Computational Biology and Chemistry 85 (2020) 107201

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/cbac

In silico identification and structure function analysis of a putative coclaurine *N*-methyltransferase from *Aristolochia fimbriata*

Roshan Ali[a,b,**], Yuannian Jiao[a,c], P. Kerr Wall[a], Simon G. Patching[d,*], Irshad Ahmad[b], Ghosia Lutfulla[e], Claude W. dePamphilis[a]

[a] Department of Biology, Life Sciences Building, The Pennsylvania State University, University Park, PA, USA
[b] Department of Molecular Biology and Genetics, Institute of Basic Medical Sciences, Khyber Medical University, Peshawar, Pakistan
[c] Institute of Botany, The Chinese Academy of Sciences, Beijing, China
[d] School of Biomedical Sciences and Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK
[e] Center of Biotechnology and Microbiology, University of Peshawar, Peshawar, Pakistan

ARTICLE INFO

Keywords:
Coclaurine *N*-methyltransferase
Aristolochia fimbriata
Basal angiosperm
Aristolochic acid
Reaction mechanism
Homology model

ABSTRACT

In this study we isolated and performed in silico analysis of a putative coclaurine *N*-methyltransferase (CNMT) from the basal angiosperm *Aristolochia fimbriata*. The Aristolochiaceae plant family produces alkaloids similar to the Papavaraceae family, and CNMTs are central enzymes in biosynthesis pathways producing compounds of ethnopharmacological interest. We used bioinformatics and computational tools to predict a three-dimensional homology model and to investigate the putative function of the protein and its mechanism for methylation. The putative CNMT is a unique (*S*)-adenosyl-*L*-methionine (SAM)-dependent *N*-methyltransferase, catalyzing transfer of a methyl group from SAM to the amino group of coclaurine. The model revealed a mixed α/β structure comprising seven twisted β-strands surrounded by twelve α-helices. Sequence comparisons and the model indicate an N-terminal catalytic Core domain and a C-terminal domain, of which the latter forms a pocket for coclaurine. An additional binding pocket for SAM is connected to the coclaurine binding pocket by a small opening. CNMT activity is proposed to follow an S_N2-type mechanism as observed for a similarly conformed enzyme. Residues predicted for the methyl transfer reaction are Tyr79 and Glu96, which are conserved in the sequence from *A. fimbriata* and in homologous *N*-methyltransferases. The isolated CNMT is the first to be investigated from any basal angiosperm.

1. Introduction

Aristolochia are a species of basal angiosperm that are a proposed model for experimental genetic studies on basal angiosperms (Jiao et al., 2011; Bliss et al., 2013). They have also been widely used in anticancer, antimalarial, antiasthmatic and slimming therapies and in the treatment of snake bite (Branch and Silva, 1983; Montes and Wilkomirsky, 1985; Li et al., 1994; Lopes et al., 2001; Ioset et al., 2003; Akindele et al., 2014; Das and Kumar, 2018). Due to the presence of aristolochic acids (Pailer, 1960; Schutte et al., 1967), natural products containing Aristolochia species have been found to be carcinogenic and nephropathic when taken internally (Arlt et al., 2002; Cosyns, 2003; Grollman et al., 2007). Aristolochic acid is synthesized through an alkaloid synthesis pathway similar to that found in opiate-producing poppies (*Papaver* spp.) (Schutte et al., 1967; Sharma et al., 1982). One of the important intermediates in the aristolochic acid biosynthetic pathway is the benzyltetrahydroisoquinoline alkaloid norlaudanosoline (tetrahydropapaveroline), which is methylated to *N*-methylnorlaudanosoline by an *N*-methyltransferase (NMT) (Schütte et al., 1967; Sharma et al., 1982). Coclaurine *N*-methyltransferase (CNMT) (Choi et al., 2001) is a unique (*S*)-adenosyl-*L*-methionine (SAM)-dependent NMT that catalyzes the transfer of a methyl group from SAM to the amino group of coclaurine and several alkaloids structurally similar to coclaurine such as (S)-norcoclaurine, (R, S)-6-*O*-methylcoclaurine and (R, S)-norlaudanosoline (Choi et al., 2001; Loeffler et al., 1995). CNMTs therefore catalyze the reaction: (*S*)-adenosyl-*L*-methionine + (S)-coclaurine ⇌ (*S*)-adenosyl-*L*-homocysteine + (S)-*N*-methylcoclaurine. Hence, CNMTs are a subclass of SAM-dependent NMTs [E.C.2.1.1.140], which are a diverse and important class of enzymes that methylate proteins, lipids, nucleic acids and other small molecules (Martin et al.,

* Corresponding author.
** Corresponding author at: Department of Molecular Biology and Genetics, Institute of Basic Medical Sciences, Khyber Medical University, Peshawar, Pakistan.
E-mail addresses: roshanali.ibms@kmu.edu.pk (R. Ali), s.g.patching@leeds.ac.uk, simonpatching@yahoo.co.uk (S.G. Patching).

https://doi.org/10.1016/j.compbiolchem.2020.107201
Received 24 June 2019; Received in revised form 31 December 2019; Accepted 8 January 2020
Available online 11 January 2020

Table 1
Comparison of amino acid compositions and other parameters of known CNMT sequences from different species of plants and an alga. The parameters include: length of the CNMT, molecular weight (MW), isoelectric point (pI), total content of negatively charged (Asp + Glu) and positively charged (Arg + Lys) residues, instability index, aliphatic index, overall mean hydropathicity.

Amino acid	Percentage contents of amino acids in CNMTs of different plants and an alga					
	A. fimbriata	*A. thaliana*	*C. japonica*	*T. flavum*	*O. sativa*	*C. reinhardtii*
Ala	6.2	5.9	5.9	5.3	9.5	7.0
Arg	3.6	3.9	4.2	3.9	5.6	7.3
Asn	2.5	2.8	4.2	4.2	3.6	5.0
Asp	6.4	5.4	5.3	5.3	4.5	3.9
Cys	2.0	2.0	2.0	2.2	1.9	1.1
Gln	5.0	1.7	4.2	3.9	2.2	3.4
Glu	9.2	8.7	10.6	9.1	8.1	6.4
Gly	5.6	5.6	4.5	5.0	3.6	6.4
His	2.8	3.1	2.8	2.8	3.3	3.1
Ile	5.3	6.8	6.1	6.1	5.6	3.6
Leu	12.0	10.1	11.7	10.8	11.4	10.4
Lys	7.8	8.7	7.0	8.3	6.1	3.4
Met	3.4	3.7	3.4	3.9	3.1	2.0
Phe	5.3	5.6	4.2	4.4	5.8	6.7
Pro	2.2	1.7	2.0	2.2	2.2	3.6
Ser	6.7	6.5	5.0	5.0	7.5	7.0
Thr	4.5	4.8	5.3	5.3	5.8	5.6
Trp	2.0	2.3	2.0	1.9	2.2	2.2
Tyr	3.9	4.8	4.5	3.9	4.7	5.3
Val	3.4	5.9	5.3	6.6	3.1	6.4
Length	357	355	358	361	359	357
MW	41,325.2 Da	41,229.7 Da	41,780.8 Da	41,903.4 Da	41,623.7 Da	41,239.8 Da
pI	5.21	6.16	5.12	5.77	6.47	7.84
Asp + Glu	56%	50%	57%	52%	45%	37%
Arg + Lys	41%	45%	40%	44%	42%	38%
Aliphatic i	83.6	89.0	90.0	90.4	84.6	80.3
Gah	−0.374	−0.164	−0.322	−0.261	−0.218	−0.275

2002; Miller et al., 2003). SAM-dependent NMTs are key enzymes in cellular biochemistry because they are involved in important biological processes including protein trafficking and sorting, metabolism, biosynthesis, signal transduction and gene expression. The benzylisoquinoline alkaloids, whose biosynthesis involves CNMTs, have been used in the development of therapeutics such as analgesics, antibiotics and antitumor agents. Enzymes from the biosynthesis of benzylisoquinoline alkaloids have also been used as biocatalysts in the production of non-natural products, which can provide different bioactive scaffolds for drug development. For example, plant tetrahydroisoquinoline alkaloids have been synthesized through an imine reductase route involving a CNMT (Yang et al., 2020). Relatively little is known about NMTs (Anaya et al., 2006; Larsson et al., 2006) and especially about CNMTs. Whilst a CNMT has been purified (Choi et al., 2001) and corresponding cDNAs have been isolated from *Tinospora cordifolia* (Loeffler et al., 1995), *Coptis japonica* (Choi et al., 2002), *Papaver somniferum* (Facchini and Park, 2003) and *Thalictrum flavum* (Samanani et al., 2005), a CNMT has not yet been investigated in Aristolochia species or in any other basal angiosperm. *Aristolochia* is a very useful genus for studying evolutionary biology (Bliss et al., 2013).

In this study, a cDNA encoding a putative CNMT was isolated from *Aristolochia fimbriata*. To get full benefit of the sequence data, we need to be able understand the structure and function of the encoded protein. The three-dimensional structure of the protein (Brenner and Levitt, 2000; Skolnick et al., 2000) can predict protein function, but the high-resolution structures of only a small fraction of proteins (Berman et al., 2002; Deshpande et al., 2005) have been solved by techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy or electron microscopy (EM), whilst a large number of sequences remain (Jimenez et al., 2005; Benson et al., 2009; Kiefer et al., 2009). Meanwhile, we can predict protein function and the structures of protein sequences by homology comparison and by using computational methods (Baker and Sali, 2001). Because the three-dimensional structure of CNMTs and their mechanism of reaction have not yet been explained, we have produced a three-dimensional homology model and a hypothetical reaction mechanism for the putative CNMT from *A. fimbriata*.

2. Materials and methods

2.1. Gene identification and isolation

For identifying all potential CNMT genes in *A. fimbriata*, the complete coding sequence of *P. somniferum* CNMT mRNA (GenBank Accession Number: AY217336) (Facchini and Park, 2003) was used for designing PCR primers. Flower buds and young fruits of *A. fimbriata*, grown under standard greenhouse conditions (Carlson et al., 2006) at the Pennsylvania State University, were collected, quick frozen in liquid nitrogen, and stored at −80 °C. Total RNA was extracted using an RNAqueous®-Midi Kit (Ambion, Inc., catalog # 1911) and quality and purity were evaluated by micro-capillary electrophoresis on an Agilent 2100 Bioanalyzer (RNA 6000 protocol). Total RNA was treated with RNase-free DNase 1 (Invitrogen DNase I cat. #18068-015) and mRNA was isolated using poly(A)Purist™ (Ambion, Inc, catalog # 1916). cDNA was amplified from mRNA using a PROMEGA Access RT-PCR System (PROMEGA, cat. #A1250) on an Eppendorf Mastercycler Gradient using the following program: For reverse transcription 48 °C for 45 min (once only); 94 °C for 2 min (once only), the amplification was done for 45 cycles of 94 °C for 30 s, 48.2 °C for 1 min and 70 °C for 2 min, finishing with 10 min at 68 °C. The forward and reverse primers had sequences of 5́-CAT GGC GTC TGA AAA GCT C-3́ and 5́-CCT ATT TTT TCT TGA AAA GCA GAT G-3́, respectively, with an expected amplicon of 1076 bp. The PCR product was run on a 2% agarose gel, stained with ethidium bromide and viewed using a Bio-Rad Gel Doc XR using Quantity One version 4.6.6 imaging software (Baxter, 2000; Ali, 2011) (Fig. S1). The expected bands were excised from the gel and extracted using a QIAquick® Gel Extraction Kit (Qiagen cat # 28704) then cloned into a pCR®II-TOPO® TA vector system (Invitrogen) using

(a)

```
MASEKLNKTE MLRRLEEGSV PDEEFRRLIR IELGRRLRWY CQKKPTYEEQ TAEIVALVKA 60
LRQMGITGDQ SDQLNSDLYD MPMSFLKITF GKLLKESGSY FKDDSMTLDE AEEAMLDLYC 120
ERAQLKDGQK ILDLGCGQGA FTLHAAQKYK KSHVTAVTNS ATQKKYIEDQ CQILELSNVE 180
VLLEDITQLT MEATFDRIIV IGLLEHMKNY GLLLQNISQW MAPADSLLFI DHVCHKSYFY 240
QCEPLDEDDW FAEYFFPPGS FAMPSASFLL YFQDDVSIVD HWILSGKHFH RTAEEWVKQL 300
DTNLEKGKEI LESKYGSKEA ALKAFNHWRG LCIFSSEIFG YNGGEEWMTS HLLFKKK    357
```

(b)

```
                                  *    *    *       * *  *   **  *
Coptis     : MAVEAKQT-- KKAAIVELLK QLELGLVPYD DIKQLIRREL ARRLQWGY-- --KPTYEEQI AEIQNLTHSL
Papaver    : ---------M QLKAKEELLR NMELGLIPDQ EIRQLIRVEL EKRLQWGY-- --KETHEEQL SQLLDLVHSL
Thalictrum : MAVEGKQVAP KKAIIVELLK KLELGLVPDD EIKKLIRIQL GRRLQWGC-- --KSTYEEQI AQLVNLTHSL
Oryza      : --MAMAARAA YLAATRAALA ALERNALPDA VTRRLTRLLL AQRLRLGYLP SSSSSAPLHL HHLLLFAHAL
Arabidopsis: --MEKIIDVA YGASVKVGLT LLEMNLLPDI VIRRLTRLLL AGRLRSGYKP ----TAEMQL SDLLRFVDSI

                  *  *     *      **  *  *     *  * * *  *     *  *  * * * * *  * **   *
Coptis     : RQMKIAT-EV ETLDSQLYEI PIEFLKIMNG SNLKGSCCYF KEDSTTLDEA EIAMLDLYCE RAQIQDGQSV
Papaver    : KGMKMAT-EM ENLDLKLYEA PMEFLKIQHG SNMKQSAGYY TDESTTLDEA EIAMLDLYME RAQIKDGQSV
Thalictrum : RQMKIAT-EV ETLDDQMYEV PIDFLKIMNG SNLKGSCCYF KNDSTTLDEA EIAMLELYCE RAQIKDGHSV
Oryza      : EEMPIAI-ET EKAKDQHYEL PTTFFKLVLG RNLKYSSCYF PDESSTLEDA EVAMLELYCE RAQLQDGQTI
Arabidopsis: KKMPIAI-HT EKPKTQHYEL PTAFFELVLG RNMKYSSCYF SNDSSSLEDA EEAILALYCE RAKVEDGQSV

                  ** ***      *  * *       *   **   *   *           * *    ***          ***
Coptis     : LDLGCGQGAL TLHVAQKYKN CRVTAVTNSV SQKEYIEEES RRRNLLNVEV KLADITTHEM AETYDRILVI
Papaver    : LDLGCGLGAV ALFGANKFKK CQFTGVTSSV EQKDYIEGKC KELKLTNVKV LLADITTYET EERFDRIFAV
Thalictrum : LDLGCGQGAL TLYVAQKYKN SRVTAVTNSV SQKEFIEEES RKRNLSNVEV LLADITTHKM PDTYDRILVV
Oryza      : LDVGCGWGSL SLYIAKKYSK CSITGICNST TQKAFIEEQC RENELSNVEI IVADISKFEM ERSFDRIISI
Arabidopsis: LDIGCGWGSL SLYIARKYSK CKLTGICNSK TQKAFIDEKC RKLGLQNIEI IVADISTFEH EGTYDRIFSI

                  *  ******   ** *   *     * ***     **** ***  *     **  *   *         **   **
Coptis     : ELFEHMKNYE LLLRKISEWI S-KDGLLFLE HICHKTFAYH YEPLDDDDWF TEYVFPAGTM IIPSASFFLY
Papaver    : ELIEHMKNYQ LLLKKISEWM K-DDGLLFVE HVCHKTLAYH YEPVDAEDWY TNYIFPAGTL TLSSASMLLY
Thalictrum : ELFEHMKNYE LLLRKIKEWM A-KDGLLFVE HICHKTFAYH YEPIDEDDWF TEYVFPAGTM IIPSASFFLY
Oryza      : EMFEHMKNYK ALLKKLSRWM K-EDSLLFVH YFCHKTFAYH FEDNNEDDWI TRYFFTGG-- TMPSANLLLY
Arabidopsis: EMFEHMKNYG ELLKKIGKWM K-EDSLLFVH YFCHKTFAYH FEDVNDDDWI TRHFFSGG-- TMPSADLLLY

                  **  **     *  ** *      *****  *                        *      **  *     * **
Coptis     : FQDDVSVVNH WTLSGKHFSR TNEEWLKRLD ANLDVIKPMF ETLMGN-EEE AVKLINYWRG FCLSGMEMFG
Papaver    : FQDDVSVVNQ WTLSGKHYSR SHEEWLKNMD KNIVEFKEIM RSITKT-EKE AIKLLNFWRI FCMCGAELFG
Thalictrum : FQDDVSVVNH WTLSGKHFSR TNEEWLKRLD ANVELIKPMF VTITGQCRQE AMKLINYWRG FCLSGMEMFG
Oryza      : FQDDVSIANH WLVSGTHYAR TSEEWLKRMD KNITSIRPIF EKTYGK--ES ATKWIAYWRT FFISVAELFG
Arabidopsis: FQENVSIVDH WLVSGTHYAK TSEEWLKRMD KEIVAVKEIM EVTYGK--EE AVKWMVYWRT FFIAVAELFG

                  * ** ***   * ** **
Coptis     : YNNGEEWMAS HVLFKKK
Papaver    : YKNGEEWMLT HLLFKKK
Thalictrum : YNNGEEWMAS HVLFKKK
Oryza      : YNNGDEWMVA HFLFRKK
Arabidopsis: YNNGEEWMVA HFLFKKK
```

Fig. 1. Conserved residues in a putative CNMT from A. fimbriata and predicted active site residues in CNMTs. (a) Conserved residues (*red*) in the sequence of a putative CNMT from A. fimbriata as identified by CASTp (Dundas et al., 2006). (b) Active site residues (*red*) in aligned sequences of CNMTs from different plant species (C. japonica, *P. somniferum*, T. flavum, O. sativa and A. thaliana) as predicted by ModBase (Pieper et al., 2014). The large majority of active site residues are conserved in all five sequences. All conserved residues are indicated by an asterisk.

One Shot® TOPO10 chemically competent *E. coli* cells (Cat. No. C4040-10). The plasmid DNA was purified using a QIAprep Spin Miniprep Kit (Qiagen Cat. #27106) and sequenced using the same primers by The Penn State Genomics Core Facility at University Park (http://www.huck.psu.edu/facilities/genomics-Core-up/). The results were analyzed by Sequencher software (version 4.9). Geospiza's FinchTV (ver 1.4.0: http://www.geospiza.com/Products/finchtv.shtml) was used to trim low-quality regions from both the chromatograms and a manual alignment was made between them.

Sequence data from this article can be found in the EMBL/GenBank data libraries under accession number HQ343195 (https://www.ncbi.nlm.nih.gov/nucCore/HQ343195).

2.2. Bioinformatic analysis

To determine the function of the putative protein, similarity searches were performed using non-redundant BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch) and BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) tools of the NCBI-BLAST package (Altschul et al., 1997) (Table S1). For protein BLAST, the sequence was first translated with the "Translate" tool of SwissProt (https://web.expasy.org/translate/) (Gasteiger et al., 2003). The ExPASy tool ProtParam (https://web.expasy.org/protparam/) (Gasteiger et al., 2005) was used to calculate the molecular weight, amino acid composition, number of negatively and positively charged residues, half life and instability index of the protein sequence (Table 1). The protein was predicted to be non-membrane, non-protease, cytoplasmic and a transferase enzyme (transferring one carbon group) using the tools Mem Type-2L (http://www.csbio.sjtu.edu.cn/bioinf/MemType/) (Chou and Shen, 2007), ProtIdent (http://www.csbio.sjtu.edu.cn/bioinf/Protease/) (Chou and Shen, 2008), Proteome Analyst Specialized Subcellular Localization Server v2.5 (https://omictools.com/proteome-analyst-specialized-subcellular-localization-server-tool) (Lu et al., 2004) and EzyPred (http://www.csbio.sjtu.edu.cn/bioinf/EzyPred/) (Shen and Chou, 2007), respectively. For further confirmation, the ExPASy tools AACompIdent (https://web.expasy.org/aacompident/) (Wilkins et al., 1996) and TagIdent (https://web.expasy.org/tagident/) (Gasteiger et al., 2005) were also used taking theoretical values of the pI (=5.21) and molecular weight (=41,325.2 Da).

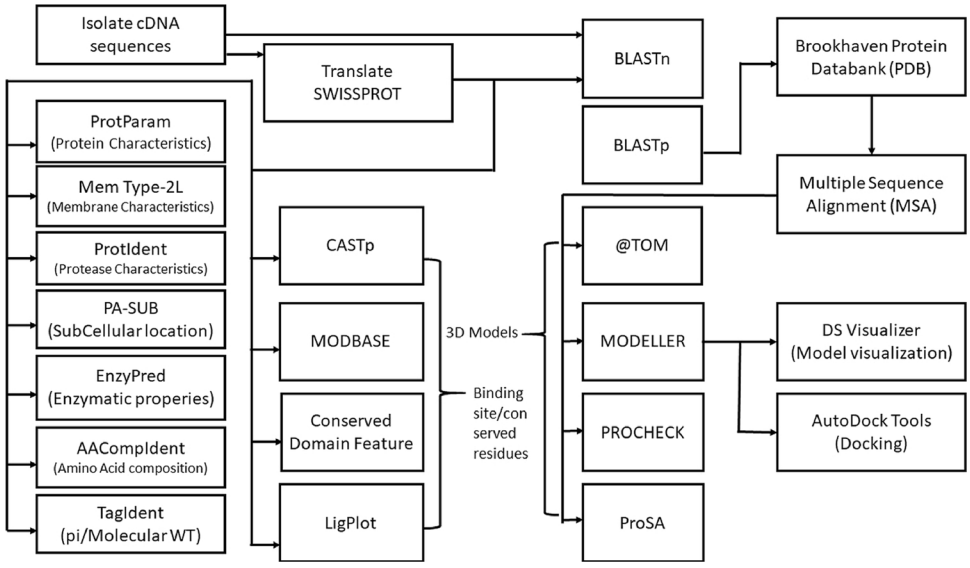

Fig. 2. Bioinformatics pipeline (details in Materials and methods).

2.3. Molecular modeling

Comparative modeling methods were employed to provide an insight into structural characteristics of the putative CNMT protein of *A. fimbriata*. To perform similarity searches for template selection, the putative CNMT sequence of *A. fimbriata* was subjected to BLAST search against the Brookhaven Protein Databank (PDB) (https://www.rcsb.org/) (Berman et al., 2000; Rose et al., 2015; Burley et al., 2017) using the program PSI-BLASTp of the NCBI-BLAST package with default parameters. Three sets of crystal structure coordinates of mycolic acid cyclopropane synthases from *M. tuberculosis* (PDB IDs: 1KPI, 1KPH and 2FK8 (Huang et al., 2002, Boissier et al., 2006), having sequence identities with the CNMT of 30%, 27% and 25%, respectively, were selected as the templates and were obtained from the PDB. Alignment of the *A. fimbriata* sequence with the template sequences was done with the help of @TOME 2.1 (Pons and Labesse, 2009), which aligns the sequences based on their sequence similarities as well as secondary structure similarities for homology modelling. Three dimensional models were built using MODELLER 9v8 (https://salilab.org/modeller) (Eswar et al., 2008). The stereochemistry of the models was evaluated with the help of PROCHECK (http://servicesn.mbi.ucla.edu/PROCHECK/) (Laskowski et al., 1993). ProSA (https://prosa.services.came.sbg.ac.at/prosa.php) (Sippl, 1993) was used for calculating energy graphs of the models. Z-score values of the putative CNMT model and templates were calculated with ProSA-web. The sequences were aligned, three dimensional models were built and the models were evaluated using @TOME 2.1 (Pons and Labesse, 2009), MODELLER 9v8 (https://salilab.org/modeller/) (Eswar et al., 2008), PROCHECK (http://servicesn.mbi.ucla.edu/PROCHECK/) (Laskowski et al., 1993) and ProSA (https://prosa.services.came.sbg.ac.at/prosa.php) (Sippl, 1993). Model No. 41 was selected as the best one. In order to examine any alteration in the Cα backbone of the model, it was superimposed on to the template 1KPI, 1KPH and 2FK8 crystal structure coordinates separately and helped by the SUPPERPOSE command of DS Visualizer® (v. 2, Accelrys Software Inc) (http://accelrys.com/products/collaborative-science/biovia-discovery-studio/requirements/technical-requirements-450.htm). Superimposition allowed us to calculate the root mean square deviation (RMSD) values for positional differences between equivalent atoms of the model and templates. The binding site and conserved residues of the protein were identified and assessed using CASTp (http://sts.bioe.uic.edu/castp/index.html?2pk9) (Fig. 1a) (Dundas et al., 2006), ModBase (https://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi) (Pieper et al., 2014) (Fig. 1b) and the Conserved Domain Database (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd_help.shtml) (Marchler-Bauer et al., 2009).

SAM and coclaurine (the substrate) were docked into the protein separately, one by one. The SAM molecule was docked into the protein model from the template structure of 2FK8 with the help of MODELLER. The three-dimensional structure of (S)-coclaurine (the substrate) was downloaded from the Chemspider database (http://www.chemspider.com). Coclaurine was docked into the model using AutoDock4 (http://autodock.scripps.edu/wiki/AutoDock4/) (Huey et al., 2007), and AutoGrid4 with help of the graphical user interface AutoDockTools http://autodock.scripps.edu/resources/adt) (Sanner, 1999). The molecules were prepared for AutoDock4 and AutoGrid4 with the help of ADT using the default parameters, Lamarkian genetic algorithm with local search, 25 million energy evaluations (medium) per run at population size of 150, and the number of GA runs was changed to 100 instead of 10 as a default. The protein was held rigid during the docking process while the ligand was allowed to be flexible. The number of rotatable bonds in the (*S*)-coclaurine was set to five. The grid-box size was 104 points × 122 points × 108 points in the x, y and z dimensions, respectively. The grid-box was centered with values of 92.713, 38.928 and 9.126 for X_center, Y_center and Z_center, respectively. The grid-box was set with AutoGrid4 and then the docking was performed with AutoDock4. AutoGrid4 and AutoDock4 were run from the graphical user interface of ADT. The results were analyzed with help of the analyze option of ADT. LIGPLOT (https://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/) (Wallace et al., 1995) was used to analyze ligand and substrate binding residues and the interactions of ligand and substrate with the protein. A summary of the whole bioinformatics pipeline is shown in Fig. 2.

3. Results and discussion

3.1. Gene identification and prediction of protein function

We isolated a 1076-bp gene from *A. fimbriata* RNA by PCR amplification (Fig. S1) and the extracted product had a sequence identical to that expected from database entries. A BLASTn search using this gene sequence showed that it is most closely related to a putative CNMT-like mRNA from *Thalictrum flavum* (Liscombe et al., 2009) (Table S1 and Fig. S2A) and the resultant translated proteins share 59.2% sequence identity. Most other similar coding sequences were NMTs, notably CNMT mRNAs from *T. flavum* (Samanani et al., 2005), *Coptis japonica* (experimentally expressed) (Choi et al., 2002), *Papaver somniferum*

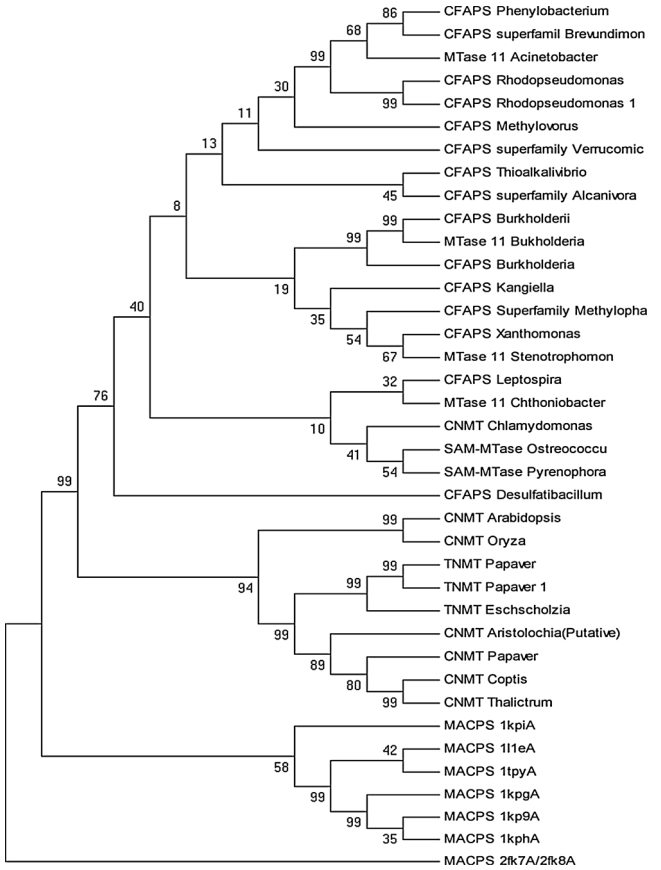

Fig. 3. Evolutionary relationships of *N*-methyltransferase enzymes. A nearest-neighbour phylogenetic tree for 38 *N*-methyltransferases including the A. fimbriata protein and cyclopropane fatty acyl phospholipid synthases (CFAPS), CNMTs, mycolic acid cyclopropane synthases (MACPS), methyltransferase_type-11 enzymes, SAM-MTases and tetrahydroprotoberberine cis-*N*-methyltransferases (TNMTs). This bootstrap consensus tree was inferred from 1000 replicates, where branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated sequences clustered together in the bootstrap test (1000 replicates) are shown next to the branches. All positions containing gaps and missing data were eliminated from the dataset (complete deletion and p-distance option). There were a total of 256 positions in the final dataset. The phylogenetic tree was constructed using Clustal X version 2.0 (Larkin et al., 2007) and MEGA 4.0 (Tamura et al., 2007; Kumar et al., 2008).

(Samanani et al., 2006) and *Arabidopsis thaliana* (Haas et al., 2002). These genes produced translated proteins with sequence identities ranging from 47.5-59.8% compared with that of the *A. fimbriata* gene. Similarly, a BLASTp search showed that almost all of the sequences were NMTs and many of them were CNMTs (Table S1). A BLAST search for the *A. fimbriata* protein sequence against the UniProt KnowledgeBase identified 167 hits with greater than 40% sequence identity. Phylogenetic analysis of the top 250 hits from the search located the *A. fimbriata* protein to a clade containing putative CNMTs from *C. japonica*, *T. flavum* and *P. somniferum* with sequence identities of 59.8%. 59.2% and 52.4%, respectively (Fig. S2B). A multiple sequence alignment identified conserved residues in the *A. fimbriata* protein and CNMTs from plants (*Oryza Sativa*, *A. thaliana*, *T. flavum*, *C. japonica*, *P. somniferum*) and a green alga (*Chlamydomonas reinhardtii*) (Fig. S3). A larger multiple sequence alignment of 38 methyltransferase enzymes including cyclopropane fatty acyl phospholipid synthases (CFAPS), CNMTs, mycolic acid cyclopropane synthases (MACPS), methyltransferase_type-11 enzymes, SAM-MTases and tetrahydroprotoberberine cis-*N*-methyltransferases (TNMTs) (Fig. S4) was used to construct a phylogentic tree for demonstrating their evolutionary relationships. The *A. fimbriata* protein was confirmed as most closely related to the group of CNMTs from *C. japonica*, *T. flavum* and *P. somniferum* (Fig. 3).

The isolated *A. fimbriata* protein contains an amino acid composition almost identical to that of other plant and algal CNMTs (Table 1). The greater number of negatively charged residues (Asp + Glu) might confer a slight acidic character to the putative CNMT, as suggested by the calculated pI value of 5.21. A high content of acidic residues also suggests that the protein is intracellular (Nishikawa et al., 1983). Similar values were observed for different parameters of the *A. fimbriata* protein compared with average values of other known CNMTs (Table 1), further confirming the protein as a putative CNMT. According to TagIdent, the closest match to the protein in terms of MW and pI was a CNMT from *C. japonica*. The closest Swiss-Prot entries (in terms of amino acid composition) and having MW and pI values in the specified range (MW: 33,060-49,590 Da, pI: 4.71-5.71) for the "COPTIS" species, was *S*-norcoclaurine synthase 1 (MW: 39,964 Da, pI: 5.27). The closest TrEMBL entries (in terms of amino acid composition) and having MW and pI values in the specified range for the "COPTIS" species was a CNMT (MW: 41,781 Da, pI: 5.12).

The Pfam database (https://pfam.xfam.org/) identified the *A. fimbriata* protein as a member of the mycolic acid cyclopropane synthetase (CMAS) family. A Pfam search for the CNMT sequence of *P. somniferum* as a query gave the same result. Scanning of ModBase (Pieper et al., 2014) for theoretical models of similar sequences, in which the option of sequence identity was changed to 40%, found only ten models out of which five were annotated as CNMTs. CASTp (Dundas et al., 2006) identified 68 conserved residues in the putative CNMT sequence (Fig. 1a) and the Conserved Domain Database (Marchler-Bauer et al., 2009) identified thirteen residues (134-LGCGQGA-140, 158-TN-159, 184-EDI-186 and 201-I) as the most conserved. The common residues predicted by both CDD and CASTp (Fig. 1a) were 133-DLGCG-137, 158-TN-159, 186-I-186 and 201-I-201. The SAM-binding residues predicted by ModBase were almost identical to these (Fig. 1b). Most of the SAM-binding residues predicted by these methods are among the conserved residues of all the selected CNMTs as indicated by the multiple alignment (Fig. 1b).

From all of these results and observations it can be hypothesized that the protein (357 residues) produced from the sequenced *A. fibriata* gene is a CNMT. Although the homologous sequences found in the searches were genes from a diverse range of eudicot (*Thalictrum*, *Coptis*, *Papaver*, *Eschscholzia*, *Populus*, *Solanum*, *Vitis*, *Arabidopsis*, *Medicago*) and monocot (*Oryza*, *Zea*) species, none of the top hits were identified as coming from a basal angiosperm. The sequence obtained from *A. fibriata* therefore appears to be the first CNMT homolog that has been cloned and characterized from a basal angiosperm lineage.

3.2. Three-dimensional homology model and active site of a CNMT

For producing a three-dimensional homology model of the putative CNMT from *A. fimbriata*, three sets of crystal structure coordinates of MACPS from *Mycobacterium tuberculosis* (PDB IDs: 1KPH, 1KPI and 2FK8) (Huang et al., 2002; Boissier et al., 2006) were selected as model templates. These proteins have sequence identities with the putative CNMT of 30%, 27% and 25%, respectively. 1KPH was selected on the basis of highest sequence identity out of all potential templates. 1KPI was selected because SWIS(S)-MODEL (Arnold et al., 2006) selected it as the most suitable template on the basis of a "HHSearch template library search". Also, the phylogenetic tree calculated using all of the top templates showed that 1KPI was the best template (Fig. 3). 2FK8 was selected because it aligns to the greater part of the *A. fimbriata* protein sequence and it contains the ligand SAM. E-values for matches with these three templates were all reliable at less than 10^{-4}.

A total of 110 homology models were built for the *A. fimbriata* protein using the "generally conserved region" (residues 76 to 348). Ramachandran plots identified model No. 41 as the best (Fig. 4), which

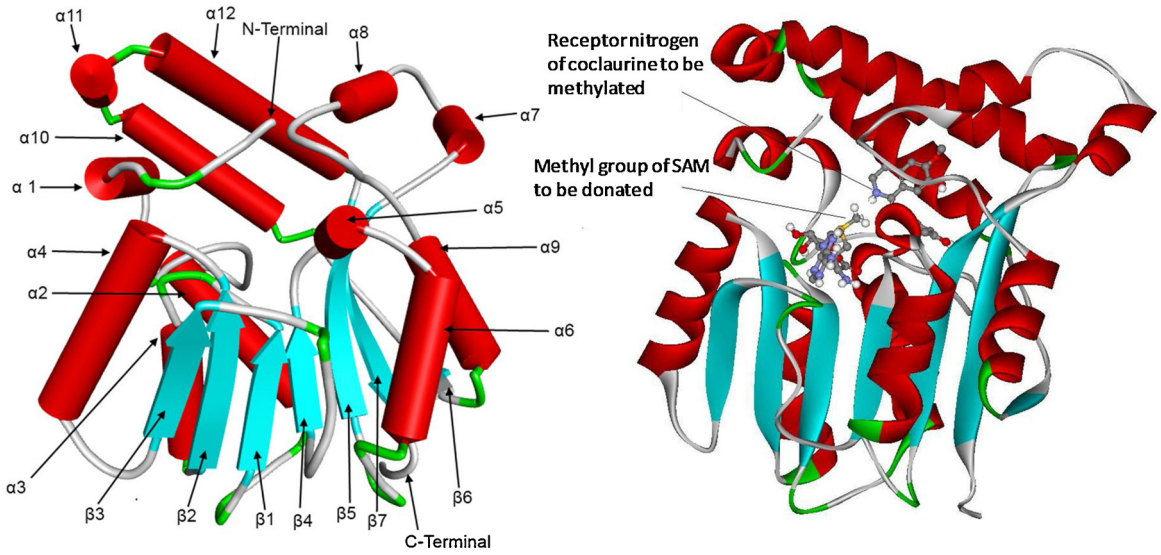

Fig. 4. Homology model (No. 41) of a putative CNMT from A. fimbriata. The model consists of 7 β-strands (*cyan*) surrounded by 12 α-helices (*red*). The first five β-strands are parallel with each other and the last two are antiparallel (*left*). The model is also shown in solid ribbon representation with the substrate and ligand in ball and stick representation coloured as follows: carbon atoms = grey, oxygen atoms = red, nitrogen atoms = purple, hydrogen atoms = white, sulfur atoms = yellow. The methyl group of SAM is to be donated to N4 of (S)-coclaurine. This figure was produced using Discovery Studio Visualizer (http://accelrys.com/products/collaborative-science/biovia-discovery-studio/visualization-download.php).

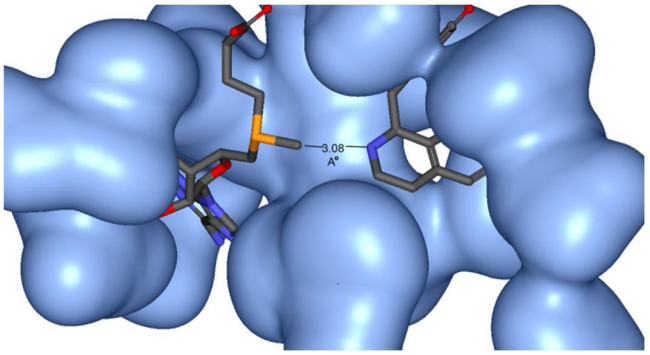

Fig. 5. Stick representations of SAM and (S)-coclaurine within their binding pockets of a putative CNMT from A. fimbriata. Each ligand is in a separate, but adjacent cavity. The two cavities are connected by a narrow opening through which methyl group transfer takes place. This figure was produced using Discovery Studio Visualizer (http://accelrys.com/products/collaborative-science/biovia-discovery-studio/visualization-download.php).

had 88.0%, 9.6%, 2.4% and 0.0% residues in the most favored, additional allowed, generously allowed and disallowed regions, respectively (Fig. S5). In comparison, Ramachandran plots for the templates 1KPI, 1KPH and 2FK8 had 89.5%, 93.2% and 94.4% residues in the most favoured region, 10.1%, 5.2% and 5.2% residues in the additional allowed region, 0.4%, 0.8% and 0.4% residues in the generously allowed region, and 0.0%, 0.8% and 0.0% residues in the disallowed region, respectively (Figs S6–S8). The overall model was superimposed best to the 1KPH crystal structure (CmaA1 from *M. tuberculosis*) with an overall rmsd of 2.90 Å. There were three regions of significant deviation (residues 207-210, 233-248 and 279-287 in the model) in an outer flexible part of the protein (Fig. S9). The ligand is nearly at the same position and in the same orientation as in template structure 2FK8 (Hma/MmaA4 from *M. tuberculosis* in complex with SAM) (Fig. S10). The small difference for the position of the ligand in the model compared with 2FK8 would be expected considering the relatively low sequence identity. In general, the model superimposed well over all of the templates, indicating correctness of the model.

The model has a mixed α/β structure with a conserved catalytic Core domain (Schluckebier et al., 1995) comprised of a central sheet of seven twisted β-strands surrounded by twelve α-helices (Fig. 4). This folding pattern is almost identical to that observed in all other SAM-dependent NMTs (Vidgren et al., 1994; Reinisch et al., 1995; Fu et al., 1996). The model also contains a substrate-recognition domain appended to the SAM-binding domain, which is variable among SAM-dependent NMTs (Martin and McMillan, 2002). In the model, the first five β-strands (order: 3-2-1-4-5) and the sixth β-strand are all parallel to each other, while the last β-strand (between strands 5 and 6) is antiparallel to them (Fig. 4). The cavity for SAM and coclaurine (Fig. 5) is surrounded by three α-helices and the C-terminal residues of β1, β4 and β5 are perpendicular to the cavity. The residues that make hydrogen bond interactions with SAM are Glu96 and Ser97 (on the loop between α1 and α2), Gly135 (on β1), Gly137 (on the loop between β1 and α3), Thr158 and Asn159 (on β2), Gln163 (on α4), Ile186 (on the loop between β3 and β4) and Ile201 (on β4). The residues that make hydrophobic interactions with SAM are Leu134, Val157, Asp185, Leu203, His206 and Phe256 (on β1, β2, loop between β3 and β4, loop between β4 and α5, α5, and α9 respectively). The substrate is surrounded by Asp231, His232, Glu96, Gly202 and four Phe (251, 256, 325 and 349) residues, arranged in such a way to form a pocket for coclaurine. The four Phe residues and the Gly cannot make hydrogen bonds and the Asp is oriented in such a way that is unable to make a hydrogen bond. However, the N_{D1} of His232 and O_{E1} of Glu96 form hydrogen bonds with O2 and N4 of coclaurine, respectively. Glu96 appears to play an important role in transfer of the methyl group. Out of the coclaurine-binding residues, only one residue (Phe256) is found amongst the overall conserved residues. Other substrate and SAM-binding residues are also conserved, although the residues are different in some of the CNMTs, but those different residues are characteristically similar.

3.3. Putative reaction mechanism of a CNMT

In the putative CNMT model, SAM (Clarke and Banfield, 2001) binds to a pocket situated next to the substrate binding pocket (Martin and McMillan, 2002), which agrees with the fact that the SAM-dependant methylation reaction involves direct transfer of the methyl group from SAM to the substrates (Ta and Kim, 2010). SAM binding is mediated through extensive hydrogen and other van der Waals interactions. There are a total of ten hydrogen-bonding interactions to SAM

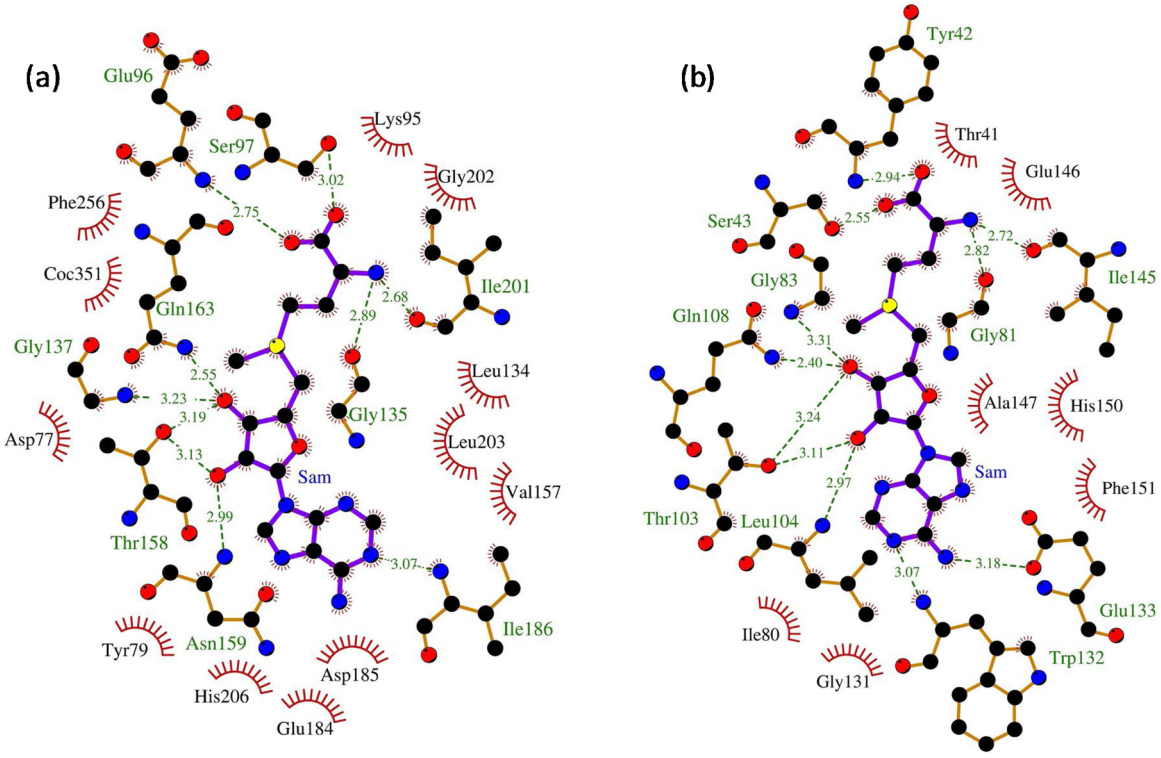

Fig. 6. SAM binding residues identified by LIGPLOT. SAM binding residues in the model of a putative CNMT from *A. fimbriata* (a) and in the crystal structure of Hma from *M. tuberculosis* in complex with SAM (PDB 2FK8) (b) identified by LIGPLOT (https://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/). Atoms are colored as follows: carbon = black, nitrogen = blue, oxygen = red, sulfur = yellow.

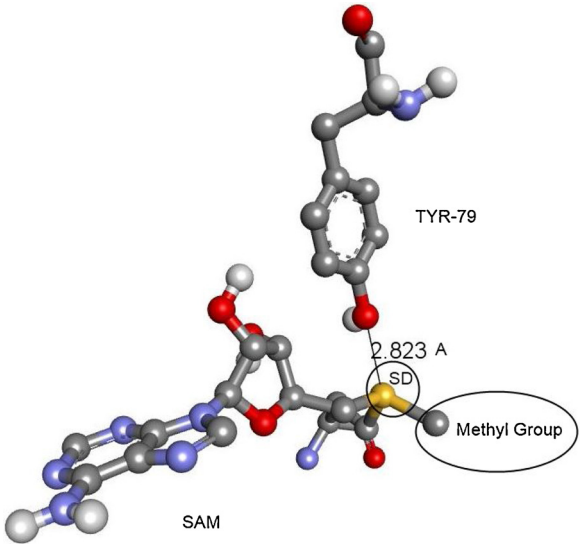

Fig. 7. Ball and stick representation of Tyr79 and SAM in the model of a putative CNMT from A. fimbriata. The OH of Tyr79 is positioned to have a charge-dipole interaction with the positively charged S_D of SAM, detracting from the covalent bond of the methyl group to SAM. This figure was produced using Discovery Studio Visualizer (http://accelrys.com/products/collaborative-science/biovia-discovery-studio/visualization-download.php).

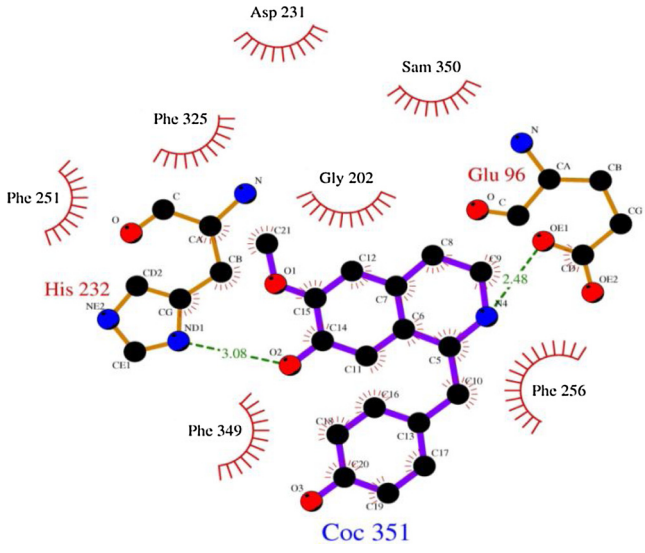

Fig. 8. Coclaurine binding residues identified by LIGPLOT. Coclaurine binding residues in the model of a putative CNMT from A. fimbriata identified by LIGPLOT (https://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/). Atoms are colored as follows: carbon = black, nitrogen = blue, oxygen = red, sulfur = yellow.

(Fig. 6a), whilst there are eleven in template 2FK8 (Fig. 6b). One less bond is due to the fact that in the putative CNMT, Glu133 (in 2FK8) is replaced with a shorter residue Thr (in the CNMT), which cannot make a hydrogen bond with SAM due to a greater distance between them (5.243 Å). The four hydrogen bonding interactions of amino and carboxyl tails of SAM (Fig. 6, a and b) have also been observed in a glycine-NMT with almost similar residues (Huang et al., 2000). The hydroxyl of Tyr79 is positioned to have a charge-dipole interaction with the positively charged S_D of SAM, detracting from the covalent bond of the methyl group to SAM (Fig. 7). There are several hydrophobic interactions between SAM and certain residues in the CNMT model (Fig. 6a), whilst there are fewer hydrophobic interactions between SAM and amino acid residues in 2FK8 (Fig. 6b). The residues surrounding the substrate binding cavity therefore change according to the nature of the substrate and its interactions with SAM (McCarthy and McCarthy,

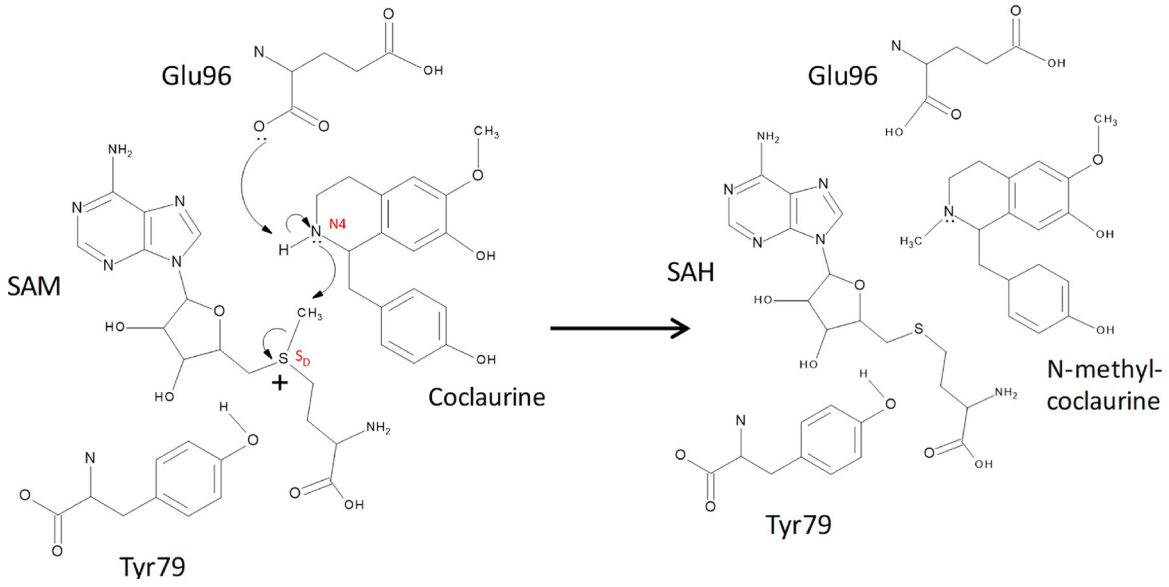

Fig. 9. Hypothetical reaction mechanism for coclaurine methylation by a CNMT. This Figure was produced using ChemDraw.

2007).

In the model of the putative CNMT, the substrate is surrounded by four phenyl rings and a guanidinium ring of His in such a way that it makes a pocket to accommodate coclaurine. One of the four Phe residues (Phe256) surrounding coclaurine is conserved in all CNMTs (Fig. 1b). Coclaurine makes a total of two hydrogen bonds (Fig. 8). There are a total of six residues that make hydrophobic interactions with coclaurine.

SAM is a highly reactive molecule and loses its methyl group on receiving electrons, thus allowing methylation of substrates. Based on the arrangement of conserved residues, the location and orientation of the bound SAM, the proposed binding mode of coclaurine and proposed mechanism of reaction for the glycine-NMT (Takata et al., 2003), a hypothetical S_N2-type mechanism is proposed for transmethylation by the putative CNMT (Fig. 9). Tyr79 (Fig. 7), Glu96 and Phe256 (Fig. 6a and b) might be the most important residues involved in transfer of the methyl group to the substrate, because these groups are at the contact opening between the pockets of the substrate and SAM.

Coclaurine binds in its active site in such a way that the lone pair of electrons on N4 of coclaurine is directly opposite the methyl carbon of SAM, which is oriented in such a way that its methyl carbon projects towards N4 of coclaurine through the small opening (Fig. 5). This orientation was first described in *Streptococcus pneumoniae* Sp1610 (Ta and Kim, 2010). It can therefore be predicted that the CNMT catalyzes the methyl transfer reaction by "proximity and orientation effects" in a single S_N2 step (Coward, 1977). Tyr79 could form a charge dipole interaction with the positively charged S_D atom of SAM, facilitating the methyl transfer reaction (Takata et al., 2003; McCarthy and McCarthy, 2007) (Fig. 7). For example, in glycine-NMTs Tyr21 has been implicated in the methyl transfer mechanism (Takata et al., 2003). Similarly, Glu96 is positioned in a way that it can take part in the removal of the hydrogen atom from N4 of coclaurine. The hydroxyl group of Tyr79 has the potential to induce a positive charge on the S_D atom of SAM. First, the hydroxyl of Tyr79 is possibly deprotonated. His206 may act as a base and has the potential to accept the proton from Tyr79 (Wada et al., 2006). The resulting oxyanion of Tyr79 can interact with the S_D atom of SAM. The electrons are attracted from the methyl group toward the S_D of SAM due to which the methyl group becomes electrophilic for the lone pair available at N4 of coclaurine. The bond between the S_D atom of SAM and the methyl group is weakened. Interactions between the sulfur atom of SAM and an oxygen atom of an enzyme have been observed in other NMTs (Bussiere et al., 1998; Jacobs et al., 2002; Wada et al., 2006), but in some cases the situation is different. For example, in BchU the Nδ atom of His150 and the sulfur atom of SAM are positioned in close vicinity (3.1 Å) of each other (Wada et al., 2006). In the model of the putative CNMT, His206 is in close proximity to the S_D atom of SAM, but the distance is 5.28 Å which is greater than that of BchU. The positive charge induced on the S_D atom of SAM also induces a positive charge on the carbon of the attached methyl group. The negatively charged unshared pair of electrons on N4 of coclaurine is attracted by the positively charged methyl group of SAM (Zubieta et al., 2003). The carboxyl oxygen of Glu96 contributes to the loss of hydrogen from N4 as the bond between the methyl group and the S_D atom of SAM is broken. A new bond is formed between the methyl group and N4 of coclaurine, allowing Glu96 to accept the hydrogen, thus completing methylation of coclaurine by SAM. All these reactions are proposed to happen simultaneously by the catalysis of SAM, which utilizes an S_N2 mechanism. This putative S_N2 reaction mechanism is identical to the experimentally determined mechanism of the glycine-NMT and remains to be confirmed experimentally in the *A. fimbriata* putative CNMT.

Conflict of interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Higher Education Commission of Pakistan.

CRediT authorship contribution statement

Roshan Ali: Conceptualization, Methodology, Software, Investigation, Validation, Formal analysis, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration, Funding acquisition. **Yuannian Jiao:** Methodology, Software, Investigation, Writing - review & editing, Visualization. **P. Kerr Wall:** Methodology, Software, Investigation, Writing - review & editing, Visualization. **Simon G. Patching:** Investigation, Validation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Irshad Ahmad:** Investigation, Validation, Formal analysis, Writing - review & editing, Visualization. **Ghosia Lutfulla:** Methodology, Software, Investigation, Writing -

review & editing, Visualization. **Claude W. dePamphilis:** Conceptualization, Validation, Formal analysis, Resources, Data curation, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Acknowledgements

The authors are thankful to Paula Ralph, Lena Landherr, Nadira Nazneen, Bindhu, Laura and Barbara Bliss for cooperation in the experimental work as well as in the computational work.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.compbiolchem.2020.107201.

References

Akindele, A.J., Wani, Z., Mahajan, G., Sharma, S., Aigbe, F.R., Satti, N., Adeyemi, O.O., Mondhe, D.M., 2014. Anticancer activity of *Aristolochia ringens Vahl.* (Aristolochiaceae). J. Tradit. Complement. Med. 5 (1), 35–41.

Ali, R., 2011. Sequencing and analysis of a cDNA encoding a putative coclaurine *N*-methyltransferase (CNMT) from *Aristolochia fimbriata*, a basal angiosperm [PhD]. University of Peshawar, Peshawar.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25 (17), 3389–3402.

Anaya, A.L., Cruz-Ortega, R., Waller, G.R., 2006. Metabolism and ecology of purine alkaloids. Front Biosci. 11, 2354–2370.

Arlt, V., Stiborova, M., Schmeiser, H., 2002. Aristolochic acid as a probable human cancer hazard in herbal remedies: a review. Mutagenesis 17 (4), 265–277.

Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22 (2), 195–201.

Baker, D., Sali, A., 2001. Protein structure prediction and structural genomics. Science 294, 93–96.

Baxter, B., 2000. Using the "Quantity One" software package for quantitative analysis of gels. [cited 2008]; Available from: http://legacy.lclark.edu/~lycan/Bio312/Quantity%20onerev%27d08.pdf.. .

Bliss, B.J., Wanke, S., Barakat, A., Ayyampalayam, S., Wickett, N., Wall, P.K., Jiao, Y., Landherr, L., Ralph, P.E., Hu, Y., Neinhuis, C., Leebens-Mack, J., Arumuganathan, K., Clifton, S.W., Maximova, S.N., Ma, H., dePamphilis, C.W., 2013. Characterization of the basal angiosperm *Aristolochia fimbriata*: a potential experimental system for genetic studies. BMC Plant Biol 13 (13).

Branch, L., Silva, M., 1983. Folk medicine of Alter do Chão, Pará. Brazil. Acta Amaz. 13 (5-6), 737–797.

Brenner, S.E., Levitt, M., 2000. Expectations from structural genomics. Prot. Sci. 9, 197–200.

Burley, S.K., Berman, H.M., Kleywegt, G.J., Markley, J.L., Nakamura, H., Velankar, S., 2017. Protein Data Bank (PDB): the single global macromolecular structure archive. Methods Mol. Biol. 1607, 627–641.

Bussiere, D.E., Muchmore, S.W., Dealwis, C.G., Schluckebier, G., Nienaber, V.L., Edalji, R.P., Walter, K.A., Ladror, U.S., Holzman, T.F., Abad-Zapatero, C., 1998. Crystal structure of ErmC, an rRNA methyltransferase which mediates antibiotic resistance in bacteria. Biochemistry 37 (20), 7103–7112.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2009. GenBank. Nucleic Acids Res. 37, D26–D31 Database issue.

Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D., Zardecki, C., 2002. The protein data bank. Acta Crystallogr. D Biol. Crystallogr. 58 (Pt 6 No 1), 899–907.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucleic Acids Res. 28 (1), 235–242.

Boissier, F., Bardou, F., Guillet, V.R., Uttenweiler-Joseph, S., Daffe, M., Quemard, A., Mourey, L., 2006. Further insight into S-adenosylmethionine-dependent methyltransferases - structural characterization of Hma, an enzyme essential for the biosynthesis of oxygenated mycolic acids in *Mycobacterium tuberculosis*. J. Biol. Chem. 281 (7), 4434–4445.

Carlson, J.E., Leebens-Mack, J.H., Wall, P.K., Zahn, L.M., Mueller, L.A., Landherr, L.L., Hu, Y., Ilut, D.C., Arrington, J.M., Choirean, S., Becker, A., Field, D., Tanksley, S.D., Ma, H., dePamphilis, C.W., 2006. EST database for early flower development in California poppy (*Eschscholzia californica Cham., Papaveraceae*) tags over 6,000 genes from a basal eudicot. Plant Mol. Biol. 62 (3), 351–369.

Choi, K.B., Morishige, T., Sato, F., 2001. Purification and characterization of coclaurine *N*-methyltransferase from cultured *Coptis japonica* cells. Phytochem. 56 (7), 649–655.

Choi, K.B., Morishige, T., Shitan, N., Yazaki, K., Sato, F., 2002. Molecular cloning and characterization of coclaurine *N*-methyltransferase from cultured cells of *Coptis japonica*. J. Biol. Chem. 277 (1), 830–835.

Chou, K.C., Shen, H.B., 2007. MemType-2L: a web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. Biochem. Biophys. Res. Commun. 360 (2), 339–345.

Chou, K.C., Shen, H.B., 2008. ProtIdent: a web server for identifying proteases and their types by fusing functional domain and sequential evolution information. Biochem. Biophys. Res. Commun. 376 (2), 321–325.

Clarke, S., Banfield, K., 2001. In: Carmel, R., Jacobsen, D.W. (Eds.), Homocysteine in health and disease. Cambridge University Press, Cambridge, UK ; New York p. 63.

Cosyns, J., 2003. Aristolochic acid and Chinese herbs nephropathy: a review of the evidence to date. Drug Saf. 26 (1), 33–48.

Coward, J.K., 1977. Chemical mechanisms of methyl transfer reactions: comparison of methylases with nonenzymic' model reactions'. In: Salvatore, F. (Ed.), The Biochemistry of adenosylmethionine : [proceedings of an international symposium on the biochemistry of adenosylmethionine, sponsored by the Accademia nazionale dei Lincei, held in Rome, Italy May 21-26, 1974]. Columbia University Press, New York, pp. 127–144.

Das, A., Kumar, G.S., 2018. Natural aristolochia alkaloid aristololactam-β-D-glucoside: interaction with biomacromolecules and correlation to the biological perspectives. Mini Rev. Med. Chem. 18 (12), 1022–1034.

Deshpande, N., Addess, K.J., Bluhm, W.F., Merino-Ott, J.C., Townsend-Merino, W., Zhang, Q., Knezevich, C., Xie, L., Chen, L., Feng, Z., Green, R.K., Flippen-Anderson, J.L., Westbrook, J., Berman, H.M., Bourne, P.E., 2005. The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. Nucleic Acids Res. 33, D233–D237 Database issue.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., Liang, J., 2006. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res. 34, W116–W118 Web Server issue.

Eswar, N., Eramian, D., Webb, B., Shen, M.Y., Sali, A., 2008. Protein structure modeling with MODELLER. Methods Mol. Biol. 426, 145–159.

Facchini, P.J., Park, S.U., 2003. Developmental and inducible accumulation of gene transcripts involved in alkaloid biosynthesis in opium poppy. Phytochem 64 (1), 177–186.

Fu, Z., Hu, Y., Konishi, K., Takata, Y., Ogawa, H., Gomi, T., Fujioka, M., Takusagawa, F., 1996. Crystal structure of glycine *N*-methyltransferase from rat liver. Biochemistry 35 (37), 11985–11993.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A., 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31 (13), 3784–3788.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis ools on the ExPASy server. In: Walker, J.M. (Ed.), Proteomics protocols handbook. Humana Press, Totowa, N.J 571-560.

Grollman, A.P., Shibutani, S., Moriya, M., Miller, F., Wu, L., Moll, U., Suzuki, N., Fernandes, A., Rosenquist, T., Medverec, Z., Jakovina, K., Brdar, B., Slade, N., Turesky, R.J., Goodenough, A.K., Rieger, R., Vukelić, M., Jelaković, B., 2007. Aristolochic acid and the etiology of endemic (Balkan) nephropathy. Proc. Natl. Acad. Sci. U. S. A. 104 (29), 12129–12134.

Haas, B.J., Volfovsky, N., Town, C.D., Troukhan, M., Alexandrov, N., Feldmann, K.A., Flavell, R.B., White, O., Salzberg, S.L., 2002. Full-length messenger RNA sequences greatly improve genome annotation. Genome Biol. 3 (6) RESEARCH0029.

Huang, Y., Komoto, J., Konishi, K., Takata, Y., Ogawa, H., Gomi, T., Fujioka, M., Takusagawa, F., 2000. Mechanisms for auto-inhibition and forced product release in glycine *N*-methyltransferase: crystal structures of wild-type, mutant R175K and S-adenosylhomocysteine-bound R175K enzymes. J. Mol. Biol. 298 (1), 149–162.

Huang, C.C., Smith, C.V., Glickman, M.S., Jacobs Jr, W.R., Sacchettini, J.C., 2002. Crystal structures of mycolic acid cyclopropane synthases from *Mycobacterium tuberculosis*. J. Biol. Chem. 277 (13), 11559–11569.

Huey, R., Morris, G.M., Olson, A.J., Goodsell, D.S., 2007. A semiempirical free energy force field with charge-based desolvation. J. Comput. Chem. 28 (6), 1145–1152.

Ioset, J., Raoelison, G., Hostettmann, K., 2003. Detection of aristolochic acid in Chinese phytomedicines and dietary supplements used as slimming regimens. Food Chem. Toxicol. 41 (1), 29–36.

Jacobs, S.A., Harp, J.M., Devarakonda, S., Kim, Y., Rastinejad, F., Khorasanizadeh, S., 2002. The active site of the SET domain is constructed on a knot. Nat. Struct. Biol. 9 (11), 833–838.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E., Schuster, S.C., Ma, H., Leebens-Mack, J., dePamphilis, C.W., 2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473 (7345), 97–100.

Jimenez, S., Rojas, S.F., Bairoch, A., Team, H., 2005. UNIPROTKB/SWISS-PROT and the human proteomics initiative. Mol. Cell. Proteom. 4 (8), S23.

Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L., Schwede, T., 2009. The SWISS-MODEL repository and associated resources. Nucleic Acids Res. 37, D387–D392 Database issue.

Kumar, S., Nei, M., Dudley, J., Tamura, K., 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief. Bioinform. 9 (4), 299–306.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23 (21), 2947–2948.

Larsson, K.A., Zetterlund, I., Delp, G., Jonsson, L.M., 2006. *N*-Methyltransferase involved in gramine biosynthesis in barley: cloning and characterization. Phytochem 67 (18), 2002–2008.

Laskowski, R.A., Macarthur, M.W., Moss, D.S., Thornton, J.M., 1993. Procheck - a program to check the stereochemical quality of protein structures. J. Appl. Crystal. 26, 283–291.

Li, H., Youji, S., Shingo, M., Chen, X., 1994. Eleven aristolochic acid derivatives from

Aristolochia cinnabarina. Phytochem 37 (1), 237–239.
Liscombe, D.K., Ziegler, J., Schmidt, J., Ammer, C., Facchini, P.J., 2009. Targeted metabolite and transcript profiling for elucidating enzyme function: isolation of novel *N*-methyltransferases from three benzylisoquinoline alkaloid-producing species. Plant J. 60 (4), 729–743.
Loeffler, S., Deusneumann, B., Zenk, M.H., 1995. S-Adenosyl-*L*-methionine-(S)-coclaurine-*N*-methyltransferase from *Tinospora cordifolia*. Phytochem 38 (6), 1387–1395.
Lopes, L., Nascimento, I., Silva, T., 2001. Phytochemistry of the Aristolochiaceae family. Res. Adv. Phytochem. 2, 19–108.
Lu, Z., Szafron, D., Greiner, R., Lu, P., Wishart, D.S., Poulin, B., Anvik, J., Macdonell, C., Eisner, R., 2004. Predicting subcellular localization of proteins using machine-learned classifiers. Bioinformatics 20 (4), 547–556.
Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Tasneem, A., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Bryant, S.H., 2009. CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Res. 37, D205–D210 Database issue.
Martin, J.L., McMillan, F.M., 2002. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Curr. Opin. Struct. Biol. 12 (6), 783–793.
McCarthy, A.A., McCarthy, J.G., 2007. The structure of two *N*-methyltransferases from the caffeine biosynthetic pathway. Plant Physiol. 144 (2), 879–889.
Miller, D.J., Ouellette, N., Evdokimova, E., Savchenko, A., Edwards, A., Anderson, W.F., 2003. Crystal complexes of a predicted S-adenosylmethionine-dependent methyltransferase reveal a typical AdoMet binding domain and a substrate recognition domain. Protein Sci. 12 (7), 1432–1442.
Montes, M., Wilkomirsky, T., 1985. Medicina traditional Chile. Editorial de la Universidad de Concepcion Chile 15 (23), 52.
Nishikawa, K., Kubota, Y., Ooi, T., 1983. Classification of proteins into groups based on amino acid composition and other characters. II. Grouping into four types. J. Biochem. 94 (3), 997–1007.
Pailer, M., 1960. Naturally occurring nitrogen compounds. Fortschr. Chem. Org. Naturstoffe 18, 55–82.
Pieper, U., Webb, B.M., Dong, G.Q., Schneidman-Duhovny, D., Fan, H., Kim, S.J., Khuri, N., Spill, Y.G., Weinkam, P., Hammel, M., Tainer, J.A., Nilges, M., Sali, A., 2014. ModBase, a database of annotated comparative protein structure models and associated resources. Nucleic Acids Res. 42, D336–D346 Database issue.
Pons, J.L., Labesse, G., 2009. @TOME-2: a new pipeline for comparative modeling of protein-ligand complexes. Nucleic Acids Res. 37, W485–W491 Web Server issue.
Reinisch, K.M., Chen, L., Verdine, G.L., Lipscomb, W.N., 1995. The crystal structure of HaeIII methyltransferase convalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. Cell 82 (1), 143–153.
Rose, P.W., Prlić, A., Bi, C., Bluhm, W.F., Christie, C.H., Dutta, S., Green, R.K., Goodsell, D.S., Westbrook, J.D., Woo, J., Young, J., Zardecki, C., Berman, H.M., Bourne, P.E., Burley, S.K., 2015. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic Acids Res. 43, D345–D356 Database issue.
Samanani, N., Alcantara, J., Bourgault, R., Zulak, K.G., Facchini, P.J., 2006. The role of phloem sieve elements and laticifers in the biosynthesis and accumulation of alkaloids in opium poppy. Plant J. 47 (4), 547–563.
Samanani, N., Park, S.U., Facchini, P.J., 2005. Cell type-specific localization of transcripts encoding nine consecutive enzymes involved in protoberberine alkaloid biosynthesis. Plant Cell 17 (3), 915–926.
Sanner, M.F., 1999. Python: a programming language for software integration and development. J. Mol. Graph Model. 17 (1), 57–61.
Schluckebier, G., O'Gara, M., Saenger, W., Cheng, X., 1995. Universal catalytic domain structure of AdoMet-dependent methyltransferases. J. Mol. Biol. 247 (1), 16–20.
Schütte, H., Orban, U., Mothes, K., 1967. Biosynthesis of aristolochic acid. Europ. J. Biochem. 1 (1), 70–72.
Sharma, V., Jain, S., Bhakuni, D., Kapil, R., 1982. Biosynthesis of aristolochic acid. J. Chem. Soc. Perkin Trans. 1 (5), 1153–1155.
Shen, H.B., Chou, K.C., 2007. EzyPred: a top-down approach for predicting enzyme functional classes and subclasses. Biochem. Biophys. Res. Commun. 364 (1), 53–59.
Sippl, M.J., 1993. Recognition of errors in three-dimensional structures of proteins. Proteins 17 (4), 355–362.
Skolnick, J., Fetrow, J.S., Kolinski, A., 2000. Structural genomics and its importance for gene function analysis. Nat. Biotechnol. 18, 283–287.
Ta, H.M., Kim, K.K., 2010. Crystal structure of *Streptococcus pneumoniae* Sp1610, a putative tRNA methyltransferase, in complex with S-adenosyl-*L*-methionine. Prot. Sci. 19 (3), 617–624.
Takata, Y., Huang, Y., Komoto, J., Yamada, T., Konishi, K., Ogawa, H., Gomi, T., Fujioka, M., Takusagawa, F., 2003. Catalytic mechanism of glycine *N*-methyltransferase. Biochemistry 42 (28), 8394–8402.
Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24 (8), 1596–1599.
Vidgren, J., Svensson, L.A., Liljas, A., 1994. Crystal structure of catechol *O*-methyltransferase. Nature 368 (6469), 354–358.
Wada, K., Yamaguchi, H., Harada, J., Niimi, K., Osumi, S., Saga, Y., Oh-Oka, H., Tamiaki, H., Fukuyama, K., 2006. Crystal structures of BchU, a methyltransferase involved in bacteriochlorophyll c biosynthesis, and its complex with S-adenosylhomocysteine: implications for reaction mechanism. J. Mol. Biol. 360 (4), 839–849.
Wallace, A.C., Laskowski, R.A., Thornton, J.M., 1995. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. 8 (2), 127–134.
Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., Williams, K.L., Hochstrasser, D.F., 1996. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Biotechnology (N Y) 14 (1), 61–65.
Yang, L., Zhu, J., Sun, C., Denga, Z., Qu, X., 2020. Biosynthesis of plant tetrahydroisoquinoline alkaloids through an imine reductase route. Chem. Sci. https://doi.org/10.1039/C9SC03773J. [Advance Article].
Zubieta, C., Ross, J.R., Koscheski, P., Yang, Y., Pichersky, E., Noel, J.P., 2003. Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family. Plant Cell 15 (8), 1704–1716.