A Journal of the Gesellschaft Deutscher Chemiker
Angewandte
GDCh
International Edition
Chemie
www.angewandte.org

Accepted Article

Title: Molecular Characterization and Structural Basis of a Promiscuous C-Glycosyltransferase from Trollius chinensis

Authors: Jun-Bin He, Peng Zhao, Zhi-Min Hu, Shuang Liu, Yi Kuang, Meng Zhang, Bin Li, Cai-Hong Yun, Xue Qiao, and Min Ye

To be cited as: *Angew. Chem. Int. Ed.* 10.1002/anie.201905505
Angew. Chem. 10.1002/ange.201905505

Link to VoR: http://dx.doi.org/10.1002/anie.201905505
http://dx.doi.org/10.1002/ange.201905505

WILEY-VCH

10.1002/anie.201905505

WILEY-VCH

RESEARCH ARTICLE

Molecular Characterization and Structural Basis of a Promiscuous *C*-Glycosyltransferase from *Trollius chinensis*

Jun-Bin He,[†,a] Peng Zhao,[†,b] Zhi-Min Hu,[a] Shuang Liu,[a] Yi Kuang,[a] Meng Zhang,[a] Bin Li,[a] Cai-Hong Yun,[*b] Xue Qiao,[*a] and Min Ye[*a]

Abstract: In this work, we explored the catalytic promiscuity of TcCGT1, a new *C*-glycosyltransferase (CGT) from the medicinal plant *Trollius chinensis*. TcCGT1 could efficiently and regio-specifically catalyze 8-*C*-glycosylation of 36 flavones and other flavonoids, and could also catalyze the *O*-glycosylation of diverse phenolics. Moreover, the crystal structure of TcCGT1 in complex with uridine diphosphate was determined at 1.85 Å resolution. Structural analysis with molecular docking revealed a new model for catalytic mechanism of TcCGT1, which was initiated by substrate spontaneous deprotonation. The spacious binding pocket explains the robust substrate promiscuity, and binding pose of the substrate determines *C*- or *O*-glycosylation activity. Site-directed mutagenesis at two residues (I94E and G284K) switched *C*- to *O*-glycosylation. This work highlights TcCGT1 as the first plant CGT with a crystal structure and the first flavone 8-*C*-glycosyltransferase, and provides a basis for protein engineering to design efficient glycosylation biocatalysts for drug discovery.

Introduction

Glycosides are a big class of bioactive natural products.[1] Compared with other types of glycosides (*O*-, *N*-, or *S*-glycosides), *C*-glycosides are stable to enzymatic or chemical hydrolysis due to the rigid C–C bond between the sugar residue and the aglycone.[1,2] Flavonoid *C*-glycosides, which are widely present in plants, exhibit significant benefits to human health.[3] Typical examples include vitexin (apigenin 8-*C*-*β*-D-glucoside, **1a**) and orientin (luteolin 8-*C*-*β*-D-glucoside, **2a**), which show anti-oxidant, anti-cancer, anti-inflammatory, antiviral, antimicrobial, and anti-diabetic activities.[3,4] Different strategies to synthesize flavonoid *C*-glycosides have been explored.[5] Chemical synthesis is usually restricted by low yield, poor selectivity, and multi-steps of protection and deprotection of functional groups.[6] In contrast, *C*-glycosylation reactions mediated by enzymes, *i.e.* *C*-glycosyltransferases (CGTs), show high efficiency and region-specificity.[7]

Recently, the enzymatic synthesis of flavonoid *C*-glycosides by plant CGTs has attracted considerable interest (Scheme S1).[7,8] The 2-hydroxyflavanone CGTs, including OsCGT from *Oryza sativa*,[8a] FeCGTa and FeCGTb from *Fagopyrum esculentum*,[8d] UGT708D1 from *Glycine max*,[8f] FcCGT from *Fortunella crassifolia*, and CuCGT from *Citrus unshiu*,[8i] utilize the open-circular form of 2-hydroxyflavanones as substrates to produce flavone *C*-glycosides after dehydration. The bifunctional glycosyltransferase (GT) UGT708A6 from *Zea mays* uses the close-circular form of 2-hydroxyflavanones as substrates.[8c] Moreover, GtUF6CGT1 from *Gentiana triflora* could directly catalyze the 6-*C*-glycosylation of flavones, and PlUGT43 from *Pueraria lobata* catalyzes the 8-*C*-glycosylation of isoflavones.[8e,8h] However, no CGTs have been discovered, so far, to catalyze the 8-*C*-glycosylation of flavones to produce important bioactive natural products like **1a** and **2a**. Moreover, the known CGTs exhibit relatively low substrate promiscuity and poor catalytic efficiency, which limit their application in the synthesis of structurally diverse *C*-glycosides. MiCGT recently reported from *Mangifera indica* could accept a variety of natural and unnatural substrates, but majority of them feature a 2,4,6-trihydroxybenzophenone-like core structure.[7c,9] Thus, it is critical to mine novel CGTs with catalytic promiscuity to improve structural diversity of natural products for drug discovery.

The catalytic mechanisms of *O*-glycosyltransferases (OGTs) and bacterial CGTs have been extensively studied on the basis of crystal structures.[10] For instance, a highly conserved catalytic dyad (His-Asp) arrangement of active-site residues are essential for glycosylation activities of plant OGTs.[10a-f] In the case of bacterial CGTs, Asp137 in UrdGT2, or Asp58 and Glu316 in SsfS6 serves as a catalytic base to accept an aromatic proton of the acceptors, thereby facilitating the nucleophilic attack at the sugar anomeric carbon to form a C–C bond.[10g,10h] However, no crystal structures have been reported for plant CGTs, so far. A few groups have tried to elucidate their catalytic mechanisms based on protein modeling and site-directed mutagenesis. Hirade *et al*. reported that Asp85 and Arg292 located in the active site of UGT708D1 were critical for the *C*-glucosylation activity.[8f] Chen *et al*. proposed that Ile152 was the critical amino acid residue for di-*C*-glycosylation of MiCGTb.[11] Gutmann and Nidetzky found the exchange of active-site motif (Ile-Asp or Asp-Ile) in PcOGT and OsCGT could achieve interconversion of *O*- and *C*-glycosylation.[12] Nevertheless, crystal structure information would remarkably improve understanding on the glycosylation mechanisms of plant CGTs.

Trollius chinensis Bunge (Ranunculaceae) is an endemic plant in China. Its flowers are widely used in traditional Chinese

[a] J. -B. He, Z. -M. Hu, S. Liu, Y. Kuang, M. Zhang, B. Li, Dr. X. Qiao,* Prof. M. 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
E-mail: qiaoxue@bjmu.edu.cn; yemin@bjmu.edu.cn
[b] Dr. P. Zhao, Prof. C. -H. Yun*
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
E-mail: yunch@hsc.pku.edu.cn
[†] These authors contributed equally to this work.

Supporting information, including detailed description of the experimental methods and additional experimental data, for this article is given via a link at the end of the document.

10.1002/anie.201905505

medicine due to significant antiviral, antibacterial, anti-inflammatory, antioxidant, and anticancer activities.[13] Flavone *C*-glycosides including vitexin (**1a**) and orientin (**2a**) are its major bioactive compounds,[14] indicating the existence of CGTs in this plant. Nevertheless, no CGTs have been reported to synthesize flavone 8-*C*-glycosides.[15] Herein, we report a novel *C*-glycosyltransferase TcCGT1 from *T. chinensis*, which represents the first CGT to catalyze 8-*C*-glycosylation of flavones. We also elucidated the crystal structure of TcCGT1, and interpreted structural mechanisms for its catalytic promiscuity. Moreover, structure-guided mutagenesis was conducted to alter the *C*-/*O*-glycosylation catalytic specificity.

Results and Discussion

Molecular Cloning and Functional Characterization of TcCGT1

We identified TcCGT1, a putative flavone 8-*C*-glycosyltransferase from the transcriptome (BioProject accession number PRJNA532685) of *Trollius chinensis* based on RNA sequencing and bioinformatics analysis. A phylogenetic tree was constructed to analyze the relationship of TcCGT1 with known CGTs. As a result, TcCGT1 was grouped together with the isoflavone 8-*C*-glycosyltransferase PlUGT43, though it shared very low sequence identity (25%–30% amino acid identity) with all reported plant CGTs (Figure S1). Similar to PlUGT43 and the flavone 6-*C*-glycosyltransferase GtUF6CGT1, TcCGT1 does not contain the conserved “DPFXL” motif for 2-hydroxyflavanone CGTs.[8] These evidences suggested TcCGT1 may be a CGT which could directly use flavonoids as substrates. To confirm this presumption, we cloned the gene from *T. chinensis* by RT-PCR (Table S1). The cDNA sequence of *TcCGT1* (GenBank accession number MK644229, Table S2) contains an open reading frame (ORF) of 1452 bp encoding 483 amino acids. Recombinant TcCGT1 was subsequently expressed in *Escherichia coli* and purified by His-tag affinity chromatography (purity > 95%, Figure S2).

To characterize the catalytic function of TcCGT1 *in vitro*, apigenin (**1**) and luteolin (**2**), the proposed biosynthetic precursors,[8e,13a] were used as sugar acceptors, and uridine 5'-diphosphate glucose (UDP-Glc) was used as the sugar donor. The reaction mixtures (50 mM pH 8.0 NaH_2PO_4-Na_2HPO_4; 0.5 mM UDP-Glc; 0.2 mM aglycone; 50 μg of purified recombinant TcCGT1; 30 °C, 12 h) were analyzed by liquid chromatography coupled with mass spectrometry (LC/MS). Heat-inactivated enzymes (100 °C, 15 min) were used as the negative control. Convincingly, TcCGT1 exhibited high *C*-glycosylation activity toward **1** and **2** (up to 100% conversion by HPLC analysis). In the reaction mixture of **1**, a new product **1a** was observed (Figure 1). Its $[M-H]^-$ ion appeared at *m*/*z* 431, which was 162 amu greater than **1**. The MS/MS spectrum showed fragment ions at *m*/*z* 341 $[M-H-90]^-$ and *m*/*z* 311 $[M-H-120]^-$, which were characteristic for *C*-glycosides.[16] The structure of **1a** was fully identified as vitexin by comparing with an authentic reference standard. Likewise, **2a** was identified as orientin. These results unequivocally established TcCGT1 as a CGT that catalyzes the *C*-glycosylation at C-8 of flavones. To our best knowledge, TcCGT1 is the first identified flavone 8-*C*-glycosyltransferase.

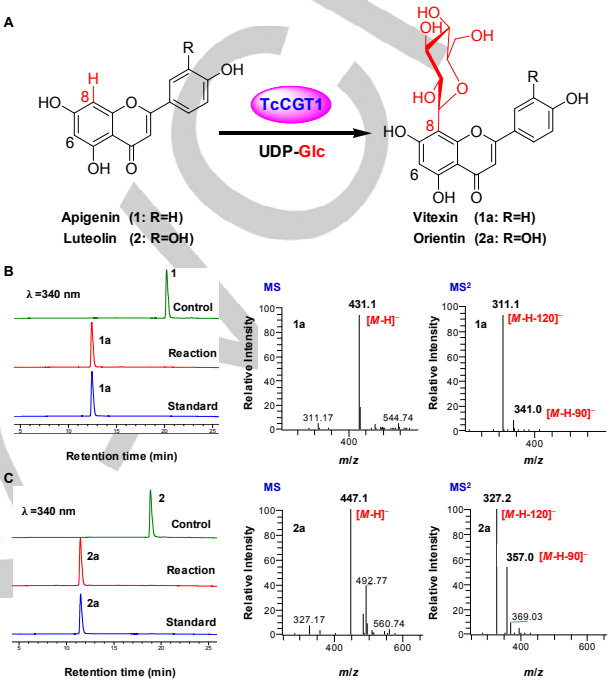


Figure 1. *C*-glycosylation of **1** and **2** catalyzed by recombinant TcCGT1. A) TcCGT1 catalyzed the *C*-glycosylation of **1** and **2**. B) LC/MS analysis of **1** and the enzymatic product **1a**. C) LC/MS analysis of **2** and the enzymatic product **2a**. The HPLC and LC/MS parameters are given in Table S3.

The biochemical characteristics of recombinant TcCGT1 were investigated using **1** as the acceptor and UDP-Glc as the sugar donor. TcCGT1 showed its maximum activity at pH 8.0 (50 mM Na_2HPO_4-NaH_2PO_4 buffer) and 30 °C, and was independent of divalent cations (Figure S3). Kinetic analysis demonstrated that TcCGT1 exhibited K_m values of 9.0 μM and 42.3 μM for **1** and UDP-Glc, respectively, and the corresponding k_{cat} values were 1.1 s^{-1} and 0.4 s^{-1}. The K_m values for **2** and UDP-Glc were 11.8 μM and 43.1 μM, and the corresponding k_{cat} values were 1.2 s^{-1} and 0.1 s^{-1}, respectively (Figure S4). These values indicated high affinity of TcCGT1 toward the flavones.

The Catalytic Promiscuity and Synthetic Applicability of TcCGT1

To explore the catalytic promiscuity of TcCGT1, an acceptor library of 114 structurally diverse substrates was tested with UDP-Glc as the sugar donor (Figure 2, Figures S5–S72). The substrates include flavonoids (**1**–**36**, **38**–**70**, **84**–**104**), hydroxynaphthalenes (**37**, **71**–**73**), lignans (**74** and **75**), stilbenes

10.1002/anie.201905505

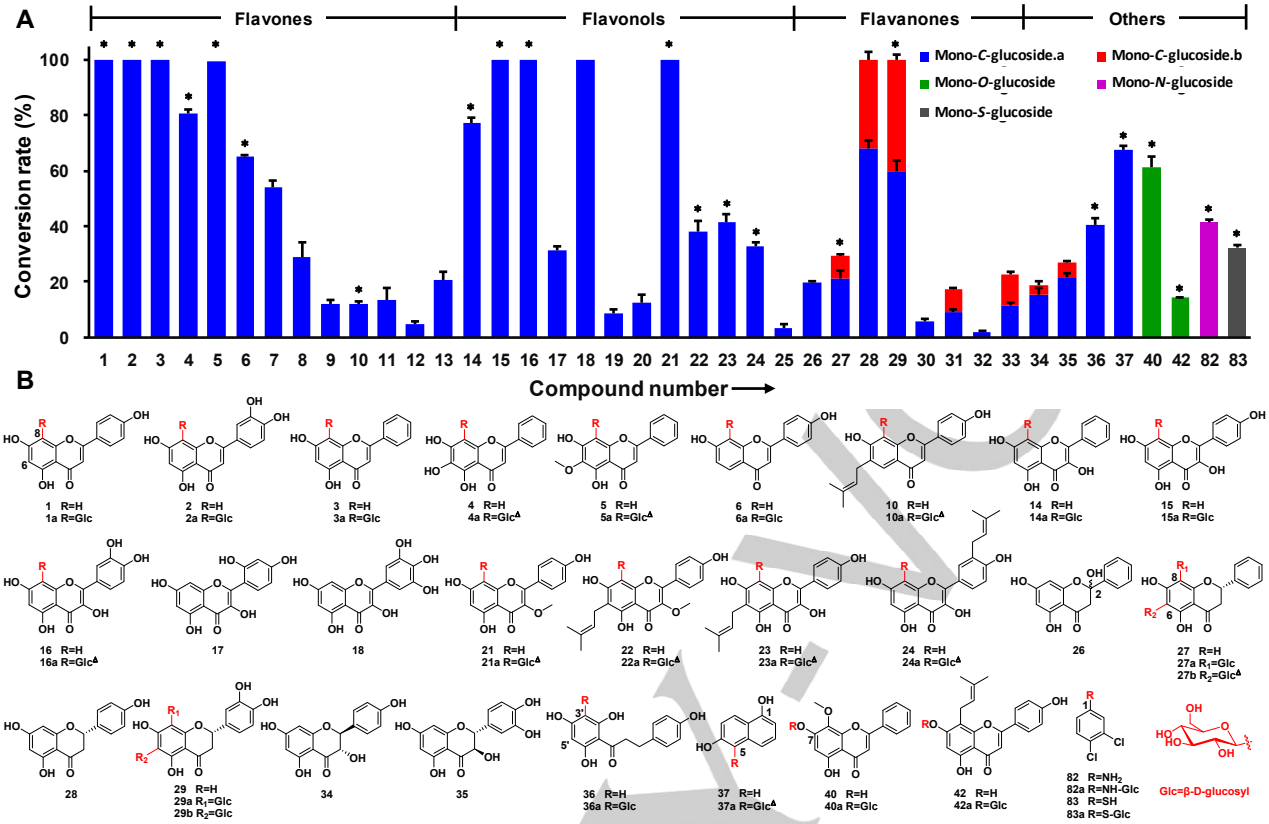


Figure 2. Catalytic promiscuity of TcCGT1. A) Conversion rates of different substrates. Experiments were performed in triplicate (n = 3). B) Structures of part of the substrates and corresponding glycosylated products. * indicates the products were isolated and confirmed by NMR spectroscopy, except that **1a**, **2a**, **36a**, **82a** and **83a** were confirmed by comparing with reference standards. △ represents new compounds. Structures of other substrates are given in Figure S5. The *O*-glucosylation conversion rates are shown in Figure S6.

(**76** and **77**), anthraquinone (**78**), benzophenone (**79**), curcumin (**80**), coumarins (**105** and **106**), triterpenoids (**107** and **108**), and simple aromatic compounds with –OH, –SH, or –NH_2 groups (**81**–**83** and **109**–**114**). LC/MS analysis revealed that TcCGT1 could catalyze the glucosylation of 83 substrates (**1**–**83**, Tables S3 and S4).

Surprisingly, TcCGT1 showed unprecedented substrate promiscuity of *C*-glycosylation. It could catalyze 36 flavonoids of different structural types, including flavones (**1**–**13**), flavonols (**14**–**25**), 2-hydroxyflavanone (**26**), flavanones (**27**–**33**), flavanonols (**34** and **35**), and dihydrochalcone (**36**). The products were identified as *C*-glycosides according to the diagnostic fragment ions $[M-H-90]^-$ and $[M-H-120]^-$ in the MS/MS spectra.[16] For 12 substrates (**1**–**5**, **14**–**16**, **18**, **21**, **28** and **29**), the conversion rates were > 77%. It is noteworthy that TcCGT1 showed high catalytic capabilities towards flavones, flavonols, and flavanones (up to 100% conversion). TcCGT1 could also catalyze hydroxynaphthalene (**37**), which is the first example of enzymatic *C*-glycosylation of this structural type. To our knowledge, most known CGTs exhibit relatively narrow substrate promiscuity.[8] Although two benzophenone CGTs, MiCGT and MiCGTb from *M. indica*, could accept a variety of substrates, majority of them feature a 2,4,6-trihydroxybenzophenone-like structure fragment.[7c,9] Thus, TcCGT1 exhibited broader substrate promiscuity than previously reported CGTs.

In addition to *C*-glycosylation, TcCGT1 also showed robust capabilities of *O*-glycosylation. It could catalyze the *O*-glucosylation of 44 substrates (**38**–**81**), including chromone (**38**), flavonoids (**39**–**70**), hydroxynaphthalenes (**71**–**73**), lignans (**74** and **75**), stilbenes (**76** and **77**), anthraquinone (**78**), benzophenone (**79**), curcumin (**80**), and simple phenolics (**81**). The conversion rates were >80% for 8 substrates (Figure S6). The products were identified as *O*-glucosides according to the diagnostic $[M-H-162]^-$ fragment ions in the MS/MS spectra. The substrate promiscuity of TcCGT1 to catalyze *O*-glycosylation was as broad as many versatile OGTs.[17] Interestingly, TcCGT1 exhibited both *C*- and *O*-glycosylation activities towards 17 substrates (**7**–**10**, **12**, **13**, **19**, **20**, **22**–**24**, **27**, **30**–**33**, **36**).

TcCGT1 also possessed *N*- and *S*-glycosylation activities toward 3,4-dichloroaniline (**82**) and 3,4-dichlorobenzenethiol (**83**), respectively. MiCGT had been reported to show *N*- but not *S*-glycosylation activity.[7c] Thus, TcCGT1 may be the first

10.1002/anie.201905505

glycosyltransferase to catalyze all the four types of glycosylation reactions, *i.e*. *C*-, *O*-, *N*-, and *S*-glycosylation.

To explore the sugar donor promiscuity of TcCGT1, we tested five other donors aside from UDP-Glc, *i.e*. UDP-xylose (UDP-Xyl), UDP-galactose (UDP-Gal), UDP-arabinose (UDP-Ara), UDP-glucuronic acid (UDP-GluA), and UDP-*N*-acetylglucosamine (UDP-GlcNAc). Compounds **1** and **2** were used as substrates. The results demonstrated that TcCGT1 could accept UDP-Xyl, UDP-Gal, and UDP-Ara, though the conversion rates for UDP-Glc and UDP-Xyl were relatively high (Figures S8, S73 and S74).

To fully identify structures of the products, we isolated 18 *C*-glycosides (**3a**–**6a**, **10a**, **14a**–**16a**, **21a**–**24a**, **27a**, **27b**, **29a**, **29b**, **37a**, **1b**) from preparative-scale enzymatic reactions. All the products are glucosides except that **1b** is a galactoside. Among them, 10 products are new compounds (**4a**, **5a**, **10a**, **16a**, **21a**–**24a**, **27b**, and **37a**). Their structures were established by HR-ESI-MS, together with 1D and 2D NMR spectroscopic analyses (Figures S75–S164). According to the HMBC spectra, the sugar moieties were attached at C-8 of flavones (**1a**–**6a**, **10a** and **1b**) and flavonols (**14a**–**16a**, **21a**–**24a**), or C-3' of dihydrochalcone (**36a**), indicating high regio-specificity of TcCGT1.[8a,15,18] Interestingly, the regio-specificity was compromised for flavanones, where TcCGT1 could generate two mono-*C*-glucosides. As exemplified by **27a**/**27b** and **29a**/**29b**, the sugar moieties were attached at C-8 and C-6, respectively. For all the obtained *C*-glycosides, the glycosidic bonds were in the *β*-configuration, according to the large coupling constants (J = 9.4−9.9 Hz) of the anomeric protons (Table S5).[7c,8a,18] Moreover, three *O*-glycosylated products (**40a**, **42a** and **72a**, Figures S165–S172) were also obtained by preparative-scale reactions.

Given the significant anti-inflammatory activities of *T. chinensis* in traditional Chinese medicine clinical practice, we evaluated compounds **1a**, **2a**, and other flavonoid 8-*C*-glycosides obtained in this study for their inhibitory activities against nuclear factor-kappa B (NF-*κ*B) and cyclooxygenase 2 (COX-2).[19,20] A number of *C*-glycosides showed similar or even higher NF-*κ*B inhibition activities than the corresponding substrates (Figure S173). Among them, six compounds (**1a**, **1b**, **2a**, **14a**, **27a** and **29b**) inhibited NF-*κ*B by > 33% at 10 μM (the positive control MG132 showed an inhibition rate of 55% at 10 μM). Moreover, seven compounds (**2a**, **14a**, **16a**, **23a**, **24a**, **29b** and **37a**) exhibited potent COX-2 inhibitory activities (> 80% at 10 μM).

Crystal Structure of TcCGT1/UDP Complex

To understand structural basis for the catalytic mechanisms and substrate promiscuity of TcCGT1, we obtained its crystal structure at 1.85Å resolution (PDB ID: 6JTD), in the presence of UDP-Glc (Table S6). Although the electron density of nearly all the protein residues were well defined, the glucose moiety was invisible in the crystal structure (Figures S174 and S175), probably because TcCGT1 bound UDP endogenously, or UDP-Glc was hydrolyzed spontaneously during crystallization. The structure was eventually identified as a TcCGT1/UDP complex. This phenomenon was similar to the ubiquitous situation in many GT structures such as UrdGT2 (PDB ID: 2P6P),[10g] UGT72B1 (PDB ID: 2VCE),[10d] UGT71G1 (PDB ID: 2ACW),[10a] UGT78G1 (PDB ID: 3HBJ),[10f] and GtfD (PDB ID: 1RRV).[21]

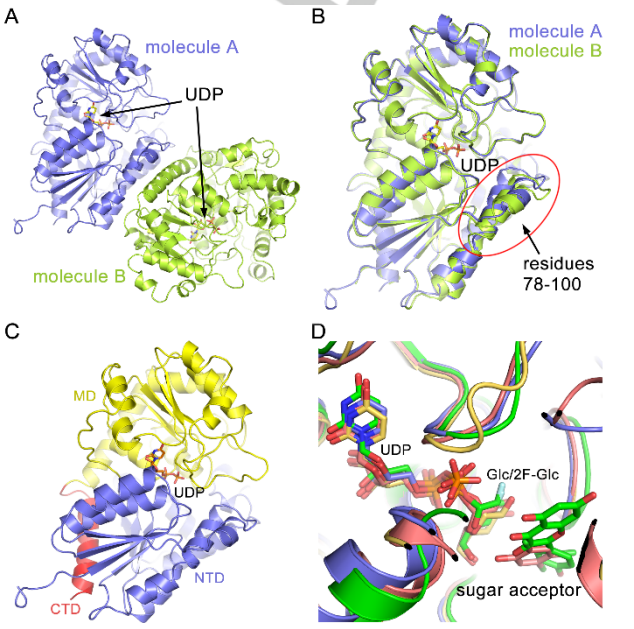


Figure 3. Crystal structure of TcCGT1 (PDB ID: 6JTD). A) The two molecules in the asymmetric unit of TcCGT1/UDP crystal structure. B) Superimposition of the two TcCGT1/UDP molecules in the asymmetric unit. C) The N-terminal domain (NTD), middle domain (MD), and C-terminal domain of TcCGT1. D) The TcCGT1/UDP-Glc model was constructed based on superimposition of MD/UDP in TcCGT1/UDP structure (slate) with those of *A. thaliana* UGT72B1 (salmon, PDB ID: 2VCE), *M. truncatula* UGT71G1 (yellow, PDB ID: 2ACW), and *V. vinifera* VvGT1 (green, PDB ID: 2C1Z). The protein molecules are shown as cartoons. UDP, UDP-sugar, and the sugar acceptor compounds are shown as sticks.

The crystal structure, in the asymmetric unit, contains two TcCGT1/UDP molecules. The two molecules are highly similar to each other, with root-mean-square deviation (RMSD) of 0.328 Å for all atoms of 471 residues. Remarkable conformational difference was observed only between residues 78–100, which participate in the formation of the putative entry channel for the sugar acceptor and thus may represent intrinsic flexibility of this region needed in entry/exit of the substrate/product (Figure 3A and 3B). TcCGT1 exhibits a typical GT-B fold structure consisting of two lobes (Figure 3C). The N-terminal domain (NTD, residues 1–251) and C-terminal domain (CTD, residues 461-483) forms one lobe (NC) responsible mainly for sugar acceptor binding, while the middle domain (MD, residues 252–460) is responsible for UDP-sugar binding. Both lobes adopt the “Rossmann-like (*β*/*α*/*β*) fold”, a conserved domain in plant glycosyltransferases,[10,22] though they do not resemble each other. The two lobes pack tightly to form a deep cleft which acts as the binding site of the sugar acceptor substrate. Sequence

10.1002/anie.201905505

WILEY-VCH

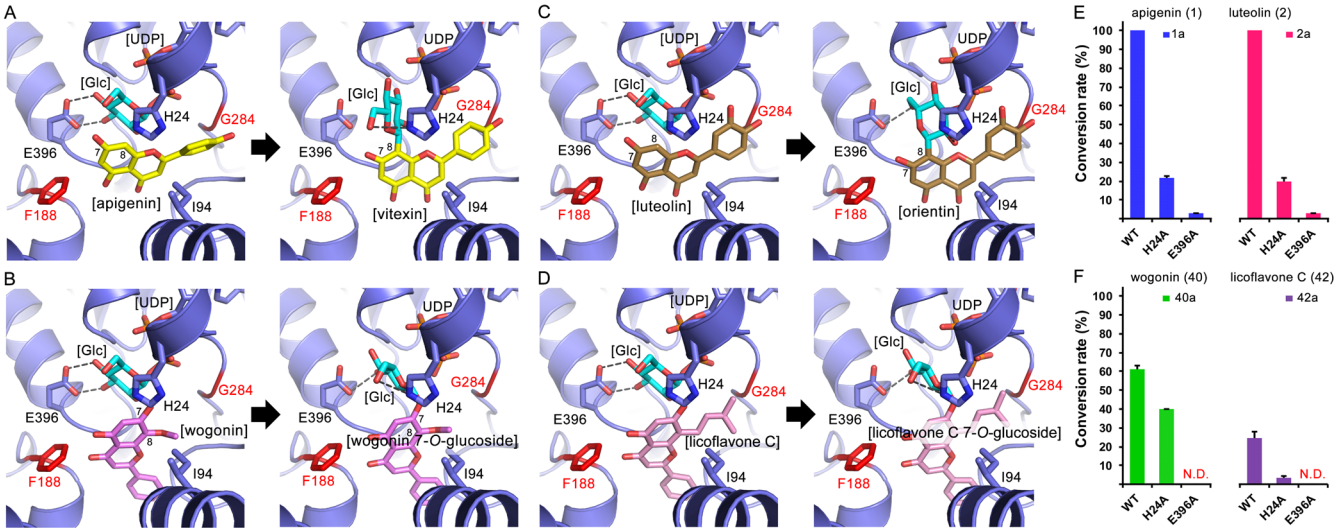


Figure 4. Molecular docking and mutagenesis studies on TcCGT1. A-D) Predicted binding modes of selected substrates and their glycosylated products to TcCGT1/UDP-Glc and TcCGT1/UDP by docking. Apigenin (**1**, A, left), vitexin (**1a**, A, right), wogonin (**40**, B, left), wogonin 7-*O*-glucoside (**40a**, B, right), luteolin (**2**, C, left), orientin (**2a**, C, right), licoflavone C (**42**, D, left) and its 7-*O*-glucoside (**42a**, D, right). The protein molecules are shown as cartoons. The key residues and substrate molecules are shown as sticks. Labels with and without brackets indicate predicted (by molecular docking) and experimentally determined models of the small molecules, respectively. Dashes indicate hydrogen bonds. E) Catalytic activities of wild-type (WT) TcCGT1 and its mutants (H24A and E396A) towards substrates **1** and **2**. F) Catalytic activities of wild-type (WT) TcCGT1 and its mutants (H24A and E396A) towards substrates **40** and **42**. N.D., products not detected (Figure S178).

alignment (Figure S176) and three-dimensional structure superimposition (Figure 3D and Figure S177) revealed that although TcCGT1 is a robust CGT, it does not show significant similarity either in primary sequence or in tertiary structure to the bacterial CGTs UrdGT2 (2P6P, sequence identity about 16% and RMSD about 13 Å) or SsfS6 (4G2T, sequence identity about 16% and RMSD about 17 Å).[10g,10h,23] Interestingly, TcCGT1 shares high homology with other plant glycosyltransferases such as UGT72B1 (2VCE, sequence identity about 40% and RMSD about 1.243Å) which is a bifunctional *N*- and *O*-glucosyltransferase (NGT and OGT) from *Arabidopsis thaliana*,[10d] UGT71G1 (2ACW, sequence identity about 27% and RMSD about 1.661 Å) which is a multifunctional triterpene/flavonoid OGT from *Medicago truncatula*,[10a] and VvGT1 (2C1Z, sequence identity about 25% and RMSD about 1.854 Å) which is a flavonoid 3-*O*-glycosyltransferase from *Vitis vinifera*.[10b] These structural comparisons indicate the catalytic mechanism of TcCGT1 may be different from bacterial CGTs.

Catalytic Mechanisms for *C*- and *O*-Glycosylation of TcCGT1

TcCGT1 exhibits both *C*- and *O*-glycosylation activities towards various flavonoids. In order to investigate the catalytic mechanisms of TcCGT1, we tried to obtain the complex structures of TcCGT1 bound by a substrate (such as **1**) or a product (such as **1a**). However, after a lot of trials including co-crystallization and soaking experiments, we were unable to obtain such a complex. Fortunately, since TcCGT1 shares high structural similarity to UGT72B1 (2VCE), UGT71G1 (2ACW) and VvGT1 (2C1Z), we could readily model the glucose moiety of UDP-Glc according to the glucose or analogue moieties observed in these structures (Figure 3D). We then built the models for TcCGT1/UDP-Glc/apigenin (**1**), TcCGT1/UDP-Glc/luteolin (**2**), TcCGT1/UDP-Glc/wogonin (**40**) and TcCGT1/UDP-Glc/licoflavone C (**42**) through computer aided molecular docking of the substrates **1** and **2** (two optimal *C*-glycosylation substrates of TcCGT1), **40** and **42** (two representative *O*-glycosylation substrates of TcCGT1) into the TcCGT1/UDP-Glc structure. We also built the TcCGT1/UDP/**1a**, TcCGT1/UDP/**2a**, TcCGT1/UDP/**40a** and TcCGT1/UDP/**42a** models by docking products **1a**, **2a**, **40a** and **42a** into the TcCGT1/UDP structure (Figure 4).

The molecular docking provided interesting information that may reveal catalytic mechanisms for the *C*- and *O*-glycosylation activities of TcCGT1. Substrate **1** or **2** binds to TcCGT1 in a direction in which C-8 is placed in the most proximity to C-1' of the glucose moiety of UDP-Glc, which could explain why the sugar is transferred to C-8 position of the substrate (8-*C*-glycosylation) (Figure 4A and 4C). In contrast, substrate **40** or **42** bound to TcCGT1 is placed in a different direction almost perpendicular to that of **1** or **2**. The binding pose of **40** and **42** clearly locates 7-OH close to C-1' of the sugar, which could explain why TcCGT1 catalyzes transfer of glucose to 7-OH of the substrates (7-*O*-glycosylation) (Figure 4B and 4D).

It had been proposed that deprotonation of a hydroxyl group activated by a histidine (His) residue is the key step to initiate GT-mediated glycosylation.[10a-c,10e,10f] In TcCGT1, this key residue was mapped to His24 (H24). However, in the TcCGT1/UDP-Glc/substrate docking models, H24 is located far away from any

10.1002/anie.201905505

WILEY-VCH

RESEARCH ARTICLE

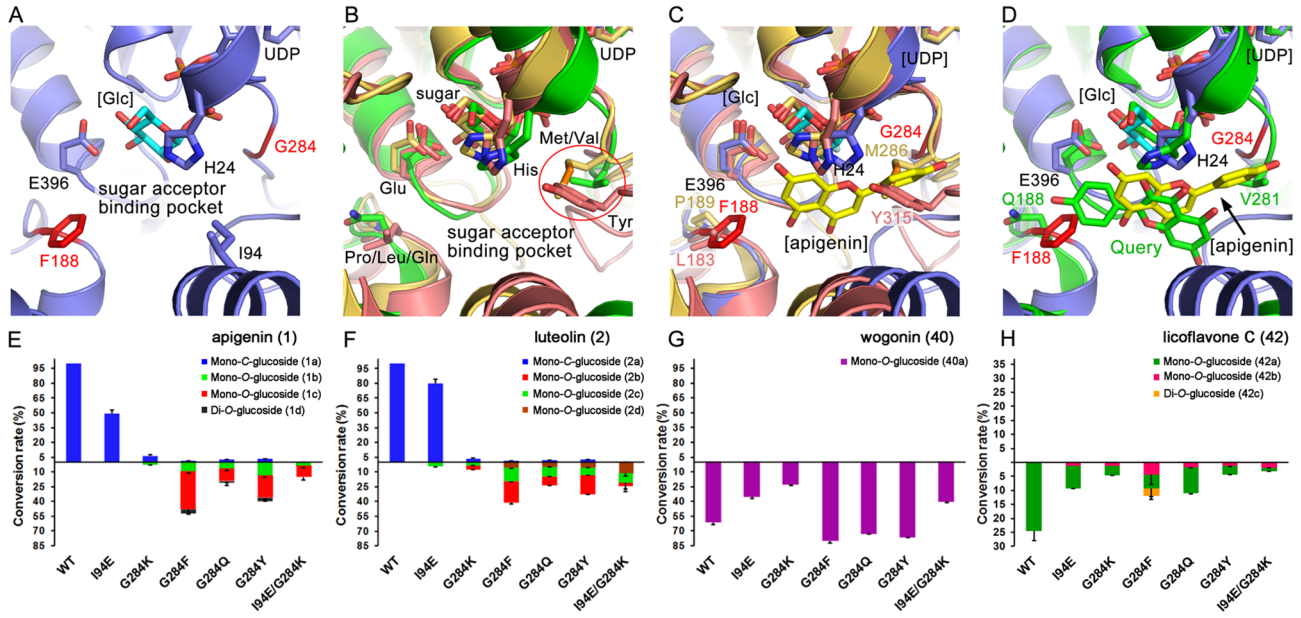


Figure 5. Comparative structural analysis and structure-guided mutagenesis of TcCGT1. A) The spacious sugar acceptor binding pocket of TcCGT1. B) The crowded (compared to TcCGT1) sugar acceptor binding pocket of OGTs, including UGT72B1 (salmon), UGT71G1 (yellow) and VvGT1 (green). C) Assuming apigenin binds to UGT72B1 (salmon) or UGT71G1 (yellow) in the same way as it binds to TcCGT1, UGT72B1 Tyr315 or UGT71G1 Met286 would clash with the substrate. D) Assuming apigenin binds to VvGT1 (green) in the same way as it binds to TcCGT1, VvGT1 Val281 would clash with the compound. The protein molecules are shown as cartoons. The key residues and substrate molecules are shown as sticks. Labels with and without brackets indicate predicted (by molecular docking) and experimentally determined models of the small molecules, respectively. E–H) Catalytic activities of wild-type (WT) TcCGT1 and its mutants towards substrates **1**, **2**, **40** and **42** (Figures S180–S183). **1a**: vitexin (apigenin 8-*C*-glucoside), **1b**: apigenin 7-*O*-glucoside, **1c**: apigenin 4'-*O*-glucoside, **1d**: apigenin 7,4'-di-*O*-glucoside. **2a**: orientin (luteolin 8-*C*-glucoside). **40a**: wogonin 7-*O*-glucoside. **42a**: licoflavone C 7-*O*-glucoside.

hydroxyl group of the sugar acceptor (including 7-OH), which did not well support the previous hypothesis. Interestingly, ligand preparation process with the Schördinger® GLIDE software package indicated that 7-OH, but not 5-OH of **1**, **2**, **40** or **42** would undergo spontaneous deprotonation in physiological condition. Therefore, although H24 may still act as a proton acceptor to facilitate 7-OH deprotonation, it does not seem to be indispensable. Further site-directed mutagenesis studies were conducted to test this hypothesis. Indeed, mutation of H24 to alanine (H24A) did weaken, but did not abolish the *C*- or *O*-glycosylation activities of TcCGT1 (Figure 4E, Figure 4F, and Figure S178). On the other hand, docking of the glycosylated products, *i.e.* **1a**, **2a**, **40a** or **42a** into the TcCGT1/UDP structure indicated that H24 may stabilize the products through hydrogen bonding and thus facilitate the catalytic process. Taken together, the docking studies indicated a new model for the catalytic mechanism of TcCGT1 which includes the following steps: 1) spontaneous deprotonation of 7-OH of the flavone substrate results in negative charge on 7-O or C-8 atom (due to electron rearrangement on the aryl ring), and H24 side-chain imidazole may stabilize the deprotonated substrate; 2) the negatively charged 7-O or C-8 atom, whichever is physically closer, would attack C-1' of the sugar to fulfill the reaction; 3) H24 side-chain imidazole interacts with and stabilizes the products through hydrogen bonding with sugar hydroxyls of the products to facilitate the reaction (Scheme S2).

Moreover, in the docking models of TcCGT1/UDP-Glc/substrate, Glu396 (E396) was very close to the sugar to form hydrogen bonds with the sugar hydroxyls. This residue was considered a key residue to stabilize the donor substrate and position the sugar in the optimal orientation for the glycosylation reaction. Mutation of E396 to alanine (E396A) indeed drastically decreased or even abolished the catalytic activity of TcCGT1 (Figure 4E, Figure 4F, and Figure S178).

Based on the model presented above, E396 plays an important role to stabilize and orient the UDP-Glc sugar. H24 acts to stabilize both the deprotonated substrate and the product sugar, though it is not indispensable for the glycosylation activity. The substrate flavonoids may undergo spontaneous deprotonation to initiate the glycosylation reaction. Most importantly, both *C*- and *O*-glycosylation activities of TcCGT1 share the common catalytic mechanism. Whether TcCGT1 acts as CGT or OGT is determined by binding pose of the substrate, *i.e.* which of the negatively charged 8-C^- or 7-O^- is located closer to C-1' of the sugar.

The Substrate Promiscuity is Enabled by the Spacious Binding Pocket

To further explore structural basis for the substrate promiscuity of TcCGT1, we analyzed structure of the active sites at the substrate binding pocket. The methoxy or dimethylallyl group at C-8 of **40** and **42** makes it impossible for these substrates to bind

to TcCGT1 in the same mode as **1** or **2** does, due to steric hindrance with the sugar of UDP-Glc or the side-chain of H24 (Figure S179). However, **1** and **2** could presumably bind to TcCGT1 in the same mode as **40** and **42** do. Comparison of the active site pocket between TcCGT1 with those of UGT72B1, UGT71G1 and VvGT1 revealed that TcCGT1 has the most spacious sugar acceptor binding pocket. This spacious pocket can readily accommodate the 2-phenyl moieties of **1** or **2** (Figure 5A), allows **1** or **2** to bind more deeply into the enzyme than **40** or **42** does (Figure 4 and Figure S179), and thus enables the *C*-glycosylation activity. In UGT72B1, UGT71G1 and VvGT1, however, the same space is occupied by Tyr315, Met286 and Val281, respectively (Figure 5B). Taken together, the docking studies indicate the spacious sugar acceptor binding pocket of TcCGT1 enables the unique binding mode of **1** or **2** to the enzyme to enable the *C*-glycosylation activity. The spacious binding pocket also explains why TcCGT1 shows broad substrate promiscuity.

Structure-Guided Engineering of the *C*- or *O*-Glycosylation Selectivity of TcCGT1

In order to explore the determination between *C*- and *O*-glycosylation activities of TcCGT1, we tended to conduct site-directed mutagenesis of residues around the spacious sugar acceptor binding pocket to shrink it. The shrinked pocket would force typical CGT substrates of TcCGT1 (such as **1** and **2**) to bind in the way as **40** and **42** do, thus leading to *O*-glycosylation. We inspected the binding pocket of typical OGTs, and found the bulky side-chains of methionine (Met286 in UGT71G1), tyrosine (Tyr315 in UGT72B1), or valine (Val281 in VvGT1) would presumably block the binding of **1** or **2** in the CGT substrate way (Figure 5C and 5D). In TcCGT1, the corresponding residue is Gly284 (G284). We then mutated Gly284 to an amino acid with a bulky side-chain, such as Phe (G284F), Gln (G284Q), Tyr (G284Y) and Lys (G284K). We also intended to build a salt-bridge between residues I94 and G284 to occupy the pocket. Thus we prepared G284F, G284Q, G284Y, G284K and I94E single mutants or I94E/G284K double mutant of TcCGT1 to see whether these modifications can convert the activity of TcCGT1 towards **1** or **2** from CGT to OGT. Consistent with our hypothesis, all the mutations drastically decreased the CGT activity of TcCGT1 towards **1** or **2**, and some mutants showed significant *O*-glycosylation activities (Figure 5E, Figure 5F, and Figures S180–S183). In the case of the natural OGT substrate **40**, these mutations did not alter the *O*-glycosylation activity or even enhanced the *O*-glycosylation activity (Figure 5G). For substrate **42**, however, the mutations partially inhibited the *O*-glycosylation activity (Figure 5H). This observation was consistent with our model. The isoprenyl group at C-8 of **42** is bulkier than the methoxyl group of **40**. Substrate **42** protrudes more deeply into the binding pocket than the latter, and was therefore affected more by the mutations.

Conclusion

In summary, we characterized a promiscuous *C*-glycosyltransferase TcCGT1 from the medicinal plant *T. chinensis*. TcCGT1 represents the first plant CGT with a crystal structure and the first flavone 8-*C*-glycosyltransferase. It could region-specifically catalyze 8-*C*-glycosylation of flavones, flavonols, and other types of flavonoids. TcCGT1 is also the first glycosyltransferase that can catalyze *C*-, *O*-, *N*-, and *S*-glycosylation reactions. The 1.85 Å resolution crystal structure of TcCGT1/UDP complex, molecular docking, and site-directed mutagenesis revealed a new model of catalytic mechanism of TcCGT1. The substrate flavonoids undergo spontaneous deprotonation to initiate the glycosylation reaction, and the residues H24 and E396 play an important role in stabilizing and orienting the small molecules. The broad substrate promiscuity of TcCGT1 is enabled by the spacious binding pocket. The selectivity between *C*- or *O*-glycosylation activities is determined by the binding pose of the substrate at the pocket, and the mutations at I94E and G284K installed the *O*-glycosylation activity of TcCGT1 while abolishing the *C*-glycosylation activity. This study provides a basis to design effective biocatalysts for efficient and directed biosynthesis of important bioactive flavonoid *C*-glycosides for drug discovery.

Acknowledgements

The authors thank Dr. Kuan Chen at Peking University for providing the authentic reference standards **1b**–**1d**, and Dr. Kai Li for collecting the plant samples. This work was supported by National Natural Science Foundation of China (81725023, 81891010/81891011, 31270769), Beijing Natural Science Foundation (JQ18027), and Young Elite Scientists Sponsorship Program by China Association for Science and Technology (2016QNRC001).

Conflict of interest

The authors declare no conflict of interest.

Keywords: glycosylation • enzyme catalysis • glycosyltransferase • crystal structure • catalytic mechanisms

[1] a) C. J. Thibodeaux, C. E. Melancon, H. W. Liu, *Nature* **2007**, *446*, 1008-1016; b) C. J. Thibodeaux, C. E. Melancon III, H. W. Liu, *Angew. Chem. Int. Ed.* **2008**, *47*, 9814-9859; c) T. Bililign, B. R. Griffith, J. S. Thorson, *Nat. Prod. Rep.* **2005**, *22*, 742-760.

[2] P. G. Hultin, *Curr. Top. Med. Chem.* **2005**, *5*, 1299-1331.

[3] a) J. Xiao, E. Capanoglu, A. R. Jassbi, A. Miron, *Crit. Rev. Food Sci. Nutr.* **2016**, *56*, S29-S45; b) O. Talhi, A. M. S. Silva, *Curr. Org. Chem.* **2012**, *16*, 859-896.

[4] a) K. Y. Lam, A. P. K. Ling, R. Y. Koh, Y. P. Wong, Y. H. Say, *Adv. Pharmacol. Sci.* **2016**, Article ID 4104595; b) M. He, J. W. Min, W. L. Kong, X. H. He, J. X. Li, B. W. Peng, *Fitoterapia* **2016**, *115*, 74-85.

[5] S. Sato, T. Akiya, H. Nishizawa, T. Suzuki, *Carbohydr. Res.* **2006**, *341*, 964-970.

10.1002/anie.201905505

[6] a) H. Satoh, S. Manabe, *Chem. Soc. Rev.* **2013**, *42*, 4297-4309; b) Y. Yang, B. Yu, *Chem. Rev.* **2017**, *117*, 12281-12356.

[7] a) E. K. Lim, *Chem. Eur. J.* **2005**, *11*, 5486-5494; b) C. Durr, D. Hoffmeister, S. E. Wohlert, K. Ichinose, M. Weber, U. von Mulert, J. S. Thorson, A. Bechthold, *Angew. Chem. Int. Ed.* **2004**, *43*, 2962-2965; c) D. W. Chen, R. D. Chen, R. S. Wang, J. H. Li, K. B. Xie, C. C. Bian, L. L. Sun, X. L. Zhang, J. M. Liu, L. Yang, F. Ye, X. M. Yu, J. G. Dai, *Angew. Chem. Int. Ed.* **2015**, *54*, 12678-12682.

[8] a) M. Brazier-Hicks, K. M. Evans, M. C. Gershater, H. Puschmann, P. G. Steel, R. Edwards, *J. Biol. Chem.* **2009**, *284*, 17926-17934; b) M. L. Hamilton, S. P. Kuate, M. Brazier-Hicks, J. C. Caulfield, R. Rose, R. Edwards, B. Torto, J. A. Pickett, A. M. Hooper, *Phytochemistry* **2012**, *84*, 169-176; c) M. L. F. Ferreyra, E. Rodriguez, M. I. Casas, G. Labadie, E. Grotewold, P. Casati, *J. Biol. Chem.* **2013**, *288*, 31678-31688; d) Y. Nagatomo, S. Usui, T. Ito, A. Kato, M. Shimosaka, G. Taguchi, *Plant J.* **2014**, *80*, 437-448; e) N. Sasaki, Y. Nishizaki, E. Yamada, F. Tatsuzawa, T. Nakatsuka, H. Takahashi, M. Nishihara, *FEBS Lett.* **2015**, *589*, 182-187; f) Y. Hirade, N. Kotoku, K. Terasaka, Y. Saijo-Hamano, A. Fukumoto, H. Mizukami, *FEBS Lett.* **2015**, *589*, 1778-1786; g) B. Hao, J. C. Caulfield, M. L. Hamilton, J. A. Pickett, C. A. O. Midega, Z. R. Khan, J. Wang, A. M. Hooper, *Phytochemistry* **2016**, *125*, 73-87; h) X. Wang, C. F. Li, C. Zhou, J. Li, Y. S. Zhang, *Plant J.* **2017**, *90*, 535-546; i) T. Ito, S. Fujimoto, F. Suito, M. Shimosaka, G. Taguchi, *Plant J.* **2017**, *91*, 187-198.

[9] D. W. Chen, L. L. Sun, R. D. Chen, K. B. Xie, L. Yang, J. G. Dai, *Chem. Eur. J.* **2016**, *22*, 5873-5877.

[10] a) H. Shao, X. Z. He, L. Achnine, J. W. Blount, R. A. Dixon, X. Q. Wang, *Plant Cell* **2005**, *17*, 3141-3154; b) W. Offen, C. Martinez-Fleites, M. Yang, E. Kiat-Lim, B. G. Davis, C. A. Tarling, C. M. Ford, D. J. Bowles, G. J. Davies, *EMBO J.* **2006**, *25*, 1396-1405; c) L. Li, L. V. Modolo, L. L. Escamilia-Trevino, L. Achnine, R. A. Dixon, X. Wang, *J. Mol. Biol.* **2007**, *370*, 951-963; d) M. Brazier-Hicks, W. A. Offen, M. C. Gershater, T. J. Revett, E.-K. Lim, D. J. Bowles, G. J. Davies, R. Edwards, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 20238-20243; e) T. Hiromoto, E. Honjo, N. Noda, T. Tamada, K. Kazuma, M. Suzuki, M. Blaber, R. Kuroki, *Protein Sci.* **2015**, *24*, 395-407; f) L. V. Modolo, L. Li, H. Pan, J. W. Blount, R. A. Dixon, X. Wang, *J. Mol. Biol.* **2009**, *392*, 1292-1302. g) M. Mittler, A. Bechthold, G. E. Schulz, *J. Mol. Biol.* **2007**, *372*, 67-76; h) F. B. Wang, M. Q. Zhou, S. Singh, R. M. Yennamalli, C. A. Bingman, J. S. Thorson, G. N. Phillips, *Proteins* **2013**, *81*, 1277-1282; i) D. Foshag, C. Campbell, P. D. Pawelek, *BBA-Proteins Proteom.* **2014**, *1844*, 1619-1630.

[11] D. W. Chen, S. Fan, R. D. Chen, K. B. Xie, S. Yin, L. L. Sun, J. M. Liu, L. Yang, J. Q. Kong, Z. Y. Yang, J. G. Dai, *ACS Catal.* **2018**, *8*, 4917-4927.

[12] A. Gutmann, B. Nidetzky, *Angew. Chem. Int. Ed.* **2012**, *51*, 12879-12883.

[13] a) M. Yuan, R. F. Wang, X. W. Wu, Y. N. An, X. W. Yang, *Chin. J. Nat. Med.* **2013**, *11*, 449-455; b) E. Witkowska-Banaszczak, *Phytother. Res.* **2015**, *29*, 475-500.

[14] a) L. Z. Wu, X. P. Zhang, X. D. Xu, Q. X. Zheng, J. S. Yang, W. L. Ding, *J. Pharm. Biomed. Anal.* **2013**, *75*, 55-63; b) Z. L. Song, Y. Hashi, H. Y. Sun, Y. Liang, Y. X. Lan, H. Wang, S. Z. Chen, *Fitoterapia* **2013**, *91*, 272-279.

[15] a) J. X. Wei, D. Y. Li, Z. L. Li, *Phytochem. Lett.* **2018**, *25*, 156-162; b) R. Yan, Y. Cui, B. Deng, J. Bi, G. Zhang, *J. Nat. Med.* **2019**, *73*, 297-302.

[16] X. Qiao, W. N. He, C. Xiang, J. Han, L. J. Wu, D. A. Guo, M. Ye, *Phytochem. Anal.* **2011**, *22*, 475-483.

[17] a) K. Xie, R. Chen, J. Li, R. Wang, D. Chen, X. Dou, J. Dai, *Org. Lett.* **2014**, *16*, 4874-4877; b) C. S. Zhang, B. R. Griffith, Q. Fu, C. Albermann, X. Fu, I. K. Lee, L. J. Li, J. S. Thorson, *Science* **2006**, *313*, 1291-1294.

[18] N. Krafczyk, M. A. Glomb, *J. Agr. Food Chem.* **2008**, *56*, 3368-3376.

[19] J. Napetschnig, H. Wu, *Annu. Rev. Biophys.* **2013**, *42,* 443-468.

[20] M. E. Turini, R. N. DuBois, *Annu. Rev. Med.* **2002**, *53*, 35-57.

[21] A. M. Mulichak, W. Lu, H. C. Losey, C. T. Walsh, R. M. Garavito, *Biochemistry* **2004**, *43*, 5170-5180.

[22] C. Breton, S. Fournel-Gigleux, M. M. Palcic, *Curr. Opin. Struct. Biol.* **2012**, *22*, 540-549.

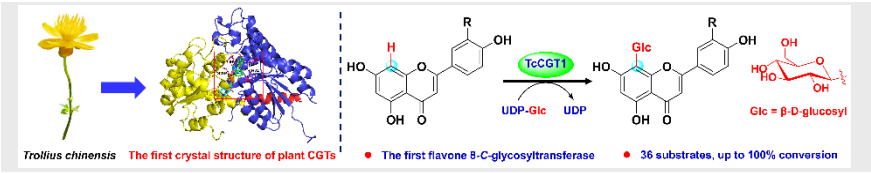
[23] X. Robert, P. Gouet, *Nucleic Acids Res.* **2014**, *42*, W320-W324.

10.1002/anie.201905505

WILEY-VCH

Entry for the Table of Contents (Please choose one layout)

RESEARCH ARTICLE



A promiscuous *C*-glycosyltransferase TcCGT1 was highlighted. TcCGT1 represents the first flavone 8-*C*-glycosyltransferase which exhibits robust substrate promiscuity toward different types of flavonoids. TcCGT1 is also the first plant CGT with a crystal structure. This work provides a basis for protein engineering to design efficient glycosylation biocatalysts for drug discovery.

Jun-Bin He,† Peng Zhao,† Zhi-Min Hu, Shuang Liu, Yi Kuang, Meng Zhang, Bin Li, Cai-Hong Yun, Xue Qiao,* and Min Ye**

Page No. – Page No.

Molecular Characterization and Structural Basis of a Promiscuous *C*-Glycosyltransferase from *Trollius chinensis*