Structure-guided engineering of the substrate specificity of a fungal β-glucuronidase toward triterpenoid saponins

Bo Lv‡1,1, Hanli Sun§1, Shen Huang‡, Xudong Feng‡, Tao Jiang§, and Chun Li‡2

From the ‡Department of Biochemical Engineering/Institute for Biotransformation and Synthetic Biosystem, School of Chemistry and Chemical Engineering, Beijing Institute of Technology, 100081 Beijing, China and §Institute of Biophysics, Chinese Academy of Science, 100101 Beijing, China

Received for publication, June 24, 2017, and in revised form, November 6, 2017. Published, Papers in Press, November 16, 2017, DOI 10.1074/jbc.M117.801910

Edited by Joseph Jez

This article contains interest with the contents of this article.

The atomic coordinates and structure factors (code SC71) have been deposited in the Protein Data Bank (http://wwpdb.org/).

*1 Both authors contributed equally to this work.

*2 To whom correspondence should be addressed. Tel./Fax: 86-01-68913273; E-mail: lichun@bit.edu.cn.

Glycosid hydrolases (GHs) have attracted special attention in research aimed at modifying natural products by partial removal of sugar moieties to manipulate their solubility and efficacy. However, these modifications are challenging to control because the low substrate specificity of most GHs often generates undesired by-products. We previously identified a GH2-type fungal β-glucuronidase from Aspergillus oryzae (PGUS) exhibiting promiscuous substrate specificity in hydrolysis of triterpenoid saponins. Here, we present the PGUS structure, representing the first structure of a fungal β-glucuronidase, and that of an inactive PGUS mutant in complex with the native substrate glucyrrhetic acid 3-O-mono-β-glucuronic acid (GAMG). PGUS displayed a homotetramer structure with each monomer comprising three distinct domains: a sugar-binding, an immunoglobulin-like β-sandwich, and a TIM barrel domain. Two catalytic residues, Glu414 and Glu505, acted as acid/base and nucleophile, respectively. Structural and mutational analyses indicated that the GAMG glycan moiety is recognized by polar interactions with nine residues (Asp962, His435, Asp414, Tyr469, Tyr473, Asp505, Arg563, Asn567, and Lys597) and that the aglycone moiety is recognized by aromatic stacking and by a π interaction with the four aromatic residues Tyr469, Phe470, Trp572, and Tyr473. Finally, structure-guided mutagenesis to precisely manipulate PGUS substrate specificity in the biotransformation of glycyrrhizin into GAMG revealed that two amino acids, Ala365 and Arg563, are critical for substrate specificity. Moreover, we obtained several mutants with dramatically improved GAMG yield (>95%). Structural analysis suggested that modulating the interaction of β-glucuronidase simultaneously toward glycan and aglycone moieties is critical for tuning its substrate specificity toward triterpenoid saponins.

Saccharides and their conjugates are one of the most abundant types of biomacromolecules in nature (1), and they perform critical roles as fundamental components, energy sources, intermediates, and cell recognition signals (2). The carbohydrate-related bioprocesses involving saccharides usually require the participation of a diverse group of glycoside hydrolases (GHs) (3) to cleave glycosidic bonds, thus yielding various functional derivatives (3, 4). Glycosides are an important class of glycoconjugates that are mainly derived from plants and microorganisms (5), and an enormous diversity of sugar moieties has been observed to attach to the aglycones in carbon frameworks, such as cyanohydrins (6), terpenoids (7), phenolics (8), and alkaloids (9). Over the past several decades, many uses of glycosides have been discovered, including the prevention of cellular damage, defense of plants against biotic stresses, facilitation of sugar storage and selective intra- and intercellular transport, and the treatment of human diseases (10–12). Triterpenoid saponins, which are often derived from traditional medicinal plants (such as licorice, ginseng, and astragalus), have recently attracted substantial attention because of their pharmacological activities, which include antitumor, anti-inflammatory, and antiviral activities (13–15). However, in preliminary clinical trials, the solubility and associated side effects of triterpenoid saponins have largely limited their intake efficiency; these problems are usually solved through partially removing a sugar moiety with GHs. For example, glycyrrhizin (GL), a major triterpenoid saponin, exhibits marked toxicity, known as pseudaldosteronism, and poor ability to cross membranes (16). However, when the distal sugar moiety is hydrolyzed, GL is transformed into glycyrrhetic acid (GA) 3-O-mono-β-d-glucuronic acid (GAMG), which has better safety and bioavailability and similar pharmacological activity (Fig. 1) (17, 18). However, most GHs have the drawback of low substrate specificity toward triterpenoid saponins, thus leading to uncontrollable processes and many undesired by-products and significantly impeding their application in modifying triterpenoid saponins.

β-Glucuronidases (GU/S; EC 3.2.1.31), which belong to GH families GH1, GH2, and GH79, cleave glucuronic acid sugar moieties from the non-reducing termini of glycosides and have

3 The abbreviations used are: GH, glycoside hydrolase; GAMG, glycyrrhetic acid 3-O-mono-β-glucuronide; TIM, triose-phosphate isomerase; GL, glycyrrhizin; GA, glycyrrhetic acid; GUS, β-glucuronidase; HGUS, GUS from human; EGUS, GUS from E. coli; PGUS, GUS from A. oryzae Li-3; HCE, high-level constitutive expression; Amp, ampicillin.
Substrate specificity of β-glucuronidase

been found in a variety of hosts, including bacteria, fungi, plants, and humans (19–22). To date, crystal structures of GUS from human (HGUS), *Escherichia coli* (EGUS), and Firmicutes in GH2 and GUS from *Acidobacterium capsulatum* in GH79 have been reported (23–26); however, GUS from fungi has not yet been solved. In addition, little progress has been made toward understanding the precise substrate recognition mechanism of GUS. The only relevant previous study has been reported by Wallace *et al.* (24) who obtained the complex structure of GUS with its inhibitor (i.e., a glycan analogue) and found that the glycan moiety is recognized by GUS through a series of polar interactions. However, how GUS recognizes and interacts with its natural substrate remains unclear, especially for the aglycone moiety, which is primarily responsible for GUS’s low substrate specificity.

Previously, we have identified a fungal GUS from *Aspergillus oryzae* Li-3 (PGUS) and heterologously produced it in *E. coli*. PGUS has strict glycan specificity because it can hydrolyze only substrates containing glucuronide groups. In contrast, PGUS does not show strict specificity toward the aglycone moiety and instead shows similar activities toward both GL and GAMG, and its final product is undesired GA (Fig. 1). These properties make PGUS an ideal model to address the substrate recognition mechanism and substrate specificity of GUS.

In this study, we determined the crystallographic structure of PGUS, which is the first structure of GUS from fungi reported to date. We also produced an inactive form (E414D/E505D) of PGUS by site-directed mutagenesis and used it to obtain crystals of the enzyme-substrate (GAMG) complex to elucidate the substrate recognition mechanism. On the basis of the results, we performed structure-guided mutagenesis to precisely manipulate the substrate specificity of PGUS, thus providing a deeper understanding of the catalytic mechanism. This study provides new insights into the relationship between the structure and function of GUS, which may facilitate the use of GUS for the functional modification of glycosides.

**Results**

*The crystal structure of PGUS*

PGUS was heterologously produced in *E. coli* and then purified by Ni²⁺-affinity chromatography, anion-exchange chromatography, and gel filtration. To facilitate purification, PGUS was fused with a His₆ tag at its N terminus, and we verified that the His tag had no obvious effects on the specific activity of PGUS. The molecular mass of PGUS was estimated to be 290 kDa on the basis of ultracentrifugation (Fig. S1). The substrate specificity of PGUS heterologously produced in *E. coli* was investigated with a focus on both glycan and aglycone moieties. As shown in Table S1, PGUS showed strict glycan specificity and hydrolyzed only the artificial substrate containing glucuronide groups (p-nitrophenol-β-D-glucuronide (pNPGa)). In contrast, PGUS did not show strict specificity toward the aglycone moiety and instead exhibited activity toward all the natural and artificial substrates (i.e., GL, GAMG, and p-nitrophenol-β-D-glucuronide) with *k*ₐᵋ/*k*ₘ values ranging from 1.93 to 127.56 mm⁻¹ s⁻¹.

The purified PGUS (purity >95%) was subsequently subjected to crystallization and X-ray assay. The structure of PGUS was solved by molecular replacement at 3.1-Å resolution with final *R*ₜₒₒₑₑ and *R*ₜₒₑₑ factors of 19.76 and 25.36%, respectively (Table 1). The final model of PGUS consists of a homotetramer containing two asymmetric units arranged with a 35° angle between them (Fig. S2). Each asymmetric unit contains two identical monomers (e.g., chains A and B) composed of 593 ordered amino acid residues. The overall conformation of the asymmetric unit resembles two fists locked together (Fig. 2a). Each monomer comprises three distinct domains: a sugar-binding domain, an immunoglobulin-like β-sandwich domain, and a triose-phosphate isomerase (TIM) barrel domain (Fig. 2b); these are the principal structural features of GH2. The 180 N-terminal residues, including 12 parallel β-strands and two parallel α-helices, resemble the sugar-binding domain of GH2

![Figure 1. Nonspecific hydrolysis of GL to GAMG and GA by GUSs (wild type) without selectivity due to the low specificity toward aglycone moieties.](image-url)
in Fig. 3, is the first complete crystal structure of GUS complexed with its natural substrate. In the PGUS-GAMG complex, the active site pocket forms a sufficiently large space to accommodate the bulky GAMG molecule. The glycan moiety is completely buried at the bottom of the active site, whereas the aglycone moiety is located partially outside of the cleft and hence is more exposed to the solvent and has fewer interactions with PGUS.

The glucuronic bond is in close proximity to the catalytic residues of sites 414 and 505 for hydrolysis. The GAMG molecule experiences distortion caused by the interaction with PGUS with the glycan residue bending almost in the same plane with the catalytic residues but in the opposite direction. The glycan is in a different plane with the aglycone moiety with an angle of 114.3°. For the aglycone, the terminal ring was also distorted by around 120° due to the interaction with neighboring residues.

Loops A and B, specifically nine residues (Asp414, Asp505, Asp416, His532, Arg534, Asn557, Asn558, Tyr469, and Tyr473), directly participate in recognizing the glycan moiety of GAMG through a series of polar interactions, including hydrogen bond and electrostatic interactions (Fig. 3). Leaving the two mutated catalytic sites, Asp414 and Asp505 intact, we introduced extensive mutations to the other seven residues to investigate their functional roles in the catalysis of PGUS.

When Asp416 was mutated to hydrophobic residues (leucine or isoleucine), PGUS was completely deactivated. When Asp416 was mutated to a residue with shorter side chains (glycine or alanine), PGUS lost most of its activity (i.e. more than 90%). Asn557 was highly conserved, and mutation of Asp416 and Asn557 led to the complete deactivation of PGUS. Interestingly, Arg563 was also an important residue responsible for glycan moiety recognition and formed two hydrogen bonds with the hydroxyl and carboxyl groups of the glycan. As expected, most mutations of Arg563 also caused PGUS deactivation. However, two mutants showed different catalytic behavior toward GL and GAMG; specifically, the retained activity of mutant R563K toward GL was 26% higher than that toward GAMG, and the retained activity of mutant R563Q toward GL was 4% higher than that toward GAMG. Therefore, we suspected that Arg563 might play an important role in determining the substrate specificity of PGUS toward GL and GAMG. The extensive polar interaction network between the enzyme and the glycan may explain the strict specificity of PGUS toward the glycan moiety (Fig. 3) in agreement with most GHs from different families in which rigorous specificity toward glycan has been reported (25, 29).

In contrast to the glycan moiety, the aglycone moiety of GAMG does not form any polar interactions with surrounding amino acids. However, the aglycone is surrounded by several aromatic residues, including Tyr469, Phe470, Trp472, and Tyr473 (Fig. 3b). Notably, Tyr469 and Tyr473 play dual functional roles in recognizing both the glycan and aglycone moieties of PGUS. We verified the functional roles of Phe470 and Trp472 through substantial mutation analysis. Generally, Phe470 and Trp472 are highly conserved, and when they were mutated to hydrophilic residues, PGUS lost most of its activity. Therefore, it can be concluded that the aglycone is recognized through aromatic stacking and π interactions with the neighboring aromatic resi-

### Table 1

| Data collection and refinement statistics | PGUS | PGUS-GAMG |
|-----------------------------------------|------|-----------|
| **Data collection and processing**      |      |           |
| X-ray source                           | SSRF BL17U | SSRF BL17U |
| Wavelength (Å)                         | 1.29996 | 0.97922 |
| Space group                            | P6,22 | P1 |
| Resolution range (Å)                   | 50.0–3.1 | 50.0–2.61 |
| Temperature (K)                        | 100   | 100 |
| Unique reflections                     | 59,403 | 92,474 |
| Rmerge (%)                             | 17.59 | 19.31 |
| Rwork (%)                              | 19.11 | 5.9 |
| Completeness (%)                       | 99.89 | 98.2 |
| Redundancy (%)                         | 12.8  | 3.9 |
| **Data refinement**                    |      |           |
| Resolution (Å)                         | 30.0–3.1 | 46.23–2.62 |
| Rwork (%)                              | 19.76 | 17.61 |
| Rmerge (%)                             | 25.36 | 24.49 |
| r.m.s.d. for bonds (Å)                 | 0.011 | 0.014 |
| r.m.s.d. for angles (°)                | 1.45  | 1.75 |
| Average B-factor (Å²)                  | 46.70 | 16.50 |
| Ramachandran plot (%)                  | 83.3  | 87.2 |
| Favored                                | 83.3  | 87.2 |
| Allowed                                | 16.7  | 12.7 |
| Forbidden                              | 0     | 0 |

Protein Data Bank code: 5C70 5C71
idues (i.e. Tyr\textsuperscript{469}, Phe\textsuperscript{470}, Trp\textsuperscript{472}, and Tyr\textsuperscript{473}), thus leading to the low specificity of PGUS toward aglycone.

In addition, loop D (361–373), which is located between the third \(\beta\)-sheet and \(\alpha\)-helix in the TIM barrel domain, was especially interesting (Fig. 4). Although this loop was not well-defined because of weak electron density, the sequence alignment suggested that this loop is highly conserved in bacterial GUSs, and it has been reported to play an important role in substrate recognition (24). We used SWISS-MODEL to reconstruct loop D. As shown in Fig. 4b, loop D was located near the aglycone moiety of GAMG, thus suggesting that this loop interacts directly with the aglycone moiety.

**Engineering the substrate specificity of PGUS on the basis of structural analysis**

Similarly to most GHs, PGUS has poor substrate specificity and shows activity toward both GL and GAMG with the main product being GA (Fig. 1). However, GAMG shows much higher sweetness than GA and has wide applications in the food industry as a functional intense sweetener (30). Therefore, on the basis of structural analysis, we aimed to limit the substrate specificity of PGUS strictly to GL by mutating critical residues in the binding pocket. The two glycans of GL are both glucuronic acids. After the terminal glycan is hydrolyzed, PGUS continues to catalyze the hydrolysis of the second glycan, so it is difficult for PGUS to discriminate the second glycan to achieve a controllable hydrolysis process. To the best of our knowledge, no such work has previously been reported.

On the basis of the above analysis of the relationship between the structure and function of PGUS, Arg\textsuperscript{563} was determined to be particularly interesting because its mutation had different effects on the activity toward GL and GAMG. Therefore, saturated mutagenesis was further performed on Arg\textsuperscript{563} (31–33). As shown in Fig. 5a, the mutation of arginine to other residues with smaller side chains at the 563 site appeared to improve the GAMG yield; indeed, the GAMG yield of mutants R563Q, R563K, and R563E reached 42, 58, and 77%, respectively. All other R563\(X\) mutants showed drastic changes and exhibited extremely low activity compared with that of wild-type PGUS. Mutant R563K was purified, and its kinetics was studied as shown in Table 2. The catalytic efficiency, \(k_{cat}/K_m\), toward GAMG decreased by 88.4%, whereas \(k_{cat}/K_m\) toward GL decreased by 12.3%, confirming site 563 had a dramatic effect on substrate specificity.

Previous research has shown that residues in the substrate channel may dramatically affect the catalytic properties of enzymes even if they are not directly involved in substrate recognition (34). Therefore, to further decrease the specificity toward GAMG, we subsequently performed saturated mutagenesis on the 21 residues of the binding pocket that did not establish direct interactions with GAMG (i.e. Trp\textsuperscript{550}, Ser\textsuperscript{562}, Ser\textsuperscript{362}, Ile\textsuperscript{363}, Gly\textsuperscript{364}, Ala\textsuperscript{365}, Gly\textsuperscript{366}, Ala\textsuperscript{367}, Tyr\textsuperscript{369}, Asn\textsuperscript{467}, Val\textsuperscript{447}, Gly\textsuperscript{448}, Tyr\textsuperscript{159}, Gln\textsuperscript{160}, Tyr\textsuperscript{164}, Gly\textsuperscript{358}, Gln\textsuperscript{175}, Tyr\textsuperscript{557}, Ala\textsuperscript{558}, and Gly\textsuperscript{560}). For each selected site, more than 200 mutants were screened to ensure high coverage, and 21 libraries containing more than 4,200 colonies were selected and characterized individually with respect to their catalytic activity and substrate specificity for GL hydrolysis. To prompt fast screening of the target mutant, an efficient screening system was set up for the high-throughput screening of PGUS mutants with the desired substrate specificity. We constructed a dual-function fusion

![Figure 2](https://example.com/image.png)

**Figure 2.** a, crystal structure of PGUS (asymmetric unit). b, structure of one monomer of PGUS (the sugar-binding domain is green, the TIM barrel domain is cyan, and the immunoglobulin-like \(\beta\)-sandwich domain is magenta). The two catalytic sites, Glu\textsuperscript{414} and Glu\textsuperscript{505}, are shown as red sticks. c, overview of the active site of PGUS. d, ribbon diagram of the six loops that form the substrate entry to the active site of PGUS: loop A (red), loop B (blue), loop C (magenta), loop D (orange), loop E (yellow), and loop F (cyan).
Figure 3. Substrate recognition and binding of PGUS elucidated from the crystal structure of PGUS complexed with GAMG. a, the active site pocket of PGUS in complex with GAMG. The electron density for GAMG is also shown. b, analysis of the interactions between GAMG and PGUS. GAMG is shown as blue sticks, and the residues of PGUS that interact with the glycan are shown as green sticks. c, a schematic diagram of the detailed interaction network drawn on the basis of the structure in b.

Figure 4. a, ribbon diagram of loop D (361–373) and GAMG. Loop D (361–373) is not well-defined due to its weak electron density. b, loop D (361–373) was reconstructed using SWISS-MODEL, and its location suggests that it is involved in aglycone recognition. The structure of PGUS is shown as green schematics, and GAMG is shown as blue sticks.
plasmid (pHCE-gus-SRRz) consisting of the GUS gene under a high-level constitutive expression (HCE) promoter and the autolytic gene cassette SRRz from bacteriophage λ under the cI857/pR promoter (see “Experimental procedures”). On the basis of the altered expression profiles, the substrate specificity of GUS mutants was screened without any prior induction or preconcentration for cell lysis.

Improved GAMG yield was observed for mutants in libraries A365X and V447X. The mutations A365T, A365Q, and V447Q were found to have the highest GAMG yield among the tested mutants with values reaching 61, 65, and 81%, respectively. Two representative mutants, A365T and V447Q, were further purified, and their kinetics were characterized. As shown in Table 2, their $k_{cat}/K_m$ toward GAMG decreased by 53.4 and 63.1%, respectively, indicating that these two sites also had an effect on substrate specificity. Therefore, Ala365 and Val447 together with Arg563 were identified as hot spots and subjected to combined site-saturation mutagenesis.

For each of these combined mutation sites, ~200 mutants were screened and characterized as described above. 

Figure 5. a, the effects of the saturated mutagenesis of Arg563 on the substrate specificity of PGUS. b, the effects of the combination of mutations of Ala365, Val447, and Arg563 on the substrate specificity of PGUS. Error bars represent S.D.
mutants with significantly improved substrate specificity were obtained. Three two-site mutants (V447Q/R563K, A365T/R563E, and A365H/R563E) exhibited much higher GAMG yield (95, 95, and 96%, respectively) than that of wild-type PGUS (<10%) (Fig. 5b), thus resulting in nearly complete alteration of the substrate selectivity of the glucuronidase from GA to GAMG formation.

To determine how substitutions affected the kinetic properties of PGUS, the kinetic parameters of the initial reaction rates of these three positive mutants and wild-type PGUS were investigated. The three mutants (V447Q/R563K, A365T/R563E, and A365H/R563E) exhibited such low catalytic activity toward GAMG that the $K_m$ values for GAMG could not be determined, thus indicating low affinity between the mutants and GAMG. In addition, the $K_m$ values toward GL of all three mutants were increased by 1.1–2.4-fold, thus indicating that their affinity to GL was lower than that of wild-type PGUS. Nonetheless, the value of $k_{cat}$ toward GL was increased by 1.2-fold for mutant V447Q/R563K (Table 2). The $k_{cat}/K_m$ of mutant V447Q/R563K (4.32 mm$^{-1}$ s$^{-1}$) was slightly higher than that of wild-type PGUS (4.14 mm$^{-1}$ s$^{-1}$). Thus, the substrate specificity of GL and activity were both significantly improved. To confirm the improved substrate specificity, a time-course kinetic study of A365H/R563E and V447Q/R563K was conducted with GL hydrolysis. The results indicated that A365H/R563E had the best substrate specificity with very little GA produced after 60 min, thus suggesting that the obtained mutant has great potential in the industrial production of GAMG (Fig. 6).

### Discussion

The substrate recognition mechanism (the aglycone moiety in particular) of GUS toward glycosides is a fundamental question that remains unclear at present due to the lack of a crystal structure of the enzyme complex. In this study, the crystal structures of PGUS and its mutant PGUS E414D/E505D in complex with the natural substrate GAMG were for the first time resolved, allowing a profound understanding of substrate-binding pocket and the roles of critical residues involved in the catalysis.

The PGUS sequence shares 54.5% identity with EGUS and 47.5% with HGUS (Fig. S3). Multiple sequence alignment of PGUS with its homologs indicated the presence of the conserved regions WSIANE (acid/base catalyst) and PIVMTEYGAD (nucleophile), which are somewhat similar to the corresponding sequences of EGUS (i.e. WSIANE and PIITQGVD) and HGUS (i.e. WSVANE and PIIQSEYGAE).

As the first crystal structure of GUS from fungus, PGUS shows high structural homology with EGUS and HGUS with calculated root mean square deviation values of 0.75 and 0.69 Å, respectively, thus reflecting the high sequence similarity (Fig. S4).

We prepared a PGUS mutant, E414D/E505D, and for the first time obtained the complex crystal structure of PGUS mutant with its natural substrate, GAMG. In the PGUS-GAMG complex, both enzyme and GAMG were clearly visualized, which allowed a precise substrate recognition mechanism analysis. The glycan is located in the bottom of the active site pocket and recognized by a series of polar interactions formed with highly reserved neighboring residues (Asp$^{114}$, Asp$^{505}$, Asp$^{162}$, His$^{332}$, Arg$^{663}$, Asn$^{567}$, Lys$^{569}$, Tyr$^{469}$, and Tyr$^{473}$). Extensive mutation was performed to verify the functional roles of these residues. The results indicate that even tiny changes in these residues would cause significant PGUS activity loss, confirming that these residues are highly critical in maintaining the correct catalytic domain for glycan recognition. This is similar to the result of EGUS complexed with its inhibitor (the only reported GUS-glycan complex structure) where the residues involved in the polar interaction network are quite similar to the residues in our case (24). The extensive polar interaction network between enzyme and glycan may explain the strict specificity of PGUS toward the glycan moiety (see Fig. 3), which is consistent with most GHs from different families where highly rigorous glycan recognition has been reported (25, 29).

We found that the recognition mechanism of GUS toward aglycone was quite different from glycan. The aglycone was mainly recognized by several aromatic residues, including Tyr$^{469}$, Phe$^{470}$, Trp$^{472}$, Tyr$^{473}$, which is responsible for the low substrate specificity of PGUS toward aglycone. Similar interactions between aglycone and aromatic residues have been found in β-glucosidases from maize and sorghum, and four aromatic residues (Phe$^{198}$, Phe$^{205}$, Phe$^{466}$, and Phe$^{578}$) are located on two sides of the substrate, forming a full pocket that completely traps the aglycone moiety (35–37).

Currently, most GHs suffer from low substrate specificity toward glycosides, thus leading to uncontrollable processes with many undesired by-products, which has significantly impeded their application in modifying glycosides. Based on in-depth structure analysis and the substrate recognition mechanism, we identified two critical residues, Arg$^{563}$ and

---

**Table 2**

Kinetic parameters of PGUS and mutants for GL and GAMG

|        | $K_m$ (mM) | $k_{cat}$ (s$^{-1}$) | $k_{cat}/K_m$ (ms$^{-1}$ s$^{-1}$) |
|--------|-----------|---------------------|---------------------------------|
|        | GL        | GAMG                | GL                              | GAMG                          |
|        |           |                     |                                |                               |
| PGUS   | 0.78 ± 0.08 | 0.92 ± 0.11 | 3.23 ± 0.17 | 7.81 ± 0.33 | 4.14 | 8.49 |
| R563K  | 0.97 ± 0.12 | 0.91 ± 0.22 | 3.52 ± 0.06 | 0.89 ± 0.13 | 3.63 | 0.98 |
| A365T  | 1.23 ± 0.24 | 1.10 ± 0.19 | 3.28 ± 0.26 | 4.36 ± 0.41 | 2.67 | 3.96 |
| V447Q  | 1.21 ± 0.08 | 1.06 ± 0.13 | 6.03 ± 0.34 | 3.32 ± 0.15 | 4.98 | 3.13 |
| A365T/R563E | 2.20 ± 0.37 | ND | 2.72 ± 0.25 | ND | 1.23 | ND |
| A365H/R563E | 2.68 ± 0.32 | ND | 3.06 ± 0.21 | ND | 1.14 | ND |
| V447Q/R563K | 1.62 ± 0.31 | ND | 7.01 ± 0.35 | ND | 4.32 | ND |

*ND, not determined.*
Ala\textsuperscript{365}, that played critical roles in determining the substrate specificity of PGUS as the double mutant A365H/R563E showed a more rigorous substrate specificity, and its affinity to GAMG was significantly decreased as the corresponding $K_m$ could not be determined. This is in agreement with the involvement of Arg\textsuperscript{563} and Ala\textsuperscript{365} in the binding sites for the glycan group and aglycone moiety, respectively. Nevertheless, Arg\textsuperscript{563} showed less interaction with GL than GAMG. Indeed, Arg\textsuperscript{563} is the only residue in PGUS that establishes two polar interactions with the glycan moiety of GAMG, and thus it plays a very important role in GAMG hydrolysis. We next generated the structures of mutants R563Q, R563K, and R564E via modeling (Fig. 7a). When Arg\textsuperscript{563} was substituted with glutamine, lysine, and glutamic acid, which have shorter side chains, the glycan moiety of GAMG could no longer be recognized by residues at site 563 due to the increased distance. Therefore, the glycosidic bond could not form a proper pose to be hydrolyzed, thus causing low affinity and activity toward GAMG and increasing the substrate GL specificity. In addition, we speculated that the interaction between Arg\textsuperscript{563} and the first glycan moiety of GL might have been weaker and hence that mutation would not significantly change the activity of PGUS toward GL.

For substitutions at position Ala\textsuperscript{365}, according to the sequence and structural analysis (Fig. 4) Ala\textsuperscript{365} is located near the aglycone, thus allowing the formation of a direct interaction between them. Here, we assumed that the mutation of Ala\textsuperscript{365} to histidine or threonine might have $\pi-\pi$ stacking or $\text{OH}-\pi$ interactions with the pentacyclic triterpene aglycone of GAMG and thus might behave as a “clamp” to hold GAMG and stop it from sliding into the active site, thereby resulting in low activity toward GAMG and increasing the GL substrate specificity (Fig. 7b). Therefore, on the basis of the proposed recognition mechanism involving Arg\textsuperscript{563} and Ala\textsuperscript{365}, we concluded that modulating the interactions between glycan and PGUS (e.g., hydrogen bonding and electrostatic interactions) and controlling the aglycone backbone movement inside the binding pocket allow for manipulation of the substrate specificity of PGUS toward GL.

Figure 6. Time course of the hydrolysis of GL by mutants A365H/R563E and V447Q/R563E.

Figure 7. a, the superimposition of structures of PGUS and mutants R563Q, R563K, and R563E docked with GAMG. Only critical residues are displayed for clarity. The mutagenesis was performed using PyMOL, and the structure was optimized by VMD (Visual Molecular Dynamics) software with energy minimization. b, the active site pocket of PGUS in complexed with GAMG (from crystal structure) and GL (from docking). The docking was performed with Autodock software.
In summary, we present the first crystal structures of PGUS and its mutant PGUS E414D/E505D in complex with the natural substrate GAMG. Structural analysis combined with mutational data indicated that the glycan moiety of GAMG is recognized by polar interactions, which are responsible for the strict specificity of PGUS toward the glycan moiety. In contrast, the aglycone group of PGUS forms aromatic stacking and π interactions with the neighboring aromatic residues (i.e. Tyr\textsuperscript{469}, Phe\textsuperscript{470}, Trp\textsuperscript{472}, and Tyr\textsuperscript{473}), thus explaining the low specificity of PGUS toward aglycone. On the basis of mutagenesis experiments, we successfully tuned the PGUS substrate specificity toward GL to improve its biotechnological applications in triterpenoid saponin modification. Three residues (Arg\textsuperscript{563}, Ala\textsuperscript{365}, and Val\textsuperscript{477}) in the binding pocket were found to be critical for the substrate specificity toward GL. Among them, mutant A365H/R563E showed 95% substrate specificity (wild-type PGUS showed only 10%), whereas activity was similar to that of the wild-type. This study provides new insight into structural analysis-based engineering of the substrate specificity of GUSs for the modification of triterpenoid saponins.

**Experimental procedures**

**Materials**

Restriction endonucleases and T4 DNA ligase were purchased from Fermentas (Beijing, China). FastPfu DNA polymerase was purchased from Transgen Biotech (Beijing, China). DNA ladder and molecular mass markers for sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were purchased from Beijing Dingguo Changsheng Biotechnology (Beijing, China). The plasmid isolation kit was purchased from BioTeke (Beijing, China). The DNA gel extraction kit was purchased from BioTeke (Beijing, China). Oligonucleotides were synthesized by GENEWIZ (Beijing, China). Sequence analysis was performed by Sangon Biotech or GENEWIZ. GL, GAMG, and GA were purchased from Sigma-Aldrich. All other chemicals used were commercially available and of analytical grade.

**Strains, plasmids, and primers**

The *E. coli* strain TOP10 was used for the propagation and amplification of plasmids, and *E. coli* strains DH5α (ΔuidA) and BL21 (DE3) were used for enzyme assays and characterization of the substrate specificity in GL transformation. Plasmids pET-28a-GUS and pUC19 were constructed and stored in our laboratory at the Beijing Institute of Technology. The pEAS-1a vector was obtained as a gift from Professor Zhanglin Lin (38). General recombinant DNA manipulations were performed using standard procedures. All polymerase chain reactions (PCRs) were conducted under the manufacturer’s recommended conditions. The primers used in the PCRs are listed in Tables S2, S3, and S4.

**Protein production and purification**

The pET28a/pgus plasmid bearing PGUS cDNA with an N-terminal His tag was transformed into *E. coli* BL21 (DE3) cells. The cells were cultured in Luria-Bertani (LB) medium containing 50 μg ml\(^{-1}\) kanamycin at 37 °C. When \(A_{600}\) of the culture reached 0.6–0.8, protein production was induced with 1 mM 1-thio-β-D-galactopyranoside for 10 h at 16 °C. The cells were collected by centrifugation (12,000 rpm) at 4 °C for 10 min, suspended in extraction buffer (50 mM Tris-HCl, pH 7.3, 150 mM NaCl), and sonicated. The crude enzyme was prepared by centrifugation at 12,000 rpm for 20 min to remove precipitated proteins and cell debris.

The crude enzyme solution was first passed through a Ni\(^{2+}\)-affinity chromatography column (His Trap\textsuperscript{TM} FF 5 ml, GE Healthcare) with a flow rate of 1 ml min\(^{-1}\) and eluted with 50, 100, and 250 mM imidazole in 50 mM Tris-HCl buffer, pH 7.3. Then the fractions showing enzyme activity were pooled and further purified with a weak anion-exchange chromatography column (HiPrep EDAE FF 16/10, GE Healthcare) with a flow rate of 1 ml min\(^{-1}\). The bound protein was eluted with a gradient NaCl solution (0–500 mM) at the same flow rate. Finally, the fractions showing the highest activity were pooled, concentrated, and passed through a gel-filtration column (Superdex\textsuperscript{TM} 200 2.6/60, GE Healthcare), which was equilibrated and eluted with 30 mM Tris-HCl, pH 7.3, 100 mM NaCl at a flow rate of 0.8 ml min\(^{-1}\). The chromatographic processes were all performed with an AKTA Purifier 10 system (GE Healthcare). The purified enzyme was stored at −4 °C until further use. The protein concentration was determined using the method of Bradford (39) at 595 nm using bovine serum albumin as the standard.

**Crystallization, data collection, and structural determination**

The initial crystallization screening of PGUS was conducted with the sitting drop vapor diffusion method at 16 °C by mixing 0.1 μl of protein solution (7.5 and 10 mg ml\(^{-1}\)) and an equal volume of reservoir solution against 50 μl of reservoir solution. Crystals of PGUS were initially obtained by vapor diffusion using a reservoir solution of 1 M succinic acid, pH 7.0, within several days at 16 °C. The diffraction power was improved after extensive optimization of the crystallization conditions. PGUS and E414D/E505D mutant protein were crystallized under the same conditions, and the mutant crystals with good quality were then soaked with GAMG independently for 3 h before incubation in the reservoir solution containing saturated sucrose as a cryoprotectant and subsequent flash freezing in liquid nitrogen. Diffraction experiments were conducted at the Shanghai Synchrotron Radiation Facility (Shanghai, China). Diffraction data collection was performed using a MAR MX 225 charge-coupled device detector (Area Detector Systems Corp., Poway, CA). The structure of PGUS was solved by the molecular replacement method with the program Molrep from the CCP4i suite. The coordinate model was *E. coli* GUS (Protein Data Bank code 3K4D). Refmac and Phenix-refine were used for refinement, and Coot was used for model building and adjustment. The structure of the PGUS mutant E414D/E505D with GAMG was solved by molecular replacement using the PGUS structure as the search model. Refmac, Phenix-refine, and Coot were used for model building and refinement.

**Enzyme assays and kinetic studies**

PGUS activity was measured using GL monoammonium salt as the substrate. The reaction mixture, which consisted of 200 μl of crude enzyme solution and 800 μl of 1.25 mM GL monoammonium salt in 50 mM sodium acetate buffer, pH 5.0, was
Substrate specificity of β-glucuronidase

incubated at 40 °C. The reaction was halted by placing the reaction mixture in boiling water, and then the reaction mixture was centrifuged at 10,000 rpm for 10 min. The supernatant was used to determine the amount of GL by high-performance liquid chromatography (HPLC). The chromatographic conditions were as follows: octadecysilyl column (Shim-pack, VP-ODS, 4.6 × 250 mm, Shimadzu Corp., Kyoto, Japan); ultraviolet (UV) detector (detection wavelength, 254 nm); flow rate, 1.0 ml min⁻¹; mobile phase, water, pH 2.85, with 0.6% acetic acid and methanol at a ratio of 19:81; and injection volume, 10 μl. The retention times for GL, GAMG, and GA were 6.0, 12.5, and 22.5 min, respectively. One unit of PGUS enzyme activity was defined as the amount of enzyme capable of converting 1 nmol of GL/min under the investigated conditions.

The Michaelis-Menten kinetic parameters of PGUS and its mutants were determined by measuring the initial rates of GL or GAMG hydrolysis at concentrations of 0.2–6.1 and 0.3–7.7 mM, respectively. An enzyme concentration of 0.17 μM was used for all kinetics tests. The reaction was performed as in the activity test described above. The kinetic parameters $K_{\text{m}}$, $v_{\text{max}}$, and $k_{\text{cat}}$ were determined using nonlinear regression with Origin software. Partial data and fits of the enzyme kinetics are presented in Fig. S5.

Construction of the dual-function vector for one-tube screening

To construct a suitable vector for PGUS gene expression and GL transformation without the addition of an inducer and extra disruption of the cell wall, the essential regions, including the constitutive target gene expression cassette, autolytic cassette, and origin of replication with the resistance gene, were retained in the final vector (Fig. S6). First, the pMB1 origin of replication with the Amp resistance gene (2125 bp) was PCR-amplified by ori-up-Kpn1-Sph1 and ori-down-NdeI from the pUC19 vector with NdeI/Kpn1. Second, the autolytic cassette containing bacteriophage λ under the cl857/p$_r$ promoter (2579 bp) was amplified by cIts857/Srrz-Up-NdeI and Srrz-down-Sph1 from the pEAS-1a vector with NdeI/Sph1. Third, the HCE promoter was used to combine PGUS fragments with the rrnBT1T2 terminator to create a promoter-gene-terminator cassette (40). The required PCR products were obtained using primer sets Phce-F and Phce-R. To amplify the promoter of Phce, a DNA fragment that consisted of the Hix$_o$ tag and rrnBT1T2 terminator fused to the PGUS gene was obtained through the overlapping linker region. The HCE promoter-gene-terminator cassette ligated to the pMB1 origin of replication with the Amp resistance gene and the autolytic cassette was named pHCE-gus. For E. coli DH5α ΔuidA cells transformed with pHCE-gus-SRRz, cell lysis was induced by a temperature shift from 35 to 42 °C, and the extent of lysis was assayed every hour after the heat induction. As shown in Fig. S7, the cells were observed to be completely disrupted with a lysis efficiency of 98% at 40 °C.

Library creation and substrate specificity screening

The mutants were prepared by whole-plasmid PCR (pHCE-gus-SRRz) with the primer containing NNN (sense strand)/NNN (antisense strand) degeneracy at the target sites. The primers are listed in Tables S3 and S4 where N represents any of the following: A, T, G, or C. PCR was performed with FastPfu DNA polymerase using plasmid PGUS as a template and a temperature program consisting of 2 min at 98 °C, 25 cycles of 20 s at 95 °C, 20 s at 65 °C, and 4 min at 72 °C; and a final 5-min extension at 72 °C. The PCR products were digested with DpnI to remove the parent plasmid and then directly transformed into E. coli DH5α (ΔuidA) cells and selected for ampicillin resistance.

Briefly, single colonies of the transformants were randomly picked with toothpicks and inoculated into 200 μl of CPGY medium (40) containing 1 g liter⁻¹ GL supplemented with ampicillin in 96-well microtiter plates and simultaneously onto LB agar plates with 100 mg ml⁻¹ ampicillin (LB/Amp plates) as replications. The cultures were incubated at 30 °C for 24 h and then at 40 °C for 24 h and subsequently subjected to a whole-cell biotransformation assay with GL (Fig. S8). The GAMG formation with PGUS and GAMG, and GA were determined by HPLC as described above. The GL conversion ($C_{\text{GL}}$) and GAMG yield ($Y_{\text{GAMG}}$) were calculated as follows.

$$C_{\text{GL}}(\%) = \frac{S_0 - S_t}{S_0} \times 100$$  \hspace{1cm} (Eq. 1)

$$Y_{\text{GAMG}}(\%) = \frac{\text{GAMG}_{\text{mol}}}{\text{GL}_{\text{mol}} + \text{GAMG}_{\text{mol}} + \text{GA}_{\text{mol}}} \times 100$$  \hspace{1cm} (Eq. 2)

where $S_0$ and $S_t$ are the concentrations of the substrate at time 0 and time $t$, respectively, and $\text{GAMG}_{\text{mol}}$ and $\text{GA}_{\text{mol}}$ are the molar concentrations of GAMG and GA, respectively.

Author contributions—B. L., H. S., H. T., and C. L. designed the experiments. B. L., H. S., and X. F. performed the gene expression, purification, and enzyme assays. H. S. and T. J. performed the protein crystallization and solved the structure. B. L., H. S., and X. F. conducted structural analysis and mutation experiments. B. L., X. F., and C. L. wrote the paper, and all authors reviewed the manuscript.

References

1. Iglesias-Fernandez, J., Raich, L., Ardevol, A., and Rovira, C. (2015) The complete conformational free energy landscape of β-xylene reveals a two-fold catalytic itinerary for β-xylanases. *Chem. Sci.* 6, 1167–1177 [CrossRef]

2. Verdoucq, L., Morinière, J., Bevan, D. R., Esen, A., Vasella, A., Henrissat, B., and Czjzek, M. (2004) Structural determinants of substrate specificity in Family 1 β-glucosidases. Novel insights from the crystal structure of sorghum dhurrinase-1, a plant β-glucosidase with strict specificity, in complex with its natural substrate. *J. Biol. Chem.* 279, 31796–31803 [CrossRef Medline]

3. Vucodac, D. J., and Davies, G. J. (2008) Mechanistic insights into glycosidase chemistry. *Curr. Opin. Chem. Biol.* 12, 539–555 [CrossRef Medline]

4. Speciale, G., Thompson, A. I., Davies, G. J., and Williams, S. J. (2014) Dissecting conformational contributions to glycosidase catalysis and inhibition. *Curr. Opin. Struct. Biol.* 28, 1–13 [CrossRef Medline]

5. Brito-Arias, M. (2007) Synthesis and Characterization of Glycosides, pp. 1–67, Springer, Boston, MA

6. Gleadow, R. M., and Muller, B. L. (2014) Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annu. Rev. Plant Biol.* 65, 155–185 [CrossRef Medline]
25. Michikawa, M., Ichinose, H., Momma, M., Biely, P., Jongkees, S., Yoshida, A., Rovira, C., Hrmova, M., and Kaneko, S. (2012) Structural and biochemical characterization of glycocside hydrolase family 79 β-glucuronidase from Acidobacterium capsulatum. J. Biol. Chem. 287, 14069–14077 CrossRef Medline

26. Wallace, B. D., Roberts, A. B., Pollet, R. M., Ingle, J. D., Biernat, K. A., Pellock, S. J., Venkatesh, M. K., Guthrie, L., O’Neal, S. K., Robinson, S. I., Dollinger, M., Figueuera, E., McShane, S. R., Cohen, R. D., Jin, J., et al. (2015) Structure and inhibition of microbium β-glucuronidases essential to the alleviation of cancer drug toxicity. Chem. Biol. 22, 1238–1249 CrossRef Medline

27. Jacobson, R. H., Zhang, X. J., DuBose, R. F., and Matthews, B. W. (1994) Three-dimensional structure of β-galactosidase from Escherichia coli. Nature 369, 761–766 CrossRef Medline

28. Tankrathok, A., Iglesias-Fernández, J., Luang, S., Robinson, R. C., Kimura, A., Rovira, C., Hrmova, M., and Ketudat Cairns, J. R. (2013) Structural analysis and insights into the glycos specificity of the rice GH1 β-glucosidase. Acta Crystallogr. D Biol. Crystallogr. 69, 2124–2135 CrossRef Medline

29. Saino, H., Shimizu, T., Hiralake, J., Nakatsu, T., Kato, H., Sakata, K., and Mizutani, M. (2014) Crystal structures of β-primeverosidase in complex with disaccharide amide inhibitors. J. Biol. Chem. 289, 16826–16834 CrossRef Medline

30. Mizutani, K., Kuramoto, T., Tamura, Y., Ohtake, N., Doi, S., Nakaura, M., and Tanaka, A. (1994) Sweetness of glycyrrhetic acid 3-O-β-D-mannoside and the related glycosides. Biosci. Biotechnol. Biochem. 58, 554–555 CrossRef Medline

31. Sun, Z., Lonsdale, R., Ille, A., Li, G., Zhou, J., and Reetz, M. T. (2016) Catalytic asymmetric reduction of difficult-to-reduce ketones: triple-code saturation mutagenesis of an alcohol dehydrogenase. ACS Catal. 6, 1598–1605 CrossRef Medline

32. Sun, Z., Lonsdale, R., Wu, L., Li, G., Li, A., Wang, J., Zhou, J., and Reetz, M. T. (2016) Structure-guided triple-code saturation mutagenesis: efficient tuning of the stereoselectivity of an epoxide hydrolase. ACS Catal. 6, 1590–1597 CrossRef Medline

33. Reetz, M. T., and Carballeira, J. D. (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. Nat. Protoc. 2, 891–903 CrossRef Medline

34. Luan, Z.-J., Li, F.-L., Dou, S., Chen, Q., Kong, X.-D., Zhou, J., Yu, H.-L., and Xu, J.-H. (2015) Substrate channel evolution of an esterase for the synthesis of cis-lactatin. Catal. Sci. Technol. 5, 2622–2629 CrossRef Medline

35. Cicek, M., Blanchard, D., Bevan, D. R., and Esen, A. (2000) The aglycone specificity-determining sites are different in 2,4-dihydroxy-7-methoxy-1,4-benzoaxin-3-one (DIMBOA)-glucosidase (maize β-glucosidase) and dhurrinase (sorghum β-glucosidase). J. Biol. Chem. 275, 20002–20011 CrossRef Medline

36. Cizzek, M., Cicek, M., Zamboni, V., Bevan, D. R., Henriass, B., and Esen, A. (2000) The mechanism of substrate (aglycone) specificity in β-glucosidases is revealed by crystal structures of mutant maize β-glucosidase-DIMBOA- DIMBOAGlc, and -dhurrin complex. Proc. Natl. Acad. Sci. U.S.A. 97, 13555–13560 CrossRef Medline

37. Cizzek, M., Cicek, M., Zamboni, V., Burmeister, W. P., Bevan, D. R., Henriass, B., and Esen, A. (2001) Crystal structure of a monocyteledon (maize ZMGlui) β-glucosidase and a model of its complex with p-nitrophenyl β-D-thioglucoside. Biochem. J. 354, 37–46 Medline

38. Xu, L., Li, S., Ren, C., Cai, Z., and Lin, Z. (2008) Heat-inducible autolytic vector for high-throughput screening. Biotechniques 41, 319–323 CrossRef Medline

39. Bradford, M. M. (1976) Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. Anal. Biochem. 72, 248–254 CrossRef Medline

40. Poo, H., Song, J., Hong, S. P., Choi, Y. H., Yun, S. W., Kim, J. H., Lee, S. C., Lee, S. G., and Sung, M. H. (2002) Novel high-level constitutive expression system, pHCE vector, for a convenient and cost-effective soluble production of human tumor necrosis factor-α. Biotechnol. Lett. 24, 1185–1189 CrossRef