ORIGINAL ARTICLE

Functional characterization, structural basis, and protein engineering of a rare flavonoid 2′-*O*-glycosyltransferase from *Scutellaria baicalensis*

Zilong Wang[a], Xueqing Du[b], Guo Ye[a], Haotian Wang[a], Yizhan Liu[a], Chenrui Liu[a], Fudong Li[c], Hans Ågren[d], Yang Zhou[e], Junhao Li[d,*], Chao He[b,*], De-An Guo[f,*], Min Ye[a,*]

[a]*State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing 100191, China*
[b]*Anhui Key Laboratory of Modern Biomanufacturing and School of Life Sciences, Anhui University, Hefei 230601, China*
[c]*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 230026, China*
[d]*Department of Physics and Astronomy, Uppsala University, Uppsala SE-751 20, Sweden*
[e]*School of Pharmacy, Jinan University, Guangzhou 510632, China*
[f]*Shanghai Research Center for Modernization of Traditional Chinese Medicine, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China*

Abstract Glycosylation is an important post-modification reaction in plant secondary metabolism, and contributes to structural diversity of bioactive natural products. In plants, glycosylation is usually catalyzed by UDP-glycosyltransferases. Flavonoid 2′-*O*-glycosides are rare glycosides. However, no UGTs have been reported, thus far, to specifically catalyze 2′-*O*-glycosylation of flavonoids. In this work, UGT71AP2 was identified from the medicinal plant *Scutellaria baicalensis* as the first flavonoid 2′-*O*-glycosyltransferase. It could preferentially transfer a glycosyl moiety to 2′-hydroxy of at least nine flavonoids to yield six new compounds. Some of the 2′-*O*-glycosides showed noticeable inhibitory activities against cyclooxygenase 2. The crystal structure of UGT71AP2 (2.15 Å) was solved, and mechanisms of its regio-selectivity was interpreted by pK_a calculations, molecular docking, MD simulation, MM/GBSA binding free energy, QM/MM, and hydrogen−deuterium exchange mass

---

*Corresponding authors.

E-mail addresses: junhao.li@physics.uu.se (Junhao Li), chaohe@ahu.edu.cn (Chao He), daguo@simm.ac.cn (De-An Guo), yemin@bjmu.edu.cn (Min Ye).

spectrometry analysis. Through structure-guided rational design, we obtained the L138T/V179D/M180T mutant with remarkably enhanced regio-selectivity (the ratio of 7-O-glycosylation byproducts decreased from 48% to 4%) and catalytic efficiency of 2′-O-glycosylation (k_{cat}/K_m, 0.23 L/(s·μmol), 12-fold higher than the native). Moreover, UGT71AP2 also possesses moderate UDP-dependent de-glycosylation activity, and is a dual function glycosyltransferase. This work provides an efficient biocatalyst and sets a good example for protein engineering to optimize enzyme catalytic features through rational design.

1. Introduction

Flavonoid glycosides are widely present in the plant kingdom, and represent an important class of bioactive natural products[1,2]. They exhibit various potential functions in improving human health. The regio-selective synthesis of flavonoid glycosides has been challenging, especially when the aglycone has multiple hydroxy groups[3]. In plants, the biosynthesis of glycosides is usually catalyzed by UDP-glycosyltransferases (UGTs). In the past two decades, a large number of flavonoid UGTs have been identified from plants[4-6]. They mainly catalyze glycosylation reactions at 3-OH, 5-OH, 7-OH, or 4′-OH of the flavonoid skeleton. Some of these UGTs exhibit high regio-selectivity, including the 3-O-glycosyltransferase Sb3GT1 from *Scutellaria baicalensis*[7], the 7,4′-di-O-glycosyltransferase ZjOGT3 from *Ziziphus jujuba* var. *spinosa*[8], and the 7-O-glucuronyltransferase EbF7GAT from *Erigeron breviscapus*[9]. These enzymes provide green, efficient, and regio-specific approaches to synthesize different types of flavonoid O-glycosides[10]. However, no UGTs have been reported, thus far, to specifically catalyze 2′-O-glycosylation of flavonoids. While OcUGT1 from *Ornithogalum caudatum* could catalyze 3,2′-dihydroxyflavone to form both the 3-O-glucoside (<10%) and the 2′-O-glucoside (<8%), the conversion rates were low, and the regio-selectivity was poor[11].

Protein crystal structures play a critical role in catalytic mechanism elucidation and functional evolution of enzymes. Thus far, more than twenty crystal structures of plant-derived UGTs have been reported[12]. However, little is known about the mechanisms of regio-selectivity of UGTs[13,14]. Complex crystal structure analysis revealed the regio-selectivities of UGT76G6 and UGT91C1 were related to substrate orientation in the active pocket. Nevertheless, it was difficult to enhance the regio-selectivities through site-directed mutagenesis[15-17]. Several strategies have been used for protein engineering of glycosyltransferases to alter catalytic functions, such as error-prone polymerase chain reaction, iterative saturation mutagenesis (ISM), and focused rational iterative site-specific mutagenesis (FRISM)[18,19]. These methods rely on large-scale screening and are time-consuming. Additionally, our group changed the regio-selectivity of a triterpene glycosyltransferase AmGT8 through semi-rational design and provided a feasible strategy for directed evolution of UGTs[20].

S. baicalensis Georgi is a popular medicinal plant distributed all around the world. Its roots are used as the traditional Chinese medicine Huang-Qin, and possess significant antiviral, anti-inflammatory, and anticancer activities. *S. baicalensis* is one of the few plants containing abundant flavonoid 2′-O-glycosides,

indicating the existence of 2′-O-glycosyltransferases[21,22]. In this work, we report a novel and efficient flavonoid 2′-O-glycosyltransferase (UGT71AP2, named by the UGT Nomenclature Committee; GenBank accession number, MK894456) from *S. baicalensis*. UGT71AP2 could preferentially catalyze the 2′-O-glycosylation of multi-hydroxy flavonoids. The mechanism for its regio-selectivity was dissected by crystal structure analysis and theoretical calculations. Moreover, a rationally designed mutant was obtained with remarkably improved catalytic efficiency and regio-selectivity.

2. Materials and methods

2.1. General remarks

Reference standards 5,7,2′-trihydroxy flavone (**1**), 7,2′-dihydroxy flavone (**2**), 3,2′-dihydroxy flavone (**5**), morusin (**8**), 2′-hydroxy flavavone (**9**), kaempferol (**13**), kaempferol 7-O-glucoside (**13a**), and astragalin (**13b**, kaempferol 3-O-glucoside) were purchased from Chengdu Must Biotechnology Co., Ltd. (China) and Aurum Pharmatech LLC. (USA). The compounds of 5,6′-dihydroxy-6,7,8,2′-tetramethoxy flavone (**3**), viscidulin III (**4**), 3,5,7,2′,6′-pentahydroxy flavone (**6**), 5,7,6′-trihydroxy-8,2′-dimethoxy flavone (**7**), viscidulin III 2′-O-glucoside (**4a**), 5,7,6′-trihydroxy flavone 2′-O-glucoside (**10**), 5,6′-dihydroxy-7,8-dimethoxy flavone 2′-O-glucoside (**11**), and 5,2′-dihydroxy-6-methoxyflavone 7-O-glucuronide (**12**) were previously purified and structurally characterized in our laboratory[23]. Thirteen UDP-sugar or GDP-sugar samples were purchased from Guangzhou AIfly Biotech Co., Ltd. (Guangzhou, China) and Sigma–Aldrich (St. Louis, MO, USA). Most of the molecular biology reagents were from Transgen Biotech (Beijing, China). Methanol and acetonitrile (Fisher Scientific, USA) were of HPLC grade. All other chemicals and reagents were purchased from Beijing Chemical Corporation (Beijing, China) unless otherwise stated.

2.2. Plant materials

The seeds of *S. baicalensis* were obtained from Anguo (China) and were sown in our laboratory under natural conditions. Ten days after sowing, the seedlings were collected and cleaned, and then immediately frozen in liquid nitrogen for RNA extraction.

2.3. Candidate gene screening

The transcriptome data of *S. baicalensis* Georgi (NCBI, SRP156996) were used. Reported UDP-glycosyltransferase gene

sequences (UGT88D1, AB479151.1; SbUFGT, AB031274.1; PfUGT57, AB362992.1; AmUGTcg10, AB362988.1; UGT88D5, AB362989.1; SiUGT23, AB362990.1; PfUGT50, AB362991.1; VvGT6, AB499075.1; VvGT5, AB499074.1) were selected as query sequences. Through BLAST analysis, more than 20 open reading frames (ORFs) were obtained as candidate genes. Five of them showed high expression levels (FPKM>200) (Supporting Information Table S1).

2.4.   Phylogenetic analysis

Molecular phylogenetic analysis was conducted using MEGA6 software with the maximum likelihood method. The bootstrap consensus tree inferred from 1000 replicates was used to represent the evolutionary history of the taxa analyzed (Supporting Information Fig. S1).

2.5.   Molecular cloning, site-directed mutagenesis, heterologous expression, and protein purification

2.5.1.   Molecular cloning
Total RNA of *S. baicalensis* was extracted from frozen seedlings by grinding the seedlings with a mortar and pestle under liquid nitrogen using TransZol method according to the manufacturer's instructions (Transgen Biotech, Beijing, China). The first-stranded complementary DNA (cDNA) was synthesized using TransScript one-step genomic DNA (gDNA) removal and cDNA synthesis SuperMix (Transgen Biotech, China). PCR was performed using 0.5 μL first strand cDNA as template, UGT-F and UGT-R as primers, and KD Plus DNA polymerase under manufacturer's instructions (Supporting Information Table S2). The amplified fragments were cloned into pET28a-(+) vector at the BamH I site by homologous recombination. The genes were sequenced by Tsingke Biological Technology Incorporation (Beijing, China). The recombinant vector was transformed into *Escherichia coli* BL21(DE3) for heterologous expression. Transformed cells were selected on agar plates containing 50 μg/mL kanamycin.

2.5.2.   Site-directed mutagenesis
The mutants of UGT71AP2 were obtained by Fast Mutagenesis system (Transgen Biotech, Beijing, China). The primers were synthesized by Tsingke Biological Technology (Beijing, China) (Table S2). The positive recombinant plasmids were transformed into *E. coli* BL21(DE3) for heterologous expression. The other conditions were the same as described above.

2.5.3.   Heterologous expression and protein purification
Single colonies harboring the target expression construct were inoculated overnight at 37 °C with shaking in LB culture medium containing 50 μg/mL kanamycin. Isopropyl thiogalactoside (0.1 mmol/L) was added into the medium for expression of recombinant protein at 18 °C when optical density at 600 nm was about 0.6. After incubation with shaking for 20 h, *E. coli* was harvested by centrifugation. The recombinant proteins were purified using a nickel-affinity column and concentrated using Amicon Ultra-15 Ultracel-30K (Merck Millipore). The concentrated proteins were used as purified recombinant enzymes. The recombinant UGT71AP2 determined by SDS-PAGE is shown in Supporting Information Fig. S2. The concentrations of recombinant enzymes were determined by micro-spectrophotometer with E 1% = 8.3.

2.6.   Effects of reaction time, pH, temperature, and divalent metal ions on enzyme activity

To characterize the enzymatic properties of UGT71AP2, its optimal reaction time, pH, temperature, and divalent metal ions were studied. To investigate the reaction time, 9 time points between 5 and 600 min were set. To optimize the reaction pH, enzymatic reactions were carried out in various reaction buffers ranging in pH values from 4.0 to 6.0 (citric acid-sodium citrate buffer), 6.0−8.0 (Na_2HPO_4/NaH_2PO_4 buffer), 7.0−9.0 (Tris-HCl buffer), and 9.0−10.0 (Na_2CO_3/$NaHCO_3$ buffer). To investigate the optimal reaction temperature, enzymatic reactions were incubated at different temperatures (4−60 °C). To test the dependence on divalent metal ions for UGT71AP2 activity, different divalent cations Ba^{2+}, Ca^{2+}, Co^{2+}, Cu^{2+}, Fe^{2+}, Mg^{2+}, Zn^{2+}, EDTA, and Blank in a final concentration of 5 mmol/L were used individually. All enzymatic reactions were conducted with UDP-glucose (UDP-Glc) as sugar donor and 5,7,2′-trihydroxy-flavone (**1**) as acceptor. All experiments were performed in triplicate and the mean value was used. The reactions were terminated with pre-cooled methanol (MeOH) and centrifuged at $21,130 \times g$ for 15 min for further HPLC analysis (Agilent 1260 instrument, Agilent Technologies, Waldbronn, Germany). The samples were separated on an Agilent Zorbax SB-C18 column (250 mm × 4.6 mm, 5 μm) at a flow rate of 1 mL/min. The mobile phase was 20%−100% MeOH in H_2O containing 0.1% formic acid in 20 min (linear gradient), followed by 100% MeOH for 5 min. The column temperature was 30 °C. The conversion rates in percentage were calculated from HPLC peak areas of glycosylated products and substrates (Supporting Information Fig. S3).

2.7.   Glycosyltransferase activity assay

The sugar acceptor and donor promiscuity of UGT71AP2 was investigated. The reaction mixtures were individually performed in a final volume of 100 μL containing 50 mmol/L Na_2HPO_4/NaH_2PO_4 buffer (pH 7), 8 μg of purified UGT71AP2, 0.1 mmol/L substrate, 0.5 mmol/L sugar-donor, 45 °C, 2 h. For unstable substrates **4**, **5**, and **6**, their optimized reaction conditions were as follows: a final volume of 100 μL containing 50 mmol/L Na_2HPO_4/NaH_2PO_4 buffer (pH 7), 2 mmol/L DTT, 0.1 mmol/L substrate, 0.5 mmol/L sugar-donor, 40 μg of purified UGT71AP2, 30 °C, 30 min. All the reactions were quenched with 200 μL ice-cold methanol. The supernatants were concentrated, resolved in 50% methanol, and then centrifuged at $21,130 \times g$ for 15 min before LC/MS analysis. Enzymatic activities of the mutants were determined with the same method as described above.

For LC/MS analysis, an Agilent 1260 HPLC instrument (Agilent Technologies, Waldbronn, Germany) was used. Samples were separated on an Agilent ZORBAX SB-C18 column (4.6 mm × 250 mm, 5 μm). The mobile phase consisted of methanol (A) and water containing 0.1% formic acid (*v/v*, B). The analytes were eluted using a linear gradient program: 0−20 min, 20%−100% A; 20%−25% min, 100% A; 26−31 min, 20% A. The flow rate was 1 mL/min. The column temperature was 30 °C. The ultraviolet absorption wavelength was 340 nm. MS analysis was performed on a Q-Exactive hybrid quadrupole-Orbitrap mass spectrometer connected to the HPLC instrument through a heated ESI source (Thermo Fisher Scientific, USA). The MS parameters were as follows: sheath gas pressure 45 arb, aux gas pressure 10 arb, discharge voltage 4.5 kV, capillary temperature 350 °C. MS resolution was set as 70,000 FWHM, AGC target 1^*E^6, maximum injection time 50 ms, and scan range *m/z* 100−1000. MS^2

resolution was set as 17,500 FWHM, AGC target 1*E^5, maximum injection time 100 ms, NCE 35.

2.8. Kinetic studies

All assays were performed in a final volume of 50 μL with 50 mmol/L Na$_2$HPO$_4$/NaH$_2$PO$_4$ buffer (pH 7.0). The reactions were incubated at 45 °C for 10 min, quenched with ice-cold MeOH, and then centrifuged at 21,130 × *g* for 15 min. The supernatants were used for HPLC analysis. The conversion rates in percentage were calculated from HPLC peak areas of glycosylated products and substrates. For UGT71AP2 native, 0.1175 μg protein, 2 mmol/L of saturated UDP-Glc, and different concentrations of compound **1** (5, 10, 20, 30, 50, 60, 80, 100, 200 μmol/L); For UGT71AP2 V179D/M180T mutant, 0.0569 μg protein, 2 mmol/L of saturated UDP-Glc, and different concentrations of compound **1** (5, 10, 20, 30, 50, 60, 80, 100, 120 μmol/L); For UGT71AP2 V179K/M180I mutant, 0.1625 μg protein, 2 mmol/L of saturated UDP-Glc, and different concentrations of compound **1** (5, 10, 20, 30, 50, 80, 100, 120, 150, 200 μmol/L); For UGT71AP2 L138D mutant, 0.425 μg protein, 2 mmol/L of saturated UDP-Glc, and different concentrations of compound **1** (5, 10, 20, 30, 50, 60, 80, 100 μmol/L); For UGT71AP2 L138T mutant, 0.0368 μg protein, 2 mmol/L of saturated UDP-Glc, and different concentrations of compound **1** (5, 10, 20, 30, 50, 100, 150 μmol/L); For UGT71AP2 L138T/V179D/M180T mutant, 0.0219 μg protein, 2 mmol/L of saturated UDP-Glc, and different concentrations of compound **1** (5, 10, 20, 30, 50, 80, 100, 120 μmol/L). Michaelis−Menten plot was fitted.

2.9. Preparative-scale reactions

The substrates (50 mmol/L) were dissolved in dimethyl sulfoxide (DMSO) as sugar acceptor, and 50 mmol/L UDP-Glc as sugar donor. The reactions contained 2 mL buffer (50 mmol/L Na$_2$HPO$_4$/NaH$_2$PO$_4$), 10−20 μL sugar acceptor (50 mmol/L), 20−40 μL sugar donor (50 mmol/L), and UGT71AP2 (100−200 μg). The reactions were incubated at 45 °C for up to 8 h and terminated by adding 6 mL methanol. The mixtures were then centrifuged at 21,130 × *g* for 15 min and the supernatants were concentrated and dissolved in 50% methanol. The glycosylated products were isolated and purified by semi-preparative HPLC (SSI, USA). NMR spectra were recorded on a Bruker AVANCE III-400 instrument at 400 (^{1}H) and 100 (^{13}C) MHz in DMSO-*d$_6$*.

2.10. Crystal structure of UGT71AP2

2.10.1. Protein expression, purification, and crystallization
The full-length cDNA of UGT71AP2 was cloned into a modified pET-28a(+) vector. A poly-histidine tag (His-tag) followed by a tobacco etch virus (TEV) protease cleavage sequence was added in front of the N-terminus of the target protein to facilitate purification. The protein was expressed in *E. coli* BL21(DE3) strain and was purified by Ni-NTA affinity chromatography. After TEV protease digestion, the protein was purified by size-exclusion chromatography on a Superdex 200 increase 10/300 GL pre-packed column (GE Healthcare). Fractions containing UGT71AP2 were collected and concentrated to 20 mg/mL using 20 mmol/L Tris-HCl (pH 7.5). Purified UGT71AP2 was incubated with 5 mmol/L UDP-Glc at 4 °C for 1 h before crystallization. The crystals were prepared by hanging drop vapor diffusion (1 μL protein and 1 μL reservoir solution). Crystals were obtained in

0.2 mol/L NaCl, 0.1 mol/L MES pH 6.0, 20% PEG6000. The crystals were flash-frozen in the reservoir solution supplemented with 30% glycerol or glycol.

2.10.2. Crystal structure determination and refinement
The diffraction data of UGT71AP2 and UGT71AP2·UDP crystals were collected at beamlines BL18U1 and BL19U1 of the Shanghai Synchrotron Radiation Facility (SSRF). All the data were processed with HKL-3000[24]. The structures were solved by molecular replacement with Phaser[25]. UGT72B1 structure (PDB ID: 2VCE) was used as the search model. Crystallographic refinement was performed repeatedly using Phenix and COOT[26,27]. The refined structures were validated by Phenix and the PDB validation server (https://validate-rcsb-1.wwpdb.org/). The final refined structures were deposited in the Protein Data Bank. The diffraction data and structure refinement statistics were summarized in Supporting Information Table S3.

2.11. In-silico modeling and calculations

2.11.1. Molecular modeling and molecular docking
The modeling of UGT71AP2 L138T/V179D/M180T mutant was carried out using UGT71AP2 (apo, 2.15 Å) as template at SWISS-MODEL online website[28]. UDP-Glc and UDP were simulated to UGT71AP2 (apo, 2.15 Å) and its mutants according to GgCGT/UDP-Glc and UGT71AP2/UDP complex structures[29]. Then we used AutoDock 4.2 software to construct a series of complex structures, which contain UGT71AP2/UDP-Glc/5,7,2′-trihydrox-yflavone (the axis of grid box is *x* = −31.574, *y* = −19.521, and *z* = −13.119), UGT71AP2 L138T/V179D/M180T mutant/UDP-Glc/5,7,2′-trihydroxyflavone (the axis of grid box is *x* = −31.574, *y* = −19.521, and *z* = −13.119), UGT71AP2/UDP/5,7,2′-tri-hydroxyflavone 2′-*O*-glucoside (the axis of grid box is *x* = −29.126, *y* = −19.021, and *z* = −8.533), UGT71AP2 L138T/V179D/M180T mutant/UDP/5,7,2′-trihydroxyflavone 2′-*O*-glucoside (the axis of grid box is *x* = −29.126, *y* = −19.021, and *z* = −8.533). Then we performed Auto-Dock analysis by Lamarckian Genetic Algorithm with default parameters for 2,500,000 evaluations in 100 cycles, and the other parameters followed the default information[30].

2.11.2. pK$_a$ prediction
The pK$_a$ values for the hydroxyl spots of **1** were calculated using the Jaguar package (Schrödinger, LLC: New York, USA). The lowest energy local conformation was used for pK$_a$ prediction at the level of B3LYP/6-31G** with implicit solvent model named polarizable continuum model (PCM)[31].

2.11.3. Molecular dynamics (MD) simulations
All the MD simulations were performed using the Desmond package (Desmond Molecular Dynamics System, D. E. Shaw Research, New York, version 2021-3) on the GPU nodes with NVIDIA Tesla T4 on Tetralith at Linköping University. The OPLS4 force field was used with the default atomic charge scheme[32]. By adding ∼13,800 TIP3P water molecules, the solvent environment is modeled explicitly with certain numbers of counter ions Na$^+$ and Cl$^-$ for mimicking the physical salt concentration of 0.15 mol/L[33]. An orthorhombic simulation box was adopted with a 10.0 Å buffering area to the solute, which is the protein−ligand complex. For maintaining the temperature and pressure of simulations at 300.0 K and 1.0 atm, the Nose-Hoover chain thermostat[34] and Martyna-Bobias-Klein barostat[35] were

used, respectively. The default minimization and equilibration protocol implemented in Desmond was used before the 100-ns production simulation for each protein−ligand system (Supporting Information Table S4), with the trajectory saved every 20 ps.

2.11.4. Binding free energy calculations

A total of 400 snapshots evenly extracted from 100-ns trajectory were used for the Prime (version 2021-3) MM/GBSA (molecular mechanics, the generalized Born model, and solvent accessibility) calculations. During the Prime calculations, the residues within 8 Å of the ligand were energy-minimized (OPLS4 force field) in the continuum solvation VSGB (variable dielectric surface generalized Born)[36]. The mean and deviation energy values calculated from the extracted snapshots were reported in this study.

2.11.5. QM/MM calculations

To understand the energy profile of the systems, we selected 3 MD snapshots from different systems for the QM/MM calculations using the ONIOM method in Gaussian 16 (rev. C.01)[37-39]. The tleap program in AmberTools20 was used to process the snapshots extracted for getting the forcefield topology files (.prmtop)[40]. To fit the ligand restraint electrostatic potential (RESP) charges by the Antechamber module of AmberTools20[41], the atomic electrostatic charges of **1** were derived from Gaussian calculations (HF/6-31G*). The extracted snapshots were energy minimized using the Sander program with the Amber99SB (protein)/GAFF (ligand) forcefield[42,43], followed by the setup of ONIOM calculations using the VMD MolUP plugin (version 1.7.0)[44]. Ions and water molecules beyond 4 Å of protein were removed. The cofactor, UDP-Glc, was truncated at the carbon atom neighboring the PA atom, while all the ligand atoms were included in the QM regions. The side-chain atoms of Asp107, His16, Glu372, Leu/Thr138, Val/Asp179, and Met/Thr180 and the water bridges between ligand and protein as well as the bridging protein atoms were also included. This results in the number of QM region atoms ranging from 126 to 139. To deal with the covalent bonds spanning between the QM and MM regions, the linking hydrogen atoms were added with defaults scaling factor for the new bond length. All residues and water molecules beyond 4 Å of **1**/UDP-Glc/**1a**/UDP or beyond 6 Å of the QM region atoms were frozen. The geometry optimizations and final energy calculations were performed at the levels of B3LYP/6-31g(d):AMBER and B3LYP-GD3BJ/6-311+g(2d,2p):AMBER, respectively. The geometry of transition state was fully optimized from an initial guess located by the flexible scanning along the reaction coordinates. Each optimized transition state was further confirmed by the unique imaginary vibrational mode, which connects the bonding and debonding atoms.

2.12. Hydrogen-deuterium exchange mass spectrometry (HDX-MS) analysis

HDX-MS analysis deuterium labeling was initiated with a 20-fold dilution into D_2O buffer (100 mmol/L phosphate, pD 7.0) of UGT71AP2/UDP, UGT71AP2/UDP/**1**, UGT71AP2/UDP/**1a**, L138T/V179D/M180T mutant/UDP/**1**, and L138T/V179D/M180T mutant/UDP/**1a**, respectively. The concentrations of protein and ligands are 1 mg/mL and 1 μmol/L, respectively. The other parameters were the same as described previously[45].

3. Results and discussion

3.1. Molecular cloning and functional characterization of the flavonoid 2′-O-glycosyltransferase UGT71AP2

More than twenty candidate UGT genes were obtained through BLAST search of the transcriptome data of *S. baicalensis*[46], using reported flavanoid UGTs as templates. In the phylogenetic tree, Sb03g16960 (UGT71AP2) was clustered with multi-site GTs (with poor regio-selectivity), indicating it was not a typical flavonoid 3-OGT, 5-OGT, or 7-OGT (Fig. S1). Three high-expression genes (Sb03g16960, Sb01g13350, and Sb03g36680) were cloned into pET-28a(+) vector and expressed in *E. coli* BL21(DE3) strain, and the proteins were purified for functional characterization (Table S2).

To discover the 2′-O-glycosyltransferase, 5,7,2′-trihydroxy-flavone (**1**) and UDP-Glc were utilized as sugar acceptor and sugar donor, respectively. The reaction mixture was analyzed by liquid chromatography coupled with mass spectrometry (LC/MS). Sb01g13350 and Sb03g36680 did not show obvious glycosylation activities. In the presence of Sb03g16960 (UGT71AP2), majority (>90%) of the substrate was transformed, and LC/MS analysis showed two major products (**1a** and **1b**) and one minor product **1c** as mono- or di-O-glycosides (Fig. 1A and B). Compounds **1a** and **1b** were then purified by preparative-scale enzyme catalysis, and were identified as 5,7,2′-trihydroxyflavone 2′-O-glucoside and 7,2′-di-O-glucoside (a new compound), respectively. The location of the glucosyl residue at 2′-OH was determined by the HMBC correlation of H-1″ (δ_H 5.12, the anomeric hydrogen) with C-2′ (δ_C 155.4). A time-course study indicated that **1** could be converted to **1a** quickly and then further converted to **1b** gradually (Fig. 1C).

The cDNA sequence of *UGT71AP2* contains an open reading frame (ORF) of 1392 bp encoding 463 amino acids. The catalytic conditions were optimized using **1** and UDP-Glc as substrates. UGT71AP2 exhibited its maximum activity at pH 7.0 (50 mmol/L Na_2HPO_4/NaH_2PO_4) and 45 °C, and the reaction was independent of divalent metal ions (Fig. S3).

3.2. Substrate and sugar donor promiscuity of UGT71AP2

Many UGTs show high catalytic promiscuity, such as OleD[47-48] and YjiC[49,50]. To explore the substrate promiscuity of UGT71AP2, we tested nine flavonoids (**1**−**9**) containing a 2′-hydroxy group (Fig. 1D). According to HPLC analysis, the conversion rates were >60% except for **8** (Fig. 1E, Supporting Information Figs. S4−S12). The products were characterized as O-glucosides according to the [M−H−162]⁻ fragment ions in the MS/MS spectra[51]. Then we purified **2a**, **2b**, **3a**, **5a**, **5b**, and **9a** from preparative-scale enzymatic reactions, and fully identified their structures by NMR and HR-ESIMS analyses (Supporting Information Figs. S13−S52). All the products were identified as 2′-O-glucosides according to the HMBC correlations between H-1″ and C-2′. These results confirmed the high regio-selectivity of UGT71AP2 towards 2′-OH. To the best of our knowledge, UGT71AP2 is the first reported flavonoid 2′-O-glycosyltransferase.

To probe the sugar donor specificity of UGT71AP2, 13 sugar donors were tested, including UDP-Glc, UDP-Gal, UDP-GlcN, UDP-GalN, UDP-Xyl, UDP-Ara, UDP-GlcNAc, UDP-GalNAc, UDP-GlcA, UDP-GalA, UDP-Rha, GDP-Man, and GDP-Fuc (Fig. 1D, Supporting Information Fig. S53). UGT71AP2 could utilize 5 UDP-sugars, with the preference
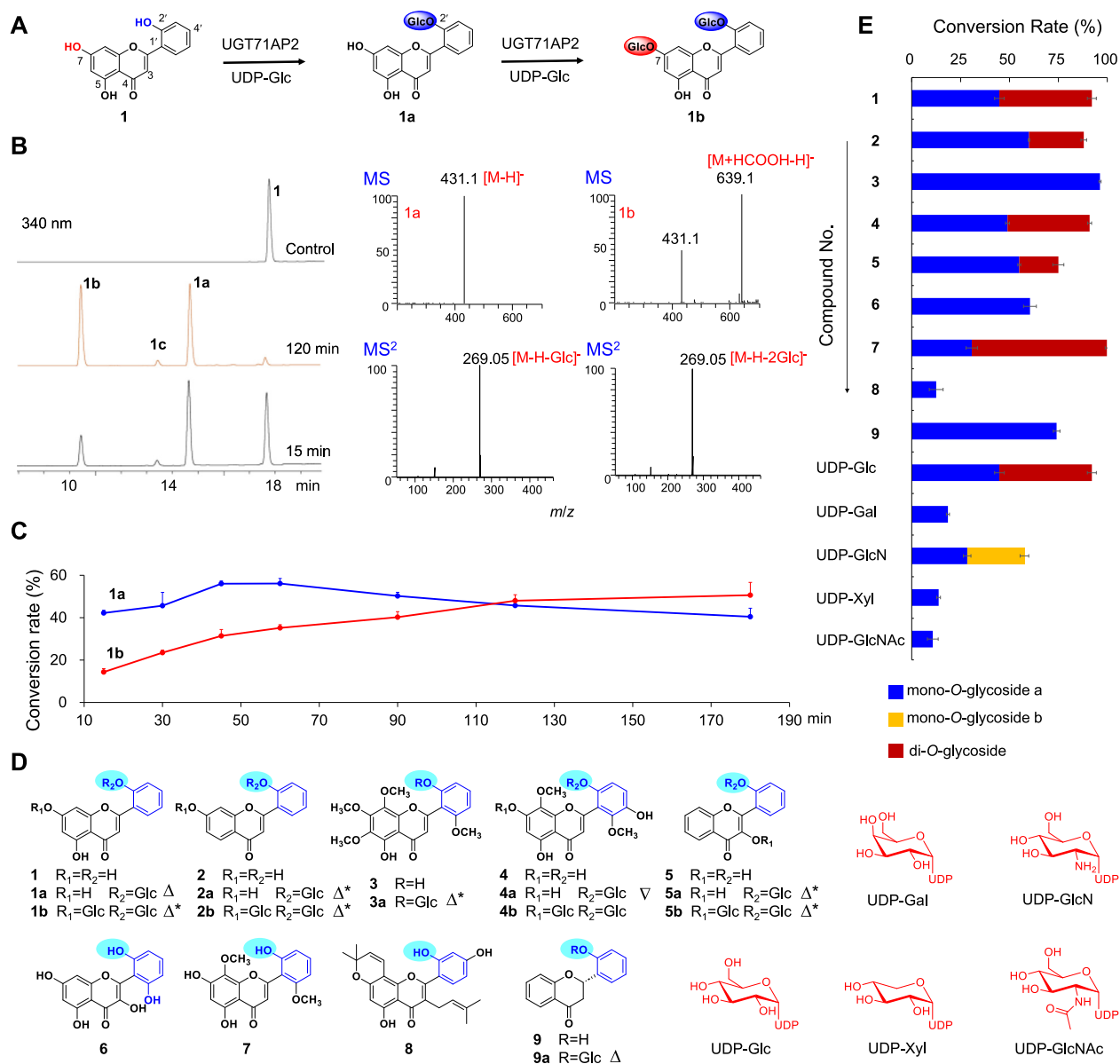
Figure 1 Substrate and donor promiscuity of UGT71AP2. (A−B) Glycosylation of 5,7,2′-trihydroxyflavone (**1**), showing HPLC chromatograms and mass spectra of **1a** and **1b**. **1c**, 5,7,2′-trihydroxyflavone 7-*O*-glucoside. (C) A time-course study of the enzymatic reaction catalyzed by UGT71AP2. (D) Structures of 9 sugar acceptors (**1**−**9**) and 5 UDP-sugar donors. Δ, the products were purified in this study and were identified by NMR. ∇, the products were identified by comparing with reference standards. *, new compounds. The blue circles highlight 2′-*O*-glycosylation. (E) Conversion rates of **1**−**9** with UDP-Glc as sugar donor, and of compound **1** with 5 sugar donors. For compounds **2**, **4**, and **6**, a very minor mono-*O*-glycoside product is present but not shown here (Supporting Information Figs. S5, S7, and S9). The experiments were performed in triplicates (*n* = 3).

order of UDP-Glc > UDP-GlcN > UDP-Gal > UDP-Xyl > UDP-GlcNAc (Fig. 1E). This preference was similar to that of GgCGT, a *C*-glycosyltransferase we had previously reported[29]. It was noteworthy that UGT71AP2 could catalyze **1** into two mono-*O*-glycoside products with similar conversion rates when UDP-GlcN was used as sugar donor. Most probably, these two products could be the 2′-*O*-glycoside and 7-*O*-glycoside. This result was different from UDP-Glc, with which UGT71AP2 showed high regio-selectivity towards 2′-OH.

Owing to the anti-inflammatory activities of *S. baicalensis*, we evaluated the inhibitory activities of **1**−**5** and their 2′-*O*-glycosides against the pro-inflammatory protein cyclooxygenase 2 (COX-2).

Compounds **2a**, **2b**, **3a**, and **4a** showed >50% inhibitions at 25 μmol/L against COX-2, and the 2′-*O*-glycosides of **2**, **4**, and **5** were significantly more potent than the aglycones (Supporting Information Fig. S54).

3.3. Crystal structure of UGT71AP2

Although the crystal structures of a number of plant glycosyltransferases have been reported, little is known about catalytic mechanisms for their regio-selectivity[13,14,52,53]. UGT71AP2 could catalyze the glycosylation of flavonoids at 2′-OH and 7-OH, though 2′-*O*-glycosylation was preferred (Supporting

Information Fig. S55). To gain insights into the regio-selectivity, we solved the crystal structure of the UGT71AP2/UDP complex (PDB ID: 8hoj, 2.9 Å), which contains two highly similar molecules with a root mean square deviation (RMSD) of 0.628 Å (Table S3). We also obtained a monomer apo structure of UGT71AP2 (PDB ID: 8hok, 2.15 Å). However, after many attempts including co-crystallization and soaking experiments, we failed to obtain the complex structure with sugar acceptor.

Similar to other UGTs, UGT71AP2 shows a typical GT-B fold with two Rossmann-like $\beta/\alpha/\beta$ domains (Fig. 2A). The N-terminal domain (NTD, residues 1−234 and 439−462) and the C-terminal domain (CTD, residues 235−438) are responsible primarily for sugar acceptor and sugar donor binding, respectively. In the meanwhile, the binding pocket is partially exposed to the solvent (Supporting Information Fig. S56).
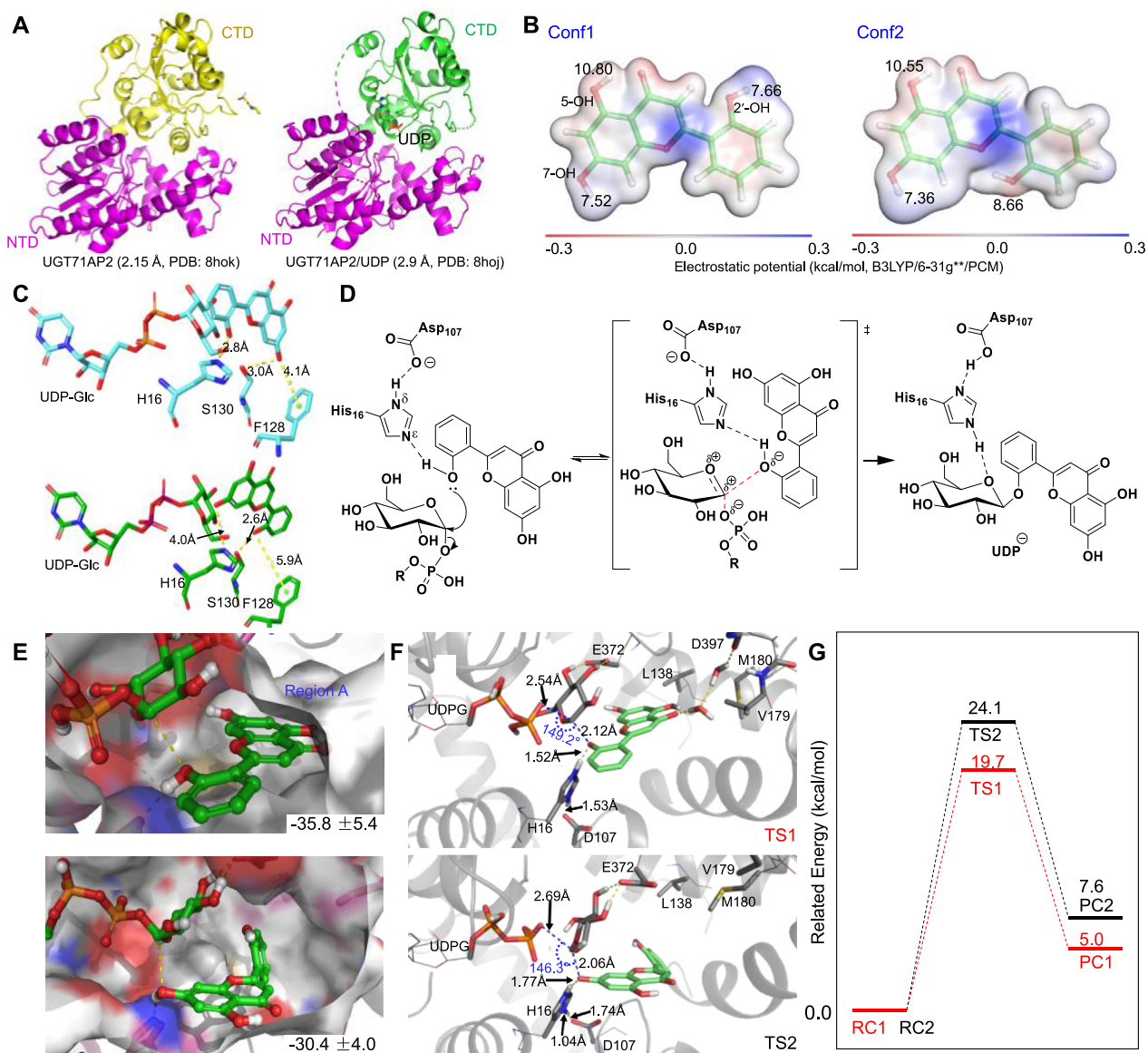

Figure 2 Structural basis for the regio-selectivity of UGT71AP2. (A) Crystal structures of UGT71AP2 (apo) and UGT71AP2·UDP. The CTD domains are depicted in yellow and green, respectively. The NTD domain is shown in pink. (B) Quantum chemical calculations predicted pK_a values of the hydroxy groups of **1**. (C) Molecular docking poses favoring 2′-O-glycosylation (top) and 7-O-glycosylation (bottom). (D) Proposed molecular mechanism of UGT71AP2-catalyzed UDP-dependent glycosylation. R, uridine monophosphate. Black and red dashed lines represent hydrogen bond interactions and reaction coordinates, respectively. (E) Representative snapshots from MD simulations for poses favoring 2′-O-glycosylation (top) and 7-O-glycosylation (bottom). The MM/GBSA binding free energies (kcal/mol) are given in the right bottom. (F) QM/MM optimized geometries for transition states TS1 and TS2. The reaction coordinates and hydrogen bonds are depicted as yellow and blue dotted lines, respectively. (G) Energy profiles for the 2′-O-glycosylation and 7-O-glycosylation reactions at the level of B3LYP-GD3BJ/6-311+g(2d,2p):amber with zero-point energy correction at the level of B3LYP/6-31G(d):amber. RC, reactant complex; TS, transition state; PC, product complex; 2′-O-glycosylation and 7-O-glycosylation poses are depicted in red and black, respectively. The energies of RC1 and RC2 are aligned to zero for the relation of energies of TS and PC in each system.

3.4. Structural basis for the 2′-O-glycosylation selectivity of UGT71AP2

To interpret whether the selectivity towards 2′-OH over other hydroxy sites is caused by chemical reactivity, we calculated the pK_a values of all the three possible glycosylation positions of 5,7,2′-trihydroxy flavone (**1**) using the Jaguar program[31]. For both configurations of **1**, the relatively small pK_a values of 7-OH indicated it was slightly more reactive than 2′-OH (Fig. 2B). We further conducted quantum chemical (QC) calculations using the anion form of **1** deprotonated at 2′- and 7-OH, respectively. The activation barrier for **1**-7-O^- bonding to UDP-Glc (16.5 kcal/mol)

is slightly lower than that of **1**-2′-O^- (17.3 kcal/mol, Supporting Information Fig. S57). The above calculation results were opposite to the experimental observations, indicating the intrinsic reactivity was not the driving force for 2′-O-glycosylation of **1**.

Through molecular docking, we obtained the structural model for UGT71AP2/UDP-Glc/**1**, which returned two binding modes favoring the glycosylation at 2′-OH and 7-OH, namely, 2′-O-pose and 7-O-pose, respectively (Fig. 2C)[30]. The 100-ns molecular dynamics (MD) simulations showed all the protein backbone and UDP-Glc atoms were stable in the two binding modes (RMSD ∼2 Å, Supporting Information Fig. S58)[54]. It is clear that **1** in the 2′-O-pose was more stable than in the 7-O-pose. In the 2′-O-pose,
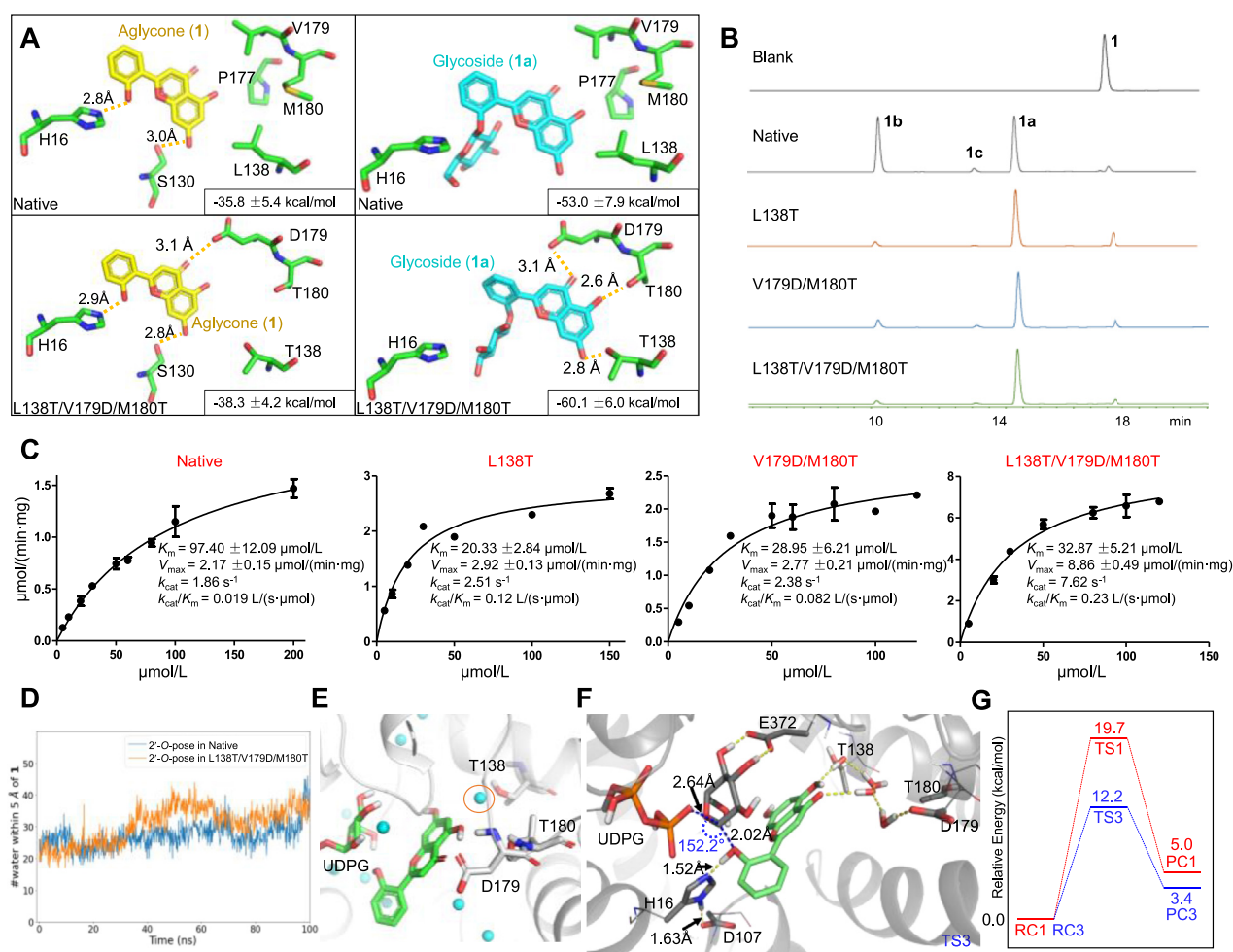

Figure 3   Rational design to improve the specificity of 2′-O-glycosylation of UGT71AP2. (A) Key residues in the docking models of 5,7,2′-trihydroxyflavone (**1**, left panels, referred to as 2′-O-pose) and 5,7,2′-trihydroxyflavone 2′-O-glucoside (**1a**, right panels, referred to as 2′-O-Glc-pose) in the active sites of native and mutant UGT71AP2. The MM/GBSA binding free energies from 100-ns MD simulations with deviations are shown at the right bottom of each subfigure. (B) HPLC chromatograms of enzymatic reactions catalyzed by the native enzyme and mutants, using **1** and UDP-Glc as the sugar acceptor and donor, respectively. **1b**, 5,7,2′-trihydroxyflavone 7,2′-di-O-glucoside; **1c**, 5,7,2′-trihydroxyflavone 7-O-glucoside. (C) Kinetic analysis of the reactions. (D) Time evolution of the number of water molecules within 5 Å of the ligand in the MD simulations. (E) Geometrically conserved (more than 20% enrichment in MD simulations) hydration sites from MD simulation for the 2′-O-pose in L138T/V179D/M180T mutant. Coordinates of these sites were calculated using our in-house toolkit written in C language. (F) QM/MM optimized transition state of 2′-O-pose of the mutant (TS3). The reaction coordinates and hydrogen bonds are depicted as yellow and blue dotted lines, respectively. (G) Energy profiles for the 2′-O-glycosylation reactions in native enzyme and L138T/V179D/M180T mutant at the level of B3LYP-D3BJ/6-311+g(2d,2p):amber with zero-point energy correction at the level of B3LYP/6-31G(d):amber. The reactant complex (RC), TS, and product complex (PC) of 2′-O-pose in the native and mutant are illustrated as RC1, TS1, PC1, and RC3, TS3, PC3, respectively. The energies of RC1 and RC3 are aligned to zero for the relation of energies of TS and PC in each system.

1 was stabilized by two hydrogen bonds with His16 and Ser130, and $\pi-\pi$ interactions of Phe128 with the A and C rings of **1**. Particularly, the δ-protonated His16 could form a hydrogen bond with Asp107, and the ε-nitrogen of His16 is close to 2'-OH (2.8 Å). Deprotonation of 2'-OH by the ε-nitrogen of His16 (from HID to HIE) could form product **1a** and protonate Asp107 (from Asp to ASH, Fig. 2D). Therefore, His16 plays dual roles as acid and base to transfer the proton from substrate to Asp107. Consistently, the H16A and D107A mutants almost showed no catalytic activity (Supporting Information Fig. S59).

We calculated the binding free energies of **1** in the two poses using the molecular mechanics generalized Born surface area (MM/GBSA) method[55]. The binding free energies of **1** in the 2'-*O*- and 7-*O*-poses were −35.8 ± 5.4 and −30.4 ± 4.0 kcal/mol, respectively (Fig. 2E, Supporting Information Figs. S60−S61), indicating the binding pose favoring 2'-*O*-glycosylation was more stable and had larger probability to occur than 7-*O*-glycosylation. From representative snapshots, we found the A/C rings of **1** could be well fitted into the hydrophobic area (region A) of the active site in the 2'-*O*-pose. For the 7-*O*-pose, however, the A/C rings of **1** were exposed to the solvent while the B ring pointed to region A.

We further constructed QM/MM model from representative snapshots of the two binding poses, which included the side chain conformations of important residues (Table S4). QM/MM calculations indicated that the deprotonation of substrate is coupled with the formation of the glycosidic bond and internal proton transfer within the His-Asp dyad[39,56]. By relating the energy of the transition state (TS) to the reactant complex (RC), QM/MM calculations showed a lower activation barrier for 2'-*O*-glycosylation (TS1, 19.7 kcal/mol) than for 7-*O*-glycosylation (TS2, 24.1 kcal/mol) (Fig. 2F−G and Supporting Information Table S5). Compared to the docking results (Fig. 2C), more residues are involved in the nonpolar interactions between the A/C rings of **1** and region A. These interactions provide extra stabilizing effects to TS1 and lower the activation barrier for 2'-*O*-glycosylation (Fig. 2F). This result interprets why 2'-*O*-glycosylation is the main catalytic reaction, and thus unravels structural basis for the regio-selectivity of UGT71AP2.

3.5. Structure-guided engineering to improve the 2'-OH selectivity

A time-course study (0−160 min) indicated UGT71AP2 could catalyze a second glycosylation reaction at both 7-OH of 2'-*O*-Glc (**1a**) and 2'-OH of 7-*O*-Glc (**1c**), and finally yielded **1b** (Supporting Information Fig. S62). To improve the 2'-OH regio-selectivity of UGT71AP2 and to improve the yield of 2'-*O*-glucoside, we used two strategies, *i.e.*, 1) to increase the stability of transition state with a lower activation barrier, and 2) to increase the stability of the product complex (PC) to suppress second-step glycosylation. During the MD simulations and MM/GBSA calculations, we found that region A was important for the regio-selectivity. Region A is mainly composed of Leu138, Pro177, Val179, and Met180. In the docking models, the carbonyl and hydroxyl groups on the A/C rings of **1** are surrounded by Region A (Fig. 3A). Particularly, the oxygen atoms are opposite to the methyl groups of L138 and V179 or the methylene groups of P177 and M180, leading to moderate interactions between **1** and region A. In the UGT71AP2/UDP/**1a** model, the above four residues showed similar interactions with **1a**. To further stabilize the binding of **1** or **1a** in region A, we considered introducing new hydrogen bonds between the ligands and surrounding key amino acids.

With this aim, we constructed a series of mutants, in which the introduction of carboxyl or hydroxyl groups to the side chain of L138 and V179/M180 remarkably improved the regio-selectivity towards 2'-*O*-glycosylation (Supporting Information Fig. S63). Particularly, the ratio of 7-*O*-glycosylation decreased from 48% of the native to 4% and 7% of the L138T and V179D/M180T mutants, respectively (Table 1, Fig. 3B). Moreover, these two mutants showed higher affinities with substrate **1** (K_m = 20.33 and 28.95 µmol/L for L138T and V179D/M180T, respectively, *vs* 97.40 µmol/L for the native), and 4.5 and 6.5 times higher catalytic efficiencies than the native enzyme (k_{cat}/K_m = 0.019 L/(s·µmol) (Fig. 3C, Supporting Information Fig. S64). We further constructed the L138T/V179D/M180T mutant. It also showed high regio-selectivity towards 2'-*O*-glycosylation, with only 4% yield of 7-*O*-glucoside. More importantly, the

Table 1   The conversion rates of 5,7,2'-trihydroxyflavone (**1**) catalyzed by UGT71AP2 (native) and its mutants, using UDP-Glc as the sugar donor.

Name	Total (%)	**1a** (%)	**1b** (%)	**1c** (%)
Native	94.2	46.2	48.0	Trace
L138D	80.5	80.5	n.d.	n.d.
L138E	89.0	85.0	Trace	4.0
L138S	87.4	53.5	22.3	11.6
L138V	83.7	63.2	20.5	Trace
L138T	90.0	86.0	4.0	n.d.
L138W	81.4	35.3	28.9	17.2
L138K	64.5	53.5	8.0	3.0
V179A/M180A	89.0	26.3	58.9	3.8
V179D/M180S	91.4	64.6	26.7	Trace
V179D/M180T	93.0	86.0	7.0	Trace
V179E/M180S	87.6	50.6	31.8	5.1
V179E/M180T	89.6	65.0	20.6	4.0
V179K/M180I	77.2	57.6	14.5	5.2
L138T/V179D/M180T	96.0	92.0	4.0	n.d.

1a, 2'-*O*-glucoside; **1b**, 7,2'-di-*O*-glucoside; **1c**, 7-*O*-glucoside. n.d., not detected.

triple-site mutant showed even increased catalytic efficiency, with a k_{cat}/K_m value of 0.23 L/(s·μmol), which was 12-fold higher than the native. The high regio-specificity and catalytic efficiency were verified by a time-course study (Fig. S55). Substrate **1** was almost completely converted to **1a** within 45 min, and the ratio of 7-*O*-glycosylation (**1b/1c**) never exceeded 5%. In addition, we found that S130 and P177 were related to 7-*O*-glycosylation. On the other hand, we also obtained the S130A/P177R mutant, which showed preference to 7-*O*-glycosylation (Supporting Information Fig. S65).

Next, we established structure of the L138T/V179D/M180T mutant by the SWISS-MODEL for dockings with **1** and **1a** by the Autodock suite (Fig. 3A, lower panels)[28]. Subsequently, 100-ns MD simulations were performed by the Desmond program. The MD-based MM/GBSA calculations demonstrated the binding free energy of **1** in the mutant as -38.3 ± 4.2 kcal/mol, which was lower than in the native (-35.8 ± 5.4 kcal/mol, Fig. 3A). During the course of MD simulation, we observed an increased number of water molecules around the active site, which could attribute to the large fluctuation of **1** in the L138T/V179D/M180T mutant (Fig. 3D). One hydration spot between **1** and region A was conserved in more than 20% of all the MD snapshots in the mutant, indicating that more water molecules were located in region A (Fig. 3E). These observations suggest the mutagenesis increased the hydrophilicity of region A, which further increased its polar interactions with the A/C rings of **1**, most probably
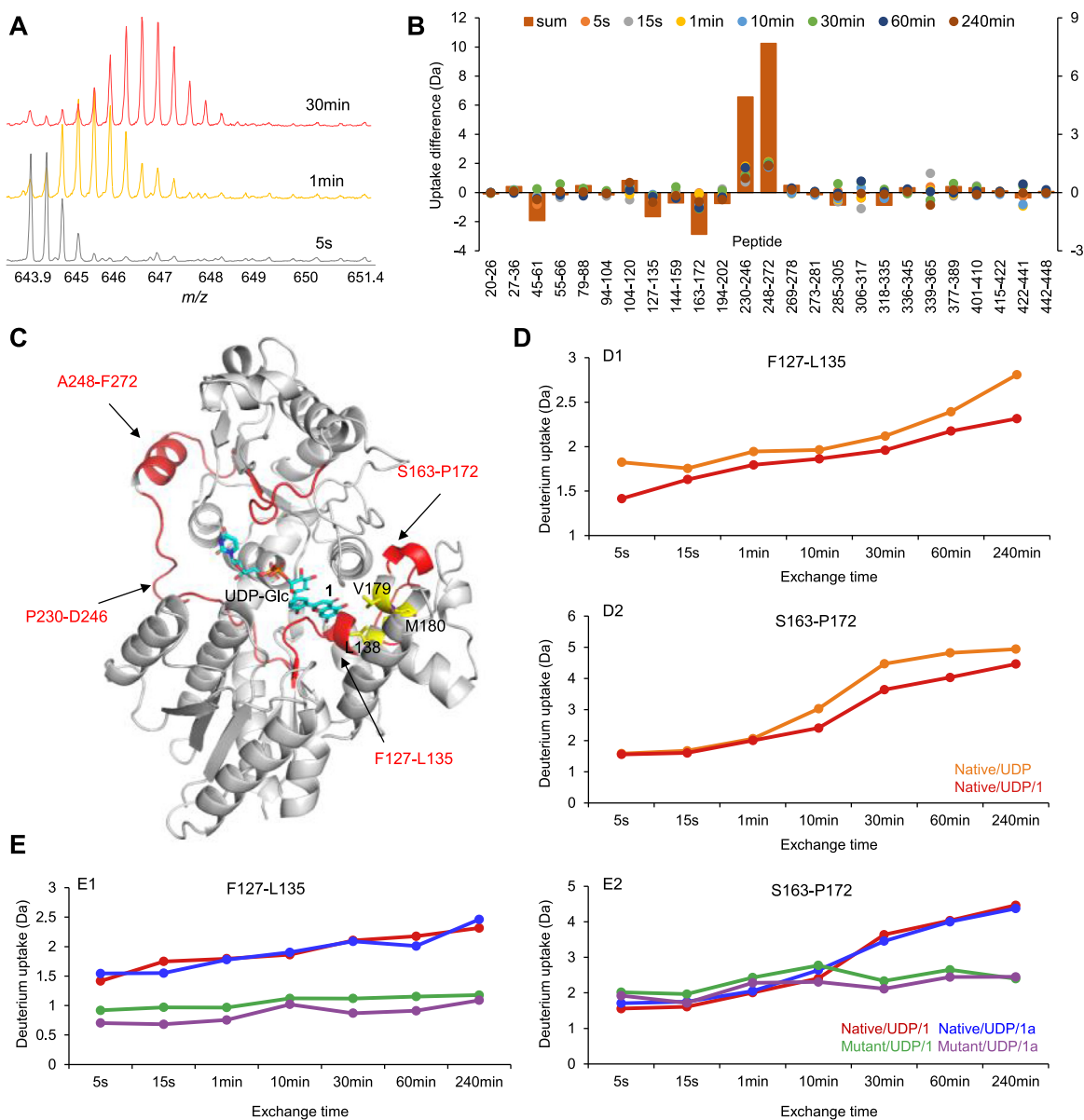

Figure 4 Structural dynamics analysis for the native and mutants of UGT71AP2 using Hydrogen−Deuterium Exchange Mass Spectrometry (HDX-MS). (A) Deuterium uptake spectra at different exchange time points. (B) Deuterium uptake differences of UGT71AP2/UDP and UGT71AP2/UDP/**1**. The uptake differences were calculated as the sum of differences for each peptide at each exchange time point. (C) Peptide fragments that exhibited significant differences in deuterium uptake. (D−E) Deuterium uptake plots of F127-L135 and S163-P172 peptide fragments for different proteins and ligands.

via water bridges. The hydrogen bonding network mediated by water molecules may be important to lower the binding free energy or activation energy[57]. By keeping the water molecules around **1** in the representative snapshots, QM/MM calculations with dispersion correction illustrate the stabilizing effect of the solvent hydrogen bonding network between **1** and UGT71AP2 for the transition state (7.5 kcal/mol lower in activation barrier, Fig. 3F−G).

We also docked the 2′-*O*-glycosylation product **1a** into the active site of L138T/V179D/M180T mutant/UDP and native/UDP models, followed by 100-ns MD simulations and MM/GBSA free energy calculations (Fig. 3A, Supporting Information Fig. S66). In the mutant system, T138, D179, and T180 could form hydrogen bonds with **1a**. These hydrogen bonds could stabilize **1a** in the active pocket, and thus suppress its re-orientation to hinder a second glycosylation at 7-OH. This is also supported by the MM/GBSA calculations, where the binding free energies for **1a** in native and L138T/V179D/M180T mutant systems are −53.0 ± 7.9 and −60.1 ± 6.0 kcal/mol, respectively. These results unraveled the mechanisms underlying the highly specific and efficient 2′-*O*-glycosylation of mutant L138T/V179D/M180T.

3.6. Hydrogen-deuterium exchange mass spectrometry (HDX-MS) analysis

To further evaluate the binding stability of the substrate (**1**) and product (**1a**) in the active pocket, we employed HDX-MS to explore the protein conformational dynamics for UGT71AP2 and its L138T/V179D/M180T mutant (Fig. 4A). The protein amino acid sequence coverage rate was >70%. When substrate **1** was added, the deuterium uptake rate of the peptide fragments K45-N61, F127-L135, and S163-P172 decreased remarkably, indicating this part of UGT71AP2 was more rigid after binding with **1** (Fig. 4B−D, Supporting Information Fig. S67). We failed to obtain information on peptide fragments containing L138, V179, and M180. However, they are close to F127-L135 and S163-P172. Changes of these peptides indicated significance of the above amino acids in substrate binding. On the other hand, the deuterium uptake rates for peptide fragments P230-D246 and A248-F272 increased, indicating they may be related to sugar donor binding (Fig. 4C).

More importantly, we analyzed the dynamic changes during 5 s and 240 min of the native and mutant when they were co-
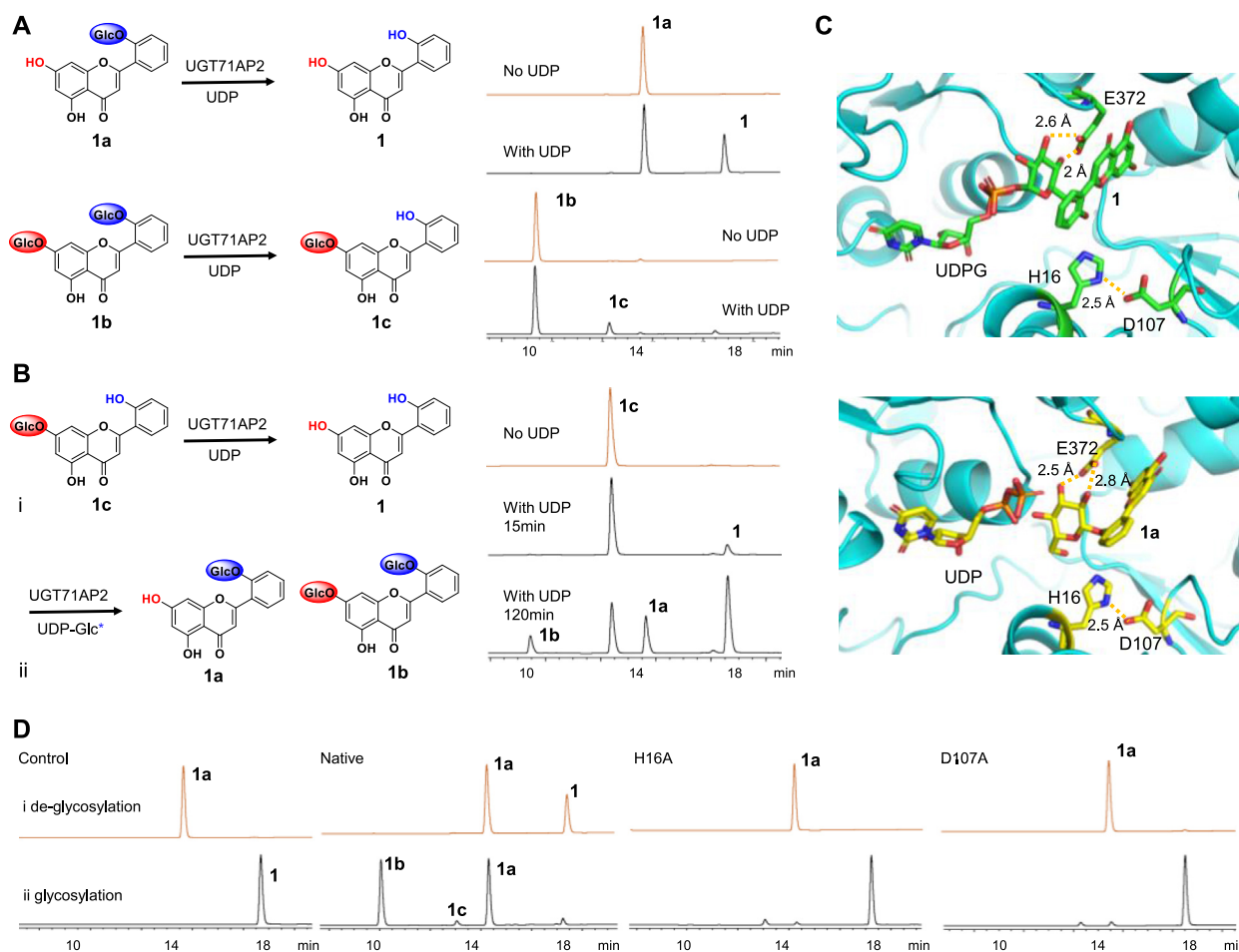

Figure 5 UDP-dependent de-glycosylation catalyzed by UGT71AP2. (A) The de-glycosylation reactions of **1a** and **1b**. The reaction mixtures were individually performed in a final volume of 100 μL containing 50 mmol/L Na_2HPO_4/NaH_2PO_4 buffer (pH 7), 8 μg of purified UGT71AP2, 0.1 mmol/L **1a** or **1b**, 0.5 mmol/L UDP, 45 °C, 15 min. (B) Dynamic equilibrium of de-glycosylation and glycosylation reactions of **1c**. UDP-Glc*, newly generated UDP-Glc. (C) Models of UGT71AP2/UDP-Glc/5,7,2′-trihydroxy flavone (**1**) and UGT71AP2/UDP/5,7,2′-trihydroxy-flavone 2′-*O*-glucoside (**1a**). (D) The catalytic activities of H16A and D107A mutants.

incubated with **1** and **1a**. By comparing the L138T/V179D/M180T mutant to the native, the deuterium uptake decreased for F127-L135 and S163-P172, indicating the mutant structures in complex with **1** and **1a** were more stable than the native (Fig. 4E). These results were consistent with theoretical calculations.

3.7. UDP-dependent de-glycosylation of UGT71AP2

Despite the high catalytic efficiency of both the native and mutant enzymes, we still could see minor residual substrates in the reaction system. We speculate that the reverse of glycosylation occurs, as bacterial GTs had been reported to possess dual functions[58,59]. Thus, we co-incubated **1a** and **1b** with the enzyme, but did not observe any products. However, when UDP was added into the system, both **1a** and **1b** could be de-glycosylated at the 2′-O-glucosyl group after 15 min, with conversion rates of 35% and 8%, respectively (Fig. 5A). When **1c** was co-incubated with UDP and UGT71AP2, we observed a dynamic equilibrium between de-glycosylation and glycosylation after 120 min (Fig. 5B). Similar de-glycosylation reactions were also observed when other 2′-O-glycosides were used as substrates (Supporting Information Figs. S68−S70). Then we tested seven known UGTs, and all of them showed UDP-dependent de-glycosylation capabilities (Supporting Information Figs. S71−S73)[7,60-62]. These results indicate that plant UGTs may generally possess UDP-dependent dual functions of glycosylation and de-glycosylation.

Glycosidases have been reported to show dual functions of glycosylation and de-glycosylation, and the reactions are independent of UDP[63,64]. The catalytic process is characterized by a two-step double-displacement of two acidic residues; one as an acid/base catalyst and the other as a nucleophile (see Supporting Information Fig. S74 as an example). In the active site of UGT71AP2, however, no similar acidic residues could be mapped. Thus, UGT71AP2 may have a new mechanism of de-glycosylation.

It is well recognized that the highly conserved catalytic dyad (His-Asp) arrangement acts as a key factor in initiating the glycosylation reaction of UGTs. According to the structural model of UGT71AP2/UDP/**1a**, we propose that the de-glycosylation and glycosylation of UGTs may share the same set of key amino acids (Fig. 5C). Thus, we constructed the H16A and D107A mutants. Both de-glycosylation and glycosylation activities were lost (Fig. 5D). Similar results were observed for saturation mutagenesis of His16 (Supporting Information Fig. S75). Glu372 also played a critical role in de-glycosylation by forming hydrogen bonds with the sugar moiety of **1a**. Consistently, the E372A mutant almost showed no de-glycosylation activity (Supporting Information Fig. S76).

As the same set of key amino acids plays the same effect on both glycosylation and deglycosylation activities (Fig. 5C and D), the de-glycosylation activity should be the reverse reaction of glycosylation[58,59]. From the mechanism shown in Fig. 2D, to suppress the reverse direction of the reaction, the protonated Asp107 in the product complex needs to be consumed by the environment. This indicates us to increase the buffer pH when the initial glycosylation reaction reaches a balance. Consistently, we found that the de-glycosylation was remarkably suppressed when the pH increased from 6 to 9 (Supporting Information Fig. S77). Thus, we incubated the reaction mixture at the optimal pH 7 for 1 h and then increased the pH using two volumes of pH 9 buffer. The glycosylation conversion rate of **1** by the L138T/V179D/M180T mutant was almost 100%, and no residual **1** could be observed (Supporting Information Fig. S78).

4. Conclusions

In summary, we characterized the first flavonoid 2′-O-glycosyltransferase UGT71AP2 from *S. baicalensis*. It could regioselectively and efficiently catalyze 2′-O-glycosylation of flavonoids. Eight 2′-O-glycosides were obtained and six of them are new compounds. We solved the crystal structure of UGT71AP2 (2.15 Å), and deduced the regio-selectivity of UGT71AP2 was mainly determined by the binding mode of the substrate in the active pocket. We obtained the L138T/V179D/M180T mutant, which showed remarkably enhanced 2′-OH regio-selectivity (the ratio of 7-O-glycosylation decreased from 48% to 4%) and improved catalytic efficiency (k_{cat}/K_m, 0.23 L/(s·μmol), 12-fold higher than the native). Furthermore, we find that UGT71AP2, joined by other plant glycosyltransferases, possesses moderate UDP-dependent de-glycosylation activity. The His-Asp dyad may be shared by the glycosylation and de-glycosylation reactions as key amino acids to initiate reaction. This work provides a new and efficient approach to prepare flavonoid 2′-O-glycosides, and elucidates the regio-selectivity mechanisms of UGT71AP2. The study also sets a good example for protein engineering to optimize the catalytic features of enzymes through structure-guided rational design and theoretical calculations.

Author contributions

Conceptualization: Min Ye, Zilong Wang, De-An Guo; Experiments: Zilong Wang, Xueqing Du, Guo Ye, Haotian Wang, Yizhan Liu, Chenrui Liu; Theoretical computations: Junhao Li, Hans Ågren, Yang Zhou; Crystal structure analysis: Fudong Li, Chao He; Original draft: Zilong Wang; Review and editing: Min Ye, Zilong Wang, Junhao Li. All the authors have given approval to the final version of the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Appendix A.　Supporting information

Supporting information to this article can be found online at https://doi.org/10.1016/j.apsb.2024.04.001.

References

1. Veitch NC, Grayer RJ. Flavonoids and their glycosides, including anthocyanins. *Nat Prod Rep* 2011;**28**:1626−95.

2. Xiao J. Dietary flavonoid aglycones and their glycosides: which show better biological significance?. *Crit Rev Food Sci Nutr* 2017;**57**: 1874−905.

3. Gupta S, Thakur K, Khare NK. Regio and stereoselectivity in *O*-glycosylation reactions—a review. *Trends Carbohydr Res* 2019;**11**: 1−42.

4. Yang B, Liu H, Yang J, Gupta VK, Jiang Y. New insights on bio-activities and biosynthesis of flavonoid glycosides. *Trends Food Sci Technol* 2018;**79**:116−24.

5. Putkaradze N, Teze D, Fredslund F, Welner DH. Natural product *C*-glycosyltransferases—a scarcely characterised enzymatic activity with biotechnological potential. *Nat Prod Rep* 2021;**38**:432−43.

6. Han BY, Wang ZL, Li JH, Jin Q, Wang HT, Chen K, et al. A highly selective *C*-rhamnosyltransferase from *Viola tricolor* and insights into its mechanisms. *Acta Pharm Sin B* 2023;**13**:3535−44.

7. Wang Z, Wang S, Xu Z, Li M, Chen K, Zhang Y, et al. Highly promiscuous flavonoid 3-*O*-glycosyltransferase from *Scutellaria baicalensis*. *Org Lett* 2019;**21**:2241−5.

8. Nakatsuka T, Sato K, Takahashi H, Yamamura S, Nishihara M. Cloning and characterization of the UDP-glucose:anthocyanin 5-*O*-glucosyltransferase gene from blue-flowered gentian. *J Exp Bot* 2008; **59**:1241−52.

9. Liu X, Cheng J, Zhang G, Ding W, Duan L, Yang J, et al. Engineering yeast for the production of breviscapine by genomic analysis and synthetic biology approaches. *Nat Commun* 2018;**9**:448.

10. Yi D, Bayer T, Badenhorst CPS, Wu S, Doerr M, Höhne M, et al. Recent trends in biocatalysis. *Chem Soc Rev* 2021;**50**:8003−49.

11. Yuan S, Yin S, Liu M, Kong JQ. Isolation and characterization of a multifunctional flavonoid glycosyltransferase from *Ornithogalum caudatum* with glycosidase activity. *Sci Rep* 2018;**8**:5886.

12. Huang W, He Y, Jiang R, Deng Z, Long F. Functional and structural dissection of a plant steroid 3-*O*-glycosyltransferase facilitated the engineering enhancement of sugar donor promiscuity. *ACS Catal* 2022;**12**:2927−37.

13. Liang DM, Liu JH, Wu H, Wang BB, Zhu HJ, Qiao JJ. Glycosyl-transferases: mechanisms and applications in natural product development. *Chem Soc Rev* 2015;**44**:8350−74.

14. Kurze E, Wüst M, Liao J, McGraphery K, Hoffmann T, Song C, et al. Structure−function relationship of terpenoid glycosyltransferases from plants. *Nat Prod Rep* 2022;**39**:389−409.

15. Lee SG, Salomon E, Yu O, Jez JM. Molecular basis for branched steviol glucoside biosynthesis. *Proc Natl Acad Sci U S A* 2019;**116**: 13131−6.

16. Yang T, Zhang J, Ke D, Yang W, Tang M, Jiang J, et al. Hydrophobic recognition allows the glycosyltransferase UGT76G1 to catalyze its substrate in two orientations. *Nat Commun* 2019;**10**:3214.

17. Zhang J, Tang M, Chen Y, Ke D, Zhou J, Xu X, et al. Catalytic flexibility of rice glycosyltransferase OsUGT91C1 for the production of palatable steviol glycosides. *Nat Commun* 2021;**12**:7030.

18. Li J, Yang J, Mu S, Shang N, Liu C, Zhu Y, et al. Efficient *O*-glycosylation of triterpenes enabled by protein engineering of plant glycosyltransferase UGT74AC1. *ACS Catal* 2020;**10**:3629−39.

19. Wen Z, Zhang ZM, Zhong L, Fan J, Li M, Ma Y, et al. Directed evolution of a plant glycosyltransferase for chemo- and regioselective glycosylation of pharmaceutically significant flavonoids. *ACS Catal* 2021;**11**:14781−90.

20. Zhang M, Yi Y, Gao BH, Su HF, Bao YO, Shi XM, et al. Functional characterization and protein engineering of a triterpene 3-/6-/2′-*O*-glycosyltransferase reveal a conserved residue critical for the regio-specificity. *Angew Chem Int Ed* 2022;**61**:e202113587.

21. Wang ZL, Wang S, Kuang Y, Hu ZM, Qiao X, Ye M. A comprehensive review on phytochemistry, pharmacology, and flavonoid biosynthesis of *Scutellaria baicalensis*. *Pharm Biol* 2018;**56**:465−84.

22. Miyaichi Y, Hanamitsu E, Kizu H, Tomimori T. Studies on the constituents of *Scutellaria* species (xxii). constituents of the roots of *Scutellaria amabilis* HARA. *Chem Pharm Bull (Tokyo)* 2006;**54**: 435−41.

23. Ji S, Li R, Wang Q, Miao WJ, Li ZW, Si LL, et al. Anti-H1N1 virus, cytotoxic and Nrf2 activation activities of chemical constituents from *Scutellaria baicalensis*. *J Ethnopharmacol* 2015;**176**:475−84.

24. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr Sect D Biol Crystallogr* 2006;**62**:859−66.

25. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr* 2007; **40**:658−74.

26. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* 2010;**66**:213−21.

27. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of coot. *Acta Crystallogr Sect D Biol Crystallogr* 2010;**66**: 486−501.

28. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018;**46**: W296−303.

29. Zhang M, Li FD, Li K, Wang ZL, Wang YX, He JB, et al. Functional characterization and structural basis of an efficient di-*C*-glycosyl-transferase from *Glycyrrhiza glabra*. *J Am Chem Soc* 2020;**142**: 3506−12.

30. Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ. Computational protein−ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc* 2016;**11**:905−19.

31. Bochevarov AD, Watson MA, Greenwood JR, Philipp DM. Multi-conformation, density functional theory-based pK_a prediction in application to large, flexible organic molecules with diverse functional groups. *J Chem Theor Comput* 2016;**12**:6001−19.

32. Lu C, Wu C, Ghoreishi D, Chen W, Wang L, Damm W, et al. OPLS4: improving force field accuracy on challenging regimes of chemical space. *J Chem Theor Comput* 2021;**17**:4291−300.

33. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;**79**:926−35.

34. Martyna GJ, Klein ML, Tuckerman M. Nosé−Hoover chains: the canonical ensemble *via* continuous dynamics. *J Chem Phys* 1992;**97**: 2635−43.

35. Martyna GJ, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. *J Chem Phys* 1994;**101**:4177−89.

36. Li J, Abel R, Zhu K, Cao Y, Zhao S, Friesner RA. The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling. *Proteins* 2011;**79**:2794−812.

37. Dapprich S, Komáromi I, Byun KS, Morokuma K, Frisch MJ. A new ONIOM implementation in Gaussian98. Part i. the calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J Mol Struct: THEOCHEM* 1999;**461−462**:1−21.

38. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. *Gaussian 16, revision C. 01; Gaussian*. Wallingford, CT: Inc; 2016.

39. Chung LW, Sameera WMC, Ramozzi R, Page AJ, Hatanaka M, Petrova GP, et al. The ONIOM method and its applications. *Chem Rev* 2015;**115**:5678−796.

40. Case DA, Berryman JT, Betz RM, Cerutti DS, Cheatham III TE, Darden TA, et al. *Amber 2020*. San Francisco, CA, USA: University of California; 2020.

41. Wang J, Cieplak P, Kollman PA. How well does a Restrained Electrostatic Potential (RESP) model perform in calculating conformational energies of organic and biological molecules?. *J Comput Chem* 2000;**21**:1049−74.

42. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general Amber force field. *J Comput Chem* 2004;**25**:1157−74.

43. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 2006;**65**:712−25.

44. Fernandes Hs, Ramos MJ, Cerqueira NMFSA. molUP: a VMD plugin to handle QM and ONIOM calculations using the Gaussian software. *J Comput Chem* 2018;**39**:1344−53.

45. Wang HT, Wang ZL, Chem K, Yao MJ, Zhang M, Wang RS, et al. Insights into the missing apiosylation step in flavonoid apiosides biosynthesis of Leguminosae plants. *Nat Commun* 2023;**14**:6658.

46. Wang ZL, Gao HM, Wang S, Zhang M, Chen K, Zhang YQ, et al. Dissection of the general two-step di-*C*-glycosylation pathway for the biosynthesis of (iso)schaftosides in higher plants. *Proc Natl Acad Sci U S A* 2020;**117**:30816−23.

47. Williams GJ, Zhang C, Thorson JS. Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. *Nat Chem Biol* 2007;**3**:657−62.

48. Zhou M, Hamza A, Zhan CG, Thorson JS. Assessing the regioselectivity of OleD-catalyzed glycosylation with a diverse set of acceptors. *J Nat Prod* 2013;**76**:279−86.

49. Pandey RP, Gurung RB, Parajuli P, Koirala N, Tuoi LT, Sohng JK. Assessing acceptor substrate promiscuity of YjiC-mediated glycosylation toward flavonoids. *Carbohydr Res* 2014;**393**:26−31.

50. Pandey RP, Parajuli P, Gurung RB, Sohng JK. Donor specificity of YjiC glycosyltransferase determines the conjugation of cytosolic NDP-sugar in *in vivo* glycosylation reactions. *Enzym Microb Technol* 2016;**91**:26−33.

51. Qiao X, Li R, Song W, Miao WJ, Liu J, Chen HB, et al. A targeted strategy to analyze untargeted mass spectral data: rapid chemical profiling of *Scutellaria baicalensis* using ultra-high performance liquid chromatography coupled with hybrid quadrupole orbitrap mass spectrometry and key ion filtering. *J Chromatogr A* 2016;**1441**:83−95.

52. Hsu TM, Welner DH, Russ ZN, Cervantes B, Prathuri RL, Adams PD, et al. Employing a biochemical protecting group for a sustainable indigo dyeing strategy. *Nat Chem Biol* 2018;**14**:256−61.

53. He JB, Zhao P, Hu ZM, Liu S, Kuang Y, Zhang M, et al. Molecular and structural characterization of a promiscuous *C*-glycosyltransferase from *Trollius chinensis*. *Angew Chem Int Ed* 2019;**58**:11513−20.

54. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE. Long-timescale molecular dynamics simulations of protein structure and function. *Curr Opin Struct Biol* 2009;**19**:120−7.

55. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 2000;**33**:889−97.

56. Tvaroška I, Kozmon S, Wimmerová M, Koča J. Substrate-assisted catalytic mechanism of *O*-GlcNAc transferase discovered by quantum mechanics/molecular mechanics investigation. *J Am Chem Soc* 2012;**134**:15563−71.

57. Zhao Y, Fang C, Zhang Q, Zhang R, Zhao X, Duan Y, et al. Crystal structure of SARS-CoV-2 main protease in complex with protease inhibitor PF-07321332. *Protein Cell* 2022;**13**:689−93.

58. Zhang C, Griffith BR, Fu Q, Albermann C, Fu X, Lee IK, et al. Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. *Science* 2006;**313**:1291−4.

59. Gantt RW, Peltier-Pain P, Singh S, Zhou M, Thorson JS. Broadening the scope of glycosyltransferase-catalyzed sugar nucleotide synthesis. *Proc Natl Acad Sci U S A* 2013;**110**:7648−53.

60. Li L, Modolo LV, Escamilla-Trevino LL, Achnine L, Dixon RA, Wang X. Crystal structure of *Medicago truncatula* UGT85H2−insights into the structural basis of a multifunctional (iso)flavonoid glycosyltransferase. *J Mol Biol* 2007;**370**:951−63.

61. Modolo LV, Li L, Pan H, Blount JW, Dixon RA, Wang X. Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso)flavonoids. *J Mol Biol* 2009;**392**:1292−302.

62. Chen K, Hu ZM, Song W, Wang ZL, He JB, Shi XM, et al. Diversity of *O*-glycosyltransferases contributes to the biosynthesis of flavonoid and triterpenoid glycosides in *Glycyrrhiza uralensis*. *ACS Synth Biol* 2019;**8**:1858−66.

63. Czjzek M, Cicek M, Zamboni V, R BD, Henrissat B, Esen A. The mechanism of substrate (aglycone) specificity in *β*-glucosidases is revealed by crystal structures of mutant maize *β*-glucosidase-DIM-BOA, -DIMBOAGlc, and -dhurrin complexes. *Proc Natl Acad Sci U S A* 2000;**97**:13555−60.

64. Lu L, Guo L, Wang K, Liu Y, Xiao M. *β*-Galactosidases: a great tool for synthesizing galactose-containing carbohydrates. *Biotechnol Adv* 2020;**39**:107465.