Electronic Supplementary Material (ESI) for Green Chemistry.

A Comparative Molecular Dynamics Approach Guides the Tailoring of Glycosyltransferases to Meet Synthetic Applications

Peng Zhang,‡[a] Shuaiqi Meng,‡[a] Zhongyu Li,‡[a] Dennis Hirtz,[b] Lothar Elling,[b] Leilei Zhu,[d] Yu Ji,*[a,c] Ulrich Schwaneberg,*[a]

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

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

c. College of Life Science and Technology, Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, 100029 Beijing, China

d. Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, 300308 Tianjin, China

‡These authors contributed equally to this work.

*Corresponding authors

Table of Contents

1. Experimental Procedures

1.1. Chemicals and reagents

All chemicals were of analytical grade and purchased from Sigma-Aldrich (Steinheim, Germany), AppliChem (Darmstadt, Germany), ExtraSynthese (Genay Cedex, France), TCI (Eschborn, Germany), Cymit (Barcelona, Spain), and VWR International (Darmstadt, Germany).

1.2. Bacterial strains, plasmids and media

Electrocompetent *E. coli* BL21 (DE3) (Agilent Technologies, Santa Clara, CA) was prepared for the protein expression. The genes encoding BarGT-1 (QEL68904), BarGT-3 (QEL68377), BsyGT (AVX09102) and BgoGT (ALC81527) with pET-28a (+) vector were 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).

1.3. Site-saturated mutagenesis libraries

The site-saturated mutagenesis (SSM) libraries of BarGT-1 and BarGT-3 were generated by one-step PCR (primers in **Table S1**). PCR mixture contained: VeriFi™ polymerase and mixes (25 µL, PCR Biosystems Ltd., London, UK), template (0.25 µL,10 ng), forward and reverse primers (400 nM, 1 µL each), and ddH_2O (up to 50 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 (1 µL, 10 U/µL) for an 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 electroporate into the electrocompetent cells *E. coli* BL21(DE3) (1900 V).

1.4. Cultivation of libraries in 96-well microtiter plates

Colonies of SSM 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). A total of 180 colonies for each single SSM library were picked in 96-

well microtiter plates. 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.

1.5. UDP-glucose-cycling high throughput screening system

The UDP-Glucose-cycling HTS system has been reported in previous protocol.[1] Briefly, 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 20 µL variants, wild type (WT) or empty vector (EV); 500 µM of 4-nitrophenyl-β-D-glucoside (4NPG); 100 µM of UDP-Glc and 500 µM of 2-naphthol (**2**) and 6-hydroxyflavone (**6**, 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 (∆Abs, fold):

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

1.6. Protein expression and purification

As reported in previous protocol,[1] 2 mL seeding transformants of wild types (BarGT-1 and BarGT-3) and variants 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 30 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, 5 s pulse, 10 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 elutes 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.

1.7. *In vitro* glycosylation reactions

The purified GTs were subjected to *in vitro* glycosylation reactions. The reaction was performed in a mixture (100 μL) containing final concentrations of 50 mM Tris-HCl buffer, 10 mM $MgCl_2$, 2 mM UDP-Glc, 1 mM substrate (dissolved in DMSO) and 200 μg/mL each protein. The glycosylation reaction of liquiritigenin (**13**) was performed in a mixture (100 μL) containing final concentrations of 50 mM Tris-HCl buffer, 10 mM $MgCl_2$, 4 mM UDP-Glc, 1 mM substrate (dissolved in DMSO) and 400 μg/mL each protein. To determine the variation of conversion rate of liquiritigenin with incubation time, the reactions were sampled at different time points (0.2, 0.4, 0.6, 0.8, 1, 2, 4, 6, 8 and 10 h). All the above reactions were terminated with an equal volume of acetonitrile. After centrifugation at 14,000 rpm for 30 min to remove protein precipitates, the reaction mixtures were analyzed by HPLC-MS.

1.8. 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 S2**. For the preparative scale reaction, a 20 mL system containing enzyme (500 μg/mL), 10 mM UDP-Glc, 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 8 h at 37°C and stopped by adding the same volume of acetonitrile. 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).

1.9. Identification of kinetic parameters

Kinetic parameters of BarGT-1 and its variant K321P for the glycosylation of substrates 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, 10 mM UDP-Glc and different concentrations of substrates at 37°C for 10 min (triplicate and repeat 3 times). The kinetic parameters of BarGT-1 and its variant K321P for sugar donor (UDP-Glc) 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, 10 mM 6-hydroxyflavone (**3**) and different concentrations of UDP-Glc at 37°C for 10 min (triplicate and repeat 3 times). The samples were quenched with an equal volume of acetonitrile and analyzed by HPLC. The kinetic values were determined by fitting the Michaelis–Menten curve to the data using the nonlinear regression method (**Fig. S8-S9**).

1.10. Phylogenetic analysis

After removing redundant sequences, a total of 254 GT sequences from nBGT-1 branch were conducted multiple sequence alignments using MAFFT (https://mafft.cbrc.jp/alignment/server/).[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]

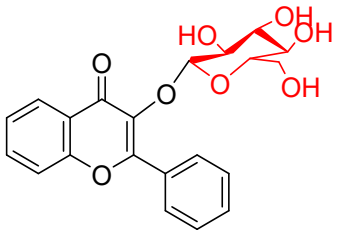
1.11. Computational analysis

The AlphaFold2 server was used to generate the structures of GT BarGT-1, -3, BsyGT, BgoGT and their variants.[6] The co-crystal structure of OleD (4M83) with the UDP and erythromycin A complex served as the reference structure. The docking and MD simulation process: YASARA software was employed to perform a structural alignment between OleD and BarGT-1 (or its variants).[7] Two docking boxes (box_{UDPG} and $box_{\mathbf{3}}$) centering on UDP and erythromycin A were defined with 5 Å size for BarGT-1 (wild type) and variant K321P. **3** was initially placed in $box_{\mathbf{3}}$ for docking, and the optimal docking $complex_{\mathbf{3}}$ was selected based on scoring and was further optimized by energy minimization and equilibrated using a 1 ns NVT ensemble followed by a 1 ns NPT ensemble. 200 ns of MD simulation on the $complex_{\mathbf{3}}$ was performed, and the dominant conformation ($DC_{\mathbf{3}}$) was clustered using a cut-off value of 0.2 nm (distance threshold) and selected as the starting conformation for the docking of sugar donor (UDP-Glc). UDP-Glc was placed in box_{UDPG} for docking, and the selected dominant conformation was further optimized by energy minimization and equilibration, and 1 μs MD simulation was finally performed for BarGT-1 complex. The molecular dynamics (MD) simulations toward BarGT-1, -3, BsyGT, BgoGT and their variants were performed using the GROMACS 2020.6 software package.[8, 9] The protein was solvated using the SPC water model.[10] The protein was centered in a 10 Å cubic box with periodic boundaries. The box was filled with around 30394 water molecules. The system was neutralized by Na^+ and Cl^- to achieve a net charge of zero. The AMBER14 force field[11] was used for the residues of GTs. The Generalized Amber Force Field (GAFF) was utilized to cope with the ligands. 5000 steps of the steepest descent followed by 5000 steps of conjugate gradients were used for energy minimization. The simulation of GTs was equilibrated by a 1 ns NVT ensemble followed by a 1 ns NPT ensemble, during which position restraints were applied to protein-heavy atoms. The production simulation was performed at 298 K for with three replicates (BarGT-3 for 1 μs; BsyGT and BgoGT for 400 ns). The binding free

energy was calculated using the molecular mechanics-poisson-boltzmann surface area (MM-PBSA) method with AMBER14. This method, known for its classical approach, is considered more appropriate for determining the free energy differences between two states compared to alternative methods.[12]

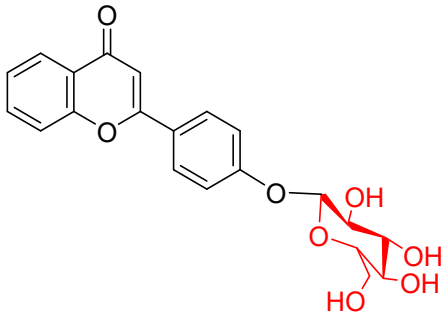
2. ^{1}H and ^{13}C NMR spectral data of products 10a-11a and 13a-13c

10a: 3-hydroxyflavone *O*-β-D-glucoside



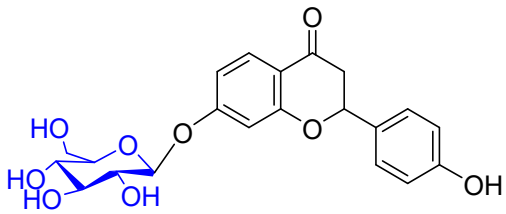
^{1}H NMR (600 MHz, DMSO-d_6) δ 8.20 – 8.15 (m, 2H), 8.12 (dd, *J* = 8.0, 1.5 Hz, 1H), 7.85 (ddd, *J* = 8.6, 7.1, 1.6 Hz, 1H), 7.76 (d, *J* = 8.2 Hz, 1H), 7.57 – 7.50 (m, 4H), 5.59 (d, *J* = 7.8 Hz, 1H), 5.39 (d, *J* = 3.6 Hz, 1H), 5.12 – 5.05 (m, 1H), 4.98 (s, 1H), 4.34 – 4.27 (m, 3H), 3.55 (dd, *J* = 11.6, 3.7 Hz, 1H), 3.17 – 3.07 (m, 3H). **^{13}C NMR (151 MHz, DMSO-d_6)** δ 173.6, 155.7, 154.8, 136.3, 134.2, 130.7, 130.7, 129.1, 128.2, 125.2, 125.1, 123.4, 118.4, 100.5, 77.5, 76.5, 74.2, 69.9, 63.5, 60.8, 56.7.

11a: 4'-hydroxyflavone *O*-β-D-glucoside



^{1}H NMR (600 MHz, DMSO-d_6) δ 8.09 (d, *J* = 8.8 Hz, 2H), 8.05 (d, *J* = 7.9 Hz, 1H), 7.83 (td, *J* = 7.7, 7.1, 1.3 Hz, 1H), 7.79 (d, *J* = 8.2 Hz, 1H), 7.50 (t, *J* = 7.4 Hz, 1H), 7.21 (d, *J* = 8.9 Hz, 2H), 6.99 (s, 1H), 5.39 (d, *J* = 4.6 Hz, 1H), 5.14 (d, *J* = 3.9 Hz, 1H), 5.07 (d, *J* = 5.0 Hz, 1H), 5.03 (d, *J* = 7.3 Hz, 1H), 4.60 (t, *J* = 5.6 Hz, 1H), 3.71 (dd, *J* = 11.9, 3.5 Hz, 1H), 3.50 – 3.39 (m, 3H), 3.27 – 3.15 (m, 2H). **^{13}C NMR (151 MHz, DMSO-d_6)** δ 176.1, 161.5, 159.2, 154.6, 133.2, 127.1, 124.4, 123.8, 123.4, 122.3, 117.5, 115.6, 104.7, 98.9, 76.2, 75.6, 72.2, 68.6, 60.6, 59.6.

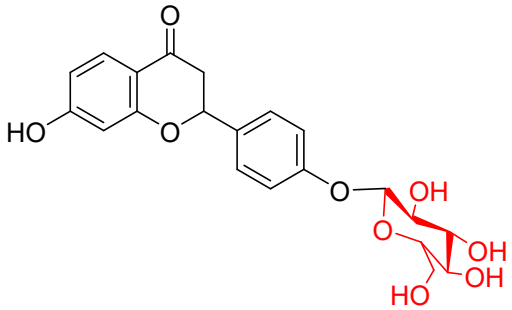
13a: neoliquiritin



^{1}H NMR (400 MHz, DMSO-d_6) δ 7.72 (d, *J* = 8.8 Hz, 1H), 7.34 (d, *J* = 8.4 Hz, 2H), 6.79 (d, *J* = 8.4 Hz, 2H), 6.71 (dt, *J* = 8.8, 1.7 Hz, 1H), 6.66 (s, 1H), 5.50 (dt, *J* = 13.0, 3.2 Hz, 1H), 5.37 (d, *J* = 3.8 Hz, 1H), 5.15 – 5.08 (m, 1H), 5.06 – 4.96 (m, 2H), 4.55 (s, 1H), 3.69 – 3.63 (m, 1H), 3.44 (dt, *J* = 11.6, 5.7 Hz, 2H),

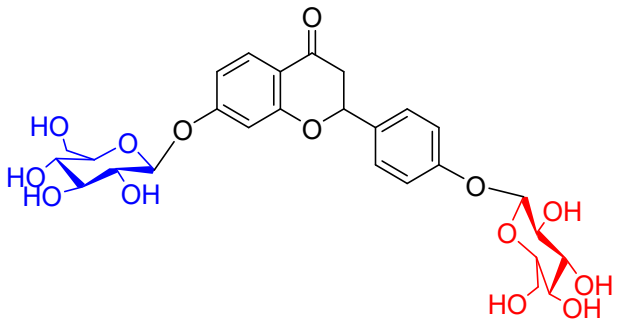
3.33 – 3.21 (m, 3H), 3.21 – 3.09 (m, 2H), 2.67 (dt, *J* = 16.6, 2.2 Hz, 1H). **^{13}C NMR (101 MHz, DMSO-d_6)** δ 190.5, 163.5, 163.4, 162.9, 157.7, 129.1, 129.0, 128.4, 128.4, 127.9, 115.3, 115.2, 110.9, 103.5, 99.6, 79.2, 77.1, 76.4, 73.1, 69.5, 60.6.

13b: liquiritin



^{1}H NMR (400 MHz, DMSO-d_6) δ 10.59 (s, 1H), 7.65 (d, *J* = 8.7 Hz, 1H), 7.45 (d, *J* = 8.7 Hz, 2H), 7.06 (d, *J* = 8.6 Hz, 2H), 6.51 (dd, *J* = 8.7, 2.2 Hz, 1H), 6.35 (d, *J* = 2.2 Hz, 1H), 5.53 (dd, *J* = 12.7, 2.6 Hz, 1H), 5.32 (d, *J* = 4.7 Hz, 1H), 5.10 (d, *J* = 4.4 Hz, 1H), 5.03 (d, *J* = 5.2 Hz, 1H), 4.89 (d, *J* = 7.2 Hz, 1H), 4.56 (t, *J* = 5.7 Hz, 1H), 3.69 (dd, *J* = 10.9, 4.4 Hz, 1H), 3.46 (dt, *J* = 11.8, 5.9 Hz, 1H), 3.33 – 3.20 (m, 3H), 3.19 – 3.02 (m, 2H), 2.67 (dd, *J* = 16.8, 2.7 Hz, 1H). **^{13}C NMR (101 MHz, DMSO-d_6)** δ 189.9, 164.6, 163.1, 157.5, 132.4, 128.4, 128.0, 128.0, 116.2, 113.6, 110.6, 102.6, 100.8, 100.3, 78.7, 77.1, 76.6, 73.2, 69.7, 60.7, 43.2.

13c: liquiritigenin 4',7-*O*-β-D-diglucoside



^{1}H NMR (400 MHz, DMSO-d_6) δ 7.72 (d, *J* = 8.7 Hz, 1H), 7.46 (d, *J* = 8.6 Hz, 2H), 7.07 (d, *J* = 8.7 Hz, 2H), 6.76 – 6.65 (m, 2H), 5.59 (dt, *J* = 12.9, 2.6 Hz, 1H), 5.35 (dd, *J* = 19.0, 4.5 Hz, 2H), 5.08 (dt, *J* = 28.2, 4.9 Hz, 4H), 4.99 (d, *J* = 7.0 Hz, 1H), 4.89 (d, *J* = 7.3 Hz, 1H), 4.56 (q, *J* = 5.3 Hz, 2H), 3.68 (td, *J* = 10.4, 4.9 Hz, 2H), 3.50 – 3.35 (m, 5H), 3.23 – 3.08 (m, 4H), 2.73 (dd, *J* = 16.6, 2.3 Hz, 1H), 2.54 – 2.48 (m, 2H). **^{13}C NMR (101 MHz, DMSO)** δ 190.3, 163.5, 163.4, 162.8, 157.5, 132.1, 128.2, 128.0, 116.2, 115.3, 111.0, 103.5, 100.3, 99.7, 99.6, 78.9, 77.1, 76.6, 76.4, 73.2, 73.1, 69.7, 69.5, 63.1, 60.7, 60.6, 43.1.

3. Additional Figures and Tables

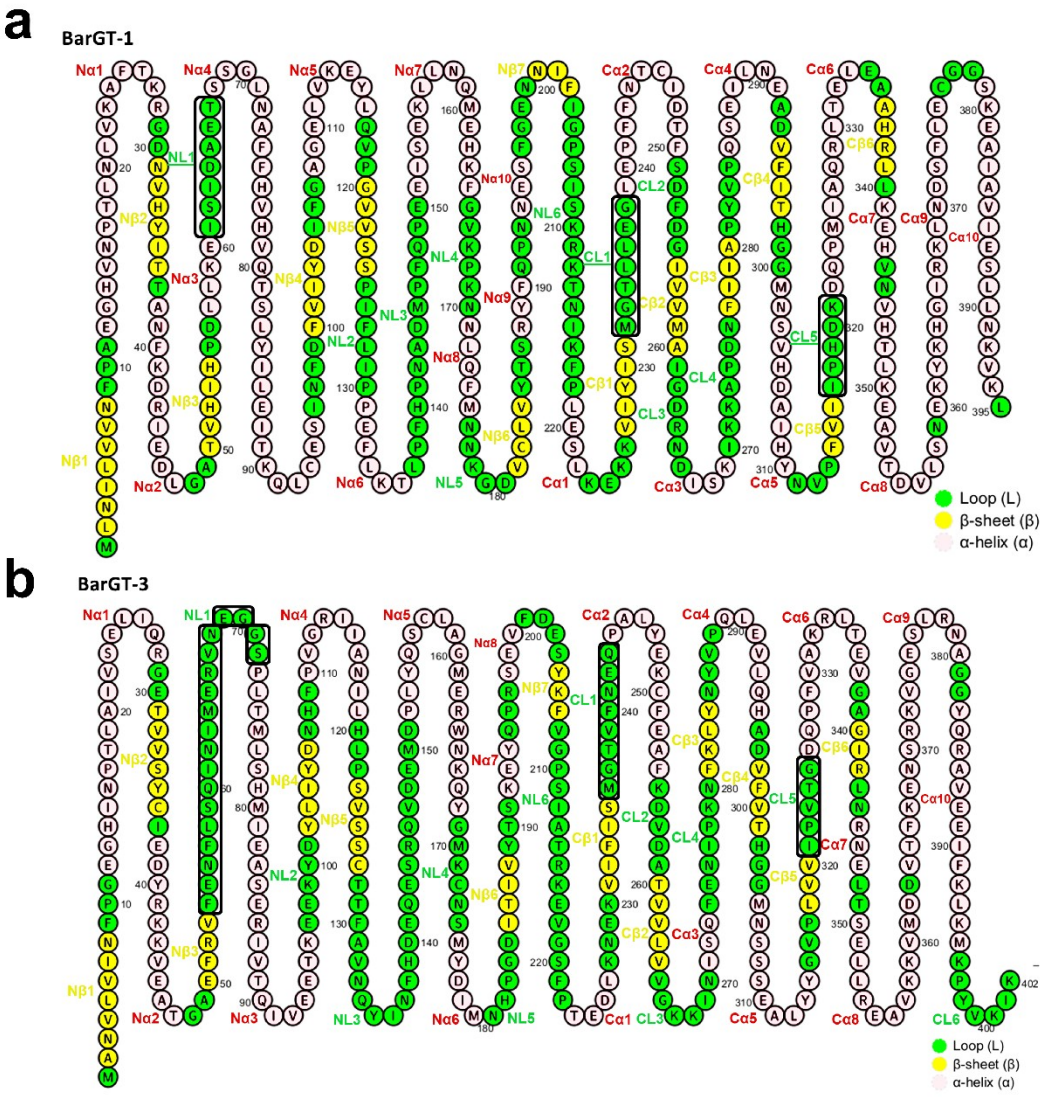


Fig. S1. The ordering of secondary components of BarGT-1 and -3. **a**. The ordering of secondary components of BarGT-1. **b**. The ordering of secondary components of BarGT-3. The α-helix, β-sheet and loop are colored by light pink, yellow and green, respectively.

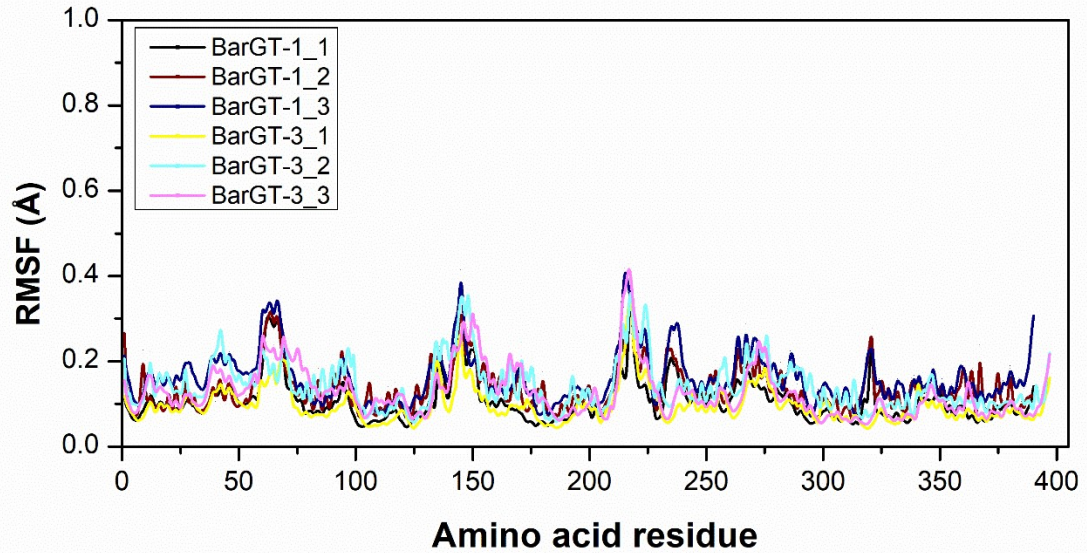


Fig. S2. The RMSF values for the α-carbon of each residue of BarGT-1 and BarGT-3 with three parallel MD simulations.

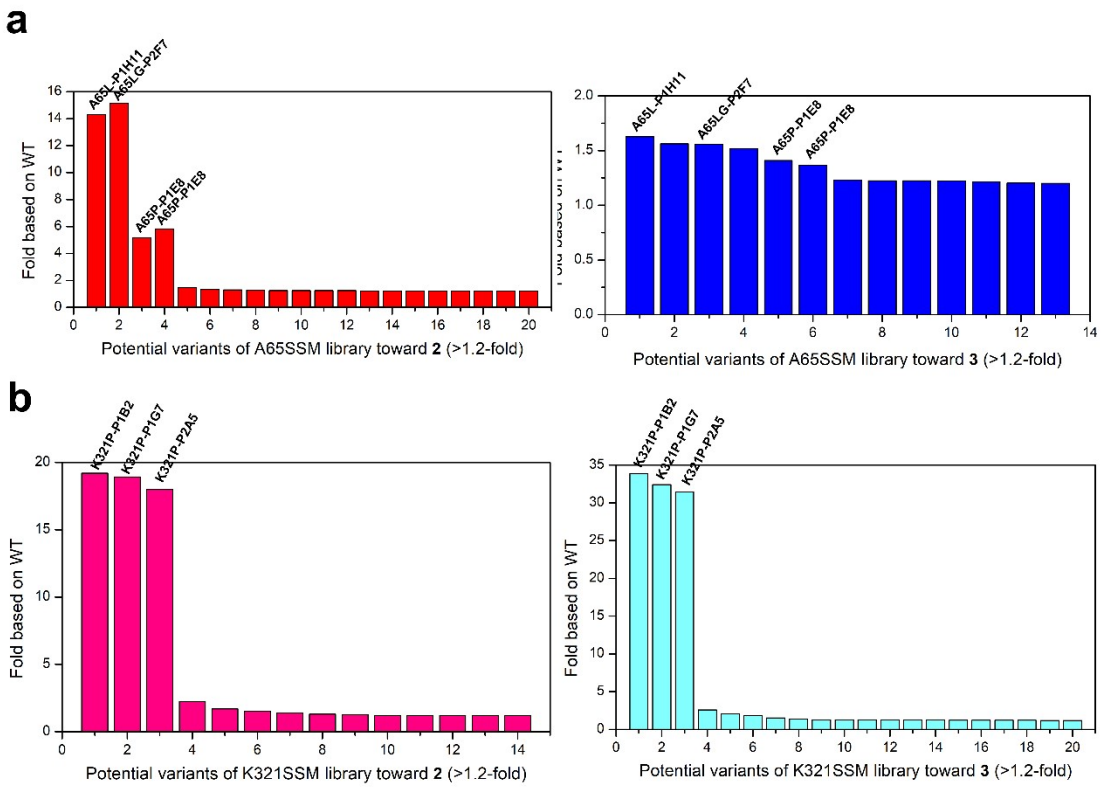


Fig. S3. Potential active variants in SSM libraries of A65 and K321. **a**. The screening of A65 SSM library towards **2** and **3**, respectively. The meaning of "P1H11" is the variant of "Row H, column 11" in plate 1. **b**. The screening of K321 SSM library towards **2** and **3** and K321P was identified with significant improvement in absorbance (405 nm) towards **2** and **3**.

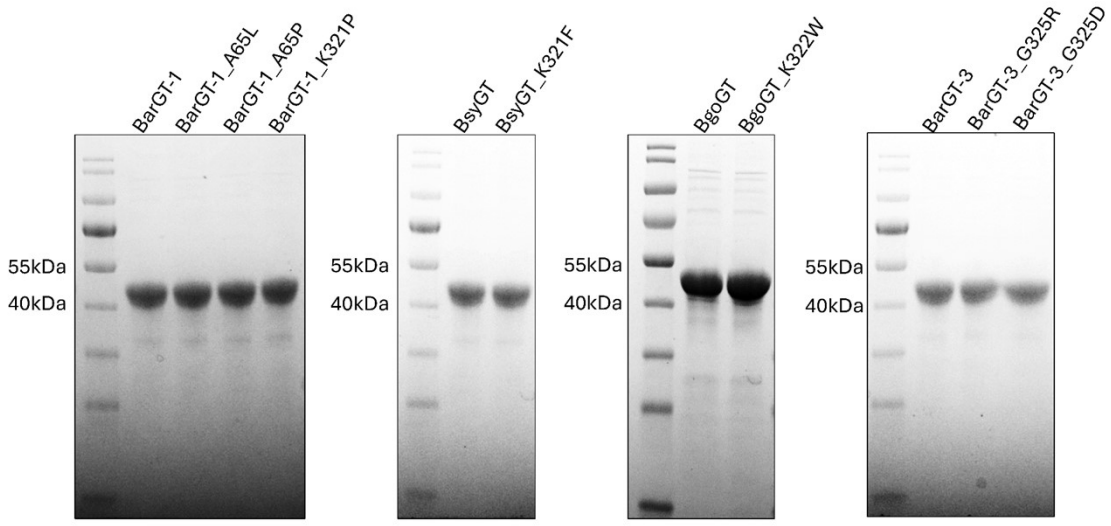


Fig. S4. The SDS-PAGE detection of the purification of proteins BarGT-1, BsyGT, BgoGT and BarGT-3 with their variants.

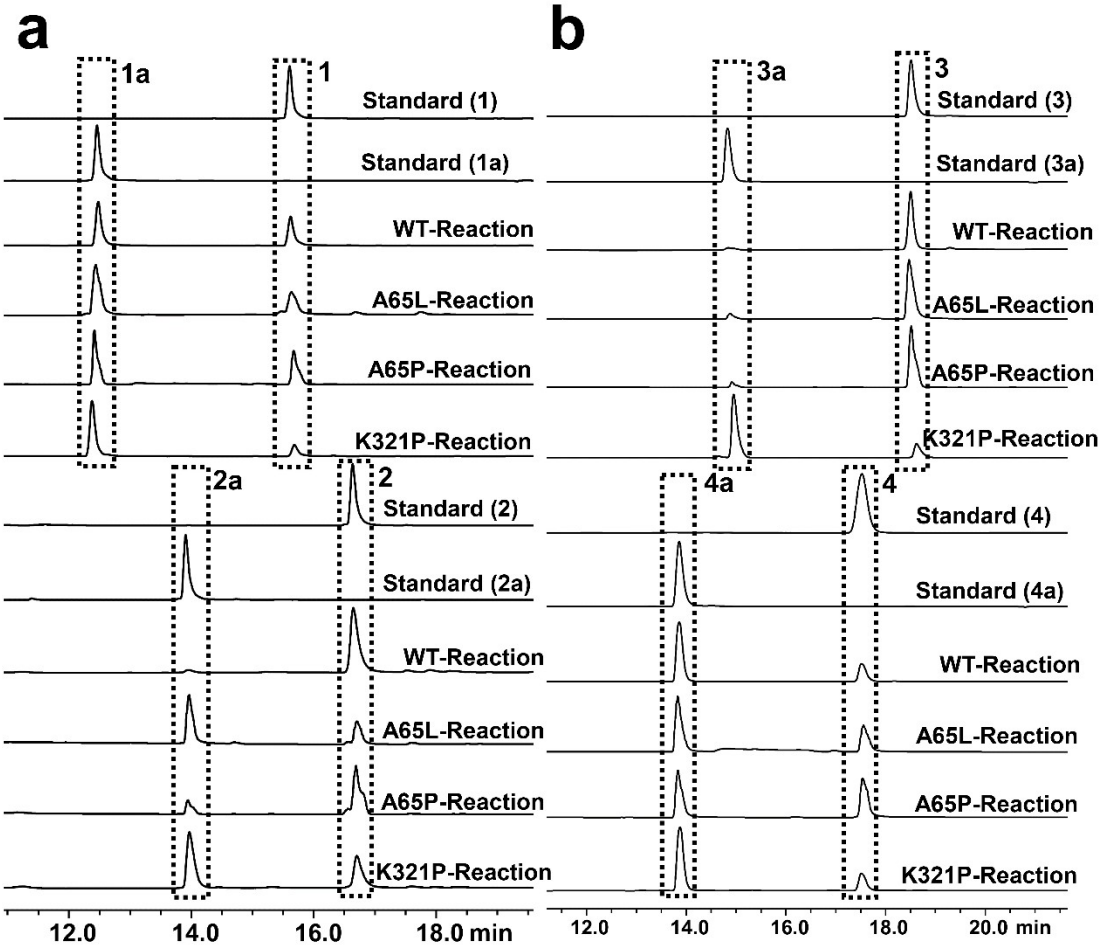


Fig. S5. The HPLC analysis of the four reactions (**a**, towards **1** and **2**; **b**, towards **3** and **4**) of glycosylated products, 1-naphthol *O*-β-D-glucoside (**1a**), 2-naphthol *O*-β-D-glucoside (**2a**), 6-hydroxyflavone *O*-β-D-glucoside (**3a**) and 7-hydroxyflavone *O*-β-D-glucoside (**4a**) obtained from glycosylation reactions (A 100 μL mixture containing final concentrations of 50 mM Tris-HCl buffer, 10 mM $MgCl_2$, 2 mM UDP-Glc, 1 mM substrate and 200 μg/mL each protein) of BarGT-1 (WT) A65L/P and K321P.

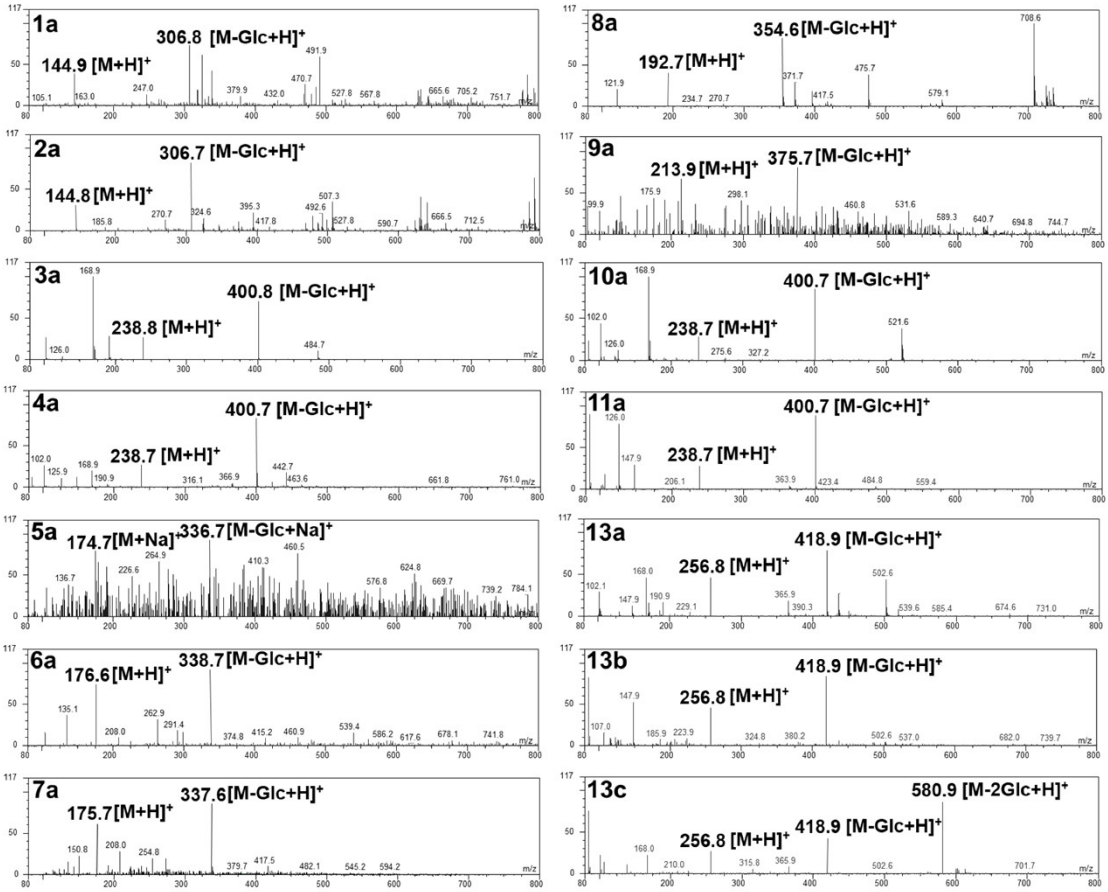


Fig. S6. The MS analysis of the glycosylated products **1a**-**11a** and liquiritigenin glycosides **13a**-**13c**. The glycosylated products were **1a**-**11a**: 1-naphthol *O*-β-D-glucoside **1a**, 2-naphthol *O*-β-D-glucoside **2a**, 6-hydroxyflavone *O*-β-D-glucoside **3a**, 7-hydroxyflavone *O*-β-D-glucoside **4a** acetaminophen *O*-β-D-glucoside **5a**, 7-hydroxy-4-methylcoumarin *O*-β-D-glucoside **6a**, 7-amino-4-methylcoumarin *N*-β-D-glucoside **7a**, 7-mercapto-4-methylcoumarin *S*-β-D-glucoside **8a**, resorufin *O*-β-D-glucoside **9a**, 3-hydroxyflavone *O*-β-D-glucoside **10a**, 4'-hydroxyflavone *O*-β-D-glucoside **11a**. The liquiritigenin (**13**) glycosides were neoliquiritin (**13a**), liquiritin (**13b**), and liquiritigenin 4',7-*O*-β-D-diglucoside (**13c**).

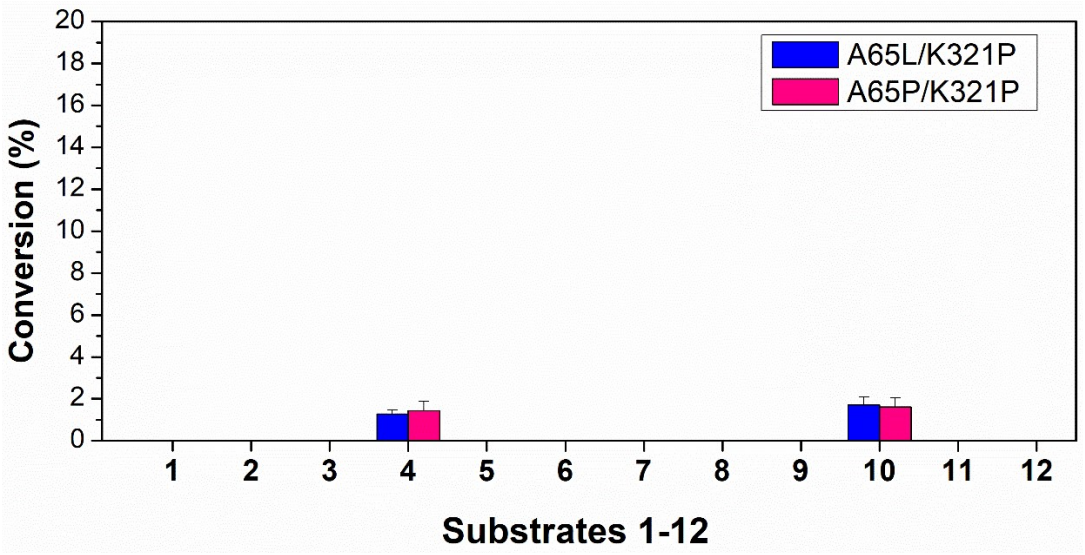


Fig. S7. The conversions of double-site variants A65L/K321P and A65P/K321P towards a panel of 12 substrates **1**-**12**.

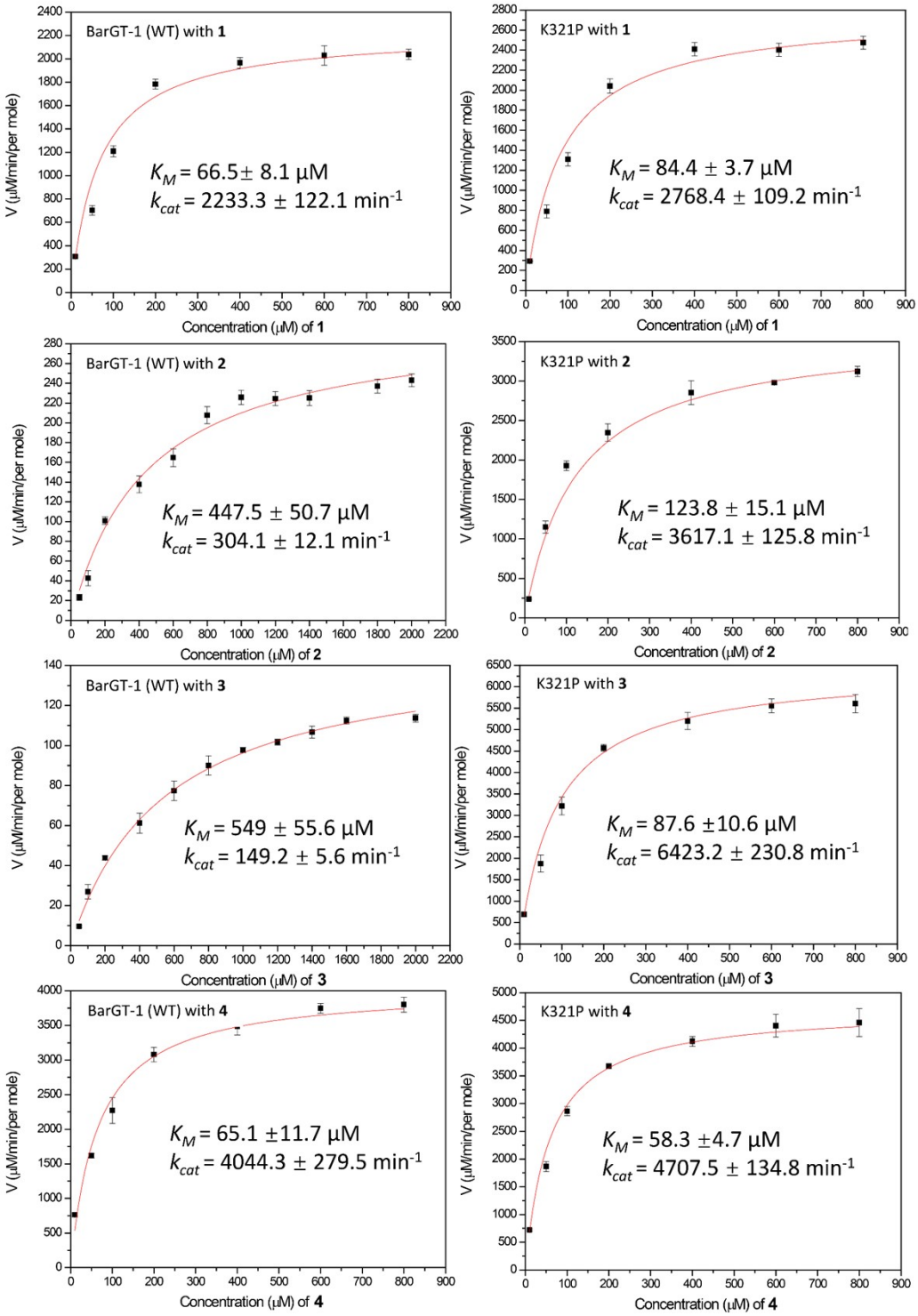


Fig. S8. Kinetic parameters and fitting curves analysis of BarGT-1 (WT) and its variant K321P. Determination of kinetic parameters for substrate **1**, **2**, **3** and **4** with saturated UDP-Glc (10 mM). GTs assays were performed in 50 mM Tris-HCl buffer (pH 7.5) containing 20 μg/mL protein and 10 mM $MgCl_2$ at 37°C for 10 min in triplicate.

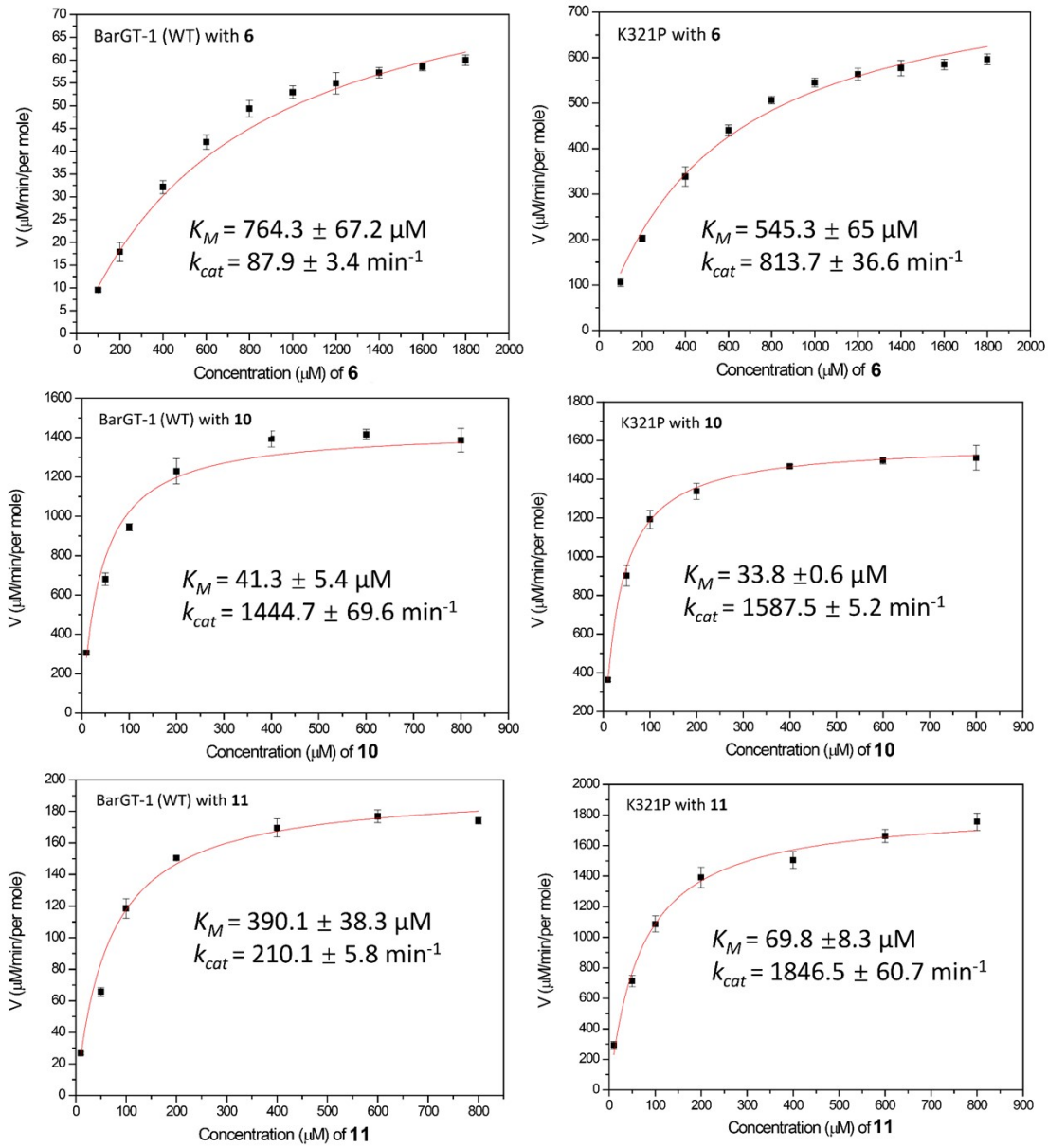


Fig. S9. Kinetic parameters and fitting curves analysis of BarGT-1 (WT) and its variant K321P. Determination of kinetic parameters for substrates **6**, **10** and **11** with saturated UDP-Glc (10 mM). GTs assays were performed in 50 mM Tris-HCl buffer (pH 7.5) containing 20 μg/mL protein and 10 mM $MgCl_2$ at 37°C for 10 min in triplicate.

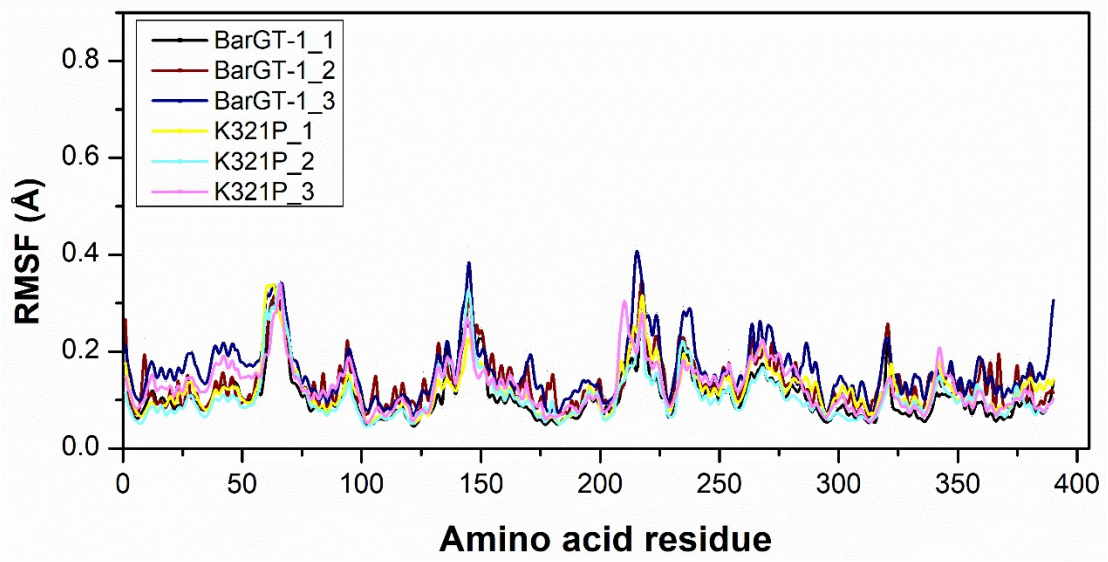


Fig. S10. The RMSF values for the α-carbon of each residue of BarGT-1 and its variant K321P with three parallel MD simulations.

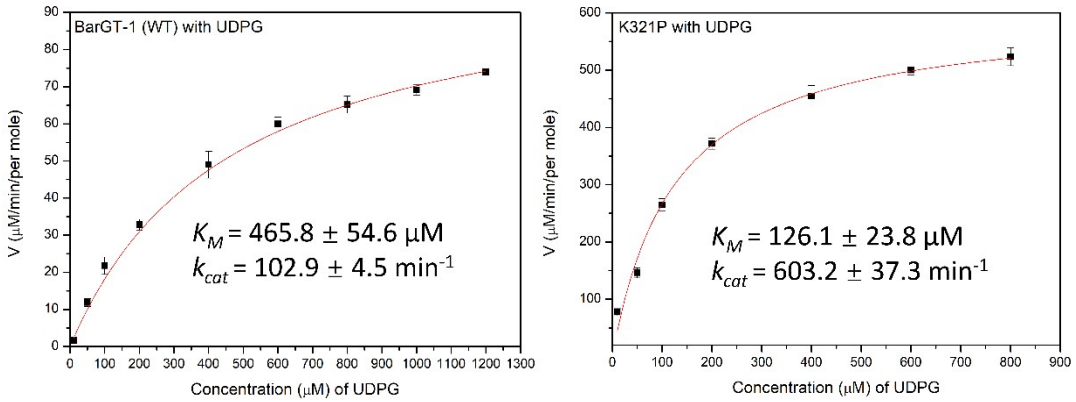


Fig. S11. Kinetic parameters and fitting curves analysis of BarGT-1 (WT) and its variant K321P. Determination of kinetic parameters for substrates UDP-Glc with saturated **3** (10 mM). GTs assays were performed in 50 mM Tris-HCl buffer (pH 7.5) containing 20 μg/mL protein and 10 mM $MgCl_2$ at 37°C for 10 min in triplicate.

Fig. S12. Multiple sequence alignment (MSA) analysis for BarGT-1 and -3 with the typical crystal structures of enzymes in GT1 family (YjiC PDB ID: 7BOV[13] and YojK PDB ID: 7VM0[14]).

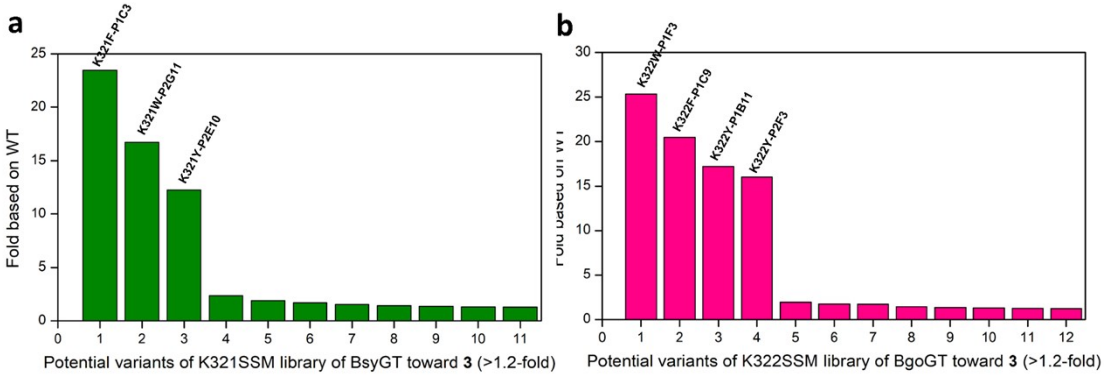


Fig. S13. Potential active variants in SSM libraries of K321 of BsyGT and K322 of BgoGT. a. The screening of K321 SSM library of BsyGT towards **3**, and K321F was identified with significant improvement in absorbance (405 nm) towards **3**. The meaning of "P1C3" is the variant of "Row C, column 3" in plate 1. b. The screening of and K322 SSM library of BgoGT towards **3** and K321W was identified with significant improvement in absorbance (405 nm) towards **3**.

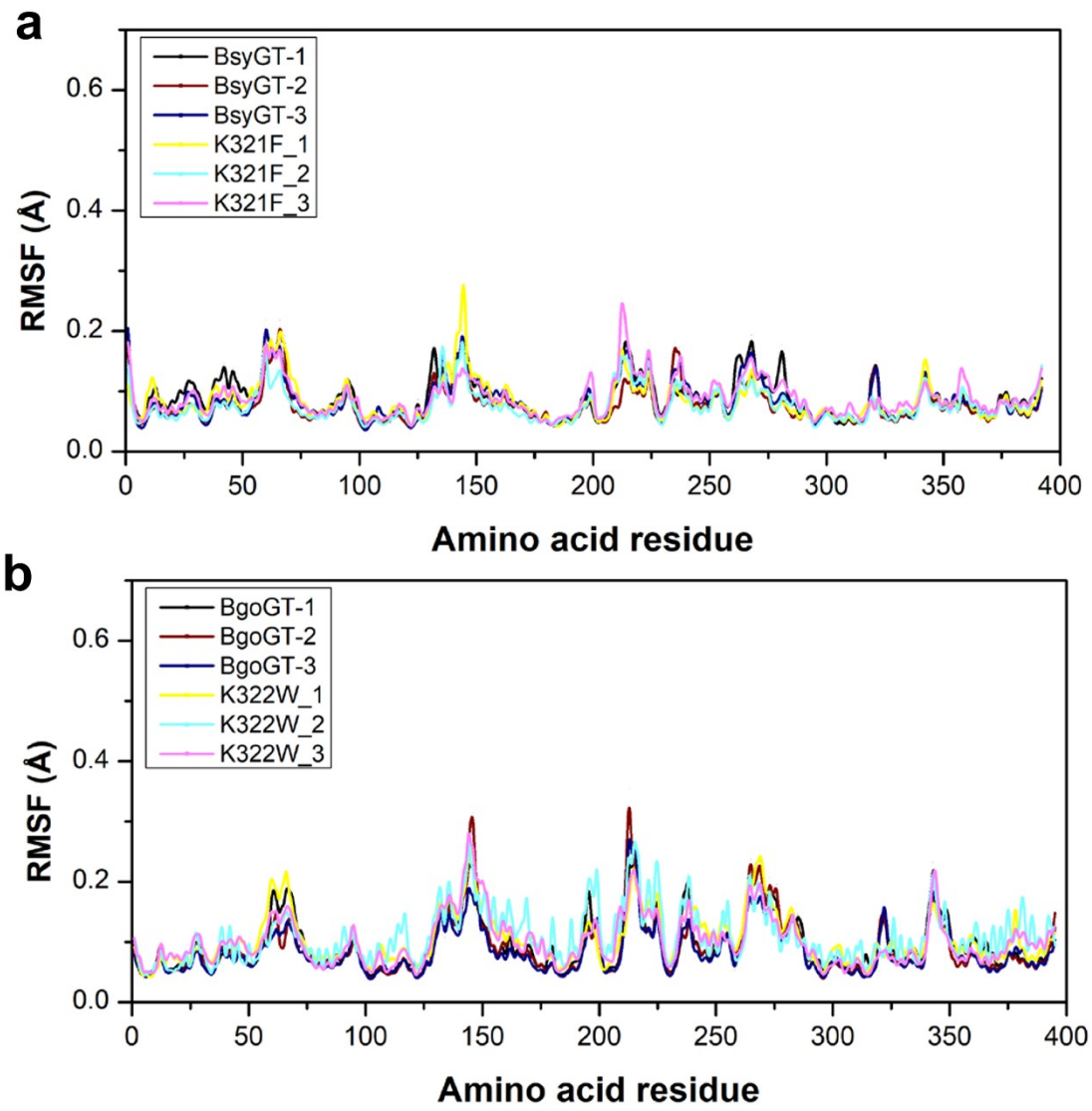


Fig. S14. a. The RMSF values for the α-carbon of each residue of BsyGT and K321F with three parallel MD simulations (400 ns). b. The RMSF values for the α-carbon of each residue of BgoGT and K322W with three parallel MD simulations.

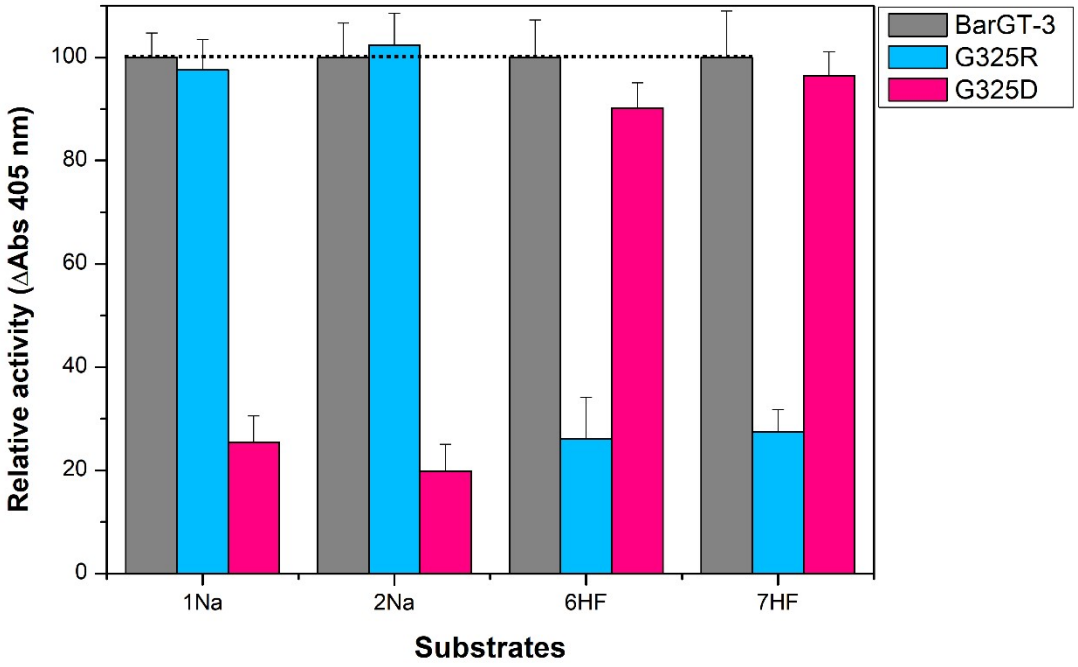


Fig. S15. The relative activity (ΔAbs, 405 nm) of BarGT-3 and its variants G325R/D towards substrates **1**, **2**, **3** and **4**. The ΔAbs (405 nm) of BarGT-3 was set as 100%.

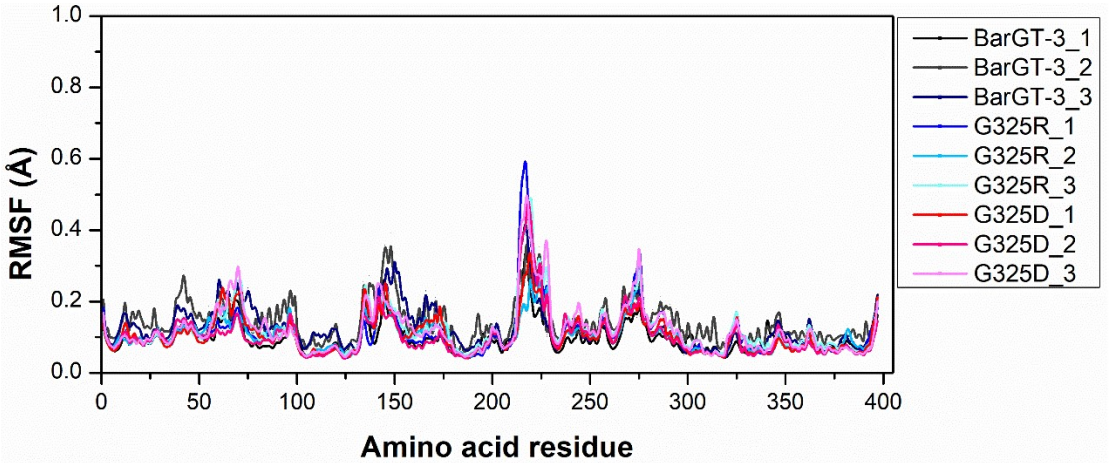


Fig. S16. The RMSF values for the α-carbon of each residue of BarGT-1 and its variants G325R/D with three parallel MD simulations.

Table S1. Primers for site-saturated mutagenesis (SSM) libraries.

Regions	Primers	Primers Sequence (5' to 3')
N-loop 1 of BarGT-1	BarGT-1-I61SSM-F	GAAGGAAnnsAGCATTGATGCGGAAACGAGCA
	BarGT-1-I61SSM-R	CAATGCTsnnTTCCTTCAGCAGATCCGGATGA
	BarGT-1-S62SSM-F	AATTnnsATTGATGCGGAAACGAGCAGCGGCC
	BarGT-1-S62SSM-R	CCGCATCAATsnnAATTTCCTTCAGCAGATCCGG
	BarGT-1-I63SSM-F	AAATTAGCnnsGATGCGGAAACGAGCAGCGGC
	BarGT-1-I63SSM-R	CGCATCsnnGCTAATTTCCTTCAGCAGATCCG
	BarGT-1-D64SSM-F	TAGCATTnnsGCGGAAACGAGCAGCGGCCTGA
	BarGT-1-D64SSM-R	TTTCCGCsnnAATGCTAATTTCCTTCAGCAGATCC
	BarGT-1-A65SSM-F	CATTGATnnsGAAACGAGCAGCGGCCTGAACG
	BarGT-1-A65SSM-R	TCGTTTCsnnATCAATGCTAATTTCCTTCAGCAG
	BarGT-1-E66SSM-F	TGATGCGnnsACGAGCAGCGGCCTGAACGCGT
	BarGT-1-E66SSM-R	TGCTCGTsnnCGCATCAATGCTAATTTCCTTCA
	BarGT-1-T67SSM-F	TGCGGAAnnsAGCAGCGGCCTGAACGCGTTTT
	BarGT-1-T67SSM-R	CGCTGCTsnnTTCCGCATCAATGCTAATTTCC
C-loop 1 of BarGT-1	BarGT-1-M232SSM-F	TATATTAGCnnsGGCACCCTGCTGGAAGGCCT
	BarGT-1-M232SSM-R	GTGCCsnnGCTAATATAAATCACTTTTTTTTCTTTCAG
	BarGT-1-G233SSM-F	ATTAGCATGnnsACCCTGCTGGAAGGCCTGGA
	BarGT-1-G233SSM-R	AGGGTsnnCATGCTAATATAAATCACTTTTTTTTCTTT
	BarGT-1-T234SSM-F	ATGGGCnnsCTGCTGGAAGGCCTGGAACCGTT
	BarGT-1-T234SSM-R	TCCAGCAGsnnGCCCATGCTAATATAAATCACTTTTT
	BarGT-1-L235SSM-F	TGGGCACCnnsCTGGAAGGCCTGGAACCGTTT
	BarGT-1-L235SSM-R	TTCCAGsnnGGTGCCCATGCTAATATAAATCAC
	BarGT-1-L236SSM-F	TGnnsGAAGGCCTGGAACCGTTTTTTAACACC
	BarGT-1-L236SSM-R	TTCCAGGCCTTCsnnCAGGGTGCCCATGCTAATATAAA
	BarGT-1-E237SSM-F	TGCTGnnsGGCCTGGAACCGTTTTTTAACACC
	BarGT-1-E237SSM-R	TTCCAGGCCsnnCAGCAGGGTGCCCATGCTAA
	BarGT-1-G238SSM-F	AAnnsCTGGAACCGTTTTTTAACACCTGCATT
	BarGT-1-G238SSM-R	AAACGGTTCCAGsnnTTCCAGCAGGGTGCCCAT
C-loop 5 of BarGT-1	BarGT-1-I317SSM-F	TGTGATTnnsCCGCATGATAAAGATCAGCCGA
	BarGT-1-I317SSM-R	CATGCGGsnnAATCACAAACGGCACGTTATAATG
	BarGT-1-P318SSM-F	GTGATTATTnnsCATGATAAAGATCAGCCGATGATT
	BarGT-1-P318SSM-R	TCATGsnnAATAATCACAAACGGCACGTTATAA
	BarGT-1-H319SSM-F	TCCGnnsGATAAAGATCAGCCGATGATTGCGC
	BarGT-1-H319SSM-R	GATCTTTATCsnnCGGAATAATCACAAACGGCAC
	BarGT-1-D320SSM-F	TCCGCATnnsAAAGATCAGCCGATGATTGCGC
	BarGT-1-D320SSM-R	GATCTTTsnnATGCGGAATAATCACAAACGGC
	BarGT-1-I317SSM-F	TGTGATTnnsCCGCATGATAAAGATCAGCCGA
	BarGT-1-I317SSM-R	CATGCGGsnnAATCACAAACGGCACGTTATAATG
Loop of BsyGT	BsyGT-K321SSM-F	TCACGGAnnsGATCAGCCGCTCGTTGCTCAAA
	BsyGT-K321SSM-R	GCTGATCsnnTCCGTGAGGTAAAACGACTAACG
Loop of BgoGT	BgoGT-K322SSM-F	CCGCATGACnnsGATCAACCAATGGTAGCACAGAGA
	BgoGT-K322SSM-R	TGATCsnnGTCATGCGGCAGAACAACCAGGGG
C-loop 5 of BarGT-3	BarGT3-G325SSM-F	GGTGACCnnsGATCAGCCGTTTGTGGCGAAAC
	BarGT3-G325SSM-R	GCTGATCsnnGGTCACCGGAATCACCACCAGC

Table S2. HPLC methods used for glycosylated products analysis.

Products	Analytical methods	Preparative methods	UV (nm)
1a	10-90% B, 30 min	35% B, 30 min	230
2a	10-90% B, 30 min	35% B, 30 min	230
3a	10-90% B, 30 min	35% B, 30 min	330
4a	10-90% B, 30 min	30% B, 30 min	330
5a	10-90% B, 30 min	-	257
6a	10-90% B, 30 min	20% B, 30 min	320
7a	10-90% B, 30 min	-	320
8a	10-90% B, 30 min	-	320
9a	10-90% B, 30 min	-	320
10a	10-90% B, 30 min	30% B, 30 min	330
11a	10-90% B, 30 min	30% B, 30 min	330
12a	10-90% B, 30 min	-	330
13a	10-90% B, 60 min	30% B, 30 min	330
13b	10-90% B, 60 min	30% B, 30 min	330
13c	10-90% B, 60 min	30% B, 30 min	330

4. Figures S17-S31. NMR spectrum of glycosylated products 10a-11a and 13a-13c.

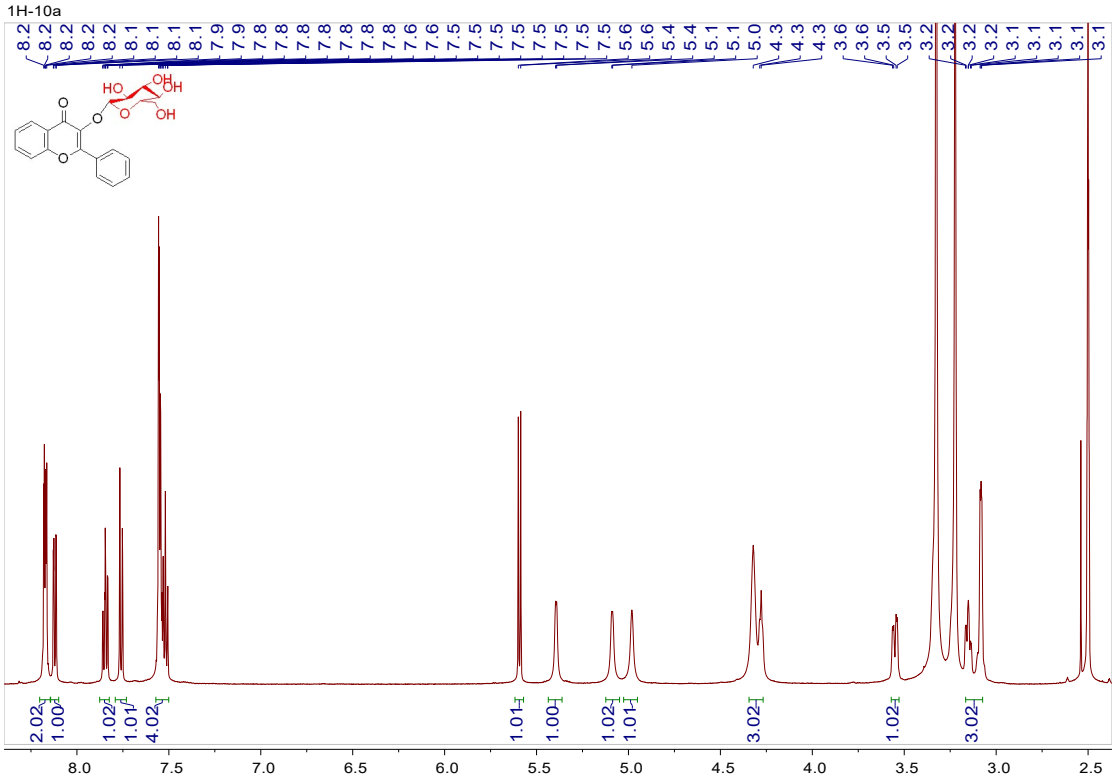


Fig. S17. ^{1}H NMR spectrum of **10a** (600 MHz, DMSO-d6)

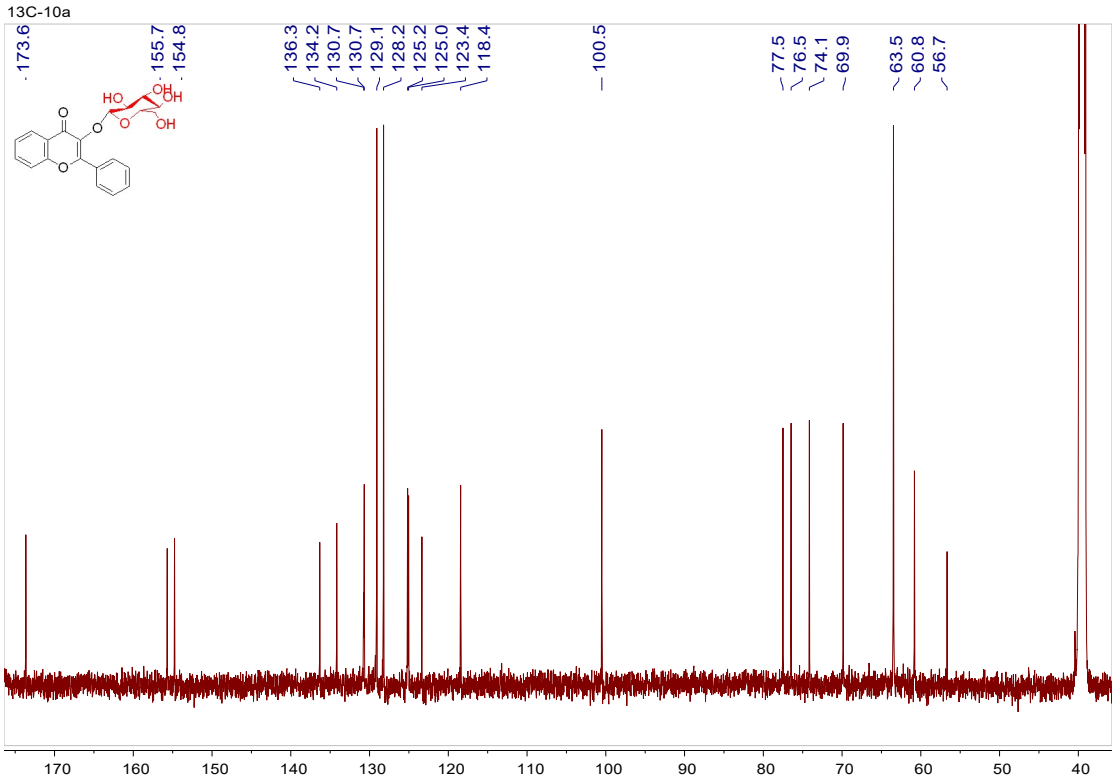


Fig. S18. ^{13}C NMR spectrum of **10a** (151 MHz, DMSO-d6)

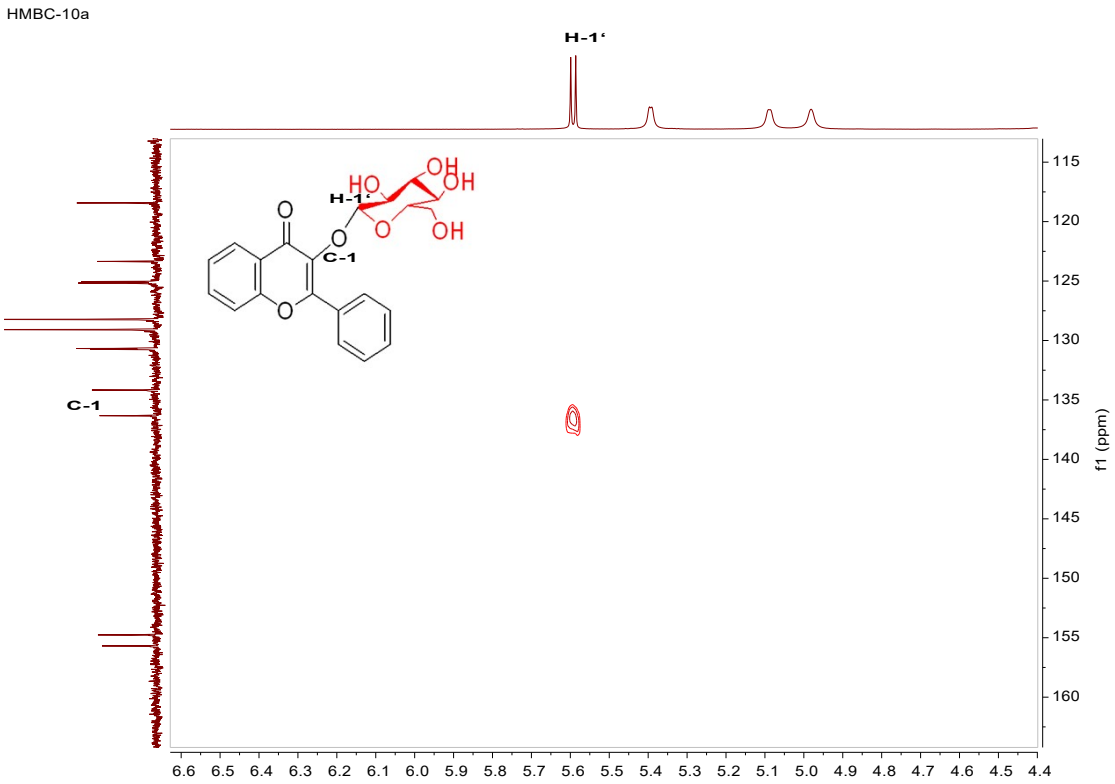


Fig. S19. HMBC spectrum of **10a** (600 MHz/151 MHz, DMSO-d6)

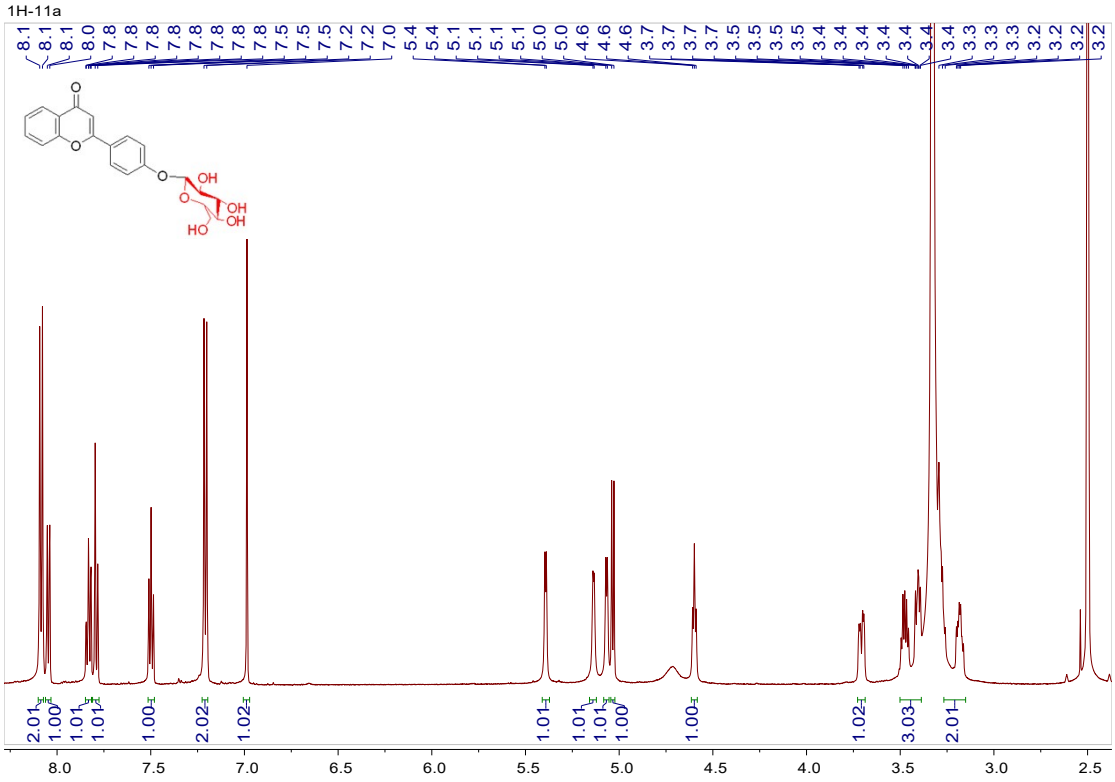


Fig. S20. ^{1}H NMR spectrum of **11a** (600 MHz, DMSO-d6)

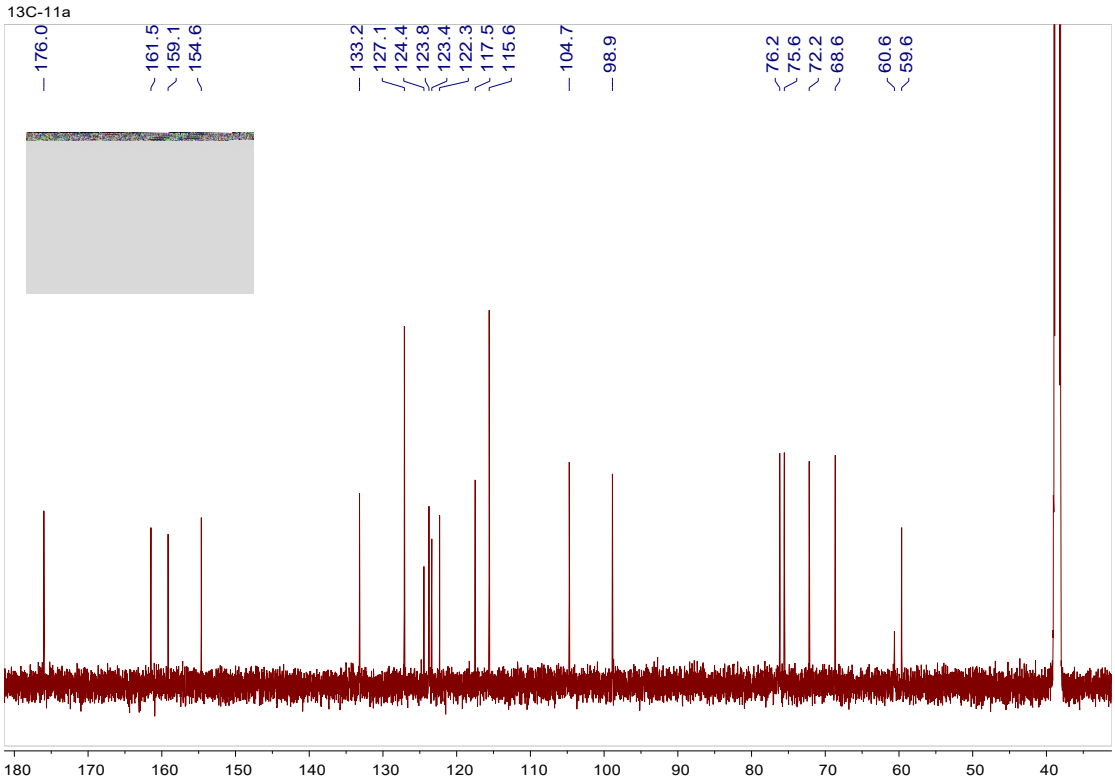


Fig. S21. ^{13}C NMR spectrum of **11a** (151 MHz, DMSO-d6)

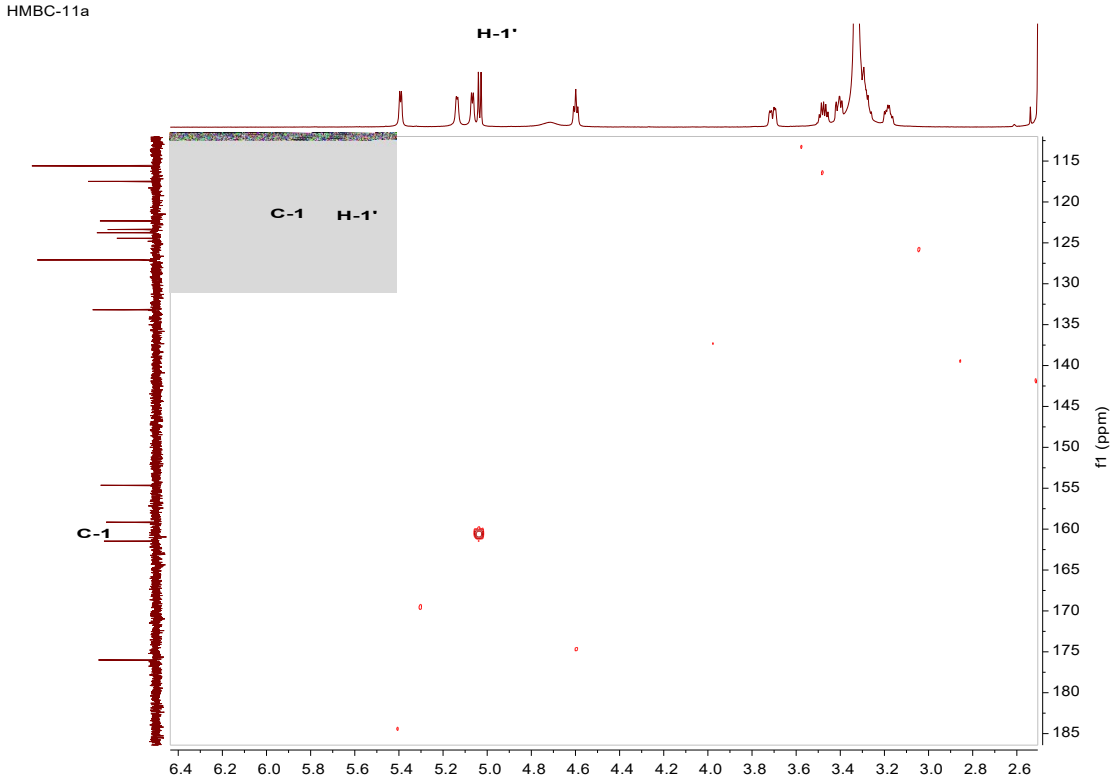


Fig. S22. HMBC spectrum of **11a** (600 MHz/151 MHz, DMSO-d6)

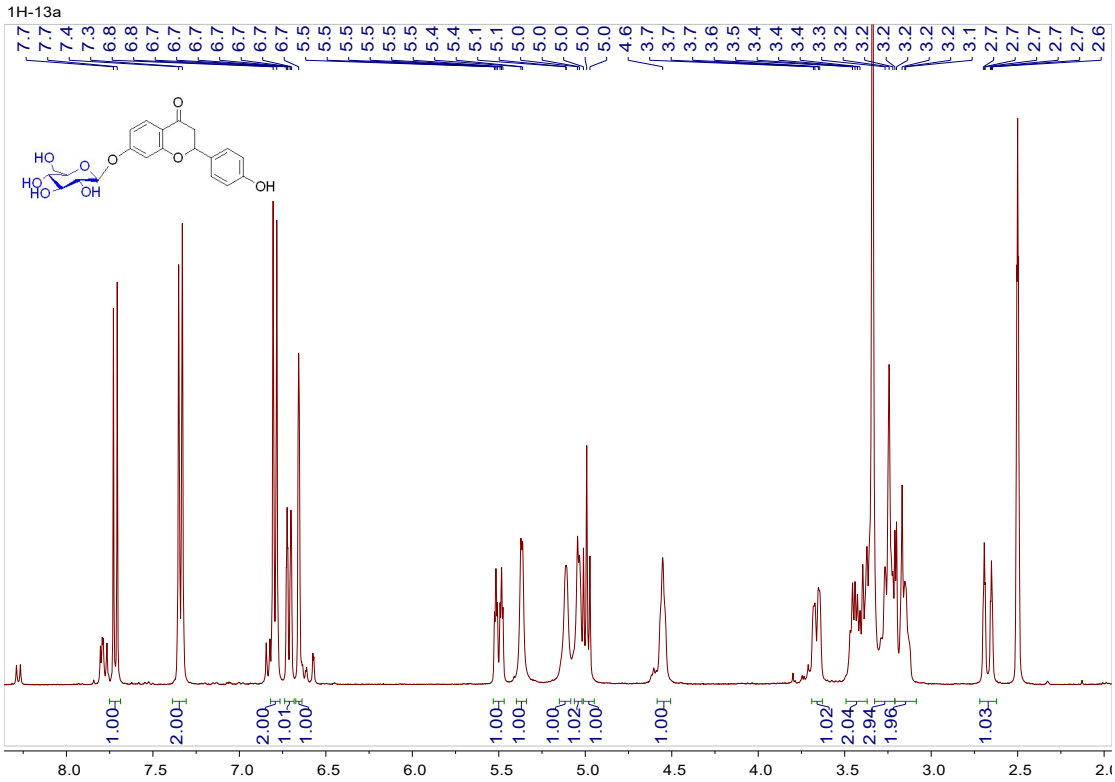


Fig. S23. ^{1}H NMR spectrum of **13a** (400 MHz, DMSO-d6)

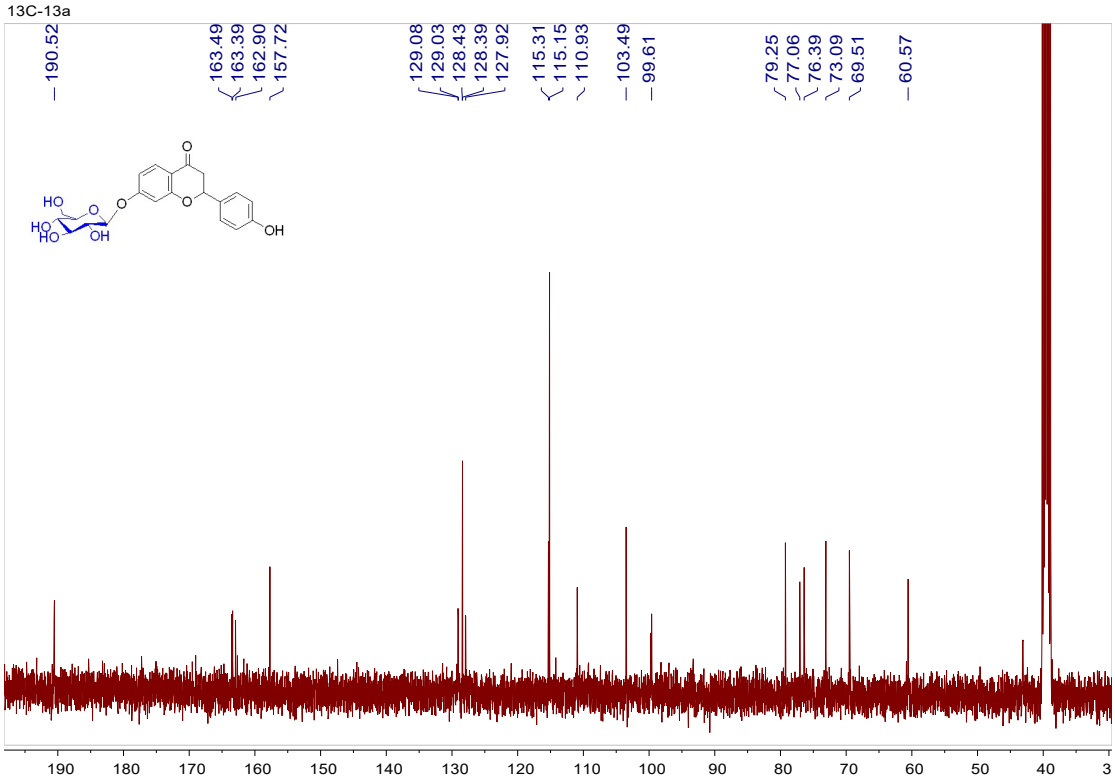


Fig. S24. ^{13}C NMR spectrum of **13a** (101 MHz, DMSO-d6)

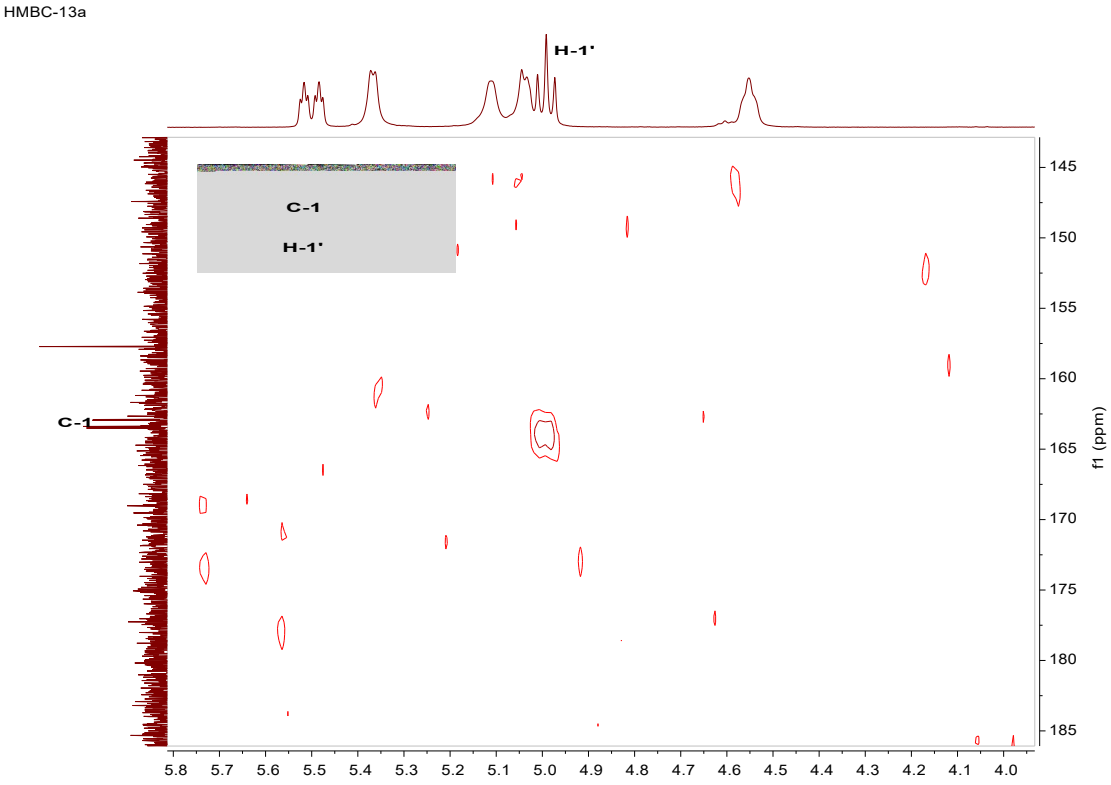


Fig. S25. HMBC spectrum of **13a** (400 MHz/101 MHz, DMSO-d6)

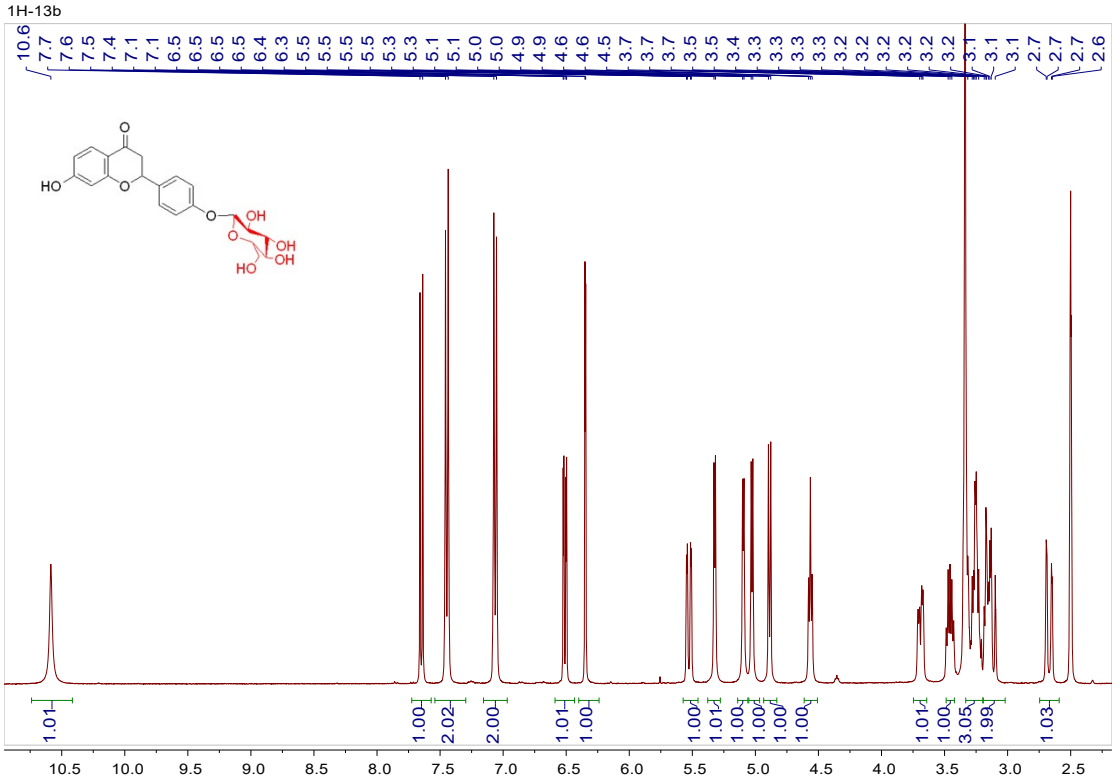


Fig. S26. ^{1}H NMR spectrum of **13b** (400 MHz, DMSO-d6)

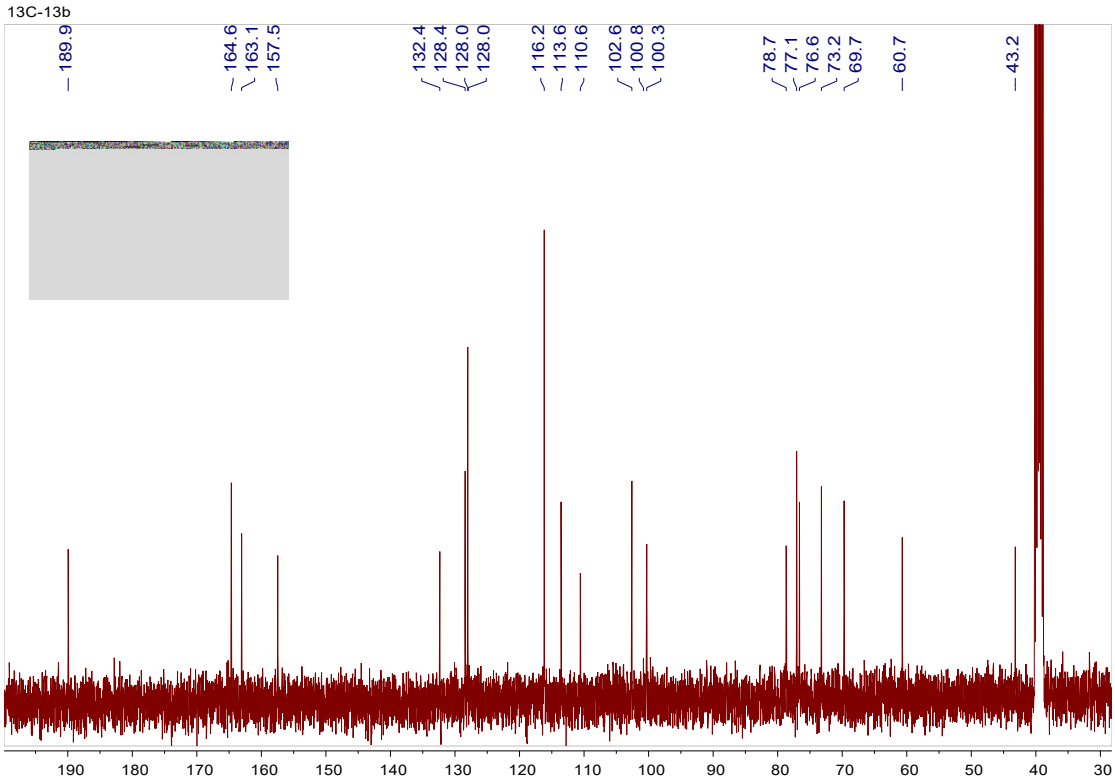


Fig. S27. ^{13}C NMR spectrum of **13b** (101 MHz, DMSO-d6)

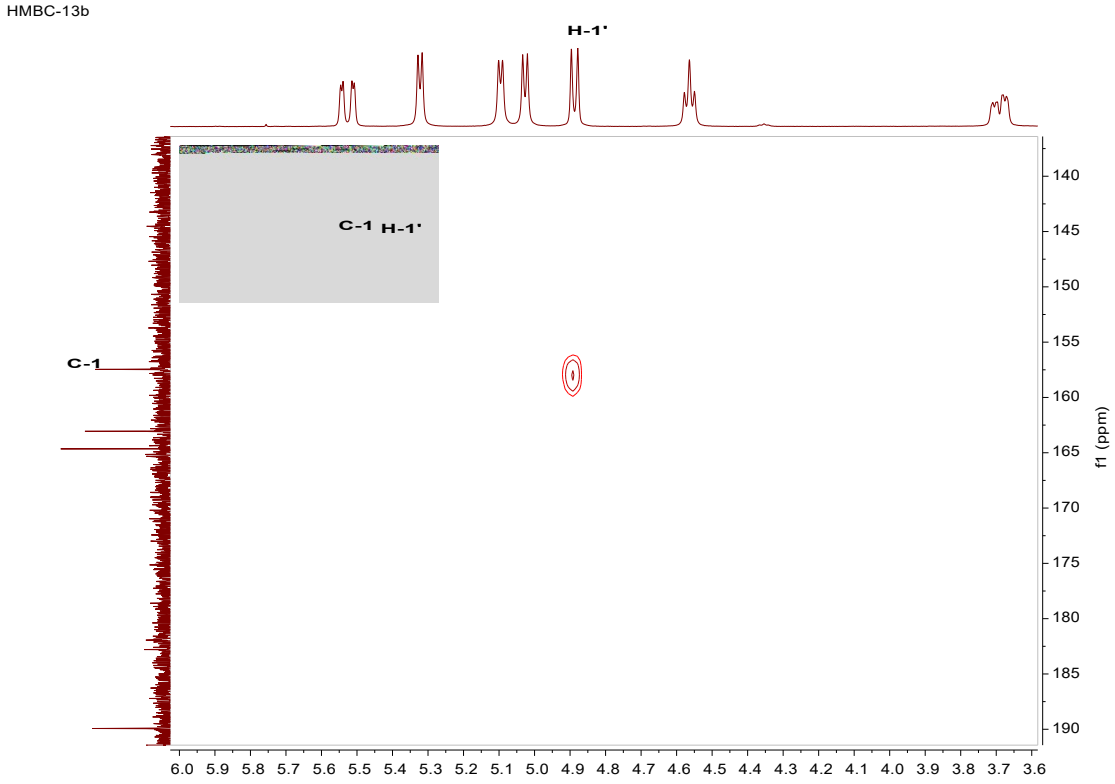


Fig. S28. HMBC spectrum of **13b** (400 MHz/101 MHz, DMSO-d6)

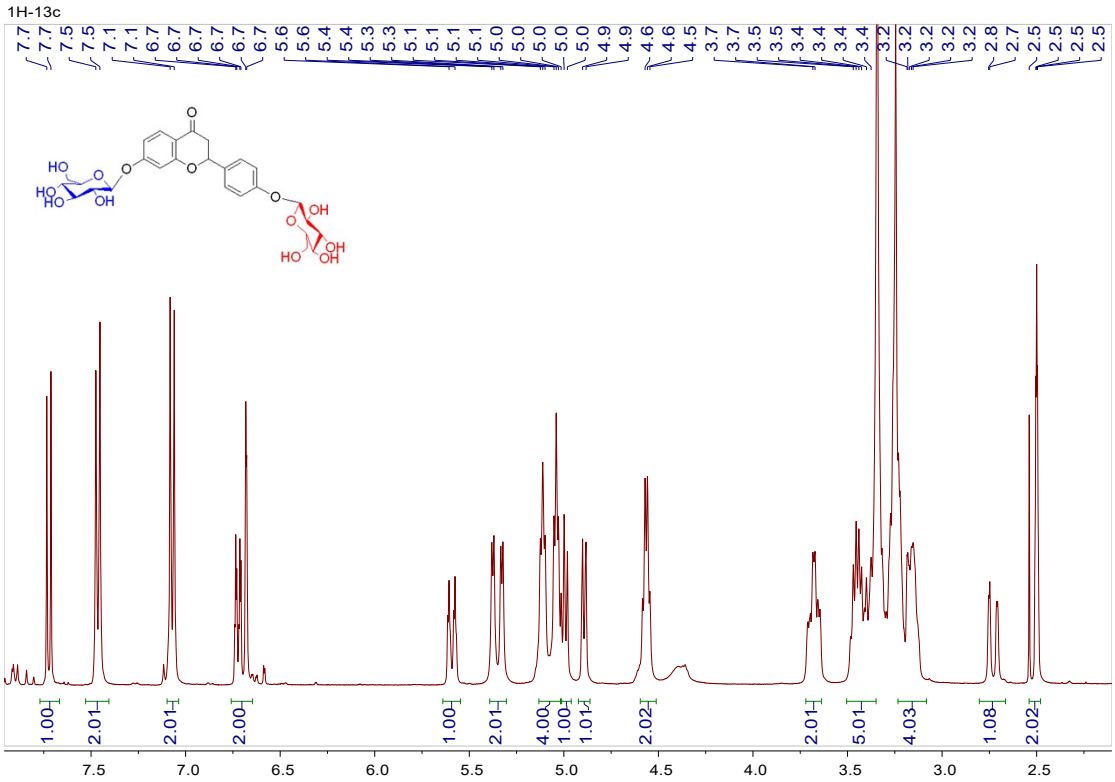


Fig. S29. ^{1}H NMR spectrum of **13c** (400 MHz, DMSO-d6)

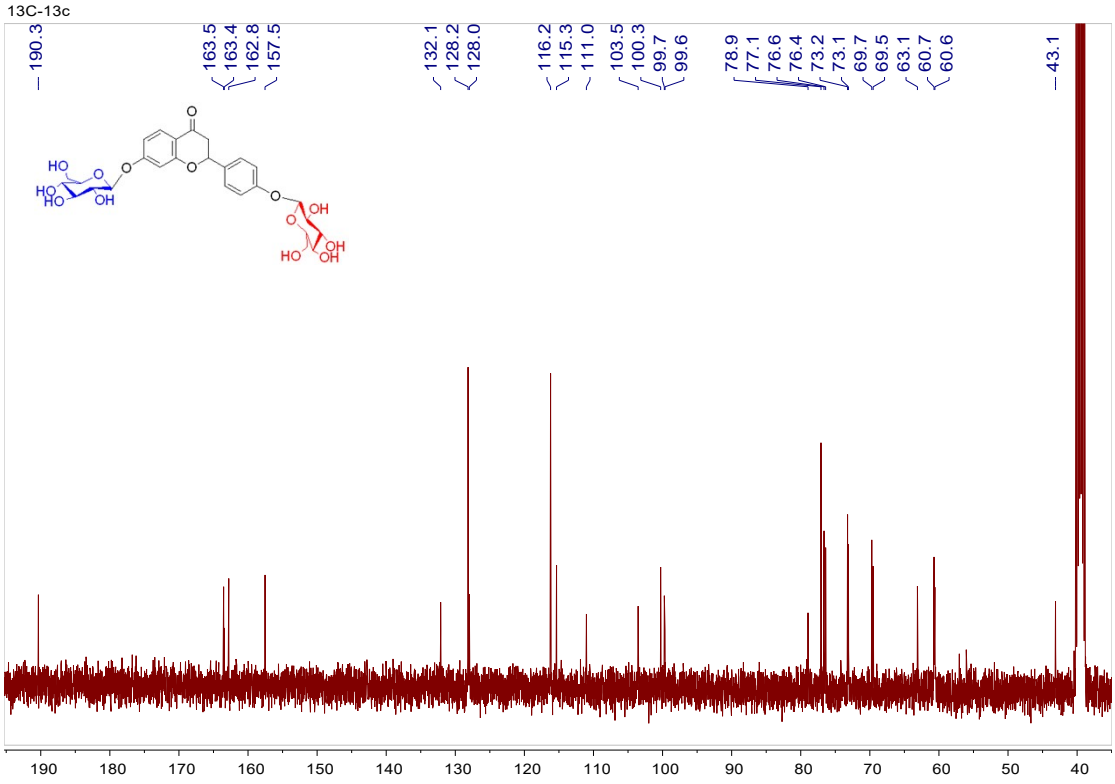


Fig. S30. ^{13}C NMR spectrum of **13c** (101 MHz, DMSO-d6)

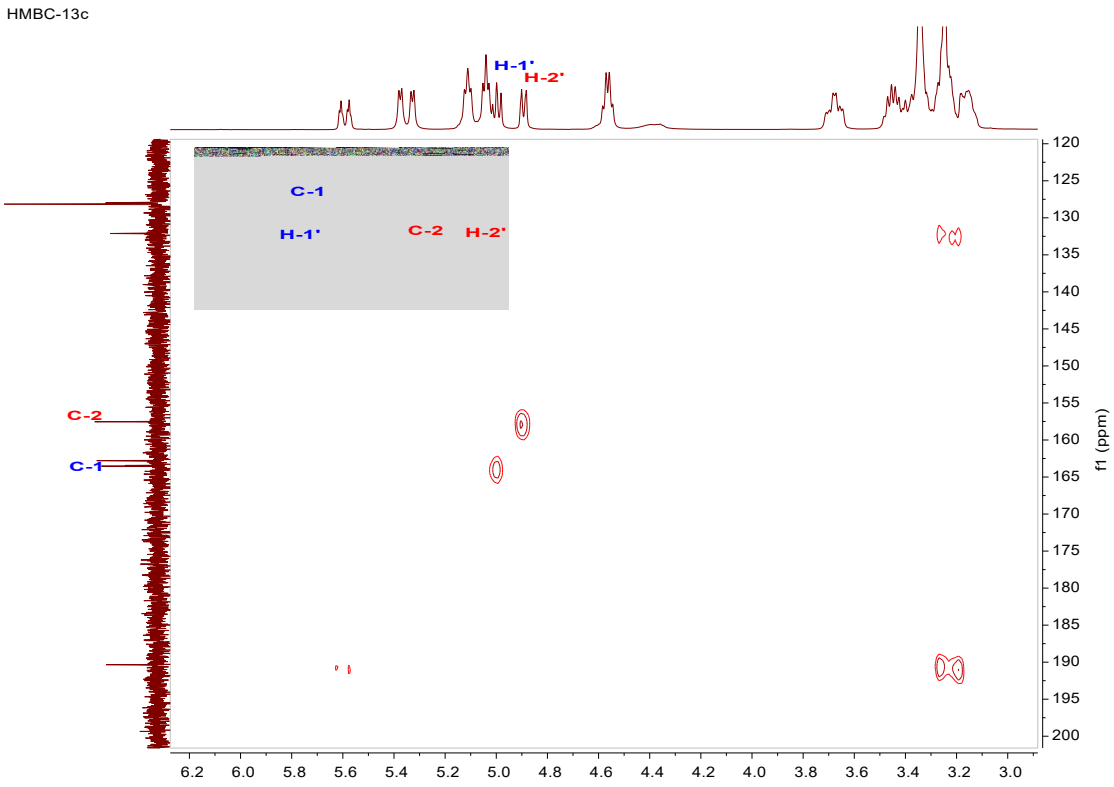


Fig. S31. HMBC spectrum of **13c** (400 MHz/101 MHz, DMSO-d6)

5. References

1. P. Zhang, Y. Ji, S. Meng, Z. Li, D. Hirtz, L. Elling and U. Schwaneberg, *Green Chemistry*, 2023.
2. K. Katoh, J. Rozewicki and K. D. Yamada, *Briefings in bioinformatics*, 2019, **20**, 1160-1166.
3. M. N. Price, P. S. Dehal and A. P. Arkin, *Molecular biology and evolution*, 2009, **26**, 1641-1650.
4. H. Shimodaira and M. Hasegawa, *Molecular biology and evolution*, 1999, **16**, 1114.
5. I. Letunic and P. Bork, *Nucleic acids research*, 2021, **49**, W293-W296.
6. 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, *Nature*, 2021, DOI: 10.1038/s41586-021-03819-2.
7. H. Land and M. S. Humble, *Protein engineering: methods and protocols*, 2018, 43-67.
8. S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845-854.
9. A. L. Justin, *Living Journal of Computational Molecular Science*, 2018, **1**.
10. J. Zielkiewicz, *J Chem Phys*, 2005, **123**, 104501.
11. J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696-3713.
12. P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan and W. Wang, *Accounts of chemical research*, 2000, **33**, 889-897.
13. B. Liu, C. Zhao, Q. Xiang, N. Zhao, Y. Luo and R. Bao, *Int J Biol Macromol*, 2020, DOI: 10.1016/j.ijbiomac.2020.10.238.
14. B. Guo, X. Hou, Y. Zhang, Z. Deng, Q. Ping, K. Fu, Z. Yuan and Y. Rao, *Frontiers in Bioengineering and Biotechnology*, 2022, **10**.