Electronic Supplementary Material (ESI) for

Discovery of *Arabidopsis* UGT73C1 as a steviol-catalyzing UDP-glycosyltransferase with chemical probes

Yiqing Zhou,[a] Weichao Li,[a] Wenjing You,[ab] Zhengao Di,[a] Mingli Wang,[a] Haiyan Zhou,[ab] Shuguang Yuan,[c] Nai-Kei Wong,[d] and Youli Xiao*[ab]

[a.] CAS Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China.

[b.] University of Chinese Academy of Sciences, Beijing 100039, China.

[c.] Laboratory of Physical Chemistry of Polymers and Membranes, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

[d.] State Key Discipline of Infection Diseases, Shenzhen Third People's Hospital, The Second Affiliated Hospital, Shenzhen University, Shenzhen 518112, China.

* Corresponding author

Prof. Youli Xiao
CAS Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China
Phone: 86-21-54924226
ylxiao@sibs.ac.cn

Table of Contents

1. Supplementary Tables and Figures

Supplementary Table S1. Proteins identified from the pull-down/LC-MS experiments.

(Please see accompanying EXCEL file)

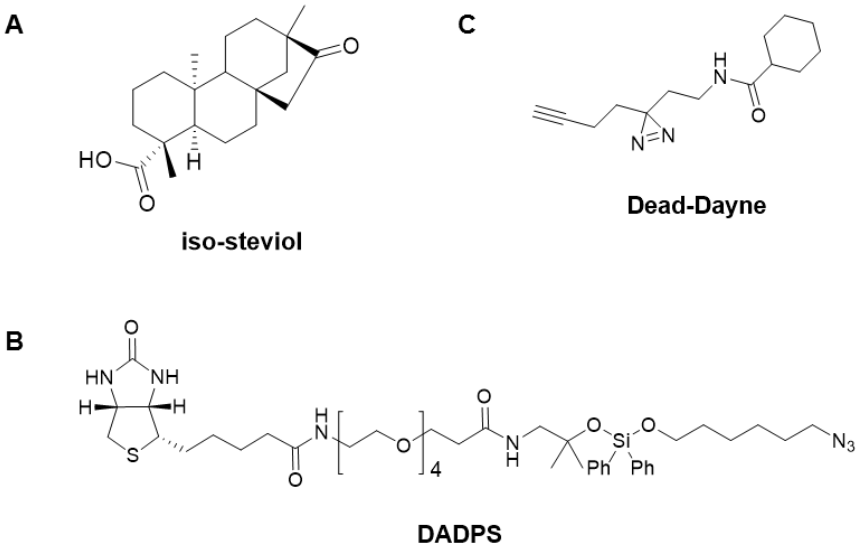


Fig. S1 Chemical structures of (A) iso-steviol; (B) acid-cleavable biotin-azide DADPS[1] and (C) negative control probe Dead-Dayne[2].

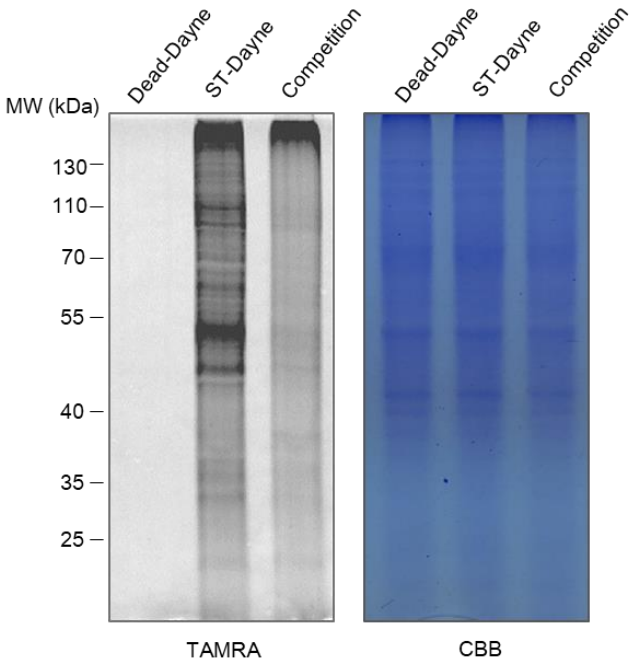


Fig. S2 Photoaffinity labeling of *Arabidopsis* leaves extract by ST-Dayne**:** fluorescence scanning (TAMRA, left) and coomassie blue staining (CBB, right). Detailed experimental procedures were provided in section 2.3 of ESI[†].

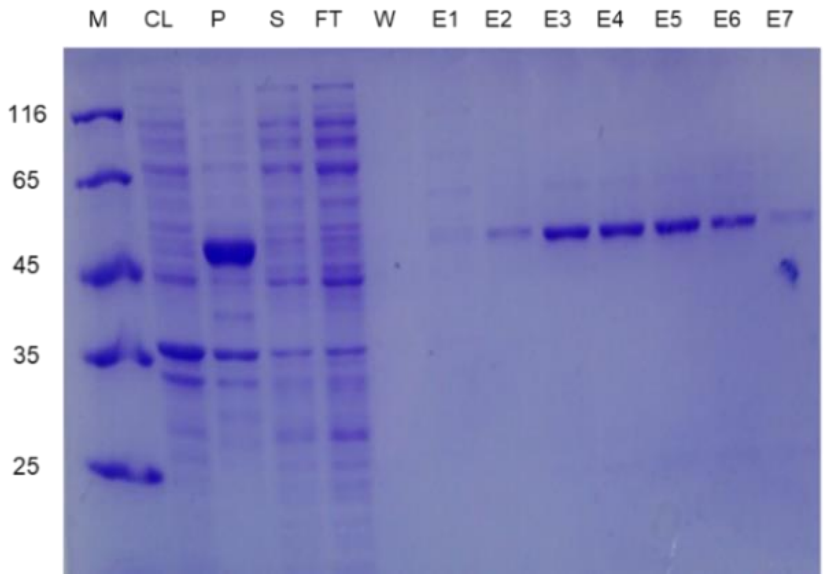


Fig. S3 The purification of *Arabidopsis* UGT73C1. CL: cell lysate; P: precipitation; S: supernatant; FT: flow through fraction; W: eluted fraction by washing buffer; E1 to E7: eluted fraction by 50 mM Tris-HCl, pH 7.4 containing 5 mM, 50 mM, 100 mM, 150 mM, 200 mM, 250 mM, and 300 mM imidazole, respectively.

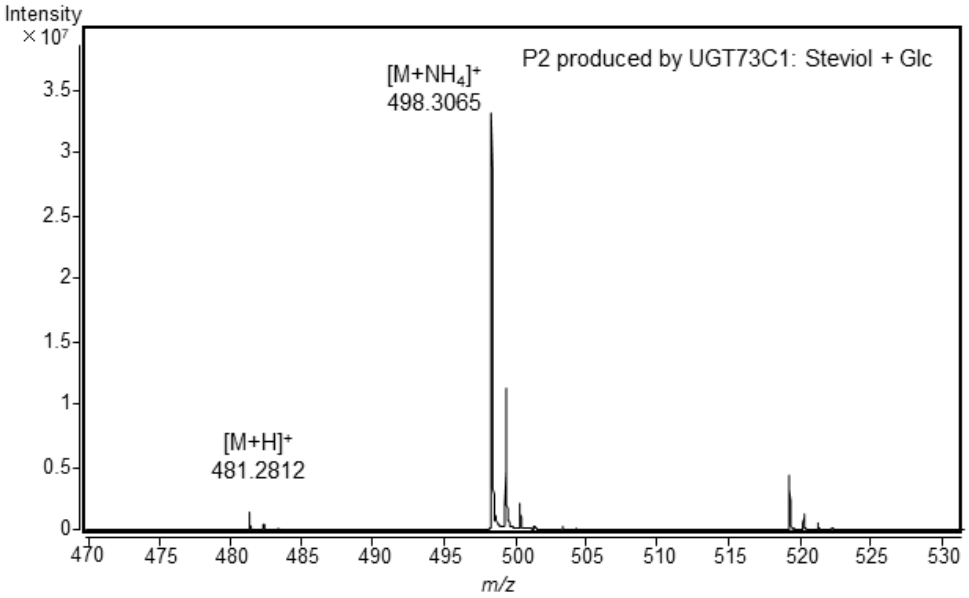
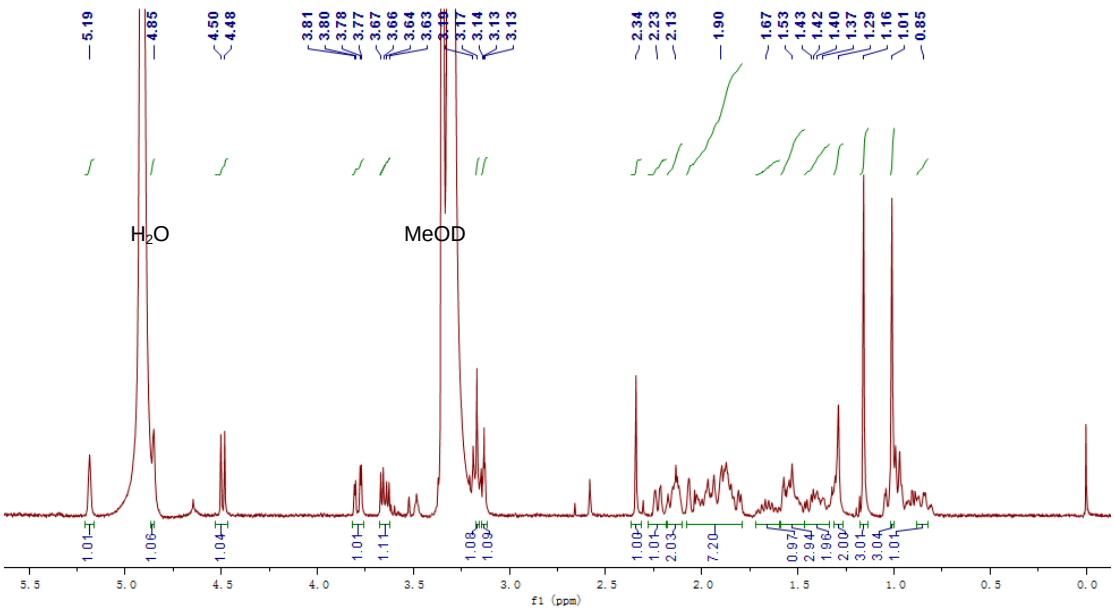


Fig. S4 The mass spectrum of **P2** (Fig. 3B, iii) produced by recombinant *Arabidopsis* UGT73C1 catalyzed glycosyltransferase reaction which suggested one glucose moiety was appended to steviol.



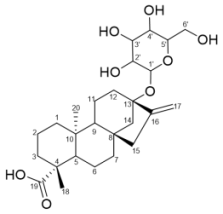
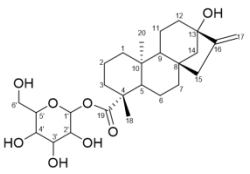
Position		 13-O-β glucopyranosol steviol (steviolmonoside)	 19-O-β glucopyranosol steviol	**P2**
Aglycone	17	4.86 (s, 1H), 5.19 (s, 1H)	4.77 (s, 1H), 4.93 (s, 1H)	4.85 (s, 1H), 5.19 (s, 1H)
	18	1.18 (s, 3H)	1.20 (s, 3H)	1.16 (s, 3H)
	20	0.98 (s, 3H)	0.98 (s, 3H)	1.01 (s, 3H)
Sugar	1’	4.49 (d, 7.6, 1H)	5.41 (d, 8.2, 1H)	4.49 (d, 7.6, 1H)
	2’	3.16 (t, 8.6, 1H)	3.35 (dd, 7.1, 7.6, 1H)	3.14 (t, 8.7, 1H)
	3’	3.34 (t, 8.9, 1H)	3.44 (dd, 8.3, 9.1, 1H)	Buried in the MeOD peak
	4’	3.32 (t, 9.1, 1H)	3.34 (dd, 8.2, 9.4, 1H)	Buried in the MeOD peak
	5’	3.18 (t, 8.7, 1H)	3.39 (ddd, 8.1, 2.1, 7.4, 1H)	3.19 (m, 1H)
	6’	3.64 (dd, 5.5, 12.2, 1H), 3.78 (dd, 2.6, 12.2, 1H)	3.66 (dd, 2.1, 12.1, 1H), 3.81 (dd, 4.2, 12.1, 1H)	3.64 (dd, 4.5, 12.2, 1H), 3.78 (dd, 3.6, 12.2, 1H)

Fig. S5 Upper: ^{1}H NMR spectrum of product **P2** catalyzed by UGT73C1. ^{1}H NMR (400 MHz, MeOD) δ 5.19 (1H, s), 4.85 (1H, s), 4.49 (1H, d, J = 7.6 Hz), 3.78 (1H, dd, J = 3.6, 12.2 Hz), 3.64 (1H, dd, J = 4.5, 12.2 Hz), 3.19 (1H, m), 3.14 (1H, t, J = 8.7 Hz), 2.34 (1H, s), 2.23 (2H, m), 2.13 (3H, m), 1.90 (7H, m), 1.67 (1H, m), 1.53 (3H, m), 1.42 (2H, m), 1.29 (2H, m), 1.16 (3H, s), 1.01 (3H, s).

Lower: Comparison of chemical shift of representative protons of product **P2** with that of 13-O-β glucopyranosol steviol (steviolmonoside)[3] and 19-O-β glucopyranosol steviol[4] reported in previous literatures. The chemical shifts of protons from aglycone C-17 and sugar ring C-1’ colored in red confirmed **P2** as steviolmonoside.

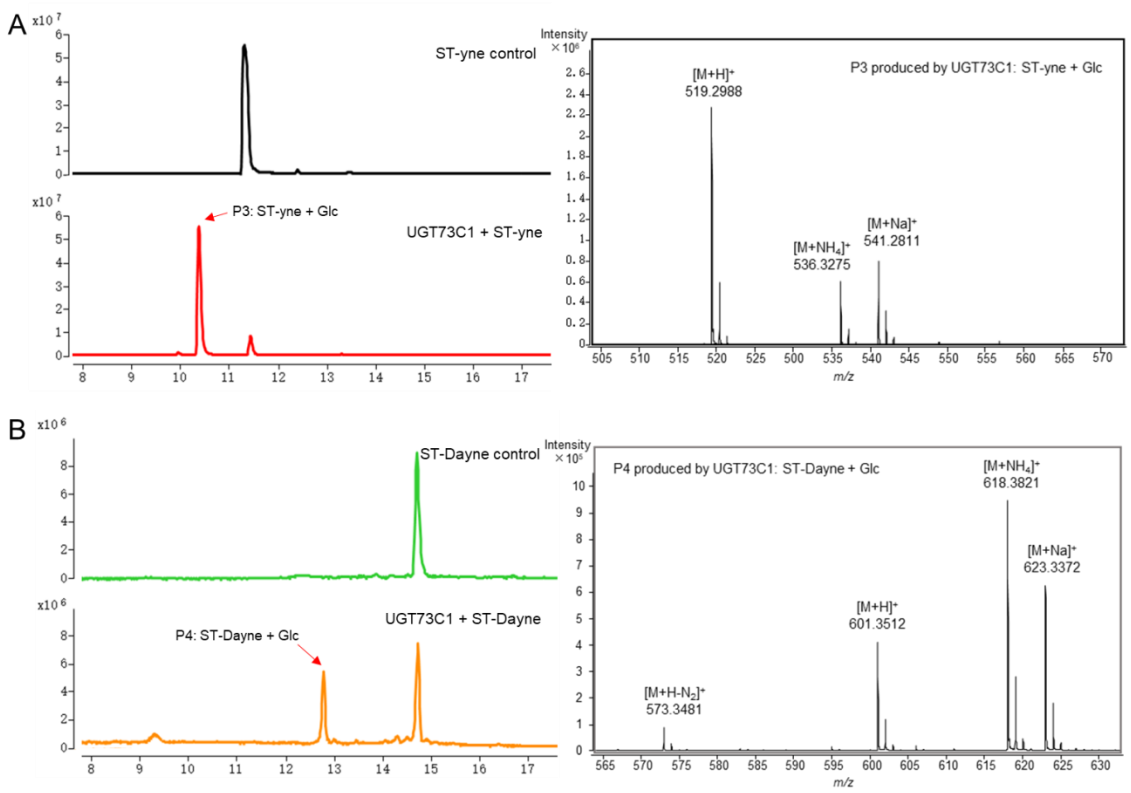
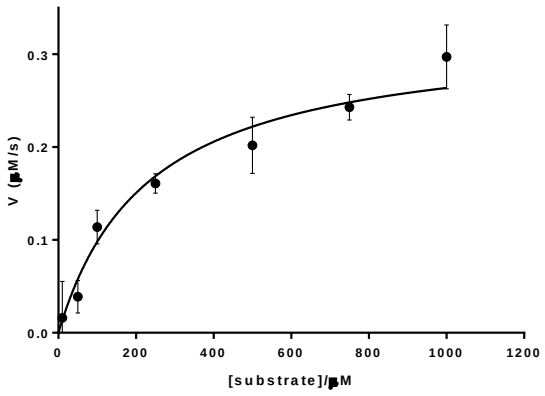


Fig. S6 LC-MS analysis of the glycosylated product of (A) ST-yne (**P3**) and (B) ST-Dayne (**P4**) catalyzed by recombinant *Arabidopsis* UGT73C1 *in vitro*. (left) Extract ion chromatography (EIC) detection; (right) Mass spectrometry result of (A) **P3** and (B) **P4** which suggested one glucose moiety was appended to ST-yne and ST-Dayne, respectively.



K_m (μM)	V_{max} (μM·s^{-1})	K_{cat} (s^{-1})	K_{cat}/K_m (mM^{-1}·s^{-1})
308 ± 91.1	0.360 ± 0.039	1.01	3.28

Fig. S7 Steady state kinetic analysis of UGT73C1 towards steviol. Experimental procedures were provided in section 2.8 of ESI†. Upper: Non-linear data fitting of Michaelis-Menten equation using GraphPad Prism 7.0. Lower: The kinetic parameters of UGT73C1 towards steviol.

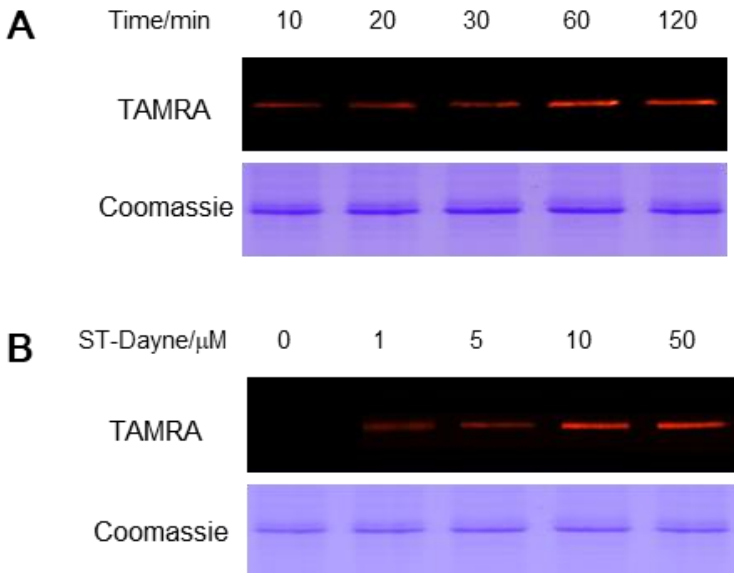


Fig. S8 Photoaffinity labeling of recombinant UGT73C1 by ST-Dayne. (A) UGT73C1 treated with 10 μM of ST-Dayne and irradiated by UV light (365 nm) for different lengths of time; (B) UGT73C1 treated with increasing concentrations of ST-Dayne and irradiated by UV light (365 nm) for 10 min.

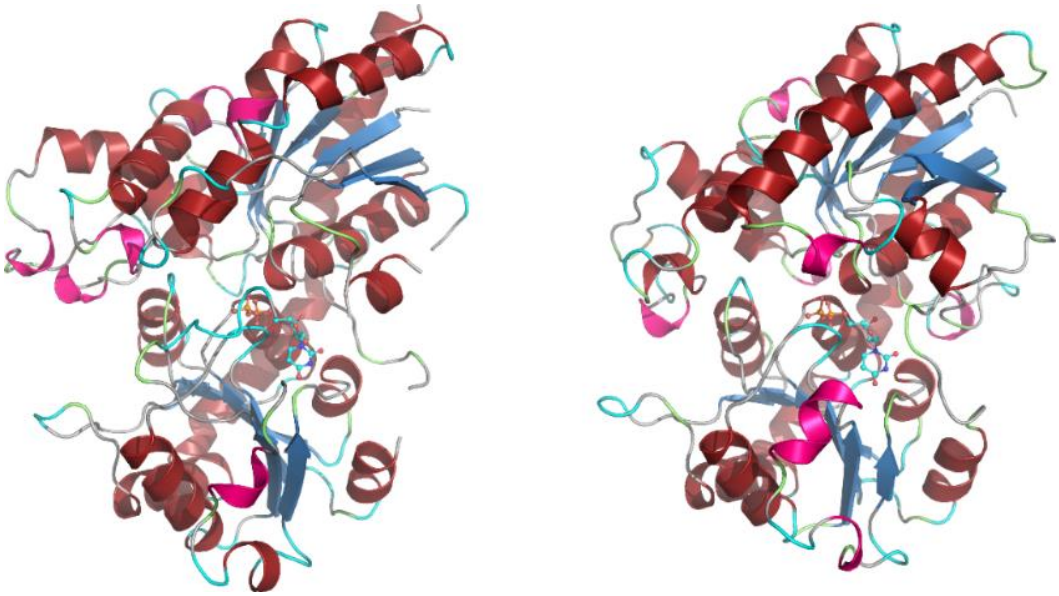


Fig. S9 Template crystal structures used in homology modeling. Left panel: crystal structure of glycosyltransferases UGT72B1 (pdb: 2VCH). Right panel: crystal structure of glycosyltransferases UGT74F2 (pdb: 5V2K). Red cartoon: α-helices. Blue cartoon: β-strands. Cyan ball-and-stick: UDP-glucose molecules.

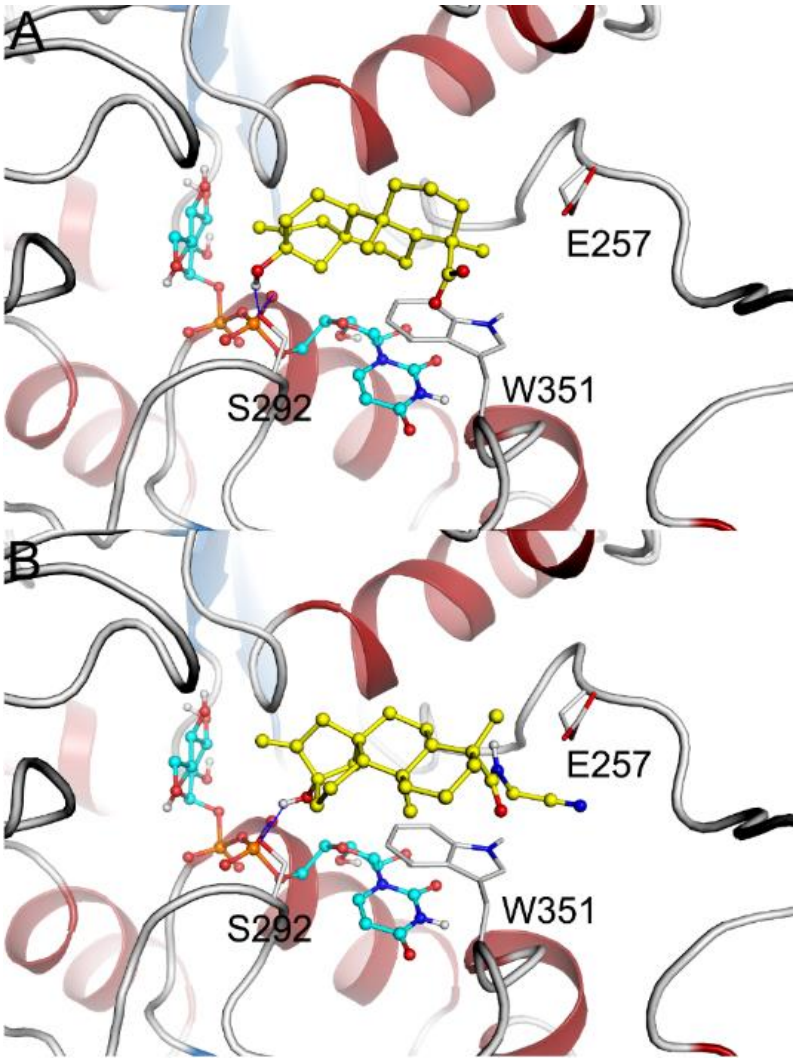


Fig. S10 Binding modes for complex (A) UGT73C1-steviol and (B) UGT73C1-(ST-yne). Blue line: H-bond interactions. Red cartoon: α-helices. Blue cartoon: β-strands. Cyan ball-and-stick: UDP-glucose molecule. Yellow ball-and-stick: substrate molecules.

2. Biological Experiments

2.1 Protein Extraction from *Arabidopsis* leaves

Twenty grams of *Arabidopsis* leaves were homogenized in a blender in liquid nitrogen and transferred to 50 mL buffer A containing 50 mM HEPES pH 7.5, 5.0 mM $MgCl_2$, 1.0 mM EDTA, 10 mM DTT, 10 g PVPP, 10 g PVP, and protease inhibitor cocktail (Roche). Insoluble substance was removed by filtering, followed by centrifuging at 18,000 g for 30 min. Then the soluble materials were precipitated by $(NH_4)_2SO_4$ (30-75% saturation) and collected by centrifuging at 15,000 g for 15 min. The precipitates were dissolved in a minimal volume of 50 mM HEPES pH 7.5, 5.0 mM $MgCl_2$, 1.0 mM EDTA and 10 mM DTT (buffer B), followed by loading onto a DEAE column (1.5 × 20 cm, Aogma Biosciences) equilibrated in buffer B and eluted by a linear gradient from BC30 (buffer B with 30 mM NaCl) to BC250 (buffer B with 250 mM NaCl). All DEAE fractions were combined and concentrated with a 10-kDa molecular weight cut-off column (Millipore) and collected for further enzymatic assays and chemoproteomics studies.

2.2 Metabolites enrichment and LC-MS analysis

Recombinant UGT85C2 (50 μg) or protein extract *of Arabidopsis* leaves (300 μg) was dissolved in a standard glycosyltransferase assay buffer (50 mM potassium phosphate, pH 7.2, 3 mM $MgCl_2$, 10 μg/mL BSA) along with 100 μM of ST-yne. The total reaction volume was 500 μL. For UGT85C2 reaction, isosteviol (100 μM) was added to the solution as internal standard. UDP-glucose (UDPG, 1 mM) was added to initiate the reactions at 30°C for 6 h and terminated by adding 200 μL of water saturated *n*-butanol. The samples were extracted three times with 200 μL of water-saturated *n*-butanol. The pooled *n*-butanol fractions were evaporated completely by a rotavapor and the residue was dissolved in MeOH/H_2O (50%/50%, v/v). The samples were click conjugated to DADPS (200 μM) with $CuSO_4$ (1.0 mM), THPTA (Tris(3-hydroxypropyltriazolylmethyl)amine, 100 μM) and ascorbic sodium (NaVc, 1.0 mM) for 6 hours at room temperature. 50 μL of streptavidin-sepharose (GE Healthcare) were added and incubated at room temperature with continuous rotation for 1 hour. The beads were washed with PBS 5 times, MeOH/H_2O (50%/50%, v/v) 5 times sequentially. Bound metabolites were eluted by 5% HCOOH at 37°C with continuous vortex overnight. The eluent was collected by filtration and the beads were washed with MeOH/H_2O (50%/50%, v/v) 2 times. After combining all eluents and washes, the sample was neutralized by NH_4HCO_3 (100 mM) and evaporated to dryness on a SpeedVac. The dried sample was resuspended in 20 μL of MeOH/H_2O (50%/50%, v/v) with sonication for LC-MS analysis. Positive-ion electrospray mass spectra were obtained by LC-MS by using an Agilent 6520 Accurate-Mass Q-TOF instrument.

2.3 Chemoproteomic profiling of *Arabidopsis* extract with ST-Dayne

Each of 300 μL of *Arabidopsis* extracts (~2.0 mg/mL) were individually incubated with 10 μM Dead-Dayne, 10 μM ST-Dayne and 10 μM ST-Dayne with excess steviol (50 μM) for 1 hour. All samples were photo-irradiated with 365 nm UV light (8W) in a 6-well plate for 10 min on ice. The resulting samples were collected to new tubes and precipitated with CH_3OH (600 μL)/$CHCl_3$ (150 μL)/H_2O (300 μL) sequentially followed by vortex. After centrifuge at 14,000 g for 3 min, the protein disk was washed twice with CH_3OH (500 μL), air-dried and re-dissolved in 200 μL of click buffer (50 mM HEPES pH 8.0, 1% SDS) by sonication. The precipitates were re-suspended in 220 μL of click buffer (50 mM HEPES pH 8.0, 1% SDS). 20 μL of samples were subjected to click reaction with TAMRA-N_3 (100 μM, Lumiprobe), $CuSO_4$ (1.0 mM), THPTA (100 μM) and NaVc (1.0 mM), followed by in-gel fluorescence scanning by FUJIFILM FLA 9000 plus DAGE fluorescence scanner; the rest samples (200 μL each) were subjected to click reaction with Biotin-N_3 (100 μM, Biomatrick Inc.), $CuSO_4$ (1.0 mM), THPTA (100 μM) and NaVc (1.0 mM). The samples were precipitated again with CH_3OH/$CHCl_3$/H_2O and re-dissolved in 200 μL of binding buffer (50 mM HEPES pH 8.0, 0.5% SDS) by sonication. 50 μL of streptavidin-sepharose (GE Healthcare) beads were added to each sample and incubated at room temperature with continuous rotation for 1 hour. The beads were washed sequentially with PBS with 0.5% SDS (w/v) three times, 4 M Urea in 50 mM trimethylamine bicarbonate (TEAB) twice, PBS once and 50 mM TEAB five times. Each wash was performed on a rotator for 15 min. The bounded proteins were subjected to on-beads reductive alkylation with 200 μL of 10 mM DTT at 56°C for 30 min and 200 μL of 55 mM iodoacetamide at 37°C in dark for another 30 min, followed by wash with 100 mM TEAB three times. Bounded proteins were digested with 0.25 μg of sequencing grade modified trypsin (Promega) reconstituted in 50 μL of 100 mM TEAB overnight at 37°C. The digests were labeled with respective tandem mass tagging (TMT-^{2}plex) Isobaric Label Reagent (Thermo Scientific) according to the manufacturer's procedures. The digested Dead-Dayne pull-down sample (negative group) was labeled with TMT reagent 126, the digested ST-Dayne pull-down sample (positive group) was labeled with reagent 127 and the competitive pull-down sample (ST-Dayne with excess steviol, competition group) was also labeled with reagent 126. Combined peptides (negative with positive groups; positive with competition groups) were desalted by Pierce C18 spin columns and evaporated to dryness on a SpeedVac. Dried peptides were suspended in 10 μL of ddH_2O containing 0.1% formic acid with sonication. The proteomics experiment was carried out in biological duplicates.

2.4 Mass spectrometry

After filtration through 22 μm membrane the clear solution was subjected to nano LC−MS/MS separation. A volume of 3.0 μL of each sample was desalted by loading on a Thermo C18 PepMap100 precolumn (300 μm × 5 mm) and eluted on a Thermo Acclaim PepMap RSLC analytical column (75 μm × 15 cm). Mobile phase A (0.1% formic acid in H_2O) and mobile phase B (0.1% formic acid in acetonitrile) were used to establish the 120 min gradient comprised of 85 min of 4−30% B, 15 min of 30−50% B, and 5 min of 90% B, followed by re-equilibrating at 4% B for 15 min. The flow rate was 0.3 μL/min. Peptides were then analyzed on a Q−Exactive proteomic mass spectrometer (Thermo Scientific) in

a data−dependent manner, with automatic switching between MS and MS/MS scans using a top 20 method. MS spectra were acquired at a resolution of 70000 with a target value of 3×10^6 ions or a maximum integration time of 50 ms. The scan range was limited from 375 to 1400 *m/z*. Peptide fragmentation was performed via higher-energy collision dissociation (HCD) with the energy set at 32 NCE. The MS/MS spectra were acquired at a resolution of 35000 with a target value of 1×10^5 ions or a maximum integration time of 100 ms. The fixed first *m/z* was 100, and the isolation window was 1.2 *m/z*.

2.5 Data process

Protein identification and TMT quantification were performed using Proteome Discoverer 2.1 software (Thermo Scientific). Peptide sequences (and hence protein identity) were determined by matching protein databases (Uniprot) with the acquired fragmentation pattern by SEQUEST HT algorithm. The precursor mass tolerance was set to 10 ppm and fragment ion mass tolerance to 0.02 Da. One missed cleavage site of trypsin was allowed. Carbamidomethyl (C) and TMT-^{2}plex (K and N-terminal) were used as a fixed modification. Oxidation (M) was used as variable modifications. All spectra were searched against protein database using a target false discovery rate (FDR) of 1%. Identified proteins were additionally filtered by at least three spectral counts and one unique peptides in each experimental replicate. Protein ratios were calculated as the median of all peptide hits belonging to a protein.

Statistical analysis was performed with Perseus 1.5.1.6 as described in previous literature.[5] In brief, TMT ratios obtained from Proteome Discoverer 2.1 were transformed with $\log_{2}$-(x) and then normalized using *Z*-score, and -$\log_{10}$-(*p*-value) of all proteins were obtained by a two-sided one sample t-test over two biological replicates. Only proteins identified have average $\log_{2}$- ratios > 1.0 and *p*-values < 0.05 (-$\log_{10}$-(*p*-value) > 1.3) were considered statistical significant targets. In the present study, only proteins whose description contain the term “glycosyltransferase” were selected for further studies.

2.6 Recombinant expression and purification of UGT73C1

UGT73C1 gene was a kind gift from Prof. Bingkai Hou of Shandong University. The gene was cloned into pET28a expression vectors. Expression was induced at an OD_{600} of 0.6 by addition of IPTG (final concentration 0.5 mM) and carried out 20 hours at 16°C in *E. coli* BL21-(DE3) cells. The cells were harvested, resuspended in buffer (25 mM Tris-HCl, pH 8.0; 150 mM NaCl) and lysed by sonication. The lysate was clarified by centrifugation (35,000 *g*, 30 min, and 4°C). The clarified lysate was subjected to His-affinity chromatography using Ni-NTA resin with a 10 CV (column volume) washing step (25 mM Tris-HCl, pH 8.0; 150 mM NaCl; 20 mM imidazole) and a 5 CV elution step (25 mM Tris-HCl, pH 8.0; 150 mM NaCl; 50 ~ 250 mM imidazole). Protein identity was confirmed by SDS-PAGE. Protein concentration was determined by the Bradford assay using bovine serum albumin as the standard.

2.7 *In vitro* glycosyltransferase assay of UGT73C1

The reaction was carried out in a total volume of 100 μL containing 100 mM Tris-HCl buffer (pH 8.0), 3 mM $MgCl_2$, 10 μg/mL BSA, 1 mM UDP-glucose along with 100 μM substrate (steviol, ST-yne, or ST-Dayne) and 100 μM recombinant UGT73C1. Reaction was performed for 4 h at 30°C and terminated by adding 200 μL of water saturated *n*-butanol. The sample was extracted three times with 200 μL of water-saturated *n*-butanol. The combined organic phase was evaporated completely by a rotavapor and the residue was dissolved in MeOH/H_2O (50%/50%, v/v) for LC-MS analysis by using an Agilent 6520 Accurate-Mass Q-TOF instrument.

2.8 Kinetics analysis of the activity of UGT73C1

The kinetic parameters were determined through the release of UDP, which can be measured using a coupled assay containing UGT, pyruvate kinase, and lactate dehydrogenase as previously reported.[6] A saturating concentration of UDPG at 5.0 mM was used to measure the kinetics parameters towards steviol. The reaction mixture (total volume of 200 μL) containing 50 mM HEPES pH 7.8, 2.5 mM $MgSO_4$, 5.0 mM UDPG, 10 mM KCl, 0.15 mM NADH, 2.0 mM phosphoenol pyruvate, 3.0 units of pyruvate kinase, and 4.0 units of lactate dehydrogenase was warmed to 30°C. The reaction was initiated by adding 4.0 μL of recombinant UGT73C1 solution (1.0 μg/μL in 50 mM HEPES, pH 7.8) at 30°C. The coupled enzyme assay was analyzed over the range of 0 to 1 mM steviol. The change of NAD^+ was monitored at 340 nm in a 0.2 mm light length quart cuvette with an Agilent Cary 60 UV-visible spectrometer. The reaction rate was converted to the unit millikatals kg^{-1} using the extinction coefficient 6.22×10^3 $M^{-1}\cdot cm^{-1}$ for NADH. The data were fit using GraphPad Prism 7.0 and the results shown represent the mean of three independent experiments ± S.D.

2.9 *In vitro* labeling of UGT73C1 with ST-Dayne

Recombinant His-tagged UGT73C1 protein was diluted to 0.1 mg/mL in PBS. For each sample, 20 μL of UGT73C1 solution was incubated with ST-Dayne (10 μM) in the presence or absence of free steviol as indicated at room temperature. Samples were irradiated with 365 nm UV light (8 Watt) on ice for 20 min and added with 1% SDS. The following click reaction and in-gel fluorescence scanning were performed as described above.

2.10 Thermal shift assay

Recombinant His-tagged UGT73C1 proteins (20 μL, 0.1 mg/mL in PBS) were individually exposed to 10 μM steviol, 10 μM UDP-glucose, or DMSO for 1 h at room temperature. The samples were aliquot into PCR tubes and heated individually at temperatures of 37°C, 47°C, 57°C, and 67°C, respectively, for 3 min followed by immediate cooling on ice. The resultant proteins were centrifuged at 18,000 g for 30 min at 4°C to pellet any precipitated

proteins. The soluble proteins in the supernatants were analyzed by SDS-PAGE followed by immuno-blotting using Anti-6×His rabbit polyclonal antibody (Sangon Biotech).

2.11 Binding site identification

5.0 μg of recombinant His-tagged UGT73C1 protein in 50 mM Tris-HCl, pH 7.4 (20 μL) were incubated with 10 μM ST-Dayne at room temperature for 1 h, followed by UV irradiation (365 nm, 8 watt) for 10 min on ice. Buffer exchange (to 25 mM NH_4HCO_3), chemical modification (reduction with 20 mM DTT and alkylation with 50 mM iodoacetamide in 25 mM NH_4HCO_3) were performed according to filter-aided sample preparation (FASP) protocol in the upper chamber of 10-kDa ultrafiltration device.[7] Then, 0.25 μg of trypsin (Thermo Scientific) was add to the sample and incubated at 37°C overnight. The digests were desalted by Ziptip desalting column (Pierce) and evaporated to dryness on a SpeedVac. The dried peptides were suspended in 8 μL ddH_2O containing 0.1% formic acid with sonication and analyzed by Thermo Orbitrap Fusion Lumos proteomic mass spectrometer as mentioned above. Data processing was performed using Proteome Discoverer 2.1 software (Thermo Scientific) and peptide sequences were determined by matching protein database with the acquired fragmentation pattern by SEQUEST HT algorithm. The precursor mass tolerance was set to 10 ppm and fragment ion mass tolerance to 0.02 Da. One missed cleavage site of trypsin was allowed. ST-Dayne (any amino acids), Carbamidomethyl (C), Oxidation (M), were used as variable modifications. All spectra were searched against protein sequence of *Arabidopsis* UGT73C1 (UniprotKB ID: Q9ZQ99) using a target false discovery rate (FDR) of 1%. Manual verification was performed to ensure confident peptide identification.

2.12 Modeling

Homology modelling of UGT73C1 was performed in Modeller V9.10 using multi-template methods. The high resolution crystal structures of glycosyltransferases UGT72B1 (pdb: 2VCH) and UGT74F2 (pdb: 5V2K) were used as templates. A total of 10,000 models were generated and a conformation with the lowest DOPE (Discrete Optimized Protein Energy) score was chosen for further study. We introduced a 300 ns molecular dynamics (MD) simulation to refine the initial homology models. MD simulations were performed in Gromacs 5.1.4. All amino acid residues of the protein were modeled according to their protonation state at neutral pH. The protein was centered in a water box with a distance of 12 Å away from the protein. The total number of atoms was approximately 116, 000: 83 Na^+ and 94 Cl^- ions, and about 36, 000 water molecules. CHARMM36m force field was assigned to the protein, water and ions, while the ligands were treated by CHARMM CgenFF force field. The ligands were submitted to GAUSSIAN 09 program for structure optimization at Hartree-Fock 6-31G* level prior to the generation of force field parameters. All bond lengths of hydrogen atoms in the system were constrained using M-SHAKE. Van der Waals and short-range electrostatic interactions were cut off at 10 Å. The whole system was heated linearly at constant volume (NVT ensemble) from 0 to 310 K over 300 ps. Ten nanoseconds equilibration was performed at constant pressure and temperature

(NPT ensemble; 310 K, 1 bar) using the Nose-Hoover coupling scheme with two temperature groups. Long-range electrostatic interactions were computed by particle mesh Ewald (PME) summation. Finally, a 300 ns MD simulation with a time step of 2.0 fs were performed for UGT73C1. A small force constant (1 kcal/mol/Å) was added to the CA atom of the protein to keep the secondary structure. The last snapshot structure of UGT73C1 from MD simulations was prepared in Schrodinger suite software under OPLS3 force field. Hydrogen atoms were added to repaired crystal structures at pH 7.0 with the PROPKA tool in Protein Preparation tool in Maestro to optimize the hydrogen bond network. Constrained energy minimizations were conducted on the full-atomic models, with heavy atom coverage to 0.5 Å. All ligand structures were built in Schrodinger Maestro software. The LigPrep module in Schrodinger software was introduced for geometric optimization by using OPLS3 force field. The ionization state of ligands were calculated with Epik tool employing Hammett and Taft methods in conjunction with ionization and tautomerization tools. The docking of a ligand to the receptor was performed using Glide. The UDP-glucose molecules were always included in the docking for all cases. Cubic boxes centered on the ligand mass center with a radius 8 Å for all ligands defined the docking binding regions. Flexible ligand docking was executed for all structures. Ten poses per ligand out of 20,000 were included in the post-docking energy minimization. The best scored pose for the ligand was chosen for further study. Figures are prepared in PyMOL and Inkscape.

3. Chemistry

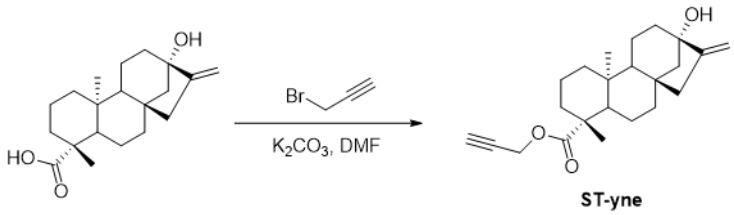
Materials

All the reagents were purchased commercially and used without further purification. Anhydrous dimethyl formamide (DMF) was distilled from calcium hydride. Brine refers to a saturated solution of sodium chloride in distilled water.

Reactions were monitored by thin-layer chromatography (TLC) carried out on 0.25 mm Huanghai silica gel plates (HSGF254) using UV light as visualizing agent . Flash column chromatography was carried out using Yantai Xinnuo silica.

^{1}H-NMR and ^{13}C-NMR spectra were recorded on a Bruker Advance ^{III} 500 (^{1}H: 500 MHz, ^{13}C: 125 MHz) and a Bruker Ultrashield™ 400 PLUS (^{1}H: 400 MHz, ^{13}C: 100 MHz) with chemical shift values in ppm relative to TMS (δ_H 0.00 and δ_C 0.00) and residual D-chloroform (δ_H 7.26 and δ_C 77.16) as standard. HR-MS were obtained using an Agilent 6520 Accurate-Mass Q-TOF instrument.

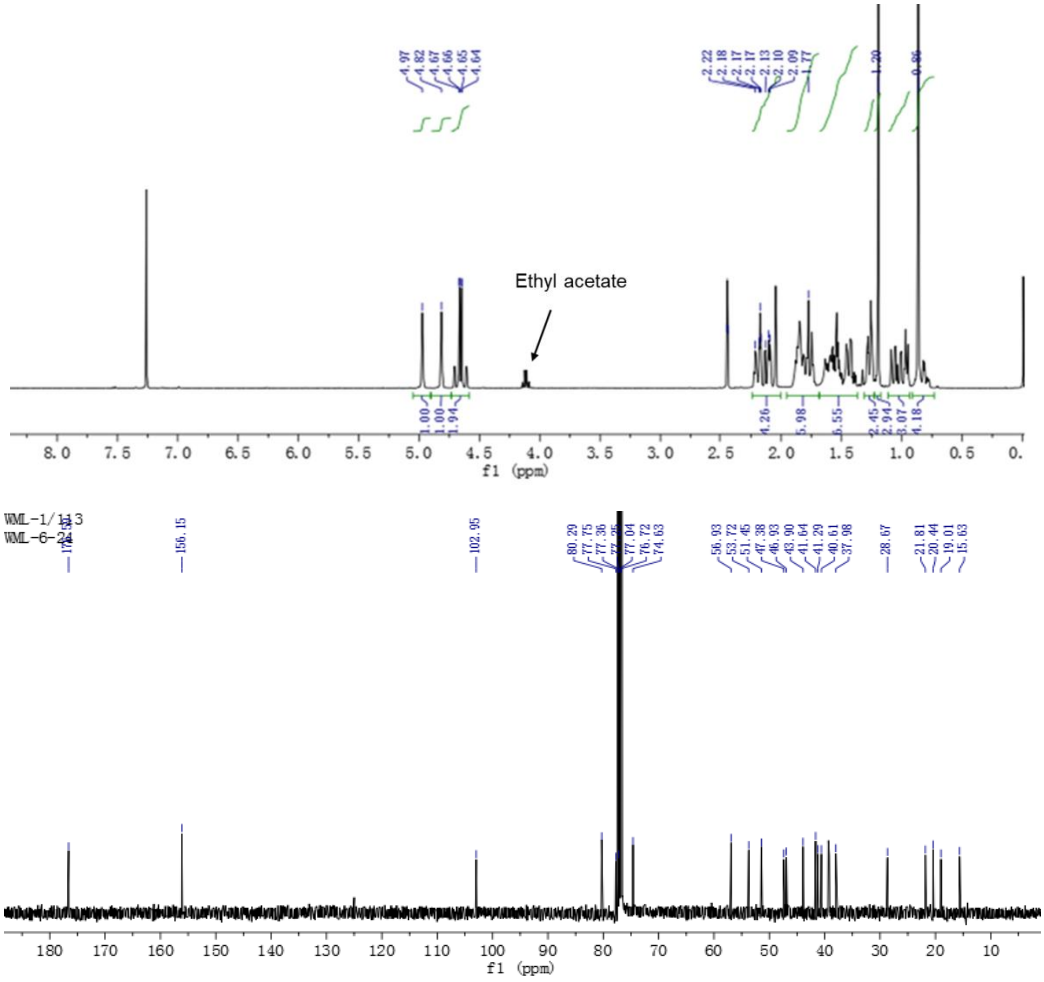
3.1 Synthesis of ST-yne



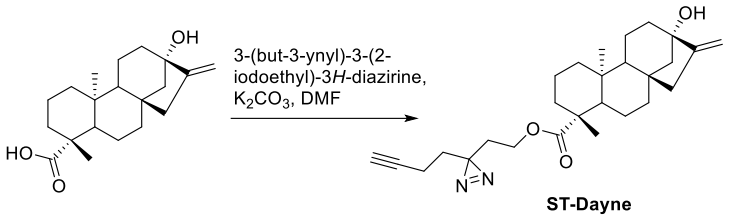
ST-yne

Steviol (32 mg, 0.09 mmol, 1 eq) was dissolved in 2 mL dry DMF and cooled to 0°C with ice bath. Propargyl bromide (16 mg, 0.13 mmol, 1.3 eq) and K_2CO_3 (30 mg, 0.2 mmol, 2 eq) were added to the reaction. The resulting solution was heated with stirring at 70°C for 6 hours. DMF was removed under vacuum and the residue was dissolved in EtOAc, washed by 1N HCl and brine, dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by flash chromatograph column to provide **ST-yne** (28 mg, 75%) as white solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ 4.97 (1H, s), 4.82 (1H, s), 4.65 (2H, td), 2.21 (2H, m), 2.18 (2H, m), 1.78 (6H, m), 1.58 (7H, m), 1.46 (2H, m), 1.22 (3H, s), 1.05 (3H, m), 0.85 (3H, s). ^{13}C NMR (100 MHz, $CDCl_3$) δ 177.50, 156.15, 102.95, 80.29, 76.72, 74.63, 56.93, 53.72, 51.45, 47.38, 46.93, 43.90, 41.64, 41.29, 40.61, 39.31, 39.24, 37.98, 28.67, 21.81, 20.44, 19.01, 15.63. HRMS-ESI *calc'd. for* $C_{23}H_{36}NO_3$ $[M+NH_4]^+$: 374.2690; Found: 374.2682.



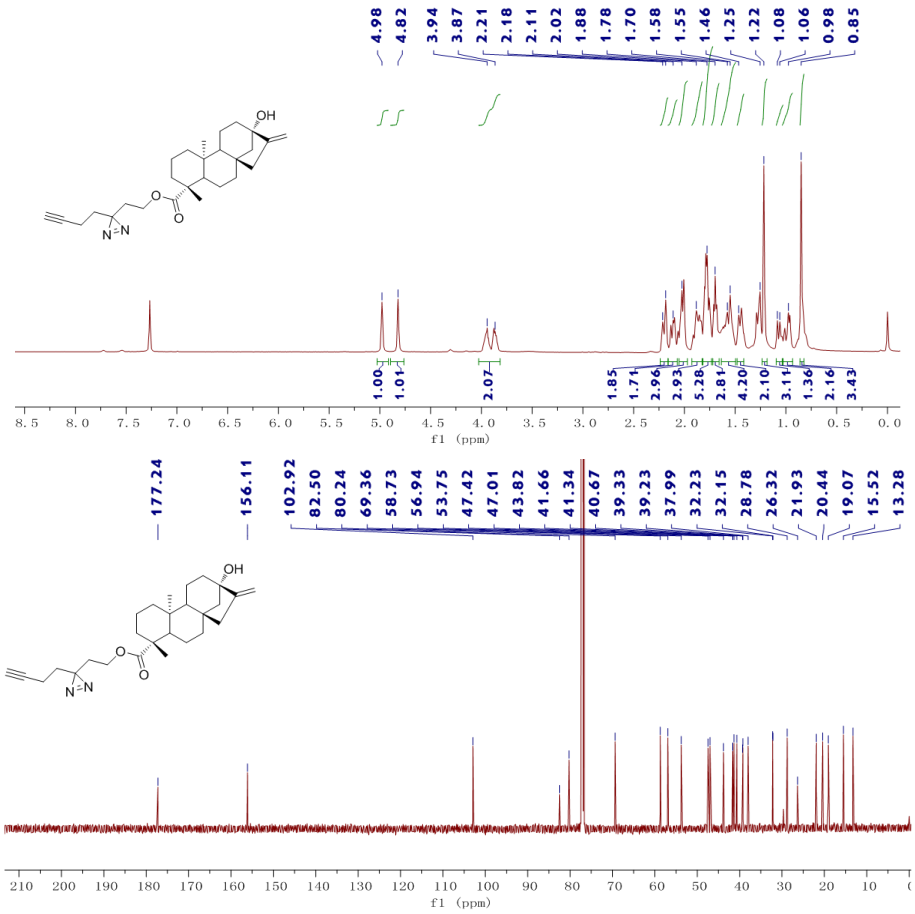
3.2 Synthesis of ST-Dayne.



Steviol (30 mg, 0.09 mmol, 1 eq) was dissolved in 2 mL dry DMF and cooled to 0°C with ice bath. 3-(but-3-ynyl)-3-(2-iodoethyl)-3H-diazirine[8] (33 mg, 0.12 mmol, 1.3 eq), K_2CO_3 (27 mg, 0.19 mmol, 2 eq) were added to the reaction. The resulting solution was heated with stirring at 70°C for 6 hr. DMF was removed under vacuum and the residue was dissolved in EtOAc, washed by 1N HCl, brine, dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by flash chromatograph column to provide **ST-Dayne** (34 mg, 82%) as white solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 4.98 (1H, s), 4.82 (1H, s), 3.90 (2H, br d), 2.21 (2H, m), 2.18 (2H, m), 2.11 (3H, m), 1.88 (3H, m), 1.78 (5H, m), 1.70 (3H, m), 1.58 (4H, m), 1.46 (2H, m), 1.22 (3H, s), 1.05 (3H, m) 0.85 (3H, s). ^{13}C NMR (125 MHz, $CDCl_3$) δ 177.24, 156.11, 102.92, 82.50, 80.24, 69.36, 58.73, 56.94, 53.75, 47.42, 47.01, 43.82, 41.66, 41.34, 40.67, 39.33, 39.23, 37.99, 32.23, 32.15, 28.78, 26.32, 21.93, 20.44, 19.07, 15.52, 13.28.

HRMS-ESI *calc'd. for* $C_{27}H_{39}N_2O_3$ $[M+H]^+$: 439.2955; Found: 439.2949.



4. **References**

1. Li, W. *et al.* Characterization of the artemisinin binding site for translationally controlled tumor protein (TCTP) by bioorthogonal click chemistry. *Bioconjugate Chem.* 2016, 27, 2828-2833.
2. Zhou Y. *et al.* Chemical proteomics reveal CD147 as a functional target of pseudolaric acid B in human cancer cells. *Chem. Commun.* 2017, 53, 8671-8674.
3. Sugimoto, N. *et al.* Analysis of Rubusoside and Related Compounds in Tenryocha Extract Sweetener. *J. Food. Hyg. Soc. Japan* 2002, 43, 250-253.
4. Chaturvedula, V. S. P. *et al.* Synthesis of *ent*-Kaurane Diterpene Monoglycosides. *Molecules* 2011, 16, 8402-8409.
5. Lehmann, J. *et al.* Synthesis of ramariolide natural products and discovery of their targets in mycobacteria. *Chem. Commun.* 2017, 53, 107-110.
6. Jackson, R. G. *et al*. Identification and biochemical characterization of an *Arabidopsis* indole-3-acetic acid glucosyltransferase. *J. Biol. Chem.* 2001, 276, 4350-4356.
7. Wisniewski, J. R. *et al.* Universal sample preparation method for proteome analysis. *Nat. Methods* 2008, 6, 359-362.
8. Li, Z. *et al*. Design and synthesis of minimalist terminal alkyne containing diazirine photo-crosslinkers and their incorporation into kinase inhibitors for cell- and tissue-based proteome profiling. *Angew. Chem. Int. Ed.* 2013, 52, 8551-8556.