# Structural Analysis of a *Papaver somniferum O*-Methyltransferase Involved in Noscapine Biosynthesis

Marc Patrick Cabry

PhD

University of York

Biology

September 2017

## Abstract

Noscapine is a major alkaloid that is found, along with the more commonly known drugs morphine and codeine, in the latex of opium poppy (*Papaver somniferum*). Noscapine has classically been used as a cough suppressant, but also has emerging antitumor properties. Unlike morphine and codeine, noscapine has no addictive properties. Recently, a 10-gene cluster for the synthesis of noscapine was discovered. The first distinct step in this pathway is an *O*-methylation of scoulerine to tetrahydrocolumbamine by a class I *S*-adenosylmethionine (SAM) dependent methyltransferase, referred to as *Papaver somniferum O*-methyltransferase 1 (PSMT1). PSMT1 is therefore an interesting target for detailed structural and functional analysis. PSMT1 has been expressed as a fusion protein, tagged to improve solubility, and purified for X-ray diffraction studies. Here we present the structures of an "apo" enzyme and an enzyme complex with SAM, bound in "open" enzyme conformations. A surface entropy reduction mutant enzyme allowed access to a closed conformation high-resolution structure in complex with *S*-adenosylhomocysteine (SAH) and the product tetrahydrocolumbamine.

These X-ray structures provide deep insight into the structure function relationships of PSMT1. PSMT1 is a homodimer with distinct dimerisation, SAM binding and substrate binding domains. PSMT1 undergoes an inward rotation of approximately 10 degrees of the SAM binding domain upon substrate binding. This results in an 8 Å closure of the active site cleft. The acceptor substrate binds in a hydrophobic cleft, held in place by two hydrogen bonds *via* the hydroxyl groups at positions 2 and 9 of scoulerine.

Further analysis of the plant *O*-methyltransferase catalytic dyad has been carried out by site-directed mutagenesis and X-ray diffraction studies. Subunit cooperativity has also been investigated with the proposal of two regulatory features. Here I present one of the highest resolution closed conformation structure of a plant *O*-methyltransferase to date. This provides the template for our understanding of, and future engineering substrate specificity in plant methyltransferases; an important class of enzymes in secondary metabolism.

# Table of Contents

# List of Figures

# List of Tables

## Acknowledgements

I would like to acknowledge and thank my supervisors, Prof. Ian A. Graham and Prof. Gideon J. Davies for the opportunity to study at the University of York and the guidance and support over the years. I also thank Prof. Neil C. Bruce and Prof. Jennifer R. Potts for their time and constructive discussions as part of my thesis advisory panel. I am very grateful for the BBSRC White Rose Consortium for financial support.

I would like to thank all those past and present members of the Graham and Davies groups especially Dr. Thilo Winzer, Dr. Marcelo Kern, Dr. Tony Larson, David Harvey and Judith Mitchell. Special thanks to Wendy A. Offen for her guidance, support and help throughout my PhD studies, especially in the first few months helping me quickly get settled and comfortable in the York Structural Biology Laboratory.

The biggest thanks go to my family and friends for their support, encouragement and love, without them it would not have been possible. My parents have always wanted the best for me and have provided both financial and emotional support throughout all my life. For this, I am truly grateful and could not wish for better parents. I would also like to thank my twin Christopher Cabry who carried out a PhD at the University of York at the same time as me. He has been a good flat mate/friend/brother and will be missed once we both move on to our next adventure in life. A big thanks goes to my partner Eve Bowdler for putting up with me over the past 7 years and making me see the better side when times were hard and always brightening up my days.

## Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

The following pieces of work which are discussed in the thesis are acknowledged along with the people who conducted them:

- Initial cloning and expression trials of PSMT1 was carried out at the Biosciences Technology Facility at the University of York prior to the commencement of this thesis.
- Initial expression, purification and structure determination experiments were carried out by Wendy A. Offen prior to the commencement of this thesis. This work resulted in the elucidation of a 3.1 Å PSMT1 structure in an open conformation by molecular replacement utilising residues 32-344 of caffeic acid *O*-methyltransferase molecule A (PDB:1KYZ) as a model for Phaser. A X-ray diffraction data set of PSMT1 co-crystallised with SAM obtained by Wendy A. Offen was also utilised by the author.
- Scoulerine was synthesised by Prof Peter J. Schammells (Monash University, Australia)

# 1. Introduction

## 1.1. *Papaver somniferum*

*Papaver somniferum* more commonly known as opium poppy is an annual herb which grows to around 1.5 m tall and has a single flower that can be white, red or blue-purple in colour. After flowering the fruit or seed pod grows to around 5 cm in diameter. Opium poppy produces a substance called latex which contains a variety of alkaloids such as codeine, morphine, noscapine, papaverine, laudanosine and sanguinarine. Latex is the cytoplasm of specialised laticifer cells which are associated with the phloem. The latex of opium poppy was historically extracted by scoring the unripe seed pod allowing the latex to seep out and dry. The dry latex is known as opium and is black in colour and typically contains around 10 % morphine, depending on the strain cultivated.

## 1.2. A Brief History of Opium Poppy

Opium poppy has been exploited by humans for thousands of years due to its unique medicinal properties. It has been proposed that opium poppy was first cultivated around 3000-5000 BC in ancient Mesopotamia, which is modern day Iraq and Kuwait.[1,2]

Philippus Aureolus Theophrastus Bombastus von Hohenheim known as Paracelsus, was a 16$^{th}$ century alchemist whom is credited as the founding father of toxicology.[3] He popularised Paracelsus' laudanum which is a tincture (an alcohol preparation of a drug) form of opium prepared with other substances. It was described as sleep inducing and possessed analgesic properties.[4] The name laudanum comes from the Latin laudare which means to praise. In the early 17$^{th}$ century Thomas Sydenham standardised the use of opium in what was known as Sydenham's laudanum. It was a tincture containing about 10 % opium, resulting in around 1 % morphine content, along with clove, saffron and cinnamon.[1,4,5] It was portrayed as a cure for all. The National Institute for Health and Care Excellence (NICE) 'Guidelines for palliative care for adults: strong opioids for pain relief' recommends a typical daily dose of between 20-30 mg of oral morphine, this would have corresponded to approximately 2-3 mL

of Sydenham's laudanum.[6] Thomas Dover, an apprentice of Sydenham produced a powdered form of opium called Dover's powder.[5] These two preparations were common place for centuries as an analgesic medicine, used to treat pain and disease. Despite the attempts to standardise the preparations, they could vary in potency due to natural differences or adultery of the material. Standardisation would only be possible once the active compound for pain relief had been identified, isolated and characterised.

It was not until 1806 that Friedrich Sertürner first reported the isolation of morphine from opium.[1,4,7] He initially named it 'principium somniferum' but later renamed to morphium after Morpheus the Greek God of dreams due to its sleep inducing effect when administered to animals.[5] Nowadays, the term morphine is used instead of morphium. The discovery of morphine is a milestone in the history of modern pharmacology: For the first time it had become possible to observe and measure the relationship between a precise dose of an active compound and its pharmaceutical effect. The discovery ushered in an intense era of discovery of pharmaceutically active molecules from plants that is still continuing today. Soon after the isolation of morphine in 1817 Pierre Jean Robiquet isolated noscapine from Derosne's salt.[8,9] Derosne's salt was a salt extracted from opium in 1803 by Jean-François Derosne, which contained noscapine and other alkaloids.[9,10] Pierre Jean Robiquet also isolated codeine in 1832.[9,11] Thebaine was discovered in 1832 by Thibouméry and papaverine was discovered by Georg Merck in 1849.[12,13]

In modern medicine the use of morphine has been reduced nearly exclusively to intravenous injection in hospital settings.[14] Codeine is widely used and is available as over the counter oral medication with paracetamol.[15] Codeine is preferred for oral consumption due to its longer acting analgesic effects. Noscapine is also still utilised as a cough suppressant in many countries.[16]

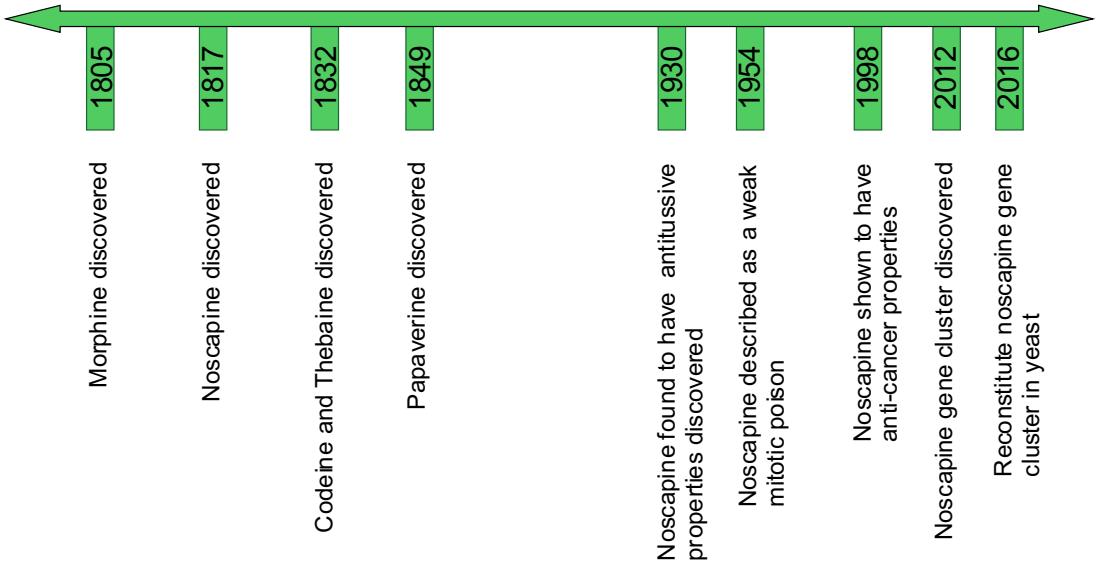



*Figure 1.1: Timeline of opium poppy major discoveries in relation to noscapine*

## 1.3. Alkaloids

Alkaloids are nitrogen containing secondary metabolites which are structurally very heterogeneous compounds most of which possess basic properties, hence the name alkaloids. Secondary metabolites are not just produced by plants, they are also synthesised in bacteria, algae, corals, sponges and lower animals.[17] Alkaloids are not essential for plant growth but are thought to play a role in plant defence against herbivores and pathogens.[18] Commonly known alkaloids are cocaine, nicotine, caffeine, atropine, morphine, codeine and noscapine, Figure 1.2. A large range of alkaloids have been exploited by humans, such as those named above. Cocaine, nicotine and caffeine are stimulants which are commonly abused. Atropine is a central and peripheral nervous system blocker used in several medical settings, such as bradycardia (slow heart rate) and as a treatment for poisoning by nerve agents and pesticides. Morphine and codeine are analgesics used to treat moderate to severe pain. Morphine can be chemically acetylated producing heroin which has been abused for many years generating a plethora of socio-economic problems. Noscapine has classically been utilised as an antitussive but also possesses weak antitumor properties.[19] Noscapine has no analgesic properties.

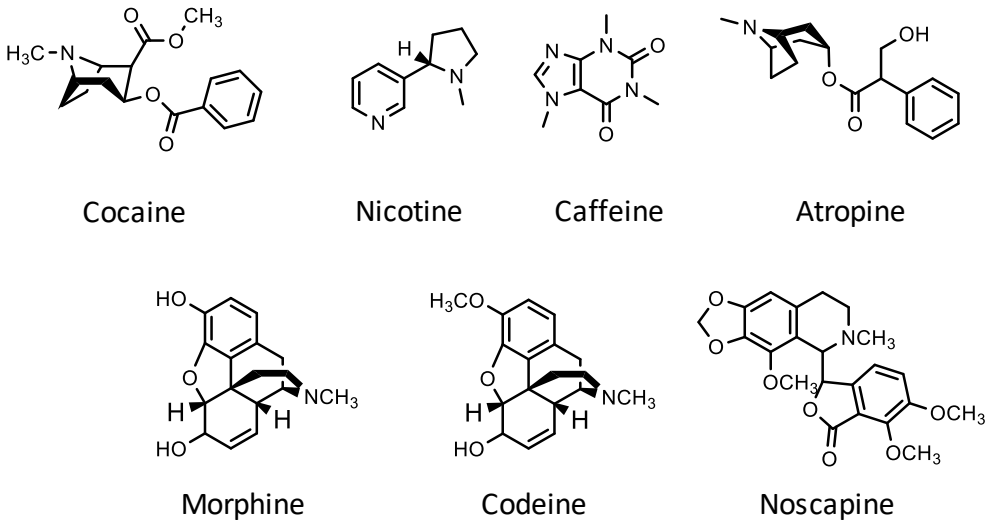



*Figure 1.2: Structures of common alkaloids*

### 1.3.1. Benzylisoquinoline Alkaloids

Alkaloids derived from opium poppy belong to the benzylisoquinoline class, Figure 1.4. The biosynthesis of benzylisoquinoline alkaloids (BIA) start with shikimic acid, produced in the shikimate biosynthetic pathway, which is subsequently converted to chorismic acid by sequential phosphorylation, alkylation, dephosphorylation and dehydration, Figure 1.3.[20] Chorismate mutase converts chorismic acid to prephenic acid which is then decarboxylated to phenylpyruvate.[21] L-phenylalanine is generated by transamination of phenylpyruvate by prealinine aminotransferase, which is subsequently hydroxylated at C4 to generate L-tyrosine by biopterin-dependent phenylalanine hydroxylase (PAH).

L-Tyrosine is the precursor of 4-hydroxyphenylacetaldehyde (4-HPAA) and dopamine (3,4-dihydroxyphenylethylamine) which are generally cited as the essential starting building blocks for BIAs. 4-Hydroxyphenylacetaldehyde is produced by deamination of L-tyrosine by the 2-oxoglutarate dependent tyrosine transaminase or by L-amino acid oxidase and subsequent decarboxylation of 4-hydroxyphenylpyruvate by 4-hydroxyphenylpyruvate decarboxylase.[22,23] Dopamine is produced by the hydroxylation of L-tyrosine followed by decarboxylation by tyrosine hydroxylase and aromatic amino acid decarboxylase, respectively.[23]

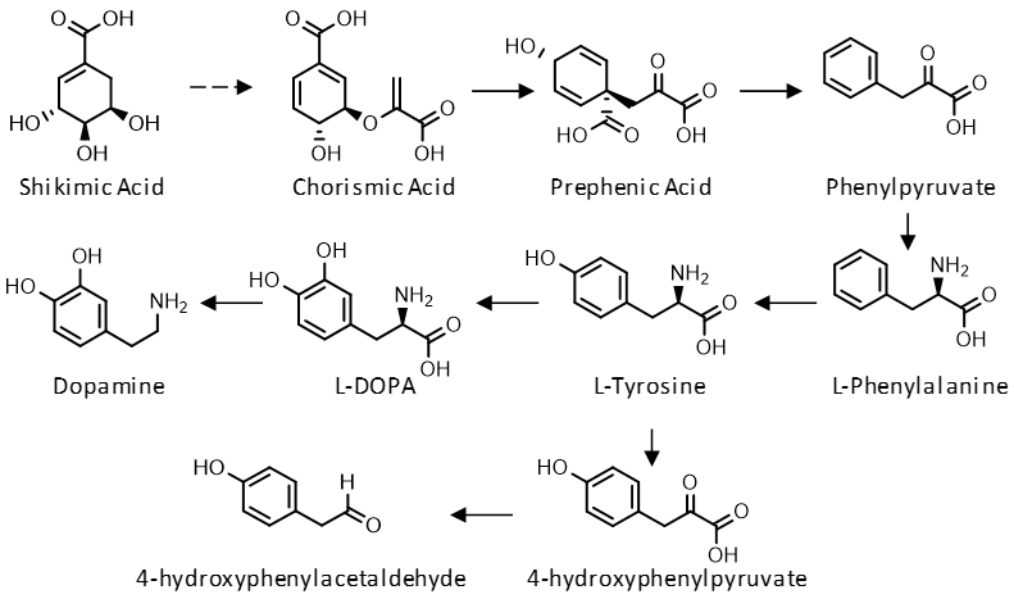



*Figure 1.3: Biosynthesis of dopamine and 4-hydroxyphenylacetaldehyde from Shikimic Acid*

The first committed biosynthetic step to BIAs is a condensation reaction between the tyrosine derivatives dopamine (red) and 4-hydroxyphenyacetaldehyde (blue) to produce (*S*)-norcoclaurine by norcoclaurine synthase (NCS), Figure 1.5.[24,25] Dopamine forms the tetrahydroisoquinoline moiety, whereas 4-HPAA forms the C1 linked benzyl component, together they form the 1-benzylisoquinoline core of all benzylisoquinoline alkaloids, Figure 1.4. Papaverine is the simplest BIA produced by opium poppy and is chemically synthesised for industrial use, whereas all other BIAs are extracted from plants.[26–28]

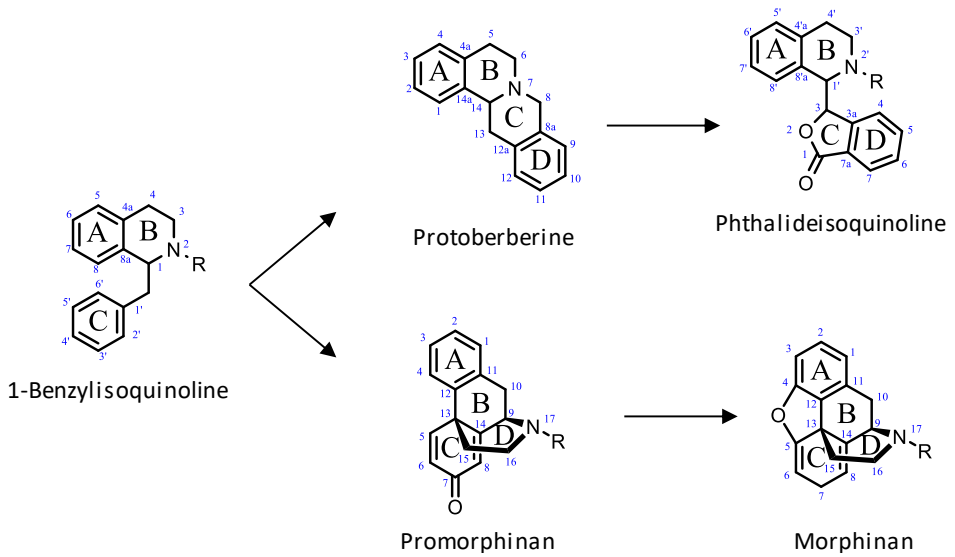



*Figure 1.4: Common benzylisoquinoline structural subgroups in opium poppy*

Several enzymatic reactions comprising of *O*-methylation, *N*-methylation and aromatic ring hydroxylation generates (S)-reticuline from (S)-norcoclaurine. (S)-reticuline, is regarded as the central intermediate from which the majority of the benzylisoquinoline alkaloids are derived, Figure 1.5.[18,28] From (S)-reticuline there are four main branch pathways in opium poppy, benzylisoquinoline alkaloid biosynthesis leading to the benzophenanthridine sanguinarine in roots, the morphinan and the phthalideisoquinoline noscapine branch pathways which are mostly found in aerial parts of the plants.[28,29] The enzymatic steps to noscapine are discussed in detail in Section 1.4.1, pg. 24.

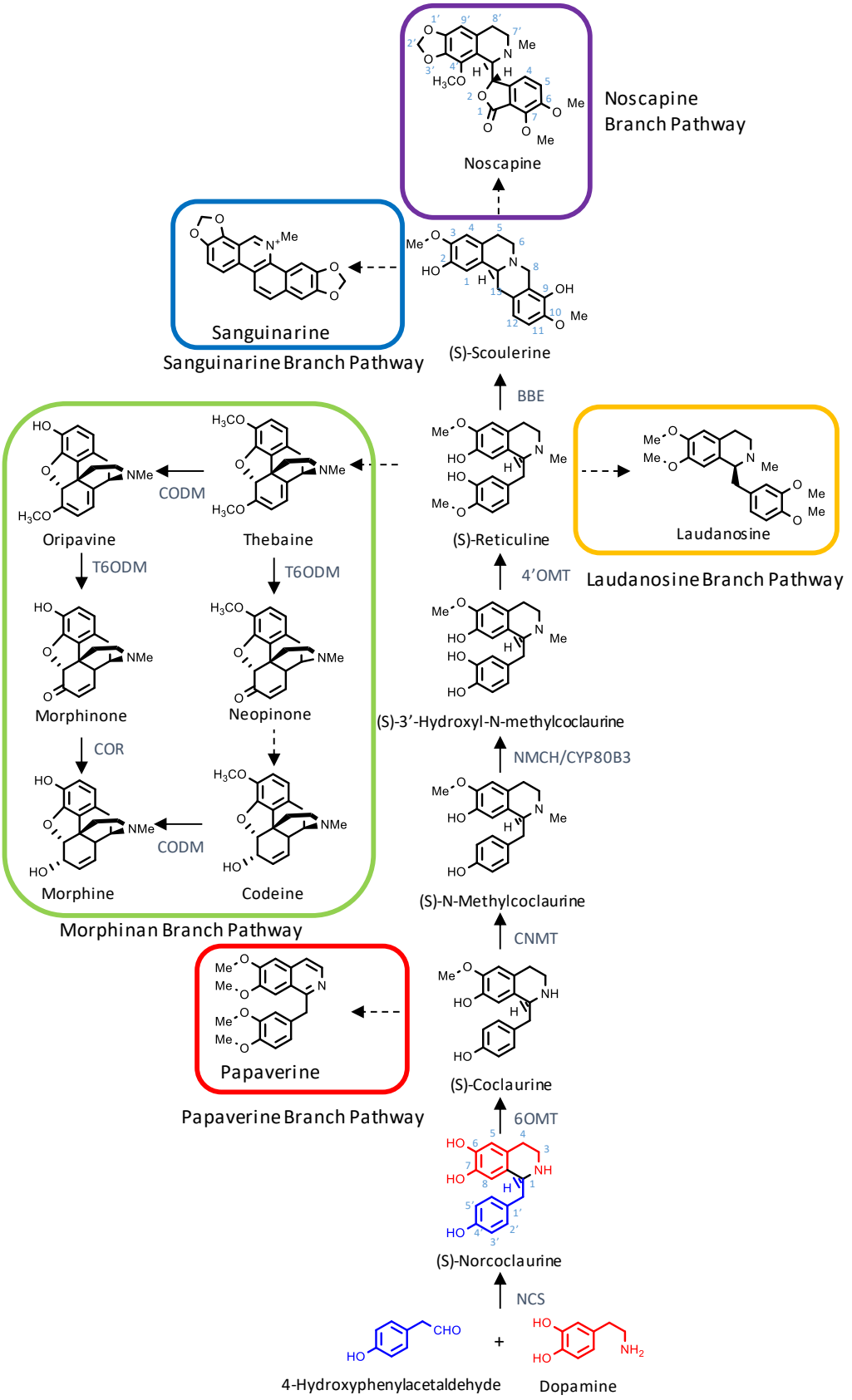



*Figure 1.5: Overview of alkaloid biosynthesis in opium poppy. 4-hydroxyphenyacetaldehyde and dopamine coloured to illustrate the formation of the tetrahydroisoquinoline moiety in (S)-norcoclaurine. Solid arrows represent a single enzymatic step while dashed arrows represent multiple enzymatic steps*

## 1.4. Noscapine

Noscapine is a phthalideisoquinoline alkaloid derived from opium poppy (*Papaver somniferum)* and was one of the first alkaloids to be isolated.[8,9] It has classically been utilised as a cough suppressant due to its antitussive activity, reported to be discovered in 1930 by review articles,[10,30] but the earliest scientific paper describing this activity is from 1954.[31] Unlike other alkaloids from opium poppy such as codeine and morphine, noscapine does not possess analgesic or sedative properties and is nonaddictive. It has been utilised as tablets and syrups to be taken orally and is still available to purchase over the counter in many countries.[16] Although the mode of action of noscapine's antitussive properties is not known, it has been shown to interact with the central nervous system of guinea pig by binding to the brain stereo-specifically with high affinity in a saturatable manner.[32]

Interest in noscapine as a drug has increased over the last 10-20 years due to its antitumor properties. In 1998 it was reported that noscapine arrests cells in metaphase and induces apoptosis in dividing cells.[19] It is postulated that noscapine binds to tubulin subunits, promoting tubulin assembly, which ultimately leads to cell death in several tumour cell lines and tumours in mice.[19,33] Noscapine is effective at treating a range of various cancers such as T-cell lymphoma, breast cancer, melanoma, ovarian carcinoma, glioblastoma, colon cancer and non-small cell lung cancer.[33–39]

Noscapine has been the subject of a variety of phase I and phase II clinical trials.[40,41] To increase the therapeutic properties of noscapine several derivatives have been synthesised, Figure 1.6. Several modifications have been tested for increased therapeutic properties including H9', N6' and O6 substitution along with O1 carbonyl reduction in conjunction with H9' substitution.[37,42–47]

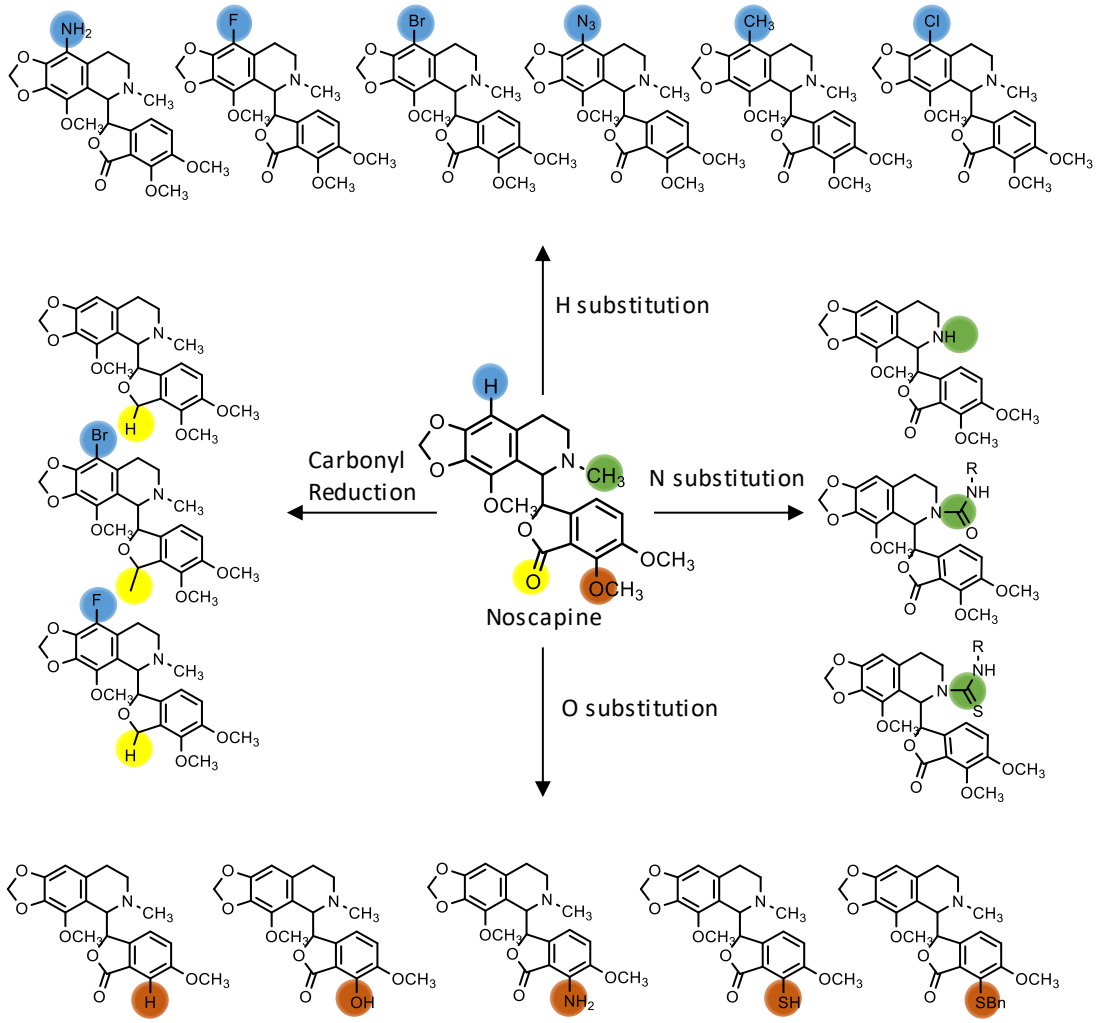



*Figure 1.6: Noscapine derivatives. Adapted from Chen et al. (2015)[10]*

### 1.4.1. Noscapine Biosynthesis

The current proposed biosynthetic pathway to noscapine is shown in Figure 1.7.[48] The first committed step to generate BIAs is the stereospecific condensation reaction of dopamine and 4-hydroxyphenyacetaldehyde (4-HPAA) by norcoclaurine synthase (NCS) generating (*S*)-norcoclaurine by a Pictet-Spengler-type reaction.[24,49–51] NCS has been proposed to have evolved from a Bet v 1 or PR10 ancestor. Bet v 1 is a major allergen protein family which has 28-38 % sequence identity and PR10 is a pathogenesis-related (PR) 10 protein family with 50-60 % sequence identity to NSC.[24] The Pictet-Spengler-type reaction mechanism based on the crystal structure involves the formation of an iminium ion by a nucleophilic attack of the dopamine amine group to the aldehyde carbonyl of 4-HPAA followed by release of water. The C5 of

the catechol moiety of dopamine undergoes electrophilic attack by the imine-carbon, resulting in ring closure followed by the deprotonation to form the product, (*S*)-norcoclaurine.[50,52] (*S*)-norcoclaurine contains the core 1-benzylisoquinoline structure that all BIAs possess.

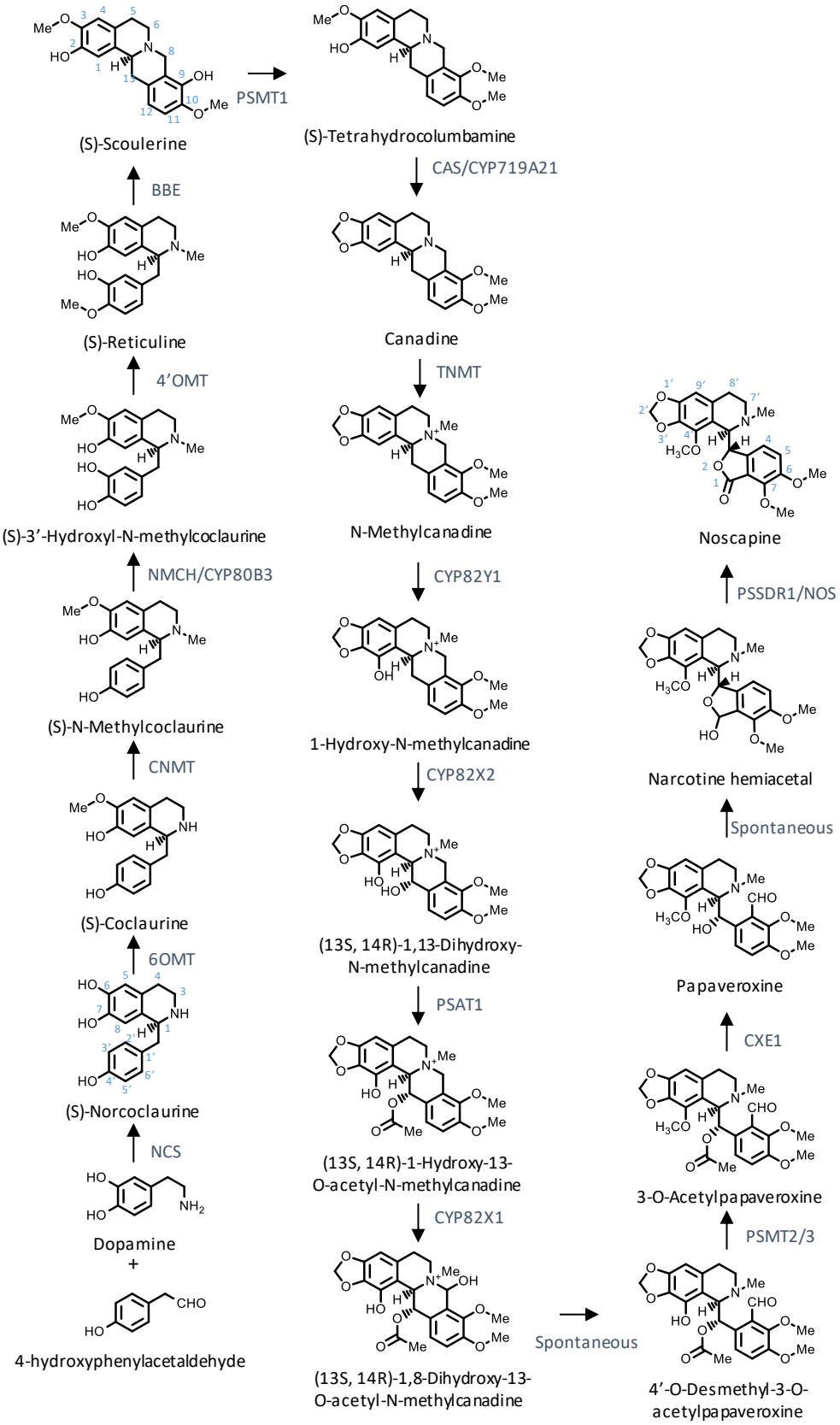



*Figure 1.7: Current proposed biosynthetic pathway to noscapine*

(*S*)-Norcoclaurine is then methylated at the C6 hydroxyl position to produce (*S*)-coclaurine by the dimeric class I *S*-adenosylmethionine (SAM) dependent *O*-methyltransferase norcoclaurine 6-*O*-methyltransferase (6OMT).[53,54] Methylation is proposed to occur by histidine acting as a catalytic base, deprotonating the hydroxyl group at position C6 of (*S*)-norcoclaurine followed by the phenolate anion performing a nucleophilic attack on the methyl-sulphonium moiety of SAM, generating (*S*)-coclaurine.[55] (*S*)-Coclaurine is the first branchpoint intermediate with the biosynthetic pathway to laudanosine branching off here, whereas the rest of the benzylisoquinoline alkaloids continue along the common pathway to (*S*)-reticuline.

(*S*)-Coclaurine is then *N*-methylated by the SAM dependent *N*-methyltransferase coclaurine *N*-methyltransferase (CNMT) generating (*S*)-*N*-methylcoclaurine.[56] This is one of two *N*-methylations involved in the biosynthesis of noscapine and is an essential modification for the cyclisation of the 1-benzylisoquinoline sub-structure to that of the protoberberines. *N*-methylation reaction mechanism is proposed to be similar to that of 6OMT as they both belong to the class I methyltransferase family where a catalytic base deprotonates the secondary amine followed by nucleophilic attack on the methyl-sulphonium ion.[57]

(*S*)-*N*-*M*ethylcoclaurine is then hydroxylated at the C3' position by *N*-methylcoclaurine 3'-hydroxylase (NMCH), a cytochrome P450 dependent monooxygenase generating 3'-hydroxy-*N*-methyl-(*S*)-coclaurine.[58] NMCH was first discovered in California poppy (*Eschscholzia californica*) and classified as CYP80B1. A putative ortholog was subsequently discovered in opium poppy and classified as CYP80B3.[58,59] Cytochrome P450 dependent monooxygenases require heme, iron and molecular oxygen as cofactors for catalysis. The consensus reaction mechanism is that the substrate binds and a cytochrome P450 reductase reduces iron(III) to iron(II) allowing molecular oxygen to bind. A second electron transfer occurs, reducing the Fe(II)-superoxide adduct generating a peroxide intermediate. Water is then released by the double protonation of the peroxide intermediate forming a highly reactive iron (V) oxo species, known as Compound 1, which reacts with the substrate to produce the hydroxylated product.[60]

3'-Hydroxy-*N*-methyl-(*S*)-coclaurine undergoes *O*-methylation of the hydroxyl group at the C4' position by the SAM dependent 3'-hydroxy-*N*-methylcoclaurine 4'-*O*-methyltransferase (4'OMT), to generate the central branch metabolite (*S*)-reticuline.[54,61]

(*S*)-Reticuline is the precursor to several different branch pathways in opium poppy, as outlined in Figure 1.5. From (*S*)-reticuline the laudanosine, sanguinarine and noscapine branch pathway emerge, along with the morphinan branch pathway which produces morphine and codeine *via* thebaine.[62] The first committed step to morphine is the isomerisation of (*S*)-reticuline to (*R*)-reticuline by STORR ((*S*)- to (*R*)-reticuline), a cytochrome P450 and oxidoreductase fusion protein.[63,64] Morphine and codeine are potent narcotics with analgesic effects, whereas thebaine is poisonous to humans. Although thebaine is not utilised as a drug, it is the starting chemical for several semi-synthetic drugs such as oxycodone and oxymorphone.[65,66]

Production of sanguinarine and noscapine involves conversion of (*S*)-reticuline to (*S*)-scoulerine by the flavin adenine dinucleotide (FAD)-linked oxidoreductase berberine bridge enzyme (BBE), also known as (*S*)-scoulerine oxidase.[67,68] BBE is responsible for the cyclisation reaction between the N2 methyl group of the tetrahydroisoquinoline moiety of (*S*)-reticuline and the C2 of its C1 linked benzyl moiety forming the C ring of the protoberberine sub-structure, Figure 1.4.

BBE catalyses a unique reaction that has so far remained elusive to organic chemists due to its challenging mechanism. The proposed reaction mechanism is the proton abstraction of the C3' hydroxyl group by a conserved base, resulting in an $S_N2$-type nucleophilic attack on the *N*-methyl group by C2'. This results in formation of the berberine bridge covalent bond and the expulsion of a hydride from the *N*-methyl group to the FAD cofactor. Rearomatisation of the protoberberine D ring occurs by a secondary base deprotonation at C8a resulting in the formation of (*S*)-scoulerine and the protoberberine sub-structure.[69] The FAD cofactor is ultimately regenerated by reoxidation utilising molecular oxygen to generate hydrogen peroxide.

(*S*)-Scoulerine is *O*-methylated at the C9 hydroxyl position producing (*S*)-tetrahydrocolumbamine by the class I SAM dependent *O*-methyltransferase

called *Papaver somniferum O*-methyltransferase 1 (PSMT1) also known as scoulerine *O*-methyltransferase (SOMT).[48,70,71]

(*S*)-Tetrahydrocolumbamine is converted to (*S*)-canadine by canadine synthase (CAS/CanSyn), also known as the cytochrome P450 CYP719A21.[72,73] CanSyn is responsible for the methylenedioxy bridge formation on the A ring of (*S*)-tetrahydrocolumbamine. The proposed reaction mechanism proceeds by the formation of Compound 1, as previously described, which abstracts a proton from the C3 methoxy group on the A ring of (*S*)-tetrahydrocolumbamine, forming an Fe(IV)-hydroxyl species and a reactive radical carbon intermediate. The reactive radical carbon intermediate is involved in a hydroxyl rebound mechanism, where the Fe(IV)-OH hydroxyl is inserted into the reactive carbon intermediate forming a hemiacetal group. Subsequent dehydration of the hemiacetal group leads to the formation of an oxonium ion. Base assisted proton abstraction on the C2 hydroxyl group leads to a nucleophilic attack on the oxonium ion resulting in the formation of the methylenedioxy bridge.[74]

(*S*)-Canadine is *N*-methylated by the SAM dependent tetrahydroprotoberberine cis-*N*-methyltransferase (TNMT) to generate *N*-methylcanadine.[75] *N*-methylcanadine is hydroxylated sequentially at the C1 position and then C13 by *N*-methylcanadine 1-hydroxylase (CYP82Y1) and 1-hydroxy-*N*-methylcanadine 13-*O*-hydroxylase (CYP82X2), respectively, to make 1-hydroxy-*N*-methylcanadine and then (13*S*, 14*R*)-1,13-dihydroxy-*N*-methylcanadine.[48,76,77] The hydroxyl group at C13 opens up (13*S*, 14*R*)-1,13-dihydroxy-*N*-methylcanadine to be acetylated by 1,13-dihydroxy-*N*-methylcandine 13-*O*-acetyltransferase (PSAT1).[77] Acetylation acts as a protecting group, such as those used in organic synthesis, postponing the formation of the hemiacetal ring leading to the phthalideisoquinoline sub-structure, but may also play a role in the transport of pathway intermediates.[77]

The resulting compound, (13*S*, 14*R*)-1-hydroxy-13-*O*-acetyl-*N*-methylcanadine, is hydroxylated at the C8 position by 1-hydroxy-13-*O*-acetyl-*N*-methylcanadine 8-hydroxylase (CYP82X1) producing (13*S*, 14*R*)-1,8-dihydroxy-13-*O*-acetyl-*N*-methylcanadine.[48,77] (13*S*, 14*R*)-1,8-dihydroxy-13-*O*-acetyl-*N*-methylcanadine

spontaneously converts to 4'-*O*-desmethyl-3-*O*-acetylpapaveroxine with the cleavage of the N7-C8 bond.

4'-*O*-Desmethyl-3-*O*-acetylpapaveroxine is subsequently *O*-methylated by the unique heterodimer *Papaver somniferum O*-methyltransferase 2 (PSMT2)/ *Papaver somniferum O*-methyltransferase 3 (PSMT3) producing 3-*O*-acetylpapaveroxine.[48,71,78] Interestingly a heterodimer of PSMT2/6OMT can also catalyse the 4'-O methylation of 4'-*O*-desmethyl-3-*O*-acetylpapaveroxine, which explains why virus induced gene silencing (VIGS) of PSMT2 resulted in the accumulation of narcotoline (4'-*O*-desmetylnoscapine) but VIGS of PSMT3 had no altered phenotype.[48,78]

3-*O*-Acetylpapaveroxine is converted to papaveroxine by carboxylesterase 1 (CXE1) which removed the O3-acetate group.[48,77,78] This leads to the spontaneous rearrangement to narcotine hemiacetal, where the C3 hydroxyl acts as a nucleophile to attack the C4 aldehyde group to form the phthalideisoquinoline sub-group hemiacetal moiety. The final biosynthetic step to noscapine is the reduction of the C1 hydroxyl group of narcotine hemiacetal to a carbonyl group by *Papaver somniferum* short-chain dehydrogenase/reductase (PSSDR1).[48,79]

The noscapine branch pathway as outlined in Figure 1.5 and described above is largely encoded for by the noscapine gene cluster, Figure 1.8.[48] The noscapine gene cluster contains 10 genes responsible for all the enzyme catalysed biosynthetic steps from (S)-scoulerine to noscapine, except for that of tetrahydroprotoberberine cis-*N*-methyltransferase (TNMT). TNMT is also involved in the sanguinarine biosynthetic pathway. Six out of the ten enzymes were characterised by virus induced gene silencing and heterologous expression in yeast.[48] Subsequent research characterised the remaining four enzymes, leading to the elucidation of PSMT2/PSMT3 heterodimer.[76–78] The identification of the gene cluster is an excellent example of gene duplication, neofunctionalisation and genome reorganisation. The clustering of genes is assumed to offer an evolutionary advantage as the gene cluster is likely to be inherited in its entirety, ensuring all the catalytic machinery is present for the biosynthesis of noscapine.

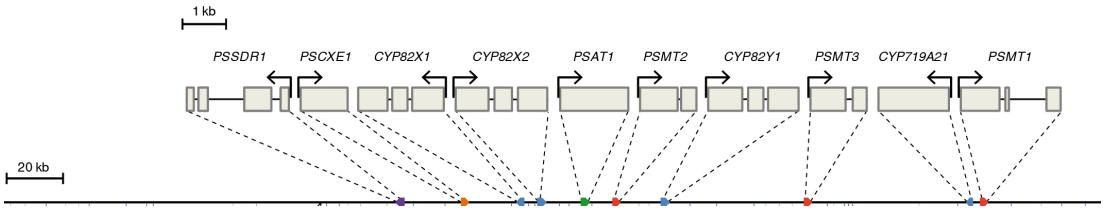



*Figure 1.8: The noscapine gene cluster in opium poppy. The intron exon structure and the positions of the ten noscapine specific genes are shown above the black line which represents 401 Kb of genomic sequence. Exons are represented by grey boxes and introns by black lines. The arrows represent the 5' to 3' orientation of each gene, adapted from Winzer et al., 2012*[48]

## 1.5. Methyltransferases

There are several distinct methods nature has implemented to methylate substrates, all utilising a co-substrate with a labile methyl group. The most common of methyltransferase are that of the *S*-adenosylmethionine (SAM) dependent methyltransferase class.[80] This is not surprising as SAM has the most favourable energetics due to the charged methyl-sulphonium centre and is said to be the second most commonly used co-substrate by enzymes, with the first being ATP.[80,81] Less common but arguably just as important are the 5-methyltetrahydrofolate and the betaine dependent methyltransferases. All three types of enzyme consume the co-substrate/methyl donor to methylate a substrate, Figure 1.9. Methyltransferases methylate a wide variety of substrates from small molecules in primary and secondary metabolism to DNA and proteins, and methylation can be on carbon, oxygen, nitrogen, sulphur and halides.[71,80,82–86]

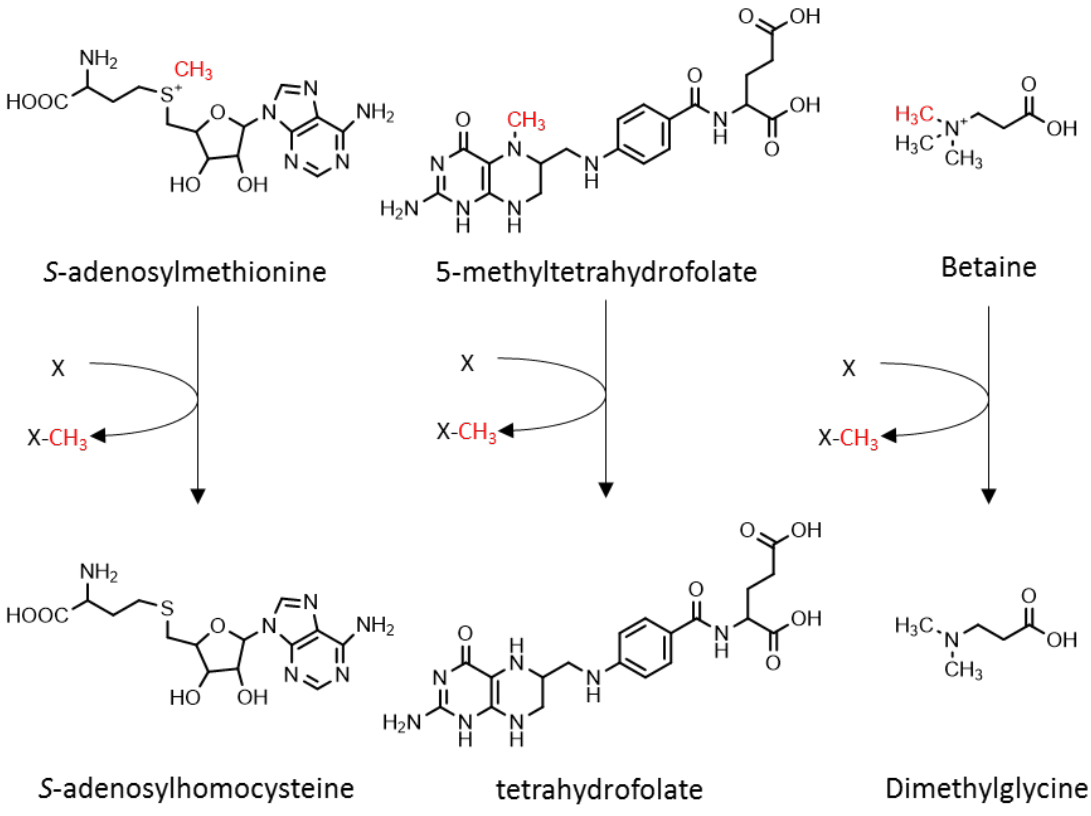



*Figure 1.9: Natural methyl donors. Labile methyl groups are coloured red*

This thesis will focus on the SAM dependent methyltransferase and only briefly mention the other types due to their role in SAM regeneration.

## 1.6. *S*-Adenosylmethionine Dependent Methyltransferases

The first SAM dependent methyltransferase to have its structure solved by X-ray crystallography was that of DNA C5-cytosine methyltransferase M.HhaI complexed with SAM to 2.5 Å in 1993 (PDB:1HMY).[86] There are five classes of SAM dependent methyltransferases which are all structurally diverse which have been numbered I-V based on the chronological order of discovery. This is a great example of functional convergence. Enzymes which catalyse the same reaction but from distinct structural families have been noted before and have been named analogous enzymes, Figure 1.10.[80,82,87]

Typically, enzymes that catalyse the same reaction have the same structural protein fold and evolutionarily related but can vary substantially in amino acid sequence, these are termed homologues. Homologous enzymes have a common ancestor and by analysing their protein sequences conserved residues can be identified. The core catalytic residues of proteins are generally highly conserved, along with residues involved in stabilisation of core folds. Homologous enzymes can be catalytically promiscuous accepting different substrates due to gene duplication and neofunctionalisation.[88–90] This can lead to new catalytic properties arising within the same structural fold.[91] In contrast, for the case of the SAM dependent methyltransferases convergent evolution has taken place where at least five distinct structural folds have evolved to catalyse the same type of reaction utilising the same co-substrate, SAM.[80]

Class I methyltransferases contain a Rossmann-like fold which consists of a seven stranded β-sheet typically with the topology 6↑7↓5↑4↑1↑2↑3↑, with 1 being the N-terminus.[80] In between β-strand 1 and 2 is an α-helix which is classed as the core β-α-β sandwich fold, this is the most conserved motif within the class I methyltansferases.[92] Typically, there is a Gly-X-Gly-X-Gly motif in the tight loop between the first β-strand and the α-helix. The final conserved feature of the superfold is an acidic residue at the end of β-strand 2 which forms hydrogen bonds with the hydroxyl groups of the ribose ring of the nucleotide. The class I methyltransferase family contains mostly monomeric enzymes made up solely of the

Rossmann-like fold domain but homodimeric, heterodimeric and tetrameric forms have also been reported with additional domains.[93]

Class II methyltransferases, also known as MetH reactivation domain, reactivate the vitamin $B_{12}$ cofactor in *E. coli* cobalamin (vitamin $B_{12}$)-dependent methionine synthase (MetH) by reductive methylation of the oxidised cobalt ($Co^{2+}$ (inactive)→ $Co^{+}$(active)).[94] Class II methyltransferases consist of a central anti-parallel β-sheet surrounded by α-helices in a C-like shape, where SAM binds in an extended conformation along the concaved region in the centre of the C-like shape forming hydrogen bond interactions with the conserved RxxxGY motif.

Class III methyltransferase structure was discovered in a homodimerc protein called cobalt-precorrin-4 methyltransferase (CbiF) from *Bacillus megaterium.*[95] This structural family consists of two α-β-α domains each containing five β-strands and four α-helices, with SAM binding within a cleft between the two domains.

Class IV methyltransferases contain the SPOUT family of RNA methytransferases.[96] This structural family consists of a knot-like structure containing a β-sheet made up of six parallel β-strands in order 321546 surrounded by seven α-helices, in an α-β-α fold.[97] The N-terminal domain is representative of half a Rossman-like fold, such as in the class I methyltransferases, but differ at the C-terminal domain with a α-helix looping back into the active site, generating a rare deep trefoil knot.[98] SAM has been shown to bind to β-strands 4-6 in a bent conformation with the conserved NxGxxxR and SxN motifs. This class of enzyme has been shown to be a dimer in solution.[97]

Class V methyltransferases are the SET domain proteins that have been shown to methylate lysines in the N-terminus of histones and ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO).[99,100] This structural family consists of three curved β-sheets made up from eight β-strands with a C-terminal α-helix knot-like structure. The binding site for SAM is located in a cleft within the SET domain. Surrounding the SET domain are pre-SET and post-SET regions that are thought to play a role in substrate specificity.

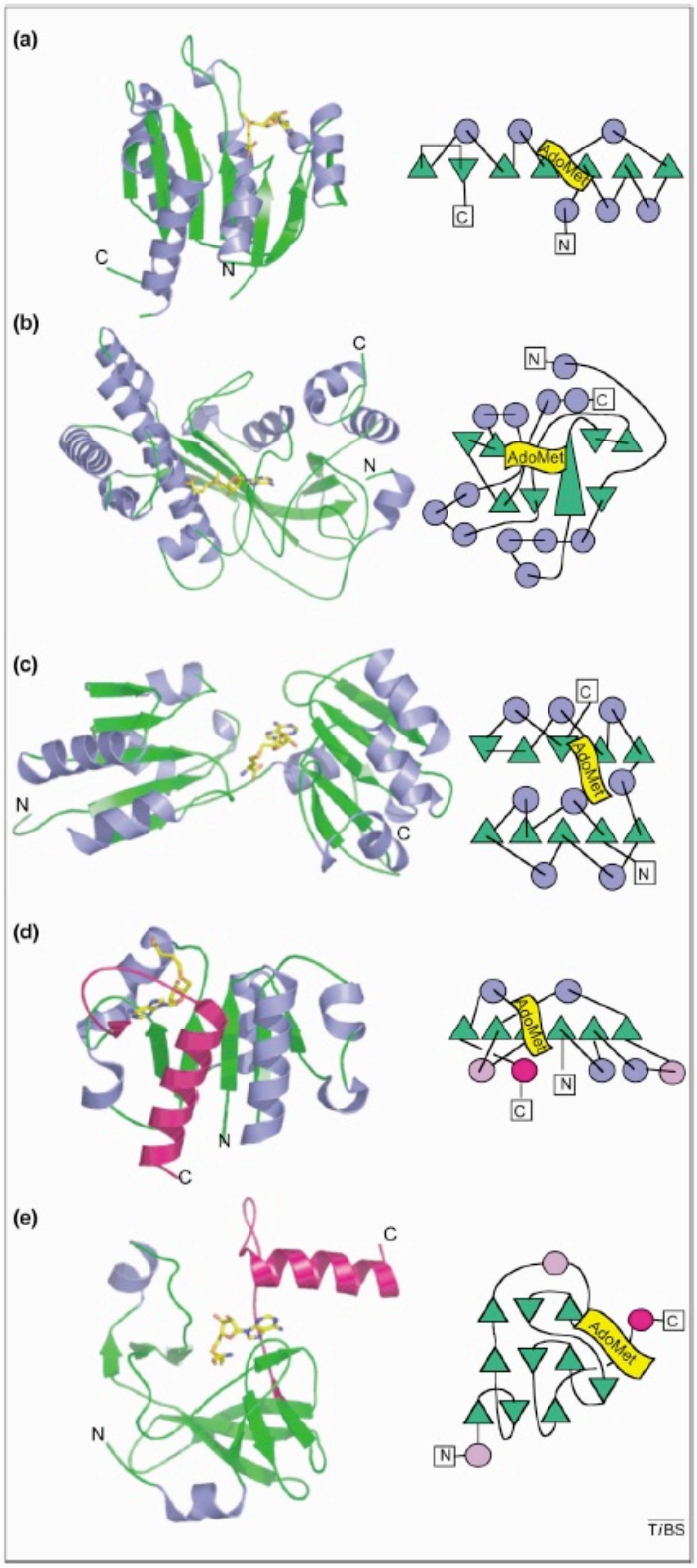



*Figure 1.10: The tertiary structure of S*-adenosylmethionine dependent methyltransferase classes I-V, left, along with their corresponding topology diagrams, right. a-class I, b-class II, c-class III, d-class IV, d-class V. Figure taken from Schubert et al. 2003[80]

### 1.6.1. *S*-Adenosylmethionine Metabolism

The consumption of SAM by SAM dependent methyltransferases generates *S*-adenosylhomocysteine (SAH), Figure 1.11.[101] SAH is a potent inhibitor of SAM dependent methyltransferases, therefore it is quickly hydrolysed by adenosylhomocysteine hydrolase to adenosine and homocysteine. To regenerate SAM, homocysteine is methylated by methionine synthase which utilises 5-methyltetrahydofolate as a methyl donor or betaine homocysteine methyltransferase which utilises betaine as the methyl donor to generate methionine. The final step to regenerate SAM is the adenylation of methionine by methionine adenosyltransferase utilising ATP.[102,103] The synthesis of SAM from methionine is responsible for around 80 % of methionine metabolism, whereas the remaining 20 % is utilised for the synthesis of proteins.[104]

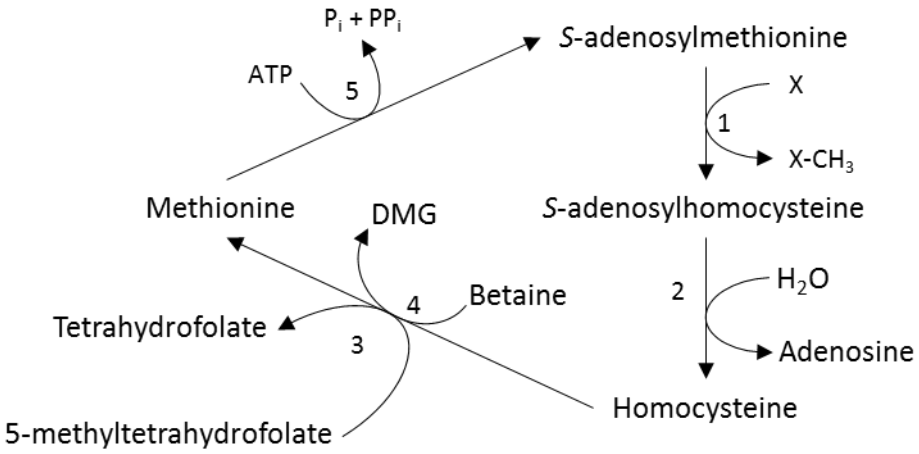



*Figure 1.11:* S*-Adenosylmethionine metabolism. 1-SAM dependent methyltransferase, 2-adenosylhomocysteine hydrolase, 3-methionine synthase, 4-betaine homocysteine methyltransferase, 5-methionine adenosyltransferase*

## 1.7. Plant O-Methyltransferases

Plant *O*-methyltransferases are a large enzymatic family, many of which are involved in lignin biosynthesis, while others methylate a large range of low molecular weight plant secondary metabolites such as alkaloids, flavonoids, phenylpropanoids and terpenoids, Figure 1.12. All known plant substrates are phenolic in nature with most having adjacent hydroxyl/catecholic groups.

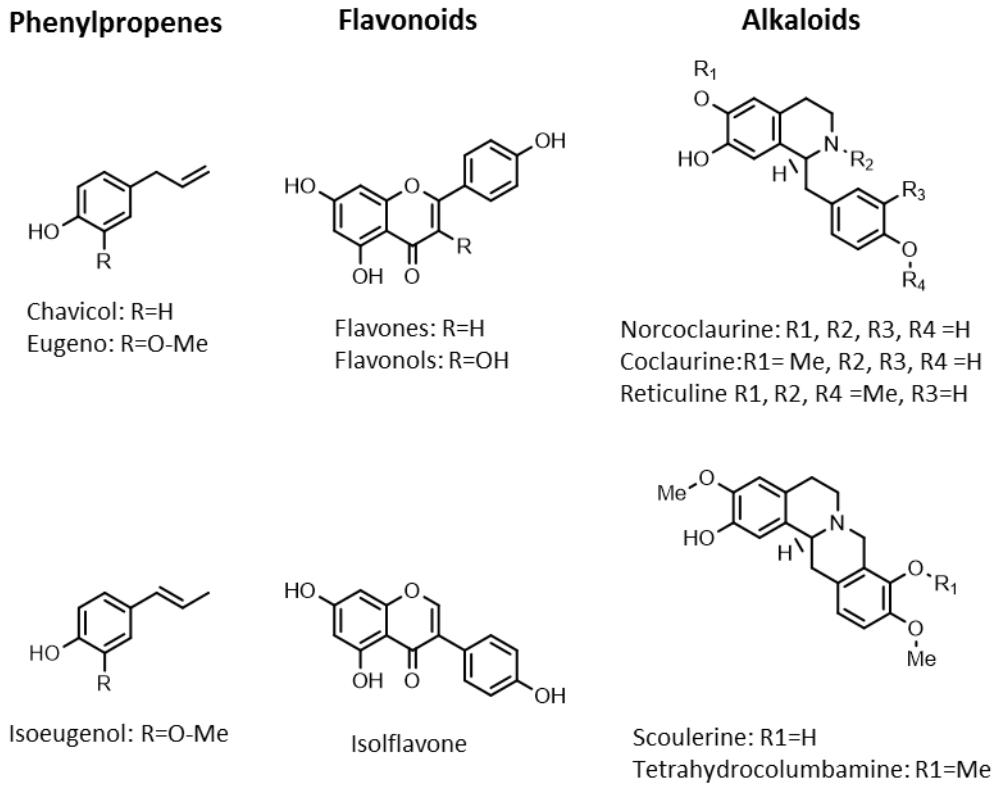



*Figure 1.12: Plant secondary metabolite chemical structures utilised by O*-methyltransferases

Plant *O*-methyltransferases are part of the class I methyltransferases family which contain the Rossmann-like superfold, Figure 1.10.[82] The first plant *O*-methyltransferase to be successfully cloned was that of caffeoyl-CoA 3-*O*-methyltransferase from parsley involved in disease resistance response.[105] Recent advantages in large-scale genomic sequencing has resulted in hundreds of plant *O*-methyltransferases available in public databases but relatively little is known about their substrate specificity.

### 1.7.1. Plant *O*-Methyltransferase Structure

The first plant *O*-methyltransferases to have their structures solved were that of chalcone *O*-methyltransferase and isoflavone *O*-methyltransferase from *Madicago sativa* by X-ray crystallography in 2001.[106] Since then the structure of several other homodimeric *O*-methyltransferases have been solved, including several caffeic acid *O*-methyltransferases, monolignol 4-*O*-methyltransferase, isoflavanone 4'-*O*-methyltransferase, coniferyl alcohol 9-*O*-methyltransferase and norcoclaurine 6-*O*-methyltransferase, Figure 1.13.[55,107–112] The majority of characterised plant

*O*-methyltransferases are homodimers but data also shows scope for heterodimeric enzymes with altered substrate specificity.[59,78,113]

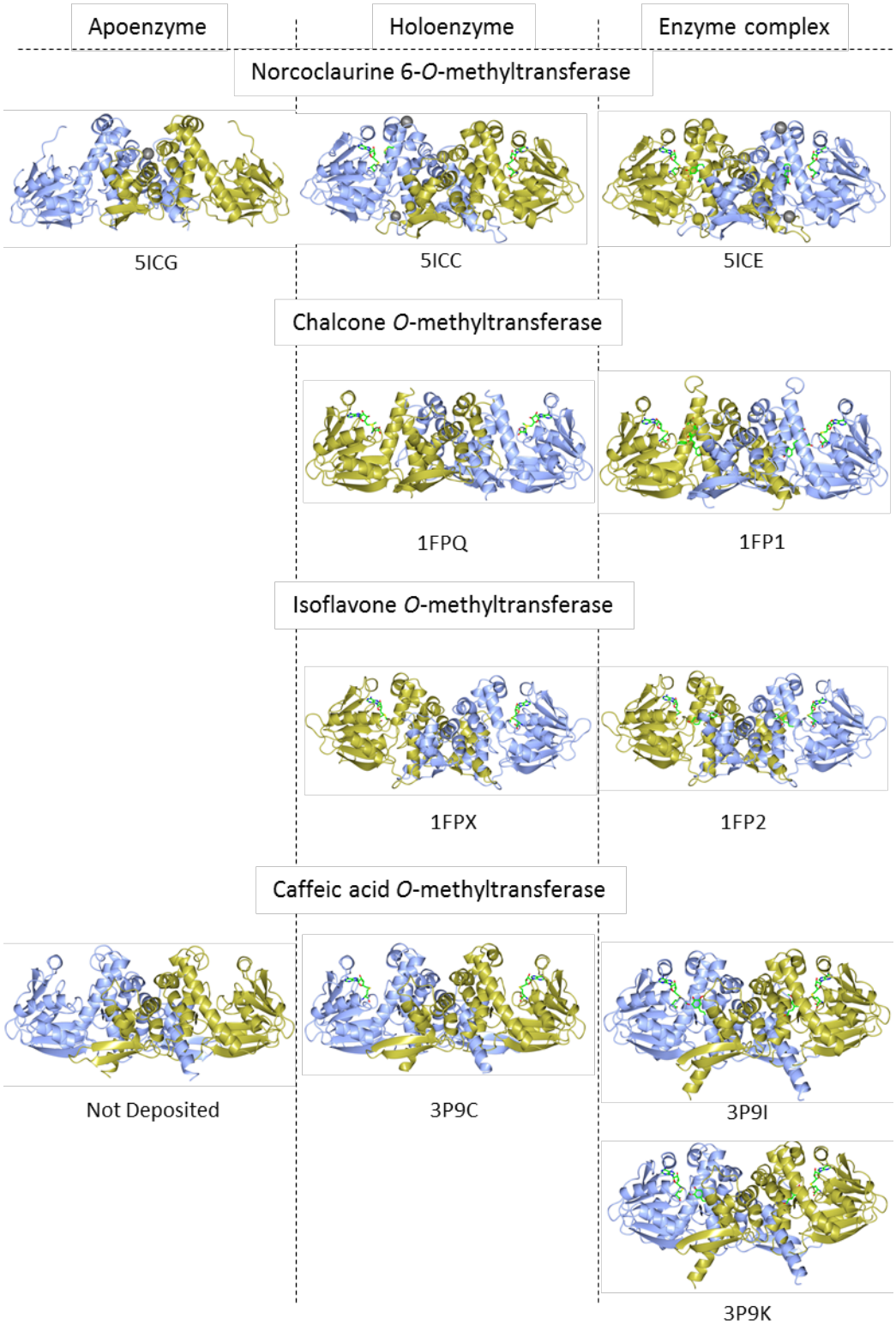



*Figure 1.13: Crystal structures of several plant O-methyltransferases. Peptides are shown as ribbons, coloured by chain. The ligands are shown as cylinders coloured by atom: carbon (green), oxygen (red), nitrogen (blue), sulphur (yellow).PDB codes listed where available*[55,107–112]

All current available structures are of homodimeric enzymes and have the same overall fold with a N-terminal dimerisation/substrate binding domain along with a conserved C-terminal SAM binding domain, Figure 1.13. They have been shown to have a large conformational change upon binding of the substrates but the causation of active site closure is unclear.

The plant *O*-methyltransferases which have their structures solved are in either a open or a closed conformation. All the "apo" structures are in an open conformation, the majority of "holo" (SAH bound) structures are in an open conformation and the majority of the complex (SAH and substrate/product bound) structures are in a closed conformation but there are exceptions.[55,106,112] For example, norcoclaurine 6-*O*-methyltransferase with SAH bound is in a closed conformation and chalcone *O*-methyltransferase with SAH and isoliquiritigenin bound is in an open conformation.[55] Probing causation of conformational changes by X-ray crystallography is difficult due to crystal growth selective pressures stabilising alternative conformations.

The dimeric structure opens up the possibility of subunit co-operativity which has been reported for *Thalictrum tuberosum O*-methyltransferases and *Sorghum bicolor* coffeoyl-CaA *O*-methyltransferase.[113,114] Subunit co-operativity can occur when an enzyme contains two or more active sites. For positive co-operativity, binding of a substrate at one active site increases the affinity at a second site. Whereas, for negative co-operativity the opposite is observed, where binding of a substrate at one site decreases the affinity at a second site. The most common method to measure co-operativity is by fitting the Hill equation to activity assay data to generate a Hill coefficient.[115] The Hill coefficient is described as 'an interaction coefficient, reflecting the extent of cooperativity among multiple ligand binding sites'. [115]

### 1.7.2. Reaction Mechanism

Several reaction mechanisms have been reported in the literature for plant *O*-methyltransferases. *Sorghum bicolor* caffeic acid *O*-methyltransferase has been reported to proceed in a random order bi-bi mechanism; where the substrates, caffeic acid and SAM, bind in a random order, catalysis occurs, and the products are

released in a random order.[111] Lineweaver-Burk plots were generated comparing activity at various concentration of methyl acceptor (caffeic acid) in respect to SAM and various concentrations of SAM in respect to methyl acceptor. Both graphs show an intercept on the x-axis indicating sequential order reaction kinetics. Product inhibition studies showed SAH was a non-competitive inhibitor towards caffeic acid but was competitive towards SAM, and the methylated product (ferulic acid) was a competitive inhibitor to both SAM and caffeic acid. This data led the authors to propose rapid equilibrium random order mechanism where caffeic acid and SAH lead to a dead end complex.

Whereas, for *Linum nodiflorum* coniferyl alcohol 9-*O*-methyltransferase it has been proposed that the methyl acceptor (coniferyl alcohol) binds first followed by SAM in a sequential order mechanism based on dissociation constants ($K_d$).[110] The dissociation constants measured by isothermal calorimetry showed that the methyl acceptor $K_d$ value lowered by a factor of 1.8 in the presence of SAM compared to without SAM present. Whereas the $K_d$ for SAM decreased by a factor of 222 in the presence of methyl acceptor compared to without.

Hydroxy-*N*-methylcoclaurine 4'-*O*-methyltransferase from *Coptis japonica* has been reported to proceed by an ordered mechanism based on product inhibition studies.[54] The inhibition studies utilising SAH as the inbibitor showed non-competitive inhibition to the methyl acceptor (norlaudanosoline) but showed competitive inhibition towards SAM. Whereas, methyl acceptor product (norreticuline) inhibition showed non-competitive inhibition in respect to both methyl acceptor and SAM.

The conflicting reaction mechanisms proposed for this class of enzyme may be due to the large conformational change upon binding on the substrates, and is most likely a random order mechanism where upon both SAM and the methyl acceptor binding the active site closes, resulting in the large decrease in $K_d$ seen by Walters et al.[110] Product inhibition studies to determine the reaction mechanism is a not an appropriate technique for this class of enzyme as the methylated product will interfere with SAM binding due to steric clashes of the methyl groups causing an

artificial competitive result. A better alternative would be to utilise a substrate analogue which can bind to the substrate pocket but cannot be methylated.

### 1.7.3. *S*-Adenosylmethionine Binding Site

Kozbial and Mushegian proposed five conserved motifs involved in SAM binding of Rossmann-like fold methyltransferases based on multiple sequence alignment and structural data. [116] Motif I consists of the first β-strand, the following loop and the first α-helix. β-strand one contains a conserved acidic residue (D or E) in the middle of the stand along with one or more positively charged residues at its N-terminus which do not interact with the substrate. There is also a conserved GxGxG sequence in the loop of motif I which forms a hydrophobic wall that interacts directly with the carboxypropyl moiety of SAM.

Motif II consists of β-strand two and the following loop with a conserved acidic residue at the C-terminal end of β-strand 2 which hydrogen bonds to the hydroxyl groups on the ribose ring. The loop following interacts with the adenine ring. Motif III is β-strand three and the following loop which contains a conserved acidic residue at the C-terminus that interacts with adenine. Motif IV is β-strand four and the following loops with a conserved D/E/N residue at the N-terminus of the loops which is not involved in substrate binding. The variable C-terminal region of motif IV interacts with the sulphonium moiety of SAM. Motif V is the α-helix following β-strand four and contains hydrophobic or aromatic residues which stabiles the adenine moiety of SAM. Motif VI is that of β-strand five and the preceding tight loop containing a conserved glycine residue with unknown function.

These motifs are conserved in plant *O*-methyltransferases along with several other conserved features based on this authors analysis of the crystal structures shown in Figure 1.13. SAM is bound by a conserved lysine which interacts extensively with the amino acid moiety of SAM, forming hydrogen bonds with the carboxylic acid of SAM *via* its terminal amine group and the amine group of SAM *via* its backbone carbonyl moiety, Figure 1.14. π Orbital hydrogen bond interactions between the aromatic ring of adenine and that of the sulphur ($S_\delta$) in methionine and on the opposing face to the second carbon ($C_\gamma$) of the alkane sidechain of a conserved leucine except for

isoflavone *O*-methyltransferase which has an arginine at this position, but the same interaction is maintained.

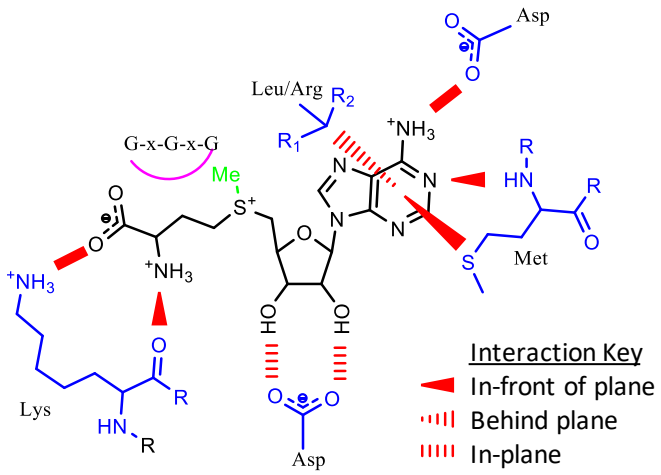



*Figure 1.14: A summary of conserved features involved in SAM binding based on chalcone-,[106] isoflavone-,[106] caffeic acid-[107] and norcoclaurine 6-[55] O-methyltransferases. SAM is shown in black with the labile methyl group in green, non-covalent bonding residues are shown in blue, non-covalent bonds are shown in red and hydrophobic interactions are shown in pink*

### 1.7.4. Phenolic Substrate Binding Site

Only a few plant *O*-methyltransferases have had their three dimensional structure determined, with even less in an closed ternary complex.[55,106,110,112] These are *Lolium perenne* caffeic acid *O*-methyltransferase (COMT) (PDB: 3P9I, 3P9K)[112], *Clarkia breweri* monolignol *O*-methyltransferase (MOMT) (PDB: 5CVJ, 5CVV), *Thalictrum flavum* norcoclaurine 6-*O*-methyltransferase (N6OMT) (PDB: 5ICE),[55] *Medicago sativa* Chalcone *O*-methyltransferase (ChOMT) (PDB: 1FP1),[106] *Medicago sativa* Isoflavone *O*-methyltransferase (IOMT) (PDB: 1FP2),[106] *Linum nodiflorum* coniferyl alcohol 9-*O*-methyltransferase (Ca9OMT)(PDB: 4EVI, 4E70).[110] ChOMT is the only structure with its natural methyl acceptor bound in the active site but is in such a conformation that catalysis cannot occur; as a water molecule is coordinated to the catalytic dyad. These six proteins all possess the characteristic C-terminal Rossman-like fold SAM binding domain along with the N-terminal dimerisation and substrate binding domain. The residues responsible for SAM binding in the C-terminal domain are well conserved as highlighted in the multiple sequence alignment, Figure 1.15, and discussed in Section 1.7.3, pg. 41. Ca9OMT implements a different catalytic

mechanism with a catalytic dyad of cysteine and aspartic acid, as opposed to histidine and aspartic acid. This is not surprising as it methylates a hydroxyl group at the end of an alkyl chain rather than a phenolic ring as described for the other enzymes. [110]

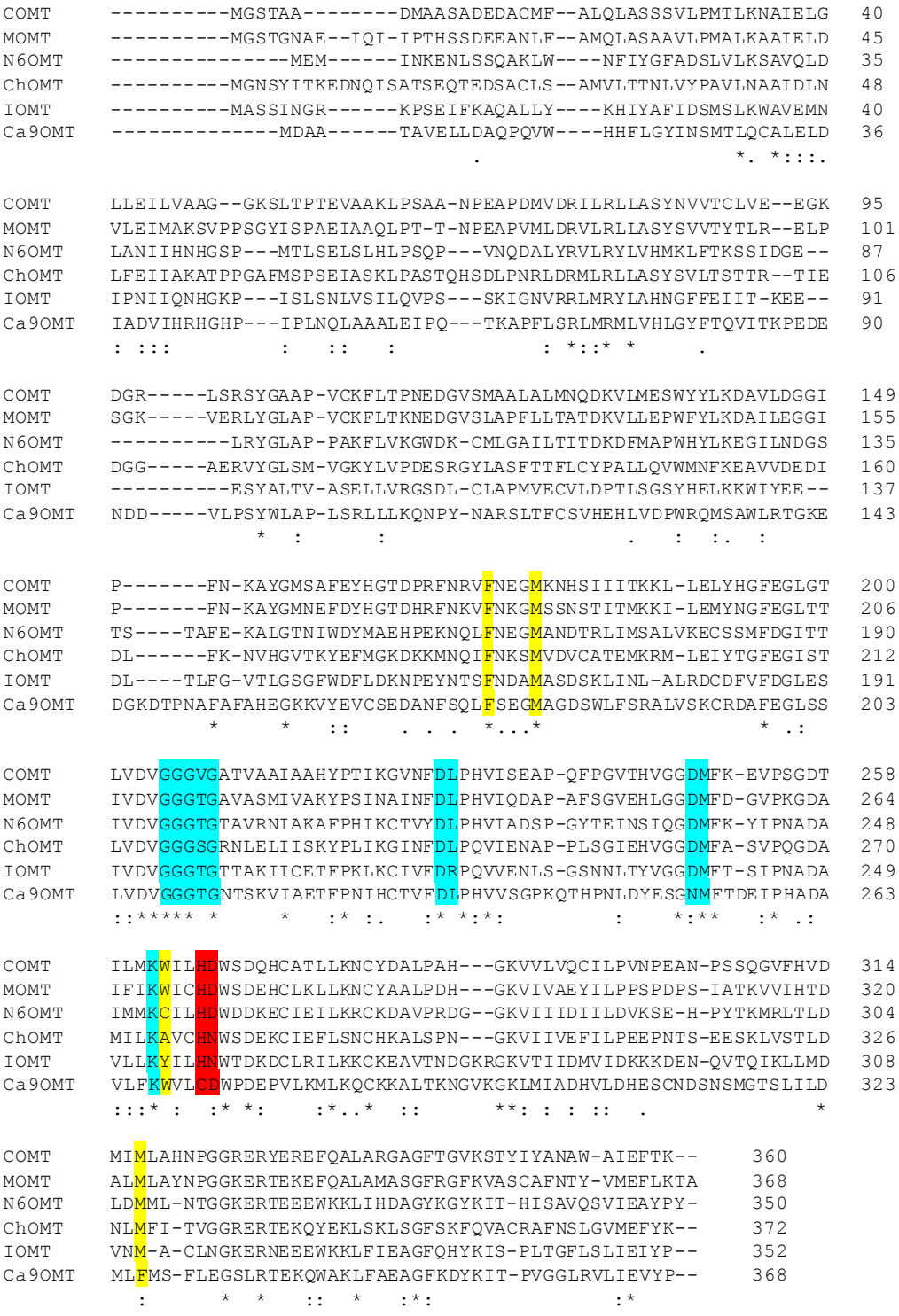



*Figure 1.15: Multiple sequence alignment of plant O-methyltransferases in a tertiary complex in closed conformation submitted to the Protein Database. Ca9OMT employs a different catalytic mechanism. Residues highlighted in cyan are the conserved residue involved in SAM binding as outlined in Figure 1.14. Red residues are the catalytic dyad and yellow residues are conserved in phenolic substrate binding. Lolium perenne caffeic acid O-methyltransferase (COMT), Clarkia breweri monolignol O-methyltransferase (MOMT) Thalictrum flavum norcoclaurine 6-O-methyltransferase (N6OMT), Medicago sativa Chalcone O-methyltransferase (ChOMT), Medicago sativa Isoflavone O-methyltransferase (IOMT), Linum nodiflorum coniferyl alcohol 9-O-methyltransferase (Ca9OMT)*

Phenolic substrate binding has not been studied in great detail, with only five conventional plant *O*-methyltransferases crystallised with the substrate, product or an analogue ligand bound. These are listed in Figure 1.15, COMT, MOMT, N6OMT, ChOMT and Ca9OMT. Conventional plant *O*-methyltransferases are those which utilise the His/Asp catalytic dyad for catalysis. The conserved residues involved in substrate binding are highlighted in yellow in Figure 1.15, along with the catalytic dyad in red, which is also involved in substrate binding. Interestingly the residues involved in phenolic substrate binding are conserved as shown in, Figure 1.16. The catalytic dyad of histidine and aspartic acid hydrogen bond to the methylated oxygen *via* $N_\varepsilon$ and $O_\delta$, respectively. The phenolic ring is clamped in place between a methionine $C_\varepsilon H$-π bond and phenylalanine on the opposing face *via* CH-π bonding. The thiol group on the methionine and the phenylalanine also interact with the transferred methyl group in the product bound structures, which would suggest that these residues are crucial in the correct orientation of the sulphonium methyl group on SAM and the positioning of the substrate for efficient methylation.

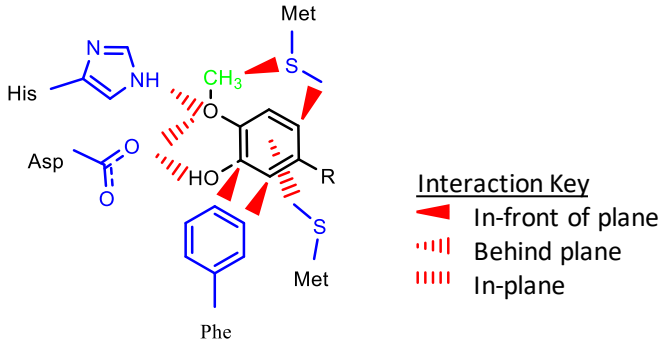



*Figure 1.16: A summary of conserved features involved in plant O-methyltransferase phenolic substrate binding. The phenolic product is shown in black, interacting residues are shown in blue, non-covalent interaction are shown in red and the transferred methyl group is shown in green*

### 1.7.5. Catalytic Mechanism

*O*-methyltransferases have a conserved catalytic dyad of histidine and aspartic acid.[55,106,112] The consensus reaction mechanism is the deprotonation of the hydroxyl group on the substrate by the conserved histidine, generating a nucleophile. Subsequently, nucleophilic attack upon the sulphonium methyl group by the

activated phenolic substrate proceeded by a $S_N2$ reaction, resulting in the formation of the methylated substrate, Figure 1.17. Site-directed mutagenesis studies have shown that mutations of the catalytic histidine result in inactive protein or protein with severely reduced activity, less than 1 %.[106,107,117] One mutation study of the aspartic acid in wheat flavone O-methyltransferase showed the loss of activity when substituted with a hydrophobic side chain and severe loss of activity when mutated to glutamic acid, but when mutated to asparagine 84 % of activity was retained.[117] They also reported mutations in the catalytic histidine "resulted in nearly complete loss of protein expression". [117]

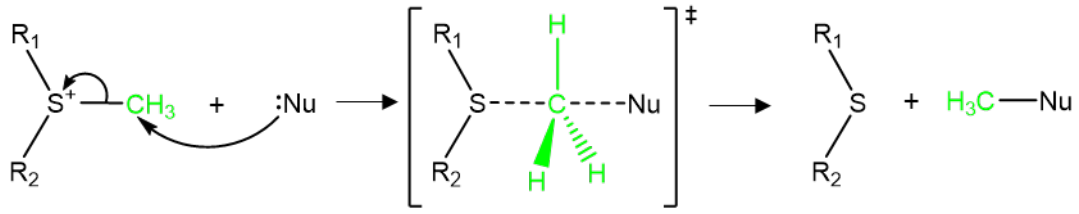



*Figure 1.17: Diagram of the conserved $S_N2$ nucleophilic attack reaction of SAM dependent methyltransferases where Nu represents the nucleophile and $R_1$ and $R_2$ represent the adenosyl and aminobutyryl moieties of SAM/SAH. The transferred methyl group is shown in green*

## 1.8. Summary and Aims

Plant *O*-methyltransferases substrate specificity has not been studied in detail due to a limited number of high resolution structures with the natural substrate or product bound in the active site. The conserved features of phenolic substrate binding have been analysed and summarised in Figure 1.16. Whereas, the residues involved in SAM binding have been studied in great detail with consensus motifs being described in the literature and summarised in Figure 1.14.

The opium poppy *O*-methyltransferase *Papaver somniferum O*-methyltransferase 1 (PSMT1) is responsible for the conversion of scoulerine to tetrahydrocolumbamine in the pathway to noscapine. Noscapine has potent antitumor properties due to its interaction with tubulin leading to cell arrest in metaphase.[19] PSMT1 catalyses the first step of the noscapine branch pathway in opium poppy and is encoded in the noscapine gene cluster, discovered in the Graham lab, which encodes for all enzymatic steps for the conversion of scoulerine to noscapine except for

tetrahydroprotoberberine cis-*N*-methyltransferase (TNMT).[48] The noscapine gene cluster also encodes for two other methyltransferases *Papaver somniferum* *O*-methyltransferase 2 (PSMT2) and *Papaver somniferum* *O*-methyltransferase 3 (PSMT3). The substrates for PSMT2 and PSMT3 were recently discovered, by heterologous gene expression in yeast, with the two polypeptides working together as a heterodimer to methylate the 4'-hydroxyl group on 4'-*O*-Desmethyl-3-*O*-acetylpapaveroxine to produce 3-*O*-acetylpapaveroxine, Figure 1.7.[78]

The major aim of this thesis was to characterise and generate structural information for PSMT1. Three-dimensional high resolution structural data provides a wealth of information allowing the rationalisation of reaction mechanism and substrate specificity. Currently there is little understanding about how the plant *O*-methyltransferase family of enzymes achieves substrate specificity, as only a few of them have been structurally elucidated. The work presented in this thesis contributes to our understanding through a detailed analysis of high resolution three-dimensional structures of PSMT1.

During the course of this PhD programme the author also carried out work over a 12 month period aimed at determining the crystallographic structure of the Iron (II)/ 2-oxoglutartae dependent dioxygenases codeine-*O*-demethylase (CODM) and thebaine-6-*O*-demethylase (T6ODM). CODM and T6ODM are responsible for the final demethylation steps on the biosynthetic pathway to morphine, Figure 1.5. [118,119] The pathway to morphine can take two routes with a bifurcation at thebaine depending on the actions of CODM and T6ODM. The major pathway, from which the enzymes are named, is *via* codeine where T6ODM demethylates thebaine to create neopinone which spontaneously rearranges to codeinone. Codeinone is then reduced by codeinone reductase (COR) to codeine and subsequently converted to morphine by 3-*O*-demethylation by CODM. The secondary route utilised the same enzymes but it an alternate order. Thebaine is demethylated by CODM at the O3 position to produce oripavine. Oripavine is subsequently demethylated by T6ODM, followed by COR to produce morphinone and morphine, respectively. T6ODM and CODM proved recalcitrant to crystallisation and this work will not be discussed in this thesis.

# 2. Materials and Methods

## 2.1. Molecular and Microbiological Methods

### 2.1.1. *E. coli* Strains

Expression and cloning hosts are commercially bought *E. coli* strains listed in Table 2.1.

| Strain | Brand | Resistance | Use |
|---|---|---|---|
| **BL21(DE3)** | Agilent | - | Expression |
| **XL10 Gold Ultracompetent** | Stratagene | Tetracycline, Chloramphenicol | Cloning |

*Table 2.1: List of commercial E. coli stains with their antibiotic resistance and use*

### 2.1.2. Plasmids

The plasmid pETFPP-3-PSMT1 containing a gene encoding a glutathione *S*-transferase (GST)-PSMT1 fusion protein separated by a 3C protease cleavage recognition site with a hexa-His N-terminal tag (6xHis-GST-r3CP-PSMT1) was generated from previous work undertaken in the Davies/Graham groups in collaboration with the Technology Facility at the Department of Biology, University of York. The pETFPP-3-PSMT1 had been shown to produce sufficient soluble PSMT1 protein for crystallisation trials. The plasmid map of pETFPP-3-PSMT1 and the protein sequence are shown in Figure 2.1 and Figure 2.2, respectively (full nucleotide sequence shown in Appendix 8.1).

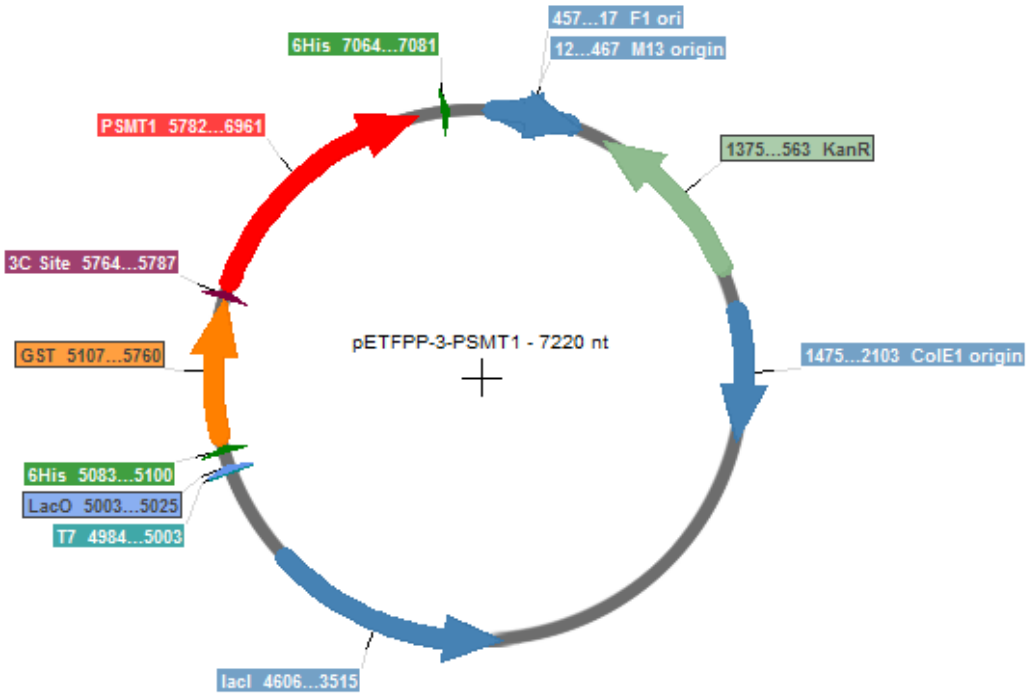



*Figure 2.1:* Plasmid map of pETFPP-3-PSMT1. The open reading frame contains a gene under the control of a T7 promotor encoding for GST-PSMT1 fusion protein separated by a 3C protease cleavage site and a hexa-His affinity tag on the N-terminus. Plasmid map generated using SerialCloner 2.6 (Serial Basic)

```
MGSSHHHHHH SSMSPILGYW KIKGLVQPTR LLLEYLEEKY EEHLYERDEG 50
DKWRNKKFEL GLEFPNLPYY IDGDVKLTQS MAIIRYIADK HNMLGGCPKE 100
RAEISMLEGA VLDIRYGVSR IAYSKDFETL KVDFLSKLPE MLKMFEDRLC 150
HKTYLNGDHV THPDFMLYDA LDVVLYMDPM CLDAFPKLVC FKKRIEAIPQ 200
IDKYLKSSKY IAWPLQGWQA TFGGGDHPPK GLEVLFQGPA MATNGEIFNT 250
YGHNHQSATV TKITASNESS NGVCYLSETA NLGKLICIPM ALRAAMELNV 300
FQLISKFGTD AKVSASEIAS KMPNAKNNPE AAMYLDRILR LLGASSILSV 350
STTKKSINRG GDDVVVHEKL YGLTNSSCCL VPRQEDGVSL VEELLFTSDK 400
VVVDSFFKLK CVVEEKDSVP FEVAHGAKIF EYAATEPRMN QVFNDGMAVF 450
SIVVFEAVFR VYDGFLDMKE LLDVGGGIGT SVSKIVAKYP LIRGVNFDLP 500
HVISVAPQYP GVEHVAGDMF EEVPKGQNML LKWVLHDWGD ERCVKLLKNC 550
WNSLPVGGKV LIIEFVLPNE LGNNAESFNA LIPDLLLMAL NPGGKERTIS 600
EYDDLGKAAG FIKTIPIPIS NGLHVIEFHK
```

*Figure 2.2: 6xHis-GST-r3CP-PSMT1 amino acid sequence in pETFFP-3-PSMT1. Green-His-tag, orange- GST, purple-3CP cleavage recognition site and red-PSMT1*

### 2.1.3. Bacterial Culture Media

*E. coli* was cultured in two different media. Luria-Broth (LB) medium, consisting of 10 g tryptone, 10 g NaCl, 5 g yeast extract in 1 L of deionised water, sterilised by autoclaving at 121 °C for 15 minutes, was used for plasmid preparations, glycerol stocks and inoculum for gene expression. Terrific-Broth (TB) medium was used for

recombinant gene expression, consisting of 12 g tryptone, 24 g yeast extract, 4 mL glycerol in 900 mL deionised water, sterilised by autoclaving at 121 °C for 15 minutes. To the sterile broth 100 mL of filter sterilised TB salts (0.17 M $KH_2PO_4$, 0.72 M $K_2HPO_4$) were added aseptically just before use.

LB Agar plates were prepared by the addition of 1.6 g of agar per 100 mL of LB media before autoclaving. The LB agar was cooled to about 50 °C, the relevant antibiotic was added under aseptic techniques and poured into sterile round plastic dishes. The plates were stored at 4 °C until needed.

### 2.1.4. Antibiotics

Antibiotics were added to the media and plates to select for *E. coli* containing a plasmid conferring antibiotic resistance. The two antibiotics used during this work were kanamycin and ampicillin. Kanamycin and ampicillin are prepared at 1000x stocks in distilled water, 50 mg $mL^{-1}$ and 100 mg $mL^{-1}$ respectively, and filter sterilised.

### 2.1.5. Cultivation of Bacteria

*E. coli* was generally cultured at 37 °C, 250 rpm in an orbital shaker. For small scale, overnight cultures 10 mL of LB media plus the relevant antibiotic was cultured in a 50 mL falcon tube. The media was inoculated by either picking a single colony from a plate or directly from a glycerol stock with a sterile 200 µL pipette tip, then placed directly into the media. The 50 mL flacon tube cap was loosely fitted to allow aeration and held in place with tape to stop the cap vibrating off. LB plates were incubated at 37 °C overnight.

For recombinant gene expression 500 mL of TB media with the appropriate antibiotic in a 2 L baffled shaker flask, was inoculated with 10 % (v/v) overnight culture. Once the optical density (OD) measured spectroscopically at 600 nm was 0.6 – 0.8 expression was induced by adding IPTG at a final concentration of 1 mM, from a 1000x (1 M) stock solution. The incubation temperature was reduced to 16 °C before induction, to aid protein folding. Cells were harvested by centrifugation (5000 rpm, 4 °C for 30 minutes) and transferred to 50 mL falcon tubes. Recombinant protein was then purified from the cells or the cell pellet was stored at -20 °C.

### 2.1.6. Preparation of Glycerol Stocks

Glycerol stocks were prepared for the long-term storage of viable *E. coli* cells at -80 °C. 1 mL of overnight culture was aliquoted into a 2 mL plastic tube containing 0.5 mL of 50 % glycerol and thoroughly mixed. Glycerol was used as a cryo-protectant to stop ice crystal formation. The glycerol stocks were flash frozen with liquid nitrogen (-196 °C), and stored at -80 °C.

### 2.1.7. Preparation of Chemically Competent *E. coli* Cells

Chemically competent *E. coli* cells are required for the transformation of foreign plasmid DNA into the cells. To prepare these, cells were sampled from a glycerol stock and spread on an LB plate with no antibiotic. A single colony from this plate was cultured overnight in 10 mL LB without antibiotic. 1 mL of overnight culture was used to inoculate 100 mL LB. Once OD reached 0.6, the cells were pelleted equally in two 50 mL falcon tubes by centrifugation at 4,000 rpm at 4 °C for 10 minutes. The supernatant was removed and the cells resuspended in 25 mL of cold, filter sterilised 0.05 M $CaCl_2$, then incubated for 20 minutes on ice. The cells were pelleted in the same manner as before, then resuspended in 10 mL of cold filter sterilised 0.1 M $CaCl_2$, 14 % (v/v) glycerol. The now chemically competent *E. coli* cells were aliquoted into 1.5 ml tubes, flash frozen with liquid nitrogen and stored at -80 °C.

### 2.1.8. DNA Transformation into *E. coli*

Plasmid DNA transformation of chemically competent *E. coli* cells was achieved using the heat shock method. 1 µL of purified plasmid DNA was mixed with 50 µL of BL21 (DE3) cells or 30 µL XL10 gold cells on ice, then incubated for 30 minutes, Table 2.2. The cells were then heat shocked by exposing them to 42 °C using a heat block for 45 seconds before returning to ice for 2 minutes. The heat shock treated cells were then mixed with 0.4 mL LB for BL21 (DE3) cells or 0.2 mL of LB for XL10 gold cells without antibiotic and incubated at 37 °C, 500 rpm. The culture was then spread onto LB plates containing the relevant antibiotic for the plasmid of choice. For site-directed mutagenesis (SDM) 5 µL of DpnI treated PCR product was mixed with 50 µL of XL10 gold cells, heat shock treated then incubated with 0.3 mL of LB.

| Cell Type | Use | Volume of cells | Volume of LB Media |
|---|---|---|---|
| **BL21(DE3)** | Expression | 50 µL | 0.4 mL |
| **XL10 Gold** | Cloning | 30 µL | 0.2 mL |
| | SDM | 50 µL | 0.3 mL |

Table 2.2: Volume and usage of chemically competent *E. coli* cells required for transformation

### 2.1.9. Isolation of Plasmid DNA from *E. coli*

Plasmid DNA was isolated exclusively form *E. coli* XL10 Gold cells cultured overnight using a QIAprep spin miniprep kit (Qiagen) as described in the manufacturer's manual. Plasmids were eluted with 30 µL of water.

### 2.1.10. DNA Sequencing

Plasmid DNA sequencing was carried out by GATC Biotech using their high-throughput Sanger sequencing service. The samples were prepared to their sample requirements: 5 µL isolated plasmid DNA (80-100 ng $\mu l^{-1}$) and 5 µL of 5 µM primer in a 1.5 mL tube. If plasmid DNA concentration was lower than 80 ng $\mu L^{-1}$, 2.5 µL of 10 µM plasmid plus 7.5 µL plasmid DNA at (50-70 ng $\mu L^{-1}$) was used.

### 2.1.11. Oligonucleotides

Oligonucleotides were designed with the aid of the Agilent QuikChange Primer Design tool (Agilent). All oligonucleotides were synthesised by Integrated DNA Technologies and purified by desalting. Upon receiving the desalted oligonucleotides, they were resuspended in molecular biology grade water to 100 µM. 10 µM working stock solutions were generated and all stored at -20 °C.

### 2.1.12. Site-Directed Mutagenesis

Site-directed mutagenesis (SDM) of plasmid DNA was carried out using the whole plasmid mutagenesis method. Overlapping oligonucleotides (primers) containing the desired mutation or mutations, flanked on either side with 15-20 complimentary nucleotides were generated with an annealing temperature of 78-79 °C. 10 ng of purified plasmid DNA was mixed with forward and reverse primers, water and KOD hot start master mix (Novagen), Table 2.3. KOD hot start master mix is a 2x mixture

containing KOD hot start polymerase, deoxynucleotides and reaction buffer containing $MgSO_4$. The whole plasmid was then amplified by PCR, typically using the thermal-cycling conditions outlined in Table 2.4. Extension time is based on KOD activity of 1 min 10 seconds per kilo base pairs at 68 °C, where optimal proofreading activity is obtained. Following the PCR, 1 µL of DpnI restriction enzyme (20 units $\mu L^{-1}$) was added to the reaction mixture and incubated at 37 °C to digest the methylated template DNA. The mutated plasmid DNA then transformed *E. coli* XL10 Gold cells. Three colonies exhibiting antibiotic resistance had their plasmid DNA extracted from overnight cultures and sequenced.

| Reagent | Volume (µL) | [Final] |
|---|---|---|
| **Molecular biology grade water** | 21 | |
| **Forward Primer (10 µM)** | 1.5 | 0.3 µM |
| **Reverse Primer (10 µM)** | 1.5 | 0.3 µM |
| **Template DNA (10 ng $\mu L^{-1}$)** | 1 | 10 ng |
| **KOD hot start master mix** | 25 | 0.02 U $mL^{-1}$ |

*Table 2.3: Quick Change site directed mutagenesis reaction constituents*

| Step | Temperature | Time | Cycles |
|---|---|---|---|
| **Hot Start** | 95 °C | 1 minute | 1 |
| **Denature** | 95 °C | 30 seconds | 20 |
| **Anneal** | 50 °C | 1 minute | |
| **Extend** | 68 °C | 8 minute 30 seconds | |
| **Storage** | 10 °C | Hold | 1 |

*Table 2.4: Thermal-cycling conditions for site directed mutagenesis*

### 2.1.13. Measurement of DNA Concentration

Nucleic acid concentration was measure using an Epoch microplate spectrophotometer (BioTek) equipped with a Take3 micro volume plate. 3 µL of each sample were pipetted onto the plate along with blanks and the absorbance at 260

nm ($A_{260}$) measured. Nucleic acid was quantified assuming an $A_{260}$ of 1 equates to 50 µg $mL^{-1}$ dsDNA.

## 2.2. Biochemical Methods

### 2.2.1. Isolation of Recombinant Protein

#### *2.2.1.1. Cell Lysis and Protein Extraction*

Cell pellets were resuspended in nickel affinity purification buffer A (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole pH 8.0, 10 mM β-mercaptoethanol) at a ratio of 8 mL of buffer to 1 g of cell pellet. A protease inhibitor cocktail tablet (cOmplete, Roshe) was added to reduce protease degradation of the target protein. The resuspended cells were lysed using mechanical methods by sonication, 15 seconds on/15 seconds off for 5 minutes. DNase I was added to the lysate to cleave the bacterial DNA, resulting in reduced viscosity of the solution. All steps were performed on ice to reduce the activity of proteases.

The cell debris was removed by centrifugation, 18,000 rpm for 45 minutes at 4 °C. The supernatant, containing all soluble proteins was filtered through a 0.22 µM syringe driven filter.

#### *2.2.1.2. Affinity Chromatography*

The clarified lysate was applied to a 5 mL HisTrap FF crude column (GE healthcare) using an ÄKTA purification system (GE healthcare), pre-equilibrated with nickel affinity purification buffer A (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole pH 8.0, 10 mM β-mercaptoethanol). The column was washed with buffer A to remove expression host proteins, leaving only the PSMT1 fusion construct bound.

#### *2.2.1.3. On-Column Cleavage*

On-column cleavage was implemented by injecting 5 mL of 0.25 mg $mL^{-1}$ hexa-histidine tagged 3C protease fusion protein directly onto the 5 mL HisTrap FF crude column, containing bound His-GST-r3CP-PSMT1, using a syringe and incubated overnight at 4 ºC. The protease cleaved the His-GST-r3CP-PSMT1 at the 3C protease cleave recognition site, liberating PSMT1 from its solubility partner, with a small three residue N-terminal overhang. The released PSMT1 was then removed from the column by washing the column with nickel affinity buffer A and collecting fractions in a 2 mL 96 deep-well collection plate. The bound proteins, containing un-cleaved His-GST-r3CP-PSMT1, cleaved His-GST and 3C protease were eluted with nickel affinity purification buffer B (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.5 M imidazole pH 8.0, 10 mM β-mercaptoethanol). Fractions with an $A_{280}$ response were analysed by SDS-PAGE and the fractions containing PSMT1 (43 kDa) were pooled.

#### *2.2.1.4. 1.1.2.1 Desalting*

After on-column cleavage the protein sample was desalted in preparation for ion exchange. This was achieved by using a HiPrep 26/60 Desalting column (GE Healthcare) according to the manufacturers protocol on an ÄKTA purification system (GE healthcare). The protein solution was concentrated to less than 10 mL and injected *via* a 15 mL loop onto the column, pre-equilibrated with 50 mM Tris pH 8.0, 10 mM β-mercaptoethanol. Multiple HiPrep 26/60 desalting columns were connected in series for large purifications.

#### *2.2.1.5. Ion Exchange Chromatography*

Anion exchange was implemented as a second purification step. Anion exchange columns contain a positively charged resin which can bind negatively charged molecules/proteins. The theoretical isoelectric point (pI) of PSMT1 is 5.46, generated using the ProtParam tool on the ExPASy server.[120] This results in PSMT1 having a net negative charge in the buffer system used, Tris pH 8.0. Desalted protein was injected using a 15 mL loop onto a HiScreen Q HP 4.7 mL (GE Healthcare) anion exchange column pre-equilibrated with 50 mM Tris pH 8.0, 10 mM β-mercaptoethanol (ion exchange Buffer A). The 15 mL loop was filled with 10 mL of protein solution multiple times until all protein solution was loaded onto the column. The column was washed

with 5 % ion exchange buffer B (50 mM Tris pH 8.0, 1 M NaCl, 10 mM β-mercaptoethanol) to remove weakly bound proteins. A 10-50 % buffer B gradient was applied over 50 mL to elute PSMT1. A step to 100% buffer B was used to remove strongly bound proteins. The eluate was fractionally collected and protein visualised by measuring absorbance at 280 nm and SDS-PAGE. Fractions containing PSMT1 (43 kDa) were pooled.

#### *2.2.1.6. Size Exclusion Chromatography*

PSMT1 was concentrated in a volume of less than 0.5 mL using a 30 kDa centrifugal concentration unit and injected onto a S200 16/600 gel filtration column (GE Healthcare) pre-equilibrated with 50 mM Tris pH 8.0, 150 mM NaCl, 10 mM β-mercaptoethanol. 1 mL fractions were collected and protein content measured by $A_{280}$. Fractions containing protein were analysed by SDS-PAGE and those containing PSMT1 were pooled.

#### *2.2.1.7. Buffer Exchange*

Purified PSMT1 was buffer exchanged into 10 mM Tris pH 8.0, 5 mM TCEP for crystallisation trials. This was achieved by repeated concentration and dilution of the protein solution using a 30 kDa centrifugal concentrating unit (Millipore).

### 2.2.2. Concentration of Protein Solution

Protein solutions were concentrated using Millipore centrifugal concentrating units according to the manufacturers protocol.

### 2.2.3. The Quantification of Protein Concentration

#### *2.2.3.1. UV spectroscopy*

The protein concentration in an aqueous solution was determined by spectroscopic methods. The theoretical molar extinction coefficient (ε) was calculated using Expasy ProtParam Server as $A = \varepsilon c l$, where A is the absorbance at 280 nm, ε is the molar extinction co-efficient ($M^{-1} cm^{-1}$), c is the concentration (M) and l is the path length, the concentration of the purified protein solution can be calculated using its absorbance at 280 nm.[120]

#### *2.2.3.2. Bradford Protein Assay*

The Bradford protein assay was used to determine protein concentration as an alternate method to that of UV spectroscopy. The Bradford protein assay exploits brilliant blue G-250 dyes ability to bind to proteins, resulting in the deprotonation of the dye. This leads to a colour change, red to blue, which was monitored by measuring absorbance at 595 nm ($A_{595}$). The 1 mL standard assay as described in the Quick Start™ Bradford protein assay instruction manual (BioRad) was followed.

### 2.2.4. Polyacrylamide Gel Electrophoresis (PAGE)

#### *2.2.4.1. SDS-Polyacrylamide Gel Electrophoresis*

Denaturing SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) was implemented to visualise and estimate the molecular weight of purified protein. The protein sample was mixed 1:2 with SDS-PAGE sample buffer (50 mM Tris pH 6.8, 10 % (v/v) glycerol, 2 % SDS, 0.05 % bromophenol blue, 0.7 M β-mercaptoethanol) and heated to 100 °C for 5 min. This unfolded the protein allowing the SDS to bind to the protein uniformly, giving the protein a net negative charge. The reducing agent β-mercaptoethanol is used to break disulphide bonds to reduce the effects of tertiary structure. The samples were loaded onto 12 % SDS-PAGE gels along with molecular weight standards. Electrophoresis was carried out using a Mini-PROTEAN Tetra Cell (BioRad) in running buffer (0.2 M glycine, 25 mM Tris, 0.1 % SDS) and ran at 200 V for 50 minutes. After electrophoresis, the gels were stained with magic dye and photographed using an InGenius LHR Gel Imaging System (Syngene). Molecular weight standards were prepared by mixing 25 µL of low range molecular weight markers with 375 µL of SDS-PAGE sample buffer. Magic dye is produced in a total volume of 2.5 L consisting of 150 mg of brilliant blue G-250 dye and 8.6 mL HCl. SDS-PAGE gels were produced in-house using the BioRad Mini-PROTEAN Tetra Cell Casting Module (BioRad). 12 % gels were made up as in Table 2.5.

| Component | Resolving Gel 12 % | Resolving Gel 10 % | Stacking Gel |
|---|---|---|---|
| **Deionised Water** | 3.2 mL | 4.0 mL | 3.2 mL |
| **Resolving gel buffer** | 2.5 mL | 2.5 mL | - |
| **Stacking gel buffer** | - | - | 1.3 mL |
| **10 % Acrylamide** | 4.2 mL | 3.3 mL | 0.5 mL |
| **10 % APS** | 50 µL | 50 µL | 25 µL |
| **TEMED** | 8 µL | 8 µL | 8 µL |

*Table 2.5: SDS-PAGE Gel recipe. Resolving gel buffer- 1.5 M Tris pH 8.8, 0.4 % SDS, stacking gel buffer- 0.5 M pH 6.8, 0.4 % SDS. APS- ammonium persulphate, TEMED- Tetramethylethylenediamine*

#### *2.2.4.2. Native-PAGE*

Native polyacrylamide gel electrophoresis (Native-PAGE) differs from SDS-PAGE as SDS is omitted from all components. This technique separates proteins based on their charge and hydrodynamic radius and was utilised to determine if a purified protein was homogeneous. Gels and reagents were made up as for SDS-PAGE but SDS is omitted and replaced with deionised water, Table 2.5. 10 % gels were used and electrophoresis was carried out at 35 V for 1 hour at 4 ºC.

### 2.2.5. Size Exclusion- Multi Angle Laser Light Scattering (SEC-MALLS)

Size exclusion multi angle laser light scattering (SEC-MALLS) was utilised to determine the molecular weight of proteins in solution. SEC-MALLS was carried out in the Technology Facility (TF) at the University of York using a Shimadzu LC-20AD liquid chromatography machine linked to a Shimadzu SIL-20A auto-sampler. Purified protein samples of 1-2 mg $mL^{-1}$ were injected onto a Superdex 200 10/300 GL column (GE Healthcare). Measurements were taken using a Shimadzu SPD-20A UV detector, a Wyatt Optilab rEX refractive index monitor, a Wyatt DAWN HELEOS light scattering detector and analysed using Wyatt ASTRA software package.

### 2.2.6. Thermal Shift Assay -Thermofluor Assay

A thermal shift assay is used to determine the temperature at which a protein unfolds thus conferring its stability in a given solution. The thermofluor assay utilises the

ability of the dye SYPRO™ orange (Life Technologies) to bind to hydrophobic patches. In aqueous solution, the dye is quenched but once the dye binds to the exposed hydrophobic regions, exposed upon protein unfolding, the dye can be excited at 492 nm and fluoresces at 610 nm.

A working stock of SYPRO orange was prepared by making a 5x solution of SYPRO orange in 10 mM Tris pH 8.0, 5 mM TCEP with a total volume of 1 mL. Purified protein at a concentration of 1 mg $mL^{-1}$ in 10 mM Tris pH 8.0, 5 mM TCEP, determined by UV spectroscopy, was mixed 1:1 with the dye in a total volume of 30 µL in a 96 well plate. The plate was heated from 25 ºC to 95 ºC at 1 ºC per 30 seconds. Fluorescence was measured every 30 seconds. Melting temperature ($T_m$) is calculated as the midpoint of the unfolding event characterised by an increase in fluorescence.

### 2.2.7. Activity Assay

The enzymatic activity of PSMT1 was determined by a stopped assay and analysed by UPLC-MS/MS. Reactions were carried out in 0.1 mM glycine-NaOH pH 9, 25 mM sodium ascorbate, 1 mM β-mercaptoethanol, 100 µM SAM and with scoulerine concentrations ranging from 0.05-10 µM. Reactions were performed in a total volume of 11.5 mL and incubated at 37 ºC in a water bath. Reactions were started by the addition of 0.5 ml of 1 µg $mL^{-1}$ enzyme. 1 mL samples were taken between 15 seconds – 5 minutes and quenched with equal volumes of methanol. The quenched samples were dried using a SpeedVac (Genevac), a centrifugal evaporator. The dried samples were then resuspended in 100 µL 10 % acetic acid.

All samples were analysed, alongside standards of scoulerine and tetrahydrocolumbamine by UPLC-MS/MS. The reverse-phase UPLC method implemented was first introduced by Winzer et al., 2012 (Winzer et al., 2012). A Waters Acquity UPLC system (Waters Ltd.) was used quipped with a Acquity UPLC BEH C18 1.7 µm 2.1 x 100 mm column (Waters Ltd.) incubated at 60 ºC. The mobile phases consisted of A: 10 mM ammonium bicarbonate pH 10.2 and B: Methanol. A gradient was used as described in Table 2.6, at a flow rate of 0.5 mL $minute^{-1}$. 2 µL of sample was injected and analysed using a Thermo TSQ Endura triple quadrupole in ESI positive mode using selective reaction monitoring (SRM). Scoulerine and

tetrahydrocolumbamine were detected using the settings in Table 2.7 and Figure 2.3. Resolution for Q1 and Q3 was 0.7 FEHM and a collision induced dissociation gas rate of 1.5 mTorr. The runs were controlled by Thermo Xcalibur software (Thermo Fisher Scientific), and analysed using Thermo Xcalibur software, Microsoft Excel (Microsoft) and GraphPad (Prism).

| Min | %A | %B | Flow rate |
|---|---|---|---|
| 0.0 | 98 | 2 | 0.5 |
| 0.2 | 98 | 2 | 0.5 |
| 0.5 | 60 | 40 | 0.5 |
| 4.0 | 20 | 80 | 0.5 |
| 4.5 | 20 | 80 | 0.5 |
| 4.6 | 98 | 2 | 0.5 |
| 5.0 | 98 | 2 | 0.5 |

*Table 2.6: UPLC gradient profile for analysis of PSMT1 assay. Solvent A: 10 mM ammonium bicarbonate pH 10.2, Solvent B: Methanol. Flow rate in ml min$^{-1}$*

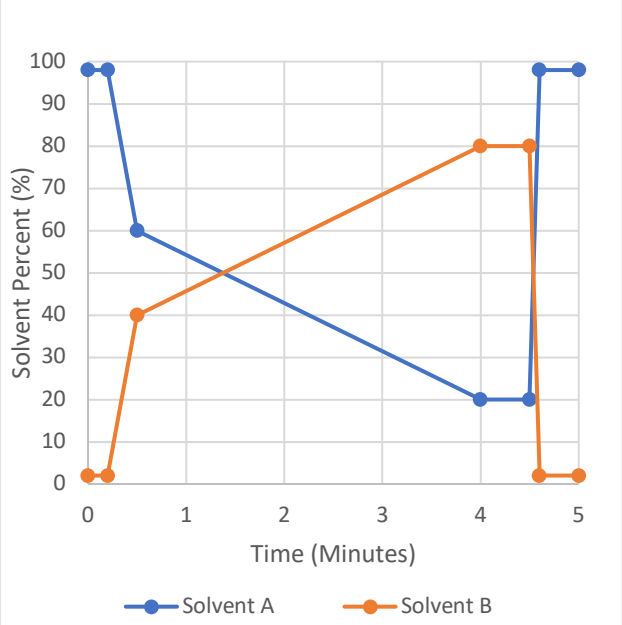



*Figure 2.3: UPLC gradient profile for analysis of PSMT1 assay. Solvent A: 10 mM ammonium bicarbonate pH 10.2 (Blue), Solvent B: Methanol (Orange)*

| Compound | Retention Time (min) | RT Window (min) | Precursor (m/z) | Product (m/z) | Collision Energy (V) | RF Lens (V) |
|---|---|---|---|---|---|---|
| **Scoulerine** | 2.2 | 0.7 | 328.049 | 178 | 24.815 | 139 |
| **Tetrahydrocol-umbamine** | 2.9 | 0.7 | 342.08 | 178 | 24.815 | 145 |

*Table 2.7: Mass spectrometer SRM settings for the analysis of scoulerine and tetrahydrocolumbamine*

#### 2.2.8. Protein Molecular Weight Determination by Mass Spectrometry

Purified protein at ranges from 1-15 mg $mL^{-1}$ were submitted to the Technology Facility at the University of York for molecular weight determination by electrospray ionisation- mass spectrometry (ESI-MS).

#### 2.2.9. Surface Entropy Reduction

Surface entropy reduction (SER) was first proposed by Derewenda in 2004[121] and is a technique that generates low entropy patches on the protein surface by site directed mutagenesis. The technique relies on the UCLA MBI surface entropy reduction prediction (SERp) server.[122] The SERp server carries out a series of processes giving a score to each residue for a given protein sequence. The secondary structure is predicted by PSIRED, a conformational entropy profile is generated and finally PSI-BLAST is utilised to predict residue conservation. Lysine and glutamic acid have been shown to localise predominantly on the surface and interfere with protein-protein interfaces.[123,124] A typical high scoring residue will be a glutamic acid or lysine on a surface exposed loop which is non-conserved. Finally, the values are plotted along the sequence and clusters of up to three high scoring residues surrounded by low scoring residues are suggested for mutation.

### 2.3. Protein Crystallisation

#### 2.3.1. Initial Screening

Initial crystallisation screening experiments were carried out using commercially bought sparse matrix screens utilising the sitting drop method in MRC 96 well dual drop plates. 46 µL of a crystallisation screen was transferred from a 2 mL 96 deep-

well block to a MRC 96 well dual drop crystallisation plate using a HYDRA II 96-channel micro-dispenser liquid handling robot (Thermo Scientific). A mosquito liquid handling robot (TTP Labtech) was utilised to dispense protein solution, typically at 10-15 mg $mL^{-1}$ and crystallisation condition into the sitting drop wells, typically 150 nl of each. The plates were sealed with a UV transparent film and stored in a Rigaku Minstrel HT UV (Rigaku) at either 20 °C or 4 °C. Plates were imaged periodically with visible and UV light and remotely monitored using Rigaku XtalTrak™ (Rigaku).

### 2.3.2. Crystallisation Optimisation

'Hits' from the initial screen are conditions which produced promising results such as small crystals, needles or clustered precipitate, these were taken forward for optimisation. Optimisation was carried out in 48 well trays by varying the pH, salt concentration and precipitant concentration around that of the initial hit. The drop volumes used were increased to a total volume of 1 µL, typically, 0.5 µL of buffering condition to 0.5 µL of protein solution. The optimisation plates were produced by hand using electronic pipettes and incubated under the same conditions as the original hit.

### 2.3.3. Data Collection

Crystals were fished/picked using a protein crystal mounting loop (Hampton CrystalCap™ Spine HT) from the crystallisation plate. If additional cryo-protectant was needed the crystal was exposed to a drop of cryo-protectant solution before flash freezing with liquid nitrogen. Initial X-ray diffraction studies were carried out in-house utilising a Rigaku MicroMax-007 HF X-ray Generator and a Rigaku R-Axis IV++ image plate detector. 0 ° and 90 ° images were taken with an oscillation of 0.5 ° and 5-minute exposures to 2.5 Å. If sufficient X-ray diffraction was observed the crystal were sent to Diamond Light Source for collection of a full data set. Data collected at Diamond Light Source was auto-processed by XIA2 on the beamline.[125]

### 2.3.4. Phasing

Molecular replacement was carried out using Molrep.[126]

### 2.3.5. Model Building and Refinement

Models were built using the CCP4i2 suite of software including BUCCANEER, REFMAC and COOT.[127–131]

# 3. Structural Characterisation of PSMT1

## 3.1. Introduction

*Papaver somniferum* *O*-methyltransferase 1 (PSMT1) is an *S*-adenosylmethionine dependent class I *O*-methyltransferase found in opium poppy with a molecular weight of 42.6 kDa, consisting of 390 amino acids. PSMT1 is responsible for the catalytic conversion of scoulerine to tetrahydrocolumbamine, Figure 3.1, by methylating the 9-hydroxyl position of the substrate in the biosynthetic pathway to noscapine, Figure 1.7.

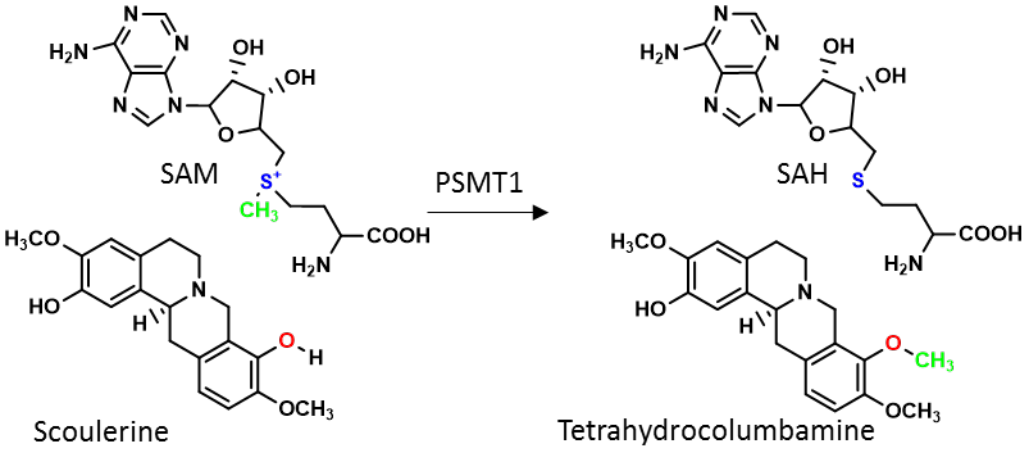



*Figure 3.1: PSMT1 reaction schematic*

PSMT1 was originally discovered and named scoulerine *O*-methyltransferase 1 alongside two other O-methyltransferase PSMT2 and PSMT3.[71] The first scoulerine *O*-methyltransferase to be discovered and characterised was that from *Coptis japonica*, a putative orthologue of PSMT1.[70] All three showed activity towards scoulerine as a substrate with PSMT1 having the highest $V_{max}$ at 2036 nmol $min^{-1}$ $mg^{-1}$ of protein and the lowest $K_m$ of 28.5 µM. PSMT1 was discovered to be part of a gene cluster for noscapine biosynthesis along with nine other genes encoding several enzyme classes including two *O*-methyltransferases (PSMT2 and PSMT3), four cytochrome P450s (CYP82X1, CYP82X2. CYP82Y1 and CYP71A21), short-chain dehydrogenase/reductase (PSSDR1), an acetyltransferase (PSAT1) and a carboxylesterase (PSCXE1), Figure 1.8.[48] The discovery together with functional genomics analysis allowed a biochemical pathway for noscapine to be proposed.[48]

PSMT1 was characterised by virus induced gene silencing and in vitro biochemical analysis as catalysing the first committed step in the noscapine branch pathway, Figure 1.5 and Figure 1.7.[48,71] At the time the function of the two other O-methyltransferases, PSMT2 and PSMT3, were not fully elucidated but PSMT2 was indicated in methylation at the 4'-OH position of a secoberbine intermediate to produce papaveroxine. Subsequent research by Li and Smolke[132] which involved the heterologous expression of the genes from the noscapine gene cluster in *Saccharomyces cerevisiae* revealed PSMT2 and PSMT3 work as a heterodimer to catalyse the 4'-*O*-methylation of 4'-*O*-desmethyl-3-*O*-acetylpapaveroxine to 3-*O*-*a*cetylpapaveroxine. Interestingly they also noted that a heterodimer of PSMT2 and norcoclaurine 6-*O*-methyltransferase (6OMT) can also catalyse this 4'-*O*-methylation reaction.

This chapter describes the heterologous expression of PSMT1, its purification, crystallisation and subsequent structure determination by X-ray crystallography.

## 3.2. Materials and Method

### 3.2.1. Expression and Purification of PSMT1

PSMT1 was expressed from pETFPP-3-PSMT1 as an N-terminally hexa-histidine tagged GST fusion protein with a 3C-protease cleave site linker (6xHis-GST-r3CP-PSMT1) in *E. coli BL21 (DE3)* cells for in vitro characterisation (as described in Section 2.2.1). The full-length construct has a predicted molecular weight of 70.5 kDa with the cleaved components having molecular weights of: 6xHis-GST 27.6 kDa and PSMT1 42.8 kDa, calculated utilising ProtParam.[120]

A single colony from a Luria-Broth (LB) agar plate containing 50 µg $mL^{-1}$ kanamycin was picked and used to inoculate 10 mL of LB medium with 50 µg $mL^{-1}$ kanamycin. This was incubated at 37 °C overnight. 5 mL of the overnight culture was used to inoculate 500 mL Terrific Broth (TB) medium with 50 µg $mL^{-1}$ kanamycin in 2.5 L baffled shaker flasks. The 500 mL culture was incubated at 37 °C, 200 rpm until cell growth entered the exponential phase indicated by an absorbance at 600 nm ($A_{600nm}$) of 0.6 – 1.0, also known as optical density (OD). Once the desired OD was achieved

the cultures were cooled to 16 °C and expression of the gene encoding 6xHis-GST-r3CP-PSMT1 was induced by addition of isopropyl-β-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The culture was incubated at 16 °C, 200 rpm overnight, for at least 18 hours, for the overexpression of the fusion protein. Cells were harvested by centrifugation at 5000 rpm at 4 °C for 30 minutes in a Beckman Avanti J-HC equipped with a JLA 8.1000 rotor. The supernatant was discarded and the cells pellet was either resuspended in 50 mM Tris pH 8.0, 500 mM NaCl, 10 mM β-mercaptoethanol, 20 mM imidazole with DNaseI and a protease inhibitor cocktail tablet (cOmplete™, Roche) for purification or stored at -20 °C.

The cells were lysed by sonication utilising a large probe with the cells solution on ice to reduce protease activity. A programme of short 15 second pulses with 15 seconds resting time was used to reduce heat induced protein damage. The cell debris was removed by centrifugation at 18,000 rpm for 45 minutes at 4 °C in a Sorvall SS34 rotor, followed by filtration using a 0.22 µM syringe driven filter.

PSMT1 was purified using a multi-step purification protocol, utilising nickel affinity chromatography on-column cleavage, desalting and ion exchange. Nickel affinity chromatography was used to capture the 6xHis-tagged fusion protein, then Human rhinovirus-3C protease cleaved the fusion protein releasing PSMT1 from the 6xHis-GST purification/solubility tag. Desalting removed the low molecular weight solutes such as NaCl and imidazole from the protein solution. Finally, anion exchange chromatography was used to remove any remaining contaminating proteins.

The clarified cell lysate was loaded onto a 5 mL HisTrap nickel affinity column (GE Healthcare) which was pre-equilibrated with nickel affinity buffer A (50 mM Tris pH 8.0, 500 mM NaCl, 10 mM β-mercaptoethanol, 20 mM Imidazole). 6xHis-GST-r3CP-PSMT1 would bind to the $Ni^{2+}$ *via* the 6xHis-tag while host proteins will not bind and were removed by flowing nickel affinity buffer A across the column until the $A_{280}$ response returned to baseline. The column was then treated with 1 column volume (CV) of 0.25 mg $mL^{-1}$ Human rhinovirus-3C protease (3CP) (6xHis-Maltose Binding Protein-3CP, 64 kDa) and incubated overnight at 4 °C. The 3CP would cleave 6xHis-GST-r3CP-PSMT1 into two at the 3CP recognition site (r3CP): the

purification/solubility tag (6xHis-GST) and the now free protein of interest, PSMT1 with a 3 residue N-terminus overhang. The cleaved PSMT1 was washed off the column with nickel affinity buffer A. The purification/solubility tag and 3CP stay bound to the column which were subsequently eluted with nickel affinity buffer B (50 mM Tris pH 8.0, 500 mM NaCl, 10 mM β-mercaptoethanol, 0.5 M Imidazole). Fractions with a high $A_{280}$ response were analysed by SDS-PAGE and fractions containing bands corresponding to PSMT1 *ca.* 43 kDa were pooled. The fractions containing PSMT1 were pooled then desalted using a HiPrep 26/60 Desalting column (GE Healthcare) pre-equilibrated with ion exchange buffer A (50 mM Tris pH 8.0, 10 mM β-mercaptoethanol).

Desalted PSMT1 solution was then loaded onto a HiScreen Q HP 4.7 ml (GE Healthcare) anion exchange column pre-equilibrated with ion exchange buffer A (50 mM Tris pH 8.0, 10 mM β-mercaptoethanol). The column was washed with 10 % ion exchange buffer B (50 mM Tris pH 8.0, 10 mM β-mercaptoethanol, 1 M NaCl) to remove weakly bound proteins. Bound PSMT1 was eluted with a 10-50 % buffer B gradient over 10 CVs and a 100 % buffer B step was utilised to elute strongly bound proteins and regenerate the column.

The fractions containing PSMT1, determined by SDS-PAGE analysis, were pooled and buffer exchanged into 10 mM Tris pH 8.0, 5 mM TCEP utilising repeated concentration and dilution of the protein solution using a 30 kDa centrifugal concentrating unit. Finally, the protein was concentrated to 15 mg $mL^{-1}$ for crystallisation trials or stored at -80 °C for future experiments. The concentration of PSMT1 was determined using its calculated extinction coefficient of 0.698 $M^{-1}$ $cm^{-1}$ at 280 nm. All steps were analysed by SDS-PAGE for purity.

### 3.2.2. Activity Assays

The activity of PSMT1 was investigated using scoulerine and SAM as substrates, Figure 3.1, by UPLC-MS/MS. Enzymatic assays were conducted using scoulerine obtained from Prof Peter J. Schammells (Monash University, Australia) and SAM (Santa Cruz Biotechnology). The rate of methylation of scoulerine to tetrahydrocolumbamine by PSMT1 was determined for both scoulerine and SAM as

the substrate. Each reaction was carried out in a total volume of 11.5 mL of 100 mM glycine-NaOH pH 9.0, 25 mM sodium ascorbate, 1 mM β-mercaptoethanol with either SAM at a constant 100 µM while varying scoulerine from 0.05 µM to 10 µM or scoulerine at a constant 5 µM while varying SAM from 0.01 µM to 100 µM. The reactions were incubated at 37 °C in a water bath and the reactions were initiated by the addition of 0.5 mL of 1 µg $mL^{-1}$ of purified PSMT1 from on-column cleavage in nickel affinity buffer A. Protein concentration of a ca. 1 mg $mL^{-1}$ was prepared by monitoring $A_{280}$ and diluted to ca. 1 µg $mL^{-1}$. The actual concentration of the ca. 1 mg $ml^{-1}$ solution was determined using the Bradford protein assay as described in (Section 2.2.3.2, pg. 57), this value was used to calculate the reaction rate. 1 mL aliquots were taken at intervals over a maximum of 5 minutes and quenched with 1 mL of methanol. The quenched samples were subsequently dried using a SpeedVac centrifugal evaporator (Genevac) and then resuspended in 100 µL of 10 % acetic acid to give a 10-fold increase in concentration for quantification by UPLC-MS/MS.

The enzymatic assay samples were analysed alongside standards of scoulerine and tetrahydrocolumbamine (Santa Cruz Biotechnology) utilising a modified method by Winzer et al., 2012.[48] A Waters Acquity UPLC system (Waters Ltd.) equipped with a Acquity UPLC BEH C18 1.7 µm 2.1 x 100 mm column (Waters Ltd.) incubated at 60 °C was utilised to separate the alkaloids by reverse phase chromatography. The mobile phases consisted of A: 10 mM ammonium bicarbonate pH 10.2 and B: Methanol with a flow rate of 0.5 mL $minute^{-1}$. 2 µL of the sample was injected onto the column and the alkaloids eluted with a gradient of solvent B over 5 minutes as shown in Figure 2.3 and Table 2.6. Scoulerine and tetrahydrocolumbamine were quantified using a Thermo TSQ Endura triple quadrupole Mass Spectrometer (MS) in ESI positive mode using selective reaction monitoring (SRM), which provided greater sensitivity over the orbitrap utilised in the published method.[48] SRM is a highly sensitive and selective method of detection with a high signal to noise ratio. This is achieved by the first quadrupole selecting for the mass to charge for either scoulerine (328 m/z) or tetrahydrocolumbamine (342 m/z). The second quadrupole fragments the precursor ion by collision induced dissociation before the third, and final quadrupole is set to select the fragmentation product which results in the highest intensity signal. In both

cases the fragment product which gave the highest signal intensity had a m/z of 178. Compound optimisation was carried out for both scoulerine (Appendix 8.2) and tetrahydrocolumbamine (Appendix 8.3), separately, by directly injecting them into the MS using a syringe driver. Scoulerine and tetrahydrocolumbamine were detected over a time window of 0.7 minutes, 1.85 - 2.55 minute and 2.55 - 3.25 minute, respectively. The time windows were selected so only one fragmentation product was detected at a time, to give the greatest response possible.

The concentration of tetrahydrocolumbamine was calculated for each sample based on a calibration curve of known concentrations versus response. Concentration against time was plotted and an initial rate calculated for each reaction. The specific rate of reaction in nmol $min^{-1}$ $mg^{-1}$ of protein was plotted against the concentration of substrate being varied. Each point represents the mean specific rate of reaction at each substrate concentration over three independent replicates with standard error of the mean shown as vertical error bars. $K_m$ and $V_{max}$ were calculated by direct fit of the Michaelis-Menten equation in GraphPad software (Prism).

### 3.2.3. Size Exclusion-Multi-Angle Laser Light Scattering

Size Exclusion Multi-Angle Laser Light Scattering (SEC-MALLS) was carried out in the Bioscience Technology Facility (TF) at the University of York using a Shimadzu LC-20AD liquid chromatography machine linked to a Shimadzu SIL-20A auto-sampler. 100 µL of purified protein samples of 2 mg $mL^{-1}$ in 10 mM Tris pH 8.0 were injected onto a Superdex 200 10/300 GL column (GE Healthcare) pre-equilibrated with 10 mM Tris pH 8.0, 150 mM NaCl. Measurements were taken using a Shimadzu SPD-20A UV detector, a Wyatt Optilab rEX refractive index monitor, a Wyatt DAWN HELEOS light scattering detector and analysed using Wyatt ASTRA software package. BSA was ran as a control.

### 3.2.4. Thermal Shift Assay -Thermofluor Assay

1 mg $mL^{-1}$ PSMT1 in 10 mM Tris pH 8.0, 5 mM TCEP which was purified by nickel affinity on-column cleavage followed by desalting and anion exchange was mixed with 5x SYPRO orange dye at a 1:1 ratio, totalling 30 µL in a 96 well plate. 5x SYPRO

orange dye was prepared by diluting 5000x SYPRO orange dye in DMSO to 5x in 10 mM Tris pH 8.0, 5 mM TCEP. Protein sample was analysed alongside that of a buffer only sample as a control. The samples were heated at 2 °C minute$^{-1}$ and fluorescence recorded every 30 seconds, excitation 492 nm, emission 610 nm. The fluorescence intensity was normalised and plotted against temperature. The melting point of the protein was determined at the midpoint of the unfolding event characterised by an increase in intensity.

### 3.2.5. Protein Molecular Weight Determination by Mass Spectrometry

The molecular weight of PSMT1 was determined by the Molecular Interactions Laboratory at the Bioscience Technology Facility based at the University of York as a service. The protein sample was buffer exchanged in to 2 mM Tris pH 8.0 using centrifugal concentrator units and submitted at a concentration of 5-10 mg ml$^{-1}$. The calculated molecular weight was corrected in reference to the external standard myoglobin.

### 3.2.6. Crystallisation of PSMT1

PSMT1 crystallisation trays were setup using a 1:1 ratio of 15 mg mL$^{-1}$ PSMT1 in 10 mM Tris, 5 mM TCEP pH 8.0, utilising the vapour diffusion method in MRC 96 well dual sitting drop crystallisation plates, and incubated at 20 °C in a Rigaku Minstrel HT UV crystal imager (Rigaku). Drops containing 150 nL of protein solution and 150 nL of crystallisation screen condition were dispensed by a mosquito liquid handling robot (TTP Labtech). The reservoir contained 46 µL of commercially bought crystallisation screen as an initial screen. Co-crystallisation experiments were carried out with 0.7 mM SAH or 2 mM scoulerine or both. Conditions with 'hits' such as small crystals or globular precipitate were optimised in 48 well sitting drop plates. Optimisation consisted of varying the components within the hit reservoir, such as pH and percipient concentrations, to obtain better crystals for X-ray diffraction studies. 48 well sitting drop trays were set-up by hand using electronic pipettes, the same ratio of reservoir solution to protein solution were used with a total volume of 1 µL, against 100 µL of reservoir solution and stored at the same temperature. Crystals were tested in house using a Rigaku MicroMax-007 HF X-ray Generator and Rigaku R-Axis

IV++ image plate detector. 0 ° and 90 ° X-ray diffraction images are taken with an oscillation of 0.5 °, 5 minute exposures to 2.5 Å resolution. Crystals which diffracted well were sent to Diamond Light Source for collection of a full data set.

#### 3.2.7. Structure Determination

Diffraction images for crystals of apoPSMT1 were auto-processed using xia2 to space group P $3_1$ 2 1 but subsequently changed to P $3_2$ 2 1 and solved by molecular replacement using a previously obtained structure of apoPSMT1 in the Davies Lab. MOLREP was used using default parameters, resulting in 4 molecules per asymmetric unit. Iterative rounds of model building in COOT and maximum likelihood-refinement utilising REFMAC were carried out to build the model.[129–131]

## 3.3. Results and Discussion

### 3.3.1. Expression and Purification of PSMT1

PSMT1 was successfully expressed as part of a fusion protein, 6xHis-GST-r3CP-PSMT1 from the plasmid pETFPP-3-PSMT1 in *E. coli* BL21(DE3) cells. Soluble PSMT1 was purified to homogeneity using a multistep process of nickel affinity on-column cleavage with 3C protease, Figure 3.2, followed by desalting and anion exchange chromatography, Figure 3.3.

Purification of PSMT1 by nickel affinity on-column cleavage with 3C protease was analysed by SDS-PAGE, Figure 3.2. 6xHis-GST-r3CP-PSMT1 (70 kDa) in the cell lysate (lane 1) was loaded onto a 5 mL nickel affinity chromatography column. 6xHis-GST-r3CP-PSMT1 bound to the column while cellular proteins did not bind and flowed through the column (lane 2). 5 mL of 0.25 mg $mL^{-1}$ HRV 3C protease (lane 3) was injected onto the column and incubated overnight at 4 °C. The HRV 3C protease would cleave the fusion construct at the HRV 3C protease cleavage site (r3CP) liberating PSMT1 from its solubility and affinity tag, 6xHis-GST. The now free PSMT1 was washed off the column with nickel affinity buffer A (lanes 4-12). The fractions consisting mostly that of PSMT1, indicated by a stained band of ca. 43 kDa were pooled. The pooled fractions also contained some contaminating HRV 3C protease ca. 64 kDa and un-cleaved fusion protein ca. 70 kDa. The proteins bound to the

column were eluted with nickel affinity buffer B (lanes 13-14). These fractions contained the full length PSMT1 fusion protein (70 kDa), HRV 3C protease (64 kDa), the cleaved solubility/purification tag (6xHis-GST, 28 kDa) and cleaved PSMT1 (43 kDa). The presence of free PSMT1 in the elution fractions is most likely due to further processing of the fusion protein after elution of the proteins from the column.

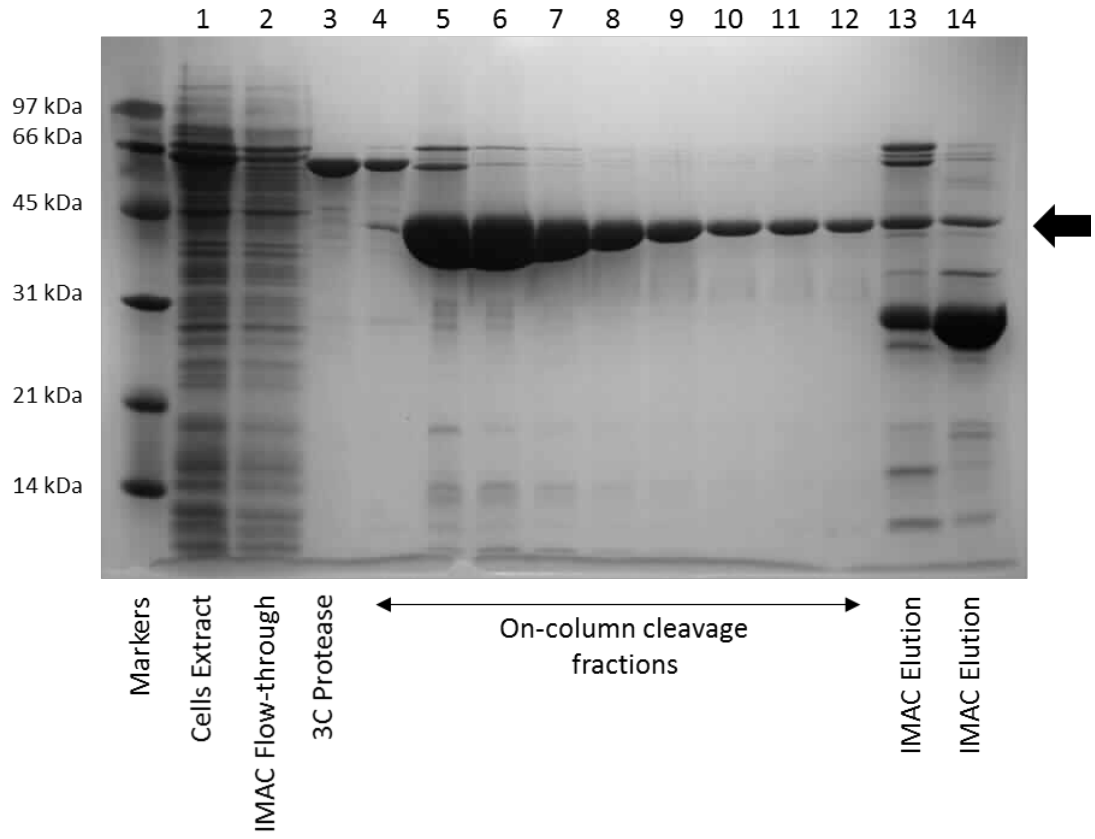



*Figure 3.2: Purification of PSMT1 by on-column cleavage analysed by SDS-PAGE. 1: Soluble fraction loaded onto IMAC column, 2: IMAC flow-through containing proteins which have not bound to the column. 3: 6xHis-MBP-3CP utilised for on-column cleavage. 4-12: Fractions collected after on-column cleavage containing cleaved PSMT1. 13-14: IMAC Elution fraction. Arrow indicates PSMT1 (43 kDa) position*

To produce a homogeneous solution of PSMT1 further purification was needed. Size exclusion and anion exchange were tested with anion exchange chromatography producing the purest sample. For anion exchange chromatography, the protein sample needed to be desalted before loading onto the anion exchange column, Figure 3.3. The protein sample was concentrated to 10 mL and loaded onto the desalting column (lane 1). The fractions with an $A_{280}$ response were pooled (lane 2). This was then loaded onto an anion exchange column and eluted with an increasing

sodium chloride concentration, utilising a gradient from 10 % - 50 % ion exchange buffer B (lanes 3-13). Fractions containing pure PSMT1 were pooled (lanes 6-11). HRV 3C protease eluted with a low salt concentration in lane 4. Strongly bound proteins were eluted with 100 % ion exchange buffer B (lane 14), this contained un-cleaved PSMT1 fusion construct along with some cleaved PSMT1. A typical purification is summarised in Figure 3.4, along with purified PSMT1 buffer exchanged into 10 mM Tris pH 8.0, 5 mM TCEP (lane 9) and subsequently concentrated to 15 mg $mL^{-1}$ (lane 10) for crystallisation trials.

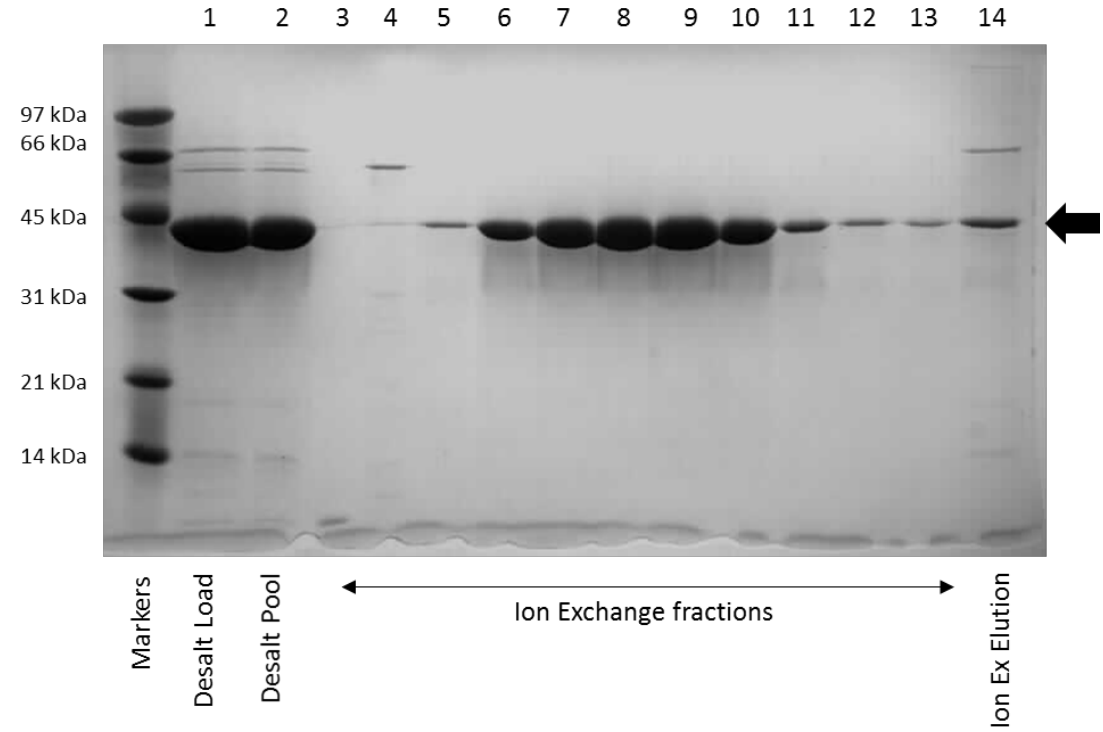



*Figure 3.3: Purification of PSMT1 by desalting and Ion Exchange analysed by SDS-PAGE. Arrow indicates PSMT1 position 1: Protein pooled from on-column cleavage containing cleaved PSMT1 (43 kDa) which was loaded onto the desalting column. 2: Pooled desalting fractions which were loaded onto an anion exchange column. 3-13: Ion exchange elution fractions obtained by increasing sodium chloride concentration. 14: Ion exchange final elution with 1 M sodium chloride to remove strongly bound proteins. Arrow indicates PSMT1 position*

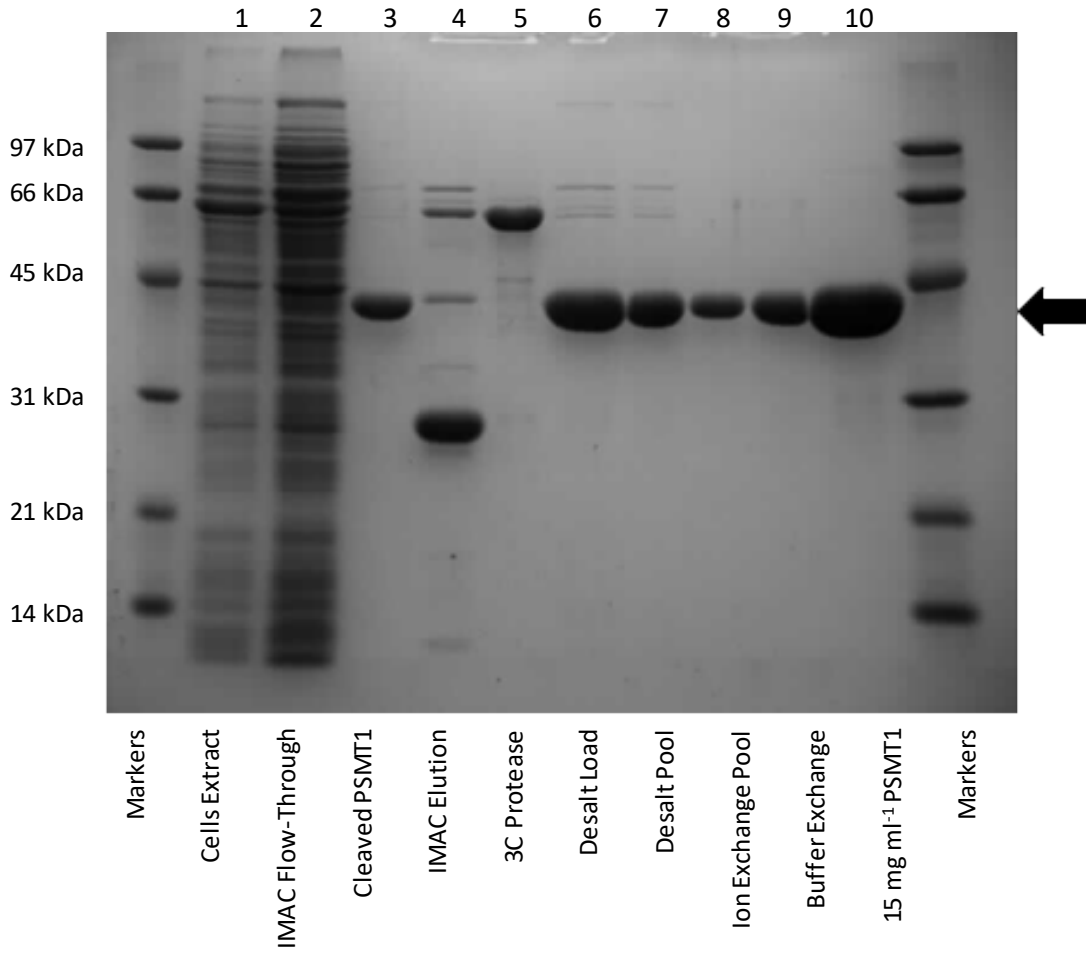



*Figure 3.4: Analysis of PSMT1 purification by 12 % SDS-PAGE. 1: Soluble fraction loaded onto IMAC column, 2: IMAC flow-through containing proteins which have not bound to the column. 3: Protein washed off the column post on-column cleavage, band ca. 43 kDa is PSMT1. 4: Elution of bound proteins post on-column cleavage containing 6xHis-GST-r3CP-PSMT1 (71 kDa), 6xHis-GST (28 kDa) and 6xHis-MBP-3CP (64 kDa). 5: Purified 6xHis-MBP-3CP utilised for on-column cleavage. 6: Protein sample loaded onto desalting column. 7: Pooled fractions from desalting column which was loaded onto an anion exchange column. 8: Pooled fractions from anion exchange chromatography containing pure PSMT1. 9: Ion exchange pool buffer exchanged into 10 mM Tris, 5 mM TECP pH 8.0. 10: A Sample of 15 mg ml$^{-1}$ PSMT1 used for crystallisation trials. Arrow indicates cleaved PSMT1 (43 kDa)*

### 3.3.2. Characterisation of Purified PSMT1

#### *3.3.2.1. Molecular Weight Determination*

15 mg mL$^{-1}$ PSMT1 in 10 mM Tris, 5 mM TCEP purified by on-column cleavage and anion exchange chromatography was buffer exchanged into 5 mM Tris pH 8.0 using a centrifugal concentrator. The resulting protein solution was submitted to the Bioscience Technology Facility based at the University of York for molecular weight determination by mass-spectrometry. The predicted molecular weight of purified

PSMT1 is 42843.4 Da as calculated by ProtParam based on the amino acid composition.[120] The corrected measured molecular weight by mass-spectrometry of PSMT1 was 42843.8 Da. This is within 1 Da of the expected molecular weight of PSMT1, therefore PSMT1 has been successfully purified and has not been degraded.

#### *3.3.2.2. Native-PAGE*

In order to determine the behaviour of PSMT1 in solution native-PAGE analysis was performed. PSMT1 purified by nickel affinity on-column cleavage followed by desalting and anion exchange in 10 mM Tris pH 8.0, 5 mM TCEP was analysed by native-PAGE, Figure 3.5. Native-PAGE separates proteins based on their charge and hydrodynamic radius, with smaller highly charged proteins migrating further than larger less charged proteins. The results show one major, large band of PSMT1 indicating a folded protein in one state. A ladder effect reducing in intensity with reduced migration is visualised, this is most likely due to multiple charged states of PSMT1 in solution or could be due to multiple oligomeric states. This protein sample was used for crystallisation trials.

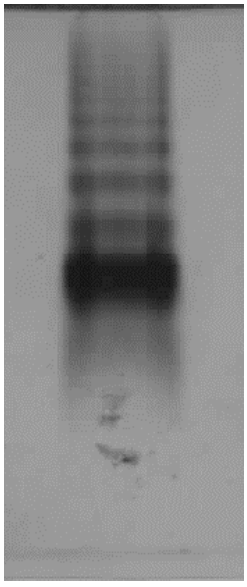


*Figure 3.5: Native-PAGE analysis of purified PSMT1. 30 µg of purified protein loaded*

### 3.3.2.1. *Size Exclusion Multi-Angle Laser Light Scattering*

In order to determine the molecular weight in solution, size exclusion-multi angle laser light scattering was carried out. 100 µL of a 2 mg mL$^{-1}$ solution of PSMT1 in 10 mM Tris pH 8.0, 5 mM TCEP was injected onto a Superdex 200 10/300 GL column (GE Healthcare) pre-equilibrated with 10 mM Tris pH 8.0, 150 mM NaCl, Figure 3.6. The elution UV trace shows a single monodispersed peak with a retention time of 14 minutes and an observed mean molecular weight of 83360 Da over the peak. The calculated molecular weight of the purified PSMT1 gene product is 42843.4 Da. As the observed molecular weight is double that of the expected molecular weight it can be concluded that PSMT1 exists as a homodimer in solution. This is consistent with other characterised plant *O*-methyltransferases such as eugenol O-methyltransferase from *Clarkia breweri, Medicago sativa* caffeic acid *O*-methyltransferase and *Thalictrum tuberosum* catechol *O*-methyltransferase, which have all be characterised as homodimers.[133,107,113]

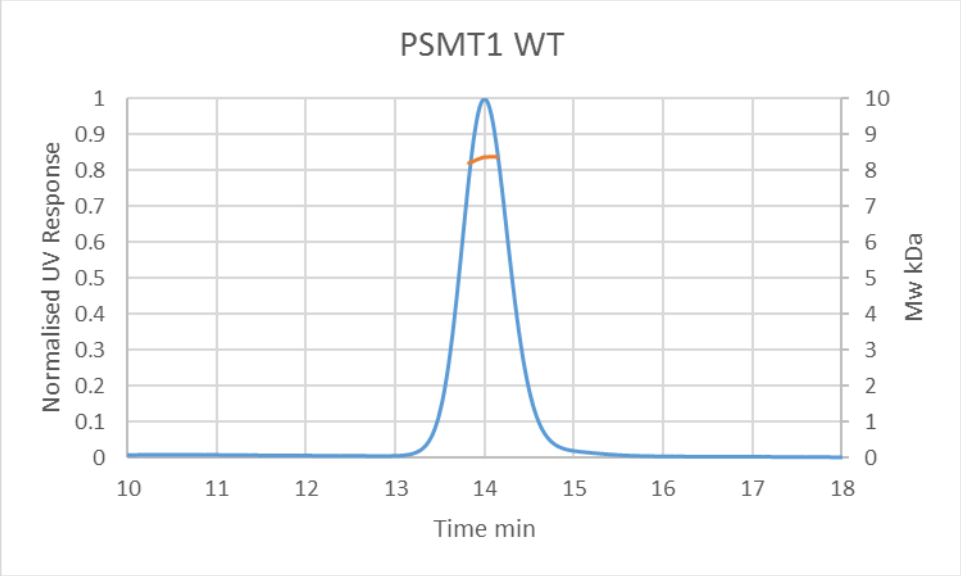



*Figure 3.6: SEC-MALLS results of PSMT1. Blue line represents normalised UV response, while the orange line represents the calculated molecular weight across the peak which has a mean molecular weight of 83360 Da*

### 3.3.2.2. *Thermal Shift Assay- Thermofluor*

The melting temperature ($T_m$) of PSMT1 was measured as 50 °C for a 1 mg mL$^{-1}$ solution of PSMT1 in 10 mM Tris pH 8.0, 5 mM TCEP, Figure 3.7. The protein solution was mixed 1:1 with 1xSYPRO orange dye, in the same buffer as PSMT1, with a total

volume of 30 µL. The PSMT1 sample showed a steady fluorescence from 25 °C to 40 °C, where upon the fluorescence started to increase until 58 °C. Over 58 °C the fluorescence decays, probably due to protein aggregation. This shows that the protein is stable and folded in the buffer conditions used. The buffer-only sample shows a steady fluorescence with a slight decrease with increased temperature, which is expected for a non-protein sample. This shows that the response recorded in the PSMT1 sample is due to the protein and not an artefact of the buffer composition used.

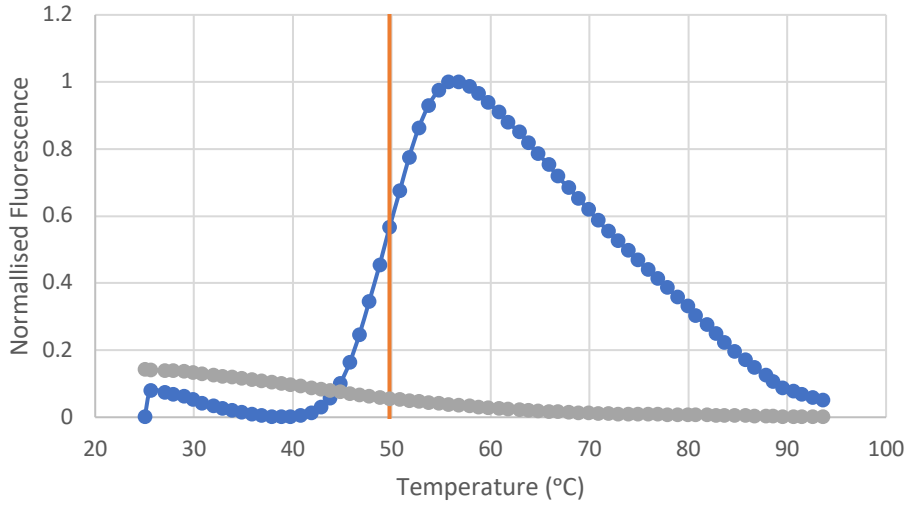



*Figure 3.7: Thermofluor analysis of PSMT1. The graph shows a normalised fluorescence response of PSMT1 against time (Blue) and the calculated melting temperature (orange), buffer-only sample (grey)*

### 3.3.3. Enzyme Kinetics of PSMT1

Kinetic analysis of PSMT1 purified by nickel affinity on-column cleavage was performed, Figure 3.8.A and B, and the $V_{max}$ and $K_m$ values were determined by Michaelis-Menten direct fit, Table 3.1. A $V_{max}$ of 492.6 nmol $min^{-1}$ $mg^{-1}$ of protein and a $K_m$ of 0.25 µM was calculated with scoulerine as the varying substrate concentration with a constant concentration of 100 µM *S*-adenosylmethionine (SAM). A $V_{max}$ of 542.4 nmol $min^{-1}$ $mg^{-1}$ of protein and a $K_m$ of 8.76 µM was calculated with SAM as the varying substrate with a constant concentration of 5 µM scoulerine. The fact that the $V_{max}$ values for scoulerine and SAM as substrates are similar at 492.6 and 542.4 nmol

$min^{-1}$ $mg^{-1}$ of protein respectively provide confidence that the methodology used for this kinetic analysis is robust.

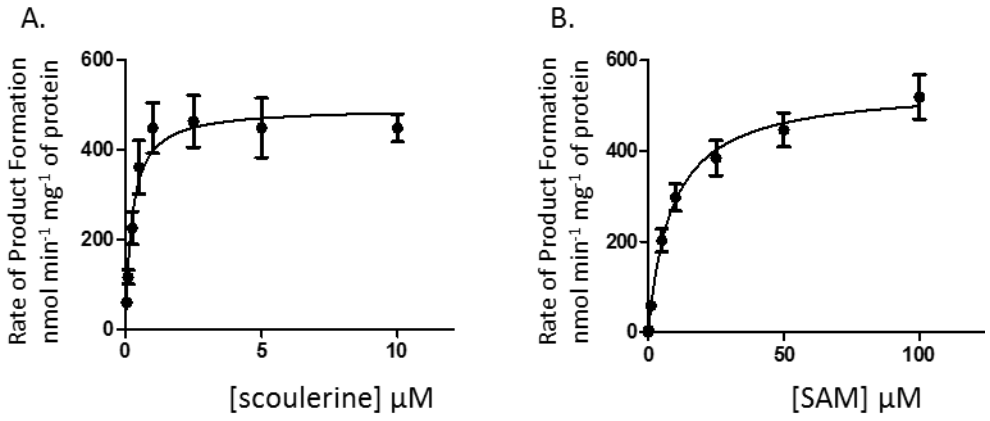



*Figure 3.8: Michaelis-Menten kinetics of PSMT1 with scoulerine and SAM as substrates. A. Varying the concentration of scoulerine with SAM at a constant 100 µM. B. Varying the concentration of SAM with a constant concentration of 5 µM scoulerine. SAM = S-adenosylmethionine*

| Substrate | | Value | std error |
|---|---|---|---|
| **Scoulerine** | $V_{max}$ (nmol $min^{-1}$ $mg^{-1}$ of protein) | 492.6 | 28.68 |
| | $K_m$ (µM) | 0.25 | 0.064 |
| **SAM** | $V_{max}$ (nmol $min^{-1}$ $mg^{-1}$ of protein) | 542.4 | 27.14 |
| | $K_m$ (µM) | 8.76 | 1.64 |

*Table 3.1: Michaelis-Menten enzyme kinetics of PSMT1 with scoulerine and SAM as substrates*

Kinetic data obtained for PSMT1 by Facchini *et al.* (2012)[71] produced a $V_{max}$ of 2036 ± 160.9 nmol $min^{-1}$ $mg^{-1}$ of protein and a $K_m$ of 28.5 µM for scoulerine as the substrate and SAM at a constant concentration of 200 µM. With SAM as the substrate and scoulerine at a constant concentration of 100 µM a $V_{max}$ of 1290 ± 73.18 nmol $min^{-1}$ $mg^{-1}$ of protein and a $K_m$ of 19 ± 2.66 µM SAM was reported.[71] These data thus report significantly higher $K_m$ and $V_{max}$ values than what have been obtained in this thesis, with the $K_m$ for scoulerine being approximately 100 fold and a $V_{max}$ that is four fold higher. For SAM the $K_m$ and $V_{max}$ are both approximately 2 fold higher in the Facchini *et al.* (2012)[71], also reported PSMT1 performing a secondary *O*-methylation of tetrahydrocolumbamine, at the 2-hydroxyl position, to produce tetrahydropalmatine but did not detect direct methylation of scoulerine at this position. This activity was not observed for PSMT1 activity assay carried out by the author of this study. Compounds were separated utilising the same UPLC method as stated previously,

Section 3.2.2 pg. 67, but mass detection was carried out in full scan mode on an LTQ Orbitrap (Themo Fisher). Apart from the substrate scoulerine with a retention time of 2.31 minutes and a m/z of 328 and the product tetrahydrocolumbamine with a retention time of 2.93 minutes and a m/z of 342 no other peaks were detected on the total ion count chromatogram, Figure 3.9. Thus, the data obtained in the study could not confirm the tetrahydrocolumbamine O-methylation activity of PSMT1. It cannot be completely ruled out though that these differences in the PSMT1 activity profile are caused by minor sequence polymorphisms. The PSMT1 protein used in this study and that used by Facchini *et al.* (2012)[71] very slightly differ in their amino acid sequence, with a percentage identity of 99.23 %, sequence alignment shown in Appendix 8.4. This equates to three residue polymorphisms, His14Arg, Ser16Thr and Val270Phe, which could be responsible for the altered substrate specificity and activity.

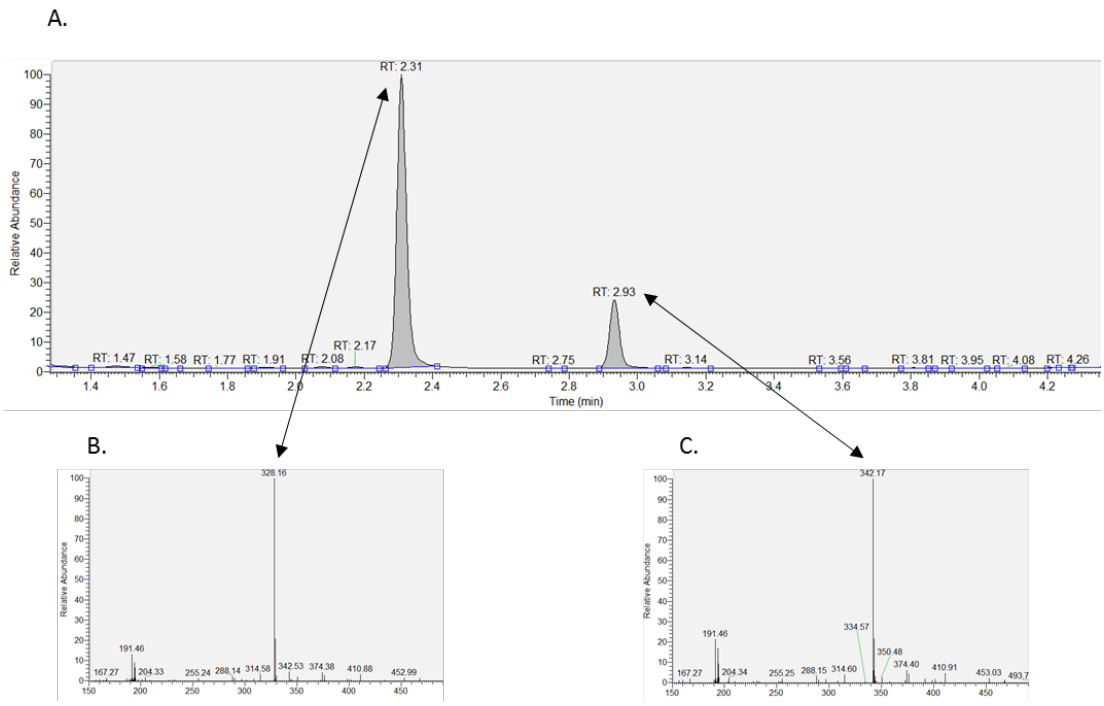



*Figure 3.9: UPLC-MS analysis of PSMT1 conversion of scoulerine to tetrahydrocolumbamine. A. Total Ion Count Chromatogram B. Mass spectrum at 2.31 min. C. Mass spectrum at 2.93 min*

$K_m$ values for other published *O*-methyltransferases vary from the low µM to low mM range but are unlikely to be their natural substrate.[112,114,134–136] Secondary plant metabolite enzymes generally have a broad substrate specificity with higher $K_m$

values than enzymes of central metabolism.[90,137] PSMT1 has one of the lowest reported $K_m$ value compared to other enzymes of plant secondary metabolism, which suggests that it has evolved to methylate only scoulerine with high efficiency and efficacy. This is corroborated by the virus induced gene silencing experiments undertaken by Winzer et al. (2012)[48], where scoulerine was only detectable in the latex when PSMT1 was silenced.[48]

The data were also analysed utilising the sigmoidal Hill plot to determine if there is cooperativity between the subunits, Table 3.2. Positive cooperativity is when binding of a substrate molecule to one active site increases the affinity toward the substrate at a second active site; this is characterised as having a Hill coefficient greater than one. Negative cooperativity is when binding of a substrate molecule to one active site decreases the affinity toward the substrate at a second active site; this is characterised as having a Hill coefficient less than one. Non-cooperativity is when binding of a molecule to one active site has no effect on affinity towards the substrate at a second active site and is characterised as having a Hill coefficient of one.

Positive cooperativity was observed for the *O*-methylation of scoulerine by PSMT1 when analysing the data with varying scoulerine concentrations, with a Hill coefficient of 1.50. This indicates that the active sites communicate with each other constructively, with binding affinity increasing at the second binding site when the first is occupied. This has also been noted in other *O*-methyltransferases, such as for *Thalictrum tuberosum* *O*-methyltransferases with Hill coefficients varying form 1.4-1.7 depending on the enzyme and substrate assayed, and for *Sorghum bicolor* caffeoyl-coenzyme A 3-*O*-methyltransferase with a reported Hill coefficient of 2.0.[113,114] The Hill coefficient value of 0.85 for SAM as the substrate indicated negative cooperativity, this could be due to the inhibitory effects of SAH.

| Substrate | | Value | std error |
|---|---|---|---|
| **Scoulerine** | $V_{max}$ (nmol $min^{-1}$ $mg^{-1}$ of protein) | 465.0 | 26.75 |
| | $K_{prime}$ (μM) | 0.106 | 0.073 |
| | Hill coefficient | 1.50 | 0.39 |
| **SAM** | $V_{max}$ (nmol $min^{-1}$ $mg^{-1}$ of protein) | 581.6 | 70.52 |
| | $K_{prime}$ (μM) | 7.51 | 1.64 |
| | Hill coefficient | 0.85 | 0.18 |

*Table 3.2: Hill plot enzyme kinetic analysis results of PSMT1 with scoulerine and SAM as substrates to probe cooperativity*

### 3.3.4. Crystallisation of PSMT1

Commercially available crystallisation screens such as PACT premier (Molecular Dimensions), JCSG-plus (Molecular Dimensions), SaltRx (Hampton Research), Index (Hampton Research), and Structure Screen 1 and 2 HT-96 (Molecular Dimensions) were set-up with PSMT1 purified by nickel affinity on-column HRV-3C Protease cleavage followed by anion exchange chromatography in 10 mM Tris, 5 mM TCEP pH 8.0. After optimisation apoPSMT1 crystals were produced in 0.1 M bis-tris propane pH 7.5, 25 % PEG3350, 0.3 M sodium citrate, Figure 3.10. The crystals were mounted on loops and flash frozen in liquid nitrogen. Initial X-ray diffraction images were collected using the in-house X-ray source at 0 ° and 90 °, Figure 3.11, and extended to approximately 4 Å resolution. The best crystals were sent to Diamond Light Source, Harwell, UK, for a full X-ray diffraction data set to be collected. The best diffracting crystal was that of apoPSMT1 and extended to 2.95 Å resolution. Co-crystallisation experiments with SAH/SAM and/or scoulerine produced crystals of poor quality which diffracted poorly except for that with SAH/SAM. Optimisation of co-crystallisation experiments of PSMT1 with SAH/SAM did not yield a data set of greater resolution than that obtained within earlier research carried out in the Davies lab prior to the start of this work. In this previous work crystals were obtained by Wendy A. Offen in 100 mM HEPES pH 7.0, 1 M sodium succinate, 10 % PEG 2000.

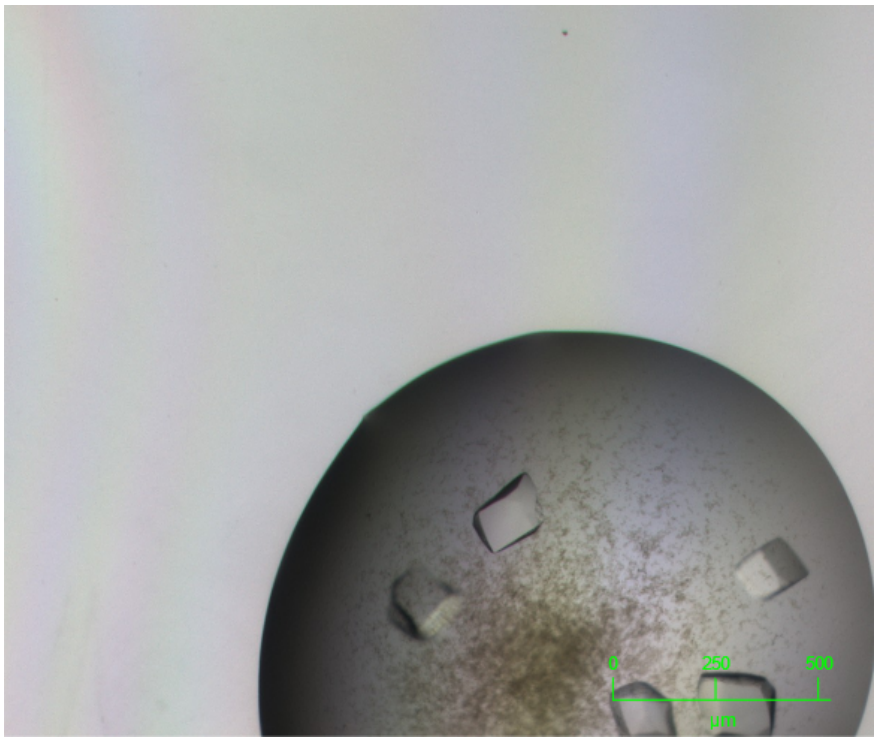



*Figure 3.10: apoPSMT1 crystals grown in 0.1 M bis-tris propane pH 7.5, 25 % PEG3350, 0.3 M sodium citrate in a 1:1 ratio of 15 mg mL$^{-1}$ PSMT1, totalling a volume 1 µL in a 48 well sitting drop plate incubated at 20 ℃*

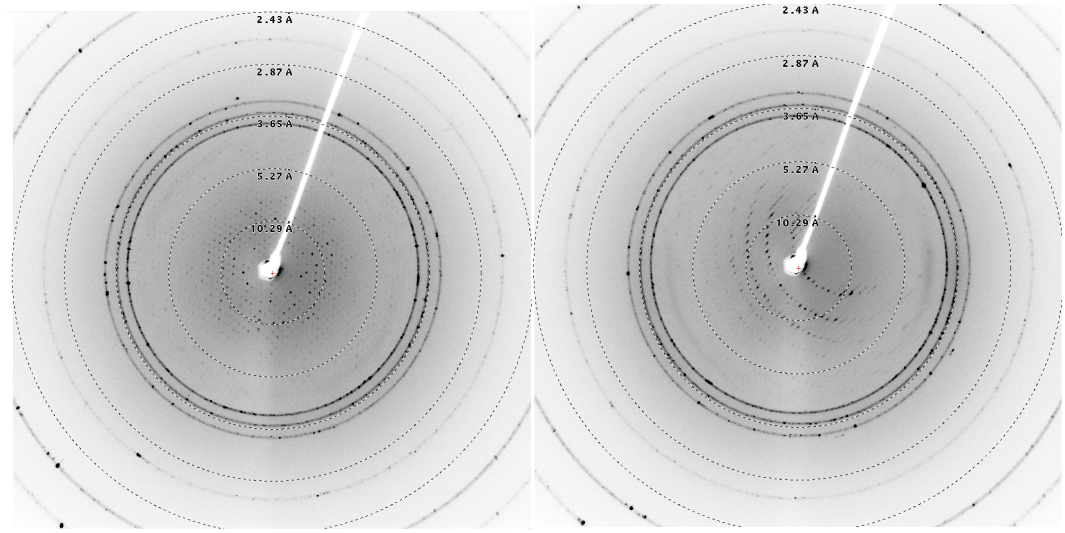



*Figure 3.11: In house diffraction images of apoPSMT1 at 0° and 90°*

### 3.3.5. Structural Determination and Analysis of PSMT1

The X-ray diffraction dataset for apoPSMT1 and PSMT1-SAM were indexed and integrated using Xia2 on the beamline to space group P $3_1$ 2 1.[125] The space group was later determined to be P $3_2$ 2 1 with a unit cell of a=111.94, b=111.94, c=305.2, α=90 °, β=90 °, γ=120 ° for apoPSMT1 and a=111.15, b=111.15, c=302.36, α=90 °, β=90 °, γ=120 ° for PSMT1 co-crystallised with SAM, with an X-ray wavelength of 0.97949 Å and 0.97626 Å, respectively. Data collection statistics are shown in Table 3.3. The apoPSMT1 crystal extended to 2.95 Å and the PSMT1-SAM crystal extended to 3.25 Å.

Both structures of PSMT1 were solved by molecular replacement using Molrep, with a subunit from a previously obtained 3.1 Å resolution unpublished structure of PSMT1 from the Davies lab by Wendy A. Offen, solved by molecular replacement, using residues 32-344 of caffeic acid *O*-methyltransferase molecule A (PDB:1KYZ) as a model for Phaser.[138] The model was built by iterative model building in COOT and refinement with REFMAC to 2.95 Å resolution for apoPSMT1 and 3.25 Å for PSMT1-SAM. The final model has an Rwork of 22 % and Rfree of 30 % for apoPSMT1, and Rwork of 17 % and Rfree of 25 % for PSMT1-SAM. The models contain four chains per asymmetric unit with electron density well defined for most amino acids.

| | apoPSMT1 | PSMT1-SAM |
|---|---|---|
| **Data collection** | | |
| Space group | P $3_2$ 2 1 | P $3_2$ 2 1 |
| Cell dimensions | | |
| a, b, c (Å) | 111.94, 111.94, 305.2 | 111.15, 111.15, 302.36 |
| (°) | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å) | 96.95(2.95) | 101(3.25) |
| Rmerge | 0.14(2.23) | 0.07(0.77) |
| Rpim | 0.05 (0.78) | 0.03(0.27) |
| I/σI | 9.0(1.1) | 23.5(3.1) |
| Completeness (%) | 100.0(100.0) | 100.0(100.0) |
| Redundancy | 9.6(9.9) | 5.1(4.9) |
| | | |
| **Refinement** | | |
| Resolution (Å) | 96.94-2.95 | 100.78-3.25 |
| No. reflections | 47543 | 34996 |
| Rwork / Rfree | 0.23/0.30 | 0.17/0.25 |
| No. atoms | | |
| Protein | 20768 | 10613 |
| Ligand/ion | 0 | 108 |
| Water | 0 | 0 |
| B-factors (Å$^2$) | | |
| Protein | 83.13 | 93.64 |
| Ligand/ion | n/a | 92.6 |
| Water | n/a | n/a |
| r.m.s. deviations | | |
| Bond lengths (Å) | 0.01 | 0.01 |
| Bond angles (°) | 1.7 | 1.8 |

*Table 3.3:Data collection and refinement statistics for PSMT1 structures*

#### *3.3.5.1. apoPSMT1*

apoPSMT1 crystals contained 4 molecules per asymmetric unit and extended to 2.95 Å. apoPSMT1 is a globular homodimer with an extensive N-terminal dimerisation domain comprising of residues 34-188, a putative substrate binding site (residues 189-222 and 334-353), and an C-terminal SAM/SAH binding domain (residues 229-325 and 358-390). There is a large cleft in-between the dimerisation/substrate domains and that of the SAM/SAH domain, indicative if an open conformation, Figure 3.12.

PSMT1 contains a SAH/SAM domain at the C-terminus containing the conserved Rossmann-like fold for nucleotide binding.[139,140] The Rossmann-like fold is a conserved super-secondary nucleotide binding structure consisting of an initial beta-alpha-beta topology with a Gly-x-Gly-x-Gly motif conserved in the loop between the first beta-strand and alpha helix and contains up to a seven-stranded beta-sheet which is usually parallel.[92,139] PSMT1 has a seven-stranded beta-sheet in mostly parallel orientation which are ordered 3214576 with stand 7 in antiparallel orientation, the strand closest to the N-terminus is numbered 1. This orientation is consistent with other plant *O*-methyltransferases such as caffeic acid *O*-methyltransferase and monolignol *O*-methyltransferase.[109,112,112,141]

The putative substrate binding site is made up of three alpha helixes. There are two alpha-helices in the substrate binding domain which sandwich the substrate in place. The front face is made up of alpha helix 11 (residues 189-222) and back face is made from alpha helix 15 (residues 334-353) which is in-between beta-strands 5 and 6 of the Rossmann-like fold. The binding pocket is capped at one end by an alpha-helix, 151-158, from the dimerisation domain and the other side is exposed to the active site cleft facing that of the SAM/SAH binding domain. As PSMT1 is in a classical open conformation without the methyl acceptor bound in the active site the catalytic residues are not discussed here but are extensively described in Chapter 4.

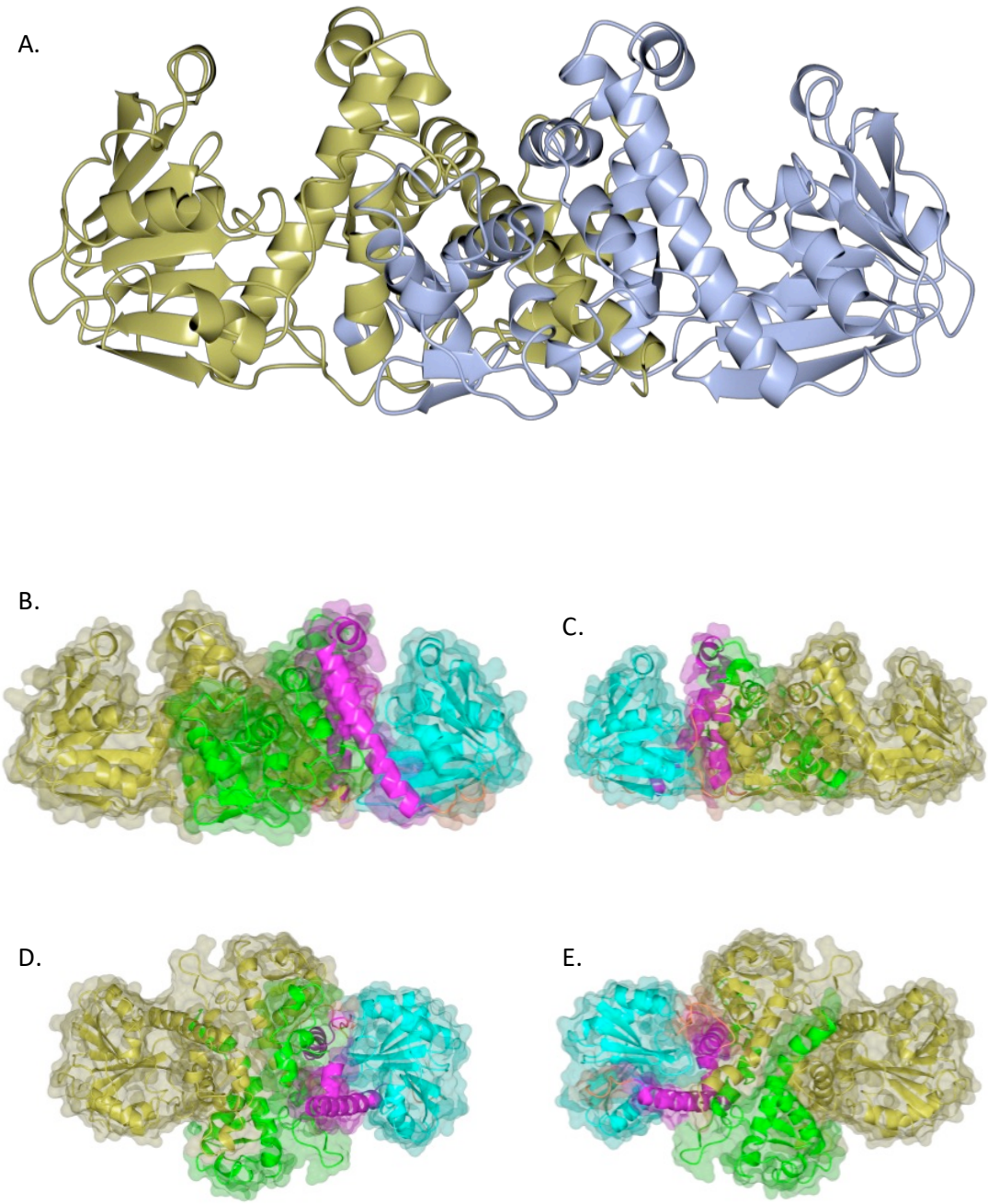



*Figure 3.12: A. Ribbon diagrams of apoPSMT1 in an open conformation, coloured by chain. B-E Ribbon diagrams of apoPSMT1 with surface displayed at 20 % transparency in different orientations. B front view, C, back view, D top view, E bottom view, orientation based on chains A to B from right to left with the putative active site cleft on the upper face. Chain A - gold. Chain B - coloured by domain, green - dimerisation domain, magenta - substrate binding domain, cyan - SAM/SAH binding domain*

Comparing the apoPSMT1 model to that other plant type-1 *O*-methyltransferases they have the same tertiary and quaternary structure, Figure 3.12 and Figure 1.13. The most structurally similar models are that of: *Medicago sativa* caffeic acid

3-*O*-methyltransferase (PDB:1KYW) with a root mean squared deviation (RMSD) of 1.22 Å over 327 equivalent residues with a 45 % sequence identity.[107] Perennial Ryegrass Lpomt1 (PDB:3P9C) also has an RMSD of 1.22 Å but over 322 equivalent residues with a sequence identity of 43 %.[112] Both the above structures are in an open conformation, caffeic acid 3-*O*-methyltransferase (1KYW) has both the methyl donor and acceptor bound, whereas Lpomt1 (3P9C) has only SAH bound in the active site. Calculating RMSD values for apoPSMT1 against plant *O*-methyltransferase models in a closed conformation such as *Clarkia breweri* monolignol *O*-Methyltransferase (PDB:3REO), monolignol 4-*O*-methyltransferase (PDB:5CVV) and perennial Ryegrass Lpomt1 (PDB:3P9K) produce RMSD values of 2.39 Å over 320 equivalent residues with 44 % sequence identity,2.40 Å over 318 equivalent residues with a sequence identity of 42 % and 2.05 Å over 320 equivalent residues with a 41 % sequence identity, respectively.[109,112] Therefore, apoPSMT1 is in an open conformation with a similar overall fold compared to that of other plant *O*-methyltransferases.

#### *3.3.5.2. PSMT1-SAH*

PSMT1 was co-crystallised with SAM by Wendy A. Offen but modelled by the author of this thesis produced a homodimer in the same open conformation as apoPSMT1 with an RMSD 0.41 Å over 341 residues, Figure 3.13. SAM is bound exclusively to the C-terminal Rossmann-like fold in an extended conformation. The nucleotide adenine group of SAM is surrounded by the non-polar side chains of Leu259, Met279, Phe280 and Trp398 and forms hydrogen bonds with the carboxyl group of Asp278 with adenine N6 and carboxyl group of Asp258 with N3 of adenine. The ribose ring is "clamped" in by hydrogen bonds between the hydroxyl groups on the ribose ring and the carboxylic acid side chain of Asp258, along with a non-polar glycine rich segment Gly235-Gly236-Gly237, Figure 3.14. This glycine rich segment also forms one wall to the carboxypropyl moiety of methionine of SAM with Trp296 on the opposing side. The amine group of methionine of SAM forms hydrogen bonds to that of the backbone oxygen on Lys292 and Gly235. The side chain of Lys292 reaches around to the carboxylic acid group of methionine making a hydrogen bond *via* its terminal (Nζ) amino group. The methyl group of the sulphonium centre is positioned by Trp293 and

Asp297 which point it into the open cleft between the SAM binding domain and that of the putative substrate binding pocket.

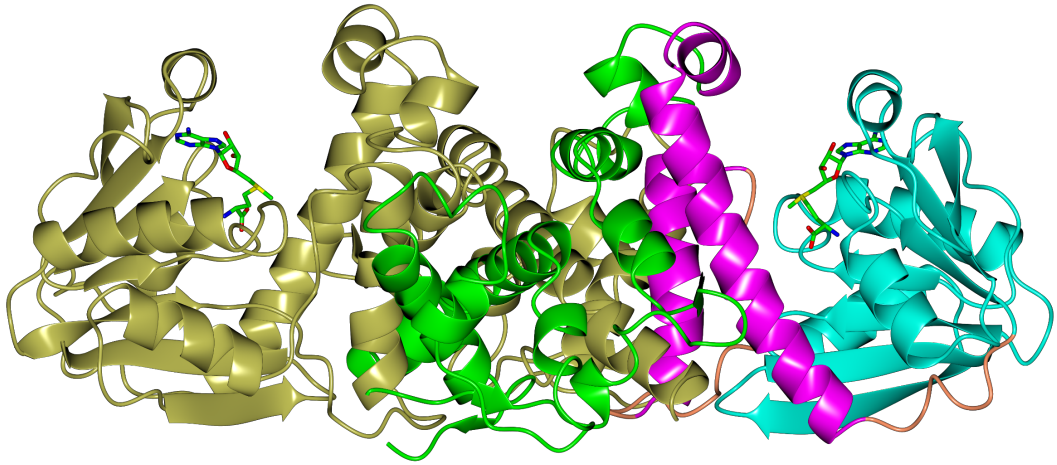


*Figure 3.13: Ribbon diagram of PSMT1 co-crystallised with SAM in an open conformation. Chain A gold. Chain B coloured by domain, green dimerisation domain, magenta substrate binding domain, cyan SAM/SAH binding domain. SAM represented by cylinders, green carbon, red oxygen, blue nitrogen, yellow sulphur*

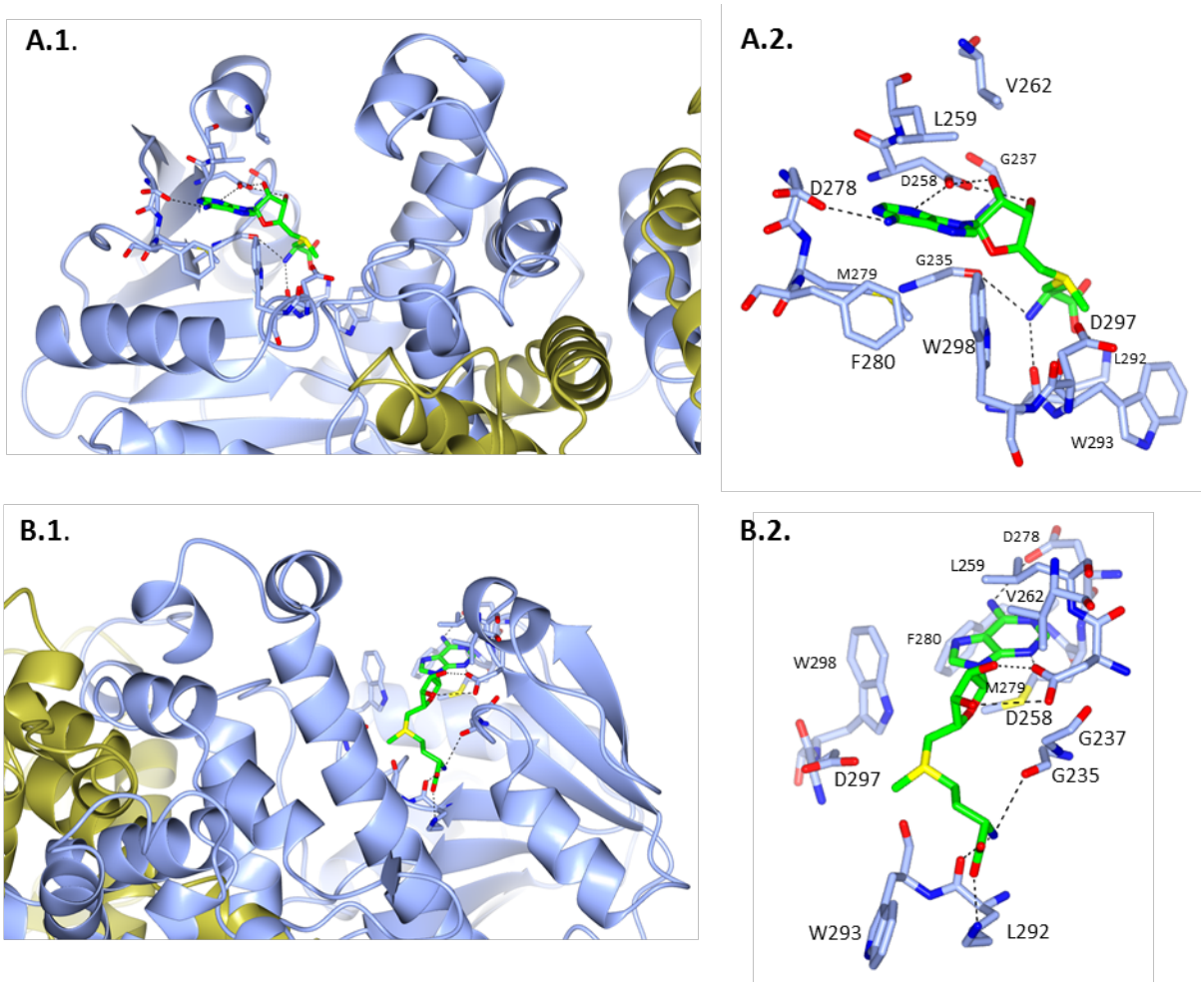



*Figure 3.14: 3.2 Å model of PSMT1-SAM showing SAM binding site. A. front view and B. rear view. 1. With peptide shown as ribbon diagram, 2. Without peptide ribbon diagram. PSMT1 represented as a ribbon diagram coloured by chain. SAM represented by cylinders coloured by atom type, green-carbon, yellow-sulphur, red-oxygen, blue-nitrogen. Residues involved in SAM binding shown as cylinders with carbon atoms coloured by chain and non-carbon atom coloured as described above*

## 3.4. Conclusion

PSMT1 has been produced heterologously in *E. coli* along with an N-terminal purification-solubility tag, Figure 3.4. PSMT1 was purified by nickel affinity on-column cleavage utilising 3C protease to liberate PSMT1 from it purification-solubility tag, resulting in a 3 residue N-terminal extension. Subsequent desalting and anion exchange chromatography resulted in pure soluble homogeneous PSMT1 as analysed by SDS-PAGE, Figure 3.4, native-PAGE, Figure 3.5, and SEC-MALLS, Figure 3.6. The calculated molecular weight by mass-spectrometry was as expected and SEC-MALLS showed that PSMT1 is a homodimer in solution, Figure 3.6. The protein is active against scoulerine producing tetrahydrocolumbamine utilising the methyl donor SAM. Direct measurement of the product tetrahydrocolumbamine by UPLC-MS/MS proved to be a valid and reproducible method for determining kinetic parameters, Table 3.1 and Table 3.2. Michaelis-Menten kinetics of PSMT1 with scoulerine and SAM as substrates resulted in the convergence of $V_{max}$ of ~ 500 nmol $min^{-1}$ $mg^{-1}$ of protein with $K_m$ values of 0.25 µM and 8.76 µM, respectively. The Hill plot analysis showed positive cooperativity between the two active site of the homodimer, with a Hill coefficient of 1.5 with scoulerine as the substrate. A Hill coefficient of 0.8 is calculated for SAM as a substrate for PSMT1 indicating negative cooperativity. Positive cooperativity is an advantageous trait for an enzyme to have as it allows the enzyme to have a rapid response to a small increase in substrate concentration around that of the $K_{prime}$. In the case of PSMT1 the $K_{prime}$ was calculated as 0.106 µM for scoulerine. The advantage of having negative cooperativity is that the enzyme will have a velocity which will be less responsive to changes in concentrations around the $K_{prime}$. The data obtained from the Hill plot analysis suggest negative cooperativity for PSMT1 with SAM as the substrate, but this could be due to the inhibitory effects of SAH.

This would suggest that PSMT1 has evolved to have a large change of velocity to a relatively small change in scoulerine concentration at around 0.1 µM. PSMT1 is acting like a scoulerine 'scavenger', converting scoulerine to tetrahydrocolumbamine as its synthesised. On the other hand, SAM is an abundant vital co-substrate whose

concentration will remain relatively constant within cells. This puts less pressure on PSMT1 to optimise for SAM binding resulting in a higher affinity constant.

The X-ray crystal structure of apoPSMT1 is a homodimer with a distinct SAM binding domain, dimerisation domain and putative substrate binding domain. The SAM binding site was confirmed by co-crystallisation with SAM. The inability to generate a protein crystal with both SAH and the methyl acceptor, scoulerine, bound in the same crystallisation conditions suggests that there is a conformational change upon binding of both substrates, such as that described for caffeic acid *O*-methyltransferase and other plant *O*-methyltransferases. This suggests that scoulerine binds to PSMT1 through an induced fit mechanism which alters the conformation of the binding pocket. Such an induced fit mechanism may have prevented crystals of PSMT1 forming with SAH and scoulerine. Further experiments would need to be carried out to obtain a tertiary complex structure and this will be discussed in Chapter 4.

# 4. PSMT1 Surface Entropy Reduction

## 4.1. Introduction

Chapter 3 described the successful purification and crystallisation of PSMT1 in an open "apo" form and with SAM bound to 2.95 Å and 3.25 Å, respectively. These relatively low-resolution structures allowed the characterisation of the protein fold in an open conformation and residues involved in SAM/SAH binding but not the methyl acceptor binding pocket because the methyl acceptor is not bound. To fully understand the substrate binding pocket the substrate or product would need to be bound in a closed conformation where the active site is fully formed. Attempts to generate crystals of PSMT1 co-crystallised with SAH/SAM and scoulerine proved unsuccessful. In an attempt to obtain a closed crystal form of PSMT1 the surface entropy reduction (SER) method was used.

Crystallisation of proteins that diffract to a high resolution, greater than 2 Å has been described as 'the rate limiting step' in structure determination by X-ray crystallography.[121,122] It has been estimated that on average less than 30 % of soluble globular protein expressed in *E. coli* will crystallise, with only a small portion providing high resolution structures.[142] Surface entropy reduction is a technique in which the high entropy residues of lysine and glutamic acid, which are generally located on the surface of a protein, are mutated by site directed mutagenesis to the low entropy residue alanine.[121,123] The technique was first proposed by Derewenda in 2004.[121]

The Eisenberg group (University of California, California, USA) generated a web based server to predict clusters of two or more high entropy residues to mutate to alanine.[122] The webserver predicts the secondary structure, generates a residue entropy profile and generates a sequence alignment. Utilising this information, clusters of high entropy residues surrounded by low entropy patches are suggested for mutation. The generated area of low entropy will remove the 'entropy shield' allowing new crystal contacts to form. This technique has been utilised to crystallise a large range of proteins that otherwise would not crystallise in their wild type

form.[122] There are several caveats to the surface entropy reduction technique. The SER mutations introduced could affect the protein's function, and the mutation of residues containing charged side chains to the hydrophobic side chain of alanine could reduce the solubility and stability of the protein. These factors were investigated for the PSMT1 surface entropy reduction mutants generated.

This chapter describes PSMT1 surface entropy reduction (SER) experiments along with the characterisation of the SER proteins. Subsequent crystallisation experiments resulted in high-resolution structures in a closed conformation with SAH and scoulerine or tetrahydrocolumbamine bound in the active sites.

## 4.2. Materials and Methods

### 4.2.1. Surface Entropy Reduction

Surface entropy reduction is achieved by using the UCLA MBI surface entropy reduction prediction (SERp) server.[122] The SERp server carries out a series of processes giving a score to each residue for a given protein sequence. The secondary structure is predicted by PSIRED, a conformational entropy profile is generated and finally PSI-BLAST is utilised to predict residue conservation. Lysine and glutamic acid have been shown to localise predominantly on protein surfaces and interfere with protein-protein interfaces.[123,124] A typical high scoring residue will be a glutamic acid or lysine on a surface exposed loop which is non-conserved. Finally, the values are plotted along the sequence and clusters of up to three high scoring residues surrounded by low scoring residues are suggested for mutation.

### 4.2.2. Site-Directed Mutagenesis

Site-directed mutagenesis of pETFPP-3-PSMT1 was carried out using the whole plasmid mutagenesis method. Overlapping oligonucleotides (primers) containing the desired mutations, flanked on either side with 15-20 complimentary nucleotides were generated with an annealing temperature of 78-79 °C. 10 ng of purified plasmid DNA was mixed with forward and reverse primers to a final concentration of 0.3 µM, water and KOD hot start master mix to a total volume of 50 µL, Table 2.3. KOD hot start master mix is a 2x mixture containing KOD hot start polymerase, dinucleotides

and reaction buffer containing $MgSO_4$. The whole plasmid was then amplified by PCR typically using the thermal-cycling conditions in Table 2.4. Extension time is based on KOD activity of 1 min 10 seconds per kilo base pair at 68 °C, where optimal proofreading activity is obtained. Following the PCR, 1 µL of DpnI restriction enzyme (20 units $\mu L^{-1}$) was added to the reaction mixture and incubated at 37 °C for at least one hour to digest the methylated template DNA. *E. coli* XL10 Gold cells were then transformed by the mutated plasmid DNA and incubated on LB agar plates containing 50 mg $mL^{-1}$ kanamycin antibiotic to select for transformed cells. Typically, three colonies exhibiting antibiotic resistance were selected, plasmid DNA was extracted from overnight cultures and the DNA sequenced. Only plasmids that carried the desired mutation and no other mutations were transformed into *E. coli BL21 (DE3)* cells for expression.

### 4.2.3. Expression and Purification

PSMT1 SER mutants were expressed from a modified pETFPP-3-PSMT1 as a N-terminally hexa-histidine tagged GST fusion protein with a 3C-protease cleave site linker (6xHis-GST-r3CP-PSMT1 SER Cx, x= 2 or 3 or 2&3) in *E. coli BL21 (DE3)* cells for in vitro characterisation (as described in Section 2.2.1, pg. 54). The predicted molecular weights of the full-length construct along with the cleaved components are shown in Table 4.1, calculated utilising ProtParam.[120]

| SER mutant | 6xHis-GST-r3CP-PSMT1 SER Cx | 6xHis-GST | PSMT1-SER Cx |
|---|---|---|---|
| Cluster 2 | 70356.2 Da | 27645.0 Da | 42729.2 Da |
| Cluster 3 | 70355.2 Da | 27645.0 Da | 42728.2 Da |
| Cluster 2&3 | 70241.0 Da | 27645.0 Da | 42614.0 Da |

*Table 4.1: PSMT1 surface entropy reduction mutant proteins predicted molecular weights based on sequence for full length construct (6xHis-GST-r3CP-PSMT1-SER Cx) and its 3C protease cleaved products (6xHis-GST and PSMT1-SER Cx)*

For each SER mutant, a single colony from a Luria-Broth (LB) agar plate containing 50 µg $mL^{-1}$ kanamycin was picked and used to inoculate 10 mL of LB medium with 50 µg

$mL^{-1}$ kanamycin. This was incubated at 37 °C overnight. 5 mL of the overnight culture was used to inoculate 500 mL Terrific Broth (TB) medium with 50 µg $mL^{-1}$ kanamycin in 2.5 L baffled shaker flasks. The 500 mL culture was incubated at 37 °C, 200 rpm until cell growth entered the exponential phase indicated by an absorbance at 600 nm ($A_{600nm}$) of 0.6 – 1.0, also known as optical density (OD). Once the desired OD was achieved the cultures were cooled to 16 °C and expression of the gene encoding 6xHis-GST-r3CP-PSMT1 SER Cx was induced by addition of isopropyl-β-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The culture was incubated at 16 °C, 200 rpm overnight, for at least 18 hours, for the overexpression of the fusion protein. Cells were harvested by centrifugation at 5000 rpm at 4 °C for 30 minutes in a Beckman Avanti J-HC equipped with a JLA 8.1000 rotor. The supernatant was discarded and the cell pellet was either resuspended in 50 mM Tris pH 8.0, 500 mM NaCl, 10 mM β-mercaptoethanol, 20 mM imidazole with DNaseI and a protease inhibitor cocktail tablet (cOmplete, Roche) for purification or stored at -20 °C.

The cells were lysed by sonication utilising a large probe with the cells solution on ice to reduce protease activity. A programme of short 15 second pulses with 15 seconds resting time was used to reduce heat induced protein damage. The cell debris were removed by centrifugation at 18,000 rpm for 45 minutes at 4 °C in a Sorvall SS34 rotor, followed by filtration using a 0.22 µM syringe driven filter.

PSMT1 SER Cx was purified using a multi-step purification protocol, utilising nickel affinity chromatography on-column cleavage, desalting and ion exchange. Nickel affinity chromatography was used to capture the 6xHis-tagged fusion protein, then Human rhinovirus-3C protease cleaved the fusion protein releasing PSMT1 SER Cx from the 6xHis-GST purification/solubility tag. Desalting removed the low molecular weight solutes such as NaCl and imidazole from the protein solution. Finally, anion exchange chromatography was used to remove any remaining contaminating proteins.

The clarified cell lysate was loaded onto a 5 mL HisTrap nickel affinity column (GE healthcare) which was pre-equilibrated with nickel affinity buffer A (50 mM Tris pH

8.0, 500 mM NaCl, 10 mM β-mercaptoethanol, 20 mM Imidazole). 6xHis-GST-r3CP-PSMT1 SER Cx would bind to the $Ni^{2+}$ *via* the 6xHis-tag while host proteins will not bind and were removed by flowing nickel affinity buffer A across the column until the $A_{280}$ response returned to baseline. The column was then treated with 1 column volume (CV) of 0.25 mg $mL^{-1}$ Human rhinovirus-3C protease (3CP) (6xHis-Maltose Binding Protein-3CP, 64 kDa) and incubated overnight at 4 °C. The 3CP would cleave 6xHis-GST-r3CP-PSMT1 SER Cx into two at the 3CP recognition site (r3CP): the purification/solubility tag (6xHis-GST) and the now free protein of interest, PSMT1 SER Cx with a 3 residue N-terminus overhang. The cleaved PSMT1 SER Cx was washed off the column with nickel affinity buffer A. The purification/solubility tag and 3CP remain bound to the column which were subsequently eluted with a nickel affinity buffer B (50 mM Tris pH 8.0, 500 mM NaCl, 10 mM β-mercaptoethanol, 0.5 M Imidazole). Fractions with a high $A_{280}$ response were analysed by SDS-PAGE and fractions containing bands corresponding to PSMT1 SER Cx ca. 43 kDa were pooled. The fractions containing PSMT1 SER Cx were pooled then desalted using a HiPrep 26/60 Desalting column (GE Healthcare) pre-equilibrated with ion exchange buffer A (50 mM Tris pH 8.0, 10 mM β-mercaptoethanol).

Desalted PSMT1 SER Cx solution was then loaded onto a HiScreen Q HP 4.7 ml (GE Healthcare) anion exchange column pre-equilibrated with ion exchange buffer A (50 mM Tris pH 8.0, 10 mM β-mercaptoethanol). The column was washed with 10 % ion exchange buffer B (50 mM Tris pH 8.0, 10 mM β-mercaptoethanol, 1 M NaCl) to remove weakly bound proteins. Bound PSMT1 SER Cx was eluted with a 10-50 % buffer B gradient over 10 CVs and a 100 % buffer B step was utilised to elute strongly bound proteins and regenerate the column.

The fractions containing PSMT1 SER Cx, determined by SDS-PAGE analysis, were pooled and buffer exchanged into 10 mM Tris, 5 mM TCEP pH 8.0 utilising repeated concentration and dilution of the protein solution using a 30 kDa centrifugal concentrating unit. Finally, the protein was concentrated to 15 mg $mL^{-1}$ for crystallisation trials or stored at -80 °C for future experiments. The concentrations of PSMT1 SER Cx was determined using their calculated extinction coefficients,

calculated with the ProtParam tool (ExPASy), shown in Table 4.2.[120] All steps were analysed by SDS-PAGE for purity.

| SER mutant | PSMT1-SER Cx extinction coefficients |
|---|---|
| Cluster 2 | 0.700 $M^{-1}$ $cm^{-1}$ |
| Cluster 3 | 0.700 $M^{-1}$ $cm^{-1}$ |
| Cluster 2&3 | 0.702 $M^{-1}$ $cm^{-1}$ |

*Table 4.2: PSMT1 surface entropy reduction mutants calculated extinction coefficients based on sequence*

### 4.2.4. Size exclusion multi angle laser light scattering

Size exclusion multi angle laser light scattering (SEC-MALLS) was carried out in the Bioscience Technology Facility (TF) at the University of York using a Shimadzu LC-20AD liquid chromatography machine linked to a Shimadzu SIL-20A auto-sampler. 100 µL of purified protein samples of 2 mg $mL^{-1}$ in 10 mM Tris pH 8.0 were injected onto a Superdex 200 10/300 GL column (GE Healthcare) pre-equilibrated with 10 mM Tris pH 8.0, 150 mM NaCl. Measurements were taken using a Shimadzu SPD-20A UV detector, a Wyatt Optilab rEX refractive index monitor, a Wyatt DAWN HELEOS light scattering detector and analysed using Wyatt ASTRA software package. BSA was ran as a control.

### 4.2.5. Thermal Shift Assay- Thermofluor

1 mg $mL^{-1}$ PSMT1 SER mutants in 10 mM Tris pH 8.0, 5 mM TCEP which were purified by nickel affinity on-column cleavage followed by desalting and anion exchange were mixed with 5x SYPRO orange dye at a 1:1 ratio, totalling 30 µL in a 96 well plate. 5x SYPRO orange dye was prepared by diluting 5000x SYPRO orange dye in DMSO to 5x in 10 mM Tris pH 8.0, 5 mM TCEP. Protein samples were analysed alongside that of a buffer only sample as a control. The samples were heated at 2 °C $minute^{-1}$ and fluorescence recorded every 30 seconds, excitation 492 nm, emission 610 nm. The fluorescence intensity was normalised and plotted against temperature. The melting

point of the protein was determined at the midpoint of the unfolding event characterised by an increase in intensity.

### 4.2.6. Molecular Weight Determination

The molecular weight of PSMT1 SER mutants were determined by the Molecular Interactions Laboratory at the Bioscience Technology Facility based at the University of York as a service by ESI-MS. The protein samples were buffer exchanged in to 2 mM Tris pH 8.0 using centrifugal concentrator units and submitted at a concentration of 5-10 mg $ml^{-1}$. The calculated molecular weights were corrected in reference to the external standard myoglobin.

### 4.2.7. Activity Assay

The activity of PSMT1 was investigated using scoulerine as the substrate, Figure 3.1, by UPLC-MS/MS. Enzymatic assays were conducted using scoulerine obtained from Prof. Peter J. Schammells (Monash University, Australia) and SAM (Santa Cruz Biotechnology). Each reaction was carried out in a total volume of 11.5 mL of 100 mM glycine-NaOH pH 9.0, 25 mM sodium ascorbate, 1 mM β-mercaptoethanol, 100 µM SAM while varying scoulerine from 0.05 µM to 10. The reactions were incubated at 37 °C in a water bath and the reactions were initiated by the addition of 0.5 mL of 1 µg $mL^{-1}$ of purified PSMT1 from on-column cleavage in nickel affinity buffer A. Protein concentration of a ca. 1 mg $mL^{-1}$ was prepared by monitoring $A_{280}$ and diluted to ca. 1 µg $mL^{-1}$. The actual concentration of the ca. 1 mg $ml^{-1}$ solution was determined using the Bradford protein assay as described in (Section 1.2.4.2), this value was used to calculate the reaction rate. 1 mL time points were taken over a maximum of 5 minutes and quenched with an equal volume of methanol. The quenched samples were subsequently dried using a SpeedVac centrifugal evaporator (Genevac) and then resuspended in 100 µL of 10 % acetic acid to give a 10-fold increase in concentration for quantification by UPLC-MS/MS.

The enzymatic assay samples were analysed alongside standards of scoulerine and tetrahydrocolumbamine (Santa Cruz Biotechnology) utilising a modified method by Winzer *et al.* (2012).[48] A Waters Acquity UPLC system (Waters Ltd.) equipped with a

Acquity UPLC BEH C18 1.7 µm 2.1 x 100 mm column (Waters Ltd.) incubated at 60 °C was utilised to separate the alkaloids by reverse phase chromatography. The mobile phases consisted of A: 10 mM ammonium bicarbonate pH 10.2 and B: Methanol with a flow rate of 0.5 mL minute$^{-1}$. 2 µL of the sample was injected onto the column and the alkaloids eluted with a gradient of solvent B over 5 minutes as shown in Figure 2.3 and Table 2.6. Scoulerine and tetrahydrocolumbamine were quantified using a Thermo TSQ Endura triple quadrupole Mass Spectrometer (MS) in ESI positive mode using selective reaction monitoring (SRM), to give greater sensitivity over the Orbitrap utilised in the published method. SRM is a highly sensitive and selective method of detection with a high signal to noise ratio. This is achieved by the first quadrupole selecting for the mass to charge for either scoulerine (328 m/z) or tetrahydrocolumbamine (342 m/z). The second quadrupole fragments the precursor ion by collision induced dissociation before the third, and final quadrupole is set to select the fragmentation product which results in the highest intensity signal. In both cases the fragment product which gave the highest signal intensity had a m/z of 178. Compound optimisation was carried out for both scoulerine (Appendix 8.2) and tetrahydrocolumbamine (Appendix 8.3), separately, by directly injecting them into the MS using a syringe driver. Scoulerine and tetrahydrocolumbamine were detected over a time window of 0.7 minutes, 1.85 - 2.55 minute and 2.55 - 3.25 minute, respectively. The time windows were selected so only one fragmentation product was detected at a time, to give the greatest response possible.

The concentration of tetrahydrocolumbamine was calculated for each sample based on a calibration cure of known concentrations vs response. Concentration against time was plotted and an initial rate calculated for each reaction. The specific rate of reaction in nmol min$^{-1}$ mg$^{-1}$ of protein was plotted against the concentration of substrate being varied. Each point represents the mean specific rate of reaction at each substrate concentration over three independent replicates with standard error of the mean shows as vertical error bars. $K_m$ and $V_{max}$ were calculated by direct fit of the Michaelis-Menten equation in GraphPad software (Prism).

### 4.2.8. Crystallisation

PSMT1 SER mutant crystallisation trays were setup using a 1:1 ratio of 15 mg $mL^{-1}$ protein in 10 mM Tris, 5 mM TCEP pH 8.0, utilising the vapour diffusion method in MRC 96 well dual sitting drop crystallisation plates, and incubated at 20 °C in a Rigaku Minstrel HT UV crystal imager (Rigaku). Drops containing 150 nL of protein solution and 150 nL of crystallisation screen condition were dispensed by a mosquito liquid handling robot (TTP Labtech). The reservoir contained 46 µL of commercially bought crystallisation screen as an initial screen such as PACT premier HT-96 (Molecular Dimensions), JCSG-plus (Molecular Dimensions), SaltRx (Hampton Research), Index (Hampton Research), and Structure Screen 1 and 2 HT-96 (Molecular Dimensions). Co-crystallisation experiments were carried out with 0.7 mM SAH and 2 mM Scoulerine. Conditions generating crystals were optimised in 48 well sitting drop plates. Optimisation consisted of varying the components within the hit reservoir, such as pH and the percipient concentrations, to obtain better crystals for X-ray diffraction studies. 48 well sitting drop trays were set-up manually using electronic pipettes, the same ratio of reservoir solution to protein solution were used with a total volume of 1 µL, against 100 µL of reservoir solution and stored at the same temperature. Crystals were tested in house using a Rigaku MicroMax-007 HF X-ray Generator and Rigaku R-Axis IV++ image plate detector. 0 ° and 90 ° X-ray diffraction images are taken with an oscillation of 0.5 °, 5 minute exposures to 2.5 Å resolution. Crystals which diffracted well were sent to Diamond Light Source for collection of a full data set.

## 4.3. Results and Discussion

### 4.3.1. Design and Generation of Surface Entropy Reduction Mutants

The SERp server proposed three clusters of high entropy residues surrounded by low entropy regions that are non-conserved for mutation, along with a SERp score, Table 4.3. The SERp score is a numerical score based on prediction confidence. The three suggested clusters were compared to that of the previously described apoPSMT1 structure, Figure 4.1. Cluster one (Glu174, Glu175 and Lys176) is involved in the dimer interface between the two polypeptide chains, for this reason cluster one was not generated. Cluster two (Lys114 and Lys115) and cluster three (Glu128 and Lys239) are on a loop which is missing from the structural model, most likely due to its flexibility. Clusters two and three were generated independently of each other along with a dual cluster mutant containing all four residue mutations (Lys114Ala, Lys115Ala, Glu128Ala and Lys129Ala.

| Cluster | Mutations | SERp Score |
|---|---|---|
| **#1** | Glu174Ala<br>Glu175Ala<br>Lys176Ala | 6.88 |
| **#2** | Lys114Ala<br>Lys115Ala | 5.05 |
| **#3** | Glu128Ala<br>Lys129Ala | 4.38 |

*Table 4.3: Proposed surface entropy reduction mutations for PSMT1*

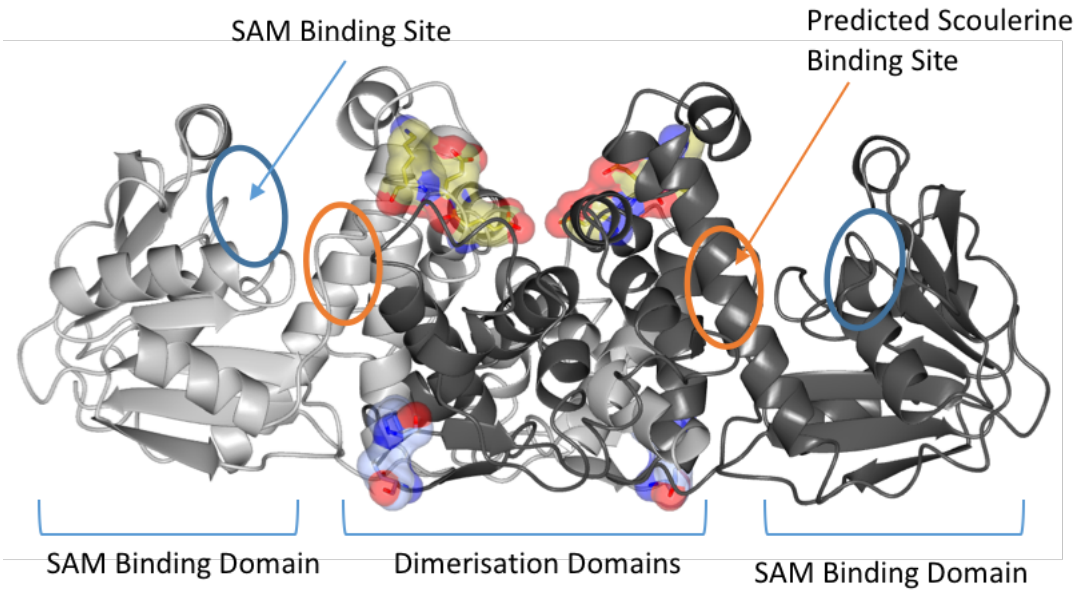



*Figure 4.1: Predicted surface entropy reduction clusters mapped onto the apoPSMT1 dimeric structure. ApoPSMT1 represented as a ribbon diagram, coloured by chain. Cluster 1 residues are represented as gold cylinders and as transparent surface. Cluster 2 and 3 are on a missing loop between residues Thr113 and Lys132 which are represented as ice blue cylinders and as transparent surface. SAM binding sites are indicated by blue circles and the predicted scoulerine binding sites are indicated by orange circles*

Cluster three was generated using a one-step site directed mutagenesis experiment utilising one set of oligonucleotide primers consisting of three nucleotide mutations a383c, a385g, a386c, Table 4.4, the resulting plasmid was named pETFPP-3-PSMT1 SER C3.

Cluster 2 proved more problematic to generate and required a two-step site-directed mutagenesis experiment, utilising two sets of primers sequentially, Table 4.4. PSMT1 Lys114Ala was initially generated by two nucleotide mutations, a340g and a341c. The second step was the introduction of Lys115Ala into pETFPP-3-PSMT1 Lys114Ala with a further two nucleotide mutations, a343g and a344c generating the plasmid pETFPP-3-PSMT1 SER C2. The double cluster mutant was generated by introducing cluster 3 mutations into pETFPP-3-PSMT1 SER C2 generating pETFPP-3-PSMT1 SER C2&3. All plasmids generated were confirmed by sanger sequencing utilising PSMT1 sequencing primer 1, Table 4.4.

| SER Cluster 3 | |
|---|---|
| Glu128Ala, Lys129Ala | A383C A385G A386C |
| Forward | 5'-gaggagatgatgtagtagtacatg**c**g**gc**gctttatgggttaacaaattcgtc-3' |
| Reverse | 5'-gacgaatttgttaacccataaagc**gc**c**g**catgtactactacatcatctcctc-3' |
| **SER Cluster 2** | |
| Lys114Ala | A340G A341C |
| Forward | 5'-gttctattctttctgtttctactaca**gc**aaaatcaatcaacagaggaggagatg-3' |
| Reverse | 5'-catctcctcctctgttgattgattttt**gc**tgtagtagaaacagaaagaatagaac-3' |
| Lys115Ala | A343G A344C |
| Forward | 5'-ttctttctgtttctactacagca**gc**atcaatcaacagaggaggagatg-3' |
| Reverse | 5'-catctcctcctctgttgattgat**gc**tgctgtagtagaaacagaaagaa-3' |
| **PSMT1 Sequencing Primer 1** | |
| | 5’-TCGGAACTGACGCAAAAGT-3’ |

*Table 4.4: Oligonucleotide primers for the generation of PSMT1 surface entropy reduction mutants and sequencing primer 1. Mutations are shown in bold*

### 4.3.2. Expression and Purification

*E. coli* BL21(DE3) cells were transformed individually by the SER mutant plasmids and the genes expressed. Gene expression and purification was carried out in the same manner as the WT enzyme. PSMT1 SER 2, 3, 2&3 proteins were purified and analysed by SDS-PAGE, Figure 4.2. All three SER proteins behaved similar to that of WT PSMT1, resulting in pure protein at 15 mg $mL^{-1}$ in 10 mM Tris pH 8.0, 5 mM TCEP.

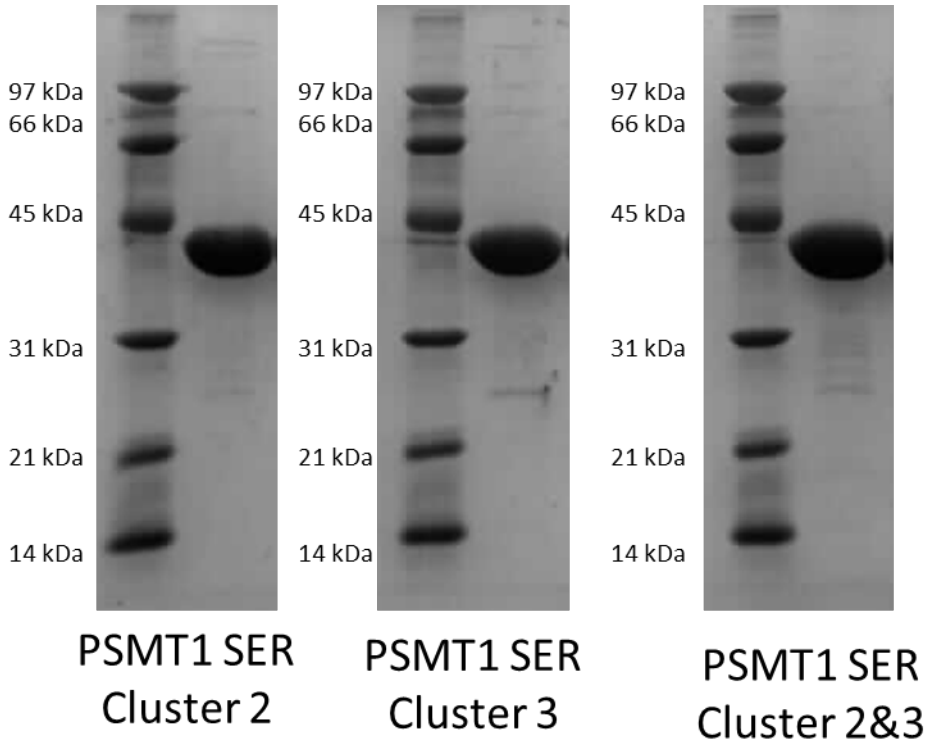



*Figure 4.2: SDS-PAGE analysis of PSMT1 surface entropy reduction cluster mutant proteins. 7.5 µg of purified protein loaded per well alongside molecular weight markers*

### 4.3.3. Characterisation of Surface Entropy Reduction Mutants

#### *4.3.3.1. Molecular Weight Determination*

15 mg $mL^{-1}$ of each PSMT1 SER mutant in 10 mM Tris, 5 mM TCEP purified by on-column cleavage and anion exchange chromatography was buffer exchanged into 5 mM Tris pH 8.0 using centrifugal concentrators. The resulting protein solutions were submitted to the Bioscience Technology Facility based at the University of York for molecular weight determination by ESI-mass-spectrometry. The expected molecular weight of the purified PSMT1 SER mutants as calculated by ProtParam[120] based on the amino acid composition are shown in Table 4.5, along with the corrected measured molecular weights determined by mass-spectrometry. They are all within 1 Da of the expected molecular weights, which is consistent with the PSMT1 SER mutants all having been successfully purified without degradation taking place.

| PSMT1 SER Mutants | Predicted Molecular weight | Measured Molecular weight |
|---|---|---|
| **Cluster 2** | 42729.2 | 42729.6 |
| **Cluster 3** | 42728.2 | 42727.7 |
| **Cluster 2&3** | 42614.0 | 42614.6 |

*Table 4.5: Molecular weight determination of PSMT1 surface entropy reduction mutants by ESI-MS*

#### *4.3.3.2. Size Exclusion-Multi-Angle Laser Light Scattering*

In order to determine if the surface entropy mutations had interfered with the dimerisation of PSMT1 size exclusion-multi angle laser light scattering (SEC-MALLS) was carried out for PSMT1 SER C2 and PSMT1 SER C2&3. PSMT1 SER C3 was not analysed due to limited availability of the protein. 100 µL of a 2 mg $ml^{-1}$ solution of the PSMT1 SER mutants in 10 mM Tris pH 8.0, 5 mM TCEP was injected onto a Superdex 200 10/300 GL column (GE Healthcare) pre-equilibrated with 10 mM Tris pH 8.0, 150 mM NaCl, Figure 4.3. The elution UV traces show one major monodispersed peak with retention times of 14 minutes and an observed mean molecular weight of 85140 Da for PSMT1 SER C2 and 85700 Da for PSMT1 SER C2&3 over their respective peaks. The calculated molecular weights of the purified PSMT1 SER mutants' gene products are 42843.4 Da and 42614.0 Da, respectively. The observed molecular weights in solution are double that of the expected molecular weights based on their sequence. This shows that the mutations introduced to PSMT1 to generate PSMT1 SER C2 and PSMT1 SER C2&3 have not effected the dimerisation state in solution.

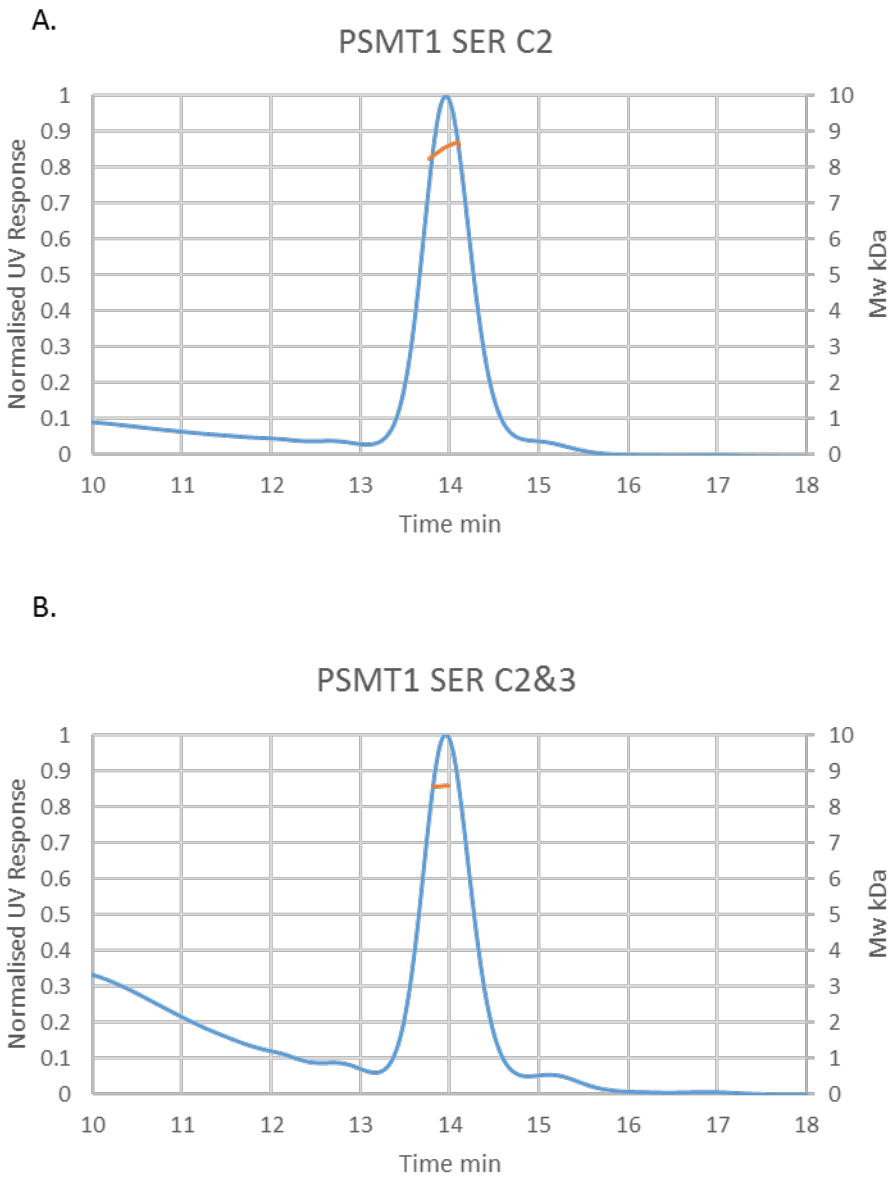



*Figure 4.3: SEC-MALLS results for PSMT1 surface entropy reduction mutants. A. PSMT1 SER C2 with a mean molecular weight of 85140 Da and B. PSMT1 SER C2&3 with a mean molecular weight of 85700 Da. Blue line represents normalised UV response, while the orange line represents calculated molecular weight across the peak*

#### *4.3.3.3. Thermal Shift Assay-Thermofluor*

Melting temperature ($T_m$) of PSMT1 SER C2 and PSMT1 SER C2&3 were measured by the thermofluor thermal shift assay. 1 mg $ml^{-1}$ of protein in 10 mM Tris pH 8.0, 5 mM TCEP that had been stored at -80 °C was mixed 1:1 with 5xSYPRO orange dye with a total volume of 30 µL, Figure 4.4. PSMT1 SER C2 and PSMT1 SER C2&3 both denature in one single denaturing event with melting temperatures of 49°C and 50 °C, respectively. PSMT1 WT has a melting temperature of 50 °C, Figure 3.7, therefore the

mutations incorporated to generate the surface entropy reduction mutants have little to no effect on the stability of the enzyme.

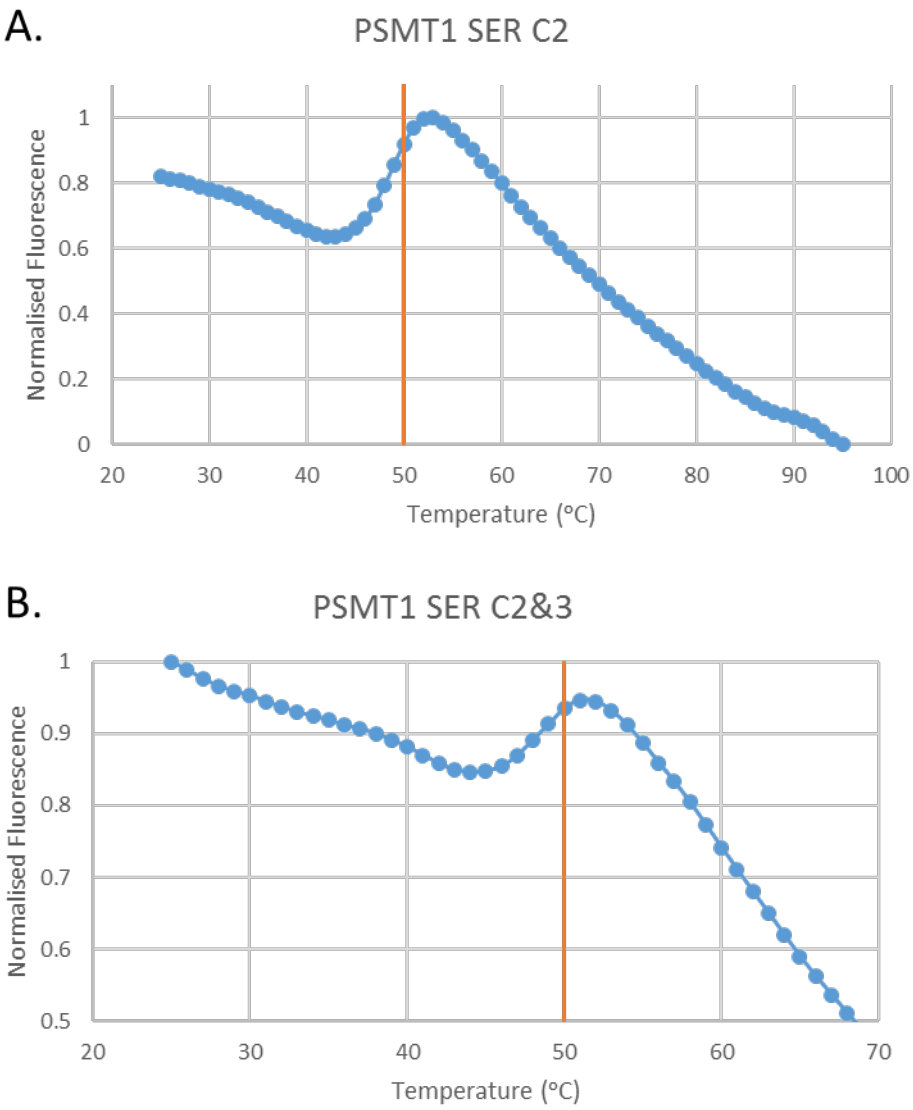



*Figure 4.4: Thermal denaturation assay of PSMT1 surface entropy reduction mutants utilising the Thermofluor method. A. PSMT1 SER C2 and B. PSMT1 SER C2&3. Graph shows normalised fluorescence response against time (Blue) and the calculated melting temperature (orange)*

#### *4.3.3.4. Initial Crystallisation Trials of SER Mutants*

Initial crystallisation trials for the PSMT1 SER mutants were carried out in MRC 96 well dual sitting drop crystallisation plates using the commercially available crystallisation screens PACT premier HT-96 (Molecular Dimensions) and JCSG-PLUS™ (Molecular Dimension). Each drop contained 150 µL of the buffer solution plus 150 µL of protein at 15 mg $mL^{-1}$ in 10 mM Tris pH 8.0, 5 mM TCEP. Two conditions were

explored per buffer solution in a single plate. The first sitting drop contained protein only and the second drop containing protein with 0.7 mM SAH and 2 mM Scoulerine in a co-crystallisation experiment. The plates were stored at 20 °C in a Rigaku Minstrel HT UV (Rigaku) and imaged periodically.

Initial crystal hits from the commercial screens were obtained in the conditions listed in Table 4.6 and visualised in Figure 4.5. The crystals were fished utilising protein crystal mounting loops (Hampton CrystalCap™ Spine HT) and flash frozen in liquid nitrogen. Initial X-ray diffraction studies were carried out utilising the in-house X-ray source collecting two images 90 ° apart with a 0.5 ° oscillation over a 5-minute exposure to 2.5 Å. The crystals which diffracted well were sent to Diamond light source for X-ray diffraction data collection, these were crystals E, F, G and I, Table 4.6. Interestingly all the crystals which diffracted well in-house were co-crystallised with scoulerine and SAH, whilst all the "apo" crystals diffracted poorly or not at all. Crystals E, F, G and I diffracted to a resolution of 1.5 Å, 2.0 Å, 3.1 Å and 2.2 Å, respectively.

Crystal E (PSMT1 SER C2 co-crystallised with SAH and scoulerine) X-ray diffraction data were eventually solved by molecular replacement utilising residues 221(Val)-390(Lys) of the apoPSMT1 WT model representing the SAM binding domain. This data set was solved first because it contained the highest resolution data of all the crystals tested. Initially molecular replacement was carried out utilising MOLREP with the full apoPSMT1 WT model along with two truncated models; one corresponding to the dimerisation domain, residues 34-292, and one corresponding to the SAM binding domain, residues 221-390. This approach was taken due to the large conformational change reported in the literature upon substrate binding. After molecular replacement and 10 rounds of refinement by REFMAC[129] the SAM domain search model (residues 221-390) gave the best results with Rwork of 0.47 and Rfree of 0.49, while the dimerisation domain (residues 34-292) gave an Rwork of 0.52 and Rfree of 0.52, and the full model (residues 34-390) gave an Rwork of 0.47 and Rfree of 0.51. Following molecular replacement, the auto-building program BUCCANEER[127] was utilised to carry out 10 rounds of iterative model building (2 cycles) and refinement (20 cycles) with 3 initial build cycles to build the model, utilising the SAM binding

domain model generated during molecular replacement as a base for chain extension. BUCCANEER[127] build 695 residues of which 678 were assigned to the sequence over 6 fragments. It predicted there to be two chains, 3 fragments per chain, of which 87 % of the residues were built of which 98 % assigned to a chain. Chain A consisted of fragment 1 residues 34-112, fragment 2 residues 129-234 and fragment 3 residues 237-392. Chain B consisted of fragment 1 residues 32-119, fragment 2 residues 129-234 and fragment 3 residues 237-392. The resulting model decreased the Rwork to 0.28 and the Rfree to 0.30, Figure 4.6.

Form the BUCCANEER[127] auto-built model manual model building was implemented utilising COOT[130,131]. Fragments two and three for chain A and chain B were joined together to generate a larger C-terminal fragment consisting of residues 129-392 for both chains. Chain B was translated to symmetry position 1,0,0 (X,Y,Z) in order to correctly build the dimer. The first fragment of each chain was assigned incorrectly due to the large unassigned region of PSMT1, this was promptly corrected by renaming the fragments. After iterative rounds of manual model building and refinement including the incorporation of water molecules the unmodeled co-crystallised ligands are clearly visible within the weighted map from refinement (2Fo-Fc) and the weighted difference map (Fo-Fc), Figure 4.7. SAH and scoulerine were modelled into the 2Fo-Fc electron density map, further refinement produced the model in Figure 4.8 with statistics listed in Table 4.7.

X-ray diffraction data from crystals F (PSMT1 SER C3) and I (PSMT1 SER C2&3) were solved by molecular replacement utilising the model of PSMT1 SER C2 1.5 Å (crystal E). PSMT1 SER C2&3 crystal G has not been solved.

| PSMT1 SER C2 | | | | | |
|---|---|---|---|---|---|
| Condition | Screen | Well | Buffer | Crystal | Resolution |
| Apo | PACT premier | F11 | 0.2 M sodium citrate, 0.1 M bis-tris propane pH 6.5, 20 % w/v PEG3350 | A | No diffraction |
| | | G1 | 0.2 M sodium fluoride 0.1 M bis-tris propane pH 7.5 20 % w/v PEG3350 | B | No diffraction |
| | | G12 | 0.2 M sodium fluoride 0.1 M bis-tris propane pH 7.5 20 % w/v PEG3350 | C | No diffraction |
| | JCSG+ | G8 | 0.15 M DL-malic acid pH 7.0, 20 % PEG 3350 | D | poor diffraction |
| SAH and Scoulerine | JCSG+ | A3 | 0.2 M di-ammonium citrate pH 5.0, 20 % PEG 3350 | E | 1.5 Å |
| PSMT1 SER C3 | | | | | |
| Condition | Screen | Well | Buffer | | Resolution |
| SAH and Scoulerine | PACT premier | B1 | 0.1 M MIB buffer pH 4.0, 25 % w/v PEG 1500 | F | 2.0 Å |
| PSMT1 SER C2&3 | | | | | |
| Condition | Screen | Well | Buffer | | Resolution |
| SAH and Scoulerine | PACT premier | B3 | 0.1 M MIB buffer pH 6.0, 25 % w/v PEG 1500 | G | 3.1 Å |
| | | D3 | 0.1 M MMT buffer pH 6.0, 25 % w/v PEG 1500 | H | Poor diffraction |
| | JCSG+ | A3 | 0.2 M di-ammonium citrate pH 5.0, 20 % PEG 3350 | I | 2.2 Å |
| | | C3 | 0.2 M ammonium nitrate pH 6.3, 20 % PEG 3350 | J | No diffraction |

*Table 4.6: Table containing crystal hits from the initial commercial screen of PSMT1 SER mutants. Images of crystals are shown in Figure 4.5*

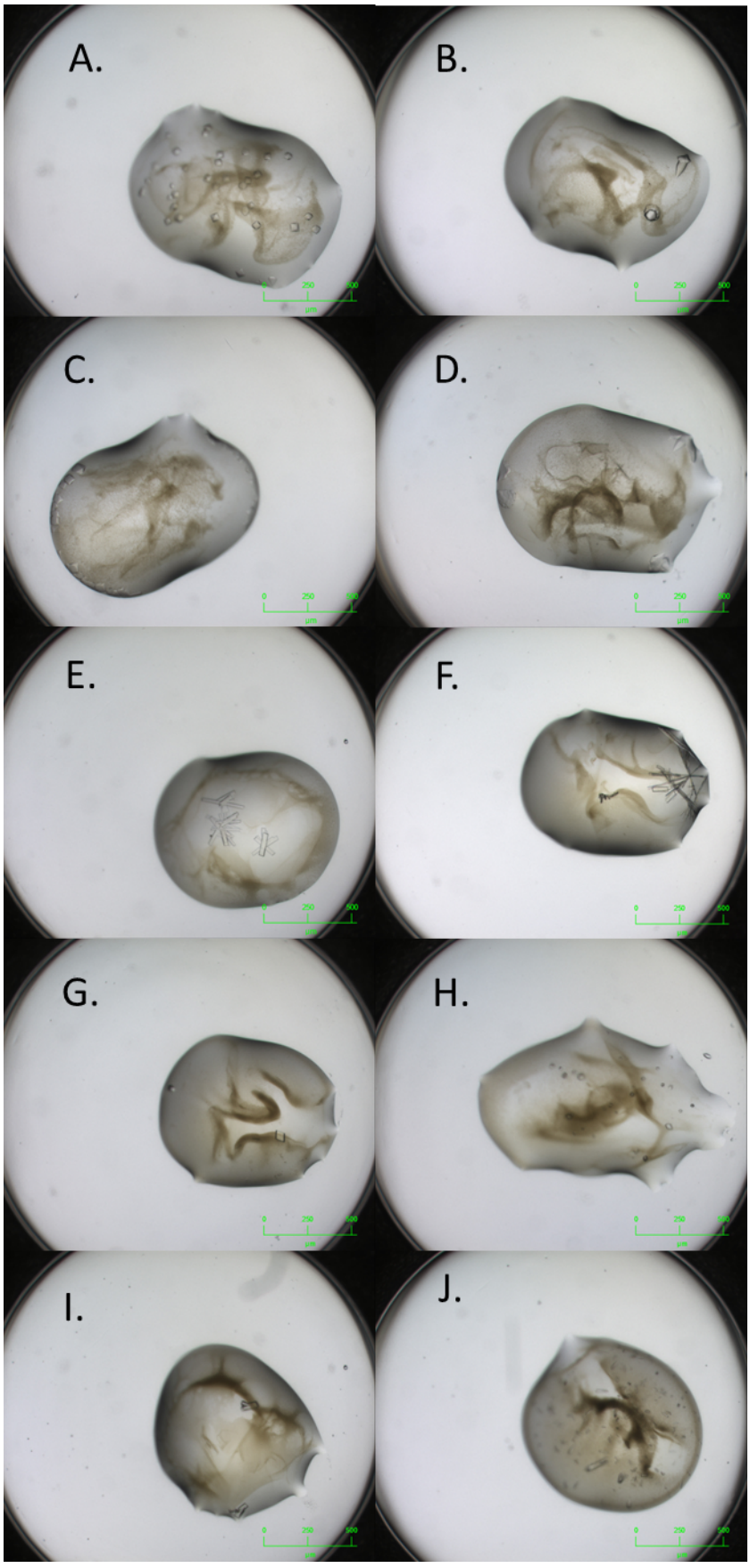



*Figure 4.5: Crystals obtained for PSMT1 SER mutants. Crystallisation conditions are described in Table 4.6*

| **PSMT1 SER** | **Cluster 2** | **Cluster 3** | **Cluster 2&3** | **Cluster 2&3** |
|---|---|---|---|---|
| **Co-crystallised with** | **Scoulerine and SAH** | **Scoulerine and SAH** | **Scoulerine and SAH** | **Scoulerine and SAH** |
| **Crystal** | **E** | **F** | **G** | **I** |
| **Data collection** | | | | |
| Space group | P $2_1$ | P $2_1$ | P $2_1$ $2_1$ $2_1$ | P $2_1$ |
| Cell dimensions | | | | |
| a, b, c (Å) | 68.04, 75.87, 76.19 | 65.71, 78.83, 68.11 | 83.01, 105.55, 168.18 | 68.15, 75.81, 76.35 |
| (°) | 90.00, 101.65, 90.00 | 90.00, 95.30, 90.00 | 90.00, 90.00, 90.00 | 90.00, 101.59, 90.00 |
| Resolution (Å) | 75.9-1.49 | 78.83-2.02 | 74.44-3.13 | 50.1-2.24 |
| Rmerge | 0.06(0.82) | 0.11(0.88) | 0.25(1.32) | 0.17(1.04) |
| Rpim | 0.04(0.53) | 0.07(0.57) | 0.13(0.55) | 0.12(0.80) |
| I / σI | 11.1(1.4) | 6.4(1.3) | 8.0(1.2) | 8.4(1.3) |
| Completeness (%) | 99.5(99.0) | 100.0(99.9) | 90.4(100.0) | 96.9(76.0) |
| Redundancy | 4.1(4.1) | 4.1(4.0) | 3.9(3.9) | 4.0(3.7) |
| Xia2 run | 3dii | dials | 3d | 3d |
| | | | | |
| **Refinement** | | | | |
| Resolution (Å) | 75.9-1.49 | 67.91-2.02 | | 50.10-2.24 |
| No. reflections | 123056 | 45410 | | 35608 |
| Rwork / Rfree | 0.11/0.18 | 0.21/0.26 | | 0.19/0.25 |
| No. atoms | | | | |
| Protein | 10719 | 10384 | | 10702 |
| Ligand/ion | 174 | 174 | | 174 |
| Water | 500 | 80 | | 466 |
| B-factors (Å$^2$) | | | | |
| Protein | 21.9 | 28.1 | | 31.3 |
| Ligand/ion | 14.9 | 20.1 | | 26.3 |
| Water | 31.6 | 20.1 | | 34.0 |
| r.m.s. deviations | | | | |
| Bond lengths (Å) | 0.03 | 0.01 | | 0.01 |
| Bond angles (°) | 2.4 | 1.8 | | 1.7 |

*Table 4.7: Data collection and refinement statistics for PSMT1 SER mutants*

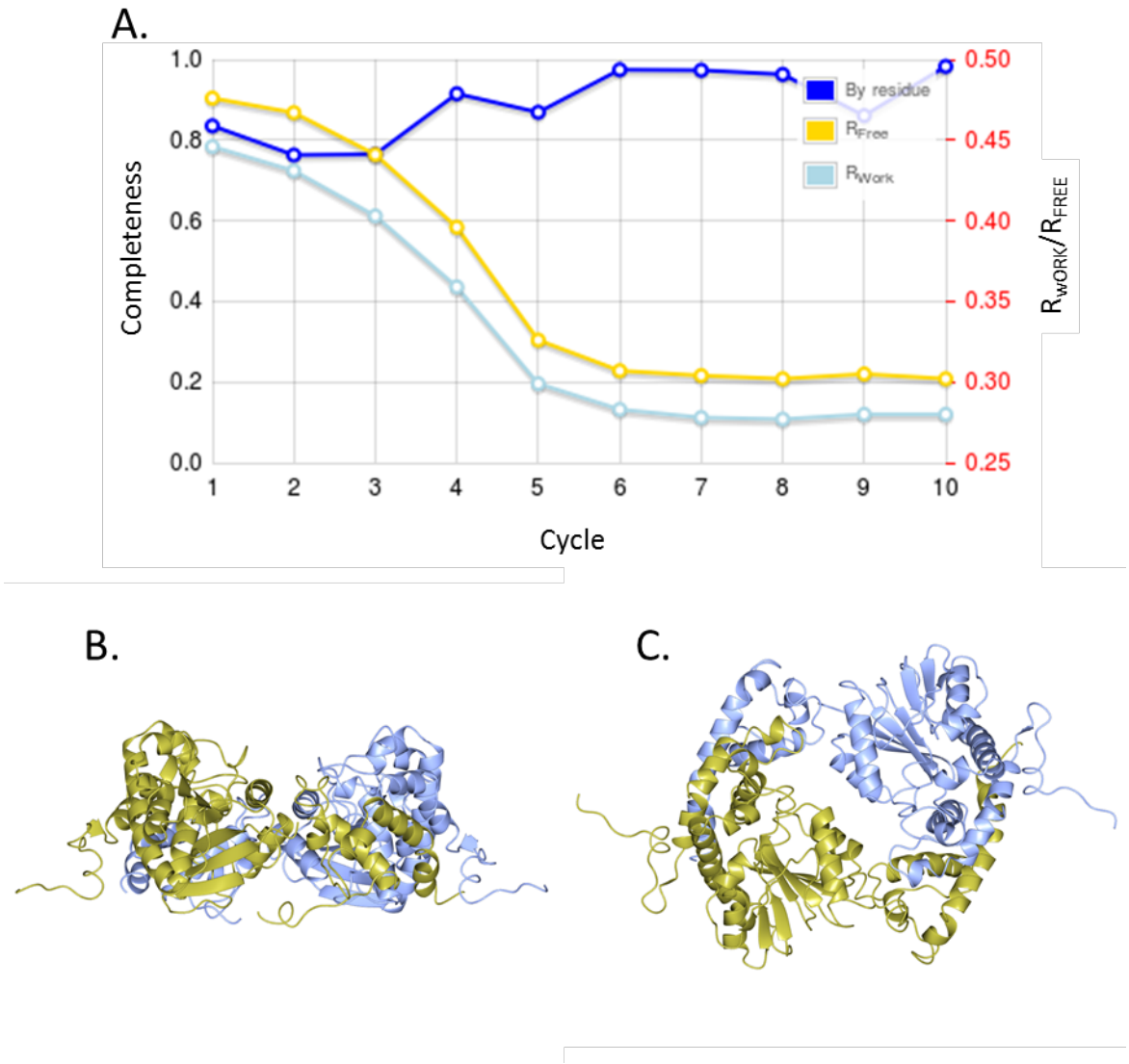



*Figure 4.6: BUCCANEER results output for PSMT1 SER C2 co-crystallised with SAH and scoulerine (Crystal E) showing, A. completeness by residue in blue, $R_{FREE}$ in yellow and $R_{WORK}$ in light blue over the 10 cycles. B. and C. are ribbon diagrams of the generated model, with C rotated 90 ° forward in relation to B*

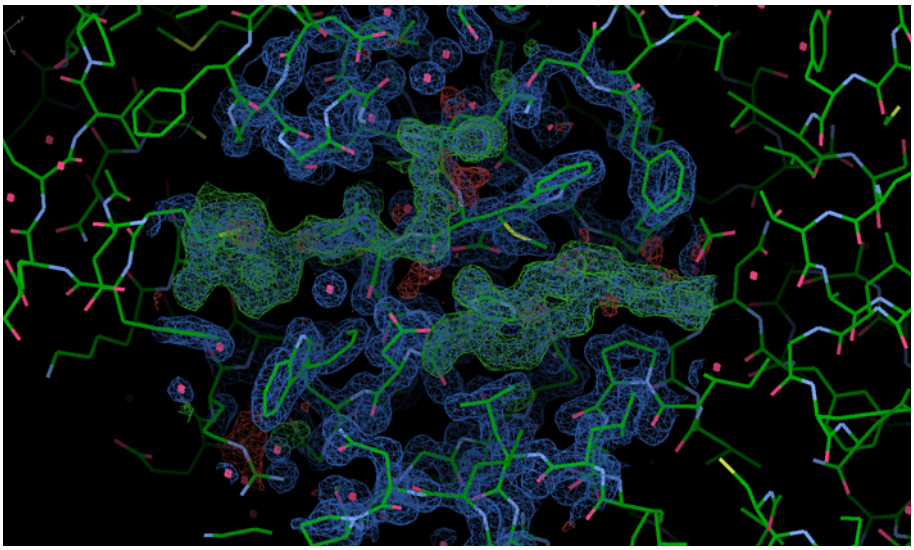


*Figure 4.7: Electron density for the unmodeled ligands bound to PSMT1 SER C2 (crystal E). Blue mesh 2Fo-Fc, green and red Fo-Fc both at 2 sigma, screenshot from COOT*

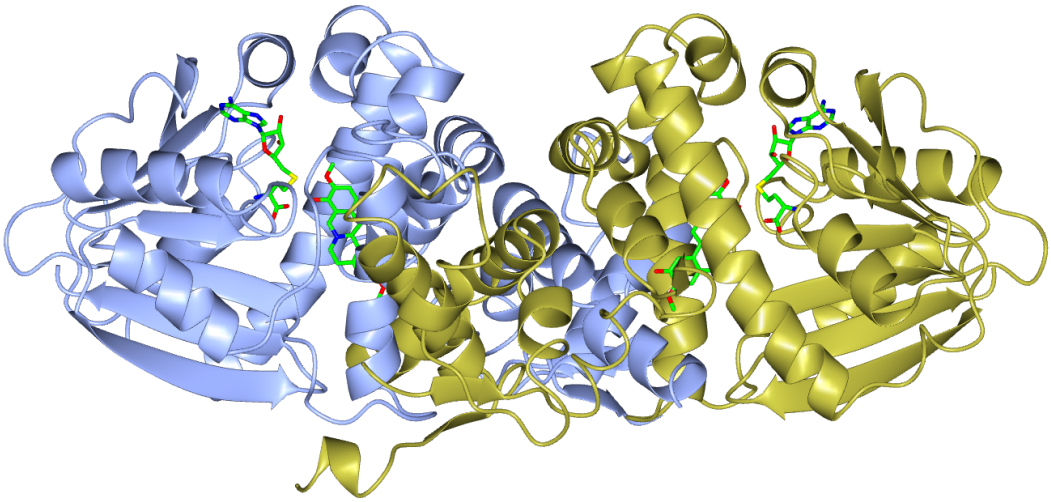


*Figure 4.8: Ribbon diagram of PSMTS SER C2 with SAH and scoulerine bound in a closed conformation, Protein shown as a ribbon diagram coloured by chain, with ligands shown as green cylinders*

As PSMT1 SER C2 generated the highest resolution model this mutant was taken forward for further characterisation to determine if the Lys114Ala and Lys115Ala mutations have affected the catalytic properties.

#### *4.3.3.5. Activity Assay of PSMT1 SER C2*

Kinetic analysis of PSMT1 SER C2 purified by on-column cleavage was performed, Figure 4.9, and the $V_{max}$ and $K_m$ were determined by Michaelis-Menten direct fit, Table 4.8. A $V_{max}$ of 446.7 nmol $min^{-1}$ $mg^{-1}$ of protein and a $K_m$ of 0.22 µM was calculated with scoulerine as the varying substrate concentration with a constant concentration of 100 µM SAM. The calculated $V_{max}$ and $K_m$ values for PSMT1 SER C2 correspond nicely to that of the activity and affinity obtained for PSMT1 WT enzyme of $492.6 \pm 28.68$ nmol $min^{-1}$ $mg^{-1}$ of protein and $0.25 \pm 0.064$ µM, respectively, Section 3.3.3 pg. 77. Therefore, the surface entropy reduction mutations incorporated into PSMT1 WT have not effected activity or affinity to the substrate, scoulerine.

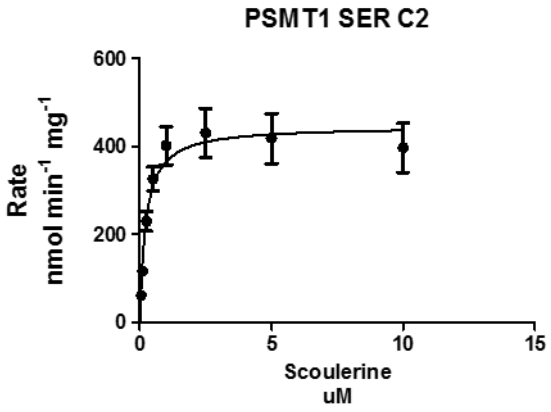



*Figure 4.9: Michaelis-Menten kinetics of PSMT1 SER C2 with scoulerine as substrate*

| **Substrate** | | **Value** | **std error** |
|---|---|---|---|
| **Scoulerine** | $V_{max}$ (nmol $min^{-1}$ $mg^{-1}$ of protein) | 446.7 | 23.99 |
| | $K_m$ (µM) | 0.22 | 0.054 |

*Table 4.8: Michaelis-Menten enzyme kinetics of PSMT1 SER C2 with scoulerine as the substrate*

#### *4.3.3.6. NATIVE-PAGE*

Purified PSMT1 SER C2 by nickel affinity on-column cleavage followed by desalting and anion exchange in 10 mM Tris pH 8.0, 5 mM TCEP was analysed by native-PAGE alongside PSMT1 WT and BSA, Figure 4.10. Native-PAGE separates protein based on their charge and hydrodynamic radius, with smaller highly charged proteins migrating further than larger less charged proteins. The results show one large band of PSMT1

SER C2 along with a ladder effect reducing in intensity with reduced migration. PSMT1 WT produced the same laddering effect but with reduced migration. Whereas, the BSA major band has migrated further. This is explained by the molecular weight and isoelectric point of the different proteins. The difference in migration of PSMT1 WT and PSMT1 SER C2 cannot be explained by the small difference in molecular weight of 114 Da but can be by their isoelectric points of 5.46 and 5.28, respectively. The lower pI of PSMT1 SER C2 would result in a greater migration towards the anode of the native-PAGE apparatus, which is located at the bottom of the gel tank. BSA has a molecular weight of 66 kDa and a pI of 4.70 and was ran alongside of the PSMT1 samples as a control. BSA possesses a lower pI than that of both PSMT1 samples and having a lower molecular weight has a two-fold effect on migration, resulting in the furthest migration of all the samples analysed.

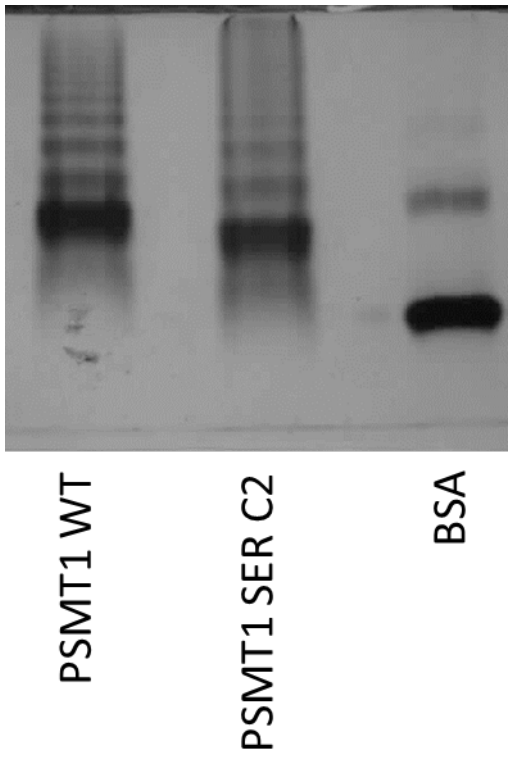



*Figure 4.10: Native-PAGE analysis of PSMT1 SER C2 alongside PSMT1 WT and BSA as a control. 30 µg of protein loaded per lane. Calculated pI for PSMT1 WT 5.46, PSMT1 SER C2 5.28, BSA 4.70*

### 4.3.4. PSMT1 SER Cluster 2 Crystal Structures

All SER models from the initial screen, Table 4.7 are in a closed conformation with SAH and scoulerine bound in their active sites. 15 mg $mL^{-1}$ PSMT1 SER C2 co-

crystallised with 0.7 mM SAH and 2 mM scoulerine in a 1:1 ratio with 0.2 M di-ammonium citrate pH 5.0, 20 % PEG 3350 crystallisation buffer in a 300 nL total volume sitting drop produced the highest diffracting crystal to 1.5 Å. As the transferred methyl group is not present in this structure co-crystallisation experiments were carried out with SAM and scoulerine (the substrates) in an effort to obtain a substrate or product (SAH and tetrahydrocolumbamine) complex. This would provide a better representation of the catalytic active site. Co-crystallisation with SAH and tetrahydrocolumbamine was not attempted as the latter was not available.

Co-crystallisation of PSMT1 SER C2 with SAM and scoulerine combined with optimisation in 48 well plates around the initial crystallisation buffer condition resulted in a crystal which diffracted to 1.3 Å at Diamond Light Source, Figure 4.11. This was produced with 15 mg $mL^{-1}$ PSMT1 SER C2, 1 mM SAM, 2 mM scoulerine in a 1:1 ratio with 0.2 M di-ammonium citrate pH 4.8, 20 % PEG 3350 with a total volume of 1 µL.

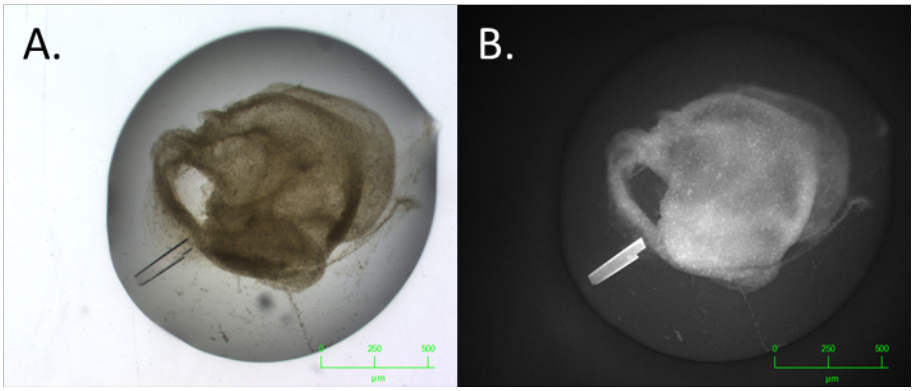



*Figure 4.11: Crystals of PSMT1 SER C2 co-crystallised with SAM and scoulerine. A. colour image, B. UV image*

PSMT1 SER C2 co-crystallised with 1 mM SAM and 2 mM scoulerine crystallised in the same space group (P $2_1$) as the previous PSMT1 SER C2 structure. X-ray diffraction data was collected at Diamond Light Source and solved by molecular replacement with the previously described 1.5 Å PSMT1 SER C2 co-crystallised with SAH and scoulerine, Table 4.9. The resulting model is a homodimer with the products, SAH and tetrahydrocolumbamine bound within the active sites, Figure 4.12.

PSMT1 has either bound the substrates, SAM and scoulerine, and turned over in crystal resulting in the products bound in the active site, or the substrates have been converted to the products and subsequently rebound during crystallisation. In reality both methods are likely to have contributed due to the time taken for crystal growth. This product bound structure is essentially identical to that of PSMT1 co-crystallised with SAH and scoulerine with an RMSD of 0.1 Å over 686 residues. Chain A consists of two fragments consisting of residues 34-112 and 128-390 while chain B also consists of two fragments of residues 34-114 and 128-390. This is most likely due to the inherent flexibility of the missing loop.

| **PSMT1 SER** | **Cluster 2** |
|---|---|
| **Co-crystallised with** | **Scoulerine and SAM** |
| | |
| **Data collection** | |
| Space group | P $2_1$ |
| Cell dimensions | |
| a, b, c (Å) | 68.21, 75.93, 76.56 |
| (°) | 90.00, 101.61, 90.00 |
| Resolution (Å) | 75.93-1.29 |
| Rmerge | 0.05(0.81) |
| Rpim | 0.03(0.52) |
| I / σI | 11.80(1.29) |
| Completeness (%) | 95.3(92.4) |
| Redundancy | 4.2(4.3) |
| Xia2 run | 3dii |
| | |
| **Refinement** | |
| Resolution (Å) | 75.11-1.29 |
| No. reflections | 182560 |
| Rwork / Rfree | 0.13/0.17 |
| No. atoms | |
| Protein | 10694 |
| Ligand/ion | 182 |
| Water | 448 |
| B-factors (Å$^2$) | |
| Protein | 19.6 |
| Ligand/ion | 14.6 |
| Water | 28.2 |
| r.m.s. deviations | |
| Bond lengths (Å) | 0.03 |
| Bond angles (°) | 2.2 |

*Table 4.9: Data collection and refinement statistics for PSMT1 SER C2 co-crystallised with scoulerine and SAM*

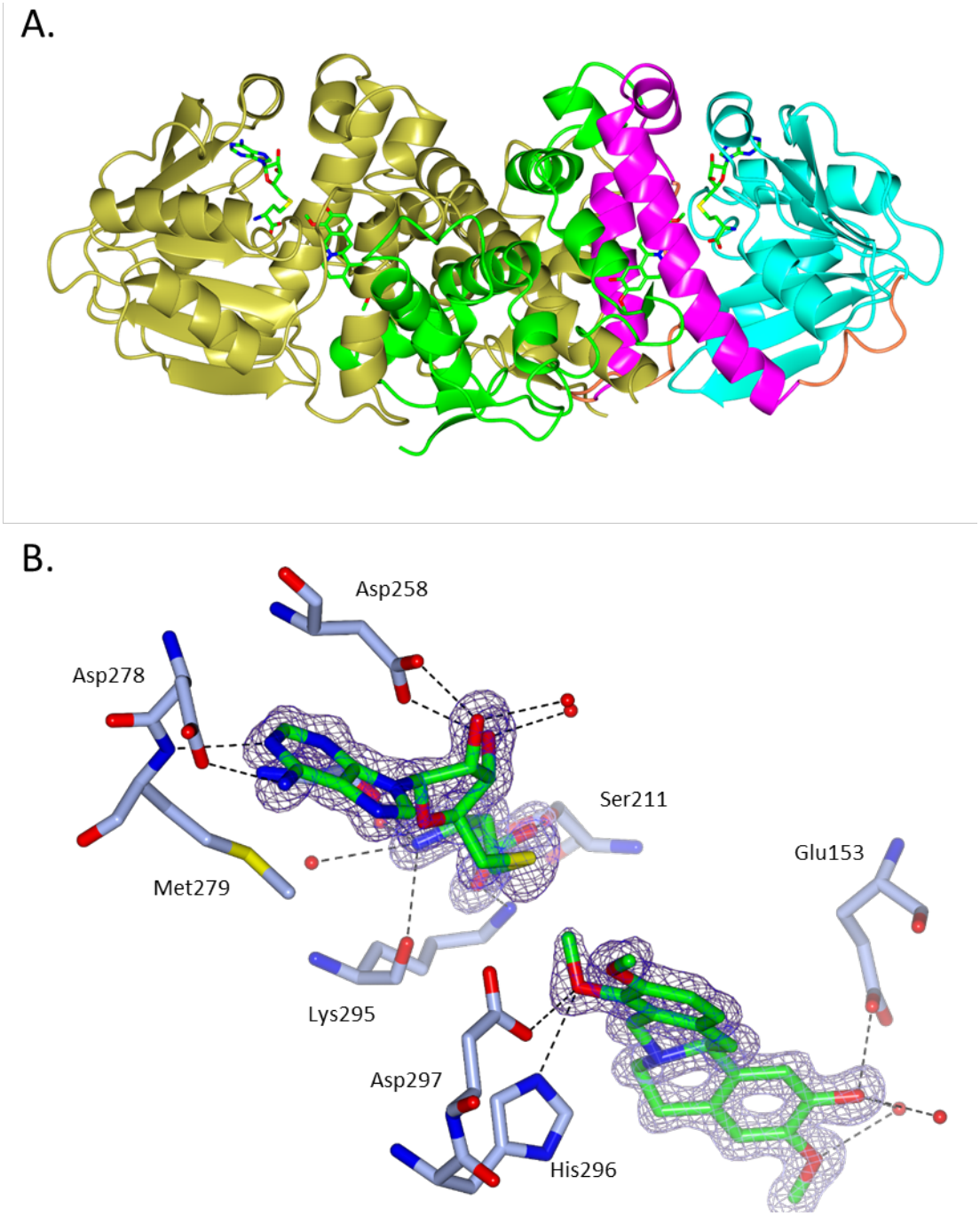



*Figure 4.12: PSMT1 SER C2 X-ray structure co-crystallised with SAM and scoulerine but with SAH and tetrahydrocolumbamine bound. A. ribbon diagram coloured with chain A coloured gold and chain B coloured by domain. Chain B dimerisation domain green, substrate binding domain magenta, SAM/SAH binding domain cyan. Ligands shown as green cylinders B. PSMT1 SER C2 active site highlighting hydrogen bonding residues. Electron density (2Fo-Fc) shown as blue mesh clipped to SAH and tetrahydrocolumbamine contoured to 1.5 σ (0.56e Å$^{-3}$)*

### 4.3.5. *S*-Adenosylhomocysteine Binding

SAH binding in the PSMT1 SER C2 structure with SAH and scoulerine bound compared to with SAH and tetrahydrocolumbamine bound are nearly identical. All the interactions are maintained as described for the PSMT1 WT structure co-crystallised with SAM in Section 3.3.5.2, pg. 87, along with several new ones. Also, due to the marked increase in resolution water molecules can easily be visualised in the model giving greater detail into secondary coordination. PSMT1 with SAH and tetrahydrocolumbamine bound is shown in Figure 4.13. The conserved features of SAM/SAH binding in plant O-methyltransferases as discussed in Section 1.7.3 pg. 41 and outlined in Figure 1.14 are also present in PSMT1.

The adenine moiety of SAH makes several hydrogen bonds with PSMT1. The carboxylic acid group of Asp278 hydrogen bonds to N6 of adenine, and Met279 backbone amine hydrogen bonds with N1 of adenine. N6 and N7 of adenine hydrogen bond to a water molecule which through an extended solvent network interacts with Glu191 Oε, Arg203 terminal amine, Gly277 backbone oxygen, Asp278 Oδ and Glu281 Oε. Several aromatic interactions are also involved in adenine recognition including Met279 Sδ-π and Leu259 CγH-π bonding to opposite faces of the adenine aromatic ring, and adenine C8 makes a CH-π interaction with Phe190. Residues Phe280, Trp298, Phe190, Asp285 and Leu259 also make up the binding pocket.

The ribose ring of SAH is recognised mainly by Asp258 which forms hydrogen bonds with the ribose hydroxyl groups. The hydroxyl groups also interact with water molecules. Ribose 3-OH hydrogen bonds to Asn204 Oδ *via* a water molecule and to the Asn204 backbone carbonyl group and to the carboxylic acid on the methionine moiety of SAH by an extended solvent network. A similar interaction is made by ribose 2-OH to His261 Nε *via* a water molecule and through a solvent network interacts with Asn200 Oδ and its backbone oxygen, and Asn204 Oδ. Trp298 Nε and Asp297 Oδ hydrogen bond with the 4-O of ribose *via* a water molecule. Residues Phe190, Gly235, Gly237 and Leu259 also interact with the ribose ring of SAH.

The homocysteine moiety of SAH is recognised by Lys292, making hydrogen bonds with the homocysteine amine group *via* its backbone carbonyl and to the

homocysteine carboxylic acid group *via* its terminal amino group. The homocysteine moieties carboxylic acid group hydrogen bonds to Ser211 Cγ-OH and to Ile238 backbone amine group. Asn204 backbone carbonyl, Arg203 Cδ, Trp240 backbone amine, Gly236 backbone carbonyl, Ser240 Cβ-OH and its backbone nitrogen, Thr240 backbone amine, Ser211 Cβ-OH, the 3-OH of the ribose ring and the methionine amino group of SAH also interact with the methionine carboxylic acid group *via* an extended solvent network. The Gly235 backbone carbonyl hydrogen bonds to the methionine amino group, along with the previous mentioned solvent network. Gly235 and Gly236 form a hydrophobic wall between the carboxylpropyl moiety of methionine and the ribose ring. Met207, Trp293, Asp297, Lys293, Trp293, Phe203 and Phe190 are responsible for the correct positioning of the methyl sulphonium functional group by forming a channel in which the methyl group extends into.

PSMT1 SER C2 with SAH and scoulerine bound is in a slightly more closed conformation than that with SAH and tetrahydrocolumbamine. This is due to the absence of the methyl group eliminating steric clashes and allowing hydrogen bonding between the SAH thioether group and the 9-hydroxyl group of scoulerine.

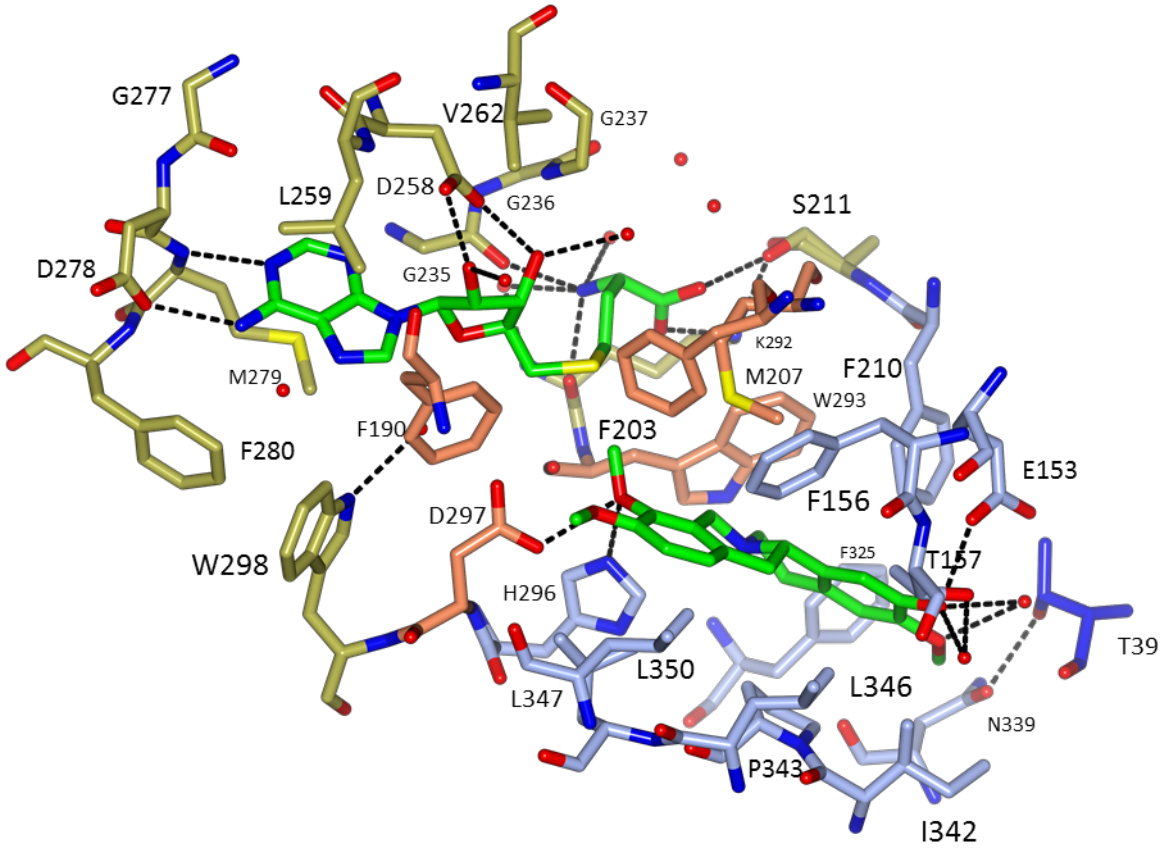



*Figure 4.13: PSMT1 SER C2 co-crystallised with SAM and scoulerine binding site. Residues and solvent within 4 Å of tetrahydrocolumbamine and SAH represented as cylinders. Tetrahydrocolumbamine and SAH are coloured green, residues involved in SAH binding shown in gold, residues from chain A involved in tetrahydrocolumbamine binding shown in ice blue with T39 from chain B coloured blue. Those residues involved in binding tetrahydrocolumbamine and SAH are coloured coral. Dashed lines indicate hydrogen bonds, water molecules are represented by red dots*

### 4.3.6. Substrate Binding

Scoulerine and tetrahydrocolumbamine bind in the active site of PSMT1 in a similar manner interacting with the same residues. Scoulerine/tetrahydrocolumbamine sits in a deep hydrophobic pocket within the enzyme, with hydrogen bonds at either end. The hydrophobic pocket is made up of residues Glu153, Phe,156, Thr157, Phe190, Phe203, Met207, Phe210, Trp293, Phe325, Asn339, Ile342, Pro343, Leu346, Leu347, Leu350 mainly from alpha helices six, twelve, sixteen and Thr39 from alpha helix one from the opposing chain. Due to the aromatic nature of the substrate and the binding pocket there is a large network of aromatic interactions.

The A-ring of scoulerine is sandwiched between a Phe210 by a CεH-π hydrogen bond and on the opposing face interacts with Ile342, Leu346 and Pro343. Thr157 Cγ makes a hydrophobic interaction with C1 and a van der Waal interaction with 2-OH of the A ring and it also makes a hydrogen bond with C3-O of scoulerine *via* a water molecule.

The carboxylic acid group Glu153 hydrogen bonds directly to the 2-hydroxyl of scoulerine and to C3-O *via* a water molecule. The O3 of scoulerine also hydrogen bonds to the backbone carbonyl of Thr39 of the opposing side chain *via* the same water molecule as Glu153. The 3-O-methyl group makes hydrophobic interactions with Thr39 from the opposing side chain and Asn339 Cα. It also makes van der Walls interactions with Asn339s backbone carbonyl group along with its Cγ-OH group.

The B-ring of scoulerine makes hydrophobic interactions with Trp293, Pro343, Phe325 and His296. His296 also forms van der Wall interactions with scoulerine N7 *via* its CεH. The C-ring makes hydrophobic interactions with Trp293 Cβ, His296 Cε and Leu346 Cδ. There is also a van der Walls interaction between His296 Nε and C8a of scoulerine.

Most plant *O*-methyltransferases methylate a hydroxyl group on a phenolic ring, the corresponding ring of scoulerine is the D-ring, and binds in a similar manner to the consensus proposed in Figure 1.16. The D-ring is held in position by Leu347 CδH-π bonding, with Met207 interacting on the opposing face. Along with, hydrophobic interactions with Leu350, Phe156 and Phe203. The 10-O-methyl position of scoulerine interacts hydrophobically with Phe190 and Phe203 and the carboxylic acid group of Asp297 *via* a hydrogen bond to its ether group. The 9-hydroxyl position of scoulerine is the hydroxyl group which is methylated by PSMT1. As expected this is highly co-ordinated making hydrogen bonds with the catalytic dyad of His296 and Asp279, as well as the backbone carbonyl of Trp293.

### 4.3.7. Catalytic Mechanism

The catalytic mechanism PSMT1 employs can be deduced from structures presented here. A product complex has been crystallised by co-crystallisation with the substrates, SAM and scoulerine resulting in turnover either in crystal or in solution

then rebinding the products. From this structure, a substrate bound state can be inferred, Figure 4.14.A. His296 acts as a catalytic base abstracting a proton from the 9-hydroxyl position of scoulerine generating a phenoxide intermediate. Subsequently the sulphonium methyl group of SAM under goes $S_N2$ nucleophilic attack by the lone pair of electrons on the reactive 9-phenoxide intermediate, Figure 4.14.B. Aspartic acid 297 may also be involved directly in catalysis or indirectly by ensuring the correct geometry for catalysis. His296 (Nδ) also hydrogen bonds to Glu356 terminal carboxylic acid group, this interaction is not only important for the correct positioning of His296 but also in promoting its alkalinity, making it a stronger base.

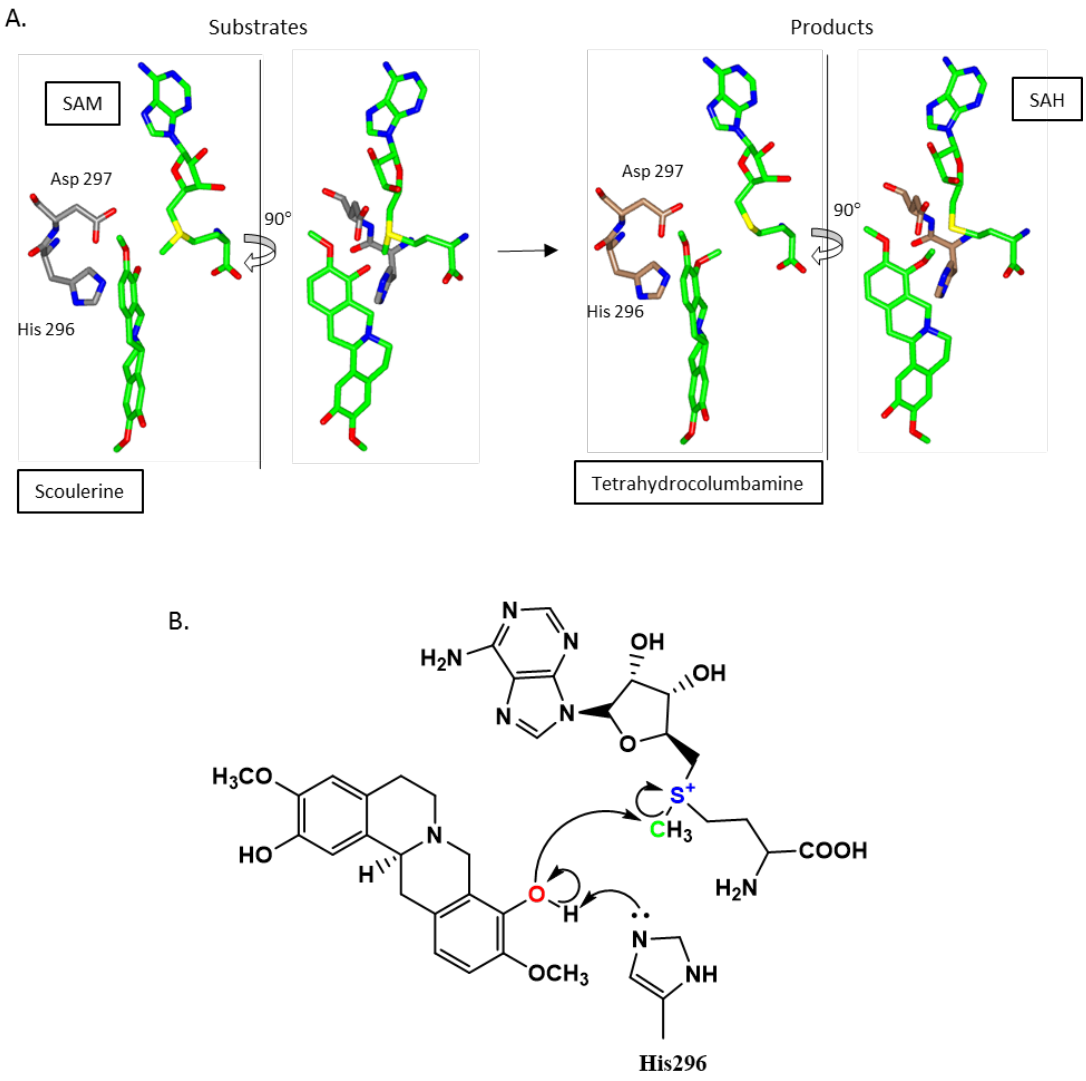



*Figure 4.14: Proposed catalytic reaction mechanism of scoulerine 9-hydroxyl methylation by PSMT1. A. Inferred orientation of the substrate complex based on PSMT1 SER C2 with SAH and tetrahydrocolumbamine bound and the product complex. B. Reaction schematic*

### 4.3.8. Surface Entropy Reduction Mutations

The surface entropy reduction mutations of Lys114Ala and Lys115Ala have allowed access to a new crystal form which diffracts to a very high resolution. To analyse how this new crystal form arose the crystal structure was studied. The previous apoPSMT1 WT structure diffracted to 2.9 Å, with a missing loop between residues 112-129 in all four chains in the asymmetric unit. Both PSMT1 SER C2 structures presented in this chapter also lack part of this loop but the loop extends to residue 118 in PSMT1 SER C2 co-crystallised with SAH and scoulerine and the SER mutations can be visualised, Figure 4.15. Ala114 forms part of the crystal contact interface with a symmetry dimer, if the Lys114Ala mutation was not present this crystal form would be impossible due to steric clashes. Lys115Ala mutation points out towards a bulk solvent channel within the crystal. It could be possible that only the Lys114Ala mutation is necessary to obtain this crystal form.

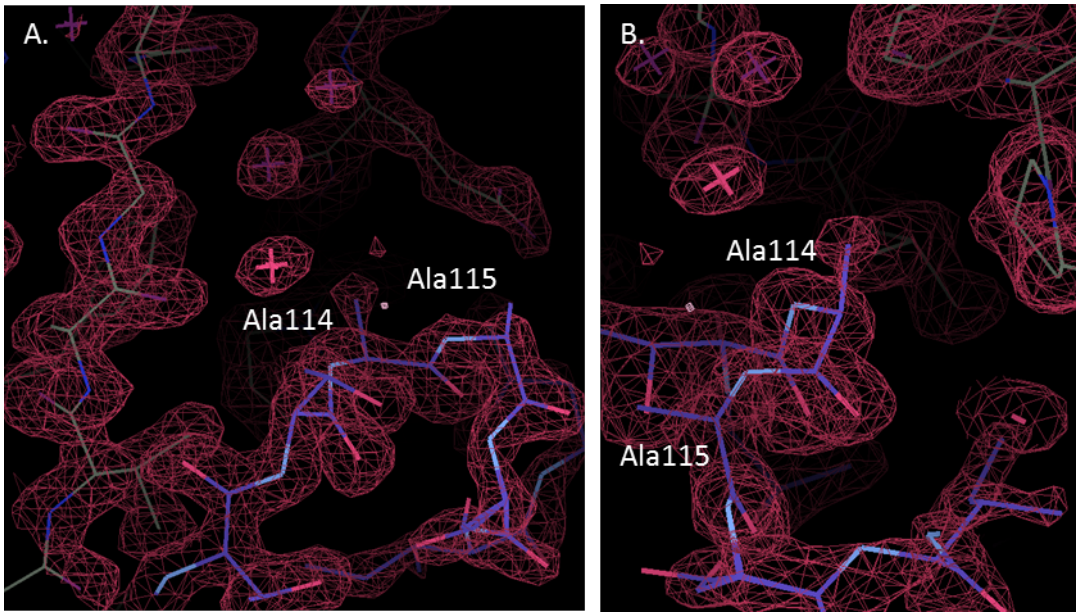



*Figure 4.15: Surface entropy reduction mutation Lys114Ala and Lys115Ala present in PSMT1 SER C2 co-crystallised with SAH and scoulerine. A and B are different orientations. Model shown in purple along with a symmetry model shown in light green, weighted electron density (2Fo-Fc) is shown as a red mesh contoured to 1.5 σ (0.55 e Å$^{-3}$), water molecules are represented as red crosses*

### 4.3.9. Open and Closed Differences

Comparison on the open and closed structures of PSMT1 will provide a deeper understanding of how the conformational change occurs. apoPSMT1 WT, PSMT1 WT

with SAM bound presented in Chapter 3 are in an open conformation and were superimposed with each other and with PSMT1 SER C2 with SAH and scoulerine bound, and PSMT1 SER C2 with SAH and tetrahydrocolumbamine bound, which are in a closed conformation, Figure 4.16. As the structure is closed with SAH and scoulerine or tetrahydrocolumbamine bound it suggests that upon binding of both substrates PSMT1 undergoes a conformational change where the SAM binding domain rotates inwards towards the dimerisation/substrate binding domain, resulting in about an 8 Å closure to the active site opening. The closure of PSMT1 leads to the formation of the functional active site binding pocket.

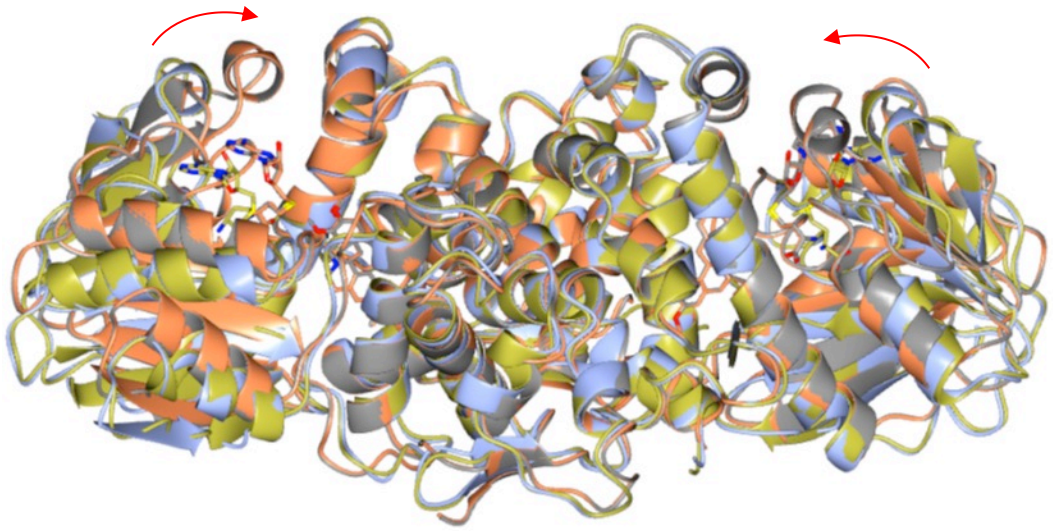


*Figure 4.16: Superposition of PSMT1 structures. apoPSMT1 WT-Open (ice blue), PSMT1 WT with SAM bound-Open (gold), PSMT1 SER C2 with SAH and scoulerine bound-Closed (coral) and PSMT1 SER C2 with SAH and tetrahydrocolumbamine bound-Closed (grey). Structures are superimposed by the dimerisation domain, peptide represented as ribbons and ligands represented as cylinders. Red arrows indicate the rotation of the SAM binding domain in relation to the dimerisation domain upon active site closure.*

PSMT1 WT with SAM bound (gold) will be compared to PSMT1 SER C2 with SAH and tetrahydrocolumbamine bound (grey), Figure 4.17-A&B. Figure 4.17.A shows the superposition of the two structural models by the dimerisation/substrate binding domains. Upon closure of the active site the labile methyl group of SAM, the 9-OH position on scoulerine and the catalytic dyad are brought into close proximity. This allows catalysis to occur where the SAM traps scoulerine in the active site which prevents the release of the phenoxide intermediate. Without the large conformational change catalysis could not occur. Upon closure, Phe210, Met207,

Phe203, Phe190 and Trp293 change conformation generating the upper face of the substrate binding pocket.

Figure 4.17.B shows the superposition of the two structural models by the SAM/SAH binding domains. SAM/SAH ligand-protein interactions are maintained with new contacts bring made by Ser211, Met207, Phe203, Phe190 and Asp297 (dark grey) on the previously solvent exposed face of SAM. The newly positioned Met207, Phe203, Phe190 and Asp297, along with Trp293 form a channel in which the transferred methyl group extends into, ensuring the correct position for catalysis.

Figure 4.17.C shows a summary of the differences between the open and closed conformations mapped onto PSMT1 SER C2 with SAH and tetrahydrocolumbamine bound. All the SAM/SAH binding site interactions are maintained during active site closure (gold). Tetrahydrocolumbamine is bound in a deep pocket at the bottom of the binding pocket. The majority of the residues responsible for substrate binding are already in the correct position for binding (ice blue), these are mostly on the bottom face of the molecule. The residues which take on new positions are shown in coral. The majority of these residue interactions are responsible for the correct positioning of the sulphonium methyl group of SAM and the 9-hydroxyl position of scoulerine for catalysis. The catalytic dyad of His296 and Asp297 is also brought into the correct position for proton abstraction at the 9-hydroxyl position of scoulerine. The large conformational change affords several advantages: In the open conformation, the catalytic dyad is distal to the sulphonium methyl group of SAM which will stop undesirable methylation of small molecules and the uncontrolled consumption of valuable SAM. It will also allow for high substrate specificity due to the buried substrate binding site and the retention of the highly reactive activated phenoxide intermediate.

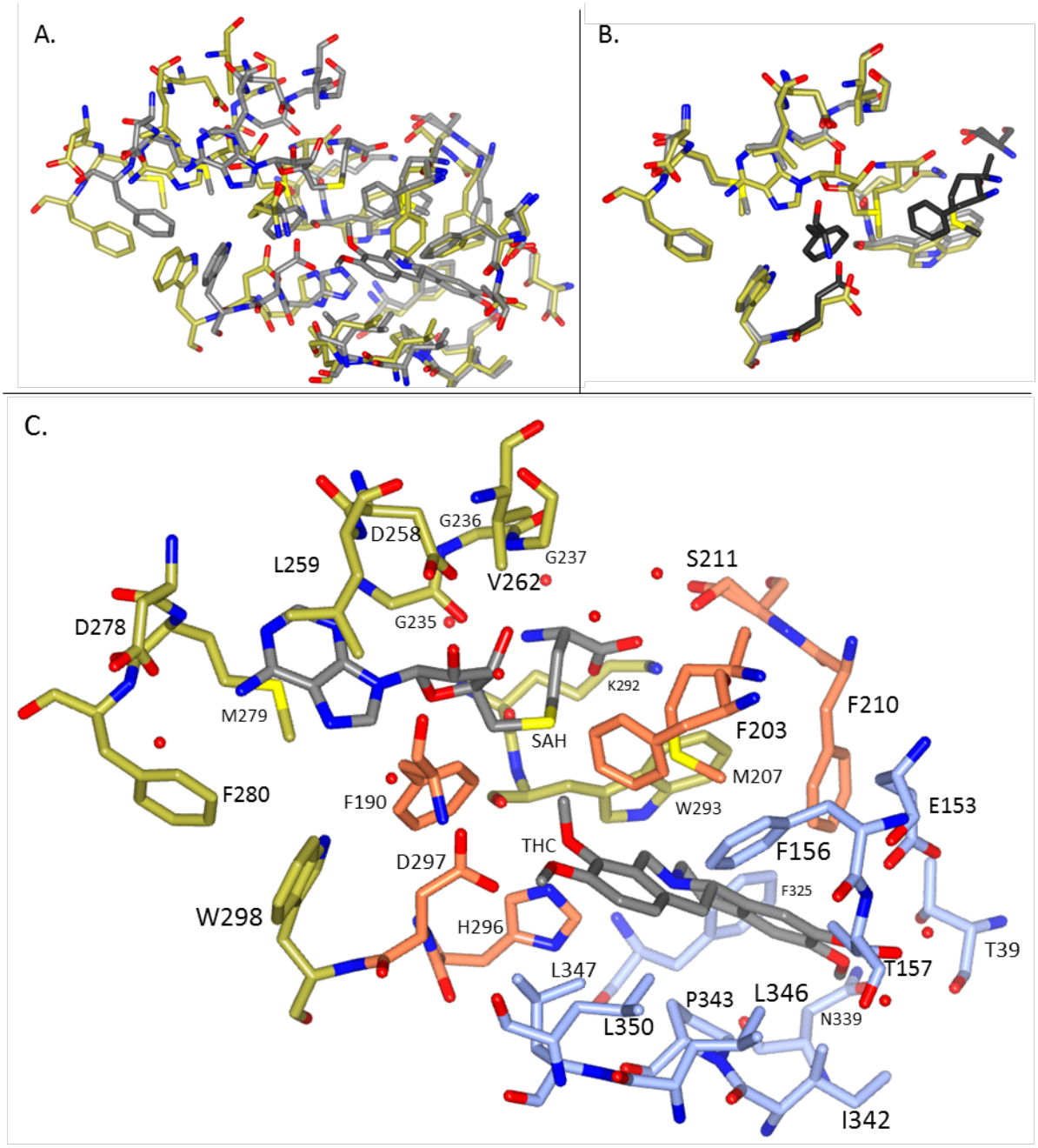



*Figure 4.17:Comparison of PSMT1 in open and closed conformations A. and B. show PSMT1 WT with SAH bound model (gold) and PSMT1 SER C2 with SAH and tetrahydrocolumbamine (THC) bound (grey) superimposed by the dimerisation/substrate binding domain and SAM binding domain, respectively. C. is a simplified representation of domain movement mapped onto PSMT1 SER C2 with SAH and tetrahydrocolumbamine bound. Gold residues represent residues which are conserved in SAM/SAH binding in both open and closed models. Ice blue residues are residues which have little conformational change upon substrate binding. Coral indicates the residues which have undergone significant conformational change*

## 4.4. Conclusion

Surface entropy reduction (SER) has allowed access to crystal forms of PSMT1 which would not be possible otherwise. Three SER constructs were generated based on sequence analysis by the surface entropy reduction prediction server, Table 4.3. All three constructs expressed and purified in the same manner as the WT enzyme, producing pure soluble homogeneous protein shown by SDS-PAGE analysis, Figure 4.2. Out of the three SER constructs PSMT1 SER cluster 2 produced the highest resolution diffracting crystals, extending to 1.5 Å, from the initial crystallisation screen trials, Table 4.6. Subsequent characterisation of PSMT1 SER C2 showed it to behave similar to that of WT protein in solution by SEC-MALLS (Figure 4.3), thermofluor (Figure 4.4) and as expected by native-PAGE (Figure 4.10) analysis. Activity assay data also show that the mutation introduced did not have an effect on catalytic activity, Figure 4.9 and Table 4.8.

Co-crystallisation of PSMT1 SER C2 with SAH and scoulerine produced a crystal with SAH and scoulerine bound in the active site. In comparison of this model with the apoWT and the WT plus SAM models it is evident that a large conformational change has taken place, Figure 4.16. The conformational change is caused by a rotation of the SAM binding domain inwards towards the dimerisation/substrate binding domain, resulting in a closure of about 8 Å of the active site opening. Co-crystallisation of PSMT1 SER C2 with the substrates SAM and scoulerine produced a crystal with the products, SAH and tetrahydrocolumbamine, bound in the active site. Not only does this provide validation of an active protein it also allowed the prediction of substrate binding and to infer the catalytic reaction mechanism, Figure 4.14.

The open and closed PSMT1 structures presented here have allowed a detailed description of how PSMT1 confers substrate specificity. It appears that both the methyl acceptor and methyl donor in either pre- or post-reaction states are need to be present for enzyme closure and active site formation. In the open form the majority of residues responsible for binding of the substrates are already in the correct conformation for binding. Upon closure and formation of the active site the

9-hydroxyl face of scoulerine and the sulphonium methyl face of SAM are brought together forming new protein interactions. This not only forms a channel into which the transferred methyl group protrudes, it also forms the catalytic dyad, with His296 and Asp297 manoeuvred into close proximity to the 9-hydroxyl position of scoulerine.

This work adds to the small number of published closed plant *O*-methyltransferases structures allowing a greater understanding of substrate and catalytic mechanism. Comparing PSMT1 in a closed conformation with the closed plant *O*-methyltransferase structures analysed in Section 1.7.4 pg. 42, Figure 1.14 and Figure 1.15 several conserved features were defined. The mode of SAM/SAH binding by the SAM binding domain is highly conserved, along with residues involved in binding of the phenolic ring to be methylated. Further analysis has shown new unreported conserved features, Figure 4.18. Glu356 in PSMT1 is a highly-conserved residue of the crystallised plant *O*-methyltransferases, Figure 4.18, therefore the feature which promotes the basicity of the catalytic histidine must also be highly conserved. The residues, Phe109, Met207, Phe203, Trp293 and the catalytic Asp297 which are responsible for the formation of the sulphonium methyl transfer channel are highly conserved. Although the hydrophobic pocket in which the substrate binds is highly variable, the substrates generally have a hydroxyl group distal to the hydroxyl group to be methylated. This distal hydroxyl group generally makes a hydrogen bond with an acidic residue or the terminal carbonyl of asparagine, although hydrogen bonding to glutamine is not observed in this set of enzymes it cannot be ruled out. This distal hydroxyl group hydrogen bond feature could aid in the prediction of possible substrates of plant *O*-methyltransferases with unknown substrates.

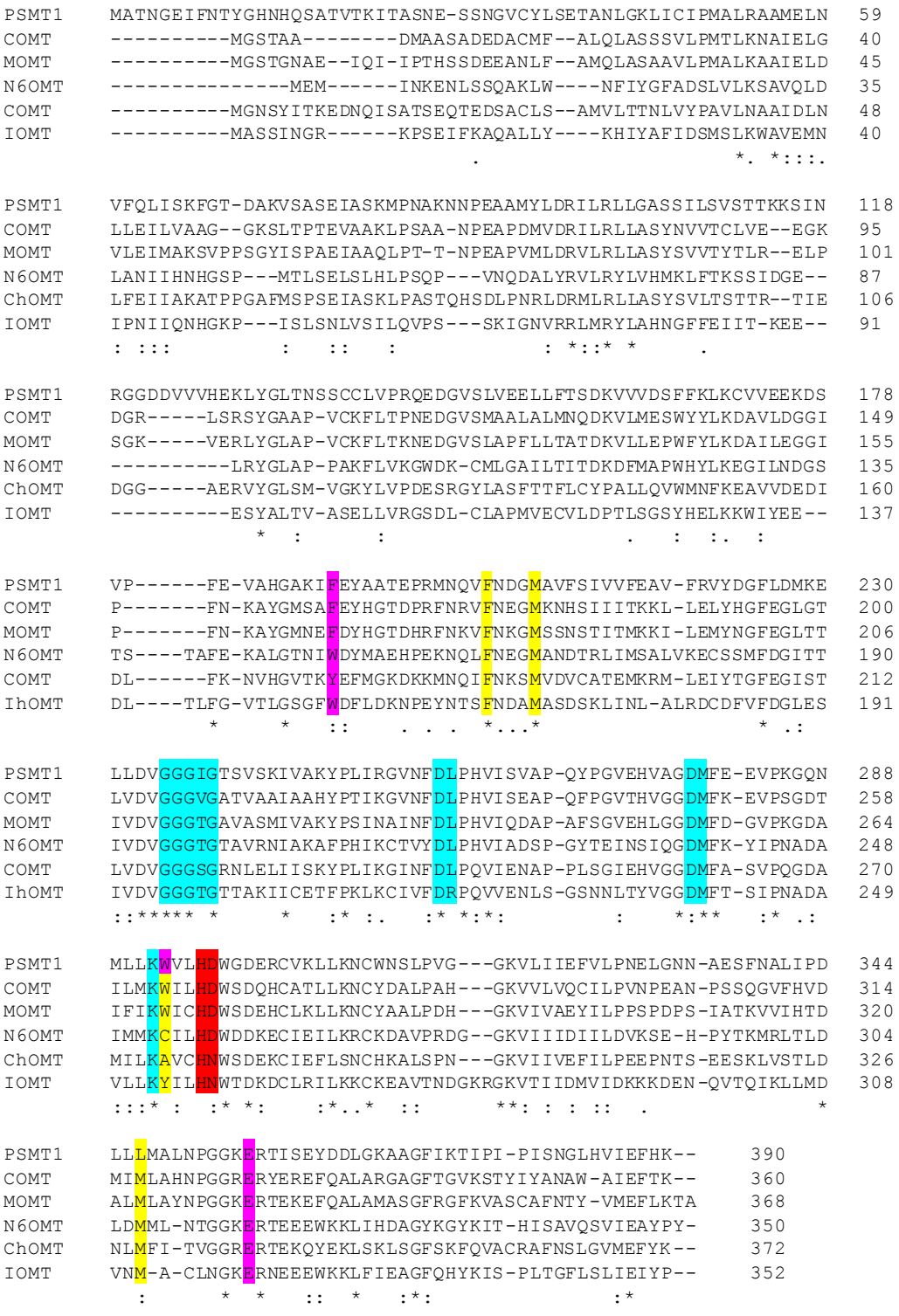



*Figure 4.18: Multiple sequence alignment of conventional plant O-methyltransferases, with a catalytic dyad of His/Asp, for which a tertiary complex in a closed conformation is available in the Protein Database along, with PSMT1. Residues highlighted in Blue are the conserved residue involved in SAM binding as outlined in Figure 1.14 Red residues is the catalytic dyad, yellow residues are conserved in phenolic substrate binding as outlined in Figure 1.16. Pink residues are newly described conserved features*

# 5. Investigating the Catalytic Mechanism of PSMT1

## 5.1. Introduction

Chapter 3 described the successful purification and crystallisation of apoPSMT1 and PSMT1 with SAM bound, both in the closed conformation. In Chapter 4, surface entropy reduction experiments allowed the crystallisation of PSMT1 SER C2 in a closed conformation with SAH and scoulerine or tetrahydrocolumbamine bound. These high resolution closed conformation structures allowed detailed analysis of the scoulerine/tetrahydrocolumbamine binding site, along with the catalytic machinery, Figure 5.1. The conserved plant *O*-methyltransferase catalytic dyad of histidine and aspartic acid were observed to be in a favourable orientation is relation to the 9-O position of scoulerine/tetrahydrocolumbamine for catalysis. To determine the importance of each of these residues site-directed mutagenesis experiments were carried out on His296 and Asp297.

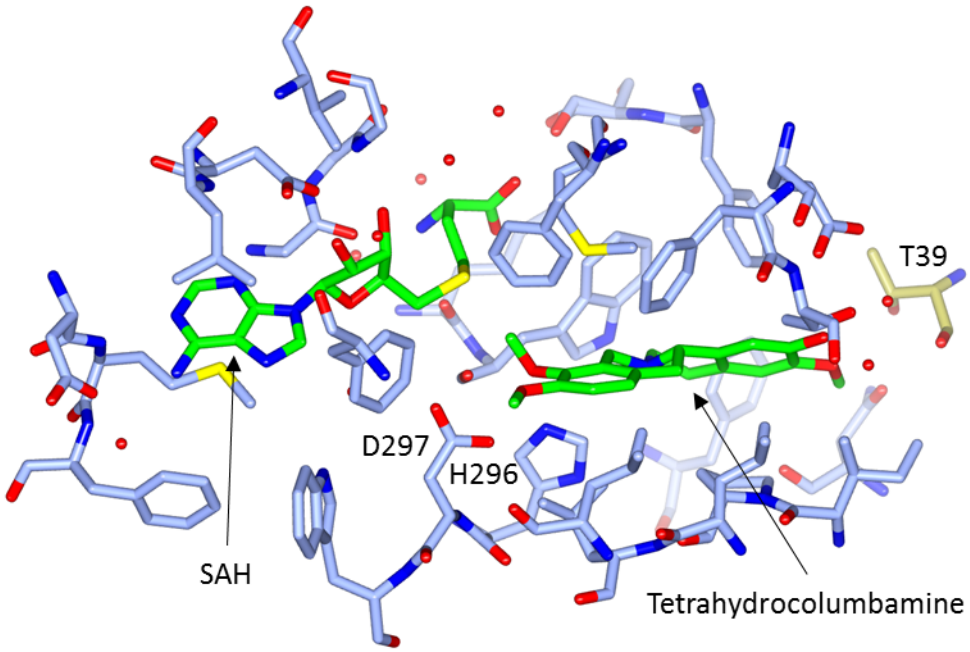



*Figure 5.1: Active site of PSMT1 SER C2 co-crystallised with SAM and tetrahydrocolumbamine with SAH and tetrahydrocolumbamine bound. Ligands shown as green cylinders, PSMT1 SER C2 active site residues shown as cylinders coloured by chain, blue and gold. Waters are represented as red spheres*

Analysis of the scoulerine/tetrahydrocolumbamine binding site in the PSMT1 SER C2 structures revealed an intriguing feature: Thr39 of the opposing subunit directly participates in substrate binding at the opposing end to that of the catalytic dyad, Figure 4.13. Kinetic analysis of PSMT1 showed positive cooperativity with a Hill coefficient of 1.50 ± 0.39 towards scoulerine, Table 3.2. As such, an obvious target for modification of subunit communication is Thr39. In this chapter the Thr39 of PSMT1 SER C2 was mutated to alanine and characterised by X-ray crystallography.

## 5.2. Materials and Methods

### 5.2.1. Generation of Active Site Mutants

Site-directed mutagenesis of pETFPP-3-PSMT1 SER C2 was carried out to generate the active site mutants, utilising the whole plasmid mutagenesis method as previously described in Section 4.2.2, pg. 93. The primers utilised are listed in Table 5.1. PSMT1 SER C2 Thr39Ala site-directed mutagenesis experiments were sequenced using PSMT1 sequencing primer 2, Table 5.1, due to PSMT1 sequencing primer 1 being downstream of the mutation.

### 5.2.1. Expression and Purification

Expression of PSMT1 SER C2 active site mutants were carried out in the same manner to that of PSMT1 SER C2, as described in Section 4.2.3, pg. 94.

Purification of the PSMT1 SER C2 active site mutants was carried out as described in Section 2.2.1.1-3, pg. 54. The purification of PSMT1 SER C2 His296Ala involved size exclusion chromatography, as described in Section 2.2.1.6 pg. 56, in the place of anion exchange chromatography.

| Asp287Ala | |
|---|---|
| | A890C, T891C |
| Forward | 5'-aaaatgggtactgcacg**ct**tggggtgatgaacgat-3' |
| Reverse | 5'-atcgttcatcaccccca**ag**cgtgcagtacccatttt-3' |
| **His296Ala** | |
| | C886G, A887C |
| Forward | 5'-gttgctaaaatgggtactg**gc**cgattggggtgatgaacga-3' |
| Reverse | 5'-tcgttcatcaccccaatcg**gc**cagtacccattttagcaac-3' |
| **His296Phe** | |
| | C886T A887T |
| Forward | 5'-gttgctaaaatgggtactgttcgattggggtgatgaacga-3' |
| Reverse | 5'-tcgttcatcaccccaatcgaacagtacccattttagcaac-3' |
| **His296Asn** | |
| | C166A |
| Forward | 5'-ttgctaaaatgggtactg**a**acgattggggtgatgaac-3' |
| Reverse | 5'-gttcatcaccccaatcgt**t**cagtacccattttagcaa-3' |
| **Thr39Ala** | |
| | A835G, G837A |
| Forward | 5'-caatggtgtctgttatctttcagaa**g**c**a**gctaacttggggaagttaata-3' |
| Reverse | 5'-tattaacttccccaagttagc**t**g**c**ttctgaaagataacagacaccattg-3' |
| **PSMT1 Sequencing Primer 2** | |
| | 5'-GTGAGCGGATAACAATTCC-3' |

*Table 5.1: Oligonucleotide primers for the generation of PSMT1 mutants to probe catalytic mechanism (D287A, H296A, H296F, H296N) and cooperativity (T39A), along with PSMT1 sequencing primer 2. Nucleotide mutations are shown in bold*

### 5.2.2. Activity Assay

The activity of PSMT1 SER C2 Asp296Ala was measured utilising a modified method of that described in Section 4.2.7, pg. 98. Only one concentration of substrates was monitored consisting of 5 µM scoulerine and 100 µM SAM, this was taken to represent $V_{max}$ based on activity assays of PSMT1 SER C2, Figure 4.9 and Table 4.8. 1 mL time points were taken at 0 min, 1 min and 10 min and quenched with equal volumes of methanol. The samples were prepared and analysed as described in Section 4.2.7, pg. 98.

#### 5.2.3. Crystallisation

PSMT1 SER C2 active site mutant crystallisation trays were setup using a 1:1 ratio of 14-15 mg $mL^{-1}$ protein in 10 mM Tris, 5 mM TCEP pH 8.0, utilising the vapour diffusion method in 48 well sitting drop plates, and incubated at 20 °C in a Rigaku Minstrel HT UV crystal imager (Rigaku). Vapour diffusion crystallisation drops were manually dispensed, with a total volume of 1 µL, against 100 µL of reservoir solution. The plates reservoir solutions were designed to explore the crystallisation space around that of the initial PSMT1 SER C2 hit (crystal E) of 0.2 M di-ammonium citrate pH 5.0, 20 % PEG 3350 by altering the pH of di-ammonium citrate between 4.2 and 5.6 along with varying concentration pf PEG 3350 between 10 – 35 %.

Crystals were tested in house using a Rigaku MicroMax-007 HF X-ray Generator and Rigaku R-Axis IV++ image plate detector. 0 ° and 90 ° X-ray diffraction images are taken with an oscillation of 0.5 °, 5-minute exposures to 2.5 Å resolution. Crystals which diffracted well were sent to Diamond Light Source (Harwell) for collection of a full data set.

### 5.3. Results and Discussion

#### 5.3.1. Probing the Active Site Catalytic Dyad

To probe the catalytic dyad of PSMT1, mutations of residues His296 and Asp297 were generated by site-directed mutagenesis of pETFPP-3-PSMT1 PSMT1 SER C2 to remove their respective functional groups. This allowed further characterisation of the catalytic mechanism.

##### *5.3.1.1. PSMT1 Surface Entropy Reduction Cluster 2 Asp297Ala Mutation*

###### 5.3.1.1.1. Expression and Purification

Asp297 of PSMT1 SER C2 was mutated by site-directed mutagenesis experiments utilising the primers outlined in Table 5.1, generating pETFPP-3-PSMT1 SER C2 Asp297Ala. Gene expression was carried out utilising the same methods as those utilised for PSMT1 SER C2. The purification differed to that of PSMT1 SER C2 as only nickel affinity on-column cleavage was implemented before being buffer exchanged

into 10 mM tris pH 8.0, 5 mM TCEP and concentrated to 14 mg $mL^{-1}$ for crystallisation trials. This method produced a protein solution containing one major band at ca. 43 kDa, representing cleaved PSMT1 SER C2 Asp297Ala protein, as analysed by SDS-PAGE, Figure 5.2. There are also two higher molecular weight contaminants of ca. 70 kDa and ca. 64 kDa, these are assumed to be uncleaved 6xHis-GST-r3CP-PSMT1 SER C2 Asp297Ala and HRV 3C protease.

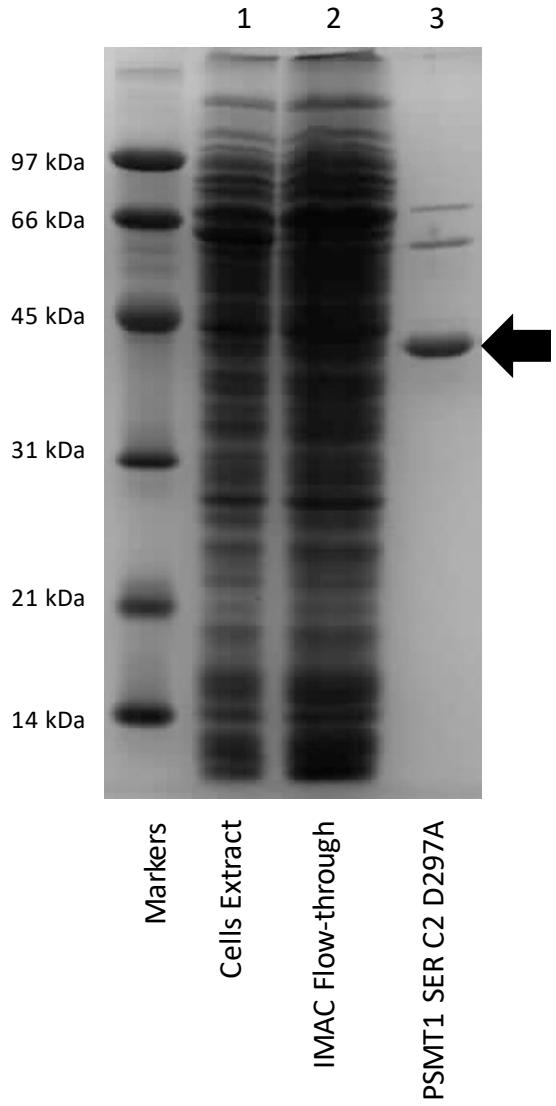



*Figure 5.2: SDS-PAGE analysis of PSMT1 SER C2 Asp297Ala mutant purification steps. Lane 1- cells extract, lane 2 - immobilised metal affinity chromatography (IMAC) flow through, lane 3 – pooled on-column cleavage fractions. Arrow indicates expected position of PSMT1 SER C2 Asp297Ala*

#### 5.3.1.1.2. Activity Assay

To determine if the Asp297Ala mutation has effected the activity of PSMT1, the rate was measured at concentrations of 5 µM scoulerine and 100 µM SAM, which are approaching the concentrations required for $V_{max}$ as determined for PSMT1 SER C2, Figure 4.9. 1 mL samples were taken at 0 min, 1 min and 10 min and quenched with 1 mL of methanol. Samples were prepared and analysed utilising the same method as described for PSMT1 wild type and PSMT1 SER C2 activity assays. The rate over 10 minutes was calculated to be 2.90 nmol $min^{-1}$ $mg^{-1}$ of protein for the PSMT1 SER C2 Asp297Ala mutation, Figure 5.3. The Asp297Ala mutation has resulted in a substantial reduction in activity to 0.6 % of that of the $V_{max}$ calculated by Michaelis-Menten analysis of PSMT1 SER C2, Table 5.2. As the mutation did not result in a complete loss of activity it suggests that Asp297 is important for catalysis but not essential.

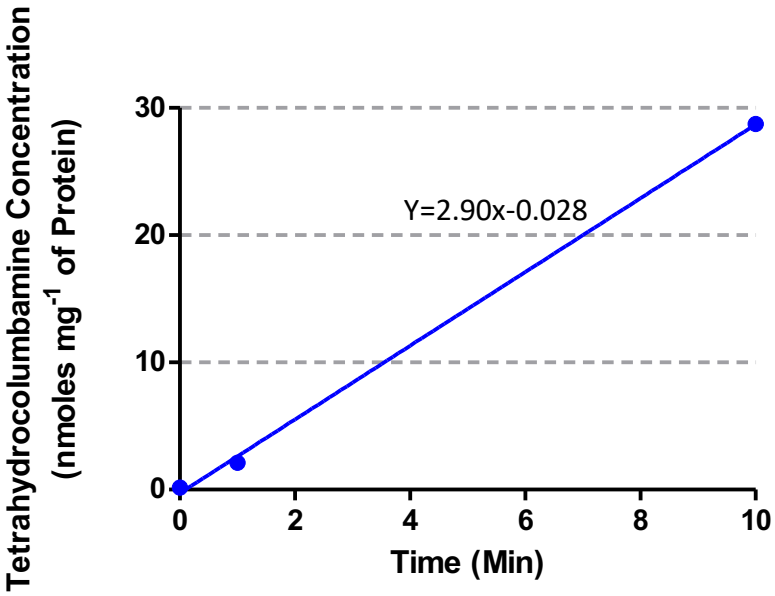



*Figure 5.3: Activity assay of PSMT1 SER C2 Asp297Ala, measuring the production of tetrahydrocolumbamine over time (blue dots), along with linear regression analysis (blue line)*

| | **PSMT1 SER C2 $V_{max}$** | **PSMT1 SER C2 Asp297Ala** |
|---|---|---|
| **Rate of product formation (nmol $min^{-1}$ $mg^{-1}$ of protein)** | 492.6±28.68 | 2.90±0.09 |
| **Relative rate (to $V_{max}$)** | 100 % | 0.6 % |

*Table 5.2: Activity Assay PSMT1 SER C2 Asp297Ala mutant with 5 µM scoulerine and 100 µM SAM measuring tetrahydrocolumbamine production, compared to PSMT1 SER C2 $V_{max}$*

#### 5.3.1.1.3. PSMT1 SER C2 Asp297Ala Structures

Crystallisation of PSMT1 SER C2 Asp297Ala was carried out utilising 48 well sitting drop plates utilising the vapour diffusion method. 0.5 µL of crystallisation buffer was mixed with 0.5 µL 14 mg $mL^{-1}$ protein solution co-crystallised with either 1 mM SAM and 2 mM scoulerine or 0.7 mM SAH and 2 mM scoulerine in 48 well trays, against 100 µL of crystallisation buffer. The crystallisation space around the original hit for PSMT1 SER C2 in JCSG-plus™ screen (Molecular Dimensions), drop A3 (0.2 mM di-ammonium citrate pH 5.0, 20 % PEG3350) was explored. Co-crystallisation experiments with SAH and scoulerine produced crystals in 0.2 M di-ammonium citrate pH 4.4, 20 % PEG3350 which extended to 1.20 Å and co-crystallisation experiments with SAM and scoulerine produced crystals in 0.2 M di-ammonium citrate pH4.6, 20 % PEG3350 which extended to 1.24 Å. Data was auto-processed utilising Xia2 on the beamline at Diamond Light Source (Harwell), statistics shown in Table 5.3. The PSMT1 SER C2 (closed conformation) co-crystallised with SAH and scoulerine structure was used, with the ligands and waters removed, as an initial model for refinement utilising REFMAC for both data sets. After several rounds of refinements, both data sets produced models of PSMT1 SER C2 Asp297Ala with SAH and scoulerine bound, Figure 5.4, refinement statistics are shown in Table 5.3. The Asp297Ala mutation in PSMT1 SER C2 has also resulted in the stabilisation of the missing loop (residues 112 to 130) in one of the subunits. The loop contains a small alpha helix (residues 117-121) and an antiparallel beta-sheet (residues 112-114 and 130-132), Figure 5.4.A.

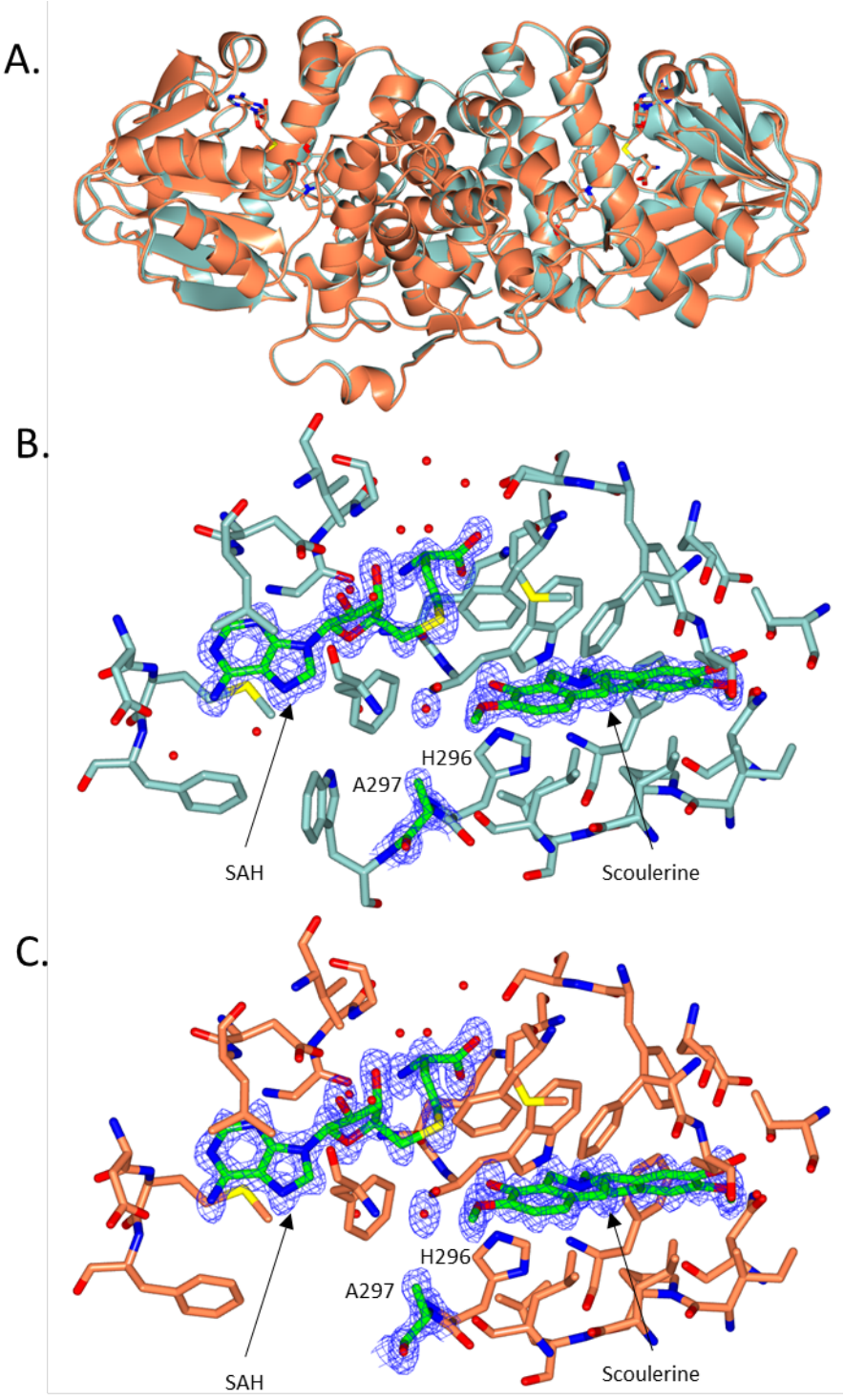



*Figure 5.4: PSMT1 SER C2 Asp297Ala mutant structures co-crystallised with SAH and scoulerine (coral) or SAM and scoulerine (sea green). A. Ribbon diagram of both structures superimposed with ligands represented as cylinders. B. Active site residues of PSMT1 SER C2 co-crystallised with SAH and scoulerine, C. Active site of PSMT1 SER C2 co-crystallised with SAM and scoulerine. Ligands and the Asp297Ala mutation are represented as green cylinders. Blue mesh represents 2Fo-Fc map contoured to 1.5 σ (B. 0.73 e Å$^{-3}$ and C. 0.64 e Å$^{-3}$)*

The PSMT1 SER C2 Asp297Ala mutation structures shown in Figure 5.4 both clearly possess electron density at position 297 which is consistent with the alanine substitution. Both structures have been modelled with SAH and scoulerine bound in the active sites showing that SAM has lost its methyl group. There is an indication that the co-crystallisation trails with SAM and scoulerine have tetrahydrocolumbamine bound at a very low occupancy due to the presence of partial positive electron density in the difference (Fo-Fc) map but the transferred methyl group is not visible in the 2Fo-Fc map so was not modelled.

In both models, the carboxyl group of aspartic acid has been replaced with a highly coordinated water molecule. This water molecule perturbs the positioning of scoulerine in such a way that might explain the marked reduction in catalysis observed in Figure 5.3. The angle observed between His296-Nε, 9-O of tetrahydrocolumbamine and the sulphur of SAH is 130.4 ° in PSMT1 SER C2 co-crystallised with SAM and scoulerine, whereas an angle of 82.2 ° was observed in PSMT1 SER C2 Asp296Ala co-crystallised with SAH and scoulerine. This, along with the reduction in activity suggests that Asp297 alone is not critical for the initial deprotonation step on the 9-hydroxyl of scoulerine but is most likely crucial for the correct positioning of the substrates for the $S_N2$ nucleophilic attack by the 9-phenoxide intermediate on the sulphonium methyl group of SAM.

| **PSMT1 SER C2** | **Asp297Ala** | **Asp297Ala** |
|---|---|---|
| Co-crystallised with | Scoulerine and SAH | Scoulerine and SAM |
| **Data collection** | | |
| Space group | P $2_1$ | P $2_1$ |
| Cell dimensions | | |
| a, b, c (Å) | 68.87, 76.62, 77.16 | 68.25, 75.93, 76.56 |
| (°) | 90.00, 101.55, 90.00 | 90.00, 101.48, 90.00 |
| Resolution (Å) | 76.62-1.20 | 50.19-1.24 |
| Rmerge | 0.04(0.71) | 0.05(0.44) |
| Rpim | 0.04(0.66) | 0.04(0.46) |
| I / σI | 12.2(1.3) | 15.0(2.3) |
| Completeness (%) | 97.5(89.0) | 74.6(18.7) |
| Redundancy | 3.6(2.4) | 3.9(2.5) |
| Xia2 run | 3dii | 3dii |
| | | |
| **Refinement** | | |
| Resolution (Å) | 75.710-1.20 | 50.189-1.24 |
| No. reflections | 238131 | 160995 |
| Rwork / Rfree | 0.13/0.16 | 0.12/0.15 |
| No. atoms | | |
| Protein | 10757 | 10736 |
| Ligand/ion | 174 | 174 |
| Water | 662 | 458 |
| B-factors (Å$^2$) | | |
| Protein | 13.3 | 16.0 |
| Ligand/ion | 7.2 | 10.5 |
| Water | 26.7 | 24.1 |
| r.m.s. deviations | | |
| Bond lengths (Å) | 0.03 | 0.03 |
| Bond angles (°) | 2.6 | 2.4 |

*Table 5.3: Data collection and refinement statistics for PSMT1 SER C2 Asp297Ala mutant*

#### *5.3.1.2. PSMT1 Surface Entropy Reduction Cluster 2 Histidine 296 Mutations*

##### 5.3.1.2.1. Expression and Purification

His296 of PSMT1 SER C2 was mutated by site-directed mutagenesis experiments utilising the primers outlined in Table 5.1, generating pETFPP-3-PSMT1 SER C2 His296Ala. Gene expression was carried out utilising the same methods as those utilised for PSMT1 SER C2. Purification of PSMT1 SER C2 His296Ala was carried out by 3C protease on-column cleavage followed by gel filtration. The first purification step of 3C protease on-column cleavage resulted in a poor yield of cleaved PSMT1 which had been liberated from its solubility/purification tag (6xHis-GST), Figure 5.4.A. The bands in the on-column cleavage fractions contain a small amount of cleaved PSMT1 (43 kDa) compared to that of contaminating proteins, such as un-cleaved 6xHis-GST-r3CP-PSMT1 (70 kDa) and 3C protease (64 kDa). A large band at 43 kDa corresponding to the molecular weight of cleaved PSMT1 was present in the first elution peak of increasing imidazole concentration. This was not observed in previously PSMT1 purification experiments. This would indicate that there has been a structural change which has allowed it to bind to the IMAC column.

To separate the cleaved PSMT1 from the co-eluting proteins in elution peak 1, Figure 5.4.A, gel filtration was implemented and analysed by SDS-PAGE Figure 5.4.B. This showed that cleaved PSMT1 mainly eluted in the void volume (lane 4) by the presence of a band at ca. 43 kDa. Cleaved PSMT1 was also present in many of the subsequent fractions. This is indicative of a misfolded, unfolded or aggregated protein. Gel filtration also known as size exclusion chromatography separates proteins based on their hydrodynamic radius, therefore a protein of the same molecular weight can elute at differing retention times due to differences in folding states. As PSMT1 is co-eluting with proteins of a higher molecular weight, as analysed by SDS-PAGE, it suggests that the mutation has destabilised the enzyme causing it to misfold and form aggregates resulting in earlier elution from the gel filtration column. Purification of PSMT1 SER C2 His296Phe also resulted in no soluble protein.

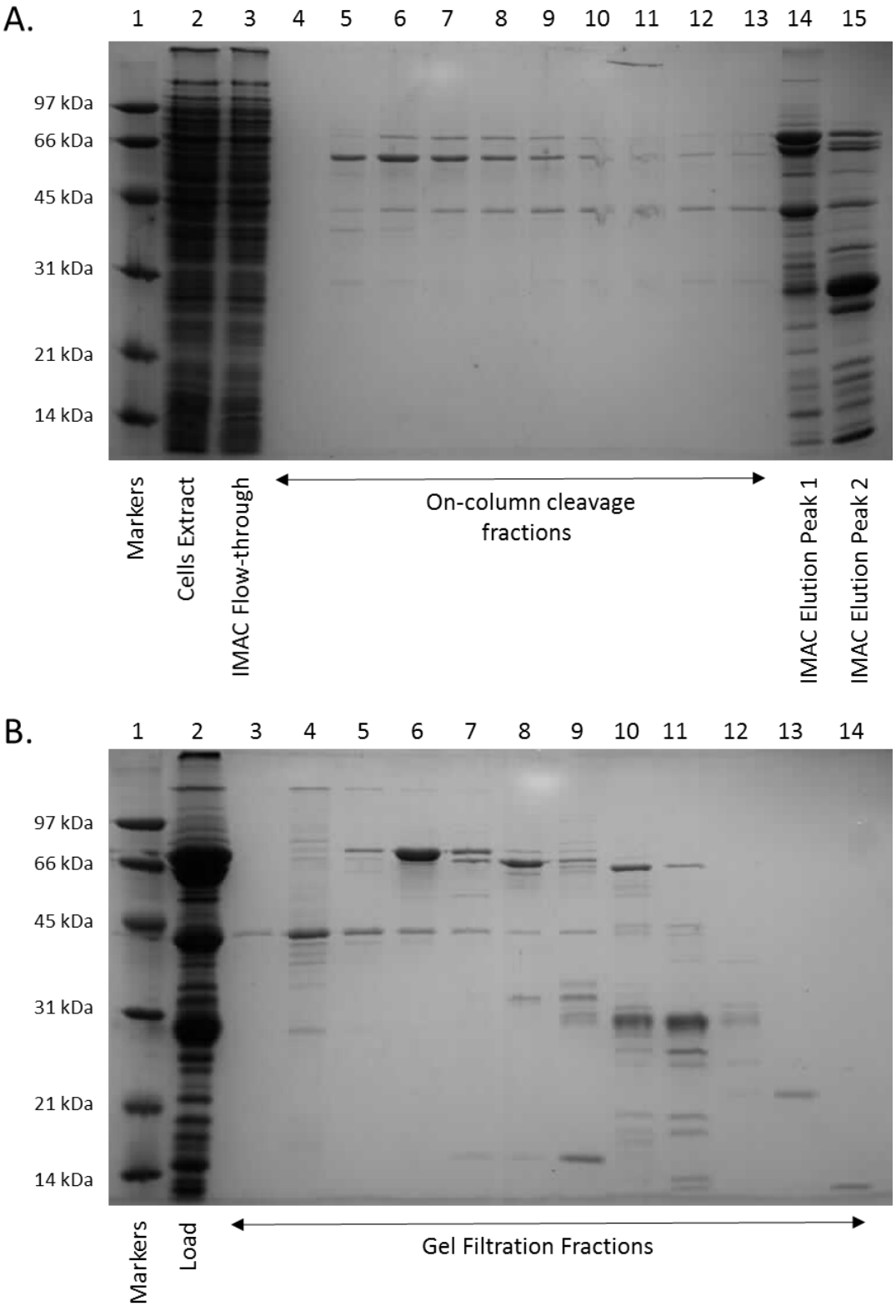



*Figure 5.5: PSMT1 SER C2 His296Ala mutant purification. A. Nickel affinity on-column cleavage, lane 1 – molecular weight markers, lane 2 - cells extract, lane 3 - immobilised metal affinity chromatography (IMAC) flow through, lanes 4-13 – on-column cleavage fractions, lanes 14-15 – IMAC elution peaks. B. Gel Filtration, lane 1 – molecular weight markers, lane 2 – concentrated IMAC elution peak 1 which was injected onto the gel filtration column, lanes 3-14 – gel filtration elution fractions*

Both His296Ala/Phe mutations have resulted in insufficient soluble protein for characterisation. Zhou et al. also reported near complete loss of protein expression when the corresponding histidine was mutated in *Triticum aestivum* L. *O*-methyltransferase to Arg, Leu and Phe.[117] To try to understand why the histidine mutations are causing the reduction in expression the previously obtained crystal structures were analysed. It was found that histidine 296 of PSMT1 forms a hydrogen bond with glutamic acid 356 Oε *via* its Nδ in all the open and closed structures of PSMT1 presented in this thesis. This bond tethers the SAM binding domain to that of the substrate binding/ dimerisation domain and is involved in an extended hydrogen bonding network, Figure 5.6. It is possible that its mutation would destabilise the quaternary structure of the enzyme.

To maintain the hydrogen bond between histidine 296 and glutamic acid 356, His296 was mutated to asparagine. The Oδ of asparagine could, in theory, hydrogen bond to the Oε of glutamic acid 356, Figure 5.7, to hopefully maintaining the hydrogen bond network shown in Figure 5.6. This proved unsuccessful, again resulting in poor protein expression. Histidine is one of the most complex of all amino acids, with it being both polar and hydrophobic, as well as having both positive and negative charges, and possessing aromatic properties. His296 makes a variety of interactions within PSMT1 and it might not be possible to express a stable His296 mutant with the 21 common amino acids. It might be possible to utilise unnatural amino acids such as substituting the imidazole ring with a pyrrole ring generating 2-amino-3-(1H-pyrrol-2-yl) propanoic acid. This might maintain the His296 to Glu356 hydrogen bond as well as maintain its aromaticity but would be unable to deprotonate the 9-hydroxyl group of scoulerine.

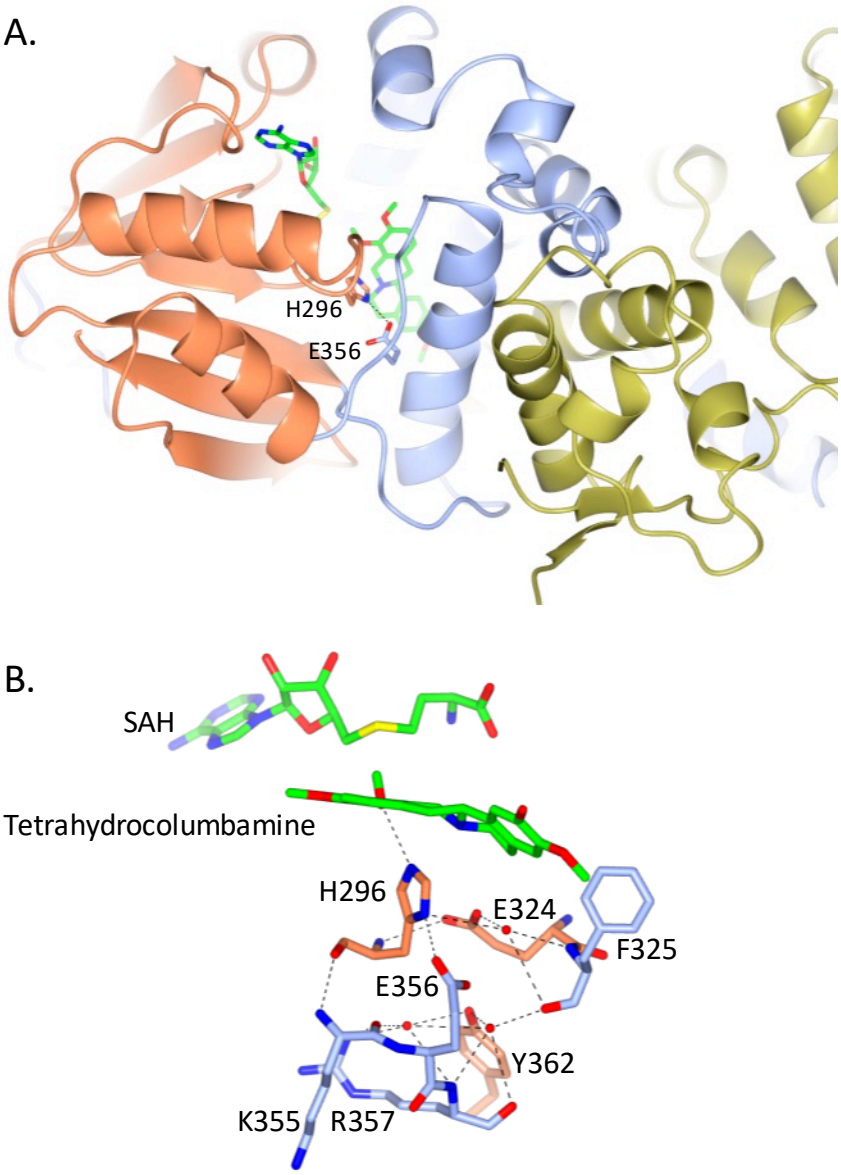



*Figure 5.6: Analysis of the environment of His296 in PSMT1 SER C2 co-crystallised with SAM and scoulerine but with SAH and tetrahydrocolumbamine bound in the active site. A. Ribbon diagram coloured by chain, chain A in ice blue and chain B in gold with chain A's SAM binding domain in coral. B. Hydrogen binding network of His296 represented as cylinders, coloured as an A. Waters shown as red spheres and hydrogen bonds represented as broken black lines*

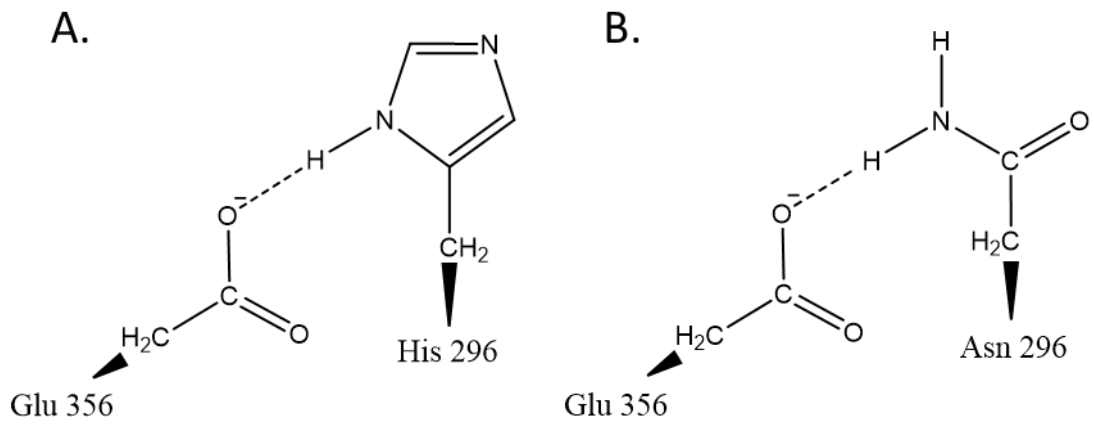



*Figure 5.7: Hydrogen binding between A. His 296 and Asp 356 based on crystal structures, B. proposed hydrogen bonding between His 296 and the Asn 356 mutant. Hydrogen bonds represented as broken black lines*

### 5.3.2. Probing the Cooperativity of PSMT1

#### 5.3.2.1.1. Expression and Purification

In Section 3.3.3, pg. 77 PSMT1 has been shown to possess positive cooperativity in respect to scoulerine with a Hill co-efficient of 1.5, Table 3.2. To probe the subunit cooperativity of the PSMT1 homodimer the X-ray crystal structures presented in this thesis were analysed. The binding site of scoulerine/tetrahydrocolumbamine in PSMT1 SER C2 co-crystallised with SAM and scoulerine but with SAH and tetrahydrocolumbamine bound is show in Figure 5.8 and coloured by peptide chain. Thr39 from the opposing chain interacts directly with tetrahydrocolumbamine by hydrophobic interactions between the 3-O-methyl group of tetrahydrocolumbamine and the Cβ of Thr39. Thr39 also interacts with the 3O of tetrahydrocolumbamine indirectly through a hydrogen bond water bridge *via* its backbone carbonyl group.

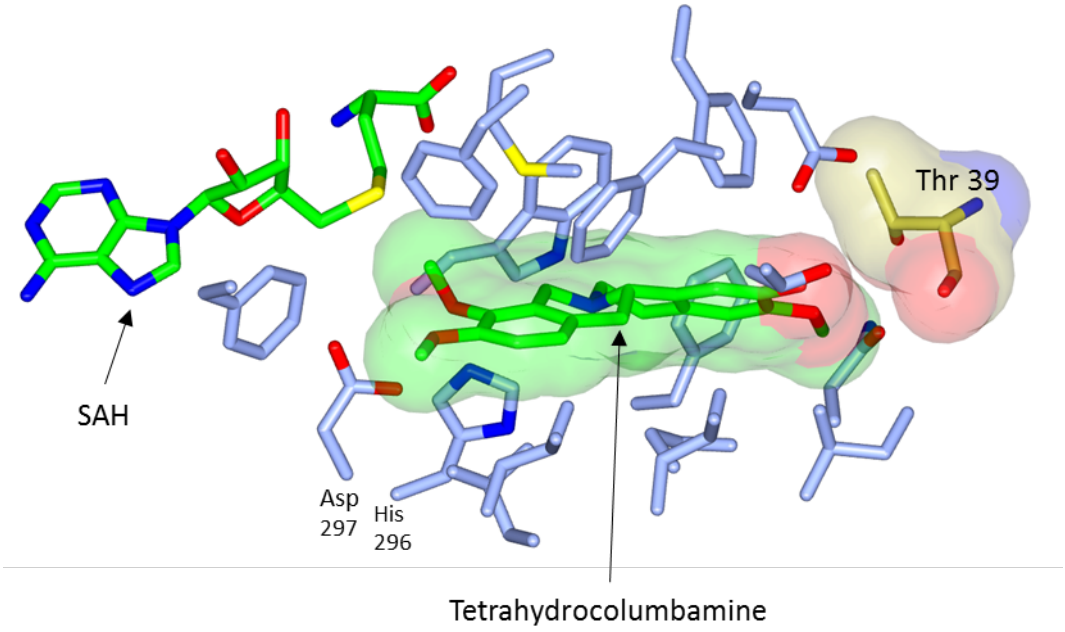



*Figure 5.8: Binding pocket of tetrahydrocolumbamine bound to PSMT1 SER C2. Residues coloured by chain (chain A blue, chain B gold), ligands coloured in green, and Tetrahydrocolumbamine and threonine 39 are represented as a transparent surface model. The catalytic dyad of Asp 297 and His 296 are also indicated*

To probe if Thr39 is a crucial residue contributing to signalling between the two active sites it was mutated to alanine by site-directed mutagenesis experiments utilising the primers in Table 5.1 to generate the plasmid pETFPP-3-PSMT1 PSMT1 SER C2 Thr39Ala.

Gene expression and protein purification were carried out utilising the same methods as those utilised for PSMT1 SER C2 and analysed by SDS-PAGE, Figure 5.9. After purification by nickel affinity on-column cleavage with 3C protease followed by desalting and anion exchange chromatography, a solution of pure PSMT1 SER C2 Thr39Ala was obtained, Figure 5.9. This was buffer exchanged into 10 mM Tris pH 8.0, 5 mM TCEP and concentrated to a concentration of 15 mg $mL^{-1}$ of PSMT1 SER C2 Thr39Ala. PSMT1 SER C2 Thr39Ala behaved the same as PSMT1 SER C2 during expression and purification, resulting in a pure protein sample analysed by SDS-PAGE, Figure 5.9

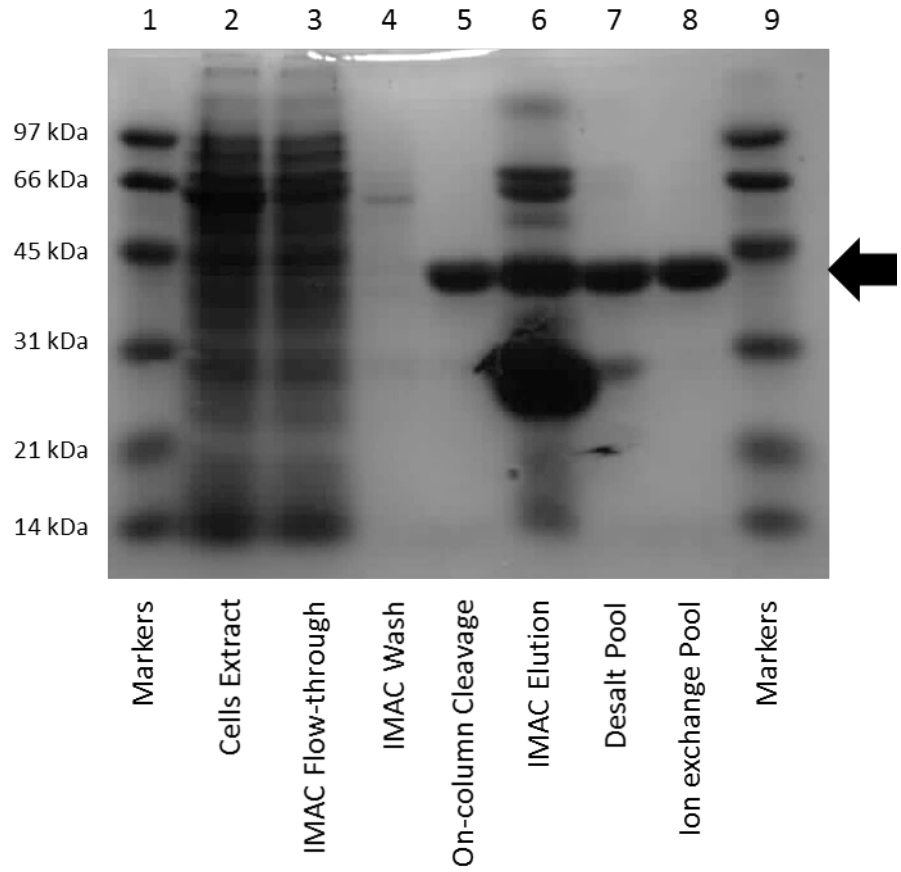



*Figure 5.9: PSMT1 SER C2 Thr39Ala purification. lanes 1 and 9 – molecular weight markers, lane 2 - cells extract, lane 3 - immobilised metal affinity chromatography (IMAC) flow through, lane 4 – IMAC wash step, lane 5- on-column cleavage pool, lane 6 – IMAC elution peak, lane 7 – desalted pool fractions, lane 8 – ion exchanged pooled fractions. Arrow indicates position of PSMT1 SER C2 Thr39Ala*

#### 5.3.2.1.2. PSMT1 SER C2 Thr39Ala Structures

Crystallisation of PSMT1 SER C2 Thr39Ala was carried out utilising 48 well sitting drop plates utilising the vapour diffusion method. 0.5 µL of crystallisation buffer was mixed with 0.5 µL 15 mg $mL^{-1}$ protein solution containing either 0.7 mM SAH and 2 mM scoulerine or 0.7 mM SAH and 2 mM tetrahydrocolumbamine. The crystallisation space around the original hit for PSMT1 SER C2 in JCSG-plus™ screen (Molecular Dimensions), drop A3 (0.2 mM di-ammonium citrate pH 5.0, 20 % PEG3350) was explored. PSMT1 SER C2 Thr39Ala co-crystallised with SAH and scoulerine crystallised in 0.2 mM di-ammonium citrate pH 5.0, 20 % PEG3350 in the space group $P\ 2_1\ 2_1\ 2_1$. Whereas, PSMT1 SER C2 Thr39Ala co-crystallised with SAH and tetrahydrocolumbamine crystallised 0.2 mM di-ammonium citrate pH 5.0, 25 %

PEG3350 in the space group P $2_1$. Refinement statistics for both crystals are shown in Table 5.4.

PSMT1 SER C2 Thr39Ala co-crystallised with SAH and tetrahydrocolumbamine crystallised in the same space group as the previous PSMT1 SER C2 structures, therefore the previous described PSMT1 SER C2 model with SAH and tetrahydrocolumbamine bound, with the ligands and waters removed, was utilised as the initial model for refinement with REFMAC. After several rounds of model building and refinement a final model was produced, Figure 5.10.A. and Table 5.4.

The data set for PSMT1 SER C2 Thr39Ala co-crystallised with SAH and scoulerine proved challenging to solve by molecular replacement utilising Molrep. Several search models were explored, these were a single, full length polypeptide chain from PSMT1 in open and closed conformations, along with truncated models consisting of only the dimerisation domain and only the SAM binding domain. In the end, the data set was solved by utilising the peptide homodimer from the PSMT1 SER C2 model co-crystallised with SAM and scoulerine. This method resulted in an initial model with a Rwork of 0.38 and Rfree of 0.42, consisting of two dimes per asymmetric unit. Analysis of the initial model in COOT showed that for each of the dimers one chain fitted the 2Fo-Fc density very well, whereas, for the second chain only the dimerisation domain had good density. After several iterative rounds of manual model building and refinement utilising COOT and REFMAC a final model was produced, Figure 5.10.B. and Table 5.4.

PSMT1 SER C2 Thr39Ala co-crystallised with SAH and tetrahydrocolumbamine crystallised with two molecules per asymmetric unit, consisting of one homodimer. Both active sites are in the closed conformation with SAH and tetrahydrocolumbamine bound in the active sites, Figure 5.10.A. This closed conformation is consistent with all previous PSMT1 SER C2 structures co-crystallised with SAH or SAM and scoulerine or tetrahydrocolumbamine. Whereas, PSMT1 SER C2 Thr39Ala co-crystallised with SAH and scoulerine generated a model with four molecules per asymmetric unit, consisting of two homodimers. Each of the homodimers have one subunit in a closed conformation with SAH and scoulerine

bound, while the other is in an open conformation with only SAH bound, Figure 5.10.B.

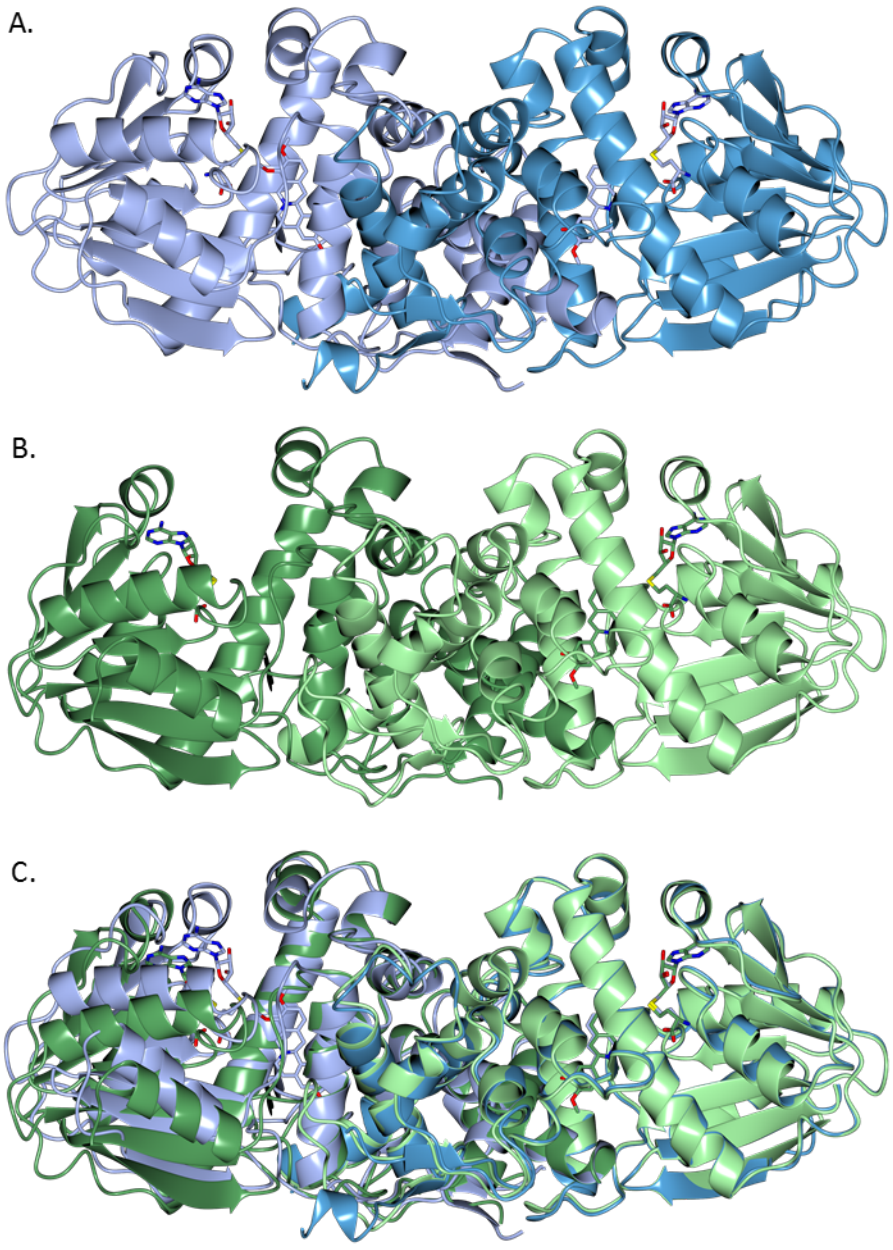



*Figure 5.10: PSMT1 SER C2 Thr39Ala structures represented as ribbon diagrams shaded by domain with ligands represented as cylinders A. PSMT1 SER C2 Thr39Ala co-crystallised with SAH and tetrahydrocolumbamine in a closed conformation with SAH and tetrahydrocolumbamine bound in the active sites, coloured blue, B. PSMT1 SER C2 Thr39Ala co-crystallised with SAH and scoulerine with one active site closed with SAH and scoulerine while the opposing active site is open with only SAH bound, coloured in green. C. Superimposition of A. and B. by the dimerisation domains*

The Thr39Ala mutation has resulted in PSMT1 SER C2 crystallising in a different space group (P $2_1$ $2_1$ $2_1$) when co-crystallised with SAH and scoulerine. Whereas, when co-crystallised with the products (tetrahydrocolumbamine and SAH) PSMT1 SER C2 Thr39Ala crystallised in the same conformation and space group (P $2_1$) as the non-mutant. This leads to the proposal that PSMT1 has a mechanism to recognise the presence the methyl group on tetrahydrocolumbamine. Upon analysis of the two PSMT1 SER C2 Thr39Ala structures, one striking feature stands out. Phe190 which is positioned at the opening of the active site on the dimerisation domain can occupy two conformational states. In all previously presented structures Phe190 is pointing into the active site, occluding the scoulerine/tetrahydrocolumbamine binding pocket. The trend is continued with PSMT1 SER C2 Thr39Ala co-crystallised with tetrahydrocolumbamine and SAH but not with scoulerine and SAH. In the latter structure Phe190 is in a different conformation for the open active site which has only SAH bound, Figure 5.11.A. This new conformation opens-up the scoulerine binding site, Figure 5.11.B&C, possibly acting like a gate. This feature suggests that Thr39 is an important residue in substrate recognition and subunit communication.

Comparison of PSMT1 SER C2 Thr39Ala co-crystallised with SAH and scoulerine (with one open and one closed subunit) to that of the closed structure of PSMT1 SER C2 Asp297Ala, which allowed the building of the flexible loop consisting of residues 110-130, exposed a second feature. Upon comparison of the loops there is a marked conformational change, with that of PSMT1 SER C2 Asp297Ala making little contact with the opposing subunit, Figure 5.12.A. Whereas, in PSMT1 SER C2 Thr39Ala co-crystallised with scoulerine and SAH, the loop from the closed subunit interacts substantially with the opposing, closed subunit, which could possibly be stabilising it in the open conformation, Figure 5.12.B.

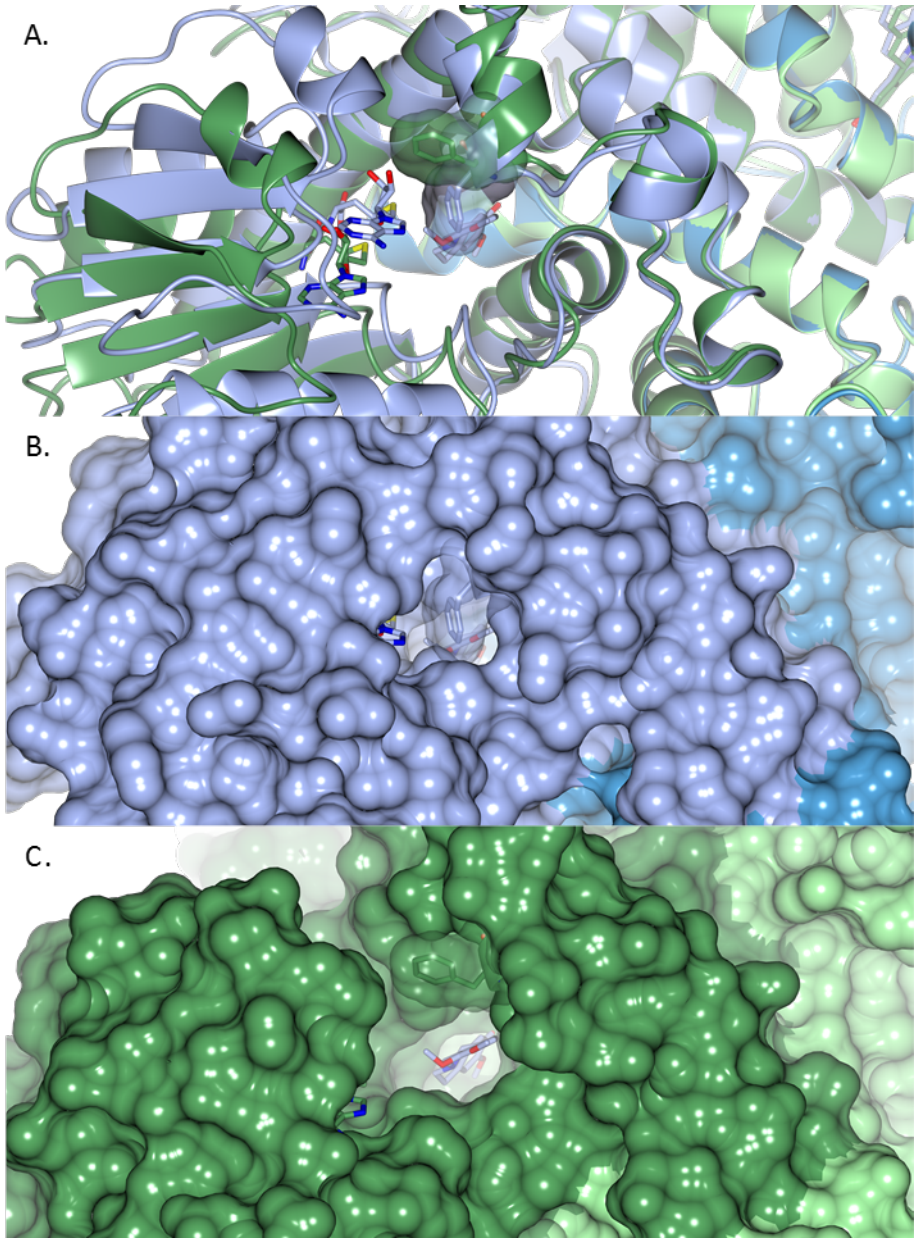



*Figure 5.11: Analysis of Phe190 in PSMT1 SER C2 Thr39Ala mutant structures. A. Ribbon diagram of PSMT1 SER C2 Thr39Ala co-crystallised with SAH and tetrahydrocolumbamine (blue) and co-crystallised with SAH and scoulerine (green). Ligands and Phe190 is represented as cylinders along with a transparent surface representation of Phe190, coloured by model. B. PSMT1 SER C2 Thr39Ala co-crystallised with SAH and tetrahydrocolumbamine (blue) in a closed conformation. Protein is represented as a solid surface with Phe190 being translucent. Ligands and Phe190 are represented as cylinders. C. PSMT1 SER C2 Thr39Ala co-crystallised with SAH and tetrahydrocolumbamine with the open conformation subunit shown in dark green and the closed subunit in light green. Protein is represented as a solid surface with Phe190 being translucent. Ligands and Phe190 are represented as cylinders. Tetrahydrocolumbamine (blue cylinders) is superimposed from B. to show the substrate binding pocket*

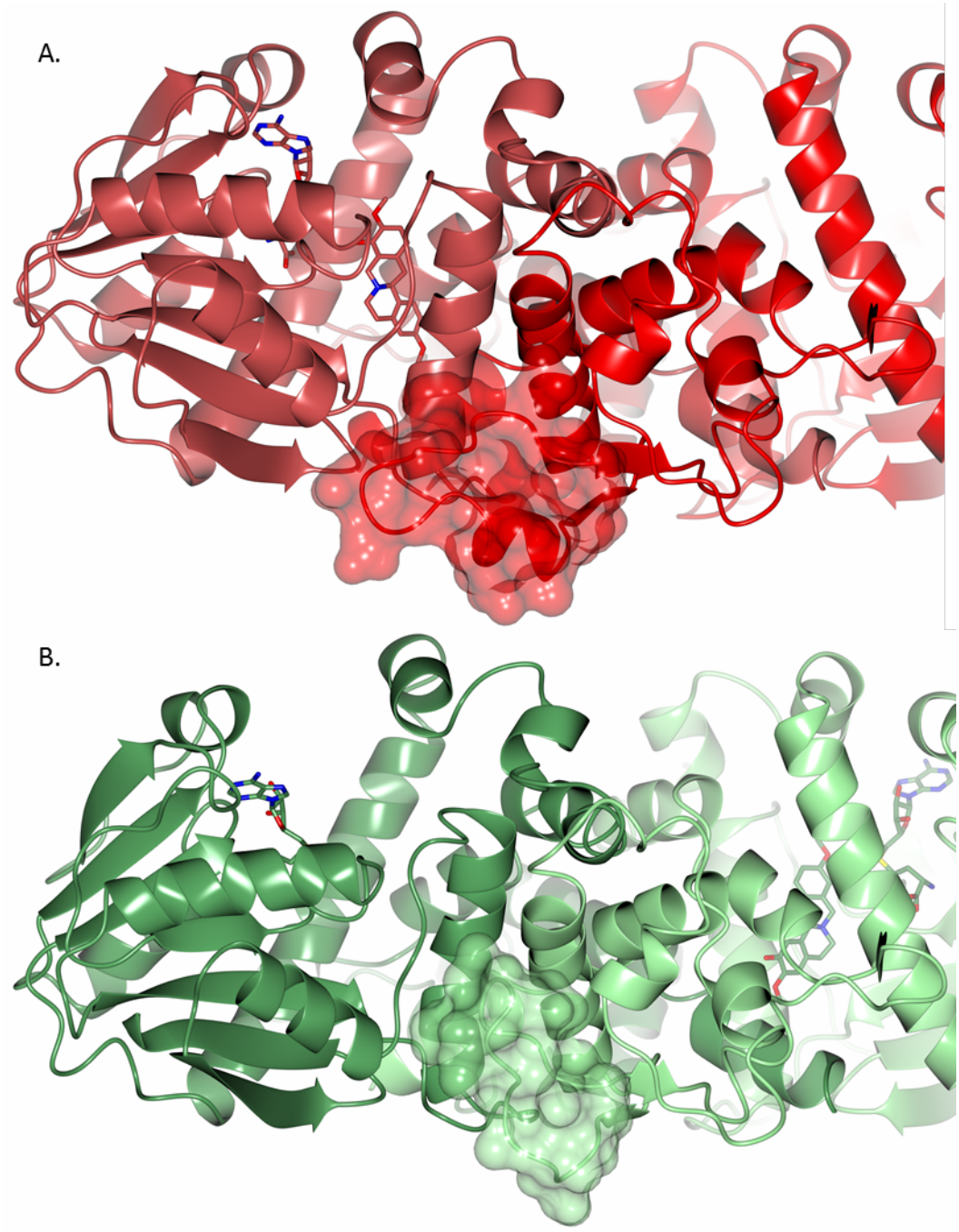



*Figure 5.12: Comparison of the loop consisting of residues 110-130 of PSMT1 SER C2 Asp297Ala co-crystallised with SAH and scoulerine in a closed conformation (A) to that of PSMT1 SER C2 Thr39Ala co-crystallised with SAH and scoulerine with one open and one closed active site (B). Protein is represented as ribbons shaded by chain, ligands are represented as cylinders and residues 110-130 are highlighted as transparent surfaces*

| **PSMT1 SER C2** | **Thr39Ala** | **Thr39Ala** |
|---|---|---|
| Co-crystallised with | Scoulerine and SAH | Tetrahydrocolumbamine and SAH |
| **Data collection** | | |
| Space group | P $2_1$ $2_1$ $2_1$ | P $2_1$ |
| Cell dimensions | | |
| a, b, c (Å) | 82.61, 102.96, 174.26 | 68.15, 75.78, 76.16 |
| (°) | 90.00, 90.00, 90.00 | 90.00, 101.84, 90.00 |
| Resolution (Å) | 88.67-1.98 | 75.78-1.61 |
| Rmerge | 0.10(1.37) | 0.12(1.08) |
| Rpim | 0.04(0.54) | 0.07(0.62) |
| I / σI | 10.0(1.5) | 5.9(1.6) |
| Completeness (%) | 99.80(95.71) | 99.79(97.87) |
| Redundancy | 7.7(8.0) | 2.1(2.1) |
| Xia2 run | DIALS | DIALS |
| | | |
| **Refinement** | | |
| Resolution (Å) | 88.64-1.99 | 74.66-1.61 |
| No. reflections | 98126 | |
| Rwork / Rfree | 0.25/0.30 | 0.15/0.20 |
| No. atoms | | |
| Protein | 20939 | 10468 |
| Ligand/ion | 262 | 182 |
| Water | 47 | 441 |
| B-factors (Å$^2$) | | |
| Protein | 38.2 | 14.9 |
| Ligand/ion | 30.4 | 11.4 |
| Water | 28.3 | 22.8 |
| r.m.s. deviations | | |
| Bond lengths (Å) | 0.02 | 0.03 |
| Bond angles (°) | 1.9 | 1.3 |

*Table 5.4: Data collection and refinement Statistics for PSMT1 SER C2 Thr39Ala Mutant*

## 5.4. Conclusion

Site-directed mutagenesis studies of key residues in PSMT1 SER C2 have proven a powerful technique to probe catalytic mechanism and subunit cooperativity. Plant *O*-methyltransferases have been described as having a catalytic dyad of histidine and aspartic acid which correspond to His296 and Asp297 in PSMT1. The proposed catalytic mechanism based on crystal structures is that His296 deprotonates the 9-hydroxyl group of scoulerine, followed by nucleophilic attack on the sulphonium methyl group of *S*-adenosylmethionine (SAM) by the 9-phenoxide intermediate, generating *S*-adenosylhomocysteine (SAH) and tetrahydrocolumbamine, Figure 4.14.

Mutations of His296 have so far been unsuccessful, resulting in poor protein expression, Figure 5.5. This is most likely due to His296 not only being a crucial catalytic residue but also being involved in structural stability. The mutation of Asp297 in PSMT1 SER C2 has proven successful with Asp297Ala expressing and purifying in the same manner to the wild-type protein. Activity assays of PSMT1 SER C2 Asp297Ala showed a reduction in activity to 0.6% to that of PSMT1 SER C2. This, along with analysis of the crystal structures lead to the proposition that Asp297 is not responsible for the deprotonation step but is crucial for the correct position of the 9-hydroxyl group on scoulerine and the sulphonium-methyl group on SAM for catalysis.

Analysing subunit cooperativity by mutating Thr39 in PSMT1 SER C2 to alanine has introduced more questions than answers. Thr39 interacts directly with scoulerine/tetrahydrocolumbamine in the active site of the opposing subunit, Figure 5.8. Co-crystallisation experiments of PSMT1 SER C2 Thr39Ala with SAH and tetrahydrocolumbamine resulted in the crystallisation of a closed conformation structure with SAH and tetrahydrocolumbamine bound in the active sites, in the same space group to that without the mutation present. In contrast, co-crystallisation experiments with SAH and scoulerine resulted in a new crystal form, under similar crystallisation conditions. The absence of the methyl group at the 9-O position of scoulerine, compared to that of tetrahydrocolumbamine, resulted in a structure with one closed active site and one open active site per homodimer. The closed active site

had both SAH and scoulerine bound, whereas the open active site only had SAH bound. PSMT1 appears to possess the ability to detect the presence or absence of the transferred methyl group on tetrahydrocolumbamine and scoulerine. The Thr39Ala mutation has led to a change in equilibrium between the open and closed conformations which should be further probed by kinetic analysis.

Phe190 has been shown to be a residue of interest due it its large conformational difference in the open active site of PSMT1 SER C2 Thr39Ala co-crystallised with SAH and scoulerine compared to that of the rest of the structures presented within this thesis, Figure 5.11.A. Phe190 appears to be acting as a gate to the scoulerine/tetrahydrocolumbamine binding site, Figure 5.11.B&C., alternatively this could be a crystallographic artefact due to crystal contacts with a symmetry molecule. Residues 110-130 form a loop which appears to act as a regulatory domain, with the loop stabilising the open conformation of one of the subunits in the PSMT1 SER C2 Thr39Ala structure co-crystallised with SAH and scoulerine, Figure 5.12.

Due to the complex nature of subunit communication in PSMT1 it has proven difficult to elucidate the mechanism upon which the observed positive cooperativity is conferred. As a plethora of X-ray crystal structures of PSMT1 have now been solved, in a variety of different conformations, molecular dynamic simulations could be carried out to predict the complicated interactions between protein subunits and protein-ligand interactions which bring about the observed large conformational changes.

# 6. Conclusions and Future Perspectives

PSMT1 with the addition of an N-terminus purification/solubility tag (6xHis-GST-r3CP-PSMT1) was heterologously expressed in *E. coli* BL21 (DE3) cells and the resulting protein purified utilising 3C protease nickel affinity on-column cleavage. This method resulted in the purification and liberation of PSMT1 from it purification/solubility tag. Subsequent purification steps included desalting and anion exchange resulting in pure protein as visualised by SDS-PAGE. Purified PSMT1 was subject to biochemical and structural analysis. Biochemical analysis showed PSMT1 is dimeric in structure with a melting temperature of 50 °C. Enzymatic assays of PSMT1 utilising its natural substrate scoulerine and the co-substrate *S*-adenosylmethionine (SAM) were performed. Direct detection of the *9*-O-methylated product tetrahydrocolumbamine utilising LC-MS/MS proved to be a valid and reproducible method for determining kinetic parameters. Michaelis-Menten analysis was performed with a calculated $V_{max}$ of 492.6 ± 28.7 nmol $min^{-1}$ $mg^{-1}$ of protein and $K_m$ of 0.25 ±0.06 µM with scoulerine as the substrate. With SAM as the substrate a $V_{max}$ of 542.4 ±27.1 nmol $min^{-1}$ $mg^{-1}$ of protein was calculated along with a $K_m$ of 8.76 ±1.6 µM. The convergence of $V_{max}$ is a good indication of a robust and reliable assay. These values differ from that published and is most likely do to the differences in the direct detection method implemented here versus the indirect measurement of product formation utilising extensive liquid-liquid extraction. Further analysis of the activity assay data utilising the sigmoidal Hill plot indicated that there is positive cooperativity in respect to scoulerine with a Hill coefficient of 1.5 ± 0.39.

Crystallisation experiments of PSMT1 produced crystals in 0.1 M bis-tris Propane pH 7.5, 25 % PEG3350, 0.3 M sodium citrate which diffracted to 2.95 Å. This apoPSMT1 structure is a homodimer with a distinct SAM binding domain, dimerisation domain and putative substrate binding domain. Co-crystallisation experiments with SAM did not improve upon a previous data set collected before the onset of this thesis, with PSMT1 co-crystallised with SAM in 0.1 M HEPES pH 7.0, 1 M sodium succinate, 10 % PEG2000 which diffracted to 3.25 Å. This structure confirmed the SAM binding

domain and revealed its mode of binding. Both of these structures have been characterised as being in an open conformation, where the sulphonium methyl group on SAM is distal to the putative scoulerine binding site. Co-crystallisation experiments with scoulerine and SAH or SAM were unsuccessful.

To obtain protein crystals of PSMT1 with the substrates bound, surface entropy reduction (SER) was implemented. High entropy side chains of lysine and glutamic acid were mutated to alanine in an attempt to increase crystal contacts. Surface entropy reduction analysis of PSMT1 resulted in three clusters being identified but cluster 1 was discounted due to it being involved in the dimerisation interface. The remaining two SER clusters were generated by site-directed mutagenesis experiments, along with a double cluster mutant, these were named PSMT1 SER C2, PSMT1 SER C3 and PSMT1 SER C2&3. All three SER enzymes were purified and subjected to biochemical analysis, alongside crystallisation trials. Biochemical analysis showed that the SER proteins were successfully expressed and purified and that the mutations had not effected stability or dimerisation. Initial crystallisation trials of the SER proteins resulted in all three producing crystals which diffracted between 3.2-1.5 Å, with PSMT1 SER C2 co-crystallised with SAH and scoulerine extended to the highest resolution.

The PSMT1 SER C2 co-crystallised with SAH and scoulerine resulted in a closed conformation structure with SAH and scoulerine bound in the active sites. Subsequent co-crystallisation of PSMT1 SER C2 with the substrates, SAM and scoulerine, resulted in crystals which diffracted to 1.3 Å. The resulting structure was also in a closed conformation but the products SAH and tetrahydrocolumbamine were bound in the active sites. As the products are bound in the active site PSMT1 SER C2 has either turned over in-crystal, trapping the products or PSMT1 SER C2 has methylated scoulerine and subsequently rebound SAH and tetrahydrocolumbamine before crystallisation. This indicated that the PSMT1 SER C2 alanine mutations of Lys114Ala and Lys115Ala had not effected activity. This was confirmed by Michaelis-Menten analysis of activity assay data with a calculated $V_{max}$ of 446.7 ± 24.0 nmol $min^{-1}$ $mg^{-1}$ of protein and a $K_m$ of 0.22 ± 0.05 μM with scoulerine as the substrate. This

is similar to activity of the WT enzyme with calculated $V_{max}$ of 492.6 ± 28.7 nmol $min^{-1}$ $mg^{-1}$ of protein and $K_m$ of 0.25 ± 0.06 µM.

Comparison of the open and closed structures revealed that PSMT1 undergoes a large conformational change, where upon binding of the substrates the SAM binding domain rotates inwards towards the substrate binding/dimerisation domains. SAM/SAH binds in an extended conformation to the SAM/SAH binding domain at the N-terminus of the protein containing the conserved Rossmann-like fold. In the open conformation, SAM/SAH binds to PSMT1 mainly on one face with major interactions between the adenine moiety N3 and N6 and the ribose hydroxyl groups with carboxylic acid side chains. The methionine moiety is clamped in place by a lysine interacting with both the carboxyl group and amino group of SAM/SAH. The carboxypropyl moiety of methionine interacts with the conserved Gly-x-Gly motif. The opposing face containing the methyl-sulphonium is exposed to bulk solvent pointing into the open cleft, which forms the active site. Upon closure of the active site, SAM/SAH makes new interaction with residues Met207, Phe203 and Phe190 with the previously exposed face of SAM/SAH. Met207, Phe203 and Phe190, along with Trp293 and Asp297 form a channel in which methyl transfer can occur.

Scoulerine/tetrahydrocolumbamine bind in a deep hydrophobic pocket within the active site with hydrogen bonds at opposing ends of the molecule. The majority of this binding pocket is constructed from the substrate binding/dimerisation domains. Upon closure the conserved catalytic dyad of His296 and Asp297 are brought into proximity to the 9-hydroxy of scoulerine, with the O9 positioned at the opposing entrance to the methyl transfer channel to that of the sulphonium methyl group of SAM. These observation lead to a model for the catalytic mechanism whereby His296 acts as a catalytic base, abstracting a proton from the 9-hydroxyl position of scoulerine generating a phenoxide intermediate. The phenoxide intermediate initiates a $S_N2$ nucleophilic attack on the methyl-sulphonium group of SAM through the methyl transfer channel. Further experiments need to be carried out to determine if His296 is responsible for the initial proton abstraction step in catalysis. This could be achieved by carrying out co-crystallisation experiments with deuterated scoulerine and determining the structure by neutron diffraction analysis. Neutron

diffraction is similar to that of X-ray diffraction but neutrons are scattered off the nucleus of an atom rather than by electrons. Neutron diffraction can differentiate between hydrogen and deuterium atoms due to their differences in scattering length of -3.7406 fm and 6.671 fm, respectively. The differences in scattering will allow the differentiation of the two hydrogen isotopes therefore the residue responsible for the deprotonation of deuterated scoulerine could be visualised. This would only work if PSMT1 binds the substrates and turns over in crystal rather than rebinding the formed products.

Aspartic acid 297 appears to be a key catalytic residue since the Asp297Ala mutation in PSMT1 SER C2 reduced the activity to 0.6 % of PSMT1 SER C2 $V_{max}$. Analysis of the X-ray structure of PSMT1 SER C2 Asp297Ala revealed Asp297 is responsible for the correct positioning of the 9-hydroxyl position on scoulerine for proton abstraction and subsequent nucleophilic attack on SAM. Glu356 terminal carboxylic acid group hydrogen bonds to His296 (Nδ), which is presumably important for the correct positioning of His296 and promoting its alkalinity. Site-directed mutagenesis studies of His296 proved unsuccessful due to insufficient soluble protein for characterisation. As His296 makes a variety of interactions within PSMT1 it might not be possible to express a stable His296 mutant with the 21 common amino acids, but optimisation of gene expression might yield enough soluble protein for biochemical characterisation. Furthermore, it might be possible to utilise unnatural amino acids such as substituting the imidazole ring with a pyrrole ring generating 2-amino-3-(1H-pyrrol-2-yl) propanoic acid. This could maintain the His296 to Glu356 hydrogen bond as well as maintain its aromaticity but would be unable to deprotonate the 9-hydroxyl group of scoulerine.

PSMT1 activity assays showed the propensity for positive cooperativity with a Hill coefficient of 1.5 ±0.39. To investigate the subunit cooperativity of the PSMT1 homodimer, threonine 39 was mutated to alanine in PSMT1 SER C2. Thr39 was chosen as it interacts direct with scoulerine/tetrahydrolcumbamine in the active site of the opposing monomer. PSMT1 SER C2 Thr39Ala co-crystallisation experiments indicated to the ability of PSMT1 to detect the presence or absence of a 9-*O*-methyl group. Co-crystallisation experiments with SAH and tetrahydrocolumbamine, with

the 9-*O*-methyl group present, crystallised in the same space group as PSMT1 SER C2 without the Thr39Ala mutation. Co-crystallisation experiments with SAH and scoulerine, without the 9-*O*-methyl group present, crystallised in an alternate space group in a different conformation but in similar crystallisation conditions. PSMT1 SER C2 Thr39Ala co-crystallised with SAH and scoulerine consisted of two homodimers per asymmetric unit with one closed active site with SAH and scoulerine bound and one open active site with only SAH bound.

Two possible regulatory features involved in active site cooperativity were identified, the first being that of Phe190 and the second that of a regulatory loop consisting of residues 110-130. Phe190 can take one of two conformations. In all PSMT1 structures, except for the open active site of PSMT1 SER C2 Thr39Ala structure co-crystallised with SAH and scoulerine, Phe190 extends inwards towards the scoulerine/tetrahydrocolumbamine binding site, ultimately forming part of the methyl transfer channel. In this inward extended position Phe190 interacts with both the adenine and ribose moieties of SAM/SAH and the 10-*O*-methyl group on scoulerine/tetrahydrocolumbamine, as well as Asp297, which has been shown to be critical for the correct positioning of the substrates for catalysis. In an open conformation Phe190 generally occluded the scoulerine binding site. However, for the open active site of PSMT1 SER C2 Thr39Ala structure co-crystallised with SAH and scoulerine, Phe190 has rotated to extend away from the scoulerine binding site, opening it up allowing unhindered access. The loop consisting of residues 110-130 has only previously been visible in the PSMT1 SER C2 Asp297Ala mutation presumably due to its inherent flexibility. In this structure the 110-130 residue loop extends across from one subunit toward the other but appears to make little contact. However, for PSMT1 SER C2 Thr39Ala co-crystallised with SAH and scoulerine the 110-130 residue loop from the closed subunit is in a different position making significant contact with the opposing subunit. This interaction could be stabilising the open conformation of the opposing open subunit. To determine if the Thr39Ala mutation has effected subunit cooperativity activity assays should be carried out and analysed.

In this thesis PSMT1 has been studied in great detail with the elucidation of high resolution X-ray crystallographic structures. These structures along with several other plant *O*-methyltransferases structures already available in the PDB provide strong evidence for a highly-conserved reaction mechanism. Plant *O*-methyltransferases accept a wide variety of substrates of varying structures with a common phenolic ring which is to be *O*-methylated and generally a hydroxyl group at the opposing end. The current work supports the view that this common phenolic ring is highly conserved and is responsible for the correct positioning of the hydroxyl group for *O*-methylation. The hydroxyl group at the opposing end to the substrates phenolic ring forms hydrogen bonds with an acidic residue or the terminal carbonyl of asparagine, although hydrogen bonding to glutamine is not observed in the current set of available structures it cannot be ruled out. The binding site for the remainder of the substrate is highly variable.

As a variety of plant *O*-methyltransferases with differing ligands bound have had their structures solved, future work needs to be targeted at analysing amino acid sequence of the plant *O*-methyltransferase class of enzymes to establish if it is possible to predict substrate specificity from primary amino acid sequence. Utilising the available crystal structures, sequence analysis and molecular modelling of plant *O*-methyltransferase it might be possible to define a set of rules to be able to predict the substrate from sequence alone based on the 3-dimensional shape of the binding pocket.

# 7. References


1. Aragón-Poce, F. *et al.* History of opium. in *International Congress Series* **1242,** 19–21 (2002).
2. Krikorian, A. D. Were the Opium Poppy and Opium Known in the Ancient near East? *J. Hist. Biol.* **8,** 95–114 (1975).
3. Doull, J. Toxicology Comes of Age. *Annu. Rev. Pharmacol. Toxicol.* **41,** 1–21 (2001).
4. Schiff Jr, P. L. Opium and its alkaloids. *Am. J. Pharm. Educ.* **66,** 186 (2002).
5. Duarte, D. F. Opium and opioids: a brief history. *Rev. Bras. Anestesiol.* **55,** 135–146 (2005).
6. Palliative care for adults: strong opioids for pain relief | Guidance and guidelines |NICE. (2012). Available at: https://www.nice.org.uk/guidance/cg140/chapter/1-Recommendations#starting-strong-opioids-titrating-the-dose. (Accessed: 8th June 2017)
7. Serturner, F. W. Darstellung der reinen Mohnsaure (Opiumsaure) nebst einer chemischen Untersuchung des Opiums unter vorzuglicher Hinsicht auf einen darin neu entdeckten Stoff und die dahin gehorigen Bemerkungen. *J. Pharm.* **14,** 47–93 (1806).
8. Robiquet, P. J. Observations sur le Mémoire de M. Sertuerner Relatif à l'Analyse de l'Opium. *Ann Chim Phys* **5,** 275–278 (1817).
9. Wisniak, J. Pierre-Jean Robiquet. *Educ. Quím.* **24,** 139–149 (2013).
10. Chen, X., Dang, T.-T. T. & Facchini, P. J. Noscapine comes of age. *Phytochemistry* **111,** 7–13 (2015).
11. Robiquet, P. J. Nouvelles Observations sur les Principaux Produits de l'Opium. *Ann Chim Phys* **51,** 225–267 (1832).

12. Uprety, H., Bhakuni, D. S. & Kapil, R. S. Biosynthesis of papaverine. *Phytochemistry* **14,** 1535–1537 (1975).
13. Merck, G. Vorläufige Notiz über eine neue organische Base im Opium. *Justus Liebigs Ann. Chem.* **66,** 125–128 (1848).
14. Medicines and Healthcare products Regulatory Agency. *Morphine Sulphate 10 mg, 15 mg, 20 mg and 30 mg/ml Solution for Injection- PL 17507/0010-3; UK/H/2512/001-4/DC*. (2011).
15. Medicines and Healthcare products Regulatory Agency. *Paracetamol and Codeine Phosphate Omega 500mg/12.8mg film-coated tablets (paracetamol and codeine phosphate hemihydrate) - PL 02855/0245; UK/H/6163/001/DC*. (2017).
16. Tripathi, M., Reddy, P. L. & Rawat, D. S. Noscapine and its analogues as anti-cancer agents. *Chem. Biol. Interface* **4,** 1–22 (2014).
17. Maplestone, R. A., Stone, M. J. & Williams, D. H. The evolutionary role of secondary metabolites — a review. *Gene* **115,** 151–157 (1992).
18. Ziegler, J. & Facchini, P. J. Alkaloid Biosynthesis: Metabolism and Trafficking. *Annu. Rev. Plant Biol.* **59,** 735–769 (2008).
19. Ye, K. *et al.* Opium alkaloid noscapine is an antitumor agent that arrests metaphase and induces apoptosis in dividing cells. *Proc. Natl. Acad. Sci.* **95,** 1601–1606 (1998).
20. Bochkov, D. V., Sysolyatin, S. V., Kalashnikov, A. I. & Surmacheva, I. A. Shikimic acid: review of its analytical, isolation, and purification techniques from plant and microbial sources. *J. Chem. Biol.* **5,** 5–17 (2011).
21. Cotton, R. G. H. & Gibson, F. The biosynthesis of phenylalanine and tyrosine; enzymes converting chorismic acid into prephenic acid and their relationships to

prephenate dehydratase and prephenate dehydrogenase. *Biochim. Biophys. Acta BBA - Gen. Subj.* **100,** 76–88 (1965).

22. Canellakis, Z. N. & Cohen, P. P. PURIFICATION STUDIES OF TYROSINE-α-KETOGLUTARIC ACID TRANSAMINASE. *J. Biol. Chem.* **222,** 53–62 (1956).
23. Martina Ruffer & Zenk. Distant Precursors of Benzylisoquinoline Alkaloids and their Enzymatic Formation : Zeitschrift für Naturforschung C. (1987). Available at: https://www.degruyter.com/view/j/znc.1987.42.issue-4/znc-1987-0402/znc-1987-0402.xml. (Accessed: 16th August 2017)
24. Samanani, N., Liscombe, D. K. & Facchini, P. J. Molecular cloning and characterization of norcoclaurine synthase, an enzyme catalyzing the first committed step in benzylisoquinoline alkaloid biosynthesis. *Plant J.* **40,** 302–313 (2004).
25. Minami, H., Dubouzet, E., Iwasa, K. & Sato, F. Functional Analysis of Norcoclaurine Synthase in Coptis japonica. *J. Biol. Chem.* **282,** 6274–6282 (2007).
26. Pictet, A. & Gams, A. Synthese des Papaverins. *Berichte Dtsch. Chem. Ges.* **42,** 2943–2952 (1909).
27. Galat, A. Synthesis of papaverine and some related compounds. *J. Am. Chem. Soc.* **73,** 3654–3656 (1951).
28. Schläger, S. & Dräger, B. Exploiting plant alkaloids. *Curr. Opin. Biotechnol.* **37,** 155–164 (2016).
29. Hagel, J. M. & Facchini, P. J. Benzylisoquinoline Alkaloid Metabolism: A Century of Discovery and a Brave New World. *Plant Cell Physiol.* **54,** 647–672 (2013).

30. Rida, P. C. G., LiVecche, D., Ogden, A., Zhou, J. & Aneja, R. The Noscapine Chronicle: A Pharmaco-Historic Biography of the Opiate Alkaloid Family and its Clinical Applications. *Med. Res. Rev.* **35,** 1072–1096 (2015).
31. Konzett, H. & Rothlin, E. Zur Wirkung von Narkotin auf den Hustenreflex und auf die Bronchialmuskulatur. *Experientia* **10,** 472–473 (1954).
32. Karlsson, M. O., Dahlström, B. & Neil, A. Characterization of high-affinity binding sites for the antitussive [3H]noscapine in guinea pig brain tissue. *Eur. J. Pharmacol.* **145,** 195–203 (1988).
33. Ke, Y. *et al.* Noscapine inhibits tumor growth with little toxicity to normal tissues or inhibition of immune responses. *Cancer Immunol. Immunother.* **49,** 217–225 (2000).
34. Aneja, R., Zhou, J., Zhou, B., Chandra, R. & Joshi, H. C. Treatment of hormone-refractory breast cancer: apoptosis and regression of human tumors implanted in mice. *Mol. Cancer Ther.* **5,** 2366–2377 (2006).
35. Landen, J. W. *et al.* Noscapine Alters Microtubule Dynamics in Living Cells and Inhibits the Progression of Melanoma. *Cancer Res.* **62,** 4109–4114 (2002).
36. Zhou, J. *et al.* Paclitaxel-resistant Human Ovarian Cancer Cells Undergo c-Jun NH2-terminal Kinase-mediated Apoptosis in Response to Noscapine. *J. Biol. Chem.* **277,** 39777–39785 (2002).
37. Verma, A. K. *et al.* Synthesis and in vitro cytotoxicity of haloderivatives of noscapine. *Bioorg. Med. Chem.* **14,** 6733–6736 (2006).
38. Aneja, R., Ghaleb, A. M., Zhou, J., Yang, V. W. & Joshi, H. C. p53 and p21 Determine the Sensitivity of Noscapine-Induced Apoptosis in Colon Cancer Cells. *Cancer Res.* **67,** 3862–3870 (2007).

39. Jackson, T., Chougule, M. B., Ichite, N., Patlolla, R. R. & Singh, M. Antitumor activity of noscapine in human non-small cell lung cancer xenograft model. *Cancer Chemother. Pharmacol.* **63,** 117–126 (2008).

40. A Study of Noscapine HCl (CB3304 ) in Patients With Relapsed or Refractory Multiple Myeloma - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT00912899?intr=noscapine&rank=1. (Accessed: 16th August 2017)

41. Study of Noscapine for Patients With Low Grade Non Hodgkin's Lymphoma or Chronic Lymphocytic Leukemia Refractory to Chemotherapy-Full Text Available at: https://clinicaltrials.gov/ct2/show/NCT00183950?intr=noscapine&rank=2. (Accessed: 16th August 2017)

42. Naik, P. K. *et al.* Rational design, synthesis and biological evaluations of amino-noscapine: a high affinity tubulin-binding noscapinoid. *J. Comput. Aided Mol. Des.* **25,** 443–454 (2011).

43. Mishra, R. C. *et al.* Second generation benzofuranone ring substituted noscapine analogs: Synthesis and biological evaluation. *Biochem. Pharmacol.* **82,** 110–121 (2011).

44. Anderson, J. T. *et al.* Identification of Novel and Improved Antimitotic Agents Derived from Noscapine. *J. Med. Chem.* **48,** 7096–7098 (2005).

45. Anderson, J. T. *et al.* Discovery of S-Phase Arresting Agents Derived from Noscapine. *J. Med. Chem.* **48,** 2756–2758 (2005).

46. DeBono, A. J. *et al.* Synthesis and Biological Evaluation of N-Substituted Noscapine Analogues. *ChemMedChem* **7,** 2122–2133 (2012).

47. DeBono, A., Capuano, B. & Scammells, P. J. Progress Toward the Development of Noscapine and Derivatives as Anticancer Agents. *J. Med. Chem.* **58,** 5699–5727 (2015).

48. Winzer, T. *et al.* A Papaver somniferum 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science* **336,** 1704–1708 (2012).

49. Samanani, N. & Facchini, P. J. Purification and Characterization of Norcoclaurine Synthase THE FIRST COMMITTED ENZYME IN BENZYLISOQUINOLINE ALKALOID BIOSYNTHESIS IN PLANTS. *J. Biol. Chem.* **277,** 33878–33883 (2002).

50. Luk, L. Y. P., Bunn, S., Liscombe, D. K., Facchini, P. J. & Tanner, M. E. Mechanistic Studies on Norcoclaurine Synthase of Benzylisoquinoline Alkaloid Biosynthesis: An Enzymatic Pictet−Spengler Reaction. *Biochemistry (Mosc.)* **46,** 10153–10161 (2007).

51. Li, J., Lee, E.-J., Chang, L. & Facchini, P. J. Genes encoding norcoclaurine synthase occur as tandem fusions in the Papaveraceae. *Sci. Rep.* **6,** (2016).

52. Ilari, A. *et al.* Structural Basis of Enzymatic (S)-Norcoclaurine Biosynthesis. *J. Biol. Chem.* **284,** 897–904 (2009).

53. Inui, T., Tamura, K., Fujii, N., Morishige, T. & Sato, F. Overexpression of Coptis japonica Norcoclaurine 6- O -Methyltransferase Overcomes the Rate-Limiting Step in Benzylisoquinoline Alkaloid Biosynthesis in Cultured Eschscholzia californica. *Plant Cell Physiol.* **48,** 252–262 (2007).

54. Morishige, T., Tsujita, T., Yamada, Y. & Sato, F. Molecular Characterization of the S-Adenosyl-l-methionine: 3′-Hydroxy-N-methylcoclaurine 4′-O-Methyltransferase Involved in Isoquinoline Alkaloid Biosynthesis in Coptis japonica. *J. Biol. Chem.* **275,** 23398–23405 (2000).

55. Robin, A. Y., Giustini, C., Graindorge, M., Matringe, M. & Dumas, R. Crystal structure of norcoclaurine-6-O-methyltransferase, a key rate-limiting step in the synthesis of benzylisoquinoline alkaloids. *Plant J.* **87,** 641–653 (2016).
56. Choi, K.-B., Morishige, T., Shitan, N., Yazaki, K. & Sato, F. Molecular Cloning and Characterization of CoclaurineN-Methyltransferase from Cultured Cells of Coptis japonica. *J. Biol. Chem.* **277,** 830–835 (2002).
57. Torres, M. A. *et al.* Structural and Functional Studies of Pavine N-Methyltransferase from Thalictrum flavum Reveal Novel Insights into Substrate Recognition and Catalytic Mechanism. *J. Biol. Chem.* **291,** 23403–23415 (2016).
58. Pauli, H. H. & Kutchan, T. M. Molecular cloning and functional heterologous expression of two alleles encoding (S)-N-methylcoclaurine 3'-hydroxylase (CYP80B1), a new methyl jasmonate-inducible cytochrome P-450-dependent mono-oxygenase of benzylisoquinoline alkaloid biosynthesis. *Plant J.* **13,** 793–801 (1998).
59. Frick, S., Kramell, R. & Kutchan, T. M. Metabolic engineering with a morphine biosynthetic P450 in opium poppy surpasses breeding. *Metab. Eng.* **9,** 169–176 (2007).
60. Ortiz de Montellano, P. R. Hydrocarbon Hydroxylation by Cytochrome P450 Enzymes. *Chem. Rev.* **110,** 932 (2010).
61. Pathak, S. *et al.* Comparative Transcriptome Analysis Using High Papaverine Mutant of Papaver somniferum Reveals Pathway and Uncharacterized Steps of Papaverine Biosynthesis. *PLOS ONE* **8,** e65622 (2013).
62. Beaudoin, G. A. W. & Facchini, P. J. Benzylisoquinoline alkaloid biosynthesis in opium poppy. *Planta* **240,** 19–32 (2014).

63. Winzer, T. *et al.* Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. *Science* **349,** 309–312 (2015).

64. Farrow, S. C., Hagel, J. M., Beaudoin, G. A. W., Burns, D. C. & Facchini, P. J. Stereochemical inversion of (S)-reticuline by a cytochrome P450 fusion in opium poppy. *Nat. Chem. Biol.* **11,** 728–732 (2015).

65. Boonstra, B., Rathbone, D. A. & Bruce, N. C. Engineering novel biocatalytic routes for production of semisynthetic opiate drugs. *Biomol. Eng.* **18,** 41–47 (2001).

66. Farrow, S. C. & Facchini, P. J. Dioxygenases Catalyze O-Demethylation and O,O-Demethylenation with Widespread Roles in Benzylisoquinoline Alkaloid Metabolism in Opium Poppy. *J. Biol. Chem.* **288,** 28997–29012 (2013).

67. Dittrich, H. & Kutchan, T. M. Molecular cloning, expression, and induction of berberine bridge enzyme, an enzyme essential to the formation of benzophenanthridine alkaloids in the response of plants to pathogenic attack. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 9969–9973 (1991).

68. Facchini, P. J., Penzes, C., Johnson, A. G. & Bull, D. Molecular Characterization of Berberine Bridge Enzyme Genes from Opium Poppy. *Plant Physiol.* **112,** 1669–1677 (1996).

69. Winkler, A. *et al.* A concerted mechanism for berberine bridge enzyme. *Nat. Chem. Biol.* **4,** 739–741 (2008).

70. Takeshita, N. *et al.* Molecular Cloning and Characterization of S-Adenosyl-L-Methionine:Scoulerine-9-O-Methyltransferase from Cultured Cells of Coptis japonica. *Plant Cell Physiol.* **36,** 29–36 (1995).

71. Dang, T.-T. T. & Facchini, P. J. Characterization of Three O-Methyltransferases Involved in Noscapine Biosynthesis in Opium Poppy1[W]. *Plant Physiol.* **159,** 618–631 (2012).

72. Ikezawa, N., Iwasa, K. & Sato, F. Molecular cloning and characterization of methylenedioxy bridge-forming enzymes involved in stylopine biosynthesis in Eschscholzia californica. *FEBS J.* **274,** 1019–1035 (2007).

73. Dang, T.-T. T. & Facchini, P. J. Cloning and characterization of canadine synthase involved in noscapine biosynthesis in opium poppy. *FEBS Lett.* **588,** 198–204 (2014).

74. Tang, M.-C., Zou, Y., Watanabe, K., Walsh, C. T. & Tang, Y. Oxidative Cyclization in Natural Product Biosynthesis. *Chem. Rev.* **117,** 5226–5333 (2017).

75. Liscombe, D. K. & Facchini, P. J. Molecular Cloning and Characterization of Tetrahydroprotoberberine cis-N-Methyltransferase, an Enzyme Involved in Alkaloid Biosynthesis in Opium Poppy. *J. Biol. Chem.* **282,** 14741–14751 (2007).

76. Dang, T.-T. T. & Facchini, P. J. CYP82Y1 Is N-Methylcanadine 1-Hydroxylase, a Key Noscapine Biosynthetic Enzyme in Opium Poppy. *J. Biol. Chem.* **289,** 2013–2026 (2014).

77. Dang, T.-T. T., Chen, X. & Facchini, P. J. Acetylation serves as a protective group in noscapine biosynthesis in opium poppy. *Nat. Chem. Biol.* **11,** 104–106 (2015).

78. Li, Y. & Smolke, C. D. Engineering biosynthesis of the anticancer alkaloid noscapine in yeast. *Nat. Commun.* **7,** 12137 (2016).

79. Chen, X. & Facchini, P. J. Short-chain dehydrogenase/reductase catalyzing the final step of noscapine biosynthesis is localized to laticifers in opium poppy. *Plant J.* **77,** 173–184 (2014).

80. Schubert, H. L., Blumenthal, R. M. & Cheng, X. Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.* **28,** 329–335 (2003).
81. Cantoni, G. L. Biological Methylation: Selected Aspects. *Annu. Rev. Biochem.* **44,** 435–451 (1975).
82. Struck, A.-W., Thompson, M. L., Wong, L. S. & Micklefield, J. S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications. *ChemBioChem* **13,** 2642–2655 (2012).
83. Attieh, J. M., (§), A. D. H. & (¶), H. S. S. Purification and Characterization of a Novel Methyltransferase Responsible for Biosynthesis of Halomethanes and Methanethiol in Brassica oleracea(*). *J. Biol. Chem.* **270,** 9250–9257 (1995).
84. OHSAWA, N., TSUJITA, M., MORIKAWA, S. & ITOH, N. Purification and Characterization of a Monohalomethane-producing Enzyme S-adenosyl-L-methionine: Halide Ion Methyltransferase from a Marine Microalga, Pavlova pinguis. *Biosci. Biotechnol. Biochem.* **65,** 2397–2404 (2001).
85. Lam, K. C., Ibrahim, R. K., Behdad, B. & Dayanandan, S. Structure, function, and evolution of plant O-methyltransferases. *Genome* **50,** 1001–1013 (2007).
86. Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W. & Roberts, R. J. Crystal structure of the Hhal DNA methyltransferase complexed with S-adenosyl-l-methionine. *Cell* **74,** 299–307 (1993).
87. Galperin, M. Y., Walker, D. R. & Koonin, E. V. Analogous Enzymes: Independent Inventions in Enzyme Evolution. *Genome Res.* **8,** 779–790 (1998).
88. O'Brien, P. J. & Herschlag, D. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **6,** R91–R105 (1999).

89. Anantharaman, V., Aravind, L. & Koonin, E. V. Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Curr. Opin. Chem. Biol.* **7,** 12–20 (2003).

90. Khersonsky, O. & Tawfik, D. S. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annu. Rev. Biochem.* **79,** 471–505 (2010).

91. Gerlt, J. A. & Babbitt, P. C. Divergent Evolution of Enzymatic Function: Mechanistically Diverse Superfamilies and Functionally Distinct Suprafamilies. *Annu. Rev. Biochem.* **70,** 209–246 (2001).

92. Hanukoglu, I. Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. *Biochem. Mol. Biol. Educ.* **43,** 206–209 (2015).

93. Marks, P., McGeehan, J., Wilson, G., Errington, N. & Kneale, G. Purification and characterisation of a novel DNA methyltransferase, M.AhdI. *Nucleic Acids Res.* **31,** 2803–2810 (2003).

94. Dixon, M. M., Huang, S., Matthews, R. G. & Ludwig, M. The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B12. *Structure* **4,** 1263–1275 (1996).

95. Schubert, H. L., Wilson, K. S., Raux, E., Woodcock, S. C. & Warren, M. J. The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase. *Nat. Struct. Mol. Biol.* **5,** 585–592 (1998).

96. Vivek Anantharaman, Eugene V Koonin & L Aravind. SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *J. Mol. Microbiol. Biotechnol.* **4,** 71–75

97. Michel, G. *et al.* The Structure of the RlmB 23S rRNA Methyltransferase Reveals a New Methyltransferase Fold with a Unique Knot. *Structure* **10,** 1303–1315 (2002).

98. Osamu Nureki *et al.* An enzyme with a deep trefoil knot for the active-site architecture. *Acta Crystallogr. Sect. D* **58,** 1129–1137

99. Zhang, X. *et al.* Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell* **111,** 117–127 (2002).

100. Trievel, R. C., Beach, B. M., Dirk, L. M., Houtz, R. L. & Hurley, J. H. Structure and catalytic mechanism of a SET domain protein methyltransferase. *Cell* **111,** 91–103 (2002).

101. Mato, J., Alvarez, L., Ortiz, P. & Pajares, M. A. S-adenosylmethionine synthesis: Molecular mechanisms and clinical implications. *Pharmacol. Ther.* **73,** 265–280 (1997).

102. Chou, T.-C. & Talalay, P. Mechanism of S-adenosyl-L-methionine synthesis by purified preparations of bakers' yeast. *Biochemistry (Mosc.)* **11,** 1065–1073 (1972).

103. Cantoni, G. L. & Durell, J. Activation of Methionine for Transmethylation Ii. the Methionine-Activating Enzyme: Studies on the Mechanism of the Reaction. *J. Biol. Chem.* **225,** 1033–1048 (1957).

104. Giovanelli, J., Mudd, S. H. & Datko, A. H. Quantitative Analysis of Pathways of Methionine Metabolism and Their Regulation in Lemna. *Plant Physiol.* **78,** 555–560 (1985).

105. Schmitt, D., Pakusch, A. E. & Matern, U. Molecular cloning, induction and taxonomic distribution of caffeoyl-CoA 3-O-methyltransferase, an enzyme involved in disease resistance. *J. Biol. Chem.* **266,** 17416–17423 (1991).

106. Zubieta, C., He, X.-Z., Dixon, R. A. & Noel, J. P. Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransferases. *Nat. Struct. Mol. Biol.* **8,** 271–279 (2001).

107. Zubieta, C., Kota, P., Ferrer, J.-L., Dixon, R. A. & Noel, J. P. Structural Basis for the Modulation of Lignin Monomer Methylation by Caffeic Acid/5-Hydroxyferulic Acid 3/5-O-Methyltransferase. *Plant Cell* **14,** 1265–1277 (2002).

108. Liu, C.-J. *et al.* Structural Basis for Dual Functionality of Isoflavonoid O-Methyltransferases in the Evolution of Plant Defense Responses. *Plant Cell* **18,** 3656–3669 (2006).

109. Zhang, K. *et al.* An Engineered Monolignol 4-O-Methyltransferase Depresses Lignin Biosynthesis and Confers Novel Metabolic Capability in Arabidopsis[C][W][OA]. *Plant Cell* **24,** 3135–3152 (2012).

110. Wolters, S. *et al.* Structural analysis of coniferyl alcohol 9-O-methyltransferase from Linum nodiflorum reveals a novel active-site environment. *Acta Crystallogr. D Biol. Crystallogr.* **69,** 888–900 (2013).

111. Green, A. R. *et al.* Determination of the Structure and Catalytic Mechanism of Sorghum bicolor Caffeic Acid O-Methyltransferase and the Structural Impact of Three brown midrib12 Mutations1[W]. *Plant Physiol.* **165,** 1440–1456 (2014).

112. Louie, G. V. *et al.* Structure-Function Analyses of a Caffeic Acid O-Methyltransferase from Perennial Ryegrass Reveal the Molecular Basis for Substrate Preference. *Plant Cell Online* **22,** 4114–4127 (2010).

113. Frick, S. & Kutchan, T. M. Molecular cloning and functional expression ofO-methyltransferases common to isoquinoline alkaloid and phenylpropanoid biosynthesis. *Plant J.* **17,** 329–339 (1999).

114. Walker, A. M. *et al.* The Structure and Catalytic Mechanism of Sorghum bicolor Caffeoyl-CoA O-Methyltransferase. *Plant Physiol.* **172,** 78–92 (2016).

115. Weiss, J. N. The Hill equation revisited: uses and misuses. *FASEB J.* **11,** 835–841 (1997).

116. Kozbial, P. Z. & Mushegian, A. R. Natural history of S-adenosylmethionine-binding proteins. *BMC Struct. Biol.* **5,** 19 (2005).

117. Zhou, J.-M. *et al.* Structure-function relationships of wheat flavone O-methyltransferase: Homology modeling and site-directed mutagenesis. *BMC Plant Biol.* **10,** 156 (2010).

118. Hagel, J. M. & Facchini, P. J. Dioxygenases catalyze the O-demethylation steps of morphine biosynthesis in opium poppy. *Nat. Chem. Biol.* **6,** 273–275 (2010).

119. Lenz, R. & Zenk, M. H. Purification and properties of codeinone reductase (NADPH) from Papaver somniferum cell cultures and differentiated plants. *Eur. J. Biochem. FEBS* **233,** 132–139 (1995).

120. Gasteiger, E *et al.* Protein Identification and Analysis Tools on the ExPASy Server. in *The Proteomics Protocols Handbook* (Humana Press, 2005).

121. Derewenda, Z. S. Rational Protein Crystallization by Mutational Surface Engineering. *Structure* **12,** 529–535 (2004).

122. Goldschmidt, L., Cooper, D. R., Derewenda, Z. S. & Eisenberg, D. Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci. Publ. Protein Soc.* **16,** 1569 (2007).

123. Baud, F. & Karlin, S. Measures of residue density in protein structures. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 12494–12499 (1999).

124. Conte, L. L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites1. *J. Mol. Biol.* **285,** 2177–2198 (1999).

125. Winter, G., Lobley, C. M. C. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69,** 1260–1273 (2013).

126. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 22–25 (2010).

127. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 1002–1011 (2006).

128. Project, C. C. & 4, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763 (1994).

129. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 355–367 (2011).

130. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132 (2004).

131. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).

132. Li, Y. & Smolke, C. D. Engineering biosynthesis of the anticancer alkaloid noscapine in yeast. *Nat. Commun.* **7,** 12137 (2016).

133. Wang, J. & Pichersky, E. Characterization ofS-Adenosyl-l-Methionine:(Iso)eugenolO-Methyltransferase Involved in Floral Scent Production inClarkia breweri. *Arch. Biochem. Biophys.* **349,** 153–160 (1998).

134. Dhar, K. & Rosazza, J. P. N. Purification and Characterization of Streptomyces griseus Catechol O-Methyltransferase. *Appl. Environ. Microbiol.* **66,** 4877–4882 (2000).

135. Petronikolou, N. & Nair, S. K. Biochemical Studies of Mycobacterial Fatty Acid Methyltransferase: A Catalyst for the Enzymatic Production of Biodiesel. *Chem. Biol.* **22,** 1480–1490 (2015).

136. Zhou, M. *et al.* Identification of determinants for tRNA substrate recognition by Escherichia coli C/U34 2'-O-methyltransferase. *RNA Biol.* **12,** 900–911 (2015).

137. Bar-Even, A. *et al.* The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry (Mosc.)* **50,** 4402–4410 (2011).

138. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

139. Rao, S. T. & Rossmann, M. G. Comparison of super-secondary structures in proteins. *J. Mol. Biol.* **76,** 241–256 (1973).

140. Joshi, C. P. & Chiang, V. L. Conserved sequence motifs in plant S-adenosyl-L-methionine-dependent methyltransferases. *Plant Mol. Biol.* **37,** 663–674 (1998).

141. Ma, Q.-H. & Xu, Y. Characterization of a caffeic acid 3-O-methyltransferase from wheat and its function in lignin biosynthesis. *Biochimie* **90,** 515–524 (2008).

142. Dale, G. E., Oefner, C. & D'Arcy, A. The protein as a variable in protein crystallization. *J. Struct. Biol.* **142,** 88–97 (2003).

# 8. Appendix

## Appendix 8.1 pETFPP-3-PSMT1 nucleotide and amino acid sequence coloured by feature

| Feature | Location |
|---|---|
| 6His | 5083 : 5100 |
| Glutathione S-transferase | 5107 : 5760 |
| 3C Protease recognition Site | 5764 : 5787 |
| PSMT1 | 5782 : 6961 |

### A8.1.1 pETFPP-3-PSMT1 nucleotide sequence, 6His-GST-r3CP-PSMT1 gene in bold

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTG
CCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTA
AATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTC
ACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGT
TCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGG
TTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTT
TCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCT
TAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCG
TTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTC
GTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACT
GAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAA
ATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGAC
AATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATAT
TCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATG
CTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTAC
CTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACA
TTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCG
TTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAA
CGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGT
AATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTC
CGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCT
TACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCA
GCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGA
AAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTA
TCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTAT
GGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTA
TCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAG
CGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCA
TATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTG
GGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACA
GACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGG
TAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAG
AAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGT
AAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAA
CATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGG
TCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACA
TAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAG
GTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACC
CCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCT
GCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGC
GACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTAC
GAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGT
TGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTG
CCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTAT
TGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGT
TGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGA
GCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTG
CGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGA
AAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCC

AGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGAT
GCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAAT
AACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACT
GACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGC
TGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCA
ACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGC
TTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGG
CATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCC
ATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGC
AGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCC
CCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATC
GGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGA
TCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAA
TTTTGTTTAACTTTAAGAAGGAGATATACC**ATGGGCAGCAGCCATCATCATCATCATCAC**AGCAGCATGTCCCCTATACT
AGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGGAATATCTTGAAGAAAAATATGAAGAGCATT
TGTATGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTAT
ATTGATGGTGATGTTAAATTAACACAGTCTATGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTG
TCCAAAAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATA
GTAAAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGT
CATAAAACATATTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTATACAT
GGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGT
ACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA
GGCCTGGAAGTTCTGTTCCAGGGACCAGCAATGGCTACCAATGGCGAAATTTTCAATACCTATGGTCATAATCATCAATC
AGCCACAGTCACTAAAATCACTGCTTCTAATGAAAGCAGCAATGGTGTCTGTTATCTTTCAGAAACGGCTAACTTGGGGA
AGTTAATATGCATTCCAATGGCACTAAGAGCTGCGATGGAGCTAAATGTGTTCCAACTTATCTCAAAGTTCGGAACTGAC
GCAAAAGTTTCGGCTTCTGAAATTGCCTCTAAAATGCCAAACGCGAAGAATAATCCTGAAGCAGCTATGTATTTGGATAG
AATTCTTCGACTGCTCGGGGCAAGTTCTATTCTTTCTGTTTCTACTACAAAAAAATCAATCAACAGAGGAGGAGATGATG
TAGTAGTACATGAGAAGCTTTATGGGTTAACAAATTCGTCGTGTTGTTTGGTCCCTCGACAAGAAGACGGGGTGTCATTA
GTCGAAGAATTGCTATTCACATCTGACAAGGTTGTTGTGGATAGTTTTTTCAAACTGAAATGTGTGGTGGAAGAAAAAGA
CAGTGTGCCATTTGAGGTTGCTCATGGTGCTAAAATCTTTGAGTATGCTGCTACAGAACCAAGAATGAATCAAGTATTTA
ACGATGGAATGGCAGTTTTCTCTATTGTTGTTTTTGAGGCTGTTTTTAGAGTTTACGATGGATTTCTTGATATGAAAGAA
TTGTTAGATGTTGGTGGTGGTATTGGTACTTCGGTTAGTAAGATTGTTGCTAAATACCCTTTGATTCGCGGTGTCAACTT
CGACTTGCCTCATGTTATTTCTGTTGCCCCTCAATACCCAGGTGTAGAGCATGTTGCAGGAGATATGTTCGAGGAAGTCC
CAAAGGGTCAAAACATGTTGCTAAAATGGGTACTGCACGATTGGGGTGATGAACGATGTGTGAAGCTGTTAAAGAATTGT
TGGAACTCATTACCTGTGGGTGGAAAAGTTTTGATAATCGAGTTTGTTCTCCCGAATGAACTTGGTAACAATGCTGAATC
ATTCAATGCGTTGATTCCCGATTTACTCCTGATGGCTCTGAATCCAGGCGGTAAAGAGCGAACGATTTCCGAATACGATG
ATTTAGGCAAAGCAGCTGGATTCATAAAAACTATACCTATCCCTATCTCCAATGGTCTTCATGTCATTGAGTTTCACAAA
TGACGCGCCTTCTCCTCACATATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGCGGATCCGAATTCGAGCTCCGTCG
ACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAG
TTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAA
AGGAGGAACTATATCCGGAT

## A8.1.2: 6xHis-GST-r3CP-PSMT1 amino acid sequence in pETFFP-3-PSMT1

MGSSHHHHHH SSMSPILGYW KIKGLVQPTR LLLEYLEEKY EEHLYERDEG 50
DKWRNKKFEL GLEFPNLPYY IDGDVKLTQS MAIIRYIADK HNMLGGCPKE 100
RAEISMLEGA VLDIRYGVSR IAYSKDFETL KVDFLSKLPE MLKMFEDRLC 150
HKTYLNGDHV THPDFMLYDA LDVVLYMDPM CLDAFPKLVC FKKRIEAIPQ 200
IDKYLKSSKY IAWPLQGWQA TFGGGDHPPK GLEVLFQGPA MATNGEIFNT 250
YGHNHQSATV TKITASNESS NGVCYLSETA NLGKLICIPM ALRAAMELNV 300
FQLISKFGTD AKVSASEIAS KMPNAKNNPE AAMYLDRILR LLGASSILSV 350
STTKKSINRG GDDVVVHEKL YGLTNSSCCL VPRQEDGVSL VEELLFTSDK 400
VVVDSFFKLK CVVEEKDSVP FEVAHGAKIF EYAATEPRMN QVFNDGMAVF 450
SIVVFEAVFR VYDGFLDMKE LLDVGGGIGT SVSKIVAKYP LIRGVNFDLP 500
HVISVAPQYP GVEHVAGDMF EEVPKGQNML LKWVLHDWGD ERCVKLLKNC 550
WNSLPVGGKV LIIEFVLPNE LGNNAESFNA LIPDLLLMAL NPGGKERTIS 600
EYDDLGKAAG FIKTIPIPIS NGLHVIEFHK

Appendix 8.2: Compound optimisation report for scoulerine

## Compound Optimization Report

Thermo SCIENTIFIC

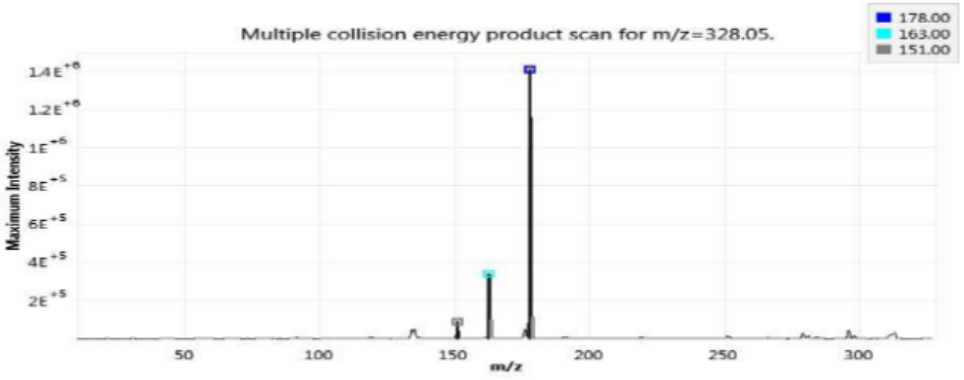
| | |
|---|---|
| Date & Time | Tuesday, December 06, 2016 12:23 PM |
| Instrument Model | TSQ Endura |
| Instrument Serial | TQH-E1-0307 |
| Software Version | 2.0.1292.15 |

| Source Parameters | |
|---|---|
| Source Type | H-ESI |
| Spray Voltage (V) | 3500 |
| Sheath Gas (Arb) | 45 |
| Aux Gas (Arb) | 13 |
| Sweep Gas (Arb) | 1 |
| Ion Transfer Tube Temp (°C) | 342 |
| Vaporizer Temp (°C) | 358 |

| Compound Optimization Input | |
|---|---|
| Source Fragmentation (V) | 0 |
| Step Collision Energy Value | 10 |
| Collision Energy Start (V) | 5 |
| Collision Energy End (V) | 55 |
| Charge State | 1 |
| Step Collision Energy | Yes |
| Unknown/Known Products | Unknown |
| Exclude Loss Masses | None |
| Compound Name | scoulerine |
| Low Mass Exclusion | 10 |
| Product | True |
| CID Gas (mTorr) | 1.5 |
| Adjust Precursor Mass | Yes |
| m/z Value | 328.11 |
| Q1 Resolution | 0.7 |
| Adjust Product Mass | Yes |
| Number of Products | 3 |
| Top N | 3 |
| Optimize RF Lens | Yes |

Compound Optimization Results

| Compound Name | Precursor m/z | Product m/z | Collision Energy (V) | RF Lens (V) | Intensity | Source Fragmentation (V) |
|---|---|---|---|---|---|---|
| scoulerine | 328.049 | 178 | 24.815 | 138.607 | 1880105.577 | 0 |
| scoulerine | 328.049 | 163 | 41.854 | 138.607 | 859826.531 | 0 |
| scoulerine | 328.049 | 151 | 24.258 | 138.607 | 150774.923 | 0 |

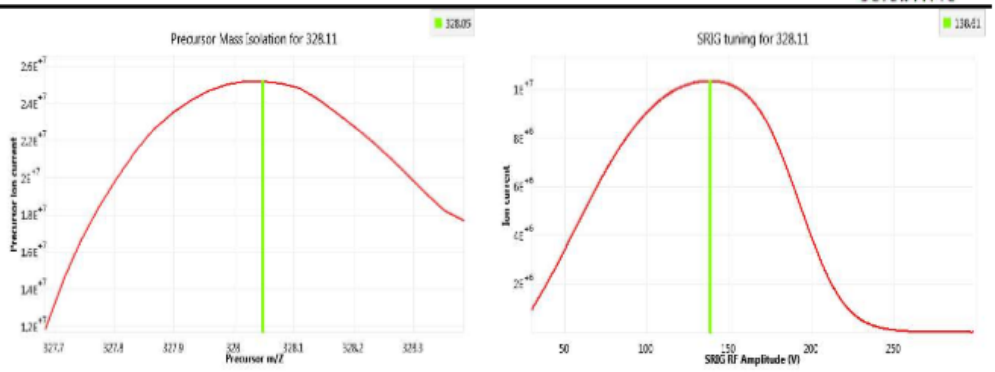


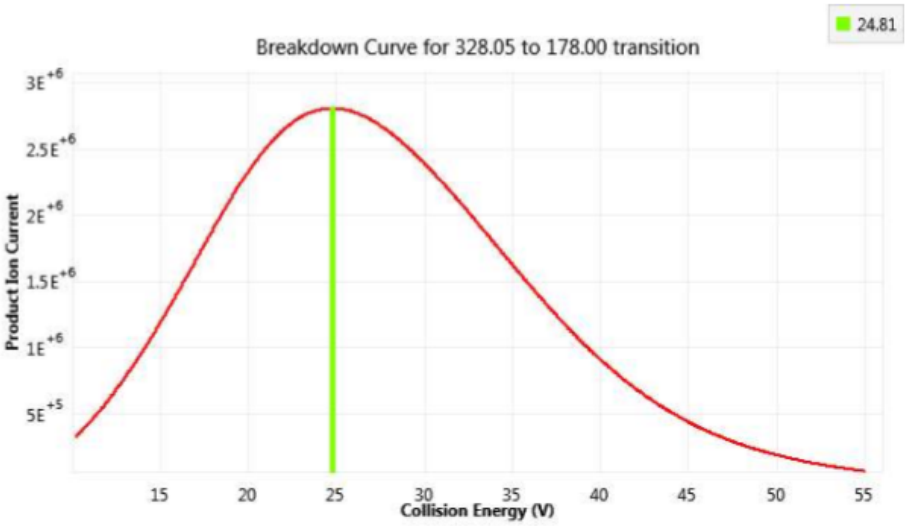

Serial Number: TQH-E1-0307

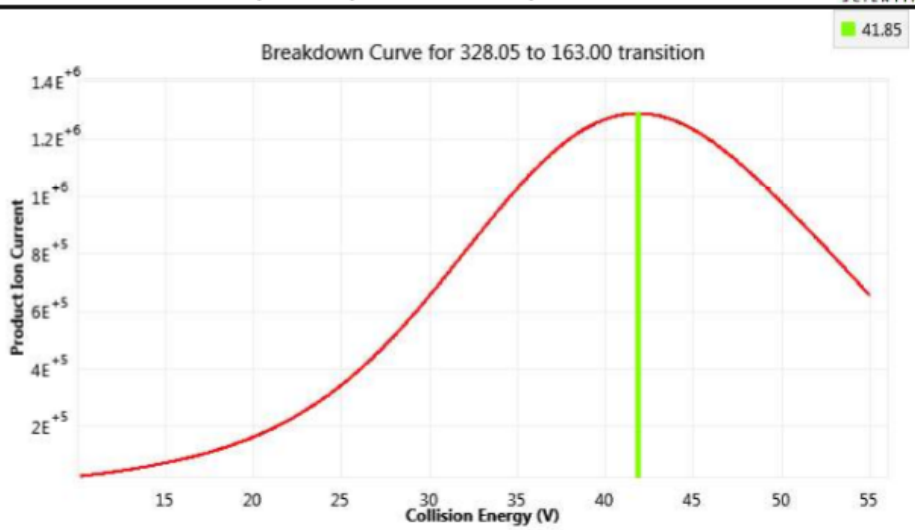
Signature:

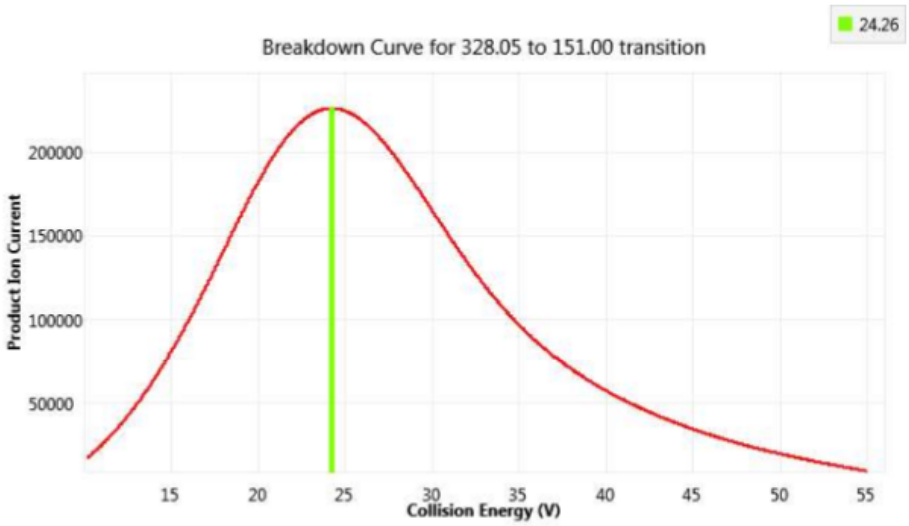
Tuesday, December 06, 2016 12:23 PM

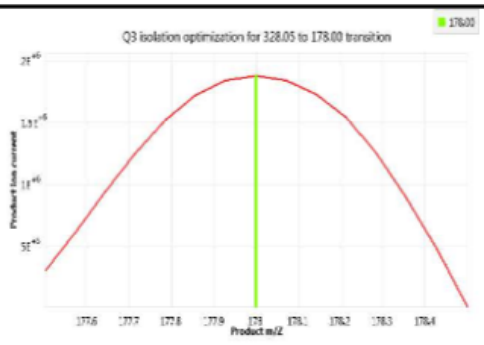
1 of 4

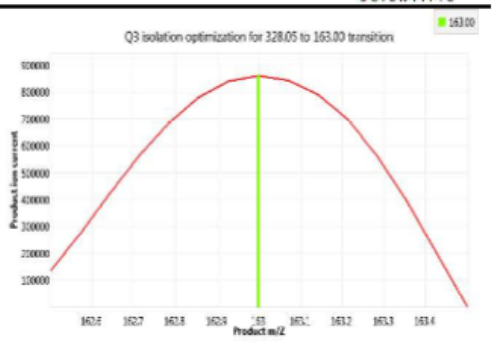

328.05
Precursor Mass Isolation for 328.11
Precursor Ion current
Precursor m/z
136.61
SRIG tuning for 328.11
Ion current
SRIG RF Amplitude (V)

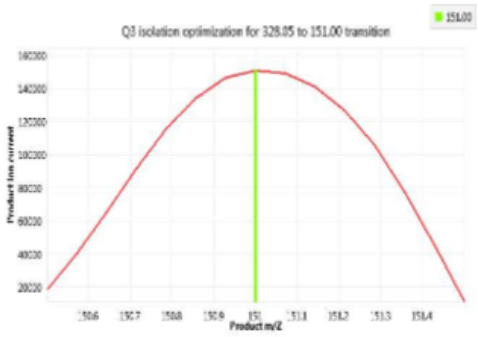

24.81
Breakdown Curve for 328.05 to 178.00 transition
Product Ion Current
Collision Energy (V)


41.85
Breakdown Curve for 328.05 to 163.00 transition
Product Ion Current
1.4E+6
1.2E+6
1E+6
8E+5
6E+5
4E+5
2E+5
15
20
25
30
35
40
45
50
55
Collision Energy (V)


24.26
Breakdown Curve for 328.05 to 151.00 transition
Product Ion Current
200000
150000
100000
50000
15
20
25
30
35
40
45
50
55
Collision Energy (V)


178.00
Q3 isolation optimization for 328.05 to 178.00 transition
Product m/z


163.00
Q3 isolation optimization for 328.05 to 163.00 transition
Product m/z


151.00
Q3 isolation optimization for 328.05 to 151.00 transition
Product m/z

Appendix 8.3 Compound optimisation report for tetrahydrocolumbamine

## Compound Optimization Report

Thermo SCIENTIFIC

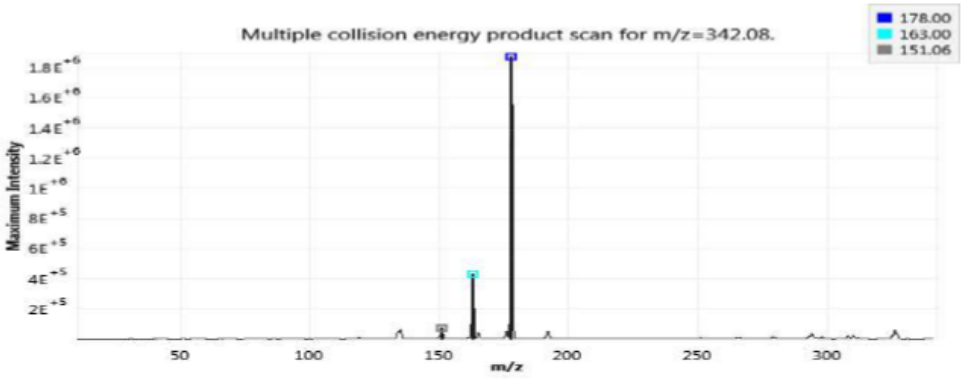
| | |
|---|---|
| Date & Time | Tuesday, December 06, 2016 12:33 PM |
| Instrument Model | TSQ Endura |
| Instrument Serial | TQH-E1-0307 |
| Software Version | 2.0.1292.15 |

| Source Parameters | |
|---|---|
| Source Type | H-ESI |
| Spray Voltage (V) | 3500 |
| Sheath Gas (Arb) | 45 |
| Aux Gas (Arb) | 13 |
| Sweep Gas (Arb) | 1 |
| Ion Transfer Tube Temp (°C) | 342 |
| Vaporizer Temp (°C) | 358 |

| Compound Optimization Input | |
|---|---|
| Source Fragmentation (V) | 0 |
| Step Collision Energy Value | 10 |
| Collision Energy Start (V) | 5 |
| Collision Energy End (V) | 55 |
| Charge State | 1 |
| Step Collision Energy | Yes |
| Unknown/Known Products | Unknown |
| Exclude Loss Masses | None |
| Compound Name | Tetrahyrocolumbamine |
| Low Mass Exclusion | 10 |
| Product | True |
| CID Gas (mTorr) | 1.5 |
| Adjust Precursor Mass | Yes |
| m/z Value | 342.11 |
| Q1 Resolution | 0.7 |
| Adjust Product Mass | Yes |
| Number of Products | 3 |
| Top N | 3 |
| Optimize RF Lens | Yes |

Compound Optimization Results

| Compound Name | Precursor m/z | Product m/z | Collision Energy (V) | RF Lens (V) | Intensity | Source Fragmentation (V) |
|---|---|---|---|---|---|---|
| Tetrahyrocolumbamine | 342.08 | 178 | 24.815 | 144.674 | 2453400.746 | 0 |
| Tetrahyrocolumbamine | 342.08 | 163 | 42.511 | 144.674 | 1123079.981 | 0 |
| Tetrahyrocolumbamine | 342.08 | 151.058 | 24.815 | 144.674 | 135435.603 | 0 |

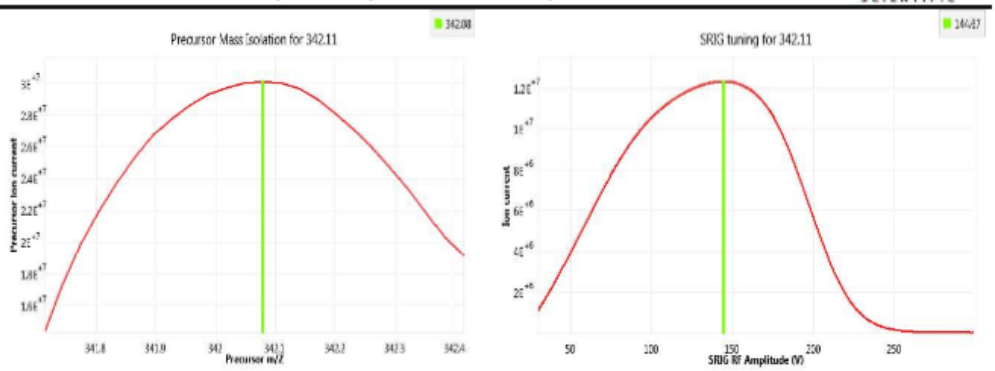


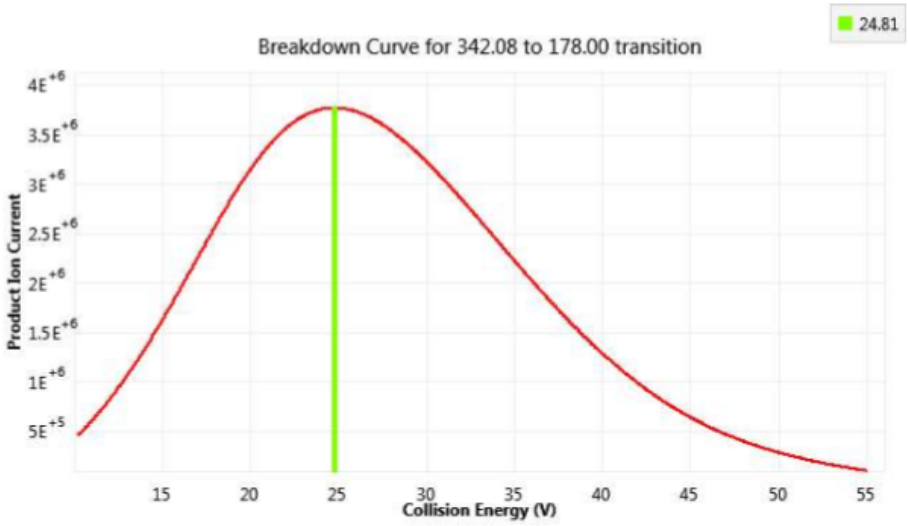

Serial Number: TQH-E1-0307

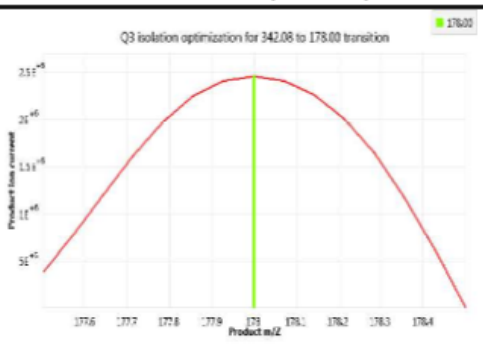
Signature:

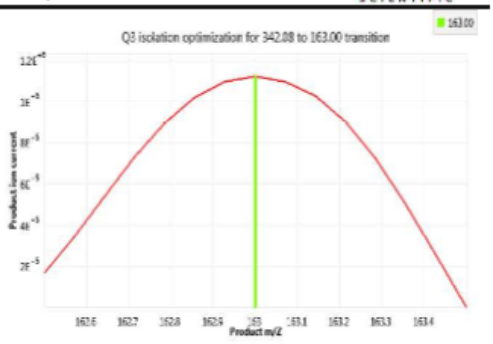
Tuesday, December 06, 2016 12:33 PM

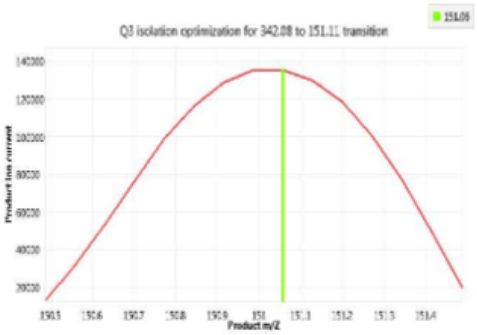
1 of 4


342.08
Precursor Mass Isolation for 342.11
Precursor Ion current
Precursor m/z
344.07
SRIG tuning for 342.11
Ion current
SRIG RF Amplitude (V)


24.81
Breakdown Curve for 342.08 to 178.00 transition
Product Ion Current
Collision Energy (V)


176.00
Q3 isolation optimization for 342.08 to 178.00 transition
Product Ion current
Product m/Z


163.00
Q3 isolation optimization for 342.08 to 163.00 transition
Product Ion current
Product m/Z


151.06
Q3 isolation optimization for 342.08 to 151.11 transition
Product Ion current
Product m/Z

Appendix 8.4: Sequence Alignment of PSMT1 and PsSOMT1 amino acid sequences by MUSCLE multisequence alignment tool. Mismatches are highlighted

```
PSMT1           MATNGEIFNTYGHNHQSATVTKITASNESSNGVCYLSETANLGKLICIPMALRAAMELNV
SOMT1           MATNGEIFNTYGHNRQTATVTKITASNESSNGVCYLSETANLGKLICIPMALRAAMELNV
                **************.*:****************************************

PSMT1           FQLISKFGTDAKVSASEIASKMPNAKNNPEAAMYLDRILRLLGASSILSVSTTKKSINRG
SOMT1           FQLISKFGTDAKVSASEIASKMPNAKNNPEAAMYLDRILRLLGASSILSVSTTKKSINRG
                ************************************************************

PSMT1           GDDVVVHEKLYGLTNSSCCLVPRQEDGVSLVEELLFTSDKVVVDSFFKLKCVVEEKDSVP
SOMT1           GDDVVVHEKLYGLTNSSCCLVPRQEDGVSLVEELLFTSDKVVVDSFFKLKCVVEEKDSVP
                ************************************************************

PSMT1           FEVAHGAKIFEYAATEPRMNQVFNDGMAVFSIVVFEAVFRVYDGFLDMKELLDVGGGIGT
SOMT1           FEVAHGAKIFEYAATEPRMNQVFNDGMAVFSIVVFEAVFREYDGFLDMKELLDVGGGIGT
                ****************************************.*******************

PSMT1           SVSKIVAKYPLIRGVNFDLPHVISVAPQYPGVEHVAGDMFEEVPKGQNMLLKWVLHDWGD
SOMT1           SVSKIVAKYPLIRGVNFDLPHVISVAPQYPGVEHVAGDMFEEVPKGQNMLLKWVLHDWGD
                ************************************************************

PSMT1           ERCVKLLKNCWNSLPVGGKVLIIEFVLPNELGNNAESFNALIPDLLLMALNPGGKERTIS
SOMT1           ERCVKLLKNCWNSLPVGGKVLIIEFVLPNELGNNAESFNALIPDLLLMALNPGGKERTIS
                ************************************************************

PSMT1           EYDDLGKAAGFIKTIPIPISNGLHVIEFHK
SOMT1           EYDDLGKAAGFIKTIPIPISNGLHVIEFHK
                ******************************
```