White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/143729/

Version: Accepted Version

Article:
Cabry, Marc P., Offen, Wendy A., Saleh, Philip et al. (2019) Structure of Papaver somniferum O-Methyltransferase 1 Reveals Initiation of Noscapine Biosynthesis with Implications for Plant Natural Product Methylation. ACS Catalysis. 3840–3848. ISSN: 2155-5435

https://doi.org/10.1021/acscatal.9b01038

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Structure of *Papaver somniferum* O-Methyltransferase 1 Reveals Initiation of Noscapine Biosynthesis with Implications for Plant Natural Product Methylation

Marc P. Cabry[1,2], Wendy A. Offen[2], Philip Saleh[2], Yi Li[1], Thilo Winzer[1], Ian A. Graham[1,*], Gideon J. Davies[2,*]

[1]Centre for Novel Agricultural Products, Department of Biology, University of York, York, YO10 5DD, United Kingdom

[2] York Structural Biology Laboratory, Department of Chemistry, University of York, York, YO10 5DD, United Kingdom

*Correspondence to ian.graham@york.ac.uk or gideon.davies@york.ac.uk

Abstract:

The opium poppy, *Papaver somniferum,* has been a source of medicinal alkaloids since the earliest civilizations; ca. 3400 B.C. The benzylisoquinoline alkaloid noscapine is produced commercially in *P. somniferum* for use as a cough suppressant and it also has potential as an anticancer compound. The first committed step in the recently elucidated noscapine biosynthetic pathway involves the conversion of scoulerine to tetrahydrocolumbamine by 9-*O*-methylation, catalysed by *O*-methyltransferase 1 (PSMT1). We demonstrate, through protein structures (obtained through rational crystal engineering at resolutions from 1.5 to 1.2Å for the engineered variants) across the reaction coordinate, how domain closure allows specific methyl transfer to generate the product. SAM-dependent methyl transfer is central to myriad natural products in plants, analysis of amino-acid sequence, now taking the three-dimensional structure of PSMT1 and low identity homologs into account, begins to shed light on the structural features that govern substrate specificity in these key, ubiquitous, plant enzymes. We propose how "gatekeeper" residues can determine acceptor regiochemistry thus allowing prediction across the wide genomic resource.

KEYWORDS: Poppy, medicinal plants, enzyme, three-dimensional structure, enzymatic catalysis, alkaloid

Introduction

The medicinal properties of *Papaver somniferum* have been known and exploited since the first civilizations[1-2]. 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 utilised as a cough suppressant[3], but recent research has been focused on its antitumor properties[4-6]. Unlike the better-known opiates, noscapine has no addictive properties[7]. The recent engineering of noscapine and a number of halogenated derivatives in *Saccharomyces cerevisiae* [8] opens up a new route to development of noscapine and its analogues as potential drug leads.

A 10-gene cluster encoding all but one of the enzymes responsible for the conversion of scoulerine to noscapine was discovered and characterised by virus induced gene silencing and heterologous gene expression and this together with other studies has led to the proposition of the full biosynthetic pathway presented in **Figure S1**[9-12]. The enzyme PSMT1, is the first enzymatic reaction catalysed by this gene cluster. The enzyme catalyses the 9-*O*-methylation of scoulerine by the *S*-adenosylmethionine (SAM) dependent *Papaver somniferum O*-methyltransferase 1 (PSMT1)[9, 13], **Figure 1a**.

PSMT1 belongs to a group of ubiquitous enzymes termed methyltransferases that catalyse *S*-adenosylmethionine dependent methyl transfer. Of particular interest, notably in plant methyltransferases, is the way in which diverse acceptors bind and are activated for catalysis. The first, seminal, insight into the structural basis for pant natural product modification was provided in 20-01 through the structure determination of two plant methyltransferases, Chalcone *O*-methyltransferase (PDB ID: 1FPQ, 1FP1) and Isoflavone *O*-methyltransferase (PDB ID: 1FPX, 1FP2) from *Medicago sativa*[14].The Noel group were able to show that these enzymes possess SAM binding domains, as observed for diverse methyltransferases, across the tree of life, and show how a second "dimerization" domain contributed both to the dimer formation and the acceptor substrate binding site.

Extensive subsequent research undertaken on diverse SAM-dependent methyltransferases has highlighted five conserved motifs involved in SAM recognition[15] including in diverse plant

methyltransferases, for example: Norcoclaurine 6-*O*-methyltransferase (PDB ID: 5ICC, 5ICE)[16], and Caffeic acid *O*-methyltransferase (PDB ID: 3P9C, 3P9I, 3P9K)[17]. In contrast to SAM binding, detailed analysis of the methyl acceptor has been hindered by the large structural diversity of known substrates and to some extent the lack of high-resolution three-dimensional structures with the methyl acceptor bound. This is of considerable importance when one considers the large diversity in sequence-space and the rapidly increasing genomic resource.

Acceptor binding and specificity are of additional interest in plant *O*-methyltransferases as these are a large family involved in biosynthetic pathways that lead to an array of biologically and medically-important compounds including lignols, flavonoids, phenylpropanoids, terpenoids and alkaloids. At the mechanistic level, it is clear that *O*-methylation is generally performed on a hydroxyl group on a phenolic or catecholic ring, and it has been proposed to proceed via S_N2 type nucleophilic attack, with acceptor activation for catalysis *via* a general base mechanism featuring an Asp/His catalytic dyad. A consensus kinetic mechanism for this class of enzymes has yet to emerge, should it indeed exist, as it has been proposed to be both a random order and an ordered mechanism[18-20]. The development of next generation sequencing technologies and transcriptomic approaches have identified hundreds of predicted *O*-methyltransferases, with only a small number of these being biochemically characterised and an even smaller number having associated structural data[14, 16-17, 19-23]. This lack of structural insight into the structural "space" further hinders genome annotation and functional prediction.

In this context, here we report on the three-dimensional structure and kinetic analysis of the *Papaver somniferum O*-methyltransferase PSMT1; which catalyses the 9-*O*-methylation of scoulerine, **Figure 1**. PSMT1 displays less than 45% sequence identity to any OMT of known 3-D structure, rendering modelling of its acceptor binding challenging. Following a surface-entropy reduction approach, crystals, diffracting beyond 1.3 Å, were obtained. We show how PSMT1 forms, as other plant OMTs, a multi-domain dimer, with domain opening and closing (16 degrees to yield 99.6% closure) reflecting the ligand-binding. High resolution studies thus provide an intimate analysis of both SAH and scoulerine binding. The 3-D structure of PSMT1, when viewed in light of previous analyses now allows us to propose genomic analysis into the specificity and signature features of the wider plant methyltransferase superfamily, including

the proposal of "gatekeeper" residue positions that determine the nature of the acceptor substrate and its substituents.

Results

Expression, purification and catalytic activity of PSMT1

The *PSMT1* cDNA was expressed in *Escherichia coli* with an N-terminus hexa-histidine-glutathione *S*-transferase (GST) purification/solubility tag and featuring a human rhinovirus 3C protease (HRV-3CP) cleavage site in the linker. PSMT1 was purified utilising a multi-step purification procedure involving nickel affinity on-column cleavage by HRV-3C Protease, followed by desalting and anion exchange. The resulting protein was visualised by SDS-PAGE analysis with a single band of ca. 43 kDa visible. Size-exclusion multi-angle laser light scattering experiments (not shown) showed that PSMT1 is a homodimer in solution with a molecular weight of 83 kDa; consistent with other structurally-characterised plant *O*-methyltransferases[16-17, 19-23].

In order to define the kinetics of the PSMT1 catalysed reaction, steady-state parameters were determined by quantitation of tetrahydrocolumbamine production using UPLC-MS/MS. Non-linear regression analysis using the Michaelis-Menten equation produced a Vmax of 580 ± 45 nmol min^{-1} mg^{-1} of protein with a K_m for scoulerine of 0.35 ± 0.07 µM. While there appears to be substrate inhibition (apparent Ki 32µM) at high scoulerine concentrations, the error values are too large to demonstrate this unambiguously, **Figure 1b**. When scoulerine was held at a fixed concentration of 5 µM, a Vmax of 532 ± 13 nmol min^{-1} mg^{-1} with a Km for SAM of 8.8 ± 0.7 µM was derived**, Figure 1c**. The kinetic values obtained for a near identical PSMT1 on the basis of an assay that relied on incorporation into tetrahydrocolumbamine of a radioactive methyl group from [methyl-^{14}C]S-adenosylmethionine were Vmax of 2000 nmol min^{-1} mg^{-1} and a *K*m for SAM of 19µM and a Km for scoulerine of 29µM [13]. These are much higher than in the current study, particularly for the *K*m values. In the current study tetrahydrocolumbamine formation was measured directly by UPLC-MS/MS and it is possibly the different assay conditions that have caused this difference in kinetic values. Published K_m values for other plant *O*-methyltransferases have been reported to vary from the µM to mM range, although the weak binders in these studies are believed not to be the natural substrate

of these enzymes[17, 24-27]. Kinetics of surface and active-centre variants are described below in light of the three-dimensional structure of PSMT1.

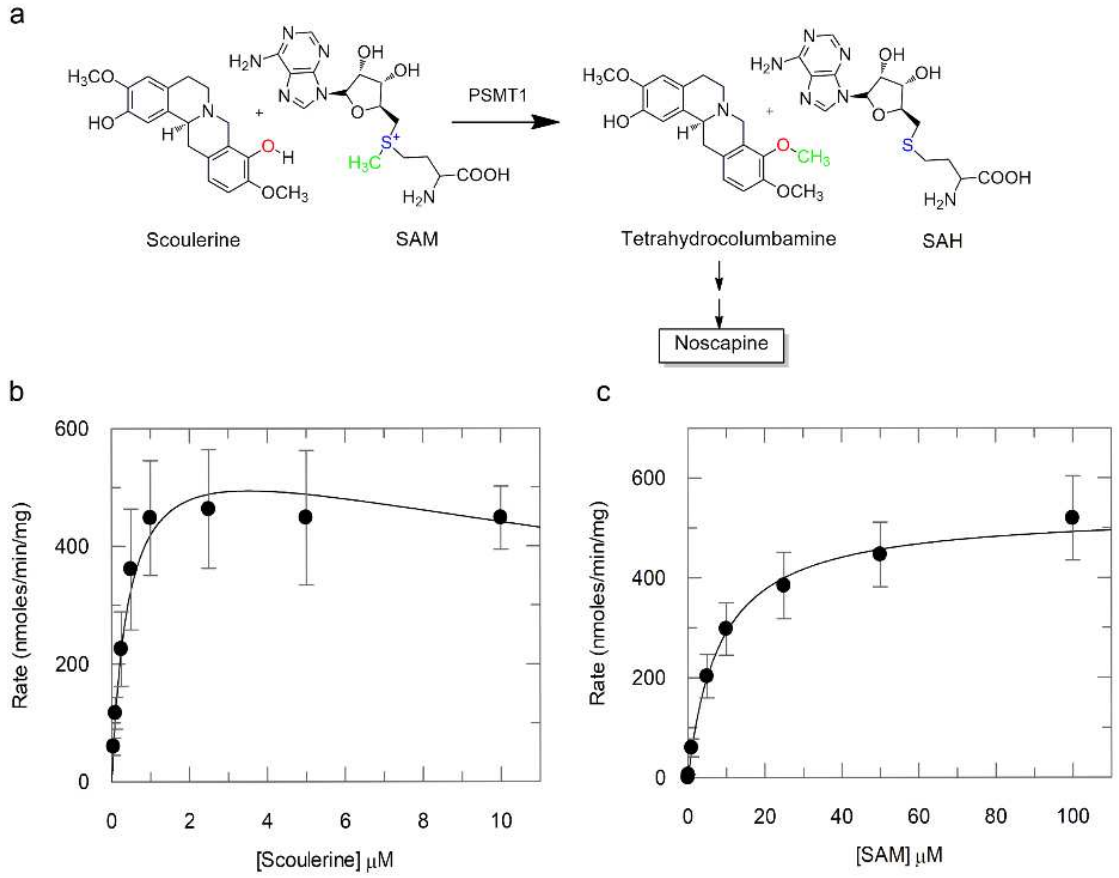

Figure 1. Reaction catalysed by PSMT1. (a) schematic of the PSMT1 catalysed reaction. Michaelis-Menten kinetics for O-methyltransferase (b) as a function of scoulerine concentration (c) as a function of SAM concentration. SAM: S-adenosylmethionine, SAH: S-adenosylhomocysteine. Error bars represent the standard deviation of triplicate repeats.

Overall structure of PSMT1

Having established through kinetics analyses that the heterologously-expressed PSMT1 is functional, we next sought to determine the three-dimensional structure of the protein, in order to shed light on specificity and catalytic mechanism. Screening of purified PSMT1 revealed an initial crystal form allowing access to an 'apo'-PSMT1 structure, crystallised in space-group $P3_221$, which diffracted to 3.05 Å, **Figure 2a, Table S1**. PSMT1 forms a symmetrical homodimer with an N-terminal dimerization domain and a SAM/SAH C-terminal Rossmann-like fold domain, separated by a putative α-helical substrate binding layer involved

with the dimerization. Co-crystallisation with SAM resulted in a, similarly weakly-diffracting, 3 Å resolution structure with SAH and/or SAM bound at varied occupancy in an extended conformation on the C-terminal Rossmann-like fold domain. Both of these structures are considered "open"; *i.e.*, not in a conformation viable for methylation of the substrate. Domain motion has been demonstrated for an array of plant OMTs including caffeic acid *O*-methyltransferase, isoflavone *O*-methyltransferase and norcoclaurine 6-*O*-methyltransferase[14, 16-17], where it is essential to bring the methyltransfer apparatus into close proximity to the acceptor hydroxyl. More discussion of ligand binding and conformational change will be given below, based upon much higher-resolution complexes.

Surface entropy reduction allows dissection of donor and acceptor binding and conformational change in PSMT1

Whilst medium-resolution analyses allowed a general description of topology, and to some extent donor binding, it was clear that detailed analysis of ligand-binding demanded a better-diffracting crystal form that was also amenable to acceptor substrate binding. After exhaustive, unsuccessful screening, access to a closed conformation of PSMT1 was achieved using a rational surface engineering approach, based upon surface entropy reduction (SER), viewed in light of our low-resolution analysis. SER was implemented utilising the UCLA MBI SER prediction (SERp) server[28]; primers in **Table S2**. Three clusters for mutagenesis were proposed and mapped onto the low resolution 'apo' structure. Two of the clusters, residues 114-115 and 128-129, were located on a flexible loop not defined in the electron density, suggesting they would make appropriate regions to reduce surface entropy and were therefore generated along with a double cluster mutant. The third cluster, residues 174-176, was located close to the dimer interface and was rejected due to the possibility of disturbing dimerization.

A SER variant, PSMT1-SER, with Lys114 and Lys115 converted to alanine, allowed the growth of crystals in a new crystal form, when co-crystallised with SAH and scoulerine, in space group $P2_1$, with diffraction extending to 1.49 Å, **Table S1.** Michaelis-Menten kinetic analysis of PSMT1-SER confirmed that the two introduced residue changes had negligible effect on

enzyme activity parameters (Vmax, 450 nmol min^{-1} mg^{-1} of protein and K_m, 0.2 µM, compared to Vmax, 580 ± 45 nmol min^{-1} mg^{-1} , K_m for scoulerine, 0.35 ± 0.07 µ for the wild type protein).

Subsequent optimisation of co-crystallisation experiments with the substrates SAM and scoulerine, resulted in a 1.29 Å resolution product-bound structure with SAH and tetrahydrocolumbamine found in the catalytic centre. Binding of the ligands resulted in a large conformational change in which the SAM/SAH binding domains had rotated inwards in relation to the dimerization domains generating the "closed-form" **Figure 2b**, causing the movement of the residues required to furnish competent active sites.

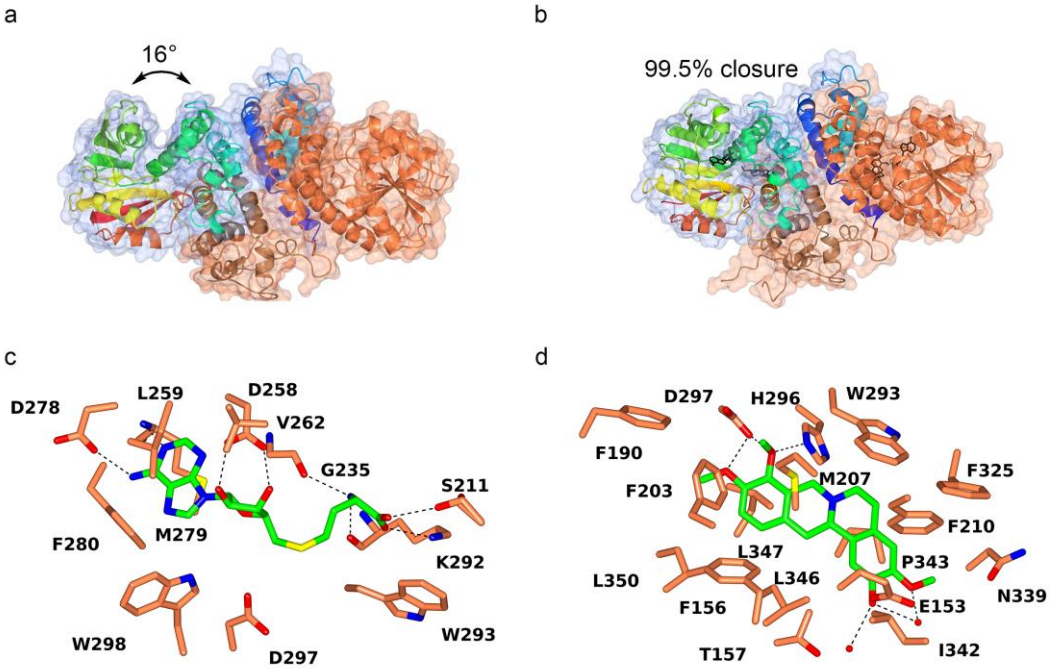

Figure 2. Three-dimensional structure and ligand binding of PsMT1. PSMT1 'open' (A) and 'closed' (B) structures represented as ribbon diagrams with transparent surfaces. In each case one monomer of the dimer is colour-ramped from N-terminus (blue) to C-terminus (red) whilst the other monomer is coloured by domain. SAH and tetrahydrocolumbamine bound in the active site, are shown as black cylinders. C and D show the interactions of SAH and tetrahydrocolumbamine, respectively in the SER variant structure co-crystallised with SAM and SCU (observed electron density is shown in Figure 3, below). Domain opening/closing analysis was performed with DYNDOM[29] from the CCP4 suite (domain movement available as a Web-Enhanced Object)

SAM/SAH binding

The structure of PSMT1 co-crystallised with SAM in the open conformation features hydrolysed SAM, and is modelled with 2 molecules of SAH (associated with chains A and B) and 1 molecule of SAM (with the methyl group modelled at half occupancy, and C, O and OXT atoms at 0.3, in chain D), and has insufficient density to allow ligand to be fitted in chain C. The nucleotide adenine group of the (hydrolysed) donor molecules interacts with the non-polar side chains of Leu259, Met279, Phe280 and Trp298 and features hydrogen bonds betweenN1 and N Met279, and N6 and OD1 Asp278 in all 3 ligands, and also N3 and OD1 Asp258 for SAM. Upon closure of the active site the interactions with SAM/SAH are maintained but with additional interactions made between Phe190, Phe203, Met207 and Ser211 from the alpha-helical layer, **Figure 2C**. The former three residues form a channel along with Trp293 and Asp297 into which the transferred methyl group extends allowing insight into catalytic mechanism, described, below. There is very little movement of the residues lining the donor binding site upon binding of SAM (or SAH) in the "open" protein conformation, apart from rotation of the side chain of Asp258 by approximately 90 degrees to allow both OD atoms to form a hydrogen bond with a ribose hydroxyl group (OD1 to O2', OD2 to O3').

Consistent with generic features of plant SAM-dependent O-methyltransferase, **Figure S2a**, the ribose ring is also clamped in by a non-polar glycine rich segment Gly235-Gly236-Gly237. This glycine rich segment also forms one wall to the carboxypropyl moiety of methionine of the donor with Trp293 on the opposing side. The amine group of the methionine occupying the donor site hydrogen bonds with the backbone oxygen of Lys292 for all 3 SAH/SAM molecules, and also the backbone oxygen of Gly235 for SAH in chain B. The side chain of Lys292 extends around to the carboxylic acid group of the methionine moiety making a hydrogen bond via its terminal (Nζ) amino group. The methyl group of the sulphonium ion (for SAM) is positioned by Trp293 and Asp297 which direct it into the open cleft between the SAM binding domain and that of the putative substrate binding pocket.

Acceptor substrate binding

One of the motivations for this work is that residues involved in the recognition of the methyl acceptor substrate have been far less well characterised than those involved in SAM binding.

PSMT1 SER (WT or variant) structures with SAH and scoulerine or tetrahydrocolumbamine bound in the active site show essentially no difference in how the compounds are bound. Scoulerine and tetrahydrocolumbamine sit in a deep hydrophobic pocket fully occluded within the enzyme braced by hydrogen bonds between the hydroxyl groups at opposing ends of the compounds with Asp297 and Glu153, **Figure 2d**. 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 seven from the dimerization domain, eleven, twelve and seventeen from the alpha-helical layer, and Thr39 from alpha helix one of the opposing chain, **Figure 2d.**

The A-ring of scoulerine is sandwiched, **Figure 2d**, between Phe210 by a CεH-π hydrogen bond and on the opposing face interacts with Ile342 and Pro343. Thr157 Cγ makes a hydrophobic interaction with C1 and a Van der Waals' 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-OH of scoulerine and to C3-O via a water molecule. The O3 of scoulerine hydrogen bonds to the backbone carbonyl of Thr39 of the opposing dimer chain via the same water molecule as Glu153. The 3-O-methyl group makes hydrophobic interactions with Thr39 from the opposing chain and Asn339 Cα. It also makes Van der Waals' interactions with the backbone carbonyl group of Asn339 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 Waals' 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. The D-ring is held in position by Leu347 CδH-π bonding and hydrophobic interactions with Leu350, along with Phe156, Phe203 and Met207 interacting on the opposing face. The 10-methoxy position of scoulerine lies in a hydrophobic sheath formed byPhe190, Phe203 and Leu350, along with 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. Further analysis of acceptor binding, in light of the diversity of plant methyltransferase will be discussed below.

Ternary complex unveils the reaction mechanism of PsMT1

Co-crystallisation with its substrates, resulted in enzymatic turnover and the products resulting in SAH and tetrahydrocolumbamine (**Figure 1**) observed at high resolution, **Figure 3a** and hence the catalytic reaction mechanism may be inferred. Consistent with similar enzymes [16-17, 21] His296 likely acts as a general catalytic base deprotonating the 9-hydroxyl position of scoulerine during nucleophilic attack at the sulphonium methyl group of SAM, resulting in the 9-methyoxy product tetrahydrocolumbamine and SAH**, Figure 3b**. Glu356 acts as a hydrogen bond acceptor for His296 Nδ1 promoting the required basicity and side-chain conformation.

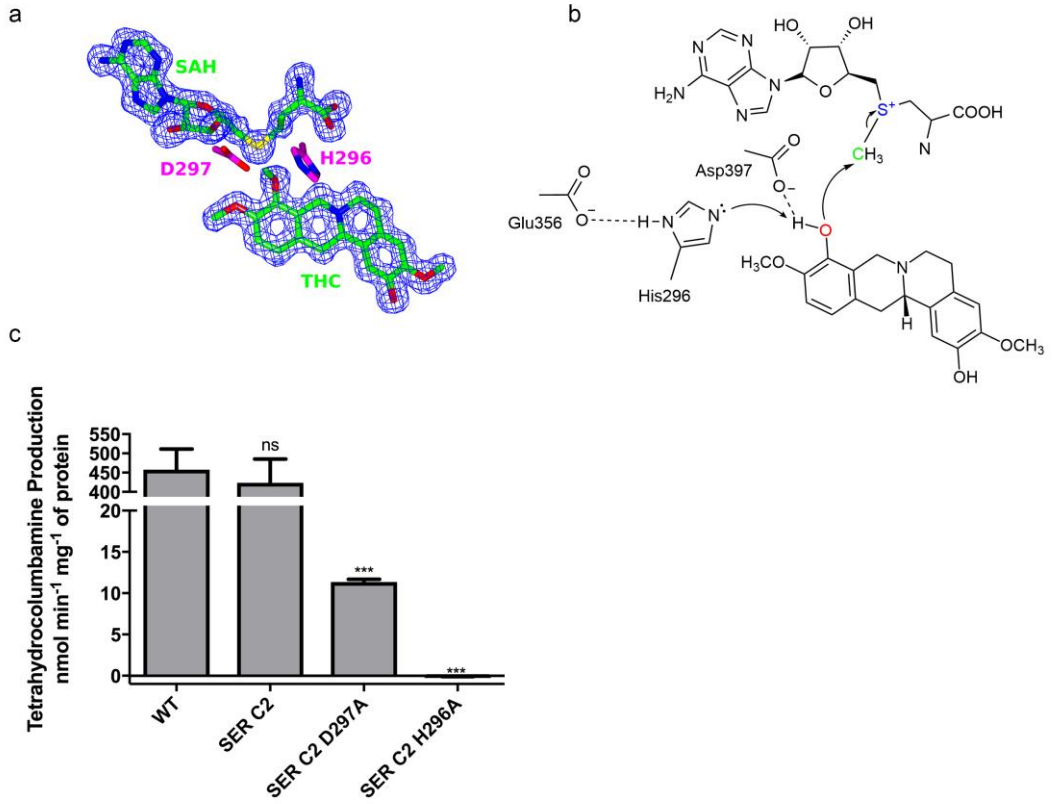

Figure 3. Ligand binding and proposed reaction mechanism of PSMT1. (a) three-dimensional structure of the SAH and tetrahydrocolumbamine binding. The electron density map is a maximum-likelihood / σ_A weighted 2Fo-Fc synthesis, contoured at 0.44 electrons / Å^3. (b) Schematic diagram of reaction mechanism (c) Bar graph showing the activity of WT PSMT1

alongside the surface entropy reduction and two active site mutants. Reactions were carried out under representative V_{max} conditions of 5 µM scoulerine and 100 µM SAM.

In order to probe the catalytic mechanism, the two residues proposed to be important for the methyl transfer event, His296 and Asp297 were mutated to alanine. Mutations were introduced into the PSMT1 SER variant and the enzymes assayed under the same reaction conditions as previously but with scoulerine and SAM at saturating concentrations, 5 µM and 100 µM, respectively. There was no significant difference between the rate of PSMT1 WT and PSMT1 SER, whereas PSMT1 SER Asp297Ala possesses approximately 2 % activity compared to PSMT1 SER. PSMT1 SER His296Ala was inactive (although the protein was produced at lower levels, likely indicating reduced stability), **Figure 3c**. We conclude that Asp297 is important but not essential for catalysis, whereas His296 is critical for activity, consistent with its likely role in the initial deprotonation of the 9-hydroxyl group on scoulerine. In order to probe substrate binding further, X-ray structures of PSMT1 SER Asp297Ala were generated with SAM and scoulerine, resulting in observation of SAH and scoulerine bound in the active site. Intriguingly, although the Asp297Ala variant has allowed removal of the methyl group from SAM, no evidence of its transfer to scoulerine is observed in-crystal (not shown), which may indicate that the donor has been hydrolysed as observed in the WT structure, or that there was SAH contamination in the commercial SAM preparations.

Three-dimensional structure reveals components of acceptor specificity in PSMT1 and beyond

O-methyltransferases are a diverse group of enzymes which have evolved to accept a wide range of substrates including alkaloids and flavonoids. The substrate specificity of these type of enzymes is difficult to deduce based on primary sequence alone, due to their varying evolutionary paths. We set out to investigate how the PSMT1 could shed light on the mechanism and specificity of plant natural product *O*-methyltransferases, which display such diverse acceptor specificities.

By harnessing a surface entropy approach our X-ray crystallographic structures have shown how the methyl donor SAM, as well as the acceptor/product scoulerine/tetrahydrocolumbamine are accommodated, which allows interrogation of the

sequence data in light of these structural facets. Consistent with other reports (*vide infra*) we can conclude that generation of OMTs with altered substrate specificity appears to occur by gene duplication and neofunctionalization, with subsequent divergence of the acceptor binding pocket; the residues involved in the recognition of the phenolic ring of the substrate to be *O*-methylated are highly conserved, **Figure S2b**. As proposed in classic work by the Noel group, Phe and Met side-chains above the phenolic ring and a Met side-chain below the ring appear to act as traps for the phenolic ring, with the conserved catalytic dyad of His and Asp providing the catalytic apparatus for methyl transfer[14, 16-17, 19-22, 26, 30]. In contrast, to this conserved methionine clamp, the residues responsible for the recognition of the variable chemical groups of the substrates are highly variable, but in light of the PSMT1 structure, and the increasing genomic and structural resource, we sought to analyse if any features were predictive of likely ligand shape and structure.

A phylogenetic tree, using the characterized plant OMT sequences was constructed (see Methods) revealing an early evolutionary split into two groups of OMTs (**Figure S3**). Analysis of the phylogenetic tree, through the telescope of molecular shape and connectivity and the known three-dimensional structures hints at some general recognition features. With a single exception, it was striking to us that one half of this evolutionary tree (sequences coloured black on **Figure S3, Figure S4**), has a small side chain (Gly, Ala, Ser, Thr, Cys, Val) at the position equivalent to Phe156 of PsMT1 – the residue that sits "opposite" the phenolic ring to be methylated. Notably, 16 of these 19 sequences act on substrates with bulky substituents opposite (to some extent "*para*") to the hydroxyl-group to be methylated strongly suggesting that this evolutionary group makes space for bulky ligands by simple steric means, reflected both in their evolutionary origin and fine sequence. A small evolutionary clade, consisting three characterized enzymes (red) bucks this trend by acting on acceptors with bulky groups para to the hydroxyl but also possessing large groups predicted to occupy the position equivalent to Phe156. These enzymes, that act on reticuline and different xanthohumols, **Figure S3, S4** have bulky groups opposite the hydroxyl (Phe, Met, Trp); unfortunately, no structural data exists for these enzymes, highlighting the need for further structural campaigns directly informed by these analyses.

The second major evolutionary group (coloured blue in **Figure S3, S4**) which includes PSMT1 primarily, but not exclusively have large amino acids in positions one would predict to lie

opposite the hydroxyl group to be methylated and to a large extent indeed work on substrates that have no steric bulk in such positions. How the exceptions function, at a structural level is as yet unknown, but inspection of the three-dimensional structures, where known, confirms that bulky groups may act as "gatekeepers" for substrate recognition; large hydrophobic groups act as gatekeeper residues, notably at positions equivalent to Phe156 and to some extent Leu350 of PSMT1, for acceptor binding. Where known structurally, large groups in these positions sterically hinder any group opposite to the target hydroxyl group, **Figure 4a**; a similar situation is observed in the *Medicago sativa* chalcone *O*-methyltransferase (PDB code 1FP1) with Phe138 and Thr332[14] playing similar roles, **Figure 4b**.

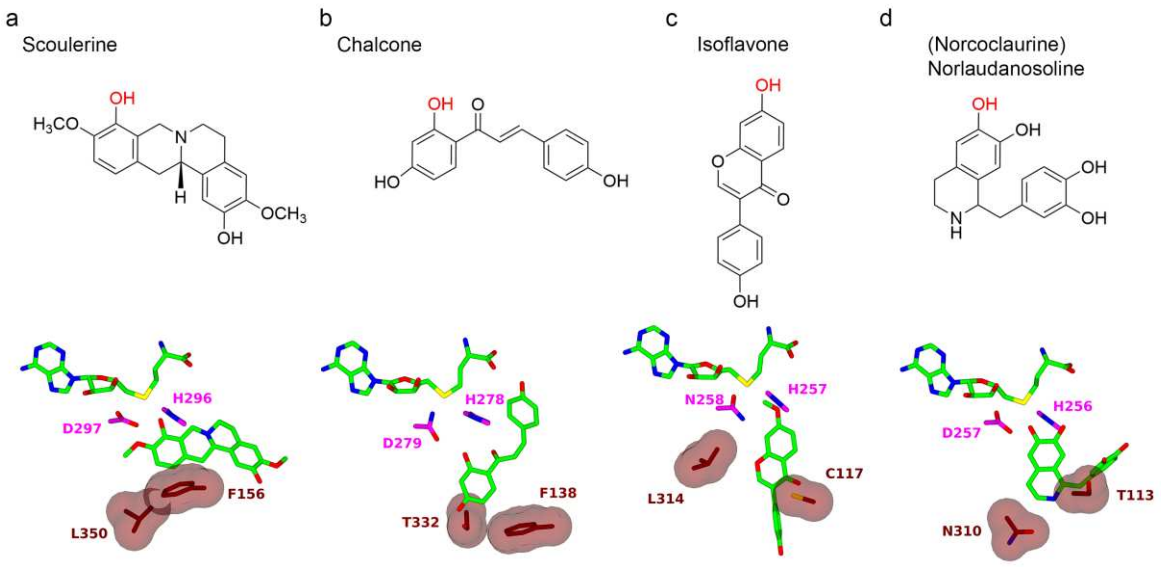

Figure 4. Residues contributing to ligand shape recognition in characterized plant O-methyltransferases. Phylogenetic analysis of characterized enzymes (**Figure S4**) and their sequence alignments (**Figure S4**). The catalytic His/Asp dyad are coloured magenta (as in **Figure 3**). Four examples are shown (a-d) with the ligand above and the three-dimensional structure of the ligand binding below (a) SAH-SCU *Papaver somniferum* O-methyltransferase 1, this work (b) ChOMT *Medicago sativa* Chalcone O-methyltransferase (PDB 1FP1; the ligand is in two position; the model with the correct *E* stereochemistry displays the best density and is shown) (c) IOMT *Medicago sativa* isoflavone O-methyltransferase (PDB 1FP2) (d) *Thalictrum flavum* norcoclaurine 6-O- methyltransferase (PDB 5ICE))

In contrast, accommodation of substrates which extend in the opposition position "*para*" to the reactive hydroxyl group is predominantly reflected in enzymes that display smaller size chains at the position equivalent to Phe156, such as Cys and Thr in *Medicago sativa* isoflavone 7-OMT (PDB code 1FP2) and *Thalictrum flavum* norcoclaurine 6-OMT (PDB code 5ICE)[14, 16], **Figure 4c,d.** This is also true in many other functionally characterised isoflavone OMTs and norcoclaurine OMTs (as reflected in the sequence alignment shown in **Figure S5**). Hence, sequence divergences open up the active site, to allow ligands of different shape and structure; buried in the primary sequence is the basis for plant natural product modification that three-dimensional structure is beginning to illuminate as increasing numbers of three-dimensional structures, coupled to enzyme characterization begin to emerge.

Plant O-methyltransferases are a fascinating group of enzymes, which display exquisite divergence in their acceptor sites to allow methyl group transfer to a wide variety of alkaloids, flavonoids, lignols, phenylpropanoids, terpenoids and other diverse natural products. Understanding how the, necessarily exquisite, acceptor specificity is achieved will be central to both enzyme and system-wide engineering campaigns to understand, modify and exploit plant secondary metabolism. As noscapine-derived and inspired compounds become increasingly important as therapeutic agents (for recent examples see Refs 31-32) and in light of breakthroughs in the biosynthesis of noscapine in heterologous hosts systems[8], we believe the dissection of PSMT1 reported here is an important step towards the goal of being able to predict substrate specificity of the growing number of methyltransferase genes in the public domain that will ultimately contribute to their exploitation in rational engineering of natural products.

Methods

Gene cloning and expression and protein purification

PSMT1 was recombinantly expressed as a His/GST-fusion protein (His-GST-r3CP-PSMT1), involving the use of the York Bioscience Technology Facility. His-GST-r3CP-PSMT1 was grown in 500 mL of Terrific-Broth (6 g tryptone, 12 g yeast extract, 2 mL glycerol, 17 mM KH_2PO_4, 72 mM K_2HPO_4) containing 100 µg mL^{-1} of kanamycin. Gene expression was induced by IPTG with a final concentration of 1 mM and cultivated at 16 °C for 18 hours. Cells were harvested by centrifugation at 5,000 rpm at 4 °C for 30 minutes and resuspended in binding buffer comprised of 100 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole, 10 mM β-mercaptoethanol, in a ratio of 1 g of cell pellet to 8 mL of binding

buffer. A protease inhibitor cocktail tablet (cOmplete™, Roche) and 100 uL DNaseI at 2 mg mL^{-1} in 50% glycerol were added before lysis by sonication. The lysate was clarified by centrifugation at 18,000 rpm for 45 minutes at 4 °C and the supernatant filtered using a syringe through a 0.22 µm filter unit, before loading onto a 5 mL HisTrap FF Crude column (GE Healthcare). After washing the column with 10 CVs of binding buffer, 5 mL 0.25 mg mL^{-1} HRV-3C protease were manually injected onto the column which was incubated overnight at 4 °C. The HRV-3C protease cleaved 6xHis-GST-r3CP-PSMT1 at the 3CP recognition site (r3CP) giving rise to the purification/solubility tag (6xHis-GST), which remained bound to the column, and PSMT1 with a 3 residue N-terminus overhang. The liberated PSMT1 was washed off the column using binding buffer, and the bound proteins were eluted with binding buffer containing 0.5 M imidazole. Fractions containing PSMT1 were pooled and buffer exchanged into buffer A (100 mM Tris pH 8.0, 10 mM β-mercaptoethanol) using a HiPrep 26/60 Desalting column (GE Healthcare), before loading onto a 4.7 mL HiScreen Q HP anion exchange column (GE Healthcare), washed with 10 % buffer B (100 mM Tris pH 8.0, 1M NaCl, 10 mM β-mercaptoethanol) and eluting with a gradient of 10-50 % buffer B. Fractions containing PSMT1 were pooled, buffer exchanged into 10 mM TRIS pH 8.0, 5 mM TCEP and concentrated to 15 mg mL^{-1} using a 30 KDa cut-off centrifugal concentrator. All PSMT1 variants and their mutants were produced and purified in the same manner as PSMT1 WT. Purification steps were monitored by SDS-PAGE using 12 % gels and protein concentrations were determined by measuring absorbance at 280 nm and calculating from their predicted extinction coefficients.

Site-directed mutagenesis experiments were carried out by PCR based whole plasmid amplification using oligonucleotides containing the target mutation, SI 8, followed by DpnI digestion of template DNA. Chemically competent *E. coli* XL10 gold cells were transformed with the products to allow plasmid amplification, and mutations were confirmed by DNA Sanger sequencing (GATC).

PSMT1 Enzymatic Assays

Standard enzyme activity assays were performed in 100 mM Gly-NaOH pH 9.0, 25 mM sodium ascorbate, 1 mM β-mercaptoethanol, 100 µM SAM and 5 µM scoulerine at 37 °C. Assays were initiated with the addition of 0.5 mL of 1 µg mL^{-1} recombinant PSMT1 protein. Protein concentration was determined using absorbance at 280 nm and corrected based on quantification using the Bio-Rad Protein Assay Kit II, which is based on the Bradford assay. 1 mL aliquots were taken over 2.5 minutes and quenched with an equal volume of methanol. O-methylation of scoulerine to tetrahydrocolumbamine was monitored by direct detection of the latter using UPLC-MS/MS. The UPLC method was carried out as described previously[9]; detection was carried out using a Thermo TSQ Endura triple quadrupole mass spectrometer in ESI positive mode using selected reaction monitoring (SRM). Samples were prepared by drying the aliquots using a SpeedVac centrifugal evaporator (Genevac) and resuspending them in 100 µL 10 % acetic acid, and they were analysed alongside authentic standards of known concentration. Scoulerine was detected at 1.85-2.55 min post injection by CID of the parental ion of 328.049 m/z, and tetrahydrocolumbamine at 2.55-3.25 min by CID of the parental ion of 342.080 m/z, and monitoring the product ions with 178 m/z. Data were processed using Thermo Xcalibur (Thermo Fisher Scientific) and GRAFIT (Erithacus Software, East Grinstead UK).

Crystallisation and Data Collection

All PSMT1 variants were crystallised using the sitting drop method at 20 °C with a 1:1 ratio of protein to crystallisation well solution. An initial PSMT1 WT 'apo' crystal was grown with 15 mg mL^{-1} protein

over a well comprised of 0.9 M sodium succinate, 0.1 M HEPES pH 7.0, 0.1 M sodium citrate, 1 % (w/v) polyethylene glycol (PEG) 2000. Co-crystallisation experiments with 12 mg mL^{-1} PSMT1 and 4.7 mM SAM produced crystalsover 0.1 M HEPES pH 7.0, 1 M sodium succinate, 1 % PEG 2000. PSMT1 SER variants were co-crystallised with 15 mg mL^{-1} protein, with either 1 mM SAM, 0.7 mM SAH, 2 mM scoulerine or 2 mM tetrahydrocolumbamine, over 0.2 M di-Ammonium citrate pH 4.4-5.0, 20-25 % PEG 3350. Crystals were harvested into nylon-fibre mounted CryoLoops™ (Hampton Research) and flash frozen in liquid nitrogen. All data were collected at the Diamond Light Source (Harwell) using a single crystal and auto-processed on the beamline with XIA2[33]. The first structure was solved by molecular replacement with the program Phaser[34] using PDB entry 1KYZ as model (residues 32-344), and improving the phases with Parrot[35], followed by model building with BUCCANEER. PSMT1 WT crystals, in the open conformation, were sufficiently isomorphous with the initial PSMT1 WT structure for it to be used as a starting model for refinement with REFMAC [36]. PSMT1 SER co-crystallised with SAH and scoulerine was solved using the SAM domain of PSMT1 WT chain A as the initial search model with MOLREP[37]. BUCCANEER[38] was used in seed chain growing mode to auto-build the remaining residues, followed by manual model building with COOT [39] and refinement with REFMAC. Ligand structures were built into the 2Fo-Fc and Fo-Fc electron density maps after the refinement of protein residues. The PSMT1 SER D297A structures were sufficiently isomorphous with the structure of PSMT1 SER co-crystallised with SAM and scoulerine to use the latter with the ligands and waters removed as a starting model for refinement with REFMAC. Manual model building of protein residues with COOT and refinement with REFMAC were carried out, followed by building the ligand molecules into the 2Fo-Fc electron density map.

Phylogenetic analysis of the characterised plant *O*-methyltransferases (OMTs) related to PSMT1

PSMT1 (AFB74611.1) protein sequence was used as query sequence in a *BLASTP* search in the curated Swissprot database via the NCBI webpage (http://www.ncbi.nlm.nih.gov/). Protein sequences showing an E value <= 1E -50 and a reported function were downloaded along with all homologous sequences from the opium poppy as well as a structurally determined (*S*)-Norcoclaurine 6-*O*-methyltransferase (5ICC) from *Thalictrum flavum*. These are: *Humulus lupulus* (Hlu) Xanthohumol 4-OMT (B0ZB56.1), Hlu Desmethylxanthohumol 6'-OMT (B0ZB55.1), *Papaver somniferum* (Pso) (*R*,*S*)-Reticuline 7-OMT (Q6WUC2.1), *Coptis japonica* (Cja) 3'-Hydroxy-*N*-methyl-(*S*)-coclaurine 4'-OMT (Q9LEL5.1), Pso 3'-Hydroxy-*N*-methyl-(*S*)-coclaurine 4'-OMT1 (Q7XB11.1), Pso 3'-Hydroxy-*N*-methyl-(*S*)-coclaurine 4'-OMT2 (Q7XB10.1), *Thalictrum flavum* (Tfl) (*S*)-Norcoclaurine 6-OMT (5ICC), Cja (R,S)-Norcoclaurine 6-OMT (Q9LEL6.1), Pso Norreticuline 7-OMT (C7SDN9.1), Pso (*R*,*S*)-Norcoclaurine 6-OMT (Q6WUC1.1), Pso PSMT3 (AFB74613.1), Cja Columbamine OMT (Q8H9A8.1), Pso PSMT2 (AFB74612.1), *Sorghum bicolor* (Sbi) 5-Pentadecatrienyl resorcinol OMT (A8QW53.1), *Pisum sativum* (Psa) Isoflavone 4'-OMT (O24305.1), *Glycyrrhiza echinate* (Gec) Isoflavone 7-OMT (Q84KK5.1), *Medicago sativa* (Msa) Isoflavone 7-OMT 8 (O24529.1), *Catharanthus roseus* (Cro) Myricetin OMT (Q8GSN1.1), Cro Tabersonine 16-OMT (B0EXJ8.1), Vitis vinifera (Vvi) Trans-resveratrol di-OMT (B6VJS4.2), *Ocimum basilicum* (Oba) Eugenol OMT (Q93WU2.1), Oba Chavicol OMT (Q93WU3.1), *Ruta graveolens* (Rgr) Anthranilate NMT (A9X7L0.1), *Pimpinella anisum* (Pan) Trans-anol OMT1 (B8RCD3.1), *Oryza sativa* (Osa) Naringenin 7-OMT (Q0IP69.2), Sbi Eugenol OMT (A8QW52.1), *Arabidopsis thaliana* (Ath) Indole glucosinolate OMT1 (Q9LPU5.1), *Mesembryanthemum crystallinum* (Mcr) Inositol 4-OMT (P45986.1), Msa Chalcone OMT (P93324.1), *Clarkia breweri* (Cbr) (Iso)eugenol OMT (O04385.2), Ath Flavone 3'-OMT (Q9FK25.1), *Chrysosplenium americanum* (Cam) Quercetin 3-OMT (P59049.1), Msa

Caffeic acid 3-OMT (P28002.1), Cja (*S*)-Scoulerine 9-OMT (Q39522.1), Pso PSMT1 (AFB74611.1), Pso predicted OMT (AKO60157.1).

Protein sequence alignments were made with ClustalX. Only conserved residue blocks in the alignment regions were used in the subsequent phylogenetic analyses. The best-scoring maximum likelihood tree of a thorough maximum likelihood analysis in conjunction with bootstrap analyses of 100 replicates was carried out with RAxML and the tree was rooted using mid-point rooting method. Groups with above 70% bootstrap value were considered as strongly supported (**Figure S3**).

Author Contributions

IG and GJD conceived and supervised the study. MC, WO, PS and WO performed experiments. All authors contributed to manuscript preparation.

Conflict of Interest.

The authors declare no conflict of interest

Accession Codes

6I5Q, 6I5Z, 6I6K, 6I6L, 6I6M, 6I6N

Supporting Information Available

Tables of X-ray data and structure refinement, details of oligonucleotide primers, the biosynthetic pathway to noscapine, conserved features of acceptor and donor substrate binding, Phylogenetic analysis of characterized plant O-Methyltransferases, key features of acceptor substrate binding and sequence alignment of characterised plant *O*-methyltransferases. A movie of domain closure is provided as a web-enhanced object. This information is available free of charge on the ACS Publications website.

Acknowledgements

This research was funded by Biotechnology and Biological Sciences Research Council (grant BB/M011151/1) for PhD support to MC. GJD is the Royal Society Ken Murray Research Professor. We thank Diamond Light Source for access to beamlines I02, I03, I04 and I04-1 (proposal numbers mx-7864 and mx-9948) that contributed to the results presented here. We thank Jared Cartwright of the York Biosciences Technology Facility for construction of pETFPP-3-PSMT1.

References

1. Aragón-Poce, F.; Martınez-Fernández, E.; Márquez-Espinós, C.; Pérez, A.; Mora, R.; Torres, L. M., History of Opium. *International Congress Series* **2002,** *1242*, 19–21.

2. Krikorian, A. D., Were the Opium Poppy and Opium Known in the Ancient near East? *J Hist Biol* **1975,** *8*, 95-114.

3. Konzett, H.; Rothlin, E., Zur Wirkung Von Narkotin Auf Den Hustenreflex Und Auf Die Bronchialmuskulatur. *Experientia* **1954,** *10*, 472-473.

4. DeBono, A.; Capuano, B.; Scammells, P. J., Progress toward the Development of Noscapine and Derivatives as Anticancer Agents. *J. Med. Chem.* **2015,** *58*, 5699-5727.

5. DeBono, A. J.; Xie, J. H.; Ventura, S.; Pouton, C. W.; Capuano, B.; Scammells, P. J., Synthesis and Biological Evaluation of N-Substituted Noscapine Analogues. *ChemMedChem* **2012,** *7*, 2122-2133.

6. Ye, K.; Ke, Y.; Keshava, N.; Shanks, J.; Kapp, J. A.; Tekmal, R. R.; Petros, J.; Joshi, H. C., Opium Alkaloid Noscapine Is an Antitumor Agent That Arrests Metaphase and Induces Apoptosis in Dividing Cells. *Proc. Natl. Acad. Sci. USA* **1998,** *95*, 1601-1606.

7. 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* **2015,** *35*, 1072-1096.

8. Li, Y.; Li, S.; Thodey, K.; Trenchard, I.; Cravens, A.; Smolke, C. D., Complete Biosynthesis of Noscapine and Halogenated Alkaloids in Yeast. *Proc Natl Acad Sci U S A* **2018,** *115*, E3922-E3931.

9. Winzer, T.; Gazda, V.; He, Z.; Kaminski, F.; Kern, M.; Larson, T. R.; Li, Y.; Meade, F.; Teodor, R.; Vaistij, F. E.; Walker, C.; Bowser, T. A.; Graham, I. A., A Papaver Somniferum 10-Gene Cluster for Synthesis of the Anticancer Alkaloid Noscapine. *Science* **2012,** *336*, 1704-1708.

10. Li, Y.; Smolke, C. D., Engineering Biosynthesis of the Anticancer Alkaloid Noscapine in Yeast. *Nature Commun* **2016,** *7*, 12137.

11. Dang, T.-T. T.; Chen, X.; Facchini, P. J., Acetylation Serves as a Protective Group in Noscapine Biosynthesis in Opium Poppy. *Nature Chem Biol* **2015,** *11*, 104-106.

12. Dang, T.-T. T.; Facchini, P. J., Cyp82y1 Is N-Methylcanadine 1-Hydroxylase, a Key Noscapine Biosynthetic Enzyme in Opium Poppy. *J Biol Chem* **2014,** *289*, 2013-2026.

13. Dang, T. T.; Facchini, P. J., Characterization of Three O-Methyltransferases Involved in Noscapine Biosynthesis in Opium Poppy. *Plant Physiol* **2012,** *159*, 618-31.

14. 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. *Nature Struct Mol Biol* **2001,** *8*, 271-279.

15. Kozbial, P. Z.; Mushegian, A. R., Natural History of S-Adenosylmethionine-Binding Proteins. *BMC Structural Biology* **2005,** *5*, 19.

16. 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.* **2016,** *87*, 641-653.

17. Louie, G. V.; Bowman, M. E.; Tu, Y.; Mouradov, A.; Spangenberg, G.; Noel, J. P., Structure-Function Analyses of a Caffeic Acid O-Methyltransferase from Perennial Ryegrass Reveal the Molecular Basis for Substrate Preference. *Plant Cell* **2010,** *22*, 4114-4127.

18. Morishige, T.; Tsujita, T.; Yamada, Y.; Sato, F., Molecular Characterization of Thes-Adenosyl-L-Methionine:3'-Hydroxy-N-Methylcoclaurine 4'-O-Methyltransferase Involved in Isoquinoline Alkaloid Biosynthesis in Coptis Japonica. *J Biol Chem* **2000,** *275*, 23398-23405.

19. Wolters, S.; Neeb, M.; Berim, A.; Schulze Wischeler, J.; Petersen, M.; Heine, A., Structural Analysis of Coniferyl Alcohol 9-O-Methyltransferase from Linum Nodiflorum Reveals a Novel Active-Site Environment. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2013,** *69*, 888-900.

20. Green, A. R.; Lewis, K. M.; Barr, J. T.; Jones, J. P.; Lu, F.; Ralph, J.; Vermerris, W.; Sattler, S. E.; Kang, C., 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* **2014,** *165*, 1440-1456.

21. 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* **2002,** *14*, 1265-1277.

22. Liu, C.-J.; Deavours, B. E.; Richard, S. B.; Ferrer, J.-L.; Blount, J. W.; Huhman, D.; Dixon, R. A.; Noel, J. P., Structural Basis for Dual Functionality of Isoflavonoid O-Methyltransferases in the Evolution of Plant Defense Responses. *Plant Cell* **2006,** *18*, 3656-3669.

23. Zhang, K.; Bhuiya, M.-W.; Pazo, J. R.; Miao, Y.; Kim, H.; Ralph, J.; Liu, C.-J., An Engineered Monolignol 4-O-Methyltransferase Depresses Lignin Biosynthesis and Confers Novel Metabolic Capability in Arabidopsis[C][W][Oa]. *Plant Cell* **2012,** *24*, 3135-3152.

24. Dhar, K.; Rosazza, J. P. N., Purification and Characterization of Streptomyces Griseus Catechol O-Methyltransferase. *Applied and Environmental Microbiology* **2000,** *66*, 4877-4882.

25. Petronikolou, N.; Nair, Satish K., Biochemical Studies of Mycobacterial Fatty Acid Methyltransferase: A Catalyst for the Enzymatic Production of Biodiesel. *Chem. Biol.* **2015,** *22*, 1480-1490.

26. Walker, A. M.; Sattler, S. A.; Regner, M.; Jones, J. P.; Ralph, J.; Vermerris, W.; Sattler, S. E.; Kang, C., The Structure and Catalytic Mechanism of Sorghum Bicolor Caffeoyl-Coa O-Methyltransferase. *Plant Physiol* **2016,** *172*, 78-92.

27. Zhou, M.; Long, T.; Fang, Z.-P.; Zhou, X.-L.; Liu, R.-J.; Wang, E.-D., Identification of Determinants for Trna Substrate Recognition by Escherichia Coli C/U34 2'-O-Methyltransferase. *RNA Biology* **2015,** *12*, 900-911.

28. Goldschmidt, L.; Cooper, D. R.; Derewenda, Z. S.; Eisenberg, D., Toward Rational Protein Crystallization: A Web Server for the Design of Crystallizable Protein Variants. *Prot Sci* **2007,** *16*, 1569.

29. Hayward, S.; Berendsen, H. J., Systematic Analysis of Domain Motions in Proteins from Conformational Change: New Results on Citrate Synthase and T4 Lysozyme. *Proteins* **1998,** *30*, 144-54.

30. Ma, Q.-H.; Xu, Y., Characterization of a Caffeic Acid 3-O-Methyltransferase from Wheat and Its Function in Lignin Biosynthesis. *Biochimie* **2008,** *90*, 515-524.

31. Meher, R. K.; Naik, M. R.; Bastia, B.; Naik, P. K., Comparative Evaluation of Anti-Angiogenic Effects of Noscapine Derivatives. *Bioinformation* **2018,** *14*, 236-240.

32. Devine, S.; Yong, C.; Amenuvegbe, D.; Aurelio, L.; Muthiah, D.; Pouton, C.; Callaghan, R.; Capuano, B.; Scammells, P., Synthesis and Pharmacological Evaluation of Noscapine-Inspired 5-Substituted Tetrahydroisoquinolines as Cytotoxic Agents. *J Med Chem* **2018,** *61*, 8444-8456.

33. Winter, G.; Lobley, C. M. C.; Prince, S. M., Decision Making in Xia2. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2013,** *69*, 1260-1273.

34. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser Crystallographic Software. *J Appl Cryst* **2007,** *40*, 658-674.

35. Cowtan, K., Recent Developments in Classical Density Modification. *Acta Crystallogr D* **2010,** *66*, 470-478.

36. Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A., Refmac5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2011,** *67*, 355-367.

37. Vagin, A.; Teplyakov, A., Molecular Replacement with Molrep. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2010,** *66*, 22-25.

38. Cowtan, K., The Buccaneer Software for Automated Model Building. 1. Tracing Protein Chains. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2006,** *62*, 1002-1011.

39. Emsley, P.; Cowtan, K., Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2004,** *60*, 2126-2132.

Supporting Information

Structure of *Papaver somniferum* O-Methyltransferase 1 Reveals Initiation of Noscapine Biosynthesis with Implications for Plant Natural Product Methylation

Marc P. Cabry[1,2], Wendy A. Offen[2], Philip Saleh[2], Yi Li[1], Thilo Winzer[1], Ian A. Graham[1,*], Gideon J. Davies[2,*]

[1]Centre for Novel Agricultural Products, Department of Biology, University of York, York, YO10 5DD, United Kingdom

[2] York Structural Biology Laboratory, Department of Chemistry, University of York, York, YO10 5DD, United Kingdom

*Correspondence to gideon.davies@york.ac.uk or ian.graham@york.ac.uk

Table S1. X-ray data and structure refinement statistics

Protein	WT	WT	SER	SER	SER D297A	SER D297A
Ligand	'apo'	SAM	SCU and SAH	SCU and SAM	SCU and SAH	SCU and SAM
Space group	P 3_2 2 1	P 3_2 2 1	P 2_1	P 2_1	P 2_1	P 2_1
Cell dimensions a, b, c (Å)	112.0, 112.0, 304.8	111.1, 111.1, 302.4	68.0, 75.9, 76.2	68.2, 75.9, 76.6	68.9, 76.6, 77.2	68.2, 75.9, 76.6
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0 120.0	90.0, 101.7, 90.0	90.0, 101.6, 90.0	90.0, 101.6, 90.0	90.0, 101.5, 90.0
Resolution (Å)	97.00 (3.05)	101.00(3.00)	74.72(1.49)	75.11(1.29)	75.71(1.20)	53.31(1.50)
Rmerge	0.151(1.607)	0.099(1.873)	0.075(0.837)	0.052(0.813)	0.069(0.694)	0.065(0.128)
R_{pim}	0.052(0.561)	0.034(0.652)	0.046(0.541)	0.037(0.604)	0.046(0.580)	0.054(0.104)
I/σI	10.7 (1.6)	16.3 (1.3)	9.3 (1.4)	**9.5 (1.7)**	7.2 (1.2)	12.2 (7.8)
Completeness (%)	100.0(100.0)	100.0(100.0)	99.7(99.1)	95.5(92.3)	97.3(88.7)	97.6(96.0)
CC(1/2)	**0.997(0.592)**	**0.999(0.530)**	**0.997(0.569)**	**0.997(0.511)**	**0.995(0.532)**	**0.993(0.982)**
Redundancy	9.9 (10.0)	10.0 (10.1)	4.2 (4.1)	3.6 (3.6)	3.1 (2.0)	**4.3 (4.4)**
Refinement						
No. reflections	41153	42222	117152	183080	225776	113281
R_{work} / R_{free}	0.21/0.26	0.20/0.27	0.13/0.18	0.12/0.16	0.12/0.13	0.10/0.14
No. atoms						
Protein	10068	10150	**5463**	**5556**	**5740**	5566
Ligand/ion	0	108	**100**	102	**100**	100
Water	0	0	**573**	597	894	768
B-factors (Å^2)						
Protein	**65.4**	**86.9**	19.0	18.0	15.3	14.5
Ligand/ion	n/a	71.2	**12.6**	12.5	9.8	9.1
Water	n/a	n/a	29.3	**30.4**	31.2	30.5
R.m.s. deviations						
Bond lengths (Å)	0.007	0.009	0.015	0.016	0.014	0.016
Bond angles (°)	1.597	1.716	1.852	1.962	1.942	1.904
Ramachandran plot residues						
In favoured regions **(%)**	91.0	90.2	97.2	97.7	97.7	98.3
In allowed regions **(%)**	7.2	6.7	2.8	2.3	2.3	1.7
In "disallowed" regions **(%)**	1.8	3.1	0	0	0	0
PDB codes	**6I5Q**	**6I5Z**	**6I6K**	**6I6L**	**6I6M**	**6I6N**

Table S2. Oligonucleotide primers used for surface entropy reduction mutants and sequencing

SER Mutations	
Lys114Ala	A340G A341C
Forward	5'-gttctattctttctgtttctactaca**gc**aaaatcaatcaacagaggaggagatg-3'
Reverse	5'-catctcctcctctgttgattgattt**gc**tgtagtagaaacagaaagaatagaac-3'
Lys115Ala	A343G A344C
Forward	5'-ttctttctgtttctactacagca**gc**atcaatcaacagaggaggagatg-3'
Reverse	5'-catctcctcctctgttgattgat**gc**tgctgtagtagaaacagaaagaa-3'
Active Site Mutants	
Asp287Ala	A890C, T891C
Forward	5'-aaaatgggtactgcacg**ct**tggggtgatgaacgat-3'
Reverse	5'-atcgttcatcacccca**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'
Sequencing Primers	
SeqPrimer 1	5’-TCGGAACTGACGCAAAAGT-3’
SeqPrimer 2	5'-GTGAGCGGATAACAATTCC-3'

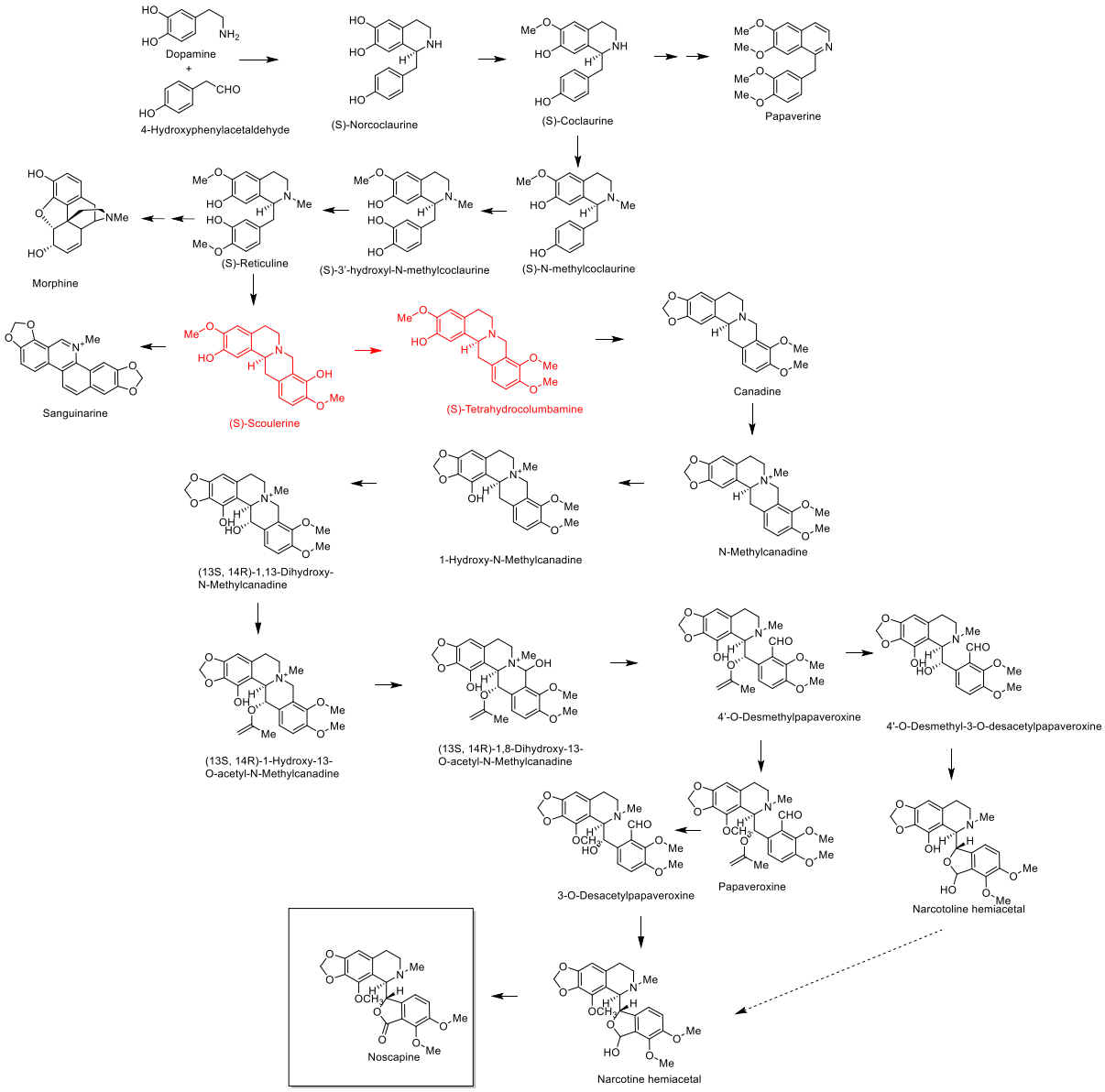

Figure S1. Current proposed biosynthetic pathway to noscapine [1-3]. Noscapine is shown in a shaded box and the PSMT1 catalysed reaction in red.

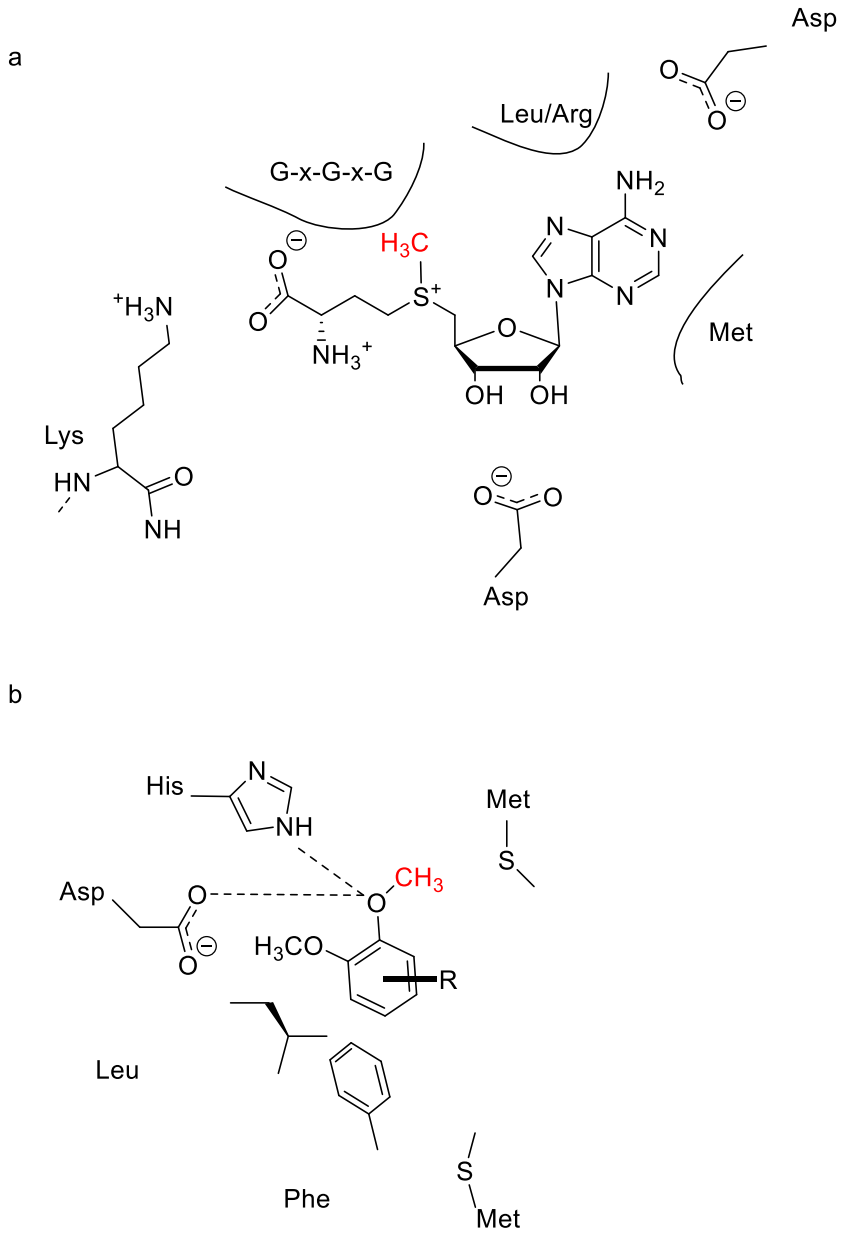

Figure S2. Conserved ligand binding features.

A. A summary of conserved features involved in SAM binding based on chalcone[4] isoflavone-[4] caffeic acid- [5] and norcoclaurine 6-[6] O-methyltransferases. B. The conserved Phe /methionine "motif" [4] involved in recognition of the acceptor phenolic group

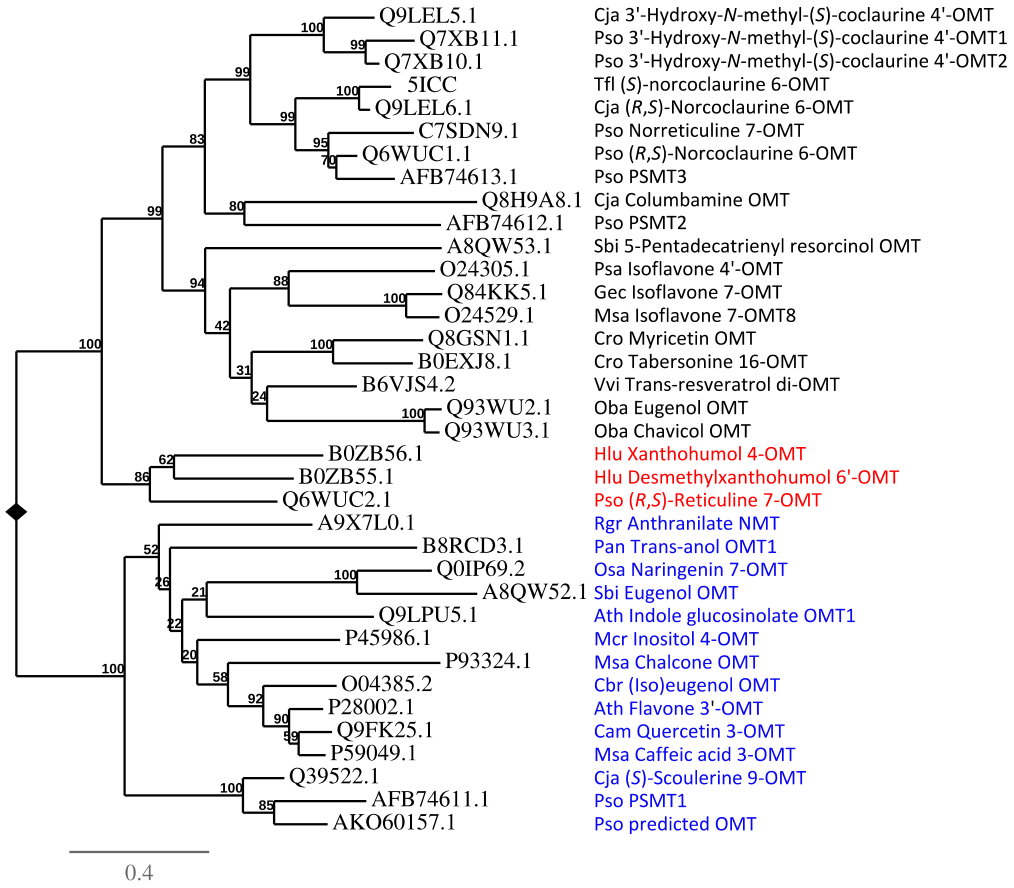

Figure S3. Phylogenetic analysis of characterized plant O-Methyltransferases.

Enzyme	Substrate extend "para" to the reactive hydroxyl group	Large/Small Gatekeeper Residues (1)
Cja 3'-Hydroxy-*N*-methyl-(*S*)-coclaurine 4'-OMT	YES	SMALL
Pso 3'-Hydroxy-*N*-methyl-(*S*)-coclaurine 4'-OMT1	YES	SMALL
Pso 3'-Hydroxy-*N*-methyl-(*S*)-coclaurine 4'- OMT2	YES	SMALL
Tfl (*S*)-Norcoclaurine 6-OMT	YES	SMALL
Cja (*R*,*S*)-Norcoclaurine 6-OMT	YES	SMALL
Pso Norreticuline 7-OMT	YES	SMALL
Pso (*R*,*S*)-Norcoclaurine 6-OMT	YES	SMALL
Pso PSMT3	NO	SMALL
Cja Columbamine OMT	YES	LARGE
Pso PSMT2	NO	SMALL
Sbi 5-Pentadecatrienyl resorcinol OMT	NO	SMALL
Psa Isoflavone 4'-OMT	YES	SMALL
Gec Isoflavone 7-OMT	YES	SMALL
Msa Isoflavone-7-OMT8	YES	SMALL
Cro Myricetin OMT	YES	SMALL
Cro Tabersonine 16-OMT	YES	SMALL
Vvi Trans-resveratrol di-OMT	YES	SMALL
Oba Eugenol OMT	YES	SMALL
Oba Chavicol OMT	YES	SMALL
Hlu Xanthohumol 4-OMT	YES	LARGE
Hlu Desmethylxanthohumol 6'-OMT	YES	LARGE
Pso (*R*,S)-Reticuline 7-OMT	YES	LARGE
Rgr Anthranilate NMT	NO	LARGE
Pan Trans-anol OMT1	YES	LARGE
Osa Naringenin 7-OMT	YES	LARGE
Sbi Eugenol OMT	YES	SMALL
Ath Indole glucosinolate OMT 1	NO	SMALL
Mcr Inositol 4-OMT	NO	SMALL
Msa Chalcone OMT	NO	LARGE
Cbr (Iso)eugenol OMT	YES	SMALL
Ath Flavone 3'-OMT	NO	LARGE
Cam Quercetin 3-OMT	YES	LARGE
Msa Caffeic acid 3-OMT	NO	LARGE
Cja (*S*)-Scoulerine 9-OMT	NO	LARGE
Pso PSMT1	NO	LARGE
Pso predicted OMT	UNKNOWN	LARGE

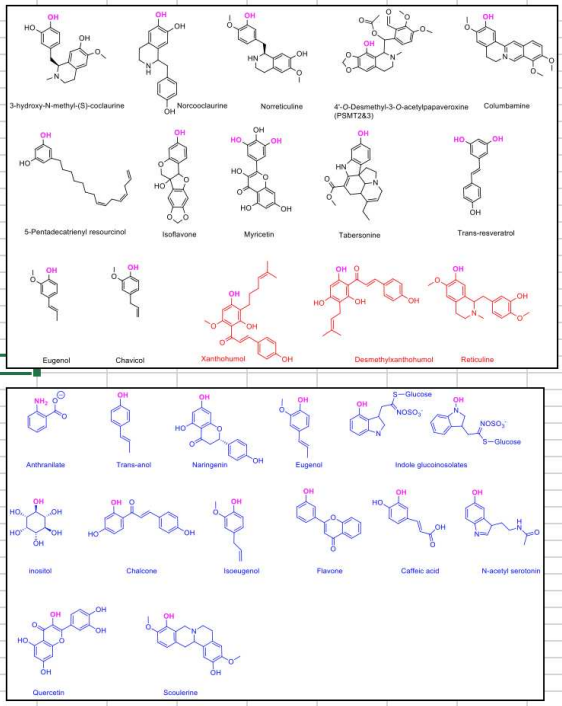

Figure S4. Proposed key features related to substrate binding of functionally characterised plant OMTs and relevant substrate chemical structures. Enzyme order follows the phylogenetic tree arrangement in Fig S3. The four structures shown in Figure 4 are shaded in yellow rows.

This Excel spreadsheet has been uploaded as a separate file.

```
Q9LEL5.1        115ILGMTQKDFMTPWHS....-LSSTRLILDIDMLVNTG-GKERTKEVWEKIVKSAGFSG333
Q7XB11.1        118ILGLSQKDFLFVWNF....-LTKARLILDIDMLVNTG-GRERTAEDWENLLKRAGFRS337
Q7XB10.1        121ILGMTQKDFMVSWHF....-LTKTRLILDIDMLVNTG-GRERTADDWENLLKRAGFRS340
5ICC            111ILTITDKDFMAPWHY....-YTKMRLTLDLDMMLNTG-GKERTEEEWKKLIHDAGYKG332
Q9LEL6.1        110ILAITDKDFMAPWHY....-YTKMRLTLDLDMMLNTG-GKERTEEEWKKLIHDAGYKG329
C7SDN9.1        120VLGIIDEDMFAPWHI....-YSKLRLTSDIDMMVNNG-GKERTEKEWEKLFDAAGFAS338
Q6WUC1.1        110ILGINDKDFLAPWHH....-YAKIRLTLDLDMMLNTG-GKERTKEEWKTLFDAAGFAS328
AFB74613.1      109ILSVTDKDFTAPWNH....-YSKSRLAMDLAMMLHTG-GKERTEEDWKKLIDAAGFAS321
Q8H9A8.1        116VMLQTHPEEFSVWSH....-FTSARLSMGMDMMLMSG--KERTKKEWEDLLRKANFTS334
AFB74612.1      118VLAISCEMMVVVWHE....-LTQAKLSLDLTVMNHGG-GRERTKEDWRNLIEMSGFSR339
A8QW53.1        127LNFVLSPFRDSPLSM....KLLETQVIYDLHLMKIG--GVERDEQEWKKIFLEAGFKD357
O24305.1        121VKGALHPSSLDMWGV....GLTELQLEYDVVMLTMFL-GKERTKKEWEKLIYDAGFSR343
Q84KK5.1        120VEGVLDPTLSGSYHQ....EITGTKLLMDVNMACLN--GKERSEEEWKKLFIEAGFRD340
O24529.1        115VEGVLDPTLSGSYHE....QVTQIKLLMDVNMACLN--GKERNEEEWKKLFIEAGFQH335
Q8GSN1.1        108VLGVNQIAELKAWNA....EAVKAQISSDIDMMVFFT-AKERTEEEWATLFREAGFSG331
B0EXJ8.1        115VLTMADPVQLKAWES....HLVKTQTSMDMAMLVNFA-AKERCEKEWAFLFKEAGFSD337
B6VJS4.2        118VLAMLDPILTKPWHY....KSTETQLFFDMTMMIFAP-GRERDENEWEKLFLDAGFSH340
Q93WU2.1        120VQVVLDPTFTNPWHH....EVLEDQLHFDMAMMCYFN-AKERTMSEWEKLIYDAGFKS240
Q93WU3.1        119VQVVLDPTFTNPWHY....EVLEDQLHFDMAMMSYFN-AKERTMNEWEKLISAAGFTS330

B0ZB56.1        122VAFMTHPYLSAPWSC....-FDDAAVMLDIALMA-LTRGKERTEKEWKRVLEEGGFPR342
B0ZB55.1        114VLMQTHPLSMAVWHF....-FDETRMVYDLLIPX-FSGGKERTELEWKRLLNEAGFTS334
Q6WUC2.1        116VLWETNPILLKPWQY....-FDKMGLIFDVLMMAHTTAGKERTEAEWKILLNNAGFPR337

A9X7L0.1        129MALPLDKVFMESWMG....SSARETSLLDVLLMTRDGGGRERTQKEFTELAIGAGFKG347
B8RCD3.1        120LKLIHHKQMQNSWEK....IIAKNISEMDIRMLLYTPGGKERTVNEFLMLGKQAGFPS339
Q0IP69.2        134GLLNLDKVFMENWYY....AAQEAFRLDV-MMLNRLAGGKERTQQEFTDLAVDAGFSG357
A8QW52.1        139GFWMTSTTNMETWHN....ASQLAFDFDLGMMLFFGASGKERTEKELLELAREAGFSG358
Q9LPU5.1        138VIVNFDSVFLNTWAQ....INANIAFDMDMLMFTQCSGGKERSRAEFEALAAASGFTH356
P45986.1        129LVLHHDKVMMESWFH....LESHMVFSLDCHTLVHNQGGKERSKEDFEALASKTGFST347
P93324.1        136TTFLCYPALLQVWMN....EESKLVSTLDNLMFIT-VGGRERTEKQYEKLSKLSGFSK354
O04385.2        131LLIATDKVLLEPWFY....IATKVVIHTDALMLAYNPGGKERTEKEFQALAMASGFRG349
Q9FK25.1        126CLMNQDKVLMESWYH....LSTKQVVHVDCIMLAHNPGGKERTEKEFEALAKASGFKG344
P59049.1        104CLMNQDKVLMESWYH....LATKGVVHIDVITVAHNPGGKERTEKEFEALAKAAGFQG322
P28002.1        128NLMNQDKVLMESWYH....LATKGVVHIDVIMLAHNPGGKERTQKEFEDLAKGAGFQG346
Q39522.1        142LLFTSDKAVVESFYN....AESFNALTPDLLMMALNPGGKERTTIEFDGLAKAAGFAE361
AFB74611.1      154LLFTSDKVVVDSFFK....AESFNALIPDLLLMALNPGGKERTISEYDDLGKAAGFIK373
AKO60157.1      147LMFVADKIVVESFYN....PESYNALTPDLLMMALNPGGKERTLLEFYDLANAAGFAK367
```

Figure S5. Sequence alignment of characterised plant *O*-methyltransferases (OMTs)

Potential "gatekeeper" residues, equivalent to Phe156 and Leu350, of PSMT1 (AFB74611.1) are coloured in red and blue, respectively. Enzymes shown: *Humulus lupulus* (Hlu) Xanthohumol 4-OMT (B0ZB56.1), Hlu Desmethylxanthohumol 6'-OMT (B0ZB55.1), *Papaver somniferum* (Pso) (*R*,*S*)-Reticuline 7-OMT (Q6WUC2.1), *Coptis japonica* (Cja) 3'-Hydroxy-*N*-methyl-(*S*)-coclaurine 4'-OMT (Q9LEL5.1), Pso 3'-Hydroxy-*N*-methyl-(*S*)-coclaurine 4'-OMT1 (Q7XB11.1), Pso 3'-Hydroxy-*N*-methyl-(*S*)-coclaurine 4'-OMT2 (Q7XB10.1), *Thalictrum flavum* (Tfl) (*S*)-Norcoclaurine 6-OMT (5ICC), Cja (R,S)-

Norcoclaurine 6-OMT (Q9LEL6.1), Pso Norreticuline 7-OMT (C7SDN9.1), Pso (*R*,*S*)-Norcoclaurine 6-OMT (Q6WUC1.1), Pso PSMT3 (AFB74613.1), Cja Columbamine OMT (Q8H9A8.1), Pso PSMT2 (AFB74612.1), *Sorghum bicolor* (Sbi) 5-Pentadecatrienyl resorcinol OMT (A8QW53.1), *Pisum sativum* (Psa) Isoflavone 4'-OMT (O24305.1), *Glycyrrhiza echinate* (Gec) Isoflavone 7-OMT (Q84KK5.1), *Medicago sativa* (Msa) Isoflavone 7-OMT 8 (O24529.1), *Catharanthus roseus* (Cro) Myricetin OMT (Q8GSN1.1), Cro Tabersonine 16-OMT (B0EXJ8.1), Vitis vinifera (Vvi) Trans-resveratrol di-OMT (B6VJS4.2), *Ocimum basilicum* (Oba) Eugenol OMT (Q93WU2.1), Oba Chavicol OMT (Q93WU3.1), *Ruta graveolens* (Rgr) Anthranilate NMT (A9X7L0.1), *Pimpinella anisum* (Pan) Trans-anol OMT1 (B8RCD3.1), *Oryza sativa* (Osa) Naringenin 7-OMT (Q0IP69.2), Sbi Eugenol OMT (A8QW52.1), *Arabidopsis thaliana* (Ath) Indole glucosinolate OMT1 (Q9LPU5.1), *Mesembryanthemum crystallinum* (Mcr) Inositol 4-OMT (P45986.1), Msa Chalcone OMT (P93324.1), *Clarkia breweri* (Cbr) (Iso)eugenol OMT (O04385.2), Ath Flavone 3'-OMT (Q9FK25.1), *Chrysosplenium americanum* (Cam) Quercetin 3-OMT (P59049.1), Msa Caffeic acid 3-OMT (P28002.1), Cja (*S*)-Scoulerine 9-OMT (Q39522.1), Pso PSMT1 (AFB74611.1), Pso predicted OMT (AKO60157.1).

References

1. Chen, X.; Dang, T.-T. T.; Facchini, P. J., Noscapine Comes of Age. *Phytochemistry* **2015,** *111*, 7-13.

2. Li, Y.; Smolke, C. D., Engineering biosynthesis of the Anticancer Alkaloid Noscapine in Yeast. *Nature Commun* **2016,** *7*, 12137.

3. Winzer, T.; Gazda, V.; He, Z.; Kaminski, F.; Kern, M.; Larson, T. R.; Li, Y.; Meade, F.; Teodor, R.; Vaistij, F. E.; Walker, C.; Bowser, T. A.; Graham, I. A., A *Papaver Somniferum* 10-Gene Cluster For Synthesis Of The Anticancer Alkaloid Noscapine. *Science* **2012,** *336*, 1704-1708.

4. 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. *Nature Struct Mol Biol* **2001,** *8*, 271-279.

5. 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* **2002,** *14*, 1265-1277.

6. 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.* **2016,** *87*, 641-653.