

## Supporting Information for

### Bifunctional Cytochrome P450 Enzymes Involved in Camptothecin Biosynthesis

Yun Yang,<sup>†,§</sup> Wei Li,<sup>†</sup> Jing Pang,<sup>†,§</sup> Liangzhen Jiang,<sup>†</sup> Xixing Qu,<sup>†</sup> Xiang Pu,<sup>†</sup> Guolin Zhang,<sup>†</sup> and Yinggang Luo <sup>\*,†,‡</sup>

<sup>†</sup> Center for Natural Products Research, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, China

<sup>‡</sup> State Key Laboratory of Bioorganic and Natural Products Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China

<sup>§</sup> University of Chinese Academy of Sciences, Beijing 100049, China

\* Corresponding author: Yinggang Luo, yinggluo@cib.ac.cn

## Contents

### Experimental section

Plant materials, seedling growth, and total RNA extraction and isolation.....	3
Molecular cloning of the candidate <i>CaCYP72As</i> .....	3
Heterologous expression of <i>CaCYP72As</i> in yeast and microsomes preparation.....	4
Enzymatic activity assay and characterization of <i>CaCYP72A565</i> and <i>CaCYP72A610</i> .....	5
Characterization and quantitation of enzymatic reaction product.....	6
Quantitative real-time PCR amplification.....	6
Analysis of biosynthetic intermediates from different tissues of <i>C. acuminata</i> .....	7

### Supplementary results

Potential genes responsible for C–C bond cleavage in <i>C. acuminata</i> .....	8
Molecular cloning and bioinformatics analyses of potent <i>CYP72As</i> from <i>C. acuminata</i> .....	8
Previously reported bifunctional CYPs.....	9
Figure S1.....	11
Figure S2.....	12
Figure S3.....	13
Figure S4.....	14
Figure S5.....	16
Figure S6.....	17
Figure S7.....	18
Figure S8.....	19
Figure S9.....	21
Figure S10.....	23
Figure S11.....	24

Figure S12.....	25
Figure S13.....	26
Figure S14.....	27
Figure S15.....	29
Figure S16.....	30
Supplementary references.....	32
Table S1.....	35
Table S2.....	36
GenBank accession numbers.....	36

## Experimental section

**Plant materials, seedling growth, and total RNA extraction and isolation.** The *C. acuminata* seeds collection, seedling growth, and total RNA extraction and isolation were performed, according to the reported experimental procedures.<sup>1-4</sup>

**Molecular cloning of the candidate *CaCYPs*.** The specific primers (Table S1) were designed on the basis of the nucleotide sequence of the *CYP* candidates from the transcriptome data of *C. acuminata*. All primers were synthesized, purified, and authenticated by TsingKe Biological Technology (Chengdu) Co., Ltd. The cDNA mixture was prepared from the total *C. acuminata* RNA by using the SMARTer™ RACE cDNA amplification kit (Clontech Laboratories, Inc., a member of TaKaRa Bio Group), according to the manufacturer's manual instructions. Using the cDNA mixture as template and the appropriate primers, the ORF of *CaCYP72A565*, *CaCYP72A610*, and *Caa\_locus\_10805* were obtained from the PCR amplification reaction by following the cycling conditions: 1 cycle of 94 °C for 5 min, 30 cycles of 94 °C for 30 s, 56 °C for 30 s, 72 °C for 1 min 30 s followed by a final extension at 72 °C for 10 min in a thermal cycler (Eppendorf AG, Hamburg, Germany). The PCR amplification products were gel-purified and ligated into the pGM-T vector (Tiangen Biotech (Beijing) Co., Ltd, China). The constructs were chemically transformed into *E. coli* DH5a competent cells. The DNA sequencing was performed in both directions by TsingKe Biological Technology (Chengdu) Co., Ltd. The nucleotide sequence was analyzed by using the similarity search BLAST program.

For *Caa\_locus\_18068*, the gene-specific primers *Caa\_locus\_18068*-F and -R (Table S1) were designed on the basis of the nucleotide sequence of *Caa\_locus\_18068* to amplify the partial fragment by PCR amplification reaction using the above prepared cDNA mixture as template and by following the PCR cycling conditions: 1 cycle of 95 °C for 5 min; 30 cycles of 95 °C for 30 s, 56 °C for 30 s, and 72 °C for 1 min followed by a final extension at 72 °C for 10 min. The PCR amplified products were gel-purified and ligated into the pGM-T vector. The construct was chemically transformed into *E. coli* DH5a competent cells and sequenced in both directions. The nucleotide sequence of the above prepared DNA was sequenced in TsingKe Biological Technology and analyzed by using the similarity search BLAST program. The 3'- ends of *Caa\_locus\_18068* were obtained by using 3'-RACE methodology, according to the manual instructions of SMARTer™ RACE cDNA amplification kit from Clontech Laboratories. In Brief, the gene-specific primers *Caa\_locus\_18068*-3'-RACE-GSP1 and -GSP2 (Table S1) were designed on the basis of the nucleotide sequence of the above confirmed amplicon of *Caa\_locus\_18068*. Using 3'-RACE-Ready cDNA as template, Universal Primer A mix (UPM) and *Caa\_locus\_18068*-3'-RACE-GSP1 as primers, the 3'-end of *Caa\_locus\_18068* was obtained by following the PCR cycling conditions: 1 cycle of 95 °C for 5 min; 30 cycles of 95 °C for 30 s, 56 °C for 30 s, and 72 °C for 30 s followed by a final extension at 72 °C for 10 min. The PCR amplification products were gel-purified and used as template for nested PCR with *Caa\_locus\_18068*-3'-RACE-GSP2 and 3'-RACE inner primer (NUP), following the cycling conditions: 1 cycle of 95 °C for 5 min, 30 cycles of 95 °C for 30 s, 56 °C for 30s, 72 °C for 30 s followed by a final extension at 72 °C for 10 min. The PCR products were gel-purified and ligated into the pGM-T vector. The *E. coli* DH5a competent cells were chemically transformed with the construct following the procedure mentioned above. The nucleotide sequences of the 3'-ends of

*Caa\_locus\_18068* were sequenced in TsingKe Biological Technology. The gene-specific primers, *Caa\_locus\_18068*-Full-F and -R (Table S1), were designed and synthesized to clone the full-length cDNA of *Caa\_locus\_18068* on the basis of the above mentioned sequences of the core amplicon fragments and 3'-RACE products. Using the above prepared cDNA mixture as template, the ORF of *Caa\_locus\_18068* was obtained by following the PCR cycling conditions: 1 cycle of 95 °C for 5 min; 30 cycles of 94 °C for 30 s, 56 °C for 30 s, 72 °C for 1 min 30 s followed by a final extension at 72 °C for 10 min. The PCR amplified products were gel-purified, ligated into the pGM-T vector, and transformed into *E.coli* DH5 $\alpha$  competent cells following the procedure mentioned above. The nucleotide sequence of the full length of *Caa\_locus\_18068* was sequenced in TsingKe Biological Technology and analyzed by using the similarity search BLAST program.

**Heterologous expression of *CaCYP72As* in yeast and microsomal preparation.** The specific primers (Table S1) were designed on the basis of the coding sequence of the *CaCYP72As* and the appropriate endonucleases (*CaCYP72A565*: *Kpn* I and *Not* I; *CaCYP72A610*: *Sal* I and *Not* I; *Caa\_locus\_10805*: *Sac* I and *Not* I; *Caa\_locus\_18068*: *Bam*H I and *Not* I) to clone and insert the *CaCYP72As* into the pYES2-CT vector. The PCR products were gel purified, digested with the corresponding endonucleases, and ligated into the pYES2-CT that was digested with the same endonucleases to afford the expression construct. *E. coli* DH5 $\alpha$  competent cells were chemically transformed with the expression constructs and sequenced following the procedure mentioned above.

*S. cerevisiae* strain WAT11 harbouring the *A. thaliana* NADPH-dependent cytochrome P450 reductase 1<sup>5</sup> was introduced as a host to overexpress the candidate *CaCYP72As*. To prepare the electroporation competent cells, a single colony of *S. cerevisiae* strain WAT11 was inoculated into 10 mL of YPDA medium (bactopeptone 2%, yeast extract 1%, dextrose 2%, and adenine hemisulfate 0.3%) and incubated overnight at 30 °C and 230 rpm in a shaking incubator. Usually, the OD<sub>600</sub> of the overnight culture was 3.0 ~ 5.0. The culture was diluted to an OD<sub>600</sub> of 0.05 ~ 0.1 in a total volume of 50 mL of YPDA medium. The yeast cells were then incubated at 30 °C and 230 rpm in a shaking incubator until the OD<sub>600</sub> reached 0.5 ~ 0.6. The yeast cells were harvested by centrifugation at 4 °C, 3500 rpm for 5 min. The cell pellets were resuspended in 20 mL of sterile Solution I (1 mL Tris-HCl buffer (pH 8.0, 100 mM Tris, 10 mM EDTA), 10 mL 1 M LiAc, and 9 mL H<sub>2</sub>O) and incubated at 30 °C and 180 rpm. Forty-five minutes later, DTT (dithiothreitol) was added to a final concentration of 25 mM, and the resulting mixture was incubated at 30 °C and 180 rpm for 15 min. The yeast cells were collected by centrifugation at 4 °C and 3500 rpm for 5 min, then resuspended in 50 mL of sterile H<sub>2</sub>O and harvested by centrifugation at 4 °C and 3500 rpm for 5 min. The resulting yeast cells were resuspended in 5 mL of sterile Solution II (1 M sorbitol) and then collected by centrifugation at 4 °C and 3500 rpm for 5 min. This step was repeated three times. The yeast cells were then resuspended in 1 mL of Solution II, divided into 10 vials (each containing 100  $\mu$ L of electrocompetent cells), and stored at – 80 °C for future use.

Both recombinant plasmids and the empty vector were independently transformed into the *S. cerevisiae* strain WAT11 via electroporation on a cell-porator with a voltage booster and 0.20 cm chambers using 100  $\mu$ L of the above prepared yeast cells and 2  $\mu$ L of plasmids. The shocked cells were added to 1 mL of YPD medium (bactopeptone 2%, yeast extract 1%, and dextrose 2%), incubated at 30 °C and 160 rpm for 2 hours, and then spread onto SC-Ura plates (yeast nitrogen base without uracil 8 g L<sup>-1</sup>, dextrose 2%, and agar 2%). The colonies that appeared on the plates

were picked and confirmed by colony PCR using the corresponding primers to verify that the plasmids were successfully transferred into the yeast cells.

For protein expression, a single colony of the recombinant strain was inoculated into 10 mL of SC-Ura medium (yeast nitrogen base without uracil 8 g L<sup>-1</sup> and dextrose 2%) and cultured at 30 °C and 230 rpm in a shaking incubator. After 24 hours later, an aliquot of this culture (1 mL) was inoculated into 100 mL of SC-Ura medium and incubated at 30 °C and 230 rpm. When the culture reached the stationary phase, the yeast cells were washed three times with sterile water, collected, and harvested by centrifugation at 3000 rpm for 5 min. The above prepared yeast cells were transferred into 100 mL of YPGal medium (bactopeptone 1%, yeast extract 1%, and galactose 2%) to induce protein expression for 6 hours as described by de Bernonville <sup>32</sup>.

The yeast culture was harvested by centrifugation at 4 °C and 3500 rpm for 5 min. The yeast cell pellets were washed twice with TEK buffer (50 mM Tris-HCl, 1 mM EDTA, 100 mM KCl, pH8.0). The cells were resuspended in TES buffer (50 mM Tris-HCl, 1 mM EDTA, 0.6 M sorbitol, pH8.0) and disrupted by using a high pressure homogenizer (Union UH-03, Shanghai, China) at 600 bar for 35 min. The debris was removed by centrifugation at 15,000 × g for 20 min, twice. The corresponding supernatant was centrifuged at 190,000 × g for 1.5 hrs (S80AT3 rotor, HITACHI Micro Ultracentrifuge C5150FNX). The resulting microsomal fraction was dissolved in TEG buffer (50 mM Tris-HCl, 1 mM EDTA, 20% glycerol, pH8.0) and stored at – 80 °C. The protein concentration was estimated by using the  $\epsilon_{280\text{nm}}$  85,620 M<sup>-1</sup> cm<sup>-1</sup> for CaCYP72A610 (6.4 mM) and the  $\epsilon_{280\text{nm}}$  79,215 M<sup>-1</sup> cm<sup>-1</sup> for CaCYP72A565 (6.4 mM) calculated from ExPASy ProtParam. The yields of CaCYP72A565 and CaCYP72A610 are ~ 1.3 mg/L. Standard Western Blot of CaCYP72As was performed to confirm their overexpression. The characteristic Soret band at 450 nm was observed in the ferrous-CO versus ferrous difference spectrum.

**Enzymatic activity assay and characterization of CaCYP72A565 and CaCYP72A610.** The enzymatic activity assays were performed in 100  $\mu$ L of the above prepared cell/microsome containing 500  $\mu$ M NADPH and 1 mM specific substrate loganin, loganic acid, or 7-deoxyloganic acid. The catalytic reaction was initiated by the addition of NADPH, and the reaction mixture was incubated at 30 °C. The reaction mixture was quenched at 10, 30, 60, 120, 180, and 240 min by the addition of 100  $\mu$ L of CH<sub>3</sub>OH. After the removal of the denatured proteins by centrifugation at 12000 rpm for 5 min, the supernatant was analysed by high-performance liquid chromatography (HPLC) on a system equipped with an Altima C<sub>18</sub> analytic column (250 mm × 4.6 mm, 5  $\mu$ m) by the gradient elution of solvent A (CH<sub>3</sub>CN) and solvent B (H<sub>2</sub>O containing 0.2% acetic acid) at a flow rate of 1 mL min<sup>-1</sup> at 35 °C, monitored at 240 nm by a diode array detector (DAD). To detect the whole enzymatic reaction with loganin as substrate, the gradient elution of solvents A (CH<sub>3</sub>CN) and B (H<sub>2</sub>O containing 0.2% acetic acid) was conducted as follows: 0 min, 10% A; 25 min, 30% A; 26 min, 98% A; and 34 min, 10% A. All experiments were performed in duplicate or triplicate. To determine the kinetics parameters, the microsomal enzymatic reactions were conducted in a total volume of 50  $\mu$ L reaction mixtures containing TE buffer (50 mM Tris-HCl, 1 mM EDTA, pH 8.0), 3.2 mM microsome, NADPH, and substrate with different concentrations. The reactions were incubated for 2 hrs at 30 °C, 250 rpm and quenched by adding 50  $\mu$ L methanol. Proteins were removed by centrifugation at 12,000 rpm for 10 min. the resulting supernatant was used for HPLC-DAD analyses. The kinetic constants were calculated with nonlinear regression analysis using Origin 8.0 software.

**Characterization and quantitation of enzymatic reaction product.** The fractions of each new individual peak from HPLC-DAD analysis were collected and characterized by high-resolution electrospray ionization mass spectrometry (HRESIMS) analysis. Each pooled product was directly analysed on a Bruker micrOTOF-Q mass spectrometer (Bremen, Germany) equipped with an electrospray ionization interface. The electrospray ionization mass spectrometer was operated in positive or negative ion mode, and the spectra were collected in the enhanced full mass scan mode from  $m/z$  50 ~ 1500.

The ultra-high-performance liquid chromatography (UPLC) - HRESIMS analyses were performed on a Waters Acquity I Class UPLC system equipped with an Acquity UPLC HSS T3 column (2.1 mm  $\times$  100 mm, 1.8  $\mu$ m) and a Vion IMS QToF mass spectrometer with electrospray ionization source, source temperature 120  $^{\circ}$ C, desolvation temperature 450  $^{\circ}$ C, cone gas flow 50 L h $^{-1}$ , desolvation gas flow 800 L h $^{-1}$ , and capillary voltage 3.00 kV. Mass spectrum acquisition was performed in positive ion mode over  $m/z$  50 to 1000 with a scan time of 0.200 s using dynamic range extension. The low collision energy was 6.00 eV, while the high collision energy ramp started at 20.00 eV and ended at 40.00 eV. The mobile phase consisted of solvent A (CH $_3$ CN) and solvent B (H $_2$ O containing 0.2% acetic acid). To detect the whole enzymatic reaction with 7-deoxyloganic acid as substrate, the gradient elution of solvents A and B was conducted as follows: 0 min, 5% A; 20 min, 17% A; 29 min, 40% A; 30 min, 98% A; and 40 min, 5% A. The UPLC separation of the sample was performed at 30  $^{\circ}$ C with a flow rate of 0.3 mL min $^{-1}$  and a sample injection volume of 10  $\mu$ L, monitored at 240 nm by a DAD. The mass spectrometer was operated in positive ion mode.

**Quantitative real-time PCR amplification.** To quantitate the relative expression level of CaCYP72A565 and CaCYP72A610 in different plant tissues from different individual plants of *C. acuminata*, 1  $\mu$ g of total RNA was used to perform reverse transcription by using an EasyScript $^{\circ}$  First-Strand cDNA Synthesis SuperMix kit and an anchored oligo(dT) $_{18}$  primer, according to the manufacturer's protocols (TransGen Biotech). The resulting cDNA mixtures were used as templates for qRT-PCR experiments by using a CFX Connect $^{\text{TM}}$  real-time PCR System (Bio-Rad, USA), according to the manufacturer's instructions. Briefly, two pairs of specific primers, rtCaa\_locus\_133-F and -R, rtCaa\_locus\_1905-F and -R (Table S1), were designed from the 3'-UTR sequence of *Caa\_locus\_133* and *Caa\_locus\_1905* for real-time PCR amplification, according to the instructions of real-time PCR system and the specificity of these primers were confirmed by blast program. The actin gene from *C. acuminata* was used as an internal reference. The real-time PCR amplification was performed in a total volume of 25  $\mu$ L SYBR Green solution containing 2.5  $\mu$ L cDNA template, 0.5  $\mu$ L SYBR Green I (Zeesan Biotech Co., Ltd, Xiamen, China), 0.5  $\mu$ L of the primers (10  $\mu$ M), 2.5  $\mu$ L 10  $\times$  TransTaq $^{\circ}$  HiFi Buffer II, 1.5  $\mu$ L dNTP (2.5 mM), 0.25  $\mu$ L TransTaq $^{\circ}$  HiFi DNA polymerase (TransGen Biotech), and ddH $_2$ O with the following cycling conditions: 1 cycle of 95  $^{\circ}$ C for 3 min; 40 cycles of 95  $^{\circ}$ C for 10 s, 56  $^{\circ}$ C for 30 s, 72  $^{\circ}$ C for 20 s. Each sample and the negative control without template were conducted in triplicate. The relative transcript levels of *CaCYP72A565* and *CaCYP72A610* in different tissues were calculated by using the corresponding transcript levels of the reference gene encoding actin in the same tissue as control.

**Analysis of biosynthetic intermediates from different tissues of *C. acuminata*.** The leaves, roots, and stems from the seedling of *C. acuminata* cultured in the laboratory described previously.<sup>1-4</sup> The young leaves of *C. acuminata* were collected from the wild plant located at 104°3'36"E, 30°40'12"N at the campus of Chengdu Institute of Biology of the Chinese Academy of Sciences in March, 2017. The flower buds of *C. acuminata* were collected from the wild plant located at 102°58'12" E, 29°58'12" N at the campus of Sichuan Agricultural University at Ya'an, China in May, 2017. The authentic standard compounds, including CAM, 10-hydroxycamptothecin, geraniol, 10-hydroxygeraniol, tryptamine, 5-hydroxytryptamine, loganin, loganic acid, secologanin, and strictosidine were commercially available. 7-Deoxyloganic acid was provided from Prof. Dr. Zhou at Northwest A&F University at Yangling, Shaanxi, China. Secologanic acid and secoxyloganin were from Prof. Dr. Song at Shenyang Pharmaceutical University, Shenyang, China.

Two solvent system, CH<sub>3</sub>CN - H<sub>2</sub>O (7 : 3, vol. / vol., Method A)<sup>6</sup> and EtOH – H<sub>2</sub>O (7 : 3, vol. /vol., Method B),<sup>7</sup> were introduced to extract the metabolites. Each plant sample was frozen and homogenized in a mortar and pestle. To the powdered plant tissue, 5 mL extraction solvent per gram sample was added. The mixture was agitated by ultrasound for 10 min and then was incubated in the dark at 4 °C for 20 hours. The mixture was separated by centrifugation at 4 °C, 12,000 rpm for 30 min. The supernatant was transferred and centrifuged again. An aliquot (500 uL) was diluted by addition of 500 uL extraction solvent. The samples were subjected to UPLC - HRESIMS analyses. A 5 µL of each extract was injected to the UPLC and analyzed by using the gradient elution with mobile phases consisting of 10 mM ammonium formate in H<sub>2</sub>O, adjusted to pH 2.85 with formic acid (solvent A) and CH<sub>3</sub>OH (solvent B). The 55-min linear gradient program was as follows: 0 min, 5% B; 15 min, 20% B; 38 min, 55%; 45 min, 99% B; and 50min, 5% B. The flow rate was 0.3 mL min<sup>-1</sup> and the column temperature was 30 °C. Metabolites were identified by the HRESIMS spectra, UV spectra, and retention time, compared with the data obtained with the standards.

## Supplementary results

### Potential genes responsible for C–C bond cleavage in *C. acuminata*

Two CYP72As, CrSLS1 and CrSLS2, designated secologanin synthase, have been cloned and functionally characterized to catalyse the specific intramolecular C–C bond cleavage conversion of loganin into secologanin in *Catharanthus roseus* (Figure 1B).<sup>8,9</sup> In *C. acuminata*, the candidate CaSLS (GenBank ID: HQ605982) was annotated from the first *C. acuminata* transcriptome dataset and was reported without biochemically catalytic function.<sup>10</sup> To characterize its enzyme function, the nucleotide sequence of the candidate CaSLS was retrieved from the NCBI database (<https://www.ncbi.nlm.nih.gov/>) and used as a query to blast the second *C. acuminata* transcriptome dataset downloaded from <http://medicinalplantgenomics.msu.edu>.<sup>11</sup> Two CYPs, Caa\_locus\_133 and Caa\_locus\_1905 (= HQ605982), were suggested to be potent CaSLSs. Additionally, seven CYP candidates, Caa\_locus\_21332, Caa\_locus\_10805, Caa\_locus\_40906, Caa\_locus\_17968, Caa\_locus\_44923, Caa\_locus\_18068, and Caa\_locus\_10311, were putatively annotated as CaSLSs from the second *C. acuminata* transcriptome dataset (<http://medicinalplantgenomics.msu.edu>).<sup>11</sup> The seven candidate CaSLSs were retrieved from the second *C. acuminata* transcriptome dataset (<http://medicinalplantgenomics.msu.edu>)<sup>11</sup> and were aligned with each other (Figure S1). Caa\_locus\_10805 and Caa\_locus\_18068 were selected as candidates according to the phylogenetic analyses and the nucleotide sequence length (Figure S1). Thus, four CaCYPs, Caa\_locus\_133, Caa\_locus\_1905, Caa\_locus\_10805, and Caa\_locus\_18068, were cloned and functionally characterized in the following works.

### Molecular cloning and bioinformatics analyses of potent CYP72As from *C. acuminata*

Multiple alignment of the amino acid residue sequence of the four candidate CaCYPs revealed that Caa\_locus\_1905 showed 95.2, 76.3, and 71.0% amino acid residue similarity and 88.3, 58.9, and 55.9% amino acid residue identity, respectively, to Caa\_locus\_133, Caa\_locus\_10805, and Caa\_locus\_18068 (Figure S2). Using the specific primers (Table S1) and the cDNA mixture prepared previously as a template,<sup>1-4</sup> the full-length cDNA of Caa\_locus\_1905, Caa\_locus\_133, and Caa\_locus\_10805 was obtained by PCR amplification. The reported sequence for Caa\_locus\_18068 was prepared by PCR amplification as described above. 3'-Rapid amplification of cDNA ends (RACE) methodology was employed to clone the 3'-end of Caa\_locus\_18068. The partial Caa\_locus\_18068 without the N-terminal ~ 25 amino acid residues was generated by overlapping the nucleotide sequences of the two fragments, which was confirmed by molecular cloning and subsequent DNA sequencing (Figure S2). The missed N-terminal amino acid residues are thought to be a membrane-anchor sequence that retains P450s on the cytoplasmic surface of the endoplasmic reticulum membrane in eukaryotic cells.<sup>12</sup> For the heterologous overexpression of plant CYPs in *Escherichia coli* or *Saccharomyces cerevisiae*, the membrane-anchor sequence of CYPs might be replaced for expression in the host.<sup>8,13</sup> Thus, the specific primers (Table S1) were designed to prepare the modified sequence instead of using the 5'-RACE methodology to clone the entire cDNA sequence from *C. acuminata*.

When aligned with the functionally characterized CrSLSs (Figure S3), the four deduced amino acid residue sequences showed characteristic features of CYPs, including a conserved PFS(G)WGPRI(V)CL(V/I)G motif (highlighted in purple) for haem binding<sup>14,15</sup> and a conserved I-helix motif (AGQETT, highlighted in yellow) involved in oxygen binding and activation.<sup>15,16</sup>



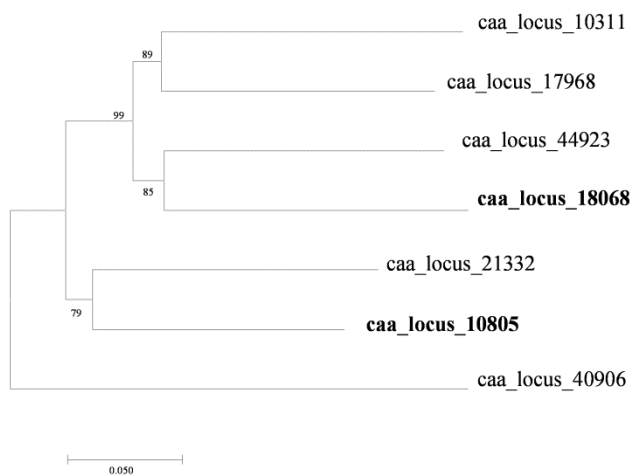
The candidate CaCYPs were classified into the CYP72A subfamily by bioinformatics and phylogenetic analyses (Figure S4). The crystal structure (PDB ID: 3NA0) of CYP11A1,<sup>17</sup> a cholesterol side chain cleavage enzyme, showed ~ 16% amino acid residue identity with the CaCYP72As and was selected by the PHYRE2 online software (Protein Homology/analogy Recognition Engine V 2.0, <http://www.sbg.bio.ic.ac.uk/phyre2/html/>) as a model template to predict the 3-dimensional structures of the CaCYP72As, as no crystal structure of the plant CYP72A subfamily could be found in the Protein Data Bank (PDB). The characteristic haem (highlighted in red sticks) involved in CYP-catalysed reactions was predicted to be present in the CaCYP72As, embraced by the conserved PFS/GWGPRICVG motif (highlighted in pink) and I-helix motif AGQETT (highlighted in yellow), by the 3DligandSite tool (Figure S5).

### Previously reported bifunctional CYPs

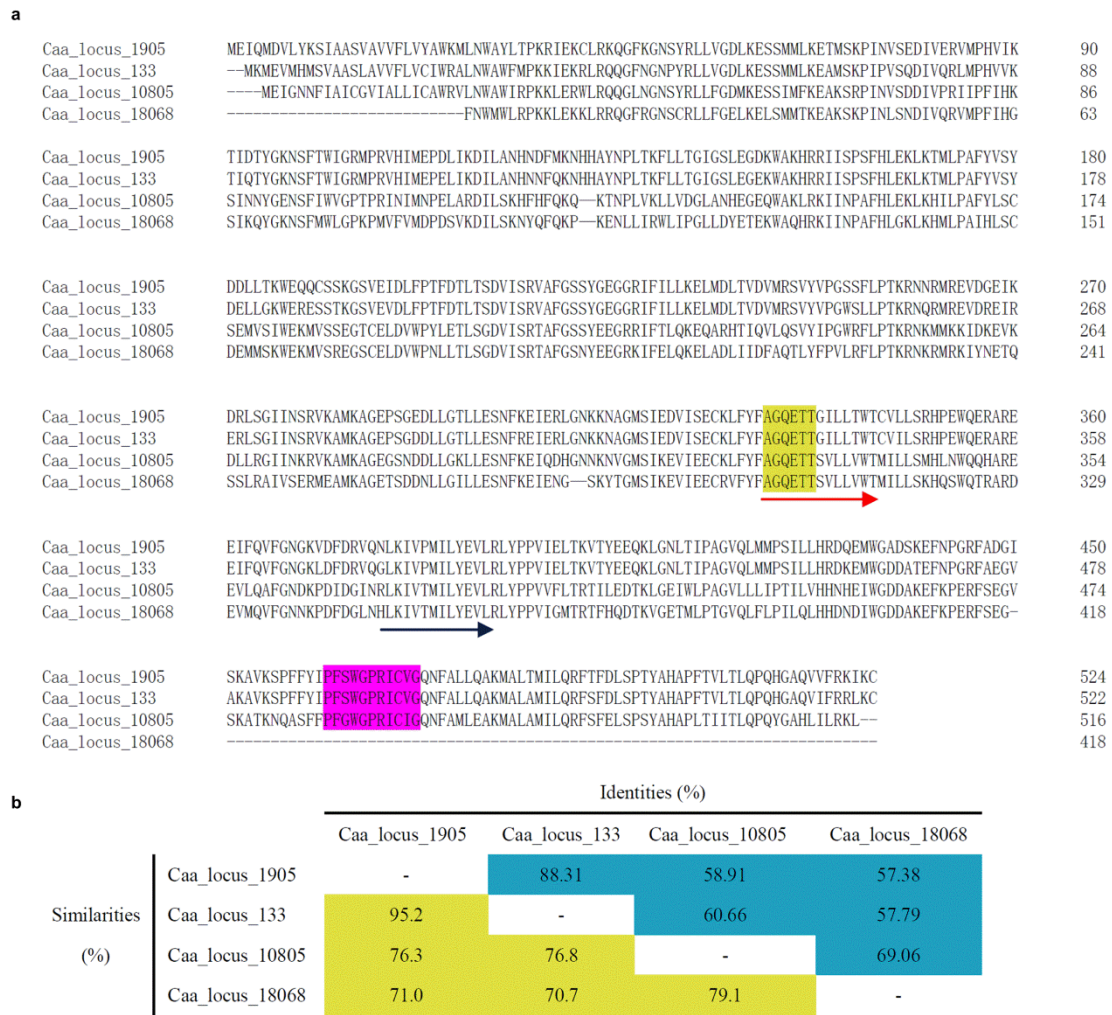
A few bifunctional CYPs were shown to catalyse both oxidative reactions and unique reactions. For instance, in microbial antibiotic biosynthesis, the CYP MycG catalyses sequential hydroxylation and epoxidation reactions in mycinamicin biosynthesis,<sup>18</sup> while the CYP GfsF catalyses epoxidation and subsequent hydroxylation in the biosynthesis of macrolide polyketide FD-891.<sup>19</sup> TamI catalyses two hydroxylations and one epoxidation in tirandamycin biosynthesis.<sup>20</sup> AurH catalyses hydroxylation and tetrahydrofuran ring formation in aureothin biosynthesis.<sup>21</sup> CYP114 was shown to catalyse the hydroxylation at C-7 of ent-kaurenoic acid and the subsequent extrusion of C-7 to generate GA<sub>12</sub>-aldehyde.<sup>22</sup> CYP170A1 has been found to be a bifunctional protein that catalyses both monooxygenase activity and terpene synthase activity by two distinct active sites in the well-established P450 protein structure.<sup>23</sup> In human steroid hormone biosynthesis, CYP11A1 catalyses stereospecific C-22 $\alpha$  and C-20 $\beta$  hydroxylation followed by a C–C bond cleavage between C-20 and C-22 to afford isocaproic aldehyde and 7-dehydropregnenolone.<sup>24</sup> Another human steroidogenic CYP, CYP17A1, is a bifunctional enzyme that catalyses 17 $\alpha$ -hydroxylation and 17,20-lyase reactions to convert 21-carbon steroids to 19-carbon androgens, the precursor for all sex steroid.<sup>25</sup>

Plant CYPs participate in a variety of biochemical pathways to produce a vast diversity of plant natural products, and CYP genes in plant genomes are estimated to represent up to 1% of the total gene annotations of each plant species, implying that plants involve many and various CYP-dependent reactions. Geraniol 10-hydroxylase (G10H), a member of the CYP76 family, hydroxylates geraniol at the C-10 position to generate 10-hydroxygeraniol, which is involved in the biosynthesis of MIA in *Catharanthus roseus*.<sup>26</sup> It was shown to catalyse the 3'-hydroxylation of naringenin, a well-known flavonoid, to produce eriodictyol.<sup>26</sup> The CYP81AA1 and CYP81AA2 from *Hypericum calycinum* and *H. perforatum* were shown to catalyse two consecutive reactions, the 3'-hydroxylation of 2,4,6-trihydroxybenzophenone and the phenolic carbon–oxygen coupling of the resulting 2,3',4,6-tetrahydroxybenzophenone.<sup>27</sup> CYP81AA1 catalyses cyclization via the *para* position to form 1,3,7-trihydroxyxanthone. CYP81AA2 directs cyclization to the *ortho* position to yield 1,3,5-trihydroxyxanthone.<sup>27</sup> In plant-derived ajmalan alkaloid pathways, CYP82S18 (vinorine hydroxylase) hydroxylates vinorine to form vomilenine and then catalyses the non-oxidative isomerization of the ajmaline precursor vomilenine to perakine.<sup>28</sup> In *Arabidopsis*, AtCYP85A2 works as a bifunctional CYP with brassinosteroid C-6 oxidase and brassinolide synthase activity to catalyse the conversion of castasterone to brassinolide in the brassinosteroid biosynthetic pathway.<sup>293</sup> Arabidopsis CYP705A1 catalyses an unusual C–C bond

cleavage reaction in which the prenyl side chain of arabidiol is cut to produce volatile (*E*)-4,8-dimethyl-1,3,7-nonatriene and a non-volatile 14-apo-arabidiol.<sup>30</sup> The CYP72As are involved in species-specific metabolic functions in some plants, and there is likely functional conservation between closely related species.<sup>31</sup>

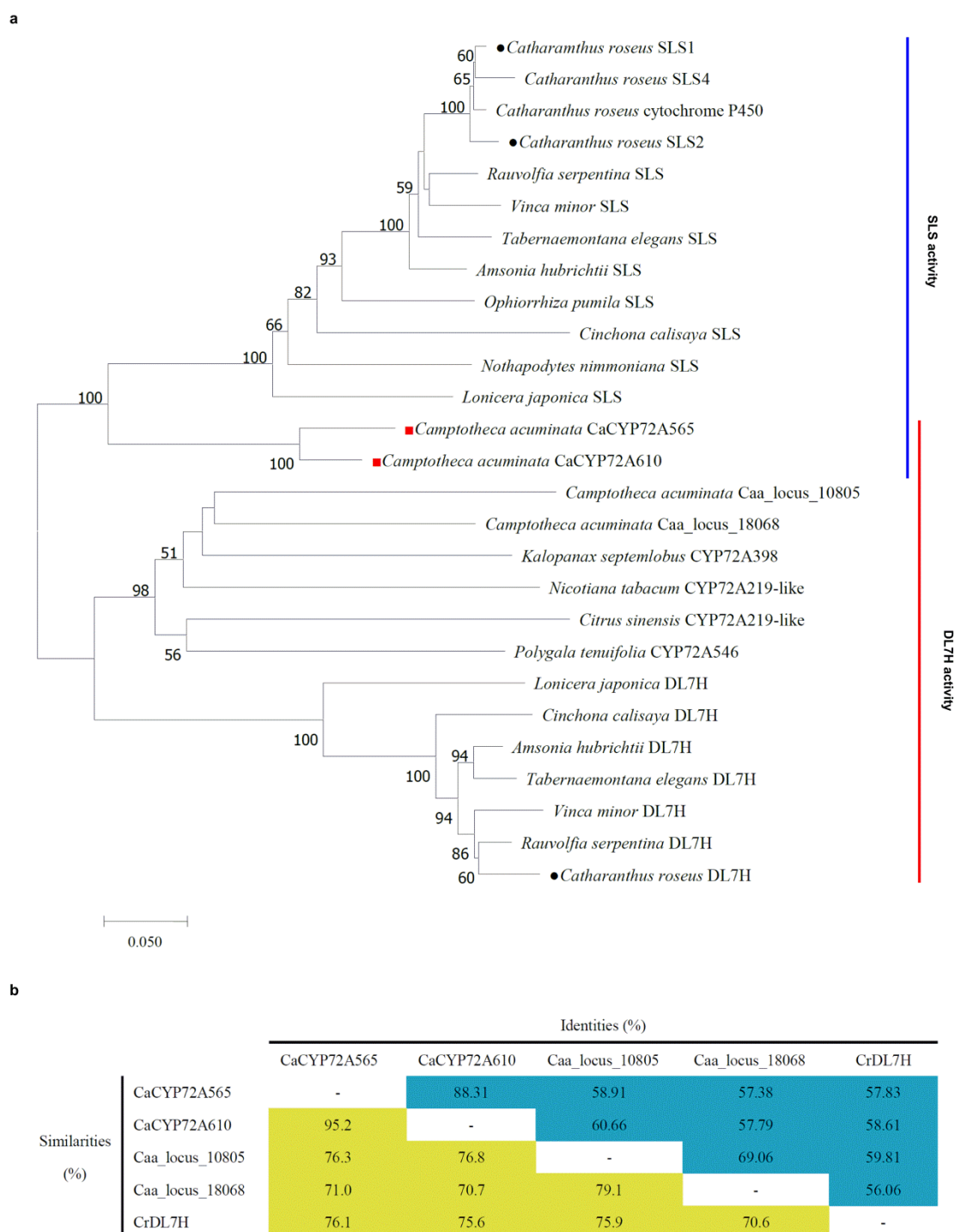


**Figure S1. Phylogenetic tree of the seven SLS candidates from the second *C. acuminata* transcriptome dataset** (<http://medicinalplantgenomics.msu.edu>). The candidates highlighted in bold form, *Caa\_locus\_10805* and *Caa\_locus\_18068*, were selected to the following experimental investigations in view of the whole full length of cDNA and the phylogenetic relationship.



**Figure S2. Multiple amino acid residue sequence alignment of the four candidate CaSLs (a) and their identities and similarities (b).** The alignment was performed by using Clustal Omega multiple alignment software. The two highly conserved characteristic domains, haem binding domain (PFS(G)WGPRICV(I)G) and I-helix motif (AGQETT), were highlighted in purple and yellow, respectively. The 3'-RACE-GSP1 and -GSP2 primers (Table S1) were designed on the conserved amino acid residues in red and black arrows, respectively, to obtain the 3'-rapid amplification of cDNA ends (RACE) of Caa\_locus\_18068 by using 3'-RACE methodology.

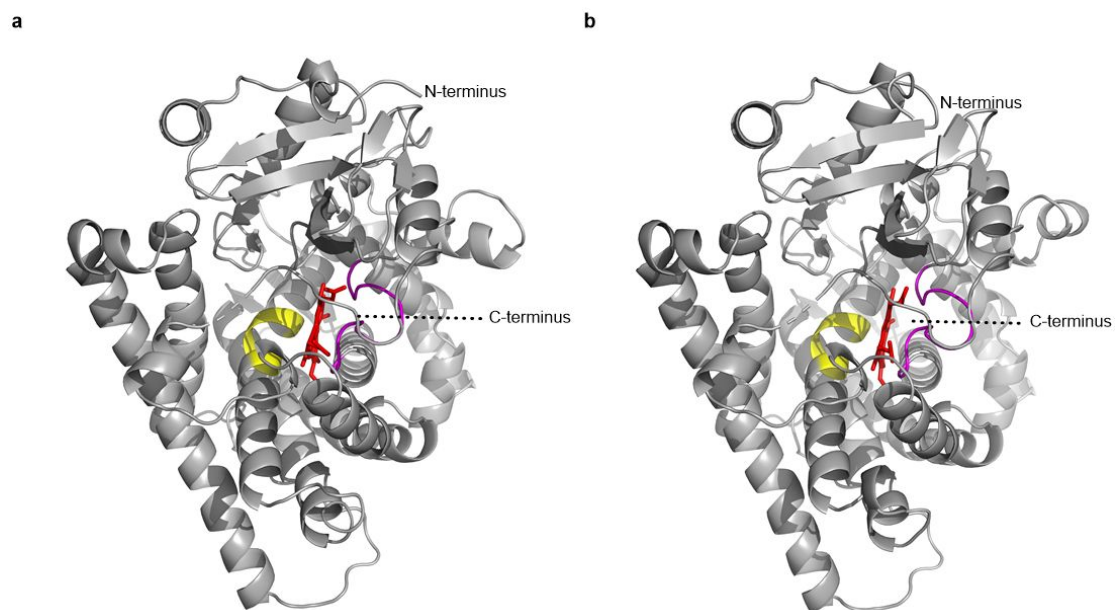




**Figure S4. Phylogenetic tree of plant secologanin synthases and 7-deoxyloganic acid-7-hydroxylases (a) and the identities and similarities between the four CaSLSs and CrDL7H (b).** The amino acid residue sequences of the plant secologanin synthases and 7-deoxyloganic acid-7-hydroxylases were retrieved from the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>) using the BLAST program. Phylogenetic analysis was performed using the MEGA 6.0 software package and neighbor-joining program (<http://www.megasoftware.net>) with 1000 boots trapped value support and a Poisson correction.

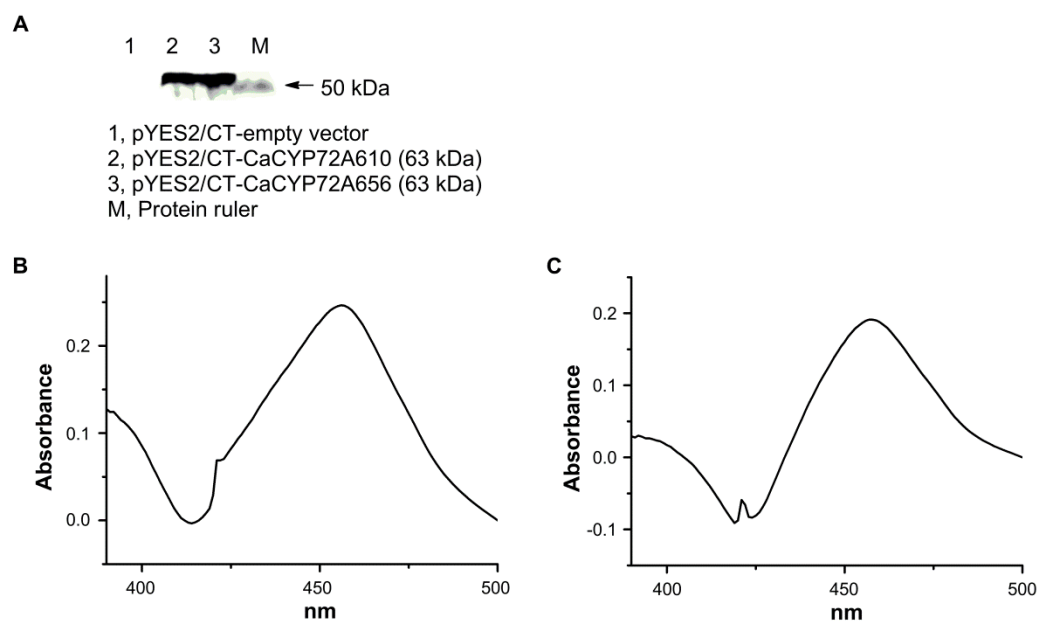
The bootstrap values higher than 50% are given at the nodes. The scale bar indicates the phylogenetic distance calculated from the number of differences. The secologanin synthases are *Catharanthus roseus* SLS1 (AAA17746.1), *Catharanthus roseus* SLS4 (AKF02530.1), *Catharanthus roseus* cytochrome P450 (AAA17732.1), *Catharanthus roseus* SLS2 (AGX93064.1), *Rauvolfia serpentina* SLS (AGX93047.1), *Vinca minor* SLS (AGX93049.1), *Tabernaemontana elegans* SLS (AGX93048.1), *Amsonia hubrichtii* SLS (AGX93046.1), *Ophiorrhiza pumila* SLS (BAP90521.1), *Cinchona calisaya* SLS (AGX93045.1), *Nothapodytes nimmoniana* SLS (AQW38832.1), and *Lonicera japonica* SLS (AGX93044.1). The 7-deoxyloganic acid-7-hydroxylases are *Kalopanax septemlobus* CYP72A398 (ALO23114.1), *Nicotiana tabacum* CYP72A219-like (NP\_001313096.1), *Citrus sinensis* CYP72A219-like (XP\_006477948.1), *Polygala tenuifolia* CYP72A546 (ASB17941.1), *Lonicera japonica* DL7H (AGX93056.1), *Cinchona calisaya* DL7H (AGX93057.1), *Tabernaemontana elegans* DL7H (AGX93060.1), *Vinca minor* DL7H (AGX93061.1), *Rauvolfia serpentina* DL7H (AGX93059.1), and *Catharanthus roseus* DL7H (AGX93062.1). The functionally characterized plant secologanin synthases and 7-deoxyloganic acid-7-hydroxylase were highlighted with black cycle. CaCYP72As reported in the present work were highlighted with red square.



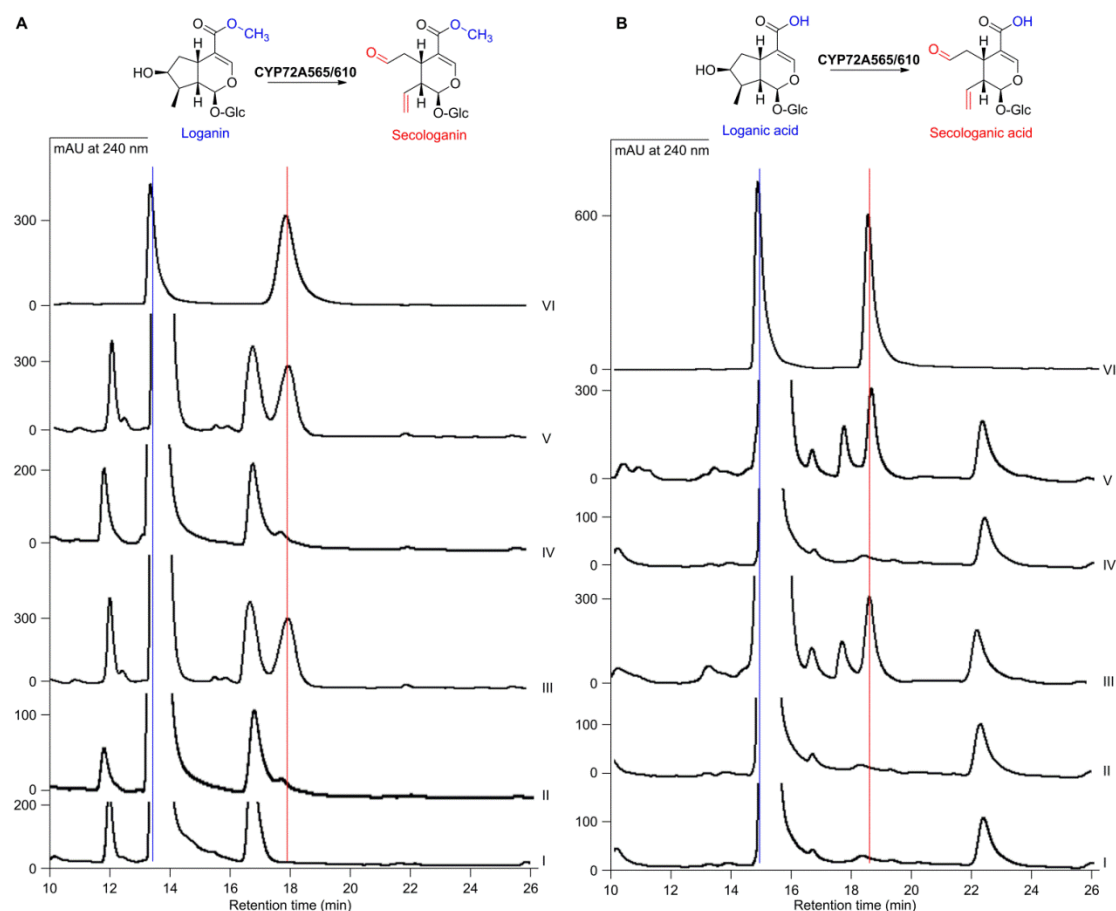


**Figure S5. Predicted 3-dimensional structure of CaCYP72A565 (a) and CaCYP72A610 (b).** The crystal structure of human steroidogenic CYP11A1 (PDB ID: 3NA0) was used as model template to construct the 3-dimensional structure of CaCYP72As. The conserved I-helix, haem-binding domain are illustrated in yellow and purple, respectively. The haem is present in red.

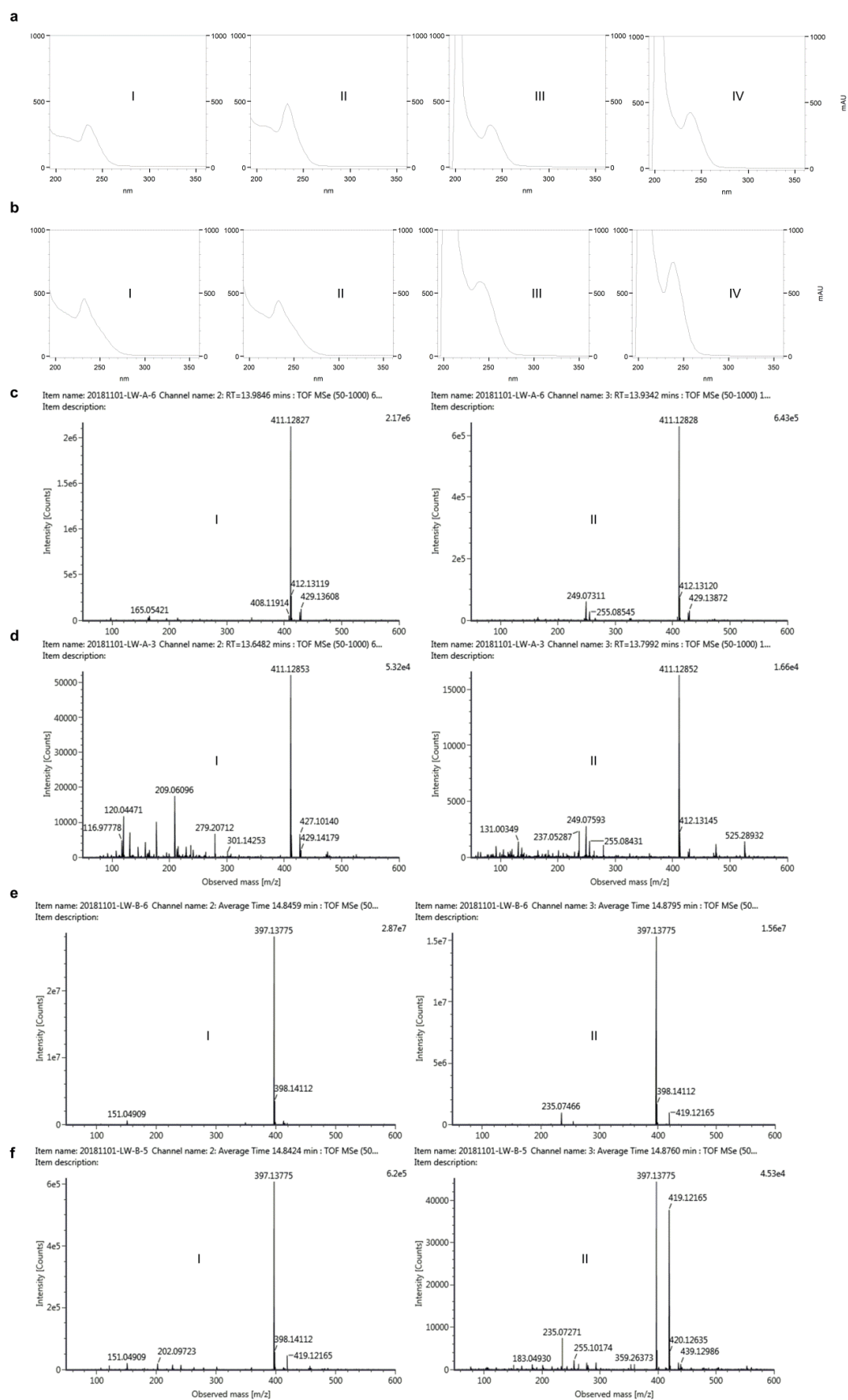




**Figure S6. CaCYP72As overexpression and Fe<sup>2+</sup>-CO binding.** **A**, Western blot of CaCYP72As overexpressed in yeast. Ferrous-CO versus ferrous difference spectrum of CaCYP72A565 (**B**) and CaCYP72A610 (**C**).

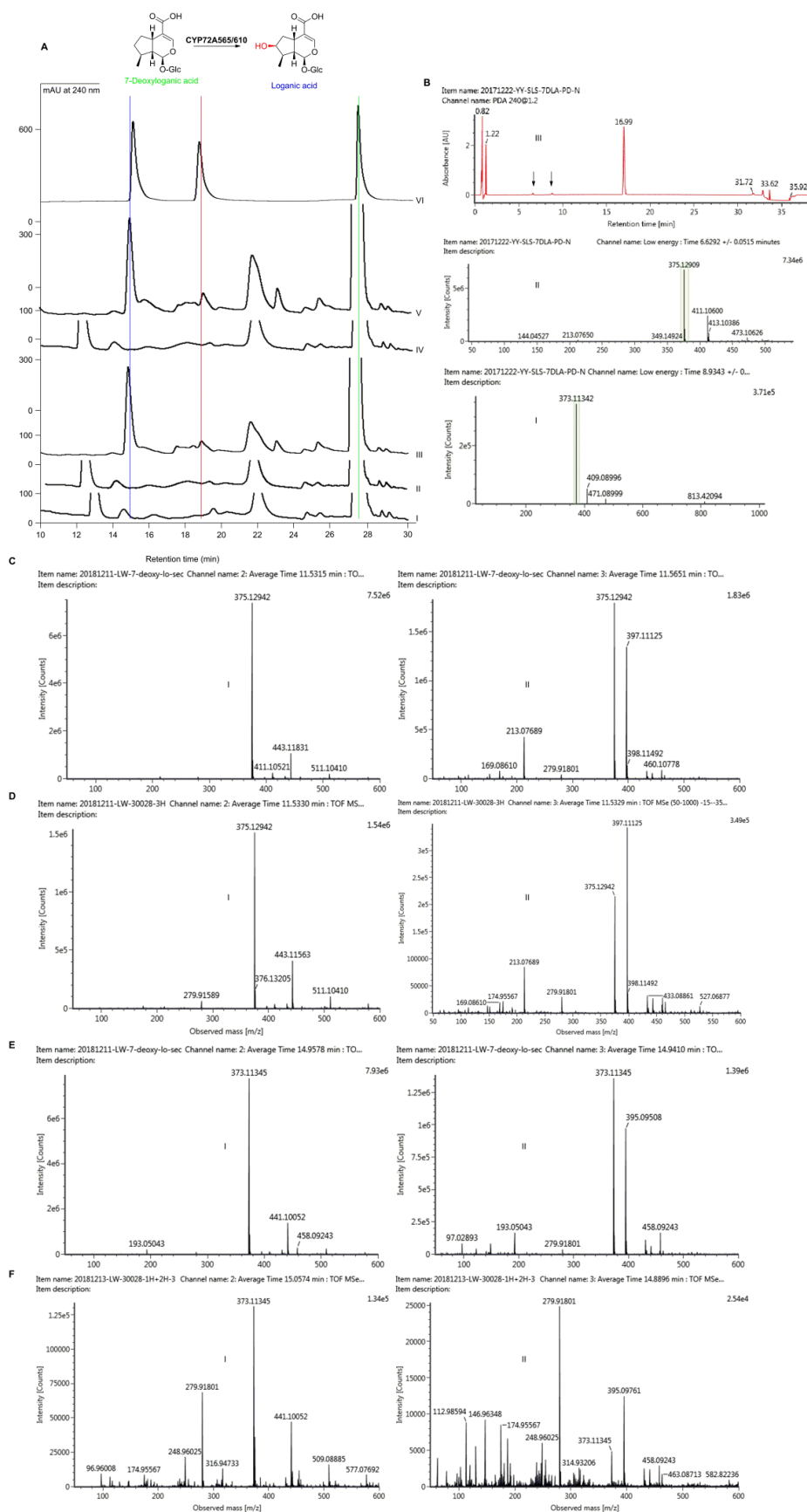


**Figure S7. Catalytic oxidative C–C bond cleavage of iridoid glucosides to generate secoiridoid glucosides by CaCYP72As.** **A**, HPLC-DAD analyses of the standard loganin (blue) and secologanin (red) (*panel VI*) and the reaction mixtures from the whole enzymatic reactions using the yeast cells harbouring the empty vector (*panel I*), boiled CaCYP72A610 (*panel II*), CaCYP72A610 (*panel III*), boiled CaCYP72A565 (*panel IV*), and CaCYP72A565 (*panel V*) as catalysts. **B**, HPLC-DAD analyses of the standard loganic acid (blue) and secologanic acid (red) (*panel VI*) and the reaction mixtures from the whole enzymatic reactions using the yeast cells harbouring the empty vector (*panel I*), boiled CaCYP72A610 (*panel II*), CaCYP72A610 (*panel III*), boiled CaCYP72A565 (*panel IV*), and CaCYP72A565 (*panel V*) as catalysts.



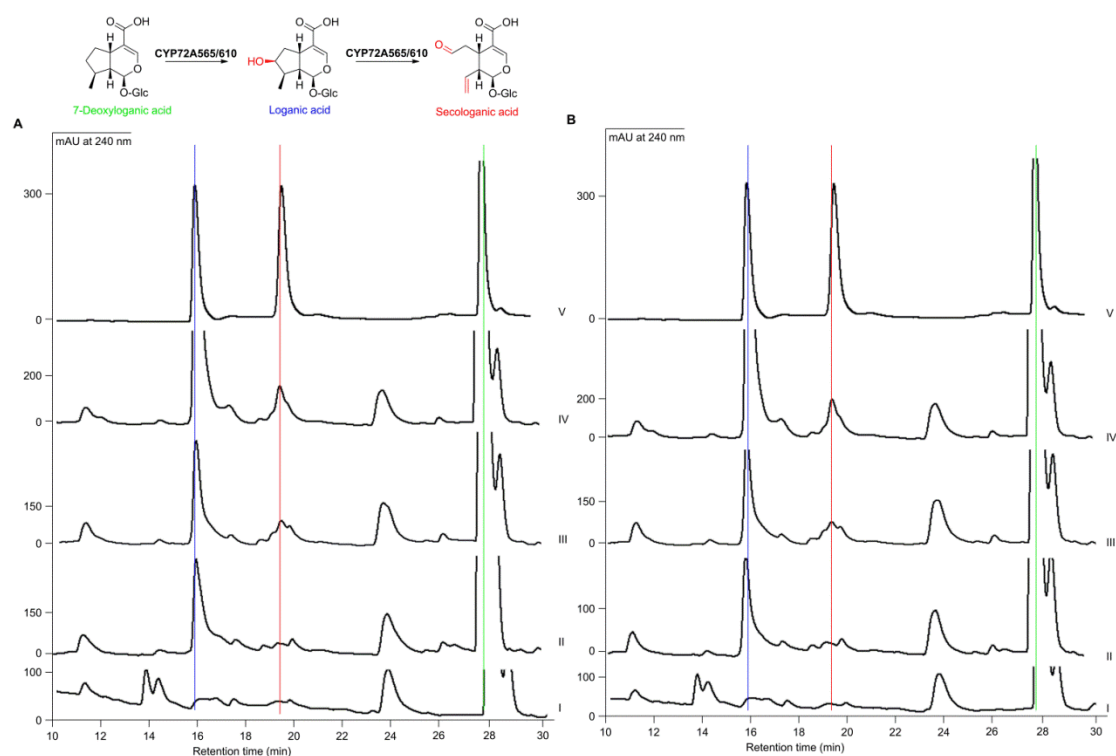
**Figure S8. UV spectra and HRESIMS of standards and enzymatic products. a,** The UV spectra of the standard loganin (*panel IV*), secologanin (*panel III*), and the enzymatic conversion

products from CaCYP72A565 (*panel II*) and CaCYP72A610 (*panel I*) using loganin as substrate. **b**, The UV spectra of the standard loganic acid (*panel IV*), secologanic acid (*panel III*), and the enzymatic conversion products from CaCYP72A565 (*panel II*) and CaCYP72A610 (*panel I*) using loganic acid as substrate. **c**, HRESIMS (positive ionization mode, *panel I*) of the standard secologanin and its fragmentation pattern (*panel II*, ). **d**, HRESIMS (positive ionization mode,  $[M + Na]^+$  at  $m/z$  411.1285, *panel I*) of the enzymatic conversion product and its fragmentation pattern ( $[M - Glc + Na]^+$  at  $m/z$  249.0759, *panel II*) using loganin as substrate. **e**, HRESIMS (positive ionization mode, *panel I*) of the standard secologanic acid and its fragmentation pattern (*panel II*). **f**, HRESIMS (positive ionization mode,  $[M + Na]^+$  at  $m/z$  397.1377, *panel I*) of the enzymatic conversion product and its fragmentation pattern ( $[M - Glc + Na]^+$  at  $m/z$  235.0727, *panel II*) using loganic acid as substrate.

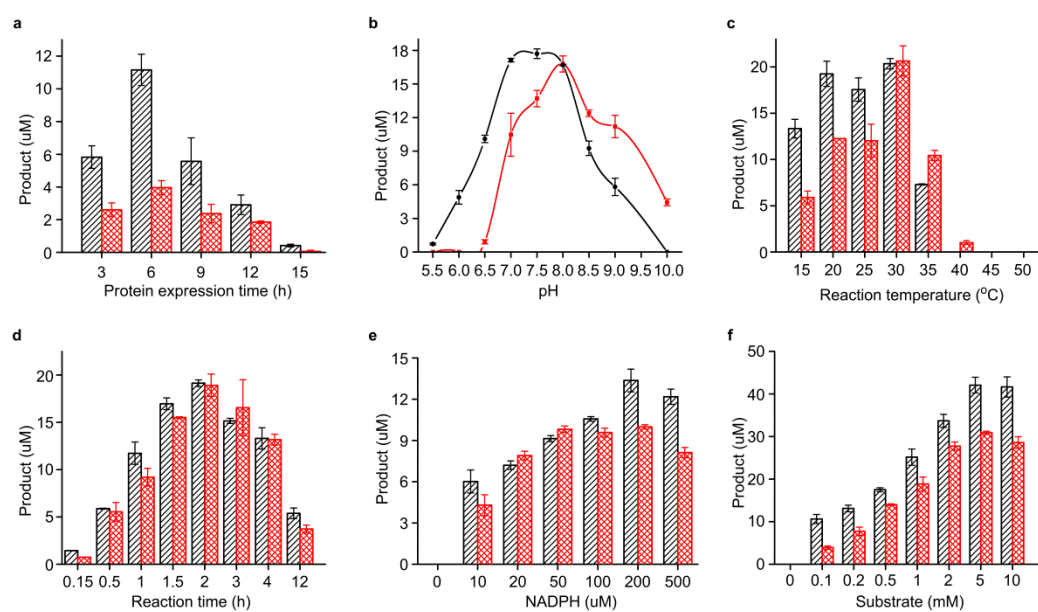


**Figure S9. Catalytic hydroxylation of 7-deoxyloganic acid to form loganic acid by CaCYP72As.** A, HPLC-DAD analyses of the standard loganic acid (blue), secologanic acid (red),

and 7-deoxyloganic acid (green) (*panel VI*) and the reaction mixtures from the whole enzymatic reactions using the yeast cells harbouring the empty vector (*panel I*), boiled CaCYP72A610 (*panel II*), CaCYP72A610 (*panel III*), boiled CaCYP72A565 (*panel IV*), and CaCYP72A565 (*panel V*) as catalysts. **B**, UPLC-DAD-HRESIMS analyses of the enzymatic reaction using CaCYP72As as catalyst. *Panel III*, UPLC-DAD of the enzymatic reaction mixture. The arrows indicate the first ( $[M - H]^-$  at  $m/z$  375.1291, *panel II*) and the second ( $[M - H]^-$  at  $m/z$  373.1134, *panel I*) product analysed by HRESIMS. **C**, HRESIMS (negative ionization mode, *panel I*) of the standard loganic acid and its fragmentation pattern (*panel II*). **D**, HRESIMS ( $[M - H]^-$  at  $m/z$  375.1294, negative ionization mode, *panel I*) of the enzymatic conversion product loganic acid and its fragmentation pattern ( $[M - Glc - H]^-$  at  $m/z$  213.0769, *panel II*) using 7-deoxyloganic acid as substrate. **E**, HRESIMS (negative ionization mode, *panel I*) of the standard secologanic acid and its fragmentation pattern (*panel II*). **F**, HRESIMS ( $[M - H]^-$  at  $m/z$  373.1134, negative ionization mode, *panel I*) of the enzymatic conversion product secologanic acid and its fragmentation pattern ( $[M - Glc - H_2O - H]^-$  at  $m/z$  193.0504, *panel II*) using 7-deoxyloganic acid as substrate.

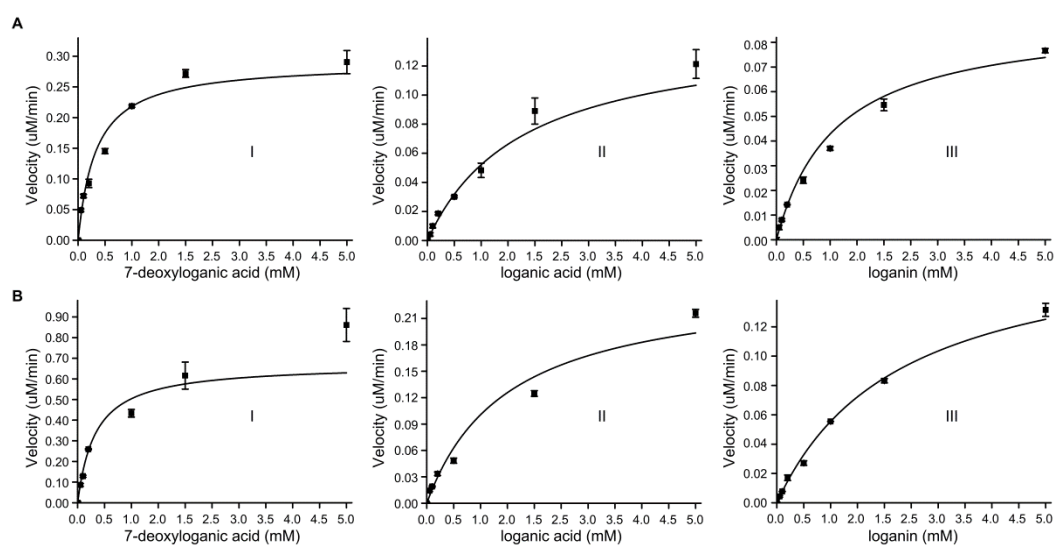


**Figure S10.** Consecutive conversion reactions from 7-deoxyloganic acid to secologanic acid via loganic acid catalysed by CaCYP72A610 (A) and CaCYP72A565 (B). HPLC-DAD analyses of the standard loganic acid (blue), secologanic acid (red), and 7-deoxyloganic acid (green) (*panels V*) and the reaction mixtures from the whole enzymatic reactions using the boiled CaCYP72As (*panels I*) and CaCYP72As as catalysts (*panels II*, 1 hour; *panels III*, 3 hour). *Panels IV*, an additional 1 mM of loganic acid was added to the reaction mixtures of the whole reaction at 1 hr and then the reaction was continued for 2 more hrs.

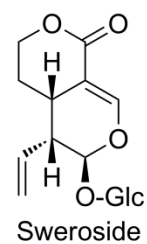
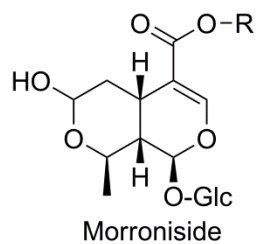
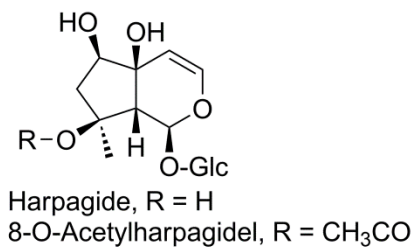
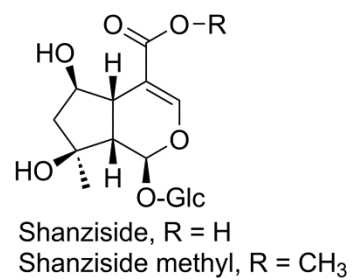
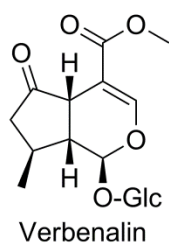
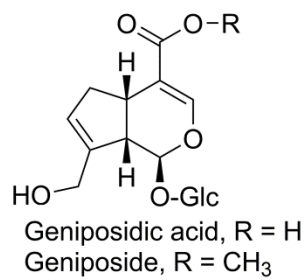


**Figure S11.** The effects of protein overexpression time (a), pH value of the reaction buffer (b), reaction temperature (c), reaction time (d), NADPH concentration (e), and substrate concentration (f) on the catalytic oxidative C – C bond cleavage catalyzed by CaCYP72As. The amount of secologanin is in black columns or line. The production of secologanic acid is indicated in red columns or line.

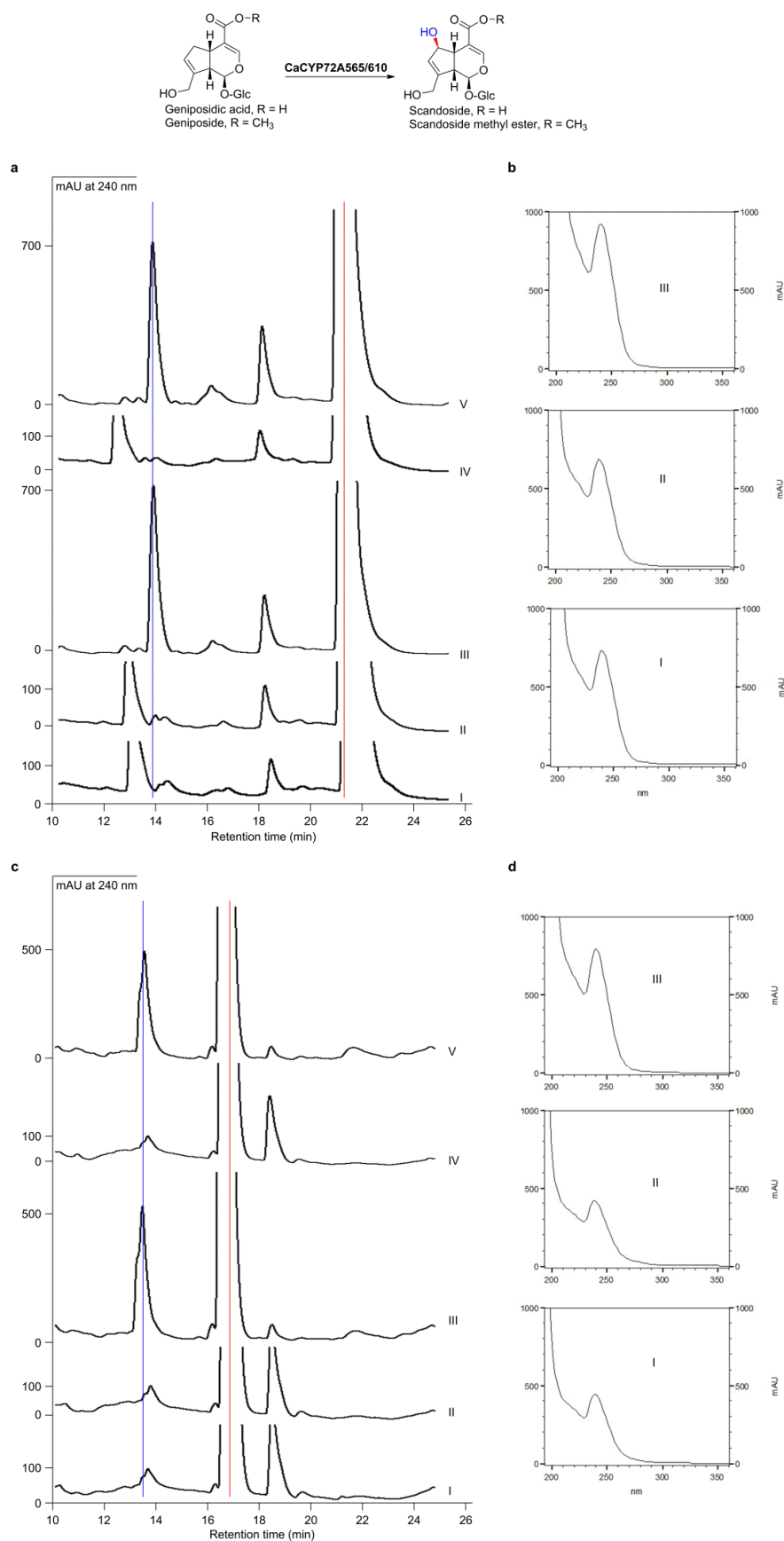




**Figure S12.** The steady-state kinetic parameters for 7-deoxyloganic acid (*panels I*), loganic acid (*panels II*), and loganin (*panels III*) in the microsomal-catalyzed reactions using CaCYP72A610 (A) and CaCYP72A565 (B), respectively.



**Figure S13. Chemical structures of the monoterpene glucosides tested in the present work.**

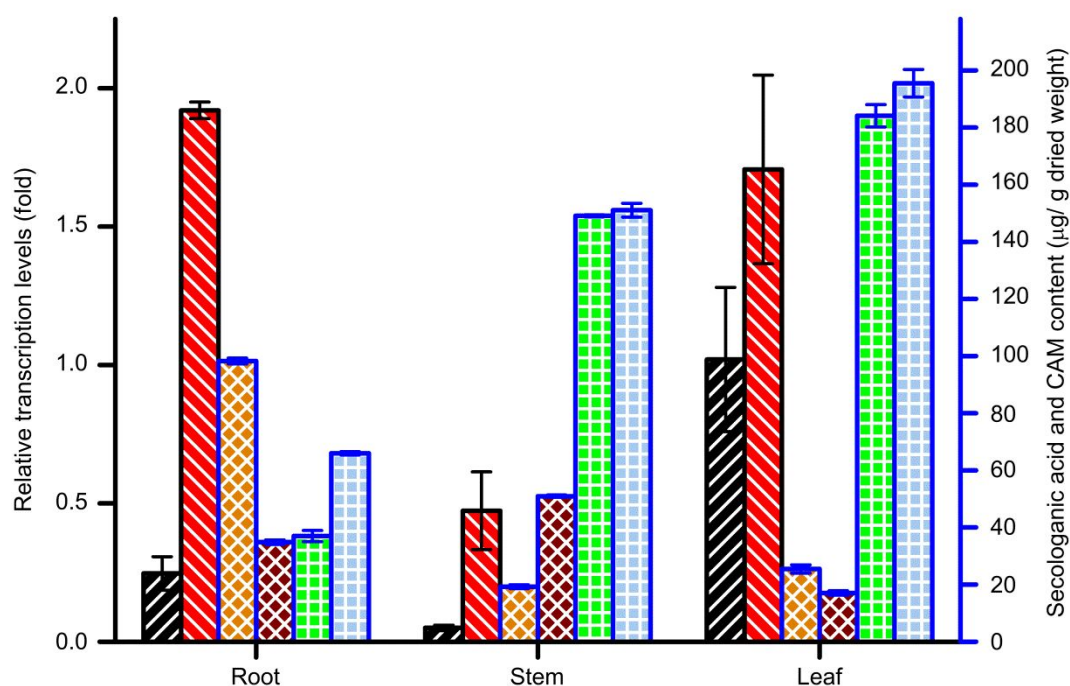


**Figure S14. Catalytic hydroxylation of geniposidic acid and its methyl ester to generate scandoside and its methyl ester by CaCYP72As.** **a**, HPLC-DAD analyses of the reaction

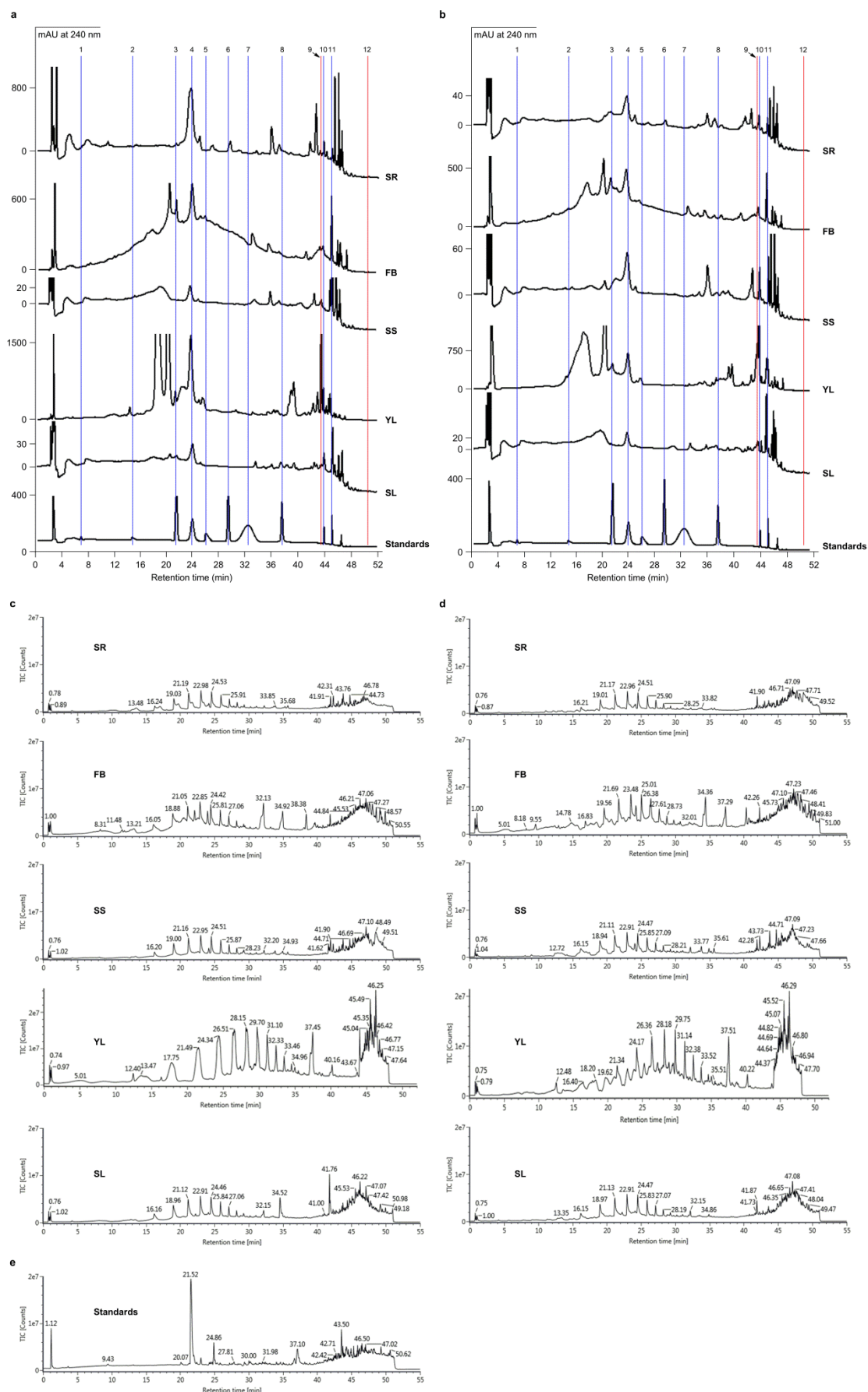
mixture from the whole enzymatic reaction using the supernatant from yeast harboring empty-vector (*panel I*), boiled CaCYP72A610 (*panel II*), CaCYP72A610 (*panel III*), boiled CaCYP72A565 (*panel IV*) and CaCYP72A565 (*panel V*) as catalysts and geniposide as substrate. **b**, The UV spectra of the standard geniposide (*panel III*) and the enzymatic conversion product from CaCYP72A565 (*panel II*) and CaCYP72A610 (*panel I*). **c**, HPLC-DAD analyses of the reaction mixture from the whole enzymatic reaction using the supernatant from yeast harboring empty-vector (*panel I*), boiled CaCYP72A610 (*panel II*), CaCYP72A610 (*panel III*), boiled CaCYP72A565 (*panel IV*) and CaCYP72A565 (*panel V*) as catalysts and geniposidic acid as substrate. **d**, The UV spectra of the standard geniposidic acid (*panel III*) and the enzymatic conversion product from CaCYP72A565 (*panel II*) and CaCYP72A610 (*panel I*). The blue and red lines indicate that these peaks have identical retention time.

Comparing with the related controls, a new product peak was observed from the HPLC-DAD analyses described above, which suggested that the CaCYP72As catalyze a conversion reaction using geniposidic acid or geniposide as substrate. An  $[M - H]^-$  at  $m/z = 389.10955$  and an  $[M + Na]^+$  at  $m/z = 427.10294$  were observed from the CaCYP72As-catalyzed conversion reaction mixture, respectively, which indicated that the two new compounds have one more oxygen atom than the corresponding substrate. To acquire sufficient amount of the compound for NMR analysis, a 4 L of culture was prepared to convert the substrate into the product. A preparative HPLC was introduced to separate the target product by using the identical analytical HPLC condition for monitoring the whole enzymatic reaction with geniposidic acid or geniposide as substrate.

The  $^1H$  NMR data of the product from CaCYP72A-catalyzed hydroxylation using geniposide as substrate are as follows.  $^1H$  NMR ( $CD_3OD$ , 600 MHz)  $\delta$  7.50 (s, 1H), 5.80 (s, 1H), 5.20 (d,  $J = 6.4$  Hz, 1H), 4.68 (d,  $J = 8$  Hz, 1H), 4.54 (dd,  $J = 5, 2$  Hz, 1H), 4.34 (d,  $J = 15$  Hz, 1H), 4.19 (d,  $J = 15$  Hz, 1H), 3.87 (dd,  $J = 12, 2$  Hz, 1H), 3.76 (s, 3H), 3.64 (dd,  $J = 12, 6$  Hz, 1H), 3.23 (t,  $J = 9$  Hz, 1H), 3.05 (t,  $J = 7$  Hz, 1H), and 3.01 (dd,  $J = 5, 1$  Hz, 1H).  $^{13}C$  NMR ( $CD_3OD$ , 125 MHz)  $\delta$  170.51, 154.04, 147.68, 130.25, 110.94, 100.42, 98.40, 82.38, 78.52, 77.99, 74.91, 71.66, 62.79, 61.14, 52.26, 47.26, and 45.63. The  $^1H$  NMR data of the product from CaCYP72A-catalyzed hydroxylation using geniposidic acid as substrate are as follows.  $^1H$  NMR ( $CD_3OD$ , 400 MHz)  $\delta$  7.51 (s, 1H), 5.84 (s, 1H), 5.05 (m, 1H), 4.72 (d,  $J = 7.7$  Hz, 1H), 4.63 (m, 1H), 4.38 (d,  $J = 16.0$  Hz, 1H), 4.21 (d,  $J = 15.1$  Hz, 1H), 3.88 (m, 1H), 3.80 (m, 1H), 3.68 – 3.44 (m), 3.26 – 3.15 (m), 3.03 (dd,  $J = 14.4, 9.1$  Hz, 1H), 2.98 (m, 1H). The NMR data of the products were in accordance with those of scandoside<sup>32</sup> and its methyl ester.<sup>33</sup> The results showed that CaCYP72As catalyze the regio- and stereo-selective  $\beta$ -hydroxylation at C-6 position of geniposidic acid and geniposide to give scandoside and its methyl ester, respectively.



**Figure S15. Relationship of the transcription levels of *CaCYP72As* with the accumulation of secologanic acid and CAM.** The transcription level of *CaCYP72A565* in the leaves was used as the control to compare the relative transcription levels of the *CaCYP72As* in different plant tissues. The transcription levels of *CaCYP72A565* and *CaCYP72A610* are indicated in black and red columns, respectively. The contents of secologanic acid extracted by using method A are shown in orange, while those extracted by method B are shown in wine. For CAM, green (method A), cyan (method B).



**Figure S16. HPLC-DAD and UPLC-HRESIMS analyses of putative biosynthetic**

**intermediates, CAM, and 10-hydroxycamptothecin in *C. acuminata*.** Five different plant tissues, including roots (SR), stems (SS), and leaves (SL) from seedlings, and young leaves (YL) and flower buds (FB) from wild mature *C. acuminata* plant, were collected and extracted with two different solvents system, CH<sub>3</sub>CN : H<sub>2</sub>O = 7 : 3 (vol. / vol., Method A) and CH<sub>3</sub>CH<sub>2</sub>OH : H<sub>2</sub>O = 7 : 3 (vol. / vol., Method B). HPLC-DAD of the standards and the extracts by using Method A (**a**) and Method B (**b**). UPLC-HRESIMS of the extracts by using Method A (**c**) and Method B (**d**). **e**, the UPLC-MS of the standards. **1**, 5-hydroxytryptamine; **2**, tryptamine; **3**, loganic acid; **4**, secologanic acid; **5**, strictosidine; **6**, loganin; **7**, secologanin; **8**, 7-deoxyloganic acid; **9**, 10-hydroxygeraniol; **10**, 10-hydroxycamptothecin; **11**, CAM; **12**, geraniol. The blue lines indicate that these standards can be detected at 240 nm. The red lines indicate that compounds **9** and **12** cannot be detected at 240 nm but can be detected at 210 nm.

## Supplementary references

1. Yang, Y.; Pu, X.; Qu, X.; Chen, F.; Zhang, G.; Luo, Y. Enhanced production of camptothecin and biological preparation of N<sup>1</sup>-acetylkynuramine in *Camptotheca acuminata* cell suspension cultures. *Appl. Microbiol. Biotechnol.* **2017**, *101*, 4053-4062.
2. Qu, X.; Pu, X.; Chen, F.; Yang, Y.; Yang, L.; Zhang, G.; Luo, Y. Molecular cloning, heterologous expression, and functional characterization of an NADPH-cytochrome P450 reductase gene from *Camptotheca acuminata*, a camptothecin-producing plant. *PLoS ONE* **2015**, *10*, e0135397.
3. Chen, F.; Li, W.; Jiang, L.; Pu, X.; Yang, Y.; Zhang, G.; Luo, Y. Functional characterization of a geraniol synthase-encoding gene from *Camptotheca acuminata* and its application in the production of geraniol in *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 1281-1292.
4. Yang, L.; Jiang, L.; Li, W.; Yang, Y.; Zhang, G.; Luo, Y. A homomeric geranyl diphosphate synthase-encoding gene from *Camptotheca acuminata* and its combinatorial optimization for production of geraniol in *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* **2017**, *44*, 1431-1441..
5. Pompon, D.; Louerat, B.; Bronine, A.; Urban, P. Yeast expression of animal and plant P450 in optimized redox environments. *Method. Enzymol.* **1996**, *272*, 51-64.
6. Sadre, R.; Magallanes-Lundback, M.; Pradhan, S.; Salim, V.; Mesberg, A.; Jones, A. D.; DellaPenna, D. Metabolite diversity in alkaloid biosynthesis: a multilane (diastereomer) highway for camptothecin synthesis in *Camptotheca acuminata*. *Plant Cell* **2016**, *28*, 1926-1944.
7. Montoro, P.; Maldini, M.; Piacente, S.; Macchia, M.; Pizza, C. Metabolite fingerprinting of *Camptotheca acuminata* and the HPLC-ESI-MS/MS analysis of camptothecin and related alkaloids. *J. Pharma. Biomed. Anal.* **2010**, *51*, 405-415.
8. Irmeler, S.; Schröder, G.; St-Pierre, B.; Crouch, N. P.; Hotze, M.; Schmidt, J.; Strack, D.; Matern, U.; Schröder, J. Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J.* **2000**, *24*, 797-804.
9. de Bernonville, T. D.; Foureau, E.; Parage, C.; Lanoue, A.; Clastre, M.; Londono, M. A.; Oudin, A.; Houillé, B.; Papon, N.; Besseau, S.; Glévarec, G.; Atehortúa, L.; Giglioli-Guivarc'h, N.; St-Pierre, B.; De Luca, V.; O'Connor, S. E.; Courdavault, V. Characterization of a second secologanin synthase isoform producing both secologanin and secoxyloganin allows enhanced *de novo* assembly of a *Catharanthus roseus* transcriptome. *BMC Genomics* **2015**, *16*, 619.
10. Sun, Y.; Luo, H.; Li, Y.; Sun, C.; Song, J.; Niu, Y.; Zhu, Y.; Dong, L.; Lv, A.; Tramontano, E.; Chen, S. Pyrosequencing of the *Camptotheca acuminata* transcriptome reveals putative genes involved in camptothecin biosynthesis and transport. *BMC Genomics* **2011**, *12*, 533.
11. Góngora-Castillo, E.; Childs, K. L.; Fedewa, G.; Hamilton, J. P.; Liscombe, D. K.; Magallanes-Lundback, M.; Mandadi, K. K.; Nims, E.; Runguphan, W.; Vaillancourt, B.; Varbanova-Herde, M.; DellaPenna, D.; McKnight, T. D.; O'Connor, S.; Buell, C. B. Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLoS ONE* **2012**, *7*, e52506.



12. Nelson, D. R.; Strobel, H. W. On the membrane topology of vertebrate cytochrome P-450 proteins. *J. Biol. Chem.* **1988**, *263*, 6038-6050.
13. Barnes, H.J.; Arlotto, M.P.; Waterman, M.R. Expression and enzymatic activity of recombinant cytochrome P450, 17 $\alpha$ -hydroxylase in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **1991**, *88*, 5597-5601.
14. Gotoh, O.; Fujii-Kuriyama, Y. Evolution, structure, and gene regulation of cytochrome P-450. In: Ruckpaul, K.; Rein, H. (Eds.) *Basis and Mechanisms of Regulation of Cytochrome P450*, *Frontiers in Biotransformation*. Taylor & Francis, London, New York, Philadelphia, **1989**, pp 195-243.
15. Durst, F.; Nelson, D. R. Diversity and evolution of plant P450 and P450-reductases. *Drug Metab. Drug Interact.* **1995**, *12*, 189-206.
16. Wachenfeldt, C. V.; Johnson, E. F. *Structures of Eukaryotic Cytochrome P450 Enzymes*. Springer, New York, **1995**, pp 183-223.
17. Strushkevich, N.; Mackenzie, F.; Cherkesova, T.; Grabovec, I.; Usanov, S.; Park, H. W. Structural basis for pregnolone biosynthesis by the mitochondrial monooxygenase system. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10139-10143.
18. Li, S.; Tietz, D. R.; Rutaganira, F. U.; Kells, P. M.; Anzai, Y.; Kato, F.; Pochapsky, T. C.; Sherman, D. H.; Podust, L. M. Substrate recognition by the multifunctional cytochrome P450 MycG in mycinamicin hydroxylation and epoxidation reactions. *J. Biol. Chem.* **2012**, *287*, 37880-37890.
19. Miyanaga, A.; Takayanagi, R.; Furuya, T.; Kawamata, A.; Itagaki, T.; Iwabuchi, Y.; Kanoh, N.; Kudo, F.; Eguchi, T. Substrate recognition by a dual-function P450 monooxygenase GfsF involved in FD-891 biosynthesis. *ChemBioChem* **2017**, *18*, 2179-2187.
20. Carlson, J. C.; Li, S.; Gunatilleke, S. S.; Anzai, Y.; Burr, D. A.; Podust, L. M.; Sherman, D. H. Tirandamycin biosynthesis is mediated by co-dependent oxidative enzyme. *Nat. Chem.* **2011**, *3*, 628-633.
21. Zocher, G.; Richter, M. E. A.; Mueller, U.; Hertweck, C. Structural fine-tuning of a multifunctional cytochrome P450 monooxygenase. *J. Am. Chem. Soc.* **2011**, *133*, 2292-2302.
22. Nett, R. S.; Montanares, M.; Marcassa, A.; Lu, X.; Nagel, R.; Charles, T. C.; Hedden, P.; Rojas, M. C.; Peters, R. J. Elucidation of gibberellin biosynthesis in bacteria reveals convergent evolution. *Nat. Chem. Biol.* **2017**, *13*, 69-74.
23. Moody, S. C.; Zhao, B.; Lei, L.; Nelson, D. R.; Mullins, J. G. L.; Waterman, M. R.; Kelly, S. L.; Lamb, D. C. Investigating conservation of the albaflavenone biosynthetic pathway and CYP170 bifunctionality in *Streptomyces*. *FEBS J.* **2012**, *279*, 1640-1649.
24. Strushkevich, N.; Mackenzie, F.; Cherkesova, T.; Grabovec, I.; Usanov, S.; Park, H. W. Structural basis for pregnolone biosynthesis by the mitochondrial monooxygenase system. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10139-10143.
25. Petrunak, E. M.; Rogers, S. A.; Aubé, J.; Scott, E. E. Structural and functional evaluation of clinically-relevant inhibitors of steroidogenic cytochrome P45017A1. *Drug Metab. Dispos.* **2017**, *45*, 635-645.
26. Sung, P.; Huang, F.; Do, Y.; Huang, P. Functional expression of geraniol 10-hydroxylase reveals its dual function in the biosynthesis of terpenoid and phenylpropanoid. *J. Agric. Food. Chem.* **2011**, *59*, 4637-4643.

27. El-Awaad, I.; Bocola, M.; Beuerle, T.; Liu, B.; Beerhues, L. Bifunctional CYP81AA proteins catalyse identical hydroxylations but alternative regioselective phenol coupling in plant xanthone biosynthesis. *Nat. Commun.* **2016**, *7*, 11472.
28. Dang, T. T.; Franke, J.; Tatsis, E.; O'Connor, S. E. Dual catalytic activity of a cytochrome P450 controls bifurcation at a metabolic branch point of alkaloid biosynthesis in *Rauwolfia serpentina*. *Angew. Chem. Int. Ed.* **2017**, *56*, 9440-9444.
29. Kim, T.; Hwang, J.; Kim, Y.; Joo, S.; Chang, S. C.; Lee, J. S.; Takatsuto, S.; Kim, S. Arabidopsis CYP85A2, a cytochrome P450, mediates the Baeyer-Villiger oxidation of castasterone to brassinolide in Brassinosteroid biosynthesis. *Plant J.* **2005**, *17*, 2397-2412.
30. Sohrabi, R.; Huh, J.; Badieyan, S.; Rakotondraibe, L. H.; Kliebenstein, D. J.; Sobrado, P.; Tholl, D. In planta variation of volatile biosynthesis: An alternative biosynthetic route to the formation of the pathogen-induced volatile homoterpene DMNT via triterpene degradation in *Arabidopsis* roots. *Plant Cell* **2015**, *27*, 874-890.
31. Prall, W.; Hendy, O.; Thornton, L. E. Utility of a phylogenetic perspective in structural analysis of CYP72A enzymes from flowering plants. *PLoS ONE* **2016**, *11*, e0163024.
32. Kim, D.; Lee, H.; Oh, Y.; Kim, M.; Kim, S.; Jeong, T.; Baek, N. Iridoid glycosides isolated from *Oldenlandia diffusa* inhibit LDL-oxidation. *Arch. Pharm. Res.* **2005**, *28*, 1156-1160.
33. Otsuka, H.; Yoshimura, K.; Yamasaki, K.; Cantoria, M. C. Isolation of 10-O-acyl iridoid glucosides from a Philippine medicinal plant, *Oldenlandia corymbosa* L. (Rubiaceae). *Chem. Pharm. Bull.* **1991**, *39*, 2049-2052.

**Table S1. List of the primers sequences used in this work**

Primer code	Sequence (5'-3')	Direction	Application
Caa_locus_1905-F	ATGGAAATTCAAATGGACGTTCTCTAC	Forward	ORF cloning
Caa_locus_1905-R	TCAACACTTGATCTTGCGGAAGACTTG	Reverse	
Caa_locus_133-F	ATGAAAATGGAGGTAATGCACAT	Forward	
Caa_locus_133-R	TCAACATTTGAGCCTGCGGAAGA	Reverse	
Caa_locus_10805-F	ATGGAAATAGGAAACAACCTTCATTG	Forward	
Caa_locus_10805-R	CTAGAGTTTACGTAAAATCAAGTGA	Reverse	
Caa_locus_18068-F	TTCAATTGGATGTGGTTGAGAC	Forward	3'-RACE
Caa_locus_18068-R	TGGCGGTACAATCTTAGAACCTCA	Reverse	
Caa_locus_18068-3'RACE-GSP1	GCTGGGCAAGAGACCACCTCAGTTTACTTG	Forward	
Caa_locus_18068-3'RACE-GSP2	ATGATTTGTATGAGGTTCTAAGATTGTAC	Forward	
Caa_locus_18068-Full-F	TTCAATTGGATGTGGTTGAGAC	Forward	ORF cloning
Caa_locus_18068-Full-R	CTAGATCTTTTTTAAAATTAAGTGAGCC	Reverse	
Caa_locus_1905-CT-F	GGGGTACCGAACACAATGTCTGCGATTGCAGCCGCGGTTG	Forward	Overexpression
Caa_locus_1905-CT-R	ATAGTTTAGCGGCCGCCAACACTTGATCTTGCGGAAGAC	Reverse	
Caa_locus_133-CT-F	CGCGGATCCAAACACAATGTCTAAAATGGAGGTAATGCAC	Forward	
Caa_locus_133-CT-R	ATAGTTTAGCGGCCGCCAACATTTGAGCCTGCGGAAG	Reverse	
Caa_locus_10805-CT-F	CGGGATCCAAACACAATGTCTGCTATTTTCATTG	Forward	
Caa_locus_10805-CT-R	ATAGTTTAGCGGCCGCCAGAGTTTACGTAAAATCAAG	Reverse	
Caa_locus_18068-CT-F	ATAGTTTAGCGGCCGCCAGATCTTTTTTAAAATTAAG	Forward	
Caa_locus_18068-CT-R	ATAGTTTAGCGGCCGCCACTAGATCTTTTTTAAAATT	Reverse	
rt-Caa_locus_1905-F	TTTGTTTGATCTTTAATTTGCTGCTG	Forward	qRT-PCR analysis
rt-Caa_locus_1905-R	ACCACAACCTTTATTAATCAAGGAGAGAAC	Reverse	
rt-Caa_locus_133-F	GCCTATTAATTTGTTATGTCCCCTAGC	Forward	
rt-Caa_locus_133-R	AGAAGGATGAAGGGTAACAATGCTAAC	Reverse	
rt-Actin-F	CATCCCTCAGCACCTTCCAG	Forward	
rt-Actin-R	CTGGGATGCAAAAACCGCC	Reverse	

**Table S2. List of the metabolites extracted by two extraction methods from different plant tissues of *C. acuminata***

Putative metabolites	Molecular formula	Retention time	Theoretical weight	Experimental weight	Method A					Method B					Standard (Y / N)
					SR	SS	SL	YL	FB	SR	SS	SL	YL	FB	
5-Hydroxytryptamine	C <sub>10</sub> H <sub>12</sub> N <sub>2</sub> O	3.60	176.095	177.1081	-	-	-	-	-	-	-	-	-	-	Y
Tryptamine	C <sub>10</sub> H <sub>12</sub> N <sub>2</sub>	8.11	160.1005	161.1068	+	+	+	+	+	+	+	+	+	+	Y
Loganic acid	C <sub>16</sub> H <sub>24</sub> O <sub>10</sub>	11.33	376.1370	399.1264	-	-	+	+	+	-	-	-	+	+	Y
Secologanic acid	C <sub>16</sub> H <sub>22</sub> O <sub>10</sub>	13.37	374.1213	397.1107	-	+	+	+	+	+	+	+	+	+	Y
Loganin	C <sub>17</sub> H <sub>26</sub> O <sub>10</sub>	20.08	390.1526	413.1419	-	-	-	-	-	-	-	-	-	-	Y
Strictosamide diol	C <sub>26</sub> H <sub>32</sub> N <sub>2</sub> O <sub>10</sub>	20.15	532.2057	533.2125	+	+	+	+	+	+	+	+	+	+	N
Strictosamide ketolactam	C <sub>26</sub> H <sub>30</sub> N <sub>2</sub> O <sub>10</sub>	20.53	530.1901	530.1886	+	+	+	+	-	+	+	+	+	-	N
Strictosamide epoxide	C <sub>26</sub> H <sub>30</sub> N <sub>2</sub> O <sub>9</sub>	21.57	514.1951	515.2024	+	+	+	+	+	+	+	+	+	+	N
Secologanin	C <sub>17</sub> H <sub>24</sub> O <sub>10</sub>	22.09	388.1370	411.1262	-	-	-	-	-	-	-	-	-	-	Y
Strictosidinic acid	C <sub>26</sub> H <sub>32</sub> N <sub>2</sub> O <sub>9</sub>	24.12	516.2108	517.2179	+	+	+	+	+	+	+	+	+	+	N
7-Deoxyloganic acid	C <sub>16</sub> H <sub>24</sub> O <sub>9</sub>	24.34	360.1420	383.1306	-	-	-	-	-	-	-	-	-	-	Y
Strictosidine	C <sub>27</sub> H <sub>34</sub> N <sub>2</sub> O <sub>9</sub>	25.50	530.2264	531.2372	-	-	-	-	-	-	-	-	-	-	Y
10-Hydroxygeraniol	C <sub>10</sub> H <sub>18</sub> O <sub>2</sub>	29.12	170.1307	193.1199	-	-	-	-	-	-	-	-	-	-	Y
10-Hydroxycamptothecin	C <sub>20</sub> H <sub>16</sub> N <sub>2</sub> O <sub>5</sub>	32.04	364.1059	365.113	+	+	+	+	+	+	+	+	+	+	Y
Pumiloside	C <sub>26</sub> H <sub>28</sub> N <sub>2</sub> O <sub>9</sub>	32.06	512.1795	512.1796	+	+	+	+	+	+	+	+	+	+	N
Camptothecin	C <sub>20</sub> H <sub>16</sub> N <sub>2</sub> O <sub>4</sub>	34.06	348.1110	349.119	+	+	+	+	+	+	+	+	+	+	Y
Strictosamide	C <sub>26</sub> H <sub>30</sub> N <sub>2</sub> O <sub>8</sub>	35.27	498.2002	499.2077	+	+	+	+	+	+	+	+	+	+	N
Deoxypumiloside	C <sub>26</sub> H <sub>28</sub> N <sub>2</sub> O <sub>8</sub>	35.67	496.1846	496.1854	+	+	+	+	-	+	+	+	+	-	N
Geraniol	C <sub>10</sub> H <sub>18</sub> O	41.34	154.1358	172.1696	-	-	-	-	-	-	-	-	-	-	Y
7-Deoxyloganic acid	C <sub>10</sub> H <sub>14</sub> O <sub>4</sub>	-	198.0892	-	-	-	-	-	-	-	-	-	-	-	N

**GenBank accession numbers:** The sequences of the *CaCYP72As* described in this work were deposited in the NCBI GenBank under the following accession numbers: CaCYP72A565 (MH763569), and CaCYP72A610 (MH763570).