Establishing a Biocatalytic Platform for Small Molecule C-Alkylation

Iain McKean

Thesis submitted for the degree of
Doctor of Philosophy

27th September 2019

Academic Supervisors: Prof. G. A. Burley & Prof. P. A. Hoskisson

Declaration of Authenticity

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 the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed: _____ (Iain JW McKean)

Date 27th September 2019

Contents

Acknowledgements

I would like to thank Glenn Burley for giving me the opportunity to carry out a PhD in his research group, Paul Hoskisson, Luke Humphreys and Radka Snajdrova for supervision, input into the ongoing direction of the project and many useful discussions. I would also like to thank Joanna Sadler for help and patience in the development of synthetic biology techniques such as overexpression and purification of enzymes which were fundamental for the work carried out.

I would like to thank both the Burley and Hoskisson groups at the University of Strathclyde for their guidance and support for the duration of my project. In particular I would like to thank all current members of the Burley group for creating an enjoyable and supportive working environment. Thanks to the Hoskisson group for helping me troubleshoot many different biology issues over the years.

Thanks to Andrea, Fergus, Jamie and Steven for their help in proof-reading countless abstracts, experimentals, presentations and for a variety of interesting scientific and non-scientific discussions! To Jack for keeping morale up with his fantastic jokes and to Giacomo for his sunny disposition.

Thanks to my family for their help and support through the past four years.

Finally, thanks to Laura for all your support through the highs and lows of the past few years – I couldn't have done this without you.

Abstract

The alkylation of aromatic small molecules is an essential synthetic process used throughout medicinal chemistry. Key limitations of synthetic methods such as the Friedel-Crafts reaction is its lack of chemoselectivity in complex organic scaffolds and stoichiometric amounts of metallic waste produced by this process. A potential solution to this issue could be the use of methyltransferase (MTase) enzymes as biocatalysts to carry out this process. In nature, S-adenosyl-L-methionine (SAM) dependent MTases carry out such reactions on a variety of substrates ranging from small molecules to DNA and proteins. SAM donates its methyl group to the target substrate with *S*-adenosyl-L-homocysteine (SAH) generated as a stoichiometric byproduct.

The MTase NovO regiospecifically transfers alkyl groups from chemically synthesised SAM analogues to a coumarin scaffold which is a precursor to the naturally-occurring antibiotic novobiocin. Chemical synthesis of SAM analogues has the issue of producing a diastereomeric mixture with epimers generated at the sulfonium centre, and display limited stability. SAM can be generated biosynthetically from the reaction of 5′-deoxy-5′-chloroadenosine (ClDA) with L-methionine (L-met), catalysed by an enzyme, SalL, from *Salinospora tropica.* This method is particularly attractive as ClDA and L-met are cheap and produce only the biologically active (S,S) diastereomer.

This thesis is a study of both SalL and NovO and their use in tandem SAM generation/alkylation (Scheme). The crystal structure of SalL has been solved to guide Site-Directed Mutagenesis (SDM). This method has then been used to better understand key residues in substrate binding and enzyme activity. Further to this, rational mutations were made in an attempt to increase space in the binding pocket to allow L-met analogues bearing groups larger than methyl to be accommodated. To this end, a range of L-met analogues were synthesised. The crystal structure of SalL revealed a channel which leaves the 2-position of ClDA solvent exposed. This raises the possibility of accommodation of 2-modified ClDA analogues as substrates of SalL. A suite of 2-modified ClDA analogues were synthesised to test this hypothesis.

The current substrate scope of NovO is narrow so it was desired to investigate coumarins which have not yet been investigated with NovO. A range of coumarins were synthesised including a precursor to a compound which has shown activity as an inhibitor of Heat Shock Protein 90 (Hsp90) which has been implicated in a number of cancers. To explore the use of

NovO as a method for late-stage functionalisation of drug molecules methylation of 7-hydroxywarfarin, a metabolite of clinically approved drug warfarin, has been carried out.
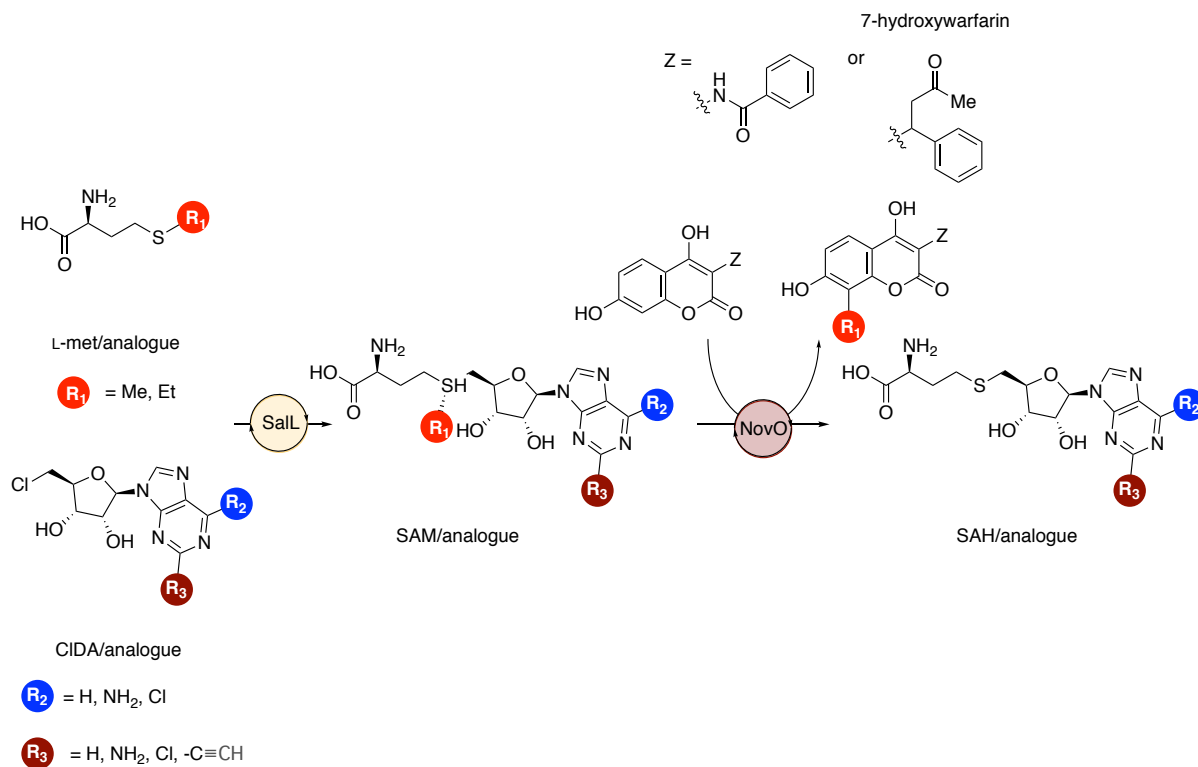
Scheme. Conversion of ClDA and L-met to SAM catalysed by SalL. Subsequently SAM is used as the cofactor of NovO to transfer a methyl or ethyl group to a target substrate such as 7-hydroxywarfarin with generation of SAH.

Abbreviations

Ade	Adenine
ADK	Adenosine Kinase
aDMA	Asymmetric Dimethylated Arginine
Ado	Adenosine
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
AMPPNP	5′-adenylyimido-triphosphate
ATase	Acyltransferase
ATP	Adenosine triphosphate
BAr$^F_4{}^-$	tetrakis[3,5-bis(trifluoromethyl)phenyl]borate
Bn	Benzyl
BSA	Bovine Serum Albumin
ClDA	5′-deoxy-5′-chloroadenosine
ClDEA	5′-chloro-5′-deoxy-2-ethynyladenosine
ClNP	2-chloro-4-nitrophenol
COMT	Catechol O-methyltransferase
CV	Column volumes
DCM	Dichloromethane
DMAP	N, N-Dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ETA	5′-Ethylthioadenosine
Et	Ethyl
EtOAc	Ethyl acetate
FDA	5′-fluoro-5′-deoxyadenosine
FDAS	5′-fluoro-5′-deoxyadenosine synthase
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HMT	Halide Methyltransferase

HRMS	High resolution mass spectrometry
HSL	Homoserinelactone
Hsp90	Heat Shock Protein 90
L-eth	L-ethionine
L-met	L-methionine
L-Se-met	L-selenomethionine
LB	Luria Broth
LCMS	Liquid chromatography mass spectrometry
MAT	Methionine Adenosyl Transferase
Me	Methyl
MeCN	Acetonitrile
MeOH	Methanol
MOM	methoxymethyl
MTA	5′-Methyl-5′-deoxythioadenosine
MTAN	5′-Methyl-5′-deoxythioadenosine Nucleosidase
MTase	Methyltransferase
MW	Molecular weight
NAF	Non-absorbed fraction
OD	Optical Density
PDB	Protein Data Bank
PET	Positron Emission Tomography
PGHS	Prostaglandin-endoperoxide synthase
PKMT	Protein Lysine Methyltransferase
PPK2	Family-2 polyphosphate kinase
PPNP	imido-triphosphate
PRMT	Protein Arginine Methyltransferase
RNA	Ribonucleic acid
RP-HPLC	Reverse Phase-High Pressure Liquid Chromatography
SAE	S-adenosyl-L-ethionine
SAHH	S-adenosyl-L-homocysteine hydrolase
SAM	S-adenosyl-L-methionine
sDMA	Symmetric Dimethylated Arginine
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC	Size Exclusion Chromatography
SeSAM	*Se*-adenosyl-L-methionine
t-AmylOH	*tert*-amyl alcohol
t-BuOH	*tert*-butanol
t-BuOOH	*tert*-butyl hydroperoxide
tet	Tetrazole
THF	Tetrahydrofuran
TLC	Thin layer chromatography
TMS	Trimethylsilyl

Chapter 1

Introduction

1 Introduction

1.1 Late-Stage Functionalization of Aromatic Small Molecules

In 2011, 67% of the top selling small molecule drugs contained at least one methyl group.[1] The presence of these groups change the physicochemical properties of pharmaceutical compounds, which in turn can enhance their pharmacokinetic properties and drug potency.[2] The underpinning reason for this is the methyl group induces a change in the three dimensional shape of a substrate to one which better complements the active site of a particular target thus increasing binding affinity.[1, 3]

The concept of modifying the drug-like properties of a molecule through introduction of a methyl group has been termed the "magic methyl" effect due to the significant change in molecular properties that is brought about by a modest chemical modification.[1] This was exemplified by Pfizer's selective inhibitor of the prostaglandin-endoperoxide synthase (PGHS) enzyme, responsible for production of prostaglandins within the body, by replacing a hydrogen atom on the drug diclofenac (**1.1**) with a methyl group (Figure 1.1a).[4] Starting from this known inhibitor of PGHS-1 and 2, a chlorine atom was substituted for fluorine and the 4-H of the phenylacetic acid unit was replaced by a methyl group. The methylated compound lumiracoxib (**1.2**) (Figure 1.1b) exhibited enhanced potency for PGHS-1 and binding affinity was reduced (K_i PGHS-1 = 0.01 vs K_i PGHS-1 = 3.2) while PGHS-2 binding remained relatively consistent (K_i PGHS-2 = 0.01 vs K_i PGHS-2 = 0.06), thus increasing overall selectivity. The rationale behind this alteration in binding affinity was that the increased volume of the methyl group relative to the hydrogen atom disrupts a hydrogen bonding interaction between the carboxylate group and Arg120 of PGHS-1. This is caused by the induction of a steric clash between the methyl group, and residues Ser353 and Tyr355.[5]
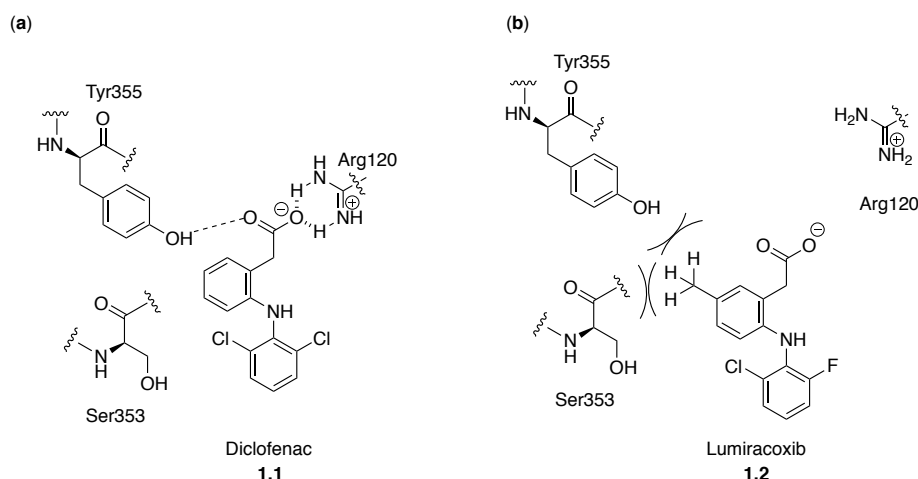
Figure 1.1a. Schematic representation of lumiracoxib in the binding pocket where hydrogen bonding to Arg120 is disrupted by a steric clash with Ser353. **b**. Schematic representation of diclofenac in the binding pocket without the steric clash observed for lumiracoxib.[2, 5]

Methylation also alters the solubility of drug candidates and varies metabolism.[6-8] The addition of a methyl group increases the overall lipophilicity of a molecule.[9] This change generally makes small molecules more cell permeable by facilitating improved interactions with lipophilic cell membranes without significantly increasing the overall molecular weight of the molecule.[10] As a result, efforts to site-specifically methylate small molecules, especially at a late-stage in a synthetic workflow are highly desirable.

1.2 Synthetic Methods of Alkylation

1.2.1 Friedel-Crafts Alkylation

The most widely used method of incorporating alkyl groups into aromatic systems is the Friedel-Crafts alkylation, first introduced in 1877.[11] In this reaction, a strong Lewis acid (such as AlCl$_3$) is required to form a C-C bond between an arene (**1.3**) and an alkyl group from an alkyl chloride (**1.4**) to give alkylated product (**1.5**) (Scheme 1.1). Although widely used on an industrial scale,[12] this method has some significant drawbacks such as the need for stoichiometric or super-stoichiometric quantities of strong Lewis acids, poor regioselectivity and potential to be over-alkylated (**1.6**).[13]
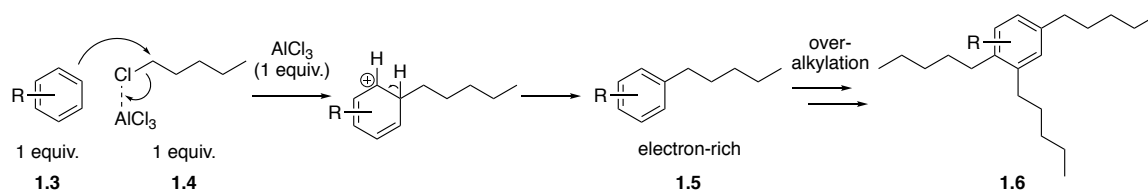
Scheme 1.1. A representative Friedel-Crafts alkylation of arene (**1.3**) and 1-chloropentane (**1.4**).

1.2.1.1   Recent advances in Friedel-Crafts alkylation

Advances have been made towards developing the selectivity of the Friedel-Crafts reaction and also new approaches to make the reaction catalytic.[14] Brønsted acid catalysis using a quaternary phosphonium salt and a tetrakis[3,5-bis(trifluoromethyl)phenyl]borate (BAr$^F_4{}^-$) (**1.9**) counterion has been used to carry out alkylations of electron-rich heteroarenes (Scheme 1.2).[15] The reaction was developed by observing the addition of an aniline derivative (**1.8**) to a ketimine (**1.7**) in high conversion. Although limited to these specific substrates at present, the reaction can tolerate mono-substitution at R^1 on the 6-membered ring of the ketimine, Me, MOM or Bn protecting groups at R^2 and various linear and cyclic alkyl groups at R^3/R^4. The reaction has also been shown to accept furan and thiophene derivatives in place of the aniline. Despite these advances, attempts to induce chirality through the presence of a chiral phosphonium salt have thus far been unsuccessful, limiting the usefulness of this transformation.

1.11	R^1	R^2	R^3	R^4	% Yield
a	H	Me	Me	Me	95
b	H	Bn	Me	Me	99
c	5-Br	Me	Me	Me	89
d	H	Me	Et	Et	99
e	H	Me	Ph	Ph	99
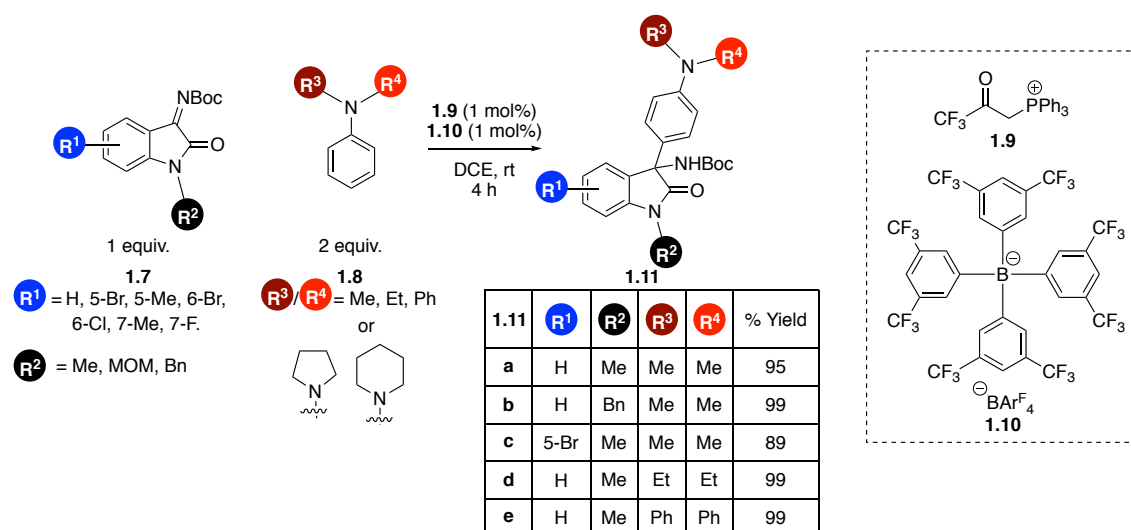
Scheme 1.2. Friedel-Crafts reaction of ketimine (**1.7**) with anilines (**1.8**). Table denotes select examples of substrate scope.

A Sc-catalysed Friedel-Crafts reaction was developed by Tsuchimoto *et al.* which avoids some of the traditional pitfalls such as their sensitivity to moisture. This also limits the use of

alcohols as substrates for this process as water generated within the reaction or the alcohols themselves may cause catalytic deactivation.[16] This in turn requires super-stoichiometric amounts of Lewis acid, which are also unrecoverable. In contrast, Sc(OTf)$_3$ is stable to moisture therefore opening up the use of alcohols as substrates (Scheme 1.3). The catalyst can also be recovered by evaporating the aqueous layer of the reaction and further drying the recovered material by heating under vacuum. A test of the recovered catalyst in the model reaction of benzyl alcohol (**1.12**) with benzene (**1.13**) to form **1.14** showed comparable yields upon repetition of the reaction (Scheme 1.3). For the given examples, all yields were ≥ 75% and substituents such as chloride on the benzyl alcohol were tolerated. However, these reactions suffer from a lack of regioselectivity typically exhibited by traditional Friedel-Crafts alkylations as mixtures of ortho, meta and para products were observed (Scheme 1.3).

1.14	R^1	R^2	% Yield
a	H	H	91
b	H	Me	quant*
c	H	OMe	87**
d	p-Me	H	91
e	p-Cl	H	93
f	o-Br	H	75
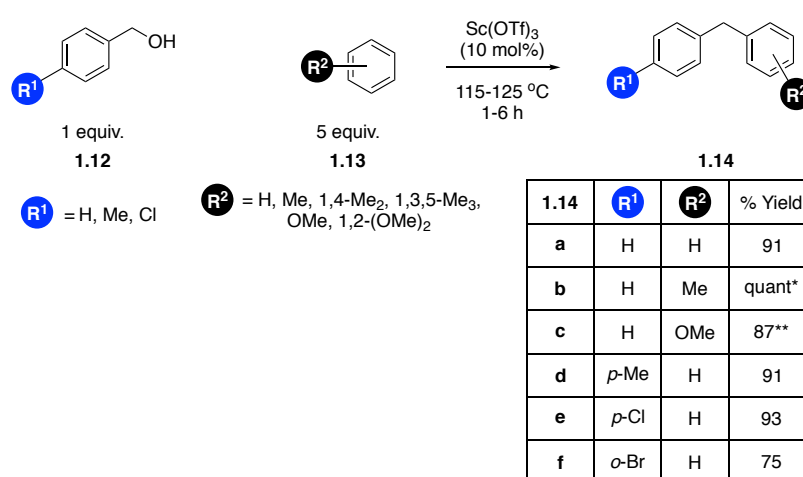
Scheme 1.3. Friedel-Crafts alkylation of benzene (**1.13**) using benzyl alcohol (**1.12**). *ratio of o:m:p = 48:7:45. **ratio of o:m:p = 57:trace:43.

Paras and MacMillan developed a method of the Friedel-Crafts reaction which regioselectively alkylated pyrrole (**1.15**) using an α,β-unsaturated aldehyde chain (**1.16**) (Scheme 1.4a) and generated a defined chiral centre (**1.17**).[17] This method utilises an organocatalyst (**1.18**) to form an iminium species by reaction with the aldehyde thus preventing direct nucleophilic attack through sheer steric bulk at the α-position (Scheme 1.4b). Instead, 1,4-addition is favoured due to the lower steric demands but the stereocentre formed is still under the influence of the steric properties of the catalyst thus resulting in formation of a defined stereocentre. A variety of alkylated pyrroles were tolerated as were a number of diversely alkylated α, β-unsaturated aldehydes.
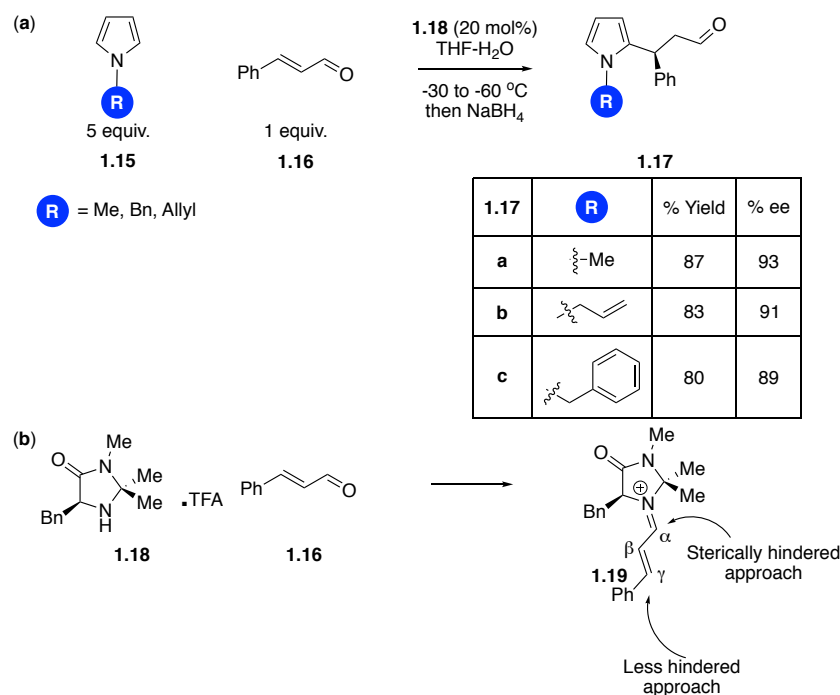
Scheme 1.4a. Select examples to illustrate the scope of enantioselective Friedel-Crafts alkylation. **b.** Formation of iminium intermediate **1.19** from condensation of **1.18** and **1.16**. Attack at the iminium carbocation is disfavoured by the steric bulk at C-α, with 1,4-addition favoured instead.

Despite these new developments, there are still a number of pitfalls associated with this methodology such as limited regioselectivity and functional group tolerance. Also, the function of some of these catalysts renders them useful for only a specific set of substrates thus limiting their broad utility. The requirement of methods for the late-stage functionalisation of small molecules has driven developments of other new methodologies for this purpose such as radical alkylations and biocatalytic transformations.

1.2.2 Radical Alkylation of Heteroarenes

A contemporary example of late-stage functionalisation is work pioneered by Langlois[18] and further developed by Baran *et al.,* which has resulted in the development of a range of zinc sulfinate salts.[19] Under mild reaction conditions, small aromatic and heterocyclic molecules such as 2-methylquinoxaline (**1.20**) can undergo radical addition *via* the generation of a radical species from zinc sulfinate salt (**1.21**) to give a mixture of functionalised products (**1.22**) (Scheme 1.5a). The reaction has been postulated to proceed by the following mechanism (Scheme 1.5b): *t*-BuOOH (**1.23**) can spontaneously break down to generate a *t*-BuO radical which can in turn generate a trifluoromethanesulfinate radical (**1.24**). Decomposition of this

species results in formation of a CF$_3$ radical which can then add onto 2-methylquinoxaline (**1.20**) to give a mixture of functionalised products (**1.22**).
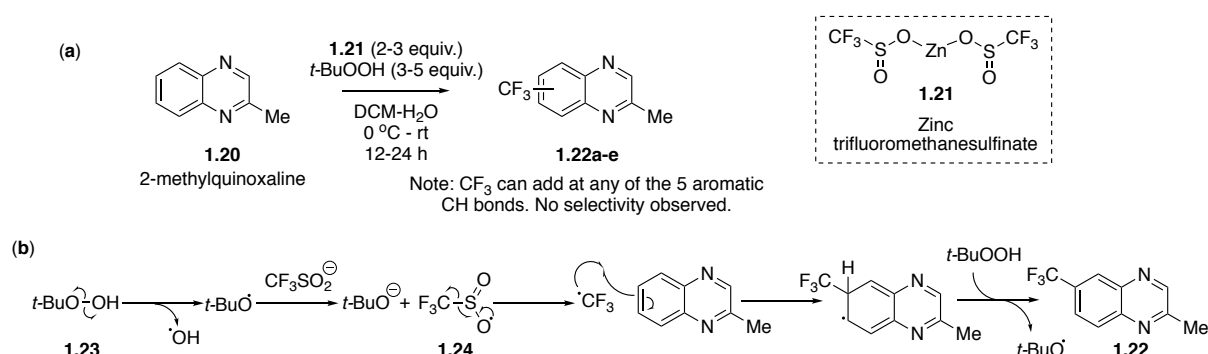
Scheme 1.5a. Radical fluoroalkylation of 2-methylquinoxaline. **b.** Representative mechanism of radical trifluoromethylation.[20]

The late stage introduction of CF$_3$ groups is of particular interest as these groups are similar in size to methyl groups but as C-F bonds are significantly stronger than C-H bonds, CF$_3$ groups are more resistant to metabolic oxidation.[21] Late-stage functionalisation allows for fine tuning drug molecule properties towards the end of a synthesis without disturbing key functionalities. However, the zinc sulfinate salt method is not without its own difficulties. It can be very effective in a situation where there is only a single C-H bond position where the radical addition can occur, but as the reaction proceeds by the generation of CF$_3$ radicals (or CF$_2$H/CF$_3$CH$_2$) poor regioselectivity is typically observed. A representative example of this is CF$_3$ addition to 2-methylquinoxaline **1.20** which results in an inseparable mixture of five compounds (**1.22a-e**).[19]

Nagib and MacMillan disclosed a process of photoredox trifluoromethylation which displayed utility across a wide range of compounds (**1.25a-e**, Scheme 1.6).[22] Visible irradiation of an iridium photocatalyst generates an excited species which in turn carries out a single electron transfer to triflyl chloride which breaks down to form trifluoromethyl radicals. After addition of the trifluoromethyl group to the arene, a single electron transfer regenerates the photocatalyst. Although effective in generation of compounds such as **1.26a**, other substrates like **1.25b** produced mixtures of regioisomers (**1.26b**) when reactions were carried out under these conditions. This process was then applied to a range of small molecules such as uracil to form the corresponding DNA base analogue (**1.26c**), a precursor to anti-Alzheimers drug Aricept (**1.26d**) and to ibuprofen (**1.26e**). Although this methodology was widely applicable,

the generation of regioisomers in a number of cases does limit the broader use of this methodology.
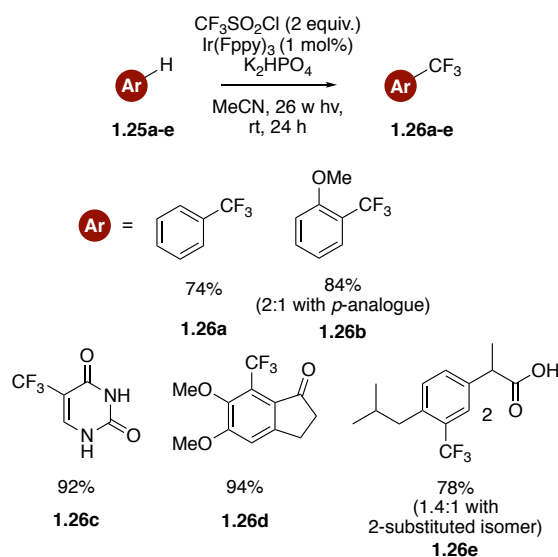
Scheme 1.6. Select examples of photoredox-mediated trifluoromethylation of arenes.

1.2.3 Alkylation of Arenes by Transition-Metal Catalysed Cross-Coupling

Over recent years, a number of methods of have been developed for carrying out selective alkylations through transition metal catalysis.[23] In 2011, Chen *et al.* developed a palladium catalysed method for the selective ortho alkylation of benzylamides (**1.27**) using alkyl halides (**1.28**) by utilizing a picolinamide protecting group (Scheme 1.7).[24] The arene starting material has been shown to maintain reactivity when a variety of substituents are present (**1.27**, Scheme 1.7a). This method tolerates a variety of alkyl chains, from methyl groups (**1.31a**) to benzyl groups (**1.31b**), and other functionalities such as esters (**1.31c**) and acetals (**1.31d**).

(a)

1.29	R¹	% Yield
a	⅟-OMe	95
b	⅟-CF₃	58*
c	⅟-Cl	86*
d	⅟-Br	86*

(b)

1.31	R²	% Yield
a	⅟-Me	90
b	phenethyl	87
c	(CH₂)ₙCO₂Me	85
d	dioxolane	73

Scheme 1.7a. Ortho-selective propylation of arenes directed by a picolinamide protecting group **b.** Ortho selective alkylation. *3-Methyl-3-pentanol as solvent, 135 °C.
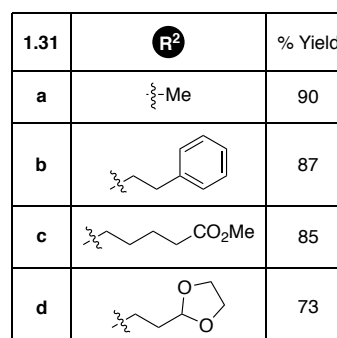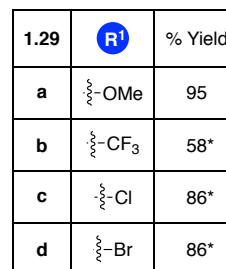
Introduction of methyl (Scheme 1.8a+b) and phenyl (Scheme 1.8c) groups to carboxylic acid bearing arenes (**1.32**/**1.34**) can be carried out *via* a palladium catalysed method developed by Yu *et al.*.[25] Although selective for the ortho position in good yield (**1.33**/**1.35**/**1.37**), in this work the scope of this transformation is limited to only methyl and phenyl groups and a limited palette of arenes have been tested as acceptors.
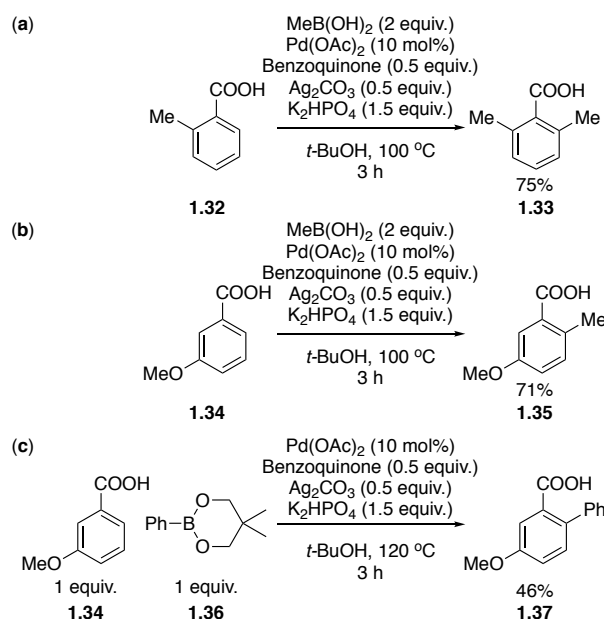
Scheme 1.8a. Palladium catalysed ortho methylation of 2-methylbenzoic acid. **b.** Palladium catalysed ortho methylation of 3-methoxybenzoic acid. **c.** Palladium catalysed ortho phenylation of 3-methoxybenzoic acid.

Some advances have been made on this work by Yang *et al.* who employed palladium catalysis and anionic oxidation to carry out C-H alkylation of arenes (Scheme 1.9).[26] This work used a pyridyl directing group (**1.38**) instead of the previously used carboxylic acid, and potassium trifluoroalkylborates as the source of alkyl group. This work showed an expansion in the types of alkyl group which could be transferred to the arene core, with ethyl, *n*-butyl and ethylbenzene all tolerated (**1.37**). However, for larger substituents the yields decrease significantly and the majority of substrates studied for this reaction had one of the ortho positions occupied so only one alkylation could take place.

1.38   **1.39**

1.39	R^1	R^2	% Yield
a	ⸯ–Me	ⸯ–Me	70
b	ⸯ–OMe	ⸯ–Me	64
c	ⸯ–OMe	ⸯ–Et	52
d	ⸯ–OMe		38
e	ⸯ–OMe		20

Scheme 1.9. Ortho alkylation of small aromatic molecules facilitated by palladium catalysis and a pyridyl directing group.

Another recent method of ortho alkylation using trifluoroalkylborates, palladium catalysis and MnF₃ as an oxidizing agent has been developed by Neufeldt *et al.* (Scheme 1.10). This work relies on having either a pyridyl ring or an amide as the directing group for C-H activation (**1.40a-b**).[27] As in the previous example, many of the reactions provided have one of the ortho positions either blocked, or partially blocked by a flanking meta group, thus calling into question whether the reaction is truly selective.
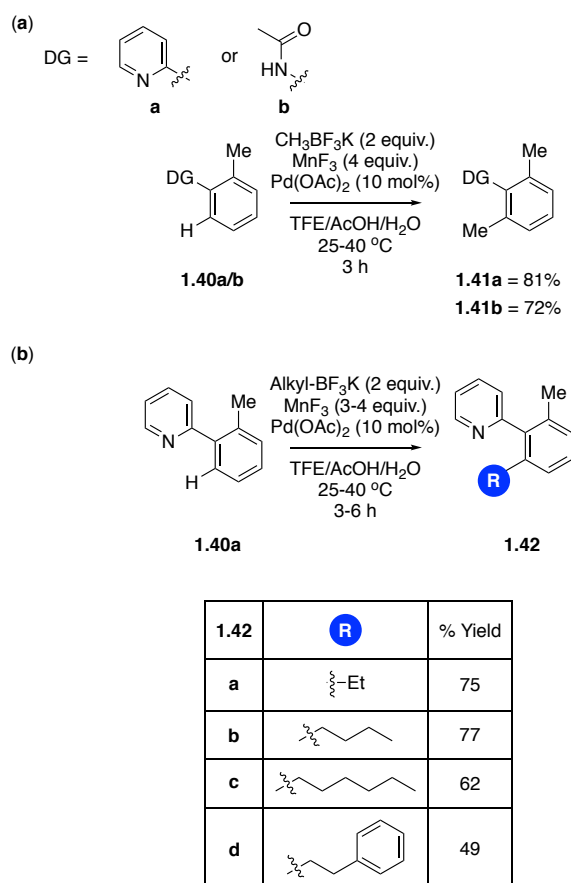
Scheme 1.10a. Selective ortho methylation of aromatic compounds. **b**. Selective ortho alkylation of pyridyl 2-methylbenzene.

Although significant progress has been made in the development of synthetic routes utilizing palladium chemistry, there is a distinct lack of a general chemical process for regioselective *C*-alkylation. In recent years, developments in biocatalysis have put it forward as an alternative to chemocatalysis. Not only do these methods avoid some of the cost/supply issues caused by use of rare metals,[28] they can also take advantage of unique properties of enzymes such as their ability to utilise mild reaction conditions while also exhibiting exquisite stereocontrol. They have been used as effective catalysts for generating functionalities such as amides,[29] reduction of alkenes without palladium[30] and formation of carbon-carbon bonds by aldolases,[31] to give a few examples.

1.3 The Application of Biocatalysis in the Synthesis of Small Molecules

Enzymes are Nature's catalysts for constructing a vast array of natural products.[32, 33] As these processes originate from living systems they are typically carried out in water at ambient temperature and pressure. They have evolved over millions of years to carry out a myriad of complex transformations with exquisite stereo- and regio-selectivity. These attributes make enzymes an attractive option for industry where complex molecules are readily sought after. For example, over 50% of drug compounds in the pharmaceutical industry possess at least one chiral centre,[34] and could therefore benefit from these attractive properties of enzymatic reactions.[35] Further to this, there is a current drive in industrial chemistry to reduce energy consumption, organic solvent waste and use of expensive and toxic metal catalysts.[36]

A recent example of the use of biocatalysis in industry is in the synthetic route to Sitagliptin (**1.45**) (Scheme 1.11)[37]. First, in the chemocatalytic route the corresponding enamine was formed with NH$_4$OAc followed by a rhodium-catalysed hydrogenation at high pressure to generate the target chiral amine with 97% enantiomeric excess (ee). Finally, the desired phosphate salt was generated by treatment with a mixture of heptane and isopropanol followed by phosphoric acid. In contrast, a biocatalytic method has been developed where a transaminase is used to directly transform the carbonyl group to the desired primary amine with 99.95% ee, thus replacing the first two steps of the chemical method.
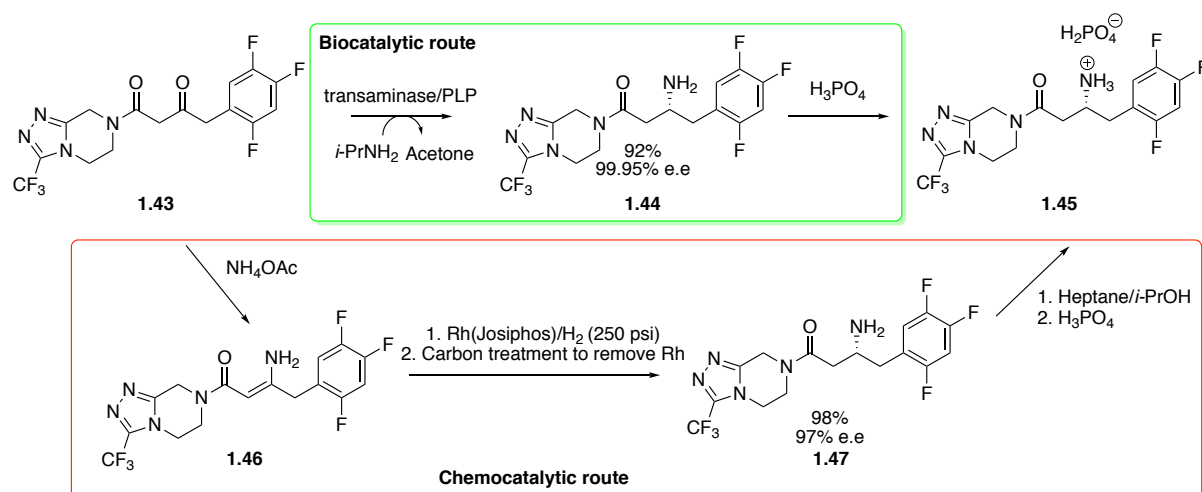
Scheme 1.11. Sitagliptin synthesis has been successfully carried out with a biocatalytic method (green) to replace the chemocatalytic method (red).[37]

The optimised enzymatic route was able to convert 200 g/L of prositagliptin ketone (**1.43**) using 6 g/L enzyme loading, resulting in 92% yield. This equates to a 10-13% increase in overall yield and an increase in productivity of 53% (units of kg/L per day). Beyond the direct benefits in generation of greater amounts of the target material, overall waste is reduced by 19%, metal catalysts are eliminated and removal of high-pressure vessels for hydrogenation also reduces cost and improves the safety of the process as a whole.

1.4 Current Methods of Biocatalytic Friedel-Crafts-like alkylation of small molecules

In recent years there have been a number of reported biocatalytic Friedel-Crafts-like C-C bond forming processes. Schmidt *et al*. have reported a Friedel-Crafts acylation of resorcinols (**1.48**) to form their corresponding aryl ketones (**1.49-1.50**) (Scheme 1.12).[38] This system employs an acyltransferase (ATase) from *Pseudomonas protegens* DSM 19095 which catalyses the reaction of 2 moles of monoacetylphloroglucinol to form a mole of phloroglucinol and a mole of 2,4-diacetylphloroglucinol (DAPG), an antibiotically active polyketide.
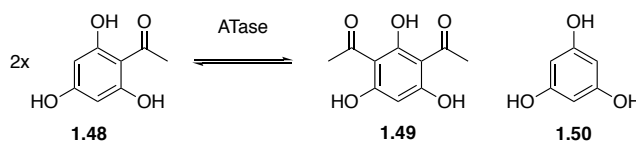
Scheme 1.12. Natural reaction of ATase *Pseudomonas aeruginosa.*[38]

The utility of this ATase was investigated using DAPG (**1.49**) as an acyl donor, and was found to tolerate aromatic compounds with modifications at both the R^1 and R^2 positions (**1.51**, Scheme 1.13). A variety of substituents are tolerated at R^2, however the R^1 position is much more restricted due to the proximity of this site to the incoming acyl group. Introduction of either a methyl or a methoxy group lowers the conversion significantly with a methyl group in particular almost entirely abolishes activity. A subsequent investigation of alternative acyl donors (to avoid the use of the non-commercially available compounds) led to discovery of donors **1.53** and **1.54**. Donor **1.54** in particular performed similarly to **1.49**, establishing the basis for further acylations of this type. The ability to tolerate moieties such as the chloride also opens up potential for further synthetic transformations.

Scheme 1.13. Scope of acyl transfer reaction with utilization of different acyl donors.

		% Conversion with different Acyl donors		
R^1	R^2	1.49	1.53	1.54
H	H	97	41	88
H	Cl	98	56	95
H	n-hex	>99	97	84
-Me	H	3	0	2
-OMe	H	41	12	18
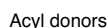
Resorcinols (**1.55**) have also been investigated as substrates for CylK, which catalyses a Friedel-Crafts-like alkylation in the biosynthetic pathway to natural product cylindrocyclophane F (**1.56**) (Scheme 1.14).[39] This reaction can be tailored to carry out a single alkylation to give a product where the alkylation has been carried out ortho- to both hydroxyl groups (Scheme 1.15). This study showed that a variety of long alkyl chains were accepted as substrates.

Scheme 1.14. Native reaction of CylK, which catalyses a dimerization process to form cylindrocyclophane F through an intermolecular Friedel-Crafts-like reaction followed by an intramolecular Friedel-Crafts-like reaction.
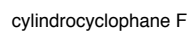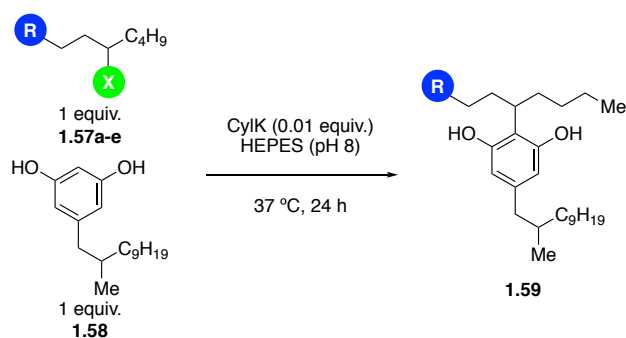
1.59	R	X	% Conversion
a	Me⎯NH⎯C(O)⎯⎯S⎯C(O)⎯	Cl	88
b	MeO⎯C(O)⎯	Cl	22
c	Me⎯NH⎯C(O)⎯⎯S⎯C(O)⎯	Br	76
d	Me⎯NH⎯C(O)⎯⎯S⎯C(O)⎯	I	29
e	Me⎯NH⎯C(O)⎯⎯S⎯C(O)⎯	F	0

Scheme 1.15. Friedel-Crafts-like alkylation of resorcinol scaffolds using CylK. Select examples of alkyl chains and halides which are accepted by the enzyme.

A primary goal of the study undertaken was to utilise this enzymatic system to synthesise benvitimod (**1.62**), a compound currently in phase II clinical trials as a treatment for psoriasis (Scheme 1.16).[40, 41] This compound displays striking similarity to known CylK substrates but requires installation of an isopropyl group adjacent to the two hydroxyls. However, in the published study of CylK this transformation is not facilitated by the enzyme, and the scope of CylK was limited to alkyl chains significantly larger than isopropyl. Nonetheless, this work demonstrates a starting point for directed evolution to facilitate this transformation by an enzymatic pathway.
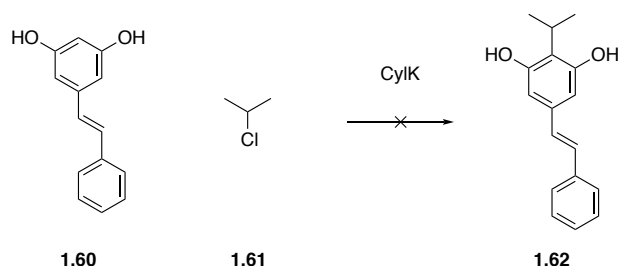
Scheme 1.16. Attempts to synthesise **1.62** (benvitimod) using CylK in the synthetic strategy above were unsuccessful.

1.5 Biocatalytic Methods of Methylation Using *S*-Adenosyl-L-Methionine Cofactors

Nature has evolved a highly efficient method to carry out late stage methylation of a vast array of substrates using methyltransferases (MTases). Effective at *O*-, *N*-, *S*- and *C*- alkylation, MTases have evolved over time to carry out diverse transformations both regio- and chemo-selectively.[42-50] One of the most studied families of MTases is the *S*-adenosyl-L-methionine (SAM) dependent family. SAM (**1.63**) acts as the cofactor to the enzyme and donates the methyl group to the target substrate (**1.64**), with generation of methylated product (**1.65**) and *S*-adenosyl-L-homocysteine (SAH) (**1.66**) as the byproduct (Scheme 1.17). SAH generation can cause issues in enzyme function as it is a known inhibitor of SAM dependent MTases. SAH can shut down the methylation reaction completely due to its high affinity for the SAM binding site of MTases.[51]
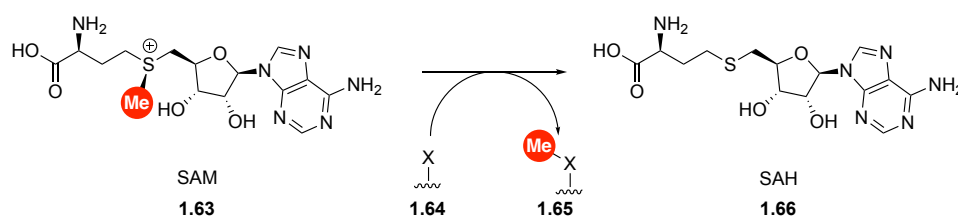
Scheme 1.17. SAM and byproduct SAH after methyl transfer has taken place. Where X = C, N, O or S.

SAM is the second most abundant cofactor in metabolism after adenosine triphosphate (ATP).[52] Originally called "active methionine" upon discovery by Cantoni, the structure of SAM consists of methionine covalently linked to the 5′ end of adenosine (Ado) (Scheme 1.17).[53] The first SAM MTase structure to be identified was catechol *O*-MTase (COMT): an enzyme involved in the metabolism of neurotransmitters in the central nervous system.[54] The structure of COMT consists of eight α-helices and one β-sheet to form the characteristic Rossman fold arrangement of the B-E helices and the 1-5 strands (Figure 1.2). This structural

fold is highly conserved across all known SAM MTases, despite the SAM binding domain varying greatly over the family.[55]
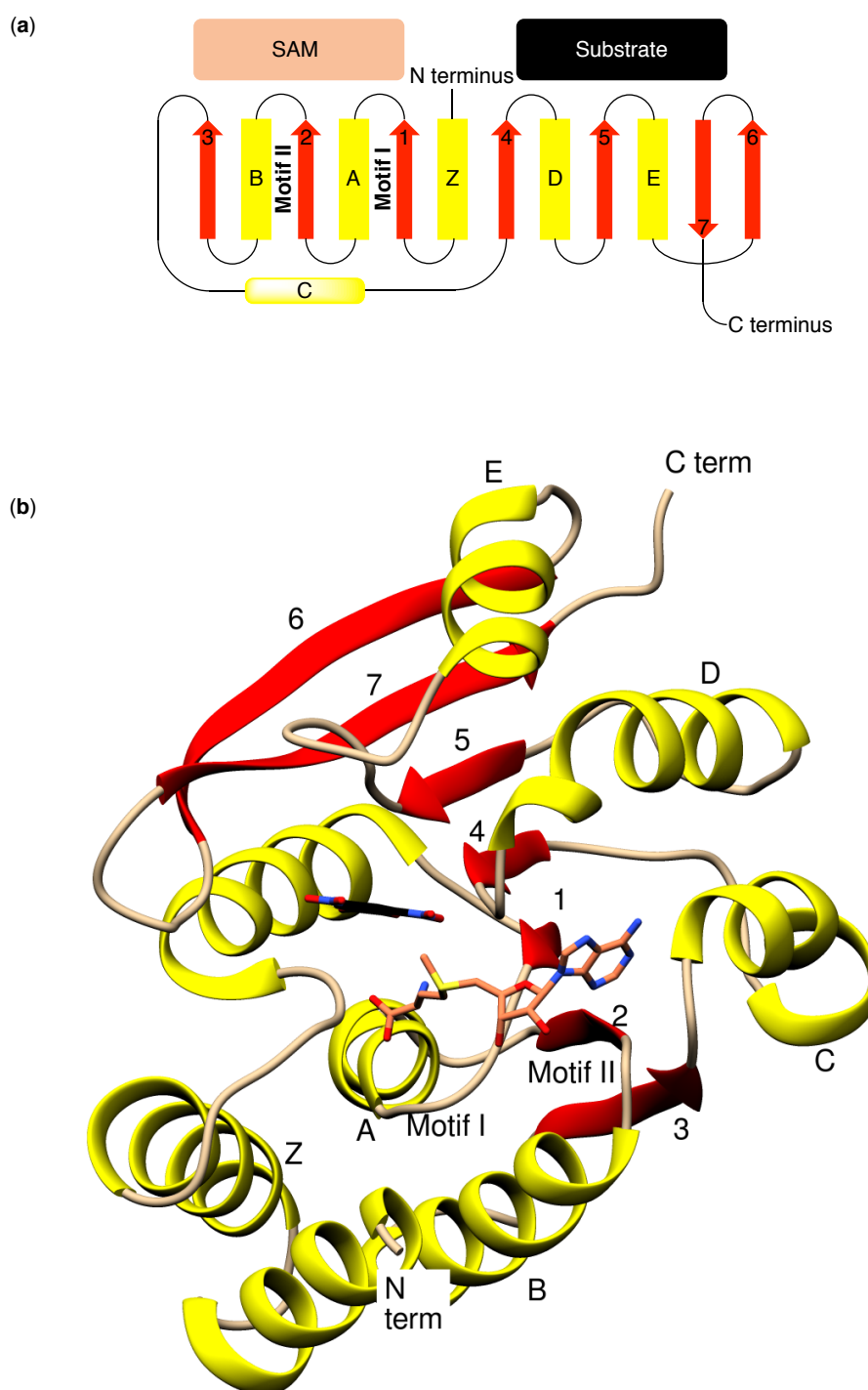
Figure 1.2a. Schematic topology of the core (Rossman type) fold of SAM-MTases. Yellow cylinders represent α-helices, red arrows represent β-strands.[55] **(b)** X-ray crystal structure of COMT, with SAM in the active site, from Rattus norvegicus (PDB accession code:1H1D).[54] Z region in yellow contains three of the eight total helices.

The location of the SAM binding site and binding conformation is conserved across the enzyme class although the chemistry of the SAM binding residues varies significantly across subfamilies.[56] There are two significant regions which appear to be common across SAM MTases. The first is a glycine-rich sequence of the format E/DXGXGXG which is implicated in the binding of the amino acid of SAM (Figure 1.3). This region is located between β-strand 1 and α-helix A and is commonly known as "Motif I". The second conserved region is a loop consisting of an acidic residue which binds the SAM cofactor through H-bonding with the ribose hydroxyls. This sequence is located between β-strand 2 and α-helix B and is known as "Motif II".[55] Across the SAM MTase family many insertions between subunits are known, however Motif I and II are generally conserved.[55]
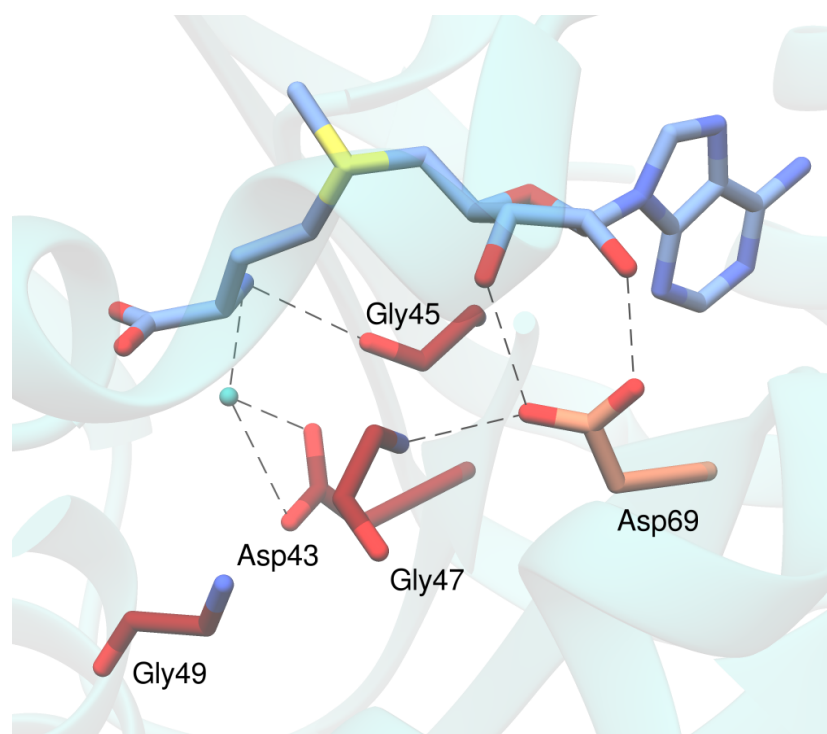
Figure 1.3. Crystal structure of a SAM-dependent MTase from Aquifex aeolicus (PDB Accession code: 3DH0).[57] Motif I is highlighted in dark red and binds the amino acid portion of SAM, Asp69 from Motif II is highlighted in coral and binds the ribose hydroxyl groups.

Despite these similarities, there are also significant structural differences between SAM MTases which act upon different substrates such as DNA, proteins, small molecules and lipids. DNA and protein MTases have their own distinct compositional features, the most significant of which is the addition of residues at the *C*-terminus of the Rossman fold. Small molecule MTases, however, have no chains of amino acids added to the *C*-terminus of the structural fold region but do have two additional helices at the *N*-terminus. Generally, this portion of the

enzyme varies greatly in composition and structure between small molecule MTases that act upon different substrates. This contributes to the variation in overall sequence between these MTases with which there is generally only a 15-20% sequence similarity. It is also common for these MTases to have a region consisting of inserts between β-strand 5 and α-helix E and β-strands 6 and 7 which form a "cover" over the active site. Lipid SAM MTases also show these typical features.[55]

In contrast, there are many different residues incorporated in each structure in the substrate binding region of the enzyme. As such there is a great variation in the chemistry of the binding interaction and therefore the substrates which can be bound.[55] It is this variation that leads to the diversity seen across the MTase family with such a wide variety of substrates acted upon: DNA (**1.67**), proteins (**1.68**), small molecules (**1.69**) and lipids (**1.70**) (Scheme 1.18).
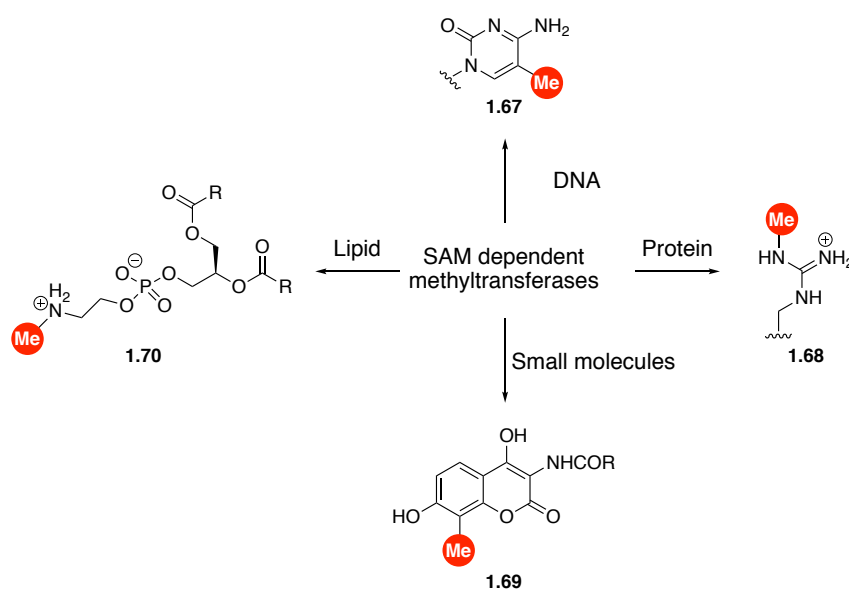
Scheme 1.18. Diversity of SAM MTase substrates. From top, methylation carried out by DNA MTases at 5-position of cytosine,[58] on proteins by PRMT,[59] on small molecules by NovO/CouO[50] and on lipids.[60]

1.6 Mechanistic Features of Methyltransferases

In each instance, a common trait of the MTase-catalysed methyl transfer is preorganisation of a nucleophilic site to be methylated to a conformation where it lies in close proximity to the electrophilic methyl group of SAM. Two possible mechanisms exist for the transfer of a methyl group onto the target nucleophile. In one case, a basic residue can remove a proton from the atom at which the transfer is occurring, thereby increasing the nucleophilicity to carry out an S_N2 type process with inversion of configuration at the CH_3 centre (Scheme 1.19 - blue).

Conversely, it could also be envisaged that if the substrate is sufficiently nucleophilic the attack step may occur prior to deprotonation. This phenomenon is likely governed by the substrate involved (Scheme 1.19 - red).[54, 61]
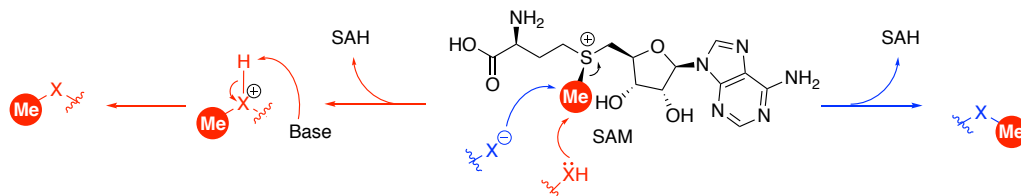
Scheme 1.19. Possible mechanisms of methyl transfer where X = C, N, O or S. **Blue pathway –** Deprotonation of substrate prior to nucleophilic attack. **Red pathway** – Deprotonation of substrate after nucleophilic attack.

Horowitz *et al.* recently proposed that this mechanism is conserved across a variety of MTases and involves the formation of abnormally strong hydrogen bonding interactions between the methylene protons flanking the sulfur group and residues within the active site (Scheme 1.20).[62] These hydrogen bonds hold SAM in a transition-state like geometry when attack of nucleophile **1.64** occurs. The resulting linear arrangement of X, Me and S promotes transfer of the methyl group.
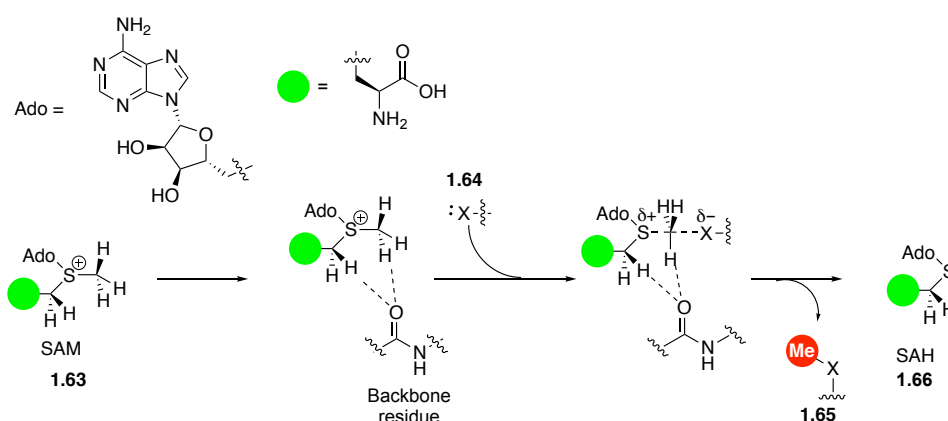
Scheme 1.20. General mechanism of methyl transfer where X = C, N, O or S. Mechanism of SAM dependent MTases proposed by Horowitz *et al.*[62] Hydrogen bonding exists between a carbonyl O from a backbone residue and methylene hydrogens adjacent to S. This constrains geometry of SAM into a transition state-like conformation, thus facilitating methyl transfer.

1.6.1 *O*-Methyltransferases

As previously discussed, one of the most studied *O*-MTases is COMT which causes degradation of neurotransmitters such as dopamine (**1.71**).[44] By methylating one of the hydroxyl groups (Scheme 1.21), **1.72** no longer functions as a neurotransmitter and is metabolically degraded within the body.[63] COMT have also been used in the biotechnology

industry as part of a microbial production of vanillin (**1.74**), the compound responsible for vanilla flavouring, from glucose **1.73** (Scheme 1.22).[64]
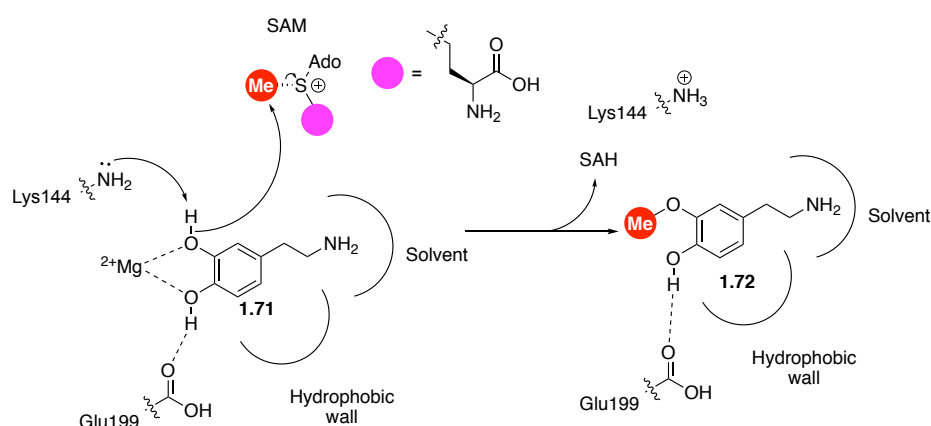
Scheme 1.21. Selective O-methylation of dopamine catalysed by COMT. Mechanism of COMT: Lys144 deprotonates the meta-phenolic proton while the molecule is held in place by hydrogen bonding to Glu199. Substrate orientation is determined by the amine tail interacting with the solvent.[65]
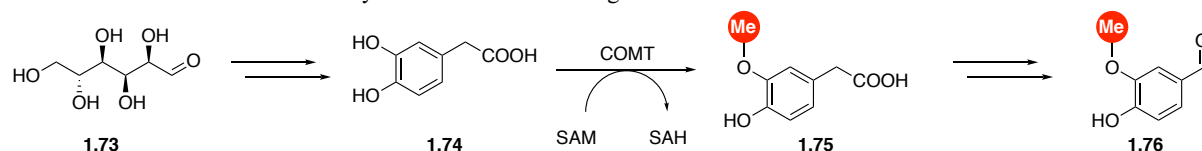
Scheme 1.22. Microbial synthesis of vanillin (**1.76**) from glucose (**1.73**).

1.6.2 *N*-methyltransferases

There are a number of *N*-MTases which have been studied, with the protein-arginine-N-MTases (PRMTs) and protein-lysine-MTases (PKMTs) two well-known examples.[42, 43] These SAM dependent enzymes operate by transfer of a methyl group on to the amine containing side-chains of arginine (**1.77**) (Scheme 1.23) and lysine (**1.82**) (Scheme 1.24).[59] Both are very common post-translational modifications of histones which play a key role in defining the structure of chromatin and in subsequent gene expression.[42, 43] There are also a variety of non-histone methylations known for both which carry out functions such as tumour suppression and regulation of transcription factors.[66, 67]

(a)

(b)
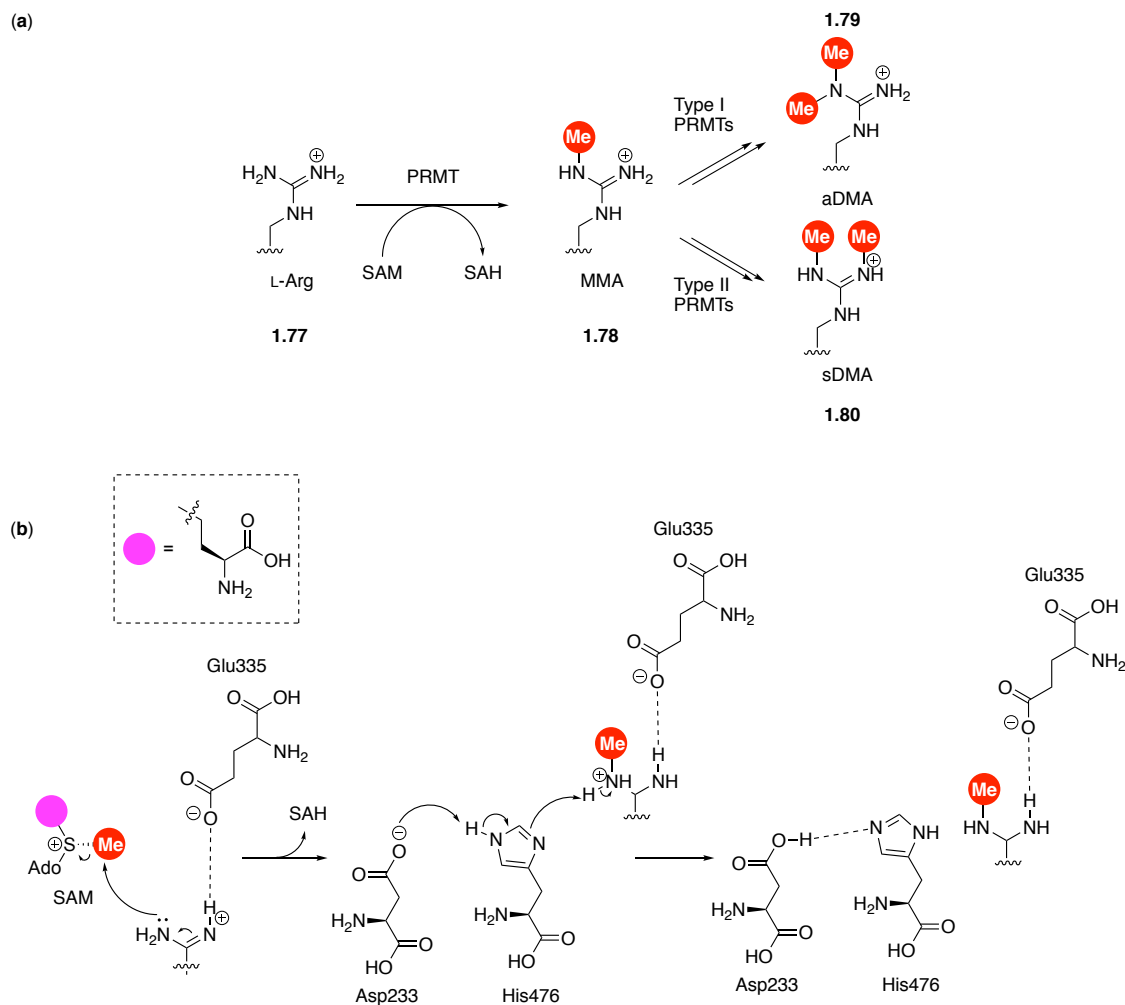
Scheme 1.23a. Methylation of arginine by protein arginine N-MTases, first to mono-methylated arginine (MMA) then on to either asymmetric or symmetric dimethylated arginine (aDMA/sDMA respectively). **b.** Putative mechanism of arginine methylation for PRMT3. Glu335 hydrogen bonds to arginine to position it towards the methyl group of SAM. Deprotonation of nitrogen is then carried out by a proton shuttle from Asp233 to His476. Loss of the Asp233 proton to bulk solvent then allows for further methylation to take place.[68]
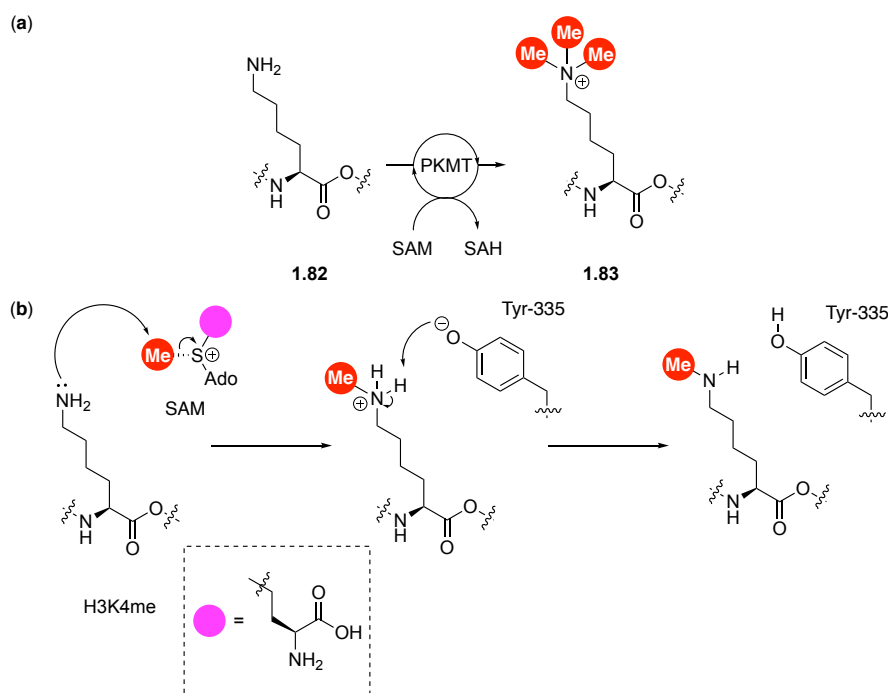
Scheme 1.24a. Methylation of lysine carried out by PKMTs. Common post translational modification of histones. **b.** Mechanism of methylation of histone H3K4 with Tyr335 residue acting as general base to deprotonate lysine, allowing for potential further methylation.[69]

1.6.3 *S*-methyltransferases

The role of S-MTases within the body appears to be exclusively within the detoxification process, with significant amounts of such enzymes located in the liver and the kidneys.[45] They are known to metabolise a number of different therapeutic compounds such as captopril (**1.84**) to give methylated derivative (**1.85**) (Scheme 1.25), and thiopurines.[46-48] They are also present in plants with involvement in the biosynthetic pathways of volatile sulfur compounds which act as a defense mechanism against certain insects.[70]
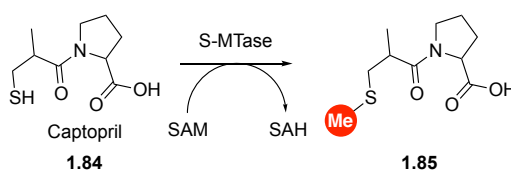
Scheme 1.25. Methylation of captopril as part of its metabolism within the body.[46] Precise mechanism of action unknown, but likely to involve positioning of thiol in close proximity to methyl group of SAM. An active site base may remove the thiol H and thus increase the nucleophilicity at this position. Subsequent nucleophilic attack by sulphur on the methyl group will provide the S-methylated captopril.

1.6.4 *C*-methyltransferases

C-MTases are less well known and understood than their O, N and S counterparts but there are still a number which have undergone study.[71] Again, they are a diverse family, with substrates ranging from small molecules to DNA. One of the most prominent examples is the HhaI MTase which methylates cytosine (**1.86**) in DNA as a means of protecting the host genome from foreign DNA and the endonuclease enzymes that digest non-methylated DNA.[49,72, 73] This occurs through initial thiolation at the C-6 position by an active site cysteine residue followed by methylation at the C-5 position with SAM donating the methyl group. Re-aromatisation then occurs at the pyridine ring by concomitant deprotonation and dethiolation (**1.87**) (Scheme 1.26).
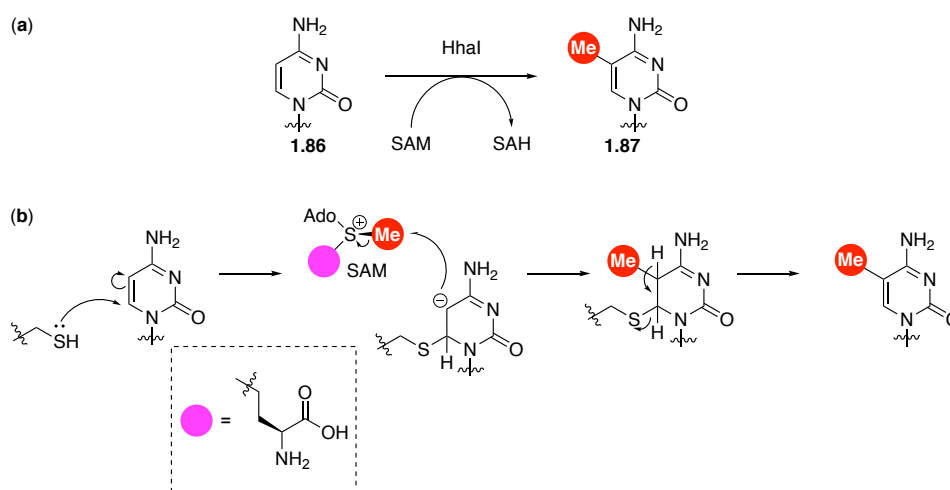
Scheme 1.26a. Methylation of cytosine by HhaI methyltransferase. **b.** Mechanism of methyl transfer to C-5 of cytosine by HhaI. Initial thiolation by an active site cysteine residue followed by methylation, then deprotonation and dethiolation.

1.6.5 Aromatic C-MTases

There are a number of examples of C-MTases acting on aromatic molecules such as in the biosynthesis of the actinomycin chromophore (**1.90**) (Scheme 1.27).[74] This compound is a known intercalator of DNA and has been utilised as a chemotherapy medication.[75, 76] This step in the biosynthetic pathway occurs by AcmI/L mediated methylation of 3-hydroxy-D-kynurenine (**1.88**) utilising SAM as a methyl donor to give 3-hydroxy-4-methyl-D-kynurenine (**1.89**). Although the precise mechanism is not known, it has been proposed that the ortho methylation is facilitated by deprotonation of the adjacent phenolic proton thus increasing the nucleophilicity of the site for methylation.[74]
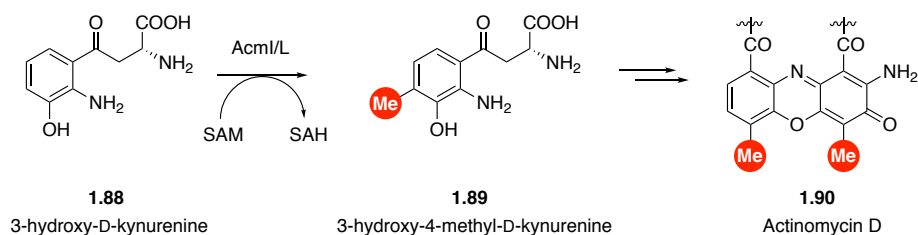
Scheme 1.27. Methylation of 3-hydroxy-D-kynurenine to give 3-hydroxy-4-methyl-D-kynurenine on the biosynthetic route to Actinomycin D.

1.6.5.1 C-methyltransferases CouO and NovO

CouO, from *Streptomyces rishiriensis,* and NovO, from *Streptomyces spheroides,* are two small molecule MTases involved in the biosynthesis of the aminocoumarin antibiotics coumermycin (**1.91c**) and novobiocin (**1.92c**) respectively (Figure 1.4). CouO and NovO have previously been studied by Walsh *et al.* with the aim of determining the stage at which methylation occurs in the biosynthetic route. For both coumermycin and novobiocin, the 8-methyl group is introduced by the respective MTase prior to functionalisation of the 7-hydroxy moieties.[50]
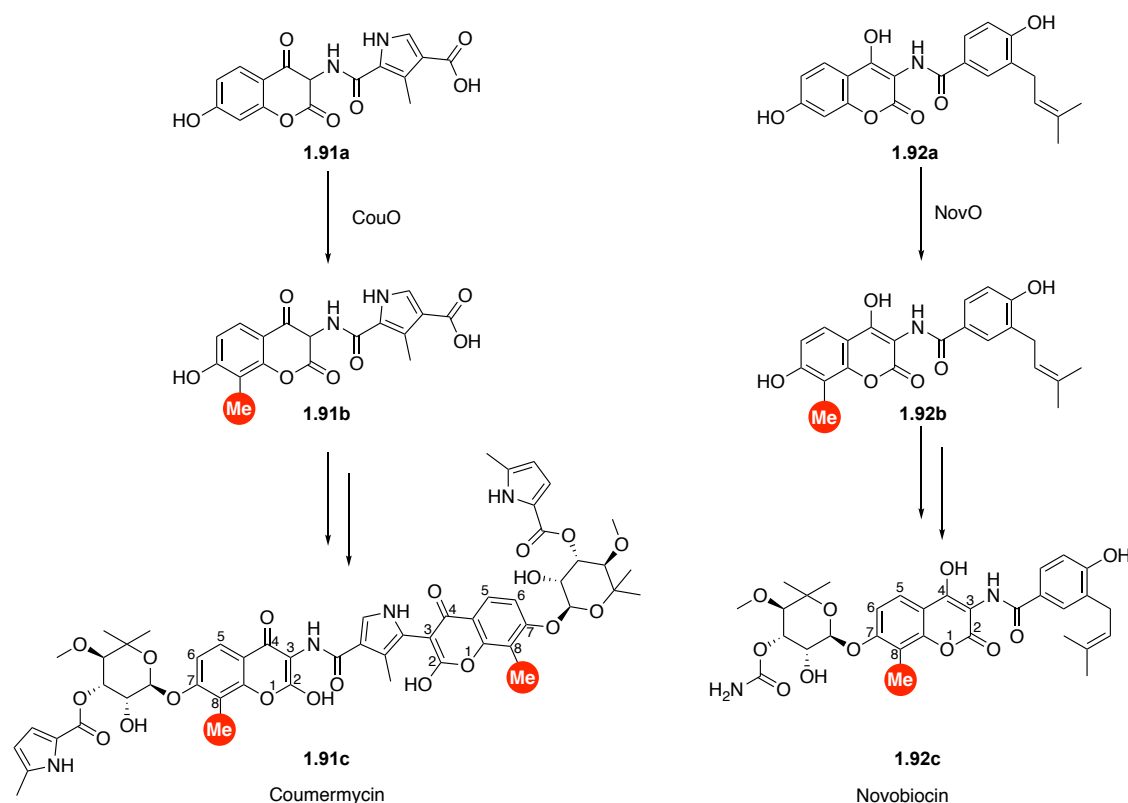
Figure 1.4. Structures of coumermycin and novobiocin with Me groups installed at the 8 position *via* the MTase enzymes CouO and NovO respectively.

Both NovO and CouO have undergone significant further studies into their activities and mechanism of action.[77] A homology model of NovO was reported by Gruber *et al.* in an attempt to rationalise the resulting activity of C-methyl transfer.[78] The SAM binding site contains the motif DLCCGSG (which has high similarity to Motif 1 DXGXGXG proposed by Martin and McMillan[55]) across residues 45-51. The importance of these residues for methyl transfer activity was determined by generation of the mutants Asp45Asn and Gly49Ala, which then showed vastly reduced k_{cat} values. It was then proposed that Asp70 and Asp96 may provide the acidic loop described previously as motif II. These regions were assumed to bind to the cofactor, whereas His15 and Tyr184 were highlighted as potential catalytic residues, with "putative" residue His15 in particular proposed to act as a catalytic base.[78]

Further studies carried out by Sadler *et al.* have since proposed an alternative mechanism of action of NovO.[77] An X-ray crystal structure was obtained for selenomethionine-substituted NovO in complex with SAH, which identified a number of key structural features. An extensive hydrogen bonding network exists, anchoring the substrate within the active site (Figure 1.5). Arg116 and Tyr25 interact with the SAH carboxylate, Glu4 and Asp96 with the free amino group of adenine (Ade), and Asp70 with the two ribose-hydroxyl groups. Backbone amide NH bonds from Ala97 and Leu71 interact with N1 and N3 of Ade respectively. Finally, Cys47 and Arg116 interact with the aliphatic amine of SAH through their backbone carbonyl groups.
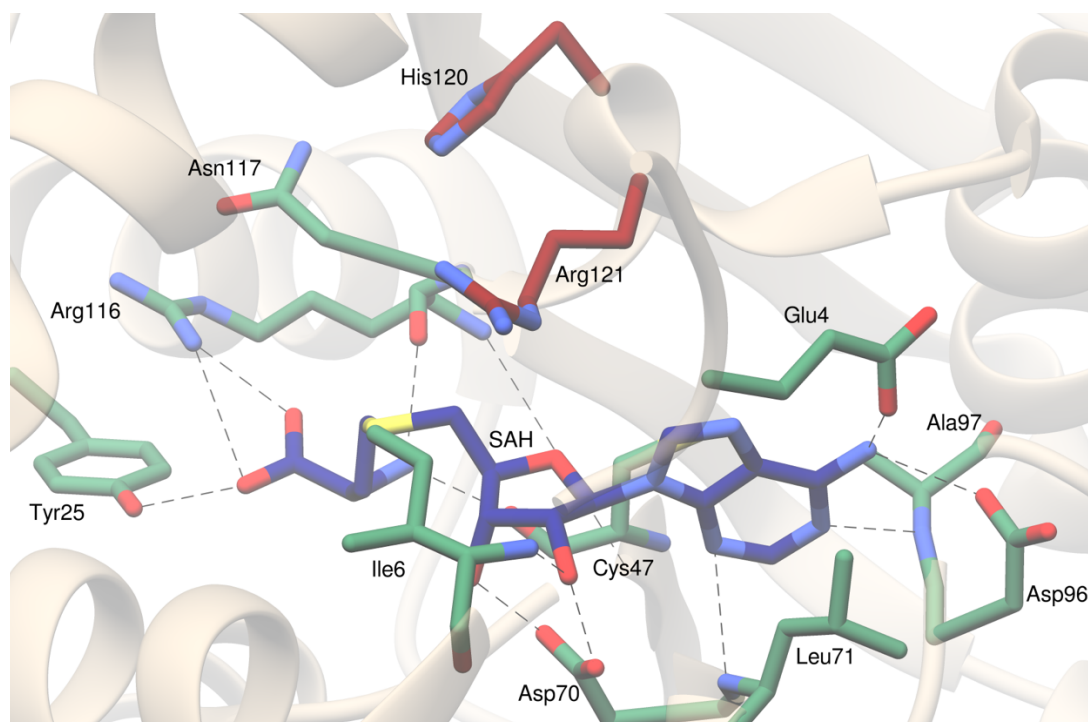
Figure 1.5. Key residues around the SAH (navy) binding site of SelMet NovO highlighted (sea green). His120 and Arg121 which are important in catalytic mechanism of NovO are highlighted in maroon.[77]

A model of NovO with aminocoumarin substrate (**1.93**) bound was designed based on the SelMet NovO structure, with a number of putative active site residues proposed (Figure 1.6). His120 and Arg121 hydrogen bond to the phenolic group at position-7 of the coumarin while Arg116 and Asn117 are in close contact to the lactone motif. Mutagenesis of these positions (Arg116Leu, Asn117Ala, His120Ala and Arg121Leu) showed complete abolition of activity, suggesting that they are indeed key residues for enzyme function.

Two further single point mutations were carried out at His120 and Arg121 to probe their role in enzyme activity. It was proposed that His120 could act as a base, thus deprotonating the phenolic group at position-7 while protonated Arg121 could H-bond to the phenolic group. His120Lys activity was tested as it would be protonated at under the neutral conditions of the assay and therefore unable to act as a base. No activity was seen for this mutant. Arg121Lys was generated and was expected to retain H-bonding to the phenolic group. There was a decreased conversion from 100% with the WT to ~ 20%. The retention of some activity is in agreement with the proposed function of the residue. These results are in contrast to the work of Gruber e*t al.* with their proposed catalytic base, His15, observed at the solvent interface of SelMet NovO rather than in the binding pocket.[78]
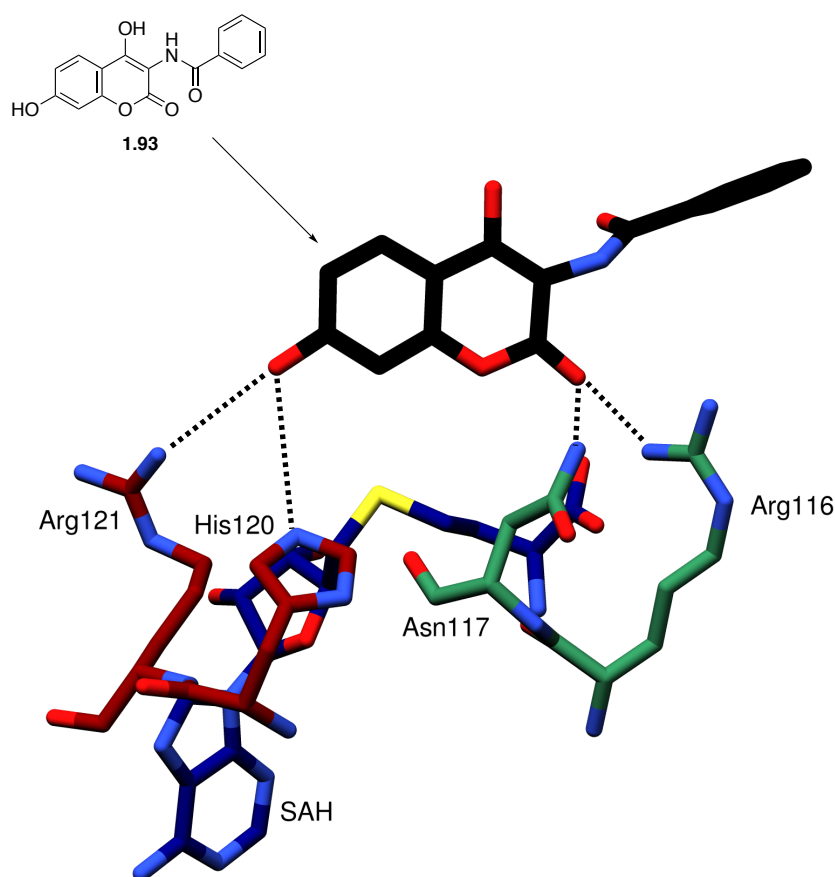
Figure 1.6. Putative key residues in methylation of aminocoumarin **1.93** (black). Arg116 and Asn117 (green) contribute to substrate binding while His120 acts as a catalytic base to deprotonate 7-OH thus increasing nucleophilicity at adjacent C8.

The mechanism proposed from this work was as follows (Scheme 1.28): Aminocoumarin (**1.93**) is bound through interaction of the C-2 carbonyl with Arg116 and Asn117. His120 acts as the catalytic base, deprotonating the phenolic OH at position-7, which is then stabilised by H-bonding to protonated His120 and Arg121. The nucleophilicity gained through this deprotonation allows for electrophilic aromatic substitution to occur at the 8-postion with a methyl group donated from SAM. Pavlov-Keller *et al.* solved the crystal structure of CouO and identified an almost identical His120-Arg121 motif that they believed was acting in the same fashion as NovO.[79]
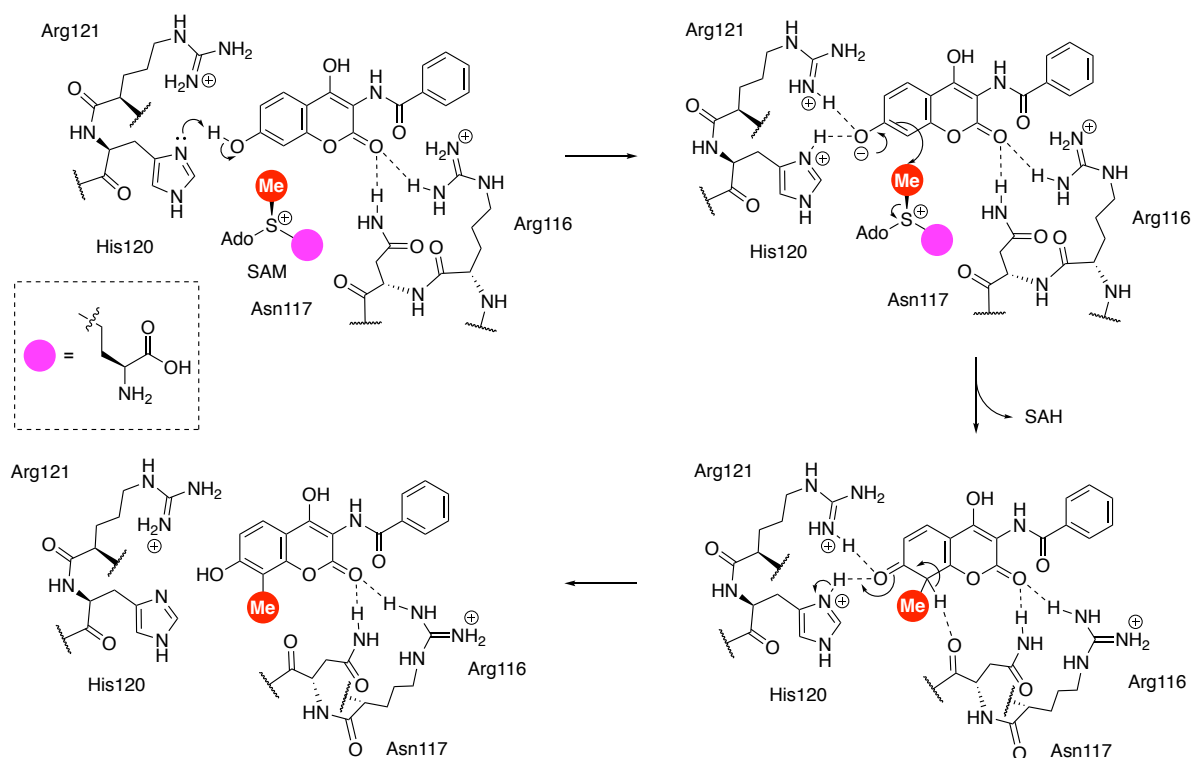
Scheme 1.28. Proposed mechanism of C-methylation of substrate **1.93** by NovO.

Both NovO and CouO have also been shown to undergo C-methylation on non-natural substrates. Of particular interest from a synthetic chemistry perspective are a series of naphthalene derivatives (**1.94-1.96**) (Figure 1.7).[80] Further to this, it has also been shown that NovO and CouO are active towards alkylation of coumarin substrate (**1.97**) using chemically synthesised non-natural SAM analogues (**1.98a-e**) (Scheme 1.29). In each case only mono-substitution was observed even in the presence of a large excess of cofactor.[80]
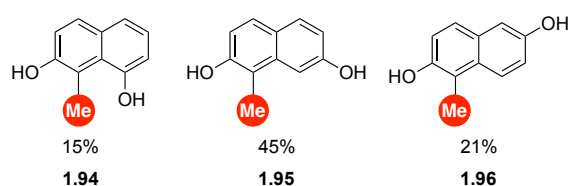
Figure 1.7. Methylation of non-natural dihydroxynaphthalene substrates of CouO.[80]
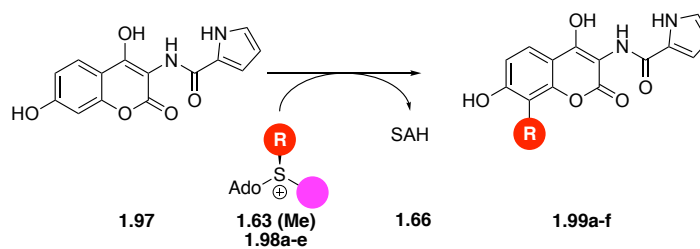
SAM analogue No.	R	R—coumarin No.	% Conversion with CouO	% Conversion with NovO
1.63	-ξ-Me	**1.99a**	>99	>99
1.98a	⌇⟋═	**1.99b**	>99	>99
1.98b	⌇⟋═	**1.99c**	38	42
1.98c	⌇⟋≡	**1.99d**	99	99
1.98d	⌇⟋≡	**1.99e**	77	41
1.98e	⌇⟋⌬	**1.99f**	45	40

Scheme 1.29. Transfer of groups other than methyl from the corresponding analogue of SAM (**1.98a-e**).[80] RP-HPLC conversions of alkyl transfer achieved by CouO and NovO on substrate **1.97**, giving products **1.99a-f**.[80]

1.7 Problems in Using SAM as a Methyl Donor

Problems with the chemical synthesis of SAM have limited advances in utilisation of MTases as effective strategies for introduction of alkyl groups.[81] The current synthetic route to SAM analogues utilises SAH (**1.66**), with an alkylation carried out using the appropriate alkyl bromide (**1.100**) in formic acid with silver triflate as a catalyst (Scheme 1.30).[80, 82] This method produces both the "*S*" and "*R*" epimers at the sulfonium centre, with only the "*S*" being accepted by a number of MTases while the "*R*" epimer can act as an inhibitor in certain MTases.[83] Another drawback to employing a biocatalytic platform for the alkylation of small molecules is the cost and purity of commercial SAM (75% purity, £41.30 for 5 mg, Sigma Aldrich 29/08/19) and its inherent instability under neutral or alkaline conditions ($t_{1/2}$ = 942 min, pH 8).[84]
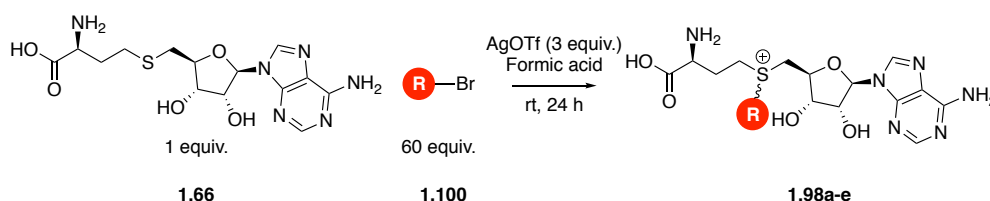
Scheme 1.30. Synthesis of SAM analogues by alkylation of SAH with R-Br where R = an alkyl chain or ring (Table 1.1).

To allow MTases to be more widely used as a method of alkylating small molecules, it is necessary to develop new methods of SAM synthesis which can circumvent some of the problems stated above. An attractive strategy would be the use of another enzyme to generate the correct SAM epimer from readily available starting materials *in situ*. SAM could then be used directly without the need for purification or long-term storage. Three viable candidates for this role are SAM generating enzymes Methionine Adenosyl Transferase (MAT), 5′-fluoro-5′-deoxyadenosine synthase (FDAS) and SalL (chlorinase from *Salinospora tropica*) (Scheme 1.31).
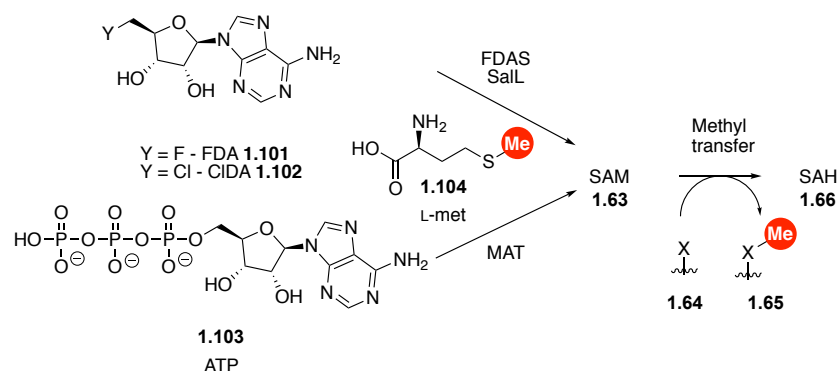
Scheme 1.31. Use of SalL or MAT with appropriate starting materials as a source of in situ SAM for subsequent methyl
transfer.

1.8   Enzymes that Catalyze the Formation of the SAM Cofactor

1.8.1   Methionine Adenosyltransferase (MAT)

The MAT enzyme has potential utility in a two-enzyme SAM generation/MTase strategy. Prevalent in nature, MAT catalyses the reaction between ATP (**1.103**) and L-methionine (**1.104**) to produce SAM (**1.63**) (Scheme 1.32a). Komoto *et al.* elucidated the mechanism of MAT mediated SAM formation in 2004 (Scheme 1.32b). They found the reaction occurs in two steps: initial displacement of the triphosphate group by L-met to form SAM, followed by hydrolysis of the triphosphate group into orthophosphate (**1.106**) and pyrophosphate (**1.107**) which then facilitates release of SAM. Their work hinged on use of 5′-adenylyimido-triphosphate (**1.105**) (AMPPNP) in crystallisation studies as this known inhibitor of ATP dependent enzymes cannot undergo the second hydrolysis event (Scheme 1.32c).[85, 86] They observed SAM and PPNP (**1.108**) trapped in the active site. From this they deduced that the second hydrolysis releases energy which allows SAM to leave the active site, without this energy SAM remains trapped in the active site.[85]
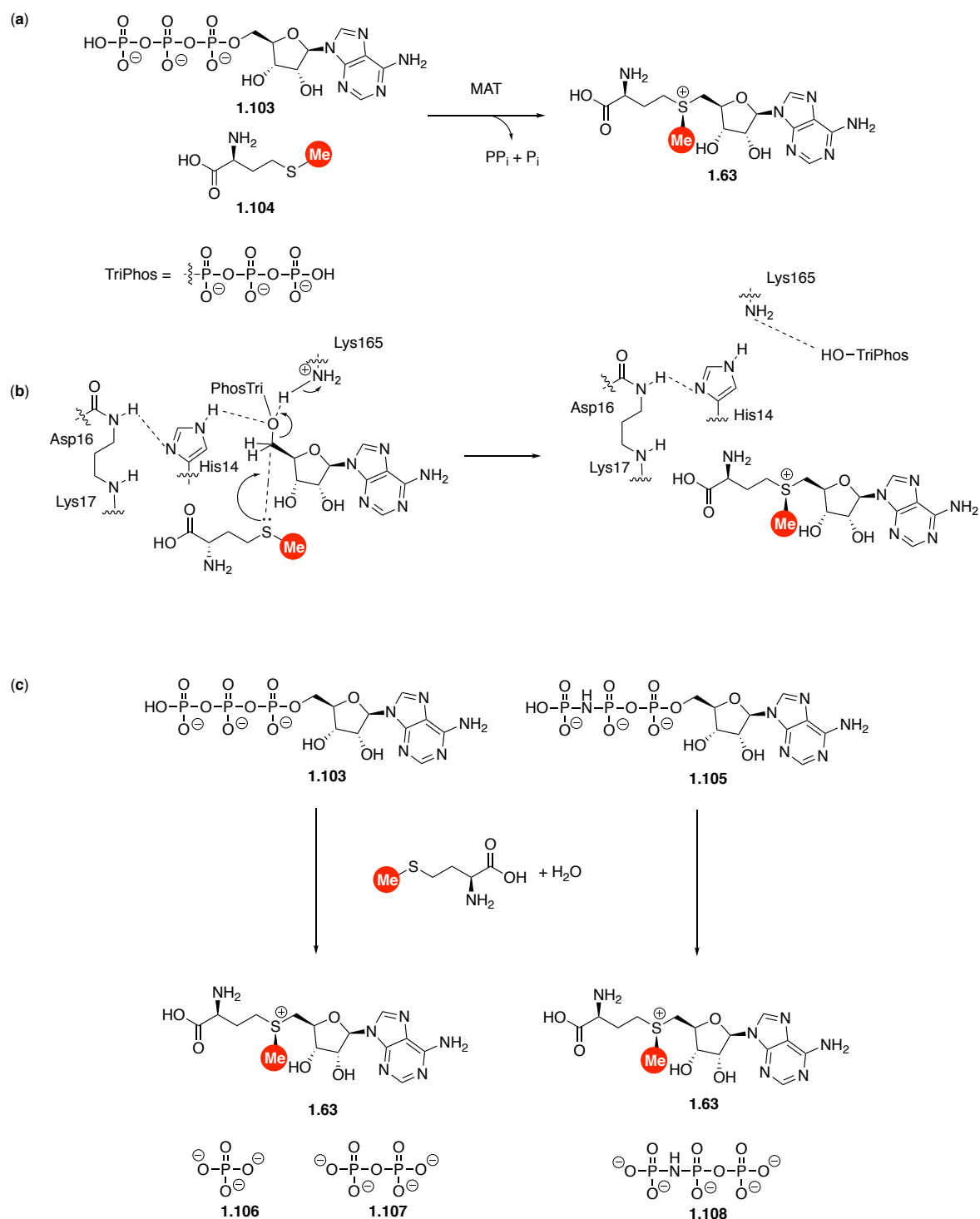
Scheme 1.32a. Reaction of methionine with ATP to give SAM catalysed by MAT. **b.** Mechanism of SAM formation from ATP and L-met as catalysed by MAT. L-met carries out S_N2 attack at the 5′-carbon, thus displacing the triphosphate and forming SAM. (**c**) ATP (**1.103**) is converted to SAM with hydrolysis of the triphosphate to **1.106** and **1.107**, releasing energy which facilitates SAM flux from active site. AMPPNP (**1.105**) forms SAM under the same conditions but the modified triphosphate (**1.108**) no longer undergoes hydrolysis thus trapping SAM in the active site.

A study carried out by Singh *et al.* investigated the use of MAT for the *in situ* generation of SAM and analogues utilising both L-met and L-selenomethionine (L-Se-met) derivatives (Scheme 1.33).[83] Seleno-analogues were investigated as SeSAM is considered to be a better alkyl donor than SAM because of the increased length and decreased strength of the Se-C bond with respect to the S-C bond.[87] They first screened a series of MATs from bacterial, archael and mammalian sources for activity towards SAM formation.[83, 88] hMAT2A was determined to have the best combination of broad specificity and high reactivity, displaying activity towards generation of a number of unnatural ethyl, allyl, propargyl and cyano SAM analogues. For this reason, hMAT2a was taken forward into a two-enzyme process.

Entry	R	S	Se
a		90	53
b		50	42
c		6	44
d		13	23
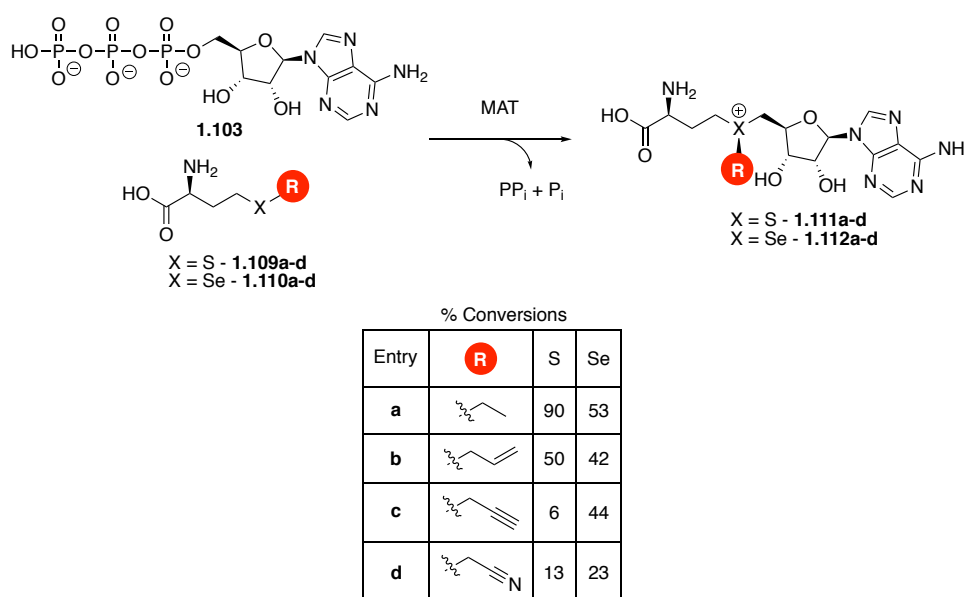
Scheme 1.33. MAT catalysed formation of SAM analogues (**1.111a-d/1.112a-d**) from ATP (**1.103**) and L-met /L-Se-met analogues (**1.109a-d**/**1.110a-d**).

Coupling of the two-enzyme SAM generation/alkylation process to the rebeccamycin MTase RebM was carried out in an attempt to introduce a variety of groups to the C4′-O of rebeccamycin (**1.113**) to give compound **1.114** (Scheme 1.34). Perhaps of greatest interest was the introduction of an acetonitrile group, the first carried out by a MTase.[83] In this study, SAH caused some issues with inhibition of the MTase, but this was circumvented by addition of an *S*-adenosyl-L-homocysteine hydrolase (SAHH) enzyme to catalyze breakdown of SAH to L-homocysteine (**1.115**) and Ado (**1.116**). This resulted in a significant increase in yield from 15-40% in each case. As such, inclusion of an enzyme such as SAHH could form a key part of any industrial MTase cycle.

Entry	R	% Conversion S	% Conversion Se
a	⌇⌇ (propyl)	40	10
b	⌇⌇ (allyl)	67	33
c	⌇⌇ (propargyl)	5	55
d	⌇⌇ N (cyanoethyl)	42	46
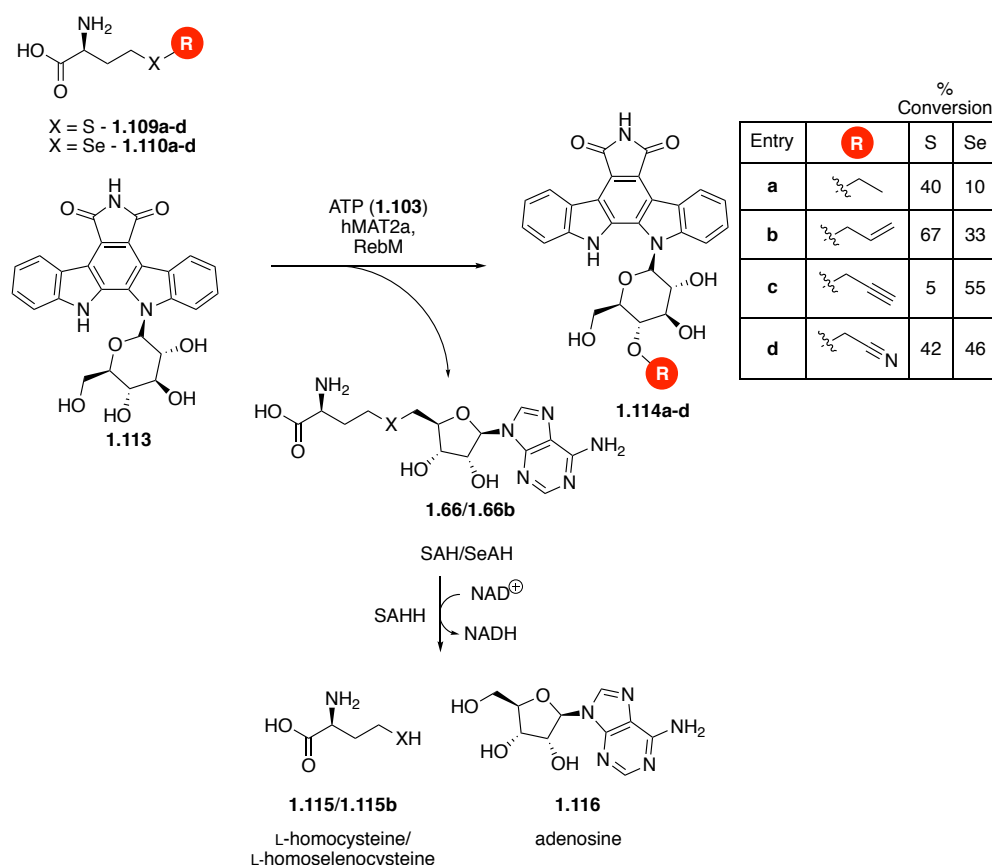
Scheme 1.34. One-pot transformation of L-met/L-Se-met (analogue) to SAM/SeSAM (analogue) with hMAT2a, then alkylation carried out by RebM with subsequent breakdown of SAH/SeAH by SAHH to prevent enzyme inhibition.

Despite the advantages of this approach in terms of broad substrate tolerance with respect to methionine analogues, this method has a key drawback. ATP is an expensive and readily hydrolysable substrate, needing to be stored at -20 ºC, and can readily degrade to either ADP or AMP. An advancement on this method would be to use an enzyme that accepts two stable and cost-effective substrates to facilitate SAM and SAM analogue formation.

1.8.2 5′-fluoro-5′-deoxyadenosine synthase (FDAS) from *Streptomyces cattelya*

5′-Fluoro-5′-deoxyadenosine synthase (FDAS) was isolated from *Streptomyces cattelya* and characterised in 2003.[89] This enzyme catalyses the reversible fluorination of SAM to generate 5′-fluoro-5′-deoxyadenosine (FDA) (**1.101**) as the first step of the biosynthetic pathway to fluoroacetate (**1.118**) and 4-fluorothreonine (**1.119**) (Scheme 1.35, Crystal structure, Figure 1.8a). In 2007, Zhu *et al.* displayed that the enzyme was active towards chloride as well as fluoride (Figure 1.8b).[90, 91] A Ser158Gly mutant with chloride bound revealed the location of

the halide binding pocket, while the wild type structure with L-met and FDA bound displayed the geometry of the products of the fluorination of SAM. The mechanism begins with binding of a fluoride ion by Ser158 and Thr80 with this anion solvated until tight binding of a SAM molecule excludes water from the active site. The fluoride anion is then trapped in an orientation which positions it for S_N2 attack to displace L-met and form FDA (Scheme 1.36).
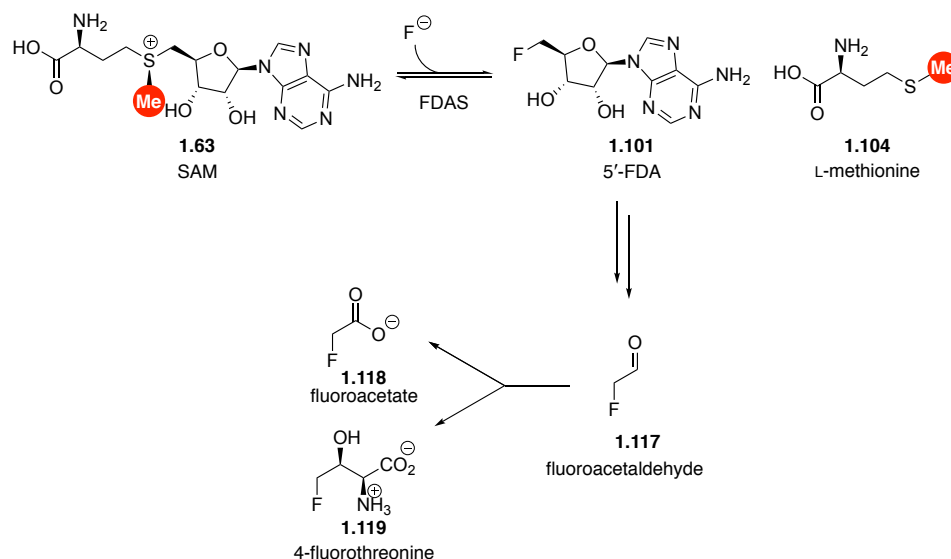
Scheme 1.35. Reversible fluorination of SAM to give FDA and L-met and subsequent biosynthesis of fluoroacetate and 4-flurothreonine.[92]
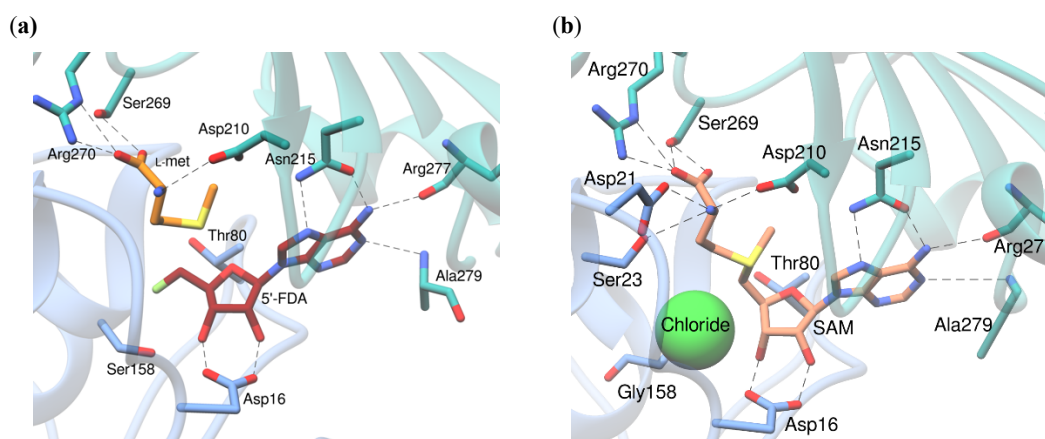
Figure 1.8a. Active site of FDAS with FDA (dark red) and methionine (orange) bound. Ser158 is key in hydrogen bonding with fluoromethyl group leading to the transition-state like orientation allowing for formation of SAM in the reverse direction (PDB Accession code: 1RQR). **b.** Active site of FDAS Ser158Gly mutant. This mutant structure displays the location of the halide binding site while maintaining the key hydrogen bonding contacts observed between enzyme and substrate in the wild-type structure (PDB Accession code: 2V7U).
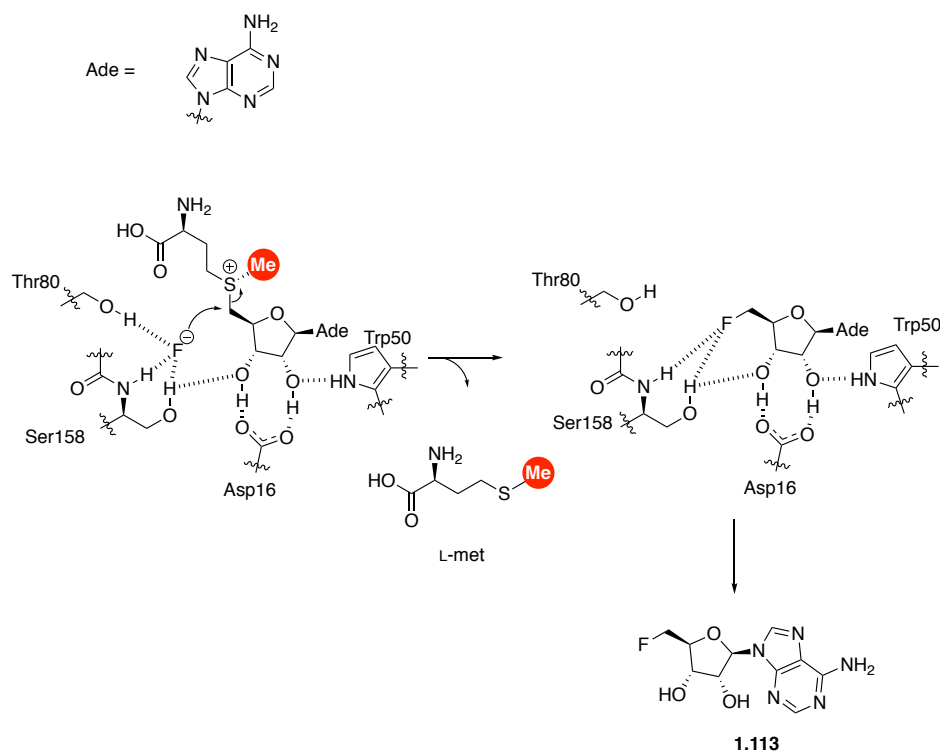
Scheme 1.36. Mechanism of FDAS-mediated fluorination of SAM.[90]

Thomsen *et al.* have shown the ability of FDAS to generate SAM and the ethyl analogue of SAM from stable FDA and either L-met or L-ethionine (L-eth). However, the turnover for this process was low with other comparable SAM forming enzymes displaying a greater propensity to generate SAM. Work by Deng *et al.* showed that the FDAS would accept chloride to form 5′-chloro-5′-deoxyadenosine (ClDA), but also perform this reaction in the reverse direction to form SAM.[91] The same group showed that transhalogenation from ClDA to FDA through a SAM intermediate was possible. L-Se-met was also found to be a substrate for FDAS, and transhalogenation could be carried out through a SeSAM intermediate (Scheme 1.37). FDAS has also been shown to accept 2′-deoxyClDA/2′-deoxyFDA to generate the corresponding SeSAM analogue (Scheme 1.38).[93]
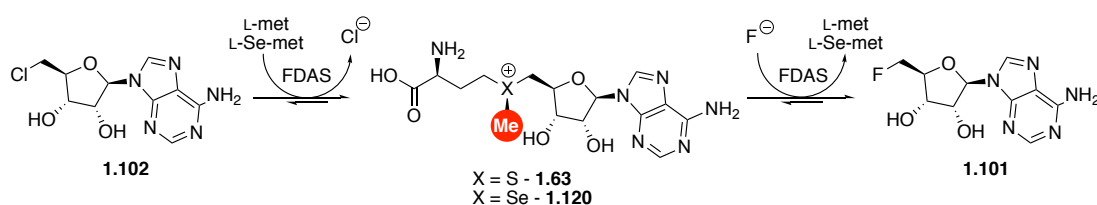
Scheme 1.37. Transhalogenation of ClDA to FDA mediated by FDAS through either a SAM or SeSAM intermediate.
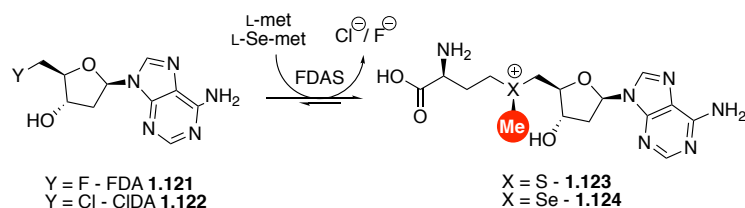
Scheme 1.38. Conversion of 2′-dClDA/FDA (**1.121**/**1.122**) to 2′-dSAM catalysed by FDAS. Conversion of 2′-ClDA and 2′-FDA to 2′-dSAM = 50%, 10% respectively, relative to natural FDA with L-met.[93]

Aside from modifications to the L-met region of the molecule, modifications to the nucleobase have also been extensively studied. Initially Thompson *et al.* explored the ability of the enzyme to carry out transhalogenation of ClDA modified with an alkyne at the 2-position, 5′-chloro-5′-deoxy-2-ethynyladenosine (ClDEA) (**1.125**) (Scheme 1.39). This enabled modification of a cyclic peptide through a PEG linker to the 2-alkyne (Scheme 1.40). Utilisation of such a strategy allowed incorporation of a ^{19}F radioisotope which was then used as a Positron Emission Tomography (PET) imaging agent.
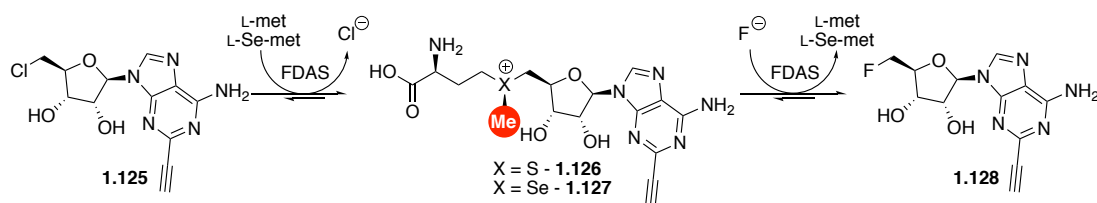
Scheme 1.39. Transhalogenation of ClDA analogue ClDEA (**1.125**) to 5′-fluoro-5′-deoxy-2-ethynyladenosine (FDEA) (**1.128**) can be carried out with either L-met or L-Se-met.
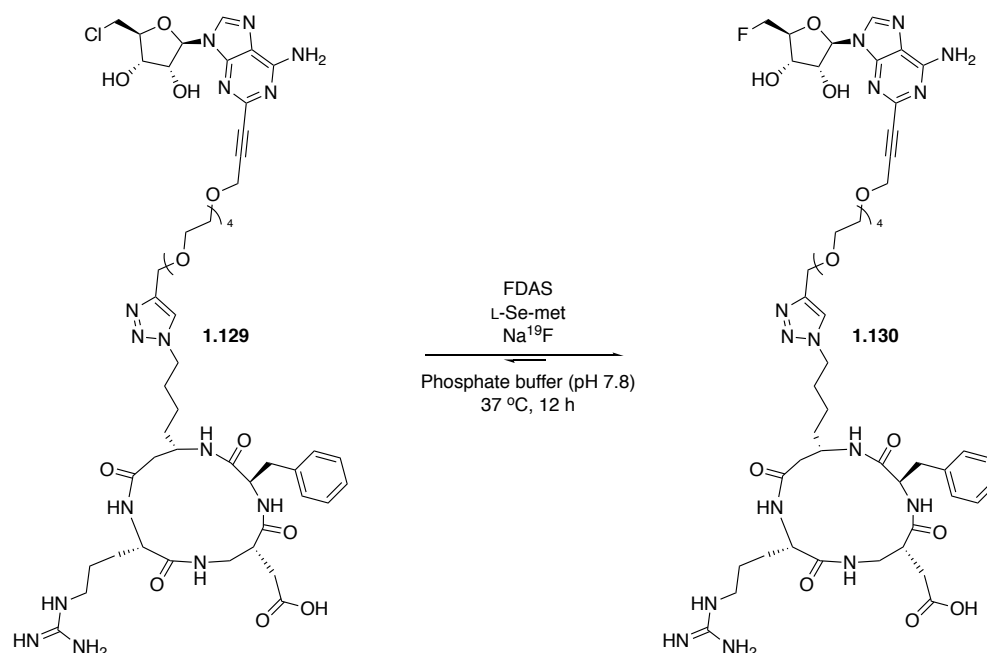
Scheme 1.40. Transhalogenation of ClDEA analogue (**1.129**) to FDEA analogue (**1.130**), which bears a macrocycle, can be carried out. When a ^{19}F labelled fluoride source is used, the macrocyclic peptide can be imaged using Positron Emission Tomography (PET).

This work was then used to probe the limits of modifications to the nucleobase which could be accepted by fluorinase enzymes. The group of Ang investigated a range of 2-modified ClDA analogues and whether or not they can be transhalogenated to the corresponding FDA analogue by a related fluorinating enzyme *flA1* (discovered by Deng *et al.*)[94] and a small panel of *flA1* mutants (Scheme 1.41).[95, 96] A number of unnatural *flA1* substrates have been identified, with compounds **1.131b** and **1.131d** showing increases in turnover of 4.36 and 2.30, relative to ClDA itself (Figure 1.9).
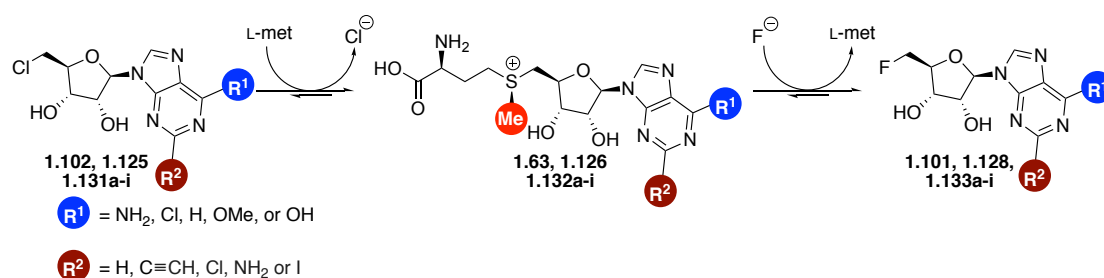
Scheme 1.41. Transhalogenation of ClDA analogues to FDA analogues through a SAM analogue intermediate catalysed by fluorinase *flA1*.
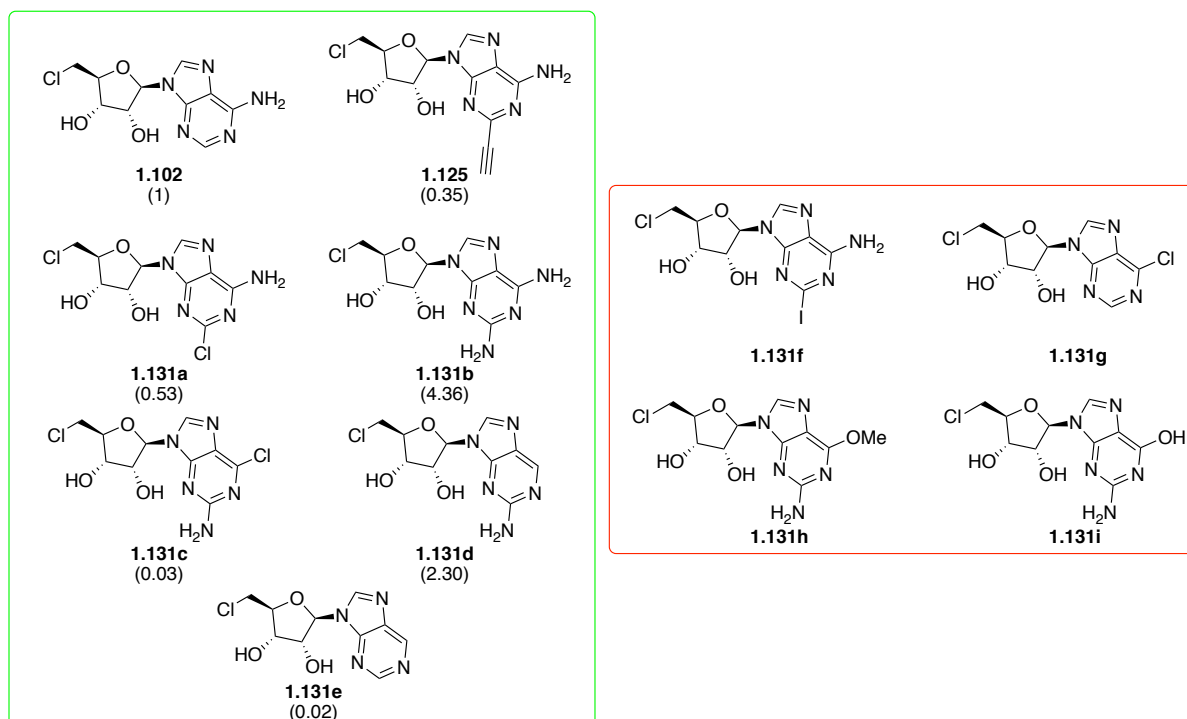
Figure 1.9. ClDA analogues which can be converted to FDA through a SAM intermediate by *flA1* (green) and substrates which show no activity (red). Conversions relative to ClDA given in parenthesis.

1.8.3 SalL from *Salinospora tropica*

In 2008, Eustaquio *et al*. discovered SalL, a homotrimeric enzyme derived from marine organism *Salinospora tropica*.[97] SalL catalyses the reversible chlorination of SAM (**1.63**), with delivery of chloride as a nucleophile to form ClDA (**1.102**) and L-met (**1.104**), in a fashion analogous to FDAS catalysed fluorination (Scheme 1.36).[97] This chlorination forms an early part of the biosynthetic pathway to Salinosporamide A (**1.133**), a compound which displays cytotoxicity towards a variety of cancer cell lines (Scheme 1.42).[98]
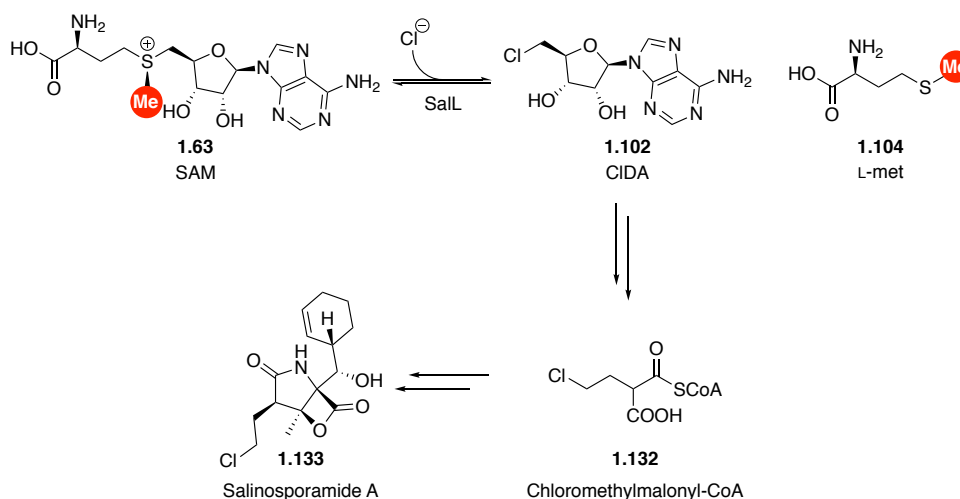
Scheme 1.42. Reversible chlorination of SAM to give ClDA and L-met and subsequent biosynthesis of anti-cancer agent Salinosporamide A (**1.133**).

SalL shares 35% homology with FDAS but displays some key structural differences in its halide binding site. Thr80 and Ser158 are replaced by Tyr70 and Gly131, respectively, in SalL and these changes do not allow SalL to act as a fluorinase, although it can accept bromide and iodide as nucleophiles to displace methionine from SAM. Significant conserved residues between the two structures are an active site Asn(215/188 in FDAS and SalL respectively) which has a hydrogen bonding interaction with both the N7 and exocyclic amine of Ade and Asp(16/11 in FDAS and SalL respectively) which hydrogen bonds to both the 2′- and 3′-hydroxyls of the ribose unit (Figure 1.10).
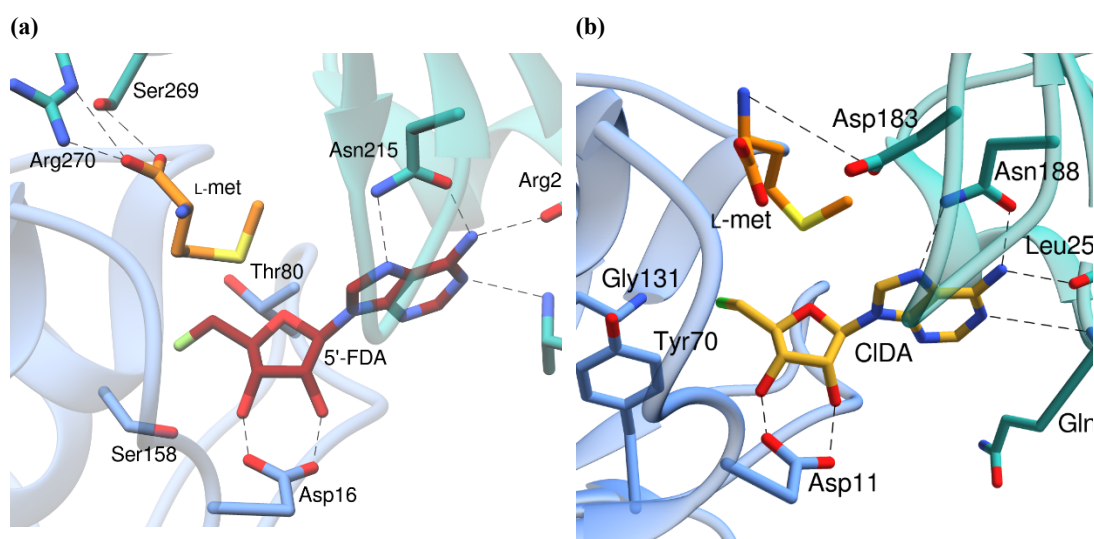
Figure 1.10a. FDAS structure for comparison with SalL structure. **b.** X-ray structure of SalL with ClDA and L-met bound, key residues highlighted. Asn188, Gln252 and Leu250 appear to H-bond to the Ade ring while Asp11 H-bonds to the ribose OH groups, thus positioning ClDA (gold) for L-met (orange) to attack and displace Cl⁻.

Structures of these SalL mutants were deposited on the Protein Data Bank (PDB) alongside the wild-type structure in Figure 1.12b and a structure where Ado is the ligand. These structures are based on crystals of the monomeric unit of SalL rather than the intact trimer. Any rational mutagenesis based on these structures may prove difficult as the active site of the enzyme lies at the monomer-monomer interface. For this reason, any structure-guided programme of mutagenesis would benefit from crystallisation of an intact trimer of SalL.

In 2013, Burkart *et al.* explored the use of SalL in an effort to produce SAM analogues.[99] This process was carried out by addition of excess methionine and an absence of chloride that is present in the native system (Scheme 1.43). The mechanism through which SalL catalyses SAM formation presumably occurs through attack by L-met at C-5′, with displacement of Cl⁻ *via* an S_N2 type process (Scheme 1.44).
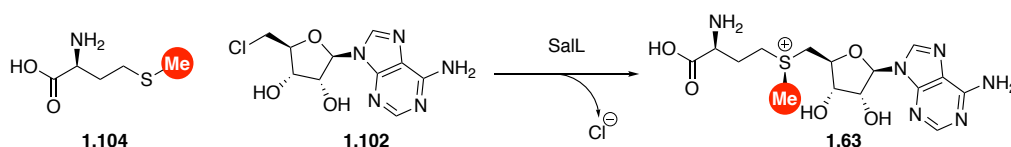
Scheme 1.43. SalL mediated reaction of ClDA with methionine to form SAM.
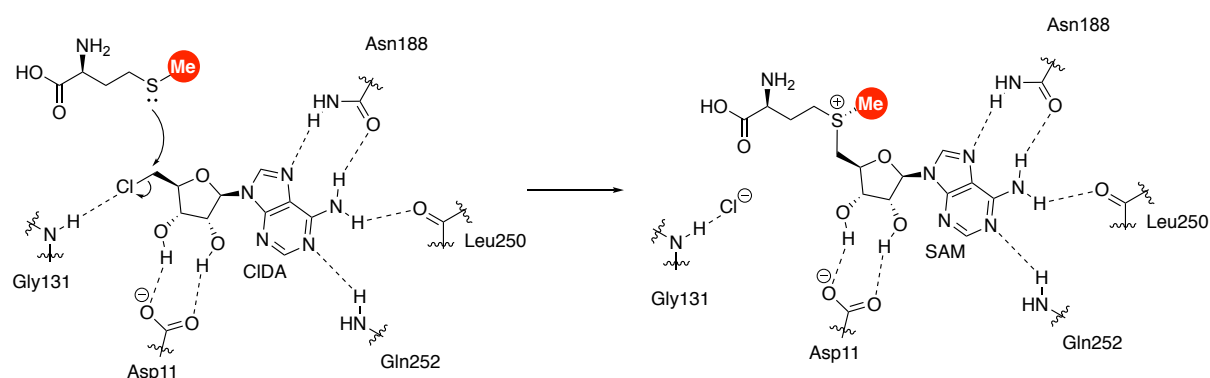
Scheme 1.44. Proposed mechanism of SAM formation catalysed by SalL.

Further work from Burkart also showed it was possible to couple ClDA (**1.102**) to different alkylated methionine analogues (**1.109**) to generate SAM analogues (**1.111**) (Scheme 1.44), with the activity for these reactions far greater for SalL than for FDAS.[100] Substrates tested included ethyl, propyl, butyl, allyl and benzyl analogues, with activity of the wild type enzyme decreasing with increasing size of the methionine analogue (Scheme 1.45). SDM of SalL was carried out as a number of active site residues were targeted to reduce steric bulk around the

bound substrates. A Trp190Ala mutant displayed a slightly increased turnover of butyl methionine of 0.934 nM/min compared to 0.318 nM/min for the wild-type.
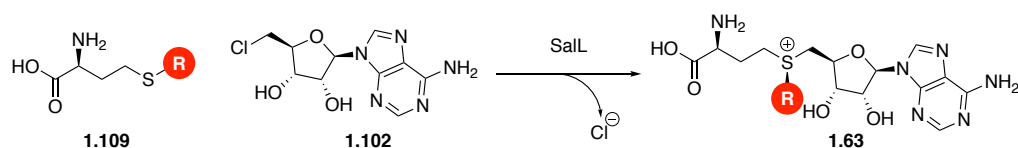
SAM analogue No.	R	Turnover (nM/min)
1.63	⌇-Me	4930
1.111a		453
1.111b		1.02
1.111e		1.01
1.111f		0.318
1.111g		0.101

Scheme 1.45. SAM analogue formation from ClDA and L-met analogues catalysed by SalL and activity of SalL towards formation of SAM analogues.

Furthermore, Burkart *et al.* demonstrated the utility of this method by using the SAM generated by SalL in conjunction with a HHal MTase to methylate DNA (**1.134-1.135**) and with MTase MtfA to install an N-methyl group on the antibiotic teicoplanin (**1.136**) to give N-methylteicoplanin (**1.137**) (Scheme 1.46). Significantly, this synthesis showed that SalL and the resulting SAM generation was compatible with both the HHal and MtfA enzymes.[99]
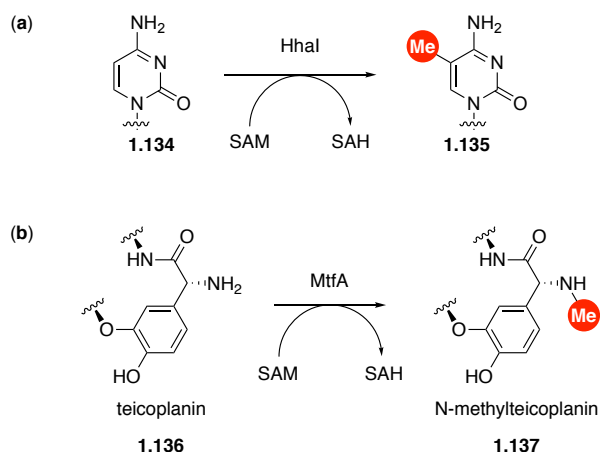
Scheme 1.46a. Methylation of cytosine carried out by HhaI MTase as discussed previously (Scheme 1.26) with SAM generated *in situ* by SalL. **b.** Methylation of teicoplanin by MtfA with SAM generated from SalL. Mechanism likely to proceed through attack of NH$_2$ group on SAM methyl with subsequent deprotonation carried out by an active site basic residue, proposed to be His228.[101]

This approach demonstrates that SalL can be used to generate SAM analogues and is compatible with MTases. However, the low turnovers associated with L-met analogues bulkier than ethyl severely limits the utility of this method for the introduction of alkyl chains *via* a two-enzyme SAM generation/MTase tandem reaction. Most recently, Sadler *et al.* have used SalL in tandem with MTase NovO to alkylate coumarins.[102] This methodology has facilitated CD_3, $^{13}CH_3$ and $^{13}CD_3$ labelling of a set of coumarins (**1.93**, **1.97**) and dihydroxynaphthalenes such as **1.138** (Scheme 1.47). This platform could perhaps be developed further by mutagenesis of the SalL active site to engineer additional space in the active site which could then accommodate methionine analogues bearing large alkyl chains.

X = O or C
Y = H or OH
Z = H, or

% Conversions

	1.139a	1.99a	1.95
R			
CH_3	97	100	65
$^{13}CH_3$	95	100	94
CD_3	88	100	91
$^{13}CD_3$	90	100	92
CH_2CH_3	34	30	7
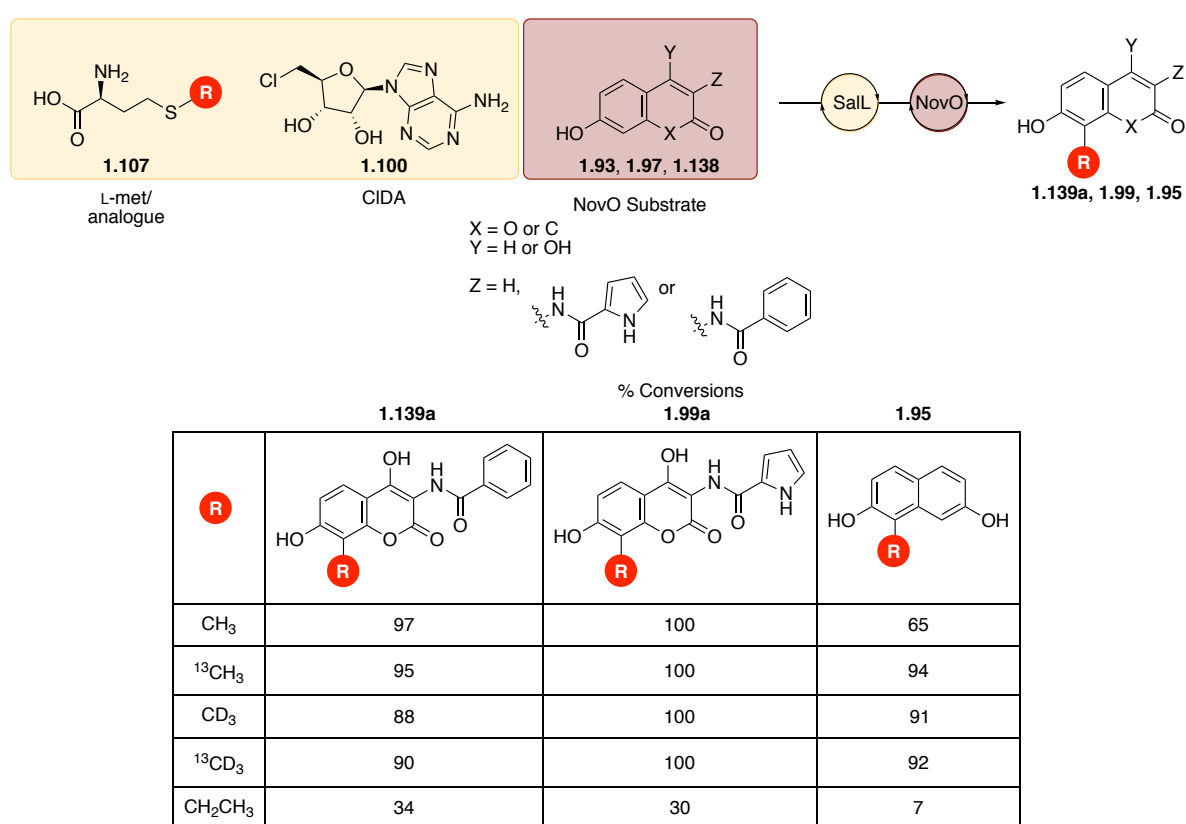
Scheme 1.47. Tandem SAM analogue generation/alkylation process mediated by SalL and NovO.[102]

1.9   Enhancing SAM stability

SAM has two primary degradation pathways: at acidic pH – depurination to form *S*-pentosylmethionine (**1.140**) and Ade (**1.141**); at neutral or basic pH – cyclisation to form 5′-methylthioadenosine (MTA) (**1.142**) and homoserinelactone (HSL) (**1.143**); (Scheme 1.48).[103]
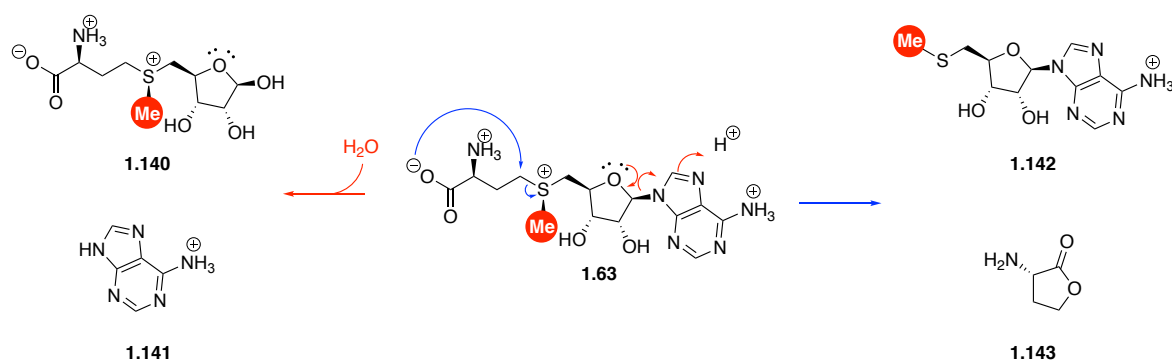
Scheme 1.48. Depurination of SAM under acidic conditions to form pentosylmethionine and Ade (red arrow). Degradation of SAM to MTA and HSL under neutral or basic conditions (blue arrow).

Huber *et al.* devised a strategy of SAM synthesis using MAT and derivatives of ATP and L-met as their starting materials. Literature precedent suggests that removing the N7 from Ade drastically reduces rates of depurination.[104-106] In addition, they proposed that the cyclisation degradation pathway (Scheme 1.48) could be suppressed by substitution of the carboxylate of SAM with a tetrazole group, as these have been reported to be less nucleophilic,[107] while maintaining a similar pKa.[108] They envisioned that combining these two strategies would lead to a stable SAM analogue (**1.146**) (Scheme 1.49, Figure 1.11) which would be similar enough to natural SAM to be accepted by MTases. The ability of MAT to carry out synthesis of 7-dz-*tet*-SAM was assessed alongside synthesis of *tet*-SAM (**1.147**) and 7-dz-SAM (**1.148**). Synthesis of *tet*-SAM was achieved with 50% conversion in a non-optimised reaction with 7-dz-SAM and 7-dz-*tet*-SAM synthesised with conversions of 92% and 57% respectively.
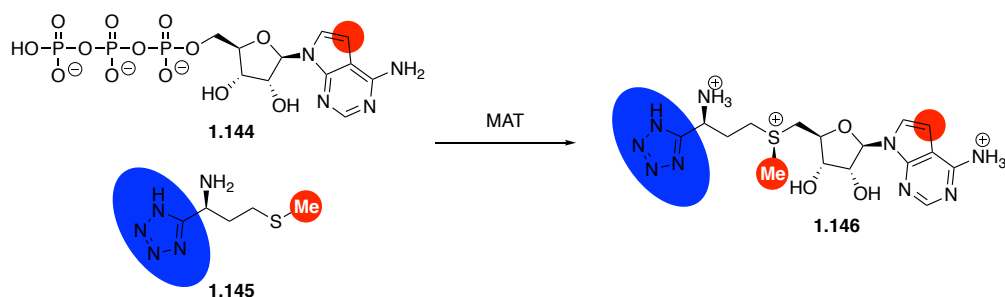
Scheme 1.49. MAT mediated synthesis of 7-dz-*tet*-SAM.
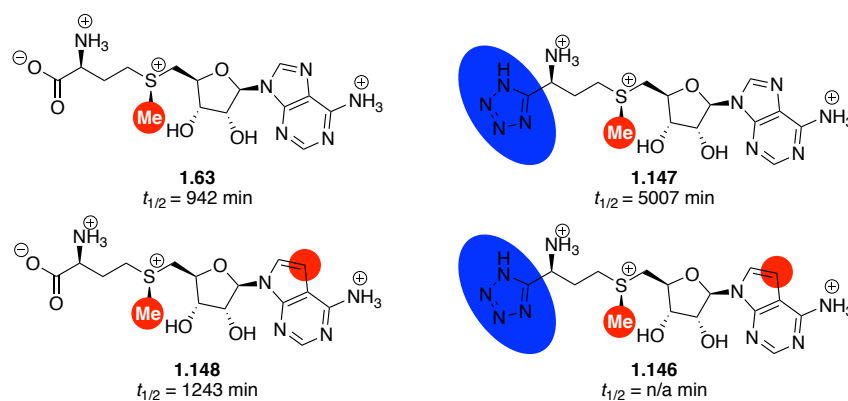
Figure 1.11. SAM analogues and their associated half-lives.

An assessment of the stability of these SAM analogues was carried out by incubation of the compounds at pH 8, 37 °C (Figure 1.11). *Tet*-SAM displayed an increase in half-life to 5007 min compared with the half-life of SAM, 942 min. In this case it was also observed that the degradation pathway to MTA and HSL was completely suppressed. 7-dz-SAM showed a slightly enhanced half-life to 1243 min and the SAM isostere containing both modifications showed no observable degradation after 3300 min.

These isosteres were then assessed for their activity towards a model MTase: DnrK, which can methylate unnatural substrate 2-chloro-4-nitrophenol (ClNP) (**1.149**). A high-throughput colorimetric assay was developed for this MTase, which allowed facile screening of the SAM isosteres and comparison of the unnatural compounds against SAM itself (Scheme 1.50). In each case, the SAM isostere showed significantly reduced activity with respect to SAM, with the greatest activity observed for 7-dz-SAM (**1.148**) (21% vs 52% for SAM). The use of non-native ClNP as the DnrK substrate will also contribute to the low conversion observed. Despite these reductions in activity, these SAM isosteres could prove useful due to their significantly enhanced stability.

Further study of additional MTases with stable SAM isosteres would reveal whether these large reductions in activity are observed for all MTases or if this is only true for DnrK. Stable SAM analogues could become particularly valuable if there were a class of MTases for which the conversion to methylated product was similar to that of SAM itself.

SAM/ analogue	% Conversion
1.63	52
1.147	19
1.148	21
1.146	13
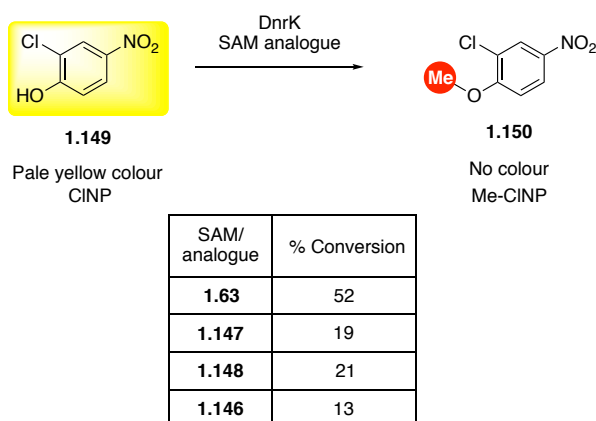
Scheme 1.50. Conversions of ClNP to Me-ClNP.

1.10 Cofactor regeneration strategies

The generation of SAH as a stoichiometric byproduct remains a problem in the use of SAM and analogues as alkylating agents due to its potential to inhibit MTase reactions. An ideal solution would be to take the SAH generated and reuse it for further SAM/analogue synthesis. Two alternative strategies for the reuse of SAH are discussed below.

1.10.1 MAT and phosphate regeneration cycle

Andexer *et al.* used a cyclic system of SAM generation and then regeneration of cofactor through sequential breakdown of SAH to a variety of simpler materials which are then reconstructed to reform SAM for further use (Scheme 1.51).[109] Initially SAM is generated by MAT, as mentioned previously, with triphosphate displaced by either L-met or an L-met analogue. After the SAM or SAM analogue produced has been used by a MTase, SAH is generated. A carbon-sulfur bond is then cleaved by SAH-hydrolase (SAHH) to produce L-homocysteine and Ado. Ado (**1.116**) is converted to AMP (**1.147**) by adenosine kinase, which utilises additional ATP as its phosphate source. ADP (**1.148**) is produced by family-2-polyphosphate kinase II (PPK2-II) by addition of phosphate from polyphosphate. Finally, ATP for utilisation in AMP synthesis and SAM synthesis is produced by family-2 polyphosphate kinase I (PPK2-I) by phosphorylation of ADP.
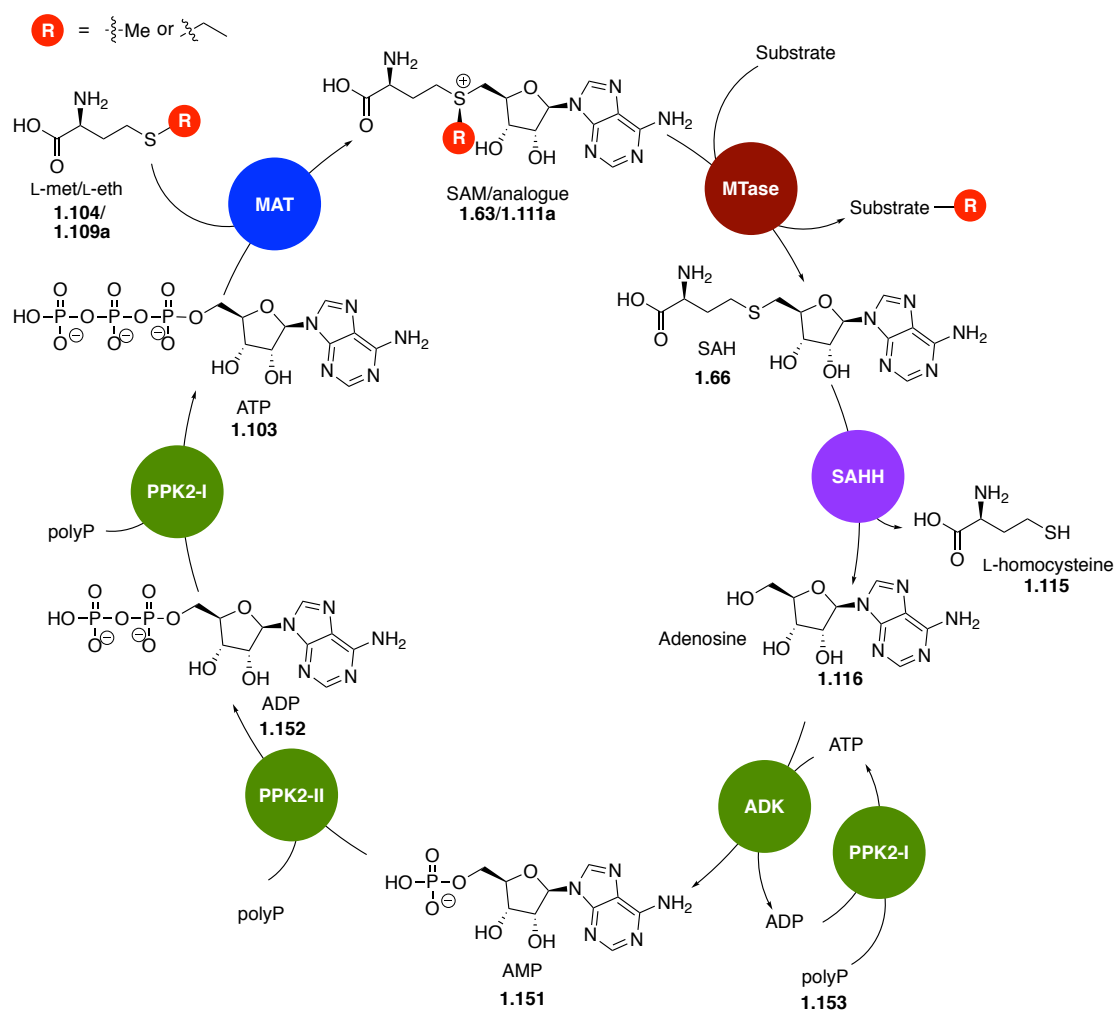
Scheme 1.51. Catalytic cycle used to regenerate SAH to SAM. Enzyme abbreviations – MAT = Methionine Adenosyl Transferase, MTase = methyltransferase, SAHH = S-adenosyl homocysteine hydrolase, ADK = adenosine kinase, PPK2 = family-2 polyphosphate kinase (PPK2-I from *Sinorhizobium meliloti*, PPK2-II from *Acinetobacter johnsonii*).[109]

This system was demonstrated to function in tandem with three different MTases: RgANMT, RnCOMT and MxSafC. These enzymes selectively methylate benzoic acid derivatives (**1.154**, **1.156** and **1.158**). For each MTase, at least ten cycles of SAM regeneration were observed (Scheme 1.52). Although this was an impressive proof of concept, the exact nature of the bottleneck in this process is not yet known. Formation of Ade was observed and proposed as a limiting factor as Ade forms a catalytic dead end for this method.
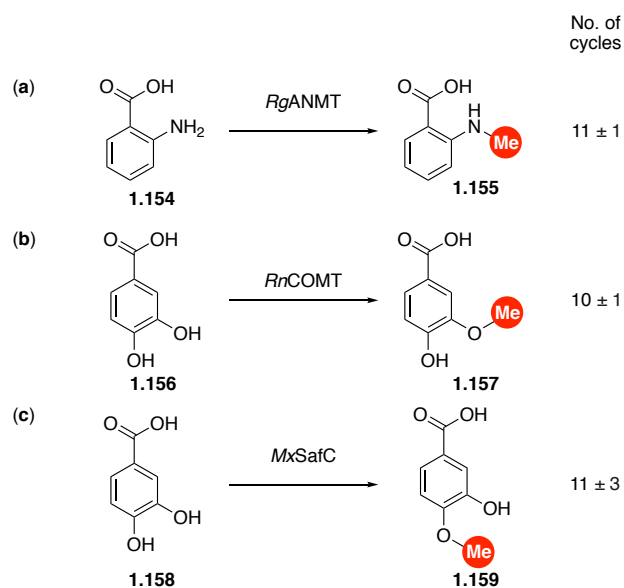
Scheme 1.52. Selective methylation of 2-aminobenzoic acid and 3,4-hydroxybenzoic acid by *Rg*ANMT, *Rn*COMT and *Mx*SafC.

1.10.2   Direct SAM generation from SAH

A simplified SAM regeneration system has been developed by Seebeck *et al.* where SAH is methylated to form SAM by a halide methyltransferase (HMT) (Scheme 1.53). In this case iodomethane is used as the methyl donor and facilitates regeneration of SAM after a MTase has transferred its methyl group.[110] Proof of concept was obtained with *O*-MTase inositol 4-MTase (IMT), *N*-MTase putrescine MTase (PMT), and *C*-MTases SgvM and NovO (Scheme 1.54). Each showed at least 35 cycles of SAM regeneration, with the IMT and NovO catalysed reactions pushed to >90% conversion by further MTase addition. This suggests that the reaction is limited by enzyme stability rather than any aspect of the tandem process. The operational simplicity of this system enables closing of the catalytic cycle in principle, for any methyltransferase.
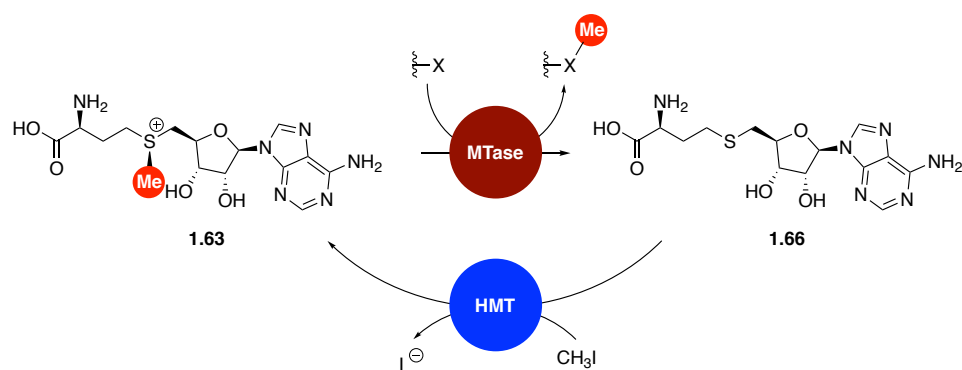
Scheme 1.53. Schematic of SAM utilised by a MTase to provide methylated product and SAH, where X = O, N or C. SAH is then regenerated by transfer of a methyl group from methyl iodide as facilitated by HMT.
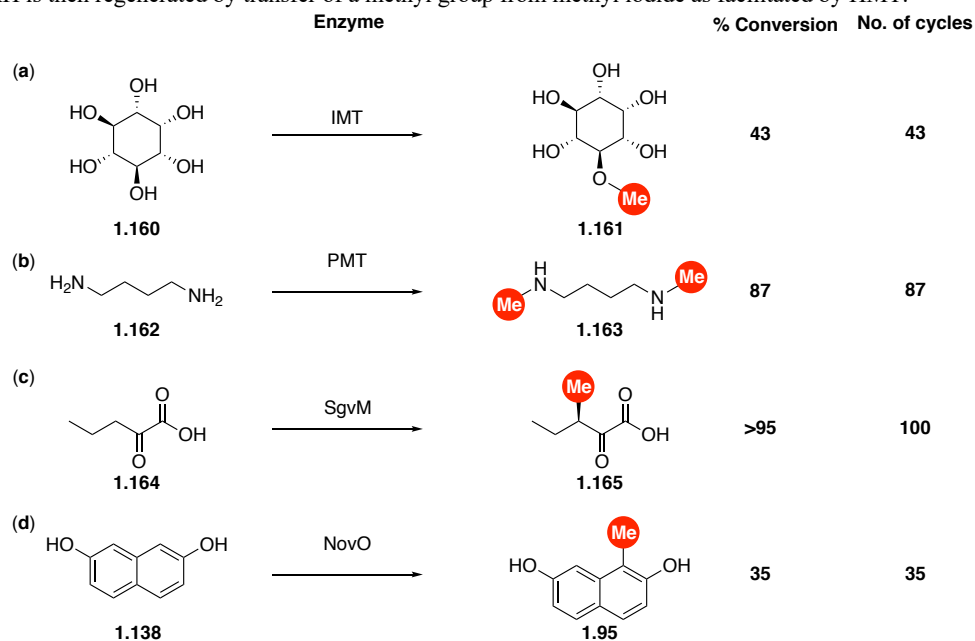
Scheme 1.54. Model MTases used to display applicability of SAM regeneration from SAH as a general platform for methylation of small molecules.

63

1.11 Hypothesis

SalL is capable of generating SAM from ClDA/L-met and has displayed utility in a tandem SAM generation/methylation system.[77, 99, 102] There is scope to expand on previous investigations into the substrate scope of SalL for turnover of L-met and ClDA analogues. The SalL reaction has a number of advantages over the MAT reaction such as the enhanced stability and lower cost of ClDA when compared with ATP.

The initial working hypothesis of this thesis is that with appropriate mutation, the active site of SalL will accommodate modifications to methionine, and therefore generate alkyl analogues of SAM (Figure 1.12a). Also, modifications to active site residues of SalL will provide further insight into residues which are important for substrate binding and the generation of SAM. SalL has thus far shown limited promiscuity towards unnatural substrates; however, investigation of ClDA analogues has not yet been undertaken. FDAS shows promiscuity towards nucleobase modified ClDA analogues and shares 35% sequence similarity with SalL (Figure 1.12b). From this it is proposed that SalL will also show promiscuity towards nucleobase modified ClDA analogues.
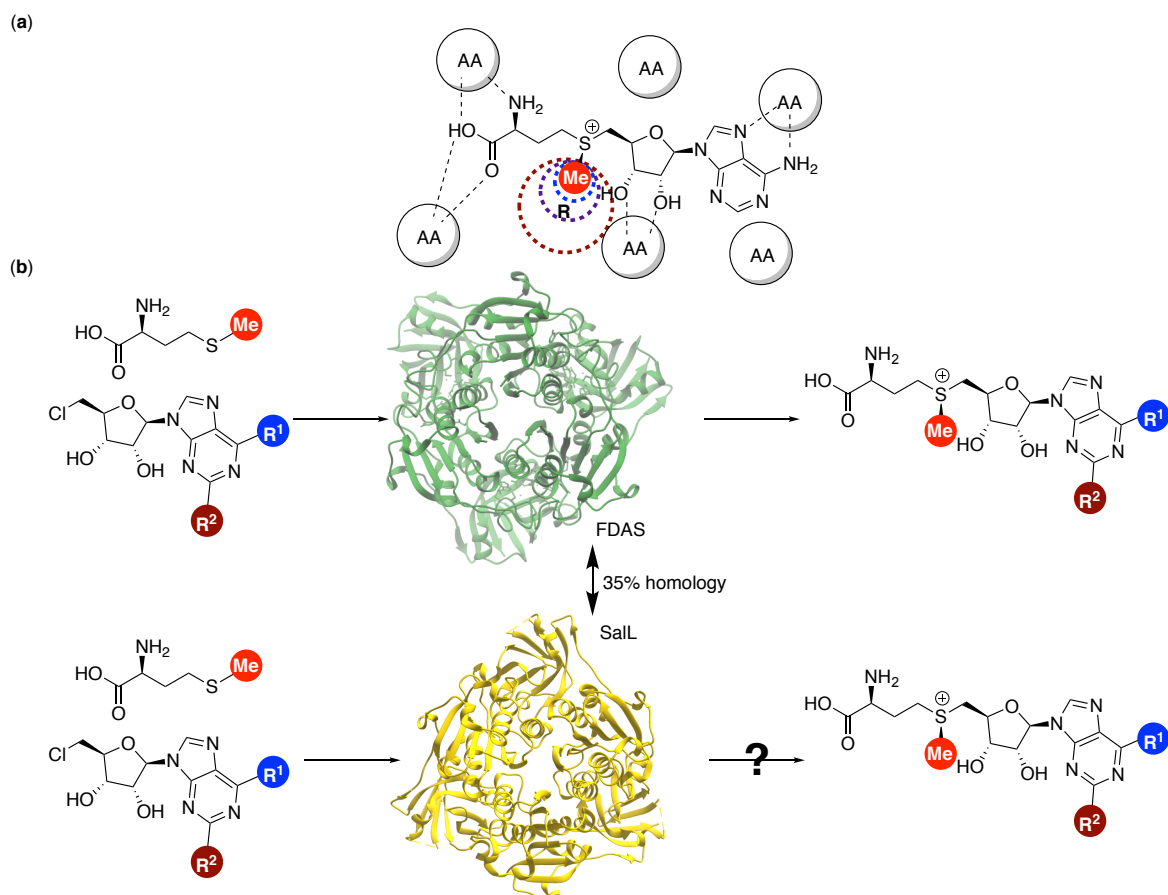
Figure 1.12a. Cartoon of SalL active-site depicting goal of expanding size of alkyl group tolerated from methyl to a larger R group and also probing function of active-site residues depicted as "AA". **b.** Modification of ClDA cofactor to probe whether SalL while show similar levels of promiscuity to modified substrates as FDAS/*flA*1 which shares 35% homology with SalL.

Finally, MTases such as NovO have already been shown to display activity towards non-natural SAM analogues with modifications at the alkyl group primed for transfer,[80] however, no nucleobase modified SAM analogues have been investigated in tandem with a SAM dependent MTase. It is surmised that modified SAM prepared by SalL will be utilised by NovO to enhance the diversity of C-alkylation (Figure 1.13).
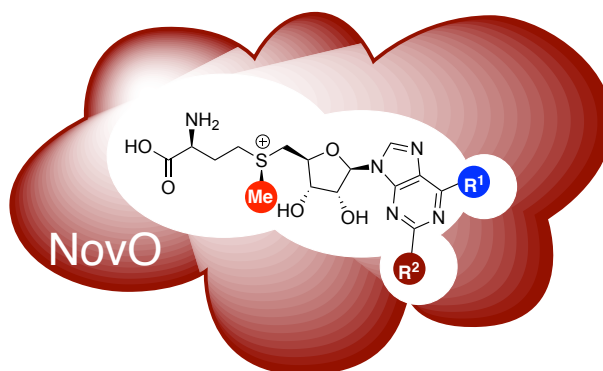
Improved binding interactions between NovO facilitated by modifications R1 and R2 may improve SAM binding and thus improve substrate diversity and alkylation efficiency of NovO.

Figure 1.13. Cartoon of NovO to depict modifications to SAM cofactor which may enhance SAM binding and thereby increase activity of such modified cofactors with respect to the natural substrate.

1.12 Thesis Aims

The specific aims of this thesis are to:

(i) Explore the functional determinants of catalysis of active-site residues in SalL through enzyme mutagenesis.

(ii) Prepare cofactor analogues to probe the scope of SAM analogue formation catalysed by SalL.

(iii) Explore the compatibility of SAM analogue formation by SalL with *C*-alkyl transfer catalysed by NovO.

Chapter 2

Mutagenesis and Structural Studies of SalL

Abstract

SalL from Salinospora tropica acts as a chlorinase in nature to generate ClDA and L-met from SAM. In the absence of high chloride concentration, the equilibrium of the reaction can be pushed towards SAM formation. A crystal structure for SalL was obtained to guide site-directed mutagenesis studies. This structure was also used to guide potential modifications to substrates supplied to SalL which could be accepted and converted to the corresponding SAM analogues. Kinetic parameters were measured for the active mutants *vs*. the wild-type enzyme.

2 Mutagenesis and Structural Studies of SalL

2.1 Enzymatic Generation of SAM Catalysed by SalL

Note: Work carried out in this section was part of a collaboration with University of York. Crystallisation trials were carried out with Amina Frese and Anibal Cuetos. The SalL crystal structure was solved by Anibal Cuetos and Gideon Grogan.

The purpose of this chapter is to investigate the generation of SalL mutants and evaluate their ability to synthesise SAM. Any mutants which retain activity could be used as part of a tandem SAM generation/alkylation process with NovO to form C-alkylated coumarin products. Determination of kinetic parameters for the mutant enzymes was used to quantify the catalytic performance of the mutants against wild-type SalL. The information gained from the kinetic studies and from inactive mutants would then be used in a future round of mutagenesis. The ultimate goal of mutagenesis of SalL is to enhance its ability to accept L-met analogues. If successful, this could then expand the tandem SalL/NovO process to a general alkylation platform rather than one simply for methylation.

2.2 The Central Dogma of Molecular Biology

DNA is composed of a sequence of four different building blocks linked by a phosphodiester backbone. Each building block contains a deoxyribose sugar and one of four nucleobases – adenine (A), cytosine (C), guanine (G), thymine (T) - the combination of which is known as a nucleotide. The sequence of a strand of DNA is made up of three nucleotide units each of which contains genetic information for the generation of proteins. These units are known as codons and it is these which govern the sequence and composition of amino acids within a protein.

The process by which the genetic information within a DNA sequence is converted into a protein sequence is known as the central dogma of molecular biology and occurs in the two steps: DNA is converted to RNA by transcription and the resulting RNA is translated into protein (Figure 2.1). During transcription, the DNA duplex is unwound by an RNA polymerase which "reads" the DNA sequence and builds an RNA strand which is composed of bases complementary to those which are being read (A complements T but this is changed to uracil – U - in RNA, C complements T and *vice versa*) but with ribose as the sugar contained in these molecules rather than the deoxyribose of DNA. The single stranded mRNA sequence is then

read and a polypeptide chain is built by the ribosome. It is this polypeptide that folds appropriately to provide the desired protein.

As a codon contains three positions for the four possible nucleotides, there are 64 possible codons. As there are only twenty canonical amino acids there are a number of codons which encode for the same amino acids, this is known as codon degeneracy. Enzyme mutagenesis methods rely on altering DNA sequences so that the resulting transcription/translation process produces a protein which has been mutated with respect to the wild-type. When a particular mutation is targeted, codon degeneracy is not a problem as any codon which produces the desired mutation can be utilised. Problems occur in random mutagenesis methods where codon degeneracy means that different DNA sequences can lead to generation of the same protein sequence.

Codon degeneracy represents a major issue for evolution of biocatalysts. Every DNA mutation which encodes for a degenerate mutant provides an additional construct for screening for which no change in activity would be observed. As screening is a major bottle-neck in programmes of mutagenesis, any way which can minimise the screening of unnecessary additional mutants is beneficial. There are a number of methods which have been developed for mutagenesis of enzymes, some of which have been developed to try and circumvent codon degeneracy.
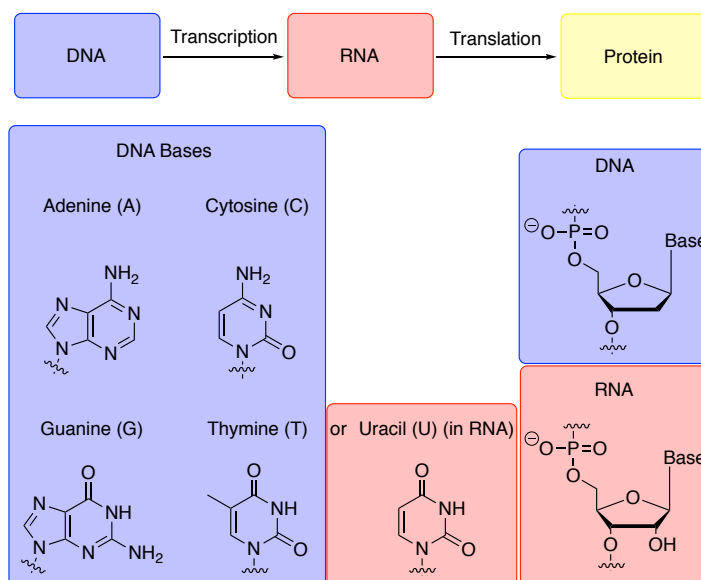
Figure 2.1. The central dogma of molecular biology.

2.3 Methods of Enzyme Mutagenesis

Genetic engineering is the process of the manipulation of DNA sequences to alter the proteins for which they encode.[111] Whilst this process has been carried out in Nature for billions of years,[112] it has also been applied in the lab and involves repeated cycles of mutagenesis on a DNA construct which encodes for a protein of interest. The starting point for this process is generally a protein which has a desired function but has significant shortcomings. This could be limited thermostability,[113] low catalytic turnover,[114] low or undesirable enantioselectivity[115] or low tolerance of organic solvents.[116]

The process begins with mutagenesis of the template DNA sequence to generate mutant DNA constructs. Protein expression is then carried out and the resulting proteins are screened for activity. If any of the proteins expressed carry out the desired function to the requirements of the user (e.g. a certain amount of conversion or ee) the cycle is finished. If the mutant has enhanced properties but the desired properties have not yet been realised, another cycle will be carried out using the best mutant as the starting construct to be mutated (Figure 2.2). This process can be repeated for as many cycles as is desired until an appropriate level of enzyme activity is achieved.
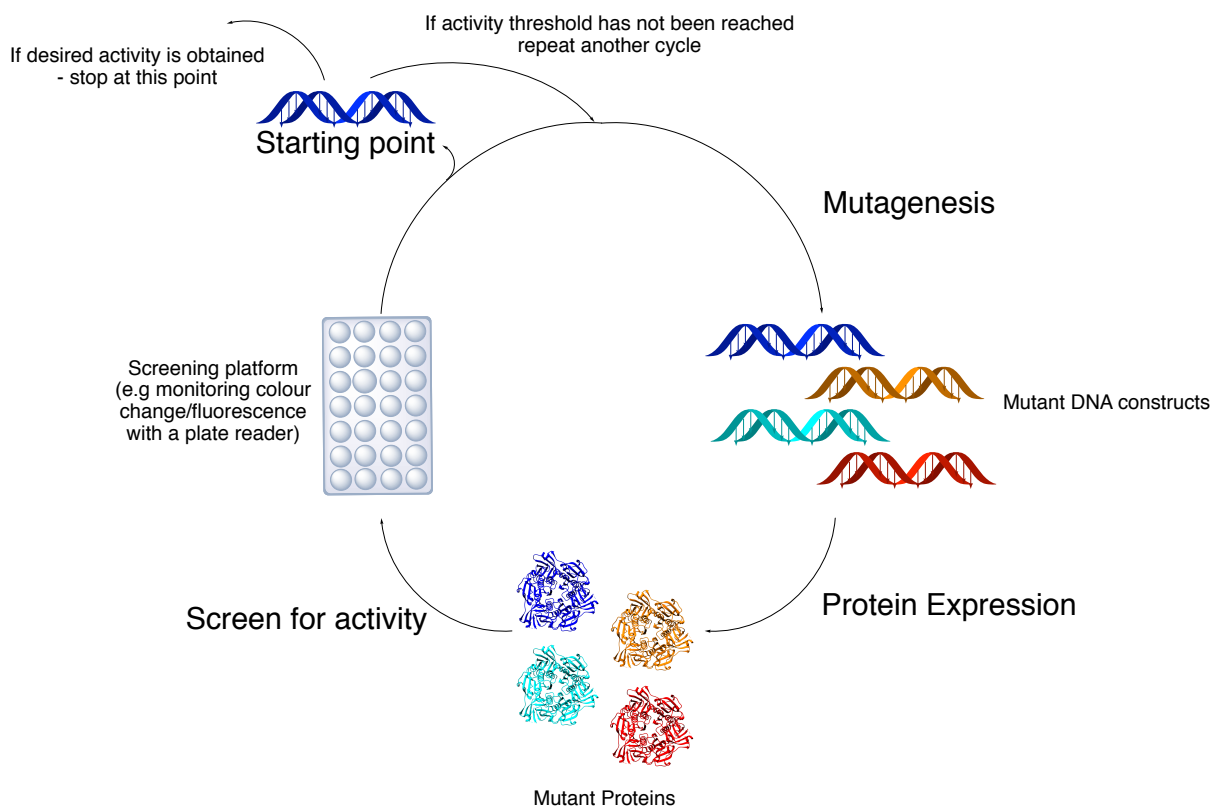
Figure 2.2. Cycle of directed evolution.

A number of different methods for carrying out mutagenesis have been employed by groups working in genetic engineering. Among the favoured approaches are error prone polymerase chain reaction (epPCR),[117] DNA shuffling[118, 119] and site/iterative saturation mutagenesis.[120, 121]

2.3.1 Error Prone Polymerase Chain Reaction (epPCR)

epPCR involves PCR amplification of a DNA sequence which encodes for the protein of interest. However, the DNA polymerase used in this case is one which has a low fidelity and as such generates numerous errors when replicating the DNA sequence thus generating sequence diversity. The number of mutations present can be controlled by factors such as the variation of concentrations of Mg^{2+} in the reaction and also the relative amounts of each deoxynucleotide triphosphate (dNTP) present.[122, 123] A library of mutagenised sequences are produced and these are then cloned into vectors for protein expression.

2.3.2 DNA shuffling

The DNA shuffling method operates by cutting a gene into a mixture of random DNA fragments using restriction enzyme DNaseI. These fragments are then recombined using PCR to form randomised DNA strands composed of fragments generated by nuclease degradation.[124] This approach has the benefit of allowing for completely random changes to the protein sequence thus potentially identifying beneficial mutations that may not be obvious from targeted mutagenesis. The primary drawback with this approach is that large libraries are generated which can be laborious to screen.

2.3.3 Site-Saturation Mutagenesis

Site-saturation mutagenesis is the process of identifying an amino acid of interest within a protein and then mutating this to every other possible amino acid. If a variant of improved activity is found, this process can then be repeated for another residue of interest using the single mutant as the template. When repeated cycles are carried out using the improved variant as the template, this is known as iterative saturation mutagenesis (ISM).

The most commonly used method for introduction of mutations at a particular site is the QuikChange protocol by Stratagene.[125] This method utilises mutagenic primers which are typically 25-45 bases long and contain a desired mutation in the middle of the primer. Codon degeneracy can cause issues in this process as there are multiple codons which can encode for

a particular amino acid leading to bias in the library formed.[126] For example, an amino acid such as Met or Trp is only encoded for by a single codon each while amino acids such as Ser or Leu are encoded for by six different codons each. For a single-site of mutation, the statistical likelihood of generating a leucine mutant rather than a tryptophan mutant is 36:1.[127] This has been investigated in detail with careful synthetic oligonucleotides design used to minimise codon degeneracy. One such codon which has been utilised at the site of mutagenesis is NNK (where N = A, C, G or T and K = G or T).[121] This approach allows for encoding all 20 canonical amino acids through the use of 32 possible codons while also minimizing redundant codons (e.g. two codons which encode for the same amino acid) and only encoding for one "stop".

Another method, developed by the Reetz group, has reduced the codon degeneracy further by using a modified mixture of oligonucleotides.[126] A total of 22 codons are utilised with the format NDT or VHG (where N = A, C, G or T; D = A, G or T; V = A, C or G) and a non-degenerate tryptophan codon, TGG. Using this set allows for coverage of all amino acids with only valine and leucine encoded for by two codons each with no stop codons incorporated. Other combinations have been developed which encode for amino acids bearing particular properties such as hydrophilicity (VRK: 12 codons – 9 amino acids R = A or G) small size (KST: 4 codons – 4 amino acids K = G or T; S = G or C) or hydrophobicity (NTT: 4 codons – 4 amino acids).[128]

2.3.4 Combinatorial Active Site Test (CASTing)

An example of the use of ISM was developed by Reetz *et al.* in the use of CASTing. This method pairs residues within the active-site of an enzyme and has been proposed to occupy the middle ground between total random mutagenesis and mutagenesis at a single site.[129] This method generates a much smaller number of mutants than total random mutagenesis. The screening effort is therefore reduced, while also covering significantly more of substrate space than single point saturation mutagenesis. The power of this method was demonstrated by evolution of a lipase to expand its substrate scope (Scheme 2.1). The natural enzyme from *Pseudomonas aeruginosa* catalyses the hydrolysis of palmitic acid *p*-nitrophenol ester (**2.01**) but cannot process bulkier substrates (such as **2.04** – **2.06**). After identifying a range of active site pairings from an available crystal structure,[130] CASTing was carried out to generate a focussed mutant library. Out of a mutant library consisting of 3000 constructs, 8 showed significant increases in either substrate scope, rate of hydrolysis of palmitic acid or both. The screening of this process was simplified greatly by the use of a *p*-nitrophenol (**2.03**)

colourimetric probe.[131] This allowed for screening to be carried out in either 96- or 384-well plates which took 8 minutes each to read.
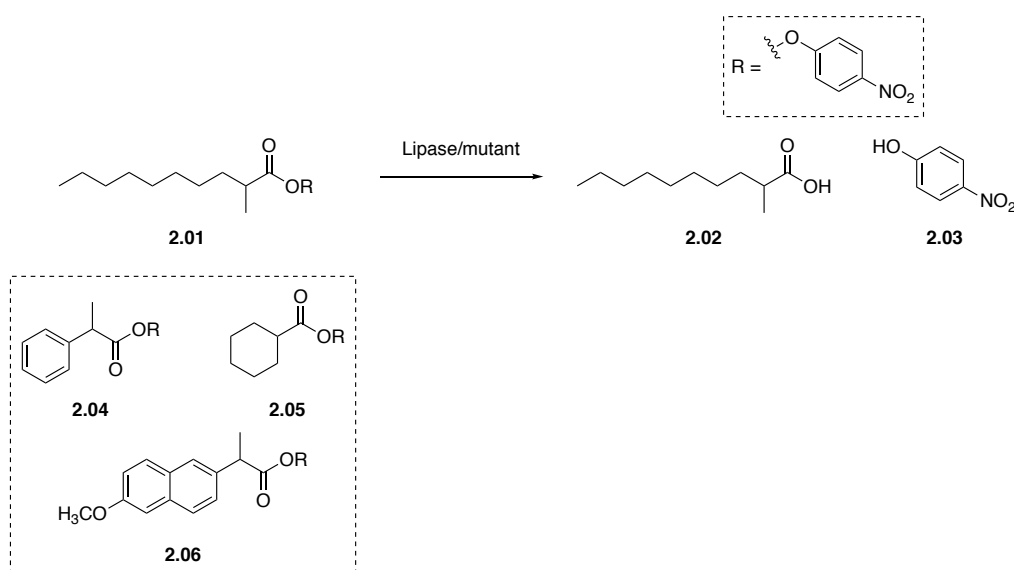
Scheme 2.1. CASTing facilitates the expansion of the substrate scope of a lipase to allow it to hydrolyse a range of racemic *p*-nitrophenyl esters which possess significantly different structures from the original substrate (**2.01**).[129]

In general, this method has been particularly effective in the evolution of enzymes where the goal of mutagenesis is altering observed stereo- or regioselectivity, as these properties are generally determined by the active site architecture.[132-134] Although effective in reducing library size and therefore screening effort, targeting of particular residues can only be carried out where some prior knowledge of the enzyme, such as an X-ray structure, is available. Even when this information is available there are certain sites where mutagenesis may be beneficial but which are impossible to identify by observation of the structure alone; these may only be identified through a random mutagenesis strategy. In each case, the method of choice may come down to the type of screen available and its throughput capability.

2.3.5 Directed Evolution Approaches for the Development of New Biocatalysts

More recently, further examples of the power of Directed Evolution as a genetic engineering tool have emerged. In recent work by Kan *et al.* Directed Evolution has been used for the evolution of naturally occurring heme-containing Cytochrome c (Cyt c) enzymes to carry out carbon-silicon bond formation: a transformation unknown in Nature.[135] A variety of studies have previously been carried out by the Arnold group, exploring the generation of carbenes by

treatment of ethyl diazoacetate (EDA) with Cytochrome enzymes.[136] An initial experiment investigated whether or not a unit of heme without the protein scaffold could facilitate the reaction of ethyl 2-diazopropanoate (Me-EDA) with phenyldimethylsilane. The target product was observed in very small amounts with no stereocontrol observed, and so this reaction was then attempted with a panel of Cytochrome enzymes. It was found that an enzyme from *Rhodothermus marinus* (Rma) could catalyse this reaction with low turnover number (≈ 35) but with 97% ee. Through sequential site saturation mutagenesis of three active site residues, the triple mutant Val75Thr, Met100Asp, Met103Glu was identified. This enzyme had greatly improved catalytic properties with a turnover number > 1500 and > 99% ee.
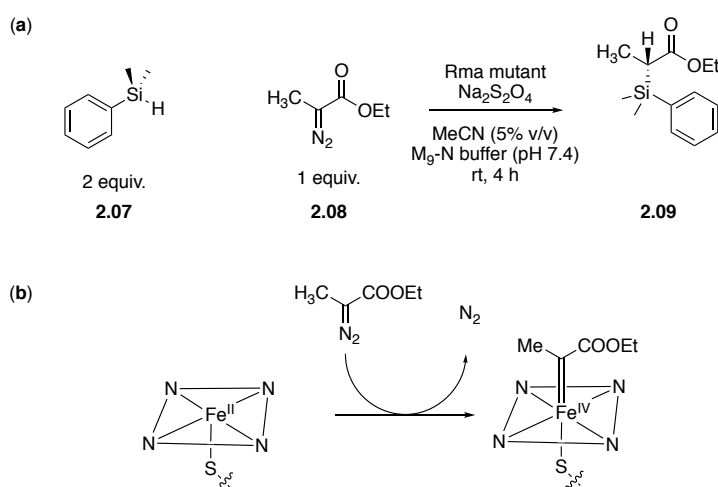
Scheme 2.2a. Enzymatic synthesis of C-Si bonds using cell lysate of Rma Val75Thr, Met100Asp, Met103Glu mutant, reaction carried out under anaerobic conditions. **b.** Proposed mechanism of carbene formation.[136]

2.4 Aims

Specific aims for this chapter are:

(i) Express, purify and crystallise SalL to determine structure.

(ii) Generate a focused library of active site mutants.

(iii) Screen their ability to form SAM.

(iv) Determine kinetic parameters for active mutants and compare them with wild-type SalL.

2.5 Site-Directed Mutagenesis of SalL

Previous work in SalL mutagenesis and assessing the ability of mutants to synthesise SAM has relied on an RP-HPLC screen to monitor the conversion of ClDA to SAM.[97, 100] RP-HPLC screening has a much lower throughput than, for example, a colorimetric method utilising a plate reader. For the purpose of this thesis, a SDM approach was undertaken as this would limit the size of library generated and therefore the screening effort required. Based on literature precedent from Moore and Burkart, RP-HPLC methods showed high sensitivity and were able to observe very low concentrations of SAM (picomolar in case of Moore *et al.*[97]) and SAM analogues.

To aid in design of SalL mutants, an X-ray structure of the intact trimer of SalL with SAM bound was obtained. Previously published structures of SalL do not contain an intact trimer and as a result of this may not provide all the essential information to design functional SalL mutants.[97] The main objective of the SDM work undertaken was to determine the effect of point mutations on catalysis. An alanine scan was carried out across a range of residues through mutating each residue of interest to alanine. Absence of activity for any of these mutants would suggest importance of the target residue to either substrate binding or enzyme activity.

In previous studies it has been observed that SalL is tightly restricted in terms of L-met analogues which can be accepted.[100] A range of alkyl chains larger than the methyl in L-met showed significantly reduced activity with respect to the natural substrate.[100] L-met was converted to SAM at a rate of 4930 nM/min with the ethyl analogue processed at a rate of 453 nM/min. Propyl, butyl, allyl and benzyl analogues were also screened for activity with each displaying < 1.02 nM/min activity (See Section 1.8.3, Scheme 1.45). Other analogues which incorporated propargyl, phenylethyl, amino acid and a primary amide showed no activity. This work also included generation of a small library of SalL mutants focused around Thr128, Trp190 and Tyr239. It was proposed that these residues facilitated hydrogen bonding of a water molecule in the active-site of the enzyme at a position which prevented accommodation of larger substrates than L-met.[100] Of the mutants investigated, Trp190Ala showed a slight increase in turnover for the butyl analogue of Met with respect to the wild-type. All other mutants showed significantly reduced activity for each substrate investigated. In the case of Trp190Ala turnover of L-met was reduced from 4930 nM/min to 30.9 nM/min. With these

results in mind, further mutagenesis was conducted to enhance the scope of SAM analogue formation bearing sterically demanding L-met analogues.

2.6 Overexpression of SalL

SalL has previously been expressed as a readily soluble trimer with yields of 60 mg/L obtained.[97] A pET26b(+) plasmid containing the gene for SalL was transformed into BL21(DE3) strain of *E. coli* by heat shock. Selection for SalL containing *E. coli.* was achieved by plating onto Kanamycin containing agar and the presence of a gene for Kanamycin resistance within the pET26b(+) vector (Appendix Figure S 2.1).[137] Literature conditions documented growth of cells at 25 °C to an optical density of 0.8 before induction of protein expression.[97] This was carried out by addition of IPTG to a final concentration of 0.25 mM and incubation of the mixture at 20 °C overnight.

Typically, protein expression is controlled by use of a plasmid which contains a T7 promotor adjacent to the desired gene for expression. This promotor is under the control of an inducible *lac* promotor. Cells are grown to a desired optical density before induction is carried out to begin protein expression. As cells grow, the T7 and *lac* promotor sequences are bound by *lac* repressor proteins from the *lacI* gene, thus inhibiting generation of T7 polymerase by the native *E. coli* polymerase. When lactose is present, allolactose can be generated by β-galactosidase and can bind to the repressor protein. This event induces a conformational change in the repressor which releases it from the *lac* promotor thereby allowing generation of T7 polymerase. This process can also be carried out by addition of a non-hydrolysable mimic of allolactose, IPTG (Figure 2.3). Once this process has occurred, the gene for T7 polymerase is exposed to *E. coli.* polymerase which carries out transcription and translation thus generating T7 polymerase. Once T7 polymerase is present it begins to transcribe the gene of interest which is then translated into the protein of interest (Figure 2.4).[138]
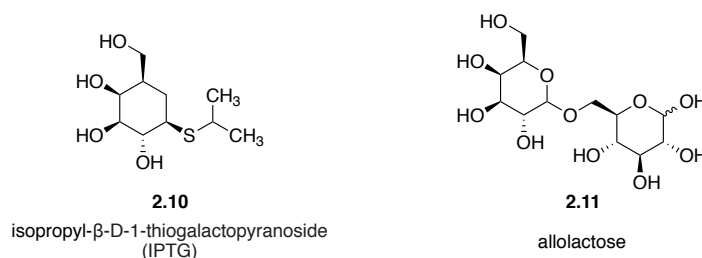
Figure 2.3. Structures of IPTG and allolactose.
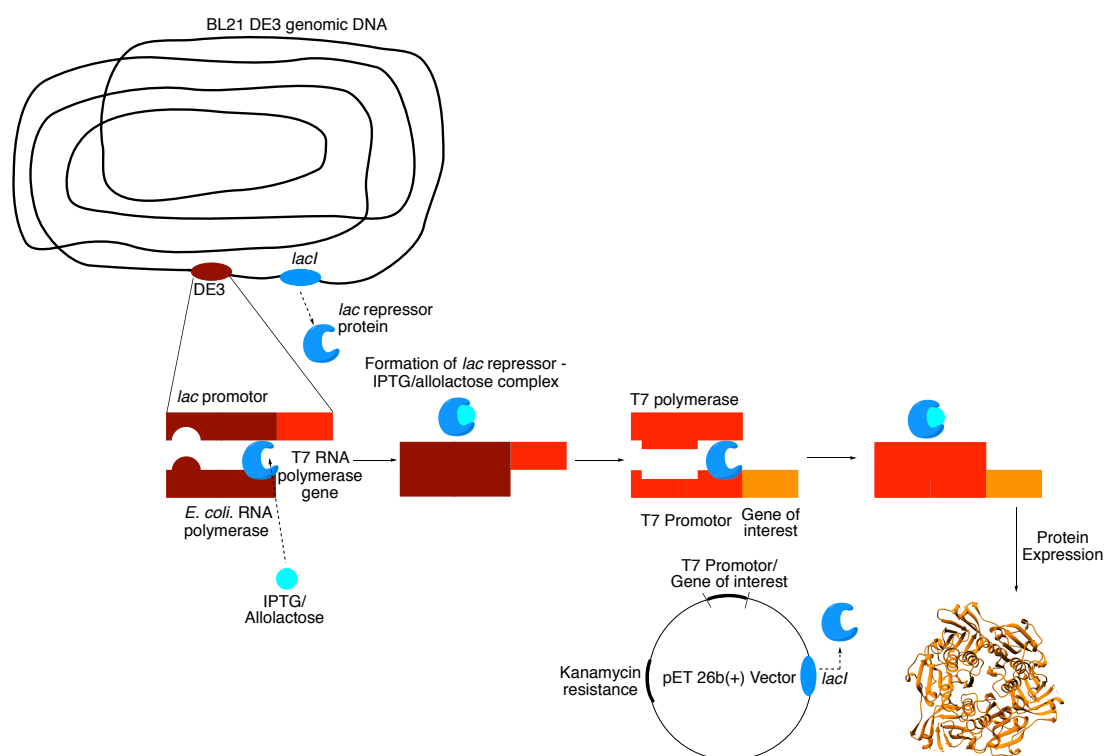
Figure 2.4. Schematic of the process of overexpression using *E. coli.*

In this work, these conditions were modified by the use of Overnight Express[TM] (OE) autoinduction media thus avoiding the addition of IPTG for protein expression.[77, 102] Autoinduction media contains controlled levels of glucose and lactose. Cells preferentially consume glucose thus allowing for significant cell growth. Once the glucose has been consumed, cells look for other carbon sources such as lactose. Lactose can be converted into allolactose by β-galactosidase and in turn can bind to the *lac* repressor proteins thus beginning the process of protein expression as above.[139]

Cells were grown at 37 ºC in OE to an optical density (OD) of ~ 1.2. The increase in temperature increased the rate of cell growth, thereby reducing the incubation time required to reach the desired OD. Once the required growth had occurred, 1 mL of the culture was sampled for SDS PAGE analysis while the remaining culture was incubated at 18 ºC overnight. Cells were harvested by centrifugation and were stored at -80 ºC.

SalL expression was checked before going through any of the purification processes. This was achieved by analysing various protein fractions by SDS PAGE. Samples of the cells grown were taken before and after induction of protein expression with cell lysis carried out. Further

to this, a sample of soluble protein was obtained by centrifugation of the post induction protein sample. Each of these three samples were mixed with SDS loading dye, heated to 90 °C and loaded onto the gel (Figure 2.5). As expected, a protein matching the mass of SalL was the major band observed in the post-induction sample and also in the post-induction soluble sample. To confirm the identity of this band, Protein Mass Fragmentation (PMF) analysis was carried out, with the resulting fragments observed (Appendix Figure S 2.2).
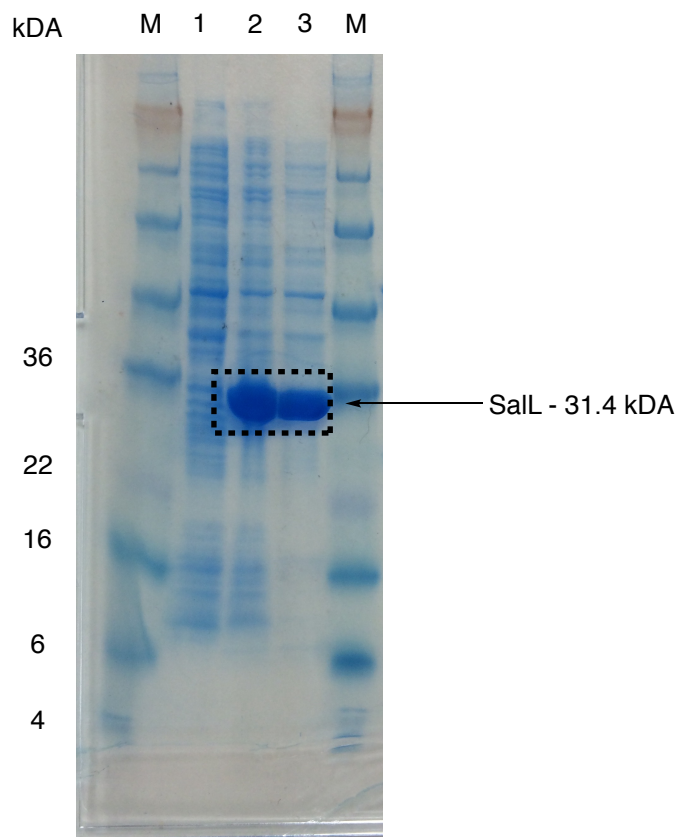
Figure 2.5. SDS PAGE of SalL expression. M - Marker, 1 - SalL pre-induction, 2 - SalL post-induction, 3 - SalL post induction soluble.

After confirmation of the presence of SalL, affinity chromatography purification was carried out. This process relies on the engineering of a 6-His tag into the protein sequence of interest. This tag has a high affinity for Ni(II) so is passed through a column packed with resin bearing this metal at its surface. The column is washed initially with a low concentration of imidazole, which competes with the His tag to bind to Ni(II), washing off of any non-specifically bound proteins. The concentration of imidazole is increased gradually and once the concentration of

imidazole is high enough to out-compete the 6-His tag the protein of interest is eluted (Appendix Figure S 2.3).

2.7 Crystallisation of SalL

Previously published structures contained either SAM and chloride or L-met and ClDA but were only of monomer units. As SalL exists as a trimer, with the active site at the monomer-monomer interface,[97] it was desirable to solve a structure of the intact trimer with either the starting materials or the product bound. To begin this process, fresh protein was prepared and purified by affinity chromatography (using the same method as in Section 2.6). SalL containing fractions were concentrated in a 10,000 MW centrifugal filter unit and were injected directly onto a Size Exclusion Sephadex 200 Increase 10/300 gl column. SalL was isolated as a single peak with a retention time of ~ 22 minutes, corresponding to an elution volume of ~ 13.2 mL (Appendix Figure S 2.4). The purity of each fraction was checked by SDS PAGE and all SalL containing fractions were concentrated to 5 mg/mL and stored on ice.

SalL samples were added to centrifugal filter units and mixed with either ClDA and L-met or ClDA and L-eth. Samples were concentrated to 10 mg/mL of SalL in 100 µL of solution with substrate concentrations of 10 mM, significantly higher than K_m for either ClDA or L-met ($(2 \pm 1) \times 10^{-4}$ mM and 2.6 ± 0.2 mM respectively).

Crystallisation plates were set up and incubated at 18 ºC overnight after which time crystals were observed (Figure 2.6). The best crystals were observed using sodium formate (0.2 M), PEG 3,350 (20 % w/v) for both L-met and L-eth. The crystals formed were then observed to diffract sufficiently (Figure 2.7). Each complex was analysed at Diamond Light Source in Didcot, Oxfordshire for collection of X-ray diffraction data. Structures of SalL with SAM and chloride and SalL with ClDA alone were solved by molecular replacement and were refined to resolutions of 1.50 Å and 1.77 Å (Appendix Table S 2.1). Unfortunately, L-eth was not present in any of the complexes analysed. It was hypothesised that this was a result of the increased K_M and therefore reduced affinity of L-eth for SalL in comparison with L-met.

(a)   **(b)**
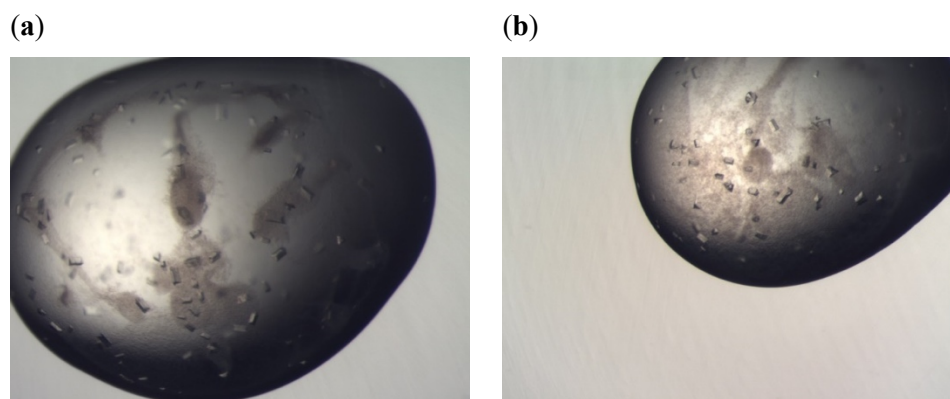
Figure 2.6. Images of crystals formed. Crystallisation conditions: Sodium formate (0.2 M), PEG 3,350 (20 % w/v). (a) SalL wild-type 10 mg/mL, L-met and ClDA (10 mM). (b) SalL wild-type 10 mg/mL, L-eth and ClDA (10 mM).
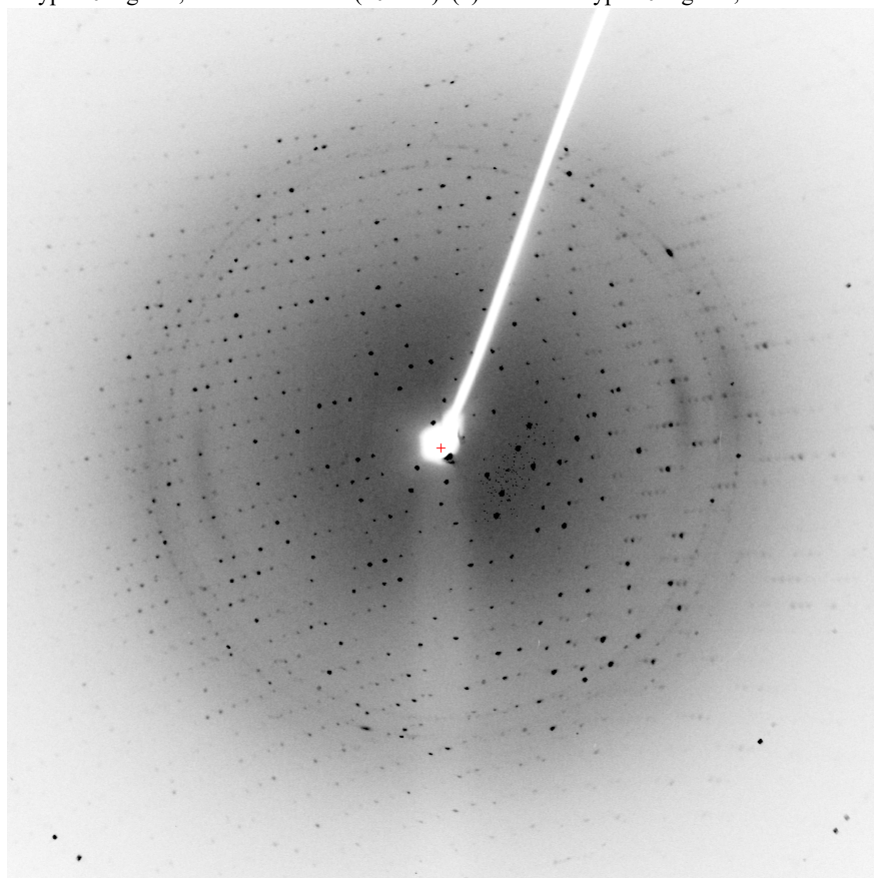
Figure 2.7. Diffraction test of SalL wild-type complex.

2.7.1 Comparative Analysis of SalL Structures

With the new SalL structures obtained, a comparison with the existing structures available for SalL in the PDB was carried out. Moore *et al.* previously published four structures for SalL (Table 2.1). A key difference in the structures obtained in this work was the presence of SAM and chloride in the wild-type enzyme (Figure 2.8). In the study of Moore *et al.* it was proposed that the reduction in the active site crowding in the Tyr70Thr mutant trapped SAM and

chloride within the active site. Thr70 bound two new water molecules at this position which prevented escape of chloride. However, with the first structure obtained in this study, it has been demonstrated that this replacement is not necessary to observe SAM and chloride within the active site. In this work, SAM and Cl⁻ were observed without turnover to ClDA and L-met. It appears that Cl⁻ is in the correct geometry to undergo nucleophilic substitution with displacement of L-met but that this has not yet occurred in the structure obtained (Figure 2.8).

Another significant difference between the structures obtained in this thesis was the presence of an interaction between Arg243 and the carboxylate of L-met. In three out of four of Moore's structures Arg243 was oriented away from the active-site of the enzyme and faced towards the interface between the solvent and the exterior of the protein (Figure 2.8). Only in the double mutant of 2Q6L was it oriented towards the bound substrates. When facing towards the active site, Arg243 is in ionic bonding distance with Glu17 from the adjacent monomer thus increasing the strength of the monomer-monomer interaction.

Table 2.1. Summary of SalL crystal structures previously published by Moore *et al.* and crystal structures of SalL obtained as part of this work.

Moore[97]				This work			
PDB	Unit	Substrate	Resolution	PDB	Unit	Substrate	Resolution
2Q6I	Monomer	ClDA + L-met	2.60 Å	6RYZ	Trimer	SAM + Cl⁻	1.50 Å
2Q6K	Monomer	Adenosine	1.55 Å				
2Q6L	Monomer (Tyr70Thr/Gly131Ser) Mutant	ClDA + L-met	2.72 Å	6RZ2	Trimer	ClDA only	1.77 Å
2Q6O	Dimer (Tyr70Thr) Mutant	SAM + Cl⁻	2.00 Å				

(a)
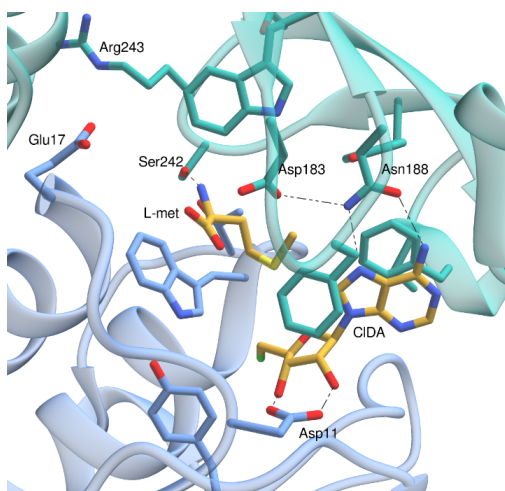
(b)
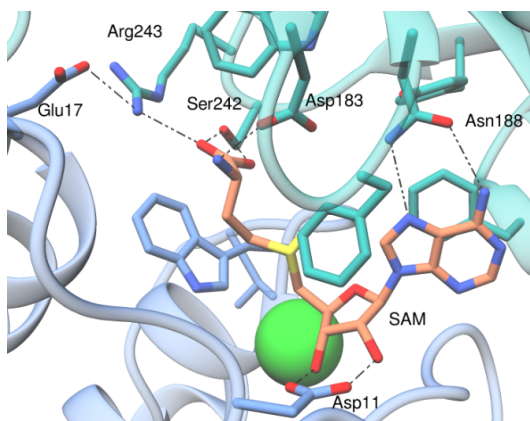
Figure 2.8a. SalL published by Eustaquio *et al.* with L-met and ClDA bound. **b.** SalL structure obtained as part of this work with SAM and chloride bound.

A significant feature observed for all SalL structures is that the 2-position of the nucleobase is solvent exposed (Figure 2.9). This raises the possibility of carrying out modifications at this position without perturbing any of the key interactions within the active site, as has been previously observed for related fluorinases FDAS and *flA1*.[95, 96, 140]
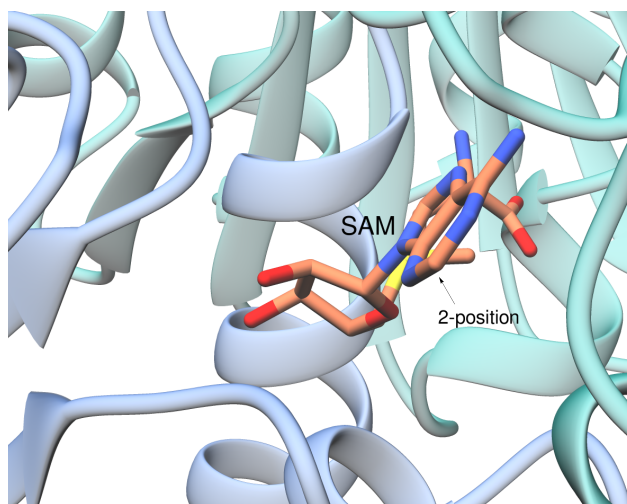
Figure 2.9. Structure of SalL with SAM bound, showing solvent exposed nature of 2-position. PDB Accession code: 6RYZ.

2.8 Preparation of Site-Directed SalL Mutants

A suite of residues were identified for a point mutagenesis study through analysis of the structures obtained by Moore[97] and in this study. Trp129 was targeted due to its proximity to the L-met substrate (Figure 2.10). Replacement with a less bulky residue such as Phe or Ala could provide a larger pocket for a L-met analogue to reside in. A similar situation could be envisioned for the Trp190 residue which again lies very close to L-met but on the opposite face to Trp129 (Figure 2.10). Trp190 has previously been studied by Burkart *et al.*, with the Phe, His and Ala mutants all investigated.[100] The Ala mutant showed some activity over a range of substrates (ethyl, propyl, butyl, allyl, and benzyl – methionine analogues) so this mutant was deemed to be an appropriate target.

Asp183 was chosen due to the close contact between its side chain and the carboxylic acid group of L-met (Figure 2.10). This residue could play a significant role in the binding of L-met in the active site which may be elucidated depending on the resulting activity of the Asp183Ala mutant. Asp183 was also substituted by Glu, with the effect of lengthening the side-chain of interest. Asn188 was replaced with Gln to probe the effect of changing the position of the hydrogen bonding interaction that exists between this residue and the N7 and exocyclic amine of the Ade moiety of ClDA (Figure 2.10). It was hypothesised that introducing greater flexibility at Asn188 by increasing the length of the amino acid sidechain may lead to movement of the substrate within the active site thus leading to a modulation in substrate

scope. Ala mutants of Asp183 and Asn188 were generated to probe the importance of hydrogen bonding at both of these positions.

Phe186 and Phe228 lie close to the Ade ring of ClDA and, as a result, could be involved in a key π-π interaction (Figure 2.10). With the Ade nucleobase sandwiched between these residues, their importance in substrate binding was probed. Mutation of Phe186 to Leu and Ala was carried out while the Phe228 residue was mutated to Ile and Ala.

The Tyr239 residue contains a hydroxyl group that lies close to the 8-position of the Ade ring (Figure 2.10). Tyr239 lies in close proximity to both the methyl group of L-met and to the 8-position of Ade. The effect of reducing steric bulk was investigated by the Tyr239Thr mutant so as to maintain any potential electronic effects which may be at play. The corresponding Tyr239Ala mutant was also targeted as by reducing steric bulk this mutation would abolish any potential electrostatic interaction which may exist. Burkart *et al.* did attempt to access a number of Tyr239 mutants but were unable to obtain sufficient levels of expression.[100]

Finally, Val12, Tyr70 and Thr75 residues were targeted as they all lie in close proximity to the 5′-position of ClDA (Figure 2.10). It was proposed that replacement of these residues may modulate activity of the enzyme towards SAM formation. Each of these residues was substituted with Met. It was hypothesised that a Met residue in close proximity to the 5′-Cl within ClDA may be able to carry out the S_N2 reaction to form SAM that is bound in the active site. This molecule could then be displaced by attack of an additional Met from solution to result in formation of unbound SAM.

Residue	Mutations carried out
Trp129	Phe/Ala
Trp190	Ala
Asp183	Glu/Ala
Asn188	Gln/Ala
Phe186	Leu/Ala
Phe228	Ile/Ala
Tyr239	Thr/Ala
Val12	Met
Tyr70	Met
Thr75	Met
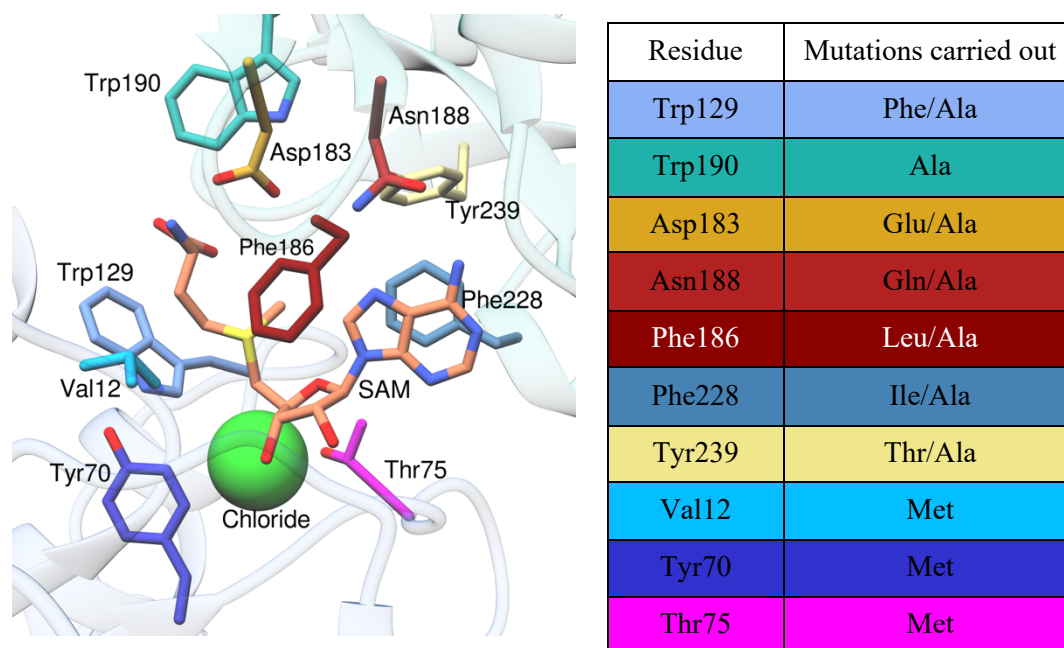
Figure 2.10. SalL structure with residues for site directed mutagenesis highlighted. Mutations which were carried out are detailed in table.

Each construct was prepared by SDM using a modified version of the QuikChange protocol by Stratagene.[125] After PCR, DpnI digestion was used to remove methylated DNA. As the DNA from the original plasmid was produced by an *E. coli* strain it will be methylated and therefore cut by Dpn1, while mutated DNA produced by primer extension during PCR will not be methylated and will therefore remain uncut. After inactivation of Dpn1 by heating to 80 °C for 20 minutes, the resulting PCR product was transformed into Top10 *E. coli.* and plated onto LB agar containing kanamycin. A control agar plate was also made up with a PCR reaction where the primers were not added and were replaced with an equal volume of water so each reaction had the same total volume. In the absence of primers, no reaction should have taken place and the parent DNA plasmid should have been degraded by DpnI. Bacterial colonies were observed for each plate except the control plate which contained no primers.

To maximise the chances of obtaining the desired mutation, 24 colonies were picked from each mutant plate and these were grown in LB. The resulting cells were lysed and their DNA harvested by a Miniprep method. Each construct obtained was then sequenced to determine whether or not the desired mutation was present. All mutants were sequenced aside from Trp129Ala where the sequence returned in each case was that of the wild type. A repeat of the SDM to form this mutant was unsuccessful with each colony picked returning the wild type sequence. From this it was assumed that the primers as designed were either not annealing to

the template plasmid or were perhaps annealing to each other.[141] With a number of other constructs available it was decided to move on with the expression and purification of the constructs which had been successfully obtained.

BL21 (DE3) *E coli.* were transformed with the remaining DNA constructs and protein expression was carried out using the same conditions as the wild type enzyme.[102] SDS PAGE of each enzyme was carried out to ensure expression had occurred and that any resulting protein was soluble in the desired purification conditions. Although this was the case for most of the constructs, a few suffered from solubility issues with neither Tyr239Thr or Tyr239Ala showing any soluble protein band by SDS PAGE. As the amide backbone of Tyr239 is in hydrogen bonding distance of the adjacent Ala247 residue (Figure 2.11), it was hypothesised that modification at this position significantly altered the steric environment around the residue. The resulting shift in the residue may perturb the backbone-backbone hydrogen bonding interaction and cause misfolding of the protein thereby adversely effecting solubility. This may also explain why Burkart *et al.* were also unable to generate mutants at this position.[100] Circular Dichroism was determined for a range of the mutants expressed (Appendix Section 2.1) with no significant deviations from the wild-type observed.
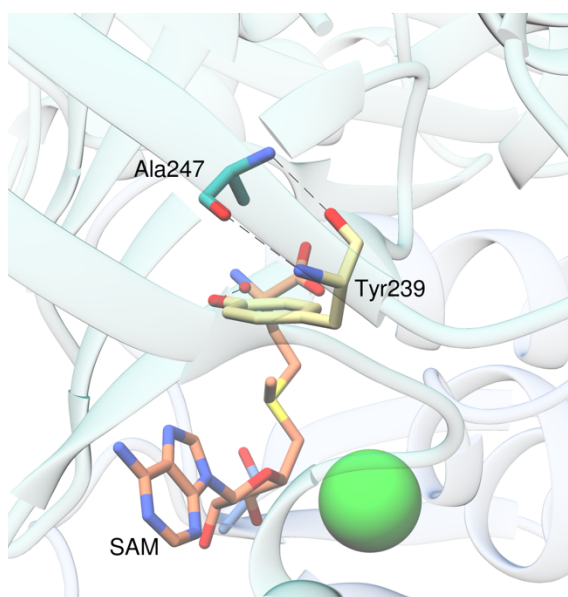
Figure 2.11. SalL structure highlighting a hydrogen bonding interaction between backbone of Tyr239 and Ala247.

In order to determine the relative activity of the mutants it is important to have an accurate measure of how much L-met is contained within each reaction, something which would be impossible in the presence of unknown amounts of endogenous L-met. Further to this, a later

goal of this study is to probe the ability of these mutants to form SAM analogues from ClDA and L-met analogues. If L-met is present from the protein expression it is highly likely this will be preferred over any unnatural L-met analogue thus preventing an accurate screen of L-met analogue formation.

All remaining mutants were purified using a gradient elution with increasing amounts of imidazole (see Experimental for details). This was successful for all the constructs aside from Asn188Gln. In the case of this mutant, all protein was eluted in the wash fraction of the affinity chromatography at 10% imidazole buffer. A repeat of the purification was carried out with 0% of the imidazole containing buffer used in the initial wash of non-specifically bound protein to see if this would facilitate binding of the His-tagged protein. Unfortunately, the same result was observed with all protein being stripped from the column in the wash fraction. From this it was assumed that either the protein was no longer soluble, or the mutation had changed the conformation of the protein so that the His-tag was no longer solvent accessible and therefore unable to bind to the column.

2.9 Assay Development for Determining SalL Mutant Activity

With a suite of pure SalL mutants in hand, initial screening was undertaken to see which mutants maintained enzymatic activity. Previously, activity of SalL was assessed by RP-HPLC so this provided a starting point for developing a method of assessing SalL activity.[97] ClDA has a strong absorbance at 254 nm, which is due to the presence of Ade. Displacement of chloride with L-met results in no change in the extinction coefficient of the product at 254 nm with respect to the starting material as Ade is unchanged by this transformation. One of the known degradation products of SAM is MTA which results from intramolecular attack of the carboxylate of L-met onto the carbon adjacent to the sulfonium (Scheme 2.3). The absorbance of MTA at 254 nm is caused by Ade and will therefore possess the same extinction coefficient as the starting material and product.[103] With this in mind, the conversion of ClDA to SAM can be determined by measuring a ratio of ClDA: SAM/MTA.
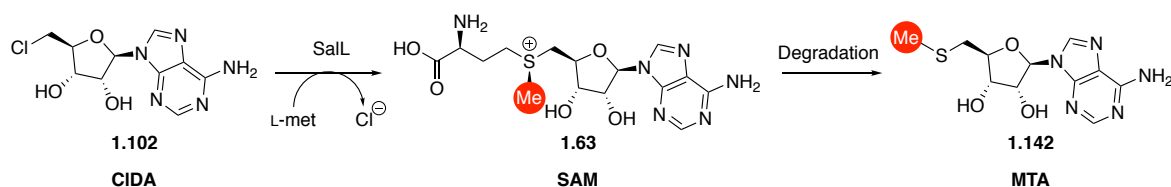
Scheme 2.3. Synthesis of SAM from ClDA, with subsequent breakdown to MTA.

To first establish a method for monitoring this transformation, authentic standards of each compound of interest were required. SAM and MTA were commercially obtained while ClDA was synthesised by carrying out the chlorination of readily available adenosine with SOCl$_2$.[142] This reaction proceeded on gram scale with high yield and was isolated by filtration. Analytical RP-HPLC of the material obtained showed it to be >95% pure and therefore suitable for the purposes of the enzymatic assay. Comparison of ClDA, MTA and SAM standards was carried out by RP-HPLC to confirm they did not overlap (see Appendix Section 2.1).
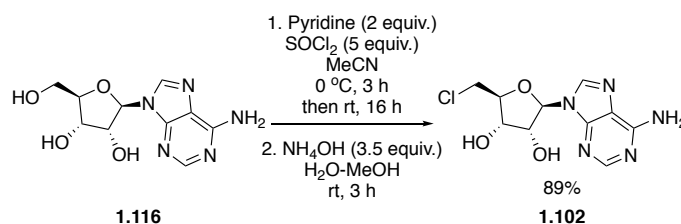
Scheme 2.4. Reaction of Ado to form ClDA.

Another consideration is the method by which enzyme is removed from the reaction mixture to allow RP-HPLC analysis to take place. A variety of methods are available for quenching enzymatic reactions such as altering the pH[100] or heating the reaction.[77] For this reason, it was essential to determine an effective way of quenching the SalL-mediated SAM formation and determining the response of ClDA and MTA to such conditions. Reaction of ClDA and L-met in the presence of SalL were carried out using literature conditions.[102] After 1 h the reaction mixture was sampled three times with three different quench methods trialled. These were acidification to pH 3 using sodium formate buffer,[100] basification to pH 10, and heating to 90 °C for 10 minutes. Each sample was analysed immediately by RP-HPLC and then re-analysed after 24 h to ensure that progress of the reaction had been halted. In the case of the basified reaction, further conversion of ClDA to MTA (via SAM) was observed upon re-analysis of the sample, thus showing that the enzyme was still active at this pH. In the case of the acidic quench or the heated quench, no further reaction was observed.

The acidification method would quench enzyme activity more rapidly than heating due to the difference in time required to thoroughly mix the sample when compared with the time taken for a sample to reach the temperature at which SalL denatures. This is particularly important in the generation of kinetic parameters for the enzymatic constructs where short intervals between time points are necessary and instantaneous quenching of reaction is required. For this reason, the acidic quench method was employed. For measuring conversions over 24 h,

heating was used to quench the reaction. This method was used due to operational simplicity when a large set of samples were being screened simultaneously to ensure that reactions were quenched at the same time. On this time-scale, the time taken for the solution to heat to a temperature where SalL is far less significant than on the time-scale of the kinetics experiment (maximum 1 h). ClDA and MTA were stable to heat and acidification to pH 3.

2.9.1 Screening of SalL Mutants for Activity

Screening of SalL mutants was carried out by setting up a range of reactions on 0.5 mL scale, each containing ClDA, L-met, DTT, BSA and the SalL construct of interest. Conditions and concentrations used are based on previously utilised conditions for SAM generation (Scheme 2.5).[102] Reactions were sampled between 0 and 24 h to probe whether the mutants generated were viable for SAM synthesis (Appendix Figure S 2.9). Conversions were calculated by analysis of the samples by RP-HPLC and comparison of the peak areas of ClDA with SAM and MTA. This method followed precedent from the Thorson group.[83, 103]

Using this assay, a number of the mutants retained activity towards SAM formation including Phe186Leu, Asp183Glu, Trp129Phe and Tyr70Met (Scheme 2.5). Each of these constructs showed >75% conversion. Trp190Ala and Val12Met showed conversions of 9% and 16% respectively. Traces of activity were observed for Phe186Ala, Phe228Ile, Phe228Ala and Thr75Met. Constructs Asp183Ala and Asn188Ala abolished activity completely.
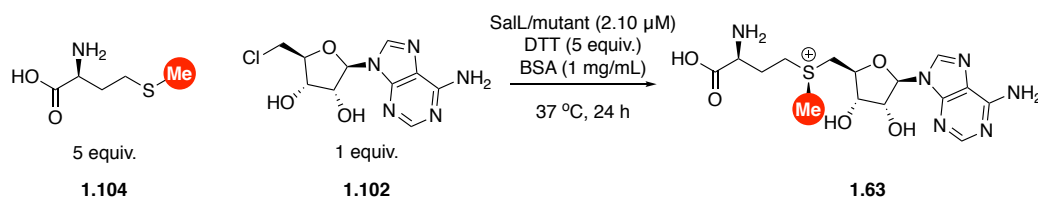
Construct	% Conversion
Wild-type	99
Phe186Leu	96
Phe186Ala	2
Phe228Ala	2
Phe228Ile	2
Asp183Ala	0
Asp183Glu	93
Asn188Ala	0
Trp190Ala	9
Trp129Phe	82
Val12Met	16
Tyr70Met	79
Thr75Met	2

Scheme 2.5. Reaction of ClDA with L-met to form SAM catalysed by wild-type and mutant constructs of SalL.

The importance of π-stacking between Ade and neighbouring aromatic residues Phe186 and Phe228 was studied with generation of Phe186Leu/Ala and Phe228Ile/Ala mutants. In the Phe186 case, activity was observed for the Leu mutant which showed conversion similar to the wild-type while the corresponding Ala mutant showed only traces of activity. In contrast, both the Phe228Ile and Ala mutants showed only traces of activity. From this it was deduced that π-stacking at Phe228 is essential to binding of ClDA but not at Phe186. Retention of the activity of the Leu mutant suggests that as long as some steric bulk is retained to direct Ade towards Phe228, SalL can still function.

Although substrate turnover was observed for Asp183Glu, activity was abolished when the corresponding Ala mutant was generated. This confirms the importance of the carboxylic acid in this position for binding of L-met through an interaction with its amine. In the case of Asn188Ala, no catalytic activity was observed, thus highlighting the essential hydrogen bonding interaction between Asn188 with both the exocyclic amine and N7 of Ade.

Previous studies showed Trp190Ala to be effective in the generation of analogues of SAM, in particular its ability to produce butyl-SAM at a comparable rate to the wild-type enzyme.[100]
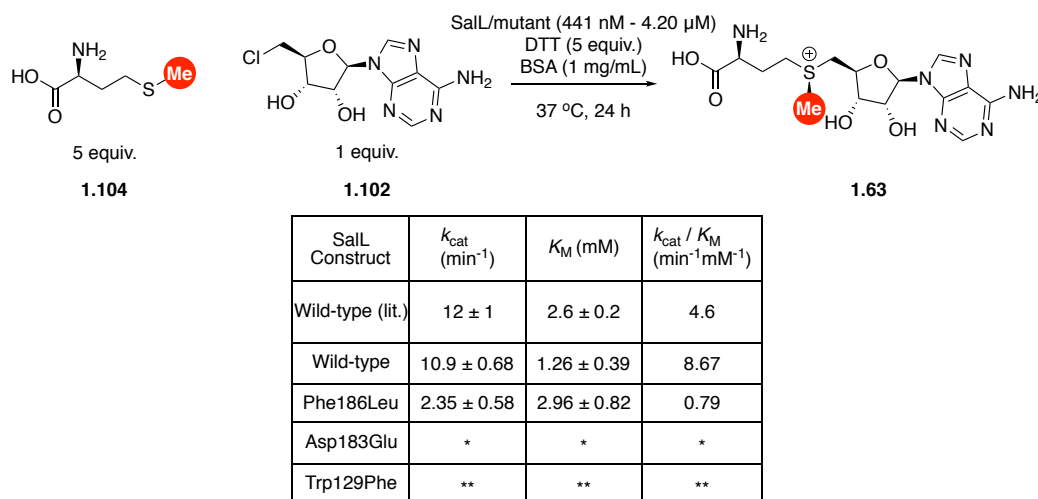
However, in our hands, this construct showed little catalytic activity with 9% conversion to SAM observed in 24 h. In comparison, Trp129Phe maintained activity (82% conversion) over 24 h.

The study of the Met mutants for residues Val12, Tyr70 and Thr75 provided a mixed set of results. Val12Met and Thr75Met showed vastly reduced turnover with 16% and 2% conversions observed respectively. Replacement of Val with Met introduces a longer, more flexible amino acid side-chain into the active site. The location of this residue in close proximity to the SAM methyl group implies that the bulkier side-chain of Met might restrict access of L-met to the active site, which would cause a significant decrease in the activity of the enzyme. The reduction in activity for the Thr75Met mutant suggests that the OH group of this residue plays a role in the formation of SAM. From the proximity of Thr75 to chloride in our SalL structure it has been hypothesised that this residue helps to stabilise chloride in the transition state of SAM formation. Replacement of this residue with Met disrupts this function and thus vastly reduces catalytic activity. In contrast, Tyr70Met displayed 79% conversion, thus implying that this residue can undergo modification without hugely effecting enzyme activity. As this residue lies very close to the SAM methyl group, this could be a site where further modifications and introduction of smaller amino acids may improve activity towards more sterically demanding L-met analogues.

For some of the constructs which retained activity, enzyme kinetics studies were undertaken. The conversion data obtained gave some insight into the impact of the mutation upon enzyme activity. However, a more in-depth investigation was required with the goal of identifying any changes in K_M and k_{cat} which had been caused by the mutations.

2.9.2 Kinetic Parameters of Active Mutants

The wild-type Michaelis-Menten parameters were determined by carrying out the SAM formation reaction at a variety of L-met concentrations (Scheme 2.6). Reactions were sampled at t = 0, 30 s, 1 min, 2 mins, 3 mins, 5 mins, and 6 mins and analysed by RP-HPLC (Appendix Figure S 2.10). Each reaction was quenched by the addition of pH 3 400 mM sodium formate buffer. Reaction progression was seen in each case and gave the following Michaelis-Menten plot from which enzymatic parameters were obtained (Appendix Figure S 2.11). These were in keeping with literature values for K_M and k_{cat} of SalL with respect to L-met (Scheme 2.6).

SalL Construct	k_{cat} (min⁻¹)	K_M (mM)	k_{cat} / K_M (min⁻¹mM⁻¹)
Wild-type (lit.)	12 ± 1	2.6 ± 0.2	4.6
Wild-type	10.9 ± 0.68	1.26 ± 0.39	8.67
Phe186Leu	2.35 ± 0.58	2.96 ± 0.82	0.79
Asp183Glu	*	*	*
Trp129Phe	**	**	**

Scheme 2.6. SAM formation from ClDA and L-met catalysed by wild-type SalL. Michaelis-Menten parameters obtained for SalL wild-type and mutant constructs and compared with SalL wild-type values from the literature.[97] *Asp183Glu was not saturated at the concentrations used, Michaelis-Menten parameters were not obtained. **Trp129Phe mutant no longer follows Michaelis-Menten kinetics at the substrate concentrations screened, instead showing agreement with the Hill model. From the values obtained k_{cat} was calculated to be 2.87 min⁻¹, the Hill coefficient was calculated to be 11.93 ± 0.44 mM.

Following this, the same procedure was repeated for construct Phe186Leu which displayed the highest turnover of any of the mutant enzymes. Unfortunately, trial reactions at 441 nM showed no observable conversion for the 1 mM L-met reaction after 6 mins. For this reason, the concentration of enzyme used in the assay was increased from 441 nM to 1.71 μM and the time over which the reaction was monitored was increased to 30 minutes, rather than 6 mins. The conversion of ClDA to SAM catalysed by Phe186Leu under these conditions was ≈ 41% (for 1 mM L-met). As the reaction progression could be monitored, these conditions were deemed to be appropriate for analysis of the kinetic parameters of this construct (Appendix Figure S 2.12).

For construct Asp183Glu the time of reaction was increased to 60 mins and the amount of enzyme was also increased to 4.20 μM to observe the progression of the reaction on a reasonable time-scale. When the initial velocity of reaction was plotted across increasing concentrations of L-met the response was linear. From this it can be deduced that saturation of the mutant enzyme was not achieved under these conditions. The upper concentration used in these assays was 15 mM L-met and the rate observed at this value was around 4 nmol/min. This value was similar to the turnover of the wild-type at this concentration, however the concentration of Asp183Glu was 9.5 × higher than in the wild-type case. The rate at 15 mM L-met was 8 nmol/min for the Phe186Leu mutant and again this was with 2.5 × less enzyme

than in the Asp183Glu reaction. As the turnover and affinity for L-met was clearly significantly lower no further investigation to determine exact k_{cat} and K_M values was carried out (Appendix Figure S 2.13).

Interestingly, in the case of the Trp129Phe construct the enzyme appeared to follow a different enzymatic model than Michaelis-Menten. At low concentrations of L-met the increase in initial velocity of reaction was slow. At 5 mM L-met there was a sudden steep increase in velocity which then began to plateau at 15 mM L-met. In this case the enzyme followed the Hill model rather than the Michaelis-Menten model thus implying cooperativity between substrate binding and activity (Appendix Figure S 2.14). [143] With three active sites per trimer - each residing at the monomer-monomer interface - it could be envisioned that binding one molecule of L-met and the subsequent production of SAM induces a conformational change in the enzyme structure. This in turn may encourage binding of substrate and production of another SAM molecule. Although noted as interesting, the catalytic turnover of this construct was still significantly lower than that of the wild-type construct therefore this phenomenon was not investigated further.

2.10 Summary and Future Work

SalL has been expressed and purified, and a crystal structure of the enzyme containing SAM and chloride was also determined. Based on the observation of structural features of the enzyme a range of mutants were expressed and their activities towards SAM formation were determined. A number of catalytically active constructs have been identified, their ability to synthesise SAM has been analysed and kinetic parameters were obtained. Although none of these mutants show enhanced activity compared to the wild-type, mutants such as Phe186Leu maintain comparable turnover. A number of the constructs expressed were observed to be catalytically inactive.

The further development of SalL as an enzyme for SAM analogue generation may benefit from a Directed Evolution approach. For this to be achieved, development of a high-throughput screen would need to be developed. A potential method of screening by colourimetric output may be the use of a chloride sensor. As SAM is formed during the reaction, chloride is displaced from ClDA and there are now a number of chloride sensing kits available such as from Sigma Aldrich[144] and BioVision.[145] Both display a linear response to chloride from 20-120 nmol and are reported to be sensitive to ≈ 0.4 mM of chloride. If the

method showed sufficient sensitivity it would allow for a high-throughput screen to be developed which in turn would facilitate Directed Evolution of SalL. This would open up a number of avenues for developing SalL, not only to expand substrate scope but also to enhance properties such as thermostability and tolerance towards organic solvent.

Chapter 3

Probing the Scope of SAM Analogue Synthesis Catalysed by SalL

Abstract

SalL mutants expressed in Chapter 2 were screened for their ability to convert analogues of L-met and ClDA into analogues of SAM. A suite of Met and ClDA analogues were prepared by chemical synthesis. Once a number of unnatural substrates had been identified, different combinations of unnatural L-met and ClDA analogues were trialled to determine if doubly modified SAM analogues could be synthesised. The outcome of this chapter was that although SalL is limited in terms of L-met analogues to the ethyl analogue, the enzyme can accommodate a variety of ClDA analogues. Successful synthesis of the corresponding SAM analogues from these compounds could then be used in a tandem SAM generation/ NovO mediated alkylation (Chapter 4).

.

3 Probing the Scope of SAM Analogue Synthesis Catalysed by SalL

3.1 Scope of Enzymatic Synthesis of SAM Analogues

Despite SalL being used to generate SAM *in situ* for MTase mediated (m)ethylation, the scope of this process has not been expanded to explore other alkyl or heteroatom containing groups. [77, 99, 100, 102] Burkart *et al.* have previously displayed competency of SalL in the generation of SAM analogues (Section 1.8.3, Scheme 1.45) but with low conversions observed for all L-met analogues containing an *S*-alkyl group larger than ethyl.[100]

Further to this, the ability of SalL to form SAM analogues bearing some groups with relevancy in medicinal chemistry and bioconjugation was investigated. Trifluoromethyl groups, for example, have been extensively utilised in medicinal chemistry due to their unique properties of acting as an isostere of a methyl group but with enhanced metabolic stability.[21] Alkynes have been used widely in bioconjugation as they provide a bio-orthogonal chemical handle for selective functionalisation. This strategy has been used extensively to tag alkyne-modified DNA and proteins using chemically synthesised alkyne analogues of SAM. The alkyne group is transferred by an appropriate DNA MTase, followed by tagging with an azide *via* a CuAAC reaction.[82, 88, 146, 147] With this in mind, establishing the substrate scope of wild-type SalL and the mutants developed in this study was desired. To achieve this, a range of L-met analogues were synthesised (Scheme 3.1).

Wider use of SAM dependent MTases has been limited by the inherent instability of SAM. Previous studies have shown that replacement of the Ade N7 of SAM prevents depurination and replacement of the carboxylate with tetrazole prevents intramolecular cyclisation and formation of MTA and HSL (Section 1.9).[103] SAM analogues bearing such modifications have been synthesised enzymatically by MAT enzymes which forms SAM from 7-DeazaATP (dz-ATP) and a tetrazole containing analogue of L-met (*tet*-L-met). Synthesis of 7-DeazaClDA (dz-ClDA) and *tet*-L-met was carried out to investigate if SalL could also produce this stable SAM analogue (Scheme 3.1).

Further to this, a range of SAM analogues bearing modifications to the Ade nucleobase have previously been synthesised by enzymes which share significant homology with SalL

(~35%).[96, 140] The ability of SalL to carry out such transformations has not been tested previously so it is desired to investigate its scope with respect to Ade modified ClDA analogues (Scheme 3.1).

MTases have been shown to accept a broad variety of SAM analogues bearing different alkyl groups at the sulfur position. The ability to form such compounds enzymatically by SalL could be used as part of a SalL mediated SAM (or analogue) generation/MTase mediated alkylation tandem process. Although nucleobase modified SAM analogues have been synthesised before, they have not been tested as cofactors for MTases. Any successfully synthesised SAM analogues from this chapter could be investigated as cofactors for NovO mediated alkylation (Scheme 3.1).
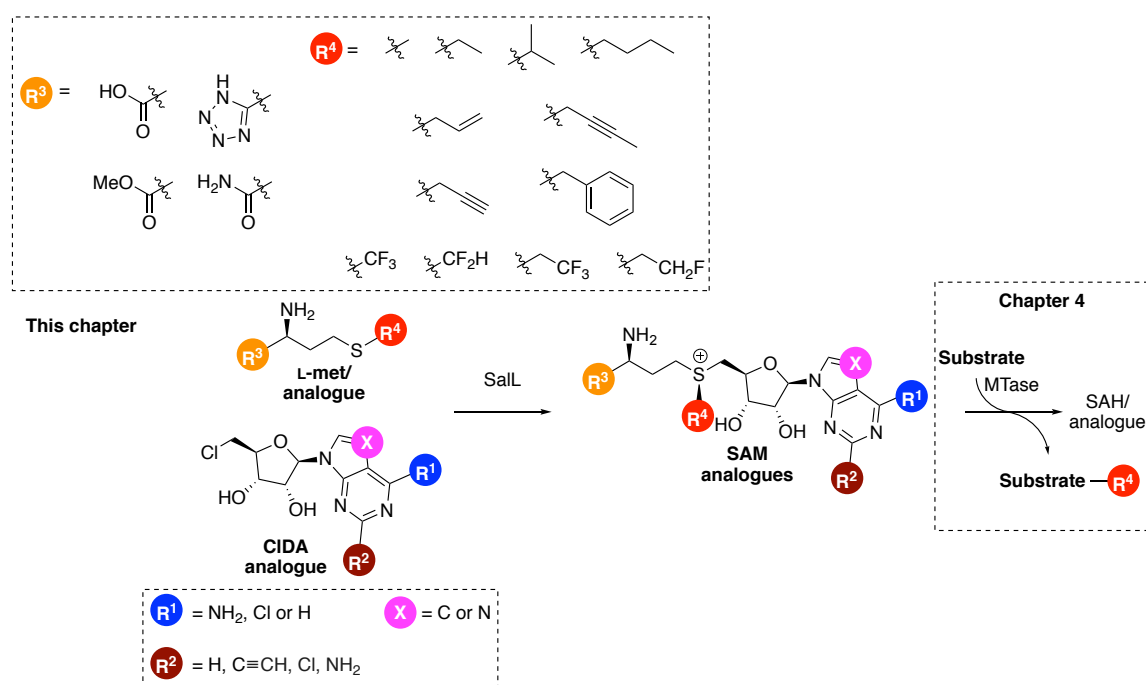
Scheme 3.1 Scope of SAM analogue formation by SalL to be explored

3.2 Aims

Specific aims for this chapter are:

(i) Synthesise L-met analogues and test with SalL/mutants for ability to form SAM analogues.

(ii) Synthesise ClDA analogues and test with SalL/mutants for ability to form SAM analogues.

(iii) Synthesise compounds for synthesis of SAM analogues with enhanced stability.

(iv) Synthesise proposed stable SAM analogues and assess stability.

(v) Combine tolerated L-met and ClDA analogues to investigate synthesis of doubly modified SAM.

3.3 Synthesis of *S*-Alkylated Methionine Analogues

Generation of alkyl analogues of L-met is typically carried out by utilisation of homocysteine as a nucleophile under basic conditions to carry out an S_N2 reaction with an appropriate alkyl bromide.[83] This strategy was employed for the synthesis of analogues **3.2a-e** (Scheme 3.2). Due to the significant expense of L-homocysteine, synthesis for an initial screen was carried out from DL-homocysteine (Sigma-Aldrich, DL-homocysteine = £49.70 per g, L-homocysteine = £2480.00 per g, 05/09/19). DL-homocysteine was alkylated at *S-* by addition of the appropriate alkyl bromide under basic conditions (Scheme 3.2). A single equivalent of alkyl bromide was used in an attempt to limit additional alkylations at either *N* or *O*. Products **3.2a**, **b** and **d** were obtained in low to moderate yields by dissolving the residue from the reaction mixture in water then adjusting the pH to ~7 after which the product was isolated by filtration. Compounds **3.2c** and **e** were isolated in moderate yields by Mass-Directed Auto Purification and semi-preparative RP-HPLC respectively.
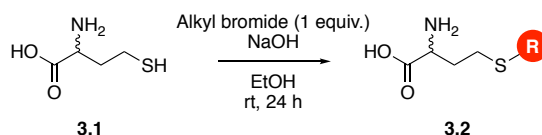
| 3.2	R	% Yield
a		11
b		68
c		36
d		28
e		66

Scheme 3.2. Alkylation of DL-homocysteine to give DL-met analogues.

Fluoroalkyl methionine analogues were synthesised following a method pioneered by Langlois[18] and further developed by Baran *et al.*, where zinc sulfinate salts were used for the

generation of alkyl radicals containing fluorine atoms.[19] Recently, the methodology of Baran has been used to carry out addition of fluoroalkyl radicals onto proteins at cysteine residues.[148] With this precedent, it was hypothesised that generation of the appropriate alkylfluoro radicals would facilitate addition on to sulfur centres. This reaction was attempted for the generation of trifluoromethyl methionine (**3.2f**) from the homocysteine dimer, homocystine (**3.3**) (Scheme 3.3). Radical addition of the CF_3 group to the sulfur results in homolytic cleavage of the disulfide bond and generation of another *S*-radical which can then form another molecule of CF_3-methione. Compound **3f** was isolated by pH controlled precipitation in good yield (88%). This reaction was repeated with different fluoroalkyl compounds, for synthesis of **3.2g** and **3.2h** however no product was isolated by alteration of pH. Instead, these compounds were both isolated in 99% yield by ion-exchange chromatography using cation exchange resin.

3.2	R	% Yield
f	CF_3	88
g	CF_3	84
h	CF_2H	83
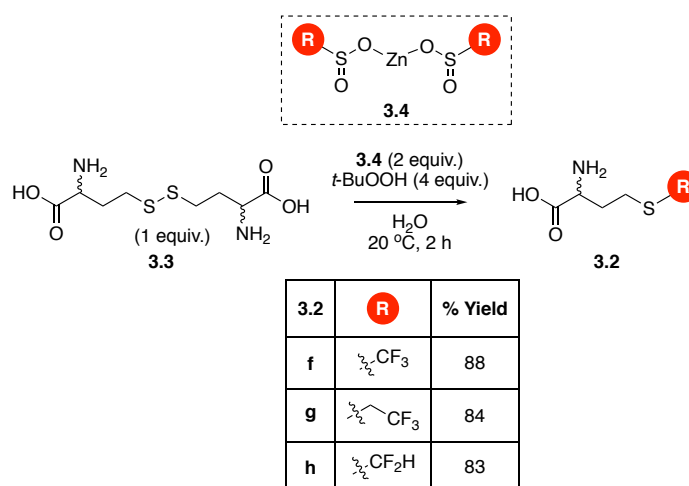
Scheme 3.3. Fluoroalkylation of homocystine *via* zinc sulfinate salts.

In addition to L-met analogues bearing different alkyl groups, a tetrazole isostere of L-met (*tet*-L-met, **3.9**, Scheme 3.4) was investigated. It has been shown that this compound can be converted to a SAM analogue in a reaction with ATP catalysed by MAT and displays enhanced stability with respect to SAM.[103] First, the carboxylic acid of Fmoc-methionine was converted to a primary amide (**3.6**) by treatment with Boc anhydride and ammonium bicarbonate, in 91% yield. Treatment of the resulting compound with a 1:1 mixture of TFA anhydride and pyridine produced nitrile (**3.7**). Following this, a [3+2] cycloaddition with sodium azide in the presence of ZnBr was carried out to provide Fmoc protected tetrazole **3.8** in 74% yield. Finally, the Fmoc group was removed using diethylamine to provide the target *tet*-L-met (**1.145**) after ion-exchange chromatography isolate *tet*-L-met as the HCl salt in 54% yield.
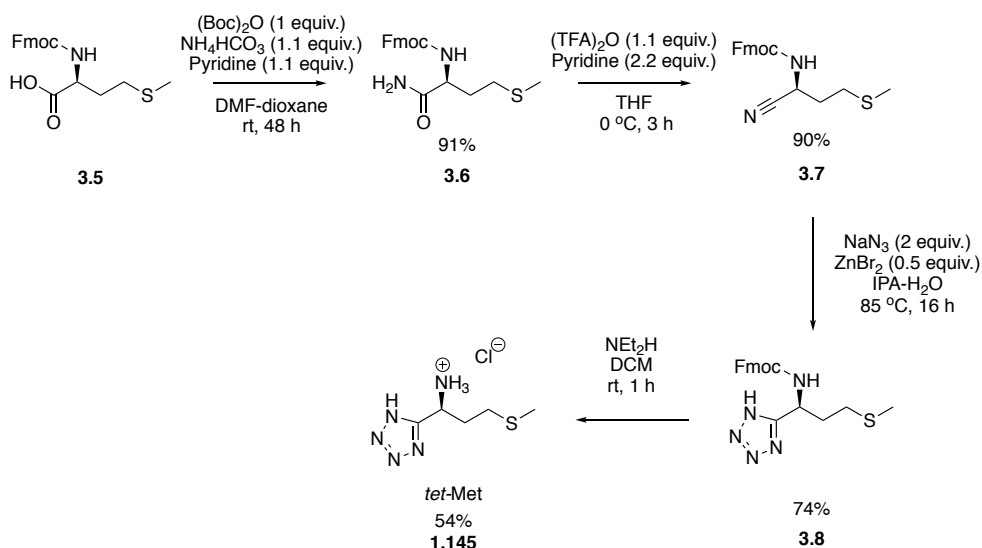
Scheme 3.4. Synthesis of *tet*-L-met analogue of L-met.

SalL has previously been shown to accept L-eth to generate *S*-adenosyl-L-ethionine (SAE).[100, 102] The tetrazole isostere of L-eth (*tet*-L-eth) was therefore prepared using the same method as for *tet*-L-met (Scheme 3.5). Initial Fmoc protection was carried out through addition of Fmoc-OSu to L-eth in aqueous sodium carbonate and acetonitrile to facilitate formation of the target material (**3.11**) in 71% yield.[149] Following this, synthesis was carried out as for *tet*-L-met, with comparable yields achieved at each step (Scheme 3.5).
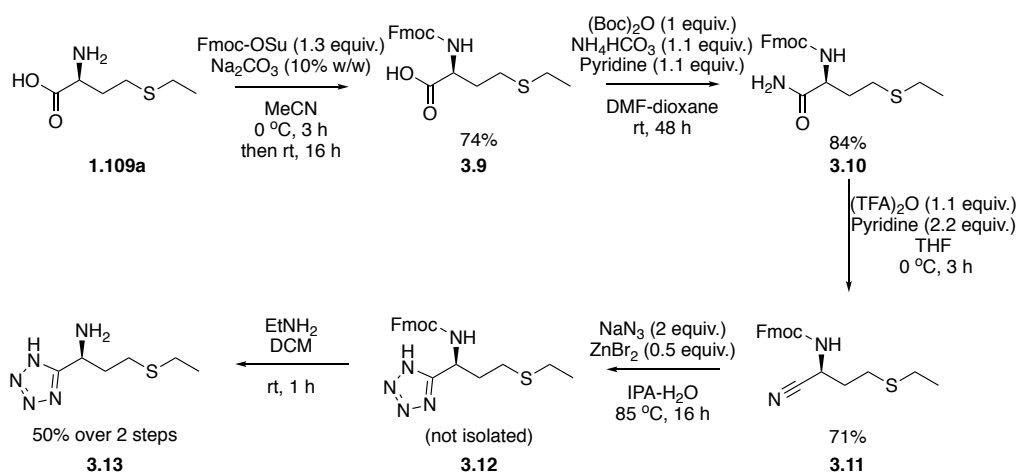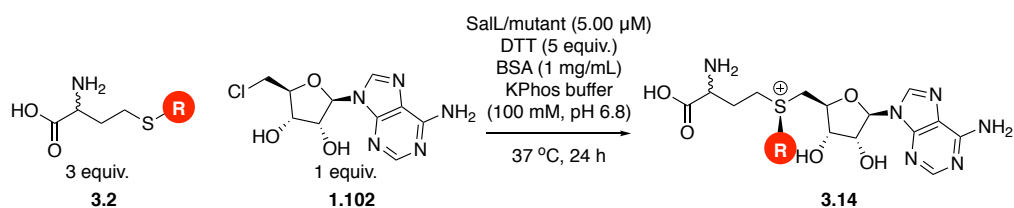

Scheme 3.5. Synthesis of *tet*-L-eth analogue of L-met.

3.4 Screening of SalL and Mutants for Activity Using Methionine Analogues

With a palette of methionine analogues and SalL mutants in hand, screening for activity was carried out (Scheme 3.6). As DL-met analogues were used rather than L-met, the number of amino acid equivalents was doubled from 1.5 equiv. to 3 equiv. from the SAM formation assays published by Sadler *et al.*[102] This would ensure that the same number of equivalents of the L-enantiomer would be present as in the L-met assays.

To allay any fears that D-amino acid analogues would inhibit SAM formation a test study was carried out where D-met was utilised instead of L-met. Conversion over 24 h under the assay conditions used previously for SAM formation was 100% with respect to ClDA. Kinetic parameters were also determined to be k_{cat} = 2.20 min^{-1} and K_M = 13.55 mM (compared with k_{cat} = 10.9 min^{-1} and K_M = 1.26 mM) (Appendix Figure S 3.1). Although turnover was reduced with respect to L-met, the 100% conversion observed confirms the D-amino acid would not be totally inhibitory to activity of the enzyme towards racemic mixtures of Met analogues. In the case of hits being observed for any of the racemic Met analogues, L-homocysteine would be used to synthesise the L-enantiomer which would then be subjected to further studies for conversion and kinetics.

Initial screening of DL-met analogues was carried out with a SalL/mutant concentration of 5.00 µM and substrate concentrations of 1 mM and 3 mM for ClDA and DL-met respectively. Reactions were sampled after 24 h, quenched by heating to 90 °C and analysed by RP-HPLC for formation of SAM analogue or MTA analogue (as in in Section 2.9.1). First, a racemate of ethionine was investigated as a substrate for SalL/mutants. In this case, Phe186Leu, Asp183Glu and wild-type SalL were observed to form *S*-adenosyl-L-ethionine (SAE) with conversions of 20%, 7% and 44% respectively. The full series of methionine analogues synthesised (**3.2a-h**) were tested alongside DL-met analogues **3.2j** and **k** which were synthesised as part of a previous study.[150] Unfortunately, no conversion was observed for any of the analogues investigated.

% Conversion

L-met analogue	R	wild-type	Phe186Leu	Trp129Phe	Asp183Glu
3.2i	ethyl/propyl	44	20	-	7
3.2a	isopropyl	-	-	-	-
3.2b	butyl	-	-	-	-
3.2c	allyl	-	-	-	-
3.2d	benzyl	-	-	-	-
3.2e	propargyl	-	-	-	-
3.2j	butynyl	-	-	-	-
3.2f	CF_3	-	-	-	-
3.2g	CF_3	-	-	-	-
3.2h	CF_2H	-	-	-	-
3.2k	CH_2F	-	-	-	-

Scheme 3.6. Reaction of ClDA with met analogues to form SAM analogues. Conversions were calculated by RP-HPLC with comparison of peak area of ClDA/SAM. Note: Only SalL mutants which maintained activity are shown in table, all mutants were screened but no conversions were observed for other constructs.

A final analogue of L-met screened for activity with SalL/mutants was L-Se-Met (**1.108**) (Scheme 3.7). SeSAM analogues have undergone significant investigation as cofactors for MTases as they have been proposed as better alkyl donors than their SAM counterparts due to the longer and weaker Se-C bond.[83] Commercially available L-Se-Met was tested as a substrate for SalL/mutants using the same conditions as the previous SAM formation assays. No conversion to SeSAM was observed.
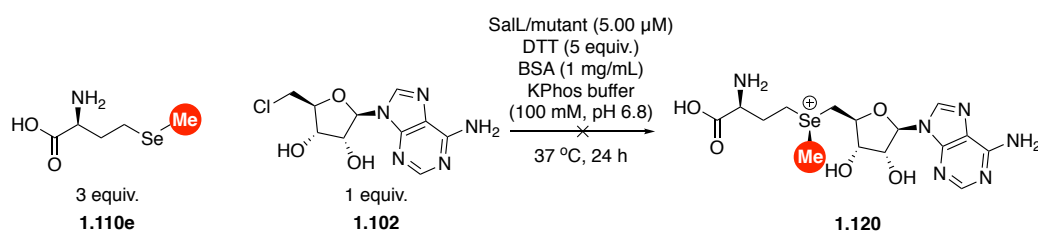
Scheme 3.7. Screening of L-Se-Met as a substrate for SalL.

Next, the tolerance of SalL/mutants to accept L-met analogues which were modified at the carboxylic acid was investigated. It was hypothesised that these compounds may confer enhanced stability to their respective SAM analogues by supressing degradation to MTA and HSL (Section 1.9). Previously synthesised *tet*-L-met (**1.145**) and *tet*-L-eth (**3.13**) were investigated alongside a primary amide (**3.15**, L-met-NH$_2$) and methyl ester analogue (**3.16**, L-met-OMe) of L-met, both of which were commercially available. Reactions were monitored as previously by following the disappearance of ClDA and the formation of SAM and related degradation products by RP-HPLC (Scheme 3.8).

	R$_1$	R$_2$	Wild-type	Phe186Leu	Trp129Phe	Asp183Glu
3.15	Me	H$_2$N (O)	37	36	0	7
3.16	Me	MeO (O)	100	100	100	100
1.145	Me	tetrazole	100	98	10	13
3.13	ethyl	tetrazole	0	0	0	0
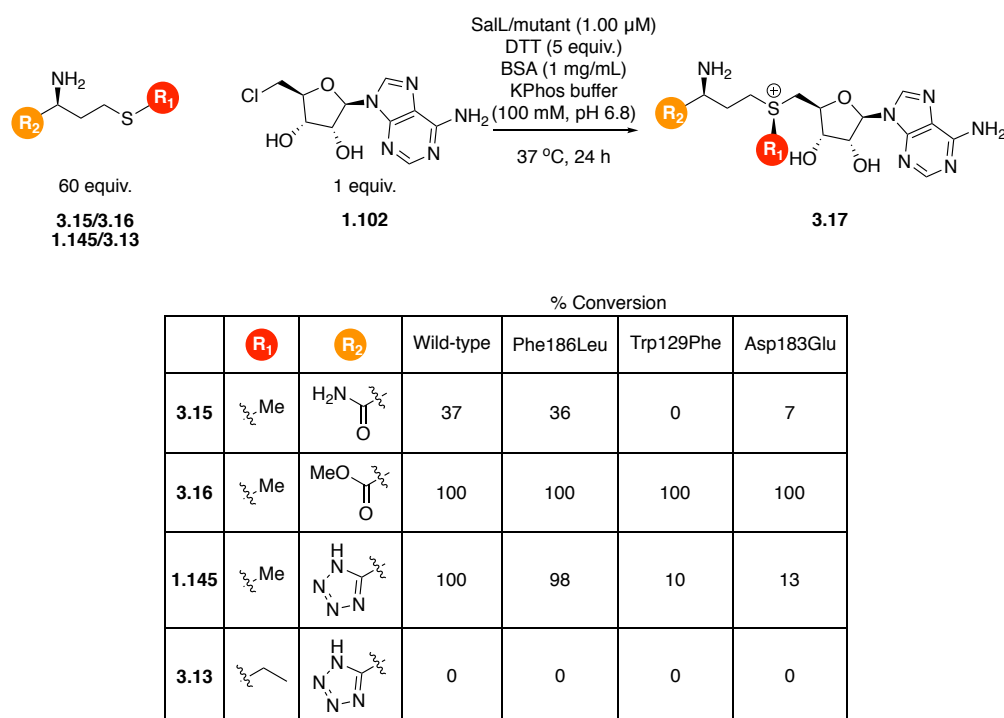
Scheme 3.8. Reaction of ClDA with L-met analogues modified at *S* and carboxylic acid to form SAM analogues. Conversions were calculated by RP-HPLC with comparison of peak area of ClDA with MTA/SAM.

Each L-met analogue showed higher conversions with either wild-type or Phe186Leu as had been observed for reaction of ClDA with L-met. Primary amide containing compound **3.15** displayed moderate conversion to its corresponding SAM analogue when compared with L-met. It also showed a significant amount of formation of MTA suggesting no enhanced stability with respect to SAM. As the conversion was significantly lower than for the natural L-met substrate no further work was carried out on this substrate. Results from methyl ester compound **3.16** were initially promising with high conversion observed across all SalL constructs tested. Upon LCMS analysis during the reaction however, it was clear that only peaks for SAM, MTA or ClDA starting material were present with no mass observed for the
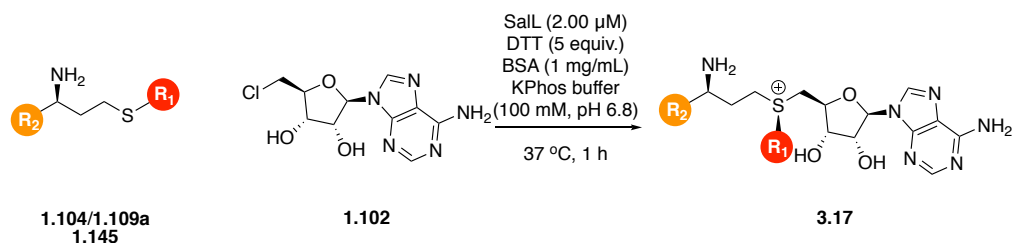
target methyl ester of SAM (Appendix Figure S 3.2). It appeared that either the methyl ester was hydrolysed *in situ* or it was removed after SAM formation. The high conversion is similar to that of L-met so suggests hydrolysis has occurred prior to reaction.

Finally, *tet* containing compounds **1.145** and *tet*-L-eth **3.13** were tested for activity with the small panel of SalL mutants. Compound **1.145** showed good conversion with both the wild-type and Phe186Leu enzymes. Conversion was significantly reduced in the case of the Trp129Phe and Asp183Glu mutants. Unfortunately, no activity was observed for **3.13** with SalL unable to tolerate the simultaneous modification of the *S*-alkyl group and carboxylic acid.

3.4.1 Enzyme Kinetics of Formation of Unnatural SAM analogues

As a result of the clear limits in SalL/mutant scope, work from this point onwards was focussed on the wild-type enzyme and the characterisation of some of the transformations it can carry out. The kinetics of SAM analogue formation were studied with respect to enantiomerically pure L-eth and *tet*-L-met (Appendix Figure S 3.3a+b, Scheme 3.9). These parameters were obtained by varying the concentration of L-met analogue at fixed ClDA and enzyme concentrations. Aliquots of the reaction were taken at t = 0, 5 mins, 10 mins, 20 mins, 30 mins, 50 mins, 60 mins and quenched with sodium formate buffer (400 mM, pH 3). The resulting samples were then analysed by RP-HPLC with the amount of SAM formed determined by comparison of the peak areas of ClDA with SAM/analogue and MTA/analogue (as in Section 2.9.2).

Scheme 3.9. Determination and comparison of kinetic parameters obtained for SalL wild-type with different L-met analogues.

The kinetic parameters showed significant increases in K_M for both unnatural L-met analogues by at least an order of magnitude. This was more pronounced for *tet*-L-met than for L-eth with K_M values increased to 35.2 mM and 20.7 mM respectively compared with 1.26 mM for L-met. k_{cat} values also suffered for both substrates as they decreased from 10.9 min^{-1} for L-met to 3.40 min^{-1} and 1.35 min^{-1}, respectively. The substrate modifications made in this case are relatively modest but the large decrease in favourable Michaelis-Menten properties exemplifies the limited tolerance of SalL. From these results, it is unsurprising that L-met analogues featuring much larger *S*-alkyl modifications are not tolerated by SalL. However, on a positive note, successful generation of L-eth and *tet*-L-met containing SAM analogues allows for later investigation of these compounds as cofactors for use with MTases.

3.5 Synthesis of ClDA Analogues

A range of ClDA analogues were targeted based on work initiated by the group of O'Hagan and further explored by the Ang group which showed these compounds to be substrates for either FDAS[92, 140] or *flA1*.[94-96] A BLAST search of FDAS and *flA1* reveals homology of 35% and 37% respectively with SalL. From this finding, the substrates which could be turned over by the fluorinase enzymes were synthesised to investigate if they were also substrates for SalL.
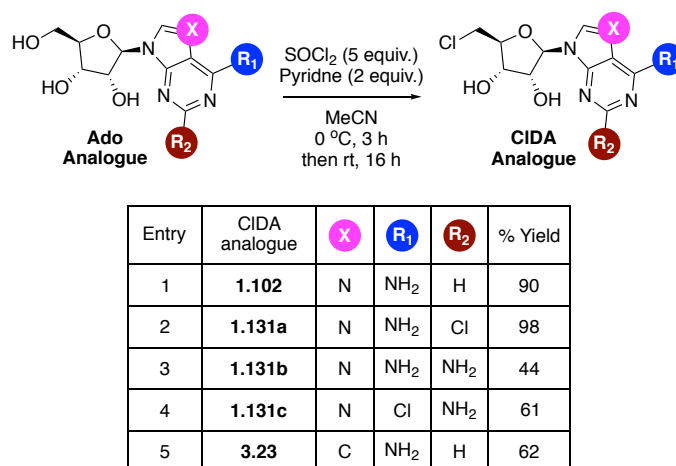
Synthesis of alkyne containing ClDA analogue ClDEA (**1.125**) was carried out according to the literature procedure of Thompson *et al.* (Scheme 3.10).[140] Starting from 6-chloroguanineriboside (**3.18**), acetate protection of the hydroxyl groups was carried out using acetic anhydride with the protected sugar (**3.19**) formed in 89% yield. After this, iodination of **3.19** was carried out to provide compound **3.20** in 60% yield. Compound **3.21** was formed through simultaneous nucleophilic aromatic substitution of the chloride group with NH_2 from ammonia and deprotection of the acetate groups in 71% yield. A Sonogashira reaction was carried out to install the desired alkyne and provide compound **3.22** in 63% yield. Finally, the target material was obtained by simultaneous chlorination and TMS deprotection to provide ClDEA (**1.125**) in 53% yield.
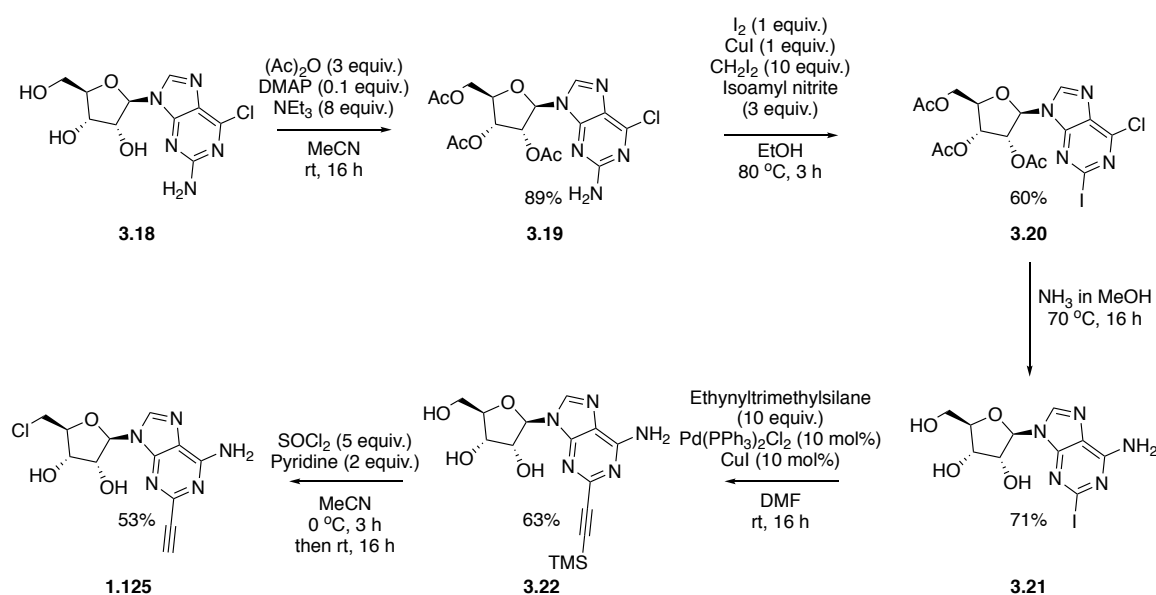
**Scheme 3.10.** Synthesis of alkyne-containing ClDA analogue ClDEA.

The Ang group carried out the synthesis of a range of 2- and 6-modified ClDA analogues following a general procedure of glycosylation of a modified nucleobase of interest onto an acetate protected ribose.[96] The Ado analogues of the precursors to the ClDA analogues depicted in Scheme 3.11 were commercially available. It was hypothesised synthesis could be carried out through the same route as ClDA from Ado by $SOCl_2$ mediated chlorination of the target Ado analogue.[142] This was attempted for Entries 1-4 with moderate to excellent yields achieved. Synthesis of 7-Dz-ClDA (**3.23**) was carried out to investigate whether SalL was capable of converting this compound to the corresponding SAM analogue. This SAM analogue had previously been shown by Huber *et al.* to possess enhanced stability with respect

to natural SAM.[103] This compound was synthesised from commercially available 7-Deazaadenosine (dzAdo) in moderate yield (Entry 5).

| Entry	CIDA analogue	X	R$_1$	R$_2$	% Yield
1	**1.102**	N	NH$_2$	H	90
2	**1.131a**	N	NH$_2$	Cl	98
3	**1.131b**	N	NH$_2$	NH$_2$	44
4	**1.131c**	N	Cl	NH$_2$	61
5	**3.23**	C	NH$_2$	H	62

Scheme 3.11. Synthesis of ClDA and analogues by chlorination of a commercial starting material.

ClDA analogues bearing a carbonyl group at the 6-position were obtained by repetition of the chlorination methodology (Scheme 3.12). Compound **1.131i** (Entry 1) was obtained in low yield due to the necessity to carry out an additional purification of the target material by semi-preparative HPLC after column chromatography which was carried out for isolation of the other ClDA analogues. As only a small amount of compound was required for enzymatic screening no further work was carried out to optimise the isolation procedure. Analogue **3.24** (Entry 2) was synthesised by the same method and was purified by column chromatography to provide the target compound in 69% yield. To complete synthesis of the desired Ade modified ClDA analogues, compound **1.131d** was achieved by hydrogenation of ClDA analogue **1.131c** in 63% yield (Scheme 3.13).

| Entry	CIDA analogue	R$_2$	% Yield
1	**1.131i**	NH$_2$	1*
2	**3.24**	H	69
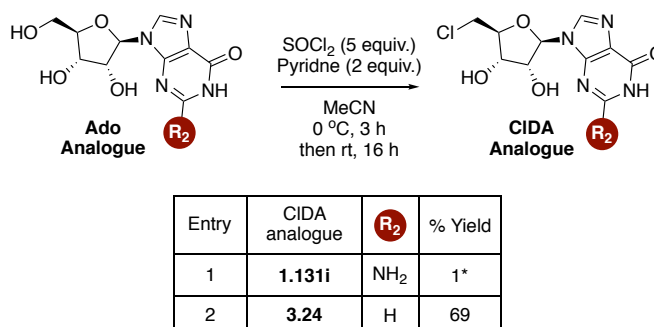
Scheme 3.12. Synthesis of ClDA and analogues by chlorination of a commercial starting material. *isolated by semi-preparative HPLC.
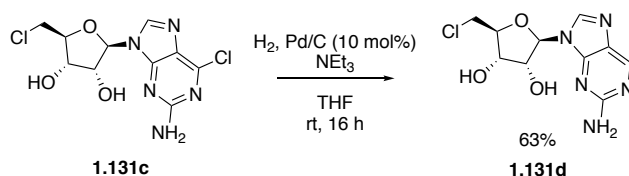
Scheme 3.13. Hydrogenation of **1.131c** to synthesise ClDA analogue **1.131d**.

To complete the analysis of molecular determinants of the ClDA cofactor for SAM formation, ClDA analogues lacking the 2′- and/or 3′-hydroxyl groups were prepared. These were 2′-DeoxyClDA (**3.25**), 3′-DeoxyClDA (**3.27**) and 2′, 3′-DideoxyClDA (**3.29**) (Scheme 3.14). Each compound was synthesised in moderate yield providing enough material to carry out subsequent enzymatic screening.
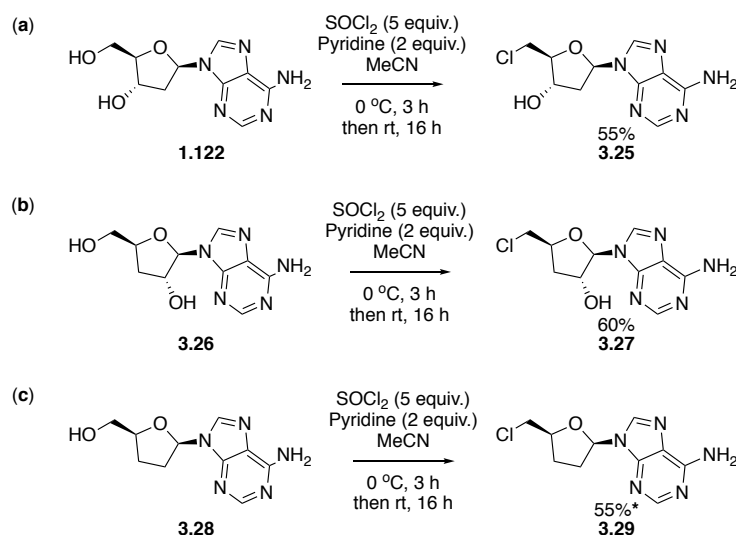
Scheme 3.14. Synthesis of **a.** 2′-DeoxyClDA, **b.** 3′-DeoxyClDA and **c.** 2′, 3′-DideoxyClDA. *isolated by semi-preparative HPLC.
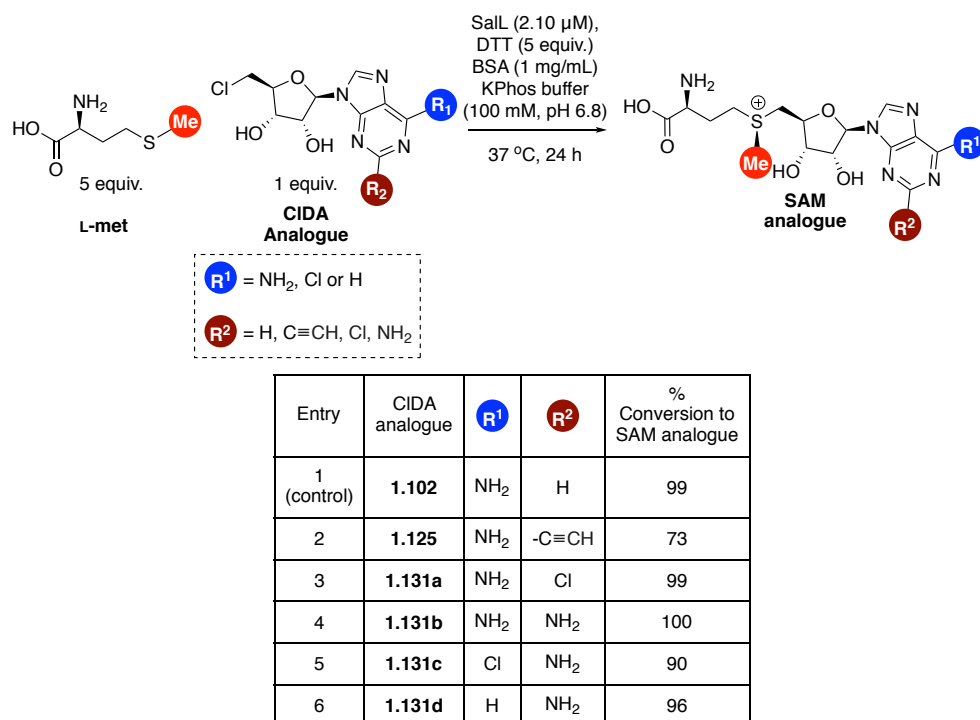
3.6 Screening of SalL for Activity Using ClDA Analogues

With a set of ClDA analogues prepared, their ability to form SAM analogues catalysed by SalL was tested. Initial trials focussed on determining kinetic parameters for SalL with respect to ClDA concentrations. However, difficulties were encountered due to sensitivity of the RP-HPLC assay being lower than necessary in light of the low K_M value of ClDA (2×10^{-4} mM),[97] it was decided that the most effective way to screen these substrates would be to measure conversions at intervals up to 24 h.

Conversions for each modified substrate were determined in the same manner as the initial screening of the SalL mutants. Each reaction was set up on 500 µL scale with ClDA/analogue (400 µM), L-met (2.00 mM) and SalL (2.10 µM). Each reaction was then sampled at various timepoints over 24 h and quenched by heating. Conversions were then calculated by comparison of SAM/MTA peak area with that of the modified nucleoside starting material. In some cases, heating of the reaction led to depurination of the MTA analogue generated from SAM analogue degradation. For these reactions conversion was calculated by comparing peak area of starting material with that of SAM/MTA/Ade analogue formation.[103]

Under the assay conditions, ClDA (**1.102**) was observed to be converted to SAM within 24 h (Entry 1, Scheme 3.15), with time course measurements of the reaction showing it to be complete after 4h (Appendix Figure S 3.3). ClDA analogues modified at the 2-position (Entries 2-4, Scheme 3.15) such as ClDEA (**1.125**), 2-Cl-ClDA (**1.131a**) and 2-Amino-ClDA (**1.131b**) were readily converted to their corresponding SAM analogues. Compound **1.125** showed conversion of 74% after 24 h (Entry 2), while **1.131a** and **1.131b** displayed conversions of 99% and 100% respectively (Entries 1 and 3). Investigation of the time course of these reactions showed full conversion after 8 h for **1.131a** and after 4 h for **1.131b** (Appendix Figure S 3.3)

Interestingly, some 6-modified ClDA analogues were successfully converted to their corresponding SAM analogues (Entries 5-6, Scheme 3.15). 6-Chloro-2-amino-ClDA (**1.131c**) reached 90% conversion over 24 h, while 2-Aminopurine ClDA analogue (**1.131d**) achieved 96% conversion. Taken collectively, the 6-amino group is not essential for SAM formation.

| Entry	CIDA analogue	R¹	R²	% Conversion to SAM analogue
1 (control)	**1.102**	NH₂	H	99
2	**1.125**	NH₂	-C≡CH	73
3	**1.131a**	NH₂	Cl	99
4	**1.131b**	NH₂	NH₂	100
5	**1.131c**	Cl	NH₂	90
6	**1.131d**	H	NH₂	96

Scheme 3.15. Reaction of ClDA analogues with L-met to form SAM analogues.

However, the presence of a carbonyl group at the 6-position is not tolerated (Scheme 3.16a+b). For both guanine (**1.131i**) and hypoxanthine (**3.24**) containing ClDA analogues no reaction was observed over 24 h. This may be caused by the electron withdrawing nature of the carbonyl group negatively influencing other interactions the nucleobase is involved in. One such interaction is that of the N7 with Asn188, the importance of which was probed by analysis of 7-DzClDA (**3.23**) (Scheme 3.16c) for SAM formation. Only ≈ 10% conversion was observed and depurination was present in the t = 0 sample, a feature not observed for any of the other ClDA analogues. Not only does this suggest poor stability of 7-Dz-ClDA under the assay conditions, it also confirms that the N7 plays an important role in binding of the ClDA analogue (Figure 3.1).
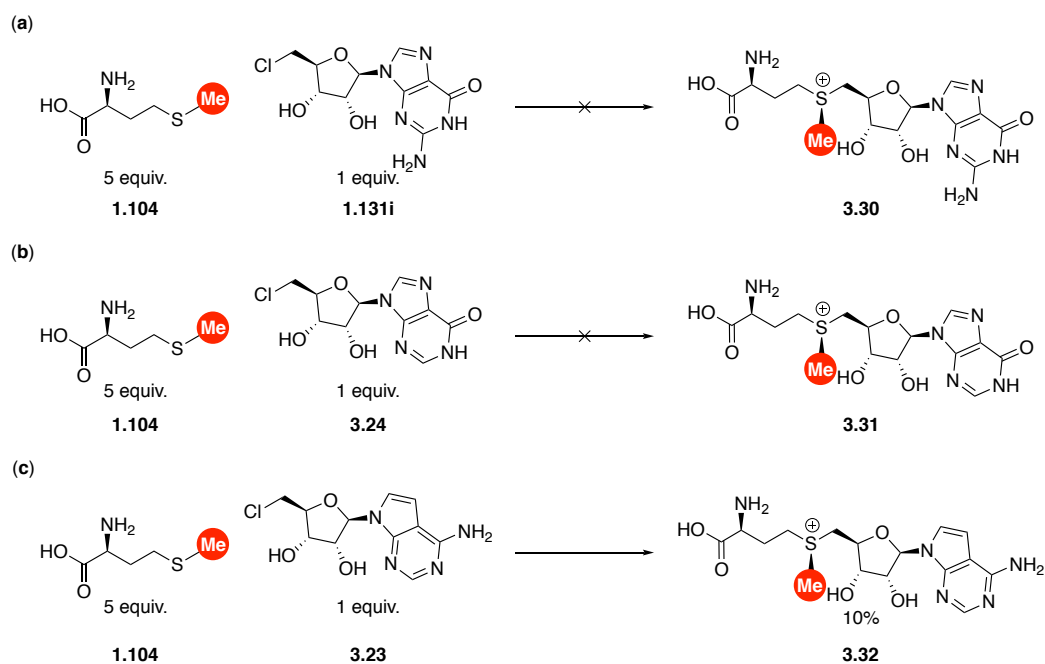
Scheme 3.16a. Attempted synthesis of SAM analogues containing **a.** Guanine. **b.** Hypoxanthine or **c**. 7-dz-Ade nucleobases.
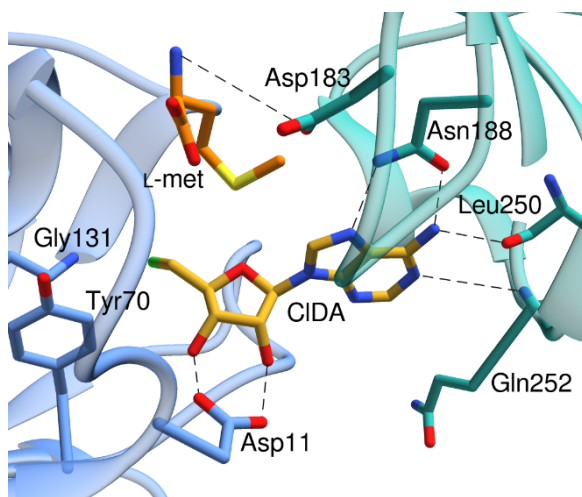
Figure 3.1. Hydrogen bonding interaction of ClDA Ade with Asn188 and ribose hydroxyls with Asp11. PDB Accession code: 2Q6I.

The same reaction conditions as in Scheme 3.15 were employed to attempt synthesis of SAM analogues bearing modification to the sugar. No reaction was observed for any of the ClDA analogues in which the hydroxyl groups of the sugar were modified (Scheme 3.17) (**3.25**, **3.27**, **3.29**). This is likely due to disruption of a hydrogen bonding interaction between Asp11 and these groups (Figure 3.1).
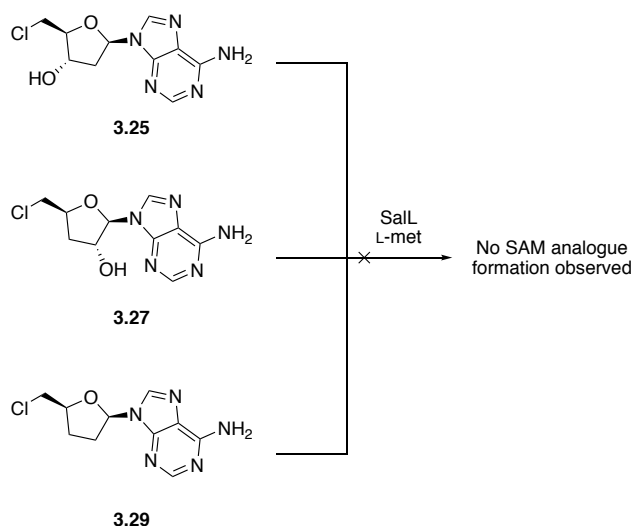
Scheme 3.17. Attempted reaction of sugar modified ClDA analogues with L-met .

3.7 Combining Unnatural Substrate Pairings

After establishing a range of unnatural substrates which can be accepted by SalL, it was desired to test the ability of the enzyme to synthesise SAM analogues which contain a methionine analogue and a ClDA analogue together. This was investigated for combinations of 2-Cl-ClDA (**1.131a**) and 2-Amino-ClDA (**1.131b**) with L-eth (**1.109a**) and *tet*-L-met (**1.145**). Each reaction was studied under the same conditions as the assays from the previous section with conversions obtained at various time points up to 24 h (Appendix Figure S 3.5, Scheme 3.18).

For ClDA (**1.102**, Entries 1, 4 and 7, Scheme 3.18), conversions are reduced when L-met is replaced with L-eth from 99% to 42%. When **1.145** is used the conversion to *tet*-SAM (**1.147**) is 99%. Compound **1.131a** (Entries 2, 5 and 8, Scheme 3.18) displays a similar reduction in conversion from 99% when **1.104** is utilised to 37% for **1.109a**. Unlike in the case of **1.102**, a reduction in conversion to 41% *tet*-2-Cl-SAM (**3.35**) is observed when **1.145** is utilised. 2-Amino-ClDA (**1.131b**, Entries 3, 6 and 9, Scheme 3.18) displayed the highest conversion when **1.109a** was used (78%). When **1.145** was utilised 99% conversion to *tet*-2-Amino-SAM (**3.36**) was observed.

R¹ = NH_2, Cl or H

R² = H, C≡CH, Cl, NH_2

R³ = HO or tetrazole

R⁴ = Me or Eth
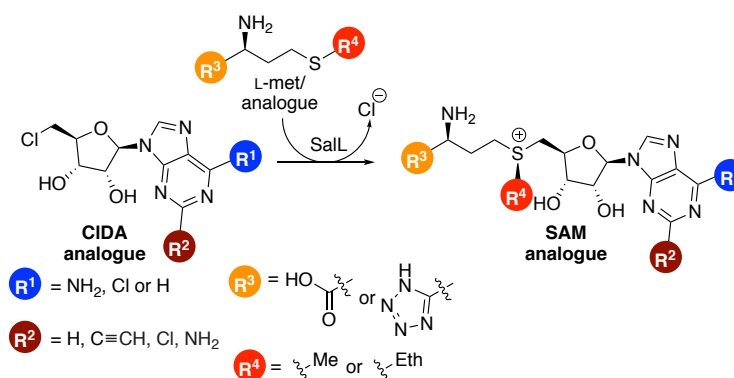
Entry	CIDA Analogue no.	L-met Analogue no.	SAM Analogue No.	R¹	R²	R³	R⁴	% Conversion
1 (control)	1.102	1.104	1.63	NH_2	H	HO-C(=O)	Me	99
2	1.131a	1.104	1.132a	NH_2	Cl	HO-C(=O)	Me	99
3	1.131b	1.104	1.132b	NH_2	NH_2	HO-C(=O)	Me	100
4	1.102	1.109a	1.111a	NH_2	H	HO-C(=O)	Eth	42
5	1.131a	1.109a	3.33	NH_2	Cl	HO-C(=O)	Eth	37
6	1.131b	1.109a	3.34	NH_2	NH_2	HO-C(=O)	Eth	78
7	1.102	1.145	1.147	NH_2	H	tetrazole	Me	99
8	1.131a	1.145	3.35	NH_2	Cl	tetrazole	Me	41
9	1.131b	1.145	3.36	NH_2	NH_2	tetrazole	Me	99

Scheme 3.18. Conversion of ClDA and L-met analogues to the corresponding SAM analogue.

3.8 Stability of *tet*-SAM Analogues in Aqueous Buffers

An assessment of the stability of *tet*-SAM was carried out in comparison with SAM. To do this, both compounds were isolated by semi-preparative HPLC and were then redissolved in the reaction buffer, potassium phosphate (100 mM, pH 6.8), in HPLC vials. Repeated injections were taken from these vials for 24 h with the amount of SAM remaining calculated by comparison of the SAM peak area with MTA peak area. In this time, the amount of SAM (**1.63**) remaining was 91% while 99% of *tet*-SAM (**1.147**) remained (Entries 1+4, Figure 3.2). Rates of degradation for 2-Cl-SAM (**1.132a**), 2-Amino-SAM (**1.132b**) and their respective

tetrazole analogues (**3.35** and **3.36**) were measured using the same method. Compounds **1.132a** and **1.132b** had degraded to 90% and 93% after 24 h (Entries 2+3, Figure 3.2) while the corresponding tetrazole compounds remained in 97% and 98% respectively (Entries 5+6, Figure 3.2, Appendix Figure S 3.6).

Entry	SAM Analogue No.	R¹	R²	R³	R⁴	% Remaining after 24 h
1 (control)	**1.63**	NH_2	H	HO–C(=O)–	Me	91
2	**1.132a**	NH_2	Cl	HO–C(=O)–	Me	93
3	**1.132b**	NH_2	NH_2	HO–C(=O)–	Me	90
4	**1.147**	NH_2	H	tetrazole	Me	99
5	**3.35**	NH_2	Cl	tetrazole	Me	97
6	**3.36**	NH_2	NH_2	tetrazole	Me	98

Figure 3.2. Degradation of SAM analogues and their corresponding tetrazole containing derivatives.

3.9 Scale-up of Tetrazole SAM Synthesis

In light of the enhanced stability observed for *tet*-SAM with respect to SAM and the other SAM analogues, the ability of SalL to prepare *tet*-SAM on a semi-preparative mg scale was investigated. If *tet*-SAM could be utilised by MTases it could be prepared on scale, purified and stored until it was required to be used. This would be in contrast to SAM which degrades readily and generally needs to be stored at -20 °C. SAM is readily utilised in a range of enzymatic kits, therefore a stable SAM analogue which displayed utility to MTases could find use in such kits as an alternative to SAM with an extended shelf-life.

Initially, the concentrations used for *tet*-SAM (**1.147**) synthesis on a small-scale were maintained and scaled up to ~ 20 mg of ClDA (**1.102**) to investigate whether these conditions provide good conversion to the target material. A first attempt (Entry 1, Scheme 3.19) showed these conditions do not maintain high conversion to **1.147** at this scale, with only 30% observed after 24 h. Although the reaction did continue to 35% after 48 h increased amounts of *tet*-L-met (**1.145**) and SalL were investigated as ideally this process would be completed over a maximum of 24 h. Increase to 2 equivalents of *tet*-L-met provided a slight increase to 60% conversion after 48 h (Entry 2). A three-fold increase in enzyme loading resulted in a modest increase to 65% conversion over 48 h (Entry 3). Increasing the number of equivalents of **1.145** to four while trialling the same enzyme loading as in Entry 2 brought about a significant improvement in conversion to 95% over 48 h. Interestingly, maintaining these conditions but doubling the enzyme loading did not increase conversion any further (Entry 5). Finally, conversion could be brought to 95% over 24 h by use of 8 equivalents of **1.145** although clearly these conditions waste significant amounts of synthesised amino acid.
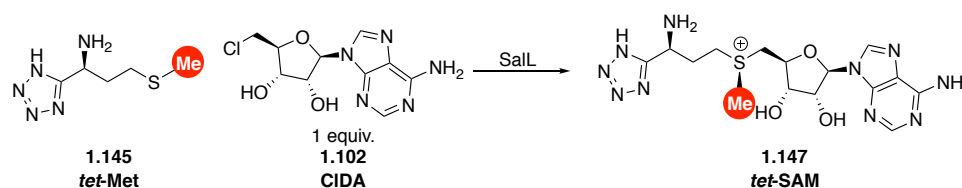
			% Conversion	
Entry	1.145 (equiv.)	[SalL]	24 h	48 h
1	1.32	840 nM	30	35
2	2	935 nM	-	60
3	2	2.81 μM	50	65
4	4	935 nM	66	95
5	4	1.87 μM	65	93
6	8	935 nM	95	100

Scheme 3.19. Optimisation of *tet*-SAM synthesis by variation of *tet*-L-met equivalents and SalL concentration.

3.10 Summary and Future Work

The synthesis of a range of Met and ClDA analogues has been displayed in this chapter with these compounds then tested as substrates for SalL. The results from this chapter have shown that there are a number of restrictions around promiscuity of SalL to modified L-met analogues, particularly around attempting to install larger alkyl groups at sulfur with only an ethyl analogue tolerated. Attempts to increase the volume of the binding pocket of the enzyme and therefore acceptance of bulkier analogues of L-met through mutagenesis was unsuccessful.

This is largely a result of being unable to predict areas of sensitivity to modification within the active-site which can maintain catalytic activity. From the results discussed in Chapter 2, it appears that an appropriate solution to this problem may be to use directed evolution until a hit is observed for a larger L-met analogue. This process could be carried out by epPCR once an appropriate screen such as utilisation of a chloride sensor had been successfully developed.

The carboxylic acid position of L-met is more susceptible to modification as displayed by acceptance of a synthesised L-met analogue where the carboxylate is replaced by a tetrazole. A primary amide containing L-met analogue is also accepted although cyclisation to a lactam and MTA appears to occur readily. L-met-OMe is converted to SAM, although LCMS evidence suggests that it hydrolyses to the free acid in solution. Assessment of stability of *tet*-SAM was assessed and showed a significant enhancement when compared with natural SAM (<1% degradation vs ~8% over 24 h).

In addition to this, promiscuity of SalL towards 2-modified ClDA analogues has been demonstrated. An area for further investigation would be to find the limit of which modifications may be tolerated at this position. Introduction of fluorine at this position would provide insight into how a non-sterically demanding, electron-withdrawing substituent may influence reactivity. It would also be interesting to probe larger groups to find the limit of tolerance at this position. Conversion is reduced for a cofactor such as ClDEA, therefore synthesis of either a 2-ethyl or 2-propyl group would prove whether or not the alkyne is the longest group which SalL will accept. Also, a number of 2-modified SAM analogues have been demonstrated to form SAM analogues which also contain a modification to the amino acid portion of the molecule. ClDA analogues containing 2-Cl- and 2-Amino-ClDA modifications were shown to form SAE and *tet*-SAM in conversions comparable to the formation of the corresponding SAM analogues formed by ClDA.

Phe186Leu was tested against a range of ClDA analogues and showed comparable reactivity to wild-type SalL. It may be interesting to investigate the structure of this mutant to observe if any changes in binding are caused by the mutation. Phe186 lies very close to the nucleobase of ClDA therefore modification to the enzyme at this position may alter how the substrate interacts with the active site.

Finally, optimisation of *tet*-SAM formation on a semi-preparative scale has been carried out. This would allow for generation of quantities of *tet*-SAM which can be stored and used as needed with degradation shown to be much less than for native SAM. Future studies would involve the investigation of this SAM analogue as a cofactor for a range of MTases.

Chapter 4

# Enhancing the Scope of Small Molecule *C*-alkylation using SAM Analogues

Abstract

SAM analogues generated *in situ* were tested as substrates of NovO to determine whether or not these compounds could retain their ability to donate a (m)ethyl group to coumarin scaffolds. Firstly, NovO and MTAN were expressed and purified, tested for activity and then utilised in a tandem SAM generation/(m)ethylation assay to alkylate some known substrates of NovO. Once this process had been established, a set of new coumarin molecules with varied electronic properties were synthesised to test whether these compounds were substrates for NovO.

Initally, SAM analogues were screened against a known substrate of NovO to identify which could be accepted by the MTase. NovO displayed a preference for 2-modified SAM analogues over natural SAM. This observation was exploited to increase conversions of known substrates of NovO and to expand the scope of NovO to previously unknown substrates. These included a precursor to an Hsp90 inhibitor and 7-hydroxywarfarin, a metabolite of clinically approved drug, warfarin. In future, it is hoped that tandem SAM/analogue generation/alkylation methodology could be applied to late-stage functionalisation of molecules of interest within the pharmaceutical industry.

4 Enhancing the Scope of Small Molecule *C*-alkylation using SAM Analogues

4.1 Current Substrate Scope of NovO

NovO has been the subject of a number of investigations for the enzymatic alkylation of aromatic small molecules, which has then been termed "Biocatalytic Friedel-Crafts Alkylations".[80] These have included work by the group of Gruber in the utilisation of chemically synthesised SAM analogues to alkylate coumarin substrates and to methylate dihydroxynaphthalenes. NovO has also been shown by Sadler *et al*. to utilise SAM (analogues) generated *in situ* by SalL to methylate, ethylate and deuterium/^{13}C label dihydroxynapthalenes and coumarins.[102] Throughout both studies, the scope of the coumarin substrates explored for NovO has been confined to coumarins (**1.93**, **1.97**, **4.1**-**4.2**) and dihydroxynaphthalenes (**4.3**-**4.4**, **1.138**) (Figure 4.1).

.
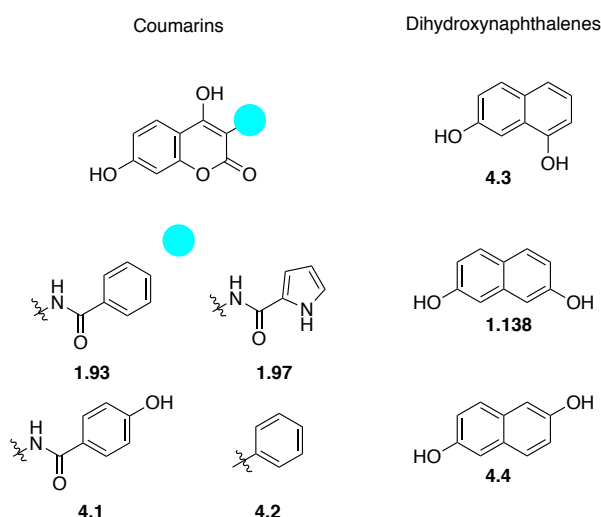
Figure 4.1. Current scope of coumarins and dihydroxynaphthalene compounds towards which NovO has shown activity.

In Chapter 3, a number of SAM analogues which were synthesised by SalL from ClDA analogues were identified. *S*-alkylated analogues have been previously investigated,[100] however SAM analogues modified at the nucleobase have not been tested as cofactors for MTases.[96] Based on structural observations of the NovO active site, it was desired to probe the modified cofactors and investigate the impact these modifications have on the efficiency of the (m)ethylation reaction. To achieve this goal, overexpression of NovO and of MTAN was

required. MTAN has been shown to enhance the yield of alkylation in previous studies by catalysing the breakdown of byproduct SAH which is a known inhibitor of NovO.[102]

4.2 Properties of Coumarin Compounds

Coumarins are naturally occurring compounds which contain the 2*H*-chromen-2-one (**4.5**) structural motif (Figure 4.2). This family of compounds have undergone study due to a variety of useful photophysical and biological properties observed for this motif. Interest in coumarins has led to the deposition of over 1000 structures containing this motif in the Cambridge Structural Database (CSD).[151] The photophysical properties of these compounds can be modulated by different substitutions to the coumarin ring. This has given rise to a large range of coumarin-based dyes[152] which have found application, for example, in optical brighteners[153] and laser dyes.[154] Further to this, these compounds have undergone many studies as fluorescent sensors for a variety of anions of environmental or biological interest such as cyanide,[155] fluoride[156] or chloride.[157]
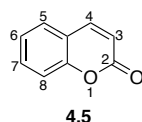

Figure 4.2. Nomenclature of the coumarin scaffold.

The coumarin motif is also found in a variety of naturally occurring antibiotics such as clorobiocin, coumermycin or novobiocin (**1.92c**) and has a number of potential uses as anti-viral or anti-inflammatory agents.[158, 159] For example, studies of **1.92c** have shown it to be an inhibitor of proteins such as Hsp90 which has been implicated in a number of cancers.[160] Structurally similar compounds to **1.92c** have also been investigated as Hsp90 inhibitors (**4.6**) by the group of Blagg[161] with the 3-position modified by appending a triazole containing compound to the 3-amino group.

Novobiocin
1.92c
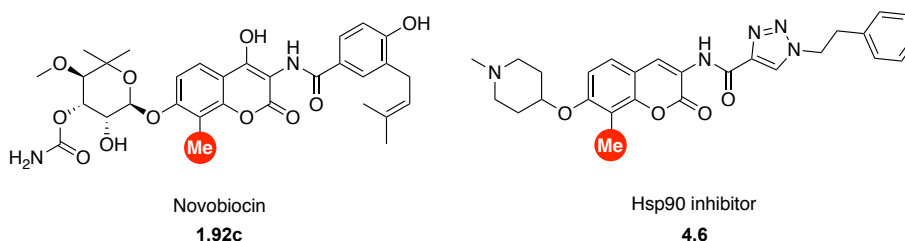
Hsp90 inhibitor
4.6

Figure 4.3. Structures of novobiocin (**1.92c**) and an Hsp90 inhibitor (**4.5**).

Further to this, perhaps the best known coumarin containing compound is the clinically-approved drug warfarin (**4.7**, Figure 4.4).[162] This compound was discovered in 1944 by Ikawa *et al.*[163] and is an anti-coagulant. Warfarin is administered as a racemate and is used as a treatment for preventing strokes in sufferers of atrial fibrillation.[164] In 2007, it was the most prescribed oral anticoagulant in the USA.[162] Within the body, warfarin is readily metabolised by Cytochrome P450 enzymes through oxidation of the aromatic ring.[165] A commonly observed metabolite is 7-hydroxywarfarin (**4.8**), which is similar in structure to known coumarin substrates of NovO, with 4- and 7-hydroxy substitution present in both.
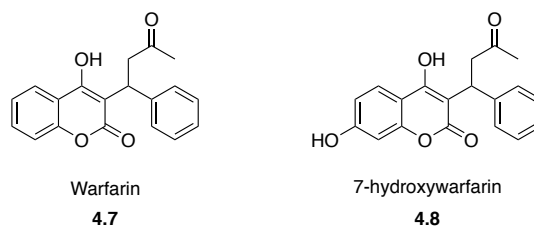
Figure 4.4. Structures of warfarin (**4.7**) and metabolite 7-hydroxywarfarin (**4.8**).

4.3   Design of Novel Coumarins for NovO (m)ethylation

Retrosynthesis of coumarin **1.93** begins with breaking of the amide bond in **1.93** and highlights a point of diversification for straightforward synthesis of 3-substituted coumarin analogues. The two fragments obtained are benzoic acid, **4.9**, and aminocoumarin precursor **4.10**. With the synthesis of **4.10** established in the literature,[80] synthesis of a diverse range of coumarin analogues can be achieved by carrying out amide couplings to this core of a variety of different benzoic acid analogues.
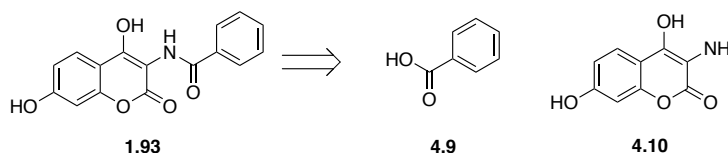

Scheme 4.1. Retrosynthesis of coumarin to starting materials **4.9** and aminocoumarin **4.10**.

A set of target coumarins was devised through analysis of a range of commercially available benzoic acid derivatives (Figure 4.5). Each compound contains either different physical or electronic properties to the coumarins which have been alkylated by NovO. Compound **4.11** contains a phenyl ring which has a 2-methyl group present which may impact the planarity of the coumarin and therefore how it fits in the binding pocket of NovO. The phenyl ring of **4.12** contains an electron-donating group and provides some extra steric bulk not observed in previously studied coumarins. Compound **4.13** has a pyridyl motif which has different

electronic properties to a phenyl ring due to the electron withdrawing nature of the N atom. Coumarins **4.14** and **4.15** also contain electron withdrawing groups through the presence of a single F atom and a CF_3 group respectively.

Further to this, it was desired to test whether NovO could carry out methylation of a 4-hydroxy analogue of a precursor to Hsp90 inhibitor **4.16**. This was deemed to be of particular interest due to the extended bulk present in the 3-position of this substrate in comparison to the previously investigated coumarin substrates. Finally, due to the similarity of 7-hydroxywarfarin (**4.8**) to the previously accepted coumarin substrates it was desired to test this drug metabolite as a substrate for NovO. Taken collectively, this set of coumarins presents a range of diverse substrates to NovO and would allow for an effective investigation of previously unexplored substrate space for this MTase.
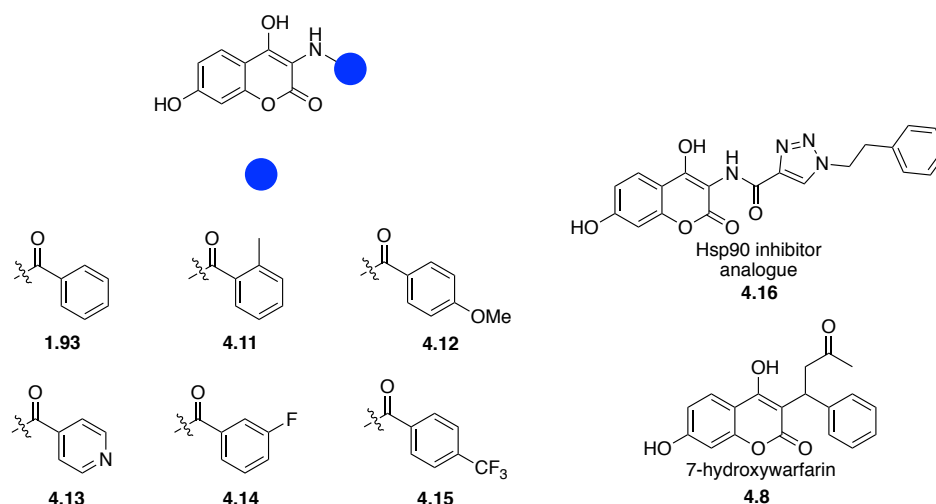
**Figure 4.5.** Set of target coumarins to probe scope of NovO for alkylation of coumarins.

4.4 Aims

Specific aims for this chapter are:

(i)   Overexpress and purify NovO and MTAN.

(ii) Synthesise a range of coumarins substituted at the 3-position.

(iii) Using previously established tandem SalL/NovO process, investigate scope of coumarins   which can be alkylated using SAM analogues generated *in situ* by SalL.

(iv) Investigate activity of NovO towards *in situ* generated SAM analogues.

(v)   Apply methodology for the *C*-alkylation of coumarins with potential medicinal relevance.

## 4.5 NovO and MTAN Overexpression

To carry out the methylation assays, overexpression of NovO and MTAN was required. NovO was expressed following the protocol of Sadler *et al.*[77] BL21 (DE3) *E. coli* were transformed with a pET26b(+) plasmid containing the gene for NovO and kanamycin resistance. After selection for the desired transformants on agar containing kanamycin, colonies were picked and starter cultures were grown in LB media. After 16 hours of cell growth, these starter cultures were transferred to 1 L of Magic Media containing 50 mL of component B (glycerol)[150] and were grown at 30 °C until an OD of 2.8 was reached. After this the culture was incubated overnight at 18 °C. The desired protein was then isolated by affinity chromatography in a similar fashion to SalL (Section 2.6) (Figure 4.6). The only difference in this case was that NovO has previously been observed to degrade at room temperature so in this case cell lysis, centrifugation and purification were all carried out at 4 °C. The resulting protein was concentrated in a centrifugal filter unit and stored in potassium phosphate buffer (20 mM, pH 6.8) with 30% glycerol.
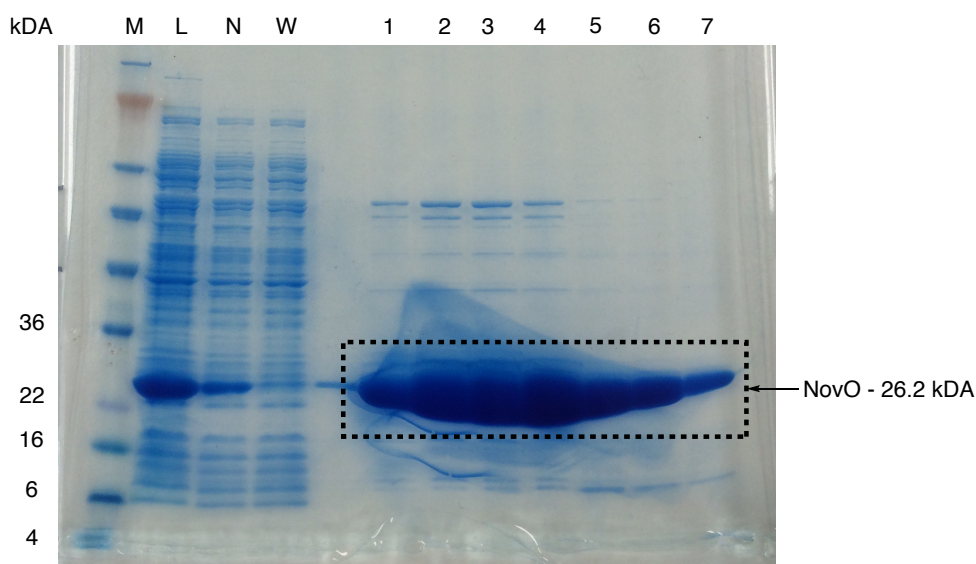
Figure 4.6. Purification SDS PAGE of NovO. Lanes: M – Marker, L – Load, N – Non-Absorbed Fraction, W – Wash, 1-7 – NovO fractions.

A plasmid for MTAN and encoding for ampicillin resistance was purchased from Addgene and was used to transform BL21 (DE3) *E. coli.* After selection for the desired transformants on agar containing ampicillin, colonies were picked and starter cultures were grown in LB media overnight. The starter cultures were transferred to 1 L of Overnight Express media and were grown at 37 °C to an OD of 2. They were then incubated overnight at 18 °C before cells were harvested by centrifugation. MTAN was much more stable than NovO so purification

was carried out using the same conditions as SalL (Section 2.6). SDS PAGE of the purified fractions of MTAN showed successful isolation of the desired protein (Figure 4.7).
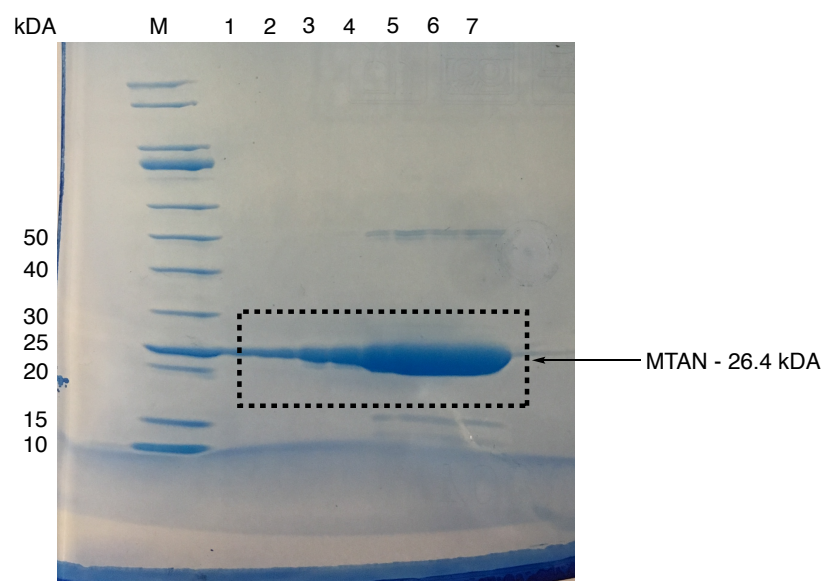

Figure 4.7. Purification SDS PAGE of MTAN. Lanes: M – Marker, 1-7 – Pure MTAN fractions.

4.6 Coumarin Synthesis

The synthesis published for the previously utilised coumarins involved construction of the coumarin core with an amino group at the 3-position (Scheme 4.2).[80] This then allowed for derivatisation of this position by coupling of different carboxylic acids to the coumarin. This approach was effective, but there was scope for the inclusion of a variety of some alternately alkylated arenes. Novel coumarins were synthesised to test if they were substrates of NovO, and subsequently if they could take part in the tandem SAM generation/MTase mediated alkylation process.

Coumarin core synthesis was envisioned as a convergent strategy starting from diethyl (Boc-amino)malonate and 2, 4-dihydroxy benzoic acid. Monohydrolysis of diethyl (Boc-amino)malonate (**4.20**) with KOH provided the target material (**4.21**) in 59% yield. Synthesis of the aromatic moiety began with acetyl protection of 2,4-Dihydroxybenzoic acid (**4.17**) with acetic anhydride providing **4.18** in 93% yield. 2,4-Acetoxybenzoic acid (**4.18**) was then converted to the corresponding acyl chloride (**4.19**), which was condensed with **4.21** in the presence of $MgCl_2$ and NEt_3 to produce β-keto ester **4.22**. Upon treatment of **4.22** with NaOH in MeOH, coumarin **4.23** was isolated in one pot through hydrolysis and intramolecular

esterification in 33% yield over three steps. The free amine (**4.10**) was subsequently generated by HCl mediated removal of the Boc protecting group thus providing the final product in 86% yield.
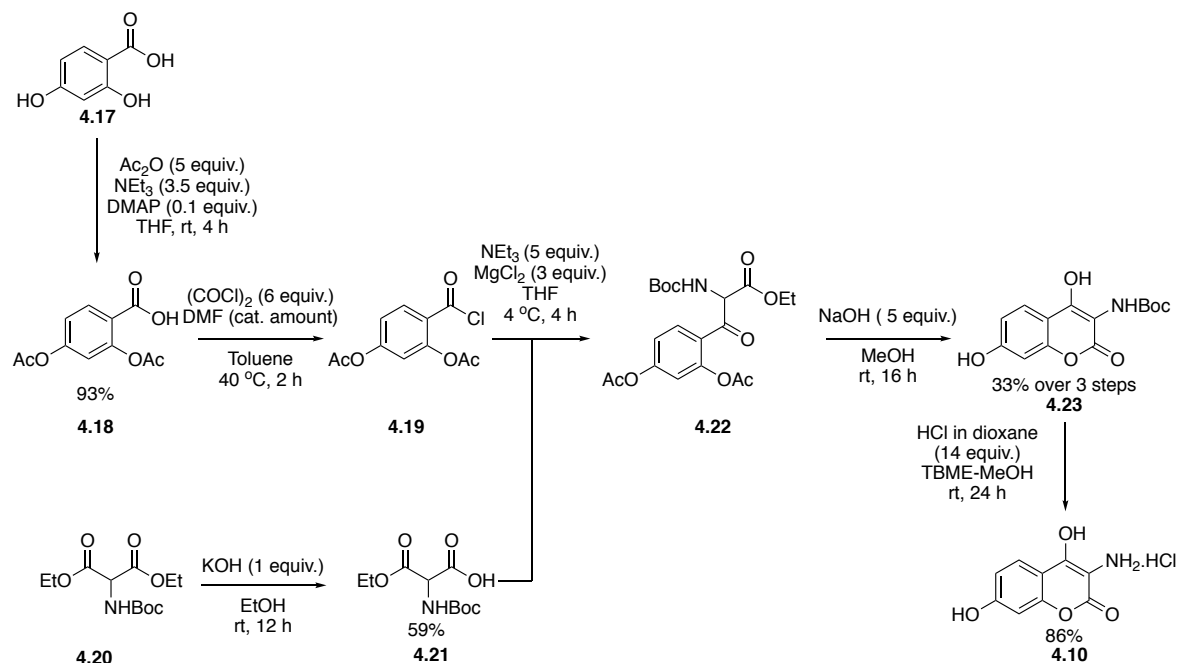

Scheme 4.2. Construction of aminocoumarin core.[80]

Once the aminocoumarin core had been successfully produced in gram-scale quantities, synthesis of a range of modified coumarins were undertaken. A set of commercially available benzoic acid derivatives bearing different substituents was identified (Section 4.2).

The final step to obtain the target coumarins was formation of an amide bond between the aminocoumarin core and the appropriate benzoic acid derivative. Activation of carboxylic acid moiety with $(COCl)_2$ to obtain the benzoyl chloride derivative was followed by amide formation in the presence of NEt_3. Finally, the reaction mixture was treated with NaOH in MeOH for 24 h to hydrolyse any ester product formed through coupling of the benzoyl chloride derivative to either of the free hydroxyl groups in addition to amide formation. After stirring with NaOH in MeOH, removal of the ester was observed for coumarins **1.93** and **4.11-4.14**. In the case of **4.15** hydrolysis had occurred at both the amide and ester linkages with the mass of the starting material observed. Attempts to modulate the hydrolysis through use of different temperatures between room temperature and 70 °C were unsuccessful with either amide and ester hydrolysis, or no hydrolysis observed in each case.

Once each coumarin had been formed and their respective esters had been hydrolysed, the reaction mixture was acidified and formation of a precipitate was observed either immediately or over the course of 24 h. The precipitate was isolated by filtration with subsequent analysis confirming isolation of the target material. This method was generally low yielding (highest yield 28% - Scheme 4.3) as significant amounts of product were left in solution. However, enough material was prepared in each case to characterise the compounds obtained and also for further use in MTase assays. As a result of this, no further optimisation of the isolation procedure was carried out.
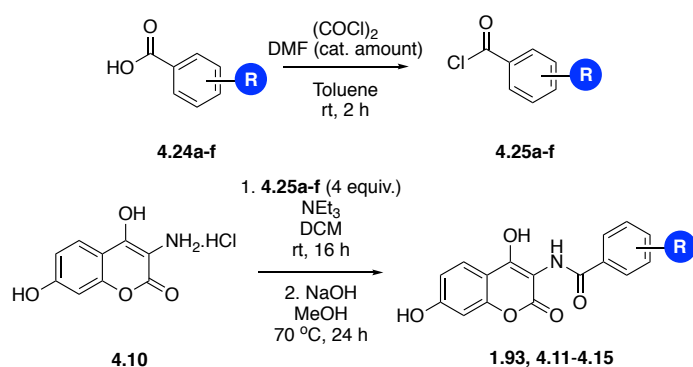
Product Compound No	O (R)	% Yield
1.93		67*
4.11		13
4.12	OMe	8
4.13	N	28
4.14	F	28
4.15	CF$_3$	-

Scheme 4.3. Coupling of benzoic acid derivatives to aminocoumarin core. *Compound **1.93** was previously synthesised as part of a related study, yield reported is from that study.[150]

To further probe the scope of acceptance of different coumarin scaffolds by NovO, a coumarin without the 4-hydroxy group was synthesised. There was precedent for this with the earlier identification by Stecher *et al.* and Sadler *et al.* that dihydroxynaphthalenes could be accepted and methylated by MTases CouO and NovO.[80, 102] NovO tolerates a bulky group at the 3-position of its natural substrate (**1.92a**) with an extended alkyl chain present (Scheme 4.4).
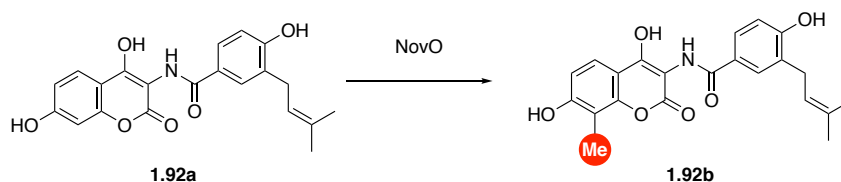
Scheme 4.4. Reaction of NovO in Nature as part of novobiocin biosynthesis.[50]

As a starting point for synthesis of a compound to fit these requirements, coumarin azide (Scheme 4.5) (**4.26**) was chosen due to its commercial availability. It contains both the 2-hydroxyl and 7-carbonyl groups and is also absent of the 4-hydroxyl. The azide group at the 3-position can be readily converted to a triazole, a known isostere of an amide, *via* a CuAAC reaction to mimic the natural substrate.[166] To further mimic the natural substrate, an alkyne bearing an extended alkyl chain, 6-heptyn-1-ol, was utilised. The desired CuAAC reaction between 6-heptyn-1-ol and **4.26** proceeded in 76% yield to provide the desired coumarin (**4.27**)
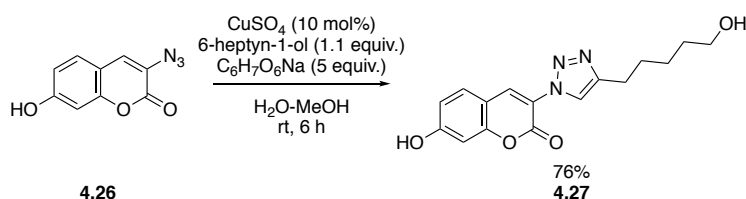
Scheme 4.5. CuAAC reaction of coumarin azide with 6-heptyn-1-ol.

At this point, a set of seven coumarins bearing a variety of modifications had been synthesised. In addition to these substrates, it was desired to attempt NovO mediated (m)ethylation of some medicinally relevant compounds. The group of Blagg have identified the coumarin scaffold as one which can be utilised as the core of inhibitors of Hsp90.[167] Triazole containing carboxylic acid (**4.28**) was converted to acid chloride **4.29** which was then coupled to coumarin fragment (**4.30**) to provide target compound **4.6** (Scheme 4.6).
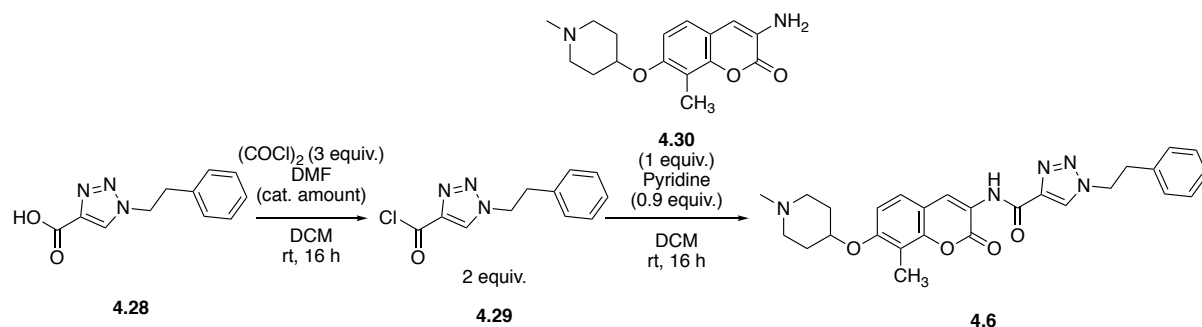
Scheme 4.6. Literature synthesis of Hsp90 inhibitor compound **4.6**.[161]

It was desired to investigate whether or not a precursor to an analogous compound could be alkylated using NovO. The carboxylic acid fragment was synthesised by the literature method (Scheme 4.7)[161] and was coupled to aminocoumarin (**4.10**) by first converting to the corresponding acyl chloride using the same methodology as previously (Scheme 4.3). The low yield of final compound **4.16** can be attributed to the use of less equivalents of carboxylic acid than in previous cases when four equivalents were used rather than two. Further to this, the issues with isolation of target material remained and the product was obtained by precipitation at acidic pH.
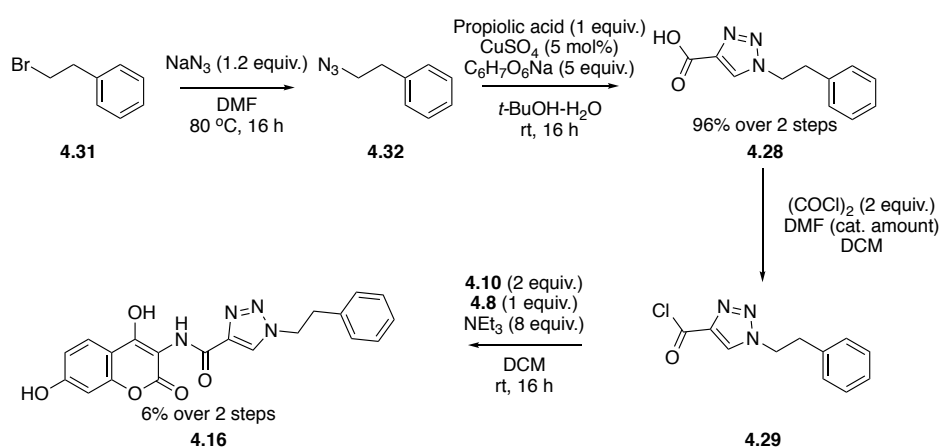
Scheme 4.7. Synthesis of Hsp90 inhibitor precursor **4.16**.[161]

4.7 Investigation of Coumarins for Compatibility with NovO

Previous work carried out by Sadler *et al.* established a tandem SalL mediated SAM generation/NovO mediated alkylation process.[102] This method was used to carry out (m)ethylation of coumarins and dihydroxynaphthalene compounds and also to deuterium and ^{13}C label such compounds through the generation of the required SAM analogues *in situ* by SalL (detailed in Section 1.6.5.1). To begin with, it was desired to repeat the literature

procedure of Sadler *et al.* to provide a benchmark for conversions of coumarin to methylated coumarin when SAM was used as the cofactor against some of the unnatural cofactors. In this process, a number of additives have been utilised. MTAN was added to degrade residual SAH which can act as an inhibitor to further methylation.[102] Further to this, BSA and DTT were added as Gruber *et al*. proposed that these additives enhance NovO stability and would therefore increase conversion.[80]
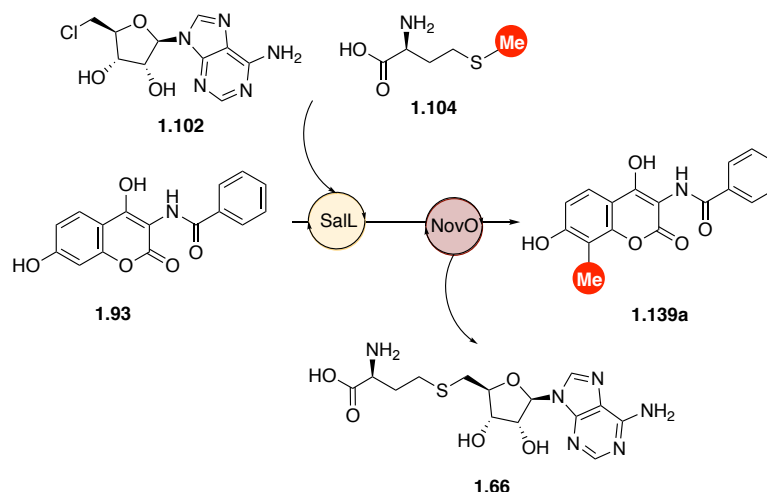
Scheme 4.8. Conditions used: Coumarin (200 μM), ClDA (400 μM), L-met (2.00 mM), SalL (2.10 μM), DTT (1.00 mM) and BSA (1.00 mg/mL), potassium phosphate buffer (100 mM, pH 6.8), 24 h, 37 °C then NovO (18.2 μM) and MTAN (132 nM).

The set of test methylations confirmed the importance of particular additives to the tandem SalL/NovO methylation reaction (Table 4.1). Entry 1 shows good conversion of **1.93** to **1.139a** (71%) in the presence of MTAN and absence of BSA/DTT. In contrast, omission of MTAN from the reaction (Entry 2) reduced the observed conversion to 32%. Inclusion of BSA and MTAN (Entry 3) returned a moderate conversion of 56%. Inclusion of DTT and MTAN within the assay provided a conversion of 94% while addition of BSA to these conditions produced 95% conversion. With these results in hand, BSA, DTT and MTAN were deemed to be essential for further assays to determine the activity of modified SAM analogues towards NovO mediated methylation of coumarins.

Table 4.1. Conversions of reaction detailed in Scheme 4.8 in the presence and absence of combinations of BSA, DTT and MTAN.

Entry	BSA	DTT	MTAN	Conversion
1	✘	✘	✔	71
2	✘	✘	✘	32
3	✔	✘	✔	56
4	✘	✔	✔	94
5	✔	✔	✔	95

A range of SAM analogues (Section 3.6) were assessed for their usefulness in the tandem process. Proof of concept was carried out by use of these compounds to (m)ethylate coumarin **1.93** and conversions obtained for SAM/SAE analogues generated *in* situ were compared with those achieved for SAM/SAE (**1.63**/**1.111a**) (Appendix Figure S 4.1, Scheme 4.9).[77] For these assays, the loading of NovO was reduced by half in an attempt to limit the amount of NovO that would be required to study all of the desired coumarins in triplicate for all cofactors which were accepted by NovO.

The results obtained from the tandem alkylation reaction with reduced NovO loading showed that for coumarin **1.93** conversion to methylated product obtained from *in situ* generated SAM was 49% (Entry 1, Scheme 4.9). Interestingly methylation was enhanced by 2-modified SAM analogues **1.132a** + **b** (77% and 85% conversion) (Entries 3-4, Scheme 4.9) which contain a 2-Cl and 2-NH_2 modification respectively. Compound **1.125** which contains a 2-alkyne modification displayed similar conversions to SAM itself (50%) (Entry 2, Scheme 4.9). Removal of the 6-NH_2 group, as was the case for the 2-Amino-6-chloro and 2-Amino-SAM analogues (**1.132c** + **1.132d**), almost completely abolishes MTase activity with 0% and 4% methylation of **1.93** observed with these cofactors (Entries 5-6, Scheme 4.9). The tandem process also showed significant differences in ethylation of **1.93** when nucleoside-SAM analogues were utilised (Entries 7-9, Scheme 4.9). For SAE (**1.111a**), 24% ethylation was observed but this conversion was enhanced by the use of either 2-Cl-SAE (**3.33**) (25%) or by 2-Amino-SAE (**3.34**) (39%).

Tetrazole containing SAM analogues were observed to carry out methylation of **1.93** however the conversions were vastly reduced with respect to the corresponding carboxylic acid containing compounds (Entries 10-12, Scheme 4.9). The greatest methylation of **1.93** was observed with tetrazole SAM analogues when *tet*-L-met was used in conjunction with **1.131b** to form corresponding SAM analogue **3.36** (22% methylation *vs.* 85% for **1.132b**). This was slightly higher when compared with methylation when cofactors **1.147** (12%) and **3.35** (6%) were used. With these results in hand, the best performing SAM analogues were taken on to investigate alkylation of the full set of coumarins.
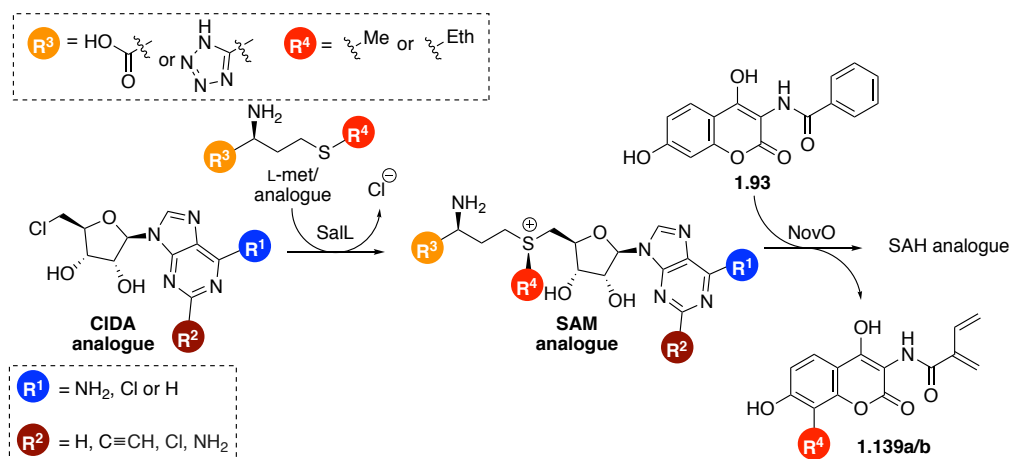
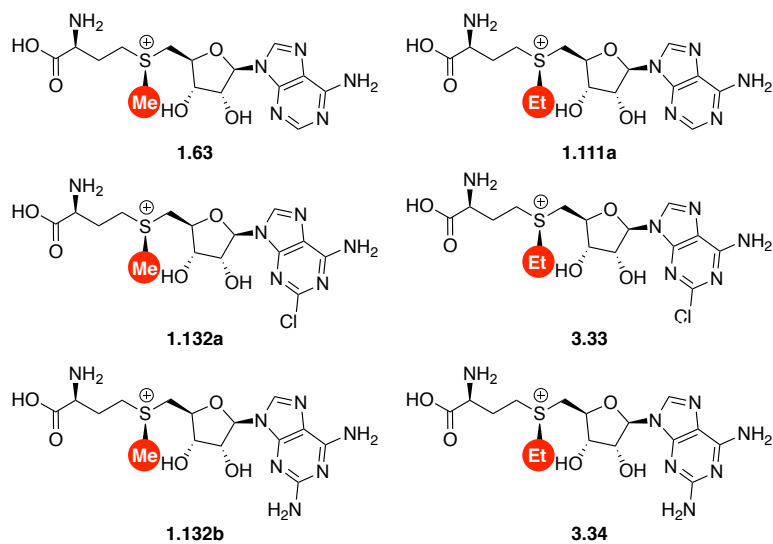| Entry	SAM analogue Compound No.	R¹	R²	R³	R⁴	% Conversion to SAM analogue	% Conversion to alkylated coumarin
1 (control)	**1.63**	NH₂	H	HO—C(=O)	Me	99	49
2	**1.125**	NH₂	-C≡CH	HO—C(=O)	Me	73	50
3	**1.132a**	NH₂	Cl	HO—C(=O)	Me	99	77
4	**1.132b**	NH₂	NH₂	HO—C(=O)	Me	100	85
5	**1.132c**	Cl	NH₂	HO—C(=O)	Me	90	0
6	**1.132d**	H	NH₂	HO—C(=O)	Me	96	4
7	**1.111a**	NH₂	H	HO—C(=O)	Eth	42	24
8	**3.33**	NH₂	Cl	HO—C(=O)	Eth	37	25
9	**3.34**	NH₂	NH₂	HO—C(=O)	Eth	78	39
10	**1.147**	NH₂	H	tetrazole	Me	99	12
11	**3.35**	NH₂	Cl	tetrazole	Me	41	6
12	**3.36**	NH₂	NH₂	tetrazole	Me	99	22

Scheme 4.9. Tandem SalL mediated SAM/analogue generation followed by NovO mediated alkylation of coumarin **1.91**. Conditions used: ClDA/ClDA analogue (400 µM), L-met/L-eth (2.00 mM), SalL (2.10 µM), DTT (1.00 mM) and BSA (1.00 mg/mL), potassium phosphate buffer (100 mM, pH 6.8), 24 h, 37 ºC then NovO (9.38 µM) and MTAN (132 nM).
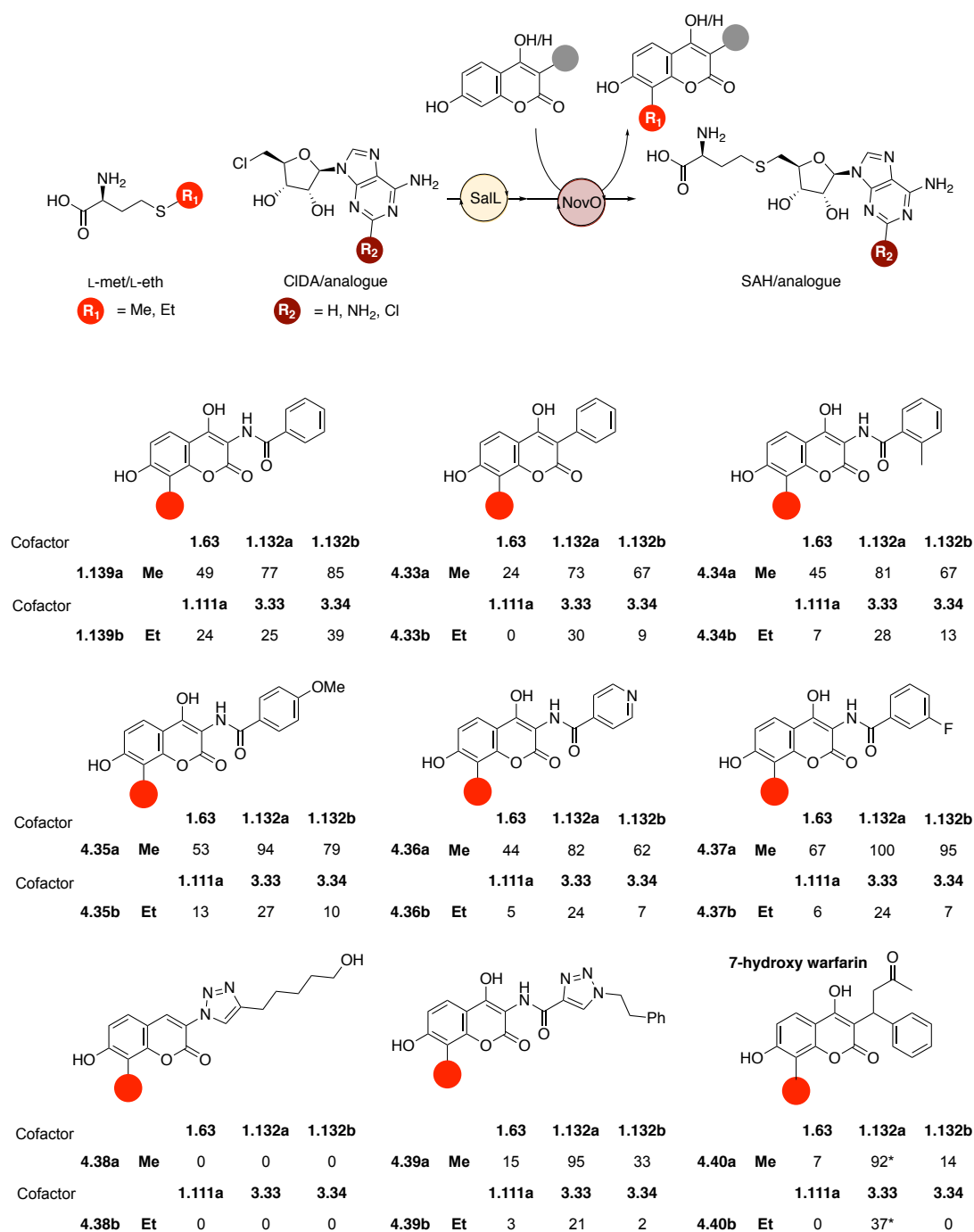
134

4.8   Screening of Methyl Transfer to Novel Coumarins

The set of coumarins obtained was subjected to the same conditions as coumarin **1.91** (Scheme 4.10) but with the panel of SAM analogues now limited to 2-Cl-SAM/SAE (**1.132a**/**3.33**) and 2-Amino-SAM/SAE (**1.132b**/**3.34**) with SAM/SAE (**1.63**/**1.111a**) (Page 136) used to benchmark the observed conversions against (Scheme 4.10). For the coumarins investigated, alkylation of all was observed aside from coumarin triazole derivative **4.27** (to form either **4.37a** or **b**) which showed no reaction for any of the cofactors investigated. For alkylation of **1.91**, 2-modified SAM analogues displayed higher conversion to the methylated and ethylated coumarins. This trend was observed across all coumarins investigated, however for these coumarins **1.132a**/**3.33** performed best in each case. For every reaction where methylation was observed, conversions were >70% when **1.132a** was utilised as the cofactor. The SAM analogue derived from ClDEA (**1.125**) was also utilised as a donor to the variety of coumarins investigated with conversions obtained generally comparable to those observed for SAM (Appendix Figure S 4.2).

In addition to alkylation to generate coumarins **1.139a/b, 4.33a/b-4.37a/b**, **1.132a**/**3.33** were found to be effective cofactors in the alkylation process to produce **4.39a/b** and **4.40a/b**. The differences in the ability of the cofactors to take part in the MTase mediated alkylation was pronounced in the case of **4.39**. Under the normal assay conditions, *in situ* generation of **1.63** and **1.132b** followed by addition of NovO leads to 15% and 33% conversion to compound **4.39a**. In contrast to this, *in situ* generation of **1.132a** and subsequent methylation produced 95% conversion to **4.39a**. This process was carried out on a 20 mg scale and product was isolated by semi-preparative RP-HPLC in 23% yield. For ethylation of this molecule, the same trend was observed as only 3% and 2% conversion were observed when **1.111a 3.34** were utilised as the ethyl donor *vs.* 21% when **3.33** was used.

In addition to this, it was noted that a metabolite of clinically approved drug warfarin, 7-hydoxywarfarin bears a striking similarity to the core of the coumarins which NovO has been shown to methylate. This compound was commercially available so was also investigated as part of this study. Initially, the conversions for this process were low for each cofactor although the previously observed trend of conversion **1.132a** > **1.132b** > SAM **1.63** ≈ 2-alkyneSAM **1.126**; 27%, 14%, 7%, 6% respectively) held true (Appendix, Figure S 4.2, Table S 4.1).

1.63

1.111a

1.132a

3.33

1.132b

3.34

Cofactor		**1.63**	**1.132a**	**1.132b**
1.139a	**Me**	49	77	85
Cofactor		**1.111a**	**3.33**	**3.34**
1.139b	**Et**	24	25	39

Cofactor		**1.63**	**1.132a**	**1.132b**
4.33a	**Me**	24	73	67
Cofactor		**1.111a**	**3.33**	**3.34**
4.33b	**Et**	0	30	9

Cofactor		**1.63**	**1.132a**	**1.132b**
4.34a	**Me**	45	81	67
Cofactor		**1.111a**	**3.33**	**3.34**
4.34b	**Et**	7	28	13

Cofactor		**1.63**	**1.132a**	**1.132b**
4.35a	**Me**	53	94	79
Cofactor		**1.111a**	**3.33**	**3.34**
4.35b	**Et**	13	27	10

Cofactor		**1.63**	**1.132a**	**1.132b**
4.36a	**Me**	44	82	62
Cofactor		**1.111a**	**3.33**	**3.34**
4.36b	**Et**	5	24	7

Cofactor		**1.63**	**1.132a**	**1.132b**
4.37a	**Me**	67	100	95
Cofactor		**1.111a**	**3.33**	**3.34**
4.37b	**Et**	6	24	7

7-hydroxy warfarin

Cofactor		**1.63**	**1.132a**	**1.132b**
4.38a	**Me**	0	0	0
Cofactor		**1.111a**	**3.33**	**3.34**
4.38b	**Et**	0	0	0

Cofactor		**1.63**	**1.132a**	**1.132b**
4.39a	**Me**	15	95	33
Cofactor		**1.111a**	**3.33**	**3.34**
4.39b	**Et**	3	21	2

Cofactor		**1.63**	**1.132a**	**1.132b**
4.40a	**Me**	7	92*	14
Cofactor		**1.111a**	**3.33**	**3.34**
4.40b	**Et**	0	37*	0

Scheme 4.10. Substrate scope of *C*-(m)ethylation to form coumarins **1.139**, **4.33-4.40**. ClDA/ClDA analogue (400 μM), L-Met/L-ethionine (2.00 mM), coumarin (200 μM), SalL (2.10 μM), DTT (1.00 mM) and BSA (1 mg/mL), potassium phosphate buffer (100 mM, pH 6.8), 24 h, 37 °C then NovO (9.38 μM) and MTAN (132 nM). *Optimised conditions for **4.40a/4.40b** with **1.132a**/**3.33**: 2-Cl-ClDA 1.60 mM, L-Met/L-ethionine (8.00 mM), 7-hydroxywarfarin (200 μM), SalL (4.20 μM), DTT (4.00 mM) and BSA (1 mg/mL), potassium phosphate buffer (100 mM, pH 6.8) 24 h, 37 °C then NovO (42.6 μM) and MTAN (528 nM). % conversions were determined by RP-HPLC using a ratio of the peak area at 300 nm of the coumarin starting material to the product.[77, 102]

In an attempt to increase conversions for (m)ethylation of warfarin, a small optimisation was carried out for methylation with **1.132a**. In this, the amount of cofactor equivalents and concentrations of enzymes were varied. The initial reaction displayed 27% conversion under normal assay conditions (Appendix Figure S 4.3, Appendix Table S4.1, Scheme 4.10). A two-fold increase in NovO and MTAN concentration produced a modest increase in conversion to 34% (Entry 2, Table 4.2). In addition, increasing the concentrations of **1.132a**, **1.104** and SalL resulted in a further increase to 51% conversion (Entry 3, Table 4.2). Increasing NovO/MTAN concentrations four-fold while utilising the concentrations of **1.132a**, **1.104** and SalL from Entry 1 resulted in a conversion of 65% (Entry 4, Table 4.2). The increased NovO/MTAN concentrations were then combined with two-fold and four-fold increases in **1.132a** and **1.104** concentration with SalL increased two-fold (Entries 5-6, Table 4.2). These modifications produced further moderate increases in conversion to 68% and 73% respectively. The conditions depicted in Entry 6 were then utilised in triplicate to the production of **4.40a**. The conversion observed across these replicates was 92% (Scheme 4.10). Repetition of the optimised conditions for ethylation of 7-hydroxywarfarin resulted in an average conversion of 37%.

Table 4.2. Reaction optimization for methylation for 7-hydroxywarfarin. 1 equiv. of 7-hydroxywarfarin in each case.

Entry	2-Cl-ClDA (μM)	L-met (mM)	DTT (mM)	SalL (μM)	NovO (μM)	MTAN (nM)	% Conversion
1	400	1	1	2.10	9.38	132	27
2	400	1	2	2.10	18.8	264	34
3	800	2	2	4.20	18.8	264	51
4	400	1	4	2.10	42.6	528	65
5	800	2	4	4.20	42.6	528	68
6	1600	4	4	4.20	42.6	528	73

Once all of the target (m)ethylated (**1.139a/b**, **4.33-4.40a/b**) coumarins had been prepared, reactions were purified on a small scale by semi-preparative RP-HPLC. ^{1}H NMR spectra were acquired to confirm the regioselectivity of the reaction with disappearance of the 8-H doublet and appearance of a singlet integrating to 3H in the case of methylation or a 2H quartet and 3H triplet in the case of ethylation (Figure 4.8). ^{1}H NMR spectra obtained for Hsp90 inhibitor (**4.16**) are compared with those for the methylated (**4.39a**) and ethylated (**4.39b**) product for reference (Figure 4.8).
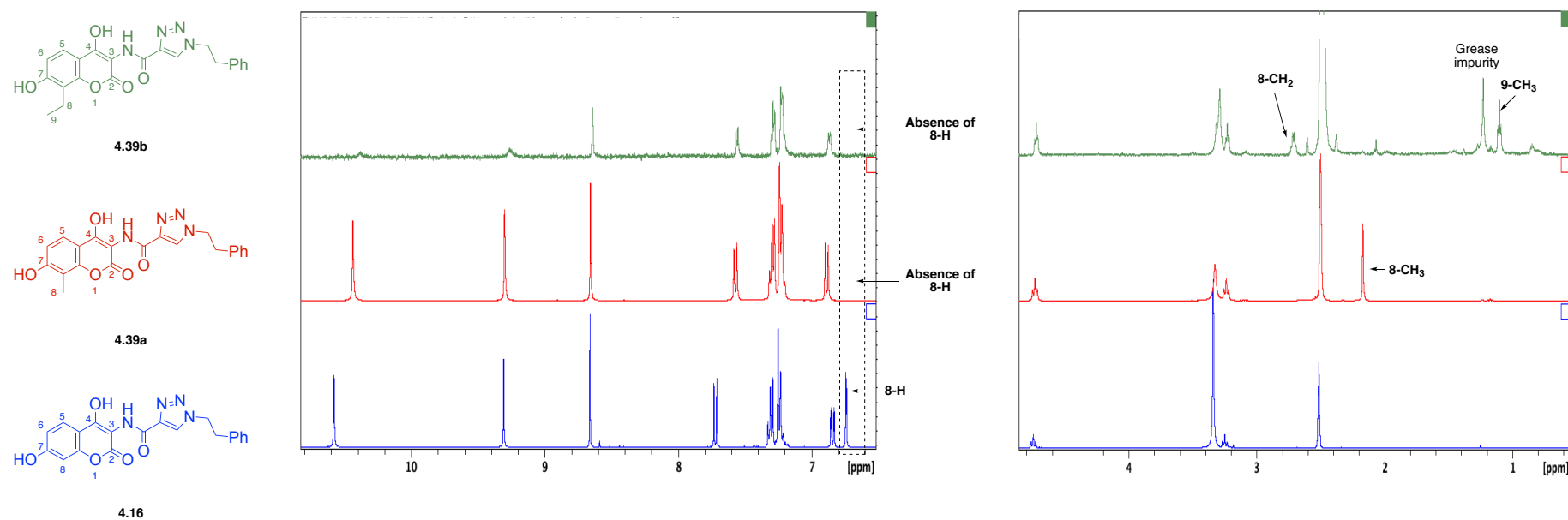
Figure 4.8. NMR spectra for coumarin **4.16** and alkylated derivatives **4.39a** and **4.39b**. Spectra expansion of 11-6.5 ppm showing absence of 8-H signal in (m)ethylated coumarins **4.83a/b.** Expansion of 5-1 ppm shows presence of (m)ethyl groups for **4.39a/b**.

4.9 Rationale for Enhanced Alkylation Using 2-modified Cofactors

With the results obtained displaying enhanced (m)ethylation of coumarins when a 2-modified cofactor is generated *in situ* rather than SAM/SAE it was desired to investigate what might cause this. A crystal structure of NovO obtained from a previous study contained SAH bound in the active-site of the enzyme but did not have any bound coumarin. Inspection of the SAH binding site, which due to their structural similarity is presumed to also be the SAM binding site, revealed some interesting structural features upon which a hypothesis was based. The close proximity of residues Leu46, Val69 and Leu 71 to the Ade part of the cofactor results in a hydrophobic pocket located directly at the 2-position (Figure 4.9). Not only does this explain the ability of the 2-modified cofactors to be tolerated by NovO, it may be the case that the presence of this pocket allows for enhanced binding affinity of the modified SAM cofactor with respect to SAM itself. Indeed closer inspection of the volume of this pocket, as visualised with Chimera, reveals it to be ~21 Å^3 which is very similar to the volume of a chlorine atom.
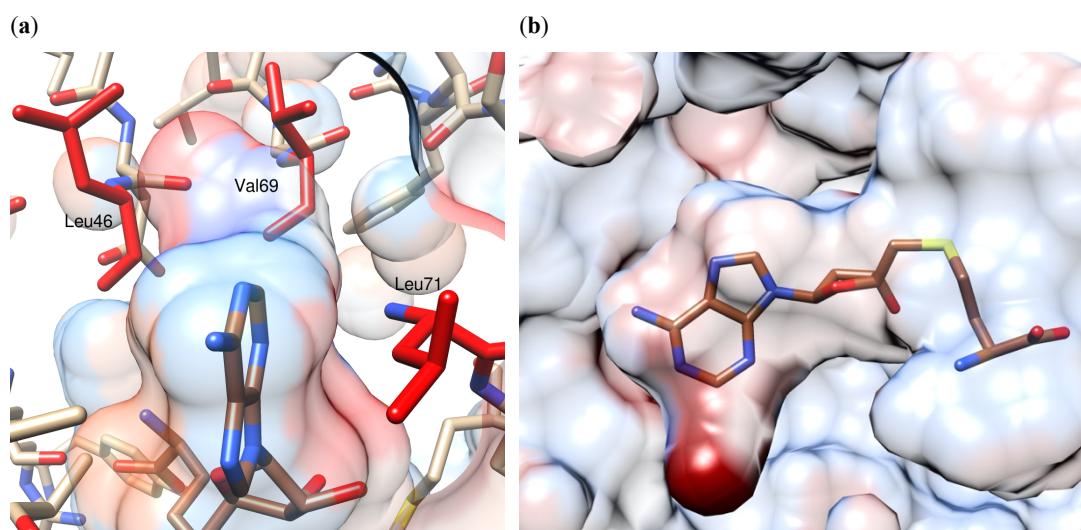
Figure 4.9. Crystal structure of NovO in complex with SAH highlighting the 2-position of adenine projecting towards a hydrophobic cleft (red. PDB: 5MGZ).

4.10 Summary

A range of novel coumarins have been synthesised and investigated as substrates for MTase NovO. As well as expanding the known substrate scope of NovO this methodology has also been applied to two compounds with potential medicinal relevance. A precursor to an Hsp90 inhibitor analogue and 7-hydroxywarfarin have been successfully (m)ethylated with this showing proof of concept of late-stage MTase mediated alkylation.

The tandem process was utilised with a range of SAM analogues generated *in situ* instead of SAM itself. A number of these compounds displayed compatibility with NovO and in the case of 2-modified SAM analogues showed either comparable conversion (2-Alkyne **1.126**) to SAM or enhanced conversion (2-Cl (**1.132a**)/2-NH$_2$ (**1.132b**)). In particular, the 2-Cl-SAM/SAE cofactor showed high conversion for each coumarin studied.

A future goal of this study could be to investigate other substituents at the 2-position of Ade. Inspection of the SalL crystal structure shows a solvent exposed channel at this position implying that the restriction of substituents at this position may be relaxed and would allow other SAM analogues to be generated. These SAM compounds could then be introduced to NovO with the conversions resulting from these cofactors then investigated. The pocket present in the NovO crystal structure is significantly smaller than the channel observed in SalL therefore it is likely that alkylation utilising the tandem process using 2-modified cofactors will be limited by NovO rather than SalL. Modifications of interest would be replacement of the 2-H from SAM with moieties such as F, Br and I. Comparison of the alkylation efficiencies of these compounds with the 2-Cl modified SAM analogue would provide interesting information into how changing the size of the halide and also the electronic change provided alters reactivity of the cofactor. It would also be interesting to probe this position with alkyl chains. A 2-alkyne modification is readily accepted, therefore investigation of methyl, ethyl and propyl substitutions in this position would give an interesting insight into the steric limitations of modifications at this position.

Chapter 5

Conclusions and Future Directions

5 Conclusions and Future Work

The purpose of this thesis was to investigate the use of SalL and NovO in a tandem enzymatic platform for the alkylation of aromatic small molecules. It was hoped that advances made in areas such as cofactor production and scope of SAM analogue formation could then be applied to other SAM dependent MTases.

Firstly, SalL was probed by carrying out mutagenesis of the active site of the enzyme. This aimed to serve the dual purpose of investigating residues which were essential to catalysis, while also attempting to increase substrate promiscuity towards analogues of L-met bearing groups larger than the methyl found in natural SAM. A number of key residues were identified for substrate binding as mutagenesis at these sites abolished enzymatic activity. A non-redundant crystal structure of the enzyme with SAM bound was obtained and this identified a novel interaction between Arg243 and Glu17 which occurs between the monomer units of the enzyme and strengthens the integrity of the active trimer. Although some active mutants were identified, none were more promiscuous than the wild-type towards methionine analogues or did indeed show any enhanced properties with respect to the wild-type.

A range of cofactor analogues were synthesised and tested for their ability to form SAM analogues. These included methionine analogues in which the S-Me was replaced by groups such as CF_3, CH_2CF_3, benzyl and butyl and ClDA analogues in which modifications were made to the sugar and to the 2-, 6- and 7-positions of the Ade nucleobase. Methionine modifications, aside from replacement of S-Me with Eth, showed no activity with SalL to form SAM analogues. In contrast, modifications to the nucleobase of ClDA were readily tolerated, with 2-modified ClDA analogues showing comparable activity towards SAM formation as ClDA. Replacement of N7 with C reduced conversion to the corresponding SAM analogue significantly (100% to 10%). Modifications to the sugar abolished activity entirely.

The SAM analogues generated were then tested for activity towards NovO mediated (m)ethylation of small aromatic molecules. Use of the tandem *in situ* SAM generation/alkylation process displayed a significant enhancement in alkylation when 2-modified SAM analogues bearing either an amino or Cl group were utilised. This process, and in particular the use of the 2-Cl-SAM analogue allowed for an expansion in the known substrate scope of NovO, with medicinally relevant compounds such as a precursor to an analogue of an Hsp90 inhibitor and a metabolite of warfarin shown to undergo (m)ethylation.

To expand on the body of work presented herein, a number of future studies are proposed:

(i) *Expansion of substrate scope of SalL* – Although this was attempted as part of this study, ultimately efforts to this end were unsuccessful. Future efforts towards this goal would revolve around identification of a high-throughput screening platform for SalL mediated SAM generation. This could be done through use of a SAM auxotroph into which random mutants of SalL could be transformed. These constructs could then be plated onto agar containing ClDA and L-met with colony growth/survival indicating presence of a functional mutant of SalL that can synthesise SAM. This could then be coupled to a second screen involving colorimetric detection of chloride ion which is produced during SAM formation. This would then allow for screening of Met analogues and would provide the basis for Directed Evolution towards more promiscuous variants of SalL.

(ii) *Use of SAM analogues* – The SAM analogues generated have been formed previously but have not been investigated as cofactors for MTases. Varied activities are shown for the SAM analogues generated in this study in tandem with MTase NovO. Investigation of these analogues and their ability to act as MTase substrates across other SAM MTase families such as DNA MTases would be of significant interest. This process could be carried out by methylating a position of a DNA strand which is susceptible to a particular endonuclease. After methylation, the DNA strand could be subjected to the given nuclease with subsequent analysis by an agarose gel identifying whether or not the methylation, and thus protection of the cleavage site, had been successful.

(iii) *Structural/Biophysical studies of NovO* – The results obtained from the tandem enzymatic process show enhanced conversion of coumarins to their respective alkylated product. A future goal would be to obtain evidence to support a hypothesis for why this might be the case. Such studies could involve obtaining a crystal structure for NovO with 2-Cl-SAM or perhaps 2-Cl-SAH bound. Further to this, work could be carried out to obtain kinetics of binding for 2-Cl-SAM with NovO. This could be compared to kinetics of SAM binding and may offer an inclination of why the observed conversions occur.

Chapter 6

Experimental

Iain McKean

6 Experimental

6.1 Experimental Procedures

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. Coumarin 1.93 was prepared via literature procedure[77] and precursor to (m)ethylated coumarin 4.33 was purchased from Fluorochem (Product No. 495141). NMR spectroscopy was carried out using either a Bruker 400 UltraShield™ B-ACS 60 spectrometer, Bruker AV500 instrument or Bruker AV600 instrument. All chemical shifts are reported in ppm and coupling constants are quoted in hertz (Hz). CDCl$_3$ referenced at 7.26 ppm (^{1}H) and 77.16 ppm (^{13}C), DMSO-d$_6$ referenced at 2.50 ppm (^{1}H) and 39.52 ppm (^{13}C) and CD$_3$OD referenced at 3.31 ppm (1H) and 49.00 ppm (^{13}C) Abbreviations for splitting patterns are s (singlet), d (doublet), t (triplet), q (quartet), quin (quintet), sxt (sextet), sept (septet) and m (multiplet). All NMR data was processed using Topspin3.5pl7 software. Proton and carbon chemical shifts were assigned using proton (^{1}H), carbon (^{13}C), Heteronuclear Single Quantum Coherence (HSQC), Heteronuclear Multiple-Bond Correlation Spectroscopy (HMBC) and Correlation Spectroscopy (COSY). LC-MS was carried out on an Aglient 1100 HPLC instrument in conjunction with a Waters Micromass ZQ 2000/4000 mass detector. Electrospray ionization (ESI) was used in all cases. High resolution mass spectra were recorded on a LTQ Orbitrap XL1 at EPSRC UK National Mass Spectrometry facility (Swansea) and on a Bruker® micrOTOF II using a direct infusion method and calibrated to sodium formate at the SIRCAMS facility, University of Edinburgh. IR data was collected on an Agilent spectrometer and the data processed using Spectrum One software. Only major absorbances are reported. Analytical reversed-phase HPLC (RP-HPLC) was carried out on a Shimadzu Prominence instrument utilising a PDA Detector scanning from 190-600 nm. Semi-preparative RP-HPLC purification was carried out on a Dionex Ultimate 3000 series instrument using a 150 x 21.2 mm Kinetex 5 µm C18 column. Protein concentrations were measured on either a Trinean Xpose micro-volume spectrometer or a Nanodrop1000 instrument. E0.1% values were determined from Expasy ProtParam (https://web.expasy.org/cgi-bin/protparam/protparam).

6.2 Synthetic procedures for the Preparation of Small Molecules
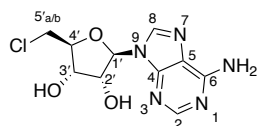
($2R,3R,4S,5S$)-2-(6-Amino-9H-purin-9-yl)-5-(chloromethyl)tetrahydrofuran-3,4-diol[142]
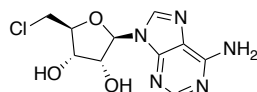
1.102

To an ice-cooled suspension of ($2R,3R,4S,5R$)-2-(6-Amino-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol (1.00 g, 3.70 mmol) in acetonitrile (50.0 mL) was added pyridine (610 μL, 7.50 mmol). Thionyl chloride (1.40 mL, 19.2 mmol) was then added dropwise and the reaction mixture was stirred at 0 °C for 3 h. The reaction was allowed to warm to rt and was stirred for a further 16 h. The resulting precipitate was isolated by filtration and dissolved in water: MeOH (20.0 mL, 5:1). Ammonium hydroxide (2.00 mL, 13.4 M, 26.8 mmol) was added and the reaction stirred at rt for 30 minutes. The solvent was removed under reduced pressure to afford the title compound as a colourless solid (950 mg, 3.33 mmol, 90%).

^{1}H NMR δ (400 MHz, DMSO-d_6) 8.36 (s, 1H, 8-H), 8.16 (s, 1H, 2-H), 7.30 (br. s, 2H, NH_2), 5.95 (d, J = 5.4 Hz, 1H, 1′-H), 5.63 (br. s, 1H, 2′-OH), 5.50 (br. s, 1H, 3′-OH), 4.76 (t, J = 5.4 Hz, 1H, 2′-H), 4.24 (t, J = 5.4 Hz, 1H, 3′-H), 4.09 (dt, J = 11.6 , 5.4 Hz, 1H, 4′-H), 3.95 (dd, J = 11.6 , 5.4 Hz, 1H, 5′-H_aH$_b$), 3.84 (dd, J = 11.6 , 5.4 Hz, 1H, 5′-H$_a$$H_b$).

^{13}C NMR: δ (101 MHz, DMSO-d_6) 156.1 (C), 152.7 (CH), 149.4 (C), 139.7 (CH), 119.1 (C), 87.4 (CH), 83.7 (CH), 72.7 (CH), 71.3 (CH), 44.9 (CH$_2$).

ν$_{max}$ (neat): 3550-3050 cm^{-1} (OH stretch), 3040 cm^{-1} (NH stretch), 1638 cm^{-1} (NH bend).

LCMS: (Method 1 – Section 6.11) t_R = 1.41 min, m/z = 286 ([M^{35}Cl+H]$^+$ 100%), 288 ([M^{37}Cl+H]$^+$, 30%).

RP-HPLC (Method 2 – Section 6.10) t_R = 4.87 min.

R_f (EtOAc) = 0.53

DL-2-Amino-4-(isopropylthio)butanoic acid[83]
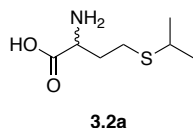
3.2a

To a solution of 2-amino-4-mercaptobutanoic acid (200 mg, 1.48 mmol) in EtOH (1.50 mL) and aqueous sodium hydroxide (3.33 mL, 2.00 M, 6.66 mmol) was added 2-bromopropane (182 mg, 1.48 mmol). The reaction mixture was stirred for 24 h before acidifying to pH 2 with conc. HCl. Solvent was removed under reduced pressure, the resulting residue was re-suspended in warm EtOH and the solids removed by filtration. The solvent was removed under reduced pressure and the residue redissolved in water (1.00 mL). The pH was adjusted to pH 4 with aqueous NaOH (2.00 M) with the resulting solid isolated by filtration and dried under vacuum to provide the title compound as a colourless solid (28.0 mg, 158 µmol, 11%).

^{1}H NMR: δ (400 MHz, D$_2$O) ppm 3.32-3.26 (m, 1H, α-CH), 2.97 (sept, J = 6.8 Hz, 1H, S-CH), 2.54 (t, J = 6.8 Hz, 2H, γ-CH_2), 1.89-1.65 (m, 2H, β-CH_2), 1.18 (d, J = 6.8 Hz, 6H, 2 × CH_3).

^{13}C NMR: δ (101 MHz, D$_2$O) ppm 171.5 (C), 51.9 (CH) 34.3 (CH) 29.8 (CH$_2$) 24.9 (CH$_2$) 22.3 (2 × CH$_3$).

M.pt.: Decomp. >250 °C.

ν_{max} (neat): 3300-2800 cm^{-1} (OH stretch).

LCMS: (**Method 1 – Section 6.11**) t$_R$ = 0.59 min, m/z = 178 [M+H]$^+$

R$_f$ (nBuOH:AcOH:H$_2$O 3:1:1) = 0.50

DL-2-Amino-4-(butylthio)butanoic acid[83]

3.2b
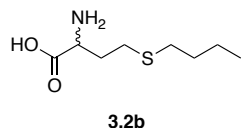
To a solution of 2-amino-4-mercaptobutanoic acid (200 mg, 1.48 mmol) in EtOH (1.50 mL) and aqueous NaOH (3.33 mL, 2 M, 6.66 mmol) was added 1-bromobutane (160 µL, 1.48 mmol). The reaction mixture was stirred for 24 h before acidifying to pH 2 with conc. HCl.

Solvent was removed under reduced pressure, the resulting residue was re-suspended in warm EtOH and the solids removed by filtration. The solvent was removed under reduced pressure and the residue redissolved in water (1.00 mL). The pH was adjusted to pH 4 with aqueous NaOH (2.00 M) with the resulting solid isolated by filtration and dried under vacuum to provide the title compound as a colourless solid (193 mg, 1.01 mmol, 68%).

^{1}H NMR: δ (400 MHz, D$_2$O) ppm 4.08 (t, *J* = 7.3 Hz, 1H, α-C*H*), 2.58 (t, *J* = 7.3 Hz, 2H, γ-C*H$_2$*), 2.45 (t, *J* = 7.3 Hz, 2H, S-C*H$_2$*), 2.18-2.00 (m, 2H, β-C*H$_2$*), 1.44-1.39 (m, 2H, C*H$_2$*CH$_3$), 1.27-1.19 (m, 2H, CH$_2$C*H$_2$*CH$_2$), 0.85 (t, *J* = 7.3 Hz, 3H, C*H$_3$*).

^{13}C NMR: δ (101 MHz, D$_2$O) ppm 182.8 (C), 55.4 (CH), 34.8 (CH$_2$), 30.9 (CH$_2$), 30.8 (CH$_2$), 27.6 (CH$_2$), 21.4 (CH$_2$), 13.1 (CH$_3$).

M.pt.: decomp 231 °C (lit = 240 °C decomposition)[83]

ν$_{max}$ (neat): 3300-2800 cm^{-1} (OH stretch).

LCMS: (Method 1 – Section 6.11) t_R = 1.44 min, m/z = 192 [M+H]$^+$

R$_f$ (nBuOH:AcOH:H$_2$O 3:1:1) = 0.53

DL-4-(Allylthio)-2-aminobutanoic acid[83]
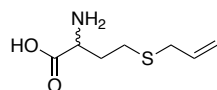
3.2c

To a solution of 2-amino-4-mercaptobutanoic acid (200 mg, 1.48 mmol) in EtOH (1.50 mL) and aqueous sodium hydroxide (3.33 mL, 2.00 M, 6.66 mmol) was added 3-bromoprop-1-ene (179 mg, 1.48 mmol). The reaction mixture was stirred for 24 h before acidifying to pH 2 with conc. HCl. Solvent was removed under reduced pressure, the resulting residue was re-suspended in warm EtOH and the solids removed by filtration. The solvent was removed under reduced pressure and the residue redissolved in water (1.00 mL). The pH was adjusted to pH 4 with aqueous NaOH (2.00 M) with the resulting solid isolated by filtration and dried under vacuum. Further purification was carried out by Mass Directed Auto Purification (see Section 6.12) to afford the title compound as a colourless solid (156 mg, 539 µmol, 36%).

¹H NMR: δ (400 MHz, D₂O) 5.85-5.72 (m, 1H, C*H*=CH₂) 5.18-5.10 (m, 2H, CH=C*H₂*) 4.07 (t, *J* = 6.4 Hz, 1H, α-C*H*) 3.18 (d, *J* = 7.1 Hz, 2H, C*H₂*-CH=CH₂) 2.61 (t, *J* = 7.4 Hz, 2H, γ-C*H₂*) 2.26-2.05 (m, 2H, β-C*H₂*).

¹³C NMR: δ (101 MHz, D₂O) 172.2 (C), 133.8 (CH), 117.8 (CH₂), 52.3 (CH), 33.4 (CH₂), 29.5 (CH₂), 25.1 (CH₂).

M.pt.: Decomp >230 °C.

ν_max (neat): 3300-2800 cm⁻¹ (OH stretch), 1744 cm⁻¹ (C=O stretch).

HRMS: C₇H₁₃NO₂S [M+H]⁺ requires *m/z* 176.0740, found 176.0738.

R_f (nBuOH:AcOH:H₂O 3:1:1) = 0.40

DL-2-Amino-4-(benzylthio)butanoic acid[100]
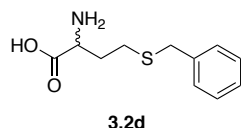
3.2d

To a solution of 2-amino-4-mercaptobutanoic acid (200 mg, 1.48 mmol) in EtOH (1.50 mL) and aqueous sodium hydroxide (3.33 mL, 2.00 M, 6.66 mmol) was added (bromomethyl)benzene (176 µL, 1.48 mmol). The reaction mixture was stirred for 24 h before acidifying to pH 2 with conc. HCl. Solvent was removed under reduced pressure, the resulting residue was re-suspended in warm EtOH and the solids removed by filtration. The solvent was removed under reduced pressure and the residue redissolved in water (1.00 mL). The pH was adjusted to pH 4 with aqueous NaOH (2.00 M) with the resulting solid isolated by filtration and dried under vacuum to provide the title compound as an off-white solid (93.0 mg, 413 µmol, 28%).

¹H NMR: δ (400 MHz, D₂O) ppm 7.26 (m, 5H, 5 × ArC*H*) 3.69 (s, 2H, S-C*H₂*) 3.21-3.13 (m, 1H, α-C*H*) 2.44-2.36 (m, 2H, γ-C*H₂*) 1.84-1.70 (m, 2H, β-C*H₂*)

¹³C NMR: δ (101 MHz, D₂O) ppm 182.7 (C), 138.6 (CH), 128.9 (2 × CH), 128.8 (2 × CH), 127.2 (C), 55.3 (CH), 35.0 (CH₂), 34.4 (CH₂), 27.2 (CH₂).

M.pt.: Decomp. >220 ºC

ν$_{max}$ (neat): 3300-2800 cm^{-1} (OH stretch), 1744 cm^{-1} (C=O stretch).

HRMS: C$_{11}$H$_{14}$NO$_2$S [M-H]$^-$ requires *m/z* 224.0751, found 224.0753.

R$_f$ (nBuOH:AcOH:H$_2$O 3:1:1) = 0.20

S-(Prop-2-yn-1-yl)homocysteine

3.2e
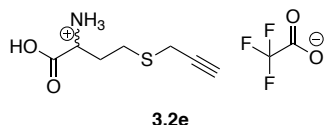
A solution of 2-amino-4-mercaptobutanoic acid (100 mg, 740 μmol) in NH$_4$OH (10.0 mL, 2.00 M) was stirred at 0 °C for 30 mins. 80% Propargyl bromide in toluene solution (96.0 μmol, 863 μmol) was added dropwise and the reaction mixture was stirred for a further 3 h. Solvent was removed under reduced pressure and the resulting residue was purified by semi-preparative RP-HPLC. Product containing fractions were dried by lyophilisation and the title compound was obtained as a pale orange solid (109 mg, 380 μmol, 51%).

^{1}H NMR: δ (500 MHz, D$_2$O) 4.10 (t, *J* = 6.9 Hz, 1H, α-C*H*), 3.34 (d, *J* = 2.5 Hz, 2H, S-C*H$_2$*C≡CH), 2.83 (t, *J* = 6.9 Hz, 2H, γ-C*H$_2$*), 2.63 (t, *J* = 2.5 Hz, 1H, S-CH$_2$C≡C*H*), 2.32-2.12 (m, 2H, β-C*H$_2$*).

^{13}C NMR: δ (125 MHz, D$_2$O) 172.2 (C), 80.3 (CH), 72.3 (C), 52.2 (CH), 29.2 (CH$_2$), 26.3 (CH$_2$), 18.0 (CH$_2$).

HRMS: C$_7$H$_{11}$NO$_2$S [M-H]$^-$ requires *m/z* 172.0438, found 172.0439.

DL-2-Amino-4-((trifluoromethyl)thio)butanoic acid[168]

3.2f
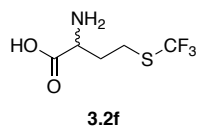
To a solution of 4,4'-disulfanediylbis(2-aminobutanoic acid) (145 mg, 540 μmol) in water (12.0 mL) was added bis{[(trifluoromethyl)sulfinyl]oxy}zinc (358 mg, 1.08 mmol). TBHP (340 μL, 2.43 mmol) was added dropwise at 20 ºC and the reaction was stirred for 2 h. A further portion of bis{[(trifluoromethyl)sulfinyl]oxy}zinc (179 mg, 540 μmol.) was then added

followed by additional TBHP (340 µL, 2.43 mmol) and the reaction mixture was stirred for a further 16 h. The reaction mixture was washed with heptane (10.0 mL) before being basified with NaOH (sat. aq) to ~ pH 9. The resulting precipitate was filtered and dried under vacuum to provide the title compound as a colourless solid (97.0 mg, 477 µmol, 88%).

^{1}H NMR: δ (400 MHz, D$_2$O) ppm 3.74 (t, J = 6.4 Hz, 1H, α-C*H*), 3.10-2.92 (m, 2H, γ-C*H$_2$*), 2.24-2.09 (m, 2H, β-C*H$_2$*).

^{13}C NMR: δ (101 MHz, D$_2$O) ppm 171.1 (COOH), 130.8 (app. d, J = 307 Hz, CF$_3$), 53.5 (CH), 30.4 (CH$_2$), 25.2 (CH$_2$).

^{19}F NMR: δ (376 MHz, D$_2$O) ppm -41.2 (CF$_3$).

ν$_{max}$ (neat): 3300-2800 cm^{-1} (OH stretch).

M.pt.: 225 °C (lit 227-228 °C)[168]

LCMS: (Method 1 – Section 6.11) t$_R$ = 1.03 min, m/z = 204 [M+H]$^+$

R$_f$ (nBuOH:AcOH:H$_2$O 3:1:1) = 0.61

DL-2-Amino-4-((2,2,2-trifluoroethyl)thio)butanoic acid[150]
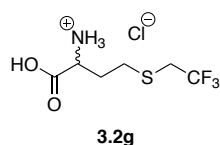
3.2g

To a solution of 4,4'-disulfanediylbis(2-aminobutanoic acid) (100 mg, 373 µmol) in water (12.0 mL) was added bis{[(2,2,2-trifluoroethyl)sulfinyl]oxy}zinc (268 mg, 745 µmol). TBHP (235 µL, 1.68 mmol) was added dropwise at 20 ºC and the reaction was stirred for 2 h. A further portion of bis{[(2,2,2-trifluoroethyl)sulfinyl]oxy}zinc (179 mg, 540 µmol) was then added followed by additional TBHP (340 µL, 2.43 mmol) and the reaction mixture was stirred for a further 16 h. The reaction mixture was washed with TBME (10.0 mL) and the aqueous fractions were collected and concentrated under reduced pressure. The residue was dissolved in aqueous HCl (1.00 mL, 1.00 M, 1.00 mmol) and purified *via* ion exchange chromatography.

Solvent was removed by lyophilisation from product containing fractions to provide the target compound as a colourless solid (80.0 mg, 315 µmol, 84%).

¹H NMR: δ (400 MHz, D$_2$O) ppm 3.80 (t, *J* = 6.8 Hz 1H, α-C*H*), 3.31 (q, *J* = 10.3 Hz, 2H, C*H$_2$*CF$_3$) 2.78 (t, *J* = 6.8 Hz, 2H, γ-C*H$_2$*) 2.02-2.22 (m, 2H, β-C*H$_2$*).

¹³C NMR: δ (101 MHz, D$_2$O) ppm 171.4 (C), 126.0 (q, *J* = 275 Hz, CF$_3$), 51.5 (CH) 32.9 (q, *J* = 32.3 Hz, CH$_2$) 29.4 (CH$_2$) 27.7 (CH$_2$).

¹⁹F NMR: δ (376 MHz, D$_2$O) ppm -66.4 (CF$_3$).

M.pt.: decomp >220 °C

ν$_{max}$ (neat): 3300-2800 cm^{-1} (OH stretch).

LC MS: (Method 1 – Section 6.11) t$_R$ = 1.03 min, m/z = 218 [M+H]$^+$

R$_f$ (nBuOH:AcOH:H$_2$O 3:1:1) = 0.50

S-(Difluoromethyl)homocysteine[169]
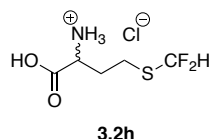
3.2h

To a solution of 4,4'-disulfanediylbis(2-aminobutanoic acid) (100 mg, 373 µmol) in water (12.0 mL) was added bis{[(difluoromethyl)sulfinyl]oxy}zinc (220 mg, 745 µmol). TBHP (235 µL, 1.68 mmol) was added dropwise at 20 °C and the reaction was stirred for 2 h. A further portion of bis{[(difluoromethyl)sulfinyl]oxy}zinc (55.0 mg, 187 µmol) was then added followed by additional TBHP (115 µL, 838 µmol) and the reaction mixture was stirred for a further 16 h. The reaction mixture was washed with TBME (10.0 mL) and the aqueous fractions were collected and concentrated under reduced pressure. The residue was dissolved in aqueous HCl (1.00 mL, 1.00 M, 1.00 mmol) and purified *via* ion exchange chromatography. Solvent was removed by lyophilisation from product containing fractions to provide the target compound as a colourless solid (68.1 mg, 307 µmol, 83%).

¹H NMR: δ (400 MHz, D_2O) ppm 6.96 (t, J = 55.4 Hz, 1H, CF₂*H*), 4.11 (t, J = 6.6 Hz 1H, α-C*H*), 2.95-2.85 (m, 2H, γ-C*H₂*), 2.32-2.10 (m, 2H, β-C*H₂*).

¹³C NMR: δ (101 MHz, D_2O) ppm 171.2 (C), 126.0 (t, J = 272 Hz, CF₂H), 51.5 (CH), 30.9 (CH₂) 22.6 (CH₂).

¹⁹F NMR: δ (376 MHz, D_2O) ppm -93.2 (d, J = 56.9 Hz, CF₂H).

LCMS: (Method 1 – Section 6.11) t_R = 0.43 min, m/z = 186 = [M+H]⁺

9*H*-Fluoren-9-yl (*S*)-(1-amino-4-(methylthio)-1-oxobutan-2-yl)carbamate[103]

3.6
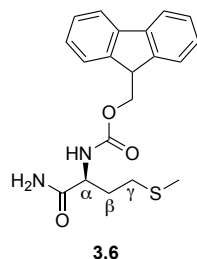
Fmoc-L-methionine (1.00 g, 2.69 mmol) was dissolved 1:1 DMF:dioxane (30.0 mL). To this solution was added $(Boc)_2O$ (646 mg, 2.96 mmol), NH_4HCO_3 (234 mg, 2.96 mmol), and pyridine (260 μL, 2.96 mmol). The reaction was stirred at rt for 48 h under an argon atmosphere. The reaction mixture was monitored by TLC ($CHCl_3$:MeOH, 9:1), and when there was a complete disappearance of starting material, the reaction mixture was poured into EtOAc (150 mL). The organic layer was washed sequentially with H_2O (2 × 150 mL), and H_2SO_4 (150 mL, 940 mM, 140 mmol), dried with Na_2SO_4, filtered and concentrated under reduced pressure to afford the title compound as an off-white solid (952 mg, 2.57 mmol, 91%).

¹H NMR: δ (500 MHz, DMSO-d_6) 7.89 (d, J = 6.8 Hz, 2H, 2 × Fmoc-C*H*), 7.73 (t, J = 6.8 Hz, 2H, 2 × Fmoc-C*H*), 7.48-7.28 (m, 6H, 4 × Fmoc-C*H* & N*H₂*), 7.02 (br. s, 1H, N*H*), 4.32-4.20 (m, 3H, Fmoc-C*H₂* & C*H*), 4.06-3.98 (m, 1H, α-C*H*), 2.48-2.40 (m, 2H, γ-C*H₂*), 2.05 (s, 3H, S-C*H₃*), 1.95-1.73 (m, 2H, β-C*H₂*).

¹³C NMR: δ (126 MHz, DMSO-d_6) 173.5 (C), 156.0 (C), 143.9 (C), 143.8 (C), 140.7 (2 × C), 127.6 (2 × CH), 127.0 (2 × CH), 125.3 (2 × CH), 120.1 (2 × CH), 65.6 (CH₂), 53.7 (CH), 46.7 (CH), 31.5 (CH₂), 29.8 (CH₂), 14.6 (CH₃).

v_{max} (neat): 3314 cm^{-1} (NH stretch), 1685 cm^{-1} (C=O stretch), 1646 cm^{-1}(C=O stretch), 1530 cm^{-1} (NH bend).

LCMS (Method 2 – Section 6.11): t_R = 7.39 min, *m/z* = 371 [M+H]$^+$

9*H*-Fluoren-9-yl (*S*)-(1-cyano-3-(methylthio)propyl)carbamate[103]

3.7
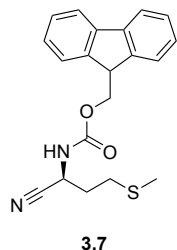
A suspension of (9*H*-fluoren-9-yl)methyl (*S*)-(1-amino-4-(methylthio)-1-oxobutan-2-yl)carbamate (800 mg, 2.16 mmol) in THF (45.0 mL) was stirred at 0 °C for 15 minutes. A solution of (TFA)$_2$O (360 µL, 2.38 mmol) and pyridine (360 µL, 4.75 mmol) was added drop-wise and the reaction was stirred at 0 °C for 3 h under argon. The reaction mixture was monitored by TLC (CHCl$_3$:MeOH, 9:1), and when there was a complete disappearance of starting material, 150 g of crushed ice was added to the reaction mixture and the corresponding precipitated nitrile product isolated as an off-white powder via vacuum filtration. The recovered filtrate was washed with H$_2$O (50.0 mL) and then lyophilised to provide the title compound as a colourless solid (685 mg, 1.94 mmol, 90%).

^{1}H NMR: δ (500 MHz, DMSO-d_6) 8.18 (d, *J* = 8.3 Hz, 1H, N*H*), 7.90 (d, *J* = 8.3 Hz, 2H, 2 × Fmoc-C*H*), 7.69 (d, *J* = 8.3 Hz, 2H, 2 × Fmoc-C*H*), 7.42 (t, *J* = 8.3 Hz, 2H, 2 × Fmoc-C*H*), 7.34 (t, *J* = 8.3 Hz, 2H, 2 × Fmoc-C*H*), 4.62 (q, *J* = 7.3 Hz, 1H, α-C*H*), 4.43 (d, *J* = 7.0 Hz, 1H, Fmoc-C*H$_2$*), 4.25 (t, *J* = 7.0 Hz, 1 H, Fmoc-C*H*), 2.53 (m, 2 H, γ-C*H$_2$* overlap with DMSO) 2.05 (s, 3 H, S-C*H$_3$*), 2.04-1.98 (m, 2H, β-C*H$_2$*).

^{13}C NMR: δ (126 MHz, DMSO-d_6) 155.4 (C), 143.6 (2 × C), 140.8 (2 × C), 127.7 (2 × CH), 127.1 (2 × CH), 125.0 (2 × CH), 120.1 (2 × CH), 119.3 (C≡N), 65.9 (CH$_2$), 46.6 (CH), 41.2 (CH), 31.1 (CH$_2$), 28.7 (CH$_2$), 14.4 (CH$_3$).

v_{max} (neat): 3308 cm^{-1} (NH stretch), 1688 cm^{-1} (C=O stretch), 1527 cm^{-1} (NH bend).

LCMS (Method 2 – Section 6.11): t_R = 8.36 min, *m/z* = 353 [M+H]$^+$

9*H*-Fluoren-9-yl (*S*)-(3-(methylthio)-1-(1*H*-tetrazol-5-yl)propyl)carbamate[103]

3.8
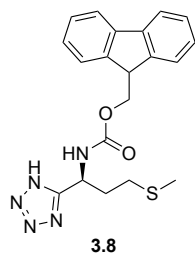
NaN$_3$ (370 mg, 5.69 mmol) and ZnBr$_2$ (319 mg, 1.42 mmol) were added to a stirred suspension of 9*H*-fluoren-9-yl-(*S*)-(1-cyano-3-(methylthio)propyl)carbamate (1.00 g, 2.83 mmol) in a mixture of H$_2$O (200 mL) and isopropanol (100 mL). The reaction mixture was heated to reflux (85 °C) with stirring for 16 h before it was acidified to pH 1 with conc. HCl. The resulting mixture was extracted with EtOAc (3 × 200 mL) with the organic phase then dried over Na$_2$SO$_4$ and concentrated under reduced pressure. The desired compound was isolated by column chromatography (CHCl$_3$:MeOH, 100:0 → 9:1) as an off-white solid (832 mg, 2.10 mmol, 74%).

^{1}H NMR: δ (500 MHz, DMSO-*d$_6$*) 8.09 (d, *J* = 7.8 Hz, 1H, N*H*), 7.89 (d, *J* = 7.5 Hz, 2H, 2 × Fmoc-C*H*), 7.73-7.67 (m, 2H, 2 × Fmoc-C*H*), 7.42 (t, *J* = 7.5 Hz, 2H, 2 × Fmoc-C*H*), 7.32 (q, *J* = 7.5 Hz, 2H, 2 × Fmoc-C*H*), 5.08-5.02 (m, 1H, α-C*H*), 4.40-4.19 (m, 3H, Fmoc-C*H* & C*H$_2$*), 2.53 (m, 2H, γ-C*H$_2$* overlap with DMSO), 2.22-2.09 (m, 2H, β-C*H$_2$*), 2.05 (s, 3H, SC*H$_3$*). *Note: One NH signal not observed.*

^{13}C NMR: δ (101 MHz, DMSO-*d$_6$*) 157.8 (C), 155.9 (C), 143.8 (C), 143.7 (C) 140.7 (C), 127.6 (CH), 127.0 (CH), 125.2 (CH), 120.1 (CH), 65.7 (CH$_2$), 46.7 (CH), 45.0 (CH), 32.3 (CH$_2$), 29.3 (CH$_2$), 14.5 (CH$_3$).

ν$_{max}$ (neat): 3293 cm^{-1} (NH stretch), 1685 cm^{-1} (C=O stretch), 1532 cm^{-1} (NH bend)

LCMS (Method 2 – Section 6.11): t$_R$ = 6.63 min, *m/z* = 394 [M-H]$^-$

(*S*)-3-(Methylthio)-1-(1*H*-tetrazol-5-yl)propan-1-amine[103]

1.145
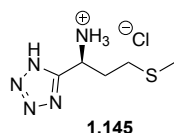
(9*H*-fluoren-9-yl)methyl-(*S*)-(3-(methylthio)-1-(1H-tetrazol-5-yl)propyl)carbamate (800 mg, 2.02 mmol) was dissolved in DCM (50.0 mL) and treated with Et_2NH (310 µL, 3.03 mmol) while stirring at room temperature for 1 h. The crude reaction was concentrated under reduced pressure with the residue dissolved in water (50.0 mL) and the mixture washed with diethyl ether (50.0 mL). Purification was carried out by ion-exchange chromatography (see Section 6.13) with the product obtained as an amorphous orange solid (228 mg, 1.09 mmol, 54%).

¹H NMR: δ (500 MHz, D_2O) 5.08-5.03 (m, 1H, α-C*H*), 2.60-2.34 (m, 4H, β-C*H₂* & γ-C*H₂*), 2.04 (s, 3H, C*H₃*).

¹³C NMR: δ (126 MHz, D_2O) 158.5 (C), 46.0 (CH), 31.4 (CH_2), 28.6 (CH_2), 14.0 (CH_3).

ν$_{max}$ (neat): 3040-2780 cm^{-1} (NH stretch), 1546 cm^{-1} (NH bend).

HRMS: $C_5H_{11}N_5S$ [M+H]$^+$ requires m/z 174.0808, found 174.0805.

N-(((9*H*-Fluoren-9-yl)oxy)carbonyl)-*S*-ethyl-*L*-homocysteine

3.9
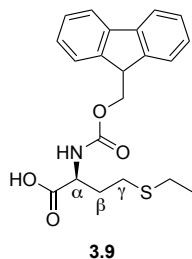
A solution of Fmoc-OSu (526 mg, 1.56 mmol) in acetonitrile (20.0 mL) was added dropwise to a solution of ʟ-ethionine (200 mg, 1.22 mmol) in 10% w/w Na_2CO_3 (40.0 mL) at 0 ºC and stirred for 16 h. The acetonitrile was removed under reduced pressure with the remaining aqueous phase acidified to pH 1 with aqueous HCl (1.00 M) and then extracted with EtOAc (3 × 50.0 mL). The organic layers were combined, washed with brine (1 × 50.0 mL) and dried over Na_2SO_4. Solvent was removed under reduced pressure and the crude residue was purified by column chromatography (Pet. Ether 40-60 ºC: EtOAc 100:0 → 9:1) to afford the title compound as a colourless solid (350 mg, 907 µmol, 74%).

¹H NMR: δ (500 MHz, CD₃OD) 7.78 (d, J = 7.2 Hz, 2H, 2 x Fmoc-C*H*), 7.67 (t, J = 7.2 Hz, 2H, 2 × Fmoc-C*H*), 7.38 (t, J = 7.4 Hz, 2H, 2 × Fmoc-C*H*), 7.30 (t, J = 7.4 Hz, 2H, 2 × Fmoc-C*H*), 4.39-4.31 (m, 3H, α-C*H* & Fmoc-C*H₂*), 4.21 (t, J = 7.2 Hz, 1H, Fmoc-C*H*), 2.67-2.59 (m, 1H, γ-C*H*$_a$H$_b$), 2.58-2.49 (m, 3H, -SC*H₂*CH₃ & γ-CH$_a$*H*$_b$), 2.14-1.89 (m, 2 H, β-C*H₂*), 1.23 (t, J = 7.2 Hz, 3H, -SCH₂C*H₃*).

¹³C NMR: δ (126 MHz, CD₃OD) 175.6 (C), 158.7 (C), 145.3 (C), 145.1 (C), 142.6 (2 × C), 128.8 (2 × CH), 128.1 (2 × CH), 126.2 (2 × CH), 120.9 (2 × CH), 67.9 (CH₂), 54.3 (CH), 48.4 (CH), 32.7 (CH₂), 28.7 (CH₂), 26.4 (CH₂), 15.1 (CH₃).

ν$_{max}$ (neat): 3317 cm⁻¹ (NH stretch), 3300-2700 cm⁻¹ (OH stretch), 1685 (C=O stretch).

HRMS: C₂₁H₂₃NO₄S [M+H]⁻ requires m/z 384.1275, found 384.1276.

9*H*-Fluoren-9-yl (*S*)-(1-amino-4-(ethylthio)-1-oxobutan-2-yl)carbamate

3.10
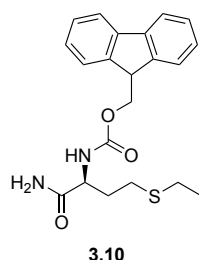
Fmoc-ʟ-ethionine (500 mg, 1.34 mmol) was dissolved in DMF:dioxane (1:1, 15.0 mL) followed by the sequential addition of (Boc)₂O (328 μL, 1.43 mmol), NH₄HCO₃ (113 mg, 1.43 mmol), and pyridine (115 μL, 1.43 mmol). The reaction was stirred at rt for 48 h under an argon atmosphere. The reaction mixture was monitored by TLC (CHCl₃:MeOH, 9:1), and when there was a complete disappearance of starting material, the reaction mixture was poured into EtOAc (100 mL) and washed sequentially with water (2 × 50.0 mL) and H₂SO₄ (20.0 mL, 0.94 M, 18.6 mmol). Solvent was removed under reduced pressure and the crude residue was purified by column chromatography (CHCl₃) to afford the title compound as a colourless solid (438 mg, 1.13 mmol, 84%).

¹H NMR: δ (500 MHz, DMSO-d_6) 7.89 (d, J = 7.3 Hz, 2H, 2 × Fmoc-C*H*), 7.73 (t, J = 6.8 Hz, 2H, 2 × Fmoc-C*H*), 7.48-7.26 (m, 6H, 4 × Fmoc-C*H* & N*H₂*), 7.02 (br. s, 1H, N*H*), 4.33-4.18 (m, 3H, Fmoc-C*H* & C*H₂*) 4.05-3.99 (m, 1H, α-C*H*), 1.94-1.72 (m, 2H, β-C*H₂*), 1.17 (t, J = 7.4 Hz, 3H, SCH₂C*H₃*).*

*γ-CH_2 & SCH_2CH$_3$ observed by HSQC (2.52 ppm & 2.51 ppm).

^{13}C NMR δ (126 MHz, DMSO-d_6) 173.4 (C=O), 156.0 (C=O), 143.9 (C), 143.7 (C), 140.7 (2 × C), 127.6 (2 × CH), 127.0 (2 × CH), 125.3 (2 × CH), 120.1 (2 × CH), 65.5 (CH$_2$), 53.7 (CH), 46.7 (CH), 32.0 (CH$_2$), 27.1 (CH$_2$), 24.7 (CH$_2$), 14.7 (CH$_3$).

ν$_{max}$ (neat): 3370 cm^{-1} (NH stretch), 3312 cm^{-1} (NH stretch), 1663 cm^{-1} (C=O stretch).

HRMS: $C_{21}H_{24}N_2O_3S$ [M+H]$^+$ requires m/z 385.1580, found 385.1583.

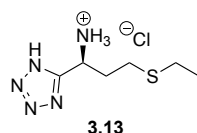9*H*-Fluoren-9-yl (*S*)-(1-cyano-3-(ethylthio)propyl)carbamate
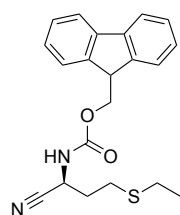
3.11

A suspension of (9*H*-fluoren-9-yl)methyl (*S*)-(1-amino-4-(methylthio)-1-oxobutan-2-yl)carbamate (263 mg, 684 μmol) in THF (15.0 mL) was stirred at 0 °C for 15 minutes. A solution of (TFA)$_2$O (310 μL, 2.20 mmol) and pyridine (310 μL, 1.50 mmol) was added dropwise and the reaction was left to stir for 3 h at 0 °C under argon. The reaction mixture was monitored by TLC (CHCl$_3$:MeOH, 9:1), and when there was a complete disappearance of starting material, 50 g of crushed ice was added to the reaction mixture and the corresponding precipitated nitrile product isolated as an off-white powder by vacuum filtration. The recovered precipitate was washed with water (200 mL) then lyophilised to provide the title compound as a colourless solid (178 mg, 486 μmol, 71%).

^{1}H NMR: δ (500 MHz, Acetone-d$_6$) 7.86 (d, *J* = 7.4 Hz, 2H, 2 × Fmoc-C*H*), 7.68 (d, *J* = 7.4 Hz, 2H, 2 × Fmoc-C*H*), 7.42 (t, *J* = 7.4 Hz, 2H, 2 × Fmoc-C*H*), 7.33 (t, *J* = 7.4 Hz, 2H, 2 × Fmoc-C*H*), 4.81 (q, *J* = 7.2 Hz, 1H, α-C*H*) 4.46 (d, *J* = 6.8 Hz, 2H, Fmoc-C*H$_2$*), 4.26 (t, *J* = 6.8 Hz, 1H, Fmoc-C*H*), 2.68 (t, *J* = 7.2 Hz, 2H, γ-C*H$_2$*), 2.57 (q, *J* = 7.2 Hz, 2H, SC*H$_2$*CH$_3$), 2.20-2.14 (m, 2H, β-CH$_2$), 1.23 (t, *J* = 7.2 Hz, 3H, SCH$_2$C*H$_3$*).

13C NMR δ (126 MHz, Acetone-d_6) 156.4 (C), 144.9 (C), 144.8 (C), 142.2 (C), 128.6 (CH), 128.0 (CH), 126.0 (CH), 120.9 (CH), 119.7 (C), 67.5 (CH$_2$), 48.0 (CH), 42.5 (CH), 33.4 (CH$_2$), 27.5 (CH$_2$), 26.0 (CH$_2$), 15.1 (CH$_3$).

HRMS: C$_{21}$H$_{22}$N$_2$O$_2$S [M+Na]$^+$ requires m/z 389.1294, found 389.1296.

(*S*)-3-(Ethylthio)-1-(1*H*-tetrazol-5-yl)propan-1-amine

3.13

NaN$_3$ (14.5 mg, 223 µmol) and ZnBr$_2$ (12.6 mg, 55.9 µmol) was added to a stirred suspension of 9*H*-fluoren-9-yl (*S*)-(1-cyano-3-(ethylthio)propyl)carbamate (44.7 mg, 122 µmol) in water:isopropanol (2:1, 30.0 mL). The reaction mixture was heated to 85 °C with stirring for 16 h Upon completion of the reaction as observed by TLC (CHCl$_3$:MeOH, 9:1), the reaction was cooled to room temperature acidified to pH 1 with conc. HCl. and EtOAc (50.0 mL) were added sequentially with continued stirring until no solid remained. The resulting mixture was extracted with EtOAc (3 × 50 mL) with the organic fractions combined, dried over Na$_2$SO$_4$ and concentrated under reduced pressure. The crude residue was purified by column chromatography (DCM:MeOH 100 → 9:1) and used directly in the next step of the synthesis. Diethylamine (3.00 mL, 29.0 mmol) was added to a solution of (9*H*-fluoren-9-yl)methyl (*S*)-(3-(methylthio)-1-(1H-tetrazol-5-yl)propyl)carbamate (50.0 mg, 122 µmol) in DCM (30.0 mL) and was stirred at rt for 30 minutes. The crude reaction was concentrated under reduced pressure and the recovered residue was dissolved in water (50.0 mL) before washing with diethyl ether (50.0 mL). The resulting aqueous phase was concentrated under reduced pressure and was purified by ion exchange chromatography (See Section 6.13). Product containing fractions were dried by lyophilisation to afford the title compound as an orange solid (11.4 mg, 60.9 µmol, 50%).

1H NMR: δ (500 MHz, CD$_3$CN) 8.61 (br. s, 2H, N*H$_2$*), 5.08 (t, *J* = 6.3 Hz, 1H, α-C*H*), 2.67-2.43 (m, 6H, 3 × C*H$_2$*), 1.18 (t, *J* = 7.3 Hz, 3H, C*H$_3$*).
Note: NH not observed.

13C NMR δ (126 MHz, CD$_3$CN) 155.9 (C), 46.7 (CH), 32.7 (CH$_2$), 27.1 (CH$_2$), 25.9 (CH$_2$), 15.0 (CH$_3$).

RP-HPLC (Method 2 – Section 6.10) t_R = 5.07 min.

Note: Detection at 214 nm for this compound.

HRMS: $C_6H_{13}N_5S$ [M+H]$^+$ requires m/z 188.0964, found 188.0964.

(2*R*,3*R*,4*R*,5*R*)-2-(Acetoxymethyl)-5-(2-amino-6-chloro-9*H*-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate[140]
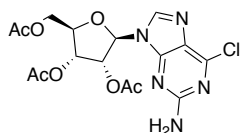
3.19

To a suspension of 6-chloroguanineriboside (1.25 g, 4.14 mmol) and DMAP (50.0 mg, 409 µmol) in anhydrous acetonitrile (50.0 mL) was added triethylamine (4.48 mL, 32.2 mmol) and acetic anhydride (1.17 mL, 12.4 mmol) under argon. The reaction was stirred at rt for 16 h before the reaction mixture was concentrated under reduced pressure. The crude residue was recrystallised from hot isopropanol to afford the title compound as a colourless solid (1.58 g, 3.68 mmol, 89%).

^{1}H NMR: δ (500 MHz, CDCl$_3$) 7.88 (s, 1H, 8-*H*), 6.00 (d, *J* = 4.8 Hz, 1H, 1′-*H*), 5.95 (t, *J* = 5.2 Hz, 2′-*H*), 5.75 (t, *J* = 4.8 Hz 1H, 3′-*H*), 5.23 (br. s, 2H, N*H₂*) 4.47-4.33 (m, 3H, 4′-*H* & 5′-*H*$_a$H$_b$ & 5′-H$_a$*H*$_b$) 2.15 (s, 3H C*H₃*), 2.11 (s, 3H, C*H₃*), 2.09 (s, 3H, C*H₃*).

^{13}C NMR: δ (126 MHz, CDCl$_3$) 170.5 (C=O), 169.6 (C=O), 169.3 (C=O), 159.1 (C), 153.1 (C), 152.0 (C), 140.7 (CH), 125.9 (C), 86.6 (CH), 80.0 (CH), 72.7 (CH), 70.5 (CH), 62.9 (CH$_2$), 20.7 (CH$_3$), 20.5 (CH$_3$), 20.4 (CH$_3$).

v_{max} (neat): 2914 cm^{-1} (Aromatic CH stretch), 2843 cm^{-1} (NH stretch), 1164 cm^{-1} (C-O stretch ester).

HRMS: $C_{16}H_{18}{}^{35}ClN_5O_7$ [M+H]$^+$ requires m/z 428.0968, found 428.0968.

(2*R*,3*R*,4*R*,5*R*)-2-(Acetoxymethyl)-5-(6-chloro-2-iodo-9*H*-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate[140]
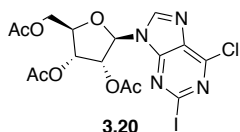
3.20

To a suspension of (2*R*,3*R*,4*R*,5*R*)-2-(acetoxymethyl)-5-(6-chloro-2-iodo-9*H*-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate (496 mg, 1.16 mmol), copper(I) iodide (442 mg, 2.32 mmol) and iodine (294 mg, 1.16 mmol) in THF (20.0 mL) was added CH_2I_2 (940 μL, 11.6 mmol) followed by isoamyl nitrite (470 μL, 3.48 mmol). The mixture was heated to reflux for 3 h, then cooled to rt and filtered through a pad of celite. The filtrate was concentrated under reduced pressure and purified by column chromatography ($CHCl_3$:EtOH 100:0 → 4:1). The product-containing fractions were combined and concentrated under reduced pressure with the resultant gum recrystallised from EtOH to afford the title compound as a yellow powder (373 mg, 692 μmol, 60%).

^{1}H NMR: δ (500 MHz, CDCl$_3$) 8.20 (s, 1H, 8-*H*), 6.20 (d, *J* = 5.3 Hz, 1H, 1′-*H*), 5.79 (t, *J* = 5.3 Hz, 1H, 2′-*H*) 5.59 (t, *J* = 5.3 Hz, 1H, 3′-*H*) 4.54-4.36 (m, 3H, 4′-*H* & 5′-*H*$_a$H$_b$ & 5′-H$_a$*H*$_b$), 2.18 (s, 3H, C*H$_3$*), 2.15 (s, 3H, C*H$_3$*), 2.11 (s, 3H, C*H$_3$*).

^{13}C NMR: δ (101 MHz, CDCl$_3$) 170.3 (C=O), 169.7 (C=O), 169.5 (C=O), 152.1 (C), 151.2 (C), 143.2 (CH), 132.4 (C), 117.1 (C), 86.8 (CH), 81.0 (CH), 73.5 (CH), 70.7 (CH), 63.1 (CH$_2$), 21.0 (CH$_3$), 20.7 (CH$_3$), 20.5 (CH$_3$).

ν$_{max}$ (neat): 1740 cm^{-1} (C=O stretch), 1207 cm^{-1} (C-O stretch).

HRMS: $C_{16}H_{16}{}^{35}ClIN_4O_7$ [M+H]$^+$ requires m/z 538.9830, found 538.9823.

(2*R*,3*R*,4*S*,5*R*)-2-(6-Amino-2-iodo-9*H*-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol[140]
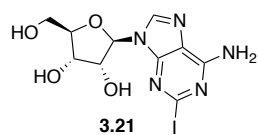
3.21

A suspension of (2*R*,3*R*,4*R*,5*R*)-2-(acetoxymethyl)-5-(6-chloro-2-iodo-9*H*-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate (500 mg, 928 μmol) in MeOH (10.0 mL) was cooled to 0 °C. Ammonia in methanol solution (500 μl, 7.00 M, 3.50 mmol) was added before the reaction vessel was sealed and heated to 60 °C for 16 h. The vessel was cooled to rt, and degassed by bubbling argon through it for 1 h. Solvent was removed under reduced pressure with the resulting residue purified by column chromatography (DCM:MeOH 100:0 → 9:1) to afford the title compound as an off white solid (260 mg, 661 μmol, 71%).

**^{1}H NMR:** δ (500 MHz, DMSO-d_6) 8.29 (s, 1H, 8-*H*) 7.71 (br. s, 2H, N*H$_2$*), 5.80 (d, *J* = 5.7 Hz, 1H, 1′-*H*), 5.48 (d, *J* = 5.7 Hz, 1H, 2′-O*H*) 5.22 (d, *J* = 5.7 Hz, 1H, 3′-O*H*), 5.06 (t, *J* = 5.7 Hz, 1H, 5′-O*H*), 4.52 (q, *J* = 5.7 Hz, 1H, 2′-*H*) 4.15 – 4.11 (m, 1H, 3′-*H*), 3.96 – 3.93 (m, 1H, 4′-*H*) 3.68 – 3.50 (m, 2H, 5′-C*H$_2$*).

^{13}C NMR: δ (126 MHz, DMSO-d_6) 156.4 (C), 150.2 (C), 139.9 (CH), 121.3 (C), 119.5 (C), 87.6 (CH), 86.3 (CH), 74.1 (CH), 71.0 (CH), 61.9 (CH$_2$).

ν$_{max}$ (neat): 3325 cm^{-1} (NH stretch), 3200-2800 cm^{-1} (OH stretch).

HRMS: C$_{10}$H$_{12}$IN$_5$O$_4$ [M+H]$^+$ requires m/z 394.0007, found 394.0009.

(2*R*,3*R*,4*S*,5*R*)-2-(6-Amino-2-((trimethylsilyl)ethynyl)-9*H*-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol[140]
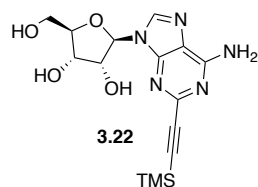
To a solution of Pd(PPh$_3$)$_2$Cl$_2$ (17.6 mg, 25.1 μmol), copper(I) iodide (4.80 mg, 25.1 μmol) and 2-iodoadenosine (100 mg, 254 μmol) in DMF (1.60 mL) was was added triethylamine (50.0 μL, 359 25.1 μmol) followed by ethynyltrimethylsilane (390 μL, 276 mg, 2.73 mmol). The mixture was stirred at rt 16 h followed by concentration under reduced pressure. The residue was diluted with EtOAc (15.0 mL) and washed with 10% w/v aqueous disodium EDTA solution (30.0 mL) before being dried over Na$_2$SO$_4$ and concentrated under reduced pressure. The crude residue was purified by column chromatography (CHCl$_3$:MeOH, 100:0 → 95:5) to afford the desired product as a light brown solid (170 mg, 468 μmol 63%).

^{1}H NMR: δ (500 MHz, DMSO-d_6) 8.44 (s, 1H, 8-*H*) 7.50 (br. s, 2H, N*H$_2$*), 5.87 (d, *J* = 5.7 Hz, 1H, 1′-*H*), 5.45 (d, *J* = 5.7 Hz, 1H, 2′-O*H*) 5.17 (d, *J* = 5.7 Hz, 1H, 3′-O*H*), 5.14-5.10 (m, 1H, 5′-O*H*), 4.48 (q, *J* = 5.7 Hz, 1H, 2′-O*H*) 4.14-4.10 (m, 1H, 3′-*H*), 3.96-3.93 (m, 1H, 4′-*H*) 3.68-3.50 (m, 2H, 5′-C*H$_2$*), 0.24 (s, 9H, (C*H$_3$*)$_3$Si).

^{13}C NMR: δ (126 MHz, DMSO-d_6) 155.8 (C), 149.3 (C), 144.8 (C), 140.3 (CH), 118.8 (C), 104.5 (C), 88.6(C), 87.0 (CH), 85.6 (CH), 73.9 (CH), 70.3 (CH) 61.3 (CH$_2$), 0.41 (3 × CH$_3$).

ν_{max} **(neat):** 3400-2600 cm^{-1} (OH stretch), 2031 cm^{-1} (C≡C stretch), 1595 cm^{-1} (N-H bend).

HRMS: C$_{15}$H$_{21}$N$_5$O$_4$Si [M+H]$^+$ requires m/z 364.1440, found 364.1441.

(2R,3R,4S,5S)-2-(6-Amino-2-ethynyl-9*H*-purin-9-yl)-5-(chloromethyl)tetrahydrofuran-3,4-diol[140]
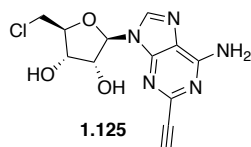
1.125

To an ice-cooled suspension of (2*R*,3*R*,4*S*,5*R*)-2-(6-amino-2-((trimethylsilyl)ethynyl)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol (50.0 mg, 138 µmol) in acetonitrile (3.00 mL) was added pyridine (22.0 µL, 273 µmol). Thionyl chloride (50.0 µL, 689 µmol) was then added dropwise with the reaction left to stir at 0 ºC for 3 h. The reaction was then allowed to warm to rt and was stirred for a further 16 h. Solvent was removed under reduced pressure and the reaction mixture was purified by column chromatography (CHCl$_3$:MeOH, 100:0 → 95:5) to afford the title compound as a colourless solid (22.5 mg, 72.6 µmol, 53%).

^{1}H NMR: δ (500 MHz, DMSO-d_6) 8.42 (s, 1H, 8-*H*) 7.51 (br. s, 2H, N*H$_2$*), 5.91 (d, *J* = 5.8 Hz, 1H, 1′-*H*), 5.59 (d, *J* = 5.8 Hz, 1H, 2′-O*H*) 5.45 (d, *J* = 5.8 Hz, 1H, 3′-O*H*), 4.70 (t, *J* = 5.8 Hz, 1H, 2′-*H*), 4.22–4.16 (m, 1H, 3′-*H*), 4.13-4.07 (m, 1H, 4′-*H*), 4.01 (s, 1H, C≡C*H*) 3.95 (dd, *J* = 11.6, 5.7 Hz, 1H, 5′-*H$_a$*H$_b$), 3.86 (dd, *J* = 11.6, 5.7 Hz, 1H, 5′-H$_a$*H$_b$*).

^{13}C NMR: δ (126 MHz, DMSO-d_6) 156.4 (C), 149.8 (C), 145.3 (CH), 141.1 (C), 119.4 (C), 87.7 (CH), 84.3 (CH), 83.8(*C*≡CH), 75.3 (C≡*C*H), 73.2 (CH), 71.7 (CH), 45.2 (CH$_2$).

ν_{max} **(neat):** 3369 cm^{-1} (NH stretch), 3300-2700 cm^{-1} (OH stretch), 2106 cm^{-1} (Alkyne CH stretch).

RP-HPLC (Method 2 – Section 6.10) t$_R$ = 6.96 min.

HRMS: C$_{12}$H$_{12}$ClN$_5$O$_3$ [M+H]$^+$ requires m/z 310.0701, found 310.0706.

(2*R*,3*R*,4*S*,5*S*)-2-(6-Amino-2-chloro-9*H*-purin-9-yl)-5-(chloromethyl)tetrahydrofuran-3,4-diol
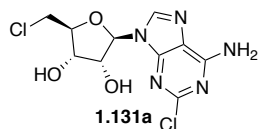
1.131a

To an ice-cooled suspension of 6-chlororiboside (500 mg, 1.66 mmol) in acetonitrile (25.0 mL) was added pyridine (250 μL, 3.10 mmol). Thionyl chloride (600 μL, 8.27 mmol) was added dropwise with the reaction left to stir at 0 °C for 3 h. The reaction was allowed to warm to rt and was stirred for a further 16 h. The reaction mixture was poured into ethyl acetate (25.0 mL) and the resulting precipitate isolated by filtration and dissolved in water:MeOH (5:1, 50.0 mL). Ammonium hydroxide (5.10 mL, 13.4 M, 68.3 mmol) was added and the reaction stirred at rt for 30 minutes. Solvent was removed under reduced pressure and the reaction mixture was purified by column chromatography (CHCl$_3$:MeOH, 100:0 → 9:1) to afford the title compound as an off-white solid (519 mg, 1.62 mmol, 98%).

¹H NMR δ (400 MHz, DMSO-*d*$_6$) 8.37 (s, 1H, 8-*H*), 7.85 (br. s., 2H, N*H*$_2$) 5.86 (d, *J* = 5.7 Hz, 1H, 1′-*H*), 5.51 (br. s, 2H, 2′-O*H* & 3′-O*H*) 4.66 (t, *J* = 5.7 Hz, 1H, 2′-*H*) 4.19-4.17 (m, 1H, 3′-*H*) 4.12-4.07 (m, 1H, 4′-*H*) 3.93 (dd, *J* = 11.5, 5.7 Hz, 1H, 5′-*H*$_a$H$_b$) 3.85 (dd, *J* = 11.5, 5.7 Hz, 1H, 5′-H$_a$*H*$_b$).
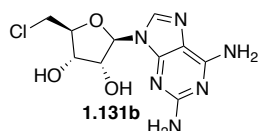
¹³C NMR δ (126 MHz, DMSO-*d*$_6$) 156.8 (C) 153.1 (C) 150.4 (C) 140.1 (CH) 118.1 (C) 87.3 (CH) 83.9 (CH) 72.7 (CH) 71.2 (CH) 44.7 (CH$_2$).

ν$_{max}$ (neat): 3400-3000 cm⁻¹ (OH stretch), 3299 cm⁻¹ (NH stretch), 1649 cm⁻¹ (NH bend).

RP-HPLC (Method 2 – Section 6.10) t$_R$ = 7.38 min.

HRMS: C$_{10}$H$_{11}$³⁵Cl$_2$N$_5$O$_3$ [M+H]⁺ requires *m/z* 320.0312, found 320.0317.

(2*S*,3*S*,4*R*,5*R*)-2-(Chloromethyl)-5-(2,6-diamino-9*H*-purin-9-yl)tetrahydrofuran-3,4-diol

1.131b

To an ice-cooled suspension of 2-aminoadenosine (1.00 g, 3.54 mmol) in acetonitrile (50.0 mL) was added pyridine (570 μL, 7.08 mmol). Thionyl chloride (1.29 mL, 17.7 mmol) was added dropwise and the reaction was stirred at 0 °C for 3 h. The reaction was then allowed to

warm to rt and was stirred for a further 16 h. Ammonium hydroxide (2.00 mL, 13.4 M, 26.8 mmol) was added and the reaction stirred at rt for 30 minutes. Solvent was removed under reduced pressure and the reaction mixture was purified by column chromatography (DCM:MeOH:NH$_4$OH -100:4:0.1) to afford the title compound as a colourless solid (462 mg, 1.54 mmol, 44%).

^{1}H NMR: δ (500 MHz, DMSO-d_6) 7.90 (s, 1H, 8-*H*) 6.71 (br. s, 2H, N*H$_2$*), 5.83 (br. s, 2H, N*H$_2$*) 5.76 (d, *J* = 5.7 Hz, 1H, 1′-*H*), 5.50 (d, *J* = 5.7 Hz, 1H, 2′-O*H*), 5.33 (d, *J* = 5.7 Hz, 1H, 3′-O*H*), 4.65 (q, *J* = 5.7 Hz, 1H, 2′-*H*), 4.16-4.12 (m, 1H, 3′-*H*) 4.05 – 4.01 (m, 1H, 4′-*H*), 3.92 (dd, *J* = 11.4, 5.7 Hz, 1 H, 5′-*H$_a$*H$_b$), 3.81 (dd, 1H, *J* = 11.4, 5.7 Hz, 5′-H$_a$*H$_b$*).

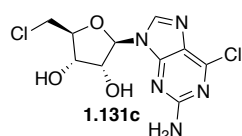**^{13}C NMR:** δ (126 MHz, DMSO-d_6) 160.3 (C), 156.2 (C), 151.9 (C), 135.9 (CH), 113.3 (C) 86.6 (CH) 83.4 (CH), 72.3 (CH), 71.3 (CH), 44.9 (CH$_2$).

ν$_{max}$ (neat): 3474 cm^{-1} and 3438 cm^{-1} (NH stretch), 3300-2700 cm^{-1} (OH stretch).

RP-HPLC (Method 2 – Section 6.10) t_R = 4.69 min.

HRMS: C$_{10}$H$_{13}$^{35}ClN$_6$O$_3$ [M+H]$^+$ requires *m/z* 301.0810, found 301.0814.

(2*R*,3*R*,4*S*,5*S*)-2-(2-Amino-6-chloro-9*H*-purin-9-yl)-5-(chloromethyl)tetrahydrofuran-3,4-diol

To an ice-cooled suspension of 6-chlororiboside (500 mg, 1.66 mmol) in acetonitrile (25.0 mL) was added pyridine (250 μL, 3.10 mmol). Thionyl chloride (600 μL, 8.27 mmol) was then added dropwise and the reaction mixture was stirred at 0 °C for 3 h. The reaction was allowed to warm to rt and was stirred for a further 16 h. The reaction mixture was poured into ethyl acetate (25.0 mL) with the resulting precipitate isolated by filtration and dissolved in water:MeOH (5:1, 50.0 mL). Ammonium hydroxide (5.10 mL, 13.4 M, 68.3 mmol) was added and the reaction stirred at rt for 30 minutes. Solvent was removed under reduced pressure and the reaction mixture was purified by column chromatography (CHCl$_3$:MeOH, 100:0 → 9:1) to afford the title compound as a pale yellow solid (324 mg, 1.01 mmol, 61%).

¹H NMR δ (500 MHz, DMSO-d_6) 8.33 (s, 1H, 8-*H*), 7.01 (br. s, 2H, N*H₂*) 5.84 (d, J = 5.8 Hz, 1H, 1′-*H*), 5.61 (d, J = 5.8 Hz, 1H, 2′-O*H*), 5.43 (d, J = 5.8 Hz, 1H, 3′-O*H*), 4.68 (q, J = 5.8 Hz, 1H, 2′-*H*) 4.19-4.15 (m, 1H, 3′-*H*) 4.10-4.05 (m, 1H, 4′-*H*) 3.93 (dd, J = 11.6, 5.8 Hz, 1H, 5′-*H*ₐHᵦ) 3.83 (dd, J = 11.6, 5.8 Hz, 1H, 5′-Hₐ*H*ᵦCl).

¹³C NMR δ (126 MHz, DMSO-d_6) 159.8 (C) 154.0 (C) 149.6 (C) 141.5 (CH) 123.6 (C) 87.0 (CH) 83.8 (CH) 72.3 (CH) 71.2 (CH) 44.7 (CH₂).

RP-HPLC (Method 2 – Section 6.10) t_R = 6.98 min.

HRMS: $C_{10}H_{11}{}^{35}Cl_2N_5O_3$ [M+H]⁺ requires m/z 320.0312, found 320.0317.

(2*R*,3*R*,4*S*,5*S*)-2-(4-Amino-7*H*-pyrrolo[2,3-d]pyrimidin-7-yl)-5-(chloromethyl)tetrahydrofuran-3,4-diol

3.23
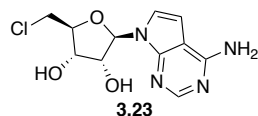
To an ice-cooled suspension of 7-Deazaadenosine (53.0 mg, 199 µmol) in acetonitrile (3.00 mL) was added pyridine (31.0 µL, 385 µmol). Thionyl chloride (73.0 µL, 1.00 mmol) was then added dropwise and the reaction mixture was stirred at 0 ºC for 3 h. The reaction was allowed to warm to rt and was stirred for a further 16 h. The solvent was removed under reduced pressure and the residue was redissolved in 5:1 MeOH:H₂O (6.00 mL). Ammonium hydroxide (2.00 mL, 1.00 M, 2.00 mmol) was added and the reaction mixture stirred for 3 h. Solvent was removed under reduced pressure and the crude residue was purified by column chromatography (CHCl₃:MeOH, 100:0 → 9:1) to provide the title compound as an off-white solid (35.0 mg, 123 µmol, 62%).

¹H NMR: δ (500 MHz, DMSO-d_6) 8.08 (s, 1H, 2-*H*), 7.32 (d, J = 3.6 Hz, 1H, 8-*H*), 7.02 (br. s, 2H, N*H₂*) 6.62 (d, J = 3.6 Hz, 1H, 7-*H*), 6.11 (d, J = 5.6 Hz, 1H, 1′-*H*), 5.42 (d, J = 6.1 Hz, 1H, 2′-O*H*), 5.35 (d, J = 6.1 Hz, 1H, 3′-O*H*) 4.50-4.45 (m, 1H, 4′-*H*) 4.15-4.10 (m, 1H, 2′-*H*) 4.08-4.01 (m, 1H, 3′-*H*), 3.92 (dd, J = 11.7, 4.6 Hz, 1 H, 5′-*H*ₐHᵦ), 3.80 (dd, J = 11.7, 4.6 Hz, 1 H, 5′-Hₐ*H*ᵦ).

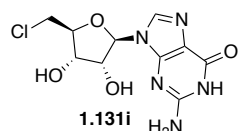**¹³C NMR:** δ (126 MHz, DMSO-d_6) 157.9 (C), 152.3 (CH), 151.0 (C), 122.0 (CH) 103.2 (C), 100.6 (CH) 87.3 (CH), 83.2 (CH), 73.6 (CH) 71.7 (CH) 45.6 (CH₂).

ν$_{max}$ (neat): 3335 cm^{-1} (NH stretch), 3300-2700 cm^{-1} (OH stretch), 1591 cm^{-1} (C=C stretch).

RP-HPLC (Method 2 – Section 6.10) t$_R$ = 4.92 min.

HRMS: C$_{11}$H$_{12}$^{35}ClN$_5$O [M+H]$^+$ requires *m/z* 285.0749, found 285.0753.

2-Amino-9-((2*R*,3*R*,4*S*,5*S*)-5-(chloromethyl)-3,4-dihydroxytetrahydrofuran-2-yl)-5,9-dihydro-6*H*-purin-6-one

1.131i

To an ice-cooled suspension of guanosine (470 mg, 1.66 mmol) in acetonitrile (30.0 mL) was added pyridine (267 µL, 3.32 mmol). Thionyl chloride (602 µL, 8.30 mmol) was then added dropwise and the reaction mixture was stirred at 0 °C for 3 h. The reaction was allowed to warm to rt and was stirred for a further 16 h. The reaction mixture was poured into ethyl acetate (25.0 mL) with the resulting precipitate isolated by filtration and dissolved in water:MeOH (5:1, 50.0 mL). Ammonium hydroxide (5.10 mL, 13.4 M, 68.3 mmol) was added and the reaction stirred at rt for 30 minutes. Solvent was removed under reduced pressure and the reaction mixture was purified by column chromatography (CHCl$_3$:MeOH, 100:0 → 9:1). Further purification was carried out by RP-HPLC to afford the title compound as a colourless solid (5.50 mg, 18.2 µmol, 1%)

^{1}H NMR: δ (500 MHz, DMSO-*d$_6$*) 10.56 (s, 1H, 1-N*H*) 7.82 (s, 1H, 8-*H*), 6.41 (br. s, 2H, N*H$_2$*) 5.66 (d, *J* = 5.7 Hz, 1H, 1′-*H*), 5.48 (d, *J* = 6.0 Hz, 1H, 2′-O*H*), 5.30 (d, *J* = 5.7 Hz, 1H, 3′-O*H*), 4.50 (q, *J* = 5.7 Hz, 1H, 4′-*H*) 4.07-4.03 (m, 1H, 2′-*H*) 3.99-3.94 (m, 1H, 3′-*H*), 3.83 (dd, *J* = 11.6, 5.7 Hz, 1 H, 5′-*H$_a$*H$_b$), 3.73 (dd, *J* = 11.6, 5.7 Hz, 1 H, 5′-H$_a$*H$_b$*).

^{13}C NMR: δ (126 MHz, DMSO-*d$_6$*) 157.2 (C), 154.2 (C), 151.9 (C), 136.2 (CH) 117.3 (C), 87.0 (CH) 84.0 (CH), 73.1 (CH), 71.7 (CH) 45.3 (CH).

RP-HPLC (Method 3 – Section 6.10) t$_R$ = 5.00 min.

HRMS: C$_{10}$H$_{12}$^{35}ClN$_5$O$_4$ [M+H]$^+$ requires *m/z* 302.0651, found 302.0657.

(2*S*,3*S*,4*R*,5*R*)-2-(Chloromethyl)-5-(6-hydroxy-9*H*-purin-9-yl)tetrahydrofuran-3,4-diol

3.24
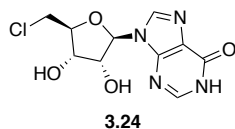
To an ice-cooled suspension of inosine (200 mg, 746 µmol) in acetonitrile (5.00 mL) was added pyridine (120 µL, 1.50 mmol). Thionyl chloride (270 µL, 3.75 mmol) was then added dropwise and the reaction mixture was stirred at 0 ºC for 3 h. The reaction mixture was allowed to warm to rt and was stirred for a further 16 h. The resulting precipitate was filtered and dissolved in water/MeOH (5:1, 12.0 mL). Ammonium hydroxide (400 µL, 6.58 M, 2.63 mmol) was added and the reaction stirred at rt for 30 minutes. The solvent was removed under reduced pressure to afford the title compound as a colourless amorphous solid (147 mg, 513 µmol, 69%).

¹H NMR δ (500 MHz, DMSO-d_6) 8.31 (s, 1H, 2-*H*), 8.07 (s, 1H, 8-*H*), 7.87 (br. s., 1H, N*H*), 5.91 (d, *J* = 5.2 Hz, 1H, 1′-*H*), 5.66 (br. s., 1H, 2′-O*H*), 5.50 (br. s., 1H, 3′-O*H*), 4.64 (t, *J* = 5.2 Hz, 1H, 4′-*H*), 4.20-4.16 (m, 1H, 2′-*H*), 4.12-4.08 (m, 1H, 3′-*H*), 3.93 (dd, *J* = 11.7, 5.2 Hz, 1H, 5′-*H*$_a$H$_b$), 3.83 (dd, *J* = 11.7, 5.2 Hz, 1H, 5′-H$_a$*H*$_b$).

¹³C NMR δ (126 MHz, DMSO-d_6) 156.5 (C) 148.3 (C) 146.0 (CH) 138.9 (CH) 124.5 (C) 87.4 (CH) 83.8 (CH) 73.1 (CH) 71.1 (CH) 44.7 (CH$_2$).

RP-HPLC (Method 3 – Section 6.10) t_R = 6.53 min.

HRMS: C$_{10}$H$_{11}$^{35}ClN$_4$O$_4$ [M+H]$^+$ requires *m/z* 287.0542, found 287.0546.

(2*R*,3*R*,4*S*,5*S*)-2-(2-Amino-9*H*-purin-9-yl)-5-(chloromethyl)tetrahydrofuran-3,4-diol

1.131d
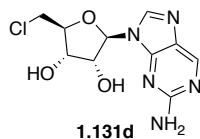
Triethylamine (1.80 mL, 12.9 mmol) was added to 5′,6-chloroguanineriboside (180 mg, 562 µmol) and Pd/C (59.6 mg, 56.2 µmol, 10 mol%) in THF (18.0 mL). The reaction mixture was stirred under a hydrogen atmosphere at rt for 16 h. The reaction mixture was filtered over celite and concentrated under reduced pressure. The crude residue was purified by column chromatography (CH$_2$Cl$_2$:MeOH 100:4) to afford the title compound as a pale yellow solid (101 mg, 353 µmol, 63%).

¹H NMR δ (500 MHz, DMSO-d_6) 8.61 (s, 1H, 6-*H*), 8.27 (s, 1H, 8-*H*) 6.62 (s, 2H, N*H₂*) 5.87 (d, *J* = 5.8 Hz, 1H, 1′-*H*), 5.59 (br. s, 1H, 2′-O*H*), 5.41 (br. s, 1H, 3′-O*H*), 4.74-4.70 (m, 1H, 2′-*H*) 4.19-4.15 (m, 1H, 3′-C*H*) 4.10-4.05 (m, 1H, 4′-*H*) 3.93 (dd, *J* = 11.7, 5.8 Hz, 1H, 5′-*H*ₐHᵦ) 3.83 (dd, *J* = 11.7, 5.8 Hz, 1H, 5′-CHₐ*H*ᵦ).

¹³C NMR δ (126 MHz, DMSO-d_6) 160.4 (C) 153.1 (C) 149.2 (CH) 141.1 (CH) 127.1 (C) 86.5 (CH) 83.7 (CH) 72.2 (CH) 71.3 (CH) 44.8 (CH₂).

ν_max (neat): 3500-2700 cm⁻¹ (OH stretch), 3334 cm⁻¹ (NH stretch), 1653 cm⁻¹ (NH bend).

RP-HPLC (Method 2 – Section 6.10) t_R = 4.92 min.

HRMS: C₁₀H₁₂³⁵ClN₅O₃ [M+H]⁺ requires m/z 286.0701, found 286.0706.

(2*S*,3*S*,5*R*)-5-(6-Amino-9*H*-purin-9-yl)-2-(chloromethyl)tetrahydrofuran-3-ol

3.25
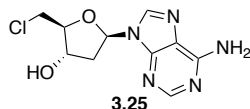
To an ice cooled suspension of 2′-deoxyadenosine (298 mg, 1.19 mmol) in acetonitrile (12.5 mL) was added pyridine (192 µL, 2.37 mmol). Thionyl chloride (433 µL, 5.93 mmol) was then added dropwise and the reaction mixture was left to stir at 0 °C for 3 h. The reaction was then allowed to warm to rt and was stirred for a further 16 h. The mixture was then extracted with ethyl acetate (3 × 100 mL) with the organic fractions combined and washed with saturated aqueous NaHCO₃ (100 mL) and water (100 mL). The organic phase was then dried over Na₂SO₄ and concentrated under reduced pressure. The crude residue was dissolved in toluene (10.0 mL) and cooled to 4 ºC. The resulting precipitate was isolated by filtration to provide the target compound as a pale yellow solid (177 mg, 656 µmol, 55%).

¹H NMR δ (500 MHz, DMSO-d_6) 8.33 (s, 1H, 2-*H*) 8.15 (s, 1H, 8-*H*) 7.27 (br. s., 2H, N*H₂*) 6.39 (t, *J* = 6.9 Hz, 1H, 1′-*H*) 5.56-5.53 (m, 1H, 3′-O*H*) 4.49-4.45 (m, 1H, 4′-*H*) 4.05-4.00 (m, 1H, 3′-*H*) 3.90 (dd, *J* = 11.3, 6.2 Hz, 1H, 5′-*H*ₐHᵦ) 3.78 (dd, *J* = 11.3, 6.2 Hz, 1H, 5′-Hₐ*H*ᵦ) 2.96-2.89 (m, 1H, 2′-*H*ₐHᵦ) 2.37-2.30 (m, 1H, 2′-Hₐ*H*ᵦ)
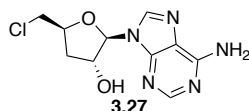
^{13}C NMR δ (126 MHz, DMSO-d_6) 156.0 (C), 152.6 (CH), 149.1 (C), 139.5 (CH), 119.2 (C), 86.2 (CH), 83.5 (CH), 71.5 (CH), 44.6 (CH$_2$), 37.9 (CH$_2$).

RP-HPLC (Method 3 – Section 6.10) t_R = 6.44 min.

HRMS: C$_{10}$H$_{12}$^{35}ClN$_5$O$_2$ [M+H]$^+$ requires *m/z* 270.0752, found 270.0756.

(2*R*,3*R*,5*S*)-2-(6-Amino-9*H*-purin-9-yl)-5-(chloromethyl)tetrahydrofuran-3-ol

3.27

To an ice-cooled suspension of 3′-deoxyadenosine (50.3 mg, 200 μmol) in acetonitrile (3.00 mL) was added pyridine (31.0 μL, 385 μmol). Thionyl chloride (73.0 μL, 1.00 mmol) was then added dropwise and the reaction mixture was stirred at 0 °C for 3 h. The reaction was allowed to warm to rt and was stirred for a further 16 h. Solvent was removed under reduced pressure and the crude residue was purified by column chromatography (CHCl$_3$:MeOH 100:0 → 95:5) to provide the title compound as a white solid (33.0 mg, 122 μmol, 61%).

^{1}H NMR: δ (500 MHz, DMSO-d_6) 8.26 (s, 1H, 2-*H*), 8.16 (s, 1H, 8-*H*), 7.27 (br. s, 2H, N*H$_2$*), 5.92 (d, *J* = 5.2 Hz 1H, 1′-*H*), 5.75 (d, *J* = 5.2 Hz 1H, 2′-O*H*) 4.75–4.72 (m, 1H, 2′-C*H*), 4.54–4.49 (m, 1H, 4′-*H*), 3.90 (dd, *J* = 11.4, 5.2 Hz 1H, 5′-*H$_a$*H$_b$), 3.84 (dd, *J* = 11.4, 5.2 Hz, 1H, 5′-H$_a$*H$_b$*), 2.39-2.33 (m, 1H, 3′-*H$_a$*H$_b$), 2.12-2.07 (m, 1H, 3′-H$_a$*H$_b$*).

^{13}C NMR: δ (126 MHz, DMSO-d_6) 156.0 (C), 152.7 (CH), 149.1 (C), 139.0 (CH), 118.9 (C) 90.6 (CH) 79.0 (CH) 74.1 (CH), 46.6 (CH$_2$), 36.5 (CH$_2$).

RP-HPLC (Method 3 – Section 6.10) t_R = 6.88 min.

HRMS: C$_{10}$H$_{12}$^{35}ClN$_5$O$_2$ [M+H]$^+$ requires *m/z* 270.0752, found *m/z* 270.0756.

9-((2*R*,5*S*)-5-(Chloromethyl)tetrahydrofuran-2-yl)-9*H*-purin-6-amine

3.29
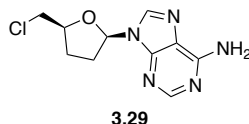
To an ice cooled suspension of ((2*S*,5*R*)-5-(6-Amino-9H-purin-9-yl)tetrahydrofuran-2-yl)methanol (47.0 mg, 200 μmol) in acetonitrile (3.00 mL) was added pyridine (31.0 μL, 385

µmol). Thionyl chloride (73.0 µL, 1.00 mmol) was added dropwise and the reaction was stirred at 0 °C for 3 h. The reaction was allowed to warm to rt and was stirred for a further 16 h. The reaction mixture was concentrated under reduced pressure and purified via RP-HPLC to provide the title compound as a white solid (27.8 mg, 110 µmol, 55%).

¹H NMR: δ (500 MHz, CD₃OD) 8.37 (s, 1 H, 2-*H*), 8.27 (s, 1 H, 8-*H*), 6.36-6.34 (m, 1H, 1′-*H*), 4.47-4.40 (m, 1H, 4′-*H*) 3.84 (dd, *J* = 11.8, 4.8 Hz, 1H, 5′-*H*ₐH_b), 3.77 (dd, *J* = 11.8, 4.8 Hz, 1H, 5′-*H*ₐH_b), 2.68-2.58 (m, 2H, 2′-C*H*₂), 2.34-2.21 (m, 2H, 3′-C*H*₂).

¹³C NMR: δ (126 MHz, CD₃OD) 154.9 (C), 150.7 (CH), 148.7 (C), 139.8 (CH) 119.1 (C) 85.5 (CH), 80.6 (CH), 45.7 (CH₂) 31.4 (CH₂) 27.2 (CH₂).

RP-HPLC (Method 3 – Section 6.10) t_R = 7.41 min.

HRMS: $C_{10}H_{12}^{35}ClN_5O$ [M+H]⁺ requires *m/z* 254.0803, found 254.0807.

2,4-Diacetoxybenzoic acid

4.18
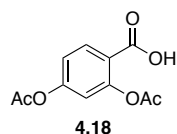
To a solution of 2,4-dihydroxybenzoic acid (3.00 g, 19.4 mmol) in tetrahydrofuran (33.0 mL) was added triethylamine (9.49 mL, 68.0 mmol), acetic anhydride (9.24 mL, 97.1 mmol) and DMAP (237 mg, 1.94 mmol). The reaction mixture was stirred at rt for 4 h after which it was acidified to pH 3 with aqueous HCl (1.00 M) and extracted with EtOAc (3 × 100 mL). The combined organic fractions were washed with brine (100 mL), dried (Na₂SO₄) and concentrated under reduced pressure. The resulting solid was recrystallised from EtOAc to provide the target compound as an off-white powder (4.27 g, 18.0 mmol, 93%).

¹H NMR δ (400 MHz, DMSO-*d₆*) 13.14 (br. s, 1H, COOH) 7.98 (d, *J* = 8.5 Hz, 1H, 6-*H*) 7.18 (dd, *J* = 8.5, 2.5 Hz, 1H, 5-*H*) 7.09 (d, *J* = 2.5 Hz, 1H, 3-*H*) 2.29 (s, 3H, C*H*₃) 2.25 (s, 3H, C*H*₃).

¹³C NMR δ (101 MHz, DMSO-*d₆*) 169.0 (C), 168.6 (C), 165.0 (C), 154.0 (C), 151.0 (C), 132.4 (C), 121.6 (CH), 119.6 (CH), 117.6 (CH), 20.8 (CH₃), 20.8 (CH₃).

ν$_{max}$ (neat): 3300-2543 cm^{-1} (OH stretch).

R$_f$: (1:1 EtOAc/hexane) 0.1.

LCMS: (Method 2 – Section 6.11): t$_R$ = 5.74 min, *m/z* 237 [M+H]$^-$

4-(Chlorocarbonyl)-1,3-phenylene diacetate

4.19
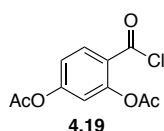
To a suspension of 2,4-diacetoxybenzoic acid (1.93 g, 8.11 mmol) in anhydrous toluene (40.0 mL) under a N$_2$ atmosphere in a flame dried flask at 22 °C was added (COCl)$_2$ (2.08 mL, 46.7 mmol) and anhydrous DMF (cat. amount). The reaction mixture was heated to 40 °C and stirred for 2 h. The reaction mixture was concentrated under reduced pressure and the resulting crude residue was used in the next stage of the synthesis without further purification.

R$_f$ (EtOAc) 0.9.

2-((*tert*-Butoxycarbonyl)amino)-3-ethoxy-3-oxopropanoic acid

4.21
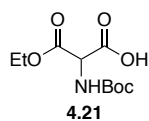
To a solution of diethyl 2-((tert-butoxycarbonyl)amino)malonate (3.60 mL, 14.1 mmol) in ethanol (25.0 mL) was added potassium hydroxide (807 mg, 14.4 mmol). The suspension was stirred at 22 °C for 12 h before concentrating under reduced pressure. The residue was dissolved in NaHCO$_3$ (75.0 mL, 1 M, 75.0 mmol) and washed with EtOAc (2 × 40.0 mL). The solution was cooled to 0 ºC, acidified to pH 2 with solid KHSO$_4$ and extracted with EtOAc (3 × 40.0 mL). The combined organic fractions were washed with brine (20.0 mL) and dried over Na$_2$SO$_4$. The solvent was removed under reduced pressure to provide the title compound as an amorphous colourless solid (2.07 g, 8.36 mmol, 59%).

^{1}H NMR δ (400 MHz, DMSO-*d$_6$*) 13.36 (br. s, 1H, COO*H*) 7.48 (d, *J* = 8.0 Hz, 1H, N*H*) 4.71 (d, *J* = 8.0 Hz, 1H, C*H*) 4.14 (m, 2H, C*H$_2$*CH$_3$) 1.39 (s, 9H, 3 × C*H$_3$*) 1.19 (t, *J* = 7.0 Hz, 3H, CH$_2$C*H$_3$*).

^{13}**C NMR** δ (101 MHz, DMSO-d_6) 167.7 (C), 167.2 (C), 155.1 (C), 78.9 (C), 61.3 (CH$_2$), 57.5 (CH), 28.1 (CH$_3$), 13.9 (CH$_3$).

ν$_{max}$ (neat): 3276 cm^{-1} (NH stretch), 3100-2500 cm^{-1} (OH stretch), 1753 cm^{-1} (C=O stretch), 1654 cm^{-1} (C=O stretch)

R$_f$: (EtOAc) 0.68.

HRMS: C$_{10}$H$_{17}$NO$_6$ [M-H]$^-$ requires *m/z* 246.0983, found 246.0975.

4-(2-((*tert*-Butoxycarbonyl)amino)-3-ethoxy-3-oxopropanoyl)-1,3-phenylene diacetate

4.22
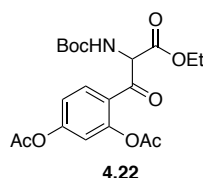
To an ice-cooled solution of 2-(*tert*-butoxycarbonylamino)-3-ethoxy-3-oxopropanoic acid (2.00 g, 8.11 mmol) in anhydrous tetrahydrofuran (20.0 mL) under an N$_2$ atmosphere was added triethylamine (5.64 mL, 40.5 mmol) and anhydrous magnesium chloride (2.31 g, 24.3 mmol). The resulting slurry was stirred vigorously at 4 °C for 2 h. A solution of the previously prepared crude acid chloride (2.06 g, 8.11 mmol) in anhydrous tetrahydrofuran (40.0 mL) was then added and the resulting suspension stirred at 4 °C for 2 h. The reaction mixture was quenched with saturated aqueous ammonium chloride (40.0 mL) and extracted with ethyl acetate (3 × 50.0 mL). The combined organic fractions were washed with brine (10.0 mL) and dried over Na$_2$SO$_4$. The solvent was removed under reduced pressure to provide a brown solid which was used in the next stage of the synthesis without further purification.

tert-Butyl (4,7-dihydroxy-2-oxo-2H-chromen-3-yl)carbamate

4.23
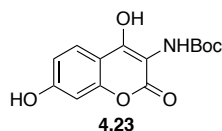
To a solution of crude coupling product (3.32 g theoretical, 8.11 mmol) in methanol (30.0 mL) was added aqueous NaOH (40.6 mL, 1.00 M, 40.6 mmol) and the reaction mixture was stirred at rt for 16 h. The reaction mixture was acidified with HCl (4.00 M) to pH 3 (~11 mL) and extracted with EtOAc (3 × 100 mL). The combined organic fractions were washed with brine

(20.0 mL), dried over MgSO$_4$ and the solvent removed under reduced pressure. Ethyl acetate (40.0 mL) was added to the residue, with a pale orange precipitate remaining. The solid was isolated by filtration to afford the target product (789 mg, 2.69 mmol, 33% over 3 steps).

^{1}H NMR: δ (500 MHz, DMSO-d_6) 11.68 (br. s, 1H, O*H*), 10.48 (br. s, 1H, O*H*), 7.79 (br. s, 1H, N*H*), 7.66 (d, *J* = 8.7 Hz, 1H, 5-*H*), 6.79 (dd, *J* = 8.5, 2.1 Hz, 1H, 6-*H*), 6.68 (d, *J* = 2.1 Hz, 1H, 8-*H*), 1.43 (s, 9H, 3 × C*H*$_3$).

^{13}C NMR: δ (126 MHz, DMSO-d_6) 161.3 (C), 161.2 (C), 160.6 (C), 154.7 (C), 153.4 (C), 124.9 (CH), 112.8 (CH), 107.9 (C), 101.8 (CH), 100.2 (C), 78.6 (C), 28.2 (3 × CH$_3$).

ν$_{max}$ (neat): 3406 cm^{-1} (NH stretch), 3350-3100 cm^{-1} (OH stretch), 1613 cm^{-1} (C=O stretch).

HRMS: C$_{14}$H$_{15}$NO$_6$ [M+H]$^+$ requires *m/z* 294.0972, found 294.0975.

3-Amino-4,7-dihydroxy-2H-chromen-2-one hydrochloride
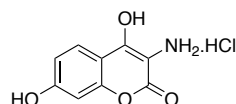
4.10

To an ice cooled solution of the *N*-Boc protected aminocoumarin (920 mg, 3.47 mmol) in *tert*-butylmethyl ether (20.0 mL) and methanol (5.00 mL) was added HCl in dioxane (11.8 mL 4 M, 47.2 mmol) and the mixture stirred at rt for 24 h. The solvent was removed under reduced pressure and the residue triturated with *tert*-butylmethyl ether (10.0 mL). The resulting precipitate was isolated by filtration, to provide the title compound as a dark orange solid (875 mg, 2.98 mmol, 86%).

^{1}H NMR: δ (400 MHz, DMSO-d_6) 7.90 (d, *J* = 8.6 Hz, 1H, 5-*H*), 6.82 (dd, *J* = 8.6, 2.2 Hz, 1H, 6-*H*), 6.72 (d, *J* = 2.2 Hz, 1H, 8-*H*).
Note: OH and NH signals not observed.
^{13}C NMR: δ (101 MHz, DMSO-d_6) 161.6 (C), 161.5 (C), 159.8 (C), 153.5 (C), 125.6 (CH), 112.8 (CH), 109.1 (C), 102.1 (CH), 94.5 (C).

ν$_{max}$ (neat): 3366 cm^{-1} (NH stretch), 3300-2800 cm^{-1} (OH stretch), 1707 cm^{-1} (C=O stretch).

LCMS (Method 2 – Section 6.11): t_R = 1.27 min, *m/z* = 194 [M+H]$^+$

Coumarin structure numbering

N-(4,7-dihydroxy-8-methyl-2-oxo-2H-chromen-3-yl)benzamide

1.139a
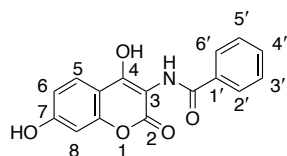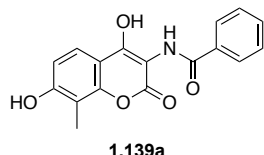
^{1}H NMR δ (600 MHz, DMSO-d_6) 11.71 (br. s, 1H, OH), 10.41 (br. s, 1H, OH), 9.40 (br. s, 1H, NH), 8.01 (d, J = 7.3 Hz, 2H, 2′-H & 6′-H), 7.60-7.48 (m, 4H, 5-H & 3′-H & 4′-H & 5′-H), 6.89 (d, J = 8.4 Hz, 1H, 6-H), 2.18 (s, 3H, CH_3).

RP-HPLC (Method 4 – Section 6.10) t_R = 5.31 min.

HRMS: $C_{17}H_{13}NO_5$ [M-H]$^-$ requires m/z 310.0721, found 310.0546.

N-(8-Ethyl-4,7-dihydroxy-2-oxo-2H-chromen-3-yl)benzamide

1.139b
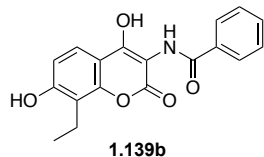
^{1}H NMR δ (600 MHz, DMSO-d_6) 10.39 (br. s, 0.2H, OH), 9.39 (br. S, 1H, NH), 8.01 (d, J = 7.4 Hz, 2H, 2′-H & 6′-H), 7.60-7.48 (m, 4H, 5-H & 3′-H & 4′-H & 5′-H), 6.88 (d, J = 8.4 Hz, 1H, 6-H), 2.73 (q, J = 7.3 Hz, 2H, CH_2), 1.11 (t, J = 7.3 Hz, 3H, CH$_3$).

Note: *One OH signal not observed.*

RP-HPLC (Method 4 – Section 6.10) t_R = 5.49 min.

HRMS: $C_{17}H_{15}NO_5$ [M-H]$^-$ requires m/z 324.0877, found 324.0873.

4,7-Dihydroxy-8-methyl-3-phenyl-2H-chromen-2-one

4.33a
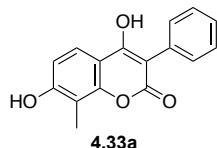
^{1}**H NMR** δ (600 MHz, DMSO-d_6) 10.97 (br. s, 1H, O*H*), 10.39 (br. s, 1H, O*H*), 7.68 (d, *J* = 8.8 Hz, 5-*H*), 7.42-7.28 (m, 5H, Ar-*H*), 6.86 (d, *J* = 8.8 Hz, 6-*H*), 2.17 (s, 3H, CH$_3$).

RP-HPLC (Method 5 – Section 12.5) t_R = 6.75

HRMS: C$_{16}$H$_{12}$O$_4$ [M+H]$^+$ requires *m/z* 269.0808, found 269.0818.

8-Ethyl-4,7-dihydroxy-3-phenyl-2H-chromen-2-one

4.33b
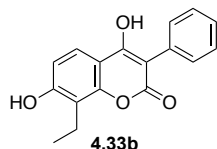
^{1}**H NMR** δ (600 MHz, DMSO-d_6) 7.68 (d, *J* = 8.4 Hz, 5-*H*), 7.44-7.24 (m, 5H, Ar-*H*), 6.82 (d, *J* = 8.4 Hz, 6-*H*) 2.72 (q, *J* = 7.5 Hz, 2H, C*H$_2$*), 1.11 (t, *J* = 7.5 Hz, 3H, C*H$_3$*).

Note: OH peaks not observed.

RP-HPLC (Method 6 – Section 6.10) t_R = 3.73 min.

HRMS: C$_{17}$H$_{14}$O$_4$ [M-H]$^-$ requires *m/z* 324.0877, found 324.0873.

N-(4,7-dihydroxy-2-oxo-2H-chromen-3-yl)-2-methylbenzamide

4.11
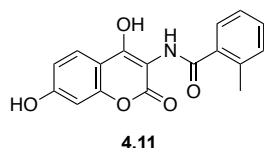
2-methyl benzoic acid (240 mg, 1.76 mmol) was suspended in anhydrous toluene (10.0 mL). Oxalyl chloride (300 μL, 3.52 mmol) and DMF (2 drops) were added with the reaction then stirred at rt for 2 h. The reaction mixture was concentrated under reduced pressure before redissolving in anhydrous DCM (30.0 mL). The crude acid chloride solution was added dropwise to a stirred solution of 3-amino-4,7-dihydroxy-2H-chromen-2-one hydrochloride (100 mg, 436 μmol) and triethylamine (490 μL, 3.52 mmol) in anhydrous DCM (30.0 mL) and left to stir at rt for 16 h. The reaction mixture was concentrated under reduced pressure

before redissolving in methanol (20.0 mL) with addition of NaOH (5.00 mL, 1.00 M, 5.00 mmol). The reaction was stirred at 70 °C for 16 h after which the pH was adjusted to 3 with HCl (1.00 M). The resulting precipitate was isolated by filtration to provide the title compound as a colourless solid (17.5 mg, 56.2 μmol, 13%).

^{1}H NMR δ (400 MHz, DMSO-d_6) 11.87 (br. s, 1H, O*H*), 10.58 (br. s, 1H, O*H*), 9.22 (br. s, 1H, N*H*), 7.74 (d, 8.3 Hz, 1H, 5-*H*), 7.70 (d, *J* = 7.5 Hz, 1H, 6′-*H*), 7.41-7.34 (m, 1H, 4′-*H*), 7.32-7.24 (m, 2H, 3′-*H*, 5′-*H*), 6.84 (dd, *J* = 8.6, 2.3 Hz, 1H, 6-*H*), 6.74 (d, *J* = 2.3 Hz, 1H, 8-*H*), 2.44 (s, 3H, 2′-C*H₃*).

^{13}C NMR: δ (101 MHz, DMSO-d_6) 169.5 (C), 161.4 (C), 160.6 (C), 159.9 (C), 153.4 (C), 136.2 (C), 135.9 (C), 130.4 (CH), 129.7 (CH), 128.0 (CH), 125.2 (CH), 125.0 (CH), 113.0 (CH), 107.9 (C), 101.9 (CH), 100.3 (C), 19.6 (CH₃).

RP-HPLC (Method 7 – Section 6.10) t_R = 5.50 min.

HRMS: C₁₇H₁₃NO₅ [M-H]⁻ requires *m/z* 310.0721, found 310.0708.

N-(4,7-dihydroxy-8-methyl-2-oxo-2H-chromen-3-yl)-2-methylbenzamide

4.34a
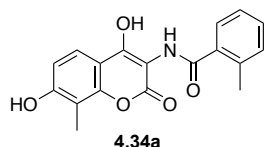
^{1}H NMR δ (600 MHz, DMSO-d_6) 11.78 (br. s, 1H, O*H*), 10.42 (br. s, 1H, O*H*), 9.22 (br. s, 1H, N*H*), 7.71 (d, *J* = 7.5 Hz, 1H, 6′-*H*), 7.61 (d, 8.5 Hz, 1H, 5-*H*), 7.41-7.34 (m, 1H, 4′-*H*), 7.32-7.24 (m, 2H, 3′-*H* & 5′-*H*), 6.89 (d, *J* = 8.5 Hz, 1H, 6-*H*), 2.44 (s, 3H, 2′-C*H₃*), 2.18 (s, 3H, 8-C*H₃*).

RP-HPLC (Method 7 – Section 6.10) t_R = 5.71

HRMS: C₁₈H₁₅NO₅ [M-H]⁻ requires *m/z* 311.0674, found 311.0657.

N-(8-ethyl-4,7-dihydroxy-2-oxo-2H-chromen-3-yl)-2-methylbenzamide

4.34b
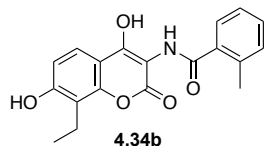
^{1}H NMR δ (600 MHz, DMSO-d_6) 10.40 (br. s, 1H, O*H*), 9.22 (br. s, 1H, N*H*), 7.71 (d, $J = 7.5$ Hz, 1H, 6′-*H*), 7.61 (d, 8.5 Hz, 1H, 5-*H*), 7.41-7.34 (m, 1H, 4′-*H*), 7.32-7.24 (m, 2H, 3′-*H* & 5′-*H*), 6.89 (d, $J = 8.5$ Hz, 1H, 6-*H*), 2.73 (q, $J = 7.5$ Hz, 2H, C*H*₂), 2.44 (s, 3H, 2′-C*H*₃), 1.11 (t, $J = 7.5$ Hz, 3H, C*H*₃). *Note: One OH signal not observed.*

RP-HPLC (Method 4 – Section 6.10) t_R = 5.55 min.

HRMS: $C_{19}H_{17}NO_5$ [M-H]⁻ requires *m/z* 338.1034, found 338.1015.

N-(4,7-dihydroxy-2-oxo-2H-chromen-3-yl)-4-methoxybenzamide

4.12
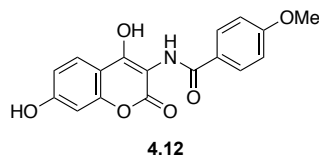
4-methoxy benzoic acid (240 mg, 1.58 mmol) was suspended in anhydrous toluene (10.0 mL). Oxalyl chloride (300 μL, 3.52 mmol) and DMF (2 drops) were added with the reaction then stirred at rt for 2 h. The reaction mixture was concentrated under reduced pressure before redissolving in anhydrous DCM (30.0 mL). The crude acid chloride solution was added dropwise to a stirred solution of 3-amino-4,7-dihydroxy-2H-chromen-2-one hydrochloride (100 mg, 436 μmol) and triethylamine (490 μL, 3.52 mmol) in anhydrous DCM (30.0 mL) and left to stir at rt for 16 h. The reaction mixture was concentrated under reduced pressure before redissolving in methanol (20.0 mL) with addition of NaOH (5.00 mL, 1.00 M, 5.00 mmol). The reaction was stirred for 2 h after which the pH was adjusted to 3 with HCl (1.00 M) and extracted with ethyl acetate (3 × 50.0 mL). Organics were combined and washed with brine (50.0 mL) before concentrating under reduced pressure. Residue was dissolved in water (10.0 mL) and left to stand for 24 h. The resulting precipitate isolated by filtration to provide the title compound as a pale orange solid (11.1 mg, 33.9 μmol, 8%).

^{1}H NMR: δ (500 MHz, DMSO-d_6) 9.88 (br. s, 1H, N*H*), 8.78 (br. s, 1H, O*H*), 7.96 (d, *J* = 7.7 Hz, 2H, 2′-*H*, 6′-*H*), 7.67 (d, *J* = 7.9 Hz, 1H, 5-*H*), 6.97 (d, *J* = 7.6 Hz, 2H, 3′-*H* & 5′-*H*), 6.60 (d, *J* = 7.9 Hz, 1H, 6-*H*), 6.49 (s, 1H, 8-*H*), 3.82 (s, 3H, OC*H*$_3$).

Note: One OH signal not observed.

^{13}C NMR: δ (101 MHz, DMSO-d_6) 165.2 (C), 162.3 (C), 161.2 (C), 159.3 (C), 154.2 (C), 129.5 (2 × CH), 127.5 (C), 126.0 (CH), 115.7 (C), 113.1 (CH), 110.4 (2 × CH), 101.0 (CH), 96.7 (C), 55.2 (CH$_3$).

Note: One C signal not observed.

RP-HPLC (Method 4 – Section 6.10) t_R = 5.31 min.

HRMS: C$_{17}$H$_{13}$NO$_6$ [M-H]$^-$ requires *m/z* 326.0670, found 326.0670.

N-(4,7-dihydroxy-8-methyl-2-oxo-2H-chromen-3-yl)-4-methoxybenzamide

4.35a
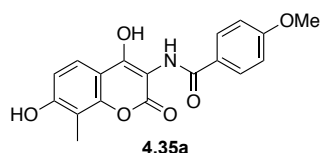
^{1}H NMR: δ (600 MHz, DMSO) 11.73 (br. s, 1H, O*H*), 10.40 (br. s, 1H, O*H*), 9.28 (br.s, 1H, N*H*) 7.99 (d, *J* = 7.7 Hz, 2H, 2′-*H*, 6′-*H*), 7.58 (d, *J* = 8.2 Hz, 1H, 5-*H*), 7.05 (d, *J* = 7.8 Hz, 2H, 3′-*H*, 5′-*H*), 6.89 (d, *J* = 7.8 Hz, 1H, 6-*H*), 3.82 (s, 3H, OC*H*$_3$), 2.18 (s, 3H, 8-C*H*$_3$).

RP-HPLC (Method 4 – Section 6.10) t_R = 5.45 min.

HRMS: C$_{18}$H$_{15}$NO$_6$ [M-H]$^-$ requires *m/z* 340.0827, found 340.0813.

N-(8-ethyl-4,7-dihydroxy-2-oxo-2H-chromen-3-yl)-4-methoxybenzamide

4.35b
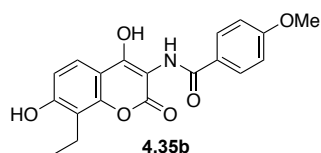
^{1}H NMR: δ (600 MHz, DMSO) 9.26 (br.s, 1H, N*H*) 7.98 (d, *J* = 8.1 Hz, 2H, 2′-*H* & 6′-*H*), 7.57 (d, *J* = 8.4 Hz, 1H, 5-*H*), 7.05 (d, *J* = 8.4 Hz, 2H, 3′-*H* & 5′-*H*), 6.87 (d, *J* = 7.8 Hz, 1H, 6-*H*), 3.82 (s, 3H, C*H*$_3$), 2.72 (q, *J* = 7.4 Hz, 2H, C*H*$_2$) 1.11 (t, *J* = 7.4 Hz, 3H, 8-C*H*$_3$). *Note: 2 × OH signals not observed.*

RP-HPLC (Method 4 – Section 6.10) t_R = 5.57 min.

HRMS: $C_{18}H_{15}NO_6$ [M-H]⁻ requires *m/z* 354.0983, found 354.0972.

N-(4,7-dihydroxy-2-oxo-2H-chromen-3-yl)isonicotinamide

4.13
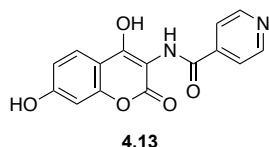
Nicotinic acid (965 mg, 7.84 mmol) was suspended in anhydrous toluene (40.0 mL). Oxalyl chloride (1.32 mL, 15.7 mmol) and DMF (2 drops) were added with the reaction then stirred at rt for 2 h. The reaction mixture was concentrated under reduced pressure before redissolving in anhydrous DCM (30.0 mL). The crude acid chloride solution was added dropwise to a stirred solution of 3-amino-4,7-dihydroxy-2H-chromen-2-one hydrochloride (450 mg, 1.96 mmol) and triethylamine (2.20 mL, 15.7 mmol) in anhydrous DCM (50.0 mL) and left to stir at rt for 16 h. The reaction mixture was acidified with HCl (1.00 M) and concentrated under reduced pressure. The precipitate remaining in the aqueous fraction was isolated by filtration to provide the title compound as a white solid (238 mg, 798 µmol, 28%).

¹H NMR δ (400 MHz, DMSO-d_6) 11.91 (br. s, 1H, O*H*), 10.59 (br. s, 1H, O*H*), 9.71 (br. s, 1H, N*H*), 8.79 (d, *J* = 5.7 Hz, 2H, 3′-*H* & 5′-*H*), 7.89 (d, *J* = 5.7 Hz, 2H, 2′-*H* & 6′-*H*), 7.74 (d, *J* = 8.6 Hz, 1H, 5-*H*), 6.85 (dd, *J* = 8.6, 2.3 Hz, 1H, 6-*H*), 6.75 (d, *J* = 2.3 Hz, 1H, 8-*H*).

¹³C NMR: δ (101 MHz, DMSO-d_6) 165.0 (C), 161.6 (C), 160.8 (C), 160.6 (C), 153.6 (C), 150.1 (2 × CH), 141.0 (C), 125.1 (2 × CH), 121.8 (CH), 113.0 (CH), 107.8 (C), 101.9 (CH), 99.2 (C).

RP-HPLC (Method 4 – Section 6.10) t_R = 2.98 min.

HRMS: $C_{15}H_{10}N_2O_5$ [M+H]⁺ requires *m/z* 299.0662, found 299.0682.

N-(4,7-dihydroxy-8-methyl-2-oxo-2H-chromen-3-yl)isonicotinamide

4.36a
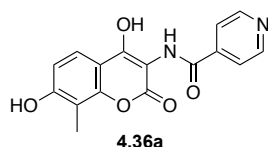
^{1}H NMR δ (600 MHz, DMSO-d_6) 11.79 (br.s, 1H, O*H*), 10.46 (br.s, 1H, O*H*), 9.71 (br.s, 1H, N*H*), 8.82-8.71 (m, 2H, 3′-*H* & 5′-*H*), 7.90 (d, *J* = 5.1 Hz, 2H, 2′-*H* & 6′-*H*), 7.61 (d, *J* = 8.4 Hz, 1H, 5-*H*), 6.91(d, *J* = 8.4 Hz, 1H, 6-*H*), 2.18 (s, 3H, C*H*$_3$).

RP-HPLC (Method 4 – Section 6.10) t_R = 3.16 min.

HRMS: C$_{16}$H$_{12}$N$_2$O$_5$ [M-H]$^-$ requires *m/z* 311.0674, found 311.0657.

N-(8-ethyl-4,7-dihydroxy-2-oxo-2H-chromen-3-yl)isonicotinamide

4.36b
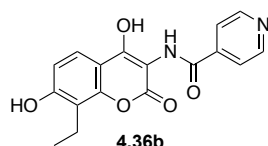
^{1}H NMR δ (600 MHz, DMSO-d_6) 10.45 (br.s, 1H, O*H*), 9.71 (br.s, 1H, N*H*), 8.82-8.76 (m, 2H, 3′-*H* & 5′-*H*), 7.90 (d, *J* = 4.7 Hz, 2H, 2′-*H* & 6′-*H*), 7.60 (d, *J* = 8.6 Hz, 1H, 5-*H*), 6.91 (d, *J* = 8.6 Hz, 1H, 6-*H*), 2.73 (q, J = 7.5 Hz, 2H, CH$_2$), 1.12 (t, *J* = 7.5 Hz, 3H, C*H*$_3$). *Note: One OH signal not observed.*

RP-HPLC (Method 4 – Section 6.10) t_R = 3.44 min.

HRMS: C$_{17}$H$_{14}$N$_2$O$_5$ [M-H]$^-$ requires *m/z* 325.0830, found 325.0823.

N-(4,7-dihydroxy-2-oxo-2H-chromen-3-yl)-3-fluorobenzamide

4.14
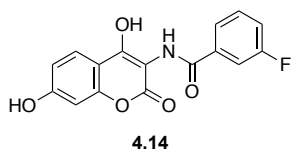
3-fluoro benzoic acid (1.01 g, 7.84 mmol) was suspended in anhydrous DCM (40.0 mL). Oxalyl chloride (1.32 mL, 15.7 mmol) and DMF (2 drops) were added with the reaction then stirred at rt for 2 h. The reaction mixture was concentrated under reduced pressure before redissolving in anhydrous DCM (30.0 mL). The crude acid chloride solution was added dropwise to a stirred solution of 3-amino-4,7-dihydroxy-2H-chromen-2-one hydrochloride

(450 mg, 1.96 mmol) and triethylamine (2.20 mL, 15.7 mmol) in anhydrous DCM (50.0 mL) and left to stir at rt for 16 h. The reaction mixture was concentrated under reduced pressure before redissolving in methanol (50.0 mL) with addition of NaOH (30.0 mL, 1.00 M, 30.0 mmol). The reaction was heated to 70 °C and stirred for 16 h after which the pH was adjusted to 5 with HCl (1.00 M) resulting in formation of a precipitate. The solid was isolated by filtration to provide the title compound as a white solid (171 mg, 542 µmol, 28%).

¹H NMR δ (400 MHz, DMSO-d_6) 11.83 (br. s, 1H, O*H*), 10.59 (br. s, 1H, O*H*), 9.51 (br. s, 1H, N*H*), 7.84 (d, J = 8.9 Hz, 1H, 6′-*H*), 7.79 (d, J = 8.9 Hz, 1H, 2′-*H*), 7.73 (d, J = 8.8 Hz, 1H, 5-*H*), 7.62-7.53 (m, 1H, 4′-*H*), 7.48-7.40 (m, 1H, 3′-*H*), 6.84 (dd, J = 8.8, 2.1 Hz, 1H, 6-*H*), 6.74 (d, J = 2.1 Hz, 1H, 8-*H*).

¹³C NMR: δ (101 MHz, DMSO-d_6) 165.2 (C), 163.0 (C), 161.5 (C), 160.7 (C), 160.6 (C), 153.5 (C), 136.4 (d, J = 7.0 Hz), 130.3 (d, J = 7.2 Hz), 125.0 (CH), 124.1 (CH), 118.3 (d, J = 21.0 Hz), 114.7 (d, J = 22.1 Hz, CH), 113.0 (CH), 107.8 (C), 101.9 (CH), 99.6 (C).

RP-HPLC (Method 4 – Section 6.10) t_R = 5.08 min.

HRMS: $C_{16}H_{10}FNO_5$ [M-H]⁻ requires *m/z* 314.0470, found 314.0470.

N-(4,7-dihydroxy-8-methyl-2-oxo-2H-chromen-3-yl)-3-fluorobenzamide

4.37a
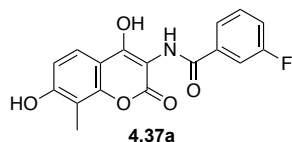
¹H NMR δ (600 MHz, DMSO-d_6) 11.71 (br. s, 1H, OH), 10.44 (br. s, 1H, OH), 9.50 (br. s, 1H, NH), 7.85 (d, J = 7.7 Hz, 1H, 6′-H), 7.80 (d, J = 10.2 Hz, 1H, 2′-*H*), 7.59 (d, J = 8.8 Hz, 1H, 5-*H*), 7.58-7.55 (m, 1H, 4′-*H*), 7.48-7.42 (m, 1H, 3′-*H*), 6.88 (d, J = 8.8 Hz, 1H, 6-*H*), 2.18 (s, 3H, CH₃).

RP-HPLC (Method 4 – Section 6.10) t_R = 5.28 min.

HRMS: $C_{17}H_{12}FNO_5$ [M-H]⁻ requires *m/z* 328.0627, found 328.0612.

N-(8-ethyl-4,7-dihydroxy-2-oxo-2H-chromen-3-yl)-3-fluorobenzamide

4.37b
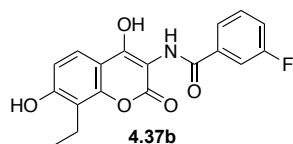
^{1}H NMR δ (600 MHz, DMSO-d_6) 9.50 (br. s, 1H, NH), 7.84 (d, J = 7.7 Hz, 1H, 6′-H), 7.79 (d, J = 9.2 Hz, 1H, 2′-H), 7.59 (d, J = 8.4 Hz, 1H, 5-H), 7.58-7.55 (m, 1H, 4′-H), 7.48-7.42 (m, 1H, 3′-H), 6.88 (d, J = 8.4 Hz, 1H, 6-H), 2.72 (q, J = 7.5 Hz, 2H, CH$_2$), 1.11 (t, J = 7.5 Hz, 3H, CH$_3$). *Note: Two OH signals not observed.*

RP-HPLC (Method 4 – Section 6.10) t_R = 5.46 min.

HRMS: $C_{18}H_{14}FNO_5$ [M-H]$^-$ requires *m/z* 342.0783, found 342.0768.

7-Hydroxy-3-(4-(5-hydroxypentyl)-1H-1,2,3-triazol-1-yl)-2H-chromen-2-one

4.27
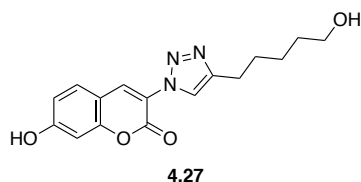
A suspension of CuSO$_4$ (8.94 mg, 56.0 μmol), coumarin azide (100 mg, 492 μmol) and 6-heptyn-1-ol (60.7 mg, 542 μmol) in methanol (10.0 mL) was prepared. Sodium ascorbate (488 mg, 2.46 mmol) was dissolved in water (5.00 mL) and added dropwise with this mixture stirred at rt for 6 h. The reaction mixture was acidified to pH 3 with HCl (1.00 M) and was extracted with ethyl acetate (3 × 50.0 mL). Organic fractions were combined, dried over Na$_2$SO$_4$ and concentrated under reduced pressure to provide the title compound as a brown solid (116 mg, 368 μmol, 76%).

^{1}H NMR δ (400 MHz, CD$_3$OD) 8.47 (s, 1H, 4-H), 8.32 (s, 1H, N-CH=C), 7.64 (d, J = 8.4 Hz, 1H, 5-H), 6.90 (dd, J = 8.4, 2.3 Hz, 1H, 6-H), 6.82 (d, J = 2.3 Hz, 1H, 8-H), 3.57 (t, J = 6.4 Hz, 2H, CH$_2$-OH) 2.80 (t, J = 7.5 Hz, 2H, CH=C-CH$_2$), 1.76 (quin, J = 7.5 Hz, CH$_2$), 1.64-1.55 (m, 2H, CH$_2$), 1.51-1.42 (m, 2H, CH$_2$).

^{13}C NMR: δ (101 MHz, CD$_3$OD) 162.4 (C), 156.3 (C), 154.6 (C), 147.2 (C), 135.2 (CH), 129.9 (CH), 122.2 (CH), 119.1 (C), 113.6 (CH), 110.1 (C), 101.5 (CH), 60.9 (CH$_2$), 31.4 (CH$_2$), 28.4 (CH$_2$), 24.5 (CH$_2$), 24.3 (CH$_2$).

ν_{max} **(neat):** 3300-2700 cm^{-1} (OH stretch), 2926 cm^{-1} (CH stretch), 1716 cm^{-1} (C=O stretch).

RP-HPLC (Method 4 – Section 6.10) t_R = 4.12 min.

HRMS: C$_{16}$H$_{17}$N$_3$O$_4$ [M-H]$^-$ requires *m/z* 314.1146, found 314.1153.

(2-Azidoethyl)benzene[161]

4.32
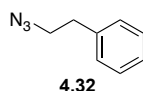
NaN$_3$ (1.05 g, 16.2 mmol) was added to a solution of phenylethyl bromide (2.50 g, 13.5 mmol) in DMF (30.0 mL) with the reaction mixture heated to 80 °C and stirred for 16 h. The reaction mixture was washed with water (20.0 mL) and was extracted with diethyl ether (2 × 50.0 mL). The organic layers were washed with brine (20.0 mL) and dried over anhydrous Na$_2$SO$_4$. The organic phase was concentrated under reduced pressure to provide the title compound as a yellow oil which was used without further purification.

1-Phenethyl-1H-1,2,3-triazole-4-carboxylic acid (procedure adapted from Zhao *et al.*)[161]

4.28
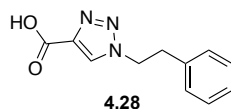
A solution of CuSO$_4$ (108 mg, 677 µmol) and sodium ascorbate (267 mg, 1.35 mmol) in water (10.0 mL) was prepared. (2-azido)ethylbenzene (1.99 g, 13.5 mmol), *t*-BuOH (10.0 mL) and propiolic acid (831 µL, 13.5 mmol) were added in respective order and the reaction mixture was stirred at rt for 16 h. The reaction mixture was acidified to pH 3 with HCl (1.00 M) and was extracted with ethyl acetate (3 × 50.0 mL). Organic fractions were combined, dried over Na$_2$SO$_4$ and concentrated under reduced pressure to provide the title compound as an off-white solid (2.83 g, 13.0 mmol, 96%)

^{1}H NMR δ (400 MHz, DMSO-d_6) 8.58 (s, 1H, =C*H*-N) 7.32-7.25 (m, 2H, 2 × ArC*H*) 7.25-7.16 (m, 3H, 3 × ArC*H*) 4.68 (t, *J* = 6.5 Hz, 2H, NC*H$_2$*CH$_2$) 3.20 (t, J = 6.5 Hz, 2H, NCH$_2$C*H$_2$*). *Note: COOH not observed.*

^{13}C NMR (101 MHz, DMSO-d_6) 161.7 (C), 139.5 (C), 137.4 (C), 128.7 (2 × CH), 128.4 (3 × CH), 126.6 (CH), 50.7 (CH$_2$), 35.4 (CH$_2$).

ν_max (neat): 3300-2500 cm^{-1} (OH stretch), 1685 cm^{-1} (C=O stretch).

LCMS (Method 2 – Section 13): t_R = 5.82 min, *m/z* = 218 [M+H]$^+$

N-(4,7-dihydroxy-2-oxo-2H-chromen-3-yl)-1-phenethyl-1*H*-1,2,3-triazole-4-carboxamide

4.16
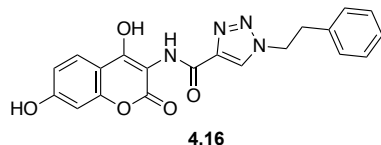
1-phenethyl-1H-1,2,3-triazole-4-carboxylic acid (851 mg, 3.92 mmol) was suspended in anhydrous DCM (20.0 mL). Oxalyl chloride (660 μL, 7.84 mmol) and DMF (2 drops) were added with the reaction then stirred at rt for 2 h. The reaction mixture was concentrated under reduced pressure before redissolving in anhydrous DCM (30.0 mL). The crude acid chloride solution was added dropwise to a stirred solution of 3-amino-4,7-dihydroxy-2*H*-chromen-2-one hydrochloride (450 mg, 1.96 mmol) and triethylamine (2.20 mL, 15.7 mmol) in anhydrous DCM (50.0 mL) and was stirred at rt for 16 h. The reaction mixture was concentrated under reduced pressure before redissolving in methanol (20.0 mL) with addition of NaOH (10.0 mL, 2.00 M, 20.0 mmol). The reaction was stirred at 80 °C for 1 h after which the pH was adjusted to 3 with HCl (1.00 M). The resulting precipitate was isolated by filtration to provide the title compound as a white solid (46.6 mg, 119 μmol, 6%).

¹H NMR δ (400 MHz, DMSO-d_6) 12.01 (br. s, 1H, O*H*) 10.57 (br. s, 1H, O*H*) 9.29 (br. s, 1H, N*H*), 8.65 (s, 1H, =C*H*-N), 7.70 (d, *J* = 8.6 Hz, 1H, 5-*H*), 7.33-7.18 (m, 5H, 5 × ArC*H*), 6.83 (dd, *J* = 8,8, 2.2 Hz, 1H, 6-*H*), 6.73 (d, *J* = 2.2 Hz, 1H, 8-*H*), 4.73 (t, *J* = 7.0 Hz, 2H, N-C*H₂*CH₂), 3.24 (t, *J* = 7.0 Hz, 2H, N-CH₂C*H₂*).

¹³C NMR: δ (101 MHz, DMSO-d_6) 161.4 (C), 160.5 (C), 159.8 (C), 159.4 (C), 153.3 (C), 142.0 (C), 137.4 (C), 128.7 (2 × CH), 128.4 (2 × CH), 127.0 (CH), 126.6 (CH), 125.0 (CH), 113.0 (CH), 107.9 (C), 101.9 (CH), 99.7 (C), 50.7 (CH₂), 35.4 (CH₂).

RP-HPLC (Method 7 – Section 6.10) t_R = 5.65 min.

HRMS: C$_{20}$H$_{16}$N$_4$O$_5$ [M-H]$^-$ requires *m/z* 391.1048, found 391.1049.

N-(4,7-dihydroxy-8-methyl-2-oxo-2H-chromen-3-yl)-1-phenethyl-1*H*-1,2,3-triazole-4-carboxamide

4.39a
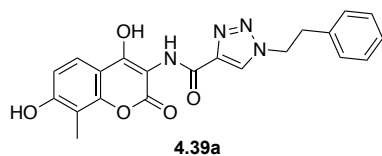
2-Cl-ClDA (32.7 mg, 102 µmol) was dissolved in DMSO (1.02 mL), L-methionine (76.0 mg, 0.509 µmol) was dissolved in aqueous sodium hydroxide (1.02 mL, 250 mM, 255 µmol) and the two solutions mixed and made up to 15 mL with aqueous potassium phosphate (100 mM, pH 6.8). A solution of N-(4,7-dihydroxy-2-oxo-2H-chromen-3-yl)-1-phenethyl-1*H*-1,2,3-triazole-4-carboxamide (19.6 mg, 0.05 mmol) was dissolved in DMSO (1.02 mL) and added to the reaction mixture. BSA (15.0 mg) and DTT (150 µL, 100 mM, 15.0 µmol) were added prior to addition of SalL (final concentration 2.10 µM). Reaction mixture was incubated at 37 ºC for 24 h after which NovO and MTAN were added (final concentrations 9.38 µM, 132 nM respectively). After a further incubation at 37 ºC for 24 h solvent was removed by lyophilisation with the resulting precipitate dissolved in 20% DMSO in water and purified by RP-HPLC to provide the title compound as a white solid (4.70 mg, 11.6 µmol, 23%).

^{1}H NMR δ (400 MHz, DMSO-d_6) 11.96 (br. s, 1H, O*H*) 10.44 (br. s, 1H, O*H*) 9.30 (br. s, 1H, N*H*), 8.66 (s, 1H, =C*H*-N), 7.57 (d, *J* = 8.4 Hz, 1H, 5-*H*), 7.33-7.18 (m, 5H, 5 × ArC*H*), 6.89 (d, *J* = 8.4 Hz, 1H, 6-*H*), 4.73 (t, *J* = 7.0 Hz, 2H, N-C*H$_2$*CH$_2$), 3.24 (t, *J* = 7.0 Hz, 2H, N-CH$_2$C*H$_2$*) 2.17 (s, 3H, 8-C*H$_3$*).

^{13}C NMR: δ (101 MHz, DMSO) 160.6 (C), 159.8 (C), 159.5 (C), 159.1 (C), 151.2 (C), 142.0 (C), 137.3 (C), 128.7 (2 × CH), 128.4 (2 × CH), 127.1 (CH), 126.6 (CH), 121.5 (CH), 111.9 (CH), 110.4 (C), 107.9 (C), 99.4 (C), 50.7 (CH$_2$), 35.4 (CH$_2$), 8.0 (CH$_3$).

RP-HPLC (Method 7 – Section 6.10) t_R = 5.88 min.

HRMS: $C_{21}H_{18}N_4O_5$ [M-H]$^-$ requires *m/z* 405.1204, found 405.1187.

N-(8-ethyl-4,7-dihydroxy-2-oxo-2H-chromen-3-yl)-1-phenethyl-1H-1,2,3-triazole-4-carboxamide

4.39b
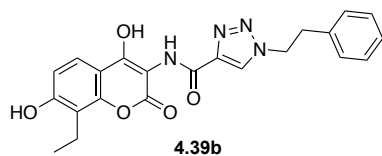
^{1}H NMR δ (600 MHz, DMSO-d_6) 9.26 (br.s, 0.18H (exchangeable proton), N*H*), 8.62 (s, 0.3H, =C*H*-N), 7.56 (d, *J* = 8.6 Hz, 1H, 5-*H*), 7.33-7.05 (m, 5H, 5 × ArC*H*), 6.87 (d, *J* = 8.6 Hz, 1H, 6-H), 4.73 (t, *J* = 6.8 Hz, 2H, N-C*H₂*CH₂), 3.24 (t, *J* = 6.8 Hz, 2H, N-CH₂C*H₂*) 2.71 (q, *J* = 7.5 Hz, 2H, C*H₂*), 1.10 (t, *J* = 7.5 Hz, 3H, CH₃). *Note: Two OH signals note observed.*

RP-HPLC (Method 4 – Section 6.10) t_R = 5.57 min.

HRMS: $C_{22}H_{20}N_4O_5$ [M-H]$^-$ requires *m/z* 419.1361, found 419.1360.

4,7-Dihydroxy-8-methyl-3-(3-oxo-1-phenylbutyl)-2H-chromen-2-one

4.40a
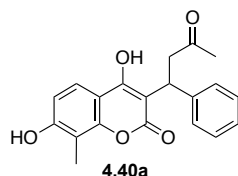
^{1}H NMR δ (600 MHz, DMSO-d_6) 7.51 (d, *J* = 8.4 Hz, 1H, 5-*H*), 7.29-7.09 (m, 5H, 5 × ArC*H*), 6.86 (d, *J* = 8.4 Hz, 1H, 6-*H*), 3.97-3.90 (m, 1H, C*H*), 2.33-2.23 (m, 1H, C*Ha*H$_b$C=O), 2.12 (s, 3H, 8-CH₃) 1.89-1.81 (m, 1H, CHa*Hb*C=O), 1.59 (s, 3H, O=CC*H₃*).

RP-HPLC (Method 7 – Section 6.10) t_R = 4.07 min.

HRMS: $C_{20}H_{18}O_5$ [M-H]$^-$ requires *m/z* 337.1082, found 337.1067.

8-Ethyl-4,7-dihydroxy-3-(3-oxo-1-phenylbutyl)-2H-chromen-2-one

4.40b
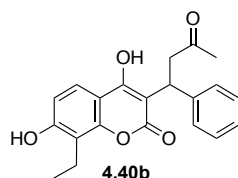
^{1}H NMR δ (600 MHz, DMSO-d_6) 7.51 (d, *J* = 8.4 Hz, 1H, 5-*H*), 7.27-7.09 (m, 5H, 5 × ArC*H*), 6.86 (d, *J* = 8.4 Hz, 1H, 6-*H*), 3.97-3.90 (m, 1H, C*H*), 2.68 (q, *J* = 7.5 Hz, 2H, C*H₂*) 2.33-2.23

(m, 1H, CH_aH$_b$C=O), 1.89-1.81 (m, 1H, CH$_a$$H_b$C=O), 1.59 (s, 3H, O=CC$H_3$), 1.08 (t, J = 7.5 Hz, 3H, 8-CH_3).

RP-HPLC (Method 7 – Section 6.10) t_R = 4.35 min.

HRMS: $C_{21}H_{20}O_5$ [M-H]$^-$ requires *m/z* 351.1238, found 351.1233.

6.3 Cloning and Expression

6.3.1 SalL

The coding sequence for SalL (GenBank: ABP73643.1) with an N-terminal 6-His tag, was synthesised by GenScript (with codon optimization for *Escherichia coli*) and subcloned into a pET26b(+) plasmid using engineered NdeI and XhoI restriction sites. The plasmids were received from GenScript in the lyophilised form. Plasmid DNA was resuspended in 40 µL nuclease-free water. *E. coli* TOP10 competent cells were transformed using the heat shock method with 5 µL of DNA solution. Transformants were plated on LB agar containing 50 µg/ mL kanamycin and grown at 37 °C overnight. Colonies were inoculated into 10 mL LB media containing 50 µg/ mL kanamycin and grown overnight at 37 °C. The DNA was then extracted and purified using a Qiagen Miniprep kit according to the manufacturer's instructions, eluting in 100 µL elution buffer.

6.3.1.1 Expression of SalL

E. coli. BL21 (DE3) competent cells (Invitrogen) were transformed with pET26b(+)-SalL plasmid. Transformants harbouring the plasmids were grown at 37 °C in LB medium supplemented with 50 μg/ mL kanamycin. This seed culture was used to inoculate Overnight Express™ supplemented with kanamycin (final concentration of 50 μg/ mL) using 2% inoculum. The cultures were incubated at 37 °C, 200 rpm and grown to an OD of ~2 before incubating at 18 °C, 200 rpm overnight. The cells were harvested by centrifugation (4000 g for 20 minutes at 4 °C) and resuspended in potassium phosphate buffer (100 mM, pH 6.8) at a volume of ~10 mL/ g of 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 before purification was carried out.

6.3.1.2   Purification of SalL

Binding Buffer (Buffer A): 50 mM potassium phosphate, 20 mM imidazole, 500 mM Na_2SO_4, pH 7.9

Binding Buffer (Buffer B): 50 mM potassium phosphate, 500 mM imidazole, 500 mM Na_2SO_4, pH 7.9

The cell pellet was resuspended in 10 mL of binding buffer per g of cells and the cells lysed by sonication on ice using a 16 mm probe for 10 minutes (1 second on, 1 second off) at 40% intensity. The resulting cell lysate was separated by centrifugation (4000 rpm, 30 mins, 4 ^{0}C) and the supernatant collected. A 5 mL HisTRAP HP column (GE Healthcare 17-5248-02) was fitted on an Akta Start protein purification system and washed with 10 column volumes (CV) of binding buffer before the supernatant of the cell lysate was loaded on to the column. The flow-through was collected as the non-absorbed fraction (NAF). The column was then washed with binding buffer for 10 CV until the UV absorbance was stable before being eluted with a gradient of imidazole (Buffer B 0-100%) over 20 CV.

The fractions were analysed by SDS-PAGE. Appropriate fractions applied to a 10,000 MW Millipore Centrifugal filter unit to be concentrated, then loaded onto a PD-10 desalting column (GE Healthcare 17-0851-01) and washed through by with phosphate buffer buffer (20 mM potassium phosphate, pH 6.8, 10% glycerol). Glycerol (1.5 mL) was added to the resulting eluent (3.5 mL) and samples were divided into 300 µL aliquots. Concentrations were determined by using an E1% value of 8.19. The identity of the protein band was confirmed by MALDI-TOF-MS.

6.3.1.3   General procedure for the purification of SalL by Size Exclusion Chromatography (SEC)

SalL for crystallisation studies was additionally purified by SEC using the following protocol (SEC buffer: 25 mM HEPES, 100 mM NaCl, 5 mM mercaptoethanol)

A Sephadex200 Increase 10/300gl SEC column was equilibrated with SEC buffer and loaded with 100 µL aliquots of the HisTrap concentrate. The column was eluted with SEC buffer with 1 mL fractions collected and analysed by SDS-PAGE. Fractions containing SalL were concentrated using a 10,000 MW Millipore Centrifugal filter unit and were stored at 4 °C for crystallisation trials.
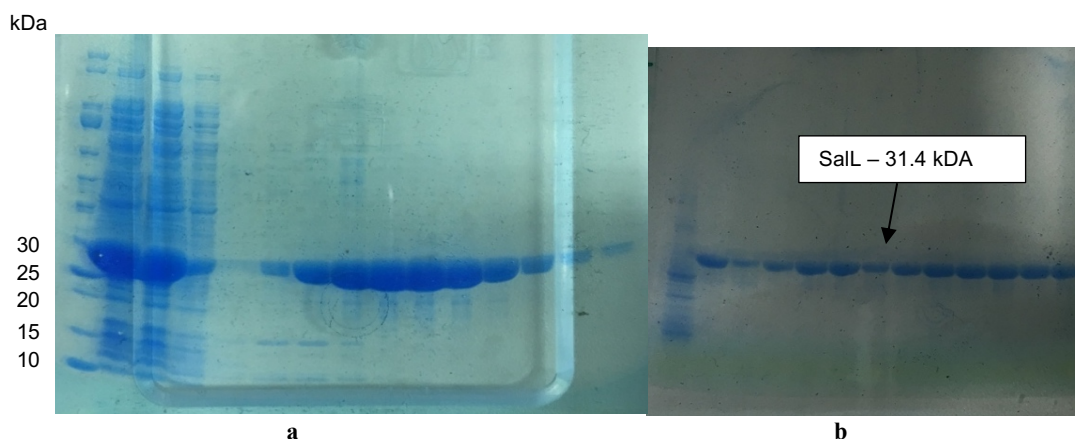
Figure S 1a. Affinity chromatography SDS PAGE gel – lane order, Marker, Load, flow through, wash, pure SalL fractions. **S4b** – SalL bands after SEC.

6.3.2 NovO

The coding sequence for NovO (Genbank: JN606326.1) with a C-terminal 6-His tag, was synthesised by GenScript (with codon optimization for *Escherichia coli*) and subcloned into a pET26b(+) plasmid using engineered NdeI and XhoI restriction sites. The plasmids were received from GenScript in the lyophilised form. Plasmid DNA 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 LB agar containing 50 ug/ mL kanamycin and grown at 37 °C overnight. Colonies were inoculated into 10 mL LB media containing 50 µg/ mL kanamycin and grown overnight at 37 °C. The DNA was then extracted and purified using a Qiagen Miniprep kit according to the manufacturer's instructions, eluting in 100 µL elution buffer.

6.3.2.1 Expression of NovO

E. coli. BL21 (DE3) competent cells (Invitrogen) were transformed with pET26b(+)-NovO plasmid. Transformants harbouring the plasmids were grown at 37 °C in LB medium supplemented with 50 µg/ mL kanamycin. This seed culture was used to inoculate Magic Media (ThermoFisher) supplemented with kanamycin (final concentration of 50 µg/ mL) and Component B (5% v/v) using 2% inoculant. The cultures were incubated at 30 °C, 200 rpm and grown to an OD of ~2 before the incubating at 18 °C, 200 rpm overnight. The cells were harvested by centrifugation (4400 rpm, 4 °C, 20 minutes), the supernatant discarded and the cell pellets stored at - 80 °C for further use.

6.3.2.2   Purification of NovO

Binding Buffer (Buffer A): 50 mM Tris-HCl, 300 mM NaCl, pH 8.0
Elution Buffer (Buffer B): 50 mM Tris-HCl, 300 mM NaCl, 500 mM imidazole, pH 8.0

The cell pellet was resuspended in 10 mL of binding buffer per gram of cells and the cells lysed by sonication on ice using a 16 mm probe for 10 minutes (1 second on, 1 second off) at 40% intensity. The resulting cell lysate was separated by centrifugation (4000 rpm, 30 mins, 4 ^{0}C) and the supernatant collected. A 5 mL HisTRAP HP column (GE Healthcare 17-5248-02) was fitted on an Akta Start protein purification system and washed with 10 column volumes (CV) of binding buffer before the supernatant of the cell lysate was loaded on to the column. The flow-through was collected as the non-absorbed fraction (NAF). The column was then washed with binding buffer for 10 CV until the UV absorbance was stable before being eluted with a gradient of imidazole (Buffer B 0-100%) over 20 CV.

The fractions were analysed by SDS-PAGE. Appropriate fractions applied to a 10,000 MW Millipore Centrifugal filter unit to be concentrated, then loaded onto a PD-10 desalting column (GE Healthcare 17-0851-01) and washed through by with phosphate buffer (20 mM potassium phosphate, pH 6.8, 10% glycerol). Glycerol (1.50 mL) was added to the resulting eluent (3.50 mL) and samples were divided into 300 µL aliquots. Concentrations were determined by using an E1% value of 11.3. The identity of the protein band was confirmed by MALDI-TOF-MS.
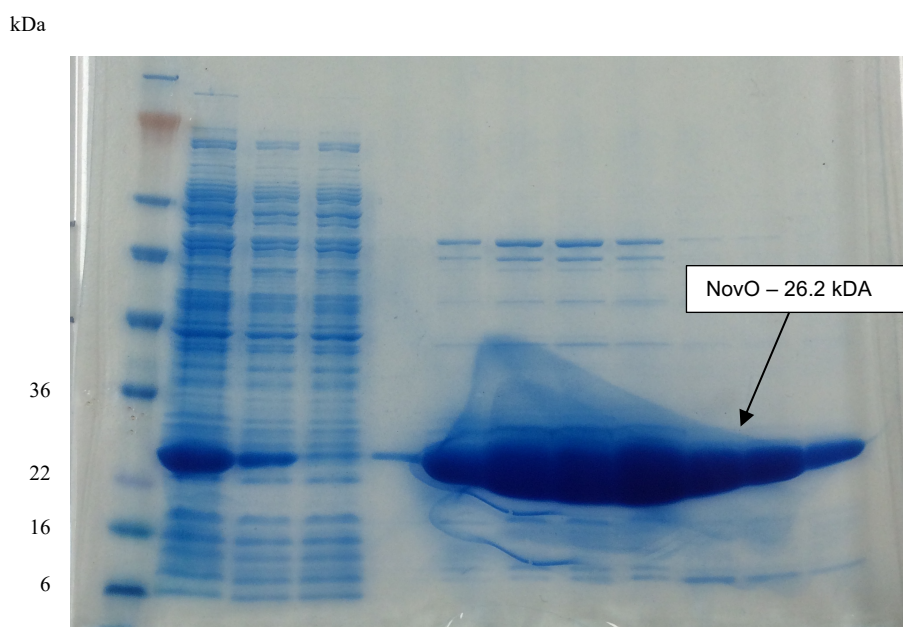
Figure S 2. NovO affinity purification gel Marker, load, non-absorbed fraction, wash, NovO fractions.

6.3.3 MTAN

Salmonella MTAN was a gift from Vern Schramm (Addgene plasmid # 64041 ; http://n2t.net/addgene:64041 ; RRID:Addgene_64041) and subcloned into a pDEST14 vector using engineered BgIII and NheI restriction sites. The plasmid was received from Addgene as a bacterial stab in agar. Cells from this stab were spread on LB agar containing 100 μg/mL ampicillin and incubated at 37 °C overnight. Colonies were inoculated into 10 mL LB media containing 100 μg/ mL kanamycin and grown overnight at 37 °C. The DNA was then extracted and purified using a Qiagen Miniprep kit according to the manufacturer's instructions, eluting in 100 μL elution buffer.

6.3.3.1 Expression of MTAN

E. coli. BL21 (DE3) competent cells (Invitrogen) were transformed with pDEST14-MTAN plasmid. Transformants harbouring the plasmids were grown at 37 °C in LB medium supplemented with 100 μg/ mL ampicillin. This seed culture was used to inoculate Overnight Express™ supplemented with ampicillin (final concentration of 100 μg/ mL) using 2% inoculant. The cultures were incubated at 37 °C, 200 rpm and grown to an OD of ~2 before incubating at 18 °C, 200 rpm overnight. The cells were harvested by centrifugation (4000 g for 20 minutes at 4 °C) and resuspended in potassium phosphate buffer (100 mM, pH 6.8) at a volume of ~10 mL/ g of 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 before purification was carried out.

6.3.3.2 Purification of MTAN

Binding Buffer (Buffer A): 50 mM potassium phosphate, 20 mM imidazole, 500 mM Na_2SO_4, pH 7.9

Binding Buffer (Buffer B): 50 mM potassium phosphate, 500 mM imidazole, 500 mM Na_2SO_4, pH 7.9

The cell pellet was resuspended in 10 mL of binding buffer per g of cells and the cells lysed by sonication on ice using a 16 mm probe for 10 minutes (1 second on, 1 second off) at 40% intensity. The resulting cell lysate was separated by centrifugation (4000 rpm, 30 mins, 4 °C) and the supernatant collected. A 5 mL HisTRAP HP column (GE Healthcare 17-5248-02) was fitted on an Akta Start protein purification system and washed with 10 column volumes (CV) of binding buffer before the supernatant of the cell lysate was loaded on to the column. The

flow-through was collected as the non-absorbed fraction (NAF). The column was then washed with binding buffer for 10 CV until the UV absorbance was stable before being eluted with a gradient of imidazole (Buffer B 0-100%) over 20 CV.

The fractions were analysed by SDS-PAGE. Appropriate fractions applied to a 10,000 MW Millipore Centrifugal filter unit to be concentrated, then loaded onto a PD-10 desalting column (GE Healthcare 17-0851-01) and washed through by with phosphate buffer (20 mM potassium phosphate, pH 6.8, 10% glycerol). Glycerol (1.50 mL) was added to the resulting eluent (3.50 mL) and samples were divided into 300 µL aliquots. Concentrations were determined by using an E1% value of 2.34. The identity of the protein band was confirmed by MALDI-TOF-MS.
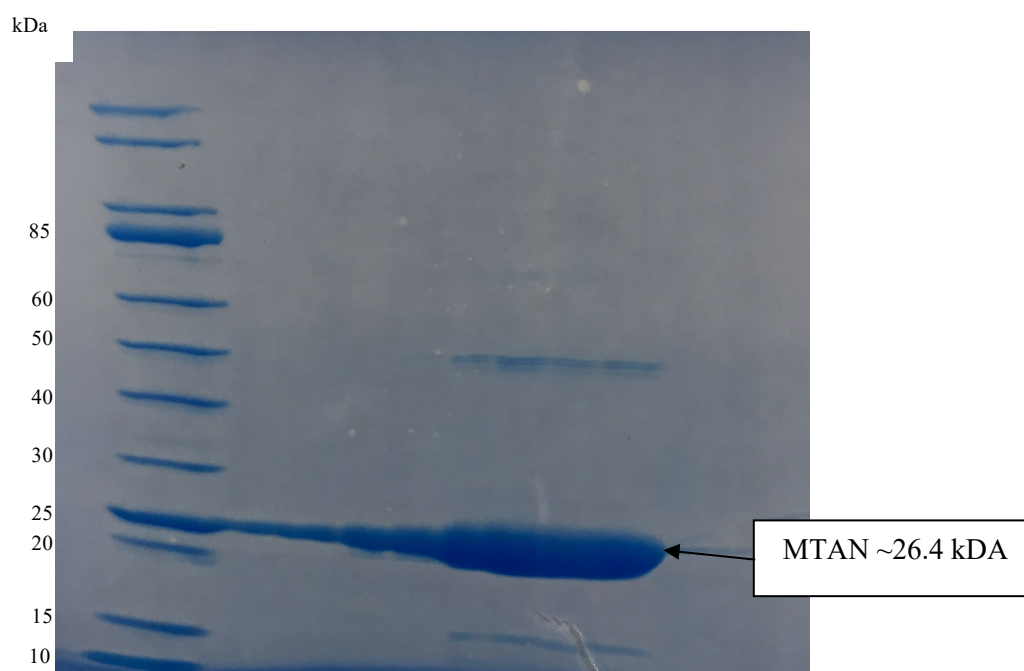
Figure S 3. MTAN affinity purification gel. Marker, MTAN fractions.

6.4 Protein Crystallisation, Data Collection and Refinement

Initial screening of crystallisation conditions was carried out using commercially available INDEX (Hampton Research), PACT premier and CSSI/II (Molecular Dimensions) screens in 96-well sitting drop format. For co-crystallisation experiments, 100 mM stock solutions of ClDA and L-met were prepared in 1:1 DMSO: ddH$_2$O. 150 nL drops consisting of 10 mg/mL enzyme and 10 mM of each substrate were prepared. Each of the solutions were then mixed 1:1 with 150 nL of crystallisation solutions and incubated at 18 ºC overnight. Crystals were obtained in conditions containing 0.2 M calcium chloride dihydrate, 0.1 M HEPES pH 7, 20%

w/v PEG6000. Crystals were flash cooled in liquid nitrogen using nylon CryoLoopsTM (Hampton Research) without any cryoprotectant.

Crystals of SalL were transferred into a cryoprotectant solution of 10% (w/v) ethylene glycol in the mother liquor using nylon CryoLoops[TM] (Hampton Research) before flash-cooling with liquid nitrogen. The datasets described in this report were collected at the Diamond Light Source, Didcot, Oxfordshire, U.K. on beamlines I04-1 (SalL plus *S*-adenosylmethionine plus chloride) and I03 (SalL plus chloroadenosine). Data were processed and integrated using XDS[170] and scaled using SCALA[171] included in the Xia2 processing system.[172] Data collection statistics are provided in Appendix Table S2.1. Crystals of both complexes were obtained in space group $P2_12_12_1$, with three molecules in the asymmetric unit comprising one trimer of SalL in both complexes. The solvent content in the crystals was 52% in each case. The structures were solved by molecular replacement using MOLREP[173] with the monomer of SalL (PDB code 2Q6I)[97] as the model. The structures were built and refined using iterative cycles in Coot[174] and REFMAC,[175] employing local NCS restraints in the refinement cycles. Following building and refinement of the protein and water molecules in the complexes, residual peaks of density were observed in the omit maps at the active sites interfaces. These could easily be modelled as *S*-adenosylmethionine plus chloride in the first instance, and as chloroadenosine in the second. The final structures exhibited % R_{cryst}/R_{free} values of 17.9.0/20.3 and 17.1/20.1 respectively. Refinement statistics for the structures are presented in Appendix Table S2.1. The Ramachandran plot for the SalL plus *S*-adenosylmethionine plus chloride showed 95.1% of residues to be situated in the most favoured regions, 4.0% in additional allowed and 0.9% residues in outlier regions. The values for the SalL plus chloroadenosine complex were 94.8%, 4.0% and 1.2% respectively. The structures have been deposited in the Protein Databank (PDB) with accession codes 6RYZ (SalL plus *S*-adenosylmethionine plus chloride) and 6RZ2 (SalL plus chloroadenosine).

6.5   General procedure for the mutation of SalL

Single point SalL mutations were carried out *via* site-directed mutagenesis using appropriate primers (**Table S1**) purchased from Sigma-Aldrich and a Q5 DNA Polymerase kit from New England Biolabs. Primers were diluted to 10 µM concentration. Each PCR reaction consisted of 10 µL 5x Q5 reaction buffer, 5 µL dNTPs, 2.5 µL of both the forward and reverse primers, 5 µL template SalL DNA and 0.5 µL of Q5 polymerase enzyme with the reaction volume made up to 50 µL with sterile distilled water.

The following cycle parameters were used on the PCR machine:

1. 95 ºC – 3 minutes

2. 95 ºC – 30 seconds

3. 60 ºC – 30 seconds

4. 72 ºC – 8 minutes – Repeat highlighted steps 12x

5. 74 ºC – 10 minutes

6. Hold at 12 ºC

With steps 2-4 repeated 12 times.

A Dpn1 digest was then carried out with subsequent transformation into Top10 competent cells carried out as previously. Colonies were inoculated into 3 mL LB media containing 50 µg/ mL kanamycin and grown overnight at 30 ˚C within a 48-well plate. The DNA was then extracted and purified using a Qiagen Miniprep kit according to the manufacturer's instructions, eluting in 100 µL elution buffer.

Table S4. List of point mutations and the SalL primers used to make each one with the mutant codon underlined.

SalL Mutant		Mutant SalL Primers
W129A	F	CGGTGACCCCGACG<u>GCG</u>TATGGTAAAGATATTG
	R	CAATATCTTTACCATA<u>CGC</u>CGTCGGGGTCACCG
W129F	F	GTGACCCCGACG<u>TTT</u>TATGGTAAAGATATTG
	R	CAATATCTTTACCATA<u>AAA</u>CGTCGGGGTCAC
D183E	F	GAAGTGGTTCGTATC<u>GAA</u>CGCGCATTCGGTAACG
	R	CGTTACCGAATGCGCG<u>TTC</u>GATACGAACCACTTC
D183A	F	GAAGTGGTTCGTATC<u>GCT</u>CGCGCATTCGGTAACG
	R	CGTTACCGAATGCGCG<u>AGC</u>GATACGAACCACTTC
F186A	F	GTATCGATCGCGCA<u>GCC</u>GGTAACGTTTGGACC
	R	GGTCCAAACGTTAC<u>CGG</u>CTGCGCGATCGATAC
F186L	F	GTATCGATCGCGCA<u>CTC</u>GGTAACGTTTGGAC
	R	GTCCAAACGTTACC<u>GAG</u>TGCGCGATCGATAC
N188A	F	GATCGCGCATTCGGT<u>GCC</u>GTTTGGACCAATATTC
	R	GAATATTGGTCCAAAC<u>GGC</u>ACCGAATGCGCGATC
N188Q	F	GATCGCGCATTCGGT<u>CAG</u>GTTTGGACCAATATTC
	R	GAATATTGGTCCAAAC<u>CTG</u>ACCGAATGCGCGATC
W190A	F	CATTCGGTAACGTT<u>GCG</u>ACCAATATTCCGACGC
	R	GCGTCGGAATATTGGT<u>CGC</u>AACGTTACCGAATG

W190G	F	CATTCGGTAACGTT<u>GGG</u>ACCAATATTCCGACGC
	R	GCGTCGGAATATTGGT<u>CCC</u>AACGTTACCGAATG
F228A	F	GTTTTGCAAAACG<u>GCG</u>GGCGAAGTGGATGAG
	R	CTCATCCACTTCGCC<u>CGC</u>CGTTTTGCAAAAC
F228I	F	GTTTTGCAAAACG<u>ATC</u>GGCGAAGTGGATGAG
	R	CTCATCCACTTCGCC<u>GAT</u>CGTTTTGCAAAAC
Y239A	F	GTCAGCCGCTGCTG<u>GCC</u>CTGAACAGCCGTG
	R	CACGGCTGTTCAG<u>GGC</u>CAGCAGCGGCTGAC
Y239T	F	GTCAGCCGCTGCTG<u>ACC</u>CTGAACAGCCGTGG
	R	CCACGGCTGTTCAG<u>GGT</u>CAGCAGCGGCTGAC

6.6 SalL Kinetics Assays

SalL/mutant (0.44 – 8.21 μM), BSA (1 mg/mL), DTT (1.00 mM.), 100 mM potassium phosphate buffer pH 6.8, ClDA (0.2 mM), L-met or analogue (0.50-30.0 mM), 37 °C, 6 mins – 60 mins. Reactions were sampled at 7 time points and quenched in 400 mM sodium formate (pH 3) buffer. Samples were directly analysed by RP-HPLC (Method 1 – Section 6.10) with conversion determined by comparison of SAM/analogue peak to ClDA peak. The identity of products was confirmed by comparing compound retention times with authentic standards or LCMS. SAM formation visualised at 254 nm, SAE formation visualised at 270 nm. Each point in the Michaelis-Menten plots was determined from averaging of three experiments with the error calculated as the standard deviation of the three. Data was plotted using OriginPro 2017 software and fitted with an appropriate kinetic model.

6.7 Screening of SalL Mutants for Activity Towards SAM/SAM analogue generation

All reactions were carried out in triplicate on a 500 µL scale.

To a solution of ClDA/ClDA analogue (100 mM stock in DMSO, final concentration 1.00 mM), L-met/ L-met analogue (500 mM stock in 250 mM NaOH NB L-eth was prepared as 250 mM stock due to low solubility. final concentration 3.00 mM), bovine serum albumin (final concentration 1.00 mg/mL) and DTT (100 mM stock in water, final concentration 1.00 mM in potassium phosphate buffer (100 mM, pH 6.8) was added SalL (final concentration 5.00 µM). The reaction was incubated at 37 °C, 500 rpm for 24 h. 50 µL aliquots were taken at a range of times (t = 0 - 24 h) and quenched by heating to 90 °C for 10 minutes. Samples were clarified by centrifugation (10 minutes, 13.2 k rpm, 4 °C) with the supernatant analysed by RP-HPLC. Conversions were calculated by comparing the ratio of the area of starting material to that of SAM/SAM analogue and MTA/MTA analogue and adenine analogues (known SAM degradation product[103]) generated.

6.8 Screening of SAM analogue formation by SalL

All reactions were carried out in triplicate on a 500 µL scale.

To a solution of ClDA/ClDA analogue (100 mM stock in DMSO, final concentration 400 µM), L-met/ L-met analogue (500 mM stock in 250 mM NaOH NB L-eth was prepared as 250 mM stock due to low solubility. final concentration 2.00 mM), bovine serum albumin (final concentration 1.00 mg/mL) and DTT (100 mM stock in water, final concentration 1.00 mM in potassium phosphate buffer (100 mM, pH 6.8) was added SalL (final concentration 2.10 µM). The reaction was incubated at 37 °C, 500 rpm for 24 h. 50 µL aliquots were taken at a range of times (t = 0 - 24 h) and quenched by heating to 90 °C for 10 minutes. Samples were clarified by centrifugation (10 minutes, 13.2 k rpm, 4 °C) with the supernatant analysed by RP-HPLC. Conversions were calculated by comparing the ratio of the area of starting material to that of SAM/SAM analogue and MTA/MTA analogue and adenine analogues (known SAM degradation product[103]) generated.

Representitive RP-HPLC chromatograms are included in Appendix Section 5.1.

6.9 Methylation/Ethylation Enzyme assays

To a 1.5 mL Eppendorf vial was added ClDA/ClDA analogue (400 μM), L-met (2.00 mM), coumarin (200 μM), SalL (2.10 μM), DTT (1.00 mM) and BSA (1.00 mg/mL) (all final concentrations) in potassium phosphate buffer (100 mM, pH 6.8). Reactions were incubated using a Thermomixer at 37 °C for 24 h at 500 rpm before addition of NovO (9.38 μM) and MTAN (132 nM). Reactions were incubated for a further 24 h before analysis by RP-HPLC and LCMS to confirm product identity. ^{1}H NMR spectra were determined for each compound obtained. % Conversion was calculated by the ratio of area/area% of starting material to product.

6.9.1 Optimised Conditions for methylation of 7-hydroxywarfarin

To a 1.5 mL Eppendorf vial was added 2-Cl-ClDA/ ethyl analogue (1.60 mM), L-met (8.00 mM), coumarin (200 μM), SalL (4.20 μM), DTT (4.00 mM) and BSA (1 mg/mL) (all final concentrations) in potassium phosphate buffer (100 mM, pH 6.8). Reactions were incubated using a Thermomixer at 37 °C for 24 h at 500 rpm before addition of NovO (42.6 μM) and MTAN (528 nM). Reactions were incubated for a further 24 h before analysis by RP-HPLC and LCMS to confirm product identity. ^{1}H NMR spectra were determined for each compound obtained. % Conversion was calculated by the ratio of area/area% of starting material to product.

Representitive RP-HPLC chromatograms are included in Appendix Section 5.2.

6.10 RP-HPLC method parameters

Method 1 -Kinetics:

Column specification: Phenomenex Luna C(18) 3μ phenyl-hexyl (150 x 4.6 mm)
Column temperature: 25 °C
Mobile phase A: 0.1% v / v TFA in water
Mobile phase B: Acetonitrile
Flow rate 1.0 mL/min
Gradient profile:

Time (min)	% A	% B
0	98.5	1.5
2	98.5	1.5
11.5	56.8	43.2
11.51	0	100
14.5	0	100
14.51	98.5	1.5
19	98.5	1.5

Method 2- SAM analogue formation

Column specification: Phenomenex Luna Polar Omega
Column temperature: 25 °C
Mobile phase A: 0.1% v / v TFA in water
Mobile phase B: Acetonitrile
Flow rate 1.5 mL/min
Gradient profile:

Time (min)	% A	% B
0	100	0
0.5	100	0
10	56.8	43.2
10.01	0	100
13.01	0	100
13.02	100	0
16	100	0

Detection: Detection wavelength of 254 nm used for SAM formation.

Method 3- SAM analogue formation formic

Column specification: Phenomenex Luna Polar Omega
Column temperature: 25 °C
Mobile phase A: 0.1% v / v Formic acid in water
Mobile phase B: Acetonitrile
Flow rate 1.5 mL/min

Gradient profile:

Time (min)	% A	% B
0	100	0
0.5	100	0
10	56.8	43.2
10.01	0	100
13.01	0	100
13.02	100	0
16	100	0

Detection: Detection wavelength of 254 nm used for SAM formation.

Methods for methylation and ethylation of coumarin substrates

Column specification: Waters Symmetry C18 (75 x 4.6 mm)
Column temperature: 25 °C
Mobile phase A: 0.1% v / v TFA in water
Mobile phase B: 0.1 % v / v TFA in acetonitrile
Mobile phase D: Acetonitrile
Flow rate 1.5 mL/min
Detection at 300 nm.
The following gradient profiles were used for different substrates.

Method 4 – GSHv2

Time (min)	% A	% B
0	95	5
0.5	95	5
4	40	60
4.1	5	95
5.8	5	95
6	95	5
9	95	5

Method 5 – GSHv4

Time (min)	% A	% D
0.01	95	5
0.5	95	5
8	45	55
4.1	5	95
5.8	5	95
6	95	5
9	95	5

Method 6– GSHv2 7-hydroxywarfarin

Time (min)	% A	% B
0	80	20
7	0	100
9	0	100
9.01	80	20
12	80	20

Method 7 – GSHv2 2-methylcoumarin

Time (min)	% A	% B
0.01	95	5
0.5	95	5
7	0	100
9	0	100
9.01	80	20
12	80	20

Method 8 – SAM stability

Note – flow rate for this method = 1 mL/min. Detection: 254 nm

Time (min)	% A	% B

0.01	99	1
5	40	60
5.1	5	95
7	5	95
7.1	99	1
10	99	1

Method 9 – GSHv2 15 minute gradient

Time (min)	% A	% B
0.01	95	5
0.50	95	5
10.00	30	70
10.01	0	100
12.00	0	100
12.01	95	5
15	95	5

6.11 LCMS methods

Method 1

Method Description: High pH Generic Analytical HPLC LC/MS 5 Minute Method.

LC Conditions: The RP-HPLC analysis was conducted on a Waters Xbridge C18 column (50mm x 4.6 mm i.d. 3 μm packing diameter) at 40 ˚C.

The solvents employed were:

A = 10 mM ammonium bicarbonate in water adjusted to pH 10 with ammonia solution.

B = Acetonitrile.

The gradient employed was:

Time (min)	Flow Rate (ml/min)	% A	% B
0	3	99	1
0.1	3	99	1
4	3	3	97
5	3	3	97

The UV detection was a summed signal from wavelength of 220 nm to 350 nm.

MS Conditions

MS: Waters Z

Ionisation mode: Positive and/or negative electrospray.

Scan Range: 100 to 1000 AMU positive, 120-1000 AMU negative.

Scan Time: 0.5 seconds

Inter scan Delay: 0.05 seconds

AmmAce (Method 2)/Formic method (Method 3)

LC Conditions: The RP-HPLC analysis was conducted on a Zorbax 45mm x 150mm C18 column at 40 ˚C.

The solvents employed were either:

AmmAce method:

A = 5 mM ammonium acetate in water.

B = 5 mM ammonium acetate in acetonitrile.

Or

Formic method:

A = 0.1 % v/v formic acid in water

B = 0.1 % v/v formic acid in acetonitrile

The gradient employed was:

Time (min)	% A	% B
0	95	5
1.48	95	5
8.5	0	100
13.5	0	100
16.5	95	5
18	95	5

The UV detection signal was recorded at either 254 or 310 nm.

MS Conditions

MS: Agilent Quadrupole

Ionisation mode: Positive and/or negative electrospray.

Scan Range: 100 to 1000 AMU positive, 120-1000 AMU negative.

6.12 Mass-Directed Auto Purification (MDAP) Method

Method Description: Trifluoroacetic Acid Focused Preparative Open Access LC/MS gradient, method A. 300 mg loading with At Column Dilution.

LC Conditions: The RP-HPLC analysis was conducted on an XSELECT CSH column (150 mm x 30 mm i.d 5 µm packing diameter) at ambient temperature.

The solvents employed were:
A = Water + 0.1 % v/v trifluoroacetic acid
B = Acetonitrile + 0.1 % v/v trifluoroacetic acid
The gradient employed was:

Time (min)	Flow Rate (ml/min)	% A	% B
0	40	100	0
3	40	100	0
3.5	30	100	0
24.5	30	75	25
25	30	1	99
32	30	1	99

The UV detection was for a signal wavelength at 254 nm.

MS Conditions

MS: Waters ZQ
Ionisation mode: Alternate-scan Positive and Negative Electrospray
Scan Range: 100 to 1000 AMU
Scan Time: 0.5 seconds
Inter scan Delay: 0.2 seconds

6.13 Ion Exchange Method

Dowex 1 x 8 100-200 mesh ion exchange resin, product no. 10276400

Ion exchange column method:

1. Load resin onto column as slurry in water

2. 2 CV water wash

3. 2 CV MeOH wash

4. 2 CV water wash

5. 1 M HCl to pH 5

6. Water to neutral

7. 1 M NH4OH to pH 9

8. Water to neutral

9. 2 CV MeOH wash

10. 2 CV water wash

11. 1 M HCl to pH 5

12. Water to neutral.

Sample was loaded in water basified to pH 10 with ammonium hydroxide (1 M). Sample was washed with 2 column volumes of ammonium hydroxide followed by water to pH 7. After this, the target material was stripped from the column by elution with HCl (1 M).

7   References

1.   H. Schönherr and T. Cernak, *Angew. Chem. Int. Ed.*, 2013, **52**, 12256-12267.
2.   E. J. Barreiro, A. E. Kümmerle and C. A. M. Fraga, *Chem. Rev.*, 2011, **111**, 5215-5246.
3.   C. Schärfer, T. Schulz-Gasch, H.-C. Ehrlich, W. Guba, M. Rarey and M. Stahl, *J. Med. Chem.*, 2013, **56**, 2016-2028.
4.   R. Esser, C. Berry, Z. Du, J. Dawson, A. Fox, R. A. Fujimoto, W. Haston, E. F. Kimble, J. Koehler, J. Peppard, E. Quadros, J. Quintavalla, K. Toscano, L. Urban, J. van Duzer, X. Zhang, S. Zhou and P. J. Marshall, *Br. J. Pharmacol.*, 2005, **144**, 538-550.
5.   M. C. Celia, F. d. P. Andre, M. S. d. S. Gilberto, M. R. S. A. Carlos and A. M. F. a. E. J. B. Carlos, *Letters in Drug Design & Discovery*, 2007, **4**, 422-425.
6.   W. F. Hoffman, A. W. Alberts, P. S. Anderson, J. S. Chen, R. L. Smith and A. K. Willard, *J. Med. Chem.*, 1986, **29**, 849-852.
7.   R. W. Friesen, C. Brideau, C. C. Chan, S. Charleson, D. Deschênes, D. Dubé, D. Ethier, R. Fortin, J. Y. Gauthier, Y. Girard, R. Gordon, G. M. Greig, D. Riendeau, C. Savoie, Z. Wang, E. Wong, D. Visco, L. J. Xu and R. N. Young, *Bioorg. & Med. Chem. Lett.*, 1998, **8**, 2777-2782.
8.   P. M. Ginnings and R. Baum, *J. Am. Chem. Soc.*, 1937, **59**, 1111-1113.
9.   P. Bazzini and C. G. Wermuth, in *The Practice of Medicinal Chemistry (Third Edition)*, Academic Press, New York, 2008, DOI: http://dx.doi.org/10.1016/B978-0-12-374194-3.00020-2, pp. 429-463.
10. K. A. Witt and T. P. Davis, *The AAPS Journal*, 2006, **8**, E76-E88.
11. J. M. F. Crafts, C., *J. Chem. Soc.*, 1877, **32**, 725-791.
12. D. J. Macquarrie, *Catalytic Asymmetric Friedel-Crafts Alkylations*, WILEY-VCH Verlag GmbH & Co. KHaA, Weinheim, 2009, pp 271-288.
13. M. Rueping and B. J. Nachtsheim, *Beilstein J. Org. Chem.*, 2010, **6**, 6.
14. M. Bandini, A. Melloni and A. Umani-Ronchi, *Angew. Chem. Int. Ed.*, 2004, **43**, 550-556.
15. L. Chen, B.-X. Xiao, W. Du and Y.-C. Chen, *Org. Lett.*, 2019, DOI: 10.1021/acs.orglett.9b02108.
16. T. Tsuchimoto, K. Tobita, T. Hiyama and S.-i. Fukuzawa, *J. Org. Chem.*, 1997, **62**, 6997-7005.
17. N. A. Paras and D. W. C. MacMillan, *J. Am. Chem. Soc.*, 2001, **123**, 4370-4371.
18. B. R. Langlois, E. Laurent and N. Roidot, *Tetrahedron Lett.*, 1991, **32**, 7525-7528.
19. Y. Fujiwara, J. A. Dixon, F. O'Hara, E. D. Funder, D. D. Dixon, R. A. Rodriguez, R. D. Baxter, B. Herle, N. Sach, M. R. Collins, Y. Ishihara and P. S. Baran, *Nature*, 2012, **492**, 95-99.
20. Y. Ji, T. Brueckl, R. D. Baxter, Y. Fujiwara, I. B. Seiple, S. Su, D. G. Blackmond and P. S. Baran, *Proc. Natl. Acad. Sci.*, 2011, **108**, 14411-14415.
21. N. A. Meanwell, *J. Med. Chem.*, 2011, **54**, 2529-2591.
22. D. A. Nagib and D. W. C. MacMillan, *Nature*, 2011, **480**, 224.
23. G. Rouquet and N. Chatani, *Angew. Chem. Int. Ed.*, 2013, **52**, 11726-11743.
24. Y. Zhao and G. Chen, *Org. Lett.*, 2011, **13**, 4850-4853.

25. R. Giri, N. Maugel, J.-J. Li, D.-H. Wang, S. P. Breazzano, L. B. Saunders and J.-Q. Yu, *J. Am. Chem. Soc.*, 2007, **129**, 3510-3511.

26. Q.-L. Yang, C.-Z. Li, L.-W. Zhang, Y.-Y. Li, X. Tong, X.-Y. Wu and T.-S. Mei, *Organometallics*, 2019, **38**, 1208-1212.

27. S. R. Neufeldt, C. K. Seigerman and M. S. Sanford, *Org. Lett.*, 2013, **15**, 2302-2305.

28. J. P. Adams, M. J. B. Brown, A. Diaz-Rodriguez, R. C. Lloyd and G.-D. Roiban, *Adv.. Synth. Catal.*, 2019, **361**, 2421-2432.

29. M. Petchey, A. Cuetos, B. Rowlinson, S. Dannevald, A. Frese, P. W. Sutton, S. Lovelock, R. C. Lloyd, I. J. S. Fairlamb and G. Grogan, *Angew. Chem. Int. Ed.*, 2018, **57**, 11584-11588.

30. R. C. Brewster, J. T. Suitor, A. W. Bennett and S. Wallace, *Angew. Chem. Int. Ed.*, 2019, DOI: 10.1002/anie.201903973.

31. D. Güclü, A. Szekrenyi, X. Garrabou, M. Kickstein, S. Junker, P. Clapés and W.-D. Fessner, *ACS Catal.*, 2016, **6**, 1848-1852.

32. J. E. Puskas, M. Y. Sen and K. S. Seo, *J. Polym. Sci. Pol. Chem.*, 2009, **47**, 2959-2976.

33. Z. C. Litman, Y. Wang, H. Zhao and J. F. Hartwig, *Nature*, 2018, **560**, 355-359.

34. K. M. Rentsch, *J. Biochem. Bioph. Meth.*, 2002, **54**, 1-9.

35. R. N. Patel, *Encyclopedia of Industrial Biotechnology*, 2010, DOI: doi:10.1002/9780470054581.eib631
10.1002/9780470054581.eib631, 1-15.

36. R. A. Sheldon and J. M. Woodley, *Chem. Rev.*, 2018, **118**, 801-838.

37. C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman and G. J. Hughes, *Science*, 2010, **329**, 305.

38. N. G. Schmidt, T. Pavkov-Keller, N. Richter, B. Wiltschi, K. Gruber and W. Kroutil, *Angew. Chem. Int. Ed.*, 2017, **56**, 7615-7619.

39. E. E. Schultz, N. R. Braffman, M. U. Luescher, H. H. Hager and E. P. Balskus, *Angew. Chem. Int. Ed.*, 2019, **58**, 3151-3155.

40. L. Zhao, X. Chen, L. Cai, C. Zhang, Q. Wang, S. Jing, G. Chen, J. Li, J. Zhang and Y. Fang, *J. Clin. Pharm. Ther.*, 2014, **39**, 418-423.

41. R. Bissonnette, L. S. Vasist, J. N. Bullman, T. Collingwood, G. Chen and T. Maeda-Chubachi, *Clin. Pharm. Drug Dev.*, 2018, **7**, 524-531.

42. M. Bremang, A. Cuomo, A. M. Agresta, M. Stugiewicz, V. Spadotto and T. Bonaldi, *Mol. Biosyst.*, 2013, **9**, 2231-2247.

43. S. M. Carr, A. Poppy Roworth, C. Chan and N. B. La Thangue, *FEBS Journal*, 2015, **282**, 4450-4465.

44. N. Pivac, P. Pregelj, M. Nikolac, T. Zupanc, G. Nedic, D. Muck Seler and A. Videtic Paska, *Genes Brain Behav.*, 2011, **10**, 565-569.

45. W. Jakoby and R. Weisiger, *Enzymatic Basis of Detoxification*, Academic, New York, 1980.

46. O. H. Drummer, P. Miach and B. Jarrott, *Biochem. Pharmacol.*, 1983, **32**, 1557-1562.

47. O. H. Drummer, L. Routley and N. Christophidis, *Biochem. Pharmacol.*, 1987, **36**, 1197-1201.

48.   R. Weinshilboum, *Fed. Proc*, 1986, **45**, 2220-2222.

49.   U. T. Sankpal and D. N. Rao, *Crit. Rev. Biochem. Mol. Biol.*, 2002, **37**, 167-197.

50.   M. Pacholec, J. Tao and C. T. Walsh, *Biochemistry*, 2005, **44**, 14969-14976.

51.   V. M. Richon, D. Johnston, C. J. Sneeringer, L. Jin, C. R. Majer, K. Elliston, L. F. Jerva, M. P. Scott and R. A. Copeland, *Chem. Biol. Drug Des.*, 2011, **78**, 199-210.

52.   G. L. Cantoni, *Annu. Rev. Biochem.*, 1975, **44**, 435-451.

53.   G. L. Cantoni, *J. Am. Chem. Soc.*, 1952, **74**, 2942-2943.

54.   J. Vidgren, L. A. Svensson and A. Liljas, *Nature*, 1994, **368**, 354-358.

55.   J. L. Martin and F. M. McMillan, *Curr. Opin. Struct. Biol.*, 2002, **12**, 783-793.

56.   R. Blumenthal and X. Cheng, *S-adenosylmethionine Dependent Methyltransferases: Structures and Functions*, World Scientific Publishing, Singapore, 1999.

57.   R. Agarwal, S. K. Burley and S. Swaminathan, Crystal structure of a SAM dependent methyltransferase from Aquifex aeolicus, 2016, PDB Accession code: 3DH0.

58.   R. Z. Jurkowska, T. P. Jurkowski and A. Jeltsch, *ChemBioChem*, 2011, **12**, 206-222.

59.   M. van Haren, L. Q. van Ufford, E. E. Moret and N. I. Martin, *Org. Biomol. Chem.*, 2015, **13**, 549-560.

60.   S. M. Watkins, X. Zhu and S. H. Zeisel, *J. Nutr.*, 2003, **133**, 3386-3391.

61.   R. W. Woodard, M.-D. Tsai, H. G. Floss, P. A. Crooks and J. K. Coward, *J. Biol. Chem.*, 1980, **255**, 9124-9127.

62.   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 and R. C. Trievel, *J. Am. Chem. Soc.*, 2013, **135**, 15536-15548.

63.   C. R. Creveling and D. R. Thakker, in *Conjugation—Deconjugation Reactions in Drug Metabolism and Toxicity*, ed. F. C. Kauffman, Springer Berlin Heidelberg, Berlin, Heidelberg, 1994, DOI: 10.1007/978-3-642-78429-3_7, pp. 189-216.

64.   E. H. Hansen, B. L. Møller, G. R. Kock, C. M. Bünner, C. Kristensen, O. R. Jensen, F. T. Okkels, C. E. Olsen, M. S. Motawia and J. Hansen, *Appl. Environ. Microbiol.*, 2009, **75**, 2765.

65.   B. J. C. Law, M. R. Bennett, M. L. Thompson, C. Levy, S. A. Shepherd, D. Leys and J. Micklefield, *Angew. Chem., Int. Ed.*, 2016, **55**, 2683-2687.

66.   K. E. Moore and O. Gozani, *BBA-Gene Regul. Mech.*, 2014, **1839**, 1395-1403.

67.   H. Wei, R. Mundade, K. C. Lange and T. Lu, *Cell Cycle*, 2014, **13**, 32-41.

68.   X. Zhang, L. Zhou and X. Cheng, *The EMBO Journal*, 2000, **19**, 3509.

69.   H.-B. Guo and H. Guo, *Proc. Natl. Acad. Sci.*, 2007, **104**, 8797-8802.

70.   J. Attieh, R. Djiana, P. Koonjul, C. Étienne, S. A. Sparace and H. S. Saini, *Plant Mol. Biol.*, **50**, 511-521.

71.   J. F. Costello and C. Plass, *J. Med. Genet.*, 2001, **38**, 285-303.

72.   H. O. Smith, *Science*, 1979, **205**, 455-462.

73.   R. J. Roberts, P. A. Myers, A. Morrison and K. Murray, *J. Mol. Biol.*, 1976, **103**, 199-208.

74.   I. Crnovcic, R. Suessmuth and U. Keller, *Biochemistry*, 2010, **49**, 9698-9705.

75. D. J. Patel, *BBA-Nucleic Acids and Protein Synthesis*, 1976, **442**, 98-108.

76. D. A. Schoenfeld, C. Rosenbaum, J. Horton, J. M. Wolter, G. Falkson and R. C. DeConti, *Cancer*, 1982, **50**, 2757-2762.

77. J. C. Sadler, C.-w. H. Chung, J. E. Mosley, G. A. Burley and L. D. Humphreys, *ACS Chem. Biol.*, 2017, **12**, 374-379.

78. M. Tengg, H. Stecher, P. Remler, I. Eiteljörg, H. Schwab and M. Gruber-Khadjawi, *J. Mol. Catal.*, 2012, **84**, 2-8.

79. T. Pavkov-Keller, K. Steiner, M. Faber, M. Tengg, H. Schwab, M. Gruber-Khadjawi and K. Gruber, *PLOS ONE*, 2017, **12**, e0171056.

80. H. Stecher, M. Tengg, B. J. Ueberbacher, P. Remler, H. Schwab, H. Griengl and M. Gruber-Khadjawi, *Angew. Chem. Int. Ed.*, 2009, **48**, 9546-9548.

81. M. R. Bennett, S. A. Shepherd, V. A. Cronin and J. Micklefield, *Curr. Opin. Chem. Biol.*, 2017, **37**, 97-106.

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

83. S. Singh, J. Zhang, T. D. Huber, M. Sunkara, K. Hurley, R. D. Goff, G. Wang, W. Zhang, C. Liu, J. Rohr, S. G. Van Lanen, A. J. Morris and J. S. Thorson, *Angew. Chem. Int. Ed.*, 2014, **53**, 3965-3969.

84. J. L. Hoffman, *Biochemistry*, 1986, **25**, 4444-4449.

85. J. Komoto, T. Yamada, Y. Takata, G. D. Markham and F. Takusagawa, *Biochemistry*, 2004, **43**, 1821-1831.

86. G. D. Markham, E. W. Hafner, C. W. Tabor and H. Tabor, *J. Biol. Chem.*, 1980, **255**, 9082-9092.

87. D. F. Iwig, A. T. Grippe, T. A. McIntyre and S. J. Booker, *Biochemistry*, 2004, **43**, 13510-13524.

88. S. Willnow, M. Martin, B. Lüscher and E. Weinhold, *ChemBioChem*, 2012, **13**, 1167-1173.

89. C. Schaffrath, H. Deng and D. O'Hagan, *FEBS Letters*, 2003, **547**, 111-114.

90. X. Zhu, D. A. Robinson, A. R. McEwan, D. O'Hagan and J. H. Naismith, *J. Am. Chem. Soc.*, 2007, **129**, 14597-14604.

91. H. Deng, S. L. Cobb, A. R. McEwan, R. P. McGlinchey, J. H. Naismith, D. O'Hagan, D. A. Robinson and J. B. Spencer, *Angew. Chem. Int. Ed.*, 2006, **45**, 759-762.

92. C. Dong, F. Huang, H. Deng, C. Schaffrath, J. B. Spencer, D. O'Hagan and J. H. Naismith, *Nature*, 2004, **427**, 561-565.

93. S. L. Cobb, H. Deng, A. R. McEwan, J. H. Naismith, D. O'Hagan and D. A. Robinson, *Org. Biomol. Chem.*, 2006, **4**, 1458-1460.

94. H. Deng, L. Ma, N. Bandaranayaka, Z. Qin, G. Mann, K. Kyeremeh, Y. Yu, T. Shepherd, J. H. Naismith and D. O'Hagan, *ChemBioChem*, 2014, **15**, 364-368.

95. 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 and E. L. Ang, *Angew. Chem. Int. Ed.*, 2016, **55**, 14277-14280.

96. W. L. Yeo, X. Chew, D. J. Smith, K. P. Chan, H. Sun, H. Zhao, Y. H. Lim and E. L. Ang, *Chem. Comm.*, 2017, **53**, 2559-2562.

97. A. S. Eustaquio, F. Pojer, J. P. Noel and B. S. Moore, *Nat. Chem. Biol.*, 2008, **4**, 69-74.

98. R. H. Feling, G. O. Buchanan, T. J. Mincer, C. A. Kauffman, P. R. Jensen and W. Fenical, *Angew. Chem. Int. Ed.*, 2003, **42**, 355-357.

99. J. M. Lipson, M. Thomsen, B. S. Moore, R. P. Clausen, J. J. La Clair and M. D. Burkart, *ChemBioChem*, 2013, **14**, 950-953.

100. M. Thomsen, S. B. Vogensen, J. Buchardt, M. D. Burkart and R. P. Clausen, *Org. Biomol. Chem.*, 2013, **11**, 7606-7610.

101. R. Shi, S. S. Lamb, B. Zakeri, A. Proteau, Q. Cui, T. Sulea, A. Matte, G. D. Wright and M. Cygler, *Chemistry & Biology*, 2009, **16**, 401-410.

102. J. C. Sadler, L. D. Humphreys, G. Burley and R. Snajdrova, *ChemBioChem*, 2017, **18**, 992-995

103. T. D. Huber, F. Wang, S. Singh, B. R. Johnson, J. Zhang, M. Sunkara, S. G. Van Lanen, A. J. Morris, G. N. Phillips and J. S. Thorson, *ACS Chem. Biol.*, 2016, **11**, 2484-2491.

104. J. K. Coward, N. C. Motola and J. D. Moyer, *J. Med. Chem.*, 1977, **20**, 500-505.

105. P. J. Berti and J. A. B. McCann, *Chem. Rev.*, 2006, **106**, 506-555.

106. J. A. B. McCann and P. J. Berti, *J. Am. Chem. Soc.*, 2007, **129**, 7055-7064.

107. C. F. Matta, A. A. Arabi and D. F. Weaver, *Eur. J. Med. Chem.*, 2010, **45**, 1868-1872.

108. Y. C. Martin, *J. Med. Chem.*, 1996, **39**, 1189-1190.

109. S. Mordhorst, J. Siegrist, M. Müller, M. Richter and J. N. Andexer, *Angew. Chem. Int. Ed.*, 2017, **56**, 4037-4041.

110. C. Liao and F. P. Seebeck, *Nat. Catal.*, 2019, DOI: 10.1038/s41929-019-0300-0.

111. J. Davies, *Trends Biotechnol.*, 1988, **6**, S7-S11.

112. U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore and K. Robins, *Nature*, 2012, **485**, 185.

113. M. T. Reetz, P. Soni, J. P. Acevedo and J. Sanchis, *Angew. Chem. Int. Ed.*, 2009, **48**, 8268-8272.

114. H. Lee, J. H. Kim, S. Han, Y. R. Lim, H. G. Park, Y. J. Chun, S. W. Park and D. Kim, *J. Toxicol. Environ. Health. Part A*, 2014, **77**, 1409-1418.

115. M. T. Reetz, *Proc. Natl. Acad. Sci.*, 2004, **101**, 5716.

116. K. Chen and F. H. Arnold, *Proc. Natl. Acad. Sci.*, 1993, **90**, 5618.

117. M. Fromant, S. Blanquet and P. Plateau, *Anal. Biochem.*, 1995, **224**, 347-353.

118. W. P. C. Stemmer, *Nature*, 1994, **370**, 389-391.

119. W. P. Stemmer, *Proc. Natl. Acad. Sci.*, 1994, **91**, 10747.

120. R. M. P. Siloto and R. J. Weselake, *Biocatalysis and Agricultural Biotechnology*, 2012, **1**, 181-189.

121. M. T. Reetz and J. D. Carballeira, *Nat. Protoc.*, 2007, **2**, 891.

122. K. R. Tindall and T. A. Kunkel, *Biochemistry*, 1988, **27**, 6008-6013.

123. K. A. Eckert and T. A. Kunkel, *Nucleic Acids Res*, 1990, **18**, 3739-3744.

124. H. Zhao and F. H. Arnold, *Nucleic Acids Res*, 1997, **25**, 1307-1308.

125. *QuikChange Site-Directed Mutagenesis Kit*. Instruction Manual (Stratagene, La Jolla, California, 2003)

126. S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z.-G. Zhang, D. J. Opperman, M. T. Reetz and J. P. Acevedo, *ACS Synth. Biol.*, 2013, **2**, 83-92.

127. M. D. Hughes, D. A. Nagel, A. F. Santos, A. J. Sutherland and A. V. Hine, *J. Mol. Biol.*, 2003, **331**, 973-979.

128. R. F. Balint and J. W. Larrick, *Gene*, 1993, **137**, 109-118.

129. M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha and A. Vogel, *Angew. Chem. Int. Ed.*, 2005, **44**, 4192-4196.

130. M. Nardini, D. A. Lang, K. Liebeton, K.-E. Jaeger and B. W. Dijkstra, *Journal of Biological Chemistry*, 2000, **275**, 31219-31225.

131. M. T. Reetz, A. Zonta, K. Schimossek, K.-E. Jaeger and K. Liebeton, *Angew. Chem. Int. Ed.*, 1997, **36**, 2830-2832.

132. J. D. Carballeira, P. Krumlinde, M. Bocola, A. Vogel, M. T. Reetz and J.-E. Bäckvall, *Chem. Comm.*, 2007, DOI: 10.1039/B700849J, 1913-1915.

133. M. T. Reetz, L.-W. Wang and M. Bocola, *Angew. Chem. Int. Ed.*, 2006, **45**, 1236-1241.

134. S. Bartsch, R. Kourist and U. T. Bornscheuer, *Angew. Chem. Int. Ed.*, 2008, **47**, 1508-1511.

135. S. B. J. Kan, R. D. Lewis, K. Chen and F. H. Arnold, *Science*, 2016, **354**, 1048.

136. P. S. Coelho, E. M. Brustad, A. Kannan and F. H. Arnold, *Science*, 2013, **339**, 307.

137. Addgene, pET-26b(+), https://www.addgene.org/vector-database/2563/, (accessed 17/09/19, 2019).

138. F. William Studier, A. H. Rosenberg, J. J. Dunn and J. W. Dubendorff, in *Methods in Enzymology*, Academic Press, 1990, vol. 185, pp. 60-89.

139. F. W. Studier, *Protein Expr. Purif.*, 2005, **41**, 207-234.

140. S. Thompson, Q. Zhang, M. Onega, S. McMahon, I. Fleming, S. Ashworth, J. H. Naismith, J. Passchier and D. O'Hagan, *Angew. Chem. Int. Ed.*, 2014, **53**, 8913-8918.

141. H. Liu and J. H. Naismith, *BMC Biotechnol.*, 2008, **8**, 91.

142. J. L. Anglin, L. Deng, Y. Yao, G. Cai, Z. Liu, H. Jiang, G. Cheng, P. Chen, S. Dong and Y. Song, *J. Med. Chem.*, 2012, **55**, 8066-8074.

143. J. S. Lolkema and D.-J. Slotboom, *J. Gen. Physiol.*, 2015, **145**, 565-574.

144. Sigma Aldrich, Chloride Assay Kit, https://www.sigmaaldrich.com/catalog/product/sigma/mak023?lang=en&region=GB, (accessed 15/08, 2019).

145. BioVision, Chloride Colorimetric Assay Kit, https://www.biovision.com/chloride-colorimetric-assay-kit.html).

146. C. Dalhoff, G. Lukinavius, S. Klimaauskas and E. Weinhold, *Nat. Protoc.*, 2006, **1**, 1879-86.

147. R. Wang, W. Zheng, H. Yu, H. Deng and M. Luo, *J. Am. Chem. Soc.*, 2011, **133**, 7648-7651.

148. M. Imiołek, G. Karunanithy, W.-L. Ng, A. J. Baldwin, V. Gouverneur and B. G. Davis, *J. Am. Chem. Soc.*, 2018, **140**, 1568-1571.

149. B. Li, J. Zhang, Y. Xu, X. Yang and L. Li, *Tetrahedron Lett.*, 2017, **58**, 2374-2377.

150. J. C. Sadler, PhD Thesis, 2017, University of Strathclyde.

151. D. Cao, Z. Liu, P. Verwilst, S. Koo, P. Jangjili, J. S. Kim and W. Lin, *Chem. Rev.*, 2019, DOI: 10.1021/acs.chemrev.9b00145.

152. G. Jones and M. A. Rahman, *J. Phys. Chem.*, 1994, **98**, 13028-13037.

153. A. Dorlars, C.-W. Schellhammer and J. Schroeder, *Angew. Chem. Int. Ed.*, 1975, **14**, 665-679.

154. G. A. Reynolds and K. H. Drexhage, *Opt. Commun.*, 1975, **13**, 222-225.

155. Z. Xu, X. Chen, H. N. Kim and J. Yoon, *Chem. Soc. Rev.*, 2010, **39**, 127-137.

156. Y. Zhou, J. F. Zhang and J. Yoon, *Chem. Rev.*, 2014, **114**, 5511-5571.

157. J. Wu, B. Kwon, W. Liu, E. V. Anslyn, P. Wang and J. S. Kim, *Chem. Rev.*, 2015, **115**, 7893-7943.

158. J. Grover and S. M. Jachak, *RSC Adv.*, 2015, **5**, 38892-38905.

159. M. Z. Hassan, H. Osman, M. A. Ali and M. J. Ahsan, *Eur. J. Med. Chem.*, 2016, **123**, 236-255.

160. M. Taipale, D. F. Jarosz and S. Lindquist, *Nat. Rev. Mol. Cell Biol.*, 2010, **11**, 515-528.

161. J. Zhao, H. Zhao, J. A. Hall, D. Brown, E. Brandes, J. Bazzill, P. T. Grogan, C. Subramanian, G. Vielhauer, M. S. Cohen and B. S. J. Blagg, *MedChemComm*, 2014, **5**, 1317-1323.

162. W. R. Porter, *J. Comput. Aid. Mol. Des.*, 2010, **24**, 553-573.

163. M. Ikawa, M. A. Stahmann and K. P. Link, *J. Am. Chem. Soc.*, 1944, **66**, 902-906.

164. E. F. Unger, Atrial fibrillation and new oral anticoagulant drugs, https://www.fda.gov/drugs/news-events-human-drugs/atrial-fibrillation-and-new-oral-anticoagulant-drugs, (accessed 23/09/19, 2019).

165. L. S. Kaminsky and Z.-Y. Zhang, *Pharmacol. Ther.*, 1997, **73**, 67-74.

166. S. Sun, Q. Jia and Z. Zhang, *Bioorg. Med. Chem. Lett.*, 2019, **29**, 2535-2550.

167. X. M. Yu, G. Shen, L. Neckers, H. Blake, J. Holzbeierlein, B. Cronk and B. S. J. Blagg, *J. Am. Chem. Soc.*, 2005, **127**, 12778-12779.

168. H. Yasui, T. Yamamoto, E. Tokunaga and N. Shibata, *J. Fluor. Chem.*, 2011, **132**, 186-189.

169. A. Hiremathad, K. Chand, A. R. Esteves, S. M. Cardoso, R. R. Ramsay, S. Chaves, R. S. Keri and M. A. Santos, *RSC Adv.*, 2016, **6**, 53519-53532.

170. W. Kabsch, *Acta Cryst. D*, 2010, **66**, 125-132.

171. P. Evans, *Acta Cryst. D, Biological Crystallography*, 2006, **62**, 72-82.

172. G. Winter, *J. Appl. Crystallogr.*, 2010, **43**, 186-190.

173. A. Vagin and A. Teplyakov, *J. Appl. Crystallogr.*, 1997, **30**, 1022-1025.

174. P. Emsley and K. Cowtan, *Acta Cryst. D, Biological Crystallography*, 2004, **60**, 2126-2132.

175. G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Cryst. D*, 1997, **53**, 240-255.