An efficient *C*-glycoside production platform enabled by rationally tuning the chemoselectivity of glycosyltransferases

Min Li [1,2,3,4,8], Yang Zhou [5,8], Zexing Wen[1], Qian Ni[1], Ziqin Zhou[5], Yiling Liu[5], Qiang Zhou[6], Zongchao Jia [7], Bin Guo[1], Yuanhong Ma [1], Bo Chen[1], Zhi-Min Zhang [5] ✉ & Jian-bo Wang [1,2,3,4] ✉

Despite the broad potential applications of *C*-glycosides, facile synthetic methods remain scarce. Transforming glycosyltransferases with promiscuous or natural *O*-specific chemoselectivity to *C*-glycosyltransferases is challenging. Here, we employ rational directed evolution of the glycosyltransferase MiCGT to generate MiCGT-QDP and MiCGT-ATD mutants which either enhance *C*-glycosylation or switch to *O*-glycosylation, respectively. Structural analysis and computational simulations reveal that substrate binding mode govern *C*-/*O*-glycosylation selectivity. Notably, directed evolution and mechanism analysis pinpoint the crucial residues dictating the binding mode, enabling the rational design of four enzymes with superior non-inherent chemoselectivity, despite limited sequence homology. Moreover, our best mutants undergo testing with 34 substrates, demonstrating superb chemoselectivities, regioselectivities, and activities. Remarkably, three *C*-glycosides and an *O*-glycoside are produced on a gram scale, demonstrating practical utility. This work establishes a highly selective platform for diverse glycosides, and offers a practical strategy for creating various types of glycosylation platforms to access pharmaceutically and medicinally interesting products.

Carbohydrate chemistry is one of the most important branches of chemistry, its impact transcending the boundaries of chemistry to influence various scientific realms encompassing life sciences, medicine, and food-related disciplines[1–3]. Glycosylation, serving as the core reaction in this field, relies on efficiency and selectivity as two critical benchmarks that demonstrate the applied potential of the respective synthetic method[4]. Subtle deviations in the formation of glycosidic bonds enable an enormous breadth of bioactivity within a core molecular skeleton[5]. Despite significant advancements in both chemical catalysis and biocatalysis within modern chemical technology[6–8], the development of catalysts for glycosylation with high stereoselectivity (reactions occurring preferentially on one stereoisomeric substrate or forming one stereoisomeric product), chemoselectivity (reactions occurring on the different functional groups) and

[1]Key Laboratory of Chemical Biology and Traditional Chinese Medicine Research (Ministry of Education) and Key Laboratory of Phytochemical R&D of Hunan Province, College of Chemistry and Chemical Engineering, Hunan Normal University, 410081 Changsha, P. R. China. [2]Department of Microbiology, Zhejiang University School of Medicine, Hangzhou 310058, P. R. China. [3]Key Laboratory of Multiple Organ Failure (Zhejiang University), Ministry of Education, Department of General Intensive Care Unit of the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310058, P. R. China. [4]Institute of Pharmaceutical Biotechnology, Zhejiang University School of Medicine, Hangzhou 310058, P. R. China. [5]State Key Laboratory of Bioactive Molecules and Draggability Assessment, Jinan University, Guangzhou 511436, P. R. China. [6]State Key Laboratory of Chemical Biology, Shanghai Institute of Organic Chemistry, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200032, China. [7]Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Ontario K7L 3N6, Canada. [8]These authors contributed equally: Min Li, Yang Zhou. ✉e-mail: zhangzm@jnu.edu.cn; jwang2023@zju.edu.cn

regioselectivity (reactions occur on the same functional groups, but at different positions) remains an ongoing challenge[9,10].

In nature, glycosyltransferases (GTs) are biocatalysts that exhibit remarkable selectivity during glycosylation. Uridine diphosphate-dependent glycosyltransferases (UGTs; EC 2.4.X.Y), a prevalent sub-class of GTs, play vital roles in natural glycoside biosynthesis[11,12]. These enzymes facilitate the transfer of activated glycosyls to various organic molecules, forming *C*-, *N*-, *O*- and *S*-glycosides, and thereby provide a direct and simple route to sugar-decorated products[13]. Among them, *C*-glycosides are frequently chosen as synthetic surrogates or mimics for other types of native glycosides in therapeutic agent development, owing to their resistance to in vivo hydrolytic enzymes (Fig. S1)[14–16]. However, the majority of known natural GTs are *O*-GTs, while the other three types occur rarely[17]. Hence, understanding the mechanism of these chemoselectivities and the engineering of *O*-UGTs to obtain UGTs with modified chemoselectivity in a controlled manner is of significant importance. Relevant studies are scarce. To the best of our knowledge, only four studies have been documented that address the modification of the initial selectivity in UGTs[18–21]. While these studies indeed revealed changes in selectivity concerning substrate specificity, only the Nidetzky group[19] and Ye group[20] actually engineered chemoselectivity (i.e., regioselectivity), whereas the other groups[18,21] focused on the modification of substrate specificity. Moreover, though the Ye group achieved successful transformation from *C*- to *O*-glycosylation, and vice versa by the Nidetzky group, both transformations incurred some loss in enzyme activity (Fig. S2).

To deepen our understanding of the chemoselectivity of UGTs and to establish the construction of a practical *C*-glycosides production platform, we reasoned that directed evolution[22–25] should be an efficient tool for the mechanism investigation via modifying the model enzyme. Additionally, leveraging insights derived from the elucidated mechanism, the rational design of chemoselectivity for a wide range of GTs and thus construction of an efficient *C*-glycosylation platform should become possible as a highly desirable goal. Based on the above analysis, in this work, we performed directed evolution targeting a trifunctional UDP-dependent CGT from *Mangifera indica* (MiCGT)[26–28]. By directing our engineering efforts to encompass both specific *C*- and

O-glycosylation, and employing crystal structure resolution and computational simulations, we uncovered the rationale for the *C*-/*O*-glycosylation switch. This led to further advancements in the rational tuning of chemoselectivity for four GTs beyond MiCGT. Ultimately, based on the most effective mutants, we established a *C*-glycosylation platform for versatile glycosides with outstanding chemoselectivities, regioselectivities, and activities (Fig. 1). Thus, we have achieved the dual goals of resolving the chemoselective mechanism of glycosyltransferases and constructing an efficient *C*-glycosylation platform.

Results

Initial characterization of MiCGT

Previous research has shown that MiCGT is an inversion UGT with promiscuous chemoselectivity and broad substrate scope[26–28]. Our initial contribution[28] was adjusting its regioselectivity and substrate specificity towards flavonoids and determining structures of MiCGT-WT and MiCGT-VFAH. Thus, owing to its inherent properties and our familiarity with it, MiCGT serves as a suitable model enzyme. Since bis(2,4-dihydroxyphenyl)-methanone (**1a**) has been reported to undergo both *O*- and *C*-glycosylation, resulting in products **2a** and **3a**, and considering its affordable commercial availability, it was selected as the model substrate in our present study.

For the sake of high-throughput screening convenience, the ability to utilize whole cells as the catalyst is necessary. Therefore, MiCGT-WT was first expressed in *E. coli*, and the resultant cells were used as catalyst directly to test the activity toward substrate **1a**. Fortunately, we observed the production of **2a** and **3a** (**2a**: **3a** = 6 : 1) in this system (Fig. S3). These results indicated that the system we selected is suitable for the further mechanism study via directed evolution.

Directed evolution of MiCGT for improved catalytic activity and chemoselectivity

Directed evolution has proven to be an invaluable tool for tailoring the catalytic functions of enzymes[22–25,29–32]. However, for UGTs, most studies report the investigation of regioselectivity[33–35], with few attempts documented to study chemoselectivity. To carry out our engineering process, substrate **1a** was docked into the crystal structure of MiCGT-
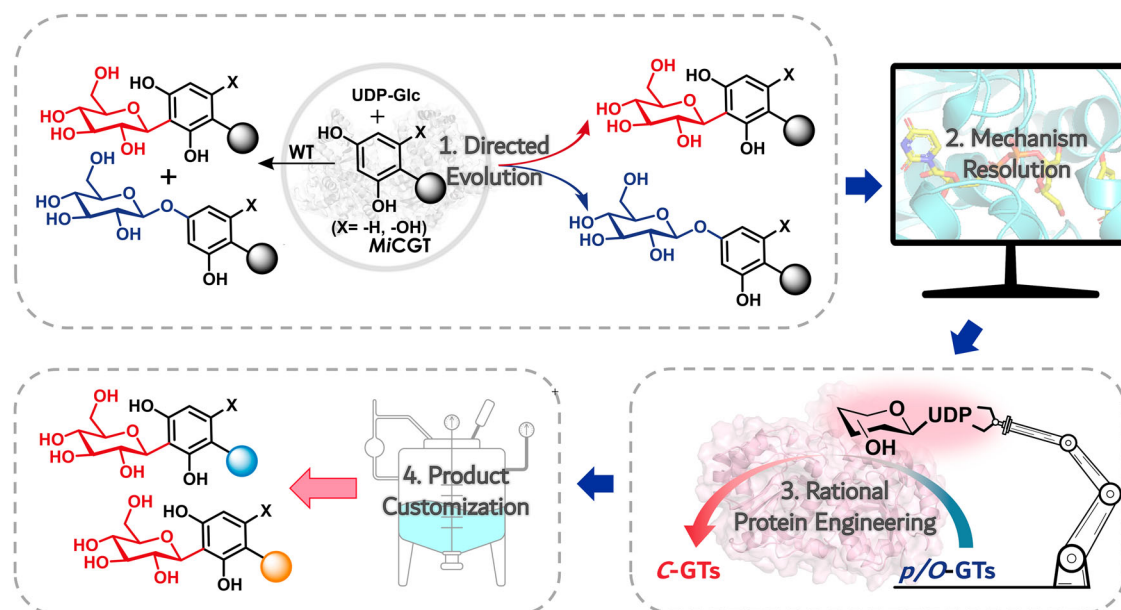

Fig. 1 | Universal strategy for construction of customizable *C*-glycosylation platform. 1. Directed evolution of glycosyltransferases with promiscuous chemoselectivity to generate mutants with enhanced or switched chemoselectivity; 2. Investigation of the mechanism of the change in chemoselectivity 3. Rational design of other glycosyltransferase under the guidance of the resolved mechanism; 4. Construction of the *C*-glycosylation platform based on the generated GTs mutants for the customization of different glycosides. WT wild type, *p*/*O*-GTs *Promiscuous*/*O*-GTs.
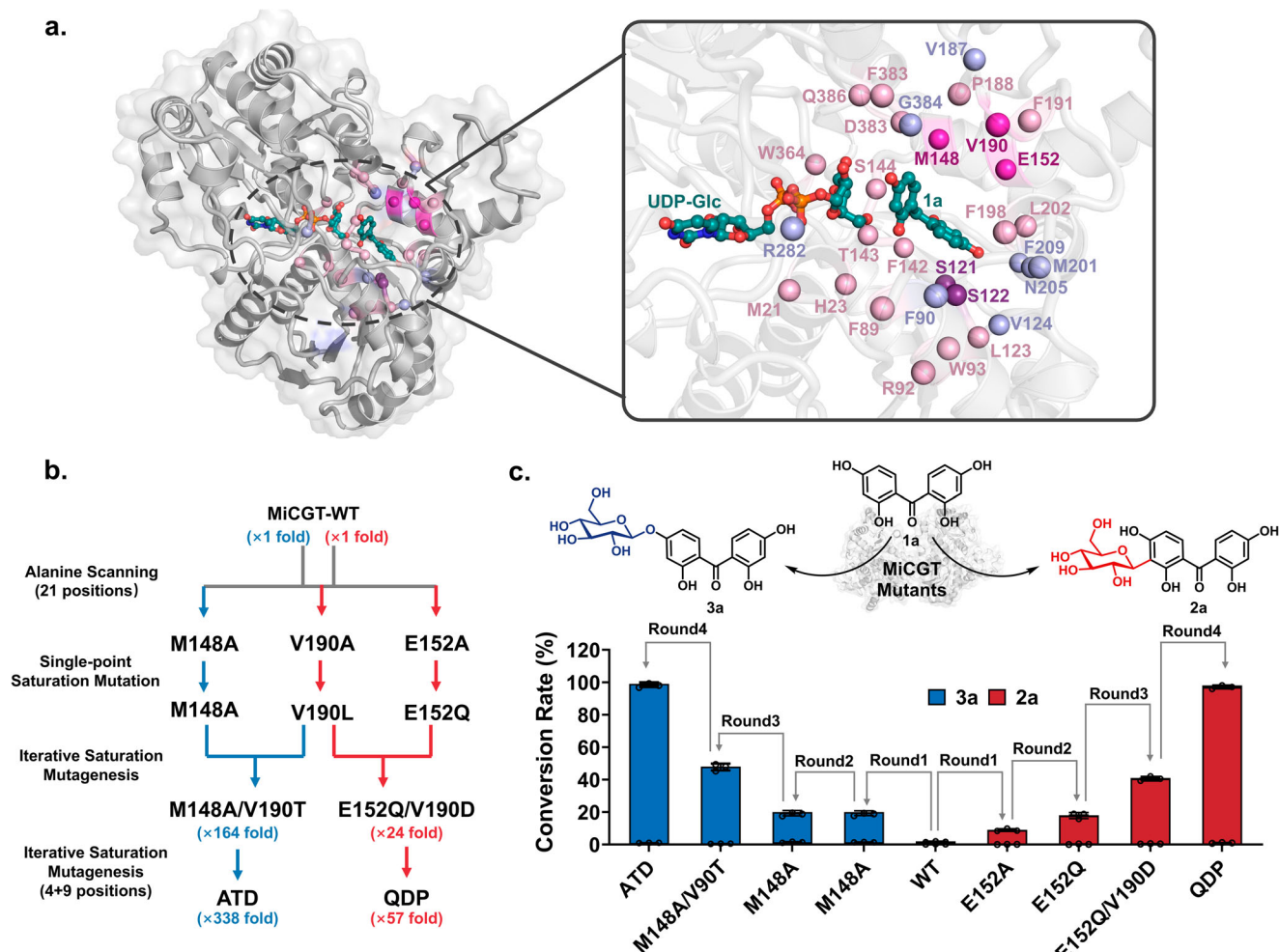
Fig. 2 | Chemoselective glycosylation catalyzed by engineered MiCGT.
a Docking of substrate bis(2,4-dihydroxyphenyl)-methanone (**1a**) and UDP-Glc at the active site of MiCGT-WT (PDB: 7VA8); **b** Design and evolution workflow used to generate mutants MiCGT-QDP (E152Q/V190D/S122P) and MiCGT-ATD (M148A/ V190T/S121D); **c** Directed evolution campaign for chemoselective glycosylation mutants. Conversion rates represent mean ± SD of three independent replicates (*n* = 3). Source data are provided as a Source Data file.

WT (PDB: 7VA8). Based on our knowledge of the MiCGT, we used the alanine scanning method[36,37] in order to identify mutational "hotspots" toward substrate **1a**[28]. Accordingly, the established alanine mutant libraries involved 21 sites (M21, H23, F89, R92, W93, S122, L123, F142, T143, S144, M148, E152, P188, V190, F191, F198, L202, W364, F383, D385, Q386) which were now re-screened with higher coverage of protein sequence space (Figs. 2a and S4).

Mutants were screened by performing biotransformation in 96-deepwell plates and products were determined by high-performance liquid chromatography (HPLC). Ultimately, six positive mutants (S122A, E152A, V190A, F191A, W364A, D385A) with improved *C*-glycosylation activities were identified. Notably, mutant E152A (previously identified) and new mutant V190A exhibited the most significant increase in *C*-glycosylation activity, which we designate as mutation "hot spots". E152, previously identified as crucial for MiCGT bis-*C*-glycosylation[27], also strongly correlates with enhanced *C*-glycosylation activity. Surprisingly, mutant M148A was observed to exhibit a remarkable increase in *O*-glycosylation activity, highlighting M148 as the key site for *O*-glycosylation (Fig. 2b and c).

To obtain the best templates for further evolution toward both *C*- and *O*-glycosylation, residues E152, V190, and M148 were selected for single-point saturation mutation (SSM)[25,38,39]. After screening using HPLC and by comparing the peak area of the corresponding products (concentration of **1a**: 0.4 mM), we found mutants E152Q and V190L to

show significantly improved *C*-glycosylation activity (22.6-folds and 12.6-folds versus WT, respectively), while mutant M148A remained as the best mutant with the highest *O*-glycosylation activity (97.6-folds versus WT). To test the synergistic effect of these three hot residues and to minimize the library size, iterative saturation mutagenesis (ISM)[25,31,37–40] was applied based on E152Q, V190L and M148A as template to screen *C*- and *O*-glycosylation enhancement mutants.

Since V190 is uniquely located at both the active pocket and the entrance of the channel (Fig. 2a), we believe that this site plays a direct role in catalytic activity, independent of the selectivity characteristics of the template enzyme. Consequently, it was logical to perform mutagenesis in combination this site with both E152Q and M148A. Thus, three NNK-based libraries encoding all 20 canonical amino acids (E152Q-V190NNK, V190L-E152NNK, and M148A-V190NNK) were constructed and screened in round 3. Double mutant, E152Q/V190D, proved to favor *C*-glycosylation activity (24-fold versus WT), and the other double mutant, M148A/V190T, was shown to prefer *O*-glycosylation activity (164-folds versus WT) enhancement mutants (concentration of **1a**: 3 mM) (Fig. 2b and c).

To further improve the glycosylation activity of the two mutants, the remaining possible sites (S122, F191, W364 and D385) and additional sites within 6 Å around substrate **1a** (F90, S121, V124, V187, M201, M205, F209, R282 and G384) were selected for ISM. After screening 2496 mutants (concentration of **1a**: 3 mM), two triple mutants QDP

(E152Q/V190D/S122P) and ATD (M148A/V190T/S121D) with 57-fold *C*-glycosylation increase and 338-fold *O*-glycosylation enhancement relative to WT were obtained, respectively (Fig. 2b and c). As these mutants present excellent chemoselectivity and activity toward substrate **1a** following the optimization of reaction conditions (Figs. S5 and S6), they served as the final mutants for conducting the chemoselective mechanism investigation.

In order to examine whether there are better evolved pathways producing the final mutants and to identify key mutations related to chemoselectivity, complete deconvolutions[41] were performed (Fig. S7). Interestingly, both pathways that we applied turned out to be the most optimized routes to the final mutants. Comparing the activity of six single mutants toward *C*- and *O*-glycosyl products, we conclude that E152Q and M148A are the most pivotal mutations for directing the chemoselectivity.

To gain insight into the mechanism of changed activity of mutants toward substrate **1a**, steady state kinetic characterization of MiCGT-WT, MiCGT-QDP and MiCGT-ATD was performed. As shown in Figs. S8–S10 and Table S5, both mutants present higher k_{cat} than WT (WT: 0.509 ± 0.088 min^{-1}, QDP: 9.473 ± 0.103 min^{-1} and ATD: 30.686 ± 0.942 min^{-1}), indicating that increased catalytic efficiency is an important factor contributing to the enhanced activity. Combining the analysis of K_m values, we observed that QDP (0.036 ± 0.003 mM) exhibits a significantly tighter affinity for substrate **1a**, whereas ATD (0.249 ± 0.030 mM) shows a similar affinity compared to WT (0.430 ± 0.155 mM). This suggests that the K_m value contributes more to the increased activity of QDP. We also performed kinetic tests on mutants involved in the evolutionary pathway of both QDP (E152Q and E152Q/V190D) and ATD (M148A and M148A/V190T) (Figs. S8–S10 and Table S5). The results revealed a clear stepwise increase in k_{cat}-values from WT to the final mutant in both pathways. However, the changes in K_m value differed significantly between the two evolutionary routes. In the QDP pathway, the K_m value for E152Q (0.176 ± 0.025 mM) and E152Q/V190D (0.189 ± 0.021 mM) were approximately 40% of that of the WT, and did not show a significant difference between them. However, with the addition of S122P mutation, the K_m value dropped dramatically to one-tenth of the WT value. Distinctly, in the ATD pathway, the K_m values (M148A: 0.269 ± 0.049 mM, M148A/V190T: 0.317 ± 0.1 mM) remained essentially unchanged compared to ATD value. These values are approximately 60% of the WT value.

Since the turnover number (TON) is a crucial parameter for evaluating the total catalytic capacity, we also ascertained this value at 30 °C and 50 °C considering the optimized reaction conditions and the long-term stability of the enzyme. The results indicate that both mutants present very high TON-values (QDP: 10375 and ATD: 9313) at 30 °C, demonstrating significantly better synthetic potential than the WT. However, the TON-values of both mutans decreased significantly at 50 °C (QDP: 4611, ATD: 1472), though they still remained higher than the WT, indicating lower long-term stability at higher temperatures (Table S6).

Structure resolution and computational analysis

In the aim of gaining an understanding of the structural basis for changed chemoselectivity and activity, we crystallized both MiCGT-QDP (PDB: 8XFH) and MiCGT-ATD (PDB: 8XFW) in the presence of UDP and substrate **1a** (Table S7), and solved the structures using MiCGT-WT (PDB: 7VA8) as the model. Regrettably, we detected UDP in both structures, but the poor electron density in the substrate binding pockets prevented us from definitively identifying substrate **1a**. This suggests that substrate **1a** may bind transiently or loosely during the reaction process. Structure overlap of MiCGT-QDP and MiCGT-ATD with MiCGT-WT reveals a backbone root mean square deviation (RMSD) of less than 0.45 Å, suggesting minimal conformational change is induced by the mutations. Except for those mutated, almost all the side-chains of residues lining the substrate binding pocket could

be well superimposed, including the conserved H23, which forms catalytic dyad with D120 and was reported to initiate the glycosylation reaction through activating deprotonation and facilitating nucleophilic attack at C-1′ of the sugar[19,20].

In MiCGT-ATD, the cavity created by the M148A mutation is not occupied by neighboring residues, resulting in a larger substrate-binding pocket. Additionally, the S121D mutation at the bottom of the pocket pushes the side-chain of H23 slightly away from D120 (Fig S11a). This structural difference contrasts with MiCGT-VFAH, a quadruple mutant of MiCGT, which enables strict 3-*O* glycosylation selectivity and a 120-fold activity enhancement toward the quercetin[28]. In MiCGT-VFAH, the binding pocket is also enlarged, but primarily due to the W93V mutation (Fig. S11b). Given the differences in *O*-glycosylation position between **1a** and quercetin, we believe that the *O*-glycosylation activity is determined by both the substrate structure and the active site configuration of the MiCGT mutant.

In MiCGT-QDP, S122P and E151Q did not cause significant shifts in the surrounding residues, suggesting that they may affect activity by subtly tuning the local environment. We compared the active site of MiCGT-QDP with that of AaCGT[42], a CGT isolated from *A. asphodeloides* that catalyzes the *C*-glycosylation of maclurin, a **1a** analogue (Fig. S11c). In MiCGT-QDP, S121 and P122 at the bottom of the pocket are replaced by I122 and M123 in AaCGT, E152 is replaced by C153, and W93 next to the H23 is substituted by F94. Since a single S122P mutation significantly increases the activity of the MiCGT-QD double-mutant, the existence of a similar hydrophobic M123 in AaCGT indicates the critical role of this position in *C*-glycosylation.

Unlike these mutated residues buried deep in the structure, V190 is located on the entrance of the pocket and is replaced by a hydrophilic residue in both mutants (Fig. 3). To understand the structural impacts of this mutation further, we carried out replica exchange with solute tempering (REST2) simulations for the surrounding "hotspots" (M21, H23, F89, R92, W93, S122, L123, F142, T143, S144, M148, E152, P188, V190, F191, F198, L202, W364, F383, D385, Q386). These simulations revealed that V190 typically engages in hydrophobic interactions with F383 and F198, as indicated by their proximity (Fig. S12). Conversely, the T190 and D190 mutations disrupt these interactions, resulting in a greater distance and broader distribution with F383 and F198, leading to a more flexible and open pocket entrance.

Because 1a was not observed in the structures, we therefore performed molecular docking in conjunction with molecular dynamics (MD) simulations to uncover the potential interactions between **1a** and the residues (Table S8). To more accurately determine the most realistic binding conformations, ensembles of binding pocket conformations generated from REST2 simulations were utilized for molecular docking. We identified forty-two distinct binding poses (14 for MiCGT-WT, 15 for MiCGT-ATD, and 13 for MiCGT-QDP), which were further validated through MD simulations to confirm the predicted binding modes (Figs. S13–S18). The stable conformations were then analyzed in depth to determine their alignment with experimental observations and to hypothesize the possible reaction mechanisms. Additionally, long time MD simulations were applied to refine these binding poses further (Fig. S19). The binding free energies of substrate **1a** to the enzymes were estimated using MM/GBSA calculations from normal MD simulations (Table S8).

In WT, the *ortho*-phenolic hydroxyl group of the 2,4-dihydroxyphenyl moiety close to UDP-Glc is buried deep in the pocket. In contrast, in QDP and ATD, it is directed towards residue H23, positioning the substrate towards the 1″-C of UDP-Glc, ready for the formation of *C*- or *O*-glycosylation product. This suggests that in both ATD and QDP, H23 may function as an assistant base for deprotonation. Further MM/GBSA energy decomposition analysis revealed that H23 positively contributes to stabilize the substrate (Table S9). To understand the function of H23 in ATD and QDP, we replaced H23 with A23 via site-directed mutagenesis. Testing the activity of the H23A
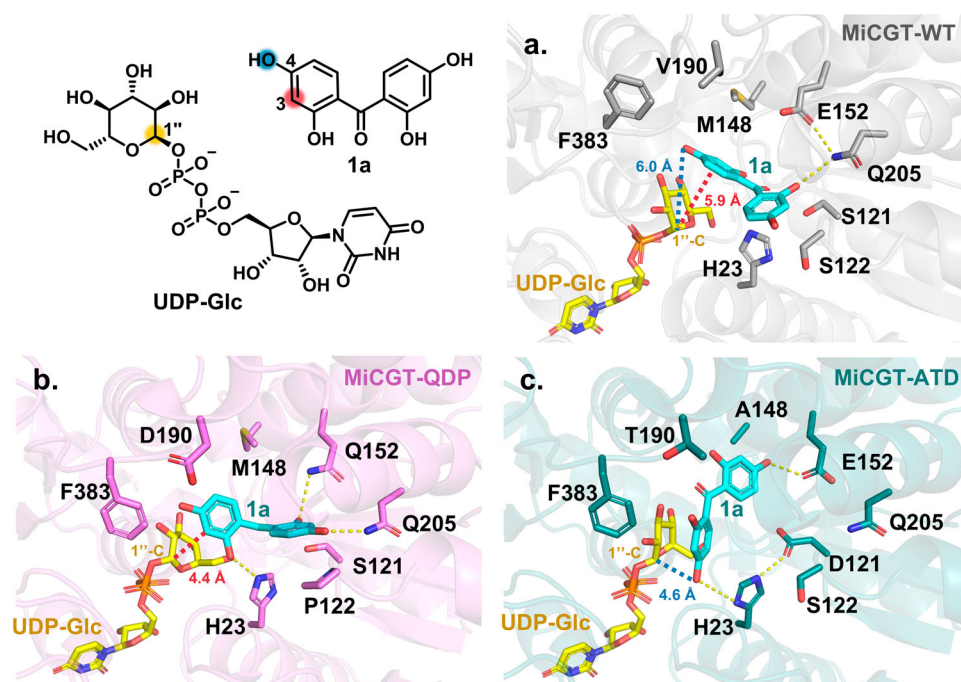
Fig. 3 | Molecular mechanism study of MiCGT-WT, MiCGT-QDP and MiCGT-ATD. Modeling of bis(2,4-dihydroxyphenyl)-methanone (**1a**) (cyan) and UDP-Glc (yellow) at the active site of MiCGT-WT (PDB: 7VA8) (**a**), MiCGT-QDP (PDB: 8XFH) (**b**) and MiCGT-ATD (PDB: 8XFW) (**c**). Hydrogen bonds are shown as yellow dash lines. The red and blue dash lines represent the distance of the acceptor reactive site to the anomeric carbon of the sugar donor (blue: d_{C-O}, red: d_{C-C}).

mutant showed that this mutation in MiCGT-ATD and MiCGT-QDP (Fig. S20) led to a more drastic decrease in activity compared to MiCGT-WT (Fig. S4), though it did not completely abolish glycosylation activity. Together with the analysis of distance between the O atom on the hydroxyl group of **1a** and His23 (Fig. S21), our current findings suggest that in MiCGT-QDP and MiCGT-ATD, H23 may play a dual role in stabilizing the acceptor at the active site[26,27] and assisting in the deprotonation of acceptor.

Interestingly, the results revealed a distinctly different binding pattern of substrate **1a** in MiCGT-WT and MiCGT-QDP compared to that of MiCGT-ATD (Fig. 3). In the active sites of WT and QDP, characterized inter alia by a bulky residue M148, substrate **1a** is accommodated in "horizontal" conformations, with the main difference appearing in the orientation of the two 2,4-dihydroxyphenyl moieties. On the other moiety, substrate **1a** forms a single hydrogen bond interaction with Q201 in WT, but in QDP, it interacts simultaneously with Q201 and the mutated Q152. Given that a single E152Q mutation significantly enhances *C*-glycosylation activity in MiCGT, we inferred that the substitution of E152 with a glutamine alters the electrostatic charge of the residue from negative to neutral, thereby changing the hydrogen bond interaction patterns. Additionally, mutation S122P is engaged in hydrophobic interactions with the benzene group of **1a** in QDP to further stabilize the conformation (Fig. 3a and b).

Conversely, we found substrate **1a** to exhibit a "vertical" conformation in ATD, in which the remote hydroxybenzene extends into a hydrophobic pocket generated by the M148A mutation, and forms a hydrogen bond with residue E152. The postulation of a "vertical" conformation of substrate **1a** is also supported by the S121D mutation, which occupies part of the space that accommodates substrate **1a** in QDP. Consequently, the 4-O position of substrate **1a** approaches the 1"-C of UDP-Glc for *O*-glycosylation. Furthermore, S121D interacts with residue H23 and may increase its nucleophilicity (Fig. 3c).

We further explored the binding modes for each round of mutations in QDP and ATD through MD simulations and MM/GBSA calculations. In the MD simulations, substrate **1a** remained stable in the binding pockets (Fig. S22). Single mutations, such as MiCGT-M148A

and MiCGT-E152Q, positioned substrate **1a** closer to 1"-C of UDP-Glc compared to MiCGT-WT. However, the average distances to 1"-C of UDP-Glc are still greater than those observed in double mutations, such as MiCGT-E152Q, MiCGT-M148A/V190T, MiCGT-E152Q/V190D or triple mutants MiCGT-QDP or MiCGT-ATD. This suggests that the conformation for *C*-glycosylation and *O*-glycosylation is likely driven by the cooperative interactions among the residues. Furthermore, the binding free energy of substrate **1a** with MiCGT-QDP (-45.7 ± 4.0 kcal/mol) was significantly lower (indicating stronger binding) than with other mutations, such as MiCGT-E152Q (-36.8 ± 5.8 kcal/mol) and MiCGT-E152Q/V190D (-25.8 ± 4.0 kcal/mol). In contrast, the binding free energy of substrate **1a** to MiCGT-ATD (-39.05 ± 3.74 kcal/mol) was only slightly improved compared to MiCGT-M148A (-31.6 ± 3.1 kcal/mol) and MiCGT-M148A/V190T (-34.5 ± 3.3 kcal/mol). These results align with kinetics data, verifying the precision of our calculation model.

Overall, the crystal structures and computational simulations rationalize the mechanisms of *C*-glycosylation and *O*-glycosylation, which primarily arise from distinctive substrate binding modes attributed to the collective effect resulting from the mutations.

Mechanism-guided design of UGTs

Now that the mechanism of chemoselectivity in the model UGT has been clarified, it is essential to confirm whether the relationship between sequence and function is unique and to illuminate how to tune the chemoselectivity against UGTs beyond MiCGT. Therefore, we selected 70 reported UGTs with *C*-glycosylation activity for sequence homology analysis, focusing on sites corresponding to mutation residues in the MiCGT mutants (both QDP and ATD) (Table S10). Surprisingly, sites corresponding to M148 in MiCGT are relatively conserved, and both are occupied by a larger amino acid residue, suggesting that this site is a key residue ubiquitously for controlling chemoselectivty in UGTs (Fig. S23a and b). Other sites including S121, S122, E152, and V190 are irregular, possibly functioning to stabilize the binding conformation favoring the formation of *C*-glycosides.
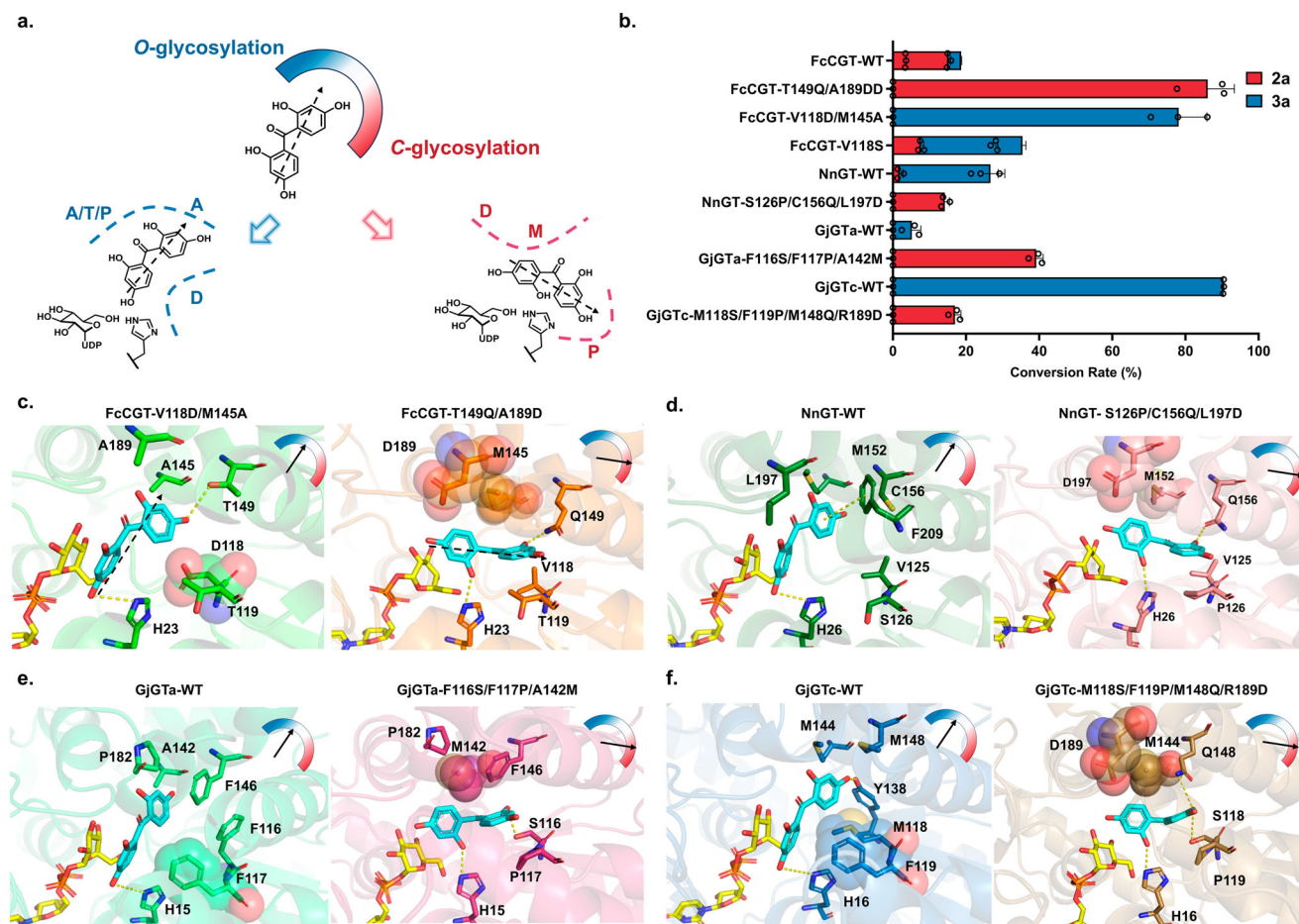
Fig. 4 | Mechanism-guided design of UGTs. a Altering substrate orientation to modulate the chemoselectivity of UGTs for glycosylation; **b** Activity tests of other UGTs variants; **c–f** Designed binding pose of the substrate in the mutated UGTs. Biotransformation was performed using 0.2 mM **1a**, 0.4 mM UDP-Glc, 10 μM of pure enzyme; Reactions were made up to 200 μL with 6% (v/v) DMSO and 100 mM Kpi (K_2HPO_4-KH_2PO_4) buffer (pH 8.0), incubated at 30 °C for 24 h. Conversion rates represent the mean ± SD of three independent replicates ($n = 3$). WT: wild type. Source data are provided as a Source Data file.

It can be inferred from the above results, to trigger the chemoselectivity switch, the initial step should involve determining the precise substrate-binding conformation based on the catalytic properties. Next, by referring to the substrate-binding modes of the two model enzymes, we can visually adjust the surrounding hotspot residues to create the desired mutant with the intended chemoselectivity (Fig. 4a).

As a proof-of-concept, we selected four UGTs including two with high homology but exhibiting different chemoselectivity (i.e, FcCGT, NnGT), and two with low homology (i.e, GjGTa, GjGTc) (Figs. S23c, S24, and Table S11). Utilizing the AlphaFold2 algorithm[43], we predicted the respective protein structures and modeled the binding conformations, followed by the poses similar to MiCGT. We attempted to manipulate the binding orientation of substrates within the active sites of these enzymes by introducing mutations that alter the size of residues in positions analogous to those in MiCGT. The binding modes for the mutations were modeled, and the resulting binding conformations were submitted to 200 ns of MD simulations for validation and MM/GBSA calculations. According to the RMSD data and binding free energies, when bound in a similar mode to that in MiCGT mutants, substrate **1a** remains essentially stable in the binding pockets (Fig. S25 and Table S12).

The predicted structures of FcCGT and NnGT indicate that they have similar substrate binding pockets and substrate channel entries. The bulky side chain at the M145 position of FcCGT may hinder the access of the receptor substrate to the active site pocket in the way of *O*-GTs channel. At the same time, the change of T149 position of FcCGT

may lead to the formation of a more spacious channel of *C*-GTs to enhance the activity of *C*-glycosylation (Fig. S26a). For NnGT, the *C*-GT channel is not only regulated by C156, but also requires mutation S126 to provide a more rigid channel scaffold (Fig. S26b). To check the precision of our theoretical predictions, we designed three FcCGT mutants for *C*-glycosyl enhancement (FcCGT-T149Q/A189D), *O*-glycosyl switching (FcCGT-V118D/M145A) and promiscuous chemoselectivity (FcCGT-V118S). We then tested their glycosylation catalytic activity. The three designed mutants perfectly met our design expectation; FcCGT-T149Q/A189D, FcCGT-V118D/M145A and FcCGT-V118S were found to achieve precise regulation of *C*-, *O*- and promiscuous selective glycosylation, respectively (Fig. 4b and c).

Considering that NnGT, originally an *O*-GT, displayed limited *C*-glycosylated activity, we proceeded to construct a triple mutant, NnGT- S126P/C156Q/L197D, guided by the sequence features of QDP, with the goal of converting it into a *C*-GT. The results validated the rationality of our design as the mutant demonstrated impeccable *C*-glycosylation selectivity, yielding **2a** (Fig. 4b and d).

The above results confirm our notion that uncovering the enzyme mechanism can precisely guide the rational design of chemoselectivity-switching for UGTs sharing high homology. To assess the universality of this model, two challenging *O*-GTs, GjGTc and GjGTa with low homology (25.97% and 28.69%, respectively) from *Gardenia jasminoides* were selected for rational enzyme design.

We first predicted their structures using AlphaFold2. A comparison of their predicted structures with the crystal structure of MiCGT-

ATD revealed that both have the spacious sugar-acceptor binding pocket, and a narrower receptor substrate entry channel. Moreover, the substrate binding is in a 'vertical' manner, which is required for *O*-glycosylation selectivity. To switch to *C*-glycosylation selectivity, adjusting the binding conformation to the 'horizontal' style manner is necessary (Fig. S26c and S26d).

Therefore, by comparing their substrate binding conformation with that of QDP, residues that differ significantly can be easily discovered. Residues F116, F117, and A142 in GjGTa, and M118, F119, M148 and R189 show a synergy effect to direct the substrate to 'stand up'. To 'lay down' the substrate, these sites can be mutated to the same residues corresponding to those in the QDP. Thus, a triple mutant (GjGTa-F116S/F117P/A142M) and a quadruple mutant (GjGTc-M118S/F119P/M148Q/R189D) were constructed to test their activities toward substrate **1a**. In line with expectations, both mutants proved to exhibit very high *C*-glycosylation selectivity. Moreover, the activity of mutant GjGTa-F116S/F117P/A142M has increased about 5-fold relative to WT (Fig. 4b−f). To study the catalytic efficiencies of the newly designed mutants with target chemoselectivity in detail, steady-state kinetic tests were performed. Finally, we obtained 4 kinetic data for 4 mutants, excluding GjGTc-M118S/F119P/M148Q/R189D due to its relatively low activity. Compared to MiCGT-QDP and MiCGT-ATD, the mutants FcCGT-T149Q/A89D $(2.445 \pm 0.096 \, \text{min}^{-1})$, FcCGT-V118D/M145A $(3.194 \pm 0.131 \, \text{min}^{-1})$, NnGT-S126P/C156Q/L197D $(0.523 \pm 0.034 \, \text{min}^{-1})$ and GjGTa-F116S/F117P/A142M $(0.021 \pm 0.00007 \, \text{min}^{-1})$ exhibit significantly lower k_{cat} values (Fig. S27 and Table S13). Analyzing the K_m values, we observed that GjGTa-F116S/F117P/A142M $(0.012 \pm 0.0005 \, \text{mM})$ demonstrates a significantly higher substrate affinity for **1a**, whereas its lower k_{cat} value results in reduced catalytic efficiency. The K_m values for FcCGT-T149Q/A189D $(0.375 \pm 0.072 \, \text{mM})$ and NnGT-S126P/C156Q/L197D $(0.319 \pm 0.061 \, \text{mM})$ are only compared to that of MiCGT-ATD. In contrast, the K_m value for FcCGT-V118D/M145A $(1.064 \pm 0.151 \, \text{mM})$ is approximately four times higher than that of MiCGT-ATD. Although these mutants exhibit better selectivity than WT, their glycosylation efficiency still falls short of achieving efficient glycoside synthesis. Rationally designing UGTs to achieve simultaneous and significant improvements in both selectivity and efficiency remains our ongoing goal.

Overall, chemoselectivity of all new design UGTs are consistent with our model. Thus, the uncovering of sequence-structure-chemoselectivity relationships of MiCGT has proven to be a valuable guide for the design of more selective GTs family members.

Exploring the substrate scope of MiCGT-mutants and preparative scale reaction with whole cell catalyst

As both mutants MiCGT-QDP and MiCGT-ATD exhibit high catalytic efficiency (TON > 9000), we were keen to explore their substrate-scopes. Typically, plant *C*-GTs demonstrate specificity towards substrates that possess a core structure of acyl phloroglucinol. This core structure is characterized by the presence of three hydroxyl groups and one acyl group[44,45]. Therefore, we divided the substrate library into four types according to the number of hydroxyls or the presence of an acyl group: acyl resorcinols (substrate **1a-1s**), acyl phenols (substrate **1t-1z**), acyl phloroglucinols (substrate **1aa-1ad**), and substrates without basic acyl groups (substrate **1ae-1ah**) (Fig. 5).

The substrate scope study indicates that the number of hydroxyl groups in the benzene ring plays a crucial role in the formation of *C*-glycoside products. Acyl phloroglucinols readily generate *C*-glycosides, while no *C*-glycoside production is observed for acyl phenols. The glycosylation of acyl resorcinols by MiCGT-WT typically yields chemoselective mixed glycoside products. In contrast, MiCGT-mutants exhibit precise control over the chemoselectivity of glycosylation modification for these substrates.

Although the regulation of *C*-glycosylation modification by acyl phenols remains elusive, MiCGT-ATD demonstrates a remarkable

enhancement in catalytic activity. Additionally, it is important to note that the absence of acyl side chains in reactions catalyzed by MiCGT mutants also significantly impacts their catalytic activity despite their chemoselective regulatory effect. This is possible due to the crucial role played by acyl side chains in substrate recognition within MiCGT.

Using this convenient glycosylation approach, we successfully synthesized a series of structurally diverse acyl resorcinol glycosides (**2a-2s** and **3a-3s**) (Fig. S28). MiCGT-QDP demonstrated exceptional chemoselective *C*-glycosylation activity and excellent tolerance towards substituents of acyl side chains. Substrates with alkyl groups, **1b-1c**, those with aromatic groups, substrates **1a, 1d-1f, and 1h-1q**, and those with heteroaromatic moieties, **1r-1s** are readily accepted by the respective mutants. Electron-withdrawing groups reduce the electron density of the benzene ring, which hinders the formation of *C*-glycosides (substrate **1n**). The halogen substitution on the benzene ring is special, with inductive and conjugation effects exerting an influence. The presence of substituents at *meta* positions of the aromatic ring bearing acyl side chains lead to higher conversions (substrate **1i, 1k,** and **1 m**) than for substituents at the *para* position (substrate **1 h, 1j** and **1 l**). Unlike MiCGT-QDP, *O*-glycosylation reactions catalyzed by MiCGT-ATD tolerate both electron-withdrawing and electron-donating groups, showing similar and acceptable catalytic activity. Substrates with larger substituents at the acyl side chains of acyl resorcinols are also accepted, but the efficiency of the reaction is relatively low (substrates **1q-1r***). These data are consistent with our steady-state kinetics results, with a larger K_m value, MiCGT-ATD presenting more promiscuous substrate specificity.

To test the practical potential of the glycosylation process, we attempted preparative reactions with whole cells as the catalyst in the production of compounds **2a, 3a, 2e** and **2ad**. This proved to be successful. As shown in Fig. 6, all products were successfully produced on gram scale, indicating the applicable power of our glycosylation platform. It is worth noting that product **2ad** is a natural product named nothofagin, which has been shown to be a molecule with biological activity, downregulating NF-*κ*B translocation through blocking calcium influx[46]. Significantly, our synthetic platform runs smoothly without any exogenous UDP-Glc, nor does it require the enhancement of any relevant intracellular enzymes in the chassis. Further improvements towards a cell-factory would facilitate the scalable production of rarer glycosyl-products.

Discussion

The methodology for the rational tuning of the chemoselectivity of GTs is crucial for the establishment of versatile glycoside production platforms. Although previous attempts to regulate the selectivity of UGTs have achieved remarkable success[33,34], most of the reported works focused on engineering regioselectivity or substrate specificity, while investigations of chemoselectivity (especially for *C*- and *O*-glycosylation) regarding engineering and mechanism investigation are rare.

Here, through an efficient directed evolution campaign of MiCGT with promiscuous chemoselectivity, together with structure resolution and computational simulations, we first succeeded in clarifying the selectivity mechanism for *C*- and *O*-glycosylation, respectively. We suggest that different binding modes of receptor substrates in UGTs correspond to the production of either *C*-glycosides ('horizontal' style) or *O*-glycosides ('vertical' style). Based on this proposal, two substrate binding models were then constructed for engineering the *O*-/*C*-glycosylation switch. Four UGTs with high to low homologous similarity were selected for rational design of the chemoselectivity switch. Guided by our models, one *C*-UGT, FcCGT, has been fine-tuned to promiscuous UGT with both *C*-/*O*-glycosylation activity, enhanced *C*-UGT activity and switched it to *O*-UGT. More significantly, three *O*-UGTs were precisely engineered into *C*-UGTs by adjusting the substrate binding conformation through mutations at sites decoded from directed evolution.
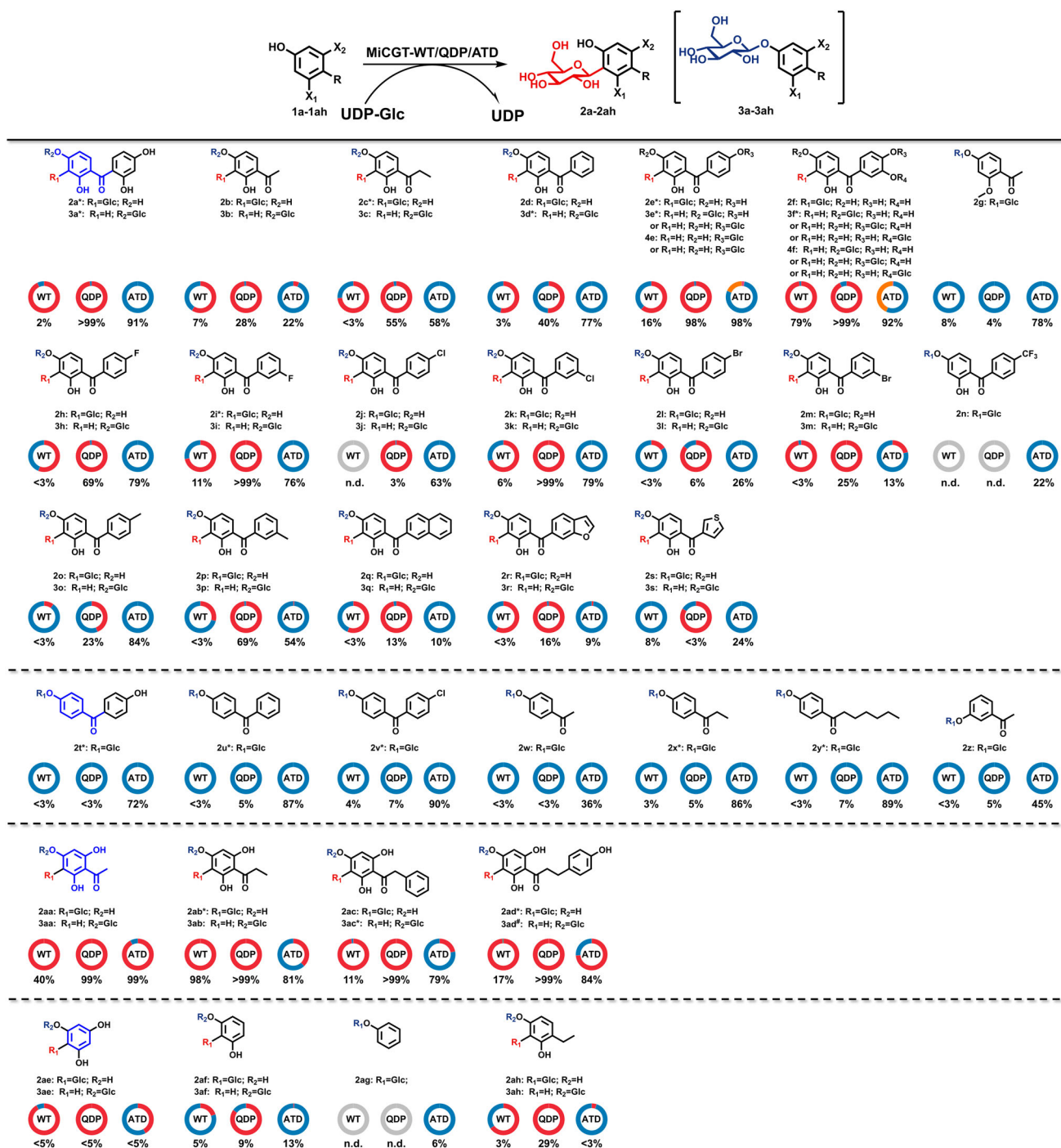
Fig. 5 | Substrate scope of glycosylation reaction. Biotransformation was carried out using 3 mM glycosyl receptor substrate, 4 mM Uridine 5′-diphosphoglucose disodium salt (UDP-Glc), 10 μM of pure enzyme. Reaction was made up to 100 μL with 6% (v/v) DMSO and 100 mM Kpi buffer corrected to pH 8.0. The reaction time of substrate **1h**-**1s** is 24 h, and the reaction time of other substrates is 2 h. Conversions given are determined by HPLC analysis. *=Glycosylated products were prepared and confirmed by MS, and ^{1}H and ^{13}C NMR spectroscopy. #=Glycosylated products were prepared and confirmed by MS and standard. n.d. = not detected. Conversion rates represent mean of three independent replicates ($n = 3$). Source data are provided as a Source Data file.

Additionally, two mutants that we created exhibit high chemo- and regioselectivity as well as broad substrate-scope, thus facilitating the establishment of a powerful glycosylation platform. With this platform, we prepared four glycosides (one O-glycosides and three C-glycosides) on a gram scale using economical whole-cell catalysts. The construction of such a highly efficient glycosylation platform breaks the bottleneck of limited sources and greatly stimulates the development and application of carbohydrate chemistry.

In order to more effectively guide the customization of glycosyl-transferases for use in building efficient glycosylation platforms, we have summarized a general four-step research strategy (Fig. S29) based on this work: *1st*, An enzyme template with promiscuous selectivity is preferred; *2nd*, Given that enzymes possess high substrate specificity, the modification of the selectivity should be based on a selected substrate skeleton; *3rd*, Similar to transforming C-GTs into O-GTs which is more attainable, the retrostrategy is preferred and the
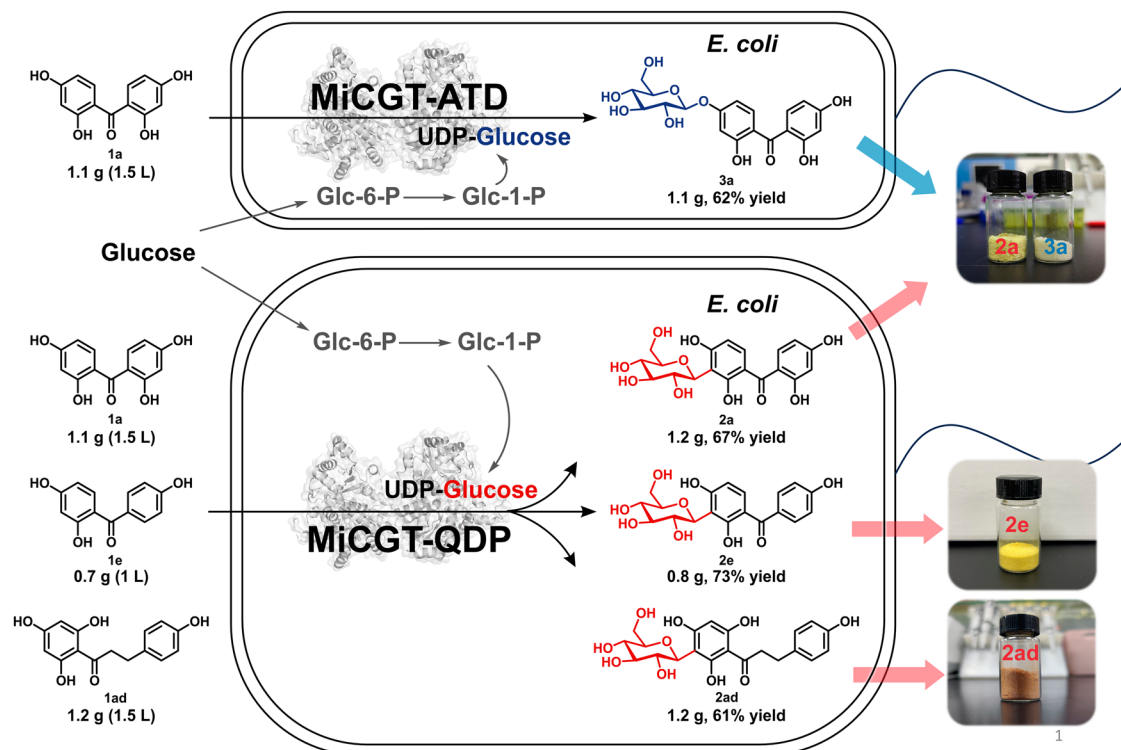
Fig. 6 | Synthesis of glycosides by whole cell of *E. coli* harboring MiCGT-mutants on gram scale. The *E. coli* (MiCGT-mutants) was resuspended by 100 mM Kpi buffer (pH 8.0) containing 2% (w/v) *D*-glucose and 6% (v/v) DMSO, and the cell density was adjusted to an $OD_{600\,nm} = 40$. Substrates were added to the reaction mixture at the final concentration of about 3 mM in a total volume of 1.5 L (Substrate **1a** and **1ad**) or 1 L (Substrate **1e**), and the bioconversion reactions were performed at 30 °C, 400 rpm for 24 h. (Glc-6-P: Glucose 6-Phosphate; Glc-1-P: Glucose 1-Phosphate).

directed evolution-based mechanism resolution can be performed in more detail. In essence, model reactions can produce many products, but the dominant product is one that is rarely found in nature. Directed evolution must encompass both directions: selectivity enhancement and selectivity switching. In this way, mechanistic resolution can cover hotspots that are associated with the initiation and enhancement of all selectivity; *4th*, Validation of the proposed 3rd step by engineering enzymes beyond the model enzyme The validation of our strategy and the elucidation of the origin of *C*- and *O*-glycosylation selectivity offer a streamlined protocol for future engineering studies of UGTs. When dealing with specific cases, based on known data, for different substrates, there are two pathways for engineering of UGTs. Firstly, as we aim to modify UGTs toward different substrates, the four-step strategy can be applied. As precise models have been constructed, optimal templates for seeking the reasonable binding conformation of any WT enzyme toward target substrate were predicted and tested. Subsequently, sites that are not suitable for the deserved selectivity can be pinpointed by comparing with the correspond binding model. Consequently, a given mutant with designed selectivity can be created via site-directed mutagenesis. Secondly, when encountering the same substrate and needing to identify a new enzyme, having the precise model enzyme on hand makes it convenient to search for enzymes with improved industrial properties, such as thermostability. Subsequently, rational design can be directly implemented based on our proposed approach.

In summary, together with the strong power of directed evolution, protein structure resolution, and computational simulations, we have unveiled the selectivity mysteries of *C*- and *O*-glycosylation in plant-derived UGTs and constructed a highly efficient *C*-/*O*-glycosylation platform. Moreover, we established a practical paradigm for the future customization of UGTs, enabling the construction of precise glycosylation platforms to meet various application needs.

Methods
All chemical reagents were purchased from Aladdin, Ark Pharm, Macklin Biochemical, 9ding Chemistry, Bide Pharm tech, Shanghai YuanYe Biotechnology Co., or Sangon Biotech (Shanghai, China) unless otherwise stated. All standard of glycosylated products were obtained from Chengdu Chroma -Biotechnology Co., Ltd. The KOD One™ PCR master Mix polymerase was obtained from Toyobo Idears & Chemistry. *Dpn*I was purchased from Thermo Fisher. The ClonExpress MultiS One-Step Cloning Kit is available from Vazyme Biotech. Primer synthesis and DNA sequencing were conducted at Sangon Biotech Company (Shanghai, China). Shimadzu Nexis LC 20AT or HPLC-MS 8050 equipped with an electrospray ionization (ESI) source was used for the measurement of substrates and products. The glycosylated products were isolated and purified by semi-preparative HPLC on an LC-8A instrument (Shimadzu). Products were characterized by ^{1}H NMR and ^{13}C NMR in DMSO-d6 on AV2 500 MHz Bruker spectrometer. Chemical shifts (δ) were given in ppm and coupling constants (J) were given in hertz (Hz).

Heterogonous expression and purification of UGTs
MiCGT, NnGT, GjGTa and GjGTc DNA sequence were synthesized by GENEWIZ (China), and then inserted into the *Nde*I-*Xho*I site of pET28a. The recombinant plasmid pET28a-UGTs were transformed into *E. coli* ROSETTA (DE3) for heterologous expression. *E. coli* cells were cultured (5 mL) overnight in Luria-Bertani (LB) medium containing 50 μg/mL kanamycin and 35 μg/mL chloramphenicol at 37 °C with shaking (220 rpm). Subsequently, 1% seeding cultures were transferred into 500 mL TB-medium containing 50 μg/mL of kanamycin and 35 μg/mL of chloramphenicol and grown at 37 °C and 220 rpm in 1 L shake flasks. After the $OD_{600\,nm}$ reached 0.8–1.0, the expression of recombinant UGTs were induced with 5 g/L *D*-Lactose monohydrate at 16 °C, 160 rpm for 18 h. The cells were harvested by centrifugation at 6,438 g

for 2 min at 4 °C and then resuspended in binding buffer (100 mM Kpi buffer, 10% glycerol, pH 8.0). After cell rupture by sonication in ice bath, the cell debris was removed by centrifugation at 15,480 g and 4 °C for 1 h. The soluble fraction was passed through a 0.45 μm syringe filter. The supernatant was applied to an ÄKTA avant 25 system equipped with a 5-mL HisTrap™ HP column following the manufacturer's instructions, using a linear gradient of 10–500 mM imidazole in 100 mM Kpi buffer, 10% glycerol, pH 8.0. For crystallization, the His-tag was cleaved, and the proteins were further purified by a Superdex 200 size exclusion column (GE Healthcare) equilibrated by a buffer containing 25 mM Tris pH 8.0, 200 mM NaCl and 5 mM DTT.

Structure determination of of MiCGT-ATD and MiCGT-QDP

Crystals of MiCGT-ATD and MiCGT-QDP were obtained at 4 °C and 20 °C, respectively, in the presence of UDP in a buffer containing 0.1 M Bis-Tris pH 6.5, 16% (v/v) PEG3350, 200 mM calcium acetate and 10% (v/v) glycerol. The crystals were soaked in the crystallization solution supplemented with 10-15% (v/v) glycerol for a short time before flash-frozen in liquid nitrogen. All data sets were collected on beam line BL02U1 at Shanghai Synchrotron Radiation Facility. The Data were indexed, integrated and scaled using HKL3000[47] package. The structures were solved by molecular replacement using PHASER[48] and the structure of WT MiCGT (PDB code: 7VA8) as the searching model. Iterative manual rebuilding and refinement of the model were performed with the use of COOT[49] and PHENIX[50], respectively.

Preparation of *E. coli* whole cell catalyst harboring UGTs mutants

After protein expression, the cells were harvested by centrifugation at 6438 g for 2 min at 4 °C and resuspended in binding buffer (100 mM Kpi buffer, pH 8.0) at $OD_{600 nm} = 40$. The wet cells were stored at −80 °C until further use.

Construction and screening of MiCGT mutant libraries

All strains, plasmids and primers in this work are shown in Tables S1 and S2. The PCR (Table S3) assay was conducted with KOD One™ PCR master Mix polymerase. After verifying the mutant sequence, the recombinant plasmid was transformed into *E. coli* ROSETTA (DE3) for heterologous expression as described above.

Alanine scans are based on mutant libraries that have been established in the laboratory[28]. Iterative saturation mutagenesis was performed using primers containing NNK codons at specific sites. To scan enough mutants, 92 colonies selected at one site were examined. The colonies were cultured in 96 deep-well plates with 0.3 mL LB medium containing 50 μg/mL kanamycin and 35 μg/mL chloramphenicol at 37 °C overnight. And then 0.7 mL TB medium containing 50 μg/mL kanamycin and 35 μg/mL chloramphenicol, 5 g/L *a*-Lactose monohydrate was added, further incubation was performed for 18 h at 18 °C. The cells were harvested and suspended in 0.2 mL reaction buffer (0.4/3 mM substrate **1a**, 50 mM Tris-HCl buffer, pH 8.0, 2% glucose) and kept shaking at 30 °C for 24 h. Subsequently, the reactions were quenched by the addition of 0.2 mL methanol. Relative activity was measured by whole-cells-catalyzed biotransformation in a total volume of 2 mL reaction buffer (3 mM substrate **1a**, 100 mM Kpi buffer, pH 8.0, 2% glucose) at 30 °C for 24 h, and the wild type served as a control.

Samples were centrifuged at 12,000 g for 15 min and analyzed by HPLC/UV on an Aglient SB-C18 column (4.6 × 250 mm, 5 μm) at a flow rate of 1 mL/ min at 30 °C. The gradient programs were used for the analyzes of the reactions (Table S4).

Reaction condition optimization

All the optimization reactions were conducted in triplicate on an analytical scale (200 μL) with 10 μM enzyme, using UDP-Glc as the donor and substrate **1a** as the acceptor. To determine the optimal reaction temperature, reactions were conducted at various temperatures (20 °C, 30 °C, 40 °C, 50 °C, 60 °C, 70 °C, 80 °C, 90 °C) for 0.5 hours or with a pre-incubation period of 0.5 hours followed by a subsequent 0.5 hours reaction at 30 °C. Although the reaction activity peaked at 60°C, the enzyme's stability was significantly compromised over prolonged periods. Additionally, we previously observed that some acceptor substrates is unstable at high temperatures[28]. Consequently, all subsequent condition optimizations were conducted at 30°C. (Fig. S5d–i). To study the optimal pH, the enzymatic reaction was performed in various reaction buffers with pH values in the range of 4.0-6.0 (citric acid-sodium citrate buffer), 6.0–8.0 (phosphate buffer), 7.0-9.0 (Tris-HCl buffer) and 9.0-11.0 (glycine-NaOH buffer) at 30 °C (Fig. S5g-S5l). Due to the optimal reaction buffer being phosphate buffer, and the incompatibility of metal ions in this buffer, further optimization of metal ion addition in the reaction is not pursued. To study the effect of the co-solvent dimethyl sulfoxide (DMSO), different concentrations of DMSO (2%, 4%, 6%, 8% and 10% v/v) were added to the enzyme-catalyzed reaction optimized conditions (Fig. S5m–o). To study the influence of different acceptor substrates, enzyme-catalyzed reactions were conducted at 30 °C, in 100 mM Kpi buffer (pH 8.0) containing 6% (v/v) DMSO. Various concentrations of acceptor substrates (1, 2, 3, 4, 5 mM) were added for the reaction. Subsequently, the reactions were quenched by the addition of 0.2 mL methanol. Supernatants were analyzed by analytical HPLC as described above (Table S4).

Determination of kinetic parameters

For kinetic studies of the MiCGT-WT, MiCGT-QDP and MiCGT-ATD toward substrate **1a**. Enzymatic assays containing 100 mM Kpi buffer, pH 8.0, 1/5/10 μM enzyme, 4 mM of saturated UDP-Glc, and varying concentrations (200, 400, 600, 800, 1000, 1500, 2000, 2500, 3000 and 3500 μM) of **1a** was performed at 30 °C for 5/10/30 min in a final volume of 100 μL. All the reactions were terminated by adding 100 μL methanol and centrifuged at 13,800 g for 15 min. Supernatants were analyzed by HPLC as described above. All experiments were performed in triplicate. The Michaelis–Menten constant (K_m) were determined from non-linear regression fitting, and the turnover number (k_{cat}) were calculated.

To determine the TON of the reactions of substrates **1a**, 3 mM substrate was mixed with the purified enzyme 0.1 μM in 100 mM Kpi buffer (pH 8.0). The total reaction system was 200 μL. which was incubated at 30 °C, 1200 rpm for at least 24 h. The formation of products was determined by HPLC and the TON was calculated according to the formula TON = $[P]_{formation}$ / $[E]$.

HPLC-MS based glycosyltransferase activity assay

In the in vitro activity assay, the enzyme assay reaction mixture consisted of 4 mM UDP-Glc, 3 mM receptor substrate, 6% (v/v) DMSO, and 10 μM purified enzyme in a final volume of 100 μL of 100 mM Kpi buffer (pH 8.0). The mixture was incubated 30 °C for 2 h or 24 h and the reaction were terminated by adding 300 μL methanol. The reactants were then centrifuged at 13,800 g for 15 min and the supernatants was analyzed by HPLC-ESIMS. Three parallel experiments were carried out. The gradient programs were used for the analyzes of the reactions (Table S4).

Products identification by NMR

Recombinant *E. coli* harboring pET28a-MiCGT-QDP or pET28a-MiCGT-ATD was prepared and induced as described above. After induction, *E. coli* whole cells were harvested by centrifugation and washed by 100 mM Kpi buffer (pH 8.0). The strains were then resuspended by 100 mM Kpi buffer (pH 8.0) containing 2% D-glucose and the cell density was adjusted to an $OD_{600 nm} = 40$. Acceptor substrates were added to the reaction mixture at the final concentration of 3 mM in a total volume of 50 mL (250 mL triangle bottle) and the bioconversion

reactions were kept shaking at 30 °C for 24 h. Subsequently, samples were centrifuged at 6438 g for 10 min and the supernatant was collected. The cells were washed by 20 mL ddH$_2$O and centrifuged twice. The supernatant was filtered, preliminary purified using macro-porous resin column chromatography with 80% EtOH. The solvent was evaporated under reduced pressure, the residue was dissolved in MeOH and purified by reverse-phase semi-preparative HPLC. The obtained products were solved in DMSO-d_6 and analyzed by ^{1}H NMR and ^{13}C NMR.

Construction and characterization of other UGTs mutants

The engineered mutants of other UGTs are constructed based on homologous recombination technology using the ClonExpress MultiS One Step Cloning Kit. The reaction mixture for the enzymatic assay contained 100 mM Kpi buffer (pH 8.0), 0.4 mM UDP-Glc, 0.2 mM acceptor substrate **1a**, 6% (v/v) DMSO and 10 μM purified enzymes in a final volume of 0.2 mL. The reation was carried out at 30 °C for 24 h and terminated by adding 0.2 mL methanol. The reactants were then centrifuged at 13,800 g for 15 min and the supernatants was analyzed by HPLC. Three parallel experiments were carried out ($n = 3$).

Gram scale preparation of *C*-glycosides by whole cell catalyst

The grame-scale preparations of glycosides were carried out with *E. coli* ROSETTA (DE3) whole cells harboring MiCGT-QDP at 30 °C and 400 rpm for 24 h. Subsequently, samples were centrifuged at 6,438 g for 5 min and the supernatant was collected. The cells were washed by 100 mL ddH$_2$O and centrifuged twice. The supernatant was filtered, preliminary purified using macro-porous resin column chromatography with 80% EtOH. The solvent was evaporated under reduced pressure and the residue was dissolved in methanol. Purification of **3a**, **2a** and **2e** was performed by silica gel chromatography (DCM:MeOH=10:1, and containing 0.1% acetic acid). **2ad** was purified by adding 0.5% flocculant (polyacrylamide), centrifuging the clarified centrifuge and filtering it with a 450 nm filter membrane. The obtained filtrate was further evaporated under reduced pressure to obtain **2ad** concentrate with a solid content of about 25%. Finally, the crystallization was placed at 0 °C to 4 °C, and the final **2ad** product was dried.

Computational method

The systems of MiCGT-WT, FcCGT, NnGT, GjGTa, and GjGTc and their mutants were prepared by Protein Preparation Wizard (Schrödinger, LLC, New York, NY, 2020). For MD simulations, the structures were placed in orthorhombic boxes with a buffer size of 10 Å and solvated using SPC water model, then neutralized with 0.15 M NaCl. The OPLS4 force field was used for the system. MD simulations were carried out using Desmond in NPT ensemble at 300 K and 1 atm using the Nosé-Hoover Langevin and Martyna-Tobias-Klein algorithms for thermostat and barostat, respectively. Each system was relaxed for 5 ns before 200 ns production simulations. Replica exchange with solute tempering (REST2)[51] simulations was carried out to sample the conformational ensemble of the active sites of the enzymes, as defined by the "hotspot" residues M21, H23, F89, R92, W93, D/S121, S122, L123, F142, T143, S144, A/M148, Q/E152, P188, T/D/V190, F191, F198, L202, W364, F383, D385, and Q386. Each replica was simulated for 200 ns, resulting in a total simulation time of 1600 ns per system. The exchange of configurations between neighboring replicas were attempted every 1000 steps (2 ps). The average exchange acceptance ratio for the simulations was 45.18% for MiCGT-WT, 44.71% for MiCGT-QDP, and 45.49% for MiCGT-ATD. Compound **1a** was prepared by Ligprep and molecular docking was carried out using Glide in extra precision (XP) mode with default settings. The docking box was defined by the centroid of "hotspot" residues (M21, H23, F89, R92, W93, D/S121, S122, L123, F142, T143, S144, A/M148, Q/E152, P188, T/D/V190, F191, F198, L202, W364, F383, D385, Q386). Post-docking minimization included 10 poses per ligand. The molecular mechanics

generalized Born surface area (MM/GBSA) method implemented in Prime was used to calculate the binding free energy between compound **1a** and the enzymes, using the 5000 snapshots extracted from the last 100 ns of 200 ns simulation trajectories, with the OPLS4 forcefield and VSGB solvation model.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The atomic coordinates and structure factors of MiCGT mutants have been deposited in the Protein Data Bank under the codes 8XFW for MiCGT-ATD [www.rcsb.org/structure/unreleased/8XFW] and 8XFH for MiCGT-QDP [www.rcsb.org/structure/unreleased/8XFH], respectively. Five glycosyltransferases are annotated as GT1 family members in the Carbohydrate-Active enZymes (CAZy) Database by searching their GenBank accession codes: A0A0M4KE44.1 (MiCGT), A0A224AM54.1 (FcCGT), XP_010258947.1 (NnGT), BAK55747.1 (GjGTa), and BAK55746.1 (GjGTc). The source data for the molecular dynamic simulations are available at the following address: https://zenodo.org/records/13305685. Source data are provided with this paper. All data supporting the findings of this study are available within the article and its supplementary information files. Source data are provided with this paper.

References

1. Crawford, C. J. et al. Advances in glycoside and oligosaccharide synthesis. *Chem. Soc. Rev.* **52**, 7773–7801 (2023).
2. Yang, D. et al. Production of carminic acid by metabolically engineered *Escherichia coli. J. Am. Chem. Soc.* **143**, 5364–5377 (2021).
3. Liu, C. et al. Inactivation of soybean bowman–birk inhibitor by stevioside: interaction studies and application to soymilk. *J. Agric. Food Chem.* **67**, 2255–2264 (2019).
4. Chang, C. W. et al. Automated quantification of hydroxyl reactivities: prediction of glycosylation reactions. *Angew. Chem. Int. Ed.* **60**, 12413–12423 (2021).
5. Xiao, J. et al. Dietary flavonoid aglycones and their glycosides: which show better biological significance? *Crit. Rev. Food Sci.* **57**, 1874–1905 (2017).
6. Deng, L. F. et al. Palladium catalysis enables cross-coupling-Like S$_N$2-glycosylation of phenols. *Science* **382**, 928–935 (2023).
7. Putkaradze, N. et al. Natural product *C*-glycosyltransferases-a scarcely characterised enzymatic activity with biotechnological potential. *Nat. Prod. Rep.* **38**, 432–443 (2021).
8. Yang, Y. et al. Recent advances in the chemical synthesis of *C*-glycosides. *Chem. Rev.* **117**, 12281–12356 (2017).
9. Pesciullesi, G. et al. Transfer learning enables the molecular transformer to predict regio- and stereoselective reactions on carbohydrates. *Nat. Commun.* **11**, 4874 (2020).
10. Yang, M. et al. Functional and Informatics Analysis Enables Glycosyltransferase Activity Prediction. *Nat. Chem. Biol.* **14**, 1109–1117 (2018).
11. Ross, J. et al. Higher plant glycosyltransferases. *Genome Biol.* **2**, 3004 (2001).
12. Zhang, C. S. et al. Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. *Science* **313**, 1291–1294 (2006).
13. Nidetzky, B. et al. Leloir glycosyltransferases as biocatalysts for chemical production. *ACS Catal.* **8**, 6283–6300 (2018).
14. Aguillón, A. R. et al. Synthetic strategies toward SGLT2 inhibitors. *Org. Process Res. Dev.* **22**, 467–488 (2018).
15. Karikó, K. et al. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. *Mol. Ther.* **16**, 1833–1840 (2008).

16. Pankiewicz, K. W. et al. From ribavirin to NAD analogues and back to ribavirin in search for anticancer agents. *Heterocycl. Commun.* **21**, 249–257 (2015).

17. Thibodeaux, C. J. et al. Unusual sugar biosynthesis and natural product glycodiversification. *Nature* **446**, 1008–1016 (2007).

18. Härle, J. et al. Rational design of an Aryl-*C*-glycoside catalyst from a natural product *O*-glycosyltransferase. *Chem. Biol.* **18**, 520–530 (2011).

19. Gutmann, A. et al. Switching between *O*- and *C*-glycosyltransferase through exchange of active-site motifs. *Angew. Chem. Int. Ed.* **51**, 12879–12883 (2012).

20. He, J. B. et al. Molecular and structural characterization of a promiscuous *C*-Glycosyltransferase from *Trollius Chinensis*. *Angew. Chem. Int. Ed.* **58**, 11513–11520 (2019).

21. Teze, D. et al. *O*-/*N*-/*S*-Specificity in Glycosyltransferase Catalysis: From Mechanistic Understanding to Engineering. *ACS Catal.* **11**, 1810–1815 (2021).

22. Arnold, F. H. et al. Innovation by evolution: bringing new chemistry to life (Nobel Lecture). *Angew. Chem. Int. Ed.* **58**, 14420–14426 (2019).

23. Zeymer, C. et al. Directed evolution of protein catalysts. *Annu. Rev. Biochem.* **87**, 131–157 (2018).

24. Wang, Y. et al. Directed evolution: methodologies and applications. *Chem. Rev.* **121**, 12384–12444 (2021).

25. Qu, G. et al. The crucial role of methodology development in directed evolution of selective enzymes. *Angew. Chem. Int. Ed.* **59**, 13204–13231 (2020).

26. Chen, D. et al. Probing the catalytic promiscuity of a regio- and stereospecific *C*-Glycosyltransferase from *Mangifera Indica*. *Angew. Chem. Int. Ed.* **54**, 12678–12682 (2015).

27. Chen, D. et al. Probing and engineering key residues for Bis-*C*-glycosylation and promiscuity of a *C*-glycosyltransferase. *ACS Catal.* **8**, 4917–4927 (2018).

28. Wen, Z. et al. Directed evolution of a plant glycosyltransferase for chemo- and regioselective glycosylation of pharmaceutically significant flavonoids. *ACS Catal.* **11**, 14781–14790 (2021).

29. Crameri, A. et al. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291 (1998).

30. Turner, N. J. et al. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5**, 567–573 (2009).

31. Reetz, M. T. et al. Laboratory evolution of stereoselective enzymes: a prolific source of catalysts for asymmetric reactions. *Angew. Chem. Int. Ed.* **50**, 138–174 (2011).

32. Kille, S. et al. Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat. Chem.* **3**, 738–743 (2011).

33. Li, J. et al. Near-perfect control of the regioselective glucosylation enabled by rational design of glycosyltransferases. *Green. Synth. Catal.* **2**, 45–53 (2021).

34. Zhang, J. et al. Catalytic flexibility of rice glycosyltransferase osugt91c1 for the production of palatable steviol glycosides. *Nat. Commun.* **12**, 7030 (2021).

35. Xia, X. L. et al. Directed evolution of the udp-glycosyltransferase ugtbl1 for highly regioselective and efficient biosynthesis of natural phenolic glycosides. *J. Agric. Food Chem.* **72**, 1640–1650 (2024).

36. Weiss, G. A. et al. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl Acad. Sci. USA* **97**, 8950–8954 (2000).

37. Acevedo-Rocha, C. G. et al. Directed evolution of proteins based on mutational scanning. *Methods Mol. Biol.* **1685**, 87–128 (2018).

38. Reetz, M. T. et al. Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angew. Chem. Int. Ed.* **44**, 4192–4196 (2005).

39. Reetz, M. T. et al. Iterative saturation mutagenesis on the basis of b factors as a strategy for increasing protein thermostability. *Angew. Chem. Int. Ed.* **118**, 7909–7915 (2006).

40. Reetz, M. T. & Carballeira, J. D. Iterative saturation mutagenesis (ism) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2**, 891–903 (2007).

41. Reetz, M. T. The importance of additive and non-additive mutational effects in protein engineering. *Angew. Chem. Int. Ed.* **52**, 2658–2666 (2013).

42. Huang, J. et al. Exploring the catalytic function and active sites of a novel *C*-glycosyltransferase from *Anemarrhena asphodeloides*. *Syn. Syst. Biotechno.* **7**, 621–630 (2022).

43. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

44. Xie, K. et al. Exploring and applying the substrate promiscuity of a c-glycosyltransferase in the chemo-enzymatic synthesis of bioactive *C*-Glycosides. *Nat. Commun.* **11**, 5162 (2020).

45. Zhang, M. et al. Functional characterization and structural basis of an efficient Di-*C*-glycosyltransferase from *Glycyrrhiza glabra*. *J. Am. Chem. Soc.* **142**, 3506–3512 (2020).

46. Kang, B. C. et al. Nothofagin suppresses mast cell-mediated allergic inflammation. *Chem. Biol. Interact.* **298**, 1–7 (2019).

47. Minor, W. et al. Optimal structure determination from sub-optimal diffraction data. *Protein Sci.* **31**, 259–268 (2022).

48. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

49. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta crystallogr. D.* **60**, 2126–2132 (2004).

50. Adams, P. D. et al. The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).

51. Wang, L. et al. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **115**, 9431–9438 (2011).

Acknowledgements

Author contributions

Z.-M.Z. and J.W. conceived the project, wrote and revised the manuscript. M. L. contributed to drafting and revising the manuscript. M. L. and Z. W. carried out the cloning and expression of the GTs, protein purification, biochemical kinetics, chemical synthesis and purification of the products. Y.Z. and Y.L. performed the computational study. Z.-M. and Z.Z. crystallized the proteins and determined the structures. Q.Z. contributed to the analysis of the products. Q.N. and Y.M. contributed to the synthesis of the substrates. B.G. and B.C. contributed to the isolation and structure determination of the substrates and products. Z.J. contributed to the analysis of the protein structure and to the revision of the manuscript. All authors approved the final version of the article.

Competing interests

The authors declare no competing interests.

Additional information