Electronic Supplementary Material (ESI) for RSC Advances.

Supplementary Information

Systems metabolic engineering of *Escherichia coli* to enhance the production of flavonoid glucuronides

Yan Yang,[a,b] Hui-Min Wang,[b] Yuan-Feng Tong,[b] Min-Zhi Liu,[b] Ke-di Cheng,[b] Song Wu[a,b*] and Wei Wang[a,b*]

Table of contents

1. Supplementary Methods 2

1.1 Plasmid construction 2

1.2 Determination of divalent metal ions, pH and temperature on enzyme activity 3

1.3 Determination of UDP-glucuronic acid yield in the engineered *Escherichia coli* strains 3

1.4 HPLC-based separation method 4

1.5 The functional analysis of engineered *E. coli* strain BPGUT harboring recombinant SbUGT 4

1.6 The procedure for hydrolysis reaction using β-glucuronidase 5

2. Supplementary Tables 6

3. Supplementary Figures 8

4. Supplementary Results 15

4.1. Identification of glucosylation and glucuronosylation products 15

4.2 Protein purification 17

4.3 Effects of divalent metal ions, pH and temperature 18

4.4 UDP-glucuronic acid yield in the engineered *Escherichia coli* strains 19

4.5 I*n vivo* functional characterization of recombinant SbUGT 19

4.6 The hydrolysis of flavonoid glucuronides by β-glucuronidase 20

5. Additional Supplementary Figures 22

6. Supplementary References 38

1. Supplementary Methods

1.1 Plasmid construction

1.1.1 Construction of plasmids containing *pgm, galU,* and *pgm-T7-galU* fusion genes

The DNA fragment encoding phosphoglucomutase (*pgm*) was amplified by polymerase chain reaction (PCR) from *E. coli* BL21 (DE3) genomic DNA using two pairs of primers (Pgm-1 to Pgm-4) derived from the *pgm* nucleotide sequence (GenBank: EG12144). The *pgm* fragment, digested with *Nco* I and *Nhe* I after verifying by sequencing, was subcloned into the expression vector pET-28a (Novagen), resulting in pET28a-Pgm. The *pgm* fusion fragment with the ribosome binding site and *T7* terminator, obtained by digestion of pET28a-Pgm with *Xba* I and *Xho* I, was inserted into the pSLB208/EG12 vector [1] to give plasmid pSLB208-Pgm.

The 1036 bp DNA fragment, encoding *E. coli* glucose-1-phosphate uridyltransferase (*galU*) gene, was obtained as that of *pgm* DNA fragment together with the two pairs of primers (GalU-1 to GalU-4) designed according to the *galU* nucleotide sequence (GenBank: NP_415752). The DNA fragment was firstly subcloned into pET28a vector to give the pET28a-GalU confirmed by DNA sequencing, and then the *galU* fragment digested with *Xba* I and *Xho* I was ligated into plasmid pSLB208/EG12 [1] to generate pSLB208-GalU.

In order to keep the uniform expression of Pgm and GalU and the high coupling of activity, it would be necessary to design a gene concatemer of which two genes contain *T7* promoter and *T7* terminator, respectively. The partial *pgm-T7* fusion fragment was amplified by SOE-PCR using the template of plasmid pET28a-Pgm and seven primers (PgmGalU-1 to PgmGalU-7). The resulting DNA fragment was cloned into the pET28a-Pgm vector to give pET28a-Pgm-T7. Subsequently, the *galU* DNA fragment, obtained from pET28a-GalU by the *Spe* I and *Bam*H I double digestion, was inserted into *Nhe* I and *Bam*H I sites of pET28a-Pgm-T7 to generate pET28a-Pgm-T7-GalU. The recombinant *pgm-T7-galU* gene fragment, isolated from plasmid pET28a-Pgm-T7-GalU by digestion with *Xba* I and *Xho* I, was inserted

into the pSLB208/EG12 vector [1] to give plasmid pSLB208-Pgm-T7-GalU.

1.1.2 Construction of plasmid with the *ugd* fusion gene

The UDP-glucose 6-dehydrogenase (*ugd*) gene was also generated by PCR as that of *pgm* DNA fragment along with two pairs of primers (Ugd-1 to Ugd-4) designed against the *ugd* sequence (GenBank: EG13407). The DNA fragment (about 1.4 kb), digested with *Nde* I and *Sal* I, was ligated into *Nde* I and *Xho* I sites of pEG12, [1] resulting in pEG-Ugd.

1.2 Determination of divalent metal ions, pH and temperature on enzyme activity

To study the necessity of divalent metal ions of SbUGT activity, $CaCl_2$, $CoCl_2$, $CuCl_2$, $MgCl_2$, $MnCl_2$, and $ZnCl_2$ were used individually in the final concentration of 5 mM. To assay for the optimal pH, the enzymatic reaction was performed in various reaction buffers with pH values in the range of 4.0–6.5 (Citric acid-sodium citrate buffer), 6.5–8.0 (PBS-Na buffer) and 8.0–9.0 (Tris-HCl buffer). To test the optimal reaction temperature, the reaction mixtures were incubated at different temperatures (4–60 °C). The assays were performed with UDPGA as a donor and baicalein (**1**) as an acceptor.

1.3 Determination of UDP-glucuronic acid yield in the engineered *Escherichia coli* strains

The engineered *E. coli* strains with the upstream and downstream modules were pre-cultured in LB broth. The overnight cultured cells were harvested and resuspended in M9 containing 2% glucose. At this step, cell density was kept at an OD_{600} of 3. The DMSO (20 μL) were added to the reaction system, which was then incubated at 30°C for 24 h. The cells were pelleted by centrifugation and resuspended in sterile water and then sonicated for 5 min in 5-s bursts at 6-s intervals. After centrifugation at 13,500 rpm for 10 min, the supernatant were analysed by analytical reverse-phase HPLC.

1.4 HPLC-based separation method

20 µL of supernatant was used for analysis on a Thermo HPLC system equipped with a Hibar® C18 reverse-phase column (Hibar, 4.60×250 mm, 5 µm particle size). The mobile phase was water with 0.05% TFA (A) and acetonitrile with 0.05% TFA (B) with a gradient program. The analysis of baicalein (**1**), scutellarein (**4**), luteolin (**6**), apignin (**7**), quercetin (**12**), genistein (**15**), daidzein (**16**), 4'-demethycalycosin (**17**), naringenin (**21**), robtin (**22**) and their corresponding glucuronides were conducted as following gradient program: 0 min (B, 20%), 25 min (B, 45%), 27 min (B, 100%), 32 min (B, 100%), 35 min (B, 20%). The separation of wogonin (**2**), oroxylin A (**3**), chrysin (**5**), kaempferol (**11**), myricetin (**13**), biochanin A (**14**), formononetin (**18**), and calycosin (**19**) and their glucuronides were achieved by a 35 min linear gradient from 20% to 60% B. It was followed by isocratic flow with 100% B for 5 min to wash off hydrophobic compounds and further an isocratic development with 20% B for 5 min to equilibrate the column. The analysis of orientin (**8**), vitexin (**9**), isovitexin (**10**) and puerarin (**20**) were carried out as below: a 25 min linear gradient from 10% to 35% B was used for separation and then followed by an isocratic flow with 100% B for 5 min and 10% B applied to elute and equilibrate the column, respectively.

The analysis of UDPGA was performed on a a Thermo HPLC system equipped with a Shodex Axpak WA-624 anion exchange column. 50 µL of injection was run at 1 ml/min of flow rate. The mobile phase was water (A) and 700 mM CH_3COONH_4 buffer (pH = 5.3 with CH_3COOH) (B). Analysis was carried out as following gradient program: 0 min (B, 2%), 10 min (B, 25%), 35 min (B, 100%), 40 min (B, 100%), 42 min (B, 2%), and 45 min (B, 2%). Detection was accomplished by measuring the UV absorption at 262 nm, and standard curves of UDPGA for quantification was also created under the same condition.

1.5 The functional analysis of engineered *E. coli* strain BPGUT harboring recombinant SbUGT

To assay for the sugar acceptor specificity of recombinant SbUGT in engineered *E. coli* strain, the

plasmid pACYC184-SbUGT-W112 and pACYC184-SbUGT-W112G76 were constructed to co-express with other three enzymes of upper module in host cells. Assays of the sugar acceptor specificity of the engineered *E. coli* strains BPGUT harboring SbUGT-W112 and SbUGT-W112G76 were performed using a standard assay process described in the text with the baicalein (**1**), apignin (**7**), 4'-demethycalycosin (**17**), and naringenin (**21**) used as sugar acceptors.

1.6 The procedure for hydrolysis reaction using β-glucuronidase

The standard assays consisted of 100 mM NaAc buffer (pH 5.0), substrate (diverse flavonoid glucuronide, about 50 μM), and 100 U of β-glucuronidase solution (dissolved in 0.2% NaCl) from bovine liver in a final volume of 500 μL. The substrates include baicalin (**1**a), oroxylin A-7-*O*-glucuronide (**3**P), scutellarein-7-*O*-glucuronide (**4**P), kaempferol-7-*O*-glucuronide (**11**P), 4'-demethycalycosin-7-*O*-glucuronide (**17**P) and narigenin-7-*O*-glucuronide (**21**P). The sample mixtures were incubated at 30°C for 15 or 30 min, and the enzymatic reaction was terminated by the addition of 500 μL of methanol. Identification of the enzyme reaction products were carried out by HPLC as described in part 1.4.

2. Supplementary Tables

Table S1. Sequence of PCR primers used in this study.

Gene	Primer	Sequence (5' to 3')	Endonuclease site (underlined)
pgm	Pgm-1	ACGTTGCAGACAAAGGACAAAGCA	
	Pgm-2	GATATACCATGGCAATCCACAATCGTGCAG	*Nco* I
	Pgm-3	TGTGTGGCTAGCTTACGCGTTTTTCAGAACTTCGCTAAC	*Nhe* I
	Pgm-4	GCGTAGCGCATCAGGCAATTCTGT	
galU	GalU-1	GGGATGCGATACAGAAATATGAAC	
	GalU-2	GGAGAAACTAGTATGGCTGCCATTAATACGAAAGTC	*Spe* I
	GalU-3	GTCATTGGGATCCGTCCGGTTTAAGACAATTTAATAAG	*Bam*H I
	GalU-4	GCACTTGCTTAAAATCCCGCCAGC	
pgm-T7-galU	PgmGalU-1	ACTGCGAAAGCTTCCTCGGTGA	*Hind* III
	PgmGalU-2	AGCTTCCTTTCGGGCTTTGTTACGCGTTTTTCAGAACTTCG	
	PgmGalU-3	TAACAAAGCCCGAAAGGAAGCT	
	PgmGalU-4	CCTATAGTGAGTCGTATTAACAAAAAACCCCTCAAGACC	
	PgmGalU-5	TTAATACGACTCACTATAGG	
	PgmGalU-6	TCGAATTCGGATCCGCGACCCATGCTAGCCATGGTATATCTCCTTCTTAAAG	*Bam*H I and *Nhe* I
	PgmGalU-7	GGTGGTGCTCGAGTGCGGCCGCGTCGACGGAGCTCGAATTCGGATCCGCGAC	*Xho* I
ugd	Ugd-1	AATAAATATCAGCTATTCTTATAAAGAAAATCTG	
	Ugd-2	GGATCCCATATGAAAATCACCATTTCCGG	*Nde* I
	Ugd-3	AAGCTTGTCGACGGAGCTCGGATCCTAGTAAATCAATCAAATCAATCTGTTC	*Sal* I
	Ugd-4	CATCTTGCCACGCCACAACTGCACT	
SbUGT	UGT-1	CGAGGACACTGACATGGACTGGACAAGGCCATGGAAGACACACTTGTGATCTACACAAC	*Nco* I
	UGT-2	GACAAGGCCATGGAAGACACACTTGTGATCTACACAACGCCGGAGCACAT	*Nco* I
	UGT-3	CAGTGTACTCGAGTTAATCCCGAGTGGCGTGAAGAAA	*Xho* I
	UGT-4	AATCCCACAATTTCTCATCTTACC	
	UGT-5	CCCTCTTTAAATCACTCATAAATCG	
MCS	Twin1-B1	GATATACCATGGGCAGCAGCCATCAT	
	Twin1-B2	TCAGTAAGATCTTTAGCAGCCGGATCTCAGTG	

Table S1. Continued.

Gene	Primer	Sequence (5' to 3')	Endonuclease site (underlined)
W112G76	W76-1	GAGCTCCCTCGTCTCAGCAAC	
	W76-2	GCTCTGCAACAGATCTCCCAG	*Bgl* II
	W76-3	TTAACGAGGATGCCTGCCGAGCCGCGCAGGTTCTTCGAAAT	
	W76-4	TCGGCAGGCATCCTCGTTAA	
	W76-5	CAGCGCCACCTGCAGATCATC	*Pst* I
	W76-6	GTTTCTCCAACTCAGCCGCCG	

3. Supplementary Figures

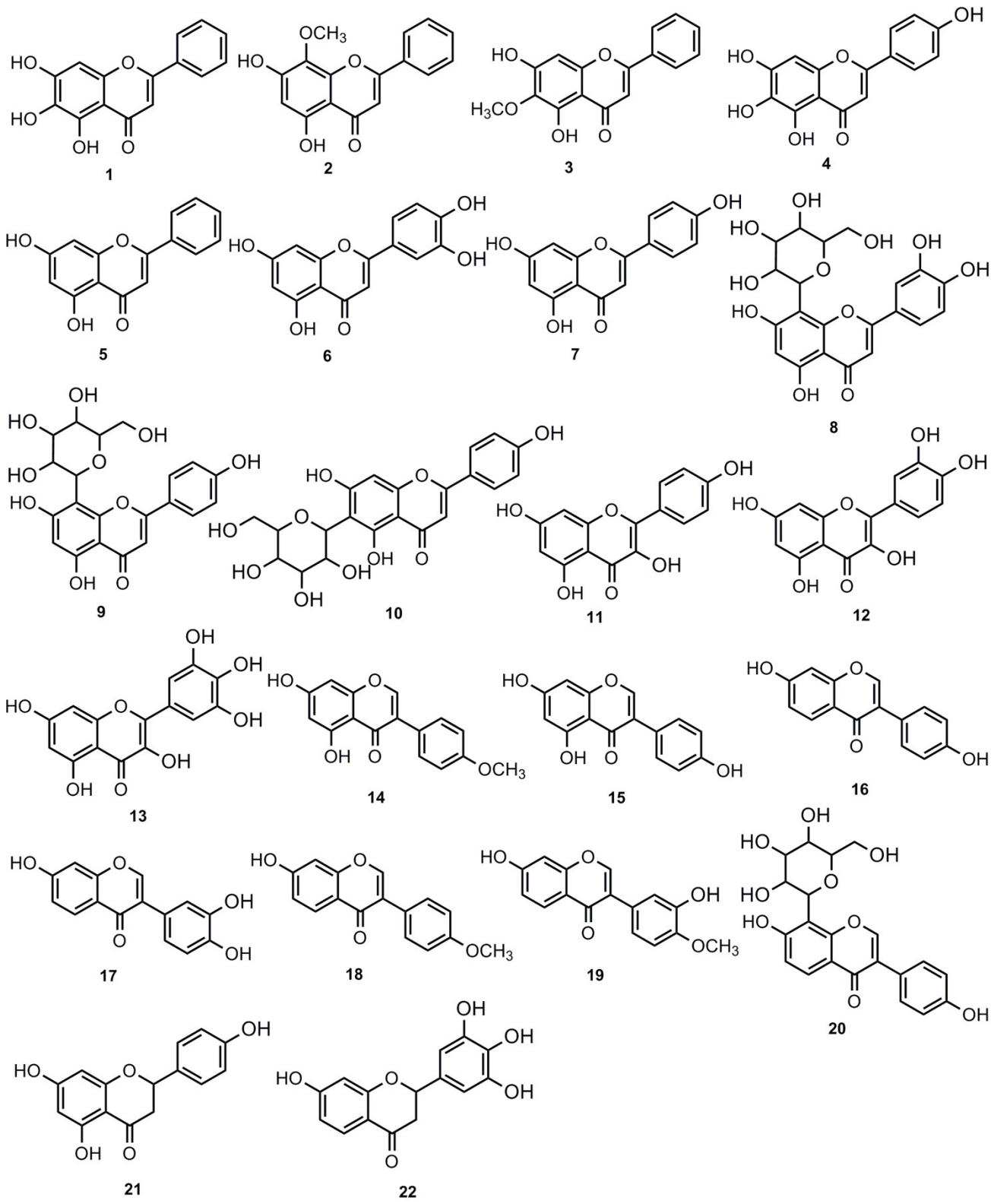


Fig. S1 List of flavonoid substrates used in this study. The name of the structures are as follows: Baicalein, **1**; Wogonin, **2**; Oroxylin A, **3**; Scutellarein, **4**; Chrysin, **5**; Luteolin, **6**; Apigenin, **7**; Orientin, **8**; Vitexin, **9**; Isovitexin, **10**; Kaempferol, **11**; Quercetin, **12**; Myricetin, **13**; Biochanin A, **14**; Genistein, **15**; Daidzein, **16**; 4'-Demethycalycosin, **17**; Formononetin, **18**; Calycosin, **19**; Puerarin, **20**; Narigenin, **21**; Robtin, **22**.

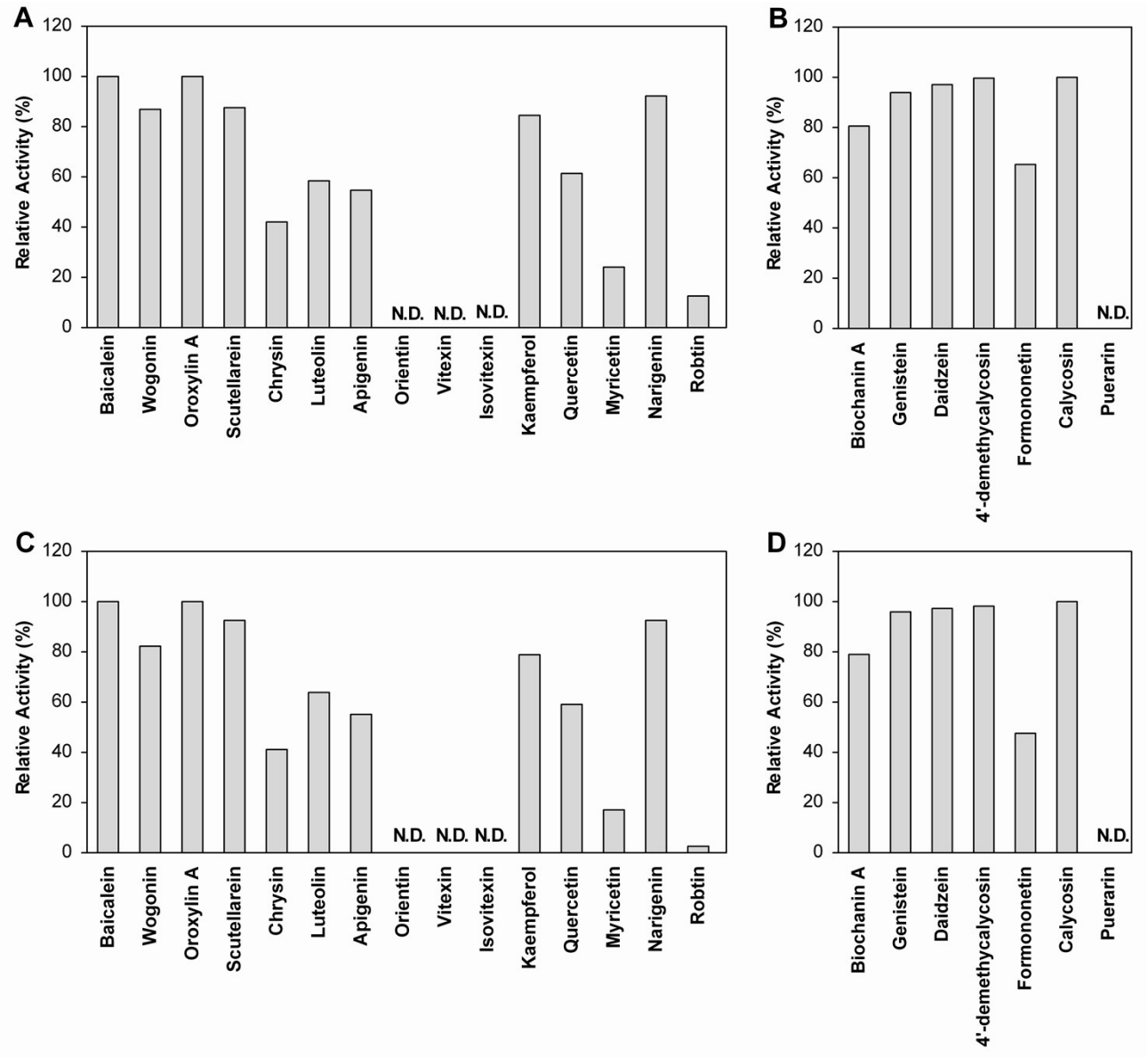


Fig. S2 Sugar acceptor specificity of of SbUGT from *S. baicalensis* Georgi. (A) The relative enzyme activities of recombinant SbUGT-W76 on flavones and flavanones detected at 280 nm, (B) The relative enzyme activities of recombinant SbUGT-W76 on isoflavones detected at 254 nm, (C) The relative enzyme activities of recombinant SbUGT -W112 on flavones and flavanones detected at 280 nm, and (D) The relative enzyme activities of recombinant SbUGT -W112 on isoflavones detected at 254 nm. The highest specific activity is set as 100%. N.D. indicates “not detected”.

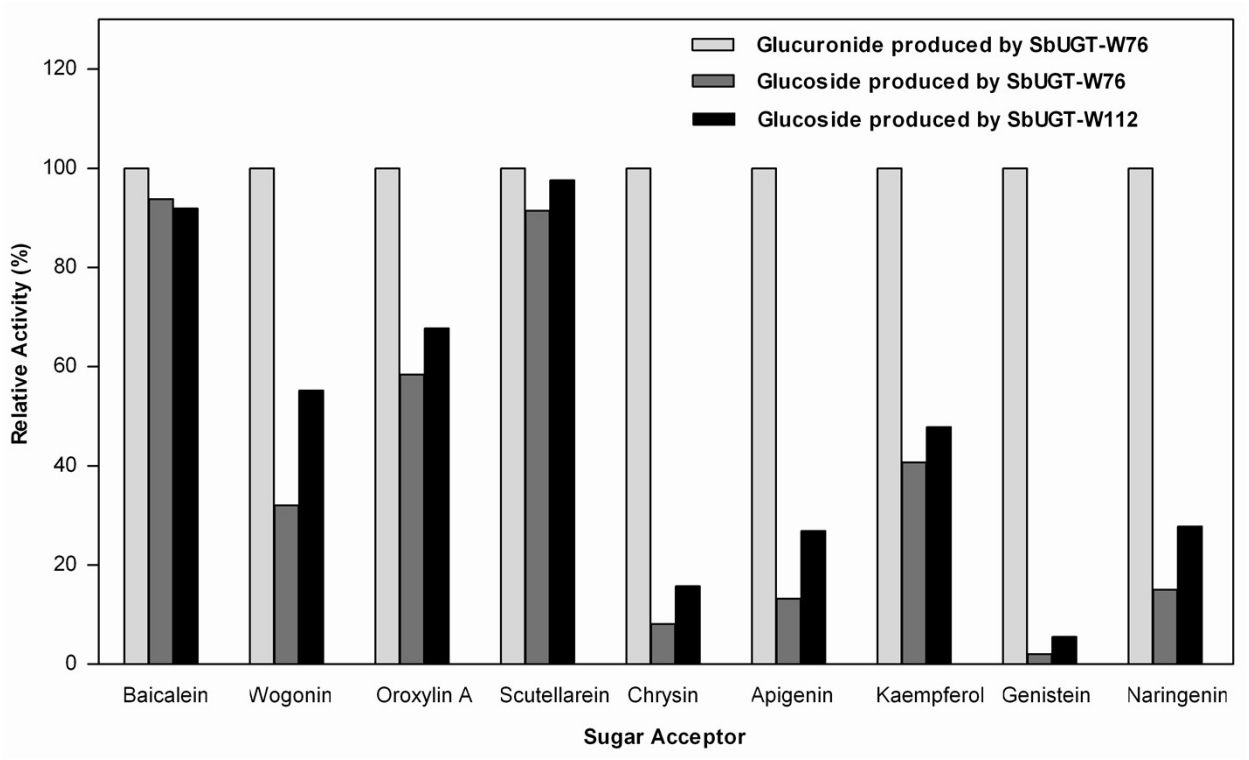


Fig. S3 Relative sugar donor specificity of SbUGT-W76 and SbUGT-W112. The glucuronosylating activity toward respective substrate is set as 100%.

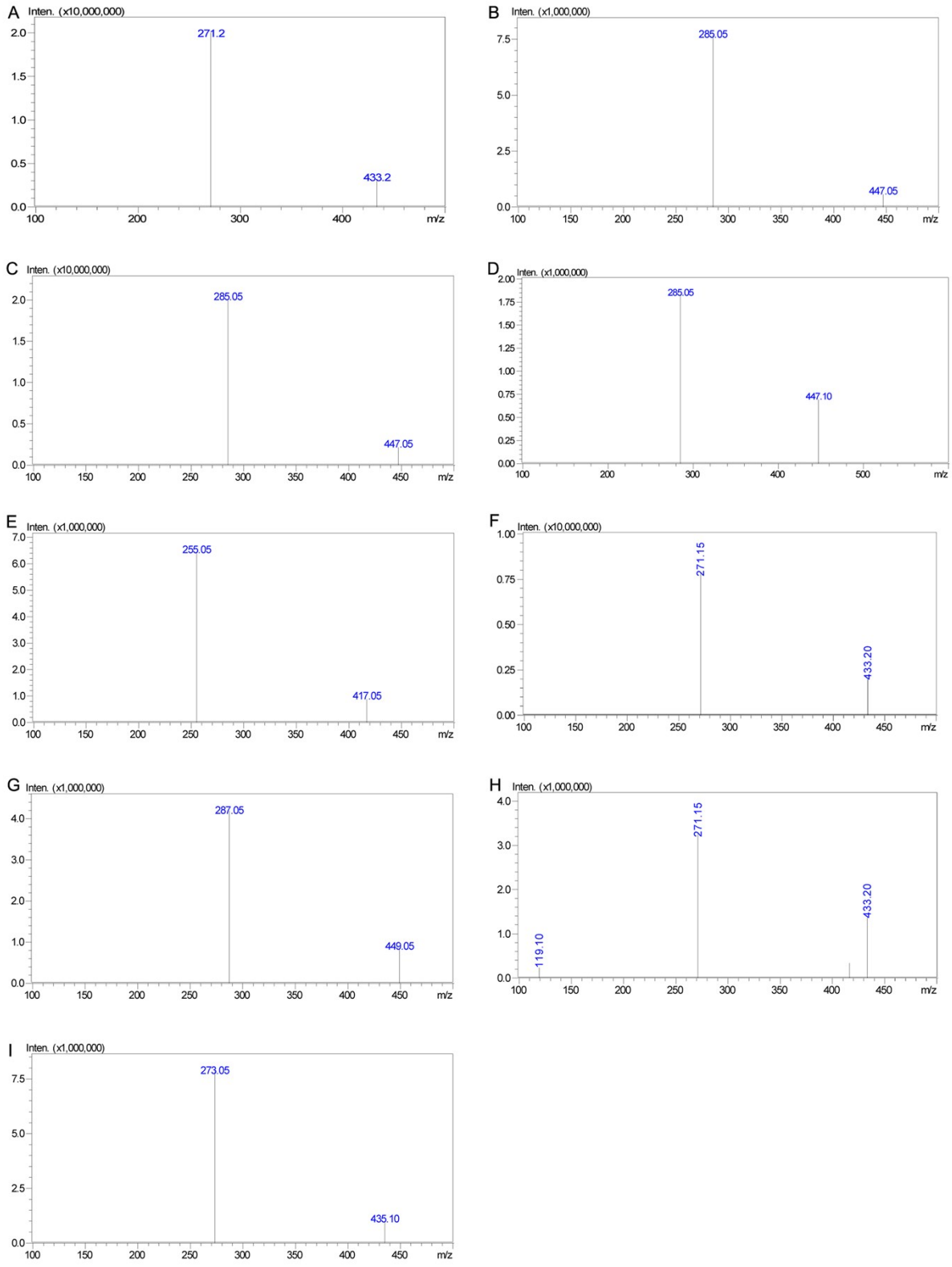


Fig. S4 The MS/MS spectra for glucosides of Baicalein (A), Wogonin (B), Oroxylin A (C), Chrysin (E), Apigenin (F), Kaempferol (G), Genistein (H), and Naringenin (I) were detected in positive ionisation mode, and Scutellarein (D) was detected in negative ionisation mode.

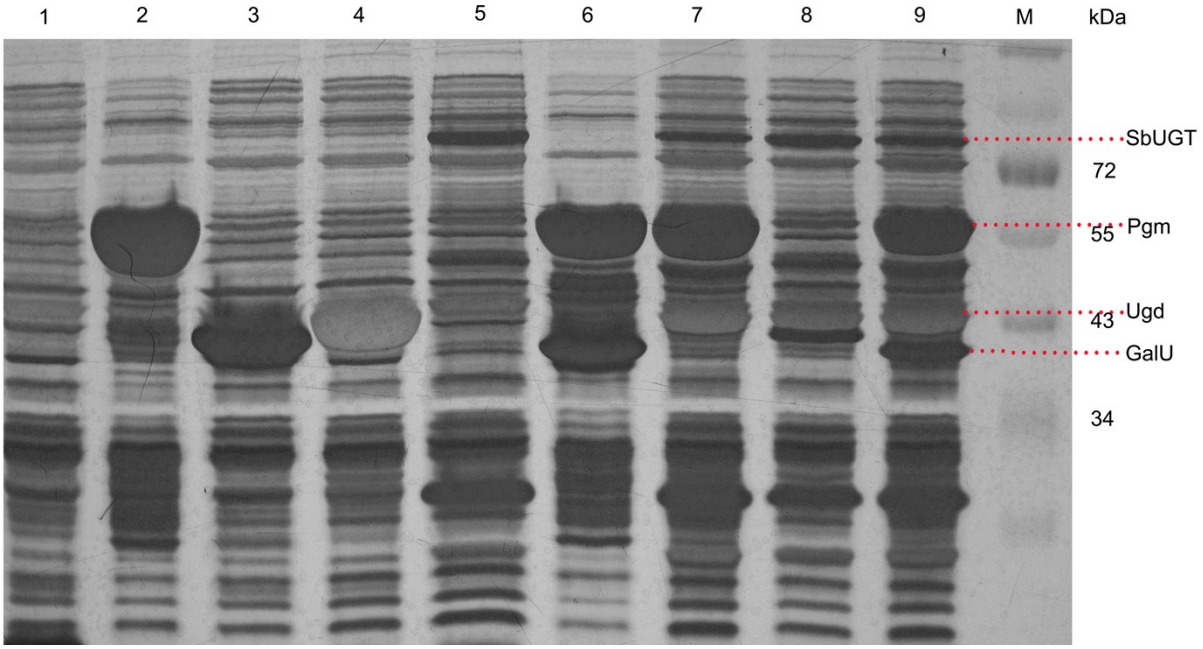


Fig. S5 SDS-PAGE analysis of recombinant Pgm, GalU, Ugd and SbUGT expressed in *E. coli.* Lane 1, The soluble fraction of *E. coli* DE3 harboring empty vector pET-28a, Lane 2, The soluble fraction of Pgm; Lane 3, The soluble fraction of GalU; Lane 4, The soluble fraction of Ugd; Lane 5, The soluble fraction of SbUGT-W76; Lane 6, The soluble fraction of Pgm and GalU; Lane 7, The soluble fraction of strain BPUT; Lane 8, The soluble fraction of strain BGUT; and Lane 9, The soluble fraction of strain BPGUT.

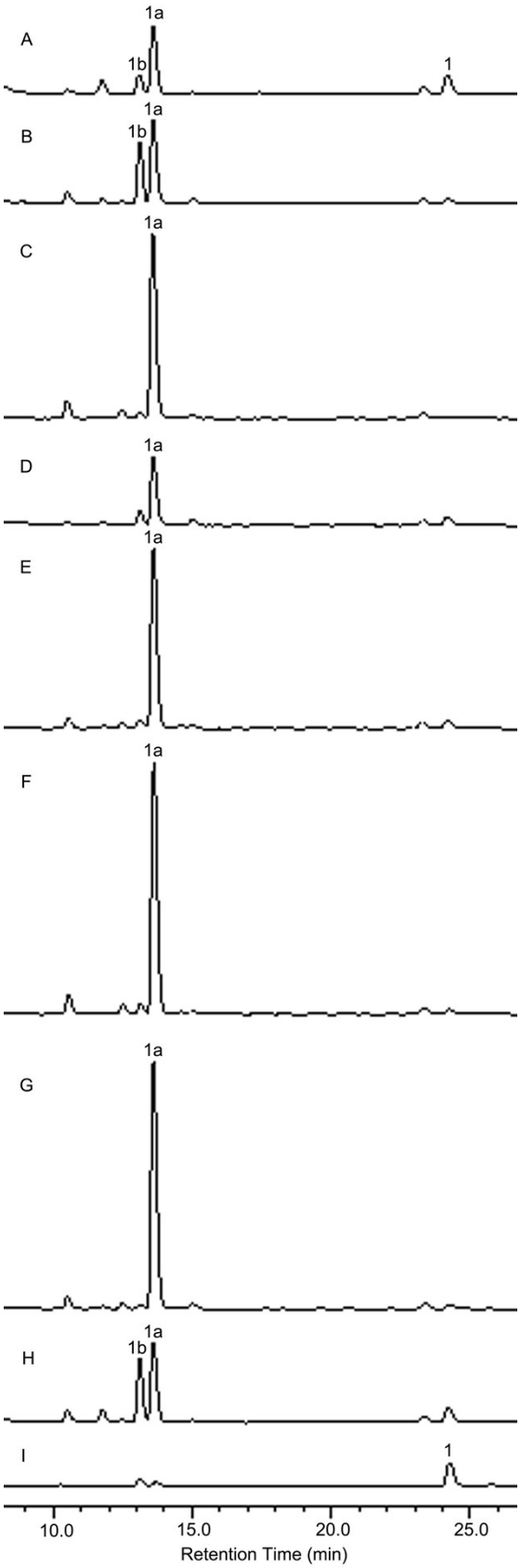


Fig. S6 The functional characterization of recombinant SbUGT-W76 in engineered *E. coli* strains with baicalein (**1**) as sugar acceptor. (A) Strain BPT, (B) Strain BGT, (C) Strain BUT, (D) Strain BPGT, (E) Strain BPUT, (F) Strain BGUT, (G) Strain BPGUT, (H) Strain BT, and (I) Strain BPGU.

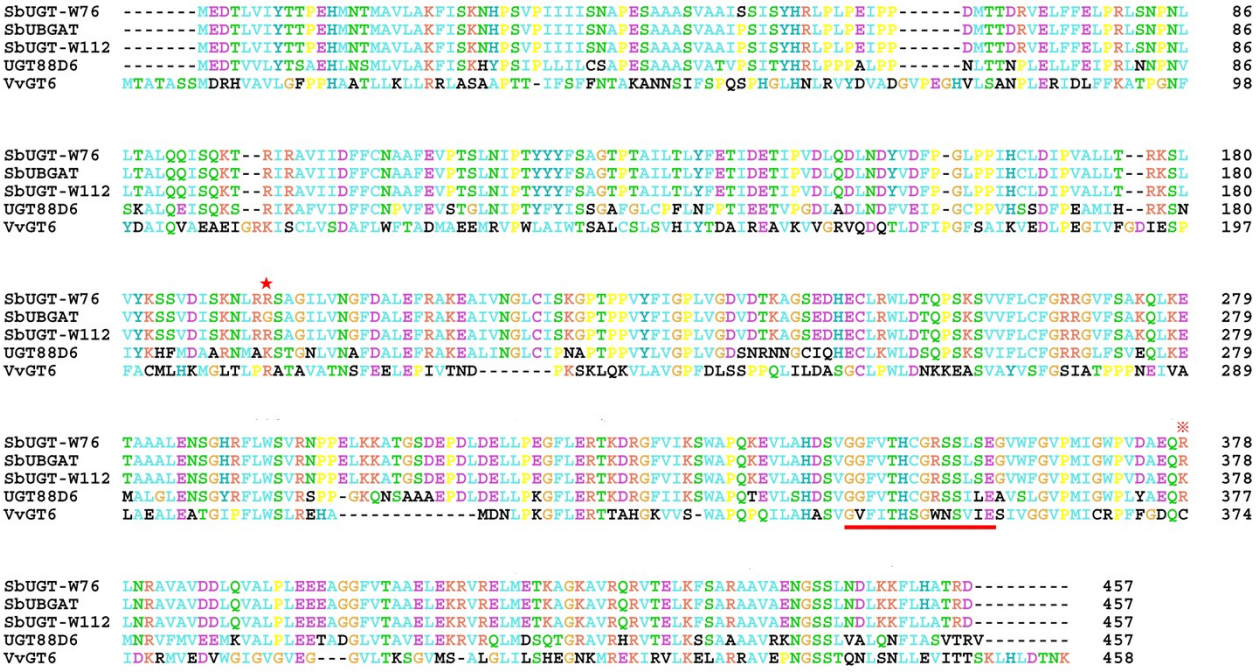


Fig. S7 A multiple alignment of the amino acid sequences of four UGTs from *Scutellaria* species along with a bifunctional VvGT6. The Arg-194 residue of SbUGT-W76 and SbUGT-W112 is marked with asterisk (★), and the Arg-378 residue of SbUGT-W76 and Lys-378 residue of SbUGT-W112 are indicated with symbol (※). The plant secondary product glycosyltransferase (PSPG) consensus sequence is underlined. Abbreviations: SbUBGAT, *S. baicalensis* Georgi UDP-dependent glucuronosyltransferase (BAH19313); UGT88D6, *S. indicum* UDP-glycosyltransferase (BAG31947); VvGT6, *Vitis vinifera* UDP-sugar flavonoid glucosyltransferase (NP_001267832).

4. Supplementary Results

4.1. Identification of glucosylation and glucuronosylation products

The 7-*O*-glucuronide products of baicalein (**1**), wogonin (**2**), oroxylin A (**3**), scutellarein (**4**) and apigenin (**7**) were identified by comparison of their LC spectrum, MS and MS/MS fragments with authentic samples. Positive ESI mass spectra were obtained from a triple quadrupole mass system (Shimadzu, Japan). ^{1}H NMR and ^{13}C NMR were recorded on Varian NMR system 600 spectrometers (Varian Inc., USA).

Baicalein-7-*O*-glucoside (1b): ESI-MS: *m/z* = 433 $[M+H]^+$, ESI-MS/MS: *m/z* = 271 $[M+H\text{-}C_6H_{10}O_5]^+$; ^{1}H NMR (DMSO-$d_6$, 600 MHz) δ = 6.99 (s, 1H; H-3), 6.92 (s, 1H; H-8), 8.02 (m, 2H; H-2', H-6'), 7.60 (m, 3H; H-3', H-4', H-5'), 4.59 (d, $J_{1",2"}$ = 7.8 Hz, 1H; H-1"), 3.23–3.69 (m, 3H; H-2", H-3", H-4"), 3.95 (m, 2H; H-5", H-6a"), 3.15 (m, 1H, H-6b"), 12.5 ppm (5-OH); ^{13}C NMR (DMSO-d_6, 150 MHz) δ = 163.4 (C-2), 104.7 (C-3), 182.5 (C-4), 146.5 (C-5), 131.6 (C-6, N. D.), 151.6 (C-7), 94.2 (C-8), 149.1 (C-9), 106.1 (C-10), 130.8 (C-1'), 126.4 (C-2', C-6'), 129.1 (C-3', C-5'), 132.0 (C-4'), 100.9 (C-1"), 73.1 (C-2"), 77.3 (C-3"), 69.6 (C-4"), 75.8 (C-5"), 60.6 ppm (C-6"). [2] N.D. indicates “not detected”.

Chrysin-7-*O*-glucuronide (5P): ESI-MS: *m/z* = 431 $[M+H]^+$, ESI-MS/MS: *m/z* = 255 $[M+H\text{-}C_6H_8O_6]^+$; ^{1}H NMR (DMSO-d_6, 600 MHz) δ = 7.07 (s, 1H; H-3), 6.50 (d, $J_{6,8}$ = 1.8 Hz, 1H; H-6), 6.92 (d, $J_{6,8}$ = 1.8 Hz, 1H; H-8), 8.10 (d, $J_{2',3';\ 5',6'}$ = 7.2 Hz, 2H; H-2', H-6'), 7.63 (m, 3H; H-3', H-4', H-5'), 5.27 (d, $J_{1",2"}$ = 6.6 Hz, 1H; H-1"), 3.02–3.67 (m, 3H; H-2", H-3", H-4"), 4.04 ppm (d, $J_{4",5"}$ = 9.0 Hz, 1H; H-5"); ^{13}C NMR (DMSO-d_6, 150 MHz) δ = 163.1 (C-2, N.D.), 105.5 (C-3), 182.2 (C-4), 161.1 (C-5), 99.6 (C-6), 163.7 (C-7), 94.8 (C-8), 157.6 (C-9), 105.8 (C-10, N.D.), 130.6 (C-1'), 129.2 (C-2', C-6'), 126.5 (C-3', C-5'), 132.2 (C-4'), 99.2 (C-1"), 72.7 (C-2"), 74.8 (C-3"), 71.3 (C-4"), 75.7 (C-5"), 170.5 ppm (C-6"). [3] N.D. indicates “not detected”.

Kaempferol-7-*O*-glucuronide (11P): ESI-MS: *m/z*: 463 $[M+H]^+$, ESI-MS/MS: *m/z*: 287 $[M+H\text{-}C_6H_8O_6]^+$; ^{1}H NMR (DMSO-d_6, 600 MHz) δ = 6.44 (d, $J_{6,8}$ = 2.4 Hz, 1H; H-6), 6.84 (d, $J_{6,8}$ = 2.4 Hz, 1H; H-8), 8.08 (d, $J_{2',3';\ 5',6'}$ = 9.0 Hz, 2H; H-2', H-6'), 6.97 (d, $J_{2',3';\ 5',6'}$ = 9.0, 2H; H-3', H-5'), 5.22 (d, $J_{1",2"}$ = 7.2 Hz, 1H; H-1"),

3.47–3.68 (m, 3H; H-2", H-3", H-4"), 3.96 ppm (s, 1H; H-5"); ^{13}C NMR (DMSO-d_6, 150 MHz) δ = 147.7 (C-2), 136.0 (C-3), 176.1 (C-4), 160.4 (C-5), 98.8 (C-6), 162.3 (C-7), 94.3 (C-8), 159.7 (C-9), 104.7 (C-10), 121.5 (C-1'), 129.6 (C-2', C-6'), 115.5 (C-3', C-5'), 155.7 (C-4'), 99.2 (C-1"), 72.8 (C-2"), 75.0 (C-3"), 71.4 (C-4"), 75.8 (C-5"), 170.3 ppm (C-6"). [4]

Quercetin-7-*O*-glucuronide (12P): ESI-MS: *m/z* = 479 $[M+H]^+$, ESI-MS/MS: *m/z* = 303 $[M+H-C_6H_8O_6]^+$; ^{1}H NMR (DMSO-d_6, 600 MHz) δ = 6.19 (d, $J_{6,8}$ = 1.8 Hz, 1H; H-6), 6.45 (d, $J_{6,8}$ = 1.8 Hz, 1H; H-8), 7.70 (d, $J_{2',6'}$ = 1.8 Hz, 1H; H-2'), 7.22 (d, $J_{5',6'}$ = 9.0 Hz, 1H; H-5'), 7.60 (dd, $J_{2',6';\ 5',\ 6'}$ = 1.8, 9.0 Hz, 1H; H-6'), 5.02 (br s, 1H; H-1"), 3.16–3.50 (m, 3H; H-2", H-3", H-4"), 3.96 ppm (s, 1H; H-5"); ^{13}C NMR (DMSO-d_6, 150 MHz): δ = 146.7 (C-2, overlapped), 136.5 (C-3), 175.9 (C-4), 160.7 (C-5), 98.2 (C-6), 164.1 (C-7), 93.5 (C-8), 156.2 (C-9), 103.0 (C-10), 125.3 (C-1'), 115.4 (C-2'), 145.9 (C-3'), 146.7 (C-4', overlapped), 116.5 (C-5'), 119.3 (C-6'), 101.5 (C-1"), 72.0 (C-2"), 75.9 (C-3"), 71.3 (C-4", N.D.), 73.1 (C-5"), 171.1 ppm (C-6"). [5] N.D. indicates "not detected".

Myricetin-7-*O*-glucuronide (13P): ESI-MS: *m/z*: 495 $[M+H]^+$, ESI-MS/MS: *m/z*: 319 $[M+H-C_6H_8O_6]^+$; ^{1}H NMR (DMSO-d_6, 600 MHz) δ = 6.18 (d, $J_{6,8}$ = 1.8 Hz, 1H; H-6), 6.56 (d, $J_{6,8}$ = 1.8 Hz, 1H; H-8), 7.21 (br s, 2H; H-2', H-6'), 5.26 (br s, 1H; H-1"), 3.42–3.67 (m, 3H; H-2", H-3", H-4"), 4.08 (br s, 1H; H-5"), 12.5 ppm (5-OH); ^{13}C NMR (DMSO-d_6, 150 MHz) δ = 150.3 (C-2), 136.8 (C-3), 176.1 (C-4), 160.7 (C-5), 98.2 (C-6), 164.1 (C-7), 93.3 (C-8), 156.2 (C-9), 103.1 (C-10), 120.8 (C-1', N.D.), 107.3 (C-2', C-6'), 145.6 (C-3', C-5'), 135.8 (C-4'), 96.9 (C-1"), 71.9 (C-2"), 73.3 (C-3"), 71.5 (C-4"), 75.5 (C-5"), 175.7 ppm (C-6"). [6] N.D. indicates "not detected".

Formononetin-7-*O*-glucuronide (18P): Its mass spectrum showed an ion peak at *m/z* 445 in positive mode and fragment ion at *m/z* 269 in the MS/MS detection, suggesting that it possesses a formononetin aglycone. Eventually, due to only one hydroxyl group at the C-7 position, the product was identified as formononetin-7-*O*-β-glucuronide. [7]

Narigenin-7-*O*-glucuronide (21P): ESI-MS: *m/z* = 449 $[M+H]^+$, ESI-MS/MS: *m/z* = 273 $[M+H-C_6H_8O_6]^+$;

^{1}H NMR (DMSO-d_6, 600 MHz) δ = 5.51 (br d, $J_{2,3a}$ = 14.1 Hz, 1H; H-2), 3.33 (br dd, $J_{2,3a;\ 3a,3b}$ = 14.1, 16.6 Hz, 1H; H-3a), 2.73 (br d, $J_{3a,3b}$ = 16.6 Hz, 1H; H-3b), 6.14 (br s, 1H; H-6), 6.19 (br s, 1H; H-8), 7.33 (d, $J_{2',3';\ 5',6'}$ = 9.0 Hz, 2H; H-2', H-6'), 6.80 (d, $J_{2',3';\ 5',6'}$ = 9.0 Hz, 2H; H-3', H-5'), 5.14 (m, 1H; H-1"), 3.17–3.45 (m, 3H; H-2", H-3", H-4"), 3.92 (br d, $J_{4",5"}$ = 9.0 Hz, 1H; H-5"), 12.0 ppm (5-OH); ^{13}C NMR (DMSO-d_6, 150 MHz) δ = 78.7 (C-2), 42.1 (C-3), 197.3 (C-4), 162.9 (C-5), 96.4 (C-6), 164.8 (C-7), 95.3 (C-8), 162.8 (C-9), 103.3 (C-10), 128.6 (C-1'), 128.4 (C-2', C-6'), 115.2 (C-3', C-5'), 157.8 (C-4'), 98.8 (C-1"), 72.7 (C-2"), 75.1 (C-3"), 71.3 (C-4"), 75.7 (C-5"), 170.3 ppm (C-6"). [8]

4.2 Protein purification

The target enzyme SbUGT could be released from the beads by cleaving off the chitin binding domain. The fusion and target enzyme SbUGT were identified as strong bands with expected sizes (75 and 51 kDa) when analyzed by SDS-PAGE.

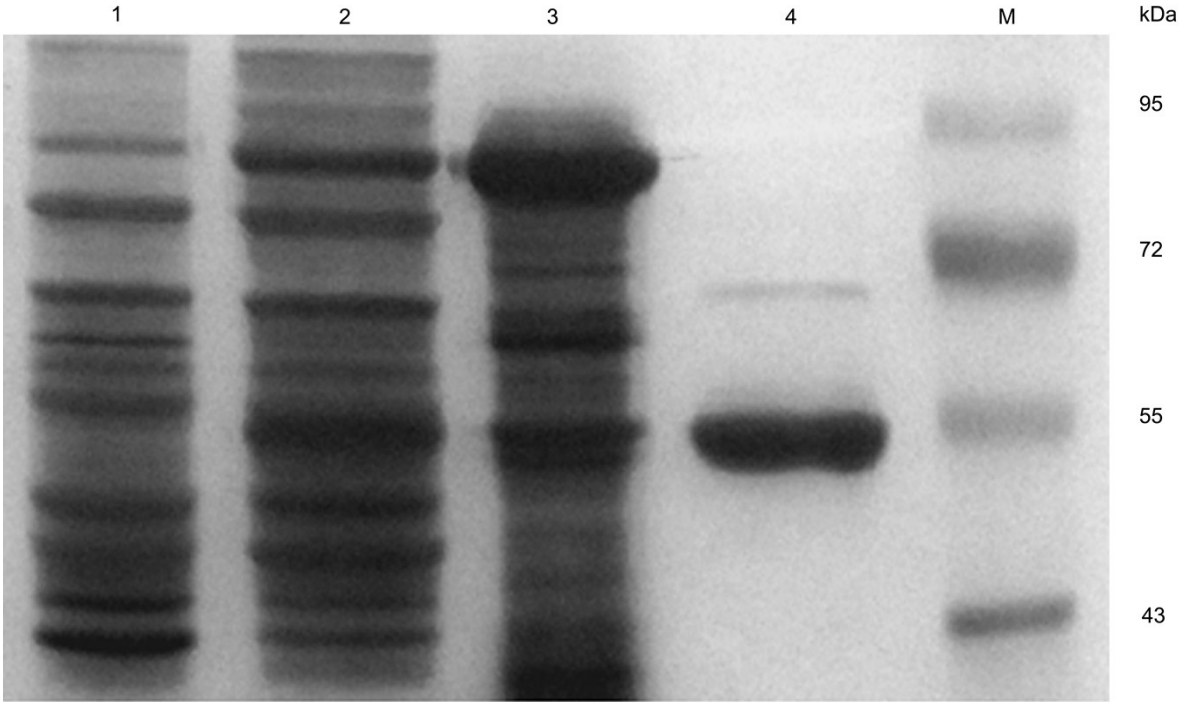


Fig. S8 SDS-PAGE analysis of expression and purification of recombinant protein SbUGT-W76. Lane 1, *E. coli* DE3; Lane 2, The soluble supernatant of SbUGT-W76; Lane 3, The SbUGT-W76 fusion protein; and Lane 4, The target protein SbUGT-W76.

4.3 Effects of divalent metal ions, pH and temperature

The effects of divalent metal ions, pH and temperature for glucuronosyl activity of purified enzyme SbUGT-W76 was investigated with UDP-glucuronic acid as sugar donor and baicalein as sugar acceptor. $CaCl_2$ and $MgCl_2$ enhanced enzyme activity, whereas $ZnCl_2$, $CoCl_2$, and $MnCl_2$ inhibited enzyme activity in some degree, and $CuCl_2$ completely inhibited SbUGT-W76 (Fig. S9A). The inhibition of flavonoid glycosyltransferases by addition of Cu^{2+} ions, which is always considered as Cu^{2+}–mediated substrate degration rather than inhibition of the enzyme. [9,10] The enzyme functioned better in the pH range of 6.5−7.5 and showed only low activity at pH＜6.0 or ＞8.0. The optimal temperature of purified SbUGT-W76 was estimated to be 30°C.

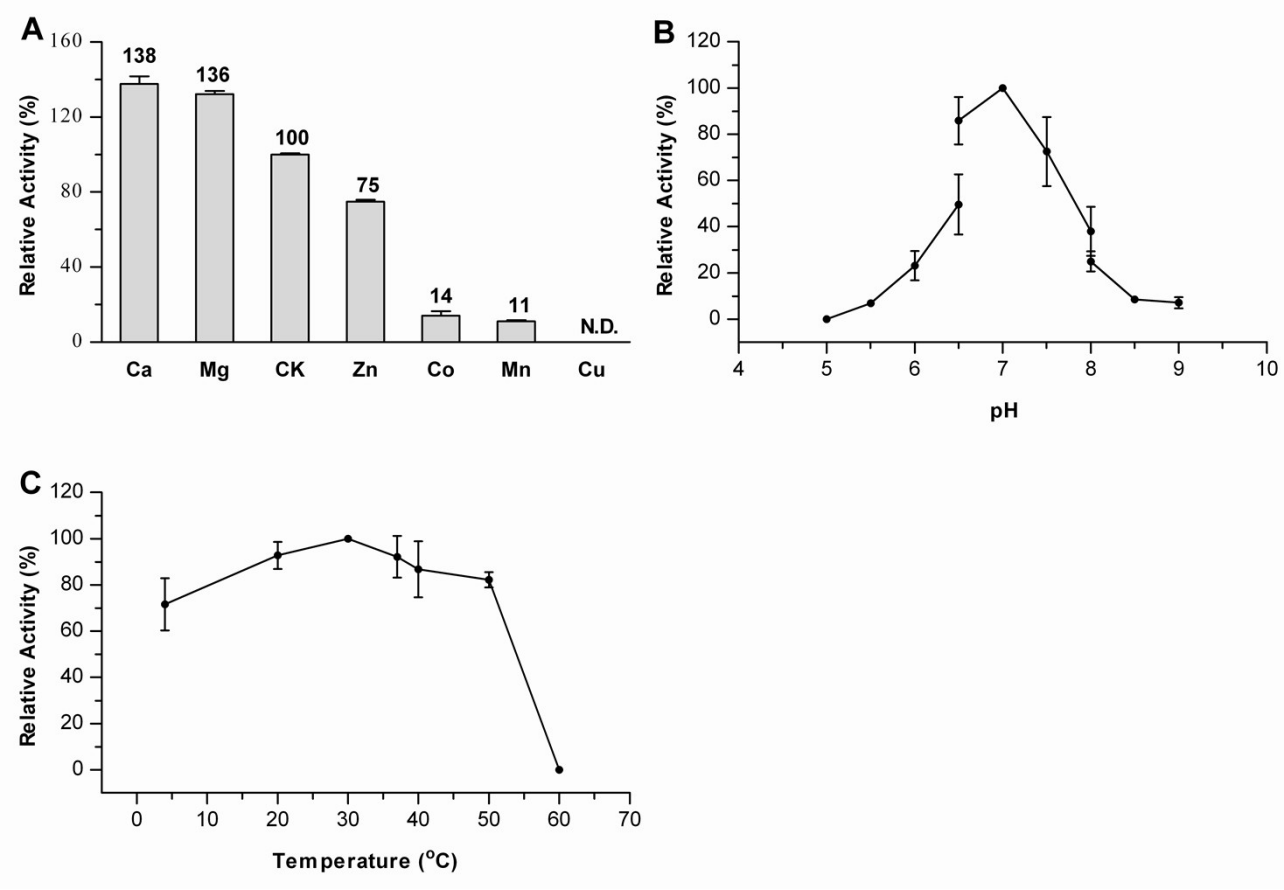


Fig. S9 Effects of various divalent metal ions (A), pH (B), and temperature (C) on enzyme activity of SbUGT-W76. Error bars represent the standard deviation from three replicates.

4.4 UDP-glucuronic acid yield in the engineered *Escherichia coli* strains

To evaluate the effect of the modified UDPGA synthetic pathway, the yield of UDPGA in the engineered *E. coli* strains were assayed. As shown in Fig. S10, UDPGA was not detected in strain BT, while strain BPGU without overexpression of SbUGT-W76 produced the highest yield of 2.69 mg/L. Most remarkably, production of UDPGA by strains BUT, BPUT, BGUT, and BPGUT increased about 4−fold compared to strains without overexpression of the *ugd* gene. These results indicated that Ugd might be the key enzyme required to boost the production of UDPGA.

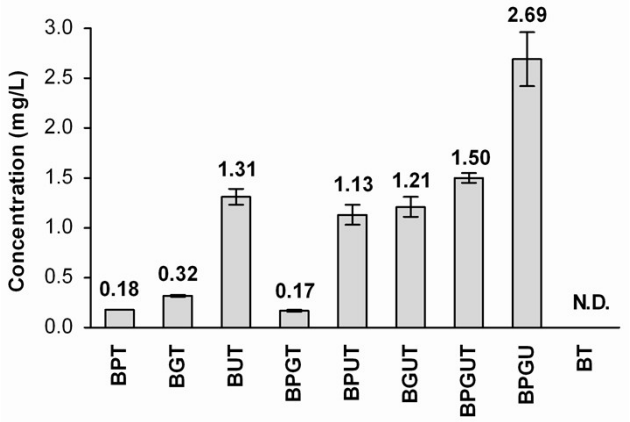


Fig. S10 Comparison of UDPGA yields in different engineered *E. coli* strains. Error bars represent the standard deviation from three replicates.

4.5 I*n vivo* functional characterization of recombinant SbUGT

The engineered *E. coli* strain BPGUT harboring recombinant SbUGT-W112 and SbUGT-W112G76 glucuronosylated 7-OH group of flavones with or without hydroxyl or methoxyl substituent at the *ortho*-position, i.e. baicalein and apigenin. Besides these, isoflavone (4'-demethycalycosin) together with one flavanone (naringenin) were also catalyzed by the engineered *E. coli* strain (Fig. S11). Both enzymatic assay of purified SbUGT-W112 and SbUGT-W112G76 and *in vivo* biocatalysis of engineered *E. coli* presented the same sugar-acceptor preference.

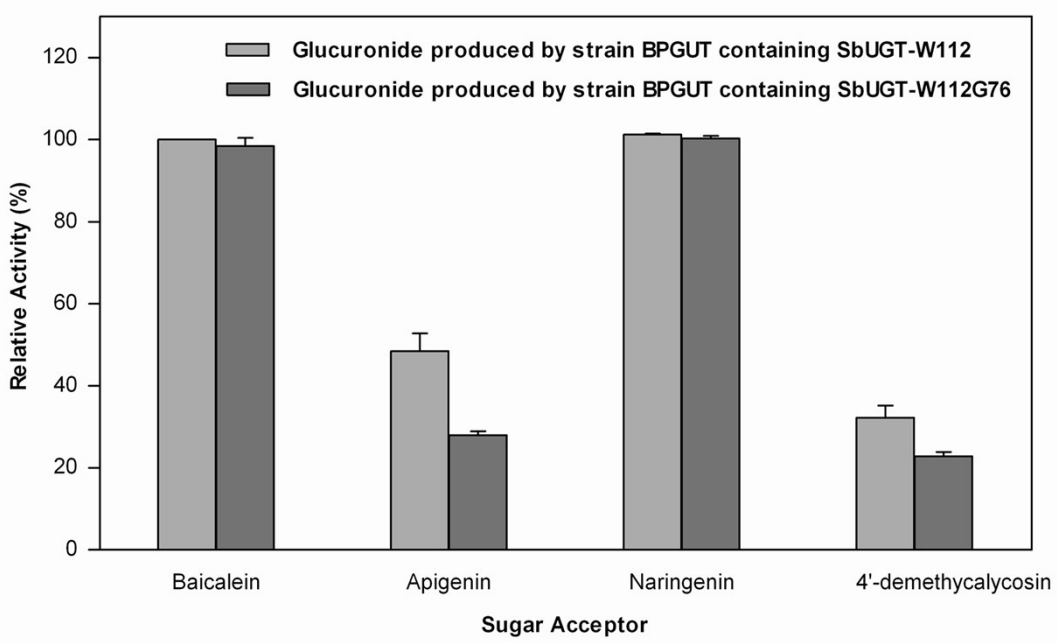


Fig. S11 Four flavonoids biotransformed by strain BPGUT containing SbUGT-W112 or SbUGT-W112G76 detected at 280 nm. The specific activity of strain BPGUT harboring SbUGT-W112 on baicalein is set as 100%. Error bars represent the standard deviation from three replicates.

4.6 The hydrolysis of flavonoid glucuronides by β-glucuronidase

Flavonoid glucuronides products including baicalin (**1**a), oroxylin A-7-*O*-glucuronide (**3**P), scutellarein-7-*O*-glucuronide (**4**P), kaempferol-7-*O*-glucuronide (**11**P), 4'-demethycalycosin-7-*O*-glucuronide (**17**P) and narigenin-7-*O*-glucuronide (**21**P) were further confirmed by its hydrolysis to the aglycones when β-glucuronidase was used (Figure S12). Conversion of these flavonoid glucuronide into aglycone increased over time. Based on these results, the flavonoid glucuronide was elucidated as β-configuration.

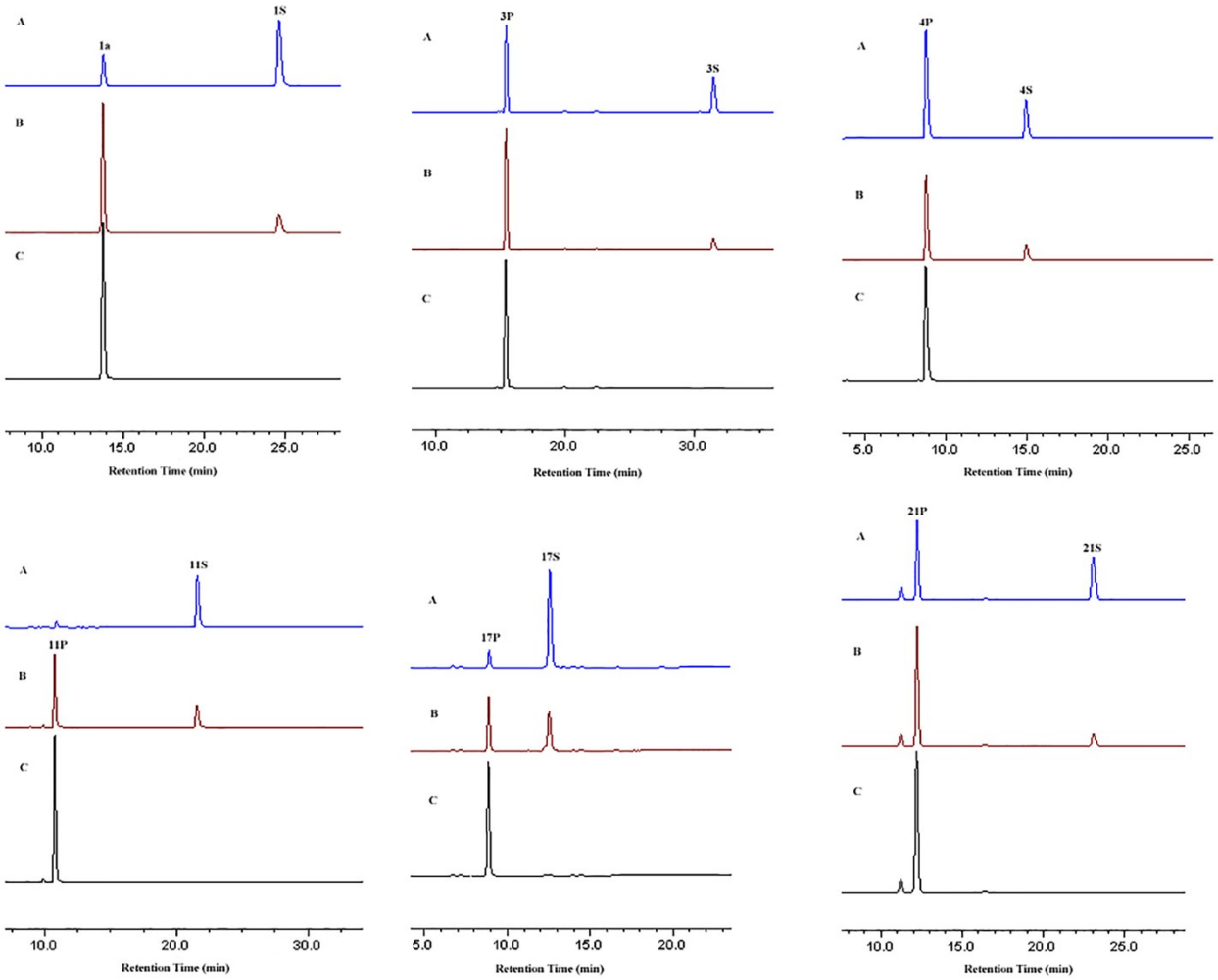


Fig. S12 HPLC analysis of hydrolytic products of flavonoid glucuronides using β-glucuronidase. The sample mixtures incubated for 15 min (A) or 30 min (B), and the negative reaction without β-glucuronidase (C).

5. Additional Supplementary Figures

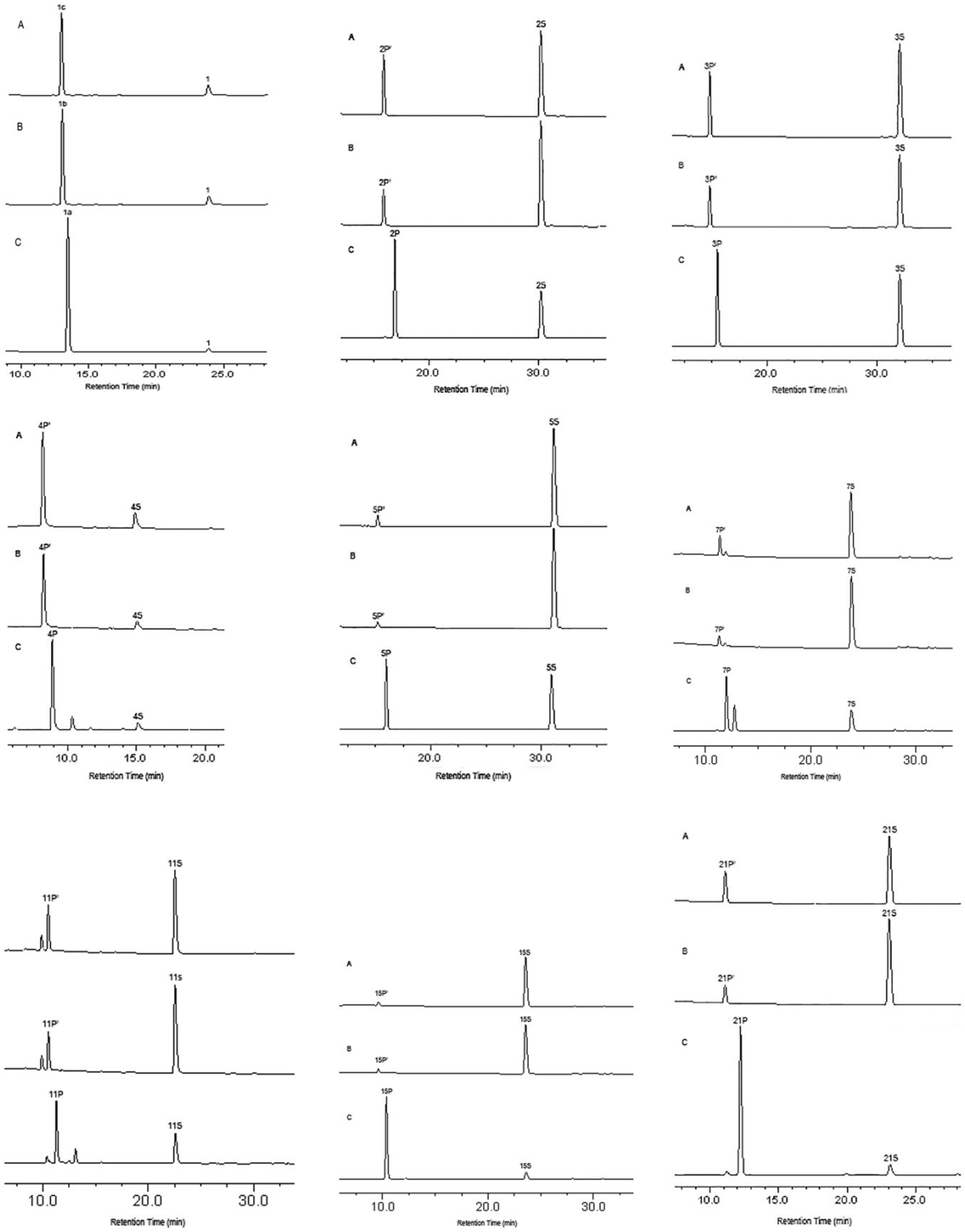
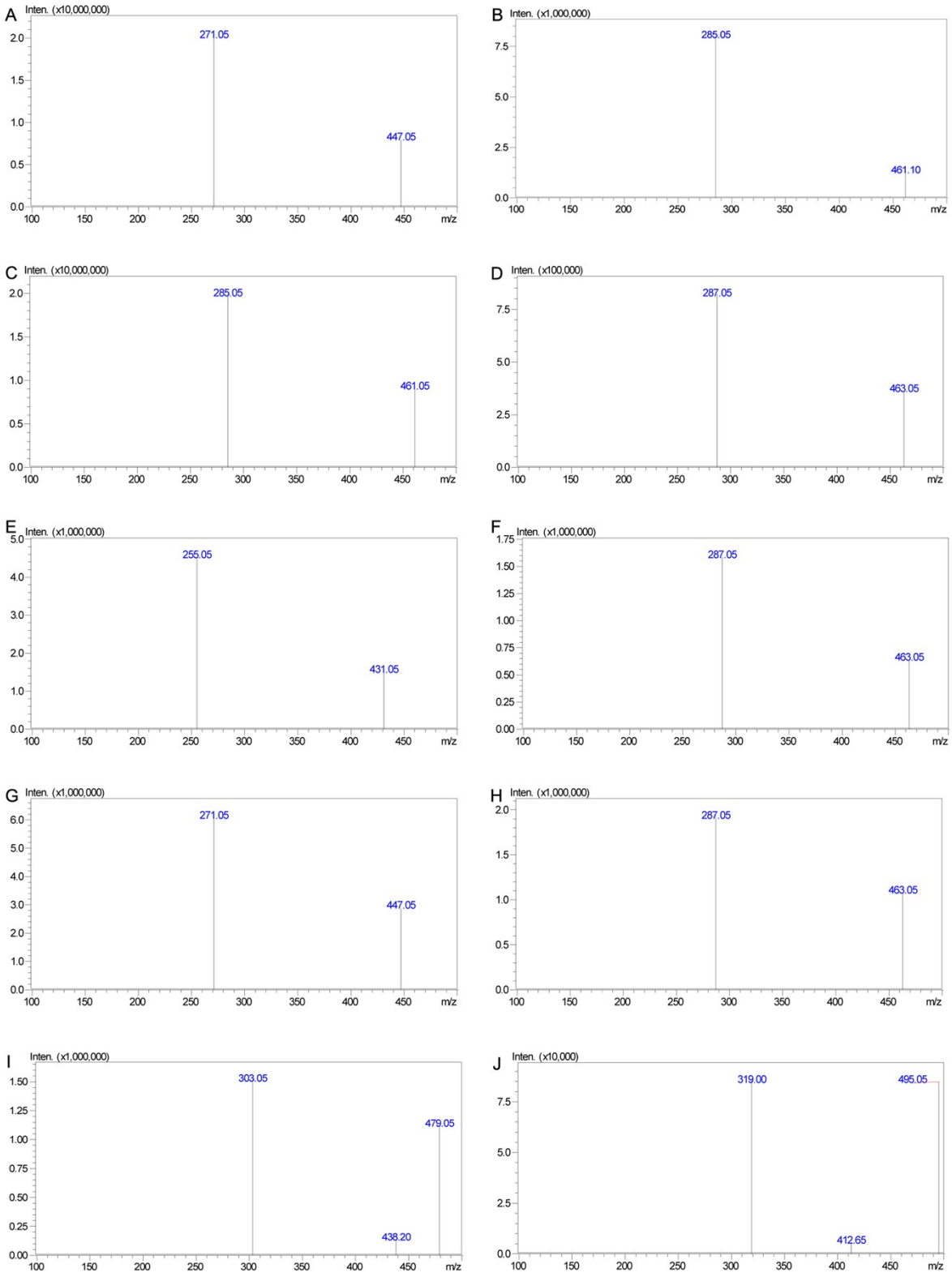


Fig. S13 The functional characterization of the purified SbUGT-W76 (A, C) and SbUGT-W112 (B) enzymes with UDPG (A, B) and UDPGA (C) as sugar donors and Baicalein (**1**), Wogonin (**2**), Oroxylin A (**3**), Scutellarein (**4**), Chrysin (**5**), Apigenin (**7**), Kaempferol (**11**), Genistein (**15**) and Naringenin (**21**) as sugar acceptors.


A
Inten. (x10,000,000)
271.05
447.05
B
Inten. (x1,000,000)
285.05
461.10
C
Inten. (x10,000,000)
285.05
461.05
D
Inten. (x100,000)
287.05
463.05
E
Inten. (x1,000,000)
255.05
431.05
F
Inten. (x1,000,000)
287.05
463.05
G
Inten. (x1,000,000)
271.05
447.05
H
Inten. (x1,000,000)
287.05
463.05
I
Inten. (x1,000,000)
303.05
479.05
438.20
J
Inten. (x10,000)
319.00
495.05
412.65
m/z

Continued.

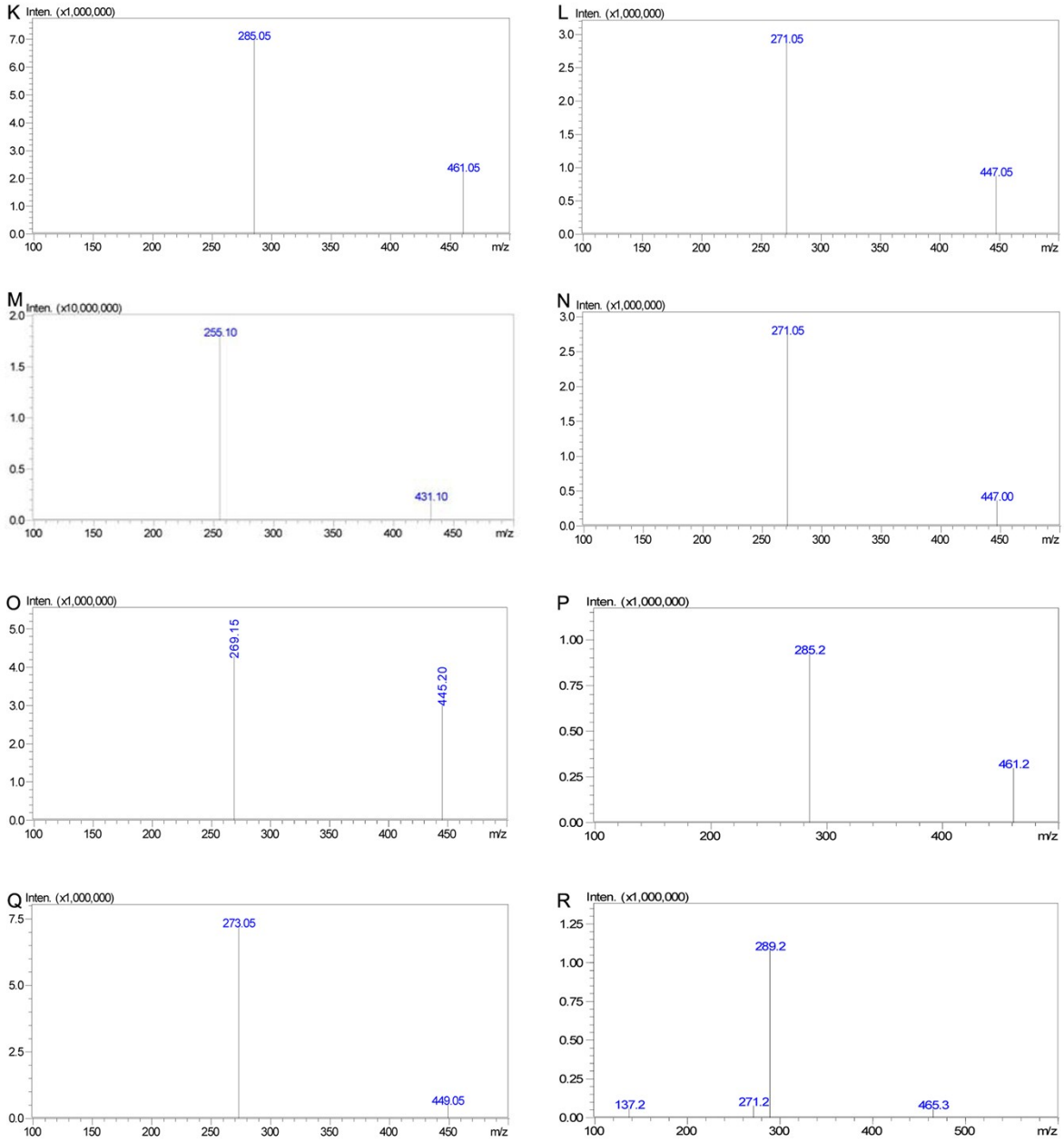
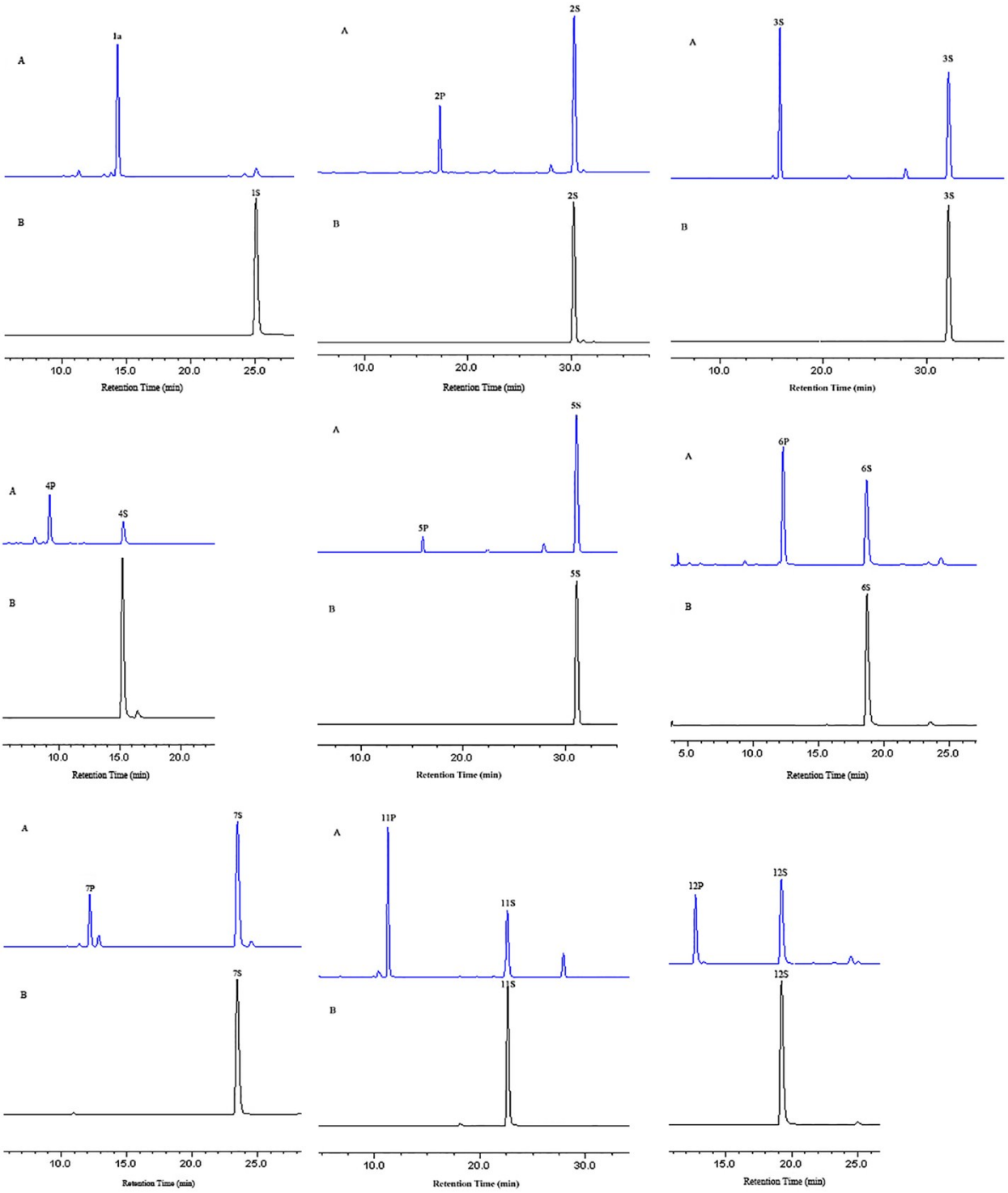


Fig. S14 The MS/MS spectra for glucuronides of Baicalein (A), Wogonin (B), Oroxylin A (C), Scutellarein (D), Chrysin (E), Luteolin (F), Apigenin (G), Kaempferol (H), Quercetin (I), Myricetin (J), Biochanin A (K), Genistein (L), Daidzein (M), 4'-demethycalycosin (N), Formononetin (O), Calycosin (P), Narigenin (Q), and Robtin (R) were detected in positive ionisation mode.


A
1a
B
1S
10.0
15.0
20.0
25.0
Retention Time (min)
A
2P
2S
B
2S
10.0
20.0
30.0
Retention Time (min)
A
3S
3S
B
3S
10.0
20.0
30.0
Retention Time (min)
A
4P
4S
B
10.0
15.0
20.0
Retention Time (min)
A
5P
5S
B
5S
10.0
20.0
30.0
Retention Time (min)
A
6P
6S
B
6S
5.0
10.0
15.0
20.0
25.0
Retention Time (min)
A
7P
7S
B
7S
10.0
15.0
20.0
25.0
Retention Time (min)
A
11P
11S
B
11S
10.0
20.0
30.0
Retention Time (min)
12P
12S
12S
15.0
20.0
25.0
Retention Time (min)

Continued.

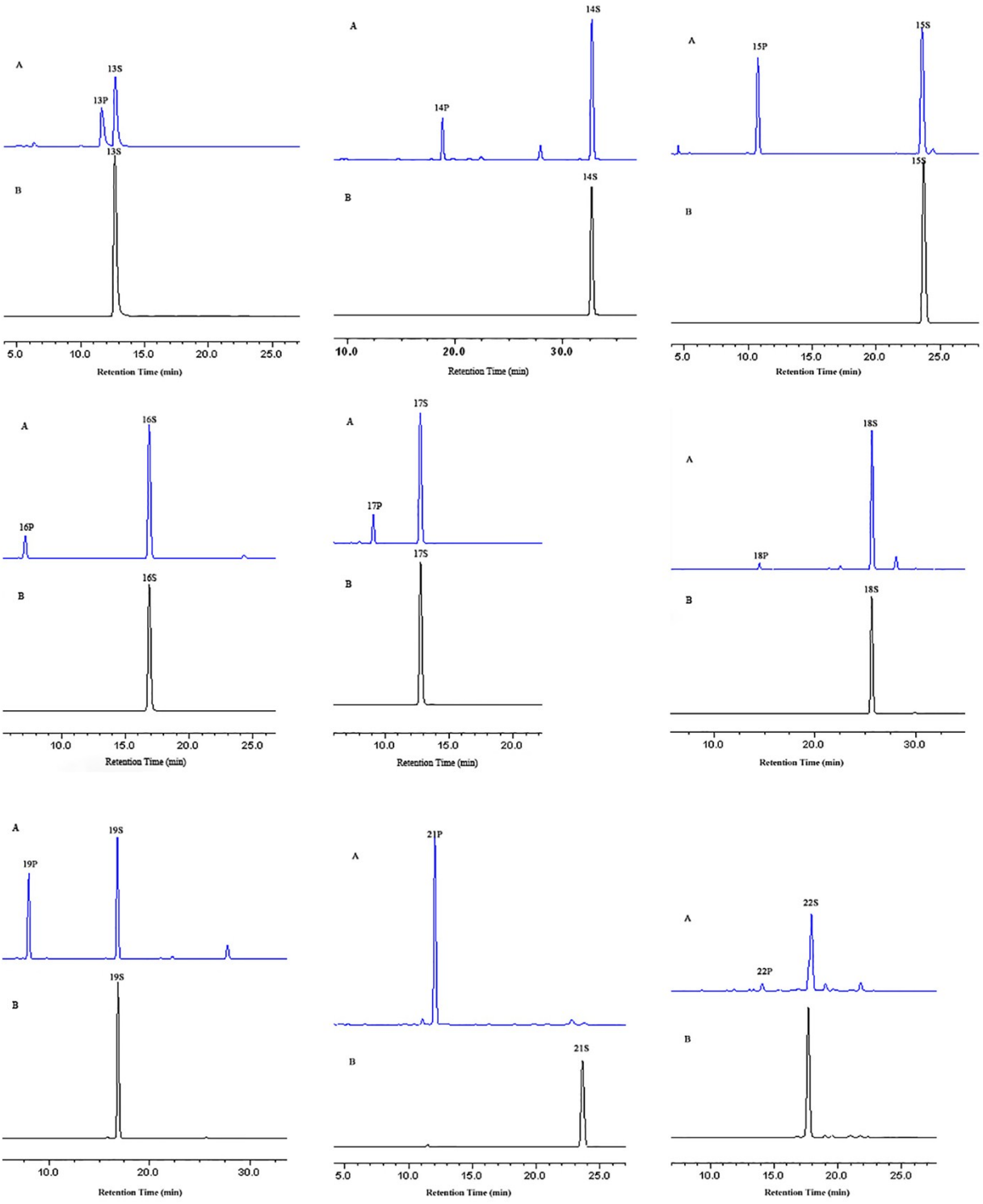


Fig. S15 HPLC analysis of the whole-cell biocatalytic products using engineered *E. coli* strain BPGUT with compounds **1** to **7, 11** to **19,** and **21** to **22** as substrates. (A) The whole-cell reaction using strain BPGUT; (B) Flavonoid substrates.

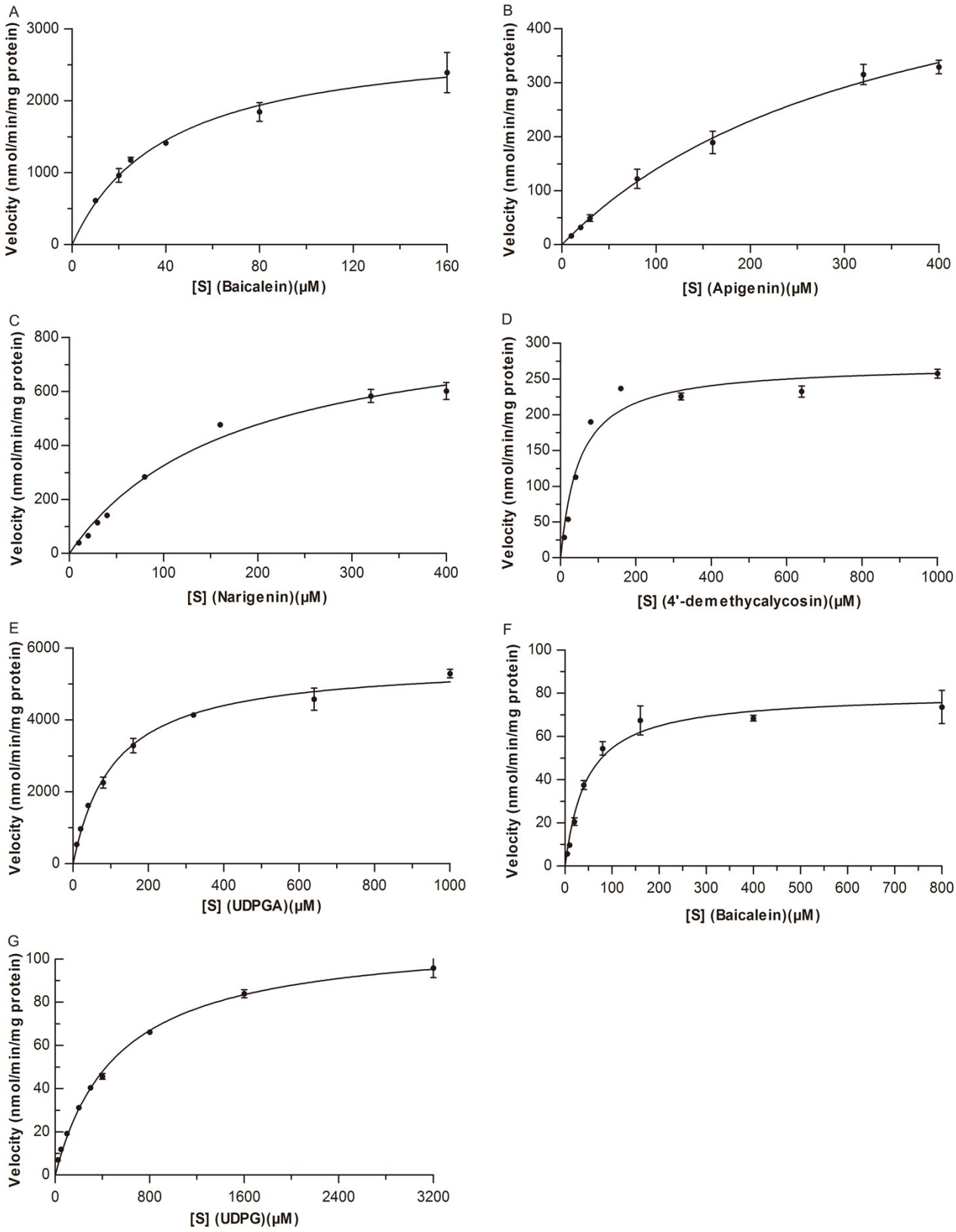


Fig. S16 Plots of initial rate versus concentration for sugar acceptors and donors with SbUGT-W76 enzyme. Baicalein (A), Apigenin (B), Narigenin (C), 4'-Demethycalycosin (D) using UDPGA as donor; UDPGA (E) and UDPG (G) using baicalein as acceptor; Baicalein (F) using UDPG as donor. The line represents the data fitted to the Michaelis–Menten equation. Error bars represent the standard deviation from three replicates.

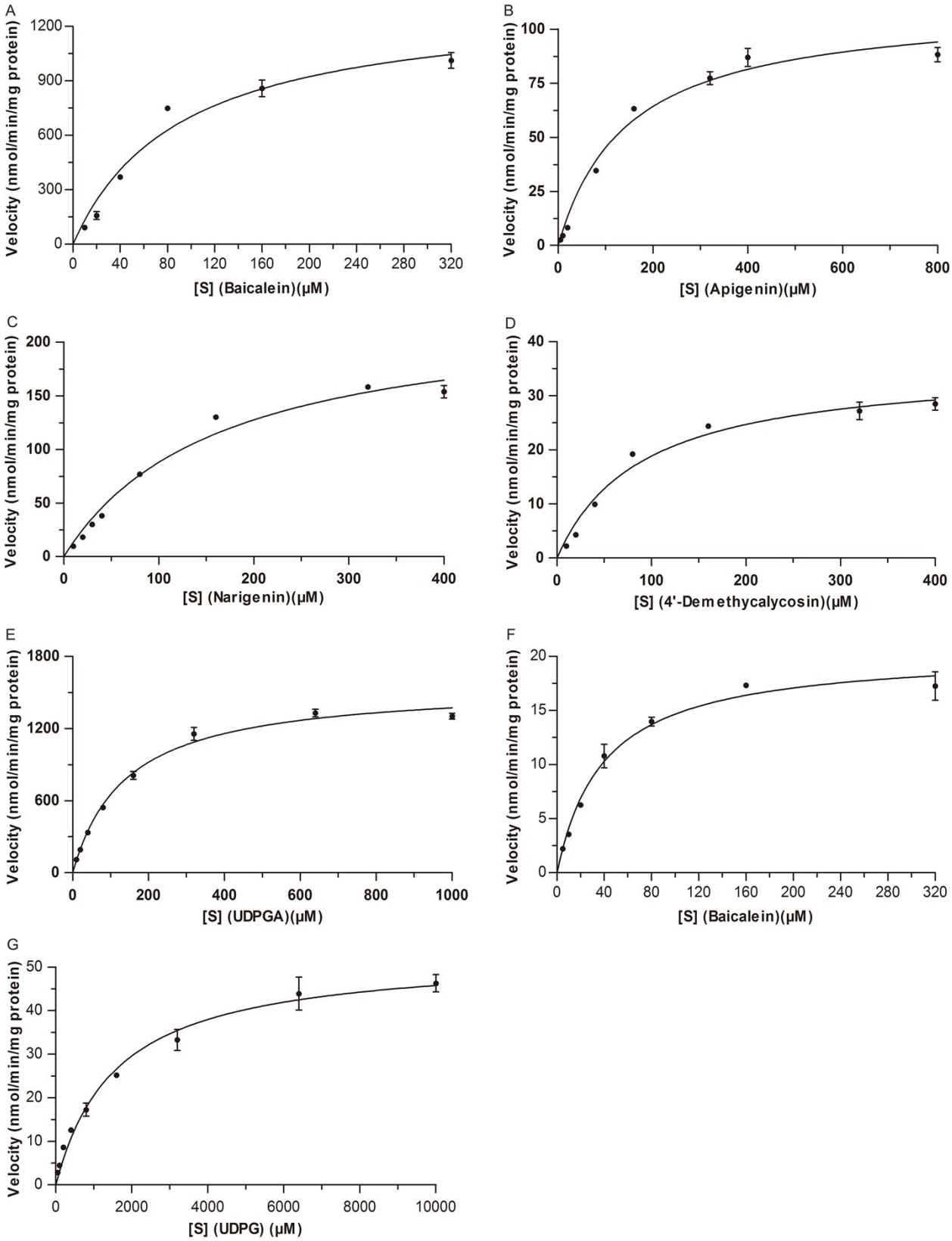


Fig. S17 Plots of initial rate versus concentration for sugar acceptors and donors with SbUGT-W112 enzyme. Baicalein (A), Apigenin (B), Narigenin (C), 4'-Demethycalycosin (D) using UDPGA as donor; UDPGA (E) and UDPG (G) using baicalein as acceptor; Baicalein (F) using UDPG as donor. The line represents the data fitted to the Michaelis–Menten equation. Error bars represent the standard deviation from three replicates.

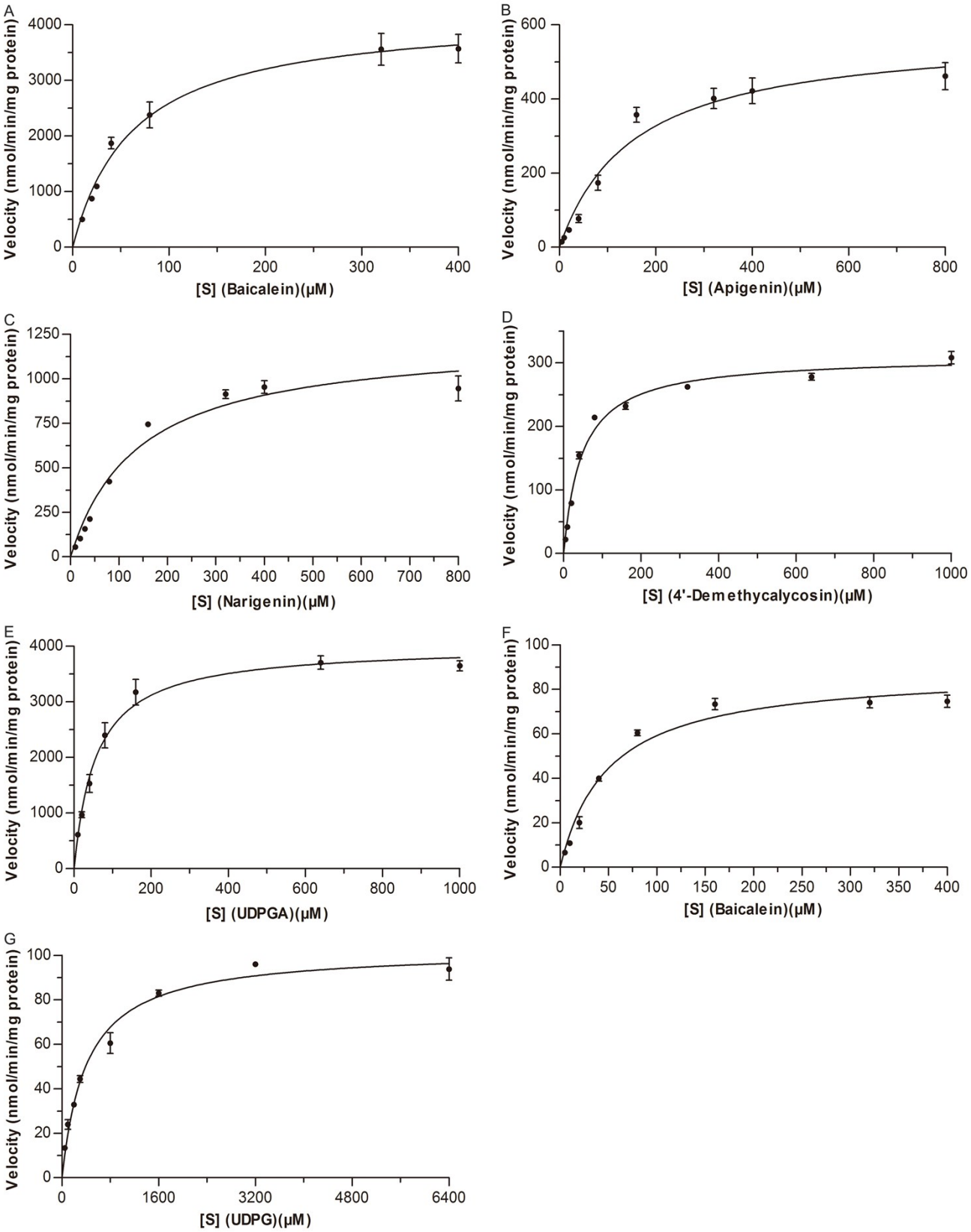


Fig. S18 Plots of initial rate versus concentration for sugar acceptors and donors with SbUGT-W112G76 enzyme. Baicalein (A), Apigenin (B), Narigenin (C), 4'-Demethycalycosin (D) using UDPGA as donor; UDPGA (E) and UDPG (G) using baicalein as acceptor; Baicalein (F) using UDPG as donor. The line represents the data fitted to the Michaelis–Menten equation. Error bars represent the standard deviation from three replicates.

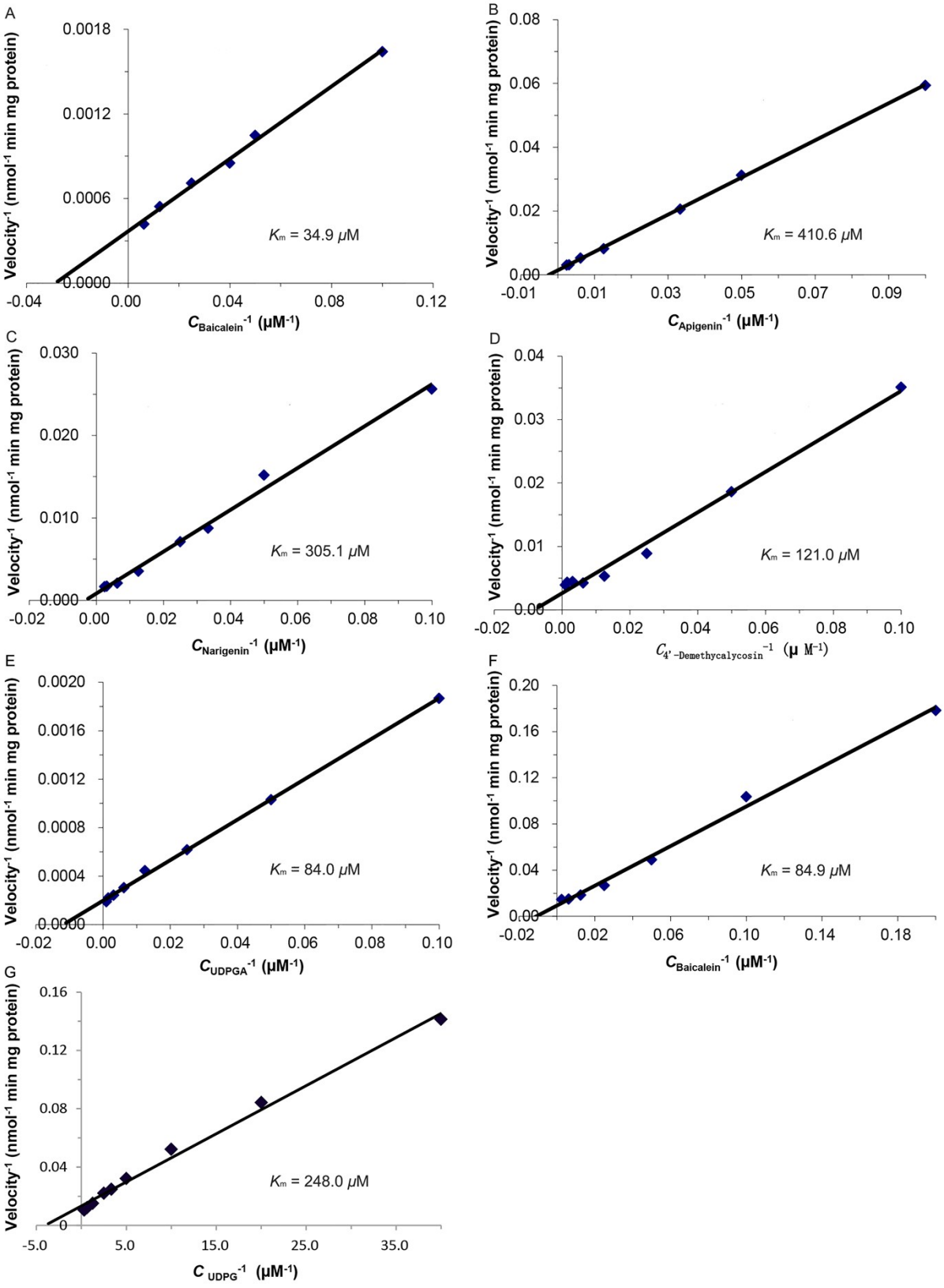


Fig. S19 The Lineweaver-Burk plots for sugar acceptors and donors with SbUGT-W76 enzyme. Baicalein (A), Apigenin (B), Narigenin (C), 4'-Demethycalycosin (D) using UDPGA as donor; UDPGA (E) and UDPG (G) using baicalein as acceptor; Baicalein (F) using UDPG as donor.

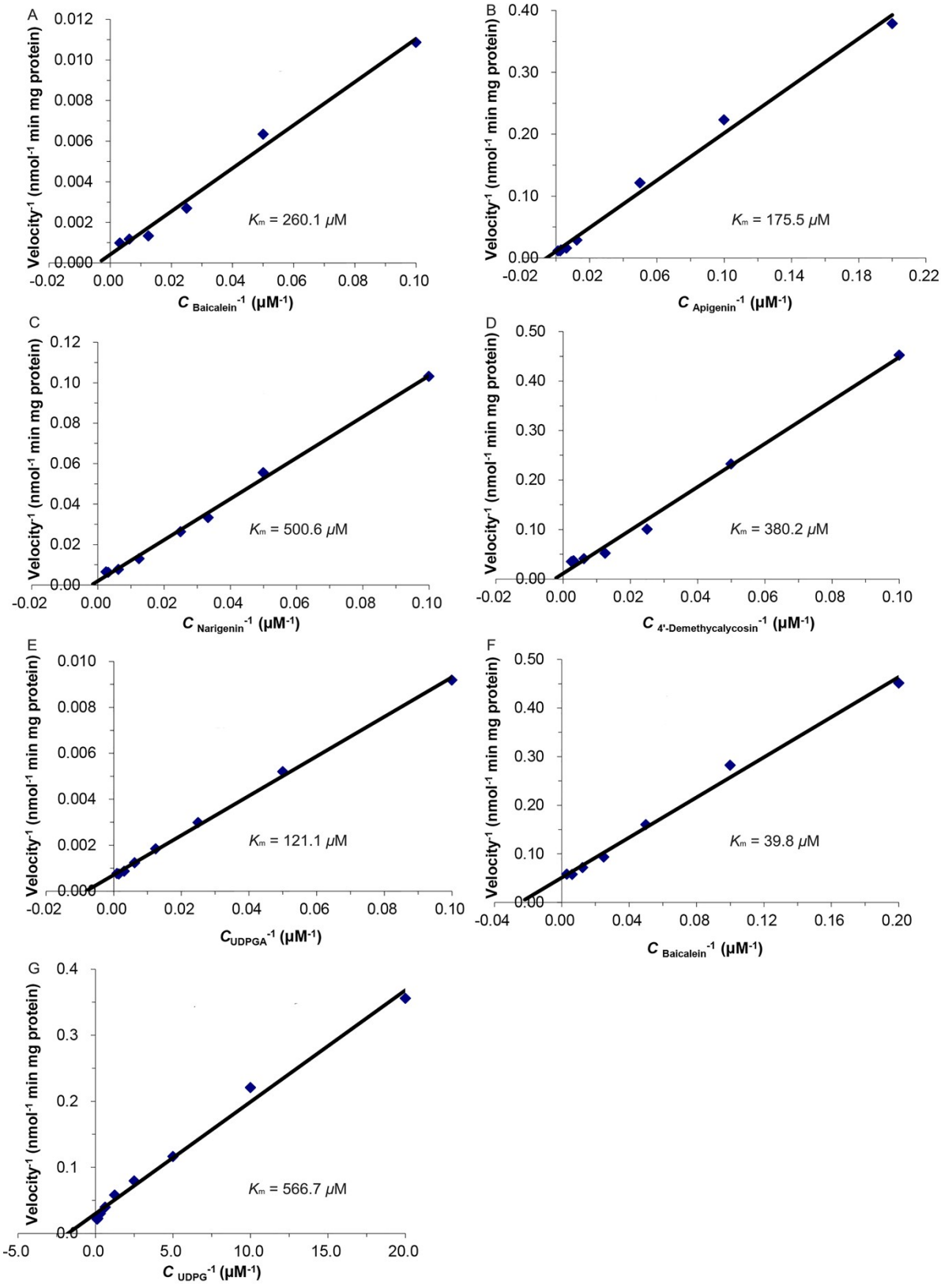


Fig. S20 The Lineweaver-Burk plots for sugar acceptors and donors with SbUGT-W112 enzyme. Baicalein (A), Apigenin (B), Narigenin (C), 4'-Demethycalycosin (D) using UDPGA as donor; UDPGA (E) and UDPG (G) using baicalein as acceptor; Baicalein (F) using UDPG as donor.

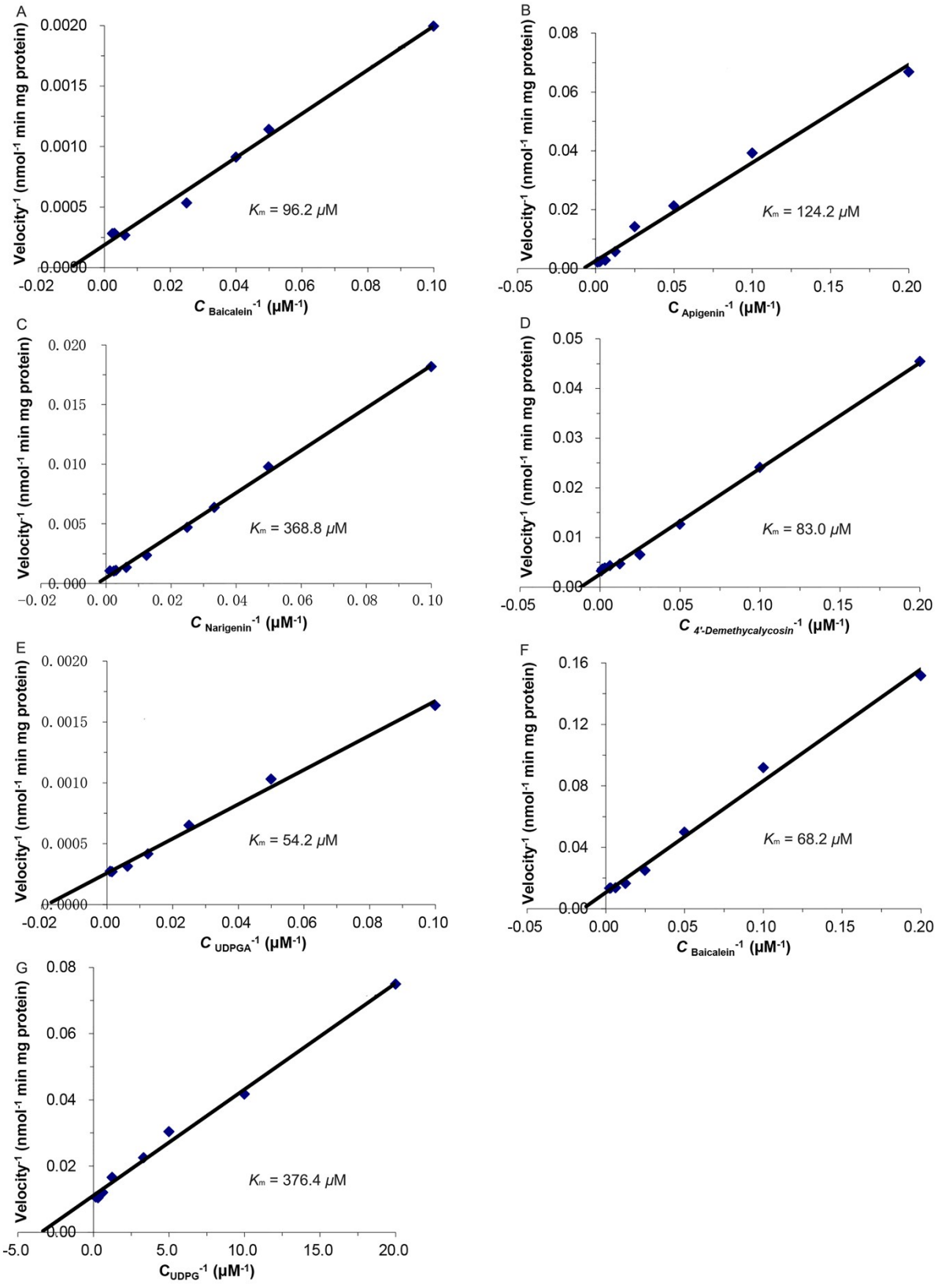


Fig. S21 The Lineweaver-Burk plots for sugar acceptors and donors with SbUGT-W112G76 enzyme. Baicalein (A), Apigenin (B), Narigenin (C), 4'-Demethycalycosin (D) using UDPGA as donor; UDPGA (E) and UDPG (G) using baicalein as acceptor; Baicalein (F) using UDPG as donor.

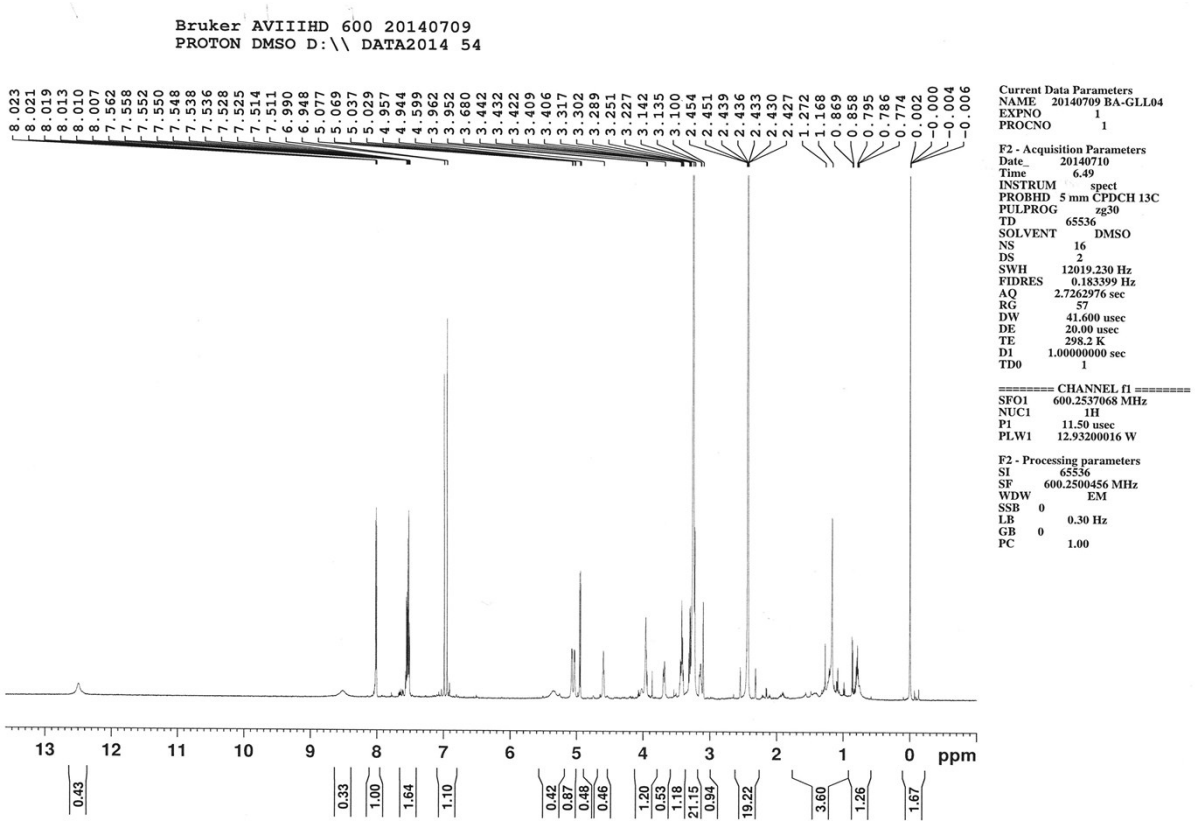


Fig. S22 The ^{1}H NMR spectrum of Baicalein-7-*O*-glucoside (**1b**) in DMSO-d_6.

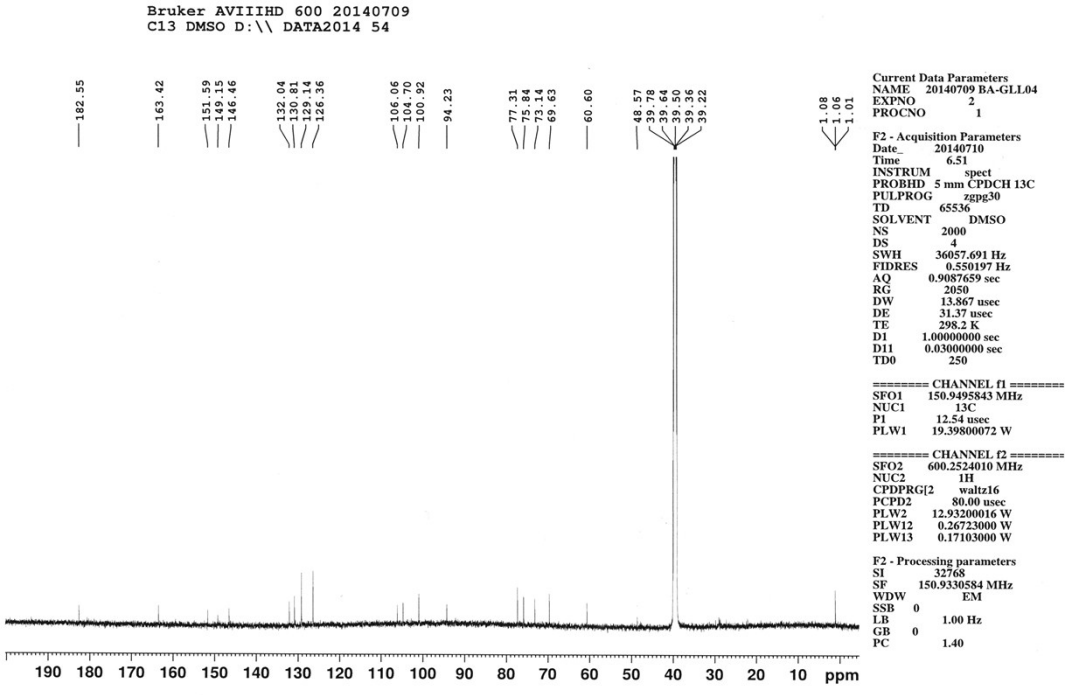


Fig. S23 The ^{13}C NMR spectrum of Baicalein-7-*O*-glucoside (**1b**) in DMSO-d_6.

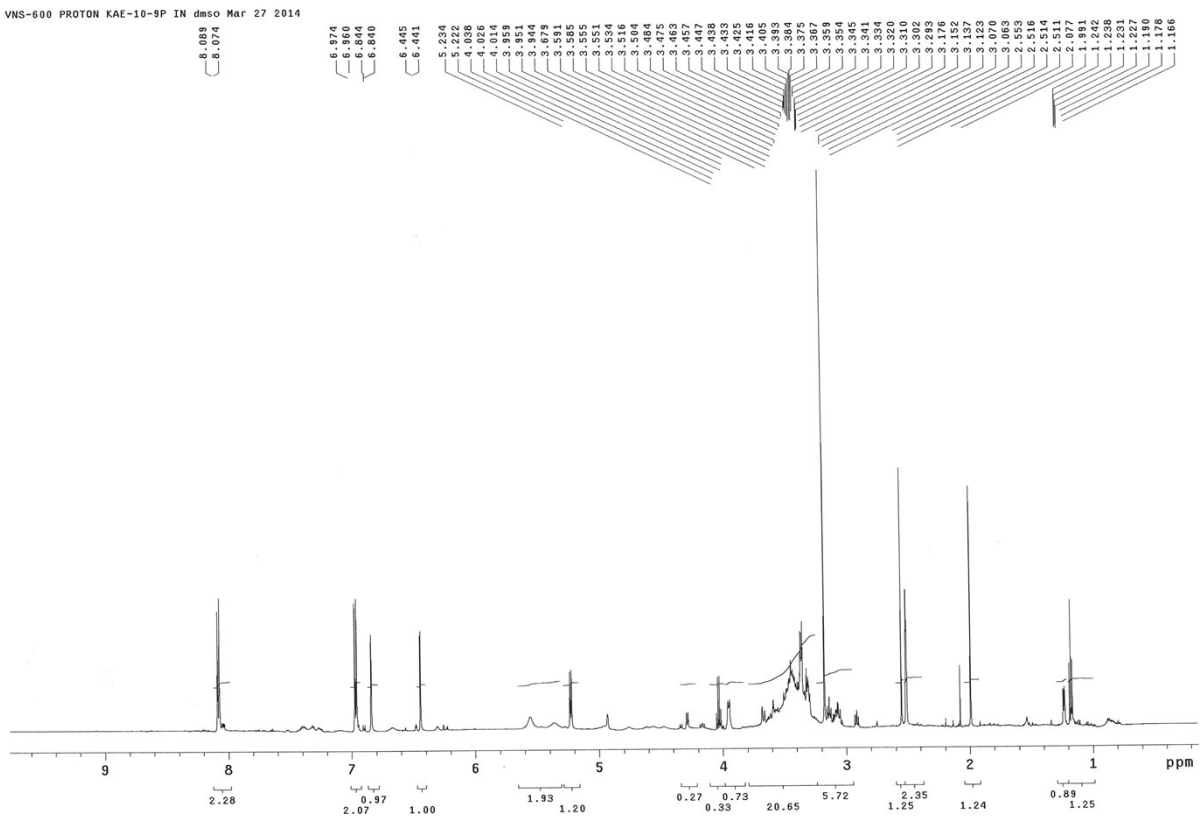


Fig. S24 The ^{1}H NMR spectrum of Kaempferol-7-*O*-glucuronide (**11P**) in DMSO-d_6.

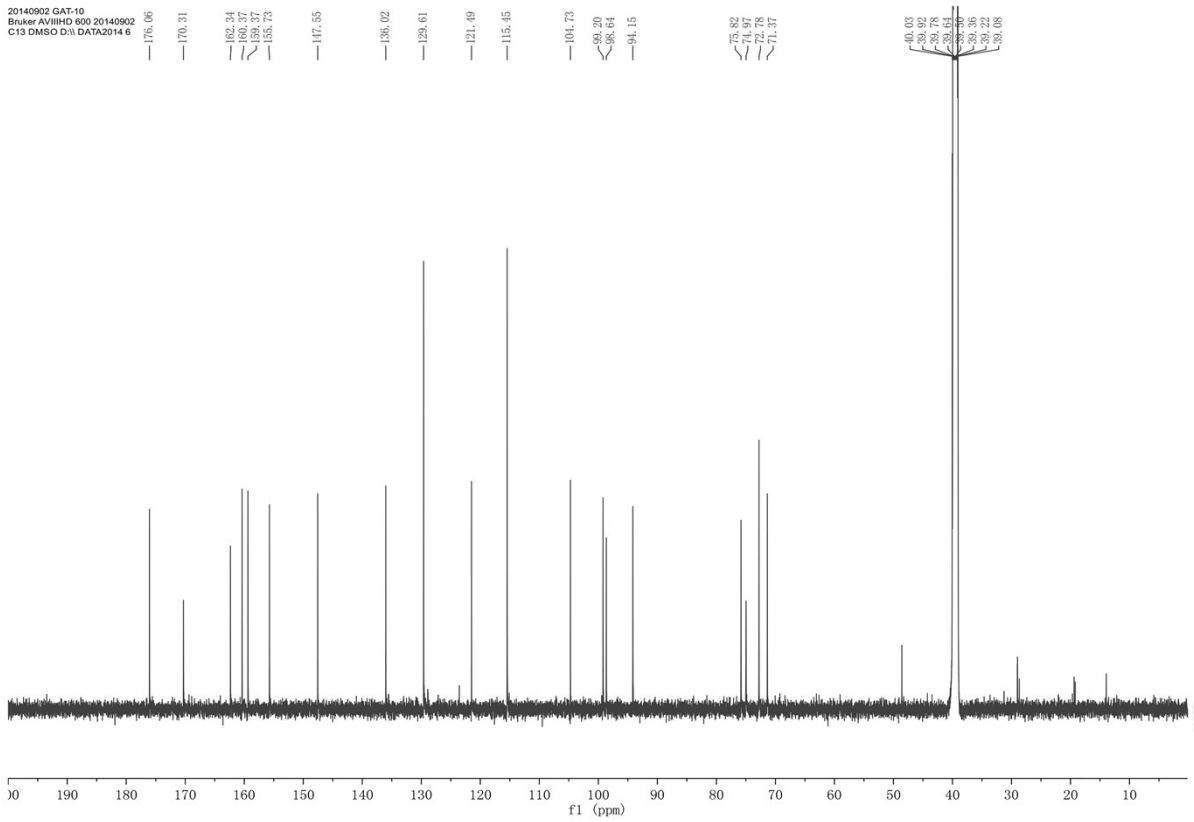


Fig. S25 The ^{13}C NMR spectrum of Kaempferol-7-*O*-glucuronide (**11P**) in DMSO-d_6.

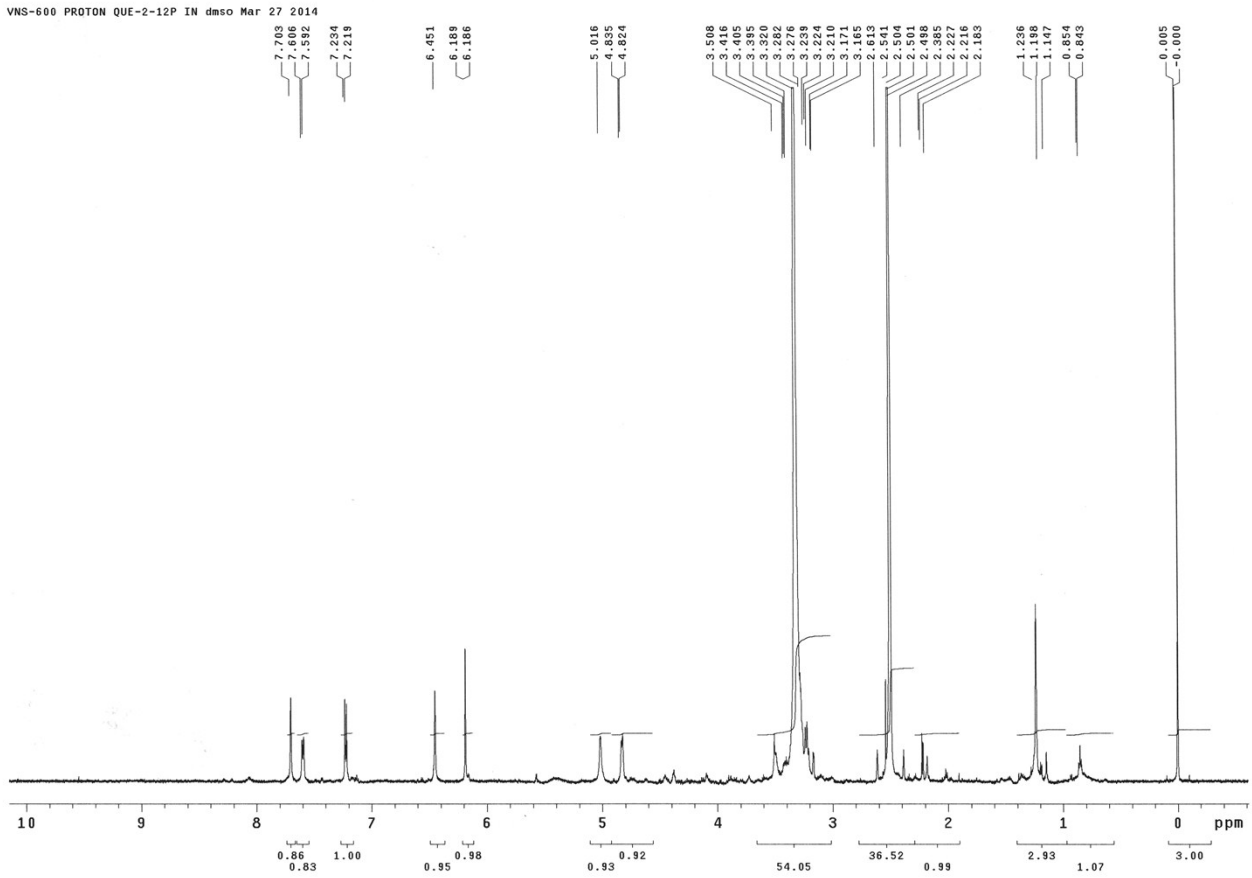


Fig. S26 The ^{1}H NMR spectrum of Quercetin-7-*O*-glucuronide (**12P**) in DMSO-d_6.

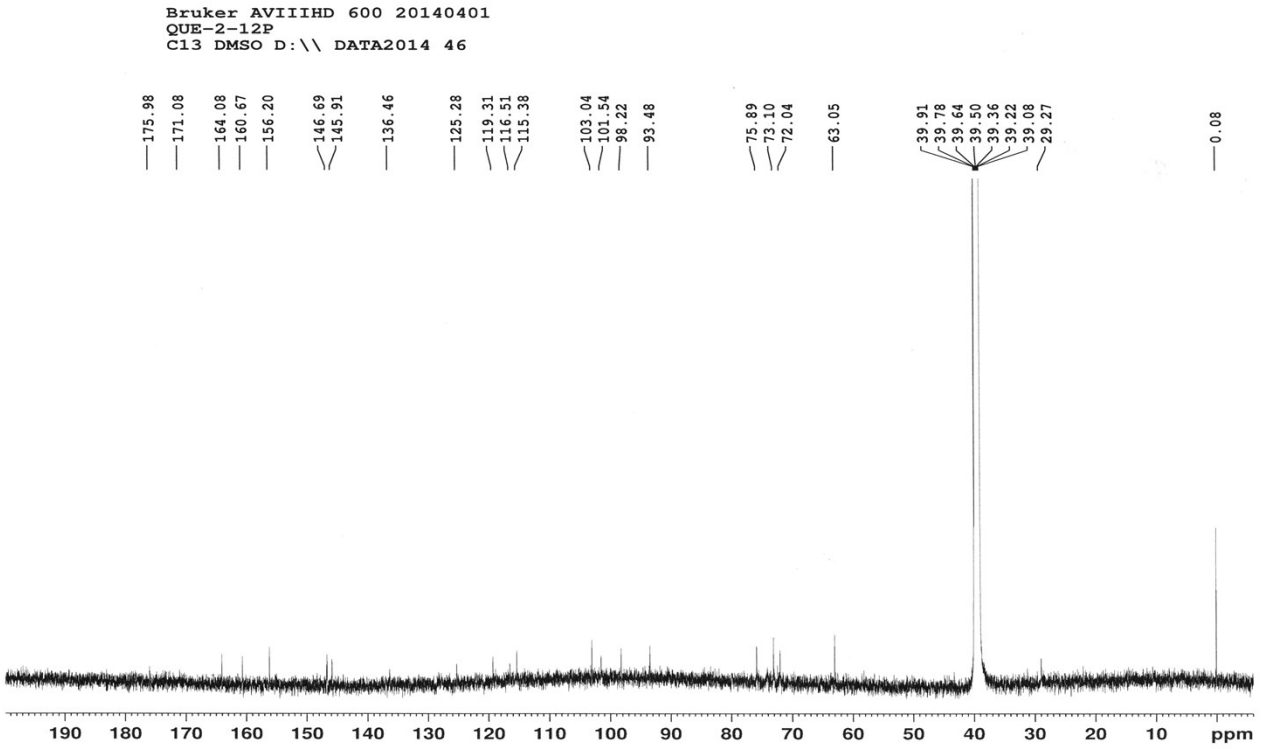


Fig. S27 The ^{13}C NMR spectrum of Quercetin-7-*O*-glucuronide (**12P**) in DMSO-d_6.

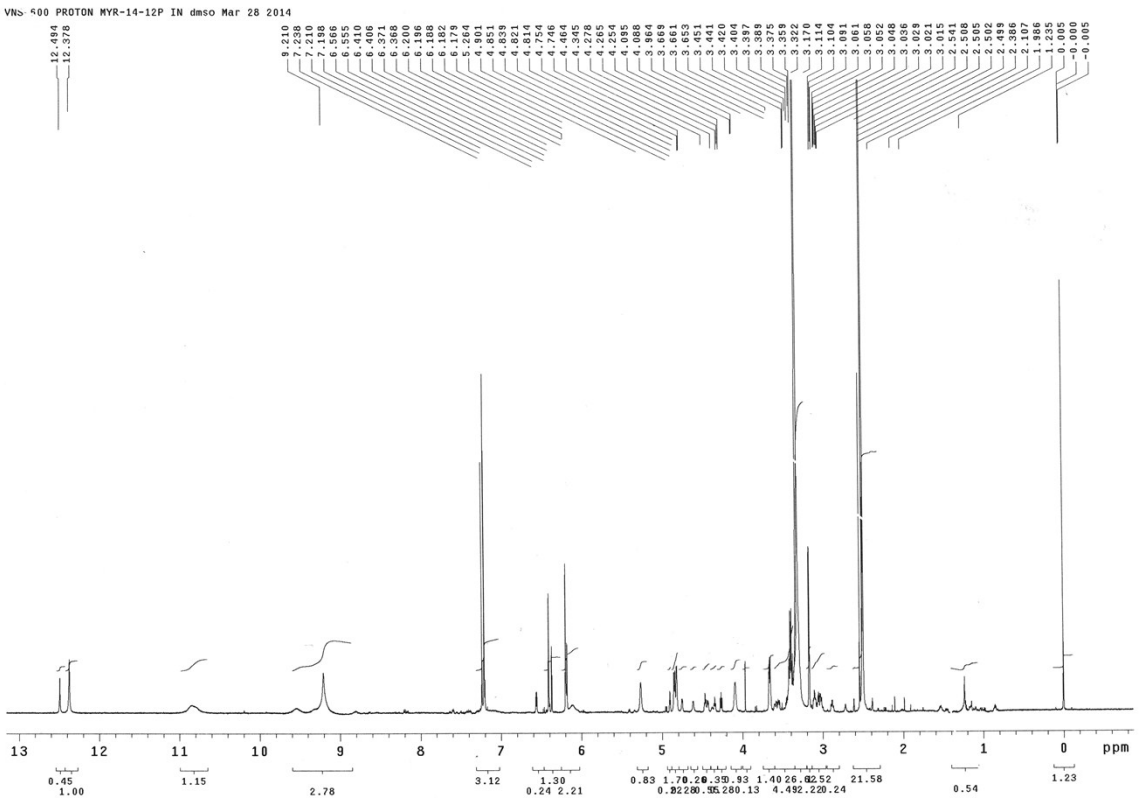


Fig. S28 The ^{1}H NMR spectrum of Myricetin-7-*O*-glucuronide (**13P**) in DMSO-d_6.

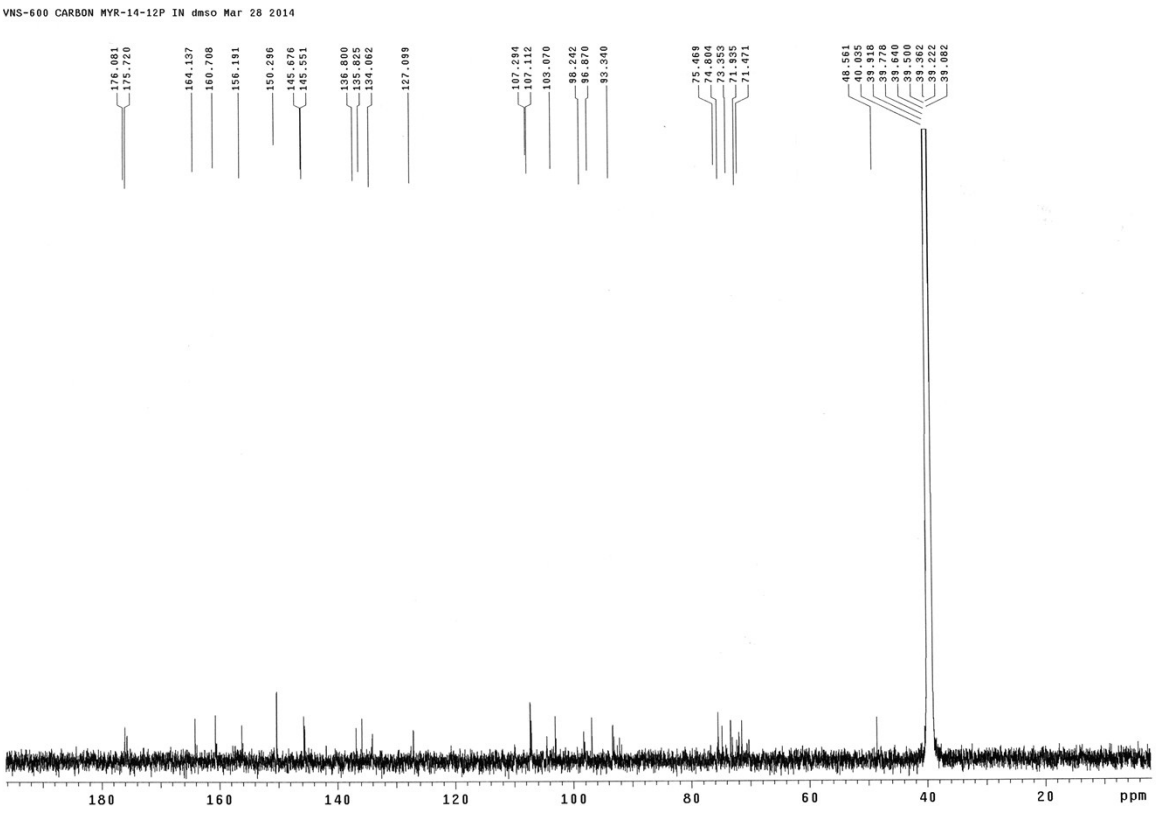


Fig. S29 The ^{13}C NMR spectrum of Myricetin-7-*O*-glucuronide (**13P**) in DMSO-d_6.

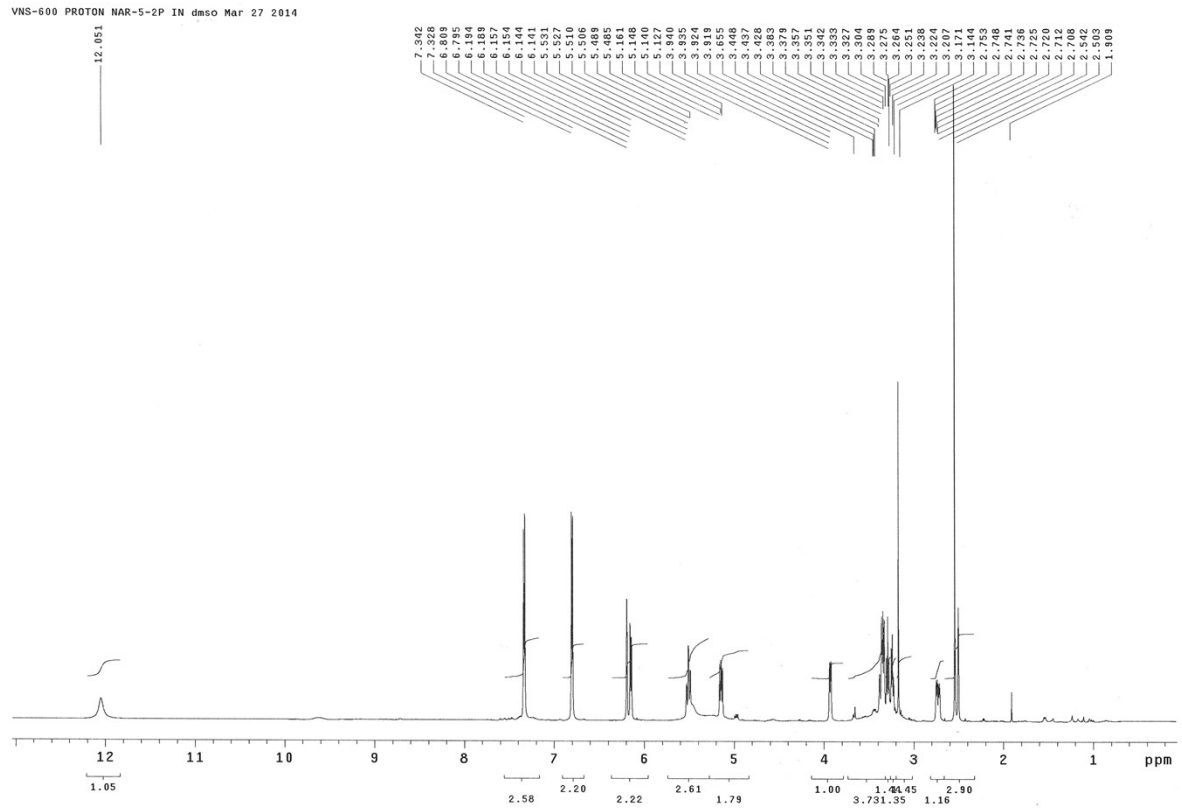


Fig. S30 The ^{1}H NMR spectrum of Narigenin-7-*O*-glucuronide (21**P**) in DMSO-d_6.

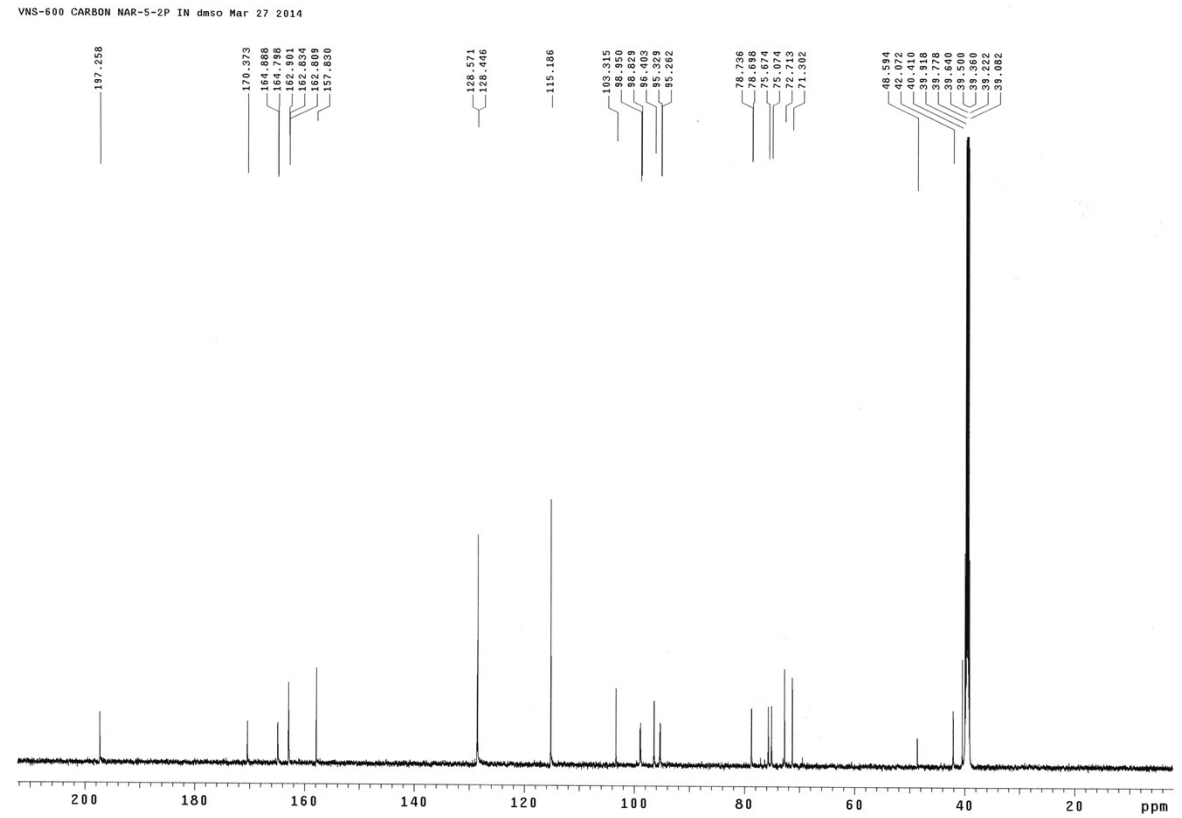


Fig. S31 The ^{13}C NMR spectrum of Narigenin-7-*O*-glucuronide (**21P**) in DMSO-d_6.

6. Supplementary References

[1] W. Wang, C. Meng, P. Zhu and K. D. Cheng, *China Biotechnology*, 2005, **25**, 35−39.

[2] L. J. Chen, D. E. Games and J. Jones, *J. Chromatogra. A*, 2003, **988**, 95−105.

[3] G. Flamini, M. Pardini, I. Morelli, K. Ertugrul, H. Dural, Y. Bagci and M. Kargioglu, *Phytochemistry*, 2002, **61**, 433–437.

[4] R. Barrington, G. Williamson, R. N. Bennett, B. D. Davis, J. S. Brodbelt and P. A. Kroon, *J. Funct. Foods*, 2009, **1**, 74–87.

[5] M. Kajjout and C. Rolando, *Tetrahedron*, 2011, **67**, 4731–4741.

[6] *China Pat*., 102875620 B, 2015.

[7] X. D. Wen, E. H. Liu, J. Yang, C. Y. Li, W. Gao, L. W. Qi, C. Z. Wang, C. S. Yuan and P. Ling, *J. Pharm. Biomed. Anal.*, 2012, **67-68**, 114–122.

[8] M. K. Khan, N. Rakotomanomana, M. Loonis and O. Dangles, *J. Agric. Food Chem.*, 2010, **58**, 8437–8443.

[9] S. Nagashima, M. Hirotani and T. Yoshikawa, *Phytochemistry*, 2000, **53**, 533−538.

[10] C. M. Ford, P. K. Boss and P. B. Hoi, *J. Biol. Chem.*, 1998, **273**, 9224−9223.