



Molecular characterization and structural basis of a promiscuous glycosyltransferase for β -(1,6) oligoglucoside chain glycosides biosynthesis

Zhennan Jiang^{1,†}, Nianhang Chen^{2,†}, Hao-Tian Wang¹, Yungang Tian¹, Xiaoyu Du³, Ruibo Wu^{2,*}, Luqi Huang^{1,3,*}, Zi-Long Wang^{1,3,*}  and Yuan Yuan^{3,*} 

¹State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, 38 Xueyuan Road, Beijing, 100191, China

²School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou, China

³State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, Experimental Research Center, China Academy of Chinese Medical Sciences, Beijing, China

Received 27 December 2024;

revised 15 February 2025;

accepted 3 March 2025.

*Correspondence (Tel +86(010)39943061; fax +86(010)39943061; email wurb3@mail.sysu.edu.cn (R.B.W.); Tel +86(10)64087649; fax +86(10)64087649; email huangluqi01@126.com (L.Q.H.); Tel +86(10)82802024; fax +86(010)82802024; email wangzilong@hsc.pku.edu.cn (Z.L.W.); Tel +86(10)64087649; fax +86(10)64087649; email y_yuan0732@163.com (Y.Y.)).

[†]These authors contributed equally to this work.

Keywords: *Platycodon grandiflorum*, β -(1,6) glycosyltransferase, glycosides, enzymatic catalysis, crystal structure, catalytic mechanism.

Summary

Sugar building blocks are crucial for the chemical diversity and biological activity of secondary metabolites. UDP-dependent glycosyltransferases (UGTs) play a pivotal role in the biosynthesis of glycosides in plants by catalysing the attachment of sugar moieties to various bioactive natural products. However, the biosynthesis of oligosaccharide-chain glycosides is often limited by the narrow substrate specificity of UGTs. In this study, we identify a regio-specific β -(1,6) glycosyltransferase, UGT94BY1, from *Platycodon grandiflorum*. UGT94BY1 exhibits broad substrate promiscuity and can transfer up to three sugar moieties to the C6-OH position of the glucosyl group in various triterpenoids and phenolic glycosides, thereby forming β -(1,6) oligoglucoside chains. To elucidate the mechanism underlying its substrate selectivity, we determined the crystal structure of the UGT94BY1 complex with UDP at a resolution of 2.0 Å. Molecular simulations revealed that a critical structural motif, comprising residues N84-M91, S141-L155 and R179-E186, plays a key role in recognizing sugar acceptors and facilitating chain elongation. Our study unveils a powerful glycosyltransferase for β -(1,6) oligoglucoside chain biosynthesis and highlights key regions involved in substrate recognition and sugar chain extension, providing valuable insights for designing UGTs with customized substrate specificities for biotechnological applications.

Introduction

Oligosaccharide chains are essential for the structural complexity, diversity, chemical stability, water solubility and biological activity of secondary metabolites (Johnson *et al.*, 2021; Silva Magedans *et al.*, 2021). These chains are typically attached to basic skeletons like terpenoids, steroid alkaloids, flavonoids and phenylethanoids to generate the corresponding glycosides in nature (Lucier *et al.*, 2024; Mrudulakumari and Lee, 2020; Seki *et al.*, 2015). The incorporation of oligosaccharide chains to these glycosides significantly enhances their biological activities and leads to important potential applications in agriculture, medicine, food and cosmetics (Gharabli *et al.*, 2023; Li *et al.*, 2023; Zaynab *et al.*, 2021). For instance, glycyrrhizin, which has an oligosaccharide chain at C-3 of the triterpene scaffold, has been developed as a treatment for viral hepatitis (Li *et al.*, 2014). QS21, a potential vaccine adjuvant with a complex oligosaccharide chain at C-28 of the triterpene scaffold, is critical for its immunostimulant activity (Martin *et al.*, 2024). Rutin and diosmin, both used in China as cardio-cerebrovascular protective drugs, contain oligosaccharide chains at the C-3 and C-7 positions of the flavonoid scaffold, respectively (Deng *et al.*, 2024; Zou *et al.*, 2023). Additionally, the type of glycosidic bonds in oligosaccharide chains plays a significant role in altering or enhancing

biological activity. For example, ginsenosides Rb1 and Rd are major active components in ginseng (Li *et al.*, 2022). While both compounds overlap in biological activity, Rb1 demonstrates greater potential as an anti-inflammatory agent, and Rd exhibits better neuroprotective effects. Rb1 differs from Rd by having an additional β -(1,6)-linked glucose (Chen *et al.*, 2022; Kim *et al.*, 2017). Moreover, mogroside V, siamenside I and mogroside IV are important natural sweeteners that contain more than one β -(1,6)-linked sugar moiety. By contrast, the basic scaffold mogroside III is tasteless (Li *et al.*, 2022).

Despite the significant progress made in chemical glycosylation reactions over the past century (Kulkarni *et al.*, 2018; Panza *et al.*, 2018), the synthesis of complex glycosides, especially plant-derived glycosides with various sugar moieties, remains challenging. These reactions often require elaborate protecting group strategies and pose difficulties in chemo-, regio- and stereoselectivity. This is exemplified by the extremely long chemical synthesis steps required for the sugar chain of QS21 (Fernández-Tejada *et al.*, 2015). By contrast, nature efficiently catalyses glycosylation reactions using carbohydrate-active enzymes (CAZY, <http://www.cazy.org/>). In plants, the biosynthesis of glycosides is typically catalysed by UDP-dependent glycosyltransferases (UGTs) (Kurze *et al.*, 2022; Mrudulakumari and Lee, 2020). Generally, the formation of a sugar chains

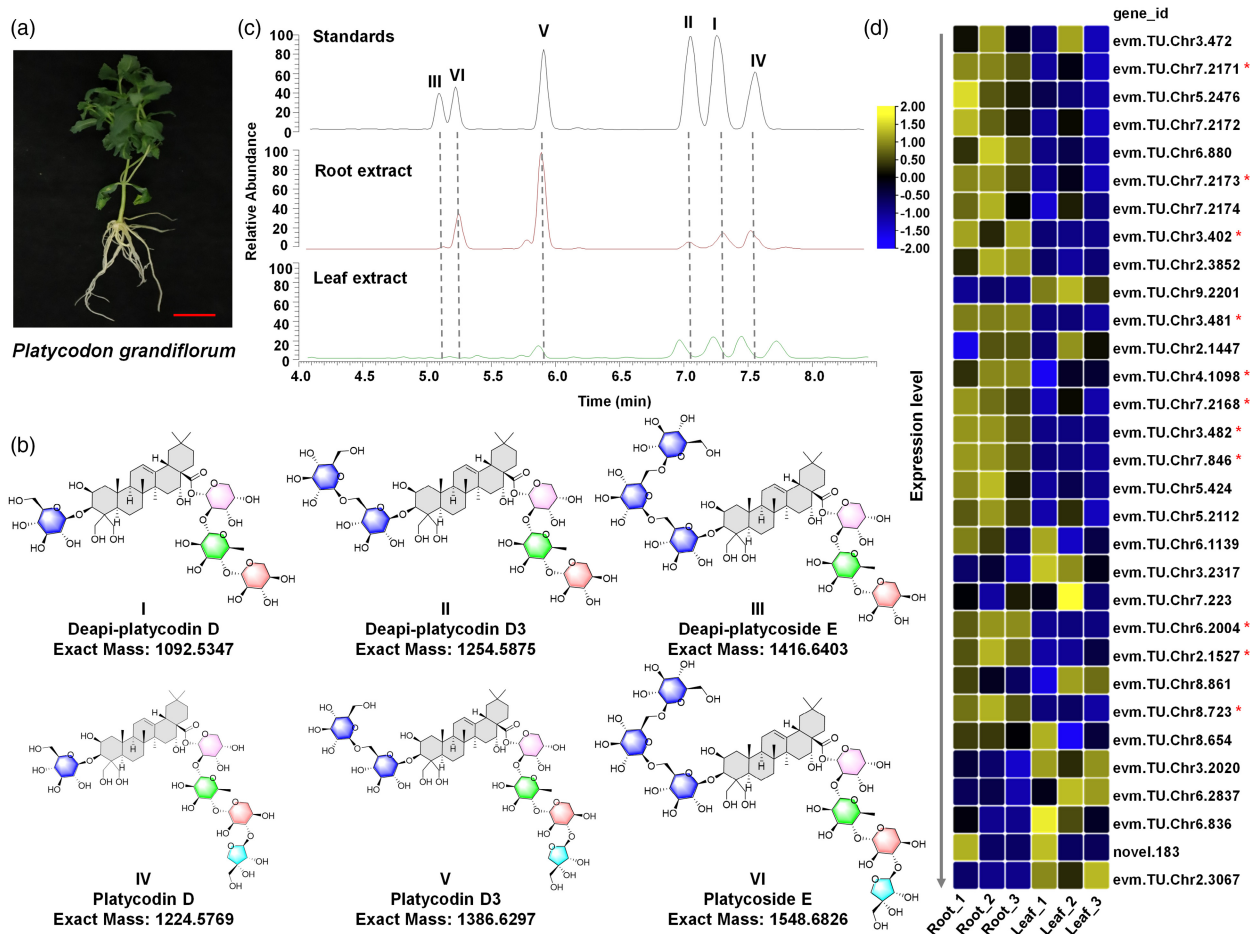


Figure 1 Screening of the candidate glycosyltransferases from *P. grandiflorum*. (a) One-year-old root of *P. grandiflorum*. The red scale bar indicates 4 cm. (b) Representative saponins with a β -(1,6) oligoglucoside chain in *P. grandiflorum*. (c) LC–MS analysis of root and leaf extracts from *P. grandiflorum*, depicting EICs (extracted ion chromatograms) of compounds I–VI. (d) The expression levels of candidate glycosyltransferase genes. Expression values represent z-score-transformed log (TPM + 1) values. * $P < 0.05$.

requires stepwise catalysis by multiple enzymes. For example, the biosynthesis of the C-28 sugar chain for QS21 involves four UGTs. However, few UGTs can catalyse multi-step glycosylation to form complex sugar chains (Liu *et al.*, 2024). A notable example is AmGT8, which catalyses a two-step glycosylation to generate a Glc(2,1)Glc oligoglucoside chain (Zhang *et al.*, 2022; Chen *et al.*, 2023). The β -(1,6) oligoglucoside chain is one of the most common active groups in natural products (Guo *et al.*, 2024; Li *et al.*, 2022; Zhang *et al.*, 2022). However, few β -(1,6) glycosyltransferases have been reported (Chen *et al.*, 2023; Tang *et al.*, 2023). Furthermore, most of these enzymes exhibit selectivity specific toward sugar acceptors and are unable to simultaneously recognize both phenolic and aliphatic sugar acceptors (Kurze *et al.*, 2022). Thus, there is an urgent need to discover UGTs with high regio-selectivity, catalytic efficiency and promiscuity for the forming oligoglucoside chains in secondary metabolites.

In the present study, we identified a β -(1,6) glycosyltransferase, UGT94BY1, which demonstrates high regio-specificity toward various glycosylated triterpenoids and phenols to produce β -(1,6) oligoglucoside chains. We also resolved the crystal structure of UGT94BY1 complexed with UDP. Through structural analysis and

molecular dynamics (MD) simulations, we pinpointed the key amino acids in UGT94BY1 responsible for substrate binding and sugar chain extension.

Results and discussion

Candidate glycosyltransferase screening and heterologous expression

Platycodon grandiflorum (Figure 1a; Figure S1), the sole species in the genus *Platycodon* (Campanulaceae), is a widely recognized medicinal plant with global recognition. Known as Jie-Geng in Chinese, Balloon flower in English and Kikyō in Japanese (Zhang *et al.*, 2022). *Platycodon grandiflorum* is rich in β -(1,6) oligoglucoside saponins, suggesting the existence of a highly efficient glycosyltransferase (Dong *et al.*, 2021; Ma *et al.*, 2023; Xie *et al.*, 2023). The primary triterpene saponins from this plant include deapi-platycodin D (I), deapi-platycodin D3 (II), deapi-platycoside E (III), platycodin D (IV), platycodin D3 (V) and platycoside E (VI) (Figure 1b). These compounds are decorated with three glucose moieties by specific β -(1,6) glucosidic bonds at the C-3 hydroxy group. Recently, PgGT1 was reported to participate in the glycosylation of platycodin D3

(V) and platycoside E (VI). However, it could not catalyse deapi-platycodin D (I) to produce deapi-platycodin D3 (II) and deapi-platycoside E (III) (Tang et al., 2023), exhibiting a narrow catalytic promiscuity.

To discover potential glycosyltransferases, we first analysed and compared the contents of six main triterpene saponins, platycodin (I–VI), all containing multiple sugar moieties in their structures, in the roots and leaves of *P. grandiflorum* (Figure 1b,c). Since these components were primarily distributed in the roots (Figure 1c) rather than in the leaves, we sequenced the transcriptomes of both roots and leaves (transcriptome_data; GSA number: CRA019288). The raw data were assembled using its genome data as a reference (Jia et al., 2022; Meng et al., 2024). A total of 121 glycosyltransferases were identified from the transcriptome, of which 31 genes were highly expressed in the roots (FPKM >10) (Figure 1d). We subsequently cloned and expressed eight candidate genes with significant differences ($p < 0.05$), excluding three incomplete sequences, in *E. coli* (Table S1).

Functional characterization and catalytic features of UGT94BY1

To investigate the function of the candidate genes, deapi-platycodin D (1) and UDP-Glc were used as the sugar acceptor and sugar donor, respectively. The reaction mixtures were prepared with 10 µg of purified protein, 0.1 mM deapi-platycodin D (1) and 0.5 mM UDP-Glc in 100 µL of 50 mM NaH₂PO₄-Na₂HPO₄ buffer (pH 8.0, 37°C). The reactions were analysed using liquid chromatography coupled with a Q Exactive hybrid quadrupole-Orbitrap mass spectrometer (LC–MS). Among the candidates, seven enzymes did not exhibit any glycosylation activity (Figure S2). However, in the presence of evm.TU.Chr7.2171 (Figure S3), most of the substrate was transformed, and LC–MS analysis revealed one major product 1b. This product exhibited an [M–H][–] ion at m/z 1415.64, which is 324 Da larger than that of compound 1. The [M–H][–] ion could lose 410 Da (representing arabinosyl, rhamnosyl and xylosyl residues at C-28 COOH) and 734 Da (corresponding to the Ara-Rha-Xyl sugar chain of C-28 COOH, with an additional two glycosyl residues), indicating that 1b is a di-*O*-glycosylated product of compound 1 (Figure 2a; Figure S4). Based on these observations, it was proposed that evm.TU.Chr7.2171 sequentially catalyses the multi-step glycosylation of compound 1. Furthermore, the exact mass of the mono-*O*-glycosylated product (1a) of 1 (Figure S4) was successfully extracted from the MS data (m/z = 1253.59). The structures of 1a and 1b were subsequently confirmed as the products of one-step and two-step glycosylation by comparison with authentic standards. Time-course experiments revealed that the first glycosylation step occurred rapidly, with the reaction proceeding so quickly that the mono-*O*-glycosylated product (1a) could not be captured even within the first 5 min (Figure 2b). In contrast, the di-*O*-glycosylated product (1b) consistently emerged as the main product throughout the entire time range from 5 to 240 min. Consequently, evm.TU.Chr7.2171 was identified as a di-*O*-triterpenoid glycosyltransferase responsible for the formation of deapi-platycodin D3 (II, 1a) and deapi-platycoside E (III, 1b).

evm.TU.Chr7.2171 was officially designated as UGT94BY1 by the UGT nomenclature committee. The cDNA sequence of UGT94BY1 contains an open reading frame (ORF) of 1392 bp,

encoding 463 amino acids. UGT94BY1 exhibited optimal activity at pH 8.0 (50 mM NaH₂PO₄-Na₂HPO₄ or 50 mM Tris–HCl) and 37°C, and its activity was independent of divalent metal ions (Figure S5). To explore its sugar donor promiscuity, seven sugar donors were tested, including UDP-Glc, UDP-Xyl, UDP-Ara, UDP-GlcNAc, UDP-Gal, UDP-GlcA and UDP-Rha, using 3-*O*-β-glucosyl platycodigenin (2) as sugar acceptor. UGT94BY1 was able to accept UDP-Glc, UDP-Xyl and UDP-Ara as sugar donors, with a significant preference for UDP-Glc (Figure S6).

Besides forming a β-(1,6) sugar chain at the C-3 position of triterpenoid, UGT94BY1 displayed unexpected catalytic functions, accepting other types of substrates. For instance, UGT94BY1 was capable of recognizing the glucose moiety at the C-28 position of the triterpenoid Chikusetsusaponin IV (16) and generate higher-polarity products (16b and 16c) (Figure 2a; Figure S7). In a time-course study, 16b and 16c were detected at 5 min and 40 min, respectively, suggesting that 16b is a precursor of 16c (Figure 2b). The corresponding mass spectra revealed the [M – H][–] ions at m/z 1249.60 and 1411.64 (Figure 2c; Figure S7), indicating that 16b and 16c are the di-*O* and tri-*O*-glycosylated products of 16, respectively. The mono-*O*-glycosylated product was also detected by extracted ion chromatography (EIC m/z = 1087.54) (Figure S7). To verify the function of UGT94BY1, the tri-*O*-glycosylated product 16c was isolated via preparative-scale enzymatic reactions. Compared with 16, 16c contains three β-(1,6) linked glycoside moieties, and its structure was fully characterized using the nuclear magnetic resonance (NMR) and high-resolution electrospray ionization mass spectrometry (HR-ESI/MS) analyses (Figures S8–S11). These results indicated that UGT94BY1 recognizes and catalyses the glucose moiety at the C-28 position of triterpenoids, facilitating multi-step glycosylation reactions to generate a oligoglucoside chain with β-(1,6) glycosidic bonds.

Flavonoid glycosides, another major group containing β-(1,6) glycosidic bonds, were also tested to determine whether UGT94BY1 is responsible for their oligoglucoside chain biosynthesis. We then tested the catalytic activity of this enzyme using a commercialized flavonoid glycoside astragalin (25). UGT94BY1 was also unexpectedly capable of catalysing the glycosylation of (25) to produce two products, 25a and 25b, with corresponding MS [M–H][–] ions at m/z 609.15 and 771.20, respectively. This suggests that 25a and 25b are mono-*O* and di-*O*-glycosylated products of 25, respectively (Figure 2a–c; Figure S12). The structure of 25a was subsequently identified as kaempferol 3-*O*-gentiobioside by comparing it with an authentic standard. These findings demonstrate that UGT94BY1 possesses both substrate promiscuity and specificity in β-(1,6) sugar chain synthesis.

Substrate promiscuity and synthetic applicability of UGT94BY1

UGT94BY1 exhibits catalytic promiscuity for deapi-platycodin D (1), Chikusetsusaponin IV (16) and astragalin (25). Based on the structural characteristics of these substrates, we hypothesize that UGT94BY1 can recognize 3-*O*-glucosylated triterpenoids, 28-*O*-glucosylated triterpenoids and glycosylated flavonoids. To explore the substrate promiscuity and recognition characteristics of UGT94BY1, an acceptor library of 39 structurally diverse substrates was tested with UDP-Glc as the sugar donor (Figure 3). The library included 20 triterpenoids (1–20), seven flavonoid glycosides (21–27), 10 triterpenoid and flavonoid aglycones (28–

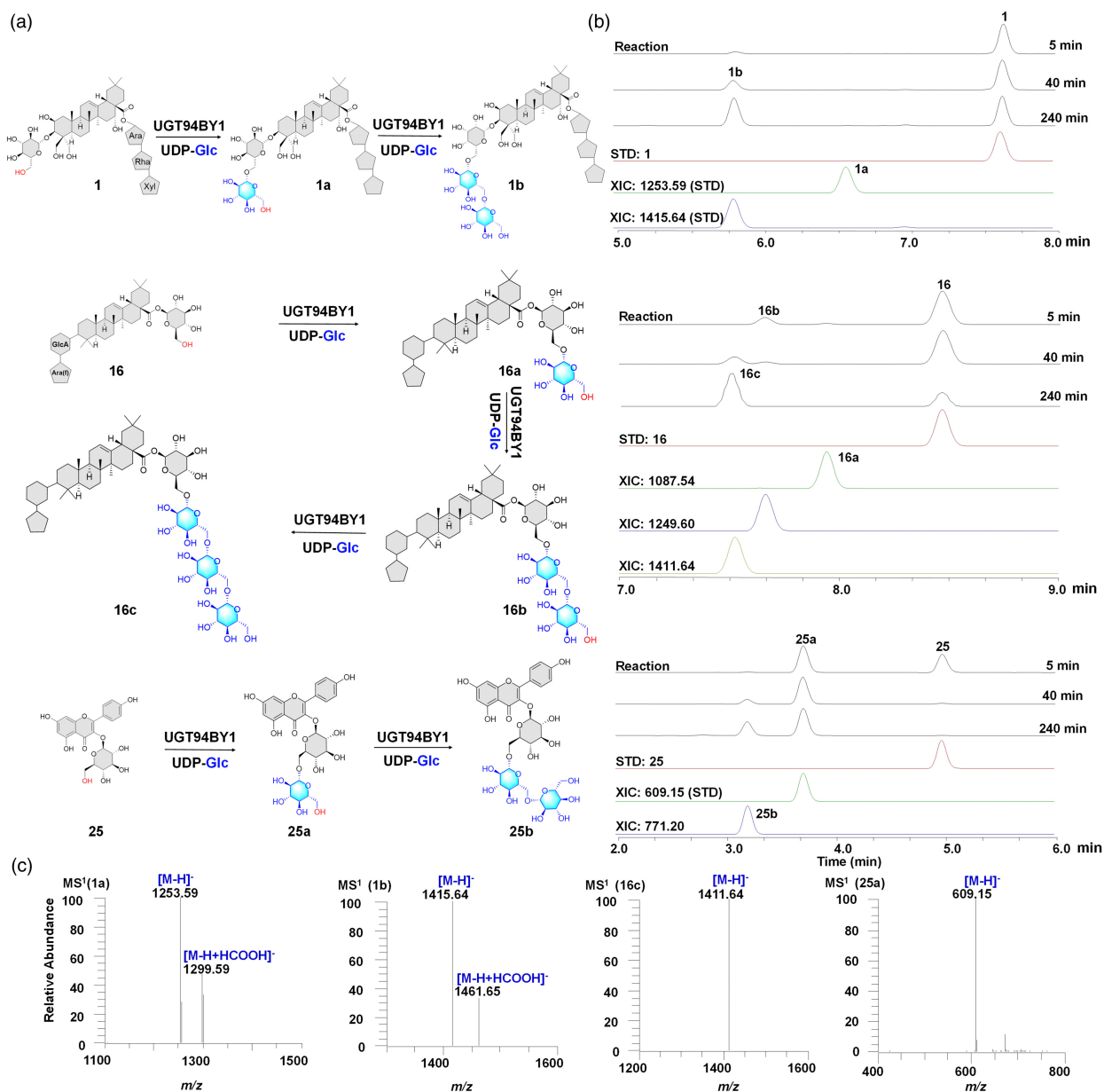


Figure 2 UGT94BY1 catalyzes multi-step glycosylation. (a) Stepwise glycosylation of compounds (**1**), (**16**) and (**25**). (b) Time-course analysis of the enzymatic reaction. The structures of **1a**, **1b**, **16c** and **25a** were fully identified by comparing with reference standards. (c) LC-MS data for **1a**, **1b**, **16c** and **25a**.

37), anthraquinone glycoside (**38**) and phenylethanoid (**39**). The reaction mixtures were analysed using LC-MS (Figure 3; Figures S4, S7, S12–S33). UGT94BY1 exhibited significant promiscuity toward triterpenoid saponins, including both tetracyclic and pentacyclic structures, as well as various types of phenolic compounds.

The C-3 position of 15 triterpenoid compounds (**1–15**) is glycosylated with one or two sugar moieties. Compounds **1–13**, which include both pentacyclic and tetracyclic triterpene scaffolds, can be efficiently catalysed by UGT94BY1, leading to the extension of sugar chain with one to three glucose moieties. A common structural feature among the substrates recognized by UGT94BY1 is the presence of a free 6-OH group in sugar moiety at the C-3 position of the triterpenoid. For instance, pentacyclic

triterpenes (**6–9**) feature two successive sugar residues at the 3-OH position, with the sugar chains consisting of Glc(6,1)Glc, Xyl(4,1)Glc, GlcA(2,1)Glc or Ara(2,1)Glc (Wang *et al.*, 2022). By contrast, compounds **14** and **15** are not substrates of UGT94BY1 as their sugar chains lack free 6-OH group in the glycosyl moieties (Figure 3b). Notably, besides the C-3 group, six of the 15 compounds contain additional sugar chains. The number of C28 carboxyl-terminal sugar chains of compounds **1**, **2** and **3** are three, zero and four sugar moieties, respectively (Figure 3b). Under the same conditions, compound **2** was able to continuously accept three glucosyl moieties via β -(1,6) glycosidic bonds (Figure S13). By contrast, up to two glycosyls could be attached to compounds **1** and **3** (Figures S4, S14). Similar results also could be seen for ginsenosides as sugar acceptors (**10–13**, Figures S21–

Figure 3 Substrate promiscuity of UGT94BY1. (a) The conversion rate of glycosylated products catalysed by UGT94BY1. Green, purple and pink represent the addition of one, two and three glucose moieties, respectively, generated by UGT94BY1. The structures of **3a** and **3b** were fully identified by comparing with reference standards. (b) Structure of substrates **1–27**. The red highlighted -OH indicates the catalytic active site. The yellow, green and purple boxes represent different types of substrates. The reaction mixture was incubated at 37°C for 8 h.

S24). The above data prove that the 6-OH group of the glycosyl moiety is crucial for UGT94BY1 to catalyse the reaction, while other substitutions can affect how many oligoglucoside chains are formed.

To verify the catalytic function of UGT94BY1 on 28-O-glycosylated triterpenoids, compounds **16–20**, all of which lack free 6-OH groups at the C-3 sugar chain, were selected. The results showed that compounds **16**, **18** and **20** could be catalysed

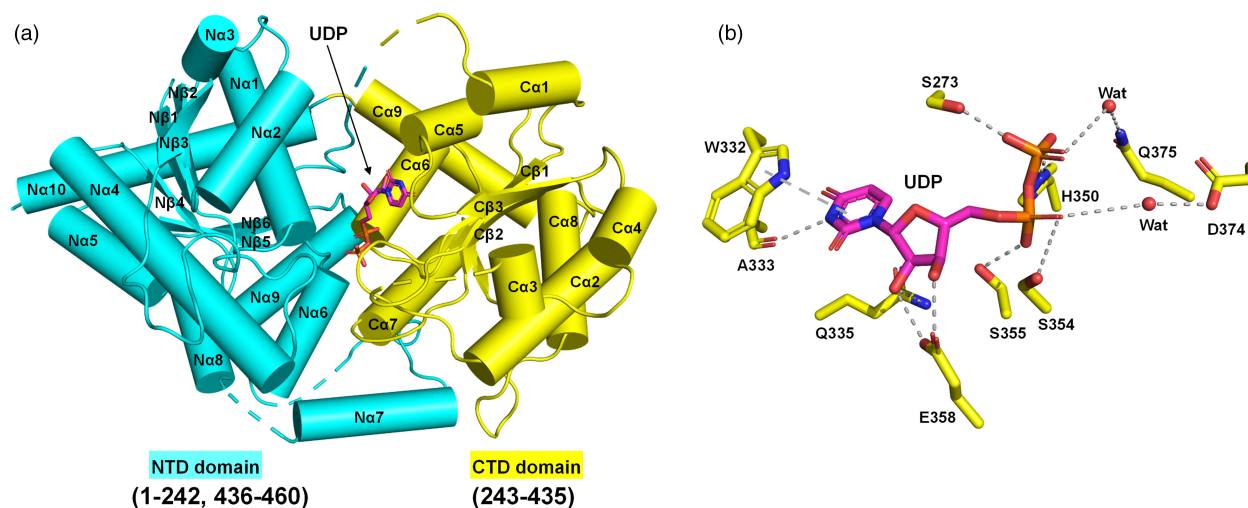


Figure 4 Crystal structure of UGT94BY1-UDP complex and the binding mode of UDP. (a) Overall structure of the UGT94BY1-UDP complex. The structure is depicted in cartoon representation, with the NTD domain coloured cyan and the CTD domain coloured yellow. The bound UDP is shown as sticks, with carbon atoms coloured magenta. (b) Binding mode of UDP in the UGT94BY1-UDP complex. UDP binding primarily involves the C-terminal domain, with hydrogen bonds indicated by dashed lines. Water molecules are shown as red spheres.

by UGT94BY1 (Figure 3; Figures S7, S25, S26). Meanwhile, the structures of compounds **17** and **19** are similar to those of **18** and **20**, respectively, with the only difference being the absence of glucose at the C28 position. Similar to when 3-*O*-glycosylated triterpenoids were used as sugar acceptors, UGT94BY1 recognized the 6-OH group of the glycosyl moiety at C-28, regardless of the substitutions at other positions.

The catalytic properties of UGT94BY1 in various flavonoid glycosides have been examined. The glucosyl moieties were attached to the A ring (**22** and **24**), B ring (**21** and **23**) and C ring (**25–27**) of flavonoids, respectively (Figure 3). UGT94BY1 was able to catalyse the conversion of compounds **21–25** to produce oligoglucoside chains (Figures S27–S30) but showed no activity toward compounds lacking a free 6-OH group (**26** and **27**). Additionally, we tested other types of phenolic sugar acceptors, including anthraquinone glycoside and phenylethanoid (Figures S31, S32). UGT94BY1 could also catalyse a stepwise three-step glycosylation of **38** (emodin 1-*O*-glucoside) and **39** (salidroside). By contrast, UGT94BY1 did not show activity for aglycones (Figure S33). Our findings indicate that the free 6-OH group of the glycosyl moiety is essential for UGT94BY1-mediated glycosylation.

To our knowledge, no enzyme with a comparable function to UGT94BY1 has been reported with the function of forming oligoglucoside chain for both triterpenes (tetracyclic and pentacyclic saponins) and phenols (flavonoid, anthraquinone and phenylethanoid glycosides). Therefore, UGT94BY1 can serve as a biocatalyst for the efficient synthesis of β -(1,6) glycosides with potential medicinal value.

Crystal structure of the UGT94BY1/UDP complex

Structure is crucial for understanding catalytic mechanisms and guiding rational enzyme engineering. Although lots of crystal structures of plant UGTs have been solved in recent years, most are specific to flavonoids as sugar acceptors. By contrast, obtaining the crystal structures of UGTs that target terpenoids remains challenging. To date, only five crystal structures of terpenoid UGTs have been reported, including those of

SgUGT94-289-3 (*Siraitia grosvenorii*) (Cui *et al.*, 2024), UGT74AC1 (*Siraitia grosvenorii*) (Li *et al.*, 2020), OsUGT91C1 (*Oryza sativa*) (Zhang *et al.*, 2021), UGT76G1 (*Stevia rebaudiana*) (Yang *et al.*, 2019) and UGT71G1 (*Medicago truncatula*) (Shao *et al.*, 2005).

To discover the mechanism of substrate promiscuity and sugar chain extension of UGT94BY1, we solved the crystal structure of UGT94BY1 complexed with UDP at a resolution of 2.0 Å (Figure 4; Table S2). UGT94BY1 is the first multi-step *O*-glycosyltransferase catalysing triterpene saponins and flavonoid glycosides with a crystal structure (PDB ID: 9IMI). It adopts a typical GT-B fold, comprising of two Rossmann-like $\beta/\alpha/\beta$ domains positioned opposite each other with a deep cleft in between. The N-terminal domain (NTD, residues 1–242 and 436–460) and the C-terminal domain (CTD, residues 243–435) are primarily responsible for the binding of the sugar acceptor and sugar donor, respectively.

In the UGT94BY1/UDP complex structure, the electron density of UDP is clearly observed in the CTD domain, and the UDP-binding pocket is conserved in other GT-B type UGTs. UDP adopts an elongated conformation and is oriented toward the sugar acceptor pocket. Within the UDP binding pocket, UDP is surrounded by the $\alpha 5$ and $\alpha 6$ helices, as well as loops at the C-terminal side of the β strands (C β 1 and C β 3). UDP interacts primarily with residues through hydrogen bonds and π - π stacking interactions. The uridine ring of UDP stacks with Trp332 via a π - π interaction and forms a hydrogen bond with the backbone carbonyl of Ala333. The 2- and 3-hydroxyl groups of the ribose ring form hydrogen bonds with Glu358, while Gln335 forms a hydrogen bond with the 2-hydroxyl of the ribose. The pyrophosphate group of UDP forms hydrogen bonds with His350, Ser354, Ser355 and Ser273. Additionally, Asp374 and Gln375 form water-mediated hydrogen bonds with the pyrophosphate of UDP.

Mechanism of the sugar chain extension and substrate promiscuity

We attempted to obtain the crystal structure of the complex with the sugar acceptor by optimizing the protein/ligand

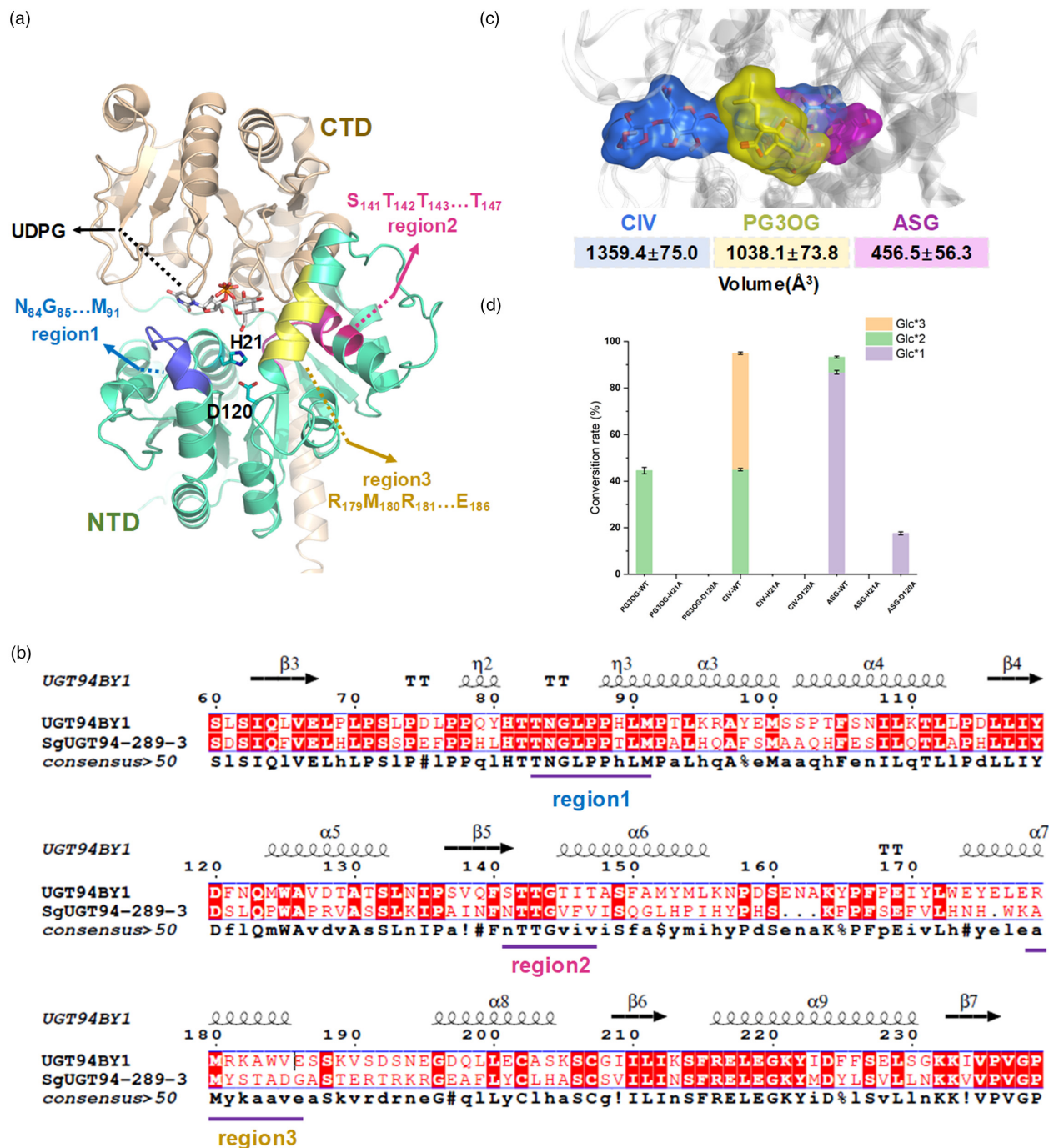


Figure 5 Binding modes of UGT94BY1 with PG3OG (**2**), CIV (**16**) and ASG (**25**). (a) The structural representation highlights the NTD (green) and CTD (dark grey) of the protein, with UPG located in the CTD (grey lines). The NTD features three regions: Region1 (blue, residues 84–91), Region2 (purple, residues 141–147) and Region3 (yellow, residues 179–186). (b) Sequence alignment of UGT94BY1 and SgUGT94-289-3. The blue surface represents CIV (Å³), with a volume of 1359.4 ± 75.0; the yellow surface represents PG3OG, with a volume of 1038.1 ± 73.8; and the purple surface represents ASG, with a volume of 456.5 ± 56.3. (d) The glycosylation activities of H21A and D120A mutants.

concentration, pH, temperature and precipitant concentration, but ultimately, we were unsuccessful. Additionally, determining the precise binding modes of various glycosyl receptors in UGTs is challenging. The molecular lengths of 3-*O*-β-Glucosyl platycodigenin (PG3OG, **2**) and Chikusetsusaponin IV (CIV, **16**) are approximately 16 Å, while Astragalin (ASG, **25**) has a

substantially shorter length of approximately 9 Å (Figure S34). Consequently, we used a computationally induced-fit approach to produce the initial conformations, refined and validated through MD simulations (Figure 5). Based on the modelling results and guided by previous catalytic reaction mechanisms (Guo *et al.*, 2022; Li *et al.*, 2020; Wang *et al.*, 2023), we propose

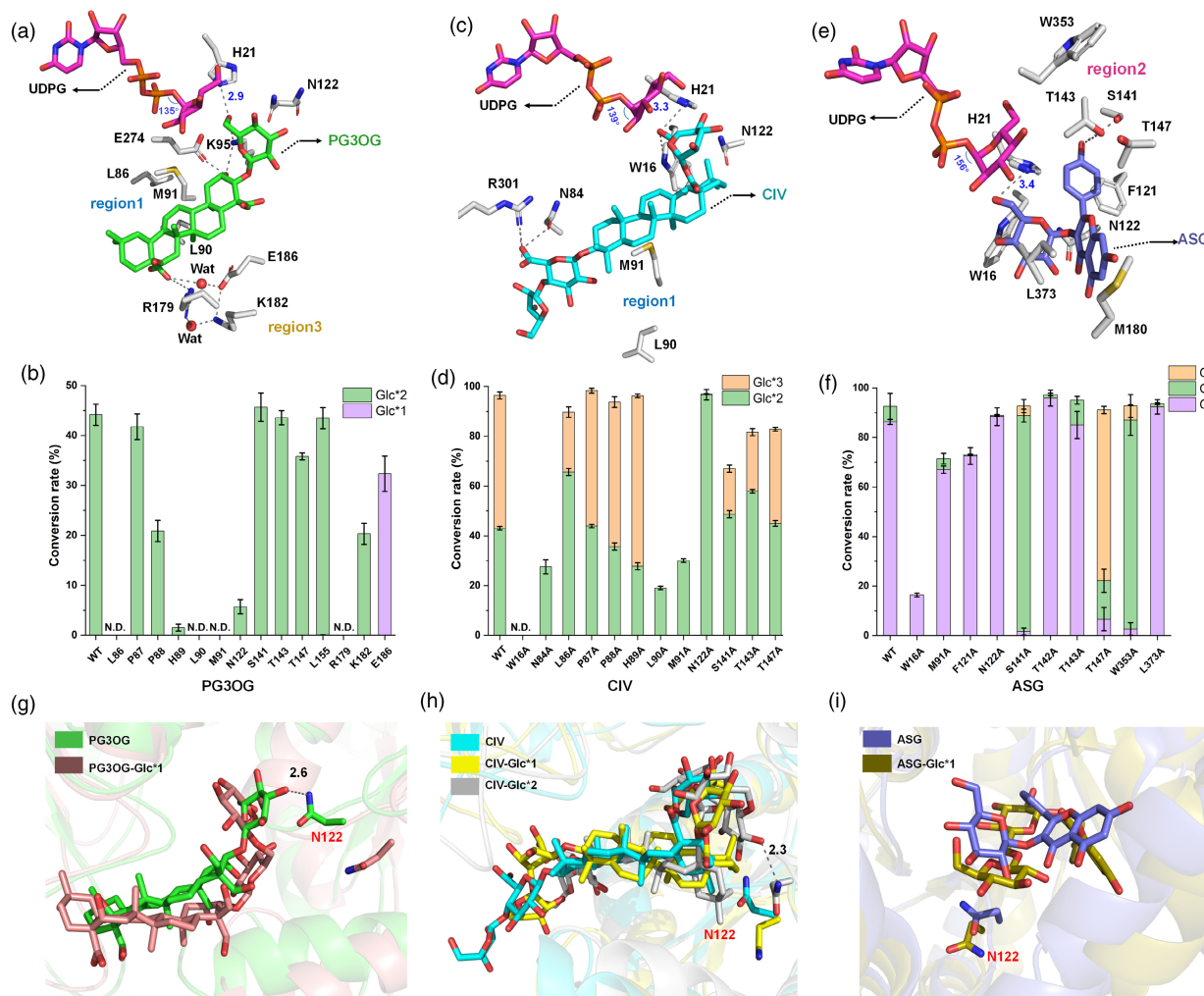


Figure 6 Multi-step strategy reveals substrate promiscuity and sugar chain formation mechanisms. (a) Binding modes of PG3OG with the protein, alongside binding free energy (MM/GBSA). Grey dashed lines represent distances and angles (the distance between the O of the substrate's 3-OH and the NE2 of H21 is less than 3.4 Å, and the angle between the O of the 3-OH, the anomeric carbon of UDPG and the O of UDPG is $140^\circ \pm 15^\circ$, indicating favourable reaction conditions, with their values indicated by blue numbers). Key residues are labelled in red font, and region1 and region3 are labelled in blue and yellow font, respectively. (b) Alanine scanning mutagenesis experiment (PG3OG was used as substrate). (c) Binding mode of CIV with the protein, highlighting key interacting residues, with region1 marked in blue and the tail region of CIV in yellow. (d) Alanine scanning mutagenesis experiment (CIV was used as substrate). (e) Binding mode of ASG with the protein, highlighting interacting residues in region2 (purple). (f) Alanine scanning mutagenesis experiment (ASG was used as substrate). (g–i) Superimposition of different sugar acceptors. All substrates were simulated with the glycosyl donor before the glycosylation reaction, and structures were selected from the conformations obtained by clustering based on MD trajectories. Residue N122 is labelled in red. Dashed lines represent hydrogen bond interactions, with numbers indicating hydrogen bond distances. The purple, green and orange bars in b, d and f indicate mono-, di- and tri-O-glycosylation products, respectively.

that three distinct regions (region1, region2 and region3, Figure 5a,b), exclusively located in the NTD, are likely the key structural motifs to modulate the substrate promiscuity of UGT94BY1. The flexibility of these residues enables the accommodation of various substrates (Figure 5c). Our simulations illustrate the adaptability of the pockets, with their volumes (\AA^3) for the protein complex of the three glycosyl receptors (PG3OG, CIV and ASG) being 1038.1, 1359.4 and 457.5, respectively. The presence of pockets adaptable for different types of sugar acceptors is similar to the recently reported SgUGT94-289-3, which adopts a dual-pocket organization in its active site to allow the two structurally distinct substrates locked in the active pocket for glycosylation (Cui *et al.*, 2024). The catalytic dyad of highly conserved H21 and D120, positioned near the receptor binding

pocket at the N-terminus, is crucial for triggering the catalytic reaction (Figure 5d).

MD simulations of the UGT94BY1/UDPG/PG3OG complex (Figure 6a) underscored the crucial role of R179 in forming electrostatic interactions (salt bridges) with the negatively charged carboxyl oxygen of PG3OG, maintaining a proximity of 4.1 Å (Figure S35) during the 200 ns MD simulation. This anchoring effect maintains the R179, E186 and K182 (region3), in a polar environment that stabilizes the polar tail of PG3OG (Figure S35), and is essential for enzymatic activity. This is consistent with experimental results summarized in (Figure 6b; Table S4), in which R179A and E186A mutations lead to enzyme inactivation, while K182A substantially limits enzyme activity. Building on this conclusion, MD simulations revealed

thermodynamic instability in the β -(1,2) conformation (Figure S36), with the 2-OH group deviating significantly from its initial position during unrestrained sampling. Binding free energy calculations demonstrated a marked energetic preference (-59.4 ± 3.3 kcal/mol) for the β -(1,6)-glycosidic bond conformation over the β -(1,2) counterpart (-18.7 ± 3.9 kcal/mol), emphasizing UGT94BY1's intrinsic selectivity for the β -(1,6)-glycosidic bond, which is governed by structural compatibility and energetic stabilization.

In addition, E274 and K95 stabilize the head of PG3OG. Regarding the interaction with the middle part of PG3OG, M91, although not directly interacting with PG3OG (Figure 6a), orchestrates its closeness by its position relative to three prolines, which modulate protein folding. Consequently, M91 serves as a keeper, preventing the sugar acceptor from the loop region. MD simulations revealed that the M91A mutant disrupts the salt bridge between R179 and PG3OG's tail, detaching the reactive site from the PG3OG head. Therefore, the binding free energy is decreased from -59.4 to -39.9 kcal/mol (Figure S37). Moreover, the hydrophobic residues near M91, including L86 and L90, were found to potentially impact the conformation of M91, leading to the loss of catalysis (Figure 6b; Table S3).

Unlike the binding mode of PG3OG, UGT94BY1/UDPG/CIV (Figure 6c; Table S4) exhibits a distinct binding pose. The C28 (head portion) of CIV is proximal to the catalytic dyad, while the middle portion of the CIV molecule is near W16, and the tail of CIV interlocks a sandwich-like structure between the N84-M91 region and R301. Notably, the stable conformation of CIV hinges on W16, as indicated by the loss of catalysis in the W16A mutant (Figure 6d). As the sugar chain extension (from CIV to CIV-Glc*2), the arabinosyl and glucuronosyl moieties at the tail approach N84 (Figure S38). During the first and second glycosylation steps, the sandwich-like structure maintained by N84-M91 (region1) and R301 remains stable. However, in the third glycosylation step, this structure shifts, with N84 engaging hydrogen bonds with arabinosyl and glucuronosyl moieties.

By contrast, M91 contributes hydrophobic interactions to anchor CIV-Glc*2 at the reaction centre. These computationally detected structural/conformational transitions were experimentally validated, as mutations N84A and M91A failed to produce CIV-Glc*3 (Figure 6d; Table S4). Regarding region2 near the dyad (Figure 5a), it comprises small-size residues that facilitate the elongation of two types of triterpenoid substrate sugar chains, with direct interactions with the substrate (Figure S38), except for CIV-Glc*2. Experimental findings (Figure 6d) demonstrate the negligible impact of mutations of S141, T143 and T147 in region2 on the first and second glycosylation of PG3OG and CIV. However, mutations impacting this region slightly reduce the yield of CIV-Glc*3. This is likely due to its proximity to the three glucose units at the head of CIV-Glc*2.

The flavonoid substrate ASG has a small molecular size (Figure 6e). In this study, two models of UGT94BY1/UDPG/ASG (Figure S39) were considered: (1) B-ring facing region2, (2) A/C-ring facing region2. Accurate conformation characterization was obtained by measuring the distances between H21 and 6''-OH of ASG (first binding mode, 3.4 Å vs second binding mode, 4.5 Å) and the binding free energies calculations (first binding mode, -35.2 kcal/mol, vs second binding mode, -26.3 kcal/mol). Thus, we confirmed priority of the 'B-ring toward region2' conformation of ASG (Figure 6e). In particular, W16 forms a crucial hydrogen bond with ASG, anchoring its orientation. S141, T143 and T147 in region2 disrupt stable π - π stacking interactions with F121 by

attracting the hydroxyl group at the end of the B-ring. Moreover, W353 brings substantial steric hindrance, allowing region2 to be closer to the binding pocket. This interaction mechanism is robustly supported by mutation experiments (Figure 6f). The W16A mutation markedly reduces the yield of the first-step glycosylation, while S141A, T147A and W353A mutants enhance the product yields. Notably, most mutations in region2 promote significant increases in higher glycosylation products, including ASG-Glc*2 and ASG-Glc*3. To uncover the mechanisms behind this complex phenomenon, further MD simulations on S141A mutants (targeting ASG and ASG-Glc*1) were performed (Figure S40). Our simulations identified subtle structural alterations in the ASG structure in the S141A mutant compared with the wild-type, positioning the critical reaction centre oxygen atoms closer to H21 to promote reactions. Additionally, we identified that a hydrophobic environment surrounding the A-ring, orchestrated by M180 and L373, emerges, significantly modulating the glycosyl acceptor's spatial arrangement. Modelling of M180A and L373A mutants was also performed (Figure S41), with simulations predicting enhanced glycosylation reaction ability, which was experimentally validated (Figure 6f; Table S4). These findings indicate that M180A and L373A mutants significantly enhanced ASG glycosylation efficiency.

Apart from region1/2/3, N122 is a crucial residue in the sugar chain extension of triterpenoid substrates (Figure 6g,h). Specifically, N122 stabilizes the arrangement of PG3OG by directly forming hydrogen bonds with its sugar moiety while not interacting with PG3OG-Glc*1. For the CIV substrate, N122 exerts a pronounced effect on the sugar acceptor CIV-Glc*2 during the third glycosylation step. Experimental findings (Figure 6b,d) demonstrate that the N122A mutation substantially decreases the conversion rate of PG3OG-Glc*2 and completely abolishes the formation of CIV-Glc*3. Regarding ASG, both of the simulations (Figure 6i) and experiments (Figure 6f) reveal that N122 does not participate in hydrogen bonding with ASG or ASG-Glc*1, exhibiting no effect on the formation of sugar chain for the N122A mutation.

Conclusion

Overall, we identified UGT94BY1 as a promiscuous and regio-specific multi-step *O*-glycosyltransferase involved in β -(1,6) glucosidic bonds biosynthesis and determined its crystal structure at a 2.0 Å resolution. UGT94BY1 is capable of simultaneously recognizing both phenolic glycosides and triterpenoid saponins to generate β -(1,6) oligoglycoside chain, demonstrating significant potential for synthesizing drug potential β -(1,6) glycosides. Through investigation of its substrate spectrum, we identified the 6-hydroxy group of the glucose moiety as the key recognition site for UGT94BY1-mediated glycosylation. Thus, sugar acceptors possessing this feature are likely to be recognized by UGT94BY1 and converted into sugar chain glycosides. Additionally, we solved the crystal structure of UGT94BY1/UDP complex and performed MD simulations and MM/GBSA calculations. These analyses revealed three key regions of residues responsible for sugar acceptor recognition and preference, which were further validated through functional assays, structural analysis and theoretical calculations. This study provides new insights into the regio-specific mechanisms of plant glycosyltransferases involved in sugar chain biosynthesis and thus offering a foundation for the design of improved enzymes with tailored substrate specificities for future biotechnological applications.

Acknowledgements

We thank Hongli Jia (Peking University) for X-ray diffraction tests, Fen Liu (Peking University) for LC–MS analyses, and the staffs from beamlines BL10U2 (NFPS, Shanghai) for assistance during data collection. This work was supported by National Science Foundation for Distinguished Young Scholars (82325049), Key project at central government level: The ability establishment of sustainable use for valuable Chinese medicine resources (2060302) and Science and Technology Innovation Project of CACMS (CI2023D001/CI2023E002-04).

Conflict of interest

The authors declare no competing financial interest.

Author contributions

Y.Y., Z.W., L.H. and Z.J. designed the research; Z.J., H.W., X.D., Y.T., Z.W. and N.C. performed the research; Z.J., H.W., R.W. and N.C. contributed new analytic and computational tools; Z.J. and Z.W. analysed data, W.Z., Z.J. and Y.Y. wrote the paper. All authors agreed to the list of authors and the identified contributions of those authors.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2021) in National Genomics Data Center (Nucleic Acids Res 2024), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA019288) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. The gene sequence of UGT94BY1 has been deposited in the NCBI database. The crystal structure in this study has been deposited in the RCSB PDB database (PDB ID: 9IMI).

References

- Chen, K., Zhang, M., Gao, B., Hasan, A., Li, J., Bao, Y.O., Fan, J. *et al.* (2023) Characterization and protein engineering of glycosyltransferases for the biosynthesis of diverse hepatoprotective cycloartane-type saponins in *Astragalus membranaceus*. *Plant Biotechnol. J.* **21**(4), 698–710.
- Chen, Y., Liu, Q., An, P., Jia, M., Luan, X., Tang, J. and Zhang, H. (2022) Ginsenoside Rd: a promising natural neuroprotective agent. *Phytomedicine* **95**, 153883.
- Chen, Y., Yan, Q., Ji, Y., Bai, X., Li, D., Mu, R., Guo, K. *et al.* (2023) Unraveling the serial glycosylation in the biosynthesis of steroidal saponins in the medicinal plant *Paris polyphylla* and their antifungal action. *Acta Pharm. Sin. B* **13**, 4638–4654.
- Cui, S., Zhang, S., Wang, N., Su, X., Luo, Z., Ma, X. and Li, M. (2024) Structural insights into the catalytic selectivity of glycosyltransferase SgUGT94-289-3 towards mogrosides. *Nat. Commun.* **15**, 6423.
- Deng, Y., Yang, P., Zhang, Q., Wu, Q., Feng, L., Shi, W., Peng, Q. *et al.* (2024) Genomic insights into the evolution of flavonoid biosynthesis and O-methyltransferase and glucosyltransferase in *Chrysanthemum indicum*. *Cell Rep.* **43**, 113725.
- Dong, X.D., Liu, Y.N., Yu, S.S., Ji, H.Y., Feng, Y.Y., Liu, A. and Yu, J. (2021) Extraction, optimization, and biological activities of a low molecular weight polysaccharide from *Platycodon grandiflorus*. *Ind. Crop. Prod.* **165**, 113427.

- Fernández-Tejada, A., Tan, D.S. and Gin, D.Y. (2015) Versatile strategy for the divergent synthesis of linear oligosaccharide domain variants of quillaja saponin vaccine adjuvants. *Chem. Commun.* **51**, 13949–13952.
- Gharabli, H., Della Gala, V. and Welner, D.H. (2023) The function of UDP-glycosyltransferases in plants and their possible use in crop protection. *Biotechnol. Adv.* **67**, 108182.
- Guo, B., Deng, Z., Meng, F., Wang, Q., Zhang, Y., Yuan, Z. and Rao, Y. (2022) Enhancement of rebudioside m production by structure-guided engineering of glycosyltransferase UGT76G1. *J. Agric. Food Chem.* **70**, 5088–5094.
- Guo, Y., Chen, X., Gong, P., Long, H., Wang, J., Yang, W. and Yao, W. (2024) *Siraitia grosvenorii* as a homologue of food and medicine: a review of biological activity, mechanisms of action, synthetic biology, and applications in future food. *J. Agric. Food Chem.* **72**, 6850–6870.
- Jia, Y., Chen, S., Chen, W., Zhang, P., Su, Z., Zhang, L., Xu, M. *et al.* (2022) A chromosome-level reference genome of Chinese balloon flower (*Platycodon grandiflorus*). *Front. Genet.* **13**, 869784.
- Johnson, J.B., Mani, J.S., Broszczak, D., Prasad, S.S., Ekanayake, C.P., Strappe, P., Valeris, P. *et al.* (2021) Hitting the sweet spot: a systematic review of the bioactivity and health benefits of phenolic glycosides from medicinally used plants. *Phytother. Res.* **35**, 3484–3508.
- Kim, J.H., Yi, Y.S., Kim, M.Y. and Cho, J.Y. (2017) Role of ginsenosides, the main active components of panax ginseng, in inflammatory responses and diseases. *J. Ginseng Res.* **41**, 435–443.
- Kulkarni, S.S., Wang, C.C., Sabbavarapu, N.M., Podilapu, A.R., Liao, P.H. and Hung, S.C. (2018) “One-Pot” protection, glycosylation, and protection–glycosylation strategies of carbohydrates. *Chem. Rev.* **118**, 8025–8104.
- Kurze, E., Wüst, M., Liao, J., McGraphery, K., Hoffmann, T., Song, C. and Schwab, W. (2022) Structure–function relationship of terpenoid glycosyltransferases from plants. *Nat. Prod. Rep.* **39**, 389–409.
- Li, J., Mu, S., Yang, J., Liu, C., Zhang, Y., Chen, P., Zeng, Y. *et al.* (2022) Glycosyltransferase engineering and multi-glycosylation routes development facilitating synthesis of high-intensity sweetener mogrosides. *iScience* **27**, 105222.
- Li, J., Yang, J., Mu, S., Shang, N., Liu, C., Zhu, Y., Cai, Y. *et al.* (2020) Efficient O-glycosylation of triterpenes enabled by protein engineering of plant glycosyltransferase UGT74AC1. *ACS Catal.* **10**, 3629–3639.
- Li, J.Y., Cao, H.Y., Liu, P., Cheng, G.-H. and Sun, M.Y. (2014) Glycyrrhizic acid in the treatment of liver diseases: literature review. *Biomed. Res. Int.* **2014**, 872139.
- Li, X., Liu, J., Zuo, T.T., Hu, Y., Li, Z., Wang, H.D., Xu, X.Y. *et al.* (2022) Advances and challenges in ginseng research from 2011 to 2020: the phytochemistry, quality control, metabolism, and biosynthesis. *Nat. Prod. Rep.* **39**, 875–909.
- Li, Y., Wang, J., Li, L., Song, W., Li, M., Hua, X., Wang, Y. *et al.* (2023) Natural products of pentacyclic triterpenoids: from discovery to heterologous biosynthesis. *Nat. Prod. Rep.* **40**, 1303–1353.
- Liu, Y., Zhao, X., Gan, F., Chen, X., Deng, K., Crowe, S.A., Hudson, G.A. *et al.* (2024) Complete biosynthesis of QS-21 in engineered yeast. *Nature* **629**, 937–944.
- Lucier, R., Kamileen, M.O., Nakamura, Y., Serediuk, S., Barbole, R., Wurllitzer, J., Kunert, M. *et al.* (2024) Steroidal scaffold decorations in solanum alkaloid biosynthesis. *Mol. Plant* **17**, 1236–1254.
- Ma, J.Q., Dong, A.B., Xia, H.Y. and Wen, S.Y. (2023) Preparation methods, structural characteristics, and biological activity of polysaccharides from *Platycodon grandiflorus*. *Int. J. Biol. Macromol.* **12**, 9106.
- Martin, L.B., Kikuchi, S., Rejzek, M., Owen, C., Reed, J., Orme, A., Misra, R.C. *et al.* (2024) Complete biosynthesis of the potent vaccine adjuvant QS-21. *Nat. Chem. Biol.* **20**(4), 493–502.
- Meng, F.B., Chu, T.Z., Hu, L.J., Zhang, M.Q., Cheng, Q., Yang, X.P., Liu, Z. *et al.* (2024) TCMPG 2.0: an enhanced database of traditional Chinese medicine plant genomes. *Med. Plant Biol.* **3**, e003.
- Mrudulakumari, V.U. and Lee, E.Y. (2020) Flavonoids, terpenoids, and polyketide antibiotics: role of glycosylation and biocatalytic tactics in engineering glycosylation. *Biotechnol. Adv.* **41**, 107550.
- Panza, M., Pistorio, S.G., Stine, K.J. and Demchenko, A.V. (2018) Automated chemical oligosaccharide synthesis: novel approach to traditional challenges. *Chem. Rev.* **118**, 8105–8150.

- Seki, H., Tamura, K. and Muranaka, T. (2015) P450s and UGTs: key players in the structural diversity of triterpenoid saponins. *Plant Cell Physiol.* **56**, 1463–1471.
- Shao, H., He, X.Z., Achnine, L., Blount, J.W., Dixon, R.A. and Wang, X.Q. (2005) Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. *Plant Cell* **17**(11), 3141–3154.
- Silva Magedans, Y.V., Phillips, M.A. and Fett-Neto, A.G. (2021) Production of plant bioactive triterpenoid saponins: from metabolites to genes and back. *Phytochem. Rev.* **20**, 461–482.
- Tang, W., Shi, J.J., Liu, W., Lu, X. and Li, B. (2023) MALDI Imaging assisted discovery of a di-O-glycosyltransferase from *Platycodon grandiflorum* root. *Angew. Chem. Int. Ed.* **135**, e202301309.
- Wang, H.T., Wang, Z.L., Chen, K., Yao, M.J., Zhang, M., Wang, R.S., Zhang, J.H. et al. (2023) Insights into the missing apiosylation step in flavonoid apiosides biosynthesis of leguminosae plants. *Nat. Commun.* **14**, 6658.
- Wang, Z.L., Zhou, J.J., Han, B.Y., Hasan, A., Zhang, Y.Q., Zhang, J.H., Wang, H.D. et al. (2022) *GuRhaGT*, a highly specific saponin 2''-O-rhamnosyltransferase from *Glycyrrhiza uralensis*. *Chem. Commun.* **58**, 5277–5280.
- Xie, L., Zhao, Y.X., Zheng, Y. and Li, X.F. (2023) The pharmacology and mechanisms of platycodin d, an active triterpenoid saponin from *Platycodon grandiflorus*. *Front. Pharmacol.* **14**, 1148853.
- Yang, T., Zhang, J., Ke, D., Yang, W. and Zhu, X. (2019) Hydrophobic recognition allows the glycosyltransferase UGT76G1 to catalyze its substrate in two orientations. *Nat. Commun.* **10**, 3214.
- Zaynab, M., Sharif, Y., Abbas, S., Afzal, M.Z., Qasim, M., Khalofah, A., Ansari, M.J. et al. (2021) Saponin toxicity as key player in plant defense against pathogens. *Toxicon* **193**, 21–27.
- Zhang, J., Tang, M., Chen, Y., Ke, D., Zhou, J., Xu, X., Yang, W. et al. (2021) Catalytic flexibility of rice glycosyltransferase OsUGT91C1 for the production of palatable steviol glycosides. *Nat. Commun.* **12**, 7030.
- Zhang, M., Yi, Y., Gao, B.H., Su, H.F., Bao, Y.O., Shi, X.M., Wang, H.D. et al. (2022) Functional characterization and protein engineering of a triterpene 3-/6-/2'-O-glycosyltransferase reveal a conserved residue critical for the regioselectivity. *Angew. Chem. Int. Ed.* **61**, e202113587.
- Zhang, S., Chai, X., Hou, G., Zhao, F. and Meng, Q. (2022) *Platycodon grandiflorum* (Jacq.) A. DC.: a review of phytochemistry, pharmacology, toxicology and traditional use. *Phytomedicine* **106**, 154422.
- Zou, J., Li, H., Wang, Z. and Ye, M. (2023) Functional characterization of two efficient glycosyltransferases catalysing the formation of rutin from *Sophora japonica* L. *Org. Biomol. Chem.* **21**, 7913–7916.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 *Platycodon grandiflorum* plants. (A) Four-year-old whole plants (aerial parts) of *P. grandiflorum*. (B) One-year-old roots of *P. grandiflorum*. The red scale bar indicates 4 cm.

Figure S2 Functional characterization of seven candidate genes. Compounds I–VI were used as sugar acceptors.

Figure S3 SDS-PAGE analysis of recombinant UGT94BY1 purified by Ni-NTA affinity chromatography, ion-exchange chromatography and Gel-filtration.

Figure S4 LC/MS analysis of the products catalyzed by UGT94BY1 using **1** as the substrate.

Figure S5 Effects of pH (A), temperature (B), and divalent metal ions (C) on enzyme activity of UGT94BY1.

Figure S6 Probing sugar donor promiscuity of UGT94BY1. Relative conversion rates of deapiplatycodin D (**1**) using different sugar donors.

Figure S7 LC/MS analysis of the products catalyzed by UGT94BY1 using **16** as the substrate.

Figure S8 The ^1H NMR spectrum of **16c** in pyridine- d_5 (600 MHz).

Figure S9 The ^{13}C NMR spectrum of **16c** in pyridine- d_5 (150 MHz).

Figure S10 The HSQC spectrum of **16c** in pyridine- d_5 (600 MHz).

Figure S11 The HMBC spectrum of **16c** in pyridine- d_5 (600 MHz).

Figure S12 LC/MS analysis of the products catalyzed by UGT94BY1 using **25** as the substrate.

Figure S13 LC/MS analysis of the products catalyzed by UGT94BY1 using **2** as the substrate.

Figure S14 LC/MS analysis of the products catalyzed by UGT94BY1 using **3** as the substrate.

Figure S15 LC/MS analysis of the products catalyzed by UGT94BY1 using **4** as the substrate.

Figure S16 LC/MS analysis of the products catalyzed by UGT94BY1 using **5** as the substrate.

Figure S17 LC/MS analysis of the products catalyzed by UGT94BY1 using **6** as the substrate.

Figure S18 LC/MS analysis of the products catalyzed by UGT94BY1 using **7** as the substrate.

Figure S19 LC/MS analysis of the products catalyzed by UGT94BY1 using **8** as the substrate.

Figure S20 LC/MS analysis of the products catalyzed by UGT94BY1 using **9** as the substrate.

Figure S21 LC/MS analysis of the products catalyzed by UGT94BY1 using **10** as the substrate.

Figure S22 LC/MS analysis of the products catalyzed by UGT94BY1 using **11** as the substrate.

Figure S23 LC/MS analysis of the products catalyzed by UGT94BY1 using **12** as the substrate.

Figure S24 LC/MS analysis of the products catalyzed by UGT94BY1 using **13** as the substrate.

Figure S25 LC/MS analysis of the products catalyzed by UGT94BY1 using **18** as the substrate.

Figure S26 LC/MS analysis of the products catalyzed by UGT94BY1 using **20** as the substrate.

Figure S27 LC/MS analysis of the products catalyzed by UGT94BY1 using **21** as the substrate.

Figure S28 LC/MS analysis of the products catalyzed by UGT94BY1 using **22** as the substrate.

Figure S29 LC/MS analysis of the products catalyzed by UGT94BY1 using **23** as the substrate.

Figure S30 LC/MS analysis of the products catalyzed by UGT94BY1 using **24** as the substrate.

Figure S31 LC/MS analysis of the products catalyzed by UGT94BY1 using **38** as the substrate.

Figure S32 LC/MS analysis of the products catalyzed by UGT94BY1 using **39** as the substrate.

Figure S33 The structures of ten triterpenoid aglycones (**28–37**).

Figure S34 The sizes of 3-O- β -Glucosyl platycodigenin (**2**), Chikusetsusaponin IV (**16**), and Astragalin (**25**).

Figure S35 The key distance changes of PG3OG and PG3OG-Glc*1 after 200 ns of molecular dynamics (MD).

Figure S36 Structural and energetic basis for UGT94BY1's preference for β -(1,6)-glycosidic bond formation.

Figure S37 Binding modes of PG3OG with wild-type UGT94BY1 (WT) and its M91A mutant, along with their binding free energy (MM/GBSA).

Figure S38 The binding mode of the CIV three-step glycosylation substrate with UGT94BY1.

Figure S39 Determining the precise binding mode of ASG.

Figure S40 The binding mode of ASG and ASG-Glc*1 in both WT and S141A.

Figure S41 The binding mode of ASG and ASG-Glc*1 in both WT, M180A, L373A.

Table S1 Primers for molecular cloning of PgOGTs.

Table S2 Data collection and refinement statistics of UGT94BY1 in complex with UDP.

Table S3 Primers for site-directed mutagenesis of UGT94BY1.

Table S4 Catalytic activities of UGT94BY1 and its mutants toward PG3OG, CIV, and AGS.