Supplementary Information

Characterization of two pterocarpan glycosyltransferases in *Astragalus membranaceus* and their application in whole-cell biocatalysis

Bai-Han Gao[1†], Meng Zhang[1†], Kuan Chen[1], Lin-Lin Wang[1], Ming-Ju Yao[1], Hong-Ye Li[1],Ming Xiong[1], Yu-Xi Huang[1], Yang Zhang[1], Quan-Li Liu[2], Juan Guo[3], Min Ye[1,4] and Xue Qiao[1,4]*

[1] State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, 38 Xueyuan Road, Beijing 100191, China

[2] Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences. No.189 Songling Road, Laoshan District, Qingdao, 266101, China

[3] State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, 100000, China

[4] Peking University-Yunnan Baiyao International Medical Research Center, 38 Xueyuan Road, Beijing 100191, China

† These authors contributed equally to this work.

*Corresponding author. E-mail addresses: qiaoxue@bjmu.edu.cn.

Table of Contents

NMR data compound 5a

Supplementary Figures

Figure S1. SDS-PAGE of His-tagged AmGT28/AmGT44 protein purified by Ni-NTA affinity chromatography.

Figure S2-15. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compounds **2-15** with UDP-Glc.

Figure S16. ^{1}HNMR spectrum of compound **5a** in DMSO-d_6 (400 MHz)

Figure S17. ^{13}CNMR spectrum of compound **5a** in DMSO-d_6 (100 MHz)

Figure S18. HMBC spectrum of compound **5a** in DMSO-d_6 (400 MHz)

Figure S19. HSQC spectrum of compound **5a** in DMSO-d_6 (400 MHz)

Figure S20. HR-MS spectrum of compound **5a**.

Figure S21. Sugar donor promiscuity of AmGT28/AmGT44

Figure S22. Effects of reaction buffer, temperature, and divalent metal ions on the activities of AmGT28.

Figure S23. Effects of reaction buffer, temperature, and divalent metal ions on the activities of AmGT44.

Figure S24. UHPLC/UV chromatograms of enzyme catalytic prodcuts demonstrating substrate preference of AmGT44 and AmGT28.

Figure S25. UHPLC/UV chromatograms of the extracts from whole-cell biocatalysis by *E. coli* cells harboring AmGT28/AmGT44 gene.

Figure S26. Calibration curves for **1a** and **2a**.

Figure S27. UHPLC/UV chromatograms of the whole-cell catalytic reaction mixtures with different substrate (**2**) concentration.

Figure S28. UHPLC/UV chromatograms of the whole-cell catalytic reaction mixture with a substrate concentration of 0.2 mM and a volume of 10 mL.

Supplementary Tables

Table S1. PCR primers used in this study.

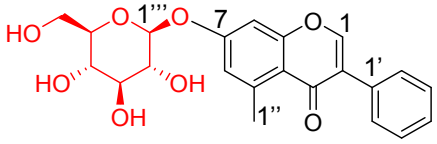
Table S2. UHPLC elution program for the enzymatic properties and the sugar donor preference experiments.

Table S3. UHPLC elution program for the substrate preference and whole-cell

catalysis experiments.

Table S4. Partition coefficients (logP) of the substrates.

NMR data compound 5a



5a: 7-Hydroxy-5-methyl-isoflavone 7-*O*-glucoside

^{1}H NMR (DMSO-d_6, 400 MHz): δ 8.34 (1H, s, H-2), 7.53 (2H, d, *J*=7.1 Hz, H-2', H-6'), 7.47-7.34 (3H, m, H-3', H-4', H-5'), 7.06 (1H, s, H-8), 6.91 (1H, s, H-6), 5.08 (1H, d, *J*=7.3 Hz, H-1'''), 3.11-3.75 (m, 6H, Glc H-2'''-H-6'''), 2.75 (3H, s, H-1''); ^{13}C NMR (100 MHz, DMSO-d_6): δ 176.5 (C-4), 160.1 (C-7), 158.6 (C-9), 152.7 (C-2), 142.1 (C-5), 132.2 (C-1'), 129.2 (C-2', C-6'), 128.1 (C-3', C-5'), 127.7 (C-4'), 124.8 (C-3), 117.5 (C-10), 117.1 (C-6), 101.7 (C-1'''), 99.9 (C-8), 77.2 (C-5'''), 76.5 (C-3'''), 73.2 (C-2'''), 69.7 (C-4'''), 60.6 (C-6'''), 22.9 (C-1''). HR-MS (ESI): *m/z* calcd. for $C_{22}H_{22}O_8$ [*M*+H]$^+$: 415.1387; found: 415.1404.

Supplementary Figures

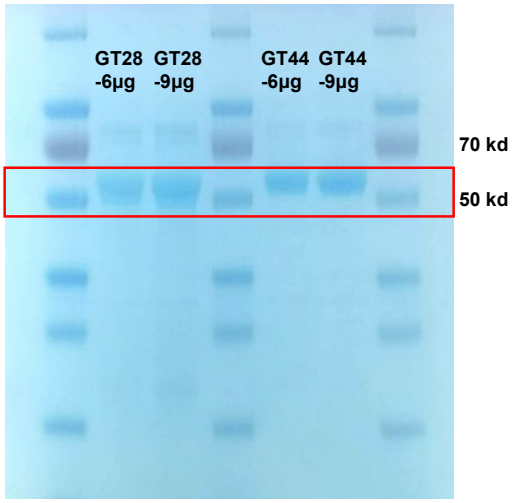


Figure S1. SDS-PAGE of His-tagged AmGT28/AmGT44 purified by Ni-NTA affinity chromatography.

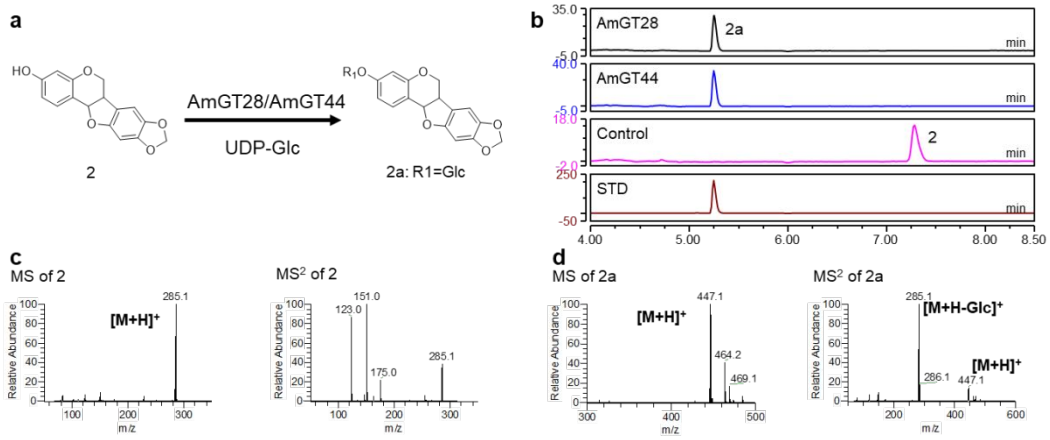


Figure S2. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **2** with UDP-Glc. (a) Reaction scheme. (b) UHPLC chromatogram of the reaction mixtures. Control, reactions conducted using boiled protein; STD, reference standard. (c) MS and MS/MS spectra for **2**. (d) MS and MS/MS spectra for **2a**.

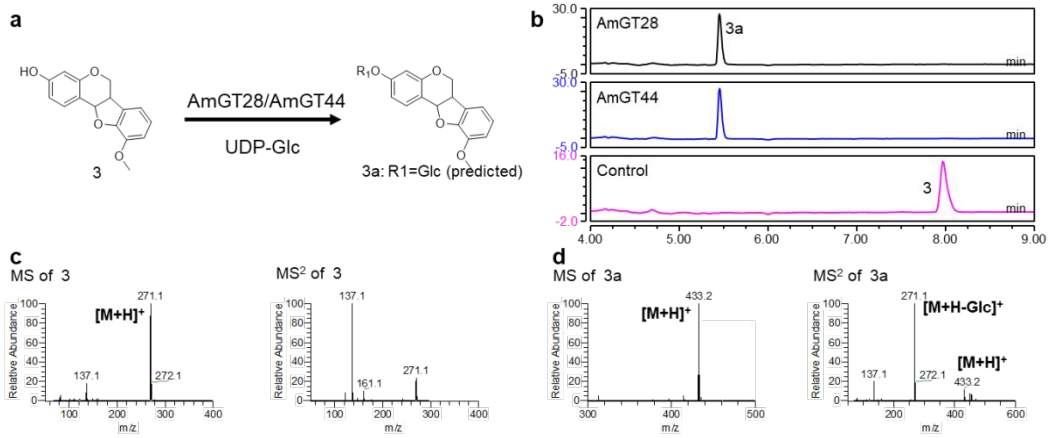


Figure S3. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **3** with UDP-Glc. (a) Reaction scheme. (b) UHPLC chromatogram of the reaction mixtures. Control, reactions conducted using boiled protein; STD, reference standard. (c) MS and MS/MS spectra for **3**. (d) MS and MS/MS spectra for **3a**.

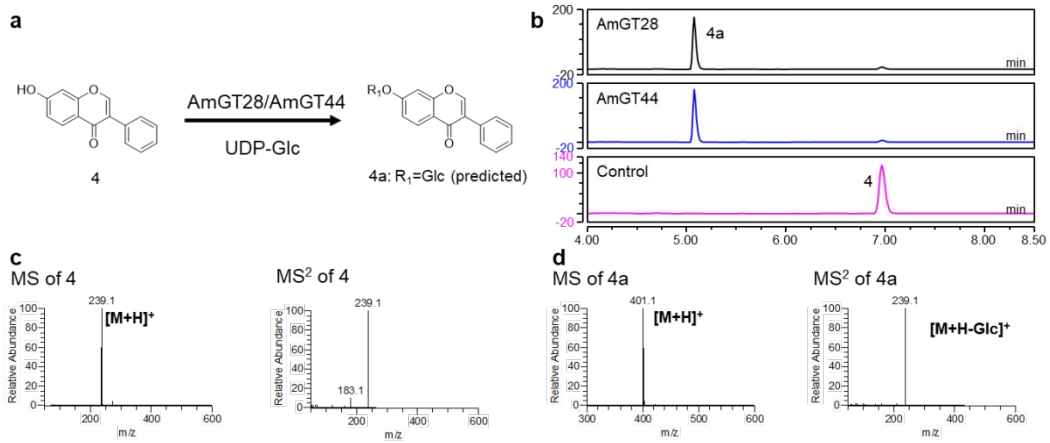


Figure S4. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **4** with UDP-Glc. (a) Reaction scheme. (b) UHPLC chromatogram of the reaction mixtures. Control, reactions conducted using boiled protein; STD, reference standard. (c) MS and MS/MS spectra for **4**. (d) MS and MS/MS spectra for **4a**.

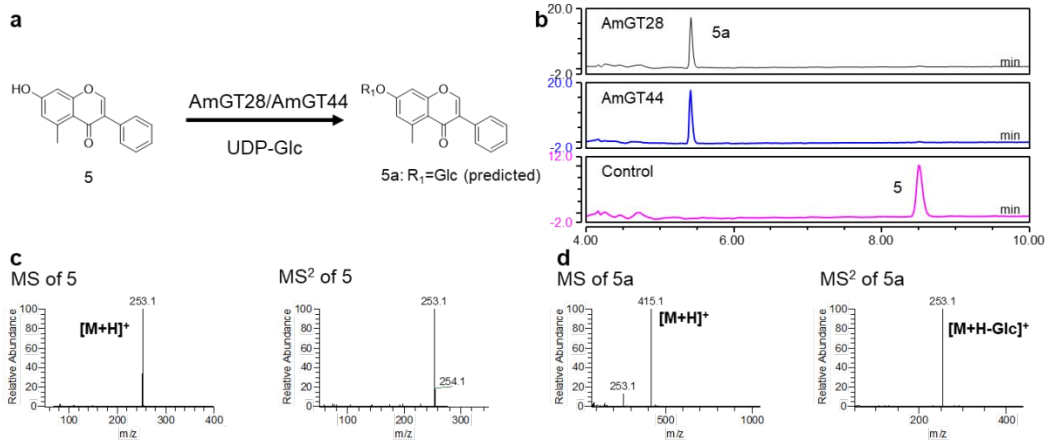


Figure S5. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **5** with UDP-Glc. (a) Reaction scheme. (b) UHPLC chromatogram of the reaction mixtures. Control, reactions conducted using boiled protein; STD, reference standard. (c) MS and MS/MS spectra for **5**. (d) MS and MS/MS spectra for **5a**.

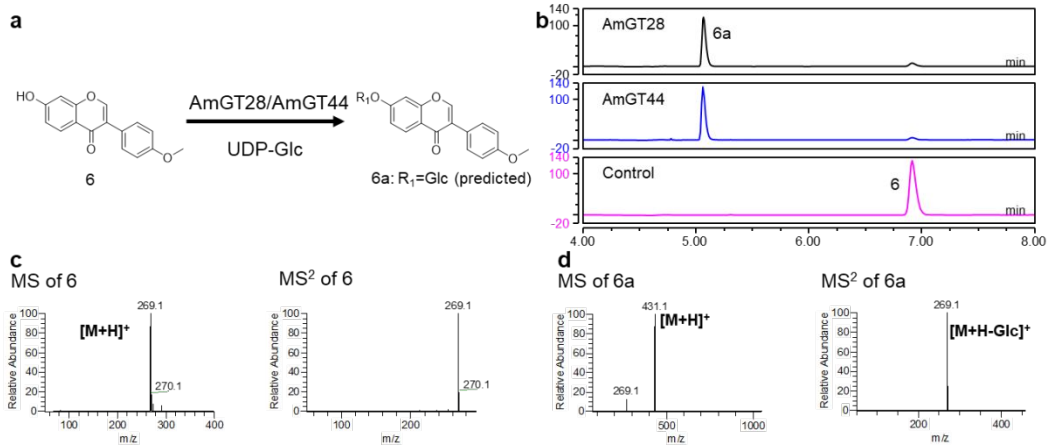


Figure S6. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **6** with UDP-Glc. (a) Reaction scheme. (b) UHPLC chromatogram of the reaction mixtures. Control, reactions conducted using boiled protein; STD, reference standard. (c) MS and MS/MS spectra for **6**. (d) MS and MS/MS spectra for **6a**.

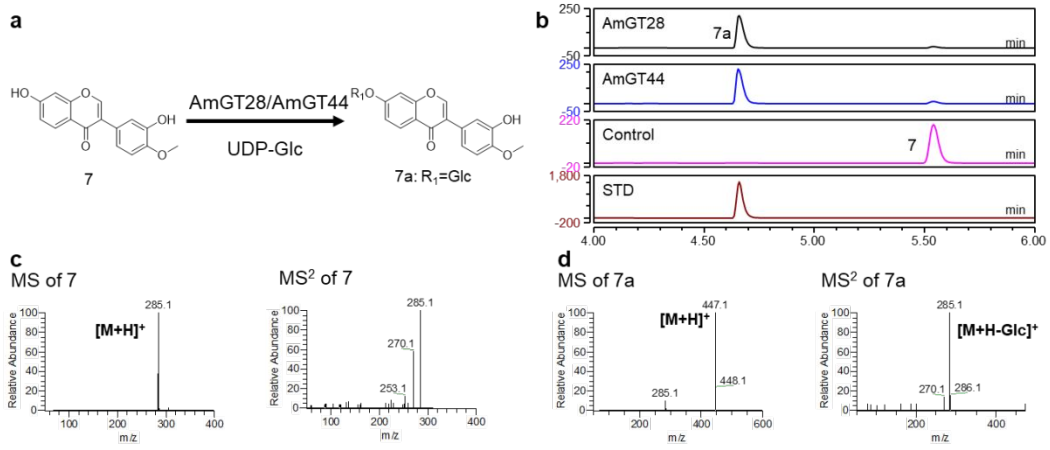


Figure S7. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **7** with UDP-Glc. (a) Reaction scheme. (b) UHPLC chromatogram of the reaction mixtures. Control, reactions conducted using boiled protein; STD, reference standard. (c) MS and MS/MS spectra for **7**. (d) MS and MS/MS spectra for **7a**.

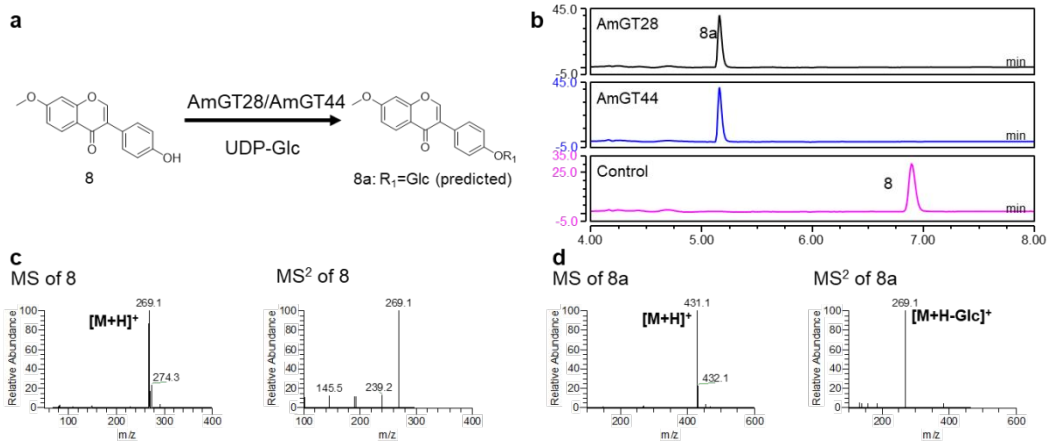


Figure S8. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **8** with UDP-Glc. (a) Reaction scheme. (b) UHPLC chromatogram of the reaction mixtures. Control, reactions conducted using boiled protein; STD, reference standard. (c) MS and MS/MS spectra for **8**. (d) MS and MS/MS spectra for **8a**.

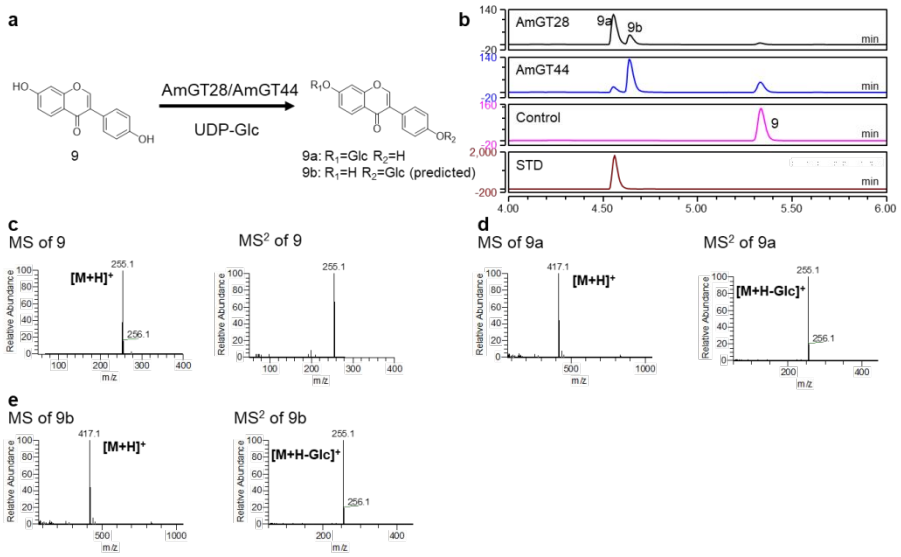


Figure S9. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **9** with UDP-Glc. (a) Reaction scheme. (b) UHPLC chromatogram of the reaction mixtures. Control, reactions conducted using boiled protein; STD, reference standard. (c) MS and MS/MS spectra for **9**. (d) MS and MS/MS spectra for **9a**. (e) MS and MS/MS spectra for **9b**.

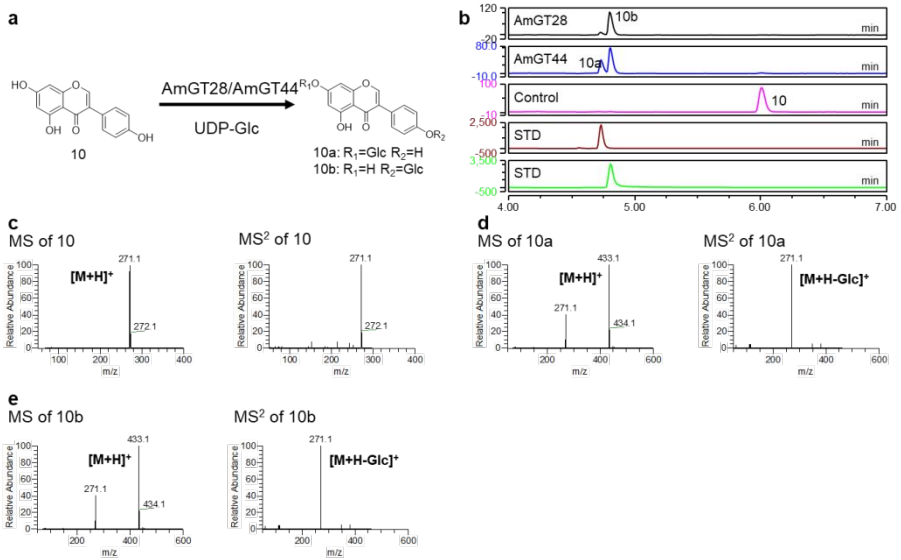


Figure S10. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **10** with UDP-Glc. (a) Reaction scheme. (b) UHPLC chromatogram of the reaction mixtures. Control, reactions conducted using boiled protein; STD, reference standard. (c) MS and MS/MS spectra for **10**. (d) MS and MS/MS spectra for **10a**. (e) MS and MS/MS spectra for **10b**.

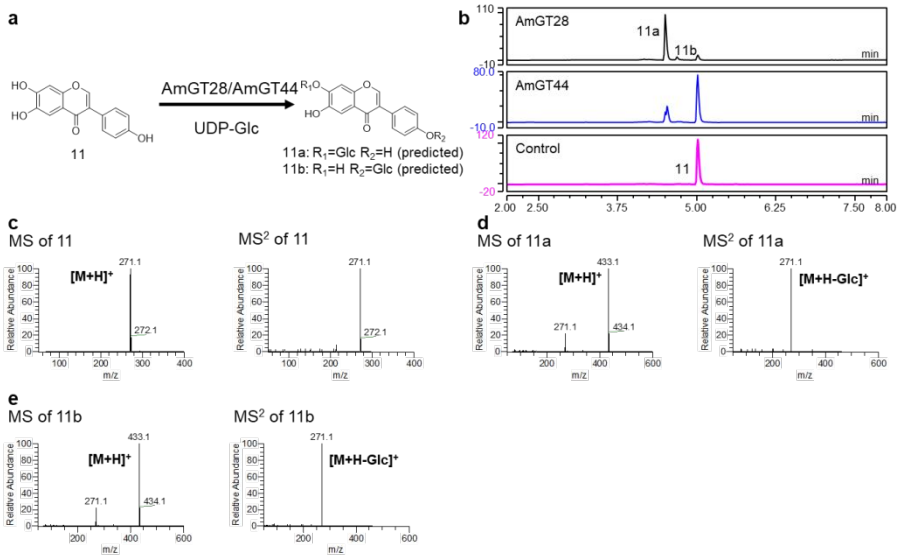


Figure S11. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **11** with UDP-Glc. (a) Reaction scheme. (b) UHPLC chromatogram of the reaction mixtures. Control, reactions conducted using boiled protein; STD, reference standard. (c) MS and MS/MS spectra for **11**. (d) MS and MS/MS spectra for **11a**. (e) MS and MS/MS spectra for **11b**.

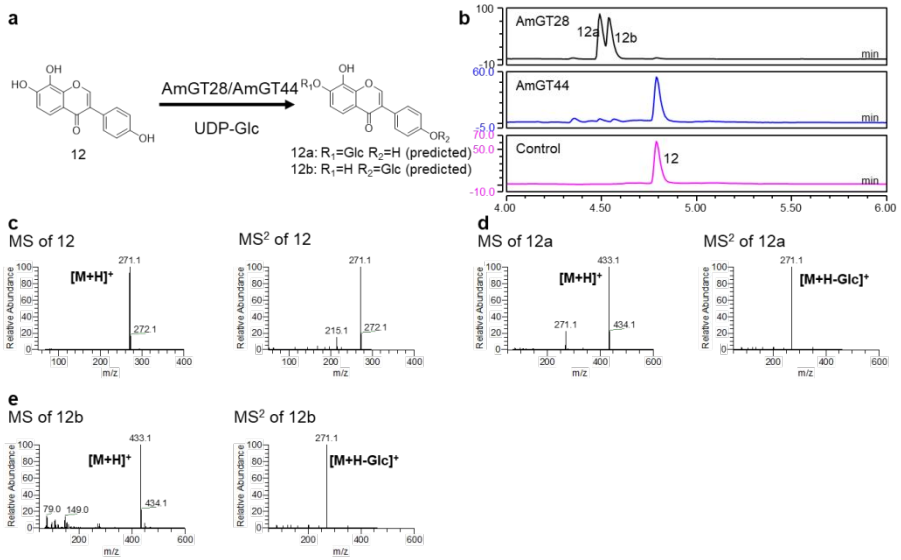


Figure S12. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **12** with UDP-Glc. (a) Reaction scheme. (b) UHPLC chromatogram of the reaction mixtures. Control, reactions conducted using boiled protein; STD, reference standard. (c) MS and MS/MS spectra for **12**. (d) MS and MS/MS spectra for **12a**. (e) MS and MS/MS spectra for **12b**.

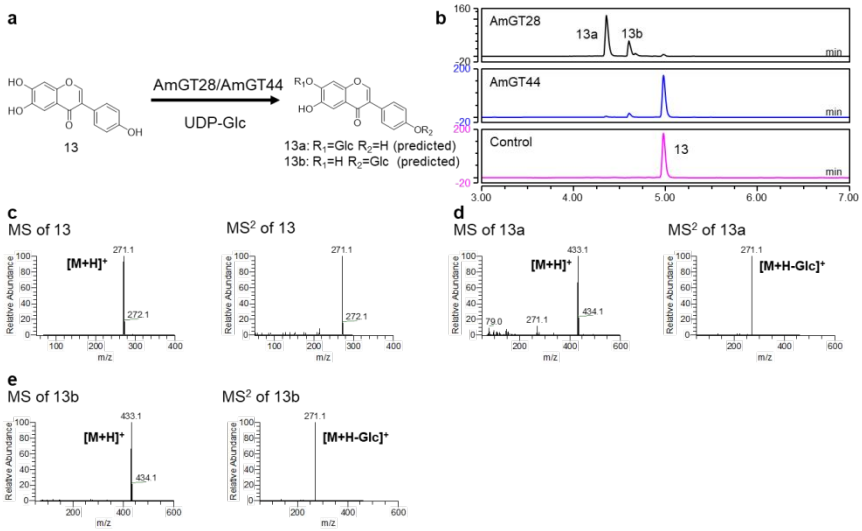


Figure S13. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **13** with UDP-Glc. (a) Reaction scheme. (b) UHPLC chromatogram of the reaction mixtures. Control, reactions conducted using boiled protein; STD, reference standard. (c) MS and MS/MS spectra for **13**. (d) MS and MS/MS spectra for **13a**. (e) MS and MS/MS spectra for **13b**.

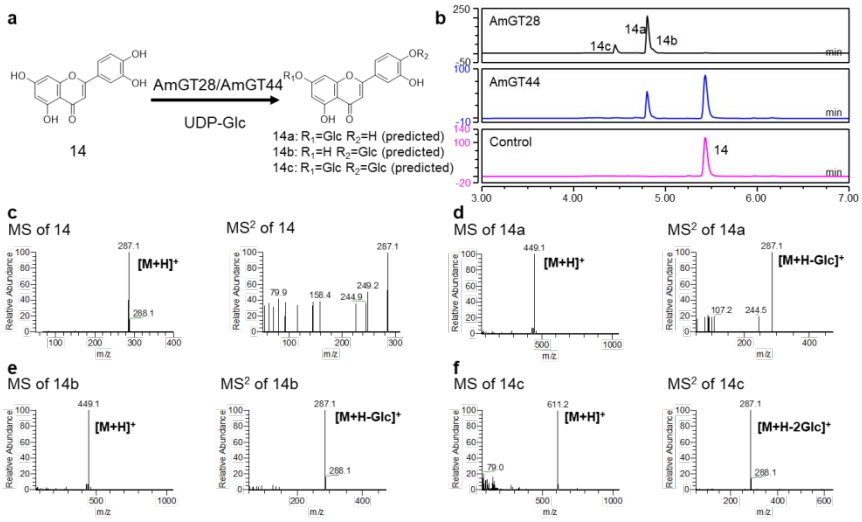


Figure S14. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **14** with UDP-Glc. (a) Reaction scheme. (b) UHPLC chromatogram of the reaction mixtures. Control, reactions conducted using boiled protein; STD, reference standard. (c) MS and MS/MS spectra for **14**. (d) MS and MS/MS spectra for **14a**. (e) MS and MS/MS spectra for **14b**. (f) MS and MS/MS spectra for **14c**.

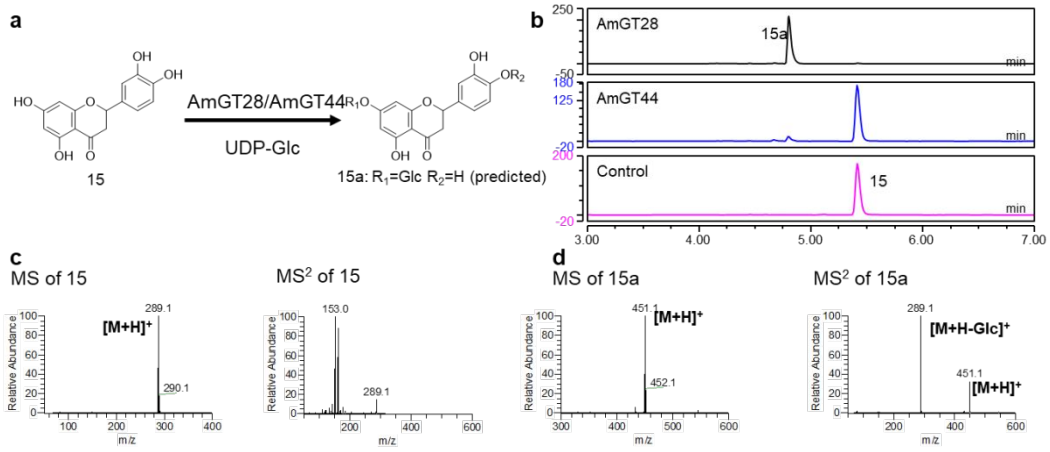


Figure S15. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **15** with UDP-Glc. (a) Reaction scheme. (b) UHPLC chromatogram of the reaction mixtures. Control, reactions conducted using boiled protein; STD, reference standard. (c) MS and MS/MS spectra for **15**. (d) MS and MS/MS spectra for **15a**.

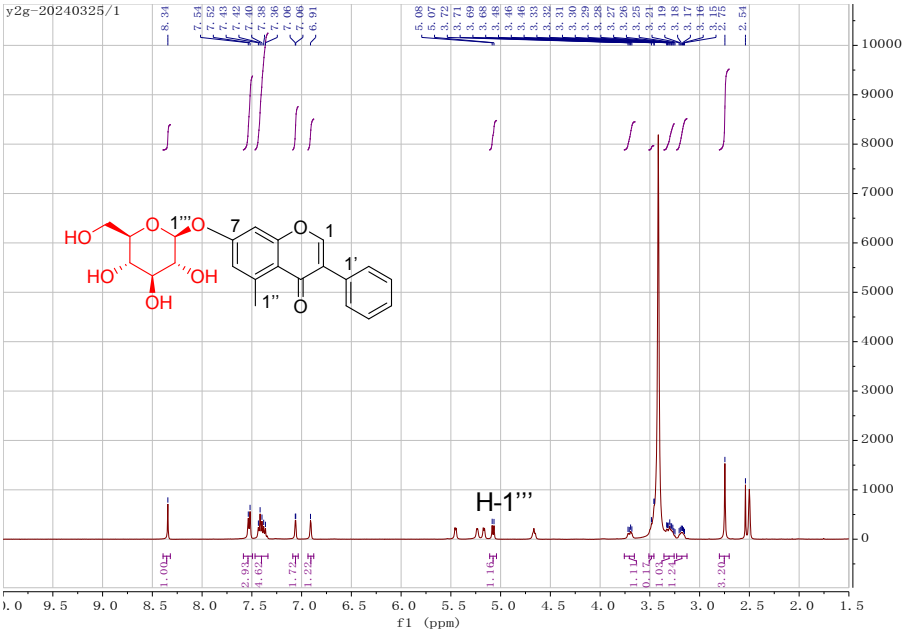


Figure S16. [1]HNMR spectrum of compound **5a** in DMSO-d_6 (400 MHz).

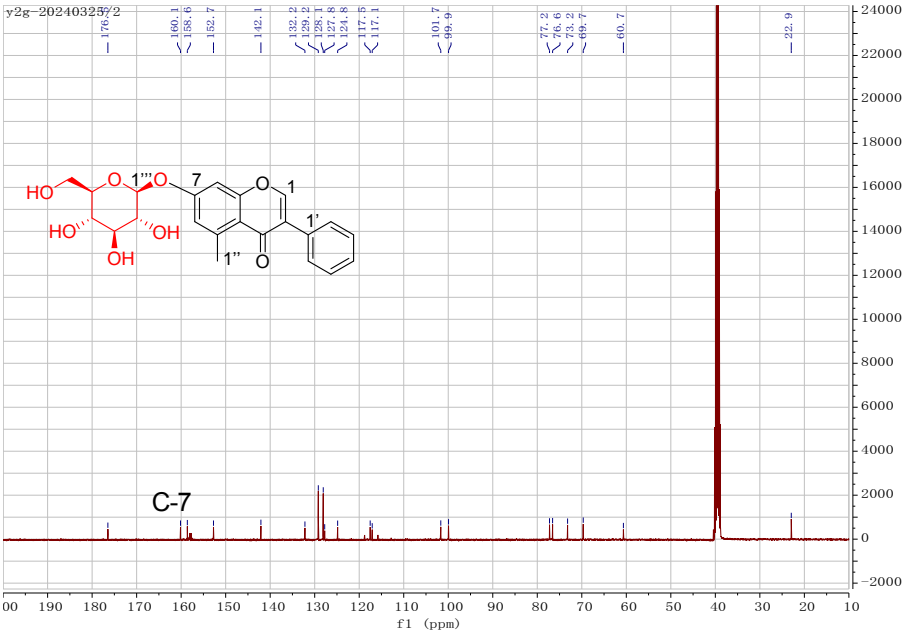


Figure S17. [13]CNMR spectrum of compound **5a** in DMSO-d_6 (100 MHz).

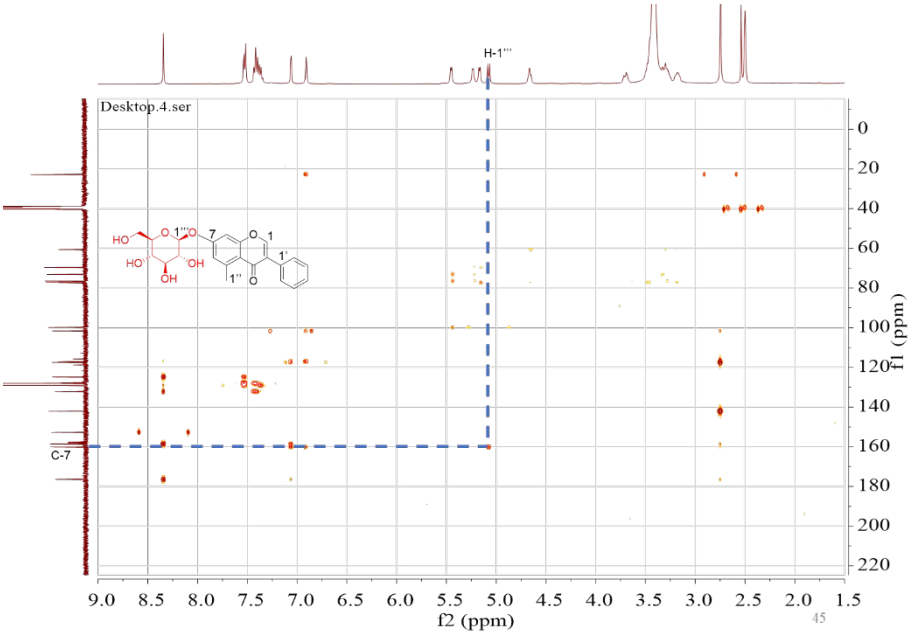


Figure S18. HMBC spectrum of compound **5a** in DMSO-d_6 (400 MHz).

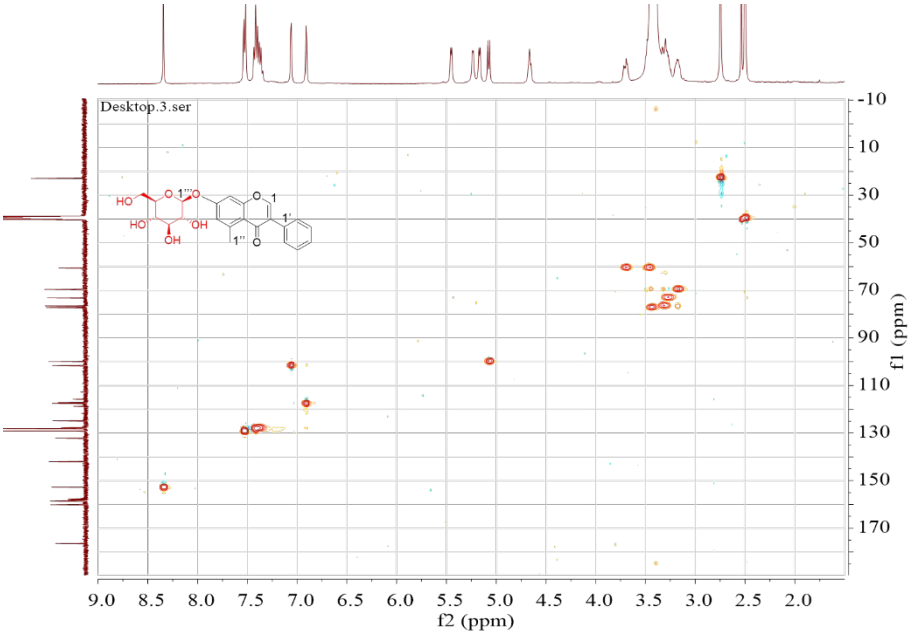


Figure S19. HSQC spectrum of compound **5a** in DMSO-d_6 (400 MHz).

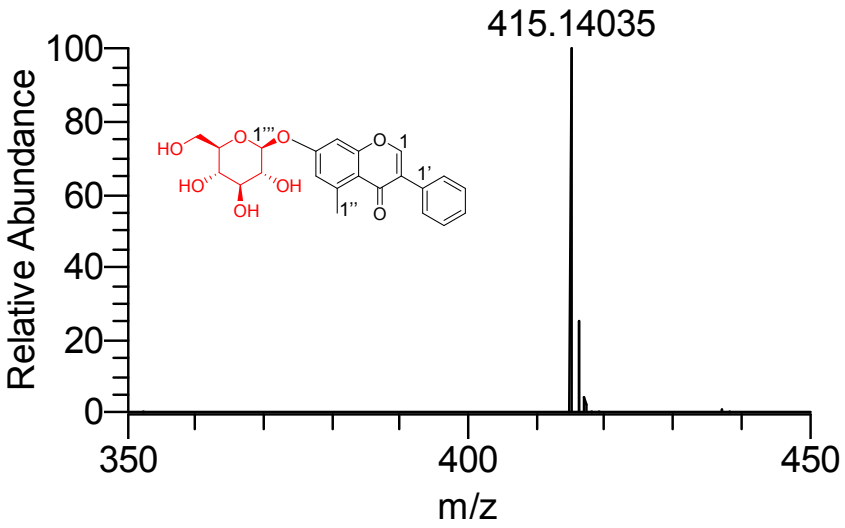


Figure S20. HR-MS (positive) spectrum of compound **5a**.

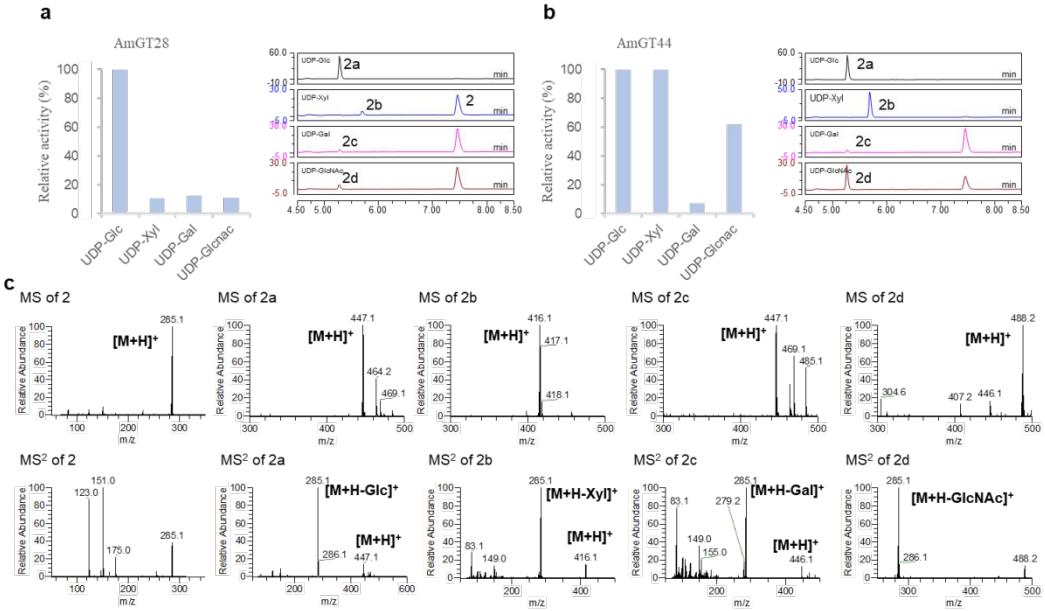


Figure S21. UHPLC/UV and UHPLC/MS analysis of AmGT28/AmGT44 catalytic reaction mixtures for compound **2** with UDP-Glc, UDP-Xyl, UDP-Gal and UDP-GlcNAc. (a, b) UHPLC chromatogram of the reaction mixtures. (c) MS and MS/MS spectra for **2**, **2a**, **2b**, **2c** and **2d**.

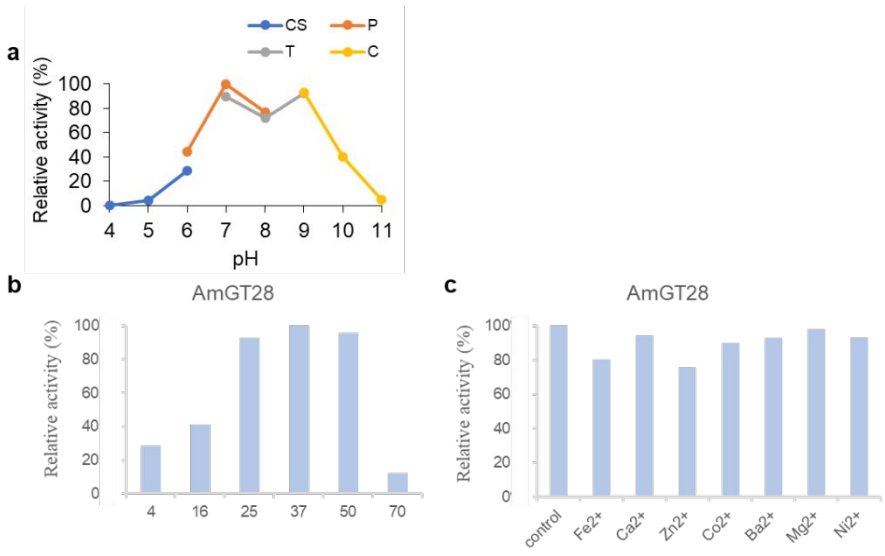


Figure S22. Effects of reaction buffer (a), reaction temperature (b), and divalent metal ions (c) on the activities of AmGT28. Maackiain (**2**) was used as the acceptor and UDP-Glc was used as the sugar donor. An optimized reaction time of 4 hour was used. AmGT28 exhibited its maximum activity at pH 7.0 (50 mM Na_2HPO_4-NaH_2PO_4) and 37°C.

To investigate the enzymatic properties of AmGT28, the pH, temperature and divalent metal ions were studied. All enzymatic reactions were carried out using UDP-Glc as the donor and maackiain (**2**) as the acceptor. The purified enzyme was added separately to the reaction solution and incubated at 37°C for 4 hours. To optimize the reaction pH, various buffers were utilized within different pH ranges: from 4.0-6.0 (citric acid-sodium citrate buffer), 6.0-8.0 (Na_2HPO_4-NaH_2PO_4 buffer), 7.0-9.0 (Tris-HCl buffer), and 9.0-11.0 (Na_2CO_3-$NaHCO_3$ buffer). To determine the optimal reaction temperature, the reactions were carried out at different temperatures (4-70°C). For testing the influence of various divalent cations (Fe^{2+}, Ca^{2+}, Zn^{2+}, Co^{2+}, Ba^{2+}, Mg^{2+}, Ni^{2+}), each cation was individually added to the reaction solution at a final concentration of 1 mM. The resulting mixtures were vacuum-dried, and the residue was dissolved in 150 μL methanol. Subsequently, the samples were centrifuged at 15,000 rpm for 30 minutes for UHPLC analysis.

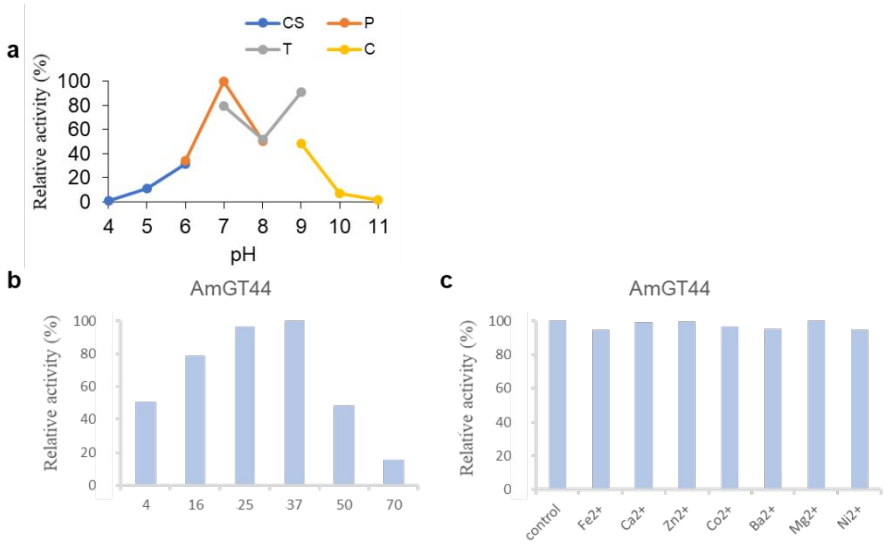


Figure S23. Effects of reaction buffer (a), reaction temperature (b), and divalent metal ions (c) on the activities of AmGT44. Maackiain (**2**) was used as the acceptor and UDP-Glc was used as the sugar donor. An optimized reaction time of 4 hour was used. AmGT44 exhibited its maximum activity at pH 7.0 (50 mM Na_2HPO_4-NaH_2PO_4) and 37°C.

To investigate the enzymatic properties of AmGT44, the pH, temperature and divalent metal ions were studied. All enzymatic reactions were carried out using UDP-Glc as the donor and maackiain (**2**) as the acceptor. The purified enzyme was added separately to the reaction solution and incubated at 37°C for 4 hours. To optimize the reaction pH, various buffers were utilized within different pH ranges: from 4.0-6.0 (citric acid-sodium citrate buffer), 6.0-8.0 (Na_2HPO_4-NaH_2PO_4 buffer), 7.0-9.0 (Tris-HCl buffer), and 9.0-11.0 (Na_2CO_3-$NaHCO_3$ buffer). To determine the optimal reaction temperature, the reactions were carried out at different temperatures (4-70°C). For testing the influence of various divalent cations (Fe^{2+}, Ca^{2+}, Zn^{2+}, Co^{2+}, Ba^{2+}, Mg^{2+}, Ni^{2+}), each cation was individually added to the reaction solution at a final concentration of 1 mM. The resulting mixtures were vacuum-dried, and the residue was dissolved in 150 μL methanol. Subsequently, the samples were centrifuged at 15,000 rpm for 30 minutes for UHPLC analysis.

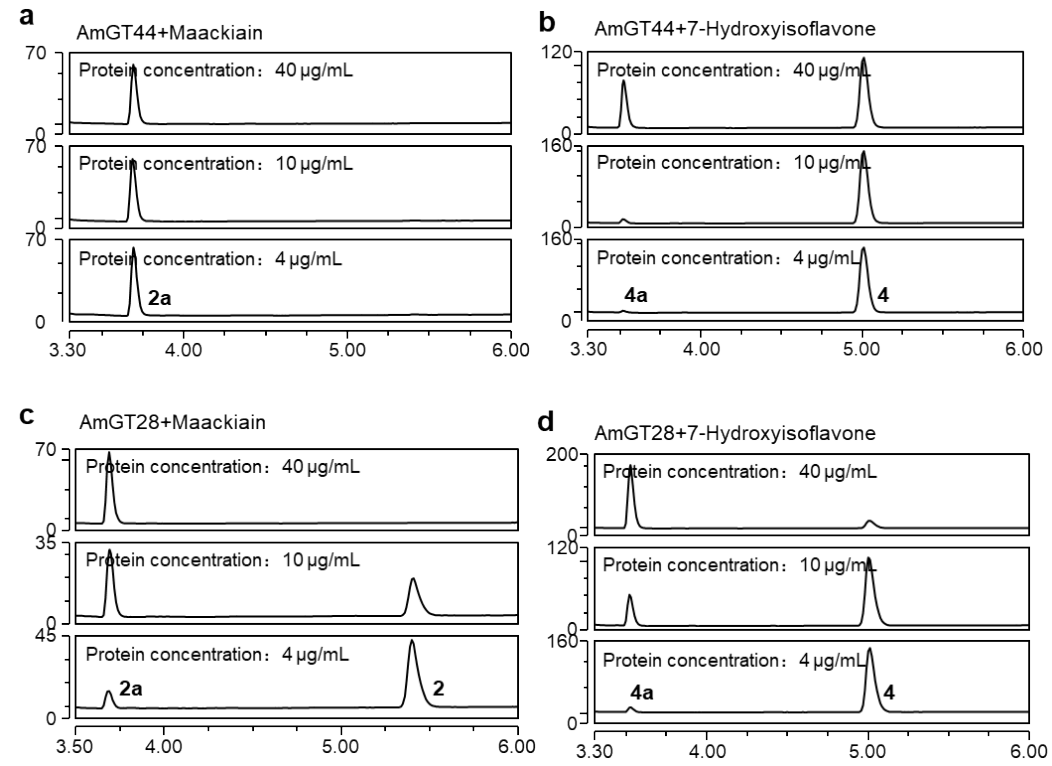


Figure S24. UHPLC/UV chromatograms of enzyme catalytic prodcuts demonstrating substrate preference of AmGT44 and AmGT28.

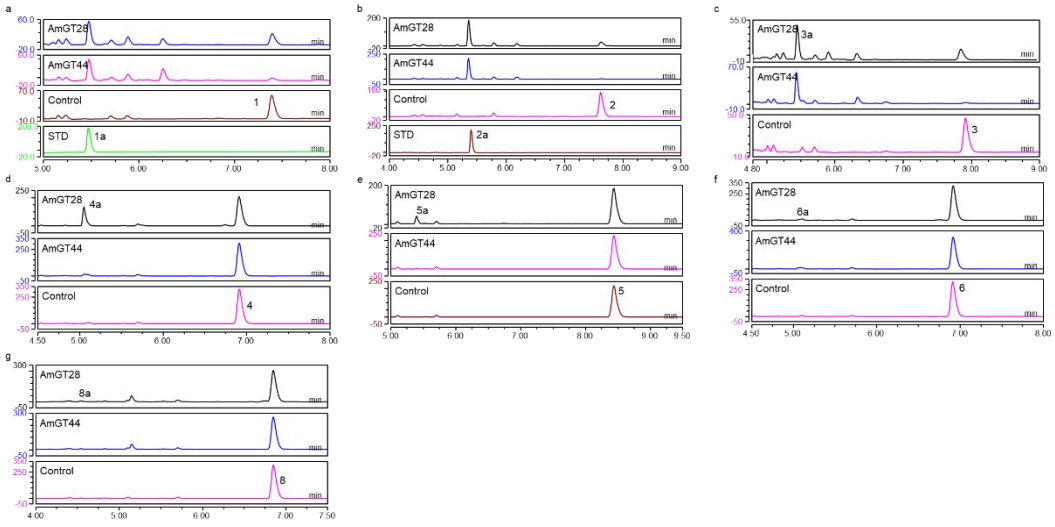


Figure S25. UHPLC/UV chromatograms of the extracts from whole-cell biocatalysis by *E. coli* cells harboring AmGT28/AmGT44 gene. Control, reactions conducted using boiled protein; STD, reference standard. The structures and numbers of substrates are given in Fig. 2c.

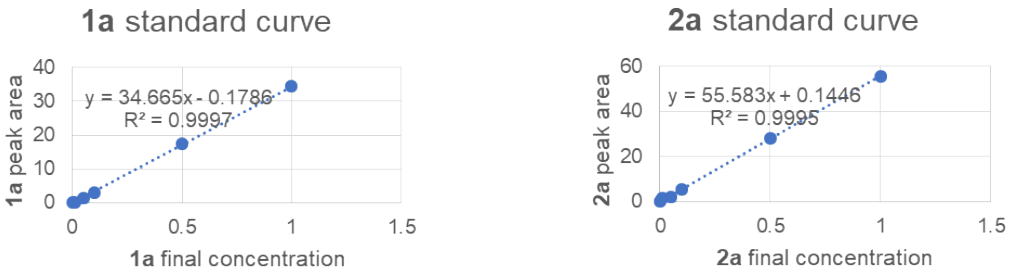


Figure S26. Calibration curves for **1a** and **2a**.

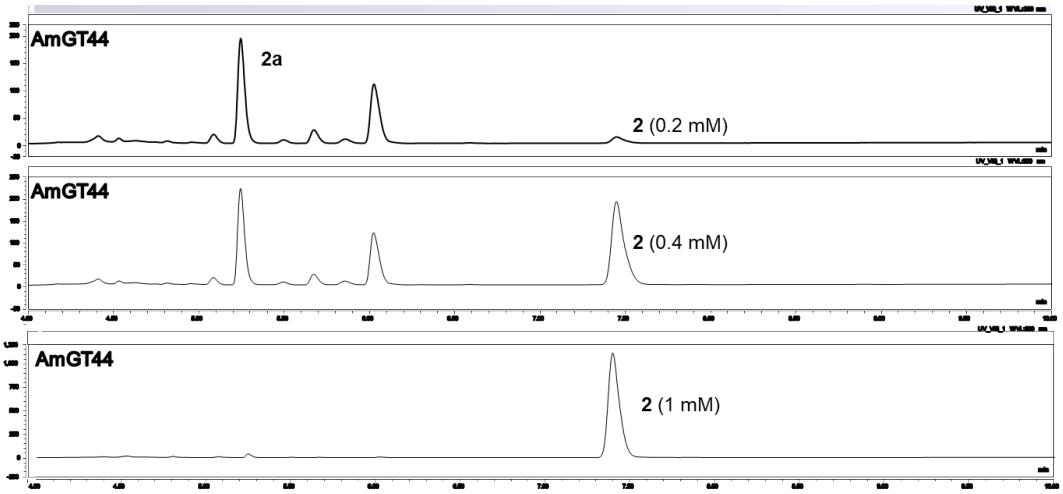


Figure S27. UHPLC/UV chromatograms of the whole-cell catalytic reaction mixtures with different substrate (**2**) concentration (0.2 mM, 0.4 mM, 1.0 mM, 500 μL). STD, reference standard.

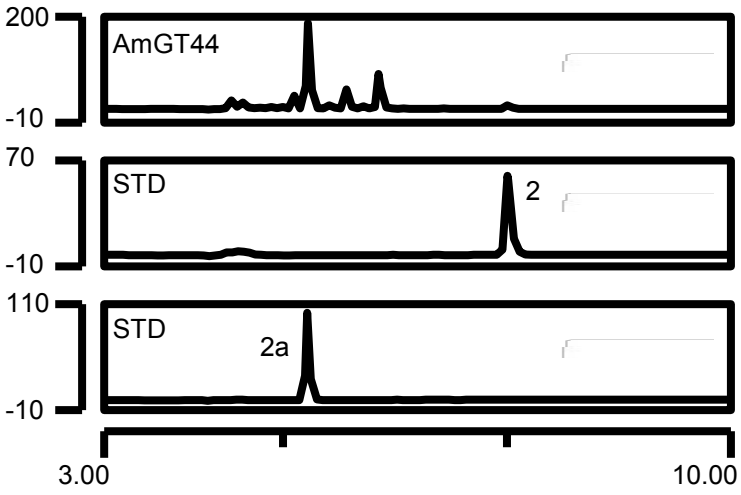


Figure S28. UHPLC/UV chromatograms of the whole-cell catalytic reaction mixture with a substrate (**2**) concentration of 0.2 mM and a volume of 10 mL. STD, reference standard.

Supplementary Tables

Table S1. PCR primers used in this study

Primers	Sequences (5' to 3')
AmGT28-F	GGTGGACAGCAAATGGGTCGCcgGATGAAGAAGAAGGACACAATAGTT
AmGT28-R	CTTGTCGACGGAGCTCGAATTCGGTTGTTTCCACAATTTGGCTAACCT
AmGT44-F	GGTGGACAGCAAATGGGTCGCcgGATGAAGGACACCATAGTTTTATAC
AmGT44-R	CTTGTCGACGGAGCTCGAATTCGGACGAAAAAGAAAAGGGGAATTAG G

Table S2. UHPLC elution program for the enzymatic properties and the sugar donor preference experiments

Time (min)	Flow (mL/min)	Aqueous phase (Water containing 0.1% formic acid)	Organic phase (Acetonitrile)
0	0.3	90	10
2	0.3	90	10
3	0.3	56	44
9	0.3	50	50
10	0.3	0	100
12	0.3	0	100
12.5	0.3	90	10
14.5	0.3	90	10

Table S3. UHPLC elution program for the substrate preference and whole-cell catalysis experiments

Time (min)	Flow (mL/min)	Aqueous phase (Water containing 0.1% formic acid)	Organic phase (Acetonitrile)
0	0.3	90	10
1	0.3	90	10
1.5	0.3	52	48
5.5	0.3	47	53
6	0.3	0	100
7	0.3	0	100
7.5	0.3	90	10
8.5	0.3	90	10

Table S4. Partition coefficients (logP) of the substrates

Substrate	Log P
1	2.35
2	2.39
3	2.48
4	2.52
5	3.01
6	2.4
7	2.01
8	3.01
9	2.13
10	1.74
11	1.74
12	1.74
13	1.74
14	1.51
15	1.24