Acta Pharmaceutica Sinica B 2023;13(1):246–255

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

ORIGINAL ARTICLE

# Structural mechanism of a dual-functional enzyme DgpA/B/C as both a *C*-glycoside cleaving enzyme and an *O*- to *C*-glycoside isomerase

**Pengfei He[a,†], Sha Wang[a,†], Sen Li[a], Siqi Liu[a], Shuqi Zhou[a], Jing Wang[a], Jiayue Tao[a], Dongdong Wang[b], Rufeng Wang[a,*], Wenfu Ma[a,*]**

[a]*School of Life Science, Beijing University of Chinese Medicine, Beijing 102488, China*
[b]*DP Technology, Beijing 100080, China*

Received 25 February 2022; received in revised form 21 April 2022; accepted 12 May 2022

**KEY WORDS**

*C*-Glycoside;
*O*-Glycoside;
*C*-Glycoside cleaving enzyme;
Isomerase;
Gut microbiota;
Flavonoid;
Puerarin and oxidoreductase

**Abstract** The *C*-glycosidic bond that connects the sugar moiety with aglycone is difficult to be broken or made due to its inert nature. The knowledge of *C*-glycoside breakdown and synthesis is very limited. Recently, the enzyme DgpA/B/C cascade from a human intestinal bacterium PUE was identified to specifically cleave the *C*-glycosidic bond of puerarin (daidzein-8-*C*-glucoside). Here we investigated how puerarin is recognized and oxidized by DgpA based on crystal structures of DgpA with or without substrate and biochemical characterization. More strikingly, we found that apart from being a *C*-glycoside cleaving enzyme, DgpA/B/C is capable of efficiently converting *O*- to *C*-glycoside showing the activity as a structure isomerase. A possible mechanistic model was proposed dependently of the simulated complex structure of DgpB/C with 3″-oxo-daidzin and structure-based mutagenesis. Our findings not only shed light on understanding the enzyme-mediated *C*-glycosidic bond breakage and formation, but also may help to facilitate stereospecific *C*-glycoside synthesis in pharmaceutical industry.



*Corresponding authors. Tel./fax: +86 10 53912152.
E-mail addresses: wrf@bucm.edu.cn (Rufeng Wang), wenfuma@bucm.edu.cn (Wenfu Ma).
[†]These authors made equal contributions to this work.
Peer review under responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2022.05.022

## 1. Introduction

Flavonoids and their derivatives belong to a family of natural compounds with more than 5000 members[1−3]. They are the most abundant polyphenolic compounds in our daily diets, most of which are in glycoside form that is defined as a molecule comprised of a sugar group and a non-sugar portion through a glycosidic linkage. Based upon the types of glycosidic bonds, they are classified as *O*-, *N*-, *S*- or *C*-glycoside. Flavonoid glycosides that are rich in foods with critical benefits to human health could be hydrolyzed by human gut microbiota into sugar and aglycone[4−8]. More than one hundred families of glycoside hydrolases have been discovered in plants and bacteria, most of which are *O*-glycosidases (CAZy database)[9]. The mechanisms of *O*-glycoside hydrolysis have been clearly elucidated with over 1000 solved structures in database[10]. The cationic transition theory from Koshland mechanism is most popularly appreciated to interpret enzyme-mediated *O*-glycoside hydrolysis[11]. Different from Koshland mechanism, GH4 and GH109 family enzymes cleave 6-phopho-*O*-glycoside using a two-step mechanism including first the oxidation of glucoside 3-hydroxyl group with $NAD^+$ as a cofactor and then cleavage of *O*-glycosidic bond by a pair of catalytic residues and water[12−14].

In comparison with *O*-glycoside, *C*-glycoside is more stable to enzymatic cleavage due to its inert C−C glycosidic bond. Typical examples of *C*-glycoside include puerarin (daidzein-8-*C*-glucoside) (**1a**), vitexin (apigenin-8-*C*-glucoside) (**2a**), orientin (luteolin-8-*C*-glucoside) (**3a**) and homoorientin (luteolin-6-*C*-glucoside) (**3b**, Supporting Information Fig. S1A). Many of these *C*-glycoside compounds are frequently used as herbal drugs in East Asia that show clinical effects against cancer, diabetes, autoimmune diseases and viral infections[15−17].

The enzymatic cleavage of *C*-glycoside recently attracts more interest[18−21]. Historically, the cleavage of *C*-glycosidic bond by human intestinal microbiota was observed decades ago[22,23]. Later, the corresponding *dfg* and *dgp* gene clusters were identified. DfgA-E in *Eubacterium cellulosolvens* could break the *C*-glycosidic bond of homoorientin and isovitexin[24]. And DgpA/B/C from PUE bacterial strain are capable of catalyzing puerarin *C*-glycoside cleavage[19,20]. The cleavage of puerarin *C*-glycoside is triggered by the oxidation of sugar moiety by DgpA that is a NAD(H)-dependent oxidoreductase in Gfo/Idh/MocA family and further cleaved by DgpB/C that is a $Mn^{2+}$-coordinated enzyme complex. DgpA oxidizes C-3″ of puerarin and the product 3″-oxo-puerarin is synergistically cleaved by DgpB/C into daidzein (Supporting Information Fig. S1B and C). Particularly, a recent published structure of homoorientin complexed with AgCGD2$\alpha/\beta$ that is a DgpB/C homolog in soil bacteria illustrated the atomic picture of the oxidized *C*-glycoside homoorientin cleaved by DgpB/C-like enzyme[21]. However, no structural information about DgpA is available and hence the molecular basis for DgpA substrate recognition is unclear. Additionally, the function of DgpA/B/C has yet not been thoroughly characterized.

Here, we first solved the structures of DgpA in substrate-free and substrate-bound forms and then explained the substrate specificity of the partial auto-inhibited DgpA. More strikingly, we found that DgpA/B/C not only breaks the *C*-glycosidic bond as a cleaving enzyme but also converts the *O*- to *C*-glycoside showing the activity as a structural isomerase. Furthermore, a possible working model for DgpA/B/C as a dual-functional enzyme was proposed.

## 2. Materials and methods

### 2.1. Chemicals and reagents

The *C*- and *O*-glycosides including puerarin (Cat. No. B20446), daidzin (Cat. No. B20226), daidzein (Cat. No. B20227), genistin (Cat. No. S31591), genistein (Cat. No. S31565), genistein-8-*C*-glucoside (Cat. No. BP2095) were all purchased from Shanghai Yuanye Bio-Technology (Shanghai, China). Orientin (Cat. No. LD-1125623), vitexin (Cat. No. LD-1125647), apigenin (Cat. No. LD-1109851), luteolin (Cat. No. LD-1131358) and isovitexin (Cat. No. YY91219) were from National Institutes for Food and Drug Control (Beijing, China). The HPLC-grade acetonitrile (Cat. No. A12147) and methanol (Cat. No. M23142) were from Thermo Fisher (Waltham, USA). Acetic acid (Cat. No. A111247) was from Shanghai MACKLIN (Shanghai, China). Manganese chloride (Cat. No. S31112) was from Shanghai Yuanye Bio-Technology (Shanghai, China). $NAD^+$ (Cat. No. N8110), DTT (Cat. No. D8220), maltotriose (Cat. No. M9690) and 2,6-dichloroindophenol (DCPIP) (Cat. No. D6350) were from Solarbio Science and Technology (Beijing, China). Glucose (Cat. No. G8270), sucrose (Cat. No. V900116), lactose (Cat. No. 17814) and soluble starch (Cat. No. V90050) were from Sigma−Aldrich (St. Louis, USA). The purity of chemicals and reagents as above is better at least than 98% except that the purity of DCPIP is better than 97%.

### 2.2. Construction of recombinant plasmids

The amino acid sequences of DgpA (GenBank: BBG22493.1), DgpB (GenBank: BBG22494.1) and DgpC (GenBank: BBG22495.1) were obtained from the gene database at the National Center for Biotechnology Information (NCBI). The DNA coding of DgpA, DgpB or DgpC was synthesized (GENEWIZ, China) after codon optimization. The genes were then cloned into a modified pET-28a vector with a 6 × His and sumo tag at the N-terminus[25]. The mutants of DgpA and DgpB/C were made with a fast mutagenesis kit (Transgen, China). All of the plasmids were confirmed with sequencing (GENEWIZ, China).

### 2.3. Protein production and purification

After plasmids of DgpA, DgpB and DgpC were transformed into *E. coli* BL21 (DE3) competent cells (Invitrogen, USA), the bacteria were cultured in Luria Broth (LB) medium supplemented with 50 μg/mL kanamycin at 37 °C. When the bacterial density ($OD_{600}$) reached 0.6, 0.2 mmol/L $\beta$-D-1-thiogalactopyranoside (IPTG) was added to induce protein production at 16 °C for 16 h. DgpA bacteria were lysed by an ultrasonic cell breaker and the bacteria for production of DgpB and DgpC recombinant proteins were mixed and lysed together. The lysis buffer was 50 mmol/L Tris-HCl, pH 8.0, 0.5 mol/L NaCl, 2 mmol/L 2-mercaptoethanol ($\beta$-ME) and 1 mmol/L phenylmethylsulfonyl fluoride (PMSF). The lysate was centrifuged at 48,000×*g* for 40 min to separate the soluble protein from pellet. Protein was then purified through a HiTrap His column (Cytiva, USA). The washing buffer for His-trap affinity chromatography contained 50 mmol/L Tris, pH 8.0, 0.5 mol/L NaCl and 5 mmol/L $\beta$-ME and the His-trap column elution buffer contained 50 mmol/L Tris, pH 8.0, 0.5 mol/L NaCl, 5 mmol/L $\beta$-ME and 1 mol/L imidazole. The eluted protein from affinity column was dialyzed against a buffer (50 mmol/L Tris-HCl, pH 8.5, 50 mmol/L NaCl and 2 mmol/L

DTT) and the sumo tag was subsequently removed with ULP1. The recombinant protein was further purified through a HiTrap Q column (Cytiva, USA). Finally, protein was purified through a size-exclusive column (the Superdex 200 increase column) (Cytiva, USA) in a buffer of 20 mmol/L Tris-HCl, pH 7.4 and 0.15 mmol/L NaCl. The resulted protein was concentrated to 20 mg/mL, aliquoted, flash frozen with liquid nitrogen and stored at −80 °C for further use. All other proteins used in the study, including mutants of DgpA K100A, R159A, D182A, H186A, Δ(308−316) or Δ(311−316) and mutants of DpgB/C complex with Y11A, H143A or E301A in DgpC were all produced in a similar manner. The yield of proteins as above were all around 20 mg per liter, except that the yield of K100A DgpA mutant was about 5 mg per liter. Protein concentration was determined based on its extinction coefficient at 280 nm, using NanoDrop 2000 (Thermofisher, USA).

### 2.4. CD spectrum data collection

All protein samples including DgpA, DgpB/C and related mutants at concentration of 0.2 mg/mL in PBS buffer (pH 7.5) were used for circular dichroism (CD) spectral data collection. The resulted curve was the average value of three measurements with the removal of PBS background, using Chirascan plus CD spectrometer (Applied Photophysics, UK). The instrument parameters are set as follows: the scanning rate was 50 nm/min, the slit width is 1 nm and the spectral data in the wavelength range of 200−260 nm were finally obtained.

### 2.5. Production and purification of selenoprotein

The plasmids coding for the expression of DgpA, DgpB and DgpC were transformed into *E. coli* B834 (DE3) competent cells (Merck, USA) and the transformed bacteria were cultured in LB supplemented with 50 μg/mL kanamycin at 37 °C. When the bacterial density ($OD_{600}$) reached 1.2, 1 L of the bacterium was pelleted at 4000×*g* centrifugation in a sterile environment at 4 °C. The bacterial cells were washed three times with sterile ultrapure water and then resuspended in 50 mL. The resuspended bacterial cells were cultured in Selenobase medium (AthenaES, USA) that was composed of 21.6 g selenobase in 1 L medium at the ratio of 25 mL resuspended bacteria per liter, supplemented with 100 μg/mL selenomethionine, 5.1 g/L selenonutrition and 50 μg/mL kanamycine. When the $OD_{600}$ of bacterial optical density was 1.2, 0.2 mmol/L IPTG was added to induce protein production at 16 °C for 20 h. The purification process of selenoprotein was in a similar way to native protein purification.

### 2.6. Protein crystallography and structure determination

The purified DgpA and its mutants were concentrated to 20 mg/mL in a buffer containing 20 mmol/L Tris-HCl pH 7.4, 0.15 mol/L NaCl and 1 mmol/L $NAD^+$. DgpA crystals grew in a buffer containing 0.2 mol/L magnesium formate and 20% PEG 3350 by hanging drop method at 16 °C. The crystal grew into full size for 2 days. DgpB/C crystals were grown in a buffer containing 0.2 mol/L magnesium acetate and 10% PEG 8000 for 36 h at concentration of 10 mg/mL. Crystals were first transferred to the soaking solution containing reservoir buffer plus 30% glucose at 37 °C for 30 min, then quickly transferred to a cryoprotective solution containing crystallization buffer plus 25% glycerol and immediately stored in liquid nitrogen for data collection.

X-ray diffraction data were collected against the 0.978 Å wavelength at the BL17U/BL18U1/BL19U1 beamlines of the Shanghai Synchrotron Radiation Facility (SSRF) of the Shanghai National Protein Science Facility (NFPS), China. Diffraction data were processed with the program HKL3000 (Supporting Information Table S1)[26]. Phasing was carried out using an automatic data-processing program Aquarium[27]. Structure refinement was carried out in Phenix[28], firstly by rigid-body refinement, following by positional and individual B factor refinements. Model of β-D-glucose was built in a final step of the structural refinement to minimize the potential phase bias. All resulted parameters for diffraction and refinement are summarized in Supporting Information Table S1.

### 2.7. Substrate docking and molecular dynamics simulation

The initial structure was prepared using the web software Hermite (https://hermite.dp.tech/) with its default parameters. The simulation was conducted using GROMACS-2019 software package[29] with AMBER99SB-ILDN force field[30] and the TIP3P water model. $Na^+$ and $Cl^-$ ions were added to neutralize the system and conferred a salt concentration of 0.15 mol/L. The system was first energy minimized using the steepest descent algorithm, and the added solvent was equilibrated with position restraints on the heavy atoms of the protein. The temperatures were maintained using the V-rescale method[31] with a relaxation time of 0.1 ps. We used the Parrinello-Rahman barostat[29] to keep the pressure at 1 bar with a time constant of 2 ps. The cutoff of electrostatic interactions and van der Waals interactions were both set to be 1.0 nm, and the particle mesh Ewald method[31] was used to treat electrostatic interactions. All bonds were constrained by the LINCS algorithms[32]. The time step was 2 fs and simulation snapshots were stored every 0.1 ns. The last frame of the trajectory was then used for docking with 3″-oxo-daidzin with LeDock that was tested to have the best capabilities to identify the correct ligand binding poses[33]. All parameters were set as default including the conformation method and the docking score function.

### 2.8. Monitoring of DgpA activity with DCPIP assay

The oxidative activity of DgpA enzyme was monitored after incubating 50 μmol/L purified protein wild type DgpA-WT, DgpA D182A, DgpA Δ(308−316) or DgpA Δ(311−316) with 10 mmol/L substrate (puerarin, vitexin, genistin, soluble starch, maltotriose, lactose, sucrose, glucose or maltose) and 0.8 mmol/L 2,6-dichlorophenol indophenol (DCPIP) in PBS buffer. After 20 h, the color shift of DCPIP with the absorbance at 600 nm was recorded by a microplate reader (SpectraMax i3x, Molecular Devices, USA). The amount of DCPIP that is reduced in the reaction was calculated according to the DCPIP reference calibration curve.

### 2.9. Monitoring of DgpA/B/C activity by HPLC

The *C*- or *O*-glycoside cleavage assay was conducted with DgpA/B/C enzymes at 20 μmol/L each subunit in PBS buffer pH 7.5 containing 1 mmol/L $Mn^{2+}$, 1 mmol/L $NAD^+$, 10 mmol/L DTT and 0.1 mmol/L substrate (puerarin, orientin, vitexin, isovitexin, daidzin or genistin). The reaction was carried out at 37 °C in PBS pH 7.5 for 8 h and ended by adding of 75% methanol. The reaction solution was centrifuged at 16,000×*g* for 15 min to remove

the denatured enzymes. After filtered through a 0.22 μm membrane, the products of the reaction were separated and analyzed with a 250 mm × 4.6 mm, 5 μm, TC-C18 column (Agilent, USA) by HPLC (Shimadzu, Japan). The chromatography was carried out using acetonitrile (Sigma, USA) as mobile phase A and 0.1% acetic acid in purified water as mobile phase B at a flow rate of 1 mL/min. The gradient elution program was as follows: 5%−20% A at 0−5 min, 20%−45% A at 5−15 min, 45%−52% A at 15−18 min, 52%−55% A at 18−19 min, and 55%−5% A at 19−23 min. The injection volume was 10 μL and the detection wavelength was set at 265 nm.

HPLC method validation including linearity, repeatability, precision, stability and recovery was performed. Calibration curves were obtained by taking the mass concentration of each compound solution as the abscissa and the peak area as the ordinate. The repeatability and precision were verified using six test solutions prepared according to the same method, and were expressed in the form of relative standard deviations (RSD). The stabilities of the six analyte references were evaluated by analyzing the sample solutions at room temperature at 0, 2, 4, 8, 12, and 24 h. The recoveries of six samples solutions were assessed by comparing the peak areas obtained from test samples with those of peak areas of test samples containing the mixed reference solution.

Each product in the assay was identified by matching its retention time and UV spectrum with those of the corresponding reference compound, and quantified in accordance with the calibration curve for each compound. The optimization of DgpA/B/C enzyme reaction conditions, the inhibition effect of glucose to puerarin *C*-glycoside degradation and potential substrate screening were all carried out in a similar way as above.

The concentration changes of intermediate compounds in DgpA/B/C mediated isomerization reaction were also performed and analyzed as above. The reaction was terminated at 0, 2, 4, 6, 8, 12, 16, 20, 24, 36 and 48 h, respectively. Each 200 μL of the terminated mixture was taken and mixed with 600 μL methanol. The mixture was centrifuged at 16,000×*g* for 15 min to remove proteins, and then the supernatant was taken and filtered through a 0.22 μm membrane for HPLC analysis.

### 2.10. *Determination of kinetic parameters*

A gradient concentration of substrate was included in the DgpA/B/C *C*-glycoside cleaving assay. The concentration of puerarin was set from 0.01 to 4 mmol/L and that of daidzin and genistin was from 0.01 to 2 mmol/L. The reaction was carried out at 37 °C for 8 h under the optimized reaction conditions (*i.e.*, PBS buffer pH 7.5 containing 1 mmol/L $Mn^{2+}$, 1 mmol/L $NAD^+$ and 10 mmol/L DTT). The $K_m$ and $k_{cat}$ kinetic parameters were derived by fitting the calculated reaction rate and substrate concentration to Michaelis−Menten model in Prism5 (GraphPad, USA).

### 2.11. *Characterization of intermediate product with NMR and MS*

To determine the intermediate compound structure, *C*- or *O*-glycoside cleaving assay with DgpA/B/C was carried out in large scale. In a 90 mL reaction assay, 1 mmol/L substrate (puerarin, daidzin or genistin), 1 mmol/L $NAD^+$, 1 mmol/L $Mn^{2+}$, 10 mmol/L DTT and 20 μmol/L each subunit of DgpA/B/C were comprised. The reaction was ended with 75% methanol and the products were purified through Sephadex LH-20 (Merck, USA) column chromatography. The purified intermediate products were dried with a vacuum dryer (Kuansons Bio-Tech, China) and then dissolved in dimethyl sulfoxide-$d_6$ for $^1H$ and $^{13}C$ NMR spectroscopy. The NMR spectra were recorded on a Bruker 400 instrument (Bruker, Germany) at 400 MHz for $^1H$, or a Bruker 700 instrument operated at 175 MHz for $^{13}C$. The purified products from the reaction of daidzin and genistin were dissolved in methanol to obtain the solutions of each compound at 50 μg/mL. After centrifugation at 13,500×*g* for 10 min, the supernatants were collected, and aliquots were analyzed by LC/MS. A Waters MicromassQ-TOF (Waters Corporation, USA) and a Shimadzu UPLC C18 column (2.1 mm × 100 mm, 2 μm) were used for LC/MS assay. The mobile phases and the gradient elution program were the same as those used in HPLC assay, and the injection volume was 5 μL.

## 3. Results

### 3.1. *Partial auto-inhibited configuration of DgpA*

First, we reconstituted the puerarin *C*-glycoside cleaving assay with recombinant DgpA and DgpB/C enzymes after condition optimization including pH, temperature and choice of reductants (Supporting Information Fig. S2). We next solved a 2.7 Å resolution DgpA free enzyme structure using single wavelength anomalous dispersion (SAD) method with a crystal of Se-Met-substituted recombinant enzyme that is in a hexameric configuration, consistent with our size-exclusive chromatographic results and native gel analysis (Fig. 1A, Supporting Information Fig. S3 and Table S1). By contrast, other oxidoreductases in Gfo/Idh/MocA family are either dimer or tetramer (Fig. 1B and C). Two significant features stand out from the DgpA free enzyme structure. First, the substrate recognition pocket is partially blocked by a loop extending from the neighbor DgpA, resulting in an auto-inhibited conformation of the enzyme in hexamer (Fig. 1D and E). Second, a poor difference electron density corresponding to glucose is observed close to the DgpA substrate recognition pocket, suggesting that glucose may be a potential substrate. To verify these, in a 2,6-dichlorophenolindophenol (DCPIP) reporter assay that is based on the color shift of DCPIP from oxidized state in blue to colorless when reduced[34], the inhibited loop deletion mutant increased about 3 fold DgpA oxidative activity to glucose than wild type (Fig. 1F), implying that DgpA activity is partially regulated by the auto-inhibited loop.

### 3.2. *Recognition of substrates by DgpA*

We next solved the crystal structure of DgpA with inhibited loop deletion mutant bound to β-D-glucose and the difference Fourier electron density in the complex map is undoubtedly fit to β-D-glucose (Fig. 2A). In the complex structure, glucose is bound to the DgpA substrate recognition pocket that is composed of H186, K100, D182, R159, E264, F189, F223 and W166 (Fig. 2A and B). Except the C-1 hydroxyl group, all other hydroxyl groups of glucose and the oxygen of 1,4-linkage are specifically recognized by substrate binding pocket, forming a hydrogen bonding network that accommodates to position glucose C-3 hydroxyl group toward the $NAD^+$ active center at a distance of 4.1 Å that is just the right distance for hydride ion ($H^-$) abstraction (Fig. 2A and B)[14]. When

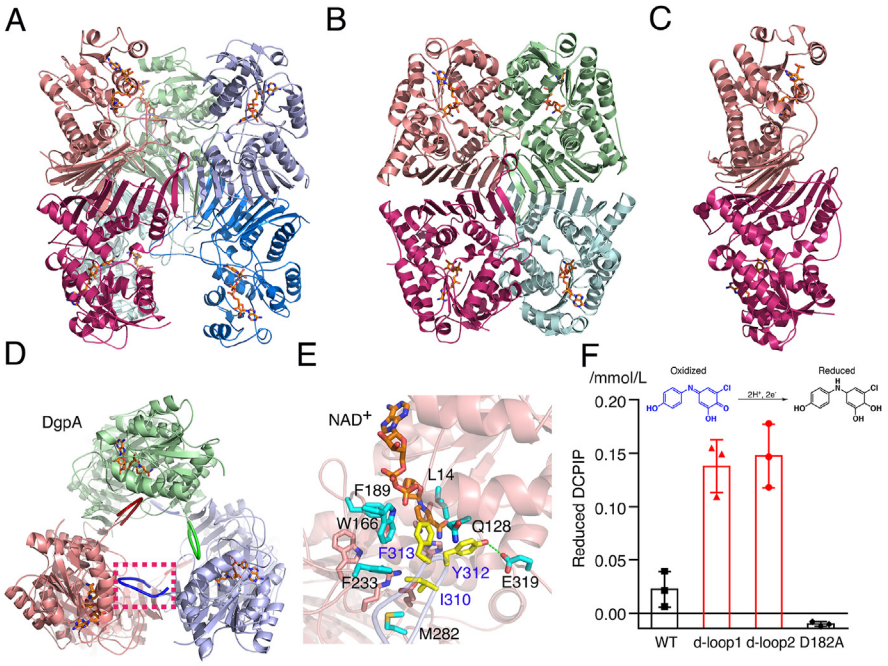



**Figure 1** The crystal structure of DgpA in an auto-inhibited conformation. (A) Free enzyme crystal structure of DgpA in homo-hexameric (PDB ID: 7XRE) and each DgpA subunit is differently colored. $NAD^+$ is displayed in stick and ball model in orange. (B−C) Representative homo-tetrameric structure (B) (PDB ID: 6a3j) or homo-dimeric structure (C) (PDB ID: 5a03) of NAD(H)-dependent oxidoreductase in Gfo/Idh/MocA family. (D) The loop that locks DgpA in an auto-inhibited conformation from the neighboring subunit is highlighted using thick purple dashed box. (E) Close-up view of the inhibiting loop with key residues in yellow bound to DgpA substrate recognition pocket with key residues in cyan. (F) The effect of the loop deletion on the glucose oxidation increase by DgpA in a glucose oxidation experiment. Two loop deletion mutants are in red bar diagram and the wild type or loss function D182A mutant are in black. The DCPIP concentration was calculated from the standard calibration curve of reference DCPIP.

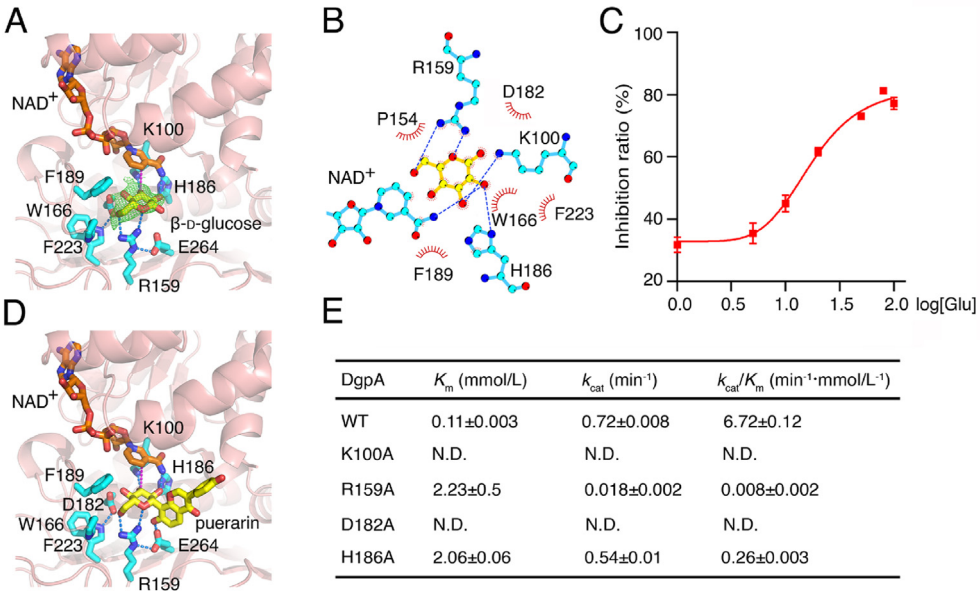
glucose was gradually added into the puerarin *C*-glycosidic bond cleavage assay, the cleaving activity of DgpA/B/C was significantly attenuated proportional to glucose concentration (Fig. 2C), suggesting that glucose and puerarin competitively bind to the same pocket on DgpA.

To further understand puerarin recognition by DgpA, we tried to get the crystal complex structure of DgpA bound to puerarin, but failed. Nevertheless, a composite model of such complex was built based on the complex structure of DgpA and glucose (Fig. 2D). In order to assess the composite model, the DgpA/B/C cleaving activity to puerarin was measured based on series of structure-based DgpA mutants. The kinetic parameters of $k_{cat}$ = 0.72 $min^{-1}$ and $K_m$ = 0.11 mmol/L were measured for wild type DgpA/B/C enzyme (Fig. 2E and Supporting Information Fig. S4). Any single point mutation except H186A in DgpA substrate recognition pocket, including K100A, R159A and D182A that form hydrogen bonds with the glycoside hydroxyl group, led to dramatically cleaving activity loss (Fig. 2E and Supporting Information Fig. S4), indicating that the sugar moiety docking to DgpA is an essential step in the *C*-glycoside cleavage. As Fig. 2A shows that 2-hydroxyl group of glucose is associated with both H186 imidazole nitrogen and $NAD^+$ primary amide to form hydrogen bonds, it may explain why the DgpA H186A mutant still remains pronounced activity (Fig. 2E and Fig. S4). The above mutations in DgpA do not change the conformation integrity of enzymes that was assessed by CD spectra of these mutants, along with wild type DgpA (Supporting Information Fig. S5).

### 3.3. *Isomeric and cleaving activity of DgpA/B/C*

Puerarin is the canonical substrate of DgpA/B/C enzyme as characterized above and also in the previous study[19,20,24,35]. We then ask what are comprised of DgpA/B/C substrate spectrum? To explore this, three types of substrate candidates including oligosaccharides, (iso)flavonoids with *C*- or *O*-coupled glycoside were first tested in the DgpA DCPIP assay. DgpA is cable of oxidizing most of the potential substrates (Supporting Information Fig. S6) that is consistent to our discovery that sugar moiety plays a dominant role in DgpA substrate recognition (Fig. 2A and D). Further, the cleaving activity of DgpA/B/C towards these substrates was quantitatively examined by high performance liquid chromatography (HPLC) analysis. For the three types of substrates we used, no noticed cleaving activity to oligosaccharides was observed. Vitexin (apigenin-8-*C*-glucoside) that also belongs to the *C*-glycoside flavonoid family is also efficiently cleaved by DgpA/B/C just as puerarin (Supporting Information Fig. S7). For *O*-glycoside flavonoids like daidzin or genistin, DpgA/B/C also display a significant catalytic efficiency with $k_{cat}$ (0.25 $min^{-1}$) and $K_m$ (0.07 mmol/L) for daidzin and $k_{cat}$ (0.026 $min^{-1}$) and $K_m$ (0.03 mmol/L) for genistin (Fig. 3A−D).

During daidzin or genistin cleavage, an intermediate compound was discovered by HPLC (Fig. 3A and C). The intermediate compound from daidzin cleavage experiment was subsequently characterized by $^1H$ and $^{13}C$ NMR. After compare with an authentic reference compound[36], we were surprised to find that the intermediate compound was unambiguously identified as




E

| DgpA | $K_m$ (mmol/L) | $k_{cat}$ (min$^{-1}$) | $k_{cat}/K_m$ (min$^{-1}$·mmol/L$^{-1}$) |
|---|---|---|---|
| WT | 0.11±0.003 | 0.72±0.008 | 6.72±0.12 |
| K100A | N.D. | N.D. | N.D. |
| R159A | 2.23±0.5 | 0.018±0.002 | 0.008±0.002 |
| D182A | N.D. | N.D. | N.D. |
| H186A | 2.06±0.06 | 0.54±0.01 | 0.26±0.003 |

**Figure 2** Recognition of sugar moiety by DgpA. (A) Crystal structure of DgpA complexed with glucose (PDB ID: 7XR9). Green lines show Fourier difference electron density map at 2.7 Å resolution, contoured at 2.3$\sigma$. NAD$^+$ is presented in stick mode in orange, key residues in substrate binding pocket are in cyan and glucose is in yellow. (B) 2D diagram presentation of glucose bound to DgpA. Some of the key hydrogen binds are displayed as dashed blue lines. (C) The inhibition ratio of glucose to puerarin DgpA/B/C cleavage normalized to the wild type enzymes. (D) The composite model of puerarin in yellow associated with DgpA. Briefly, the structural alignment of glucose to the glucosyl group of puerarin using PyMOL (www.pymol.com) led to the composition model of DgpA bound to puerarin. (E) Summary of kinetic parameters ($k_{cat}$ and $K_m$) for *C*-glycoside cleavage of puerarin with wild type DgpA/B/C and DgpA substrate binding deficiency mutants.

puerarin (Supporting Information Fig. S8 and NMR data analysis in Fig. S8 caption). Although we failed to get enough pure intermediate material from genistin cleaving reaction for high quality NMR analysis, the corresponding LC/MS data shows molecular weight of the intermediate compound is identical to genistein-8-*C*-glucoside (Supporting Information Fig. S10A and B). To further confirm the identification of these compounds, we compared the intermediate compounds in Fig. 3A and C with the corresponding references that are puerarin and genistein-8-*C*-glucoside respectively *via* LC/MS (Supporting Information Figs. S9 and S10). The high similar pattern of MS profile between intermediate compounds and the references is further indicative of the intermediates as *C*-glycoside (Figs. S9 and S10). Moreover, the concentration change of *O*-glycosides, intermediate

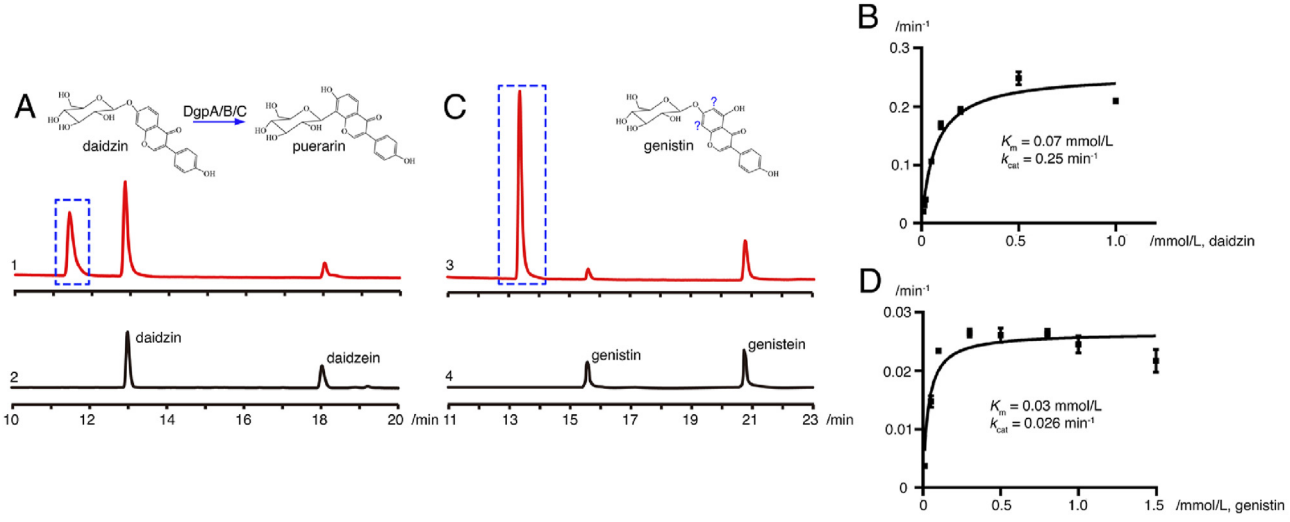



**Figure 3** Intermediate compounds identification during *O*-glycoside hydrolysis with DgpA/B/C. (A) HPLC chromatography of the products from daidzin (daidzein-7-*O*-*β*-D-glucoside) cleavage (1) in red and the reference compounds (daidzin and daidzein) used in the assay in black (2). The reaction scheme is drawn based on the NMR identification in Supporting Information Fig. S8. (B) Kinetic parameters for daidzin cleaved into daidzein and sugar. (C) HPLC analysis of the products from genistin (genistein-7-*O*-*β*-D-glucoside) cleavage (3) in red and the reference compounds (genistin and genistein) used in the assay in black (4). The sugar moiety might be isomerized to either C-6 or C-8 position of the flavone. (D) Kinetic parameters for genistin cleaved into genistein. The intermediate compounds from the reaction are highlighted with dashed blue boxes.

compounds and aglycones in the reaction was monitored in time course. The methodological tests showed that the HPLC method applied to assay the content of 6 compounds (puerarin, daidzin, daidzein, genistein-8-*C*-glucoside, genistin, and genistein) was acceptable (Supporting Information Tables S2 and S3). The concentration of the intermediates first kept rising until the concentration of substrates that is *O*-glycoside daidzin and genistin here significantly dropped. Thereafter, the intermediates were next cleaved by DgpA/B/C enzyme (Supporting Information Fig. S11). Collectively, the only difference in the two compounds structure is that the glucosyl group migrates from 7-*O* in daidzin to 8-*C* in puerarin or genistin to genistein-8-*C*-glucoside (Fig. 3A and C). The conversion of *O*- to *C*-glycoside in daidzin and genistin strongly supports that DgpA/B/C is not only a *C*-glycoside cleaving enzyme but also an *O*- to *C*-glycoside isomerase.

### 3.4. *Modelling structure of DgpB/C bound to 3″-oxo-daidzin*

To further investigate the DgpA/B/C isomeric mechanism, we solved a 2.1 Å resolution crystal structure of DgpB/C complex (Fig. 4A and Supporting Information Table S1) that is in a hetero-octameric configuration consistent to the recent publication[21]. We then tried to model the 3″-oxo-daidzin into the DgpB/C free enzyme structure based on the structure of AgCGD2$\alpha/\beta$ complexed with homoorientin where AgCGD2$\alpha/\beta$ is a DgpB/C homolog in soil bacterium *Arthrobacter globiformis*[21]. 1-ns molecular dynamics (MD) simulation was conducted to relax the experimental structure, in which some of the protein side-chains are poorly positioned. The last frame of the trajectory was then used for docking with the software LeDock[33]. In the simulated DgpB/C bound to substrate structure, 3″-*O*-daidzin is recognized by two patches formed by DgpB/C compelx. The oxidized sugar moiety is specifically bound to $Mn^{2+}$ and the surrounding polarized residues on DgpC that make up the first binding patch (Fig. 4B). The recognition mode of the oxidized sugar moiety is similar to the well-established substrate binding pattern of *O*-glycosidases in GH4 and GH109 family[12–14]. The polarized first patch is proposed to function as the catalytic core region. The flavone group that is daidzein in the case of daidzin is preferentially associated with the second patch that is composed by a cluster of aromatic and hydrophobic residues between DgpC and DgpB interface, suggesting that the substrate selectivity of DgpB/C relies on the second patch (Fig. 4B and C).

### 3.5. *Possible mechanism for DgpA/B/C as an O- to C-glycoside isomerase*

A possible two-step mechanism of DgpA/B/C as a *C*-glycoside structural isomerase is proposed. The breakage of the *O*-glycosidic bond is first initiated with oxidation of glucoside 3-hydroxyl group by DgpA complexed with $NAD^+$. The resulted product 3″-oxo-daidzin will be then bound to DgpB/C complex (Fig. 4A–C). The deprotonation of 2-carbon on glucoside by H143 accommodates the cleavage of *O*-glycosidic bond with the catalytic residues of H143 and Glu301[11,13,14] (Fig. 4B–C and Supporting Information Fig. S12 a–c). Further, the activation of the aromatic 8-carbon nucleophile is coupled with the deprotonation of adjacent phenyl group both by H143 and E301. The attack of the 1-carbon on glucoside that is associated with DgpB/C in an intermediate state to the partially negative charged aromatic 8-carbon leads to the formation of puerarin (daidzein-8-*C*-glucoside, Supporting Information Fig. S12 d–f). The process of *C*-glycoside isomerization by DgpA/B/C highly resembles the extensively studied mechanism of *C*-glycoside transferases in plants and bacteria[37–39] (Fig. S12).

To assess this model, the structure-based DgpC mutations were made to examine the cleaving and isomeric activity using *C*-glycoside puerarin and *O*-glycoside daidzin as substrates, respectively. All the mutations don't change the conformational integrity of DgpB/C enzyme complex, which is verified by CD spectra (Supporting Information Fig. S13). The mutation of H143A that is located at the first patch and Y11A that is one of the key residues from the second patch dramatically reduce both the cleaving and isomeric activity, consistent to the fact that both of them are key residues either for catalysis or for substrate recognition (Fig. 4A–I). An interesting mutant of DgpC E301A that is also localized on the first patch displayed pronounced inhibition on *C*-glycoside cleavage, but showed little effect on *O*-glycoside cleavage or isomerization, suggesting that *O*-glycoside hydrolysis are required less catalytic residues than *C*-glycoside and the breakage of *O*-glycosidic bond is mechanistically a prerequisite for *C*-glycosidic bond formation (Fig. 4D–I).

## 4. Discussion

One of the basic building blocks in chemical world is constructed with C–C bond that is difficult to be broken or made in physiological condition. Currently, only very few *C*-glycoside cleaving enzymes have been identified[18–21,24]. As one of the *C*-glycoside cleaving enzyme, the biochemical characterization of DgpA/B/C enzyme is only partially investigated. DgpA participates in the process by abstracting hydride ion from glucoside 3-hydroxyl group that may lessen the stability of the *C*-glycosidic bond. However, structural biology information about this process remains unsolved. In this study, we first carried out the structural characterization of DgpA in substrate free and complex state. Our structures and DgpA activity assays explain the mechanism of DgpA substrate specificity (Figs. 1 and 2). Despite of as a *C*-glycoside cleaving enzyme, our enzymatic activity assays and the intermediate compound identification by NMR clearly confirmed that DgpA/B/C enzymes are also an *O*- to *C*-glycoside isomerase, leading to a possible two-step isomerization working model that involves the breakage of *O*-glycosidic bond and electrophilic substitution on aromatic hydrogen (Fig. S12).

The primary function of isomerase is to transform a molecule between its isomers that can be categorized as stereoisomerism and structural isomerism. Stereoisomerism retains molecular bond connection but performs as an enantiomer inverter. By contrast, structural isomerism rearranges the intramolecular bond connection[40,41]. Because isomeric reaction rate is faster and the regional specificity is much higher than traditional chemical synthesis methods, isomerase enzymes are widely used in industry. For instance, glucose isomerase (GI) that converts glucose into fructose is one of the most widespread enzymes in industry for fructose manufacturing[42–45].

Due to the better stability of *C*-glycoside than *O*-glycoside, *C*-glycosides are harnessed as the surrogates for native *O*-glycoside drugs in clinics[46–48]. Although recent years witnesses a significant progress in the field of *C*-glycoside chemical synthesis, it is still strongly hampered by poor stereoselectivity and laborious protection and deprotection of functional groups[49,50]. Conversely, enzyme-mediated *C*-glycoside synthesis displays a high efficiency with great regiospecificity[42,45]. Moreover, DgpA/B/C-like enzymes may prevail in the bacterial kingdom like human gut microbiota and soil bacteria and thus they may be a rich source for

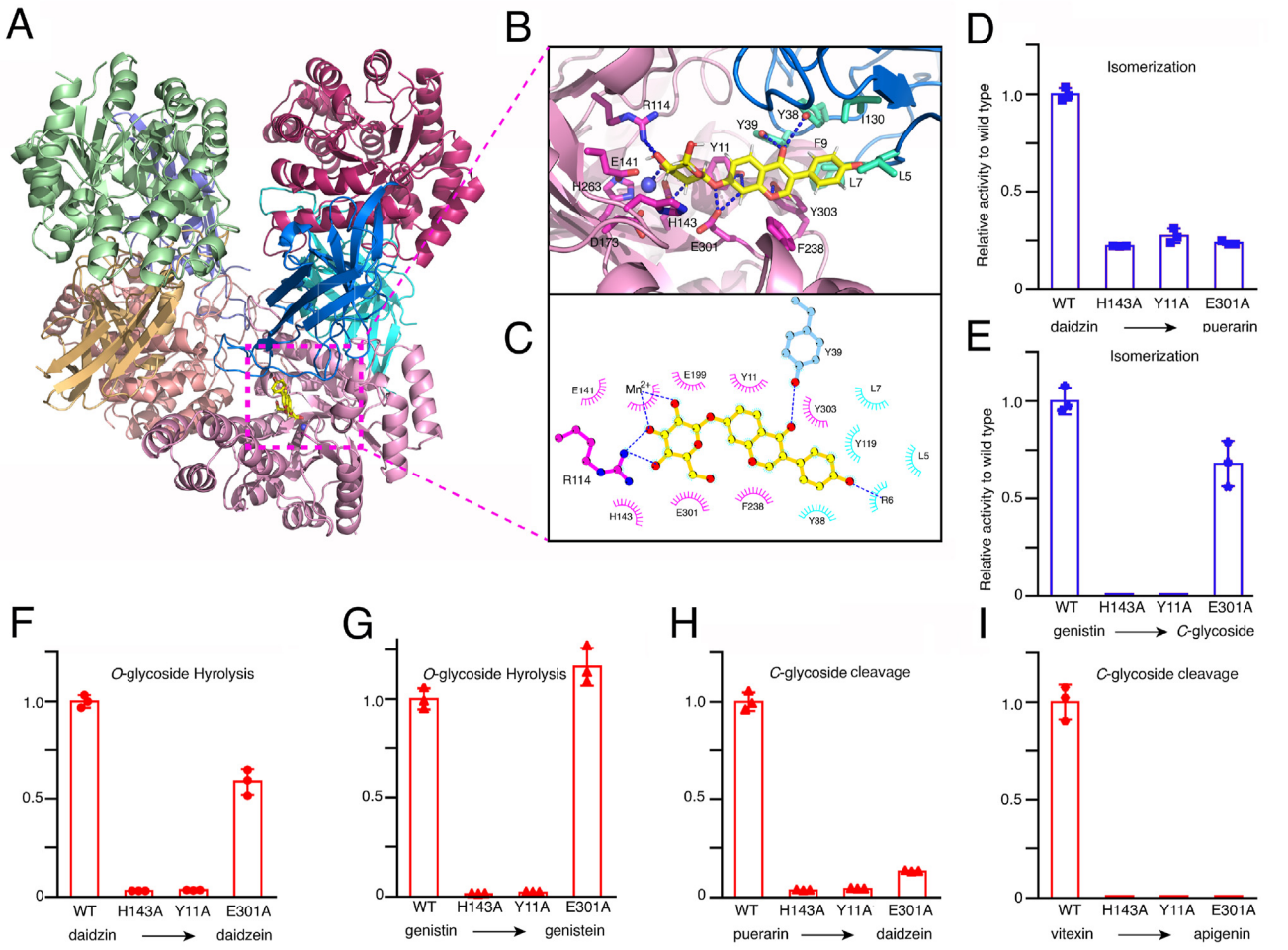



**Figure 4** The modelled structure of DgpB/C complexed with 3″-oxo-daidzin and isomeric or cleaving activity regulation by structure-based DgpC mutagenesis. (A) Structure of DgpB/C complexed with 3″-oxo-daidzin. The coordinates for structure of DgpB/C complex is deposited under the PDB ID as 7XRF. (B) Interaction details of 3″-oxo-daidzin with DgpB/C. 3″-oxo-daidzin is shown as stick in yellow. Key DgpC residues in the pocket are in cyan and DgpB residues are in blue. (C) 2D view of (B). (D—I) Cleavage or isomerization activity regulated by mutations on DgpC including H143A, Y11A or E301A. Reaction was carried out at optimized conditions for 4 h and the products were further analyzed with HPLC purification. (D, E) The blue bar diagrams are displayed as normalized isomerization activity against wild type DgpA/B/C enzyme for daidzin (D) and genistin (E). (F, G) The bar diagrams representing the *O*-glycoside hydrolysis for daidzin (F) and genistin (G) are in red. (H, I) *C*-Glycoside cleavage for puerarin (H) and vitexin (I) are also in red bar diagram.

isomerase-mediated *C*-glycoside synthesis with the synthetic biology strategies[19–21,24].

Currently, *C*-glycoside transferases (CGTs) are the only known enzymes that are able to devise *C*-glycosides in plants and bacteria. The sugar moiety that is activated in the nucleotide-diphospho (NDP) bound form is made a bond to the carbon on the flavonoid aromatic ring by CGTs. A highly conserved catalytic dyad motif is located in active pockets of almost all CGTs. Particularly, E396 and H24 in the structure of TcCGT1 that is a flavone-8-*C*-glycosyltransferase from *Trollius chinensis*[39] and D58 and E316 in SsfS6 that selectively adds a sugar group to C-9 position of the anhydrotetracycline aglycone[37,51] are both the critical residues for CGT activity. Fig. 4 shows the H143 and E301 on DgpC play a comparable role as the catalytic dyad residues in CGTs from the perspectives of catalytic pocket structures and enzyme activity. When the structure of DgpC was searched against the protein structure database using Dali server, sugar isomerases are the top hits with RMSD around 2.5 Å, both of which share a typical TIM barrel-like domain and a divalent cation at the catalytic center[52]. Together, our study here revealed a unique family of enzymes in nature that not only breaks the *C*-glycosidic bond, but also make it from *O*-glycosidic bond.

## Acknowledgments

This study was supported by grants from National Natural Science Foundation of China (No. 81073018 and 81274044) to Rufeng Wang and Startup fund program at Beijing University of Chinese Medicine (90011451310011) and key research fund for drug discovery in Chinese medicine at Beijing University of Chinese Medicine (1000061223476) to Wenfu Ma. We thank the staffs from BL17B1/BL18U1/BL19U1/BL19U2/BL01B beamline of National Facility for Protein Science in Shanghai (NFPS) at Shanghai Synchrotron Radiation Facility, for assistance during data collection. We also thank all the WANG and MA lab members for helpful discussions.

## Author contributions

Rufeng Wang and Wenfu Ma designed research; Pengfei He, Sha Wang, Siqi Liu, Shuqi Zhou, Jing Wang and Jiayue Tao performed research; Dongdong Wang performed the computational simulation; all the authors analyzed data; and Wenfu Ma wrote the paper.

## Conflicts of interest

The authors declare no conflicts of interest.

## Appendix A. Supporting information

Supporting data to this article can be found online at https://doi.org/10.1016/j.apsb.2022.05.022.

## References


1. Kumar S, Pandey AK. Chemistry and biological activities of flavonoids: an overview. *Sci World J* 2013;**2013**:162750.
2. Nijveldt RJ, van Nood E, van Hoorn DE, Boelens PG, van Norren K, van Leeuwen PA. Flavonoids: a review of probable mechanisms of action and potential applications. *Am J Clin Nutr* 2001;**74**:418–25.
3. Falcone Ferreyra ML, Rius SP, Casati P. Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front Plant Sci* 2012;**3**:222.
4. Braune A, Blaut M. Bacterial species involved in the conversion of dietary flavonoids in the human gut. *Gut Microb* 2016;**7**:216–34.
5. Koppel N, Maini Rekdal V, Balskus EP. Chemical transformation of xenobiotics by the human gut microbiota. *Science* 2017;**356**:eaag2770.
6. Riva A, Kolimar D, Spittler A, Wisgrill L, Herbold CW, Abranko L, et al. Conversion of rutin, a prevalent dietary flavonol, by the human gut microbiota. *Front Microbiol* 2020;**11**:585428.
7. Xiao J, Capanoglu E, Jassbi AR, Miron A. Advance on the flavonoid *C*-glycosides and health benefits. *Crit Rev Food Sci Nutr* 2016; **56**(Suppl 1):S29–45.
8. Bililign T, Griffith BR, Thorson JS. Structure, activity, synthesis and biosynthesis of aryl-*C*-glycosides. *Nat Prod Rep* 2005;**22**:742–60.
9. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* 2022;**50**:D571–7.
10. Berman H, Henrick K, Nakamura H. Announcing the worldwide protein data bank. *Nat Struct Biol* 2003;**10**:980.
11. Koshland Jr DE. Stereochemistry and the mechanism of enzymatic reactions. *Biol Rev* 1953:416–36.
12. Rajan SS, Yang X, Collart F, Yip VL, Withers SG, Varrot A, et al. Novel catalytic mechanism of glycoside hydrolysis based on the structure of an $NAD^+/Mn^{2+}$-dependent phospho-alpha-glucosidase from. *Bacillus subtilis. Structure* 2004;**12**:1619–29.
13. Yip VL, Withers SG. Family 4 glycosidases carry out efficient hydrolysis of thioglycosides by an alpha,beta-elimination mechanism. *Angew Chem Int Ed Engl* 2006;**45**:6179–82.
14. Yip VL, Varrot A, Davies GJ, Rajan SS, Yang X, Thompson J, et al. An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 beta-glycosidase from Thermotoga maritima. *J Am Chem Soc* 2004;**126**:8354–5.
15. He M, Min JW, Kong WL, He XH, Li JX, Peng BW. A review on the pharmacological effects of vitexin and isovitexin. *Fitoterapia* 2016; **115**:74–85.
16. Lam KY, Ling AP, Koh RY, Wong YP, Say YH. A Review on medicinal properties of orientin. *Adv Pharmacol Sci* 2016:4104595. 2016.
17. Zhou YX, Zhang H, Peng C. Puerarin: a review of pharmacological effects. *Phytother Res* 2014;**28**:961–75.
18. Kumano T, Hori S, Watanabe S, Terashita Y, Yu HY, Hashimoto Y, et al. FAD-dependent *C*-glycoside-metabolizing enzymes in microorganisms: screening, characterization, and crystal structure analysis. *Proc Natl Acad Sci U S A* 2021;**118**:e2106580118.
19. Nakamura K, Zhu S, Komatsu K, Hattori M, Iwashima M. Deglycosylation of the isoflavone *C*-glucoside puerarin by a combination of two recombinant bacterial enzymes and 3-oxo-glucose. *Appl Environ Microbiol* 2020;**86**:e00607–20.
20. Wei B, Wang YK, Qiu WH, Wang SJ, Wu YH, Xu XW, et al. Discovery and mechanism of intestinal bacteria in enzymatic cleavage of C–C glycosidic bonds. *Appl Microbiol Biotechnol* 2020;**104**: 1883–90.
21. Mori T, Kumano T, He H, Watanabe S, Senda M, Moriya T, et al. *C*-Glycoside metabolism in the gut and in nature: identification, characterization, structural analyses and distribution of C–C bond-cleaving enzymes. *Nat Commun* 2021;**12**:6294.
22. Hattori M, Kanda T, Shu YZ, Akao T, Kobashi K, Namba T. Metabolism of barbaloin by intestinal bacteria. *Chem Pharm Bull (Tokyo)* 1988;**36**:4462–6.
23. Hattori M, Shu YZ, el-Sedawy AI, Namba T, Kobashi K, Tomimori T. Metabolism of homoorientin by human intestinal bacteria. *J Nat Prod* 1988;**51**:874–8.
24. Braune A, Engst W, Blaut M. Identification and functional expression of genes encoding flavonoid *O*- and *C*-glycosidases in intestinal bacteria. *Environ Microbiol* 2016;**18**:2117–29.
25. Wang Z, Song J, Milne TA, Wang GG, Li H, Allis CD, et al. Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression. *Cell* 2010;**141**: 1183–94.
26. 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 D Biol Crystallogr* 2006;**62**:859–66.
27. Yu F, Wang QS, Li MJ, Zhou H, Liu K, Zhang KH, et al. Aquarium: an automatic data-processing and experiment information management system for biological macromolecular crystallography beamlines. *J Appl Crystallogr* 2019;**52**:472–7.
28. Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* 2019;**75**:861–77.
29. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 2013;**29**:845–54.
30. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 2010;**78**:1950–8.
31. York DM, Wlodawer A, Pedersen LG, Darden TA. Atomic-level accuracy in simulations of large protein crystals. *Proc Natl Acad Sci U S A* 1994;**91**:8715–8.
32. Hess B, Bekker H, Berendsen H, Fraaije J. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 1997;**18**:1463–72.
33. Wang Z, Sun H, Yao X, Li D, Xu L, Li Y, et al. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys Chem Chem Phys* 2016;**18**:12964–75.
34. Jahn B, Jonasson NSW, Hu H, Singer H, Pol A, Good NM, et al. Understanding the chemistry of the artificial electron acceptors PES, PMS, DCPIP and Wurster's blue in methanol dehydrogenase assays. *J Biol Inorg Chem* 2020;**25**:199–212.
35. Nakamura K, Zhu S, Komatsu K, Hattori M, Iwashima M. Expression and characterization of the human intestinal bacterial enzyme which cleaves the *C*-glycosidic bond in 3″-oxo-puerarin. *Biol Pharm Bull* 2019;**42**:417–23.
36. Yu L, Gao F, Yang L, Xu L, Wang Z, Ye H. Biotransformation of puerarin into puerarin-6″-*O*-phosphate by Bacillus cereus. *J Ind Microbiol Biotechnol* 2012;**39**:299–305.
37. Wang F, Zhou M, Singh S, Yennamalli RM, Bingman CA, Thorson JS, et al. Crystal structure of SsfS6, the putative *C*-glycosyltransferase involved in SF2575 biosynthesis. *Proteins* 2013;**81**:1277–82.

38. Mittler M, Bechthold A, Schulz GE. Structure and action of the C–C bond-forming glycosyltransferase UrdGT2 involved in the biosynthesis of the antibiotic urdamycin. *J Mol Biol* 2007;**372**:67–76.
39. 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 Engl* 2019;**58**:11513–20.
40. Martinez Cuesta S, Furnham N, Rahman SA, Sillitoe I, Thornton JM. The evolution of enzyme function in the isomerases. *Curr Opin Struct Biol* 2014;**26**:121–30.
41. Martinez Cuesta S, Rahman SA, Thornton JM. Exploring the chemistry and evolution of the isomerases. *Proc Natl Acad Sci U S A* 2016;**113**:1796–801.
42. Bhosale SH, Rao MB, Deshpande VV. Molecular and industrial aspects of glucose isomerase. *Microbiol Rev* 1996;**60**:280–300.
43. Hartley BS, Hanlon N, Jackson RJ, Rangarajan M. Glucose isomerase: insights into protein engineering for increased thermostability. *Biochim Biophys Acta* 2000;**1543**:294–335.
44. Dai C, Miao T, Hai J, Xiao Y, Li Y, Zhao J, et al. A novel glucose isomerase from *Caldicellulosiruptor bescii* with great potentials in the production of high-fructose corn syrup. *BioMed Res Int* 2020;**2020**:1871934.
45. Jin LQ, Jin YT, Zhang JW, Liu ZQ, Zheng YG. Enhanced catalytic efficiency and thermostability of glucose isomerase from *Thermoanaerobacter ethanolicus via* site-directed mutagenesis. *Enzym Microb Technol* 2021;**152**:109931.
46. Chao EC, Henry RR. SGLT2 inhibition—a novel strategy for diabetes treatment. *Nat Rev Drug Discov* 2010;**9**:551–9.
47. Compain P, Martin OR. Carbohydrate mimetics-based glycosyltransferase inhibitors. *Bioorg Med Chem* 2001;**9**:3077–92.
48. Yang G, Schmieg J, Tsuji M, Franck RW. The *C*-glycoside analogue of the immunostimulant alpha-galactosylceramide (KRN7000): synthesis and striking enhancement of activity. *Angew Chem Int Ed Engl* 2004;**43**:3818–22.
49. Yang Y, Yu B. Recent Advances in the chemical synthesis of *C*-glycosides. *Chem Rev* 2017;**117**:12281–356.
50. Cavezza A, Boulle C, Gueguiniat A, Pichaud P, Trouille S, Ricard L, et al. Synthesis of pro-xylane: a new biologically active *C*-glycoside in aqueous media. *Bioorg Med Chem Lett* 2009;**19**:845–9.
51. Li L, Wang P, Tang Y. *C*-Glycosylation of anhydrotetracycline scaffold with SsfS6 from the SF2575 biosynthetic pathway. *J Antibiot (Tokyo)* 2014;**67**:65–70.
52. Holm L. Using Dali for protein structure comparison. *Methods Mol Biol* 2020;**2112**:29–42.