Article

Efficient O-glycosylation of Triterpenes Enabled by Protein Engineering of Plant Glycosyltransferase UGT74AC1

Jiao Li, Jiangang Yang, Shicheng Mu, Na Shang, Cui Liu, Yueming Zhu, Yi Cai, Pi Liu, Jianping Lin, Weidong Liu, Yuanxia Sun, and Yanhe Ma

ACS Catal., **Just Accepted Manuscript** • DOI: 10.1021/acscatal.9b05232 • Publication Date (Web): 21 Feb 2020

Efficient *O*-glycosylation of Triterpenes Enabled by Protein Engineering of Plant Glycosyltransferase UGT74AC1

Jiao Li[1,2‡], Jiangang Yang[1‡], Shicheng Mu[1], Na Shang[1], Cui Liu[1], Yueming Zhu[1], Yi Cai[1], Pi Liu[1], Jianping Lin[1], Weidong Liu[1]*, Yuanxia Sun[1,2]* and Yanhe Ma[1]*

[1] National Engineering Laboratory for Industrial Enzymes, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, 32 Xi Qi Dao, Tianjin Airport Economic Area, Tianjin 300308, China

[2] University of Chinese Academy of Sciences, Beijing 100049, China

ABSTRACT: Triterpene *O*-glycosylation has attracted significant interest from the pharmaceutical industry as a valuable means for drug design and development. Plant glycosyltransferases, which catalyze this glycosylation reaction, play the key step in preparing structure diverse and valuable triterpene glycosides. However, this class of enzymes usually suffered from low catalytic efficiency. To address this problem, triterpene glycosyltransferase UGT74AC1 from *Siraitia grosvenorii* was chosen and its crystal structure was solved and employed as the molecular basis to implement directed evolution and sequence/structure-based engineering. Several resultant UGT variants exhibit a 10^2- to 10^4-fold improvement in catalytic efficiency for triterpene glycosylation. Specially, one variant exhibited up to 4.17×10^4-fold increase in catalytic efficiency towards mogrol and 1.53×10^4-fold increase to UDP-glucose, respectively. Moreover, the mutants also displayed extended substrate promiscuity compared with wild type enzyme and conserved regioselectivity. Based on the results of molecular docking and molecular dynamics simulations, it was proposed that the improved enzymatic activity and substrate promiscuity were likely owing to the stable hydrophobic interactions and favorite conformations between the enzyme and substrates. This work has also laid a foundation for the engineering of other plant UGTs for their practical application to the synthesis of valuable triterpene saponins.

KEYWORDS: biocatalysis · catalytic efficiency · protein engineering · plant glycosyltransferase · triterpene saponins

1. INTRODUCTION

Glycosylation is a key step in the synthesis of complex and structure diverse glycosides. Chemical glycosylation usually requires tedious protection and deprotection steps and suffers

from low specificities and yields.[1] Enzymatic glycosylation, on the other hand, exhibit high efficiency and stereo/regiospecificity to create glycosidic linkages rapidly in a direct and an environmental benign manner and are thus particularly attractive for the synthesis of glycosides. Terpenoids represent a large class of natural products. [2] It has been disclosed that glycosylation modification of terpenoids can change their pharmacological activity[3,4] and other properties, such as taste[5] and bioavailability[6] as well, which has motivated considerable interest in terpene glycosylation for drug development in the pharmaceutical industry.[7]

Glycosyltransferases (GTs) (EC 2.4.x.y) catalyze the transfer of glycosyl moiety from an activated nucleotide sugar to a variety of acceptor biomolecules. These enzymes can be classified into 110 families based on amino acid sequence similarities collected in the Carbohydrate Active Enzyme database (CAZy, http://www.cazy.org/).[8] The CAZy GT1 family often referred to as UDP glycosyltransferases (UGTs), which accept UDP sugars as sugar donor,[9] are widely detected in the synthetic pathway of natural glycosides including terpene glycosides. Plant UGTs which are belong to GT1 family displayed excellent properties (e.g. high regioselectivity and substrate specificity)[10] and have attracted substantial attention due to their great potential application in biotechnology.[2] Until now, lots of plant UGT genes from *Arabidopsis thaliana*,[11] *Medicago truncatul*,[12] *Glycyrrhiza uralensis*,[13] *Camellia sinensis*,[14] *Panax ginseng*,[15] *Siraitia grosvenorii*,[16,17] *Glycine max*,[18] *Carthamus tinctorius*[19] and so on were annotated in databases. However, only a few plant UGTs have been characterised to be responsible for triterpene glycosylation.[20] Furthermore, the majority of these UGTs have low catalytic activity, narrow substrate scope and difficult soluble heterologous expression, which hinder their practical application.[21, 22] Therefore, it is necessary to understand the catalytic mechanism of plant UGTs furtherly and engineer them to improve their property and applicability.

Crystal structures not only provide structural basis for understanding enzyme-substrate interaction but also lay the foundation for protein engineering. With the rapid development of protein crystallization techniques, incresing number of GTs structure has been solved and disclosed in PDB database. And three main kinds of structural topologies of GTs (termed as GT-A, GT-B and GT-C folds) were identified until now.[7] In GT1 family, there are 35 UGT crystal structures included (statistics in CAZy), among which 13 plant UGTs structures have been solved, including UGT74F2, UGT72B1 and UGT89C1 from *A. thaliana*,[23-25] Os79 from *Oryza sativa*,[26] UGT78K6 and Ct3GT-A from *C. ternatea*,[27,28] VvGT1 from *V. vinifera*,[29] UGT78G1, UGT85H2 and UGT71G1 from *M. truncatula*,[30-32] UGT76G1 from *Stevia rebaudiana*,[33] TcCGT1 from *Trollius chinensis*,[34] and PtUGT1 from *P. tinctorium*.[35] All of these UGTs contain a GT-B fold like others in GT1, consisting of two β/α/β Rossmann-like domains. Among those UGT structures in database, there was only one plant UGT (*Mt*UGT71G1) except for UGT51 from *Saccharomyces cerevisiae* (*Sc*UGT51), exhibiting catalytic glycosylation of triterpenoids.[32,36] Limited crystal structures restricted the detailed understanding of structure–function relationships and glycosylation mechanisms between UGTs and triterpenes.

Nowadays, several protein engineering strategies are available to increase the catalytic activity and regioselectivity of UGTs. For example, directed evolution of UGT OleD from *Streptomyces antibioticus* was performed using a simple high-throughput screen method based on a fluorescent surrogate acceptor substrate.[37] The triple mutant (P67T/S132F/A242V) displayed marked improvement in catalytic efficiency and substrate promiscuity.[38] The *Sc*UGT51 has also been engineered and its catalytic efficiency to protopanoxadiol increased about 1800-fold after nine rounds of iterative saturation mutagenesis.[39] However, only UGTPg45 from *Panax ginseng* has been engineered using direct evolution method, resulting in a minimal

(1.5-fold) increase in catalytic efficiency to protopanoxadiol.[40] Therefore, dramatically enhancing the catalytic efficiency of plant UGTs remains a challenge probably due to the lack of sufficient molecular templates for their structural modelling and in-depth understanding of their reaction mechanisms especially concerning the enzyme-substrate binding and interaction.

We recently reported that UGT74AC1 from *Siraitia grosvenorii* (*Sg*UGT74AC1) catalyzed the high regiospecific glycosylation of 3-OH of mogrol (**1a**) (Figure 1) and had good soluble expression in *Escherichia coli* BL21(DE3). It showed high application in preparing triterpene glycosides such as Rh2 which possesses versatile anti-tumour and anti-inflammatory activities.[41-43] However, it showed rather weak enzymatic activity and conversion rates.[17] In this research, we analyzed the crystal structures of *Sg*UGT74AC1 and on the basis of this information, this enzyme was engineered to address the challenges in the *O*-glycosylation of triterpenes using **1a** as the model substrate. Remarkable increase in catalytic efficiency to **1a** and high substrate promiscuity were observed for the engineered mutants. Crystal structural studies and molecular dynamics (MD) simulations of wild type (WT) and mutants were also employed to reveal their molecular basis in glycosylation.

2. RESULTS AND DISCUSSION

2.1 Overall Crystal Structure of UGT74AC1. Herein, to better understand the molecular basis of *Sg*UGT74AC1 in glycosylation of **1a**, the crystal structure of *Sg*UGT74AC1 (PDB ID: 6L90) and its complex with UDP-glucose (UDPG) (PDB ID: 6L8Z) were solved by X-ray crystallography at a resolution of 2.02 and 2.10 Å, respectively (Table S1 and Figure S1). As shown in Figure 1A, the structure of *Sg*UGT74AC1 displays the conserved GT-B fold[7] which features the N- (Met1–Lys226) and C-terminals (Cys259–Thr454) β/α/β Rossmann domains and

a less conserved connecting loop region (Thr227-Val258). Moreover, the C-terminal domain consists of ten α-helices and six β-strands and possesses the highly conserved plant secondary product glycosyltransferase (PSPG) motif (Trp330-Gln373), which determines the recognition and binding of the UDP-sugar donor (Figure S2). The UDPG which is located across two helices, Cα5 and Cα6, is perpendicular to six β-sheets of the C-terminal domain and interacts with PSPG through the formation of several hydrogen bonds (Figure S3). Therefore, the residues around these regions can be engineered to increase sugar specificity and catalytic activity.

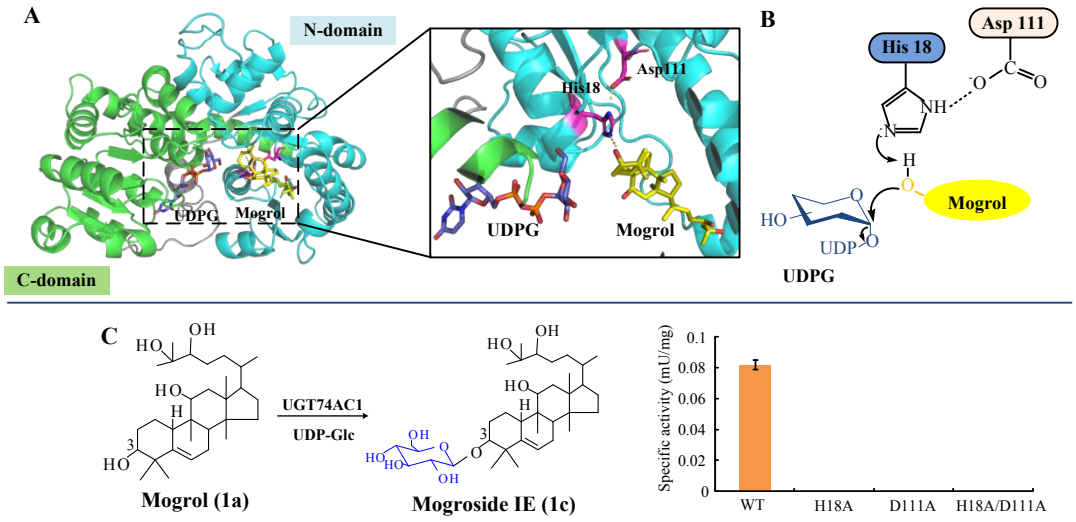

Figure 1. (A) Crystal structures of UGT74AC1 with UDPG (blue sticks), and mogrol (**1a**) was docked into the binding pocket and shown as yellow sticks. The secondary structures within N-domain (Met1-Lys226) and C-domain (Cys259-Thr454) are coloured cyan and green, respectively. The linker between C-domain and N-domain from residues Thr227–Val258 is coloured gray. A detailed active site view of UGT74AC1-UDPG-**1a**. The catalytic residues His18 and Asp111 are shown as magenta sticks. Yellow dotted lines indicate hydrogen bonds. (B) The S_N2 mechanism of UGT74AC1 catalysis. His18, with the aid of Asp111, specifically deprotonates the 3-hydroxyl of mogrol and makes it a nucleophile to attack the sugar donor UDPG. (C) The glycosylation of mogrol (**1a**) catalyzed by UGT74AC1 using UDPG as sugar donor. And specific activity of UGT74AC1 wild type (WT) and mutants (H18A and D111A) in catalysis sites.

The N-terminal domain interacting with the acceptor consists of seven α-helices and seven β-strands. Docking and MD simulation with **1a** and complex enzyme-UDPG showed that **1a**

could form a hydrogen bond with His18 at the acceptor pocket (Figure 1A). The dyad, His18-Asp111, is located close to the sugar donor and the acceptor in *Sg*UGT74AC1. Mutations of His18 and Asp111 to alanine completely eliminated the activity on mogrol (Figure 1C). We proposed that the residue His18 as the general base abstracts a proton from the 3-hydroxyl group of **1a**, and the generated nucleophile attacks the C1 carbon of glucose in UDPG, using S_N2-like mechanism.[32,34,35] Site Asp111 stabilises the catalytic conformation and balances the charge (Figure 1B). The inactivity of the His18 and Asp111 mutation indicated their important catalytic role and confirmed that these two residues are irreplaceable.

2.2 Mutations for Improving Activity. To increase the catalytic efficiency of *Sg*UGT74AC1 in the glycosylation of **1a**, directed evolution and sequence/structure-based design approaches were employed (Figure 2). The error-prone PCR method was executed initially to identify "hot sites". Mutation library containing approximately 5000 clones were selected based on a colour change-based screening method,[39] and five activity-improved mutants (L109I, R28H, F203I, L48M and T79Y) were obtained (Figure 2A). The best hit in this round of screening was variant T79Y which exhibited an eightfold enhancement in specific activity. These five residues are located at the N-domain, and the distance between residue T79 and substrate **1a** is within 4 Å. Subsequently, iterative mutagenesis approaches were applied to test their synergistic effect. Mutant T79Y served as the template to combine with four other mutants in the second-round screening. Mutant M2 (L48M/T79Y) was chosen due to its 48-fold improvement in specific activity compared with WT and served as the template for the sequential incorporation of the remaining sites (R28H, L109I and F203I) (Table S2). The preferable mutant M4 (T79Y/L48M/R28H/L109I), which has a 379-fold higher catalytic efficiency (k_{cat}/K_m value) to **1a** than WT, was discovered (Table 1). Although the three mutations (L48M, R28H and L109I)

are located far (12–21 Å) from **1a**, their incorporation in M4 remarkable amplified the specific activity relative to that of mutant T79Y, indicating that the residues without forming a direct interaction with the substrate also influence the enzymatic activity of UGTs. F203I and L109I mutations based on M3 (T79Y/L48M/R28H) led to improved activity, respectively, but their combination in M5-1 exhibited a negative effect (Figure 2A).

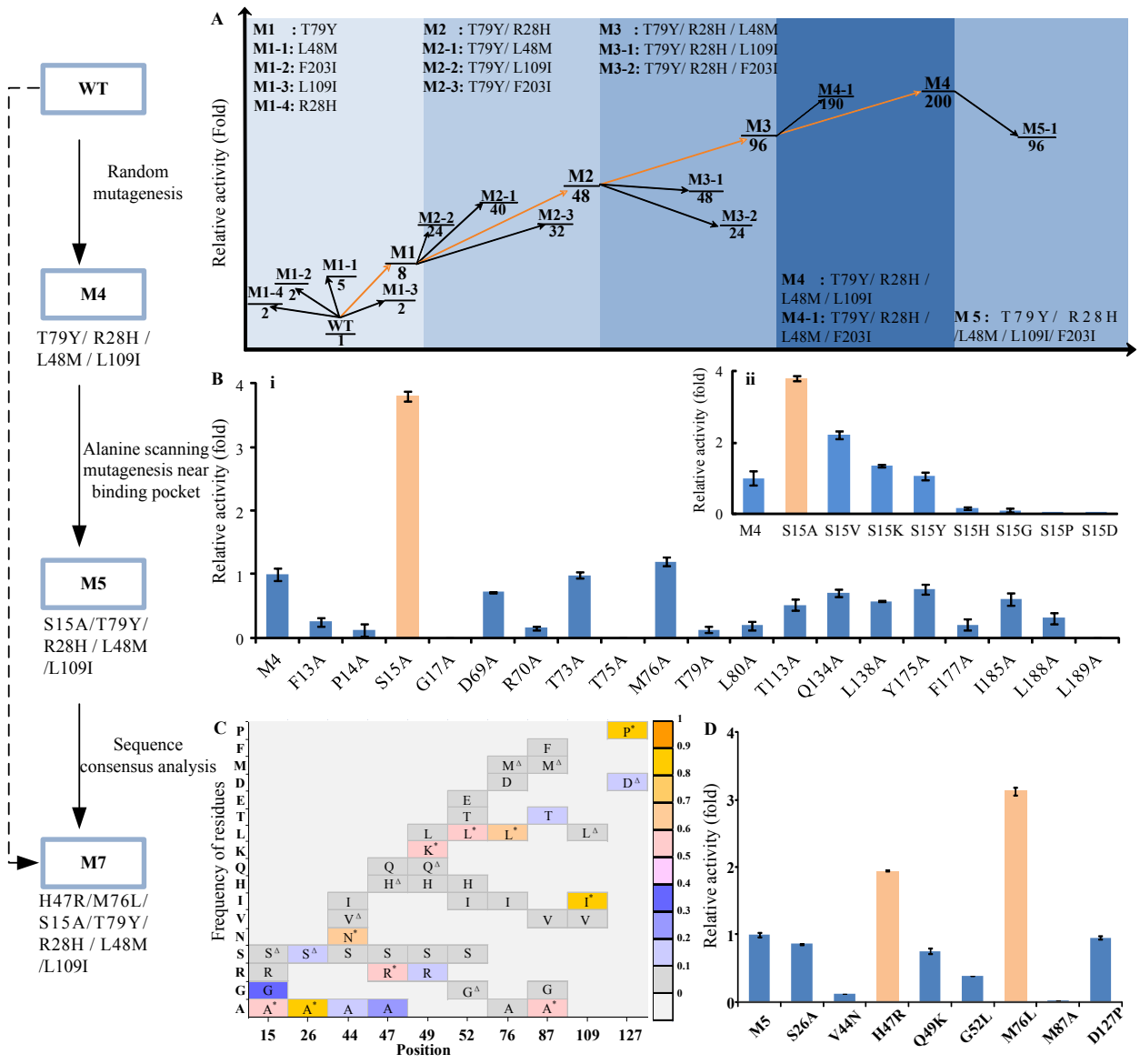

Figure 2. Evolution workflow of UGT74AC1. (A) Five mutants (L109I, R28H, F203I, L48M and T79Y) are screened from approximately 5,000 clones, followed by combination of five sites in different ways. The

numbers below the horizontal lines show the relative activity of randomisation mutants of UGT74AC1 (WT) and combined mutants compared with the WT. The best combination route is presented in the pink line, leading to the best mutant M4 (T79Y/R28H/L48M/L109I, ~200-fold higher activity than WT) in the first round. (B) Relative activity of alanine-scanning mutants of UGT74AC1-M4 around 4 Å to mogrol. Mutant M5 (T79Y/R28H/L48M/L109I/S15A, ~3.8-fold higher than M4) is obtained from alanine-scanning mutagenesis near the binding pocket. Error bars represent standard deviation from three repeats. (C) and (D) Active-based sequence conservation analysis of N-terminal of 13 UGTs that show activity to mogrol or mogrosides. (C) The frequency distribution of amino acid occurrence for mutated targets. * and Δ stands for the mutants and amino acid in WT at each point, respectively. (D) Relative activities of the selected mutants of UGT74AC1-M5 (UGT74AC1-T79Y/L48M/R28H/L109I/S15A). Error bars represent standard deviation from three repeats. The relative activities of mutants were determined on crude extracts.

Inspired by the beneficial mutation T79Y, we focused on the mutations near the catalytic centre. Nineteen residues (F13, P14, S15, G17, D69, R70, T73, T75, M76, T79, L80, T113, Q134, L138, Y175, F177, I185, L188, L189) adjacently located (<4 Å) were selected for alanine scanning mutagenesis with M4 and WT as the template (Figure S4), and the results showed that additional S15A mutation resulted in 3.8-fold and 5.1-fold increased activity than M4 and WT, respectively (Figure 2B-i and Table S3). Then, mutants of M4 having S15 replaced with seven other amino acids, including valine, lysine, tyrosine, histidine, glycine, proline and aspartic acid were investigated. The result (Figure 2B-ii) showed that S15Y, S15K and S15V displayed 1- to 2-fold increased activity than M4. Especially, mutant M5 (T79Y/L48M/R28H/L109I/S15A) showed a 5.15-fold decrease in K_m value and a 5.57×10^3-fold increase in k_{cat}/K_m value, compared with WT (Table 1 and Figure S5, S6), suggesting that M5 showed higher affinity and considerably improved catalytic efficiency to **1a** than WT.

Table 1. Apparent kinetic parameters of UGT74AC1 and its mutants towards mogrol (**1a**) and UDPG.

Substrate	Enzyme	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}\cdot s^{-1}$)	Fold[a]
Mogrol	WT	49.40 ± 5.9	1.02×10^{-4}	2.06×10^{-3}	1
	M4	70.20 ± 4.5	0.05 ± 0.005	0.78 ± 0.08	3.79×10^{2}
	M5	9.59 ± 0.5	0.11 ± 0.03	11.47 ± 0.5	5.57×10^{3}
	M6	15.10 ± 0.3	0.55 ± 0.06	36.42 ± 0.7	1.77×10^{4}
	M7	10.10 ± 0.3	0.86 ± 0.04	86.00 ± 6.6	4.17×10^{4}
UDPG	WT	890.20 ± 5	1.48×10^{-4}	1.67×10^{-4}	1
	M4	64.10 ± 3	0.05 ± 0.004	0.78 ± 0.05	4.67×10^{3}
	M5	66.08 ± 6	0.10 ± 0.002	1.51 ± 0.04	9.06×10^{3}
	M6	219.50 ± 9	0.37 ± 0.01	1.68 ± 0.02	1.01×10^{4}
	M7	148.50 ± 5	0.38 ± 0.03	2.56 ± 0.06	1.53×10^{4}

Notes: WT: wild-type; M4: T79Y/L48M/R28H/L109I; M5: T79Y/L48M/R28H/L109I/S15A; M6: T79Y/L48M/R28H/L109I/S15A/M76L; M7: T79Y/L48M/R28H/L109I/S15A/M76L/H47R.

[a] Fold change over catalytic efficiency of WT.

Amino acids positively affecting the function of protein would survive preferentially during natural selection.[44,45] Consensus sequence design, which replaced non-consensus residues with consensus ones at each positions after multiple sequence alignment (MSA), is a powerful strategy to improve protein thermostability.[46] In this work, we employed an activity-based sequence conservative analysis (ASCA) approach to increase the catalytic efficiency of UGT74AC1. Twelve plant UGT sequences (seven from *Siraitia grosvenorii* and five from *Arabidopsis thaliana*) characterised by their glycosylation activity to **1a** or its glycosides were collected and aligned with UGT74AC1 by using MSA (Figure S1).[47] Analysis of the degree of conservation and accessibility of amino acids interacting with **1a** in the N-terminal domain helped to identify ten residues that had the highest frequency at individual positions. Besides S15A and L109I as mentioned in mutant M5, eight other mutations were executed with M5 as template (Figure 2C). The results in Figure 2D showed that the additional mutation of H47R and M76L displayed 1.95- and 3.20-fold higher activities than M5, respectively. The k_{cat}/K_m value of

resulting mutant M6 (T79Y/L48M/R28H/L109I/S15A/M76L) was 3.17-fold higher than that of M5. Variant M7 (T79Y/L48M/R28H/L109I/S15A/M76L/H47R) was disclosed to be the best mutant with a 4.17×10^4-fold increase in catalytic efficiency to **1a** compared with WT (Table 1). The mutation sites were located at the N-terminal of the enzyme (Figure S7). The ASCA approach gave four positive mutations (S15A, L109I, M76L and H47R), and their combination in M7 led to a 96-fold increase in specific activity compared with M3 (Table S2). This result showed that the ASCA method could be an effective and promising tool to engineer UGTs to enhance the catalytic efficiency from small-sized mutation libraries.

2.3 The Molecular Basis of Enhanced glycosylation Activity. To gain an insight into the source of high catalytic efficiency of the enzyme mutants to **1a**, we obtained the crystal structures of mutant M7 (PDB ID: 6L8W, 2.05 Å) and performed molecular docking and unconstrained MD simulations (200 ns) of complexes WT-UDPG-**1a** and M7-UDPG-**1a** (Figure S8). According to UGT mechanism,[48] a catalytic conformation, which supports catalysis, occurs when the distance between the 3-hydroxyl-O of **1a** and the NE2 nitrogen of the catalytic residue His18 is less than 3.6 Å and the angle of NE2/3-hydroxyl-H/3-hydroxyl-O is larger than 135°. The last 100 ns MD trajectory with 10000 frames in total was utilized to calculate the emerging frequency of catalytic conformations for each complex.[49] The results showed that the frequency value in complex M7-UDPG-**1a** was 74.93 %, approximately five times that observed in WT-UDPG-1a (15.09 %) (Figure S9). This finding revealed that M7-UDPG-**1a** is more favourable than WT-UDPG-**1a** for the formation of catalytically competent poses to produce corresponding glycosides. Catalytic conformations of WT-UDPG-**1a** and M7-UDPG-**1a** are used to calculate the binding free energy by using molecular mechanics generalised Born surface area (MM/GBSA) method implemented in AMBER16.[50] The calculated binding free energies for **1a**

in complexes WT-UDPG-**1a** and M7-UDPG-**1a** were −41.3 and −49.9 kcal/mol, respectively. The lower binding free energy of M7-UDPG-**1a** is in line with the experiment result that M7 have higher affinity than WT towards **1a**.

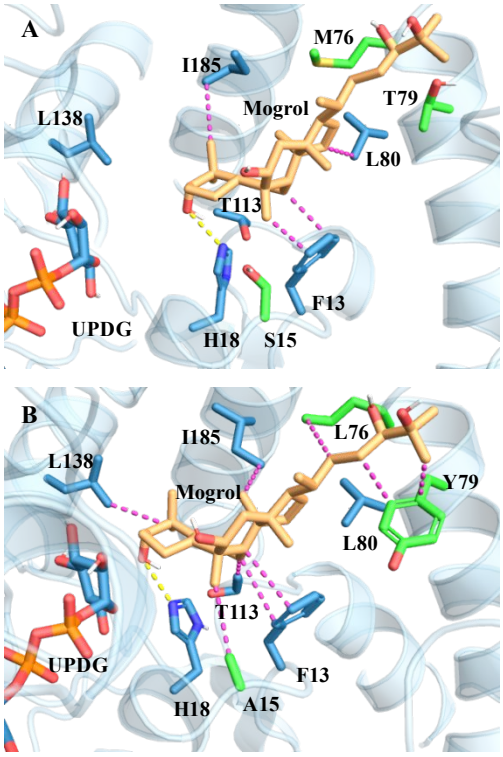

Figure 3. Catalytic conformations of WT-UDPG-**1a** complex (A) and M7-UDPG-**1a** complex (B) in MD simulations. Mogrol and UDPG are shown as orange and blue sticks, respectively. Mutant residues are shown as green sticks and amino acids around **1a** are shown as blue sticks. Yellow dotted lines indicate hydrogen bonds. Hydrophobic interactions are shown as magenta dashes.

The structural characteristics of catalytic conformation of complexes WT-UDPG-**1a** and M7-UDPG-**1a** were then studied to explore the conformational changes. Their most representative conformations were obtained by aligning and clustering the MD trajectory into three clusters based on the backbone atoms of the protein. The representative catalytic conformation of WT-UDPG-**1a** showed that residues His18 formed hydrogen bonds with hydroxyl groups at C3 positions of **1a**, respectively, and hydrophobic interactions were observed

among F13, L80, I185 and **1a** (Figure 3A). However, in M7-UDPG-**1a**, more hydrophobic interactions between enzyme and **1a** were observed than that in WT (Figure 3B). In addition to the hydrophobic interactions of F13, T113, L138 and I185 with **1a**, the replacement of Ser15 with an alanine resulted in a new hydrophobic interaction which facilitated nucleophile attacks. The benzene ring of T79Y in M7 formed new hydrophobic interactions with **1a** and anchored the tail of **1a**. Replacing flexible Met76 with leucine also resulted in additional hydrophobic interaction, which stabilized the hydrophobic tail of **1a**. These hydrophobic interactions resulted in stronger binding affinity than WT-UDPG-**1a** which was reflected in the significantly increased specific activity (Table 1).

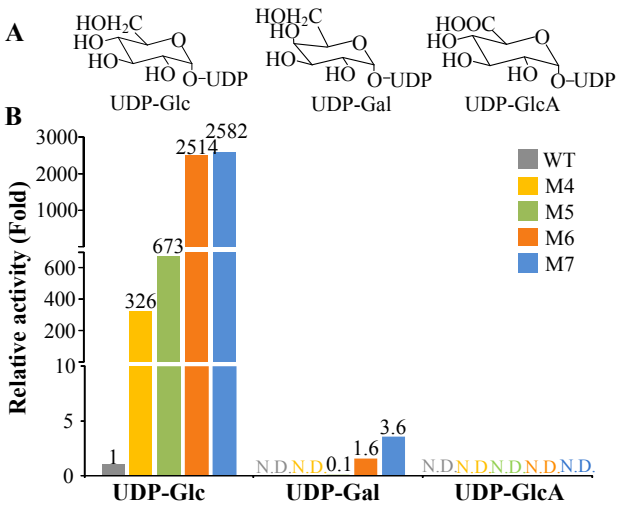

Figure 4. Determinants of sugar donor specificity of UGT74AC1. (A) Structures of UDP-sugar. (B) Comparison of the glycosylation activity of UGT74AC1 WT and mutants for UDP-Glc, UDP-Gal, and UDP-GlcA. N.D.:No detected activity.

2.4 Sugar Donor Scope of SgUGT74AC1 and its Variants. The kinetic parameters of UGT74AC1 and its mutants towards UDPG were further measured. Mutant M4 (T79Y/L48M/R28H/L109I) displayed 14-fold lower K_m value and 4.67×10^3-fold higher k_{cat}/K_m value than WT. M5 had the similar K_m value but increased k_{cat}/K_m value compared to M4. The

k_{cat}/K_m value of M7 was 1.69-fold higher than that of M5 to give an excellent catalytic efficiency to UDPG, which was 1.53×10^4-fold higher than that of WT (Table 1). Next, we investigated the sugar donor scopes for UGT74AC1 and its mutants employing two other UDP-sugars with **1a** as acceptor (Figure 4). WT just accepted UDPG as glycosyl donor. And this enzyme showed no activity to UDP-galactose (UDP-Gal) and UDP-glucuronic acid (UDP-GlcA). Surprisingly, the S15A mutation creatively endowed mutant M5 the catalytic ability to glycosylate **1a** with UDP-Gal as a sugar donor. Moreover, the catalytic activity of M7 with UDP-Gal sugar donor increased 36-fold compared to M5. Structure analysis of complex WT-UDP-Gal-**1a** after MD simulation showed that the catalytic residue His18 formed a hydrogen bond with 4/6–OH of galactose residue in WT-UDP-Gal-**1a**, thereby impeding the catalytic conformation of **1a** (Figure 5A, S10). However, the binding of galactose residue was stabilized in M7-UDP-Gal-**1a** and **1a** was observed in the stable catalytic conformation (Figure 5B). A salt bridge and a hydrogen bond were introduced between H47R and galactose. A psi–psi interaction was found between W330 and the uracil ring of UDP-Gal. The binding free energy of M7-UDP-Gal-**1a** was −103.9 kcal/mol, which is 69.6 kcal/mol lower than that of WT (−34.3 kcal/mol). These structural changes led to stable enzyme-substrate binding to expand sugar donor promiscuity.

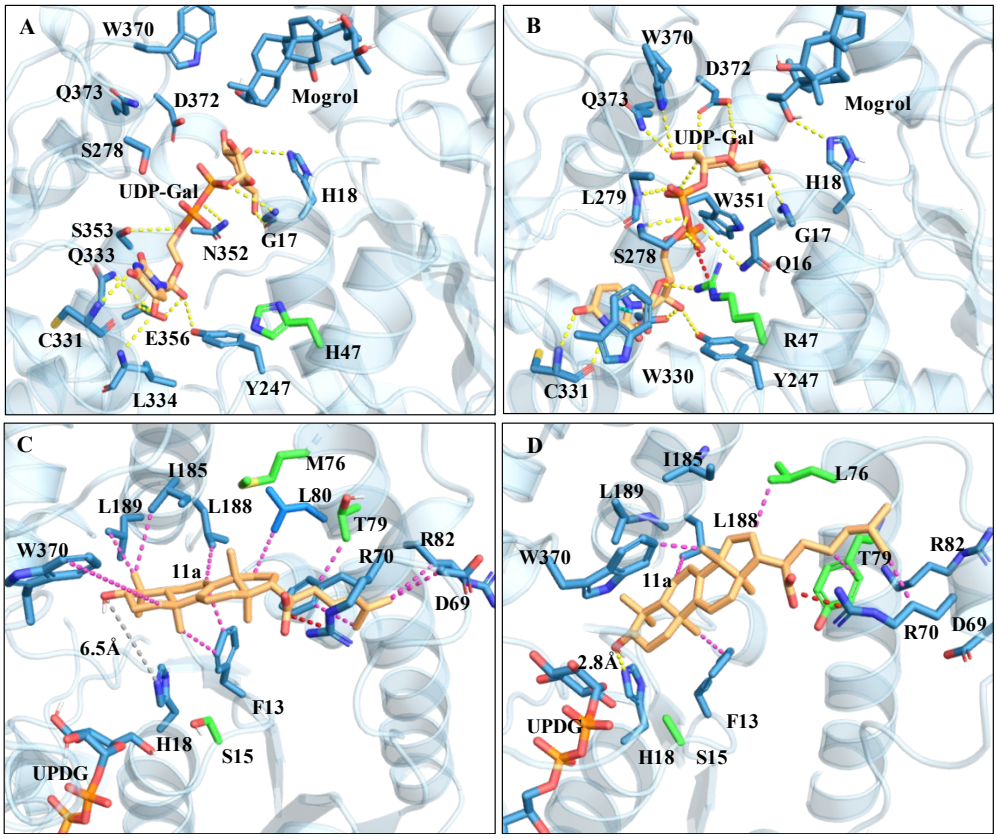

Figure 5. (A) and (B) Representative binding conformations of WT-UDP-Gal-**1a** and M7-UDP-Gal-**1a** in MD simulations, respectively. Mutant residues are shown as green sticks and amino acids that interact with donor UDP-Gal are shown as blue sticks. Yellow, magenta and red dotted lines indicate hydrogen bonds, hydrophobic interactions and salt bridges, respectively. Psi-psi interaction is shown in blue line. The substrate **1a** (upper structure) and UDPG (lower structure) are shown as blue and orange stick models, respectively. (C) and (D) Representative conformations of WT-UDPG-**11a** and M7-UDPG-**11a** in MD simulations, respectively. The substrate **11a** (upper structure) and UDPG (lower structure) are shown as orange and blue stick models, respectively.

2.5 Triterpenes Acceptor Scope of SgUGT74AC1 and Variants. To test the application scope of UGT74AC1 and their mutants in the glycosylation of other structurally different tetracyclic triterpenes, we further examined the enzymatic reactions between **2a**-**11a** and UDPG. Delightfully, although WT UGT74AC1 only had slight catalytic ability to **2a**–**10a**, the engineered mutants displayed 10^2–10^4-fold increase in their catalytic ability (Figure 6, Table S4

and Figure S14–S25). Especially, the highest improvement (1.3×10^4-fold) in specific activity to **10a** was observed under the catalysis of mutant M6. Compounds **1a**, **3a**, **4a**, **8a** and **9a** proved to be good substrates for mutant M7, and substrates **5a**–**7a** were highly suitable for mutant M5. It was noted that substrates bearing alkyl or alkene group at tail, e.g. **1a**-**4a**, were easily glycosylated by mutants in high efficiency, in contrast to triterpenes bearing a carboxyl group at various positions of the side chain, e.g. **5a**-**11a**. The hydroxyl and carbonyl of triterpene skeleton, such as **3a**/**4a** and **7a**/**8a**, also influenced the catalytic ability probably due to the change in hydrophobic interactions between the enzyme and substrates. Unexpectedly, mutants M6 and M7 displayed catalytic activity towards dehydrotrametenolic acid (**11a**) which cannot be accepted by WT. We noted that UGT51 mutant with seven mutations (M7-1) achieved 610-fold higher enzyme activity to **3a** than that of the wild type.[39] However, this mutant showed no activity to **1a**. It was noted that the mutants in our study exhibited not only higher multiples (>3000-fold) to **3a** (Figure 6) but also larger sugar donor and triterpene acceptor promiscuity compared with their work.

To gain insights into the basis for enzyme mutants to catalyze the glycosylation of different triterpenes, the catalytic conformation of **11a** was also explored. The orientation of **11a** can sustain a relatively stable catalytic conformation in M7-UGT-**11a** with frequency of catalytic conformation for 86.67%, while it can't form catalytic conformation in WT-UDPG-**11a** (Figure S11, S12). In WT-UDPG-**11a**, the long distance between catalytic residues His18 and C3–OH of **11a** (6.5 Å) resulted in non-catalytic conformations and non-activity (Figure 5C). However, the mutations of M76L and T79Y formed hydrophobic interaction with the hydrophobic tail of **11a**. Mutations M76L and T79Y anchored the substrate in the proper orientation, pushing the C3–OH of **11a** being close to the catalytic site of His18 (2.8 Å) and enabling the catalysis (Figure 5D).

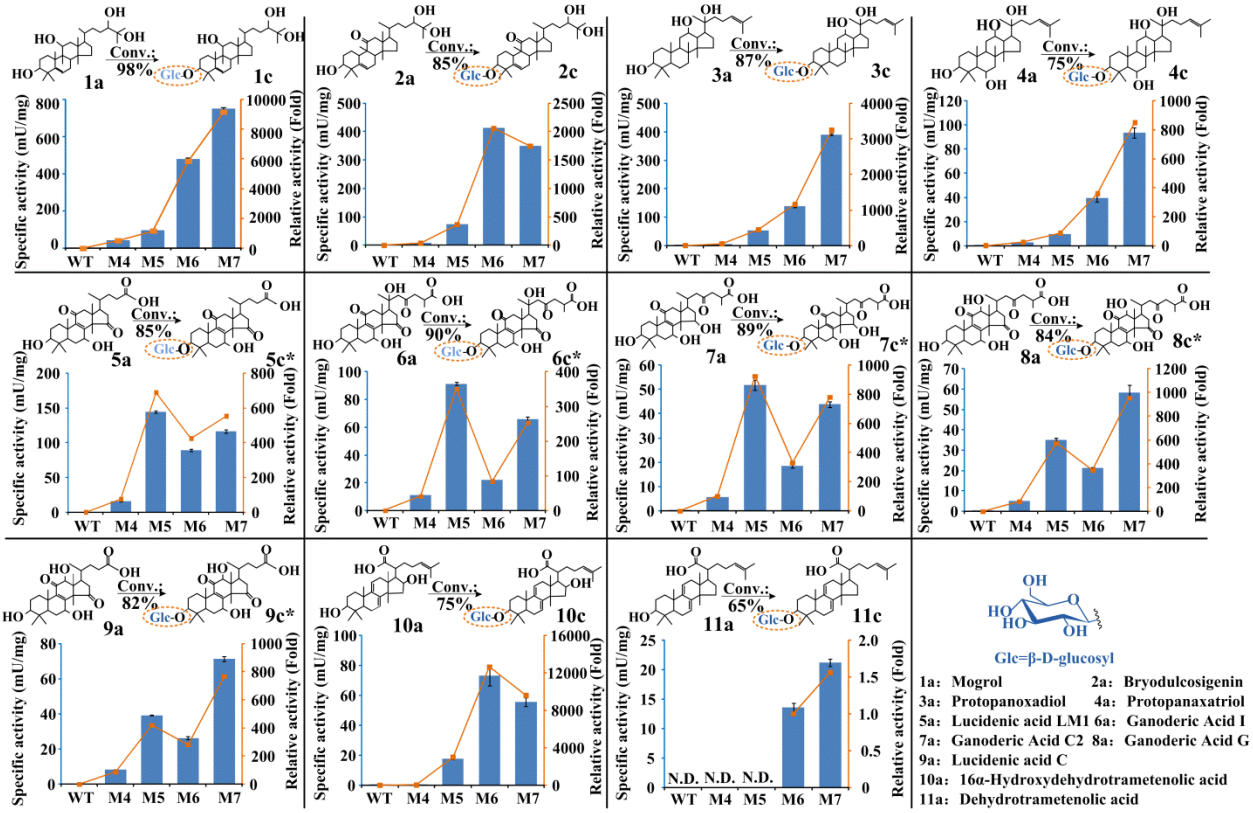

Figure 6. Specific and relative activity of UGT74AC1 (WT) and its mutants using different substrates (**1a**–**11a**) and UDPG as acceptor and donor, separately. "*" indicate new compounds.

The glycosylated products were prepared on a preparative scale and their structures were analysed by HPLC-MS and NMR (Figure S14-S70). The glycosyl moieties were β–linked to the C3 hydroxyl group of triterpenes, suggesting that the engineered proteins conserved the same high regioselectivity as WT enzyme. Five new triterpenes glycosides **5c**–**9c**, which have not been achieved by other UGTs previously, were acquired. Those novel triterpene saponins may have application potential in pharmaceutical study.

3. Conclusions

In summary, we have successfully engineered a plant UGT, *Sg*UGT74AC1, to attain several mutants that exhibited dramatically increased catalytic abiltity (10^2-10^4 folds) in triterpene

glycosylation by means of combining error-prone PCR, structure-based semi-rational design and activity-based sequence conservative analysis approaches. Moreover, compared with WT enzyme, the engineered mutants also displayed extended donor and acceptor promiscuity while maintaining the regioselectivity for C3 hydroxyl group of triterpene substrates. Crystal structure analysis and MD simulations provided a better understanding of the origins of the improved catalytic activity and substrate promiscuity of these UGT mutants. It is the hydrophobic interactions between enzyme and substrate that play a critical role in anchoring triterpenes at the active centre. This research has also provided a model and some guidance for the engineering of other plant UGTs. Given that these mutant enzymes exhibited different catalytic activities for different triterpenes, our future work will be focused on the engineering of the enzymes to suit specific triterpenes and further improvement of their catalytic activities and application scopes.

4. Experimental section

4.1 Materials and chemicals

Terpenoids (mogrol (**1a**), mogroside IE (**1c**), Bryodulcosigenin (**2a**), protopanaxadiol (**3a**), protopanaxatriol (**4a**), Lucidenic acid LM_1 (**5a**), Ganoderic Acid I (**6a**), Ganoderic Acid C2 (**7a**), Ganoderic Acid G (**8a**), Lucidenic acid C (**9a**), 16α-Hydroxydehydrotrametenolic acid (**10a**) and Dehydrotrametenolic acid (**11a**)) were purchased from Chengdu Biopurify Phytochemicals. And all other reagents were purchased from Sigma–Aldrich.

4.2 Heterogonous expression and purification of enzymes

The coding regions of UGT74AC1 were cloned from the fruits of *S. grosvenorii* and inserted into the pET-32a expression vector. The gene of macrolide-inactivating glycosyltransferase (OleD) variant TDP-16, which has been applied in high throughput screening of UGT74AC1 mutants,

was synthesized by Genscript (nanjing) Co., Ltd and inserted into expression plasmid pET21. The recombinant vectors pET32a-UGT74AC1 or pET21-TDP-16 were then transformed into *E. coli* BL21 (DE3) for expression. The strains harboring appointed plasmid were grown in LB medium containing 100 μg/mL of ampicillin at 37 ºC. When OD_{600} reached 0.6-0.8, the cells were induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 16 ºC, 200 rpm for 20 h. The cell pellets were harvested by centrifugation at 5000 g for 10 min at 4 °C, then, suspended in lysis buffer containing 50 mM Tris-HCl (pH 7.4), 150 mM NaCl and 25 mM imidazole. After sonication, the cell debris was centrifuged at 17000 g and 4 ºC. The supernatant was applied to Ni-NTA agarose affinity column to pure the recombinant proteins.

4.3 Construction of UGT74AC1 mutant libraries and library screening

GeneMorph II Random Mutagenesis Kit (Agilent Technologies) was used to introduce mutations during error-prone PCR process. The pET32a-UGT74AC1 plasmid served as template. All procedures were performed according to manufactures' protocol. Approximately 5000 mutants were examined. Site-saturation mutations were applied using primers containing NNK codon at specified sites. All primers used are provided in Table S5. About 100 colonies at one site were screened. Colonies were cultured in 96 deep-well plates with 300 μL LB medium (100 μg/mL of ampicillin) in each well at 37 ºC overnight. After that, 10 μL culture of each cell was transferred to a new plate with 1mL LB containing 100 μg/mL of ampicillin, and incubated at 37 ºC for 3 h. when OD_{600} reached 0.6-0.8, the culture was applied with 0.4 mM IPTG at 16 ºC for 24 h. The cell was harvested and suspended in 200 μL lysis buffer (20 mM Tris-HCl, 0.3 mg/ml lysozyme, pH 8.0) and suffered freeze-thaw three times. Cells were removed and the supernatant obtained was used for high-throughput assay. The reaction was in 0.1 mL Tris-HCl (20 mM, pH 8.0) containing 0.25 mM mogrol, 1 mM UDP-Glucose, 0.5 mM 2-chloro-4-nitrophenyl β-D-

glucopyranoside, 1% Tween 80, 11 μM OleD variant TDP-16 and 50 μL supernatant. The reaction was incubated at 30 ºC for 20 min and monitored at 410 nm.[51]

4.4 Specific activities of UGT74AC1 and its mutants

To detect the activity of UGT74AC1 and its mutants, the UGT activity assay was conducted in 300 μL reaction buffer, including 10 mM UDPG, 1 mM $MgCl_2$, 50 mM Tris-HCl (pH 8.0) and 0.8 mg purified enzyme of WT or ~100 μg purified enzyme of mutants (depending on the activity of mutants). The reaction was incubated at 40 ºC for 10 min (mutants) and 3h (WT) and then stopped by adding same volume of Methanol. The reactants were centrifuged and filtered through 0.22 μm filter and analyzed by HPLC or HPLC-MS. A single unit of enzymatic activity was defined as the amount of enzyme that consuming 1 μmol of substrates acceptor per min.

4.5 Kinetic analysis of UGT74AC1 and its mutants

For the kinetic analysis of UGT74AC1 and its mutants towards mogrol, the reactions were measured in 300 μL reaction buffer (10 mM UDPG, 1 mM $MgCl_2$, 50 mM Tris-HCl pH 8.0 and 0.8 mg purified enzyme of WT or 100 μg purified enzyme of mutants) at 40 ºC for 4 h (WT) or 30 min (mutants), with the concentration of mogrol varied from 0.02-0.4 mM and then stopped by adding same volume of Methanol. For the kinetic analysis of UGT74AC1 and its mutants towards UDPG, the reactions were measured in 300 μL reaction buffer (0.2 mM mogrol, 1 mM $MgCl_2$, 50 mM Tris-HCl pH 8.0, 0.8 mg purified enzyme of WT; 100 μg purified enzyme of mutants) at 40 ºC for 4 h (WT) or 30 min (mutants), with the concentration of UDPG varied from 0.02-0.4 mM and then stopped by adding same volume of Methanol. Reaction mixture was analyzed by HPLC or HPLC-ESI-MS, using a reversed-phase column (Ultimate C18 column, 4.6 mm × 250 mm, 5 μm particle, Welch, Shanghai, China). ESI methods were as described before.[34,35] For HPLC analyses, solvent A (H_2O with 0.1% formic acid) and solvent B

(acetonitrile with 0.1% formic acid) was utilized as the mobile phase at flow rate of 1 mL/min and appointed UV wavelength and gradient program are listed in Table S6.

4.6 Preparation of glycosides

The reactions were conducted in 200 mL reaction buffer (50 mg acceptor, 200 mg UDPG, 1 mM $MgCl_2$, 50 mM Tris-HCl pH 8.0 and 1L cell lysates of M7) at 40 ºC for 5 h and same volume of Methanol was added. Then, the reaction mixture was evaporated to 5 mL. To identify the structure of products, glycosides were purified by a preparative HPLC system using reverse-phase Ultimate C18 column and redissolved in dimethylsulfoxide-d_6. And finally, purified glycosylated products were subjected to AVANCE III 600MHz spectrometer and characterized by ^{1}H NMR, ^{13}C NMR, COSY, TOCSY, HSQC and HMBC.

4.7 Crystallization and structure solution

All crystallization experiments were conducted at 25 ºC using the sitting-drop vapor-diffusion method. In general, 1 μL UGT74AC1 containing solution (25 mM Tris-HCl, pH 7.5, 150 mM NaCl; 30 mg/mL) was mixed with 1 μL of reservoir solution in 48-well Cryschem Plates, and equilibrated against 100 μL of the reservoir solution. The optimized crystallization condition of UGT was 20% PEG8000, 0.2 M magnesium acetate tetrahydrate, 0.1M sodium acodylate pH 6.5. Within 3 to 4 days, the crystals reached dimensions suitable for X-ray diffraction. The mutant UGT74AC1 crystals were grown in the same condition as wild-type enzyme. The crystals in complex with UDPG were obtained by soaking with mother liquor containing 5 mM corresponding molecules for 3-5 h before data collection. All of the X-ray diffraction data sets were collected at beam line BL17U1, BL17B1, BL18U1 and BL19U1 of the Shanghai Synchrotron Radiation Facility (SSRF). The crystals were mounted in a cryoloop and soaked with cryoprotectant solution (25% PEG8000, 5% Glycerol, 0.2 M magnesium acetate

tetrahydrate, 0.1M sodium acodylate pH 6.5) prior to data collection at 100 K. The diffraction images were processed by using HKL2000.[52] The crystal structure of wild-type UGT74AC1 was solved by molecular replacement (MR) method with MrBUMP[53] from the CCP4[54] using the structure of the UGT85H2 from *Medicago truncatula* (PDB ID code 2PQ6, 27 % sequence identity with UGT74AC1) as a search model, most of the residues were built by using Buccaneer[55] and ARP/wARP,[56] the further manual model building and refinement was carried out using Refmac5[57] and Coot.[58] Prior to structural refinements, 5% randomly selected reflections were set aside for calculating R_{free} as a monitor.[59] The mutant and complex structures were determined by MR method with Phaser[60] using the refined UGT74AC1 structure as a search model. The $2F_o$-F_c difference Fourier map showed clear electron densities for most amino acid residues. Subsequent refinements by incorporating ligands and water molecules were according to 1.0 σ map level. Data collection and refinement statistics are summarized in Table S1. All figures were prepared by using the PyMOL program (http://pymol.sourceforge.net/).

4.8 Molecular docking and molecular dynamics simulation

In order to shed light on possible reasons for high catalytic efficiency or new activity of mutant M7, UDPG from crystal structure of 2ACW was import into the UDPG binding site of UGTs by superimposition. The Glide module in Schrödinger program was used to dock substrates into the enzymes.[61] The molecular dynamics (MD) simulation was carried out with Amber 16 molecular dynamics package. Substrates parameters were generated using the antechamber module and GAFF2 with AM1-BCC charges.[62] The complex system was solvated in TIP3P with rectangular water box with a buffer distance of 10 Å. Neutralization of the system was achieved by adding Na^+ counter-ions. A two-stage geometry optimization was performed first for the water and ions, then the entire system, using steepest descent algorithm for the first 2000 steps followed by

conjugated gradient algorithm for another 1000 steps. Then, the system was slowly heated up to 300 K and well equilibrated. A 10 ns MD simulation was carried out with a distance constraint (3.0 Å, 50 kcal/mol) between the coordinating N atom of catalytic residue His18 and the O atom of the **1a** or O atom of **11a**, thereby simulating the induced-fit process of substrate binding. Finally, 200 ns unconstrained production MD was performed at 300 K and 1 atm with 2 fs integration time step. The ff14SB force field was used to model the protein[63] and the general AMBER force field (GAFF) for substrates.[64] The SHAKE algorithm[65] was employed to fix all bonds involving hydrogen atoms and the particle mesh Ewald (PME) method[66] was applied to treat long range electrostatic interactions. Substrate binding free energies were calculated using molecular mechanics generalized Born surface area (MM-GBSA) method implemented in AMBER16. Catalytic conformation was defined according to the reaction mechanism[48] and the emerging frequency of catalytic conformations was calculated in the last 100 ns of MD simulations. Protein–ligand interaction profiler[67] and Pymol was used to visualize models and construct graphical illustrative figures.

ASSOCIATED CONTENT

SUPPORTING INFORMATION

This information is available free of charge on the ACS Publications website.

Supporting tables and figures showing details of specific activity of variant, crystallographic data collection, SDS-PAGE of pure protein, HPLC chromatograms, NMR analysis and molecular dynamics simulations.

AUTHOR INFORMATION

Corresponding Author

* E-mail: sun_yx@tib.cas.cn; liu_wd@tib.cas.cn; lin_jp@tib.cas.cn

Author Contributions

‡ J.L. and J.Y. contributed equally to this work. They performed the protein engineering experiments, analyzed the data and prepared the manuscript. N.S. and W.L. sovled the crystal structure of *Sg*UGT74AC1 and its mutant. J. L., P. L., and C. L. performed the MD simulation and analyszed the data. Y. Z., S.M., and Y.C. purified the products and analysized their structure. Y.S and Y.M designed the experiment and revised the manuscript.

Notes

The authors declare no conflict of interest.

ABBREVIATIONS

UDP, Uridine diphosphat; UGT, Uridine diphosphate glycosyltransferases; MD, molecular dynamics; WT, wild type; UDPG, UDP-glucose; PSPG, plant secondary product glycosyltransferase; MSA, multiple sequence alignment; MM-GBSA, molecular mechanics generalised Born surface area; UDP-Gal, UDP-galactose; UDP-GlcA, UDP-glucuronic acid; HPLC-MS, High Performance Liquid Chromatography-Mass Spectrometry; NMR: Nuclear Magnetic Resonance.

ACKNOWLEDGMENT

This work was supported by the National Key Research and Development Program of China (2019YFA0905100), Natural Science Foundation of China (No. 31771909), Tianjin synthetic biotechnology innovation capacity improvement action.

REFERENCES

(1) Hsu, C. H.; Hung, S. C.; Wu, C. Y.; Wong, C. H. Toward Automated Oligosaccharide Synthesis. *Angew. Chem., Int. Ed.* **2011**, *50*, 11872-11923.

(2) Tiwari, P.; Sangwan, R. S.; Sangwan, N. S. Plant Secondary Metabolism Linked Glycosyltransferases: an Update on Expanding Knowledge and Scopes. *Biotechnol. Adv.* **2016**, *34*, 714-739.

(3) Makino, T.; Shimizu, R.; Kanemaru, M.; Suzuki, Y.; Moriwaki, M.; Mizukami, H. Enzymatically Modified Isoquercitrin, α-oligoglucosyl Quercetin 3-O-glucoside, is Absorbed More Easily than Other Quercetin Glycosides or Aglycone after Oral Administration in rats. *Biol. Pharm. Bull.* **2009**, *32*, 2034-2040.

(4) Liang, C.; Zhang, Y.; Jia, Y.; Wang, W.; Li, Y.; Lu, S.; Jin, J. M.; Tang, S. Y. Engineering a Carbohydrate-processing Transglycosidase into Glycosyltransferase for Natural Product Glycodiversification. *Sci. Rep.* **2016**, *6*, 21051.

(5) Prakash, I.; Markosyan, A.; Bunders, C. Development of Next Generation Stevia Sweetener: Rebaudioside M. *Foods* **2014**, *3*, 162-175.

(6) Dai, L.; Li, J.; Yang, J.; Men, Y.; Zeng, Y.; Cai, Y.; Sun, Y. Enzymatic Synthesis of Novel Glycyrrhizic Acid Glucosides Using a Promiscuous *Bacillus* Glycosyltransferase. *Catalysts* **2018**, *8*, 615.

(7) Liang, D.; Liu, J.; Wu, H.; Wang, B.; Zhu, H.; Qiao, J. Glycosyltransferases: Mechanisms and Applications in Natural Product Development. *Chem. Soc. Rev.* **2015**, *44*, 8350-8374.

(8) Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. The Carbohydrate-Active Enzymes Data-Base (CAZy): an Expert Resource for Glycogenomics. *Nucleic Acids Res.* **2009**. *37*, D233–D238.

(9) Yonekura-Sakakibara, K.; Hanada, K. An Evolutionary View of Functional Diversity in Family 1 Glycosyltransferases. *Plant J.* **2011**, *66*, 182-193.

(10) McArthur, J. B.; Chen, X. Glycosyltransferase Engineering for Carbohydrate Synthesis. *Biochem. Soc. Trans.* **2016**, *44*, 129-142.

(11) Li, Y.; Baldauf, S.; Lim, E.-K.; Bowles, D. J. Phylogenetic Analysis of the UDP-Glycosyltransferase Multigene Family of *Arabidopsis thaliana*. *J. Biol. Chem.* **2001**, *276*, 4338-4343.

(12) Achnine, L.; Huhman, D. V.; Farag, M. A.; Sumner, L. W.; Blount, J. W.; Dixon, R. A. Genomics-Based Selection and Functional Characterization of Triterpene Glycosyltransferases from the Model Legume *Medicago truncatula*. *Plant J.* **2005**, *41*, 875-887.

(13) Mochida, K.; Sakurai, T.; Seki, H.; Yoshida, T.; Takahagi, K.; Sawai, S.; Uchiyama, H.; Muranaka, T.; Saito, K. Draft Genome Assembly and Annotation of *Glycyrrhiza uralensis*, a Medicinal Legume. *Plant J.* **2017**, *89*, 181-194.

(14) Cui, L.; Yao, S.; Dai, X.; Yin, Q.; Liu, Y.; Jiang, X.; Wu, Y.; Qian, Y.; Pang, L.; Gao, L.; Xia, T. Identification of UDP-Glycosyltransferases Involved in the Biosynthesis of Astringent Taste Compounds in Tea (*Camellia sinensis*). *J. Exp. Bot.* **2016**, *67*, 2285-2297.

(15) Kang, K. B.; Jayakodi, M.; Lee, Y. S.; Park, H.-S.; Koo, H. J.; Choi, I. Y.; Kim, D. H.; Chung, Y.J.; Ryu, B.; Dong, Y. L.; Sung, S. H.; Yang, T.-J. Identification of Candidate UDP-Glycosyltransferases Involved in Protopanaxadiol-Type Ginsenoside Biosynthesis in *Panax ginseng*. *Sci. Rep.* **2018**, *8*, 1-10.

(16) Itkin, M.; Davidovich-Rikanati, R.; Cohen, S.; Portnoy, V.; Doron-Faigenboim, A.; Oren, E.; Freilich, S.; Tzuri, G.; Baranes, N.; Shen, S.; Petreikov, M.; Sertchook, R.; Ben-Dor,

S.; Gottlieb, H.; Hernandez, A.; Nelson, D.; Paris, H.; Tadmor, Y.; Burger, Y.; Lewinsohn, E.; Katzir, N.; Schaffer, A. The Biosynthetic Pathway of the Nonsugar, High-Intensity Sweetener Mogroside V from *Siraitia grosvenorii*. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E7619-E7628.

(17) Dai, L.; Liu, C.; Zhu, Y.; Zhang, J.; Men, Y.; Zeng, Y.; Sun, Y. Functional Characterization of Cucurbitadienol Synthase and Triterpene Glycosyltransferase Involved in Biosynthesis of Mogrosides from *Siraitia grosvenorii*. *Plant Cell Physiol.* **2015**, *56*, 1172-1182.

(18) Yin, Q.; Shen, G.; Di, S.; Fan, C.; Chang, Z.; Pang, Y. Genome-Wide Identification and Functional Characterization of UDP-Glucosyltransferase Genes Involved in Flavonoid Biosynthesis in *Glycine max*. *Plant Cell Physiol*. **2017**, *58*, 1558-1572.

(19) Xie, K.; Chen, R.; Li, J.; Wang, R.; Chen, D.; Dou, X.; Dai, J. Exploring the Catalytic Promiscuity of a New Glycosyltransferase from *Carthamus tinctorius*. *Org. Lett.* **2014**, *16*, 4874-4877.

(20) Rahimi, S.; Kim, J.; Mijakovic, I.; Jung, K.-H; Choi, G.; Kim, S.-C.; Kim, Y.-J. Triterpenoid-Biosynthetic UDP-Glycosyltransferases from Plants. *Biotechnol. Adv.* **2019**, *37*, 107394.

(21) Fan, B.; Chen, T.; Zhang, S.; Wu, B.; He, B. Mining of Efficient Microbial UDP-Glycosyltransferases by Motif Evolution Cross Plant Kingdom for Application in Biosynthesis of Salidroside. *Sci. Rep.* **2017**, *7*, 1-9.

(22) Dai, L.; Liu, C.; Li, J.; Dong, C.; Yang, J.; Dai, Z.; Zhang, X.; Sun, Y. One-pot Synthesis of Ginsenoside Rh2 and Bioactive Unnatural Ginsenoside by Coupling Promiscuous Glycosyltransferase from *Bacillus Subtilis* 168 to Sucrose Synthase. *J. Agric. Food Chem.* **2018**, *66*, 2830-2837.

(23) Thompson, A. M. G.; Iancu, C. V.; Neet, K. E.; Dean, J. V.; Choe, J.-Y. Differences in Salicylic Acid Glucose Conjugations by UGT74F1 and UGT74F2 from *Arabidopsis thaliana*. *Sci. Rep.* **2017**, *7*, 46629.

(24) Brazier-Hicks, M.; Offen, W. A.; Gershater, M. C.; Revett, T. J.; Lim, E.-K.; Bowles, D. J.; Davies, G. J.; Edwards, R. Characterization and Engineering of the Bifunctional N- and O-Glucosyltransferase Involved in Xenobiotic Metabolism in Plants. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 20238-20243.

(25) Zong, G.; Fei, S.; Liu, X.; Li, J.; Gao, Y.; Yang, X.; Wang, X.; Shen, Y. Crystal Structures of Rhamnosyltransferase UGT89C1 from *Arabidopsis thaliana* Reveal the Molecular Basis of Sugar Donor Specificity for UDP-β-l-Rhamnose and Rhamnosylation Mechanism. *Plant J.* **2019**, *99*, 257-269.

(26) Wetterhorn, K. M.; Newmister, S. A.; Caniza, R. K.; Busman, M.; McCormick, S. P.; Berthiller, F.; Adam, G.; Rayment, I. Crystal Structure of Os79 (Os04g0206600) from *Oryza sativa*: A UDP-Glucosyltransferase Involved in the Detoxification of Deoxynivalenol. *Biochemistry* **2016**, *55*, 6175-6186.

(27) Hiromoto, T.; Honjo, E.; Noda, N.; Tamada, T.; Kazuma, K.; Suzuki, M.; Blaber, M.; Kuroki, R. Structural Basis for Acceptor-Substrate Recognition of UDP-Glucose: Anthocyanidin 3-O-Glucosyltransferase from *Clitoria ternatea*. *Protein Sci*. **2015**, *24*, 395-407.

(28) Hiromoto, T.; Honjo, E.; Tamada, T.; Noda, N.; Kazuma, K.; Suzuki, M.; Kuroki, R. Crystal Structure of UDP-Glucose: Anthocyanidin 3-O-Glucosyltransferase from *Clitoria ternatea*. *J. Synchrotron Radiat.* **2013**, *20*, 894-898.

(29) Offen, W.; Martinez-Fleites, C.; Yang, M.; Kiat-Lim, E.; Davis, B. G.; Tarling, C. A.; Ford, C. M.; Bowles, D. J.; Davies, G. J. Structure of a Flavonoid Glucosyltransferase Reveals the Basis for Plant Natural Product Modification. *EMBO J.* **2006**, *25*, 1396-1405.

(30) Modolo, L. V.; Li, L.; Pan, H.; Blount, J. W.; Dixon, R. A.; Wang, X. Crystal Structures of Glycosyltransferase UGT78G1 Reveal the Molecular Basis for Glycosylation and Deglycosylation of (iso) Flavonoids. *J. Mol. Biol*. **2009**, *392*, 1292-1302.

(31) Li, L.; Modolo, L. V.; Escamilla-Trevino, L. L.; Achnine, L.; Dixon, R. A.; Wang, X. Crystal Structure of *Medicago truncatula* UGT85H2–Insights into the Structural Basis of a Multifunctional (iso) Flavonoid Glycosyltransferase. *J. Mol. Biol*. **2007**, *370*, 951-963.

(32) Shao, H.; He, X.; Achnine, L.; Blount, J. W.; Dixon, R. A.; Wang, X. Crystal Structures of a Multifunctional Triterpene/Flavonoid Glycosyltransferase from *Medicago truncatula*. *The Plant Cell*, **2005**, *17*, 3141-3154.

(33) Yang, T.; Zhang, J.; Ke, D.; Yang, W.; Tang, M.; Jiang, J.; Cheng, G.; Li, J.; Cheng, W.; Wei, Y.; Li, Q.; Naismith, J. H.; Zhu, X. Hydrophobic Recognition Allows the Glycosyltransferase UGT76G1 to Catalyze its Substrate in Two Orientations. *Nat. Commun.* **2019**, *10*, 1-12.

(34) He, J.-B.; Zhao, P.; Hu, Z.-M.; Liu, S.; Kuang, Y.; Zhang, M.; Li, B.; Yun, C.-H.; Qiao, X.; Ye, M. Molecular and Structural Characterization of a Promiscuous C-Glycosyltransferase from *Trollius chinensis*. *Angew. Chem.* **2019**, *131*, 11637-11644.

(35) Hsu, T. M.; Welner, D. H.; Russ, Z. N.; Cervantes, B.; Prathuri, R. L.; Adams, P. D.; Dueber, J. E. Employing a Biochemical Protecting Group for a Sustainable Indigo Dyeing Strategy. *Nat. Chem. Biol.* **2018**, *14*, 256.

(36) Chen, L.; Zhang, Y.; Feng, Y. Structural Dissection of Sterol Glycosyltransferase UGT51 from *Saccharomyces cerevisiae* for Substrate Specificity. *J. Struct. Biol.* **2018**, *204*, 371-379.

(37)Williams, G. J.; Thorson, J.S. A High-Throughput Fluorescence-Based Glycosyltransferase Screen and its Application in Directed Evolution. *Nat. Protoc.* **2008**, *3*, 357.

(38) Williams, G. J.; Zhang, C.; Thorson, J. S. Expanding the Promiscuity of a Natural-Product Glycosyltransferase by Directed Evolution. *Nat. Chem. Biol.* **2007**, *3*, 657-662.

(39) Zhuang, Y.; Yang, G.-Y.; Chen, X.; Liu, Q.; Zhang, X.; Deng, Z.; Feng, Y. Biosynthesis of Plant-Derived Ginsenoside Rh2 in Yeast via Repurposing a Key Promiscuous Microbial Enzyme. *Metab. Eng.* **2017**, *42*, 25-32.

(40) Wang, P.; Wei, W.; Ye, W.; Li, X.; Zhao, W.; Yang, C.; Li, C.; Yan, X.; Zhou, Z. Synthesizing Ginsenoside Rh2 in *Saccharomyces cerevisiae* Cell Factory at High-Efficiency. *Cell discovery* **2019**, *5*, 5.

(41) Hwang, J.-T.; Kim, S.-H.; Lee, M.-S.; Kim, S. H.; Yang, H.-J.; Kim, M.-J.; Kim, H.-S.; Ha, J.; Kim, M.; Kwon, D. Y. Anti-Obesity Effects of Ginsenoside Rh2 are Associated with the Activation of AMPK Signaling Pathway in 3T3-L1 Adipocyte. *Biochem. Biophys. Res. Commun.* **2007**, *364*, 1002-1008.

(42) Lin, W.-C.; Deng, J.-S.; Huang, S.-S.; Lin, W.-R.; Wu, S.-H.; Lin, H.-Y.; Huang, G.-J. Anti-Inflammatory Activity of Sanghuangporus Sanghuang by Suppressing the TLR4-mediated PI3K/AKT/mTOR/IKKβ Signaling Pathway. *RSC Adv.* **2017**, *7*, 21234-21251.

(43) Chen, W.; Qiu, Y. Ginsenoside Rh2 Targets EGFR by Up-Regulation of MiR-491 to Enhance Anti-Tumor Activity in Hepatitis B Virus-Related Hepatocellular Carcinoma. *Cell Biochem. Biophys.* **2015**, *72*, 325-331.

(44) King, J. L.; Jukes, T. H. Non-darwinian Evolution. *Science* **1969**, *164*, 788-798.

(45) Dyer, K. F. The Quiet Revolution: a New Synthesis of Biological Knowledge. *J. Biol. Educ.* **1971**, *5*, 15-24.

(46) Modarres, H. P.; Mofrad, M. R.; Sanati-Nezhad, A. Protein Thermostability Engineering. *RSC Adv.* **2016**, *6*, 115252-115270.

(47) Liu, Y.; Lee, J.Y.; Khare, M. Methods and Materials for Enzymatic Synthesis of Mogrolside Compounds. U.S. Patent 9,920,349, Mar 20, 2018.

(48) Nidetzky, B.; Gutmann, A.; Zhong, C. Leloir Glycosyltransferases as Biocatalysts for Chemical Production. *ACS Catal.* **2018**, *8*, 6283-6300.

(49) Li, G.; Yao, P; Gong, R; Li, J.; Liu, P.; Lonsdale, R.; Wu, Q.; Lin, J.; Zhu, D.; Reetz, M. T. Simultaneous Engineering of an Enzyme's Entrance Tunnel and Active Site: the Case of Monoamine Oxidase MAO-N. *Chem. Sci*. **2017**, *8*, 4093-4099.

(50) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 198-210.

(51) Gantt, R. W.; Peltier-Pain, P.; Cournoyer, W. J.; Thorson, J. S. Using Simple Donors to Drive the Equilibria of Glycosyltransferase-Catalyzed Reactions. *Nat. Chem. Biol*. **2011**, *7*, 685.

(52) Otwinowski, Z.; Minor, W. Processing of X-Ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol*. **1997**, *276*, 307-326.

(53) Keegan, R. M.; Winn, M. D. MrBUMP: an Automated Pipeline for Molecular Replacement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2008**, *64*, 119-124.

(54) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, J. R.; Vagin, A.; Wilson, K. S. Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 235-242.

(55) Cowtan, K. The Buccaneer Software for Automated Model Building. 1. Tracing Protein Chains. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 1002-1011.

(56) Cohen, S. X.; Jelloul, M. B.; Long, F.; Vagin, A.; Knipscheer, P.; Lebbink, J.; Sixma, T. K.; Lamzin, V. S.; Murshudov, G. N.; Perrakis, A. ARP/wARP and Molecular Replacement: the Next Generation. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2008**, *64*, 49-60.

(57) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 355-367.

(58) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126-2132.

(59) Brünger, A. T. Assessment of Phase Accuracy by Cross Validation: the Free R Value. Methods and Applications. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1993**, *49*, 24-36.

(60) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr*. **2007**, *40*, 658-674.

(61) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K. Glide: a New Approach for Rapid,

Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739-1749.

(62) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem*. **2000**, *21*, 132-146.

(63) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput*. **2015**, *11*, 3696-3713.

(64) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem*. **2004**, *25*, 1157-1174.

(65) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys*. **1977**, *23*, 327-341.

(66) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.***1995**, *103*, 8577-8593.

(67) Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M. PLIP: Fully Automated Protein–Ligand Interaction Profiler. *Nucleic Acids Res*. **2015**, *43*, W443-W447.

673 **Table of Contents**

674