PCP
PLANT & CELL PHYSIOLOGY

Rapid Paper

Allylic Hydroxylation Activity Is a Source of Saponin Chemodiversity in the Genus *Glycyrrhiza*

Much Z. Fanani [1,†], Satoru Sawai[1,2,3,4,†], Hikaru Seki [1,2], Masato Ishimori[3], Kiyoshi Ohyama[2,5,7], Ery O. Fukushima [1,6], Hiroshi Sudo[4,8], Kazuki Saito[2,3,*] and Toshiya Muranaka [1,2,*]

[1]Department of Biotechnology, Graduate School of Engineering, Osaka University, Suita, 565-0871 Japan
[2]RIKEN Center for Sustainable Resource Science, Yokohama, 230-0045 Japan
[3]Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, 260-8675 Japan
[4]Tokiwa Phytochemical Co., Ltd, Chiba, 285-0801 Japan
[5]Department of Chemistry and Materials Science, Tokyo Institute of Technology, Tokyo, 152-8551 Japan
[6]Translational Plant Research Group, Universidad Regional Amazónica IKIAM, Tena, Ecuador
[7]Present address: Japan Tobacco Inc., Leaf Tobacco Research Center, 1900, Idei, Oyama, Tochigi 323-0808, Japan.
[8]Present address: School of Pharmacy and Pharmaceutical Sciences, Hoshi University, 2-4-41, Ebara, Shinagawa-ku, Tokyo 142-8501, Japan.
[†]These authors contributed equally to this work.
*Corresponding authors: Kazuki Saito, E-mail, kazuki.saito@riken.jp; Fax, +81-45-503-9489; Toshiya Muranaka, E-mail, muranaka@bio.eng.osaka-u.ac.jp, Fax, +81-6-6879-7426.
(Received 21 October 2020; Accepted 18 December 2020)

Licorice (*Glycyrrhiza*) produces glycyrrhizin, a valuable triterpenoid saponin, which exhibits persistent sweetness and broad pharmacological activities. In the genus *Glycyrrhiza*, three species, *Glycyrrhiza uralensis*, *Glycyrrhiza glabra* and *Glycyrrhiza inflata*, produce glycyrrhizin as their main triterpenoid saponin, which has a ketone group at C-11. Other *Glycyrrhiza* species produce mainly oleanane-type saponins, which harbor homoannular or heteroannular diene structures that lack the C-11 ketone. Although the glycyrrhizin biosynthetic pathway has been fully elucidated, the pathway involving saponins with diene structures remains unclear. CYP88D6 from *G. uralensis* is a key enzyme in glycyrrhizin biosynthesis, catalyzing the sequential two-step oxidation of β-amyrin at position C-11 to produce 11-oxo-β-amyrin. In this study, we evaluated the functions of CYP88D6 homologs from the glycyrrhizin-producing species *G. glabra* and *G. inflata* and from the non-glycyrrhizin-producing species *Glycyrrhiza pallidiflora* and *Glycyrrhiza macedonica*, using yeast engineered to supply β-amyrin as a substrate. Yeast expressing *CYP88D6* homologs from glycyrrhizin-producing species produced 11-oxo-β-amyrin. However, yeast expressing *CYP88D6* homologs (such as CYP88D15) from the non-glycyrrhizin-producing *Glycyrrhiza* species accumulated olean-9(11),12-dien-3β-ol and oleana-11,13(18)-dien-3β-ol; these diene compounds are non-enzymatic or yeast endogenous enzymatic dehydration derivatives of 11α-hydroxy-β-amyrin, a direct reaction product of CYP88D15. These results suggest that the activities of CYP88D6 homologs, particularly their ability to catalyze the second oxidation, could influence glycyrrhizin productivity and diversify the chemical structures of saponins in *Glycyrrhiza* plants. A synthetic biological approach to engineer CYP88D15 could enable the production of pharmacologically active saponins with diene structures, such as saikosaponins, whose biosynthetic pathways have yet to be fully characterized.

Keywords: Allylic hydroxylation • Chemodiversity • CYP88D subfamily • Cytochrome P450 • Licorice • Triterpenoid.

Accession number: The nucleotide sequences reported in this article have been submitted to GenBank under accession numbers MW071146, MW071147 and MW071148.

Introduction

Plants produce a diverse array of specialized metabolites in response to biotic and abiotic stresses. The chemical structures of these compounds can vary greatly depending on plant taxonomy, often exhibiting order-, family-, genus-, or species-specific properties (Moghe and Kruse 2018). Plant specialized metabolites have important uses in our daily lives, functioning as health-promoting molecules (Butelli et al. 2008), pigments (Polturak et al. 2016), flavors (Schwab et al. 2008), perfumes (Caniard et al. 2012) and medicines (Wani et al. 1971). Approximately one quarter of drugs prescribed worldwide are derived from plants (Rates 2001), and their uses in society are increasing, not only to improve human health, but also for renewable energy and sustainable green chemistry (Facchini et al. 2012). However, the mechanisms that drive the vast chemodiversity observed in specialized plant metabolites are poorly understood.

Licorice, *Glycyrrhiza* species belonging to the Leguminosae family, is an important medicinal plant with high economic value. Since ancient times, licorice has been used to treat

Editor-in-Chief's Choice

diseases (Gibson 1978, Shibata 2000). The underground parts of licorice, i.e. the roots and stolons, are used in traditional medicine and are where the main bioactive compound, glycyrrhizin, accumulates (Hayashi et al. 2000). Glycyrrhizin is a triterpenoid saponin known for its long-lasting sweetness (170 times sweeter than sucrose) (Mizutani et al. 1994), and it is responsible for the broad pharmaceutical activities of licorice (Kao et al. 2014, Wang et al. 2015, Yang et al. 2015). Thus, glycyrrhizin is a high-value specialized metabolite.

Of the 21 accepted species constituting the genus *Glycyrrhiza* (The Plant List 2013), only three (*Glycyrrhiza uralensis*, *Glycyrrhiza glabra* and *Glycyrrhiza inflata*) are known to produce glycyrrhizin. Other *Glycyrrhiza* species, including *Glycyrrhiza pallidiflora*, *Glycyrrhiza macedonica* and *Glycyrrhiza yunnanensis*, primarily accumulate homoannular or heteroannular diene-type saponins, such as yunganoside L2 and macedonoside C (**Fig. 1**) (Ohtani, et al. 1994, Hayashi et al. 2000). The chemical structure of glycyrrhizin contains a ketone group at C-11, whereas yunganoside L2 and macedonoside C have homoannular diene and heteroannular diene structures, respectively, in place of the C-11 ketone.

Cytochrome P450 monooxygenases (CYPs) contribute to triterpene chemodiversity by introducing various functional groups, such as hydroxyl, carbonyl, carboxyl and epoxy moieties (Ghosh 2017). CYP88D6, a CYP from *G. uralensis* that belongs to the CYP88D subfamily, catalyzes a sequential two-step oxidation reaction at the C-11 position of β-amyrin to first produce 11α-hydroxy-β-amyrin and then 11-oxo-β-amyrin during glycyrrhizin biosynthesis (Seki et al. 2008). The enzymes involved in glycyrrhizin biosynthesis, including CYP88D6, have previously been identified (**Fig. 1**); e.g. CYP72A154 is a C-30 oxidase (Seki et al. 2011), licorice cellulose synthase-like enzyme, named GuCSyGT (Chung et al. 2020) or GuCsl (Jozwiak et al. 2020), is a 3-*O*-glucuronosyltransferase and GuUGT73P12 is a uridine diphosphate-dependent glycosyltransferase that catalyzes the second glucuronic acid transfer (Nomura et al. 2019). However, the biosynthetic mechanisms of homoannular and heteroannular diene-type saponins are poorly understood.

Here, we show that CYP88D15, a CYP88D6 homolog from non-glycyrrhizin-producing *Glycyrrhiza* species, catalyzes allylic hydroxylation at C-11 of β-amyrin and that subsequent non-enzymatic or yeast-endogenous-enzymatic dehydration of the allylic hydroxyl group acts as a trigger to generate homoannular and heteroannular diene structures. The expression of *CYP88D15* in β-amyrin-producing yeast yields oleana-9(11),12-dien-3β-ol (**3**) and oleana-11,13(18)-dien-3β-ol (**4**) as homoannular and heteroannular diene compounds, respectively. Furthermore, mutation analyses of the CYP88D6 homolog indicate that M221 and F300 residues are essential for the second oxidation at C-11 during glycyrrhizin production.

Results

Non-glycyrrhizin-producing species have distinct CYP88D6 homologs

We initially hypothesized that the oxidation activities of CYP88D6 homologs influence glycyrrhizin production in *Glycyrrhiza* plants. To evaluate their oxidation activities, we isolated *CYP88D6* homologs from several *Glycyrrhiza* species. Based on their chemotaxonomy and the availability of plant material in Japan, four *Glycyrrhiza* species were selected: *G. pallidiflora* and *G. macedonica* as non-glycyrrhizin-producing *Glycyrrhiza* species and *G. glabra* and *G. inflata* as glycyrrhizin-producing species. We also included *CYP88D6* from *G. uralensis* (*Gu*CYP88D6) to complete the set of three known glycyrrhizin-producing species.

To assign names to the newly isolated CYPs, the protein sequences of the CYP88D6 homologs were submitted to a nomenclature committee (**https://drnelson.uthsc.edu/CytochromeP450.html**; Nelson 2009). CYP88D6 homologs from glycyrrhizin-producing species shared high sequence identity (98.58%) with *Gu*CYP88D6 and thus were named CYP88D6 in both *G. glabra* (*Gg*CYP88D6) and *G. inflata* (*Gi*CYP88D6) (**Supplementary Table S1**). The CYP88D6 homologs from non-glycyrrhizin-producing *Glycyrrhiza* species displayed lower sequence identities to *Gu*CYP88D6 (81.01%) and thus were given a different isoform number: CYP88D15 in both *G. pallidiflora* (*Gp*CYP88D15) and *G. macedonica* (*Gm*CYP88D15). For further analyses, we used *Gp*CYP88D15 as a representative of the non-glycyrrhizin-producing *Glycyrrhiza* species because *Gp*CYP88D15 and *Gm*CYP88D15 share identical nucleotide and protein sequences (**Supplementary Table S1**).

To gain insight into the evolutionary relationships among members of the CYP88D subfamily, we performed multiple alignment and phylogenetic analyses. A heme-binding cysteine residue and a glutamate residue involved in electron-chain relay were found to be conserved throughout the subfamily (**Supplementary Fig. S1**). Although the CYP88D6 homologs from *Glycyrrhiza* species were grouped together in the *Glycyrrhiza* clade in the phylogenetic analysis, they were divided into two subgroups, the glycyrrhizin-producing and non-glycyrrhizin-producing groups (**Fig. 2**). These results suggest that the enzymatic function of CYP88D15 from non-glycyrrhizin-producing *Glycyrrhiza* species may differ from that of CYP88D6 from glycyrrhizin-producing species.

Engineered yeast expressing *GpCYP88D15* predominantly yield dehydrated products, not ketolated 11-oxo-β-amyrin

To investigate the enzymatic activity of CYP88D6 homologs on the substrate β-amyrin, *CYP88D6* homologs were each expressed under the control of a galactose-inducible promoter in yeast engineered to produce β-amyrin (Seki et al. 2008). After the 4 d of the induction culture with galactose, β-amyrin [compound **1**, retention time (Rt) = 23.2 min, *m*/*z* 498] was detected in all yeast strains, but compounds **2**, **3** and **4** were uniquely detected in yeast expressing *CYP88D6* homologs; this indicated that the formation of compounds **2**, **3** and **4** was dependent on the *CYP88D6* homolog (**Fig. 3A**). Compounds **2**, **3** and **4** had identical Rts and mass spectra as those of the following authentic standards: 11-oxo-β-amyrin (compound **2**, Rt = 31.1 min, *m*/*z* 512), oleana-9(11),12-dien-3β-ol (compound **3**, Rt = 21.3 min, *m*/*z* 496) and oleana-11,13(18)-dien-3β-ol (compound **4**, Rt =

23.1 min, *m/z* 496) (**Fig. 3B**). Since in the case of the induction period shortened to 2 d as the previous report (Seki et al. 2008), the β-amyrin-producing yeast expressing *CYP88D6* homologs (*GuCYP88D6* and *GpCYP88D15*) accumulated 11α-hydroxy-β-amyrin (**Supplementary Fig. S2**), the homoannular diene compound (**3**) and the heteroannular diene compound (**4**) may have been derived from the dehydration of 11α-hydroxy-β-

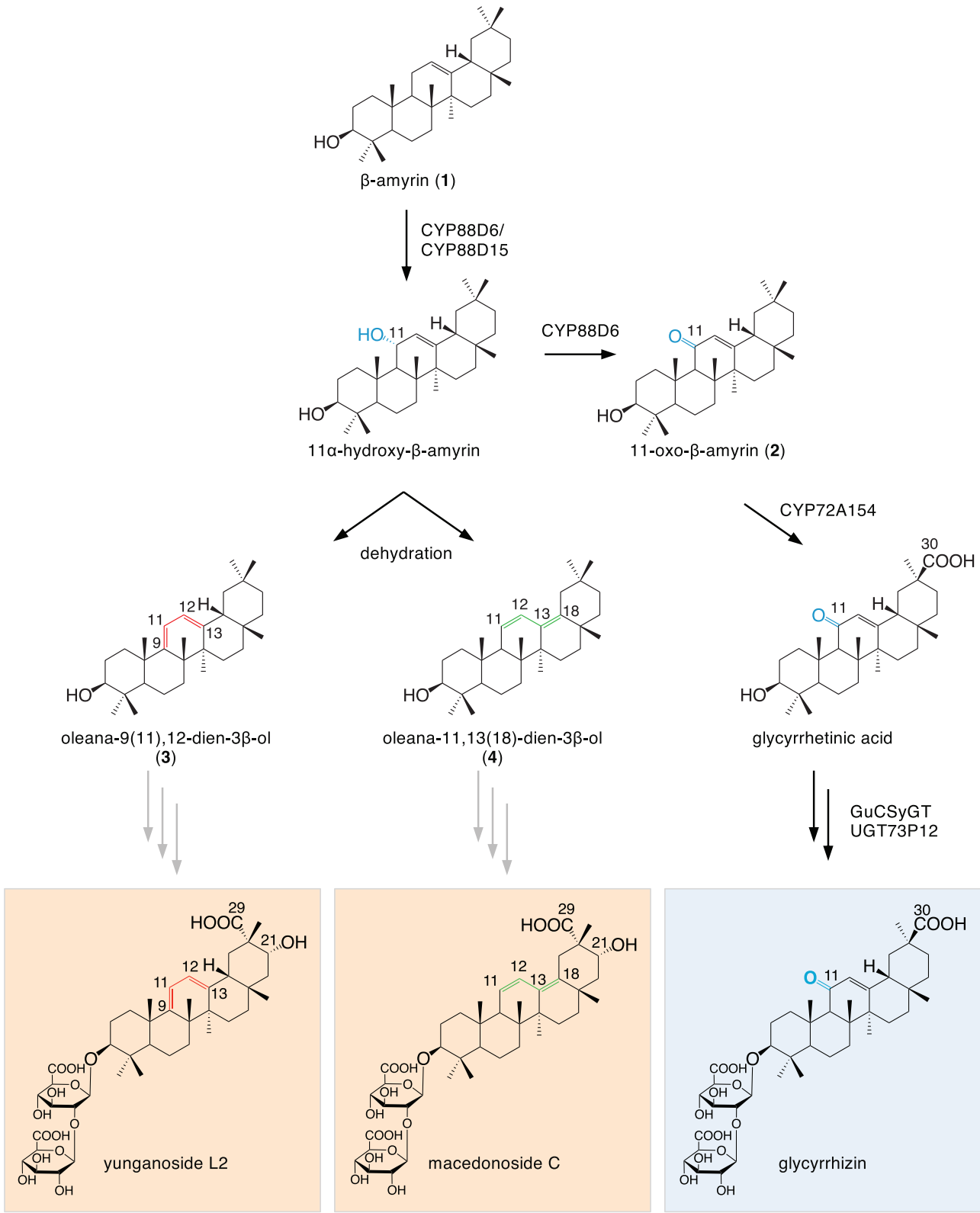

Fig. 1 Saponin chemodiversity in *Glycyrrhiza* species and associated biosynthetic enzymes. Black arrows indicate known biosynthetic enzymes, and gray arrows indicate uncharacterized biosynthetic enzymes.

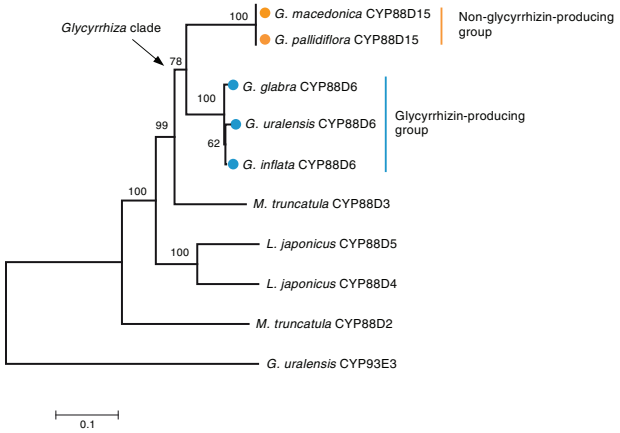

Fig. 2 Phylogenetic analysis of the CYP88D subfamily. The filled circles indicate the major saponins produced by each species: blue for glycyrrhizin-producing species and orange for non-glycyrrhizin-producing *Glycyrrhiza* species. CYP93E3 from *G. uralensis* was used as an outgroup.

amyrin, which possesses an allylic hydroxyl group at C-11 (Fig. 3C).

To evaluate product specificity, the relative peak area of each compound was calculated by comparison with the peak area of spiked internal standards (Fig. 3D). Yeast expressing the *CYP88D6* from glycyrrhizin-producing species showed higher yields of 11-oxo-β-amyrin compared with yeast expressing *GpCYP88D15* from non-glycyrrhizin-producing *Glycyrrhiza* species (Fig. 3D). By contrast, yeast expressing *GpCYP88D15* produced higher yields of the dehydrated compounds **3** and **4**. These results indicate that CYP88D15 from non-glycyrrhizin-producing *Glycyrrhiza* species preferentially catalyzes allylic hydroxylation at C-11 of β-amyrin and rarely catalyzes the second oxidation step.

Identification of protein regions containing essential residues involved in 11-oxo-β-amyrin formation

To identify which protein regions are involved in the two-step oxidation of β-amyrin, chimeric proteins were constructed by

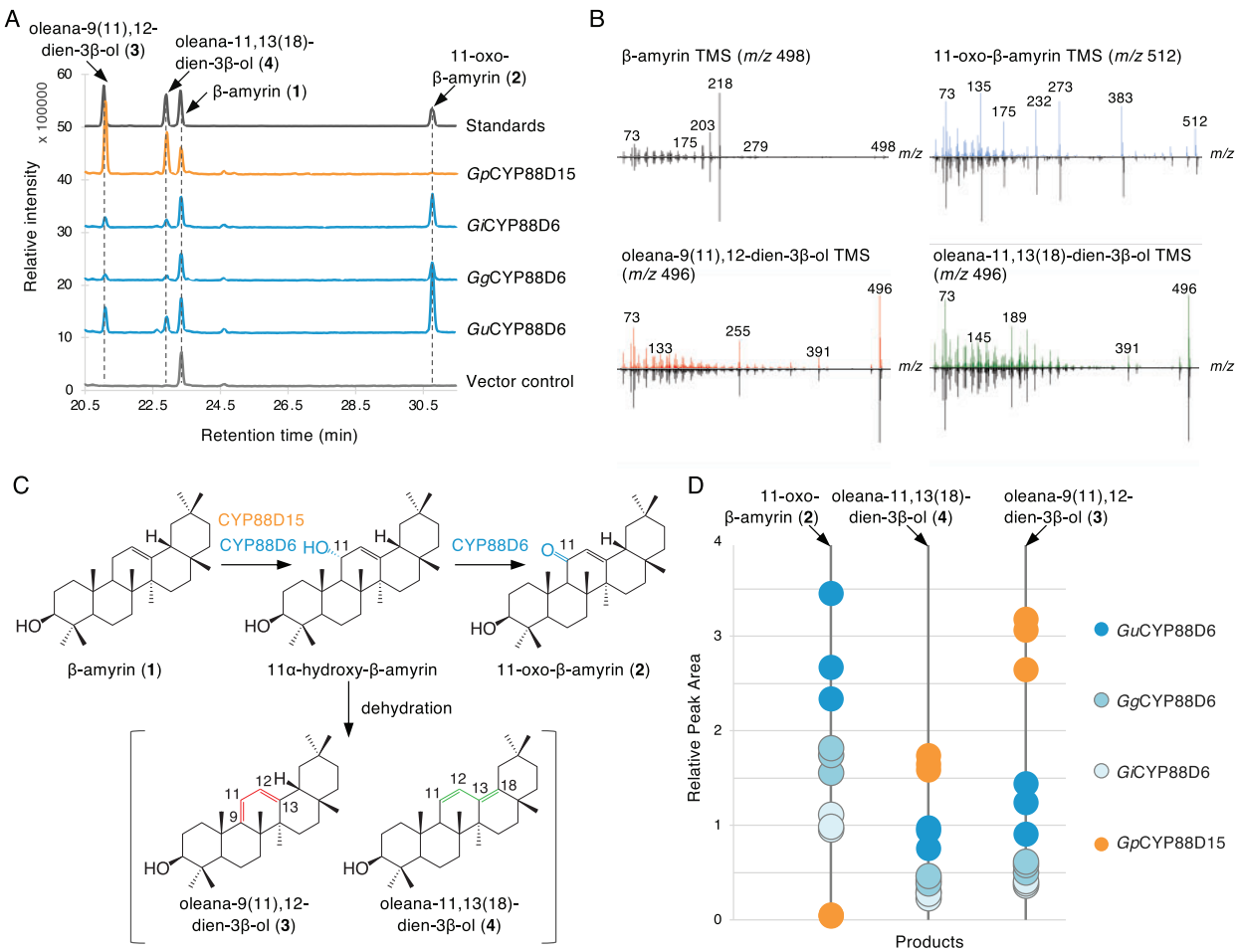

Fig. 3 Functional characterization of *CYP88D6* homologs from *Glycyrrhiza* species. (A) GC–MS analyses of extracts from engineered yeast expressing *CYP88D6* homologs. β-Amyrin-producing yeast that transformed with pELC-GW empty vector was used as the vector control. The total ion counts are shown. Blue spectra indicate glycyrrhizin-producing species, and orange spectra indicate non-glycyrrhizin-producing *Glycyrrhiza* species. (B) Mass spectra of compounds **1**–**4** detected by GC–MS. Mass spectra are shown as follows: top side for compounds detected in yeast extracts and bottom side for authentic standards. (C) Oxidation activity of CYP88D6 homologs toward β-amyrin. (D) Relative peak area of compounds **2**–**4**. Triplicate samples are shown as filled circles.

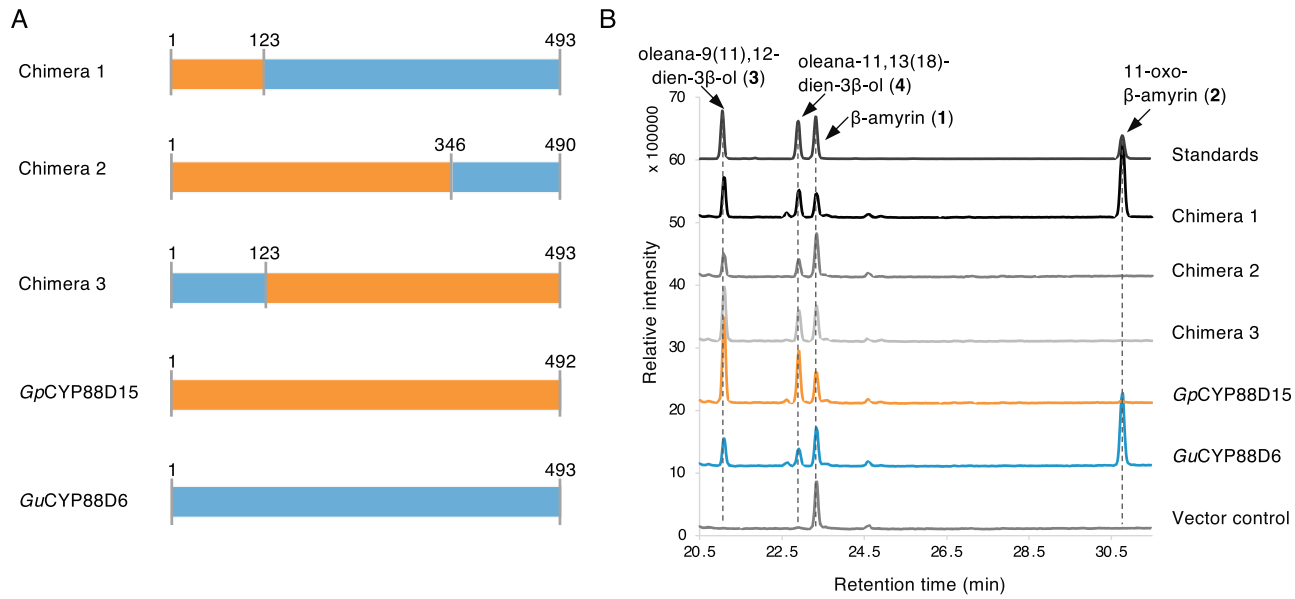

Fig. 4 Functional characterization of protein chimeras. (A) Map of the chimeras. The colors indicate the parent segment. (B) GC–MS analyses of extracts from engineered yeast expressing *chimeras*. β-Amyrin-producing yeast that transformed with pELC-GW empty vector was used as the vector control. The total ion counts are shown.

segment swapping between *Gu*CYP88D6 and *Gp*CYP88D15, and the oxidation activities of the chimeras were assessed (**Fig. 4A**). Each segment contained at least one predicted substrate recognition site (SRS): residues 1–123 contained SRS1, residues 123–346 contained SRS2, SRS3 and SRS4, and residues 346–492 contained SRS5 and SRS6. The gas chromatography–mass spectrometry (GC–MS) peak area for 11-oxo-β-amyrin from yeast expressing *Chimera 1* was comparable with the *GuCYP88D6*-expressing positive control; however, no 11-oxo-β-amyrin was detected in yeast strains expressing *Chimera 2* or *Chimera 3* (**Fig. 4B**). Chimera 1 comprised residues 1–123 from *Gp*CYP88D15 and residues 123–493 from *Gu*CYP88D6 (**Fig. 4A**), suggesting that the region between residues 123 and 493 is necessary for the second oxidation step catalyzed by *Gu*CYP88D6 during 11-oxo-β-amyrin formation.

M221V or F300L substitutions in *Gu*CYP88D significantly reduce 11-oxo-β-amyrin formation

To identify essential residues involved in the second oxidation of C-11 during 11-oxo-β-amyrin biosynthesis, we compared the protein sequences of CYP88D6 homologs from glycyrrhizin-producing species and non-glycyrrhizin-producing *Glycyrrhiza* species (**Supplementary Fig. S1**, **Fig. 5A**). Candidate residues for mutagenesis were selected based on the following criteria: (i) located between amino acids 123 and 493, (ii) located in a putative SRS, (iii) conserved in homologs of glycyrrhizin-producing species but not in homologs of non-glycyrrhizin-producing *Glycyrrhiza* species and (iv) oriented toward the heme pocket (**Fig. 5B**). We selected six residues, L217, M221, I296, F300, S473 and V476, and applied the site-directed mutagenesis to convert them to their corresponding residue in *Gp*CYP88D15 (**Fig. 5C**). Of the six yeast strains, each expressing a mutated version of *CYP88D6*, two mutations (M221V and F300L) resulted in dramatically decreased 11-oxo-β-amyrin formation and increased levels of dehydrated products (**Fig. 5D**). This indicated that residues M221 and F300 of *Gu*CYP88D6 are essential for the second oxidation step during 11-oxo-β-amyrin biosynthesis.

Substitution of essential *Gu*CYP88D residues into corresponding *Gp*CYP88D15 residues does not enhance 11-oxo-β-amyrin biosynthesis

The *Gu*CYP88D6 residues that we found to be essential for second oxidation step during 11-oxo-β-amyrin biosynthesis (M221 and F300) were substituted into their corresponding positions in *Gp*CYP88D15 (V221 and L297), to assess the ability of these residues to elicit 11-oxo-β-amyrin formation. The two mutations in *Gp*CYP88D15, *V221M* and *L297F*, were expressed in engineered yeast (**Fig. 6A, B**). Extracts from yeast cultures expressing either *V221M* or *L297F GpCYP88D15* displayed no significant difference in 11-oxo-β-amyrin production compared with the unmutated *GpCYP88D15*-expressing control. Similarly, a double substitution *GpCYP88D15* mutant harboring both *V221M* and *L297F* mutations (*V221M/L297F*) presented similar spectra to the *GpCYP88D15* control, showing no significant difference in 11-oxo-β-amyrin formation (**Fig. 6A, B**). This suggested that further residue modifications to *GpCYP88D15* are required to enhance 11-oxo-β-amyrin formation.

Discussion

Several *Glycyrrhiza* species, such as *G. uralensis*, *G. glabra* and *G. inflata*, produce glycyrrhizin as their main saponin, which is ketolated at C-11. By contrast, other *Glycyrrhiza* species produce mainly oleanane-type saponins that have diene structures and lack the C-11 ketone group. The present work suggests that glycyrrhizin production in *Glycyrrhiza* plants is likely

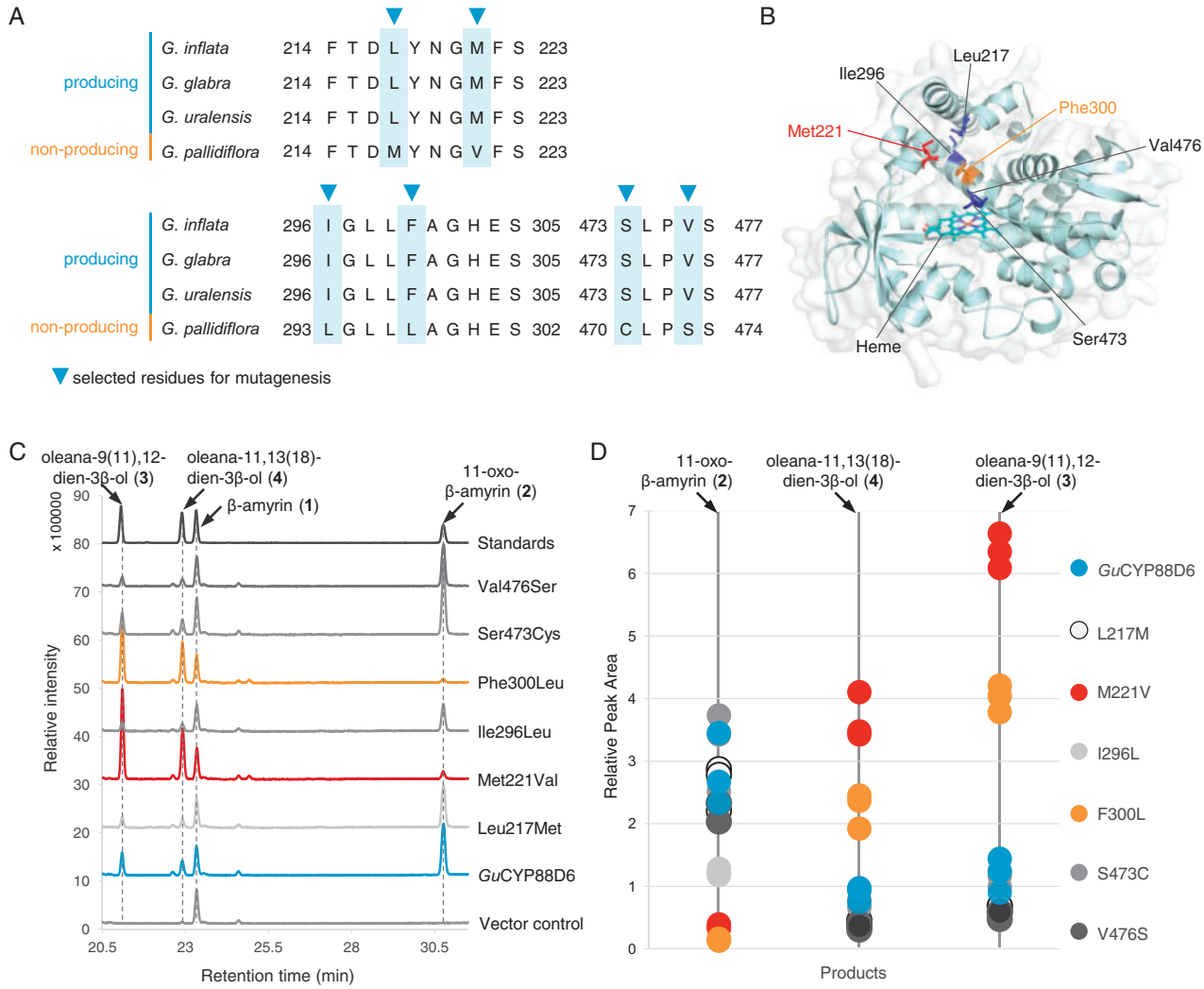

Fig. 5 The effect of mutations on 11-oxo-β-amyrin production. (A) Comparison of putative substrate recognition sites. (B) Positions of selected residues in the *Gu*CYP88D6 model. The residues selected for mutagenesis are indicated. (C) GC–MS analyses of extracts from engineered yeast expressing *GuCYP88D6* mutants. β-Amyrin-producing yeast that transformed with pELC-GW empty vector was used as the vector control. The total ion counts are shown. (D) Relative peak areas of compounds **2**–**4**. Triplicate samples are shown as filled circles.

determined by the activity of CYP88D6 homologs and their ability to perform the second oxidation step at the C-11 position. Our results also indicate that differences in CYP oxidation ability could be a source of saponin chemodiversity in *Glycyrrhiza* plants.

Recombinant yeast expressing *CYP88D15*, a homolog of CYP88D6 from *G. pallidiflora*, accumulated predominantly oleana-9(11),12-dien-3β-ol (**3**) and oleana-11,13(18)-dien-3β-ol (**4**) as reaction products although the recombinant yeast cultured for the short period as the previous report (Seki et al. 2008) accumulated 11α-hydroxy-β-amyrin. 11α-Hydroxy-β-amyrin is a labile compound and is easily converted into oleana-9(11),12-dien-3β-ol (**3**) and oleana-11,13(18)-dien-3β-ol (**4**) by non-enzymatic dehydration; this is because the hydroxylated position, C-11, is in an allylic position (**Fig. 7**). There is also a possibility that the dehydration is catalyzed by a yeast endogenous dehydrase. Thus, the production of oleana-9(11),12-dien-3β-ol (**3**) and oleana-11,13(18)-dien-3β-ol (**4**) by *CYP88D15*-expressing yeast suggests that CYP88D15 cannot catalyze the second oxidation reaction necessary to produce 11-oxo-β-amyrin, a glycyrrhizin biosynthetic intermediate. The CYP88D15-catalyzed allylic hydroxylation of β-amyrin and its subsequent dehydration should be a source the biosynthetic pathways of major saponins with homoannular or heteroannular diene structures in non-glycyrrhizin-producing *Glycyrrhiza* species. A recent study revealed that allylic hydroxylation is a key enzymatic transformation during the production of bitter compounds in bitter melon (Takase et al. 2019).

The phylogenetic analysis of CYP88D homologs from *Glycyrrhiza* species performed in this study showed that the homologs were divided into two groups corresponding to the *Glycyrrhiza* plant categories: glycyrrhizin-producing species and non-glycyrrhizin-producing species. This suggests that the differences in the catalytic properties of the CYP88D6 homologs affect the productivity of glycyrrhizin and other saponins in *Glycyrrhiza* species. The mutation analyses of the CYP88D6 homologs indicated that M221 and F300 are essential residues for glycyrrhizin production but that they are not the only two

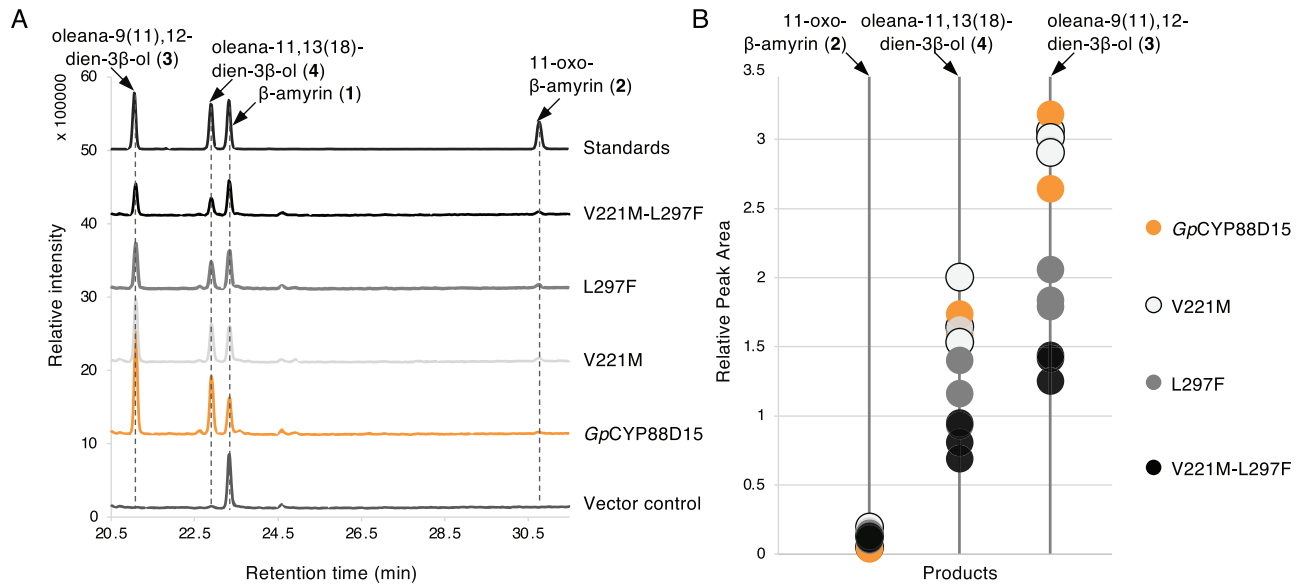

Fig. 6 Functional characterization of *GpCYP88D15* mutants. (A) GC–MS analyses of extracts from engineered yeast strains expressing *GpCYP88D15* mutants. β-Amyrin-producing yeast that transformed with pELC-GW empty vector was used as the vector control. (B) Relative peak areas of the compounds. Each triplicate sample is displayed.

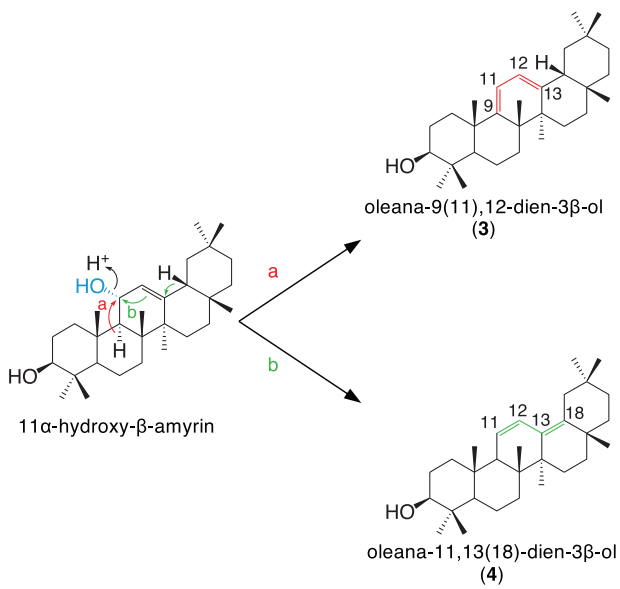

Fig. 7 The proposed mechanism for allyl alcohol dehydration. Red (A) indicates the route of homoannular diene formation. Green (B) indicates the route of heteroannular diene formation.

residues necessary for second oxidation step of 11-oxo-β-amyrin biosynthesis. To date, a limited number of CYP88D subfamily members have been functionally characterized; among them, only CYP88D6 homologs from *Glycyrrhiza* species have displayed oxidation activity toward β-amyrin. Future studies should aim to elucidate the sequence–function relationship within the CYP88D subfamily by functionally comparing CYP88D enzymes from various leguminous plants.

Triterpenoid saponins with homoannular and heteroannular dienes can also be found in non-leguminous plants, such as Myrtaceae (Begum et al. 2002), Ranunculaceae (Zhao et al. 2012), Celastraceae (Cáceres-Castillo et al. 2008), Araliaceae (Mutsuga et al. 1997), Apiaceae (Shimaoka et al. 1975), Euphorbiaceae (Tanaka and Matsunaga 1988), Caryophyllaceae (Koike et al. 1999) and Malvaceae (Mahato et al. 1988) (**Supplementary Fig. S3**). Non-leguminous plants may produce oleanane-type saponins with homoannular or heteroannular diene structures via the same mechanism as leguminous plants, i.e. by the introduction of a hydroxyl group at C-11, followed by its dehydration. However, the CYP88D subfamily is limited to leguminous plants, which suggests the existence of other mechanisms that generate saponins with diene structures in non-leguminous plants. However, a synthetic biological approach using CYP88D15 could be applied to produce pharmacologically active saponins with diene structures whose biosynthetic enzymes have not been identified yet.

Materials and Methods

Triterpene standards

β-Amyrin was purchased from Extrasynthese (Lyon, France). 11-Oxo-β-amyrin was synthesized as described in our previous report (Seki et al. 2008). Oleana-9(11),12-dien-3β-ol (**3**) and oleana-11,13(18)-dien-3β-ol (**4**) were synthesized from oleanolic acid (**Supplementary Methods**).

Plant materials

Glycyrrhiza inflata was obtained from the Health Sciences University of Hokkaido (Hokkaido, Japan).

Isolation of *CYP88D6* homologs

The preparations of cDNA templates for PCR from underground parts of *G. glabra*, *G. pallidiflora* and *G. macedonica* were described in our previous report

(Fanani et al. 2019). The cDNA template for PCR from underground part of *G. inflata* was also prepared as that of *G. macedonica*. The PCR primers used are listed in **Supplementary Table S1**. The primer set used to amplify *GgCYP88D6* and *GpCYP88D15* cDNA fragments was designed based on the *GuCYP88D6* gene. The full-length coding sequences (CDSs) of *GgCYP88D6* and *GpCYP88D15* were amplified after performing the rapid amplification of cDNA ends PCR, as described in our previous report (Seki et al. 2008). The full-length CDSs of *GmCYP88D15* and *GiCYP88D6* were amplified using primer sets based on the full-length CDSs of *GpCYP88D15* and *GuCYP88D6*, respectively (**Supplementary Table S1**). The full-length CDSs of *CYP88D6* homologs were cloned into the pENTR/D-TOPO vector using either the pENTR™/D-TOPO™ Cloning Kit (Thermo Fisher Scientific, Waltham, MA, USA) or NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs, Beverly, MA, USA).

Protein sequence comparisons

To compare the protein sequences of CYP88D6 homologs, multiple sequence alignment was performed using the Align feature in Uniprot (**www.uniprot.org/align/**) (Pundir et al. 2016). The percentage identities between each CYP88D6 homolog are summarized in **Supplementary Table S2**. The putative substrate recognition sites were assigned as described previously (Seki et al. 2008).

Phylogenetic analysis

The protein sequences of CYP88D subfamily members (**Supplementary Table S3**) were used for phylogenetic analysis. The protein sequence of CYP93E3 from *G. uralensis* was included as an outgroup. A neighbor-joining tree (Saito and Nei 1987) was estimated by the p-distance model and a bootstrap analysis with 1,000 replicates (Felsenstein 1985) using MEGA7 (Kumar et al. 2016).

In vivo enzyme activity assay

To characterize the oxidation activities of CYP88D6 homologs using β-amyrin as the substrate, in vivo enzyme activity assays were performed using the engineered yeast strain *Saccharomyces cerevisiae* INVSc1 (*MATa his3D1 leu2 trp1-289 ura3-52*; Thermo Fisher Scientific) transformed with β-amyrin synthase from *Lotus japonicus* (pYES3-ADH-OSC1) (Seki et al. 2008). To generate yeast expression vectors for the dual expression of each *CYP88D6* homolog and cytochrome P450 reductase (*L. japonicus CPR1*), each full-length CDS of the CYP88D6 homolog was cloned into pELC-GW (Seki et al. 2011) using Gateway™ LR Clonase™ II Enzyme Mix (Thermo Fisher Scientific). Four individual pELC-CYP88D6 homologs, and pELC-GW as a negative control, were transformed into the engineered yeast (Seki et al. 2008) using the Frozen-EZ Yeast Transformation II™ kit (Zymo Research, Orange, CA, USA).

All transformant yeast strains were cultured in synthetic defined (SD) medium containing a yeast nitrogen base without amino acids or ammonium sulfate (Becton, Dickinson and Company, Sparks, MD), 5 g/l ammonium sulfate (Nacalai Tesque, Kyoto, Japan) and −Leu/−Trp Drop-out supplement (Takara Bio USA, Mountain View, CA, USA) and were incubated at 30°C with shaking at 200 rpm. Culture seeds were prepared by inoculating a fresh single colony into 2 ml of fresh SD medium supplemented with 2% glucose (Nacalai Tesque) and incubated for 18 h. Then, 100 μl of culture seeds were transferred into 5 ml of fresh SD medium supplemented with 2% glucose and incubated for 18 h. To perform galactose induction, yeast cells were collected by centrifugation and resuspended into 5 ml of fresh SD medium supplemented with 2% galactose (Nacalai Tesque). Upon galactose induction, the optical density of yeast cultures at 600 nm was adjusted to approximately 2. The yeast cultures were then incubated for 4 d. Sample preparation for thin-layer chromatography (TLC) analysis was performed as described in **Supplementary Methods**. All assays were performed in triplicate by culturing three independent colonies to confirm our results.

Triterpene metabolite analyses

Analyses of reaction products were performed as described in our previous report (Fanani et al. 2019), with the addition of an internal standard. Prior to yeast extraction, 50 μl of 100 mg/l ursolic acid was spiked into the yeast culture as an internal standard. GC–MS was performed for qualitative analyses as described in our previous report (Fanani et al. 2019). For quantitative analyses, GC–MS analysis was performed as follows: the initial oven temperature was 120°C, increasing from 120 to 300°C at 20°C/min, and then held for 20 min. Samples were injected in splitless mode with an injection temperature of 250°C, using helium as the carrier gas, and a flow rate of 1.0 ml/min. The Rts and mass fragmentation patterns of detectable compounds were compared with those of authentic standards to assign the peaks. The peak areas of the total ion current chromatograms were calculated using manual integration of Qualitative Analysis (B.070.00; Agilent Technologies, Santa Clara, CA, USA). The relative peak areas of compounds were calculated by dividing the peak areas of the individual compounds by the peak area of the internal standard. TLC analysis of yeast extract was performed as described in **Supplementary Methods**.

Chimeric protein construction

To identify which CYP88D6 homolog protein regions contain residues essential for the second oxidation activity, three protein segments were generated: segment 1–123 for SRS1, segment 123–346 for SRS2, SRS3 and SRS4 and segment 346–492 for SRS5 and SRS6 (**Supplementary Fig. S1**). To generate chimeric proteins, the CDS segments of *GuCYP88D6* and *GpCYP88D15* were amplified by PCR using the primer pairs listed in **Supplementary Table S1**. The segments were assembled and cloned into pENTR/D-TOPO using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). Evaluation of the oxidation activities of the chimeric proteins toward β-amyrin was performed as described above.

Homology modeling analyses

Homology modeling of *Gu*CYP88D6 was performed using CYP90B1 of *Arabidopsis thaliana* (Protein Data Bank ID 6A17) (Fujiyama et al. 2019) as a template in SWISS-MODEL (Arnold et al. 2006). Despite the low sequence identity, structural assessment of the *Gu*CYP88D6 model using a Ramachandran plot indicated that it was 94.44% favorable. The *Gu*CYP88D6 model was analyzed and visualized using PyMol version 2.3.1 (Schrödinger LLC 2017).

Site-directed mutagenesis

Site-directed mutations in *GuCYP88D6* and *GpCYP88D15* were introduced using the PrimeSTAR Mutagenesis Kit (Takara Bio). The primer pairs used are listed in **Supplementary Table S1**.

Supplementary Data

Supplementary data are available at PCP online.

Funding

The Grants-in-Aid for Scientific Research of the Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant Number JP19H02921); the Scientific Technique Research Promotion Program for Agriculture, Forestry, Fisheries, and Food Industry, Japan; The Program for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry (BRAIN); The Special Fund from the Director of RIKEN Yokohama Institute; the RIKEN Rijicho Fund; and the Monbukagakusho Scholarship.

Acknowledgments

We thank Toshio Aoki and Tomoyoshi Akashi (Nihon University, Japan) for valuable discussions and technical advice, Hideaki Shimizu and Shigeyuki Yokoyama (RIKEN, Japan) for valuable discussions, Mareshige Kojoma (Health Sciences University of Hokkaido, Japan) for providing *G. glabra* and *G. inflata*, Makio Shibano (Osaka University of Pharmaceutical

Sciences, Japan) for providing *G. macedonica*, Yoshinori Fujimoto and Takao Kirita (Tokyo Institute of Technology, Japan) for kindly give the authentic standards, David R. Nelson (University of Tennessee, USA) for the naming of P450s, Hiromi Hashinokuchi (RIKEN, Japan) for technical assistance with GC–MS analysis, and Kyoko Inoue (Osaka University, Japan) for technical assistance with NMR analysis.

Disclosures

The authors Hiroshi Sudo and Satoru Sawai were employed by company Tokiwa Phytochemical Co., Ltd. All other authors declare no competing interests.

References

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

Begum, S., Sultana, I., Siddiqui, B.S., Shaheen, F. and Gilani, A.H. (2002) Structure and spasmolytic activity of eucalyptanoic acid from *Eucalyptus camaldulensis* var. obtusa and synthesis of its active derivative from oleanolic acid. *J. Nat. Prod.* 65: 1939–1941.

Butelli, E., Titta, L., Giorgio, M., Mock, H.P., Matros, A., Peterek, S., et al. (2008) Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.* 26: 1301–1308.

Cáceres-Castillo, D., Mena-Rejón, G.J., Cedillo-Rivera, R. and Quijano, L. (2008) 21β-Hydroxy-oleanane-type triterpenes from *Hippocratea excelsa*. *Phytochemistry* 69: 1057–1064.

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

Chung, S.Y., Seki, H., Fujisawa, Y., Shimoda, Y., Hiraga, S., Nomura, Y., et al. (2020) A cellulose synthase-derived enzyme catalyses 3-*O*-glucuronosylation in saponin biosynthesis. *Nat. Commun*. 11. 5664.

Facchini, P.J., Bohlmann, J., Covello, P.S., De Luca, V., Mahadevan, R., Page, J.E., et al. (2012) Synthetic biosystems for the production of high-value plant metabolites. *Trends Biotechnol.* 30: 127–131.

Fanani, M.Z., Fukushima, E.O., Sawai, S., Tang, J., Ishimori, M., Sudo, H., et al. (2019) Molecular basis of C-30 product regioselectivity of legume oxidases involved in high-value triterpenoid biosynthesis. *Front. Plant Sci*. 10: 1520.

Felsenstein, J. (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783–791.

Fujiyama, K., Hino, T., Kanadani, M., Watanabe, B., Lee, H.J., Mizutani, M., et al. (2019) Structural insights into a key step of brassinosteroid biosynthesis and its inhibition. *Nat. Plants* 5: 589–594.

Ghosh, S. (2017) Triterpene structural diversification by plant cytochrome P450 enzymes. *Front. Plant Sci.* 8: 1886.

Gibson, M.R. (1978) *Glycyrrhiza* in old and new perspectives. *Lloydia* 41: 348–354.

Hayashi, H., Hosono, N., Kondo, M., Hiraoka, N., Ikeshiro, Y., Shibano, M., et al. (2000) Phylogenetic relationship of six *Glycyrrhiza* species based on rbcL sequences and chemical constituents. *Biol. Pharm. Bull.* 23: 602–606.

Jozwiak, A., Sonawane, P.D., Panda, S., Garagounis, C., Papadopoulou, K.K., Abebie, B., et al. (2020) Plant terpenoid metabolism co-opts a component of the cell wall biosynthesis machinery. *Nat. Chem. Biol.* 16: 740–748.

Kao, T.C., Wu, C.H. and Yen, G.C. (2014) Bioactivity and potential health benefits of licorice. *J. Agric. Food Chem.* 62: 542–553.

Koike, K., Jia, Z. and Nikaido, T. (1999) New triterpenoid saponins and sapogenins from *Saponaria officinalis*. *J. Nat. Prod.* 62: 1655–1659.

Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33: 1870–1874.

Mahato, S.B., Pal, B.C. and Sarkar, S.K. (1988) New triterpenoid saponins from *Corchorus acutangulus*. *Phytochemistry* 27: 1433–1437.

Mizutani, K., Kuramoto, T., Tamura, Y., Ohtake, N., Doi, S., Nakaura, M., et al. (1994) Sweetness of glycyrrhetic acid 3-*O*-β-D-monoglucuronide and the related glycosides. *Biosci. Biotechnol. Biochem.* 58: 554–555.

Moghe, G.D. and Kruse, L.H. (2018) The study of plant specialized metabolism: challenges and prospects in the genomics era. *Am. J. Bot.* 105: 959–962.

Mutsuga, M., Kojima, K., Saracoglu, I. and Ogihara, Y. (1997) Minor saponins from *Tetrapanax papyriferum*. *Chem. Pharm. Bull.* 45: 552–554.

Nelson, D.R. (2009) The cytochrome p450 homepage. *Hum. Genomics* 4: 59–65.

Nomura, Y., Seki, H., Suzuki, T., Ohyama, K., Mizutani, M., Kaku, T., et al. (2019) Functional specialization of UDP-glycosyltransferase 73P12 in licorice to produce a sweet triterpenoid saponin, glycyrrhizin. *Plant J.* 99: 1127–1143.

Ohtani, K., Kasai, R., Yang, C.R., Yamasaki, K., Zhou, J. and Tanaka, O. (1994) Oleanane glycosides from roots of *Glycyrrhiza yunnanensis*. *Phytochemistry* 36: 139–145.

Polturak, G., Breitel, D., Grossman, N., Sarrion-Perdigones, A., Weithorn, E., Pliner, M., et al. (2016) Elucidation of the first committed step in betalain biosynthesis enables the heterologous engineering of betalain pigments in plants. *New Phytol.* 210: 269–283.

Pundir, S., Martin, M.J. and O'Donovan, C.; UniProt Consortium. (2016) UniProt tools. *Curr. Protoc. Bioinformatics* 53: 1.29.1–1.29.15.

Rates, S.M. (2001) Plants as source of drugs. *Toxicon* 39: 603–613.

Saito, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. 4: 406–425.

Schrödinger LLC (2017) The PyMOL Molecular Graphics System, Version 2.0.

Schwab, W., Davidovich-Rikanati, R. and Lewinsohn, E. (2008) Biosynthesis of plant-derived flavor compounds. *Plant J.* 54: 712–732.

Seki, H., Ohyama, K., Sawai, S., Mizutani, M., Ohnishi, T., Sudo, H., et al. (2008) Licorice β-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. *Proc. Natl. Acad. Sci. USA* 105: 14204–14209.

Seki, H., Sawai, S., Ohyama, K., Mizutani, M., Ohnishi, T., Sudo, H., et al. (2011) Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. *Plant Cell* 23: 4112–4123.

Shibata, S. (2000) A drug over the millennia: pharmacognosy, chemistry, and pharmacology of licorice. *Yakugaku Zasshi* 120: 849–862.

Shimaoka, A., Seo, S. and Minato, H. (1975) Saponins isolated from *Bupleurum falcatum* L.; components of saikosaponin b. *J. Chem. Soc, Perkin Trans*. 1: 2043–2048.

Takase, S., Kera, K., Nagashima, Y., Mannen, K., Hosouchi, T., Shinpo, S., et al. (2019) Allylic hydroxylation of triterpenoids by a plant cytochrome P450 triggers key chemical transformations that produce a variety of bitter compounds. *J. Biol. Chem.* 294: 18662–18673.

Tanaka, R. and Matsunaga, S. (1988) Triterpene dienols and other constituents from the bark of *Phyllanthus flexuosus*. *Phytochemistry* 27: 2273–2277.

The Plant List (2013). Version 1.1. Published on the Internet. **http://www.theplantlist.org/** (2 June 2020).

Wang, L., Yang, R., Yuan, B., Liu, Y. and Liu, C. (2015) The antiviral and antimicrobial activities of licorice, a widely-used Chinese herb. *Acta Pharm. Sin. B* 5: 310–315.

Wani, M.C., Taylor, H.L., Wall, M.E., Coggon, P. and McPhail, A.T. (1971) Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chem. Soc.* 93: 2325–2327.

Yang, R., Wang, L.Q., Yuan, B.C. and Liu, Y. (2015) The pharmacological activities of licorice. *Planta Med.* 81: 1654–1669.

Zhao, C., Shao, J. and Fan, J. (2012) A new triterpenoid with antimicrobial activity from *Anemone rivularis*. *Chem. Nat. Compd.* 48: 803–805.