Pak. J. Bot., 48(1): 231-239, 2016.

MUTATIONAL, PHYLOGENY AND EVOLUTION ANALYSES OF *SALVIA* COPALYL DIPHOSPHATE SYNTHASE

DA CHENG HAO[1*], RAMESHA B. THIMMAPPA[2] AND PEI GEN XIAO[3]

[1]*Biotechnology Institute, School of Environment, Dalian Jiaotong University, Dalian 116028, China*
[2]*Department of Metabolic Biology, John Innes Centre, Norwich NR4 7UH, UK*
[3]*Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Beijing 100193, China*
[*]*Corresponding author's email: hao@djtu.edu.cn*

Abstract

The cyclization of geranylgeranyl diphosphate (GGPP) is catalyzed by copalyl diphosphate synthase (CPS), a class II diterpene synthase (diTPS), to form copalyl diphosphate (CPP), which is an essential substrate of a variety of diterpenes in secondary metabolism of angiosperm including *Salvia* medicinal plants. The protein environment of the N-terminal class II active site stabilizes the carbocation intermediates and maintains the catalytic activity of angiosperm class II diTPS. The virtual modeling and mutagenesis of the class II diTPS of *Salvia miltiorrhiza* (SmCPS) were accomplished to illuminate the catalytic activity of SmCPS. Terminal truncations and point mutations established the role of the βγ domain and α domain, i.e., they facilitate the flexible conformational change of the class II active site after substrate binding. E203 and K238 in the N-terγ domain of SmCPS1are functional in the substrate binding and conformational transition and might be essential in catalysis. Similar to other CPSs, the ensuing protonation of the GGPP substrate and coordination of the diphosphate group are governed by highly conserved residues in the DxDD motif of SmCPS, e.g., D372 of CPS1. Moreover, F256 andY505stabilize the carbocation and control the enzymatic activity during CPP formation. The amino acids of the predicted active sites, despite under purifying selection, vary greatly, corresponding to the functional flexibility of angiosperm CPSs. Molecular phylogeny and evolution analyses suggest early and ongoing evolution of labdane-related diterpenoid metabolism in angiosperm.

Key words: Copalyl diphosphate synthase, Mutagenesis, *Salvia*, Molecular phylogeny, Evolution.

Abbreviations: CPP, copalyl diphosphate; CPS, CPP synthase; Ta, wheat (*Triticum aestivum*); Hv, barley (*Hordeum vulgare*); Os, rice (*Oryza sativa*); At, *Arabidopsis thaliana*; KS, ent-kaurene synthase; AgAS, grand fir (*Abies grandis*) abietadiene synthase; GC-MS, gas chromatography with mass spectrometric detection; GGPP, ((E,E,E)-geranylgeranyl diphosphate; PBS, phosphate buffered saline.

Introduction

Salvia miltiorrhiza (red sage; Sm) is a highly valuable traditional Chinese medicinal plant of the family Lamiaceae like many other species reported as highly important (Shinwari, 2010; Shinwari *et al*., 2011) which synthesizes the abietane-type norditerpenoidquinone, i.e., tanshinones (Gao *et al*., 2009), which is abundant in the root of Sm and has been found to have various bioactivities, including antibacterial, anti-inflammatory and anticancer activities. The cyclization of geranylgeranyl diphosphate (GGPP) is catalyzed by the class II diterpene synthase (diTPS), e.g., Sm copalyl diphosphate synthase 1 (CPS1), to form copalyl diphosphate (CPP) of normal stereochemistry, which is an intermediate in the formation of the abietane-type diterpene miltiradiene and various tanshinones. While ent-CPP is a halfway product of the biosynthetic pathway of gibberellins, some of which have been characterized (Nakagiri *et al*., 2005; Koksal *et al*., 2011), SmCPS1 is the first normal CPP specific CPS identified from angiosperm (Gao *et al*., 2009). In addition, SmCPSs 2-5 were predicted from the unpublished Sm genomic sequences (Ma *et al*., 2012), but has not been fully characterized.

The *Arabidopsis thaliana* (At) ent-CPS has three α-helical domains (α, β and γ), similar to the related diterpene cyclase taxadiene synthase (Koksal *et al*., 2011). However, active sites of ent-CPS are at the crossing point of the βγ domains, in contrast with those in the α domain of taxadiene synthase. Modular domain style of plant diTPSs allows the development of alternate active sites and chemical tactics for catalyzing isoprenoid cyclization. A further work (1.55 Å-resolution) illuminates structure-function associations which were equivocal in the 2.25 Å-resolution structure (Koksal *et al*., 2014). The positions of the diphosphate group, the tertiary ammonium cation of the substrate, and the extensive solvent structure, are well described. However, the crystal structure of normal CPS has not been reported for any plant, nor has that of any Lamiaceae CPS. Besides SmCPS1, *Isodoneriocalyx*CPS1 and CPS2 are the only Lamiaceae CPSs that catalyze the formation of CPP (Li *et al*., 2012), but the stereochemistry of this product is unknown.

In the present study, the qualified structural and functional investigation of SmCPS, and the homology modeling, were performed, while the molecular substrate docking, truncation, and site directed mutagenesis, were also accomplished to obtain a deeper understanding of structure-activity relationship of class II diTPSs of secondary metabolism. Molecular phylogeny and evolutionary analyses were performed to envision the expansion and functional divergence of CPSs.

Materials and Methods

***Salvia* materials and their treatment:** Seeds of *S. miltiorrhiza* were obtained from Chinese Academy of Medical Sciences. Plants were grown in a greenhouse at the John Innes Center, UK, at 21°C with 16 h of supplemental light. MeJA of 0.75 mM was sprayed onto

the leaves and poured into root-containing soil. Eight-week old sprouts were treated for 0, 10, 26, and 73 h or six days before leaves and roots were gathered discretely. Leaves/roots of two plantlets were collected together, which were subjected to RNA extraction instantly.

RNA isolation, cDNA synthesis and CPS gene amplification: Total RNA was extracted from *Salvia* leaves and roots by the RNeasy Mini Kit (Qiagen, Germany). TURBO™ DNase (Ambion, USA) was used to eliminate genomic DNA from the extracted RNA sample, followed by the RNA check on a 1% agarose gel. RNA concentration was defined by the Nano Drop 2000C spectrophotometer (Thermo Scientific, USA). RNA with $OD_{260/280}$ 1.8~2.2 and $OD_{260/230}$>1.8 was used in cDNA synthesis and quantitative real-time reverse transcription (qRT) -PCR. Superscript II reverse transcriptase (Invitrogen) was used to synthesize the first strand cDNA from total RNA. The *S. miltiorrhiza* CPS1 gene was amplified using primers FLf and FLr (sequences available upon request) under the following conditions: 98°C 30 s, 38 cycles of [98°C 10 s, 59°C 60 s, 72°C 2.4 min], 72°C 10 min. The annealing temperature and the extension time at 72°C for CPS3 were 52°C and 2 min 10 s respectively. The PCR products resolved by agarose gel electrophoresis were purified and transferred into the pCR-Blunt II-TOPO vector (Invitrogen).

Preparation of protein constructs and site directed mutagenesis: CPS gene was transferred into the Gateway entry vector pDONR207 (Invitrogen) via BP reaction. Site directed mutagenesis was performed according to the mutagenesis procedure (Liu and Naismith, 2008), based on the pseudomature CPS1 beginning at MPQVH after removing the plastidial transit peptide. The mutagenesis primers were designed to promote primer-template annealing and excluding primer dimerization. The newly synthesized DNA was the template in successive PCR cycles. The 50 µl reaction mix had 10 ng of pDONR-CPS1 (template), 1 µM primer pair, 200 µM dNTPs and 1 u of Phusion DNA polymerase. The PCR cycles: 94°C 2 min, followed by 20 PCR cycles. Each cycle: 98°C 10sec, $T_{m\ no}$ (calculated based on the primer sequence annealed to the template) -5°C 40 sec, 72°C 3 min. The cycles were finished with 72°C 10 min. The purified PCR products were digested with 20 u *Dpn* I at 37°C 1.5 hr. An aliquot of 6 µl PCR products was transformed into *E. coli* DH5α competent cells (Invitrogen) via heat shock. The transformed cells were spread on a LB plate having 50µg/ml gentamycin and incubated at 37°C overnight. Plasmid DNA was isolated from colony and sent for DNA sequencing to check the desired mutation. The mutated CPS DNAs were transferred into the Gateway expression vector pH9GW (Yu and Liu, 2006) via LR reaction. N-ter and C-ter truncations were obtained via PCR using CPS as the template.

Expression and purification of recombinant CPS: The *E. coli* C41(DE3) (Lucigen) containing SmCPS in pH9GW was grown to A_{600} = 0.5 at 37°C and 200rpm in 100 ml LB medium and 50 µg/ml kanamycin. IPTG (final concentration 0.5 mM) was supplemented and the bacteria were cultured for a further 16 h at 16°C and 200 rpm. The cells were precipitated by centrifugation (8000g, 10min, 4°C) and resuspended in lysis buffer (30 ml/L culture, PBS pH 7.4, 100 mM NaCl, 10% glycerol, 5 mM β-ME, 0.5 mg/ml lysozyme). The cells were disrupted by a diagenode sonicator (M level, 1×180s) in the ice-cold lysis buffer, and the rough lysate was obtained by centrifugation (20,000 g, 20 min, 4°C). The lysate was purified by Ni-NTA affinity chromatography (Qiagen). The eluted protein-containing portions were pooled and desalted on the Biomax-10 membrane (Millipore). At ent-CPS was expressed with pGGeC plasmid as previously described (Cyr *et al*., 2007).

CPS enzyme assay: The 500 µl assay buffer consisted of PBS pH 7.4, 10 mM β-ME, 5 mM $MgCl_2$, 5% [v/v] glycerol, 88.8 µM GGPP and 34 µg purified CPS, which was covered by 500 µl hexane to separate the reaction products. After 1 h incubation at 30°C, the mixture was hydrolyzed by 50 u calf intestinal alkaline phosphatase (Promega) for 3 h at 37°C and pH 9.3. Hexane (3×500 µl) was used to extract the mixture. The pooled hexane phase was decreased to 200 µl under nitrogen and evaluated in GC–MS.

GC–MS enquiry: GC analyses were undertaken on the Agilent (Palo Alto, CA) 6890N GC instrument by using a ZB5-HT column with a 5973 mass selective detector under the parameters: EI 70eV, source 200°C; injection 250°C, interface 300°C; flow speed of He 1.0 ml/min, constant flow; splitless injection, column temperature: 40°C 3 min, then increased to 300°C at a speed of 20°C/ min and then 300°C 3 min. The EI-MS spectra of all products were compared with those of the NIST library (V1.6d) or bought genuine standards.

Analysis of inducible CPS gene transcription: DNase-treated RNA was utilized to create first strand cDNA by SuperScript II reverse transcriptase (200 u/µl; Invitrogen). The qRT-PCR reactions were prepared using SYBR® Green JumpStart™ TaqReadyMix™ (Sigma-Aldrich), and were performed in three technical replicates in the CFX96 Real-Time PCR Detection System (Bio-Rad) using 0.2 ml non-skirted low profile 96-well PCR plates (Thermo Scientific) and PCR sealers Microseal B film (Bio-Rad). qRT-PCRs were performed as follows: 95ºC 2 min, 43 cycles of 95ºC 10 sec, 60ºC 40 sec, and 72ºC 15 sec. After PCR, the melting curve was generated by 65ºC 5 sec and increasing to 95ºC to examine the amplicon specificity. The qRT-PCR primer information is available upon request. CFX Manager software (Bio-Rad) was used for data analysis. This test was duplicated thrice with individually grown plantlets.

Circular dichroism (CD) spectroscopy: CD spectra were obtained by the Chirascan-plus spectropolarimeter (Applied Photophysics, UK) using purified CPSs at 0.2 mg/ml in PBS without Cl^-, pH 7.6. Data acquisition was performed in the quartz cuvette with a 0.5-mm path. To acquire whole CD spectra, wavelength scans were collected at 20°C with 30 nm/min scanning speed, 2 nm bandwidth, 0.5 nm step size and a range between 190 and 260 nm. Thermal melt curves were obtained at 1°C intervals by increasing the temperature from 20 to 90°C with a speed of 1°C/min. The millidegree-unit original

records between 201 and 260 nm were evaluated by the Global 3 software (Applied photophysics), which fits the full-spectrum data to create a comprehensive analysis of CPS unfolding, and defines transition temperatures (melting points: T_m) and their associated Van't Hoff enthalpies (ΔH) (Walden *et al*., 2014).

Computational structure analysis: Modelling was done in Phyre2 (Kelley and Sternberg, 2009). Among a list of models obtained in the server, ones that have been modelled using At-CPS as template were chosen for further docking analysis. In docking GGPP was used as a ligand. Both the protein models and GGPP ligand co-ordinates were submitted in SWISSDOCK server. From this, we got around 256 different docked poses for each of the protein. All these 256 docked poses were individually examined in CHIMERA (http://www.cgl.ucsf.edu/chimera) to select ones that match the conformation as seen in At-CPS-AG8 structure (Koksal *et al*., 2014).

Phylogenetic and evolutionary analysis: CPS nucleotide coding sequence and amino acid sequence were retrieved from NCBI by: 1. searching "copalyl diphosphate synthase" in GenBank; 2. searching database Transcriptome Shotgun Assembly using Tblastn (search translated nucleotide databases using a protein query). More CPS sequences were obtained from Ensembl (http://plants.ensembl.org/index.html). Decrease redundancy (http://web.expasy.org/decrease_redundancy/) was used to reduce the redundancy of CPS sequences. The CPS sequence alignment was constructed using MULTALIN (Combet *et al*., 2000). The domain boundary of SmCPS1 was inferred according to the alignment between Atent-CPS (Koksal *et al*., 2011) and SmCPS1. The MEGA6 software (Tamura *et al*., 2013) was used to construct the phylogenetic tree, which was based on the amino acid sequence alignment and the neighbor joining method. The CPS codon sequence alignment was analyzed on Selecton server to identify site-specific positive and negative selections (Stern *et al*., 2007).

Results and Discussion

Structural modeling of SmCPS: SmCPS1 (NCBI GenBankEU003997) (Cui *et al*., 2011) and CPS3 (GenBankKJ461673 from this study) were modeled in the predicted mature forms without transit peptides. In models CPS1GG1, 1.205 and CP3GG1, 1.200 we could see C14-15 double bond of GGPP being nearer to DIDD and EVDD motifs respectively (Fig. 1), indicating these are appropriate docked conformations and are comparable to At-CPS (Koksal *et al*., 2014). The structures displayed the characteristicα–helical TPS fold, consisting of the βγ and α domains. Both enzymes possess the D (E in CPS3) xDD motif, which is essential for the protonation-dependent cyclization of GGPP.

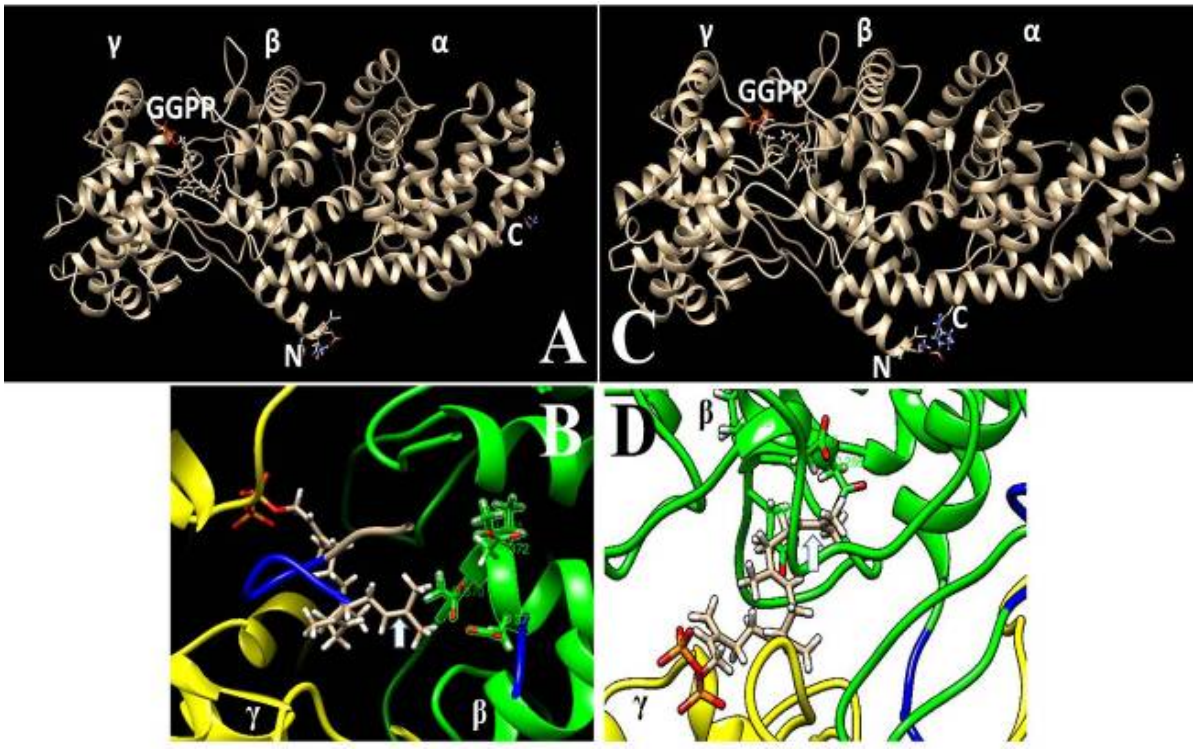

Fig. 1. Homology modeling of *Salvia* CPSs with the docked substrate GGPP. A. Overall structure of the CPS1/GGPP complex; the N- and C-termini are indicated by "N" and "C" respectively; GGPP is shown as a stick figure bound in the active site at the interface of the and domains. B. zoomed-in view showing the C14-15 double bond of GGPP is nearer to the DIDD motif (aa 370-373) of CPS1;β domain = green, γ domain = yellow; white arrow denotes the C14-15 double bond of GGPP and is nearer to the protonating aspartate (D). C. Overall structure of theCPS3/GGPP complex. D. zoomed-in view showing the C14-15 double bond of GGPP is nearer to the EVDD motif (aa 285-288) of CPS3.

Amplification of CPS genes and induction of gene transcription: The gene sequence of SmCPS1 amplified from Sm cDNA was identical to EU003997 (Cui *et al*., 2011). The amplified CPS3 (KJ461673), without plastid transit peptide, was seven amino acids longer than JN831115, which was predicted from unpublished draft genome of Sm (Ma *et al*., 2012). Their amino acid sequences are 95% identical.

The basal expression level of CPS1 and CPS3 genes in the rootwas much higher than that in the leaf (Fig. 2A). The basal expression of CPS3 gene was extremely low and it could be regarded as the root specific. After 0.75mM MeJA treatment, mRNA levels of some CPSs were induced and quantified by qRT-PCR. CPS1 expression was induced significantly in the root but not in the leaf, while CPS3 expression had the most dramatic upregulation in both root and leaf. CPS4 expression was inducible in both tissues, while CPS2 was not inducible in the root and CPS5 (ent-CPS) was not inducible in the leaf.

***E. coli* expression of CPS enzymes and their catalytic activity:** The first 63 aa of CPS1 correspond to the plastid transit peptide and thus Δ63 was expressed as the pseudomature form (Fig. 2B). CPS1 Δ79, Δ100 and CΔ16 (C-terminal truncation of 16 aa) were well expressed.CPS3 expression was obtained in both soluble and insoluble parts. All site-directed mutants were expressed well in *E. coli*.

The *In vitro* CPS reaction product CPP ($C_{20}H_{33}O_7P_2$; MW 450.44) was dephosphorylated to copalol for GC-MS analysis (Fig. 2C, D). The peak area of the 275 characteristic MS ion of copalol ($C_{20}H_{34}O$; MW 290.48), after calibrated with that of the characteristic ion of the internal standard ergosterol, was used to represent the product yield and to compare the catalytic activity of the wild type (WT Δ63) and the mutant (Fig. 2E and Table 1). Compared to Δ63, Δ79 showed the similar level of product formation (Table 1). In contrast, Δ100 and CΔ16 showed no product formation, suggesting that aa 80-100 of β domain (Q76-I103 and P318-Q527), as well as the C-terminal 16 aa of α domain (A528-A793), are essential for the class II catalysis in SmCPS1. The β domain is part of the class II active site in the closed conformation, while the C-terminal of α domain might facilitate the shutting of the class II active site. Interestingly, most residues of these two parts are subject to the purifying selection.

With the aims of investigating the function of amino acids of highly conserved sites and predicted active sites, we made site-specific CPS mutants. CD spectra showed no significant change in secondary structure when comparing WT and mutant protein (Fig. 3), and the thermal stability is not substantially different between WT and mutant (Table available upon request), indicating that alterations of catalytic activity can be ascribed to the specific point mutations. Thirteen residues were predicted by I-TASSER server (Roy *et al*., 2010) as the substrate binding site. Eleven recombinant CPSs were obtained to characterize the functional consequence of mutations on these residues (Fig. 2E). S362W, D372N, D373-(deletion) andK457W/W458K double mutant abolished the catalytic activity of each CPS mutant respectively, and thus no product was detected in GC-MS.D372of the DXDD motif is the general acid responsible for substrate protonation (Koksal *et al*., 2014). D372 and D373 are strictly conserved in all CPSs (Fig. 4).

C200G/G201C double mutant and K238Q almost eliminated the enzyme activity (5.39% and 5.02% product formation compared to WT, respectively).The diphosphate binding region is rather basic largely due to the conserved "basic segment" V234-R241 (Fig. 4). These basic residues could interact with the substrate diphosphate when bound in a fully productive conformation. In 1.55 Å-resolution structure of the At CPS-substrate complex, the substrate diphosphate group accepts hydrogen bonds directly from conserved K245 (K238 in SmCPS1) and K463 (K457 in SmCPS1) (Koksal *et al*., 2014). F256H, Y505H and D370E significantly decreased the SmCPS1 activity (18.43%, 26.29% and 32.36% product formation respectively). In contrast, I411T increased the CPS product formation (161%). Of note, this site has the highest K_a/K_s value (0.37, confidence interval 0.28-0.62) among the predicted active sites, and the amino acid at this site is variable (Fig. 4), implying that it is under relaxed purifying selection (Figure available upon request). D370 is not consistent in all CPSs, e.g., D is replaced by E in SmCPS3 and maize ent-CPS (Fig. 4), implying that the negatively charged side chain was selected during long term evolution. This site has a much lower ω value (0.065, CI 0.036-0.13) than I411 and is subject to strong purifying selection. The residue of F256 is capricious among the angiosperm CPSs, e.g., the corresponding residues in *Salvia sclarea* LPPS and wheat CPS1 are a tyrosine and a histidine respectively (Fig. 4), indicating that the function of the side chain in F256 might be the particular conformational limitation of the carbocation intermediates.

Thirteen recombinant CPSs were heterologously expressed to characterize the functional consequence of mutations of nine highly conserved residues (Table 1). Substitution of W266 to alanine, D141 to asparagine, E203 to glutamine and W579 to isoleucine severely compromised CPP formation. Substitution of W266 to leucine and Y108 to phenylalanine substantially decreased CPP production. On the contrary, replacement of 324R with histidine and M507 with leucine had trivial influence on product formation. W266 and W579 located distant from GGPP might have an ancillary outcome on the carbocation maintenance. Substitutions of W266 with phenylalanine, leucine or alanine caused 31%, 76% or 97% losses of product, and substitution of W579 with isoleucine caused 88% reduction of CPP, suggesting that it could be the global hydrophobicity and dimension of the side chain that are critical in these two positions.

A divalent cation important for CPS activity may play a role in substrate binding, e.g., for binding the GGPP diphosphate group. Possibly, E203 (corresponding to E211 of At ent-CPS; Fig. 4) could be part of the putative metal-binding site in CPS (Koksal *et al*., 2011). Exchange of E203, with the negatively charged side chain, to Q (with uncharged R group) might eradicate the binding between E and the catalytically obligatory Mg^{2+}, thus interfering with the coordination between the GGPP diphosphate group and Mg^{2+}. R386 (R379 in SmCPS1) participates in the formation of the hydrogen bond networks with general acid D379 (Koksal *et al*., 2014). Substitution of R379 to non-functional alanine might impair hydrogen-bonded solvent channel leading to D372 of SmCPS1 and negatively impact proton transfer to and from D372 upon catalysis.

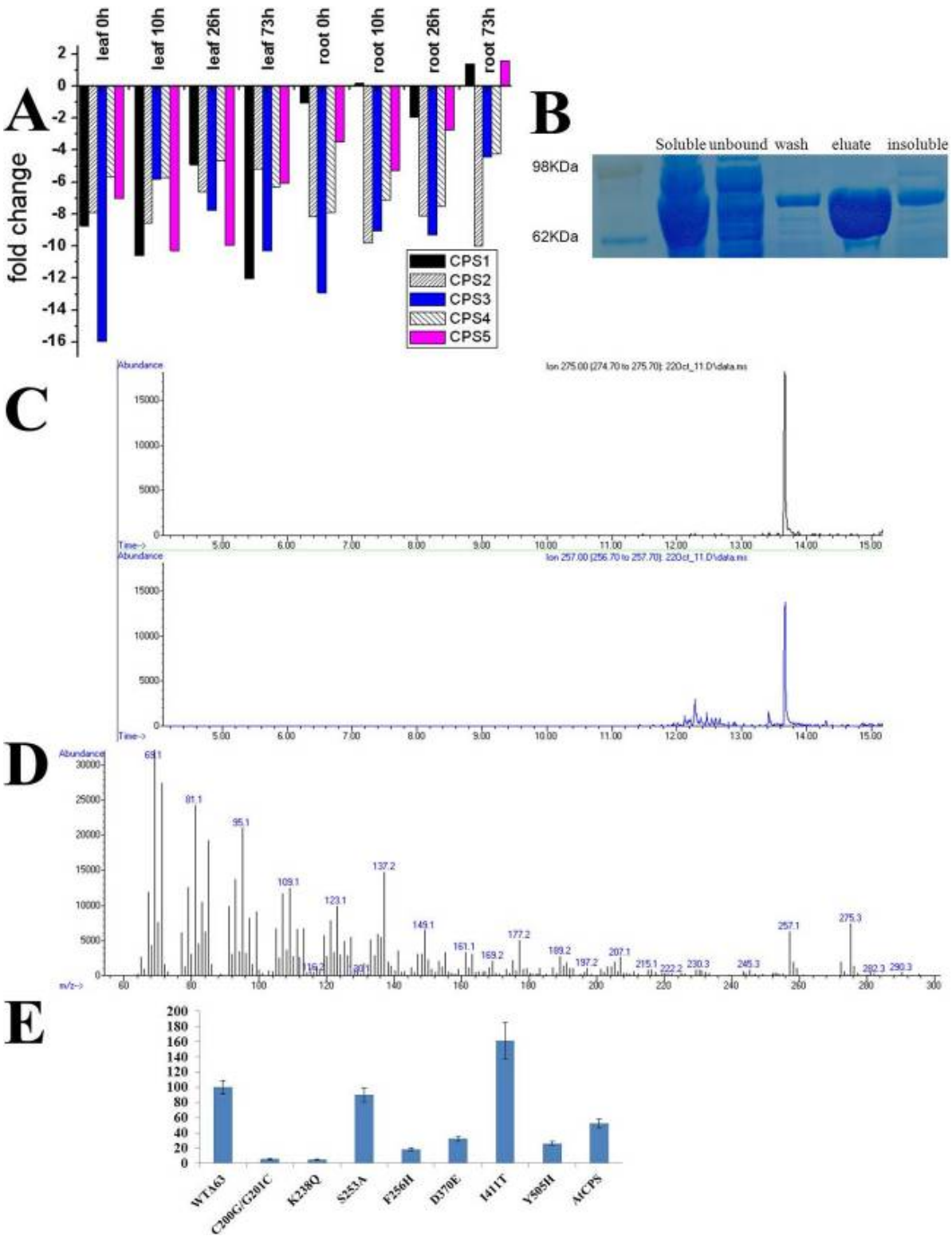

Fig. 2A. Tissue-specific expression of *Salvia* CPSs in response to MeJA determined by qRT-PCR. Analysis of biological replicates demonstrates the same overall expression pattern. B. SDS-PAGE analysis of affinity-purified His-tagged CPSs. GC-MS analysis of the dephosphorylated reaction product formed by CPS1 wild type and mutant respectively, with GGPP as substrate. C. 275 (copalol) and 257(copalol) m/z extracted ion chromatograms. D. The corresponding mass spectra show the product peak (retention time 13.7 min) matching copalol in mass fragmentation patterns included in the database. E. Product yield of mutants of predicted SmCPS1active sites (not including conserved sites outside of catalytic center). S362W, D372N, D373- (deletion) and K457W/W458K double mutant abolished the catalytic activity of each CPS mutant respectively, and thus are not shown. Bars represent the standard error.

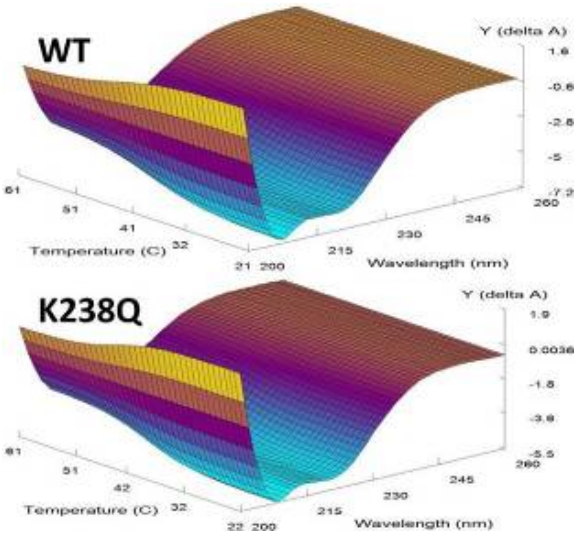

Fig. 3. Far-UV CD spectrum of wild type and variants of SmCPS1. The temperature-wavelength surface is a 3-D representation of the calculated property. Each of the variants adopts an essentially identical structure to the wild type protein, as assayed by Far-UV CD.ΔA, difference in absorption.

The corresponding position of SmCPS1 R324 is regarded as a regulatory switch in monocot CPSs (Wu *et al*., 2012), as the residue is arginine in CPSs operating in phytoalexin metabolism but it is histidine in those devoted to gibberellin biosynthesis. However, in the present study substitution of R324 to histidine had no effect on product profile and substitution to glutamine showed a 44% reduced production of normal CPP (Table 1), implying that the stereochemical diversity arose separately in monocot and dicot.

The recombinant SmCPS3 failed to produce detectable amount of CPP or other relevant compound, although the inducible CPS3 gene expression by MeJA (Fig. 2A) implies that CPS3 might be involved in secondary metabolism. There are at least two anomalous amino acids, i.e., A333 (N425 in At ent-CPS) and T411 (D503 in At ent-CPS), in CPS3 sequence (Fig. 4). In At ent-CPS, general acid D379 is stabilized by a hydrogen bond with N425 (Koksal *et al*., 2011), which is conserved in all CPSs except SmCPS3. The catalytic activity of N425A At CPS shows a 13-fold decrease in k_{cat} (Koksal *et al*., 2014). D503, conserved in all CPSs except SmCPS3, is a key residue of the hydrogen bond network linking D379, solvent channel and the solvent-open surface of the β domain adjoining the boundary with the γ domain. Threonine at this position might abrogate the active site solvent network and abolish the catalysis.

Molecular phylogeny and evolution of CPS: A wide range of characterized and predicted CPSs are included in the phylogenetic tree (Fig. 5). Fungi CPSs are at the most basal part, followed by dicot, gymnosperm, fern and monocot. Both monocot and dicot CPSs are undergoing rapid radiation and multiple gene duplications. Gene duplication and functional divergence might occur in the common ancestor of dicot CPSs, as well as within the respective species. SmCPS2 is closest to four *S. sclarea* diTPSs. The cyclization of GGPP to (8R)-hydroxy-CPP is catalyzed by SsTPS1 (Gunnewich *et al*., 2013). SsLPPS (NCBI accession AFU61897) yielded labda-13-en-8-ol diphosphateand non-hydroxylated analogue, (9 S, 10 S)-CPP, as the major and minor products from GGPP, respectively (Caniard *et al*., 2012). The other two LPPSs produced 13-labden-8, 15-diol pyrophosphate (Schalk *et al*., 2012). The above five are closely related to the normal CPP specific SmCPS1, followed by putatively nonfunctional SmCPS3. The above CPSs are sister to a four-member cluster, in which only tobacco CPS2 is known to synthesize 8-hydroxy-CPP (Sallaud *et al*., 2012). In contrast, other SmCPSs are grouped in another major branch of dicots (Fig. 5). Sm ent-CPS is closest to *Isodon eriocalyx* CPS1, which might be more relevant to gibberellin formation but also participate in ent-kaurane production (Li *et al*., 2012). These two are closely related to *Andrographis paniculata* ent-CPS, followed by *Scoparia dulcis* ent-CPS (Nakagiri *et al*., 2005). The above four CPSs are sister to SmCPS4 and *I. eriocalyx* CPS2, which might be involved in the biosynthesis of pharmaceutically useful diterpenoids (Li *et al*., 2012). Taken together, these results suggest both early and continued CPS expansion and functional specification. These results also suggest that it is not reliable to infer the CPS function with regard to stereochemistry simply based on the amino acid sequence relatedness.

Table 1. Product yield analysis of SmCPS1 mutants of highly conserved sites (not including active sites).

	Domain that the mutation lies in	Average 275[a] peak area (13.7 min)	Average 253[b] peak area (17.789 min)	Normalized peak area	Product percentage
WT Δ63		36285	141713	36285	100
Δ79		27901	163236	32138.53	88.57
Y108F	γ	12554	148854	13186.6	36.34
D141N	γ	2147	136199	2063.46	5.68
E203Q	γ	4278	138291	4174.69	11.50
W266A	γ	1319	136414	1269.67	3.49
W266F	γ	24477	145134	25067.88	69.08
W266L	γ	7999	155750	8791.31	24.22
R324H	β	31841	155886	35025.48	96.52
R324Q	β	19251	148622	20189.55	55.64
R379A	β	27915	118118	23267.19	64.12
A423S	β	20974	135152	20002.95	55.12
M507A	β	17011	150326	18044.89	49.73
M507L	β	16950	377542	17702.76	90.86
W579I	α	4301	146943	4459.73	12.29

Data from a representative experiment (run in duplicate) are shown
[a]the characteristic MS ion of copalol; [b]the characteristic ion of the internal standardergosterol ($C_{28}H_{44}O$, MW 396.65)

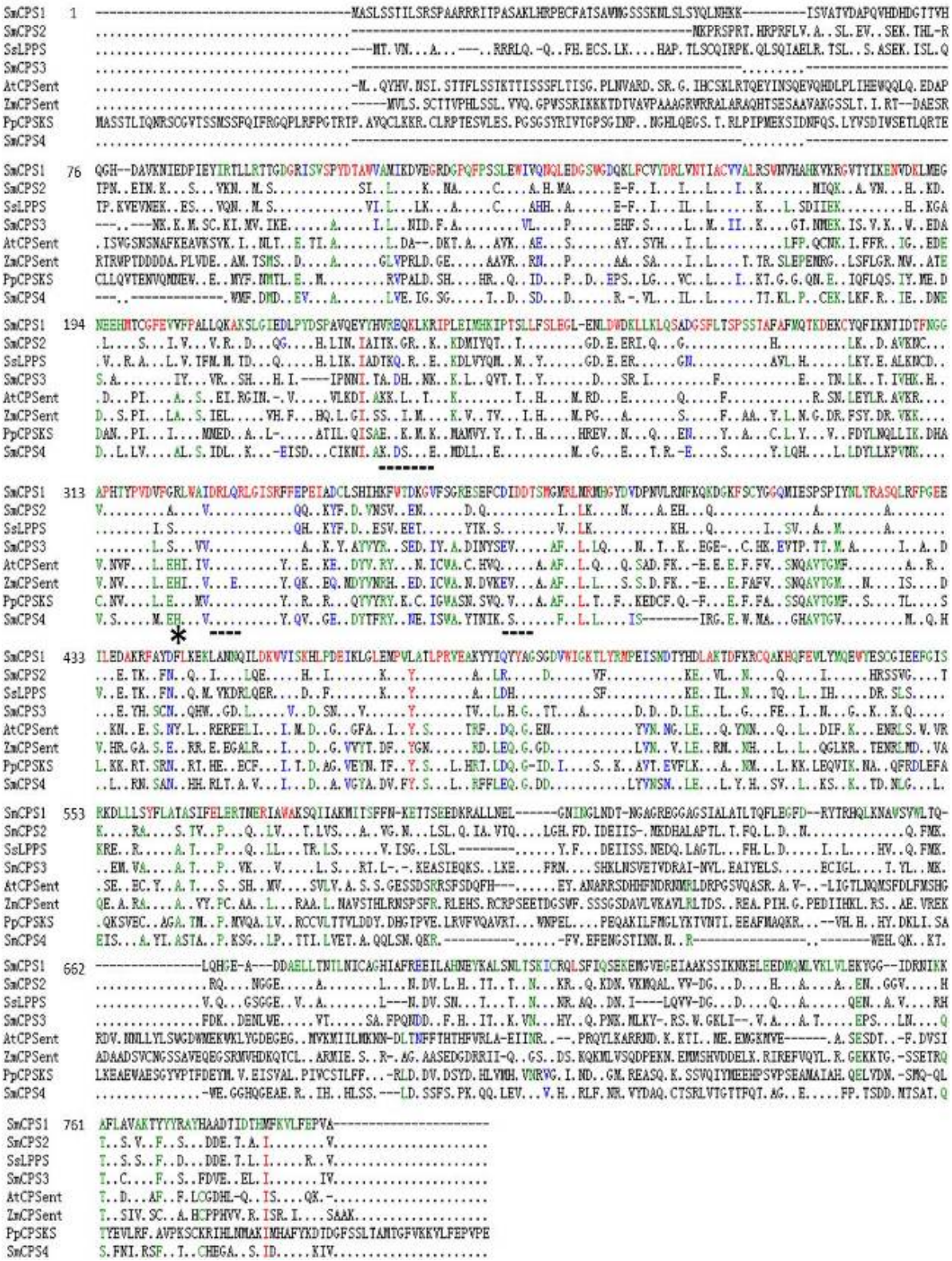

Fig. 4. Amino acid sequence alignment of eight CPSs, with the catalytic "DXDD" motif (SmCPS1 aa 370-373), "basic segment" (V234-R241) and solvent channel (D329-R333) underlined and the '*' underneath the regulatory switch position described in the text. Residues that are not identical to those of SmCPS1 are shown. Zm, *Zea mays*; Pp, *Physcomitrella patens*. S104-Y317 represent γ domain of SmCPS1. The proposed metal binding site of class II diterpene cyclase, i.e., the "EDXXD-like" motif in the γ domain (Cao *et al*., 2010), seems untenable.

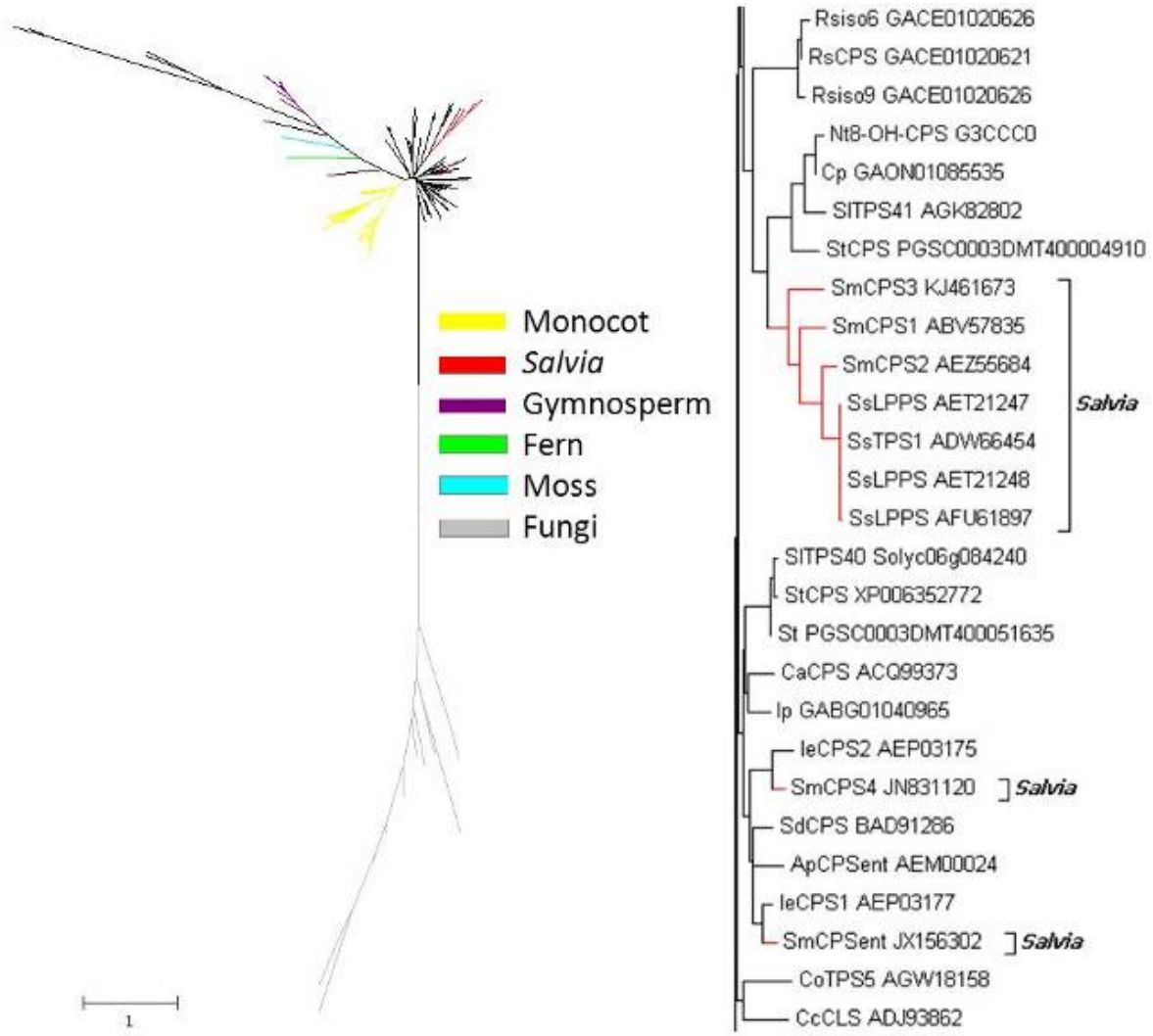

Fig. 5. Phylogenetic tree constructed from 151 amino acid sequence alignment using the Neighbor Joining method, with the fungi CPSs as the outgroup sequences. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the JTT+G model (shape parameter = 1) and are in the units of the number of amino acid substitutions per site. All gap positions were removed for each sequence pair. Bootstrap values of all major branches on the right tree (partially enlarged left tree) are well above 50%. Rs, *Rauvolfia serpentine* (Indian snakeroot); Nt, *Nicotiana tabacum* (tobacco); Cp, *Cuscuta pentagona* (a parasitic plant of the family Convolvulaceae); Sl, *Solanum lycopersicum* (tomato); St, *Solanum tuberosum* (potato); Ss, *Salvia sclarea*; Ca, *Coffea arabica*; Ip, *Ipomoea purpurea* (common morning glory); Ie, *Isodon eriocalyx* (Xiang Cha Cai); Sd, *Scoparia dulcis* (goatweed); Ap, *Andrographis paniculata* (Chuan Xin Lian); Co, *Copaifera officinalis* (legume family); Cc, *Cistus creticus subsp. Creticus* (pink rock-rose). CLS, copal-8-ol diphosphate synthase.

Conclusion

Structural modeling of SmCPS, involved in diterpene biosynthesis, identified amino acids that might be responsible for specific reaction mechanisms of the class II cyclization reaction. The N-terminal βγ domain (residues Q76–Q527) of SmCPS1 has the active site at their interface and is relevant for catalysis. Amino acids of the predicted active sites are under purifying selection. K238 and E203 are especially important for binding substrate. R379 helps maintain the highly polar hydrogen bond networks with the general acid D372.W266 and W579 located distant from GGPP might have a lesser effect on the carbocation maintenance. Molecular phylogeny and evolution analyses provide insights into the early and ongoing evolution of labdane-related diterpenoid metabolism in dicots.

Acknowledgements

This study is supported by Natural science fund of Liaoning province (2015020663), China Scholarship Council (201208210122) and the Ministry of Science and Technology national support program (No. 2012BAI29B01). We thank Dr. Lijia Xu (Institute of Medicinal Plant Development, Chinese Academy of

Medical Sciences, China) and Prof. Reuben Peters (Iowa State University) for providing *S. miltiorrhiza* seeds and pGGeC respectively. We also thank Dr. Miriam Walden (Department of Biological Chemistry, John Innes Centre) and Dr. Robert Minto (Indiana University-Purdue University) for suggestions in circular dichroism and GC-MS respectively. Some experiments were performed in Anne Osbourn's lab, John Innes Centre, UK.

References

Caniard, A., P. Zerbe, S. Legrand, A. Cohade, N. Valot, J.L. Magnard, J. Bohlmann and L. Legendre. 2012. Discovery and functional characterization of two diterpene synthases for sclareol biosynthesis in *Salvia sclarea* (L.) and their relevance for perfume manufacture. *B.M.C. Plant Biol*., 12:119.

Cao, R., Y. Zhang, F.M. Mann, C. Huang, D. Mukkamala, M.P. Hudock, M.E. Mead, S. Prisic, K. Wang, F.Y. Lin, T.K. Chang, R.J. Peters and E. Oldfield. 2010. Diterpene cyclases and the nature of the isoprene fold. *Proteins*, 78:2417-2432.

Combet, C., C. Blanchet, C. Geourjon and G. Deléage. 2000. NPS@: network protein sequence analysis. *Trends Biochem. Sci*., 25: 147-150.

Cui, G., L. Huang, X. Tang and J. Zhao. 2011. Candidate genes involved in tanshinone biosynthesis in hairy roots of *Salvia miltiorrhiza* revealed by cDNA microarray. *Mol. Bio. Rep*., 38: 2471-2478.

Cyr, A., P.R. Wilderman, M. Determan and R.J. Peters. 2007. A modular approach for facile biosynthesis of labdane-related diterpenes. *J. Am. Chem. Soc*., 129: 6684-6685.

Gao, W., M.L. Hillwig, L. Huang, G. Cui, X. Wang, J. Kong, B. Yang and R.J. Peters. 2009. A functional genomics approach to tanshinone biosynthesis provides stereochemical insights. *Org Lett*., 11:5170-5173.

Günnewich, N., Y. Higashi, X. Feng, K.B. Choi, J. Schmidt and T.M. Kutchan. 2013. A diterpene synthase from the clary sage *Salvia sclarea* catalyzes the cyclization of geranylgeranyl diphosphate to (8R)-hydroxy-copalyl diphosphate. *Phytochem*., 91:93-99.

Kelley, L.A. and M.J. Sternberg. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc*., 4: 363-371.

Köksal, M., H. Hu, R.M. Coates, R.J. Peters and D.W. Christianson. 2011. Structure and mechanism of the diterpene cyclase ent-copalyl diphosphate synthase. *Nat. Chem. Biol*., 7:431-433.

Köksal, M., K. Potter, R.J. Peters and D.W. Christianson. 2014. 1.55 Å-resolution structure of ent-copalyl diphosphate synthase and exploration of general acid function by site-directed mutagenesis. *Biochim. Biophys. Acta*., 1840:184-190.

Li, J.L., Q.Q. Chen, Q.P. Jin, J. Gao, P.J. Zhao, S. Lu and Y. Zeng. 2012. IeCPS2 is potentially involved in the biosynthesis of pharmacologically active *Isodon* diterpenoids rather than gibberellin. *Phytochem*., 76:32-39.

Liu, H. and J.H. Naismith. 2008. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *B.M.C. Biotechnol*., 8: 91.

Ma, Y., L. Yuan, B. Wu, X. Li, S. Chen and S. Lu. 2012. Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in *Salvia miltiorrhiza*. *J. Exp. Bot*., 63:2809-2823.

Nakagiri, T., J.B. Lee and T. Hayashi. 2005. cDNA cloning, functional expression and characterization of ent-copalyl diphosphate synthase from *Scoparia dulcis* L. *Plant Sci*., 169: 760-767.

Roy, A., A. Kucukural and Y. Zhang. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc*., 5:725-738.

Sallaud, C., C. Giacalone, R. Töpfer, S. Goepfert, N. Bakaher, S. Rösti andA. Tissier. 2012. Characterization of two genes for the biosynthesis of the labdanediterpene Z-abienol in tobacco (*Nicotianatabacum*) glandular trichomes. *Plant J*., 72:1-17.

Schalk, M., L. Pastore, M.A. Mirata, S. Khim, M. Schouwey, F. Deguerry, V. Pineda, L. Rocci and L. Daviet.2012. Toward a biosynthetic route to sclareol and amber odorants. *J. Am. Chem. Soc*., 134:18900-18903.

Shinwari, Z.K. 2010. Medicinal plants research in Pakistan. *J. Med. Pl. Res*., 4(3): 161-176.

Shinwari, Z.K., S. Sultan and T. Mehmood. 2011. Molecular and morphological characterization of selected *Mentha* species. *Pak. J. Bot*., 43(3): 1433-1436

Stern, A., A. Doron-Faigenboim, E. Erez, E. Martz, E. Bacharach and T. Pupko. 2007. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Res*., 35: W506-511.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Bio. Evol*., 30: 2725-2729.

Walden, M., A. Crow, M.D. Nelson and M.J. Banfield. 2014. Intramolecular isopeptide but not internal thioester bonds confer proteolytic and significant thermal stability to the *S. pyogenes* pilus adhesin Spy0125. *Proteins*, 82: 517-527.

Wu, Y., K. Zhou, T. Toyomasu, C. Sugawara, M. Oku, S. Abe, M. Usui, W. Mitsuhashi, M. Chono, P.M. Chandler and R.J. Peters. 2012. Functional characterization of wheat copalyl diphosphate synthases sheds light on the early evolution of labdane-related diterpenoid metabolism in the cereals. *Phytochem*., 84:40-46.

Yu, X.H. and C.J. Liu. 2006. Development of an analytical method for genome-wide functional identification of plant acyl-coenzyme A-dependent acyltransferases. *Anal Biochem*., 358: 146–148.

(Received for publication 20 March 2015)