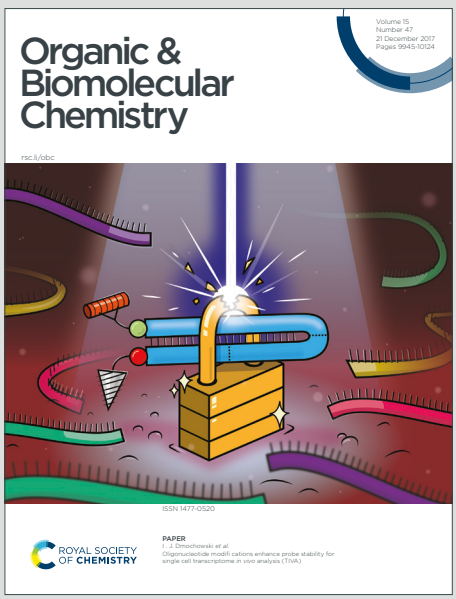
Organic & Biomolecular Chemistry

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: K. Feng, R. Chen, K. Xie, D. Chen, J. Liu, W. Du, L. Yang and J. Dai, *Org. Biomol. Chem.*, 2019, DOI: 10.1039/C9OB01352K.



ROYAL SOCIETY OF CHEMISTRY

rsc.li/obc

View Article Online
DOI: 10.1039/C9OB01352K

Journal Name

ARTICLE

Ep7GT, a glycosyltransferase with sugar donor flexibility from *Epimedium pseudowushanense*, catalyzes the 7-*O*-glycosylation of baohuoside

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Keping Feng,[a, b, c] Ridao Chen,[b , c, d] Kebo Xie, [b, c, d] Dawei Chen, [b, c, d] Jimei Liu, [b, c, d] Wenyu Du, [b, c, d] Lin Yang [a, *] and Jungui Dai [b, c, d, *]

Icariin (**1a**), a 7-*O*-glycosylated flavonoid glycoside, is recognized as the major pharmacologically active ingredient of *Epimedium* plants, which have been used in traditional Chinese medicine for thousands of years. However, no glycosyltransferase (GT) responsible for the 7-*O*-glycosylation of flavonoids has been identified from *Epimedium* plants to date. Herein, a GT, Ep7GT, was identified from *E. pseudowushanense* B. L. Guo, which can regiospecifically transfer a glucose moiety to baohuoside (**1**) at 7-OH to form icariin (**1a**). Ep7GT showed a rare broad donor substrate spectrum, including UDP-glucose, UDP-xylose, UDP-*N*-acetylglucosamine, UDP-rhamnose, UDP-galactose, UDP-glucuronic acid and TDP-glucose. Moreover, two new derivatives of icariin (**1a**), 7-*O*-β-$_D$-[2-(acetylamino)-2-deoxy-glucopyranosyl]-baohuoside (**1b**) and 7-*O*-β-$_D$-xylosyl-baohuoside (**1c**), were biosynthesized by Ep7GT *in vitro*. Engineered *Escherichia coli* harbouring *Ep7GT* was constructed, and 10.1 μg/mL icariin (**1a**) was yielded by whole-cell biotransformation with baohuoside (**1**) as substrate. The present work not only characterizes the GT responsible for the 7-*O*-glycosylation in the biosynthesis of icariin in *Epimedium* plants, but also indicates the significant potential of an enzymatic approach for the production of glycosylated baohuoside derivatives with different sugar moieties. What's more, these findings also provide a promising alternative for producing natural/unnatural bioactive flavonoid glycosides by metabolic engineering.

Introduction

Icariin (**1a**, Figure 1) is the major pharmacologically active ingredient of *Epimedium* plants. It is a flavonoid glycoside with multiple biological activities including antioxidant, anti-inflammatory, anti-osteoporotic, neuroprotective, angiogenesis stimulating, post-stroke dementia improving, testosterone mimicking, anti-depressant and tumour multidrug resistance reversal activities[1-9]. Chemical synthesis of icariin (**1a**) doesn't seem practicable because of its low yield[10]; and isolating icariin (**1a**) from *Epimedium* plants is also limited by the low rate of recovery[11]. Thus, the biosynthetic processing of icariin (**1a**) could represent a promising alternative and can also yield novel derivatives of icariin (**1a**).

In our previous work, a rhamnosyltransferase and a UDP-rhamnose synthase from *E. pseudowushanense* B. L. Guo that catalyses the rhamnosylation of anhydroicaritin to form baohuoside (**1**, Figure 1) have been characterized[12]. Furthermore, a hypothesized biosynthetic pathway for icariin (**1a**) in *Epimedium* has been proposed, in which the 7-*O*-glucosylation is supposed to occur after the 3-*O*-rhamnosylation[12]. As a part of our ongoing work, we have isolated and functionally characterized a glycosyltransferase (GT) Ep7GT from *E. pseudowushanense* catalysing the 7-*O*-glucosylation of baohuoside (**1**) to form icariin (**1a**), which accomplishes a step forward in the biosynthesis of icariin (**1a**). Moreover, an engineered *Escherichia coli* strain containing *Ep7GT* was constructed to produce icariin (**1a**) through whole-cell bioconversion with baohuoside (**1**) as substrate.

While most of the plant GTs are found to be highly specific to sugar donor substrates[13], Ep7GT was found to possess donor-flexibility, enabling it to recognize various NDP-sugars as donors to yield novel derivatives of icariin (**1a**). This has broadened the applications of GTs in plants to synthesize and modify flavonoids with new structures and more potent biological activities.

Results and discussion

Cloning candidate cDNAs from *E. pseudowushanense*

A transcriptome database of *E. pseudowushanense* containing a total of 52,967 unigenes was acquired[12]. To narrow down the range of candidate genes, local BLAST (basic local alignment search tool) queries were performed to search the assembled unigenes using the amino acid sequences of AtUGT73C6 from *Arabidopsis thaliana*, which catalysed the transfer of glucose

[a] College of Life and Environmental Sciences, Minzu University of China, 27 Zhong Guan Cun Southern Street, Beijing 100081, China
E-mail: 15116995486@163.com; jgdai@imm.ac.cn
[b] State Key Laboratory of Bioactive Substance and Function of Natural Medicines; [c] CAMS Key Laboratory of Enzyme and Biocatalysis of Natural Drugs; [d] NHC Key Laboratory of Biosynthesis of Natural Products, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, 1 Xian Nong Tan Street, Beijing, 100050, China

†Electronic Supplementary Information (ESI) available: See DOI: 10.1039/x0xx00000x

from UDP-glucose to 7-OH substituents of kaempferol-3-*O*-rhamnoside and quercetin-3-*O*-rhamnoside[14]. Sequences with high similarity to the query probe from the database were considered as candidate genes for further functional identification. After obtaining their full-length sequences by RACE (rapid-amplification of cDNA ends), their ORFs (open reading frames) were isolated by nested PCR (polymerase chain reaction). Finally, the ORFs of the 9 candidate GT cDNAs, which were named *EpGT1-9* (*EpGT8* was renamed *Ep7GT* after functional characterization), were retrieved for further investigation.

Functional characterization of recombinant EpGTs *in vitro*.

The coding regions of *EpGTs* were cloned into the expression vector pET-28a and overexpressed in *E. coli*. For the glycosylation activity assay, the crude extracts containing the recombinant EpGTs were incubated with baohuoside (**1**) as sugar acceptor in the presence of UDP-glucose. Control reactions were performed with crude extracts prepared from *E. coli* harbouring an empty vector. The assay results showed that EpGT2, EpGT4 and EpGT8 (renamed Ep7GT after functional characterization) were able to catalyze the conversion of baohuoside (**1**) to a single glucosylated product. Among them, EpGT8 exhibited the highest conversion rate, and thus was selected for further study (Figure S2).

To clarify the glycosylation properties, the enzymatic product (**1a**) was prepared and further characterized by extensive spectroscopic data analysis. One product with higher polarity was detected in the reaction mixture (0.4 mM **1**, 0.8 mM UDPG, 5 μg of purified enzyme) by HPLC-MS (Figure 1). Selected ion monitoring for *m/z* 721 $[M–H+HCOOH]^-$ suggested a mono-glucosylated product with a molecular weight that was 162 amu higher than that of **1** (Figure 1B, 1D). No activity was observed in the control group. In the ^{1}H NMR spectrum, the appearance of a proton at δ_H 5.00 (1H, d, J = 9.0 Hz, H-1‴) with large coupling constant indicated the introduction of a glucose moiety with β type glycosidic bond. The observation of an upfield shift for C-7 (δ_C 160.5) in the ^{13}C NMR spectrum compared to that of **1** (δ_C 162.2) suggested that the glucose moiety was introduced to 7-OH[15]. All of the data are in good agreement with the previously reported data for icarrin[11]. Thus, the enzymatic product **1a** was unequivocally identified as 7-*O*-β-$_D$-glucosyl-baohuoside, icariin. Accordingly, EpGT8 is a glycosyltransferase which can regiospecifically transfer a glucose moiety to baohuoside (**1**) at 7-OH to form icariin (**1a**). What's more, a phylogenetic tree of EpGT8 and other plant GTs was constructed and revealed that EpGT8 was grouped into the 7GT cluster (GTs catalysing the 7-*O*-glycosylation of flavonoids) clade (Figure 2). Thus, EpGT8 was assigned as a flavonoid 7-*O*-glucosyltransferase and renamed Ep7GT (*Epimedium pseudowushanense* 7-*O*-glucosyltransferase).

Ep7GT shows an ORF of 1440 bp encoding a protein of 479 amino acids with a calculated molecular mass of 54.1 kDa. A multiple sequence alignment of the encoded polypeptides of

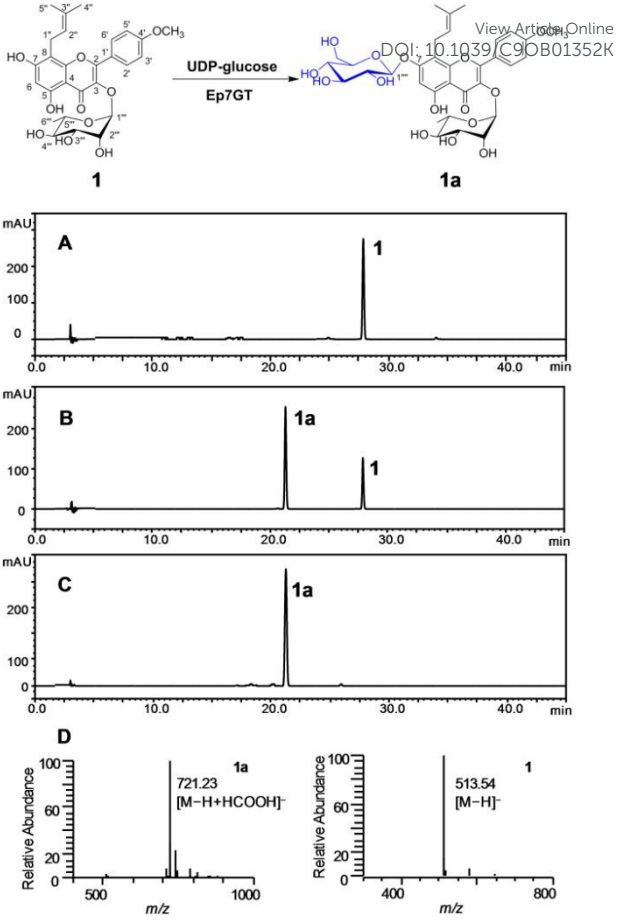


Figure 1. Functional characterization of recombinant Ep7GT. A) Control group; B) Enzymatic reaction catalysed by Ep7GT with **1** as sugar accepter and UDP-glucose as sugar donor; C) Standard of **1a**; D) MS spectra of **1a** and **1** in negative mode. The maximum of UV absorption peak of **1a** is 270 nm.

Ep7GT and the reported glycosyltransferase AtUGT73C6 from *A. thaliana* and EpPF3RT from *E. pseudowushanense* are performed (Figure S3). Highly conserved sequences of the GTs, PSPG (plant secondary product glycosyltransferase) motifs, were observed in these polypeptides. Moreover, Ep7GT shared identities of 42.7% at the amino acid level with AtUGT73C6 and 24.4% with EpPF3RT.

Substrate specificity of Ep7GT

To investigate the sugar donor specificity of Ep7GT, six other sugar donors including UDP-xylose, UDP-*N*-acetylglucosamine, UDP-rhamnose, UDP-galactose, UDP-glucuronic acid and TDP-glucose along with UDP-glucose were tested (Figure 3). Surprisingly, Ep7GT also showed activity towards UDP-xylose, UDP-*N*-acetylglucosamine and TDP-glucose with baohuoside (**1**) as the acceptor (Figure 4). When testing other putative intermediates involved in the biosynthetic pathway of icariin (**1a**), Ep7GT displayed activity towards all tested sugar donors with kaempferol (**2**) as the acceptor to form tri-glycosylkaempferol, di-glycosylkaempferol or mono-glycosylkaempferol (Figure 5). Ep7GT showed activity towards

DOI: 10.1039/C9OB01353K

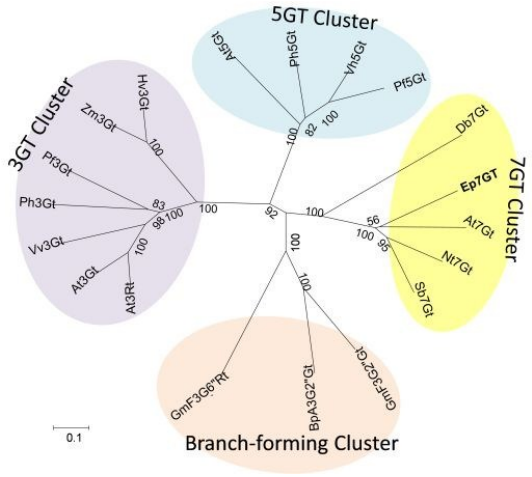


Figure 2. The phylogenetic relationships between Ep7GT and the related plant flavonoid glycosyltransferases. The protein sequences were aligned using ClustalW. The neighbour-joining phylogenetic tree was drawn using MEGA 5. The bootstrap value was 1000, and the branch lengths represent the relative genetic distances. The abbreviations of the protein sequences and their accession numbers are as follows: (At3Rt (*Arabidopsis thaliana*; NM_102790); At3Gt (*A. thaliana*; NM_121711); Vv3Gt (*Vitis vinifera*; AF000371); Ph3Gt (*Petunia hybrid*; AB027454); Pf3Gt (*Perilla frutescens*; AB002818); Hv3Gt (*Hordeum vulgare*; X15694); Zm3Gt (*Zea mays*; X13501); At5Gt (*A. thaliana*; NM_117485); Pf5Gt (*P. frutescens*; AB013596); Vh5Gt (*Verbena hybrid*; BAA36423); Ph5Gt (*P. hybrid*; AB027455); Db7Gt (*Dorotheanthus bellidiformis*; CAB56231); Nt7Gt (*Nicotiana tabacum*; AAB36653); Sb7Gt (*Scutellaria baicalensis*; BAA83484); At7Gt (*A. thaliana*; NM_129234); BpA3G2''Gt (*Bellis perennis*; AB190262); GmF3G2''Gt (*Glycine max*; BAR88077); GmF3G6''Rt (*G. max*; BAN91401)).

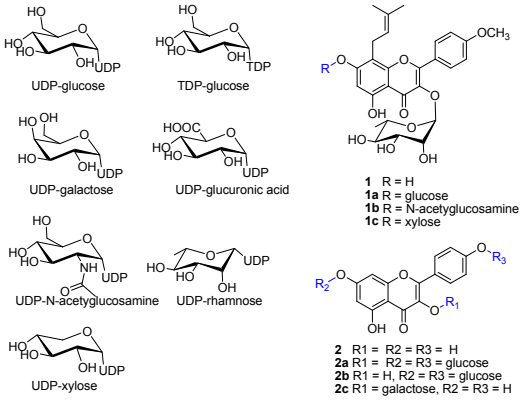


Figure 3. Structures of sugar donors and enzymatic products of Ep7GT-catalysed glycosylating reactions.

glycosylkaempferol (Figure 5). Ep7GT showed activity towards UDP-glucose, UDP-galactose, UDP-N-acetylglucosamine and UDP-xylose with 8-prenylkaempferol (3) (Figure S4) as the acceptor. With anhydroicaritin as acceptor, Ep7GT can recognize UDP-glucose and TDP-glucose (Figure S5); with baohuoside II as acceptor, Ep7GT can recognize UDP-glucose, UDP-N-acetylglucosamine, and UDP-xylose (Figure S6). It's interesting that there are no secondary metabolites with UDP-N-acetylglucosamine and/or UPD-glucuronic acid. It is a rare example that a GT exhibited such broad specificity towards sugar donors with different acceptors. Moreover, it's worthy to note that Ep7GT exhibited different substrate spectra towards sugar donors for different acceptors. And, Ep7GT showed higher activity to UDP-glucose with K_m 146.9 μM than UDP-N-acetylglucosamine with Km 348.0 μM and UDP-xylose with K_m 541.3 μM with baohuoside (1) as sugar acceptor. As for the relative activities of Ep7GT with anhydroicaritin and baohuoside, Ep7GT showed higher activity when UDP-Glc was used as donor. However, anhydroicaritin has more hydroxy to be glycosylated and resulted in more products. Most plant GTs are highly specific to the acceptor and donor substrates in vivo probably due to the limitation of diversity of substrates in cell. In fact, some GTs displayed relatively broad substrate spectra in vitro. The catalytic promiscuity makes these GTs be multi-functional GTs. Once the plant needs some kind of glycosides, only corresponding aglycons need to be synthesized and the already existed GTs will achieve the glycosylation reactions. So the evolution of GTs with broad specificity might be a more effective way for its secondary metabolism, also be a reaction against the different environment.

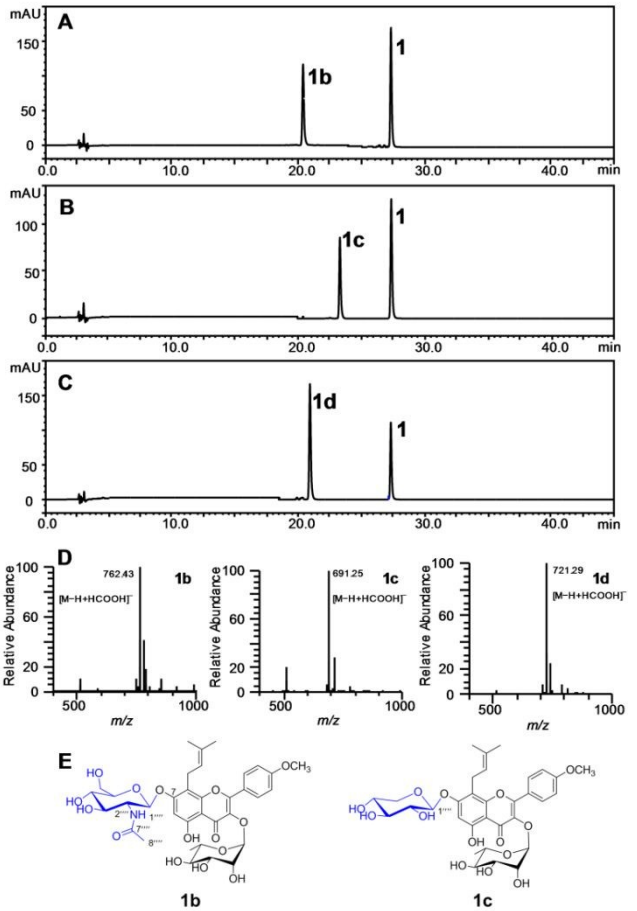


Figure 4. Ep7GT-catalysed glycosylation of **1** with UDP-*N*-acetylglucosamine (A), UDP-xylose (B) and TDP-glucose (C) as sugar donors; D) MS spectra of **1b**, **1c** and **1d** in negative mode; (E) Structures of **1b** and **1c**. The maximum of UV absorption peaks of **1b**, **1c** and **1d** are 270 nm.

View Article Online
DOI: 10.1039/C9OB01352K

Journal Name

ARTICLE

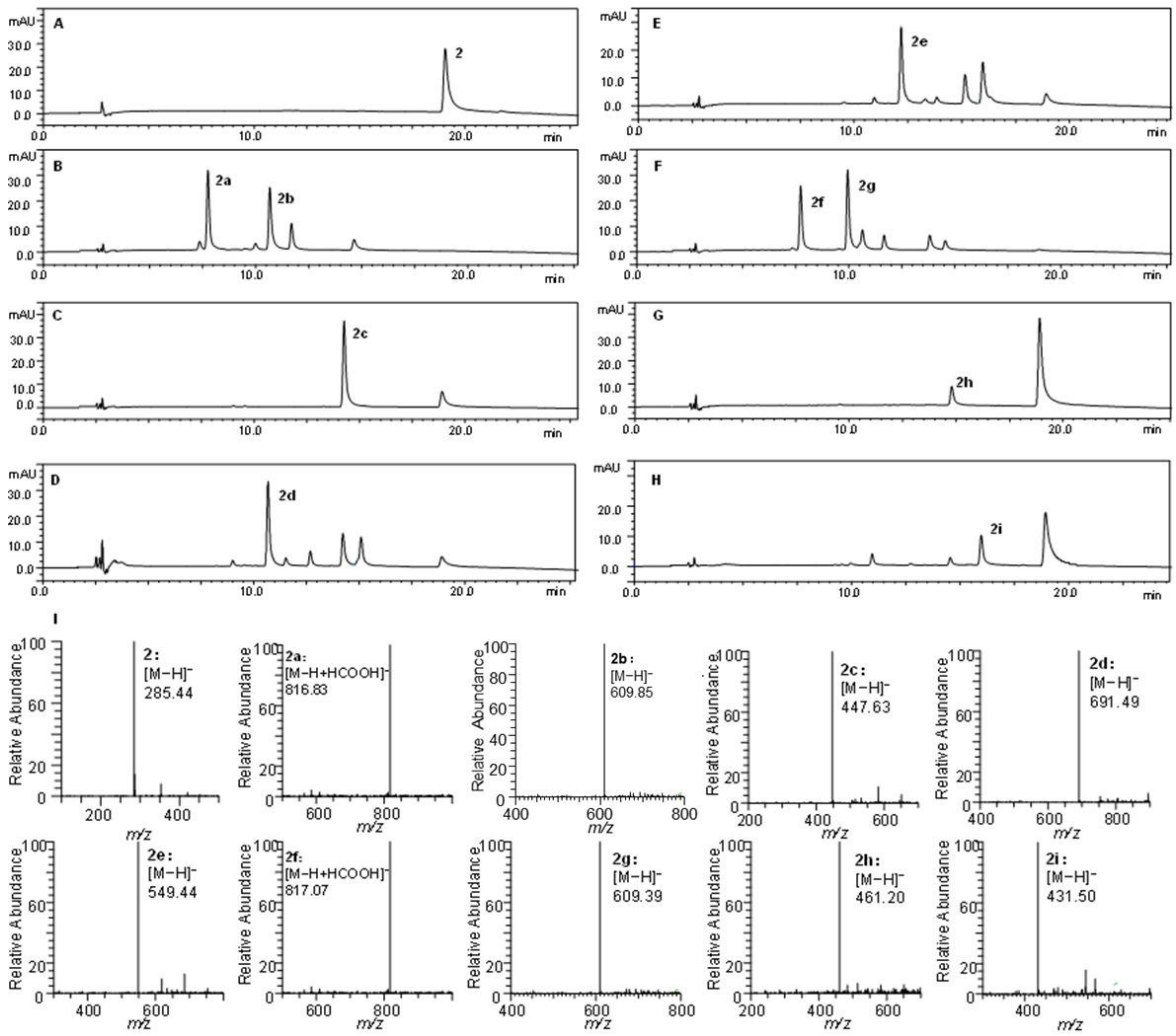


Figure 5. Ep7GT-catalysed glycosylation of **2** with different sugar donors**.** A), Control group; Reactions with sugar donors: (B), UDP-glucose; (C), UDP-galactose; (D), UDP-*N*-acetylglucosamine; (E), UDP-xylose; (F), TDP-glucose; (G), UDP-glucuronic acid and (H) UDP-rhamnose; I) MS spectra of the main products in negative mode. The maximum of UV absorption peaks of all the products are 270 nm.

The representative enzymatic products (**1b**, **1c**, **2a, 2b** and **2c**) were prepared and further characterized to clarify the glycosylation properties. For **1b**, one product with HRESIMS peak at *m/z* 762.2261 $[M–H+COOH]^-$ suggested the introduction of an *N*-acetylglucosamine moiety with a molecular weight that was 204 amu higher than that of **1** (Figure 4). The appearance of protons at δ_H 5.02 (1H, d, *J* = 9.0 Hz, H-1‴), 7.84 (1H, d, *J*= 9.2 Hz, H-N*H*), and 1.81 (3H, s, H-8‴) in its ^{1}H NMR spectrum indicated the introduction of a *N*-acetylglucosamine moiety with β type glycosidic bond. According to the HMBC correlation of H-1‴/C-7, the *N*-acetylglucosamine moiety was deduced to be attached at 7-OH. Thus, the enzymatic product **1b** was unequivocally identified as 7-*O*-β-$_D$-[2-(*N*-acetylamino)-2-deoxy-glucopyranosyl]-baohuoside. The ^{1}H NMR spectrum of product **1c** showed protons at δ_H 5.06 (1H, d, *J*= 7.2 Hz, H-1⁗) with a large coupling constant, and the 132 amu higher molecular weight than that of **1** also indicated the appearance of a xylose moiety with β type glycosidic bond. The 7-*O*-xylose moiety was confirmed by HMBC correlation of H-1⁗/C-7. Thus, the enzymatic product **1c** was unequivocally identified as 7-*O*-β-$_D$-xylosyl-baohuoside. The 486 amu higher molecular mass than that of **2** and the appearances of protons at δ_H 5.49 (1H, d, *J*=7.1 Hz, H-1′), 5.09 (1H, d, *J*=7.2 Hz, H-1‴), and 5.03 (1H, d, *J*=7.3 Hz, H-1⁗)

DOI: 10.1039/C9OB01352K

indicated the appearance of three β type glycosidic bonds of **2a**. According to the HMBC correlations of H-1''/C-3, H-1'''/C-7 and H-1''''/C-4', the glucose moieties were determined to attach at 3-OH, 7-OH and 4'-OH, respectively. Thus, **2a** was unequivocally identified as 3,7,4'-tri-*O*-β-$_D$-glucosyl-kaempferol. All of the data are in good agreement with the previously reported data for 3,7,4'-tri-β-$_D$-*O*-glucosyl-kaempferol[16]. The 324 amu higher molecular weight and the appearances of protons at δ_H 5.08 (1H, d, *J*=7.3 Hz, H-1'') and 5.01 (1H, d, *J*=6.9 Hz, H-1''') indicated the existence of two

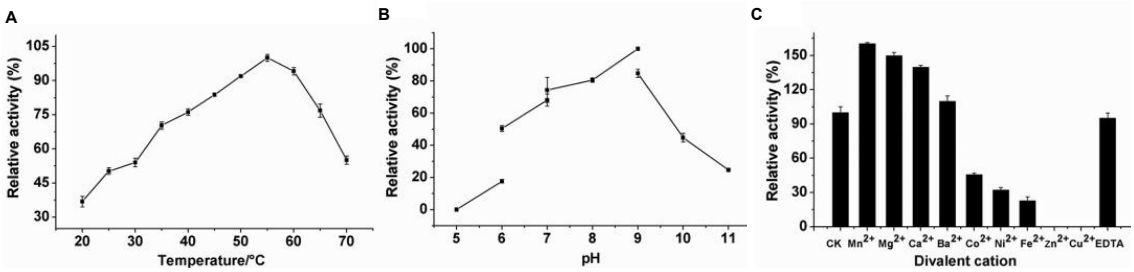


Figure 6. Effects of temperature (A), pH; (B) and various divalent metal ions; "CK" is a control group without adding divalent cations or EDTA to the reaction system; (C) on enzyme activity of Ep7GT. UDP-glucose and baohuoside (**1**) were used as sugar donor and acceptor. The values are presented as the means, and the error bars show the SD (*n* = 3).

glucose moieties of **2b**. After analysing by HMBC correlations, the moieties were attached to C-7 and C-4' positions, respectively; therefore, **2b** was characterized as 7,4'-di-*O*-β-$_D$-glucosyl-kaempferol. The enzymatic product **2c** showed a molecular mass of 448, which was 162 higher than that of **2**. The proton at δ_H 5.14 (1H, d, *J*=7.7 Hz, H-1'') and the correlation of H-1''/C-3 at HMBC obviously suggested that **2c** was a 3-*O*-galacosylated product, 3-*O*-β-$_D$-galacosyl-kaempferol, astragalin[17].

Biochemical properties of recombinant Ep7GT

As indicated in Figure 6, when UDP-glucose was used as sugar donor and baohuoside (**1**) was used as sugar acceptor, Ep7GT exhibited a broad optimum temperature range of 20-70 °C (pH 7.4), and its maximum activity was observed at 55 °C. Analysis of the enzyme activity ranging from pH 5.0 to pH 11.0 at 37 °C showed that the maximum activity was achieved at pH 9.0. The effect of divalent cations on the activity displayed that this enzyme was metal ion-independent, where Mn^{2+}, Mg^{2+}, Ca^{2+} and Ba^{2+} can improve the activity, while the other divalent cations decreased the activity. The glucosylation activity was observed to decrease in the order of Mn^{2+}>Mg^{2+}>Ca^{2+}>Ba^{2+}>Co^{2+}>Ni^{2+}> Fe^{2+}>Sn^{2+}>Cu^{2+} at 37 °C and pH 7.4. According to the k_{cat}/K_m values for different sugar donors (Table S3 and Figure S30), Ep7GT showed higher affinity to UDP-glucose than UDP-xylose and UDP-*N*-acetylglucosamine with baohuoside (**1**) as sugar acceptor.

Bioconversion with engineered *E. coli*

Whole-cell biocatalysis can overcome the supply of expensive exogenous donors such as UDP-glucose and the tedious isolation steps of recombinant proteins. And, there is no SignalP for Ep7GT, so the recombinant enzyme is still inside the cell. And the substrates can go through the cell membrane and then be glycosylated by Ep7GT inside the cell. So there is

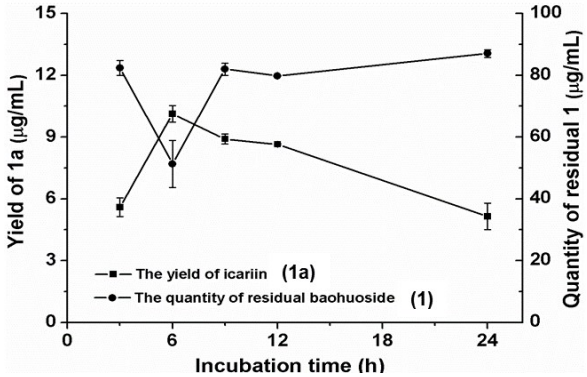


Figure 7. Time course for the biotransformation of baohuoside (**1**). The values are presented as the means, and the error bars show the SD (*n* = 3).

no need to lyse the cells in the bioconversion reactions. Thus, engineered *E. coli* with an intrinsic biosynthesis pathway of UDP-glucose was constructed to produce icariin (**1a**) by whole-cell biotransformation. The engineered *E. coli* strain which contained the pET-28a vector carrying *Ep7GT* produced a single product, which was determined to be icariin (**1a**), whereas no product was detected with the strain containing the empty pET-28a vector. As a result, 10.1 μg/mL of icariin (**1a**) with a conversion rate of 11.7% was produced by this biosystem after 6 h of incubation with 200 μM baohuoside (**1**) added into 1 mL cell cultures (Figure 7). Surprisingly, the yield of icariin (**1a**) became decreased with incubation durations longer than 6 h, while the amount of baohuoside (**1**) increased accordingly, Ep7GT was indeed able to catalyze the deglycosyaltion of icariin (**1a**) to generate baohuoside (**1**) (Figure S32). However, Icariin (**1a**) can't be hydrolyzed when incubated with Ep7GT alone. To explore the degradation of Icariin (**1a**) in the biotransformation system, Icariin (**1a**) was incubated with whole cell catalyst and the relative amout of **1a** was analyzed. In the whole-cell biotransformation system, the concentration of Icariin (**1a**) was indeed decreased without generating baohuoside (**1**) when incubated with whole cell catalyst according to the results of

experiments (Figure S33), suggesting the reversible chemical reaction equilibrium in this system. This module could also be used as a part of *de novo* biosynthesis of icariin (**1a**) through further optimization strategies including enzyme engineering, fermentation engineering, etc.

Experimental section

General methods

UDP-glucose, UDP-*N*-acetylglucosamine, UDP-galactose, and UDP-glucuronic acid were purchased from Sigma-Aldrich (St. Louis, USA). UDP-xylose was purchased from Beijing BG Biotech Co. Ltd. (Beijing, China), UDP-rhamnose was prepared from UDP-glucose by EpRhS[12] and TDP-glucose was obtained as a gift. All of the acceptor substrates were purchased from Nantong Feiyu Biological Technology Co., Ltd. (Nantong, China). The analyses of the substrate specificity and the determinations of the conversion rates were performed using an Agilent 1200 series HPLC system (Agilent Technologies, Germany) coupled with an LCQ Fleet ion trap mass spectrometer (Thermo Electron Corp., USA) equipped with an electrospray ionization (ESI) source (Table S1, Supporting Information, SI). Compounds were characterized by HRESIMS using an Agilent Technologies 6520 Accurate Mass Q-TOF LC/MS spectrometer (Agilent Technologies, Santa Clara, USA) and ^{1}H NMR at 400 or 600 MHz and ^{13}C NMR at 100 or 150 MHz using a Mercury-400 and a Bruker AVIIIHD-600 spectrometers. The chemical shifts (δ) were referenced to internal solvent resonances and are given in parts per million (ppm). Coupling constants (J) are given in hertz (Hz).

Plant materials

Fresh leaves of *E. pseudowushanense* were collected and immediately frozen with liquid nitrogen and stored at −80 °C until use. *E. pseudowushanense* was cultivated in Guizhou Province, China, and identified by Prof. Baolin Guo, Institute of Medicinal Plant Development, Peking Union Medical College & Chinese Academy of Medical Sciences.

Functional annotation of unigenes from *E. pseudowushanense* transcriptome data

Transcriptome sequencing of *E. pseudowushanense* was performed by the Beijing Genomics Institute. In brief, a cDNA library derived from mRNA of *E. pseudowushanense* was sequenced using Illumina HiSeq™ 2000. A large number of short reads was hence generated from the sequencing. These resulting raw reads were then combined by Trinity, an assembling programme, to form longer fragments that could not be extended at either end, which were defined as unigenes. The unigenes were then annotated by blastx against GenBank databases such as NR (non-redundant proteins), Swiss-Prot, KEGG (Kyoto Encyclopedia of Genes and Genomes) and COG (cluster of orthologous groups of proteins) (e-value < 0.00001). The unigene sequences were first aligned to protein databases such as NR, Swiss-Prot, KEGG and COG (e-value < 0.00001) by blastx and to the nucleotide database NT (Nucleotide Sequence Database) (e-value < 0.00001) by blastn, retrieving the proteins with the highest sequence similarity to the given unigenes along with their protein functional annotations.

Isolation, expression and purification of the Ep7GT

Total RNA of the leaves of *E. pseudowushanense* was prepared by using the E.Z.N.A.™ Plant RNA Kit (Omega Bio-Tek, USA) and reverse-transcribed to cDNA with SmartScribe reverse transcriptase (Clontech, USA) following the manufacturer's instructions. Local BLAST queries were performed with the bioinformatics software BioEdit to search the assembled unigenes using the amino acid sequences of AtUGT73C6 from *Arabidopsis thaliana*[14]. The unigenes were analysed using the online bioinformatics tools from NCBI. The open reading frame (ORF) finding was performed using the online programme (https://www.ncbi.nlm.nih.gov/orffinder/). The missing sequences of the tentative partial cDNAs were obtained by RACE (rapid amplification of cDNA end). After obtaining the full-length sequences, the ORFs of the candidate genes were also isolated by nested PCR. Finally, the ORFs of the candidate GT cDNAs, which were named *EpGT1-9*, were retrieved for further study (Table S2).

The PCR products were purified and cloned into the pET-28a vector using an eFusion Cloning Kit (Biophay, China). After verification of the sequences, the recombinant plasmids and empty pET-28a were transformed into Transetta (DE3) *E. coli* (TransGen Biotech, China) for heterologous expression. The described *E. coli* strains were cultivated in 1 L shake flasks at 37 °C and 200 rpm using 200 mL LB (Luria-Bertani) medium containing 50 μg/mL kanamycin and 34 μg/mL chloromycetin until they reached an OD_{600} (optical density at 600 nm) of 0.4–0.6. The recombinant N-terminal His_6-EpGTs expression was performed by induction with 0.1 mM IPTG (isopropyl-β-$_D$-thiogalactoside) for 16 h at 18 °C.

The cells were harvested by centrifugation at 10,000 g for 5 min at 4 °C. The cells were resuspended in 20 mL Tris-HCl buffer (50 mM Tris-HCl buffer, 50 mM NaCl, 1 mM DTT, 5% glycerol, pH 7.4) and disrupted by sonication in an ice bath. The cell debris was removed by centrifugation at 10,000 g and 4 °C for 60 min, and the supernatant was collected as crude extract. For purification, the harvested cells were resuspended in 20 mL binding buffer (20 mM phosphate buffer, 0.5 M NaCl, 20 mM imidazole, pH 7.4) containing 1 mM PMSF (phenylmethylsulfonyl fluoride). The cells were also disrupted by sonication in an ice bath, and the cell debris was removed by centrifugation at 10,000 g and 4 °C for 60 min. The soluble fraction was passed through a 0.45 μm syringe filter unit, and the cleared supernatant was immediately applied to a 1 mL column of Ni-NTA resin (GE, USA) that was pre-equilibrated with binding buffer. The resin was subsequently eluted with 5 mL of washing buffer (20 mM phosphate buffer, 0.5 M NaCl, 50 mM imidazole, pH 7.4). The elution was carried out with 5 mL of different elution buffers (20 mM phosphate buffer, 0.5 M NaCl, 75–250 mM imidazole, pH 7.4). The protein purification was performed with a flow rate of 1 mL/min and a temperature of 4 °C. The proteins were concentrated and buffer exchanged to a desalting buffer (50

View Article Online
DOI: 10.1039/C9OB01352K

mM Tris-HCl buffer, 50 mM NaCl, 1 mM DTT, 5% glycerol, pH 7.4) using an Amicon Ultra-30K centrifugal concentrator (Millipore, USA). The protein purity was confirmed by by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and the protein concentration for all of the studies was determined using the Protein Quantitative Kit (TransGen Biotech, China). The final protein was flash-frozen in liquid nitrogen and stored at −80 °C. For example, the approximate protein yield for EpGT8 (Ep7GT) (54.1 kDa, Figure S1) was 5.4 mg/L.

Activity assays *in vitro*

0.4 mM UDP-glucose was incubated with 0.2 mM aglycon, 50 mM Tris-HCl (pH 7.4) and crude extracts of 20 mL strains containing the recombinant EpGTs in a final volume of 100 µL. Control experiments were performed with the crude extracts of strains containing the empty vector. The activity assays, initiated by the addition of the enzymes, were performed at 30 °C for up to 6 h and terminated by the addition of 200 µL of ice-cold MeOH. For activity assay *in vitro* of purified EpGT8 (Ep7GT), EpGT2 and EpGT4, the reaction mixture containing 0.4 mM baohuoside (**1**), 0.8 mM UDPG, 5 µg of recombinant enzymes were incubated in a total volume of 100 µL at 30 °C and pH7.4 for 30 min. Subsequently, the samples were centrifuged at 15,000 g for 30 min to collect the supernatant and then analysed by HPLC-UV/ESIMS as described in the general methods. For quantification, three parallel assays were routinely carried out. The HPLC peak areas were integrated with Agilent Workstation Software, and the total percent conversion was calculated as a percentage of the total peak areas of the substrate and the product. The HPLC analyses were performed on a Shiseido CAPCELL PAK C18 MG III column (250 mm × 4.6 mm I.D., 5 µm, Shiseido Co., Ltd., Tokyo, Japan) at a flow rate of 1 mL/min, with column temperature maintained at 30 °C. The mobile phase was a gradient elution of solvent A (0.1% formic acid aqueous solution) and solvent B (MeOH). Gradient programmes were used for the analyses of the reactions (Table S1).

Preparative scale reactions

Respectively, 19.5 µmol, 23.3 µmol and 23.3 µmol of baohuoside (**1**), 14.0 µmol and 70.0 µmol of kaempferol (**2**) were dissolved in 200 µL DMSO and diluted with a buffer solution (50 mM Tris-HCl, pH 7.4, 25 mL total volume). UDP-glucose, UDP-*N*-acetylglucosamine, UDP-xylose and UDP-galacose were added to the same molar amount of aglycon **1** and twice the molar amount of aglycon **2**, along with 25 mL of the purified enzyme Ep7GT, respectively. The reactions were performed at 30 °C for up to 12 h, followed by extraction with ethyl acetate (100 mL × 5). The organic phase was evaporated to dryness under reduced pressure, and the residue was dissolved in 1.5 mL MeOH and purified by reverse-phase semi-preparative HPLC. The obtained products were weighed, and yields of 32.8% (**1a**, 6.4 µmol), 19.7% (**1b**, 4.6 µmol), 6.0% (**1c**, 1.4 µmol), 11.4% (**2a**, 8.0 µmol), 14.7% (**2b**, 10.3 µmol) and 17.1% (**2c**, 2.4 µmol) were calculated, respectively. The structures were identified by MS, ^{1}H NMR and ^{13}C NMR spectroscopic data analysis (Figures S7–S29).

The Spectroscopic Data of Prepared Glucuronides

Icariin (**1a**, isolated yield 32.8%). ^{1}H NMR (600 MHz, DMSO-d_6): δ_H 12.56 (1H, s, 5-OH), 7.89 (2H, d, *J*=8.1 Hz, H-2′/6′), 7.13 (2H, d, *J*=8.2 Hz, H-3′/5′), 6.63 (1H, s, H-6), 5.28 (1H, d, *J*=1.9 Hz, H-1‴), 5.17 (1H, t, *J*=9.0 Hz, H-2″), 5.00 (1H, d, *J*=9.0 Hz, H-1⁗), 3.86 (1H, s, 3H, O-CH_3), 3.34 (2H, H-1″), 3.07–4.00 (protons in rhamnose and glucose), 1.69 (3H, s, H-4″), 1.60 (3H, s, H-5″), 0.79 (3H, d, *J*=7.2 Hz, H-6‴); ^{13}C NMR (125 MHz, DMSO-d_6): δ_C 178.3 (C-4), 161.4 (C-4′), 160.5 (C-7), 159.1 (C-5), 157.3 (C-2), 153.0 (C-9), 134.6 (C-3), 131.1 (C-3″), 130.6 (C-2′/6′), 122.3 (C-1′), 122.1 (C-2″), 114.1 (C-3′/5′), 108.3 (C-8), 105.6 (C-10), 102.0 (C-1‴), 100.5 (C-1⁗), 98.1 (C-6), 77.2 (C-5⁗), 76.6 (C-3⁗), 73.4 (C-2⁗), 71.1 (C-4‴), 70.7 (C-5‴), 70.3 (C-3‴), 70.1 (C-2‴), 69.7 (C-4⁗), 60.6 (C-6⁗), 55.5 (OCH_3-4′), 25.5 (C-5″), 21.4 (C-1″), 17.9 (C-4″), 17.5 (C-6‴). ESI-MS: *m/z* 721 [M–H+COOH]$^-$. All of these data are in good agreement with the previously reported data for icariin[11].

7-*O*-β-$_D$-[2-(*N*-acetylamino)-2-deoxy-glucopyranosyl]-baohuoside (**1b**, novel, isolated yield 19.7%). ^{1}H NMR (600 MHz, DMSO-d_6): δH 7.86 (2H, d, *J*=8.9 Hz, H-2′/6′), 7.84 (1H, d, *J*=9.2 Hz, H-NH), 7.11 (2H, d, *J*=8.9 Hz, H-3′/5′), 6.61 (1H,s, H-6), 5.27 (1H,d, *J*=1.6 Hz, H-1‴), 5.05 (1H, t, *J*=6.6 Hz, H-2″), 5.02 (1H, d, *J*=9.0 Hz, H-1⁗), 3.85 (3H, s, O-CH_3), 3.37 (2H, H-1″), 3.04–3.98 (protons in rhamnose and *N*-acetylglucosamine), 1.81 (3H, s, H-8⁗), 1.64 (3H, s, H-4″), 1.59 (3H, s, H-5″), 0.77 (3H, d, *J*=6.1 Hz, H-6‴); ^{13}C NMR (125 MHz, DMSO-d_6): δC 178.2 (C-4), 169.2 (C-7⁗), 161.4 (C-4′), 160.4 (C-7), 159.1 (C-5), 157.5 (C-2), 153.0 (C-9), 134.7 (C-3), 131.0 (C-3″), 130.6 (C-2′/6′), 122.3 (C-1′), 122.2 (C-2″), 114.0 (C-3′/5′), 108.4 (C-8), 105.7 (C-10), 102.0 (C-1‴), 99.4 (C-1⁗), 97.9 (C-6), 77.4 (C-5⁗), 74.0 (C-3⁗), 71.1 (C-4‴), 70.7 (C-5‴), 70.3 (C-3‴), 70.2 (C-2‴), 70.1 (C-4⁗), 60.6 (C-6⁗), 55.5 (OCH_3-4′), 55.0 (C-2⁗), 25.5 (C-4″), 23.0 (C-8⁗), 21.0 (C-1″), 17.9 (C-5″), 17.5 (C-6‴). HRESIMS: *m/z* 762.2261 [M–H+COOH]$^-$.

7-*O*-β-$_D$-Xylosyl-baohuoside (**1c**, novel, isolated yield 6.0%). ^{1}H NMR (600 MHz, DMSO-d_6): δ_H 7.90 (2H, d, *J*=9.0 Hz, H-2′/6′), 7.13 (2H, d, *J*=9.0 Hz, H-3′/5′), 6.60 (1H, s, H-6), 5.28 (1H, d, *J*=1.7 Hz, H-1‴), 5.16 (1H, t, *J*=6.6 Hz, H-2″), 5.06 (1H, d, *J*=7.2 Hz, H-1⁗), 3.86 (3H, s, OCH_3), 3.07–4.00 (protons in rhamnose and xylose), 1.68 (3H, s, H-4″), 1.60 (3H, s, H-4″), 0.79 (3H, d, *J*=6.1 Hz, H-6‴); ^{13}C NMR (125 MHz, DMSO-d_6): δ_C 178.3 (C-4), 161.3 (C-4′), 160.2 (C-7), 159.0 (C-5), 157.3 (C-2), 153.1 (C-9), 134.7 (C-3), 131.1 (C-3″), 130.5 (C-2′/6′), 122.3 (C-1′), 122.0 (C-2″), 114.1 (C-3′/5′), 108.4 (C-8), 105.6 (C-10), 102.0 (C-1‴), 100.7 (C-1⁗), 97.9 (C-6), 77.2 (C-5⁗), 76.3 (C-3⁗), 73.1 (C-2⁗), 71.1 (C-4‴), 70.7 (C-5‴), 70.3 (C-3‴), 70.1 (C-2‴), 69.3 (C-4⁗), 55.5 (OCH_3-4′), 25.5 (C-5″), 21.4 (C-1″), 17.9 (C-4″), 17.5 (C-6‴). HRESIMS: *m/z* 691.1915 [M–H+COOH]$^-$.

3,7,4′-Tri-*O*-β-$_D$-glucosyl-kaempferol (**2a**, isolated yield 11.4%). ^{1}H NMR (400 MHz, DMSO-d_6): δ_H 8.14 (2H, d, *J*=9.0 Hz, H-2′/6′), 7.17 (2H, d, *J*=9.0 Hz, H-3′/5′), 6.82 (1H, d, *J*=2.2 Hz, H-8), 6.45 (1H, d, *J*=2.2 Hz, H-6), 5.49 (1H, d, *J*=7.1 Hz, H-1″), 5.09 (1H, d, *J*=7.2 Hz, H-1‴), 5.03 (1H, d, *J*=7.3 Hz, H-1⁗), 3.15–4.0 (protons

in glucose); ^{13}C NMR (100 MHz, Methanol-d_4): δ_C 177.8 (C-4), 162.9 (C-7), 160.9 (C-5), 159.4 (C-4′), 156.1 (C-9), 156.1 (C-2), 134.0 (C-3), 130.7 (C-2′/6′), 123.6 (C-1′), 115.8 (C-3′/5′), 105.8 (C-10), 100.7 (C-6), 94.5 (C-8), 100.7 (C-1″), 74.2 (C-2″), 77.2 (C-3″), 69.9 (C-4″), 76.5 (C-5″), 60.7 (C-6″), 99.9 (C-1‴), 73.2 (C-2‴), 77.7 (C-3‴), 69.8 (C-4‴), 76.4 (C-5‴), 60.6 (C-6‴), 99.7 (C-1⁗), 73.1 (C-2⁗), 77.1 (C-3⁗), 69.6 (C-4⁗), 76.4 (C-5⁗), 60.6 (C-6⁗). HRESIMS: *m/z* 773.2108 $[M+H]^+$ [16].

7,4′-Di-*O*-β-$_D$-glucosyl-kaempferol (**2b**, isolated yield 14.7%). ^{1}H NMR (400 MHz, DMSO-d_6): δ_H 8.19 (2H, d, *J*=8.9 Hz, H-2′/6′), 7.21 (2H, d, *J*=9.0 Hz, H-3′/5′), 6.84 (1H, d, *J*=2.2 Hz, H-8), 6.43 (1H, d, *J*=2.2 Hz, H-6), 5.08 (1H, d, *J*=7.3 Hz, H-1″), 5.01 (1H, d, *J*=6.9 Hz, H-1‴), 3.15–4.0 (protons in glucose); ^{13}C NMR (100 MHz, Methanol-d_4): δ_C 176.4 (C-4), 162.8 (C-7), 160.4 (C-5), 158.6 (C-4′), 155.9 (C-9), 146.7 (C-2), 136.9 (C-3), 129.3 (C-2′/6′), 124.3 (C-1′), 116.1 (C-3′/5′), 104.8 (C-10), 98.8 (C-6), 94.8 (C-8), 99.9 (C-1″), 73.2 (C-2″), 77.2 (C-3″), 69.7 (C-4″), 76.6 (C-5″), 60.7 (C-6″), 99.8 (C-1‴), 73.1 (C-2‴), 77.2 (C-3‴), 69.6 (C-4‴), 76.4 (C-5‴), 60.6 (C-6‴). HRESIMS: *m/z* 611.1585 $[M+H]^+$ [16].

Astragalin (**2c**, isolated yield 17.1%). ^{1}H NMR (600 MHz, Methanol-d_4): δ_H 8.09 (2H, d, *J*=6.8 Hz, H-2′/6′), 6.88 (2H, d, *J*=8.2 Hz, H-3′/ 5′), 6.42 (1H, s, H-8), 6.21 (1H, s, H-6), 5.14 (1H, d, *J*=7.7 Hz, H-1″), 3.42-3.80 (protons in galactose); ^{13}C NMR (100 MHz, Methanol-d_4): δ_C 179.7 (C-4), 166.1 (C-7), 163.1 (C-5), 161.6 (C-4′), 159.1 (C-9), 158.5 (C-2), 135.6 (C-3), 132.3 (C-2′/6′), 122.7 (C-1′), 116.1 (C-3′/5′), 105.7 (C-10), 105.0 (C-1″), 100.0 (C-6), 94.7 (C-8), 77.1 (C-3″), 75.0 (C-5″), 73.0 (C-2″), 70.0 (C-4″), 62.0 (C-6″). HRESIMS: *m/z* 449.1068 $[M+H]^+$ [17].

Biochemical properties of recombinant Ep7GT

To assay for the optimal reaction temperature, the reaction mixtures which contain pure protein were incubated at different temperatures (20–70 °C, pH 7.4). To study the optimal pH, the enzymatic reaction was performed in various reaction buffers at 37 °C with pH values in the range of 5.0–7.0 (citric acid-sodium citrate buffer), 7.0–9.0 (Tris-HCl buffer) and 9.0–11.0 (Na_2CO_3-$NaHCO_3$ buffer). To test the effects of divalent metal ions on Ep7GT activity, $BaCl_2$, $CaCl_2$, $CoCl_2$, $FeCl_2$, $CuSO_4$, $MgCl_2$, $SnCl_2$, $MnCl_2$ and $NiCl_2$ were used individually at final concentrations of 5 mM at 37 °C with pH 7.4; . The assays were performed with UDP-glucose as the donor and baohuoside (**1**) as the acceptor. The reactions were analyzed by HPLC-UV as described above and three independent experiments were performed.

Apparent K_m value assays

To compare the apparent *Km* values for UDP-glucose, UDP-xylose and UDP-*N*-acetylglucosamine using baohuoside (**1**) as sugar accepter, 19 μg (for UDP-glucose), 38 μg (for UDP-xylose) and 114 μg (for UDP-*N*-acetylglucosamine) of recombinant Ep7GT were incubated in a total volume of 100 μL at 55 °C for 30 min (for UDP-glucose and UDP-xylose, where the rate of product formation was determined to be linear) and 60 min (for UDP-*N*-acetylglucosamine) with various concentrations of sugar donors (200–800 μM) and a fixed concentration of baohuoside (**1**) (1 mM). To determine the apparent *Km* values for baohuoside (**1**) and kaempferol (**2**), 5 μg of recombinant Ep7GT was incubated in a total volume of 100 μL at 55 °C for 30 min with various concentrations of acceptors (20–400 μM) and a fixed concentration of UDP-glucose (400 mM). All experiments were performed in triplicate, and enzyme activity was evaluated by the production of related products (nmol) per mg of recombinant protein per minute. Apparent K_m values were calculated from Lineweaver–Burk plots (Figure S30).

Bioconversion of baohuoside (1) with engineered *E. coli*

The pET-28a vector carrying *Ep7GT* was transformed into *E. coli* strain Transetta (DE3), and the resulting transformant was used to biotransform baohuoside (**1**). One hundredth of the volume of the overnight-induced culture was inoculated in fresh LB media, and the culture was grown until the OD_{600} reached 0.8. The enzymes were induced by adding 1 mM IPTG to the culture, and the culture was incubated at 20 °C for 20 h. The cells were harvested by centrifugation and resuspended to obtain a cell concentration corresponding to an OD_{600} of 3.0 in 1 mL M9 medium containing 2% glucose[12]. After the substrate (200 μM) was added to the cells, the reaction mixture was incubated at 30 °C for 36 h. The reaction mixture (20 μL) was collected at 3, 6, 9, 12, 24 and 36 h. The samples were terminated by the addition of 100 μL of ice cold MeOH and centrifuged. The supernatant was used for HPLC analysis. Three independent experiments were performed, and the quantitative analysis was further defined by the external standard method. The HPLC peak areas were integrated using Agilent Workstation Software. The linear regression models between the peak area and the quantity of standards were established for quantitative analysis of the bioconversion. The models and the regression equations are given in Figure S31.

Accession numbers

The nucleotide sequences of *Ep7GT* (*EpGT8*) are available in the GenBank database under accession number MH261365.

Conclusions

In summary, Ep7GT, a novel flavonoid GT from *E. pseudowushanense* that catalyses the 7-*O*-glucosylation of baohuoside (**1**) to form icariin (**1a**) in the biosynthesis of icariin (**1a**) was characterized. It possesses sugar donor substrate flexibility and can be used as an environmentally friendly and efficient biocatalyst for the preparation of natural bioactive flavonol glycosides, icariin (**1a**) and its derivatives. Icariin (**1a**) was also economically synthesized from baohuoside (**1**) by whole-cell transformation with engineered *E. coli* harbouring *Ep7GT*. Furthermore, the glycosylation module in this study might serve as a component of biosystems for the production of bioactive compounds such as icariin (**1a**) from kaempferol (**2**) or other precursors by further metabolic engineering work. The work of identifying the prenyltransferase responsible in the biosynthesis of icariin is in progress.

Journal Name ARTICLE

View Article Online
DOI: 10.1039/C9OB01352K

Conflicts of interest

The authors declare no competing financial interest.

Acknowledgements

This work was financially supported by the Drug Innovation Major Project (2018ZX09711001-006), National Natural Science Foundation of China (No. 81573317), CAMS Innovation Fund for Medical Sciences (Nos. CIFMS-2016-I2M-3-012 and 2017-I2M-3-013), Natural Science foudation of Shaanxi Province, China (2019JQ-963), and Beijing Key Laboratory of Non-Clinical Drug Metabolism and PK/PD Study (No. Z141102004414062).

References

1 H. Ma, X. Ma, B. Ge, P. Zhen, J. Zhou, Y. Gao, C. Xian and K. Chen, *Cell Prolif.*, 2014, **47**, 527–539.
2 C. Xu, B. Liu, J. Wu, Y. Xu, X. Duan, Y. Cao and J. Dong, *Eur. J. Pharmacol.,* 2010**, 642***,* 146–153.
3 Y. Zhai, X. Guo, B. Ge, P. Ma. X. Zhen, J. Zhou, H. Ma, C. Xian and K. Chen, *Bone,* 2014, **66**, 189–198.
4 M. Chen, J. Hao, Q. Yang and G. Li. *Molecules,* 2014, **19**, 9502–9514.
5 X. Wang, J. Li, L. Qian, X. Zang, S. Zhang, X. Wang, J. Jin, X. Zhu, X. Zhang, Z. Wang and Y. Xu, *Neuroscience,* 2013, **236**, 281–288.
6 D. Liu, Y. Ye, L. Xu, W. Yuan and Q. Zhang, *Biomed. Pharmacother.*, 2018, **108**, 663–669.
7 A. W. Shindel, Z. C. Xin, G. Lin, T. M. Fandel, Y. C. Huang, L. Banie, B. N. Breyer, M. M. Garcia, C. S. Lin and T. F. Lue, *J. Sex. Med*., 2010, **7**, 1518–1528.
8 J. Wu, J. Du, C. Xu, J. Le, Y. Xu, B. Liu and J. Dong, *Pharmacol. Biochem. Behav.,* 2011, **98**, 273–278.
9 L. Sun, W. Chen, L. Qu, J. Wu and J. Si, *Mol. Med. Rep*., 2013, **8**, 1883–1887.
10 Q. Mei, C. Wang, Z. Zhao, W. Yuan and G. Zhang, *Beilstein J. Org. Chem.*, 2015, **11**, 1220–1225.
11 J. Pan and B. Guo. *Molecules*, 2016, **21**, 1475.doi: 10.3390/molecules21111475.
12 K. Feng, R. Chen, K. Xie, D. Chen, B. Guo, X. Liu, J. Liu, M. Zhang and J. Dai, *Org. Biomol. Chem.* 2018, **16**, 452–458.
13 P. Parajuli, R. P. Pandey, N. T. H. Trang, T. J. Oh and J. K. Sohng, *Carbohydr. Res.*, 2015, **418**, 13–19.
14 P. Jones, B. Messner, J. I. Nakajima, A. R.Schäffner and K. Saito, *J. Biol. Chem.*, 2003, **278**, 43910–43918.
15 Q. Xia, D. Xu, Z. Huang, J. Liu, X.Wang, X. Wang and S. Liu, *Fitoterapia*, 2010, **81**, 437–442.
16 L. O. Manguro, I. Ugi, P. Lemmen and R. Hermann, *Phytochemistry,* 2003, **64**, 891–896.
17 S. M. Kim, K. Kang, E. H. Jho, Y. J. Jung, C. W. Nho, B. H. Um and C. H. Pan, *Phytother. Res.*, 2011, **25**,1011–1017.