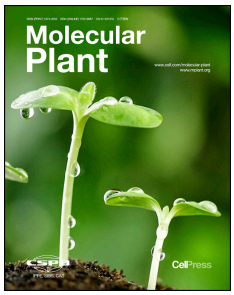
Accepted Manuscript



Characterizations of Panax ginseng UDP-glycosyltransferases catalyzing protopanaxatriol and biosyntheses of bioactive ginsenosides F1 and Rh1 in metabolically engineered yeasts

Wei Wei, Pingping Wang, Yongjun Wei, Qunfang Liu, Chengshuai Yang, Guoping Zhao, Jianmin Yue, Xing Yan, Zhihua Zhou

PII: S1674-2052(15)00261-0

DOI: 10.1016/j.molp.2015.05.010

Reference: MOLP 141

To appear in: *MOLECULAR PLANT*

Received Date: 28 April 2015

Revised Date: 22 May 2015

Accepted Date: 24 May 2015

Please cite this article as: **Wei W., Wang P., Wei Y., Liu Q., Yang C., Zhao G., Yue J., Yan X., and Zhou Z.** (2015). Characterizations of Panax ginseng UDP-glycosyltransferases catalyzing protopanaxatriol and biosyntheses of bioactive ginsenosides F1 and Rh1 in metabolically engineered yeasts. Mol. Plant. doi: 10.1016/j.molp.2015.05.010.

Characterizations of *Panax ginseng* UDP-glycosyltransferases catalyzing protopanaxatriol and biosyntheses of bioactive ginsenosides F1 and Rh1 in metabolically engineered yeasts

Wei Wei[‡1], Pingping Wang[‡1], Yongjun Wei[1], Qunfang Liu[2], Chengshuai Yang[1], Guoping Zhao[1], Jianmin Yue[*2], Xing Yan*[1], Zhihua Zhou*[1]

1. CAS-Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China
2. State Key Laboratory of Drug Research, Institute of Materia Medica, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, 201203, China

[‡]These authors contributed equally to this work

*Corresponding authors:
Zhihua Zhou
Email: zhouzhihua@sippe.ac.cn
Tel: 86-21-54924050, Fax: 86-21-54924049.
Xing Yan
Email: yanxing@sibs.ac.cn
Tel: 86-21-54924049, Fax: 86-21-54924049.
Jianmin Yue
E-mail: jmyue@mail.shcnc.ac.cn
Telephone: 86-21-50806718. Fax: 86-21-50806718.

Running title: UGTs glucosylating protopanaxatriol

Short summary

UDP-glycosyltransferase (UGT) encoding genes catalyzing the glycosylation of protopanaxatriol were cloned and characterized. Some key amino acid residues of the UGTs responsible for the activities and substrate regio-specificities were identified. The construction of yeast cell factories producing the bioactive ginsenosides F1 and Rh1 paved the way to manufacture these two high-value natural compounds through microbial fermentation in the future.

Abstract

Ginsenosides, the main pharmacological active natural compounds in ginseng (*Panax ginseng*), are mostly the glycosylated products of protopanaxadiol (PPD) and protopanaxatriol (PPT). No uridine diphosphate glycosyltransferase (UGT) that catalyzes PPT to produce PPT-type ginsenosides has yet been reported. Here, we show that UGTPg1, which was previously demonstrated to regio-specifically glycosylate the C20-OH of PPD, also specifically glycosylates the C20-OH of PPT to produce bioactive ginsenoside F1. We report the characterizations of four novel UGT genes isolated from *P. ginseng*, sharing high deduced amino acid identities (>84 %) with UGTPg1. We demonstrate that UGTPg100 specifically glycosylates the C6-OH of PPT to produce bioactive ginsenoside Rh1, and UGTPg101 catalyzes PPT to produce F1, followed by the generation of ginsenoside Rg1 from F1. However, UGTPg102 and UGTPg103 are found to have no detectable activity on PPT. Through structural modeling and site-directed mutagenesis, we identified several key amino acids of these UGTs which may play important roles in determining their activities and substrate regio-specificities. Moreover, we constructed yeast recombinants to biosynthesize F1 and Rh1, respectively, by introducing the genetically engineered PPT-producing pathway and UGTPg1 or UGTPg100. Our study reveals the possible biosynthetic pathways of PPT-type ginsenosides in *Panax* plants, and provides a sound manufacturing approach for bioactive PPT-type ginsenosides in yeast *via* synthetic biology strategies.

Key words: UDP-glycosyltransferase; triterpenoids; protopanaxatriol; ginsenoside F1; ginsenoside Rh1; *Panax ginseng*

Introduction

Ginsenosides, a group of triterpene saponins, are the main pharmacologically active constituents of ginseng (*Panax ginseng*) roots (Shibata, 2001). Ginseng has been used in Asia for approximately 5000 years, as traditional medicine and dietary supplements (Yun, 2001). More than 150 naturally occurring ginsenosides have been identified from *Panax* spp. (Christensen, 2009). Most ginsenosides are glycosylated products of the tetracyclic triterpenoid aglyones, protopanaxadiol (PPD) and protopanaxatriol (PPT) (Supplementary Figure 1). Glycosylation of natural products in plants is catalyzed by glycosyltransferases (GTs), leading to alteration of the hydrophilicity, stability, and subcellular localization, and thus the chemical properties and bioactivity, of the natural products (Bowles et al., 2006).

GTs are members of a multigene superfamily that catalyzes the transfer of a sugar moiety to a specific acceptor. The GTs that use uridine-diphosphate (UDP)-activated sugar molecules as donors are referred to as UDP-glycosyltransferases (UGTs). The diversity of plant UGTs has been demonstrated by comparing genomic and complementary DNA (cDNA) sequences from different plant species. More than 100 UGT-encoding genes are thought to exist in each plant genome, according to plant genomic or cDNA data. For example, the Carbohydrate Active enZyme database (CAZy) contains 122 UGT genes from *Arabidopsis thaliana*. However, only a limited number of UGTs characterized to glycosylate triterpenoid aglycones were uncovered from plants, such as *Medicago truncatula* (Achnine et al., 2005), *Saponaria vaccaria* (Meesapyodsuk et al., 2007), *Barbarea vulgaris* (Augustin et al., 2012), *Glycine max*

(Shibuya et al., 2010) and *P. ginseng* (Jung et al., 2014; Wang et al., 2015; Yan et al., 2014).

More than 100 unique UGT-encoding cDNA sequences were identified in each of the transcriptomes of *P. ginseng*, *P. notoginseng* and *P. quinguefolium* (Chen et al., 2011; Li et al., 2013; Luo et al., 2011; Sun et al., 2010), which may contribute to the diversity of ginsenosides in *Panax* plants. Although only EST data are available for *P. ginseng*, 42 full-length UGTs from *P*. *ginseng* were assembled in our previous study (Yan et al., 2014). UGTPg1 from *P. ginseng* was characterized to regio-specifically glycosylate the C20-OH of PPD and its derived ginsenosides (Yan et al., 2014). More recently, two UGTs from *P*. *ginseng* were characterized to catalyze the glycosylation of C3-OH of PPD to yield Rh2 and to elongate a glucose moiety of Rh2 to generate Rg3, respectively (Jung et al., 2014; Wang et al., 2015). However, until now no UGT which could glycosylate PPT to produce PPT-derived ginsenosides has been reported.

Some characterized plant UGTs are promiscuous and can act on substrates belonging to various classes. For example, UGT71G1 from *M. truncatula* can glycosylate flavonoid (quercetin) and triterpenoid (medicagenic acid, hederagenin) (Shao et al., 2005). UGT73C5 of *A. thaliana* can catalyze 13-*O*-glucosylation of the brassinosteriods brassinolide and castasterone, but it also transfers a glucose moiety to the C3-OH of deoxynivalenol (Poppenberger et al., 2003; Poppenberger et al., 2005). Some UGTs, however, are regio-specific or substrate specific. The three characterized *Stevia rebaudiana* UGTs (UGT74G1, UGT76G1 and UGT85C2) catalyze the glycosylation of steviol in a regio-selective manner (Richman et al., 2005). UGT73A14,

a UGT from *Gentiana triflora*, glycosylates C3'-OH of delphinidin 3,5-diglucoside, but it has no activity on other anthocyanins such as pelargonidin 3,5-diglucosides or cyanidin 3,5-diglucosides (Fukuchi-Mizutani et al., 2003).

Results from some studies suggest that a change in key amino acid residues of an UGT might alter its sugar donor specificity or substrate regio-specificity. Mutation of the last amino acid in the PSPG box (from H to Q) of a UDP-galactosyltransferase, ACGaT, confers glucosyltransferase activity, in addition to its inherent galactosyltransferase activity (Kubo et al., 2004). The F148V and Y02A mutants of UGT71G1 drastically changed the regio-selectivity for quercetin glycosylation from predominantly the C3'-OH of the B-ring to the C3-OH of the C ring (He et al., 2006).

In this study, four novel UGT-encoding genes (*UGTPg100*, *UGTPg101*, *UGTPg102* and *UGTPg103*) were isolated from *P. ginseng*. Catalytic activities of the four novel UGTs and our previously identified UGTPg1 (Yan et al., 2014) on PPT were evaluated and compared. UGTPg1 could only glycosylate the C20-OH of PPT. UGTPg100 could catalyze the glycosylation of C6-OH of PPT and its derived ginsenosides in a regio-specific manner. UGTPg101 could transform PPT to F1 and Rg1 sequentially. However, UGTPg102 and UGTPg103 possessed no detected activity. Through protein structure modeling, site-directed mutagenesis and enzyme assays, the key amino acid residues affecting the catalytic function of these UGTs were identified. The genetically engineered yeasts to produce PPT-type ginsenosides F1 and Rh1 were constructed and *in vivo* functions of UGTPg1 and UGTPg100 were further confirmed in these genetically engineered yeasts.

Results

Cloning of novel UDP-glycosyltransferase genes involved in the biosynthesis of ginsenosides

By functionally screening the ORFs from our constructed database, we were able to report the identification of the first tetracyclic triterpenoid UGT (UGTPg1), which was characterized to glycosylate the C20-OH of PPD and its derived ginsenosides (Yan et al., 2014). We re-assembled the cDNA contigs in the database and focused on the candidates belonging to the UGT71 subfamily, of which UGTPg1 is a member. We obtained another UGT-encoding gene, *UGTPg100*, which showed high deduced amino acid identity (84.1%) to UGTPg1 (Supplemental Table 1). An additional three UGT ORFs were amplified from mRNAs of *P. ginseng* by RT-PCR, using primer pairs against UGTPg1- and UGTPg100-encoding genes and identified as *UGTPg101*, *UGTPg102* and *UGTPg103*. The deduced amino acid sequences of the five cloned UGT genes were aligned (Figure 1). The amino acid sequences of *UGTPg101* and *UGTPg102* were mostly identical to that of UGTPg1, with identities up to 98.1% and 95.4%, respectively (Supplemental Table 1). The deduced amino acid sequence of *UGTPg103* was highly identical to that of *UGTPg100*, with an amino acid identity up to 98.7% (Supplemental Table 1). Besides, phylogenetic analysis of the five UGTPgs and other characterized plant UGTs belonging to UGT71 subfamily demonstrated that the five UGTPgs were divided into two branches and phylogenetically far from UGT71G1 (Shao et al., 2005), which was characterized to glycosylate triterpenoid substrates (Supplemental Figure 2).

Enzymatic characterizations of the five UGTs towards PPT and PPT-type ginsenosides

ORFs of the four UGT (UGTPg100, UGTPg101, UGTPg102 and UGTPg103) encoding genes were cloned into pET28a and heterologously expressed in *Escherichia coli* BL21 (DE3) individually. Western blotting was employed to verify successful expression of the recombinant C-terminally 6×His-tagged proteins as soluble products (Figure 2A). The activities and substrate specificities of the four newly expressed UGTs as well as UGTPg1 towards PPT and PPT-type ginsenosides were examined. Crude cell extracts containing each of the recombinant UGTs were incubated with UDP-glucose as a sugar donor and PPT or different PPT-type ginsenosides (F1, Rh1, Rg1 and Rg2) as putative sugar acceptors (Supplemental Figure 3A). As UGTPg1 was demonstrated to transfer a glucosyl moiety to the C20-OH of dammarenediol-II, PPD and PPD-type ginsenosides (Rh2 and Rg3), the activities of the four newly expressed UGTs were also tested against PPD or PPD-type ginsenosides (CK, Rh2, F2 and Rg3; Supplemental Figure 3B).

The crude UGTPg100 preparation could readily convert PPT into Rh1, and F1 into Rg1, respectively, as monitored by thin layer chromatography (TLC; Supplemental Figure 3A), high-performance liquid chromatography (HPLC; Figure 2B and Supplemental Figure 4) and high-performance liquid chromatography/electrospray ionization mass spectroscopy (HPLC/ESIMS; Figure 2C) analyses. After purification, the structure of Rh1 was confirmed by ^{13}C nuclear magnetic resonance (NMR; Supplemental Table 2). The above results indicated that UGTPg100 transferred a glucosyl moiety to the free C6-OH of PPT or PPT-type ginsenosides *in vitro*. UGTPg100 had no detected activity towards PPD or any of the tested PPD-type ginsenosides (Supplemental Figure 3B). UGTPg100 also could not glycosylate Rh1, Rg1, Rg2, the PPT-type ginsenosides without a free hydroxyl at C6 (Supplemental Figure 3A and Supplemental Figure 4). Namely, UGTPg100 neither catalyzed the

glycosylation of the C3and/or C20 hydroxyls of ginsenosides (PPT/PPD/CK/Rh2/Rh1) nor transferred a glucosyl moiety to extend the sugar chain at C6 of Rh1, Rg1 or Rg2. These findings indicate that UGTPg100 catalyzes the C6-OH glycosylation of PPT and the derived ginsenoside F1 in a regio-selective manner.

UGTPg101 had catalytic activities towards PPD and PPD-type ginsenosides, similar to the activities of UGTPg1 (Yan et al., 2014). Both UGTPg101 and UGTPg1 catalyzed C20-OH glycosylation of PPD, Rh2 and Rg3 to produce the ginsenosides CK, F2 and Rd, respectively (Supplemental Figure 3B). When incubated with PPT, UGTPg101 glycosylated the C20-OH of PPT to produce F1, as monitored by TLC (Supplemental Figure 3A) and HPLC (Figure 2B). After incubating UGTPg101 with PPT for a longer period, an additional glycosylated product could be detected in the TLC profile (Supplemental Figure 3A), which was confirmed by HPLC to be Rg1 (Figure 2B). These results suggest that UGTPg101 is a bifunctional UGT. UGTPg101 could glycosylate the C6-OH of F1, but it could not glycosylate the C20-OH of Rh1 or Rg2 (Figure 2B; Supplemental Figure 3A). Rh1 was not detected when UGTPg101 was incubated with PPT (Figure 2B), indicating that it could not glycosylate the C6-OH of PPT. All the above data indicates that UGTPg101 has the capability to glycosylate sequentially the C20-OH of PPT to produce F1 and then the C6-OH of F1 to produce Rg1.

When incubated with PPT, UGTPg1 glycosylated the C20-OH of PPT to produce F1, as monitored by TLC (Supplemental Figure 3A), HPLC (Figure 2B) and HPLC/ESIMS (Figure 2C) analyses. After purification, the structure of F1 was confirmed by ^{13}C NMR (Supplemental Table 2). As was observed for UGTPg101, UGTPg1 could not glycosylate the C20-OH of Rh1or Rg2 and the C6-OH of PPT (Supplemental Figure 3A). Unlike UGTPg101, UGTPg1 was unable to glycosylate the C6-OH of F1, either

(Supplemental Figure 3A and Supplemental Figure 4). These results demonstrate that UGTPg1 can only glycosylate the C20-OH of PPT. By contrast, UGTPg102 and UGTPg103 had no activity towards any of the aglycones or ginsenosides evaluated when UDP-glucose was used as the sugar donor (Supplemental Figure 3 and Figure 2B).

The kinetic parameters of UGTPg1 and UGTPg100 towards the substrate PPT were assessed (Table 1). Their activities and catalytic efficiencies were similar. In addition, the proposed biosynthetic pathway of ginsenoside Rg1 in *Panax* plants based on the functions of UGTPg1, UGTPg100 and UGTPg101 was illustrated in Figure 2D.

The substitutions between H144 of UGTPg1 and Y144 of UGTPg102 shifts their functions

UGTPg1 and UGTPg102 shared a 95.4% amino acid sequence identity, with only 17 amino acid substitutions. UGTPg1 glycosylated the C20-OH of PPT (Figure 2B); however, UGTPg102 had no detectable activity towards the substrates evaluated (Supplemental Figure 3). SWISS-MODEL was employed to conduct structural modeling, using UGT71G1 from *M. truncatula* (PDB: 2ACW) as the template. When the locations of the amino acid substitutions were mapped to the protein 3D structure, six residues located around the substrate binding pocket and the presumed catalytic residues were noted (Figure 3A).

The amino acid residues L38, A40, F85, T134, H144 and Q388 of UGTPg1 corresponded to F38, S40, L85, I134, Y144 and H388 of UGTPg102, respectively. These amino acid residues of UGTPg1 were independently converted into the

corresponding ones of UGTPg102 by site-directed mutagenesis, yielding six UGTPg1 mutated proteins: UGTPg1-L38F, UGTPg1-A40S, UGTPg1-F85L, UGTPg1-T143I, UGTPg1-H144Y and UGTPg1-Q388H. All crude cell extracts of the recombinant *E. coli* expressing these proteins were subjected to Western blotting analysis to confirm expression (Figure 3B).

The functions of these mutated proteins were assayed. UGTPg1-L38F, UGTPg1-A40S, UGTPg1-F85L, UGTPg1-T134I and UGTPg1-Q388H retained the ability to transfer a glucose residue to the C20-OH of PPT in a manner similar to the wild-type UGTPg1 (Figure 3C). However, the mutant UGTPg1-H144Y could not catalyze the glycosylation of PPT (Figure 3C). Meanwhile, when Y144 of UGTPg102 was converted into H144, the mutated protein UGTPg102-Y144H gained the ability to transfer a glucose residue to the C20-OH of PPT. UGTPg102-Y144H still contained 16 amino acid residues not present in UGTPg1 (Figure 3C). These results suggest that H144 may be vital for the glycosylation activity of UGTPg1 towards the C20-OH of PPT.

A142, L186 and G338 are the key amino acid residues determining the function of UGTPg100

UGTPg100 and UGTPg103 were found to share a very high (98.7%) amino acid sequence identity, differing in only six amino acids. The six different residues lead to completely different activities, suggesting that some of them played critical roles in the function of UGTPg100.

Structural modeling of UGTPg100 was conducted using UGT71G1 from *M. truncatula* (PDB: 2ACW) as the template. We started by investigating four amino acid substitutions in detail: A142 and L186 located around the substrate binding pocket and the catalytic center and W205 and G338 located in the N- and C- terminal Rossmamn fold domains, respectively (Figure 4A). Amino acid residues A142, L186, W205 and G338 of UGTPg100 corresponded to T142, S186, R205 and R338 of UGTPg103, respectively. These four amino acids in UGTPg100 were independently converted to the corresponding ones in UGTPg103 by site-directed mutagenesis, yielding four mutated proteins: UGTPg100-A142T, UGTPg100-L186S, UGTPg100-W205R and UGTPg100-G338R, respectively. All crude cell extracts of recombinant *E. coli* expressing these proteins were subjected to Western blotting analysis to confirm expression (Figure 4B).

The activity of these mutated proteins towards PPT was investigated. UGTPg100-A142T, UGTPg100-L186S and UGTPg100-G338R showed no activity towards PPT, but UGTPg100-W205R retained the ability to transfer a glucose residue to the C6-OH of PPT, similarly to the wild-type UGTPg100 (Figure 4C). It is suggested that one of the three substitutions (A142T, L186S and G338R) leads to complete loss of UGTPg100 activity and that the three amino acids are critical for the catalytic architecture.

The single-site mutated proteins of UGTPg103, UGTPg103-T142A, as well as the double-site mutated protein UGTPg103-T142A/S186L, were not activated to transfer a glucose residue to the C6-OH of PPT. Only the triple-site mutant of UGTPg103,

T142A/S186L/R338G, could obtain an activity to glycosylate the C6-OH of PPT (Figure 4C), which was weaker than that of UGTPg100. The other two amino acid residues, V111 and I349 in UGTPg100, corresponding to I111 and V349 in UGTPg103, respectively, may also contribute to this activity, even if these amino acids bear structurally and characteristically similar side chains.

Mutated proteins UGTPg1-H82C and UGTPg1-H144F as well as a chimeric protein of UGTPg1 and UGTPg100 alter their substrate regio-specificity towards PPT

UGTPg1 and UGTPg100 showed relatively low amino acid identity (84.1%), sharing more than 70 amino acid substitutions, insertions and deletions. They were observed to be regio-specific UGTs, preferring different hydroxyls on PPT. Four amino acid residues around the substrate binding pocket and the catalytic center were selected for further investigation, based on structural modeling (Figure 5A). These amino acids of UGTPg1 were independently converted into the corresponding ones in UGTPg100. All crude cell extracts of recombinant *E. coli* expressing these proteins were subjected to Western blotting analysis to confirm expression (Figure 5B).

Activities of the mutated proteins UGTPg1-A10V and UGTPg1-I13F responsible for glycosylation of the C20-OH of PPT were decreased significantly compared with the wild-type UGTPg1 (Figure 5C). The mutated protein UGTPg1-H82C could not only glycosylate the C20-OH of PPT to produce F1, but also C6-OH of PPT to produce Rh1 (Figure 5C). UGTPg1-H82C altered the substrate regio-specificity, concurrently

glycosylating the two hydroxyls on PPT. Similarly, UGTPg1-H144F showed a weak activity towards the C6-OH of PPT, without compromising its activity towards the C20-OH of PPT (Figure 5C). By contrast, the mutated proteins UGTPg100-F144Y, UGTPg100-F144H and UGTPg100-C82H did not show activity towards PPT (Figure 4C and Figure 5C).

To evaluate the sequence-based functional differences, six chimeric proteins of UGTPg1 and UGTPg100 were constructed (Figure 6A). As illustrated in Figure 6B and 6C, although all the chimeric proteins were successfully expressed in *E. coli* BL21 (DE3), Chi_1, Chi_2, Chi_3, Chi_4 and Chi_5 did not show activity towards PPT. Chi_6, which was closely matched to a cloned UGT gene from *P. ginseng* (Jung et al., 2014) and fell into the cluster of UGTPg1, UGTPg101 and UGTPg102 in the phylogenetic tree (supplementary Figure 2), possessed a strong glucose transfer activity toward the C20-OH of PPT to produce F1 and a weak glycosylation activity toward the C6-OH of PPT to produce Rh1 (Figure 6C). Based on these results, it seems that the C-terminal domain, including the PSPG box of UGTPg100, was critical for the transfer of a glucose moiety to the C6-OH of PPT. These results indicated that the PSPG box with the C-terminal sequence of UGTPg100 must be compatible to ensure the glycosylation of the C6-OH of PPT.

The above results demonstrated that the mutated proteins UGTPg1-H82C, UGTPg1-H144F and Chi_6 altered the substrate regio-specificity, catalyzing the C20-OH and C6-OH of PPT at the same time to produce F1 and Rh1, respectively. Besides, UGTPg1-H144F and Chi_6 could produce a small amount of Rg1 as well as F1 and Rh1 when they were separately incubated with PPT (Figure 5C and Figure 6C). However, no Rg1 was detected when UGTPg1-H82C was incubated with PPT (Figure 5C).

The *in vivo* syntheses of ginsenoside Rh1 and F1 in genetically engineered yeasts

The reconstruction of PPT biosynthetic pathway in yeast was achieved *via* co-expressing the encoding genes of a dammarenediol synthase (PgDDS) (Han et al., 2006; Tansakul et al., 2006), two cytochrome P450 (CYP716A47 and CYP716A53v2) (Han et al., 2011; Han et al., 2012), and an NADPH-cytochrome P450 reductase (PgCPR1) (Yan et al., 2014) isolated from *P. ginseng*. The recombinants to produce Rh1 and F1 were constructed by introducing *UGTPg100* and *UGTPg1* into the PPT producing yeast strain, respectively. Twenty Rh1 and F1 producing transformants were individually selected and their ginsenoside yields in *n*-butanol extract of fermentation supernatant and yeast cells were measured independently by HPLC (Supplementary Figure 5), which represents the intracellular and secreted ginsenosides, respectively. For all of the selected transformants, most of the produced Rh1 or F1 could be secreted to the fermentation supernatant (Supplementary Figure 5). Transformants ZW-F1-17 and ZW-Rh1-20 produced the highest total amount of F1 and Rh1, which were measured as 42.1 mg/L and 92.8 mg/L, respectively (Supplementary Figure 5 and Figure 7A; Table 2). The ginsenoside Rh1 produced in ZW-Rh1-20 and F1 in ZW-F1-17 were further confirmed by HPLC (Figure 7B) and HPLC/ESIMS (Figure 7C). The yields of other ginsenoside intermediates produced in the recombinant yeast cells ZW-F1-17 and ZW-Rh1-20, such as dammarenediol II (DM), PPD, PPT, compound K (CK) and 20(*S*)-*O*-β-(D-glucosyl)-dammarenediol II (DMG) (Yan et al. 2014), were also listed in Table 2. The successful biosyntheses of ginsenoside Rh1 and F1 in the genetically engineered yeasts further confirm that UGTPg1 and UGTPg100 catalyze the glycosylation of C20-OH and C6-OH of PPT, respectively, both *in vitro* and *in vivo*. *De novo* production of Rh1 and F1 *via*

metabolically engineered yeast strains ZW-Rh1-20 and ZW-F1-17 would pave the way to produce these two high-value natural compounds through microbial fermentation in the future.

Discussion

Diversity and similarity among UGTPg genes

All of the five UGTPg genes investigated in this study were cloned from the cDNA sample from a *P. ginseng* plant. It means the five UGTPg genes with high similarity but with different functions or activities coexist in the same genome. The similar phenomena have also been observed in other plants. Three UGTs with the similarities >91%, VvGT5, VvGT6 and VvGT3, were cloned from *Vitis vinifera*. VvGT3 showed no activity to the tested substrates while VvGT5 and VvGT6 were characterized to glycosylate flavonoid using UDP-glucuronic acid and UDP-glucose/UDP-galactose as sugar donors, respectively (Ono et al., 2010). VvGT5, VvGT6 and VvGT3 were found to be located in tandem on chromosome 11, suggesting that one of them may have arisen from the other by gene duplication (Jaillon et al., 2007; Ono et al., 2010). It was possible that VvGT3 might result from the gene duplication followed by a nonfunctionalization event. Augustin et al. (2012) cloned and characterized five UGTs belonging to the UGT73C gene subfamily in *B. vulgaris* with varied substrate specificities. All five UGT sequences were mapped to an existing linkage map of *B. vulgaris* and were found to be located within a region that corresponded to the genome sequences ranged from 13.5 Mb to 19.6 Mb of *A. thaliana* chromosome 2 containing a tandem repeat cluster of UGT73C genes (Augustin et al., 2012; Kuzina et al., 2011). These facts not only supported the speculation that some UGT genes with high sequence identity, but different functions, might have evolved by

gene duplication from a common ancestor, but also suggested that these clustered UGT genes might be co-regulated in secondary metabolic processes.

The process of gene duplication to generate multiplicity in gene copy number results in the common genetic redundancy in plants. The five UGT genes from *P. ginseng* that we have cloned and characterized shared high amino acid identity and might have also undergone similar gene duplication events followed by neofunctionalization or nonfunctionalization. The bifunctional gene UGTPg101, which could catalyze C20-OH glycosylation of PPT and C6-OH glycosylation of F1, might be regarded to be in a promiscuous state, serving as a starting point in evolution (Aharoni et al., 2005; O'Brien and Herschlag, 1999; Weng et al., 2012). This evolutionary intermediate might undergo functional specialization to become UGTPg1 and UGTPg100, which specifically catalyzed the glycosylation of C20-OH and C6-OH, respectively. UGTPg102 and UGTPg103 might evolve to nonfunctionalization or other specialized functions to use acceptors and/or sugar donors which we have not tested in this study. However, the possibility that some of the five UGT genes merely represent allelic variation could not be excluded because *P. ginseng* is a tetraploid plant. Genome sequencing of *P. ginseng* would truly elucidate the chromosomal locations and genetic relationship of the five UGT genes.

Complex substrate regio-specificities of UGTPgs and their key amino acid residues

Glycosylation is the main modification in the complex plant secondary metabolism. Some UGTs were shown to use structurally different compounds as substrates, and some UGTs recognized their substrate in a regio-selective manner. UGTPg1 regio-specifically glycosylated the C20-OH of PPD and its derived ginsenosides (Yan

et al., 2014), and it was also found to regio-specifically glycosylate the C20-OH of PPT in this study. However, it could not catalyze the C20-OH of Rh1 or Rg2, which are PPT-derived ginsenosides with the C6-OH glucosylated (Supplementary Figure 3A). These results indicated that glucose or other sugar moiety at the C6 of PPT might interfere with the glycosylation at C20-OH. By contrast, glucose or other sugar moiety at the C3 of PPD did not interfere with glycosylation of C20-OH (Yan et al., 2014). It seemed that the substrate-binding pocket of UGTPg1 could flexibly accommodate PPD-derived ginsenosides with the C3-OH glycosylated but could not flexibly accommodate PPT-derived ginsenosides with the C6-OH glycosylated. UGTPg100 regio-selectively catalyzed the C6-OH glycosylation of PPT and its derived ginsenoside F1 (Figure 2 and Supplementary Figure 4). The glucose moiety at C20-OH did not seem to interfere with the glycosylation of C6-OH, and UGTPg100 could flexibly accommodate PPT-derived ginsenoside with C20-OH glycosylated. However, the bifunctional UGTPg101 could not catalyze the C6-OH of PPT directly. It initially catalyzed the C20-OH glycosylation of PPT to generate F1 and then catalyzed the C6-OH glycosylation of F1 to produce Rg1. The glycosylation of C20-OH might make UGTPg101 accommodate the C6-OH of F1. Similar sequential glycosylations have also been reported in other plants (Kogawa et al., 2007; Ogata et al., 2005).

The mechanism underlying the varying degrees of flexibility and regio-specificities of UGTPgs to PPD- and PPT-type ginsenosides could be elucidated by combining the resolution of the UGTPgs crystal structure with different substrates and the investigation of the key amino acid residues responsible for their activities and substrate regio-specificities. At present, only six plant UGTs, namely, UGT71G1 (Shao et al., 2005), UGT85H2 (Li et al., 2007) and UGT78G1 (Modolo et al., 2009) from *M. truncatula*, *Vv*GT1 (Offen et al., 2006) from *V. vinifera*, UGT72B1 (Brazier-Hicks et

al., 2007) from *A. thaliana*, and Ct3GT-A (Hiromoto et al., 2013) from *Clitoria ternatea*, have been structurally analyzed. The five UGTPgs in our present study share the highest amino acid identity (43%) with UGT71G1 from *M. truncatula*, among the six reported plant UGTs with available 3D crystal structure information. It has been demonstrated that H22 and D121 are the key catalytic residues of UGT71G1 (Shao et al., 2005). These two residues corresponded to the highly conserved H15 and D117 residues within the five UGTPgs. Three amino acid residues located in the PSPG box, W360, E381 and Q382, were proposed to interact directly with the glucose residue on the sugar donor (Osmani et al., 2009). These three residues were also conserved among the five UGTPgs, corresponding to W364/365, E385/386 and Q386/387, respectively. Moreover, the conserved residues P11, F42, L78, L/F85 and L119 of the five UGTPgs corresponded to P18, F49, L85, I82 and F113 of UGT71G1, which were reported to provide hydrophobic environment for the acceptor binding pocket and to facilitate the accommodation of triterpenoid substrates, such as medicagenic acid and hederagenin (Shao et al., 2005). The above conserved amino acid residues in UGTPg1, UGTPg100 and UGTPg101 may contribute to their ability to catalyze PPT.

Some different amino acid residues between the five UGTPgs were also testified to be critical for their activities or substrate regio-specificities to PPT based on the site-directed mutagenesis results in this study. For example, amino acid residues H144, A10 and I13 were believed to be critical for the UGTPg1's activity towards the C20-OH of PPT (Figure 3 and Figure 5). Besides, amino acid residues H144 and H82 affected the substrate regio-specificity of the UGTs to catalyze PPT (Figure 5C). Similarly, the mutated proteins UGTPg100-C82H, UGTPg100-F144H, UGTPg100-F144Y, UGTPg100-A142T, UGTPg100-L186S and UGTPg100-G338R, showed no activity toward PPT. It seems that amino acid residues F144, C82, A142,

L186 and G338, are also critical for the activities of UGTPg100 to glycosylate the C6-OH of PPT (Figure 4C and Figure 5C).

F148 of UGT71G1, being corresponded to H/F/Y144 of the five UGTPgs, was located at one end of the substrate-binding pocket together with Y202 accommodating the sugar acceptor quercetin (Shao et al., 2005). The mutated protein UGT71G1-F148V switched the glycosylation preference from C3'-OH to C3-OH (He et al., 2006). The mutated protein UGTPg1-H144F was able to glycosylate both the C20-OH and C6-OH of PPT (Figure 5C). These results suggest that the amino acid at this site might determine the substrate regio-specificity of the UGTs in UGT71 subfamily, although other amino acid residues were also detected to exist at the corresponding site of these characterized UGTs in UGT71 subfamily besides H/F144 (Supplementary Figure 6). The difference of UGTPg100 G338 and UGPg103 R338 played a critical role in determining their different activities (Figure 4C). G338/339, located close to the PSPG box, was highly conserved among the characterized UGTs in UGT71 subfamily (Supplementary Figure 6). These results suggest the substitution of the glycine (G), a neutral and small amino acid, to an arginine (R), a positive charged and large one, might disturb the structure of the nearby PSPG box, and thus affect the binding of sugar donor. Besides, the amino acid residues A142 and L186/187 are also relatively conserved in these characterized plant UGTs of UGT71 subfamily (Supplementary Figure 6). By contrast, several amino acid types were detected at the residue sites 10, 13 and 82 of these members from UGT71 subfamily. More case studies are required to make sure whether these conserved amino acid residues, such as G338/339, A142 and L186/187, could also be critical for the activity of other plant UGTs or whether the diverse amino acid types at the sites 10, 13 and 82 represent the diverse substrate selectivities.

The site-directed mutagenesis results from this study give us clues to understand the key amino acid residues which are responsible not only for the activities and substrate regio-specificities of UGTPgs for PPT glycosylation but also for those characteristics of other UGTs in UGT71 subfamily. However, because neither the five UGTPgs was crystallized nor their closely related UGT71G1 was crystallized with triterpenoid substrates, it is still hard to predict the interaction between UGTPgs and their substrates PPT or PPT-type ginsenosides through homologous modeling. In future studies, it is necessary to uncover the crystal structures of UGTPg1, UGTPg100 and UGTPg101 and then to further elucidate the mechanisms for their regio-selectivity or substrate specificity based on systems biology and protein engineering approaches. These studies would accumulate more knowledge to understand the relations between the structure and characterizations of plant UGTs, which may give clues to design and rebuild UGTs by protein engineering to increase enzyme catalytic efficiencies or expand their substrate selectivities to create non-natural therapeutic compounds.

The possible biosynthetic pathways of PPT-type ginsenosides *in vivo*

UGTPg1 and UGTPg100 were characterized to specifically catalyze the glucosylation of C20-OH and C6-OH of PPT to produce F1and Rh1in *in vitro* reactions, respectively (Figure 2B). The successful biosyntheses of ginsenoside Rh1 and F1 in the genetically engineered yeasts further referred that UGTPg1 and UGTPg100 should be the key enzymes to produce F1 and Rh1 from PPT in *Panax* plants, respectively. In *Panax* plants, being possibly like in yeast recombinant ZW-F1-17, UGTPg1 have three potential substrates, DM, PPD and PPT. According to the V_{max}, K_m, k_{cat} and k_{cat}/K_m for UGTPg1 to catalyze the glycosylation of PPT (Table 1), PPD

and DM (Yan et al. 2014), UGTPg1 is found to prefer PPT over DM or PPD as its substrate, because both its affinity and the turn-over number towards PPT are higher than those of DM or PPD. That is why the yield of F1 is much higher than that of CK in ZW-F1-17 (Table 2). The UGTs with higher affinity and the turn-over number towards PPT, such as UGTPg1 and UGTPg100, together with the relatively high activity of CYP716A53v2 catalyzing PPD to produce PPT, make sure there is a certain ratio of PPT-type ginsenosides in *Panax* plants (Shibata, 2001). In our previous study, UGTPg1 could glycosylate the C20-OH of DM to produce DMG, however, DMG was not detected in ZW-F1-17. This is possibly due to that (i) the DM production is too low and (ii) DMG can be subsequently converted into CK by CYP716A47 as proved by our previous work (Yan et al. 2014). DMG has also never been identified in *Panax* plants.

Rg1 and Re are the main PPT-type ginsenosides in *Panax* plants (Shibata, 2001). According to the substrate specificities of the cloned UGTs toward PPT from *P. ginseng* in this study, it is possible that the biosynthetic pathway of ginsenoside Rg1 is an orderly sequential transfer of glucose moieties to the aglycone PPT in *Panax* plants (Figure 2D): (i) UGTPg1 initially glycosylates the C20-OH of PPT to produce F1, then UGTPg100 glycosylates the C6-OH of F1 to produce Rg1; (ii) UGTPg101 initially glycosylates the C20-OH of PPT to produce F1, then glycosylates the C6-OH of F1 to produce Rg1 alone. Re is supposed to be produced from Rg1 by transferring a rhamnose moiety to its C6-O-Glc or be produced from Rg2 by transferring a glucosyl moiety to its C20-OH. However, because UGTPg1 and UGTPg101 could not glycosylate the C20-OH of Rg2 (Supplemental Figure 3A), it is mostly possible that Re is produced from Rg1.

Compared with PPD and PPD-type ginsenosides, PPT and PPT-type ginsenosides

have an extra free hydroxyl, which may alter their hydrophilicity and thus subcellular localization. It is not surprised that the PPT-type ginsenosides, F1 and Rh1, can be more efficiently secreted to the fermentation supernatant compared with CK, a PPD-type ginsenoside (Supplementary Figure 5; Yan et al. 2014). The biosynthetic pathways of Rh1 and F1 and intermediate metabolites present in yeast cell as well as their subcellular localization may also reflect their biosyntheses and hydrophilicity in *Panax* plants.

The results in this study provide not only the scientific bases for understanding the biosynthetic pathways of PPT-type ginsenosides within *Panax* plants, but also the potential approach to biosynthesize the bioactive PPT-type ginsenosides Rh1, F1 and Rg1 in yeast or in plants outside *Panax* genus with well-established genetic manipulation systems, such as in tobacco (*Nicotiana benthamiana*), rice (*Oryza sativa*) or tomato (*Lycopersicon esculentum* Mill). The transgenic plants producing these ginsenosides may act as plant bioreactors or dietary supplements.

Methods

Cloning of UDP-glycosyltransferase encoding cDNAs

The RNA was extracted from *P. ginseng* callus using Trizol (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instruction and then converted into cDNA using the PrimeScript RT reagent Kit with gDNA eraser (Takara, Dalian, China). The cDNAs potentially encoding UDP-glycosyltransferase with long ORFs (likely to be full-length) were PCR amplified using the *P. ginseng* cDNA as the template and cloned into the pMD18T vector (Takara, Dalian, China).

Site-directed mutagenesis and construction of chimeric UGT-encoding genes

The UGT genes *UGTPg1*, *UGTPg100*, *UGTPg101*, *UGTPg102* and *UGTPg103*, with a C-terminal 6×His-tag, were ligated into the pET28a vector using T4 ligase (NEB, Ipswich, MA, USA). Site-directed mutagenesis was conducted using the Hieff Mut™ Site-Directed Mutagenesis Kit (Yeasen Biotech, Shanghai, China), and the primers are listed in Supplemental Table 3.

To construct the chimeric UGT-encoding genes of UGTPg1 and UGTPg100, four splitting points not located in the protein secondary structures, such as α-helices and β-sheets, were identified according to structural modeling. Primers C1-F/C1-R, C2-F/C2-R, C3-F/C3-R and C4-F/C4-R were designed to amplify the four split DNA fragments of UGTPg1 and UGTPg100. To construct Chi_1, DNA fragment Chi_1_up was amplified with the primers C1-F/C1-R using UGTPg100-pMD18T as the template. DNA fragment Chi_1_down was amplified with primers C2-F/C4-R using UGTPg1-pMD18T as the template. A linearized vector was prepared by PCR using primers pet-F/pet-R and template pET28a. Then, the Chi_1 expressing vector, Chi_1-pET28a, was constructed by *in vitro* recombination of linearized vector,

Chi_1_up and Chi_1_down using the ClonExpress® MultiS One Step Cloning Kit (Vazyme, Nanjing, China). The other chimeric UGT genes were constructed similarly.

Heterologous expression of the UGT genes in *E. coli* BL21 (DE3)

Heterologous expression of the UGT genes in *E.coli* BL21 (DE3) was carried out as described previously (Yan et al., 2014). Briefly, UGT genes with a C-terminally 6×His-tag was ligated into the pET28a vector and expressed in *E. coli* BL21 (DE3) under 100 μM IPTG induction. The cells were collected by centrifugation and suspended in 100 mM phosphate buffer (pH 8.0) supplemented with 1 mM PMSF and then disrupted with a French Press (25 kpsi). The cell debris was removed by centrifugation at 12000 g for 20 min, and the supernatant was used for enzymatic assays and Western blot. The pET28a-transformed *E. coli* BL21 (DE3) cells were treated in parallel as a control. Western blotting was conducted using an anti-6×His primary antibody and HRP-conjugated secondary antibody (Abmart, Shanghai, China). Blots were visualized by chemiluminescence with ECL Western Blot Substrate (Pierce, Rockford, IL, USA) according to the manufacturer's protocol.

Quantification of the UGTPg1 and UGTPg100 in the crude cell extract was carried out as described previously (Yan et al., 2014). Serially diluted samples (0, 2, 4, 6, 8, 10 ng/μL) of a C-terminal 6×His-tagged xylanase (53 kDa) were used to generate a standard curve for quantitative reference.

Enzymatic assays for UGTPgs

Enzymatic assays for glycosyltransferases were performed as described by (Yan et al., 2014). The reaction was performed in a 100-μL volume containing 100 mM phosphate buffer (pH 8.0), 1% Tween-20, 5 mM UDP-glucose, 0.5 mM acceptor substrate and

recombinant *E. coli* extract in a 35°C water bath overnight and was terminated by adding 100 μL *n*-butanol. The product was extracted into the organic phase, which was then evaporated. The residue was dissolved in methanol for TLC and HPLC analyses.

For the kinetic analysis of UGTPg1 and UGTPg100, the reaction mixture contained 100 mM phosphate buffer (pH 8.0), 1% Tween-20, 5 mM UDP-glucose, acceptor substrate (40-500 μM PPD) and 100 ng UGTPg1 or UGTPg100 in a final volume of 100 μL. The reaction was incubated at 40°C for 20 min. HPLC analysis was used to quantify the target product in each reaction. The Michaelis-Menten parameters were calculated by least-squares fitting of the kinetic model using Prism 5 (GraphPad, San Diego, CA, USA). All data are presented as means ± SD of three independent experiments.

Construction of yeast strains to produce ginsenoside Rh1 and F1

Ginsenoside Rh1 and F1 producing yeasts were constructed based on the PPD producing chassis yeast ZW-PPD-B constructed in our previous study (Wang et al., 2015). Ginsenoside Rh1 producing strains, namely ZW-Rh1 transformants, were constructed by introducing *CYP716A53v2* (Han et al. 2012), *PgCPR1* (Yan et al. 2014) and *UGTPg100* genes (under the control of constitutive promoter *TEF1*, *TDH3* and *PGK1* and followed by the terminator *FBA1*, *ADH1* and *TPI1*, respectively) into the δ DNA site of ZW-PPD-B. Similarly, ginsenoside F1 producing strains, namely ZW-F1 transformants, was constructed by introducing *CYP716A53v2*, *PgCPR1* and *UGTPg1* genes (under the control of constitutive *TEF1*, *TDH3* and *PGK1* promoter and followed by the terminator *FBA1*, *ADH1* and *TPI1*, respectively) into the δ DNA site of ZW-PPD-B. Twenty transformants of ZW-Rh1 and ZW-F1 were selected to be cultured in flasks for 6 days and the concentrations of their ginsenosides and

intermediates in the *n*-butanol extract of culture supernatant and yeast cells were analyzed individually by HPLC, just as described in our previous study (Yan, et al. 2014)

TLC and HPLC analysis

TLC analysis was performed using silica gel 60 F254 plates (Merck KGaA, Darmstadt, Germany) as described previously (Yan et al., 2014). A mixture of PPD-type ginsenoside authentic samples (PPD, CK, Rh2, F2, Rg3, Rd,) or a mixture of PPT-type ginsenoside authentic samples (PPT, F1, Rh1, Rg1), 1 mg/mL of each dissolved in methanol, was included as a marker. Ginsenoside authentic samples were purchased from Nanjing Zelang Medical Technology Co., Ltd. (Nanjing, China).

A Shimadzu LC-20A prominence system (Shimadzu, Kyoto, Japan) was used for the HPLC analysis. Chromatographic separations were carried out at 35°C on a Shodex C18-120-5 4E column (5 μm, 4.6 mm × 250 mm). The gradient elution system consisted of water (A) and acetonitrile (B). The HPLC program for the extracted yeast fermentation was similar as described previously (Yan et al., 2014). The HPLC program for the extracted compounds from the *in vitro* reactions by incubating UGTPgs with PPT or PPT-type ginsenosides as substrates was: 0 min (22.5% B), 0-55 min (62.5% B), and 55-60min(22.5% B). The flow rate was kept at 0.8 mL/min.

HPLC/ESIMS analysis

The chromatographic analysis was performed using an Agilent 1200 Series LC system (Agilent Technologies, Waldbronn, Germany) equipped with a Shodex C18-120-5 4E column. ESIMS data were acquired using a Bruker Daltonics Esquire 3000plus (Bruker Daltonics, MA, USA) in the positive ionization mode.

NMR analysis

^{13}C NMR experiments were performed in pyridine-d_5 for F1 and Rh1 on Bruker Avance III 500 (Bruker, Billerica, MA, USA) with reference to the solvent peaks.

Accession numbers

Sequence data from this article can be found in the EMBL/GenBank data libraries under accession number KP795113, KP795114, KP795115 and KP795116.

Author contributions

W. W. and P.W. performed the biological experiments; Q. L., X. Y. and C. Y. carried out the chemical analyses; Y. W. performed the bioinformatic analyses; Z. Z., X. Y., J. Y., and G. Z. supervised and coordinated the experiments; W. W., P. W. and Z. Z. wrote the manuscript. W. W. and P. W. contributed equally to this work.

Acknowledgement

This work was financially supported by National Basic Research Program of China (973 program: 2012CB721103 and 2015CB755703) and Science and Technology Commission of Shanghai Municipality (14JC1407000).

References

Achnine, L., Huhman, D.V., Farag, M.A., Sumner, L.W., Blount, J.W., and Dixon, R.A. (2005). Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. Plant J. **41**:875-887.

Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S.M., Roodveldt, C., and

Tawfik, D.S. (2005). The 'evolvability' of promiscuous protein functions. Nat Genet. **37**:73-76.

Augustin, J.M., Drok, S., Shinoda, T., Sanmiya, K., Nielsen, J.K., Khakimov, B., Olsen, C.E., Hansen, E.H., Kuzina, V., and Ekstrøm, C.T. (2012). UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyze sapogenin 3-O-glucosylation in saponin-mediated insect resistance. Plant Physiol. **160**:1881-1895.

Bowles, D., Lim, E.-K., Poppenberger, B., and Vaistij, F.E. (2006). Glycosyltransferases of lipophilic small molecules. Annu. Rev. Plant Biol. **57**:567-597.

Brazier-Hicks, M., Offen, W.A., Gershater, M.C., Revett, T.J., Lim, E.-K., Bowles, D.J., Davies, G.J., and Edwards, R. (2007). Characterization and engineering of the bifunctional N-and O-glucosyltransferase involved in xenobiotic metabolism in plants. Proc. Natl. Acad. Sci. U. S. A. **104**:20238-20243.

Chen, S., Luo, H., Li, Y., Sun, Y., Wu, Q., Niu, Y., Song, J., Lv, A., Zhu, Y., and Sun, C. (2011). 454 EST analysis detects genes putatively involved in ginsenoside biosynthesis in *Panax ginseng*. Plant Cell Rep. **30**:1593-1601.

Christensen, L.P. (2009). Ginsenosides chemistry, biosynthesis, analysis, and potential health effects. Adv. Food Nutr. Res. **55**:1-99.

Fukuchi-Mizutani, M., Okuhara, H., Fukui, Y., Nakao, M., Katsumoto, Y., Yonekura-Sakakibara, K., Kusumi, T., Hase, T., and Tanaka, Y. (2003). Biochemical and molecular characterization of a novel UDP-glucose: anthocyanin 3′-O-glucosyltransferase, a key enzyme for blue anthocyanin biosynthesis, from gentian. Plant Physiol. **132**:1652-1663.

Han, J.-Y., Kim, H.-J., Kwon, Y.-S., and Choi, Y.-E. (2011). The Cyt P450 enzyme CYP716A47 catalyzes the formation of protopanaxadiol from dammarenediol-II during ginsenoside biosynthesis in *Panax ginseng*. Plant Cell Physiol. **52**:2062-2073.

Han, J.Y., Hwang, H.S., Choi, S.W., Kim, H.J., and Choi, Y.E. (2012). Cytochrome P450 CYP716A53v2 catalyzes the formation of protopanaxatriol from

protopanaxadiol during ginsenoside biosynthesis in *Panax ginseng*. Plant Cell Physiol **53**:1535-1545.

Han, J.Y., Kwon, Y.S., Yang, D.C., Jung, Y.R., and Choi, Y.E. (2006). Expression and RNA interference-induced silencing of the dammarenediol synthase gene in *Panax ginseng*. Plant Cell Physiol. **47**:1653-1662.

He, X.Z., Wang, X., and Dixon, R.A. (2006). Mutational analysis of the *Medicago* glycosyltransferase UGT71G1 reveals residues that control regioselectivity for (iso)flavonoid glycosylation. J Biol Chem **281**:34441-34447.

Hiromoto, T., Honjo, E., Tamada, T., Noda, N., Kazuma, K., Suzuki, M., and Kuroki, R. (2013). Crystal structure of UDP-glucose:anthocyanidin 3-*O*-glucosyltransferase from *Clitoria ternatea*. J. Synchrotron Radiat. **20**:894-898.

Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449**:463-U465.

Jung, S.C., Kim, W., Park, S.C., Jeong, J., Park, M.K., Lim, S., Lee, Y., Im, W.T., Lee, J.H., Choi, G., et al. (2014). Two ginseng UDP-glycosyltransferases synthesize ginsenoside Rg3 and Rd. Plant Cell Physiol **55**:2177-2188.

Kogawa, K., Kato, N., Kazuma, K., Noda, N., and Suzuki, M. (2007). Purification and characterization of UDP-glucose: anthocyanin 3',5'-O-glucosyltransferase from *Clitoria ternatea*. Planta **226**:1501-1509.

Kubo, A., Arai, Y., Nagashima, S., and Yoshikawa, T. (2004). Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. Arch Biochem Biophys **429**:198-203.

Kuzina, V., Nielsen, J.K., Augustin, J.M., Torp, A.M., Bak, S., and Andersen, S.B. (2011). *Barbarea vulgaris* linkage map and quantitative trait loci for saponins, glucosinolates, hairiness and resistance to the herbivore *Phyllotreta nemorum*. Phytochemistry **72**:188-198.

Li, C., Zhu, Y., Guo, X., Sun, C., Luo, H., Song, J., Li, Y., Wang, L., Qian, J., and

Chen, S. (2013). Transcriptome analysis reveals ginsenosides biosynthetic genes, microRNAs and simple sequence repeats in *Panax ginseng* CA Meyer. BMC genomics **14**:245.

Li, L., Modolo, L.V., Escamilla-Trevino, L.L., Achnine, L., Dixon, R.A., and Wang, X. (2007). Crystal structure of *Medicago truncatula* UGT85H2–insights into the structural basis of a multifunctional (Iso) flavonoid glycosyltransferase. J. Mol. Biol. **370**:951-963.

Luo, H., Sun, C., Sun, Y., Wu, Q., Li, Y., Song, J., Niu, Y., Cheng, X., Xu, H., and Li, C. (2011). Analysis of the transcriptome of *Panax notoginseng* root uncovers putative triterpene saponin-biosynthetic genes and genetic markers. BMC genomics **12**:S5.

Meesapyodsuk, D., Balsevich, J., Reed, D.W., and Covello, P.S. (2007). Saponin biosynthesis in *Saponaria vaccaria*. cDNAs encoding β-amyrin synthase and a triterpene carboxylic acid glucosyltransferase. Plant Physiol. **143**:959-969.

Modolo, L.V., Li, L., Pan, H., Blount, J.W., Dixon, R.A., and Wang, X. (2009). Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso) flavonoids. J. Mol. Biol. **392**:1292-1302.

O'Brien, P.J., and Herschlag, D. (1999). Catalytic promiscuity and the evolution of new enzymatic activities. Chem. Biol. **6**:R91-R105.

Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B.G., Tarling, C.A., Ford, C.M., Bowles, D.J., and Davies, G.J. (2006). Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. EMBO J. **25**:1396-1405.

Ogata, J., Kanno, Y., Itoh, Y., Tsugawa, H., and Suzuki, M. (2005). Plant biochemistry: anthocyanin biosynthesis in roses. Nature **435**:757-758.

Ono, E., Homma, Y., Horikawa, M., Kunikane-Doi, S., Imai, H., Takahashi, S., Kawai, Y., Ishiguro, M., Fukui, Y., and Nakayama, T. (2010). Functional differentiation of the glycosyltransferases that contribute to the chemical diversity of bioactive flavonol glycosides in grapevines (*Vitis vinifera*). Plant

Cell **22**:2856-2871.

Osmani, S.A., Bak, S., and Moller, B.L. (2009). Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. Phytochemistry **70**:325-347.

Poppenberger, B., Berthiller, F., Lucyshyn, D., Sieberer, T., Schuhmacher, R., Krska, R., Kuchler, K., Glossl, J., Luschnig, C., and Adam, G. (2003). Detoxification of the *Fusarium* mycotoxin deoxynivalenol by a UDP-glucosyltransferase from *Arabidopsis thaliana*. J. Biol. Chem. **278**:47905-47914.

Poppenberger, B., Fujioka, S., Soeno, K., George, G.L., Vaistij, F.E., Hiranuma, S., Seto, H., Takatsuto, S., Adam, G., and Yoshida, S. (2005). The UGT73C5 of *Arabidopsis thaliana* glucosylates brassinosteroids. Proc. Natl. Acad. Sci. U. S. A. **102**:15253-15258.

Richman, A., Swanson, A., Humphrey, T., Chapman, R., McGarvey, B., Pocs, R., and Brandle, J. (2005). Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of *Stevia rebaudiana*. Plant J. **41**:56-67.

Shao, H., He, X., Achnine, L., Blount, J.W., Dixon, R.A., and Wang, X. (2005). Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. Plant Cell **17**:3141-3154.

Shibata, S. (2001). Chemistry and cancer preventing activities of ginseng saponins and some related triterpenoid compounds. J. Korean Med. Sci. **16**:S28-S37.

Shibuya, M., Nishimura, K., Yasuyama, N., and Ebizuka, Y. (2010). Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in *Glycine max*. FEBS Lett **584**:2258-2264.

Sun, C., Li, Y., Wu, Q., Luo, H., Sun, Y., Song, J., Lui, E., and Chen, S. (2010). De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. BMC genomics **11**:262.

Tansakul, P., Shibuya, M., Kushiro, T., and Ebizuka, Y. (2006). Dammarenediol-II

synthase, the first dedicated enzyme for ginsenoside biosynthesis, in *Panax ginseng*. FEBS Lett. **580**:5143-5149.

Wang, P., Wei, Y., Fan, Y., Liu, Q., Wei, W., Yang, C., Zhang, L., Zhao, G., Yue, J., Yan, X., et al. (2015). Production of bioactive ginsenosides Rh2 and Rg3 by metabolically engineered yeasts. Metab Eng **29**:97-105.

Weng, J.K., Philippe, R.N., and Noel, J.P. (2012). The rise of chemodiversity in plants. Science **336**:1667-1670.

Yan, X., Fan, Y., Wei, W., Wang, P., Liu, Q., Wei, Y., Zhang, L., Zhao, G., Yue, J., and Zhou, Z. (2014). Production of bioactive ginsenoside compound K in metabolically engineered yeast. Cell Res **24**:770-773.

Yun, T.-K. (2001). *Panax ginseng*-a non-organ-specific cancer preventive? Lancet Oncol. **2**:49-55.

Table 1. The kinetic parameters of UGTPg1 and UGTPg100 towards PPT.

Enzyme	V_{max} (nmol/min/mg)	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (μM^{-1} s^{-1})
UGTPg1	386±14	104±12	17.9	0.17
UGTPg100	426±20	153±10	9.1	0.06

Table 2. The yields of the ginsenosides and the intermediates in yeast cell factories.

Strains	Rh1 mg/L	F1 mg/L	PPT mg/L	CK mg/L	PPD mg/L	DMG mg/L	DM mg/L	Total mg/L
ZW-Rh1-20	92.8±12.5	-	3.5±0.5	-	43.4±2.7	-	8.8±0.3	148.5
ZW-F1-17	-	42.1±3.2	13.9±1.5	7.5±1.7	49.2±1.4	ND	3.5±0.4	116.2

ND represents no detected.

Figure legend

Figure 1. The amino acid alignment of the five UGTs. ClustalW was used to do the alignment. Identical amino acids are shaded in black and similar ones are in gray. Dashes represents the gaps introduced to improve the alignment.

Figure 2. Characterizations of the five UGTPgs from *P. ginseng* to catalyze protopanaxatriol (PPT) and PPT-type ginsenosides. **A**, Western blotting of the five UGTs heterologously expressed in *E. coli* BL21 (DE3). The cell extract of recombinant *E. coli* BL21 harboring empty vector pET28a was used as a negative control. **B**, HPLC analyses of the *in vitro* reactions of UGTs catalyzing the substrates PPT, F1 and Rh1. **C**, the MS spectrum of F1 and Rh1 produced by an *in vitro* reaction catalyzed by UGTPg1 and UGTPg100, respectively. **D**, the proposed pathway for Rg1 biosynthesis based on the functions of the three UGTs, UGTPg1, UGTPg100 and UGTPg101. Solid arrows represent characterized steps, dashed arrow represents unidentified step.

Figure 3. Characterizations of the UGTPg1 and UGTPg102 mutated proteins catalyzing the substrate PPT. **A**, structural modeling of UGTPg1 based on UGT71G1 from *M. truncatula*. The two conserved catalytic residues, H15 and D117, were marked in red. The six amino acid residues of UGTPg1 that differed from those of UGTPg102 located around the substrate binding pocket and catalytic center were labeled in purple. **B**, Western blotting of the UGTs and the mutated proteins heterologously expressed in *E. coli* BL21 (DE3). The cell extract of recombinant *E. coli* BL21 harboring empty vector pET28a was used as a negative control. **C**, HPLC analyses of the *in vitro* reactions of the UGTs and the mutated proteins catalyzing the substrate PPT.

Figure 4. Characterizations of UGTPg100 and UGTPg103 mutated proteins catalyzing the substrate PPT. **A**, structural modeling of UGTPg100 based on UGT71G1 from *M. truncatula*. The two conserved catalytic residues, H15 and D117, were labeled in red. The six amino acid residues of UGTPg100 differing from those of UGTPg103 were

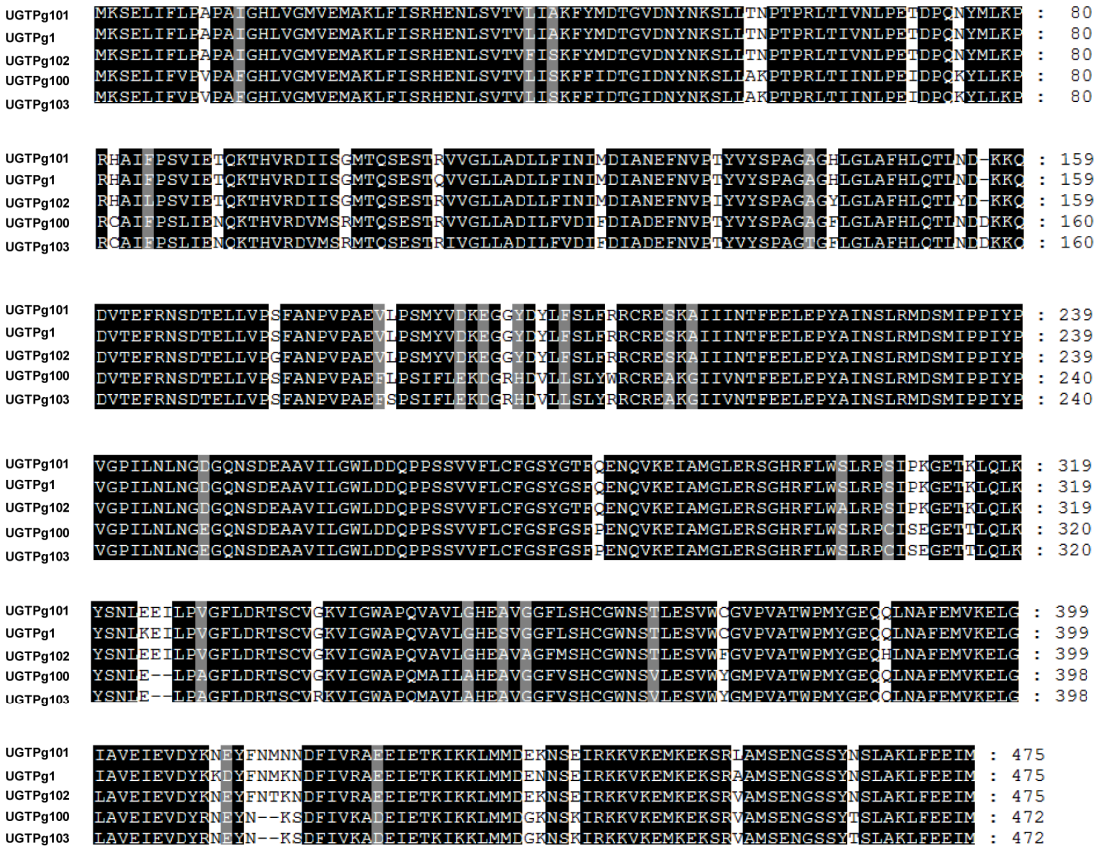
labeled in purple and blue. **B**, Western blotting of the UGTs and the mutated proteins heterologously expressed in *E. coli* BL21 (DE3). The cell extract of recombinant *E. coli* BL21 harboring empty vector pET28a was used as a negative control. UGTPg103-double mutation and UGTPg103-triple mutation represent UGTPg103-T142A/S186L and UGTPg103-T142A/S186L/R338G, respectively. **C**, HPLC analyses of the *in vitro* reactions of UGTs and the mutated proteins catalyzing PPT.

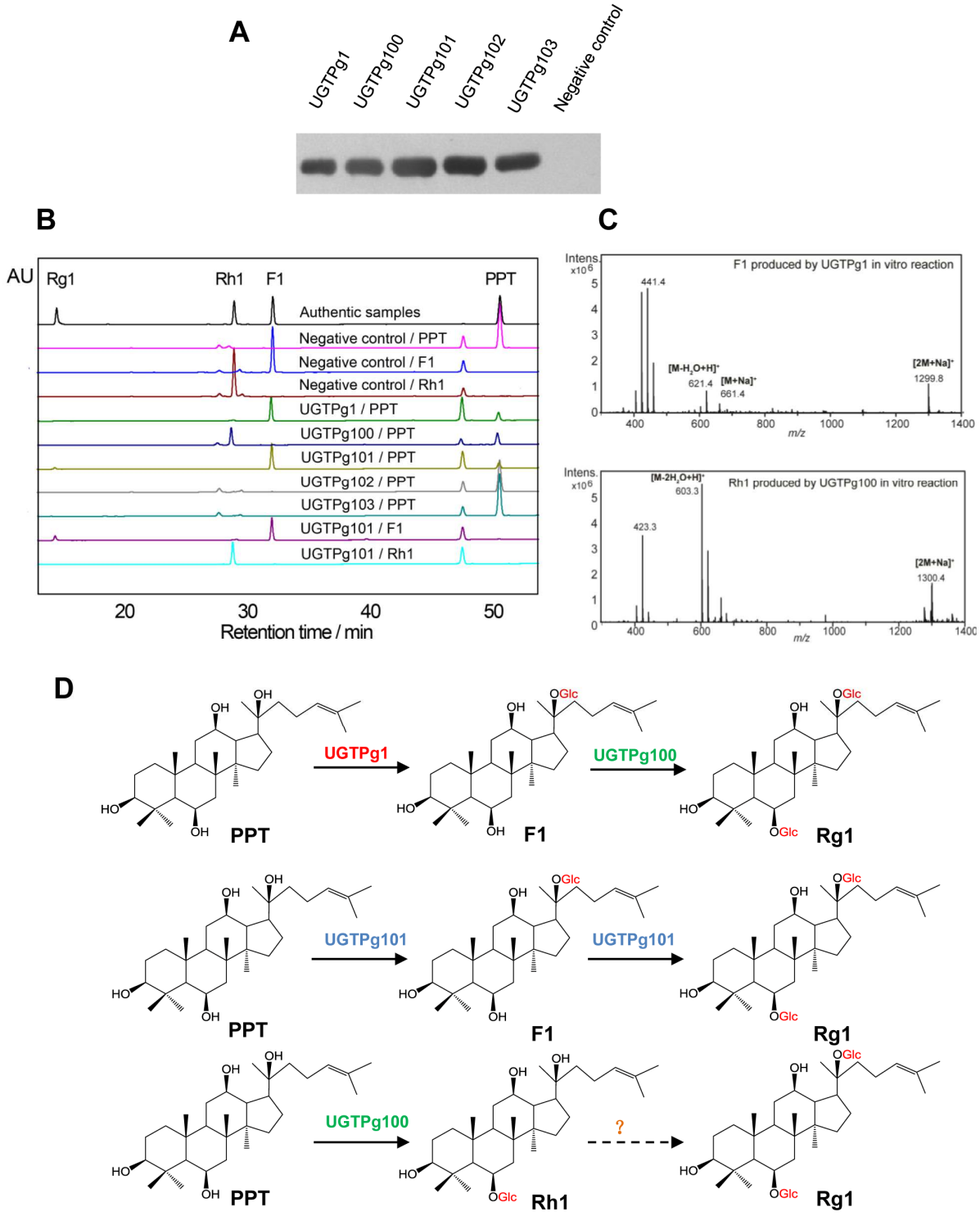
Figure 5. Characterizations of the UGTPg1 and UGTPg100 mutated proteins catalyzing the substrate PPT. **A**, structural modeling of UGTPg1 based on UGT71G1 from *M. truncatula*. The two conserved catalytic residues, H15 and D117, were labeled in red. The four amino acid residues of UGTPg1 differing from those of UGTPg102 located around the substrates binding pocket and catalytic center were labeled in purple. **B**, Western blotting of the UGTs and the mutated proteins heterologously expressed in *E. coli* BL21 (DE3). The cell extract of recombinant *E. coli* BL21 harboring empty vector pET28a was used as a negative control. **C**, HPLC analyses of the *in vitro* reactions of the UGTs and the mutated proteins catalyzing the substrate PPT.

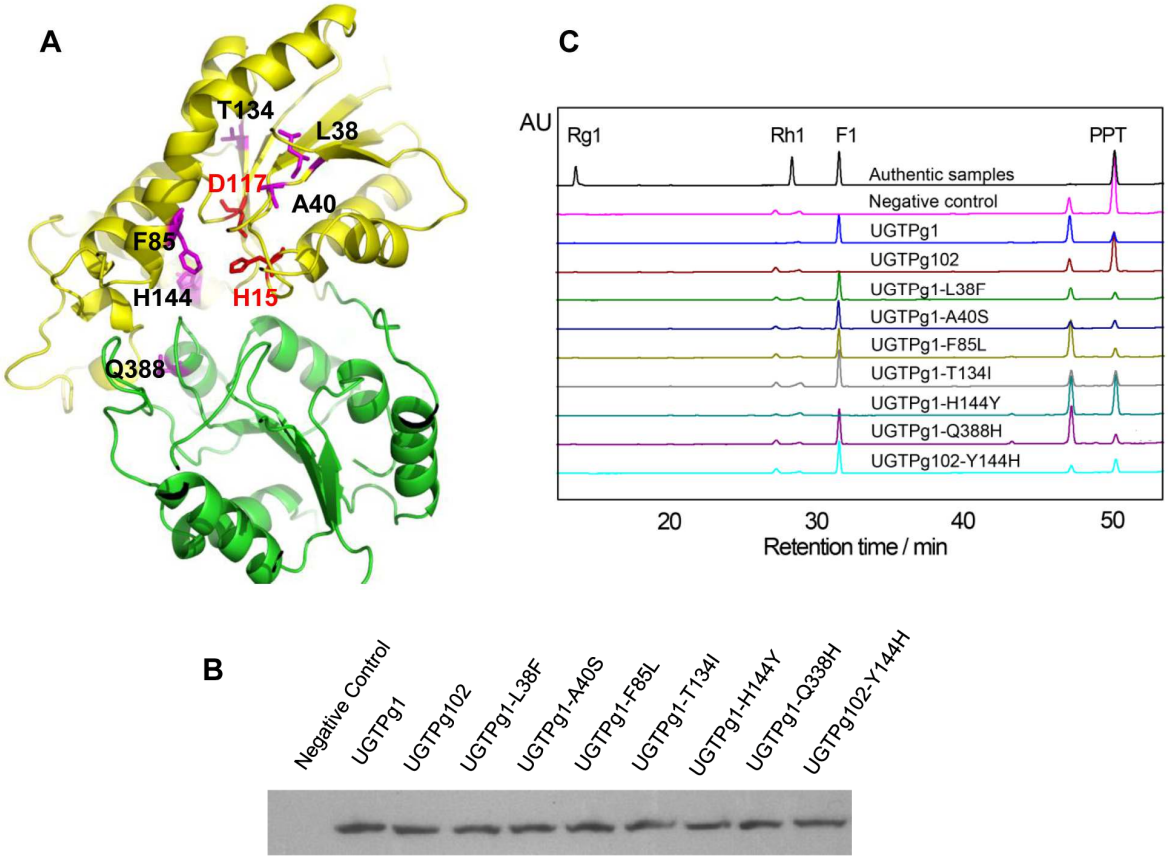
Figure 6. Construction and characterizations of chimeric proteins of UGTPg1 and UGTPg100 catalyzing the substrate PPT. **A**, schematic representation of the six chimeric proteins with the key amino acid residues labeled. **B**, Western blotting of the UGTs and the chimeric proteins heterologously expressed in *E. coli* BL21 (DE3). The cell extract of recombinant *E. coli* BL21 harboring empty vector pET28a was used as a negative control. **C**, HPLC analyses of the *in vitro* reactions of the UGTs and chimeric proteins catalyzing the substrate PPT.

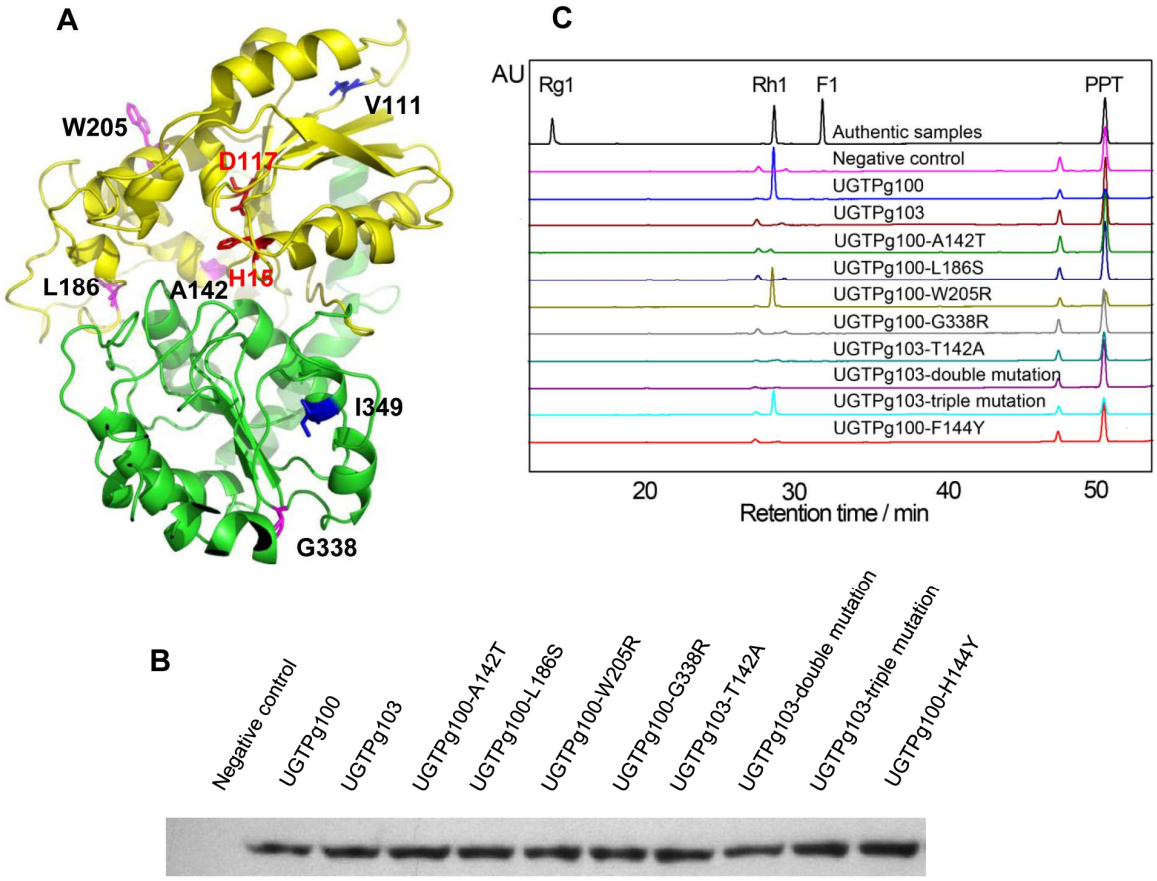
Figure 7. Productions of F1 and Rh1 in genetically engineered yeasts. **A**, the production of ginsenosides and the intermediates in ZW-F1-17 (up panel) and ZW-Rh1-20 (down panel). Blue and red bars represent intracellular and extracellular production of the corresponding metabolites, respectively. **B**, HPLC analyses of the

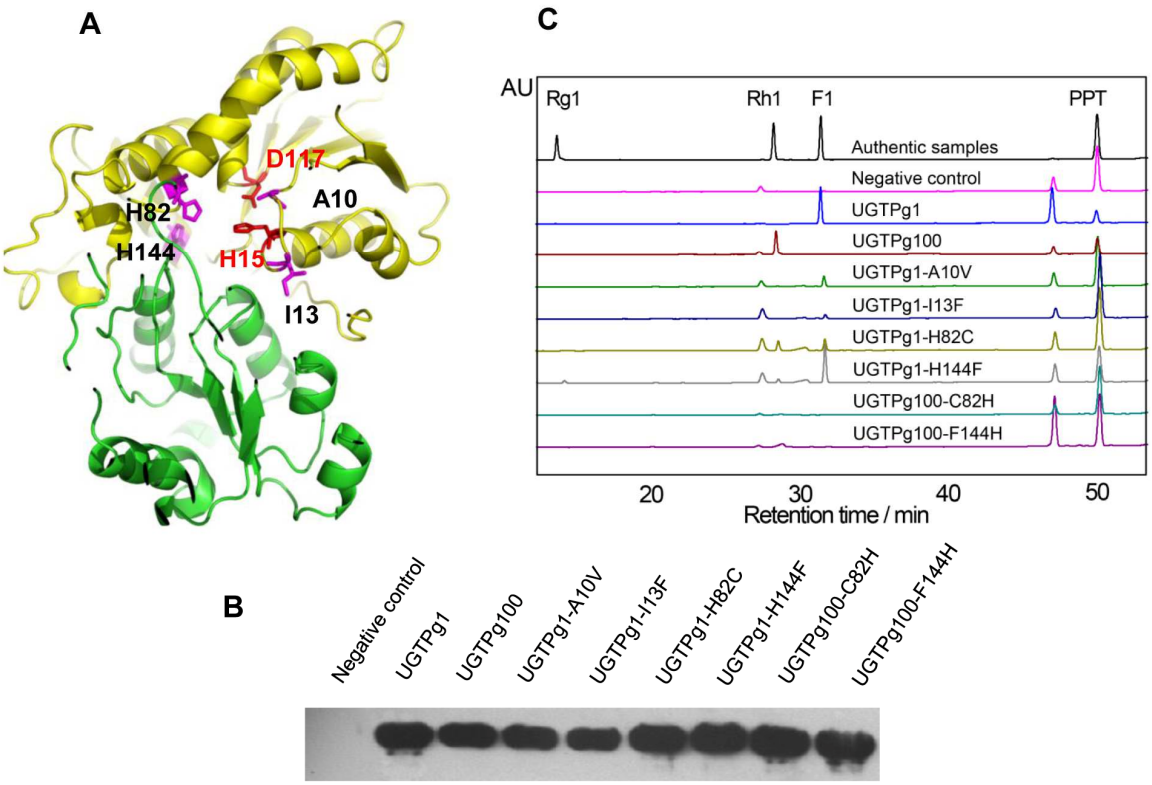
n-butanol extract of yeast strains producing ginsenoside F1 or Rh1. 1, authentic samples of DM, PPD, DMG and PPD-type ginsenoside; 2, authentic samples of PPT and PPT-type ginsenoside; 3, the strain ZW-F1-17; 4, the strain ZW-Rh1-20; 5, the wild-type strain BY4742 as a control. **C**, the MS spectrum of F1 and Rh1 produced by ZW-F1-17 (up panel) and ZW-Rh1-20 (down panel).

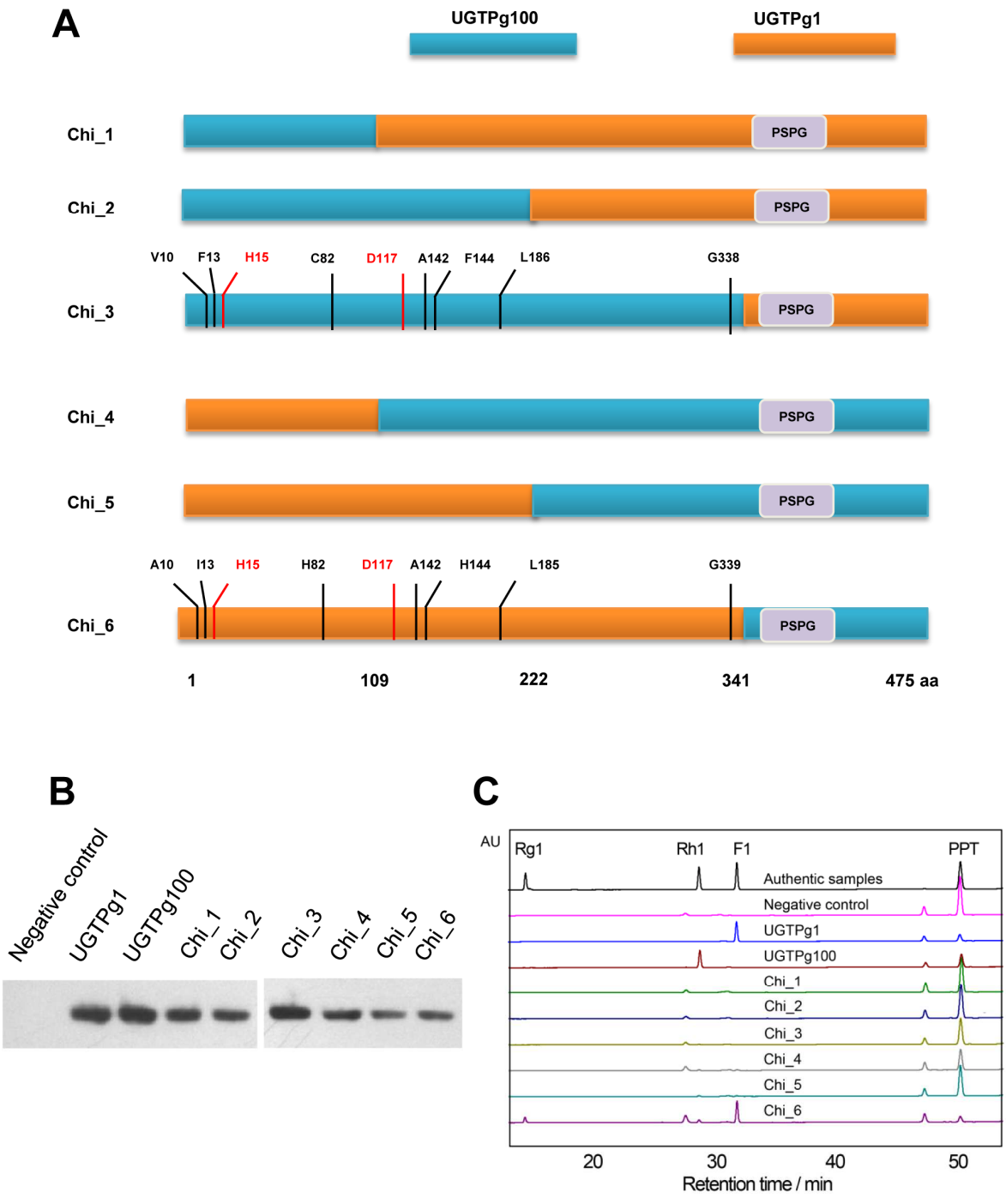


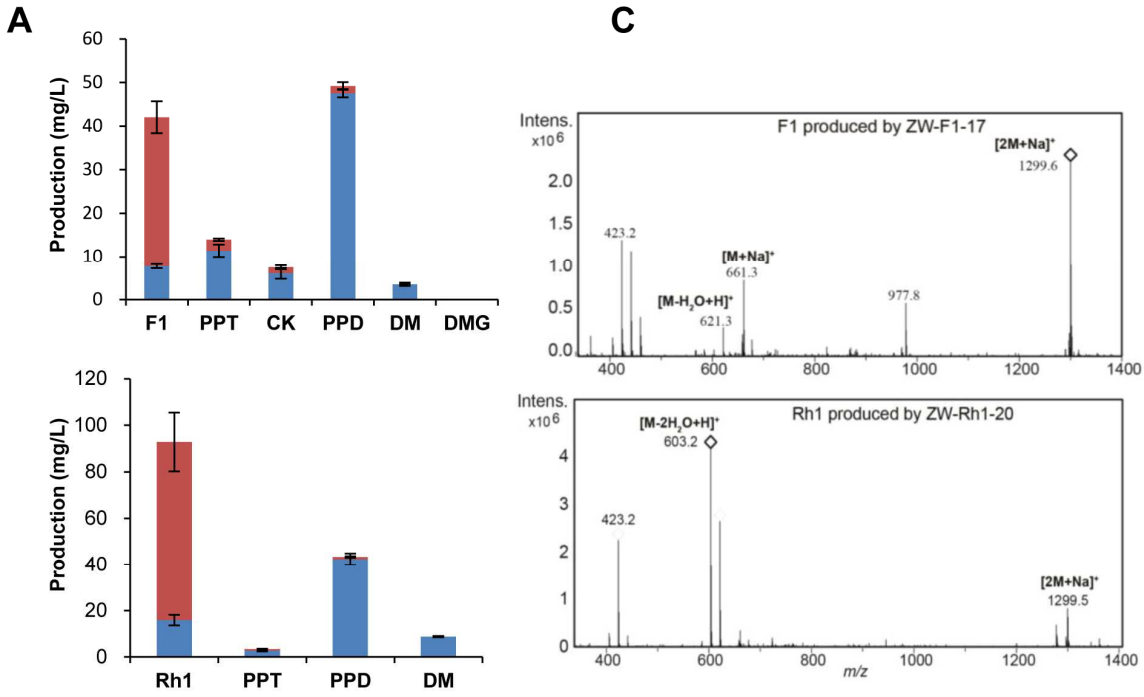

A
UGTPg1
UGTPg100
UGTPg101
UGTPg102
UGTPg103
Negative control
B
AU
Rg1
Rh1
F1
PPT
Authentic samples
Negative control / PPT
Negative control / F1
Negative control / Rh1
UGTPg1 / PPT
UGTPg100 / PPT
UGTPg101 / PPT
UGTPg102 / PPT
UGTPg103 / PPT
UGTPg101 / F1
UGTPg101 / Rh1
20
30
40
50
Retention time / min
C
F1 produced by UGTPg1 in vitro reaction
441.4
621.4
661.4
1299.8
Rh1 produced by UGTPg100 in vitro reaction
603.3
423.3
1300.4
m/z
D
UGTPg1
UGTPg100
UGTPg101
UGTPg101
UGTPg100
?
PPT
F1
Rg1
PPT
F1
Rg1
PPT
Rh1
Rg1

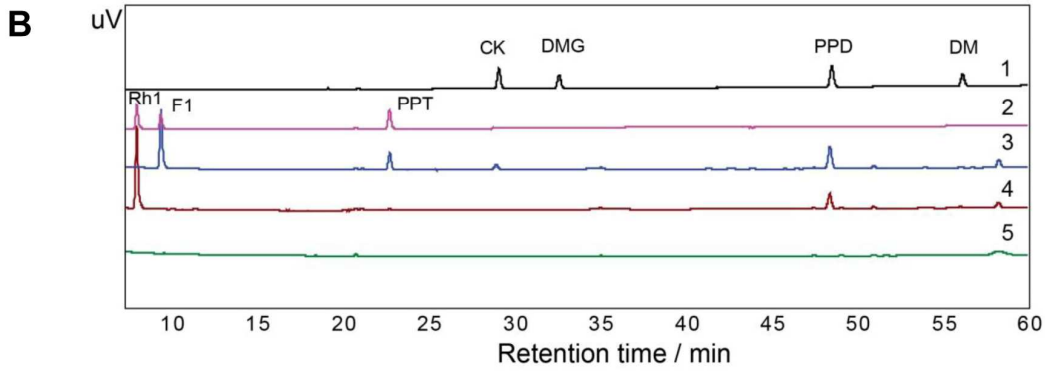







A
UGTPg100
UGTPg1
Chi_1
Chi_2
Chi_3
Chi_4
Chi_5
Chi_6
PSPG
V10 F13 H15 C82 D117 A142 F144 L186 G338
A10 I13 H15 H82 D117 A142 H144 L185 G339
1
109
222
341
475 aa
B
Negative control
UGTPg1
UGTPg100
Chi_1
Chi_2
Chi_3
Chi_4
Chi_5
Chi_6
C
AU
Rg1
Rh1
F1
PPT
Authentic samples
Negative control
UGTPg1
UGTPg100
Chi_1
Chi_2
Chi_3
Chi_4
Chi_5
Chi_6
20
30
40
50
Retention time / min


A
Production (mg/L)
F1
PPT
CK
PPD
DM
DMG
Production (mg/L)
Rh1
PPT
PPD
DM
C
Intens.
x10^6
F1 produced by ZW-F1-17
[2M+Na]+
1299.6
423.2
[M+Na]+
661.3
[M-H2O+H]+
621.3
977.8
Intens.
x10^6
Rh1 produced by ZW-Rh1-20
[M-2H2O+H]+
603.2
423.2
[2M+Na]+
1299.5
m/z


B
uV
CK
DMG
PPD
DM
Rh1
F1
PPT
1
2
3
4
5
Retention time / min