Sequence mining yields 18 phloretin C-glycosyltransferases from plants for the efficient biocatalytic synthesis of nothofagin and phloretin-di-C-glycoside

Putkaradze, Natalia; Gala, Valeria Della; Vaitkus, Dovydas; Teze, David; Welner, Ditte Hededam

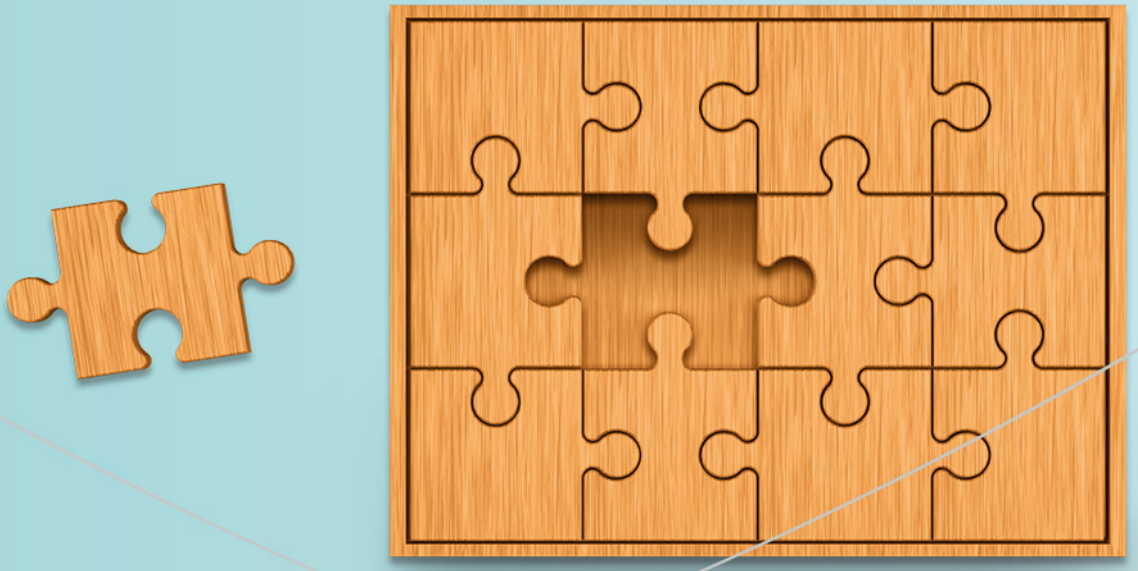
Published in:
Biotechnology Journal

Link to article, DOI:
10.1002/biot.202200609

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

SHIMADZU
Excellence in Science



Fits perfectly into your workflow

Due to automated sample preparation, our new CLAM-2040 makes analysis easier and more efficient than ever before – perfect also for less experienced users. You save time and effort and avoid errors resulting from a manual pretreatment. In addition, you get full flexibility because the CLAM-2040 can be combined with other devices like the LCMS-9050 Q-TOF and compatible with LIS & LAS (option).

Easy to use and to maintain:
reduced workload for operators.

Operational productivity:
automates all steps from pretreatment to result.

Full flexibility:
compatible with a variety of Shimadzu devices.

www.shimadzu.eu/perfect-workflow

Received: 5 December 2022 | Revised: 23 February 2023 | Accepted: 16 March 2023

DOI: 10.1002/biot.202200609

Biotechnology Journal

RESEARCH ARTICLE

Sequence mining yields 18 phloretin C-glycosyltransferases from plants for the efficient biocatalytic synthesis of nothofagin and phloretin-di-C-glycoside

Natalia Putkaradze | **Valeria Della Gala** | **Dovydas Vaitkus** | **David Teze** | **Ditte Hededam Welner**

The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark

Correspondence
Ditte Hededam Welner, The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK- 2800 Kongens Lyngby, Denmark.
Email: diwel@biosustain.dtu.dk

Abstract

C-glycosyltransferases (*C*-GTs) offer selective and efficient synthesis of natural product *C*-glycosides under mild reaction conditions. In contrast, the chemical synthesis of these *C*-glycosides is challenging and environmentally harmful. The rare occurrence of *C*-glycosylated compounds in Nature, despite their stability, suggests that their biosynthetic enzymes, *C*-GTs, might be scarce. Indeed, the number of characterized *C*-GTs is remarkably lower than *O*-GTs. Therefore, discovery efforts are crucial for expanding our knowledge of these enzymes and their efficient application in biocatalytic processes. This study aimed to identify new *C*-GTs based on their primary sequence. 18 new *C*-GTs were discovered, 10 of which yielded full conversion of phloretin to its glucosides. Phloretin is a dihydrochalcone natural product, with its mono-*C*-glucoside, nothofagin, having various health-promoting effects. Several of these enzymes enabled highly selective production of either nothofagin (UGT708A60 and UGT708F2) or phloretin-di-*C*-glycoside (UGT708D9 and UGT708B8). Molecular docking simulations, based on structural models of selected enzymes, showed productive binding modes for the best phloretin *C*-GTs, UGT708F2 and UGT708A60. Moreover, we characterized UGT708A60 as a highly efficient phloretin mono-*C* glycosyltransferase (k_{cat} = 2.97 s^{-1}, K_M = 0.1 μM) active in non-buffered, dilute sodium hydroxide (0.1–1 mM). We further investigated UGT708A60 as an efficient biocatalyst for the bioproduction of nothofagin.

KEYWORDS
biocatalysis, C-glycosyltransferase, glycosylation, nothofagin, phloretin

1 | INTRODUCTION

C-glycosylation seems to occur scarcely in Nature, with *C*-glycosides being not as abundant as *O*-glycosides and the number of reported natural *C*-glycosyltransferase (*C*-GT) enzymes few.[1–3] *C*-GTs have considerable biotechnological potential due to their unique ability to form *C*-glucosides with antidiabetic, antioxidant, antinociceptive, anticancer, anti-inflammatory, and hepatoprotective activities, as well as

having high metabolic stability.[4,5] The first studies of *C*-GTs were published in the early 2000s[6–8] with the majority of known *C*-GTs discovered within the last few years.[9,10] Most reported *C*-GTs are from plants, display the GT-B structural fold,[11] utilize UDP-Glc as glycosyl donor, and react on the aromatic ring of specific hydroxylated acceptors.[1,3] These *C*-GTs belong to the glycosyltransferases family 1 in the Carbohydrate Active enZyme (CAZy) classification.[12] Early *C*-GT discovery studies were motivated by the presence of *C*-glycosylated compounds in particular organisms such as *Pueraria lobata*[13] and *Glycyrrhiza glabra*[14] and species of bamboo and cereals.[9,15] Fundamental properties that make *C*-GT discovery particularly challenging, include their high sequence similarity with related *O*-glycosyltransferases (*O*-GTs), and the occurrence of both *C*- and *O*-glycosylation activities in a single enzyme towards related compounds or even the same aglycon.[14,16] Moreover, the *C*-glycosylation activity of a *C*-GT might be shifted to *O*-glycosylation by simple amino acid substitutions;[17] with some *C*-GTs capable of forming all four types of glycosidic linkage (*C*-, *O*-, *N*-, and *S*-).[14,18] In this study, we aimed to discover new *C*-GTs based exclusively on the protein primary sequence. *C*-GT candidates from different plants were selected based on sequence identity (>53%) with known *C*-GTs and/or specific sequence motifs; DPF and DPFXL motifs were previously described in the majority of reported *C*-GTs.[1] In addition, we selected for the PSPG (plant secondary product glycosyltransferase) consensus motif, which is conserved among UDP-dependent glycosyltransferases.[19] *C*-GT candidates were recombinantly overproduced in *Escherichia coli*, partially purified, and screened for activity towards common *C*-glycosylation acceptor substrates. 18 new phloretin *C*-GTs were discovered, with four of them yielding a single product, that is, either nothofagin or phloretin-di-*C*-glycoside. These four were further investigated functionally and structurally using AlphaFold and molecular docking. Specifically, this study identifies UGT708A60 from *Hordeum vulgare* (barley) as a highly active phloretin mono-*C*-GT forming nothofagin, and characterizes it biochemically by estimating its kinetic parameters as well as pH and temperature optima.

2 | MATERIALS AND METHODS

2.1 | Reagents and chemicals

Phloretin, phlorizin dehydrate (both ≥99%) and naringenin (≥95%) were purchased from Sigma-Aldrich. Nothofagin (>98%), vitexin (≥97%), and isovitexin (≥98%) were purchased from Carbosynth, Santa Cruz Biotechnology and Merck chemicals, respectively. All reagents were analytical grade.

2.2 | Sequence selection

Full-length UGT sequences were selected from UniProt/GenBank databases according to the following criteria: (i) a sequence identity > 50% with at least one of six reported C-GTs (*Fc*CGT from *Citrus japonica*, *Gg*CGT from *Glycyrrhiza glabra*, *Mi*CGT from *Mangifera indica*, *Os*CGT from *Oryza sativa*, *Gm*CGT from *Glycine max* and *Zm*CGT from *Zea mays*); (ii) the presence of the PSPG and the DPF motifs; (iii) a SolutProt score > 0.25 (https://loschmidt.chemi.muni.cz/soluprot/); and (iv) no more than one sequence from each species. In addition, two sequences from *Solanum* species (UGT708AE2 and UGT708AE1, from *Solanum lycopersicum* and *Solanum tuberosum*, or tomato and potato, respectively) with the DPFIM motif were included, since this motif was not present in any other sequence. The proteins were then named by the UGT Nomenclature Committee.[20]

2.3 | Protein production in *E. coli* and purification

Plasmids encoding the UGTs were synthesized by Biomatik (Canada) and cloned in the pET28a(+) vector (Novagen) using the NcoI and XhoI restriction sites. The protein coding sequence was lacking the start methionine and contained an N-terminal sequence consisting of a 6xHis-tag and a TEV cleavage site as follows: MGHHHHHHDYDIPTTENLYFQGS. *E. coli* BL21 Star (DE3) cells were transformed with the plasmids and precultures grown overnight in LB medium supplemented with 50 mg/L kanamycin at 37°C. Main cultures were prepared by adding 5 mL (1:100 v/v) overnight culture to 500 ml of 2xYT culture with 50 mg/L kanamycin and grown at 37°C and 220 rpm until OD_{600} = 0.5–0.7. The cultures were induced with 0.25–0.5 mM IPTG and grown for approximately 22 h at 20°C. The cells were harvested by centrifugation and at 4°C and resuspended in lysis buffer (50 mM HEPES, 300 mM NaCl, pH 7.6, with 20 mM imidazole) supplemented with 0.4 mg DNAse I. The resuspended cells were subjected to cell lysis using 2–2.5 runs of Avestin emulsiflex C5 (ATA Scientific Pty Ltd., Canada), and the lysate was clarified by centrifugation at 14,500 × *g* at 4°C for 50 min. The supernatant containing the target proteins was filtered with 0.22 μM filter and loaded into 1 ml prepacked HisTrap FF column. The column was previously washed and equilibrated with 10 column volume (CV) of the same buffer (50 mM HEPES, 300 mM NaCl, pH 7.6, with 20 mM imidazole). After protein binding, the column was washed with 18 CV of wash buffer: (50 mM HEPES, 300 mM NaCl, pH 7.6, with 35 mM imidazole). Proteins were eluted with 14 CV imidazole gradient from 35 to 500 mM. The elution peak fractions were pooled together and concentrated using centrifugal filters (MW cutoff 30 kDa). The total concentration of partially purified proteins was measured by spectrophotometric measurements at 280 nm in a NanoDrop Instrument (ThermoFisher Scientific, Wilmington, USA). The sample purity was assessed using SDS-PAGE.

For the enzyme characterization, UGT708A60 was further purified using a 1 ml prepacked HisTrap FF column. The UGT708A60 peak fractions, from the previous purification as described above, were concentrated, and buffer exchanged using a 15-ml centrifugal filter (MW cutoff 30 kDa) to the following buffer: 25 mM HEPES, 150 mM NaCl, pH 7.5, with 0.5 mM EDTA and 1 mM DTT. Tobacco Etch Virus (TEV) protease (10 μg/ml final, produced in house using the pRK793 plasmid) was added to the sample and TEV cleavage was carried out at 10°C for 5 h. The sample was then loaded in the HisTrap FF column,

which was equilibrated with 25 mM HEPES, 150 mM NaCl, pH 7.5, and purification performed using a gradient from 0 to 250 mM imidazole with the same buffer. The flow-through fractions with UGT708A60 were collected and concentrated in the storage buffer (25 mM HEPES 150 mM NaCl, pH 7, 1 mM DTT). The sample purity was analyzed using SDS-PAGE and was >85% (Figure S2).

2.4 | In vitro activity screening

In vitro glycosylation reactions with the partially purified protein samples were carried out in a total reaction volume of 25 μl. The reaction mixture contained 25 mM HEPES buffer, pH 7, 100 μM phloretin or 50 μM nothofagin, phlorizin, apigenin or naringenin, 250 μM UDP-glucose and 50–250 μg/ml of enzymes. The control was made with the reaction mixture without the enzyme. The reaction was carried out at 30°C overnight. After incubation, the reaction was quenched by adding 25 μl of ice-cold acetonitrile.

2.5 | Phloretin *C*-glucosylation by selected UGTs

To investigate phloretin mono and di-*C*-glycosylation reactions over time, phloretin concentration was increased 10-fold with respect to the screening experiment, from 100 μM to 1 mM, and total protein concentration decreased to 75 μg/ml. These reactions were performed in a 25 μl starting volume, with 5 μl samples taken after 5 min, 2 h, 4 and 22 h. The reaction mixture consisted of 25 mM HEPES buffer, pH 7, 1 mM of phloretin, 5 mM of UDP-Glc and 75 μg/ml total protein (partially purified samples) and the reactions performed at 30°C and overnight. The different time point samples were quenched with acetonitrile and analyzed via HPLC.

2.6 | Biochemical characterization of UGT708A60

For the identification of the pH optimum of the enzyme, activity was investigated in 100 mM Tris-BisTris Buffer, pH 5–10, with 0.1 μg UGT708A60, 25 μM phloretin and 100 μM UDP-Glc. 10 μl of reaction were quenched with 240 μl 0.1% acetic acid after a few seconds, 2, 4, 6, and 8 min. The activity of UGT708A60 at higher pHs were also checked in either in 0.03 and 1 mM sodium hydroxide with 100 mM NaCl, 1 μg enzyme, 100 μM phloretin (dissolved in 400 μM NaOH) and 250 μM UDP-Glc.

For the identification of the temperature optimum, the activity of 0.1 μg of enzyme towards 200 μM phloretin was assayed in 100 mM Tris-BisTris buffer, pH 10, with 800 μM UDP-Glc. These reactions were carried out at 30–52°C in a thermocycler. The reactions were stopped by addition of 240 μl acetic acid (0.1%) to 10 μl of the reaction mix after 5, 15, 60, and 180 min.

For the kinetic characterization, the activity of 0.6 ng enzyme was assayed towards 0.1–6 μM phloretin in 100 mM Tris-BisTris buffer, pH 10, with 500 μM UDP-Glc at 45°C. The reaction was initiated by addition of phloretin and quenched with acetic acid after 5 min. The μM of product formed was obtained from the peak area and the corresponding substrate concentration. K_M and k_{cat} values were determined using the Michaelis-Menten or Haldane (substrate inhibition) models, using the drc package in Rstudio 2022.02.0+443.

Time dependent conversion was carried out in 100 mM Tris-BisTris buffer, pH 10, with 10 μg/ml of enzyme, 250 μM phloretin and 1 mM UDP-Glc in 1 ml reactions. After addition of phloretin, the sample was immediately placed in the HPLC sampler at 22°C and 2 μl of the reaction media were injected and measured every 6 min on the HPLC (see below).

2.7 | HPLC analysis

The reaction mixture was analyzed by reversed-phase HPLC using an Ultimate 3000 Series apparatus (Thermo Fisher) and a kinetex 2.6 μm C18 100 Å 100 × 4.6 mm analytical column (Phenomenex). Mixtures of water and acetonitrile containing 0.1% formic acid were used as the mobile phase. A gradient from 5% to 25% acetonitrile for 1.5 min followed by gradients from 25% to 80% for 2 min and from 80% to 100% for 1.5 min were applied for the separation of phloretin and its glucosides as well as apigenin and naringenin reactions. The flow rate was 1 ml/min, and most analytes were detected at 290 nm. Apigenin, vitexin and isovitexin were detected at 320 nm. Authentic standards of phloretin, nothofagin, phlorizin (phloretin 2′-β-*O*-glucoside), vitexin (apigenin 8-β-*C*-glucoside) and isovitexin (apigenin 6-β-*C*-glucoside) were used for the identification of glycosylation specificity. *O*-glucosylation specificity for naringenin reactions were identified by using UGT708A6 reaction product.[21] The HPLC data was monitored and quantified via the Chromeleon software (Thermo Fisher) using the area under a peak.

2.8 | Sequence analysis and structure modelling

The CLC Main Workbench software (version 22.0.1) was used for the multiple sequence alignment (MSA) and maximum likelihood phylogenetic tree amino acid sequences of *C*-GTs and *O*-GTs. The alignment mode was very accurate (slow), gap open cost was 10.0 and gap extension cost was 1.0. The following parameters were applied for maximum likelihood phylogeny: neighbor joining method for the start tree construction, WAG protein substitution model and bootstrap analysis with 100 replicates. Protein structural models were generated by using AlphaFold v2.0, using all available structural homologs, and the database search preset was set to "reduced_dbs."[22] Only the highest ranking (in pLDDT score) models were used in further analyses (Table S2).

2.9 | Molecular docking

Binary complexes of protein and sugar donor were obtained by structurally aligning protein model structures on the crystal structure of *Pt*UGT1 from *Polyngonum tinctorium*, which has a

TABLE 1 Conversion of common *C*-glycosylation substrates by *C*-GTs candidates

C-GT	Phloretin (*C*-/*O*-)	Nothofagin (-*C*)	Phlorizin (*O*- to *C*-)	Apigenin (*C*-/*O*-)	Naringenin (*O*-)
UGT708G4	+, *C*-	++	-	-	+
UGT708AC1	+, *C*-	++	-	-	-
UGT708N1	+++, *C*-	+++	++	++, *O*-	+
UGT708AC2	+++, *C*-	+++	-	+, *C*-	++
UGT708AE1	++, *C*-	++	-	-	-
UGT708AD1	++, *C*-	+	-	-	-
UGT708H7	++, *C*-	-	-	-	+
UGT708H8	+, *C*-	-	-	-	-
UGT708F1	+, *C*-	-	-	-	-
UGT708B1	+++, *C*-	+++	+	-	++
UGT708D9	+++, *C*-	+++	++	-	+
UGT708AF1	+++, *C*-	+	-	-	+
UGT708D10	+++, *C*-	+++	+	-	++
UGT708AE2	+, *C*-, *O*-	++	-	-	+
UGT708F2	+++, *C*-	+++	-	-	+
UGT708B8	+++, *C*-	+++	+	++, *O*-	++
UGT708AF2	+++, *C*-	++	-	-	+
UGT708A60	+++, *C*-	++	+	++, *O*-	++

Note: -no conversion, +conversion < 10%, ++ > conversion > 10%, +++conversion > 90%.

bound UDP-glucose molecule in its active site (6SU6.pdb).[23] The phloretin molecule was added by docking into the acceptor binding site of the binary complexes, using gnina v1.0.1 software,[24] a fork of smina,[25] itself a fork of AutoDock Vina.[26] PyMOL (v2.4.0) was used for superimposition and visualization of resulting structures.[27]

2.10 | Estimation of raw material costs

Reaction conditions for the cyclodextrin-containing process were taken from Schmölzer et al.[28] as follows: 120 mM phloretin, 300 mg/L UGT, 1 mM UDP, 150 mg/L SuSy, 115 mM cyclodextrin inclusion complexes, 97% conversion, and theoretically scaled to produce 1 g of nothofagin. Reaction conditions for the process described in this work were identical to those used for phloretin *C*-glucosylation by selected UGTs: 1 mM phloretin, 75 mg/L UGT, 1 mM UDP, 25 mM HEPES, pH 7, 100% conversion, and scaled to produce 1 g of nothofagin. The price of phloretin (4000 USD/kg) was the average of the selling price from vendors found through PubChem. The price of UGT enzymes (25 USD/kg) was conservatively taken from previously published studies.[29,30] The prices of UDP (50 USD/kg) and UDP-Glc (84.9 kUSD/kg) were the cheapest available online (AliBaba and Carbosynth, respectively). The cost of water (0.0077 USD/kg) is based on proprietary information about RO purified water from BioBased Europe pilot facility in Belgium (http://www.bbeu.org/pilotplant). The price of HEPES (0.059 USD/kg) is that of bulk traded (https://www.zauba.com).

3 | RESULTS AND DISCUSSION

To explore the activity of putative plant *C*-GT candidates, we selected UGT sequences (Table S1) with > 53% sequence similarity to reported *C*-GTs (FcCGT from *Citrus japonica*,[10] GgCGT from *Glycyrrhiza glabra*,[14] MiCGT from *Mangifera indica*,[16] OsCGT from *Oryza sativa*,[15] GmCGT from *Glycine max*[31] and ZmCGT from *Zea mays*[21]). All sequences contained DPF and PSPG motifs. Candidates were expressed in *E. coli* and partially purified for in vitro assaying. Their activities on the common *C*-glycosylation substrates phloretin, its *C*-glucoside nothofagin as well as phlorizin, apigenin and naringenin were measured (Table 1). There were only trace amounts of apigenin *C*-glucoside detectable in the reaction with UGT708AC2, whereas almost all investigated enzymes showed *C*-glycosylation activity towards phloretin and nothofagin. Ten *C*-GT candidates yielded complete glucosylation of phloretin, whereas the others did not reach complete conversion of 100 μM phloretin after 22 h (Table 1 and Figure S1). Phloretin di-*C*-glucoside was detected as the final product for most of the *C*-GT reactions with 100 phloretin and as the sole product with 50 μM nothofagin. Seven enzymes could *C*-glycosylate 50 μM nothofagin with conversion > 90%. Six enzymes could use phlorizin (phloretin 2′-β-*O*-glucoside) instead of phloretin as aglycon, however, their activities were rather low (Table 1). Besides that, several enzymes were identified as bifunctional (*C*- and *O*-) displaying low *O*-glycosylation activity towards a flavone substrate apigenin and a flavanone naringenin as well. Since the samples used for this initial activity screening were only partially purified, absolute comparison of the enzymatic

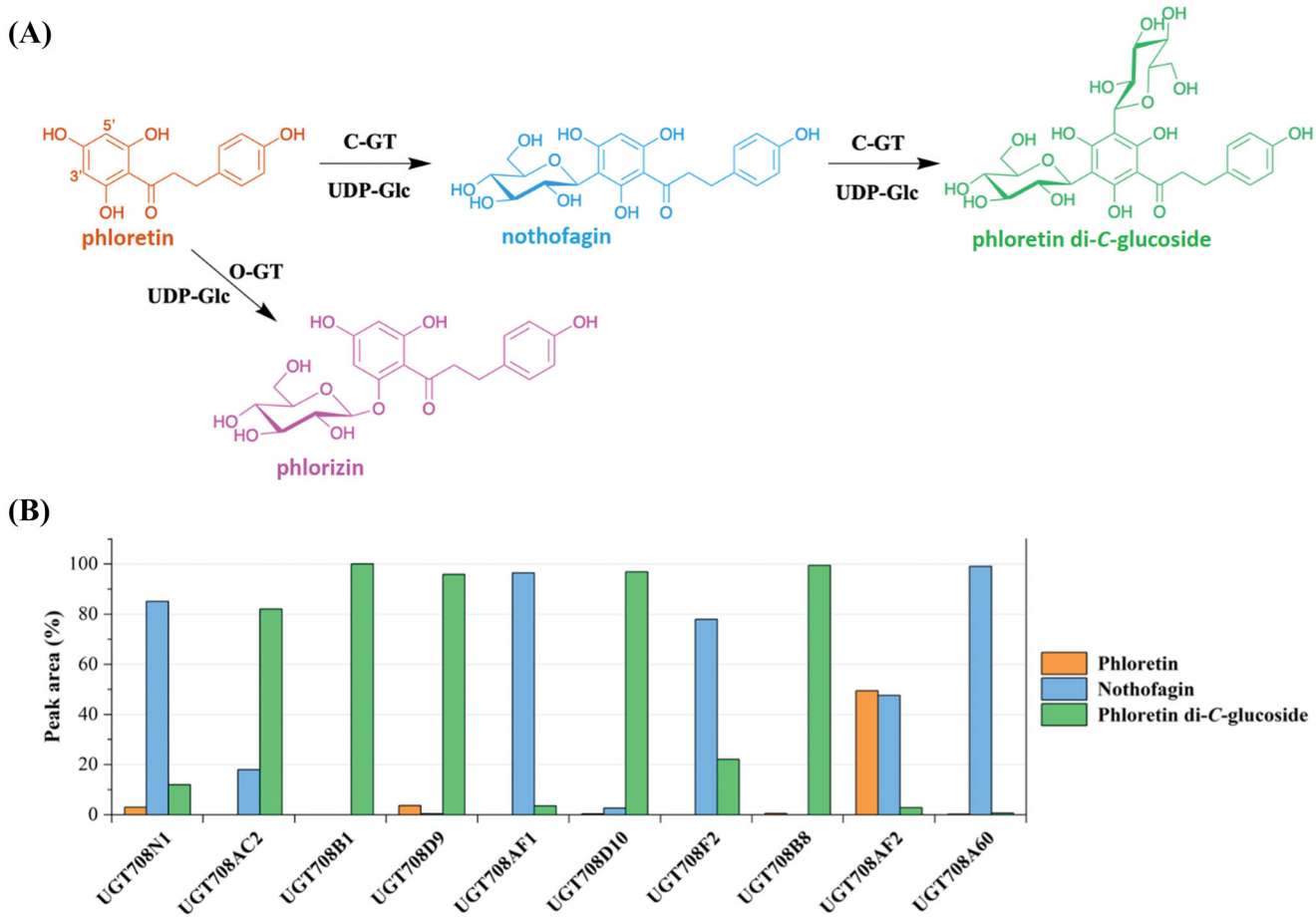


FIGURE 1 *C*-glucosylation of phloretin (orange) to nothofagin (blue) and phloretin di-*C*-glucoside (green). (A) Chemical structures involved in the glucosylation reactions of phloretin, phlorizin (phloretin *O*-glucoside) in pink. (B) Substrate and product distribution in reactions by 10 selected *C*-GTs after 22 h. Peak area shown as percentage of the total peak of phloretin and its glucosides.

activities was not possible at this point. However, we further characterized the 10 enzymes yielding complete glycosylation of phloretin. First, we followed nothofagin and phloretin di-*C*-glucoside formation over time, with modified reaction conditions, in order to observe differences in reaction rates of these enzymes. Indeed, five *C*-GTs (UGT708AC2, UGT708B1, UGT708D9, UGT708D10, and UGT708B8) were able to form predominantly (>80%) di-*C*-glucoside after 22 h whereas the mono-*C*-glucoside nothofagin was present as the main product for UGT708N1, UGT708AF1, UGT708F2 and UGT708A60 and UGT708AF2 showed only ≈50% conversion of phloretin (Figure 1B). These enzymes could also further convert nothofagin to phloretin di-*C*-glucoside to a lesser extent (12.0% for UGT708N1, 3.5% for UGT708AF1, 22.1% for UGT708F2, 0.7% for UGT708A60, and 2.9% for UGT708AF2).

Further, we selected 2 mono- and 2 di-*C*-glycosyltransferases with highest protein productivity per liter of *E. coli* culture (>1 mg/L) to observe their product formation at different time points; after 15 min, 2, 4, and 22 h (Figure 2). Interestingly, phloretin di-*C*-glucoside was detected as the major product for UGT708D9 and UGT708B8 after 2 h (Figure 2A,C), whereas UGT708F2 and UGT708A60 produced almost only nothofagin at all investigated time points. UGT708F2 displayed only low di-*C*-glucosylation activity within 22 h (Figure 2B) and it was negligible for UGT708A60 (Figure 2D).

To investigate structural differences responsible for the different mono- and di-*C*-glucosylation preferences for the four enzymes, structural models using AlphaFold2 were generated.[22] Structural analysis of four selected UGTs was assessed, including the binding of phloretin in presence of both donor (UDP-glucose, superimposed) and the acceptor phloretin. We investigated three parameters to assess if a docking pose could be reactive, relative nucleophilic attack distance (distance $C1_{glc}$-$C3'_{phlo}$, Figure 3A), deprotonation of the phenolic oxygen by the catalytic histidine (distance $N\delta_{His}$-$OH2'_{phlo}$), and orbital angles (angle between the vector perpendicular to the A aromatic ring of phloretin and the $C1_{glc}$-$O1_{glc}$ bond, Figure 3B). Reactive poses for phloretin were obtained for both UGT708F2 and UGT708A60, presenting coplanarity angles of 13.5° and 17.2°, respectively, $C1_{glc}$-$C3'_{phlo}$ distances of 4.5 Å and 4.4 Å, respectively, and $N\delta_{His}$-$OH2'_{phlo}$ distances of 3.2 Å for both (Figure 3). On the other hand, the pose yielding phloretin best positioned for a *C*-glucosylation reaction, according to angles and distances, in UGT708D9 and UGT708B8 presented the aromatic A ring in a tilted orientation (41.1° and 31.6° angles), and in UGT708B8, the nucleophilic attack distance was too long (5.5 Å, Figure 3). These

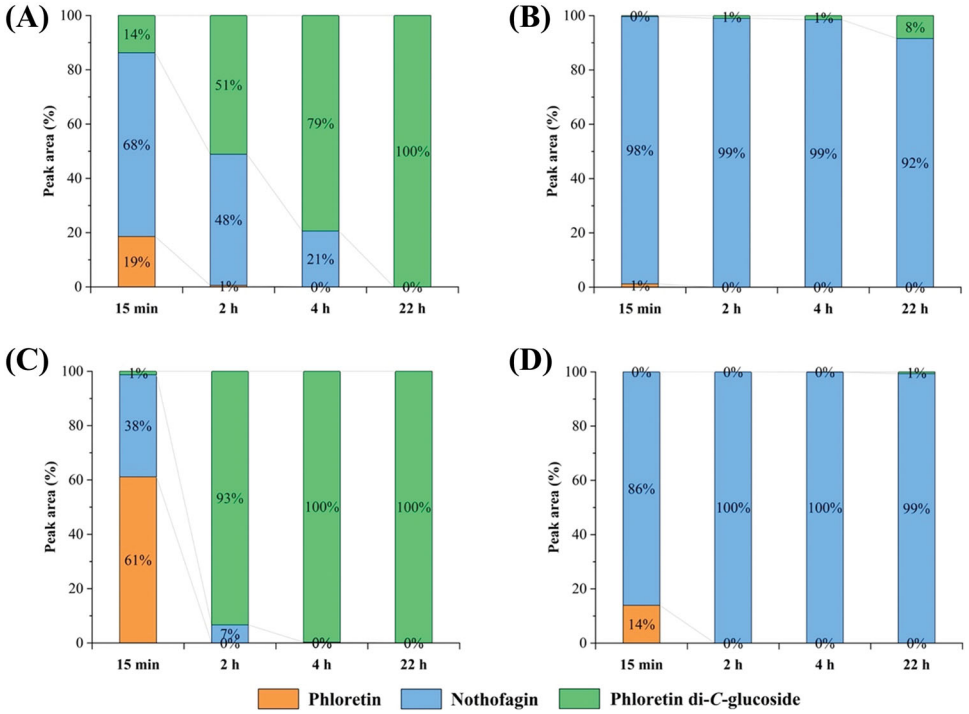


FIGURE 2 Mono- and di-*C*-glucosylation of phloretin by: (A) UGT708D9, (B) UGT708F2, (C) UGT708B8, and (D) UGT708A60 within 22 h. Peak area shown as percentage of the total peak of phloretin and its glucosides.

observations are consistent with UGT708F2 having the most efficient phloretin mono-*C*-glycosylation activity, and UGT708B8 having the worst (Figure 2B,C).

All four enzymes have the residues known to enable *C*-glycosylation: the DPF motif (90-92), the catalytic dyad H24 and D121, and R283 (indexing from UGT708A60).[31] Notably, in all modelled structures, R283 forms a salt bridge with the D90 of the DPF motif (only D is shown in Figure 3). Interestingly, this Arg, which is not otherwise conserved in the GT1 family, was found to be present in all 18 sequences selected based on this DPF motif (Figure S3). Chen and colleagues analyzed two highly similar enzymes (91% sequence identity) from *Mangifera indica*, *Mi*CGT and *Mi*CGTb, which catalyze either mono- (*Mi*CGT) or di-*C*-glycosylation (*Mi*CGTb).[16,18,32] It was shown that I152 of *Mi*CGTb had a crucial role in di-*C*-glycosylation, and its mutation to glutamate converted the enzyme to a mono-*C*-GT. Moreover, the *Mi*CGT mutant E152I efficiently catalyzed di-*C*-glucosylation, convincingly pinpointing residue 152 as the determining factor between these two enzymes.[18] They also reported that *Mi*CGT mutants E152F and E152C produced di-*C*-glucosides with ~20% and ~10% conversion rates, respectively. However, UGT708D9 and UGT708B8, where the analogous residues are F150 and F148, respectively (Figure 3), achieve complete di-*C*-glycosylation, while UGT708F2 (F153) and UGT708A60 (C153) have severely reduced di-*C* capability. Hence, while the residue 152 was the essential discriminant in the *Mangifera indica* enzymes, it does not expand to the proteins analyzed here.

Chen and co-workers proposed that the 152 position had an effect on the active site size, and that di-*C*-glucosylation was allowed by the wider active site. A similar observation was reported in a sperate study with another di-*C*-glycosyltransferase GgCGT.[14] However, the comparison of modelled structures revealed that bi-*C*-glycosylation capable UGT708D9 and UGT708B8 had smaller binding sites, when compared with UGT708F2 and UGT708A60. Indeed, UGT708D9, the most proficient at di-*C*, presents bulky hydrophobic residues (W88 and F120 instead of Phe and Ala/Thr in other three, respectively) that considerably reduced the acceptor binding site size. UGT708B8, which favors mono- over di-*C* by an order of magnitude, presents slightly more space with M117 instead of Val/Ile. In contrast, all residues in the binding site vicinity of UGT708F2 were equal or smaller in size than in UGT708D9/UGT708B8, and UGT708A60, having the largest site, also comprised C153 instead of phenylalanine, and A123 instead of threonine/phenylalanine. Hence, between these four enzymes, the smaller the active site, the more proficient at di-*C*-glycosylation the enzyme is found to be.

We also performed amino acid sequence analysis of new phloretin *C*-GTs in comparison with in-house phloretin *O*-GTs and flavone/isoflavone *C*-GTs (Figure S3 and S4). Important to mention, that the investigated phloretin GTs showed exclusively either *O*- (data not shown) or *C*-glucosylation activity on phloretin. Interestingly, none of the UGTs catalyzing solely the *O*-glucosylation of phloretin contain the DPF motif and D-R combination as shown in Figure 3C-F. Moreover, despite of the known close evolutionary relationship of *O*- and *C*-GTs, all 18 *C*-GTs showed different grouping from phloretin *O*-GTs and flavone/isoflavone *C*-GTs (Figure S4).

For the kinetic characterization we selected UGT708A60 due to its ability for selective formation of nothofagin. pH optimum of the enzyme was found to be at 9.5–10 and temperature optimum at 45°C

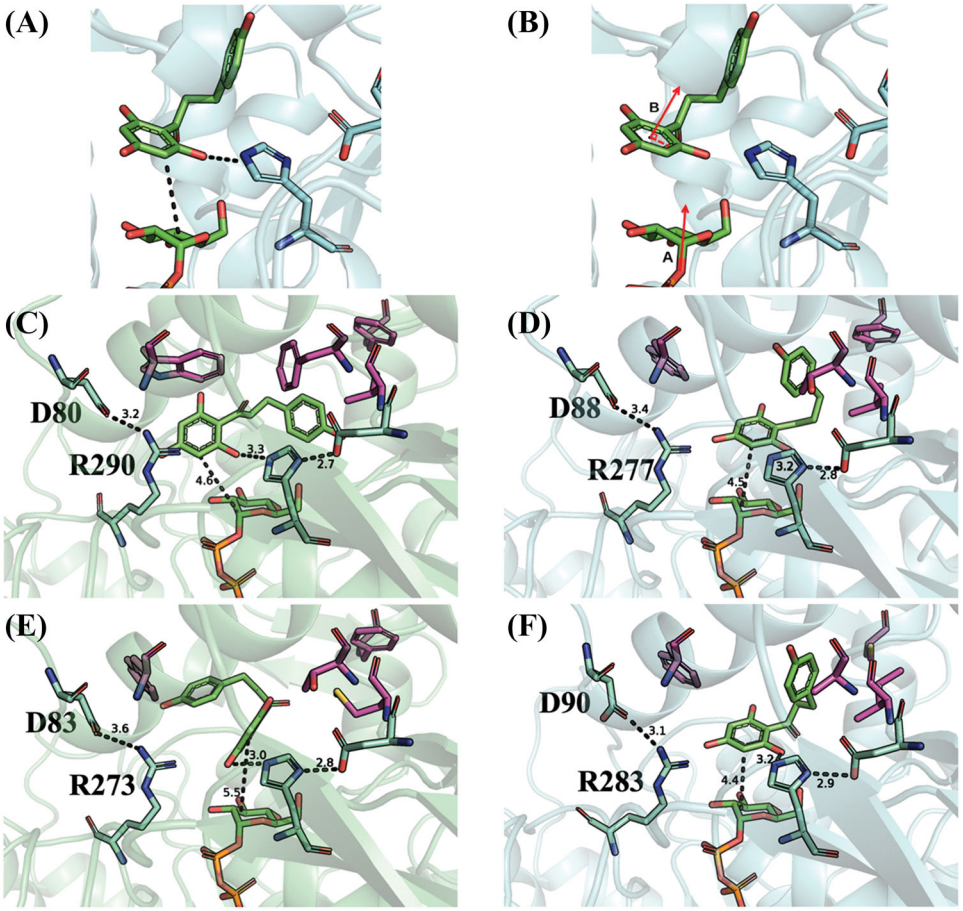


FIGURE 3 Parameters for reactivity assertion: (A) nucleophilic attack distance $C1_{glc}$-$C3'_{phlo}$ and hydrogen bond distance Nd_{His}-$OH2'_{phlo}$ are shown in dotted lines, (B) coplanarity angle between vectors A and B and active sites of: (C) UGT708D9, (D) UGT708F2, (E) UGT708B8, (F) UGT708A60 with phloretin docked and key residues and differences between enzymes highlighted. Variable residues that form the binding pocket are shown in magenta. Chemically relevant distances are measured and presented in Å.

(Figure 4A,B). Peculiarly, the enzyme is also well active in pure NaOH 0.1–10 mM solutions. This curiously high pH optimum is thought to not be physiologically relevant as phloretin glycosylation by UGT708A60 should be cytoplasmic, at a pH of circa 7.5. Besides that, a time course reaction was also performed at room temperature (Figure 4D), with 85% conversion of 250 μM phloretin achieved in 2 h. Further, we estimated the kinetic constants of UGT708A60, with the apparent K_M and k_{cat} values at pH 10 and 45°C found to be 0.1 μM and 2.97 s^{-1}, respectively. According to the BRENDA database on July 8th, 2022, it is the lowest K_M ever reported for a GT1 enzyme against any acceptor.[33] The enzyme efficiently catalyzes synthesis of nothofagin showing a catalytic efficiency of 2.8×10^7 s^{-1} M^{-1} which is the highest reported so far for selective nothofagin-producing enzymes (Table 2). The enzyme also presents a slight substrate inhibition, that quickly plateaus at 1.60 s^{-1} (Figure S5). Another phloretin *C*-GT with a comparable catalytic efficiency (>2.4×10^7 s^{-1} M^{-1}) to UGT708A60 is *Fc*CGT,[10] however, it was characterized as a di-*C*-GT producing both: phloretin and nothofagin.

Although enzymatic conversion of phloretin to nothofagin is likely competitive with conventional chemistry with respect to environmental sustainability, it must also be economically feasible to be implemented. The state-of-the-art of enzymatic nothofagin production uses cyclodextrin to solubilize the hydrophobic phloretin.[28] We estimate the raw materials of this approach to amount to 1536 USD per gram of nothofagin produced, driven by the high price of cyclodextrin (Table S3). If, instead, we propose to produce nothofagin at lower titers with higher water consumption, as demonstrated here, thus omitting the need for cyclodextrin, we arrive at a raw material cost of 556.8 USD per gram, or 5.9 if UDP-Glc is recycled with the SuSy system, if compatible conditions could be found (e.g., regarding pH, fructose concentrations, etc).[28] This paves the way for biotechnological production of this valuable compound.

4 | CONCLUSIONS

C-glycosyltransferases (*C*-GTs) are attractive enzymes for biotechnological production of valuable *C*-glycosylated polyphenols. However, information about their substrates, specificity and reaction mechanism are limited, due in part by the small number of these enzymes discovered to date. To discover new *C*-GTs, we mined sequence databases and successfully identified and produced 18 *C*-GTs, all presenting activity

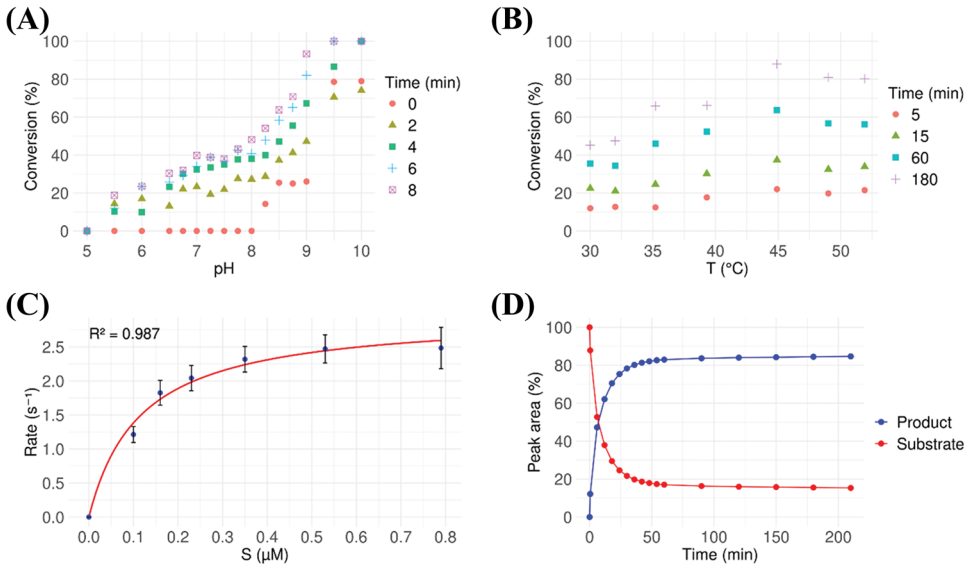


FIGURE 4 Biochemical characterization of UGT708A60. pH (A) and temperature (B) optima, Michaelis-Menten plot (C) and time course reaction (D) within 210 min at 298 K.

TABLE 2 Estimated catalytic values of UGT708A60 and previously reported *C*-GTs selective towards nothofagin production

C-GT	k_{cat} [s^{-1}]	K_M [μM] phloretin	Catalytic efficiency [s^{-1} M^{-1}]	Specific activity [nkat mg protein^{-1}]
*Os*CGT[34]	4.4	9	0.5×10^6	85 (5.1 ± 0.3 U/mg)
*Os*CGT[15]	10.84	4.78	2.2×10^6	0.587
*Mi*CGTb[32]	0.79	166	0.047×10^5	-
UGT708A60	2.97	0.1	2.8×10^7	-

on the polyphenol phloretin. While the molecular discriminants that govern the mono- and di-*C*-glucosylation balance could not be determined, we showed that a narrower active site does not necessarily favor the mono-*C*-glucosylation. Our results suggest that previously identified residues and properties do not translate to all the *C*-GT enzymes. Hence, it is important to discover and report more systems presenting either mono or di-*C*-glycosylation, to obtain a comprehensive view of the mechanistic determinants. Moreover, we fully characterized UGT708A60, and propose it as an efficient biocatalyst for green and economically feasible production of the mono-*C*-glucoside nothofagin.

AUTHOR CONTRIBUTIONS

Natalia Putkaradze: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Supervision; Validation; Writing – original draft; Writing – review & editing. Valeria Della Gala: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing – review & editing. Dovydas Vaitkus: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Visualization; Writing – review & editing. David Teze: Formal analysis; Investigation; Methodology; Supervision; Validation; Visualization; Writing – review & editing. Ditte Hededam Welner: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing – review & editing.

ACKNOWLEDGMENTS

The authors thank Dr. Cameron James Hunt for language edition, Dr. Michael Court from UGT Nomenclature Committee (Washington State University) for assigning unique names to the new *C*-GTs, as well as the Novo Nordisk Foundation for financial support through grants NNF21OC0071624, NNF18OC0034744, NNF10CC1016517 and NNF20CC0035580.

CONFLICT OF INTEREST

The authors declare no competing interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the article and/or the supplementary material of this article.

REFERENCES

1. Putkaradze, N., Teze, D., Fredslund, F., & Welner, D. H. (2021). Natural product C-glycosyltransferases—A scarcely characterised enzymatic activity with biotechnological potential. *Natural Product Reports*, *38*(3), 432–443. https://doi.org/10.1039/d0np00040j
2. Jones, P., & Vogt, T. (2001). Glycosyltransferases in secondary plant metabolism: Tranquilizers and stimulant controllers. *Planta*, *213*(2), 164–174. https://doi.org/10.1007/s004250000492
3. Tegl, G., & Nidetzky, B. (2020). Leloir glycosyltransferases of natural product *C*-glycosylation: Structure, mechanism and specificity. *Biochemical Society Transactions*, *48*(4), 1583–1598. https://doi.org/10.1042/BST20191140
4. He, M., Min, J. W., Kong, W. L., He, X. H., Li, J. X., & Peng, B. W. (2016). A review on the pharmacological effects of vitexin and isovitexin. *Fitoterapia*, *115*, 74–85. https://doi.org/10.1016/j.fitote.2016.09.011
5. Bucar, F., Xiao, J. B., & Ochensberger, S. (2020). *Flavonoid C-glycosides in diets*. Springer Singapore, 10.1007/978-981-15-4148-3_6
6. Trefzer, A., Hoffmeister, D., Künzel, E., Stockert, S., Weitnauer, G., Westrich, L., Rix, U., Fuchser, J., Bindseil, K., Rohr, J., & Bechthold, A. (2000). Function of glycosyltransferase genes involved in urdamycin A biosynthesis. *Chemistry & Biology*, *7*(2), 133–142. https://doi.org/10.1016/S1074-5521(00)00079-X
7. Hoffmeister, D., Dräger, G., Ichinose, K., Rohr, J., & Bechthold, A. (2003). The *C*-glycosyltransferase UrdGT2 is unselective toward D- and L-configured nucleotide-bound rhodinoses. *Journal of the American Chemical Society*, *125*(16), 4678–4679. https://doi.org/10.1021/ja029645k
8. Mittler, M., Bechthold, A., & Schulz, G. E. (2007). Structure and action of the C–C Bond-forming Glycosyltransferase UrdGT2 involved in the biosynthesis of the antibiotic urdamycin. *Journal of Molecular Biology*, *372*(1), 67–76. https://doi.org/10.1016/j.jmb.2007.06.005
9. Sun, Y., Chen, Z., Yang, J., Mutanda, I., Li, S., Zhang, Q., Zhang, Y., Zhang, Y., & Wang, Y. (2020). Pathway-specific enzymes from bamboo and crop leaves biosynthesize anti-nociceptive *C*-glycosylated flavones. *Communications Biology*, *3*(1), 1–11. https://doi.org/10.1038/s42003-020-0834-3
10. Ito, T., Fujimoto, S., Suito, F., Shimosaka, M., & Taguchi, G. (2017). *C*-Glycosyltransferases catalyzing the formation of di-*C*-glucosyl flavonoids in citrus plants. *Plant Journal*, *91*(2), 187–198. https://doi.org/10.1111/tpj.13555
11. Lairson, L. L., Henrissat, B., Davies, G. J., & Withers, S. G. (2008). Glycosyltransferases: Structures, Functions, and Mechanisms. *Annual Review of Biochemistry*, *77*(1), 521–555. https://doi.org/10.1146/annurev.biochem.76.061005.092322
12. Wang, X., Li, C., Zhou, C., Li, J., & Zhang, Y. (2017). Molecular characterization of the *C*-glucosylation for puerarin biosynthesis in Pueraria lobata. *Plant Journal*, *90*(3), 535–546. https://doi.org/10.1111/tpj.13510
13. Zhang, M., Li, F.-D., Li, K., Wang, Z.-L., Wang, Y.-X., He, J.-B., Su, H.-F., Zhang, Z.-Y., Chi, C.-B., Shi, X.-M., Yun, C.-H., Zhang, Z.-Y., Liu, Z.-M., Zhang, L.-R., Yang, D.-H., Ma, M., Qiao, X., & Ye, M. (2020). Functional characterization and structural basis of an efficient di-*C*-glycosyltransferase from Glycyrrhiza glabra. *Journal of the American Chemical Society*, *142*, 3506–3512. https://doi.org/10.1021/jacs.9b12211
14. Brazier-Hicks, M., Evans, K. M., Gershater, M. C., Puschmann, H., Steel, P. G., & Edwards, R. (2009). The *C*-glycosylation of flavonoids in cereals. *Journal of Biological Chemistry*, *284*(27), 17926–17934. https://doi.org/10.1074/jbc.M109.009258
15. Chen, D., Chen, R., Wang, R., Li, J., Xie, K., Bian, C., Sun, L., Zhang, X., Liu, J., Yang, L., Ye, F., Yu, X., & Dai, J. (2015). Probing the catalytic promiscuity of a regio- and stereospecific *C*-Glycosyltransferase from Mangifera indica. *Angewandte Chemie*, *127*(43), 12869–12873. https://doi.org/10.1002/ange.201506505
16. Gutmann, A., & Nidetzky, B. (2013). Enzymatic *C*-glycosylation: Insights from the study of a complementary pair of plant *O*- and *C*-glucosyltransferases. *Pure and Applied Chemistry*, *85*(9), 1865–1877. https://doi.org/10.1351/pac-con-12-11-24
17. Chen, D., Fan, S., Chen, R., Xie, K., Yin, S., Sun, L., Liu, J., Yang, L., Kong, J., Yang, Z., & Dai, J. (2018). Probing and engineering key residues for bis-*C*-glycosylation and promiscuity of a *C*-Glycosyltransferase. *ACS Catalysis*, *8*(6), 4917–4927. https://doi.org/10.1021/acscatal.8b00376
18. Osmani, S. A., Bak, S., & Møller, B. L. (2009). Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. *Phytochemistry*, *70*(3), 325–347. https://doi.org/10.1016/j.phytochem.2008.12.009
19. Mackenzie, P. I., Owens, I. S., Burchell, B., Bock, K. W., Bairoch, A., Belanger, A., Gigleux, S. F., Green, M., Hum, D. W., Iyanagi, T., Lancet, D., Louisot, P., Magdalou, J., Roy Chowdhury, J., Ritter, J. K., Tephly, T. R., Schachter, H., Tephly, T., Tipton, K. F., & Nebert, D. W. (1997). The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics*, *7*(4), 255–269. https://doi.org/10.1097/00008571-199708000-00001
20. Ferreyra, M. L. F., Rodriguez, E., Casas, M. I., Labadie, G., Grotewold, E., & Casati, P. (2013). Identification of a bifunctional Maize *C*- and *O*-glucosyltransferase. *Journal of Biological Chemistry*, *288*(44), 31678–31688. https://doi.org/10.1074/jbc.M113.510040
21. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … , Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2
22. Teze, D., Coines, J., Fredslund, F., Dubey, K. D., Bidart, G. N., Adams, P. D., Dueber, J. E., Svensson, B., Rovira, C., & Welner, D. H. (2021). *O-/N-/S*-specificity in glycosyltransferase catalysis: From mechanistic understanding to engineering. *ACS Catalysis*, *11*(3), 1810–1815. https://doi.org/10.1021/acscatal.0c04171
23. McNutt, A. T., Francoeur, P., Aggarwal, R., Masuda, T., Meli, R., Ragoza, M., Sunseri, J., & Koes, D. R. (2021). GNINA 1.0: Molecular docking with deep learning. *Journal of Cheminformatics*, *13*(1), 1–20. https://doi.org/10.1186/s13321-021-00522-2
24. Koes, D. R., Baumgartner, M. P., & Camacho, C. J. (2013). Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *Journal of Chemical Information and Modeling*, *53*(8), 1893–1904. https://doi.org/10.1021/ci300604z
25. Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455–461. https://doi.org/10.1002/jcc.21334
26. Schrödinger, L. (2010). The PyMOL molecular graphics system, version 2.4.
27. Schmölzer, K., Lemmerer, M., & Nidetzky, B. (2018). Glycosyltransferase cascades made fit for chemical production: Integrated biocatalytic process for the natural polyphenol *C*-glucoside nothofagin. *Biotechnology and Bioengineering*, *115*(3), 545–556. https://doi.org/10.1002/bit.26491
28. Tournier, V., Topham, C. M., Gilles, A., David, B., Folgoas, C., Moya-Leclair, E., Kamionka, E., Desrousseaux, M.-L., Texier, H., Gavalda, S., Cot, M., Guémard, E., Dalibey, M., Nomme, J., Cioci, G., Barbe, S., Chateau, M., André, I., Duquesne, S., & Marty, A. (2020). An engineered PET depolymerase to break down and recycle plastic bottles. *Nature*, *580*(7802), 216–219. https://doi.org/10.1038/s41586-020-2149-4
29. Liu, G., Zhang, J., & Bao, J. (2016). Cost evaluation of cellulase enzyme for industrial-scale cellulosic ethanol production based on rigorous Aspen Plus modeling. *Bioprocess and Biosystems Engineering*, *39*(1), 133–140. https://doi.org/10.1007/s00449-015-1497-1
30. Hirade, Y., Kotoku, N., Terasaka, K., Saijo-Hamano, Y., Fukumoto, A., & Mizukami, H. (2015). Identification and functional analysis of 2-

hydroxyflavanone C-glucosyltransferase in soybean (Glycine max). *Febs Letters*, *589*(15), 1778–1786. https://doi.org/10.1016/j.febslet.2015.05.010

31. Chen, D., Sun, L., Chen, R., Xie, K., Yang, L., & Dai, J. (2016). Enzymatic synthesis of acylphloroglucinol 3-*C*-Glucosides from 2-*O*-Glucosides using a *C*-Glycosyltransferase from Mangifera indica. *Chemistry: A European Journal*, *22*(17), 5873–5877. https://doi.org/10.1002/chem.201600411
32. Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D., & Schomburg, D. (2021). BRENDA, the ELIXIR core data resource in 2021: New developments and updates. *Nucleic Acids Research*, *49*(D1), D498–D508. https://doi.org/10.1093/nar/gkaa1025
33. Bungaruang, L., Gutmann, A., & Nidetzky, B. (2013). Leloir glycosyltransferases and natural product glycosylation: Biocatalytic synthesis of the *C*-glucoside nothofagin, a major antioxidant of redbush herbal tea. *Advanced Synthesis & Catalysis*, *355*(14–15), 2757–2763. https://doi.org/10.1002/adsc.201300251
34. Nidetzky, B., Gutmann, A., & Zhong, C. (2018). Leloir glycosyltransferases as biocatalysts for chemical production. *ACS Catalysis*, *8*(7), 6283–6300. https://doi.org/10.1021/acscatal.8b00710

SUPPORTING INFORMATION

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

How to cite this article: Putkaradze, N., Gala, V. D., Vaitkus, D., Teze, D., & Welner, D. H. (2023). Sequence mining yields 18 phloretin C-glycosyltransferases from plants for the efficient biocatalytic synthesis of nothofagin and phloretin-di-C-glycoside. *Biotechnology Journal*, *18*, e2200609. https://doi.org/10.1002/biot.202200609