Article

Functional characterization and structural basis of an efficient di-*C*-glycosyltransferase from *Glycyrrhiza glabra*

Meng Zhang, Fu-Dong Li, Kai Li, Zi-Long Wang, Yu-Xi Wang, Jun-Bin He, Hui-Fei Su, Zhong-Yi Zhang, Chang-Biao Chi, Xiao-Meng Shi, Cai-Hong Yun, Zhi-Yong Zhang, Zhen-ming Liu, Liang-Ren Zhang, Dong-Hui Yang, Ming Ma, Xue Qiao, and Min Ye

J. Am. Chem. Soc., **Just Accepted Manuscript** • DOI: 10.1021/jacs.9b12211 • Publication Date (Web): 27 Jan 2020

Functional characterization and structural basis of an efficient di-*C*-glycosyltransferase from *Glycyrrhiza glabra*

Meng Zhang,[†,Δ] Fu-Dong Li,[‡,Δ] Kai Li,[†,Δ] Zi-Long Wang,[†] Yu-Xi Wang,[†] Jun-Bin He,[†] Hui-Fei Su,[†] Zhong-Yi Zhang,[†] Chang-Biao Chi,[†] Xiao-Meng Shi,[†] Cai-Hong Yun,[§] Zhi-Yong Zhang,[‡] Zhen-Ming Liu,[†] Liang-Ren Zhang,[†] Dong-Hui Yang,[†] Ming Ma,*[,†] Xue Qiao,*[,†] and Min Ye*[,†]

[†] State Key Laboratory of Natural and Biomimetic Drugs & Key Laboratory of Molecular Cardiovascular Sciences of Ministry of Education, School of Pharmaceutical Sciences, Peking University, 38 Xueyuan Road, Beijing 100191, China

[‡] National Science Center for Physical Sciences at Microscale Division of Molecular & Cell Biophysics and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China

[§] Department of Biochemistry and Biophysics & Department of Integration of Chinese and Western Medicine, School of Basic Medical Sciences, Peking University, 38 Xueyuan Road, Beijing 100191, China

ABSTRACT: A highly efficient di-*C*-glycosyltransferase GgCGT was discovered from the medicinal plant *Glycyrrhiza glabra*. GgCGT catalyzes a two-step di-*C*-glycosylation of flopropione-containing substrates with conversion rates of >98%. To elucidate the catalytic mechanisms of GgCGT, we solved its crystal structures in complex with UDP-Glc, UDP-Gal, UDP/phloretin, and UDP/nothofagin, respectively. Structural analysis revealed that the sugar donor selectivity was controlled by the hydrogen-bond interactions of sugar hydroxyl groups with D390 and other key residues. The di-*C*-glycosylation capability of GgCGT was attributed to a spacious substrate-binding tunnel, and the G389K mutation could switch di- to mono-*C*-glycosylation. GgCGT is the first di-*C*-glycosyltransferase with a crystal structure, and the first *C*-glycosyltransferase with a complex structure containing a sugar acceptor. This work could benefit the development of efficient biocatalysts to synthesize *C*-glycosides with medicinal potential.

INTRODUCTION

C-glycosides are a class of important natural products with noticeable druggability due to their potent bioactivities and high stability against gastro-intestinal hydrolysis metabolism.[1-3] For instance, puerarin (daidzein 8-*C*-glucoside) is widely used to treat cardiovascular diseases, and vicenin-2 (apigenin 6,8-di-*C*-glucoside) could inhibit diabetic vascular inflammation.[4,5] Most *C*-glycosides are derived from plants. Biosynthetically, they are formed by the catalysis of *C*-glycosyltransferases (CGTs). While a big family of *O*-glycosyltransferases (OGTs) are known, a limited number of CGTs have been discovered from plants. This is partly due to the low gene sequence similarities of CGTs. Up to now, thirteen CGTs have been characterized from plants, with isoflavone, flavone, resorcinol, or 2′,4′,6′-trihydroxyacetophenone units as substrates (sugar acceptors).[6-16] It is particularly noteworthy that only three of the plant CGTs could catalyze di-*C*-glycosylation, i.e. FcCGT and CuCGT from citrus plants, and MiCGTb from *Mangifera indica*.[6,7] MiCGTb and its mutants could specifically modify compounds with trihydroxyacetophenone units.

On the other hand, little is known about the catalytic mechanisms of CGTs. Recently, our group reported the first crystal structure for plant CGTs (TcCGT1, a flavone 8-*C*-glycosyltransferase from *Trollius chinensis*).[8] Through structural analysis and molecular docking, we found a new mode of catalytic mechanism for TcCGT1, which was initiated by substrate spontaneous deprotonation. Moreover, we interpreted the mechanisms for its broad substrate promiscuity and for the determination between *C*- and *O*-glycosylation activities. Several other research groups have tried to elucidate the catalytic mechanisms of plant CGTs by homology modeling and molecular docking. Hirade *et al*. found that H20, D85 and R292 of UGT708D1 were evolutionally conserved and were critical for the *C*-glycosylation activity.[9] Chen *et al*. reported that I152 of MiCGTb was the key amino acid residue for the second *C*-glycosylation.[7] However, due to the lack of crystal structures, the catalytic mechanisms for di-*C*-glycosylation still need to be further confirmed.

Glycyrrhiza glabra L. is a popular herbal medicine worldwide.[17] It contains abundant bioactive flavonoid and triterpenoid glycosides, particularly flavonoid di-*C*-glycosides such as vicenin-2, schaftoside, and isoschaftoside, which indicate the presence of di-*C*-glycosyltransferases.[18] While a series of OGTs have been reported, no CGTs have been discovered from *Glycyrrhiza* species, thus far.[19] Here we report a new di-*C*-

glycosyltransferase GgCGT from *G. glabra*. Its crystal structures in complex with UDP-Glc, UDP-Gal, UDP/phloretin, and UDP/nothofagin were determined, respectively. The molecular catalytic mechanisms for its di-*C*-glycosylation activity and sugar donor selectivity were elucidated.

■ RESULTS AND DISCUSSION

Molecular cloning and functional characterization of GgCGT. In this work, we analyzed the transcriptome of *G. glabra* in the EST database (ERR706841). The CGT gene *UGT708D1* from soybean (*Glycine max*, Leguminosae family) was used as a template.[9] Through analysis by conserved sequence, a cDNA fragment was discovered, and the target gene was designated as *GgCGT* (GenBank accession No. MH998596). As the stop codon was lacking, 3'-RACE was used to obtain a full-length open reading frame (ORF) of 1419 bp encoding 472 amino acids. It was cloned into a pET28a(+) vector. Recombinant GgCGT was expressed in the *E. coli* BL21(DE3) strain and was purified by Ni-NTA affinity chromatography (Figure S1). The nucleotide sequence identity of GgCGT and UGT708D1 was 62.7%.

The function of GgCGT was characterized by co-incubating 20 μg purified protein, 0.1 mM phloretin (**1**), and 0.5 mM UDP-glucose (UDP-Glc) in 100 μL of 50 mM NaH_2PO_4-Na_2HPO_4 buffer (pH 8.0, 37 °C, 2 h). As shown in Figure 1, at the presence of UDP-Glc, GgCGT could completely convert phloretin (**1**) into a more polar product **1a**. The mass spectrum of **1a** showed an $[M-H]^-$ ion at *m/z* 597, which could produce fragments at *m/z* 477 ($[M-H-120]^-$) and *m/z* 357 ($[M-H-240]^-$), indicating **1a** as a di-*C*-glucoside.[20] Product **1a** was purified from a preparative-scale reaction, and was identified as phloretin 3′,5′-di-*C*-β-D-glucoside by NMR analyses (SI).[21] These results confirmed GgCGT as a di-*C*-glycosyltransferase. It was named as UGT708B4 by the UGT Nomenclature Committee.[22]

The biochemical properties of GgCGT were investigated using **1** as the acceptor and UDP-Glc as the sugar donor. GgCGT showed a maximal activity at pH 8.0 (50 mM NaH_2PO_4-Na_2HPO_4) and 37 °C. The activity was independent of divalent metal ions (Figure S2). The di-*C*-glucosylation of **1** was a highly efficient two-step reaction. The first step was completed within 10 min to produce **8** (nothofagin). The kinetic parameters could not be determined. The second step was completed within 35 min, and the apparent K_m value for **8** was 4.5 μM (Figure S3).

Substrate specificity of GgCGT. To explore the substrate specificity of GgCGT, a library consisting of 51 phenolic compounds was screened by enzymatic assay using UDP-Glc as the donor (Figure 2). The reaction mixtures were analyzed by liquid chromatography coupled with mass spectrometry (LC/MS) (SI). GgCGT showed high promiscuity by catalyzing the *C*-glycosylation of 33 substrates. It is particularly noteworthy that GgCGT could completely convert six substrates (**1-4**, **6-7**) into di-*C*-glucosides (conversion rates >98%). Interestingly, all the six substrates contain a flopropione unit (shown in blue color in the structures). For **5** which has an acetophenone structure, although all the substrates were glycosylated, only 80% of the products were di-*C*-glucosides. GgCGT could also efficiently catalyze substrates with a 2′,4′-dihydroxyacetophenone (**9-11**) or 2′,4′,6′-trihydroxybenzaldehyde (**12-15**) unit, though only mono-*C*-glucosides were produced. These results indicated flopropione as the minimum unit required for di-*C*-glycosylation.

GgCGT also catalyzed the mono-*C*-glucosylation of various phenolic compounds (**16-33**), though the conversion rates were relatively low. For **23-49**, GgCGT also showed OGT activities (Figure 2). With substrates **50** and **51**, it showed potent *S*- and *N*-glycosylation activities, respectively. Thus, GgCGT is the second glycosyltransferase after TcCGT1 to form all the four types of glycosidic bonds.[8]

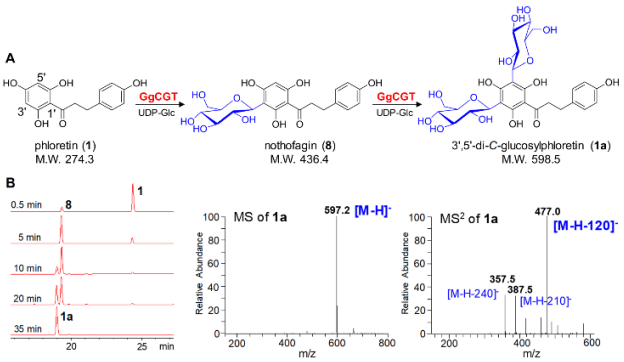


Figure 1. Di-*C*-glucosylation of phloretin (**1**) catalyzed by recombinant GgCGT. (A) Two-step glucosylation of **1** to produce **8** and **1a** sequentially. (B) Time-course study of the enzymatic reaction, and LC/MS analysis of the product **1a**. The HPLC chromatograms were recorded at 300 nm.

To confirm structures of the products, five di-*C*-glucosides (**1a**–**5a**) and two mono-*C*-glucosides (**5b**, **10a**) were purified. Three of them (**2a**, **4a**, **10a**) are new compounds. Their structures were unambiguously identified by NMR spectroscopic analyses (SI). All the *C*-glucosides contain a β-glycosidic bond, determined by the large coupling constants (J = 9.6-10.2 Hz) of the anomeric protons.[23] Compound **2a** exhibited moderate inhibitory activities against sodium-dependent glucose cotransporter 2 (SGLT2) with an IC_{50} value of 7.5 μM.[24]

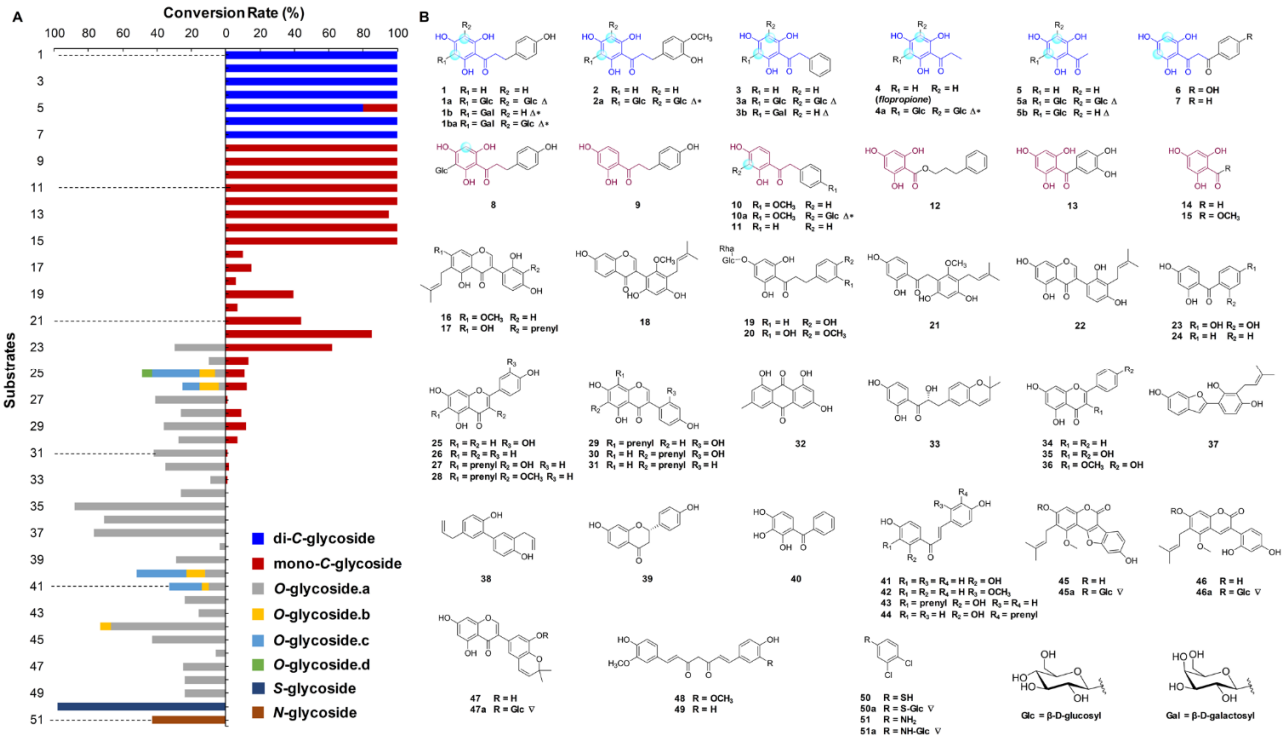


Figure 2. Substrate promiscuity of GgCGT. (A) Conversion rates (%) of glycosylated products for substrates **1**-**51**, using UDP-Glc as the sugar donor. (B) Structures of **1**-**51** and part of the glycosylated products. "Δ" means the products were purified and identified by NMR; "*" represents new compounds. The conversion rates were calculated by HPLC peak area ratio (SI). "▽" represents the glycosylation products were identified by comparing with reference standards. The reaction mixtures were incubated at 37 °C for 2h.

The biosynthetic precursors of di-*C*-glycosylflavones in plants have been proposed to be 2-hydroxyflavanones.[6] In this study, GgCGT could efficiently convert 2-hydroxyflavanones (**6**, **7**) into their di-*C*-glucosides. With the addition of hydrochloric acid, **6a** was readily converted into vicenin-2, a major *C*-glucoside of *G. glabra* (Figure S4). These results suggested that GgCGT might be involved in the biosynthesis of vicenin-2 in *G. glabra*.

Sugar donor specificity of GgCGT. To probe the sugar donor specificity of GgCGT, UDP-xylose (UDP-Xyl), UDP-galactose (UDP-Gal), UDP-arabinose (UDP-Ara), UDP-*N*-acetylglucosamine (UDP-GlcNAc), and UDP-glucuronic acid (UDP-GlcA) were tested. When phloretin (**1**) was used as the substrate, GgCGT could efficiently utilize all the donors except for UDP-GlcA to generate mono-*C*-glycosides (**1b**/**1c**/**1d**/**1e**) with conversion rates of >95% (Figure 3). Similar results were obtained using **6** as the substrate (Figure S5). Two *C*-galactosides (**1b**, **3b**) were purified and characterized by NMR spectroscopic analyses (SI). When the protein concentration was reduced to 30 μg/mL and the reaction time was reduced to 30 min, GgCGT showed the preference order of UDP-Glc > UDP-Xyl > UDP-Gal > UDP-Ara (Figure 3). The stereo-configuration of sugar 4-OH appeared to play a critical role in the catalytic efficiency.

Furthermore, GgCGT could efficiently convert **1b**/**1c**/**1d**/**1e** into di-*C*-glycosides with the addition of UDP-Glc (Figure S6). When compounds **2** and **3** were used as acceptors, similar high conversion rates (>98%) were observed in analytical reactions to generate di-*C*-glycosides with galactosyl and glucosyl moieties. A new compound (**1ba**) was purified and identified as phloretin 3′-*C*-β-D-galactoside-5′-*C*-β-D-glucoside. On the other hand, GgCGT could also convert nothofagin (**8**) into di-*C*-glycosides with different sugars, which had not been reported for CGTs. These results indicated that GgCGT had more efficient and versatile catalytic activities than previously reported CGTs, especially in the second step of *C*-glycosylation.

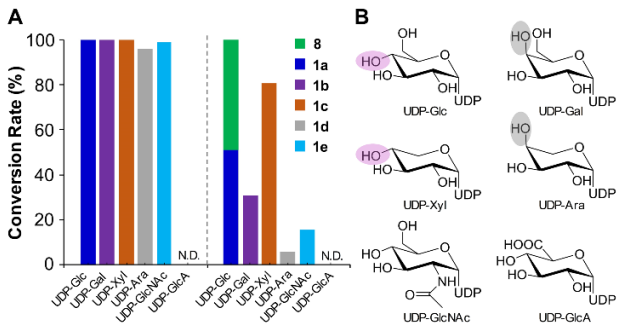


Figure 3. Probing sugar donor promiscuity of GgCGT. (A) Relative conversion rates of phloretin (**1**) using different sugar donors at high enzyme concentration (left, 200 μg/mL, 2h) and low enzyme concentration (right, 30 μg/mL, 0.5h). (B) Structures of sugar donors. The structures of **1b**-**1e** are shown in Figure S6.

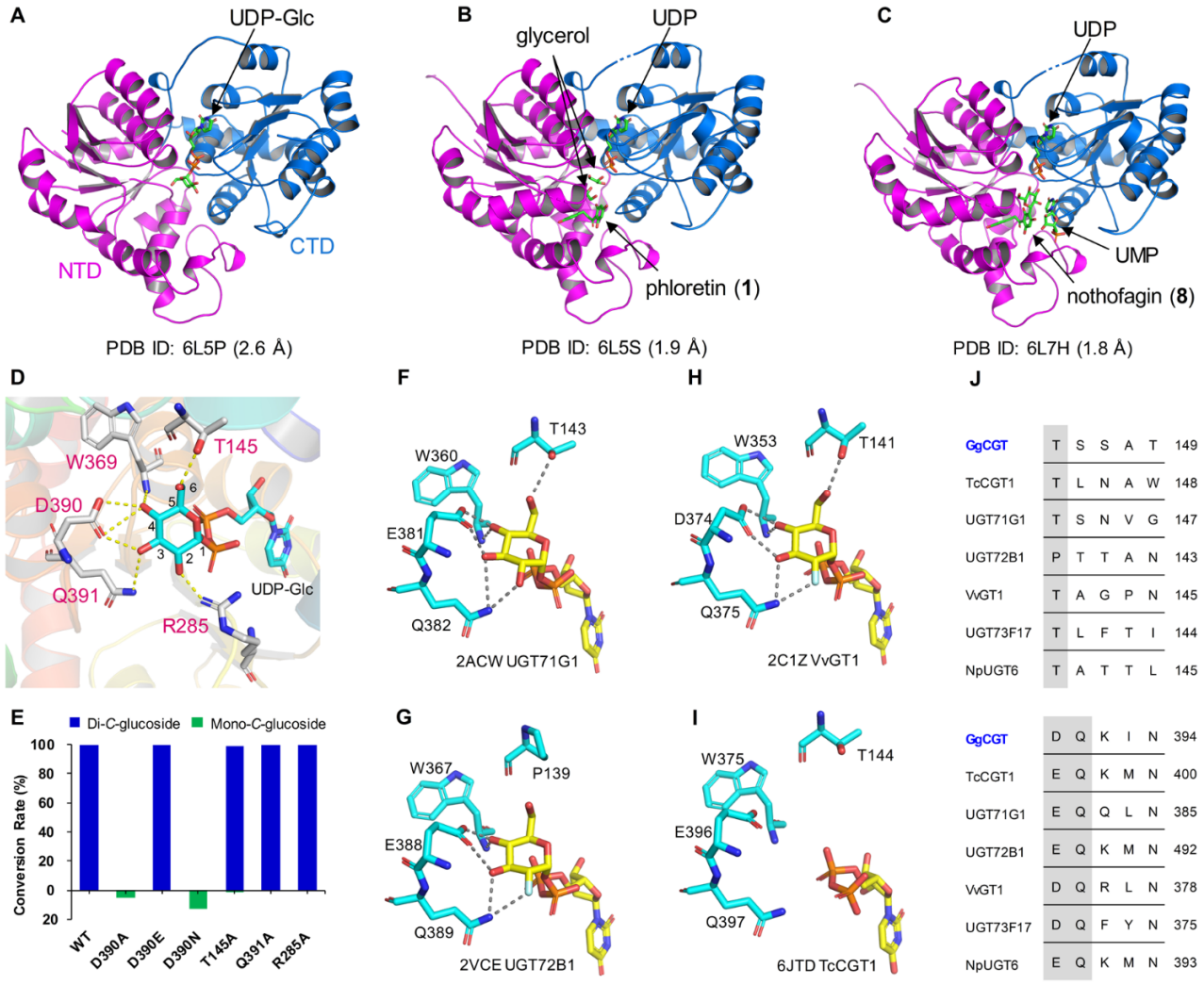


Figure 4. The crystal structures of GgCGT and its sugar donor selectivity analysis. (A) The crystal structures of GgCGT/UDP-Glc, highlighting an NTD domain (purple) and a CTD domain (blue). The sugar donor is shown as green sticks. (B/C) The crystal structures of GgCGT/UDP/phloretin (**1**) and GgCGT/UDP/nothofagin (**8**). (D) The hydrogen-bond interactions of the Glc moiety (yellow dashes) with surrounding residues of GgCGT. (E) The *C*-glycosylation conversion rates of wild-type (WT) GgCGT and mutants, using phloretin (**1**) and UDP-Glc as the sugar acceptor and donor, respectively. The reaction mixtures were incubated at 37 °C for 2h. (F-I) Other structures of UGTs with conserved residues contributing sugar recognition. The yellow sticks represent UDP-Glc molecules (UDP for TcCGT1), and the cyan sticks represent the conserved binding residues. The hydrogen-bond interactions are shown as grey dashes. (J) Sequence alignment of GgCGT with other UGTs, and the conserved residues for sugar donor binding are labeled in grey shade.

Crystal structures of GgCGT in complex with different ligands. To further investigate the catalytic mechanisms of GgCGT, we solved the complex crystal structures of GgCGT/UDP-Glc (PDB ID: 6L5P), GgCGT/UDP-Gal (PDB ID: 6L5Q), GgCGT/UDP/phloretin (PDB ID: 6L5S/6L5R), and GgCGT/UDP/nothofagin (PDB ID: 6L7H) at 2.6, 2.9, 1.9/2.9, and 1.8 Å resolution, respectively (Figure 4A, Table S1). Different from previously reported UDP-glycosyltransferase (UGT) crystals, the structures obtained in this study showed clear electron density of the intact UDP-sugar and the substrates. The 6L5S structure contains two molecules of glycerol in the active pocket because the crystals were frozen with 30% glycerol. We then solved a new structure of GgCGT/UDP/phloretin (6L5R) with no glycerol, by freezing the crystals in 30% PEG400. Although the resolution of 6L5R was relatively low, alignment of the two structures revealed that the position of phloretin was not remarkably affected by glycerol (Figure S7). Thus, the 6L5S structure was used for further analysis in this study.

To our knowledge, GgCGT is the first di-*C*-glycosyltransferase with a crystal structure. GgCGT adopts a canonical GT-B fold consisting of two Rossmann-like $\beta/\alpha/\beta$ domains that face each other and are separated by a deep cleft.[25] The N-terminal domain (NTD, residues 1-248 and 454-469) and the C-terminal

domain (CTD, residues 249-453) are responsible primarily for acceptor and sugar donor binding, respectively. The structure of GgCGT in different complexes varied slightly, indicating that combination with different ligands may affect the protein structure.

Catalytic mechanism for sugar donor selectivity of GgCGT. The GgCGT/UDP-Glc structure indicated that the 2-OH, 6-OH, 3-OH, and 4-OH groups of the glucose moiety could form hydrogen bonds with R285, T145, Q391, and D390, respectively. Particularly, the terminal carboxyl group of D390 could interact with both 3-OH (one interaction) and 4-OH (two interactions) through three hydrogen bonds, and thus plays an essential role in stabilizing UDP-Glc (Figure 4D). Consistently, the D390A and D390N mutations almost abolished the *C*-glycosylation activity (Figure 4H). However, the D390E mutant retained strong di-*C*-glycosylation activity, for that glutamic acid (E) has a similar side chain carboxyl group as aspartic acid (D). Moreover, the 4-OH of Glc also forms weak hydrogen-bonding with W369. These interactions render UDP-Glc the preferred sugar donor for GgCGT.

Although the GgCGT/UDP-Gal structure has relatively low resolution, electron density of the Gal moiety could be recognized. The orientation of Gal was slightly different from that of Glc, thus may lead to weak interactions with the protein. Particularly, there should be no significant hydrogen-bonding between 4-OH of Gal and surrounding residues. These speculations were confirmed by the remarkably decreased activities of the mutants (Figure S7). Consistently, SPR (surface plasmon resonance) analysis revealed that GgCGT had strong combination with UDP-Glc, but not UDP-Gal (Figure S8). The affinity (K_D, equilibrium dissociation constant) of GgCGT binding to UDP-Glc was 15.5 μM.

Furthermore, we speculated the combination mode of UDP-Xyl and UDP-Ara based on the structures of GgCGT/UDP-Glc and GgCGT/UDP-Gal, respectively (Figure S9). The sugar moieties had few hydrogen-bond interactions and thus led to low conversion rates. These analyses supported the sugar donor preference of GgCGT: UDP-Glc (seven interactions) > UDP-Xyl (six) > UDP-Gal (four) > UDP-Ara (three). Intriguingly, a number of CGTs and OGTs have the same sugar preference as GgCGT. These GTs include UGT71G1, UGT72B1, VvGT1, and TcCGT1, whose crystal structures had been solved, together with UGT73F17 and NpUGT6, OGTs recently discovered by our group.[8,19,26-28] The amino acid residues T145, D390, and Q391 are highly conserved in these enzymes (Figure 4F-J). These results supported our deductions on the mechanism of sugar donor selectivity of GgCGT.

Catalytic mechanism for di-*C*-glycosylation of GgCGT. To elucidate the mechanism of di-*C*-glycosylation, we superimposed the structures of GgCGT/UDP/phloretin and GgCGT/UDP-Glc (Figure 5A). A hydrophobic tunnel was formed by a series of hydrophobic amino acid residues (F91/F92/F95/F144/F151/F153/F154/F199/F203) to accommodate the substrate (Figure S10). In the

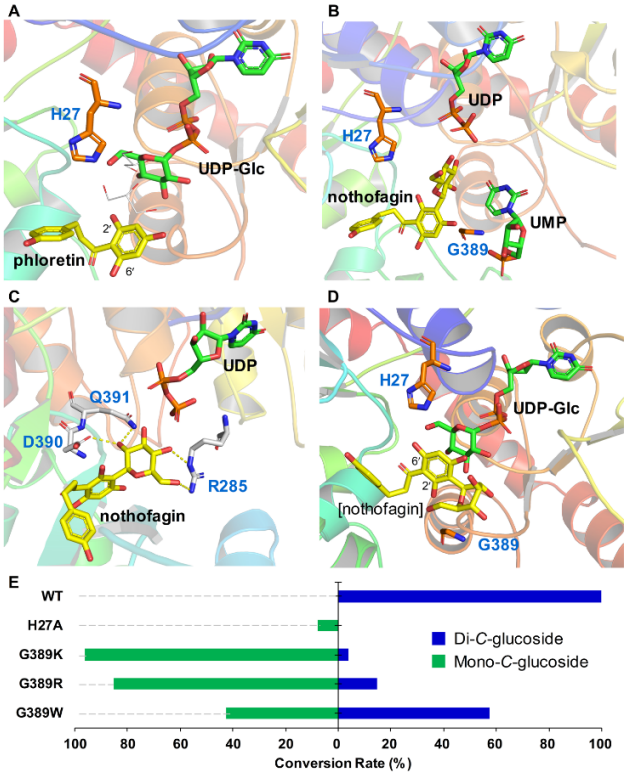


Figure 5. Sugar acceptor binding analysis. (A) Structural analysis of the *C*-glycosylation through structure alignment of GgCGT/UDP-Glc and GgCGT/UDP/phloretin. Glycerol is shown as thin sticks (grey). (B) Structural analysis of GgCGT/UDP/nothofagin. (C) Hydrogen-bond interactions of the sugar moiety of nothofagin with surrounding residues. (D) Molecular docking with nothofagin (**8**) (yellow sticks) into the substrate binding pocket of GgCGT. (E) Conversion rates of wild-type (WT) GgCGT and its mutants, using **1** and UDP-Glc as the sugar acceptor and donor, respectively. Labels with brackets indicate predicted (by molecular docking) models of the small molecule. The reaction mixtures were incubated at 37 °C for 2h.

superimposed structure, C-3′ of phloretin (**1**) was close to the anomeric carbon of UDP-Glc, and could facilitate the reaction. In the structure of GgCGT/UDP/nothofagin, the position of nothofagin (**8**) was similar to that of phloretin (Figure 5B). The sugar moiety of **8** was located near UDP, and could interact with surrounding residues through hydrogen bonding (Figure 5C). This structure may demonstrate the state after the first *C*-glycosylation.

H27 was close to 2′-OH of phloretin (**1**), and could initiate the glycosylation reaction by deprotonation to produce nothofagin (**8**).[9,26] This speculation was proved by the H27A mutant, which showed weak *C*-glycosylation activity (Figure 5E). In addition, the mutants H27D, H27E, and H27F showed very weak or no *C*-glycosylation activities (Figure S11). In contrast, H27K remained strong *C*-glycosylation activity, probably because lysine (K) is a basic amino acid like histidine (H). Moreover, supplementary imidazole or its analogues in the reaction mixture could increase the catalytic activities of H27A

(Figure S11).[29] According to these results, we speculate that His27 in GgCGT acts as a base to facilitate deprotonation of the substrate. This key histidine residue is generally present in plant glycosyltransferases.[30]

Similarly, H27 was close to 6′-OH of nothofagin (**8**) in the molecular docking model of GgCGT/UDP-Glc/**8**, and could continue to initiate the second step *C*-glycosylation (Figure 5D).[31] For sugar acceptors with a resorcinol unit (**9**-**11**), the second *C*-glycosylation cannot be initiated due to the absence of 6′-OH. This deduction explained why the flopropione unit was the preferred substrate structure for di-*C*-glycosylation.

G389 also played a critical role in the di-*C*-glycosylation of GgCGT. In the GgCGT/UDP/nothofagin structure, there was a UMP (uridine monophosphate) molecule near G389, indicating GgCGT has an adequate substrate-binding tunnel to accommodate the 3′-*C*-glucosyl moiety of **8**. Consistently, in the docking model of GgCGT/UDP-Glc/nothofagin, the sugar moiety of **8** was located near G389. When G389 was substituted by other amino acids with a long side chain, such as lysine, arginine, or tryptophan, the di-*C*-glycosylation activity of GgCGT decreased remarkably (Figure 5E). Particularly, the G389K mutation could almost switch di- to mono-*C*-glycosylation. The spacious binding pocket also explained the broad substrate promiscuity of GgCGT.

Based on the above evidences, we proposed the reaction process for the di-*C*-glycosylation catalyzed by GgCGT (Figure S12). Initially, UDP-Glc and phloretin (**1**) are bound to GgCGT in the optimal positions to complete the first *C*-glycosylation. After the nucleophilic substitution, UDP and **8** are released as the protein structure changed dynamically during the reaction. The protein structure then changes to the original form to prepare for the second step reaction. Compound **8** then binds to GgCGT with its 6′-OH close to H27 (the benzene ring rotated) to accomplish the second *C*-glycosylation.

CONCLUSION

In summary, we characterized a new di-*C*-glycosyltransferase GgCGT from the medicinal plant *Glycyrrhiza glabra*. We solved its crystal structures in complex with UDP-Glc, UDP-Gal, UDP/phloretin, and UDP/nothofagin, respectively. GgCGT represents the first di-*C*-glycosyltransferase with a crystal structure, and the first CGT with complex structures with both UDP-sugar and the substrate. It could efficiently catalyze di-*C*-glycosylation of at least six substrates containing a flopropione unit, as well as the mono-*C*-glycosylation of at least 27 phenolic compounds. A total of 10 products were purified, and 5 of them are new compounds, including **2a** as a moderate SGLT2 inhibitor. GgCGT showed the sugar donor preference of UDP-Glc > UDP-Xyl > UDP-Gal > UDP-Ara. Structural analysis indicated this preference may be determined by the hydrogen-bonding interactions of sugar hydroxyl groups with key amino acid residues including R285, T145, D390, and Q391, especially those of 4-OH with the carboxyl group of D390 side chain. The flopropione unit was the minimum required unit for the di-*C*-glycosylation due to the interactions of its 2′-/6′-OH with H27. The spacious substrate-binding tunnel near G389 was critical for the di-*C*-glycosylation activity and the broad substrate promiscuity of GgCGT. This work provides insights into the catalytic mechanisms of *C*-glycosyltransferases, which could be valuable in developing efficient biocatalysts to synthesize *C*-glycosides of medicinal potential.

ASSOCIATED CONTENT

Supporting Information. The Supporting Information is available free of charge on the ACS Publications website at DOI:
Detailed descriptions of all experiments: molecular cloning, expression, and purification of GgCGT; phylogenetic analysis; effects of pH, temperature and divalent metal ions; determination of kinetic parameters; substrates screening of GgCGT; scaled-up reactions; the NMR and HRESIMS data of glycosylated products; inhibitory activity assay against SGLT2; crystal data of GgCGT; mutagenesis and enzyme assay (PDF).

AUTHOR INFORMATION

Corresponding Author

* E-mail: mma@bjmu.edu.cn
* E-mail: qiaoxue@bjmu.edu.cn
* E-mail: yemin@bjmu.edu.cn

Author Contributions

[Δ]These authors contributed equally to this work.

Notes

Any additional relevant notes should be placed here.

ACKNOWLEDGMENT

The authors thank Dr. Jing Wang and Dr. Qian Wang at State Key Laboratory of Natural and Biomimetic Drugs of Peking university for technical help in the SPR experiments. This work was supported by National Natural Science Foundation of China (No. 81725023, 81891010/81891011, 81973448, 81921001, and 21907002), Beijing Natural Science Foundation (JQ18027), and the Science & Technology Department of Xinjiang Uygur Autonomous Region (2018AB012).

REFERENCES

[1] Williams, G. J.; Zhang, C.; Thorson, J. S. Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. *Nat. Chem. Biol.* **2007**, *3*, 657-662.

[2] Langenhan, J. M.; Griffith, B. R.; Thorson, J. S. Neoglycorandomization and chemoenzymatic glycorandomization: two complementary tools for natural product diversification. *J. Nat. Prod.* **2005**, *68*, 1696-1711.

[3] Talhi, O.; Silva, A. M. S. Advances in *C*-glycosylflavonoid research. *Curr. Org. Chem.* **2012**, *16*, 859-896.

[4] Wei, S. Y.; Chen, Y.; Xu, X. Y. Progress on the pharmacological research of puerarin: a review. *Chin. J. Nat. Med.* **2014**, *12*, 407-414.

[5] Ku, S. K.; Bae, J. S. Vicenin-2 and scolymoside inhibit high-glucose-induced vascular inflammation *in vitro* and *in vivo*. *Can. J. Physiol. Pharmacol.* **2016**, *94*, 287-295.

[6] Ito, T.; Fujimoto, S.; Suito, F.; Shimosaka, M.; Taguchi, G. *C*-glycosyltransferases catalyzing the formation of di-*C*-glucosyl flavonoids in citrus plants. *Plant J.* **2017**, *91*, 187-198.

[7] Chen, D. W.; Fan, S.; Chen, R. D.; Xie, K. B.; Yin, S.; Sun, L. L.; Liu, J. M.; Yang, L.; Kong, J. Q.; Yang, Z. Y.; Dai, J. G. Probing and engineering key residues for bis-*C*-glycosylation and promiscuity of a *C*-glycosyltransferase. *ACS Catal.* **2018**, *8*, 4917-4927.

[8] He, J. B.; Zhao, P.; Hu, Z. M.; Liu, S.; Kuang, Y.; Zhang, M.; Li, B.; Yun, C. H.; Qiao, X.; Ye, M. Molecular and structural characterization of a promiscuous *C*-glycosyltransferase from *Trollius chinensis*. *Angew. Chem. Int. Ed.* **2019**, *58*, 11513-11520.

[9] Hirade, Y.; Kotoku, N.; Terasaka, K.; Saijo-Hamano, Y.; Fukumoto, A.; Mizukami, H. Identification and functional analysis of 2-hydroxyflavanone *C*-glucosyltransferase in soybean (*Glycine max*). *FEBS Lett.* **2015**, *589*, 1778-1786.

[10] Wang, X.; Li, C. F.; Zhou, C.; Li, J.; Zhang, Y. S. Molecular characterization of the *C*-glucosylation for puerarin biosynthesis in *Pueraria lobata*. *Plant J.* **2017**, *90*, 535-546.

[11] Sasaki, N.; Nishizaki, Y.; Yamada, E.; Tatsuzawa, F.; Nakatsuka, T.; Takahashi, H.; Nishihara, M. Identification of the glucosyltransferase that mediates direct flavone *C*-glucosylation in *Gentiana triflora*. *FEBS Lett.* **2015**, *589*, 182-187.

[12] Chen, D. W.; Chen, R. D.; Wang, R. S.; Li, J. H.; Xie, K. B.; Bian, C. C.; Sun, L. L.; Zhang, X. L.; Liu, J. M.; Yang, L.; Ye, F.; Yu, X. M.; Dai, J. G. Probing the catalytic promiscuity of a regio- and stereospecific *C*-glycosyltransferase from *Mangifera indica*. *Angew. Chem. Int. Ed.* **2015**, *54*, 12678-12682.

[13] Brazier-Hicks, M.; Evans, K. M.; Gershater, M. C.; Puschmann, H.; Steel, P. G.; Edwards, R. The *C*-glycosylation of flavonoids in cereals. *J. Biol. Chem.* **2009**, *284*, 17926-17934.

[14] Falcone Ferreyra, M. L.; Rodriguez, E.; Casas, M. I.; Labadie, G.; Grotewold, E.; Casati, P. Identification of a bifunctional maize *C*- and *O*-glucosyltransferase. *J. Biol. Chem.* **2013**, *288*, 31678-31688.

[15] Nagatomo, Y.; Usui, S.; Ito, T.; Kato, A.; Shimosaka, M.; Taguchi, G. Purification, molecular cloning and functional characterization of flavonoid *C*-glucosyltransferases from *Fagopyrum esculentum* M. (buckwheat) Cotyledon. *Plant J.* **2014**, *80*, 437-448.

[16] Mashima, K.; Hatano, M.; Suzuki, Hideyuki.; Shimosaka, M.; Taguchi, G. Identification and characterization of apigenin 6-*C*-Glucosyltransferase involved in biosynthesis of isosaponarin in Wasabi (*Eutrema japonicum*). *Plant Cell Physiol.* **2019**, *60*, 2733-2743.

[17] Montero, L.; Ibanez, E.; Russo, M.; di Sanzo, R.; Rastrelli, L.; Piccinelli, A. L.; Celano, R.; Cifuentes, A.; Herrero, M. Metabolite profiling of licorice (*Glycyrrhiza glabra*) from different locations using comprehensive two-dimensional liquid chromatography coupled to diode array and tandem mass spectrometry detection. *Anal. Chim. Acta* **2016**, *913*, 145-159.

[18] Song, W.; Qiao, X.; Chen, K.; Wang, Y.; Ji, S.; Feng, J.; Li, K.; Lin, Y.; Ye, M. Biosynthesis-based quantitative analysis of 151 secondary metabolites of licorice to differentiate medicinal *Glycyrrhiza* species and their hybrids. *Anal. Chem.* **2017**, *89*, 3146-3153.

[19] (a) He, J. B.; Chen, K.; Hu, Z. M.; Li, K.; Song, W.; Yu, L. Y.; Leung, C. H.; Ma, D. L.; Qiao, X.; Ye, M. UGT73F17, a new glycosyltransferase from *Glycyrrhiza uralensis*, catalyzes the regiospecific glycosylation of pentacyclic triterpenoids. *Chem. Commun.* **2018**, *54*, 8594-8597. (b) Chen, K.; Hu, Z. M.; Song, W.; Wang, Z. L.; He, J. B.; Shi, X. M.; Cui, Q. H.; Qiao, X.; Ye, M. Diversity of *O*-glycosyltransferases contributes to the biosynthesis of flavonoid and triterpenoid glycosides in *Glycyrrhiza uralensis*. *ACS Synth. Biol.* **2019**, *8*, 1858-1866. (c) Nagashima, S.; Inagaki, R.; Kubo, A.; Hirotani, M.; Yoshikawa, T. cDNA cloning and expression of isoflavonoid-specific glucosyltransferase from *Glycyrrhiza echinata* cell-suspension cultures. *Planta* **2004**, *218*, 456-459. (d) Xu, G. J.; Cai, W.; Gao, W.; Liu, C. S. A novel glucuronosyltransferase has an unprecedented ability to catalyse continuous two-step glucuronosylation of glycyrrhetinic acid to yield glycyrrhizin. *New Phytol.* **2016**, *212*, 123-135. (e) Nomura, Y.; Seki, H.; Suzuki, T.; Ohyama, K.; Mizutani, M.; Kaku, T.; Tamura, K.; Ono, E.; Horikawa, M.; Sudo, H.; Hayashi, H.; Saito, K.; Muranaka, T. Functional specialization of UDP-glycosyltransferase 73P12 in licorice to produce a sweet triterpenoid saponin, glycyrrhizin. *Plant J.* **2019**, *99*, 1127–1143.

[20] Abad-García, B.; Garmón-Lobato, S.; Berrueta, L. A.; Gallo, B.; Vicente, F. New features on the fragmentation and differentiation of *C*-glycosidic flavone isomers by positive electrospray ionization and triple quadrupole mass spectrometry. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1834–1842.

[21] Kokotkiewicz, A.; Luczkiewicz, M.; Sowinski, P.; Glod, D.; Gorynski, K.; Bucinski, A. Isolation and structure elucidation of phenolic compounds from *Cyclopia subternata* Vogel (honeybush) intact plant and *in vitro* cultures. *Food Chem.* **2012**, *133*, 1373-1382.

[22] Homepage of the UGT Nomenclature Committee, http://prime.vetmed.wsu.edu/resources/udp-glucuronsyltransferase-homepage.

[23] Krafczyk, N.; Glomb, M. A. Characterization of phenolic compounds in rooibos tea. *J. Agric. Food Chem.* **2008**, *56*, 3368-3376.

[24] Jesus, A. R.; Vila-Vicosa, D.; Machuqueiro, M.; Marques, A. P.; Dore, T. M.; Rauter, A. P. Targeting type 2 diabetes with *C*-glucosyl dihydrochalcones as selective sodium glucose co-transporter 2 (SGLT2) inhibitors: synthesis and biological evaluation. *J. Med. Chem.* **2017**, *60*, 568-579.

[25] Breton, C.; Fournel-Gigleux, S.; Palcic, M. M. Recent structures, evolution and mechanisms of glycosyltransferases. *Curr. Opin. Struct. Biol.* **2012**, *22*, 540-549.

[26] (a) Shao, H.; He, X. Z.; Achnine, L.; Blount, J. W.; Dixon, R. A.; Wang, X. Q. Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. *Plant Cell* **2005**, *17*, 3141-3154. (b) Offen, W.; Martinez-Fleites, C.; Yang, M.; Kiat-Lim, E.; Davis, B. G.; Tarling, C. A.; Ford, C. M.; Bowles, D. J.; Davies, G. J. Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J.* **2006**, *25*, 1396-1405. (c) Li, L.; Modolo, L. V.; Escamilia-Trevino, L. L.; Achnine, L.; Dixon, R. A.; Wang, X. Q. Crystal structure of *medicago truncatula* UGT85H2 - insights into the structural basis of a multifunctional (iso)flavonoid glycosyltransferase. *J. Mol. Biol.* **2007**, *370*, 951-963. (d) Brazier-Hicks, M.; Offen, W. A.; Gershater, M. C.; Revett, T. J.; Lim, E. K.; Bowles, D. J.; Davies, G. J.; Edwards, R. Characterization and engineering of the bifunctional *N*- and *O*-glucosyltransferase involved in xenobiotic metabolism in plants. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 20238-20243. (e) Gutmann, A.; Nidetzky, B. Switching between *O*- and *C*-glycosyltransferase through exchange of active-site motifs. *Angew. Chem. Int. Ed.* **2012**, *51*, 12879-12883.

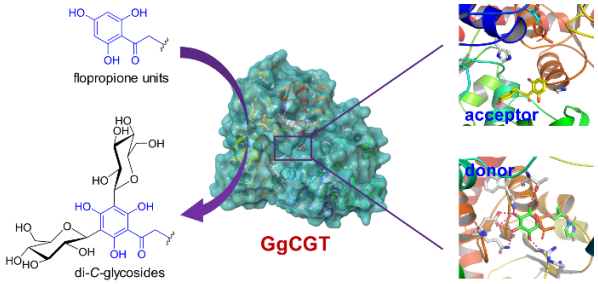
[27] Hu, Z. M.; He, J. B.; Chen, K.; Wang, Z. L.; Liu, J. Y.; Qiao, X.; Ye, M. Molecular cloning and biochemical characterization of a new flavonoid glycosyltransferase from the aquatic plant lotus. *Biochem. Biophys. Res. Commun.* **2019**, *510*, 315-321.

[28] Brazier-Hicks, M.; Edwards, R. Functional importance of the family 1 glucosyltransferase UGT72B1 in the metabolism of xenobiotics in *Arabidopsis thaliana*. *Plant J.* **2005**, *42*, 556-566.

[29] Newmyer, S. L.; deMontellano, P. R .O. Rescue of the catalytic activity of an H42A mutant of horseradish peroxidase by exogenous imidazoles. *J. Biol. Chem.* **1996**, *271*, 14891-14896.

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

[31] Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem*. **2010**, *31*, 455-461.



Insert Table of Contents artwork here

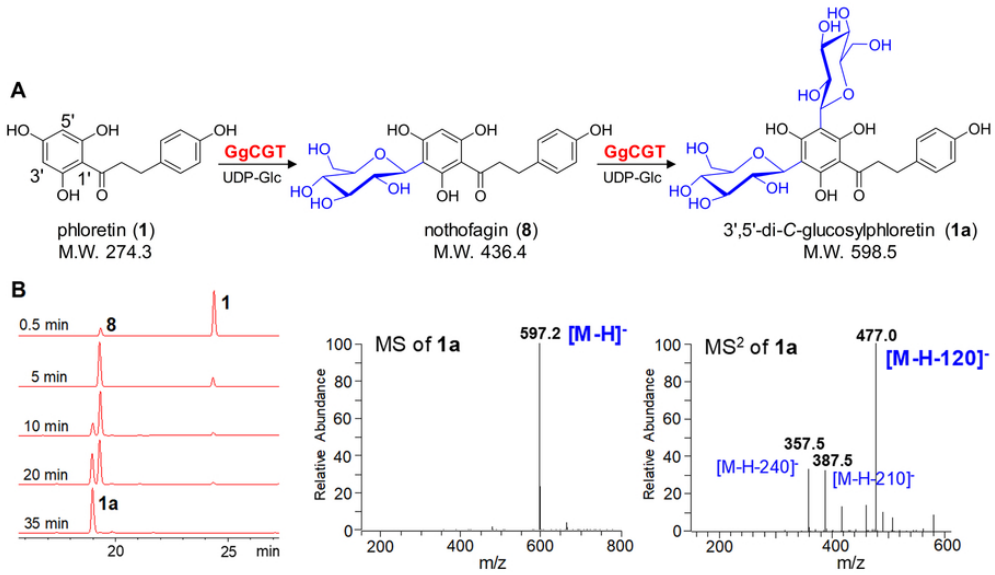


Figure 1. Di-*C*-glucosylation of phloretin (**1**) catalyzed by recombinant GgCGT. (A) Two-step glucosylation of **1** to produce **8** and **1a** sequentially. (B) Time-course study of the enzymatic reaction, and LC/MS analysis of the product **1a**. The HPLC chromatograms were recorded at 300 nm.

38x22mm (600 x 600 DPI)

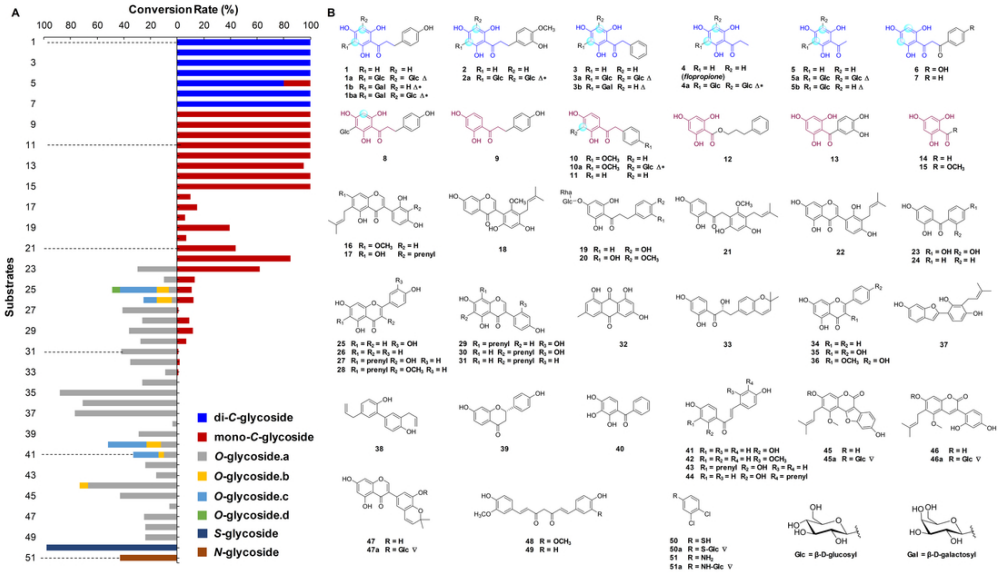


Figure 2. Substrate promiscuity of GgCGT. (A) Conversion rates (%) of glycosylated products for substrates **1**-**51**, using UDP-Glc as the sugar donor. (B) Structures of **1**-**51** and part of the glycosylated products. "Δ" means the products were purified and identified by NMR; "∗" represents new compounds. The conversion rates were calculated by HPLC peak area ratio (SI). "▽" represents the glycosylation products were identified by comparing with reference standards. The reaction mixtures were incubated at 37 °C for 2h.

46x26mm (600 x 600 DPI)

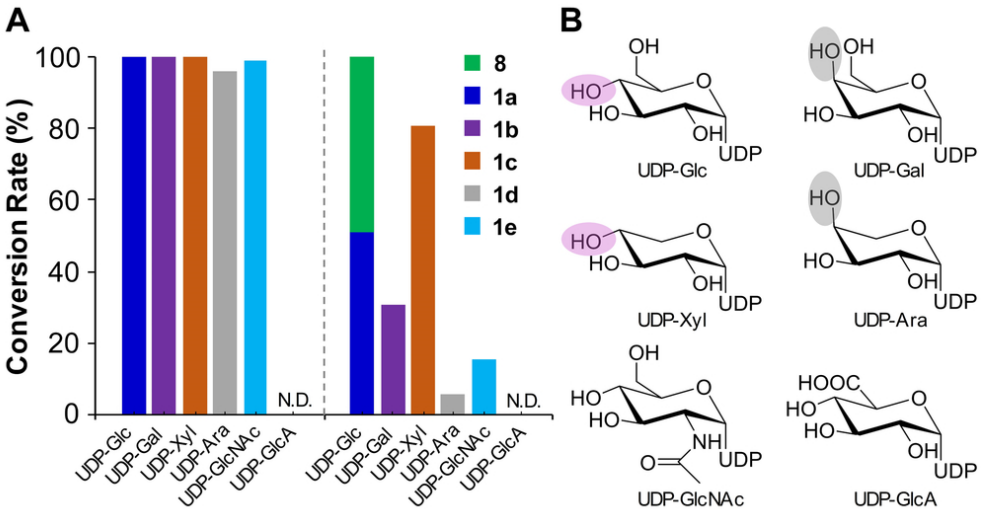


Figure 3. Probing sugar donor promiscuity of GgCGT. (A) Relative conversion rates of phloretin (**1**) using different sugar donors at high enzyme concentration (left, 200 μg/mL, 2h) and low enzyme concentration (right, 30 μg/mL, 0.5h). (B) Structures of sugar donors. The structures of **1b**-**1e** are shown in Figure S6.

42x23mm (600 x 600 DPI)

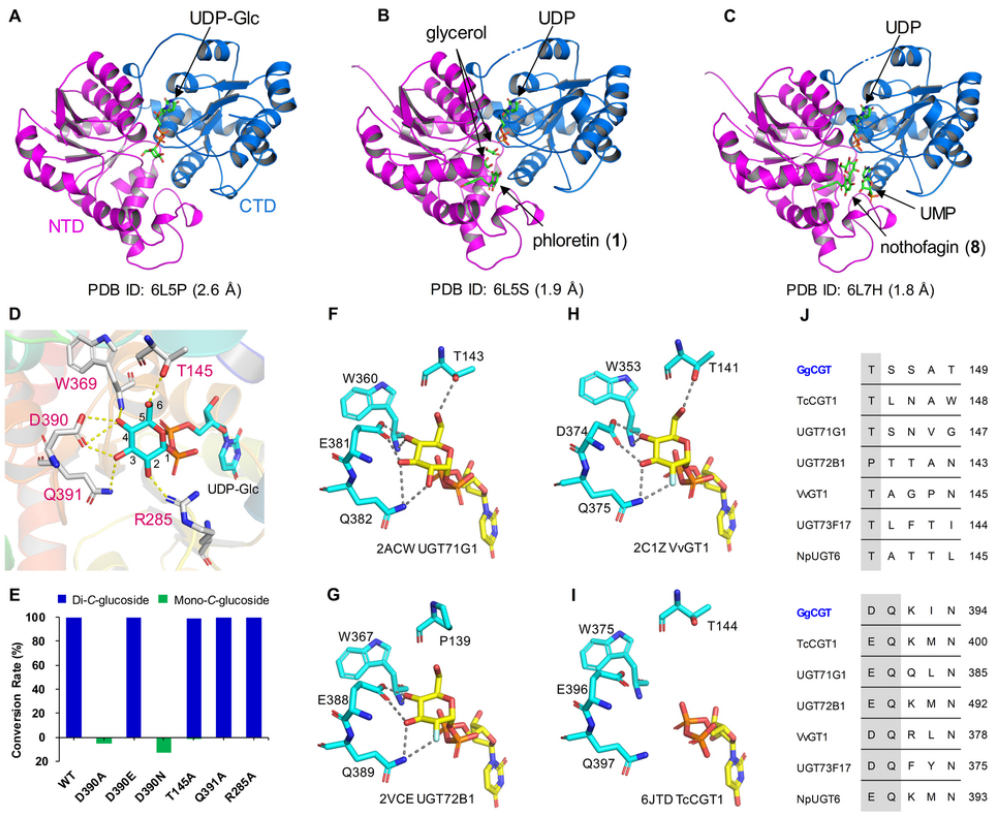


Figure 4. The crystal structures of GgCGT and its sugar donor selectivity analysis. (A) The crystal structures of GgCGT/UDP-Glc, highlighting an NTD domain (purple) and a CTD domain (blue). The sugar donor is shown as green sticks. (B/C) The crystal structures of GgCGT/UDP/phloretin (**1**) and GgCGT/UDP/nothofagin (**8**). (D) The hydrogen-bond interactions of the Glc moiety (yellow dashes) with surrounding residues of GgCGT. (E) The *C*-glycosylation conversion rates of wild-type (WT) GgCGT and mutants, using phloretin (**1**) and UDP-Glc as the sugar acceptor and donor, respectively. The reaction mixtures were incubated at 37 °C for 2h. (F-I) Other structures of UGTs with conserved residues contributing sugar recognition. The yellow sticks represent UDP-Glc molecules (UDP for TcCGT1), and the cyan sticks represent the conserved binding residues. The hydrogen-bond interactions are shown as grey dashes. (J) Sequence alignment of GgCGT with other UGTs, and the conserved residues for sugar donor binding are labeled in grey shade.

40x33mm (600 x 600 DPI)

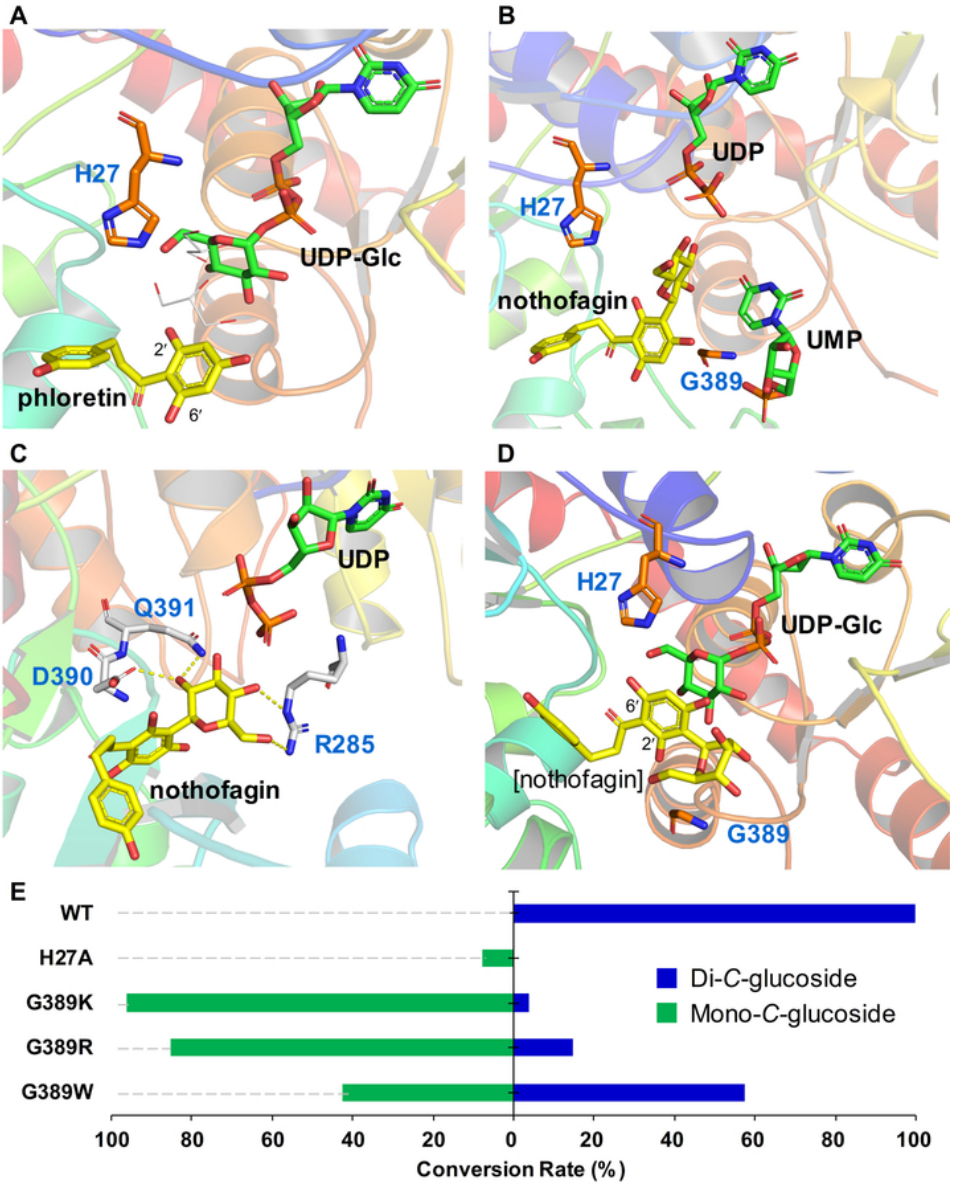
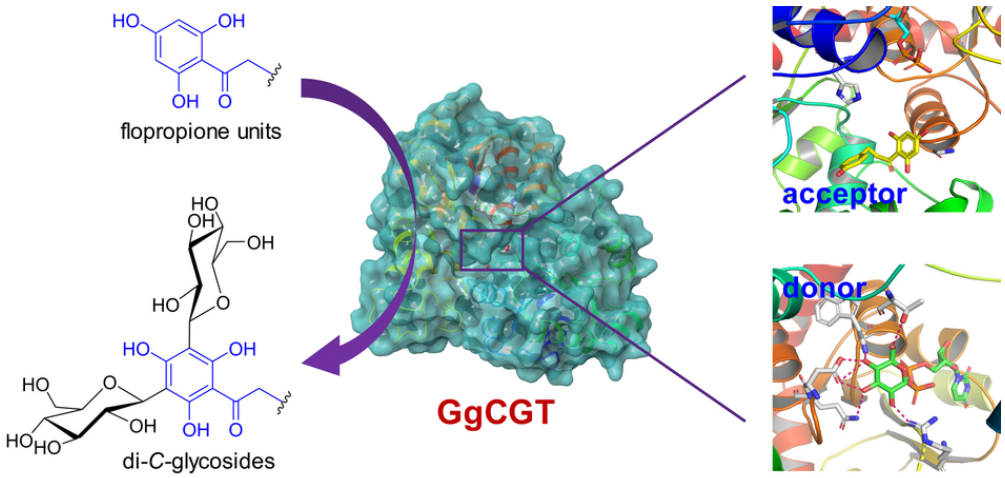


Figure 5. Sugar acceptor binding analysis. (A) Structural analysis of the *C*-glycosylation through structure alignment of GgCGT/UDP-Glc and GgCGT/UDP/phloretin. Glycerol is shown as thin sticks (grey). (B) Structural analysis of GgCGT/UDP/nothofagin. (C) Hydrogen-bond interactions of the sugar moiety of nothofagin with surrounding residues. (D) Molecular docking with nothofagin (**8**) (yellow sticks) into the substrate binding pocket of GgCGT. (E) Conversion rates of wild-type (WT) GgCGT and its mutants, using **1** and UDP-Glc as the sugar acceptor and donor, respectively. Labels with brackets indicate predicted (by molecular docking) models of the small molecule. The reaction mixtures were incubated at 37 °C for 2h.

30x38mm (600 x 600 DPI)



40x19mm (600 x 600 DPI)