Synthetic and Systems Biotechnology 7 (2022) 621–630

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: www.keaipublishing.com/en/journals/synthetic-and-systems-biotechnology

Exploring the catalytic function and active sites of a novel *C*-glycosyltransferase from *Anemarrhena asphodeloides*

Jia Huang [a,1], Yaru She [a,1], Jingyang Yue [a], Yidu Chen [a], Yu Li [a], Jing Li [a], Yonger Hu [a], Deying Yang [a], Jiabo Chen [a], Lu Yang [a], Zhongqiu Liu [a,***], Ruibo Wu [b], Pengfei Jin [c,**], Lixin Duan [a,*]

[a] *Joint Laboratory for Translational Cancer Research of Chinese Medicine of the Ministry of Education of the People's Republic of China, International Institute for Translational Chinese Medicine, Guangzhou University of Chinese Medicine, Guangzhou, 510006, PR China*
[b] *Guangdong Provincial Key Laboratory of New Drug Design and Evaluation, School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou, 510006, PR China*
[c] *Department of Pharmacy, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing Key Laboratory of Assessment of Clinical Drugs Risk and Individual Application (Beijing Hospital), Beijing, 100730, PR China*

ARTICLE INFO

Keywords:
C-glycosides
Anemarrhena asphodeloides
C-glycosyltransferases
Benzophenone
Active sites

ABSTRACT

Anemarrhena asphodeloides is an immensely popular medicinal herb in China, which contains an abundant of mangiferin. As an important bioactive xanthone *C*-glycoside, mangiferin possesses a variety of pharmacological activities and is derived from the cyclization reaction of a benzophenone *C*-glycoside (maclurin). Biosynthetically, *C*-glycosyltransferases are critical for the formation of benzophenone *C*-glycosides. However, the benzophenone *C*-glycosyltransferases from *Anemarrhena asphodeloides* have not been discovered. Herein, a promiscuous *C*-glycosyltransferase (AaCGT) was identified from *Anemarrhena asphodeloides*. It was able to catalyze efficiently mono-*C*-glycosylation of benzophenone, together with di-*C*-glycosylation of dihydrochalcone. It also exhibited the weak *O*-glycosylation or potent *S*-glycosylation capacities toward 12 other types of flavonoid scaffolds and a simple aromatic compound with –SH group. Homology modeling and mutagenesis experiments revealed that the glycosylation reaction of AaCGT was initiated by the conserved residue H23 as the catalytic base. Three critical residues H356, W359 and D380 were involved in the recognition of sugar donor through hydrogen-bonding interactions. In particular, the double mutant of F94W/L378M led to an unexpected enzymatic conversion of mono-*C*- to di-*C*-glycosylation. This study highlights the important value of AaCGT as a potential biocatalyst for efficiently synthesizing high-value *C*-glycosides.

1. Introduction

Herbs and herbal extracts are an important class of medicinal resources, which are often used to treat and prevent a wide range of diseases [1]. *Anemarrhena asphodeloides,* known as the only perennial plant belonging to the Anemarrhena genus of the monocotyledonous Liliaceae family, is one of the most popular medicinal herb in China. The dried rhizome of *A. asphodeloides* is a traditional Chineses herbal medicine, called 'Zhimu', and often used for the treatment of febrile diseases, fever, cough, and diabetes [2]. In addition, the herbal extracts of *A. asphodeloides* are acknowledged as a potential drug candidate for regulating mood due to its beneficial effects on central nervous system [3].

Flavonoid *C*-glycosides are a massive group of bioactive secondary metabolites with sugar moieties directly attached to the flavonoid backbones [4–6]. The rigid C–C bonds of flavonoid *C*-glycosides, featuring the distinguished stability against hydrolysis, have drawn considerable scientific interest for their significant value in pharmaceutical application [7,8]. In *A. asphodeloides*, an abundant of bioactive flavonoid *C*-glycosides, such as mangiferin and neomangiferin, have

Peer review under responsibility of KeAi Communications Co., Ltd.
* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: liuzq@gzucm.edu.cn (Z. Liu), jinpengfei3674@bjhmoh.cn (P. Jin), nlizn@gzucm.edu.cn (L. Duan).
[1] These authors contributed equally to this article.

https://doi.org/10.1016/j.synbio.2022.01.003
Received 12 November 2021; Received in revised form 28 December 2021; Accepted 6 January 2022

been isolated from its rhizomes [9]. Among these, mangiferin is one of the main bioactive ingredients, which has been reported to exhibit various pharmacological activities, including anti-inflammatory [10], antivirus [11], anti-cardiac vascular diseases [12], and antitumor activities [13]. Different from common flavonoid glycosides, mangiferin belongs to a canonical xanthone *C*-glycoside, which is produced from the cyclization reaction of a benzophenone *C*-glycoside in plant secondary metabolism and rare in monocotyledons [14].

Biosynthetically, benzophenone *C*-glycosides are derived from the catalysis of *C*-glycosyltransferases (CGTs). By far, with isoflavone, flavone, 2-hydroxyflavanone, chalcone and anthraquinone as substrates, plentiful CGTs have been identified from various organisms, such as plants [15], bacteria [16] and insect [17], and some significant progresses have been achieved in the field of *C*-glycosylation of natural and unnatural substrates [18]. For example, TcCGT1, identified from *Trollius chinensis,* was the first CGT that could catalyze 8-*C*-glycosylation of flavones [19]; GgCGT, identified from *Glycyrrhiza glabra*, was a powerful CGT that could efficiently catalyze di-*C*-glycosylation of flopropione-containing substrates [20]. However, the studies about benzophenone CGTs are extremely limited. At present, only two benzophenone CGTs have been characterized from plants, and both of them are from the same dicotyledon [4]. MiCGT was the first benzophenone CGT identified from *Mangifera indica*, which could efficiently convert various natural and unnatural benzophenone scaffolds to the corresponding mono-*C*-glycosides [14]. MiCGTb, the other efficient benzophenone CGT from *Mangifera indica*, also showed potent catalytic activities toward various benzophenone aglycones [21]. Besides, three other plant CGTs, such as FcCGT from *Fortunella crassifolia* [22], CuCGT from *Citrus unshiu* [22]and AbCGT from *Aloe barbadensis* [23], also displayed the *C*-glycosylation activities toward the benzophenone aglycone (maclurin) due to their possessing substrate promiscuities, but they are not generally recognized as benzophenone CGTs for the reason that maclurin is not their endogenous substrate, and the conversion rates that FcCGT and CuCGT catalyzed *C*-glycosylation of maclurin were rather low [22]. It is particularly noteworthy that the active sites for *C*-glycosylation of benzophenone CGTs, especially for determining interconversion of diverse types of glycosides, remain elusive.

A. asphodeloides contains abundant mangiferin, indicating that there exist benzophenone CGTs in this plant. However, the corresponding CGTs from *A. asphodeloides* have not yet been discovered and characterized. Herein, with the aim to identify the desired benzophenone CGTs, the transcriptomic data of *A. asphodeloides* were investigated. A candidate benzophenone CGT gene was screened out for cloning and heterologous expression in *Escherichia coli*. The benzophenone CGT enzyme, designated AaCGT, was functionally characterized through enzymatic assays *in vitro*. The biochemical properties and the substrate promiscuity of AaCGT were also explored. Moreover, molecular docking and site-directed mutagenesis were conducted to further investigate the active sites for *C*-glycosylation, sugar donor recognition and mono-*C*- or di-*C*-glycosylation selectivity of AaCGT. This study provides insights into the benzophenone CGTs from monocotyledons and provide a potential route for the directed synthesis of bioactive *C*-glycosides in drug discovery.

2. Material and methods

2.1. Chemicals, reagents and general methods

The standard compound **2** was a gift from associate professor Peng Wu and the standard compound **4** was obtained from Jie Zhang, Guangzhou University of Chinese Medicine. The other standards were purchased from Sichuan Weikeqi Bio-Technology CO. LTD. (Chengdu, China) and Chengdu Must Bio-Technology Co. Ltd (Chengdu, China). Chemical reagents were obtained from Thermo Fisher Scientific Co. Ltd. (Shanghai, China), and Sigma-Aldrich (St. Louis, MO, USA). UDP-glucose, UDP-rhamnose, UDP-galactose were purchased from Tianjin Heowns Bio-Technology Co. Ltd. (Tianjin, China). Glycosylated products were analyzed using an Agilent 1290 Series UHPLC system (LC-MS) coupled with a hybrid quadrupole time of flight (Q-TOF) mass spectrometer (6540, Agilent Technologies, Inc. CA, USA) equipped with a dual jet stream electrospray ionization (ESI) source. For quantitative analysis, 2 μL of a filtered sample was sent to a SB-C18 RRHD reversed-phase column (2.1 × 100 mm, 1.8 μm; Agilent) with the system operated in positive or negative ion mode at the flow rate 0.4 mL/min using solvent A (water with 0.1% formic acid), and solvent B (acetonitrile). The exact gradient programs were shown in Table S4. An optimal MS and MS^2 signal information was acquired in the following instrument and scan source parameters: full scan range, 50–1600 *m/z*; gas temperature, 320 °C; skimmer voltage, 65 V; sheath gas flow, 11 L/min; nebulizer, 35 psi; capillary voltage, 3,500 V; fragmentor, 130 V. The glycosylated products were determined by comparing with the retention times, ESI-MS and ESI-MS^2 of reference standards or identified by the characteristic MS^2 fragment. Conversion rates between glycosylated products and substrates were calculated based on their peak areas obtained from LC-MS.

2.2. Plant materials, RNA extraction, transcriptome sequencing and phylogenetic analysis

A. asphodeloides was collected from the Beijing Botanical Garden (China) while *Mangifera indica and Glycyrrhiza glabra* L were obtained from the garden of School of Pharmaceutical Sciences, Guangzhou University of Chinese Medicine. All of them were cultivated in the authors' laboratory. The RNA extraction was conducted using the Quick RNA Isolation Kit (Waryong Biotech. Beijing, China). The RNA samples from leaves, short-stems and roots of *A. asphodeloides* were sent to perform transcriptome sequencing using Illumina HiSeq™ 2000 platform (Biomarker Technologies, Beijing, China). The transcriptome data was assembled on the default parameters using Trinity[2] software [24]. All the unigene sequences were functionally annotated in the public databases, including NCBI non-redundant protein database (Nr), Swiss-Prot protein database, Kyoto Encyclopedia of Genes and Genomes (KEGG) database, Gene Ontology (GO) database and Clusters of Orthologous Groups (COG) database (E-value$<10^{-5}$). A phylogenetic tree was constructed by MEGA 7.0 software with neighbor-joining method after multiple alignments of protein sequences performed by ClustalW. The bootstrap value was set 1,000 replicates by default.

2.3. Molecular cloning, heterologous expression and purification of recombinant proteins

AaCGT was screened out through sequences analysis and cluster in phylogenetic tree though lacked 3′-cDNA ends. The incomplete sequence of *AaCGT* was amplified by PCR using the SMARTer® RACE 3′ Kit (Clontech, USA). A pair of nested gene-specific primers were applied to PCR for obtaining the 3′-cDNA ends of *AaCGT*. The amplified fragments were ligated into the pEASY-Blunt vector (Transgen, Beijing, China) and verified by sequencing. The full-length cDNA sequence of *AaCGT* was amplified by PCR using gene-specific primer pairs and cloned into pET-32a (+) vector (Invitrogen, US) using the ClonExpress® II One Step Cloning Kit (Vazyme Biotech, China). The nucleotide sequences of *MiCGT* and *GgCGT* were downloaded from GenBank Search database. The full-length coding sequences of *MiCGT* and *GgCGT* were respectively reverse-transcribed to cDNA with the PrimeScript™ II 1st Strand cDNA Synthesis Kit (TaKaRa Biotech, Dalian, China). The resulting cDNA was amplified by PCR using Phanta MAX Super-Fidelity DNA Polymerase (Vazyme Biotech, Nanjing, China) with gene-specific primer pairs. The PCR product of *MiCGT* was inserted into pET-32a (+) vector (Invitrogen, US) while that of *GgCGT* was ligated into pET-28a (+) vector (Invitrogen, US) using the ClonExpress® II One Step Cloning Kit (Vazyme Biotech, China). The full-length ORFs of *AaCGT, MiCGT* and *GgCGT* were individually validated by sequencing. The

resulting recombinant plasmids were transformed into *E.coli* BL21 (DE3) (Transgen Biotech, China) and transformants harboring the desired plasmid were cultured in 5 mL Luria-Bertani (LB) medium (100 μg mL^{-1} ampicillin) for 6–8 h at 37 °C with shaking at 220 rpm. The cells were inoculated into the 500 mL LB medium (100 μg mL^{-1} ampicillin) and cultured until an optical density (OD) at 600 nm ranged from 0.4 to 0.6. The cultures were then induced with 0.1 mM isopropyl β-D-1-thio-galactopyranoside (IPTG) at 18 °C for 18 h with shaking at 150 rpm. The cell cultures were harvested by centrifugation at 4500 rpm for 20 min at 4 °C, and then resuspended in 15 mL lysis buffer (50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, pH 8.0). A final concentration of 1 mg mL^{-1} lysozyme was incubated with resuspension for 30 min on ice, followed by ultrasonic decomposition for 12 min in an ice bath. The soluble supernatants were incubated with Ni-NTA resin (TaKaRa Biotech, Dalian, China) for 1 h at 4 °C after centrifugation at 14000 rpm for 30 min at 4 °C. The mixture was applied to a Superflow column and eluted with a gradient of imidazole (50, 80, 150, 250, and 300 mM). The proteins concentrated by Amicon Ultra-30K (Merck Millipore, Germany) were stored at −80 °C.

2.4. Effects of pH, temperature and divalent metal ions of AaCGT

To study the optimal reaction pH for AaCGT activity, the enzymatic reactions were performed in various reaction buffers ranged in pH values from 4.0 to 5.0 (citric acid-sodium citrate buffer), 6.0–8.0 (Tris-HCl buffer) and 9.0–11.0 (Na_2CO_3–$NaHCO_3$ buffer). To determine the optimal reaction temperature for AaCGT activity, the enzymatic reactions were carried out at different temperatures (4, 25, 30, 37, 45, 60 °C). To investigate the dependence of divalent metal ions for AaCGT activity, a final concentration of 5 mM $CaCl_2$, $BaCl_2$, $MgCl_2$, $ZnCl_2$, and EDTA were added into the reaction mixtures individually. All the reactions were terminated with an equal volume of ice-cold methanol and centrifuged at 12,000 rpm for 30 min. The supernatants were then analyzed by LC-MS. The above reactions were conducted with maclurin as a glycosylated acceptor and UDP-glucose as a sugar donor and performed in triplicate.

2.5. Determination of kinetic parameters of AaCGT

To determine the kinetic parameters of AaCGT, enzymatic assays were performed in a final volume of 200 μL reaction mixture, containing 50 mM Tris-HCl buffer (pH 7.0), 10 μg of purified AaCGT, 500 μM of saturated UDP-Glc, and varying concentrations (30, 50, 80, 100, 150, 200, 250, 300 μM) of Maclurin. The above reactions were conducted at 45 °C for 30 min, and quenched with an equal volume of ice-cold methanol. The supernatants were collected by centrifugation at 12,000 rpm for 30 min and further analyzed by LC-MS. Three parallel reactions were performed for quantifying glycosylated products of AaCGT. The value of K_m was calculated using Michaelis-Menten plot method.

2.6. Glycosyltransferase activity assays

All reactions containing 50 mM Tris-HCl buffer (pH 8.0), 20 μg of purified enzyme, 100 μM aglycone and 500 μM UDP-Glc were individually performed in a final volume of 200 μL. The reactions were carried out at 37 °C for 2 h and terminated with an equal volume of ice-cold methanol. Samples were centrifuged at 12,000 rpm for 30 min to obtain supernatants for further analysis through LC-MS. The enzymes heated at 99 °C for 12 min were used as the negative control. All the assays were conducted in triplicate. UDP-galactose (UDP-Gal) and UDP-rhamnose (UDP-Rha) were also tested to explore the sugar donor selectivity of AaCGT with maclurin, phloretin and norathyriol used as acceptors. All the assays were performed and analyzed as mentioned above.

2.7. Molecular modeling and molecular docking

The protein models of AaCGT, AbCGT and their sugar donors were established by SWISS-MODEL [25]. The crystal structure of GgCGT (PDBID: 6L5R), sharing 46.68% sequential identity with AaCGT and having the highest global model quality estimation (GMQE) value (0.80), was chosen as a template for molecular docking using Glide [26] (Schrödinger LLC, New York). Maclurin and phloretin were respectively docked into the binding pockets of AaCGT or AbCGT and the best scoring models were accepted based on ranking given by the software.

2.8. Classical MD simulation

The MD simulation was performed with the previous methods from our laboratory as the reference [27]. The corresponding parameters of the ligand in the force field was generated using the general AMBER force field (GAFF) [28], coupled with the Amber FF14SB force field [29] employed for AaCGT and the TIP3P model used for water molecules [30]. The restrained electrostatic potential (RESP) [31] charge was applied to establish the partial atomic charge of sugar acceptors, with the HF/6-31G* calculation through the Gaussian 09 package [32]. During our generating original coordinates and topology files process through the tleap program in AMBER12 [33], cubic models were employed to define the periodic boundary condition for operating the MD simulations. To relax the solvent and protein-ligand complex, the routine should be minimized with all the solute atoms constrained, followed by protein backbone and finally all of them restored to a stable state with no constraint. The program of heating each system was carried out ranging from 0 to 300 K under the NVT ensemble for 100 ps, and then equilibrated using NPT ensemble MD simulations at 300 K for 100 ps with the target pressure of 1.0 atm. The final program ended with a 100ns MD simulation for each system operating under the NPT ensemble with no constraint. The high-frequency stretching vibration of covalent bonds involving hydrogen atoms was constrained using the SHAKE algorithm integration [34], and a cutoff of 12 Å was set for the nonbonded interaction while the long-range electrostatic interactions dealt with the particle mesh Ewald algorithm [35]. MD simulation contained a 2 fs time step with the frames saved every 5000 steps for analysis.

2.9. Construction of AaCGT mutants and enzyme activity assays

The desired mutants were constructed through site-directed mutagenesis of AaCGT with the wild-type expression plasmid used as the template for PCR. Phanta MAX Super-Fidelity DNA Polymerase (Vazyme Biotech, China) was applied to PCR with the degenerate primers presented in Table S2. The exact PCR conditions are as follows: 95 °C for 5 min, followed by 20 cycles of 95 °C for 30 s, 57 °C for 30 s, and 72 °C for 4 min 30 s, end with 72 °C for 10 min. The restriction enzyme DpnI (Thermo Scientific, Shanghai, China) was used to digest for acquiring the correct mutants. The mutants verified by sequencing were transformed into *E.coli* BL21 (DE3) (Transgen Biotech, China) for heterologous expression and purification of recombinant proteins as described above. Enzyme activity assays of AaCGT mutants were performed with the same method as that of the wild-type enzyme.

3. Results and discussion

3.1. Molecular cloning and phylogenetic analysis of AaCGT

In this study, a BLAST search was carried out for mining candidate CGT genes from the transcriptome data of *A. asphodeloides*. Seven functionally-identified plant CGT protein sequences were used as a query (Table S2), such as OsCGT (*Oryza sativa*), FeUGT708C1 (*Fagopyrum esculentum*), GgCGT (*Glycyrrhiza glabra*), MiCGT (*Mangifera indica*), MiCGTb (*Mangifera indica*), PlUGT43 (*Pueraria lobata*), and TcCGT1 (*Trollius chinensis*). Two unigenes (universal genes, *CL650.*

Contig1 and *Unigene8580*) annotated as CGTs based on the public databases were preliminarily screened out even though *CL650.Contig1* lacked a short 3'-cDNA sequence. Only CL650.Contig1 showed glycosyltransferase activity in the following assays and shared the conserved PSPG box sequence with previously reported CGTs (Fig. S1). Thus, the potential *CGT* gene was tentatively defined as *AaCGT*. The full-length open reading frame (ORF) of *AaCGT* (1386 bp, GenBank accession number MZ358121), encoding 461 amino acids, was obtained through 3′-RACE.

The neighbor-joining phylogenetic tree was constructed for assessing the evolutionary relationship of AaCGT with known CGTs and canonical flavonoid OGTs (*O*-glycosyltransferases). Bacterial CGTs were also added into the phylogenetic tree for further description about the evolutionary origins of CGTs in different species. Intriguingly, AaCGT was grouped into the clade of monocotyledon CGTs in the UGT708-family CGTs (Fig. 1), which are an orthologous group of enzymes that catalyze the *C*-glycosylation using the signature substrates as the glycosyl acceptors, such as chalcones, closed-circular or open-circular 2-hydroxyflavanones, or a 2,4,6-trihydroxybenzophenone-like core structure. Dicotyledon CGTs, as the paralogous genes of UGT708-family members, were another predominant clade of UGT708-family CGTs and shared the similar substrates with monocotyledon CGTs. Notably, the two classes of enzymes, representing two independent gene duplication events, were descended from the common ancestral CGT gene, and AaCGT may share the similar *C*-glycosylation capacity with them. TcCGT1 and PlUGT43 were clustered into a clade of the flavonoid 8C CGTs, while WjGT1 and GtUF6CGT1 belonged to a cluster of the flavonoid 6C CGTs. These distinctive CGTs share a direct *C*-glycosylation capability toward flavone or isoflavone scaffolds, clearly different from UGT708-family CGTs in the pattern of glycosylation. In addition, bacterial CGTs were phylogenetically divergent from plant CGTs. Likewise, flavonoid OGTs were also distinct from plant CGTs, and could be divided into three independent clades, including 3-OH, 5-OH and 7-OH OGTs. These phylogenetic results indicate that plant CGTs and bacterial CGTs evolved from different ancestral CGT genes independently and in parallel with flavonoid OGTs [36].

3.2. Heterologous expression and functional characterization of AaCGT

The full-length ORF of AaCGT was cloned into a pET32a(+) vector for expression in the *E. coli* BL21(DE3) strain. The resulting recombinant AaCGT contained fused Trx•tag, S•tag and 6 × His• tag and the molecular weight was thus assigned as 66.24 kDa, in line with what was presented by SDS-PAGE. Ni-NTA affinity chromatography was used to earn the purified AaCGT (Fig. S2). The glycosylation capability of AaCGT was characterized by enzymatic activity assay *in vitro* with maclurin (**1**), a proposed biosynthetic precursor of mangiferin, used as the acceptor and UDP-glucose (UDP-Glc) as a sugar donor. The detecting reaction mixture (50 mM pH 8.0 Tris-HCl, 100 μM aglycones, 500 μM UDP-Glc and 20 μg of purified recombinant AaCGT) was further analyzed by liquid chromatography-mass spectrometry (LC-MS) after coincubating at 37 °C for 2 h. Convincingly, AaCGT could completely

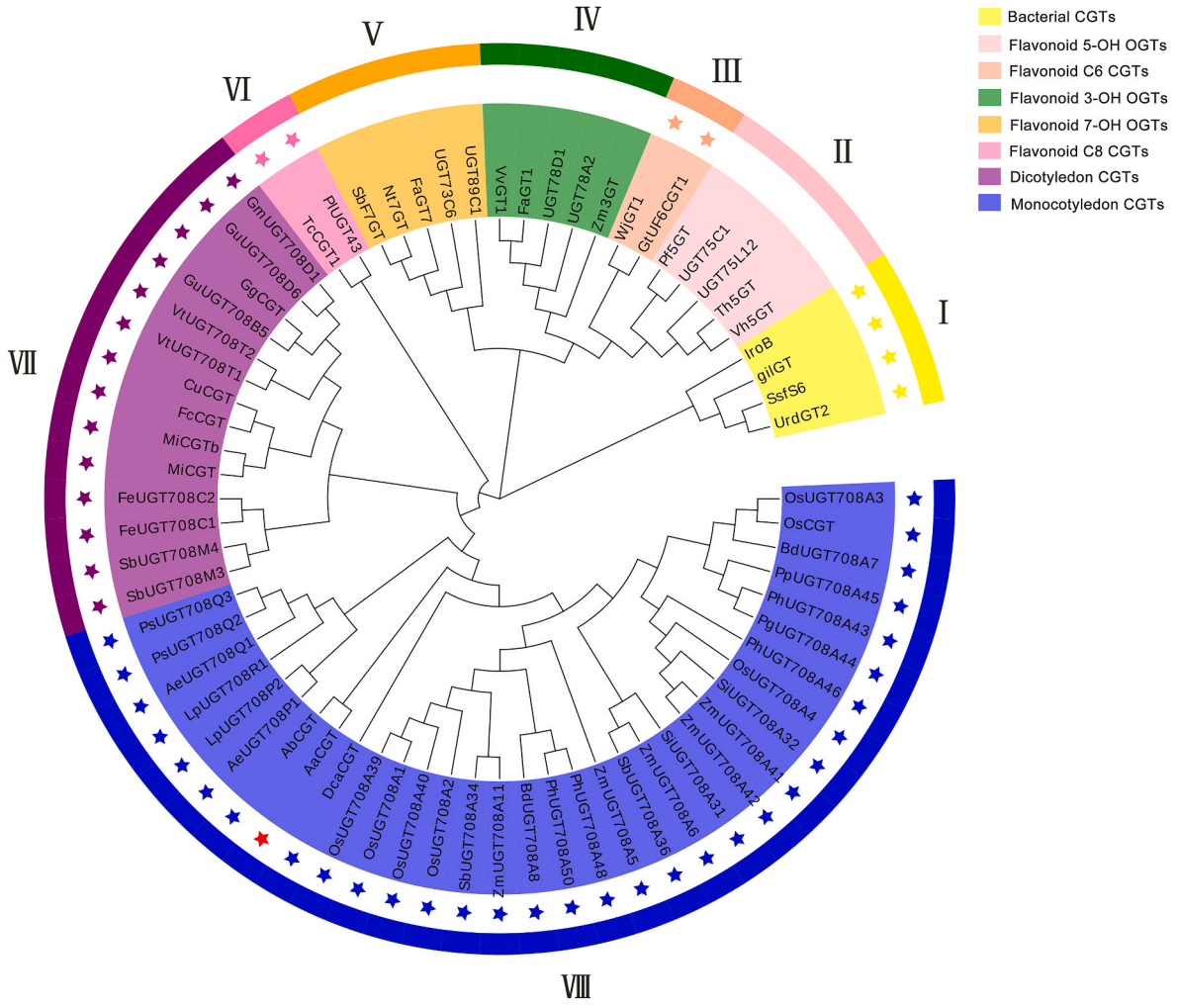

Fig. 1. Phylogenetic analyses of glycosyltransferases. AaCGT and other 69 glycosyltransferases, including the reported CGTs from plants and bacteria and the canonical OGTs from plants, are shown in the neighbor-joining tree. AaCGT is labeled with the red star.

converted it into a new product **1a** (Fig. 2B). The mass spectrum of **1a** showed its $[M-H]^-$ ion at *m/z* 423.0933, which was 162 amu more than that of **1**. The characteristic MS/MS fragment ions of **1a** appeared at *m/z* 303.0522 $[M-H-120]^-$ and *m/z* 333.0615 $[M-H-90]^-$ (Fig. 2C), indicating that **1a** was a *C*-glycoside. The *C*-glycosylation can be distinguished from other types of glycosylation by the characteristic MS/MS fragment ions $[M-H-120]^-$ and $[M-H-90]^-$, while that of *O*-, *S*-, and *N*-glucoside is $[M-H-162]^-$ /$[M+H-162]^+$ [23]. The structure of product **1a** was further identified as 3-*C*-glucosylmaclurin by comparing with the glycosylated product of the reported enzyme MiCGT (Fig. 2A) [14]. Hence, AaCGT was a novel benzophenone CGT that might be involved in the biosynthesis of mangiferin in *A. asphodeloides* (Fig. S24). AaCGT was subsequently named as UGT708Z2 by the UGT Nomenclature Committee [37]. As far as we know, only MiCGT and MiCGTb from *Mangifera indica* were reported as benzophenone CGTs. They shared highly similar 90% sequence identity, but there are only 47% sequence identity between MiCGT and AaCGT. Moreover, AaCGT was the first benzophenone CGT characterized from monocotyledons. This discovery enriches the diversity of *C*-glycosyltransferases.

3.3. The biochemical properties and kinetic parameters of AaCGT

To screen out the optimum conditions for determining the kinetic parameters of AaCGT, the enzymatic reactions were performed with the compound **1** used as the acceptor and UDP-Glc as the sugar donor. The maximum activity of the resulting recombinant AaCGT appeared at pH 7.0 (Fig. 3A) and 45 °C in the Tris-HCl buffer (Fig. 3B). In spite of the catalytic activity of AaCGT slightly improved by Ba^{2+}, a majority of divalent metal ions decreased the enzymatic activity of AaCGT (Fig. 3C), convincing us that it could be tentatively categorized into the GT-B type enzyme, a class of enzymes that are independent of divalent cations. The kinetic parameters of AaCGT showed that the *K*m values for **1** was 139.9 μM (Fig. 3D).

3.4. The exploration of substrate promiscuity and sugar donor specificity for AaCGT

Plant CGTs are usually able to accept various natural and unnatural compounds as substrates. Thus, in order to further explore the substrate promiscuity of AaCGT, the other 21 structurally diverse substrates, including dihydrochalcone (**2**), xanthone (**4**), isoflavones (**6**, **7**, **16**), flavones (**5**, **15**), flavanones (**8**–**10**), flavonols (**11**–**14**), anthraquinone (**17**), steroid sapogenin (**19**, **20**), triterpenoid (**21**, **22**) and simple aromatic compound with –SH, or $–NH_2$ groups (**3**, **18**), were individually added into enzymatic activity assay with UDP-Glc as the sugar donor. LC-MS analysis showed that AaCGT could exhibit catalytic activities toward 13 substrates.

Unexpectedly, AaCGT was able to completely catalyze the substrate **2** into **2a** and **2aa**, the product **2a** of which was even close to 80% in the conversion rate (Fig. 4, A). The two characteristic MS/MS fragment ions of **2a** appeared at *m/z* 315.0872 $[M-H-120]^-$ and *m/z* 345.0977 $[M-H-90]^-$, which corresponded to nothofagin, a mono-*C*-glycoside intermediate produced by GgCGT (Fig. S3). In contrast, the product **2aa** shared the common MS/MS fragmentation pattern with 3',5'-di-*C*-glucosylphloretin, a final product resulted from the enzymatic assay of GgCGT with the compound **2** as the accepted substrate. The MS/MS fragmentation pattern features four signature MS/MS fragment ions, including *m/z* 357.0972 $[M-H-240]^-$, *m/z* 387.1068 $[M-H-210]^-$, *m/z* 417.1171 $[M-H-180]^-$ and *m/z* 477.1362 $[M-H-120]^-$ (Fig. S3), all of which were characteristic MS/MS fragments of 3',5'-di-*C*-glucosylphloretin [20,38]. Both the intermediate **2a** and the final product **2aa** exhibited the same retention time with the respective reference standards that were generated from GgCGT. A previous study reported that 3',5'-di-*C*-glucosylphloretin was derived from two-step glycosylation reaction that GgCGT catalyzed the compound **2** into nothofagin, followed by further *C*-glycosylating the intermediate, finally resulting in the production of the di-*C*-glycoside [20]. Hence, we speculated that AaCGT performed an

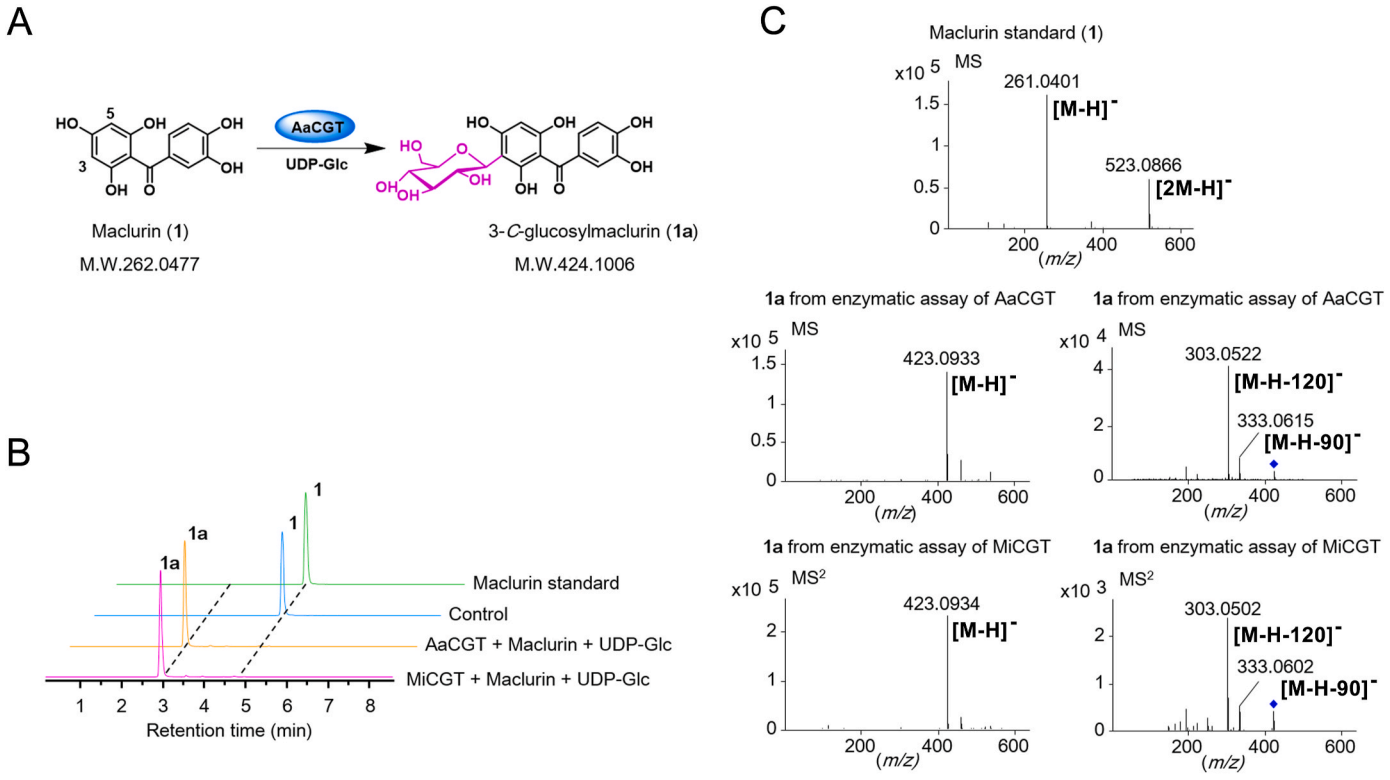

Fig. 2. *C*-glycosylation of **1** catalyzed by recombinant AaCGT. (A) AaCGT catalyzed *C*-glycosylation of **1** to form **1a**. (B) The selected ion chromatograms of **1** and the glycosylated product **1a** $[M-H]^-$. "Control" refers to the reaction with the addition of heat-inactivated AaCGT. (C) Typical negative ion MS and MS/MS spectra of the reference standard **1** and the product **1a.** AaCGT was inactivated by heating at 99 °C for 12 min. The enzymatic reactions were performed at 37 °C for 2 h.

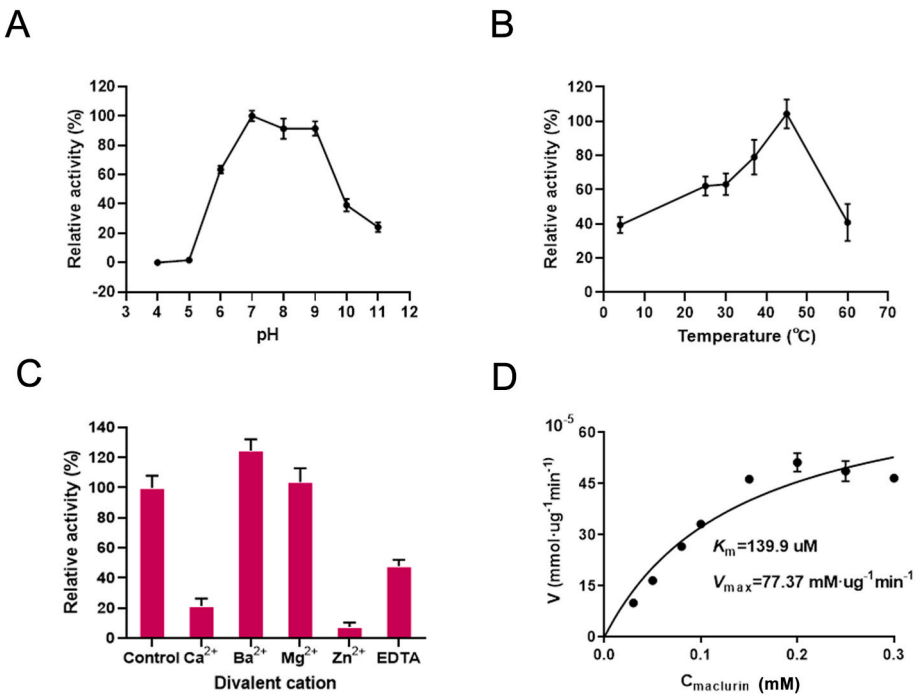

Fig. 3. The biochemical properties and kinetic parameters of recombinant AaCGT. (A) Effects of pH on AaCGT activities. (B) Effects of various temperature on AaCGT activities. (C) Dependence of AaCGT activities on various divalent metal ions. (D) Kinetic parameters of recombinant AaCGT. The enzymatic reactions were carried out with maclurin (**1**) as the acceptor and UDP-Glc as the sugar donor. The error bars of conversion rates refer to mean ± SD of three independent replicates (n = 3).

identical di-*C*-glucosylation of **2** based on the fact that AaCGT and GgCGT shared the identical MS/MS fragmentation pattern and retention time.

AaCGT also showed the robust promiscuity toward other structurally diverse flavonoid scaffolds excluding the compounds **15** and **16** (Fig. 4A, B). The signature MS/MS fragment ion of *O*-glycoside is $[M+H-162]^+$ (Figs. S5-S15), revealing that all of them were *O*-glycosides. In contrast to *C*-glycosides, *O*-glycosides were apparently low in the conversion rates, indicating that the closed C rings of *O*-glycosides played a critical role in the *C*- or *O*-glycosylation selectivity of AaCGT through hindering the binding of the enzyme and substrates.

In addition to flavonoid substrates, 3,4-dichlorobenzenethiol (**3**) could also be accepted by AaCGT. For *S*-glycosylation, the signature MS/MS fragment also appears at $[M+H-162]^+$ [23]. Based on ESI-MS analysis, together with the MS/MS spectrum that the product **3a** showed the characteristic MS/MS fragment ion at *m/z* 176.9330 $[M-H-162]^-$, AaCGT was found to be able to convert **3** into the corresponding *S*-glycoside at a close to 100% conversion rate (Fig. 4A, S4). In contrast, AaCGT was incapable to glycosylate 3,4-dichloroaniline (**18**). Besides, we also test the activities of AaCGT for anthraquinone, steroid sapogenin and triterpenoid, while no new product was observed, implying that AaCGT could not accept these metabolites derived from other biosynthesis pathways as substrates. Overall, these results indicated that AaCGT was a promiscuous benzophenone CGT possessing di-*C*-, *O*-, *S*-glycosylation activities.

Furthermore, we also investigated whether AaCGT could utilize other UDP-sugars as donors. When UDP-Gal and UDP-Rha used as the sugar donors, AaCGT was inactive to the substrates **1** and **2** (Fig. S16), as opposed to the above result that AaCGT could efficiently utilize UDP-Glc to produce corresponding glycosides (**1a**/**2a**/**2aa**) with **1** or **2** as the substrate. Remarkably, AaCGT could only accept UDP-Glc as a donor, indicating that it possessed sugar donor specificity.

3.5. The active sites for C-glycosylation of AaCGT

Recently, active sites for C-glycosylation of many 2-hydroxyflavanone CGTs have emerged as a hot issue, while active sites for *C*-glycosylation of benzophenone CGTs were reported rarely. Thus, in order to explore the key sites for *C*-glycosylation of AaCGT, we employed molecular modeling and molecular docking to construct the model for AaCGT/UDP-Glc/maclurin (**1**). In the structure model, both the substrate **1** and the sugar donor (UDPG) are located in the active pocket of AaCGT (Fig. 5A, B). In previous studies, a conserved His residue of N-terminal (10–30 amino acids) was reported to initiate the glycosylation reaction through activating deprotonation and facilitating nucleophilic attack at C-1′ of the sugar [20,39]. A similar residue H23 was observed in AaCGT based on the alignment of amino acid sequences (Fig. S1) and the subsequent docking results showed that the residue H23 was sterically close to 4-OH of a slightly bent maclurin (**1**) (Fig. 5A and B). Meanwhile, C-3 of **1** was proximal to the anomeric carbon of UDP-Glc. The mutation result of H23A exhibited that *C*-glycosylation activity of **1** was completely eliminated (Fig. 5C, S18), substantiating the reliability of this hypothesis. Different from our result, Hirade et al. [40] reported that the mutant of UGT708D1 gained the *O*-glycosylation when the active site H20 was mutated to Ala. H20 of UGT708D1 was therefore inferred as the key amino acid mediating enzymatic regioselectivity. It seems that the role of general base His in UGT708-family CGTs differed between 2-hydroxyflavanone and other flavonoid substrates [20,40]. D121 serving as another critical residue was close to H23 (Fig. 5B), whose role was to balance the charge of catalytic residue H23 after deprotonation [18]. The mutant D121A was also devoid of *C*-glycosylation, demonstrated whichever site of the His-Asp dyad was indispensable for the *C*-glycosylation reaction of AaCGT.

The *C*-glycosylation activities of UGT708-family CGTs are also subject to some conserved amino acids. For the reported UGT708C1, D96 and R280 stabilized the active center through forming a salt bridge [39]. The two amino acids were highly conserved in the UGT708-family CGTs and respectively mapped to D88 and R276 of AaCGT, the residue R276 of which was located near the acceptor (Fig. 5B). The long side chain of R276 was probably stimulative to the departure of the UDP molecule, assisting the His-Asp dyad to activate *C*-glycosylation of AaCGT [40]. Consistently, the inactivity of the mutants of D88A and R276A displayed

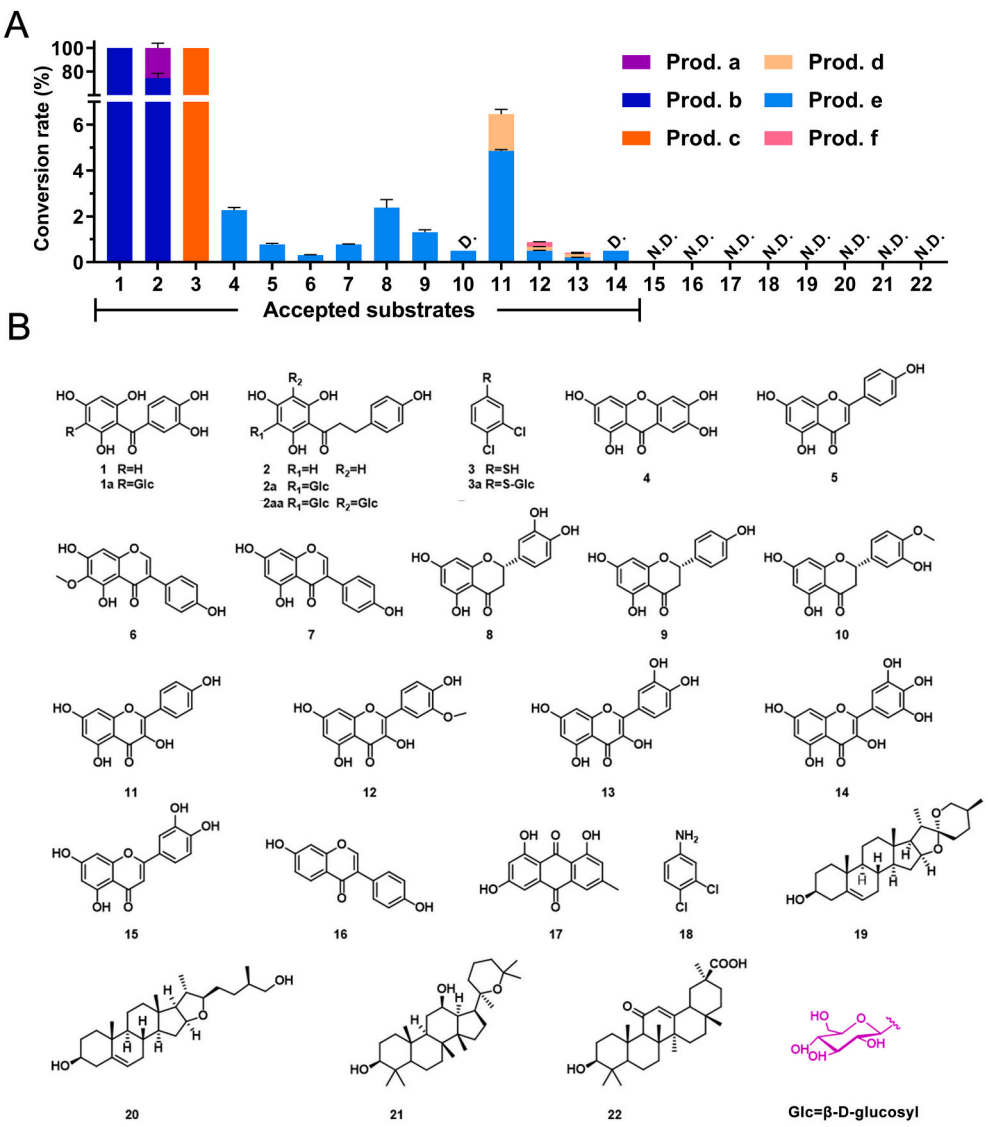

Fig. 4. Substrate promiscuity of recombinant AaCGT. (A) Conversion rates (%) of diverse substrates. Prod. a indicates conversion rate of the di-*C*-glycoside and Prod. b represents conversion rate of the mono-*C*-glycoside; Prod. c refers to *S*-glycoside while Prod. d, Prod. e and Prod. f refer to conversion rates of structurally diverse mono-*O*-glycosides. "D." means products detected in MS and MS/MS. "N.D." means that no products are detected. (B) Structures of various acceptors and glycosylated products for AaCGT. The enzymatic reactions were performed at 37 °C for 2 h. The error bars of conversion rates refer to mean ± SD of three independent replicates (n = 3).

their critical role in facilitating the enzymatic activity of AaCGT.

We also investigated the catalytic activities of these mutants toward the substrates phloretin (**2**) and norathyriol (**4**). For the substrate **2**, the catalytic activities of the mutants H23A and R276A were almost lost, while the D88A and D121A mutations abolished their glycosylation activities completely (Figs. S17A and S19). In addition, all of these mutants were incapable of catalyzing the *O*-glycosylation of **4** (Figs. S17B and S20). These results consolidated our viewpoint that the glycosylation activities of AaCGT are at the mercy of these crucial active sites, including H23, D88, D121 and R276.

3.6. The active sites for sugar donor recognition of AaCGT

The amino acids around the sugar moiety of donor are of great essential to enzymatic activity of CGTs. To explore the structural basis for sugar donor recognition of AaCGT, we analyzed the interactions between the corresponding residues of enzyme and the sugar moiety of UDP-Glc based on the model (Fig. 5D). The structural analysis revealed that the amidogen located at the terminal of W359 could directly form hydrogen bond with 4-OH group of the sugar moiety, while the residue H356 taking a far more sinuous path to interact with 2-OH group (Fig. 5E). Additionally, two hydrogen bonds could be observed when

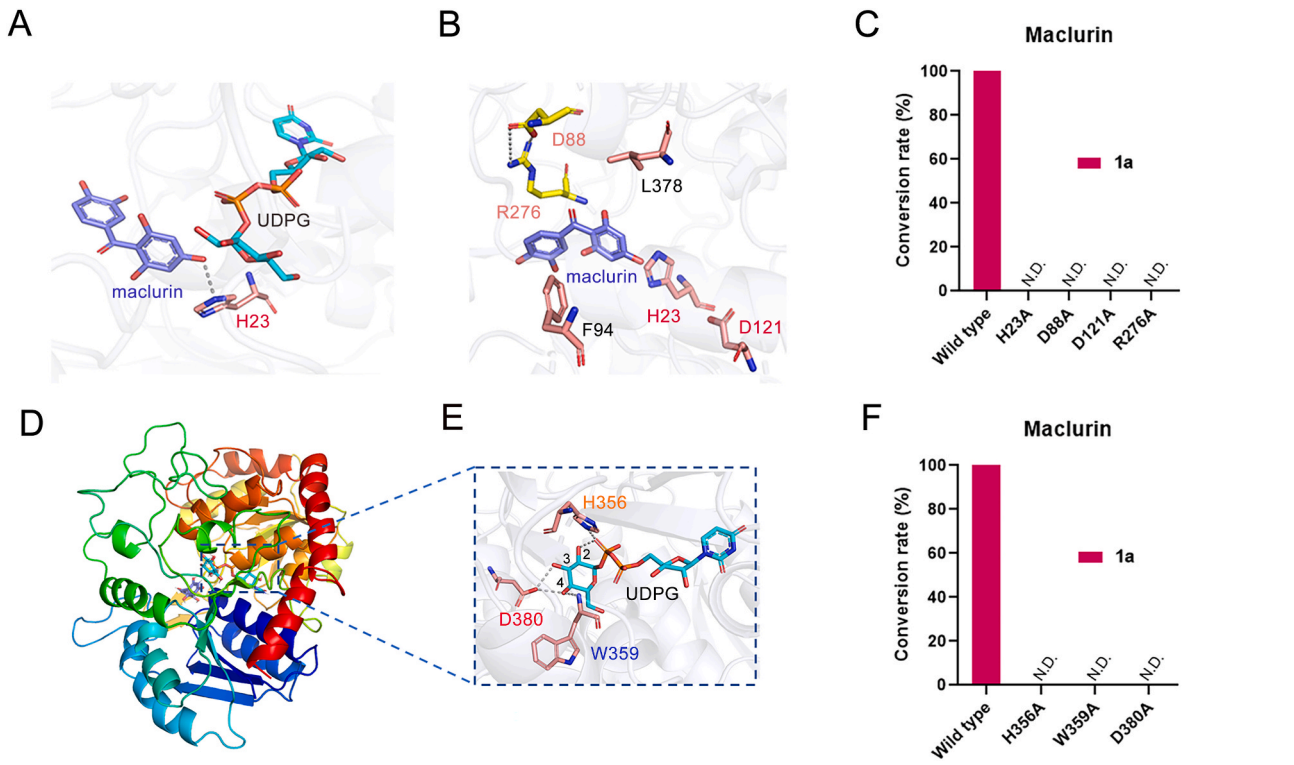

Fig. 5. The active sites for *C*-glycosylation of AaCGT and recognition of sugar donor. The key amino acids for *C*-glycosylation of AaCGT are shown in part (A) and (B). (C) Conversion rates of the wild type AaCGT and the mutants for *C*-glycosylation using **1** as the substrate. The key amino acids for recognition of sugar donor are shown in part (D) and (E). (F) Conversion rates of the wild type AaCGT and the mutants for recognition of sugar donor using **1** as the substrate. The enzymatic reactions were performed at 37 °C for 2 h. The error bars of conversion rates refer to mean ± SD of three independent replicates (n = 3).

D380 interacting with both 3-OH and 4-OH, which would intensify the stability of UDP-Glc. This was similar to the prominent role of D390 uncovered from GgCGT [20]. Our structural analysis were subsequently validated by the mutation results of H356A, W359A and D380A that all of the mutants abolished thoroughly the *C*-glycosylation activities toward **1** (Fig. 5F, S21). All of these indicated that the key residues around UDP-Glc stabilized sugar moiety through hydrogen-bonding interactions, rendering UDP-Glc recognized by AaCGT. Moreover, we also observed that these mutants were inactive toward **2** (Figs. S17C and S22). Notably, the amino acids H356, W359 and D380, highly conserved at the PSPG box in plant CGTs (Fig. S1), were important determinants for the *C*-glycosylation activity of AaCGT.

3.7. The active sites involved in the switching for mono-C- or di-C-glycosylation selectivity of AaCGT

Plant CGTs share similar protein structures but differ in catalytic activities. It is reported that both GgCGT and AbCGT showed robust di-*C*-glycosylation activities toward **2** (up to 100% conversion) [20,23], while AaCGT exhibited both mono-*C*- and di-*C*-glycosylation activities toward **2** according to our study, suggesting that some key amino acids around the acceptor-binding pocket should contribute to the mono-*C*- or di-*C*-glycosylation selectivity. In order to investigate the crucial active sites involved in this selectivity, molecular modeling and site-directed mutagenesis were performed. During the conducting molecular modeling process, we found that AbCGT shared 55.76% sequential similarity with AaCGT, higher than that of GgCGT and AaCGT (46.68%). Given both AaCGT and AbCGT were originated from monocot *liliaceae* family plants, and clustered into the same small clade in the monocotyledon CGTs (far away from the clade of GgCGT, Fig. 1), AbCGT was thus considered as a better control structure than GgCGT. Afterward, we constructed models for AaCGT/UDP-Glc/phloretin (**2**) and AbCGT/UDP-Glc/phloretin (**2**) by molecular docking. Molecular dynamic (MD) simulation showed that the binding modes of AaCGT/UDP-Glc/**2** and AbCGT/UDP-Glc/**2** were stable (RMSD ≤3.0 Å for the protein backbone, Figs. S25–26), based on which the structures of AaCGT/UDP-Glc/**2** and AbCGT/UDP-Glc/**2** were subsequently superimposed. The corresponding amino acid residues around their acceptor-binding pockets were highly coincident, except two residues, namely, the residues F94 (W93) and L378 (M385) of AaCGT (AbCGT) (Fig. 6A). The mutants of F94W, L378 M and F94W/L378 M for AaCGT were then constructed. The results demonstrated that the F94W mutant slightly improved the di-*C*-glycosylation activity of AaCGT, while the L378 M mutant weakened it a bit (Fig. 6B, S23). Moreover, the F94W/L378 M mutant surprisingly switched mono-*C*- to di-*C*-glycosylation. This delightful outcome may result from the conformational alteration around acceptor-binding pocket. However, works on crystal structure of AaCGT are still essential for dissection of this potential mechanism.

4. Conclusion

In summary, a promiscuous *C*-glycosyltransferase AaCGT was identified from monocot medicinal plant *A. asphodeloides* through functional analysis. AaCGT showed a highly efficient *C*-glycosylation ability toward benzophenone and dihydrochalcone, leading to the production of mono-*C*- or di-*C*-glycoside compound. AaCGT also exhibited the weak *O*-glycosylation or potent *S*-glycosylation capacities that 12 structurally diverse compounds could be accepted as substrates, including other types of flavonoid scaffolds and a simple aromatic compound with –SH group. Homology modeling and mutagenesis experiments revealed that the glycosylation reaction of AaCGT was initiated by the conserved residue H23 as the catalytic base and the resulting charge was accordingly balanced by D121. The residues D88 and D121 were critical for *C*-glycosylation activities of AaCGT. The recognition mechanism of sugar donor was mediated by three critical residues H356, W359 and D380

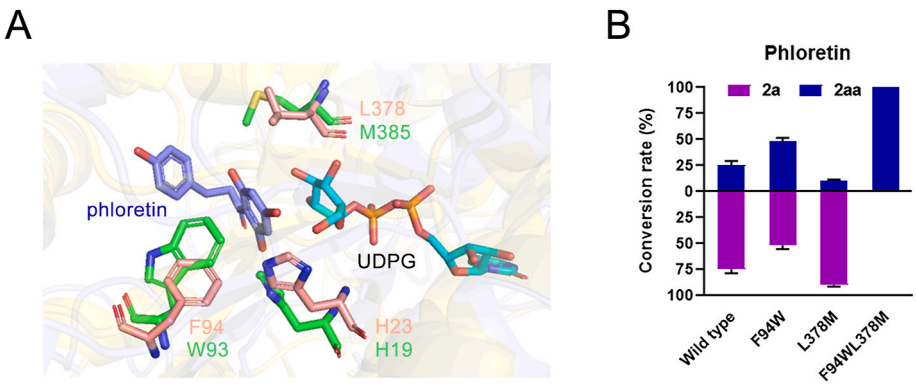

Fig. 6. Structural analysis and structure-guided mutagenesis of AaCGT. (A) The two differential residues and the catalytic bases (His23 and His19) located near acceptor-binding pockets of AaCGT (salmon) and AbCGT (green). (B) Percent conversion of the wild type AaCGT and the mutants, with **2** as the substrate. The enzymatic reactions were performed at 37 °C for 2 h. The error bars of conversion rates refer to mean ± SD of three independent replicates (n = 3).

through hydrogen-bonding interactions. More importantly, an unexpected enzymatic conversion of mono-*C*- to di-*C*-glycosylation was generated from the stronger interaction of acceptor and the mutational amino acids of F94W and L378 M. This study paves a way for efficiently synthesizing high-value *C*-glycosides utilizing CGTs as available biocatalysts. Further engineering based on crystal sources may produce diverse, even rare *C*-glycosides, bringing about their wider application in the field of drug discovery.

Data availability

All data supporting the findings of this work are available within this paper and the supplementary materials. All data is available from the authors upon reasonable request. The protein sequence of AaCGT (accession no. MZ358121) is deposited in GenBank. Source data are provided with this paper.

CRediT authorship contribution statement

Jia Huang: Conceptualization, Formal analysis, Investigation, Visualization, Writing – original draft. **Yaru She:** Software, Visualization, Data curation. **Jingyang Yue:** Visualization, Data curation. **Yidu Chen:** Investigation, Data curation. **Yu Li:** Investigation, Formal analysis. **Jing Li:** Investigation, Visualization. **Yonger Hu:** Validation, Data curation. **Deying Yang:** Formal analysis, Data curation. **Jiabo Chen:** Data curation. **Lu Yang:** Investigation. **Zhongqiu Liu:** Validation, Resources. **Ruibo Wu:** Resources. **Pengfei Jin:** Funding acquisition, Project administration. **Lixin Duan:** Conceptualization, Funding acquisition, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgments

We thank associate professor Peng Wu and Jie Zhang (Guangzhou University of Chinese Medicine, Guangzhou) for providing the available compounds, phloretin and norathyriol. This work was supported by the National Natural Science Foundation of China No.81874333 and the Guangdong Foundation for Basic and Applied Basic Research No.2020A1515010926.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2022.01.003.

References

[1] Rape M. Plant biology informs drug discovery. Nat Rev Mol Cell Biol 2014;15:501. https://doi.org/10.1038/nrm3842.
[2] Wang Y, Dan Y, Yang D, et al. The genus Anemarrhena Bunge: a review on ethnopharmacology, phytochemistry and pharmacology. J Ethnopharmacol 2014; 153:42–60. https://doi.org/10.1016/j.jep.2014.02.013.
[3] Cui Y, Dai YX, Feng J. Application of Timosaponin for treating anxiety. 2007. CN101002879A 20070725.
[4] Tegl G, Nidetzky B. Leloir glycosyltransferases of natural product *C*-glycosylation: structure, mechanism and specificity. Biochem Soc Trans 2020;48:1583–98. https://doi.org/10.1042/bst20191140.
[5] Kitamura K, Ando Y, Matsumoto T, et al. Total synthesis of aryl *C*-glucoside natural products: strategies and tactics. Chem Rev 2018;118:1495–598. https://doi.org/10.1021/acs.chemrev.7b00380.
[6] Dai LH, Hu Y, Chen CC, et al. Flavonoid *C*-glycosyltransferases: function, evolutionary relationship, catalytic mechanism and protein engineering. ChemBioEng Rev 2021;8:15–26. https://doi.org/10.1002/cben.202000009.
[7] Lalitha K, Muthusamy K, Prasad YS, et al. Recent developments in β-*C*-glucosides: synthesis and applications. Carbohydr Res 2015;402:158–71. https://doi.org/10.1016/j.carres.2014.10.008.
[8] Yang Y, Yu B. Recent advances in the chemical synthesis of *C*-glucosides. Chem Rev 2017;117:12281–356. https://doi.org/10.1021/acs.chemrev.7b00234.
[9] Xie YY, Wang XM, Wang SH, et al. Metabolism and pharmacokinetics of major polyphenol components in rat plasma after oral administration of total flavonoid tablet from *Anemarrhenae Rhizoma*. J Chromatogr B Analyt Technol Biomed Life Sci 2015;1026:134–44. https://doi.org/10.1016/j.jchromb.2015.12.003.
[10] Saha S, Mahalanobish S, Dutta S, et al. Mangiferin ameliorates collateral neuropathy in tBHP induced apoptotic nephropathy by inflammation mediated kidney to brain crosstalk. Food Funct 2019;10:5981–99. https://doi.org/10.1039/c9fo00329k.
[11] Rechenchoski DZ, Agostinho KF, Faccin-Galhardi LC, et al. Mangiferin: a promising natural xanthone from *Mangifera indica* for the control of acyclovir – resistant herpes simplex virus 1 infection. Bioorg Med Chem 2020;28:115304. https://doi.org/10.1016/j.bmc.2020.115304.
[12] Song J, Meng Y, Wang M, et al. Mangiferin activates Nrf2 to attenuate cardiac fibrosis via redistributing glutaminolysis-derived glutamate. Pharmacol Res 2020; 157:104845. https://doi.org/10.1016/j.phrs.2020.104845.
[13] Lemaire S, Houpis IN, Xiao T, et al. Stereoselective *C*-glycosylation reactions with arylzinc reagents. Org Lett 2012;14:1480–3. https://doi.org/10.1021/ol300220p.
[14] Chen DW, Chen R, Wang R, et al. Probing the catalytic promiscuity of a regio- and stereospecific *C*-glycosyltransferase from *Mangifera indica*. Angew Chem Int Ed 2015;54:12678–82. https://doi.org/10.1002/anie.201506505.
[15] Wang ZL, Gao HM, Wang S, et al. Dissection of the general two-step di-*C*-glycosylation pathway for the biosynthesis of (iso)schaftosides in higher plants. Proc Natl Acad Sci USA 2020;117:30816–23. https://doi.org/10.1073/pnas.2012745117.
[16] Wang F, Zhou M, Singh S, et al. Crystal structure of SsfS6, the putative *C*-glycosyltransferase involved in SF2575 biosynthesis. Proteins 2013;81:1277–82. https://doi.org/10.1002/prot.24289.
[17] Kannangara R, Siukstaite L, Borch-Jensen J, et al. Characterization of a membrane-bound *C*-glucosyltransferase responsible for carminic acid biosynthesis in *Dactylopius coccus Costa*. Nat Commun 2017;8:1987. https://doi.org/10.1038/s41467-017-02031-z.
[18] Putkaradze N, Teze D, Fredslund F, et al. Natural product *C*-glycosyltransferases – a scarcely characterised enzymatic activity with biotechnological potential. Nat Prod Rep 2021;38:432–43. https://doi.org/10.1039/d0np00040j.
[19] He JB, Zhao P, Hu ZM, et al. Molecular and structural characterization of a promiscuous *C*-glycosyltransferase from *Trollius chinensis*. Angew Chem Int Ed 2019;58:11513–20. https://doi.org/10.1002/anie.201905505.

[20] Zhang M, Li FD, Li K, et al. Functional characterization and structural basis of an efficient di-*C*-glycosyltransferase from *Glycyrrhiza glabra*. J Am Chem Soc 2020; 142:3506–12. https://doi.org/10.1021/jacs.9b12211.
[21] Chen DW, Sun L, Chen R, et al. Enzymatic synthesis of acylphloroglucinol 3-*C*-glucosides from 2-*O*-glucosides using a *C*-glycosyltransferase from *Man-gifera indica*. Chem Eur J 2016;22:5873–7. https://doi.org/10.1002/chem.201600411.
[22] Ito T, Fujimoto S, Suito F, et al. *C*-glycosyltransferases catalyzing the formation of di-*C*-glucosyl flavonoids in citrus plants. Plant J 2017;91:187–98. https://doi.org/10.1111/tpj.13555.
[23] Xie K, Zhang X, Sui S, et al. Exploring and applying the substrate promiscuity of a *C*-glycosyltransferase in the chemo-enzymatic synthesis of bioactive *C*-glucosides. Nat Commun 2020;11:5162. https://doi.org/10.1038/s41467-020-18990-9.
[24] Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 2011;29:644–52. https://doi.org/10.1038/nbt.1883.
[25] Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 2018;46:W296–303. https://doi.org/10.1093/nar/gky427.
[26] Friesner RA, Murphy RB, Repasky MP, et al. Extra precision glide:dockin-g and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem 2006;49:6177–96. https://doi.org/10.1021/jm051256o.
[27] Wu SD, Zhang F, Xiong W, Molnár I, et al. An unexpected oxidosqualene cyclase active site architecture in the *Iris tectorum* multifunctional α-Amyrin synthase. ACS Catal 2020;10:9515–20. https://doi.org/10.1021/acscatal.0c03231.
[28] Wang JM, Wolf RM, Caldwell JW, et al. Development and testing of a general amber force field. J Comput Chem 2004;25:1157–74. https://doi.org/10.1002/jcc.20035.
[29] Maier JA, Martinez C, Kasavajhala K, et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theor Comput 2015;11:3696–713. https://doi.org/10.1021/acs.jctc.5b00255.
[30] Jorgensen WL, Chandrasekhar J, Madura JD, et al. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–35. https://doi.org/10.1063/1.445869.
[31] Bayly CI, Cieplak P, Cornell WD, et al. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J Phys Chem 1993;97:10269–80. https://doi.org/10.1021/J100142A004.
[32] Frisch MJ, Trucks GW, Schlegel HB, et al. Gaussian, Inc. Wallingford CT. 2016.
[33] Case DAD, Cheatham TE, Simmerling CL, et al. AMBER12. San Francisco, CA: University of California; 2012.
[34] Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of alkanes. J Comput Phys 1977;23:327–41. https://doi.org/10.1016/0021-9991(77)90098-5.
[35] Darden T, York D, Pedersen L. Particle mesh Ewald: an N.log(N) methodfor Ewald sums in large systems. J Chem Phys 1993;98:10089–92. https://doi.org/10.1063/1.464397.
[36] Feng CY, Li SS, Taguchi G, et al. Enzymatic basis for stepwise *C*-glycosylation in the formation of flavonoid di-*C*-glucosides in sacred lotus (*Nelumbo nucifera Gaertn*). Plant J 2021;106:351–65. https://doi.org/10.1111/tpj.15168.
[37] Homepage of the UGT nomenclature committee. http://prime.vetmed.wsu.ed-u/resources/udp-glucuronsyltransferase-homepage. 021/06/09.
[38] Abad-García B, Garmón-Lobato S, Berrueta LA, et al. New features on the fragmentation and differentiation of *C*-glycosidic flavone isomers by positive electrospray ionization and triple quadrupolemass spectrometry. Rapid Commun Mass Spectrom 2008;22:1834–42. https://doi.org/10.1002/rcm.3560.
[39] Liu M, Wang D, Li Y, et al. Crystal structures of the *C*-glycosyltransferase UGT708C1 from Buckwheat provide insights into the mechanism of *C*-glycosylation. Plant Cell 2020;32:2917–31. https://doi.org/10.1105/tpc.20.00002.
[40] Hirade Y, Kotoku N, Terasaka K, et al. Identification and functional analysis of 2-hydroxyflavanone *C*-glucosyltransferase in soybean (*Glycine max*). FEBS Lett 2015; 589:1778–86. https://doi.org/10.1016/j.febslet.2015.05.010.