Electronic Supplementary Material (ESI) for Green Chemistry.

A Phylogeny-based Directed Evolution Approach to Boost the Synthetic Applications of Glycosyltransferases

Peng Zhang[1], Yu Ji[1,*], Shuaiqi Meng[1], Zhongyu Li[1], Dennis Hirtz [3], Lothar Elling[3], Ulrich Schwaneberg[1,2,*]

1. Institute of Biotechnology, RWTH Aachen University, Worringerweg 3, 52074 Aachen, Germany

2. DWI-Leibniz Institute for Interactive Materials, Forckenbeckstraße 50, 52074 Aachen, Germany

3. Laboratory for Biomaterials, Institute of Biotechnology and Helmholtz-Institute for Biomedical Engineering, RWTH Aachen University, Pauwelsstraße. 20, 52074 Aachen, Germany

*Corresponding Authors: Ulrich Schwaneberg, u.schwaneberg@biotec.rwth-aachen.de; Yu Ji, yu.ji@biotec.rwth-aachen.de

Table of Contents

1. Experimental section

1.1. Retrieve GT1 family *Bacillus* glycosyltransferases from CAZy database

The information of GT1 family *Bacillus* glycosyltransferases (GTs) was retrieved from the release (until June 2021) of CAZy database (http://www.cazy.org/). After removing redundant sequences, a total of 1931 *Bacillus* GT sequences were further downloaded from the NCBI (GenBank) database using the Batch Entrez (https://www.ncbi.nlm.nih.gov/sites/batchentrez) according to the CAZy-supplied accession number.

1.2. Phylogenetic analysis

The selected 1931 GT sequences were conducted multiple sequence alignments using MAFFT (https://mafft.cbrc.jp/alignment/server/).[1, 2] FastTree and Jukes-Cantor evolution model were used to construct maximum-likelihood phylogenetic trees.[3] To infer a tree for a protein alignment with the JTT+CAT model and to quickly estimate the reliability of each clade in the tree with the Shimodaira–Hasegawa test.[4] The resulting phylogenetic trees were visualized by iTOL.[5]

1.3. Chemicals and reagents

All chemicals were of analytical grade and purchased from Sigma-Aldrich (Steinheim, Germany), AppliChem (Darmstadt, Germany), ABCR (Karlsruhe, Germany), TCI (Eschborn, Germany), and VWR International (Darmstadt, Germany).

1.4. Bacterial strains, plasmids and media

(Electro)competent *E. coli* BL21 (DE3) (Agilent Technologies, Santa Clara, CA) was prepared for the protein expression. The genes encoding BarGT-1 (QEL68904), BarGT-2 (QEL69054) and BarGT-3 (QEL68377) with pET-28a (+) vector were retrieved from the CAZy database and synthesized in GenScript (GenScript Biotech, Rijswijk, Netherlands). LB media: 10 g/L tryptone, 5 g/L yeast extract and 10 g/L NaCl; TB media: 800 mL solution A (12 g/L tryptone, 24 g/L yeast extract and 4 g/L glycerol) mixed with 200 mL solution B (2.31 g KH_2PO_4 and 12.54 g K_2HPO4); SOC media: 20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose

(Carl Roth GmbH, Karlsruhe, Germany).

1.5. Casting *error-prone* PCR and site-saturated mutant libraries

The BarGT-3 contains two domains (N- and C- domain), accordingly, the gene *pET28a-BarGT-3* was also divided into two fragments (fragment1: 604 bp and fragment2: 605 bp, **Table S4**). Cloning of the two fragments was further amplified by Megawhop.[6] The details for constructing casting *error-prone* PCR (cepPCR) mutant libraries were described in previous protocols.[7] For site-saturated mutagenesis (SSM), the SSM libraries of BarGT-3 were generated by one-step PCR (primers in **Table S4**). PCR mixture contained: VeriFi™ polymerase and mixes (50 μL, PCR Biosystems Ltd., London, UK), template (0.5 μL,10 ng), forward and reverse primers (500 nM, 2 μL each), and ddH_2O (up to 100 μL). The PCR condition was 95°C for 1 min, 95°C for 15 s/60°C for 15 s/72°C for 3.5 min (30 cycles), 72°C for 5 min. The amplified PCR products were digested by DpnI restriction endonuclease (2 μL, 10 U/μL) for additional 2 h to remove the template. The digested PCR products were purified by the Nucleospin Extract II kit (Macherey-Nagel, Dueren, Germany) and then transformed into the electrocompetent cells *E. coli* BL21(DE3) (1900 V). A total of 2160 colonies were picked in 96-well microtiter plate for the cepPCR mutant libraries (1080 colonies for N- and C- domain, respectively) and 180 colonies were for each single-site saturation mutagenesis library.

1.6. Cultivation of libraries in 96-well microtiter plates

Colonies of cepPCR and SSM mutant libraries were transferred from LB agar plates into 96-well microtiter plates (MTPs, PS-F-bottom, Greiner Bio-One, Frickenhausen, Germany) containing 150 μL/well LB media supplemented with kanamycin (40 μg/mL). After cultivation (37°C, 900 rpm, 70% humidity, 24 h) in a MTP shaker (Multitron II, Infors GmbH, Einsbach, Germany). The libraries were further transferred to the fresh 96-well MTPs (150 μL/well LB media, kanamycin 40 μg/mL) using MTP replicators for another cultivation (37°C, 900 rpm, 70% humidity, 24 h). Then, 50 μL glycerol (60%, w/w) was added and the MTPs were stored at −80°C. For expression, 5 μL of MTP seedings were transferred into MTPs (PS-V-bottom, Greiner Bio-One, Frickenhausen, Germany) with TB media (145 μL, kanamycin 40 μg/mL). After cultivation for 2 h (37°C, 900 rpm, 70% humidity), IPTG was added for protein expression with a final concentration of 0.1 mM (16°C, 900 rpm, 70% humidity, 18 h). Cells were

harvested by using a centrifuge (4°C, 4000 rpm, 30 min, 5810R Eppendorf, Hamburg, Germany) and stored at −20°C for 2 h. Cell pellets were thawed and resuspended in 150 μL lysozyme solution (5 mg/mL in Tris-HCl buffer 50 mM, pH 7.5) and incubated in a MTP shaker (37°C, 900 rpm, 70% humidity, 2 h). After centrifugation (4°C, 4000 rpm, 30 min), the clarified cell lysates were used for screening assay.

1.7. UDPG-recycling HTS system for GT variants screening

The reverse reactions from 4NPG to UDP or UDPG catalyzed by BarGT-3 (WT) were performed in a final volume of 100 μL Tris-HCl buffer (50 mM, pH 7.5, 10 mM $MgCl_2$) with 500 μM of 4-nitrophenyl-β-D-glucoside (4NPG), 100 μM of UDP or UDPG. For the UDPG-recycling HTS system, each MTP well reaction was performed in a final volume of 200 μL Tris-HCl buffer (50 mM, pH 7.5, 10 mM $MgCl_2$) with variants (or wild type (WT), empty vector (EV), 20 μL), 500 μM of 4-nitrophenyl-β-D-glucoside (4NPG), 100 μM of UDPG and 500 μM of substrate **1** (dissolved in DMSO). The reaction mixtures were incubated in a microtiter plate reader (TECAN Sunrise, Männerdorf, Switzerland) at 37°C, and the absorbance measurements were recorded every 10 min at 405 nm for 2 h, with the plate shaken within the reader for 10 s before collection of each time point. Improvement (ΔAbs, fold):

$$\frac{\text{Slope (Variant)}_{405\ nm} - \text{Slope (EV)}_{405\ nm}}{\text{Slope (WT)}_{405\ nm} - \text{Slope (EV)}_{405\ nm}}$$

1.8. Protein expression and purification

The recombinant vectors of *pET28a-BarGT-1*, *pET28a-BarGT-2* and *pET28a-BarGT-3* were transformed into *E. coli* BL21 (DE3) chemical competent cells for protein expression. 2 mL seeding transformants (active variants or WT) were transferred into a 300 mL flask containing 100 mL LB media (kanamycin, 40 μg/mL) for large-scale cultivation at 37°C until OD reached 0.4~0.6 and induced by adding IPTG (0.1 mM) for protein expression at 16°C for extra 20 h. The cell pellets were harvested by centrifugation at 4000 rpm for 20 min, washed twice with a Tris–HCl buffer (50 mM Tris–HCl, pH 7.5) and resuspended in 10 mL of the same buffer. The cells were sonicated in ice-water mixture (10 min, 10 s pulse, 20 s pause, 60% amplitude), and the supernatant was collected by centrifugation at 4000 rpm for 30 min at 4°C. The target proteins were further

analyzed by the 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

After being treated with 0.45 μm filters, the supernatants were purified using a His-Tag Ni-IDA column (MACHEREY-NAGEL, Dueren, Germany). The binding proteins were eluted with a Tris–HCl buffer (50 mM Tris–HCl, pH 7.5) containing different concentrations of imidazole (20 mM, 50 mM, 100 mM, 150 mM and 200 mM), and the eluates were analyzed by SDS-PAGE. Imidazole was removed from the protein solution by using ultrafiltration tubes (30 kDa, Millipore, Schwalbach, Germany). The concentrated protein samples were stored at −80°C for purified GTs assays.

1.9. *In vitro* glycosylation reactions

The purified GTs were subjected to *in vitro* glycosylation reactions. Each reaction was performed in a mixture (100 μL) containing final concentrations of 50 mM Tris-HCl buffer, 10 mM $MgCl_2$, 2 mM UDPG, 1 mM substrate (dissolved in DMSO) and 200 μg/mL each protein. All the above reactions were terminated with an equal volume of methanol. After centrifugation at 14,000 rpm for 30 min to remove protein precipitates, the reaction mixtures were analyzed by HPLC-MS.

1.10. HPLC-PDA, MS and NMR assays

The samples were subjected to HPLC-PDA (HPLC, Nexera X2, Shimadzu Deutschland GmbH, Duisburg, Germany) equipped with a C18 column (YMC CO., LTD, C18, 250 mm×4.6 mm, 5 μm), and equipped with an ESI-MS single quadrupole mass detector (ThermoFisher Scientific Surveyor MSQ Plus, Illinois, USA). Masses were determined via the mass/charge ratio (m/z) in a positive mode with 55 V cone voltage and 3 KV needle voltage at 400°C. Products were detected by UV absorbance from 200 to 400 nm. The binary mobile phase consisted of solvent A (HPLC grade purity water) and solvent B (HPLC grade purity acetonitrile) at a flow rate of 1 mL/min. HPLC methods used for analysis of glycosylated products are detailed in **Table S5**. For the preparative scale reaction, a 20 mL system containing enzyme (500 μg/mL), 10 mM sugar donors, 5 mM substrate (dissolved in DMSO), 50 mM Tris-HCl (pH 7.5) buffer, and 10 mM $MgCl_2$ was used. The reaction was incubated for 6 h at 37°C and stopped by adding the same volume of methanol. After centrifugation at 14,000 rpm and 4°C for 30 min, the supernatant was concentrated by rotary

evaporator, and separated by reversed-phase semipreparative HPLC (Shimadzu Deutschland GmbH, Duisburg, Germany) using a C_{18} column (CS-Chromatographie Service GmbH, Dueren, Germany, 125 mm × 8 mm, 10 μm) at a flow rate of 1 mL/min. 1D and 2D NMR spectra of each product (~ 10 mg) were recorded on a Bruker Avance III 400 NMR spectrometer (Bruker BioSpin, Germany).

1.11. Identification of kinetic parameters

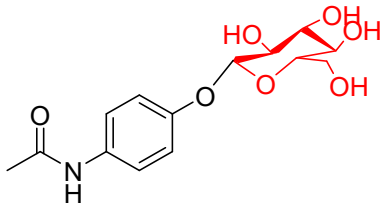
Kinetic parameters of BarGT-3 and its variants for the glycosylation of substrate **1** (acetaminophen) were measured in a reaction mixture (100 μL) with Tris–HCl buffer (50 mM Tris–HCl, pH 7.5, 10 mM $MgCl_2$) containing 20 μg enzyme, 4 mM UDPG and different concentrations of substrate **1** at 37°C for 10 min (triplicate and repeat 3 times). The samples were quenched with an equal volume of methanol and analyzed by HPLC. The kinetic values were determined by fitting the Michaelis–Menten curve to the data using the nonlinear regression method (**Figure S38**).

1.12. Computational analysis

The AlphaFold2 server is used to generate the structures of GT BarGT-3 and its variant I63P/H108G.[8] PROCHECK was used for the Ramachandran plot analysis.[9] The molecular dynamics (MD) simulations toward BarGT-3 or the variant were carried out by Yet Another Scientific Artificial Reality Application (YASARA, version 21.12.19).[10] The hydrogen bonding network and a pKa prediction were optimized to increase the solute stability and fine-tune the protonation states of protein residues (pH of 7.4 was selected), respectively.[11] The system was balanced in a net charge of zero, which was neutralized by Na^+ and Cl^-. The AMBER14 force field was used for the residues of BarGT-3.[12] The BarGT-3 or its variant were centered in a 10 Å simulation cell by using a periodic boundary. The box was filled with around 30394 water molecules. After the steepest descent and simulated annealing minimizations to remove clashes, the simulation system was performed at 298 K, pH 7.4 for 100 ns. The simulation snapshots were captured every 100 ps. The atomic type and force field parameters are taken from GAFF force field.[13] GROMACS tools and YASARA “md_analyze.mcr” macro were utilized for trajectory analysis.

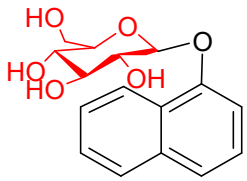
2. ^{1}H and ^{13}C NMR spectral data for glycosylated products 1a-9a

1a



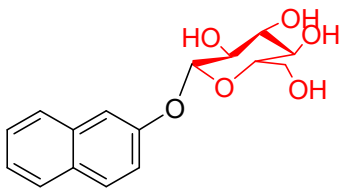
^{1}H NMR (400 MHz, DMSO-d_6) δ 9.83 (s, 1H), 7.47 (d, *J* = 9.0 Hz, 2H), 6.96 (d, *J* = 9.0 Hz, 2H), 5.29 (s, 1H), 5.05 (d, *J* = 25.8 Hz, 2H), 4.77 (d, *J* = 7.4 Hz, 1H), 4.57 (s, 1H), 3.68 (d, *J* = 11.5 Hz, 1H), 3.45 (d, *J* = 11.6 Hz, 1H), 3.37 (s, 1H), 3.27 – 3.10 (m, 3H), 2.00 (s, 3H). **^{13}C NMR (101 MHz, DMSO)** δ 167.9, 153.2, 133.7, 120.2, 116.4, 100.8, 77.0, 76.7, 73.3, 69.7, 62.1, 60.7, 58.1, 23.9.

2a



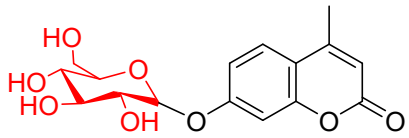
^{1}H NMR (400 MHz, DMSO-d_6) δ 8.33 (d, *J* = 7.4 Hz, 1H), 7.88 (d, *J* = 8.7 Hz, 1H), 7.52 (dq, *J* = 12.8, 6.8, 5.6 Hz, 3H), 7.42 (t, *J* = 8.0 Hz, 1H), 7.19 (d, *J* = 7.6 Hz, 1H), 5.47 (d, *J* = 5.4 Hz, 1H), 5.13 (d, *J* = 4.8 Hz, 1H), 5.05 (d, *J* = 5.2 Hz, 1H), 5.01 (d, *J* = 7.7 Hz, 1H), 4.59 (t, *J* = 5.7 Hz, 1H), 3.73 (dd, *J* = 11.6, 5.3 Hz, 1H), 3.54 – 3.47 (m, 1H), 3.45 – 3.41 (m, 1H), 3.40 – 3.33 (m, 2H), 3.23 (td, *J* = 9.1, 5.3 Hz, 1H). **^{13}C NMR (101 MHz, DMSO)** δ 153.0, 134.0, 127.3, 126.4, 126.1, 125.3, 122.1, 121.3, 109.0, 101.1, 77.2, 76.6, 73.4, 69.8, 60.7.

3a



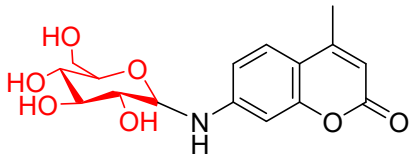
^{1}H NMR (400 MHz, DMSO-d_6) δ 7.86 (d, *J* = 3.2 Hz, 1H), 7.84 (s, 1H), 7.79 (d, *J* = 8.2 Hz, 1H), 7.46 (t, *J* = 5.9 Hz, 2H), 7.37 (t, *J* = 7.1 Hz, 1H), 7.26 (dd, *J* = 8.9, 2.4 Hz, 1H), 5.38 (d, *J* = 4.6 Hz, 1H), 5.13 (d, *J* = 4.3 Hz, 1H), 5.06 (d, *J* = 5.3 Hz, 1H), 5.03 (d, *J* = 7.3 Hz, 1H), 4.60 (t, *J* = 5.7 Hz, 1H), 3.73 (dd, *J* = 11.6, 3.7 Hz, 1H), 3.50 (dt, *J* = 11.8, 6.0 Hz, 1H), 3.45 – 3.39 (m, 1H), 3.31 (q, *J* = 5.7 Hz, 2H), 3.26 – 3.14 (m, 1H). **^{13}C NMR (101 MHz, DMSO)** δ 155.7, 134.5, 129.7, 129.5, 128.0, 127.4, 126.8, 124.5, 119.3, 110.9, 101.0, 77.6, 77.2, 73.8, 70.2, 61.2.

4a



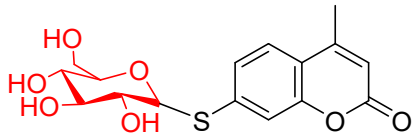
^{1}H NMR (400 MHz, DMSO-d_6) δ 7.70 (d, *J* = 9.1 Hz, 1H), 7.03 (d, *J* = 7.7 Hz, 2H), 6.25 (s, 1H), 5.39 (d, *J* = 4.7 Hz, 1H), 5.12 (d, *J* = 4.5 Hz, 1H), 5.04 (dd, *J* = 8.6, 6.3 Hz, 2H), 4.58 (t, *J* = 5.5 Hz, 1H), 3.70 (dd, *J* = 10.1, 5.3 Hz, 1H), 3.52 – 3.38 (m, 2H), 3.30 – 3.23 (m, 2H), 3.22 – 3.09 (m, 1H), 2.41 (s, 3H). **^{13}C NMR (101 MHz, DMSO)** δ 160.6, 160.6, 154.9, 153.8, 126.9, 114.6, 113.9, 112.2, 103.7, 100.5, 77.6, 77.0, 73.6, 70.1, 61.1, 18.6.

5a



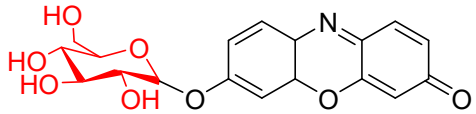
^{1}H NMR (400 MHz, DMSO-d_6) δ 7.49 (d, *J* = 8.7 Hz, 1H), 7.27 (d, *J* = 7.5 Hz, 1H), 6.73 (dd, *J* = 8.7, 2.1 Hz, 1H), 6.57 (d, *J* = 2.0 Hz, 1H), 5.99 (s, 1H), 5.01 (d, *J* = 25.8 Hz, 4H), 4.45 (t, *J* = 8.0 Hz, 2H), 3.66 (d, *J* = 11.4 Hz, 1H), 3.27 (t, *J* = 8.5 Hz, 2H), 3.19 (t, *J* = 8.6 Hz, 1H), 3.12 (t, *J* = 9.1 Hz, 1H), 2.33 (s, 3H). **^{13}C NMR (101 MHz, DMSO)** δ 161.1, 155.6, 154.2, 151.8, 126.3, 111.5, 110.5, 108.9, 98.9, 84.5, 78.1, 73.4, 70.6, 61.4, 61.4, 18.5.

6a



^{1}H NMR (400 MHz, DMSO-d_6) δ 7.66 (d, *J* = 8.4 Hz, 1H), 7.42 (s, 1H), 7.37 (d, *J* = 9.9 Hz, 1H), 6.33 (s, 1H), 5.46 (d, *J* = 5.1 Hz, 1H), 5.17 (s, 1H), 5.10 – 5.00 (m, 1H), 4.84 (d, *J* = 9.7 Hz, 1H), 4.64 (t, *J* = 5.0 Hz, 1H), 3.77 – 3.61 (m, 1H), 3.36 – 3.21 (m, 3H), 3.14 (t, *J* = 8.6 Hz, 2H), 2.40 (s, 3H). **^{13}C NMR (101 MHz, DMSO)** δ 160.2, 153.6, 153.6, 141.6, 125.9, 124.5, 117.8, 115.5, 113.9, 86.2, 81.5, 78.5, 72.9, 70.2, 61.4, 18.5.

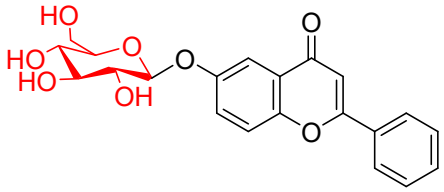
7a



^{1}H NMR (400 MHz, DMSO-d_6) δ 7.78 (d, *J* = 8.8 Hz, 1H), 7.53 (d, *J* = 9.8 Hz, 1H), 7.24 – 6.98 (m, 2H), 6.79 (dd, *J* = 9.8, 2.0 Hz, 1H), 6.28 (d, *J* = 1.9 Hz, 1H), 5.46 (d, *J* = 4.5 Hz, 1H), 5.17 (d, *J* = 4.4 Hz, 1H), 5.11 (t, *J* = 6.5 Hz, 2H), 4.62 (t, *J* = 5.4 Hz, 1H), 3.71 (dd, *J* = 9.6, 5.1 Hz, 1H), 3.53 – 3.40 (m, 2H), 3.35 – 3.24 (m, 4H), 3.18 (q, *J* = 7.3, 6.3 Hz, 1H). **^{13}C NMR (101 MHz, DMSO)** δ

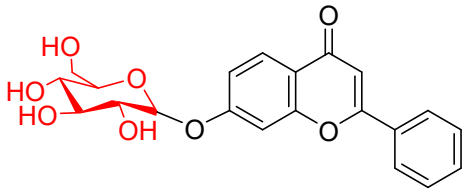
185.9, 161.4, 150.2, 146.2, 145.4, 135.4, 134.4, 131.7, 129.0, 115.3, 106.2, 103.1, 100.5, 77.6, 76.9, 73.6, 70.0, 61.1.

8a



^{1}H NMR (400 MHz, DMSO-d_6) δ 8.20 – 8.03 (m, 2H), 7.79 (d, *J* = 9.1 Hz, 1H), 7.57 (ddt, *J* = 21.7, 9.1, 4.1 Hz, 5H), 7.03 (s, 1H), 5.39 (d, *J* = 4.4 Hz, 1H), 5.12 (d, *J* = 3.5 Hz, 1H), 5.04 (d, *J* = 5.0 Hz, 1H), 4.98 (d, *J* = 7.2 Hz, 1H), 4.59 (t, *J* = 5.6 Hz, 1H), 3.70 (dd, *J* = 11.7, 3.4 Hz, 1H), 3.51 (dt, *J* = 11.6, 5.7 Hz, 1H), 3.40 – 3.35 (m, 1H), 3.30 – 3.16 (m, 3H). **^{13}C NMR (101 MHz, DMSO)** δ 177.3, 162.9, 155.2, 151.5, 132.3, 131.7, 129.6, 127.1, 126.8, 124.6, 124.5, 120.5, 109.9, 108.1, 106.7, 101.5, 77.6, 76.9, 73.8, 70.0, 61.1.

9a



^{1}H NMR (400 MHz, DMSO-d_6) δ 8.10 (dd, *J* = 8.0, 1.6 Hz, 2H), 7.97 (d, *J* = 8.8 Hz, 1H), 7.64 – 7.57 (m, 3H), 7.41 (d, *J* = 2.3 Hz, 1H), 7.15 (dd, *J* = 8.8, 2.3 Hz, 1H), 6.99 (s, 1H), 5.14 (d, *J* = 7.3 Hz, 1H), 3.72 (d, *J* = 9.9 Hz, 2H), 3.51 – 3.46 (m, 3H), 3.35 – 3.29 (m, 3H), 3.20 (t, *J* = 8.9 Hz, 2H). **^{13}C NMR (101 MHz, DMSO)** δ 176.9, 162.9, 162.2, 157.6, 132.2, 131.7, 129.6, 126.7, 126.7, 118.5, 116.0, 107.3, 104.3, 100.5, 77.7, 77.0, 73.6, 70.0, 61.3, 61.1, 40.9.

3. Additional Tables and Figures

Table S1. Genbank numbers of GTs in the nBGT branches (nBGT-1, -2, and -3 branch).

nBGT-1 branch				nBGT-2 branch			nBGT-3 branch			
AKO92307	ASI83653	QIE37132	AWU53427	QCJ40852	QEL79606	QGY35731	AXI30570	QFQ25983	QDQ05736	ACP14091
QDZ79798	ARO61460	ALL24124	ASE27575	AKP45588	AMX77761	QEF17374	AXI29309	QCC40235	AFQ12710	ACK91009
AEN88948	AKE17019	QKH57482	AQM46545	ASS95367	ALL59140	QCX94685	QCY28547	QEL73654	AAP25969	QKH72916
AWD67921	QFQ26430	AJH76572	APT26090	AGK54606	AKJ61866	ASK14909	AJI25776	QEL68377	ASI77642	AJK35990
QFY73879	AOM05681	AIE79619	ANR58054	AMM93225	AJK43082	ARV92898	QJX81068	QCT44488	AOY15547	QKU27705
AKP78182	QFR29855	AIY75337	ANR52759	QIZ06182	AJA20125	ANS48396	QCY28428	QBZ25094	QCU11396	QKU12805
AXI30102	QEF17221	AJG96512	ANR47460	AIE59987	AIE34220	AAP09752	AJI25618	ASI83079	ADY21406	QKU11542
AUO10047	ADH07188	AJI06405	ANR42170	AFU13409	AHZ51797		AYF07158	ASI72506	QEL84803	QKH65586
ADF39976	QCT44954	ARO18384	ANR36874	AXK18725	QKH15880		QKO33960	ARO65285	AIW85562	QKH26912
QCY26091	QBZ25613	BAL18312	ANR31570	QHA18026	QKH06032		AOY16482	AKE16462	QBP91709	QIY27416
AJI21740	ASI72979	ASI78066	ANR26269	AHA06092	QDZ74023		QDQ06588	AVR31858	ARJ21792	QIY24192
QJX76752	QGY35592	AAT54769	ANR20969	QEQ17642	QIZ42715		ASI78483	QCY61166	AJH19105	QIY20227
AYE50448	QCC40734	AFQ12216	ANR10369	ARC27942	ARX66651		AFQ11761	ARV92241	ABY43166	QIY14818
ADE70132	AZR77437	QHH88309	ANR05072	ABY43802	BCA31822		AIE79969	QGY35100	QIW18740	QHD21881
QLC87973	BAR84180	QCU10944	ANH86767	ARJ22432	AHX18834		AAT31200	QEF16730	AZJ20240	QHA64704
QDZ85886	QLF01873	ADY21879	AMC04577	QBP91130	QHV04619		QCY64928	QCX93999	AYF06111	QHA63532
QCR28179	ASK14747	AAU17876	ALC34716	AIW84865	QHV46425		ACM15767	AAP09035	AZQ47999	QHA58412
AVX09102	ALZ61621	AUD26034	BAR76650	AJH21198	AND08241		ASZ17681	AZR76915	AJG92909	QHA50907
AQU74603	QJU34994	AXO98681	AJM81107	QEL85441	ACK97105		QKH10521	ATI59305	AIY77413	QHA44171
ANF46916	AQQ63262	QBJ68344	AJI38545	AYF06867	AQY39298		QKH04759	QJU35461	ARO17947	QHA39772
BBK95448	AAP09581	AXO93311	AJH94493	ALQ68388	AXR17389		ANC22789	AQQ62754	AJI05851	QCX56408
AJI14555	QCX94525	AOY16041	AJH90397	QDF22468	AXR23120		AYF85303	ANC19090	QDF23177	AWU52923
AIK35241	ARV92760	QKI18600	AJH97227	AXY09564	QKH21836		AMR05901	ANP81107	AXY10282	ASE32253
QCS53509	AUB64188	AJH84742	AJH53825	ATI53055	QFP89258		AWC60795	AYF79901	QKI24588	AQM46029
AQX55196	AJA19978	QKI14904	AJH38570	ALL24258	QBM62490		AWC44565	AMR02476	AJI36413	APT25574
ALC81527	AMX77913	QKH02285	AJH36431	QIE37262	QBC62443		AWC36541	ACK63108	ATI52459	AHE83486
AKR09638	AKJ61711	AJI11247	AJH28258	QKH57353	QBC57218		AWC32513	QLF01436	AAU18367	ANR57567
AOM11265	AIE34382	AJG62268	AJG88662	AJH73272	QBC51988		ABS21876	ASK14182	AXO92776	ANR52272
QFY00254	AHZ51966	QKQ42541	AJG80787	QBJ68481	AND24623		AWC56688	ADH06668	QBJ67832	ANR46975
ANS48205	AGE78948	QKH31406	AJG69806	AXO93445	AFQ26497		AWC5255	ASZ65770	ALL23551	ANR41683

							6			
ANC13929	ALL58984	ABK85666	AJG65727	ARZ62874	QCJ49685		AWC4805 9	BAR84744	ASZ18155	ANR36387
ANC08107	AJK43074	AJH66958	AJG46980	QKH66248	AJH03721		AWC4012 8	QFR29366	QHH84611	ANR31085
AHA72074	AIM31272	ACO30839	AJG29314	AJH65579	AKR09792		AWC2848 8	ALZ62237	AIE79238	ANR25784
AVP48380	AKR35599	ASZ17703	AJF90192	ACK90322	ASL65524		AWC6481 5	ARP57438	QKH59125	ANR20484
ARP57993	ASJ48848	QHH85006	AJA86820	QKI18471	AOM11443		AIK36662	AGG00769	QIE36672	ANR09882
AGG01380	QDZ73877	QKH27401	AIK54753	QKH72254	AVP48227		AJI18252	AFV17833	AJH74544	ANR04585
AFV18433	QDD83876	AJG77446	AIK62286	AJK37207	QFY00403		AXK18081	AEA15711	QHH92190	ANH86289
AEA16298	ANE85799	QKH72394	AIK60228	AJH83040	AHA72249		AFU12716	AUB64715	EEL00975	AMC04096
QCY61747	AHX18685	QKH66108	AIK34279	QDQ06294	AST01546		QEQ17017	ASL64755	BAL17790	ALC34232
AYF84325	QKH15725	AJK33306	AIK08693	ALZ61473	ARP58158		ARC28624	AVP43818	QGG19884	AHE89379
ANP81578	QKH06185	AJH63079	AIF56819	ASI78190	AGG01550		AHA10508	AOM10741	QHG34794	AJM80618
AMR02971	ANV71901	ACK88962	AHE89936	ACM13064	AFV18596		QHA17341	ALC53357	QHA26401	AJI38763
AXR17223	QEL79454	QKE07672	AHE84053	QIW22576	AEA16461		QDD83276	QFX99623	AYY26653	AJH96844
AXR22945	QIZ42564	ARZ62737	AAP26486	AZJ20890	ANC14087		QEL78925	ANS47663	ACM12493	AJH88375
AFQ14800	QEL74182	QKI25077	ACQ48454	AJG96177	ANC08266		QHV47063	ANC13314	ACJ77544	AJI01235
AQY39098	QEL68904	AJI33394	ACP14023	ARO18515	ADH07315		QHV03977	ANC07491	ADK04697	AJH62205
AMR85067	ARX66490	ADK05248		AJI05809	AQQ63417		ALL59929	AHA71485	ACO26709	AJH56121
AJQ59336	ANN32615	AJH57017		AIY73504	QJU34852		AJK42731	AKR09134	AEW55175	AJH52019
ACK96959	BCA31658	BBK95449		AFQ12067	QKO33670		AIM32357	AMR84536	ABK85173	AJH38450
QGV08483	QHV46578	AHK38686		ADY22023	QBZ25778		AMX76971	AJQ58787	AJH69939	AJH37021
QKH21635	QHV04464	AIM11887		QCU10820	ASI73129		AKJ57264	AQY38635	QKQ42038	AJH30070
QFP89064	AFU13263	AIM06450		AOY16180	ARO60391		AJA19468	QGV08958	QKI15452	AJG89233
QCJ49880	QHA17887	AFH83899		ASZ17565	QEL74332		AIE33375	QCJ50360	QKH02781	AJG83672
QBM62311	QEQ17508	AAT31754		QHH85108	QEL69054		AHZ50965	AJH08239	QKH30902	AJG76214
QBC62255	ARC28077	QKU12352		QHG35329	QCT45110		AGE77890	AFQ15382	AJI10755	AJG72481
QBC57019	AJH22268	QKU07083		QHA26970	AVR32550		QKH06722	ACK96743	AJG57414	AJG63227
QBC51797	ARJ22349	QIY26965		AYY27255	ASI83816		QKH15190	QKH21152	AUD25561	AJG48661
AND24420	AIW84961	QIY23748		BAL18449	AKE17176		QDZ73376	QJU67094	AXO98138	AJG28830
AND08095	QEL85354	QIY15808		QGG19290	ARO62529		AHX18210	QFP88588	QKE07142	AJF89710
AJH04559	ABY43707	QIY15271		ACJ81078	ASZ66567		ANE85312	QBM61883	ARZ62256	AJA86366
AFQ26279	ALQ68242	QHD22336		ANE85944	BAR84029		ASJ48342	QBC61786	AIM11353	AIK52342
AZV66524	QIW22107	QHA64258		QDD84047	ALC52657		AKR34955	QBC56538	BBK94905	AIK64512
ASL65356	AYF06729	QHA63090		AGE78765	ACK60575		ARX65978	QBC51322	BBB72256	AIK57480
ASZ66412	AZJ20754	QHA54018		AIM31459	AOM05866		QKO33032	AZV66002	BAR77186	AIK31256
ANC19652	AZQ47631	QHA51354		ANV71756	AZR77585		BCA36388	AXR16522	AIM05932	AIF56332
QKO33526	QDQ06158	QHA48563		ASJ48991	QFR30009		ANN32142	AXR22255	AIK08141	AAT54250
ACK62605	QDF22625	QHA40219		AKR35750	QCC40886		ANV72408	AND23938	AHK38148	
ARO62609	AXY09721	QCX56919		AIM31458	ATI59942		QIZ42088	AND07562	AFH83363	

AVR32392	ATI52929	BBB72787		AGE78766	QLF02022		AOM05164	AFQ25764	ACQ47473	

Table S2. GT sequence identity of BarGT-1, -2 and -3 in the nBGT branches with the typical GTs (YjiC, YojK and YdhE) in the main branches.

Identity (%)	BarGT-1	BarGT-2	BarGT-3	YjiC	YojK
BarGT-2	31.4				
BarGT-3	33.7	37.6			
YjiC	27.6	29.1	30.9		
YojK	30.8	37.7	43.6	30.8	
YdhE	52.7	30.1	32.8	28.2	33.9

Table S3. The specific activities (U/mg) of BarGT-3 (WT), I63P and I63P/H108G towards 9 substrates (Each reaction was performed in a 100 μL mixture containing final concentrations of 50 mM Tris-HCl buffer, 10 mM $MgCl_2$, 2 mM UDPG, 1 mM substrate and 200 μg/mL each protein for 2 h at 37°C).

	Specific activity (U/mg)		
	WT	**I63P**	**I63P/H108G**
1a	0.004	0.02	0.02
2a	0.03	0.04	0.04
3a	0.04	0.04	0.04
4a	0.03	0.04	0.04
5a	0.01	0.01	0.03
6a	0.01	0.02	0.03
7a	0.01	0.02	0.03
8a	0.03	0.04	0.04
9a	0.03	0.04	0.04

Table S4. The primers for casting *error-prone* PCR and SSM libraries.

Primers	Primers Sequence (5' to 3')
BarGT3-fragment1-F	CGGCCTGGTGCCGCGCGGCAGCCAT
BarGT3-fragment1-R	GGGCCCACAAATTTATAGCTTTCAT
BarGT3-fragment2-F	TATCAGCCGCGCAGCGAAGTGTTTG
BarGT3-fragment2-R	GGTGCTCGAGTGCGGCCGCAAGCTT
BarGT-3-I63SSM-F	GATTAACnnsATGGAACGCGTGAACGAAGGCG
BarGT-3-I63SSM-R	GTTCCATsnnGTTAATCTGACTCAGAAAGTTTTCAAACA
BarGT-3-H108SSM-F	TGATAACnnsTTTCCGGTGGGCCGCATTATTG
BarGT-3-H108SSM-R	CCGGAAAsnnGTTATCATAAATCAGATAATCATATTTTTCTTC

Table S5. HPLC methods and physical parameters for glycosylated products **1a-9a**.

Substrate	Analytical methods	Preparative methods	UV (nm)	Melting points (℃)	α/β nature of glycosidic linkage
1	5% solvent B, 25 min	10% solvent B, 25 min	257	198	*O*-**β**-D-linkage
2	20-90% B, 30 min	35% B, 30 min	230	182	*O*-**β**-D-linkage
3	20-90% B, 30 min	35% B, 30 min	230	183	*O*-**β**-D-linkage
4	20-90% B, 30 min	20% B, 30 min	320	260	*O*-**β**-D-linkage
5	20-90% B, 30 min	20% B, 30 min	320	254	*O*-**β**-D-linkage
6	20-90% B, 30 min	20% B, 30 min	320	243	*O*-**β**-D-linkage
7	20-90% B, 30 min	30% B, 30 min	320	220	*O*-**β**-D-linkage
8	20-90% B, 30 min	35% B, 30 min	330	256	*O*-**β**-D-linkage
9	20-90% B, 30 min	30% B, 30 min	330	258	*O*-**β**-D-linkage
4NPG	20% B, 30 min	-	350		-

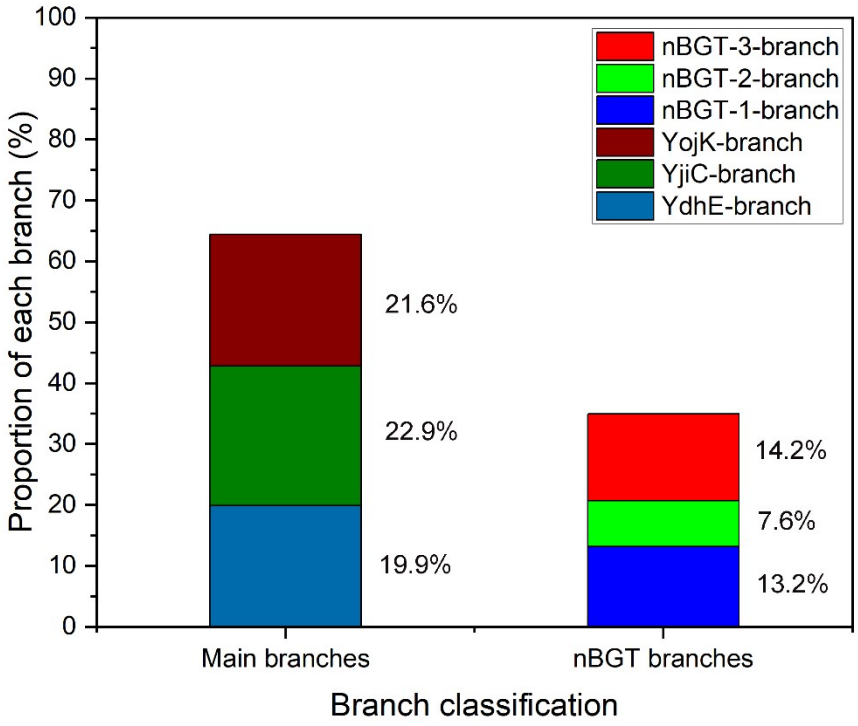


Figure S1. The proportion (%) of each branch in *Bacillus* GT1 family enzymes. YdhE-, YojK-, YjiC-branch (main branches) account for 19.9%, 22.9% and 21.6% of all GTs, respectively. nBGT-1-, nBGT-2-, nBGT-3-branch (novel Bacillus GTs branches) account for 13.2%, 7.6% and 14.2% of all GTs, respectively.

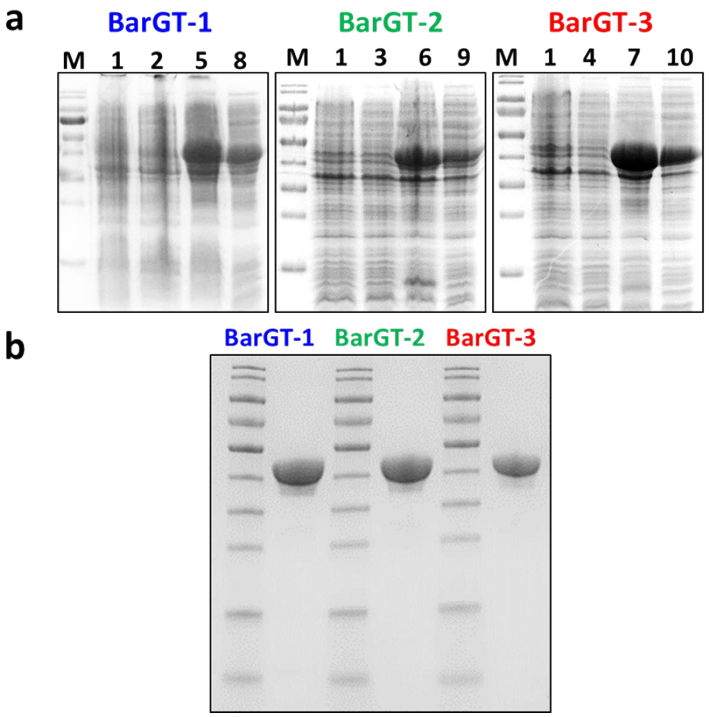


Figure S2. SDS-PAGE detection of the expression of BarGT-1, -2 and -3. (a). All of the GTs were soluble expression. M is the protein molecular weight marker, Lane 1, the control of pET 28a vector; Lanes 2-4, the *E. coli* lysate before induction; Lanes 5-7, the *E. coli* lysate after induction; Lanes 8-10, the soluble supernatants after induction. (b). the purified proteins of BarGT-1, -2 and -3.

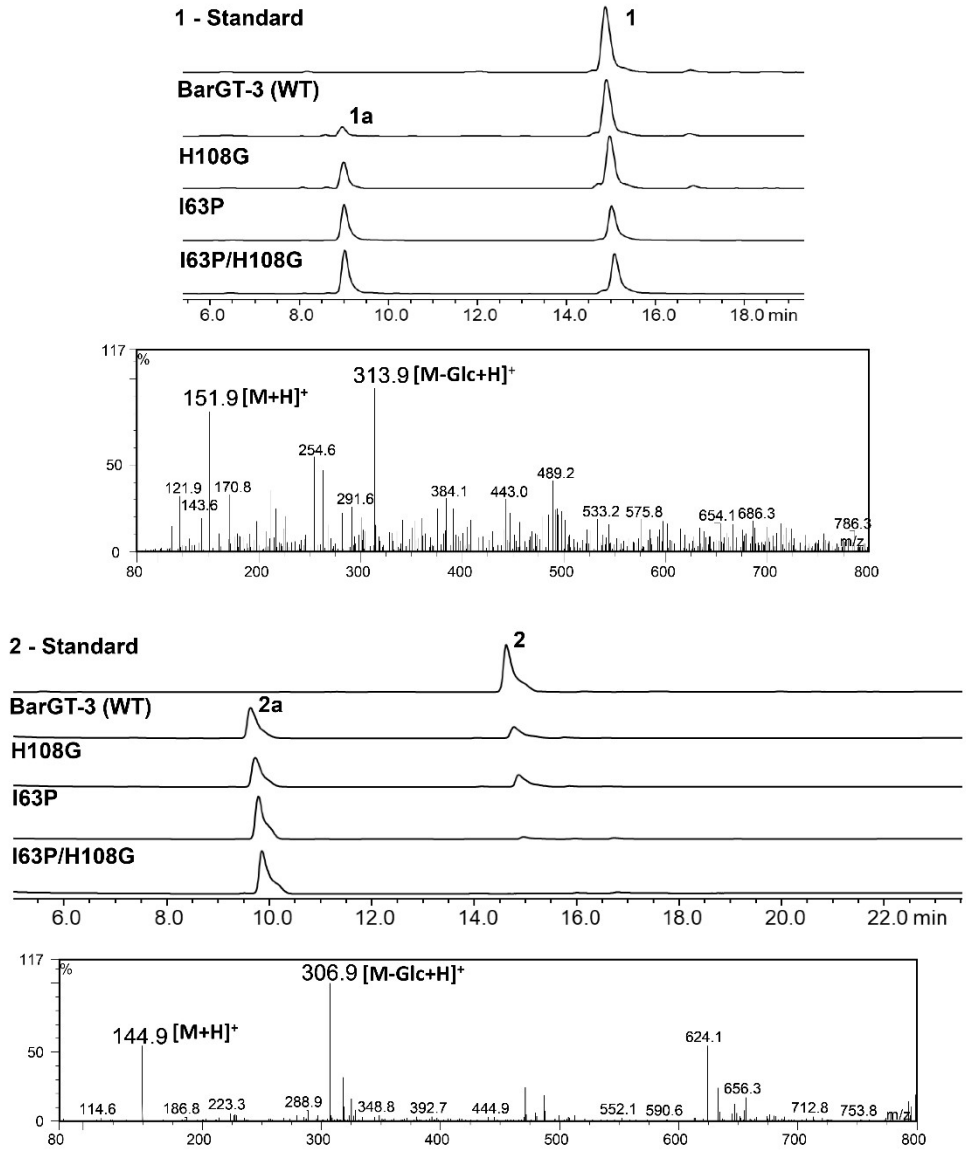


Figure S3. HPLC and positive ion mass spectrometry (MS) analysis for the products **1a** and **2a**. The activities toward substrates **1** and **2** ($[M+H]^+$) of BarGT-3 and its variants were determined by HPLC, and the products **1a** and **2a** ($[M-Glc+H]^+$) were identified by positive ion MS.

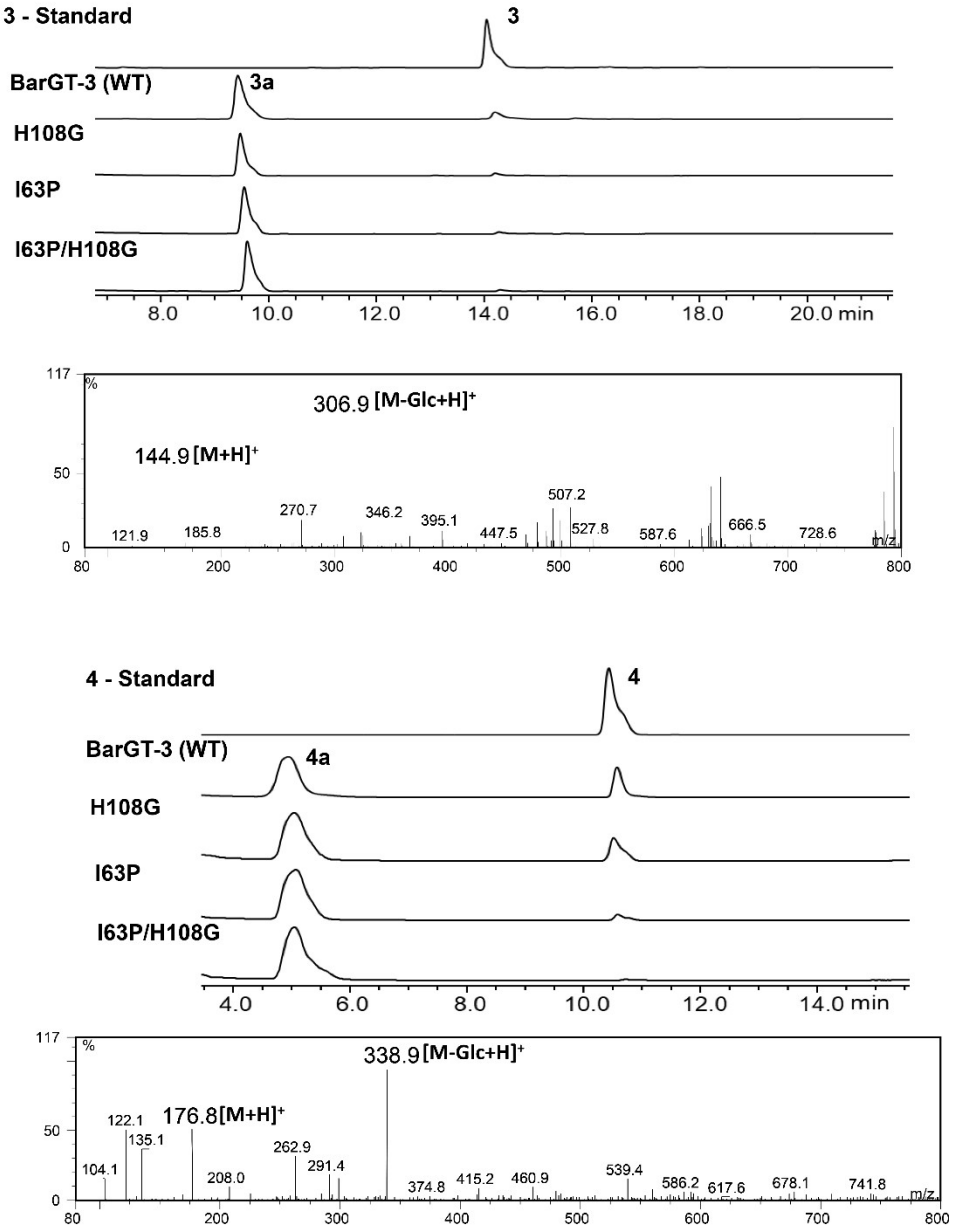


Figure S4. HPLC and positive ion mass spectrometry (MS) analysis for the products **3a** and **4a**. The activities toward substrates **3** and **4** ($[M+H]^+$) of BarGT-3 and its variants were determined by HPLC, and the products **3a** and **4a** ($[M-Glc+H]^+$) were identified by positive ion MS.

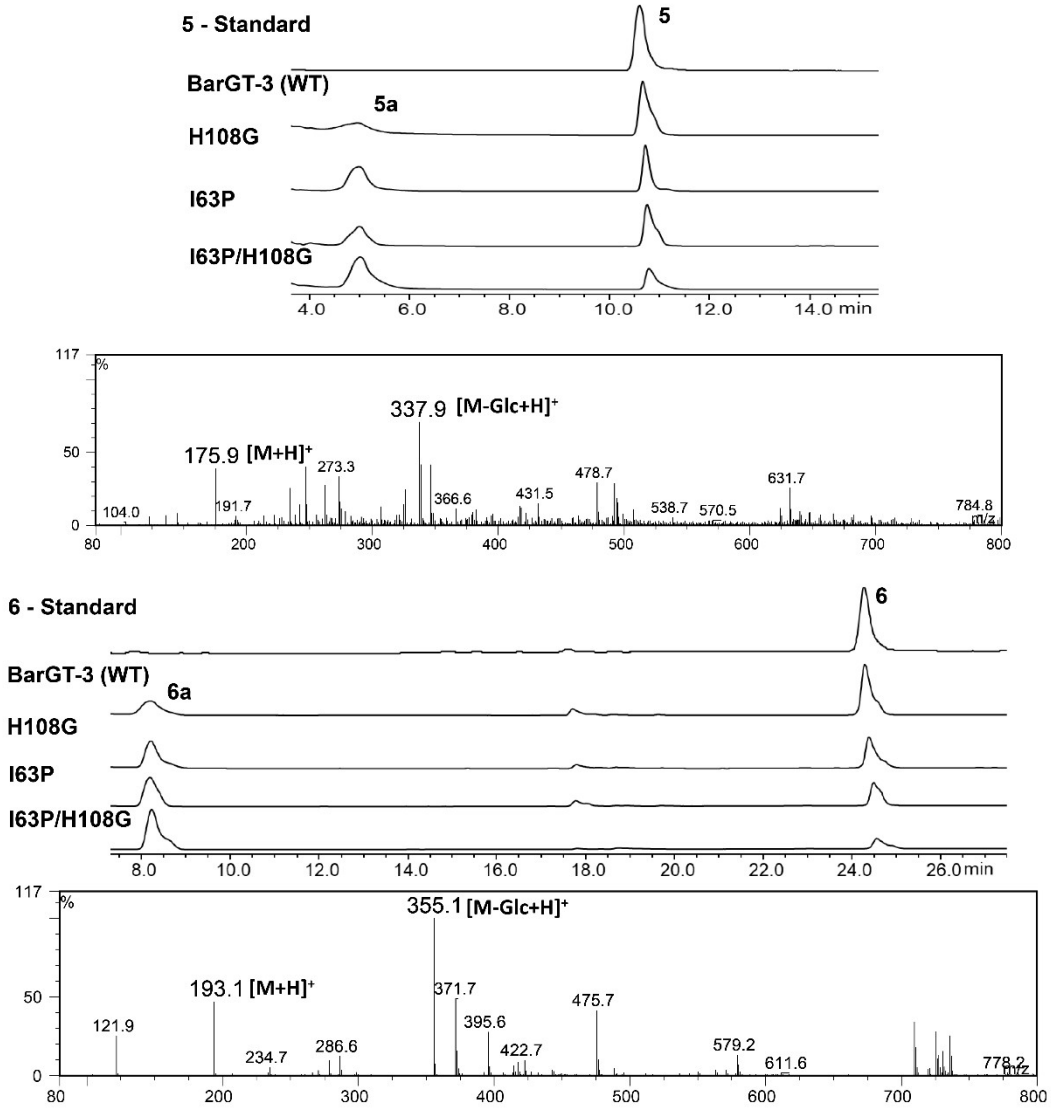


Figure S5. HPLC and positive ion mass spectrometry (MS) analysis for the products **5a** and **6a**. The activities toward substrates **5** and **6** ($[M+H]^+$) of BarGT-3 and its variants were determined by HPLC, and the products **5a** and **6a** ($[M-Glc+H]^+$) were identified by positive ion MS.

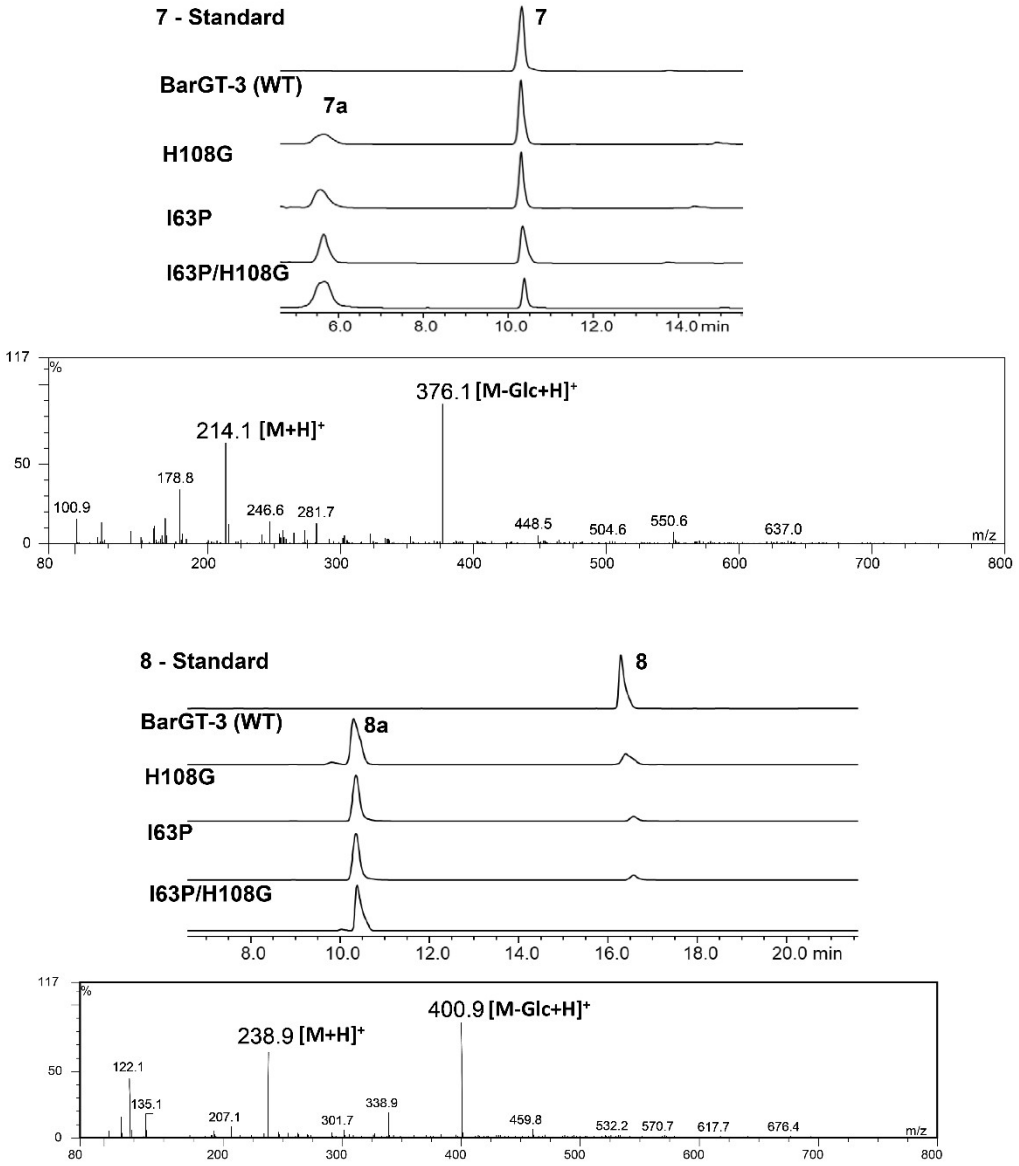


Figure S6. HPLC and positive ion mass spectrometry (MS) analysis for the products **7a** and **8a**. The activities toward substrates **7** and **8** ($[M+H]^+$) of BarGT-3 and its variants were determined by HPLC, and the products **7a** and **8a** ($[M-Glc+H]^+$) were identified by positive ion MS.

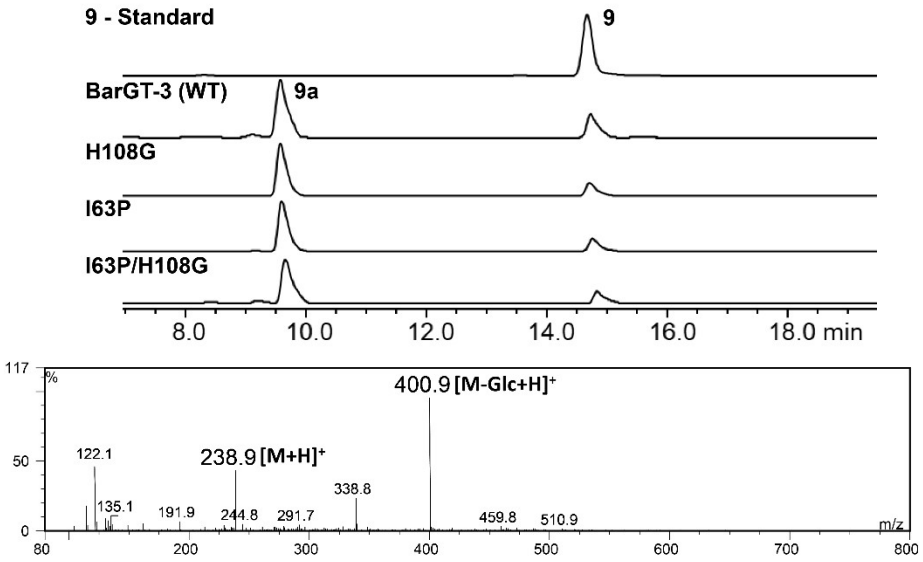


Figure S7. HPLC and positive ion mass spectrometry (MS) analysis for the product **9a**. The activities toward substrate **9** ($[M+H]^+$) of BarGT-3 and its variants were determined by HPLC, and the product **9a** ($[M-Glc+H]^+$) were identified by positive ion MS.

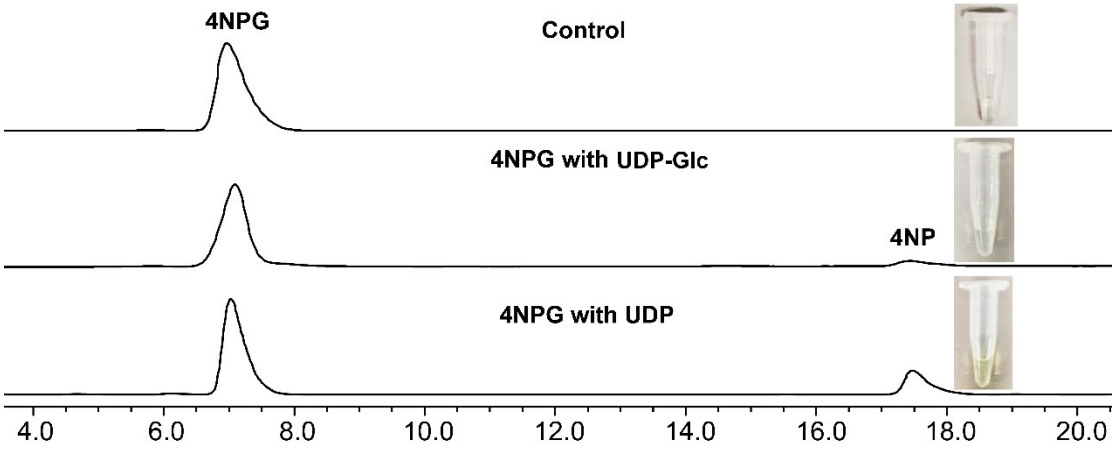


Figure S8. The verification of the catalytic reversibility of BarGT-3. The reverse reactions from 4NPG to UDP or UDPG catalyzed by BarGT-3 (WT), 100 μL Tris-HCl buffer (50 mM, pH 7.5, 10 mM $MgCl_2$) with 500 μM of 4-nitrophenyl-β-D-glucoside (4NPG), 100 μM of UDP or UDPG.

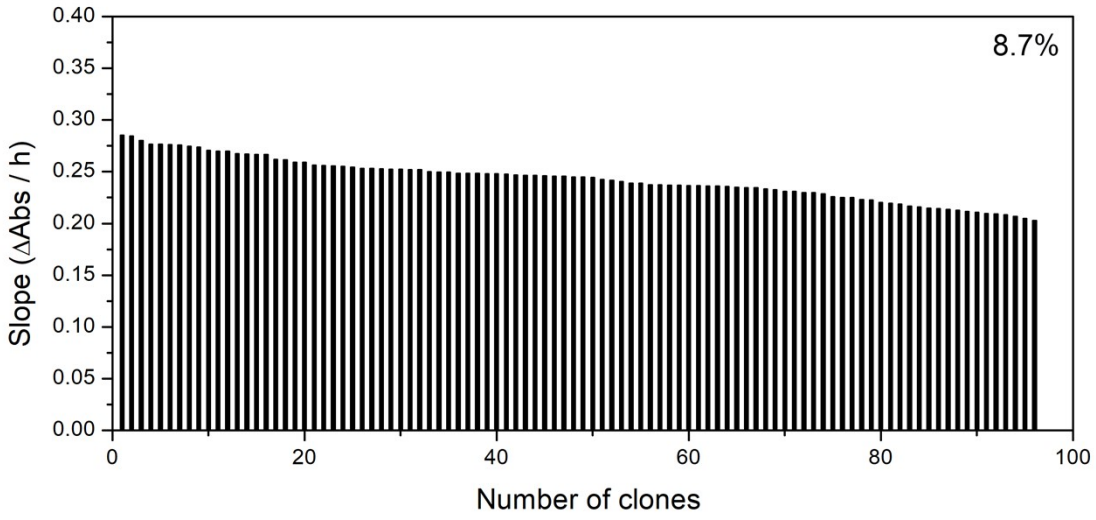


Figure S9. Optimization of coefficient of variation (CV) in the UDPG-recycling high-throughput system. CV is evaluated by a 96-well plate of BarGT-3 (WT). The lowest CV value is 8.7% in the condition of 20 μL lysate and pH 7.5.

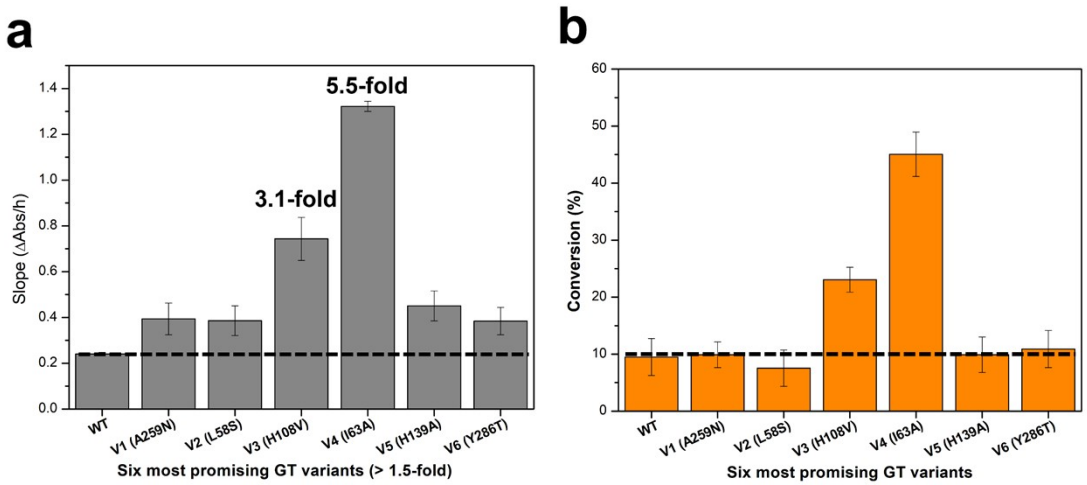


Figure S10. Six most promising variants screening from the cepPCR library. (a). 6 most promising GT variants (> 1.5-fold) were screened in high-throughput screening (HTS) system. (b). The conversions of lysates of 6 promising variants toward acetaminophen (substrate **1**).

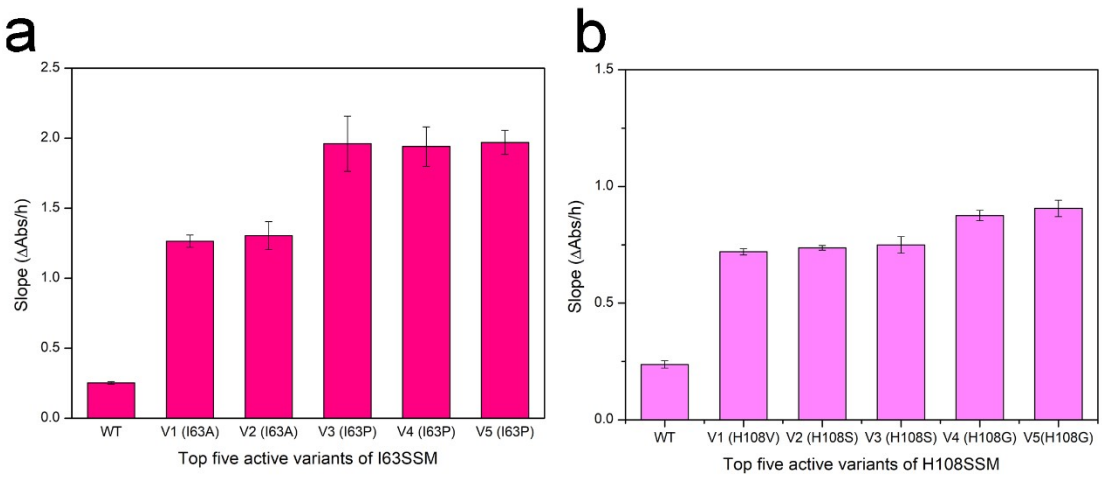


Figure S11. Top five active variants of site-saturation mutagenesis (SSM) libraries of I63 and H108. The best single variants for substrate **1** catalysis are proven to be I63P and H108G.

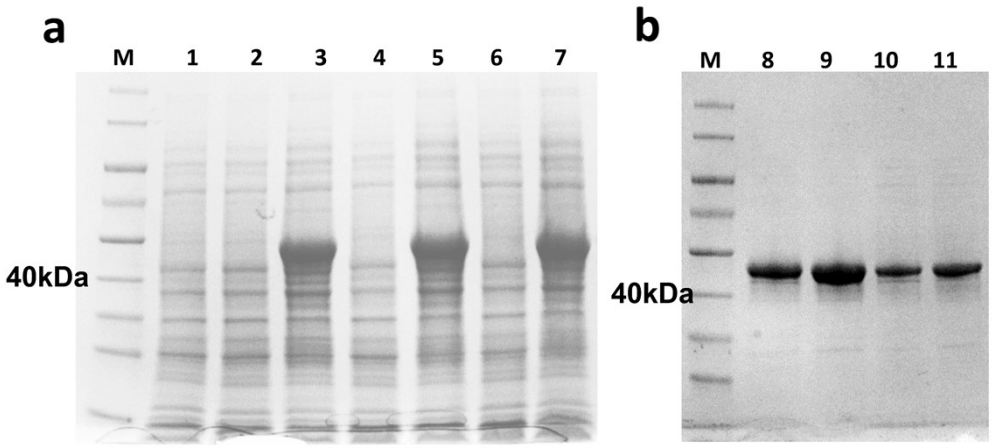


Figure S12. SDS-PAGE detection of the expression of variants I63P, H108G and I63P/H108G. (a). M is the protein molecular weight marker, Lane 1, the control of pET 28a vector; Lanes 2, 4 and 6, the *E. coli* lysate before induction of I63P, H108G and I63P/H108G; Lanes 3, 5 and 7, the soluble supernatants after induction of I63P, H108G and I63P/H108G. (b). M is the protein molecular weight marker, lanes 8-11, the purified proteins of BarGT-3, I63P, H108G and I63P/H108G.

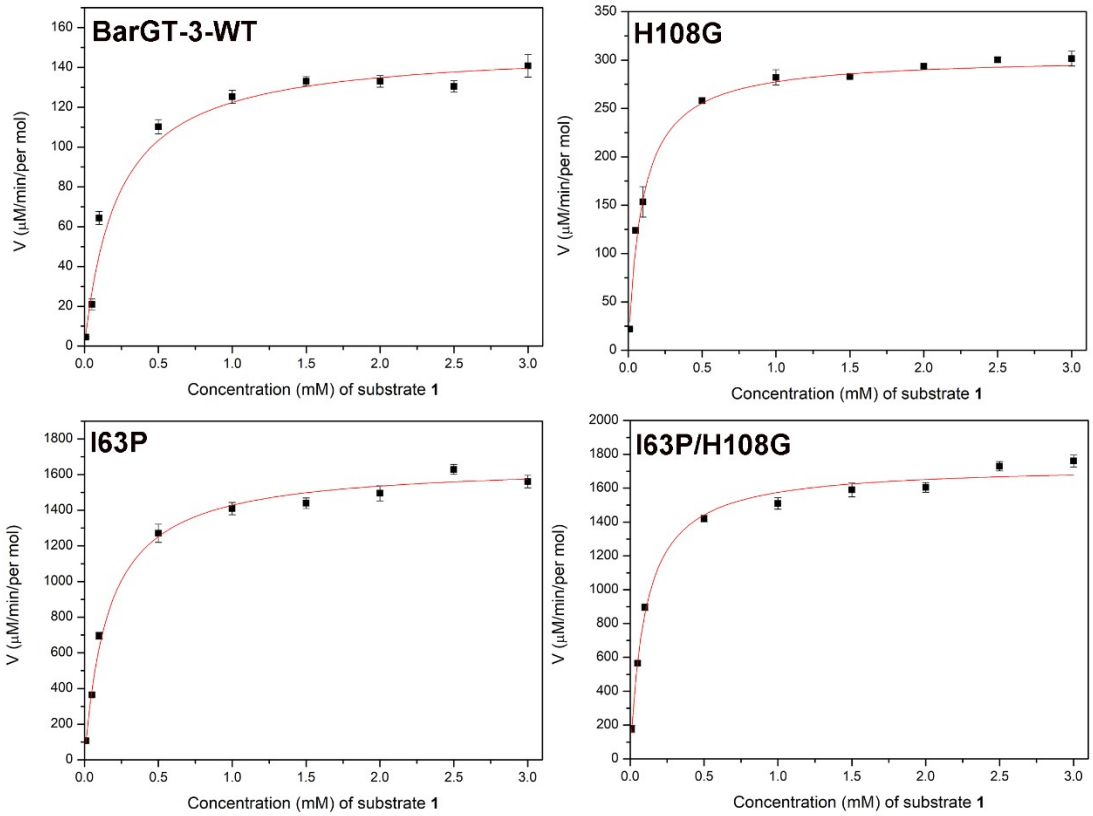


Figure S13. Kinetic parameters and fitting curves analysis of BarGT-3 (WT) and its variants. Determination of kinetic parameters for substrate **1** with saturated UDPG (4 mM). Enzyme assays were performed in 50 mM Tris- HCl buffer (pH 7.5) containing 20 μg/mL protein and 10 mM $MgCl_2$ at 37°C for 10 min in triplicate.

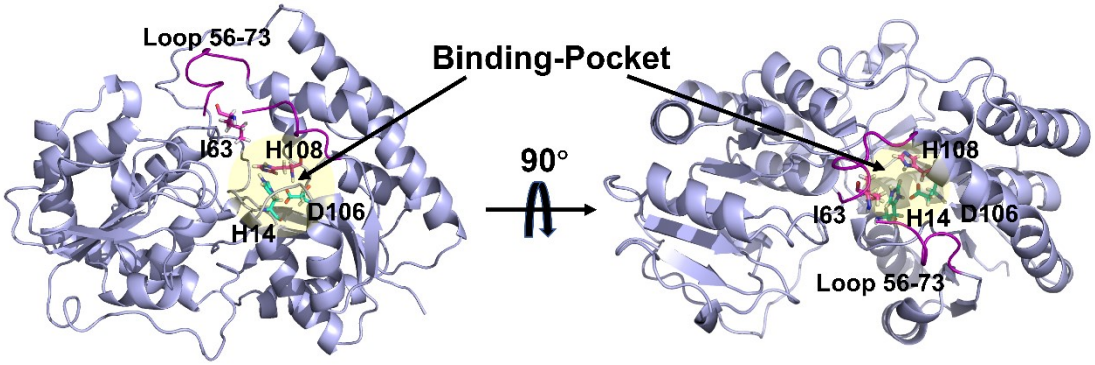


Figure S14. The positions of two identified residues I63 and H108 in the structural model of BarGT-3. The structure of BarGT-3 was constructed by AlphaFold2. The residue I63 was on a loop region (residue 56-73 at the N-domain, purple) above the substrate-binding pocket (yellow region). The identified residue H108 was located at the binding pocket of BarGT-3.

BarGT-3_WT **I63P/H108G**

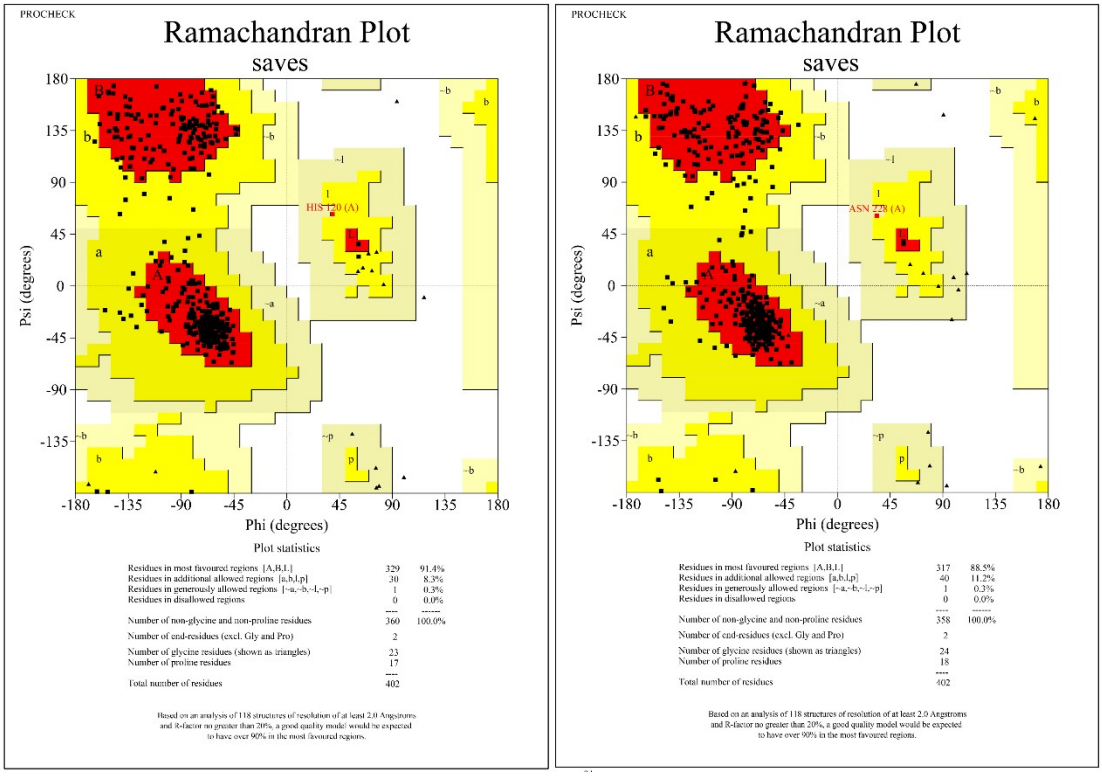


Figure S15. The Ramachandran plot assessment. The Ramachandran plot was used for evaluating model qualities of BarGT-3 (WT) and its double-variant (I63P/H108G) constructed by AlphaFold2. No residue of WT and I63P/H108G was in the disallowed region. The parameters of WT and I63P/H108G were marked below.

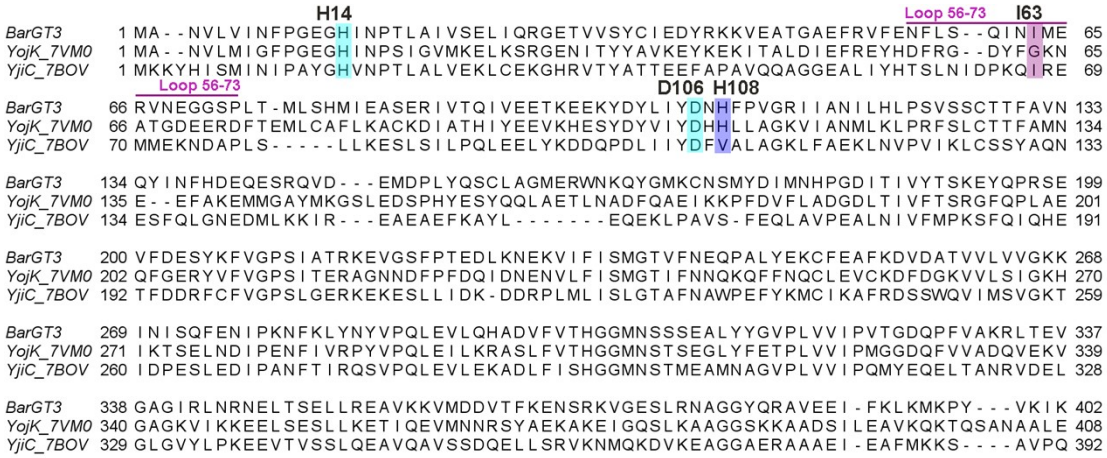


Figure S16. Multiple sequence alignment (MSA) analysis for BarGT-3. BarGT-3 was conducted MSA with the typical crystal structures of enzymes in GT1 family (YjiC PDB ID: 7BOV[14] and YojK PDB ID: 7VM0[15]).

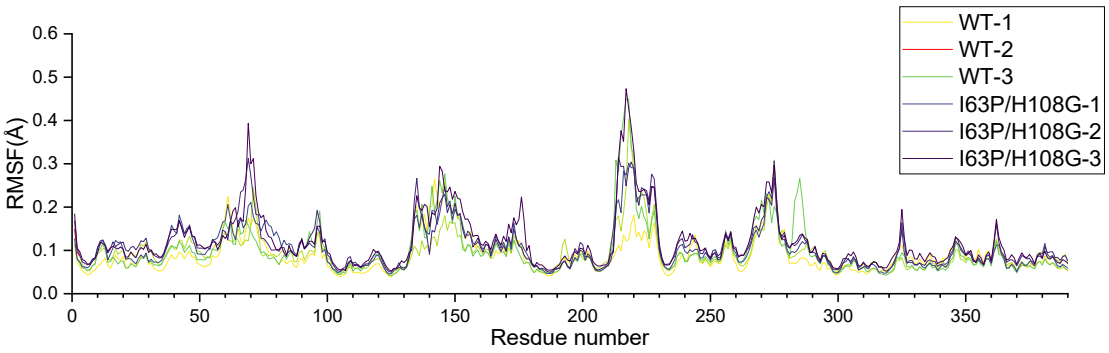


Figure S17. The comparison of RMSF values of WT and the double-variant I63P/H108G. Three parallel sets of MD simulations for WT and the double-variant I63P/H108G were performed with 200 ns, respectively.

4. Figures S18-S44. NMR spectrum of glycosylated products 1a-9a.

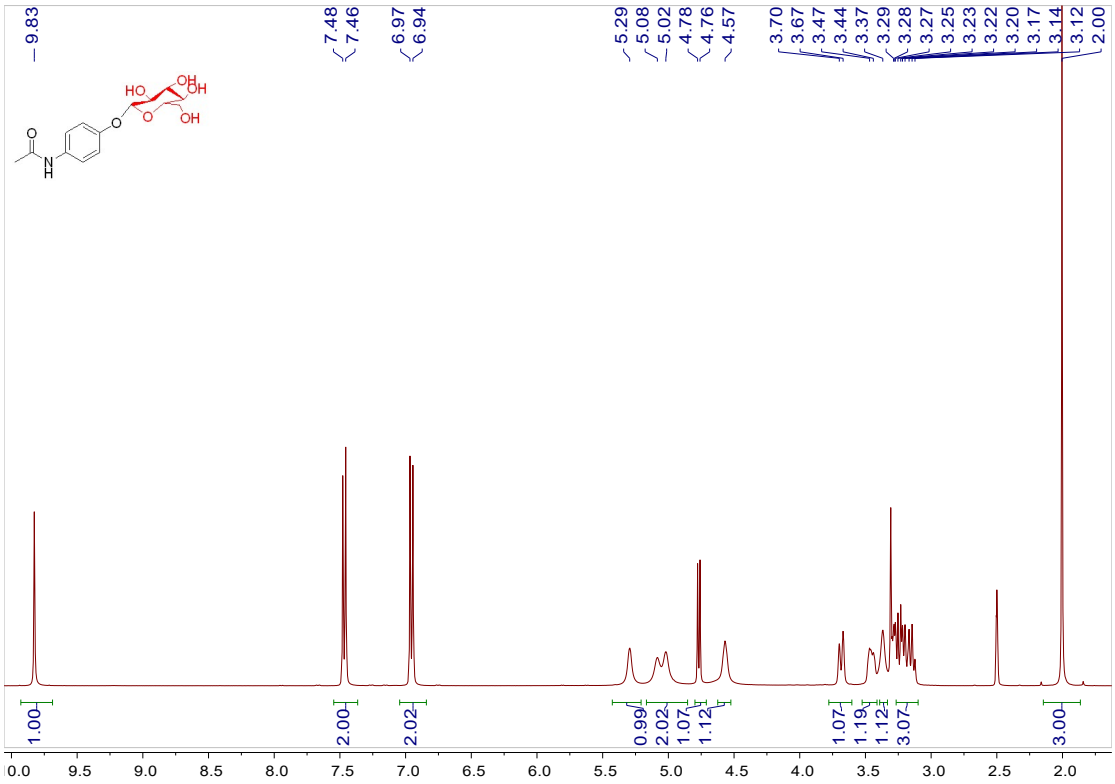


Figure S18. ^{1}H NMR spectrum of compound **1a** (400 MHz, DMSO-d6)

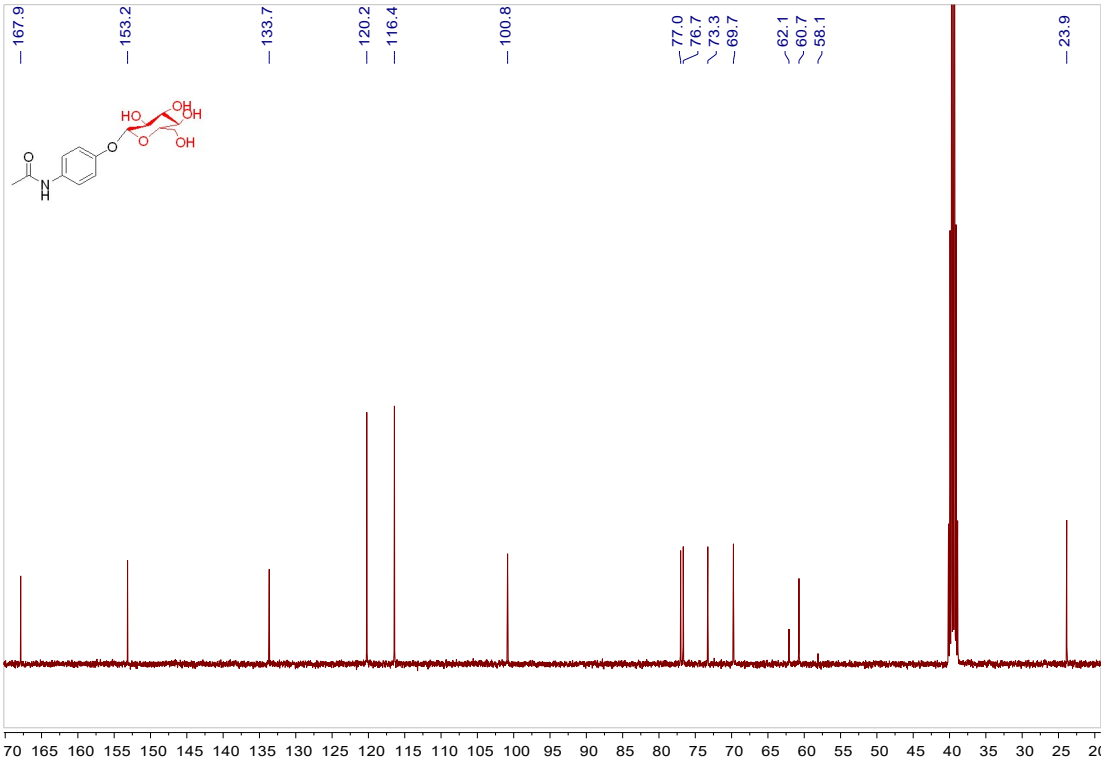


Figure S19. ^{13}C NMR spectrum of compound **1a** (101 MHz, DMSO-d6)

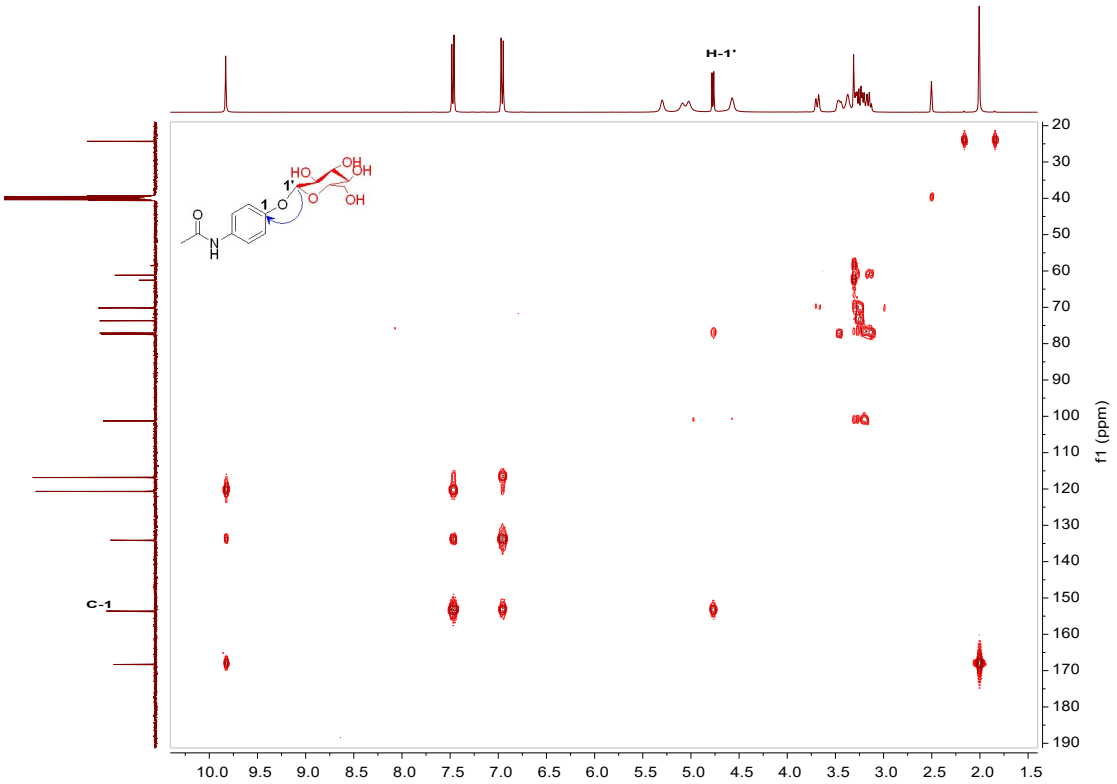


Figure S20. HMBC spectrum of compound **1a** (400 MHz/101 MHz, DMSO-d6)

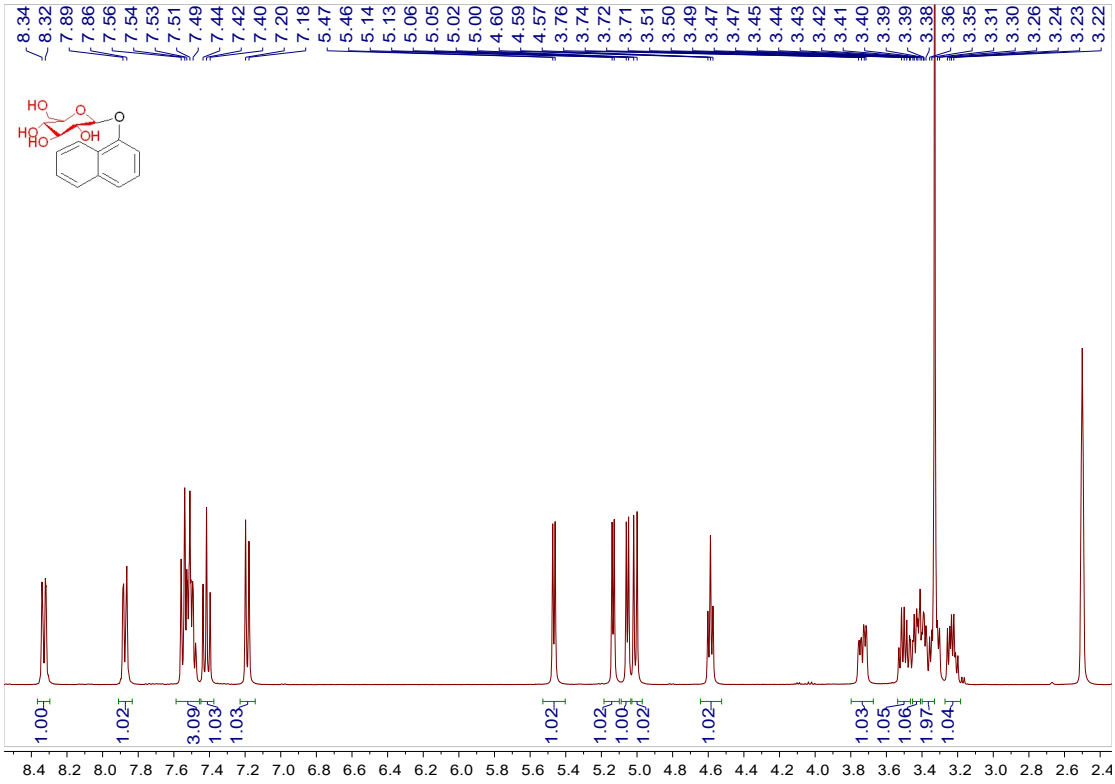


Figure S21. ^{1}H NMR spectrum of compound **2a** (400 MHz, DMSO-d6)

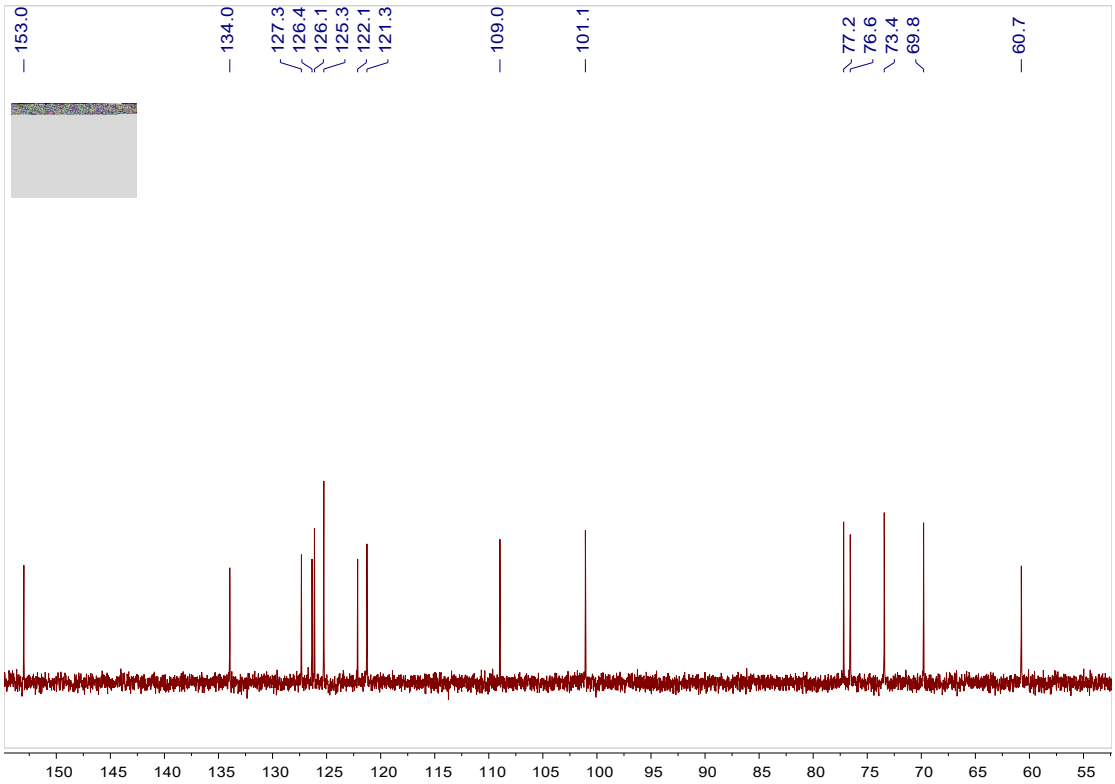


Figure S22. ^{13}C NMR spectrum of compound **2a** (101 MHz, DMSO-d6)

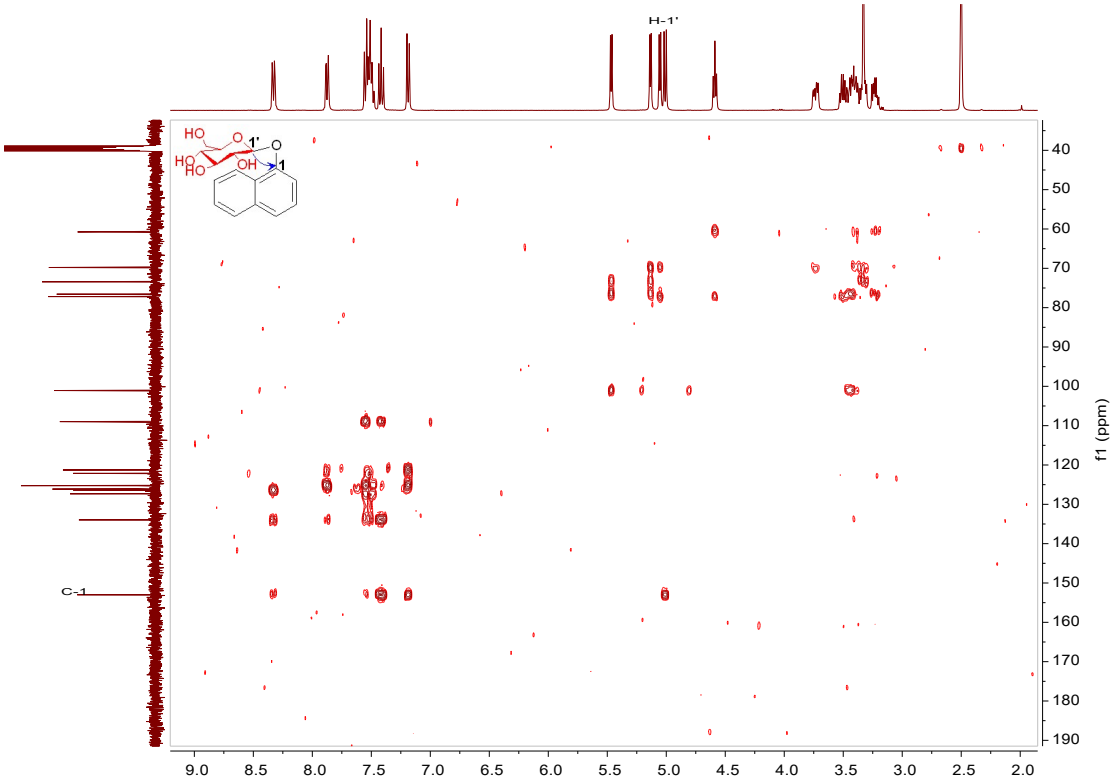


Figure S23. HMBC spectrum of compound **2a** (400 MHz/101 MHz, DMSO-d6)

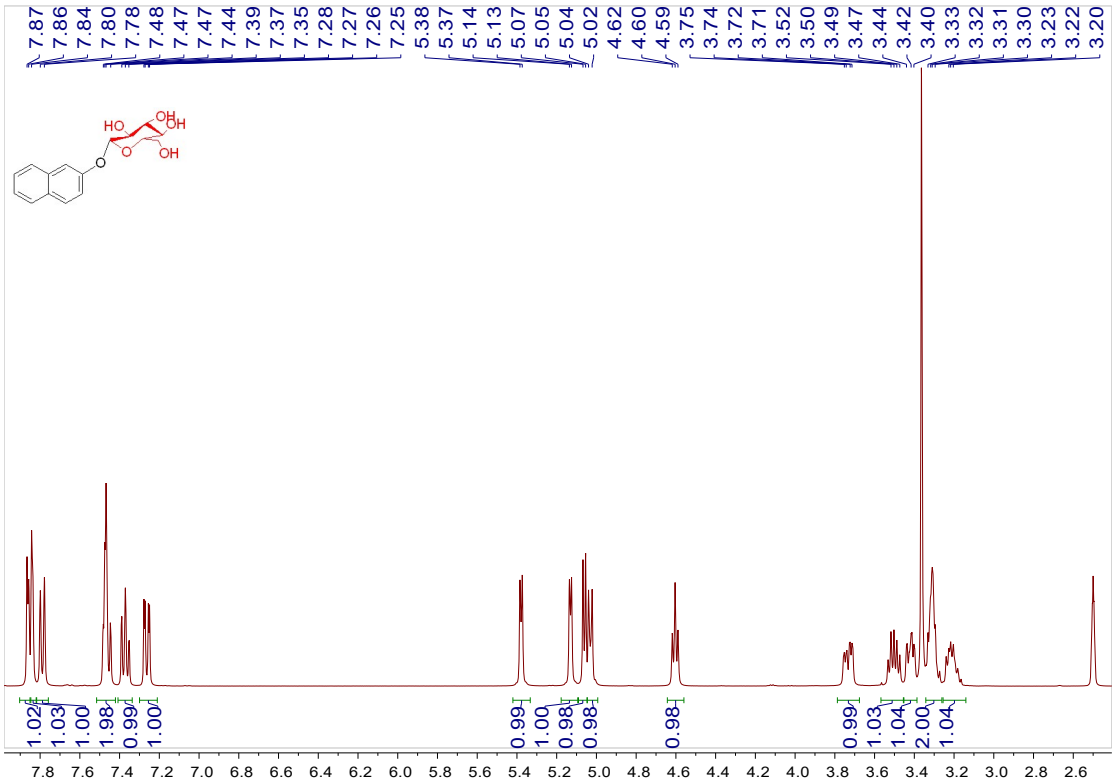


Figure S24. ^{1}H NMR spectrum of compound **3a** (400 MHz, DMSO-d6)

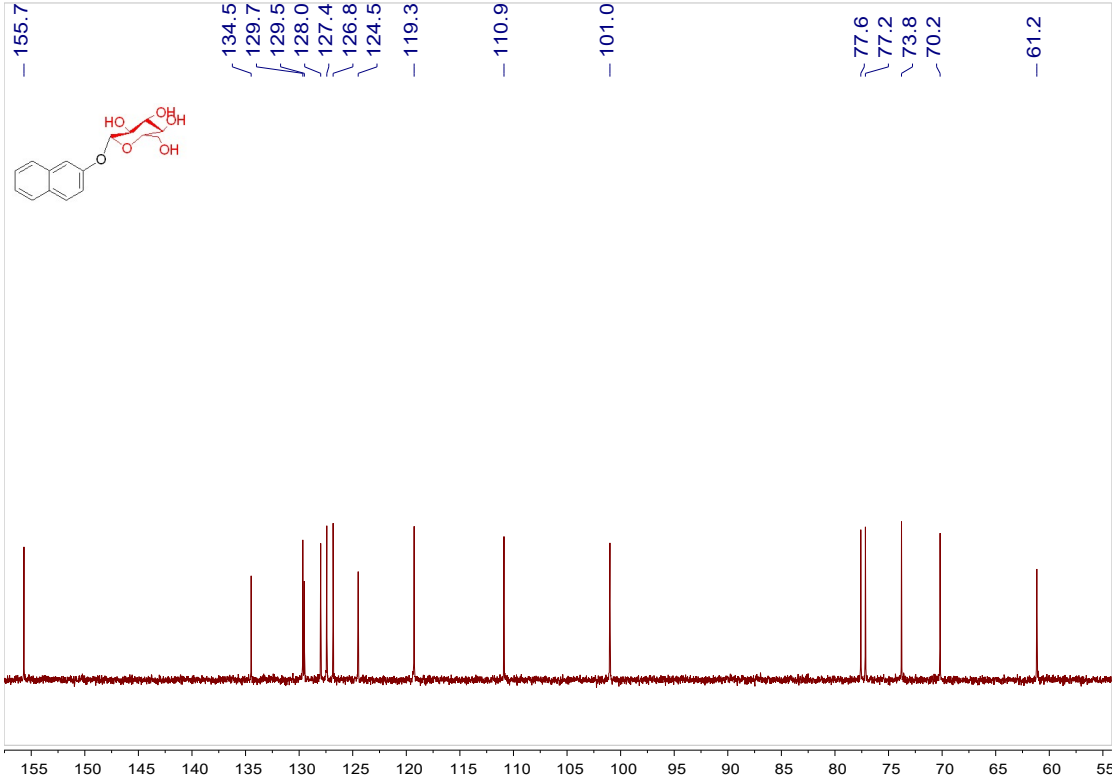


Figure S25. ^{13}C NMR spectrum of compound **3a** (101 MHz, DMSO-d6)

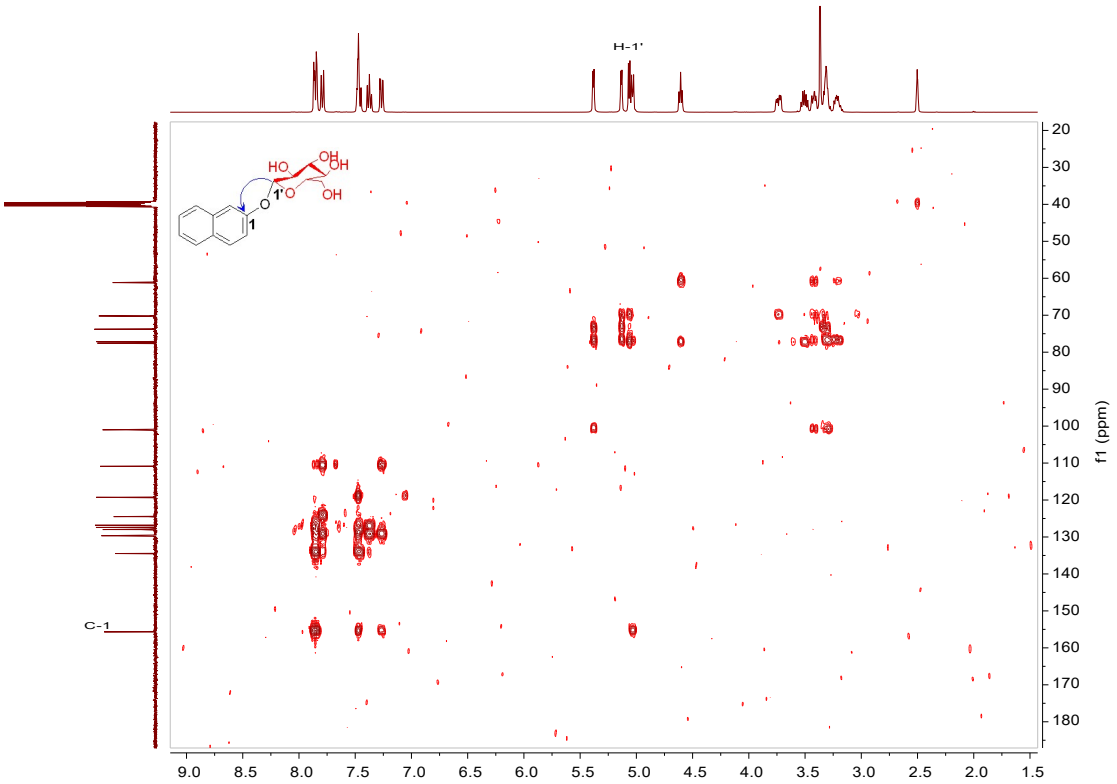


Figure S26. HMBC spectrum of compound **3a** (400 MHz/101 MHz, DMSO-d6)

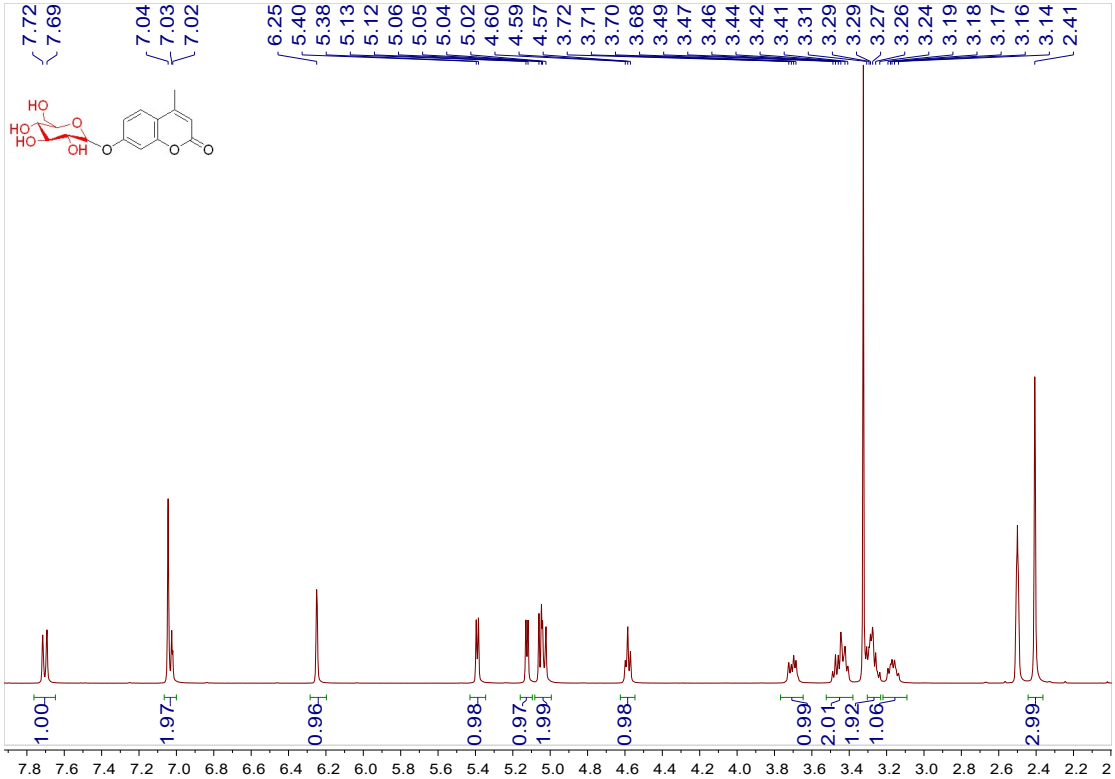


Figure S27. ^{1}H NMR spectrum of compound **4a** (400 MHz, DMSO-d6)

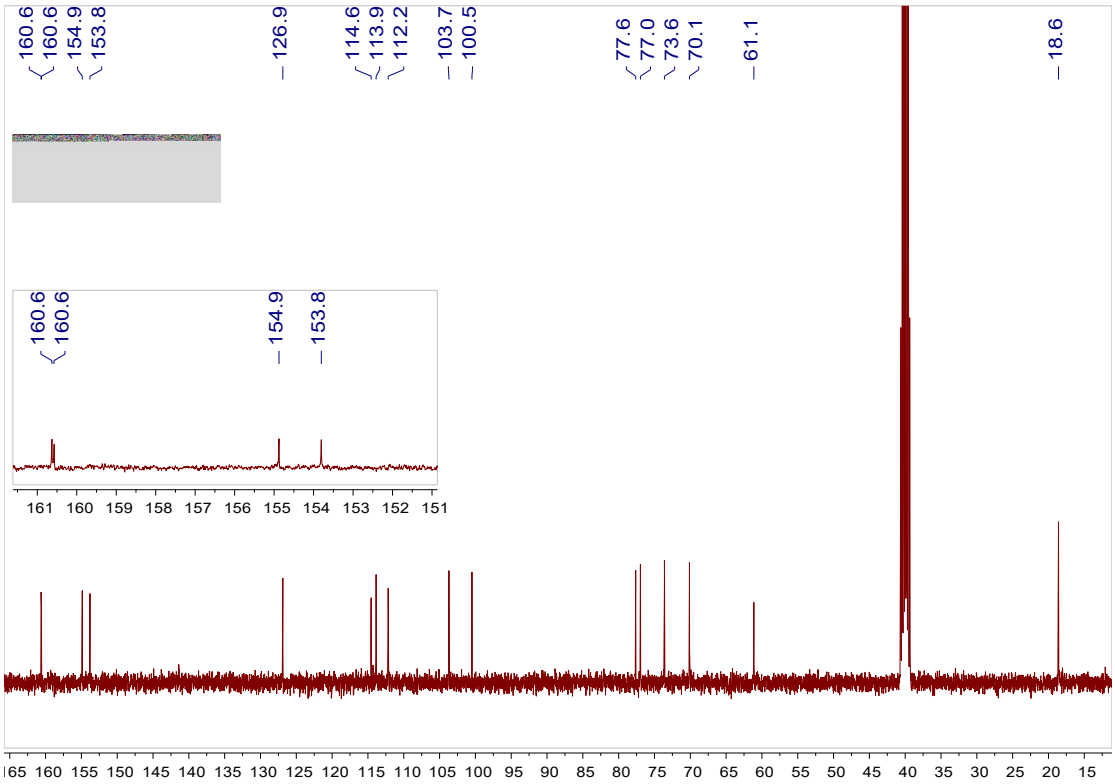


Figure S28. ^{13}C NMR spectrum of compound **4a** (101 MHz, DMSO-d6)

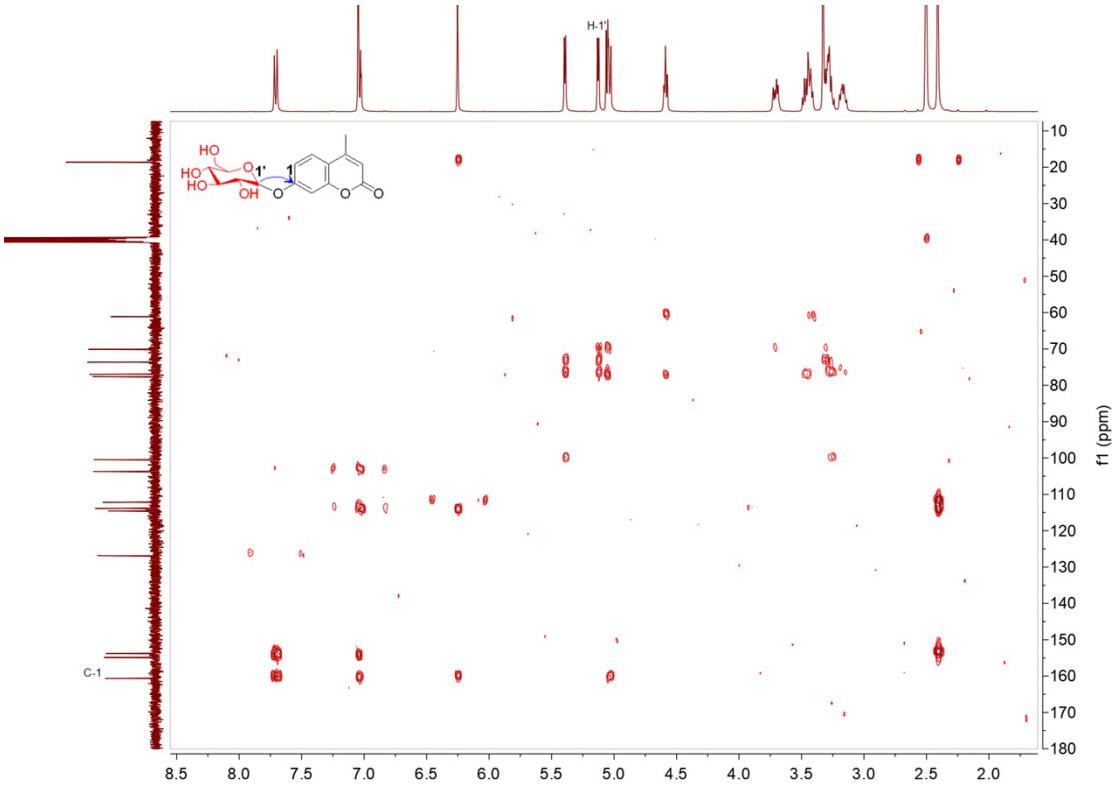


Figure S29. HMBC spectrum of compound **4a** (400 MHz/101 MHz, DMSO-d6)

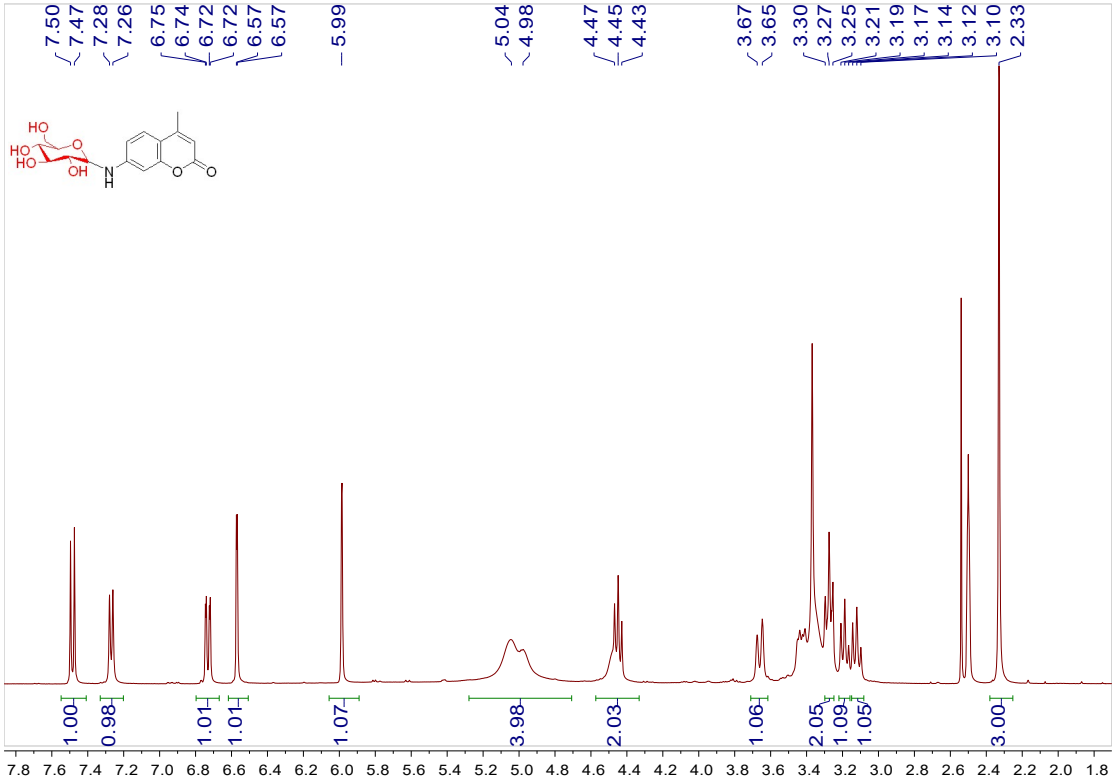


Figure S30. ^{1}H NMR spectrum of compound **5a** (400 MHz, DMSO-d6)

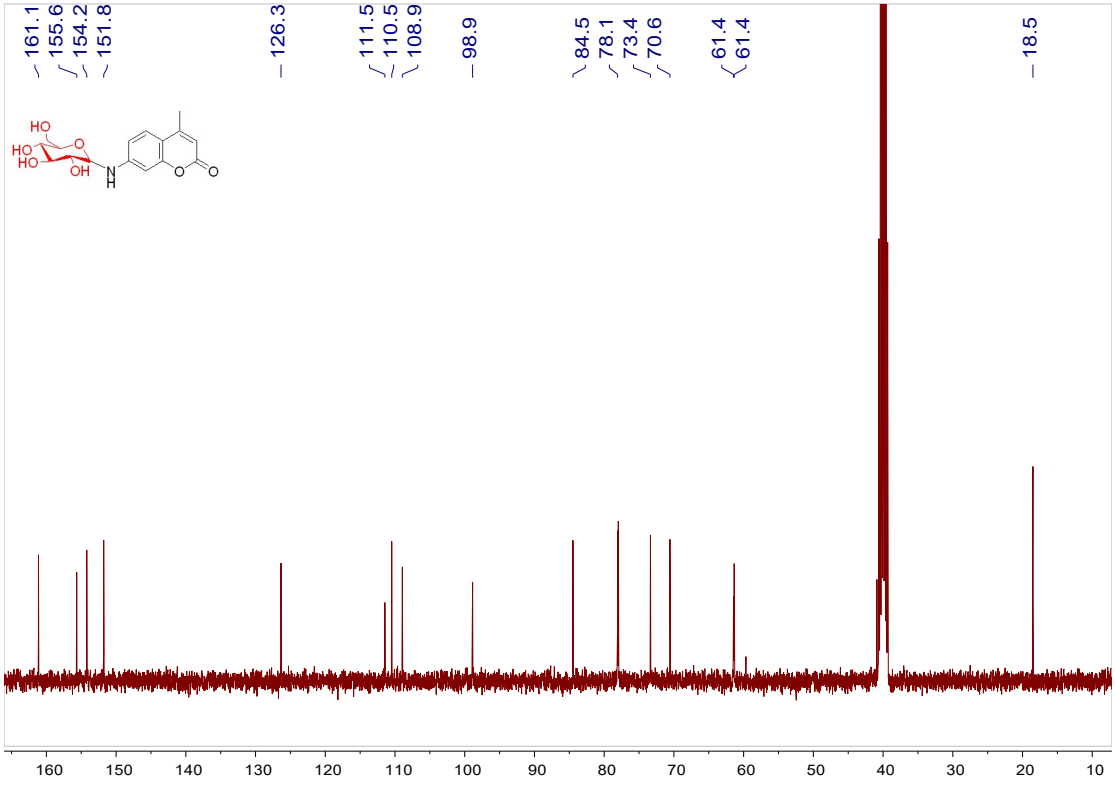


Figure S31. ^{13}C NMR spectrum of compound **5a** (101 MHz, DMSO-d6)

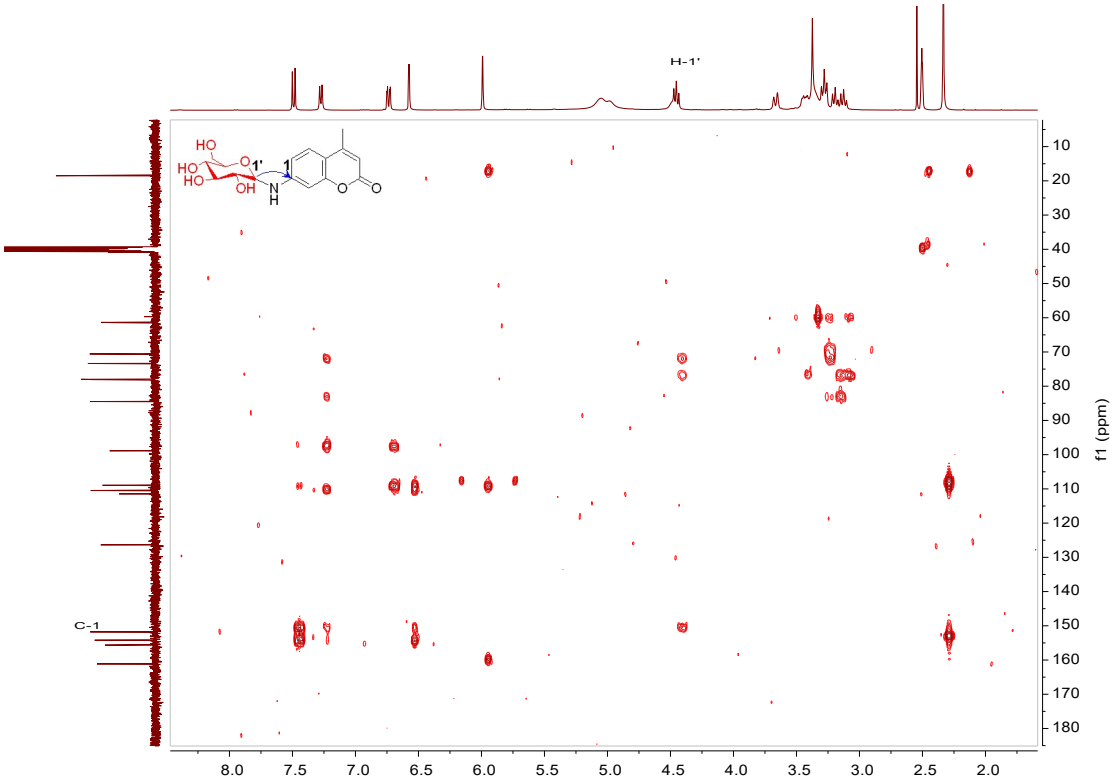


Figure S32. HMBC spectrum of compound **5a** (400 MHz/101 MHz, DMSO-d6)

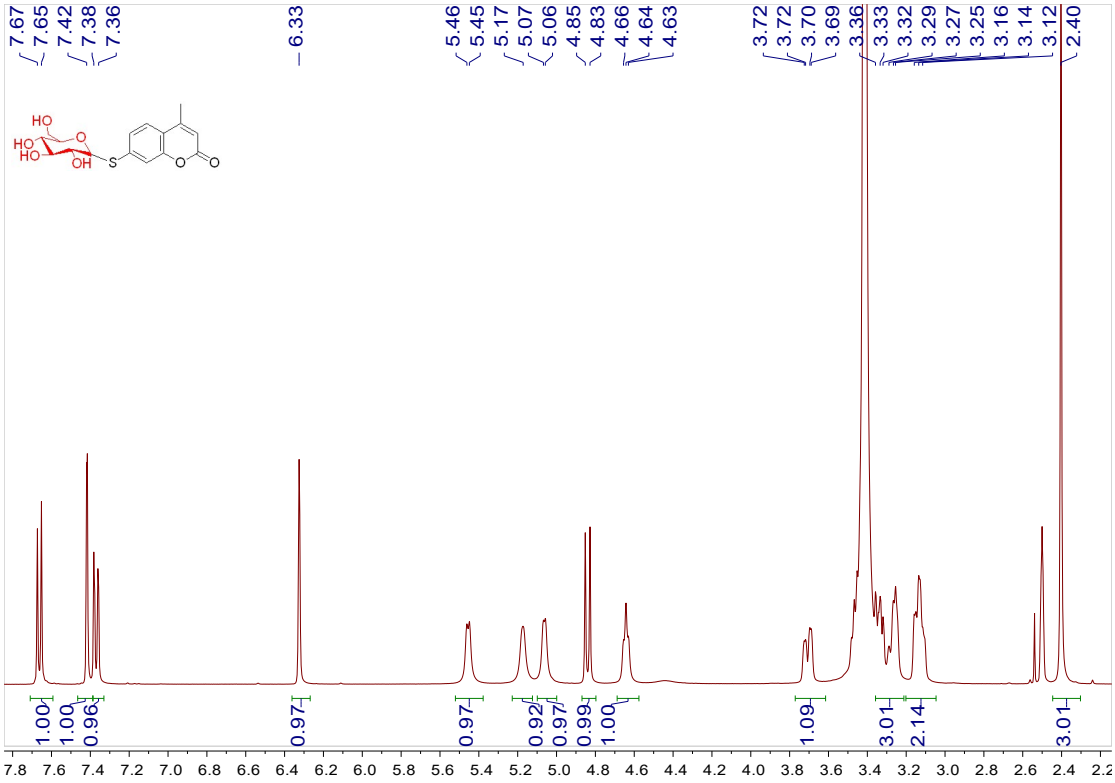


Figure S33. ^{1}H NMR spectrum of compound **6a** (400 MHz, DMSO-d6)

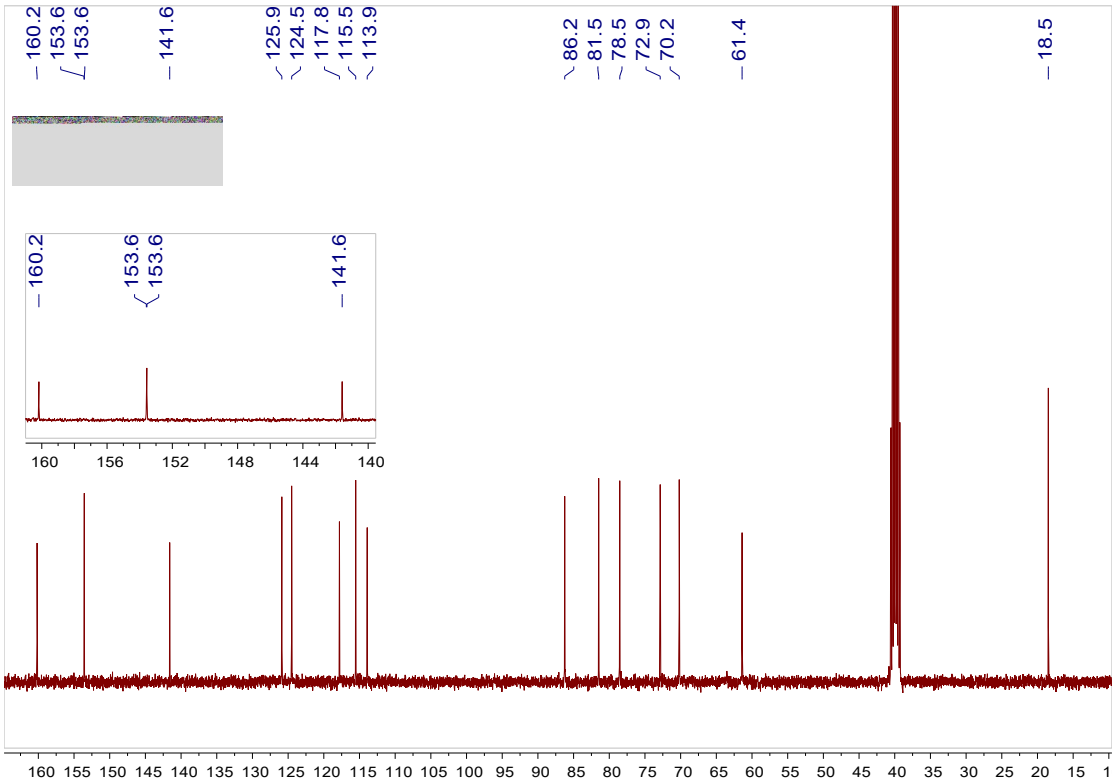


Figure S34. ^{13}C NMR spectrum of compound **6a** (101 MHz, DMSO-d6)

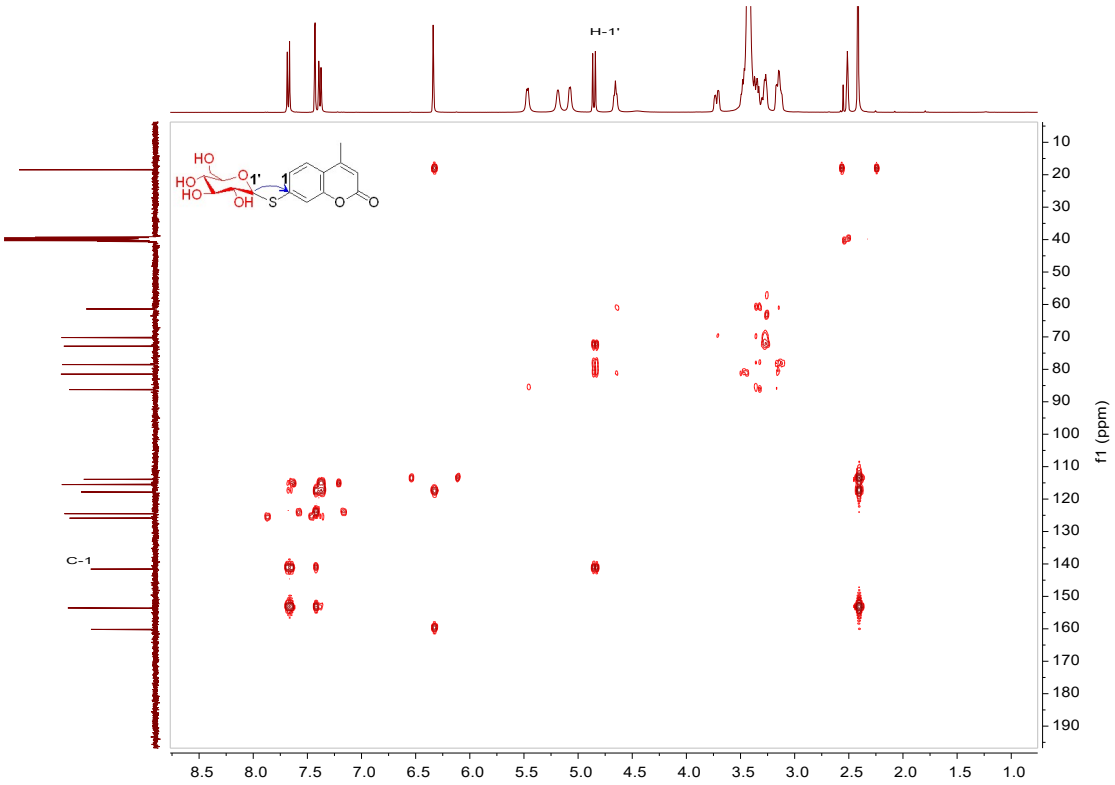


Figure S35. HMBC spectrum of compound **6a** (400 MHz/101 MHz, DMSO-d6)

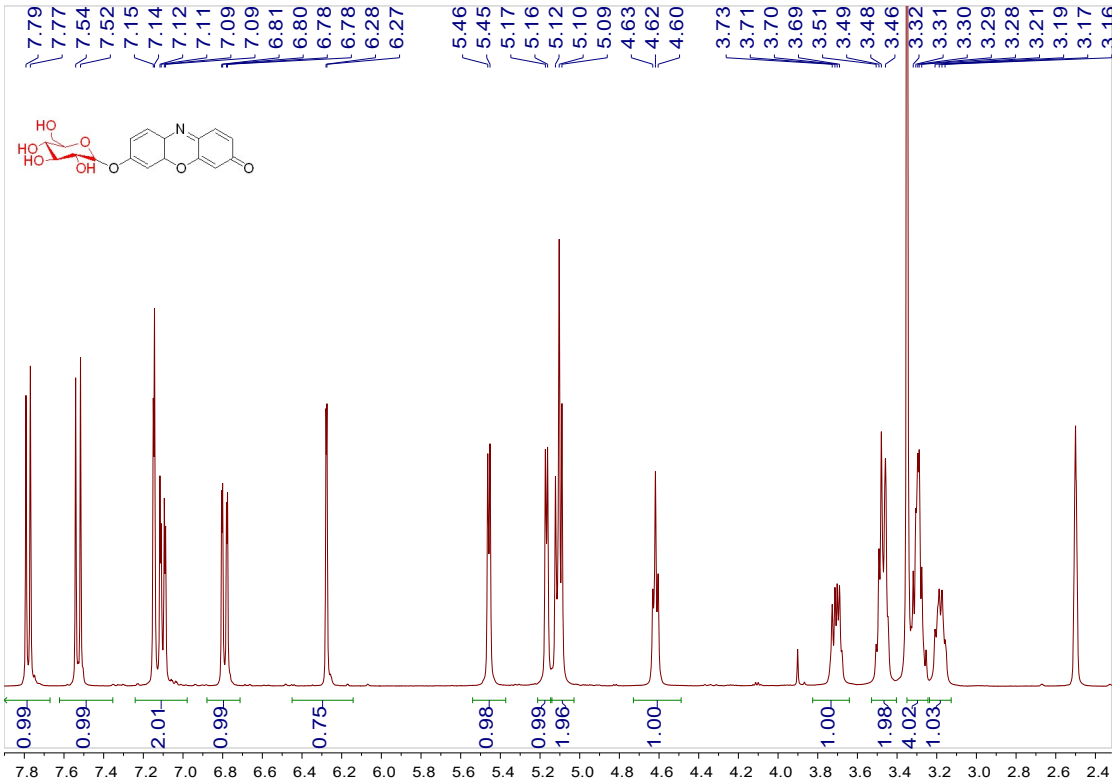


Figure S36. ^{1}H NMR spectrum of compound **7a** (400 MHz, DMSO-d6)

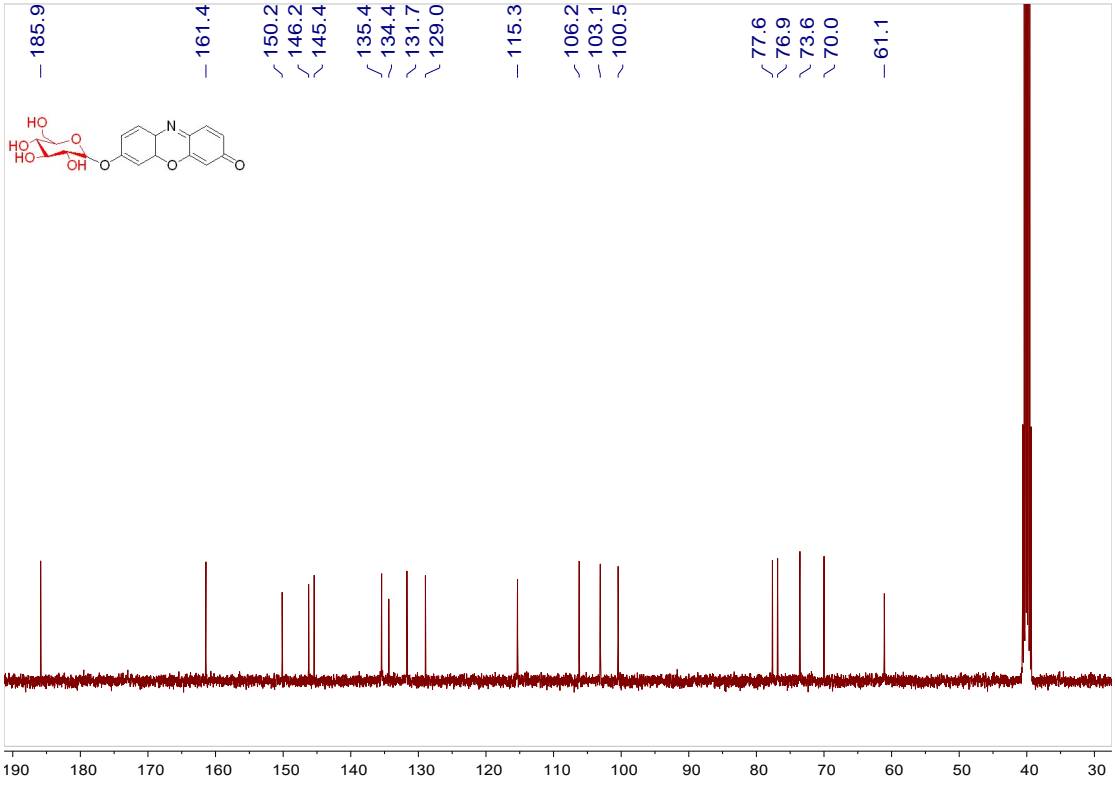


Figure S37. ^{13}C NMR spectrum of compound **7a** (101 MHz, DMSO-d6)

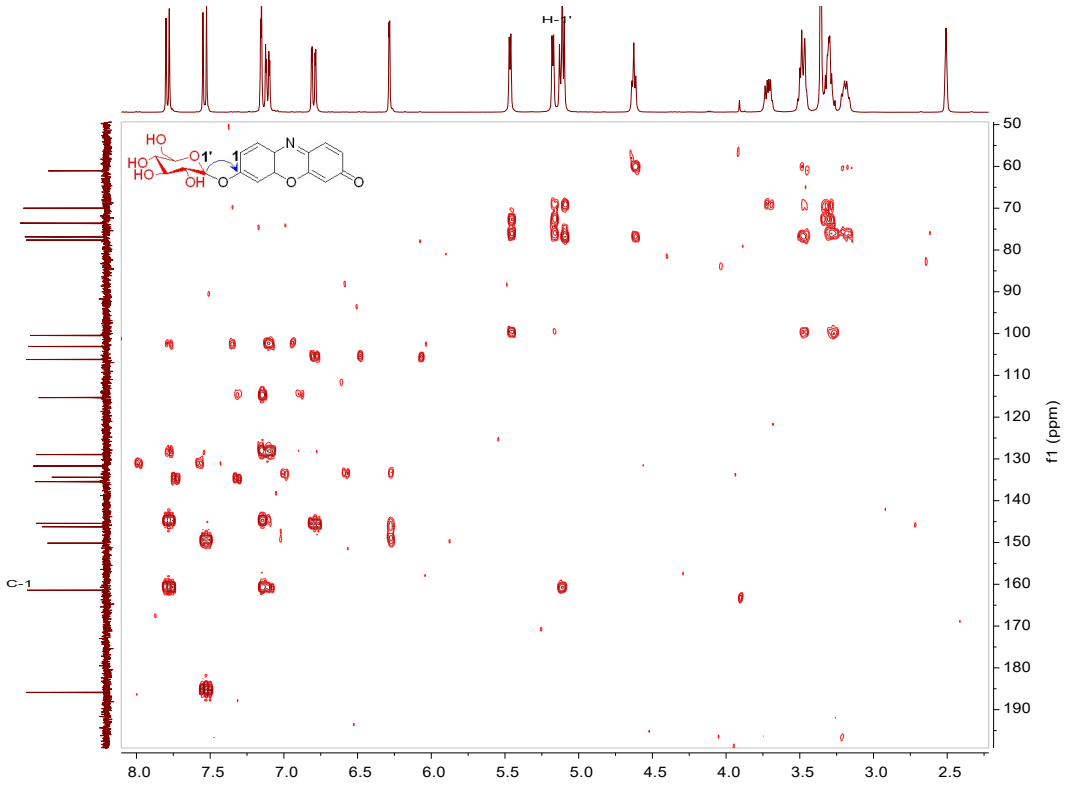


Figure S38. HMBC spectrum of compound **7a** (400 MHz/101 MHz, DMSO-d6)

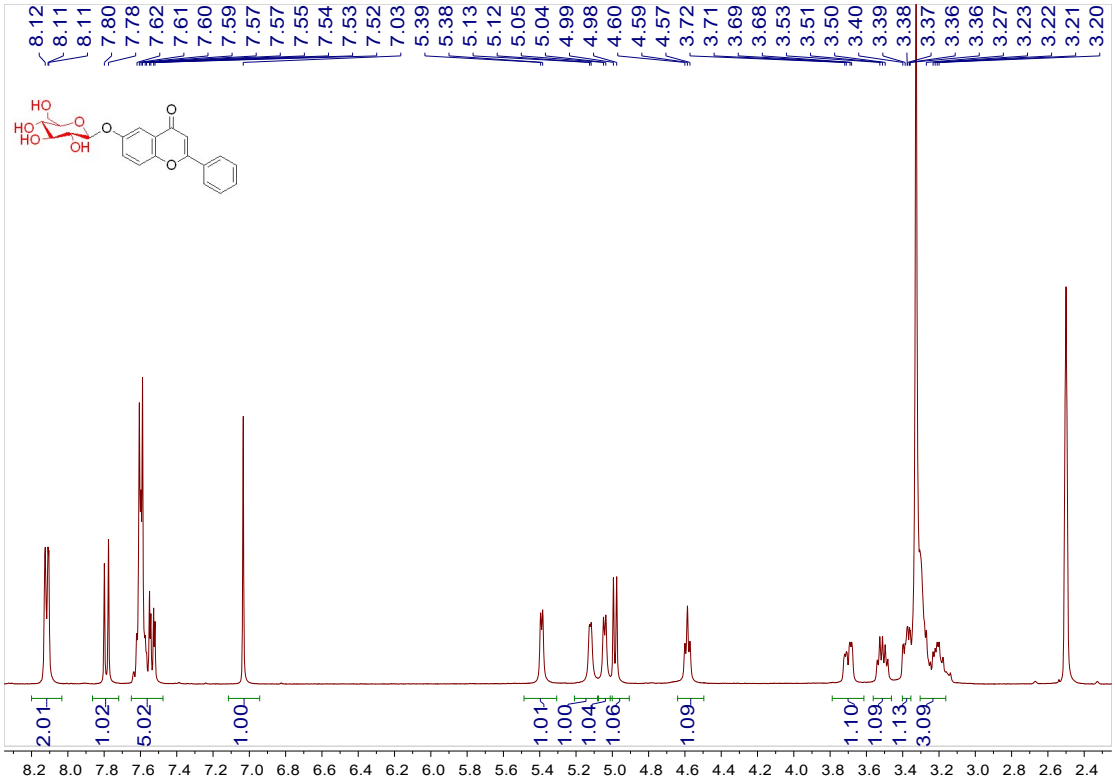


Figure S39. ^{1}H NMR spectrum of compound **8a** (400 MHz, DMSO-d6)

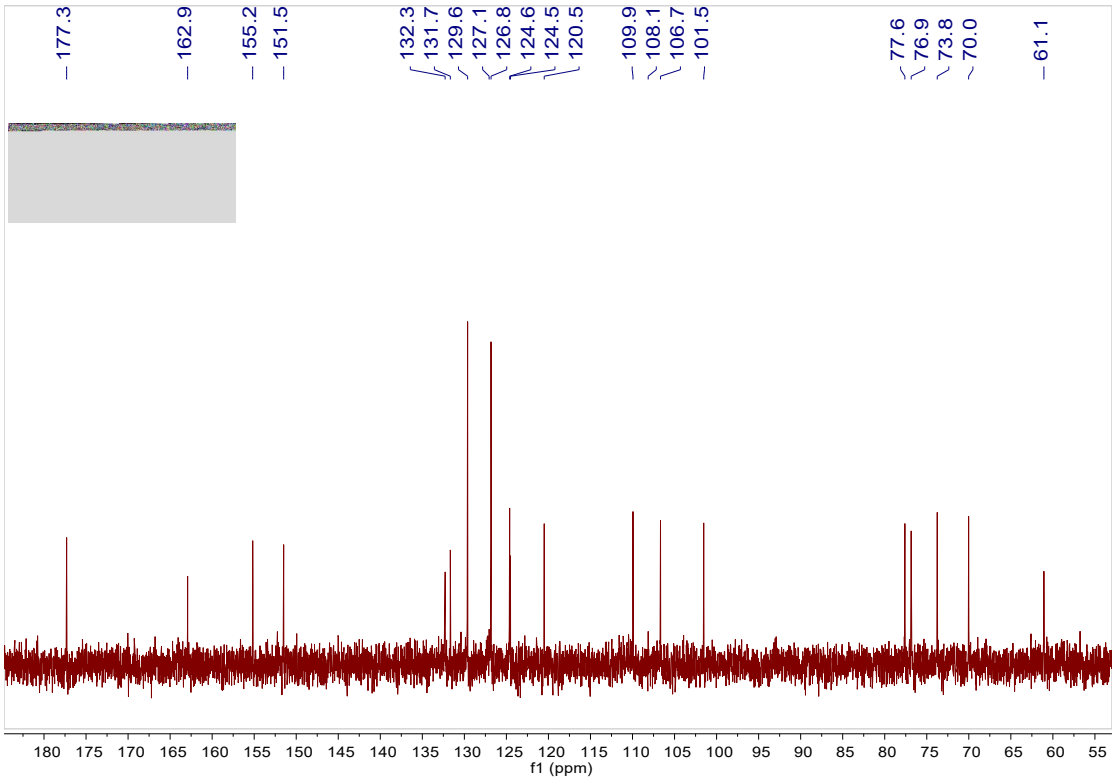


Figure S40. ^{13}C NMR spectrum of compound **8a** (101 MHz, DMSO-d6)

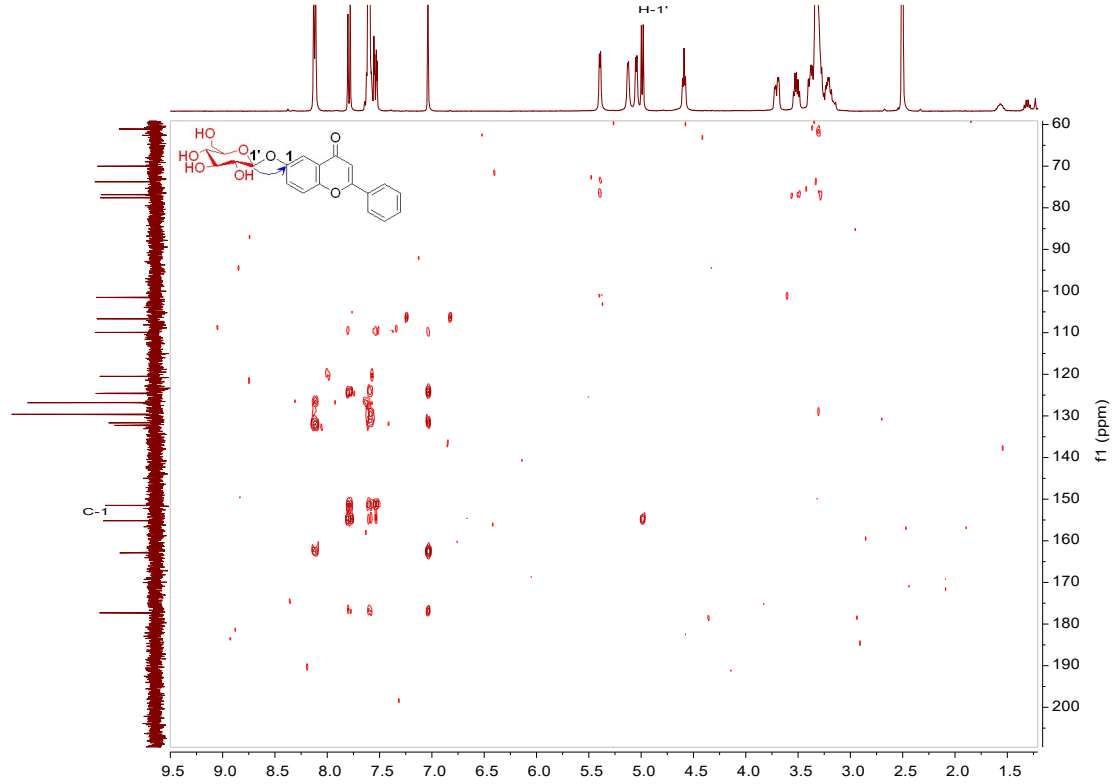


Figure S41. HMBC spectrum of compound **8a** (400 MHz/101 MHz, DMSO-d6)

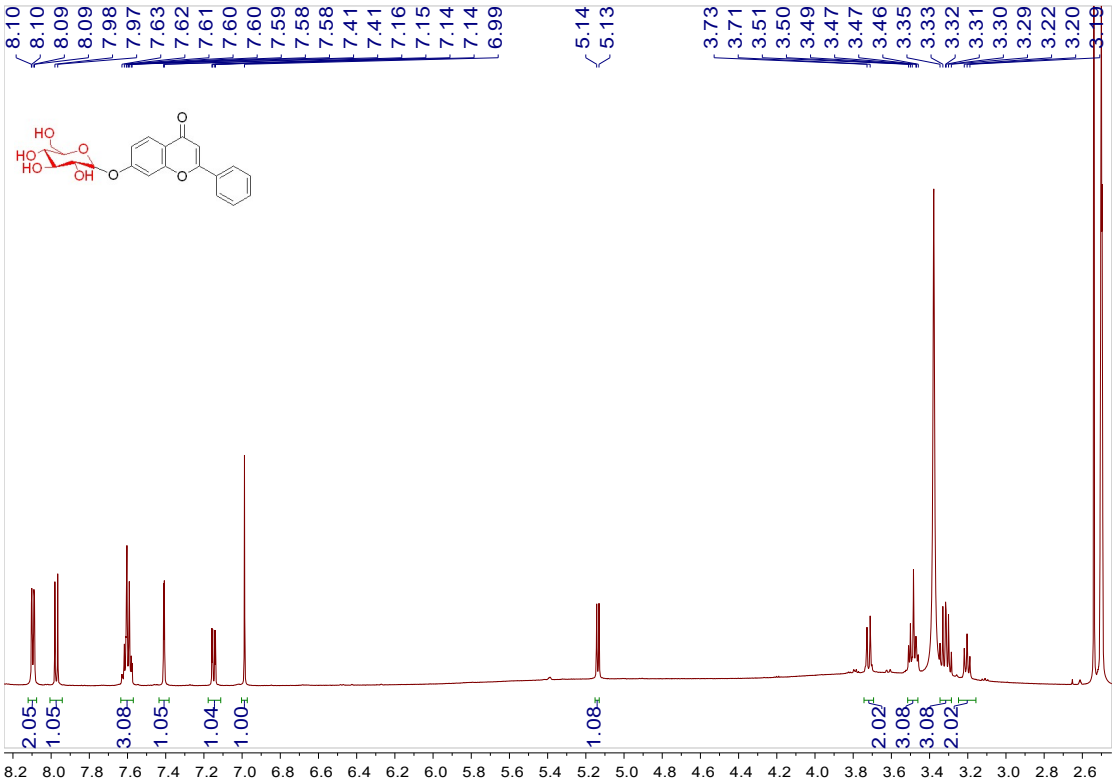


Figure S42. ^{1}H NMR spectrum of compound **9a** (400 MHz, DMSO-d6)

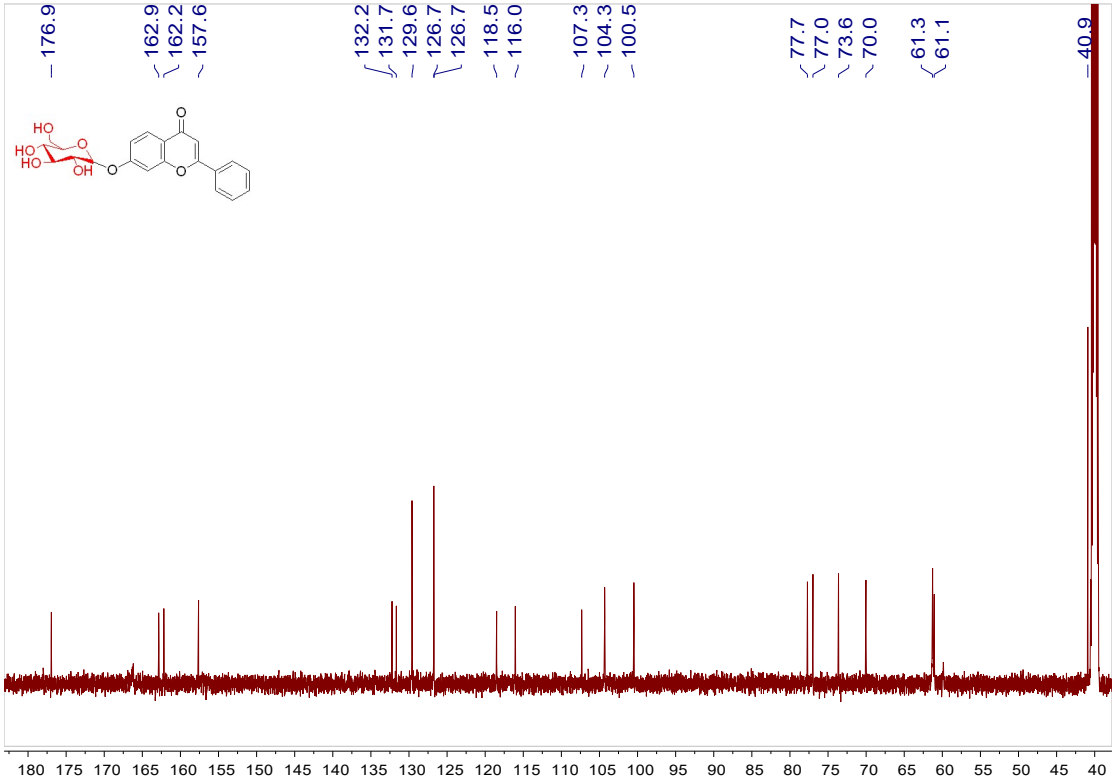


Figure S43. ^{13}C NMR spectrum of compound **9a** (101 MHz, DMSO-d6)

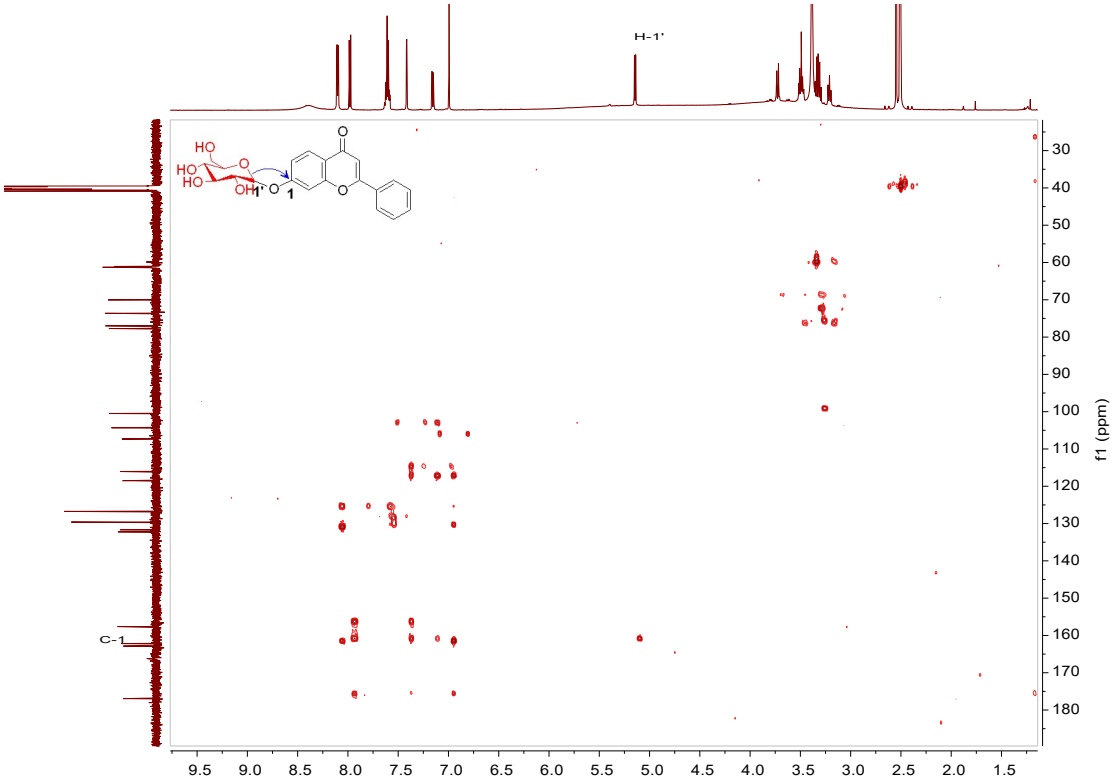


Figure S44. HMBC spectrum of compound **9a** (400 MHz/101 MHz, DMSO-d6)

5. The DNA and protein sequences of BarGT-1, -2, and -3.

>BarGT-1 (DNA sequence)

ATGCTGAACATTCTGGTGGTGAACTTTCCGGCGGAAGGCCATGTGAACC
CGACCCTGAACCTGGTGAAAGCGTTTACCAAACGCGGCGATAACGTGCA
TTATATTACCACCGCGAACTTTAAAGATCGCATTGAAGATCTGGGCGCGA
CCGTGCATATTCATCCGGATCTGCTGAAGGAAATTAGCATTGATGCGGA
AACGAGCAGCGGCCTGAACGCGTTTTTTCATGTGCATGTGCAGACGAGC
CTGTATATTCTGGAAATTACCAAACAGCTGTGCGAAAGCATTAACTTTGA
TTTTGTGATTTATGATATTTTTGGCGCGGGCGAACTGGTGAAAGAATATC
TGCAAGTGCCGGGCGTGGTGAGCAGCCCGATTTTTCTGATTCCGCCGG
AATTTCTGAAAACCCTGCCGTTTCATCCGAACGCGGATATGCCGTTTCAG
CCGGAAGAAATTAGCGAAAAACTGCTGAATCAGATGGAACATAAATTTGG
CGTGAAACCGAAAAACAACCTGCAGTTTATGAACAACAAAGGCGATGTG
TGCCTGGTGTATACGAGCCGCTATTTTCAGCCGAACAACGAGAGCTTTG
GCGAAAACAACATTTTTATTGGCCCGAGCATTAGCAAACGCAAAACCAAC
ATTAAATTTCCGCTGGAAAGCCTGAAAGAAAAAAAAGTGATTTATATTAGC
ATGGGCACCCTGCTGGAAGGCCTGGAACCGTTTTTTAACACCTGCATTG
ATACCTTTAGCGATTTTGATGGCATTGTGGTGATGGCGATTGGCGATCG
CAACGATATTAGCAAAATTAAAAAAGCGCCGGATAACTTTATTATTGCGC
CGTATGTGCCGCAGAGCGAAATTCTGAACGAAGCGGATGTGTTTATTAC
CCATGGCGGCATGAACAGCGTGCATGATGCGATTCATTATAACGTGCCG
TTTGTGATTATTCCGCATGATAAAGATCAGCCGATGATTGCGCAGCGCCT
GACCGAACTGGAAGCGGCGCATCGCCTGCTGAAAGAACATGTGAACGT
GCATACCCTGAAAGAAGCGGTGACCGATGTGCTGAGCAACGAAAAATAT
AAACATGGCATTCGCAAACTGAACGATAGCTTTCTGGAATGCGGCGGCA
GCAAAGAAGCGATTGCGGTGATTGAAAGCCTGCTGAACAAAGTGAAACT
GTAA

>BarGT-1 (Protein sequence)

MLNILVVNFPAEGHVNPTLNLVKAFTKRGDNVHYITTANFKDRIEDLGATVHI
HPDLLKEISIDAETSSGLNAFFHVHVQTSLYILEITKQLCESINFDFVIYDIFGA
GELVKEYLQVPGVVSSPIFLIPPEFLKTLPFHPNADMPFQPEEISEKLLNQME
HKFGVKPKNNLQFMNNKGDVCLVYTSRYFQPNNESFGENNIFIGPSISKRKT
NIKFPLESLKEKKVIYISMGTLLEGLEPFFNTCIDTFSDFDGIVVMAIGDRNDIS
KIKKAPDNFIIAPYVPQSEILNEADVFITHGGMNSVHDAIHYNVPFVIIPHDKDQ
PMIAQRLTELEAAHRLLKEHVNVHTLKEAVTDVLSNEKYKHGIRKLNDSFLEC
GGSKEAIAVIESLLNKVKL

>BarGT-2 (DNA sequence)

ATGGCGCGCGTGCTGTTTATTAACGCGGGCAGCGAAGGCCATATTAACC
CGACCCTGCAAGTGGTGGAAGAACTGATTAGCCGCGGCGAAGAAATTGT
GTATTTTAGCATTGAAGCGTTTCGCGAACGCATTGAAAAAACCGGCGCG
ACCGTGCGCACCATTGATGATCAGAAATTTATTAAAGCGTTTCTGAGCGG
CGGCCGCAACTATCTGCAAGAACGCATTAACGGCCTGCTGCATACCGCG
GATATTGTGATTCCGAGCGTGCTGGAACAGATTGAAGGCGAACATTTTG
ATTATATTATTCATGATAGCATGTTTGGCTGCGGCCATCTGATTGCGCAG
ATTCTGAAACTGCCGGCGATTAACAGCTGCACGAGCTTTGCGCAAGATG
AAAAGAGCTTTGAACAGATGCTGGGCCATCTGAGCAAAAACATTAGCGT
GGAAATTTATGATAAAATTCAGAACGATTTTCAGAACCTGACCAAAGGCA
TTGCGGAAAAATATGGCGTGGAAATTAAAAGCAGCTATGAAGTGTTTTGC
AACCCGGCGCCGCTGACCATTGTGTATACCATTAAAGAATTTCAGCCGTT
TGGCGATACCTTTGATGAAACCTATAAATTTGTGGGCCCGAGCATTAGCG
CGCAGATGAAAAACCGCGATGTGGATTTTACGAGCATTGAAGAAAAAAG
CCCGATTTATATTAGCCTGGGCACCGTGTTTAACGAAGCGATTGATTTTT
ATAAACTGTGCATGAAAGCGTTTGAAAACAGCGAACATACCATTGTGATG
AGCATTGGCAGCAAAACCAAAATTAGCGATCTGGGCGAAATTCCGAAAA
ACTTTATTGTGAAAAACTATGTGCCGCAGACCGAACTGCTGACCTATACC
AAACTGTTTATTACCCATGGCGGCATGAACAGCGCGCATGAAGGCCTGT
ATAACGGCATTCCGCTGGTGGTGATTCCGCAGAGCGCGGATCAGCCGG
TGGTGGCGAAACAAGTGGAAAGCCTGGGCGCGGGCATTAAACTGCAGA
TGCAAGGCCTGACCGCGGATCAGCTGAGCGAAAGCGTGGAAATGGTGC
TGAACAACCCGAGCTTTAAAGAAGTGGCGCTGAACCTGAAAAAGAGCTT
TCAGAAAAGCGGCGGCTATAAAGAAGCGGTGGATGAAATTTTTATTTTTG
TGGGTCAGTAA

>BarGT-2 (Protein sequence)

MARVLFINAGSEGHINPTLQVVEELISRGEEIVYFSIEAFRERIEKTGATVRTID
DQKFIKAFLSGGRNYLQERINGLLHTADIVIPSVLEQIEGEHFDYIIHDSMFGC
GHLIAQILKLPAINSCTSFAQDEKSFEQMLGHLSKNISVEIYDKIQNDFQNLTK
GIAEKYGVEIKSSYEVFCNPAPLTIVYTIKEFQPFGDTFDETYKFVGPSISAQM
KNRDVDFTSIEEKSPIYISLGTVFNEAIDFYKLCMKAFENSEHTIVMSIGSKTKI
SDLGEIPKNFIVKNYVPQTELLTYTKLFITHGGMNSAHEGLYNGIPLVVIPQSA
DQPVVAKQVESLGAGIKLQMQGLTADQLSESVEMVLNNPSFKEVALNLKKS
FQKSGGYKEAVDEIFIFVGQ

>BarGT-3 (DNA sequence)

ATGGCGAACGTGCTGGTGATTAACTTTCCGGGCGAAGGCCATATTAACC
CGACCCTGGCGATTGTGAGCGAACTGATTCAGCGCGGCGAAACCGTGG
TGAGCTATTGCATTGAAGATTATCGCAAAAAAGTGGAAGCGACCGGCGC
GGAATTTCGCGTGTTTGAAAACTTTCTGAGTCAGATTAACATTATGGAAC
GCGTGAACGAAGGCGGCAGCCCGCTGACCATGCTGAGCCACATGATTG
AAGCGAGCGAACGCATTGTGACGCAGATTGTGGAAGAAACCAAAGAAGA
AAAATATGATTATCTGATTTATGATAACCATTTTCCGGTGGGCCGCATTAT
TGCGAACATTCTGCATCTGCCGAGCGTGAGCAGCTGCACCACCTTTGCG
GTGAATCAGTATATTAACTTTCATGATGAACAAGAAAGCCGCCAAGTGGA
TGAAATGGATCCGCTGTATCAGAGCTGCCTGGCGGGCATGGAACGCTG
GAACAAACAGTATGGCATGAAATGCAACAGCATGTATGATATTATGAACC
ATCCGGGCGATATTACCATTGTGTATACGAGCAAAGAATATCAGCCGCG
CAGCGAAGTGTTTGATGAAAGCTATAAATTTGTGGGCCCGAGCATTGCG
ACCCGCAAAGAAGTGGGCAGCTTTCCGACCGAAGATCTGAAAAACGAAA
AAGTGATTTTTATTAGCATGGGCACCGTGTTTAACGAACAGCCGGCGCT
GTATGAAAAATGCTTTGAAGCGTTTAAAGATGTGGATGCGACCGTGGTG
CTGGTGGTGGGCAAAAAAATTAACATTAGTCAGTTTGAAAACATTCCGAA
AAACTTTAAACTGTATAACTATGTGCCGCAGCTGGAAGTGCTGCAGCATG
CGGATGTGTTTGTGACCCATGGCGGCATGAACAGCAGTAGCGAAGCGC
TGTATTATGGCGTGCCGCTGGTGGTGATTCCGGTGACCGGCGATCAGC
CGTTTGTGGCGAAACGCCTGACCGAAGTGGGCGCGGGCATTCGCCTGA
ACCGCAACGAACTGACGAGCGAACTGCTGCGCGAAGCGGTGAAAAAAG
TGATGGATGATGTGACCTTTAAAGAAAACAGCCGCAAAGTGGGCGAAAG
CCTGCGCAACGCGGGCGGCTATCAGCGCGCGGTGGAAGAAATTTTTAA
ACTGAAAATGAAACCGTATGTGAAAATTAAATAA

>BarGT-3 (Protein sequence)

MANVLVINFPGEGHINPTLAIVSELIQRGETVVSYCIEDYRKKVEATGAEFRV
FENFLSQINIMERVNEGGSPLTMLSHMIEASERIVTQIVEETKEEKYDYLIYDN
HFPVGRIIANILHLPSVSSCTTFAVNQYINFHDEQESRQVDEMDPLYQSCLA
GMERWNKQYGMKCNSMYDIMNHPGDITIVYTSKEYQPRSEVFDESYKFVG
PSIATRKEVGSFPTEDLKNEKVIFISMGTVFNEQPALYEKCFEAFKDVDATVV
LVVGKKINISQFENIPKNFKLYNYVPQLEVLQHADVFVTHGGMNSSSEALYY
GVPLVVIPVTGDQPFVAKRLTEVGAGIRLNRNELTSELLREAVKKVMDDVTF
KENSRKVGESLRNAGGYQRAVEEIFKLKMKPYVKIK

6. References

1. S. Kuraku, C. M. Zmasek, O. Nishimura and K. Katoh, aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity, *Nucleic acids research*, 2013, **41**, W22-28.
2. K. Katoh, J. Rozewicki and K. D. Yamada, MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization, *Briefings in bioinformatics*, 2017, DOI: 10.1093/bib/bbx108.
3. M. N. Price, P. S. Dehal and A. P. Arkin, FastTree: computing large minimum evolution trees with profiles instead of a distance matrix, *Molecular biology and evolution*, 2009, **26**, 1641-1650.
4. H. Shimodaira and M. Hasegawa, Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference, *Mol.biol.evol*, 1999, 8.
5. I. Letunic and P. Bork, Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation, *Bioinformatics*, 2007, **23**, 127-128.
6. K. Miyazaki, in *Method Enzymol*, Elsevier, 2011, vol. 498, pp. 399-406.
7. J. Yang, A. J. Ruff, M. Arlt and U. Schwaneberg, Casting epPCR (cepPCR): A simple random mutagenesis method to generate high quality mutant libraries, *Biotechnol Bioeng*, 2017, **114**, 1921-1927.
8. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature*, 2021, DOI: 10.1038/s41586-021-03819-2.
9. R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J Appl Crystallogr*, 1993, **26**, 283-291.
10. H. Land and M. S. Humble, in *Protein Eng*, Springer, 2018, pp. 43-67.
11. E. Krieger, R. L. Dunbrack, R. W. Hooft and B. Krieger, in *Computational Drug Discovery and Design*, Springer, 2012, pp. 405-421.
12. J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB, *J. Chem. Theory Comput.*, 2015, **11**, 3696-3713.
13. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, Development and testing of a general amber force field, *J. Comput. Chem.*, 2004, **25**, 1157-1174.
14. B. Liu, C. Zhao, Q. Xiang, N. Zhao, Y. Luo and R. Bao, Structural and biochemical studies of the glycosyltransferase Bs-YjiC from Bacillus subtilis, *Int J Biol Macromol*, 2020, DOI: 10.1016/j.ijbiomac.2020.10.238.
15. B. Guo, X. Hou, Y. Zhang, Z. Deng, Q. Ping, K. Fu, Z. Yuan and Y. Rao, Highly efficient production of rebaudioside D enabled by structure-guided engineering of bacterial glycosyltransferase YojK, *Frontiers in Bioengineering and Biotechnology*, 2022, **10**.