Supporting Information

Facile Chemoenzymatic Synthesis of Biotinylated Heparosan Hexasaccharide

Table of Contents

Bacterial strains, plasmids, and materials

E. coli electrocompetent DH5α and chemically competent BL21(DE3) cells were purchased from Invitrogen (Carlsbad, CA). *P. multocida* P-1059 (ATCC#15742) was from American Type Culture Collection (ATCC, Manassas, VA, USA). KfiA synthetic gene with codons optimized for *E. coli* expression (GenBank accession number: KC538946) was customer synthesized by GeneArt (Grand Island, NY) based on KfiA gene sequence from *E. coli Nissle* 1917 (GenBank accession number: AJ586888, ORF79). Vector plasmids pET15b and pET22b (+) were from Novagen (EMD Biosciences Inc. Madison, WI, USA). Vector pMAL-c4X and restriction enzymes *Nde*I, *Bam*HI, *Eco*RI, and *Hin*dIII were purchased from New England Biolabs, Inc. (Beverly, MA, USA). Nickel–nitrilotriacetic acid agarose (Ni^{2+}–NTA agarose), QIAprep spin mini-prep kit, and QIAEX II gel extraction kit were from Qiagen (Valencia, CA, USA). Herculase-enhanced DNA polymerase was from Stratagene (La Jolla, CA, USA). T4 DNA ligase and 1 Kb DNA ladder were from Promega (Madison, WI, USA).

Cloning of PmHS2 and KfiA.

PmHS2 (GenBank accession number: AAQ55110) was cloned as N- and C-His_6-tagged fusion proteins in pET15b and pET22b(+) vectors, respectively, using the genomic DNA of *P. multocida* P-1059 (ATCC#15742, type A strain) as the template for polymerase chain reactions (PCR). KfiA was cloned as a fusion protein with an N-terminal maltose binding protein (MBP) and a C-terminal His_6-tag in pMAL-c4X vector using the KfiA synthetic gene as the PCR template. Primers used for cloning are shown in Table 1. PCR was performed in a reaction mixture of 50 μL containing the genomic DNA (1 μg) or the synthetic gene, forward and reverse primers (1 μM each), 10× Herculase buffer (5 μL), dNTP mixture (1 mM), and 5 U (1 μL) of Herculase-enhanced DNA polymerase. The reaction mixture was subjected to 30 cycles of PCR amplification with an annealing temperature of 55 °C (for PmHS2) or 52 °C (for KfiA). The resulting PCR products were purified, digested, and ligated with the corresponding predigested vectors. The ligation products were transformed into electrocompetent *E. coli* DH5α cells. Plasmids containing the target genes as confirmed by DNA sequencing (performed by UC-Davis Sequencing Facility) were selected and transformed into *E. coli* BL21(DE3) chemically competent cells.

Table 1. Primers used for cloning PmHS2 and KfiA

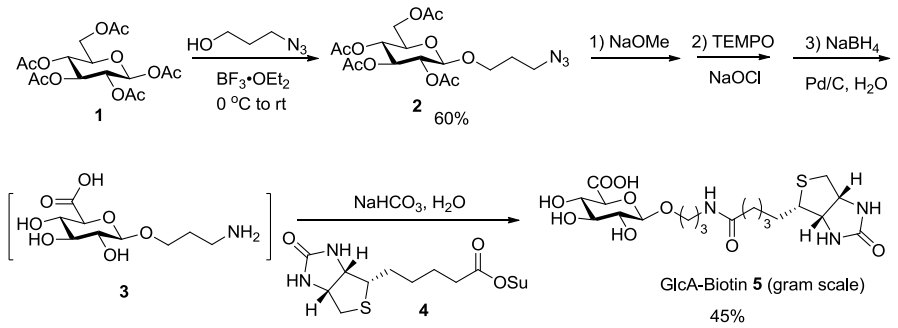
Primers	Sequences
KfiA_pMAL-c4X_F_*Eco*RI	5'-GACC*GAATTC*ATGATTGTTGCAAATATGAGC-3'
KfiA_pMAL-c4X_R_*Hind*III	5'-GTCG*AAGCTT*TTAGTGGTGGTGGTGGTGGTGACCTTCCACATTA TAC-3'
PmHS2_pET15b/22b(+)_F_*Nde*I	5'-GATC*CATATG*AAGGGAAAAAAAGAGATGAC-3'
PmHS2_pET15b_R_*Bam*HI	5'-AAG*GGATCC*TTATAAAAAATAAAAAGGTAAACAGG-3'
PmHS2_pET22b(+)_R_*Bam*HI	5'-AAG*GGATCC*TTAGTGGTGGTGGTGGTGGTGTAAAAAATAAAAA GGTAAACAGG

Expression and purification of PmHS2 and KfiA.

E. coli strains were cultured in LB rich medium (10 g/L tryptone , 5 g/L yeast extract, and 10 g/L NaCl) supplemented with ampicillin (100 μg/mL). Over-expression of PmHS2 was achieved by inducing the *E. coli* BL21 (DE3) cell culture with 0.1 mM of isopropyl-1-thio-β-D-galactopyranoside (IPTG) when the

$OD_{600\,nm}$ of the culture reached 0.8–1.0 followed by incubation at 20 °C with shaking (250 rpm) for 20 h. Overexpression of KfiA was performed by inoculating 10 mL of a fresh overnight bacterial culture grown in LB containing 50 μg/mL ampicillin and 20 μg/mL chloramphenicol into 1 L of LB (containing 50 μg/mL ampicillin, 20 μg/mL chloramphenicol, and 2 mg/mL L-arabinose). The culture was incubated at 37 °C with shaking at 250 rpm. When the $OD_{600\,nm}$ of the culture reached 0.4–0.6, expression was induced by adding IPTG to a final concentration of 0.3 mM followed by incubation at 20 °C with shaking (250 rpm) for 20 h. Bacterial cells were harvested by centrifugation at 4 °C in a Sorvall Legend RT centrifuge with a hanging bucket rotor at 4,000 rpm for 2 h. Harvested cells were resuspended in lysis buffer (Tris–HCl buffer, 100 mM, pH 8.0 containing 0.1 % Triton X-100) (20 mL for cells collected from one liter cell culture). Lysozyme (100 μg/mL) and DNaseI (5 μg/mL) were added to the cell resuspension. The resulting mixture was incubated at 37 °C for 1 h with shaking at 200 rpm. Cell lysate (supernatant) was obtained by centrifugation at 12,000 rpm for 15 min. Purification was carried out by loading the supernatant onto a Ni^{2+}-NTA column pre-equilibrated with 10 column volumes of binding buffer (10 mM imidazole, 0.5 M NaCl, 50 mM Tris–HCl, pH 7.5). The column was washed with 10 column volumes of binding buffer and 10 column volumes of washing buffer (20–50 mM imidazole, 0.5M NaCl, 50 mM Tris–HCl, pH 7.5). The target protein was eluted with Tris–HCl buffer (50 mM, pH 7.5) containing imidazole (200 mM) and NaCl (0.5 M). The fractions containing the purified enzymes were collected and dialyzed against Tris–HCl buffer (20 mM, pH 7.5) containing 10 % glycerol. Dialyzed proteins were stored at 4 °C.

Chemical Synthesis of GlcA-biotin conjugate 5.



3-Azido-propyl 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranoside (2).

To a stirred solution of β-D-glucose pentaacetate (10.0 g, 0.027 mol), 3-azido-propanol (6.7 g, 0.066 mol) in dry CH_2Cl_2 (100 mL) was added $BF_3{\cdot}Et_2O$ (16 mL, 0.081 mol) dropwise at 0 °C, the reaction mixture was stirred at 0 °C for 2 h, then at rt overnight, and was then quenched by slowly adding aq $NaHCO_3$ solution. The mixture was diluted with CH_2Cl_2, the organic layer was washed with aq $NaHCO_3$ solution, brine, dried over Na_2SO_4, concentrated. The residue was purified by column chromatography on silica gel (hexane/ethyl acetate, v/v, 2:1) to isolate the title compound (6.88 g, 60%), whose NMR characterization data is identical with the literature reported value.[1]

Synthesis of GlcA-biotin conjugate 5.

Among the steps for the preparation of GlcA-biotin conjugates **5**, the selective oxidation of Glucose into Glucuronic acid and azide reduction steps were performed based on our optimization of the procedure developed by Lefeber *et. al*.[2] The experimental details are shown below.

To a stirred solution of 3-azido-propyl 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranoside (2.62 g, 6.08 mmol) in anhydrous methanol (80 mL) was added 0.2 mL of freshly prepared solution of NaOMe in methanol (0.5 M) at rt. The reaction mixture was stirred at rt for 2h, then was neutralized with acid resin (Dowex® 50WX2 H^+–form) until pH around 7, filtered, and washed with methanol (20 mL x 3). The collected filtrate was concentrated providing a residue (1.6 g) which was used directly in next selective oxidation step without purification.

To the intermediate tetra-ol prepared above in water (150 mL) was added TEMPO (23 mg, 0.15 mmol), followed by KBr (0.37 g, 3.17 mmol). In a separate reaction flask, aqueous NaOCl (13% Cl active, 19.8 mL) was adjusted to pH 10 by addition of 4 M HCl and cooled to 5 ^oC. The pre-adjusted NaOCl solution was then added into the reaction mixture dropwise while the reaction mixture's pH (monitored by digital *Mettler-Toledo* pH Meter) was maintained at 9.5±0.5 by adding 1 M NaOH solution. After stirring for 4 h at 5 ^oC and subsequent TLC analysis (*n*-BuOH/AcOH/H_2O, 2:1:1) indicating the reaction cleanly generated a more polar spot, the reaction mixture was then neutralized with 4 M HCl to pH around 7 and concentrated to dryness. Desalting of product was performed by extraction of the residual solid with methanol (50 mL x 3). The combined organic solution was concentrated under reduced pressure, and the crude residue was used directly for next step reaction (azide reduction).

A solution of the above prepared 3-azido-propyl-β-D-glucopyranosiduronic acid in 0.1 M NaOH (30 mL) was added dropwise into a suspension of 10 % Pd/C (5 mg, catalytic) and $NaBH_4$ (1.15 g, 30.27 mmol) in water (15 mL) at rt. After 1h of stirring at rt and subsequent TLC analysis (*n*-BuOH/AcOH/H_2O, 2:1:1) indicating the reaction cleanly generated a more polar spot, the reaction mixture was neutralized with acid resin (Dowex® 50WX2 H^+–form) until pH around 7, filtered through Celite to remove Pd/C catalyst. The filtrate was lypholized to afford the amine product, which was directly used for next coupling step without purification.

To a stirred solution of biotin-OSu in acetonitrile (2.07 g, 6.05 mmol) was added a solution of the above prepared amine in water (0.16 M, 38 mL), followed by 0.5 mL of Sat. aq $NaHCO_3$ solution. After stirring at rt for 4 h, TLC (EA/MeOH/H_2O, 2:1:1) indicated the reaction was complete, the reaction was then mixed with 40 gram of silica gel, and solvents was removed under reduced pressure till dryness. The dry silica powder containing product was then loaded onto silica gel column and the product was purified by chromatography (EA/MeOH/H_2O, 2:1:1) to obtain the final GlcA-biotin conjugate **5** (1.5 g, 45% over 4 steps).

Representative procedure for one-pot three enzymes catalyzed reaction adding GlcNTFA residue forming disaccharide:

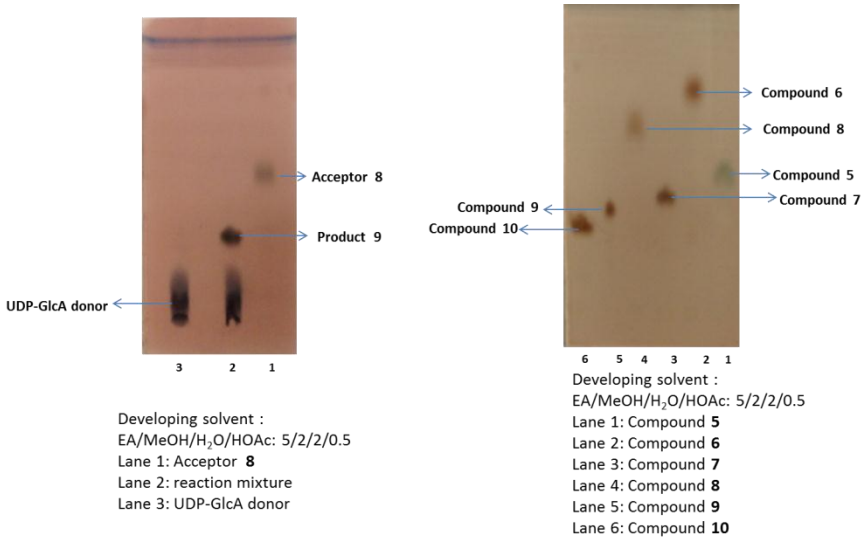
A total volume of 13 mL of aqueous solution in a centrifuge tube containing the following reaction components (final concentrations for each components indicated): GlcA-β-biotin **5** (10 mM), UTP (12 mM), ATP (12 mM), $MgCl_2$ (20 mM) and Tris-HCl buffer (200 mM, pH 7.0), PmGlmU (1-2 mg), NahK (1-2 mg), PmHS2 (2-3 mg) was incubated for 2 days at 37^oC with gentle shaking. Reaction progress and

product formation was monitored by thin layer chromatography (EtOAc/MeOH/H_2O/AcOH=5:2:1:0.2, v/v) using *p*-anisaldehyde staining solution. The reaction was quenched by adding an equal volume of ice-cold ethanol once the acceptor was monitored to be consumed completely, and stored at 4^oC for 30 min. The precipitated proteins were then removed by centrifugation (7000 rpm, 40 min). The clear solution was concentrated to around 2 mL and loaded onto BioGel P-2 gel filtration column eluting with pure water in a natural gravity-drove speed to obtain the desired product in pure form. The synthesis of tetrasaccharide and hexasaccharide followed essentially same procedure with same concentrations for respective components. Reaction progress and product formation was monitored by thin layer chromatography (EtOAc/MeOH/H_2O/AcOH=5:2:2:0.5, v/v) using *p*-anisaldehyde staining solution.

Representative procedure for enzymes catalyzed reaction adding GlcA residue forming trisaccharide:

A total volume of 11 mL of aqueous solution in a centrifuge tube containing the following reaction components (final concentrations for each components indicated): GlcNTFA-α-(1,4)-GlcA-β-biotin **6** (10 mM), UDP-GlcA (10 mM), $MnCl_2$ (10 mM) and Tris-HCl buffer (200 mM, pH 7.0), PmHS2 (2-3 mg) was incubated at 37^oC overnight with gentle shaking. Reaction progress and product formation was monitored by thin layer chromatography (EtOAc/MeOH/H_2O/AcOH=5:2:1.5:0.5, v/v) using a *p*-anisaldehyde staining solution. The reaction was quenched by adding an equal volume of ice-cold ethanol once the acceptor was monitored to be consumed completely and stored at 4^oC for 30 min. The precipitated proteins were then removed by centrifugation (7000 rpm, 40 min). The clear solution was concentrated to around 2 mL and loaded onto BioGel P-2 gel filtration column eluting with pure water in a natural speed driven by gravity to obtain the desired product in pure form. The synthesis of pentasaccharide backbone followed essentially same procedure with same concentrations for respective components. Reaction progress and product formation was monitored by thin layer chromatography (EtOAc/MeOH/H_2O/AcOH=5:2:2:0.5, v/v) using a *p*-anisaldehyde staining solution.

A typical TLC analysis for synthesis of pentasaccharide 9 and TLC analysis of compounds 5-10



NMR, HRESIMS and HPLC characterization datas of the synthesized compounds 5-10:

GlcA-biotin conjugate 5: ^{1}H NMR (D_2O, 400 MHz): δ 4.57 (dd, J = 4.8, 8.0 Hz, 1H, Biotin), 4.48 (d, J = 8.0 Hz, 1H, H-1-GlcA), 4.38 (dd, J = 4.8, 8.0 Hz, 1H, Biotin), 3.97 (d, J = 9.2 Hz, 1H, H-5-GlcA), 3.93-3.87 (m, 1H, OC*H*HCH$_2$CH$_2$NH), 3.68-3.62 (m, 1H, OCH*H*CH$_2$CH$_2$NH), 3.57-3.47 (m, 2H, H-3, 4-GlcA), 3.31-3.23 (m, 4H, H-2-GlcA, OCH$_2$CH$_2$*CH$_2$*NH, Biotin), 2.95 (dd, J = 4.8, 13.2 Hz, 1H, Biotin), 2.74 (d, J = 13.2 Hz, 1H, Biotin), 2.21 (t, J = 7.2 Hz, 2H, Biotin), 1.82-1.76 (m, 2H, OCH$_2$*CH$_2$*CH$_2$NH), 1.69-1.49 (m, 4H, Biotin), 1.41-1.30 (m, 2H, Biotin); ^{13}C NMR (D_2O, 100 MHz): δ 176.78 (CO), 172.06 (CO), 165.28 (NH*CO*NH), 102.21 (C-1-GlcA), 75.20, 74.38, 72.68, 71.22, 67.80 (O*CH$_2$*CH$_2$CH$_2$NH), 62.09, 60.23, 55.38, 39.65, 36.05 (OCH$_2$CH$_2$*CH$_2$*NH), 35.44, 28.19 (OCH$_2$*CH$_2$*CH$_2$NH), 27.87, 27.66, 25.14. HRESIMS calcd for $C_{19}H_{30}N_3O_9S$ [M-H]$^-$: 476.5215, found 476.5210; Anal. RP-HPLC: t_R = 4.06 min (Eluent A: H_2O (containing 0.05% TFA, v/v), Eluent B: CH_3CN (containing 0.05% TFA, v/v); gradient: 0–10 min 0–60% B; purity: 97.25 %).

GlcNTFA-α-(1,4)-GlcA-β-biotin 6: ^{1}H NMR (D_2O, 400 MHz): δ 5.40 (d, J = 3.6 Hz, 1H, H-1-GlcNTFA), 4.59 (dd, J = 4.8, 8.0 Hz, 1H, Biotin), 4.48 (d, J = 8.0 Hz, 1H, H-1-GlcA), 4.40 (dd, J = 4.4, 8.0 Hz, 1H, Biotin), 4.06 (d, J = 9.6 Hz, 1H, H-5-GlcA), 4.00 (dd, J = 3.6, 11.2 Hz, 1H, H-2-GlcNTFA), 3.93-3.87 (m, 1H, OC*H*HCH$_2$CH$_2$NH), 3.83-3.75 (m, 4H, H-6, H-6', 5, 3-GlcNTFA), 3.73-3.63 (m, 2H, OCH*H*CH$_2$CH$_2$NH, H-4-GlcNTFA), 3.60-3.52 (m, 2H, H-3, 4-GlcA), 3.33-3.22 (m, 4H, OCH$_2$CH$_2$*CH$_2$*NH, H-2-GlcA, Biotin), 2.97 (dd, J = 4.8, 12.8 Hz, 1H, Biotin), 2.76 (d, J = 12.8 Hz, 1H, Biotin), 2.23 (t, J = 7.2 Hz, 2H, Biotin), 1.83-1.77 (m, 2H, OCH$_2$*CH$_2$*CH$_2$NH), 1.73-1.53 (m, 4H, Biotin), 1.41-1.32 (m, 2H, Biotin); ^{13}C NMR (D_2O, 100 MHz): δ 176.78 (CO), 171.79 (CO), 165.31 (NH*CO*NH), 161.00 (q, CF$_3$*CO*, J = 35.6 Hz), 116.00 (q, J = 50.7 Hz, *CF$_3$*CO), 102.30 (C-1-GlcA), 97.29 (C-1-GlcNTFA), 76.95, 75.74, 74.12, 73.09, 72.38, 69.95, 69.14, 67.86 (O*CH$_2$*CH$_2$CH$_2$NH), 62.09, 60.22, 59.70, 55.38, 54.16, 39.66, 36.04 (OCH$_2$CH$_2$*CH$_2$*NH), 35.46, 28.19 (OCH$_2$*CH$_2$*CH$_2$NH), 27.86, 27.66, 25.13. HRESIMS calcd for $C_{27}H_{40}F_3N_4O_{14}S$ [M-H]$^-$: 735.2214, found 735.2175; Anal. RP-HPLC: t_R = 4.52 min (Eluent A: H_2O (containing 0.05% TFA, v/v), Eluent B: CH_3CN (containing 0.05% TFA, v/v); gradient: 0–10 min 0–60% B; purity: 99.49 %).

GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 7: ^{1}H NMR (D_2O, 400 MHz): δ 5.41 (d, J = 3.6 Hz, 1H, H-1-GlcNTFA), 4.60-4.59 (m, 2H, H-1-GlcA'), 4.49 (d, J = 8.0 Hz, 1H, H-1-GlcA), 4.41 (dd, J = 4.8, 8.0 Hz, 1H, Biotin), 4.08-4.01(m, 3H, H-5-GlcA, H-5-GlcA', H-2-GlcNTFA), 3.98-3.88 (m, 3H, OC*H*HCH$_2$CH$_2$NH, H-6, 3-GlcNTFA), 3.83-3.67 (m, 6H, H-6', 5, 4-GlcNTFA, OCH*H*CH$_2$CH$_2$NH, H-3, 4-GlcA'), 3.62-3.53 (m, 2H, H-3, 4-GlcA), 3.40-3.25 (m, 5H, OCH$_2$CH$_2$*CH$_2$*NH, H-2-GlcA', H-2-GlcA, Biotin), 2.98 (dd, J = 4.8, 13.2 Hz, 1H, Biotin), 2.77 (d, J = 12.8 Hz, 1H, Biotin), 2.24 (t, J = 6.8 Hz, 2H, Biotin), 1.84-1.78 (m, 2H, OCH$_2$*CH$_2$*CH$_2$NH), 1.74-1.53 (m, 4H, Biotin), 1.44-1.35 (m, 2H, Biotin); ^{13}C NMR (D_2O, 100 MHz): δ 175.79 (CO), 171.04 (CO), 170.85 (CO), 164.32 (NH*CO*NH), 160.00 (q, CF$_3$*CO*, J = 35.3 Hz), 115.05 (q, J = 53.9 Hz, *CF$_3$*CO), 101.31 (C-1-GlcA'), 101.24 (C-1-GlcA), 95.99 (C-1-GlcNTFA), 77.24, 76.08, 74.74, 73.98, 73.12, 73.08, 72.09, 71.70, 70.11, 70.05, 67.41, 66.87 (O*CH$_2$*CH$_2$CH$_2$NH), 61.10, 59.23, 58.07, 54.39, 52.92, 38.68, 35.03 (OCH$_2$CH$_2$*CH$_2$*NH), 34.47, 27.19 (OCH$_2$*CH$_2$*CH$_2$NH), 26.87, 26.68, 24.15. HRESIMS calcd for $C_{33}H_{48}F_3N_4O_{20}S$ [M-H]$^-$: 909.2535, found 909.2522; Anal. RP-HPLC: t_R = 4.46 min (Eluent A: H_2O (containing 0.05% TFA, v/v), Eluent B: CH_3CN (containing 0.05% TFA, v/v); gradient: 0–10 min 0–60% B; purity: 97.26 %).

GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 8: ^{1}H NMR (D_2O, 600 MHz): δ 5.29 (d, *J* = 3.6 Hz, 1H, H-1-GlcNTFA'), 5.26 (d, *J* = 3.0 Hz, 1H, H-1-GlcNTFA), 4.45 (dd, *J* = 5.4, 7.2 Hz, 1H, Biotin), 4.42 (d, *J* = 7.8 Hz, 1H, H-1-GlcA'), 4.34 (d, *J* = 7.8 Hz, 1H, H-1-GlcA), 4.26 (dd, *J* = 4.8, 7.2 Hz, 1H, Biotin), 3.94 (d, *J* = 10.2 Hz, 1H, H-5-GlcA'), 3.90 (d, *J* = 9.6 Hz, 1H, H-5-GlcA), 3.88-3.85 (m, 2H, H-2-GlcNTFA', H-2-GlcNTFA), 3.80-3.75 (m, 3H, OC*H*HCH$_2$CH$_2$NH, H-6-GlcNTFA', H-3-GlcNTFA'), 3.69-3.61 (m, 6H, H-6'-GlcNTFA', H-6, 6'-GlcNTFA, H-5-GlcNTFA', H-5-GlcNTFA, H-3-GlcNTFA), 3.58-3.46 (m, 6H, OCH*H*CH$_2$CH$_2$NH, H-3-GlcA', H-3-GlcA, H-4-GlcNTFA', H-4-GlcNTFA, H-2-GlcA'), 3.40 (t, *J* = 9.6 Hz, 1H, H-4-GlcA'), 3.21 (t, *J* = 9.0 Hz, 1H, H-4-GlcA), 3.08-3.17 (m, 4H, OCH$_2$CH$_2$*CH$_2$*NH, H-2-GlcA, Biotin), 2.83 (dd, *J* = 4.8, 13.2 Hz, 1H, Biotin), 2.62 (d, *J* = 12.6 Hz, 1H, Biotin), 2.09 (t, *J* = 7.2 Hz, 2H, Biotin), 1.68-1.64 (m, 2H, OCH$_2$*CH$_2$*CH$_2$NH), 1.58-1.39 (m, 4H, Biotin), 1.27-1.21 (m, 2H, Biotin); ^{13}C NMR (D_2O, 150 MHz): *δ* 176.86 (CO), 174.95 (CO), 174.90 (CO), 165.35 (NH*CO*NH), 161.61 (q, *J* = 34.5 Hz, CF$_3$*CO*), 115.32 (q, *J* = 52.6 Hz, *CF$_3$*CO), 102.36 (C-1-GlcA'), 101.99 (C-1-GlcA), 96.27 (C-1-GlcNTFA'), 96.06 (C-1-GlcNTFA), 78.17, 76.64, 76.41, 76.31, 76.16, 76.09, 76.04, 75.78, 73.42, 73.03, 71.92, 70.63, 70.10, 69.54, 68.54, 67.43 (O*CH$_2$*CH$_2$CH$_2$NH), 62.08, 60.22, 59.96, 59.24, 55.36, 54.24, 53.82, 39.66, 35.94 (OCH$_2$CH$_2$*CH$_2$*NH), 35.50, 28.16 (OCH$_2$*CH$_2$*CH$_2$NH), 27.87, 27.66, 25.17. HRESIMS calcd for $C_{41}H_{58}F_6N_5O_{25}S$ $[M-H]^-$: 1168.3046, found 1168.2958; Anal. RP-HPLC: t_R = 4.72 min (Eluent A: H_2O (containing 0.05% TFA, v/v), Eluent B: CH_3CN (containing 0.05% TFA, v/v); gradient: 0–10 min 0–60% B; purity: 99.14%).

GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 9: ^{1}H NMR (D_2O, 600 MHz): *δ* 5.25 (d, *J* = 3.6 Hz, 1H, H-1-GlcNTFA'), 5.22 (d, *J* = 3.6 Hz, 1H, H-1-GlcNTFA), 4.43-4.39 (m, 3H, H-1-GlcA'', H-1-GlcA', Biotin), 4.31 (d, *J* = 7.8 Hz, 1H, H-1-GlcA), 4.23 (dd, *J* = 4.8, 7.8 Hz, 1H, Biotin), 3.93 (d, *J* = 9.6 Hz, 1H, H-5-GlcA''), 3.90-3.83 (m, 4H, H-5-GlcA', H-5-GlcA, H-2-GlcNTFA', H-2-GlcNTFA), 3.79-3.72 (m, 5H, OC*H*HCH$_2$CH$_2$NH, H-6-GlcNTFA', H-6-GlcNTFA, H-5-GlcNTFA', H-3-GlcNTFA), 3.66-3.47 (m, 11 H, OCH*H*CH$_2$CH$_2$NH, H-6'-GlcNTFA', H-6'-GlcNTFA, H-5-GlcNTFA, H-3-GlcNTFA', H-3-GlcA'', H-3-GlcA', H-3-GlcA, H-4-GlcNTFA', H-4-GlcNTFA, H-2-GlcA''), 3.42-3.36 (m, 2H, H-4-GlcA'', H-4-GlcA'), 3.20-3.07 (m, 6H, OCH$_2$CH$_2$*CH$_2$*NH, H-4-GlcA, H-2-GlcA', H-2-GlcA, Biotin), 2.80 (dd, *J* = 4.8, 13.2 Hz, 1H, Biotin), 2.59 (d, *J* = 12.6 Hz, 1H, Biotin), 2.06 (t, *J* = 7.2 Hz, 2H, Biotin), 1.64-1.62 (m, 2H, OCH$_2$*CH$_2$*CH$_2$NH), 1.56-1.36 (m, 4H, Biotin), 1.22-1.20 (m, 2H, Biotin); ^{13}C NMR (D_2O, 150 MHz): *δ* 176.64 (CO), 171.91 (CO), 171.77 (CO), 171.70 (CO), 165.11 (NH*CO*NH), 162.71 (q, *J* = 35.6 Hz, CF$_3$*CO*), 158.93 (q, *J* = 37.8 Hz, CF$_3$*CO*), 117.71 (q, *J* = 66.9 Hz, *CF$_3$*CO), 113.88 (q, *J* = 61.4 Hz, *CF$_3$*CO), 102.11 (overlap two carbon, C-1-GlcA'', C-1-GlcA'), 102.06 (C-1-GlcA), 96.90 (C-1-GlcNTFA'), 96.69 (C-1-GlcNTFA), 78.04, 77.74, 76.94, 76.61, 75.52, 75.44, 74.74, 73.89, 73.86, 73.59, 72.99, 72.90, 72.47, 70.94, 70.85, 68.14, 68.09, 67.60 (O*CH$_2$*CH$_2$CH$_2$NH), 61.95, 60.07, 58.82, 58.76, 55.26, 53.74, 53.67, 39.53, 35.84 (OCH$_2$CH$_2$*CH$_2$*NH), 35.27, 27.96 (OCH$_2$*CH$_2$*CH$_2$NH), 27.75, 27.53, 25.02. HRESIMS calcd for $C_{47}H_{66}F_6N_5O_{31}S$ $[M-H]^-$: 1342.3367, found 1342.3306; Anal. RP-HPLC: t_R = 4.63 min (Eluent A: H_2O (containing 0.05% TFA, v/v), Eluent B: CH_3CN (containing 0.05% TFA, v/v); gradient: 0–10 min 0–60% B; purity: 98.25%).

GlcNAc-α-(1,4)GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 10

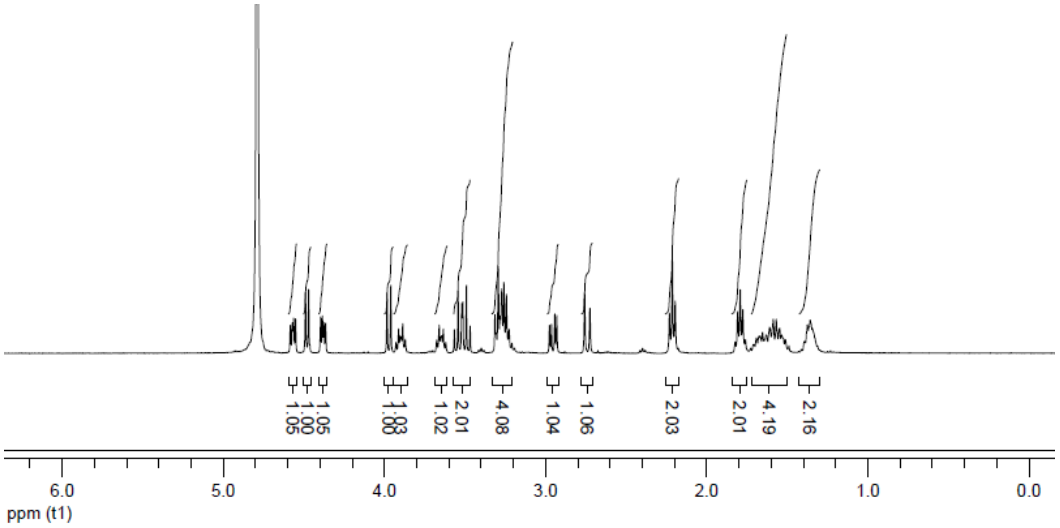
^{1}H NMR (D_2O, 600 MHz): δ 5.45 (d, J = 3.6 Hz, 1H, H-1-GlcNTFA'), 5.43 (d, J = 3.6 Hz, 1H, H-1-GlcNTFA), 5.38 (d, J = 3.6 Hz, 1H, H-1-GlcNAc), 4.63-4.58 (m, 3H, H-1-GlcA", H-1-GlcA', Biotin), 4.50 (d, J = 7.8 Hz, 1H, H-1-GlcA), 4.43 (dd, J = 4.8, 7.8 Hz, 1H, Biotin), 4.11-4.03 (m, 5H, H-5-GlcA", H-5-GlcA', H-5-GlcA, H-2-GlcNTFA', H-2-GlcNTFA), 3.97-3.89 (m, 6H, OC*H*HCH$_2$CH$_2$NH, H-2-GlcNAc, H-6-GlcNTFA', H-6-GlcNTFA, H-3-GlcNTFA', H-3-GlcNTFA), 3.85-3.80 (m, 6H, H-6'-GlcNTFA', H-6'-GlcNTFA, H-6-GlcNAc, H-5-GlcNTFA', H-5-GlcNTFA, H-3-GlcNAc), 3.77-3.67 (m, 10H, OCH*H*CH$_2$CH$_2$NH, H-6'-GlcNAc, H-5-GlcNAc, H-3-GlcA", H-3-GlcA', H-3-GlcA, H-4-GlcNTFA', H-4-GlcNTFA, H-4-GlcNAc, H-2-GlcA"), 3.61 (t, J = 10.4 Hz, 1H, H-4-GlcA"), 3.53 (t, J = 9.6 Hz, 1H, H-4-GlcA'), 3.40-3.25 (m, 6H, OCH$_2$CH$_2$*CH$_2$*NH, H-4-GlcA, H-2-GlcA', H-2-GlcA, Biotin), 3.00 (dd, J = 5.4, 13.2 Hz, 1H, Biotin), 2.79 (d, J = 13.2 Hz, 1H, Biotin), 2.26 (t, J = 7.2 Hz, 2H, Biotin), 2.06 (s, 3H, *CH$_3$*CONH), 1.85-1.81 (m, 2H, OCH$_2$*CH$_2$*CH$_2$NH), 1.76-1.56 (m, 4H, Biotin), 1.45-1.38 (m, 2H, Biotin); ^{13}C NMR (D_2O, 150 MHz): δ 176.70 (CO), 174.29 (CO), 172.35 (CO), 172.17 (CO), 172.14 (CO), 165.19 (NH*CO*NH), 162.96 (q, J = 34.3 Hz, CF$_3$*CO*), 158.97 (q, J = 37.7 Hz, CF$_3$*CO*), 117.77 (q, J = 55.8 Hz, *CF$_3$*CO), 114.91 (q, J = 73.7 Hz, *CF$_3$*CO), 102.15 (C-1-GlcA"), 102.09 (overlap two carbon, C-1-GlcA', C-1-GlcA), 97.43 (C-1-GlcNAc), 96.75 (C-1-GlcNTFA'), 96.60 (C-1-GlcNTFA), 77.81, 77.76, 76.73, 76.46, 76.14, 75.68, 75.60, 75.57, 74.21, 74.03, 73.95, 73.07, 72.96, 72.07, 70.89, 70.23, 69.05, 68.10, 68.07, 67.56 (O*CH$_2$*CH$_2$CH$_2$NH), 61.95, 60.07, 59.56, 58.82, 58.81, 55.27, 53.73, 53.70, 53.33, 39.54, 35.79 (OCH$_2$CH$_2$*CH$_2$*NH), 35.30, 27.97 (OCH$_2$*CH$_2$*CH$_2$NH), 27.74, 27.53, 25.02, 21.66 (*CH$_3$*CONH). HRESIMS calcd for $C_{55}H_{79}F_6N_6O_{36}S$ $[M-H]^-$: 1545.4160, found 1545.4065; $[M-H]^{2-}$ 772.2041, found 772.2002. Anal. RP-HPLC: t_R = 4.56 min (Eluent A: H_2O (containing 0.05% TFA, v/v), Eluent B: CH_3CN (containing 0.05% TFA, v/v); gradient: 0–10 min 0–60% B; purity: 99.02%).

References

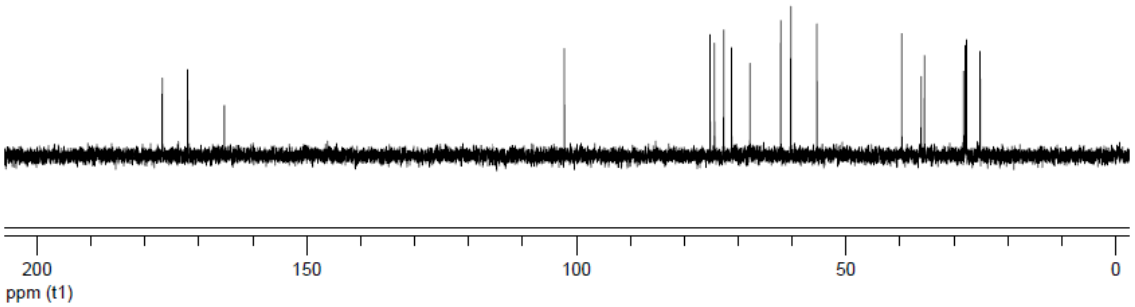
(1) Peng, L.; DeSousa, J.; Su, Z.; Novak, B. M.; Nevzorov, A. A.; Garland, E. R.; Melander, C. *Chem. Commun.*, 2011, **47**, 4896.

(2) Lefeber, D. J.; Kamerling, J. P.; Vliegenthart, J. F. G. *Chem.-Eur. J.*, 2001, **7**, 4411.

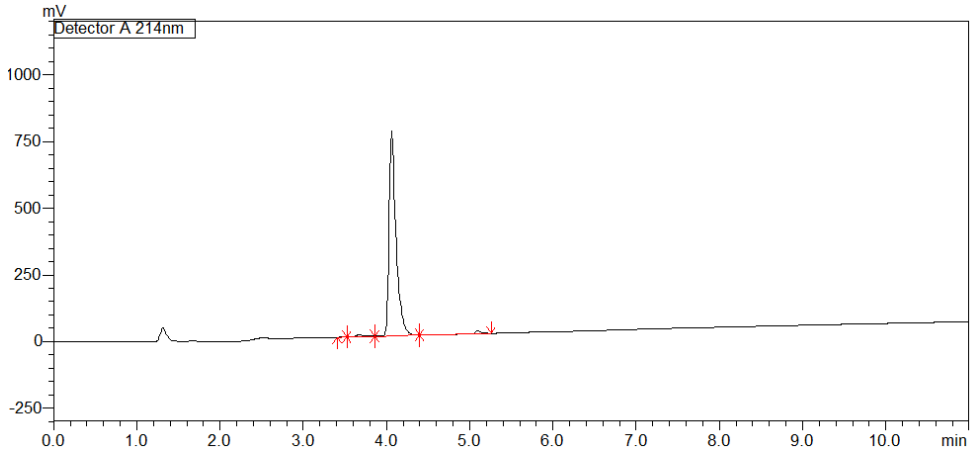
^{1}H-NMR of GlcA-biotin conjugate 5: (400 MHz, D_2O)



^{13}C-NMR of GlcA-biotin conjugate 5 (100 MHz, D_2O)

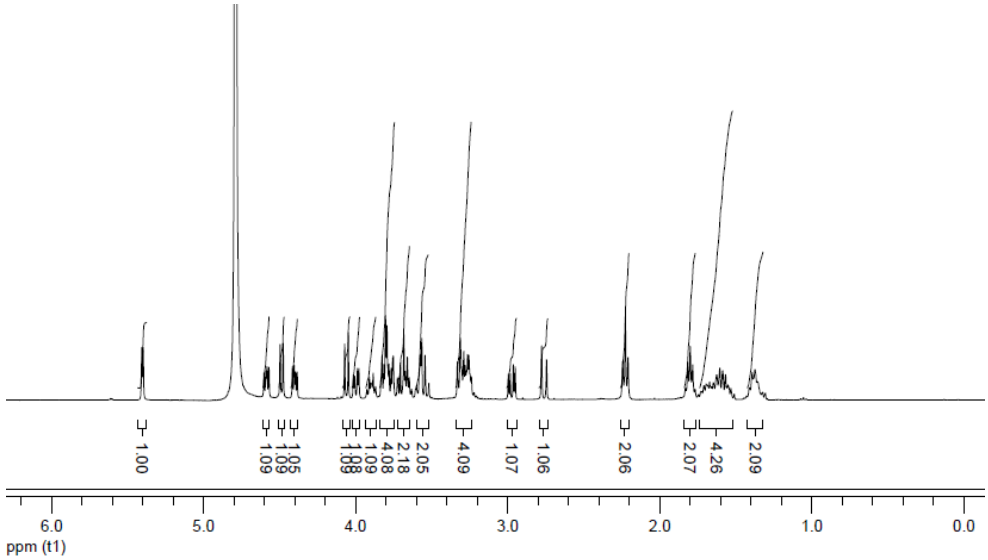


HPLC spectra of GlcA-biotin conjugate 5

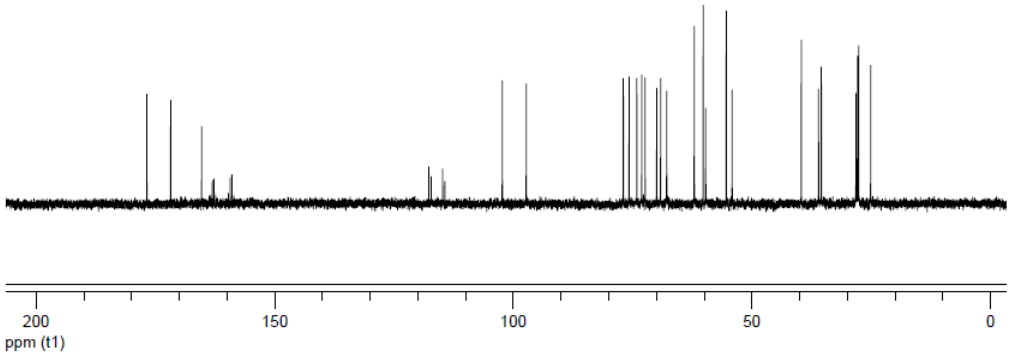


Purity: 97.25%. HPLC instrument: Shimadzu Prominence 20A coupled with UV detector; Detection: UV absorption at 214 nm; Column: Agilent Plus C18 (4.6 × 100 mm, 3.5 µm). Mobile Phase: A, H_2O (containing 0.05% TFA, v/v). B, Acetonitrile (containing 0.05% TFA, v/v). Gradient elution: B from 0 to 60% in 10 min. Flow rate: 1 mL/min.

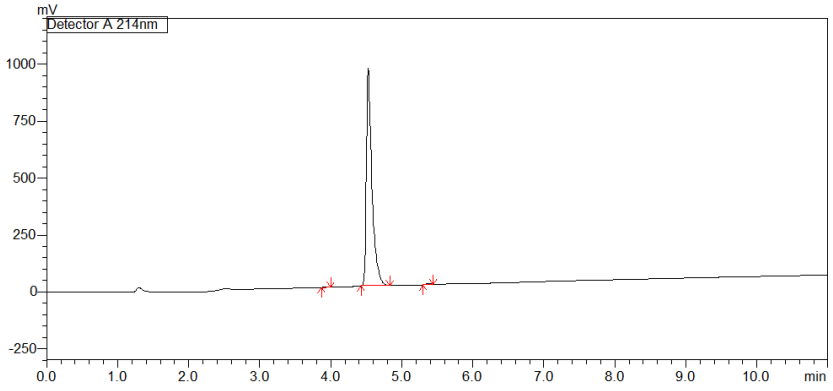
^{1}H-NMR of GlcNTFA-α-(1,4)-GlcA-β-biotin 6 (400 MHz, D_2O)



^{13}C-NMR of GlcNTFA-α-(1,4)-GlcA-β-biotin 6 (100 MHz, D_2O)

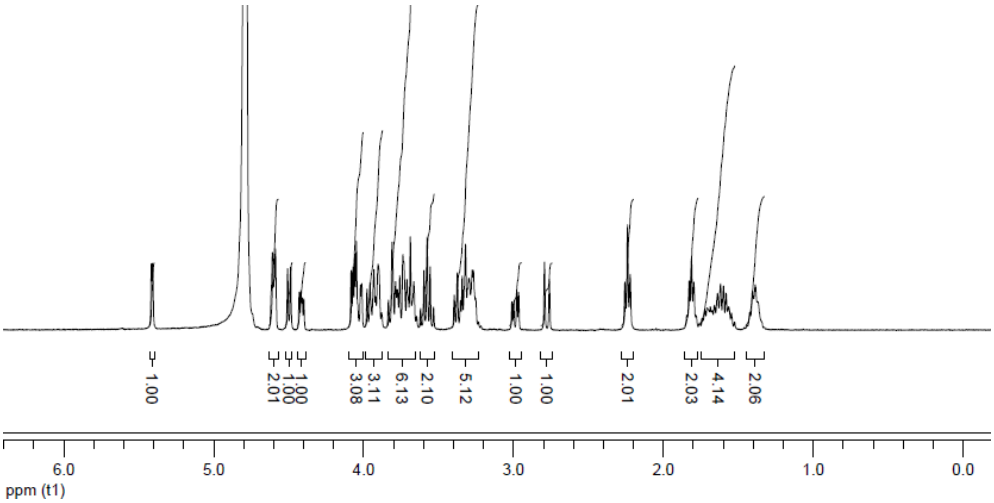


HPLC spectra of GlcNTFA-α-(1,4)-GlcA-β-biotin 6

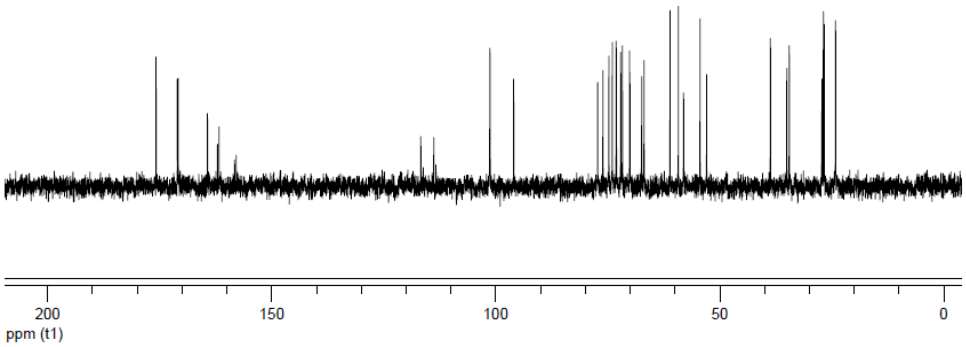


Purity: 99.49%. HPLC instrument: Shimadzu Prominence 20A coupled with UV detector; Detection: UV absorption at 214 nm; Column: Agilent Plus C18 (4.6 × 100 mm, 3.5 μm). Mobile Phase: A, H_2O (containing 0.05% TFA, v/v). B, Acetonitrile (containing 0.05% TFA, v/v). Gradient elution: B from 0 to 60% in 10 min. Flow rate: 1 mL/min.

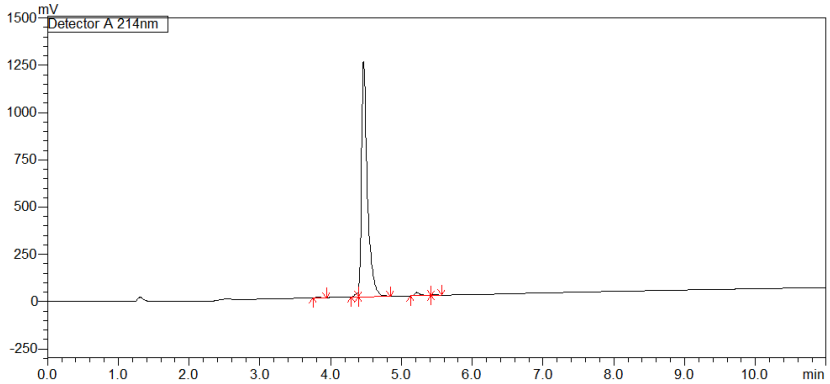
^{1}H-NMR of GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 7 (400 MHz, D_2O)



^{13}C-NMR of GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 7 (100 MHz, D_2O)

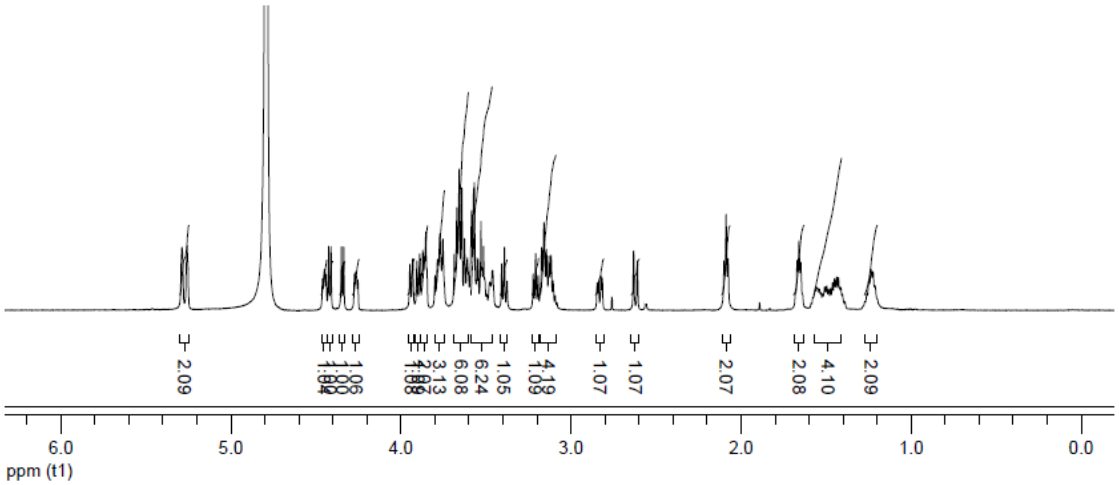


HPLC spectra of GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 7

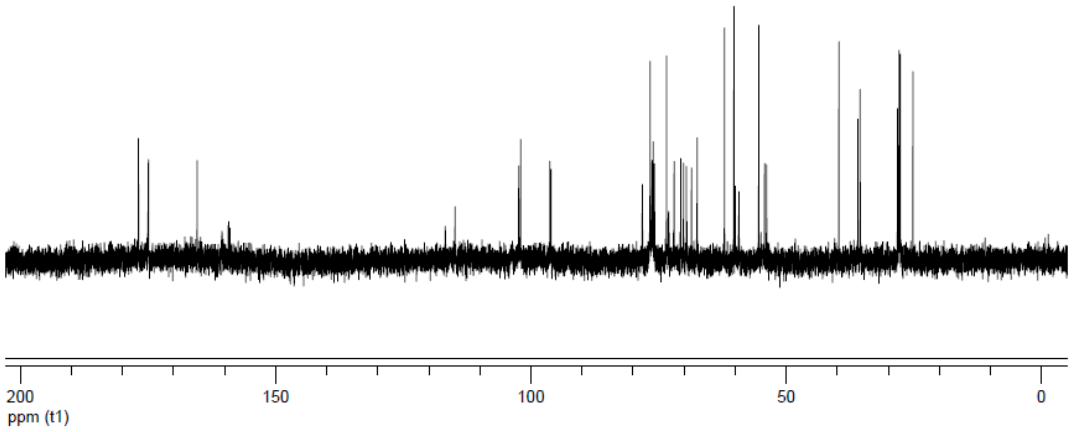


Purity: 97.26%. HPLC instrument: Shimadzu Prominence 20A coupled with UV detector; Detection: UV absorption at 214 nm; Column: Agilent Plus C18 (4.6 × 100 mm, 3.5 μm). Mobile Phase: A, H_2O (containing 0.05% TFA, v/v). B, Acetonitrile (containing 0.05% TFA, v/v). Gradient elution: B from 0 to 60% in 10 min. Flow rate: 1 mL/min.

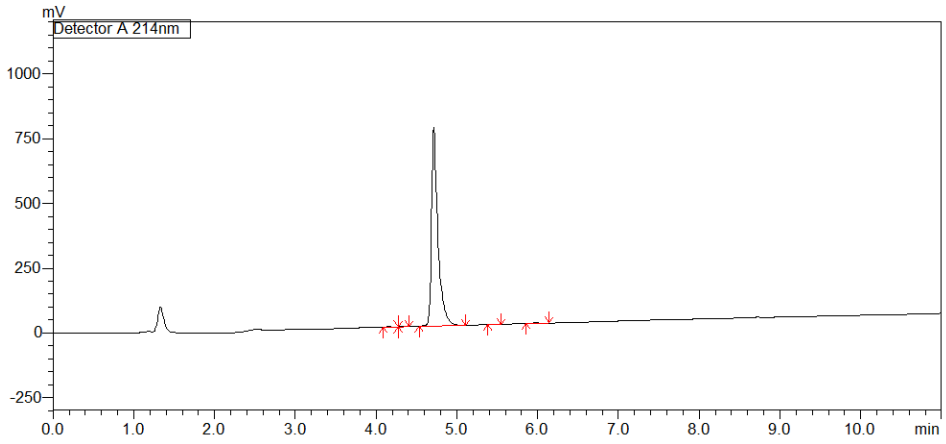
^{1}H-NMR of GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 8 (600 MHz, D_2O)



^{13}C-NMR of GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 8 (150 MHz, D_2O)

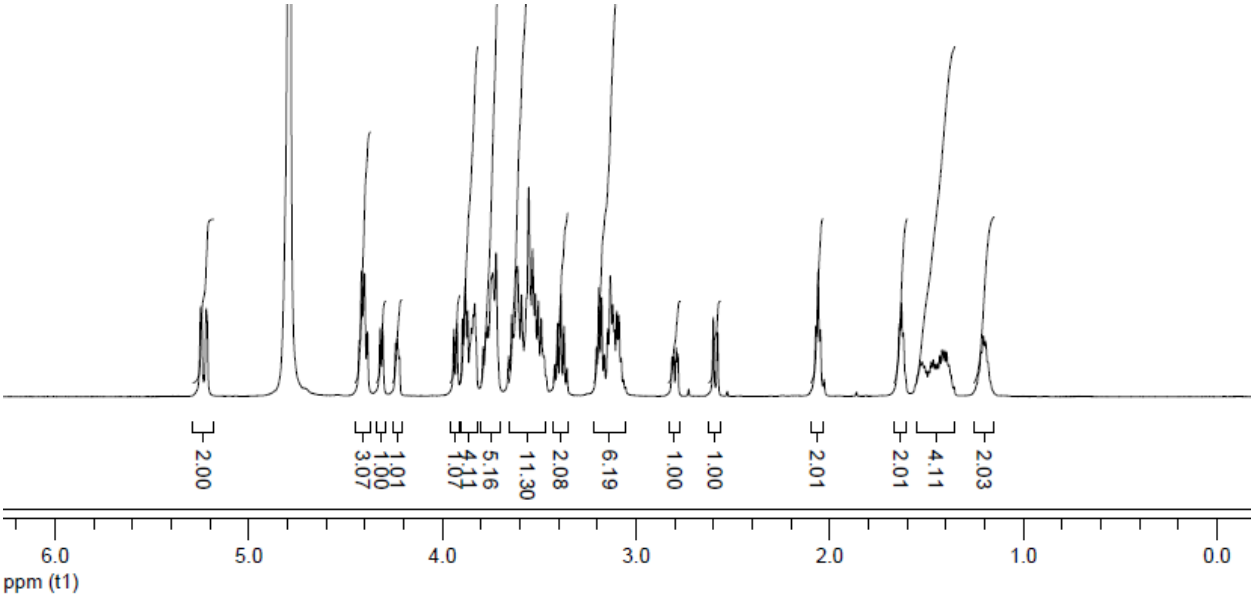


HPLC spectra of GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 8

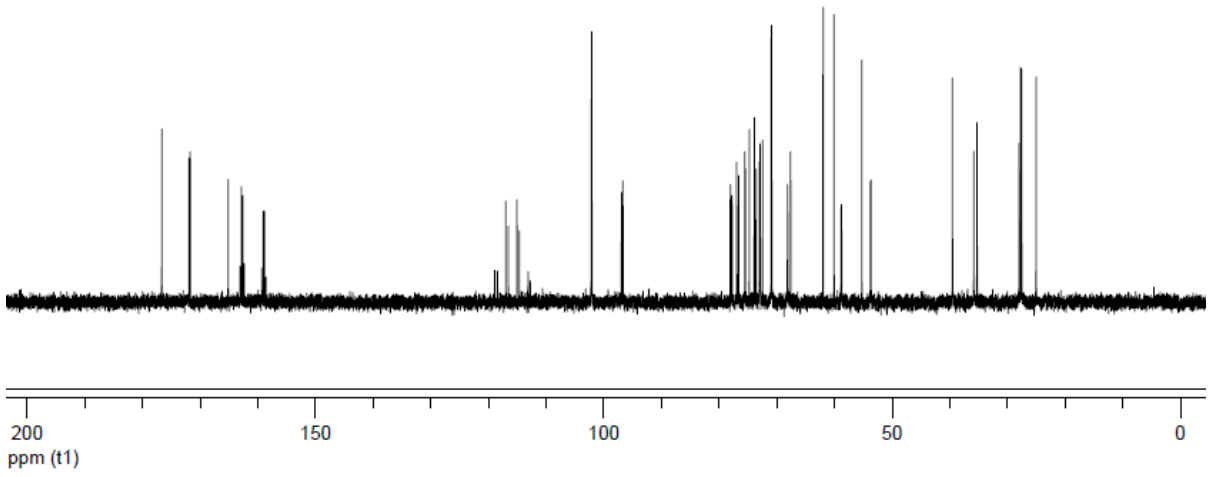


Purity: 99.14%. HPLC instrument: Shimadzu Prominence 20A coupled with UV detector; Detection: UV absorption at 214 nm; Column: Agilent Plus C18 (4.6 × 100 mm, 3.5 μm). Mobile Phase: A, H_2O (containing 0.05% TFA, v/v). B, Acetonitrile (containing 0.05% TFA, v/v). Gradient elution: B from 0 to 60% in 10 min. Flow rate: 1 mL/min.

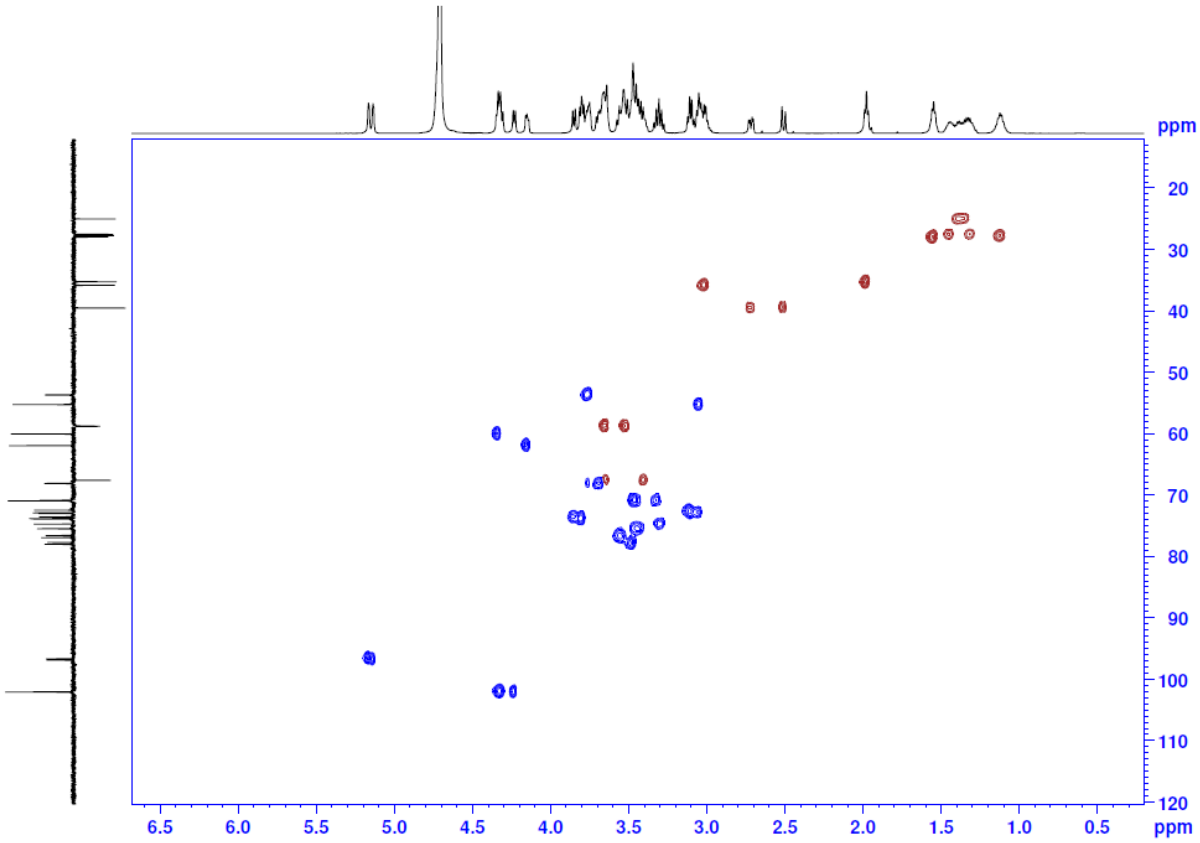
^{1}H-NMR of GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 9 (600 MHz, D_2O)



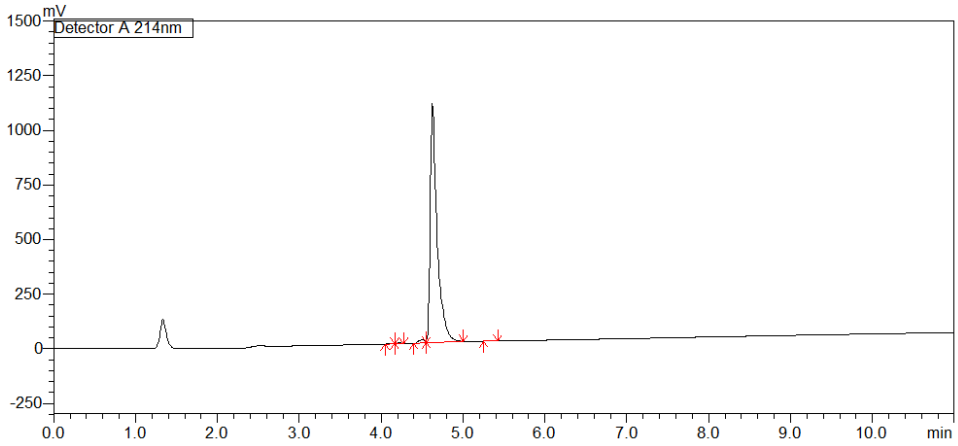
^{13}C-NMR of GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 9 (150 MHz, D_2O)



HSQC of GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 9 (600 MHz, D_2O)

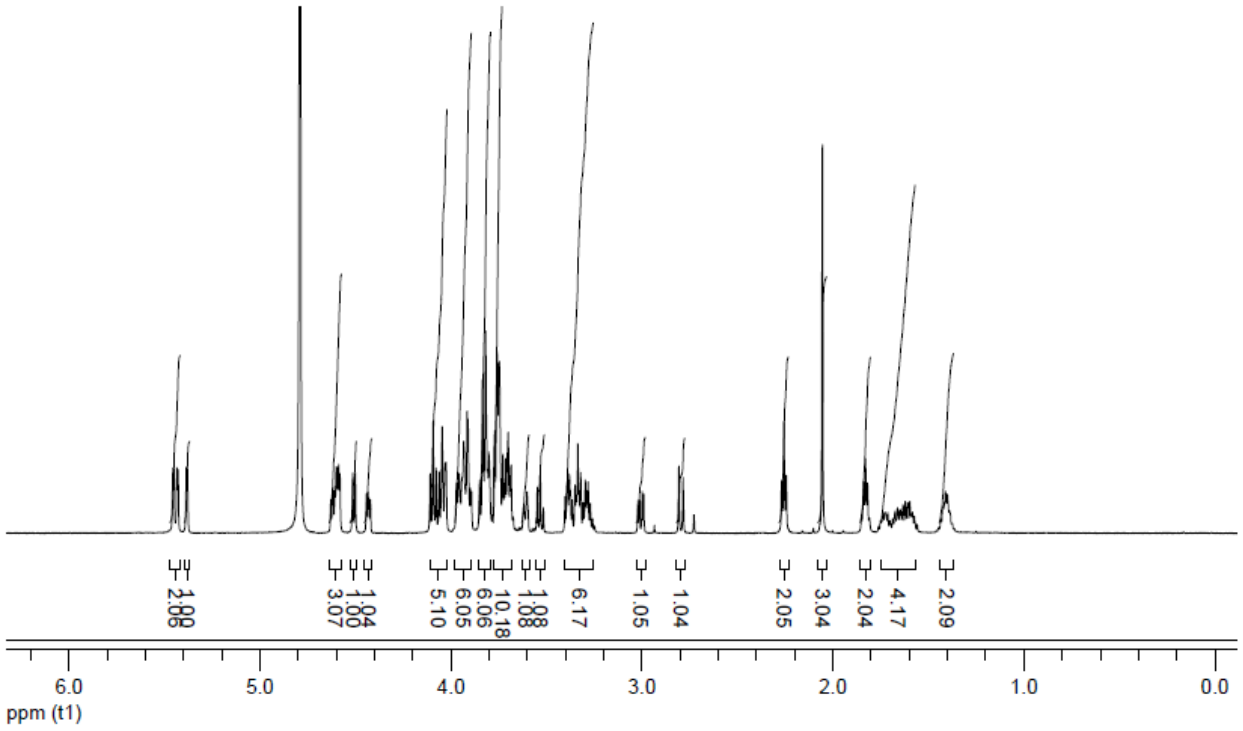


HPLC spectra of GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 9

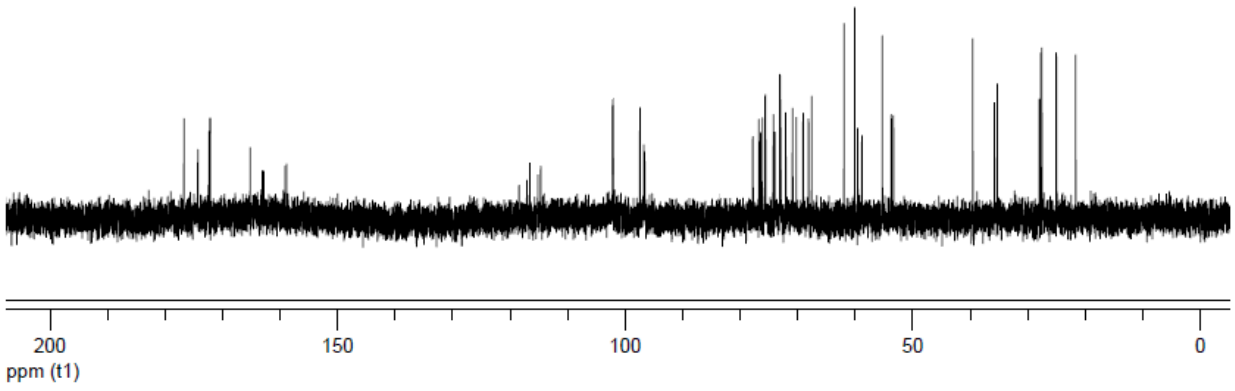


Purity: 98.25%. HPLC instrument: Shimadzu Prominence 20A coupled with UV detector; Detection: UV absorption at 214 nm; Column: Agilent Plus C18 (4.6 × 100 mm, 3.5 μm). Mobile Phase: A, H_2O (containing 0.05% TFA, v/v). B, Acetonitrile (containing 0.05% TFA, v/v). Gradient elution: B from 0 to 60% in 10 min. Flow rate: 1 mL/min.

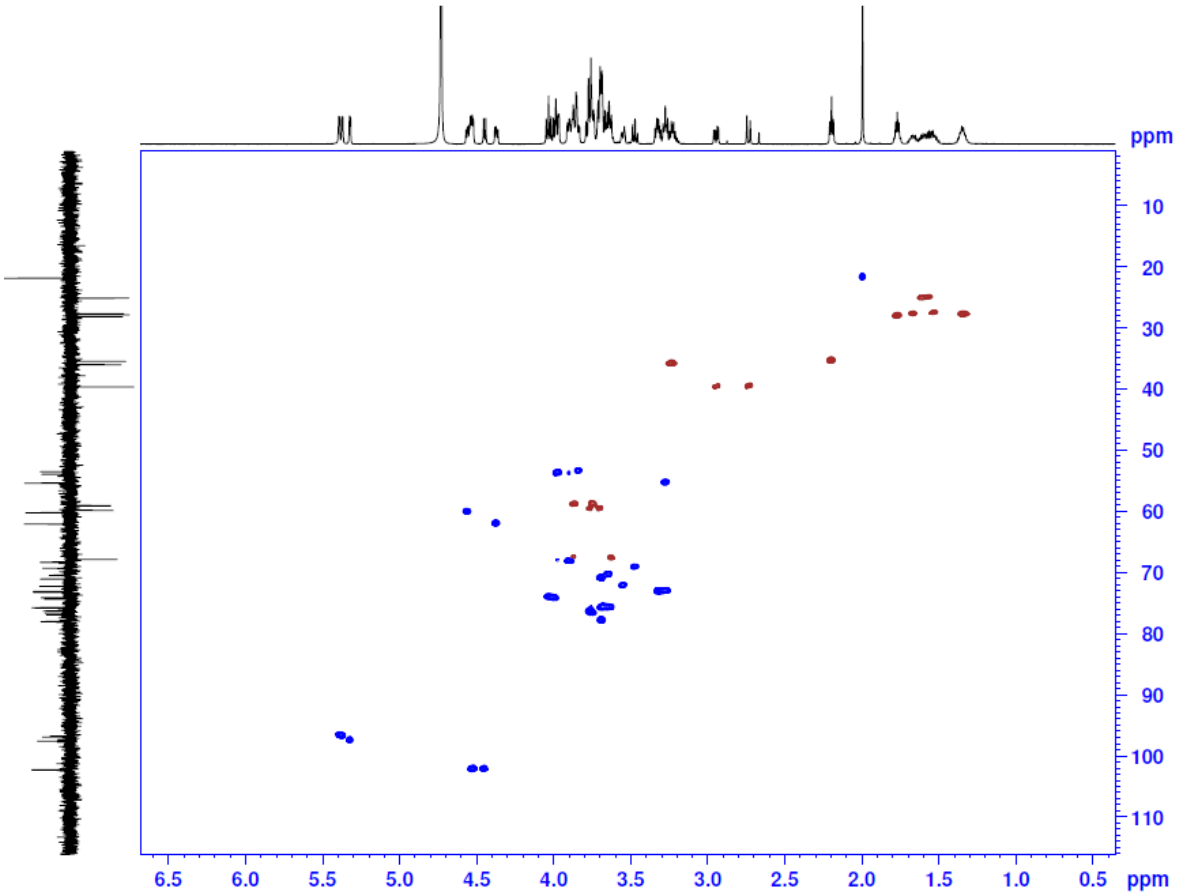
^{1}H-NMR of GlcNAc-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 10 (600 MHz, D_2O)



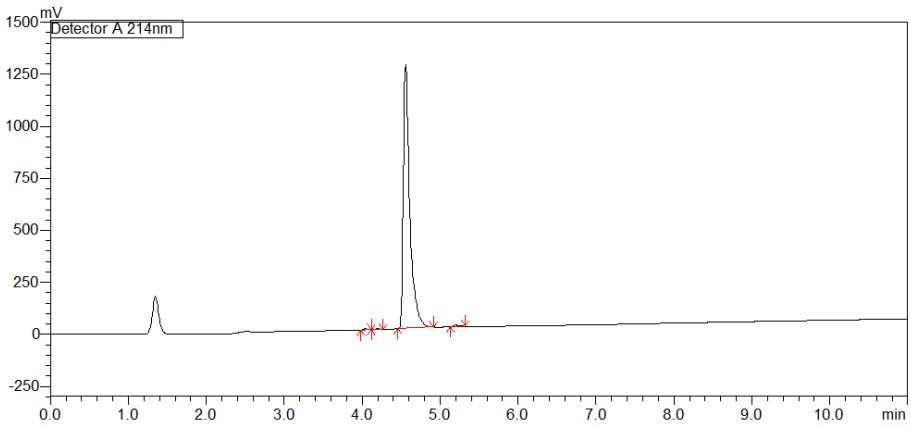
^{13}C-NMR of GlcNAc-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 10 (100 MHz, D_2O)



HSQC of GlcNAc-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 10 (600 MHz, D_2O)



HPLC spectra of GlcNAc-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 10



Purity: 99.02%. HPLC instrument: Shimadzu Prominence 20A coupled with UV detector; Detection: UV absorption at 214 nm; Column: Agilent Plus C18 (4.6 × 100 mm, 3.5 μm). Mobile Phase: A, H_2O (containing 0.05% TFA, v/v). B, Acetonitrile (containing 0.05% TFA, v/v). Gradient elution: B from 0 to 60% in 10 min. Flow rate: 1 mL/min.

HRESIMS of GlcNAc-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-(1,4)-GlcNTFA-α-(1,4)-GlcA-β-biotin 10