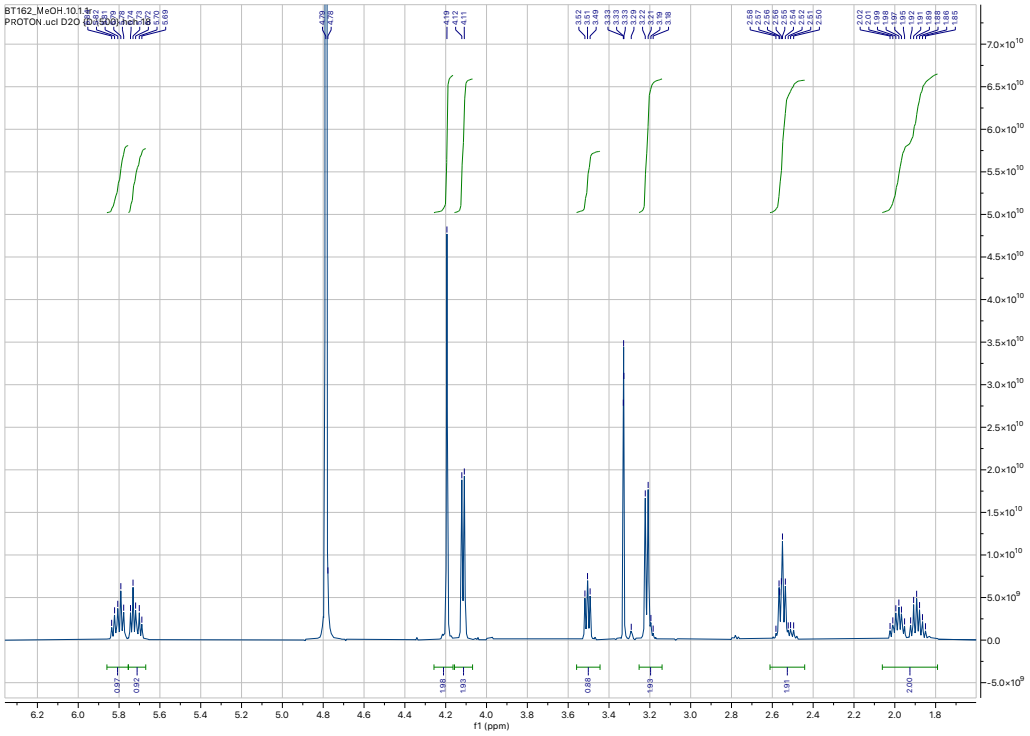


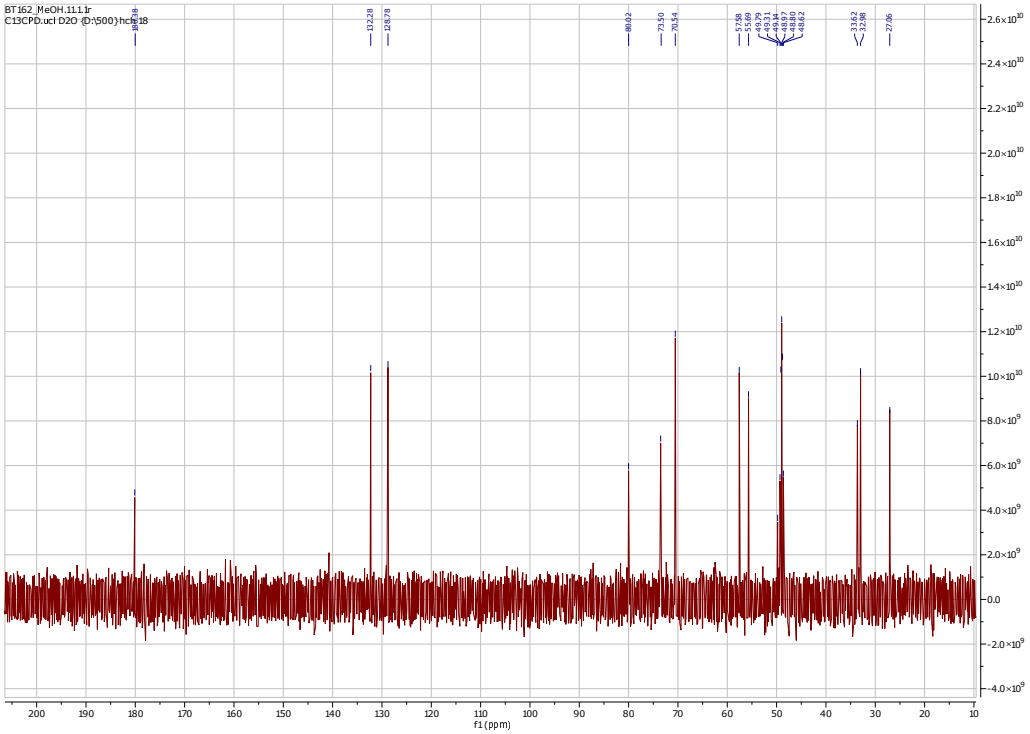


105



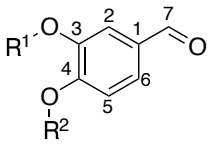
Spectrum	Solvent	Field (MHz)
^{1}H	$D_2O + CD_3OD$	500
^{13}C	$D_2O + CD_3OD$	500

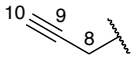
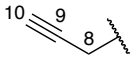




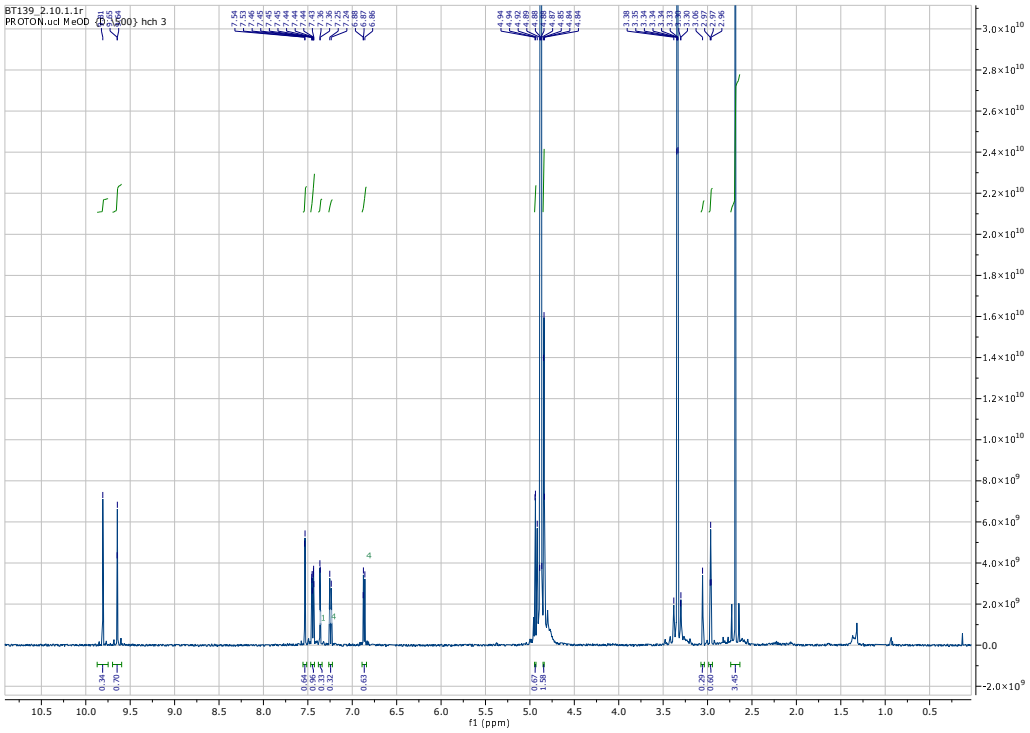
E.6 Other enzyme products

93

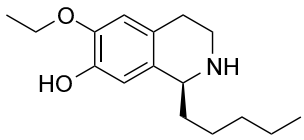


	R^1	R^2
93a		OH
93b	OH	

Spectrum	Solvent	Field (MHz)
^1H	CD_3OD	500

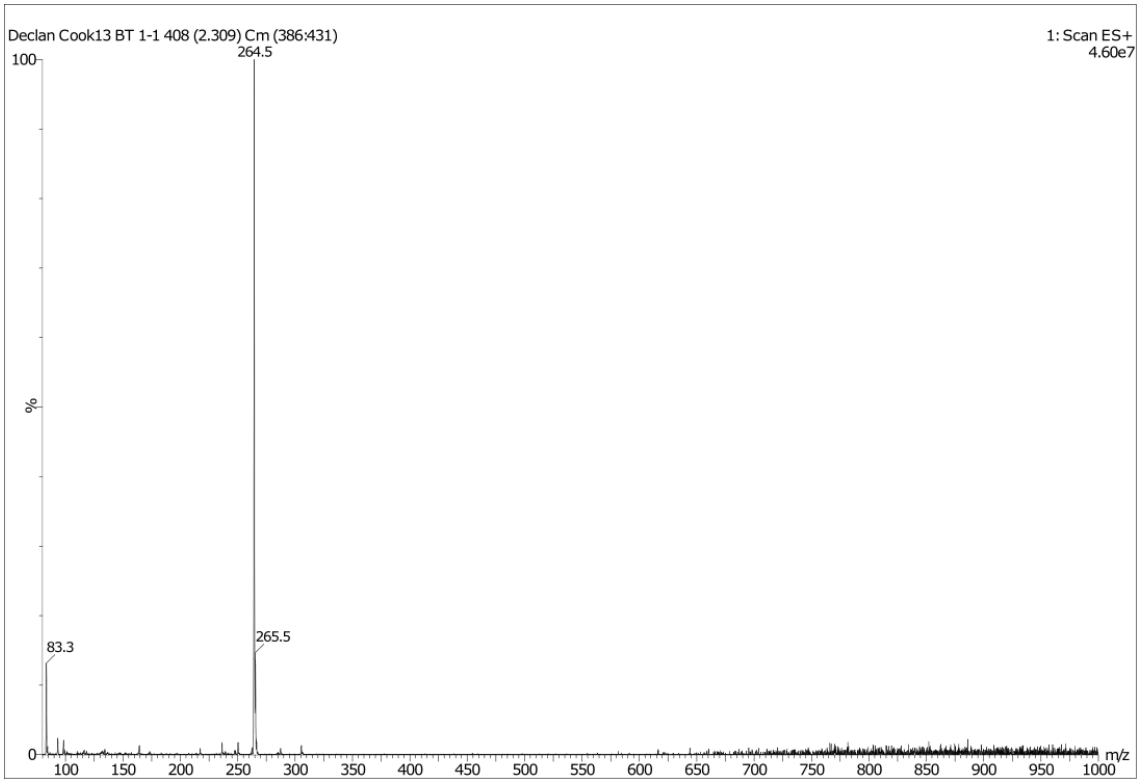


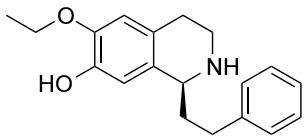
Appendix F: Mass Spectra



M+H = 264

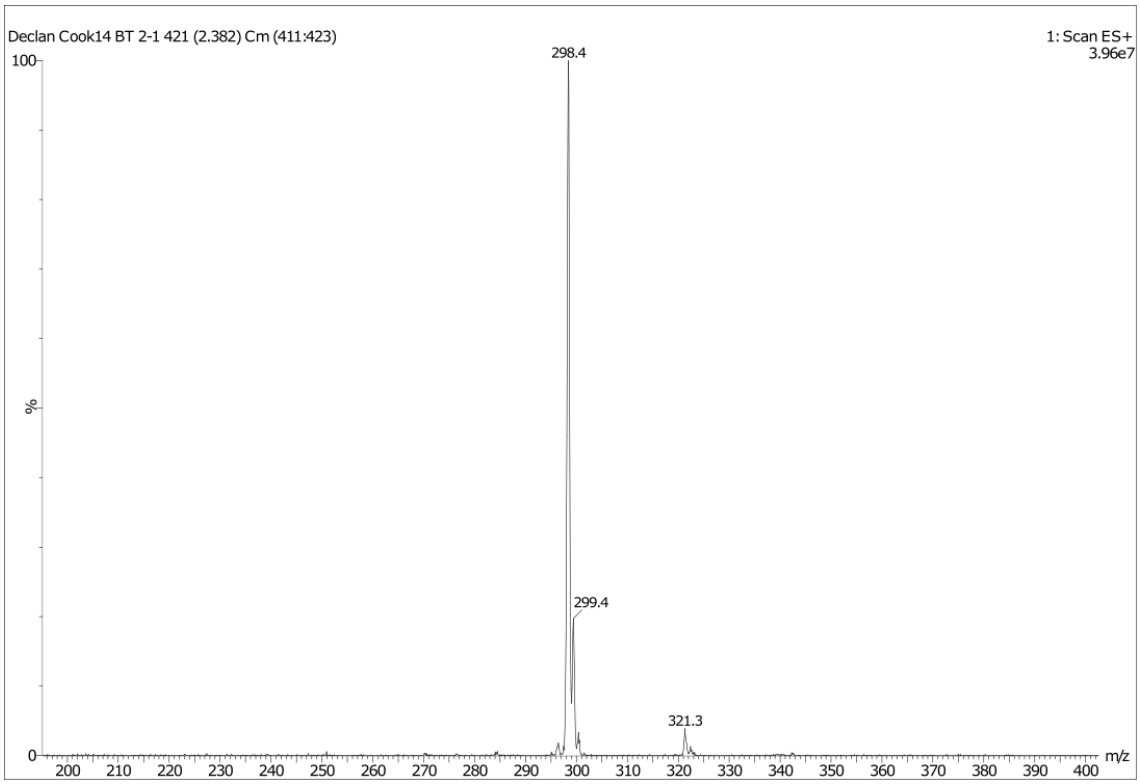
73

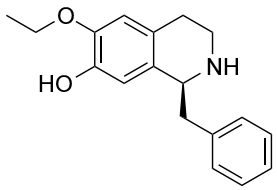



O
HO
NH

M+H = 298

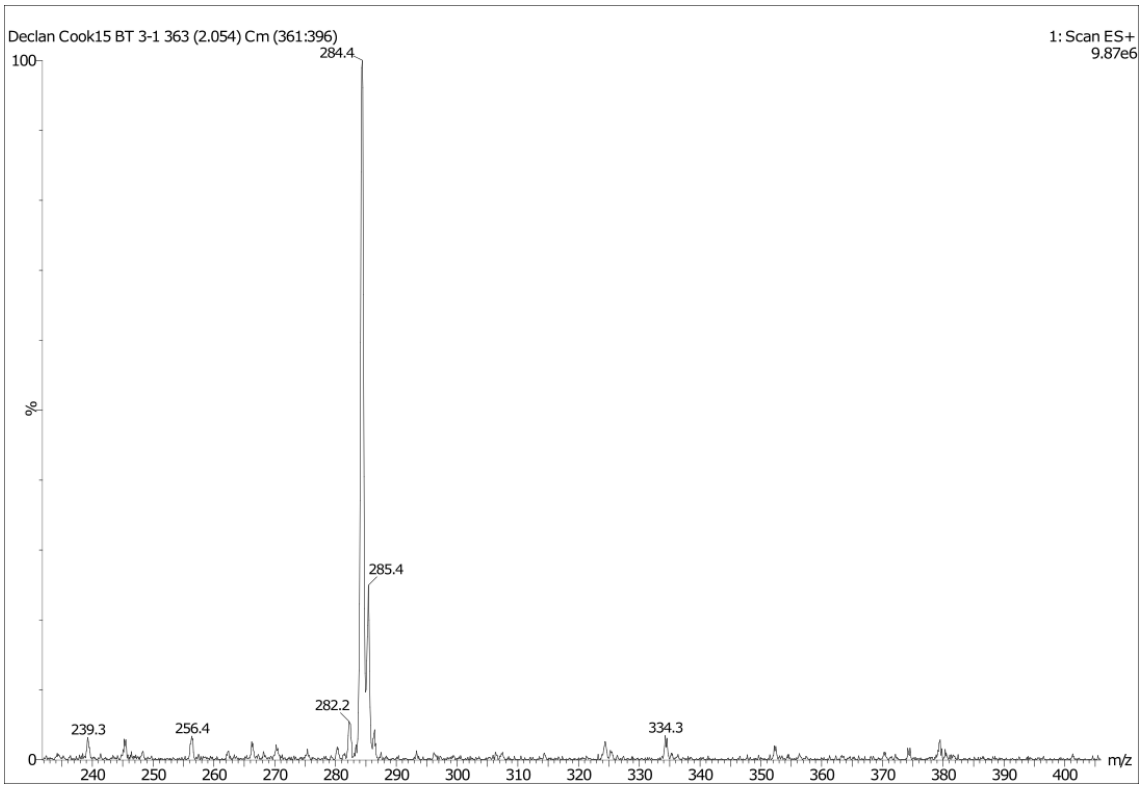
74

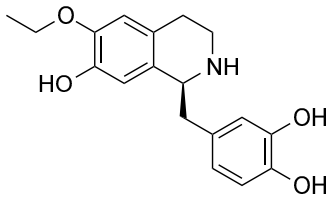

Declan Cook14 BT 2-1 421 (2.382) Cm (411:423)
1: Scan ES+
3.96e7
100
%
0
298.4
299.4
321.3
200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360 370 380 390 400
m/z


O
HO
NH

M+H = 284

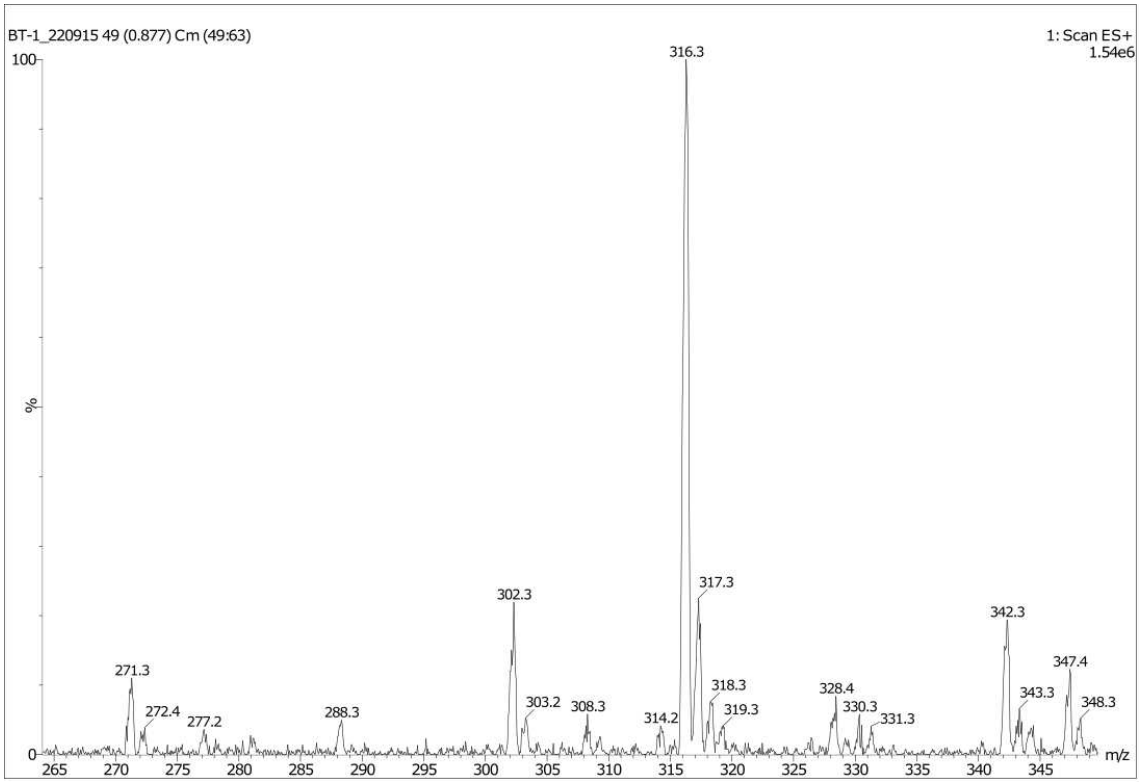
75

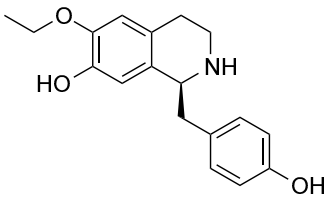

Declan Cook15 BT 3-1 363 (2.054) Cm (361:396)
1: Scan ES+
9.87e6
100
%
0
284.4
285.4
282.2
239.3
256.4
334.3
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
m/z


O
HO
NH
OH
OH

M+H = 316

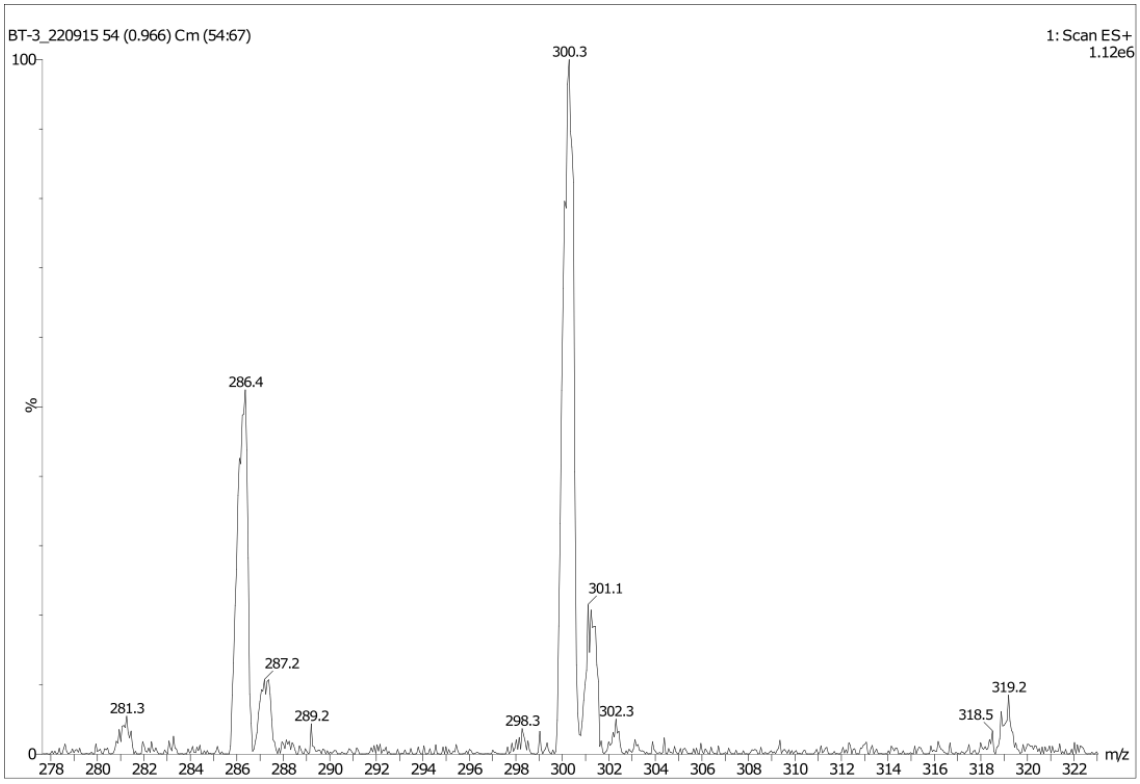
76a

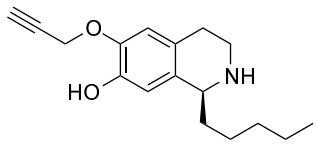

BT-1_220915 49 (0.877) Cm (49:63)
1: Scan ES+
1.54e6
100
%
0
316.3
317.3
302.3
342.3
347.4
271.3
272.4
277.2
288.3
303.2
308.3
314.2
318.3
319.3
328.4
330.3
331.3
343.3
348.3
265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345
m/z


O
HO
NH
OH

M+H = 300

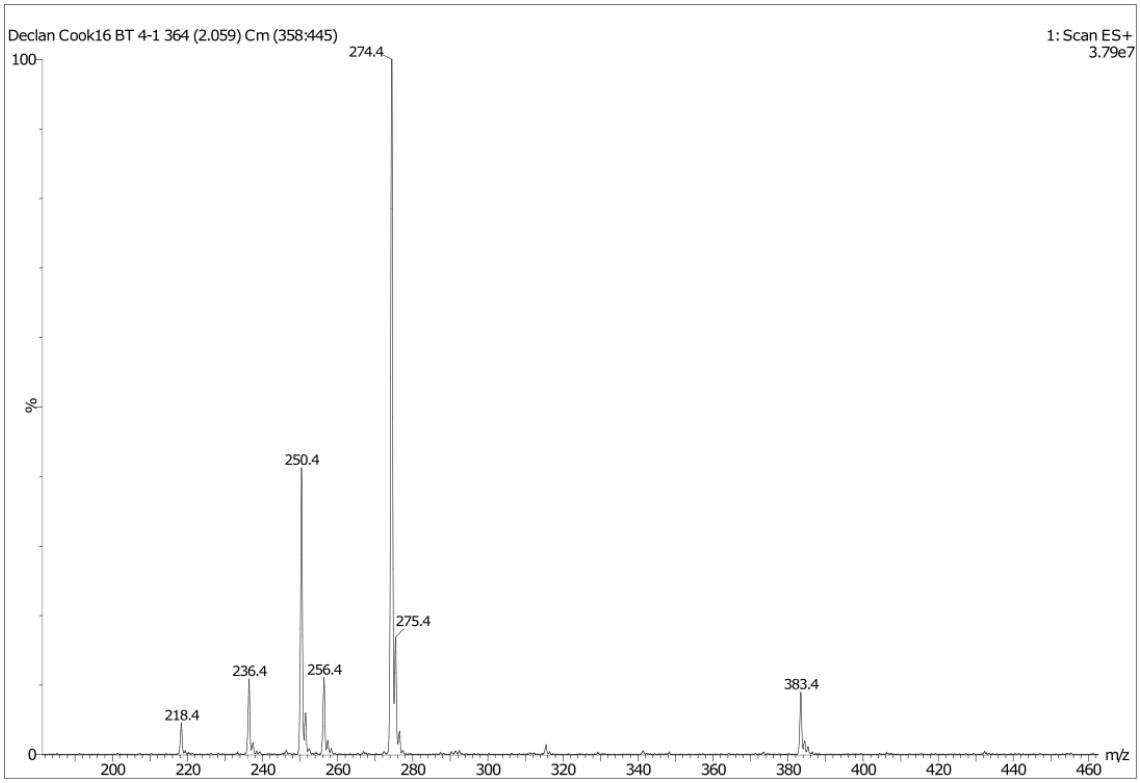
78

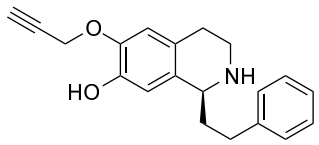

BT-3_220915 54 (0.966) Cm (54:67)
1: Scan ES+
1.12e6
100
%
0
300.3
286.4
301.1
287.2
281.3
289.2
298.3
302.3
319.2
318.5
278 280 282 284 286 288 290 292 294 296 298 300 302 304 306 308 310 312 314 316 318 320 322
m/z


O
HO
NH

M+H = 274

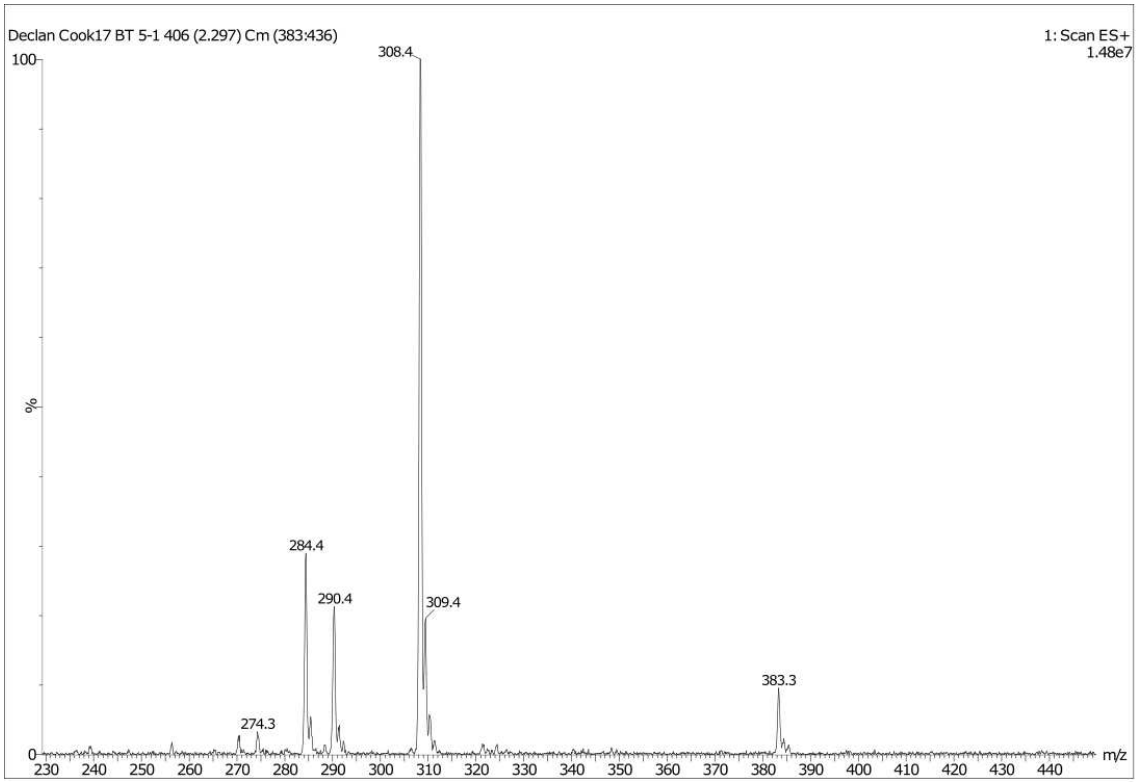
96

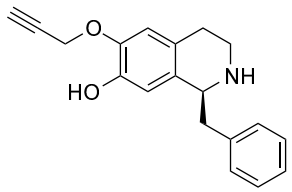

Declan Cook16 BT 4-1 364 (2.059) Cm (358:445)
1: Scan ES+
3.79e7
100
%
0
274.4
250.4
275.4
236.4
256.4
218.4
383.4
200
220
240
260
280
300
320
340
360
380
400
420
440
460
m/z


O
HO
NH

M+H = 308

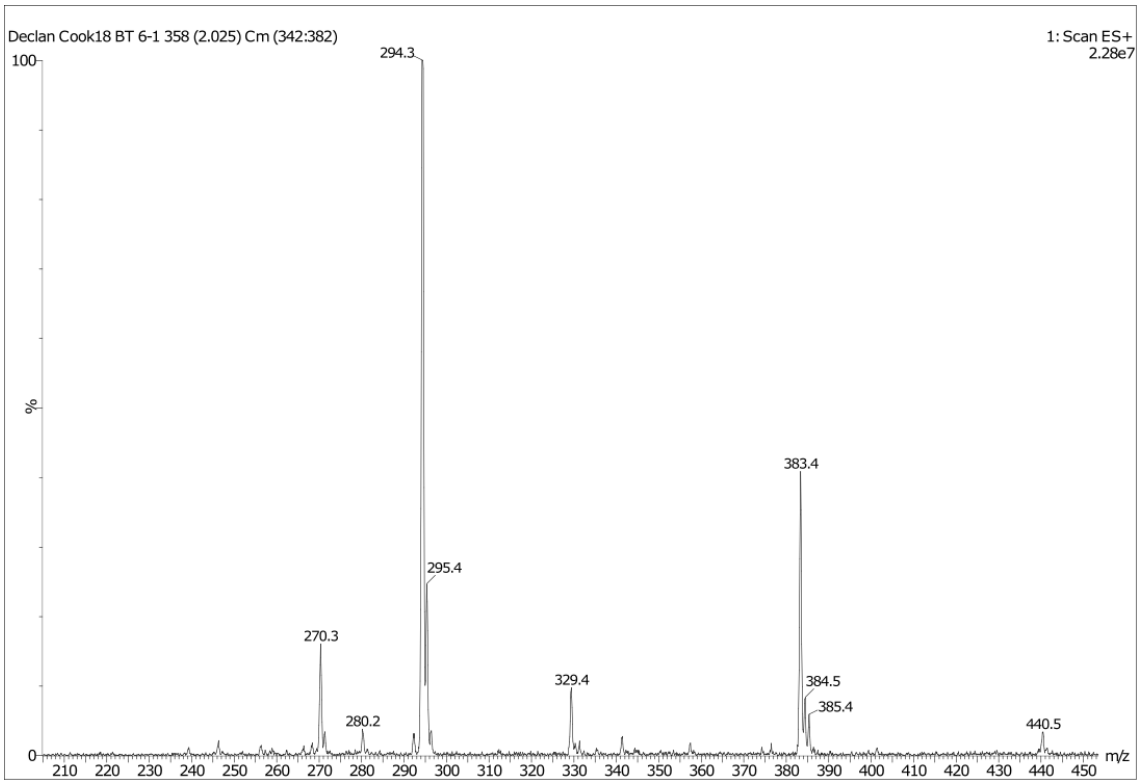
97

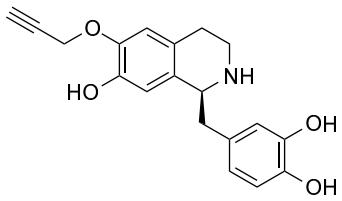

Declan Cook17 BT 5-1 406 (2.297) Cm (383:436)
1: Scan ES+
1.48e7
100
%
0
308.4
284.4
290.4
309.4
383.3
274.3
230 240 250 260 270 280 290 300 310 320 330 340 350 360 370 380 390 400 410 420 430 440
m/z


O
HO
NH

M+H = 294

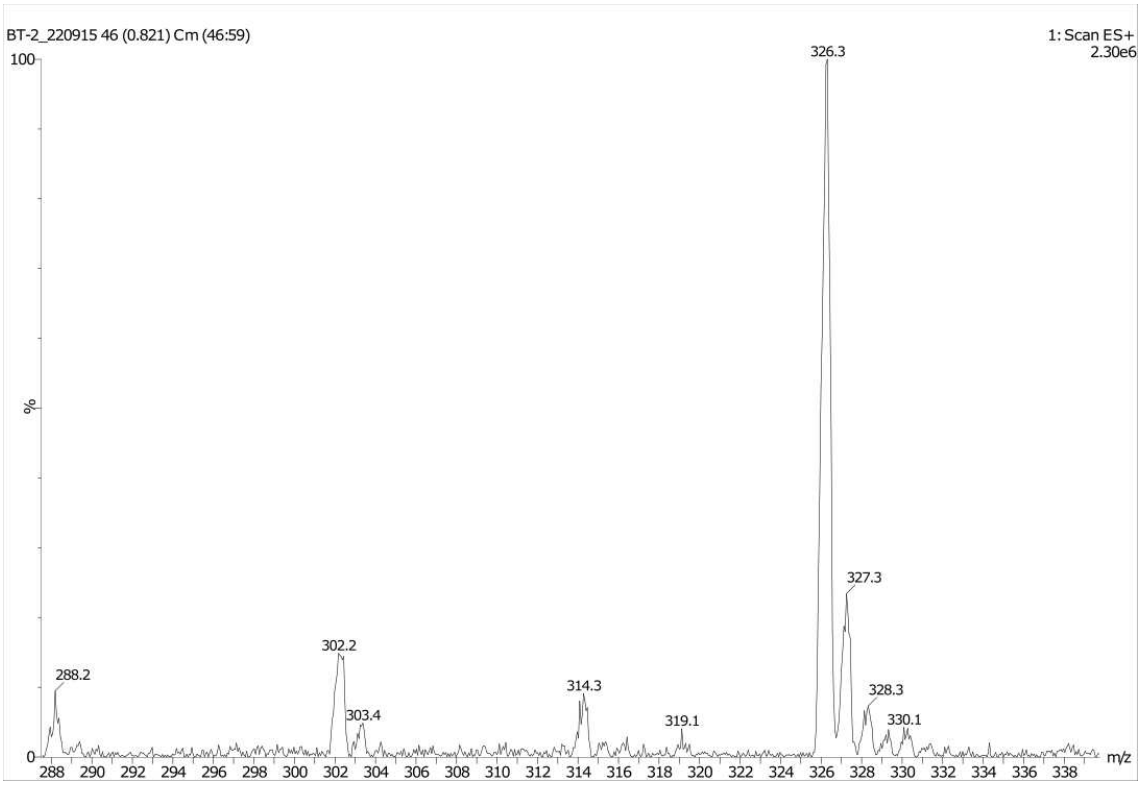
98

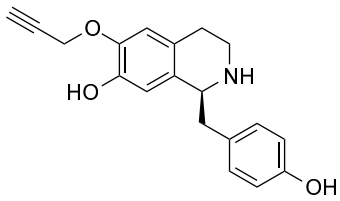

Declan Cook18 BT 6-1 358 (2.025) Cm (342:382)
1: Scan ES+
2.28e7
100
%
0
294.3
295.4
270.3
280.2
329.4
383.4
384.5
385.4
440.5
210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360 370 380 390 400 410 420 430 440 450
m/z


O
HO
NH
OH
OH

M+H = 326

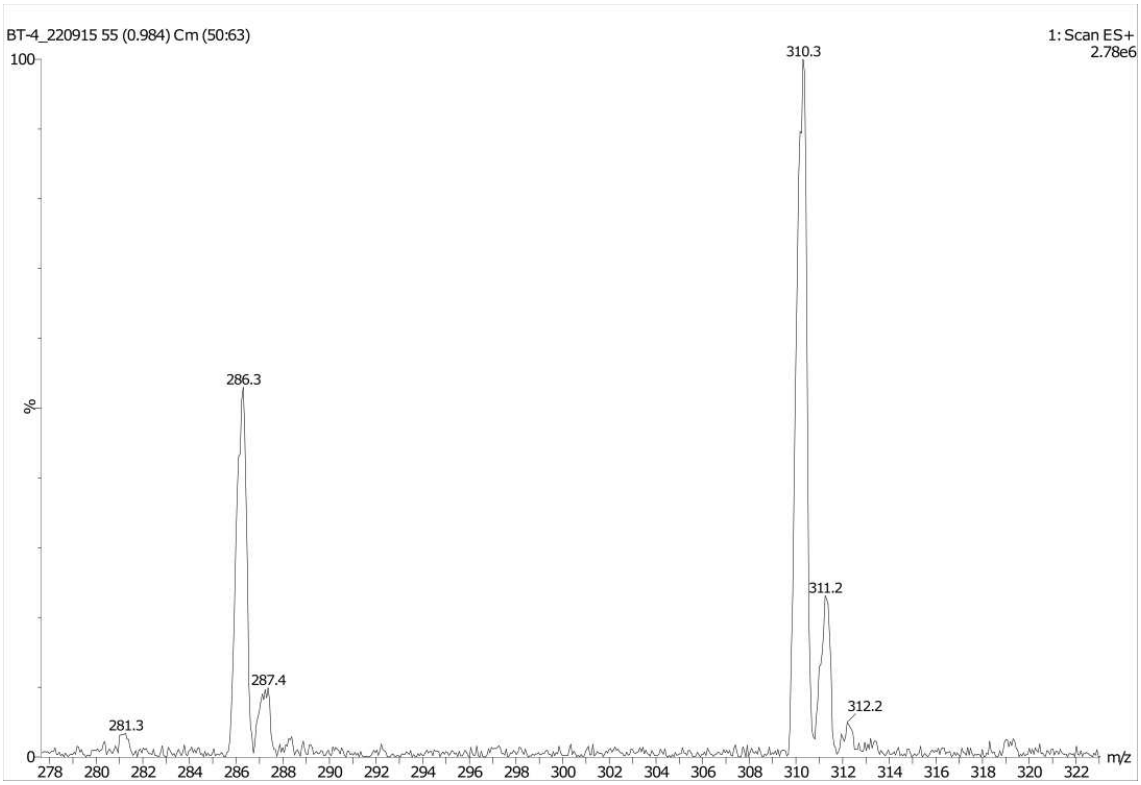
99


BT-2_220915 46 (0.821) Cm (46:59)
1: Scan ES+
2.30e6
100
%
0
326.3
327.3
302.2
288.2
314.3
328.3
303.4
319.1
330.1
m/z
288 290 292 294 296 298 300 302 304 306 308 310 312 314 316 318 320 322 324 326 328 330 332 334 336 338

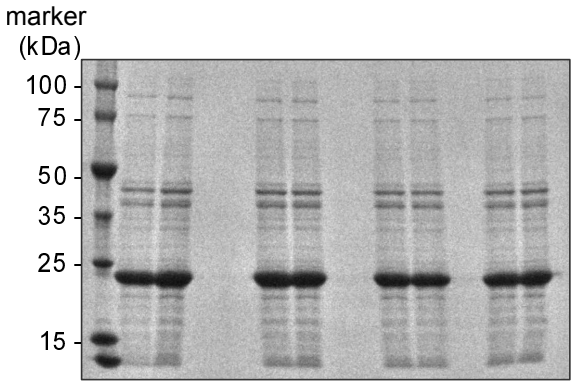

O
HO
NH
OH

M+H = 310

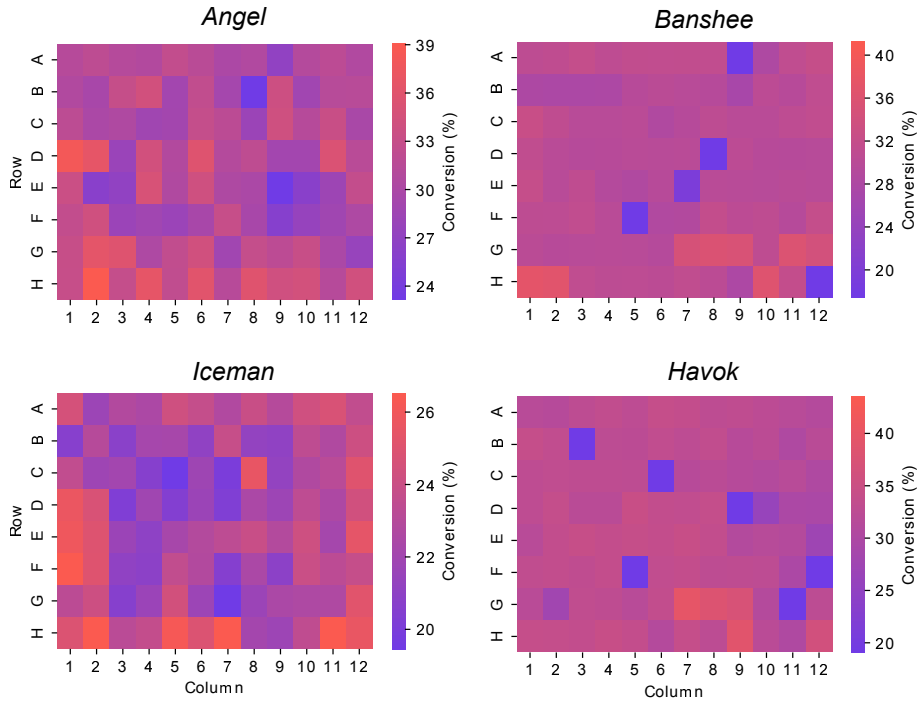
100


BT-4_220915 55 (0.984) Cm (50:63)
1: Scan ES+
2.78e6
100
%
0
310.3
286.3
311.2
287.4
312.2
281.3
278 280 282 284 286 288 290 292 294 296 298 300 302 304 306 308 310 312 314 316 318 320 322
m/z

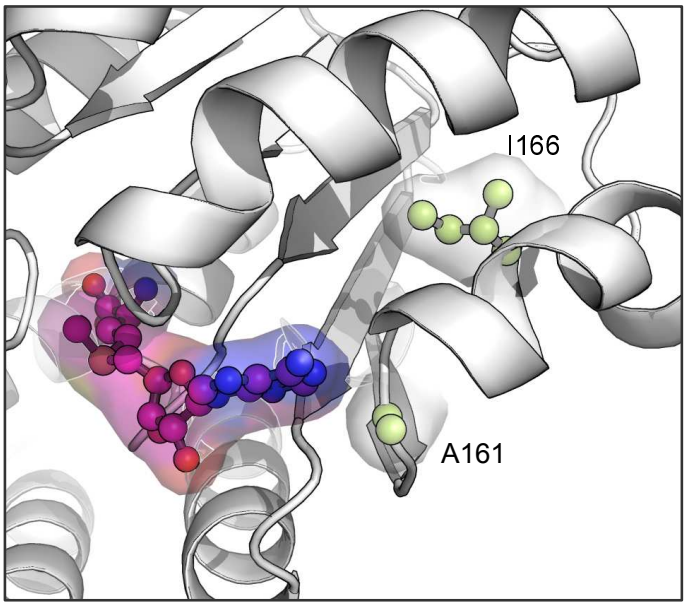
Appendix G: Random Mutagenesis Screen Analysis



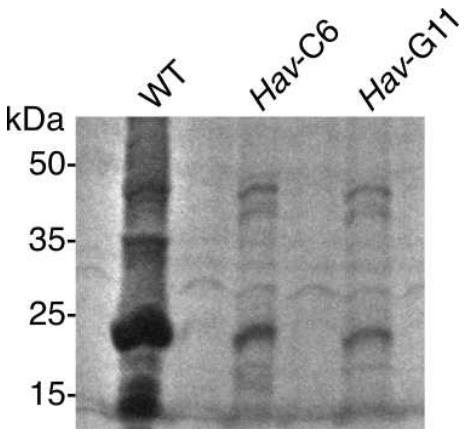
Supplementary Figure G.1: SDS-PAGE analysis of a sample of clarified lysates produced from *Rn*COMT mutant stock plates.



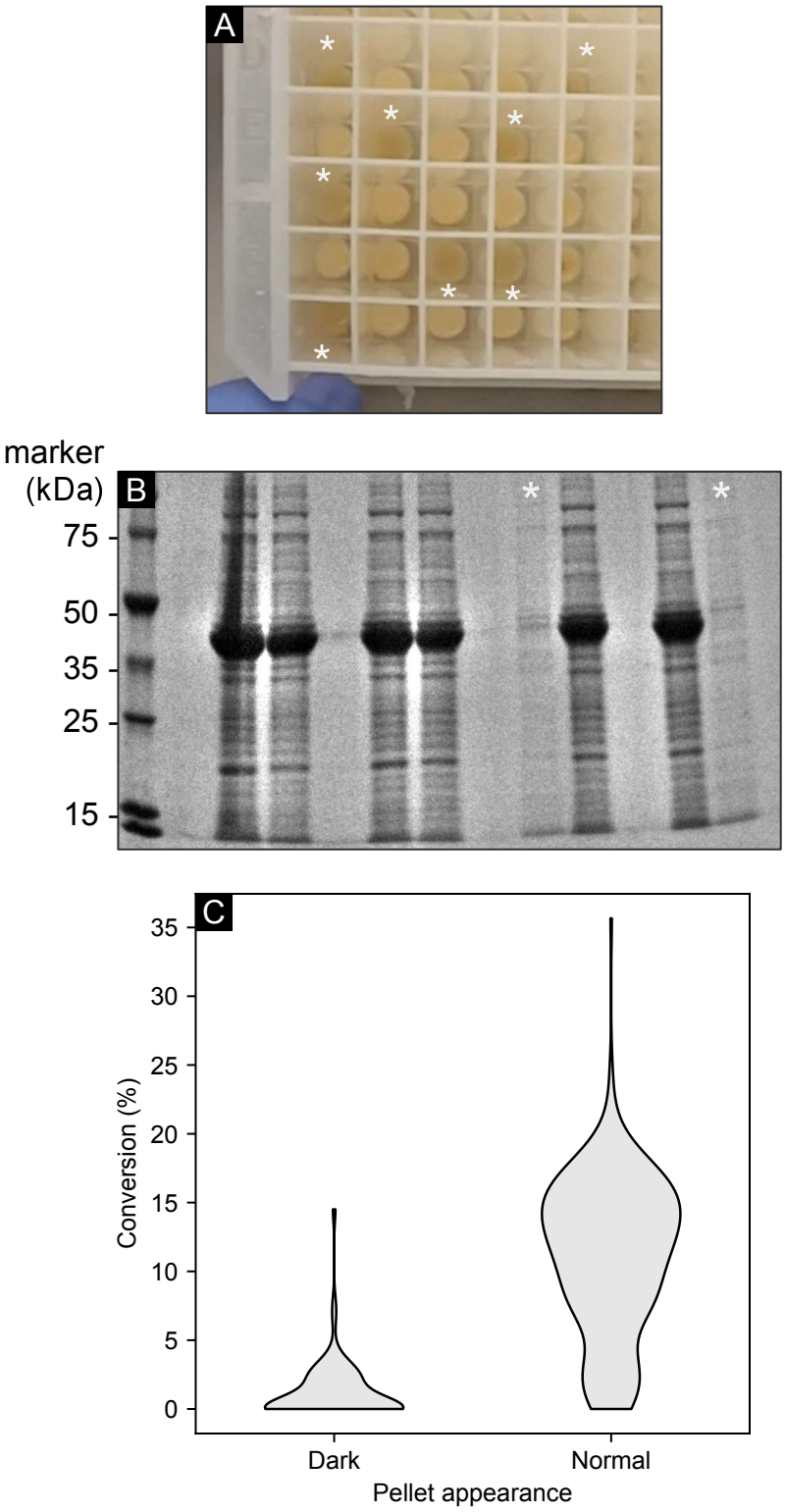
Supplementary Figure G.2: Heatmaps showing conversions across assay microplates to help visualise potential biases affecting rows G and H. Colour scale minima and maxima for each data set are at two standard deviations below and above the mean, respectively.



Supplementary Figure G.3: Graphic showing the positions of wild-type residues which are substituted in mutant *Angel*-E9 (*Rn*COMT (A161V, I166T))



Supplementary Figure G.4: SDS-PAGE analysis of *Rn*COMT mutants *Hav*-C6 and *Hav*-G11



Supplementary Figure G.5: Analysis of *Uu*MAT expression in random mutagenesis screen. A) Frame from video used to identify 'dark' pellets. Those categorised as such are marked with asterisks. B) SDS-PAGE analysis of a sample of clarified lysates produced from *Uu*MAT (mass = 45 kDa) mutant stock plates. Lanes showing lysates from dark pellets are marked with asterisks. C) Violin plot showing distribution of conversions from lysates of pellets with dark and normal appearances across all plates.