MUTATIONAL, PHYLOGENY AND EVOLUTION ANALYSES OF SALVIA COPALYL DIPHOSPHATE SYNTHASE

DA CHENG HAO¹*, RAMESHA B. THIMMAPPA² AND PEI GEN XIAO³

¹Biotechnology Institute, School of Environment, Dalian Jiaotong University, Dalian 116028, China
²Department of Metabolic Biology, John Innes Centre, Norwich NR4 7UH, UK
³Institute 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-ter domain of SmCPS1 are 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 and Y505 stabilize 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 norterpenoidquinone, 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 ext-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, Isodomeriocalyx 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 NanoDrop 2000C (Thermo Scientific, USA). OD 260/280 1.8~2.2 and OD 260/230>1.8 was used in cDNA spectrophotometer (Thermo Scientific, USA). RNA with concentration was defined by the Nano Drop 2000C (Germany). TURBO™ DNase (Ambion, USA) was used to leaves and roots by the RNeasy Mini Kit (Qiagen, Germany). The eluted protein-containing portions were pooled and desalted on the Biomax-10 affinity chromatography (Qiagen). Ent-CP3 was expressed with pGGeGc plasmid as previously described (Cyr et al., 2007).

**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 μg of pDONR-CPS1 (template), 1 μM primer pair, 200 μM dNTPs and 1 μ of Phusion DNA polymerase. The PCR cycles: 94°C 2 min, followed by 20 PCR cycles. Each cycle: 98°C 10 sec, Tm 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 A600 = 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, 10 min, 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 affinity chromatography (Millipore). At ent-CP3 was expressed with pGGeGc 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 MgCl2, 5% [v/v] glycerol, 88.8 μg 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 ZBS-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: 95°C 2 min, 43 cycles of 95°C 10 sec, 60°C 40 sec, and 72°C 15 sec. After PCR, the melting curve was generated by 65°C 5 sec and increasing to 95°C 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 250 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 (\( \Delta H \)) (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 (Comb 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 GenBank EU003997) (Cui et al., 2011) and CPS3 (GenBank KJ461673 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 characteristic α-helical TPS fold, consisting of the \( \beta \gamma \) and \( \alpha \) 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 \( \alpha \) and \( \gamma \) domains. B. zoomed-in view showing the C14-15 double bond of GGPP is nearer to the DIDD motif (aa 370-373) of CPS1\( \beta \) domain = green, \( \gamma \) domain = yellow; white arrow denotes the C14-15 double bond of GGPP and is nearer to the protonating aspartate (D). C. Overall structure of the CPS3/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 peptidase, 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 root was 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, A100 and CA16 (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 637) was dephosphorylated to copalol for GC-MS analysis (Fig. 2C, D). The peak area of the 275 MW 450.44) was dephosphorylated to copalol for GC-MS. D372 and D373 are strictly conserved in all CPSs (Fig. 4). The DXDD motif is the general acid responsible for substrate binding (Koksal et al., 2014). D372 and D373 are also conserved 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 α 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 carboxylation 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 SmCPS1 active 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.
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 γ domain adjoining the boundary with the β domain. Thrreonine 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). SaLPPS (NCBI accession AFU61897) yielded labda-13-ene-8-sol diphosphate and 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 (Sialauda eriocalyx ent-CPS, as the major and minor products from GGPP, respectively (Caniard 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 ent-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 ent-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).

<table>
<thead>
<tr>
<th>Domain that the mutation lies in</th>
<th>Average 275° peak area (13.73 min)</th>
<th>Average 253° peak area (17.789 min)</th>
<th>Normalized peak area</th>
<th>Product percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT ΔA63</td>
<td>36285</td>
<td>141713</td>
<td>36285</td>
<td>100</td>
</tr>
<tr>
<td>Δ79</td>
<td>27901</td>
<td>163236</td>
<td>32138.53</td>
<td>88.57</td>
</tr>
<tr>
<td>Y108F</td>
<td>12554</td>
<td>148854</td>
<td>13186.6</td>
<td>36.34</td>
</tr>
<tr>
<td>D141N</td>
<td>2147</td>
<td>136199</td>
<td>2063.46</td>
<td>5.68</td>
</tr>
<tr>
<td>E203Q</td>
<td>4278</td>
<td>138291</td>
<td>4174.69</td>
<td>11.50</td>
</tr>
<tr>
<td>W266A</td>
<td>1319</td>
<td>136414</td>
<td>1269.67</td>
<td>3.49</td>
</tr>
<tr>
<td>W266F</td>
<td>24477</td>
<td>145134</td>
<td>25067.88</td>
<td>69.08</td>
</tr>
<tr>
<td>W266L</td>
<td>7999</td>
<td>155750</td>
<td>8791.31</td>
<td>24.22</td>
</tr>
<tr>
<td>R324H</td>
<td>31841</td>
<td>155886</td>
<td>35025.48</td>
<td>96.52</td>
</tr>
<tr>
<td>R324Q</td>
<td>19251</td>
<td>148622</td>
<td>20189.55</td>
<td>55.64</td>
</tr>
<tr>
<td>R379A</td>
<td>27915</td>
<td>118118</td>
<td>17702.76</td>
<td>49.73</td>
</tr>
<tr>
<td>A423S</td>
<td>20974</td>
<td>135152</td>
<td>20002.95</td>
<td>55.12</td>
</tr>
<tr>
<td>M507A</td>
<td>17011</td>
<td>18044.89</td>
<td>40.96</td>
<td>11.50</td>
</tr>
<tr>
<td>M507L</td>
<td>16950</td>
<td>17702.76</td>
<td>90.86</td>
<td></td>
</tr>
<tr>
<td>W579I</td>
<td>4301</td>
<td>146943</td>
<td>4459.73</td>
<td>12.29</td>
</tr>
</tbody>
</table>

Data from a representative experiment (run in duplicate) are shown

*a*the characteristic MS ion of copalol; *b*the characteristic ion of the internal standard ergosterol (C28H44O, 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.
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


(Received for publication 20 March 2015)