

**DEVELOPMENT OF THE  
C-METHYLTRANSFERASES NOVO  
AND COUO FOR BIOCATALYTIC  
FRIEDEL-CRAFTS ALKYLATION**

**Joanna C. Sadler**

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to GSK in accordance with the author's contract of engagement with GSK under the terms of the United Kingdom Copyright Acts. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

**Signed:**

**Date:**

## Abstract

Regiospecific monoalkylation and fluoroalkylation of aromatic compounds is of key importance in medicinal chemistry, yet can be difficult to accomplish chemically. Enzymatic methods may offer a viable alternative to traditional C-C bond forming reactions, such as the Friedel-Crafts alkylation. This thesis describes a number of key advances in the use of two S-adenosyl-L-methionine (SAM) dependent methyltransferases (MTs) NovO (*Streptomyces spheroides*) and CouO (*Streptomyces rishiriensis*) as biocatalysts for the Friedel-Crafts alkylation. Firstly, NovO was expressed, purified and crystallised and the X-ray crystal structure of SelMet-NovO was solved to 1.9 Å resolution (PDB accession code: 5MGZ). Additionally, a homology model of CouO (85% sequence identity to NovO) was generated using the NovO structure. Both NovO and CouO are Class I MTs, comprising a Rossmann fold with additional  $\alpha$ -helices at the C- and N-termini. Secondly, a catalytic mechanism for NovO was proposed based on the X-ray crystal structure, substrate docking studies, kinetic isotope effect (KIE) data and mutational analysis. This work identified a His-Arg motif to be central to the catalytic mechanism, with initial deprotonation by H120 and stabilisation of the resulting intermediate by R121. Additionally, the KIE data showed that methyl transfer was the rate limiting step. Finally, comparison with CouO identified position 117 in both proteins to be key for mediating the substrate scope of both enzymes. A third major contribution towards the development of NovO and CouO as biocatalysts was preliminary work on their directed evolution (DE) towards a wider substrate scope. Although a more efficient screening or selection strategy will be necessary to increase the substrate scope further, one mutant of NovO (N117M) was identified with higher activity toward 1, 6-dihydroxynaphthalene relative to the WT enzyme, serving as proof of concept for the ‘evolvability’ of these MTs. The development of a platform for biocatalytic fluoroalkylation using MTs was also explored, with extensive attempts towards the synthesis of a suitable fluoroalkylated cofactor, including the synthesis of novel fluorinated amino acids. A final key contribution towards the development of a biocatalytic Friedel-Crafts alkylation platform was the development of an *in situ* cofactor synthesis system using the halogenase SalL (*Salinospora tropica*). This was demonstrated on a preparative scale using crude cell lysates (methyl transfer) or purified enzymes ( $\text{CD}_3$ ,  $^{13}\text{CH}_3$ ,  $^{13}\text{CD}_3$  and Et transfer). This work addressed the issue of cofactor cost and instability associated with SAM dependent MTs.

## Acknowledgements

First and foremost, I would like to thank my supervisors Dr. Glenn Burley (University of Strathclyde), Dr. Luke Humphreys (GSK, currently at Gilead), Dr. Radka Snajdrova (GSK, currently at Novartis) and Dr. Joe Adams (GSK) for their ongoing support, guidance and input over the last three and a half years. Their comments regarding this thesis were also gratefully received and I am particularly grateful to Dr. Joe Adams for his part in carrying out the data integrity check. Finally, I would like to express my gratitude towards the Prof. Harry Kelly and Prof. William Kerr for accepting me onto the Industrial PhD programme, which gave me the opportunity to carry out the research described herein.

This thesis is multi-disciplinary, spanning organic chemistry, molecular biology, crystallography and bioinformatics. As such, a number of people were involved to varying degrees in helping me to carry out this research. Firstly, thank you to Dr. Chun-wa Chung (GSK), who solved the crystal structure of SelMet-NovO. This was a key milestone for the project that enabled the large majority of the work described in this thesis. I would also like to extend my gratitude towards Dr. Julie Mosley (GSK) for her help in setting up endless crystallisation screens, harvesting crystals and preliminary X-ray diffraction data analysis. Finally, I want to thank Dr. Colin Edge (GSK) for very helpful discussions regarding the design of mutants to probe the mechanism of NovO, in addition to providing useful feedback on the bioinformatics aspects of this thesis.

I am also very grateful to Dr. Andrew Fosberry (GSK) for sharing his expertise in molecular biology and in particular for his guidance in designing the CASTing libraries and with site-directed mutagenesis experiments. Thank you also to Dr. Emma Jones (GSK) and Angita Shrestha (GSK) for their help with enzyme purification. I would also like to express my appreciation for Pat Marder, who kindly prepared all the media for enzyme expression.

For their help with analytical chemistry at GSK, I would like to thank Dr. Andy Roberts, Larissa Perim Calmon, Dr. Helen Taylor, Dr. Tom Atherton, Dr. Nick Taylor, Dr. Petra Booij and Graham Stevens. I would also like to express my gratitude towards the Synthetic Biochemistry group at GSK Stevenage for creating a fun, dynamic and productive atmosphere and welcoming me into the group. Thank you also to Clive Mountain for witnessing my electronic lab notebook experiments and for his meticulous comments on a data integrity check for the *in situ* cofactor synthesis work.



Thank you also to the Burley group at the University of Strathclyde for welcoming me into the group. I am particularly grateful for useful discussions with Iain McKean, who is continuing work based on some of the ideas discussed in this thesis.

I am also very grateful towards Dr. Mandana Gruber (Austrian Centre of Industrial Biotechnology) for providing useful information regarding her previous work on NovO and CouO. From the University of Stuttgart, I would also like to thank Dr. Bettina Nestl and Philipp Trauzettel for useful discussions regarding the use of MATs for the synthesis of fluorinated Met analogues.

Finally, I have received very kind and patient support from my family and friends over the past three and a half years, all willing me to succeed and encouraging me through the highs and lows. I am hugely grateful for this and would like to extend my deepest thanks to every one of them.

# Contents

Abstract.....	3
Acknowledgements.....	4
Abbreviations.....	10
1 Introduction.....	15
1.1 Enzyme classes .....	16
1.2 Methyltransferases (MTs).....	16
1.3 <i>S</i> -adenosyl-L-methionine (SAM).....	20
1.4 Mechanism of methyl transfer .....	22
1.4.1 DNA MTs .....	24
1.4.2 Radical SAM MTs .....	25
1.4.3 Small molecule MTs .....	28
1.5 C-MTs.....	29
1.5.1 Aliphatic C-MTs.....	29
1.5.2 Aromatic C-MTs.....	31
1.6 Friedel-Crafts alkylation using NovO and CouO.....	35
1.7 Aromatic trifluoromethylation .....	37
1.8 Directed evolution.....	39
1.8.1 Rational design strategies.....	41
1.8.2 Semi-rational evolution strategies.....	42
1.8.3 Random mutagenesis strategies .....	45
1.8.4 Computational protein design .....	46
1.9 Screening and selection.....	46
1.9.1 Screening.....	47
1.9.2 Selection.....	50
1.10 Strategies for the synthesis of SAM and non-natural SAM analogues .....	50
1.10.1 MATs .....	51

---

1.10.2	Halogenases .....	55
1.10.3	Stable SAM isosteres .....	58
1.11	Aims and objective .....	59
2	Expression, purification and structural and functional characterisation of NovO and CouO.....	63
2.1	Expression and purification of NovO and CouO .....	63
2.2	Synthesis of aminocoumarin substrates and activity assays of NovO and CouO ...	70
2.3	Crystallisation of NovO .....	73
2.4	Selenomethionine labelling and structure elucidation of NovO .....	75
2.5	Crystallisation of SelMet NovO.....	78
2.6	Analysis of the X-ray crystal structure of SelMet NovO complexed with SAH ....	80
2.6.1	Cofactor binding in NovO.....	83
2.7	Attempts towards obtaining a crystal structure of NovO with substrate bound in the active site .....	86
2.8	Crystallisation screens with substrate and non-reactive SAM analogues .....	89
3	Probing the Active Site of NovO and CouO.....	93
3.1	Preparation and analysis of a NovO-substrate docking model .....	93
3.2	Mutational analysis of putative substrate binding residues.....	96
3.3	Methyltransferase assay with NovO and NovO mutants .....	100
3.4	Investigation into the functions of H120 and R121 .....	102
3.5	Substrate scope of NovO and CouO .....	106
3.6	Kinetic isotope effects of NovO.....	111
3.6.1	Proposed catalytic mechanism of NovO .....	117
3.6.2	Comparison of proposed mechanism to previous studies .....	118
3.7	Homology modeling of CouO based on NovO.....	119
4	Attempts towards expanding the substrate scope of NovO by directed evolution.....	129
4.1	Preparation of CASTing libraries .....	129
4.2	Development of a MT activity screening assay .....	132

---

4.3	Synthesis and evaluation of HCys selective probes .....	145
4.4	Purification and screening of purified CASTing libraries .....	156
4.5	Screening of CASTing libraries by UPLC.....	162
5	Cofactor Synthesis and <i>In Situ</i> Cofactor Generation .....	168
5.1	Cofactor synthesis.....	168
5.1.1	Synthesis of a CF <sub>3</sub> analogue of SAM.....	168
5.1.2	Electrophilic trifluoromethylation.....	174
5.1.3	Nucleophilic trifluoromethylthiolation .....	176
5.1.4	Radical trifluoromethylation .....	177
5.1.5	Attempts towards the S <sub>N</sub> 2 reaction between an activated adenosine derivate and fluorinated Met analogue .....	182
5.2	Attempts towards the direct trifluoroalkylation of SAH.....	184
5.3	Enzymatic strategies for the synthesis of SAM and SAM analogues .....	187
5.4	Tandem SAM-synthesis and substrate alkylation .....	190
5.5	Synthesis of Met analogues.....	195
5.6	Application of tandem SAM synthesis and methylation to isotope labelling .....	205
	Conclusions.....	211
6	Experimental .....	217
6.1	General experimental techniques and procedures.....	217
6.2	Synthetic procedures .....	218
7	General Procedure for the one-pot enzymatic cascade methylation using E. coli cell free extract and E. coli cell lysate .....	253
7.1	General Procedure for the one-pot enzymatic cascade methylation using purified enzymes .....	254
7.2	Gene synthesis and enzyme expression .....	260
	Gene synthesis and cloning (SalL, NovO and CouO).....	260
	Enzyme expression .....	260
	Gene synthesis and cloning (SAHH, MTAN and LuxS) .....	262

---

7.2.1	Enzyme purification.....	263
7.3	Enzyme assays .....	268
7.4	Determination of SelMet NovO crystal structure .....	273
	Bibliography .....	275
	Appendix.....	293
	HPLC, LC-MS and UPLC Methods.....	294
	DNA sequences .....	297
	Media preparation.....	305
	SelMet-NovO X-ray diffraction data collection and refinement statistics .....	309

## Abbreviations

<b>5'-dA'</b>	5'-Deoxyadenosine radical
<b>5-MeTHF</b>	5-methyltetrahydrofolate
<b>AcCl</b>	Acetyl chloride
<b>AcOH</b>	Acetic acid
<b>ATP</b>	Adenosine triphosphate
<b>B-FIT</b>	B-Factor Iterative Test
<b>BLAST</b>	Basic local alignment search tool
<b>Boc</b>	<i>tert</i> -Butyloxycarbonyl
<b>BrDA</b>	5'-Bromo-5'-deoxyadenosine
<b>BSA</b>	Bovine serum albumin
<b>BuOH</b>	<i>n</i> -Butanol
<b>BzCl</b>	Benzoyl chloride
<b>CAST</b>	Combinatorial Active-Site Saturation Test
<b>CCoAOMT</b>	Caffeoyl coenzyme A 3-O-methyltransferase
<b>CFE</b>	Cell free extract
<b>CIDA</b>	5'-chloro-5'-deoxy adenosine
<b>COMT</b>	Catechol <i>O</i> -methyltransferase
<b>CPME</b>	Cyclopentylmethyl ether
<b>CV</b>	Column volumes
<b>Cx-SAM</b>	Carboxy- <i>S</i> -adenosyl- <i>L</i> -methionine
<b>DDMQH<sub>2</sub></b>	2-Methoxy-6-poly- prenyl-1,4-benzoquinone
<b>DE</b>	Directed evolution
<b>DEAD</b>	Diethyl azodicarboxylate
<b>DFHBI</b>	3, 5-Difluoro-4-hydroxybenzylidene imidazolinone
<b>DHN</b>	Dihydroxynaphthalene
<b>DIPEA</b>	<i>N, N</i> -Diisopropylethylamine
<b>DMAP</b>	<i>N, N</i> -Dimethylaminopyridine
<b>DMF</b>	<i>N, N</i> -Dimethylformamide
<b>DMKH<sub>2</sub></b>	Demethylmenaquinol
<b>DMQH<sub>2</sub></b>	2-Methoxy-5-methyl-6-polyprenyl-1,4-benzoquinone
<b>DMSO</b>	Dimethylsulfoxide
<b>DNA</b>	Deoxyribonucleic acid

---

<b>DNMT1</b>	Deoxyribonucleic acid methyltransferase 1
<b>dNTP</b>	Deoxy-nucleoside triphosphate
<b>DTNB</b>	5,5'-Dithiobis-(2-nitrobenzoic acid)
<b>DTT</b>	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
<b>EC</b>	Enzyme classification
<b>EDTA</b>	Ethylenediaminetetraacetic acid
<i>ee</i>	Enantiomeric excess
<b>EI</b>	Electron ionisation
<b>epPCR</b>	Error-prone polymerase chain reaction
<b>ES</b>	Electrospray
<b>EtOAc</b>	Ethyl acetate
<b>EtOH</b>	Ethanol
<b>FACS</b>	Fluorescence activated cell sorting
<b>FDA</b>	5'-fluoro-5'-deoxy adenosine
<b>GC-MS</b>	Gas chromatography-mass spectrometry
<b>GSH</b>	Glutathione
<b>GSK</b>	GlaxoSmithKline
<b>HCys</b>	Homocysteine
<b>HEPES</b>	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid)
<b>HPLC</b>	High pressure liquid chromatography
<b>HRMS</b>	High resolution mass spectrometry
<b>IDA</b>	5'-Iodo-5'-deoxyadenosine
<b>IPTG</b>	Isopropyl $\beta$ -D-1-thiogalactopyranoside
<b>KIE</b>	Kinetic isotope effect
<b>KMT</b>	Lysine methyltransferase
<b>KRED</b>	Ketoreductase
<b>LB</b>	Luria Bertani
<b>LC-MS</b>	Liquid chromatography-mass spectrometry
<b>MAT</b>	Methionine adenosyl transferase
<b>MDAP</b>	Mass-directed automated purification
<b>Mes</b>	2-( <i>N</i> -morpholino)ethanesulfonic acid
<b>MOE</b>	Molecular Operating Environment
<b>Mpt.</b>	Melting point
<b>MSA</b>	Multiple sequence alignment

---

<b>MsCl</b>	Methanesulfonyl (mesyl) chloride
<b>MT</b>	Methyltransferase
<b>MTA</b>	5'-Methylthioadenosine
<b>MTAN</b>	5'-Methylthioadenosine nucleosidase
<b>MTB</b>	Modified Terrific broth
<b>MV<sup>2+</sup></b>	Methyl viologen
<b>NBS</b>	<i>N</i> -Bromosuccinimide
<b>NMR</b>	Nuclear magnetic resonance
<b>Ns</b>	2-Nitrobenzenesulfonyl chloride (nosyl)
<b>NTA</b>	Nitrilotriacetic acid
<b>OD</b>	Optical density
<b>OE</b>	Overnight Express™
<b>PCR</b>	Polymerase chain reaction
<b>PD</b>	Proximity and desolvation
<b>PDB</b>	Protein Data Bank
<b>PEG</b>	Poly(ethylene) glycol
<b>PLB</b>	Propensity for ligand binding
<b>PMF</b>	Protein mass fragmentation
<b>PMT</b>	Protein methyltransferase
<b>PRMT</b>	Protein arginine methyltransferase
<b>PSMS</b>	Zinc bis(phenylsulfonylmethanesulfinate)
<b>QM</b>	Quantum mechanical
<b>RDS</b>	Rate determining step
<b>RLS</b>	Rate limiting step
<b>RNA</b>	Ribonucleic acid
<b>rRNA</b>	Ribosomal ribonucleic acid
<b>RSMT</b>	Radical <i>S</i> -adenosyl-L-methionine methyltransferase
<b>SAH</b>	<i>S</i> -Adenosylhomocysteine
<b>SAHH</b>	<i>S</i> -Adenosylhomocysteine hydrolase
<b>SAM</b>	<i>S</i> -Adenosyl-L-methionine
<b>SAMT</b>	Salicylic acid carboxyl methyltransferase
<b>SDM</b>	Site-directed mutagenesis
<b>SDS PAGE</b>	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
<b>SEC</b>	Size exclusion chromatography
<b>SelMet</b>	Selenomethionine



---

<b>SET</b>	Single electron transfer
<b>SI</b>	Stable isotope
<b>SPOUT</b>	Superfamily of MTs comprising SpoU and TrmD
<b>TA</b>	Transaminase
<b>TB</b>	Terrific Broth
<b>TBHP</b>	<i>tert</i> -Butylhydroperoxide
<b>TBME</b>	<i>tert</i> -Butylmethylether
<b>THF</b>	Tetrahydrofuran
<b>TIM</b>	Triosephosphateisomerase
<b>TMG</b>	Tetramethylguanidine
<b>TOF</b>	Turnover frequency
<b>TS</b>	Transition state
<b>TsCl</b>	4-Toluenesulfonyl (tosyl) chloride
<b>UPLC</b>	Ultra-high pressure liquid chromatography
<b>UV</b>	Ultra-Violet
<b>WT</b>	Wild type
<b>ZPE</b>	Zero-point energy

# **Chapter 1.**

## **Introduction**

# 1 Introduction

The development of low cost, environmentally benign and efficient synthetic strategies is crucial for the sustainable production of high value organic compounds such as pharmaceuticals. Although there are many examples of this being achieved using traditional synthetic organic chemistry approaches, there are still many transformations for which there is no alternative to harsh reaction conditions or low yields due to poor chemo-, regio- or stereoselectivity. A number of new technologies have evolved over recent years in order to tackle this. For example, research in the fields of metal and organocatalysis has provided a myriad of catalysts for a vast range of reactions, such as in the popular field of C-H bond activation.<sup>[1,2]</sup> Additionally, new manufacturing technologies such as flow chemistry are emerging as highly efficient and easily tunable systems for chemical synthesis.<sup>[3,4]</sup> Flow chemistry has been shown to be particularly useful for reactions which are not suitable for large scale batch synthesis, such as exothermic reactions.<sup>[5]</sup> A third area of interest is the use of enzymes as catalysts, which is known as biocatalysis.

Biocatalysis offers significant potential as a strategy for organic synthesis, due to the inherent regioselectivity, high stereoselectivity, substrate specificity and mild reaction conditions associated with the use of enzymes. Furthermore, there are now many effective strategies for modifying and designing enzymes to perform specific reactions with improved characteristics such as higher stereoselectivity and higher activity, as well as for improving stability and solubility in both water and organic solvents. This field is known as directed enzyme evolution and is discussed in detail in Section 1.8.<sup>[6-13]</sup> As a result, biotransformations are becoming an economically viable and environmentally attractive alternative for the industrial scale synthesis of organic small molecules.<sup>[14]</sup> However, the widespread use of biocatalysis is currently restricted by limitations such as narrow substrate scope of enzymes, enzyme instability and cofactor availability.<sup>[15]</sup> As such, there is a need to expand the existing biocatalytic toolbox with robust and versatile biocatalysts to cater for a variety of chemical transformations, through a combination of identifying new biocatalysts from Nature and the directed evolution of existing enzymes. Additionally, the *de novo* design of enzymes for biocatalysis *in silico* is a promising approach and is predicted to become increasingly important over the next decade as developments in synthetic biology aid *de novo* design and *vice versa*.<sup>[16]</sup> Finally, despite significant advances in the field of cofactor regeneration for biocatalysis,<sup>[17]</sup> there is still a need for the development of cost effective

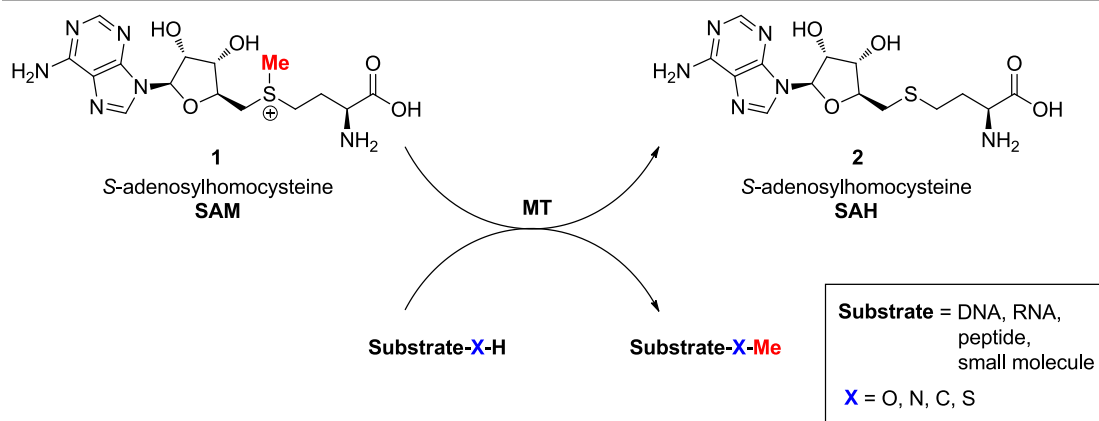
methods for cofactor synthesis or recycling for some classes of enzymes in order for biocatalysis to be an economically viable option.

## 1.1 Enzyme classes

According to the enzyme classification (EC) system, there are six major classes of naturally occurring enzymes, which have evolved to perform every transformation necessary to synthesise the myriad of organic molecules found in Nature. These are: oxidoreductases (oxidation and reduction); transferases (transfer of groups such as acetyl, phosphate, amino and methyl); hydrolases (hydrolysis); lyases (addition and elimination of small molecules); isomerases (catalysing reactions such as racemisation and epimerisation) and ligases (coupling of molecules to form C-C, C-O, C-S, P-O or C-N bonds) and have EC numbers 1-6, respectively.<sup>[18]</sup> Enzymes from all 6 classes have been employed as biocatalysts, in some cases on an industrial scale. For example, ketoreductases (KREDs, from the oxidoreductase family, EC 1) are commonly employed in the enantioselective reduction of ketones for the synthesis of chiral alcohols,<sup>[19]</sup> whilst transaminases (TAs, from the transferase family, EC 2) are a popular method for the transformation of a ketone to a chiral amine,<sup>[20]</sup> which can be difficult to achieve using traditional chemical approaches. Contrary to KREDs and TAs, methyltransferases (MTs, EC 2) are much less well developed for use in synthetic chemistry, yet have significant potential as a valuable addition to the biocatalytic toolbox owing to their ability to perform regiospecific methylation reactions. Furthermore, MTs avoid the issue of genotoxicity associated with many chemical alkylating agents. As such, the development of two C-MTs (NovO and CouO) for use in organic synthesis is the subject of this thesis.

## 1.2 Methyltransferases (MTs)

MTs catalyse the stereospecific transfer of a one carbon unit from a methyl donor (the cofactor) to a substrate. Whilst some MTs employ methyl donors such as cobalamin,<sup>[21,22]</sup> L-methylfolate<sup>[23–25]</sup> and dimethylthetin,<sup>[26]</sup> the vast majority of MTs found in Nature are *S*-adenosyl-L-methionine (SAM, **1**) dependent. SAM dependent MTs are responsible for the methylation of a structurally diverse set of substrates, spanning small molecules such as catechol and salicylic acid, through to biomacromolecules such as deoxyribonucleic acid (DNA), peptides and lipids, and can perform *N*-, *O*-, *C*- and *S*-transmethylation reactions to provide the methylated product and *S*-adenosylhomocysteine (SAH, **2**) as the by-product (**Scheme 1**).<sup>[27–29]</sup>



**Scheme 1.** General scheme for methylation of substrate by SAM dependent MTs.

Over the past decade there has been a growing interest in MTs from two different perspectives. Firstly, MTs have been identified as key mediators of epigenetic regulation of gene transcription, for which two classes of MTs have been identified.<sup>[30]</sup> The first is DNA MTs, which can silence gene transcription directly *via* methylation at the 5-position of cytosine.<sup>[31–33]</sup> The second is protein MTs (PMTs), which regulate the post-translational modification of histone proteins *via* protein arginine methyltransferases (PRMTs) and lysine methyltransferases (KMTs).<sup>[33–35]</sup> Post-translational modifications of histones control conformational transitions between condensed and relaxed states of chromatin, which in turn regulates gene transcription.<sup>[33]</sup> Disregulation of epigenetic control *via* both of these mechanisms has been implicated in a number of diseases, including many cancers.<sup>[36,37]</sup> As such, both DNA MTs and PMTs have attracted considerable attention as drug targets, especially for cancer therapies. Indeed, two small molecule DNA MTs inhibitors (5-azacitidine and decitabine) have been approved in the United States for the treatment of myelodysplastic syndrome<sup>[38]</sup> and the development of a selective inhibitor of different PRMTs for the treatment of a variety of cancers is an active area of research.<sup>[39–41]</sup>

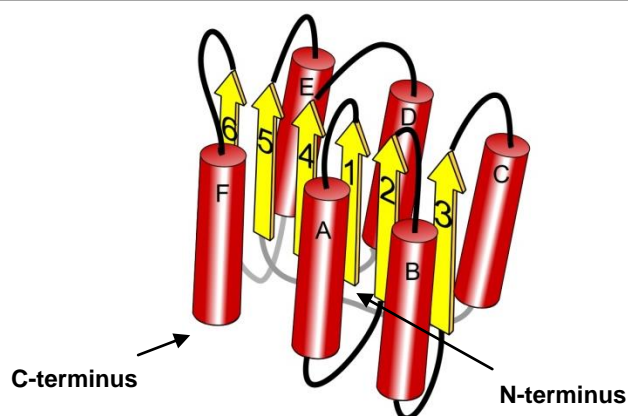
The second main area of interest in MTs is for use in synthetic organic chemistry. Methyl groups are ubiquitous in pharmaceutically active compounds, owing to their beneficial effect on binding affinity to a given target protein, which has become known as the ‘magic methyl effect’.<sup>[42]</sup> Methylation of a small molecule increases its lipophilicity, thereby reducing the free energy of desolvation upon the transfer of a substrate from an aqueous environment into the binding pocket of the target protein.<sup>[43]</sup> Depending upon the position of the methyl group in the substrate relative to the binding site, methylation of a small molecule has been shown to result in a 3.5- to 43-fold improvement binding affinity.<sup>[44–46]</sup> Despite the prevalence of the

methyl group medicinal chemistry, the regiospecific addition of a methyl group to small molecules can be challenging and addressing the dearth of synthetic methods to achieve this is an active area of research.<sup>[42,47]</sup> SAM dependent MTs have attracted considerable attention as biocatalysts to this end, due to their ability to perform late-stage, regiospecific methylation of small molecules under mild reaction conditions and without the need for protecting groups.<sup>[48,49]</sup>

To further develop MTs for applications in both chemical biology and biocatalysis, the structural and mechanistic characterisation of many of these enzymes has been carried out. For chemical biology applications, structural and mechanistic information can aid the rational design of selective inhibitors of MTs.<sup>[50,51]</sup> For applications in biocatalysis, this insight may facilitate or aid the directed evolution of MTs towards synthetically useful substrates and/or tolerance towards process conditions.<sup>[52]</sup> As MTs catalyse the transfer of a methyl group to a wide range of substrates, there is a very large degree of structural and mechanistic variation across the class, which has been divided into 9 sub-classes in two successive publications:

- Class I: Rossmann-fold MTs (SAM dependent, alpha/beta)<sup>[53]</sup>
- Class II: TIM beta/alpha-barrel (alpha/beta)<sup>[53]</sup>
- Class III: tetrapyrrole methylase (alpha/beta)<sup>[53]</sup>
- Class IV: SPOUT (alpha/beta)<sup>[53]</sup>
- Class V: SET domain (all beta)<sup>[53]</sup>
- Class VI: transmembrane (all alpha)<sup>[54]</sup>
- Class VII: DNA/RNA-binding 3-helical bundle (all alpha)<sup>[54]</sup>
- Class VIII: SSo0622-like (alpha/beta)<sup>[54]</sup>
- Class IX: thymidylate synthetase (alpha/beta)<sup>[54]</sup>

Out of these, Class I MTs are by far the most abundant and can be identified by the characteristic Rossmann fold in their tertiary structure.<sup>[55]</sup> The Rossmann fold consists of a seven membered  $\beta$ -sheet with a  $\beta$ -hairpin at the C-terminus, which is sandwiched by seven  $\alpha$ -helices (**Figure 1**).



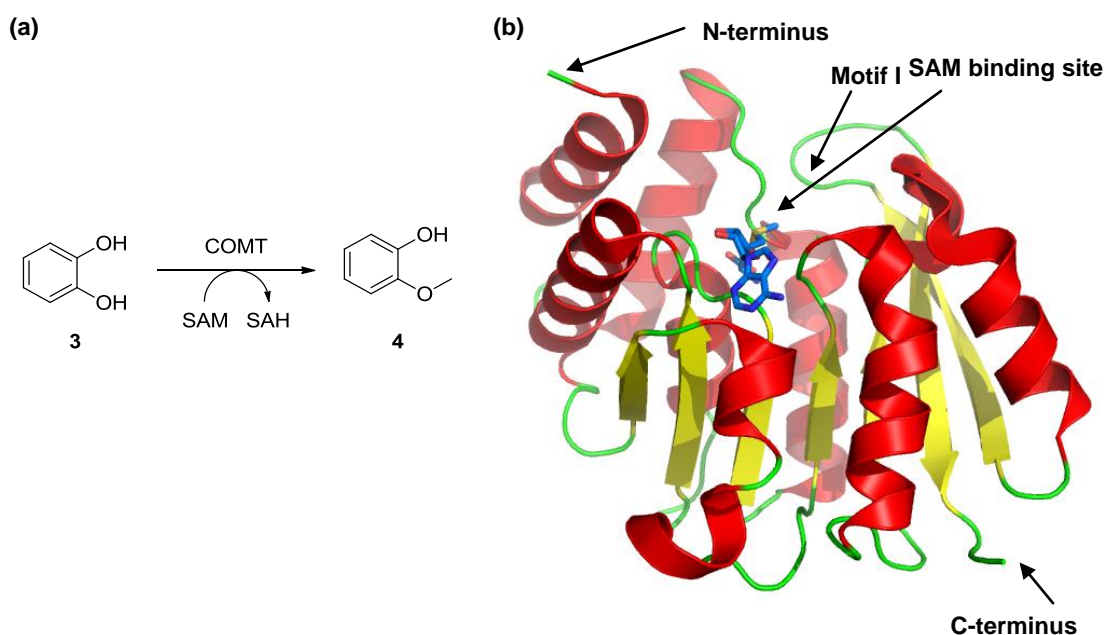
**Figure 1.** Diagrammatic topology of the core fold (or Rossmann fold) in Class I SAM-dependent MTs. Helices are shown as red cylinders and strands as yellow arrows.<sup>[55]</sup>

The  $\beta$ -strand order 3 2 1 4 5 (7) 6 is most commonly observed, although arrangements with the reverse order of  $\beta$ -strands have been reported.<sup>[56]</sup> Within this structure, binding domains for SAM and the substrate are clearly defined, although many of the residues involved in SAM binding are poorly conserved across the class.<sup>[55]</sup> Despite the high degree of sequence and structural variation between Class I MTs, Kozbial and Mushegian have defined six distinctive motifs within the characteristic Rossmann fold:<sup>[57]</sup>

- Motif I: The consensus sequence DL/VGXGXG (where X=any amino acid) found at the N-terminus of the first  $\beta$ -strand and the loop connecting the first  $\beta$ -strand and the adjacent  $\alpha$ -helix.
- Motif II: The second  $\beta$ -strand after Motif I.
- Motif III: The third  $\beta$ -strand located at the edge of the Rossmann fold.
- Motif IV: The fourth  $\beta$ -strand and the flanking loops.
- Motif V: The helix following the fourth  $\beta$ -strand.
- Motif VI: The motif that corresponds to  $\beta$ -strand 5.

This can be visualized by analysis of an archetypal Class 1 MT, such as catechol *O*-MT (COMT). As shown in **Figure 2a**, COMT catalyses the transfer of a methyl group from SAM to a hydroxy group on catechol (**3**) to make guaiacol (**4**), a small volatile compound which is a precursor to many natural flavourants.<sup>[58]</sup> The crystal structure of COMT from *Rattus norvegicus* (**Figure 2b**) highlights the Rossmann fold described above. Seven  $\beta$ -sheets form the core of the structure, with a series of  $\alpha$ -helices forming the outer shell. Most small-molecule MTs characterized to date display modifications at the N-terminus, whilst the C-terminus is largely unmodified. This is also exemplified by COMT, which displays an

additional two  $\alpha$ -helices at the N-terminus.<sup>[55,59,60]</sup> Conversely, for macromolecule MTs such as DNA and RNA MTs, the Rossmann fold is further decorated with additional secondary structural motifs at the C-terminus, but generally not the N-terminus.<sup>[55]</sup>



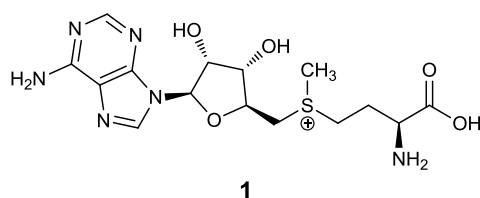
**Figure 2.** (a) Conversion of catechol (3) to guaiacol (4) by COMT. (b) Crystal structure of COMT from *Rattus norvegicus* (PDB accession code: 1H1D) showing classic Class 1 MT fold and SAM (shown in blue) in the active site. Helices are shown in red; strands in yellow and flexible loops in green.<sup>[59]</sup>

Out of the 6 motifs of secondary structure listed above, the most well known is ‘Motif I’, a glycine rich region near the N-terminus, which is known to form part of the binding pocket for the methionine portion of SAM.<sup>[55]</sup> Additionally, there are conserved acidic residues in  $\beta$ -strands 1, 2 and 3 which have all been identified to form part of the SAM binding site either directly or *via* a through-water H-bonding network.<sup>[57]</sup>

### 1.3 S-adenosyl-L-methionine (SAM)

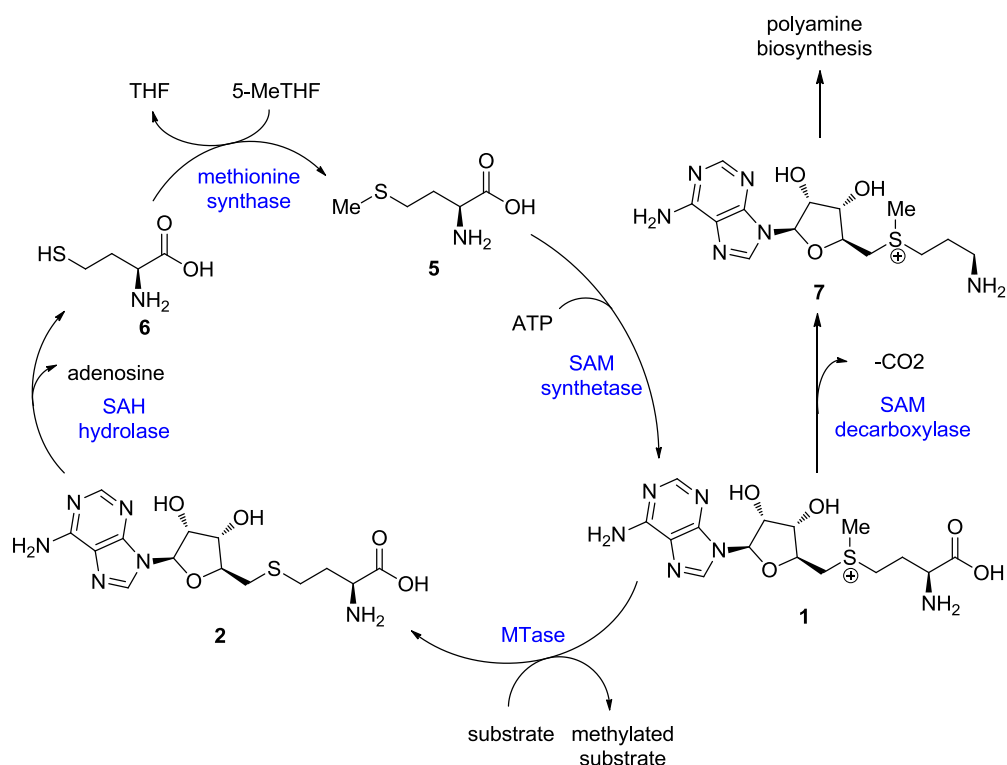
S-adenosyl-L-methionine (SAM, 1, **Figure 3**) is the most common cofactor for MTs and the second most abundant cofactor in Nature after adenosine triphosphate (ATP).<sup>[61]</sup> SAM was first identified in 1953 by Cantoni, who recognised that methyl transfer took place in the absence of ATP.<sup>[62]</sup> It is a conjugate of adenosine and methionine, in which the reactivity originates from the sulfonium centre bearing a pendant electrophilic methyl group originating from methionine.





**Figure 3.** Structure of SAM.

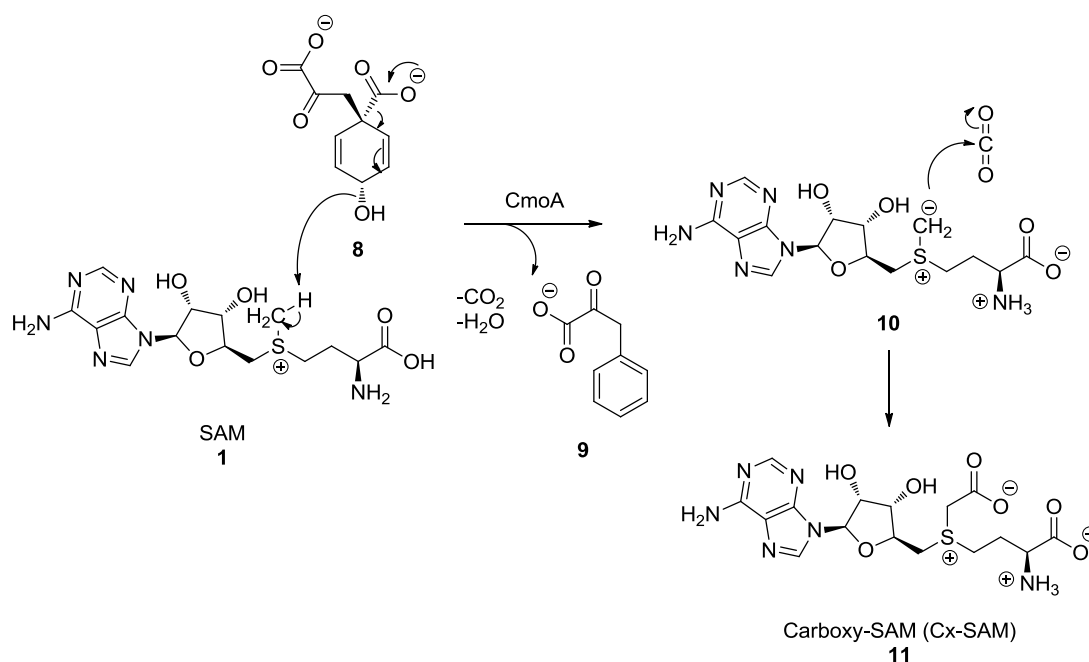
Although it is most well known for its role as a cofactor for SAM dependent MTs, SAM is also involved in other metabolic pathways, such as in the synthesis of the polyamines derived from spermidine (**Scheme 2**).<sup>[63]</sup> SAM is synthesised from methionine (Met, **5**) and adenosine triphosphate (ATP) by SAM synthetase with the loss of the triphosphate group. Following methyl transfer to a substrate by a MT, *S*-adenosyl homocysteine (SAH, **2**) is broken down by SAH hydrolase to provide adenosine and homocysteine (**6**). The homocysteine is methylated by methionine synthase using 5-methyltetrahydrofolate (5-MeTHF) as the methyl donor.<sup>[64]</sup> This regenerates methionine once more, which can be converted into another molecule of SAM by SAM synthetase to complete the cycle (**Scheme 2**). SAM is also extensively involved in the biosynthesis of natural products, such as the



**Scheme 2.** Biosynthesis and metabolism of SAM. 5-MeTHF: 5-methyltetrahydrofolate, THF: tetrahydrofolate.

polyamine pathway, as indicated in **Scheme 2**.<sup>[57]</sup> Following decarboxylation by SAM decarboxylase, the pendant aminopropyl group of product **7** can be transferred to an amine group by a polyamine synthase.

In a recent development into the research of SAM and its derivatives, the metabolite carboxy-SAM (Cx-SAM, **11**) has been identified.<sup>[29]</sup> Cx-SAM has been proposed to be formed from the deprotonation of SAM by prephenate (**8**), followed by the addition of the resulting carbanion (**10**) into carbon dioxide and is catalysed by the enzyme CmoA (**Scheme 3**).



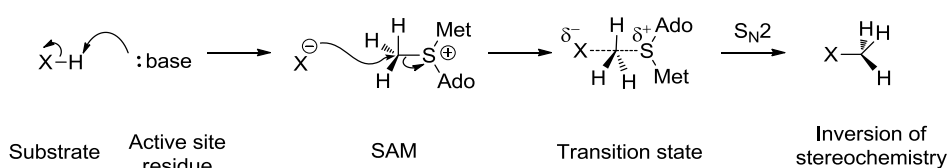
**Scheme 3.** Biosynthesis of carboxy-SAM (Cx-SAM, **11**), catalysed by CmoA.<sup>[29]</sup>

Cx-SAM is of interest to the chemical biology community, as it is used by the Cx-SAM dependent MT CmoB to transfer a carboxymethyl group to the wobble-position of transfer-ribonucleic acid (tRNA), thereby regulating the translation of the genetic code.<sup>[65]</sup> Additionally, this transformation is also of interest from a synthetic chemistry perspective, as chemical reagents to perform carboxymethylations, such as iodoacetic acid, are highly toxic and often provide a complex mixture of products.

## 1.4 Mechanism of methyl transfer

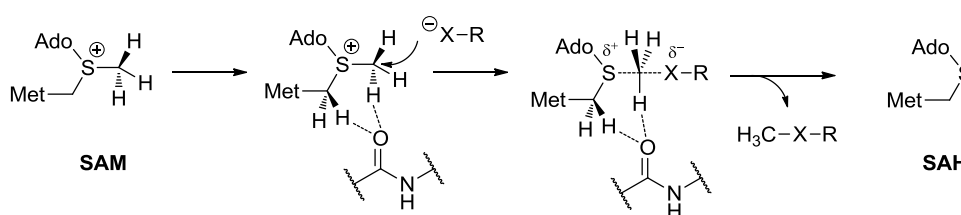
Central to developing methyltransferases as synthetically useful biocatalysts is an understanding of the reaction mechanism of methyl transfer. Indeed, some classes of MTs,

such as DNA MTs, have been well studied and the mechanism of methyl transfer is well understood.<sup>[66]</sup> For the large majority of MTs, methyl transfer reactions are  $S_N2$  in character, proceeding *via* a trigonal planar transition state and with inversion of stereochemistry at the transferred carbon centre.<sup>[67,68]</sup> Generally, this reaction requires activation of the substrate by a basic residue or Lewis acidic metal ion within the active site to provide a suitable nucleophile to attack the carbon centre of the pendant *S*-methyl group of the cofactor (**Scheme 4**). The regioselectivity of the reaction is determined by the position of the substrate in the active site relative to the pendant methyl group of SAM, which is typically within 3 Å of the acceptor atom.<sup>[69]</sup>



**Scheme 4.** General mechanism for methyl transfer by SAM-dependent methyltransferases. Alternatively, deprotonation of X-H may occur after nucleophilic attack. This is likely to be substrate dependent. Ado: adenosine. Met: methionine.

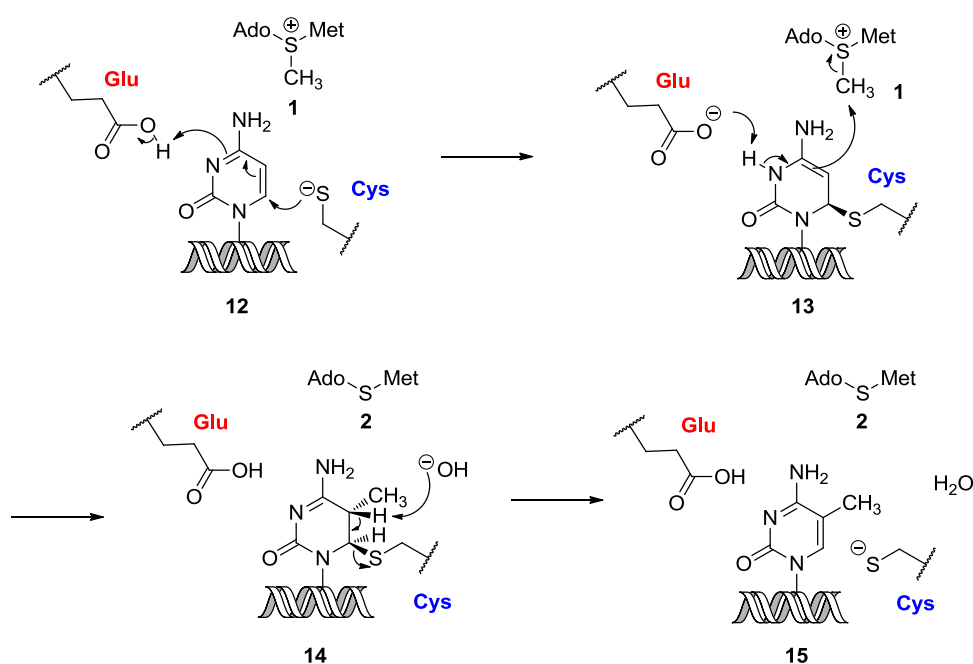
Horowitz *et al.* have recently proposed a universal mechanism for methyl transfer involving unusually strong carbon-oxygen hydrogen bonds between active site residues and SAM, which were found to be conserved over a range of MT classes.<sup>[70]</sup> These hydrogen bonds restrict movement of the SAM methyl group, thereby aiding the linear arrangement of  $R_2S-CH_3-X$  during methyl transfer (**Scheme 5**). The strongest hydrogen bonds were observed with backbone carbonyl groups, although tyrosine hydroxyl groups were also found to contribute.



**Scheme 5.** CH-O hydrogen bonding assisted mechanism for methyl transfer by SAM-dependent MTs as proposed by Horowitz *et al.*<sup>[70]</sup> H-bonding between SAM and a backbone carbonyl group of the enzyme is shown here, however hydrogen bonding between SAM and tyrosine hydroxyl groups has also been proposed.

### 1.4.1 DNA MTs

The mechanism of methyl transfer by DNA MTs has been extensively studied due to involvement of these enzymes in epigenetic regulation. Three classes of DNA MTs are known: m6A *N*-MTs, m4C *N*-MTs and m5C *C*-MTs. From a mechanistic point of view, m5C MTs are particularly interesting, as they catalyse the formation of a  $sp^2$ - $sp^3$  C-C bond, which can be difficult to achieve chemically. All members of the m5C MT family operate *via* a highly conserved mechanism involving an active site Cys residue found in the conserved ‘Motif IV’ region. As shown in **Scheme 6**, the mechanism is initiated by the nucleophilic attack of a cysteine residue at C-6 of the cytosine ring (**12**). The resulting resonance stabilised, covalently linked intermediate **13** now has a nucleophilic centre at C-5, which attacks the electrophilic methyl group of SAM in the key C-C bond forming step to provide intermediate **14**. Finally, a  $\beta$ -elimination of H-5 affords the enzyme-product complex **15**. The final proton abstraction step has been proposed to be carried out by a hydroxide ion provided by a proton wire through a water channel.<sup>[32]</sup>



**Scheme 6.** Mechanism of DNA methylation catalysed by m5C MTs such as hHal and DNMT1.<sup>[32,71]</sup>

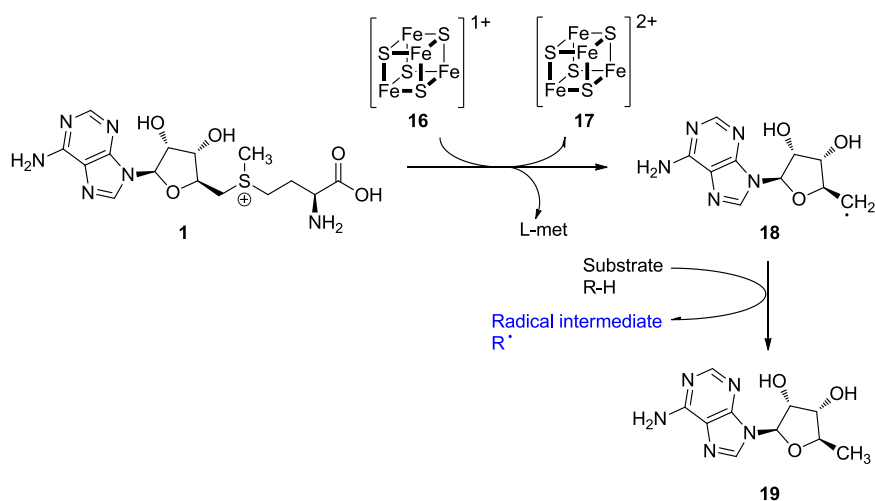
Schramm and coworkers have recently reported a detailed study into the transition state structure of human DNA MT 1 (DNMT1) based on a series of kinetic isotope effect (KIE) experiments and quantum mechanical (QM) calculations.<sup>[71]</sup> Contrary to previous studies, which have proposed  $\beta$ -elimination to be the rate limiting step (RLS), the authors found the

largest energy barrier to be associated with the methyl transfer step, which occurs in a stepwise fashion after the attack of Cys at C-6.<sup>[71]</sup>

### 1.4.2 Radical SAM MTs

Over the past decade, a new family of SAM-dependent enzymes has emerged that operate *via* a single electron transfer (SET) process, representing the largest known enzyme superfamily with more than 165 000 members, which are found in all three domains of life.<sup>[72]</sup> Radical SAM MTs lie within this superfamily, which is known as the ‘radical SAM superfamily’.<sup>[73–76]</sup> Members of the radical SAM superfamily are identified by the characteristic cysteine motif CXXXCXXC, known as the radical-SAM motif, where X is any residue other than Cys. This motif binds a [4Fe-4S] cluster which is also conserved across the superfamily and serves as a radical initiator in the catalytic mechanism. The [4S-4Fe] cluster is also present in other enzyme families which also use SET processes, such as cytochrome P450 enzymes.<sup>[77]</sup>

Four classes of radical SAM MTs (RSMTs) fall within the radical SAM superfamily, which have been determined according to the protein’s structure, cofactor requirement and predicted catalytic mechanism.<sup>[75]</sup> Each of Classes A-D has a distinct reaction mechanism, but share the requirement of two molecules of SAM for the transfer of one methyl group. The first is required for substrate activation and the second for methyl transfer.<sup>[74,75]</sup> The reaction is initiated by the formation of a 5’-deoxyadenosine radical (5’-dA<sup>•</sup>, **18**) *via* the

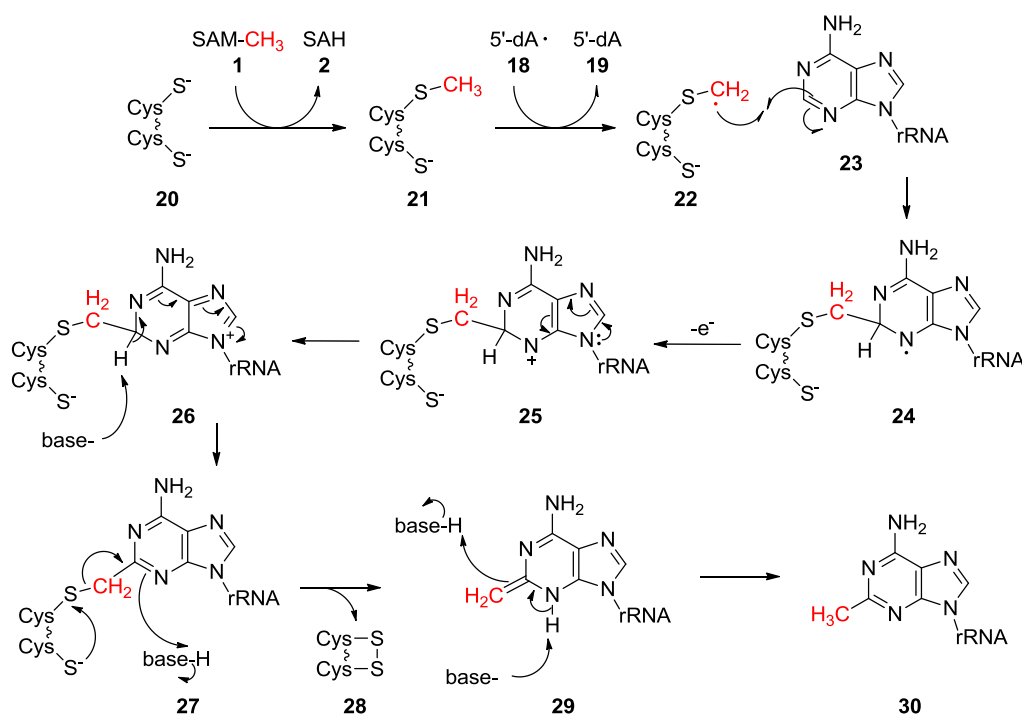


**Scheme 7.** Mechanism of the radical intermediate formation in [4Fe-4S] (**16**) mediated methyl transfer. L-met: L-methionine.

oxidation of the [4Fe-4S] cluster from 1<sup>+</sup> (**16**) to 2<sup>+</sup> (**17**) as shown in **Scheme 7**.

Class A comprises the enzymes that methylate ribosomal RNA at the C-2 and C-8 positions of adenosine *via* an sp<sup>2</sup>-sp<sup>3</sup> C-C bonding forming reaction and are represented by the two well characterised enzymes RlmN and Cfr.<sup>[74,78]</sup> Class B represents the largest and most versatile class of RSMTs. Unlike other RSMT classes, they have an N-terminal cobalamin (Vitamin B-12) binding domain in addition to the radical-SAM motif. This class of RSMTs methylates a broad range of substrates such as aliphatic and aromatic carbon centers and phosphinates.<sup>[79–81]</sup> Class C RSMTs contain a coproporphyrin III oxidase motif in addition to the radical-SAM motif and methylate heteroaromatic substrates.<sup>[82]</sup> Finally, Class D RSMTs, which currently only contains one family, employ folate as the methyl donor, as exemplified by MJ0619 from *Methanocaldococcus jannaschii*.

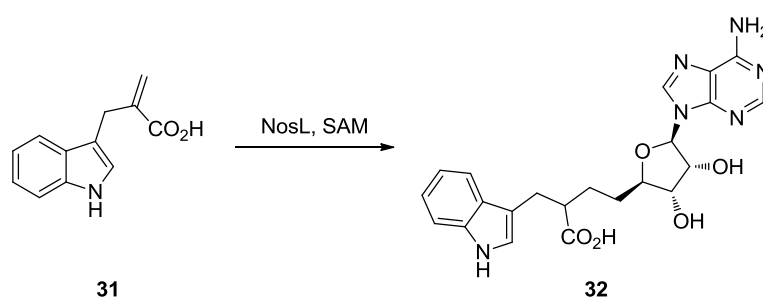
In the case of Class A RSMTs, such as RlmN, substrate activation has been proposed to occur after the transfer of a methyl group from SAM to an active site cysteine residue (**20**). A hydrogen radical is then abstracted from the resulting methylated residue **21** by 5'-dA<sup>•</sup> (**18**), which is generated according to the mechanism described in **Scheme 7**. This forms an enzyme-bound CH<sub>2</sub> radical (**22**), which is responsible for substrate activation in the carbon-



**Scheme 8.** Proposed mechanism of Class A RSMTs for the methylation of rRNA at the C-8 position.<sup>[83]</sup>

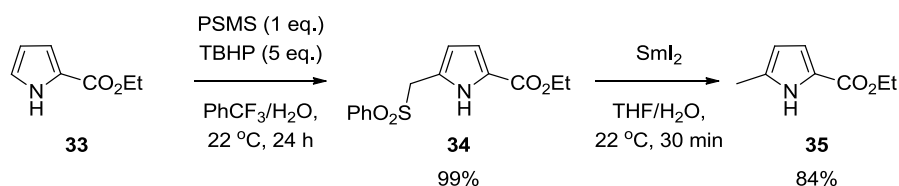
carbon bond forming step with the RNA substrate **23** to provide intermediate **24**. This is followed by a one electron oxidation to provide intermediate **25**. A series of base-catalysed proton transfer steps lead to the methylated product **30** following cleavage of the substrate-protein bond by intramolecular S<sub>N</sub>2 displacement (**Scheme 8**).<sup>[78,83,84]</sup>

Insights such as these into the mechanism of RSMTs have enabled the development of new methodologies for C-C bond formation in synthetic chemistry, as demonstrated by two notable examples. In the first, a change in the fate of the 5'-dA' radical (**18**) from hydrogen abstraction to radical addition in the radical SAM enzyme NosL was achieved by using a substrate analogue containing an olefin functional group (**31**), to provide the adenosylated product **32** (**Scheme 9**).<sup>[85]</sup>



**Scheme 9.** NosL catalysed reaction with non-natural substrate **31** to provide adenosylated product **32**.

The second example is a C-H methylation of heteroarenes inspired by RSMTs using zinc bis(phenylsulfonylmethanesulfinate) (PSMS) as the methyl donor, which was developed by Baran and coworkers.<sup>[86]</sup> In this system, decomposition of PSMS is initiated by the addition of *tert*-butylhydroperoxide (TBHP) to generate a carbon centered phenylsulfonyl methyl radical, which reacts with a heterocycle to provide a phenylsulfonyl methylated intermediate. Subsequent desulfonylation by treatment with SmI<sub>2</sub> reveals the methyl group to furnish the desired methylated product (**Scheme 10**). This methodology was demonstrated on a wide range of heterocyclic substrates, out of which the example with pyrrole **33** is shown in **Scheme 10**. In this example, excellent regioselectivity of the radical addition reaction was observed, affording 99% yield of phenylsulfonyl methylated intermediate **34**. Desulfonylation afforded the methylated product **35** in 84% yield.



**Scheme 10.** Radical C-H methylation of heterocycles using zinc sulfinate salts. PSMS: zinc bis(phenylsulfonylmethanesulfinate); TBHP: *tert*-butylhydroperoxide; THF: tetrahydrofuran.<sup>[86]</sup>

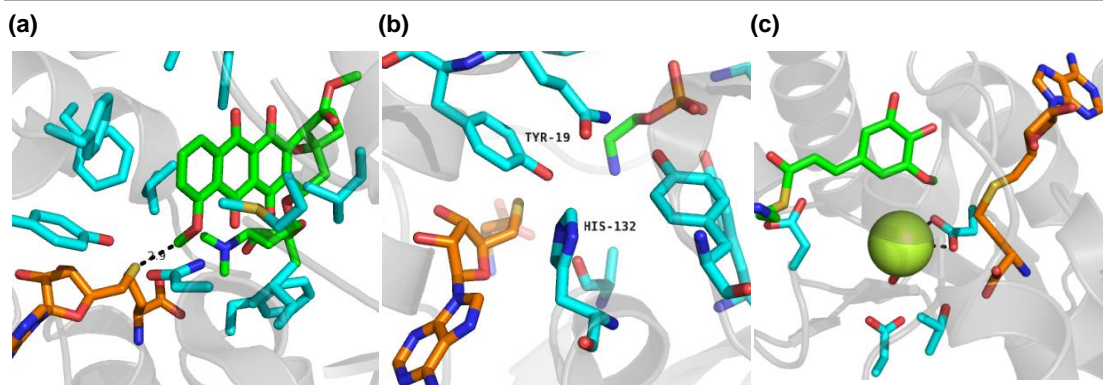
### 1.4.3 Small molecule MTs

Within the general  $S_N2$  mechanism outlined above, three mechanistically distinct modes of catalysis have been proposed for small molecule MTs: proximity and desolvation (PD), general acid/base mediated catalysis and metal dependent mechanisms. In PD catalysis, the transmethylation reaction does not involve direct participation of active site residues, but rather relies on the alignment of reactants and exclusion of water within the active site to facilitate the transmethylation reaction. One example of a small molecule MT operating by PD effects is DnrK, which is an *O*-MT involved in the biosynthesis of the anti-tumor agent daunorubicin (**Figure 4a**).<sup>[87]</sup> In a mutational analysis study of a putative catalytic base of DnrK, no significant change in catalytic activity was observed, leading the authors to conclude that rate enhancement of the transmethylation reaction relative to the enzyme free control was mainly due to orientational and proximity effects.<sup>[87]</sup>

Conversely, acid/base mediated catalysis involves a basic residue within the active site that is responsible for activating the substrate *via* deprotonation, which generates the nucleophile for nucleophilic attack at the electrophilic methyl group of SAM. In some examples, a series of active site residues can work in tandem to form a proton shuttle system. An example of this is phosphoethanolamine *N*-MT from *Plasmodium falciparum*, which has been proposed to operate *via* H132 mediated deprotonation of Y19, which in turn deprotonates the substrate in preparation for nucleophilic attack of SAM (**Figure 4b**).<sup>[88]</sup>

Metal-dependent MTs fall into two further categories; those that rely on a divalent metal cation for anchoring the substrate into the active site and otherwise operate *via* an acid/base mediated mechanism; and those that use the metal ion to perturb the  $pK_a$  of the substrate (which is almost exclusively a plant-derived phenol), thereby promoting loss of the proton to generate a nucleophilic phenolate anion. The latter is exemplified by caffeoyl





**Figure 4.** Examples of the three types of catalysis found in SAM dependent MTs. **(a)** Proximity and desolvation, exemplified by the *O*-MT DnrK1 (PDB: 1TW2). **(b)** Acid/base mediated catalysis, exemplified by phosphoethanolamine *N*-MT (PDB: 4R6X). **(c)** Metal dependence, exemplified by CCoAOMT (PDB: 1SUI),  $\text{Mg}^{2+}$  shown as green sphere.

coenzyme A 3-*O*-methyltransferases (CCoAOMTs), which contain a well-ordered active site  $\text{Mg}^{2+}$  binding site, comprising an octahedral arrangement of hydroxy and carboxy ligands, of which one is the 3-hydroxy group of the caffeoyl substrate. The  $\text{Mg}^{2+}$  ion has been proposed to mediate deprotonation of the substrate, generating a phenolate which is held by the metal cation in close proximity to the reactive methyl group on SAM, thereby facilitating the transmethylation reaction (**Figure 4c**).<sup>[89]</sup>

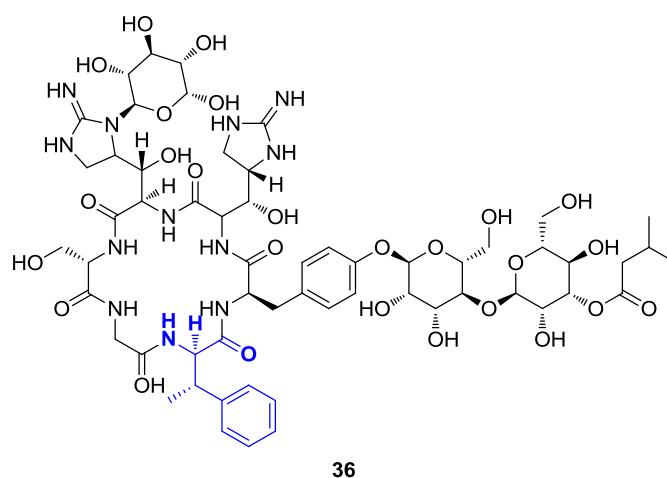
## 1.5 C-MTs

As discussed above, MTs are of interest to the synthetic organic chemistry community for their potential as biocatalysts for regiospecific methylation reactions. *C*-MTs are particularly appealing due to the importance of C-C bond forming reactions in organic chemistry. Many classes of enzyme are known to catalyse C-C bond formation, including aldolases (EC 4.1.2), transketolases (EC 2.2.1), hydroxynitrile lyases (EC 4.1.2.47), cyclases (ECs 1.14, 1.21, 2.3 and 4.6.1) and methyltransferases (EC 2.1.1), out of which small molecule MTs remain one of the least well understood and developed for use in organic synthesis.<sup>[90,91]</sup> *C*-MTs can be broadly categorised into enzymes methylating  $\text{sp}^3$  (aliphatic) or  $\text{sp}^2$  (aromatic) carbon centres, out of which most known *C*-MTs fall into the aromatic *C*-MT group.

### 1.5.1 Aliphatic C-MTs

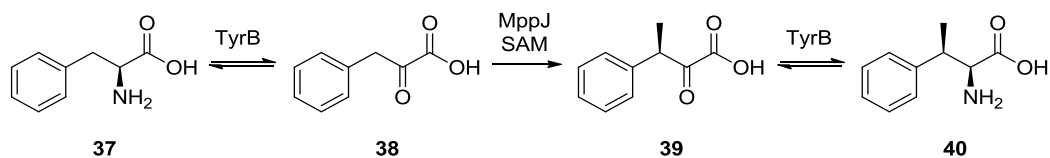
Aliphatic *C*-MTs have been identified in the biosynthesis of some natural products, such as the antibiotics indolmycin and mannopeptimycin.<sup>[92,93]</sup> Out of these, the most well

characterised is the biosynthesis of  $\beta$ -methylphenylalanine, which is a key intermediate in the synthesis of the glycopeptide antibiotic mannopeptimycin (**36**, **Figure 5**).



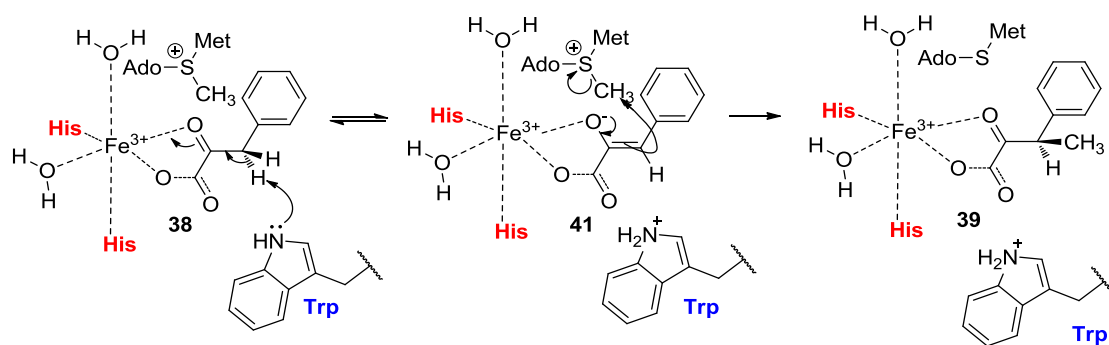
**Figure 5.** Structure of mannopeptimycin (**36**) with the (2*S*, 3*S*)- $\beta$ -methylphenylalanine motif highlighted in blue.

In the biosynthetic pathway of **36**, phenylalanine (Phe, **37**) is converted to the  $\alpha$ -keto acid intermediate **38** by the transaminase TyrB. This increases the acidity of the benzylic protons and activates the position for enantiospecific methylation by the *C*-MT MppJ to provide **39**. The methylated intermediate is then converted to the target amino acid **40** by the transaminase TyrB (**Scheme 11**).



**Scheme 11.** Biosynthesis of (2*S*, 3*S*)- $\beta$ -methylphenylalanine (**40**), a key intermediate in the synthesis of mannopeptimycin (**36**).

MppJ represents the first ferric ion-dependent *C*-MT to be characterised.<sup>[94]</sup> In the mechanism proposed by Liu and coworkers,  $\text{Fe}^{3+}$  acts as a Lewis acid to promote enolate formation *via* deprotonation at the benzylic position of **38** by an active site tryptophan residue. The resulting intermediate **41** is then methylated *via* nucleophilic attack of the enolate through the benzylic carbon to the electrophilic methyl group of SAM to provide product **39** (**Scheme 12**).<sup>[94]</sup>

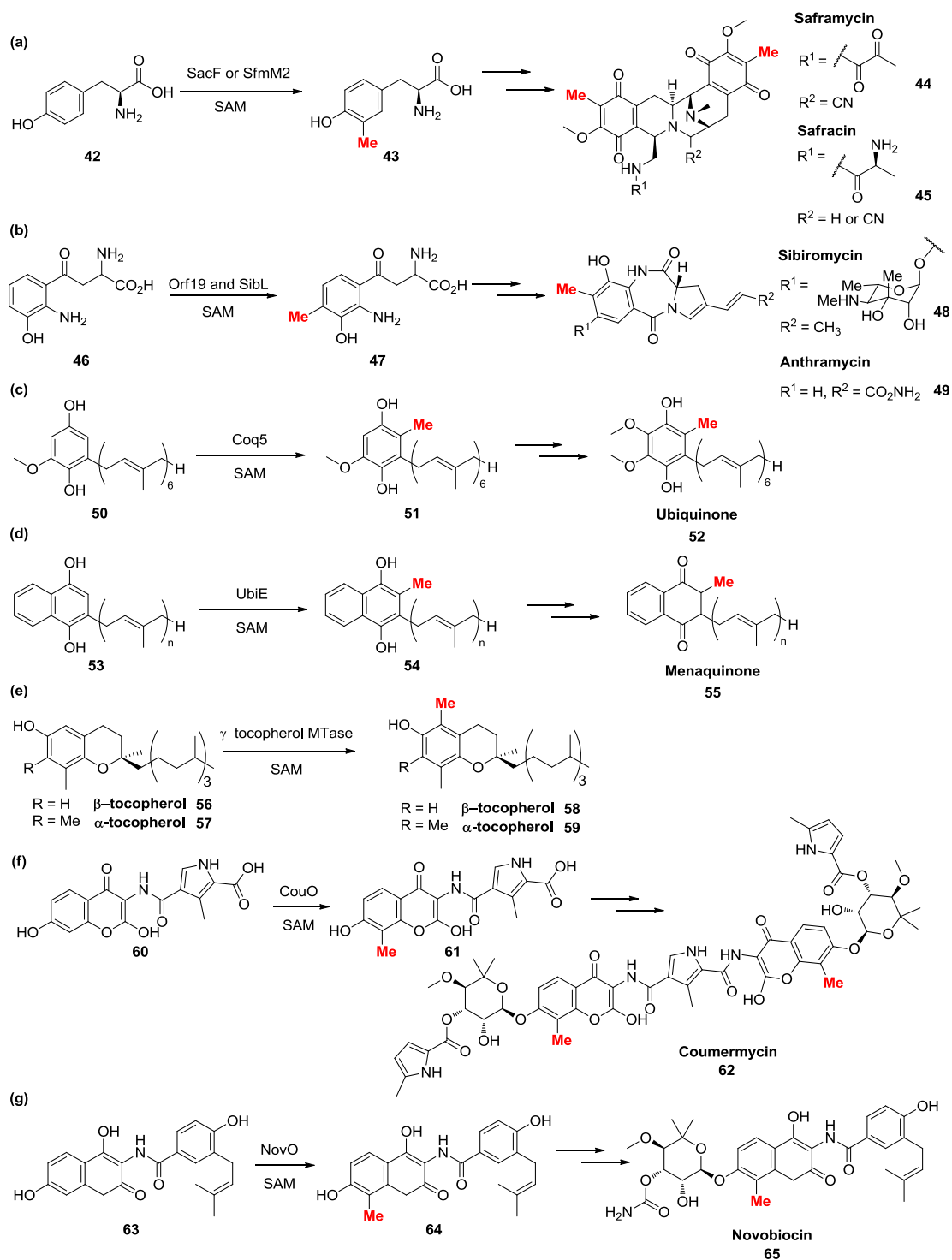


**Scheme 12.** Methyl transfer mechanism at the benzylic position of Phe derivative **38** catalysed by MppJ as proposed by Liu *et al.*<sup>[94]</sup>

With reference to the classification of the mechanism of methyl transfer in small molecule MTs described above, MppJ operates *via* a metal-dependent acid/base mediated mechanism.

### 1.5.2 Aromatic C-MTs

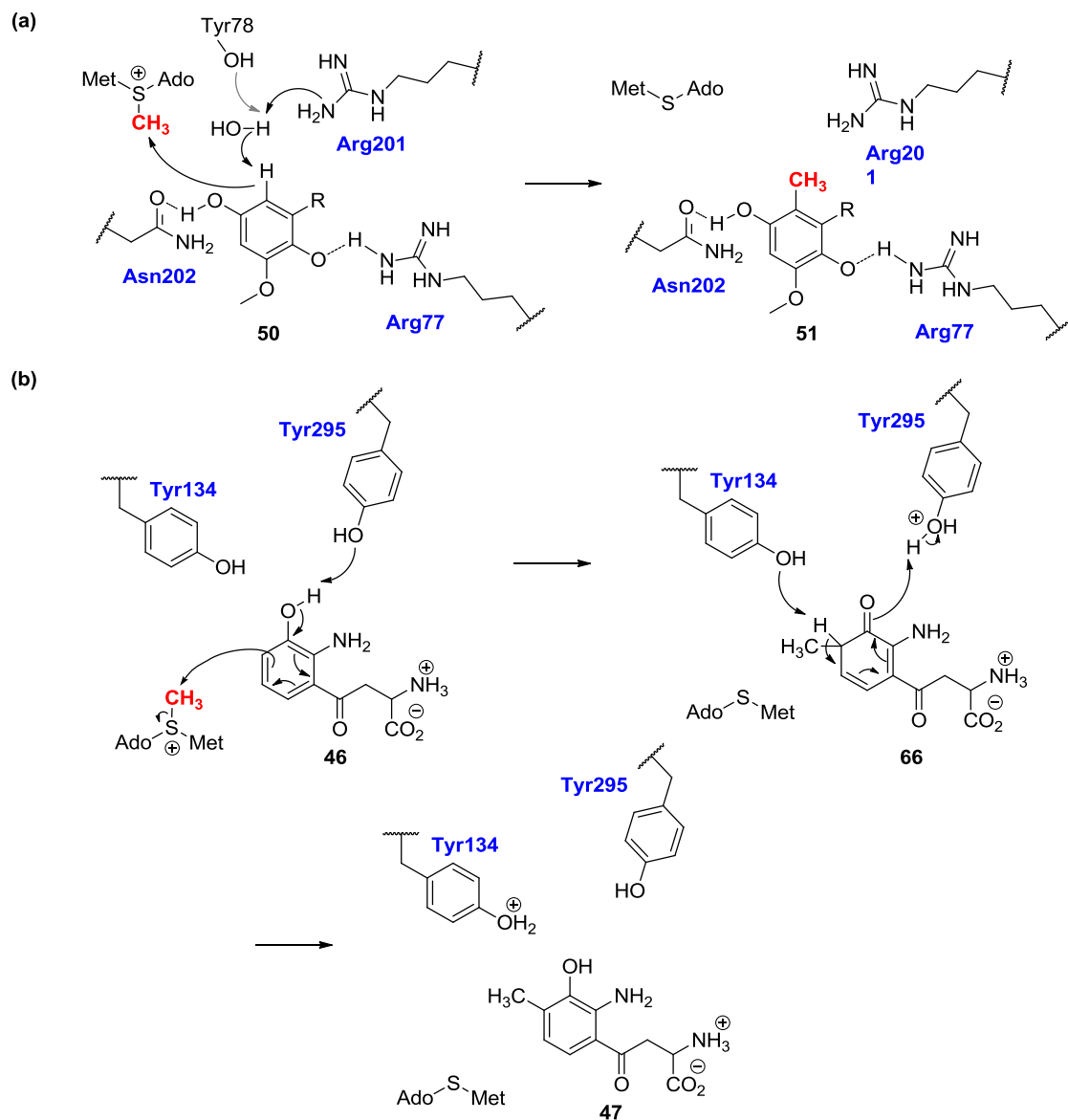
Aromatic MTs are the most well known of the small molecule C-MTs. A number of aromatic C-MTs have now been identified from the biosynthetic pathways of natural products and are summarised in **Figure 6**. SacF and SfmM2 have been implicated in the biosynthesis of the antibiotics saframycin (**44**) and safracin (**45**) by *Streptomyces lavendulae* and *Pseudomonas fluorescens*, respectively.<sup>[95,96]</sup> Both enzymes are highly specific for the methylation of tyrosine (**42**), with both D- and L-enantiomers being accepted (**Figure 6a**).<sup>[48]</sup> Orf19 (*Streptomyces refuineus*) and SibL (*Streptosporangium sibiricum*) are involved in the biosynthesis of benzodiazepine antibiotics sibiromycin (**48**) and anthramycin (**49**), respectively (**Figure 6b**).<sup>[97,98]</sup> Moreover, both enzymes have been demonstrated to be effective for the regiospecific methylation of amino acid **46** *in vitro*.<sup>[48,99]</sup> Coq5 catalyses the methylation of 2-methoxy-6-poly-prenyl-1,4-benzoquinone (DDMQH<sub>2</sub>, **50**) to 2-methoxy-5-methyl-6-poly-prenyl-1,4-benzoquinone (DMQH<sub>2</sub>, **51**) in the biosynthesis of coenzyme Q (Ubiquinone, **52**), which is a component of the electron transport chain in most eukaryotic cells (**Figure 6c**).<sup>[100,101]</sup> The related enzyme, UbiE catalyses the methylation of the related substrate demethylmenaquinol (DMKH<sub>2</sub>, **53**) (**Figure 6d**).<sup>[102]</sup> The enzyme  $\gamma$ -tocopherol MT is involved in the biosynthesis of both  $\alpha$ - and  $\beta$ -tocopherol (**58** and **59**, respectively), which are exclusively synthesised by photosynthetic organisms as lipophilic anti-oxidants (**Figure 6e**).<sup>[103]</sup> Finally, the C-MTs NovO and CouO, which are the subject of this thesis, have been isolated from the biosyntheses of the antibiotics novobiocin (**65**) and coumermycin (**63**) by *Streptomyces spheroides* and *Streptomyces rishiriensis*, respectively, and have been shown to methylate coumarin and dihydroxynaphthalene scaffolds *in vitro* (**Figure 6f-g**).<sup>[104–107]</sup>



**Figure 6.** Summary of functionally characterised aromatic small molecule C-MTs.

A feature common to all of these transformations is the presence of a hydroxy group adjacent to the site of methylation. Whilst a detailed mechanistic understanding of these enzymes is lacking, the crystal structures of SibL in complex with SAH and 3HK (PDB: 4U1Q)<sup>[99]</sup> and

Coq5 in complex with SAM from which a substrate binding model was generated (PDB: 4OBW)<sup>[101]</sup> have provided a basis for putative catalytic mechanisms for both enzymes (**Scheme 13**). In the case of Coq5, the authors have proposed that R201 or Y78 deprotonates an active site water molecule to provide a hydroxide ion which is responsible for deprotonating the substrate (**50**), providing a nucleophilic centre for subsequent attack of the

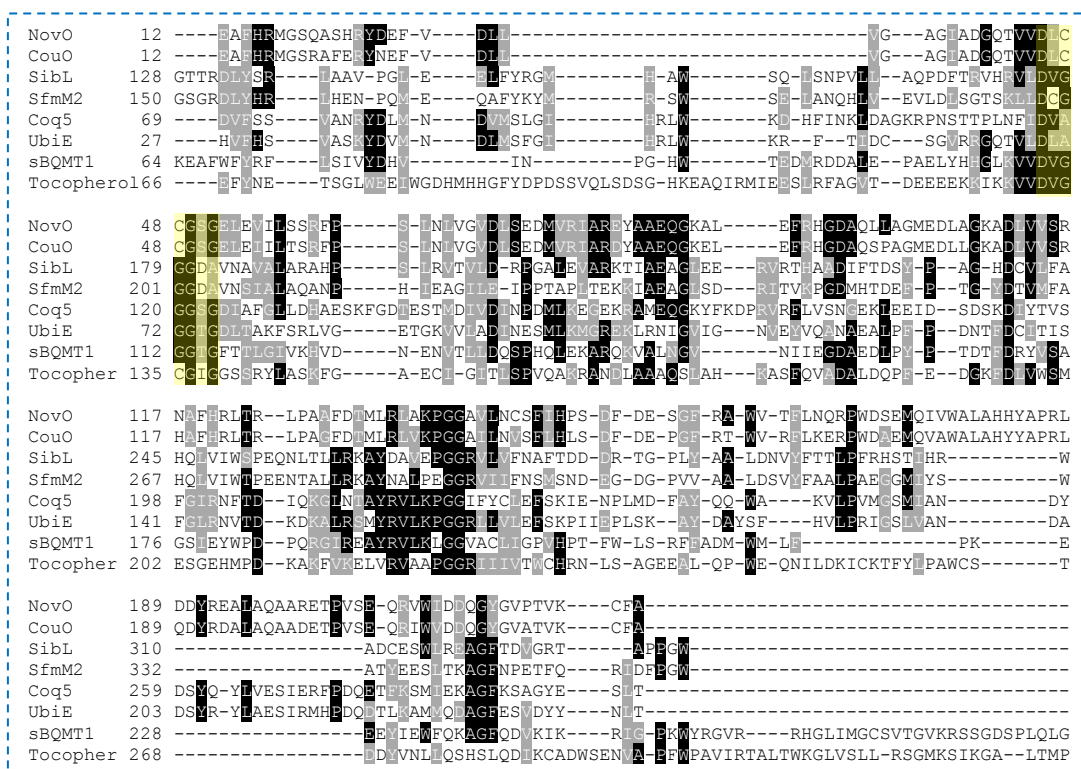


**Scheme 13.** Putative mechanisms for methyl transfer by the small molecule C-MTs (a) Coq5 (grey arrow signifies an alternative proposed mechanism)<sup>[101]</sup> and (b) SibL.<sup>[99]</sup>

electrophilic methyl group of SAM to provide product **51** (**Scheme 13a**).<sup>[101]</sup> However, given the approximate  $pK_a$ s of the protons involved (protonated Arg: 12.48; protonated Tyr: 10.07; water: 15.7)<sup>[108]</sup>, this mechanism appears unfeasible<sup>[108]</sup> and would require further probing by

mutational analysis, kinetics and/or QM calculations to be verified.<sup>[32,71,109,110]</sup> In the example of SibL, Chen and coworkers proposed activation of substrate **46** *via* deprotonation of the phenolic proton by Y295, followed by nucleophilic attack of the adjacent carbon atom at the electrophilic methyl group of SAM (**Scheme 13b**). Rearomatisation is then achieved by deprotonation of intermediate **66** by Y134 to provide the product **47**.<sup>[99]</sup> With reference to the three general mechanisms for methyl transfer in small molecule MTs, this falls into the acid/base mediated category.

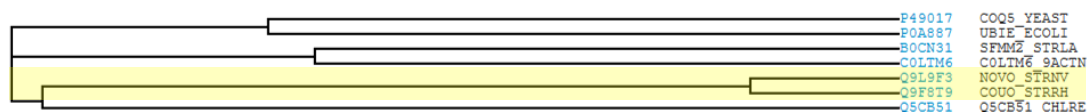
As demonstrated by Coq5 and SibL, the active site residues are poorly conserved between small molecule C-MTs. The extent of this is demonstrated by a multiple sequence alignment (MSA) of these enzymes (**Figure 7**). In general, there is very poor alignment of the sequences, although the previously discussed ‘Motif 1’ region is moderately well conserved (**Figure 7**, highlighted yellow).



**Figure 7.** Section of multiple sequence alignment (MSA) of small molecule C-MTs. ‘Motif 1’ region highlighted in yellow.

This is promising from a biocatalysis perspective, as each of these small molecule C-MTs may offer a unique entry point into a toolbox of biocatalysts for C-C bond formation.<sup>[48]</sup> Out of the enzymes discussed above, NovO and CouO are of particular interest. As shown in the

phylogenetic tree in **Figure 8**, these enzymes are very distinct from the rest of the group. Indeed, a basic local alignment search tool (BLAST) search of NovO reveals an 85%

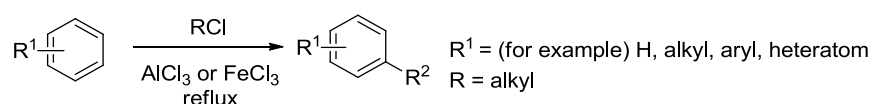


**Figure 8.** Phylogenetic tree of aromatic small molecule C-MTs. Position of NovO and CouO highlighted in yellow.

sequence identity match to CouO, whilst all other matches are <35% sequence identity.<sup>[111]</sup> Although no crystal structure of NovO or CouO has been reported prior to this project, preliminary biochemical characterisation of NovO indicated that it is a homodimer which operates *via* an acid/base mediated mechanism, with H15 being proposed as the catalytic base.<sup>[112]</sup> With regards to CouO, the crystallisation and preliminary analysis of the X-ray diffraction data has been reported, however it was not possible for structure elucidation due to the very low homology of CouO to other structures in the PDB.<sup>[68]</sup>

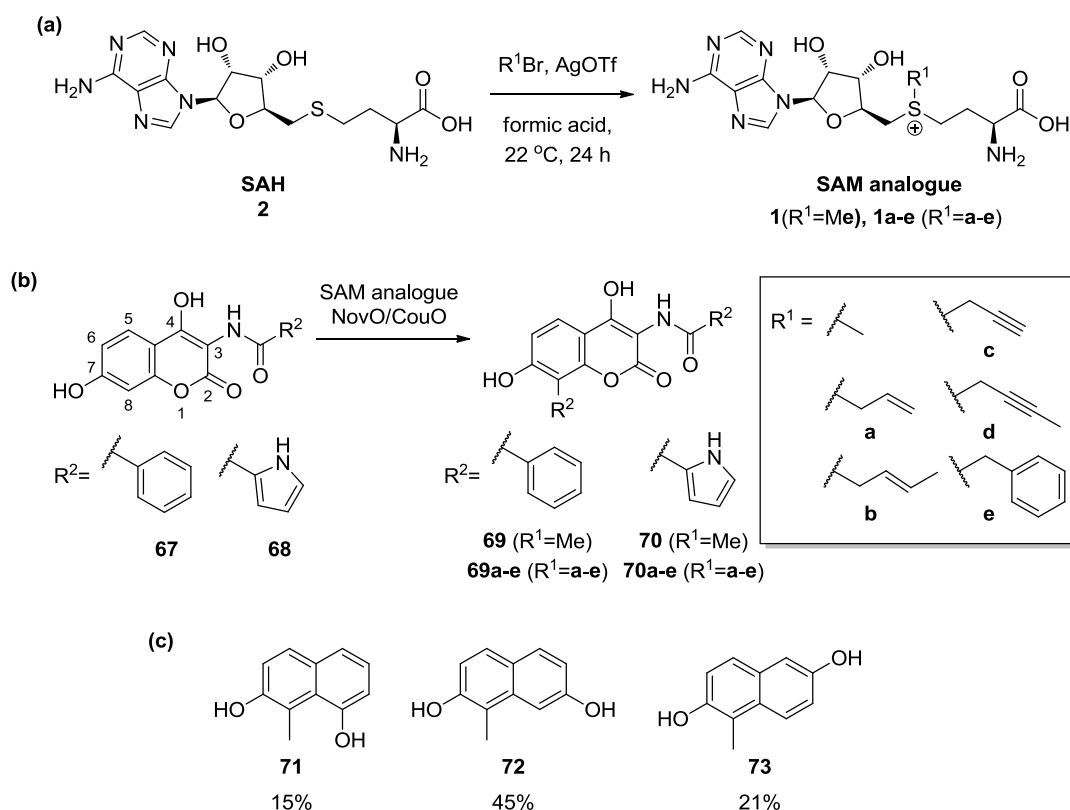
## 1.6 Friedel-Crafts alkylation using NovO and CouO

Due to the prevalence of alkyl groups in medicinal chemistry, methods for their installation on small molecules in a site-specific manner are highly sought after.<sup>[42,113]</sup> Traditionally, the alkylation of aromatic rings is carried out chemically by the Friedel-Crafts reaction, in which an electrophilic alkyl cation is generated from the corresponding alkyl halide and a Lewis acid, such as  $\text{AlCl}_3$  (**Scheme 14**).<sup>[114]</sup> However, obtaining the monoalkylated product with the correct regiochemistry is often problematic. Firstly, the alkylated product is more reactive to further alkylation than the starting material due to the positive inductive effect of alkyl substituents.<sup>[113]</sup> Additionally, rearrangements of the carbocation intermediate to favour more stable secondary and tertiary carbocations are common, leading to a mixture of alkylated products which can be difficult to separate.<sup>[115]</sup> Furthermore, harsh reaction conditions may be unsuitable for some substrates and the disposal of the metal salts generated can make this a very costly process. Therefore, there is a need for the development of environmentally benign and regioselective alternative methodologies.<sup>[113]</sup>



**Scheme 14.** General scheme for the Friedel-Crafts alkylation.

To address some of the limitations of the Friedel-Crafts alkylation, Gruber and coworkers have pioneered the use of NovO and CouO as biocatalysts for the *C*-alkylation of small molecules.<sup>[107]</sup> A library of non-natural SAM analogues (**1a-e**), which varied in the nature of the electrophilic alkyl group at the sulfonium centre, were synthesised by alkylation of SAH (**2**) using the corresponding alkyl bromide in formic acid in the presence of a Lewis acid (**Scheme 15a**). These analogues were then used as cofactors in the MT reaction with NovO or CouO, to transfer the corresponding alkyl group to aminocoumarin substrates **67** and **68** (**Scheme 15b**). Additionally, a range of dihydroxynaphthalene substrates were also accepted by CouO to provide the methylated products **71-73** (**Scheme 15c**).<sup>[107]</sup>



**Scheme 15.** General reaction scheme for biocatalytic Friedel-Crafts type alkylation using CouO or NovO and non-natural cofactors.

This work provides a promising foundation for developing a platform for biocatalytic Friedel-Crafts alkylation reactions. However, the methodology is currently limited by the relatively narrow substrate scope of NovO and CouO and by the availability of the non-natural SAM-analogues, the synthesis of which has only been carried out on a small scale and is not amenable to large scale synthesis. Additionally, the low  $k_{cat}$  of NovO and CouO (NovO:  $0.008 \text{ s}^{-1}$  and CouO:  $0.004 \text{ s}^{-1}$  for the natural substrates **63** and **60**, respectively)<sup>[104]</sup>



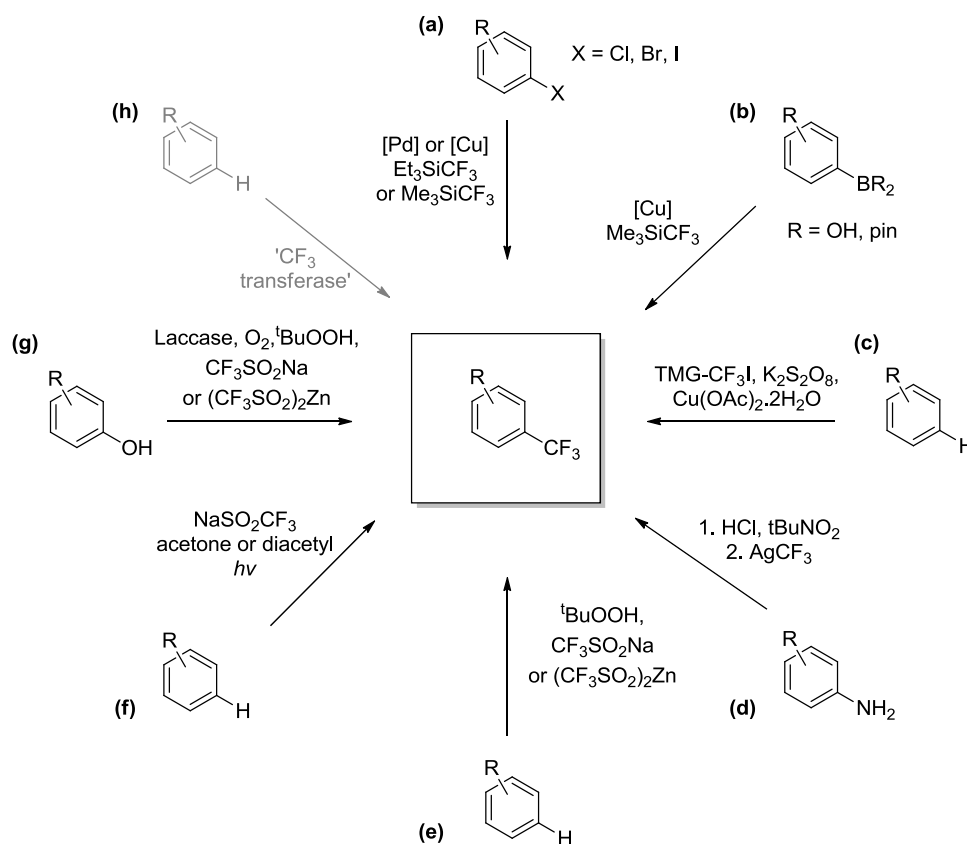
and their poor stability to organic solvents, temperature and pH<sup>[112,116]</sup> currently render these enzymes too underdeveloped for routine use in organic synthesis.

## 1.7 Aromatic trifluoromethylation

Biocatalytic aromatic methylation has been demonstrated as an alternative approach to the Friedel-Crafts alkylation reaction. However, there is currently no literature precedent for an analogous biocatalytic aromatic trifluoromethylation, which could offer a mild alternative to the existing chemical alternatives. The trifluoromethyl group is widely recognised as a ‘privileged motif’ in medicinal chemistry. This is because fluorinated molecules often show enhanced physicochemical, adsorption, distribution, metabolism and excretion properties, in addition to an increased binding efficiency and selectivity to target proteins.<sup>[117]</sup> A major challenge in medicinal chemistry is overcoming high clearance rates. Drug clearance is, in part, facilitated by drug metabolism by cytochrome P450 oxidases, as oxidation of small molecules facilitates renal excretion due to increased polarity and solubility of the compounds.<sup>[118]</sup> A common strategy to overcome this is the incorporation of an electron withdrawing group such as the trifluoromethyl motif, as P450 oxidases have shown decreased activity towards electron deficient substrates.<sup>[117]</sup> Therefore, compounds bearing a trifluoromethyl group often have longer half lives than their non-fluorinated analogues, thereby increasing the time available for the drug to exert its therapeutic effect.<sup>[117]</sup> As such, methods of introducing fluorinated alkyl groups, in particular aromatic trifluoromethylation, have been the focus of extensive research efforts.<sup>[119–121]</sup>

One of the oldest and most widely used reagents in this area is (CF<sub>3</sub>)SiMe<sub>3</sub>, also known as Ruppert’s or the Ruppert-Prakash reagent.<sup>[122,123]</sup> This is a source of nucleophilic CF<sub>3</sub> and has been shown to react with a large range of electrophiles such as aldehydes, ketones, acid chlorides, imines and sulfones.<sup>[123]</sup> Furthermore, methodology has now been developed to use Ruppert’s reagent for transition metal-assisted aromatic trifluoromethylation of aryl halides (**Scheme 16a**).<sup>[124]</sup> This has also been extended to copper-mediated coupling with aryl boronic esters or acids by Buchwald *et al.* (**Scheme 16b**).<sup>[125]</sup> A much newer class of trifluoromethylation reagent has recently been developed by Ritter and coworkers, who reported the use of condensed phase, halogen bonded dimethylsulfoxide (DMSO) and tetramethylguanidine (TMG) adducts of CF<sub>3</sub>I in the presence of Cu<sup>2+</sup> for the direct C-trifluoromethylation of a variety of aromatic substrates (**Scheme 16c**).<sup>[126]</sup> To extend aromatic trifluoromethylation to anilines, Wang *et al.* have developed a silver-mediated one-pot procedure for the trifluoromethylation of an aniline *via* the diazonium salt (**Scheme**

**16d).**<sup>[127]</sup> A fourth class of trifluoromethylating reagent which has attracted considerable interest in recent years are trifluoromethylated sodium and zinc sulfinate salts, which decompose in the presence of a suitable initiator to generate trifluoromethyl radicals, which



**Scheme 16.** Approaches to aromatic trifluoromethylation. (a) Pd or Cu catalysed trifluoromethylation of aryl halides;<sup>[131,132]</sup> (b) Cu catalysed trifluoromethylation of aryl boronic esters or acids;<sup>[131]</sup> (c) Use of condensed phase halogen bonded CF<sub>3</sub>I-TMG (tetramethylguanidine) adducts for direct trifluoromethylation.<sup>[126]</sup> (d) Ag catalysed trifluoromethylation of anilines *via* the diazonium salt.<sup>[127]</sup> (e) Radical aromatic trifluoromethylation.<sup>[133,134]</sup> (f) Photoinduced trifluoromethylation.<sup>[135]</sup> (g) Laccase mediated trifluoromethylation of phenols.<sup>[136]</sup> (h) Hypothetical (shown in grey) biocatalytic trifluoromethylation using a small molecule C-MT.

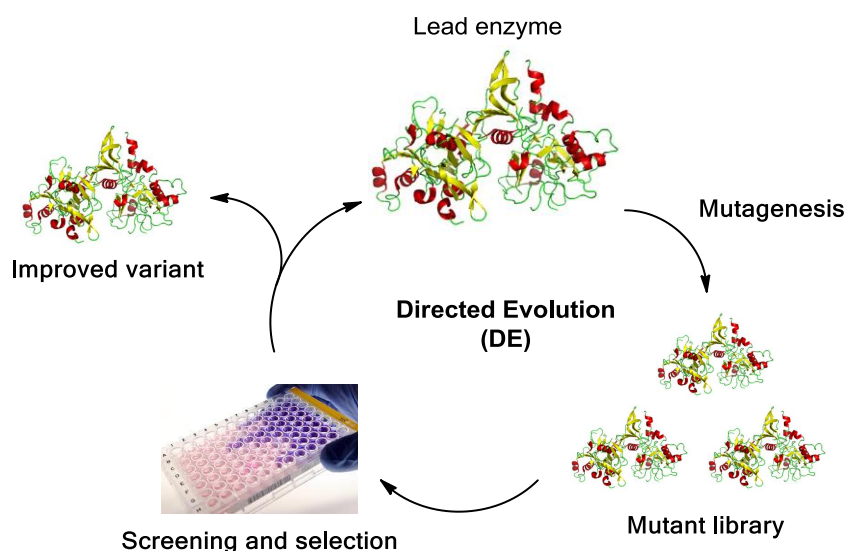
can then be trapped by the substrate in a C-C bond forming reaction.<sup>[120,128]</sup> To date, two main strategies towards the generation of these radicals have been employed. The first is the use of a peroxide, most often *tert*-butylhydroperoxide (TBHP), which decomposes in the presence of trace metals in the reaction to produce the radical initiator.<sup>[129]</sup> Whilst this procedure was first reported by Langlois and coworkers in 1994 for the synthesis of trifluoromethylated amino acids,<sup>[130]</sup> it has now been extended to the trifluoromethylation of arenes and heterocycles by Baran and coworkers, who have also developed analogous difluoromethyl and trifluoroethyl zinc sulfinate salts for direct difluoromethylation and

trifluoroethylation, respectively (**Scheme 16e**).<sup>[129,137]</sup> The second strategy for generation of CF<sub>3</sub> radicals is by photoredox catalysis, first demonstrated by MacMillan and coworkers in 2011.<sup>[118]</sup> Recently, however, it has been shown that this can also be achieved in the absence of a metal catalyst using visible light irradiation and acetone or diacetyl, providing a clean and operationally simple route to access a wide range of trifluoromethylated arenes with moderate to good regioselectivity (**Scheme 16e**).<sup>[135]</sup> Finally, in an interesting application of the first radical strategy, Kroutil and coworkers have demonstrated the use of a laccase to generate a radical cation of a phenol substrate, which is then trapped by a CF<sub>3</sub> radical generated from the peroxide mediated decomposition of a zinc sulfinate salt in a C-C bond forming reaction (**Scheme 16g**).<sup>[136]</sup> To date, this represents the first example of a biocatalytic C-trifluoromethylation reaction.

An alternative strategy to a biocatalytic C-trifluoromethylation methodology would be to use a C-MT in conjunction with a suitable trifluoromethylated cofactor (**Scheme 16h**). To date however, there have been no reports of the synthesis of such a cofactor, which would most likely be a S-trifluoromethylated analogue of SAM.

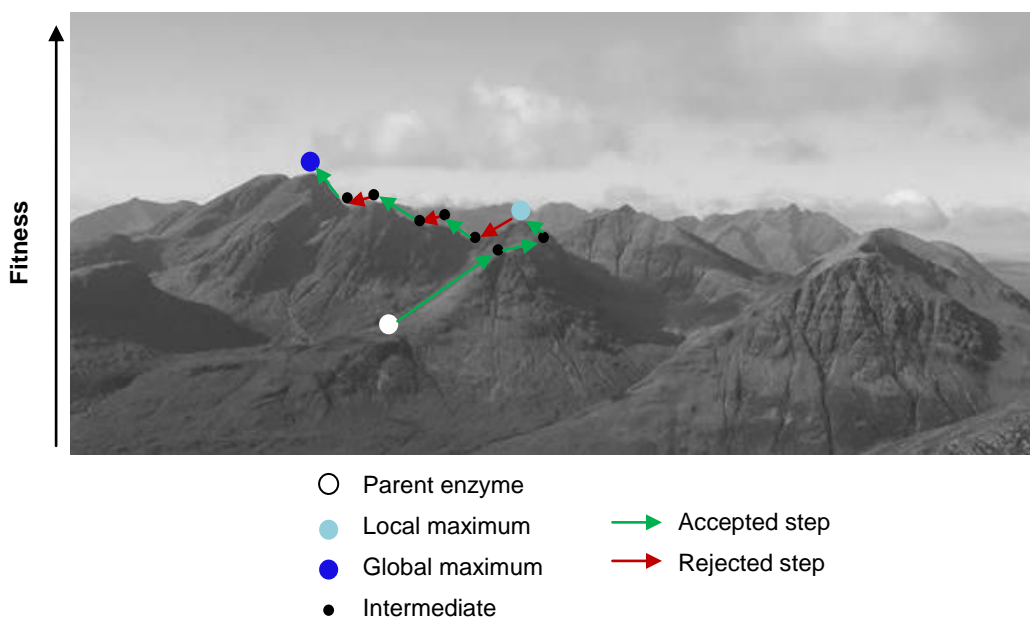
## 1.8 Directed evolution

The C-MTs NovO and CouO have potential to be useful biocatalysts for the regiospecific alkylation of small molecules.<sup>[48]</sup> However, their use is currently limited by their narrow substrate scope, low  $k_{cat}$  and instability towards organic solvents, high temperatures and non-neutral pHs.<sup>[112,116]</sup> Directed evolution (DE) of biocatalysts to identify variants of the WT enzyme harbouring beneficial mutations has emerged as a powerful strategy to address many of these limitations. Starting with the WT enzyme that one wishes to evolve, mutagenesis is carried out at the DNA level to introduce genetic diversity within a library of mutant genes. This library is translated into a corresponding library of proteins and screened or selected for functional variants harbouring beneficial mutations. When an improved variant is identified, it can be used directly as a biocatalyst, or used as the lead enzyme in a second round of evolution (**Scheme 17**).<sup>[138]</sup>



**Scheme 17.** General scheme for the workflow of a typical directed enzyme evolution project.

An important concept in the field of DE is navigation of the sequence space of the protein one wishes to evolve.<sup>[139]</sup> If all residues in a sequence are simultaneously randomised for any of the 20 naturally occurring amino acids, for a sequence of  $N$  residues there would be  $20^N$  possible sequences. Even for a small protein of 100 amino acids, this would result in  $\sim 1.3 \times 10^{130}$  different sequences, which is greater than the number of atoms in the known universe.<sup>[139]</sup> Therefore, it is necessary to navigate the sequence space intelligently in order to reduce the number of variants that require screening or selection to a number that is experimentally feasible. A popular analogy to visualise this is that of the ‘3D sequence landscape’, in which navigating the fitness landscape of a sequence is likened to navigating a mountain range (**Figure 9**). Starting from the parent enzyme (**Figure 9**, white circle), it is straightforward to navigate to the local maximum, through a series of steps each leading to an enzyme with improved fitness with respect to the starting point (**Figure 9**, green arrows). Once the local maximum is reached, however, it is not possible to reach the global maximum without encountering disadvantageous mutations, which would typically be rejected in a DE project in which only improved variants are used for subsequent rounds of evolution (**Figure 9**, red arrows). Consequently, finding (and recognising) the global maximum in a fitness landscape of a given enzyme sequence space is non-trivial and developing strategies to navigate the sequence space intelligently is an active area of research.<sup>[139]</sup> Different approaches for achieving this can generally be organised into three categories: mutagenesis by rational design, semi-rational mutagenesis and random



**Figure 9.** Navigating a hypothetical 3D fitness landscape.

mutagenesis. Choosing which approach to adopt for a given DE project depends on how much structural and mechanistic information is available about the parent enzyme, the size of the library generated, how mutants will be screened or selected and which property or properties of the enzyme the experimenter wishes to alter.<sup>[138]</sup> Finally, the *de novo* design of new biocatalysts *in silico* is also emerging as a promising future direction for the development of biocatalysts with enhanced properties and even novel functions.<sup>[16]</sup>

### 1.8.1 Rational design strategies

DE by rational design of an enzyme can be an effective approach in examples when detailed structural characterisation, such as an X-ray crystal structure of the parent enzyme, is available. In particular, rational DE strategies based on site-directed mutagenesis (SDM) have proven to be effective for altering an enzyme's substrate specificity, enantioselectivity, regioselectivity or stability and for introducing novel functionality to an enzyme.<sup>[52,140–144]</sup> For example, Nicotra and coworkers rationally mutated two phenylalanine residues to valine to alter the regioselectivity of acylation of monosaccharides by a lipase.<sup>[141]</sup> In a second example, Halpert *et al.* rationally mutated cytochrome P450 2B1 to mimic the active site of the structural homologue cytochrome P450 2C5 to alter the regioselectivity of the enzyme.<sup>[143]</sup>

Another type of rational DE is the incorporation of unnatural amino acids (UAAs) into an enzyme.<sup>[145–149]</sup> For example, Fasan and coworkers demonstrated that incorporation of the UAAs *para*-amino-phenylalanine, *para*-acetyl-phenylalanine, *O*-benzyl-tyrosine or 3-(2-naphthyl)alanine into the active site of a cytochrome P450 enzyme could alter the regioselectivity of oxidation of (*S*)-ibuprofen methyl ester and (+)-nootkatone.<sup>[150]</sup> In a second example, Chen *et al.* demonstrated the inversion of enantioselectivity of diketoreductase by site directed mutagenesis of an active site tryptophan residue to one of the UAAs 4-cyano-L-phenylalanine, 4-methoxy-L-phenylalanine, 4-phenyl-L-phenylalanine or *O*-*tert*-butyl-L-tyrosine.<sup>[151]</sup>

## 1.8.2 Semi-rational evolution strategies

Owing to the complexity and often unpredictability of enzyme catalysis, semi-rational approaches to DE tend to be most common when structural and/or mechanistic information on an enzyme is known.<sup>[138,152,153]</sup> Many approaches to semi-rational DE employ saturation mutagenesis techniques. Saturation mutagenesis is the random exchange of a defined amino acid for another, undefined amino acid. Maximum diversity is introduced when a degenerate codon for all 20 canonical amino acids is used (NNK or NNS), although reduced amino acid alphabets can also be employed if a particular functionality is required.<sup>[139,154]</sup> The decision of where to focus this randomisation normally depends on which property of the parent enzyme the experimenter would like to evolve. For example, thermostability and organic solvent tolerance may be enhanced by targeting positions identified by a consensus approach based on MSA or by identifying sites with a high B-factor,<sup>[10,155]</sup> whilst residues lining the binding pocket of an enzyme would be targeted for manipulation of the stereo- and regioselectivity, substrate scope or even to introduce non-native enzyme function.<sup>[153]</sup>

This has recently been demonstrated by Arnold and coworkers, who rationally engineered a cytochrome *c* heme from *Rhodothermus marinus* (Rma cyt. *c*) to catalyse the formation of C-Si bonds with high chemo- and enantioselectivity.<sup>[156]</sup> Interestingly, this protein is not known to have any catalytic function in living systems, yet when applied for the catalysis of carbene (which is generated from diazo compound **75**) insertion into a Si-H bond of a silane substrate (**74**), 97% enantiomeric excess (*ee*) with a turnover frequency (TOF) of 7 was achieved before evolution (**Scheme 18**).<sup>[156,157]</sup> Upon interrogation of the crystal structure of Rma cyt. *C*, the authors identified a Met residue which is coordinatively labile with respect to the iron centre of the heme group.<sup>[158]</sup> The authors hypothesised that mutation at this position could facilitate formation of an ‘active site’ cavity to enhance catalysis. By



detected by UV spectroscopy, enabling a medium-to-high throughput screening strategy. Using this approach, the authors were able to identify mutants with activity towards 11 substrates with which the WT enzyme was inactive or had very low activity.<sup>[159]</sup> In a subsequent publication, Reetz *et al.* further developed this methodology to the use of iterative rounds of CASTing, whereby the genes of the hits obtained in the first round are used as the parent gene for successive rounds of evolution.<sup>[160,161]</sup> By combining beneficial mutations identified in the first round of screening to produce a library of double mutants, the authors were able to identify mutants providing up to 145-fold increase in rate with non-natural substrates of PAL relative to the WT enzyme.<sup>[160]</sup> In other examples, Reetz and coworkers applied iterative rounds of CASTing to evolve P450-BM3 for the regio- and enantioselective hydroxylation of cyclohexene-1-carboxylic acid methyl ester,<sup>[162]</sup> whilst Jiang and coworkers used a similar strategy to evolve a variant of monoamine oxidase from *Aspergillus niger* (MAO-N-D5) for use in the kinetic resolution of *rac*-mexiletine.<sup>[163]</sup>

Another application of saturation mutagenesis is the B-Factor Iterative Test (B-FIT) method, which has been applied for engineering biocatalysts with enhanced thermostability and organic solvent tolerance.<sup>[10]</sup> In protein X-ray crystallography, the B-factor of a given scattering centre describes the displacement of atoms from their mean position in the crystal structure, which is used as a crude measure of the degree of flexibility of a given region of the structure.<sup>[164]</sup> Using the B-FIT method, saturation mutagenesis is carried out on residues with a high B-factor in the crystal structure, based on the assumption that a higher degree of rigidity in an enzyme will render it more stable to higher temperatures and/or organic solvents. As with CASTing, the best results have often been obtained when this is carried out in an iterative fashion, meaning that the gene corresponding to the best hit from the first library of mutants based on one site is used as the parent gene for a second round of saturation mutagenesis at one of the other selected sites.<sup>[10]</sup> The effectiveness of this methodology was demonstrated by the evolution of a lipase from *Bacillus subtilis* (BSL), from which a mutant was identified with a  $T_{50}^{60}$  (the temperature at which a heat treatment for 60 minutes results in 50% activity loss) of 93 °C, compared to 48 °C for the WT BSL.<sup>[10]</sup> In a subsequent study, Reetz and coworkers showed that these mutants were also more tolerant to up to 50% v/v acetonitrile, *N,N*-dimethylformamide (DMF) and dimethylsulfoxide (DMSO).<sup>[165]</sup>



### 1.8.3 Random mutagenesis strategies

In random mutagenesis strategies, mutations are introduced in an unbiased manner throughout the entire gene encoding the protein of interest. Random mutagenesis is often used if there is no structural data available on the protein of interest. A popular method for achieving this *in vitro* is the error-prone polymerase chain reaction (epPCR), in which a low fidelity DNA polymerase (ie one that lacks proof reading ability) is used for DNA replication during the PCR reaction. The *Taq* polymerase from *Thermus aquaticus* is the most commonly used DNA polymerase, which typically introduces a mutation every  $0.1-2 \times 10^4$  bases.<sup>[166]</sup> This value can be adjusted by varying the  $MgCl_2$  or  $MnCl_2$  concentrations, using unbalanced amounts of dNTPs or by increasing the amount of *Taq* polymerase used.<sup>[167,168]</sup> Alternatively, chemical mutagens such as nitrous acid, formic acid, hydrazine or ethyl methane sulfonate can be used to introduce random mutations in base pairings.<sup>[169]</sup> A useful comparison of different random gene mutagenesis techniques has been reported by Schwaneberg and coworkers.<sup>[170]</sup> Following amplification, the PCR products are cloned into an appropriate vector, which is then used for recombinant protein expression. In one of the first examples of using epPCR for a DE project, Arnold and Chen were able to evolve the protease subtilisin E to hydrolyse a peptide substrate in 60% DMF 256 times more efficiently than the WT enzyme over three iterative rounds of epPCR and screening.<sup>[171]</sup> In a more recent example, Janssen and coworkers used epPCR to evolve an epoxide hydrolase towards higher enantioselectivity in the hydrolysis of a range of epoxide substrates, with the best mutant showing a 4-fold improvement over the WT enzyme.<sup>[172]</sup>

Another popular method for randomly generating diversity in a DNA library is the use of a mutator strain such as *E. coli* XL1-red, which contains deactivated proof reading and repair enzymes and has a mutation frequency of ~1-2 base changes per gene.<sup>[138]</sup> This process involves transforming a plasmid containing the gene of interest into the mutator strain, propagating the culture and isolating the library DNA, which can then be transformed into the desired strain for screening. This approach has been successfully demonstrated by Turner and coworkers, who used *E. coli* XL1-red to generate diversity in a library of genes encoding the monoamine oxidase from *Aspergillus niger* (MAO-N).<sup>[173]</sup> Following transformation and screening of ~20 000 clones from the library against a set of secondary amines using an agar-plate based colorimetric screen, a mutant with up to 5.5-fold higher  $k_{cat}$  relative to WT was identified. Moreover, the mutant showed enhanced activity towards chiral secondary amines, enabling a deracemisation methodology to be established.<sup>[173]</sup>

Although random mutagenesis methods remain an effective and popular approach to introducing genetic diversity into a DNA library, both methods can be prone to deletions, insertions and amino acid bias due to the redundancy of the genetic code.<sup>[139]</sup> An alternative approach to random mutagenesis is DNA shuffling, or recombination. In this case, multiple homologous parent genes, which could be generated from epPCR or from homologous proteins, are randomly fragmented using DNase I and the fragments recombined using overlap-extension PCR.<sup>[174,175]</sup> Camararo and coworkers have successfully demonstrated this approach by shuffling the DNA from two fungal laccases for the directed evolution of chimeras with enhanced thermostability and were active across a wider pH range.<sup>[176]</sup>

### 1.8.4 Computational protein design

Over the past decade, bioinformatics has played an increasingly important role in guiding DE projects. A large number of online tools, databases and programmes are now available to aid library design, analyse mutational effects in proteins and even design new protein sequences from scratch.<sup>[16,177]</sup> For example, the HotSpot Wizard server (<http://loschmidt.chemi.muni.cz/hotspotwizard/>) identifies functional and stability hotspots in a given protein based on the integration of structural, functional and evolutionary information obtained from further databases and bioinformatics tools.<sup>[178]</sup> Having created this mutability map, the tool also aids library design and suggests the most suitable degenerate codon based on MSAs with other similar proteins. Additionally, there have now been significant advances towards designing *de novo* proteins using programmes such as the Rosetta software suite.<sup>[16,179]</sup> However, due to the very low sequence homology of the methyltransferases NovO and CouO to other known sequences, computational protein design was not explored further in this thesis.

## 1.9 Screening and selection

As discussed, there are many different ways in which genetic diversity can be introduced into a gene encoding a protein of interest. In many cases, the bottleneck of a DE project is identifying a suitable assay that is capable of identifying mutants harbouring beneficial mutations. For DE projects that generate relatively small libraries, it may be possible to screen mutants using high-pressure liquid chromatography (HPLC) or liquid-chromatography-mass spectrometry (LC-MS), which has successfully been demonstrated in a number of cases.<sup>[52,180,181]</sup> However, despite significant advances in the design of smaller, focused libraries, screening large libraries remains an essential component of the majority of

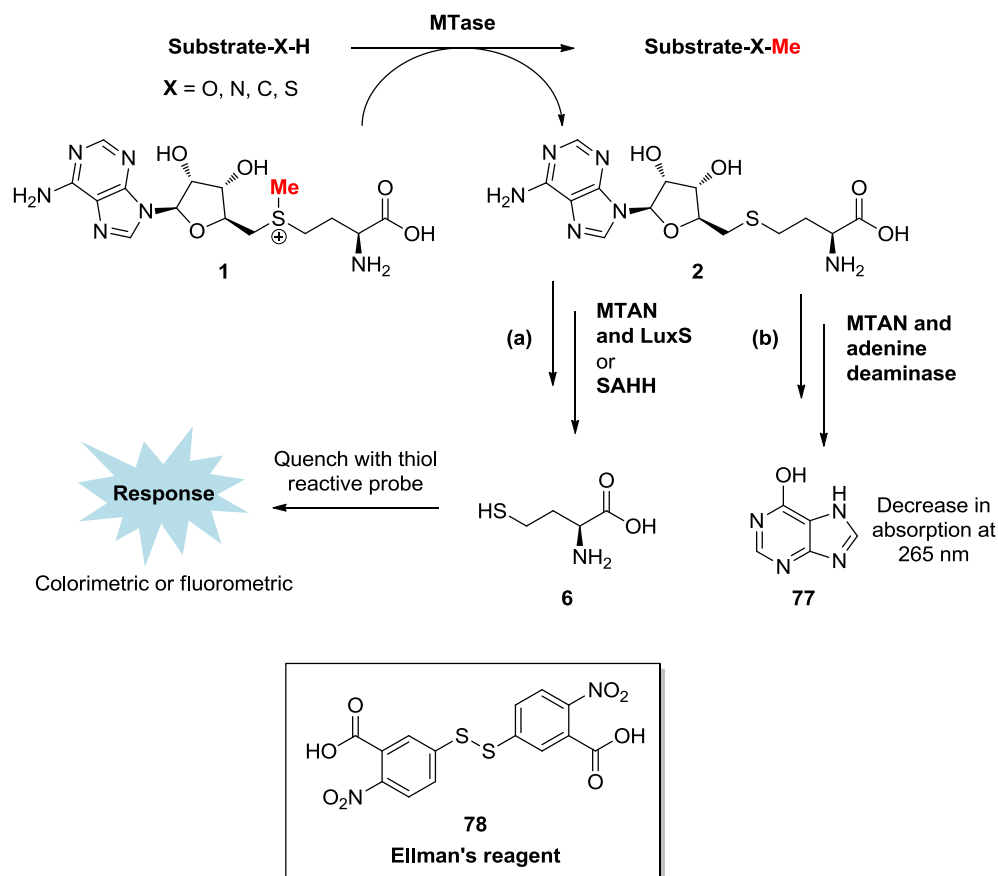
DE projects. Although sometimes used interchangeably, the terms ‘screening’ and ‘selection’ refer to two distinct approaches to identifying variants harbouring favourable mutations. Whereas ‘screening’ refers to the measurement of a given enzyme property such as activity, enantio- or regioselectivity, ‘selection’ refers to the use of a system in which the host organism has a growth or survival advantage because it harbours an enzyme variant with the desired property or properties.<sup>[182]</sup>

### 1.9.1 Screening

Beyond the use of chromatography based screening methods such as HPLC, UPLC and LC-MS, most screening strategies used in DE can broadly be categorised into plate based colorimetric or fluorometric systems (medium throughput), colony based screens (high throughput) and ultrahigh-throughput methods such as fluorescence-activated cell sorting (FACS) and the use of microcapillary arrays.<sup>[138,183]</sup> Despite the potential of the ultrahigh-throughput techniques, plate (96-well or 384-well) and colony based systems remain the most popular as they do not require any specialised equipment and are generally straightforward to set up.

A myriad of plate-based screens have been reported to cater for a range of biocatalysed transformations.<sup>[184]</sup> In the simplest case, a substrate is selected which releases a coloured or fluorescent product upon reaction or induces an easily detectable change in solution, such as precipitation. For example, in their model study to demonstrate CASTing as a DE strategy, Reetz and coworkers used a library of *p*-nitrophenyl esters as model substrates. Upon ester hydrolysis, *p*-nitrophenol was produced, which has a strong absorbance at 405 nm and was easily detected by UV spectroscopy.<sup>[159]</sup> In a second example, Turner and coworkers used an amine donor for a transaminase reaction which, following the transamination reaction, spontaneously polymerised to provide a dark precipitate, which could be used in either 96-well plate or agar plate based format.<sup>[185]</sup> In some cases however, including *C*-methylation by *C*-MTs, it is not possible to monitor enzyme activity directly in this way. In this case, chemo- and enzyme-coupled assays have been developed to probe enzyme activity indirectly. For SAM dependent MTs, plate-based enzyme coupled assays have focused on the indirect detection of SAH either by a) decomposition to homocysteine (**6**) using methylthioadenosine nucleosidase (MTAN) and the hydrolase LuxS or SAH hydrolase (SAHH) (**Scheme 19a**) or b) to hypoxanthine (**77**) using MTAN and adenine deaminase (**Scheme 19b**).<sup>[186–188]</sup> If the first option is selected, the homocysteine generated can be trapped with a thiol reactive probe such as Ellman’s reagent<sup>[189]</sup> (5,5'-dithio-bis-(2-

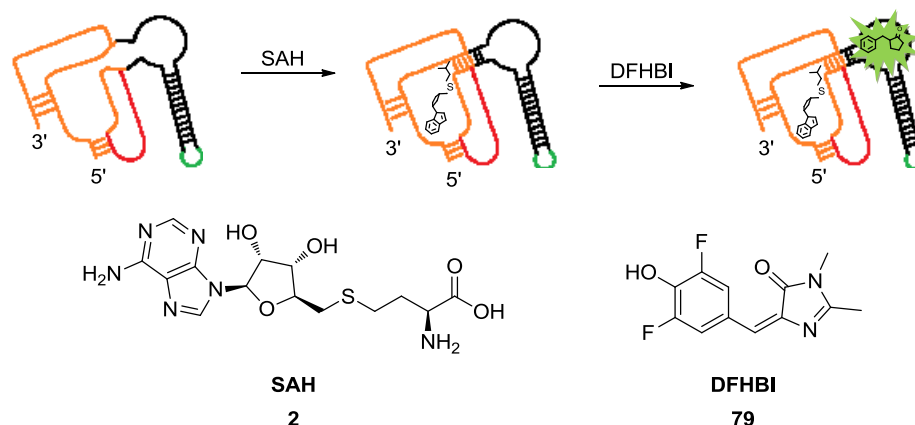
nitrobenzoic acid, **78**), triggering a colorimetric or fluorometric response. This assay (MTAN/LuxS coupled) has been used by Zhou and coworkers to measure the kinetics of salicylic acid carboxyl MT (SAMT),<sup>[186]</sup> whilst the SAHH coupled version was employed by Trievel and coworkers to measure activity of histone MTs.<sup>[190]</sup> If the MTAN/adenine deaminase option is chosen, MT activity is detected by measuring the decrease in absorption at 265 nm, caused by conversion of adenine into hypoxanthine by adenine deaminase. This approach was demonstrated by Zhou and coworkers to measure the kinetic parameters of PRMT1.<sup>[187]</sup>



**Scheme 19.** Plate based, enzyme coupled assays for the detection of SAM dependent MT activity.

Very recently, a novel approach to SAH detection for use in SAM dependent MT assays using a SAH selective riboswitch has been reported.<sup>[191]</sup> In this work, a SAH riboswitch was conjugated to a circular permutant of the Spinach2 aptamer (cpSpinach2). When SAH is present it binds to the RNA riboswitch, inducing a conformational change which allows the fluorophore 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI, **79**) to bind, upon which fluorescence is turned on (**Scheme 20**). Importantly, this sensor was reported to have

>1000-fold selectivity for SAH over SAM, rendering it a very sensitive tool for detecting MT activity, as demonstrated by the detection of SAH in live *E. coli* cells.<sup>[191]</sup>



**Scheme 20.** SAH detection using an RNA fluorescent bioswitch based on a SAH riboswitch.<sup>[191]</sup> SAH riboswitch shown in orange; cpSpinach2 shown in black. SAH: *S*-adenosylhomocysteine; DFHBI: 3, 5-difluoro-4-hydroxybenzylidene imidazolinone.

So far, most of the screening methods discussed are primarily used in 96-well or 384-well plate format. In order to screen larger libraries more efficiently however, higher throughput assays on an agar plate have also been developed. Not only does this allow more mutants to be screened at once, but it also negates the need for picking colonies to a plate, culture propagation and protein expression stages. One disadvantage compared to plate based screening, however, is that results are typically qualitative rather than quantitative. Nonetheless, agar plate based assays have enabled successful DE projects in a number of cases. Microbial cultures harbouring active lipases form a clearing zone on an agar plate prepared with tributyrin is a classic example of this, which has now been developed into a number of more specialised systems.<sup>[192]</sup> In a more recent example, a solid-phase assay for screening a library of (*R*)- or (*S*)-selective transaminases using glyoxylate as the amino acceptor was developed.<sup>[193]</sup> A further example of the effective use of a colony based screen is in the development of a toolbox of MAO-N variants for the oxidation of a library of chiral amines.<sup>[194]</sup> In both examples, the assay was based on detection of hydrogen peroxide by a peroxidase in the presence of a substrate that provides a highly coloured product upon oxidation.<sup>[173,193,195]</sup> To date however, there are no published reports of a colony based screen for SAM dependent MT activity.

As tailoring a biocatalyst to suit a particular transformation by DE is becoming increasingly popular in both academic and industrial settings, there is need to increase the efficiency of

the screening process, especially in projects involving very large libraries. As such, ultrahigh-throughput screening strategies are emerging. Two notable examples are fluorescence activated cells sorting (FACS) and the use of microcapillary arrays.<sup>[196]</sup> For FACS to be viable for use in DE, there must be a fluorescent output arising from enzyme activity. Examples of the successful application of FACS in DE projects have addressed this requirement by designing the substrate such that a fluorescent product is released from the enzymatic transformation<sup>[197,198]</sup> and *via* substrate tethering to a fluorescent dye.<sup>[199,200]</sup> In contrast to the plate and colony based screens outlined above, FACS can sort mutants at rates of up to  $10^7$  hour<sup>-1</sup> and also does not require lengthy colony picking and culturing stages.<sup>[196]</sup>

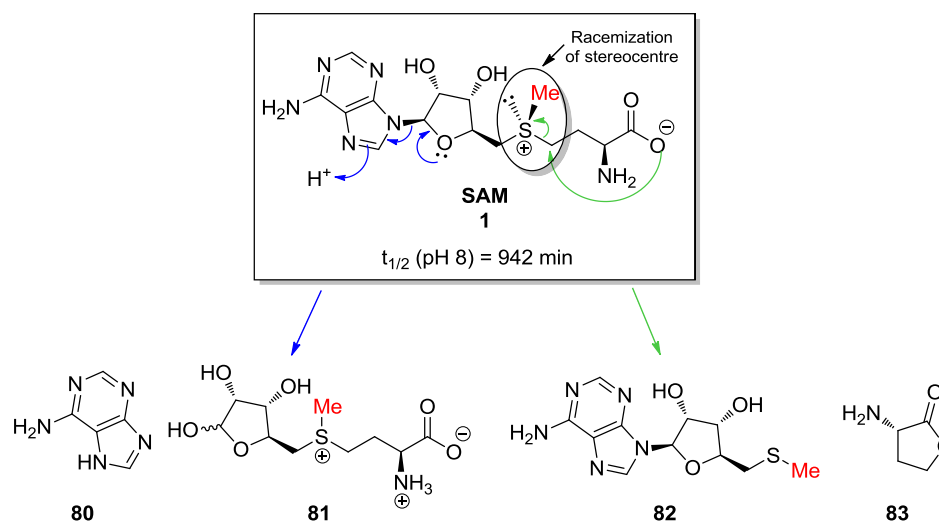
## 1.9.2 Selection

As an alternative to screening a variant library for mutants with beneficial properties, selection pressures may be placed on the system to favour the survival of cells or colonies harbouring advantageous mutations.<sup>[182,183]</sup> An early example of this is the DE of an esterase which liberates glycerol as a carbon source. When the colony based assay was carried out on minimal media agar plates, colonies bearing the desired esterase mutants grew faster and were easier to detect.<sup>[201]</sup> A second early example is the DE of an aspartate aminotransferase towards broader substrate acceptance, which used an auxotrophic strain of *E. coli* deficient in the branched-chain amino acid transferase gene.<sup>[202]</sup> More recently, a number of new strategies have been introduced including affinity selection<sup>[203]</sup> and phage assisted continuous evolution,<sup>[204]</sup> although their application in the DE of biocatalysts is currently limited.

## 1.10 Strategies for the synthesis of SAM and non-natural SAM analogues

A major limitation to the use of MTs *in vitro* for large scale synthesis is the availability, cost and instability of the cofactor SAM. SAM is produced commercially in a yeast fermentation process, with yields of up to 8.55 g/L reported for *Pichia pastoris*.<sup>[205]</sup> This makes it very expensive to buy SAM in large quantities; at the time of writing, 500 mg of SAM as the *p*-toluenesulfonate salt costs £467.00 from Sigma Aldrich.<sup>[206]</sup> Furthermore, SAM is hygroscopic and unstable, with a half life of 942 minutes at pH 8 at 37 °C.<sup>[207]</sup> Owing to the high density of functional groups within the SAM molecule, there are numerous possible mechanisms by which degradation could occur. Out of these, the two most commonly observed pathways are extrusion of the adenine ring (**Scheme 21**, blue arrows) to release adenine (**80**) and decomposition product **81**, and intramolecular attack of the carboxylate

group to produce 5'-methylthioadenosine (MTA, **82**) and homoserine lactone **83** (**Scheme 21**, green arrows).<sup>[208]</sup> Additionally, the sulfonium centre of the biologically active *S*, *S*-diastereomer of SAM epimerises at a rate of  $\sim 10^{-16} \text{ s}^{-1}$  at physiological conditions to produce the biologically inactive *R*, *S*-diastereomer.<sup>[208,209]</sup> Not only do some of these degradation products have inhibitory effects on many MTs,<sup>[50]</sup> but from a synthetic chemistry perspective it is also difficult to determine the exact stoichiometry of SAM relative to substrate in a given reaction.



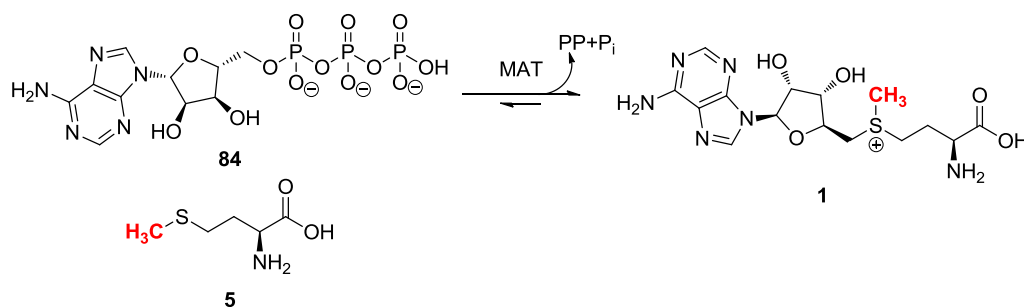
**Scheme 21.** Common degradation pathways of SAM.

Consequently, for uses in organic synthesis it is preferable to synthesise SAM and SAM analogues *in situ*, as it avoids issues associated with cost, handling and instability. Two enzymatic approaches have been adopted to achieve this: methionine adenosyl transferases (MATs) and halogenases, both of which offer the possibility of being coupled directly to a MT in a one-pot, two enzyme tandem process. Moreover, both enzymes exclusively produce the *S*, *S*- diastereomer of SAM or SAM analogues, which avoids the need for separation of diastereomers.<sup>[209]</sup>

### 1.10.1 MATs

In Nature, SAM is synthesised by methionine adenosyl transferases (MATs) from adenosine triphosphate (ATP, **84**) and methionine (Met, **5**) (**Scheme 22**).<sup>[210]</sup> As SAM is a vital cofactor for virtually all living organisms, MATs are a very large family of enzymes, with over 12 890 different sequences in UniProtKB at the time of writing. Nonetheless, the sequence

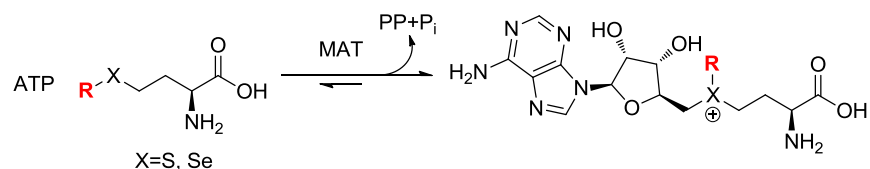
of MATs are generally well conserved across different species, especially with regards to the ATP binding site.<sup>[211]</sup>




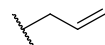

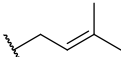
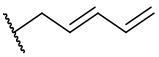


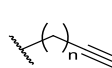
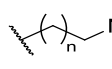
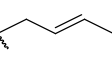
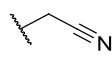
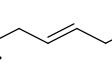
**Scheme 22.** Synthesis of SAM from ATP and Met by MAT.

MATs have attracted considerable attention as biocatalysts for the *in situ* synthesis of SAM and SAM analogues. For example, Singh and coworkers have comprehensively evaluated a set of MATs from bacterial, archaeal and mammalian sources for their ability to catalyse this reaction with Met and a variety of non-natural Met analogues to provide a range of non-natural *S*-alkylated SAM analogues.<sup>[212]</sup> Within this study, human MAT II catalytic  $\alpha$ -subunit (hMAT2A) and the archaeal thermophilic MAT from *Methanocaldococcus jannaschii* (mMAT) were found to accept the widest range of Met analogues, which included those bearing primary and secondary alkyl groups, alkynes, nitriles, allylic systems, azides and an amino substituent.<sup>[212,213]</sup> Additionally, the synthesis of Se analogues was also investigated *via* the use of *Se*-alkylated selenomethionine derivatives.<sup>[212]</sup> *Se*-SAM analogues are of interest as they have been shown to be both more stable<sup>[214,215]</sup> and better alkyl donors<sup>[216,217]</sup> than SAM. In a similar manner, Phillips and coworkers have reported the substrate promiscuity of a MAT from *Sulfolobus solfataricus* (sMAT) with regards to the same set of Met analogues.<sup>[218]</sup> The range of non-natural SAM analogues that have been synthesised using WT MAT enzymes is summarised in **Table 1**.





R	MAT	% Conversion					
		X=S			X=Se		
		n=1	n=2	n=3	n=1	n=2	n=3
 n = 1, 2, 3	eMAT <sup>140</sup>	8	2	nd	5	-	-
	hMAT1A <sup>140</sup>	75	5	nd	57	6	6
	hMAT2A <sup>140</sup>	90	54	12	53	24	26
	hMAT2 <sup>140</sup>	88	55	13	82	74	49
	mMAT <sup>140</sup>	94	21	12	73	27	32
	sMAT <sup>146</sup>	83	45	28	76	65	69
	eMAT <sup>140</sup>	-	-	-	-	5	-
	hMAT1A <sup>140</sup>	-	2	-	-	5	-
	hMAT2A <sup>140</sup>	-	2	-	-	11	-
	hMAT2 <sup>140</sup>	-	2	-	-	10	-
	mMAT <sup>140</sup>	-	8	-	-	10	-
	sMAT <sup>146</sup>	-	55	-	-	40	-
 n = 1, 2	mMAT <sup>140</sup>	1	1	-	-	-	nd
	sMAT <sup>146</sup>	60	36	-	-	-	-
	eMAT <sup>140</sup>	-	-	-	-	-	-
	eMAT <sup>140</sup>	-	2	-	-	2	-
	hMAT1A <sup>140</sup>	-	5	-	-	10	-
	hMAT2A <sup>140</sup>	-	50	-	-	42	-
	hMAT2 <sup>140</sup>	-	52	-	-	59	-
	mMAT <sup>140</sup>	-	16	-	-	15	-
	sMAT <sup>146</sup>	-	28	-	-	30	-
	eMAT <sup>140</sup>	-	-	-	-	3	-
	hMAT1A <sup>140</sup>	-	-	-	-	5	-
	hMAT2A <sup>140</sup>	-	2	-	-	11	-
	hMAT2 <sup>140</sup>	-	2	-	-	12	-
	mMAT <sup>140</sup>	-	6	-	-	14	-
	sMAT <sup>146</sup>	-	15	-	-	45	-
	eMAT <sup>140</sup>	-	3	-	-	5	-
	hMAT1A <sup>140</sup>	-	-	-	-	3	-
	hMAT2A <sup>140</sup>	-	4	-	-	9	-
	hMAT2 <sup>140</sup>	-	6	-	-	15	-
	mMAT <sup>140</sup>	-	13	-	-	16	-
	sMAT <sup>146</sup>	-	12	-	-	46	-
	eMAT <sup>140</sup>	-	4	-	-	-	-
	hMAT1A <sup>140</sup>	-	2	-	-	-	-
	hMAT2A <sup>140</sup>	-	4	-	-	-	nd
	hMAT2 <sup>140</sup>	-	44	-	-	-	-
mMAT <sup>140</sup>	-	8	-	-	-	-	

		n=1	n=2	n=1	n=2
 $n = 1, 2$	eMAT <sup>140</sup>	2	nd	-	2
	hMAT1A <sup>140</sup>	2		7	6
	hMAT2A <sup>140</sup>	6	12	44	22
	hMAT2 <sup>140</sup>	20	40	58	55
	mMAT <sup>140</sup>	6	16	14	17
	sMAT <sup>146</sup>	35	31	28	10
 $n = 1, 2$	hMAT1A <sup>140</sup>	4	-		
	hMAT2A <sup>140</sup>	18	19		nd
	hMAT2 <sup>140</sup>	49	46		
	mMAT <sup>140</sup>	29	31		
	sMAT <sup>146</sup>	51	62		
	eMAT <sup>140</sup>		-		5
	hMAT1A <sup>140</sup>		2		4
	hMAT2A <sup>140</sup>		5		9
	hMAT2 <sup>140</sup>		12		15
	mMAT <sup>140</sup>		32		18
	sMAT <sup>146</sup>		30		26
	eMAT <sup>140</sup>		2		2
	hMAT1A <sup>140</sup>		2		4
	hMAT2A <sup>140</sup>		13		23
	hMAT2 <sup>140</sup>		20		33
	mMAT <sup>140</sup>		6		0
	sMAT <sup>146</sup>		69		12
	eMAT <sup>140</sup>				8
	hMAT1A <sup>140</sup>				7
	hMAT2A <sup>140</sup>		nd		16
	mMAT <sup>140</sup>				25
					30

**Table 1.** Summary of Met analogues accepted by WT MAT enzymes.

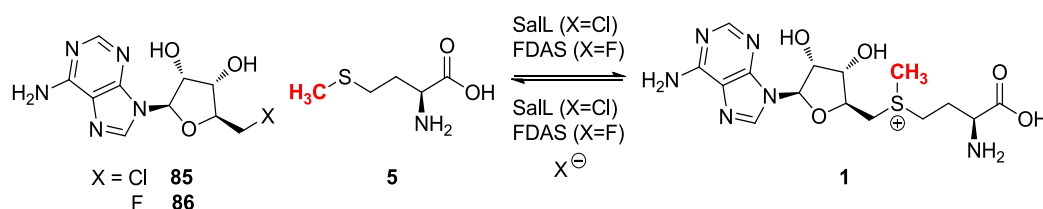
As seen in **Table 1**, there is a general trend of a lower conversion to the SAM analogue as the size of the R group increases. This arises from the fact that in Nature, MATs have evolved to accept Met as the natural substrate, which has a very small, unhindered *S*-alkyl group (R=Me). From both a chemical biology and synthetic chemistry perspective, however, a biocatalyst capable of synthesising SAM analogues with larger R groups is desirable. To this end, Wessjohann and coworkers have rationally designed variants of MAT from *Bacillus subtilis*. In this study, the authors targeted two isoleucine (I) residues (I105 and I317) which appeared to make close contact with the Me group of Met in the X-ray crystal structure of the enzyme.<sup>[144]</sup> Relative to the WT enzyme, the variants I317V and I317A showed higher activity with ethyl, propyl, butyl and prop-1-enyl *S*-alkylated Met analogues. On a preparative scale, the I317V mutant provided 84% and 89% conversion with Met and ethionine, respectively, whilst the I317A mutant provided 43% and 28% conversion with *S*-propyl- and *S*-butyl- Met analogues, respectively. Interestingly, in the case of the double mutant I105A/I317A, the substrate specificity was inverted (propyl > butyl > ethyl >

methyl).<sup>[144]</sup> In a different study, Luo and coworkers engineered variants of hMAT2A to accept two Met analogues bearing a terminal alkyne (*S*-(*E*)-hex-2-en-5-yn-1-yl and *S*-(*E*)-pent-2-en-4-yn-1-yl).<sup>[219]</sup> In this example, analysis of the X-ray crystal structure (PDB code: 2P02) identified V121, I117 and I322 to be restricting the pocket site within the active site. Mutation of these three residues to alanine (A), glycine (G), leucine (L) and valine (V) individually and in combination identified I117A to have the highest activity towards both Met analogues.<sup>[219]</sup>

Further to being effective for the synthesis of SAM and SAM analogues, MATs have also been coupled to MTs in a one-pot, biocatalytic tandem alkylation process. For example, Luo and coworkers demonstrated that the hMAT2A I117A mutant described above could be employed *in vivo* for the synthesis of *S*-alkynated SAM analogues, which were accepted by engineered protein MTs for the alkylation of chromatin.<sup>[219]</sup> Using this strategy, it was possible to site-selectively label chromatin, thereby enabling genome-wide profiling of methylation in living cells.<sup>[219]</sup> In a second example, Micklefield and coworkers demonstrated the site-specific bioalkylation of the antibiotic rapamycin using hMAT2A I322V for cofactor synthesis and the *O*-MT RapM for site-specific transfer of a methyl, ethyl or allyl group according to the Met analogue used.<sup>[220]</sup>

### 1.10.2 Halogenases

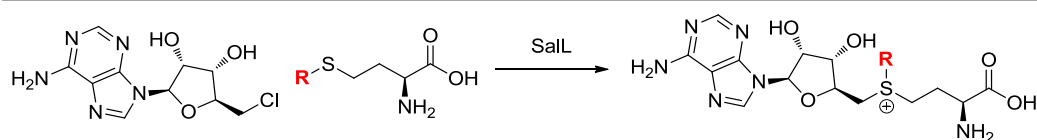
An alternative to the use of MATs for SAM synthesis is the use of the halogenases SalL (*Salinospora tropica*)<sup>[221]</sup> or FDAS (*Streptomyces cattleya*).<sup>[222]</sup> In their native environments, SalL and FDAS catalyse the decomposition of SAM by the nucleophilic attack of Cl<sup>-</sup> (SalL) or F<sup>-</sup> (FDAS) at the 5'-position of the adenosine moiety of SAM.<sup>[221,222]</sup> However, under conditions with low halide ion concentrations, it is possible to push the equilibrium in the synthesis direction to synthesise SAM from Met and 5'-chloro-5'-deoxy adenosine (CIDA, **85**) or 5'-fluoro-5'-deoxy adenosine (FDA, **86**) using SalL or FDAS, respectively (**Scheme 23**). Owing to the ease of preparation of CIDA from adenosine by treatment with thionyl chloride,<sup>[223]</sup> SalL has emerged as a promising biocatalyst for the *in situ* synthesis of SAM from inexpensive starting materials.



**Scheme 23.** Enzymatic production of SAM by SalL. When a high concentration of chloride or fluoride ions are present, the reverse reaction is favoured.

To demonstrate the potential of SalL for *in situ* cofactor synthesis, Burkhardt and coworkers successfully coupled this process with the two SAM dependent MTs HhaI (a DNA mC5 MT) and MftA (teicoplanin O-MT).<sup>[224]</sup>

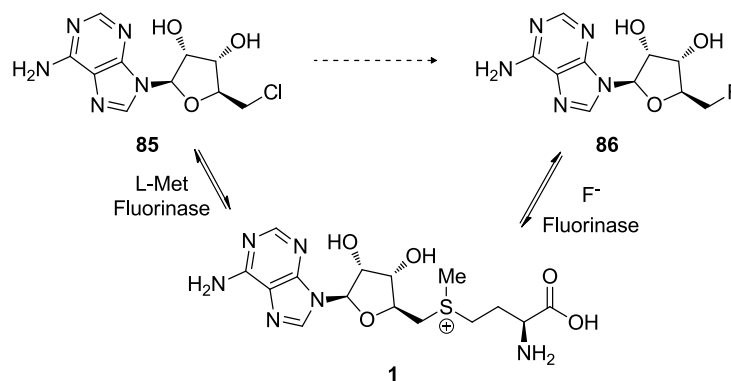
To further investigate the scope of SalL and FDAS for the synthesis of SAM and SAM analogues, Clausen and coworkers measured their activity with a library of six alkyl donors.<sup>[181]</sup> Whilst FDAS only showed very low activity with the *S*-ethyl analogue, SalL was active with the *S*-ethyl, -propyl, -butanyl, -allyl and -benzyl analogues, albeit with a drastic decrease in activity with increasing size of the R group (**Table 2**).<sup>[181]</sup> In order to increase activity of SalL with these bulkier substrates, the authors designed a series of single point mutants designed to destabilise two water molecules which were bound in the active site of the X-ray crystal structure of SalL (PDB code: 2Q6I). The authors hypothesised that dislodging these water molecules from the active site by exchanging nearby amino acids for hydrophobic residues would increase the size of the substrate binding pocket to accommodate larger alkyl groups.<sup>[181]</sup> As shown in **Table 2** however, nearly all the mutants of SalL displayed lower activity with all the tested substrates relative to the WT enzyme. The one notable exception to this was SalL W190A, which showed a three-fold increase in activity with the *S*-butyl substrate relative to the WT enzyme.<sup>[181]</sup> To demonstrate the application of SalL and SalL W190A for *in situ* cofactor synthesis, the authors coupled this process to a PRMT to modify the RGG peptide by alkylation in a one-pot, two enzyme coupled process.<sup>[181]</sup>



R	Enzyme	Activity (nM/min)		
	WT SaIL	4930		
	SaIL T128S	356		
	SaIL T128G	123		
	SaIL W190F	843		
	SaIL W190A	31		
	SaIL Y239F	915		
			n=1	n=2
WT SaIL		453	1	0.3
SaIL T128S		6	-	-
SaIL T128G		9	-	-
SaIL W190F		3	0.15	0.12
SaIL W190H		-	0.1	0.04
SaIL W190A		2	0.9	0.9
SaIL Y239F		68	0.4	0.4
	WT SaIL	1.02		
	SaIL W190F	0.3		
	SaIL W190A	0.5		
	SaIL Y239F	0.4		
	WT SaIL	0.1		
	SaIL W190A	0.075		

**Table 2.** Summary of substrate scope of SaIL and SaIL single point mutants investigated by Clausen and coworkers.<sup>[181]</sup>

Recently, an interesting new application of SAM-dependent halogenases has been reported by Ang and coworkers. The fluorinase FIA1 (*Streptomyces* sp MA37) has 87% homology to FDAS (*Streptomyces cattleya*) and also catalyses the fluorination of SAM to produce FDA, but with a markedly higher  $k_{cat}$  than its known homologues.<sup>[225]</sup> With this in mind, Ang and coworkers hypothesised that ClDA (**85**) could be converted into FDA (**86**) via SAM (**1**) using an evolved mutant of FIA1, Met and a fluoride source (**Scheme 24**).<sup>[180]</sup>

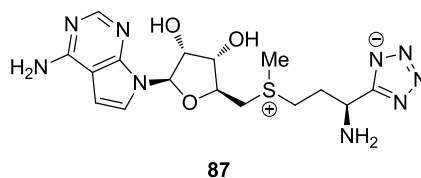


**Scheme 24.** Conversion of CIDA (**85**) into FDA (**86**) using the fluorinase FIA1 reported by Ang and coworkers.<sup>[180]</sup>

The authors used FIA1 as the parent enzyme in a semi-rational DE project, in which they targeted 23 residues within 5 Å from SAM in a homology model of FIA1 based on the homologue FIA (PDB code: 1RQP) towards saturation mutagenesis. The two most beneficial mutations were combined in a second round of evolution to provide the double mutant FIA1 F213Y/A279L, which provided a 2.7-fold increase in conversion to FDA from CIDA after 90 minutes, using NaF as the fluoride source.<sup>[180]</sup> Moreover, it was possible to extend this methodology to the synthesis of <sup>18</sup>F radiolabelled FDA, with overall radiochemical conversions of 34% for FIA1 F213Y/A279L.

### 1.10.3 Stable SAM isosteres

The use of MATs and halogenases for *in situ* cofactor synthesis represents a promising step towards establishing a platform for biocatalytic alkylation reactions. However, the problem of cofactor instability still remains. To address this, Thorson and coworkers have recently reported the synthesis and use of an analogue of SAM which is resistant to the two most common degradation pathways discussed above (**Scheme 21**).<sup>[207]</sup> In this work, a SAM analogue (**87**) containing the adenine isostere 7-deazaadenine and the carboxylate isostere



**Figure 10.** Structure of SAM isostere **87** designed to be resistant to the two most common degradation pathways.<sup>[207]</sup>

tetrazolate (**Figure 10**) was synthesised using hMAT2A for the key C-S bond forming step to provide the sulfonium centre. Moreover, **87** was demonstrated to be an effective methyl donor for the prototypic SAM dependent MT DnrK and showed no degradation after 2 days at pH 8.<sup>[207]</sup>

## 1.11 Aims and objective

The overall objective of this thesis is to develop the use of the C-MTs NovO and CouO as biocatalysts for the Friedel-Crafts alkylation and fluoroalkylation. Currently, their utility in organic synthesis is limited due to their narrow substrate scope, cost and instability of SAM and non-natural SAM analogues. It is hypothesised that the substrate scope may be improved by carrying out DE on NovO and CouO. However, there is currently very limited structural and mechanistic understanding of these enzymes which limits the feasibility of semi-rational DE, which is most effective for altering the substrate scope of an enzyme. Therefore, five primary aims will be addressed:

### 1. Structural characterisation of NovO and CouO

At present, the structures of NovO and CouO are not known and as such there is no structural information available to support a semi-rational DE project. Therefore the first aim is to obtain an X-ray crystal structure of NovO and CouO, which would ideally have substrate bound in the active site so that active site residues can clearly be identified.

### 2. Understanding of mechanistic features associated with methyl transfer in NovO and CouO

Having obtained structural information about NovO and CouO, the mechanism of methyl transfer within their active sites will be interrogated to provide information on which residues are required for catalytic activity. This would also aid design of the variant library in a subsequent DE project.

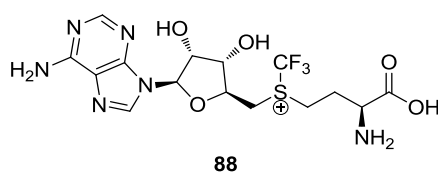
### 3. Evolve NovO and CouO towards a broader substrate scope using directed evolution

With structural and mechanistic information about NovO and CouO in hand, a DE project will be undertaken with the primary goal of expanding the substrate scope, which is currently limited to a small range of aminocoumarins and dihydroxynaphthalenes. Ideally, this will be

extended to substrates which are more commonly used in organic synthesis such as substituted phenols. However, in order to efficiently identify variants with a broader substrate scope, first it will be necessary to establish a suitable screening assay, ideally with medium to high throughput. To date, the only literature precedent for the DE of a MT has involved a very small variant library, which was therefore amenable to screening by HPLC.<sup>[52]</sup> However, if a semi-rational approach such as CASTing is adopted for DE of NovO and CouO, then a higher throughput assay such as a colourimetric or fluorometric system is preferable.

#### 4. Synthesis of a fluorinated analogue of SAM as a cofactor for biocatalytic trifluoromethylation

As discussed, there is currently no literature precedent for a biocatalytic trifluoromethylation reaction using MTs. In order to achieve this, a suitable cofactor is required. To this end, the trifluoromethylated analogue of SAM, **88** (Figure 11), was identified as a key synthetic target to investigate the feasibility of a biocatalytic trifluoromethylation reaction using MTs.



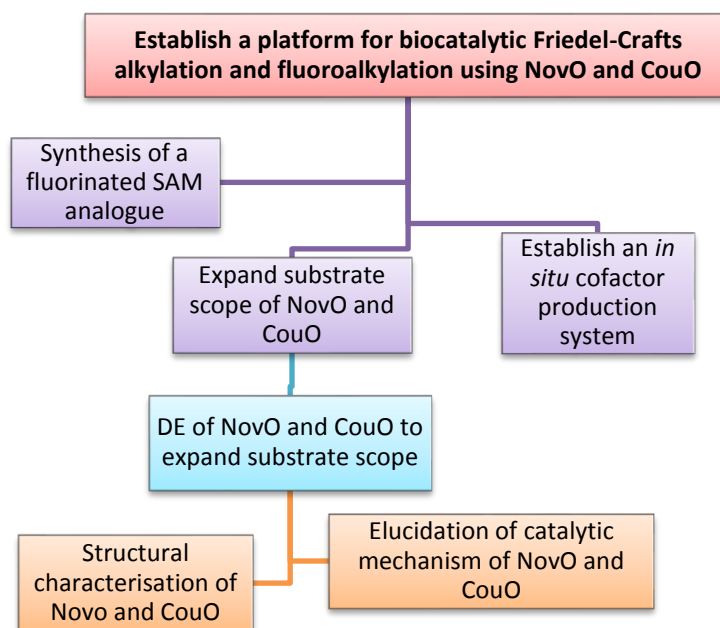
**Figure 11.** Trifluoromethylated analogue of SAM (**88**) proposed as a synthetic target for use as a cofactor for a biocatalytic trifluoromethylation reaction using NovO or CouO.

#### 5. Development of a robust and scalable system for *in situ* cofactor synthesis

Finally, the issue of the cost and availability of the cofactor, SAM will be addressed. As discussed, both MATs and the halogenase SalL have been demonstrated to be effective for the *in situ* production of SAM and SAM analogues from more widely available starting materials. Out of these, SalL was identified to be particularly appealing for use in this project, as CIDA can be easily prepared in one pot from adenosine, which is inexpensive and readily available. Furthermore, in comparison to ATP, which is the adenosyl donor in the MAT system, the use of CIDA is far more atom efficient. Although the substrate scope of SalL is currently more limited than that of the MAT enzymes, there is scope for the DE of SalL to address this limitation. Therefore, a one-pot, two-enzyme cascade reaction employing SalL and NovO or CouO will be investigated.



The aims outlined above are summarised in **Figure 12**.



**Figure 12.** Summary of project aims.

## **Chapter 2.**

# **Expression, purification and structural and functional characterisation of NovO and CouO**

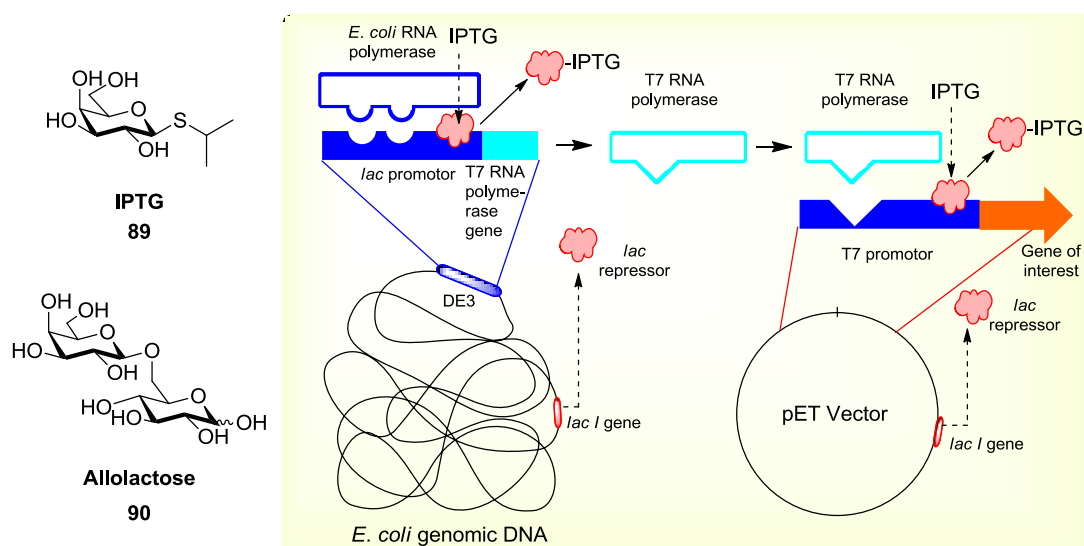
## 2 Expression, purification and structural and functional characterisation of NovO and CouO

The C-MTs NovO and CouO are of interest as biocatalysts for Friedel-Crafts-type alkylation and fluoroalkylation reactions, yet are currently underdeveloped for routine use in synthetic organic chemistry. In particular, a major limitation is their narrow substrate scope.<sup>[107,226]</sup> To rationalise this and provide a detailed structural understanding of these MTs, which would support a directed evolution project, X-ray crystal structures of NovO and CouO were required. The first stage to achieve this was to determine suitable conditions for the expression of NovO and CouO, which would enable purification of sufficient quantities of the protein to support crystallisation studies.

### 2.1 Expression and purification of NovO and CouO

The genes for NovO and CouO from *Streptomyces spheroides* and *Streptomyces rishiriensis*, respectively, were codon optimised for *E. coli*, synthesized and sub-cloned into a pET26b(+) vector with a C-terminal 6xHis-tag at GenScript™. The constructs were transformed into *E. coli* BL21 (DE3) and the overexpression of the genes was attempted under the conditions reported by Stecher *et al.*<sup>[107]</sup> The cultures were grown in Luria Bertani (LB) broth at 37 °C to an optical density (OD) of ~0.6, before inducing with isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubating at 25 °C overnight. IPTG (**89**) is an analogue of allolactose (**90**) and is commonly used as an inducer of the *lac* operon. Its mechanism of action is summarised in **Figure 13**.

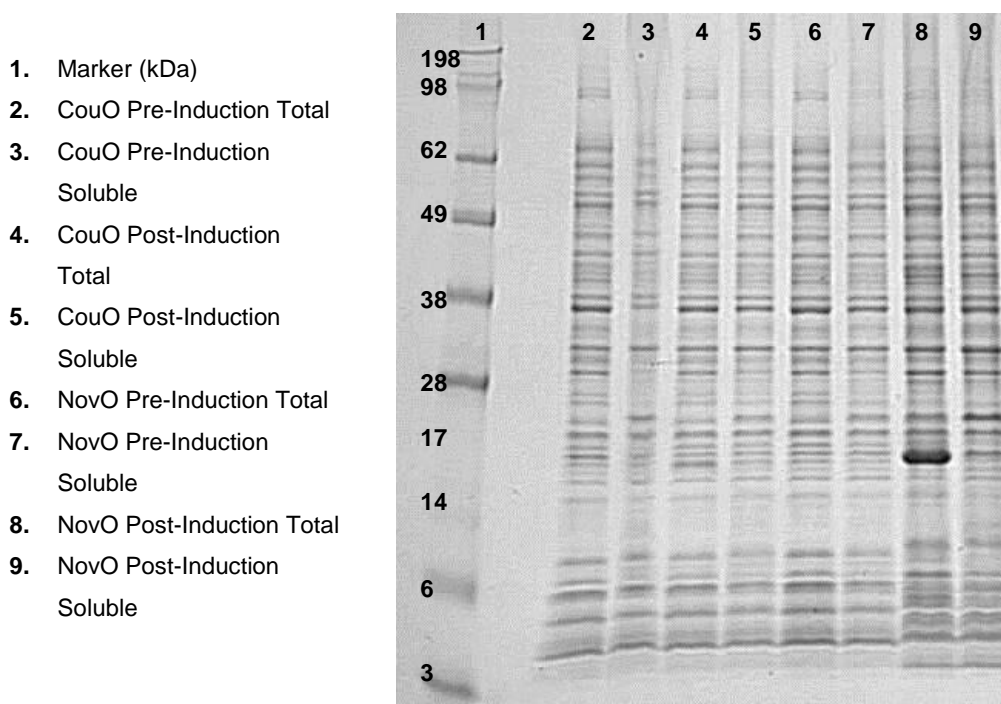
Briefly, the *lac I* gene in the *E. coli* genomic DNA encodes the *lac* repressor protein, which binds to the *lac* promoter in the *E. coli* DNA. Similarly, the *lac I* gene in the pET vector encodes the *lac* repressor protein which binds the T7 promoter of the pET vector. These repressor proteins prevent RNA polymerase binding and subsequent transcription of the following gene. IPTG binds to the *lac* repressor, causing a conformational change in the protein and dissociation from the *lac* and T7 promoters. RNA polymerase then binds to the *lac* promoter and the T7 RNA polymerase gene is transcribed and translated to produce T7 RNA polymerase. This is now able to bind to the T7 promoter so that the gene of interest is transcribed and translated to generate the protein of interest (**Figure 13**). Unlike the natural substrate for the repressor protein (allolactose, **90**), IPTG is not hydrolysed *in vivo*, causing IPTG levels to remain constant, leading to increased levels of overexpression.<sup>[227]</sup>



**Figure 13.** Mechanism of gene expression with IPTG using the *lac* operon and T7 expression system.

Analysis of the expression levels of the genes for CouO and NovO under the LB and IPTG induction conditions by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) showed very good overexpression of both enzymes (**Figure 14**, lanes 4 and 8). However, CouO was only present in the total cell lysate sample and NovO only showed very low levels of soluble expression (**Figure 14**, lanes 4-5 and 8-9, respectively). As high levels of soluble protein expression are desirable for purification and subsequent crystallisation, alternative expression conditions were investigated.

The overexpression of NovO and CouO was next attempted in Overnight Express (OE) media at 37 °C, until an OD of ~2 was reached and the temperature was lowered to 18 °C overnight. In contrast to LB, OE is an autoinduction media, which does not require the addition of an inducer such as IPTG. Instead, glucose present in the broth is metabolised preferentially during growth to a high cell density (OD ~2). As glucose levels are depleted, lactose is taken up by the cells and converted to the inducer allolactose by  $\beta$ -galactosidase, triggering protein expression in a manner analogous to that described in **Figure 13**.<sup>[228]</sup> However, analysis by SDS PAGE revealed this system also produced very low levels of expression of NovO, whilst CouO was not visually detectable (**Figure 14**, lanes 7-9 and 3-6, respectively).



**Figure 14.** Analysis of heterologous expression of NovO and CouO. Enzymes expressed in *E. coli* BL21 (DE3) in OE Media at 37 °C to OD 2, then 18 °C overnight and analysed by SDS PAGE. Theoretical mass of 6xHis-CouO: 26.5 kDa; 6xHis-NovO: 26.2 kDa.

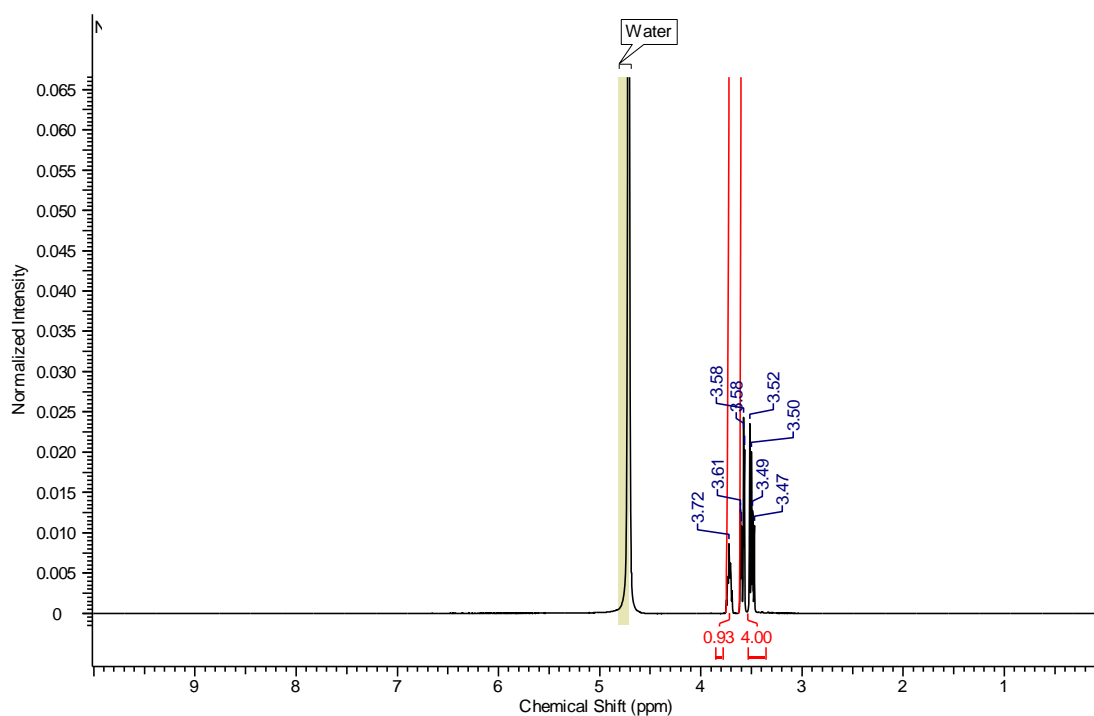
As both methods failed to produce reasonable levels of soluble NovO and CouO, a screen of protein expression conditions was carried out. In order to identify expression conditions with increased levels of soluble expression, a range of different media and two different induction temperatures were investigated (**Table 3**). In some cases, reducing the post-induction temperature has been shown to give higher levels of soluble expression.<sup>[229]</sup> This is thought to be due to a slower rate of production of inclusion bodies, so that when the protein of interest is produced it is less likely to aggregate to another protein and therefore more likely to be soluble. A repeat of the existing conditions was also carried out as a control (**Table 3**, entries 1 and 6).

Entry	Enzyme	Media	OD at induction or temp. drop	Expression temp. (°C)
1	CouO	LB	1.8	25
2	CouO	TB	1.1	18
3	CouO	MTB	1.0	18
4	CouO	OE	3.6	18
5	CouO	Magic	1.8	18
6	NovO	LB	1.9	25
7	NovO	TB	0.7	18
8	NovO	MTB	0.7	18
9	NovO	OE	3.4	18
10	NovO	Magic	1.5	18
11	CouO	LB	0.8	30
12	CouO	TB	1.0	14
13	CouO	MTB	1.1	14
14	CouO	OE	3.6	14
15	CouO	Magic	1.9	14
16	NovO	LB	1.8	30
17	NovO	TB	0.6	14
18	NovO	MTB	0.4	14
19	NovO	OE	1.3	14
20	NovO	Magic	1.6	14

**Table 3.** Conditions tested in overexpression optimization. LB: Luria-Bertani, TB: Terrific Broth, MTB: Modified Terrific Broth, OE: Overnight Express. Glycerol (to a final concentration of 1% v/v) was added to LB, TB and MTB media and ‘Component B’ (to a final concentration of 5% v/v) was added to Magic Media. IPTG (to a final concentration of 100  $\mu$ M) was added to experiments in LB, TB and MTB after the specified OD was reached. Conditions providing highest levels of soluble NovO or CouO are highlighted in grey.

The different media chosen varied in the peptone, yeast extract, sodium chloride and glycerol levels. For example, TB has 20% more peptone and 5 times more yeast extract than LB. Additionally, Overnight Express and Magic Media are autoinduction systems and are generally grown to a higher OD (~2) before the temperature is lowered. Magic Media was supplemented with ‘Component B’, which is provided with the media concentrate by the commercial suppliers. In order to determine the composition of ‘Component B’, a sample

was analysed by NMR spectroscopy. The spectrum was consistent with that of an aqueous glycerol solution, which is a commonly used additive for protein expression (**Figure 15**).<sup>[230]</sup> Glycerol was also added as a supplement to LB, TB and MTB prior to inoculation.

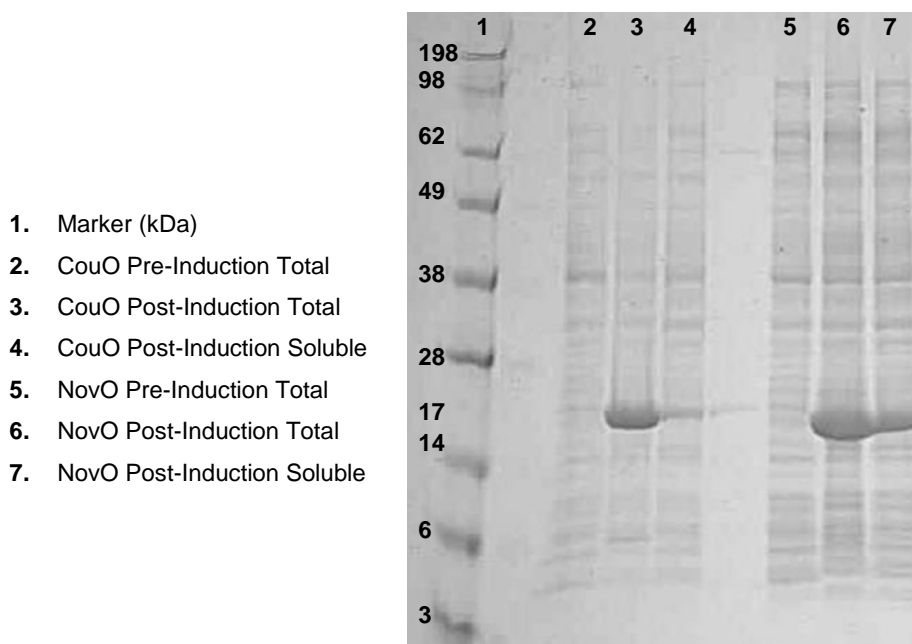


**Figure 15.**  $^1\text{H}$  NMR spectrum ( $\text{D}_2\text{O}$ , 400 MHz) of 'Component B' Magic Media supplement.

Samples of the resulting total and soluble fractions from the cell lysates were analysed by SDS PAGE. In contrast to the hypothesis discussed above, higher levels of both total and soluble protein expression were consistently observed at 18 °C than at 14 °C. Growth of cultures harboring the CouO gene grew faster than those with the NovO gene, however levels of soluble protein expression were higher for NovO than for CouO. Overall, the highest levels of soluble NovO and CouO expression were obtained using Magic Media with an overnight incubation temperature of 18 °C. Therefore, unless otherwise stated, these conditions were used throughout the project for the overexpression of both NovO and CouO and any mutants thereof.

The overexpression of NovO and CouO in Magic Media was carried out on a 5 L scale to provide soluble protein for crystallography studies. Analysis of the samples from the overexpression experiments (**Figure 16**) showed very high levels of CouO and NovO in the total protein samples (**Figure 16**, lanes 3 and 6) and high levels of NovO in the soluble fraction (**Figure 16**, lane 7). Levels of soluble CouO were lower, although higher than

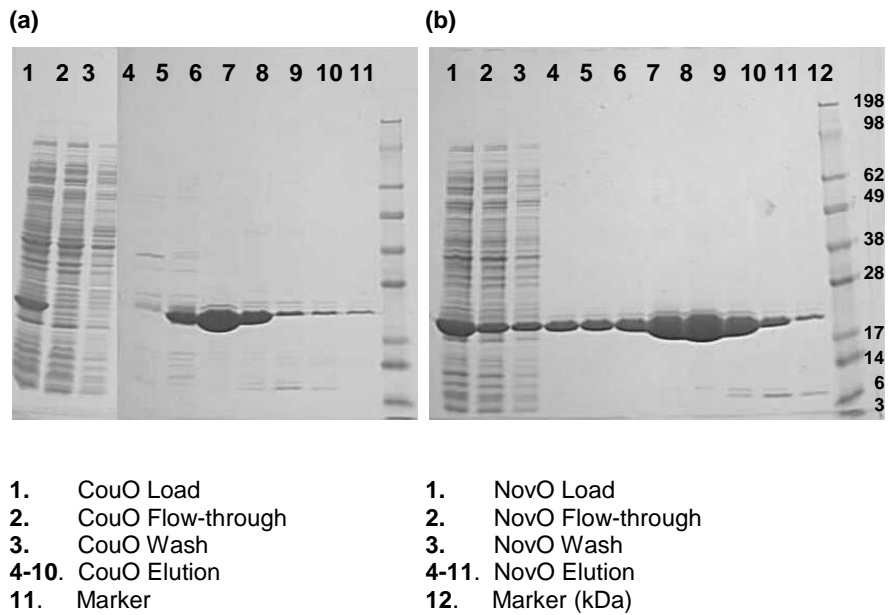
observed before optimisation. As the overexpression was carried out on a 5 L scale, this provided sufficient soluble enzyme in both samples for to take forwards for purification to support crystallographic studies.



**Figure 16.** Analysis of heterologous expression of NovO and CouO under optimized expression conditions. Enzymes expressed in *E.coli* BL21 (DE3) in Magic Media at 30 °C to an OD of 1.6 (CouO) and 1.7 (NovO), then 18 °C overnight and analysed by SDS PAGE. Theoretical mass of 6xHis-CouO: 26.5 kDa; 6xHis-NovO: 26.2 kDa.

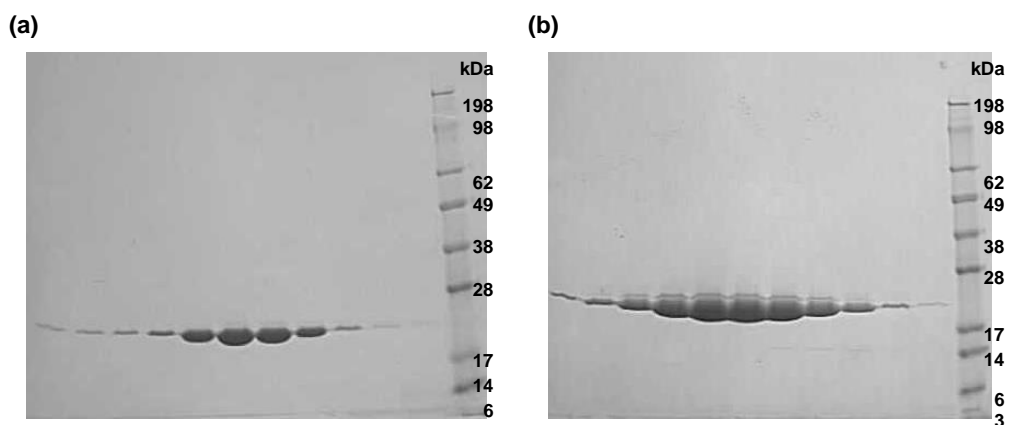
Typically, >99% pure enzyme is necessary for protein crystallography. To achieve this, the lysed cell pellets were first purified by affinity chromatography based on the C-terminal 6xHis-tag, followed by size exclusion chromatography (SEC). First reported by Hochuli *et al.* in 1988, affinity chromatography of His-tagged proteins is now a widely used method for the purification of recombinant proteins.<sup>[231]</sup> The procedure is based on the strong binding affinity of histidine residues for nickel. A sequence of six histidine residues (the 6xHis-tag) is inserted at the end of the polypeptide sequence (C- or N-terminus) and the cell lysate is loaded onto a cartridge of Ni beads. The 6xHis-tagged protein binds to the Ni-nitrilotriacetic acid (NTA) beads, allowing the contaminating, untagged proteins to be washed through the cartridge. The protein of interest is then displaced from the Ni beads by eluting with an imidazole solution. This procedure was carried to purify NovO and CouO and the eluted fractions were analysed by SDS PAGE (**Figure 17**).





**Figure 17.** Analysis of purification of (a) CouO and (b) NovO by Ni-NTA affinity chromatography by SDS PAGE. Theoretical mass of 6xHis-CouO: 26.5 kDa; 6xHis-NovO: 26.2 kDa.

In both cases, this method was successful in removing a large proportion of host cell proteins from the target protein, as seen by comparing the load sample with the eluted fractions (Figure 17). However, some contaminating proteins were found to persist in both cases. Therefore, the samples were purified further by size exclusion chromatography (SEC). This purification method separates proteins based on size by passing the sample through a silica column. Larger molecules are poorly retained by the pores in the silica and elute first,

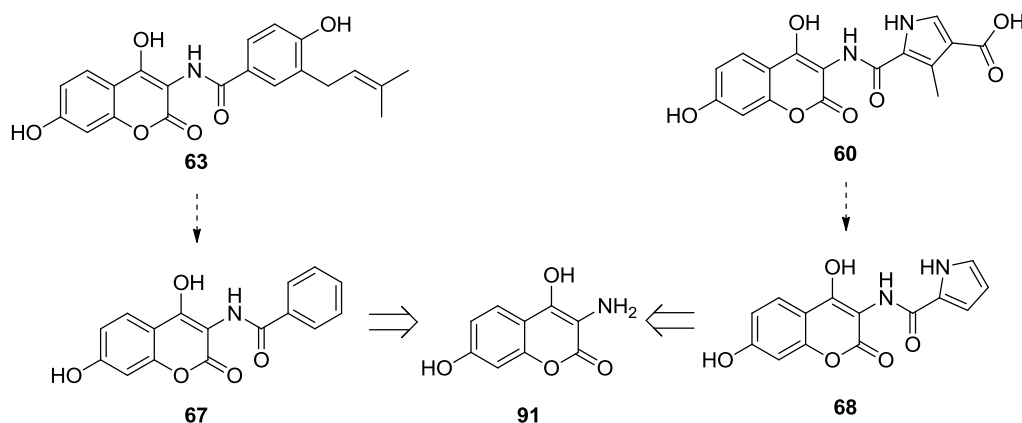


**Figure 18.** Analysis of purification of (a) CouO and (b) NovO by SEC by SDS PAGE. Theoretical mass of 6xHis-CouO: 26.5 kDa; 6xHis-NovO: 26.2 kDa.

whereas smaller molecules are retained better and elute later. This approach provided ~30 mg of CouO and ~100 mg of NovO of very high purity to support crystallographic studies (**Figure 18**).

## 2.2 Synthesis of aminocoumarin substrates and activity assays of NovO and CouO

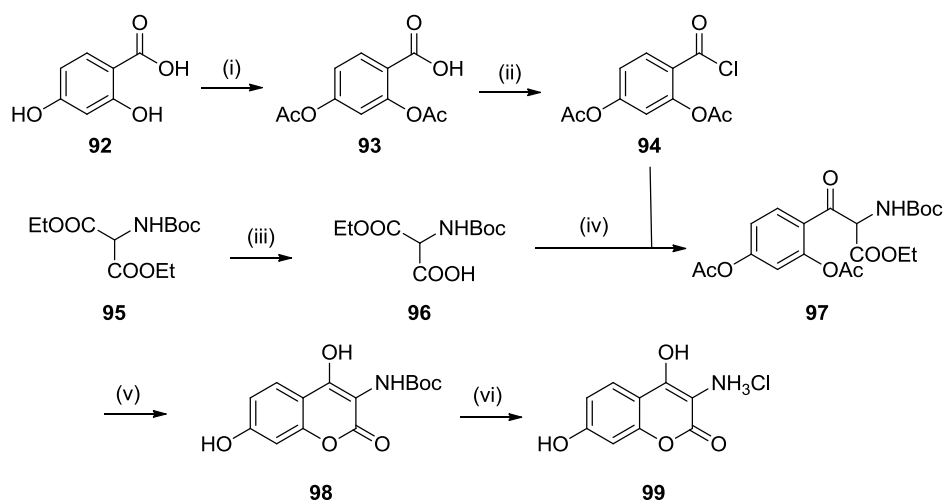
Prior to attempting the crystallisation of NovO and CouO, the methyltransferase activity of both enzymes was tested. To this end, it was necessary to synthesise appropriate substrates with which to test the activity. In Nature, the natural substrates for NovO and CouO are aminocoumarins **63** and **60**, respectively. However, Gruber and coworkers have previously reported that truncated analogues **67** and **68** are also accepted by both NovO and CouO and are therefore suitable for use to confirm enzyme activity. Both substrates **67** and **68** were synthesised from the common aminocoumarin intermediate **91** (**Figure 19**).



**Figure 19.** Natural substrates for NovO (**63**) and CouO (**60**) and the truncated analogues **67** and **68**, both of which can be synthesised from the common intermediate **91** and are accepted by NovO and CouO.<sup>[107]</sup>

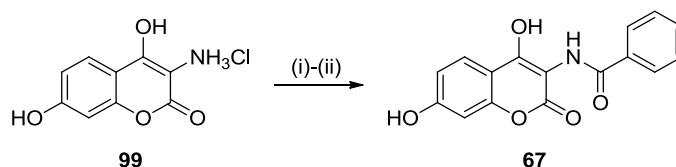
Aminocoumarin **91** was synthesised according to the route reported by Gruber *et al.* (**Scheme 25**).<sup>[107]</sup> First, the commercially available benzoic acid derivative **92** was acetylated by treatment with Ac<sub>2</sub>O in the presence of *N,N*-dimethylaminopyridine (DMAP) as a nucleophilic catalyst. The resulting carboxylic acid **93** was then activated for coupling with amino acid derivative **96** by conversion to the acid chloride **94**, which was taken through to the next stage of the synthesis without further purification. The other coupling partner, ethyl *N*-Boc malonate, **96**, was synthesised from the commercially available precursor **95** by partial hydrolysis under basic conditions. **96** was then decarboxylated under by treatment with triethylamine to generate the nucleophile which was subsequently reacted with the acid

chloride **94** to provide intermediate **97**. Following base mediated acetyl deprotection in the next step, the coumarin scaffold was constructed by intramolecular attack of the hydroxy group into the ethyl ester to provide **98**. This ring closure step is thermodynamically favoured due to the formation of a highly conjugated bicyclic ring system. The key intermediate **98** was isolated in 60% yield over three steps from **93**. The *N*-Boc-protected amine **98** was then deprotected under acidic conditions to provide the aminocoumarin scaffold and isolated as the hydrochloride salt **99** in 99% yield.



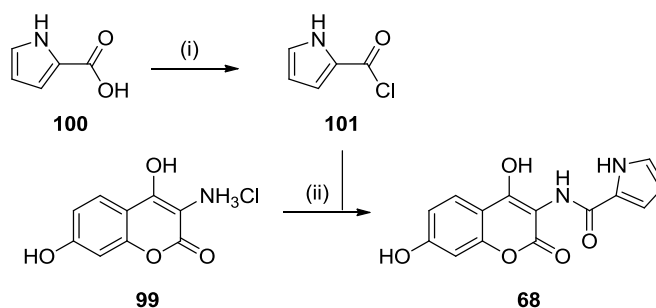
**Scheme 25.** Synthesis of aminocoumarin scaffold. *Reagents and conditions:* (i)  $\text{Ac}_2\text{O}$ , DMAP,  $\text{NEt}_3$ , THF, 22 °C, 1 h, 80%. (ii)  $(\text{COCl})_2$ , DMF, toluene, 70 °C, 2 h, *product not isolated*. (iii) KOH, methanol/ $\text{H}_2\text{O}$ , 22 °C, 12 h, 77%. (iv) **96**,  $\text{MgCl}_2$ ,  $\text{NEt}_3$ , THF, 22 °C, 2 h, then **97**, THF, 22 °C, 2 h, *product not isolated*. (v)  $\text{NaOH}_{(\text{aq})}$ , methanol, 20 °C, 2 h, 60% from **94**. (vi) HCl in cyclopentylmethylether (CPME), *tert*-butylmethylether (TBME)/methanol, 22 °C, 24 h, 99%.

Having synthesised the aminocoumarin scaffold **91** (as the hydrochloride salt **99**), it was then necessary to synthesise substrates **67** and **68** by derivatisation of the amine *via* amide bond formation. For the benzamide coumarin substrate **67**, conversion to the free base **91** and subsequent treatment with the corresponding acid chloride provided the *N*-acyl derivative **67**. As the reaction of **91** with benzoyl chloride provided a mixture of the desired product, along with di- and tri- *O*-acylated compounds, the reaction mixture was treated with aqueous NaOH in a second step to selectively hydrolyse the external ester groups over the amide and lactone groups (**Scheme 26**). This strategy successfully afforded the target compound **67** in 69% isolated yield over the two steps from **99**.



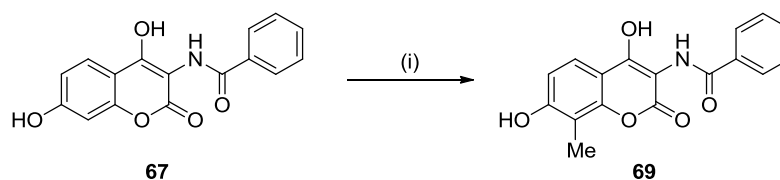
**Scheme 26.** Synthesis of aminocoumarin substrate **67**. *Reagents and conditions:* (i) BzCl, NEt<sub>3</sub>, EtOAc, 22 °C, 3 h. (ii) NaOH<sub>(aq)</sub>, methanol, 22 °C, 12 h, 69% over two steps from **99**.

As the corresponding acid chloride for the synthesis of substrate **68** is not commercially available, a variety of amide coupling reagents were screened for the formation of **68** from the corresponding carboxylic acid **100**. However, in all cases it was not possible to isolate the desired product and the starting material and a complex mixture of products was observed by LC-MS in all attempts. Therefore, carboxylic acid **100** was converted to the corresponding acid chloride **101** by treatment with oxalyl chloride, followed by coupling of the crude reaction product to aminocoumarin **99**. Subsequent base mediated ester hydrolysis provided substrate **68** in 47% isolated yield over two steps.



**Scheme 27.** Synthesis of aminocoumarin substrate **68**. *Reagents and conditions:* (i) (COCl)<sub>2</sub>, DMF, toluene, 70 °C, 2 h, *product not isolated*. (ii) NEt<sub>3</sub>, CH<sub>2</sub>Cl<sub>2</sub> then MeOH, NaOH, 22 °C, 1 h, 47% over 2 steps.



With the appropriate substrates in hand, it was now possible to carry out an activity assay of the purified NovO and CouO. At this stage, only activity with one of the substrates (**67**) was assessed to confirm activity of both enzymes. The assay was carried out using commercially available SAM as the methyl donor (**Scheme 28**) under conditions reported previously.<sup>[107]</sup> Under these unoptimised conditions, 18% and 20% conversion of aminocoumarin **67** to the methylated product **69** was observed with CouO and NovO, respectively. No methylation of **67** was observed in the absence of either NovO or CouO. This confirmed that the enzymes were active and could therefore be taken forward into crystallisation screens.



**Scheme 28.** Activity assay for NovO and CouO. *Reagents and conditions:* (i) *S*-adenosylmethionine *p*-toluenesulfonate salt (10 eq.), DTT (0.5 eq.), BSA (1 mg/mL), CouO (7  $\mu$ M) or NovO (10  $\mu$ M), 50 mM potassium phosphate buffer, pH 6.5 (NovO) or pH 7 (CouO), 36 h, 20% (NovO) and 18% (CouO) conversion by area/area% by HPLC.

## 2.3 Crystallisation of NovO

To enable enzyme evolution by rational or semi-rational design, the crystal structures of NovO and CouO were required. To this end, following the purification of active NovO, a crystallisation screen was set up using 8 commercially available sets of 92 conditions. Using a two-well crystallisation plate, NovO was incubated both with and without SAM in each condition for each of the 8 screens. These initial screens highlighted the PEGsII screen to be particularly successful, with two conditions (wells B10 and D7) producing protein crystals (**Table 4**). These conditions were both 2-(*N*-morpholino)ethanesulfonic acid (Mes) buffered at pH 6.5 and contained 30% poly(ethylene) glycol (PEG), which is a commonly used precipitant for protein crystallisation.<sup>[232]</sup>

Well Ref.	Image	Drop	Condition	Results
D7		NovO, SAM	0.1 M Mes pH 6.5, 30% PEG-24, 0.1 M NaOAc	N.D.
B10		NovO, SAM	0.1 M Mes pH 6.5, 30% PEG-44	Spots to 1.5 Å. Smear spots. Crystal disordered.

**Table 4.** Hits from initial crystallisation conditions screen (round 1). Screens carried with NovO from 3.8 mg/mL stock and a total concentration of 2 mM SAM. Mes: 2-(*N*-morpholino)ethanesulfonic acid buffer.

The crystals were harvested and subjected to preliminary analysis to determine the unit cell 0 and 90°. Whilst the crystal from D7 was highly disordered, a full X-ray diffraction data set was collected from the crystal from the B10 condition with a resolution of 1.5 Å. However,

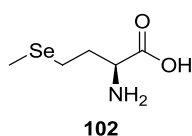
it was not possible to find a structural solution using molecular replacement due to a combination of issues including a low homology model (approximately 30% sequence identity match) and the poor quality of the collected data. Despite good resolution, the data was anisotropic and also the diffraction spots were smeared. This could be due to the freezing conditions or the crystal growth conditions or a combination of both. Therefore, further crystallisation studies were carried out with the objective of obtaining a better quality of crystal. A second round of conditions screening was set up in which the protein concentration was varied (3.8 mg/mL and 7.5 mg/mL stock). Additionally, experiments were seeded using a crystal obtained from the initial round of screening. Seeding provides a point of nucleation and can promote crystal formation.<sup>[233]</sup> Three crystals were frozen and tested for the unit cell at 0 and 90°, however all crystals were found to be of poorer quality, showing a disordered, weak or multiple lattice with spots to 2.2–3.7 Å.

A third round of screening was carried out, this time incubating NovO with SAH and with or without one of two aminocoumarin substrates **67** or **68**. This produced 11 hits, of which 9 crystals were large enough for freezing and testing the unit cell. However, in all cases there was poor resolution, a multiple lattice, anisotropic packing and/or a disordered unit cell. As the best results were obtained from the original screen under conditions B10 and D7, it was decided to return to these conditions and optimise the conditions for the crystallisation to obtain higher quality data. Four different preparations of the buffer in conditions B10 and D7 were screened, each using NovO (3.8 mg/mL) in two wells, one with and one without seed. The protein was incubated with SAH and without SAH and one of the substrates **67** or **68**. From this experiment, one crystal was frozen and X-ray diffraction data was collected with resolution to 1.5 Å. Despite good quality data, once more, it was not possible to solve the structure due low sequence similarity of the homology model and having 12 molecules in the asymmetric unit.

At this point, an alternative approach was explored. One method to aid structure elucidation is the incorporation of heavy atoms, such as selenium, into the protein.<sup>[234]</sup> This can be achieved by overexpressing the protein in a growth medium supplemented with selenomethionine (SelMet). This has been reported to help solve the phase problem in X-ray crystallography.<sup>[234]</sup>

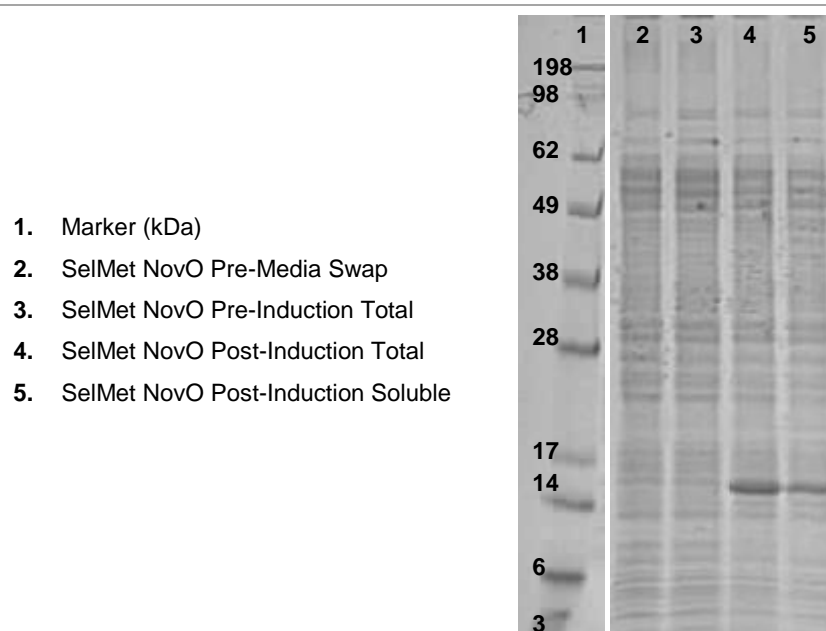
## 2.4 Selenomethionine labelling and structure elucidation of NovO

Selenium is most commonly incorporated into proteins by substitution of Met residues with selenomethionine (SelMet, **102**), a naturally occurring amino acid found in foods including Brazil nuts, cereal grain and soybeans (**Figure 20**).<sup>[235]</sup>



**Figure 20.** Structure of selenomethionine.

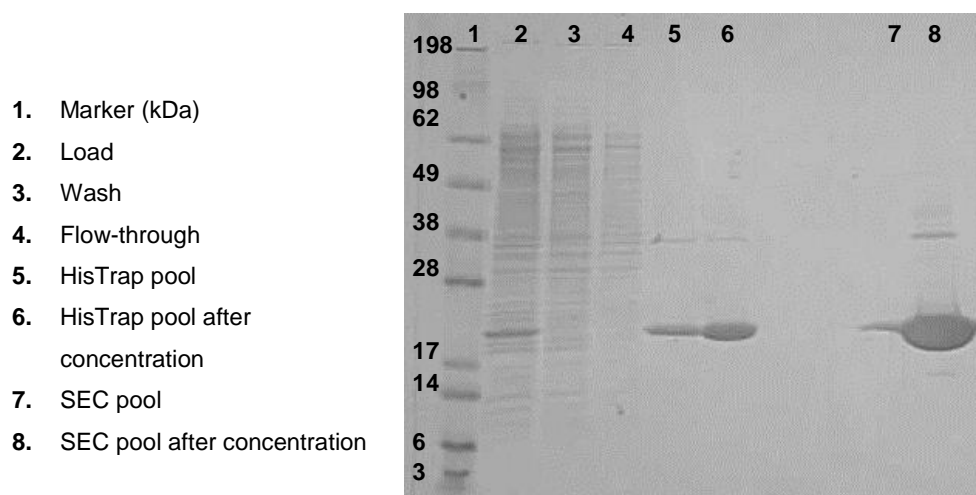
Sequence analysis revealed NovO to have 6 Met residues in each monomer chain, rendering it an appropriate candidate for SelMet labelling. Gruber and coworkers have previously reported NovO to exist as a homodimer based on gel-filtration chromatography, in which case there would be 12 potential sites for SelMet incorporation in the protein.<sup>[112]</sup> To express SelMet labelled NovO, a culture of *E. coli* BL21 harboring the gene for NovO was grown overnight in Modified Terrific Broth (MTB) media to an OD of ~9. The culture was then aseptically transferred into a minimal media supplemented with SelMet before inducing protein expression with 100  $\mu$ M IPTG. No expression of NovO prior to induction (known as ‘leaky expression’) was visible by SDS PAGE, as seen in samples taken both before and after the media swap (**Figure 21**, lanes 2 and 3).<sup>[236]</sup> After inducing with IPTG and incubating overnight at 18 °C, post induction total and soluble expression levels were analysed by SDS PAGE (**Figure 21**, lanes 4 and 5). Gratifyingly, overexpression of NovO was observed, and moreover, levels of soluble expression were sufficient to support crystallographic studies.



**Figure 21.** Analysis of heterologous expression of SelMet NovO in *E. coli* BL21 (DE3) by SDS PAGE. Culture grown in MTB at 30 °C overnight then aseptically transferred into in SelMet supplemented media and grown at 37 °C for 30 min before inducing protein expression with 100  $\mu$ M IPTG and growing overnight at 18 °C. Theoretical mass of WT 6xHis-NovO: 26.2 kDa.

The NovO obtained from the overexpression in SelMet supplemented media was purified by affinity chromatography, followed by size exclusion chromatography as previously described for WT NovO. As shown in **Figure 22**, the affinity chromatography using a Ni-NTA column was successful in removing the majority of contaminating proteins (**Figure 22**, lane 2 vs lanes 5 and 6). Additionally, by analysis of the wash and unbound fractions, it can be seen that all of the SelMet labelled NovO was loaded onto the Ni column bound and was not displaced in the wash. Fractions containing SelMet NovO were pooled and further purified by SEC, which was successful in removing the contaminant seen at ~50 kDa (**Figure 22**, lane 5 vs lane 9). Fractions containing SelMet NovO were pooled and concentrated to 12 mg/mL to provide protein sample for crystallographic studies (**Figure 22**, lane 10). At this concentration, low levels of contaminating proteins were visible by SDS PAGE, although these were at low enough levels in comparison to the amount of SelMet NovO to be used directly in crystallographic studies. Crystallisation work was carried out immediately to avoid freeze thawing of the protein prior to crystallisation, which has been reported to increase the rate of denaturation for some proteins.<sup>[237]</sup>

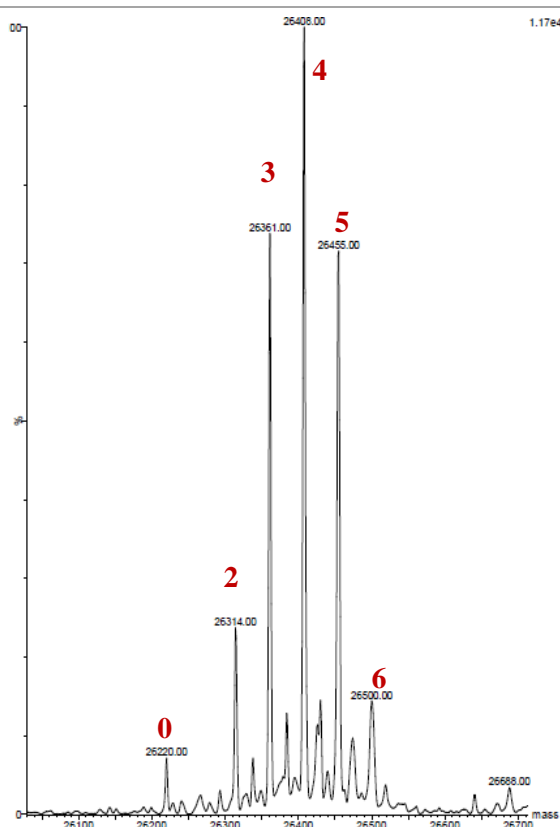




**Figure 22.** Analysis of purification of SelMet NovO by Ni-NTA affinity chromatography and SEC by SDS PAGE. Theoretical mass of WT 6xHis-NovO: 26.2 kDa.

To determine the level of SelMet incorporation, a sample of the purified SelMet NovO was analysed by protein LC-MS. There is a difference of 47 Da between the molecular weights of Met and SelMet, resulting in a maximum molecular weight of 26500 for 6xSelMet-6xHis-NovO. Although Se has a number of naturally occurring stable isotopes, these were not visible in the mass spectrum of the SelMet NovO sample.

The distribution of SelMet-incorporation could be clearly seen in the mass spectrum (**Figure 23**). The most abundant mass ion observed was 26.4 kDa, corresponding to 4 SelMet substitutions out of a maximum of 6, giving an approximate incorporation level of 67%. Additionally, the mass ions corresponding to 0, 2, 3, 5 and 6 SelMet substitutions of 26220, 26314, 26361, 26408, 26445 and 26500 Da, respectively, were observed (**Figure 23**). Peaks corresponding to the sodium ion adduct  $[M+Na^+]$  for each mass were also visible for each of these peaks in the spectrum. This result confirmed that expression of SelMet NovO had been successful, and that the level of SelMet incorporation was sufficient to improve the phasing of X-ray diffraction data to aid structure elucidation.



**Figure 23.** SelMet incorporation analysis by protein LC-MS. Numbers above peaks indicate the number of Met/SelMet substitutions out of a maximum of 6 in each monomer chain.

## 2.5 Crystallisation of SelMet NovO

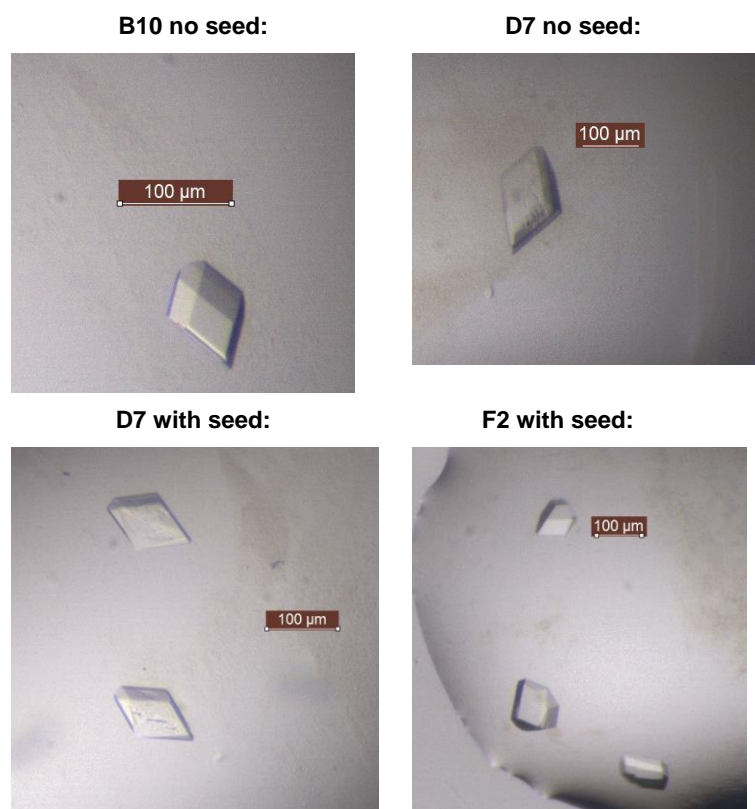
Conditions for the crystallisation of NovO to provide crystals with good morphology and provide high quality data have been previously identified. Nonetheless, due to the low homology of NovO to any other structurally characterised proteins, it was not possible to solve the structure previously. Based on previous results for the crystallisation of non-labelled NovO, two conditions from the PEGsII screen (conditions B10 and D7) were used initially for crystallisation. Additionally, from previous studies, it was found that seeding and co-crystallising with *S*-adenosylhomocysteine (SAH) was also found to produce higher quality crystals. With this in mind, four seed solutions of varying concentrations were prepared and used in the crystallisation studies. It was hypothesised that using a more dilute solution of seed stock would result in larger crystals by reducing the number of nucleation sites. Therefore, a crystallisation screen was set up with SelMet NovO and SAH, screening different formulations of conditions B10 and D7 (Tray 1) both with no seed (drop 1) and with varying concentrations of seed added (drop 2). Additionally, a full PEGsII screen was also carried out with SAH (Tray 2), without (drop 1) or with (drop 2) seed as it was also

possible that SelMet NovO could crystallise in a different set of conditions to the non-labelled protein. From these experiments, a large number of crystals were obtained, out of which those with the best morphology were frozen and submitted for data collection (**Table 5**).

Entry	Crystal	Condition	Size and Morphology	Diffraction results
1	Tray 1 D3 drop 1	0.1 M Mes pH 6.5, 30% PEG-24, 0.1 M NaOAc	Medium, average morphology	Good diffraction, resolution: 2.5 Å
2	Tray 1 C8 drop 1	0.1 M Mes pH 6.5, 30% PEG-44	Big, multiple	Low resolution; poor quality data
3	Tray 2 F2 drop 2	0.1 M Mes pH 6.5, 20% PEG-44, 0.6 M NaCl	Medium, good morphology	Good diffraction, resolution: 1.7 Å
4	Tray 2 D7 drop 2	0.1 M Mes pH 6.5, 30% PEG-44	Small, good morphology	Weak diffraction

**Table 5.** SelMet NovO crystals obtained from SelMet NovO crystallisation experiments.

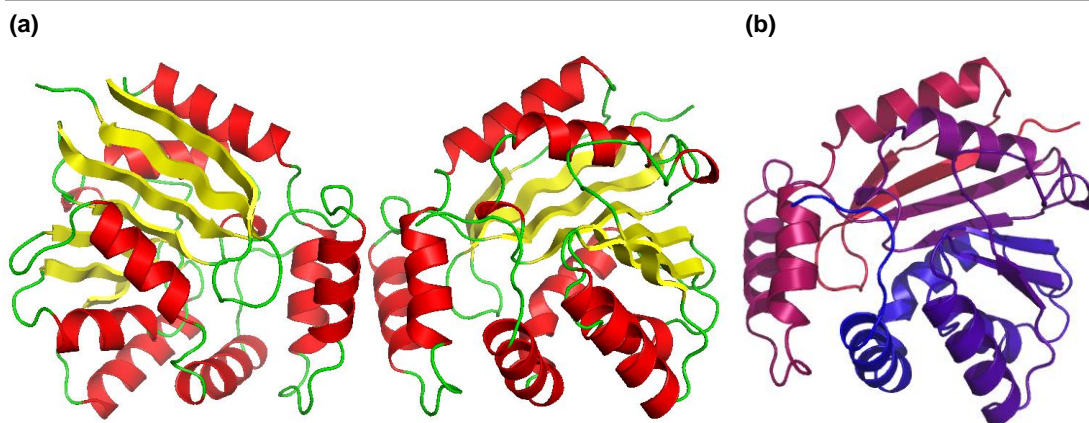
In general, crystals grown in the B10 condition had better morphology than those grown in the D7 condition, which were more often multiple or disordered. Additionally, a new condition (well condition F2) was identified which produced medium sized crystals with good morphology (**Figure 24**). An X-ray diffraction data set was collected on a crystal from this position with resolution of 1.9 Å at the European Synchrotron Radiation Facility Beamline ID23-1. Gratifyingly, the SelMet labelling was found to improve the quality of the data to the extent that it was possible for Dr. C-W. Chung to find a solution by iterative rounds of molecular replacement based on the data collected from the crystal from the F2 condition. The resulting X-ray crystal structure of SelMet NovO was deposited in the PDB with the accession code 5MGZ.



**Figure 24.** Images of crystals obtained from SelMet NovO crystallisation.

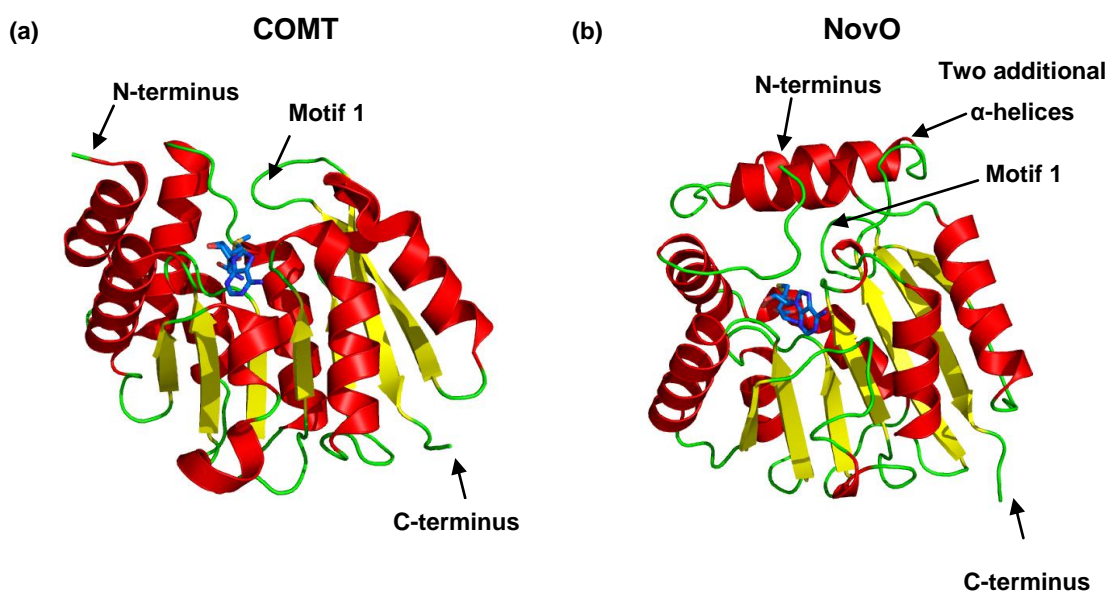
## 2.6 Analysis of the X-ray crystal structure of SelMet NovO complexed with SAH

SelMet NovO crystallised as a homodimer, with the C-termini pointing into solvent and the N-termini close to the centre of the dimer structure (**Figure 25a**). This is consistent with a previous study reported by Gruber and coworkers in which gel-filtration chromatography of NovO revealed a molecular weight corresponding to dimeric state.<sup>[112]</sup> Each monomer comprised a Rossmann fold (a seven membered  $\beta$ -sheet with a  $\beta$ -hairpin at the C-terminus, which is sandwiched by six  $\alpha$ -helices)<sup>[238]</sup> with two additional  $\alpha$ -helices at the N-terminus. This tertiary structure is typical for a small molecule methyltransferase, in which additional structural motifs at the N-termini are common.<sup>[55,239]</sup>



**Figure 25.** Solution to structure of SelMet NovO solved to 1.9 Å resolution coloured by (a) secondary structure and (b) terminus (N-terminus: blue, C-terminus: red).

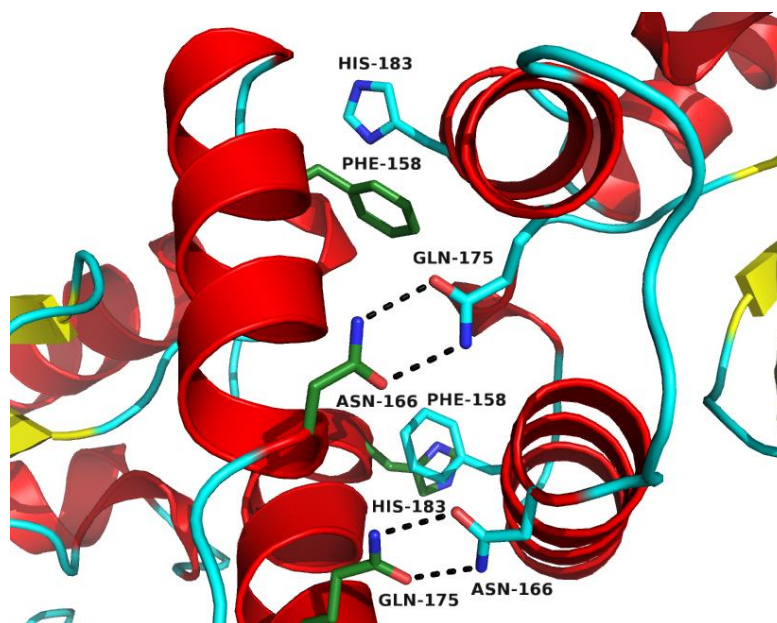
In comparison to the structure of COMT, which was discussed as an example of an archetypal Class I MT in Chapter 1, NovO is very similar in that it has two additional  $\alpha$ -helices at the N-terminus and ‘Motif 1’ is located in a flexible loop near the cofactor binding site. However, NovO also has two additional  $\alpha$ -helices between the first  $\alpha$ -helix and the second  $\beta$ -strand, as shown in **Figure 26**. It is these two  $\alpha$ -helices that form key interactions with the second monomer chain to form the homodimeric complex.



**Figure 26.** Comparison between COMT and NovO. SAM (COMT) and SAH (NovO) shown in blue.

With reference to the crystal structure, hydrogen bonds between the two monomer chains were observed between Q175 and N166 (of chains 1 and 2, respectively). In addition, F158

and H183 of chains 1 and 2 were found to lie in close proximity and may be involved in  $\pi$ -stacking interactions (**Figure 27**). However, it is not known whether the dimer structure is necessary for activity. This could be determined by point mutations at these residues to prevent H-bond formation between the two monomers and could be the subject of further structural characterisation work.



**Figure 27.** Interactions between monomer chains in SelMet NovO.

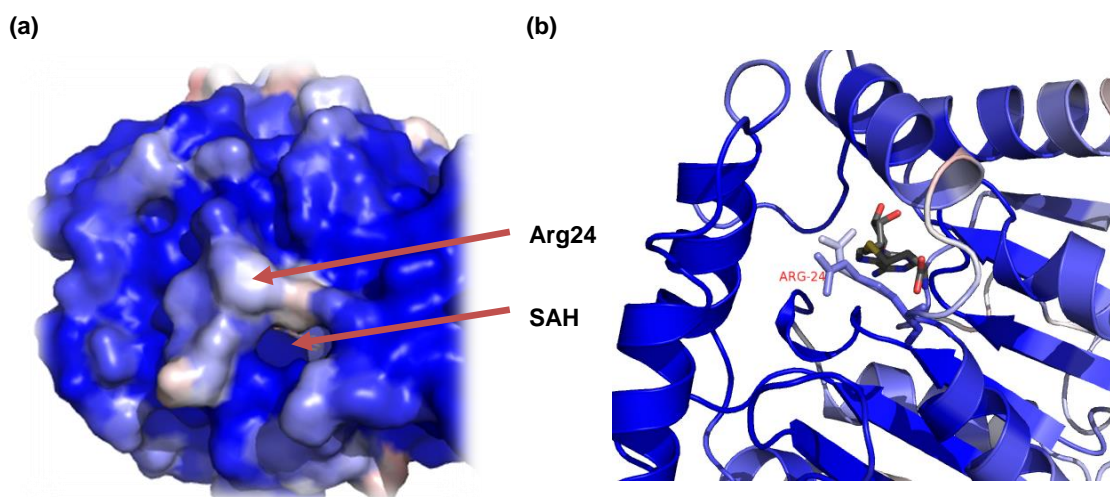
A second notable feature of the SelMet NovO crystal structure is the absence of an obvious pore for entrance of a substrate and exit of the product, despite the SAH binding site being buried within the centre of the protein. As the reaction occurs at the *S*-centre of the SAH molecule which is located near the centre of the protein structure, it can be reasonably assumed that the active site is also near the centre of the protein. With this in mind, it was proposed that the entrance to the active site may be *via* Arg24, which is a long, flexible amino acid and lies on the surface of the protein. The degree of flexibility of a residue in a protein is often quantified by B-factor.<sup>[164]</sup> The B-factor of a given residue arises from the displacement of atoms from their mean position in a crystal structure, which diminishes the scattered X-ray intensity. The B-factor is directly related to the mean square isotropic displacement ( $u$ ) of the atom, according to **Equation 1**.



$$B = 8\pi^2\langle u^2 \rangle$$

**Equation 1.** Dependence of B-factor on atomic displacement ( $u$ ) from the mean position of a scattering centre.

Based on this theory, analysis of the SelMet NovO crystal structure by B-value revealed Arg24 to be highly flexible with an average B-factor of ~32, compared to that of ~16 for surrounding residues (**Figure 28a**), and was further confirmed by the existence of two conformations in the crystal structure (**Figure 28b**).



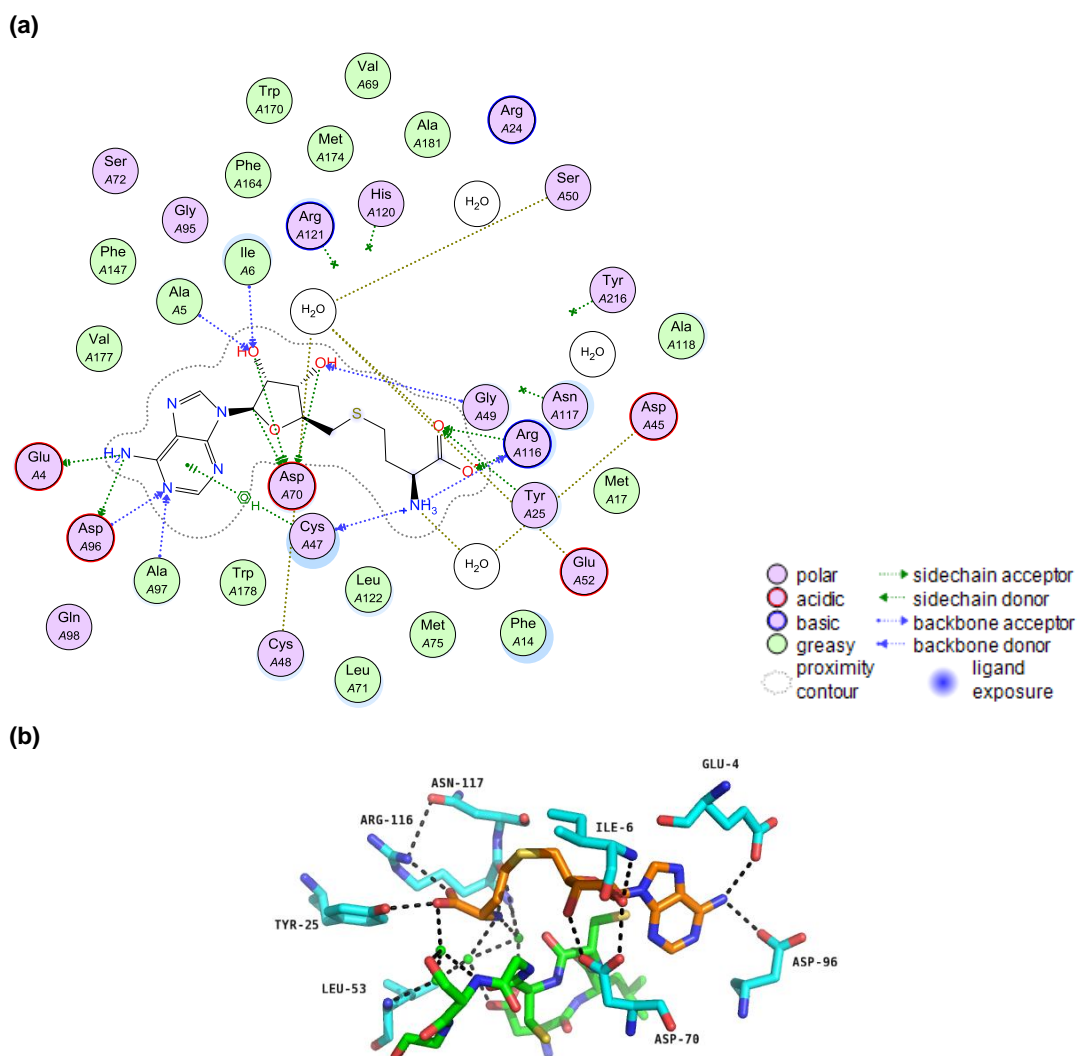
**Figure 28.** (a) Secondary structure of NovO coloured by B-factor (blue: low, white: high) showing the position of R24. (b) Surface map of whole protein showing R24 as ‘gatekeeper’ residue, with SAH shown in grey.

P169 was also found to have two possible conformations in the crystal structure (flipped up or down). This region of high flexibility was postulated to be the reason for a large degree of disorder in many of the crystals obtained, as the residue lies on the outside of the protein and is likely to be involved in intermolecular interactions.

### 2.6.1 Cofactor binding in NovO

A schematic of SAH binding in NovO is shown in **Figure 29a**. The strongest interaction between NovO and the methionine portion of the cofactor was seen between the carboxylate and guanidine moiety of Arg116. This residue is not well conserved across the methyltransferase family, despite being preceded by the highly conserved region from D111-S115. The adenosine portion of the cofactor has strong interactions with D96 and E4, which form H-bonding interactions with the  $-\text{NH}_2$  group of adenosine (**Figure 29a and 21b**). Interestingly, neither of these residues are conserved across the methyltransferase family

according to the sequence alignment in **Figure 30**. Finally, the ribose moiety of the cofactor was found to have strong interactions with D70, which is highly conserved across the superfamily. A feature common to Class 1 MTs is a highly conserved, glycine rich region known as ‘Motif 1’ consisting of E/DXXXGXG, where X is any residue.<sup>[55,112,239]</sup> Gruber *et al.* have previously proposed this motif was located between D45 and G51, with D45 and G49 being crucial for enzymatic function, whilst C47 and C48 make key hydrogen bonds

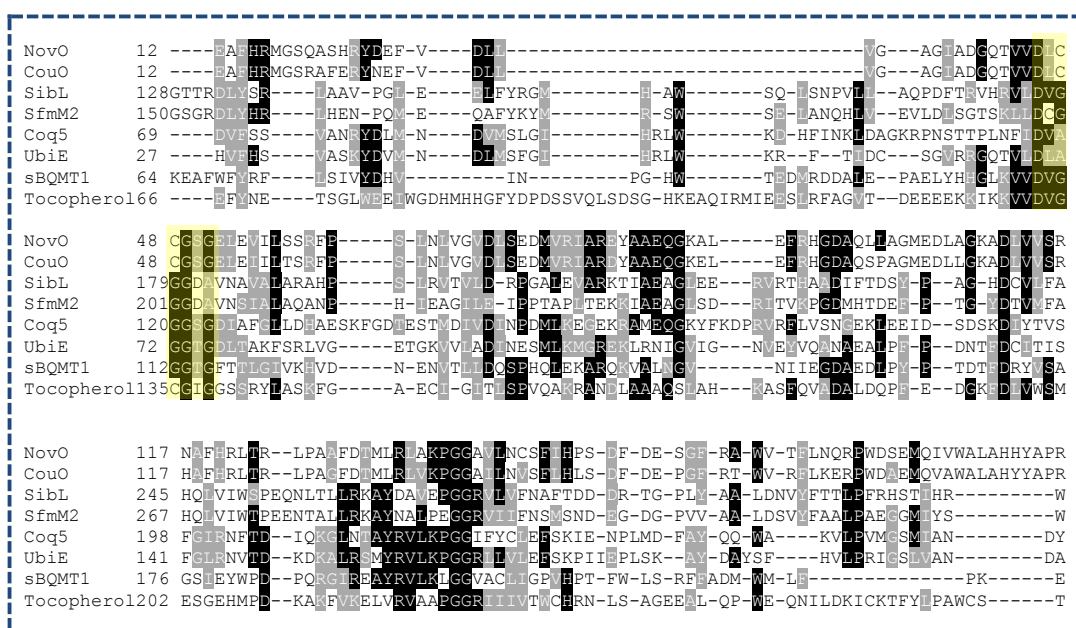


**Figure 29.** (a) Interactions map of residues involved in SAH binding in SelMet-NovO crystal structure. (b) SAH binding site with ‘Motif 1’ highlighted in green (D45, L46, C47, C48, G49, S50, G51), SAH highlighted in orange and other residues interacting with SAH highlighted in blue.

with the amino acid moiety of SAH *via* two water molecules (**Figure 29b**). Additionally, G49 and S50 (*via* an H<sub>2</sub>O molecule) also form H-bonding interactions with the cofactor. Residues in ‘Motif 1’ are highlighted in **Figure 29b**.



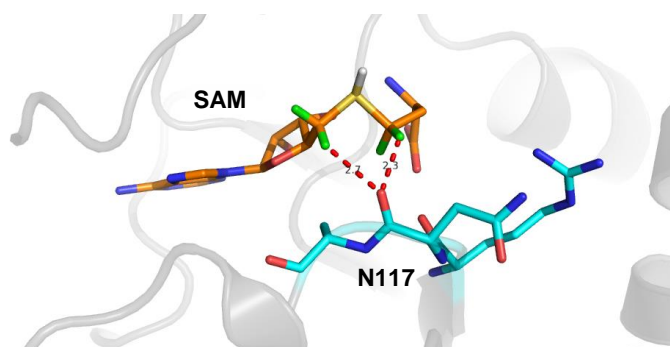
Analysis of the SAH binding pocket of NovO is further evidence of the very high degree of sequence and structural variation across the SAM-dependent MT family. To further investigate this for small molecule methyltransferases, the sequences of NovO and CouO were aligned with six other known small molecule MTs, a section of which is shown in **Figure 30**. The results revealed very low homology between the sequences, with only a few highly conserved residues across the alignment. Whilst the majority of these fall within the ‘Motif 1’ region (**Figure 30**, shaded in yellow), other conserved residues such as L57, D70 and G87 are found in isolated regions of the alignment. Finally, a highly conserved sequence of two glycine regions was found at positions 139 and 140 in NovO, which are located on a loop region between a  $\beta$ -sheet and an  $\alpha$ -helix in the Rossmann fold.



**Figure 30.** Section of sequence alignment of eight small molecule aromatic C-MTs including NovO and CouO. Results shaded coloured according to amino acid similarity (black: high; white: low). The highly conserved sequence E/DXXXGXG known as ‘Motif 1’ is highlighted in yellow. Multiple sequence alignment carried out using T-Coffee.<sup>[240]</sup>

The presence of CH-O hydrogen bonding between SAH and the active site was also analysed. Horowitz *et al.* proposed that this hydrogen bond network is highly conserved across the MT family irrespective of the enzyme’s class, active site structure or cofactor binding conformation.<sup>[70]</sup> As the crystal structure obtained was with SAH rather than SAM as the cofactor, the naturally occurring diastereomer of SAM was modeled using MOE and the potential for CH-O hydrogen bonds examined. As can be seen in **Figure 32**, the carbonyl oxygen atom of N117 acts as a potential H-bond acceptor, with distances of 2.3 Å and 2.7 Å

to the H-bond donors. This tentative evidence supports the theory proposed by Horowitz *et al.* and further suggests that such CH-O hydrogen bonds are conserved across the whole MT superfamily.



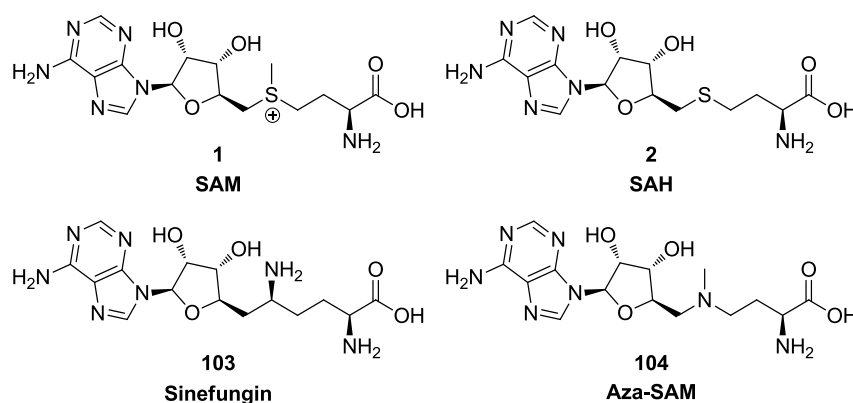
**Figure 31.** Proposed CH-O hydrogen bonding between N117 carbonyl oxygen and protons adjacent to the sulfonium centre of SAM.

## 2.7 Attempts towards obtaining a crystal structure of NovO with substrate bound in the active site

Having obtained a crystal structure of NovO complexed with SAH, efforts were then directed towards obtaining a crystal structure with both SAH and substrate bound in the active site of NovO. The two most common approaches to visualise ligand binding in X-ray crystallography are substrate soaking of protein crystals and cocrystallisation of the ligand and protein.<sup>[233]</sup> To investigate both of these strategies, crystals from the previously identified conditions D7 and B10 were grown for use in substrate soaking studies, in which the substrate was added to a final concentration of 50 mM as a solution in dimethylsulfoxide (DMSO). However, in all attempts the substrate precipitated out of solution as soon as it was added to the drop. This was thought to be due to the very poor aqueous solubility of substrates **67** and **68**. With this in mind, the cocrystallisation strategy was more promising. To this end, a PEGsII screen with NovO and both SAH and aminocoumarin substrate **67** or **68** was set up. Although some crystals were obtained, analysis at 0 and 90° indicated that the substrate ligand was not present in the active site.

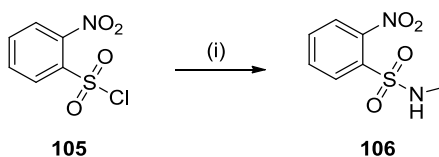
Cocrystallisation with non-reactive SAM analogues was then investigated. Due to the inherent reactivity of SAM, the use of non-reactive structural analogues has been a popular approach for studying SAM dependent enzymes. The two most commonly used analogues are the two regioisomers sinefungin (**103**) and ‘Aza-SAM’ (**104**) (**Figure 32**).<sup>[241–244]</sup>

Sinefungin is a naturally occurring antibiotic produced by *Streptomyces griseus* and *Streptomyces incarnatus* and is a potent inhibitor of SAM dependent enzymes.<sup>[245]</sup> Whereas SAM bears an electrophilic methyl group on the central sulfonium centre that is responsible for its reactivity, sinefungin displays a primary amine, rendering this a basic and nucleophilic centre. On the other hand, the regioisomer of sinefungin, Aza-SAM, is a non-natural synthetic analogue of SAM first reported by Thompson and coworkers in 1996.<sup>[246]</sup> In Aza-SAM, the sulfonium centre has been replaced with a tertiary amine, giving rise to a basic centre. Unlike sinefungin, Aza-SAM is not commercially available and therefore had to be synthesised for use in this project.



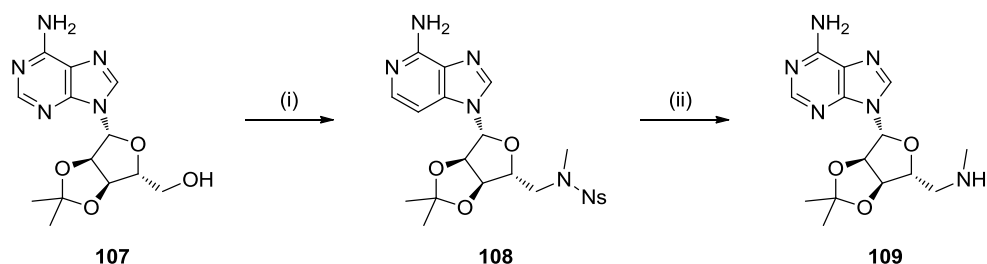
**Figure 32.** Structures of SAM (**1**), SAH (**2**), Aza-SAM (**103**) and sinefungin (**104**).

Based on the route developed by Nelson *et al.*, diastereomerically pure Aza-SAM was prepared in 13% yield over 5 steps from commercially available starting materials.<sup>[247]</sup> Initially, sulfonamide **106** was prepared in 90% isolated yield from the corresponding sulfonyl chloride **105** (Scheme 29).



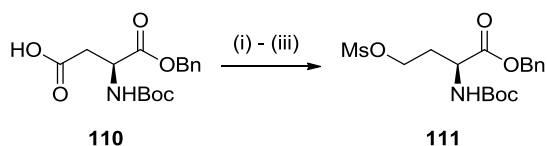
**Scheme 29.** Synthesis of sulfonamide **106**. Reagents and conditions: MeNH<sub>2</sub>, EtOH, CH<sub>2</sub>Cl<sub>2</sub>, 0-22 °C, 30 mins, 90%.

Next, an aza-Fukuyama-Mitsunobu reaction<sup>[247]</sup> was employed to couple adenosine acetamide **107** with the sulfonamide **106** to provide *N*-nosyl protected intermediate **108** in 48% yield. The resulting sulfonamide was deprotected on treatment with thiophenol to provide the desired secondary amine **109** in 98% yield (Scheme 30).



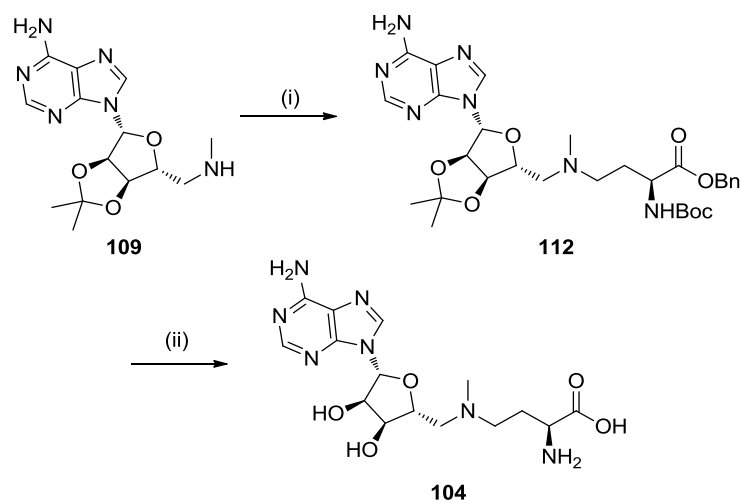
**Scheme 30.** Synthesis of *N*-methyl adenosine derivative **109**. *Reagents and conditions:* (i) **106**, PPh<sub>3</sub>, diethyl azodicarboxylate (DEAD), THF, 0–22 °C, 16 h, 48%, (ii) PhSH, MeCN, 22 °C, 16 h, 98%.

With the adenosyl fragment of the molecule in hand, it was necessary to prepare the amino acid fragment with an electrophilic centre to take part in the subsequent coupling reaction. Therefore, mesylate **111** was prepared from the commercially available *N*-Boc-*L*-aspartic *O*-benzyl ester (**110**) by reduction of the carboxylic acid to the primary alcohol, followed by treatment with MsCl to provide derivative **111** in 76% yield over 3 steps (**Scheme 31**).



**Scheme 31.** Synthesis of amino acid **111** from aspartic acid derivative **110**. *Reagents and conditions:* (i) isobutyl chloroformate, *N*-methylmorpholine, THF, –20 °C, 10 min. (ii) NaBH<sub>4</sub>, MeOH, –20 °C, 10 min. (iii) MsCl, NEt<sub>3</sub>, CH<sub>2</sub>Cl<sub>2</sub>, 0–22 °C, 30 min, 76% over three steps.

Next, the two fragments **109** and **111** were coupled *via* an iodide catalysed S<sub>N</sub>2 reaction to provide the *O*- and *N*- protected Aza-SAM **112** in 44% yield. Finally, the molecule was globally deprotected by treatment with base then acid to provide Aza-SAM (**104**) in 83% isolated yield (**Scheme 32**). This provided sufficient material to support crystallographic studies.

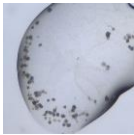



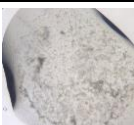







**Scheme 33.** Coupling and deprotection to synthesis Aza-SAM (**104**). *Reagents and conditions:* (i) **111**, DIPEA, MeCN, 70 °C, 24 h, 44%. (ii) aq. NaOH, MeCN, 22 °C then aq. HCl, 22 °C, 90 min, 83%.

## 2.8 Crystallisation screens with substrate and non-reactive SAM analogues

With the SAM analogues SAH, Aza-SAM and sinefungin in hand, further crystallisation screens were set up in an attempt to obtain a crystal structure of NovO with substrate bound. As previous attempts at cocrystallisation had been unsuccessful, a substrate soaking approach was pursued to obtain a crystal structure with substrate bound in the active site. In order to obtain suitable crystals for soaking, three crystallisation screens were set up with SAH, Aza-SAM and sinefungin, each with two different protein concentrations. For each ligand of the three cofactor analogues, a full PEGsII screen was run in addition to an optimisation screen around the PegsII D7 condition (PegsIIOptiD7).

Whilst the majority of experiments resulted in precipitation, good quality crystals were obtained in three drops (**Table 6**, entries **6**, **7** and **9**), out of which the largest crystals were obtained in the presence of SAH (**Table 7**, entries **6** and **7**). No crystals were observed in the presence of sinefungin. All three drops yielding crystals contained the same buffer (0.1 M Mes, pH 6.2) and either 30.91% w/v or 36.36% w/v PEG-4000 and did not contain any additional inorganic salts. Whilst crystals were grown from both concentrations of NovO, larger crystals were obtained from the higher concentration (7 mg/mL stock concentration).

Entry	Screen	Ligand	Image	Condition
1	PEGsII	SAH		3.4 mg/mL NovO, 0.2 M MgCl <sub>2</sub> , 0.1 M Tris pH 8.5, 30%w/v PEG-4000
2	PEGsII	Aza-SAM		3.4 mg/mL NovO, 0.1 M CaCl <sub>2</sub> , 0.1 M NaOAc pH 4.6, 15%w/v Peg-400
3	PEGsII	Aza-SAM		7 mg/mL NovO, 0.2 MgCl <sub>2</sub> , 0.1 M Mes pH 6.5, 25%w/v PEG-4000
4	PEGsII	Sinefungin		3.4 mg/mL NovO, 0.1 M CaCl <sub>2</sub> , 0.1 M NaOAc pH 4.6, 15%w/v PEG-400
5	PEGsII	Sinefungin		7 mg/mL NovO, 0.1 M CaCl <sub>2</sub> , 0.1 M NaOAc pH 4.6, 15%w/v PEG-400
6	PEGsIIOptiD7	SAH		3.4 mg/mL NovO, 0.1 M Mes pH 6.2, 31%w/v PEG-4000
7	PEGsIIOptiD7	SAH		7 mg/mL NovO, 0.1 M Mes pH 6.2, 31%w/v PEG-4000
8	PEGsIIOptiD7	Aza-SAM		7 mg/mL NovO, 0.1 M Mes pH 6.5, 38%w/v PEG-4000
9	PEGsIIOptiD7	Aza-SAM		3.4 mg/mL NovO, 0.1 M Mes pH 6.4, 36%w/v PEG-4000
10	PEGsIIOptiD7	Sinefungin		7 mg/mL NovO, 0.1 M Mes pH 6.1, 20%w/v PEG-4000

**Table 7.** Results of crystallisation screens with SAH, Aza-SAM (**104**) and sinefungin (**103**).

Based on these results, a further matrix optimisation experiment was carried out to determine the optimum PEG concentration (range: 30-40% w/v) and buffer pH (range: pH 6.1-6.25).

The higher concentration of NovO stock was used throughout the experiment. This screen yielded many crystals, but the largest and highest quality crystal grew in 0.1 M Mes buffer at pH 6.25 with 33% w/v PEG-4000 as the precipitant (**Figure 33**).



**Figure 33.** Image of NovO crystal grown under optimised conditions. *Growth conditions:* 3.5 mg/mL NovO, 20 mM SAH, 0.1 M Mes pH 6.25, 32.73% w/v PEG-4000. Total drop volume: 200 nL.

Next, the crystals were harvested and substrate soaks attempted. Unfortunately, due to the very poor aqueous solubility of substrates **67** and **68**, precipitation was observed as soon as a solution of the substrate (as a concentrated solution in DMSO) was added to the drop. Nonetheless, it was possible to harvest one crystal that had been soaked in substrate **68**. A dataset was collected with resolution of  $\sim 2.5$  Å, however no substrate was observed in the active site of the protein. The experiments were repeated, but unfortunately all attempts resulted in the crystals disintegrating upon collection or precipitation of the substrate.

Therefore, despite extensive efforts to obtain a crystal structure with substrate bound in the active site, attempts were unsuccessful. With this in mind, it was decided instead to carry out computational ligand docking studies using the SelMet crystal structure of NovO in order to understand the mode of substrate binding within the active site.

## **Chapter 3.**

# **Probing the Active Site of NovO and CouO**



### 3 Probing the Active Site of NovO and CouO

This chapter describes the characterisation of substrate binding within the active site of NovO using aminocoumarin **67**. As attempts to obtain a crystal structure of NovO with substrate **67** or **68** bound in the active site were unsuccessful, computational ligand docking was carried out to generate a substrate binding model.

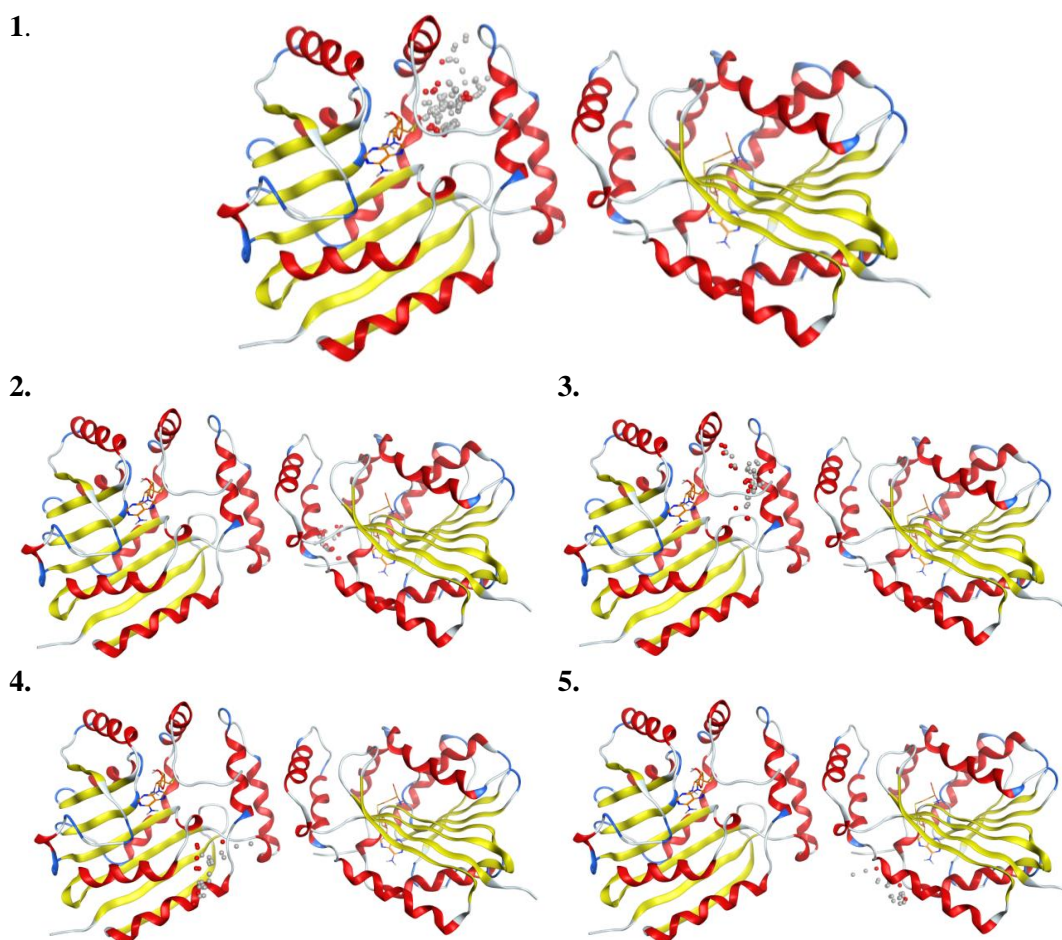
#### 3.1 Preparation and analysis of a NovO-substrate docking model

Substrate docking studies were carried out using Molecular Operating Environment (MOE)<sup>[248]</sup> software to provide a model of the substrate binding in the active site of NovO. This would be used as a starting point from which to probe substrate-protein interactions by mutational analysis and structure-activity relationships. Initially, the X-ray crystal structure was prepared using the Structure Preparation command followed by optimisation of the H-bonding network using the Protonate 3D command, and finally calculation of partial charges. Next, putative binding sites in NovO were identified using the Site Finder application.<sup>[249]</sup> The default settings for these procedures were used in all cases. The Site Finder application ranks putative binding sites based on their propensity for ligand binding (PLB),<sup>[250]</sup> also taking into account the size, number of hydrophobic residues and number of side chain contact atoms for each site (**Table 8**).

These putative binding sites were further analysed manually based on their proximity to the SAH present in the crystal structure, from which only Site 1 was found to be in a suitable location relative to SAH (**Figure 34**). Therefore, this site was selected for carrying out the substrate docking experiment, using dummy atoms to mark the putative binding site.

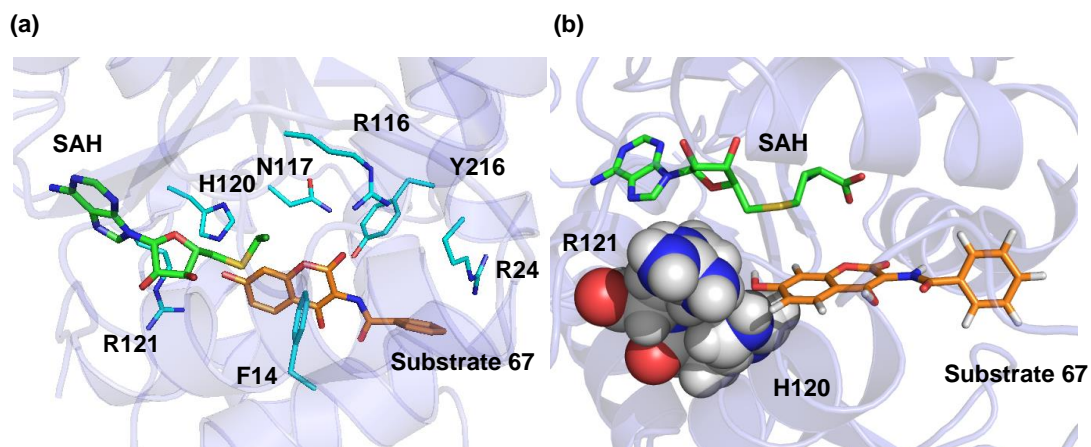
Site	PLB	Residues
1	3.6	I6, F14, M17, G18, Q20, A21, H23, R24, Y25, R116, N117, A118, H120, R121, F127, F164, Q167, R168, P169, W170, D171, M174, V177, W178, A181, Y216.
2	1.05	M1, K2, I3, E4, A6, I6, T7, E10, A11, H120, R121, I176, V177, L180, A181, W184, A185
3	1.05	E155, R159 (Chain 1), M1, K2, I3, E4, A6, I6, T7, E10, A11, H120, R121, I176, V177, L180, A181, W184, A185 (Chain 2)
4	0.77	K2, F119, H120, L122, T123, R124, L125, P126, Y184, A185, P186, D190, E193, A194, Q197
5	0.13	K2, L122, T123, R124, L125, P126, Y184, A185, P186, D190, E193, A194

**Table 8.** Top 5 ligand binding sites identified by the Site Finder application using MOE. PLB: propensity for ligand binding.<sup>[250]</sup>



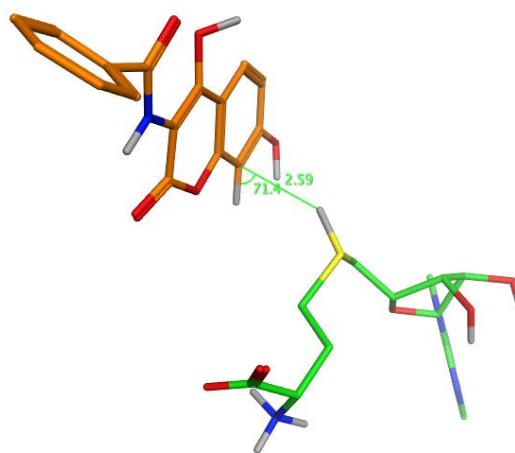
**Figure 34.** Top 5 results from Site Finder search using MOE software relative to the position of SAH, which is shown in orange. Dummy atoms are represented as white and red spheres.

Next, the Dock application was employed to model substrate **67** into the putative binding site. The top 5 hits were analysed manually and the final model selected based on proximity of the coumarin C-8 to the sulfur centre of SAH and the potential energy of the substrate, giving rise to the final model shown in **Figure 35a**, in which the coumarin ring adopted a planar conformation.



**Figure 35.** (a) Model of substrate **67** bound in the active site of NovO with SAH obtained from docking studies. (b) ‘Sandwiching’ effect of substrate hydroxyl group between H120 and R121.

When a methyl group was modeled onto the sulfur centre of SAH to model the naturally occurring *S*- diastereomer of SAM,<sup>[251]</sup> the pendant methyl group was 2.59 Å from C-8, at an angle of 109° from the coumarin scaffold (**Figure 36**).



**Figure 36.** Conformation of substrate **67** in the substrate binding model, showing the distance between coumarin C-8 and the electrophilic methyl group of SAM.

Next, the residues forming the substrate binding pocket were analysed and a set of residues were identified that may be involved in substrate binding (F14, R116, N117, H120, R121 and Y216) (**Figure 35a**). Additionally, R24, which has previously been proposed as a gatekeeper residue, was found at the entrance to the proposed substrate binding site. Out of these residues, H120 and R121 were of particular interest as they were observed to ‘sandwich’ the hydroxy group *ortho* to the site of methylation (**Figure 35b**). Based on this model, it was proposed that one or both of these basic residues were involved in the catalytic mechanism and/or were necessary for substrate activation. Furthermore, this would rationalise the strong dependence of catalytic activity on the presence of a hydroxy group *ortho* to the site of methylation that was observed during substrate screening experiments.

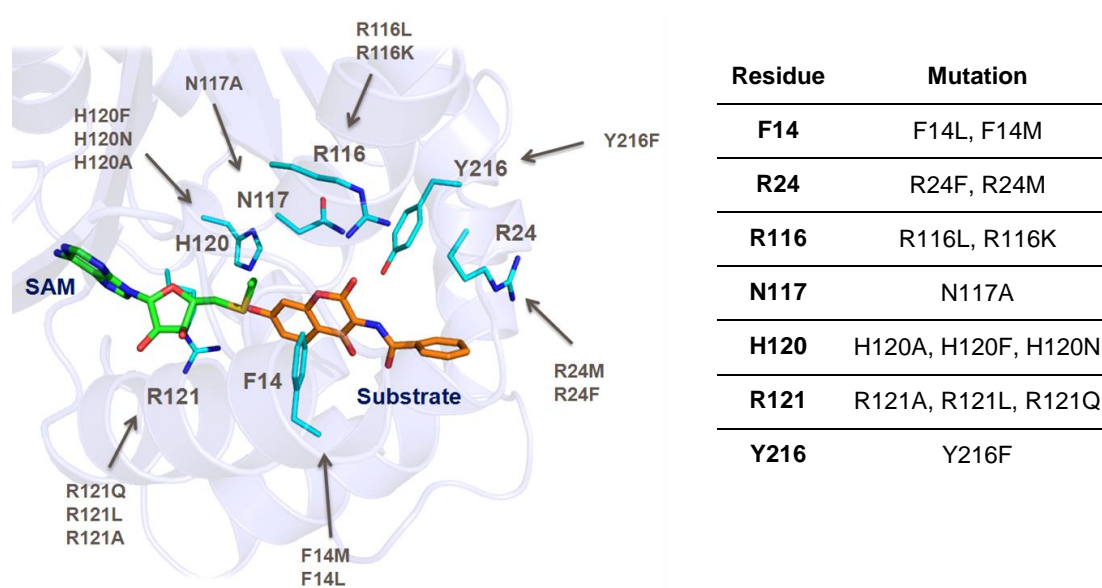
### 3.2 Mutational analysis of putative substrate binding residues

Mutational analysis is a commonly used strategy to probe the function of individual residues in a protein. For example, Diver and Long have used scanning mutagenesis to elucidate the active site region of isoprenylcysteine carboxyl methyltransferase and Sankpal and Rao studied the effect of mutating conserved residues on methyltransferase activity of the DNA methyltransferase HhaI.<sup>[109,110]</sup> Therefore, to interrogate the substrate binding model, a library of single point mutants of NovO (**Figure 37**) was designed to address the following hypotheses:

- **F14** was proposed to have edge-to-face interactions with the coumarin ring. Mutating this residue to leucine (**F14L**) was predicted to remove this interaction and destabilise the enzyme-substrate complex. **F14M** was designed in order to restore a small degree of enzyme-substrate interaction at this position, as the ‘Methionine-Aromatic Motif’ has been shown to have an important role in protein stabilisation.<sup>[252]</sup> Specifically, the slightly acidic  $-S-CH_3$  protons have been shown to interact with the  $\pi$ -system of neighboring aromatic rings, thereby mimicking the proposed edge-to-face interactions.
- **R24** is the proposed ‘gatekeeper’ residue to the active site. **R24M** was chosen to maintain steric bulk and flexibility, whilst removing H-bond acceptor properties of arginine. **R24F** was predicted to have a much lower degree of flexibility and was hypothesised to inhibit substrate entrance to the active site.
- **R116** has been predicted to H-bond to the coumarin carbonyl oxygen, in addition to its involvement in SAM binding. **R116L** was designed to remove both of these interactions whilst maintaining a similar level of steric hindrance. With the aid of molecular

modeling, **R116K** was designed to maintain interaction with SAM, whilst removing H-bond donation to the coumarin carbonyl oxygen.

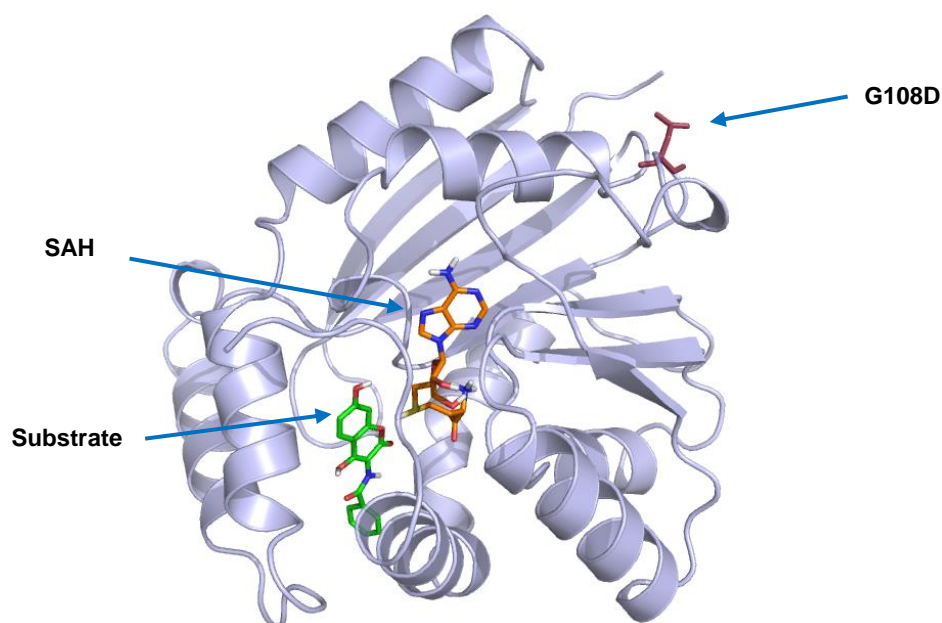
- **N117** was also hypothesised to form an H-bond to the coumarin carbonyl oxygen. Therefore, **N117A** was designed to investigate the extent to which this interaction contributes to substrate binding.
- **H120** and **R121** have been observed to form very close interactions with the 7-hydroxy group and are proposed to be key residues in substrate binding and recognition. Thus, we wished to investigate a range of mutations at these positions. In addition to the alanine mutants, **H120F** and **R121L** were proposed to maintain a similar shape of substrate binding pocket whilst removing H-bond acceptor properties. Additionally, **H120N** and **R121Q** were designed to investigate the effect of having an H-bond donor at this position whilst maintaining a similar level of steric bulk around the substrate.
- The substrate binding model showed **Y216** to be in close proximity to the amide carbonyl and coumarin ring. Therefore, it was hypothesised that H-bonds between the tyrosine hydroxy group and amide nitrogen or lactone carbonyl of coumarin **67** may be involved in substrate binding. **Y216F** was designed to test the effect of removing this hydroxy group from the active site on the rate of methylation.



**Figure 37.** Proposed library of single point mutants to investigate the substrate binding model.

Site-directed mutagenesis (SDM) was carried using KOD XL DNA polymerase with the pET26b(+) plasmid harbouring the gene for wild-type (WT) NovO to prepare the desired NovO mutants. From the 14 desired mutations, 10 hits were identified and confirmed by

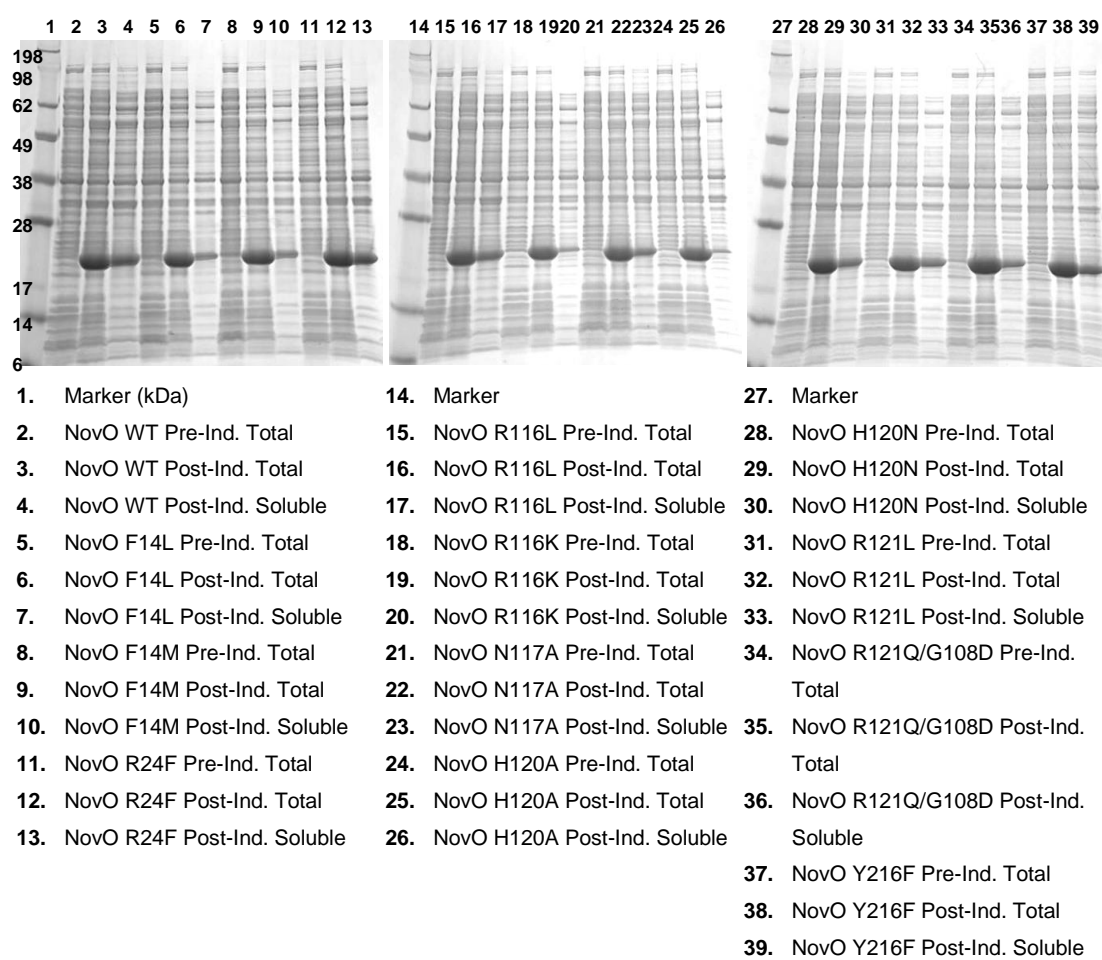
gene sequencing. Out of the remaining 4 mutations (R24M, H120F, R121A and R121Q), one hit was identified for R121Q, however a second mutation (G108D) was also introduced, generating the double mutant R121Q/G108D. Referring to the crystal structure of NovO, G108 was found to lie on a flexible loop at the enzyme-solvent interface (**Figure 38**). Therefore, no direct effect on the substrate binding or the catalytic cycle of the enzyme was predicted for G108D. Based on this rationale, it was decided to use this mutant for investigations of R121Q.



**Figure 38.** Position of G108D in NovO monomer of the SelMet NovO crystal structure.

Despite repeated attempts at the unsuccessful SDM experiments, no hits were obtained for mutants R24M, H120F or R121A. Therefore, work on preparing these mutants was stopped, as other mutants to investigate these residues (R24F, H120A, H120N, R121L and R121Q/G108D) had successfully been prepared.

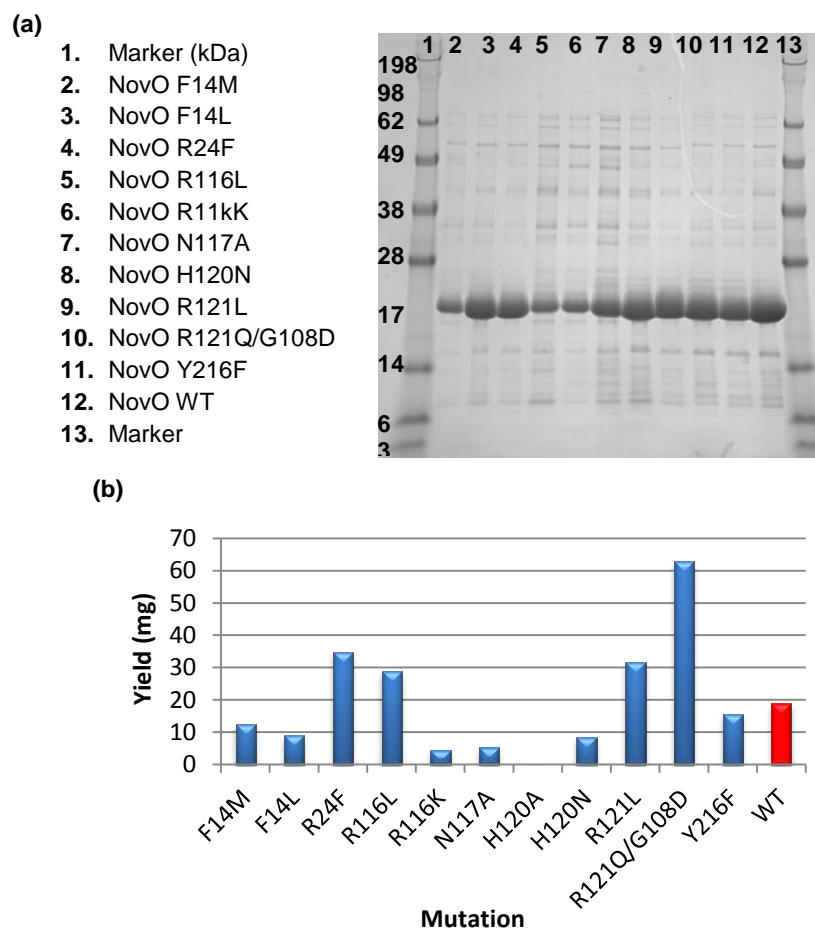
With the plasmid DNA in hand, the mutants of NovO were overexpressed in *E. coli* using the previously optimised expression conditions (Chapter 2). Overexpression of the desired mutants was observed in all cases, however levels of soluble expression varied between mutants. In particular, H120A had very low levels of soluble expression. In contrast, there was a high level of soluble expression for R121Q/G108D compared to the WT enzyme (**Figure 39**). As G108D lies at the solvent interface on the crystal structure of NovO, it was hypothesised that this could be a beneficial mutation for soluble expression levels.



**Figure 39.** Analysis of overexpression of NovO WT and NovO mutants by SDS-PAGE. Theoretical mass of WT 6xHis-NovO: 26.2 kDa. Ind.: induction.

Next, the mutants were purified by affinity chromatography and the total protein yield for each mutant was calculated according to the protein concentration after desalting each sample using a PD-10 cartridge. As shown in **Figure 40**, the highest yield was obtained for R121Q/G108D (approximately three times higher than for WT NovO). This correlates with the high levels of soluble expression observed for this mutant and is hypothesised to be due to the presence of an acidic residue at the solvent interface of the enzyme. The next highest yields were obtained for R24F, R116L and R121L, followed by R14M, Y216F and WT. Comparatively low yields were obtained for R116K, N117A and H120N, presumably due to much lower levels of soluble expression for these mutants. The purification of H120A was unsuccessful on three successive attempts, which is thought to be due to very low levels of soluble expression. Therefore, it was not possible to obtain a sample of purified NovO H120A and so the crude cell lysate was used instead for the activity assay with this mutant.





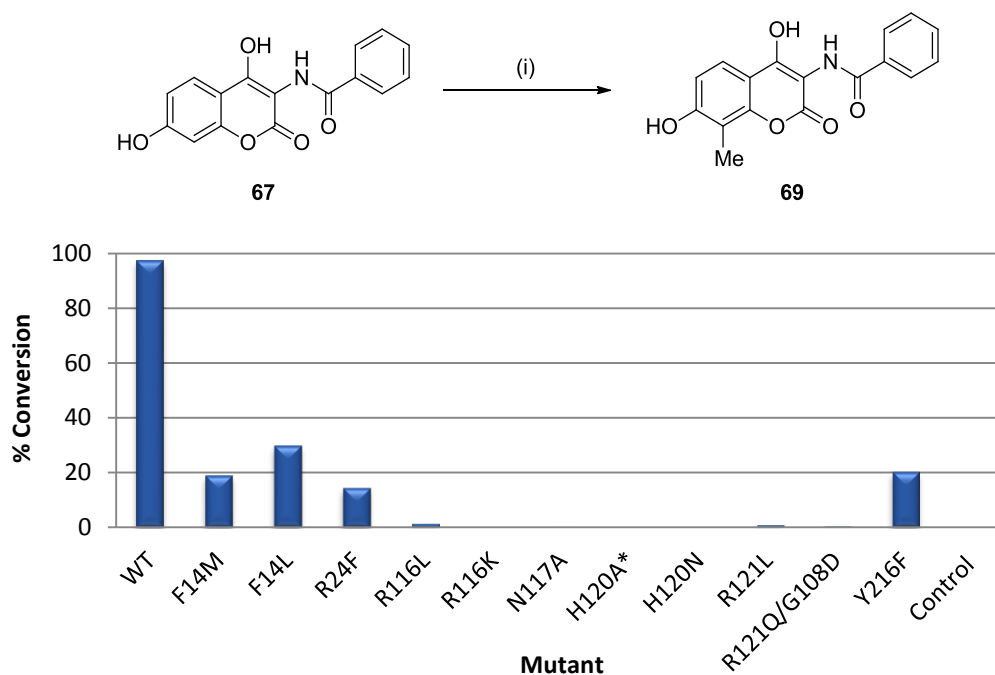
**Figure 40.** (a) SDS PAGE of purified NovO mutants. Theoretical mass of 6xHis-NovO WT: 25.5 kDa. (b) Total protein yield from 100 mL culture for expression of NovO WT and single point mutants.

### 3.3 Methyltransferase assay with NovO and NovO mutants

Having prepared purified samples of NovO and a series of mutants designed to probe the residues comprising the proposed substrate binding site, the effect of each mutation relative to the WT enzyme was investigated. In order to draw a quantitative comparison between the activity of NovO and the mutants, kinetic data on each of the mutants was collected over a range of substrate concentrations, running experiments in triplicate in each case. However, in all cases it was not possible to generate a Michaelis-Menten plot. This was due to very low levels of conversion (<2%) being observed, resulting in a high degree of error in the data collected by HPLC. In instances where meaningful data was collected, the activity of the mutants was found to be so low that the substrate concentrations required to provide sufficient data points to generate Michaelis-Menten plots overloaded the HPLC instrument, introducing a high degree of error into the data. Therefore, the activity of mutants was



assessed according to percentage conversion of **67** to methylated product **69** by area/area% HPLC after 24 hours (**Figure 41**). Due to the insolubility of mutant H120A, the cell lysate from the overexpression of this mutant was used to measure percentage conversion.

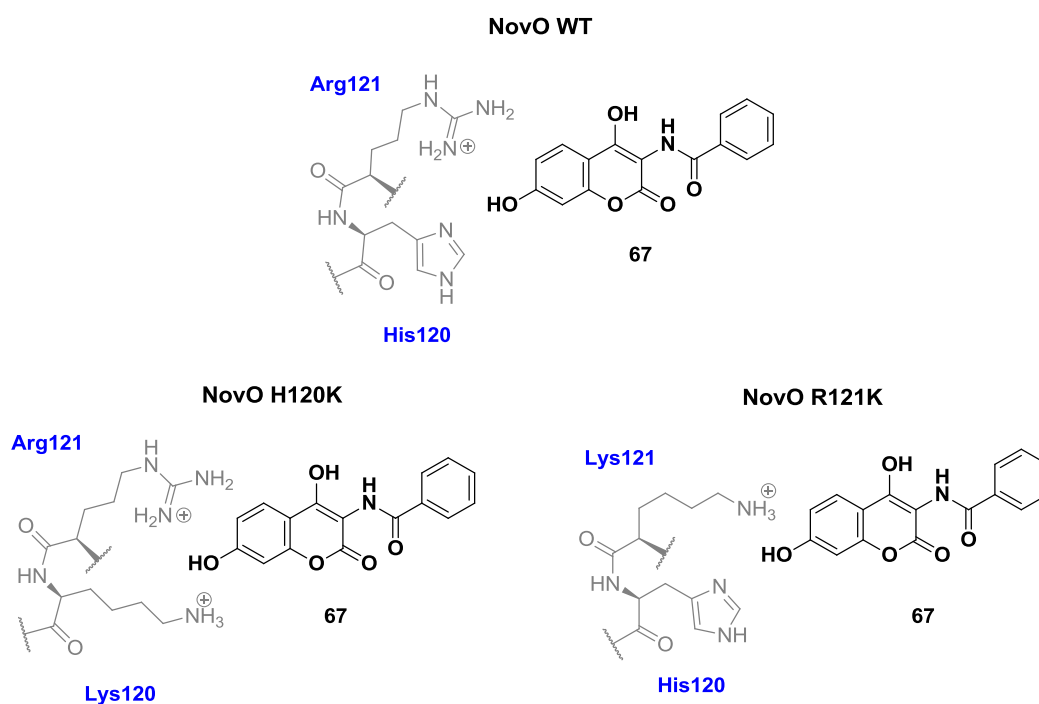


**Figure 41.** Percentage (%) methylation of substrate **67** after 24 hours by NovO WT relative to NovO mutants. \*Due to the low levels of soluble expression of H120A, cell lysate was used. Control: No protein added. *Reagents and conditions:* (i) NovO WT or mutant (19  $\mu$ M), 2 mM SAM, 0.1 mg/mL BSA, potassium phosphate buffer pH 6.5 containing 5 v/v% DMSO, 37  $^{\circ}$ C, 24 h.

Whilst the WT protein exhibited nearly quantitative conversion of substrate **67** (at a reaction concentration of 0.1 mM) in 24 hours, mutants F14M, F14L, R24F and Y216F showed reduced activity with only ~15-30% conversion. Conversely, mutations at positions R116, N117, H120 and R121 nearly abolished all methyltransferase activity (**Figure 41**). These data indicate that whilst residues F14, R24F and Y216 may be involved in substrate binding, they are not critical to the catalytic mechanism of NovO. On the other hand, residues R116, N117, H120 and R121 were hypothesised to be critical for substrate binding or the catalytic mechanism. H120 and R121 were of particular interest, as they were observed to make H-bonding interactions with the 7-hydroxy group which is adjacent to the site of methylation and were therefore the subject of further investigation.

### 3.4 Investigation into the functions of H120 and R121

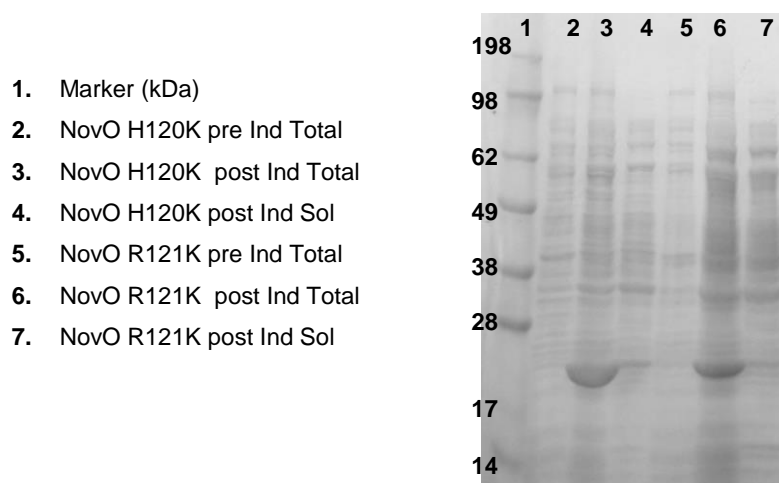
Based on the results of the SDM experiments and substrate binding model, which indicated that H120 and R121 were important for catalytic function, the role of these two residues was probed. It was hypothesised that this could be investigated by sequentially mutating H120 and R121 for a lysine (K) residue, to provide the single point mutants NovO H120K and R121K, respectively. The pKa values of the protonated side chain amine of lysine, guanidine of arginine and of histidine are 10.53, 12.48 and 6.0, respectively.<sup>[108]</sup> As such, under the assay conditions of pH 6.5, it was hypothesised that H120 would be deprotonated, whilst R121, H120K and R121K would be protonated. Therefore, if the catalytic mechanism involved a deprotonation of the 7-hydroxy group prior to methyl transfer, it was predicted that conversion would only be observed in the case of NovO R121K, which would still contain H120 as the free base. In the case of NovO H120K, there would be no free base available at this position for carrying out a deprotonation event and the enzyme was predicted to be inactive (**Figure 42**).



**Figure 42.** Proposed single point mutants NovO H120K and R121K and their interaction with substrate 67.

Based on this hypothesis, the single point mutants H120K and R121K were prepared by SDM using Q5 DNA polymerase, as two attempts using the previously employed KOD XL DNA polymerase had been unsuccessful. Expression of the two mutants was carried out

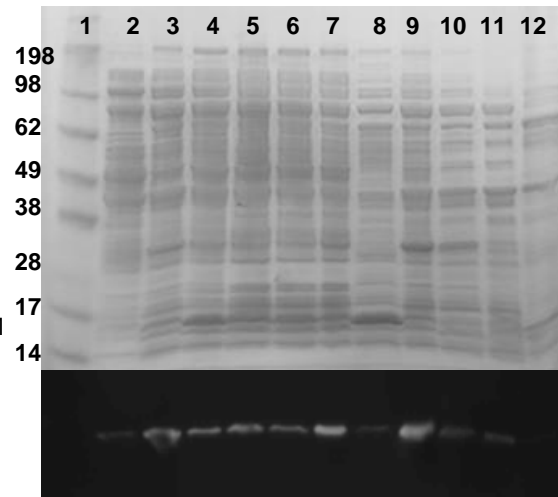
using the previously identified expression conditions in Magic Media. Although levels of total expression were high in both cases, very low levels of soluble expression were observed, especially for the H120K mutant (**Figure 43**).



**Figure 43.** Analysis of overexpression of NovO H120 and NovO R121K. Enzymes expressed in *E. coli* BL21 (DE3) in Magic Media at 30 °C to OD ~2 then at 18 °C overnight and analysed by SDS PAGE. Theoretical mass of WT 6xHis-NovO: 26.2 kDa.

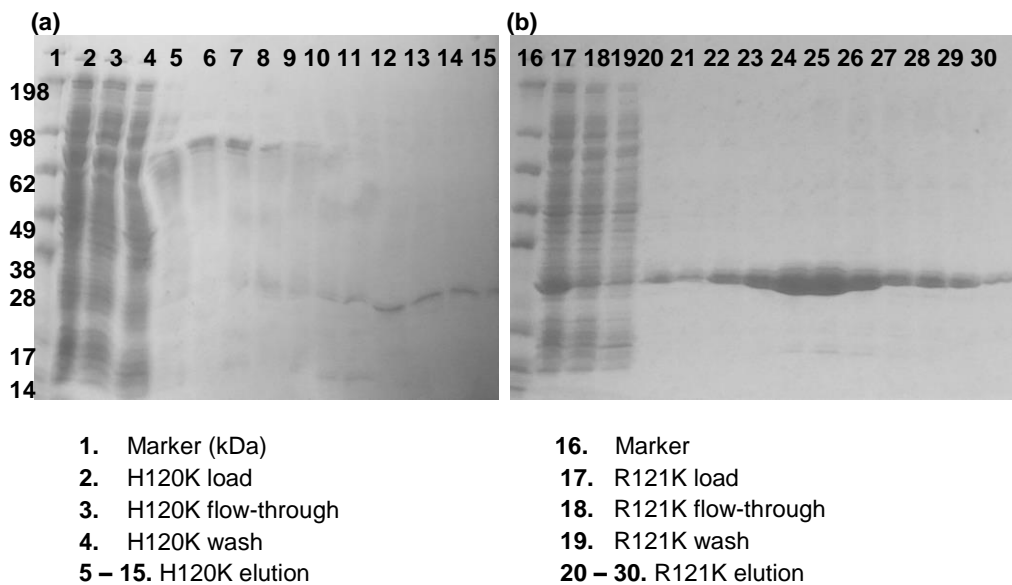
Therefore, a screen of alternative lysis conditions was undertaken to identify a lysis buffer that provided enough soluble protein for purification by affinity chromatography. Many additives have been reported to improve the solubilisation and stability of a protein. For example, isopropanol has been demonstrated to break up non-specific aggregation and glycerol can improve stability and solubilisation.<sup>[253–255]</sup> These additives have been employed by Lindwall *et al.* to design a sparse matrix approach to the solubilisation of overexpressed proteins, which they demonstrated with  $\alpha$ -tubulin.<sup>[256]</sup> Ten lysis buffers from this study were prepared and screened with NovO H120K and the soluble fraction in each case analysed by SDS PAGE. BugBuster™ was also tested as it had been used previously to analyse the level of total and soluble protein expression of NovO H120K and R121K (**Figure 43**) and was therefore used a control experiment. In addition, a Western Blot was carried out to confirm the small bands observed in the SDS PAGE were His-tagged protein (**Figure 43**).<sup>[257,258]</sup>

1. Marker (kDa)
2. BugBuster™
3. 100 mM Tris, 10% glycerol, pH 7.6
4. 100 mM Tris, 50 mM LiCl, pH 7.6
5. 100 mM HEPES, 50 mM (NH<sub>4</sub>)SO<sub>2</sub>, 10% glycerol, pH 7
6. 100 mM HEPES, 100 mM KCl, pH 7
7. 100 mM Tris, 50 mM NaCl, 10% isopropanol, pH 8.2
8. 100 mM potassium phosphate, 50 mM (NH<sub>4</sub>)SO<sub>2</sub>, 1% Triton X-100, pH 6
9. 100 mM triethanolamine, 100 mM KCl, 10 mM DTT, pH 8.5
10. 100 mM Tris, 100 mM NaCl, 100 mM glutamine, 10 mM DTT, pH 8.2
11. 100 mM triethanolamine, 50 mM LiCl, 5 mM EDTA, pH 8.5
12. 10 mM sodium acetate, 100 mM KCl, 10 mM DTT, pH 5.5



**Figure 44.** Analysis of lysis buffer conditions for the solubilisation of NovO H120K by SDS-PAGE (top) and Western Blot (bottom). Theoretical mass of WT 6xHis-NovO: 26.2 kDa.

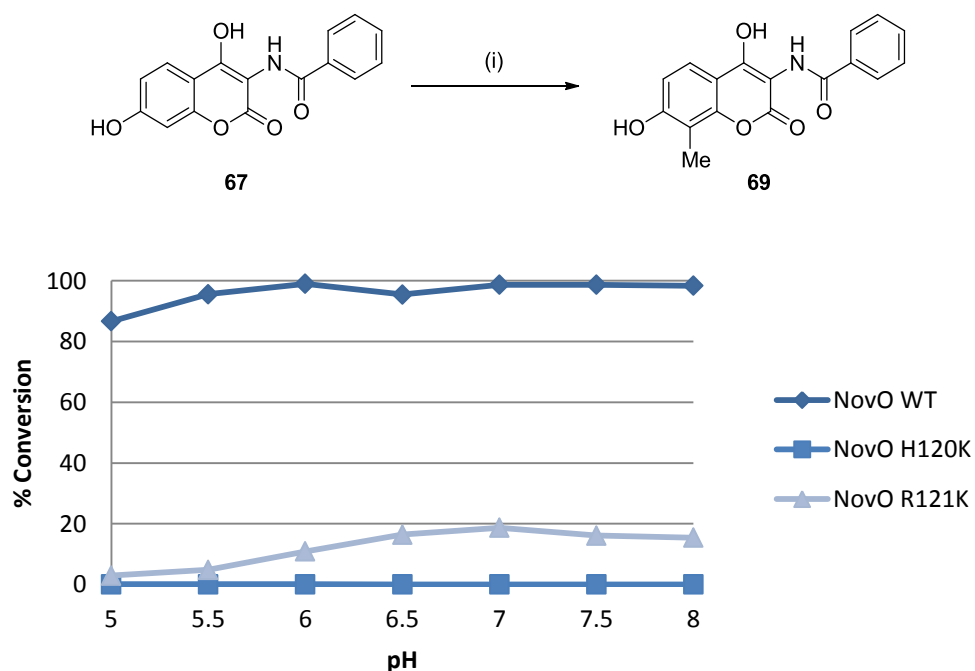
This experiment identified a number of lysis buffers that resulted in higher levels of soluble protein than BugBuster™, from which the conditions in lanes **7** and **9** (**Figure 44**) gave the highest levels of soluble protein. From these two conditions, 100 mM Tris, 50 mM NaCl and



**Figure 45.** Analysis of purification of (a) NovO H120K and (b) NovO R121K by Ni-NTA affinity chromatography by SDS PAGE. Theoretical mass of WT 6xHis-NovO: 26.2 kDa.

10% isopropanol was selected as it was more similar to the previously employed lysis buffer and did not contain DTT, which can cause denaturation of proteins at high concentration.<sup>[259]</sup> The cell pellets from the overexpression of NovO H120K and NovO R121K were resuspended in the isopropanol lysis buffer and lysed by sonication before purification by Ni-NTA affinity chromatography. The load, flow-through, wash and elution fractions were analysed by SDS PAGE (**Figure 45**), which showed that both proteins had been purified. In particular, high levels of NovO R121K were obtained (**Figure 45a**) due to higher levels of soluble expression.

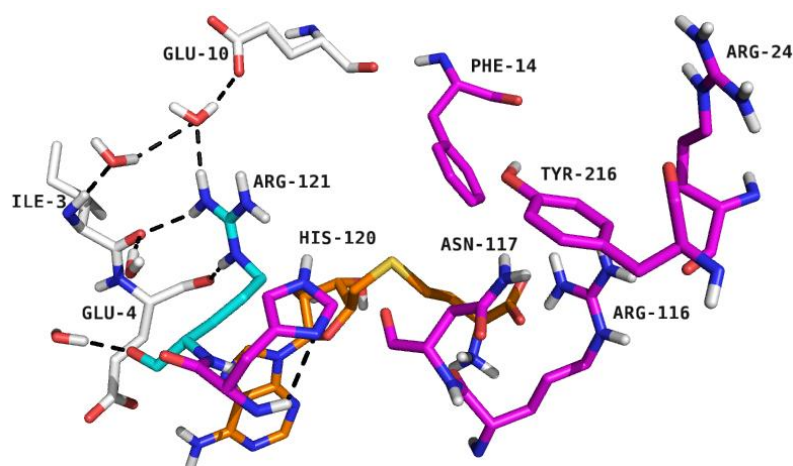
With purified samples of the lysine mutants in hand, the effect of each mutation on enzyme activity was investigated. Furthermore, the dependence of enzyme activity on the pH of the reaction mixture was of interest. Therefore, the percentage conversion of **67** to methylated product **69** with each mutant was recorded over pH 5-8 (**Figure 46**). As predicted, NovO H120K was inactive over this pH range. In the case of NovO R121K, however, a maximum of 19% conversion was observed at pH 7, with significant drop-off in activity being observed below pH 6.5, to just 3% at pH 5. Interestingly, this trend was mirrored by the WT enzyme and correlates to



**Figure 46.** Enzymatic activity of C-methylation of **67** by NovO H120K and R121K mutants relative to WT over pH values 5-8. *Reagents and conditions:* NovO mutant (19  $\mu$ M), 2 mM SAM, 0.1 mg/mL BSA, potassium phosphate buffer pH 5-8 containing 5 v/v% DMSO, 37  $^{\circ}$ C, 24 h.

the pKa of protonated histidine (6.0). These data suggests that a basic residue in its free base form is required for enzyme activity and supports the theory that the mechanism involves initial deprotonation of the 7-hydroxy group of the substrate by H120.

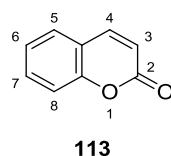
These data also showed that the R121K mutant was much less active than the WT, indicating that R121 is also important for enzyme function. Analysis of the environment surrounding R121 in the crystal structure revealed this residue to be involved in an extensive H-bonding network and was directly H-bonding with surrounding residues I3, E4 and E10 (**Figure 47**). In the case of R121K however, much of this H-bonding network is lost and therefore the enzyme is proposed to be less effective in stabilising the intermediate formed following deprotonation by histidine. Therefore, it was proposed that R121K is involved in stabilisation of the resulting phenoxide intermediate following deprotonation by H120.



**Figure 47.** H-bonding network surrounding R121.

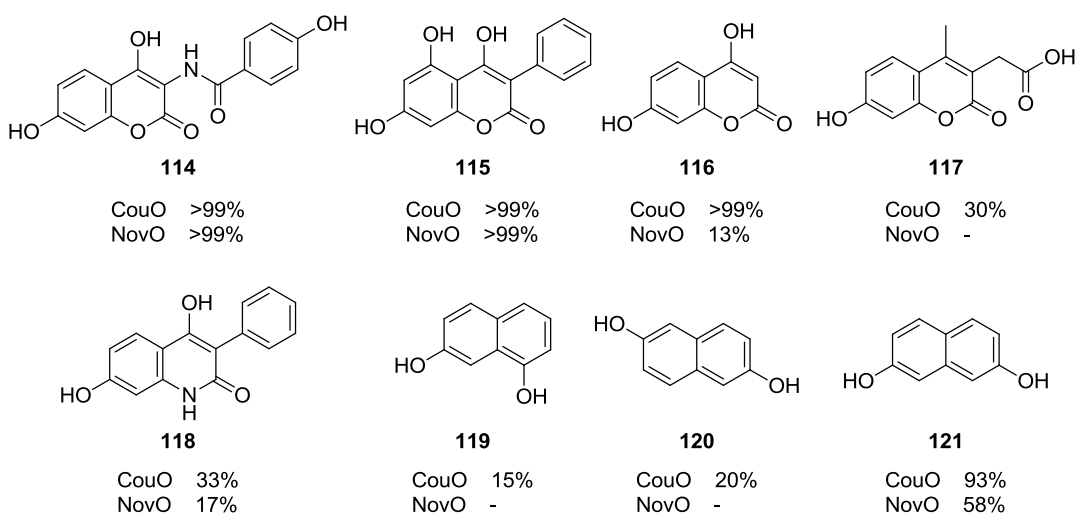
### 3.5 Substrate scope of NovO and CouO

An extensive substrate screen for the methylation of coumarin derivatives by NovO and CouO has previously been carried out by Stecher *et al.*<sup>[107,260]</sup> In particular, they focused on substitution at positions 3, 4, 5, 6 of the coumarin ring (**Figure 48**). Additionally, the oxygen at position 1 was substituted for carbon and nitrogen and monocyclic substrates were also screened.



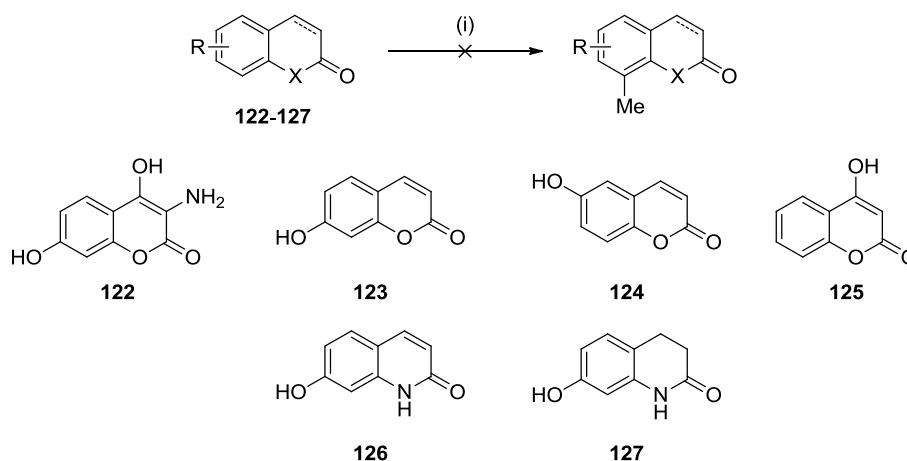
**Figure 48.** Numbering of coumarin (**113**) scaffold.

Out of a large library of compounds tested, only a small number of substrates (**114-121**) were methylated by NovO or CouO and provided an initial understanding of the substrate specificity of these enzymes (**Figure 49**).<sup>[107,260]</sup> In all cases, a 7-hydroxy group was required for methylation and, in general, CouO showed a wider substrate scope than NovO. Interestingly, it was found that the oxygen in position 1 could be replaced with a carbon, as demonstrated by the methylation of three dihydroxynaphthalene substrates (**119-121**).



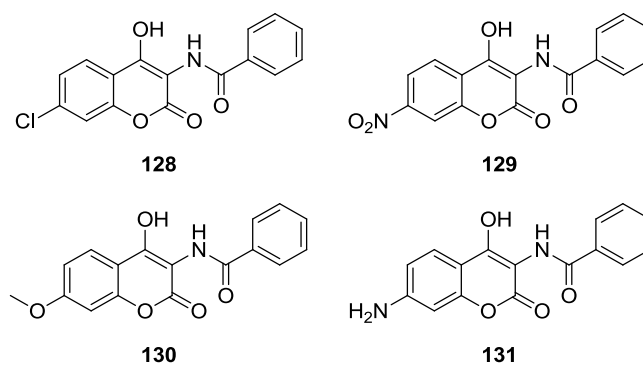
**Figure 49.** Substrates accepted by NovO and CouO in a substrate screen carried out by Stecher *et al.* with the conversion for each enzyme after 24 hours by area % by HPLC annotated.<sup>[107,260]</sup>

In this work, a further substrate screen was carried out to investigate substrates that have not been previously reported. As shown in **Scheme 34**, none of the substrates tested (**122-127**) were methylated by NovO or CouO, even when the substrate displayed a 7-hydroxy group. This further demonstrates the narrow substrate scope of NovO and CouO.



**Scheme 34.** Further substrate screening with NovO and CouO. *Reagents and conditions:* (i) *E. coli* cell lysate harbouring NovO or CouO, 2 mM SAM, bovine serum albumin (1 mg/mL), 1 mM DTT, 50 mM sodium phosphate buffer pH 6.5, 35 °C, 24 h.

Next, the requirement for a 7-hydroxy group was probed by varying the substituent at this position. The 7-chloro (**128**), 7-nitro (**129**), 7-methoxy (**130**) and 7-amino (**131**) analogues of substrate **67** were chosen to probe the effect of having an electron withdrawing (7-chloro and 7-nitro) or electron donating group (7-amino) at this position. Additionally, the 7-methoxy analogue **130** was of interest to investigate the need for an acidic proton at this position (**Figure 50**).

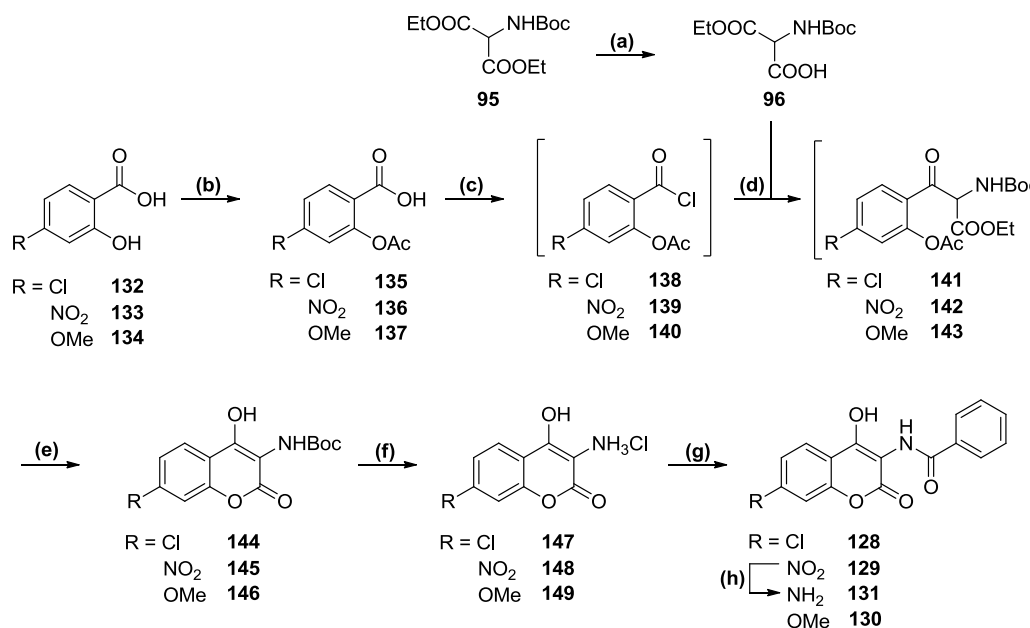


**Figure 50.** Substrate analogues **128-131** to probe 7-OH requirement of the substrate for methylation by NovO and CouO.

To this end, substrate analogues **128-131** were prepared in an analogous fashion to the preparation of substrates **67** and **68** (**Scheme 35**). In most cases, the syntheses proceeded smoothly under the same conditions as previously described in Chapter 2. An exception was the acetylation of chloro- and nitro- benzoic acid derivatives **132** and **133**, respectively, in which very low yields (<10%) were obtained using acetic anhydride, DMAP and



triethylamine. Therefore, acetyl chloride (AcCl) was used as the acylating agent in these cases to afford the acetylated benzoic acid derivatives **135** and **136**.

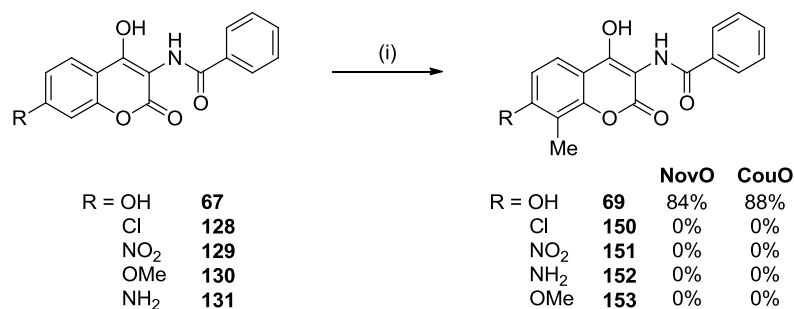


**Scheme 35.** Synthetic route to aminocoumarin substrate analogues **128-131**.<sup>[107]</sup> *Reagents and conditions:* (a) KOH, aq. EtOH, 22 °C, 12 h, 77%. (b) Ac<sub>2</sub>O, 4-dimethylaminopyridine (DMAP), NEt<sub>3</sub>, THF, 0-22 °C, 4 h (79%, **137**) or AcCl, NEt<sub>3</sub>, THF, 0-22 °C, 16 h (87% (**135**), 89% (**136**)). (c) (COCl)<sub>2</sub>, *N,N*-dimethylformamide (DMF), toluene, 40 °C, *products not isolated*. (d) MgCl<sub>2</sub>, NEt<sub>3</sub>, THF, 4 °C, then acid chloride, *products not isolated*. (e) aq. NaOH, MeOH, 22 °C, 4 h, 28% from **135** (**144**), 70% from **135** (**145**), 23% from **137** (**146**). (f) HCl in cyclopentylmethyl ether (CPME), *tert*-butylmethyl ether, MeOH, 22 °C, 24 h, 95% (**147**), 97% (**148**), 89% (**149**). (g) BzCl, NEt<sub>3</sub>, EtOAc, 22 °C, 3 h, then aq. NaOH, MeOH 22 °C, 12 h, 58% (**128**), 63% (**129**), 39% (**130**). (h) Fe, AcOH, aq. HCl, 40 °C, 4 h, 69% (**131**).

To synthesise the 7-amino substrate **131**, the nitro group of **129** was chemoselectively reduced using iron under acidic conditions.<sup>[261]</sup> Although clean and quantitative conversion to 7-amino substrate **131** was observed by HPLC and LC-MS, poor solubility of the product in a variety of organic solvents led to an isolated yield of 49%. As the synthesis provided enough material to support screening with NovO and CouO, however, further optimisation of the work-up conditions was not carried out.

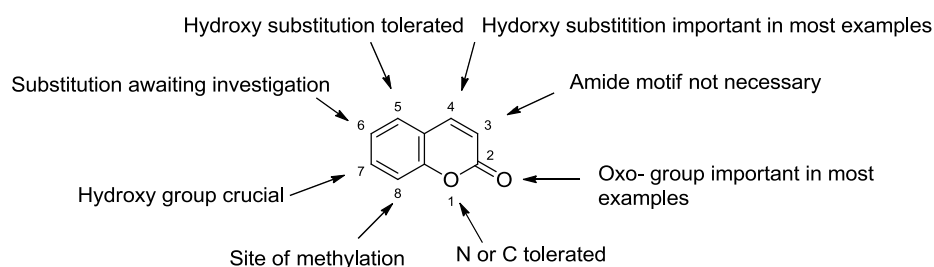
With a library of substrate analogues in hand, each was tested in the methyltransferase assay. The reaction mixtures were analysed by HPLC after 24 hours to determine the percentage conversion to methylated products **69** and **150-153**. As shown in **Scheme 36**, none of the substrate analogues were accepted by NovO or CouO. This indicated that the substrate must be electron rich and requires an acidic proton at the 7-position to be accepted by NovO. This

supports the hypothesis that there is an initial deprotonation of the substrate to initiate the catalytic mechanism of NovO.



**Scheme 36.** Methylation of aminocoumarin substrates **67** and **128-131** to products **69** and **150-153** by NovO WT. *Reagents and conditions:* (a) *E. coli* cell lysate harbouring NovO, 2 mM SAM, bovine serum albumin (1 mg/mL), 50 mM sodium phosphate buffer pH 6.5, 35 °C, 24 h.

Combining these results with those reported by Stecher *et al*, it is possible to draw some conclusions on substrate requirements for acceptance and methylation by NovO and CouO. As summarised in **Figure 51**, the most pressing requirement for substrate acceptance is a hydroxy group at the 7-position. Interestingly, this mirrors the observation that the substrates of other known small molecule aromatic C-MTs also have a hydroxy group adjacent to the site of methylation (Chapter 1).<sup>[48]</sup> Additionally, a carbonyl or hydroxy group at position 2 was found to be necessary in most examples.<sup>[107,260]</sup> Interestingly, it was shown that an amide group at the 3 position is not necessary, as exemplified in **Figure 49**. It was also shown that quinoline and naphthalene scaffolds are tolerated, with higher conversions observed with CouO than NovO in all examples. This understanding of limitations to the substrate scope of NovO and CouO will serve as a useful tool for the directed evolution of these enzymes in



**Figure 51.** Overview of substrate requirements by NovO and CouO.

future work and the dependence on the 7-hydroxy group supports a catalytic mechanism involving an initial deprotonation step.

### 3.6 Kinetic isotope effects of NovO

Kinetic isotope effect (KIE) data can provide valuable insights into the rate determining step (RDS) and transition state (TS) of reaction mechanisms.<sup>[262]</sup> As such, KIE data on NovO could provide additional information of the mechanism of methyl transfer by this enzyme, in particular to identify the RDS of the mechanism. The KIE is defined as the change in rate of a reaction when one of the atoms in the reactants is substituted with one of its isotopes and is formally calculated from the respective rate constants, according to **Equation 2**.<sup>[262]</sup>

$$KIE = \frac{k_L}{k_H}$$

**Equation 2.** Calculation of KIE, where  $k_L$  is the rate constant using the light isotope and  $k_H$  is the rate constant with the heavy isotope.

The observed difference in rate arises from the difference in zero-point energy (ZPE) between the reactants and the transition states of the isotopologues involved. This is a direct result of a quantum mechanical treatment of energy levels, in which all bonds have quantized vibrational energy levels. The vibrational energy levels ( $E_n$ ) are dependent on the frequency of the bond stretch ( $\nu$ ) according to **Equation 3**.

$$E_n = (n + 1)h\nu$$

**Equation 3.** Calculation of vibrational energy levels ( $E_n$ ), where  $h$  is Planck's constant.

The vibrational frequency of the bond stretch is dependent on the reduced mass of the two connected atoms ( $\mu$ ) according to **Equation 4**, such that the greater the reduced mass, the smaller the vibrational frequency.

$$\nu = \frac{1}{2\pi c} \sqrt{\frac{k}{\mu}}$$

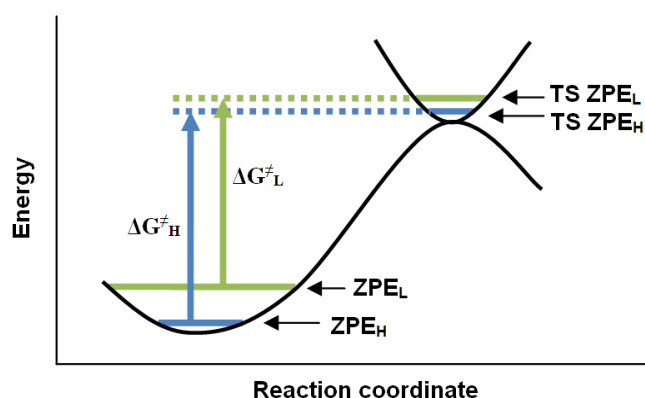
**Equation 4.** Calculation of bond stretch frequency ( $\nu$ ), where  $c$  is the speed of light and  $k$  is a constant.

Finally, the reduced mass is calculated from the two nuclei involved as shown in **Equation 5**. Thus, the heavier the nuclei involved in the bond stretch, the lower the  $E_n$ . As ZPE equates to  $n=0$ , the heavier the nuclei, the lower the ZPE and the more 'classically' the system behaves.

$$\mu = \frac{m_1 \cdot m_2}{m_1 + m_2}$$

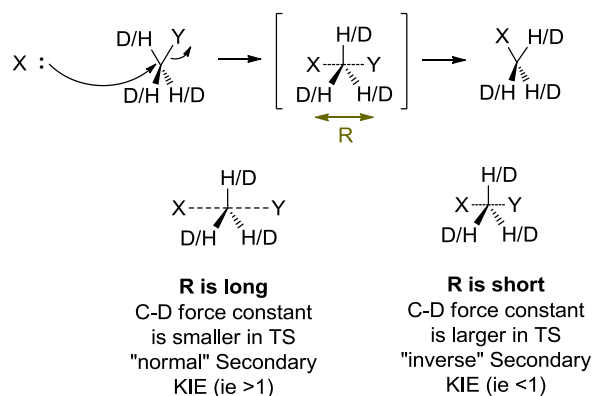
**Equation 5.** Calculation of reduced mass ( $\mu$ ) of the two atoms involved in the bond stretch, where  $m_1$  and  $m_2$  refer to the masses of the two atoms involved.

As the reaction proceeds and the TS is formed, the force constant between the breaking bond typically decreases and the ZPE difference between the isotopologues diminishes. This results in a larger activation energy (ie  $\Delta G^\ddagger_{\text{H}} > \Delta G^\ddagger_{\text{L}}$ , **Figure 52**) for the heavier isotope and therefore a slower rate of reaction if the bond breaking or bond forming step involving the isotopically labelled atom is the RDS of the reaction.<sup>[263]</sup> Therefore, primary KIEs are typically  $>1$  (termed ‘normal KIE’), although examples of inverse KIEs (KIE  $<1$ ) are also known.<sup>[263,264]</sup>



**Figure 52.** Difference in activation energy between isotopologues resulting from lower ZPE for heavy isotope.

In addition to primary KIEs in which the isotopically labelled atom is directly involved in a bond breaking or bond forming step, secondary KIEs can also be observed. This occurs if a difference in rate is observed when the isotopically labelled atom is not directly involved in the bond breaking or bond forming step and are typically much smaller in magnitude than primary KIEs and can also be normal or inverse.<sup>[264]</sup> For an  $S_{\text{N}}2$  reaction involving methyl transfer, smaller force constants for the C–H/D bonds in the transition state than in the reactant leads to an isotope effect  $>1$  (‘normal secondary KIE’). Conversely, force constants that are greater in the transition state than in the reactant result in an isotope effect  $<1$  (‘inverse secondary KIE’) (**Figure 53**). Secondary KIEs are most commonly observed when the isotope is one bond away from the bond breaking or bond forming atom ( $\alpha$ -KIE), although  $\beta$ -KIEs (two bonds away) have also been reported.<sup>[265]</sup>



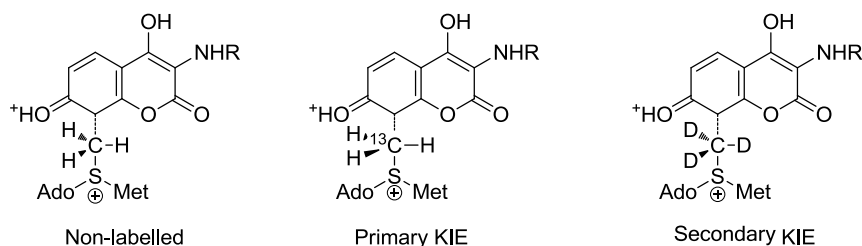
**Figure 53.** Effect of C-H/D force constant on secondary KIE.

Both primary and secondary KIEs have been used as tools to study methyl transfer reactions, in particular to study the nature of the transition state of a given reaction. In particular, the mechanism of the *O*-MT COMT has been the subject of debate since Hegazi and coworkers published their study on primary ( $^{13}\text{C}$ ) and  $\alpha$ -deuterium isotope effects in the methyl transfer reaction catalyzed by COMT. In this study, primary and secondary KIEs of  $1.09 (\pm 0.05)$  and  $0.83 (\pm 0.05)$ , respectively, were reported for the methylation of 3, 4-dihydroxyacetophenone by COMT.<sup>[266]</sup> From these data, the authors surmised the rate-determining step to be transfer of the methyl group with an  $\text{S}_{\text{N}}2$ -like transition state. Furthermore, they hypothesised that this near-maximum value for the primary KIE indicates symmetrical transition state and the low value for the secondary KIE indicates a tight transition state. This was reinforced by studying the secondary KIE for a closely related, uncatalysed  $\text{S}_{\text{N}}2$  reaction between methoxide and *S*-methylthiophenium, for which they measured a secondary KIE of  $0.975 (\pm 0.02)$ .<sup>[267]</sup> The authors concluded that the smaller KIE for the enzyme catalysed reaction was evidence of a tighter transition state, which was proposed to be due to a compression effect caused by the enzyme.<sup>[268,269]</sup>

Subsequent computational studies have both supported<sup>[270]</sup> and disputed these conclusions, the latter of which propose that similar primary KIEs will be found for early, symmetric and late transition states and that it is therefore not appropriate to use primary KIE data to estimate transition state structure.<sup>[271,272]</sup> Recently, Lameira and coworkers reported a study indicating that the catalytic efficiency of COMT is largely due to electrostatic preorganisation effects and that compression does not account for the catalytic effect.<sup>[273]</sup>

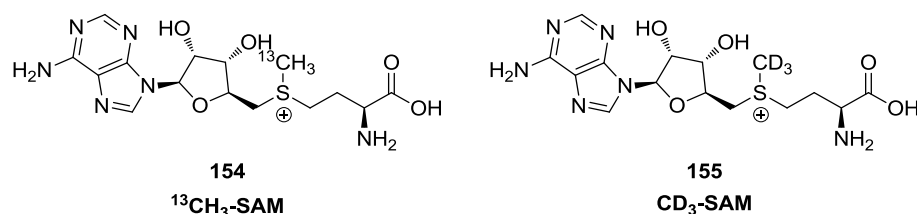
Despite this ongoing dispute regarding the interpretation of primary and secondary KIEs with respect to enzyme catalysed methyl transfer, KIE data was proposed to provide a

valuable insight into the catalytic mechanism of NovO, particularly in determining whether methyl transfer was the rate limiting step.



**Figure 54.** Proposed  $S_N2$  transition state during NovO catalysed methyl transfer reaction to substrate **67**.

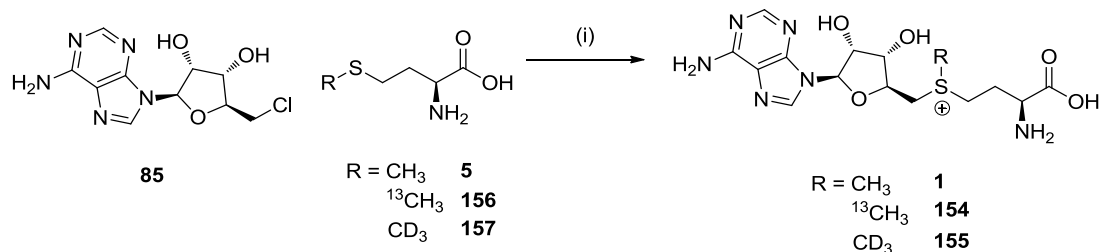
As discussed, MTs are generally accepted to operate *via* an  $S_N2$  mechanism (Chapter 1). Bearing in mind the requirement of an electron-rich substrate bearing a 7-hydroxy group, we hypothesised an  $S_N2$  type transition state between substrate **67** and SAM (**1**) or isotopically labelled SAM (**154** and **155**) with a planar centre at the methyl group being transferred (**Figure 54**) was hypothesised. To obtain KIE data, it was necessary to measure the  $k_{cat}$  of the methyl transfer reaction using labelled SAM analogues **154** and **155** using Michaelis-Menten kinetics.<sup>[274]</sup> In addition, the  $k_{cat}$  of the reaction using non-labelled SAM was measured in parallel to ensure that the KIE would be calculated between two experiments that had been carried out under the same conditions. As such, the isotopically labelled SAM analogues **154** and **155** (**Figure 55**) were required to determine the primary and secondary KIEs, respectively.



**Figure 55.** Structures of isotopically labelled SAM analogues  $^{13}\text{CH}_3$ -SAM (**154**) and  $\text{CD}_3$ -SAM (**155**).

To this end, the enzyme SalL was employed to synthesise **154** and **155** *in situ* prior to the addition of NovO (**Scheme 37**) in a one-pot, two step process.<sup>[224,275,276]</sup> A further advantage of preparing the cofactors enzymatically *in situ* was that only the naturally occurring, *S*-diastereomer of SAM and isotopically labelled SAM analogues was formed in each case.<sup>[144,251]</sup> The development of this methodology is described in detail in Chapter 5. In these experiments, a large excess of CIDA (**85**, 10 eq.) and Met (**5**) or isotopically labelled Met (**156** or **157**, 12.5 eq.) was used to ensure there was a large excess of SAM or SAM analogue

present in the reaction. In line with the Michaelis-Menten model for enzyme kinetics, this would ensure that substrate and enzyme concentration would be the only rate limiting factors in the experiment. The synthesis of SAM (**1**), **154** or **155** was confirmed by HPLC before carrying out kinetic measurements.

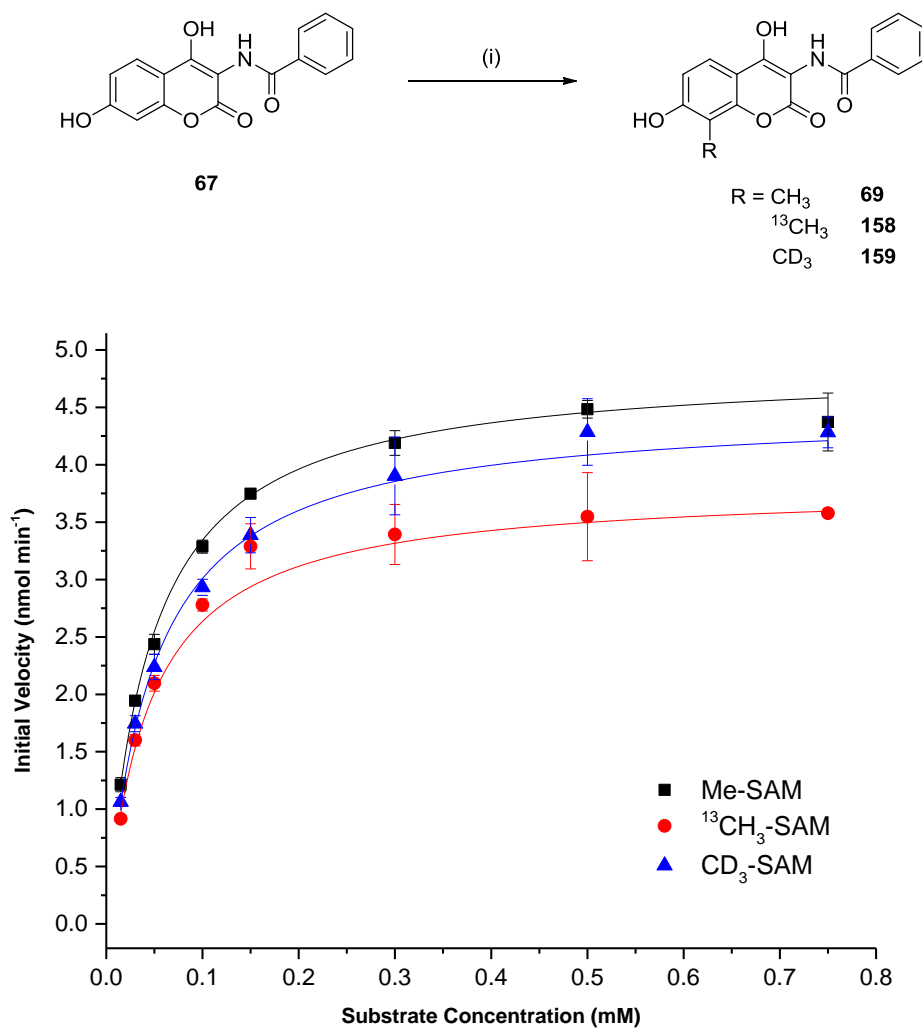


**Scheme 38.** Enzymatic synthesis of SAM (**1**),  $^{13}\text{CH}_3$ -SAM (**154**) and  $\text{CD}_3$ -SAM (**155**) using SalL. *Reagents and condition:* potassium phosphate buffer, pH 6.5, SalL (0.1 mg/ mL), 37 °C, 2 h, *products not isolated.*

Kinetics experiments were carried out in triplicate for each of the three cofactor analogues, using a range of substrate concentrations (15-750  $\mu\text{M}$ ) and were initiated by the addition of a known amount of NovO. Reactions samples were quenched after 3 minutes by heating and the percentage conversion to methylated product **69**, **158** or **159** in each case was analysed by HPLC. Using this data, the mean initial rate of reaction ( $\text{nmol min}^{-1}$ ) at each substrate concentration was calculated and plotted on a Michealis-Menten graph relative to the standard deviation of the dataset for each given substrate concentration (**Figure 56**). Next, a non-linear curve was fitted to each dataset using the Michaelis-Menten model available in Origin, which follows the formula shown in **Equation 6**. Adjusted R-square values very close to one (**Table 9**), which is a measure of fit of the curve to the data points, confirmed this model to be a good fit to the data points.

$$y = \frac{V_{max} * x}{K_m + x}$$

**Equation 6.** Michaelis-Menten rate equation, where y = initial velocity ( $V_0$ ) and x = substrate concentration [S].



**Figure 56.** Michaelis-Menten plot of NovO kinetics using Me-SAM (**1**), <sup>13</sup>CH<sub>3</sub>-SAM (**154**) and CD<sub>3</sub>-SAM (**155**). *Reagents and conditions:*(i) NovO (19 μM), SAM, <sup>13</sup>CH<sub>3</sub>-SAM or CD<sub>3</sub>-SAM (10 eq., pre-formed *in situ* [Scheme 38]), 50 mM sodium phosphate buffer pH 6.5, 35 °C, 3 min.

Using this model, the kinetic parameters  $K_m$  and  $V_{max}$  for the NovO catalysed methyl transfer reaction with each of Me-SAM, <sup>13</sup>CH<sub>3</sub>-SAM and CD<sub>3</sub>-SAM were obtained. Using these values,  $k_{cat}$  for each reaction was calculated according to **Equation 7**.

$$k_{cat} = \frac{V_{max}}{no. mol enzyme}$$

**Equation 7.** Calculation of  $k_{cat}$ .

Finally, the primary and secondary KIEs were calculated relative to the  $k_{cat}$  with Me-SAM



according to **Equation 2**. This provided primary and secondary KIE values of 1.22 and 1.03, respectively (**Table 9**).

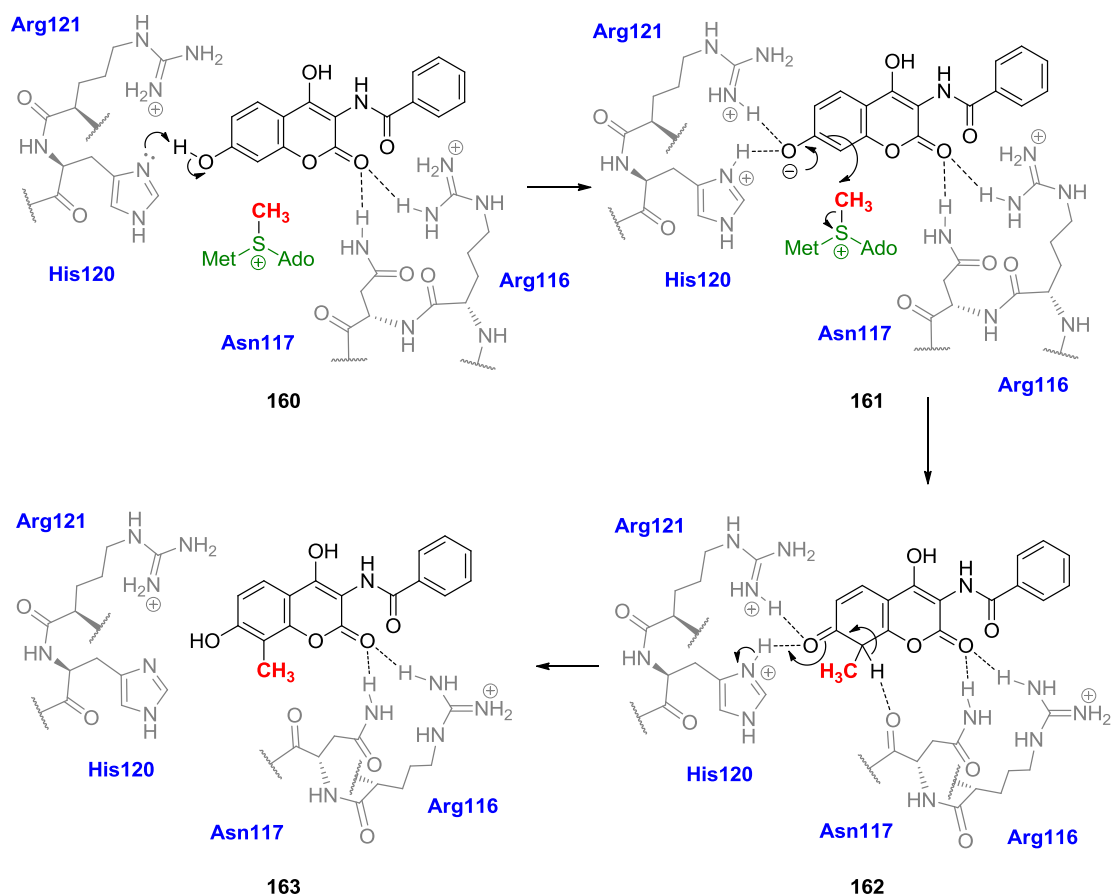
	$k_m$ ( $\mu\text{M}$ )	$V_{max}$ (nmol $\text{min}^{-1}$ )	Error (Adj. R-Square)	$k_{cat}$ ( $\text{s}^{-1}$ )	KIE
<b>Me-SAM</b>	43.7	4.6	0.998	$2.1 \times 10^{-3}$ ( $\pm 0.0004$ )	-
<b><math>^{13}\text{C}</math>-SAM</b>	41.6	3.8	0.997	$1.7 \times 10^{-3}$ ( $\pm 0.0008$ )	1.22 ( $\pm 0.04$ )
<b><math>\text{CD}_3</math>-SAM</b>	52.8	4.5	0.998	$2.0 \times 10^{-3}$ ( $\pm 0.0009$ )	1.03 ( $\pm 0.03$ )

**Table 9.** Kinetic parameters for the NovO catalysed methyl transfer reaction of substrate **67** using Me-SAM,  $^{13}\text{CH}_3$ -SAM and  $\text{CD}_3$ -SAM.

The value obtained for the primary KIE is notably higher than the primary KIEs for catechol COMT and human DNMT1 reported by Schowen and Schramm, respectively.<sup>[71,266]</sup> However, NovO represents a mechanistically distinct small molecule C-MT and at the time of writing, this is the only report of KIE data for this class of MT. This KIE is indicative of methyl transfer being the rate limiting step in the catalytic cycle, as has been reported for COMT and DNMT1. Furthermore, a large, normal KIE is characteristic of a symmetric transition state in the  $\text{S}_{\text{N}}2$  reaction. A smaller secondary KIE value of 1.03 was obtained, which was within the limits of the error margins and therefore was concluded not to be statistically significant.

### 3.6.1 Proposed catalytic mechanism of NovO

With the crystal structure, substrate binding model and kinetics data in hand, a catalytic mechanism for methyl transfer by NovO was proposed as follows. Firstly, an initial enzyme-substrate complex (**160**) is formed through H-bonding of the C-2 carbonyl of coumarin **67** with R116 and N117. Based on the pKa of histidine and the proximity of H120 to the phenolic proton ( $\sim 1.75$  Å based on the substrate docking model) and the study of mutants H120K and R121K, H120 is proposed to deprotonate the phenolic -OH in the 7-position, which activates the substrate for methyl transfer. The resulting phenoxide is stabilized by H-bonding to protonated H120 and R121 residues (complex **161**) and has sufficient nucleophilicity to effect electrophilic aromatic substitution at the 8-position with SAM as the methyl donor and SAH as the leaving group. Finally, rapid deprotonation of the resulting intermediate (**162**) to regain aromaticity of the system provides the enzyme-product complex **163** (**Scheme 39**).



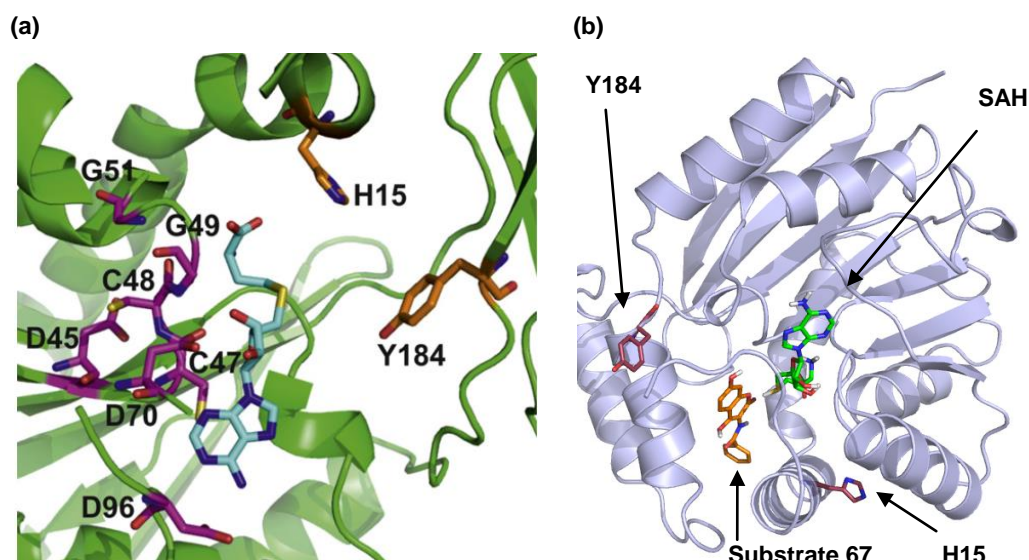
**Scheme 39.** Proposed catalytic mechanism for the methylation of substrate **67** by NovO.

The His-Arg motif has previously been proposed to be implicated in structurally and functionally unrelated proteins, such as in the catalytic mechanism of the periplasmic alginate epimerase AlgG from *Pseudomonas aeruginosa*.<sup>[277]</sup> In this study, site directed mutagenesis analysis and functional analysis suggests that the catalytic mechanism employs an active site His residue as the catalytic base and a protonated Arg to neutralise the resulting negative charge.<sup>[277]</sup> At the time of writing, this work demonstrates the first example of a His-Arg motif being implicated in the catalytic mechanism of a C-MT.

### 3.6.2 Comparison of proposed mechanism to previous studies

Prior to this work, Teng *et al.* have proposed two putative active site residues, H15 and Y184, based on a homology model of NovO from *Pyrococcus horikoshii* OT3 (PDB code 1WZN).<sup>[112]</sup> This model was supported by their finding that mutation of H15 to Asp or Glu reduced enzyme activity to 6% of that of the wild type. By placing a Lys or Arg residue at the position, the authors were able to restore activity to 25% and 22%, respectively. Based on these results, it was proposed that H15 is likely to be the general base in the methyl

transfer reaction.<sup>[112]</sup> However, investigation of the position of H15 and Y184 on the crystal structure of NovO (**Figure 57**) revealed both the residues to be >10 Å from the sulfur atom of SAH and therefore from the site of methyl transfer and, furthermore, both are pointing away from the active site. Based on these observations, it was concluded that these residues are not active site residues and that H15 is not directly involved in the catalytic mechanism. With this in mind, it was hypothesised that the drop in activity for H15D and H15N is more likely to be due to interference with protein folding.



**Figure 57.** (a) Homology model of NovO with bound SAH based on the SAM-dependent MT derived from *Pyrococcus horikoshii* OT3 (PDB code 1WZN) as proposed by Teng *et al.*<sup>[112]</sup> Model generated using Yasara software.<sup>10</sup> (b) Position of H15 and Y184 in NovO relative to the substrate and SAH according to the substrate binding model generated from our crystal structure of NovO.

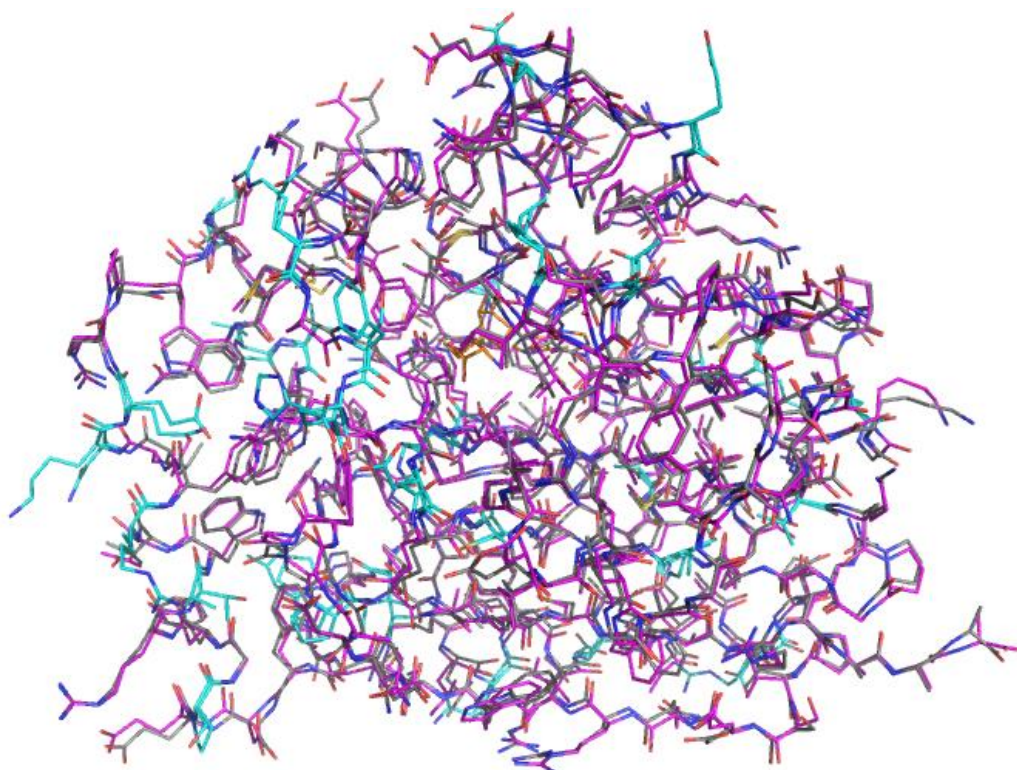
### 3.7 Homology modeling of CouO based on NovO

Having thoroughly investigated the active site of NovO and proposed a catalytic mechanism for methyl transfer, a homology model of CouO based on NovO was built. The two sequences share 85% sequence identity and excellent alignment (**Figure 58**) and are known to have very similar substrate specificity.<sup>[104,107]</sup>

NovO	1	MKIEAITGSEAEAFHRMGSQASHRYDEFVDLLVGAGIADGQTVVDLCCGSGELEVILSSR
CouO	1	MKIEPITGSEAEAFHRMGSRAFERYNDEFVDLLVGAGIADGQTVVDLCCGSGELEIILTSR
NovO	61	FPSLNLVGVDLSEDMVRIAREYAAEQKALEFRHGDAQLLAGMEDLAGKADLVVSRNAFH
CouO	61	FPSLNLVGVDLSEDMVRIARDYAAEQGRELEFRHGDAQSEFAGMEDLLGKADLVVSRHAFH
NovO	121	RLTRLPAAFDTMLRLAKPGGAVLNCSFIHPSDFDESGFRAWVTFLNQRPWDSEMQIVWAL
CouO	121	RLTRLPAGFDTMLRLVKPGGAILNVSFHLSDDFDEPGFRTWVRFLEKRPWDAEMQVAWAL
NovO	181	AHHYAPRLDDYREALAQAAARETPVSEQRVWIDDQGYGVPTVKCFARRAA
CouO	181	AHYAPRLQDYRDALAAADETPVSEQRVWIDDQGYGVATVKCFARRAA

**Figure 58.** Sequence alignment of NovO and CouO shaded according to sequence similarity.

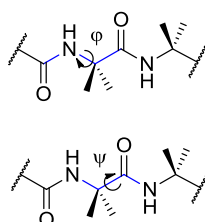
A homology model of CouO based on our crystal structure of NovO was built using the MOE Homology Model builder and is shown in **Figure 59** (CouO shown in magenta). For comparison with NovO, the structure was superposed onto the crystal structure of NovO (**Figure 59**, shown in grey) with the amino acids that differ between the two sequences highlighted in cyan. Analysis of the position of these sites revealed that the majority of different residues lie at the solvent interface of the structure. Out of the four residues



**Figure 59.** Homology model of CouO from NovO. CouO: magenta; NovO; grey; sites of residue variation: cyan.

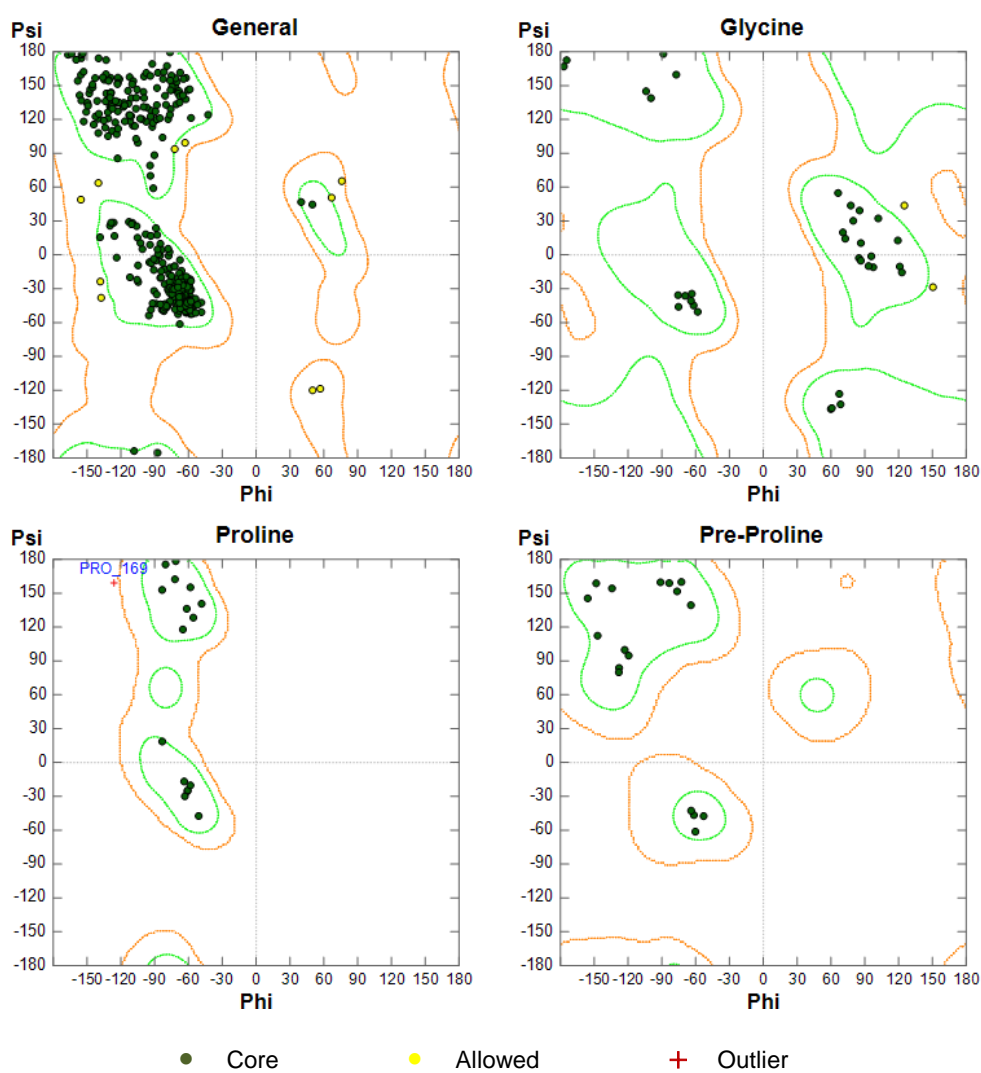
identified to be involved in intermolecular interactions between chains in NovO, three were conserved in CouO. The fourth, NovO N166 superposed with K166 in CouO, which could also be involved in H-bonding interactions.

A commonly used tool for homology model validation is analysis of the psi-phi ( $\psi$ - $\phi$ ) plot (Ramachandran plot) of a structure.<sup>[278]</sup> This is based on the theory that the backbone torsion angles ( $\psi$  and  $\phi$ , **Figure 60**) are restricted by steric hindrance of sidechain atoms.<sup>[279]</sup>



**Figure 60.** Psi ( $\psi$ ) and phi ( $\phi$ ) torsion angles on a polypeptide chain.

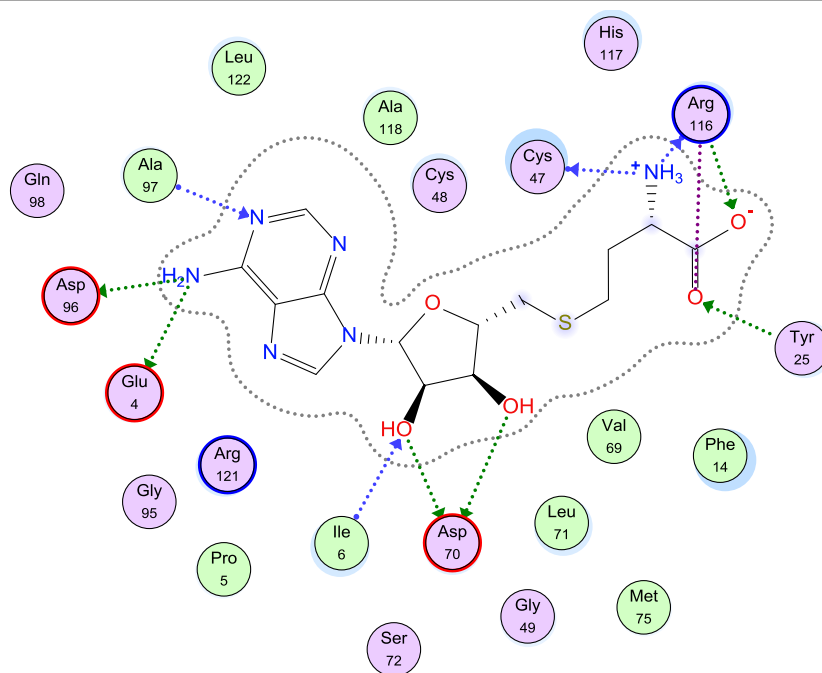
Depending on the secondary structure motif (left or right handed  $\alpha$ -helix and parallel or anti-parallel  $\beta$ -sheets), the 2D data points  $\psi$ - $\phi$  cluster in favourable regions and tend to be excluded from disallowed regions. As glycine has no sidechain and proline has restrictions arising from its 5-membered ring ( $\phi$  is restricted to  $\sim -60^\circ$  and  $\psi$  to  $-45^\circ$  and  $+135^\circ$  in the helical and sheet regions, respectively), different disallowed regions are observed.<sup>[280,281]</sup> Additionally, the amino acids preceding proline in the backbone chain also show a different distribution pattern on the plot.<sup>[281]</sup> As such, four different plots from the CouO homology model were generated to represent these four categories (**Figure 61**). The colour scheme corresponds to how closely a residue lies within a certain region for each map (core: green; allowed: yellow; outlier: red). Out of 230 residues, only one outlier (P169) was identified in the Proline plot, with the large majority of residues falling in the core region on all four plots. As P169 in NovO was found to have two conformations in the crystal structure of NovO (Chapter 2), this outlier in the CouO homology model was presumed to be carried through from the NovO crystal structure. These data indicate a high topological quality of the model, validating it as a tool to better understand the structure and function of CouO.



**Figure 61.** Ramachandran plot of CouO homology model based on the NovO crystal structure.

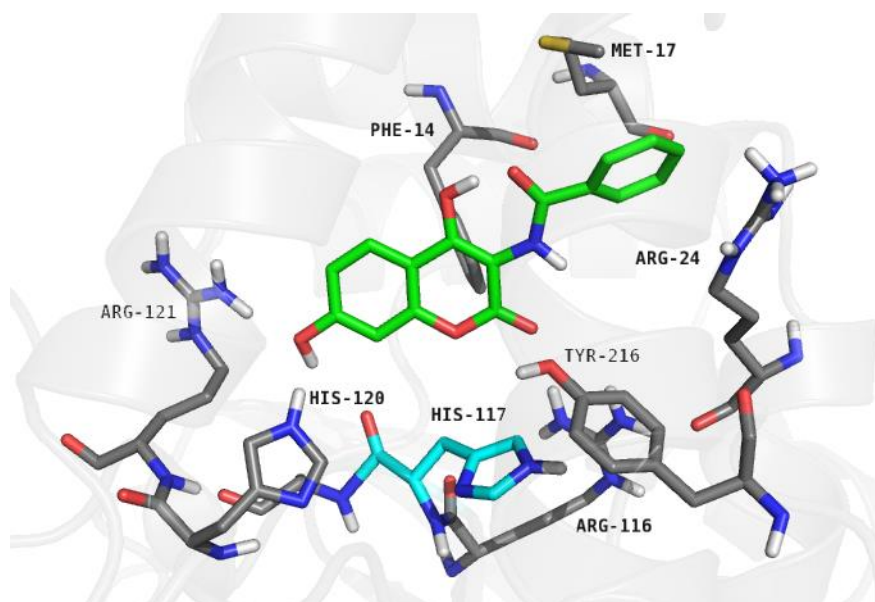
Having validated the homology model of CouO, the SAH binding pocket for this enzyme was analysed, which can be visualised by an interactions diagram (**Figure 62**). The residues involved in SAH binding in NovO were all conserved in CouO, in which they were involved in the same H-bonding interactions with the cofactor. In particular, D96 and E4 formed H-bonds with the adenine moiety, R116 with the amino acid portion and D70 with the *cis*-diol of the ribose ring.





**Figure 62.** Interactions map of residues involved in SAH binding in CouO homology model.

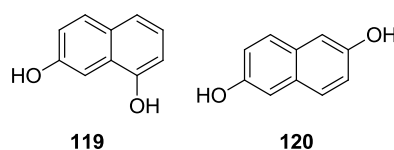
Next, the substrate binding site of CouO was defined. This was of particular interest, as Gruber and coworkers have reported a slight difference in substrate specificity between NovO and CouO. With this in mind, substrate **67** was modeled into the proposed active site of CouO, based on the results of our investigation into the NovO binding site and the high



**Figure 63.** Putative substrate binding site of CouO with H117 highlighted in cyan.

structural and sequence homology between the two proteins. The residues implicated in substrate binding were found to mirror the NovO substrate binding model, with the exception of the residue at position 117, which is an asparagine (N) in CouO and a histidine (H) in NovO (**Figure 63**, shown in cyan).

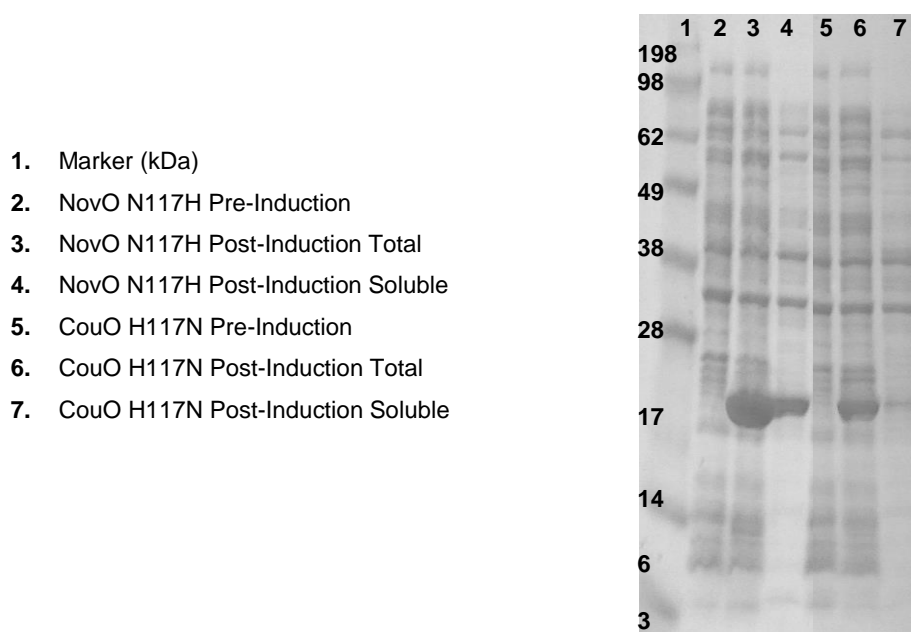
With this in mind, it was hypothesised that this residue could be responsible for the observed differences in substrate specificity between NovO and CouO which was reported by Gruber *et al.*<sup>[107]</sup> For example, dihydroxynaphthalene substrates **119** and **120** (**Figure 64**) were methylated in 15% and 21% conversion, respectively, under the reported conditions, whilst NovO did not methylate either substrate.<sup>[107]</sup>



**Figure 64.** Substrates accepted by CouO but not NovO, as reported by Gruber and coworkers.<sup>[107]</sup>

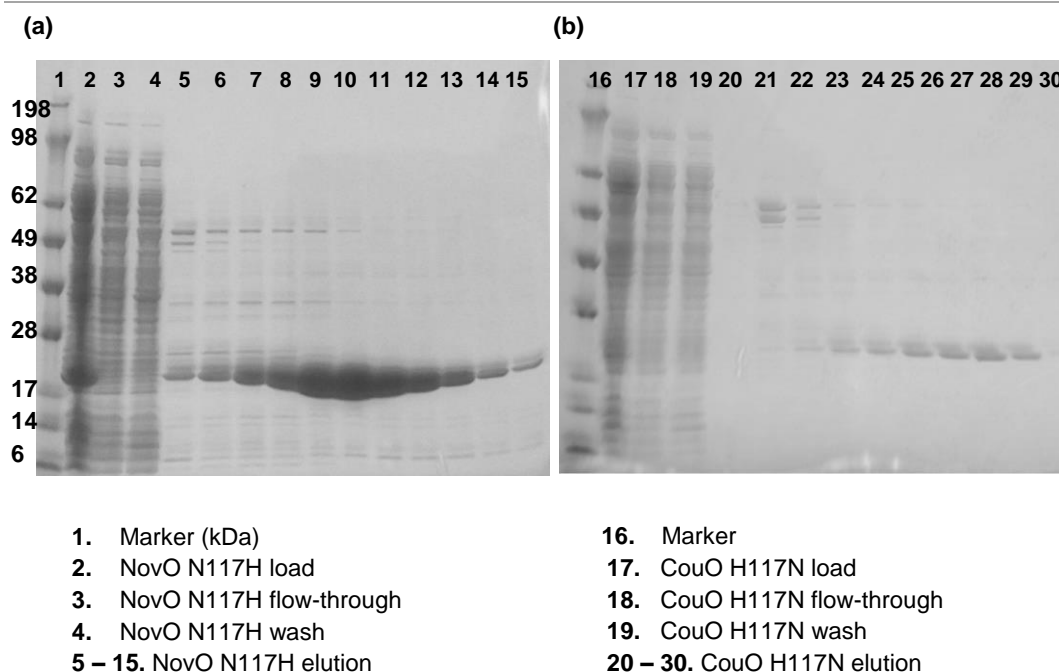
Based on these results, it was hypothesised that analysis of ‘swap-over’ mutants at this position could further probe the role of CouO H117 in accepting these substrates. Therefore, SDM was carried out on the WT DNA to prepare single point mutants NovO N117H and CouO H117N. The enzymes were overexpressed in Magic Media, using the conditions previously identified for the WT enzymes (Chapter 2). Analysis of the post-induction total and soluble fractions of both constructs showed both proteins to have been expressed. Whilst NovO N117H was also visible in high levels in the post-induction soluble fraction, CouO H117N was only visually detectable in trace amounts (**Figure 65**). Nonetheless, as only a small amount of protein was required for screening reactions, it was decided to carry these samples through to purification.





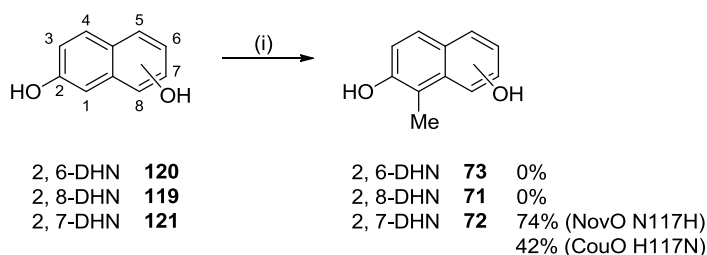
**Figure 65.** Analysis of heterologous expression of NovO N117H and CouO H117N. Enzymes expressed in *E. coli* BL21 (DE3) in Magic Media at 30 °C to OD ~2, then 18 °C overnight and analysed by SDS PAGE. Theoretical mass of WT 6xHis CouO: 26.5 kDa; WT 6xHis NovO: 26.2 kDa.

To aid solubilisation of CouO H117N, lysis was carried out in the previously identified lysis buffer containing 10 v/v% isopropanol, which was found to give good results for poorly soluble mutants of NovO (see Chapter 2, **Figure 44**). For comparison, NovO N117H was also purified under these same conditions. Following lysis in the isopropanol buffer, the proteins were purified by Ni-NTA chromatography using the same method as previously described for NovO WT, CouO WT and NovO mutants. This method provided a sufficient amount of protein for use in subsequent screening experiments (**Figure 66**).



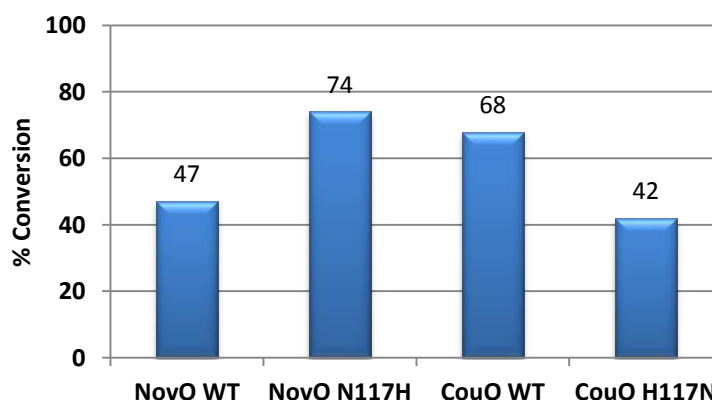
**Figure 66.** Analysis of purification of (a) NovO N117H and (b) CouO H117N by Ni-NTA affinity chromatography by SDS PAGE. Theoretical mass of WT 6xHis-NovO: 25.5 kDa; WT 6xHis-CouO: 25.6 kDa.

With purified samples of NovO N117H and CouO H117N in hand, both enzymes were investigated for their ability to catalyse the methylation of the three dihydroxynaphthalene (DHN) regioisomers **119-121**. The assays were carried out on 0.1 mM scale in the presence of a large excess of SAM (2 mM). The percentage conversion to methylated products **71-73** was analysed by HPLC after 24 h. Under these conditions, no conversion of 2, 6-DHN (**120**) or 2, 8-DHN (**71**) was observed, which was hypothesised to be due to a lower enzyme loading than that used by Gruber and coworkers, who used crude *E. coli* cell lysate for the transformations. However, under our assay conditions using the same concentration of purified enzyme for both reactions, 2, 7-DHN (**121**) was methylated by NovO N117H and CouO H117N in 74% and 42% conversion, respectively (**Scheme 40**).



**Scheme 40.** Investigation into substrate scope of NovO N117H and CouO H117N with three DHN regioisomers **119-121**. *Reagents and conditions:* NovO N117H or CouO H117N (19  $\mu$ M), SAM (2 mM), 0.1 mg/ mL BSA, potassium phosphate buffer pH 6.5 containing 5 v/v% DMSO, 37  $^{\circ}$ C, 24 h.

For comparison, the assay was also carried out using WT NovO and CouO. Interestingly, NovO WT showed a similar activity to CouO H117N (47% and 42%, respectively), whilst NovO N117H and CouO WT displayed similar activities (74% and 68%, respectively) (**Figure 67**). These results support the hypothesis that residue 117 in both NovO and CouO is involved in substrate binding and could be responsible for the observed difference in substrate scope.



**Figure 67.** Comparison of activity of NovO WT, NovO N117H, CouO WT and CouO H117N by methylation of dihydroxynaphthalene substrate **121**.

Taken collectively, the results of the investigation of the substrate binding sites of NovO and CouO could form a valuable platform for rational or semi-rational approaches to directed evolution towards a synthetically useful substrate scope. In particular, the strict requirement for a hydroxy group at the 7-position of coumarin substrates has been rationalised based on a proposed mechanism involving initial deprotonation of this hydroxy group by an active site histidine residue. This mechanistic insight enabled a better informed choice of substrates for a subsequent directed enzyme evolution project.

## **Chapter 4.**

# **Attempts towards expanding the substrate scope of NovO by directed evolution**

## 4 Attempts towards expanding the substrate scope of NovO by directed evolution

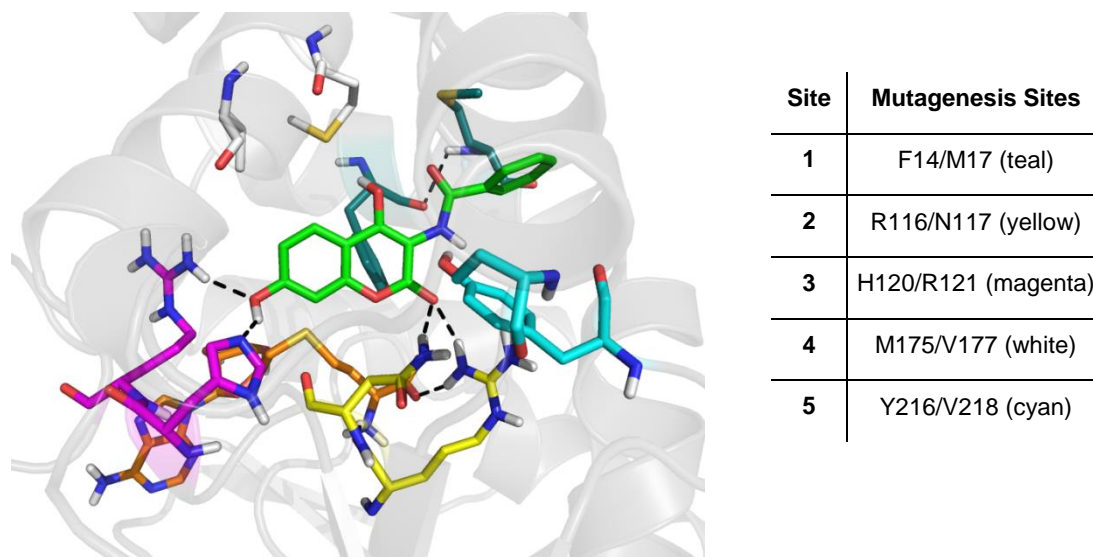
The use of the C-MTs CouO and NovO as biocatalysts for the Friedel-Crafts alkylation of aromatic substrates is currently limited by their narrow substrate scope. Directed evolution of biocatalysts is an effective strategy to alter or broaden the substrate specificity for a range of classes of enzymes. Therefore, the next stage in the project was to use the structural information obtained about NovO (Chapters 2 and 3) to direct an enzyme evolution project to expand the substrate scope of this MT. As such, this chapter describes the work towards the directed evolution of NovO towards a synthetically useful substrate scope to make C-MTs a more attractive and environmentally benign alternative to traditional Friedel-Crafts chemistry for the alkylation of aromatic compounds.

A number of different strategies for semi-rational evolution have been reported and there are advantages and disadvantages associated with each, which are summarised in a number of review articles and in Chapter 1.<sup>[139,152,153,282]</sup> In selecting which strategy to use for the directed evolution of NovO, an important consideration is the balance between library size (statistically, a larger library of random mutants has a higher probability of containing a mutant with a beneficial mutation) and screening effort. This was particularly pertinent in this study, as there is very limited literature precedent of high throughput screening methodologies for methyltransferases and most reported procedures are plate based assays, which are now considered to be low-to-medium throughput.<sup>[188,190,283,284]</sup> As such, the strategy presented herein focused on generating relatively small libraries. With this in mind, a Combinatorial Active-site Saturation Testing (CASTing) approach was chosen, which has been shown to be particularly successful for expanding the range of substrate acceptance of enzymes.<sup>[159]</sup> First introduced by Reetz and coworkers, this involves randomising defined pairs of residues which are close in 3D space in the active site of an enzyme and has been discussed in detail in Chapter 1.<sup>[6,7,9,159,285]</sup>

### 4.1 Preparation of CASTing libraries

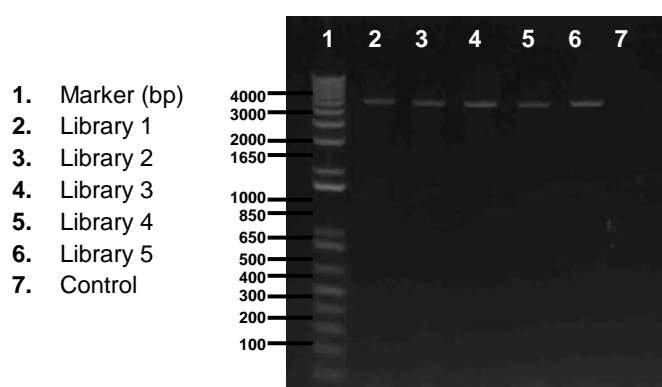
In order to identify suitable CASTing pairs, the substrate binding model of aminocoumarin **67** in the active site of the crystal structure of SelMet NovO (Chapter 3) was used to identify residues which formed the substrate binding pocket. Five pairs of residues were selected that

were proposed to interact with a different face or edge of the substrate. Each of these pairs would be randomly mutated to produce five CASTing libraries (**Figure 68**).



**Figure 68.** Binding model of aminocoumarin substrate **67** in the active site of NovO showing position of CASTing library pairs in the active site of NovO.

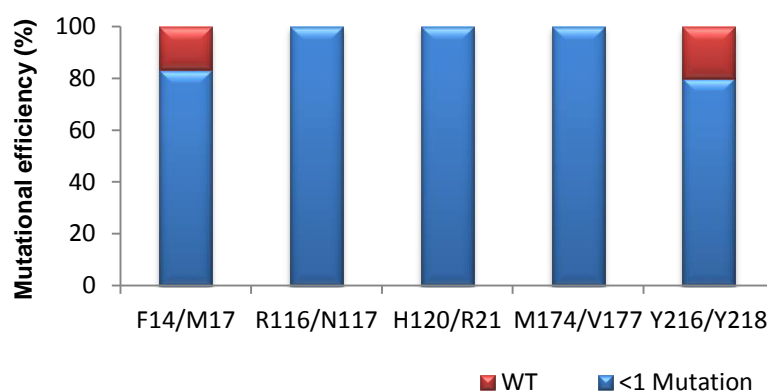
To generate the CASTing libraries, site-directed mutagenesis (SDM) was carried out with Q7 DNA polymerase, using primers with degenerate codons at the desired mutagenesis sites (**Figure 68**). When the PCR was complete, each reaction was incubated with Dpn1 to digest the template DNA. To confirm the Dpn1 digest had been successful, a control reaction was carried out to which primers had not been added. The PCR products were subsequently



**Figure 69.** E-gel of purified PCR products from SDM to generate CASTing libraries 1-5 with control reaction in which no primers were added. Theoretical size of pET26b(+)-NovO: 5945 base pairs (bp).

purified using a PCR product purification kit and the products visualised on an E-gel (**Figure 69**). A band at ~6000 bp was clearly visible for each of the five libraries (**Figure 69**, lanes 2-6), whilst no bands were visible in the control reaction (**Figure 69**, lane 7), indicating the PCR reaction and Dpn1 digestions had been successful.

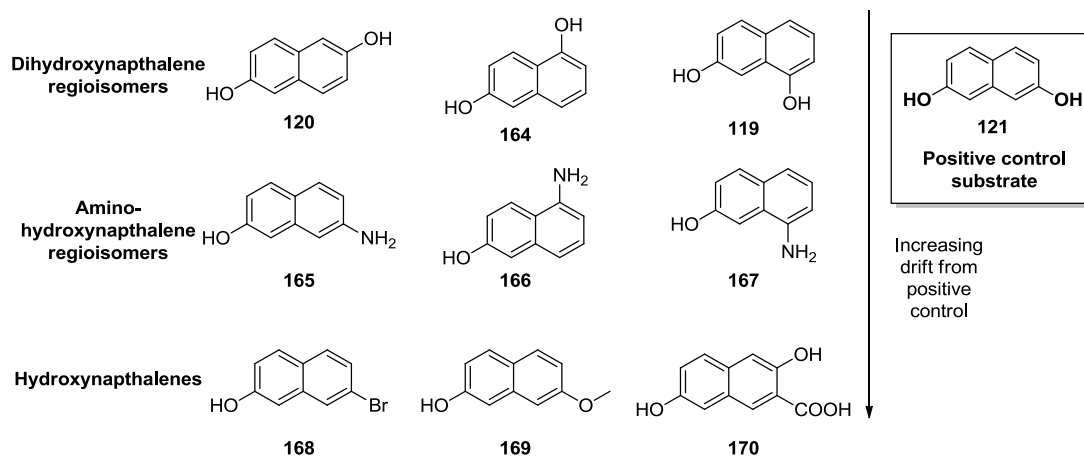
To determine the mutational efficiency of the SDM, a sample of the purified PCR products from each library was transformed into *E. coli* Top10. For each library, 12 colonies were picked and the plasmid DNA was isolated by MiniPrep and sequenced. As shown in **Figure 70**, all 5 libraries had over 80% mutational efficiency, with all 12 colonies picked from libraries 2-4 containing at least one mutation. These data indicated that the SDM introduced a high degree of sequence variation in the specified loci of the NovO gene.



**Figure 70.** Mutational efficiency of SDM based on sequencing of 12 colonies from each library.

Having prepared plasmid DNA for each of the 5 CASTing libraries, the next objective was to screen each CASTing library against a suite of substrates with which WT NovO had very low or no activity. Based on the substrate specificity of WT NovO, a set of nine substrates was proposed, which were further divided into three subcategories (**Figure 71**). The first subset was a series of dihydroxynaphthalene (DHN) regioisomers (**119**, **120** and **164**), each bearing a hydroxy group at the 2- position and the second hydroxy group on the other ring. As the 2, 7-DHN regioisomer (**121**) is a known substrate of NovO, it was hypothesised that its regioisomers (which show low levels of conversion with CouO) would be good candidates for directed evolution.<sup>[107]</sup> The second subset comprised three aminohydroxynaphthalene regioisomers (**165-167**). These were predicted to be more electron rich than the DHN substrates whilst maintaining a similar steric profile. Finally, the third subset (**168-170**) comprised hydroxynaphthalenes bearing additional functional motifs that could serve as handles for further modifying the molecule. For example,

bromohydroxynaphthalene **168** was of interest as it could be a substrate for a Suzuki cross coupling reaction following methylation.<sup>[124]</sup> Furthermore, in all cases, the regioselective methylation by a mutant of NovO could provide access to a product which would be difficult to access using traditional synthetic organic chemistry transformations.



**Figure 71.** Proposed substrate library for use in the directed evolution of NovO.

With a substrate library and CASTing DNA libraries in hand, the next objective was to establish a suitable medium-to-high throughput assay which would efficiently identify variants harboring beneficial mutations.

## 4.2 Development of a MT activity screening assay

Despite the potential of small molecule MTs as useful biocatalysts for late-stage, regioselective alkylation of small molecules, there is very little literature precedent for their directed evolution towards a wider substrate scope, higher  $k_{cat}$ , thermostability or organic solvent tolerance. One exception is a recent study by Micklefield and coworkers, who have recently reported rationally designed active site modifications and quaternary structure changes to alter the regioselectivity of COMT.<sup>[286]</sup> Due to the relatively small number of mutants generated in this study, the authors were able to use HPLC, which is generally considered a low-throughput technique, to determine the activity of each mutant.

However, for the evolution of NovO, a much higher throughput assay was desirable in order to screen a high proportion of each CASTing library against a library of substrates. As each library contains two randomised sites with fully degenerate codons (NNN), the number of possible mutants in each library is  $20^2=400$  variants. Therefore, it would be preferable to screen ~400 clones from each library to ensure the whole library is screened. Although the



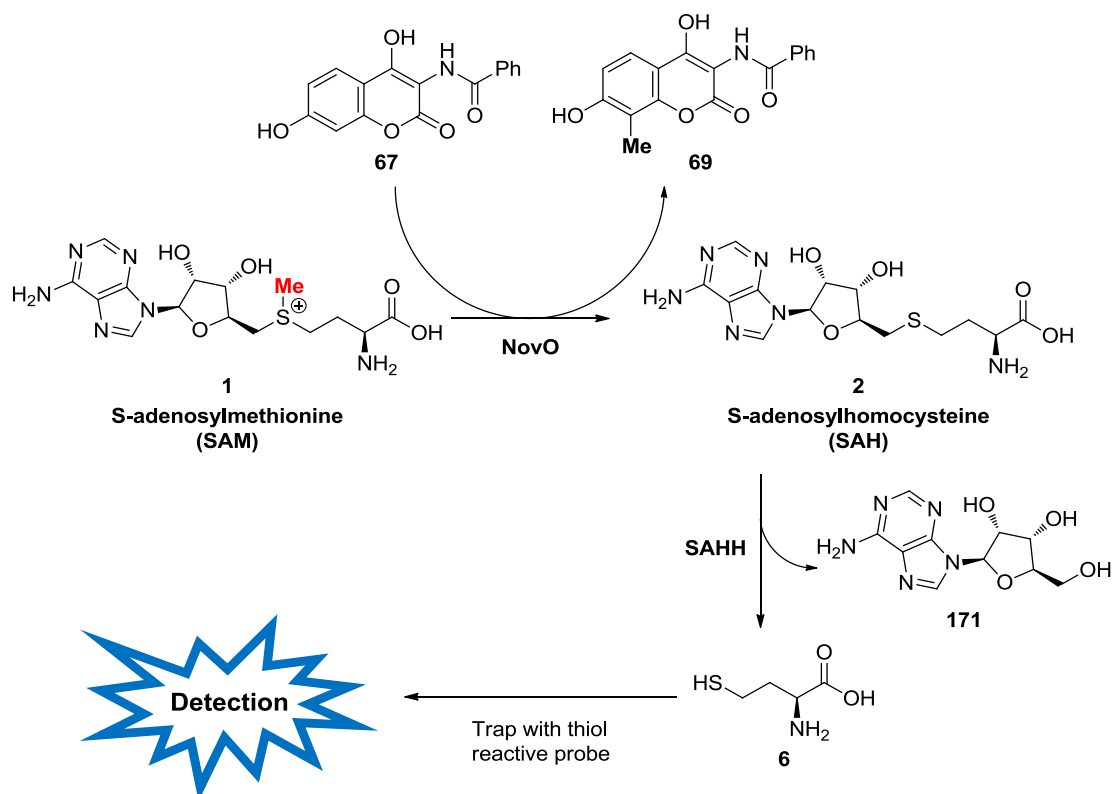
addition of a methyl group to a substrate can have a profound effect on its interaction with a target protein in drug discovery,<sup>[287]</sup> methylation usually does not alter the physical or electronic properties of the molecule enough such that it can be the basis for a sensitive, high throughput assay. As such, most examples of SAM dependent MT assays in the literature monitor methyl transfer indirectly *via* the detection of SAH, which is produced as a byproduct of the reaction. To date, three main approaches for the detection of SAH in the context of methyltransferase assays have been reported:

1. Direct detection of SAH using a riboswitch based fluorescent biosensor selective for SAH.<sup>[191]</sup>
2. Indirect detection of SAH *via* detection of hypoxanthine (MTAN and adenine deaminase coupled assay).<sup>[187]</sup>
3. Indirect detection of SAH *via* detection of homocysteine (SAHH or MTAN and LuxS coupled assay).<sup>[186,283,284]</sup>

The direct detection of SAH using a SAH-selective riboswitch is an attractive option due to the operational simplicity and very high sensitivity and selectivity of the method. However, this method was published in March 2016 and therefore there was not sufficient time during this project to develop this approach into a high throughput screening assay for evolution of NovO, although this will be the subject of future work within the research group. Therefore, an indirect detection method of SAH was investigated, for which there was much more literature precedent. Out of options **2** and **3** above, the detection of homocysteine was preferred as there was literature precedent for carrying out the assay in crude cell lysate using a colourimetric, commercially available thiol reactive probe.<sup>[284]</sup> This was particularly attractive as it would eliminate the need to purify each mutant, which is both costly and time consuming.

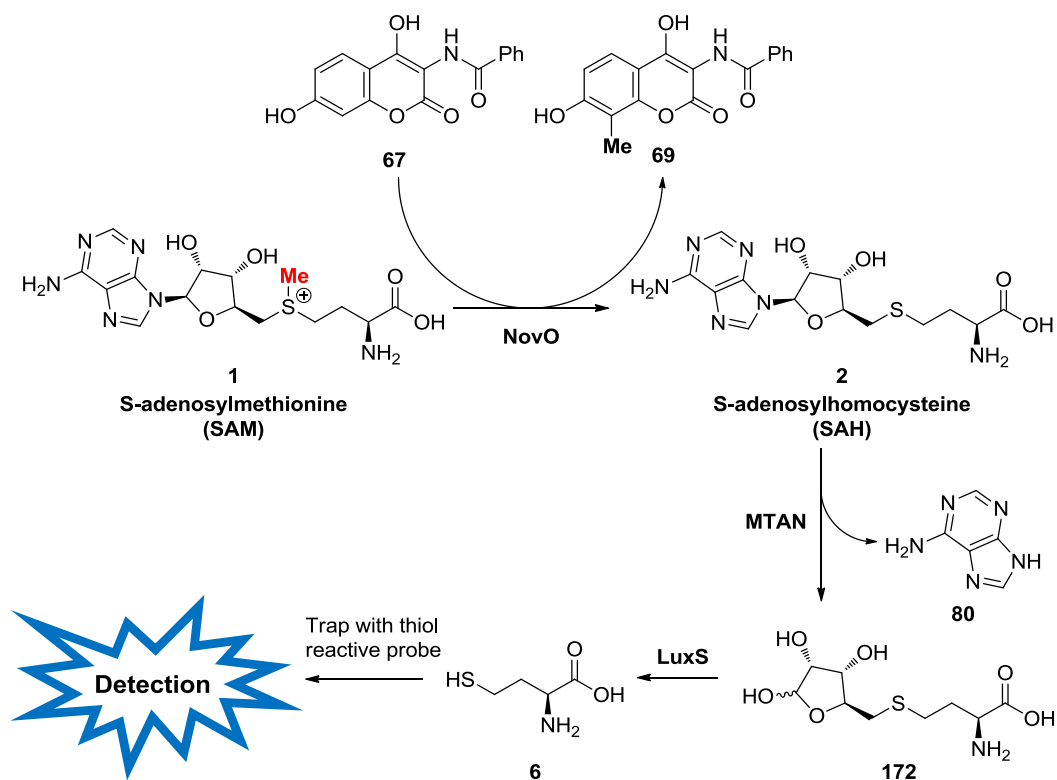
Two different enzyme coupled assays have been reported for the detection of homocysteine produced *via* the enzymatic degradation of SAH. The first employs SAH hydrolase (SAHH) to degrade SAH into adenosine (**171**) and homocysteine (**6**). The homocysteine can then be trapped by a suitable thiol reactive probe, resulting in a change in the absorbance, emission or fluorescence spectrum of system (**Scheme 41**). Due to the nucleophilic nature of the thiol group, the thiol reactive probe typically possesses an electrophilic centre which is susceptible to nucleophilic attack. Although the equilibrium for the decomposition of SAH by SAHH is known to favour the synthesis direction, it was hypothesised that the equilibrium could be

shifted in the hydrolysis direction by *in situ* removal of homocysteine by trapping with a thiol reactive probe.<sup>[283,288]</sup>



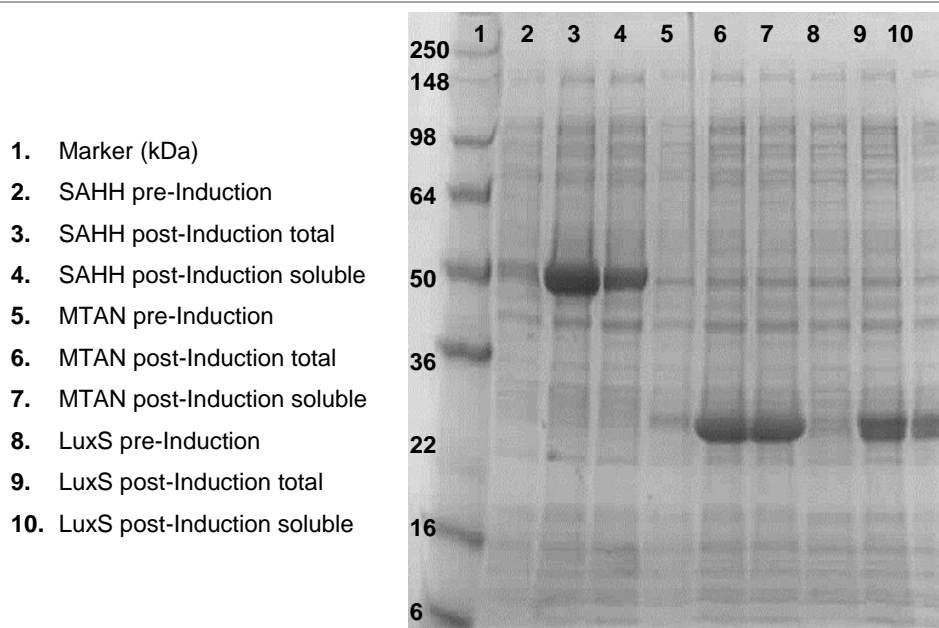
**Scheme 41.** SAHH hydrolase mediated methyltransferase activity assay.

The second approach involves coupling two enzymes to the MT reaction. Initially, SAH (2) is degraded *via* cleavage of the adenosine ring (80) using methylthioadenosine nucleosidase (MTAN). In contrast to the SAHH coupled assay, the degradation of SAH by MTAN is irreversible.<sup>[289]</sup> Next, the enzyme LuxS is employed to cleave the C-S bond of thioether 172 to liberate homocysteine (6), which is then trapped by a thiol reactive probe as described above (Scheme 42).



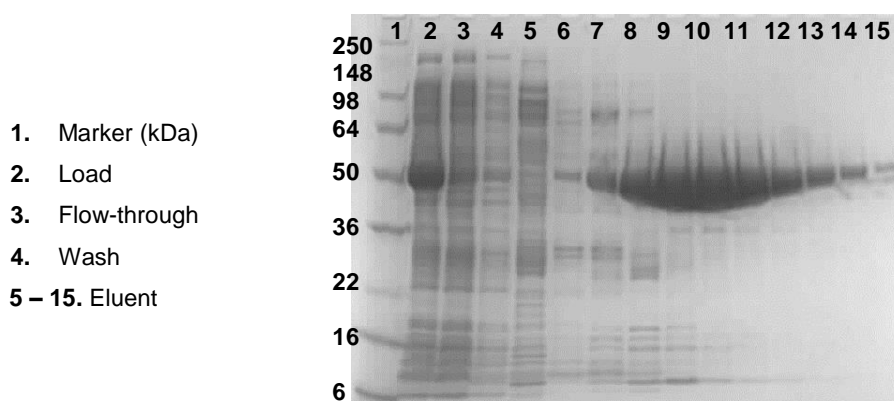
**Scheme 42.** MTAN and LuxS mediated methyltransferase activity assay.

In order to evaluate both of these strategies and identify which was better suited for use in this project, the overexpression of each of the three proteins SAHH, MTAN and LuxS was carried out in *E. coli*. The genes for SAHH (*homo sapiens*), MTAN (*E. coli*) and LuxS (*Bacillus subtilis*) were codon optimised for *E. coli*, synthesized and sub-cloned into a pET21b vector with an N-terminal 6xHis-tag and N-terminal Tev cleavage site at GenScript™. The constructs were transformed into *E. coli* BL21 (DE3) and the overexpression of the genes was carried out using Terrific Broth (TB), growing at 37 °C to an OD of ~0.6 and then inducing with IPTG and incubating at 15 °C overnight. Although the post-induction ODs were low (ODs ~2-3), analysis by SDS PAGE revealed high levels of expression of the target proteins. Furthermore, the soluble fraction contained high levels of the target protein (**Figure 72**), indicating that these conditions were effective for the expression of this protein. Therefore, despite the poor growth of the cultures, these overexpression experiments provided enough protein to support initial proof of concept investigations.



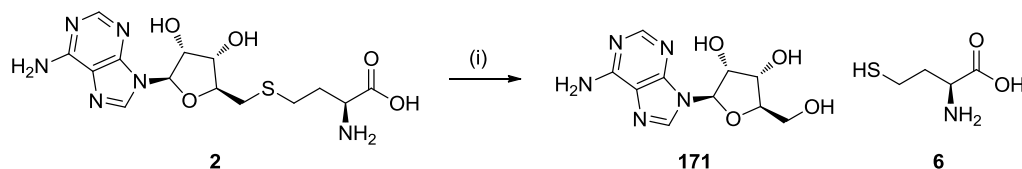
**Figure 72.** Analysis of heterologous expression of SAHH, MTAN and LuxS. Enzyme expression in *E. coli* BL21 (DE3) in TB at 37 °C to OD ~0.6 then induced with 500 mM IPTG and incubated at 15 °C overnight and analysed by SDS PAGE. Theoretical mass of 6xHis-Tev-SAHH: 49.4 kDa, 6xHis-Tev-MTAN: 30.0 kDa, 6xHis-Tev-LuxS: 19.4 kDa.

Having expressed SAHH, MTAN and LuxS, the enzymes were purified and their activity confirmed by investigating the activity in the degradation of SAH, both independently and in the context of the enzyme couple MT assay. First, SAHH was purified by Ni-NTA affinity chromatography (**Figure 73**). Although very low levels of contaminating proteins were present in the eluted fractions, it was judged to be pure enough to carry through into subsequent activity assay experiments.



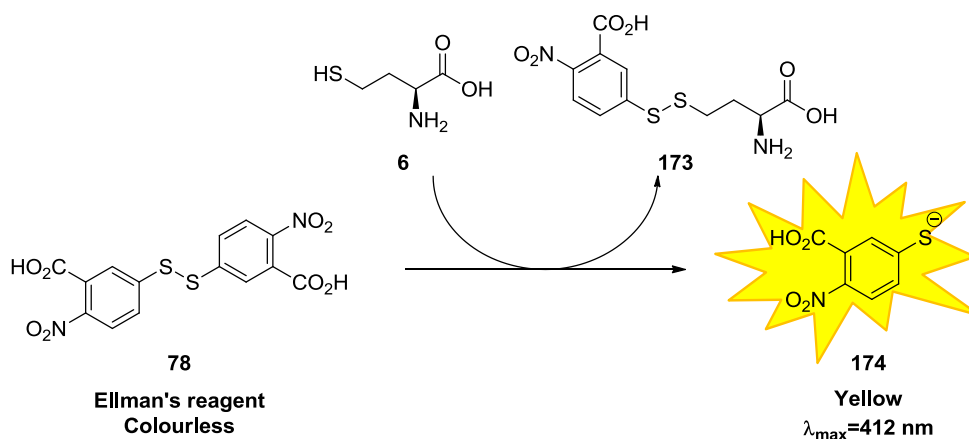
**Figure 73.** Analysis of purification of SAHH by Ni-NTA affinity chromatography by SDS PAGE. Theoretical mass of 6xHis-Tev-SAHH: 49.4 kDa.

To determine whether the SAHH was active, SAH (**2**) was incubated with three different concentrations of SAHH and the reactions analysed by LC-MS after one hour. In addition, a negative control reaction was carried out in which no SAHH was added. After this time, complete consumption of the SAH was observed for all three concentrations of SAHH, thereby confirming enzyme activity (**Scheme 43**).



**Scheme 43.** Activity assay for SAHH. *Reagents and conditions:* (i) SAHH (72, 36 or 14  $\mu\text{M}$ ), potassium phosphate buffer (50 mM, pH 7), 37  $^{\circ}\text{C}$ , 1 h, 100% conversion detected by HPLC and LC-MS.

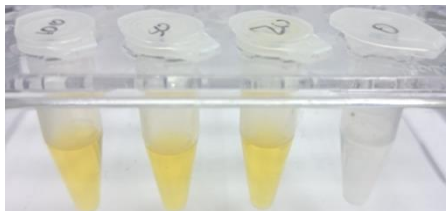
To confirm that homocysteine (**6**) was being produced in the reaction, 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB, also known as Ellman's reagent, **78**) was added to each of the four reactions. Ellman's reagent is an aromatic disulfide linked compound and is used extensively to quantify the concentration of thiol groups in a given sample.<sup>[189]</sup> As shown in **Scheme 44**, the thiol of homocysteine attacks the disulfide bond of Ellman's reagent, triggering a transthioylation reaction to provide disulfide **173** and the thiolate ion **174**, which has a strong absorbance maximum at 412 nm and a yellow colour which is clearly visible to the naked eye.



**Scheme 44.** Mechanism of thiol detection of HCys by Ellman's reagent (**78**).

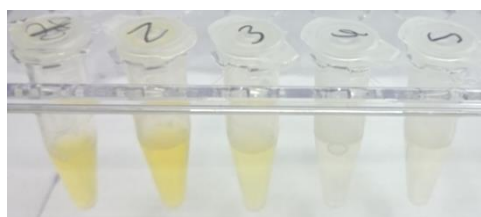
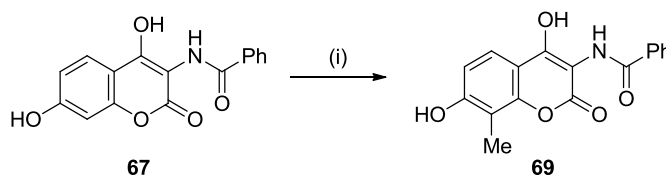
Ellman's reagent was added to each of the four SAHH activity assay reactions at an equimolar concentration to the original amount of SAH added at the start of the reaction

(500  $\mu\text{M}$ ). After 5 minutes, all three reactions to which SAHH had been added had turned yellow, whereas the negative control reaction remained colourless (**Figure 74**). This experiment confirmed both that HCys had been produced in the reaction and that Ellman's reagent was an effective method for its visual detection.



**Figure 74.** Visual detection of HCys produced from the degradation of SAH by SAHH using Ellman's reagent (78).

Having confirmed the SAHH was active for the production of HCys from SAH, the next stage was to determine whether it was active in the SAHH coupled MT assay. For this, and other proof of concept experiments, aminocoumarin substrate **67** was used as it was a good substrate for both NovO and CouO. A series of five reactions were carried out to test the assay with each of NovO and CouO (**Table 10**, entries **1** and **2**); to analyse whether there was any background reaction with SAHH (**Table 10**, entry **3**); and to determine whether the

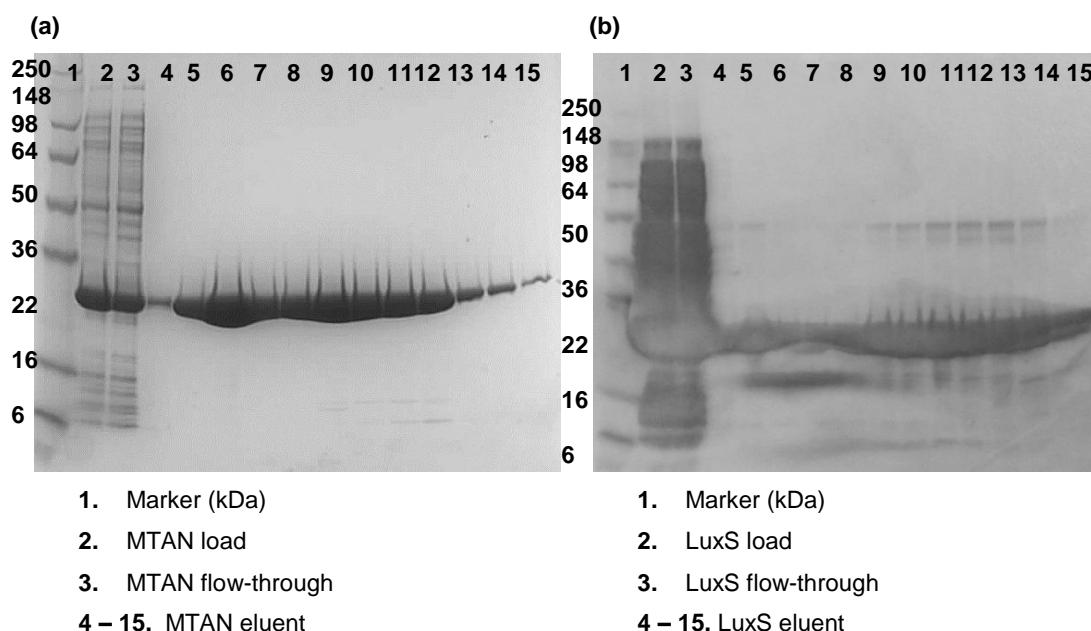


Entry	Enzymes	Ellman's Test	% Conversion
1	NovO/SAHH	+	22
2	CouO/SAHH	+	46
3	SAHH	+	-
4	NovO	-	19
5	CouO	-	42

**Table 10.** SAHH coupled MT assay using purified enzymes. *Reagents and conditions:*(i) SAM (2.5 eq.), NovO or CouO (0.5 mg/mL), SAHH (0.25 mg/mL), DTNB (2.5 eq.), 50 mM potassium phosphate buffer, pH 7, 37 °C, 24 h. % Conversion measured by area/area% by HPLC.

addition of SAHH to the MT reaction had an effect on the levels of conversion to methylated product **69** (Table 10, entries 4 and 5). In each case, the percentage conversion to methylated product **69** was analysed by HPLC after 24 hours. At the end of each reaction, Ellman's reagent was added as a visual test for HCys production. As shown in Table 10, low (22%) to moderate (46%) levels of conversion were observed under the assay conditions for NovO and CouO, respectively, and the HCys product was confirmed by a colour change from colourless to yellow upon addition of Ellman's reagent. Interestingly however, a faint yellow colour was also observed for the negative control reaction to which no MT had been added (Table 10, entry 3), indicating a background reaction of Ellman's reagent with SAHH. In contrast, neither of NovO or CouO appeared to show significant levels of background reaction with Ellman's reagent (Table 10, entries 4 and 5). Interestingly, these final two experiments also showed that the addition of SAHH does not have an effect on the levels of conversion to methylated product **69**, with 19% and 42% conversion being observed for NovO and CouO, respectively.

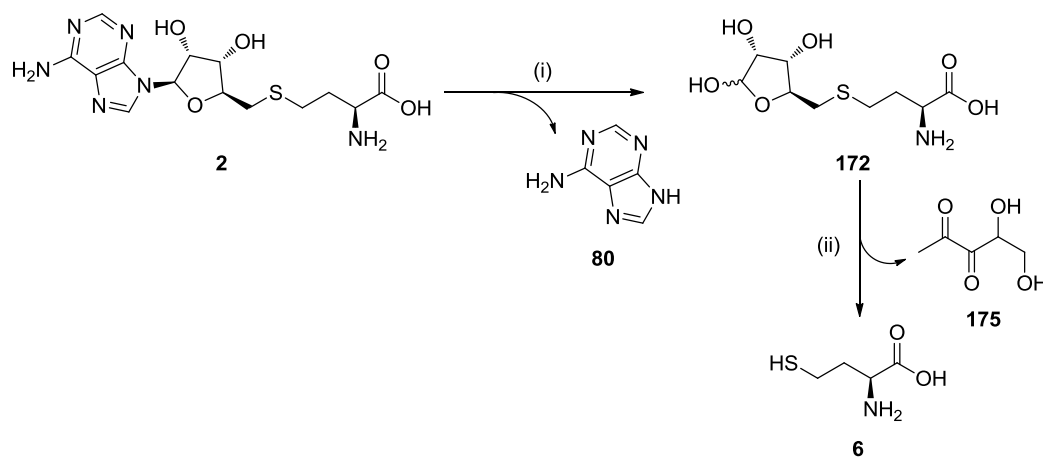
In order to draw a comparison between the SAHH coupled system and the MTAN/LuxS coupled system, the next stage was to purify and confirm activity of MTAN and LuxS in an analogous manner to that described above for SAHH. Therefore, MTAN and LuxS were purified by Ni-NTA affinity chromatography (Figure 75). Analysis by SDS PAGE revealed



**Figure 75.** Analysis of purification of (a) MTAN and (b) LuxS by Ni-NTA affinity chromatography by SDS PAGE. Theoretical mass of 6xHis-Tev-MTAN: 30.0 kDa, 6xHis-Tev-LuxS: 19.4 kDa.

this had been successful in both cases, although a protein band corresponding to the target protein in the flow-through fraction indicates overloading of the column in both cases. Nonetheless, these experiments provided purified samples of both MTAN and LuxS for further use in the enzyme coupled assay.

Initially, the activity of MTAN and LuxS was investigated in a stepwise fashion. SAH was incubated with three different concentrations (65, 32 and 16  $\mu\text{M}$ ) of MTAN for one hour and the reaction analysed by HPLC. Additionally, a negative control reaction was carried out in which no MTAN was added (**Scheme 45** (i)). In all three reactions to which MTAN had been added, complete consumption of SAH was observed in one hour, whilst no conversion was observed in the negative control reaction. Next, to confirm activity of LuxS, one of the MTAN reactions was aliquotted into three portions and either 62, 12 or 0  $\mu\text{M}$  LuxS added to each (**Scheme 45** (ii)). Additionally, a negative control reaction was set up which did not contain MTAN or LuxS.

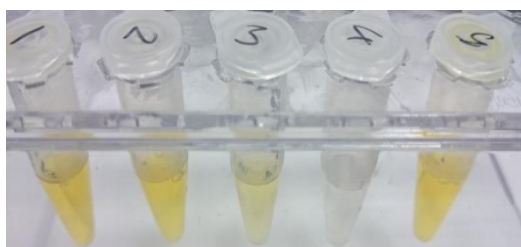


**Scheme 45.** Activity assay for MTAN and LuxS. *Reagents and conditions:* (i) MTAN (65, 32 or 16  $\mu\text{M}$ ), potassium phosphate buffer (50 mM, pH 7), 37  $^{\circ}\text{C}$ , 1 h, 100% conversion by HPLC. (ii) LuxS (62, 12 or 0  $\mu\text{M}$ ), potassium phosphate buffer pH 7, 37  $^{\circ}\text{C}$ , 1 h, HCys detected by Ellman's test.

Addition of Ellman's reagent to each reaction confirmed the production of homocysteine in the two reactions to which LuxS had been added (**Table 11**, entries **1** and **2**). Interestingly, a faint yellow colour was also observed for the reaction to which no LuxS had been added relative to the reaction which contained neither enzyme, indicating a background reaction between MTAN and Ellman's reagent (**Table 11**, entries **3** and **4**). Finally, a fifth reaction was also carried out in which SAH, MTAN and LuxS were combined, to confirm that both enzymes could operate in a tandem one-pot reaction to produce homocysteine (**Table 11**,



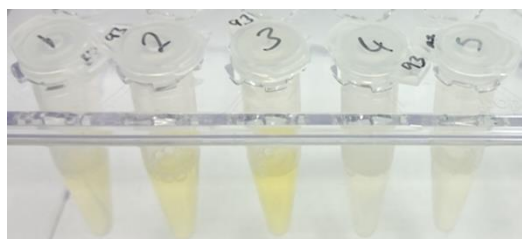
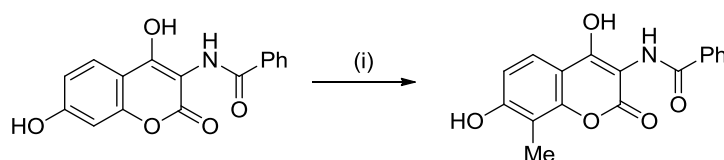
entry **5**). Gratifyingly, a positive readout using Ellman's reagent confirmed both enzymes to be active under these conditions.



Entry	Reaction
1	MTAN reaction + 62 $\mu$ M LuxS
2	MTAN reaction +12 $\mu$ M LuxS
3	MTAN reaction only
4	No LuxS or MTAN
5	LuxS + MTAN, SAH

**Table 11.** Activity assay for MTAN and LuxS visualised by the addition of Ellman's reagent.

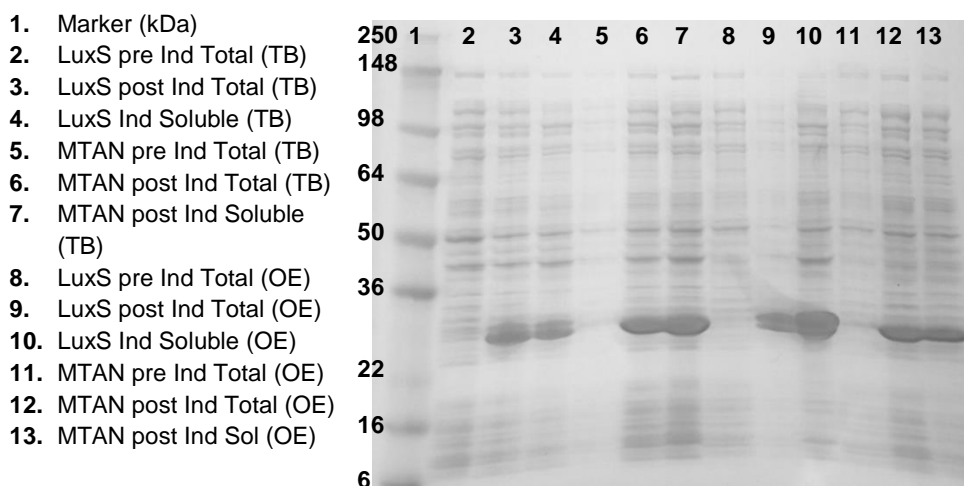
Having confirmed that the MTAN and LuxS were both active, the next stage was to test both enzymes in the MTAN/LuxS coupled methyltransferase assay with NovO and CouO and aminocoumarin substrate **67** in an analogous manner to that for SAHH as described above. To this end, five reactions were set up in parallel (**Table 12**). In the first two, the assay was run with either of NovO or CouO (**Table 12**, entries **1** and **2**), in which a positive result for thiols using Ellman's reagent was obtained, mirroring the 70% and 80% conversion of **67** to **69** using NovO and CouO, respectively. The assay was also run in the absence of a MT as a negative control (**Table 12**, entry **3**). However, despite no conversion to product **69**, a strong positive result for thiols was observed, indicating a background reaction with Ellman's reagent. Finally, the assay was also run in the absence of MTAN and LuxS to determine whether the addition of these enzymes had an effect on conversion levels of aminocoumarin **67** to methylated product **69** (**Table 12**, entries **4** and **5**). Interestingly, a drastic drop in percentage conversion was observed, with only 24% and 30% being observed for NovO and CouO, respectively. These data indicated that SAH is a potent inhibitor of both NovO and CouO, as irreversibly removing it from the system using MTAN drastically improved the activity of the MTs.



Entry	Enzymes	Ellman's Test	% Conversion
1	NovO/MTAN/ LuxS	+	70
2	CouO/MTAN/ LuxS	+	80
3	MTAN/LuxS	+	-
4	NovO	-	24
5	CouO	-	30

**Table 12.** MTAN and LuxS coupled MT assay using purified enzymes. *Reagents and conditions:* (i) SAM (2.5 eq.), NovO or CouO (0.5 mg/mL), SAHH (0.25 mg/mL), DTNB (2.5 eq.), 50 mM potassium phosphate buffer, pH 7, 37 °C, 24 h. % Conversion measured by area/area% by HPLC.

In comparison to the SAHH coupled assay, the MTAN/LuxS coupled assay gave much higher levels of conversion to methylated product **69**. This was presumed to be due to MTAN being much more efficient at removing SAH from the system than SAHH, thereby reducing byproduct inhibition of NovO or CouO caused by SAH. Based on these results, it was decided to take the MTAN/LuxS coupled assay through into further development. As such, expression conditions for MTAN and LuxS providing higher yields of MTAN and LuxS than previously obtained (**Figure 72**) were required. To this end, the expression of both enzymes was trialed in both TB and OE, this time using a lower pre-induction temperature of 30 °C. Analysis by SDS PAGE revealed in high levels of expression in each case (**Figure 76**), with slightly higher levels of soluble expression of LuxS in OE and or MTAN in TB.



**Figure 76.** Analysis of heterologous expression of MTAN and LuxS in TB and OE. Enzyme expression in *E. coli* BL21 (DE3) in TB at 30 °C to OD ~0.6 then induced with 500 mM IPTG and incubated at 15 °C overnight (columns 2-7) or in OE at 30 °C to OD ~2 then at 18 °C overnight (columns 8-13) and analysed by SDS PAGE. Theoretical mass of 6xHis-Tev-MTAN: 30.0 kDa, 6xHis-Tev-LuxS: 19.4 kDa.

The post induction ODs were higher in OE than TB for both constructs, however, with a significant increase from 5.4 to 15.8 being observed for LuxS (**Table 13**).

Entry	Construct	Expression Media	Post Induction OD
1	LuxS	TB	5.4
2	MTAN	TB	5.4
3	LuxS	OE	15.8
4	MTAN	OE	5.9

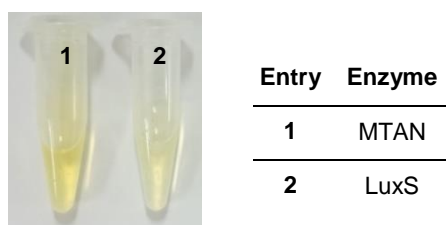
**Table 14.** Post induction ODs of cultures harbouring LuxS and MTAN expressed in TB and OE on 200 mL scale.

Based on these data, it was decided to use OE for the expression of both constructs, which was scaled up to 1 L scale and carried out in duplicate for MTAN and LuxS. In all cases, very high soluble expression was observed. Moreover, higher post induction ODs were measured, demonstrating the growth dependence on the scale of the culture (**Table 15**).

Entry	Enzyme	Post Induction OD
1	MTAN a	18.7
2	MTAN b	17.3
3	LuxS a	13.8
4	LuxS b	16.8

**Table 15.** Post induction ODs of cultures harbouring LuxS and MTAN expressed in OE on 1 L scale.

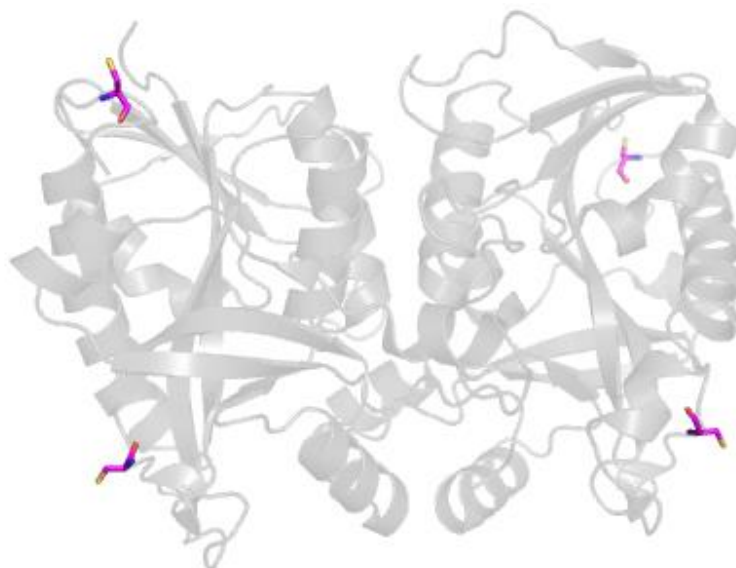
To further probe the cause of the background reaction observed in the negative control (**Table 12**, entry 3), each of MTAN and LuxS was separately incubated with Ellman's reagent and the colour change recorded after 1 hour (**Table 16**). A strong positive result was observed for MTAN, with a fainter yellow colour being observed for LuxS.



**Table 16.** Investigation into background reaction of Ellman's reagent with MTAN and LuxS.

To understand this result, the crystal structure of MTAN was analysed (PDB accession code: 3BSF).<sup>[290]</sup> Solvent accessible Cys residues were of particular interest, as it was hypothesised that this would be the most likely cause of background reaction with Ellman's reagent. Interrogation of the crystal structure revealed two such residues on each monomer chain (C113 and C250). As MTAN exists as a dimer in solution, this is equivalent to four solvent accessible Cys residues per unit of MTAN (**Figure 77**, highlighted in magenta).

With this in mind, it was hypothesised that identifying a thiol reactive probe that was selective for HCys over Cys may reduce the risk of identifying false positives caused by surface accessible Cys residues both on MTAN and in the *E. coli* cell free extract from the expression of NovO CASTing libraries.



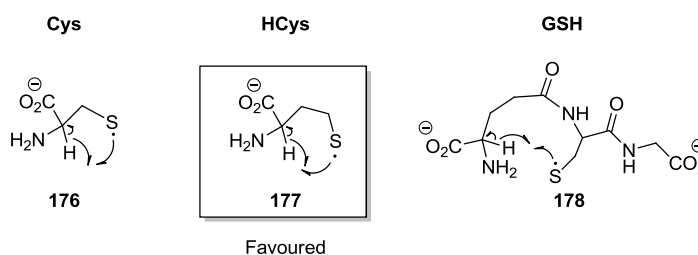
**Figure 77.** Position of the solvent accessible Cys residues C113 and C250 in MTAN. PDB accession code: 3BSF.<sup>[290]</sup>

### 4.3 Synthesis and evaluation of HCys selective probes

Thiols such as Cys, HCys, glutathione (GSH) and hydrogen sulfide are ubiquitous throughout Nature. Moreover, fluctuations in the levels of these compounds *in vivo* have been implicated in a number of diseases. For example, variations in HCys concentration can be indicative or causative of certain cardiovascular and neurodegenerative diseases, such as coronary artery disease and Alzheimer's disease.<sup>[24,291]</sup> As such, the identification of efficient methods for the detection of thiols is an active area of research and the subject of a series of review articles.<sup>[292–295]</sup> Thiol reactive probes have been developed based on a range of different mechanisms, including nucleophilic substitution, Michael addition, cyclisation, cleavage of a disulfide bond, metal complexation and redox chemistry.<sup>[189,292,296–299]</sup> Due to the structural similarity between Cys and HCys, which differ only by an extra methylene unit in the case of HCys, a particular challenge has been faced when developing a probe which is selective for one over the other. For the purposes of identifying a suitable probe for the MTAN/LuxS coupled MT assay, a probe which was selective for HCys and was compatible with the previously described assay system in *E. coli* cell free extract was sought after for this project.

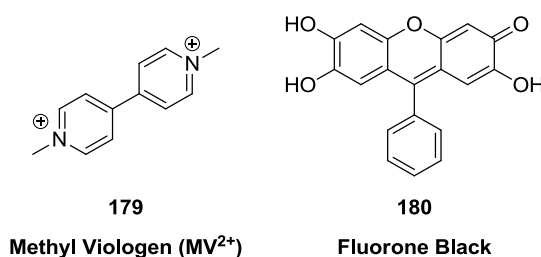
One approach to the development of HCys selective probes is to take advantage of the different rates of formation of  $\alpha$ -amino carbon-centered radicals from thiyl radicals. A number of different mechanisms for the formation of thiyl radicals from the thiol groups of

Cys, HCys and GSH are known, such as single electron oxidation of the thiolate ion, homolytic cleavage of a disulfide bond and hydrogen atom abstraction.<sup>[300]</sup> Once formed, the thiyl radical can rearrange to the  $\alpha$ -amino carbon-centered radical, which is stabilised by the geminal amine and carboxylate groups. This intramolecular H atom transfer reaction is favoured for the HCys radical **177** compared to the Cys and GSH radicals **176** and **178** due to the formation of a 5 membered ring transition state (TS) compared to a 4- and 9-membered ring TS for Cys and GSH, respectively (**Figure 78**). Consequently, HCys selective probes have been developed that operate by single electron oxidation of the probe by HCys, resulting in a change in the probe's absorption or emission spectrum.



**Figure 78.** Formation of  $\alpha$ -amino carbon-centered radicals from thiyl radicals *via* intramolecular H atom abstraction, which is favoured for HCys over Cys and GSH.

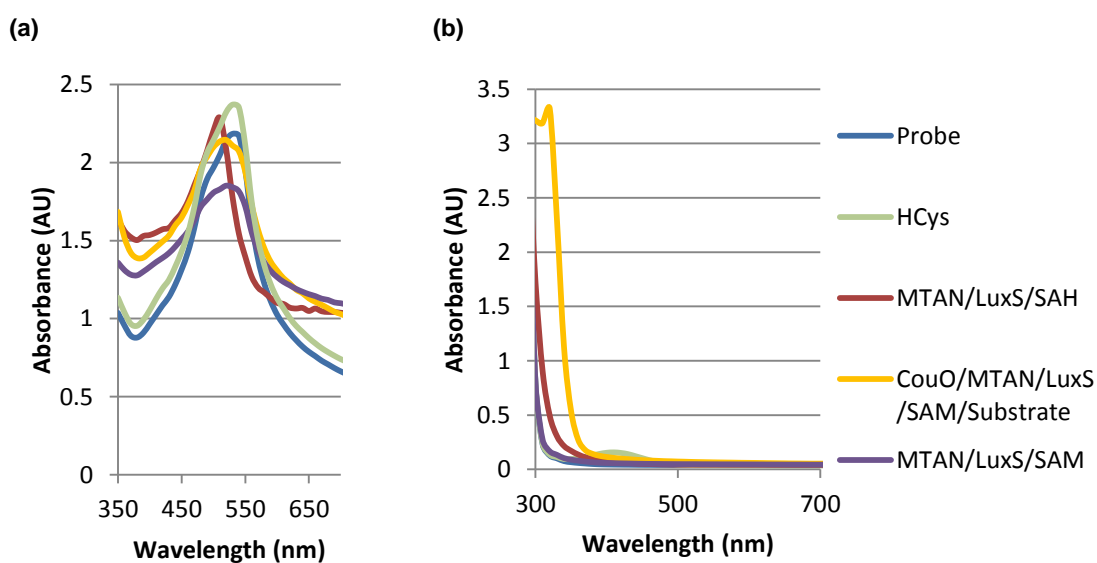
Two such examples are methyl viologen ( $MV^{2+}$ , **179**) and fluorone black (**180**) (**Figure 79**), which are both commercially available. In 2005, Strongin and coworkers reported the selective detection of HCys with  $MV^{2+}$  in plasma at neutral pH under refluxing conditions, which resulted in a colour change from colourless to blue and was not observed with Cys or GSH.<sup>[301]</sup> Additionally, the xanthene dye fluorone black (**180**) showed modest selectivity for HCys at pH 7.3, with an increase in absorbance at 510 nm being observed at room temperature.<sup>[294]</sup>



**Figure 79.** Structures of the HCys selective probes methyl viologen ( $MV^{2+}$ ) and fluorone black.

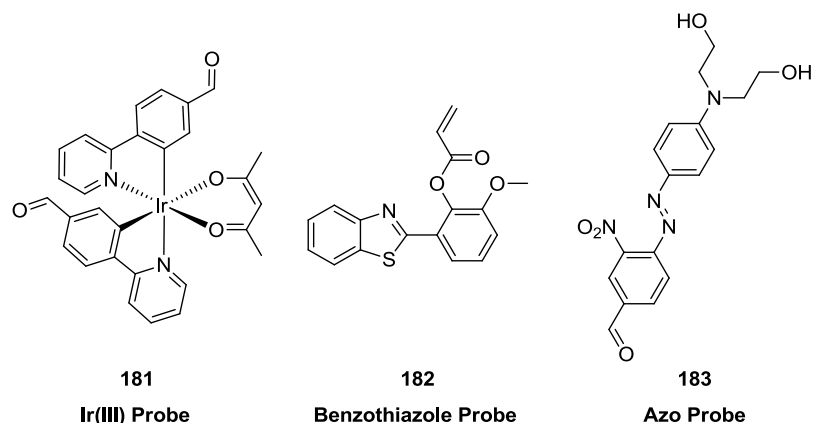
With this in mind,  $MV^{2+}$  and fluorone black were both investigated for their compatibility with the MTAN/LuxS coupled MT assay. Both compounds were incubated with HCys

(positive control); MTAN, LuxS and SAH (positive control); CouO, MTAN, LuxS, SAM and aminocoumarin substrate **67** (assay conditions); and MTAN, LuxS and SAM (negative control) and the absorbance spectra recorded after 4 hours (**Figure 80**). In the case of fluorone black, a small increase in absorbance at 510 nm was observed in the positive control with HCys, however this was not observed under the assay conditions or in the positive control with MTAN, LuxS and SAH (**Figure 80a**). In the case of  $MV^{2+}$ , an increase in absorption at 410 nm was observed in the positive control with HCys, which was not observed under the assay conditions or in the MTAN/LuxS/SAH positive control system (**Figure 80b**). Therefore, it was concluded that neither of these probes were suitable for this system.



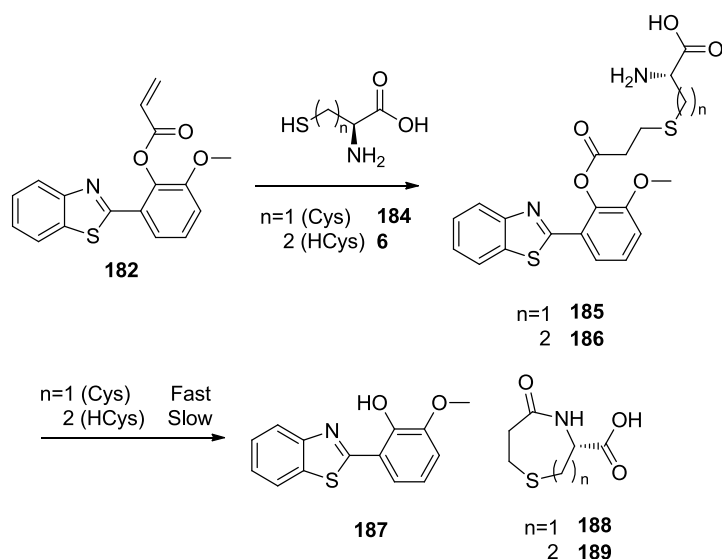
**Figure 80.** Absorption spectra of (a) fluorone black (**180**) and (b)  $MV^{2+}$  (**179**) after 4 hours incubation at 37 °C with the MTAN/LuxS assay components.

Based on these results, it was decided that detection of HCys based on formation of an  $\alpha$ -amino carbon-centered radical was not compatible with the MTAN/LuxS coupled assay system. Therefore, probes based on an alternative mechanism were investigated. Thiol reactive sensors that work *via* the nucleophilic attack of the thiol group into an electrophilic centre of the probe molecule have been a popular approach due to the excellent nucleophilicity of the thiol group. Three examples that were of particular interest for application in this project are shown in **Figure 81**.



**Figure 81.** Structures of HCys probes investigated for use in the MT assay based on nucleophilic attack of thiol group.

Huang and coworkers have reported the Ir(III) probe **181** as a selective phosphorescent sensor for HCys, which is activated by the nucleophilic attack of HCys into the aldehyde, followed by cyclisation to form a thiazinane ring.<sup>[302]</sup> The observed selectivity for HCys over Cys and GSH was hypothesised to be due to a photo-induced electron transfer process not seen for HCys. The second example, **182**, is a benzothiazole compound and was designed by Strongin *et al.* for the differentiation of HCys and Cys *via* a conjugate addition and cyclisation sequence.<sup>[297]</sup> In this example, the thiol attacks the enone in a 1, 4- fashion followed by cyclisation and extrusion of benzothiazole **187** (Scheme 46). As the rate of the



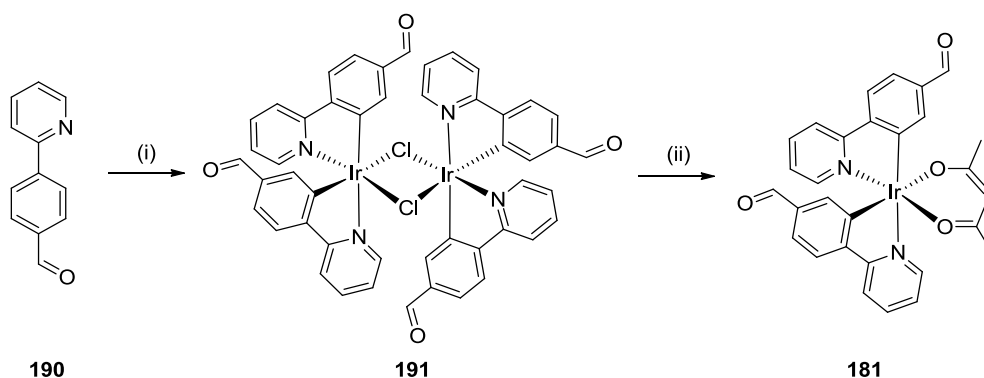
**Scheme 46.** Mechanism of simultaneous detection of Cys and HCys using benzothiazole probe **182**.



cyclisation in the second stage is faster for Cys than for HCys due to the formation of 7-membered ring (**188**) for Cys compared to an 8-membered ring (**189**) for HCys, it was possible to use the same probe to distinguish between the two analytes.<sup>[297]</sup>

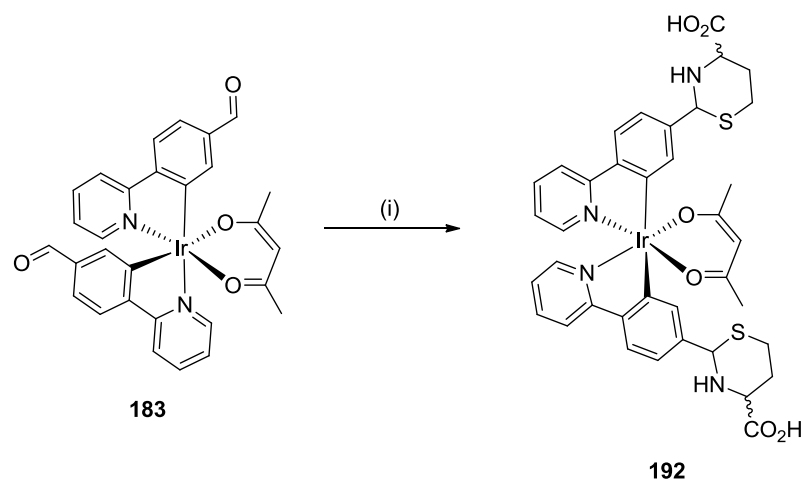
Finally, azo probe **183** was also of interest. Although it was not reported to differentiate between Cys and HCys, the probe can operate in aqueous conditions at neutral pH and moreover, the colour change from red to orange arising from formation of the thiazinane ring after addition of HCys or Cys was reported to be visible to the naked eye, which could facilitate the development of a colourimetric assay, which would simplify the screening process.

In order to evaluate the probes, each was synthesised according to the reported procedure and the reaction with HCys investigated. Firstly, Ir(III) probe **181** was prepared in 70% isolated yield over two steps from Ir(III)Cl<sub>3</sub>·xH<sub>2</sub>O and pyridine **190** via the chloro-bridged dimer **191**, followed by ligand exchange by heating with acetylacetonone (acac) (**Scheme 47**).



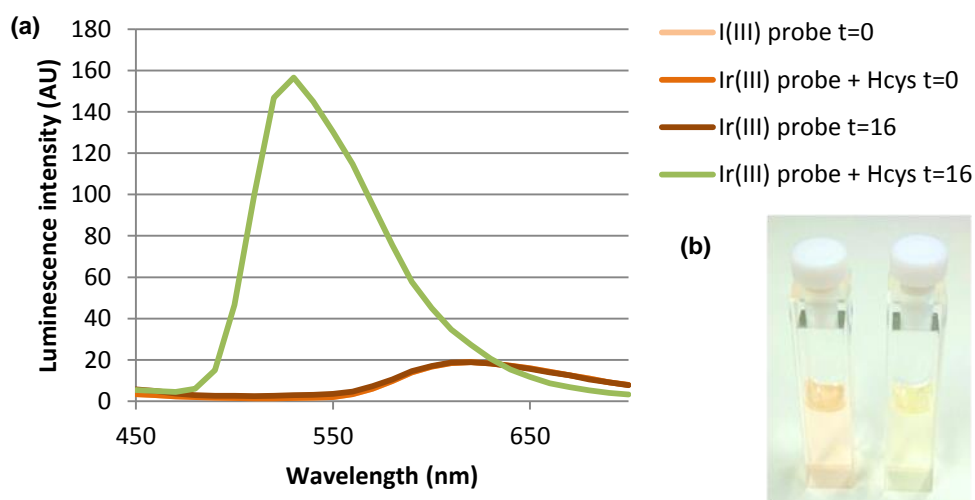
**Scheme 47.** Synthesis of Ir(III) **181**. *Reagents and conditions:* (i) Ir(III)Cl<sub>3</sub>·xH<sub>2</sub>O, water/2-ethoxy ethanol (1:3 v/v%), reflux, 24 h. (ii) acetylacetonone, Na<sub>2</sub>CO<sub>3</sub>, 2-ethoxy ethanol, 90 °C, 1 h, 70% over 2 steps.

With the Ir(III) probe in hand, the literature results were reproduced to confirm that the compound was able to detect HCys in solution. To this end, a sample of probe **181** was incubated with an excess (200 eq.) of HCys in DMSO/HEPES (**Scheme 48**). In addition, a negative control was carried out in which no HCys was added.



**Scheme 48.** Reaction of Ir(III) probe **181** with HCys. *Reagents and conditions:* (i) HCys (200 eq.), DMSO:HEPES (50 mM pH 7.2) (9:1 v/v%), 37 °C, 16 h, *product not isolated*.

The phosphorescence intensity of the resulting solution was measured and compared to the negative control (**Figure 82**). A strong emission peak was observed at 530 nm after 16 hours of incubation with HCys, whilst very low levels of luminescence were observed in the negative control, which mirrored the result reported by Huang and coworkers. Furthermore, a colour change from orange to green was visible to the naked eye.

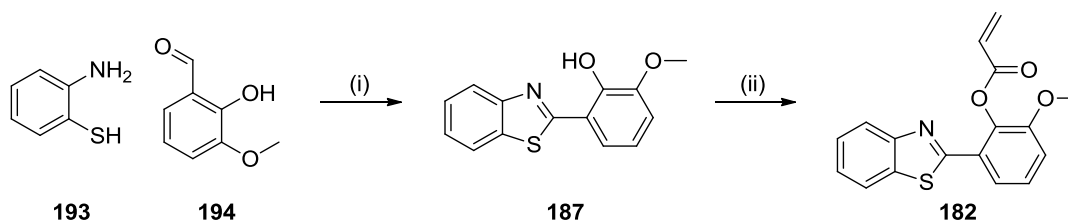


**Figure 82.** (a) Change in the phosphorescence emission spectra of Ir(III) probe **183** (20  $\mu$ M) in DMSO-HEPES buffer (50 mM, pH 7.2, 9:1 v/v) after incubation at 37 °C for 16 h with (green) and without (orange) 200 eq. HCys.  $\lambda_{\text{ex}}=360$  nm. (b) Colour change of Ir(III) probe **183** after incubation with 200 eq. HCys.

Encouraged by this result, the Ir(III) probe **183** was tested with the MTAN/LuxS coupled MT assay. However, **183** was found to be insoluble under the assay conditions, which

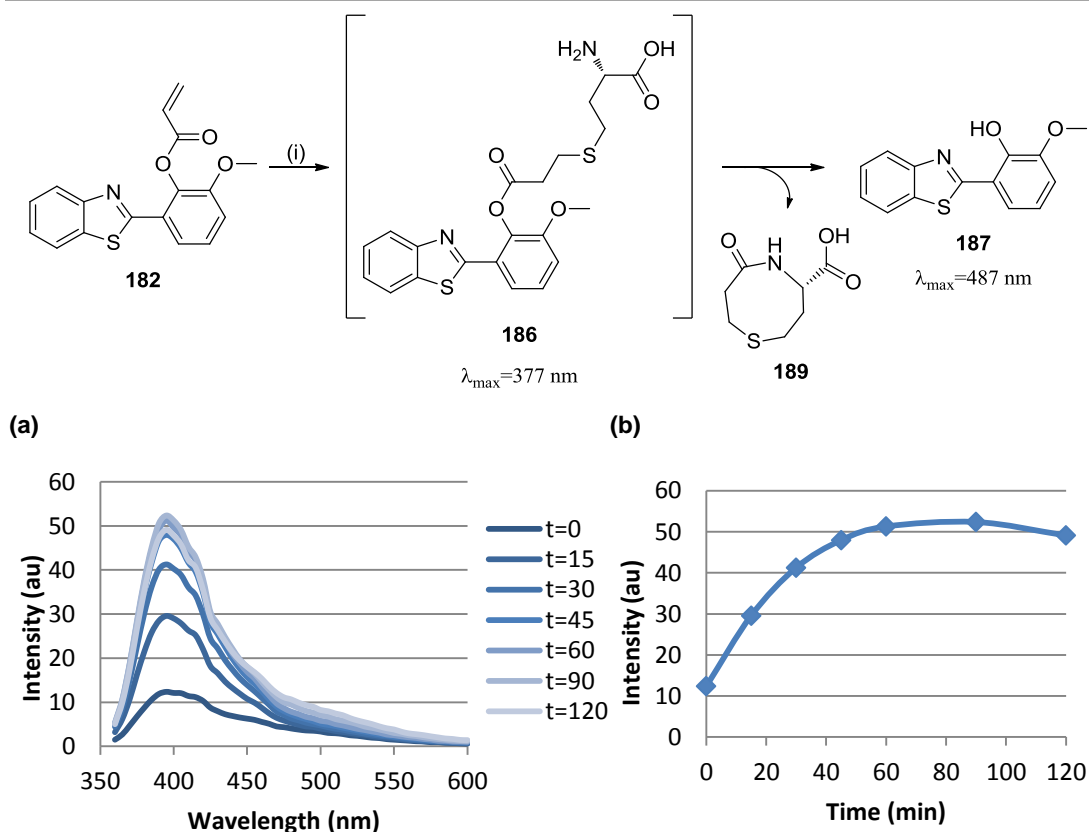
only contained 5% v/v DMSO compared to 90% v/v DMSO in the literature conditions. It has previously been shown that the activity of NovO rapidly decreases above 10% v/v DMSO,<sup>[260]</sup> and consequently it was concluded that this probe was not suitable for use in the assay. Furthermore, Huang and coworkers reported a maximum response with 200 eq. HCys with respect to **183** and thus it was predicted that the probe may not be sensitive enough at very low concentrations of HCys in the MTAN/LuxS coupled MT reaction.

Next, the suitability of benzothiazole probe **182** for the MTAN/ LuxS coupled MT assay was evaluated. To this end, **182** was synthesised in two steps *via* condensation of 2-mercapto aniline **193** and salicylic acid derivative **194** under oxidative conditions to form the benzothiazole **187**. Subsequent acylation of the phenol with acryloyl chloride afforded the target compound **182** in 41% over two steps (**Scheme 49**).



**Scheme 49.** Synthesis of benzothiazole probe **182**. *Reagents and conditions:* (i) H<sub>2</sub>O<sub>2</sub>, aq. HCl, 22 °C, 1.5 h, 59%. (ii) acryloyl chloride, NEt<sub>3</sub>, CH<sub>2</sub>Cl<sub>2</sub>, 22 °C, 4 h, 69%.

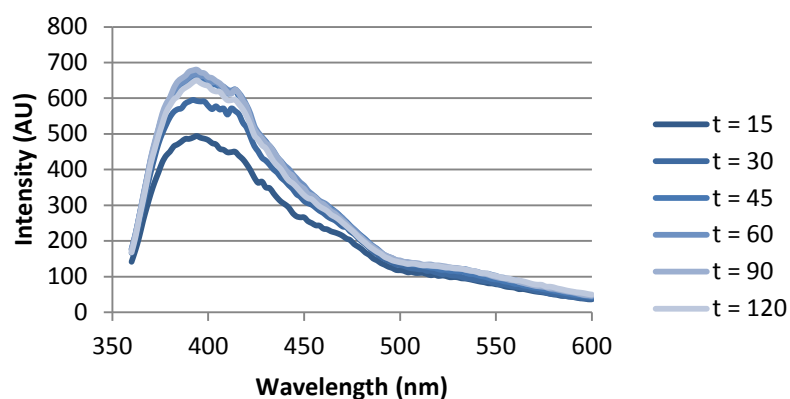
Next, the probe **182** was incubated with HCys under the same conditions as those reported by Strongin and coworkers and the fluorescence intensity ( $\lambda_{\text{ex}}=304$  nm) of the reaction was monitored over 2 hours. As shown in **Figure 83**, an increase in fluorescence intensity at 395 nm was observed over the first 60 minutes after addition of HCys, which correlates to the formation of HCys adduct **186**. After 60 minutes, a gradual decrease in fluorescence intensity was observed, which was concurrent with an increase in fluorescent intensity between 450-500 nm. These data were consistent with the slow degradation of HCys adduct by cyclisation and degradation into products **187** and **189** (**Scheme 46**).



**Figure 83.** (a) Change in fluorescence intensity spectrum of Strongin probe **182** over 2 hours after the addition of HCys ( $\lambda_{\text{ex}}=304$  nm). (b) Change in fluorescence intensity at 395 nm over 2 hours after the addition of HCys ( $\lambda_{\text{ex}}=304$  nm). Reagents and conditions: (i) HCys (20  $\mu\text{M}$ ), EtOH/50 mM potassium phosphate buffer, pH 7.4 (2:8 v/v), 22  $^{\circ}\text{C}$ , 2 h, products not isolated.

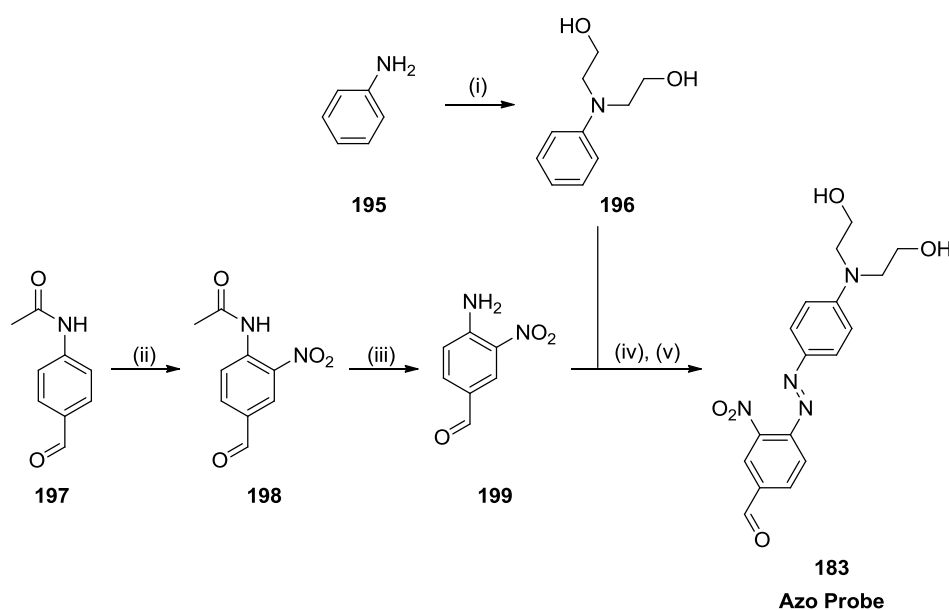
Next, probe **182** was incubated with *E. coli* cell free extract in order to determine whether there would be any background reactions or degradation of **182** under the assay conditions. To this end, **182** was incubated with *E. coli* cell free extract and the fluorescence intensity measured over 2 hours (**Figure 84**). Unfortunately, a strong increase in fluorescence intensity at 395 nm was observed, indicating that there were other nucleophiles present in the cell free extract capable of reacting with **182** in an analogous manner to HCys. Therefore, it was concluded that **182** was unsuitable for use in the MTAN/LuxS coupled MT assay.

As identification of a suitable HCys-selective probe had not been possible, it was decided to synthesise and evaluate the azo probe **183**, which is reported to detect both Cys and HCys. To this end, aniline (**195**) was first *N, N*-dialkylated in 27% isolated yield to provide aniline derivative **196**, which was coupled with aldehyde **199** to provide the target compound. Synthesis of aldehyde **199** was achieved by nitration of the commercially available aldehyde



**Figure 84.** Change in fluorescence intensity spectrum of Strongin probe **182** incubated in *E. coli* cell free extract over 2 hours ( $\lambda_{\text{ex}}=304$  nm).

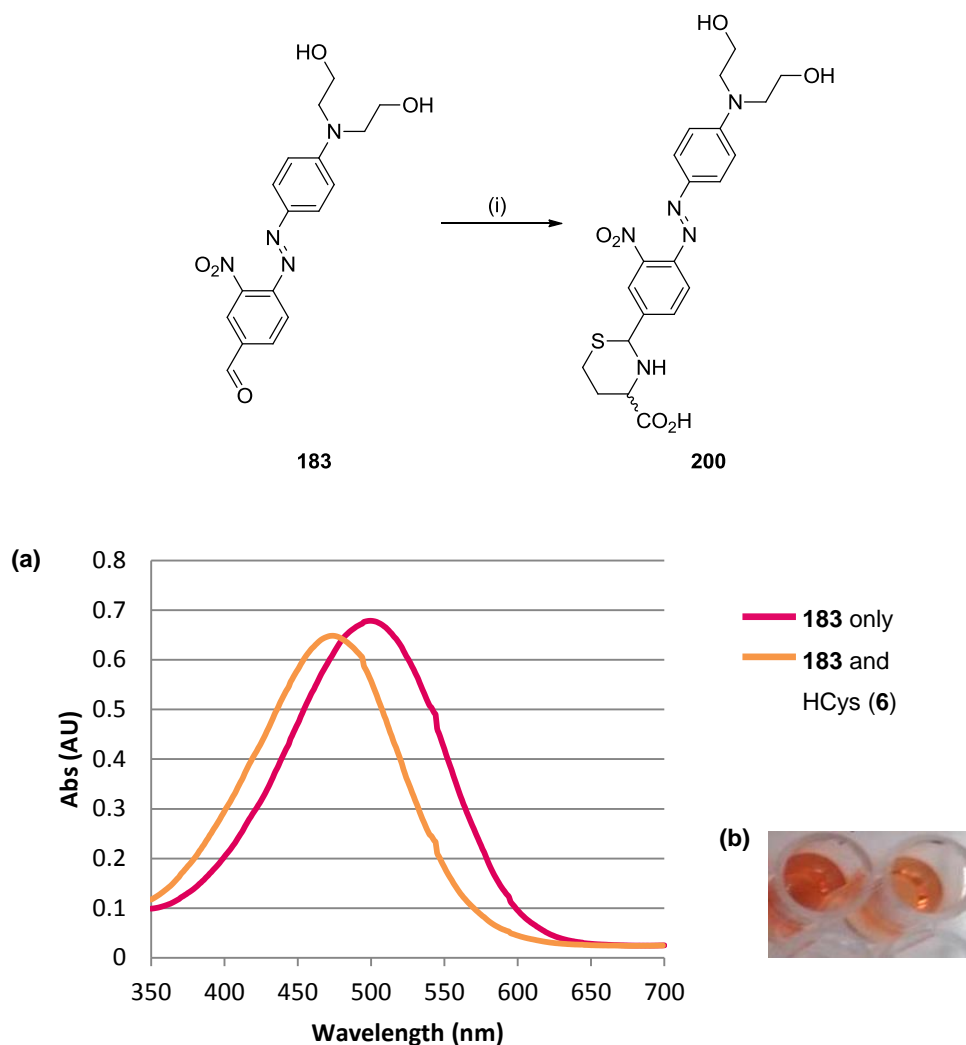
**197** in 27% isolated yield. The resulting nitro compound **198** was deacetylated under acidic conditions to provide aniline **199** in 98% isolated yield. Finally, the two fragments were coupled *via* diazotization of **199** and subsequent addition of **196** to provide the desired azo linked product **183** in 70% isolated yield from **199** (Scheme 50).



**Scheme 50.** Synthesis of azo probe **183**. Reagents and conditions: (i) 2-chloroethanol,  $\text{CaCO}_3$ , KI,  $\text{H}_2\text{O}$ , reflux, 19 h, 27%. (ii)  $\text{HNO}_3$ ,  $\text{H}_2\text{SO}_4$ ,  $0^\circ\text{C}$ , 20 min, 35%. (iii)  $\text{HCl}/\text{MeOH}$ , reflux, 1 h, 98%. (iv)  $\text{NaNO}_2$ , aq.  $\text{HCl}$ ,  $\text{H}_2\text{O}$ ,  $0^\circ\text{C}$ , 1 h. (v) **196**, 16 h,  $0-22^\circ\text{C}$ , 70% over 2 steps from **199**.

Next, **183** was incubated with HCys and the change in the absorption spectrum between 350-700 nm was recorded (Figure 85). Analysis of the reaction mixture by LC-MS confirmed quantitative conversion to the thiazinane product **200** in 15 minutes at room temperature.

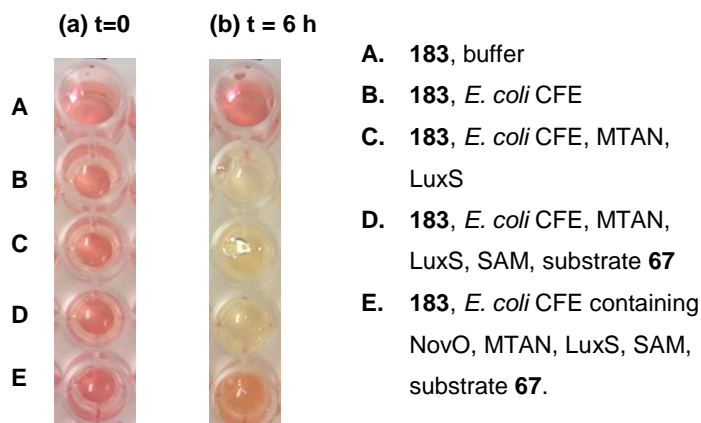
Furthermore, a blue-shift in the absorbance spectrum from 509 nm to 466 nm was observed (**Figure 85a**), which was concurrent with a visible colour change from red to orange (**Figure 85b**).



**Figure 85.** Reaction of azo probe **183** with HCys. *Reagents and conditions:* (i) HCys, DMF: water (9:1 v/v), 15 min, 22 °C, quantitative conversion by LC-MS, *product not isolated*.

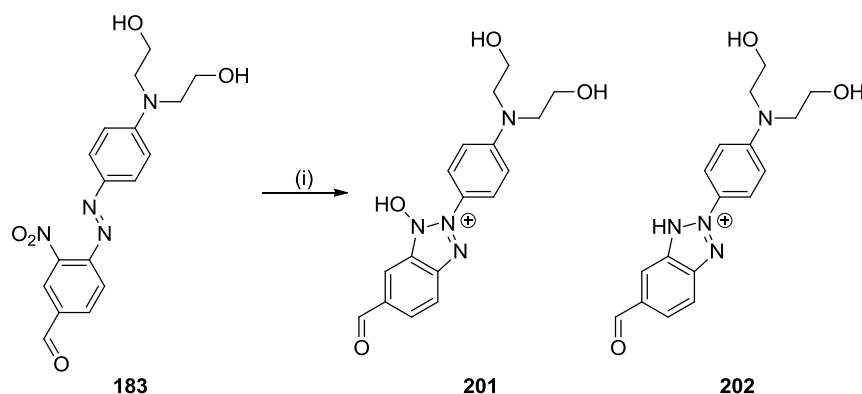
Having established that **183** was an effective colourimetric probe for HCys, it was next necessary to determine whether there were any competing reactions with other assay components. To this end, **183** was incubated with *E. coli* cell free extract; *E. coli* cell free extract, MTAN and LuxS; *E. coli* cell free extract, MTAN, LuxS, SAM and aminocoumarin substrate **67**; and *E. coli* cell free extract containing NovO, MTAN, LuxS, SAM and aminocoumarin substrate **67** (**Figure 86**), the latter of which was designed as a positive control to mimic the MTAN/LuxS coupled MT assay. Interestingly, after 6 hours incubation

at 37 °C, discolouration of **183** was observed for experiments containing *E. coli* cell free extract (**Figure 86**, rows **B-D**), whilst a colour change from red to orange was observed in the positive control reaction. These results strongly indicated that the azo compound **183** was not stable under the assay conditions.



**Figure 86.** Bleaching of azo probe **183** in *E. coli* cell free extract.

To further investigate the cause of this bleaching effect, a sample of **183** was incubated in *E. coli* cell free extract at 37 °C for 16 hours and the products separated and isolated by silica gel chromatography (**Scheme 51**). Characterisation of the two products by LC-MS and NMR spectroscopy identified the two products to be benzotriazole derivatives **201** and **202**. It was hypothesised that these compounds may have been formed *via* partial reduction of the nitro group of **183**, followed by cyclisation to form the benzotriazole core. Biodegradation of both nitroaromatic compounds and azo compounds is well known.<sup>[303–305]</sup> In particular, type 1 nitroreductases from *E. coli* catalyse the two-electron reduction of nitro groups *via* nitroso and hydroxylamine intermediates.<sup>[306]</sup> With this in mind, it was hypothesised that a nitroreductase in the *E. coli* cell free extract was responsible for the observed bleaching of **183**. Consequently, it was concluded that **183** was not appropriate for the MTAN/LuxS coupled assay.



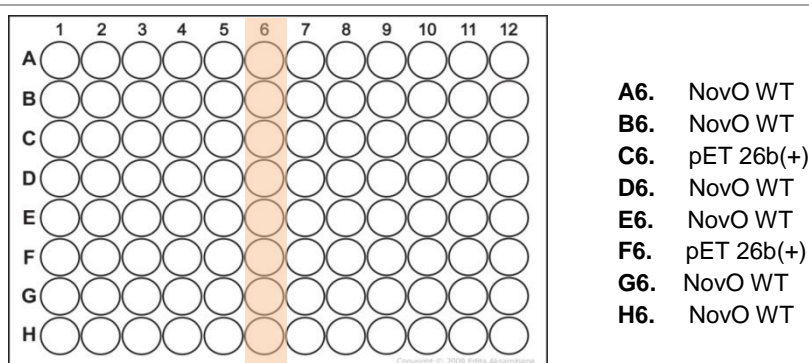
**Scheme 51.** Decomposition of azo probe **183** in *E. coli* cell free extract. *Reagents and conditions:* (i) *E. coli* cell free extract, 50 mM potassium phosphate buffer pH 6.5, 37 °C, 16 h.

In summary, it was concluded that background reaction of the thiol reactive probes with the *E. coli* cell free extract led to non-reproducible results and as such, it was not possible to design a robust, medium-to-high throughput assay using in *E. coli* cell free extract. It was hypothesised that purification of the CASTing libraries following expression may be necessary. Although this would greatly reduce the throughput of the assay, it was thought that this strategy could facilitate the screening of a small number of plates against a smaller substrate library, thereby providing proof of concept for the evolution of NovO towards a wider substrate scope using a CASTing approach.

#### 4.4 Purification and screening of purified CASTing libraries

Before purifying plates of mutants for all 5 CASTing libraries, it was decided to first obtain proof of concept for this approach by testing the overexpression and purification of a single plate of mutants from library 1. To this end, a sample of the DNA library of CASTing library 1 was transformed into BL21 (DE3). Additionally, WT NovO DNA and empty pET26b(+) were transformed separately into BL21 (DE3) for use as positive and negative control reactions, respectively. Next, a master plate was prepared by picking 88 single colonies into a 96-well plate containing LB broth, the antibiotic selection marker (kanamycin) and glycerol and growing the colonies for 8 hours at 37 °C. The remaining 8 wells (column 6) were inoculated with a colony from the transformation of WT NovO or empty pET26b(+) according to **Figure 87**.



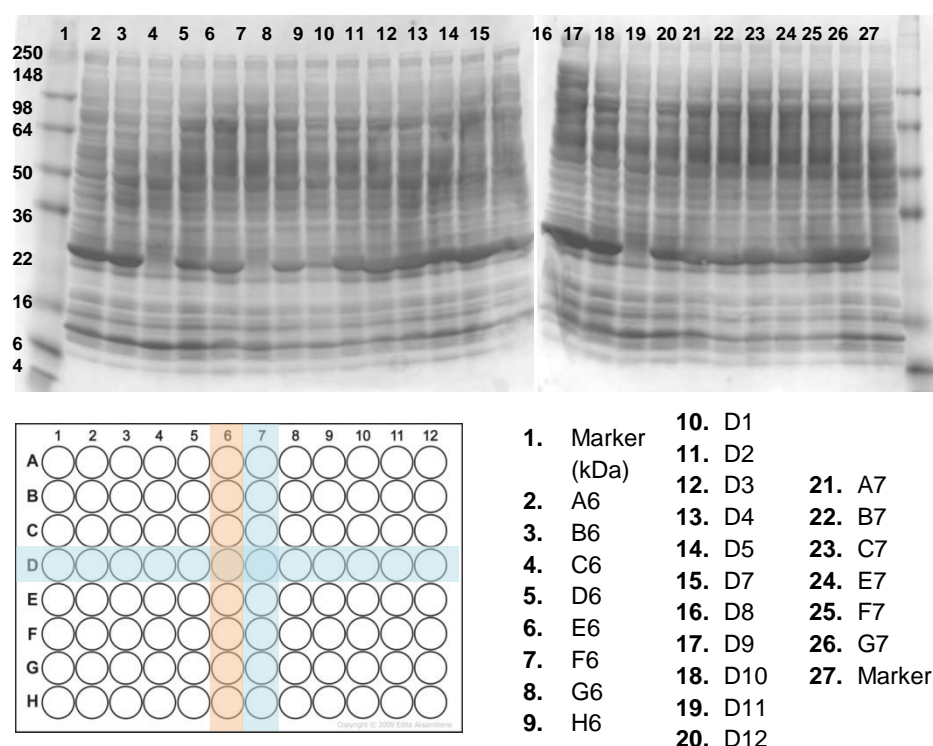


**Figure 87.** Format of MT assay control lane in expression of CASTing library 1 in BL21 (DE3). All other wells were inoculated with a colony from the transformation of the CASTing DNA library. Control lane highlighted in orange.

Next, a deep-well 96-well plate containing Magic Media, ‘Component B’ and kanamycin was inoculated using seed colonies from the master plate. Cultures were grown at 30 °C to an OD ~2 before lowering the incubation temperature to 18 °C and incubating for a further 16 hours. The agitation rate was found to have a significant effect on the rate of growth of the cultures, with faster growth and higher post induction ODs being obtained at 800 rpm compared to 220 rpm (OD ~7-8 and ~5-6, respectively), presumably due to more efficient oxygen transfer into the cultures at the higher speed. Therefore, all further expression in 96-well plates was done with an agitation rate of 800 rpm. After harvesting the cell pellets by centrifugation, the cultures were resuspended in potassium phosphate buffer and lysed using a 96 pronged ultrasound probe. However, lysis by this method was found to be unreliable and led to contamination between wells by ‘jumping’ of the samples upon sonication. Therefore, alternative lysis conditions were sought.

Out of the many conditions for bacterial cell lysis that are known, the most commonly employed involve subjecting the cultures to one or more freeze/thaw cycles and/or the use of lysozyme.<sup>[256]</sup> Lysozyme catalyses the hydrolysis of 1, 4- $\beta$  linkages between *N*-acetylmuramic acid and *N*-acetyl-D-glucosamine in peptidoglycan, which is a structural component of bacterial cell walls,<sup>[307]</sup> whilst the freeze/thaw cycle is well known to disrupt cell membranes by ice crystal formation and is widely used to promote cell lysis. To investigate this approach, plates were stored at -80 °C, before thawing at room temperature and resuspending in a lysis buffer containing lysozyme (0.5 mg/mL) and benzonase (25 U/mL). Benzonase is a non-proteolytic endonuclease, which is commonly used additive to degrade DNA and RNA and has the effect of reducing the viscosity of the cell lysate, improving ease of handling in subsequent experiments.<sup>[308]</sup> The plate was incubated at room

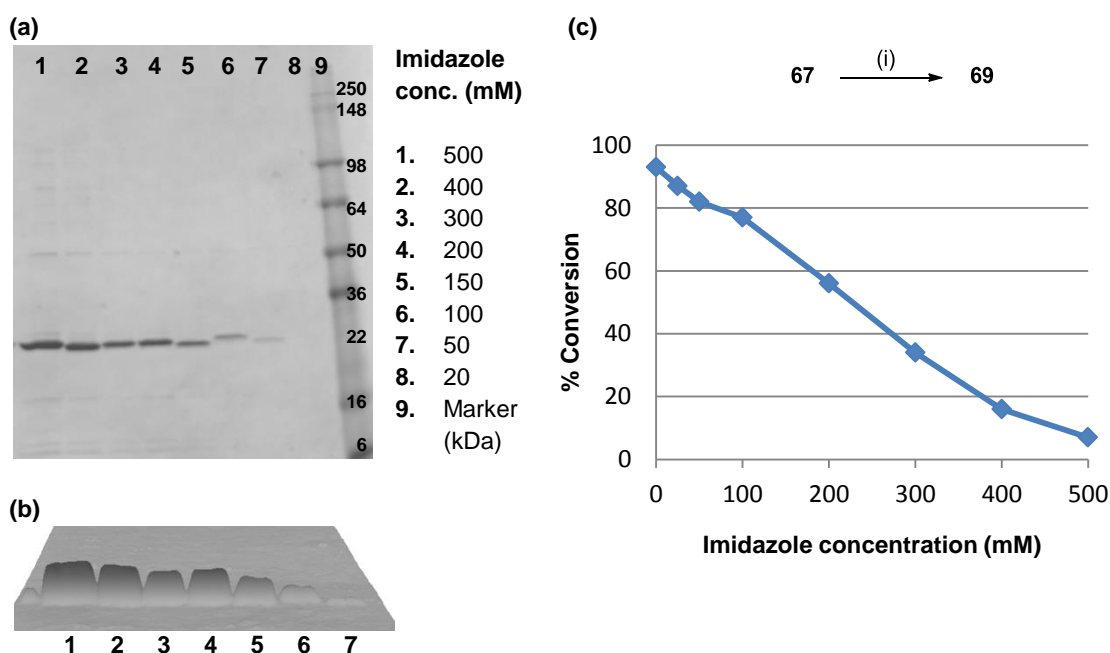
temperature for 1 hour before clarifying the lysates by centrifugation and analysing the supernatant of the control lane and a cross section from the plate by SDS PAGE (**Figure 88**). Protein bands corresponding to NovO were clearly visible in all samples, with the exception of wells C6 and F6 resulting from the ‘expression’ of empty pET26b(+). Additionally, no expression of NovO or a mutant of NovO was observed in wells D9 and G7, which could be due to the presence of a stop codon in these mutants. Nonetheless, it was concluded that this was an appropriate method for expression and cell lysis of CASTing libraries and that there were good levels of soluble protein expression.



**Figure 88.** Analysis of soluble protein fractions from the heterologous expression of NovO CASTing library 1 in 96 well plate format by SDS PAGE. Enzymes were expressed in *E. coli* BL21 (DE3) in Magic Media at 30 °C to OD ~2 then 18 °C overnight. Theoretical mass of WT 6xHis-NovO: 26.5 kDa. Control lane highlighted in orange. Additional wells analysed by SDS PAGE highlighted in blue.

Having confirmed the overexpression of NovO mutants in 96-well plate format, it was then necessary to determine the optimum imidazole concentration for elution of the protein in the final stage of purification by Ni-NTA affinity chromatography. Therefore, the purification of a plate of NovO mutants from the expression of CASTing library 1 was carried out with a varying concentration of imidazole in the elution step (**Figure 89a**) and the amount of purified protein obtained using each concentration of imidazole was visualised by SDS

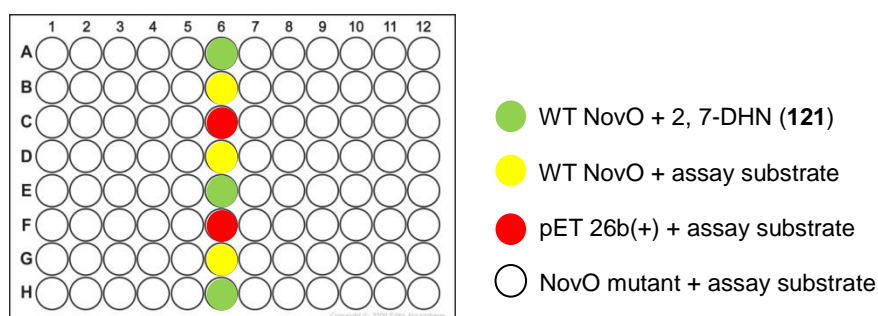
PAGE. This was aided by analysis of the 3D representation of the gel, in which it could clearly be seen that a rapid drop off in protein concentration was obtained for imidazole concentrations below 200 mM (**Figure 89b**). As it was preferable to use the resulting eluent directly in the MT assay without an additional desalting step, it was also necessary to determine whether imidazole had an inhibitory effect on the MT reaction. To this end, the methylation of aminocoumarin **67** using WT NovO was carried out in 8 parallel reactions, each of which was spiked with an increasing concentration of imidazole (**Figure 89c**). The imidazole was prepared as a stock solution in potassium phosphate buffer and adjusted to pH 6.5 prior to use. This was to ensure that a change in activity was not due to variation in pH caused by the addition of imidazole. Analysis of each reaction by HPLC revealed an almost linear negative correlation between imidazole concentration and percentage conversion to methylated product **69**. These data strongly indicated that imidazole had an inhibitory effect on the activity of the MT. Drawing together the results from both of these experiments, 150 mM imidazole was chosen as a compromise between elution efficiency and MT activity inhibition.



**Figure 89.** Investigation into optimal imidazole concentration in elution buffer for the purification of NovO mutants. **(a)** Analysis of effect imidazole concentration on elution efficiency by 2D SDS PAGE. **(b)** 3D image of SDS PAGE gel showing effect of imidazole concentration on elution efficiency. **(c)** Effect of imidazole concentration on activity of NovO. % Conversion of substrate **67** to methylated product **69** measured by area/area% conversion by HPLC.

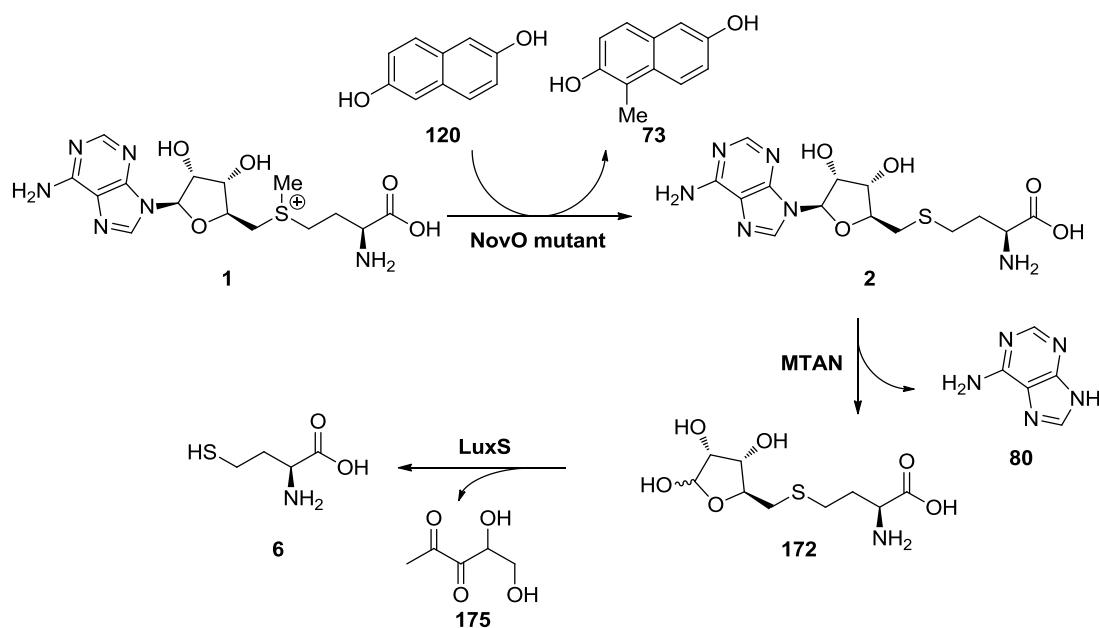
Based on these results, a 96-well plate from the expression of NovO mutants from CASTing library 1 was purified using a KingFisher™ Flex purification system (an automated protein purification system), eluting the enzyme with 150 mM imidazole. The resulting eluent was used directly in the MTAN/LuxS coupled MT assay. In order to determine whether the assay was functional, a control lane (column 6) was included on the assay plate (**Figure 90**), which was carried through from the overexpression of WT NovO or empty pET26b(+) in the previous stage. The control lane contained three alternating reactions:

1. WT NovO and 2, 7-dihydroxynaphthalene (2, 7-DHN, **121**), which is a known substrate for WT NovO (**Figure 90**, shown in green).<sup>[107]</sup> A positive readout from the assay was expected for these wells.
2. WT NovO and the assay substrate (**Figure 90**, shown in yellow). A negative readout from the assay was expected for these wells.
3. No enzyme control (from use of empty pET26b(+) in these wells) and the assay substrate (**Figure 90**, shown in red). A negative readout from the assay was expected for these wells.



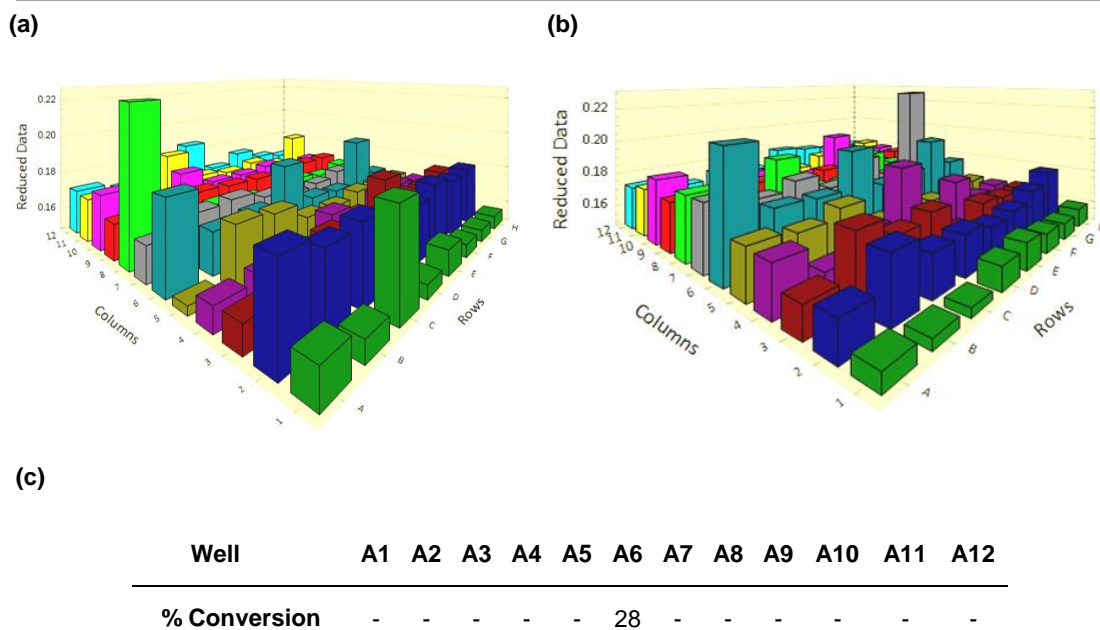
**Figure 90.** Plate format for screening of CASTing libraries using MTAN/LuxS coupled MT assay and purified enzyme.

The methyltransferase assay was carried out using 2, 6-DHN (**120**) as the assay substrate and purified NovO mutants from CASTing library 1 (**Scheme 52**).



**Scheme 52.** MTAN/LuxS coupled MT assay using purified enzymes from expression of CASTing library 1.

After 16 hours incubation at 37 °C, 2x50  $\mu$ L aliquots of the assay reaction from each well were transferred into two 96-well plates and mixed with 10 eq. Ellman's reagent to detect HCys present. This process was carried out in duplicate to assess the reproducibility of the results. After incubating the plates at room temperature for 30 minutes, the absorbance at 405 nm was measured for both plates, which were arbitrarily labelled **a** and **b** (**Figure 91a** and **b**). Additionally, the percentage conversion to methylated product **73** in row A was analysed by HPLC in order to assess the accuracy of the assay (**Figure 91c**). Although a positive response was obtained for the positive control wells in both plates, comparison of the two duplicates revealed a large degree of variation across the rest of the plate. Furthermore, analysis by HPLC showed no conversion in any of the Row A samples (with the exception of the positive control in well A6), whilst the absorbance data from plate **a** suggested that well A8 was a hit with higher conversion than the positive control. Consequently, it was concluded that the MTAN/LuxS coupled assay was not robust enough for screening the CASTing libraries, even using purified enzyme and that further development work would be necessary to overcome the issues of reproducibility and robustness, which was not possible within the timeframe of this project.

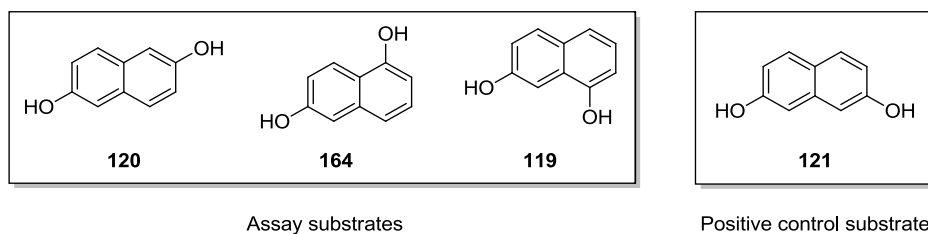


**Figure 91.** Comparison of absorption at 414 nm and HPLC analysis, where ‘Columns’ and ‘Rows’ refer to columns 1-12 and rows A-H on the 96-well plate.

Despite extensive efforts to establish a screening platform for the CASTing libraries based on the MTAN/LuxS coupled assay, it was not possible to develop a robust assay in the timeframe available. Consequently, it was decided to screen a small number of plates by UPLC using *E. coli* cell free extract from the expression of NovO CASTing plates.

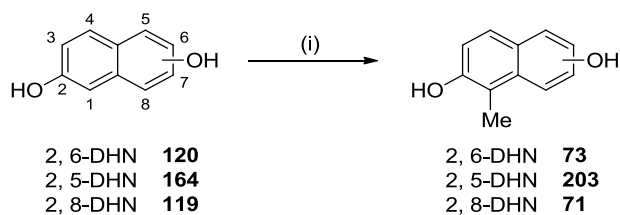
## 4.5 Screening of CASTing libraries by UPLC

Due to the relatively low throughput of screening plates by UPLC, the size of the substrate library was decreased to the three dihydroxynaphthalene (DHN) regioisomers **120**, **164** and **119** (**Figure 92**). Additionally, **121** was used as positive control substrate in all screening experiments.



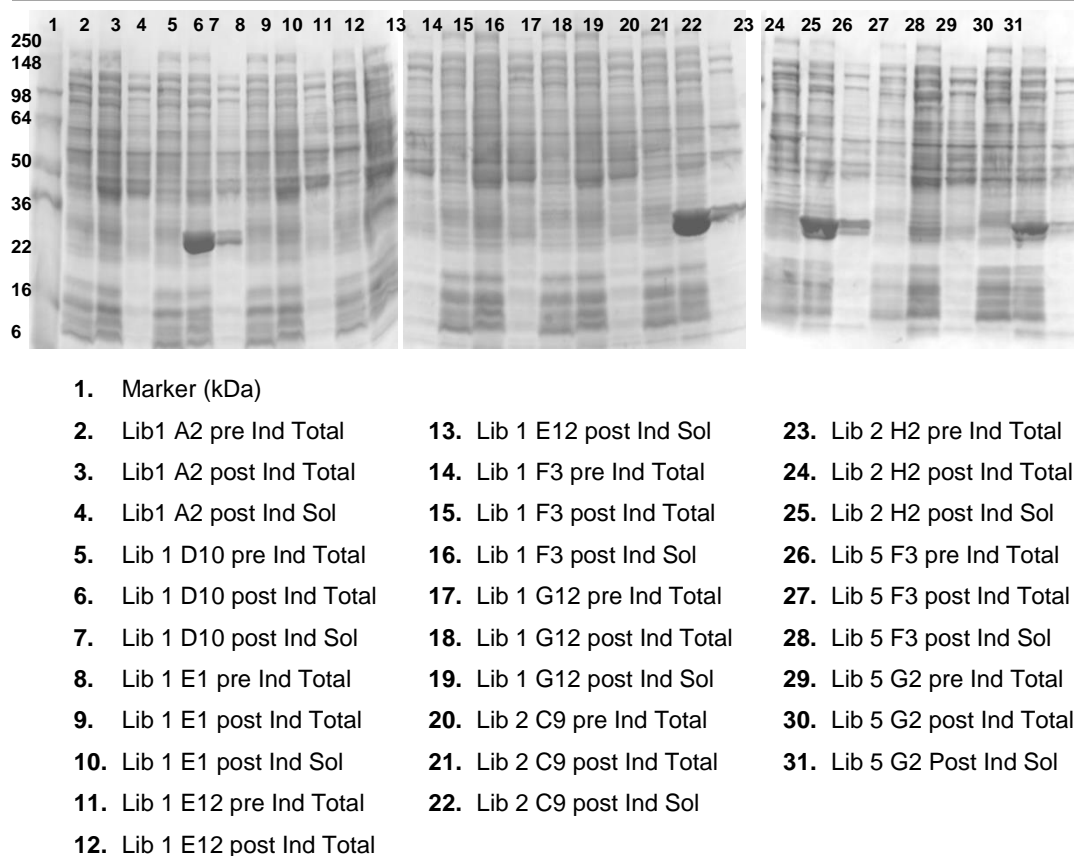
**Figure 92.** Substrate library for screening of CASTing libraries by HPLC.

Each of the three substrates was screened separately against one 96-well plate of NovO mutants for each of the five CASTing libraries, resulting in a total of 15 screening plates. The *E. coli* cell free extract from the overexpression of the CASTing libraries in plate format was prepared as described above and used for all screening experiments. The MT assays were incubated at 37 °C for 16 hours before quenching with acetonitrile, clarifying by centrifugation and analysing the supernatant by UPLC (**Scheme 53**). MTAN was included in the assay mixture as it has been shown to improve conversions, presumably by reducing byproduct inhibition *via* the *in situ* degradation SAH.<sup>[289]</sup>



**Scheme 53.** Screening of CASTing libraries by UPLC. *Reagents and conditions* (i) *E. coli* cell free extract from expression of CASTing libraries, substrate (1 mM), SAM (2 mM), MTAN (1  $\mu$ L), 50 mM potassium phosphate buffer, pH 6.5, 37 °C, 16 hours.

From these experiments, putative hits were identified from libraries 1, 2 and 5. Hits were putative as without reference samples of the methylated products **73**, **203** and **71**, it was not possible to confirm product formation in the UPLC assay. Therefore, putative hits were further analysed by LC-MS. In order to confirm that these mutants did display higher activity than WT NovO for the given substrate, the top 10 mutants were expressed in B121 (DE3) and the protein expression in post induction total and soluble fractions was analysed by SDS PAGE (**Figure 93**). Interestingly, out of the 10 expression experiments, only four of the mutants were expressed in *E. coli* under the standard expression conditions (Library 1 D10, Library 2 C9, Library 2 H2 and Library 5G2).



**Figure 93.** Analysis of heterologous expression of putative hits from CASTing libraries 1-5 in *E. coli* BL21 (DE3) in Magic Media at 30 °C to an OD~2 then at 18 °C overnight and analysed by SDS PAGE. Theoretical mass of WT 6xHis-NovO: 26.2 kDa. Ind.: induction.

To gain more information on the four proteins that were overexpressed, samples of the plasmid DNA in each case was isolated by MiniPrep and sequenced using T7 forward and reverse primers (**Table 17**). Out of the four constructs sequenced, one (Library 2 C9) was WT, whilst the others were a single or double mutants of NovO: the library 1 D10 mutant

Entry	Identifier	Sequencing results	Translated sequence
1	Lib1_D10	AGA/AGG	R/R
2	Lib2_C9	CGT/AAT	R/N (WT)
3	Lib2_H2	CGA/ATG	R/M
4	Lib5_G2	CAT/CCG	H/P

**Table 17.** Sequencing results of four hits from screening of CASTing libraries against three dihydroxynaphthalene regioisomer substrates.



was found to be the double arginine mutant NovO F14R/M17R, whilst library 2 H2 was the single mutant NovO N117M and library 5 G2 the double mutant Y216H/V218P.

In order to confirm that these proteins were active in the methylation of DHN substrates, the crude *E. coli* cell lysates from the overexpression of the NovO mutants were used in the MT assay. In these experiments, SAM was prepared *in situ* from methionine and 5'-deoxy-5'-chloroadenosine (CIDA) using SalL, as described in detail in Chapter 5. Additionally, the reactions were also carried out with WT NovO to determine which mutants gave higher levels of conversion. As shown in **Table 18**, in all but one of the experiments, a lower level of methylation was observed with the mutants compared to WT NovO. The one exception to this was the methylation of 2, 5-DHN with NovO N117M, which provided 18% conversion to methylated product **203** compared to 7% with WT NovO.

(i)

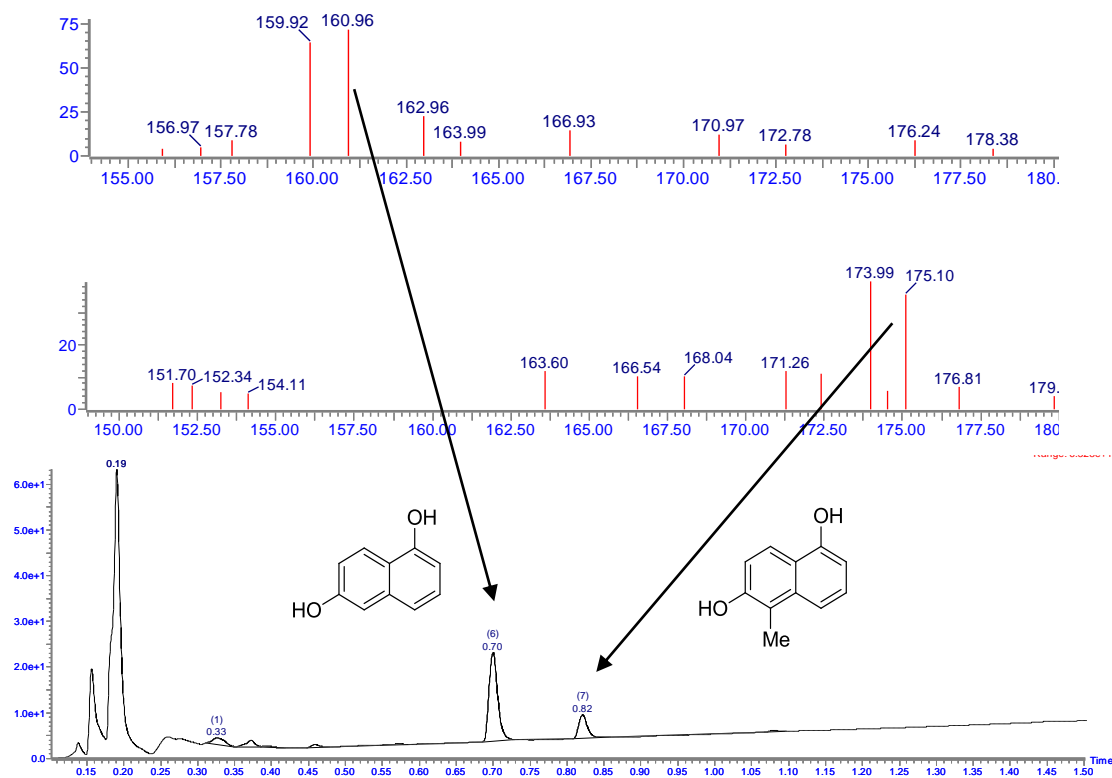
2, 6-DHN	<b>120</b>	2, 6-DHN	<b>73</b>
2, 5-DHN	<b>164</b>	2, 5-DHN	<b>203</b>
2, 8-DHN	<b>119</b>	2, 8-DHN	<b>71</b>
2, 7-DHN	<b>121</b>	2, 7-DHN	<b>72</b>

Mutant	% Conversion			
	2, 6-DHN	2, 5-DHN	2, 8-DHN	2, 7-DHN
	<b>120</b>	<b>164</b>	<b>119</b>	<b>121</b>
<b>F14R/M17R</b>	9	1	1	1
<b>N117M</b>	13	18	4	99
<b>Y216H/V218P</b>	9	1	1	1
<b>WT</b>	11	7	11	100

**Table 18.** Screening of hits from CASTing experiments against dihydroxynaphthalene regioisomers. *Reagents and conditions:* (i) L-Met (2 eq.), 5'-deoxy-5'-chloroadenosine (1.5 eq.), SalL (30  $\mu$ M), MTAN (5  $\mu$ M), *E. coli* cell lysate containing WT NovO or NovO mutants, resuspended at 10 mL/g cell pellet, DTT (1 eq.), BSA (1 mg/mL), 37  $^{\circ}$ C, 24 h. % Conversion determined by area/area% by HPLC.

To confirm this result, the sample was analysed by LC-MS. Gratifyingly, mass peaks corresponding to the methylated product **203** were clearly visible for the peak at 0.82 minutes in the LC trace (**Figure 94**). As such, the mutant NovO N117M was concluded to exhibit higher activity towards 1, 6-DHN compared to the WT enzyme.



**Figure 94.** LC-MS analysis (ES<sup>+</sup>) of methylation of 1, 6-DHN (**164**) with NovO N117M.

This work demonstrated proof of principle for the evolution of NovO towards a wider substrate scope. Despite the extensive setbacks in establishing a medium-to-high throughput screening platform to identify mutants harbouring beneficial mutations, it was still possible to screen a small number of mutants against three substrates. As such, the probability of identifying significantly improved variants of NovO was greatly diminished. With this in mind, work to establish a high throughput assay for MT activity using a SAH selective riboswitch is ongoing in the research group.

## **Chapter 5.**

# **Cofactor Synthesis and *In Situ* Cofactor Generation**

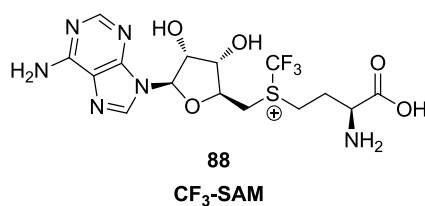
## 5 Cofactor Synthesis and *In Situ* Cofactor Generation

### 5.1 Cofactor synthesis

A cost-effective and scalable method for the synthesis of SAM and a variety of non-natural SAM analogues is required for biocatalytic alkylation using SAM-dependent methyltransferases (MTs) to be an economically viable and environmentally benign methodology for use in synthetic chemistry.<sup>[48]</sup> As discussed in Chapter 1, the development of a biocatalytic fluoroalkylation methodology is also of interest due to the privileged status of the fluorinated compounds in medicinal chemistry.<sup>[117,309]</sup> In order to develop a platform for biocatalytic trifluoromethylation, a trifluoromethylated analogue of SAM ('CF<sub>3</sub>-SAM') was required. Whilst there are a number of examples of both the chemical and enzymatic synthesis of non-natural SAM analogues,<sup>[107,224,310]</sup> the synthesis of a trifluoromethylated analogue has not been reported. Therefore, the two primary objectives for this project were to synthesise and utilise CF<sub>3</sub>-SAM with NovO and CouO; and to establish a scalable, one-pot procedure for the *in situ* synthesis of SAM and SAM analogues in tandem with NovO or CouO.

#### 5.1.1 Synthesis of a CF<sub>3</sub> analogue of SAM

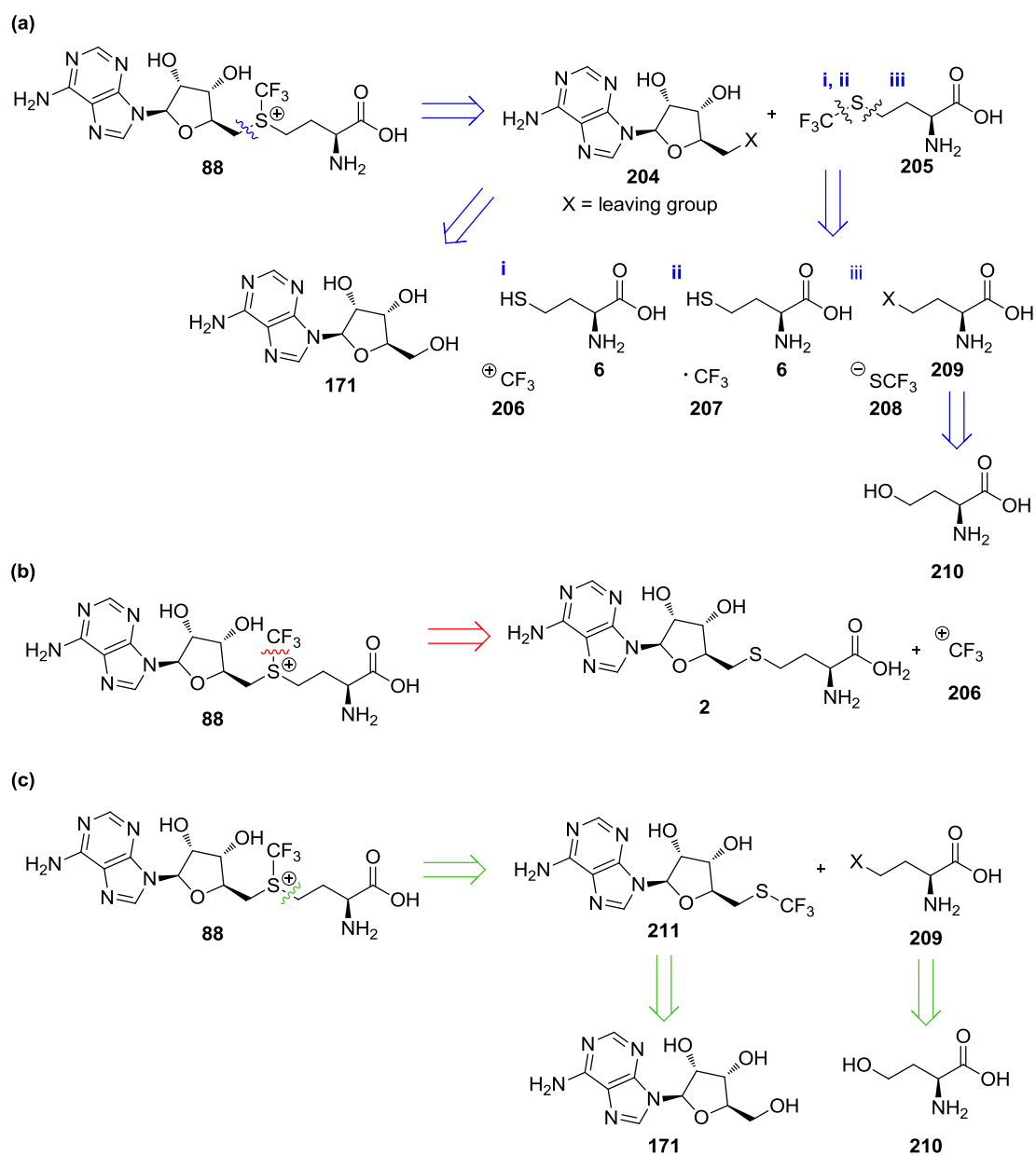
In order to develop a methodology for biocatalytic trifluoromethylation, a suitable cofactor was required as the donor of the trifluoromethyl group. As both NovO and CouO are SAM-dependent MTs, the trifluoromethylated SAM analogue **88** was proposed as a suitable cofactor for use in a biocatalytic trifluoromethylation reaction using NovO or CouO.



**Figure 95.** Proposed cofactor required for trifluoromethylation using NovO or CouO ('CF<sub>3</sub>-SAM').

A retrosynthetic analysis of CF<sub>3</sub>-SAM identified three potential disconnections around the sulfonium centre. The first (**Scheme 54**, disconnection **a**) corresponds to a S<sub>N</sub>2 reaction between a derivative of adenosine with a leaving group at the 5'- position (**204**) and a trifluoromethylated Met analogue (CF<sub>3</sub>-Met, **205**). It was hypothesised that CF<sub>3</sub>-Met (**205**)

could be synthesised either by electrophilic or radical trifluoromethylation of homocysteine (**6**) or alternatively, by nucleophilic trifluoromethylthiolation of an activated homoserine derivative (**209**) (**Scheme 55**, disconnections **i-iii**).

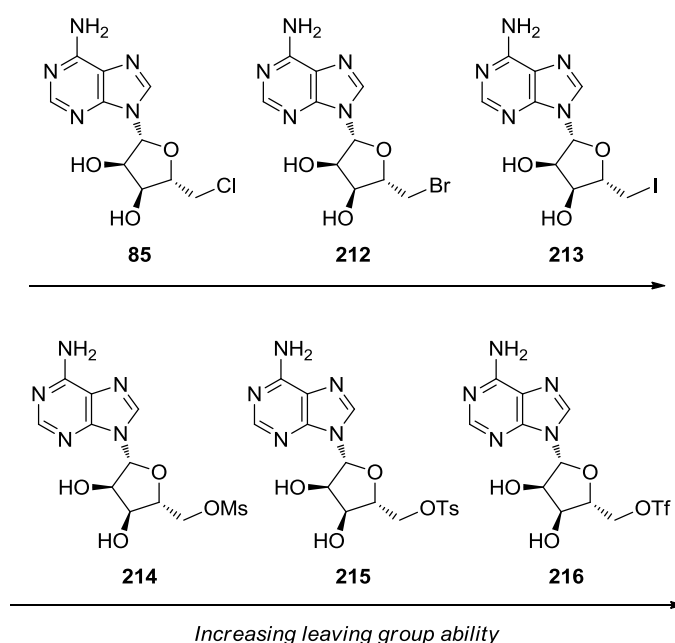


**Scheme 56.** Retrosynthetic analysis of  $\text{CF}_3$ -SAM (**88**). Initial disconnections are denoted by (a) (blue), (b) (red) and (c) (green) and disconnections for the synthesis of  $\text{CF}_3$ -Met (**205**) are denoted by i, ii and iii. X = leaving group.

Given the literature precedent for the enzymatic synthesis of Met analogues using methionine adenosyl transferases (MATs) or SaL (*Salinospora tropica*), it was proposed

that the coupling of **204** and **205** could be carried out chemically or enzymatically to provide the target molecule **88**.<sup>[224,310,311]</sup> The second disconnection (**Scheme 57b**) corresponds to the direct *S*-trifluoromethylation of SAH (**2**), which is commercially available. This route was interesting as it may be possible to recycle the SAH produced as a byproduct of the proposed trifluoromethyl transfer reaction. Finally, a third disconnection between a trifluoromethylthiolated adenosine derivative **211** and an activated homoserine derivative **209** was proposed (**Scheme 58c**). Due to the high cost of homoserine (**210**), this approach was less attractive and, unlike disconnection **a**, there is not an enzyme known that is capable of carrying out a similar reaction which could be carried out in tandem with methyl transfer or, unlike disconnection **b**, scope for the development of a SAH recycling system. Therefore, disconnections **a** and **b** were the subject of initial investigations.

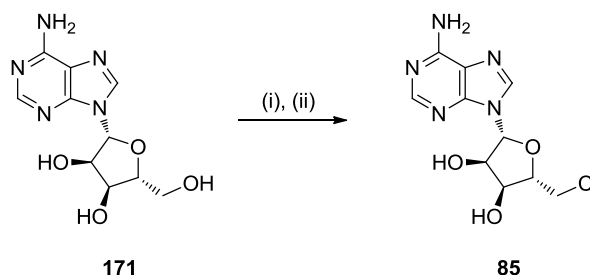
The synthesis of CF<sub>3</sub>-SAM was first attempted *via* disconnection **a**, for which two coupling partners were required: a derivative of adenosine with a leaving group at the 5'-position (**204**) and an *S*-trifluoromethylated analogue of methionine (**205**). To evaluate a range of reactivities of **204**, adenosine derivatives **85** and **212-216** were synthesised (**Figure 96**),



**Figure 96.** 5'-deoxy-5'-activated adenosine derivatives **85** and **212-216** in increasing order of leaving group ability.

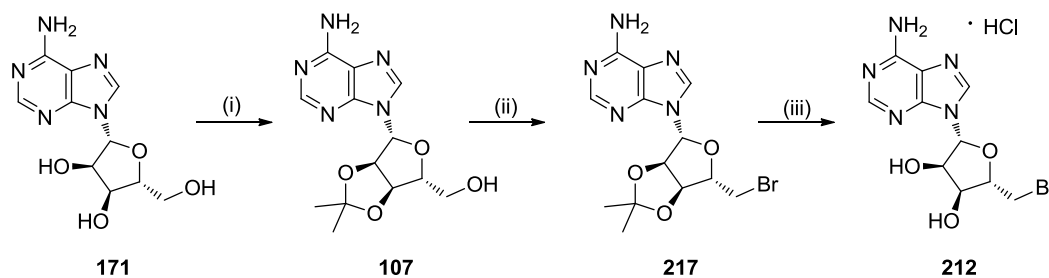
which display a gradient of increasing leaving group ability from 5'-chloro-5'-deoxyadenosine (CIDA, **85**) through to the triflate derivative **216**.

The synthesis of 5'-chloro-5'-deoxyadenosine (CIDA, **85**) is well preceded and was attempted first.<sup>[223]</sup> Using an excess of thionyl chloride under basic conditions and subsequent treatment with ammonium hydroxide to hydrolyse the intermediate sulfonyl ester, the desired product **85** was isolated in a 98% yield on gram scale (**Scheme 59**).



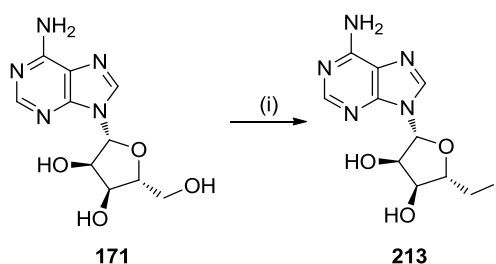
**Scheme 59.** Synthesis of 5'-chloro-5'-deoxyadenosine (CIDA, **50**). *Reagents and conditions:* (i) SOCl<sub>2</sub>, pyridine, MeCN, 0 °C, 3 h, then 22 °C, 16 h. (ii) NH<sub>4</sub>OH, methanol/H<sub>2</sub>O, 30 min, 98% from **171**.

In an analogous manner, the synthesis 5'-bromo-5'-deoxyadenosine (BrDA, **212**) was attempted starting from adenosine by treatment with thionyl bromide. Despite evidence for product formation by LC-MS, it was not possible to isolate any of the desired product. This was proposed to be due to the instability of **212** to the strong base used in the second step. Therefore, alternative conditions including the use of PPh<sub>3</sub>Br<sub>2</sub>, PPh<sub>3</sub>, *N*-bromosuccinimide (NBS) and tetrabutylammonium bromide and PPh<sub>3</sub> were also attempted. However, in all cases no product was isolated, either due to very low levels of conversion to product or degradation of the product during the work up. To improve the stability of the product to the work up conditions and to ensure a selective reaction of the primary over the secondary alcohol, the *syn*-diol in **171** was protected as the acetal to provide isopropylidene **107** in 87% yield (**Scheme 60**). This had the additional benefit of improved the solubility of the product in organic solvents. Acetal protected BrDA (**217**) was then prepared by treatment with PPh<sub>3</sub> and NBS under basic conditions to provide the desired product in a 33% yield. The low yield was thought to be due to the instability of the product upon purification by column chromatography. However, due to the presence of triphenylphosphine oxide in the reaction mixture, it was not possible to obtain **217** without this purification step. Finally, acid mediated acetal deprotection provided the target compound BrDA (**212**) as the hydrochloride salt in 96% yield (**Scheme 60**).



**Scheme 60.** Synthesis of BrDA from adenosine. *Reagents and conditions:* (i) Acetone,  $\text{HClO}_4$ , 0–22 °C, 3 h, 87%. (ii)  $\text{PPh}_3$ , *N*-bromosuccinimide (NBS),  $\text{BaCO}_3$ ,  $\text{CH}_2\text{Cl}_2$ , reflux, 2 h, 33%. (iii) aq. HCl, THF, 22 °C, 96%.

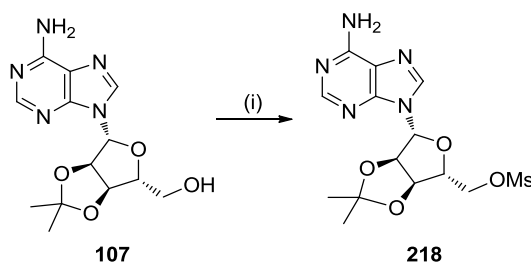
Next, 5'-iodo-5'-deoxyadenosine (IDA, **213**) was synthesised according to the procedure reported by Perrone *et al.*, by treatment of adenosine (**171**) with elemental iodine in the presence of  $\text{PPh}_3$  and pyridine (**Scheme 61**).<sup>[312]</sup> However, the product degraded during purification by flash chromatography in all attempts, resulting in low yields of up to 15%. Therefore, the material used in subsequent  $\text{S}_{\text{N}}2$  reactions was obtained from isolating the crude product by filtration and washing with an excess of acetone to remove the triphenylphosphine oxide and therefore contained a low level of impurities.



**Scheme 61.** Synthesis of IDA from adenosine. *Reagents and conditions:*  $\text{PPh}_3$ ,  $\text{I}_2$ , pyridine, 22 °C, 2 h, 15%.

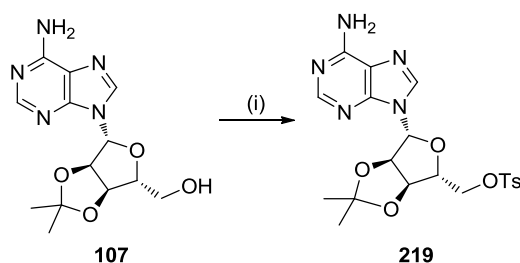
The synthesis of the mesylate, tosylate and triflate adenosine derivatives **214–216** was also attempted. Due to the very poor solubility of adenosine in organic solvents, the *syn*-diol protected acetal **107** was used. The adenosine acetonide mesylate derivative **218** was prepared by the treatment of **107** with mesyl chloride (MsCl) in the presence of triethylamine (**Scheme 62**). As the material was found to be unstable to silica gel chromatography, **214** was taken through into subsequent coupling experiments in its crude form.





**Scheme 62.** Synthesis of adenosine acetamide mesylate **218**. *Reagents and conditions:* (i) MsCl, NEt<sub>3</sub>, CH<sub>2</sub>Cl<sub>2</sub>, 0-22 °C, 30 min, *product used without purification*.

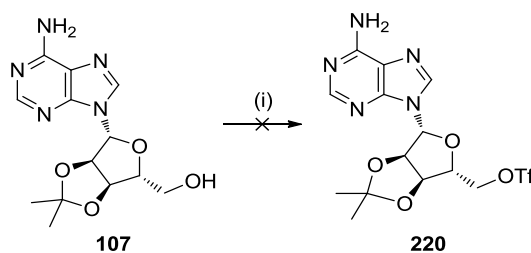
Tosylate adenosine derivative **219** was synthesised in an analogous fashion by the treatment of adenosine derivative **107** with tosyl chloride (TsCl) in the presence of triethylamine. As before, the material was found to be unstable to purification and was therefore also used in its crude form (**Scheme 63**).



**Scheme 64.** Synthesis of adenosine acetamide tosylate **219**. *Reagents and conditions:* (i) TsCl, NEt<sub>3</sub>, CH<sub>2</sub>Cl<sub>2</sub>, 0-22 °C, 90 min, *product used without purification*.

Finally, the synthesis of adenosine acetamide triflate **220** was attempted by treatment of **107** with triflic anhydride in the presence of triethylamine. Unfortunately, analysis of the reaction mixture by HPLC revealed a complex mixture of products and product isolation was not attempted (**Scheme 65**). However, as a range of alternative activated adenosine derivatives had been synthesised, it was decided not to pursue the synthesis of triflate **220** and instead to move forward with the other, more stable analogues.

Next, the synthesis of CF<sub>3</sub>-Met (**205**) was investigated. There is substantial literature precedent for the synthesis of CF<sub>3</sub>-Met, which is most commonly synthesised *via* the direct trifluoromethylation of homocysteine using CF<sub>3</sub>I and sodium in liquid ammonia. This was first reported in 1992 by Soloshonok *et al.* and later modified by Shibata *et al.* in 2011 to start from the corresponding amino acid dimer, homocystine.<sup>[313,314]</sup> However, this route is

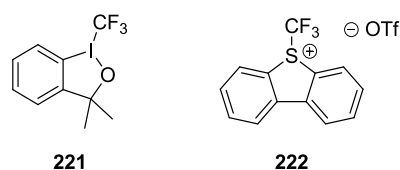


**Scheme 65.** Synthesis of adenosine triflate **216**. *Reagents and conditions:* (i)  $\text{ Tf}_2\text{O}$ ,  $\text{Net}_3$ ,  $\text{CH}_2\text{Cl}_2$ , 0–22 °C, 2 h, *product not isolated*.

operationally complicated due to low temperatures and the use of liquid ammonia, elemental sodium and  $\text{CF}_3\text{I}$ , which is a gas at room temperature. Therefore, alternative syntheses under milder conditions were investigated. As shown in **Scheme 56**, disconnections **i-iii**, it was hypothesised that the  $\text{CF}_3$  group could be introduced by electrophilic or radical trifluoromethylation, or nucleophilic trifluoromethylthiolation. To determine the most effective route towards these compounds, each of these three strategies was investigated.

### 5.1.2 Electrophilic trifluoromethylation

The development of mild, bench stable electrophilic trifluoromethylation reagents is a popular research area due the importance of the  $-\text{CF}_3$  group in pharmaceutically active compounds.<sup>[121]</sup> Two such reagents were selected for investigation based on their stability, cost and predicted reactivity. The first, **221**, is a hypervalent iodine reagent developed by Togni *et al.* and has been reported to be effective in the trifluoromethylation of thiols under mild reaction conditions.<sup>[315]</sup> The second, **222**, was developed by Umemoto *et al.* and has been reported to transfer a  $-\text{CF}_3$  group to alkyl and aryl thiolates from the sulfonium centre to afford *S*-trifluoromethylated products (**Figure 97**).<sup>[316]</sup>



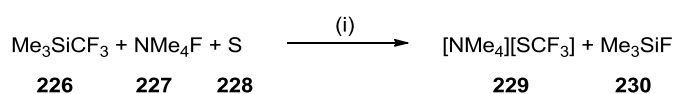
**Figure 97.** Structure of the two electrophilic trifluoromethylating agents investigated (Togni's reagent **221** and Umemoto's reagent **222**).

Due to the cost and limited availability of homocysteine (HCys), cysteine (Cys) was used for scoping reactions. As Cys differs from HCys only by a  $-\text{CH}_2-$  unit in the amino acid sidechain, they were predicted to have a very similar reactivity profile. Additionally, it was



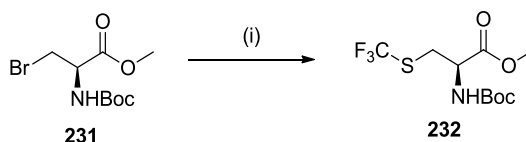
### 5.1.3 Nucleophilic trifluoromethylthiolation

An alternative approach to the synthesis of **205** is by nucleophilic trifluoromethylthiolation of an activated serine derivative. The synthesis and reactions of a number of trifluoromethylthiolate sources have been reported, both with transition metal counterions (eg. AgSCF<sub>3</sub> and CuSCF<sub>3</sub>) and organic counterions (eg. [NMe<sub>4</sub>][SCF<sub>3</sub>]).<sup>[317–319]</sup> Due to its facile synthesis, [NMe<sub>4</sub>][SCF<sub>3</sub>] (**229**) was chosen for scoping experiments. To this end, **229** was synthesised in 80% yield according to the reported procedure, using Rupert's reagent (**226**)<sup>[122]</sup> as the trifluoromethyl source in combination with elemental sulfur and tetramethylammonium fluoride (**227**) (**Scheme 66**).<sup>[317]</sup>



**Scheme 66.** Synthesis of trifluoromethylthiolate salt **229**. *Conditions:* THF, -78 – 22 °C, 16 h, 80%.

Initial experiments indicated that protection of the carboxylic acid was necessary to prevent quenching of the nucleophile by deprotonation of the free acid. Therefore, the commercially available *N*-Boc *O*-methyl ester protected serine derivative **231** was used for proof of concept of this strategy. On treatment with salt **229** in acetonitrile at room temperature, it was possible to obtain product **232** in 81% isolated yield (**Scheme 67**).

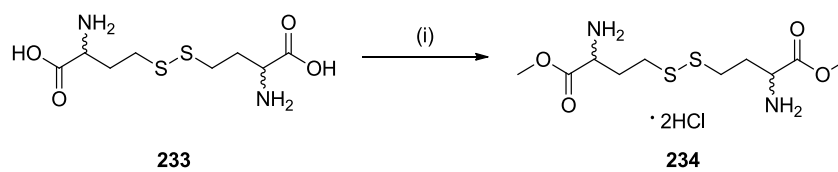


**Scheme 67.** Synthesis of *S*-trifluoromethylated Cys derivative **232** by nucleophilic trifluoromethylthiolation. *Reagents and conditions:* (i) ([NMe<sub>4</sub>][SCF<sub>3</sub>]), MeCN, 22 °C, 2 h, 81%.

Despite this positive result and proof of concept of a valid disconnection, the need for protecting groups and prefunctionalisation of the starting material to generate a leaving group, coupled with the very high cost of homoserine and its derivatives rendered this strategy less attractive than its alternatives. As such, no further development work was carried out for this route.

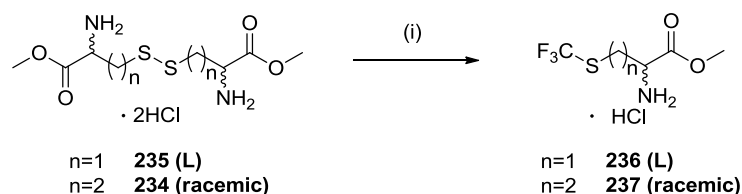
### 5.1.4 Radical trifluoromethylation

Radical trifluoromethylation of aromatic substrates is an attractive transformation as it offers a direct approach to trifluoromethylation without the need for prefunctionalisation of substrates and as such, is a very active area of research which has been discussed in Chapter 1.<sup>[120,134]</sup> The trifluoromethyl radical has also been employed for the synthesis of *S*-trifluoromethylated amino acids by the homolytic cleavage of a disulfide bond, starting from the respective dimer (cystine or homocystine) and was first reported by Langlois *et al.* in 1994.<sup>[130]</sup> The trifluoromethyl radical was generated from the decomposition of sodium trifluoromethyl sulfinate in the presence of *tert*-butylhydroperoxide (TBHP) to provide a small library of *S*-trifluoromethylated Cys and HCys derivatives in 48-56% yield, with the carboxylic acid protected as the ethyl ester in all cases.<sup>[130]</sup> To investigate this approach towards the synthesis of CF<sub>3</sub>-Met, first the literature results were reproduced. Whilst cystine dimethyl diester was commercially available, the dimethyl diester hydrochloride salt **234** was prepared in quantitative yield from homocystine (**233**) (**Scheme 68**).



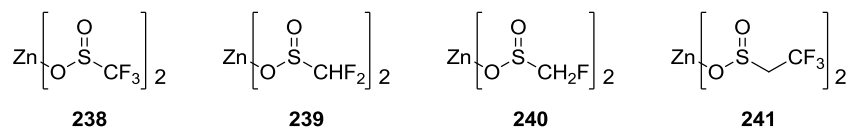
**Scheme 68.** Carboxylic acid protection of homocystine. *Reagents and conditions:* (i) SOCl<sub>2</sub>, MeOH, 22 °C, 16 h, 100%.

Both *O*-protected cystine and homocystine were subjected to the conditions reported by Langlois *et al.* to provide trifluoromethylated products **236** and **237**. In the case of cystine, the CF<sub>3</sub>-Met methyl ester (**95**) was isolated in a 21% yield (**Scheme 69**), whilst 11% conversion to **237** was observed in the case of homocystine. Due to the small scale of the reaction, isolation of the product was not attempted in this case. Although the reaction conditions and work up were unoptimised and therefore a low conversion or yield was obtained, this served as proof of principle to develop a methodology in which a fluorinated alkyl radical is used to cleave a dithiol bond to generate fluorinated methionine analogues. Furthermore, analysis of reaction samples by <sup>19</sup>F and <sup>1</sup>H NMR spectroscopy was found to be an effective method to monitor reaction progress.



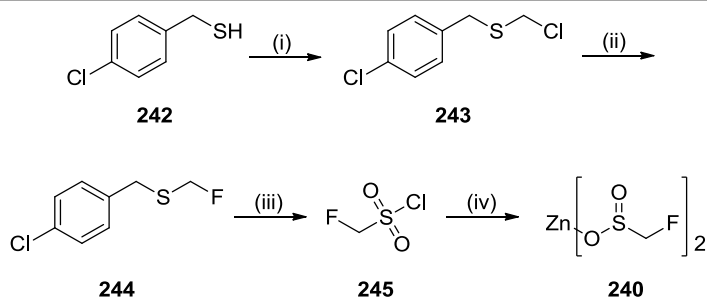
**Scheme 69.** Synthesis of  $\text{CF}_3$ -Met under Langlois conditions. *Reagents and conditions:* (i)  $\text{CF}_3\text{SO}_2\text{Na}$ , *tert*-BuOOH, water, 22 °C, 22 h, 21% isolated yield (**236**,  $n=1$ ), 11% conversion by  $^1\text{H}$  NMR spectroscopy (**237**,  $n=2$ ).

As the aim was to develop methodology for a library of fluorinated Met analogues, a method which could be applied to a range of fluoroalkyl radicals was sought after. To this end, Baran *et al.* have reported the use of zinc fluoroalkyl sulfinate salts **238-240** (**Figure 98**) as fluoroalkyl radical sources for the fluorination of heterocycles.<sup>[133,134,320,321]</sup> Therefore, the zinc sulfinate salts offered a distinct advantage over the corresponding sodium salt, as they enable the preparation of a library of fluorinated Met analogues. This would serve as a useful tool for probing the relative reactivity of fluorinated Met analogues with adenosine derivatives activated at the 5'-position in subsequent investigations into the synthesis of fluorinated SAM analogues.



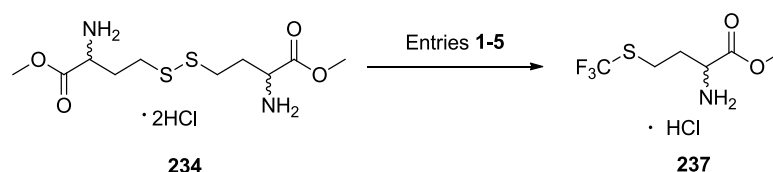
**Figure 98.** Fluoroalkylated zinc sulfinate salts developed by Baran *et al.*

Out of the compounds shown in **Figure 98**, all are commercially available with the exception of the monofluorinated analogue **240**. Therefore, this compound was synthesised according to the procedure reported by Baran and coworkers.<sup>[320]</sup> First, the commercially available thiol **242** was treated with paraformaldehyde and HCl gas (generated *in situ*) to provide thioether **243**. Next, the monofluoro- substituent was introduced *via* an  $\text{S}_{\text{N}}2$  reaction with potassium fluoride in the presence of dicyclohexano-18-crown-6, which has a high affinity for the chelation of  $\text{K}^+$ , thereby improving the solubility and nucleophilicity of the  $\text{Cl}^-$  ion.<sup>[322]</sup> Next, sulfonyl chloride **245** was prepared by oxidative chlorination with sodium hypochlorite and perchloric acid.<sup>[323]</sup> Finally, the zinc sulfinate salt **240** was prepared in 26% yield from **242** by treatment of sulfonyl chloride **245** with elemental zinc (**Scheme 70**). Due to the poor stability of compounds **243-245**, the synthesis was telescoped and only the final compound **240** was isolated and characterised.



**Scheme 70.** Synthesis of fluoromethyl zinc sulfinato salt **240**. *Reagents and conditions:* (i) paraformaldehyde,  $\text{HCl}_{(\text{g})}$  generated *in situ* from the addition of conc.  $\text{H}_2\text{SO}_4$  to  $\text{NaCl}$ ,  $\text{CH}_2\text{Cl}_2$ ,  $-15\text{ }^\circ\text{C}$ , 2 h. (ii)  $\text{KF}$ , dicyclohexano-18-crown-6,  $\text{MeCN}$ , reflux, 72 h. (iii)  $\text{HClO}$ ,  $\text{NaClO}$ ,  $\text{CH}_2\text{Cl}_2$ ,  $-10\text{ }^\circ\text{C}$ , 6 h. (iv)  $\text{Zn}$ , water,  $0\text{--}10\text{ }^\circ\text{C}$ , 3 h, 26% from **242**.

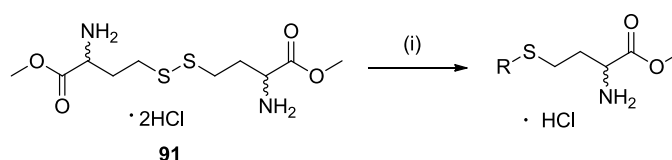
With a library of zinc sulfinato salts in hand, their use as fluoroalkyl radical precursors for the preparation of a library of *S*-fluoroalkylated amino acids was investigated. For proof of concept and reaction optimisation, the trifluoromethyl zinc sulfinato salt **238** was used. Initially, reaction conditions were chosen based on those reported by Baran *et al.* for the radical trifluoromethylation of heterocycles, with the difference of carrying the reactions out in water as opposed to an organic/aqueous mixture.<sup>[133,134,321]</sup> The trifluoromethylation of dimethyl diester **234** was attempted with 2 eq. zinc sulfinato salt **238** and 4.5 eq. TBHP, which was added dropwise to the reaction mixture over 5 minutes (**Table 20**, entry **1**) and gave 51% conversion to the desired product **237**. With proof of concept in hand, the effect of the order of addition, the peroxide used and solvent on percentage conversion was investigated (**Table 20**, entries **2-4**). Adding a solution of the salt to the reaction mixture containing the peroxide initiator (**Table 20**, entry **2**) gave a lower conversion of 40%. Secondly, hydrogen peroxide was investigated as an alternative radical initiator as it is readily available and decomposes to water soluble byproducts (**Table 20**, entry **3**). However, no conversion was observed after an extended reaction time. The need for a protic solvent was then investigated by carrying the reaction out in DMF, a polar aprotic solvent. In this instance, no conversion to product was observed after an extended reaction time (**Table 20**, entry **4**). Considering the poor solubility of the product in organic solvents, water was concluded to be the best solvent for this reaction. Finally, the number of equivalents of the zinc sulfinato salt **238** was investigated. Following consumption of the first 2 equivalents of salt (as monitored by  $^{19}\text{F}$  NMR spectroscopy), a further portion of 2 eq. was added along with a further addition of 4.5 eq. peroxide (**Table 20**, entry **5**). In this way, it was possible to achieve quantitative conversion to the desired product.



Entry	Solvent	Eq. 238	Peroxide	Eq. Peroxide	Addition order	% Conversion <sup>a</sup>
1	water	2	TBHP	4.5	peroxide to reaction	51
2	water	2	TBHP	4.5	<b>238</b> to reaction	40
3	water	2	H <sub>2</sub> O <sub>2</sub>	4.5	peroxide to reaction	0
4	DMF	2	TBHP	4.5	peroxide to reaction	0
5	water	2 x 2 eq.	TBHP	2 x 4.5 eq	peroxide to reaction	100

**Table 20.** Conditions screening for synthesis of **237** by radical trifluoromethylation with zinc sulfinate salts. <sup>a</sup>% conversion determined by <sup>1</sup>H NMR spectroscopy following consumption of zinc sulfinate salt as monitored by <sup>19</sup>F NMR spectroscopy. Peroxide added at a rate of 1.7 mL/h. Reactions all carried out at 20 °C.

Next, the reaction was carried out on preparative scale (200 mg starting material) under the optimised reaction conditions to provide the CF<sub>3</sub>-Met methyl ester **237** in 48% isolated yield. The procedure was also carried out using the other zinc sulfinate salts **239-241** to provide the corresponding *S*-fluoroalkylated amino acids. This was successful for the -CF<sub>2</sub>H and CH<sub>2</sub>CF<sub>3</sub> analogues and provided the desired compounds **241** and **242** in 56% and 68% isolated yield, respectively. Conversely, no product formation of the -CH<sub>2</sub>F analogue **243** was detected by <sup>19</sup>F NMR and no product was isolated (**Table 21**). Interestingly, Baran and coworkers also



Entry	No.	R	Yield <sup>a</sup> (%)
1	<b>237</b>	-CF <sub>3</sub>	48
2	<b>241</b>	-CF <sub>2</sub> H	56
3	<b>242</b>	-CH <sub>2</sub> CF <sub>3</sub>	68
4	<b>243</b>	-CH <sub>2</sub> F	-

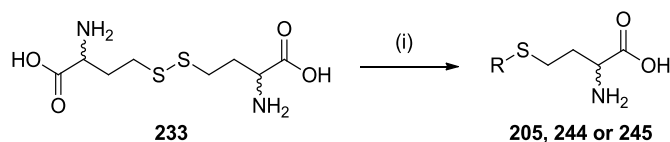
**Table 21.** Synthesis of CF<sub>3</sub>-Met and analogues by radical trifluoromethylation from dimethyl diester **91** with zinc sulfinate salts. *Reagents and conditions:* (i) (RSO<sub>2</sub>)<sub>2</sub>Zn, *tert*-butylhydroperoxide, water, 22 °C, 4 h. <sup>a</sup>Isolated yield.



did not observe any reaction between zinc sulfinate salt **240** and a number of their heterocyclic substrates, although the authors did not speculate on the reason for this.<sup>[320]</sup>

This work demonstrated the use fluoroalkylated zinc sulfinated salts as an effective method for the synthesis of fluorinated Met analogues. Furthermore, the procedure provided the products in excellent purity, removing the need for difficult and time consuming purification of amino acids. In comparison to the methods for electrophilic trifluoromethylation, this methodology operates under very mild conditions and was concluded to be a preferable strategy for the synthesis of a library of fluorinated Met analogues.

Encouraged by these results, the next objective was to determine whether this methodology could be extended to the synthesis of unprotected amino acids, starting from the disulfide linked dimer **233**. To this end, **233** was treated with 2 eq. of zinc sulfinate salt **238**, with the slow addition of 4 eq. TBHP *via* syringe pump. Gratifyingly, this provided the desired product **205** in 86% isolated yield (**Table 22**, entry 1). Moreover, using zinc sulfinate salts **239** and **241**, it was possible to isolate the difluoromethyl and trifluoroethyl Met analogues **244** and **245** in 99% and 95% isolated yield, respectively. As quantitative conversion was observed by <sup>1</sup>H NMR in each case, the variation in yield is thought to be due to loss of product during work up.



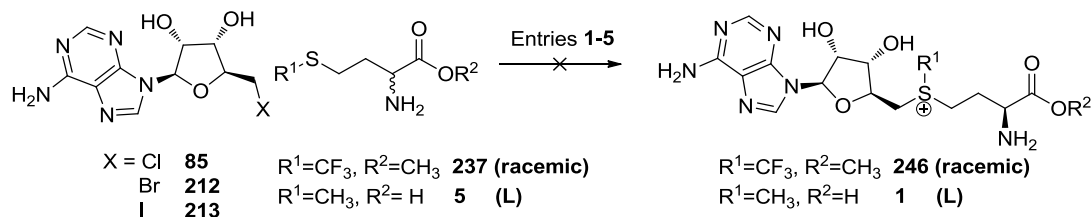
Entry	No.	R	Yield <sup>a</sup> (%)
1	205	-CF <sub>3</sub>	86
2	244	-CF <sub>2</sub> H	99
3	245	-CH <sub>2</sub> CF <sub>3</sub>	95

**Table 22.** Synthesis of CF<sub>3</sub>-Met and analogues by radical trifluoromethylation of free diacid **233** with zinc sulfinate salts. *Reagents and conditions:* (i) (RSO<sub>2</sub>)<sub>2</sub>Zn, *tert*-butylhydroperoxide, water, 22 °C, 4 h. <sup>a</sup>Isolated yield.

Overall, radical fluoroalkylation using zinc sulfinate salt proved to be an effective methodology for the synthesis of a library of fluorinated Met analogues. Moreover, not only does this strategy operate with the homocysteine methyl ester, but is also compatible with the free amino acid, avoiding the need for extra protection and deprotection steps.

## 5.1.5 Attempts towards the S<sub>N</sub>2 reaction between an activated adenosine derivative and fluorinated Met analogue

The synthesis of sulfonium compounds by the nucleophilic attack of a thioether at an electrophilic centre is widely reported.<sup>[324,325]</sup> Therefore, this was proposed as a plausible approach to synthesise fluoroalkylated SAM analogues. To investigate this route, the synthesis of CF<sub>3</sub>-SAM was attempted *via* the reaction of CF<sub>3</sub>-Met methyl ester **237** with the activated adenosine derivatives **85**, **212** and **213** in the presence of silver triflate. This was proposed to provide a triflate counterion to the resulting sulfonium centre and promote the S<sub>N</sub>2 reaction by precipitation of the corresponding silver halide salt. Initially, this reaction was attempted in formic acid at room temperature (**Table 23**, entry 1). These conditions were selected based on the conditions used by Gruber and coworkers to synthesise non-natural *S*-alkylated Met analogues from the direct alkylation of SAH.<sup>[107]</sup> However, analysis of the reaction mixture by LC-MS, HPLC and <sup>19</sup>F NMR spectroscopy showed only starting materials to be present in all three cases (CIDA **85**, BrDA **212** and IDA **213**). The reactions were next attempted in DMF and DMSO (polar aprotic solvents) in which both starting

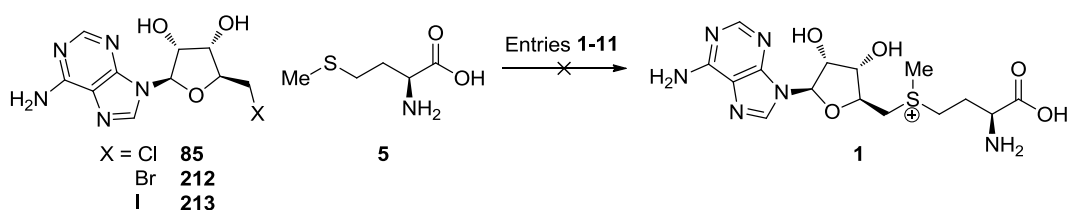


Entry	Reagents	Temp. (°C)	Solvent	% Conversion <sup>a</sup>
1	<b>85/212/213</b> , Ag(OTf) <sub>2</sub>	22	formic acid	-
2	<b>85/212/213</b> , Ag(OTf) <sub>2</sub>	22 - 50 <sup>b</sup>	DMF	-
3	<b>85/212/213</b> , Ag(OTf) <sub>2</sub>	22 - 50 <sup>b</sup>	DMSO	-
4	<b>85/212/213</b>	22 - 50 <sup>b</sup>	DMF	-
5	<b>85/212/213</b>	22 - 50 <sup>b</sup>	DMSO	-

**Table 24.** Conditions screened for the synthesis of CF<sub>3</sub>-SAM and SAM by S<sub>N</sub>2 reaction between CF<sub>3</sub>-Met or Met and activated adenosine derivatives, CIDA, BrDA and IDA (**85**, **212** and **213**, respectively). <sup>a</sup>Conversion determined by LC-MS, HPLC and <sup>19</sup>F NMR spectroscopy. <sup>b</sup>Reactions stirred at 22 °C for 3 h prior to heating to the specified temperature *via* microwave irradiation for 3 h. DMF: *N,N*-dimethylformamide. DMSO: dimethylsulfoxide.

materials were soluble (**Table 25**, entries **2** and **3**). After no product formation was observed after 3 hours at room temperature, the reactions were heated to 50 °C for 3 hours. However, after this time it was still not possible to detect the formation of **88**. Finally, the reaction was also attempted in the absence of Ag(OTf)<sub>2</sub> in both DMF and DMSO (**Table 26**, entries **4** and **5**) but once more, no product formation was observed and no product was isolated.

Due to the extreme electronegativity of the trifluoromethyl group, it was hypothesised that there may not be sufficient electron density on the sulfur atom to carry out the S<sub>N</sub>2 reaction. Whilst a library of *S*-fluoroalkylated alkylated had been prepared to investigate this effect, the analogous reaction with Met was investigated first, as Met was predicted to have a much more electron rich sulfur centre compared to its fluorinated analogues due

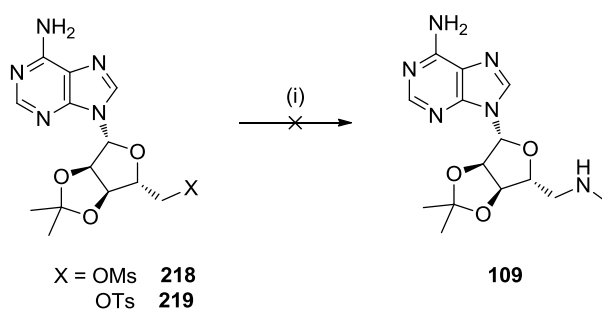


Entry	Reagents	Temp. (°C)	Solvent	% Conversion
<b>1</b>	<b>85</b>	22	water	-
<b>2</b>	<b>85</b>	22	formic acid	-
<b>3</b>	<b>85</b>	22	acetonitrile	-
<b>4</b>	<b>85</b>	22	acetone	-
<b>5</b>	<b>85</b>	22	DMSO	-
<b>6</b>	<b>85</b>	22	DMF	-
<b>7</b>	<b>85</b> , AgOTf, K <sub>2</sub> CO <sub>3</sub>	22	water	-
<b>8</b>	<b>85</b> , AgOTf, K <sub>2</sub> CO <sub>3</sub>	22	DMSO	-
<b>9</b>	<b>85</b> , AgOTf, K <sub>2</sub> CO <sub>3</sub>	22	acetonitrile	-
<b>10</b>	<b>85</b> , AgOTf, K <sub>2</sub> CO <sub>3</sub>	22	DMF	-
<b>11</b>	<b>85/ 212/ 213</b> , AgOTf	22 - 80 <sup>b</sup>	formic acid	-

**Table 27.** Conditions screened for the synthesis of SAM by S<sub>N</sub>2 reaction between Met and activated adenosine derivatives, ClDA, BrDA and IDA (**85**, **212** and **213**, respectively). <sup>a</sup>Conversion determined by LC-MS and HPLC. <sup>b</sup>Reactions stirred at 22 °C for 3 h prior to heating to the specified temperature *via* microwave irradiation for 3-4 h. DMF: *N,N*-dimethylformamide. DMSO: dimethylsulfoxide.

to a weak inductive effect from the *S*-methyl substituent. CIDA (**85**) was chosen for initial screening reactions due to relative ease of synthesis and availability. A range of solvents were screened, as were both basic and acidic conditions, with and without the addition of Ag(OTf)<sub>2</sub> (**Table 27**, entries **1-10**). However, no product formation was observed under any of these conditions. Finally, a full screen of adenosine derivatives **85**, **212** and **213** was carried out and the reactions heated to 80 °C for 4 hours. Again, only starting material was observed and no product was isolated.

To further investigate whether it was possible to carry out an S<sub>N</sub>2 reaction at this centre, the reaction between the *syn*-diol protected adenosine -mesylate and -tosylate derivatives **218** and **219** and methylamine was attempted. This reaction was predicted to be more favourable than using methionine as the nucleophile due to the higher stability of the secondary amine product **109** compared to that of the sulfonium product **88** (CF<sub>3</sub>-SAM). Furthermore, methylamine is a small, unhindered nucleophile and was predicted to be a better nucleophile for this reaction. However, after an extended reaction time at reflux, there was no evidence for product formation in any of the three reactions (**Scheme 71**).



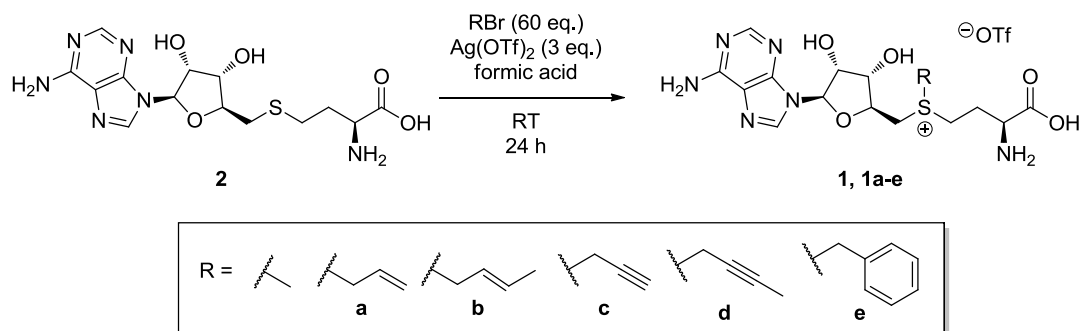
**Scheme 72.** S<sub>N</sub>2 reaction between adenosine derivatives **218** and **219** and methylamine. *Reagents and conditions:* (i) Methylamine (3 eq.), THF, 22 °C –reflux, 20 h.

It was concluded that adenosine derivatives activated at the 5'-position are poor electrophiles for S<sub>N</sub>2 reactions of this type and therefore this route towards the synthesis of *S*-fluorinated SAM analogues was not considered to be viable.

## 5.2 Attempts towards the direct trifluoroalkylation of SAH

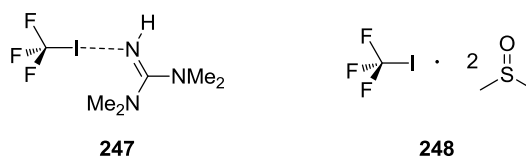
As all attempts at the synthesis of an *S*-fluorinated SAM analogue *via* an S<sub>2</sub>N reaction between an activated adenosine derivative and fluorinated Met analogue had been unsuccessful, an alternative approach was required. Based on the initial retrosynthetic analysis of CF<sub>3</sub>-SAM, the direct trifluoromethylation of SAH using an electrophilic

trifluoromethylation reagent was investigated. This disconnection had some literature precedent, as Gruber and coworkers have disclosed the *S*-alkylation of SAH using an alkyl bromide in the presence of AgOTf in neat formic acid to prepare SAM analogues **1a-e** (**Scheme 73**).<sup>[107]</sup>



**Scheme 73.** Synthesis of *S*-alkylated SAM analogues reported by Gruber and coworkers. Quantitative conversion reported in 24 h.

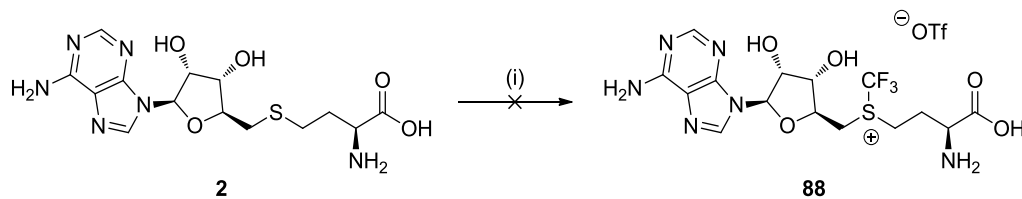
Owing to the poor availability, toxicity and very low boiling point of trifluorobromomethane ( $\text{CF}_3\text{Br}$ ), trifluoroiodomethane ( $\text{CF}_3\text{I}$ ) was selected as a more suitable trifluoromethylated alkyl halide for use in the synthesis of  $\text{CF}_3$ -SAM. Moreover, Ritter *et al.* have recently reported condensed phase adducts of  $\text{CF}_3\text{I}$  with tetramethylguanidine (TMG) or DMSO *via* halogen bonding (**Figure 99**) which are stable at room temperature.<sup>[126]</sup> These reagents are an attractive alternative to  $\text{CF}_3\text{I}_{(\text{g})}$  due to their relative ease of handling and are reported to display similar reactivity to  $\text{CF}_3\text{I}$ .<sup>[126]</sup>



**Figure 99.** Bench-stable  $\text{CF}_3\text{I}$ -TMG (**247**) and  $\text{CF}_3\text{I}$ -2DMSO (**248**) adducts developed by Ritter and coworkers.<sup>[126]</sup>

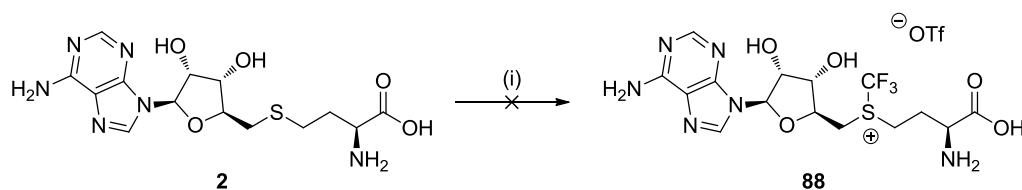
Therefore, the direct *S*-trifluoromethylation of SAH (**2**) was attempted using the  $\text{CF}_3\text{I} \cdot 2\text{DMSO}$  adduct (**248**), which was commercially available at the time of experimentation. Initially, the reaction was attempted using 5 eq. of **248** in the presence of 3 eq. AgOTf (**Scheme 74**). However, after an extended reaction time of 16 hours, no evidence for product formation was observed by LC-MS. Therefore, the reaction was repeated with 60 eq. **248**, mimicking the conditions reported by Gruber and coworkers

(Scheme 73). However, it was still not possible to detect formation of the desired product **88** and therefore no further development of this reaction was carried out.



**Scheme 74.** Direct trifluoromethylation of SAH (**2**) using  $\text{CF}_3\text{I}\cdot 2\text{DMSO}$  (**248**). Reagents and conditions: (i)  $\text{CF}_3\text{I}\cdot 2\text{DMSO}$  (5 eq. and 60 eq.), AgOTf (3 eq.), 22 °C, 16 h, no product formation observed by LC-MS.

As this had been unsuccessful, electrophilic trifluoromethylating agents **221** and **222** were also screened for reaction with SAH to synthesise  $\text{CF}_3$ -SAM. Two Lewis acids, AgOTf and  $\text{Cu}(\text{OTf})_2$ , were screened, however, after an extended reaction time of 3 days, no evidence for formation of **88** was observed by LCMS and therefore no product was isolated (Table 28).

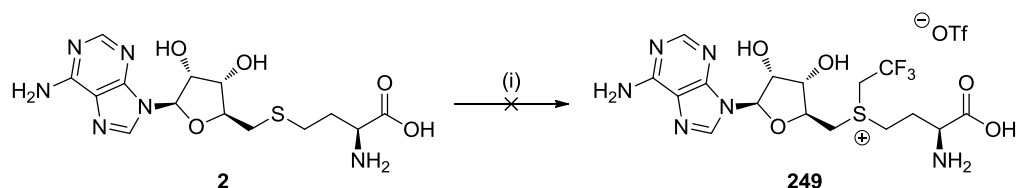


Entry	Reaction	Conditions	Conversion (%)
1	SAH + <b>221</b>	$\text{Cu}(\text{OTf})_2$ , MeOH/ DMSO (6:1), 22 °C, 3 days	-
2	SAH + <b>221</b>	AgOTf, MeOH/ DMSO (6:1), 22 °C, 3 days	-
3	SAH + <b>222</b>	$\text{Cu}(\text{OTf})_2$ , MeOH/ DMSO (6:1), 22 °C, 3 days	-
4	SAH + <b>222</b>	AgOTf, MeOH/ DMSO (6:1), 22 °C, 3 days	-

**Table 28.** Electrophilic trifluoromethylation of SAH using **36** and **37**.

Based on these results, it was hypothesised that  $\text{CF}_3$ -SAM could be unstable under reaction, analysis or work-up conditions. With this in mind, it was hypothesised that the *S*-trifluoroethyl analogue of SAM (**249**) may be more stable as the highly electronegative  $-\text{CF}_3$  centre would be further from the sulfonium centre. Therefore, the trifluoroethylation of SAH under analogous conditions to those reported by Gruber and coworkers (Scheme 73) using both trifluoroethyl iodide and trifluoroethyl bromide as the alkylating agent (Scheme 75) was attempted. However, after an extended reaction time of 48 hours, no product formation

was observed by HPLC, LC-MS or  $^{19}\text{F}$  NMR spectroscopy and the reactions were terminated.



**Scheme 75.** Direct *S*-trifluoroethylation of SAH using an alkyl halide. *Reagents and conditions:*  $\text{CF}_3\text{CH}_2\text{I}$  or  $\text{CF}_3\text{CH}_2\text{Br}$ ,  $\text{Ag}(\text{OTf})_2$ , formic acid, 48 h, *product not isolated*.

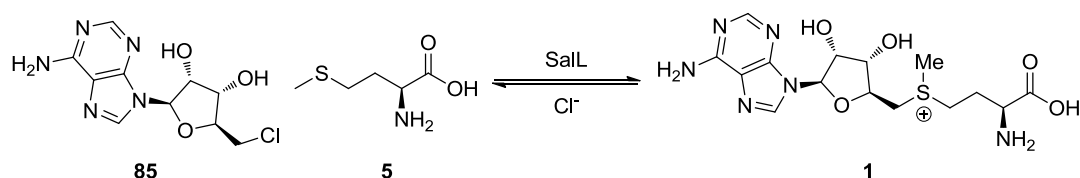
Having systematically investigated a number of syntheses towards an *S*-fluoroalkylated SAM analogue without success, it was decided not to further pursue this further *via* chemical means. If such cofactors are not stable under reaction, analysis or work up conditions, it was hypothesised that an *in situ* approach may be more appropriate, as a long half life would not be required. With this in mind, the development of an *in situ* cofactor production system was pursued, which has been demonstrated to be well suited to enzymatic strategies.<sup>[181,224,311]</sup>

### 5.3 Enzymatic strategies for the synthesis of SAM and SAM analogues

Enzymatic strategies for synthesis of SAM and *S*-alkylated SAM analogues are well precedented and have been discussed in detail in Chapter 1. An enzymatic approach to the synthesis of SAM and SAM analogues was attractive for this project due to the mild reaction conditions and the use of low cost or readily prepared starting materials. Furthermore, it may be possible to couple cofactor synthesis with alkyl transfer using NovO or CouO *via* an *in situ* cofactor production strategy. This would eliminate the need for handling of SAM and SAM analogues, which are very hygroscopic and readily degrade to a number of by-products, some of which are known inhibitors of MTs.<sup>[242,326]</sup>

Two approaches to enzymatic SAM and SAM analogue synthesis have been reported. The first utilises methionine adenosyl transferases (MATs) to synthesise SAM from ATP and methionine.<sup>[210]</sup> MATs are ubiquitous in Nature and are responsible for SAM synthesis *in vivo* and have been a popular approach to the *in vitro* enzymatic synthesis of SAM and SAM analogues. For example, Singh and coworkers have disclosed a tandem process for the *O*-alkylation of indocarbazoles *via* the use of engineered variants of methionine adenosyl transferases (MAT) for SAM analogue production, followed by the *O*-MT RebM to effect *O*-

alkylation.<sup>[144,212,219,327]</sup> However, a limitation of this method is the high cost of ATP for SAM production. Therefore, the second alternative was considered attractive, which utilises 5'-deoxy-5-chloroadenosine (CIDA, **85**) and Met to form SAM and is catalyzed by the chlorinase SaL from the marine organism *Salinospora tropica*.<sup>[221,328]</sup> In the marine environment, SaL degrades SAM to produce **85** as part of the biosynthesis of the anticancer agent salinosporamide A. However, under low chloride ion concentrations, the reaction equilibrium is biased towards the generation of SAM (**1**) (Scheme 76).<sup>[181,224,329]</sup>

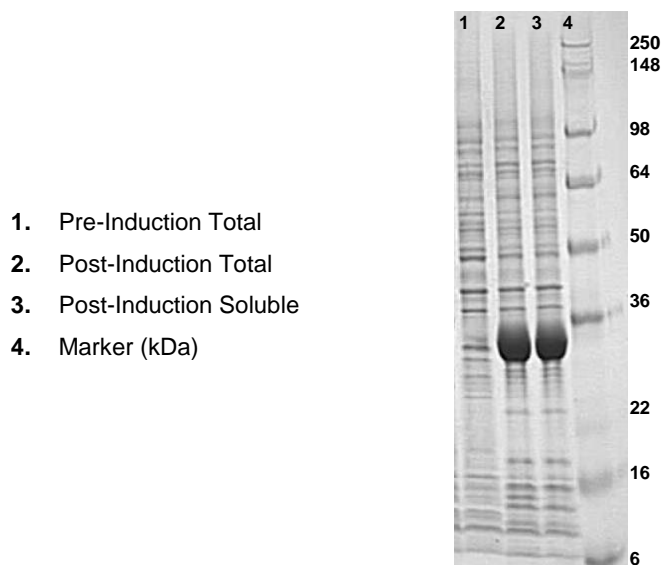


**Scheme 76.** Synthesis of SAM catalysed by SaL from *Salinospora tropica*.

As CIDA can be readily prepared from adenosine in a one-pot process, the use of SaL offers an inexpensive alternative to the use of MATs for *in situ* SAM synthesis.<sup>[181]</sup> With this in mind, the next stage was to investigate whether SaL could be utilised in tandem with NovO or CouO in order to develop a one-pot methodology for the C-alkylation of aminocoumarin and dihydroxynaphthalene substrates from readily available, inexpensive starting materials.

To this end, the gene for SaL from *Salinospora tropica* was codon optimised for *E. coli*, synthesized and sub-cloned into a pET26b(+) vector with an N-terminal 6xHis-tag at GenScript™. The construct was transformed into *E. coli* BL21 (DE3) and overexpression of the gene was carried out using Overnight Express Media, growing at 37 °C to an OD of ~2 and then at 18 °C overnight. Analysis by SDS PAGE revealed high levels of expression. Furthermore, the soluble fraction contained high levels of the target protein (**Figure 100**) indicating that these conditions were effective for the expression of SaL.

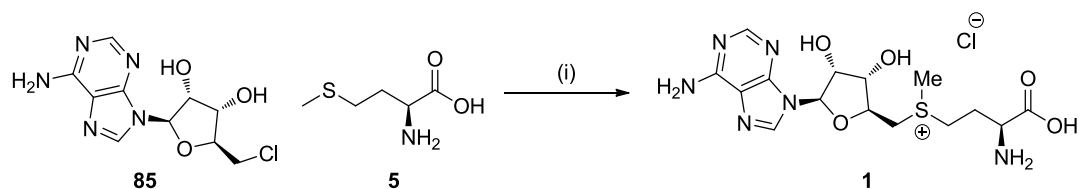




**Figure 100.** Analysis of heterologous expression of SalL in *E. coli* BL21 (DE3) expressed in Overnight Express media. Cells were grown to an OD of 1.9 at 37 °C then at 18 °C overnight. Theoretical mass of 6xHis-SalL: 30.1 kDa.

With a view to the development of an operationally simple platform for *in situ* SAM synthesis, a process using *E. Coli* cell free extract containing SalL for SAM synthesis was preferred to the use of purified enzyme. As such, the cell pellet harbouring SalL was resuspended in phosphate buffer and the cells lysed by sonication. The lysate was clarified by centrifugation and the resulting supernatant used for biotransformations.

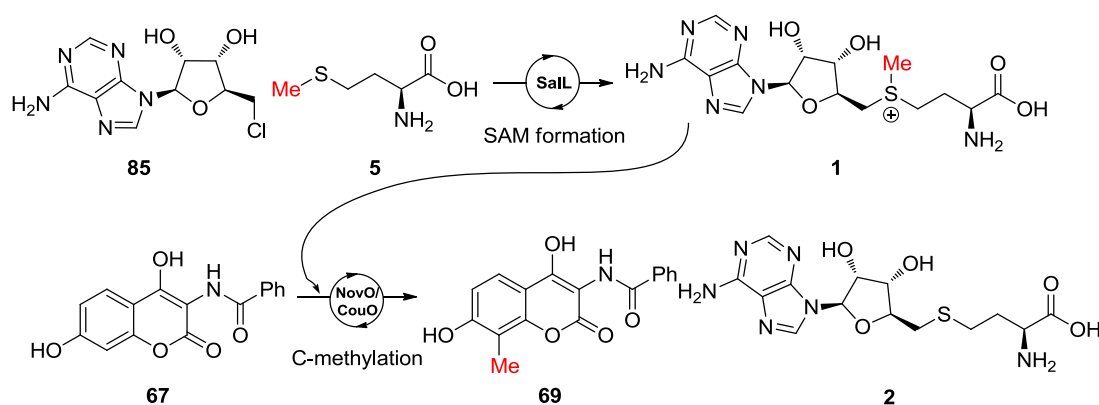
To confirm that the previously overexpressed SalL was active, the assay for SAM synthesis reported by Burkart *et al.* was repeated using the cell free extract containing SalL and analysed by HPLC and LCMS to confirm SAM formation. Gratifyingly, quantitative conversion to SAM was observed in one hour under these conditions (**Scheme 77**). This reaction was carried out on analytical scale in order to assess the feasibility of this approach and as such, isolation of SAM was not attempted.



**Scheme 77.** Synthesis of SAM using *E. coli* cell free extract containing SalL. Reagents and conditions: (i) *E. coli* cell free extract harbouring SalL, 37 °C, 1 h, 100% conversion by HPLC.

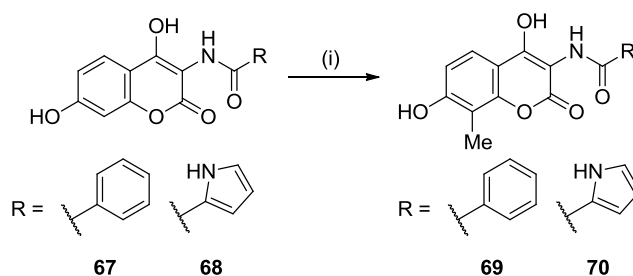
## 5.4 Tandem SAM synthesis and substrate alkylation

Having demonstrated that the *E. coli* cell free extract containing SalL was active for the synthesis of SAM, the next objective was to incorporate this into a tandem SAM synthesis and methylation process by coupling SalL with a C-MT in a one-pot system. It was hypothesised that *E. coli* cell free extract harbouring SalL could be utilised to prepare SAM *in situ* from CIDA (**85**) and Met (**5**). Next, NovO or CouO would be used to catalyse C-methylation of a suitable substrate, such as aminocoumarin **67**, using the freshly prepared SAM as the corresponding methyl donor (**Scheme 78**). This concept has previously been demonstrated by Burkart and coworkers, who coupled SalL catalysed SAM synthesis to the DNA MT HhaI.<sup>[224]</sup> However, this concept had not been demonstrated on small molecule C-MTs at preparative scale. Such a process would eliminate the need to add commercial SAM (which is typically ~80% pure) to the reaction, which is both expensive and unstable and it is therefore difficult to precisely control the amount that is added to a reaction.<sup>[310]</sup>



**Scheme 78.** Proposed one-pot tandem SAM synthesis and methyl transfer system using SalL and NovO or CouO.

The first stage in developing this process was to determine whether SalL and CouO or NovO were active under the same reaction conditions. To this end, *E. coli* cell free extract harbouring SalL was combined with *E. coli* cell lysate harbouring NovO or CouO and one of aminocoumarin substrates **67** or **68**. For proof of concept experiments, a large excess of Met (**5**, 50 eq.) and CIDA (**85**, 5 eq.) were added for SAM synthesis. Additionally, BSA and DTT were added to aid protein stability. In addition to screening the aminocoumarin substrates **67** and **68** with NovO or CouO and SalL, a reaction was run in the absence of SalL as a negative control (**Table 29**). The reactions were incubated at 37 °C for 24 hours and the conversion to methylated products was analysed by HPLC. Gratifyingly, in the presence of both NovO or CouO and SalL, 95-97% conversion to methylated products **69** and **70** was observed after 24



Entry	Substrate	Enzymes	% Conversion <sup>a</sup>
1	67	SalL, CouO	96
2	67	SalL, NovO	95
3	68	SalL, CouO	97
4	68	SalL, NovO	96
5	67	NovO	0

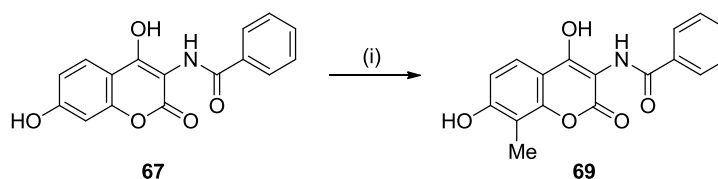
**Table 29.** One-pot tandem methylation reaction using SalL and CouO or NovO. *Reagents and conditions:* (i) L-Met (**5**, 50 eq.), CIDA (**85**, 5 eq.), *E. coli* cell free extract containing SalL, *E. coli* cell lysate containing CouO or NovO, DTT, BSA, potassium phosphate buffer (50 mM, pH 6.8), 35 °C. <sup>a</sup>% conversion measured by area % by HPLC after 24 h.

hours (**Table 29**, entries **1-4**). In the absence of SalL, no conversion to product was observed (**Table 29**, entry **5**). This negative result confirmed the need for SalL in the system to catalyse SAM synthesis.

With proof of concept in hand, the reaction conditions were optimised in order to develop this into an efficient and cost effective process. In particular, it was desirable to minimise the number of equivalents of Met and CIDA, as this would improve the atom efficiency of the process.<sup>[330]</sup>

In the initial experiments, 5 eq. CIDA and 50 eq. of amino acid were used (**Table 29**). These values were based on the literature precedent for the enzymatic synthesis of Met analogues using SalL.<sup>[224,310]</sup> To minimise this, the number of equivalents of CIDA and Met necessary to achieve high levels of conversion were investigated sequentially (**Table 30**). The reaction was found to proceed with quantitative conversion using 2 eq. CIDA and 2 eq. Met, which would form 2 eq. of SAM *in situ*, assuming quantitative conversion (**Table 30** entry **10**). However, ~30% conversion to the methylated product was also observed in the negative control reaction in which no Met was added to the system (**Table 30** entry **12**). Conversely, no conversion was observed when the addition of CIDA was omitted (**Table 30** entry **13**).

Based on these experiments, it was hypothesised that L-Met was present in the system from the cell lysate and cell free extract. In the presence of CIDA and SalL, this was being converted to SAM which was causing background methylation catalysed by NovO or CouO.



Entry	Eq. CIDA	Eq. Met	% Conversion CouO	% Conversion NovO
1	5	50	100	100
2	2.5	50	100	100
3	2	50	100	100
4	1.5	50	92	95
5	1	50	69	72
6	2	50	100	100
7	2	25	100	100
8	2	10	100	100
9	2	5	100	100
10	2	2	99	100
11	2	1	78	77
12	2	0	28	28
13	0	2	0	0

**Table 30.** Optimisation of CIDA and amino acids equivalents. Conversion calculated by area/area% in HPLC chromatogram after 24 h. *Reagents and conditions:* (i) CIDA (**85**), Met (**5**), BSA, DTT, *E. coli* cell lysate containing CouO or NovO in 50 mM phosphate buffer, pH 7 (CouO) and pH 6.5 (NovO) diluted at 10 mL/g pellet, *E. coli* cell free extract containing SalL in 50 mM phosphate buffer, pH 6.8 diluted at 5 mL/g pellet. Optimised conditions highlighted in grey.

The long term goal of this project was to develop a platform for biocatalytic C-alkylation, which would include the transfer of alkyl groups other than methyl. However, Clausen and coworkers have previously reported a drastic decrease in enzyme activity of SalL with S-alkylated Met analogues compared to with Met.<sup>[181]</sup> Therefore, it was hypothesised that background Met in the system would outcompete the formation of non-natural SAM

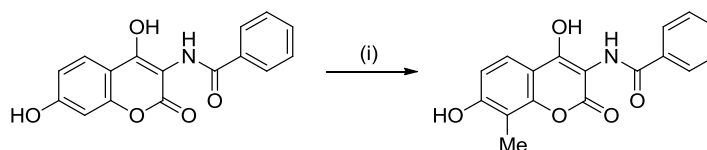
analogues. Furthermore, based on the work by Gruber and coworkers, NovO and CouO were also predicted to have much higher activity with SAM than its non-natural analogues, which would further limit this methodology to methyl transfer if residual methionine was present.<sup>[107]</sup> Therefore, an operationally simple and scalable protocol for the removal of residual Met was sought after.

The intracellular concentration of metabolites in *E. coli* has been studied by Rabinowitz and coworkers using LC-MS/MS.<sup>[331]</sup> Nearly half the measured metabolome comprised amino acids, out of which the vast majority is glutamate with an intracellular concentration of ~96 mM. In contrast, the intracellular concentration of methionine was estimated to be significantly lower at 0.15 mM.<sup>[331]</sup> Although it has been shown that this can be decreased further by suppressing biosynthesis by addition of Met to the growth medium,<sup>[332]</sup> this was predicted to inhibit recombinant protein expression as Met is required to initiate the translation process.<sup>[333]</sup> Therefore, methods to remove Met from the *E. coli* cell lysate and cell free extracts after protein expression were investigated.

A screen of methods to remove Met was carried out (**Table 31**). In each case, the conversion of substrate **67** to methylated product **69** was measured both without and with the addition of 2 eq. Met. The latter was carried out as a positive control to determine whether the enzymes retained activity after each treatment. Additionally, the assay was repeated using *E. coli* cell lysate containing NovO or CouO and *E. coli* cell lysate containing SalL without the addition of Met, which resulted in ~30% conversion to **69** (**Table 31**, entry **1**). Using the cell free extract for both NovO or CouO and SalL also produced similar levels of background methylation (**Table 31**, entry **2**).

To determine whether the expression medium was the source of the background Met, the cell pellet was washed twice with reaction buffer prior to lysis (**Table 31**, entry **3**). However, ~20% background methylation was still observed. Therefore, it was concluded that the majority of the Met originated from the *E. coli* metabolome. Next, the *E. coli* cell free extracts containing NovO or CouO and SalL was filtered using a centrifugal filter with a 10 kDa retention limit. It was predicted that Met (Mw: 149 Da) would pass through the filter, whilst NovO, CouO and SalL would be retained (Mw: 26.2, 26.5 and 30.1 kDa, respectively). Whilst this was successful in reducing background methylation by ~20%, a decrease in activity of ~15% was also observed in the positive control experiment (**Table 31**, entry **4**). This could be due to instability of the enzyme towards spin filtration or loss of enzyme by sticking to the filter device. Next, the reaction was carried out with purified SalL,

both with washed *E. coli* cell lysate containing NovO or CouO and with washed and filtered *E. coli* cell lysate containing NovO or CouO (Table 31, entries 5 and 6). This decreased the amount of background methylation by half, indicating that the residual methionine was present in roughly equal amounts in the NovO or CouO and SaLL preparations. Finally, dialysis was investigated as a method for Met removal (Table 31, entry 7). Gratifyingly, this was successful in removing residual Met and no background methylation was observed after 24 hours. However, conversion to the methylation product in the both control reactions was reduced to 50-60% and a white precipitate was observed throughout the reaction. This



Entry	NovO/ CouO	SaLL	- Met	- Met	+ Met	+ Met
			% Conversion CouO	% Conversion NovO	% Conversion CouO	% Conversion NovO
1	Cell lysate	CFE	28	30	-	-
2	CFE	CFE	26	29	-	-
3	Washed cell lysate	Washed CFE	22	20	79	70
4	Filtered CFE	Filtered CFE	6	9	74	83
5	Washed cell lysate	Purified	15	16	83	98
6	Washed and filtered CFE	Purified	16	25	64	92
7	Dialysed CFE	Dialysed CFE	0	0	58	55

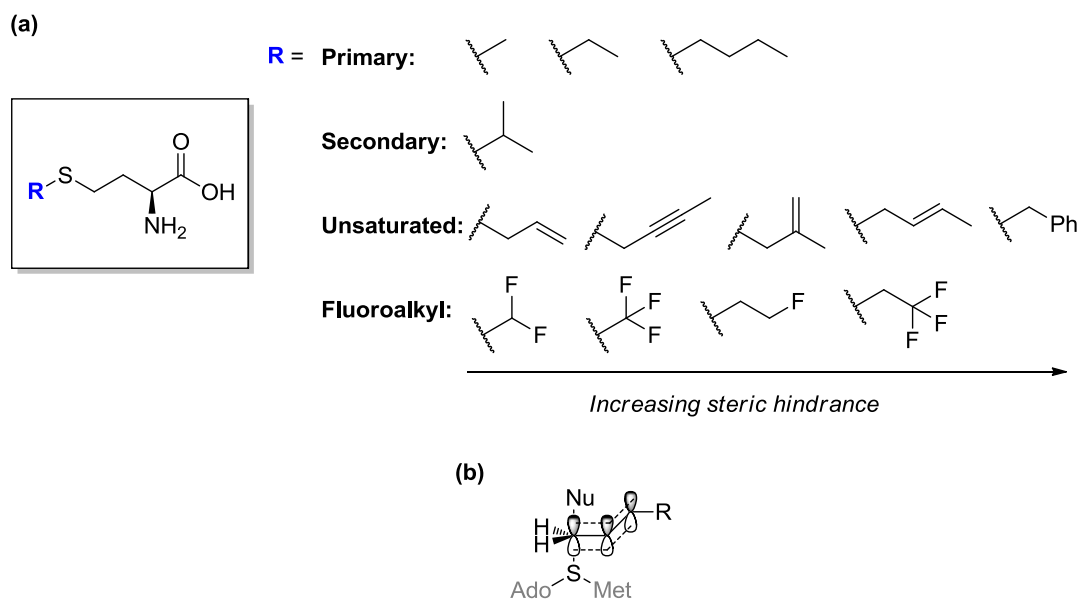
**Table 31.** Screening of methods to remove residual Met from tandem reaction system. MTs were diluted at 10 mL/g cell pellet; SaLL was diluted at 5 mL/g cell pellet. CFE: cell free extract. Filtered CFE: CFEs were filtered twice using an Amicon Ultra Centrifugal Filter with a 10 kDa MW limit. Washed CFE: cell pellet was washed twice with 50 mM phosphate buffer prior to lysing. Dialysed CFE: CFE dialysed for 16 hours at 4 °C in phosphate buffer using 10 KDa MW limit dialysis tubing. Purified: Enzyme purified by HisTRAP affinity chromatography.

suggested that the enzymes were not stable to dialysis and denaturation was occurring, resulting in a decrease in activity.

Out of the methods screened, dialysis was found to be most effective in reducing background methylation, although a protein stability problem had also been identified. However, it was concluded that reducing background methylation was a priority in order to eliminate competition in the Sall active site with non-natural Met analogues. Therefore, the dialysis was scaled up to 50 mL for each protein and the resulting protein solutions used directly in further experiments.

## 5.5 Synthesis of Met analogues

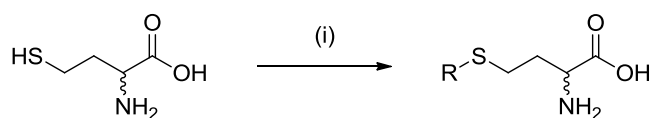
With these results in hand, the scope of this methodology regarding the alkyl group that can be transferred was explored by screening the system against a library of *S*-alkylated methionine analogues. The library was designed to investigate both primary and secondary alkyl transfers and to explore the extent of steric tolerance by Sall, NovO and CouO (**Figure 101a**). Additionally, unsaturated analogues were also investigated. Not only do these provide a functional handle for further synthetic manipulation of the alkylated product, but allylic and propargylic systems are ‘double activated’ towards the S<sub>N</sub>2 reaction, resulting in an enhanced rate of alkyl transfer compared to the corresponding saturated system. The double activation effect arises from stabilisation of the S<sub>N</sub>2 sp<sup>2</sup> hybridised transition by overlap of the reacting p-orbital with the π-system of the adjacent carbon atom (**Figure 101b**), which is known as conjugative stabilisation.<sup>[334]</sup>



**Figure 101.** (a) Proposed library of *S*-alkylated Met analogues for screening with one-pot tandem SAM synthesis and alkyl transfer system. (b) Mechanism of ‘double activation’ of unsaturated SAM analogues.

With the exception of the *S*-fluoroethyl analogue, the preparation of *S*-fluoroalkyl Met analogues has been described previously. The remaining Met analogues were synthesised by *S*-alkylation using the corresponding alkyl bromide in the presence of NaOH (**Table 32**). Despite this being a straightforward *S*-alkylation reaction which in most cases proceeded with quantitative conversion and no by-product formation, isolation of the amino acid products was challenging and resulted in a range of isolated yields from moderate (60%) to good (77%). To isolate the amino acids, ion exchange chromatography, reverse phase chromatography and precipitation of the product by pH control were all explored. Out of these, precipitating the product out of water or methanol by adjusting the pH and collecting by filtration was found to provide material with the highest purity. This also avoided the need for lengthy ion exchange or reverse phase chromatography procedures. Using racemic homocysteine as the starting material, this method provided sufficient amounts of Met analogues to support further investigations (**Table 32**). Methionine (**5**) and ethionine (**250**) were not synthesised as they are both commercially available.

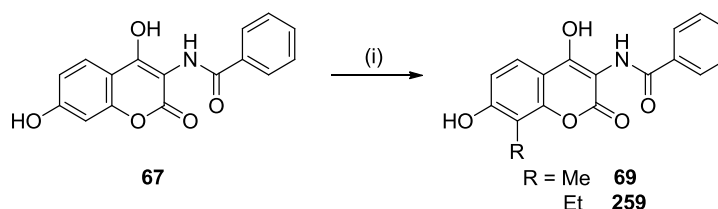




Entry	R	Compound No.	% Yield
1		5	Commercial
2		250	Commercial
3		251	60
4		252	68
5		253	68
6		254	58
7		255	70
8		256	77
9		257	66
10		258	73

**Table 32.** Synthesis of Met analogues. *Reagents and conditions:* (i) RBr, NaOH<sub>(aq)</sub>, EtOH, 22 °C, 24 h.

With a library of *S*-alkylated and *S*-fluoroalkylated Met analogues in hand, the scope of the tandem SAM synthesis and alkyl transfer methodology using SaIL and NovO or CouO was explored. However, before screening the entire library, the commercially available acid L-ethionine (L-Et) was tested in parallel with L-Met as a positive control. Unfortunately, analysis of the reaction mixture by HPLC after 24 h revealed that there was still a significant level of background methylation. Whilst conversions to methylated product **69** of 21% and 10% were seen for CouO and NovO, respectively, only 13% and 7% ethylation were observed for CouO and NovO, respectively (**Table 33**). This indicated that the previous efforts to remove residual Met from the system had been unsuccessful. Furthermore, lower levels of methylation when Met was added (CouO: 49% and NovO: 69%) compared to those obtained during optimisation experiments (CouO and NovO: 100%) suggested that the enzymes were not stable to dialysis, rendering this process unsuitable.

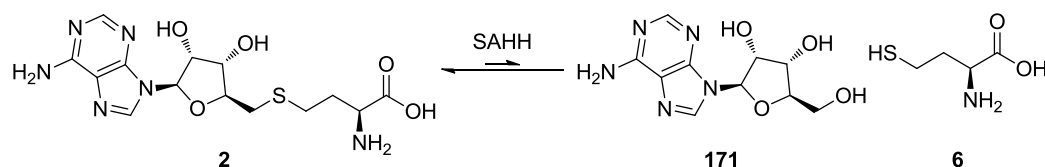


Entry	R	% Alkylation CouO	% Alkylation NovO	% Methylation CouO	% Methylation NovO
1		49	69	NA	NA
2		13	7	21	10

**Table 33.** Screen of Met and Et in tandem alkylation system. *Reagents and conditions:* (i) CIDA, Met, BSA, DTT, CouO/NovO dialysed CFE in 50 mM phosphate buffer, pH 7 (CouO) and pH 6.5 (NovO) at 10 mL/g pellet, SalL dialysed CFE in 50 mM phosphate buffer, pH 6.8 at 5 mL/g pellet. % Conversion by area/area% by HPLC reported after 24 h.

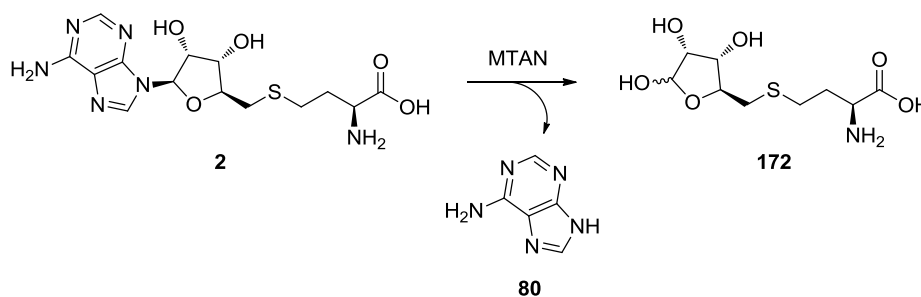
As it had not been possible to suppress background methylation using *E. coli* cell free extract harbouring SalL and *E. coli* cell lysate harbouring NovO and CouO, it was decided to develop the system using purified enzymes. Any Met present in the cell free extract prior to purification by Ni-NTA was predicted to be removed during the wash step and should therefore be an effective method to avoid the issue of background methylation. For proof of concept and development purposes, this work focused on the use of SalL with NovO, although it predicted that a similar approach could be taken with CouO.

Due to the cost and time associated with enzyme purification, a process that used as little enzyme as possible was desirable. This was particularly important for NovO, which has a tenfold lower  $k_{cat}$  with substrate **67** than SalL (NovO:  $0.02\text{ s}^{-1}$  and SalL:  $0.2\text{ s}^{-1}$ ).<sup>[275]</sup> To this end, the reduction of inhibition of NovO by SAH was investigated. SAH is the byproduct of the reaction and a known inhibitor of most SAM dependent MTs.<sup>[335,336]</sup> Two strategies have been reported to suppress inhibition of MTs by SAH. The first is the use of the SAH hydrolase (SAHH) to decompose the byproduct SAH (**2**) into adenosine (**171**) and homocysteine (**6**) (**Scheme 79**). Although the equilibrium of this reaction is known to favour the synthesis rather than the decomposition reaction,<sup>[337]</sup> SAHH has been demonstrated to be a useful additive to SAM dependent MT reactions.<sup>[188,190,283,284]</sup>



**Scheme 79.** Conversion of SAH into adenosine and homocysteine catalysed by SAHH.

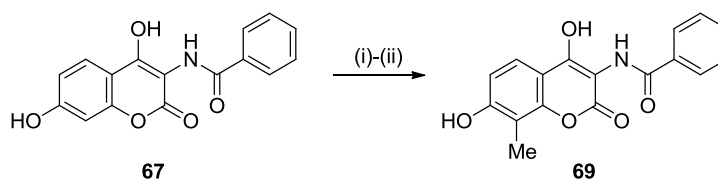
The second strategy to degrade SAH is the use of methylthioadenosine nucleosidase (MTAN), which cleaves the adenine (**80**) ring from SAH, thereby greatly reducing its binding affinity to the MT SAM binding site.<sup>[289]</sup> Unlike the degradation of SAH by SAHH, the MTAN catalysed reaction is irreversible and has a very high  $k_{cat}$  of  $\sim 2.1 \text{ s}^{-1}$  (calculated for MTAN from *E. coli*).<sup>[338]</sup> As with SAHH, the use of MTAN to suppress by product inhibition in MTs is also well precedented.<sup>[187]</sup>



**Scheme 80.** Degradation of SAH by cleavage of the adenine ring catalysed by MTAN.

Based on this literature precedent, the tandem SAM synthesis and methyl transfer methodology using purified enzymes with the addition of either SAHH or MTAN was investigated. In both cases, SAHH or MTAN was added after full consumption of CIDA (monitored by HPLC), as CIDA is a potent inhibitor of SAHH and MTAN.<sup>[289,339]</sup> For screening purposes, the reaction was carried out with 2 eq. CIDA and 2-10 eq. Met. As shown in **Table 34**, both systems were effective in carrying out the desired transformation. However, much higher conversions to methylated product **69** were obtained with MTAN than with SAHH.

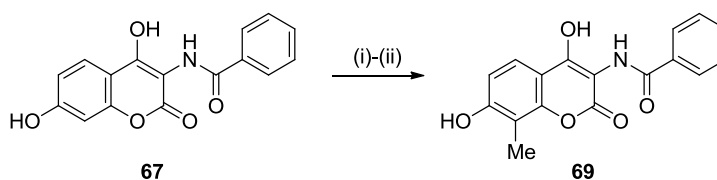
Whilst the addition of SAHH resulted in a maximum conversion of  $\sim 70\%$  when 2.5, 5 or 10 eq. Met was added (**Table 34**, entries **1-4**), reactions to which MTAN had been added all gave quantitative conversion to **69**, even with 2 eq. Met (**Table 34**, entries **5-8**). Based on these preliminary results, it was decided to use MTAN as an additive in all further development work on this methodology.



Entry	Eq. CIDA	Eq. Met	% Conversion
<b>1<sup>a</sup></b>	2	2	8
<b>2<sup>a</sup></b>	2	2.5	72
<b>3<sup>a</sup></b>	2	5	68
<b>4<sup>a</sup></b>	2	10	68
<b>5<sup>b</sup></b>	2	2.0	100
<b>6<sup>b</sup></b>	2	2.5	100
<b>7<sup>b</sup></b>	2	5.0	100
<b>8<sup>b</sup></b>	2	10.0	100

**Table 34.** Preliminary conditions screening for tandem alkylation using purified SalL and NovO. *Reagents and conditions:* (i) L-Met (**5**), CIDA (**85**), BSA (1 mg/mL), SalL (0.03 mg/mL), 37 °C, 2 h, then (ii) NovO (0.6 mg/mL) and <sup>a</sup>SAHH (0.16 mg/mL) or <sup>b</sup>MTAN (0.03 mg/mL), 37 °C, 24 h. % Conversion determined by area/area% HPLC.

Next, the minimum number of equivalents of CIDA and Met required to achieve quantitative conversion was investigated. Whilst it was possible to limit the amount of CIDA to 1.5 eq., decreasing the amount of Met below 2 eq. resulted in a drop off in conversion (**Table 35**, entries **1-5**). As such, 1.5 eq. CIDA and 2 eq. Met were used in all further experiments. Next, the use of DTT as an additive to improve protein stability was explored. Gratifyingly, the addition of DTT to a final concentration of 1 mM to the reaction provided quantitative conversion to methylated product **69** (**Table 35**, entry **6**). A reaction was also run in the absence of MTAN in order to quantify its effectiveness on suppressing byproduct inhibition. In this case, only 61% conversion was observed (**Table 35**, entry **7**), which indicates that MTAN is effective in removing SAH from the system and, in doing so, suppressing inhibition of NovO by SAH.

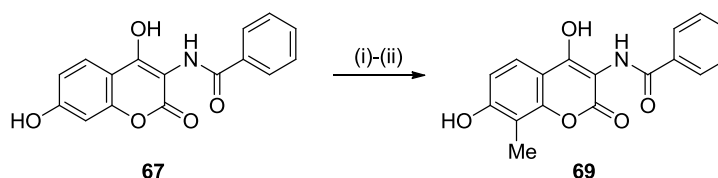


Entry	Eq. CIDA	Eq. Met	Additive(s)	% Conversion
1	2	2	MTAN	99
2	1.5	2	MTAN	98
3	1.5	1.5	MTAN	92
4	1.1	1.1	MTAN	66
5	1.1	1.5	MTAN	75
6	1.5	2	DTT+MTAN	100
7	1.5	2	-	61

**Table 35.** Optimisation of tandem SAM synthesis and methylation using purified enzymes. *Reagents and conditions:* (i) L-Met (**5**), CIDA (**85**), BSA (1 mg/mL), SaIL (0.03 mg/mL), 37 °C, 2 h, then (ii) NovO (0.6 mg/mL) and MTAN (0.03 mg/mL), 37 °C, 24 h. % Conversion determined by area/area% HPLC. Optimised conditions highlighted in grey.

Having identified conditions that enabled quantitative methylation of aminocoumarin substrate **67**, the scope of the methodology with regard to the alkyl group that could be transferred was investigated using the previously prepared library of Met analogues.

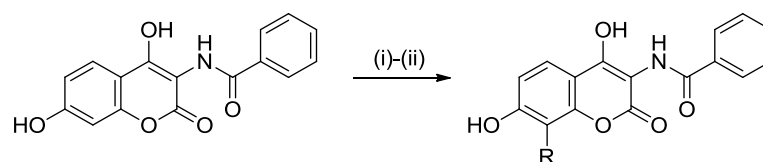
However, as the Met analogues were racemic, it was first necessary to investigate whether the presence of the non-natural, D-Met had any inhibitory effect on the system. Therefore, the reaction was carried out with L-Met, L-Met and D-Met and D-Met only (**Table 36**). 83% conversion was obtained with L-Met, compared to 77% conversion with a 1:1 ratio of L-Met to D-Met. Considering the small scale on which these experiments were carried out, this difference of 6% conversion was deemed to be within experimental error and as such, D-Met was concluded not to have a significant inhibitory effect on the system. Interestingly, 3% conversion to methylated product **69** was observed when exclusively D-Met was added, indicating that D-Met is accepted by SaIL and the resulting non-natural diastereomer of SAM is accepted by NovO, albeit at a drastically decreased rate (**Table 36**, entry **3**).



Entry	Eq. L-Met	Eq. D-Met	% Conversion
1	2	0	83
2	2	2	77
3	0	2	3

**Table 36.** Investigation into effect of D-Met on the tandem SAM synthesis and alkylation system. *Reagents and conditions:* (i) L/D-Met (**5**), CIDA (**85**, 2 eq.), BSA (1 mg/mL), SalL (0.03 mg/mL), 37 °C, 2 h, then (ii) NovO (0.6 mg/mL) and MTAN (0.03 mg/mL), 37 °C, 24 h. % Conversion determined by area/area% HPLC.

Based on this data, it was concluded that the racemic *S*-alkylated Met analogues could be used in this system without need for enantiomeric separation. Therefore, the library of *S*-alkylated Met analogues was screened under the optimised reaction conditions (**Table 37**). To compensate for the 1:1 ratio of D- to L- amino acid in the samples, 4 eq. of the racemate was used in the screening reactions. Whilst methyl and ethyl groups were successfully transferred in 65% and 28% conversion, respectively, only traces (<1%) of product were observed by HPLC with all of the remaining Met analogues (**Table 37**). As Gruber and coworkers have previously demonstrated that NovO can accept and transfer non-natural cofactors, it was concluded that these results were due to the inability of SalL to accept the more sterically hindered Met-analogues.<sup>[107]</sup> Indeed, SalL has been shown to have limited substrate scope and our results reinforce this conclusion.<sup>[181]</sup> As such, work on the directed evolution of SalL towards a wider substrate scope is ongoing in the research group.



Entry	R	% Conversion
1		65
2		28
3		<1
4		<1
5		<1
6		<1
7		<1
8		<1
9		<1
10		<1
11		<1
12		<1
13		<1

**Table 37.** Substrate scope of tandem SAM synthesis and methylation using purified enzymes. *Reagents and conditions:* (i) Met analogue (*rac*), CIDA (**85**, 2 eq.), BSA (1 mg/mL), SalL (0.03 mg/mL), 37 °C, 2 h, then (ii) NovO (0.6 mg/mL) and MTAN (0.03 mg/mL), 37 °C, 24 h. % Conversion determined by area/area% HPLC.

Having established an *in situ* cofactor production system that was compatible with NovO and effective for the transfer of a methyl or ethyl group to aminocoumarin substrate **67**, the scalability of the process was explored by demonstrating the process on preparative scale. Therefore, the methylation and ethylation of **67** using Met or ethionine, respectively, was carried out using 20 mg of starting material. Additionally, preparative scale experiments were also carried out with pyrrole aminocoumarin substrate **68** and dihydroxynaphthalene (DHN) substrate **121**, which have also been shown to be accepted by NovO. Whilst crude *E. coli* cell lysate harbouring NovO and *E. coli* cell free extract harbouring SalL were used for Me transfer, purified

$R^1 = \text{Me} \quad \mathbf{5}$   
 $\text{Et} \quad \mathbf{250}$

Substrate	R <sup>1</sup>					
	Product no.	Conversion (%)	Isolated yield (%)	Product no.	Conversion (%)	Isolated yield (%)
 <b>67</b>	<b>69</b>	97 <sup>a</sup>	52 <sup>a</sup>	<b>259</b>	34 <sup>b</sup>	18 <sup>b</sup>
 <b>68</b>	<b>70</b>	100 <sup>a</sup>	76 <sup>a</sup>	<b>260</b>	30 <sup>b</sup>	27 <sup>b</sup>
 <b>121</b>	<b>72</b>	65 <sup>a</sup>	46 <sup>a</sup>	<b>261</b>	7% <sup>b</sup>	-

**Table 38.** Preparative scale tandem SAM or SAM analogue synthesis and alkyl transfer using SalL and NovO. <sup>a</sup>Reaction carried out using NovO cell lysate and SalL cell free extract and purified MTAN and 7 h pre-incubation period with SalL. <sup>b</sup>Reaction carried out using purified enzymes and 32 hour pre-incubation with SalL. *Reagents and conditions:* (i) SalL (0.5 mol%), BSA (1 mg/mL), DTT (1 eq.), 50 mM potassium phosphate buffer pH 6.5, CIDA (1.5 eq.), Met or Met analogue (2 eq.), 37 °C, 7 hours (CH<sub>3</sub>) or 32 hours (ethyl). (ii) NovO (4 mol%), MTAN (0.05 mol%), 37 °C, 16 hours. Reactions carried out on 20 mg substrate scale. Yields of isolated products in parentheses.

enzymes were used for ethyl transfer to avoid issues with background methylation. Due to the much lower activity of SalL with ethionine (~0.8 nM/min) compared to methionine (~10 nM/min),<sup>[181]</sup> a pre-incubation period of 32 hours before the addition of NovO and MTAN was used for ethyl transfer reaction. As shown in **Table 38**, excellent levels of conversion to methylated products were obtained for aminocoumarin substrates **67** and **68** (97% and 100%, respectively) and a moderate conversion of 65% for the DHN substrate **121**. As expected based on the previous screening work (**Table 37**), much lower conversions to ethylated products **259-261** were obtained. Due to the small scale of these reactions, the isolated yields showed some variation. In particular, low isolated yields were obtained for benzamide aminocoumarin **69**, due to very poor aqueous solubility and poor solubility in all



commonly used extraction solvents, resulting in significant loss of yield during work-up. As only 7% ethylation of DHN substrate **261** was achieved and the reaction had been carried out on a small scale, product isolation was not attempted (**Table 38**).

Despite the limitations of this methodology regarding substrate scope, this work nonetheless represents the first example of SaIL being used in tandem with a C-methyltransferase in a scalable, one-pot process. This work therefore provides proof of concept for a biocatalytic platform for the Friedel-Crafts alkylation under mild reaction conditions, starting from inexpensive starting materials.

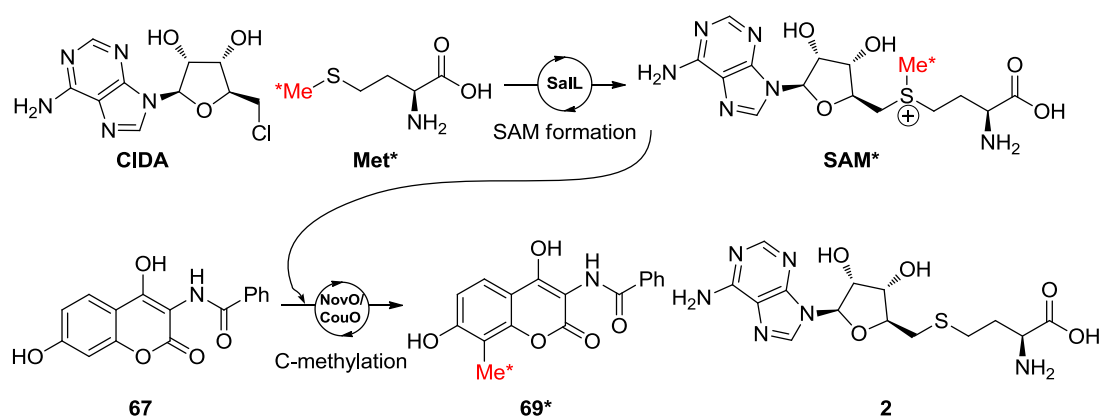
## 5.6 Application of tandem SAM synthesis and methylation to isotope labelling

Having developed a methodology for a one-pot tandem SAM synthesis and methyl or ethyl transfer system, it was hypothesised that this process could be extended to the transfer of stable isotope labelled methyl groups by use of the corresponding isotopically labelled Met analogues. The many applications of stable isotope (SI) labelling span a range of disciplines. For example,  $^{13}\text{C}$  labelling has been used extensively to elucidate biosynthetic pathways, especially in the popular field of polyketide synthases.<sup>[340,341]</sup> Another popular application of SI labelling is in the study of drug metabolism. Understanding the metabolic and pharmacokinetic fate of a drug molecule is a crucial step in the drug development process.<sup>[342]</sup> A common approach is the *in vitro* analysis of drug metabolites in biofluids. Due to its high sensitivity and selectivity, mass spectrometry (MS) has evolved as the analytical method of choice for such studies. If drug metabolites have been labelled with SIs, MS filtering techniques can be used to quickly identify metabolites, thereby rationalising the fate of the drug *in vivo*.<sup>[342–345]</sup> For example, SI labelling/isotope filter MS strategies have been proposed a general approach to screen for glutathione adducts of drugs.<sup>[345,346]</sup> Finally, there has also been considerable interest in the use of stable isotopes, in particular deuterium, in the drug molecule itself.<sup>[347]</sup> The rationale for this stems from the benefits of the deuterium kinetic isotope effect on the safety and efficacy of drugs metabolised by C-H bond scission, whilst maintaining a very similar size and shape to the non-deuterated analogue.<sup>[347,348]</sup>

Despite the widespread application of SI labelling, methods for the incorporation of SIs into compounds of interest are limited. In particular, a method for the regiospecific, late stage introduction of SIs would be very attractive. Owing to the prevalence of the methyl group in pharmaceutically active compounds, incorporation of stable isotopes into a molecule by

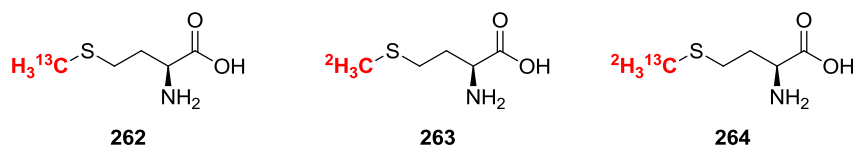
methylation using a labelled methylating agent could be a useful strategy for the preparation of labelled drug molecules.<sup>[287]</sup>

Therefore, the application of the one-pot tandem SAM synthesis and methyl transfer methodology to the synthesis of stable isotope labelled aminocoumarins was investigated. In an analogous fashion to the previously described process, the use of CIDA (**85**) and an SI labelled Met analogue (Met\*) with SaIL to synthesise an SI labelled SAM analogue (SAM\*) was envisaged. The isotopically labelled methyl group would then be transferred to the 8-position of the coumarin ring of substrate **67** using NovO, thereby providing the aminocoumarin substrate bearing a SI label (**Scheme 81**). It was hypothesised that this could provide proof of principle for a general strategy for late stage, regiospecific stable isotope labelling of any substrate for a SAM dependent MT.



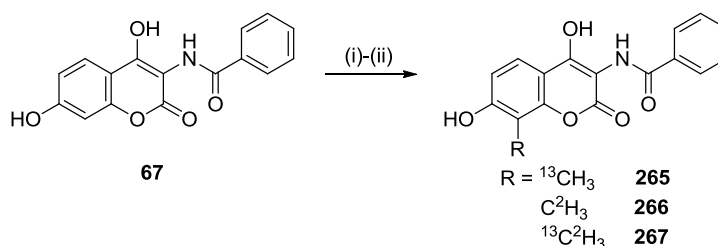
**Scheme 81.** Proposed one-pot tandem SIL SAM synthesis and SIL methyl transfer system using SaIL and NovO.

For proof of principle studies, the tandem SAM synthesis and SI labelled methyl group transfer was carried out on a screening scale and the reactions monitored by HPLC and LC-MS. Three different SI labelled Met analogues were used, **262-264** (**Figure 102**), all of which were commercially available.



**Figure 102.** Isotopically labelled Met analogues.

Using the optimised conditions from the development of the tandem system with each of Met analogues **262-264**, the SI labelling of aminocoumarin substrate **67** was investigated. Additionally, the system was also tested using CouO in place of NovO to further explore the scope of this methodology. Purified enzymes were used in all cases to ensure there was no residual non-labelled methionine in the system. Gratifyingly, analysis by HPLC revealed good levels conversion (72-82%) to SI labelled products **265-267** was observed with all three Met analogues (**Table 39**).

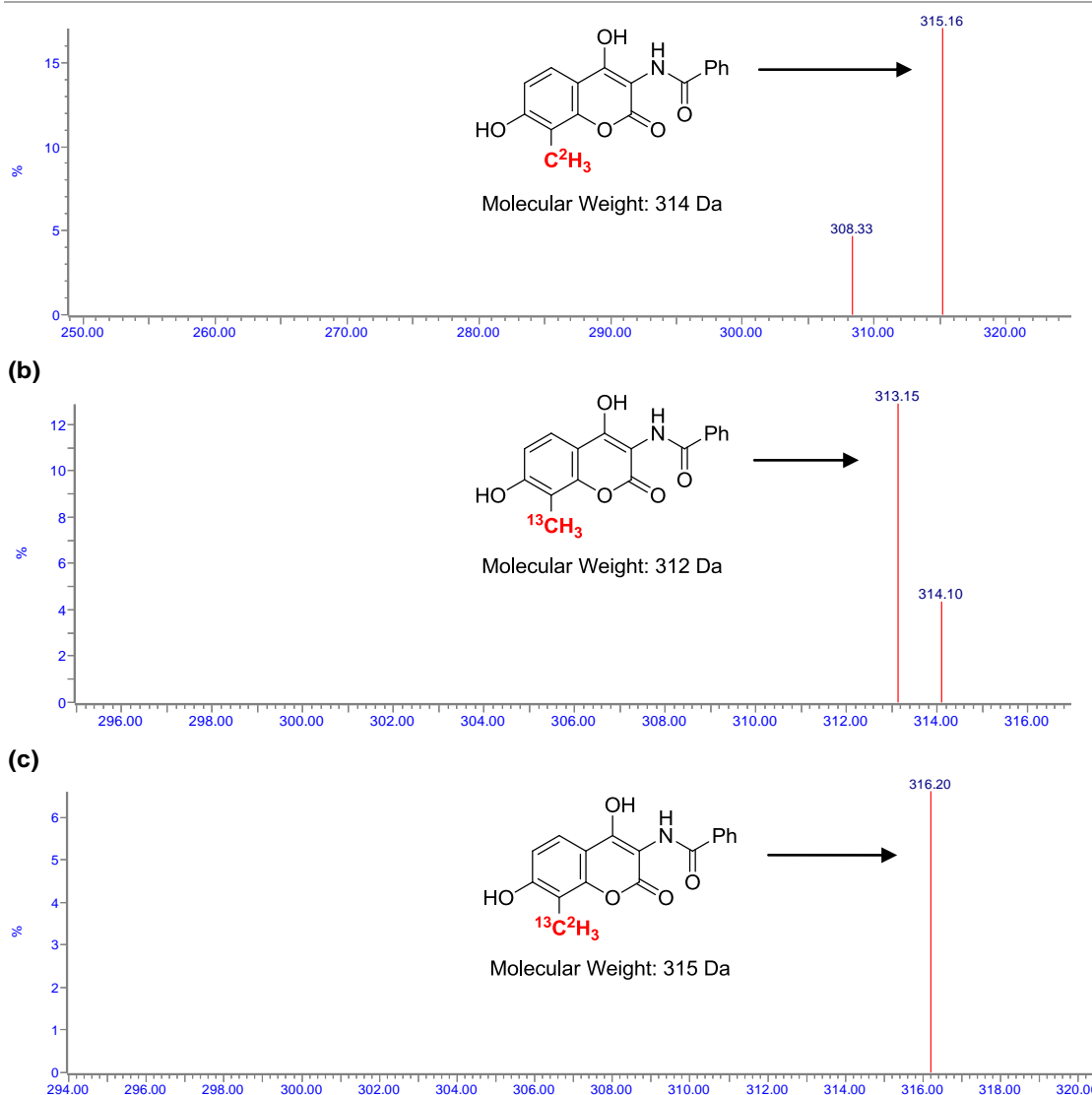


Entry	MT	R	% Conversion
1	NovO	$^{13}\text{CH}_3$	76
2	CouO	$^{13}\text{CH}_3$	75
3	NovO	$\text{C}^2\text{H}_3$	73
4	CouO	$\text{C}^2\text{H}_3$	72
5	NovO	$^{13}\text{C}^2\text{H}_3$	76
6	CouO	$^{13}\text{C}^2\text{H}_3$	82

**Table 39.** Tandem SAM synthesis and methylation using isotopically labelled Met analogues. *Reagents and conditions:* (i) **262**, **263** or **264** (2 eq.), ClDA (1.5 eq.), BSA (1 mg/mL), SalL (0.03 mg/mL), 37 °C, 2 h, then (ii) NovO (0.6 mg/mL) and MTAN (0.03 mg/mL), 37 °C, 24 h. % Conversion determined by area/area% HPLC.

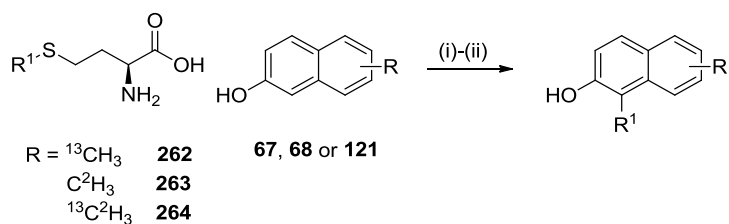
To confirm whether the SI labelled methyl group had been transferred to aminocoumarin **67**, the samples were further analysed by LC-MS. Pleasingly, the  $[\text{M}+\text{H}]^+$  mass ion for each of the three SI labelled products was clearly visible (**Figure 103**), confirming that the SI labelled methyl group had successfully been transferred to the substrate. Furthermore, the  $[\text{M}+\text{H}]^+$  mass ion corresponding to the non-labelled product ( $m/z=312$ ) was not observed in any of the three spectra, indicating that the desired SI labelled isotopologue had been formed exclusively.

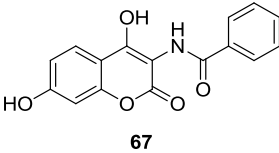
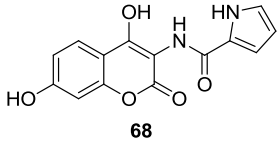
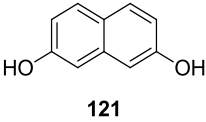
(a)



**Figure 103.** Section of mass spectra of isotopically labelled aminocoumarin products (a) **256** (b) **266** and (c) **267** produced *via* tandem SAM synthesis and methyl transfer reaction using SalL and NovO.

Having obtained proof of concept for this methodology on a small scale, the reactions were carried out on preparative scale, such that the isotopically labelled products could be isolated and fully characterised. Although NovO and CouO have both been shown to be compatible with this system, the preparative scale experiments were only carried out with NovO, as the same products would be obtained in each case. At the time of writing, this is the only report of an *in situ* cofactor production system using SalL and a methyltransferase for the preparative scale of isotopically



Substrate	R								
	$^{13}\text{CH}_3$			$\text{C}^2\text{H}_3$			$^{13}\text{C}^2\text{H}_3$		
	Product no.	Conv. (%)	Isolated Yield (%)	Product no.	Conv. (%)	Isolated Yield (%)	Product no.	Conv. (%)	Isolated Yield (%)
 <b>67</b>	<b>256</b>	95	76	<b>266</b>	88	53	<b>267</b>	90	48
 <b>68</b>	<b>268</b>	100	91	<b>269</b>	100	81	<b>270</b>	100	76
 <b>121</b>	<b>271</b>	94	83	<b>272</b>	91	69	<b>273</b>	92	87

**Table 40.** Preparative scale tandem SAM synthesis and methyl transfer using SalL and NovO with SI labelled Met analogues. *Reagents and conditions:* (i) SalL (0.5 mol%), BSA (1 mg/mL), DTT (1 eq.), 50 mM potassium phosphate buffer pH 6.5, CIDA (1.5 eq.), Met or Met analogue (2 eq.), 37 °C, 7 hours. (ii) NovO (4 mol%), MTAN (0.05 mol%), 37 °C, 16 hours. Reactions carried out on 20 mg substrate scale.

labelled compounds. In addition to aminocoumarin substrate **67**, this transformation was also demonstrated with aminocoumarin substrate **68** and DHN **121** (Table 40). All reactions were carried out using 20 mg of starting material, and the percentage conversion to SI labelled product after 24 h analysed in each case. Additionally, the products were isolated from the reaction mixture by extraction and subsequent mass directed automated purification (MDAP). Gratifyingly, good (88%) to excellent (100%) levels of conversion were observed for all substrates. In particular, the pyrrole aminocoumarin substrate **68** was converted in quantitative conversion with all three Met analogues. Due to the small scale of these reactions, the isolated yields showed

more variation. In particular, lower isolated yields were obtained for benzamide aminocoumarin **67**, due to very poor aqueous solubility and poor solubility of the products in all commonly used extraction solvents, resulting in significant loss of yield during work-up. Nonetheless, in all cases it was possible to obtain sufficient material for full structural characterisation.

This work demonstrates a biocatalytic, one-pot procedure for the *in situ* synthesis of stable isotope labelled SAM analogues and the subsequent MT mediated labelling of an appropriate substrate. Moreover, the system is scalable for the isolation of SI labelled compounds, which can be characterised and could be used for further applications such as those discussed above. Although this system has been demonstrated with the C-MT NovO, it is anticipated that the principle could be applied to any SAM dependent MT, thereby lending itself not only to the preparation of isotopically labelled small molecules, but also to the site specific labelling of macromolecules such as DNA and proteins.<sup>[31,215,334,349]</sup>

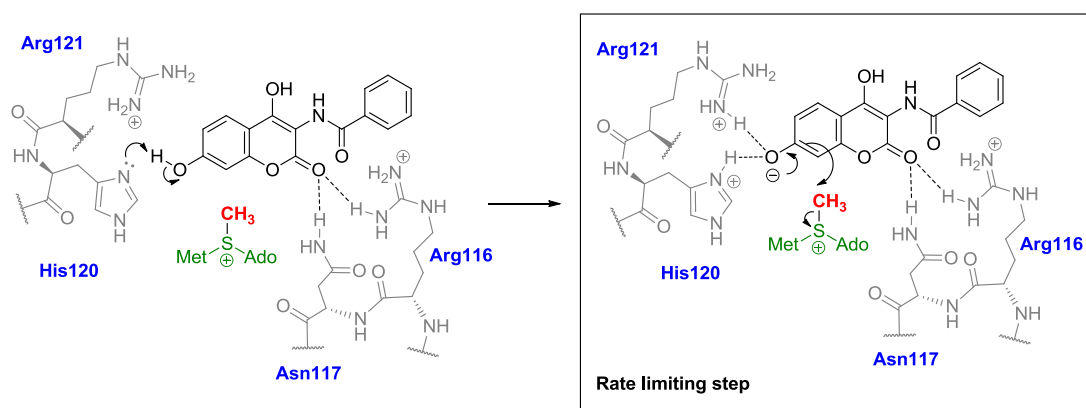
## Conclusions

The overall aim of this thesis was to develop the use of the C-MTs NovO and CouO as biocatalysts for the Friedel-Crafts alkylation and fluoroalkylation. Five objectives were identified in order to achieve this: the structural characterisation of NovO and CouO; elucidation of the catalytic mechanism of NovO and CouO; evolution of NovO and CouO towards a broader substrate scope; synthesis of 'CF<sub>3</sub>-SAM'; and to establish a system for *in situ* cofactor production. To this end, a number of key advances have been made that significantly advance understanding of NovO and CouO, in addition to the establishment of an efficient *in situ* cofactor production system.

The first major objective was to elucidate the structure of NovO and CouO. This would provide a valuable foundation on which to base a DE project using a semi-rational library design strategy. To this end, the X-ray crystal structure of SelMet-NovO was successfully elucidated to 1.9 Å resolution with SAH bound in the active site. As previously predicted by Gruber and coworkers based on multiple sequence alignments, the crystal structure revealed NovO to be a homodimeric Class I MT, in which each monomer comprised a Rossmann fold with two additional  $\alpha$ -helices at the N-terminus, as is typical for small molecule MTs.<sup>[55]</sup> Interestingly however, NovO also displayed two additional C-terminal  $\alpha$ -helices, which were involved in dimerisation with the second monomer. This makes it structurally distinct from the archetypal Class I MT COMT, which was described in Chapter 1. Based on these results, a homology model of CouO was generated using the crystal structure of NovO, as the sequences of these two enzymes share 85% sequence identity and have excellent homology. The homology model indicated that NovO and CouO share a very similar tertiary structure. As NovO and CouO have a unique primary and tertiary structure compared to other known MTs, they represent a unique entry point into a toolbox of biocatalysts for performing C-C bond forming reactions, which is particularly interesting from a biocatalysis point of view.<sup>[55,239]</sup>

Having structurally characterised NovO, the mechanism of methyl transfer within its active site was investigated. Based on a substrate docking model of an aminocoumarin bound to NovO, a series of active site residues were proposed and supported by a series of mutational analysis experiments. Out of these, H120 and R121 appeared to make key contacts with the 7-hydroxy group, which was found to be an essential requirement for substrate recognition. Testing the methyltransferase activity of the two lysine mutants H120K and R121K strongly

indicated H120 to be the general base in the catalytic mechanism, with R121 being involved in stabilisation of the resulting intermediate. Additionally, kinetic isotope effect data showed the methyl transfer to be the rate limiting step (**Scheme 82**).



**Scheme 82.** Key step in the catalytic mechanism of methyl transfer by NovO.

Out of the three types of catalysis by MTs outlined in Chapter 1 (proximity and desolvation; acid/base mediated; and metal dependent), NovO therefore falls the acid/base mediated group. To date, this is the first example of a His-Arg motif being implicated in the key step of methyl transfer for a small molecule MT and is mechanistically distinct from the two previous proposed mechanisms for methyl transfer by Coq5 and SibL.<sup>[99,101]</sup> Additionally, the active site of CouO was investigated using a homology model generated from NovO. Comparison of the model with the active site of NovO revealed very high similarity, with the exception of the residue at position 117, which was an asparagine in NovO and a histidine in CouO. Analysis of the ‘swap over mutants’ (CouO H117N and NovO N117H) indicated that this residue is involved in substrate binding and may account for the difference in substrate acceptance between the two enzymes that was observed by Gruber and coworkers.<sup>[107,260]</sup> As such, position 117 of both NovO and CouO may be an activity ‘hotspot’ which could be targeted towards saturation mutagenesis in a DE project aimed at altering the substrate specificity of these MTs.<sup>[161,178]</sup>

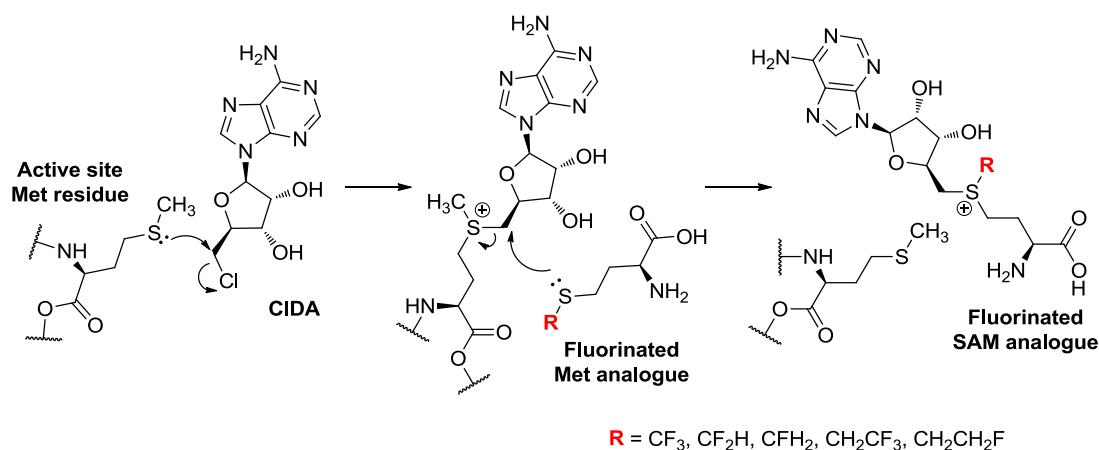
The structural and mechanistic insight into NovO and CouO provided a strong foundation upon which to carry out a DE project. Although DE can be effective for altering a range of properties of an enzyme including thermostability and higher  $k_{cat}$ , in this thesis the aim was to evolve the enzymes towards a wider substrate scope. The reason for this was to expand the toolbox of MTs available for use in organic synthesis.<sup>[226]</sup> DE work focused on NovO, as much more detailed structural and mechanistic information had been obtained for this



enzyme compared to CouO. Using a CASTing approach and analysing the mutants by UPLC, it was possible to identify a mutant (NovO N117M) with higher activity to 1, 6-dihydroxynaphthalene than the WT enzyme. At the time of writing, this represents the first example of the DE of a small molecule C-MT. Interestingly, this mutant was also at position 117, further suggesting that this position is key to mediating substrate acceptance and future work could involve generating a saturation mutagenesis library at this position. Unfortunately, it was not possible to establish a robust medium- or high-throughput assay for screening of the CASTing libraries. The use of an MTAN/LuxS coupled assay was explored in detail, and despite successfully coupling all three enzymes (NovO, MTAN and LuxS), it was not possible to identify a suitable thiol reactive probe for homocysteine detection that gave reproducible results. If further DE of NovO and CouO is attempted in the future, it will first be necessary to address this issue. To this end, a SAH selective riboswitch conjugated to the cpSpinach2 aptamer has recently been reported as a highly sensitive fluorescent reported of MT activity.<sup>[191]</sup> Based on this literature precedent, which was described in detail in Chapter 1, there may be scope to develop this into a high throughput screen for use in the DE of NovO and CouO, perhaps using a FACS approach.<sup>[196]</sup>

Methods for the regiospecific fluoroalkylation of small aromatic molecules are highly sought after due to the significance of fluoroalkyl groups (in particular the trifluoromethyl group) in medicinal chemistry.<sup>[117,309]</sup> The first biocatalytic trifluoromethylation of phenols has recently been reported by Kroutil and coworkers using a laccase and trifluoromethyl radicals generated *in situ*.<sup>[136]</sup> However, there is no literature precedent for a biocatalytic trifluoromethylation using MTs. To this end, the synthesis of a trifluoromethylated analogue of SAM (CF<sub>3</sub>-SAM) was extensively investigated based on a retrosynthetic analysis of the target molecule. Two approaches were investigated; the direct trifluoromethylation of SAH and the S<sub>N</sub>2 reaction between an adenosine derivative activated at the 5'-position and an *S*-trifluoromethylated analogue of methionine. To support the latter, a library of *S*-fluoroalkylated Met analogues, including four novel compounds, were prepared using zinc sulfinate salts developed by Baran and coworkers.<sup>[134]</sup> However, it was not possible to detect formation of CF<sub>3</sub>-SAM using either of these approaches. The enzymatic synthesis of CF<sub>3</sub>-SAM was also investigated using the halogenase SalL, but it was still not possible to detect product formation. Based on these results, it was concluded that the electron withdrawing effect of the fluorine substituents on the fluorinated Met analogues caused the sulfur to be not nucleophilic enough to carry out the S<sub>N</sub>2 reaction required to synthesise the target molecule. With this in mind, further work would be required to evolve SalL towards

accepting these more deactivated substrates. One possible approach to this is currently being investigated within the research group is outlined in **Scheme 83**. By engineering a Met residue into the active site of the protein, it may be possible to form an enzyme bound SAM intermediate by reaction with CIDA (**Scheme 83**). This would render the 5'-position more nucleophilic, such that it may be susceptible to nucleophilic attack by a fluorinated Met analogue to produce the desired fluorinated SAM analogue.



**Scheme 83.** Proposed design of an engineered SalL mutant capable of accepting fluorinated Met analogues.

The final objective was to establish a robust and scalable system for *in situ* cofactor synthesis that was compatible with both NovO and CouO. The halogenase SalL was identified as a promising starting point, due to the low cost of the starting materials and higher atom economy in comparison with the MAT system, which employs ATP as the adenosyl donor. By combining CIDA, Met and SalL in one pot with NovO or CouO and an aminocoumarin substrate, it was possible to achieve methylation in quantitative conversion, demonstrating that these enzymes can operate under the same reaction conditions. However, when *S*-alkylated analogues were investigated for use in this system to transfer a range of alkyl groups, ~30% background methylation was observed, indicating high levels of residual methionine were present in the *E. coli* cell lysates. In order to circumvent this issue, purified enzymes were utilised for the transfer of groups other than methyl, which included a range of isotopically labelled methyl groups and ethyl. This methodology was demonstrated on preparative scale and outlines a general procedure for the site specific stable isotope labelling of any substrate of a SAM dependent MT.

In summary, significant advances have been made in the development of NovO and CouO as biocatalysts for the Friedel-Crafts alkylation. A key milestone was elucidation of the X-ray

crystal structure of NovO, which enabled a catalytic mechanism of NovO to be proposed based on a substrate docking model, KIE data and mutational analysis studies. This, in turn, informed a DE project using a CASTing approach, from which a mutant with enhanced activity towards a dihydroxynaphthalene substrate was identified. To identify a toolbox of variants of NovO and CouO to cater for a wider range of Friedel-Crafts alkylation reaction, it will be necessary to establish a platform for the medium- or high-throughput screening of mutants. Nonetheless, this thesis serves as proof of concept that NovO is ‘evolvable’ and that there is scope to alter its substrate scope. In a second key outcome, an efficient and scalable *in situ* cofactor product system has been established, which addresses the issue of cofactor cost and instability. Taken collectively, these contributions have developed the use of NovO and CouO as biocatalysts for the Friedel-Crafts alkylation and furthermore, will enable future work on their directed evolution to address the need for biocatalysts in organic synthesis.

## **Chapter 6.**

# **Experimental**

## 6 Experimental

### 6.1 General experimental techniques and procedures

#### Reagents and Solvents

All reagents and solvents were used as supplied from commercial sources and used without further purification unless otherwise stated. Solvents were all HPLC grade and were used without further purification, unless otherwise stated. Unless otherwise stated, potassium phosphate buffer (made from solid  $K_2HPO_4$  and  $KH_2PO_4$  to pH 7.4 at 22 °C) was used and adjusted to the required pH using aq. 2 M HCl or aq. 2 M NaOH prior to use in biotransformations.

#### NMR Spectroscopy

NMR spectroscopy was carried out using a Bruker 400 UltraShield™ B-ACS 60 spectrometer. All chemical shifts ( $\delta$ ) were referenced to the deuterium lock and are reported in parts per million (ppm) and coupling constants are quoted in hertz (Hz). Abbreviations for splitting patterns are app. (apparent), s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet) and spt (septet). All NMR data was processed using ACD Labs 12.0 software. Proton and carbon chemical shifts were assigned using proton ( $^1H$ ), carbon ( $^{13}C$ ), Heteronuclear Single Quantum Coherence (HSQC), Heteronuclear Multiple-Bond Correlation Spectroscopy (HMBC) and Correlation Spectroscopy (COSY).

#### Liquid Chromatography-Mass Spectrometry (LC-MS), High Performance Liquid Chromatography (HPLC) and Gas Chromatography-Mass Spectrometry (GC-MS)

LC-MS was carried out on an Agilent 1100 HPLC instrument in conjunction with a Waters Micromass ZQ 2000/4000 mass detector using electrospray (ES) ionization. HPLC was carried out on an Agilent 1100 series instrument (see Appendix for method and column parameters). GC-MS was carried out on a Thermo Electron PolarisQ Series GC-MS instrument using electron ionisation (EI).

#### Infra-Red (IR) Spectroscopy and Melting Point Determination

IR data was collected on a Perkin Elmer Spectrum One FTIR spectrometer and the data processed using Spectrum One software. Only major absorbances are reported. Melting points of solids were obtained using a Stuart automatic SMP40 melting point machine.

## Centrifugation

Centrifugation was carried out on three different instruments depending on the sample volume and required speed. For sample volumes of 2 mL or less, an Eppendorf Centrifuge 5415R with a 45° fixed angle FA-45-24-11 rotor. For samples above 2 mL, a ThermoScientific Heraeus Multifuge X3R using a TX-750 rotor fitted with round buckets or plate carriers as required. For the centrifugation of samples above 2 mL requiring centrifugation at 100 000 g, a Beckmann Coulter Avanti centrifuge was used with a JA 30.50 rotor.

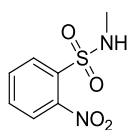
## DNA and protein concentration determination

DNA and protein concentrations were determined using a Trinean DropSense™ 16 instrument. For calculation of protein concentrations, the following mass extinction coefficient (E1%) values were used:

Protein	E1%
NovO, SelMet-NovO and NovO mutants	11.31
CouO and CouO mutants	11.69
SalL	8.11
SAHH	11.12
MTAN	2.86
LuxS	6.68

## 6.2 Synthetic procedures

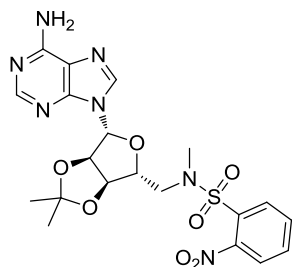
### *N*-methyl-2-nitrobenzenesulfonamide (106)<sup>[242]</sup>



To an ice cooled solution of 2-nitrobenzene-1-sulfonyl chloride (5.00 g, 22.6 mmol) in dichloromethane (50 mL) was added methylamine (33% in EtOH, 6.93 mL, 56.4 mmol) dropwise. The reaction was warmed to 22 °C and stirred for 30 minutes before washing with aq. 1 M HCl (50 mL), aq. sat. Na<sub>2</sub>CO<sub>3</sub> (50 mL) and drying over MgSO<sub>4</sub>. The filtrate was concentrated under reduced pressure and the residue recrystallised from EtOH to provide the title compound as a pale yellow crystalline solid (4.39 g, 90%). <sup>1</sup>H NMR: δ (400 MHz, CDCl<sub>3</sub>) 8.14 (m, 1H, Ar-H), 7.87 (m, 1H, Ar-H), 7.75 (m, 2H, Ar-H), 5.22 (br. s, 1H, NH),

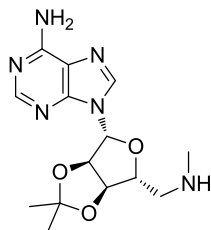
2.80 (d,  $J=5.4$  Hz, 3H, CH<sub>3</sub>). <sup>13</sup>C NMR:  $\delta$  (101 MHz, CDCl<sub>3</sub>) 148.3 (C), 133.7 (CH), 132.7 (CH), 132.5 (C), 131.5 (CH), 125.5 (CH), 29.8 (CH<sub>3</sub>). **Mpt.:** 112-114 °C (lit. 106 °C<sup>[350]</sup>).  **$\nu_{\max}$  (neat):** 3335, 1534, 1338, 1167, 730 cm<sup>-1</sup>.  **$m/z$ :** (ES<sup>+</sup>) 217 (M+H<sup>+</sup>, 60%), 186 (100%).  **$R_f$ :** 0.92 (10% MeOH in dichloromethane).

***N*-(((3*aR*,4*R*,6*R*,6*aR*)-6-(6-amino-9*H*-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)methyl)-*N*-methyl-2-nitrobenzenesulfonamide (108)**<sup>[242]</sup>



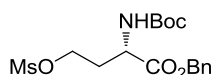
To an ice cooled solution of ((3*aR*,4*R*,6*R*,6*aR*)-6-(6-amino-9*H*-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)methanol (**107**) (3.00 g, 9.76 mmol), *N*-methyl-2-nitrobenzenesulfonamide (**106**) (2.32 g, 10.7 mmol) and triphenylphosphine (2.82 g, 10.7 mmol) in THF (90 mL) was added diethyl azodicarboxylate (0.77 mL, 4.88 mmol) dropwise. The reaction was gradually warmed to 22 °C and stirred for 16 hours before adding a further portion of triphenylphosphine (1.28 g, 4.88 mmol) and diethyl azodicarboxylate (0.77 mL, 4.88 mmol) and stirring for a further 3 hours. The reaction was concentrated under reduced pressure and the residue triturated with hot MeOH. The resulting material was purified by column chromatography (0-10% MeOH in dichloromethane) to provide the title compound as a colourless amorphous solid (2.02 g, 41%). <sup>1</sup>H NMR:  $\delta$  (400 MHz, CDCl<sub>3</sub>) 8.36 (s, 1H, Ar-H), 7.84 (s, 1H, Ar-H), 7.85 (dd,  $J_1=7.8$ ,  $J_2=1.2$  Hz, 1H, Ar-H), 7.48-7.62 (m, 3H, 3xAr-H), 6.03 (d,  $J=2.0$  Hz, 1H, CH), 5.53 (br. s, 2H, NH<sub>2</sub>), 5.46 (dd,  $J=6.4$ ,  $J_2=2.0$  Hz, 1H, CH), 5.13 (dd,  $J=6.4$ ,  $J_2=3.7$  Hz, 1H, CH), 4.43 (dt,  $J=7.9$ ,  $J_2=4.2$  Hz, 1H, CH), 3.68 (dd,  $J=14.8$ ,  $J_2=4.8$  Hz, 1H, CH<sub>2</sub>), 3.50 (dd,  $J=14.8$ ,  $J_2=7.9$  Hz, 1H, CH<sub>2</sub>), 2.85 (s, 3H, CH<sub>3</sub>), 1.60 (s, 3H, CH<sub>3</sub>), 1.38 (s, 3H, CH<sub>3</sub>). <sup>13</sup>C NMR:  $\delta$  (101 MHz, DMSO-*d*<sub>6</sub>) 156.6 (C), 153.2 (CH), 149.1 (C), 148.2 (C), 140.6 (CH), 135.0 (C), 132.7 (CH), 130.8 (CH), 130.2 (CH), 124.7 (CH), 119.7 (C), 114.1 (C), 89.4 (CH), 84.8 (CH), 83.7 (CH), 82.3 (CH), 51.8 (CH<sub>2</sub>), 36.1 (CH<sub>3</sub>), 27.4 (CH<sub>3</sub>), 25.7 (CH<sub>3</sub>). **Mpt.:** 228-230 °C.  **$\nu_{\max}$  (neat):** 1678, 1609, 1544, 1350, 1172, 767, 695 cm<sup>-1</sup>.  **$m/z$ :** (ES<sup>+</sup>) 506 (M+H<sup>+</sup>, 100%).  **$R_f$ :** (10% MeOH in dichloromethane) 0.37.

**9-((3a*R*,4*R*,6*R*,6a*R*)-2,2-dimethyl-6-((methylamino)methyl)tetrahydrofuro[3,4-  
d][1,3]dioxol-4-yl)-9H-purin-6-amine (109)**<sup>[242]</sup>



To a solution of **108** (1.00 g, 1.98 mmol) and cesium carbonate (1.29 g, 3.96 mmol) in acetonitrile (35 mL) was added thiophenol (0.41 mL, 3.96 mmol). The reaction was stirred for 16 hours before adding aq. 1 M NaOH (20 mL) and extracting the reaction mixture with dichloromethane (2x100 mL). The combined organic fractions were washed with brine (100 mL), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure. The resulting yellow oil was purified by column chromatography (100% dichloromethane then 10% aq. 7 N NH<sub>3</sub> in MeOH) to provide the title compound as a pale yellow foam (0.62 g, 98%). **<sup>1</sup>H NMR:** δ (400 MHz, CDCl<sub>3</sub>) 8.35 (s, 1H, Ar-H), 7.91 (s, 1H, Ar-H), 6.01 (d, *J*=3.2 Hz, 1H, CH), 5.80 (br. s, 2H, NH<sub>2</sub>), 5.48 (d, *J*=3.2 Hz, 1H, CH), 5.02 (dd, *J*<sub>1</sub>=3.4, *J*<sub>2</sub>=6.4 Hz, 1H, CH), 4.37 (dd, *J*<sub>1</sub>=1.7, *J*<sub>2</sub>=3.4 Hz, 1H, CH), 2.87 (t, *J*=5.9 Hz, 2H, CH<sub>2</sub>), 2.43 (s, 3H, CH<sub>3</sub>), 1.62 (s, 3H, CH<sub>3</sub>), 1.39 (s, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:** δ (101 MHz, CDCl<sub>3</sub>) 155.6 (C), 153.2 (CH), 149.5 (C), 139.9 (CH), 120.5 (C), 114.6 (C), 90.9 (CH), 85.3 (CH), 83.4 (CH), 82.4 (CH), 53.6 (CH<sub>2</sub>), 36.6 (CH<sub>3</sub>), 27.3 (CH<sub>3</sub>), 25.4 (CH<sub>3</sub>). **Mpt.:** 134.4-135.0 °C. **v<sub>max</sub> (neat):** 3136, 1679, 1605, 1208, 1074. **m/z:** (ES<sup>+</sup>) 321 (M+H<sup>+</sup>, 100%). **R<sub>f</sub>:** (10% aq. 7 N NH<sub>3</sub> in MeOH in dichloromethane) 0.24.

**(*S*)-Benzyl 2-((*tert*-butoxycarbonyl)amino)-4-((methylsulfonyl)oxy)butanoate (111)**<sup>[242]</sup>

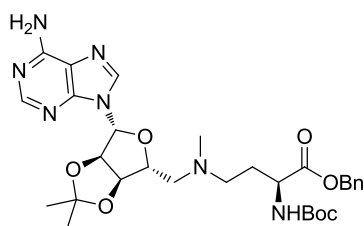


A solution of (*S*)-4-(benzyloxy)-3-((*tert*-butoxycarbonyl)amino)-4-oxobutanoic acid (3.00 g, 9.28 mmol) in anhydrous THF (90 mL) was cooled to -20 °C and 4-methylmorpholine (1.02 mL, 9.28 mmol), then isobutyl chloroformate (1.20 mL, 9.28 mmol) was added dropwise. The reaction was stirred for 10 minutes before adding NaBH<sub>4</sub> (1.05 g, 27.8 mmol) portionwise over 5 minutes. Anhydrous MeOH (90 mL) was then added over 5 minutes and the reaction was stirred for 10 minutes before concentrating under reduced pressure. The residue was partitioned between EtOAc (30 mL) and water (10 mL) and the organic fraction



washed with water (2x30 mL), sat. aq. NaHCO<sub>3</sub> (30 mL) and brine (30 mL), dried over MgSO<sub>4</sub> and concentrated under reduced pressure. The residue was dissolved in dichloromethane (75 mL) and cooled to 0 °C. Triethylamine (2.59 mL, 18.6 mmol) was added, followed by methanesulfonic anhydride (4.04 g, 23.2 mmol) and the reaction was stirred for 30 minutes whilst warming to 22 °C. The reaction was concentrated under reduced pressure and the resulting oil was purified by column chromatography (15-60% EtOAc in heptane) to provide the title compound as a colourless oil (2.74 g, 76%). <sup>1</sup>H NMR: δ (400 MHz, CDCl<sub>3</sub>) 7.28-7.45 (m, 5H, Ar-H), 5.19 (s, 2H, CH<sub>2</sub>), 4.47 (dd, *J*<sub>1</sub>=2.3, *J*<sub>2</sub>=1.3 Hz, 1H, CH), 4.20-4.38 (m, 2H, CH<sub>2</sub>), 2.95 (s, 3H, CH<sub>3</sub>), 2.23-2.38 (m, 1H, CH<sub>2</sub>), 2.14 (ddd, *J*<sub>1</sub>=14.5, *J*<sub>2</sub>=7.3, *J*<sub>3</sub>=5.5 Hz, 1H, CH<sub>2</sub>), 1.43 (s, 9H, 3xCH<sub>3</sub>). *NH* signal not detected. <sup>13</sup>C NMR: δ (101 MHz, CDCl<sub>3</sub>) 171.6 (CO), 155.7 (CO), 135.0 (C), 128.7 (4xCH), 128.5 (CH), 80.8 (C), 67.7 (CH<sub>2</sub>), 65.8 (CH), 50.5 (CH<sub>2</sub>), 37.2 (CH<sub>3</sub>), 31.8 (CH<sub>2</sub>), 28.2 (3xCH<sub>3</sub>). *v*<sub>max</sub> (neat): 3369, 1740, 1711, 1352, 1172, 733 cm<sup>-1</sup>. *m/z*: (ES<sup>+</sup>) 388 (M+H<sup>+</sup>, 6%), 288 (100%). *R*<sub>f</sub>: (1:1 EtOAc/heptane) 0.35. [α]<sub>D</sub><sup>20</sup>: -8.90°.

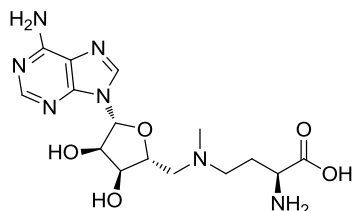
**(*S*)-Benzyl 4-(((3*aR*,4*R*,6*R*,6*aR*)-6-(6-amino-9*H*-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)methyl)(methyl)amino)-2-((*tert*-butoxycarbonyl)amino)butanoate (**112**)<sup>[242]</sup>**



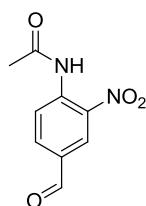
To a solution of **109** (1.00 g, 3.12 mmol) in acetonitrile (30 mL) was added tetrabutylammonium iodide (1.15 g, 3.12 mmol) and DIPEA (0.55 mL, 3.12 mmol). To this was added a solution of **111** (1.21 g, 3.12 mmol) in acetonitrile (10 mL) and the reaction mixture was stirred at 70 °C for 24 hours, before adding a further portion **111** (0.60 g, 1.56 mmol) in acetonitrile (5 mL) and stirring at 70 °C for a further 16 hours. The reaction mixture was concentrated under reduced pressure and redissolved in dichloromethane (50 mL). The solution was washed with water (2x20 mL) and the organic fraction washed with brine (20 mL) and dried over a hydrophobic frit. The filtrate was concentrated under reduced pressure and purified by column chromatography (10% MeOH in dichloromethane) to provide the title compound as a colourless oil (0.84 g, 44%). <sup>1</sup>H NMR: δ (400 MHz, MeOD-*d*<sub>4</sub>) 8.40 (s, 1H, Ar-H), 8.39 (s, 1H, Ar-H), 7.31-7.36 (m, 5H, Ar-H), 6.36 (d,

$J=2.1$  Hz, 1H, CH), 5.39 (dd,  $J_1=6.2$ ,  $J_2=2.1$  Hz, 1H, CH), 5.15 (d,  $J=2.1$  Hz, 2H, CH), 5.12 (dd,  $J_1=6.2$ ,  $J_2=4.0$  Hz, 1H, CH), 4.66 (dt,  $J_1=10.2$ ,  $J_2=3.2$  Hz, 1H, CH), 4.14 (m, 1H, CH), 3.80 (m, 1H, CH<sub>2</sub>), 3.58 (m, 1H, CH<sub>2</sub>), 2.86 (s, 3H, CH<sub>3</sub>), 2.20 (m, 1H, CH<sub>2</sub>), 1.93 (m, 1H, CH<sub>2</sub>), 1.63 (s, 3H, CH<sub>3</sub>), 1.38 (s, 12H, 4xCH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, CDCl<sub>3</sub>) 172.5 (CO), 158.0 (CO), 153.3 (C), 149.4 (C), 147.5 (CH), 144.4 (CH), 137.0 (C), 129.6 (2xCH), 129.5 (2xCH), 129.4 (CH), 121.2 (C), 116.5 (C), 92.3 (CH), 85.5 (CH), 83.9 (CH), 82.9 (CH), 81.1 (C), 68.4 (CH<sub>2</sub>), 58.7 (CH<sub>2</sub>), 52.7 (CH), 28.6 (CH<sub>3</sub>), 27.4 (CH<sub>3</sub>), 25.5 (CH<sub>3</sub>).  **$\nu_{\max}$  (neat):** 1673, 1178, 1131, 1079, 721 cm<sup>-1</sup>.  **$m/z$ :** (ES<sup>+</sup>) 612 (M+H<sup>+</sup>, 100%).  **$R_f$ :** (5% MeOH in dichloromethane) 0.23.  **$[\alpha]_D^{20}$ :** -144.3°.

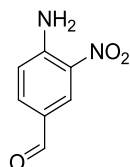
**(S)-2-amino-4-(((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(methylamino)butanoic acid (104)**<sup>[242]</sup>



To a solution of **112** (0.20 g, 0.33 mmol) in acetonitrile (1 mL) was added aq. 2 M NaOH (2.50 mL, 5.00 mmol). The reaction was stirred at 22 °C for 2 hours before neutralising to pH 6 with aq. conc. HCl and stirring for a 1 hour before adding a further portion of aq. conc. HCl (5 mL, 25.0 mmol) and stirring for a further 30 minutes. The reaction mixture was concentrated under reduced pressure and redissolved in MeOH (5 mL). The suspension was filtered and concentrated under reduced pressure. The resulting residue was treated with 4 M HCl in dioxane (2 mL, 8.00 mmol) and the solvent removed under reduced pressure to give a yellow solid. The solid was purified using Dowex-X8-50 ion exchange resin, eluting with aq. 1 M NH<sub>4</sub>OH to provide the title compound as an amorphous pale yellow solid (103 mg, 83%). **<sup>1</sup>H NMR:**  $\delta$  (400 MHz, D<sub>2</sub>O) 8.21 (s, 1H, Ar-H), 8.14 (s, 1H, Ar-H), 5.99 (d,  $J=4.9$  Hz, 1H, CH), 4.73 (app. t,  $J=5.1$  Hz, 1H, CH), 4.31 (m, 1H, CH), 4.22 (app. t,  $J=5.1$  Hz, 1H, CH), 3.66 (dd,  $J_1=5.4$ ,  $J_2=7.1$  Hz, 1H, CH), 2.92 (m, 2H, CH<sub>2</sub>), 2.77 (t,  $J=6.7$ , 2H, CH<sub>2</sub>), 2.37 (s, 3H, CH<sub>3</sub>), 1.89-2.05 (m, 2H, CH<sub>2</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, D<sub>2</sub>O) 174.6 (CO), 155.6 (C), 152.8 (C), 148.7 (C), 140.3 (CH), 119.0 (CH), 88.2 (CH), 80.5 (CH), 73.1 (CH), 72.0 (CH), 59.1 (CH<sub>2</sub>), 54.4 (CH<sub>2</sub>), 54.2 (CH<sub>2</sub>), 40.8 (CH), 26.3 (CH<sub>3</sub>). **Mpt.:** decomp >178 °C.  **$\nu_{\max}$  (neat):** 3145, 1643, 1599, 1332, 1130, 1046 cm<sup>-1</sup>.  **$m/z$ :** (ES<sup>+</sup>) 382 (M+H<sup>+</sup>, 100%).  **$R_f$ :** (3:1:1 <sup>n</sup>BuOH:AcOH:H<sub>2</sub>O) 0.05.  **$[\alpha]_D^{20}$ :** -144.3°.

***N*-(4-formyl-2-nitrophenyl)acetamide (198)**<sup>[351]</sup>

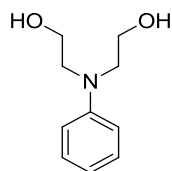
Sulfuric acid (0.65 mL, 12.3 mmol) was cooled to 0 °C before adding *N*-(4-formylphenyl)acetamide (0.20 g, 1.23 mmol) portionwise. After complete dissolution of the starting material, nitric acid (0.06 mL, 1.23 mmol) was added dropwise at 0 °C. The reaction was stirred for 10 minutes before pipetting the reaction mixture onto ice. The resulting yellow precipitate was collected by filtration and purified by column chromatography (100% dichloromethane) to provide the title compound as a yellow solid (89 mg, 35%). **<sup>1</sup>H NMR:** δ (400 MHz, CDCl<sub>3</sub>) 10.62 (br. s, 1H, NH), 9.99 (s, 1H, COH), 9.03 (d, *J*=9 Hz, 1H, Ar-H), 8.74 (d, *J*=2 Hz, 1H, Ar-H), 8.15 (dd, *J*<sub>1</sub>=9, *J*<sub>2</sub>=2 Hz, 1H, Ar-H), 2.35 (s, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:** δ (101 MHz, CDCl<sub>3</sub>) 188.8 (CO), 169.2 (CO), 139.5 (C), 135.6 (CH), 130.9 (C), 128.0 (CH), 122.3 (CH), 25.8 (CH<sub>3</sub>). **Mpt.:** 155.6-156-2 °C. **v<sub>max</sub> (neat):** 3349, 1687, 1573, 1512, 1193, 835, 657 cm<sup>-1</sup>. **m/z:** (ES<sup>+</sup>) 167 ([M-Ac]+H<sup>+</sup>, 100%). **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.16.

**4-Amino-3-nitrobenzaldehyde (199)**<sup>[351]</sup>

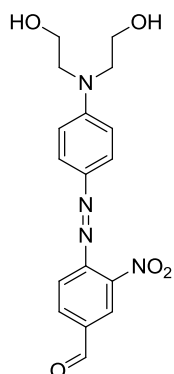
*N*-(4-formyl-2-nitrophenyl)acetamide (**198**) (50.0 mg, 0.24 mmol) was dissolved in methanol (4 mL) and aq. 1 M hydrochloric acid (4 mL, 4.00 mmol) was added. The reaction was heated to reflux and stirred for 1 hour, before cooling to 22 °C and neutralising with aq. 2 M NaOH. The reaction was extracted with dichloromethane (3x5 mL) and the combined organic fractions were washed with brine (5 mL), dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated under reduced pressure to provide the title compound as an amorphous yellow solid (39 mg, 98%). **<sup>1</sup>H NMR:** δ (400 MHz, CDCl<sub>3</sub>) 9.82 (s, 1H, CHO), 8.63 (s, 1H, Ar-H), 7.92 (dd, *J*<sub>1</sub>=8.7, *J*<sub>2</sub>=1.8 Hz, 1H, Ar-H), 6.91 (d, *J*=8.7 Hz, 1H, Ar-H), 6.40-6.79 (br. s, 2H, NH<sub>2</sub>). **<sup>13</sup>C NMR:** δ (101 MHz, CDCl<sub>3</sub>) 188.7 (CO), 148.5 (C), 133.7 (CH), 131.6 (CH),

126.6 (C), 119.5 (CH). *One C signal not detected.* **Mpt.:** 189.7-190.2 °C.  $\nu_{\text{max}}$  (**neat**): 3319, 1557, 1251, 830, 724  $\text{cm}^{-1}$ . **m/z:** ( $\text{ES}^+$ ) 167 ( $\text{M}+\text{H}^+$ , 65%), 149 (100%). **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.33.

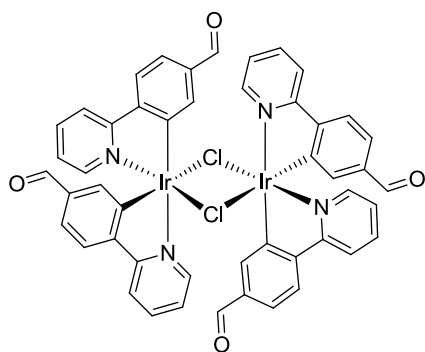
**2,2'-(Phenylazanediyl)diethanol (196)**<sup>[351]</sup>



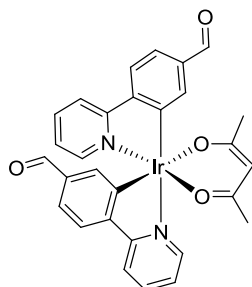
To a flask containing water (14 mL) was added aniline (0.91 mL, 10.0 mmol), calcium carbonate (2.00 g, 20.0 mmol), 2-chloroethanol (1.41 mL, 21.0 mmol) and potassium iodide (166 mg, 1.00 mmol). The reaction mixture was heated to reflux for 16 hours before cooling to 22 °C and filtering. The filtrate was extracted with EtOAc (3x15 mL) and the combined organic layers were washed with brine (15 mL), dried over  $\text{Na}_2\text{SO}_4$ , filtered and concentrated under reduced pressure. The residue was purified by column chromatography (100% EtOAc) to provide the title compound as an off-white crystalline solid (489 mg, 27%). **<sup>1</sup>H NMR:**  $\delta$  (400 MHz,  $\text{CDCl}_3$ ) 7.23 (dd,  $J_1=8.8$ ,  $J_2=7.3$  Hz, 2H, Ar-H), 6.67-6.78 (m, 3H, A-H), 3.85 (t,  $J=4.9$  Hz, 4H, 2x $\text{CH}_2$ ), 3.57 (t,  $J=4.9$  Hz, 4H, 2x $\text{CH}_2$ ). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz,  $\text{CDCl}_3$ ) 147.9 (C), 129.3 (2xCH), 117.0 (CH), 112.7 (2xCH), 60.8 (2x $\text{CH}_2$ ), 55.4 (2x $\text{CH}_2$ ). **Mpt.:** 58.8-60.4 °C.  $\nu_{\text{max}}$  (**neat**): 3324, 1595, 1504, 1185, 1027, 752  $\text{cm}^{-1}$ . **m/z:** ( $\text{ES}^+$ ) 182 ( $\text{M}+\text{H}^+$ , 100%). **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.12.

**4-((4-(Bis(2-hydroxyethyl)amino)phenyl)diazenyl)-3-nitrobenzaldehyde (183)**<sup>[351]</sup>


A solution of sodium nitrite (20.1 mg, 0.29 mmol) in water (1.70 mL) was added dropwise over 5 minutes to an ice cooled solution of 4-amino-3-nitrobenzaldehyde (**199**) (50.0 mg, 0.30 mmol) in aq. 6 M HCl (3 mL, 18.0 mmol). The reaction was stirred at 0 °C for 2 hours and then added dropwise over 5 minutes to an ice cooled solution of **196** (54.5 mg, 0.30 mmol) in ethanol (8 mL). The reaction was warmed to 22 °C and stirred for 18 hours before concentrating under reduced pressure. The residue was partitioned between sat. aq. NaHCO<sub>3</sub> (15 mL) and dichloromethane (20 mL). The aqueous layer was extracted with dichloromethane (3x20 mL) and the combined organic fractions dried over Na<sub>2</sub>SO<sub>3</sub> and purified by column chromatography (100% EtOAc) to provide the title compound as a dark red solid (63 mg, 58%). **<sup>1</sup>H NMR:** δ (400 MHz, CDCl<sub>3</sub>) 10.07 (s, 1H, CHO), 8.32 (d, *J*=1.6 Hz, 1H, Ar-H), 8.11 (dd, *J*=8.3, 1.6 Hz, 1H, Ar-H), 7.89 (d, *J*=9.2 Hz, 2H, Ar-H), 7.86 (d, *J*=8.2 Hz, 1H, Ar-H), 6.79 (d, *J*=9.2 Hz, 2H, Ar-H), 3.98 (t, *J*=4.9 Hz, 4H, 2xCH<sub>2</sub>), 3.76 (t, *J*=4.9 Hz, 4H, 2xCH<sub>2</sub>). **<sup>13</sup>C NMR:** δ (101 MHz, CDCl<sub>3</sub>) 189.3 (CO), 152.3 (C), 149.2 (C), 147.4 (C), 144.6 (C), 135.6 (C), 132.9 (CH), 126.9 (CH), 125.3 (CH), 119.5 (2xCH), 112.3 (2xCH), 60.6 (2xCH<sub>2</sub>), 55.1 (2xCH<sub>2</sub>). **Mpt.:** 117.9-119.4 °C. **v<sub>max</sub> (neat):** 1735, 1591, 1512, 1308, 1039, 821 cm<sup>-1</sup>. **m/z:** (ES<sup>+</sup>) 359 (M+H<sup>+</sup>, 100%). **R<sub>f</sub>:** (EtOAc) 0.19.

**Ir(III) (di(4-(pyridin-2-yl)benzaldehyde)) chloro-bridged dimer (191)**<sup>[302]</sup>

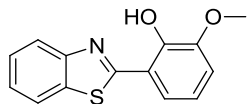
To a solution of Ir(III)Cl<sub>3</sub>.xH<sub>2</sub>O (352 mg, 1.00 mmol) in water and 2-ethoxy ethanol (1:3 v/v%, 8 mL) at 22 °C was added 4-(pyridin-2-yl)benzaldehyde (458 mg, 2.50 mmol). The reaction was heated to reflux and stirred for 24 hours before cooling to 22 °C. The precipitate was collected by filtration and washed with EtOH:H<sub>2</sub>O (95:1, 25 mL) to provide the title compound as an orange solid, which was taken through to the next stage of the synthesis without further purification.

**Ir(III) (di(4-(pyridin-2-yl)benzaldehyde))(acac) (181)**<sup>[302]</sup>

To a mixture of **191** (theoretical amount 50.0 mg, 0.04 mmol) and sodium carbonate (30.4 mg, 0.29 mmol) in degassed 2-ethoxy ethanol (1 mL) was added pentane-2,4-dione (10.5 μL, 0.10 mmol). The reaction was stirred at 90 °C for 1 hour under a N<sub>2</sub> atmosphere before cooling to 22 °C and collecting the precipitate by filtration. The solid was washed with ice-cooled ethanol and dried under N<sub>2</sub> to provide the title compound as red amorphous solid (50.0 mg, 70% over two steps). <sup>1</sup>H NMR: δ (400 MHz, CDCl<sub>3</sub>) 9.64 (s, 2H, 2xCHO), 8.57 (d, *J*=5.2 Hz, 2H, 2xAr-H), 8.00 (d, *J*=8.1 Hz, 2H, 2xArH), 7.81-7.92 (m, 2H, 2xAr-H), 7.70 (d, *J*=8.1 Hz, 2H, 2xAr-H), 7.28-7.40 (m, 4H, 4xAr-H), 6.70 (d, *J*=1.4 Hz, 2H, 2xAr-H), 5.26 (s, 1H, CH), 1.80 (s, 6H, 2xCH<sub>3</sub>). <sup>13</sup>C NMR: δ (101 MHz, DMSO-d<sub>6</sub>) 193.1 (2xCOH), 184.4 (2xCO), 165.7 (2xC), 151.7 (2xC), 148.3 (2xCH), 147.1 (2xC), 138.6 (2xCH), 135.0 (2xC), 132.1 (2xCH), 124.3 (4xCH), 123.6 (2xCH), 120.7 (2xCH), 100.4

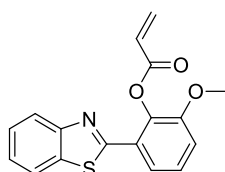
(CH), 28.1 (2xCH<sub>3</sub>). **Mpt.:** decomp. >290 °C.  $\nu_{\max}$  (**neat**): 1682, 1562, 1514, 1474, 1395, 1193, 778 cm<sup>-1</sup>. **m/z:** (ES<sup>+</sup>) 656 (M+H<sup>+</sup>, 100%). **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.14.

**2-(Benzo[d]thiazol-2-yl)-6-methoxyphenol (187)**<sup>[297]</sup>



To a solution of 2-hydroxy-3-methoxybenzaldehyde (0.48 g, 3.15 mmol) in ethanol (10 mL) was added 2-aminobenzenethiol (0.45 mL, 4.20 mmol), hydrogen peroxide (1.93 mL, 18.9 mmol) and aq. conc. hydrochloric acid (0.78 mL, 9.45 mmol). The reaction mixture was stirred at 22 °C for 90 minutes before quenching with water (10 mL). The resulting precipitate was collected by filtration, dried under vacuum and recrystallised from ethanol to provide the title compound as a mauve crystalline solid (477 mg, 59%). **<sup>1</sup>H NMR:**  $\delta$  (400 MHz, CDCl<sub>3</sub>) 12.72 (br. s, 1H, OH), 8.02 (d,  $J=8.1$  Hz, 1H, Ar-H), 7.90 (d,  $J=8$  Hz, 1H, Ar-H), 7.51 (td,  $J_1=8$ ,  $J_2=1$  Hz, 1H, Ar-H), 7.41 (dt,  $J_1=8$ ,  $J_2=1$  Hz, 1H, Ar-H), 7.34 (dd,  $J_1=8$ ,  $J_2=1$  Hz, 1H, Ar-H), 6.99 (dd,  $J_1=8$ ,  $J_2=1$  Hz, 1H, Ar-H), 6.91 (t,  $J=8$  Hz, 1H, Ar-H), 3.96 (s, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, CDCl<sub>3</sub>) 169.4 (C), 151.7 (C), 149.0 (C), 148.3 (C), 132.7 (C), 126.7 (CH), 125.6 (CH), 122.3 (CH), 121.5 (CH), 120.1 (CH), 119.2 (CH), 116.8 (C), 114.2 (CH), 56.3 (CH<sub>3</sub>). **Mpt.:** 163.9-164.9 °C.  $\nu_{\max}$  (**neat**): 1459, 1246, 1003, 756, 724, 702 cm<sup>-1</sup>. **m/z:** (ES<sup>+</sup>) 258 (M+H<sup>+</sup>, 100%). **R<sub>f</sub>:** (EtOAc) 0.67.

**2-(Benzothiazol-2-yl)-6-methoxyphenyl acrylate (182)**<sup>[297]</sup>



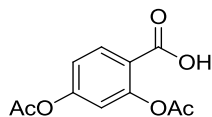
To an ice cooled solution of **187** (150 mg, 0.58 mmol) in anhydrous dichloromethane (10 mL) was added triethylamine (0.16 mL, 1.17 mmol) and a solution of acryloyl chloride (0.06 mL, 0.73 mmol) in dichloromethane (4 mL) dropwise. The reaction was stirred at 0 °C for 90 minutes before warming to 22 °C and stirring for a further 4 hours. The reaction mixture was diluted with dichloromethane (10 mL) and washed three times with water (3x10 mL). The organic fraction was dried over a hydrophobic frit and concentrated under reduced pressure to provide a brown oil, which was triturated with TBME (10 mL) to provide the title compound as a pale brown solid (125 mg, 69%). **<sup>1</sup>H NMR:**  $\delta$  (400 MHz,

CDCl<sub>3</sub>) 8.08 (d,  $J=8.0$  Hz, 1H, Ar-H), 7.99 (dd,  $J_1=8$ ,  $J_2=1$  Hz, 1H, Ar-H), 7.90 (d,  $J=8$  Hz, 1H, Ar-H), 7.49 (td,  $J_1=8$ ,  $J_2=1$  Hz, 1H, Ar-H), 7.31-7.44 (m, 2H, 2xAr-H), 7.11 (dd,  $J_1=8$ ,  $J_2=1$  Hz, 1H, Ar-H), 6.73 (dd,  $J_1=17$ ,  $J_2=1$  Hz, 1H, CH<sub>2</sub>), 6.51 (dd,  $J_1=17$ ,  $J_2=11$  Hz, 1H, CH), 6.13 (dd,  $J_1=11$ ,  $J_2=1$  Hz, 1H, CH<sub>2</sub>), 3.89 (s, 3H, CH<sub>3</sub>). <sup>13</sup>C NMR:  $\delta$  (101 MHz, CDCl<sub>3</sub>) 163.3 (CO), 162.1 (C), 152.7 (C), 151.9 (C), 138.0 (C), 135.6 (C), 133.3 (CH<sub>2</sub>), 127.6 (CH), 127.5 (C), 126.7 (CH), 126.2 (CH), 125.3 (CH), 123.4 (CH), 121.3 (2xCH), 114.1 (CH), 56.3 (CH<sub>3</sub>).  $\nu_{\max}$  (neat): 1745, 1270, 1237, 1125, 991, 756 cm<sup>-1</sup>. **Mpt.:** 116.5-117.8 °C. ***m/z*:** (ES<sup>+</sup>) 312 (M+H<sup>+</sup>, 70%), 258 (100%). ***R<sub>f</sub>*:** (EtOAc) 0.65.

### General procedure for the synthesis of 2, 4-diacetoxybenzoic acid and 2-acetoxy 4-methoxybenzoic acid

To a solution of 2, 4-dihydroxybenzoic acid or 2-hydroxy 4-methoxybenzoic acid (1 eq.) in tetrahydrofuran (10 mL/g) was added triethylamine (3.5 eq.), 4-dimethylaminopyridine (0.1 eq.) and acetic anhydride (5 eq.) and the reaction was stirred at 22 °C for 4 hours. The reaction mixture was acidified with aq. 1 M HCl to pH 3 and extracted with EtOAc (3x100 mL). The combined organic fractions were washed with brine (100 mL) and dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was removed under reduced pressure and the resulting solid was recrystallised from EtOAc to provide the title compounds.

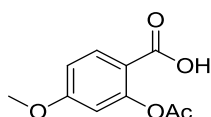
### 2,4-Diacetoxybenzoic acid (93)<sup>[352]</sup>



Colourless crystalline solid (12.6 g, 82%).

<sup>1</sup>H NMR:  $\delta$  (400 MHz, DMSO-d<sub>6</sub>) 13.12 (br. s, 1H, COOH), 7.97 (d,  $J=8.6$  Hz, 1H, Ar-H), 7.19 (dd,  $J_1=8.6$ ,  $J_2=2.3$  Hz, 1H, Ar-H), 7.09 (d,  $J=2.3$  Hz, 1H, Ar-H), 2.29 (s, 3H, CH<sub>3</sub>), 2.24 (s, 3H, CH<sub>3</sub>). <sup>13</sup>C NMR:  $\delta$  (101 MHz, DMSO-d<sub>6</sub>) 168.9 (CO), 168.6 (CO), 164.9 (CO), 154.0 (C), 151.0 (C), 132.4 (C), 121.6 (CH), 119.6 (CH), 117.6 (CH), 20.8 (CH<sub>3</sub>), 21.0 (CH<sub>3</sub>). **Mpt.:** 135-137 °C (lit. 136-137 °C).<sup>[352]</sup>  $\nu_{\max}$  (neat): 2859, 1773, 1680, 1607 cm<sup>-1</sup>. ***m/z*:** (ES<sup>+</sup>) 261 (M+Na<sup>+</sup>, 5%), 179 (100%). ***R<sub>f</sub>*:** (1:1 EtOAc/heptane) 0.16.



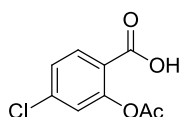
**2-Acetoxy 4-methoxybenzoic acid (137)**<sup>[353]</sup>

Colourless crystalline solid (4.96 g, 79%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO- $d_6$ ) 12.71 (br. s, 1H, COOH), 7.90 (d,  $J=8.9$  Hz, 1H, Ar-H), 6.93 (dd,  $J_1=8.9$ ,  $J_2=2.5$  Hz, 1H, Ar-H), 6.77 (d,  $J=2.5$  Hz, 1H, Ar-H), 3.82 (s, 3H, CH<sub>3</sub>), 2.23 (s, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO- $d_6$ ) 169.0 (CO), 165.1 (CO), 163.4 (C), 152.2 (C), 133.0 (C), 115.9 (CH), 111.7 (CH), 109.2 (CH), 55.8 (CH<sub>3</sub>), 20.9 (CH<sub>3</sub>). **Mpt.:** 144-145 °C.  **$\nu_{\max}$  (neat):** 2936, 1760, 1694, 1611, 1335, 1210  $\text{cm}^{-1}$ .  **$m/z$ :** (ES<sup>+</sup>) 211 (M+H<sup>+</sup>, 2%), 151 (100%). **R<sub>f</sub>:** (EtOAc) 0.62.

**General procedure for the synthesis of 2-acetoxy 4-chlorobenzoic acid and 2-acetoxy 4-nitrobenzoic acid**

To an ice-cooled solution of 4-chloro-2-hydroxybenzoic acid or 4-nitro-2-hydroxybenzoic acid (1 eq.) in tetrahydrofuran (20 mL) was added triethylamine (3 eq.) and acetyl chloride (2 eq.) dropwise and the reaction mixture was stirred at 22 °C for 16 hours. The reaction mixture was acidified with aq. 2 M HCl to pH 2 and extracted with EtOAc (3x15 mL). The combined organic fractions were washed with brine (15 mL) and dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was removed under reduced pressure and the resulting solid purified as described below.

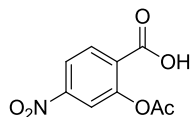
**2-Acetoxy 4-chlorobenzoic acid (135)**<sup>[354]</sup>

Purified by recrystallisation from EtOAc to provide the title compound as a colourless crystalline solid (0.54 g, 87%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, CDCl<sub>3</sub>) 8.06 (d,  $J=8.5$  Hz, 1H, Ar-H), 7.35 (dd,  $J_1=8.5$ ,  $J_2=2.0$  Hz, 1H, Ar-H), 7.18 (d,  $J=2.0$  Hz, 1H, Ar-H), 2.35 (s, 3H, CH<sub>3</sub>). *COOH signal not detected.* **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO- $d_6$ ) 168.9 (CO), 164.8 (CO), 150.9 (C), 137.6 (C), 132.8 (CH), 126.3 (CH), 124.1 (CH), 123.1 (C), 20.8 (CH<sub>3</sub>). **Mpt.:** 147 °C (lit.: 133-135 °C).<sup>[354]</sup>

$\nu_{\max}$  (neat): 1763, 1686, 1191, 936  $\text{cm}^{-1}$ .  $m/z$ : ( $\text{ES}^+$ ) 171 ( $\text{M}-\text{CO}_2^-+\text{H}^+$ , 75%), 372 (100%).  
 $R_f$ : (1:1 EtOAc/heptane) 0.22.

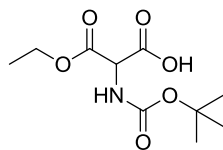
**2-Acetoxy 4-nitrobenzoic acid (136)**<sup>[355]</sup>



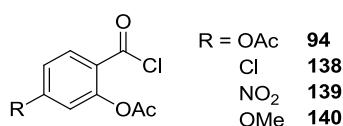
Purified by column chromatography (20-100% EtOAc in heptane) to provide the title compound as colourless amorphous solid (0.55 g, 89%).

$^1\text{H NMR}$ :  $\delta$  (400 MHz,  $\text{CDCl}_3$ ) 8.28 (d,  $J=8.6$  Hz, 1H, Ar-H), 8.19 (dd,  $J_1=8.6$ ,  $J_2=2.1$  Hz, 1H, Ar-H), 8.02 (d,  $J=2.1$  Hz, 1H, Ar-H), 2.39 (s, 3 H,  $\text{CH}_3$ ). *COOH signal not detected.*  
 $^{13}\text{C NMR}$ :  $\delta$  (101 MHz,  $\text{DMSO}-d_6$ ) 169.0 (CO), 166.9 (CO), 151.6 (C), 151.5 (C), 133.4 (CH), 127.9 (C), 120.8 (CH), 119.7 (CH), 20.9 ( $\text{CH}_3$ ). **Mpt.:** 146-147 °C.  $\nu_{\max}$  (neat): 1768, 1701, 1523, 1350  $\text{cm}^{-1}$ .  $m/z$ : ( $\text{ES}^-$ ) 224 ( $\text{M}-\text{H}^+$ , 7%), 182 (100%).<sup>[356]</sup>  $R_f$ : (EtOAc) 0.62.

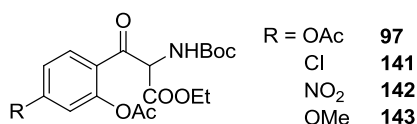
**2-(tert-Butoxycarbonylamino)-3-ethoxy-3-oxopropanoic acid (96)**<sup>[357]</sup>



To a solution of diethyl 2-((tert-butoxycarbonyl)amino)malonate (10.0 mL, 39.2 mmol) in ethanol (50 mL) was added solid potassium hydroxide (2.25 g, 40.1 mmol). The suspension was stirred at 22 °C for 12 hours before concentrating under reduced pressure. The suspension was redissolved in aq. 1 M  $\text{NaHCO}_3$  (75 mL) and washed with EtOAc (2x40 mL). The solution was cooled to 0 °C, acidified with solid  $\text{KHSO}_4$  to pH 2 and extracted with EtOAc (3x40 mL). The combined organic fractions were washed with brine (20 mL) and dried over  $\text{Na}_2\text{SO}_4$ . The solvent was removed under reduced pressure to provide the title compound as an amorphous colourless solid (7.66 g, 77%).  $^1\text{H NMR}$ :  $\delta$  (400 MHz,  $\text{DMSO}-d_6$ ) 13.41 (br. s, 1H, COOH), 7.47 (d,  $J=7.9$  Hz, 1H, NH), 4.71 (d,  $J=7.9$  Hz, 1H, CH), 4.15 (q,  $J=7.0$  Hz, 2H,  $\text{CH}_2$ ), 1.381 (s, 9H, 3x $\text{CH}_3$ ), 1.19 (t,  $J=7.0$  Hz, 3H,  $\text{CH}_3$ ).  $^{13}\text{C NMR}$ :  $\delta$  (101 MHz,  $\text{DMSO}-d_6$ ) 167.8 (CO), 167.2 (CO), 155.1 (CO), 78.9 (C), 61.3 ( $\text{CH}_2$ ), 57.6 (CH), 28.1 ( $\text{CH}_3$ ), 13.9 (3x $\text{CH}_3$ ). **Mpt.:** 84-85 °C (lit.: 93-95 °C [diethyl ether/hexane]).<sup>[357]</sup>  $\nu_{\max}$  (neat): 3264, 2984, 1756, 1733, 1652  $\text{cm}^{-1}$ .  $m/z$ : ( $\text{ES}^+$ ) 248 ( $\text{M}+\text{H}^+$ , 6%), 148 (100%).  $R_f$ : (EtOAc) 0.53.

**General procedure for acid chloride formation**

To a suspension of carboxylic acid (1 eq.) in anhydrous toluene (10 mL/g) under a N<sub>2</sub> atmosphere in a flame dried flask at 22 °C was added (COCl)<sub>2</sub> (2 eq.) and anhydrous DMF (0.1 eq.). The reaction mixture was heated to 40 °C and stirred for 2 hours. The reaction mixture was concentrated under reduced pressure and the resulting residue was used in the next stage of the synthesis without further purification.

**General procedure for the coupling of 2-(*tert*-butoxycarbonylamino)-3-ethoxy-3-oxopropanoic acid with acid chlorides 94 and 138-140**

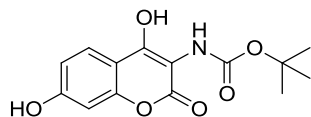
To an ice-cooled solution of **96** (1 eq.) in anhydrous tetrahydrofuran (8 mL/g) under an N<sub>2</sub> atmosphere was added triethylamine (5 eq.) and anhydrous magnesium chloride (3 eq.). The resulting slurry was stirred vigorously at 4 °C for 2 hours. A solution of the previously prepared acid chloride (theoretical amount 1 eq.) in anhydrous tetrahydrofuran (5 mL/mmol) was then added and the resulting suspension stirred for 2 hours. The reaction mixture was quenched with saturated aq. ammonium chloride (5 mL/mmol) and extracted with ethyl acetate (3x50 mL). The combined organic fractions were washed with brine (10 mL) and dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was removed under reduced pressure and the resulting residue was used in the next stage of the synthesis without further purification.

**General procedure for the synthesis of *N*-Boc aminocoumarin derivatives **98** and **144-146**<sup>[107]</sup>**

To a solution of crude coupling product **97** or **141-143** (theoretical amount 1 eq.) in methanol (10 mL/g) was added aq. 1 M NaOH (5 eq.) and the reaction mixture was stirred at 22 °C for 4 hours. The reaction mixture was acidified with aq. 2 M HCl to pH 3 and extracted with EtOAc (3x100 mL). The combined organic fractions were washed with brine (20 mL), dried over MgSO<sub>4</sub> and the solvent removed under reduced pressure. The resulting

solid was recrystallised from EtOAc or purified by column chromatography to provide the title compounds.

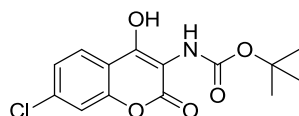
***tert*-Butyl (4,7-dihydroxy-2-oxo-2*H*-chromen-3-yl) carbamate (98)<sup>[107]</sup>**



Purified by recrystallisation from EtOAc and isolated as a pale orange crystalline solid (3.57 g, 58% over 3 stages from **93**).

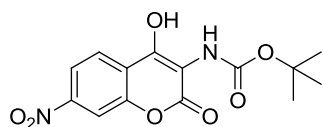
**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO-*d*<sub>6</sub>) 11.68 (br. s, 1H, OH), 10.51 (br. s, 1H, OH), 7.78 (br. s, 1H, NH), 7.66 (d,  $J=8.7$  Hz, 1H, Ar-H), 6.81 (dd,  $J_1=8.7$ ,  $J_2=2.2$  Hz, 1H, Ar-H), 6.69 (d,  $J=2.2$  Hz, 1H, Ar-H), 1.43 (s, 9H, 3xCH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO-*d*<sub>6</sub>) 161.3 (CO), 161.2 (C), 154.7 (C), 153.4 (C), 124.9 (CH), 112.8 (CH), 107.9 (C), 101.8 (CH), 100.49 (C), 78.6 (C), 28.2 (3xCH<sub>3</sub>). *One CO signal not detected.* **Mpt.:** 208-210 °C.  **$\nu_{\max}$  (neat):** 3370, 1612, 1575 cm<sup>-1</sup>. ***m/z*:** (ES<sup>+</sup>) 294 (M+H<sup>+</sup>, 9%), 238 (100%). ***R<sub>f</sub>*:** (1:1 EtOAc/heptane) 0.38.

***tert*-Butyl (7-chloro-4-hydroxy-2-oxo-2*H*-chromen-3-yl) carbamate (144)**



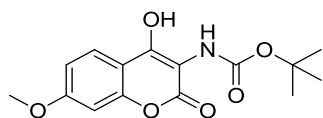
Purified by recrystallisation from EtOAc and isolated as a colourless crystalline solid (0.20 g, 28% over 3 stages over 3 stages from **135**).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO-*d*<sub>6</sub>) 12.24 (br. s, 1H, OH), 7.98 (br. s, 1H, NH), 7.86 (d, 1H,  $J=8.5$  Hz, Ar-H), 7.59 (d,  $J=2.0$  Hz, 1H, Ar-H), 7.43 (dd,  $J_1=8.5$ ,  $J_2=2.0$  Hz, 1H, Ar-H), 1.42 (s, 9H, 3xCH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO-*d*<sub>6</sub>) 159.6 (CO), 157.9 (CO), 154.0 (C), 151.5 (C), 135.8 (C), 124.8 (CH), 124.0 (CH), 115.8 (CH), 115.2 (C), 103.3 (C), 78.8 (C), 27.7 (3xCH<sub>3</sub>).  **$\nu_{\max}$  (neat):** 3287, 1705, 1602 cm<sup>-1</sup>. **Mpt.:** 162 °C. **HRMS:** C<sub>14</sub>H<sub>15</sub>ClN<sub>2</sub>O<sub>5</sub> [M+H<sup>+</sup>] requires *m/z* 312.0633, found 312.0643. ***R<sub>f</sub>*:** (1:1 EtOAc/heptane) 0.54.

***tert*-Butyl (7-nitro-4-hydroxy-2-oxo-2*H*-chromen-3-yl) carbamate (145)**

Purified by recrystallisation from EtOAc and isolated as a yellow crystalline solid (0.50 g, 70% over 3 steps from **136**).

**<sup>1</sup>H NMR**  $\delta$  (400 MHz, DMSO-*d*<sub>6</sub>) 12.58 (br. s, 1H, OH), 8.22 (d,  $J=2.2$  Hz, 1H, Ar-H), 8.18 (dd,  $J_1=2.2$ ,  $J_2=8.7$  Hz, 1H, Ar-H), 8.12 (s, 1H, NH) 8.09 (d,  $J=8.7$  Hz, 1H, Ar-H), 1.42 (s, 9H, 3xCH<sub>3</sub>). **<sup>13</sup>C NMR**  $\delta$  (101 MHz, DMSO-*d*<sub>6</sub>) 160.0 (CO), 157.9 (C), 154.1 (CO), 151.0 (C), 148.9 (C), 125.2 (CH), 121.8 (C), 118.7 (CH), 111.8 (CH), 105.5 (C), 79.1 (C), 28.1 (3xCH<sub>3</sub>). **Mpt.:** decomp. >210 °C.  **$\nu_{\max}$  (neat):** 3336, 1695, 1627, 1574, 1538, 1524, 1345, 706 cm<sup>-1</sup>. **HRMS:** C<sub>14</sub>H<sub>15</sub>N<sub>2</sub>O<sub>7</sub> [M+H<sup>+</sup>] requires  $m/z$  323.0874, found 323.0883. **R<sub>f</sub>:** (20% EtOAc in heptane) 0.23.

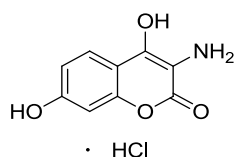
***tert*-Butyl (7-methoxy-4-hydroxy-2-oxo-2*H*-chromen-3-yl) carbamate (146)**

Purified by column chromatography (5-20% EtOAc in heptane) and isolated as a colourless crystalline solid (0.68 g, 23% over 3 steps from **137**).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO-*d*<sub>6</sub>) 11.82 (br. s, 1H, OH), 7.84 (br. s, 1H, NH), 7.75 (d,  $J=8.9$  Hz, 1H, Ar-H), 6.96 (m, 2H, Ar-H), 3.86 (s, 3H, CH<sub>3</sub>), 1.43 (s, 9H, 3xCH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO-*d*<sub>6</sub>) 162.5 (CO), 161.1 (CO), 154.6 (C), 153.3 (C), 124.7 (CH), 112.1 (CH), 109.1 (C), 100.4 (CH), 78.7 (C), 55.9 (CH<sub>3</sub>), 28.2 (3xCH<sub>3</sub>). *Two C signals not detected.* **Mpt.:** 149-150 °C.  **$\nu_{\max}$  (neat):** 3299, 1675, 1610, 1372, 1161, 688 cm<sup>-1</sup>. **HRMS:** C<sub>15</sub>H<sub>18</sub>NO<sub>6</sub> [M+H<sup>+</sup>] requires  $m/z$  312.0867, found 312.0883. **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.48.

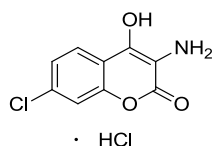
**General procedure for synthesis of aminocoumarins 99 and 147-149**

To an ice cooled solution of the *N*-Boc protected aminocoumarin (1 eq.) in *tert*-butylmethyl ether (TBME) (20 mL/g) and methanol (5 mL/g) was added 3M HCl in cyclopentylmethyl ether (15 eq.) and the mixture stirred at 22 °C for 24 hours. The solvent was removed under reduced pressure and the residue triturated with TBME (10 mL). The suspension was filtered to provide the title compounds.

**3-Amino-4,7-dihydroxy-2H-chromen-2-one hydrochloride (99)**<sup>[107]</sup>

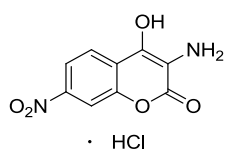
Isolated as an amorphous off-white solid (1.94 g, 99%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO- $d_6$ ) 10.52 (br. s, 1H, OH), 8.99 (br. s, 4H, NH<sub>3</sub> and OH), 7.90 (d,  $J=8.6$  Hz, 1H, Ar-H), 6.83 (dd,  $J_1=8.6$ ,  $J_2=2.3$  Hz, 1H, Ar-H), 6.72 (d,  $J=2.3$  Hz, 1H, Ar-H). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO- $d_6$ ) 162.0 (CO), 160.6 (C), 159.7 (C), 153.5 (C), 125.8 (CH), 113.1 (CH), 108.2 (C), 102.2 (CH), 94.9 (C). **Mpt.:** 240-241 °C.  **$\nu_{\max}$  (neat):** 3336, 1710  $\text{cm}^{-1}$ .  **$m/z$ :** (ES<sup>+</sup>) 194 (M<sup>+</sup>, 100%). **R<sub>f</sub>:** (EtOAc) 0.03.

**3-Amino-7-chloro-4-hydroxy-2H-chromen-2-one hydrochloride (147)**

Isolated as an off-white amorphous solid (136 mg, 95%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO- $d_6$ ) 7.88 (d,  $J=8.3$  Hz, 1H, Ar-H), 7.64 (br. s, 2H, NH<sub>2</sub>), 7.43 (s, 1H, Ar-H), 7.32 (d,  $J=8.3$ , 1H, Ar-H). *OH signal not detected.* **<sup>13</sup>C NMR:**  $\delta$  (176 MHz, DMSO- $d_6$ ) 159.8 (C), 152.4 (C), 134.5 (C), 125.6 (CH), 123.2 (CH), 120.5 (CH), 116.0 (C). *One C and CO and signal not detected.* **Mpt.:** 241 °C.  **$\nu_{\max}$  (neat):** 2900, 1733  $\text{cm}^{-1}$ . **HRMS:** C<sub>9</sub>H<sub>7</sub>ClNO<sub>3</sub> [M<sup>+</sup>] requires  $m/z$  212.0109, found 212.0115. **R<sub>f</sub>:** (EtOAc) 0.08.

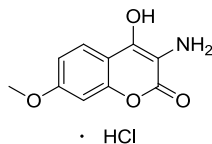
**3-Amino-7-nitro-4-hydroxy-2H-chromen-2-one hydrochloride (148)**

Isolated as a pale yellow solid (78 mg, 97%).

**<sup>1</sup>H NMR**  $\delta$  (400 MHz, DMSO- $d_6$ ) 8.35 (br. s, 3H, NH<sub>3</sub>), 8.10 (m, 2H, 2xAr-H), 8.02 (m, 1H, Ar-H). *OH signal not detected.* **<sup>13</sup>C NMR**  $\delta$  (176 MHz, DMSO- $d_6$ ) 159.6 (C), 147.2 (C), 125.9 (CH), 124.3 (CH), 118.1 (C), 111.5 (CH). *Two C signals and CO signal not detected.*

**Mpt.:** Decomp. >215 °C.  $\nu_{\text{max}}$  (neat): 2968, 1668, 1539, 1356  $\text{cm}^{-1}$ . **HRMS:**  $\text{C}_9\text{H}_7\text{N}_2\text{O}_5$  [ $\text{M}^+$ ] requires  $m/z$  223.0350, found 223.0354. **R<sub>f</sub>:** (EtOAc) 0.06.

### 3-Amino-7-methoxy-4-hydroxy-2H-chromen-2-one hydrochloride (149)



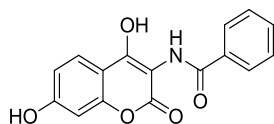
Isolated as a pale yellow solid (0.35 g, 89%).

**<sup>1</sup>H NMR**  $\delta$  (400 MHz, DMSO- $d_6$ ) 9.20 (br. s, 4H, OH and NH<sub>3</sub>), 7.93 (d,  $J=8.2$  Hz, 1H, Ar-H), 6.94 (m, 2H, Ar-H), 3.85 (s, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO- $d_6$ ) 162.4 (CO), 159.9 (C), 153.5 (C), 125.4 (CH), 124.9 (C), 111.7 (CH), 100.6 (CH), 96.1 (C), 94.8 (C), 55.8 (CH<sub>3</sub>). **Mpt.:** decomp. >250 °C.  $\nu_{\text{max}}$  (neat): 2914, 1706, 1243, 1194  $\text{cm}^{-1}$ . **HRMS:**  $\text{C}_{10}\text{H}_{10}\text{NO}_4$  [ $\text{M}^+$ ] requires  $m/z$  208.0604, found 208.0605. **R<sub>f</sub>:** (EtOAc) 0.04.

### General procedure for the synthesis of aminocoumarin substrates

To a solution of **99** or **146-149** (1 eq.) in EtOAc (12 mL/mmol) was added benzoyl chloride (5 eq.) and triethylamine (10 eq.). The suspension was stirred at 22 °C for 3 hours before being filtered. The filtrate was concentrated under reduced pressure and the residue dissolved in methanol (25 mL). Aq. 1 M NaOH (25 mL, 25.0 mmol) was added and the reaction was stirred at 22 °C for 12 hours before acidifying to pH 3 with aq. 2 M HCl. The reaction mixture was extracted with EtOAc (3x50 mL) and the combined organic fractions were washed with brine (50 mL) and dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was removed under reduced pressure and the crude solid washed with EtOAc (10 mL) and filtered to provide the title compounds.

### *N*-(4,7-Dihydroxy-2-oxo-2H-chromen-3-yl) benzamide (**67**)<sup>[107]</sup>

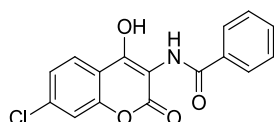


Isolated as a pale yellow amorphous solid (0.62 g, 68%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO- $d_6$ ) 11.81 (br. s, 1H, OH), 10.54 (s, 1H, OH), 9.41 (s, 1H, NH), 8.01 (d,  $J=7.3$  Hz, 2H, 2xAr-H), 7.73 (d,  $J=8.7$  Hz, 1H, Ar-H), 7.59 (app. t,  $J=5.9$  Hz,

1H, Ar-H), 7.52 (t,  $J=8.2$  Hz, 2H, 2xAr-H), 6.83 (dd,  $J=8.7$ ,  $J_2=2.2$  Hz, 1H, Ar-H), 6.73 (d,  $J=2.2$  Hz, 1H, Ar-H).  $^{13}\text{C NMR}$ :  $\delta$  (101 MHz, DMSO- $d_6$ ) 166.5 (CO), 161.4 (CO), 160.8 (C), 160.3 (C), 153.5 (C), 133.9 (C), 131.5 (CH), 128.1 (2xCH), 128.0 (2xCH), 125.0 (CH), 112.9 (CH), 108.0 (C), 101.9 (CH), 100.1 (C). **Mpt.**: 283-285 °C (lit.: 283-285 °C).<sup>[107]</sup>  $\nu_{\text{max}}$  (neat): 3176, 1665, 1625  $\text{cm}^{-1}$ .  $m/z$ : (ES<sup>+</sup>) 298 (M+H<sup>+</sup>, 100%). **R<sub>f</sub>**: (1:1 EtOAc/heptane) 0.31.

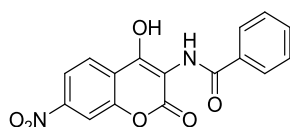
***N*-(4-Chloro, 7-hydroxy-2-oxo-2*H*-chromen-3-yl) benzamide (150)**



Isolated as a pale yellow amorphous solid (74 mg, 58%).

$^1\text{H NMR}$ :  $\delta$  (400 MHz, DMSO- $d_6$ ) 12.34 (br. s, 1H, OH), 9.55 (s, 1H, NH), 8.01 (d,  $J=7.3$  Hz, 2H, 2xAr-H), 7.91 (d,  $J=8.5$  Hz, 1H, Ar-H), 7.64 (d,  $J=2.0$  Hz, 1H, Ar-H), 7.59 (d,  $J=7.3$  Hz, 1H, Ar-H), 7.56-7.50 (m, 2H, 2xAr-H), 7.47 (dd,  $J_1=2.0$ ,  $J_2=8.5$  Hz, 1H, Ar-H).  $^{13}\text{C NMR}$ :  $\delta$  (101 MHz, DMSO- $d_6$ ) 166.4 (CO), 159.9 (CO), 159.2 (C), 152.0 (C), 136.5 (C), 133.7 (C), 131.7 (CH), 128.2 (2xCH), 128.0 (2xCH), 125.3 (CH), 124.6 (CH), 116.4 (CH), 115.4 (C), 103.0 (C). **Mpt.**: 233 °C.  $\nu_{\text{max}}$  (neat): 3352, 1695, 1623, 701  $\text{cm}^{-1}$ . **HRMS**:  $\text{C}_{16}\text{H}_{11}\text{ClNO}_4$  [M+H<sup>+</sup>] requires  $m/z$  316.0371, found 316.0377. **R<sub>f</sub>**: (1:1 EtOAc/heptane) 0.25.

***N*-(4-Nitro, 7-hydroxy-2-oxo-2*H*-chromen-3-yl) benzamide (151)**



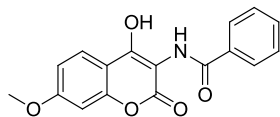
Isolated as a pale yellow amorphous solid (40 mg, 63%).

$^1\text{H NMR}$ :  $\delta$  (400 MHz, DMSO- $d_6$ ) 9.71 (s, 1H, NH), 8.27 (d,  $J=2.1$  Hz, 1H, Ar-H), 8.21 (dd,  $J_1=2.1$ ,  $J_2=8.7$  Hz, 1H, Ar-H), 8.15 (d,  $J=8.7$ , 1H, Ar-H), 8.02 (d,  $J=7.2$  Hz, 2H, Ar-H), 7.61 (t,  $J=7.3$  Hz, 1H, Ar-H), 7.53 (t,  $J=7.6$  Hz, 2H, Ar-H). *OH signal not detected.*  $^{13}\text{C NMR}$ :  $\delta$  (101 MHz, DMSO- $d_6$ ) 166.3 (CO), 159.7 (C), 158.2 (CO), 151.1 (C), 149.0 (C), 133.6 (C), 131.8 (CH), 128.2 (2xCH), 128.0 (2xCH), 125.4 (CH), 121.9 (C), 118.8 (CH), 111.8 (CH),



105.3 (C). **Mpt.:** 296 °C.  $\nu_{\text{max}}$  (**neat**): 1698, 1630, 1524, 1341  $\text{cm}^{-1}$ . **HRMS:**  $\text{C}_{16}\text{H}_{11}\text{N}_2\text{O}_6$   $[\text{M}+\text{H}^+]$  requires  $m/z$  327.0612, found 327.0622. **R<sub>f</sub>:** (1:1 EtOAc/heptane): 0.28.

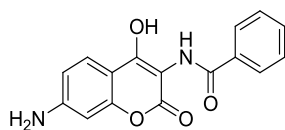
***N*-(4-Methoxy, 7-hydroxy-2-oxo-2*H*-chromen-3-yl) benzamide (153)**



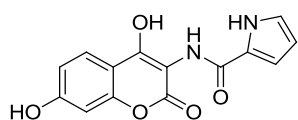
Isolated as a pale yellow solid (50 mg, 39%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO- $d_6$ ) 11.95 (s, 1H, OH), 9.45 (s, 1H, NH), 8.01 (d,  $J=7.2$  Hz, 2H, 2xAr-H), 7.81 (d,  $J=8.7$  Hz, 1H, Ar-H), 7.58 (t,  $J=7.2$  Hz, 1H, Ar-H), 7.51 (d,  $J=7.6$  Hz, 2H, 2xAr-H), 7.02 (m, 2H, Ar-H), 3.88 (s, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO- $d_6$ ) 166.5 (CO), 162.6 (CO), 153.5 (C), 133.9 (CH), 131.6 (CH), 128.2 (2xCH), 128.0 (2xCH), 124.8 (C), 112.2 (CH), 109.4 (C), 100.5 (CH), 55.9 (CH<sub>3</sub>). *Three C signals not detected.* **Mpt.:** decomp. >240 °C.  $\nu_{\text{max}}$  (**neat**): 3299, 1674, 1609, 1372, 1294, 688  $\text{cm}^{-1}$ . **HRMS:**  $\text{C}_{17}\text{H}_{14}\text{NO}_5$   $[\text{M}+\text{H}^+]$  requires  $m/z$  312.0867, found 312.0883. **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.40.

***N*-(4-Amino, 7-hydroxy-2-oxo-2*H*-chromen-3-yl) benzamide (152)**

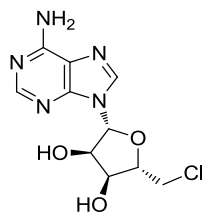


To a suspension of **151** (100 mg, 0.31 mmol) in acetic acid (5 mL) was added iron powder (171 mg, 3.07 mmol). The reaction mixture was heated to 50 °C and stirred for 4 hours before concentrating under reduced pressure. The mixture was acidified to pH 1 with aq. 2 M HCl and EtOAc (5 mL) was added. The resulting suspension was filtered to provide the title compound as a yellow solid (63 mg, 69%). **<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO- $d_6$ ) 11.48 (s, 1H, OH), 9.32 (s, 1H, NH), 8.00 (d,  $J=5.7$  Hz, 2H, 2xAr-H), 7.53 (m, 4H, 4xAr-H), 6.58 (d,  $J=7.0$  Hz, 1H, Ar-H), 6.43 (s, 1H, Ar-H). *NH<sub>2</sub> signal not detected.* **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO- $d_6$ ) 166.6 (C), 161.0 (CO), 160.7 (CO), 154.0 (C), 153.2 (C), 134.0 (C), 131.5 (CH), 128.1 (2xCH), 128.0 (2xCH), 124.6 (CH), 111.3 (CH), 104.2 (C), 98.4 (C), 98.2 (CH). **Mpt.:** decomp. >260 °C.  $\nu_{\text{max}}$  (**neat**): 3349, 1664, 1608, 1539, 1381  $\text{cm}^{-1}$ . **HRMS:**  $\text{C}_{16}\text{H}_{13}\text{N}_2\text{O}_4$   $[\text{M}+\text{H}^+]$  requires  $m/z$  297.0870, found 297.0879. **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.20.

***N*-(4,7-dihydroxy-2-oxo-2H-chromen-3-yl)-1H-pyrrole-2-carboxamide (68)**<sup>[107]</sup>

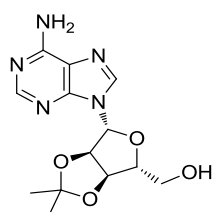
To a suspension of 1H-pyrrole-2-carboxylic acid (968 mg, 8.71 mmol) in anhydrous toluene (50.0 mL) was added oxalyl chloride (1.53 mL, 17.4 mmol) and *N,N*-dimethylformamide (0.02 mL, 0.22 mmol). The reaction was heated to 65 °C and stirred for 1 hour before cooling to 22 °C and concentrating under reduced pressure. The residue was redissolved in anhydrous dichloromethane (10.0 mL) and added to a solution of **99** (500 mg, 2.18 mmol) and triethylamine (2.43 mL, 17.4 mmol) in dichloromethane (30.0 mL). The reaction was stirred at 22 °C for 1 hour before acidifying to pH 2 with aq. 2 M HCl and extracting with EtOAc (3x30 mL). The combined organic fractions were dried over a hydrophobic frit and concentrated under reduced pressure. The residue was redissolved in methanol (30 mL) and sodium hydroxide (10 mL, 10.0 mmol) was added. The reaction was stirred at 22 °C for 1 hour before concentrating under reduced pressure and acidifying to pH 5 using aq. conc. HCl. The mixture was extracted with EtOAc (3x30 mL), the combined organic fractions washed with brine (30 mL) and dried over a hydrophobic frit. The filtrate was concentrated under reduced pressure and the product triturated from MeCN to provide the title compound as an amorphous brown solid (294 mg, 47%). **<sup>1</sup>H NMR:** δ (400 MHz, DMSO-*d*<sub>6</sub>) 12.00 (br. s, 1H, OH), 11.66 (br. s, 1H, NH), 10.55 (s, 1H, OH), 9.04 (s, 1H, NH), 7.73 (d, *J*=8.7 Hz, 1H, Ar-H), 6.95-7.08 (m, 2H, 2xAr-H), 6.85 (dd, *J*<sub>1</sub>=8.7, *J*<sub>2</sub>=2.3 Hz, 1H, Ar-H), 6.75 (d, *J*=2.3 Hz, 1H, Ar-H), 6.18 (m, 1H, Ar-H). **<sup>13</sup>C NMR:** δ (101 MHz, DMSO-*d*<sub>6</sub>) 161.3 (C), 160.8 (CO), 159.4 (C), 153.2 (C), 125.4 (CH), 125.0 (C), 122.4 (CH), 113.0 (CH), 112.3 (CH), 109.0 (CH), 108.1 (C), 101.9 (CH), 100.4 (C). *One CO signal not detected.* **Mpt.:** decomp. > 290 °C. ***v*<sub>max</sub> (neat):** 3355, 1679, 1642, 1530, 1134, 739 cm<sup>-1</sup>. ***m/z:*** (ES<sup>+</sup>) 287 (M+H<sup>+</sup>, 37%), 145 (100%). ***R<sub>f</sub>:*** (1:1 EtOAc/heptane) 0.31.

**(2R, 3R, 4S, 5S)-2-(6-amino-9H-purin-9-yl)-5-(chloromethyl)tetrahydrofuran-3,4-diol (85)**<sup>[223]</sup>



To an ice cooled suspension of adenosine (1.00 g, 3.74 mmol) in acetonitrile (10 mL) was added pyridine (0.61 mL, 7.48 mmol) and thionyl chloride (1.40 mL, 18.7 mmol) dropwise over 5 minutes. The reaction mixture was stirred at 0 °C for 3 hours before being warmed to 22 °C and stirred for 16 hours. The resulting precipitate was filtered and dissolved in water/methanol (5:1). 25% aq. ammonia (2.00 mL) was added and the reaction mixture was stirred at 22 °C for 30 minutes. The solvent was removed under reduced pressure to provide the title compound as a colourless amorphous solid (1.05 g, 98%). <sup>1</sup>H NMR: δ (400 MHz, DMSO-d<sub>6</sub>) 8.35 (s, 1H, Ar-H), 8.17 (s, 1H, Ar-H), 7.31 (s, 2H, NH<sub>2</sub>), 5.95 (d, *J*=5.7 Hz, 1H, CH), 5.59 (d, *J*=6.0 Hz, 1H, OH), 5.45 (d, *J*=5.2 Hz, 1H, OH), 4.77 (q, *J*=5.6 Hz, 1H, CH), 4.18-4.32 (m, 1H, CH), 4.03-4.16 (m, 1H, CH), 3.78-4.02 (m, 2H, CH<sub>2</sub>). <sup>13</sup>C NMR: δ (101 MHz, DMSO-d<sub>6</sub>) 156.1 (C), 152.7 (CH), 149.4 (C), 139.7 (CH), 119.1 (C), 87.4 (CH), 83.6 (CH), 72.6 (CH), 71.3 (CH), 44.8 (CH<sub>2</sub>). **Mpt.:** 72 °C (lit. 75-80 °C).<sup>[358]</sup> **v**<sub>max</sub> (neat): 3148, 1648, 1601 cm<sup>-1</sup>. **m/z:** (ES<sup>+</sup>) 286 (M<sup>35</sup>Cl+H<sup>+</sup>, 100%), 288 (M<sup>37</sup>Cl+H<sup>+</sup>, 35%). **R<sub>f</sub>:** (EtOAc) 0.20.

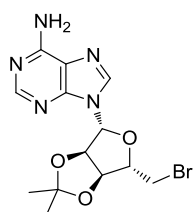
**((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)methanol (107)**<sup>[359]</sup>



To an ice cooled suspension of adenosine (10.0 g, 37.4 mmol) in acetone (200 mL) was added perchloric acid (3.22 mL, 37.4 mmol) dropwise. The reaction was stirred at 0 °C for 1 hour before being warmed to 22 °C and stirred for 2 hours. The reaction mixture was neutralised to pH 7 with solid NaHCO<sub>3</sub> before filtering. The filtrate was concentrated under reduced pressure and the crude residue was purified by column chromatography (0-10% MeOH in dichloromethane) to provide the title compound as a colourless amorphous solid

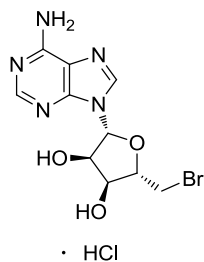
(10.0 g, 87%). **<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO-*d*<sub>6</sub>) 8.34 (s, 1H, Ar-H), 8.15 (s, 1H, Ar-H), 7.33 (s, 2H, NH<sub>2</sub>), 6.12 (d,  $J$ =3.1 Hz, 1H, CH), 5.34 (dd,  $J_1$ =3.1,  $J_2$ =6.2 Hz, 1H, OH), 5.22 (t,  $J$ =5.5 Hz, 1H, CH), 4.97 (dd,  $J_1$ =2.6,  $J_2$ =6.2 Hz, 1H, CH), 4.21 (dt,  $J_1$ =2.6,  $J_2$ =4.8 Hz, 1H, CH), 3.58 (dd,  $J_1$ =5.1,  $J_2$ =11.7 Hz, 1H, CH<sub>2</sub>), 3.52 (dd,  $J_1$ =5.6,  $J_2$ =11.7 Hz, 1H, CH<sub>2</sub>), 1.55 (s, 3H, CH<sub>3</sub>), 1.33 (s, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO-*d*<sub>6</sub>) 156.1 (C), 152.6 (CH), 148.8 (C), 139.7 (CH), 119.1 (C), 113.0 (C), 89.6 (CH), 86.3 (CH), 83.2 (CH), 81.3 (CH), 61.6 (CH<sub>2</sub>), 27.1 (CH<sub>3</sub>), 25.2 (CH<sub>3</sub>). **Mpt.:** 221 °C (lit. 221-222 °C). **<sup>360</sup>v<sub>max</sub>** (neat): 3209, 1683, 1605, 1116, 1084, 710 cm<sup>-1</sup>. **m/z:** (ES<sup>+</sup>) 308 (M+H<sup>+</sup>, 100%). **R<sub>f</sub>:** (EtOAc) 0.14.

**9-((3*aR*,4*R*,6*S*,6*aS*)-6-(bromomethyl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9H-purin-6-amine (217)**<sup>[361]</sup>



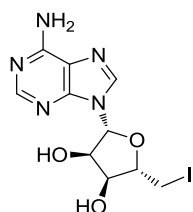
To a suspension of **107** (0.50 g, 1.63 mmol) and *N*-bromosuccinimide (0.58 g, 3.25 mmol) in dichloromethane (10 mL) at 22 °C was added triphenylphosphine (0.85 g, 3.25 mmol) portionwise. BaCO<sub>3</sub> (0.50 g, 2.53 mmol) was then added and the reaction mixture was heated to reflux for 1 hour before being cooled and filtered. The filtrate was washed with water (2x10 mL) and the combined aqueous fractions were extracted with dichloromethane (2x10 mL). The combined organic layers were washed with brine (10 mL) and dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was removed under reduced pressure to provide a brown oil which was purified by column chromatography (100% EtOAc) to provide the title compound as a pale brown oil (0.20 g, 33%). **<sup>1</sup>H NMR:**  $\delta$  (400 MHz, CDCl<sub>3</sub>) 8.38 (s, 1H, Ar-H), 7.93 (s, 1H, Ar-H), 6.13 (d,  $J$ =2 Hz, 1H, CH), 5.56 (br. s, 2H, NH<sub>2</sub>), 5.51 (dd,  $J_1$ =2,  $J_2$ =6 Hz, 1H, CH), 5.19 (dd,  $J_1$ =3 Hz,  $J_2$ =6 Hz, 1H, CH), 4.49-4.56 (m, 1H, CH), 3.68 (dd,  $J_1$ =8,  $J_2$ =10 Hz, 1H, CH<sub>2</sub>), 3.47 (dd,  $J_1$ =5,  $J_2$ =10 Hz, 1H, CH<sub>2</sub>), 1.64 (s, 3H, CH<sub>3</sub>), 1.42 (s, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, CDCl<sub>3</sub>) 155.6 (C), 153.2 (CH), 149.2 (C), 140.1 (CH), 114.6 (C), 91.2 (CH), 86.8 (CH), 84.2 (CH), 83.6 (CH), 75.3 (C), 31.8 (CH<sub>2</sub>), 27.1 (CH<sub>3</sub>), 25.3 (CH<sub>3</sub>). **v<sub>max</sub>** (neat): 3059, 1659, 1208, 1090, 831 cm<sup>-1</sup>. **m/z:** (ES<sup>+</sup>) 370 (M<sup>79</sup>Br+H<sup>+</sup>, 98%), 372 (M<sup>81</sup>Br+H<sup>+</sup>, 100%). **R<sub>f</sub>:** (EtOAc) 0.09.

**(2R, 3R, 4S, 5S)-2-(6-amino-9H-purin-9-yl)-5-(bromomethyl)tetrahydrofuran-3,4-diol hydrochloride (212)**<sup>[362]</sup>



To a solution of **217** (300 mg, 0.81 mmol) in water/tetrahydrofuran (1:1, 4 mL) was added 25% aq. HCl (0.49 mL, 4.05 mmol) and the reaction was stirred at 22 °C for 1 hour. The solvent was removed under reduced pressure to provide the title compound as a pale brown gum (0.29 g, 96%). **<sup>1</sup>H NMR:**  $\delta$  (400 MHz, D<sub>2</sub>O) 8.55 (s, 1H, Ar-H), 8.44 (s, 1H, Ar-H), 6.18 (d,  $J=4.8$  Hz, 1H, CH), 4.89 (t,  $J=5.2$ , 1H, CH), 4.48 (t,  $J=5.1$  Hz, CH), 4.42 (q,  $J=4.8$  Hz, 1H, CH), 3.82 (dd,  $J_1=4.3$ ,  $J_2=11.6$  Hz, 1H, CH<sub>2</sub>), 3.78 (dd,  $J_1=5.2$ ,  $J_2=11.6$  Hz, 1H, CH<sub>2</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, D<sub>2</sub>O) 150.0 (C), 148.4 (CH), 144.6 (C), 142.9 (CH), 118.9 (C), 88.4 (CH), 83.5 (CH), 73.7 (CH), 71.7 (CH), 32.5 (CH<sub>2</sub>).  **$\nu_{\max}$  (neat):** 3254, 1700, 692 cm<sup>-1</sup>.  **$m/z$ :** (ES<sup>+</sup>) 330 (M<sup>79</sup>Br+H<sup>+</sup>, 100%), 332 (M<sup>81</sup>Br+H<sup>+</sup>, 98%). **R<sub>f</sub>:** (3:1:1 water/*n*-BuOH/AcOH) 0.65.

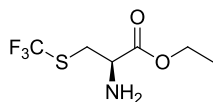
**(2R, 3R, 4S, 5S)-2-(6-amino-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol (213)**<sup>[312]</sup>



To a solution of adenosine (1.00 g, 3.74 mmol) in pyridine (10 mL) was added iodine (1.43 g, 5.61 mmol) and triphenylphosphine (1.47 g, 5.61 mmol). The reaction was stirred at 22 °C for 2 hours, before quenching with saturated Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>. The layers were separated and the organic layer concentrated under reduced pressure. The crude solid was washed with acetone (10 mL) and purified by column chromatography (10% MeOH in dichloromethane) to provide the title compound as a yellow solid (0.17 g, 12%). **<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO-*d*<sub>6</sub>) 8.43 (s, 1H, Ar-H), 8.21 (s, 1H, Ar-H), 7.60 (br. s, 2H, NH<sub>2</sub>), 5.93 (d,  $J=6$  Hz, 1H, CH), 4.80 (app. t,  $J=5$  Hz, 1H, CH), 4.17 (dd,  $J=4$ ,  $J_2=5$  Hz, 1H, CH), 4.04-3.99 (m, 1H,

CH), 3.61 (dd,  $J_1=6$ ,  $J_2=11$  Hz, 1H, CH<sub>2</sub>), 3.47 (dd,  $J_1=7.0$ ,  $J_2=11$  Hz, 1H, CH<sub>2</sub>). *OH signals not detected.* <sup>13</sup>C NMR:  $\delta$  (101 MHz, D<sub>2</sub>O) 154.6 (CH), 150.9 (C), 149.2 (CH), 140.5 (C), 119.0 (C), 87.5 (CH), 84.0 (CH), 73.1 (CH), 72.9 (CH), 7.7 (CH<sub>2</sub>). **Mpt.:** 80 °C.  $\nu_{\text{max}}$  (neat): 3050, 1652, 667 cm<sup>-1</sup>. *m/z:* (ES<sup>+</sup>) 378 (M+H<sup>+</sup>, 100%). **R<sub>f</sub>:** (3:1:1 water/*n*-BuOH/AcOH) 0.74.

**(*R*)-ethyl 2-amino-3-((trifluoromethyl)thio)propanoate (225)**<sup>[315]</sup>



**Synthesis A (Togni reagent, 221):**

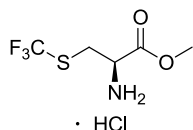
To a solution of 3,3-dimethyl-1-(trifluoromethyl)-1,2-benziodoxole (**221**) (98.0 mg, 0.30 mmol) in MeOH (1 mL) at -78 °C was added a solution of (*R*)-ethyl 2-amino-3-mercaptopropanoate hydrochloride (50.0 mg, 0.27 mmol) in MeOH (1 mL) in a dropwise manner. The reaction was stirred at -78 °C and left warm up slowly to 22 °C over 16 hours. The solvent was removed under reduced pressure and the product washed with heptane/EtOAc (20:1, 5 mL). The resulting suspension was filtered and washed with a further portion of heptane/EtOAc (20:1, 10 mL). The resulting white solid was redissolved in MeOH (10 mL) and concentrated under reduced pressure to provide the title compound as an amorphous colourless solid (63.4 mg, 93%). <sup>1</sup>H NMR:  $\delta$  (400 MHz, D<sub>2</sub>O) 4.51 (dd,  $J_1=4.8$ ,  $J_2=7.1$  Hz, 1H, CH), 4.33 (q,  $J=7.1$  Hz, 2H, CH<sub>2</sub>), 3.66 (dd,  $J_1=4.8$ ,  $J_2=15.8$  Hz, 1H, CH<sub>2</sub>), 3.52 (dd,  $J_1=7.1$ ,  $J_2=15.8$  Hz, 1H, CH<sub>2</sub>), 1.31 (t,  $J=7.1$  Hz, 3H, CH<sub>3</sub>). <sup>13</sup>C NMR:  $\delta$  (101 MHz, D<sub>2</sub>O) 167.7 (CO), 64.3 (CH), 52.5 (CH<sub>2</sub>), 28.7 (CH<sub>2</sub>), 13.1 (CH<sub>3</sub>). <sup>19</sup>F NMR:  $\delta$  (377 MHz, D<sub>2</sub>O) -41.26 (CF<sub>3</sub>). **Mpt.:** 152 °C.  $\nu_{\text{max}}$  (neat): 3049, 1741, 1651, 1116 cm<sup>-1</sup>. *m/z:* (EI) 144 (M-CO<sub>2</sub>Et, 100%). **R<sub>f</sub>:** (3:1:1 water/*n*-BuOH/AcOH) 0.79.

**Synthesis B (Umemoto reagent, 222):**

To an ice cooled solution of (*R*)-ethyl 2-amino-3-mercaptopropanoate hydrochloride (20.0 mg, 0.11 mmol) in anhydrous DMF (0.5 mL) was added a suspension of NaH (8.62 mg, 0.22 mmol) in DMF (0.5 mL) dropwise. The reaction mixture was stirred at 0 °C for 20 minutes before 5-(trifluoromethyl)-5H-dibenzo[b,d]thiophen-5-ium triflate (**222**) (43.3 mg, 0.11 mmol) was added. The reaction was stirred at 0 °C for 10 minutes before being warmed to 22 °C using a water bath and stirred for 16 hours. The solvent was removed under reduced pressure and the product washed with heptane/EtOAc (20:1, 5 mL). The suspension was filtered and the product redissolved in MeOH (2 mL). The solvent was

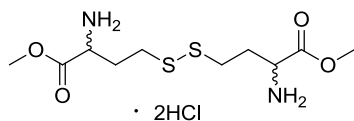
removed under reduced pressure to provide the title compound as an amorphous colourless solid (16 mg, 68 %). Data match to above.

**(R)-Methyl 2-amino-3-((trifluoromethyl)thio)propanoate hydrochloride (236)**<sup>[130]</sup>



To a solution of cystine dimethyl ester dihydrochloride (0.50 g, 1.47 mmol) and sodium trifluoromethanesulfinate (0.69 g, 4.40 mmol) in water (10 mL) was added *tert*-butylhydroperoxide (TBHP) (0.20 mL, 1.47 mmol) dropwise at a rate of 1.70 mL/hour via a syringe pump. The reaction temperature was maintained at 22 °C and the reaction was stirred for 16 hours at 22 °C before acidifying with sat. aq. Na<sub>2</sub>S<sub>2</sub>O<sub>5</sub> to pH 1. The reaction mixture was washed with *tert*-butylmethyl ether TBME (10 mL) before the aqueous fraction was basified to pH 9 with aq. 2 M NaOH. The reaction mixture was extracted with TBME (2x10 mL) and 3 M HCl in cyclopentylmethyl ether (5 mL) was added. The aqueous fraction was collected and evaporated to dryness to provide the title compound as a colourless amorphous solid (0.15 g, 21%). <sup>1</sup>H NMR: δ (400 MHz, D<sub>2</sub>O) 4.54 (dd, *J*=7, 5 Hz, 1H, CH), 3.88 (s, 3H, CH<sub>3</sub>), 3.70 (dd, *J*=16, 5 Hz, 1H, CH<sub>2</sub>), 3.57 (dd, *J*=16, 7 Hz, 1H, CH<sub>2</sub>). <sup>13</sup>C NMR: δ (101 MHz, D<sub>2</sub>O) 168.2 (CO), 54.0 (CH<sub>2</sub>), 52.4 (CH), 28.8 (CH<sub>3</sub>). CF<sub>3</sub> signal not detected. <sup>19</sup>F NMR: δ (377 MHz, D<sub>2</sub>O) -41.3. *m/z*: (EI) 144 (M-CO<sub>2</sub>Me, 100%). **Mpt.:** Decomp. >120 °C. *v*<sub>max</sub> (neat): 2859, 1747, 1490, 1247, 1069 cm<sup>-1</sup>. **R<sub>f</sub>**: (3:1:1 water/*n*-BuOH/AcOH) 0.82.

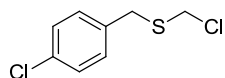
**DL-Dimethyl (2*S*,2'*S*)-4,4'-disulfanediylbis(2-aminobutanoate) (234)**<sup>[130]</sup>



To a dry flask under a N<sub>2</sub> atmosphere was added racemic homocystine (1.00 g, 3.73 mmol) and anhydrous methanol (10 mL). To the vigorously stirred suspension was added thionyl chloride (1.36 mL, 18.6 mmol) dropwise. The reaction was stirred for 24 hours at 22 °C before being concentrated under reduced pressure to give a yellow residue. Toluene (5 mL) was added and the solution concentrated under reduced pressure to provide the title compound as a pale yellow solid (1.37 g, 100%). <sup>1</sup>H NMR: δ (400 MHz, D<sub>2</sub>O) 4.27 (app. t,

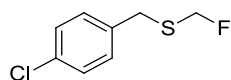
$J=6.2$  Hz, 2H, 2xCH), 3.82 (s, 6H, 2xCH<sub>3</sub>), 2.84 (m, 4H, 2xCH<sub>2</sub>), 2.36 (m, 4H, 2xCH<sub>2</sub>). <sup>13</sup>C NMR:  $\delta$  (101 MHz, D<sub>2</sub>O,) 170.3 (2xCO), 53.7 (2xCH), 51.4 (2xCH<sub>2</sub>), 32.0 (2xCH<sub>2</sub>), 28.9 (2xCH<sub>3</sub>). **Mpt.:** 165 °C.  $\nu_{\text{max}}$  (neat): 2863, 1744 cm<sup>-1</sup>. **m/z:** (ES<sup>+</sup>) 297 (M+H<sup>+</sup>, 100%). **R<sub>f</sub>:** (3:1:1 water/*n*-BuOH/AcOH) 0.44.

**(4-Chlorobenzyl)(chloromethyl)sulfane (243)**<sup>[320]</sup>



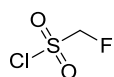
To an oven-dried three-necked round bottomed flask was added paraformaldehyde (1.02 g, 34.1 mmol). One neck was connected to a Schott bottle filled with sat. aq. NaHCO<sub>3</sub>, to the middle neck was placed a suba seal and to the third neck a tap joint. Anhydrous dichloromethane (20 mL) was added, followed by (4-chlorophenyl)methanethiol (**242**) (4.50 mL, 34.1 mmol) and the suspension was cooled to -15 °C. In a separate flask, dry HCl gas was generated from the dropwise addition of sulfuric acid (25 mL, 235 mmol) to sodium chloride (60 g, 1.03 mol) and was bubbled through the suspension *via* a needle connected to the tap joint. This was continued for 2 hours, until the solution became clear. Next, N<sub>2</sub> gas was bubbled through the mixture for 30 minutes before transferring the reaction mixture into a separating funnel, washing with water (20 mL), brine (20 mL) and drying over Na<sub>2</sub>SO<sub>2</sub>. The organic fraction was concentrated under reduced pressure to provide the crude product as a colourless oil, which was taken through into the next stage of the synthesis without further purification.

**(4-Chlorobenzyl)(fluoromethyl)sulfane (244)**<sup>[320]</sup>

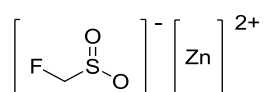


To an oven dried flask under an N<sub>2</sub> atmosphere was added potassium fluoride (3.96 g, 68.2 mmol) and dicyclohexano-18-crown-6 (0.92 mL, 2.73 mmol) was added followed by a solution of (4-chlorobenzyl)(chloromethyl)sulfane (**243**) (theoretical amount 7.06 g, 34.1 mmol) in acetonitrile (40 mL) *via* a cannula. The reaction mixture was stirred vigorously and heated at reflux for 72 hours before cooling to 22 °C, diluting with water (50 mL) and extracting with dichloromethane (2x50 mL). The combined organic fractions were washed with brine (50 mL), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure to provide the title compound as a yellow oil which was taken through into the next stage of the synthesis without purification.



**Fluoromethanesulfonyl chloride (245)**<sup>[320]</sup>

To a solution of (4-chlorobenzyl)(fluoromethyl)sulfane (**244**) (5.27 g, 27.6 mmol) in dichloromethane (40 mL) was added aq. 1 M hydrochloric acid (221 mL, 221 mmol). The solution was cooled to -10 °C and a solution of 14% aq. sodium hypochlorite (82 mL, 193 mmol) was added dropwise over 15 minutes. The reaction mixture was stirred for 6 hours at -10 °C before extracting three times with dichloromethane (3x50 mL). The combined organic fractions were washed with water (50 mL), brine (50 mL), dried over a hydrophobic frit and concentrated under reduced pressure at 10 °C. The crude product was isolated by Kugelrohr distillation (50 °C, 8 mbar) to provide the title compound as a colourless oil which was taken through to the next stage of the synthesis without further purification.

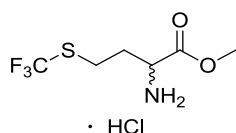
**Zinc fluoromethanesulfinate (240)**<sup>[320]</sup>

To an ice cooled suspension of zinc dust (2.76 g, 42.3 mmol) and water (10 mL) was added fluoromethanesulfonyl chloride (**245**) (theoretical amount 0.80 g, 6.04 mmol) dropwise. The reaction mixture was stirred for 3 hours with gradual warming to 22 °C. The reaction mixture was filtered through sand and the filtrate concentrated under reduced pressure. The resulting solid was dried under vacuum for 16 hours to provide the title compound as a colourless solid (400 mg, 26% from **242**). **<sup>1</sup>H NMR:** δ (400 MHz, D<sub>2</sub>O) 4.46 (d, *J*=48.0 Hz, 2H, CH<sub>2</sub>). **<sup>13</sup>C NMR:** δ (101 MHz, D<sub>2</sub>O) 97.8 (d, *J*=206.4 Hz, CH<sub>2</sub>). **<sup>19</sup>F NMR:** δ (377 MHz, D<sub>2</sub>O) -216.97.

**General procedure for the synthesis of *O*-Me protected fluorinated methionine analogues using zinc sulfinate salts**

To a solution of homocystine dimethyl ester dihydrochloride (**234**) (200 mg, 0.54 mmol) in water (12 mL) was added the appropriate zinc sulfinate salt (2 eq.) and TBMP (0.34 mL, 2.44 mmol) at a rate of 1.70 mL/hour *via* a syringe pump. The reaction was maintained at 20 °C for the duration of the addition and reaction. The reaction was stirred vigorously for 2 hours before a further portion of zinc sulfinate salt (1 eq.) and TBHP (0.34 mL, 2.44 mmol) were added and the reaction mixture was stirred for a further 2 hours. The reaction mixture was acidified with sat. aq. Na<sub>2</sub>S<sub>2</sub>O<sub>5</sub> to pH 1. The reaction mixture was washed with 2,2,3-trimethylpentane (10 mL) before being basified to pH 9 with solid K<sub>2</sub>CO<sub>3</sub> and extracted with diethyl ether (2x10 mL). The organic fractions were combined and extracted with aq. 2 M HCl (2x5 mL). The aqueous fractions were combined and the solvent removed under reduced pressure to provide the title compounds.

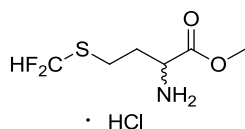
**DL-Methyl 2-amino-4-((trifluoromethyl)thio)butanoate hydrochloride (**237**)**<sup>[130]</sup>



Compound isolated as colourless amorphous solid (135 mg, 49%).

**<sup>1</sup>H NMR:** δ (400 MHz, D<sub>2</sub>O) 4.27 (app. t, *J*=6.6 Hz, 1H, CH), 3.83 (s, 3H, CH<sub>3</sub>), 3.20-3.03 (m, 2H, CH<sub>2</sub>), 2.47-2.22 (m, 2H, CH<sub>2</sub>). **<sup>13</sup>C NMR:** δ (101 MHz, D<sub>2</sub>O) 169.9 (CO), 53.8 (CH), 51.4 (CH<sub>3</sub>), 30.3 (CH<sub>2</sub>), 25.2 (CH<sub>2</sub>). **<sup>19</sup>F NMR:** δ (377 MHz, D<sub>2</sub>O) -41.1 (CF<sub>3</sub>). **Mpt.:** 162 °C.  **$\nu_{\max}$  (neat):** 2842, 1747, 1226, 1094 cm<sup>-1</sup>. ***m/z*:** (EI) 157 ([M-CO<sub>2</sub>Me]<sup>+</sup>, 100%). ***R<sub>f</sub>*:** (3:1:1 water/*n*-BuOH/AcOH) 0.54.

**DL-Methyl 2-amino-4-((difluoromethyl)thio)butanoate hydrochloride (**241**)**

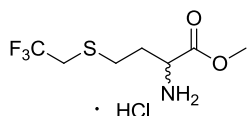


Compound isolated as colourless amorphous solid (173 mg, 56%).

**<sup>1</sup>H NMR:** δ (400 MHz, D<sub>2</sub>O) 7.04 (app. t, *J*=55.7 Hz, 1H, CF<sub>2</sub>H), 4.27 (app. t, *J*=6.6 Hz, 1H, CH), 3.83 (s, 3H, CH<sub>3</sub>), 2.99 (m, 2H, CH<sub>2</sub>), 2.29 (m, 2H, CH<sub>2</sub>). **<sup>13</sup>C NMR:** δ (101 MHz, D<sub>2</sub>O) 170.1 (CO), 131.1 (t, *J*=271.0 Hz, CF<sub>2</sub>H), 53.7 (CH), 51.5 (CH<sub>3</sub>), 30.9 (CH<sub>2</sub>), 22.5

(CH<sub>2</sub>). <sup>19</sup>F NMR: δ (377 MHz, D<sub>2</sub>O) -93.2 (CF<sub>2</sub>H). **Mpt.:** 162 °C. **v<sub>max</sub>** (neat): 2878, 1747, 1227, 1056 cm<sup>-1</sup>. **HRMS:** C<sub>6</sub>H<sub>12</sub>O<sub>2</sub>NF<sub>2</sub>S [M+H]<sup>+</sup> requires *m/z* 200.0551, found 200.0546. **R<sub>f</sub>**: (3:1:1 water/*n*-BuOH/AcOH) 0.56.

#### DL-Methyl 2-amino-4-((2,2,2-trifluoroethyl)thio)butanoate hydrochloride (242)



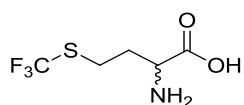
Compound isolated as a colourless amorphous solid (173 mg, 68%).

<sup>1</sup>H NMR: δ (400 MHz, D<sub>2</sub>O) 4.27 (t, *J*=6.5 Hz, 1H, CH), 3.82 (s, 3H, CH<sub>3</sub>), 3.31 (q, *J*=10.2 Hz, 2H, CH<sub>2</sub>CF<sub>3</sub>), 2.83 (m, 2H, CH<sub>2</sub>), 2.24 (m, 2H, CH<sub>2</sub>). <sup>13</sup>C NMR: δ (101 MHz, D<sub>2</sub>O) 170.3 (CO), 131.1 (t, *J*=271.0 Hz, CF<sub>2</sub>H), 53.7 (CH), 51.5 (CH<sub>3</sub>), 32.8 (CH<sub>2</sub>), 29.3 (CH<sub>2</sub>), 27.6 (CH<sub>2</sub>). <sup>19</sup>F NMR: δ (377 MHz, D<sub>2</sub>O) -66.4 (CF<sub>3</sub>). **Mpt.:** 102 °C. **v<sub>max</sub>** (neat): 2861, 1748, 1235, 1093 cm<sup>-1</sup>. **HRMS:** C<sub>7</sub>H<sub>13</sub>O<sub>2</sub>NF<sub>3</sub>S [M+H]<sup>+</sup> requires *m/z* 232.0614, found 232.0604. **R<sub>f</sub>**: (3:1:1 water/*n*-BuOH/AcOH) 0.58.

#### General procedure for the synthesis of fluorinated methionine analogues using zinc sulfinate salts

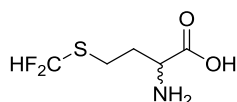
To a solution of racemic homocysteine (1 eq.) and zinc sulfinate salt (2 eq.) in water (3 mL) was added *tert*-butyl hydroperoxide (4 eq.) dropwise over 5 minutes and the reaction stirred at 20 °C for 4 hours before washing the reaction mixture with trimethylpentane (5 mL) and basifying with solid K<sub>2</sub>CO<sub>3</sub> to pH 7. The resulting solid was collected by filtration and washed with ice-cold water to provide the title compounds.

#### DL-2-Amino-4-((trifluoromethyl)thio)butanoic acid (205)<sup>[314]</sup>



Isolated as a colourless amorphous solid (65.4 mg, 86 % yield).

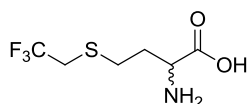
<sup>1</sup>H NMR: δ (400 MHz, D<sub>2</sub>O) 4.16 (app. t, *J*=6.6 Hz, 1H, CH), 2.97-3.17 (m, 2H, CH<sub>2</sub>), 2.16-2.46 (m, 2H, CH<sub>2</sub>). <sup>13</sup>C NMR: δ (101 MHz, D<sub>2</sub>O) 168.4 (CO), 129.5 (CF<sub>3</sub>), 48.7 (CH), 27.6 (CH<sub>2</sub>), 22.5 (CH<sub>2</sub>). <sup>19</sup>F NMR: δ (377 MHz, D<sub>2</sub>O) -41.2 (CF<sub>3</sub>). **v<sub>max</sub>** (neat): 3272, 1601, 1090, 742 cm<sup>-1</sup>. **Mpt.:** Decomp. >200 °C. **HRMS:** C<sub>5</sub>H<sub>9</sub>O<sub>2</sub>NF<sub>3</sub>S [M+H]<sup>+</sup> requires *m/z* 204.0301, found 204.0298. **R<sub>f</sub>**: (3:1:1 water/*n*-BuOH/AcOH) 0.45.

**DL-2-Amino-4-((difluoromethyl)thio)butanoic acid (244)**

Isolated as a colourless amorphous solid (70.3 mg, 87%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, D<sub>2</sub>O) 7.07 (t,  $J=55.9$  Hz, 1H, CF<sub>2</sub>H), 3.82 (app. t,  $J=6.3$  Hz, 1H, CH), 2.97 (t,  $J=7.6$  Hz, 2H, CH<sub>2</sub>), 2.07-2.36 (m, 2H, CH<sub>2</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, D<sub>2</sub>O) 174.3 (CO), 121.3 (t,  $J=270.7$  Hz, CF<sub>2</sub>H), 53.7 (CH), 31.9 (CH<sub>2</sub>), 22.9 (CH<sub>2</sub>). **<sup>19</sup>F NMR:**  $\delta$  (377 MHz, D<sub>2</sub>O) -93.04.  **$\nu_{\max}$  (neat):** 3325, 1599, 1120, 739 cm<sup>-1</sup>. **Mpt.:** Decomp. >190 °C.

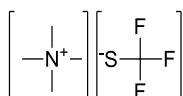
**HRMS:** C<sub>5</sub>H<sub>10</sub>O<sub>2</sub>NF<sub>2</sub>S [M+H]<sup>+</sup> requires  $m/z$  186.0395, found 186.0390. **R<sub>f</sub>:** (3:1:1 water/*n*-BuOH/AcOH) 0.48.

**DL-2-Amino-4-((2,2,2-trifluoroethyl)thio)butanoic acid (245)**

Isolated as a colourless amorphous solid (75 mg, 93 % yield).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, D<sub>2</sub>O) 4.13 (t,  $J=6.5$  Hz, 1H, CH), 3.24 (q,  $J=10.3$  Hz, 2H, CH<sub>2</sub>), 2.77 (t,  $J=7.3$  Hz, 2H, CH<sub>2</sub>), 2.05-2.29 (m, 2H, CH<sub>2</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, D<sub>2</sub>O) 168.6 (CO), 123.3 (q,  $J=275.4$  Hz, CF<sub>3</sub>), 48.73 (CH), 30.32 (CH<sub>2</sub>), 26.59 (CH<sub>2</sub>), 24.88 (CH<sub>2</sub>). **<sup>19</sup>F NMR:**  $\delta$  (377 MHz, D<sub>2</sub>O) -64.3.  **$\nu_{\max}$  (neat):** 3278, 1599, 1135, 857 cm<sup>-1</sup>. **Mpt.:** Decomp. >200 °C.

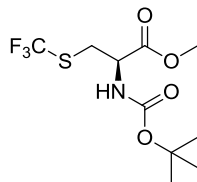
**HRMS:** C<sub>6</sub>H<sub>11</sub>O<sub>2</sub>NF<sub>3</sub>S [M+H]<sup>+</sup> requires  $m/z$  218.0457, found 218.0455. **R<sub>f</sub>:** (3:1:1 water/*n*-BuOH/AcOH) 0.51.

**Tetramethylammonium trifluoromethanethiolate (229)<sup>[318]</sup>**

To a solution of sulfur (0.20 g, 6.24 mmol) in anhydrous tetrahydrofuran (50 mL) at -78 °C under an N<sub>2</sub> atmosphere was added trimethyl(trifluoromethyl)silane (1.11 mL, 7.48 mmol) and tetramethylammonium fluoride (0.58 g, 6.24 mmol). The reaction was stirred at -78 °C for 1 hour before being warmed to 22 °C and stirred for 16 hours. The resulting suspension was filtered and dried under a N<sub>2</sub> atmosphere to provide the title compound as a beige amorphous solid (0.88 g, 80 %). **<sup>1</sup>H NMR:**  $\delta$  (400 MHz, Acetonitrile-d<sub>3</sub>) 3.12 (s, 12H,

4xCH<sub>3</sub>). <sup>13</sup>C NMR: δ (101 MHz, Acetonitrile-d<sub>3</sub>) 54.8 (CH<sub>3</sub>). <sup>19</sup>F NMR: δ (377 MHz, Acetonitrile-d<sub>3</sub>) -7.78 (SCF<sub>3</sub>). **Mpt.:** 101 °C. **v**<sub>max</sub> (**neat**): 3408, 1490, 1220, 949 cm<sup>-1</sup>.

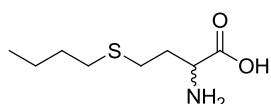
**(R)-Methyl 2-((tert-butoxycarbonyl)amino)-3-((trifluoromethyl)thio)propanoate (232)**<sup>[315]</sup>



To an ice cooled suspension of (*R*)-methyl 3-bromo-2-((*tert*-butoxycarbonyl)amino)propanoate (0.10 g, 0.35 mmol) in anhydrous acetonitrile (4 mL) under a N<sub>2</sub> atmosphere was added tetramethylammonium trifluoromethanethiolate (68.4 mg, 0.39 mmol). The reaction was stirred at 22 °C for 2 hours before being filtered and concentrated under reduced pressure. The residue was purified by column chromatography (0-20% EtOAc in heptane) to provide the title compound as a brown oil (43.3 mg, 81%). <sup>1</sup>H NMR: δ (400 MHz, CDCl<sub>3</sub>) 5.42 (br. s, 1H, NH), 4.77 (app. t, *J*<sub>1</sub>=3.9 Hz, 1H, CH), 3.83 (dd, *J*<sub>1</sub>=3.4, *J*<sub>2</sub>=10.6 Hz, 1H, CH<sub>2</sub>), 3.81 (s, 3H, CH<sub>3</sub>), 3.79 (dd, *J*<sub>1</sub>=3.9, *J*<sub>2</sub>=10.6 Hz, 1H, CH<sub>2</sub>), 1.46 (s, 9H, 3xCH<sub>3</sub>). <sup>13</sup>C NMR: δ (101 MHz, D<sub>2</sub>O) 169.7 (CO), 155.0 (CO), 80.5 (C), 54.0 (CH), 53.0 (CH<sub>2</sub>), 34.1 (CH<sub>3</sub>), 28.3 (3xCH<sub>3</sub>). *CF*<sub>3</sub> signal not observed. <sup>19</sup>F NMR: δ (337 MHz, CDCl<sub>3</sub>) -40.9 (CF<sub>3</sub>). **v**<sub>max</sub> (**neat**): 2980, 1749, 1710, 1114 cm<sup>-1</sup>. **m/z:** (ES<sup>+</sup>) 304 (M+H<sup>+</sup>, 4%), 193 (100%). **R<sub>f</sub>:** (EtOAc) 0.79.

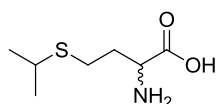
**General procedure for the synthesis of *S*-alkylated Met analogues**

To a solution of racemic homocysteine (200 mg, 1.48 mmol) in ethanol (2 mL) and aq. 2 M sodium hydroxide (3.33 mL, 6.66 mmol) was added the corresponding alkyl bromide (1.48 mmol). The reaction mixture was stirred for 24 hours before acidifying with conc. aq. HCl to pH 2 and concentrating under reduced pressure. The residue was resuspended in warm ethanol and the precipitate removed by filtration. The filtrate was concentrated under reduced pressure and the residue redissolved in water (2 mL). The mixture was adjusted to ~pH 4 with aq. 2 M NaOH and the resulting suspension filtered to afford the title compounds.

**DL-2-amino-4-(butylthio)butanoic acid (251)**<sup>[311]</sup>

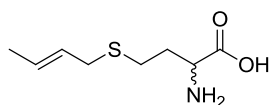
Compound isolated as a colourless amorphous solid (169 mg, 60%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, D<sub>2</sub>O) 4.14 (app. t,  $J=6.4$  Hz, 1H, CH), 2.49-2.64 (dt,  $J_1=6.4$ ,  $J_2=48.4$  Hz, 4H, 2xCH<sub>2</sub>), 2.19-2.08 (m, 2H, CH<sub>2</sub>), 1.48-1.45 (m, 2H, CH<sub>2</sub>), 1.30-1.25 (m, 2H, CH<sub>2</sub>), 0.78 (t,  $J=7.4$ , 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, D<sub>2</sub>O) 171.6 (COOH), 51.7 (CH), 30.7 (CH<sub>2</sub>), 30.5 (CH<sub>2</sub>), 29.5 (CH<sub>2</sub>), 26.3 (CH<sub>2</sub>), 21.2 (CH<sub>2</sub>), 12.8 (CH<sub>3</sub>). **Mpt.:** 251.4-252.8 °C.  **$\nu_{\max}$  (neat):** 2954, 1579, 1413, 1341 cm<sup>-1</sup>.  **$m/z$ :** (ES<sup>+</sup>) 192 (M+H<sup>+</sup>, 100%). **R<sub>f</sub>:** (3:1:1 water/*n*-BuOH/AcOH) 0.63.

**DL-2-amino-4-(isopropylthio)butanoic acid (252)**<sup>[311]</sup>

Compound isolated as a colourless amorphous solid (179 mg, 68%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, D<sub>2</sub>O) 3.78 (dd,  $J_1=6.7$ ,  $J_2=5.7$  Hz, 1H, CH), 3.01 (dt,  $J_1=13.4$ ,  $J_2=6.7$  Hz, 1H, CH), 2.64 (d,  $J=15.6$  Hz, 2H, CH<sub>2</sub>), 1.95-2.19 (m, 2H, CH<sub>2</sub>), 1.21 (d,  $J=6.6$  Hz, 6H, 2xCH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, D<sub>2</sub>O) 174.3 (COOH), 54.1 (CH), 34.4 (CH), 30.7 (CH<sub>2</sub>), 25.2 (CH<sub>2</sub>), 22.4 (CH<sub>3</sub>). **Mpt.:** Decomp. >215 °C.  **$\nu_{\max}$  (neat):** 2957, 1578, 1410, 1339 cm<sup>-1</sup>.  **$m/z$ :** (ES<sup>+</sup>) 178 (M+H<sup>+</sup>, 100%). **R<sub>f</sub>:** (3:1:1 water/*n*-BuOH/AcOH) 0.34.

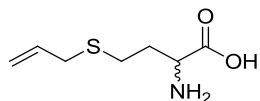
**DL-(E)-2-amino-4-(but-2-en-1-ylthio)butanoic acid (253)**<sup>[311]</sup>

Compound isolated as a colourless amorphous solid (186 mg, 66%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, D<sub>2</sub>O) 5.34-5.74 (m, 2H, 2xCH), 3.77 (dd,  $J_1=5.5$ ,  $J_2=6.9$ , 1H, CH), 3.13 (d,  $J=7.4$  Hz, 2H, CH<sub>2</sub>), 2.54-2.60 (m, 2H, CH<sub>2</sub>), 1.97-2.17 (m, 2H, CH<sub>2</sub>), 1.65 (dd,  $J_1=6.4$ ,  $J_2=1.4$  Hz, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, D<sub>2</sub>O) 174.2 (COOH), 129.9 (CH), 126.1 (CH), 54.0 (CH<sub>2</sub>), 32.6 (CH<sub>2</sub>), 30.2 (CH<sub>2</sub>), 25.3 (CH<sub>2</sub>), 16.9 (CH<sub>3</sub>). **Mpt.:** Decomp.

>200 °C.  $\nu_{\max}$  (neat): 2917, 1576, 1411, 1339  $\text{cm}^{-1}$ .  $m/z$ : ( $\text{ES}^+$ ) 190 ( $\text{M}+\text{H}^+$ , 100%).  $R_f$ : (3:1:1 water/*n*-BuOH/AcOH) 0.39.

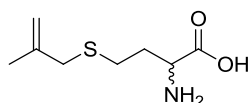
**DL-4-(allylthio)-2-aminobutanoic acid (254)**<sup>[311]</sup>



Compound isolated as a colourless amorphous solid (146 mg, 56%).

$^1\text{H NMR}$ :  $\delta$  (400 MHz,  $\text{D}_2\text{O}$ ) 5.80 (ddt,  $J_1=17.1$ ,  $J_2=9.9$ ,  $J_3=7.2$  Hz, 1H, CH), 5.05-5.22 (m, 2H, 2x $\text{CH}_2$ ), 3.78 (dd,  $J_1=6.8$ ,  $J_2=5.6$  Hz, 1H, CH), 3.18 (d,  $J=7.3$  Hz, 2H,  $\text{CH}_2$ ), 2.54-2.60 (m, 2H,  $\text{CH}_2$ ), 1.99-2.18 (m, 2H,  $\text{CH}_2$ ).  $^{13}\text{C NMR}$ :  $\delta$  (101 MHz,  $\text{D}_2\text{O}$ ) 174.2 (COOH), 133.8 ( $\text{CH}_2$ ), 117.8 (CH), 54.0 (CH), 33.4 ( $\text{CH}_2$ ), 30.1 ( $\text{CH}_2$ ), 25.4 ( $\text{CH}_2$ ). **Mpt.:** Decomp. >230 °C.  $\nu_{\max}$  (neat): 2923, 1575, 1410, 1339  $\text{cm}^{-1}$ .  $m/z$ : ( $\text{ES}^+$ ) 167 ( $\text{M}+\text{H}^+$ , 100%).  $R_f$ : (3:1:1 water/*n*-BuOH/AcOH) 0.37.

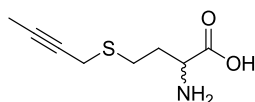
**DL-2-amino-4-((2-methylallyl)thio)butanoic acid (255)**<sup>[311]</sup>



Compound isolated as a colourless amorphous solid (196 mg, 70%).

$^1\text{H NMR}$ :  $\delta$  (400 MHz,  $\text{D}_2\text{O}$ ) 4.88 (d,  $J=4.6$  Hz, 2H,  $\text{CH}_2$ ), 3.89 (app. t,  $J=6.3$  Hz, 1H, CH), 3.19 (s, 2H,  $\text{CH}_2$ ), 2.55 (t,  $J=7.6$  Hz, 2H,  $\text{CH}_2$ ), 2.01-2.21 (m, 2H,  $\text{CH}_2$ ), 1.77 (s, 3H,  $\text{CH}_3$ ).  $^{13}\text{C NMR}$ :  $\delta$  (101 MHz,  $\text{D}_2\text{O}$ ) 173.4 (COOH), 141.5 (C), 113.9 ( $\text{CH}_2$ ), 53.5 (CH), 38.0 ( $\text{CH}_2$ ), 29.8 ( $\text{CH}_2$ ), 25.7 ( $\text{CH}_2$ ), 19.8 ( $\text{CH}_3$ ). **Mpt.:** Decomp. >210 °C.  $\nu_{\max}$  (neat): 2917, 1576, 1411, 1339  $\text{cm}^{-1}$ .  $m/z$ : ( $\text{ES}^+$ ) 190 ( $\text{M}+\text{H}^+$ , 100%).  $R_f$ : (3:1:1 water/ *n*-BuOH/ AcOH) 0.42.

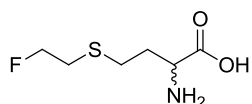
**DL-2-amino-4-(prop-1-yn-1-ylthio)butanoic acid (256)**<sup>[363]</sup>



Compound isolated as an orange amorphous solid (162 mg, 59%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, D<sub>2</sub>O) 3.85 (app. t,  $J=6.3$  Hz, 1H, CH), 3.29 (s, 2H, CH<sub>2</sub>), 2.75-2.80 (m, 2H, CH<sub>2</sub>), 2.04-2.31 (m, 2H, CH<sub>2</sub>), 1.78 (s, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, D<sub>2</sub>O) 173.8 (COOH), 80.6 (C), 75.0 (C), 53.7 (CH), 29.8 (CH<sub>2</sub>), 26.5 (CH<sub>2</sub>), 18.6 (CH<sub>2</sub>), 2.5 (CH<sub>3</sub>). **Mpt.:** Decomp. >200 °C.  $\nu_{\max}$  (neat): 2916, 1578, 1411, 1339 cm<sup>-1</sup>. **m/z:** (ES<sup>+</sup>) 188 (M+H<sup>+</sup>, 100%). **R<sub>f</sub>:** (3:1:1 water/*n*-BuOH/AcOH) 0.37.

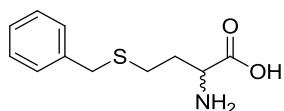
**DL-2-amino-4-((2-fluoroethyl)thio)butanoic acid (257)**<sup>[364]</sup>



Compound isolated as a colourless amorphous solid (195 mg, 73%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, D<sub>2</sub>O) 4.68-4.54 (dt,  $J_1=44.0$ ,  $J_2=5.8$  Hz, 2H, CFH<sub>2</sub>), 3.78 (dd,  $J=7.0$ , 5.7 Hz, 1H, CH), 2.95-2.89 (dt,  $J_1=20.0$ ,  $J_2=5.8$  Hz, 2H, CH<sub>2</sub>), 2.72 (t,  $J=7.8$  Hz, 2H, CH<sub>2</sub>), 2.05-2.23 (m, 2H, CH<sub>2</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, D<sub>2</sub>O) 174.5 (COOH), 84.2 (CH<sub>2</sub>), 82.6 (CH<sub>2</sub>), 53.9 (CH), 30.6 (CH<sub>2</sub>), 27.1 (CH<sub>2</sub>). **<sup>19</sup>F NMR:**  $\delta$  (377 MHz, D<sub>2</sub>O) -214.1 (CH<sub>2</sub>F). **Mpt.:** Decomp. >225 °C.  $\nu_{\max}$  (neat): 2923, 1574, 1409, 1338, 984 cm<sup>-1</sup>. *Compound unstable to ES<sup>+</sup> and EI.* **R<sub>f</sub>:** (3:1:1 water/*n*-BuOH/AcOH) 0.24.

**DL-2-amino-4-(benzylthio)butanoic acid (258)**<sup>[363]</sup>



Compound isolated as a colourless amorphous solid (256 mg, 77%).

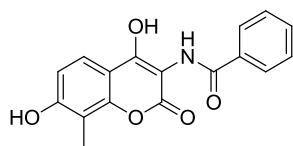
**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, D<sub>2</sub>O) 7.09-7.13 (m, 4H, Ar-H), 7.02-7.09 (m, 1H, Ar-H), 3.89 (app. t,  $J=6.3$  Hz, 1H, CH), 3.51-3.54 (s, 2H, CH<sub>2</sub>), 2.35-2.31 (m, 2H, CH<sub>2</sub>), 1.71-2.01 (m, 2H, CH<sub>2</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, D<sub>2</sub>O) 171.2 (COOH), 138.0 (C), 128.8 (2xCH), 128.7 (2xCH), 127.2 (CH), 51.5 (CH), 34.6 (CH<sub>2</sub>), 29.1 (CH<sub>2</sub>), 25.7 (CH<sub>2</sub>). **Mpt.:** Decomp. >240 °C.  $\nu_{\max}$  (neat): 2923, 1576, 1413, 1338, 697 cm<sup>-1</sup>. **m/z:** (ES<sup>+</sup>) 226 (M+H<sup>+</sup>, 15%), 188 (100%). **R<sub>f</sub>:** (3:1:1 water/*n*-BuOH/AcOH) 0.42.



### General procedure for the one-pot enzymatic cascade methylation using *E. coli* cell free extract and *E. coli* cell lysate

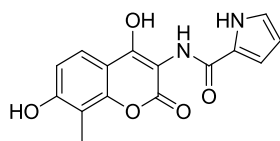
To a 50 mL Falcon tube was added substrate **67**, **68** or **121** (20.0 mg, 0.13 mmol, as a 100 mM solution in DMSO), **85** (53.6 mg, 0.19 mmol, as a 100 mM solution in DMSO), L-methionine (0.25 mmol, as a 500 mM solution in aq. 250 mM NaOH) and DTT (19.3 mg, 0.13 mmol, as a 100 mM solution in potassium phosphate buffer). BSA (to a final concentration of 1 mg/mL) was added before initiating SAM synthesis with SaL cell free extract (15 mL, cell pellet resuspended in 5 mL/g). The reaction was incubated at 35 °C, 750 rpm for 2 hours, before adding NovO cell lysate (35 mL, cell pellet resuspended in 10 mL/g) and MTAN (0.5 μM) and incubating the reaction at 35 °C, 750 rpm, for a further 16 hours. The reaction mixture was acidified to pH 5 with aq. conc. HCl and clarified by centrifugation (20 minutes, 4000 rpm, 4 °C). The supernatant was extracted 3 times with EtOAc (3x50 mL), each time clarifying the emulsion by centrifugation (10 minutes, 4000 rpm, 4 °C) and the combined organic fractions purified by mass-directed automated purification (MDAP) to provide the title compounds.

#### *N*-(4,7-dihydroxy-8-methyl-2-oxo-2H-chromen-3-yl)benzamide (**69**)<sup>[107]</sup>



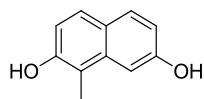
97% Conversion by HPLC. Isolated as a colourless solid (11 mg, 53%).

**<sup>1</sup>H NMR:** δ (400 MHz, DMSO-*d*<sub>6</sub>) 11.72 (s, 1H, OH), 10.44 (s, 1H, OH), 9.43 (s, 1H, NH), 8.03 (d, *J*=7.3 Hz, 2H, 2xAr-H), 7.57-7.64 (m, 2H, 2xAr-H), 7.50-7.57 (m, 2H, 2xAr-H), 6.91 (d, *J*=8.7 Hz, 1H, Ar-H), 2.20 (s, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:** δ (101 MHz, DMSO-*d*<sub>6</sub>) 166.5 (CO), 160.8 (CO), 160.4 (C), 159.1 (C), 151.5 (C), 133.9 (C), 131.6 (CH), 128.2 (CH), 128.0 (CH), 121.5 (CH), 111.8 (CH), 110.4 (C), 107.9 (C), 99.9 (C), 8.1 (CH<sub>3</sub>). **Mpt.:** Decomp. > 310-312 °C. **v<sub>max</sub> (neat):** 3215, 1567, 1534, 1097, 695 cm<sup>-1</sup>. **m/z:** (ES<sup>+</sup>) 312 (M+H<sup>+</sup>, 100%). **R<sub>f</sub>:** (EtOAc) 0.28.

***N*-(4,7-dihydroxy-8-methyl-2-oxo-2H-chromen-3-yl)-1H-pyrrole-2-carboxamide (70)**<sup>[107]</sup>

100% Conversion by HPLC. Isolated as an amorphous pale yellow solid (16 mg, 76%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO-*d*<sub>6</sub>) 11.89 (s, 1H, OH), 11.66 (br. s, 1H, NH), 10.41 (s, 1H, OH), 9.03 (s, 1H, NH), 7.58 (d, *J*=8.6 Hz, 1H, Ar-H), 7.04 (br. s, 1H, Ar-H), 6.96 (br. s, 1H, Ar-H), 6.89 (d, *J*=8.7 Hz, 1H, Ar-H), 6.17 (d, *J*=3.0 Hz, 1H, Ar-H), 2.17 (s, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO-*d*<sub>6</sub>) 160.8 (C), 160.8 (CO), 159.5 (C), 159.0 (C), 151.2 (C), 125.4 (C), 122.4 (CH), 121.5 (CH), 112.3 (CH), 111.9 (CH), 109.0 (CH), 108.0 (C), 100.1 (C), 8.1 (CH<sub>3</sub>). *One CO signal not detected.* **Mpt.:** Decomp. >300 °C.  **$\nu_{\max}$  (neat):** 3340, 1584, 1532, 1404, 731 cm<sup>-1</sup>. ***m/z:*** (ES<sup>+</sup>) 301 (M+H<sup>+</sup>, 100%). ***R<sub>f</sub>:*** (1:1 EtOAc/heptane) 0.33.

**1-Methylnaphthalene-2,7-diol (72)**<sup>[107]</sup>

65% Conversion by HPLC. Isolated as a pink solid (10 mg, 46%).

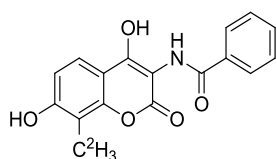
**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, MeOD-*d*<sub>4</sub>) 7.56 (d, *J*=8.9 Hz, 1H, Ar-H), 7.43 (d, *J*=8.9 Hz, 1H, Ar-H), 7.12 (d, *J*=2.2 Hz, 1H, Ar-H), 6.87 (d, *J*=8.8 Hz, 1H, Ar-H), 6.83 (dd, *J*<sub>1</sub>=8.8, *J*<sub>2</sub>=2.2 Hz, 1H, Ar-H), 2.38 (s, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, MeOD-*d*<sub>4</sub>) 156.8 (C), 153.7 (C), 137.3 (C), 131.0 (CH), 127.9 (CH), 125.3 (C), 115.9 (CH), 115.6 (CH), 106.0 (CH), 10.8 (CH<sub>3</sub>). *One C signal not detected.* **Mpt.:** 148.1 °C.  **$\nu_{\max}$  (neat):** 3221, 1626, 1185, 1098, 822 cm<sup>-1</sup>. ***m/z:*** (ES<sup>-</sup>) 173 (M-H<sup>+</sup>, 45%), 347 (100%). ***R<sub>f</sub>:*** (1:1 EtOAc/heptane) 0.28.

### General procedure for the one-pot enzymatic cascade methylation using purified enzymes

To a 50 mL Falcon tube was added substrate **67**, **68** or **121** (20.0 mg, 0.13 mmol, as a 100 mM solution in DMSO), **85** (53.6 mg, 0.19 mmol, as a 100 mM solution in DMSO), L-methionine (or isotopically labelled analogue, purchased from Sigma Aldrich) or L-ethionine (0.25 mmol, as a 500 mM solution in aq. 250 mM NaOH), potassium phosphate buffer (50 mM, pH 6.5, 40 mL) and DTT (19.3 mg, 0.13 mmol, as a 100 mM solution in potassium

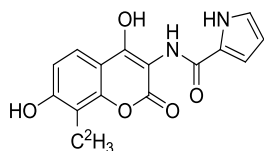
phosphate buffer). BSA (to a final concentration of 1 mg/mL) was added before initiating SAM or SAM-analogue synthesis with SalL (5  $\mu$ M). The reaction was incubated at 35  $^{\circ}$ C, 750 rpm for 7 hours ( $^{13}\text{CH}_3$ ,  $^{13}\text{CD}_3$ ,  $\text{CD}_3$  transfer) or 32 hours (Et transfer) before adding NovO (40  $\mu$ M) and MTAN (0.5  $\mu$ M) and incubating the reaction at 35  $^{\circ}$ C, 750 rpm, for a further 16 hours. The reaction mixture was acidified to pH 5 with aq. conc. HCl and the solids removed by centrifugation (20 min, 4000 rpm, 4  $^{\circ}$ C). The supernatant was extracted 3 times with EtOAc (3x50 mL), each time clarifying the emulsion by centrifugation (10 minutes, 4000 rpm, 4  $^{\circ}$ C) and the combined organic fractions purified by MDAP to provide the title compounds.

***N*-(4,7-dihydroxy-8- $^{2}\text{H}_3$ -methyl-2-oxo-2H-chromen-3-yl)benzamide (266)**



88% conversion by HPLC. Product isolated as an amorphous colourless solid (11 mg, 52%).  $^1\text{H NMR}$ :  $\delta$  (400 MHz, DMSO- $d_6$ ) 11.72 (s, 1H, OH), 10.43 (s, 1H, OH), 9.42 (s, 1H, NH), 7.98-8.05 (d,  $J=3$  Hz, 2H, 2xAr-H), 7.55-7.63 (m, 2H, 2xAr-H), 7.48-7.55 (m, 2H, 2xAr-H), 6.89 (d,  $J=8.7$  Hz, 1H, Ar-H).  $^{13}\text{C NMR}$ :  $\delta$  (101 MHz, DMSO- $d_6$ ) 166.5 (CO), 160.8 (CO), 160.4 (C), 159.1 (C), 151.5 (C), 133.9 (C), 131.5 (CH), 128.1 (CH), 128.0 (CH), 121.5 (CH), 111.8 (CH), 110.3 (C), 107.9 (C), 99.8 (C). **Mpt.:** Decomp. > 265  $^{\circ}$ C.  $\nu_{\text{max}}$  (neat): 3214, 1565, 1538  $\text{cm}^{-1}$ . **HRMS:**  $\text{C}_{17}\text{H}_{11}^2\text{H}_3\text{O}_5\text{N}$   $[\text{M}+\text{H}]^+$  requires  $m/z$  315.1055, found 315.1047. **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.39.

***N*-(4,7-dihydroxy-8- $^{2}\text{H}_3$ -methyl-2-oxo-2H-chromen-3-yl)-1H-pyrrole-2-carboxamide (269)**

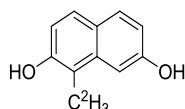


100% conversion by HPLC. Product isolated as an amorphous pale yellow solid (17 mg, 80%).

$^1\text{H NMR}$ :  $\delta$  (400 MHz, DMSO- $d_6$ ) 11.89 (s, 1H, OH), 11.65 (br. s, 1H, NH), 10.42 (s, 1H, OH), 9.03 (s, 1H, NH), 7.58 (d,  $J=8.6$  Hz, 1H, Ar-H), 7.04 (br. s, 1H, Ar-H), 6.96 (br. s, 1H,

Ar-H), 6.89 (d,  $J=8.7$  Hz, 1H, Ar-H), 6.17 (d,  $J=3.3$  Hz, 1H, Ar-H).  $^{13}\text{C}$  NMR:  $\delta$  (101 MHz, DMSO- $d_6$ ) 160.8 (C), 160.8 (CO), 159.5 (C), 159.0 (C), 151.2 (C), 125.4 (C), 122.4 (CH), 121.5 (CH), 112.3 (CH), 111.9 (CH), 109.0 (CH), 108.0 (C), 100.1 (C). One CO signal and  $\text{C}^2\text{H}_3$  signal not observed. **Mpt.:** Decomp.  $>300$  °C.  $\nu_{\text{max}}$  (neat): 3343, 1587, 1554, 1400, 728  $\text{cm}^{-1}$ . **HRMS:**  $\text{C}_{15}\text{H}_{10}^2\text{H}_3\text{O}_5\text{N}_2$   $[\text{M}+\text{H}]^+$  requires  $m/z$  304.1007, found 304.0996. **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.33.

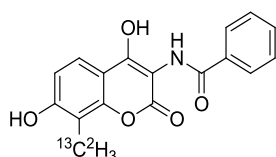
### 1-( $^2\text{H}_3$ )-methyl-naphthalene-2,7-diol (272)



91% conversion by HPLC. Product isolated as a dark pink solid (15 mg, 68%).

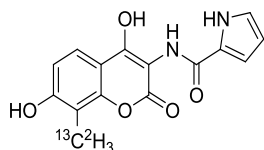
$^1\text{H}$  NMR:  $\delta$  (400 MHz, MeOD- $d_4$ ) 7.56 (d,  $J=8.7$  Hz, 1H, Ar-H), 7.43 (d,  $J=8.7$  Hz, 1H, Ar-H), 7.12 (d,  $J=2.3$  Hz, 1H, Ar-H), 6.86 (d,  $J=8.7$  Hz, 1H, Ar-H), 6.84 (dd,  $J=8.7, 2.3$  Hz, 1H, Ar-H). OH signals not detected.  $^{13}\text{C}$  NMR:  $\delta$  (101 MHz, MeOD- $d_4$ ) 156.8 (C), 153.7 (C), 137.4 (C), 131.0 (CH), 127.9 (CH), 125.3 (C), 115.9 (CH), 115.6 (CH), 106.0 (CH). One C and  $\text{C}^2\text{H}_3$  signals not detected. **Mpt.:** Decomp.  $>170$  °C.  $\nu_{\text{max}}$  (neat): 3284, 1634, 1188, 1027, 823  $\text{cm}^{-1}$ . **HRMS:**  $\text{C}_{11}\text{H}_8^2\text{H}_3\text{O}_2$   $[\text{M}+\text{H}]^+$  requires  $m/z$  178.0942, found 178.0935. **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.28.

### N-(4,7-dihydroxy-8- $^{13}\text{C}^2\text{H}_3$ )-methyl-2-oxo-2H-chromen-3-yl)benzamide (267)



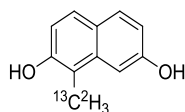
90% conversion by HPLC. Product isolated as an amorphous colourless solid (10 mg, 47%).  $^1\text{H}$  NMR:  $\delta$  (400 MHz, DMSO- $d_6$ ) 11.72 (br. s, 1H, OH), 10.43 (s, 1H, OH), 9.42 (s, 1H, NH), 8.01 (d,  $J=7.1$  Hz, 2H, 2xAr-H), 7.55-7.62 (m, 2H, 2xAr-H), 7.49-7.55 (m, 2H, 2xAr-H), 6.89 (d,  $J=8.7$  Hz, 1H, Ar-H).  $^{13}\text{C}$  NMR:  $\delta$  (101 MHz, DMSO- $d_6$ ) 166.5 (CO), 160.8 (CO), 160.4 (C), 159.1 (C), 151.5 (C), 133.9 (C), 131.5 (CH), 128.1 (CH), 128.0 (CH), 121.5 (CH), 111.8 (CH), 110.1 (C), 107.9 (C), 99.8 (C), 7.3 (spt,  $J=19.7$  Hz,  $^{13}\text{C}^2\text{H}_3$ ). **Mpt.:** Decomp.  $>300$  °C.  $\nu_{\text{max}}$  (neat): 3219, 1567, 1539, 1139, 695  $\text{cm}^{-1}$ . **HRMS:**  $\text{C}_{16}^{13}\text{CH}_{11}^2\text{H}_3\text{O}_5\text{N}$   $[\text{M}+\text{H}]^+$  requires  $m/z$  316.1088, found 316.1079. **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.39.

***N*-(4,7-dihydroxy-8-[ $^{13}\text{C}^2\text{H}_3$ ]-methyl-2-oxo-2H-chromen-3-yl)-1H-pyrrole-2-carboxamide (270)**

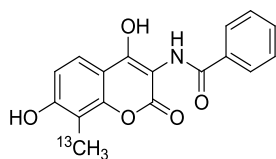


100% conversion by HPLC. Product isolated as a colourless amorphous solid (16 mg, 75%).  
 $^1\text{H NMR}$ :  $\delta$  (400 MHz, DMSO- $d_6$ ) 11.89 (s, 1H, OH), 11.66 (br. s, 1H, NH), 10.41 (s, 1H, OH), 9.03 (s, 1H, NH), 7.58 (d,  $J=8.6$  Hz, 1H, Ar-H), 7.04 (br. s, 1H, Ar-H), 6.96 (br. s, 1H, Ar-H), 6.89 (d,  $J=8.6$  Hz, 1H, Ar-H), 6.17 (d,  $J=3.2$  Hz, 1H, Ar-H).  $^{13}\text{C NMR}$ :  $\delta$  (101 MHz, DMSO- $d_6$ ) 160.8 (C), 160.8 (CO), 159.5 (C), 159.0 (C), 151.2 (C), 125.4 (C), 122.4 (CH), 121.5 (CH), 112.3 (CH), 111.9 (CH), 109.0 (CH), 108.0 (C), 100.1 (C), 7.3 (spt,  $J=19.7$  Hz,  $^{13}\text{C}^2\text{H}_3$ ). *One CO signal not observed.* **Mpt.**: Decomp.  $>300$  °C.  $\nu_{\text{max}}$  (**neat**): 3182, 1574, 1551, 1403, 735  $\text{cm}^{-1}$ . **HRMS**:  $\text{C}_{14}^{13}\text{CH}_{10}^2\text{H}_3\text{O}_5\text{N}_2$   $[\text{M}+\text{H}]^+$  requires  $m/z$  305.1041, found 305.1030. **R<sub>f</sub>**: (1:1 EtOAc/heptane) 0.33.

**1-( $^{13}\text{C}^2\text{H}_3$ )-methylnaphthalene-2,7-diol (273)**

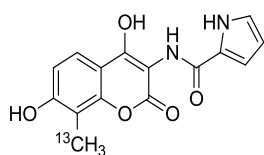


92% conversion by HPLC. Product isolated as a dark pink solid (19 mg, 85%).  
 $^1\text{H NMR}$ :  $\delta$  (400 MHz, MeOD- $d_4$ ) 7.56 (d,  $J=8.7$  Hz, 1H, Ar-H), 7.43 (d,  $J=8.7$  Hz, 1H, Ar-H), 7.12 (d,  $J=2.2$  Hz, 1H, Ar-H), 6.87 (dd,  $J=8.7, 1.0$  Hz, 1H, Ar-H), 6.83 (dd,  $J=8.7, 2.3$  Hz, 1H, Ar-H). *OH signals not detected.*  $^{13}\text{C NMR}$ :  $\delta$  (101 MHz, MeOD- $d_4$ ) 156.8 (C), 153.7 (C), 137.3 (C), 131.0 (CH), 127.9 (CH), 125.3 (d,  $J=2.5$  Hz, C), 115.9 (d,  $J=1.9$  Hz, CH), 115.6 (CH), 106.0 (d,  $J=4.0$  Hz, CH), 10.1 (spt,  $J=19.3$  Hz,  $^{13}\text{C}^2\text{H}_3$ ). *One C signal not detected.* **Mpt.**: Decomp.  $>170$  °C.  $\nu_{\text{max}}$  (**neat**): 3228, 1627  $\text{cm}^{-1}$ . **HRMS**:  $\text{C}_{10}^{13}\text{CH}_8^2\text{H}_3\text{O}_2$   $[\text{M}+\text{H}]^+$  requires  $m/z$  179.0975, found 179.0970. **R<sub>f</sub>**: (1:1 EtOAc/heptane) 0.28.

***N*-(4,7-dihydroxy-8-<sup>13</sup>C]-methyl-2-oxo-2H-chromen-3-yl)benzamide (256)**

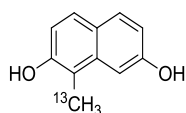
95% conversion by HPLC. Product isolated as a colourless amorphous solid (16 mg, 76%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO-*d*<sub>6</sub>) 11.73 (br. s, 1H, OH), 10.44 (s, 1H, OH), 9.43 (s, 1H, NH), 8.03 (d, *J*=7.3 Hz, 2H, 2xAr-H), 7.57-7.65 (m, 2H, 2xAr-H), 7.50-7.57 (m, 2H, 2xAr-H), 6.91 (d, *J*=8.7 Hz, 1H, Ar-H), 2.19 (d, *J*=128.7 Hz, 3H, <sup>13</sup>CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO-*d*<sub>6</sub>) 166.5 (CO), 160.8 (CO), 160.4 (C), 159.1 (C), 151.4 (C), 133.9 (C), 131.5 (CH), 128.1 (CH), 128.0 (CH), 121.5 (CH), 111.8 (CH), 110.6 (C), 107.9 (C), 99.8 (C), 8.1 (<sup>13</sup>CH<sub>3</sub>). **Mpt.:** Decomp. >300 °C.  **$\nu_{\max}$  (neat):** 3203, 1565, 1538, 1096, 695 cm<sup>-1</sup>. **HRMS:** C<sub>16</sub><sup>13</sup>CH<sub>14</sub>NO<sub>5</sub> [M+H]<sup>+</sup> requires *m/z* 313.0867, found 313.0892. **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.39.

***N*-(4,7-dihydroxy-8-<sup>13</sup>C]-methyl-2-oxo-2H-chromen-3-yl)-1H-pyrrole-2-carboxamide (268)**

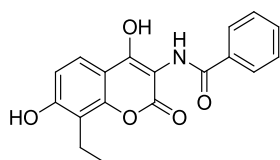
100% conversion by HPLC. Product isolated as an amorphous pale yellow solid (19 mg, 90%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO-*d*<sub>6</sub>) 11.89 (s, 1H, OH), 11.65 (br. s, 1H, NH), 10.41 (s, 1H, OH), 9.03 (s, 1H, NH), 7.58 (d, *J*=8.6 Hz, 1H, Ar-H), 7.04 (br. s, 1H, Ar-H), 6.96 (d, *J*=1.0 Hz, 1H, Ar-H), 6.89 (d, *J*=8.7 Hz, 1H, Ar-H), 6.17 (d, *J*=3.1 Hz, 1H, Ar-H), 2.17 (d, *J*=128.7 Hz, 3H, <sup>13</sup>CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO-*d*<sub>6</sub>) 160.8 (C), 160.8 (CO), 159.5 (C), 159.0 (C), 151.2 (C), 125.4 (C), 122.4 (CH), 121.5 (CH), 112.3 (CH), 111.9 (CH), 109.0 (CH), 108.1 (C), 100.1 (C), 8.1 (<sup>13</sup>CH<sub>3</sub>). *One CO signal not observed.* **Mpt.:** Decomp. > 300 °C.  **$\nu_{\max}$  (neat):** 3343, 1587, 1555, 1525, 1401, 729 cm<sup>-1</sup>. **HRMS:** C<sub>14</sub><sup>13</sup>CH<sub>13</sub>N<sub>2</sub>O<sub>5</sub> [M+H]<sup>+</sup> requires *m/z* 302.0853, found 302.0840. **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.33.

**1-(<sup>13</sup>C)-methylanthralene-2,7-diol (271)**

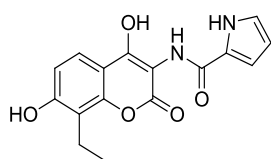
94% conversion by HPLC. Product isolated as a dark pink solid (18 mg, 83%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, MeOD-d<sub>4</sub>) 7.56 (d,  $J=8.7$  Hz, 1H, Ar-H), 7.43 (d,  $J=8.9$  Hz, 1H, Ar-H), 7.13 (d,  $J=2.3$  Hz, 1H, Ar-H), 6.85 (d,  $J=8.9$  Hz, 1H, Ar-H), 6.83 (dd,  $J_1=8.7$ ,  $J_2=2.3$  Hz, 1H, Ar-H), 2.38 (d,  $J=126.6$  Hz, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, MeOD-d<sub>4</sub>) 156.8 (C), 153.6 (C), 137.3 (C), 131.0 (CH), 127.9 (CH), 125.3 (d,  $J=2.5$  Hz, C), 115.9 (d,  $J=1.9$  Hz, CH), 115.6 (CH), 106.0 (d,  $J=4.0$  Hz, CH), 10.8 (<sup>13</sup>CH<sub>3</sub>). *One C signal not detected.* **Mpt.:** Decomp. >170 °C.  **$\nu_{\max}$  (neat):** 3284, 1634, 1185, 1068, 833 cm<sup>-1</sup>. **HRMS:** C<sub>10</sub><sup>13</sup>CH<sub>11</sub>O<sub>2</sub> [M+H]<sup>+</sup> requires  $m/z$  175.0754, found 175.0704. **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.28.

**N-(8-ethyl-4,7-dihydroxy-2-oxo-2H-chromen-3-yl)benzamide (259)**

34% conversion by HPLC. Product isolated as an amorphous colourless solid (4 mg, 18%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO-d<sub>6</sub>) 11.72 (br. s, 1H, OH), 10.43 (s, 1H, OH), 9.41 (s, 1H, NH), 8.01 (d,  $J=7.1$  Hz, 2H, 2xAr-H), 7.56-7.66 (m, 2H, 2xAr-H), 7.47-7.56 (m, 2H, 2xAr-H), 6.90 (d,  $J=8.7$  Hz, 1H, Ar-H), 2.74 (q,  $J=7.3$  Hz, 2H, CH<sub>2</sub>), 1.12 (t,  $J=7.4$  Hz, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO-d<sub>6</sub>) 166.6 (CO), 160.9 (CO), 160.5 (C), 158.8 (C), 151.1 (C), 133.9 (C), 131.5 (CH), 128.1 (CH), 128.0 (CH), 121.8 (CH), 116.6 (C), 112.1 (CH), 108.0 (C), 99.9 (C), 15.8 (CH<sub>2</sub>), 13.5 (CH<sub>3</sub>). **Mpt.:** Decomp. >265 °C.  **$\nu_{\max}$  (neat):** 3331, 1668, 1571, 1543, 1103, 703 cm<sup>-1</sup>. **HRMS:** C<sub>18</sub>H<sub>16</sub>NO [M+H]<sup>+</sup> requires  $m/z$  326.1023, found 326.1012. **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.45.

***N*-(8-ethyl-4,7-dihydroxy-2-oxo-2H-chromen-3-yl)-1H-pyrrole-2-carboxamide (260)**

30% conversion by HPLC. Product isolated as an amorphous pale yellow solid (6 mg, 27%).

**<sup>1</sup>H NMR:**  $\delta$  (400 MHz, DMSO-*d*<sub>6</sub>) 11.88 (s, 1H, OH), 11.65 (br. s, 1H, NH), 10.40 (s, 1H, OH), 9.02 (s, 1H, NH), 7.58 (d, *J*=8.6 Hz, 1H, Ar-H), 7.04 (br. s, 1H, Ar-H), 6.96 (br. s, 1H, Ar-H), 6.89 (d, *J*=8.7 Hz, 1H, Ar-H), 6.17 (d, *J*=3.2 Hz, 1H, Ar-H), 2.73 (q, *J*=7.4 Hz, 2H, CH<sub>2</sub>), 1.11 (t, *J*=7.4 Hz, 3H, CH<sub>3</sub>). **<sup>13</sup>C NMR:**  $\delta$  (101 MHz, DMSO-*d*<sub>6</sub>) 160.9 (C), 160.8 (CO), 159.6 (C), 158.7 (C), 150.9 (C), 125.4 (C), 122.4 (CH), 121.7 (CH), 116.5 (C), 112.3 (CH), 112.1 (CH), 109.0 (CH), 100.1 (C), 15.8 (CH<sub>2</sub>), 13.5 (CH<sub>3</sub>). *One CO signal not observed.* **Mpt.:** Decomp. > 260 °C.  **$\nu_{\max}$  (neat):** 3354, 1576, 1555, 1405, 736 cm<sup>-1</sup>. **HRMS:** C<sub>16</sub>H<sub>15</sub>N<sub>2</sub>O<sub>5</sub> [M+H]<sup>+</sup> requires *m/z* 315.0976, found 315.0965. **R<sub>f</sub>:** (1:1 EtOAc/heptane) 0.42.

### 6.3 Gene synthesis and enzyme expression

#### Gene synthesis and cloning (SalL, NovO and CouO)

Genes for NovO, CouO and SalL (see Appendix) were codon optimised for *E. coli* and subcloned into a pET26b(+) plasmid using NdeI and XhoI restriction sites with a C-terminal (NovO and CouO) or N-terminal (SalL) 6xHis tag at GenScript. The plasmids were received from GenScript in the lyophilised form. Plasmid DNA was resuspended in 40  $\mu$ L water. *E. coli* TOP10 competent cells were transformed using the heat shock method with 5  $\mu$ L of DNA solution. Transformants were plated on LB agar containing kanamycin (50  $\mu$ g/mL) and grown at 37 °C overnight. Colonies were inoculated into 10 mL LB media containing kanamycin (50  $\mu$ g/mL) and grown overnight at 30 °C. The DNA was then extracted and purified using a Qiagen Miniprep kit according to the manufacturer's instructions, eluting in 100  $\mu$ L elution buffer.

#### Enzyme expression

##### SalL

The plasmid pET26b(+)-SalL was transformed into *E. Coli* BL21 (DE3) competent cells (Invitrogen) for protein expression. Transformants harbouring the plasmids were grown at 37 °C in LB medium supplemented with kanamycin (50  $\mu$ g/mL) to provide the seed culture.



The seed culture was used to inoculate 1 L Overnight Express™ media and grown to an optical density (OD) of 2, then incubated for a further 16 hours at 18 °C. The cell pellet was harvested by centrifugation (1100 g, 20 minutes, 4 °C) and stored at -80°C.

#### **Preparation of *E. coli* cell free extract containing SalL for biotransformations**

The cell pellet from the overexpression of SalL was resuspended in potassium phosphate buffer (50 mM, pH 6.8) at a volume of 5 mL per g of wet cell pellet. The resuspended cell pellets were sonicated using a 16 mm probe, on ice, for a total of 5 minutes (9.0 sec on, 10.0 sec off), at 40% amplitude. The resulting cell lysate was separated by centrifugation (1100 g, 10 minutes, 4 °C) and the supernatant stored at -80 °C until further use in biotransformations.

#### **CouO and NovO (WT and mutants)**

The plasmids pET26b(+)-CouO and pET26b(+)-NovO or mutants obtained from SDM of this plasmid DNA were transformed into *E. Coli* BL21 (DE3) competent cells (Invitrogen) for protein expression. Transformants harbouring the plasmids were grown at 30 °C in LB media containing kanamycin (50 µg/mL) to provide the seed cultures. For each construct, Magic media containing ‘Component B’ (5% v/v) and kanamycin (50 µg/mL) was inoculated with 20 mL seed culture (2% v/v). The flasks were incubated at 30 °C at 200 rpm and grown to an OD of ~2 before the temperature was reduced to 18 °C and incubated at 200 rpm overnight. The cells were harvested by centrifugation (4400 rpm, 4 °C, 20 minutes) and stored at -80 °C.

#### **Preparation of *E. coli* cell free extracts containing NovO or CouO for biotransformations**

Crude cell lysates were prepared by resuspending a portion of the cell pellet from the overexpression of CouO or NovO in potassium phosphate buffer (50 mM, pH 6.5 [NovO], pH 7.0 [CouO]) at a volume of 10 mL per g wet cell pellet. The resuspended cell pellets were sonicated using a 16 mm probe, on ice, for a total of 5 minutes (9.0 sec on, 10.0 sec off), at 40% amplitude. The resulting cell lysate solution was stored at -80 °C until further use in biotransformations.

#### **SelMet NovO**

The plasmid pET26b(+)-NovO was transformed into *E. coli* BL21 (DE3) competent cells (Invitrogen) for protein expression. Transformants harbouring the plasmids were grown at 37 °C in 100 mL LB medium containing kanamycin (50 µg/mL) and glucose (1% v/v) for

7 hours at 37 °C, 240 rpm. The seed culture was used to inoculate 1 L Modified Terrific Broth (MTB) supplemented with kanamycin (50 µg/mL) and glucose (4% v/v), using 2% inoculant. The culture was grown overnight at 30 °C, 180 rpm. The culture was then spun aseptically (4700 rpm, 30 minutes, 22 °C) and the supernatant was discarded. The pellet was resuspended in G1X minimal media supplemented with sorbitol (440 mM), betaine (1.5 mM), glycerol (1% v/v), kanamycin (50 µg/mL) and the amino acids threonine, lysine, phenylalanine (each at 100 µg/mL) and leucine, isoleucine, valine and selenomethionine (each at 50 µg/mL). The culture was incubated at 37 °C, 180 rpm for 30 minutes and then at 18 °C, 180 rpm for 30 minutes before inducing with isopropyl β-D-1-thiogalactopyranoside (IPTG) (final concentration of 100 µM) and incubating overnight at 18 °C, 180 rpm. The cell pellet was harvested by centrifugation (4400 rpm, 4 °C, 20 minutes) and stored at -80 °C.

### **Gene synthesis and cloning (SAHH, MTAN and LuxS)**

Genes for SAHH, MTAN and LuxS (see Appendix) were codon optimised for *E. coli* and subcloned into a pET21b(+) plasmid using NdeI and XhoI restriction sites with an N-terminal Tev cleavage site followed by a 6xHis tag at GenScript. The plasmids were received from GenScript in the lyophilised form, which was resuspended in 40 µL water. *E. coli* TOP10 competent cells were transformed using the heat shock method with 5 µL of DNA solution. Transformants were plated on to LB agar containing ampicillin (100 µg/mL) and grown at 37 °C overnight. Colonies were inoculated into 10 mL LB media containing ampicillin (50 µg/mL) and grown overnight at 37 °C. The DNA was isolated using a Qiagen Miniprep kit according to the manufacturer's instructions, eluting in 100 µL elution buffer.

### **SAHH, MTAN and LuxS (Method A)**

The plasmids pET21b(+)-SAHH, pET21b(+)-MTAN and pET21b(+)-LuxS were transformed into *E. Coli* BL21 (DE3) competent cells (Invitrogen) for protein expression. Transformants harbouring the plasmids were grown at 37 °C in LB medium containing ampicillin (100 µg/mL) overnight to provide the seed cultures. For each construct, Terrific Broth (TB) containing ampicillin (100 µg/mL) was inoculated with seed culture (2% v/v) and incubated at 37 °C, 200 rpm and grown to an OD of ~0.6 before adding IPTG (500 mM) and incubating at 15 °C, 200 rpm overnight. The cell pellets were harvested by centrifugation (4400 rpm, 20 minutes, 4 °C) and stored at -80 °C.

### **MTAN and LuxS (Method B)**

The plasmids pET21b(+)-MTAN and pET21b(+)-LuxS were transformed into *E. Coli* BL21 (DE3) competent cells (Invitrogen) for protein expression. Transformants harbouring the

plasmids were grown at 37 °C in LB medium supplemented with ampicillin (100 µg/mL) overnight to provide seed cultures. For each construct, Overnight Express (OE) containing ampicillin (100 µg/mL) was inoculated with seed culture (2% v/v) and incubated at 30 °C, 200 rpm and grown to an OD of ~2 before lowering the temperature to 18 °C and incubating at 200 rpm overnight. The cell pellets were harvested by centrifugation (4400 rpm, 20 minutes, 4 °C) and stored at -80 °C.

### **General procedure for the expression of CASTing libraries in 96-well plate format**

2 µL CASTing library plasmid DNA was transformed into *E. Coli* BL21 (DE3) competent cells (Invitrogen) for protein expression *via* the heat shock method. Each well of a sterile 96-well plate containing 200 µL LB containing kanamycin (50 µg/mL) and glycerol (10% v/v) was inoculated with a single colony from the transformation and the seed cultures were incubated at 37 °C, 220 rpm for 4 hours. 15 µL of seed culture was used to inoculate 800 µL Magic Media containing 'Component B' (5% v/v) and kanamycin (50 µg/mL). The cultures were incubated at 30 °C, 800 rpm and grown to an OD of ~2 before incubating at 18 °C, 800 rpm overnight. The cultures were harvested by centrifugation (4000 rpm, 20 minutes, 4°C) and stored at -80 °C.

#### **6.3.1 Enzyme purification**

**General procedure for the purification of NovO WT, NovO mutants F14L, R24F, H120N, N117A, Y216F, R116L, F14M, R24M, H120F, R121Q, R121L, R116K, CouO WT and SelMet-NovO.**

#### **Affinity chromatography**

Binding Buffer: 50 mM Tris-HCl, 300 mM NaCl, pH 8.0

Elution Buffer: 50 mM Tris-HCl, 300 mM NaCl, 500 mM imidazole, pH 8.0

The cell pellet harbouring NovO or CouO was resuspended in binding buffer at 10 mL per g wet cell pellet and lysed by sonication (5 minutes, 9.9 seconds on, 9.9 seconds off, 40% intensity). The cell lysate was clarified by centrifugation (100 000 g, 90 minutes, 4 °C) and loaded onto a pre-equilibrated HisTRAP HP 5 mL column at a flow rate of 1.5 mL/min, collecting the flow through. The protein was eluted at a flow rate of 1.5 mL/min using the one of following gradients as specified:

#### **Method A (NovO WT)**

Step 1: 0-20 mM imidazole (0-4% elution buffer) for 10 CV

Step 2: 20-50 mM imidazole (4-10% elution buffer) for 10 CV

Step 3: 50-500 mM imidazole (10-100% elution buffer) for 10 CV

**Method B (CouO)**

Step 1: 0-20 mM imidazole (0-4% elution buffer) for 10 CV

Step 2: 20-250 mM imidazole (4-50% elution buffer) for 20 CV

**Method C (NovO mutants F14L, R24F, H120N, N117A, Y216F, R116L, F14M, R24M, H120F, R121Q, R121L, R116K and SelMet-NovO)**

Step 1: 0-20 mM imidazole (0-4% elution buffer) for 10 CV

Step 2: 20-500 mM imidazole (4-100% elution buffer) for 15 CV

Step 3: 500 mM imidazole (100% elution buffer) for 6 CV

Fractions containing the protein of interest were determined by SDS PAGE, pooled and concentrated to 3 mL. The sample was desalted using a PD-10 column (GE healthcare) according to the supplier's manual, eluting with 100 mM Tris-HCl, 150 mM NaCl, 40% glycerol, pH 8. Protein identity was confirmed by PMF (WT NovO: 53% coverage, score: 150; CouO: 54% coverage, score: 132). The protein was portioned into 0.5 mL aliquots and stored at -80 °C.

**Size exclusion chromatography (SEC) of WT NovO, WT CouO and SelMet-NovO**

SEC buffer for CouO: 100 mM Tris-HCl, 150 mM NaCl, pH 8

SEC buffer for NovO: 100 mM Tris-HCl, 150 mM NaCl, 5% glycerol, pH 8

A HiLoad 26/60 320 mL Superdex 75 Prep Grade SEC Column was equilibrated with SEC buffer and loaded with 10 mL of the pooled fractions from the affinity chromatography purification step. The column was eluted with SEC buffer over 1.2 CV at 1.5 mL/min, collecting 2 mL fractions, which were analysed by SDS PAGE. Fractions containing pure protein were pooled and the protein concentration determined by Bradford Assay. Solutions containing NovO and CouO after SEC were concentrated as required using an Amicon Ultra 15 centrifugal Ultracel with a 10 kDa retention limit.

**General Procedure for the purification of NovO H120K, NovO R121K, NovO N117H and CouO H117N**

Lysis buffer: 100 mM Tris, 50 mM NaCl, 10% isopropanol, lysozyme (1 mg/mL), pH 8.2

Binding buffer: 50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 8

The cell pellet harbouring NovO H120K or NovO R121K was resuspended in binding buffer at 10 mL per g wet cell pellet and lysed by sonication (5 minutes, 0.5 seconds on, 0.5 seconds off, 40% amplitude). The cell lysate was incubated on ice for a further 10 minutes before being clarified by centrifugation (100 000 g, 90 minutes, 4 °C). The clarified lysate was loaded onto a pre-equilibrated HisTRAP HP 5 mL column at a flow rate of 1 mL/min, collecting the flow through. The column was washed with binding buffer until UV absorbance was stable. The protein was eluted with a gradient of 0-100% elution buffer over 20 CV to a final concentration of 500 mM imidazole, collecting 2 mL fractions. Fractions containing the desired protein were determined by SDS PAGE, pooled and concentrated to 3 mL. The sample was desalted using a PD-10 column (GE healthcare), eluting with 100 mM Tris, 150 mM NaCl, 40% glycerol, pH 8. Total yield from 200 mL culture: 1.56 mg (NovO H120K); 28.07 mg (NovO R121K); 1.56 mg (NovO H120K); 28 mg (NovO R121K). The protein was portioned into 0.5 mL aliquots and stored at -80 °C for further use.

#### **General Procedure for the Purification of Sall**

Binding Buffer: 50 mM potassium phosphate, 20 mM imidazole, 500 mM Na<sub>2</sub>SO<sub>4</sub>, pH 7.9

Elution Buffer: 50 mM potassium phosphate, 500 mM imidazole, 500 mM Na<sub>2</sub>SO<sub>4</sub>, pH 7.9

The cell pellet harbouring Sall was resuspended in binding buffer at 10 mL per g wet cell pellet and lysed by sonication (5 minutes, 9.9 seconds on, 9.9 seconds off, 40% amplitude). The cell lysate was clarified by centrifugation (100 000 g, 90 minutes, 4 °C) and loaded onto a pre-equilibrated HisTRAP HP 5 mL column at a flow rate of 1.5 mL/min, collecting the flow through. The column was washed with binding buffer until UV absorbance was stable. The protein was eluted with a gradient of 0-100% elution buffer at 1 mL/min over 20 CV to a final concentration of 500 mM imidazole, collecting 2 mL fractions. Fractions containing Sall were determined by SDS PAGE, pooled and concentrated to 3 mL. The sample was desalted using a PD-10 column (GE healthcare), eluting with 50 mM phosphate buffer, pH 7.5 containing 10% glycerol. Total yield from 1 L culture: 25 mg. Protein identify was confirmed by PMF (% coverage: 43%, score: 56). The protein was portioned into 0.5 mL aliquots and stored at -80 °C.

#### **General Procedure for the Purification of MTAN**

Binding Buffer: 50 mM potassium phosphate buffer pH 7.5

Elution buffer: 50 mM potassium phosphate buffer pH 7.5, 500 mM imidazole

The cell pellet harbouring MTAN was resuspended in binding buffer at 10 mL per g wet cell pellet and lysed by sonication (5 minutes, 0.5 seconds on, 0.5 seconds off, 40% amplitude). The cell lysate was clarified by centrifugation (100 000 g, 90 min, 4 °C) and loaded onto a pre-equilibrated HisTRAP HP 1 mL column at a flow rate of 1 mL/min, collecting the flow through. The column was washed with 25 mM imidazole until UV absorbance was stable. The protein was eluted with a gradient of 4-50% elution buffer over 20 CV to a final concentration of 250 mM imidazole, collecting 2 mL fractions. Fractions containing MTAN were determined by SDS PAGE, pooled and concentrated to 3 mL. The sample was desalted using a PD-10 column (GE healthcare), eluting with 50 mM HEPES, pH 8 containing 10% glycerol. Total yield from 1 L culture: 20 mg. Protein identify was confirmed by PMF (50% coverage, score: 54). The protein was portioned into 0.5 mL aliquots and stored at -80 °C for further use.

#### **General Procedure for the Purification of LuxS**

Binding Buffer: 50 mM potassium phosphate buffer pH 7.2

Elution buffer: 50 mM potassium phosphate buffer pH 7.2, 500 mM imidazole

The cell pellet harbouring LuxS was resuspended in binding buffer at 10 mL per g wet cell pellet and lysed by sonication (5 minutes, 0.5 seconds on, 0.5 seconds off, 40% amplitude). The cell lysate was clarified by centrifugation (100 000 g, 90 min, 4 °C) and loaded onto a pre-equilibrated HisTRAP HP 1 mL column at a flow rate of 1 mL/min, collecting the flow through. The column was washed with 25 mM imidazole until UV absorbance was stable. The protein was eluted with a gradient of 4-100% elution buffer over 20 CV to a final concentration of 500 mM imidazole, collecting 2 mL fractions. Fractions containing LuxS were determined by SDS PAGE, pooled and concentrated to 2.5 mL. The sample was desalted using a PD-10 column (GE healthcare), eluting with HEPES (50 mM, pH 8) containing 10% glycerol. Total yield from 100 mL culture: 42 mg. Protein identify was confirmed by PMF (44% coverage, score: 52). The protein was portioned into 0.5 mL aliquots and stored at -80 °C.

#### **General Procedure for the Purification of SAHH**

Binding Buffer: 50 mM Phosphate buffer pH 7, 300 mM NaCl, 10% glycerol

Elution buffer: 50 mM Phosphate buffer pH 7, 300 mM NaCl, 10% glycerol, 500 mM imidazole

The cell pellet harbouring SAHH was resuspended in binding buffer at 10 mL per g wet cell pellet and lysed by sonication (5 minutes, 0.5 seconds on, 0.5 seconds off, 40% amplitude). The cell lysate was clarified by centrifugation (100 000 g, 90 min, 4 °C) and loaded onto a pre-equilibrated HisTRAP HP 1 mL column at a flow rate of 1 mL/min, collecting the flow through. The column was washed with 25 mM imidazole until the UV absorbance was stable. The protein was eluted with a gradient of 4-100% elution buffer over 20 CV to a final concentration of 500 mM imidazole, collecting 2 mL fractions. Fractions containing SAHH were determined by SDS PAGE, pooled and concentrated to 3 mL. The sample was desalted using a PD-10 column (GE healthcare), eluting with 20 mM potassium phosphate buffer pH 7.2, 100 mM NaCl, 1 mM EDTA and 10% glycerol. Total yield from 100 mL culture: 42 mg. Protein identify was confirmed by PMF (35% coverage, score: 134). The protein was portioned into 0.5 mL aliquots and stored at -80 °C.

#### **Purification of CASTing library 2 using KingFisher™ automated purification system**

Lysis buffer: 50 mM potassium phosphate, 300 mM NaCl, 0.5 mg/mL lysozyme, 25 mU/mL benzonase, pH 6.5

Wash buffer: 50 mM potassium phosphate, 20 mM imidazole, pH 6.5

Elution buffer: 50 mM potassium phosphate buffer, 500 mM imidazole, pH 6.5

Plate setup:

- Bead/wash plate (loaded with 900 µL wash buffer and 40 µL Promega MagneHis Ni-Particles #V854B/ well)
- 4 x Wash plates (loaded with 500 µL wash buffer/ well)
- Elution plate (loaded with 200 µL wash buffer/elution buffer to final concentration of 150 mM imidazole/ well).

Method:

The cell pellets from the overexpression of the CASTing library (in a deep-well 96 well plate) were resuspended in 700 µL lysis buffer and incubated at 22 °C, 800 rpm for 1 hour.

The plate was purified using the KingFisher™ purification system (see Appendix), eluting in 150 mM imidazole and the resulting purified protein used directly in the MTAN/LuxS coupled MT assay.

## 6.4 Enzyme assays

### General procedure for the synthesis of SAM using SalL

To an eppendorf vial containing 900  $\mu$ L supernatant containing SalL (see 'Preparation of *E. coli* cell free extract containing SalL for biotransformations' above) was added L-Met (15 mM, from a 150 mM stock solution in 250 mM aq. NaOH) and CIDA (**85**) (180  $\mu$ M, from a 300 mM stock solution in DMSO) giving a net pH of 7.4. The reaction mixture was incubated using a thermomixer at 750 rpm, 37 °C for 2 hours or until consumption of CIDA was observed by HPLC (Appendix, Method C).

### General procedure for methylation of aminocoumarin substrates with WT CouO, WT NovO, CouO mutants and NovO mutants

To an eppendorf vial containing 700  $\mu$ L *E. coli* cell lysate containing NovO or CouO (see Preparation of *E. coli* cell lysate containing NovO or CouO for biotransformations' above) was added DTT (100  $\mu$ L, from a 10 mM stock solution in 50 mM potassium phosphate buffer, pH 6.5 [NovO] or 7.0 [CouO] to a final concentration of 1 mM), SAM (100  $\mu$ L, from a 20 mM stock solution in phosphate buffer, pH 6.5 [NovO] or 7.0 [CouO] to a final concentration of 2 mM), substrate (10  $\mu$ L, from a 100 mM stock solution in DMSO to a final concentration of 1 mM) and bovine serum albumin (BSA) (1 mg/mL). The reaction mixture was incubated using a thermomixer at 750 rpm, 35 °C for 24 hours and the percentage conversion to methylated product analysed by UPLC (Appendix, Method A with 320 nm detection for aminocoumarin substrates and 220 nm detection for dihydroxynaphthalene substrates).

### General procedure for tandem SAM synthesis and methylation of aminocoumarins using SalL (cell free extract) and NovO or CouO (crude cell lysates) on screening scale

To an eppendorf vial was added 450  $\mu$ L *E. coli* cell lysate containing NovO or CouO (see Preparation of *E. coli* cell lysate containing NovO or CouO for biotransformations' above) and 450  $\mu$ L *E. coli* cell free extract containing SalL (see 'Preparation of *E. coli* cell free extract containing SalL for biotransformations' above). L-Met (4  $\mu$ L, from a 500 mM stock solution in 250 mM aq. NaOH to a final concentration of 2 mM), **85** (20  $\mu$ L, from a 100 mM stock solution in DMSO to a final concentration of 2 mM), substrate (10  $\mu$ L, from a 100 mM stock solution in DMSO to a final concentration of 1 mM), DTT (10  $\mu$ L, from a 100 mM stock solution in 50 mM potassium phosphate buffer to a final concentration of 1 mM) and BSA (1 mg/mL) were added and the total reaction volume was made up to 1 mL with 50 mM potassium phosphate buffer at pH 6.5 (NovO) or pH 7.0 (CouO). The reaction volume



mixture was incubated using a thermomixer at 35 °C, 750 rpm and reaction samples were quenched with the addition of an equal volume of aq. NaOH (250 mM). Samples were clarified by centrifugation (13 200 rpm, 15 minutes, 4 °C) and analysed by UPLC (Appendix, Method A).

**General procedure for the tandem SAM or SAM analogue synthesis and alkylation of aminocoumarins using SalL and NovO or CouO (purified enzymes; screening scale)**

To an eppendorf vial was added 50 mM potassium phosphate buffer pH 6.5 (NovO) or pH 7.0 (CouO) (438 µL), aminocoumarin substrate **67** (5 µL, from a 100 mM stock solution in DMSO to a final concentration of 1 mM), L-Met (2 µL, from a 500 mM stock solution in 250 mM aq. NaOH to a final concentration of 2 mM) or DL-Met analogue (4 µL, from a 500 mM stock solution in 250 mM aq. NaOH to a final concentration of 4 mM), **85** (7.5 µL, from a 100 mM stock solution in DMSO to a final concentration of 1.5 mM), DTT (5 µL, from a 100 mM stock solution in potassium phosphate buffer [50 mM, pH 6.5] to a final concentration of 1 mM), BSA (1 mg/mL) and SalL (5 µM). The reaction was incubated at 37 °C, 750 rpm for 4 hours before adding NovO (40 µM) and MTAN (0.5 µM) and incubating the reactions for a further 16 hours at 37 °C, 750 rpm. Reaction samples were quenched by addition of an equal volume of MeCN, clarified by centrifugation (13 200 rpm, 4 °C, 15 min) and analysed by HPLC (Appendix method A) and LC-MS (Appendix Methods A and/or B).

**General Procedure for KIE measurements**

All reactions were carried out in triplicate in 96-well plate on format 200 µL (total reaction volume) scale using a multichannel pipette for reagent addition and sampling of aliquots.

To a solution of **67** (from a 5 or 25 mM stock solution in DMSO to a final reaction volume of 15, 30, 50, 100, 150, 300, 500 or 1000 µM), **85** (10 µL, from a 100 mM stock solution in DMSO to a final concentration of 2 mM) and methionine, methionine-<sup>13</sup>C or methionine-d<sub>3</sub> as appropriate (5 µL, from a 100 mM stock solution in 250 mM aq. NaOH to a final concentration of 2.5 mM) in phosphate buffer (50 mM, pH 6.5 to a final reaction volume of 200 µL) was added SalL<sup>[221,224]</sup> (final concentration 0.1 mg/mL). The reaction was incubated at 37 °C, 750 rpm for 2 hours. Methyl transfer reactions were initiated by the addition of 20 µL NovO (final concentration 0.5 mg/mL) and quenched after 3 minutes by transferring 50 µL of the reaction into a preheated vial and incubated at 97 °C for 10 minutes before clarifying by centrifugation (15 minutes, 13 200 rpm, 4 °C). The supernatant was transferred

to an UPLC insert vial and 5  $\mu$ L injected directly onto the UPLC for analysis (Appendix Method A with 320 nm detection). Percentage conversion for each substrate concentration was calculated by the ratio of area/area% of starting material to product and plotted on a Michaelis-Menten plot.

#### **SAHH activity assay**

To a solution of SAH (500  $\mu$ M, from a 25 mM stock solution in DMSO) in 50 mM potassium phosphate buffer (pH 7) was added SAHH (72, 36 or 14  $\mu$ M) and 5,5'-dithiobis-(2- nitrobenzoic acid) (DTNB) (500 mM, from a 100 mM solution in 50 mM potassium phosphate buffer (pH 8). The reaction was incubated at 37  $^{\circ}$ C, 750 rpm for 1 hour before quenching reaction samples with MeOH (2:1 MeOH: reaction mixture), clarifying by centrifugation (13 200 rpm, 10 minutes, 4  $^{\circ}$ C) and analysing by HPLC (Appendix Method B).

#### **MTAN and LuxS activity assays**

To a solution of SAH (500  $\mu$ M, from a 25 mM stock solution in DMSO) in 50 mM potassium phosphate buffer (pH 7) was added MTAN (65, 32 or 16  $\mu$ M) and the reaction was incubated at 37  $^{\circ}$ C, 750 rpm for 1 hour before quenching reaction samples with MeOH (2:1 MeOH: reaction mixture), clarifying by centrifugation (13 200 rpm, 10 minutes, 4  $^{\circ}$ C) and analysing by HPLC (Appendix Method B). To the remaining reaction mixture was added LuxS (62 or 12  $\mu$ L) and DTNB (500 mM, from a 100 mM solution in 50 mM potassium phosphate buffer [pH 8]). The reaction was incubated at 37  $^{\circ}$ C, 750 rpm for 1 hour. The presence of homocysteine was confirmed by a colour change from colourless to yellow resulting from the reaction of DTNB with homocysteine.

#### **General Procedure for MTAN/LuxS coupled MT assay using purified enzyme**

To a shallow well 96-well plate was added 50  $\mu$ L purified enzyme from the KingFisher<sup>TM</sup> purification of the CASTing library and 50  $\mu$ L assay buffer containing 2, 7-, 1, 6- or 1, 7-dihydroxynaphthalene substrate (**67**, **68** or **121**) (1 mM, from a 100 mM stock solution in DMSO), SAM (2 mM, from a 20 mM stock solution in 50 mM potassium phosphate buffer [pH 6.5]), MTAN (0.45  $\mu$ M) and LuxS (80  $\mu$ M) in 50 mM potassium phosphate buffer (pH 6.5). The plate was incubated at 37  $^{\circ}$ C, 220 rpm for 16 hours. To detect homocysteine, 30  $\mu$ L of the assay solution from each well was added to 90  $\mu$ L of quench solution, containing 1 mM DTNB in 50 mM potassium phosphate buffer (pH 8). After 30 minutes incubation at 22  $^{\circ}$ C, the absorbance at 405 nm was measured for each well.

**General procedure for generation of single point mutants by SDM**

Site-directed mutagenesis (SDM) was accomplished by a modified QuickChange method A, B or C or B (see below) using primers bearing desired mutations. Colonies containing the desired mutation were identified by gene sequencing using T7 forward and reverse primers and subsequently amplified using a Qiagen Miniprep kit according to the manufacturer's instructions, eluting with 100  $\mu$ L elution buffer. The gene sequence for all mutants was confirmed by gene sequencing using gene sequencing primers for NovO (see Appendix).

**SDM Method A (NovO F14L, R24F, H120N, H120A, N117A, Y216F, R116L)**

To the PCR reaction was added buffer for KOD XL polymerase (5  $\mu$ L, final concentration 1x), dNTPs (5  $\mu$ L, final concentration 200  $\mu$ M), forward and reverse primers (1  $\mu$ L each, final concentration 2  $\mu$ M each), pET26b(+)-NovO-WT (2.5  $\mu$ L, final concentration ~5 ng/ $\mu$ L), KOD XL polymerase (1  $\mu$ L, 2.5 U) and water (to a final volume of 50  $\mu$ L). The PCR reaction was carried out using the following parameters:

1. 95 °C for 3 minutes
2. 95 °C for 30 seconds
3. 60 °C for 30 seconds
4. 72 °C for 6 minutes
5. Repeat steps 2-4 12 times (13 cycles total)
6. 74 °C for 10 minutes
7. Hold at 12 °C

**SDM Method B (NovO F14M, R24M, H120F, R121Q, R121L and R116K)**

As above, but with altered PCR cycle parameters:

1. 95 °C for 3 minutes
2. 95 °C for 30 seconds
3. 65 °C for 30 seconds
4. 72 °C for 6 minutes
5. Repeat steps 2-4 9 times (10 cycles total)
6. 74 °C for 10 minutes
7. Hold at 12 °C

**SDM Method C (NovO H120K and NovO R121K and CASTing libraries 1-5)**

To the PCR reaction was added buffer for Q5 DNA polymerase (10  $\mu$ L, final concentration 1x), DMSO (5% v/v), dNTPs (5  $\mu$ L, final concentration 100  $\mu$ M), forward and reverse primers (2.5  $\mu$ L each, final concentration 10  $\mu$ M each), pET26b(+)-NovO-WT (1  $\mu$ L, final

concentration ~0.8 ng/ $\mu$ L), KOD XL polymerase (0.5  $\mu$ L, 1 U) and water (to a final volume of 50  $\mu$ L). The PCR reaction was carried out using the following parameters:

1. 90 °C for 3 minutes
2. 90 °C for 30 seconds
3. 60 °C for 30 seconds
4. 72 °C for 8 minutes
5. Repeat steps 2-5 11 times (12 cycles in total)
6. 74 °C for 10 minutes
7. 12 °C for ever

### **Dpn1 digestion and Transformation**

1  $\mu$ L Dpn1 added to each PCR reaction product and incubated at 37 °C for 2 h (105 °C lid temperature). 2  $\mu$ L from the Dpn1 digestion was added to ~50  $\mu$ L *E. Coli* Top10 competent cells and incubated on ice for 15 minutes. The cells were then heat shocked at 42 °C for 45 seconds then incubated on ice for 2 minutes. 200  $\mu$ L SOC media was added and the cultures were incubated for 1 hour at 37 °C, 1400 rpm in a thermomixer. The cultures were plated onto LB agar containing kanamycin (50  $\mu$ g/mL) and incubated overnight at 37 °C.

### **Seed culture**

24 colonies for each construct were picked to a 48 deep-welled plate containing 3 mL LB supplemented with glycerol (10% v/v) and Kanamycin (50  $\mu$ g/mL) in each well. The plate was incubated overnight at 37 °C, 200 rpm. From this was taken 200  $\mu$ L from each well to a 96 well plate which was stored at -20 °C until future use. The remaining culture was spun down (15 minutes, 4000 rpm) and the supernatant discarded and the plasmid DNA was extracted from each of the 24 clones for each construct by MiniPrep using either the Kingfisher robot, manifold vacuum line or manually. Each clone was sequenced using T7 forward and reverse primers for identification of colonies bearing the desired mutation.

### **Amplification of hits for gene sequencing**

For each SDM sequencing hit, mL LB supplemented with glycerol (10% v/v) and Kanamycin (50  $\mu$ g/mL) was inoculated with 20  $\mu$ L seed culture and incubated overnight at 240 rpm, 37 °C. The cultures were pelleted by centrifugation (4 °C, 4400 rpm, 20 minutes) and resuspended in 1 mL resuspension buffer. 1 mL lysis buffer was added, followed by 1.2 mL neutralisation buffer. The suspension was aliquoted into 4 x 800  $\mu$ L in Eppendorf vials and the MiniPrep continued according to the supplied protocol. The resulting DNA was sequenced using gene sequencing primers (see Appendix).

### **Preparation of CASTing DNA libraries 1-5**

1  $\mu\text{L}$  Dpn1 was added to each PCR reaction product and incubated at 37 °C for 2 hours (105 °C lid temperature) then heated to 80 °C for 20 minutes. PCR products were purified using a Qiagen PCR purification kit according to the supplier's manual. 2  $\mu\text{L}$  of the resulting PCR products was transformed into  $\sim 50$   $\mu\text{L}$  *E. coli* Top10 using standard transformation procedure. The colonies from the transformation experiments were washed off the agar plate with 2 mL LB containing kanamycin (50  $\mu\text{g}/\text{mL}$ ) glycerol (10% v/v) and was used to inoculate 12 mL LB containing kanamycin (50  $\mu\text{g}/\text{mL}$ ) glycerol (10% v/v). The culture was incubated at 37 °C for 4 hours, 220 rpm and the plasmid DNA isolated by MiniPrep according to suppliers manual.

## **6.5 Determination of SelMet NovO crystal structure**

### **Crystallisation of NovO with SAH**

C-terminal His-tagged NovO 1-230 from *Streptomyces spheroides* at 12 mg/mL in trisaminomethane hydrochloride (Tris-HCl) (0.1 M, pH 8.0) containing 150 mM NaCl was incubated with 2 mM *S*-adenosylhomocysteine (SAH) and left on ice for 30 minutes. Microseeding was used to obtain reproducible SelMet crystals of good quality. Crystallisation was achieved using drops containing 120 nL protein, 120 nL well solution (0.1 M 2-[*N*-morpholino]ethanesulfonic acid, [Mes] pH 6.5, 20% w/v poly[ethylene glycol] [PEG]-4K, 0.6 M NaCl) and 60 nL seed solution in sitting drop Medical Research Council (MRC) plates at 20 °C. Crystals grew over a period of 18 days and were harvested into mother liquor supplemented with 15% glycerol before flash freezing.

### **Data collection, Structure solution and refinement**

X-ray diffraction data were collected at 100 K at the European Synchrotron Radiation Facility (ESRF, beamline ID23.1). Selenomethionine (SelMet) datasets were collected at the SelMet absorption peak (0.97916 Å  $f' = -7.82$ ,  $f'' = 4.98$ ), 1800 images taken with 0.2 degree oscillations to give 360 degrees in total. The data were processed and scaled using autoPROC.<sup>[365]</sup> As there was no known structure of NovO the structure was solved by Single-wavelength Anomalous Dispersion (SAD). AUTOSHARP<sup>[366]</sup> and CCP4<sup>[367]</sup> were used for heavy atom detection, refinement, phasing and density modification. Two molecules of a homodimer were present within the asymmetric unit of the P2<sub>1</sub> crystal, with a solvent fraction of 45.3%. All 12 SelMet sites were found. The autobuilt coordinates provided a good starting point for manual rebuilding of NovO using COOT.<sup>[368]</sup> Structure refinement was carried out using AUTOBUSTER<sup>[369]</sup> utilising TLS refinement and automatic water

---

fitting, with additional round of refinement and modelling were achieved using COOT and `refmac`<sup>[370]</sup> with TLS refinement. The *S*-adenosylhomocysteine (SAH) was present in the active site of both chains (see Appendix). MOGUL and MOLPROBITY were used to confirm the quality of the refined structure. Refinement statistics of the 1.9 Å X-ray homodimeric complex of NovO with SAH are shown in the Appendix.

# **Bibliography**

- 
- [1] N. Kuhl, M. N. Hopkinson, J. Wencel-Delord, F. Glorius, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 10236–54.
- [2] D. Morton, H. M. L. Davies, *J. Org. Chem.* **2016**, *81*, 343–350.
- [3] D. T. McQuade, P. H. Seeberger, *J. Org. Chem.* **2013**, *78*, 6384–9.
- [4] R. Porta, M. Benaglia, A. Puglisi, *Org. Process Res. Dev.* **2016**, *20*, 2–25.
- [5] C. E. Brocklehurst, H. Lehmann, L. La Vecchia, *Org. Process Res. Dev.* **2011**, *15*, 1447–1453.
- [6] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185–94.
- [7] M. Goldsmith, D. S. Tawfik, *Curr. Opin. Struct. Biol.* **2012**, *22*, 406–12.
- [8] N. J. Turner, *Nat. Chem. Biol.* **2009**, *5*, 567–73.
- [9] M. T. Reetz, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 138–74.
- [10] M. T. Reetz, J. D. Carballeira, A. Vogel, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 7745–51.
- [11] J. F. Chaparro-Riggers, K. M. Polizzi, A. S. Bommarius, *Biotechnol. J.* **2007**, *2*, 180–91.
- [12] J. D. Bloom, S. T. Labthavikul, C. R. Otey, F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5869–74.
- [13] E. M. Brustad, F. H. Arnold, *Curr. Opin. Chem. Biol.* **2011**, *15*, 201–10.
- [14] C. Wandrey, A. Liese, D. Kihumbu, *Org. Process Res. Dev.* **2000**, *4*, 286–290.
- [15] M. T. Reetz, *J. Am. Chem. Soc.* **2013**, *135*, 12480–12496.
- [16] H. Kries, R. Blomberg, D. Hilvert, *Curr. Opin. Chem. Biol.* **2013**, *17*, 221–228.
- [17] W. Liu, P. Wang, *Biotechnol. Adv.* **2007**, *25*, 369–384.
- [18] A. G. McDonald, K. F. Tipton, *FEBS J.* **2014**, *281*, 583–92.
- [19] G. W. Huisman, J. Liang, A. Krebber, *Curr. Opin. Chem. Biol.* **2010**, *14*, 122–9.
- [20] M. Fuchs, J. E. Farnberger, W. Kroutil, *European J. Org. Chem.* **2015**, *32*, 6965–6982.
- [21] W. Ding, Q. Li, Y. Jia, X. Ji, H. Qianzhu, Q. Zhang, *ChemBioChem* **2016**, *17*, 1191–7.
- [22] E. Nzymes, R. Banerjee, S. W. Ragsdale, *Annu. Rev. Biochem.* **2003**, *72*, 209–247.
- [23] W. M. Patrick, *Microbiology* **2014**, *160*, 1571–1584.
- [24] R. Clarke, A. D. Smith, K. A. Jobst, H. Refsum, *Arch. Neurol.* **1998**, *55*, 1449–1455.
- [25] K. D. Allen, H. Xu, R. H. White, *J. Bacteriol.* **2014**, *196*, 3315–3323.



- [26] Alton Meister, in *Adv. Enzymol. Relat. Areas Mol. Biol.*, **1951**, pp. 485–486.
- [27] H. Toda, N. Itoh, *Phytochemistry* **2011**, *72*, 337–343.
- [28] L. A. Wessjohann, J. Keim, B. Weigel, M. Dippe, *Curr. Opin. Chem. Biol.* **2013**, *17*, 229–235.
- [29] J. Kim, H. Xiao, J. B. Bonanno, C. Kalyanaraman, S. Brown, X. Tang, N. F. Al-Obaidi, Y. Patskovsky, P. C. Babbitt, M. P. Jacobson, et al., *Nature* **2013**, *498*, 123–6.
- [30] S. L. Berger, *Nature* **2007**, *447*, 407–412.
- [31] U. T. Sankpal, D. N. Rao, **2002**, *37*, 167–197.
- [32] J. Yang, L. Lior-Hoffmann, S. Wang, Y. Zhang, S. Broyde, *Biochemistry* **2013**, *52*, 2828–2838.
- [33] K. Helin, D. Dhanak, *Nature* **2013**, *502*, 480–8.
- [34] C. H. Arrowsmith, C. Bountra, P. V. Fish, K. Lee, M. Schapira, *Nat. Rev. Drug Discov.* **2012**, *11*, 384–400.
- [35] E. L. Greer, Y. Shi, *Nat. Rev. Genet.* **2012**, *13*, 343–57.
- [36] F. H. Andrews, B. D. Strahl, T. G. Kutateladze, *Nat. Chem. Biol.* **2016**, *12*, 662–668.
- [37] T. J. Gonda, R. G. Ramsay, *Nat. Rev. Cancer* **2015**, *15*, 686–694.
- [38] E. H. Estey, *Leukemia* **2013**, *27*, 1803–12.
- [39] A. Mazzolari, D. De Maddis, M. Carini, *Future Med. Chem.* **2016**, *8*, 1589–1607.
- [40] R. A. Copeland, M. E. Solomon, V. M. Richon, *Nat. Rev. Drug Discov.* **2009**, *8*, 724–732.
- [41] K. W. Kuntz, J. E. Campbell, H. Keilhack, R. M. Pollock, S. K. Knutson, M. Porter-Scott, V. M. Richon, C. J. Sneeringer, T. J. Wigle, C. J. Allain, et al., *J. Med. Chem.* **2016**, *59*, 1556–1564.
- [42] H. Schönherr, T. Cernak, *Angew. Chemie - Int. Ed.* **2013**, *52*, 12256–12267.
- [43] N. T. Southall, K. A. Dill, a D. J. Haymet, *J. Phys. Chem.* **2002**, *106*, 521–533.
- [44] P. R. Andrews, D. J. Craik, J. L. Martin, *J. Med. Chem.* **1984**, *27*, 1648–1657.
- [45] C. S. Leung, S. S. F. Leung, J. Tirado-Rives, W. L. Jorgensen, *J. Med. Chem.* **2012**, *55*, 4489–500.
- [46] M. Boehringer, H. Fischer, M. Hennig, D. Hunziker, J. Huwyler, B. Kuhn, B. M. Loeffler, T. Luebbers, P. Mattei, R. Narquizian, et al., *Bioorg. Med. Chem. Lett.* **2010**, *20*, 1106–8.
- [47] G. Yan, A. J. Borah, L. Wang, M. Yang, *Adv. Synth. Catal.* **2015**, n/a-n/a.
- [48] M. Tengg, H. Stecher, L. Offner, K. Plasch, F. Anderl, H. Weber, H. Schwab, M. Gruber-Khadjawi, *ChemCatChem* **2016**, *8*, 1354–1360.

- [49] A.-W. Struck, M. L. Thompson, L. S. Wong, J. Micklefield, *Chembiochem* **2012**, *13*, 2642–55.
- [50] H. U. Kaniskan, K. D. Konze, J. Jin, *J. Med. Chem.* **2015**, *58*, 1596–1629.
- [51] H. Ü. Kaniskan, M. S. Eram, J. Liu, D. Smil, M. L. Martini, Y. Shen, V. Santhakumar, P. J. Brown, C. H. Arrowsmith, M. Vedadi, et al., *Med. Chem. Commun.* **2016**, *7*, 1793–1796.
- [52] B. J. C. Law, M. R. Bennett, M. L. Thompson, C. Levy, S. A. Shepherd, D. Leys, J. Micklefield, *Angew. Chemie Int. Ed.* **2016**, *128*, 2733–2737.
- [53] H. L. Schubert, R. M. Blumenthal, X. Cheng, *Trends Biochem. Sci.* **2003**, *28*, 329–35.
- [54] T. Wlodarski, J. Kutner, J. Towpik, L. Knizewski, L. Rychlewski, A. Kudlicki, M. Rowicka, A. Dziembowski, K. Ginalski, *PLoS One* **2011**, *6*, e23168.
- [55] J. L. Martin, F. M. McMillan, *Curr. Opin. Struct. Biol.* **2002**, *12*, 783–793.
- [56] R. Gana, S. Rao, H. Huang, C. Wu, S. Vasudevan, *BMC Struct. Biol.* **2013**, *13*, 6.
- [57] P. Z. Kozbial, A. R. Mushegian, *BMC Struct. Biol.* **2005**, *5*, 19.
- [58] M. H. Mageroy, D. M. Tieman, A. Floystad, M. G. Taylor, H. J. Klee, *Plant J.* **2012**, *69*, 1043–51.
- [59] J. Vidgren, L. A. Svensson, A. Liljas, *Nature* **1994**, *368*, 354–8.
- [60] J. Martin, J. Begun, M. McLeish, J. Caine, G. Grunewald, *Structure* **2001**, *9*, 977–985.
- [61] W. A. M. Loenen, N. Kresge, H. Tabor, R. D. Simoni, R. L. Hill, N. E. Murray, W. A. M. Loenen, M. Fontecave, M. Atta, E. Mulliez, et al., *Biochem. Soc. Trans.* **2006**, *34*, 330–3.
- [62] G. L. Cantoni, *J. Am. Chem. Soc.* **1952**, *74*, 2942–2943.
- [63] S. Merali, A. B. Clarkson, *FEMS Microbiol. Lett.* **2004**, *237*, 179–86.
- [64] H. Hesse, R. Hoefgen, *Trends Plant Sci.* **2003**, *8*, 259–62.
- [65] J. Kim, H. Xiao, J. Koh, Y. Wang, J. B. Bonanno, K. Thomas, P. C. Babbitt, S. Brown, Y. S. Lee, S. C. Almo, *Nucleic Acids Res.* **2015**, *43*, 4602–4613.
- [66] X. Cheng, S. Kumar, J. Posfai, J. W. Pflugrath, R. J. Roberts, *Cell* **1993**, *74*, 299–307.
- [67] D. K. Ho, J. C. Wu, D. V. Santi, H. G. Floss, *Arch. Biochem. Biophys.* **1991**, *284*, 264–9.
- [68] A. Lyskowski, M. Teng, G. Steinkellner, H. Schwab, M. Gruber-Khadjawi, K. Gruber, *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **2012**, *68*, 698–700.
- [69] D. K. Liscombe, G. V. Louie, J. P. Noel, *Nat. Prod. Rep.* **2012**, *29*, 1238–50.
- [70] S. Horowitz, L. M. A. Dirk, J. D. Yesselman, J. S. Nimtz, U. Adhikari, R. A. Mehl, S. Scheiner, R. L. Houtz, H. M. Al-Hashimi, R. C. Trievel, *J. Am. Chem. Soc.* **2013**,

- 135, 15536–48.
- [71] Q. Du, Z. Wang, V. L. Schramm, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2916–2921.
- [72] J. B. Broderick, B. R. Duffus, K. S. Duschene, E. M. Shepard, *Chem. Rev.* **2014**, *114*, 4229–317.
- [73] S. C. Wang, P. A. Frey, *Trends Biochem. Sci.* **2007**, *32*, 101–110.
- [74] T. L. Grove, J. S. Benner, M. I. Radle, J. H. Ahlum, B. J. Landgraf, C. Krebs, S. J. Booker, *Science* **2011**, *332*, 604–7.
- [75] D. G. Fujimori, *Curr. Opin. Chem. Biol.* **2013**, *17*, 597–604.
- [76] J. Wang, R. P. Woldring, G. D. Román-Meléndez, A. M. McClain, B. R. Alzua, E. N. G. Marsh, *ACS Chem. Biol.* **2014**, DOI 10.1021/cb5004674.
- [77] B. Meunier, S. P. de Visser, S. Shaik, *Chem. Rev.* **2004**, *104*, 3947–80.
- [78] F. Yan, J. M. LaMarre, R. Röhrich, J. Wiesner, H. Jomaa, A. S. Mankin, D. G. Fujimori, *J. Am. Chem. Soc.* **2010**, *132*, 3953–64.
- [79] W. J. Werner, K. D. Allen, K. Hu, G. L. Helms, B. S. Chen, S. C. Wang, *Biochemistry* **2011**, *50*, 8986–8988.
- [80] W. L. Kelly, L. Pan, C. Li, *J. Am. Chem. Soc.* **2009**, *131*, 4327–34.
- [81] R. D. Woodyer, G. Li, H. Zhao, W. a van der Donk, *Chem. Commun. (Camb)*. **2007**, 359–61.
- [82] G. Layer, J. Moser, D. W. Heinz, D. Jahn, W.-D. Schubert, *EMBO J.* **2003**, *22*, 6214–24.
- [83] K. K. J. Chan, S. Thompson, D. O’Hagan, *Chembiochem* **2013**, *14*, 675–7.
- [84] Q. I. Zhang, W. A. V. A. N. D. E. R. Donk, W. E. N. Liu, *Acc. Chem. Res.* **2012**, *45*, 555–564.
- [85] X. Ji, Y. Li, L. Xie, H. Lu, W. Ding, Q. Zhang, *Angew. Chemie Int. Ed.* **2016**, *201605917*, 1–5.
- [86] J. Gui, Q. Zhou, C. Pan, Y. Yabe, A. C. Burns, M. R. Collins, M. a Ornelas, Y. Ishihara, P. S. Baran, R. Michael, *J. Am. Chem. Soc.* **2014**, *136*, 4853–4856.
- [87] A. Jansson, H. Koskiniemi, P. Mäntsälä, J. Niemi, G. Schneider, *J. Biol. Chem.* **2004**, *279*, 41149–41156.
- [88] S. G. Lee, Y. Kim, T. D. Alpert, A. Nagata, J. M. Jez, *J. Biol. Chem.* **2012**, *287*, 1426–34.
- [89] J.-L. J. Ferrer, C. Zubieta, R. R. A. Dixon, J. J. P. Noel, *Plant Physiol.* **2005**, *137*, 1009–1017.
- [90] K. Fesko, M. Gruber-Khadjawi, *ChemCatChem* **2013**, *5*, 1248–1272.

- [91] R. J. Mikolajek, A. C. Spiess, M. Pohl, J. Büchs, *Biotechnol. Prog.* **n.d.**, 25, 132–8.
- [92] M. K. Speedie, U. Hornemann, H. G. Floss, *J. Biol. Chem.* **1975**, 250, 7819–25.
- [93] Y.-T. Huang, S.-Y. Lyu, P.-H. Chuang, N.-S. Hsu, Y.-S. Li, H.-C. Chan, C.-J. Huang, Y.-C. Liu, C.-J. Wu, W.-B. Yang, et al., *Chembiochem* **2009**, 10, 2480–7.
- [94] 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, et al., *Acta Crystallogr. D. Biol. Crystallogr.* **2014**, 70, 1549–60.
- [95] A. Velasco, P. Acebo, A. Gomez, C. Schleissner, P. Rodríguez, T. Aparicio, S. Conde, R. Muñoz, F. De La Calle, J. L. Garcia, et al., *Mol. Microbiol.* **2005**, 56, 144–154.
- [96] L. Li, W. Deng, J. Song, W. Ding, Q. F. Zhao, C. Peng, W. W. Song, G. L. Tang, W. Liu, *J. Bacteriol.* **2008**, 190, 251–263.
- [97] Y. Hu, V. Phelan, I. Ntai, C. M. Farnet, E. Zazopoulos, B. O. Bachmann, *Chem. Biol.* **2007**, 14, 691–701.
- [98] W. Li, A. Khullar, S. Chou, A. Sacramo, B. Gerratana, *Appl. Environ. Microbiol.* **2009**, 75, 2869–78.
- [99] S.-C. Chen, C.-H. Huang, S.-J. Lai, J.-S. Liu, P.-K. Fu, S.-T. Tseng, C. S. Yang, M.-C. Lai, T.-P. Ko, Y. Chen, *Sci. Rep.* **2015**, 5, 10100.
- [100] M. Bentinger, M. Tekle, G. Dallner, *Biochem. Biophys. Res. Commun.* **2010**, 396, 74–79.
- [101] Y.-N. Dai, K. Zhou, D.-D. Cao, Y.-L. Jiang, F. Meng, C.-B. Chi, Y.-M. Ren, Y. Chen, C.-Z. Zhou, *Acta Crystallogr. D. Biol. Crystallogr.* **2014**, 70, 2085–92.
- [102] P. T. Lee, A. Y. Hsu, H. T. Ha, C. F. Clarke, P. T. Lee, A. Y. Hsu, H. T. Ha, C. F. Clarke, *J. Appl. Microbiol.* **1997**, 179, 1748–1754.
- [103] M. Koch, R. Lemke, K.-P. Heise, H.-P. Mock, *Eur. J. Biochem.* **2002**, 270, 84–92.
- [104] M. Pacholec, J. Tao, C. T. Walsh, *Biochemistry* **2005**, 44, 14969–76.
- [105] M. Steffensky, A. Mühlenweg, Z. Wang, S. Li, A. Mu, L. Heide, *Antimicrob Agents Chemother.* **2000**, 44, 1214–1222.
- [106] Z. X. Wang, S. M. Li, L. Heide, *Antimicrob. Agents Chemother.* **2000**, 44, 3040–8.
- [107] H. Stecher, M. Tengg, B. J. Ueberbacher, P. Remler, H. Schwab, H. Griengl, M. Gruber-Khadjawi, *Angew. Chem. Int. Ed. Engl.* **2009**, 48, 9546–8.
- [108] W. H. E. and K. M. J. M C Dawson, D C Elliott, *Data for Biochemical Research*, Oxford, Clarendon Press, **1959**.
- [109] U. T. Sankpal, D. N. Rao, *Nucleic Acids Res.* **2002**, 30, 2628–38.
- [110] M. M. Diver, S. B. Long, *J. Biol. Chem.* **2014**, 289, 26007–26020.
- [111] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**,

215, 403–10.

- [112] M. Tengg, H. Stecher, P. Remler, I. Eiteljörg, H. Schwab, M. Gruber-Khadjawi, *J. Mol. Catal. B Enzym.* **2012**, *84*, 2–8.
- [113] M. Rueping, B. J. Nachtsheim, *Beilstein J. Org. Chem.* **2010**, *6*, 6.
- [114] T. P. Smyth, B. W. Corby, *Org. Process Res. Dev.* **1997**, *1*, 264–267.
- [115] R. R. Naredla, D. A. Klumpp, *Chem. Rev.* **2013**, *113*, 6905–48.
- [116] M. Tengg, *PhD Thesis* **2011**.
- [117] K. Müller, C. Faeh, F. Diederich, *Science* **2007**, *317*, 1881–6.
- [118] D. A. Nagib, D. W. C. MacMillan, *Nature* **2011**, *480*, 224–8.
- [119] M. Schlosser, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 5432–46.
- [120] A. Studer, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 8950–8.
- [121] N. Shibata, A. Matsnev, D. Cahard, *Beilstein J. Org. Chem.* **2010**, *6*, 65.
- [122] I. Ruppert, K. Schlich, W. Volbach, *Tetrahedron Lett.* **1984**, *25*, 2195–2198.
- [123] G. K. S. Prakash, A. K. Yudin, *Chem. Rev.* **1997**, *97*, 757–786.
- [124] N. Miyaura, A. Suzuki, *Chemistry* **1995**, *95*, 2457–2483.
- [125] T. D. Senecal, A. T. Parsons, S. L. Buchwald, *J. Org. Chem.* **2011**, *76*, 1174–6.
- [126] F. Sladojevich, E. McNeill, J. Börgel, S.-L. Zheng, T. Ritter, *Angew. Chemie Int. Ed.* **2015**, *54*, 3712–3716.
- [127] X. Wang, Y. Xu, F. Mo, G. Ji, D. Qiu, J. Feng, Y. Ye, S. Zhang, Y. Zhang, J. Wang, *J. Am. Chem. Soc.* **2013**, *135*, 10330–3.
- [128] C. Zhang, *Adv. Synth. Catal.* **2014**, *356*, 2895–2906.
- [129] Y. Ji, T. Brueckl, R. D. Baxter, Y. Fujiwara, I. B. Seiple, S. Su, D. G. Blackmond, P. S. Baran, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 14411–14415.
- [130] B. Langlois, D. Montègre, N. Roidot, *J. Fluor. Chem.* **1994**, *68*, 63–66.
- [131] T. Besset, C. Schneider, D. Cahard, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 5048–50.
- [132] E. J. Cho, T. D. Senecal, T. Kinzel, Y. Zhang, D. a Watson, S. L. Buchwald, *Science* **2010**, *328*, 1679–81.
- [133] F. O'Hara, D. G. Blackmond, P. S. Baran, *J. Am. Chem. Soc.* **2013**, *135*, 12122–34.
- [134] Y. Ji, T. Brueckl, R. D. Baxter, Y. Fujiwara, I. B. Seiple, S. Su, D. G. Blackmond, P. S. Baran, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 14411–5.
- [135] L. Li, X. Mu, W. Liu, Y. Wang, Z. Mi, C.-J. Li, *J. Am. Chem. Soc.* **2016**, DOI 10.1021/jacs.6b02782.

- [136] R. C. Simon, E. Busto, N. Richter, V. Resch, K. N. Houk, W. Kroutil, *Nat. Commun.* **2016**, *7*, 13323.
- [137] F. O'Hara, R. D. Baxter, a G. O'Brien, M. R. Collins, J. a Dixon, Y. Fujiwara, Y. Ishihara, P. S. Baran, *Nat. Protoc.* **2013**, *8*, 1042–1047.
- [138] M. S. Packer, D. R. Liu, *Nat. Rev. Genet.* **2015**, *16*, 379–394.
- [139] A. Currin, N. Swainston, P. J. Day, D. B. Kell, *Chem. Soc. Rev.* **2015**, *44*, 1172–239.
- [140] K. Steiner, H. Schwab, *Comput. Struct. Biotechnol. J.* **2012**, *2*, e201209010.
- [141] L. Cipolla, M. Lotti, L. De Gioia, F. Nicotra, *J. Carbohydr. Chem.* **2003**, *22*, 631–644.
- [142] K. Pollmann, V. Wray, H. J. Hecht, D. H. Pieper, *Microbiology* **2003**, *149*, 903–913.
- [143] S. Kumar, E. E. Scott, H. Liu, J. R. Halpert, *J. Biol. Chem.* **2003**, *278*, 17178–17184.
- [144] M. Dippe, W. Brandt, H. Rost, a. Porzel, J. Schmidt, L. a. Wessjohann, *Chem. Commun.* **2015**, *51*, 3637–3640.
- [145] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413.
- [146] N. Voloshchuk, J. K. Montclare, *Mol. Biosyst.* **2010**, *6*, 65–80.
- [147] Q. Wang, A. R. Parrish, L. Wang, *Chem. Biol.* **2009**, *16*, 323–336.
- [148] Y. Ravikumar, S. P. Nadarajan, T. Hyeon Yoo, C. S. Lee, H. Yun, *Biotechnol. J.* **2015**, *10*, 1862–1876.
- [149] S. Zheng, I. Kwon, *Biotechnol. J.* **2012**, *7*, 47–60.
- [150] J. N. Kolev, J. M. Zaengle, R. Ravikumar, R. Fasan, *ChemBioChem* **2014**, *15*, 1001–1010.
- [151] H. Ma, X. Yang, Z. Lu, N. Liu, Y. Chen, *PLoS One* **2014**, *9*, DOI 10.1371/journal.pone.0103792.
- [152] R. A. Chica, N. Doucet, J. N. Pelletier, *Curr. Opin. Biotechnol.* **2005**, *16*, 378–384.
- [153] S. Lutz, *Curr. Opin. Biotechnol.* **2011**, *21*, 734–743.
- [154] M. T. Reetz, D. Kahakeaw, R. Lohmer, *Chembiochem* **2008**, *9*, 1797–804.
- [155] R. K. P. Kuipers, H.-J. Joosten, E. Verwiël, S. Paans, J. Akerboom, J. van der Oost, N. G. H. Leferink, W. J. H. van Berkel, G. Vriend, P. J. Schaap, *Proteins* **2009**, *76*, 608–16.
- [156] S. B. J. Kan, R. D. Lewis, K. Chen, F. H. Arnold, *Science (80-. )*. **2016**, *354*, 1048–1051.
- [157] J. G. Kleingardner, K. L. Bren, *Acc. Chem. Res.* **2015**, *48*, 1845–1852.
- [158] B. D. Levin, K. A. Walsh, K. K. Sullivan, K. L. Bren, S. J. Elliott, *Inorg. Chem.* **2015**, *54*, 38–46.

- [159] M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha, A. Vogel, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 4192–6.
- [160] M. T. Reetz, J. D. Carballeira, J. Peyralans, H. Höbenreich, A. Maichele, A. Vogel, *Chem. - A Eur. J.* **2006**, *12*, 6031–6038.
- [161] M. T. Reetz, J. D. Carballeira, *Nat. Protoc.* **2007**, *2*, 891–903.
- [162] R. Agudo, G. D. Roiban, M. T. Reetz, *ChemBioChem* **2012**, *13*, 1465–1473.
- [163] Z. Chen, Y. Ma, M. He, H. Ren, S. Zhou, D. Lai, Z. Wang, L. Jiang, *Appl. Biochem. Biotechnol.* **2015**, *176*, 2267–2278.
- [164] K. N. Trueblood, H. B. Bürgi, H. Burzlaff, J. D. Dunitz, C. M. Gramaccioli, H. H. Schulz, U. Shmueli, S. C. Abrahams, H.-B. Bürgi, *Acta Crystallogr. Sect. A* **1996**, *52*, 770–781.
- [165] M. T. Reetz, P. Soni, L. Ferna, *Chem. Commun.* **2010**, *46*, 8657–8658.
- [166] K. R. Tindall, T. A. Kunkel, *Biochemistry* **1988**, *27*, 6008–6013.
- [167] J. Kaur, R. Sharma, *Crit. Rev. Biotechnol.* **2006**, *26*, 165–199.
- [168] M. Zaccolo, D. M. Williams, D. M. Brown, E. Gherardi, *J. Mol. Biol.* **1996**, *255*, 589–603.
- [169] D. P. Nannemann, W. R. Birmingham, R. A. Scism, B. O. Bachmann, *Future Med. Chem.* **2011**, *3*, 809–19.
- [170] T. S. Wong, D. Roccatano, M. Zacharias, U. Schwaneberg, *J. Mol. Biol.* **2006**, *355*, 858–871.
- [171] K. Chen, F. H. Arnold, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5618–5622.
- [172] Bert van Loo, J. H. L. Spelberg, J. Kingma, T. Sonke, M. G. Wubbolts, and D. B. Janssen, *Chem. Biol.* **2004**, *11*, 961–990.
- [173] R. Carr, M. Alexeeva, M. J. Dawson, V. Gotor-Fernández, C. E. Humphrey, N. J. Turner, *ChemBioChem* **2005**, *6*, 637–639.
- [174] W. P. Stemmer, *Nature* **1994**, *370*, 389–391.
- [175] W. M. Coco, W. E. Levinson, M. J. Crist, H. J. Hektor, a Darzins, P. T. Pienkos, C. H. Squires, D. J. Monticello, *Nat. Biotechnol.* **2001**, *19*, 354–359.
- [176] I. Pardo, A. I. Vicente, D. M. Mate, M. Alcalde, S. Camarero, *Biotechnol. Bioeng.* **2012**, *109*, 2978–2986.
- [177] R. Verma, U. Schwaneberg, D. Roccatano, *Comput. Struct. Biotechnol. J.* **2012**, *2*, e201209008.
- [178] J. Bendl, J., Stourac, J., Sebestova, E., Vavra, O., Musil, M., Brezovsky, J., Damborsky, *Nucleic Acids Res.* **2016**, *44*, 487.
- [179] F. Richter, A. Leaver-Fay, S. D. Khare, S. Bjelic, D. Baker, *PLoS One* **2011**, *6*, 1–12.

- [180] H. Sun, W. L. Yeo, Y. H. Lim, X. Chew, D. J. Smith, B. Xue, K. P. Chan, R. C. Robinson, E. G. Robins, H. Zhao, et al., *Angew. Chemie Int. Ed.* **2016**, *138673*, 14277–14280.
- [181] M. Thomsen, S. B. Vogensen, J. Buchardt, M. D. Burkart, R. P. Clausen, *Org. Biomol. Chem.* **2013**, *11*, 7606–10.
- [182] C. G. Acevedo-Rocha, R. Agudo, M. T. Reetz, *J. Biotechnol.* **2014**, *191*, 3–10.
- [183] H. Xiao, Z. Bao, H. Zhao, *Ind. Eng. Chem. Res.* **2015**, *54*, 4011–4020.
- [184] J.-L. Reymond, V. S. Fluxà, N. Maillard, *Chem. Commun.* **2008**, 34–46.
- [185] A. P. Green, N. J. Turner, E. O'Reilly, *Angew. Chemie - Int. Ed.* **2014**, *53*, 1–5.
- [186] C. L. Hendricks, J. R. Ross, E. Pichersky, J. P. Noel, Z. S. Zhou, *Anal. Biochem.* **2004**, *326*, 100–105.
- [187] K. M. Dorgan, W. L. Wooderchak, D. P. Wynn, E. L. Karschner, J. F. Alfaro, Y. Cui, Z. S. Zhou, J. M. Hevel, *Anal. Biochem.* **2006**, *350*, 249–255.
- [188] R. Wang, G. Ibáñez, K. Islam, W. Zheng, G. Blum, C. Sengelaub, M. Luo, *Mol. Biosyst.* **2011**, *7*, 2970–81.
- [189] G. L. Ellman, *Arch. Biochem. Biophys.* **1959**, *82*, 70–77.
- [190] E. Collazo, J.-F. Couture, S. Bulfer, R. C. Trievel, *Anal. Biochem.* **2005**, *342*, 86–92.
- [191] Y. Su, S. F. Hickey, S. G. L. Keyser, M. C. Hammond, *J. Am. Chem. Soc.* **2016**, *138*, 7040–7047.
- [192] R. Gupta, P. Rathi, N. Gupta, S. Bradoo, *Biotechnol Appl Biochem* **2003**, *37*, 63–71.
- [193] M. S. Weiß, I. V. Pavlidis, C. Vickers, M. Hohne, U. T. Bornscheuer, *Anal. Chem.* **2014**, *86*, 11847–11853.
- [194] D. Ghislieri, A. P. Green, M. Pontini, S. C. Willies, I. Rowles, A. Frank, G. Grogan, N. J. Turner, *J. Am. Chem. Soc.* **2013**, *135*, 10863–10869.
- [195] Y. Ahn, S. B. Ko, M. J. Kim, J. Park, M. Alexeeva, A. Enright, M. J. Dawson, M. Mahmoudian, N. J. Turner, M. Carson, et al., *Tetrahedron Lett.* **2008**, *43*, 1247–1250.
- [196] G. Yang, S. G. Withers, *ChemBioChem* **2009**, *10*, 2704–2715.
- [197] S. J. Kwon, R. Petri, A. L. DeBoer, C. Schmidt-Dannert, *Chembiochem* **2004**, *5*, 1069–74.
- [198] S. Becker, H. Hübenreich, A. Vogel, J. Knorr, S. Wilhelm, F. Rosenau, K. E. Jaeger, M. T. Reetz, H. Kolmar, *Angew. Chemie - Int. Ed.* **2008**, *47*, 5085–5088.
- [199] A. Aharoni, K. Thieme, C. P. C. Chiu, S. Buchini, L. L. Lairson, H. Chen, N. C. J. Strynadka, W. W. Wakarchuk, S. G. Withers, *Nat. Methods* **2006**, *3*, 609–614.
- [200] H. Shim, A. Karlström, S. M. Touami, R. P. Fuller, C. F. Barbas, *Bioorganic Med. Chem. Lett.* **2004**, *14*, 4065–4068.



- [201] U. T. Bornscheuer, J. Altenbuchner, H. H. Meyer, *Biotechnol. Adv.* **1998**, *58*, 554–559.
- [202] H. I. K. Agamiyama, *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 5511–5515.
- [203] P. H. Bessette, J. J. Rice, P. S. Daugherty, *Protein Eng. Des. Sel.* **2004**, *17*, 731–739.
- [204] B. C. Dickinson, M. S. Packer, A. H. Badran, D. R. Liu, *Nat Commun* **2014**, *5*, 5352.
- [205] J. Zhang, X. Wang, E. Su, G. Fang, Y. Ren, D. Wei, *Biochem. Eng. J.* **2008**, *41*, 74–78.
- [206] “S-(5'-Adenosyl)-L-methionine p-toluenesulfonate salt,” can be found under <http://www.sigmaaldrich.com/catalog/product/sigma/a2408?lang=en&region=GB>, **2016**.
- [207] T. D. Huber, F. Wang, S. Singh, B. R. Johnson, J. Zhang, M. Sunkara, S. G. Van Lanen, A. J. Morris, G. N. Phillips, J. S. Thorson, *ACS ChemBio* **2016**, *11*, 2484–2491.
- [208] C. Desiderio, R. A. Cavallaro, A. De Rossi, F. D'Anselmi, A. Fuso, S. Scarpa, *J. Pharm. Biomed. Anal.* **2005**, *38*, 449–456.
- [209] S. E. Wu, W. P. Huskey, R. T. Borchardt, R. L. Schowen, *Biochemistry* **1983**, *22*, 2828–2832.
- [210] Z. J. Lu, G. D. Markham, *J. Biol. Chem.* **2002**, *277*, 16624–16631.
- [211] S. Horikawa, J. Sasuga, K. Shimizu, H. Ozasa, K. Tsukada, *J. Biol. Chem.* **1990**, *265*, 13683–6.
- [212] S. Singh, J. Zhang, T. D. Huber, M. Sunkara, K. Hurley, R. D. Goff, G. Wang, W. Zhang, C. Liu, J. Rohr, et al., *Angew. Chemie Int. Ed.* **2014**, *53*, 3965–3969.
- [213] T. D. Huber, B. R. Johnson, J. Zhang, J. S. Thorson, *Curr. Opin. Biotechnol.* **2016**, *42*, 189–197.
- [214] I. R. Bothwell, K. Islam, Y. Chen, W. Zheng, G. Blum, H. Deng, M. Luo, *J. Am. Chem. Soc.* **2012**, *134*, 14905–12.
- [215] S. Willnow, M. Martin, B. Lüscher, E. Weinhold, *Chembiochem* **2012**, *13*, 1167–73.
- [216] D. F. Iwig, A. T. Grippe, T. A. McIntyre, S. J. Booker, *Biochemistry* **2004**, *43*, 13510–24.
- [217] D. F. I. And, S. J. Booker, *Biochemistry* **2004**, *43*, 13496–13509.
- [218] 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, et al., *FEBS J.* **2014**, *281*, 4224–4239.
- [219] R. Wang, K. Islam, Y. Liu, W. Zheng, H. Tang, N. Lailier, G. Blum, H. Deng, M. Luo, *J. Am. Chem. Soc.* **2013**, *135*, 1048–1056.
- [220] B. J. C. Law, A.-W. Struck, M. R. Bennett, B. Wilkinson, J. Micklefield, *Chem. Sci.* **2015**, *6*, 2885–2892.

- [221] A. S. Eustáquio, F. Pojer, J. P. Noel, B. S. Moore, *Nat. Chem. Biol.* **2008**, *4*, 69–74.
- [222] D. O’Hagan, C. Schaffrath, S. L. Cobb, J. T. G. Hamilton, C. D. Murphy, *Nature* **2002**, *416*, 279.
- [223] J. L. Anglin, L. Deng, Y. Yao, G. Cai, Z. Liu, H. Jiang, G. Cheng, P. Chen, S. Dong, Y. Song, *J. Med. Chem.* **2012**, *55*, 8066–74.
- [224] J. M. Lipson, M. Thomsen, B. S. Moore, R. P. Clausen, J. J. La Clair, M. D. Burkart, *Chembiochem* **2013**, *14*, 950–3.
- [225] H. Deng, L. Ma, N. Bandaranayaka, Z. Qin, G. Mann, K. Kyeremeh, Y. Yu, T. Shepherd, J. H. Naismith, D. O’Hagan, *ChemBioChem* **2014**, *15*, 364–368.
- [226] M. Teng, H. Stecher, L. Offner, K. Plasch, F. Anderl, H. Weber, H. Schwab, M. Gruber-Khadjawi, *ChemCatChem* **2016**, *8*, 1354–1360.
- [227] A. Marbach, K. Bettenbrock, *J. Biotechnol.* **2012**, *157*, 82–8.
- [228] A. Grabski, M. Mehler, D. Drott, *Nat. Methods* **2005**, *2*, 233–235.
- [229] X. Yang, Y. Zhang, *Appl. Microbiol. Biotechnol.* **2013**, *97*, 10423–33.
- [230] T. Sambrook, J. Fritsch E. F., Maniatis, *Molecular Cloning: A Laboratory Manual. 2nd Edition*, New York, **1989**.
- [231] D. Koma, T. Sawai, S. Harayama, K. Kino, *Appl. Microbiol. Biotechnol.* **2006**, *73*, 172–80.
- [232] M. Yamanaka, K. Inaka, N. Furubayashi, M. Matsushima, S. Takahashi, H. Tanaka, S. Sano, M. Sato, T. Kobayashi, T. Tanaka, *J. Synchrotron Radiat.* **2011**, *18*, 84–7.
- [233] A. McPherson, *Methods* **2004**, *34*, 254–65.
- [234] H.-Y. Wu, Y.-S. Cheng, *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2014**, *70*, 378–383.
- [235] P. D. Whanger, *J. Am. Coll. Nutr.* **2002**, *21*, 223–32.
- [236] G. L. Rosano, E. A. Ceccarelli, *Front. Microbiol.* **2014**, *5*, 172.
- [237] E. Cao, Y. Chen, Z. Cui, P. R. Foster, *Biotechnol. Bioeng.* **2003**, *82*, 684–90.
- [238] S. T. Rao, M. G. Rossmann, *J. Mol. Biol.* **1973**, *76*, 241–56.
- [239] H. L. Schubert, R. M. Blumenthal, X. Cheng, *Trends Biochem. Sci.* **2003**, *28*, 329–335.
- [240] C. Notredame, D. G. Higgins, J. Heringa, *J. Mol. Biol.* **2000**, *302*, 205–217.
- [241] J. F. Couture, G. Hauk, M. J. Thompson, G. M. Blackburn, R. C. Trievel, *J. Biol. Chem.* **2006**, *281*, 19280–19287.
- [242] C. Joce, J. Caryl, P. G. Stockley, S. Warriner, A. Nelson, *Org. Biomol. Chem.* **2009**, *7*, 635–638.
- [243] A. Jansson, H. Koskiniemi, A. Erola, J. Wang, P. Mäntsälä, G. Schneider, J. Niemi, *J.*

- Biol. Chem.* **2005**, *280*, 3636–44.
- [244] W. Zheng, G. Ibáñez, H. Wu, G. Blum, H. Zeng, A. Dong, T. Hajian, A. Allal-hassani, M. F. Amaya, A. Siarheyeva, et al., *J* **2012**, *134*, 18004–18014.
- [245] M. P. Maguire, P. L. Feldman, H. Rapoport, *J. Org. Chem.* **1990**, *55*, 948–955.
- [246] M. J. Thompson, A. Mekhalfia, D. L. Jakeman, S. E. V. Phillips, K. Phillips, J. Porter, G. M. Blackburn, *Chem. Commun.* **1996**, 791.
- [247] T. Fukuyama, C.-K. Jow, M. Cheung, *Tetrahedron Lett.* **1995**, *36*, 6373–6374.
- [248] C. C. G. Inc., **2015**.
- [249] Paul Labute and Martin Santavy, “Locating Binding Sites in Protein Structures,” can be found under <https://www.chemcomp.com/journal/sitefind.htm>, **n.d.**
- [250] S. Soga, H. Shirai, M. Kobori, N. Hirayama, *J. Chem. Inf. Model.* **2007**, *47*, 400–406.
- [251] J. W. Cornforth, S. A. Reichard, P. Talalay, H. L. Carrell, J. P. Gluskerld, **1977**, 7292–7300.
- [252] C. C. Valley, A. Cembran, J. D. Perlmutter, A. K. Lewis, N. P. Labello, J. Gao, J. N. Sachs, *J. Biol. Chem.* **2012**, *287*, 34979–34991.
- [253] E. Asakura, T., Adachi, K. and Schwartz, *J. Biol. Chem.* **1978**, *253*, 6423–6425.
- [254] S. N. Gekko, K. and Timasheff, *Biochemistry* **1981**, *20*, 4667–4676.
- [255] S. . Arakawa, T. and Timasheff, *Methods Enzymol.* **1985**, *114*, 49–77.
- [256] G. Lindwall, M.-F. Chau, S. R. Gardner, L. A. Kohlstaedt, *Protein Eng. Des. Sel.* **2000**, *13*, 67–71.
- [257] G. J. Towbin H, Staehelin T, *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 4350–54.
- [258] S. G. Renart J, Reiser J, *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 3116–3120.
- [259] R. Stevens, L. Stevens, N. Price, *Biochem. Educ.* **1983**, *11*, 70.
- [260] H. Stecher, *PhD Thesis n.d.*
- [261] Y. Liu, Y. Lu, M. Prashad, O. Repi?, T. J. Blacklock, *Adv. Synth. Catal.* **2005**, *347*, 217–219.
- [262] K. C. Westaway, *Adv. Phys. Org. Chem.* **2006**, *41*, 217–273.
- [263] B. K. Carpenter, *Nat. Chem.* **2010**, *2*, 80–82.
- [264] D. B. Northrop, *Biochemistry* **1975**, *14*, 2644–2651.
- [265] D. W. Parkin, V. L. Schramm, *Biochemistry* **1987**, *26*, 913–20.
- [266] M. F. Hegazi, R. T. Borchardt, R. L. Schowen, *J. Am. Chem. Soc.* **1979**, *101*, 4359–4365.
- [267] C. H. Gray, J. K. Coward, K. B. Schowen, R. L. Schowen, *J. Am. Chem. Soc.* **1979**,

- 101, 4351–4358.
- [268] I. Williams, *J. Am. Chem. Soc.* **1984**, *106*, 7206–7212.
- [269] I. Mihel, J. O. Knipe, J. K. Coward, R. L. Schowen, *J. Am. Chem. Soc.* **1979**, *101*, 4349–4351.
- [270] V. Moliner, I. H. Williams, *J. Am. Chem. Soc.* **2000**, *122*, 10895–10902.
- [271] O. Matsson, a Dybala-Defratyka, M. Rostkowski, P. Paneth, K. C. Westaway, *J. Org. Chem.* **2005**, *70*, 4022–7.
- [272] G. D. Ruggiero, I. H. Williams, M. Roca, V. Moliner, I. Tuñón, *J. Am. Chem. Soc.* **2004**, *126*, 8634–8635.
- [273] J. Lameira, R. P. Bora, Z. T. Chu, A. Warshel, *Proteins Struct. Funct. Bioinforma.* **2015**, *83*, 318–330.
- [274] L. Michaelis, *Biochem. Z.* **1913**, *49*, 333–369.
- [275] A. S. Eustáquio, F. Pojer, J. P. Noel, B. S. Moore, *Nat. Chem. Biol.* **2008**, *4*, 69–74.
- [276] M. Thomsen, S. B. Vogensen, J. Buchardt, M. D. Burkart, R. P. Clausen, *Org. Biomol. Chem.* **2013**, *11*, 7606–10.
- [277] F. Wolfram, E. N. Kitova, H. Robinson, M. T. C. Walvoort, J. D. C. Codée, J. S. Klassen, P. L. Howell, *J. Biol. Chem.* **2014**, *289*, 6006–6019.
- [278] G. N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, *J. Mol. Biol.* **1963**, *7*, 95–9.
- [279] R. W. Hooft, C. Sander, G. Vriend, *Comput. Appl. Biosci.* **1997**, *13*, 425–430.
- [280] B. K. Ho, E. A. Coutsiias, C. Seok, K. A. Dill, *Protein Sci.* **2005**, *14*, 1011–8.
- [281] B. K. Ho, R. Brasseur, *BMC Struct. Biol.* **2005**, *5*, 14.
- [282] J. L. Porter, P. L. S. Boon, T. P. Murray, T. Huber, C. A. Collyer, D. L. Ollis, *ACS Chem. Biol.* **2013**, *10*, 611–621.
- [283] C. Wang, S. Leffler, D. H. Thompson, C. A. Hrycyna, *Biochem. Biophys. Res. Commun.* **2005**, *331*, 351–6.
- [284] D. Schulz, A. Rentmeister, *RNA Biol.* **2012**, *9*, 577–586.
- [285] J. Sylvestre, H. Chautard, F. Cedrone, M. Delcourt, *Org. Process Res. Dev.* **2006**, *10*, 562–571.
- [286] B. J. C. Law, M. R. Bennett, M. L. Thompson, C. Levy, S. A. Shepherd, D. Leys, J. Micklefield, *Angew. Chemie - Int. Ed.* **2016**, *55*, 2683–2687.
- [287] H. Schönherr, T. Cernak, *Angew. Chem. Int. Ed. Engl.* **2013**, 12256–12267.
- [288] A. Shokar, A. Au, S. H. An, E. Tong, G. Garza, J. Zayas, S. F. Wnuk, K. M. Land, *Bioorganic Med. Chem. Lett.* **2012**, *22*, 4203–4205.
- [289] V. Z. and A. E. P. F. Della Ragione, M. Porcelli, M. Carteni-Farina, *J. Biochem.* **1985**, *232*, 335–341.

- [290] H. . Song, E. . Park, W.-S. Choi, **2010**, DOI 10.2210/pdb3bsf/pdb.
- [291] T. T. Fung, E. B. Rimm, D. Spiegelman, N. Rifai, G. H. Tofler, W. C. Willett, F. B. Hu, **2001**.
- [292] K. Wang, H. Peng, B. Wang, *J. Cell. Biochem.* **2014**, *115*, 1007–22.
- [293] Y. Zhou, J. Yoon, *Chem. Soc. Rev.* **2012**, *41*, 52.
- [294] W. Wang, O. Rusin, X. Xu, K. K. Kim, J. O. Escobedo, S. O. Fakayode, K. A. Fletcher, M. Lowry, C. M. Schowalter, C. M. Lawrence, et al., *J. Am. Chem. Soc.* **2005**, *127*, 15949–58.
- [295] X. Chen, Y. Zhou, X. Peng, J. Yoon, *Chem. Soc. Rev.* **2010**, *39*, 2120–35.
- [296] Z. Guo, S. Nam, S. Park, J. Yoon, *Chem. Sci.* **2012**, *3*, 2760.
- [297] X. Yang, Y. Guo, R. M. Strongin, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 10690–3.
- [298] P. Wang, J. Liu, X. Lv, Y. Liu, Y. Zhao, W. Guo, *Org. Lett.* **2012**, *14*, 520–3.
- [299] P. Xie, G. Gao, J. Liu, Q. Jin, G. Yang, *J. Fluoresc.* **2015**, *25*, 1315–1321.
- [300] R. Zhao, J. Lind, G. Merenyi, T. E. Eriksen, *J. Am. Chem. Soc.* **1994**, *116*, 12010–12015.
- [301] O. Rusin, N. N. St Luce, R. A. Agbaria, J. O. Escobedo, S. Jiang, I. M. Warner, F. B. Dawan, K. Lian, R. M. Strongin, *J. Am. Chem. Soc.* **2004**, *126*, 438–9.
- [302] H. Chen, Q. Zhao, Y. Wu, F. Li, H. Yang, T. Yi, C. Huang, *Inorg. Chem.* **2007**, *46*, 11075–81.
- [303] F. Chem, *Annu. Rev. Microbiol.* **1995**, *49*, 523–555.
- [304] J. T. Chacko, K. Subramaniam, *Int. J. Environmetal Sci.* **2011**, *1*, 1250–1260.
- [305] C. Mercier, V. Chalansonnet, S. Orenga, C. Gilbert, *J. Appl. Microbiol.* **2013**, *115*, 1012–1022.
- [306] P. L. D. W. Byrant, D. R. McCalla, M. Leeksa, *Can. J. Microbiol.* **1981**, *27*, 81–86.
- [307] a J. Kirby, *Nat. News Views* **2001**, *8*, 737–739.
- [308] K. L. K. M D Miller, J Tanner, M Alpaugh, M J Benedik, *Nature* **1994**, *1*, 638–653.
- [309] E. P. Gillis, K. J. Eastman, M. D. Hill, D. J. Donnelly, N. A. Meanwell, *J. Med. Chem.* **2015**, *58*, 8315–59.
- [310] M. Thomsen, S. B. Vogensen, J. Buchardt, M. D. Burkart, R. P. Clausen, *Org. Biomol. Chem.* **2013**, *11*, 7606–10.
- [311] S. Singh, J. Zhang, T. D. Huber, M. Sunkara, K. Hurley, R. D. Goff, G. Wang, W. Zhang, C. Liu, J. Rohr, et al., *Angew. Chemie - Int. Ed.* **2014**, *53*, 3965–3969.
- [312] P. Perrone, F. Daverio, R. Valente, S. Rajyaguru, J. A. Martin, V. L  v  que, S. Le Pogam, I. Najera, K. Klumpp, D. B. Smith, et al., *J. Med. Chem.* **2007**, *50*, 5463–70.

- [313] V. Soloshonok, V. Kukhar, Y. Pustovit, V. Nazaretian, *Synlett* **1992**, 1992, 657–658.
- [314] H. Yasui, T. Yamamoto, E. Tokunaga, N. Shibata, *J. Fluor. Chem.* **2011**, 132, 186–189.
- [315] I. KIELTSCH, P. Eisenberger, A. Togni, *Angew. Chem. Int. Ed. Engl.* **2007**, 46, 754–7.
- [316] T. Umemoto, S. Ishihara, *J. Am. Chem. Soc.* **1993**, 115, 2156–2164.
- [317] W. Tyrra, D. Naumann, B. Hoge, Y. L. Yagupolskii, *J. Fluor. Chem.* **2003**, 119, 101–107.
- [318] D. Kong, Z. Jiang, S. Xin, Z. Bai, Y. Yuan, Z. Weng, *Tetrahedron* **2013**, 69, 6046–6050.
- [319] L. Zhai, Y. Li, J. Yin, K. Jin, R. Zhang, X. Fu, C. Duan, *Tetrahedron* **2013**, 69, 10262–10266.
- [320] Y. Fujiwara, J. A. Dixon, F. O. Hara, E. D. Funder, D. D. Dixon, R. A. Rodriguez, R. D. Baxter, N. Sach, M. R. Collins, Y. Ishihara, et al., *Nature* **2012**, 492, 5–10.
- [321] Y. Fujiwara, J. A. Dixon, R. A. Rodriguez, R. D. Baxter, D. D. Dixon, M. R. Collins, D. G. Blackmond, P. S. Baran, *J. Am. Chem. Soc.* **2012**, 134, 1494–1497.
- [322] D. J. Sam, H. E. Simmons, *J. Am. Chem. Soc.* **1974**, 96, 2252–2253.
- [323] U. Lange, A. Senning, *Chem. Ber.* **1991**, 124, 1879–1880.
- [324] V. K. Aggarwal, A. Thompson, R. V. H. Jones, *Tetrahedron Lett.* **1994**, 35, 8659–8660.
- [325] S. Oae, *Organic Chemistry of Sulfur*, Plenum Press, New York And London, **1977**.
- [326] J. Zhang, Y. G. Zheng, *ACS Chem. Biol.* **2016**, 11, 583–597.
- [327] R. Wang, W. Zheng, M. Luo, *Anal. Biochem.* **2014**, 450, 11–19.
- [328] A. S. Eustáquio, R. P. McGlinchey, Y. Liu, C. Hazzard, L. L. Beer, G. Florova, M. M. Alhamadsheh, A. Lechner, A. J. Kale, Y. Kobayashi, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106, 12295–300.
- [329] C. Vranken, A. Fin, P. Tufar, J. Hofkens, M. Burkart, Y. Tor, *Org. Biomol. Chem.* **2016**, 14, 6189–6192.
- [330] B. M. Trost, *Angew. Chemie Int. Ed. English* **1995**, 34, 259–281.
- [331] B. Bennett, E. Kimball, M. Gao, *Nat. Chem. Biol.* **2009**, 5, 593–599.
- [332] B. Y. S. Wijestjndera, D. D. Woods, *J. gen. Microbiol.* **1960**, 22, 229–241.
- [333] C. Neill, *Biology*, Company, The Benjamin Cummings Publishing, **1996**.
- [334] S. Klimasauskas, E. Weinhold, *Trends Biotechnol.* **2007**, 25, 99–104.
- [335] R. T. Borchardt, J. A. Huber, Y. S. Wu, *J. Med. Chem.* **1976**, 19, 1094–1099.
- [336] M. Luo, *ACS Chem. Biol.* **2012**, 7, 443–63.

- [337] X. Yang, Y. Hu, D. H. Yin, M. A. Turner, M. Wang, R. T. Borchardt, P. L. Howell, K. Kuczera, R. L. Schowen, *Biochemistry* **2003**, *42*, 1900–9.
- [338] V. Singh, W. Shi, S. C. Almo, G. B. Evans, R. H. Furneaux, P. C. Tyler, G. F. Painter, D. H. Lenz, S. Mee, R. Zheng, et al., *Biochemistry* **2006**, *45*, 12929–12941.
- [339] T. Zimmerman, G. Wolberg, G. Duncan, G. Elion, *Biochemistry* **1980**, *19*, 2252–2259.
- [340] Y. A. Chan, A. M. Podevels, B. M. Kevany, M. G. Thomas, *Nat. Prod. Rep.* **2009**, *26*, 90–114.
- [341] H. Seto, T. Sato, H. Yonehara, *J. Am. Chem. Soc.* **1973**, *95*, 8461–8462.
- [342] J. Iglesias, L. Sleno, D. A. Volmer, *Curr. Drug Metab.* **2012**, *13*, 1213–1225.
- [343] A. Chokkathukalam, D.-H. Kim, M. P. Barrett, R. Breitling, D. J. Creek, *Bioanalysis* **2014**, *6*, 511–24.
- [344] C. S. Elmore, R. A. Bragg, *Bioorganic Med. Chem. Lett.* **2015**, *25*, 167–171.
- [345] C. J. Unkefer, R. A. Martinez, *Drug Test. Anal.* **2012**, *4*, 303–307.
- [346] T. Rousu, O. Pelkonen, A. Tolonen, *Rapid Commun. Mass Spectrom.* **2009**, *23*, 843–55.
- [347] T. G. Gant, *J. Med. Chem.* **2014**, *57*, 3595–3611.
- [348] S. L. Harbeson, R. D. Tung, *Medchem news* **2014**, 8–22.
- [349] X. Zhang, X. Cheng, *Structure* **2003**, *11*, 509–520.
- [350] Y. Ichikawa, K. Hirata, M. Ohbayashi, M. Isobe, *Chemistry* **2004**, *10*, 3241–51.
- [351] D. Zhang, M. Zhang, Z. Liu, M. Yu, F. Li, T. Yi, C. Huang, *Tetrahedron Lett.* **2006**, *47*, 7093–7096.
- [352] M. Simokoriyama, *Bull. Chem. Soc. Jpn.* **1941**, *16*, 284–91.
- [353] G. Athanasellis, G. Melagraki, A. Afantitis, K. Makridima, O. Igglessi-Markopoulou, *Arkivoc* **2006**, *2006*, 28.
- [354] M. da Silva, C. M. S. Menezes, E. I. Ferreira, C. Q. F. Leite, D. N. Sato, C. C. Correia, C. P. Pimenta, K. C. A. Botelho, *Chem. Biol. Drug Des.* **2008**, *71*, 167–72.
- [355] A. R. Fersht, A. J. Kirby, *J. Am. Chem. Soc.* **1967**, *89*, 4857–4863.
- [356] J. J. Brophy, V. Diakiw, R. J. Goldsack, D. Nelson, J. S. Shannon, *Org. Mass Spectrom.* **1979**, *14*, 201–203.
- [357] M. Kawai, R. Nyfeler, J. M. Berman, M. Goodman, *J. Med. Chem.* **1982**, *25*, 397–402.
- [358] W. Jahn, *Chem. Ber.* **1965**, *98*, 1705–1708.
- [359] C. Moreau, T. Kirchberger, J. M. Swarbrick, S. J. Bartlett, R. Fliegert, T. Yorgan, A. Bauche, A. Harneit, A. H. Guse, B. V. L. Potter, *J. Med. Chem.* **2013**, *56*, 10079–102.

- [360] R. Saladino, P. Carlucci, M. C. Danti, C. Crestini, E. Mincione, *Tetrahedron* **2000**, *56*, 10031–10037.
- [361] K. Murakami, Masuo; Takahashi, Kozo; Murase, *5'-Deoxyadenosine Derivatives*, **1970**, JP 45030336.
- [362] K. L. Brown, X. Zou, *J. Am. Chem. Soc.* **1992**, *114*, 9643–9651.
- [363] M. Thomsen, S. B. Vogensen, J. Buchardt, M. D. Burkart, R. P. Clausen, *Org. Biomol. Chem.* **2013**, *11*, 7606–7610.
- [364] T. Bourdier, C. J. R. Fookes, T. Q. Pham, I. Greguric, A. Katsifis, *J. Label. Compd. Radiopharm.* **2008**, *51*, 369–373.
- [365] W. T. and B. G. Vonnrhein C., Flensburg C., Keller P., Sharff A., Smart O., Paciorek W., *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 293–302.
- [366] B. G. Vonnrhein C., Blanc E., Roversi P., *Methods in Molecular Biology*, **2007**.
- [367] Collaborative Computational Project, *Acta Cryst. D50* **1994**, *4*, 760–763.
- [368] S. W. G. and C. K. Emsley P., Lohkamp B., *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 486–501.
- [369] G. et al Bricogne, **2009**.
- [370] L. F. and V. A. A. Murshudov G. N., Skubák P., Lebedev A. A., Pannu N. S., Steiner R. A., Nicholls R. A., Winn M. D., *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 355–367.



# Appendix

## HPLC, LC-MS and UPLC Methods

### Method A: 5 minute UPLC method

<b>Instrument</b>	Agilent 1260 Infinity					
<b>Column</b>	Waters X-select CSH C18 column (XP), 2.5µm particle size, 2.1mm x 30mm, 130A					
<b>Mobile Phase</b>	<b>A</b>	0.05% v/v TFA in water				
	<b>B</b>	0.05% v/v TFA in acetonitrile				
<b>Flow Rate</b>	1 mL min <sup>-1</sup>					
<b>Gradient Profile</b>	<b>Time / min</b>	0	3.7	4.0	4.1	5.5
	<b>% A</b>	97	5	5	97	97
	<b>% B</b>	3	95	95	3	3
<b>Column Temperature</b>	40 °C					
<b>UV Detection</b>	220 nm or 320 nm (aminocoumarin substrates only)					
<b>Injection Volume</b>	0.5 µL					

### Method B: 8 minute HPLC method

<b>Instrument</b>	Hewlett Packard Agilent 1100			
<b>Column</b>	Phenomenex Luna C18(2), 50 x 2.0 mm, 3 µm			
<b>Mobile Phase</b>	<b>A</b>	0.05% v/v TFA in water		
	<b>B</b>	0.05% v/v TFA in acetonitrile		
<b>Flow Rate</b>	1 mL min <sup>-1</sup>			
<b>Gradient Profile</b>	<b>Time / min</b>	0	8	8.01
	<b>% A</b>	100	5	100
	<b>% B</b>	0	95	0
<b>Column Temperature</b>	40 °C			
<b>UV Detection</b>	220			
<b>Injection Volume</b>	0.5 µL			

### Method C: 10 minute HPLC method

<b>Instrument</b>	Hewlett Packard Series 1100					
<b>Column</b>	Phenomenex Luna C18(2), 50 x 2.0 mm, 3 m.					
<b>Mobile Phase</b>	<b>A</b>	0.05% v/v TFA in water				
	<b>B</b>	0.05% v/v TFA in acetonitrile				
<b>Flow Rate</b>	1 mL min <sup>-1</sup>					
<b>Gradient Profile</b>	<b>Time / min</b>	0	2	6	6.10	10
	<b>% A</b>	100	100	40	100	100
	<b>% B</b>	0	0	60	0	0
<b>Column Temperature</b>	40 °C					
<b>UV Detection</b>	254 nm					
<b>Injection Volume</b>	2 µL					

**Method D: UPLC 5 minute method**

<b>Instrument</b>	Agilent 1290 Infinity II					
<b>Column</b>	X Select CSH C18, 30 x 2.1 mm, 2.5 µM particle size.					
<b>Mobile Phase</b>	<b>A</b>	0.05% v/v TFA in water				
	<b>B</b>	0.05% v/v TFA in acetonitrile				
<b>Flow Rate</b>	1 mL min <sup>-1</sup>					
<b>Gradient Profile</b>	<b>Time / min</b>	2.5	2.7	2.71	5.0	
	<b>% A</b>	5	5	0	0	
	<b>% B</b>	95	95	100	100	
<b>Column Temperature</b>	60 °C					
<b>UV Detection</b>	220 nm					
<b>Injection Volume</b>	0.5 µL					

**LC-MS****Method A: 5 minute high pH**

<b>LC Instrument</b>	Agilent 1100 series HPLC				
<b>Column</b>	Waters Xbridge C18 (50 mm × 4.6 mm × 3.0 μm)				
<b>Mobile Phase</b>	<b>A</b>	Ammonium bicarbonate (10 mM aqueous solution, adjusted to pH 10 with ammonia solution)			
	<b>B</b>	Acetonitrile			
<b>Flow Rate</b>	3 mL min <sup>-1</sup>				
<b>Gradient Profile</b>	<b>Time / min</b>	0	0.1	4	5
	<b>% A</b>	99	99	3	3
	<b>% B</b>	1	1	97	97
<b>Column Temperature</b>	40 °C				
<b>UV Detection</b>	220 nm				
<b>Injection Volume</b>	1 μL				
<b>MS Instrument</b>	Waters ZQ				
<b>Ionisation Mode</b>	Positive and/or negative electrospray.				
<b>Scan Range</b>	100 to 1000 AMU positive, 120-1000 AMU negative.				
<b>Scan Time</b>	0.5 seconds				
<b>Interscan Delay</b>	0.05 seconds				

**Method B: 8 minute low pH**

<b>LC Instrument</b>	Agilent 1100 series HPLC					
<b>Column</b>	Phenomenex Luna C18 column (50 mm x 2.1 mm i.d. 3 µm packing diameter)					
<b>Mobile Phase</b>	<b>A</b>	Water + 0.05% v/v trifluoroacetic acid				
	<b>B</b>	Acetonitrile + 0.05% v/v trifluoroacetic acid				
<b>Flow Rate</b>	1 mL min <sup>-1</sup>					
<b>Gradient Profile</b>	<b>Time / min</b>	0	8.0	8.5	8.6 3	10
	<b>% A</b>	100	5	5	100	100
	<b>% B</b>	0	95	95	0	0
<b>Column Temperature</b>	40 °C					
<b>UV Detection</b>	220 nm					
<b>Injection Volume</b>	1 µL					
<b>MS Instrument</b>	Waters ZQ					
<b>Ionisation Mode</b>	Positive electrospray.					
<b>Scan Range</b>	100 to 1000 AMU					
<b>Scan Time</b>	1 second					
<b>Interscan Delay</b>	0.05 seconds					

**DNA sequences****6xHis-NovO (*Streptomyces spheroides*)**

ATGAAAATCG AAGCAATCAC GGGTTCCGAA GCAGAAGCAT TCCACCGCAT  
 GGGCTCGCAA GCATCGCACC GCTATGACGA ATTTGTTGAT CTGCTGGTTG  
 GCGCCGGTAT TGCAGATGGC CAGACCGTGG TTGACCTGTG CTGTGGCAGC  
 GGTGAACTGG AAGTGATCCT GAGCTCTCGT TTTCCGTCGC TGAACCTGGT  
 TGGTGTTCGAT CTGAGCGAAG ACATGGTTCG TATTGCGCGC GAATATGCGG  
 CCGAACAGGG CAAAGCGCTG GAATTTTCGTC ATGGCGATGC CCAACTGCTG  
 GCAGGCATGG AAGATCTGGC TGGTAAAGCG GACCTGGTTCG TGAGTCGTAA  
 TGCCTTTCAC CGTCTGACCC GCCTGCCGGC AGCTTTCGAT ACGATGCTGC  
 GCCTGGCTAA ACCGGGCGGT GCAGTTCTGA ACTGCAGTTT TATTCATCCG  
 TCCGATTTTG ACGAATCAGG CTTCCGTGCA TGGGTTACGT TCCTGAATCA

GCGCCCGTGG GATTCTGAAA TGCAAATCGT CTGGGCCCTG GCACATCACT  
 ATGCGCCGCG TCTGGATGAC TACCGTGAAG CTCTGGCACA GGCAGCCCGT  
 GAAACCCCGG TCTCCGAACA GCGCGTGTGG ATCGATGACC AAGGCTACGG  
 TGTGCCGACG GTTAAATGTT TTGCGCGTCCG CGCAGCTGCG CATCACCATC  
 ACCATCACTA A

**6xHis-CouO (*Streptomyces rishiriensis*)**

ATGAAAATCG AACCGATCAC CGGCTCAGAA GCAGAAGCAT TCCACCGTAT  
 GGGCTCACGC GCCTTTGAAC GCTATAACGA ATTTGTGGAT CTGCTGGTTG  
 GCGCGGGTAT TGCCGATGGC CAGACCGTGG TTGACCTGTG CTGTGGCAGC  
 GGTGAACTGG AAATTATCCT GACGAGCCGT TTTCCGTCTC TGAACCTGGT  
 TGGTGTGCGAT CTGTCTGAAG ACATGGTTCG TATCGCGCGC GATTATGCGG  
 CCGAACAGGG CAAAGAAGCTG GAATTTTCGTC ATGGCGACGC ACAATCACCG  
 GCTGGCATGG AAGATCTGCT GGGTAAAGCA GACCTGGTCG TGTCGCGTCA  
 TGCTTTTCAC CGTCTGACCC GTCTGCCGGC AGGTTTCGAT ACGATGCTGC  
 GCCTGGTTAA ACCGGGCGGT GCCATTCTGA ATGTCAGTTT TCTGCATCTG  
 TCCGATTTTG ACGAACC GGG TTTCCGTACC TGGGTGCGCT TCCTGAAAGA  
 ACGTCCGTGG GATGCAGAAA TGCAGGTTGC TTGGGCACTG GCACACTATT  
 ACGCACCGCG TCTGCAGGAT TATCGCGACG CACTGGCTCA AGCAGCTGAT  
 GAAACCCCGG TCTCCGAACA GCGCATCTGG GTGGATGACC AAGGCTACGG  
 TGTGGCGACG GTTAAATGCT TTGCACGTCG CGCAGCAGCA CATCACCATC  
 ACCATCACTA A

**6xHis-SalL (*Salinospora tropica*)**

ATGAAAAAAG GTCACCACCA CCACCACCAC TCGCAGCACA ATCTGATCGC  
 ATTCCTGTGCG GATGTTGGCT CGGCGGACGA AGCACACGCG CTGTGCAAAG  
 GCGTTATGTA TGGTGTGCGC CCGGCGGCCA CCATTGTCGA TATCACGCAT  
 GACGTCGCCC CGTTTGATGT GCGTGAAGGC GCTCTGTTCC TGGCGGACGT  
 GCCGCATAGC TTTCCGGCTC ACACCGTTAT TTGTGCGTAT GTGTACCCGG  
 AAACCGGCAC GGCCACCCAC ACGATCGCAG TTCGCAACGA AAAGGGTCAG  
 CTGCTGGTGC GCCCGAACAA TGGTCTGCTG TCATTCGCGC TGGATGCCTC  
 GCCGGCAGTG GAATGCCATG AAGTTCTGTC TCCGGACGTC ATGAATCAAC  
 CGGTGACCCC GACGTGGTAT GGTAAGATA TTGTTGCAGC TTGTGCAGCA  
 CACCTGGCAG CTGGTACCGA TCTGGCAGCA GTCGGTCCGC GTATTGACCC  
 GAAACAGATC GTGCGCCTGC CGTACGCAAG CGCTTCTGAA GTTGAAGGCG

GTATTCGTGG CGAAGTGGTT CGTATCGATC GCGCATTCCG TAACGTTTGG  
 ACCAATATTC CGACGCATCT GATCGGCAGT ATGCTGCAAG ATGGTGAACG  
 CCTGGAAGTG AAAATCGAAG CGCTGTCCGA CACCGTTCTG GAACTGCCGT  
 TTTGCAAAAC GTTCGGCGAA GTGGATGAGG GTCAGCCGCT GCTGTACCTG  
 AACAGCCGTG GCCGCCTGGC ACTGGGTCTG AACCAATCTA ACTTCATCGA  
 AAAATGGCCG GTCGTGCCGG GCGATAGTAT CACCGTGTCC CCGCGTGTTC  
 CGGACAGCAA TCTGGGCCCG GTGCTGGGTT AA

**6xHis-Tev-SAHH (*Homo sapiens*)**

ATGGGTCATC ATCATCATCA TCATGAAAAT CTGTACTTCC AAGGCATGAG  
 CGATAAACTG CCGTATAAGG TGGCGGACAT CGGTCTGGCG GCGTGGGGTC  
 GTAAGGCGCT GGACATTGCG GAGAACGAAA TGCCGGGTCT GATGCGTATG  
 CGTGAGCGTT ATAGCGCGAG CAAGCCGCTG AAAGGTGCGC GTATCGCGGG  
 CTGCCTGCAC ATGACCGTGG AGACCGCGGT TCTGATTGAA ACCCTGGTGA  
 CCCTGGGCGC GGAAGTTCAG TGGAGCAGCT GCAACATCTT TAGCACCCAA  
 GATCATGCGG CGGCGGCGAT TGC GAAGGCG GGTATTCCGG TGTACGCGTG  
 GAAAGGCGAG ACCGATGAGG AATACCTGTG GTGCATTGAA CAGACCCTGT  
 ATTTCAAGGA CGGTCCGCTG AACATGATCC TGGACGATGG TGGCGATCTG  
 ACCAACCTGA TTCACACCAA ATGCCCGCAA CTGCTGCCGG GTATCCGTGG  
 CATTAGCGAG GAAACCACCA CCGGTGTGCA CAACCTGTAC AAGATGATGG  
 CGAACGGCAT CCTGAAAGTT CCGGCGATTA ACGTGAACGA CAGCGTTACC  
 AAGAGCAAAT TTGATAACCT GTATGGTTGC CGTGAGAGCC TGATCGACGG  
 CATTAAAGCGT GCGACCGATG TGATGATCGC GGGTAAAGTT GCGGTGGTTG  
 CGGGTTATGG TGATGTGGGC AAGGGTTGCG CGCAGGCGCT GCGTGGTTTC  
 GGTGCGCGTG TTATCATTAC CGAAATCGAT CCGATTAACG CGCTGCAGGC  
 GGCGATGGAG GGTTATGAAG TGACCACGAT GGACGAGGCG TGCCAAGAAG  
 GCAACATCTT CGTTACCACC ACCGGTTGCA TTGATATCAT TCTGGGCCGT  
 CACTTTGAGC AGATGAAGGA CGATGCGATC GTTTGCAACA TTGGTCACTT  
 CGACGTGGAA ATCGATGTTA AATGGCTGAA CGAGAACGCG GTGGAAAAGG  
 TTAACATTAA ACCGCAAGTG GACCGTTATC GTCTGAAGAA CGGTGCTCGT  
 ATCATTCTGC TGGCGGAGGG CCGTCTGGTT AACCTGGGTT GCGCGATGGG  
 CCACCCGAGC TTCGTGATGA GCAACAGCTT TACCAACCAG GTTATGGCGC  
 AAATCGAACT GTGGACCCAC CCGGACAAAT ACCCGGTGGG TGTTCACTTT  
 CTGCCGAAGA AACTGGATGA AGCGGTGGCG GAAGCGCACC TGGGCAAGCT

GAACGTAAAG CTGACCAAAC TGACCGAGAA ACAAGCGCAA TACCTGGGCA  
 TGAGCTGCGA TGGCCCGTTC AAGCCGGACC ACTACCGTTA CTAG

**6xHis-Tev-LuxS (*Bacillus subtilis*)**

ATGCCGAGCG TTGAGAGCTT CGAGCTGGAT CATAATGCGG TTGTTGCGCC  
 GTATGTTCGT CACTGCGGTG TTCATAAAGT TGGTACCGAT GGTGTGGTTA  
 ACAAGTTCGA CATCCGTTTT TGCCAGCCGA ACAAGCAAGC GATGAAACCG  
 GATACCATCC ACACCCTGGA ACACCTGCTG GCGTTCACCA TTCGTAGCCA  
 CGCGGAGAAA TACGACCACT TTGATATCAT TGACATCAGC CCGATGGGTT  
 GCCAGACCGG CTA CTACTATCTG GTGGTTAGCG GCGAACCAGC CAGCGCGGAG  
 ATTGTGGATC TGCTGGAAGA CACCATGAAA GAGGCGGTTG AAATCACCGA  
 GATTCCGGCG GCGAACGAAA AGCAGTGCGG TCAAGCGAAA CTGCACGACC  
 TGGAGGGCGC GAAACGTCTG ATGCGTTTCT GGCTGAGCCA GGATAAAGAA  
 GAACTGCTGA AGGTTTTCGG CGAGAATCTG TATTTCCAAG GCCATCATCA  
 TCATCATCAT TAA

**6xHis-Tev-MTAN (*E. coli*)**

ATGGGTCATC ATCATCATCA TCATGAAAAT CTGTACTTCC AAGGCATGAA  
 AATCGGCATT ATCGGTGCGA TGGAGGAAGA GGTTACCCTG CTGCGCGACA  
 AAATTGAGAA TCGTCAGACC ATTAGCCTGG GCGGTTGCGA GATCTACACC  
 GGTCAGCTGA ACGGCACCGA AGTTGCGCTG CTGAAGAGCG GTATTGGCAA  
 AGTTGCGGCG GCGCTGGGTG CGACCCTGCT GCTGGAGCAC TGCAAGCCGG  
 ACGTTATCAT TAACACCGGT AGCGCGGGTG GCCTGGCGCC GACCCTGAAA  
 GTGGGTGACA TCGTGGTTAG CGATGAGGCG CGTTATCACG ACGCGGATGT  
 TACCGCGTTC GGTTACGAAT ATGGTCAACT GCCGGGTTGC CCGGCGGGTT  
 TTAAGGCGGA CGATAAACTG ATTGCGGCGG CGGAGGCGTG CATTGCGGAA  
 CTGAACCTGA ACGCGGTTTCG TGGTCTGATC GTGAGCGGCG ACGCGTTCAT  
 TAACGGTAGC GTTGGCCTGG CGAAGATCCG TCACAACCTT CCGCAAGCGA  
 TTGCGGTGGA GATGGAAGCG ACCGCGATCG CGCACGTTTG CCACAACCTC  
 AACGTGCCGT TTGTGGTTGT GCGTGCGATT AGCGACGTGG CGGATCAGCA  
 AAGCCACCTG AGCTTCGATG AATTTCTGGC GGTGGCGGCG AAGCAGAGCA  
 GCCTGATGGT GGAGAGCCTG GTTCAAAAAC TGGCGCACGG TTAA

**SDM Primers:**

**F14M fwd** GGTTCGAAGCAGAAGCAATGCACCGCATGGGCTCG  
**F14M rev** CGAGCCCATGCGGTGCATTGCTTCTGCTTCGGAACC



<b>F14L fwd</b>	GGTTCCGAAGCAGAAGCACTCCACCGCATGGGCTCG
<b>F14L rev</b>	CGAGCCCATGCGGTGGAGTGCTTCTGCTTCGGAACC
<b>R24M fwd</b>	GGCTCGCAAGCATCGCACATGTATGACGAATTTGTT
<b>R24M rev</b>	AACAAATTCGTCATACATGTGCGATGCTTGCGAGCC
<b>R24F fwd</b>	GGCTCGCAAGCATCGCACUTCTATGACGAATTTGTT
<b>R24F rev</b>	AACAAATTCGTCATAGAAGTGCGATGCTTGCGAGCC
<b>H120F fwd</b>	GTGAGTCGTAATGCCTTTTTCCGTCTGACCCGCCTG
<b>H120F rev</b>	CAGGCGGGTCAGACGGAAAAAGGCATTACGACTCAC
<b>H120N fwd</b>	GTGAGTCGTAATGCCTTTAACCGTCTGACCCGCCTG
<b>H120N rev</b>	CAGGCGGGTCAGACGGTTAAAGGCATTACGACTCAC
<b>H120A fwd</b>	GTGAGTCGTAATGCCTTTGCCCGTCTGACCCGCCTG
<b>H120A rev</b>	CAGGCGGGTCAGACGGGCAAAGGCATTACGACTCAC
<b>N117A fwd</b>	CTGGTCGTGAGTCGTGCTGCCTTTCACCGTCTGACC
<b>N117A rev</b>	GGTCAGACGGTGAAAGGCAGCACGACTCACGACCAG
<b>R121Q fwd</b>	CGTAATGCCTTTCACCAGCTGACCCGCCTGCCG
<b>R121Q rev</b>	CGGCAGGCGGGTCAGCTGGTGAAAGGCATTACG
<b>R121L fwd</b>	CGTAATGCCTTTCACCTTCTGACCCGCCTGCCG
<b>R121L rev</b>	CGGCAGGCGGGTCAGAAGGTGAAAGGCATTACG
<b>Y216F fwd</b>	GTGTGGATCGATGACCAAGGCTTCGGTGTGCCGACGGTT
<b>Y216F rev</b>	AACCGTCGGCACACCGAAGCCTTGGTCATCGATCCACAC
<b>H120K fwd</b>	CGTAATGCCTTTAAACGTCTGACCCGCCTGC
<b>H120K rev</b>	GCAGGCGGGTCAGACGTTTAAAGGCATTACG
<b>R121K fwd</b>	CGTAATGCCTTTCACAAACTGACCCGCCTGC
<b>R121K rev</b>	GCAGGCGGGTCAGTTTGTGAAAGGCATTACG

#### Gene sequencing primers (5' to 3')

<b>NovO 1 fwd</b>	CAATCACGGGTTCCGAAG
<b>NovO 1 rev</b>	CTTCGGAACCCGTGATTG
<b>NovO 2 fwd</b>	CGGTATTGCAGATGGCC
<b>NovO 2 rev</b>	GGCCATCTGCAATACCG
<b>NovO 3 fwd</b>	GTCGATCTGAGCGAAGAC
<b>NovO 3 rev</b>	GTCTTCGCTCAGATCGAC
<b>NovO 4 fwd</b>	AGGCATGGAAGATCTGG
<b>NovO 4 rev</b>	CCAGATCTTCCATGCCT
<b>NovO 5 fwd</b>	GCTGCGCCTGGCTAAAC
<b>NovO 5 rev</b>	GTTTAGCCAGGCGCAGC
<b>NovO 6 fwd</b>	CATGGGTTACGTTCTTG
<b>NovO 6 rev</b>	CAGGAACGTAACCCATG
<b>NovO 7 fwd</b>	CTATGCGCCGCGTCTGG
<b>NovO 7 rev</b>	CCAGACGCGGCGCATAG
<b>NovO 8 fwd</b>	GTGCCGACGGTTAAATG
<b>NovO 8 rev</b>	CATTTAACCGTCGGCAC

**CouO 1 fwd** CGAACCGATCACCGGCTC  
**CouO 1 rev** GAGCCGGTGATCGGTTCG  
**CouO 2 fwd** GATGGCCAGACCGTGGTTG  
**CouO 2 rev** CAACCACGGTCTGGCCATC  
**CouO 3 fwd** GACATGGTTTCGTATCGC  
**CouO 3 rev** GCGATACGAACCATGTC  
**CouO 4 fwd** CTGCTGGGTAAAGCAGAC  
**CouO 4 rev** GTCTGCTTTACCCAGCAG  
**CouO 5 fwd** GTCAGTTTTCTGCATCTG  
**CouO 5 rev** CAGATGCAGAAAAGTGC  
**CouO 6 fwd** GCTTGGGCACTGGCACAC  
**CouO 6 rev** GTGTGCCAGTGCCCAAGC  
**CouO 7 fwd** GGTCTCCGAACAGCGCATC  
**CouO 7 rev** GATGCGCTGTTCGGAGACC  
**CouO 8 fwd** GCACATCACCATCACCATC  
**CouO 8 rev** GATGGTGATGGTGATGTGC

**SDM primers for CASTing libraries:**

**F14/M17 fwd** GAAGCAGAAGCANNNCACCGCENNNGGCTCGCAAGC  
**F14/M17 rev** GCTTGCAGCCNNNGCGGTGNNNTGCTTCTGCTTC  
**R116/N117 fwd** CTGGTCGTGAGTNNNNNNGCCTTTCACCGTC  
**R116/N117 rev** GACGGTGAAAGGCNNNNNNNACTCACGACCAG  
**H120/R121 fwd** CGTAATGCCTTTNNNNNNCTGACCCGCCTG  
**H120/R121 rev** CAGGCGGGTCAGNNNNNNAAAGGCATTACG  
**M174/V177 fwd** GTGGGATTCTGAANNNCAAATCENNNTGGGCCCTGGCAC  
**M174/V177 rev** GTGCCAGGGCCANNNGATTTGNNNTTCAGAATCCCAC  
**Y216/V218 fwd** GATGACCAAGGCNNNGGTNNNCCGACGGTTAAATG  
**Y216/V218 rev** CATTTAACCGTCGGNNNACCNNNGCCTTGGTCATC

Section of sequencing results for production of single point mutants

F14M

970 980 990 1000 1010 1020  
 AGAAGCAATGCACCGCATGGGCTCGCAAGCATCGCACCCGCTATGACGAATTTGTTGATCTGCTG

F14M\_14\_NovO\_rev\_2\_NovO\_rev\_2\_B2(1>927) ← AGAAGCAATGCACCGCATGGGCTCGCAAGCATCGCACCCGCTATGACGA  
 F14M\_14\_NovO\_rev\_3\_NovO\_rev\_3\_C2(1>945) ← AGAAGCAATGCACCGCATGGGCTCGCAAGCATCGCACCCGCTATGACGAATTTGTTGATCTGCTG  
 F14M\_14\_NovO\_rev\_4\_NovO\_rev\_4\_D2(3>933) ← AGAAGCAATGCACCGCATGGGCTCGCAAGCATCGCACCCGCTATGACGAATTTGTTGATCTGCTG  
 F14M\_14\_NovO\_rev\_5\_NovO\_rev\_5\_E2(3>935) ← AGAAGCAATGCACCGCATGGGCTCGCAAGCATCGCACCCGCTATGACGAATTTGTTGATCTGCTG  
 F14M\_14\_NovO\_rev\_6\_NovO\_rev\_6\_F2(1>932) ← AGAAGCAATGCACCGCATGGGCTCGCAAGCATCGCACCCGCTATGACGAATTTGTTGATCTGCTG  
 F14M\_14\_NovO\_rev\_7\_NovO\_rev\_7\_G(36>933) ← AGAAGCAATGCACCGCATGGGCTCGCAAGCATCGCACCCGCTATGACGAATTTGTTGATCTGCTG  
 F14M\_14\_NovO\_rev\_8\_NovO\_rev\_8\_H2(1>921) ← AGAAGCAATGCACCGCATGGGCTCGCAAGCATCGCACCCGCTATGACGAATTTGTTGATCTGCTG  
 Novo 6H C term codon opti.seq(1>711) → AGAAGCAATGCACCGCATGGGCTCGCAAGCATCGCACCCGCTATGACGAATTTGTTGATCTGCTG  
 F14M\_14\_NovO\_fwd\_1\_NovO\_fwd\_1\_A1(1>917) → GCGAGGGC-NNC-AGCATCGCACCCGCTATGACGAATTTGTTGATCTGCTG

F14L

1000 1010 1020 1030 1040 1050 1060  
 GAGATATACATATGAAAATCGAAGCAATCACGGGTTCCGAAGCAGAAGCACTCCACCGCATGGG

R14F\_11\_NovO\_rev\_1\_NovO\_rev\_1\_A(1>1000) ← GAGA-ACA  
 R14F\_11\_NovO\_rev\_2\_NovO\_rev\_2\_B(1>1003) ← GAGATATACATATGAAAATCGAAGCAATCACGGGTTCCGAAGCAGAAGCACTCCACCGCATGGG  
 R14F\_11\_NovO\_rev\_3\_NovO\_rev\_3\_C2(1>956) ← GAGATATACATATGAAAATCGAAGCAATCACGGGTTCCGAAGCAGAAGCACTCCACCGCATGGG  
 R14F\_11\_NovO\_rev\_4\_NovO\_rev\_4\_D2(1>979) ← GAGATATACATATGAAAATCGAAGCAATCACGGGTTCCGAAGCAGAAGCACTCCACCGCATGGG  
 R14F\_11\_NovO\_rev\_5\_NovO\_rev\_5\_E2(1>959) ← GAGATATACATATGAAAATCGAAGCAATCACGGGTTCCGAAGCAGAAGCACTCCACCGCATGGG  
 R14F\_11\_NovO\_rev\_6\_NovO\_rev\_6\_F2(1>992) ← GAGATATACATATGAAAATCGAAGCAATCACGGGTTCCGAAGCAGAAGCACTCCACCGCATGGG  
 R14F\_11\_NovO\_rev\_7\_NovO\_rev\_7\_G(1>1003) ← GAGATATACATATGAAAATCGAAGCAATCACGGGTTCCGAAGCAGAAGCACTCCACCGCATGGG  
 R14F\_11\_NovO\_rev\_8\_NovO\_rev\_8\_H2(8>991) ← GAGATATACATATGAAAATCGAAGCAATCACGGGTTCCGAAGCAGAAGCACTCCACCGCATGGG  
 Novo 6H C term codon opti.seq(1>711) → ATGAAAATCGAAGCAATCACGGGTTCCGAAGCAGAAGCACTCCACCGCATGGG  
 R14F\_11\_NovO\_fwd\_1\_NovO\_fwd\_1\_A1(1>988) → A-GGG

R24F

1040 1050 1060 1070 1080 1090 1100  
 AGAAGCATTCCACCGCATGGGCTCGCAAGCATCGCACTTCTATGACGAATTTGTTGATCTGCTG

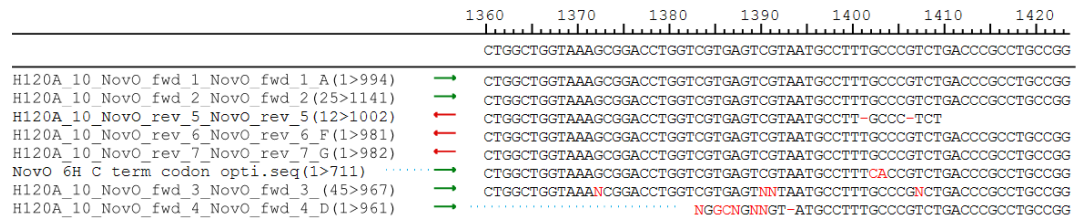
R24L\_10\_NovO\_rev\_2\_NovO\_rev\_2\_B(1>1022) ← AGAAGCATTCCACCGCATGGGCTCGCAAGCATCGCACTTCTATGACGA-~~TT~~-GT  
 R24L\_10\_NovO\_rev\_3\_NovO\_rev\_3\_C(1>1071) ← AGAAGCATTCCACCGCATGGGCTCGCAAGCATCGCACTTCTATGACGAATTTGTTGATCTGCTG  
 R24L\_10\_NovO\_rev\_5\_NovO\_rev\_5\_E(1>1053) ← AGAAGCATTCCACCGCATGGGCTCGCAAGCATCGCACTTCTATGACGAATTTGTTGATCTGCTG  
 R24L\_10\_NovO\_rev\_6\_NovO\_rev\_6\_F8(1>990) ← AGAAGCATTCCACCGCATGGGCTCGCAAGCATCGCACTTCTATGACGAATTTGTTGATCTGCTG  
 R24L\_10\_NovO\_fwd\_1\_NovO\_fwd\_1\_A(1>1004) → GGGC-N-GC-AGCATCGCACTTCTATGACGAATTTGTTGATCTGCTG  
 R24L\_10\_NovO\_rev\_4\_NovO\_rev\_4\_D8(1>988) ← AGAAGCATTCCACCGCATGGGCTCGCAAGCATCGCACTTCTATGACGAATTTGTTGATCTGCTG  
 R24L\_10\_NovO\_rev\_7\_NovO\_rev\_7\_G(21>948) ← AGAAGCATTCCACCGCATGGGCTCGCAAGCATCGCACTTCTATGACGAATTTGTTGATCTGCTG  
 R24L\_10\_NovO\_rev\_8\_NovO\_rev\_8\_H8(3>944) ← AGAAGCATTCCACCGCATGGGCTCGCAAGCATCGCACTTCTATGACGAATTTGTTGATCTGCTG  
 Novo 6H C term codon opti.seq(1>711) → AGAAGCATTCCACCGCATGGGCTCGCAAGCATCGCACTTCTATGACGAATTTGTTGATCTGCTG

H120N

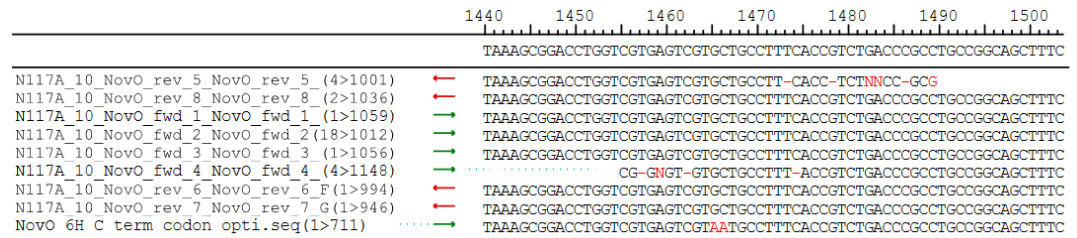
1350 1360 1370 1380 1390 1400 1410  
 -CCTGGTCGTGAGTCGTAATGCCTTTAACCGTCTGACCCGCTGCGCGCAGCTTTCGATAC-GA

H120N\_10\_NovO\_rev\_5\_NovO\_rev\_5\_E(1>997) ← ACCTGGTCGTGAGTCGTAATGCCTTTAACCGTCTGACCCGCTGCGCGCAGCTTTCGATAC-GA  
 H120N\_10\_NovO\_fwd\_2\_NovO\_fwd\_2(39>1035) → -CCTGGTCGTGAGTCGTAATGCCTTTAACCGTCTGACCCGCTGCGCGCAGCTTTCGATAC-GA  
 H120N\_10\_NovO\_rev\_6\_NovO\_rev\_6\_F(1>977) ← -CCTGGTCGTGAGTCGTAATGCCTTTAACCGTCTGACCCGCTGCGCGCAGCTTTCGATAC-GA  
 H120N\_10\_NovO\_rev\_7\_NovO\_rev\_7\_G(1>949) ← -CCTGGTCGTGAGTCGTAATGCCTTTAACCGTCTGACCCGCTGCGCGCAGCTTTCGATAC-GA  
 H120N\_10\_NovO\_rev\_8\_NovO\_rev\_8\_H(1>952) ← -CCTGGTCGTGAGTCGTAATGCCTTTAACCGTCTGACCCGCTGCGCGCAGCTTTCGATAC-GA  
 Novo 6H C term codon opti.seq(1>711) → -CCTGGTCGTGAGTCGTAATGCCTTTAACCGTCTGACCCGCTGCGCGCAGCTTTCGATAC-GA  
 H120N\_10\_NovO\_fwd\_1\_NovO\_fwd\_1(16>1003) → -CCTGGTCGTGAGTCGTAATGCCTTTAACCGTCTGACCCGCTGCGCGCAGCTTTCGATAC-GA  
 H120N\_10\_NovO\_fwd\_3\_NovO\_fwd\_3(43>936) → -CCTGGTCGTGAGTCGTAATGCCTTTAACCGNCTGACCCGCTGCGCGCAGCTTTCNATACTGA  
 H120N\_10\_NovO\_fwd\_4\_NovO\_fwd\_4\_D(3>965) → GNGGNNNTA-TGCCTTT-ACCCTGACCCGCTGCGCGCAGCTTTCGATAC-GA

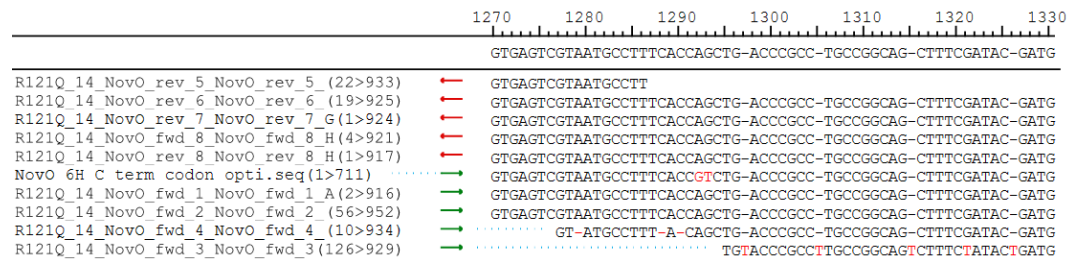
### H120A



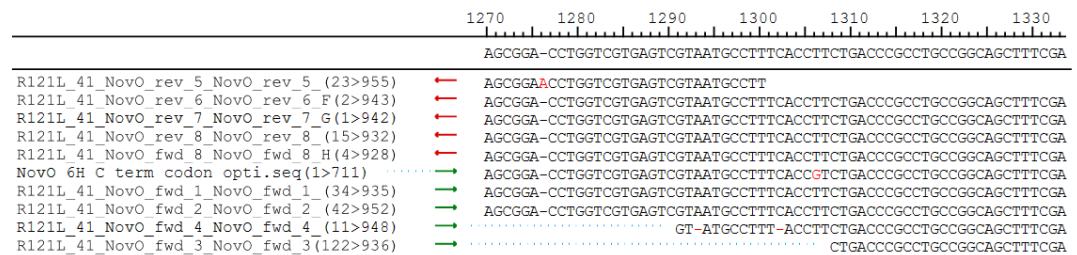
### N117A



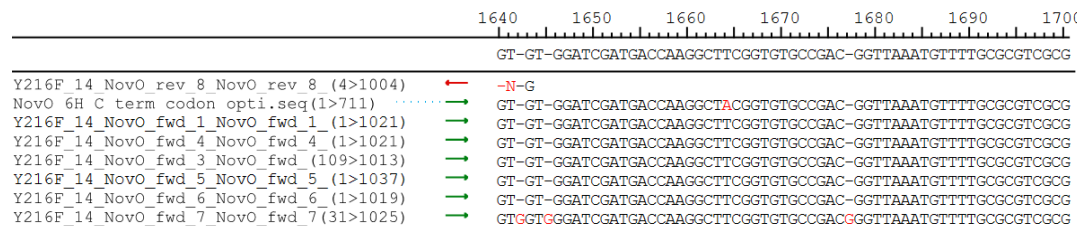
### R121Q



### R121L



### Y216F



**H120K**

		1280	1290	1300	1310	1320	1330	1340
		AGCGGA-CCTGGTC-GTGAGTCGTAATGCCTTAAACGCTGACCCGCCTGCCGGCAGCTTTTCG						
NovO_H120K_4_NovO_rev5_E2_w11-81(1>901)	←	AGCGGA	ACCTGGTC-GTGAGTCGTAATGCCTT-AAAC-TCT-NCCCG-C-GCC-GC					
NovO_H120K_4_NovO_rev6_F2_w11-8(15>940)	←	AGCGGA-CCTGGTC-GTGAGTCGTAATGCCTTAAACGCTGACCCGCCTGCCGGCAGCTTTTCG						
NovO_H120K_4_NovO_rev7_G2_w11-8(25>931)	←	AGCGGA-CCTGGTC-GTGAGTCGTAATGCCTTAAACGCTGACCCGCCTGCCGGCAGCTTTTCG						
NovO_H120K_4_NovO_fwd8_H1_w11-81(5>914)	←	AGCGGA-CCTGGTC-GTGAGTCGTAATGCCTTAAACGCTGACCCGCCTGCCGGCAGCTTTTCG						
NovO_6H_C term codon opti.seq(1>711)	→	AGCGGA-CCTGGTC-GTGAGTCGTAATGCCTTAAACGCTGACCCGCCTGCCGGCAGCTTTTCG						
NovO_H120K_4_NovO_fwd1_A1_w11-81(1>865)	→	AGCGGA-CCTGGTC-GTGAGTCGTAATGCCTTAAACGCTGACCCGCCTGCCGGCAGCTTTTCG						
NovO_H120K_4_NovO_fwd2_B1_w11-8(39>940)	→	AGCGGA-CCTGGTC-GTGAGTCGTAATGCCTTAAACGCTGACCCGCCTGCCGGCAGCTTTTCG						
NovO_H120K_4_NovO_fwd3_C1_w11-8(32>910)	→	AGCGGA-CCTGGTC-GTGAGTCNTAATGCCTTAAACGCTGACCCGCCTGCCGGCAGCTTTTCG						
NovO_H120K_4_NovO_fwd4_D1_w11-81(1>893)	→	CNCGG-GNNGT-A-GCCTTT--ACGCTGACCCGCCTGCCGGCAGCTTTTCG						

**R121K**

		1250	1260	1270	1280	1290	1300
		T-CGTGAGTCGTAATGCCTTTCACAAACTGACCCGCCTGCCGGCAGCTTTTC					
NovO_R121K_1_NovO_rev5_E6_w11-8(21>899)	←	TTCGTGAGTCGTAATGCCTT					
NovO_R121K_1_NovO_rev6_F6_w11-81(1>898)	←	T-CGTGAGTCGTAATGCCTTTCACAAACTGACCCGCCTGCCGGCAGCTTTTC					
NovO_R121K_1_NovO_rev7_G6_w11-8(25>930)	←	T-CGTGAGTCGTAATGCCTTTCACAAACTGACCCGCCTGCCGGCAGCTTTTC					
NovO_R121K_1_NovO_fwd8_H5_w11-81(4>936)	←	T-CGTGAGTCGTAATGCCTTTCACAAACTGACCCGCCTGCCGGCAGCTTTTC					
NovO_R121K_1_NovO_rev8_H6_w11-81(1>933)	←	T-CGTGAGTCGTAATGCCTTTCACAAACTGACCCGCCTGCCGGCAGCTTTTC					
NovO_6H_C term codon opti.seq(1>711)	→	T-CGTGAGTCGTAATGCCTTTCACCGTCTGACCCGCCTGCCGGCAGCTTTTC					
NovO_R121K_1_NovO_fwd1_A5_w11-81(1>784)	→	T-CGTGAGTCGTAATGCCTTTCACAAACTGACCCGCCTGCCGGCAGCTTTTC					
NovO_R121K_1_NovO_fwd2_B5_w11-8(43>983)	→	T-CGTGAGTCGTAATGCCTTTCACAAACTGACCCGCCTGCCGGCAGCTTTTC					
NovO_R121K_1_NovO_fwd3_C5_w11-8(30>905)	→	T-CGTGAGTCGTAATGCCTTTCACAAACTGACCCGCCTGCCGGCAGCTTTTC					
NovO_R121K_1_NovO_fwd4_D5_w11-8(29>836)	→	ACCCGCCTGCCGGCAGCTTTTC					

**R116L**

		1260	1270	1280	1290	1300	1310	1320
		ATGCCCAACTGCTGGCAGGCATGGAAGATCTGGCTGGTAAAGCCGA-CCTGGTCGTGAGTCTTA						
R116L_10_NovO_rev_5_NovO_rev_5_(12>941)	←	ATGCCCAACTGCTGGCAGGCATGGAAGATCTGGCTGGTAAAGCCGA	ACCTGGTCGTGAGTCTTA					
R116L_10_NovO_rev_6_NovO_rev_6_F(1>951)	←	ATGCCCAACTGCTGGCAGGCATGGAAGATCTGGCTGGTAAAGCCGA-CCTGGTCGTGAGTCTTA						
R116L_10_NovO_fwd_8_NovO_fwd_8_(2>1091)	←	ATGCCCAACTGCTGGCAGGCATGGAAGATCTGGCTGGTAAAGCCGA-CCTGGTCGTGAGTCTTA						
R116L_10_NovO_rev_7_NovO_rev_7_G(1>938)	←	ATGCCCAACTGCTGGCAGGCATGGAAGATCTGGCTGGTAAAGCCGA-CCTGGTCGTGAGTCTTA						
R116L_10_NovO_rev_8_NovO_rev_8_(16>994)	←	ATGCCCAACTGCTGGCAGGCATGGAAGATCTGGCTGGTAAAGCCGA-CCTGGTCGTGAGTCTTA						
NovO_6H_C term codon opti.seq(1>711)	→	ATGCCCAACTGCTGGCAGGCATGGAAGATCTGGCTGGTAAAGCCGA-CCTGGTCGTGAGTCTTA						
R116L_10_NovO_fwd_1_NovO_fwd_1_A(1>981)	→	ATGCCCAACTGCTGGCAGGCATGGAAGATCTGGCTGGTAAAGCCGA-CCTGGTCGTGAGTCTTA						
R116L_10_NovO_fwd_2_NovO_fwd_2_(53>1049)	→	ATGCCCAACTGCTGGCAGGCATGGAAGATCTGGCTGGTAAAGCCGA-CCTGGTCGTGAGTCTTA						
R116L_10_NovO_fwd_4_NovO_fwd_4_D(1>955)	→	GNGGNGNNC-T-						

**Media preparation**

The following procedures prepare 1 L of the given medium.

**Luria Bertani (LB) Broth**

1. Dissolve Bacto Tryptone (10.0 g, # BD211701), Bacto Yeast Extract (5.0 g, # BD212730) and NaCl (5.0 g) in de-ionised water.
2. Make up to 1L with de-ionised water.
3. Adjust to pH 7.0 with aqueous 5 M NaOH.
4. Dispense as required and sterilise by autoclaving (fluid cycle 121°C, 20 min)

**Overnight Express (OE)**

1. Dissolve in Overnight Express powder (60.0 g, Novagen #71491-5) in de-ionised water.

2. Make up to 1 L with de-ionised water.
3. Adjust to pH 6.9 with aqueous 5 M NaOH.
4. If aliquoting into Ultra Flasks add 1 mL Antifoam (Dow Corning #632436X).
5. Autoclave 121°C for 15mins

### **Magic Media**

NB: For 950mls (50 mL Component B (Invitrogen #K6801) added directly before use):

1. Dissolve contents of Component A pouch (Invitrogen #K6801) in 950 mL de-ionised water.
2. Add 1 mL Antifoam (Dow Corning #DC1520)
3. Autoclave on Liquid cycle for 121°C for 20mins.

### **Terrific Broth**

1. Dissolve Bacto Tryptone (12.0 g, BD #211701), Bacto Yeast Extract (24.0 g, BD #212730),  $\text{KH}_2\text{PO}_4$  (2.3 g, BDH #29608CR) and  $\text{K}_2\text{HPO}_4$  (12.5 g, Merck #1.05101.1000) in 800 mL de-ionised water and make up to 1 L with de-ionised water.
2. If aliquoting into Ultra Flasks, add 1 mL Antifoam (Dow Corning #632436X).
3. Autoclave 121° for 15mins

### **Modified Terrific Broth**

4. Dissolve Bacto Tryptone (12.0 g, BD #211701), Bacto Yeast Extract (24.0 g, BD #212730), Bacto Peptone (2.0 g, BD #211830),  $\text{KH}_2\text{PO}_4$  (2.3 g, BDH #29608CR) and  $\text{K}_2\text{HPO}_4$  (12.5 g, Merck #1.05101.1000) in 800 mL de-ionised water and make up to 1 L with de-ionised water.
5. If aliquoting into Ultra Flasks, add 1 mL Antifoam (Dow Corning #632436X).
6. Autoclave 121° for 15mins

### **G1X Minimal Media**

1. Dissolve  $\text{KH}_2\text{PO}_4$  (17.0 g),  $\text{K}_2\text{HPO}_4$  (15.3 g) and  $(\text{NH}_4)_2\text{SO}_4$  (3.0 g) in 830 mL de-ionised water.
2. Add 6 mL of a solution containing  $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$  (3.37 g/L),  $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$  (0.3 g/L),  $\text{MgSO}_4 \cdot 2\text{H}_2\text{O}$  (37.5 g/L),  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$  (0.99 g/L),  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$  (0.63 g/L),  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  (0.13 g/L),  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$  (0.13 g/L) and  $\text{H}_3\text{BO}_3$  (0.04 g/L).
3. Add 300  $\mu\text{L}$  of a solution containing D-biotin (0.01 mg/mL), thiamine (1 mg/mL) and  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$  (1 g/mL).

4. Adjust to pH 7.4
5. Autoclave 121° for 15mins

### **Purification of CASTing libraries using KingFisher™ Flex purification system**

Pick-up	His Tip comb	
<b>Bead Wash</b>		
Beginning of step	Precollect	No
	Release beads	Yes
Mixing / heating	Mixing time, speed	00:02:00, medium
	Heading during mixing	No
End of step	Postmix	No
	Collect count	3
	Collect time (s)	5
<b>Binding</b>		
Beginning of step	Precollect	No
	Release beads	Yes
Mixing / heating	Mixing time, speed	00:15:00, medium
	Heating during mixing	No
End of step	Postmix	No
	Collect count	3
	Collect time (s)	5
<b>Wash (x4)</b>		
Beginning of step	Precollect	No
	Release beads	Yes
Mixing / heating	Mixing time, speed	00:02:00, medium
	Heating during mixing	No
End of step	Postmix	No
	Collect count	3
	Collect time (s)	5
<b>Elution</b>		
Beginning of step	Precollect	No
	Release beads	Yes
Mixing / heating	Mixing time, speed	00:10:00, medium
	Heating during mixing	No

---

End of step	Postmix	No
	Collect count	3
	Collect time (s)	5

---

Release Beads	Wash 4	
	Release time, speed	00:00:30, fast

---

Leave	His Tip comb	
-------	--------------	--



## SelMet-NovO X-ray diffraction data collection and refinement statistics

(collection on a single crystal)	NovO/SAH
Data collection	
Space group	P2 <sub>1</sub>
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	43.724, 66.970, 80.563
$\alpha$ , $\beta$ , $\gamma$ (°)	90.000, 92.99, 90.000
Resolution (Å)	66.97-1.90 (2.00-1.90)
<i>R</i> <sub>merge</sub>	0.056 (0.080)
<i>I</i> / $\sigma$ <i>I</i>	22.6 (10.6)
Mn(I) CC (1/2)	0.997 (0.988)
Anomalous completeness (%)	88.6 (53.2)
Completeness (%)	90.7 (58.0)
Anomalous multiplicity	3.0 (1.8)
Redundancy	6.0 (3.8)
Wilson B-factor	17.79
Refinement	
Resolution (Å)	66.97-1.90
No. reflections	201009 (11593)
No. uniq reflections	33239 (3062)
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.178/0.202
No. atoms	4330
Protein	3618
Ligand/Other	52/10
Water	650
Ramachandran	
Favored (%)	97
Allowed (%)	3
Outliers (%)	0.22
Rotamer outliers	0.54
B-factors	
Protein	23.24
Ligand/Other	11.13/26.50
Water	31.68
R.m.s deviations	
Bond lengths (Å)	0.005
Bond angles (°)	1.032

\*Highest resolution shell is shown in parenthesis.