

# SCIENTIFIC REPORTS



OPEN

## Purification and characterization of a novel phloretin-2'-O-glycosyltransferase favoring phloridzin biosynthesis

Received: 28 May 2016  
Accepted: 28 September 2016  
Published: 12 October 2016

Tingjing Zhang<sup>1</sup>, Jianqiang Liang<sup>2</sup>, Panxue Wang<sup>3</sup>, Ying Xu<sup>4</sup>, Yutang Wang<sup>1</sup>, Xinyuan Wei<sup>1</sup> & Mingtao Fan<sup>1</sup>

Phloretin-2'-O-glycosyltransferase (P2'GT) catalyzes the last glycosylation step in the biosynthesis of phloridzin that contributes to the flavor, color and health benefits of apples and processed apple products. In this work, a novel P2'GT of *Malus x domestica* (MdP2'GT) with a specific activity of 46.82  $\mu$ kat/Kg protein toward phloretin and uridine diphosphate glucose (UDPG) at an optimal temperature of 30 °C and pH 8.0 was purified from the engineered *Pichia pastoris* broth to homogeneity by anion exchange chromatography, His-Trap affinity chromatography and gel filtration. The purified MdP2'GT was low N-glycosylated and secreted as a stable dimer with a molecular mass of 70.7 kDa in its native form. Importantly, MdP2'GT also exhibited activity towards quercetin and adenosine diphosphate glucose (ADPG), kaempferol and UDPG, quercetin and UDP-galactose, isoliquiritigenin and UDPG, and luteolin and UDPG, producing only one isoquercitrin, astragalol, hyperoside, isoliquiritin, or cynaroside, respectively. This broad spectrum of activities make MdP2'GT a promising biocatalyst for the industrial preparation of the corresponding polyphenol glycosides, preferably for their subsequent isolation and purification. Besides, MdP2'GT displayed the lowest  $K_m$  and the highest  $k_{cat}/K_m$  for phloretin and UDPG compared to all previously reported P2'GTs, making MdP2'GT favor phloridzin synthesis the most.

Phloretin and its glycosides belong to a subclass of polyphenols within a larger group of plant-based phenolic bioactive natural products. They exhibit a wide variety of beneficial biological activities, such as anti-oxidant<sup>1</sup>, anti-inflammatory<sup>2</sup>, anti-cancer<sup>3</sup> and anti-diabetic activities<sup>4</sup>. Phloretin, a precursor of phloridzin, exhibits additional pharmacologically biological functions like anti-tumor<sup>5</sup> and anti-estrogen activities<sup>6</sup>, and inhibition of cardiovascular disease<sup>3</sup>. However, the use of phloretin as a drug and food additive has been limited due to its weak aqueous solubility, chemical and/or biological instability, and low absorbability<sup>1,3,7,8</sup>. Attachment of sugar moieties to phloretin by glycosylation reactions can increase its aqueous solubility and *in vivo* half life, and typically exerts profound direct or indirect effects on its biological activity<sup>1,3,7,9,10</sup>.

Phloridzin, a phloretin glycoside, is biosynthesized via a glycosylation reaction towards phloretin and UDPG catalyzed by phloretin-2'-O-glycosyltransferases (P2'GTs)<sup>11,12</sup>. Phloridzin is comparable to phloretin in the bioactive properties, but it exhibits greater bioavailability and lower toxicity<sup>1,2,10,13,14</sup> and is thus widely used in diverse commercial products as an alternative to phloretin<sup>7,8</sup>. To date, commercial available phloridzin is extracted from non-renewable sources using traditional solvent-based approaches<sup>15–18</sup>, which restrict its efficient availability. Selective glycosylation of phloretin by P2'GT provides a convenient access to phloridzin through bottom-up synthesis. However, only two P2'GTs, UGT88F1 (accession no. **ACZ44840**) from *M. x domestica* and UGT88F2 (accession no. **ACZ44838**) from *Pyrus communis*, can selectively glycosylate phloretin at the 2'-OH to produce analytical amounts of phloridzin<sup>19,20</sup>. Whereas these two P2'GTs appear to exclusively accept phloretin as a glycosyl acceptor. And, the UGT88F2, orthologous to UGT88F1 with 99% homologous sequence identity, can also

<sup>1</sup>College of Food Science and Engineering, Northwest A&F University, Yangling, Shaanxi, 712100, China. <sup>2</sup>College of Life Sciences, Northwest A&F University, Yangling, Shaanxi, 712100, China. <sup>3</sup>Department of Food Science, University of Massachusetts, Amherst, MA01003, USA. <sup>4</sup>College of Life Science and Engineering, Shaanxi University of Science & Technology, Xi'an, Shaanxi, 710021, China. Correspondence and requests for materials should be addressed to M.F. (email: fanmt@nwsuaf.edu.cn)

deglycosylate the formed phloridzin to phloretin and UDP-glucose again, as increasing the incubation time or the concentration of UDP-glucose<sup>7</sup>. These shortcomings limit the availability and usable scope of UGT88F1 and UGT88F2 in the actual production process. Other reported P2'GTs can accept a number of phenylpropanoid substrates, but glycosylation of phloretin by these plant P2'GTs yields mixtures of mono-glycosides<sup>14,20,21</sup>. A more promiscuous bacterial GT glycosylates phloretin at the 2'-, 4'-, 6'- and 4-OH positions and produces a complex mixture of glycosidic compounds, including two monoglycosides, two diglycosides and one triglycoside<sup>8</sup>. In summary, all reported P2'GTs suffer from a narrow substrate range or low region selectivity, which could reduce economic benefits and complicate the separation and purification of the formed glycosides. Additionally, industrialized societies are moving toward industrial biotechnology, which has proven to be environmentally sound and commercially efficient<sup>22</sup>. This move poses a continuous demand for novel biocatalysts, preferably that can accept a wide range of substrates but do not produce by-products during reactions. Furthermore, P2'GTs belong to the UDP-dependent glycosyltransferases (UGTs) with a supergene family and multiple P2'GTs in apples have been reported to be involved in apple phloridzin biosynthesis<sup>11,12,19–21</sup>, other isoenzymes might be able to glycosylate phloretin.

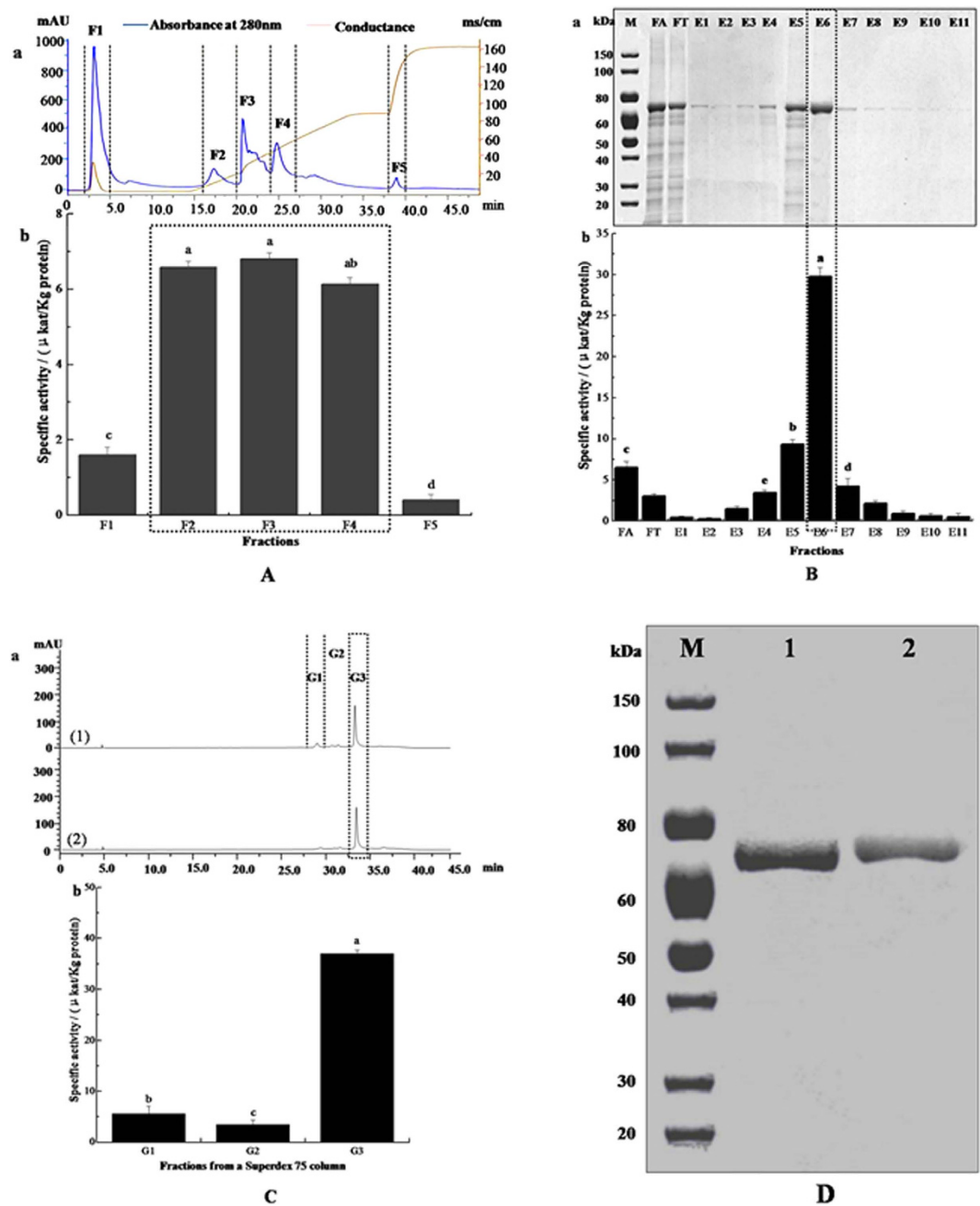
Herein, we report a novel recombinant P2'GT that exhibits predominant region specificity for the sugar attachment site and broad-range acceptance for substrates. This P2'GT was produced by an engineered *P. pastoris* GS115 strain harboring a gene (Accession no. **KT444675**) from *M. x domestica* (*MdP2'GT*). In order to broaden its application range in the actual production process, potential donor substrates and potential acceptor substrates (giving a maximum of 225 combinations) was assayed to evaluate their abilities to act as *MdP2'GT* substrates. The formed glycosides in combinations were sequentially identified by reversed-phase high performance liquid chromatography equipped with a diode array detector (HPLC-DAD), liquid chromatography tandem mass spectrometry (LC-MS/MS), <sup>1</sup>H Nuclear Magnetic Resonance (<sup>1</sup>H NMR), and <sup>13</sup>C NMR. Finally, the binding of preferred substrates and the most potent inhibitor with the enzyme active sites was predicted to elucidate key interactions within the *MdP2'GT*/substrate complex. To the best of our knowledge, it is the first time that reports donor substrate and acceptor substrate combinations specificity of P2'GT. Also, our data indicates that *MdP2'GT* is a promising and preferable biocatalyst for efficient glycosylation in natural product synthesis and also simplifies the subsequent isolation and purification of the glycosylated products.

## Results

**Purification of *MdP2'GT*.** The three-step purified process used for *MdP2'GT* purification is summarized in Fig. 1 and Table 1. The first purification step (DEAE anion-exchange chromatography) resulted in five enzymatically active fractions (F1–F5) that exhibited significant differences ( $P < 0.05$ ) in specific activity (Fig. 1A), which led to an approximately 59-fold purification with 49.85% yield and a specific activity of 6.48  $\mu\text{kat/Kg}$  protein for *MdP2'GT* (Table 1). His-trap affinity chromatography with 80 mM imidazole as the eluent (E6, Fig. 1B) further increased the purification to 301-fold and the specific activity to 33.13  $\mu\text{kat/Kg}$  protein, with 13.13% yield (Table 1). The gel filtration was finally performed to remove the imidazole in E6, resulting in three fractions (G1–G3) (Fig. 1C(1)). G3, which exhibited the highest specific activity, was loaded again onto a G75 Superdex gel filtration column and yielded a single peak in chromatography (Fig. 1C(2)). Accordingly, approximately 110 mg of purified *MdP2'GT* was obtained from 1 L of engineered *P. pastoris* GS115 culture, the specific activity of which increased from an average of 0.11  $\mu\text{kat/Kg}$  protein in the crude broth to 40.45  $\mu\text{kat/Kg}$  protein, representing a purification of 368-fold with 9.02% recovery. The single peak was further analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting and both yielded a single homogeneous band (Fig. 1D), indicating the purity of *MdP2'GT* was more than 95%. However, the size of purified *MdP2'GT* appeared larger than the theoretical value (53.9 kDa) deduced from the amino acid sequence of mature *MdP2'GT* fused to the His<sub>6</sub> tag (Fig. S1).

**Molecular weight evaluation and glycosylation analysis.** According to the size exclusion chromatography (SEC) and the protein standard curve (Fig. 2A,B), the molecular weight (MW) of purified *MdP2'GT* was approximately 70.7 kDa that is much larger than the theoretical value of 53.9 kDa. To explain this discrepancy, the glycosylation pattern was investigated by deglycosylation with Peptide N-glycosidase F (PNGase F) and Endoglycosidase H (Endo H). Asymmetric tailing of the protein peak in SEC (Fig. 2A2) and the broad signal in SDS-PAGE (Fig. 2C1) suggested a heterologous glycosylation pattern in *MdP2'GT*. The recombinant *MdP2'GT* was released as a stable dimer with an approximate molecular mass of 65.3 kDa after deglycosylation with PNGase F or Endo H in its native states (Figs 2A3 and 5B,C2,C5). MALDI-TOF mass spectrometric analysis further confirmed that the *MdP2'GT* signal was decreased from 70 kDa to 65 kDa after deglycosylation under native conditions (Fig. 2D1,2). In contrast, under denaturing conditions, *MdP2'GT* was deglycosylated by PNGase F generating two subunits with MWs of 35.2 kDa and 24.9 kDa (Fig. 2). However, when *MdP2'GT* was treated with Endo H under identical conditions, only a monomeric molecular mass of 24.5 kDa was observed (Fig. 2). Hence, approximately 23.8% of the molecular mass is attributable to N-linked carbohydrates, but approximately 30.4% of the glycosylation is resistant to PNGase F treatment under denaturing conditions. Additionally, MALDI-TOF mass spectrometric analysis revealed that asparagine glycosylation was the dominant site for the addition of mannose-containing oligosaccharide and the candidate sites included Asn77, 95, 132, and 195. Finally, the specific activities of native and natively deglycosylated *MdP2'GT* toward phloretin and UDPG were determined. Interestingly, removal of the *MdP2'GT* glycosylation sites resulted in no significant change in specific activity (Fig. 2C). Thus, the purified *MdP2'GT* in this study could be used directly in practice without additional treatments.

**Biochemical characterization of recombinant *MdP2'GT*.** Purified *MdP2'GT* displayed maximal activity at pH 8.0, and up to 80% of its maximal activity remained within the pH range of 7.0–9.0 (Fig. 3A). Moreover,



**Figure 1. Purification steps of recombinant *MdP2'GT*.** (A) (a) DEAE ion-exchange chromatography fractions and (b) their corresponding specific activity. (B) (a) SDS/PAGE analysis of Ni<sup>2+</sup> affinity chromatography fractions (FA: mixture of F2-F4 from (A); FT: FA flow-through; E1-E11: fractions eluted by buffer C containing 5 mM, 10 mM, 20 mM, 40 mM, 60 mM, 80 mM, 100 mM, 200 mM, 300 mM, 400 mM, and 500 mM imidazole, respectively) and (b) their corresponding specific activity. (C) Further purified of E6 in (B) by G75 Superdex gel filtration (a(1)elution fractions G1-G3 from E6, a(2)elution fraction G3 purified by gel filtration again) and (b) the corresponding specific activity. (D) (1) SDS/PAGE and (2) Western Blot analysis of purified *MdP2'GT*, M: SuperSignal Molecular Weight Protein Ladder.

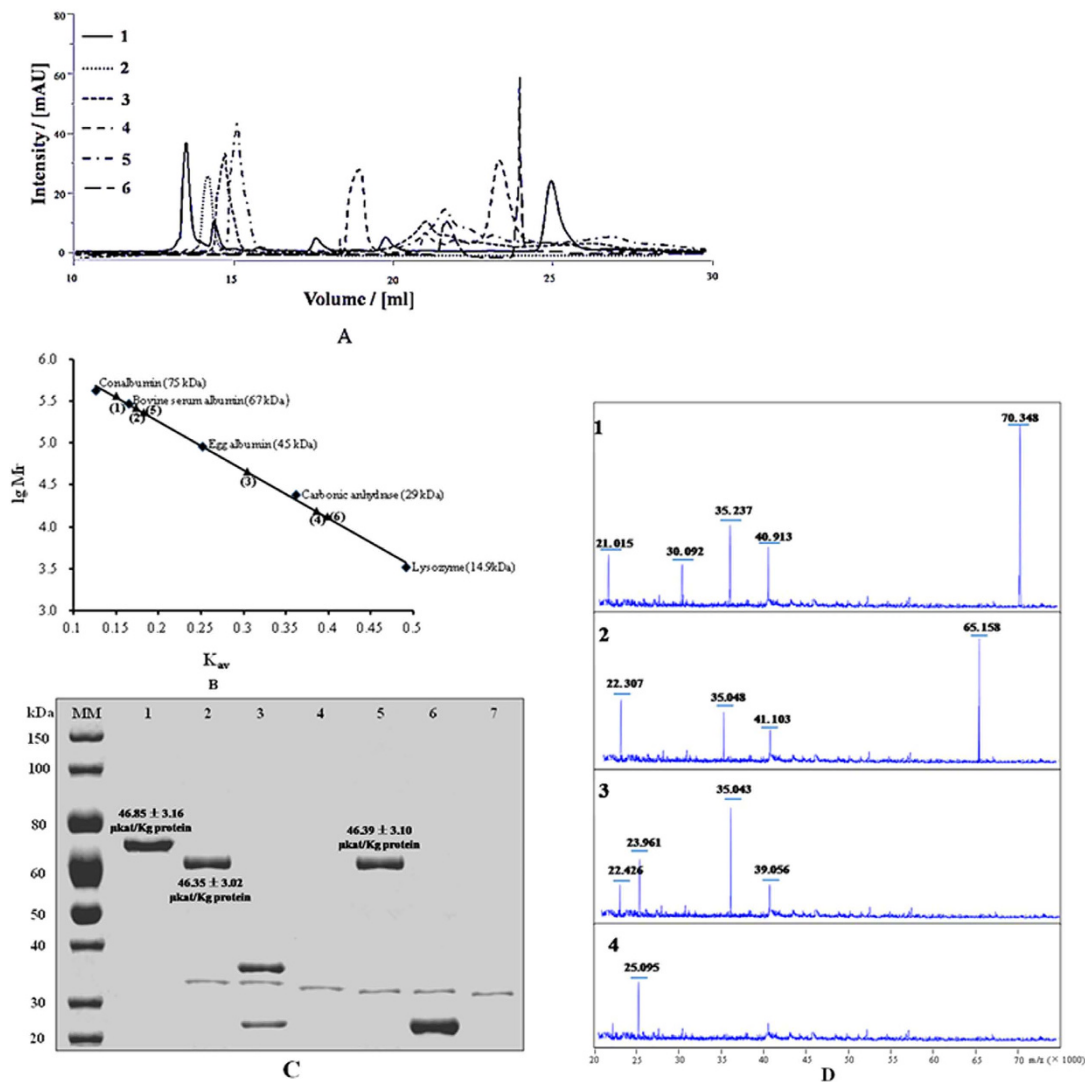
*MdP2'GT* remained stable in the pH range of 7.0–9.0 incubated at 30 °C for 24 h (Fig. 3B). At pH 8.5, the maximal residual activity of *MdP2'GT* was >90% of its original activity, but at pH < 6.5, the residual activity was reduced to <50% (Fig. 3B), indicating that purified *MdP2'GT* preferred alkaline conditions to acidic conditions. Additionally, purified *MdP2'GT* exhibited higher activity and stability in Tris/HCl buffer than in potassium phosphate buffer at pH 7.0 (Fig. 3A,B). Therefore, subsequent biochemical characterizations were performed in Tris/HCl buffer system.

Purified *MdP2'GT* exhibited activity in a broad temperature range from 15 °C to 65 °C, with a maximal activity of 46.18 μkat/Kg protein at 30 °C (Fig. 3C). Its activity was not obviously influenced by incubation at temperatures

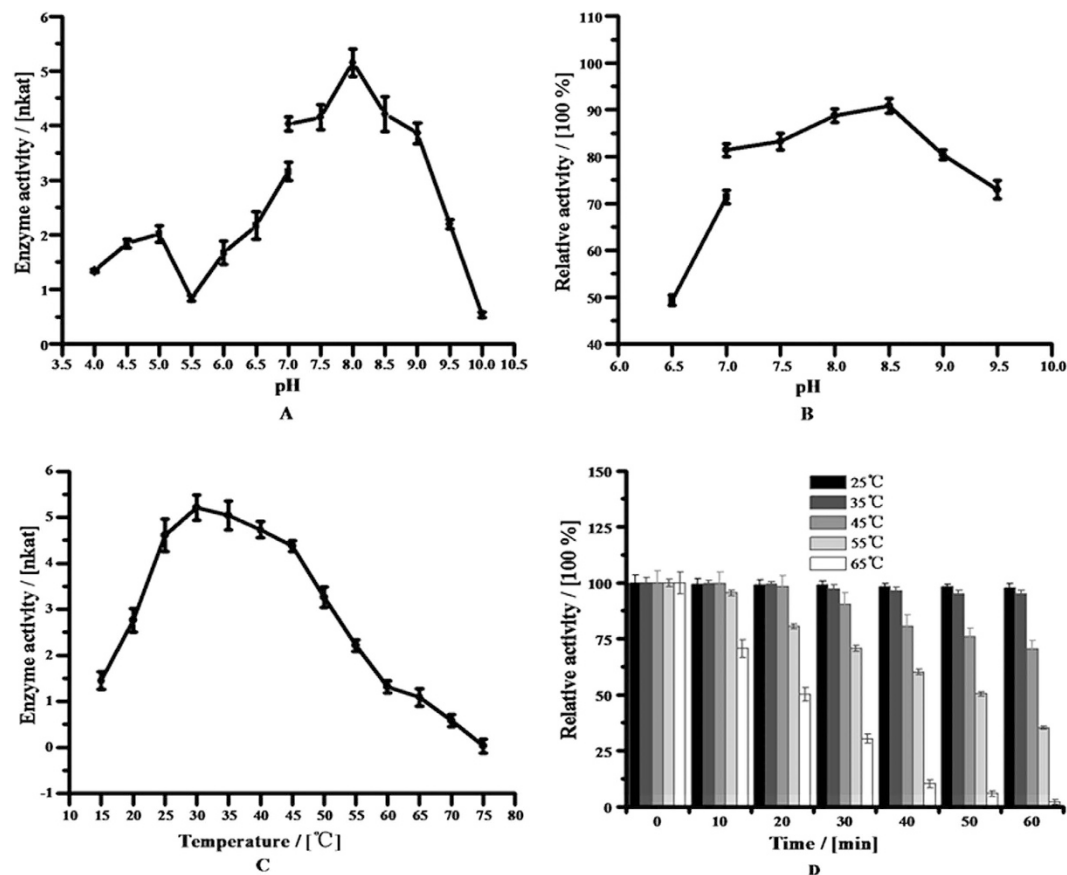
Purification step	Total protein (mg)	Total activity (nkat)	Specific activity ( $\mu\text{kat/Kg protein}$ )	<sup>b</sup> Purification (Fold)	Yield (100%)
Crude	1240.71 $\pm$ 50.46	0.14 $\pm$ 0.01	0.11 $\pm$ 0.01	1	1
<sup>a</sup> DEAE Sepharose Fast Flow	618.49 $\pm$ 31.73	4.01 $\pm$ 0.18	6.48 $\pm$ 0.39	59	49.85
Ni <sup>2+</sup> affinity (80 mM imidazole)	162.85 $\pm$ 18.96	5.40 $\pm$ 0.34	33.13 $\pm$ 1.23	301	13.13
Gel filtration (Superdex 75)	111.89 $\pm$ 8.93	4.53 $\pm$ 0.37	40.45 $\pm$ 2.01	368	9.02

**Table 1. Purification and activity of *MdP2'*GT from the bioengineered *P. pastoris* GS115 supernatant.**

<sup>a</sup>Determined data represents the pooled fractions of F2, F3, and F4 that displayed the highest specific activity in Fig. 1A. <sup>b</sup>Calculated according to the mean of the specific activity and expressed as an integer.



**Figure 2. Molecular weight and glycosylation analysis.** (A) SEC of marker proteins (1, 75 kDa conalbumin, 67 kDa bovine serum albumin, 45 kDa egg albumins, 29 kDa carbonic anhydrase, and 14.9 kDa lysozyme), native *MdP2'*GT (2, calculated MW 70.7 kDa), native (3, calculated MW 65.3 kDa) and denatured (4, calculated MW 35.2 kDa and 24.9 kDa) deglycosylated *MdP2'*GT with PNGase F, and native (5, calculated MW 65 kDa) and denatured (6, calculated MW 24.5 kDa) deglycosylated *MdP2'*GT with Endo H, and (7) 10 U of Endo H. (B) The calibration curve of the marker proteins. (C) SDS-PAGE analysis of (1) native *MdP2'*GT, (2) native and (3) denatured deglycosylated *MdP2'*GT with PNGase F, (4) 1 U PNGase F, (5) native and (6) denatured deglycosylation *MdP2'*GT with Endo H, and (7) 10 U of Endo H. The specific activity of the native and native deglycosylated *MdP2'*GT under optimal conditions are indicated near the respective lanes. (D) MALDI-TOF spectra of (1) the native, (2) the native deglycosylated *MdP2'*GT with PNGase F or Endo H, and denatured deglycosylated *MdP2'*GT with (3) PNGase F or (4) Endo H.



**Figure 3.** Effects of pH and temperature on enzyme activity and stability. (A) Effect and optimum pH. (B) pH stability. (C) Effect and optimum temperature. (D) Temperature stability.

as high as 45 °C for 20 min but decreased at higher temperatures or longer incubation times (Fig. 3D). Incubated at 45 °C for 60 min, the enzyme retained 70% of its original activity, whereas incubated at 55 °C for 60 min decreased its relative enzyme activity to <40%. Under identical conditions at 65 °C, enzyme activity was only 2% of its original activity (Fig. 3D).

As shown in Table 2, enzyme activity was completely inhibited by 5 mM  $\text{Co}^{2+}$  and  $\text{Cu}^{2+}$ , partially inhibited by 5 mM  $\text{Ca}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Al}^{3+}$ , and  $\text{Fe}^{3+}$ , and negligibly affected by 0.5 mM  $\text{K}^{+}$ ,  $\text{Ni}^{2+}$  and  $\text{Ba}^{2+}$ , while 2 mM  $\text{Mg}^{2+}$  and 5 mM  $\text{Fe}^{2+}$  could increase enzyme activity to 130% and 120% of the original activity, respectively. Sodium chloride, used in the purification protocol, had no obvious inhibitory effects even at the concentrations as high as 500 mM. Also, EDTA had no obvious inhibitory effects, suggesting *MdP2'GT* is not a metalloenzyme. However, *MdP2'GT* can be degraded by endogenous *Pichia* proteases and is susceptible to sulfhydryl oxidation, suggesting that adding glycerol, 4-(2-Aminoethyl) benzenesulfonyl fluoride (AEBSE, a protease inhibitor) and  $\beta$ -mercaptoethanol ( $\beta$ -ME) or dithiothreitol (DTT) to the buffers used during purification is essential. No obvious effects of 5% glycerol, 0.5 mM AEBSE, 5 mM  $\beta$ -ME and 5 mM DTT on *MdP2'GT* activity were observed. Due to maximum glycoside product concentrations were restricted by the aqueous solubility of the comparably hydrophobic aglycone acceptors, use of co-solvent conducting the reaction in an aqueous-organic two-phase system can enhance the effective acceptor concentration<sup>7,8</sup>. The co-solvents 20% methanol and 20% dimethyl sulfoxide (DMSO) also had no effects on enzyme activity. Whereas the strong inhibition was observed with 1 mM *p*-hydroxymercuribenzoate (PHMB) and 1 mM *N*-Ethylmaleimide (NEM), indicating the presence of an SH group in the active site of *MdP2'GT*. *MdP2'GT* was also completely inhibited by UDP (by-product), strongly inhibited by the other tested product analogs except for uridine at a final concentration of 5 mM. In addition, Higher  $\text{IC}_{50}$  values indicate lower inhibition, and thus  $\text{Cu}^{2+}$  was the most effective metal ion inhibitor, UDP was the most potent product inhibitor.

**Substrate specificity of *MdP2'GT*.** In this study, a maximum of 225 combinations, resulting from the nine potential glycosyl donor (NDP-sugars) substrates and twenty-five potential glycosyl acceptor (polyphenols) substrates, were evaluated as the substrates for *MdP2'GT*. Only seven combinations were able to act as the enzyme substrates for the production of the corresponding polyphenol glycosides. Phloridzin (phloretin-2'-*O*-glucoside), produced in UDPG/phloretin and ADPG/phloretin combinations, was identified in our previous study (Zhang, *et al.*, unpublished data); its chromatography data and structural formula were shown in Table 3 and Fig. 4, respectively. The glycosides formed in the other five combinations were firstly identified by HPLC-DAD assays, with retention times ( $R_t$ ) of 30.43 (1), 30.54 (2), 28.81 (3), 28.75 (4), and 27.60 min (5), respectively (Table 3). The

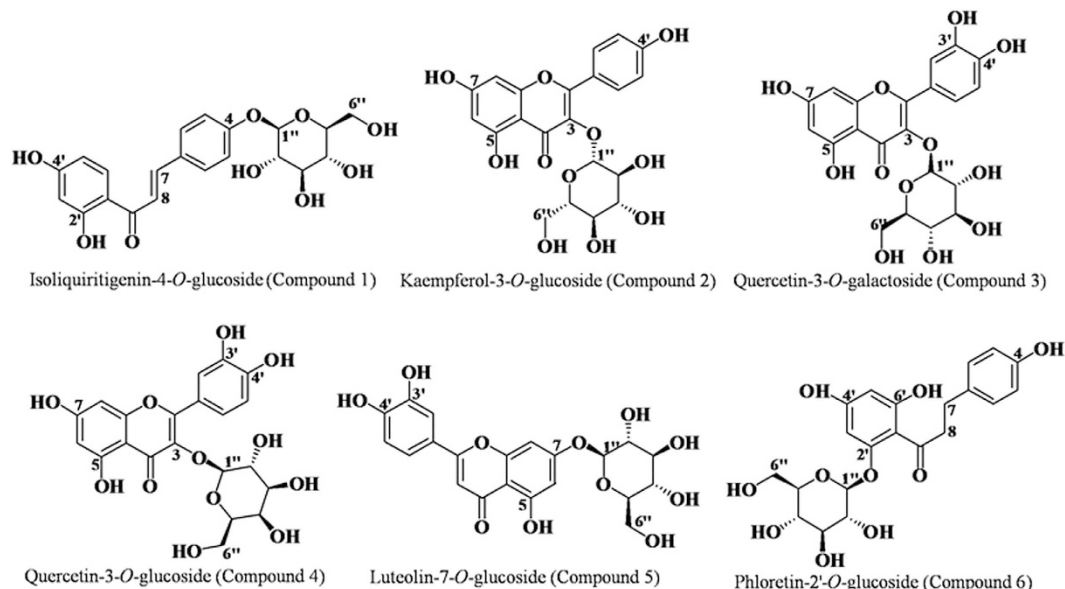
Addition	Concentration/(mM)	Relative activity/(100%)	IC <sub>50</sub> /(mM)
None	—	100	—
NaCl	500	100 ± 1.00	ND
KCl	5	99.58 ± 2.31	ND
MgCl <sub>2</sub>	2	130.35 ± 3.53	Activator
NiCl <sub>2</sub>	5	98.26 ± 2.08	ND
CoCl <sub>2</sub>	5	0.00 ± 0.00	0.05 ± 0.01
CuCl <sub>2</sub>	5	0.00 ± 0.00	0.01 ± 0.01
CaCl <sub>2</sub>	5	54.21 ± 1.05	5.50 ± 0.12
BaCl <sub>2</sub>	5	99.03 ± 2.54	ND
MnSO <sub>4</sub>	5	47.41 ± 1.13	4.50 ± 0.11
FeSO <sub>4</sub>	5	119.43 ± 3.61	Activator
AlCl <sub>3</sub>	5	67.81 ± 1.03	8.00 ± 0.50
FeCl <sub>3</sub>	5	60.78 ± 1.01	6.50 ± 0.23
EDTA	5	99.16 ± 2.21	ND
Glycerol	686	98.13 ± 2.15	ND
AEBSF	0.5	99.51 ± 2.36	ND
β-mercaptoethanol	5	99.02 ± 1.47	ND
Dithiothreitol	5	99.15 ± 1.43	ND
Methanol	1236	99.23 ± 1.41	ND
Dimethyl sulfoxide	704	99.54 ± 1.40	ND
<i>p</i> -hydroxymercuribenzoate	1	18.45 ± 1.02	0.24 ± 0.07
<i>N</i> -Ethylmaleimide	1	20.53 ± 1.08	0.28 ± 0.01
UDP	5	0 ± 0.00	0.10 ± 0.01
UTP	5	3.05 ± 0.26	2.5 ± 0.10
UMP	5	38.25 ± 1.07	4.00 ± 0.13
Uridine	5	98.17 ± 2.12	ND
Trilobatin	5	21.70 ± 1.15	3.00 ± 0.21
3-hydroxyphloridzin	5	15.65 ± 0.58	2.50 ± 0.20

**Table 2. Effects of various reagents on purified *MdP2'*GT activity.** Relative activities were calculated by comparison to a control in which none is added. No effect of addition on *MdP2'*GT activity was defined as relative activities  $\geq 95\%$ . ND: no effects detected at selected concentrations. Activator: additions increased *MdP2'*GT activity at selected concentrations.

HPLC-ESI/MS/MS analysis of each typical peak in negative ion mode for these glycosides were  $m/z^+ = 417.118$  (1), 447.092 (2), 463.087 (3), 463.091 (4), and 447.92 (5) (Table 3), and for their corresponding flavone aglycones were  $m/z^+ = 254.653$  (isoliquiritigenin), 286.081 (kaempferol), 301.932 (quercetin), 301.929 (quercetin), and 284.308 (luteolin), respectively (data not shown), signifying that these glycosides (1–5) were all monoglycosides. Then, the structures of the formed glycosides (1–5) were further elucidated from the <sup>1</sup>H and <sup>13</sup>C NMR spectra. As shown in Table 3, the <sup>1</sup>H and <sup>13</sup>C NMR spectral data of the glycoside 1 was consistent with the reported literature<sup>23</sup> and identified as isoliquiritin (Isoliquiritigenin-4-*O*-glucoside), the glycoside 2 was consistent with the reported literature<sup>24</sup> and identified as astragaline (Kaempferol-3-*O*-glucoside), the glycoside 3 was consistent with the reported literature<sup>25</sup> and identified as hyperoside (Quercetin-3-*O*-galactoside), the glycoside 4 was consistent with the reported literature<sup>24</sup> and identified as isoquercitrin (Quercetin-3-*O*-glucoside), and the glycoside 5 was consistent with the reported literature<sup>26</sup> and identified as cynaroside (Luteolin-7-*O*-glucoside). The structural formulas of these glycosides were provided in Fig. 4.

Chromatographic data demonstrated that the purified *MdP2'*GT exhibited additional activities towards the chalcone (isoliquiritigenin), flavonols (kaempferol and quercetin), and the flavone (luteolin) to produce the corresponding glycosides isoliquiritin, astragaline, hyperoside, isoquercitrin, and cynaroside. Besides, the formed glycosides in seven donor/acceptor combinations were all *O*-glycosidic compounds, indicating *MdP2'*GT is an *O*-glycosyltransferase (OGT). Remarkably, only one hydroxy group on the glycosyl acceptor was glycosylated by *MdP2'*GT, resulting in the production of a single glycoside in each donor/acceptor combination. These unique features suggest *MdP2'*GT a special *P2'*GT that is not only high region specificity regarding its sugar attachment site but also accept a wide range of substrates and do not produce by-products during reactions.

**Kinetic parameters of *MdP2'*GT.** The apparent kinetic parameters of *MdP2'*GT for the seven accepted substrate combinations were determined under optimal conditions to evaluate its substrate preference. As shown in Table 4, *MdP2'*GT exhibited the lowest  $K_m$  for UDP-glucose (7.53  $\mu$ M) and phloretin (0.50  $\mu$ M). The  $k_{cat}$  for UDPG and phloretin were 21.33  $s^{-1}$  and 9.87  $s^{-1}$ , respectively, and were the highest among the seven combinations. Accordingly, the calculated catalytic efficiencies ( $k_{cat}/K_m$ ) were the highest for UDPG (2.83  $s^{-1}\mu M^{-1}$ ) and phloretin (19.73  $s^{-1}\mu M^{-1}$ ). Besides, the highest specific activity (46.82  $\mu$ kat/Kg protein) was also observed in the



**Figure 4.** Structural formulas of the glycosides synthesized by *MdP2'GT*.

UDPG/phloretin combination. Thus, UDPG and phloretin were the most preferred donor/acceptor combination for *MdP2'GT* among the seven donor/acceptor combinations. Compared to reported P2'GTs, *MdP2'GT* exhibited a lower  $K_m$  and a higher  $k_{cat}/K_m$  than that of UGT88F1 in the UDPG/phloretin combination, which is the lowest  $K_m$  value and the highest  $k_{cat}/K_m$  reported (Table 5)<sup>19–21</sup>, demonstrating it is a novel P2'GT and favors phloridzin (2'-O-glycoside) synthesis the most. Besides to the 2'-O-glycoside (phloridzin), *MdP2'GT* also could efficiently produce 3-O-glycosides (isoquercitrin, hyperosidin and astragaline) with  $k_{cat}/K_m$  of  $6.99 \text{ s}^{-1} \cdot \mu\text{M}^{-1}$ ,  $5.34 \text{ s}^{-1} \cdot \mu\text{M}^{-1}$ , and  $2.70 \text{ s}^{-1} \cdot \mu\text{M}^{-1}$  in the ADPG/Quercetin, UDP-Gal/Quercetin, and UDPG/Kaempferol combinations, respectively (Table 4), while it displayed relatively lower  $k_{cat}/K_m$  in the UDPG/Isoliquiritigenin and UDPG/Luteolin combinations to produce small amount of 4-O-glycoside (isoliquiritin) and 7-O-glycoside (Cynaroside) (Table 4). These highly site-specific glycosylation characterizations and divergent product profiles favor the potential use of *MdP2'GT* as a commercial biocatalyst. Additionally, purified *MdP2'GT* exhibited optimal activity at pH values of 7.5 to 8.5 and temperatures from 30 °C to 40 °C depending on the combination of substrates. Strikingly, the specific activity and kinetic parameters of *MdP2'GT* were dramatically influenced by the sugar donor and acceptor combinations, e. g. *MdP2'GT* displayed maximal specific activity and optimal kinetic parameters in the UDPG/phloretin combination but only 10% of its maximal specific activity and poorest kinetic parameters in the ADPG/phloretin combination, while nearly 70% of the maximal specific activity was restored in the ADPG/quercetin combination. This donor-acceptor interaction is beneficial for *MdP2'GT* as a promising biocatalyst because the affinity and turnover rates toward the substrate can be adjusted by modulating the type of NDP-sugars or acceptor substrates.

**Molecular modeling and docking.** Until now, there is no available crystal-based 3D structures for P2'GTs<sup>3,14,19–21</sup>. The homology-based structure for *MdP2'GT* was model using coordinates from the crystal structure of one plant GT (Accession no. **2vg8A**) which shares sequence identity with *MdP2'GT* as high as 34.07%<sup>27</sup>. As expected, *MdP2'GT* consisted of N- and C-terminal domains with a similar  $\alpha/\beta/\alpha$  fold (often referred to as a Rossmann fold). These two domains were connected by the interdomain linker (Fig. 5A). The “putative secondary plant glycosyltransferase (PSPG)” consensus sequence of 44 amino acids found in all Leloir GTs that is thought to be involved in binding of the activated NDP-sugar donor<sup>9,19</sup> was observed in the C-terminal domain (Fig. 5A). We also explored molecular docking studies to elucidate the interactions between *MdP2'GT* and its preferred substrates (UDPG and phloretin) as well as its most potent product inhibitor (UDP). Five residues (W359, A360, S382, E385, and Y399) in PSPG and two residues (Y55 and S290) out of PSPG were predicted to interact with the uridine diphosphate moiety of UDPG in the form of hydrogen bonds, the residues E401, Q402, W380, H15, and T140 were predicted to interact with the 2''-OH, 3''-OH, and 6''-OH on glucose residue of the UDPG in the form of hydrogen bonds (Fig. 5B). For sugar acceptor phloretin, 2'-OH group gave the best fitness score by stabilizing H-bond formation to the sugar donor, 4'- and 4-OH groups were predicted to interact with the residues P11 and Q189 in the form of H-bonds, respectively, the 6'-OH group pointed in the opposite direction from the UDPG (Fig. 5B). It is therefore not surprising that only one 2'-O-glycoside (phloridzin) was synthesized by *MdP2'GT* in UDPG/Phloretin and ADPG/Phloretin combinations. Besides, UDP, one by-product generated during the enzyme reaction, was observed to interact with the same active site residues as that with the uridine diphosphate moiety of UDPG (Fig. 5C). In this case, UDP will compete with UDPG to irreversibly bind to their active sites, leading to the reduction of *MdP2'GT*'s affinity to UDPG.

Donor/acceptor combinations	Glycosides	R <sub>t</sub> (min)	m/z [M-H <sup>+</sup> ]	NMR (500 MHz, DMSO- <i>d</i> <sub>6</sub> )	
				δ <sup>1</sup> H (ppm)	δ <sup>13</sup> C (ppm)
UDPG/Isoliquiritigenin	1	30.43	417.118	8.009 (1H, d, J = 8.8 Hz, H-6'), 7.837 (1H, d, J = 15.5 Hz, H-8), 7.806 (2H, d, J = 8.0 Hz, H-2,6), 7.720 (1H, d, J = 15.3 Hz, H-7), 7.182 (2H, d, J = 8.7 Hz, H-3,5), 6.438 (1H, dd, J = 2.0, 8.5 Hz, H-5'), 6.314 (1H, d, J = 2.0 Hz, H-3'), 5.011 (1H, d, J = 7.2 Hz, H-1''), 4.013 (1H, dd, J = 2.0, 11.8 Hz, H-6''), 3.745 (1H, dd, J = 5.3, 11.9 Hz, H-6'')	191.97 (C-9), 166.19 (C-4'), 165.16 (C-2'), 159.66 (C-4), 143.43 (C-8), 132.10 (C-6'), 130.02 (C-2,6), 129.13 (C-1), 118.65 (C-7), 116.60 (C-3,5), 113.28 (C-1'), 107.85 (C-5'), 102.43 (C-3'), 100.41 (C-1''), 76.86 (C-3''), 76.55 (C-5''), 73.44 (C-2''), 69.90 (C-4''), 61.07 (C-6'')
UDPG/Kaempferol	2	30.54	447.092	7.949 (2H, d, J = 9.0 Hz, H-2', 6'), 6.798 (2H, d, J = 8.5 Hz, H-3', 5'), 6.295 (1H, d, J = 2.0 Hz, H-8), 6.110 (1H, d, J = 2.0 Hz, H-6), 5.113 (1H, d, J = 7.0 Hz, H-1''), 3.088-3.573 (6H, m, sugar protons)	178.15 (C-4), 164.62 (C-7), 161.67 (C-5), 160.15 (C-4'), 157.81 (C-9), 157.14 (C-2), 134.18 (C-3), 130.87 (C-2', 6'), 121.49 (C-1'), 114.71 (C-3', 5'), 104.40 (C-10), 102.97 (C-1''), 98.58 (C-6), 93.43 (C-8), 76.97 (C-5''), 76.72 (C-3''), 74.38 (C-2''), 70.09 (C-4''), 61.37 (C-6'')
ADPG/Quercetin	3	28.81	463.087	7.502 (1H, dd, J = 7.5, 2.5 Hz, H-6'), 7.497 (1H, d, J = 2.4 Hz, H-2), 6.774 (1H, d, J = 7.0 Hz, H-5'), 6.326 (1H, d, J = 2.1 Hz, H-8), 6.120 (1H, d, J = 2.0 Hz, H-6), 5.194 (1H, d, J = 7.4 Hz, H-1'')	177.97 (C-4), 164.46 (C-7), 161.49 (C-5), 157.09 (C-2), 156.89 (C-9), 148.46 (C-4'), 144.46 (C-3'), 134.15 (C-3), 121.60 (C-6'), 121.46 (C-1'), 116.35 (C-5'), 114.88 (C-2), 104.20 (C-10), 103.40 (C-1''), 98.51 (C-6), 93.38 (C-8), 75.80 (C-5''), 73.55 (C-3''), 71.64 (C-2''), 68.48 (C-4''), 60.47 (C-6'')
UDP-Gal/Quercetin	4	28.75	463.091	7.491 (2H, dd, J = 7.8, 2.0 Hz, H-2', 6'), 6.773 (1H, d, J = 7.8 Hz, H-5'), 6.278 (1H, d, J = 2.0 Hz, H-8), 6.091 (1H, d, J = 2.0 Hz, H-6), 5.271 (1H, d, J = 7.3 Hz, H-1''), 3.706-3.285 (6H, m, sugar protons)	177.61 (C-4), 164.17 (C-7), 161.17 (C-5), 157.13 (C-2), 156.58 (C-9), 147.98 (C-4'), 144.03 (C-3'), 133.75 (C-3), 121.34 (C-6'), 121.20 (C-1'), 115.71 (C-5'), 114.13 (C-2), 103.81 (C-10), 102.46 (C-1''), 98.03 (C-6), 92.85 (C-8), 76.52 (C-5''), 76.25 (C-3''), 73.87 (C-2''), 69.34 (C-4''), 60.68 (C-6'')
UDPG/Luteolin	5	27.60	447.092	7.41 (1H, d, J = 2.4 Hz, H-2'), 7.35 (1H, dd, J = 8.4, 2.4 Hz, H-6'), 6.89 (1H, d, J = 8.4 Hz, H-5'), 6.77 (1H, d, J = 2.4 Hz, H-8), 6.75 (1H, s, H-3), 6.43 (1H, d, J = 2.4 Hz, H-6), 5.07 (1H, d, J = 7.2 Hz, H-1'')	181.68 (C-4), 164.34 (C-2), 162.60 (C-7), 160.72 (C-5), 156.75 (C-9), 149.10 (C-4'), 145.05 (C-3'), 121.37 (C-1'), 118.53 (C-6'), 115.07 (C-2'), 112.50 (C-5'), 105.01 (C-10), 102.47 (C-3), 99.59 (C-6), 99.01 (C-1''), 94.10 (C-8), 76.47 (C-3''), 75.81 (C-5''), 72.68 (C-2''), 69.14 (C-4''), 60.29 (C-6'')
<sup>a</sup> UDPG/Phloretin <sup>a</sup> ADPG/Phloretin	6	28.04	437.256	2.79 (2H, t, J = 7.4 Hz, H-7), 3.03 (2H, dd, J = 11.4 Hz, H-2'', 4''), 3.11 (1H, m, J = 5.1 Hz, H-5''), 3.20 (1H, m, J = 4.7 Hz, H-3''), 3.38 (2H, m, J = 8.3 Hz, H-8), 3.62 (2H, dd, J = 5.4, 11.8 Hz, H-6''), 4.94 (1H, d, J = 6.8 Hz, H-1''), 5.86 (2H, d, J = 1.9 Hz, H-3, 5), 6.08 (1H, d, J = 1.8 Hz, H-3'), 6.59 (2H, d, J = 8.1 Hz, H-2, 6), 6.95 (1H, d, J = 8.4 Hz, H-5')	29.43 (C-7), 45.56 (C-8), 61.02 (C-6'), 69.68 (C-4''), 73.30 (C-2''), 77.00 (C-5''), 77.08 (C-3''), 94.05 (C-5'), 96.96 (C-3'), 100.66 (C-1''), 105.37 (C-1'), 114.68 (C-2 and C-6), 128.99 (C-3 and C-5), 132.48 (C-1), 154.97 (C-4), 160.91 (C-4'), 164.58 (C-6'), 166.16 (C-2'), 205.14 (C-9)

**Table 3. Analytical data of formed glycosides in the other five donor/acceptor combinations with purified *MdP2'*/GT.** <sup>a</sup>The formed glycoside in UDPG/Phloretin and ADPG/Phloretin combinations was determined in our previous unpublished study and identified as phloridzin.

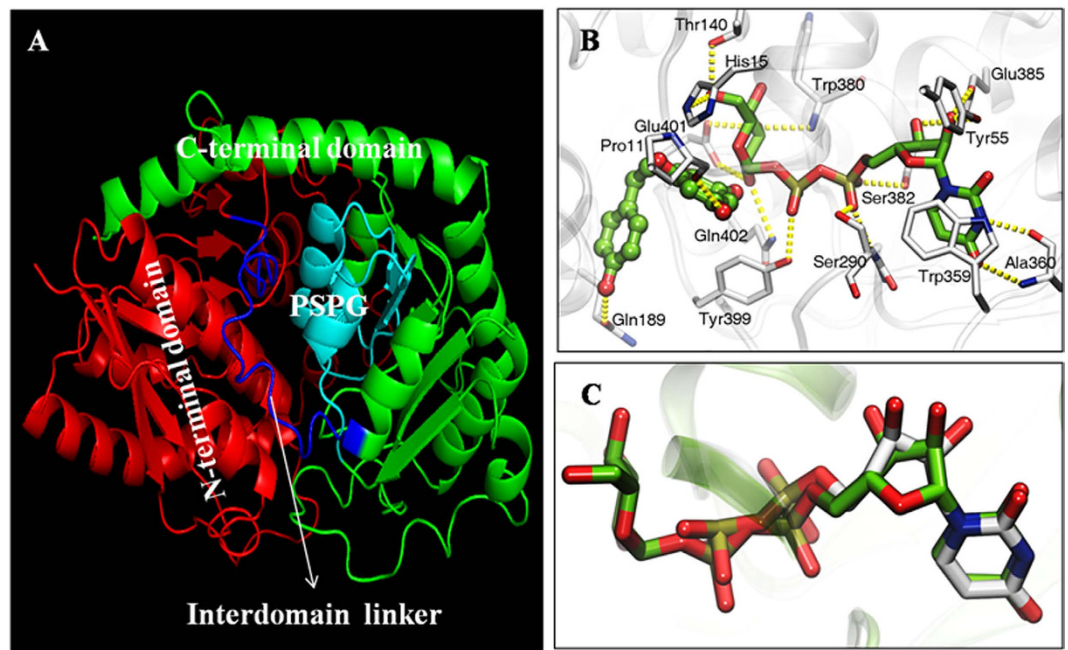
## Discussion

The predominant phenolic compound phloridzin (phloretin 2'-*O*-glucoside) in *Malus* species has various beneficial biological and pharmacological activities<sup>3,11,12,19–21</sup>. Compelling evidence has suggested that P2'/GT is a key enzyme catalysing the rate-limiting step in the phlorizin biosynthetic pathway<sup>12,21</sup>. Recently, it has reported seven plant P2'/GTs (UGT88F1, UGT88F2, UGTA15 (Accession no. **AAZ80472**), UGT71K1 (Accession no. **ACZ44835**), UGT71K2 (Accession no. **ACZ44837**), UGTA16 (Accession no. **ACZ44836**), and DicGT4 (Accession no. **BAD5200**)<sup>14,19,20</sup> and one bacterial P2'/GT (YijC, Accession no. **AAU40842**)<sup>3</sup>, but they suffer from narrow substrate range or low region selectivity. Besides, these reported P2'/GTs from *Malus* display low sequence identities with each other (34.1–46.3%)<sup>20</sup>. Therefore, the presence of possible further phloretin glycosylating enzymes with novel characterizations cannot be excluded. In this work, a novel P2'/GT (*MdP2'*/GT) with highly region specificity regarding the sugar attachment site and relatively broader substrate acceptance, was purified from the generally recognized as safe (GRAS) organism engineered *P. pastoris* GS115 and then characterized.

DEAE-anion-exchange chromatography, the first use in P2'/GT purification to our knowledge, effectively removed more than 50% of impurities and increased enzyme specific activity to approximately 60% via the effective adsorption, demonstrating it was a powerful tool for *MdP2'*/GT purification. Besides, DEAE Sepharose is a weak anion-exchange resin showing much higher affinity and capacity for the negatively charged bio-molecules than that for the positive or neutral charged ones<sup>28</sup>, signifying *MdP2'*/GT was a negatively charged P2'/GT, in good agreement with the deduced pI of 5.8<sup>19,20</sup>. However, the molecular mass of the purified *MdP2'*/GT was a little larger than the expected MW due to the low glycosylation pattern. Similar results have been observed in other GTs<sup>29,30</sup>. They reported that foreign enzyme expressed in *P. pastoris* can be hyper- or low- glycosylated when directed to secretion. This hyper- or low- glycosylation pattern can increase the MW and may affect the recombinant enzyme activity if its native structure does not contain the sugars. Fortunately, the catalytic activity of purified *MdP2'*/GT was not affected by the low-glycosylated pattern, which might be explained by the glycosylation sites that positioned far away from the predicted active sites in space (Fig. 5B). Instead, the low N-glycosylated pattern is favorable because it frequently increases the half-lives of foreign proteins expressed in *P. pastoris*<sup>30,31</sup>. In addition, the simple deglycosylation treatments in the present study might facilitate the determination of favorable conditions for advanced structural research on *MdP2'*/GT, as the flexibility and heterogeneity of carbohydrate moieties potentially affect the crystallization<sup>32,33</sup>.

Our data showed that purified *MdP2'*/GT preferred alkaline conditions, suggesting the amino acids ionized at alkaline pH might be present in the catalytic site<sup>34</sup>. Similar behavior was observed in UGT88F1, UGT88F2, and YijC, but different from UGT71A15, UGT71K1, UGT71A16, UGT71K2, and DicGT4, which have the optimal pH values at 7.7, 8.0, 8.0, 6.75, 6.75, 6.75, 6.25, and 6.0, respectively (Table 5)<sup>3,14,19,20</sup>. Moreover, the optimal and stable temperatures of purified *MdP2'*/GT dramatically differed from those for reported P2'/GTs (UGT88F1, UGT88F2, UGT71K1, and UGT71K2), which have optimum temperatures of 25 °C, 45 °C, 40 °C, and 25 °C, respectively





**Figure 5. Structural analysis of *MdP2'GT*.** (A) Ribbon diagram of the *MdP2'GT* homology-based structure with the N-terminal domain (red), C-terminal domain (green), interdomain linker region (blue), and the PSPG motif (cyan). (B) Overlay of the UDPG/phloretin complex with *MdP2'GT*. (C) Overlay of the inhibitor (UDP) with *MdP2'GT*.

(Table 5)<sup>19,20</sup>, and exhibit thermo-stability at temperatures as high as 80 °C<sup>19,20</sup>. These differences might result from the glycosylation in the amino acid sequences, the binding ability of the substrates to the enzyme active sites, the backbone structure determining the overall shape of the acceptor pocket, or the isoenzyme form and the amino acid residues under ambient conditions<sup>9,31,35</sup>. Other factors, such as expression vector and host strain, organisms, apple cultivars, genotype, as well as the edaphic and environmental (temperature, salinity, water stress and light intensity) conditions<sup>1,3,14,20</sup>, may also affect these properties. However, the observed enzyme inhibition by Co<sup>2+</sup> and Cu<sup>2+</sup> ions may be due not only to their effects on *MdP2'GT* itself but also the destruction of phloretin, since the Co<sup>2+</sup>, Cu<sup>2+</sup> and Hg<sup>2+</sup> have been reported to damage polyphenol anthocyanins that has properties similar to those of phloretin<sup>36,37</sup>. Notably, similar GT inhibition by UDP, just 5 mM was sufficient to completely inactivate *MdP2'GT*, has been reported previously and addressed by using the glycosyltransferase-catalyzed cascade reactions that utilize UDP dependent conversion of sucrose to regenerate the UDP-glucose donor with sucrose synthase<sup>7,8,37</sup>. The sucrose synthase reaction not only served to overcome the UDP inhibitory effect but also to solve the problem of cost-effective supply of the NDP-sugar substrate.

P2'GTs are in complexes with substrates and their corresponding products, such as UGT88F1 and UGT88F2 with phloretin as substrate and strictly region-specific for the position 2' to produce phloridzin, other reported P2'GTs with phloretin as substrate producing a mixture of mono-glycosides, di-glycosides and tri-glycosides<sup>3,14,19–21</sup>. In addition to phloretin, UGT1A15, UGT71K1, and UGT71K2 also accept isoliquiritigenin, kaempferol, and quercetin as substrates<sup>20,21</sup>, and DicGT4 accept kaempferol and quercetin as substrates<sup>9</sup>. Another polyphenol luteolin can be utilized by UGT71A15 and UGT71A16 as well<sup>20</sup>. Besides to these polyphenols, the reported P2'GTs (except UGT88F1) also accept other compounds as substrates (Table 5)<sup>14,20,38</sup>. In present study, *MdP2'GT* could transfer the glucose moiety from the activated UDPG to phloretin specifically at the position 2'-OH to form phloridzin, which was similar to UGT88F1 and UGT88F2 but different from UGT71A15, UGT71K1, UGT71K2, UGT71A16, YijC, and DicGT4. Similar to other reported P2'GTs (except UGT88F1 and UGT88F2), *MdP2'GT* exhibited additional high activity towards isoliquiritigenin, kaempferol, quercetin, and luteolin, the formed corresponding glycosides, however, have remarkable and significant differences. E. g. phloretin and isoliquiritigenin only differ by a single C7-C8 bond and a C7-C8 double bond respectively on the flexible open-chain three-carbon linker connecting the two aromatic rings (Fig. 6)<sup>12,23</sup>. Due to this high structural similarity, 2'-OH on isoliquiritigenin should have been glycosylated in the UDPG/isoliquiritigenin combination, whereas only 4-O-glucoside (isoliquiritin) was detected (Fig. 4). It might be attributable to that the amount of 2'-O-glucoside was too low to be detected, but in scaled up reactions, it was still not detected. We speculate the presence of a C7-C8 double bond on the flexible open-chain three-carbon linker might create a steric hindrance on the 2'-OH and made the 4-OH amenable to be glycosylated based on the similar finding in previous studies<sup>9,10,39,40</sup>. Besides, isoquercitrin and hyperoside, which are native or accumulated as their aglycones in apples<sup>41,42</sup>, could be synthesized by *MdP2'GT* but couldn't be reported P2'GTs. Reported P2'GTs such as UGT1A15, UGT71K1, UGT71K2, and DicGT4 can accept quercetin as substrate, but the sugar donor must be UDPG, and the formed glycosides were quercetin-3-O-glucoside (quercitrin) and 7-O-glucoside not isoquercitrin<sup>14,20</sup>. Furthermore, *MdP2'GT*'s affinities to UDPG and phloretin, the substrate for phloridzin synthesis<sup>12,21</sup>, are also relatively high compared to

Donor/acceptor combinations	Acceptor class	Products	Donors			Acceptors			Specific activity ( $\mu\text{kat}/\text{Kg protein}$ )	pH optimum	Temperature optimum ( $^{\circ}\text{C}$ )
			$K_m$ ( $\mu\text{M}$ )	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$k_{\text{cat}}/K_m$ ( $\text{s}^{-1} \cdot \mu\text{M}^{-1}$ )	$K_m$ ( $\mu\text{M}$ )	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$k_{\text{cat}}/K_m$ ( $\text{s}^{-1} \cdot \mu\text{M}^{-1}$ )			
UDPG/Phloretin	Dihydrochalcone	Phloridzin	$7.53 \pm 1.35$	$21.33 \pm 0.68$	2.83	$0.50 \pm 0.02$	$9.87 \pm 0.89$	19.73	$46.82 \pm 3.45$	8.0	30
ADPG/Phloretin	Dihydrochalcone	Phloridzin	$58.67 \pm 5.63$	$4.65 \pm 0.14$	0.08	$4.09 \pm 0.14$	$2.95 \pm 0.10$	0.72	$4.90 \pm 0.98$	8.0	30
UDPG/Isoliquiritigenin	Chalcone	Isoliquiritin	$36.19 \pm 4.09$	$8.68 \pm 0.26$	0.24	$2.40 \pm 0.08$	$4.35 \pm 0.15$	1.81	$10.37 \pm 1.56$	7.5	40
UDPG/Kaempferol	Flavonol	Astragaline	$30.43 \pm 3.11$	$10.83 \pm 0.47$	0.36	$2.09 \pm 0.12$	$5.65 \pm 0.65$	2.70	$17.13 \pm 1.43$	8.5	35
ADPG/Quercetin	Flavonol	Isoquercitrin	$14.57 \pm 2.45$	$15.22 \pm 0.56$	1.04	$1.06 \pm 0.08$	$7.41 \pm 0.70$	6.99	$32.18 \pm 2.95$	8.5	35
UDP-Gal/Quercetin	Flavonol	Hyperoside	$26.19 \pm 3.23$	$13.18 \pm 0.52$	0.50	$1.81 \pm 0.10$	$6.18 \pm 0.23$	5.24	$23.13 \pm 2.03$	8.0	35
UDPG/Luteolin	Flavone	Cynaroside	$42.59 \pm 4.81$	$5.95 \pm 0.09$	0.14	$2.89 \pm 0.18$	$3.24 \pm 0.18$	1.12	$7.80 \pm 1.13$	7.5	40

**Table 4. Enzyme characters and kinetic parameters of purified *MdP2'GT*.**

Enzyme	$^a K_m$ ( $\mu\text{M}$ )	$^a k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$^a k_{\text{cat}}/K_m$ ( $\text{s}^{-1} \cdot \mu\text{M}^{-1}$ )	$^a V_{\text{max}}$ (mmol/s. kg protein)	$^a$ Optimal pH	$^a$ Optimal temperature ( $^{\circ}\text{C}$ )	Substrate specificity (Donor/acceptor combinations)	Organisms	References
<i>MdP2'GT</i>	0.50	9.87	19.74	167.21	8.0	30	UDPG/Phloretin, ADPG/Phloretin, UDPG/Isoliquiritigenin, UDPG/Kaempferol, ADPG/Quercetin, UDP-Gal/Quercetin, and UDPG/Luteolin	<i>M. x domestica cv. Golden delicious</i>	This study
UGT88F1	0.62	9.72 <sup>c</sup>	15.68	164.75 <sup>b</sup>	7.7	25	UDPG/Phloretin, UDP-galactose/Phloretin	<i>M. x domestica cv. Royal Gala</i>	19
UGT88F2	—	—	—	—	8.0	45	UDPG/Phloretin, UDP/Phloridzin	<i>P. communis cv. Abbé Fetel</i>	7, 8, 20
UGT71A15	82	0.42 <sup>d</sup>	0.005	$8 \times 10^{-3}$	6.75	30	UDPG/Phloretin, UDPG/Naringenin chalcone, UDPG/Eriodictyol chalcone, UDPG/Isoliquiritigenin, UDPG/Butein, UDPG/Naringenin, UDPG/Eriodictyol, UDPG/Luteolin, UDPG/Kaempferol, and UDPG/Quercetin	<i>M. x domestica cv. Rebella</i>	20, 21
UGT71K1	—	—	—	—	6.75	40	UDPG/Phloretin, UDPG/Kaempferol, UDPG/Quercetin, UDPG/Isoliquiritigenin, UDPG/Butein	<i>M. x domestica cv. Rebella</i>	20
UGT71K2	—	—	—	—	6.25	25	UDPG/Phloretin, UDPG/Kaempferol, UDPG/Quercetin, UDPG/Isoliquiritigenin, UDPG/Butein	<i>P. communis cv. Abbé Fetel</i>	20
UGT71A16	—	—	—	—	6.75	30	UDPG/Phloretin, UDPG/Eriodictyol, UDPG/Apiogenin, UDPG/Luteolin, and UDPG/Caffeic acid	<i>P. communis cv. Abbé Fetel</i>	20
<sup>c</sup> YjiC	—	—	—	—	8.0	30	UDPG/Phloretin, UDPG/geldanamycin analogs	<i>Bacillus licheniformis</i>	3, 38
DicGT4	—	—	—	—	6.0	30	UDPG/Phloretin, UDPG/Naringenin, UDPG/Kaempferol, UDPG/Quercetin, UDPG/apigenin, and UDPG/cyanidin	<i>Dianthus caryophyllus</i>	14

**Table 5. Enzyme characters and Kinetic parameters of *MdP2'GT* and reported *P2'GTs*.** <sup>a</sup>Data was determined for the glycosylation of 2'-OH group on phloretin using the UDPG/Phloretin combination as substrate. <sup>b</sup>Calculated according to the turnover rate of  $9.72 \times 10^{-4} \text{ mol.s}^{-1} \cdot \text{mol}^{-1}$  and MW = 59 kDa. <sup>c</sup>Calculated according to  $K_m$  and MW = 59 kDa. <sup>d</sup>Calculated according to  $K_m$  and MW = 52 kDa. <sup>e</sup>Geldanamycin analogs act as the acceptors of YjiC included 18-dehydroxy-17-demethoxyreblastatin, 18-dehydroxy-17-O-demethylreblastatin, 18-dehydroxy-17-O-demethyl-4,5-dehydroreblastatin, 17-demethoxy-reblastatin, and 7-demethoxy-15-hydroxyreblastatin.

reported *P2'GTs*. The  $K_m$  of *MdP2'GT* at optimal pH 8.0 and temperature of 30  $^{\circ}\text{C}$  are 0.61% to 80.65% of the  $K_m$  of reported *P2'GTs* from other apple cultivars or organisms (Table 5)<sup>19,21</sup>. The  $k_{\text{cat}}/K_m$  of *MdP2'GT* are 1.25-fold to three orders of magnitude to reported *P2'GTs* for the phloridzin synthesis (Table 5)<sup>19,21</sup>. Therefore, *MdP2'GT* shows the best enzymatic efficiency and favors phloridzin synthesis the most. Also *MdP2'GT* shows extraordinary catalysis properties for phloridzin synthesis. 370 mg/L phloridzin (850  $\mu\text{M}$ ) was obtained from 1.0 mM phloretin with purified *MdP2'GT* as catalyst, which was the highest among all reported phloridzin yield of  $1.29 \times 10^{-2} \text{ mg/L}^{19}$ , 310 mg/L<sup>7</sup>,  $1.18 \times 10^{-4} \text{ mg/L}^{21}$ , 5.27 mg/L<sup>3</sup>, 13.2 mg/L<sup>14</sup>. These significant differences between *MdP2'GT* and reported *P2'GTs* couldn't be solely attributable to their primary sequence identity (13.2–92.6%, Table 6), which can be illustrated by the two reported plant *P2'GTs* named UGT88F1 and UGT88F2. These two *P2'GTs* both belong to the UGT 88family and share 99.0% amino acid sequence identity (Table 6)<sup>20</sup> but the UGT88F2 displays higher optimal pH, temperature, and phloridzin yield as well as exhibits additional activity to hydrolyze phloridzin to phloretin (Table 5)<sup>7,8,20</sup>, whereas the UGT88F1 exclusively glycosylates 2'-OH on phloretin<sup>19</sup>. The two other UGTs designated UGT73A5 and UGT71F2 from *Dorotheanthus bellidiformis* provide an example of the contrary. These two UGTs show ~20% amino acid sequence identity but glucosylate the same set of acceptor molecules forming the same products<sup>9,40</sup>. Another two plant *P2'GTs* designated UGT71K1 and UGT71K2 share

93.1% amino acid sequence identity (Table 6) but they also use the same compounds as sugar donors and acceptors and produce the same glycosides<sup>20</sup>.

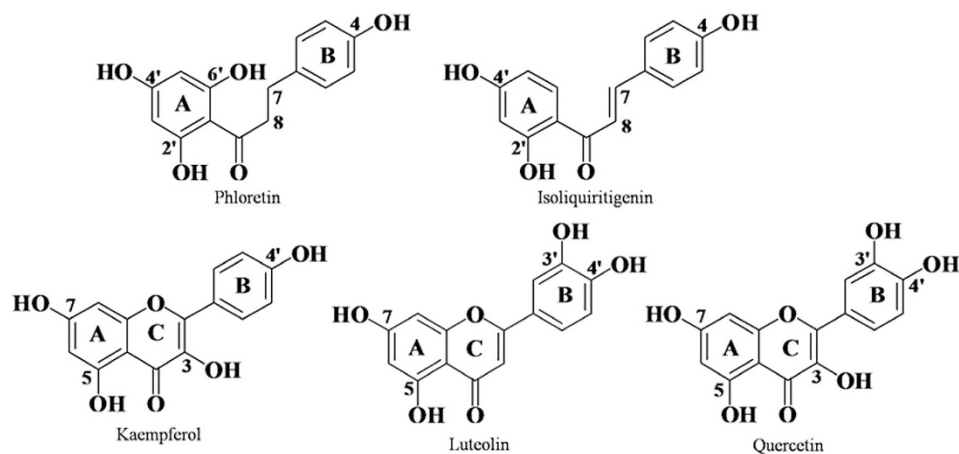
It has reported that the highly conserved secondary and tertiary structures are the key factors that influence the substrate specificity and  $k_{\text{cat}}/K_m$  values for GTs<sup>9,10,27,43</sup>. In structural analysis, *MdP2'*GT features an  $\alpha/\beta/\alpha/\beta$  structure (Fig. 5A) while the reported P2'GTs adopt a  $\beta/\alpha/\beta$  fold in PSPG motif<sup>14,19,20</sup>. This extra  $\alpha$ -helix structure can create a more favorable hydrophobic environment to offer stabilizing interactions between substrates and *MdP2'*GT than reported P2'GTs. Moreover, the glycine residues in PSPG motif (Fig. S1) provide a small size of the side chains, which can confer *MdP2'*GT enough space availability to the substrates for binding the active packet. These unique structural features might not only give *MdP2'*GT a lower  $K_m$  toward UDPG than the reported P2'GTs but might also offer it additional activities to the other UDP-sugar donor UDP-Gal. The interdomain linker is another crucial factor determining the substrates binding due to the highly flexibility with respect to length and sequence for UGTs<sup>9,10,27,40,44</sup>. Both linker spans and sequences of *MdP2'*GT were the same to UGT88F1 and UGT88F2 but different from other P2'GTs<sup>3,14,19,20</sup>, and *MdP2'*GT shared high sequence identity with UGT88F1 (92.6%) and UGT88F2 (92.1%) but low with other P2'GTs (13.2–33.6%, Table 6), which might result in that ADPG could be utilized by *MdP2'*GT, UGT88F1 and UGT88F2 as another sugar donor but couldn't by the other P2'GTs. Changes in the PSPG motif and the interdomain linker region affecting the affinity and types of sugar donor to the enzyme binding sites are also proposed for other crystallized GTs<sup>9,10,39,40</sup>. The highly divergent residues in the acceptor pocket confer GTs large differences in their individual range of acceptors<sup>9,10,27,40</sup>. For *MdP2'*GT, the functional 2'-OH group on phloretin can be positioned near to the His residue (H15) acting as the general base facilitating deprotonation of the acceptor while reported P2'GTs such as YjiC, UGT71K1 and UGT71K2 carry a Gly residue, UGTA15 and UGTA16 carry a Ile residue, and DicT4 carries a large phenylalanine residue at this position. These residues pointing into the acceptor binding pocket may hamper deprotonation of the acceptor and therefore affecting the catalytic efficiency ( $k_{\text{cat}}/K_m$ ). Similar findings were observed in other previous reports<sup>9,10</sup>. Besides, inter- and intradomain interactions, confer stabilization in the form of S-S bridges, salt bridges and H-bond formation to the secondary and tertiary structure<sup>9,40</sup>, are important for activity and specificity as well. Furthermore, the differences in  $k_{\text{cat}}/K_m$  between *MdP2'*GT and reported P2'GTs may also be related to the enzyme conformation stabilized by the glycosylation procedure<sup>33</sup>. The usual binding of the enzyme to the substrate will not be imbedded or prevented by ambience factors after glycosylation, which is commonly reflected by an increase in  $k_{\text{cat}}/K_m$ . Similar results were observed in previous reports<sup>31,35</sup> but the opposite results exist as well<sup>29,33</sup>. These opposite researches reported that the glycosylation of the enzymes is usually accompanied by a reduction in the substrate affinity due to the steric crowding caused by the presence of carbohydrate molecules near the substrate binding site. This discrepancy need further investigated.

Among the accepted poplyphenols, *MdP2'*GT preferred phloretin the most, followed by the kaempferol and isoliquiritigenin, and the worst one is luteolin when used UDPG as sugar donor (Table 4). Compared to phloretin, kaempferol and luteolin contain an extra C ring (Fig. 6) that may affect their binding affinity. The large aromatic residue Tyr399 in the predicted acceptor binding pocket (Fig. 5B) could create a steric hindrance to the extra C ring and locally affect the binding affinity. In contrast, there is no steric hindrance to the flexible open-chain three-carbon linker, making the functional 2-OH 2'-OH group on phloretin more amenable to be glycosylated. However, *MdP2'*GT exhibits a lower  $k_{\text{cat}}/K_m$  toward isoliquiritigenin than kaempferol, although isoliquiritigenin has similar structure to phloretin with a flexible open-chain three-carbon linker. This discrepancy might be related to the presence of a C7-C8 double bond, lacking OH group on 6' position of the A ring, and the absence of isoliquiritigenin in apples<sup>41,42</sup>. Luteolin lacks one OH group at the C3 position of the C ring but has one extra OH group at the C3' position of the B ring compared to kaempferol (Fig. 6), which leads to the sugar moiety of UDPG transferred to the 7-OH group on the A ring. Also, luteolin is not natural constituents of apple fruit but kaempferol is<sup>41,42</sup>. All these differences make *MdP2'*GT exhibit a lowest  $k_{\text{cat}}/K_m$  toward luteolin. In addition, quercetin contains one extra OH group at the C3' position of the B ring compared to kaempferol (Fig. 6), it speculate that the residue Tyr399 can produce a stronger hindrance to quercetin than kaempferol making *MdP2'*GT exhibit a higher  $k_{\text{cat}}/K_m$  toward kaempferol than quercetin. However, the  $k_{\text{cat}}/K_m$  with UDPG/kaempferol combination was much lower than that with ADPG/ quercetin and UDP-Gal/ quercetin combinations for *MdP2'*GT (Table 4). On the other hand, the results signify the activated NDP-sugars may also influence the regioselectivity and catalytic efficiency of the Leloir GTs. One previous study also reports the activated NDP-sugars can influence the conformation of enzyme showing improved bioactivity<sup>45</sup>. This is also the first to report the donor-acceptor interactions during the final glycosylation step of phloridzin using a phloretin glycosyltransferase originating from apples.

As a novel P2'GT showing predominantly region-specific concerning the sugar attachment site and favoring phloridzin biosynthesis the most, *MdP2'*GT illustrates a distinctive set of key residues for substrates recognition. Moreover, the outstanding region-specific and broader substrate acceptance makes *MdP2'*GT an interesting enzyme for a promising catalyst in industrial preparation of phloridzin, isoquercitrin, hyperoside, and astragaline. Due to the no information on its crystal structure, the catalysis mechanism of *MdP2'*GT should be further elucidated by site directed mutagenesis or resolving its crystal structure or *MdP2'*GT-substrate complex in future study.

## Methods

**Chemicals.** NEM and PHMB with purity  $\geq 98\%$  were purchased from the Aladdin Reagent Co. (Shanghai, China). Marker proteins, PNGase F, Endo H, nucleoside diphosphate (NDP)-sugars, AEBSEF, Yeast Nitrogen Base (YNB, molecular biology grade), biotin and Geneticin (G418) with purity  $\geq 98\%$  were purchased from Sigma-Aldrich (St. Louis, MO, USA). The glycosyl acceptors evaluated in this study, 3-hydroxyphloridzin and trilobatin,



**Figure 6.** Structural formulas of the polyphenols acting as the acceptors of *MdP2'*/GT.

Enzyme	UGT88F1	UGT88F2	UGT71A15	UGT71K1	UGT71K2	UGT71A16	YjiC	DicGT4	References
UGT88F2	<b>I: 99.0% S: 99.4%</b>								19, 20
UGT71A15	<i>I: 34.6% S: 51.0%</i>	<i>I: 34.3% S: 50.9%</i>							19, 20, 21
UGT71K1	<i>I: 36.1% S: 53.6%</i>	<i>I: 36.1% S: 53.6%</i>	<i>I: 46.3% S: 67.4%</i>						19, 20, 21
UGT71K2	<i>I: 35.8% S: 53.0%</i>	<i>I: 35.8% S: 53.0%</i>	<i>I: 45.8% S: 67.0%</i>	<b>I: 93.1% S: 95.4%</b>					19, 20, 21
UGT71A16	<i>I: 34.4% S: 51.0%</i>	<i>I: 34.1% S: 53.9%</i>	<b>I: 91.7% S: 96.0%</b>	<i>I: 45.1% S: 66.6%</i>	<i>I: 44.3% S: 65.6%</i>				19, 20, 21
<sup>a</sup> YjiC	<i>I: 14.2% S: 46.7%</i>	<i>I: 14.2% S: 46.7%</i>	<i>I: 16.8% S: 44.4%</i>	<i>I: 15.7% S: 45.2%</i>	<i>I: 15.7% S: 45.2%</i>	<i>I: 16.4% S: 45.5%</i>			3, 19, 20, 21
<sup>b</sup> DicGT4	<i>I: 27.1% S: 42.3%</i>	<i>I: 27.3% S: 42.3%</i>	<i>I: 26.5% S: 41.4%</i>	<i>I: 27.0% S: 39.1%</i>	<i>I: 26.8% S: 39.4%</i>	<i>I: 26.5% S: 43.1%</i>	<i>I: 20.2% S: 48.2%</i>		3, 14, 19, 20, 21
<sup>c</sup> <i>MdP2'</i> /GT	<b>I: 92.6% S: 96.1%</b>	<b>I: 92.1% S: 95.6%</b>	<i>I: 31.6% S: 35.6%</i>	<i>I: 32.6% S: 34.6%</i>	<i>I: 33.6% S: 35.0%</i>	<i>I: 31.8% S: 37.1%</i>	<i>I: 13.2% S: 26.6%</i>	<i>I: 27.1% S: 33.1%</i>	This study

**Table 6.** Amino acid sequence identity (I) and similarity (S) of *MdP2'*/GT and other reported *P2'*/GTs.

Values for orthologous sequences are highlighted in bold; those for paralogous sequences are in italics.

<sup>a</sup>Calculated according to the sequences of YjiC following the method described as Gosch *et al.*<sup>20</sup>. <sup>b</sup>Calculated according to the sequences of DicGT4 following the method described as Gosch *et al.*<sup>20</sup>. <sup>c</sup>Calculated according to the sequences of *MdP2'*/GT shown in Fig. S1 following the method described as Gosch *et al.*<sup>20</sup>.

were purchased from Chengdu Must Bio-Technology Co., Ltd. (Chengdu, China) with purity >98%. All other reagents were of analytical grade.

***MdP2'*/GT activity.** The *MdP2'*/GT activity assay was performed following a previous method with minor modifications<sup>19</sup>. Reaction mixtures consisted of 0.5 ml buffer A (50 mM Tris/HCl, pH 7.5, and 5 mM DTT), 270 μM UDPG, 20 μM phloretin (dissolved in DMSO) and 20 μg purified *MdP2'*/GT. After incubation at 30 °C for 30 min, the reactions were immediately terminated by freezing in liquid N<sub>2</sub> and lyophilized. Methanol (0.25 ml) was then added to dissolve the components, and the product phloridzin was quantified by the HPLC-DAD method on a WondaSil<sup>®</sup> C<sub>18</sub> column (4.6 × 250 mm, ID = 5 μm, Shimadzu, Kyoto, Japan) as described in our previous study<sup>42</sup>. One unit of *MdP2'*/GT activity was defined as the amount of enzyme needed to produce 1 μmol product per second (s) at 30 °C and pH 7.5, and specific activity was expressed as μkat/Kg protein.

**Expression and purification of *MdP2'*/GT.** The open reading frame of *MdP2'*/GT was sub-cloned into the *Sna*BI and *Not*I sites of pPIC9K (Novagen) and transformed into *P. pastoris* GS115 (Novagen) by electroporation<sup>46,47</sup>. Engineered *P. pastoris* GS115 carrying pPIC9K/*MdP2'*/GT was cultured in a 5 L bioreactor (Applikon<sup>®</sup> Biotechnology, Foster City, USA) with a working volume of 1 L buffered minimal sorbitol medium (BMSM, 100 mM potassium phosphate, pH 7.0, 1.34% YNB, 4 × 10<sup>-5</sup>% biotin, and 5.0% sorbitol) in the presence of G418 (4.0 mg/ml) overnight at 28 °C, 280 rpm, and were induced, at the optical density (OD<sub>600</sub>) of 10.0–15.0, with 0.75% methanol at 25 °C for 122 h<sup>46,47</sup>. The resulting supernatant was harvested, dialyzed and lyophilized to a powder. The powder was dissolved in buffer B (50 mM Tris/HCl, pH 7.5, 2 mM MgCl<sub>2</sub>, 5 mM DTT, 0.5 mM AEBSF, and 5% glycerol), filtered through a 0.22-μm membrane (Millipore) and then loaded onto a pre-equilibrated diethyl-aminoethanol (DEAE) Sepharose column (HiTrap-DEAE-FF, 1 ml, Amersham Biosciences) operated with an AKTA purifier system (Amersham Biosciences, Uppsala, Sweden)<sup>28</sup>. The column was first washed with buffer B for 5 column volumes (CVs) and then eluted with buffer B containing NaCl (500 mM) a linear gradient (0 to 0.5 M) at a flow rate of 1.0 ml/min. The fractions of interest (identified by activity towards UDPG and Phloretin) were collected and dialyzed against buffer C (50 mM Tris/HCl, pH 7.5, 200 mM NaCl, 2 mM MgCl<sub>2</sub>, 5 mM DTT, and 0.5 mM AEBSF) at 4 °C. The dialyzed solution was further purified under native conditions on a gravity flow column (10 ml, GE Healthcare) containing 1 ml cOmplete His-Tag Purification Resin (Roche CH

05-893-682-001) pre-equilibrated with buffer C<sup>48</sup>. The column was stepwise washed with 20 CVs of 40–500 mM imidazole in buffer C. The fractions of interest, judged by SDS-PAGE and enzyme activity assays, were applied to the gel filtration on a Superdex 75 10/300 GL column (24 ml, Amersham Biosciences) using buffer C as the eluent<sup>49</sup>. The final purified fractions were pooled, desalted, and concentrated using a 10 K membrane ultracentrifugation system (Ultracel-10, Millipore, Massachusetts, USA)<sup>49</sup>. The enzyme purity was estimated by specific activity as described in enzyme activity assay, Coomassie Blue R-250 (Bio-Rad, CA, USA) stained 8% SDS-PAGE, and Western blot with an anti-His antibody<sup>19</sup>, respectively. The total protein was determined by the Bradford method using a Bradford kit with bovine serum albumin as the standard<sup>47,48</sup>.

**Molecular weight evaluation.** The concentration of the purified *MdP2'*GT solution was adjusted to 5 mg/ml and loaded onto a Sephacryl S-300 column (GE Healthcare, Piscataway, USA; 1.6 × 60 cm) pre-equilibrated with buffer D (50 mM Tris-HCl, pH 7.5, 200 mM NaCl, and 5 mM DTT) for SEC<sup>31</sup>. Five marker proteins (75 kDa conalbumin, 67 kDa bovine serum albumin, 45 kDa egg albumins, 29 kDa carbonic anhydrase, and 14.9 kDa lysozyme) were used to calibrate the column and generate a standard curve based on the partition coefficient of  $K_{av}$  (abscissa) and  $\lg M_r$  (ordinate).  $K_{av}$  is calculated on the the molecular mass and the elution volume of each protein;  $M_r$  is the molecular mass of the corresponding marker proteins.

**Glycosylation analysis.** Glycosylation analysis was performed by enzyme treatment with PNGase F and Endo H according to the methods described by Luciana Facchinetti *et al.*<sup>31</sup> and Schlenzig *et al.*<sup>35</sup>. The sizes of the deglycosylated *MdP2'*GT, PNGase F and Endo H were analyzed by SDS-PAGE and SEC as described in the molecular weight analysis. The native and deglycosylated *MdP2'*GT bands in the SDS-PAGE gel were excised and analyzed by matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF/MS) in the linear mode as described by Schlenzig *et al.*<sup>35</sup>.

**Optimum pH and temperature and stability.** The optimal pH was assessed in the pH range of 4.0–10.0 using 100 mM potassium phosphate buffer (pH 4.0–7.0) and Tris/HCl buffer (pH 7.0–10.0) at 30 °C. The optimal temperature was determined in the temperature range of 15–75 °C at pH 8.0. For the pH stability evaluation, *MdP2'*GT was incubated at 30 °C for 24 h at pH values varying from 6.5 to 9.5. For the temperature stability assessment, *MdP2'*GT was incubated at pH 8.0 at several temperatures (25, 35, 45, 55, and 65 °C) for different periods of time (0, 10, 20, 30, 40, 50, and 60 min). The residual activity of *MdP2'*GT was determined as described in the enzyme activity assay section and expressed as the relative activity compared with the initial activity, which was considered 100%.

**Effects of metal ions and other compounds.** The effects of metal ions (monovalent, divalent, and trivalent metal ions), chelating agent (EDTA), protectants (glycerol and AEBSF), thiol-containing compounds (3-ME and DTT), thiol inhibitors (PHMB and NEM), substrate co-solvents (methanol and DMSO), substrate analogs (UDP, uridine triphosphate (UTP), uridine monophosphate (UMP), and uridine) and product analogs (trilobatin and 3-hydroxyphloridzin) on the activity of purified *MdP2'*GT were examined as described in the enzyme activity assay section. The inhibition effects were calculated and expressed as relative activity (100%). The half-maximal inhibitory concentration (IC<sub>50</sub>) values, defined as the concentration of inhibitor required to reduce the original activity of *MdP2'*GT activity by 50%, were used to evaluate the effects.

**Substrate specificity assays.** Nine potential sugar donors, including UDPG, ADPG, guanosine diphosphate glucose, cytidine diphosphate glucose, thymidine diphosphate glucose, UDP-Gal, UDP-xylose, UDP-rhamnose, and UDP-pentose, and five types of potential sugar acceptors, including hydroxycinnamic acids (chlorogenic acid, caffeic acid, 4-coumaric acid, 3-coumaric acid, and 2-coumaric acid), chalcones (phloretin, isoliquiritigenin, naringenin chalcone, eriodictyol chalcone, and butein), flavanols (catechin, epicatechin, epigallocatechin, epicatechin gallate, and epigallocatechin gallate), flavonols (quercetin, dihydroquercetin, rutin, kaempferol, and myricetin), and flavones (baicalein, luteolin, eriodictyol, chrysin, and apigenin) were evaluated as potential *MdP2'*GT substrates. For chromatographic identification of the enzyme products, substrate specificity assays were scaled up to a total volume of 20 ml containing approximately 1.0 mg of purified *MdP2'*GT, 1.0 mM acceptor substrate, and 10.8 mM donor substrate in buffer A. The reactions were performed at 30 °C and pH 7.5 for 15 h and terminated by adding 50  $\mu$ l of glacial acetic acid. After lyophilization in liquid N<sub>2</sub>, the powder containing the enzyme products was re-dissolved in 10 ml of methanol and then filtered through a 0.22- $\mu$ m membrane (Millipore). The formed glycosides were first evaluated using the same HPLC-DAD method as described in our previous study<sup>42</sup>. Then, the target HPLC peaks were subjected to LC-MS/MS (Foster City, CA, USA) on a 4000 QTrap system from Applied Biosystems equipped with an electrospray ionization (ESI) interface and a HPLC system (Shimadzu, Tokyo, Japan)<sup>50</sup>. The chemical structures of target glycosides were further confirmed by NMR assays in DMSO-*d*<sub>6</sub> solvent (Sigma-Aldrich, St. Louis, USA)<sup>24</sup>. <sup>1</sup>H NMR and <sup>13</sup>C NMR spectroscopic analysis was performed on a Bruker Avance 500 MHz spectrometer (Bruker Corp., Billerica, USA) with. Chemical shifts were expressed in  $\delta$  ppm with coupling constants (J) determined using tetramethylsilane (TMS) as an internal standard.

**Kinetic parameters.** The kinetic parameters of *MdP2'*GT were determined by measuring its activity in the presence of various concentrations of sugar donor and acceptor substrates. For the donor substrates, the acceptor substrates concentration was fixed at 1.0 mM, with donors concentrations varying from 27  $\mu$ M to 25.4 mM. For the acceptor substrates, the donor substrates concentration was fixed at 10.8 mM, with acceptor concentration varying from 4.0  $\mu$ M to 2.0 mM. Kinetic constants ( $K_m$ ,  $V_{max}$ ,  $K_{cat}$ ) were calculated from Lineweaver-Burk plots. All assays were performed in triplicate.

**Molecular modeling and docking.** The 3D structures of *MdP2'*GT was modeled by submitting its full-length amino acid sequences (Fig. S1) to the I-TASSER (<http://zhanglab.ccmb.med.umich.edu/>). Structures of the substrates and inhibitor were built by ChemBioDraw Ultra 12.0 and their ground state conformation predicted was docked into the possible catalytic pocket of the appropriate *MdP2'*GT model according to the previous report<sup>10,27</sup>.

## References

- Overwin, H., Wray, V. & Hofer, B. Biotransformation of phloretin by amylosucrase yields three novel dihydrochalcone glucosides. *J Biotechnol.* **211**, 103–106 (2015).
- Chang, W. T., Huang, W. C. & Liou, C. J. Evaluation of the anti-inflammatory effects of phloretin and phlorizin in lipopolysaccharide-stimulated mouse macrophages. *Food Chem.* **134**, 972–979 (2012).
- Pandey, R. P. *et al.* Enzymatic synthesis of novel phloretin glucosides. *AEM.* **79**, 3516–3521 (2013).
- Xintaropoulou, C. *et al.* A comparative analysis of inhibitors of the glycolysis pathway in breast and ovarian cancer cell line models. *Clin trials* **4**, 6 (2015).
- Abkin, S. V. *et al.* Phloretin increases the anti-tumor efficacy of intratumorally delivered heat-shock protein 70 kDa (HSP70) in a murine model of melanoma. *Cancer Immunol Immun.* **65**, 83–92 (2016).
- San Miguel, S. M., Opperman, L. A., Allen, E. P., Zielinski, J. & Svoboda, K. K. Bioactive polyphenol antioxidants protect oral fibroblasts from ROS-inducing agents. *Arch Oral Biol.* **57**, 1657–1667 (2012).
- Gutmann, A. *et al.* Towards the synthesis of glycosylated dihydrochalcone natural products using glycosyltransferase-catalysed cascade reactions. *Green Chem.* **16**, 4417–4425 (2014).
- Gutmann, A., Krump, C., Bungaruang, L. & Nidetzky, B. A two-step O- to C-glycosidic bond rearrangement using complementary glycosyltransferase activities. *Chem Commun.* **50**, 5465–5468 (2014).
- Osmani, S. A., Bak, S. & Møller, B. L. Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. *Phytochemistry.* **70**, 325–347 (2009).
- Li, L. *et al.* Crystal structure of *Medicago truncatula* UGT85H2—insights into the structural basis of a multifunctional (iso) flavonoid glycosyltransferase. *J Mol Biol.* **370**, 951–963 (2007).
- Gosch, C., Halbwirth, H., Kuhn, J., Miosic, S. & Stich, K. Biosynthesis of phloridzin in apple (*Malus domestica* Borkh.). *Plant Sci.* **176**, 223–231 (2009).
- Gosch, C., Halbwirth, H. & Stich, K. Phloridzin: biosynthesis, distribution and physiological relevance in plants. *Phytochemistry.* **71**, 838–843 (2010).
- Brazier-Hicks, M. & Edwards, R. Metabolic engineering of the flavone-C-glycoside pathway using polyprotein technology. *Metab Eng.* **16**, 11–20 (2013).
- Werner, S. R. & Morgan, J. A. Expression of a *Dianthus* flavonoid glycosyltransferase in *Saccharomyces cerevisiae* for whole-cell biocatalysis. *J Biotechnol.* **142**, 233–241 (2009).
- Lee, J. *et al.* Phloridzin isolated from *Acanthopanax senticosus* promotes proliferation of  $\alpha 6$  integrin (CD 49f) and  $\beta 1$  integrin (CD29) enriched for a primary keratinocyte population through the ERK-mediated mTOR pathway. *Arch Dermatol Res.* **305**, 747–754 (2013).
- Sun, Y. *et al.* Preparative separation of five flavones from flowers of *Polygonum cuspidatum* by high-speed countercurrent chromatography. *J Sep Sci.* **37**, 1703–1709 (2014).
- Sun, Y., Li, W. & Liu, Z. Preparative isolation, quantification and antioxidant activity of dihydrochalcones from Sweet Tea (*Lithocarpus polystachyus* Rehd.). *J Chromatogr B.* **1002**, 372–378 (2015).
- Wang, Q. W., Liu, J. J., Liu, X., Zha, H. W. & Ren, N. Purification of phloridzin from *Lithocarpus polystachyus* Rehd leaves using polymeric adsorbents functionalized with glucosamine and  $\beta$ -cyclodextrin. *Chem Pap.* **68**, 1521–1531 (2014).
- Jugd , H., Nguy, D., Moller, I., Cooney, J. M. & Atkinson, R. G. Isolation and characterization of a novel glycosyltransferase that converts phloretin to phlorizin, a potent antioxidant in apple. *FEBS J.* **275**, 3804–3814 (2008).
- Gosch, C., Halbwirth, H., Schneider, B., H lscher, D. & Stich, K. Cloning and heterologous expression of glycosyltransferases from *Malus x domestica* and *Pyrus communis*, which convert phloretin to phloretin 2'-O-glucoside (phloridzin). *Plant Sci.* **178**, 299–306 (2010).
- Gosch, C. *et al.* Substrate specificity and contribution of the glycosyltransferase UGT71A15 to phloridzin biosynthesis. *Trees.* **26**, 259–271 (2012).
- Mohamed, Y. M. *et al.* Isolation and characterization of a heavy metal-resistant, thermophilic esterase from a Red Sea Brine Pool. *Sci Rep.* **3**, 3358 (2013).
- Fu, B., Li, H., Wang, X., Lee, F. S. & Cui, S. Isolation and identification of flavonoids in licorice and a study of their inhibitory effects on tyrosinase. *J Agr Food Chem.* **53**, 7408–7414 (2005).
- Braunberger, C. *et al.* LC-NMR, NMR, and LC-MS identification and LC-DAD quantification of flavonoids and ellagic acid derivatives in *Drosera peltata*. *J Chromatogr B.* **932**, 111–116 (2013).
- Hicks, J. M. *et al.* Quantification of chlorogenic acid and hyperoside directly from crude blueberry (*Vaccinium angustifolium*) leaf extract by NMR spectroscopy analysis: single-laboratory validation. *J AOAC Int.* **95**, 1406–1411 (2012).
- Žemlička, L. *et al.* Physicochemical and biological properties of luteolin-7-O- $\beta$ -D-glucoside (cynaroside) isolated from *Anthriscus sylvestris* (L.) Hoffm. *Monatshefte für Chemie-Chemical Monthly.* **145**, 1307–1318 (2014).
- Brazier-Hicks, M. *et al.* Characterization and engineering of the bifunctional N- and O-glycosyltransferase involved in xenobiotic metabolism in plants. *P Natl Acad Sci.* **104**, 20238–20243 (2007).
- Zhao, Y., Dong, X., Yu, L. & Sun, Y. Protein adsorption to poly (ethylenimine)-modified Sepharose FF: VI. Partial charge neutralization drastically increases uptake rate. *J Chromatogr A.* **1427**, 102–110 (2016).
- Liu, B. *et al.* Characterization and immunological activity of different forms of recombinant secreted Hc of botulinum neurotoxin serotype B products expressed in yeast. *Sci Rep.* **5**, 7678 (2015).
- Zou, S. *et al.* N-linked glycosylation influences on the catalytic and biochemical properties of *Penicillium purpurogenum*- $\beta$ -D-glucuronidase. *J Biotechnol.* **157**, 399–404 (2012).
- Facchinetti, C. G. *et al.* *Saccharomyces cerevisiae* asparaginase II, a potential antileukemic drug: Purification and characterization of the enzyme expressed in *Pichia pastoris*. *Protein Expres Purif.* **120**, 118–125 (2016).
- Ohashi, T., Nakakita, S. I., Sumiyoshi, W. & Takegawa, K. Production of heterologous glycoproteins by a glycosylation-defective *alg3och1* mutant of *Schizosaccharomyces pombe*. *J Biotechnol.* **150**, 348–356 (2010).
- Abdel-Naby, M. A., Fouad, A. A. & El-Refai, H. Catalytic and thermodynamic properties of glycosylated *Bacillus cereus* cyclodextrin glycosyltransferase. *Int J Biol Macromol.* **76**, 132–137 (2015).
- García-Cano, I. *et al.* Expression, purification, and characterization of a bifunctional 99-kDa peptidoglycan hydrolase from *Pediococcus acidilactici* ATCC 8042. *AMB.* **99**, 8563–8573 (2015).
- Schlenzig, D., Wermann, M., Ramsbeck, D., Moenke-Wedler, T. & Schilling, S. Expression, purification and initial characterization of human meprin  $\beta$  from *Pichia pastoris*. *Protein Expres Purif.* **116**, 75–81 (2015).

36. Ford, C. M., Boss, P. K. & Høj, P. B. Cloning and characterization of vitis vinifera UDP-Glucose: flavonoid 3-O-glucosyltransferase, a homologue of the enzyme encoded by the maize bronze-1 Locus that may primarily serve to glucosylate anthocyanidins *in vivo*. *J Biol Chem.* **273**, 9224–9233 (1998).
37. Sawada, S. y. *et al.* UDP-glucuronic acid: Anthocyanin glucuronosyltransferase from red daisy (*Bellis perennis*) flowers enzymology and phylogenetics of a novel glucuronosyltransferase involved in flower pigment biosynthesis. *J Biol Chem.* **280**, 899–906 (2005).
38. Wu, C. Z. *et al.* Enzymatic glycosylation of nonbenzoquinone geldanamycin analogs via *Bacillus* UDP-glycosyltransferase. *AEM.* **78**, 7680–7686 (2012).
39. Vogt, T. & Jones, P. Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends Plant Sci.* **5**, 380–386 (2000).
40. Vogt, T. Substrate specificity and sequence analysis define a polyphyletic origin of betanidin 5- and 6-O-glucosyltransferase from *Dorotheanthus bellidiformis*. *Planta* **214**, 492–495 (2002).
41. Xu, Y. *et al.* Variation in phenolic compounds and antioxidant activity in apple seeds of seven cultivars. *Saudi J Biol Sci.* **23**, 379–388 (2016).
42. Zhang, T. *et al.* Screening for antioxidant and antibacterial activities of phenolics from *Golden Delicious* apple pomace. *Chem Cent J.* **10**, 47–52 (2016).
43. Channale, S. M. *et al.* Characterization of two coleopteran alpha-amylases and molecular insights into their differential inhibition by synthetic alpha-amylase inhibitor, acarbose. *Insect Biochem Molec.* **74**, 1–11 (2016).
44. Mohamed, A. A., Ceunen, S., Geuns, J. M., Van den Ende, W. & De Ley, M. UDP-dependent glycosyltransferases involved in the biosynthesis of steviol glycosides. *J Plant Physiol.* **168**, 1136–1141 (2011).
45. Baig, I. *et al.* Mithramycin analogues generated by combinatorial biosynthesis show improved bioactivity. *J Nat Prod.* **71**, 199–207 (2008).
46. Yu, P., Yan, Y., Gu, Q. & Wang, X. Codon optimisation improves the expression of *Trichoderma viride* sp. endochitinase in *Pichia pastoris*. *Sci Rep.* **3**, 3043 (2013).
47. Yu, P., Yang, J. & Gu, H. Expression of HpaI in *Pichia pastoris* and optimization of conditions for the heparinase I production. *Carbohydr Pol.* **106**, 223–229 (2014).
48. Neef, J. *et al.* Versatile vector suite for the extracytoplasmic production and purification of heterologous His-tagged proteins in *Lactococcus lactis*. *AMB.* **99**, 9037–9048 (2015).
49. Hassan, N. *et al.* Biochemical and structural characterization of a thermostable  $\beta$ -glucosidase from *Halothermothrix orenii* for galacto-oligosaccharide synthesis. *Applied microbiology and biotechnology* **99**, 1731–1744 (2015).
50. Wang, Y. *et al.* Rapid Determination of Trace Sulfonamides in Milk by Graphene Oxide-Based Magnetic Solid Phase Extraction Coupled with HPLC-MS/MS. *Food Anal Method.* **9**, 2521–2530 (2016).

## Acknowledgements

This project was supported by Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant no. 20130204110032). The project was also supported by the program- Comprehensive utilization of the by-products of horticultural crops processing (Grant no. 201503142-10). We are grateful to Prof. Ge hong Wei (College of Life Science, Northwest A & F University) for providing relevant experimental platform and greatly appreciate the staff of State Key Laboratory of Crop Stress Biology for Arid Areas, NWAU.

## Author Contributions

Conceived and designed the experiments: M.F., T.Z. and X.W. Performed the experiments: T.Z., J.L. and Y.W. Analyzed the data: T.Z., J.L., Y.W. and M.F. Wrote the paper: T.Z., J.L., P.W., Y.X. and M.F.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Zhang, T. *et al.* Purification and characterization of a novel phloretin-2'-O-glycosyltransferase favoring phloridzin biosynthesis. *Sci. Rep.* **6**, 35274; doi: 10.1038/srep35274 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016