Supporting Information

Convergent Chemoenzymatic Synthesis and Biological Evaluation of a Heparan Sulfate Proteoglycan Syndecan-1 Mimetic

Jia Gao,[†] Yongmei Xu,[‡] Jian Liu,[‡] and Xuefei Huang[*, †, ¶, ∇]

[†]Department of Chemistry, [¶]Department of Biomedical Engineering, and [∇]Institute for Quantitative Health Science and Engineering, Michigan State University, East Lansing, Michigan 48824, United States

[‡]Department of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States

Email: huangxu2@msu.edu

Supporting Information

Table of Contents

Materials and Methods S3

- Materials S3
- Synthesis of Oligosaccharide **4** S3
- General Procedure for Automated Solid-Phase Peptide Synthesis S4
- High-Performance Liquid Chromatography S5
- Sortase A Expression, Purification and Quantification S5
- General Procedure for Sortase A-Mediated Ligation S6
- Size-Exclusion Purification of HS Glycopeptide S6
- BLI Binding Experiment S6
- Wound-Healing Assay Procedure S6
- Identification of Ligand Binding Sites and Biomolecule Visualization S7

Solid-Phase Synthesis of Synstatin Peptide **3** S8

Wound-Healing Assay Result S9

Computer Docking Simulation Result and Biomolecule Size Comparison S10-11

Product Characterization Data and Spectra S12

- Compound **1** S12
- Compound **2** S18
- Compound **3** S27
- Compound **5** S29

References S31

Materials and Methods

Materials. Sortase A-expressing BL21 cells were obtained from Prof. Xue-long Sun (Cleveland State University, OH). Gibco LB broth and LB agar were purchased from Thermo Fischer Scientific (Waltham, MA). Nickel columns and Nickel resins were purchased from Bio-rad (Hercules, California). SDS-PAGE gels and 10x Tris/Glycine/SDS electrophoresis buffer were purchased from Bio-rad (Hercules, California). Tris-HCl buffer was purchased from MilliporeSigma (St. Louis, MO). Sephadex G-15 and G-25 were purchased from MilliporeSigma (St. Louis, MO). EZ-Link™ Sulfo-NHS-LC-Biotin was purchased from Thermo Fischer Scientific (Waltham, MA). Recombinant human integrin αvβ3 was purchased from R&D Systems (Minneapolis, MN). Heparin sodium salt was purchased from MilliporeSigma (St. Louis, MO). MDA-MB-231 breast carcinoma cells were obtained from Prof. Kathy Gallo (Michigan State University, MI). DMEM (Dulbecco's Modified Eagle Medium) was purchased from MilliporeSigma (St. Louis, MO). Fetal Bovine Serum was purchased from Thermo Fischer Scientific (Waltham, MA). Human EGF was purchased from Alomone labs Ltd. (Jerusalem, Israel). Human vitronectin protein was purchased from R&D Systems (Minneapolis, MN).

Synthesis of Oligosaccharide 4. The starting compound **6** was dissolved in H_2O (5 mg/ml), to which Pd/C (10 mg/ml) was added. The mixture was then placed under a hydrogen balloon and stirred at room temperature for 1 h. After completion of the reaction, the mixture was filtered through a PTFE syringe filter (0.2 mm, 13 mm). The filtrate was concentrated, and the desired product **7** was subjected to a Sephadex G-10 column, after which it was dissolved in aqueous solution of $NaHCO_3$ at pH 8.5. 1.5 equivalents of 6-azidohexanoic acid NHS ester **10** in anhydrous *N,N*-dimethylformamide (DMF) was added. The reaction was then stirred at room temperature for 6 hours. Upon completion, the reaction mixture was directly loaded onto a Sephadex G-10 column for purification.

General Procedure for Automated Solid-Phase Peptide Synthesis. All the peptides reported were synthesized on a Liberty Blue™ Automated Microwave Peptide Synthesizer following standard Fmoc-based solid-phase peptide synthesis protocol. The 2-chlorotrityl resins with or without Fmoc-amino acid loaded were purchased from Chem-Impex. The Liberty Blue software from CEM Corporation was used to program the synthesis, including resin swelling, amino acid loading, couplings and Fmoc- removals. Commercially available DMF from Fischer Chemical was supplied to the synthesis module as reaction and washing solvent. Peptide synthesis was enabled by sequential couplings of Fmoc-amino acid (purchased from Chem-Impex) which was

preactivated by *N,N,N′,N′*-tetramethyl-O-(1H-benzotriazol-1-yl) uronium hexafluorophosphate (HBTU), *N*-hydroxybenzotriazole (HOBt), *N,N*-diisopropylethylamine (DIPEA), at 50 °C for 10 min, and deprotections with 20% piperidine in DMF at 60 °C for 4 min. In-between each coupling/deprotection step, resin-bound peptide was thoroughly washed with DMF. For the synthesis of Fmoc-Gly5-OH peptide, Fmoc-glycine was preactivated by 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU) and DIPEA instead. Resin-bound peptides were cleaved off the solid support with a cocktail solution of trifluoroacetic acid (TFA), triisopropylsilane (TIPS) and water (TFA/TIPS/H_2O, 95:2.5:2.5). The crude peptides were then purified with reverse-phase C18 preparative HPLC. Compound purity was each confirmed with chromatograms from C18 analytical HPLC.

High-Performance Liquid Chromatography. LC-8A Solvent Pumps, DGU-14A Degasser, SPD-10A UV-Vis Detector, SCL-10A System Controller (Shimadzu Corporation, JP) and Vydac 218TP 10 μm C18 Preparative HPLC column (HICHROM Limited, VWR, UK) or 20RBAX 300SB-C18 Analytical HPLC column (Agilent Technologies, CA) were installed for HPLC purifications with HPLC-grade acetonitrile (EMD Millipore Corporation, MA) and Milli-Q water (EMD Millipore Corporation, MA). A variety of eluting gradients were set up on LabSolutions software. Dual-wavelength UV detector was set to 220 nm and 254 nm for monitoring the absorbance from amide and Fmoc respectively. The eluted compounds were analyzed with ESI-MS to confirm their identities. The aqueous solutions of purified compounds were then lyophilized to obtain the dry solid.

Sortase A Expression, Purification and Quantification. Sortase A-expressing BL21 competent cells (2 μL) was transferred onto a kanamycin/chloramphenicol containing petri dish and incubated at 37 °C overnight. One colony of BL21 cells was picked and inoculated into 10 mL starter culture, which contained kanamycin (35 mg/L). The cell culture was incubated at 37 °C for 12-16 hours until OD_{600} value reached 0.85, and transferred into autoclaved culture medium (1L, 35 mg/L kanamycin). After roughly 5 hours, the OD_{600} reached 0.85. IPTG (0.5 mM) was added to induce protein expression. The cells were further incubated for another 4 hours at 37 °C, and centrifuged at 4 °C, 5000 rpm for 10 min. The cells were suspended into the lysis buffer (40 mL, 20 mM Tris, 250 mM NaCl, pH 8.0), then lysed through sonication. The lysate was centrifuged at 20,000 g for 20 min. The supernatant was purified by Nickel column (a. washing buffer: 20 mM Tris, 0.5 M NaCl and 250 mM imidazole; b. eluting buffer: 20 mM Tris, 0.5 M NaCl and 250 mM imidazole). Dialysis in buffer (20 mM Tris, 150 mM NaCl, pH 8.0) was performed to harvest the protein solution. Protein purity was confirmed with SDS-

PAGE gel electrophoresis and standard Bradford assay was used to determine the concentration and the expression yield of sortase A.

General Procedure for Sortase A-Mediated Ligation. The pH of the 10X reaction buffer (500 mM Tris-HCl, 1.5 M NaCl, 50 mM $CaCl_2$, 5 mM mercaptoethanol and 2mM Ni(II) sulfate) was adjusted to 8.5 with the addition of NaOH or HCl. Ligation substrates (250-500 μM) were dissolved and added into Tris-HCl reaction buffer, followed by the addition of sortase A. The reaction vessel was kept at 25 °C until reaction completion, with the reaction progress monitored with LC-MS. After the reaction, the enzyme was deactivated and precipitated out with addition of ethanol. Clear aqueous solutions from centrifuge were loaded onto G-15/G-25 size exclusion column for purification.

Size-Exclusion Purification of HS Glycopeptide. Samples were prepared in distilled water (500 μL) and then slowly transferred onto a G-15 or G-25 size-exclusion column. Fractions of 1 mL elute were collected. Fractions containing desired compounds were identified by ESI-MS analysis. The purified compounds were lyophilized to obtain the dry solid.

BLI Binding Experiment. BLI Octet K2 instrument (ForteBio, Molecular Devices, CA) was installed for binding experiments. Polypropylene black 96-well plates (Greiner Bio-one, Austria) and streptavidin (SA) sensor chips (ForteBio, Molecular Devices, CA) were used to assist sample preparations and detections of binding activities. Assay buffer was phosphate buffered saline (PBS) unless otherwise noted. Integrin αvβ3 protein solutions were prepared according to the assay design. To prepare biotinylated analytes, 1 mM of each amine-containing ligand compound and EZ-Link™ Sulfo-NHS-LC-Biotin (1.2 equiv.) were added to 0.1 M $NaHCO_3$ solution (pH 8.5). The reaction was allowed to proceed overnight. Upon completion, reaction mixture was passed through G-10 column to remove unconjugated biotin reactant. Sensors were then loaded with biotin-labeled compounds. The binding activity (including association and dissociation) between ligand and protein was measured under the monitoring via BLI. Biotin was used as negative control for all BLI assays. The assay results were then processed through the licensed software. Various concentrations of protein were tested against each ligand to obtain the kinetic data. The curve fitting was achieved using a 2:1 heterogenous ligand binding model provided by the data-processing software.

Wound-Healing Assay Procedure. MDA-MB-231 breast carcinoma cells were cultivated in the 6-well plate until 90% confluency was reached. After 24-hour starvation with serum-free medium, wounds were created by scratching the monolayer by sterile P200 pipet tips. This process was done carefully to make sure that all wounds were similar in size. Microscopic

images (taken with Zeiss Axionvert 200 Pred Axio Observer microscopy, Boston Industries, Inc.) were collected right after the wounding process (T=0). Then, starving serum-free medium was replaced by the growth medium that contains certain concentration of the analyte (growth medium without analytes was treated as control group). And human EGF was added to stimulate cancer cell migrations. After 20-hour (T = 20) incubation at 37 °C, microscopic images were collected with same microscopy instrument. Images at T=0 and T=20 of the same site were processed with GraphPad Prism Version 5.0c to obtain the cell migration results.

Identification of Ligand Binding Sites. To initiate the search for ligand binding sites on the integrin αvβ3 protein surface, synstatin peptide $SSTN_{92-119}$ model was constructed *de novo* using an open computation platform developed by Tuffery group.[1] Integrin protein (PDB:4G1M) was used as the receptor reference to facilitate the model construction and improve the subsequent docking simulations. 200 rounds of independent model simulations with sOPEP force field were applied to get good-quality peptide conformation predictions. The best candidate models were selected for the ligand-receptor molecular docking simulations.

Hot spot on the protein surface for synstatin binding was identified through examining the docking results. For the heparan sulfate binding simulations, a heparan sulfate tetrasaccharide structure was utilized to identify potential HS binding sites on integrin αvβ3. After uploading the integrin coordinate file to ClusPro docking platform, binding simulations were initiated under the built-in 'Heparin Ligand' mode.[2] Simulation results were then visualized and processed with UCSF Chimera software to pinpoint potential HS binding sites.[3]

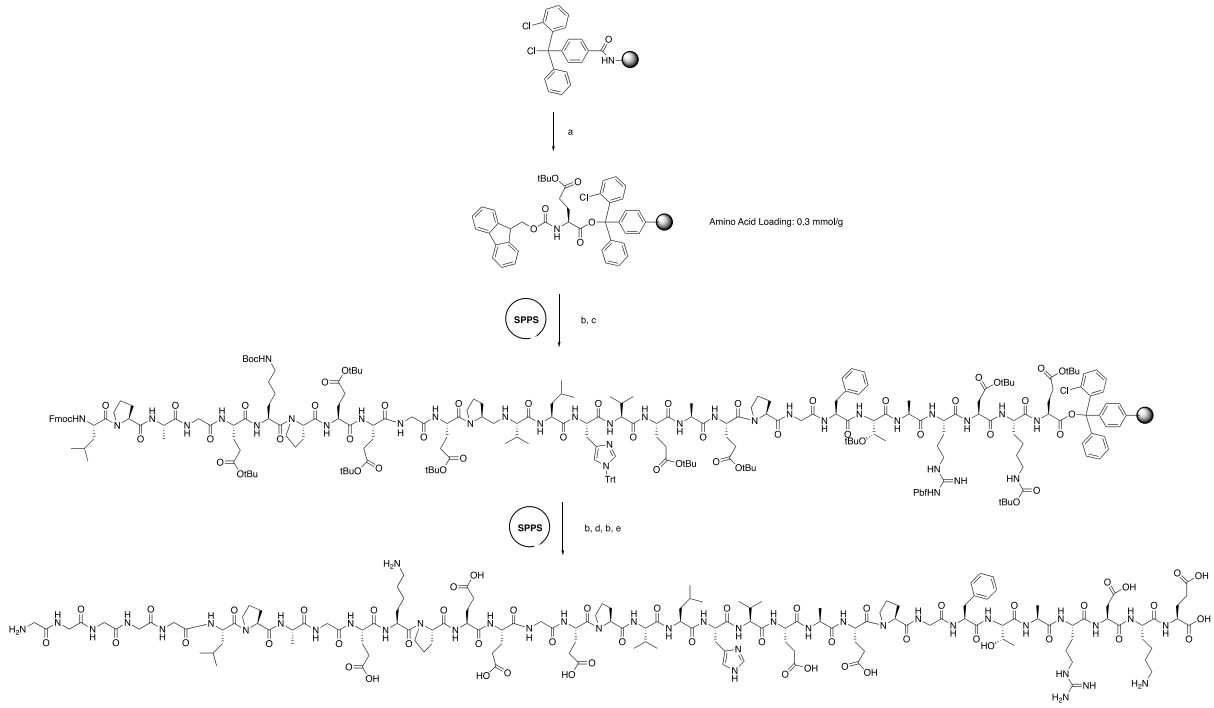
Biomolecule Visualization. The construction of syndecan-1 mimetic started with the heparan sulfate octasaccharide moiety. Its structure was prepared through 'GAG Builder' program at GLYCAM-Web.[4] Counter ions were added to the negatively charged sulfate groups and the HS octasaccharide was solvated into a cube of water molecules. Structural optimization was accomplished with GLYCAM force field. The generated PDB file of HS octasaccharide was later used to construct the glycopeptide.

Ab initio modeling of the peptide backbone was achieved with QUARK program.[5] Top model that adopted a more extended conformation was selected for the syndecan-1 mimetic construction.

The structure coordinates of HS octasaccharide and peptide backbone were input into Maestro software.[6] The artificial linkage connecting HS and peptide backbone was manually created. The resulted syndecan-1 mimetic structure was then optimized using all-atom minimization function to approximate its conformation.

The dimensions of syndecan-1 mimetic and integrin αvβ3 were measured with UCSF Chimera to provide a rough estimation of their sizes.

Solid-Phase Synthesis of Synstatin Peptide 3



Scheme S1. Solid-phase synthesis of $Gly_5SSTN_{92\text{-}117}$ peptide **3**. Reagents and conditions: (a) Amino acid loading: Fmoc-Glu(O-tBu)-OH, DIPEA, DMF; (b) Fmoc- cleavage: 20% piperidine/DMF, 50 °C, 2 min, microwave; (c) Amino acid coupling: 5 eq. Fmoc-AA-OH, HBTU, HOBt, DIPEA, DMF, 50 °C, 10 min, microwave; (d) Oligopeptide coupling: 5 eq. Fmoc-Gly5-OH **6**, HATU, DIPEA, DMF, 50 °C, 10 min, microwave; (e) Resin cleavage: TFA/TIPS/H_2O (95:2.5:2.5, v/v/v), 24 % overall.

Wound Healing Assay Result

Gly_5SSTN						
Concentration (μM)	Relative Migration Area (Artificial Unit)					
6	103.92	99.61	100.60	107.91	98.96	96.18
2	102.01	102.77	99.73	97.85	122.89	109.22
1	111.56	109.67	118.42	115.84	115.74	113.54
0	100.70	98.40	94.45	99.88	103.13	100.35

Heparin						
Concentration (μM)	Relative Migration Area (Artificial Unit)					
6	82.62	95.3	83.28	80.77	73.85	81.61
2	88.3	83.49	85.63	83.17	82.23	79.42
1	97.95	98.89	88.91	83.87	88	83.28
0	100.70	98.40	94.45	99.88	103.13	100.35

Syndecan-1 Mimetic						
Concentration (μM)	Relative Migration Area (Artificial Unit)					
6	65.04	65.37	80.11	52.96	65.27	71.59
2	82.81	84.61	92.50	81.69	82.47	80.88
1	82.26	85.77	103.42	94.08	106.81	84.19
0	100.70	98.40	94.45	99.88	103.13	100.35

Table S1. Summary of wound-healing assay results. The results presented were from 6 biological replicates.

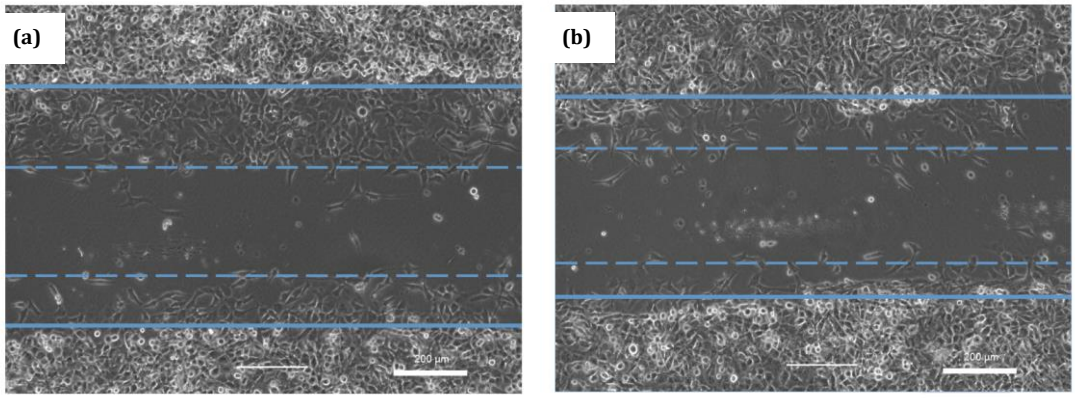


Figure S1. Representative microscopy images of MDA-MB-231 treated with (a) PBS as control and (b) glyco-polypeptide mimetic **1** (6 μM) after 20-hour incubation (solid lines for cell frontiers at T=0 and dashed lines for T=20 hours; 10X magnification; scale bar, 200 μm). The addition of glyco-polypeptide mimetic **1** significantly reduced the migratory abilities of the cancer cells, leaving larger relative wound areas.

Computer Docking Simulation Result and Biomolecule Visualization

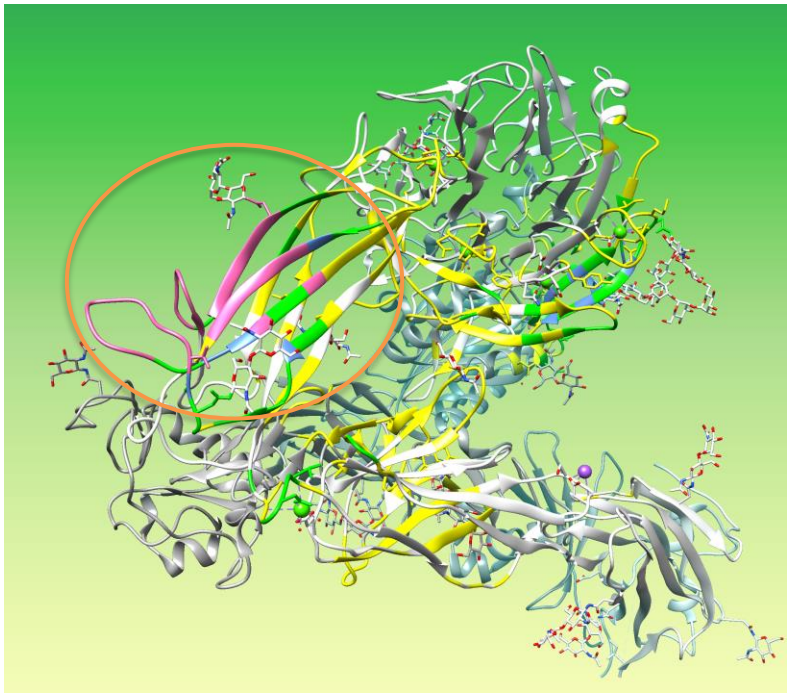


Figure S2. Identified synstatin peptide binding site (as circled) on the surface of integrin αvβ3 (PDB: 4G1M).

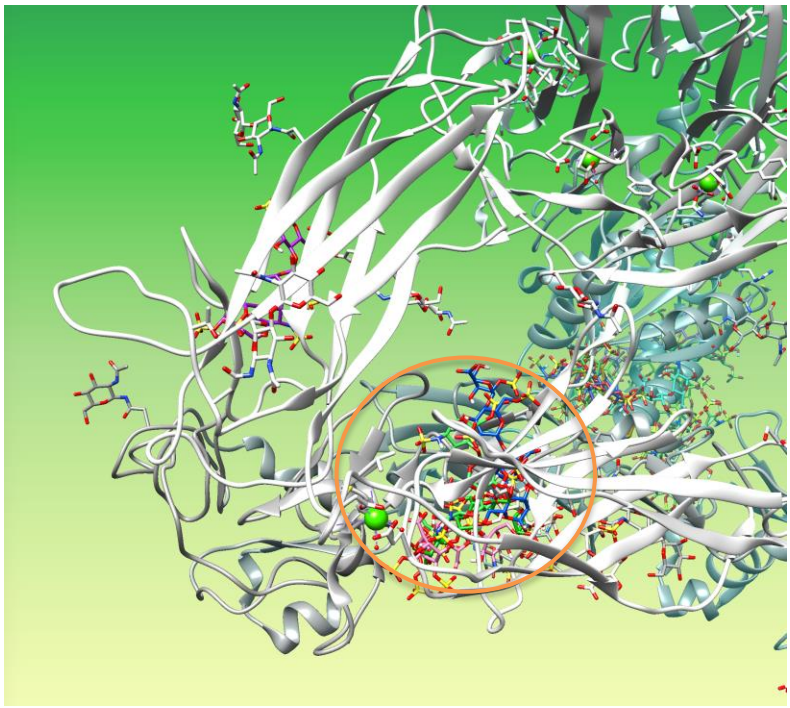


Figure S3. One of the identified heparan sulfate binding sites (as circled) on the surface of integrin αvβ3 (PDB: 4G1M).

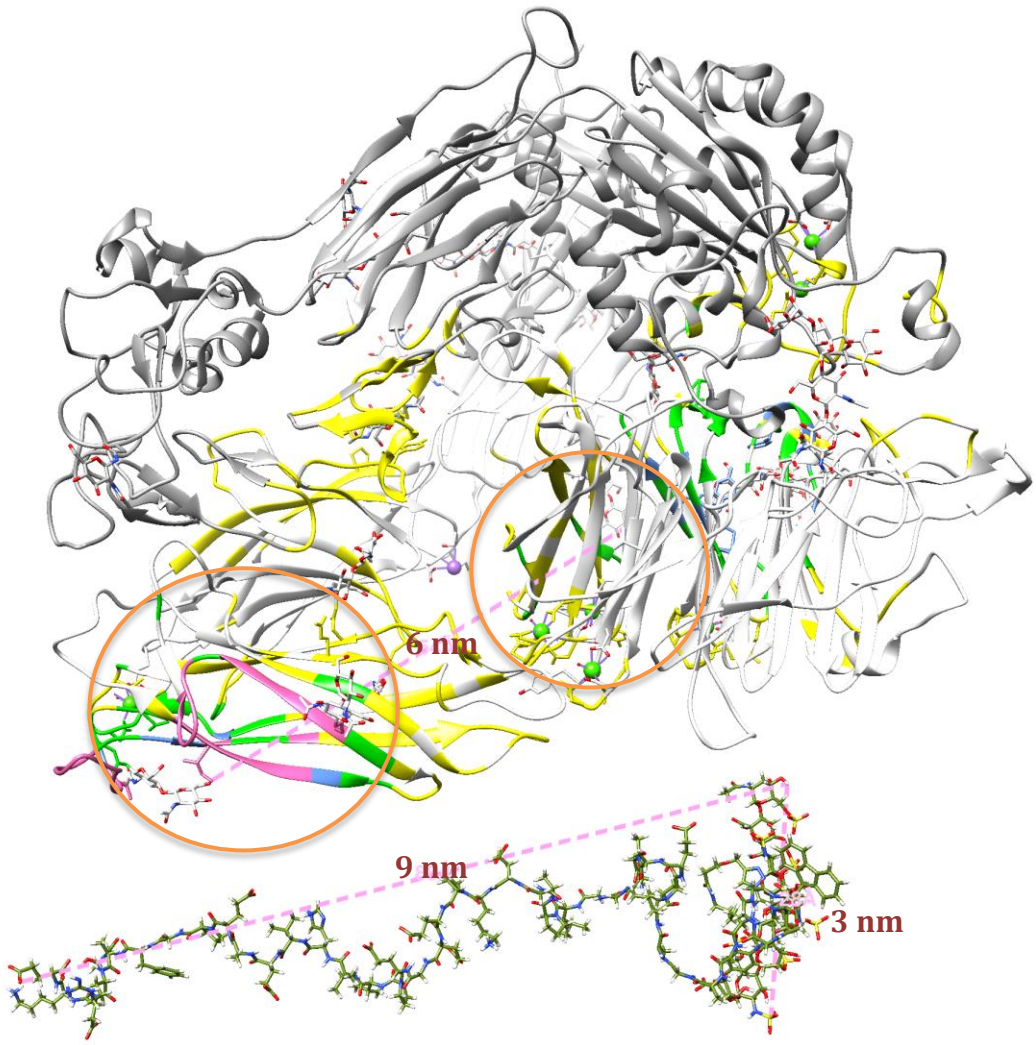


Figure S4. Biomolecule visualization and approximate size comparison of syndecan-1 mimetic **1** and integrin αvβ3 (PDB: 4G1M). Predicted binding areas of synstatin peptide and heparan sulfate are highlighted with orange circles.

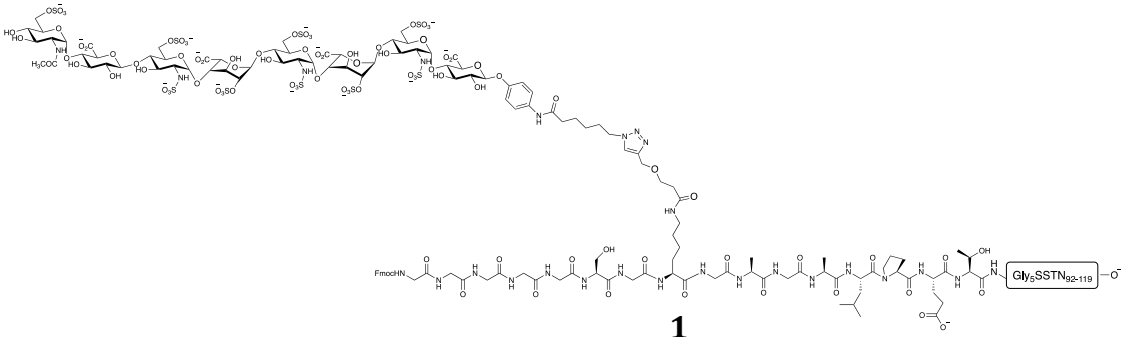
	Estimated Center-to-Center Distance
The Spotted Binding Sites	**6 nm**

Table S2. Measured estimated distance of the spotted synstatin and heparan sulfate binding sites

	Longitudinal	**Transversal**
Syndecan-1 Mimetics	**9 nm**	**3 nm**

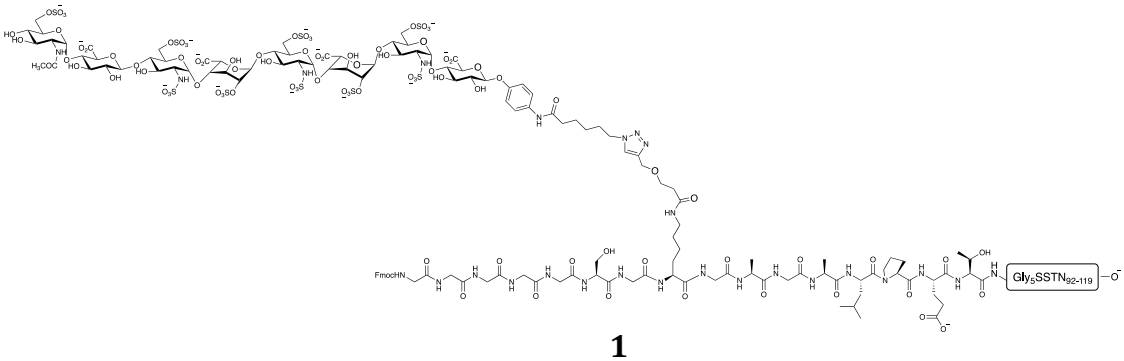
Table S3. Measured approximate dimensions of syndecan-1 mimetic.

Product Characterization Data and Spectra



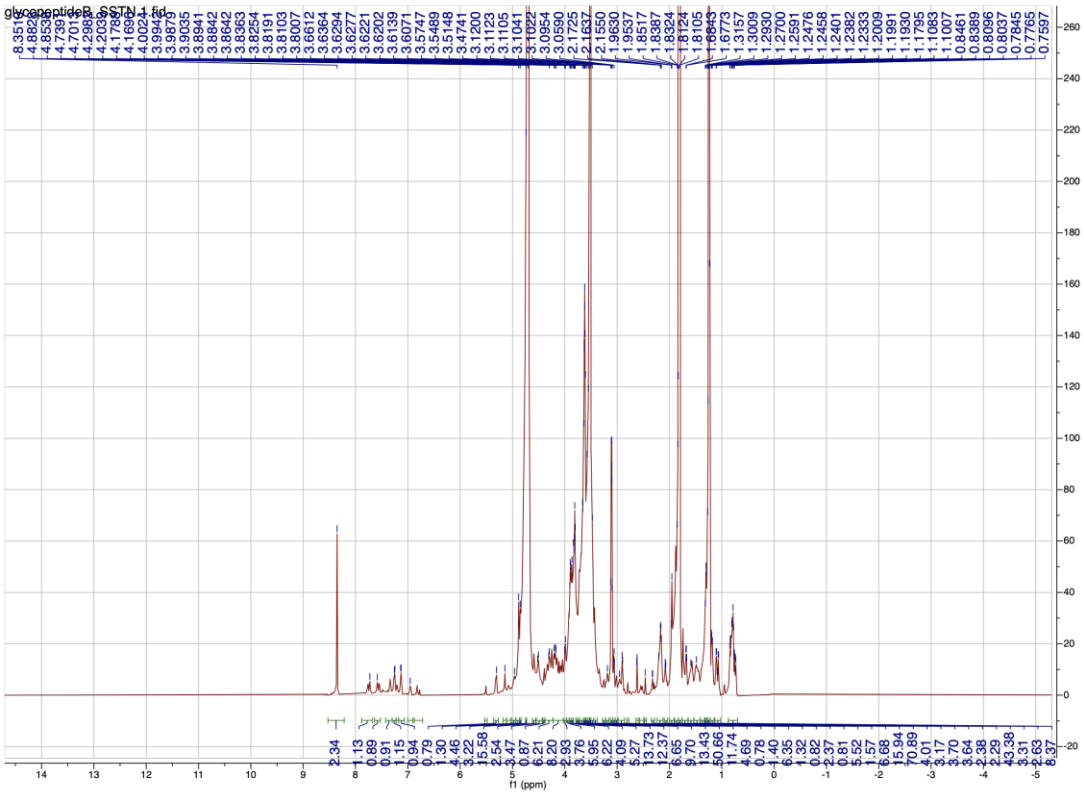
1

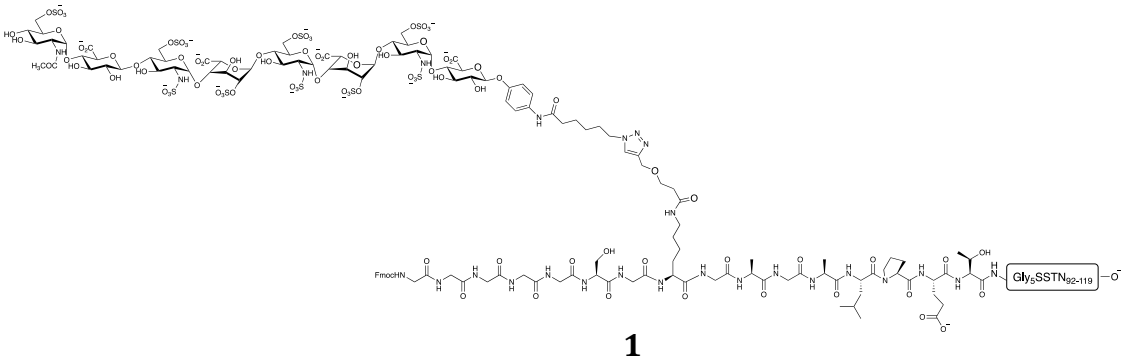
The purity of glycopeptide **1** was verified with analytical C-18 HPLC (5-100% acetonitrile/water; 0.1% trifluoroacetic acid). ^{1}H-NMR (900 MHz, D_2O), δ 0.70 – 0.88 (8 H, m), 1.02 – 1.13 (3 H, m), 1.14-1.22 (3 H, m), 1.22 – 1.27 (43 H, m), 1.27-1.28 (2 H, m), 1.31 – 1.33 (2 H, m), 1.33-1.42 (4 H, m), 1.42-1.54 (4H, m), 1.54-1.64 (3H, m), 1.65-1.76 (4 H, m), 1.76-1.81 (71 H, m), 1.83 – 1.90 (16 H, m), 1.91 – 2.00 (7 H, m), 2.03-2.10 (2H, m), 2.13-2.24 (6 H, m), 2.28-2.35 (1H, m), 2.49-2.57 (1H, m), 2.59-2.64 (1 H, m), 2.86-2.98 (2H, m), 2.99-3.04 (1 H, m), 3.04-3.08 (1H, m), 3.09 – 3.13 (6 H, m), 3.16-3.22 (1 H, m), 3.22-3.28 (1H, m), 3.38-3.44 (5 H, m), 3.44-3.50 (12H, m), 3.51-3.53 (51 H, m), 3.53 – 3.56 (13 H, m), 3.56-3.59 (10 H, m), 3.59-3.61 (7 H, m), 3.61 – 3.65 (12 H, m), 3.65-3.72 (14H, m), 3.72-3.75 (5H, m), 3.76-3.78 (4H, m), 3.83-3.87 (6 H, m), 3.88-3.92 (6H, m), 3.92-3.97 (4H, m), 3.97-4.02 (3H, m), 4.02-4.22 (8H, m), 4.22-4.39 (6H, m), 4.39-4.41 (1H, m), 4.42-4.55 (3H, m), 4.55-4.61 (3H, m), 4.77-4.82 (16 H, m), 4.82-4.86 (3 H, m), 4.88-4.92 (4 H, m), 4.94-5.01 (1H, m), 5.03-5.11 (1H, m), 5.11-5.18 (1H, m), 5.27-5.37 (1H, m), 7.07-7.17 (1 H, m), 7.23 – 7.30 (1 H, m), 7.30-7.42 (1 H, m), 7.52 – 7.64 (1 H, m), 7.67 – 7.88 (1 H, m), 8.30-8.40 (2 H, m). ^{13}C-NMR (225 MHz, D_2O), δ12.1, 12.2, 12.3, 12.8, 12.9, 13.3, 15.7, 16.4, 16.5, 16.5, 16.6, 16.7, 17.2, 17.7, 17.8, 17.9, 18.0, 18.2, 18.2, 18.4, 18.6, 18.7, 18.8, 19.3, 20.0, 20.6, 20.9, 21.4, 21.7, 21.9, 22.0, 22.0, 22.4, 22.4, 22.7, 23.0, 23.3, 24.0, 24.1, 24.2, 24.3, 24.4, 24.6, 24.7, 26.2, 27.1, 28.6, 28.9, 29.2, 29.3, 30.0, 30.2, 30.5, 33.4, 33.5, 33.7, 33.8, 34.0, 34.4, 35.9, 36.0, 37.0, 37.1, 38.6, 38.7, 38.9, 38.9, 39.0, 39.1, 39.8, 39.9, 40.4, 42.3, 42.4, 42.6, 43.1, 43.2, 43.2, 43.6, 43.6, 44.4, 44.5, 46.7, 46.8, 47.4, 47.9, 49.5, 49.6, 49.8, 50.1, 50.2, 50.3, 51.6, 51.7, 51.9, 52.0, 53.3, 534, 53.5, 53.6, 53.7, 53.8, 54.2, 54.3, 54.4, 55.0, 55.1, 55.2, 55.7, 55.8, 57.5, 57.7, 58.0, 58.2, 58.3, 58.5, 58.8, 59.0, 59.1, 59.2, 59.2, 59.4, 59.6, 60.3, 60.4, 60.5, 61.1, 61.2, 62.0, 62.4, 62.5, 62.8, 62.9, 64.0, 64.3, 65.4, 65.6, 66.0, 66.1, 66.2, 66.5, 66.8, 67.1, 67.2, 67.3, 67.4, 67.5, 68.6, 68.8, 68.9, 69.0, 69.1, 69.4, 69.5, 69.6, 69.8, 70.1, 70.2, 70.7, 71.1, 71.3, 71.4, 71.7, 71.8, 72.1, 72.5, 73.2, 73.3, 73.4, 74.0, 75.1, 75.3, 75.4, 75.7, 75.8, 75.9, 76.2, 76.4, 76.5, 79.9, 93.1, 95.2, 95.3, 95.6, 96.4, 96.5, 96.9, 97.6, 98.3, 98.4, 99.0, 100.1, 100.2, 101.6, 101.6, 116.8, 120.2, 123.4, 123.5, 125.0, 127.1, 127.2, 127.4, 128.0, 128.5, 128.6, 128.6, 129.1, 145.9, 146.5, 146.8, 146.9, 147.0, 147.1, 147.2, 147.3, 147.5, 147.6, 147.7, 148.0, 152.2, 152.3, 152.3. ESI-MS: $C_{277}H_{411}N_{65}O_{145}S_9$ $[M+10H]^{4-}$ calcd: 1451.0859, obsd: 1451.0818 (2.84 ppm).



1

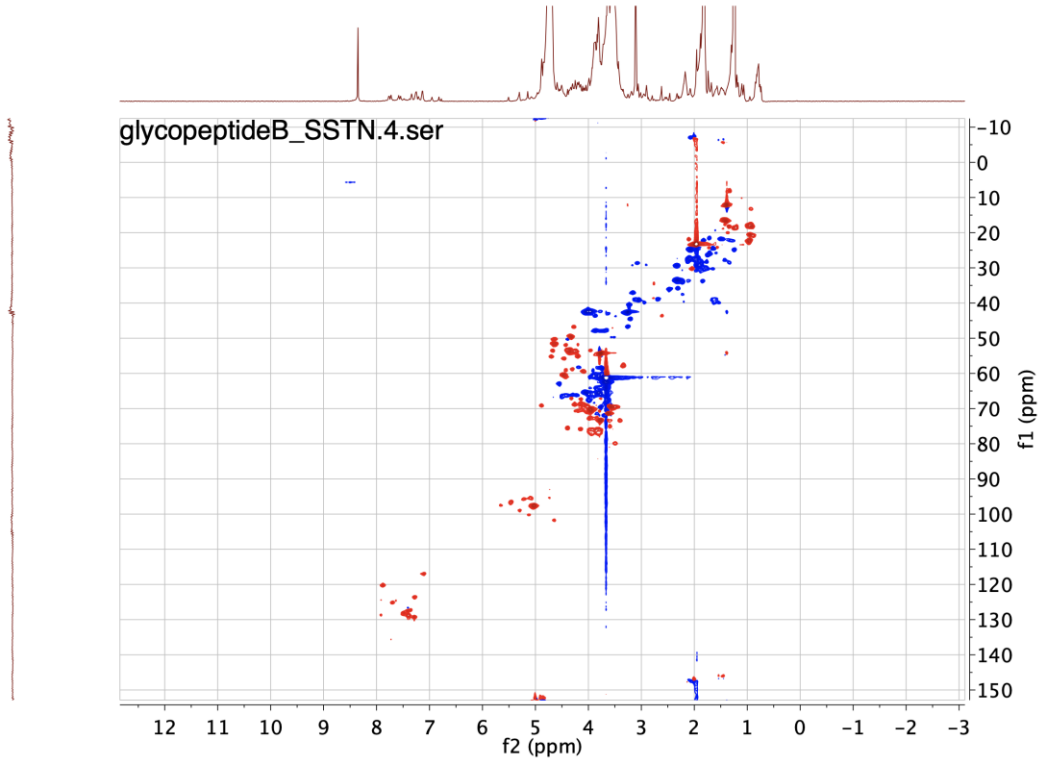
^{1}H-NMR of **1** (900 MHz D_2O)

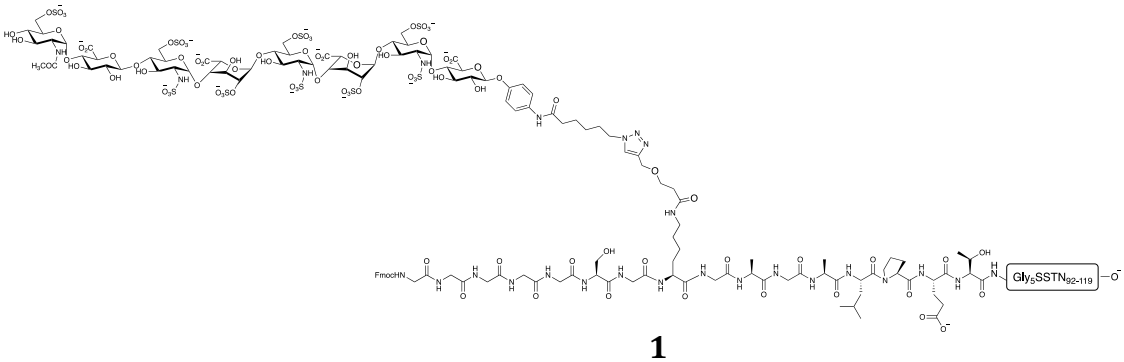




1

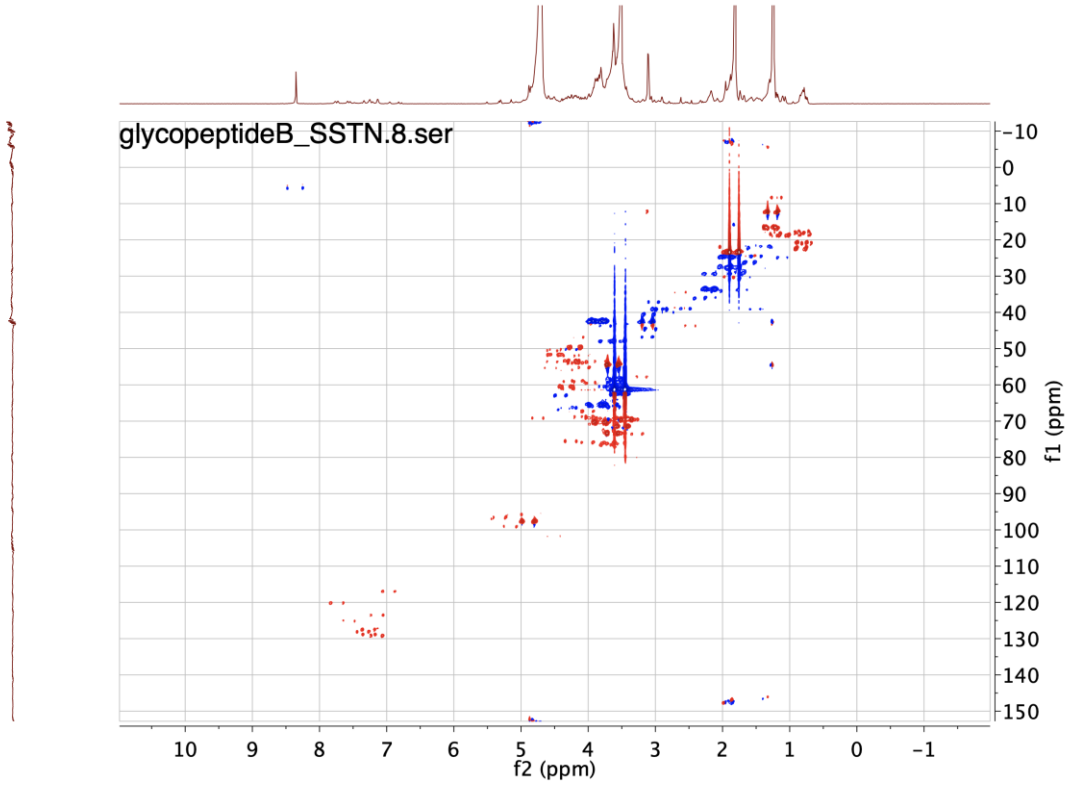
^{1}H-^{13}C gHSQCAD of **1** (900 MHz D_2O)

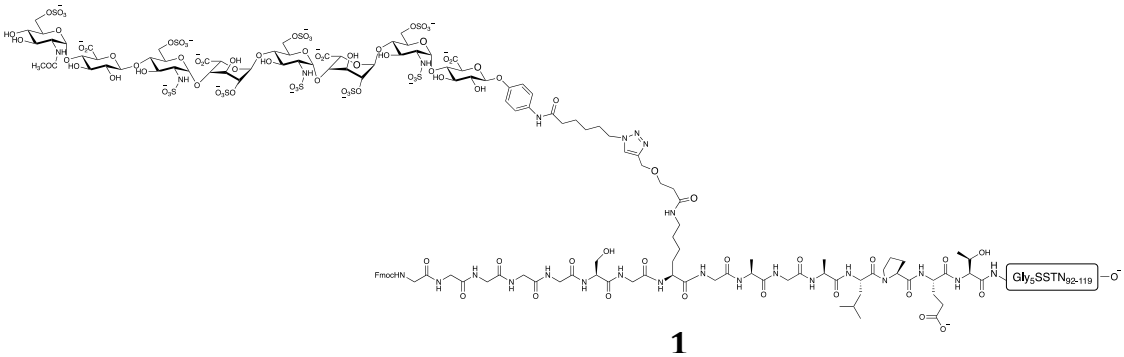




1

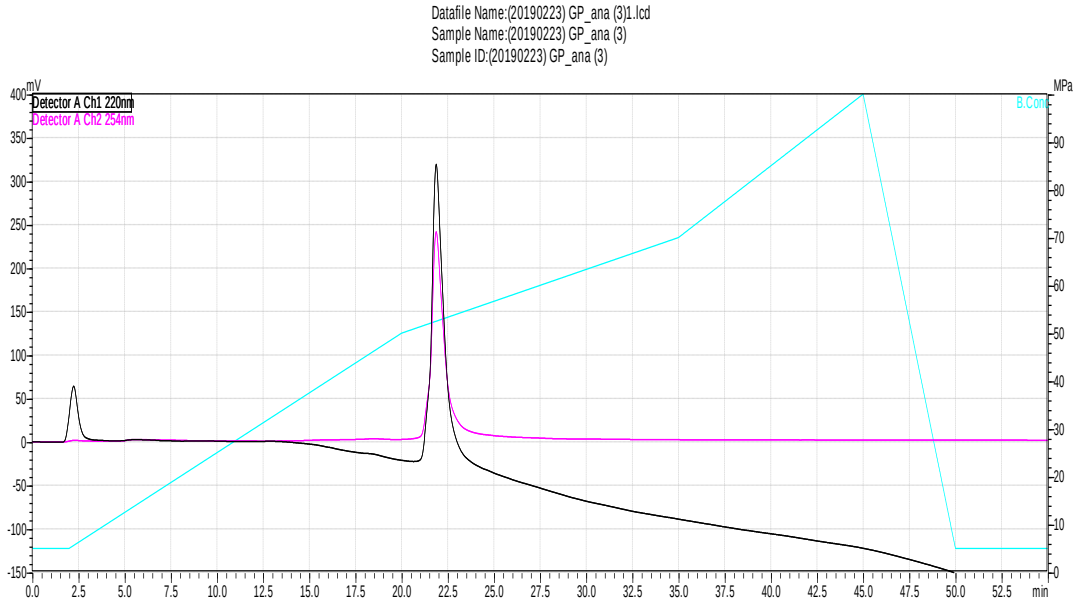
^{1}H-^{13}C Coupled gHSQCAD of **1** (900 MHz D_2O)

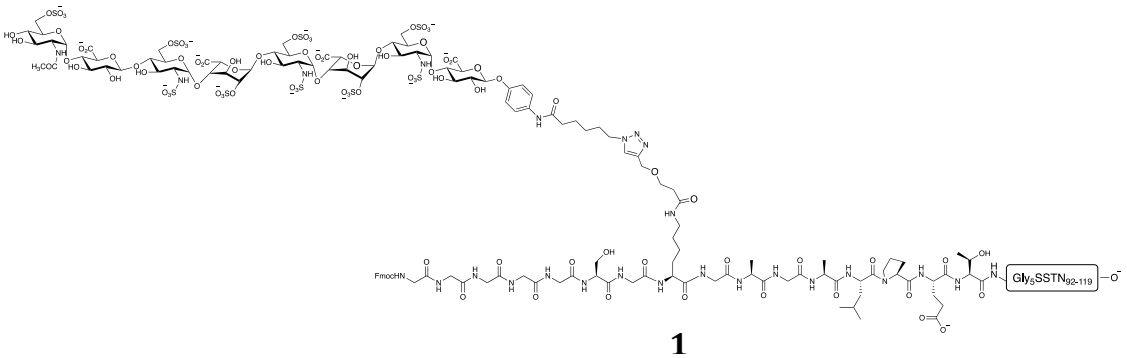




1

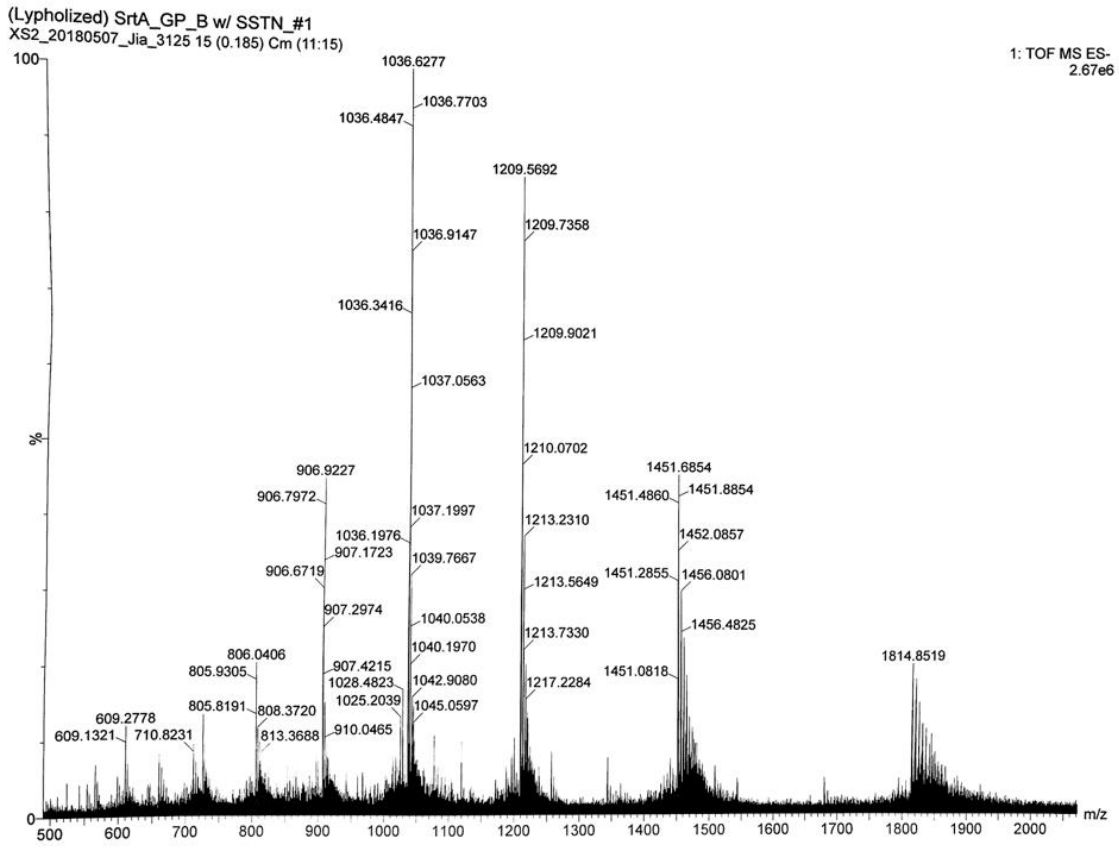
HPLC Chromatogram of **1**

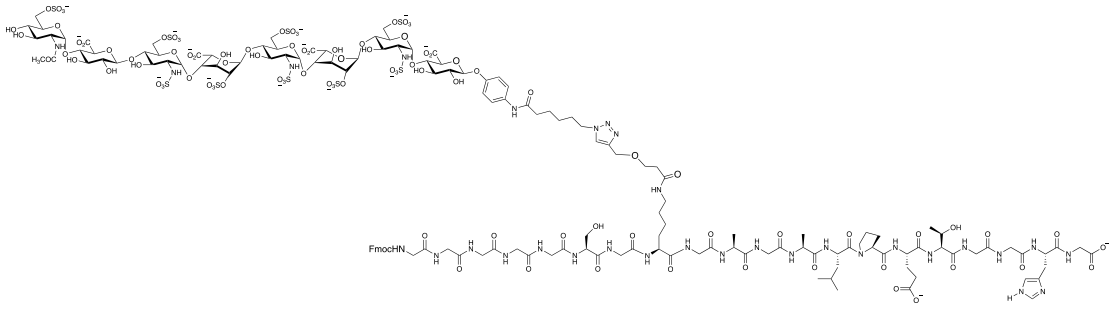



FmocHN
GlyS SSTN92-119

1

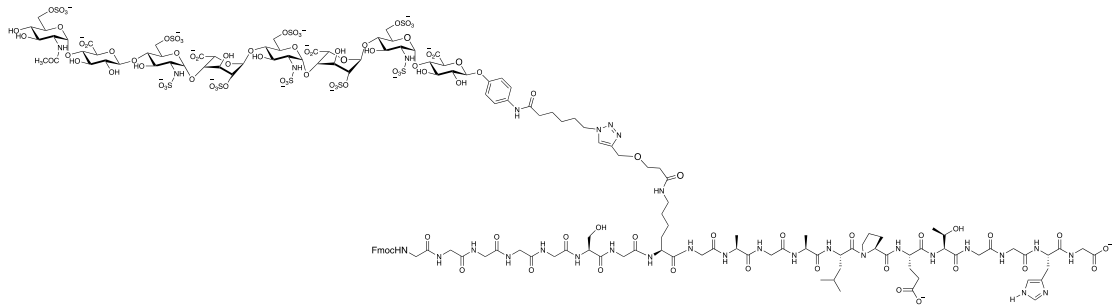
ESI-MS of **1**


(Lypholized) SrtA_GP_B w/ SSTN_#1
XS2_20180507_Jia_3125 15 (0.185) Cm (11:15)
1: TOF MS ES-
2.67e6
100
%
0
1036.6277
1036.7703
1036.4847
1209.5692
1209.7358
1036.9147
1036.3416
1209.9021
1037.0563
1210.0702
906.9227
1451.6854
906.7972
1451.4860
1451.8854
1037.1997
1213.2310
1036.1976
907.1723
1452.0857
1039.7667
906.6719
1213.5649
1451.2855
1456.0801
907.2974
1040.0538
1213.7330
1456.4825
806.0406
1040.1970
1814.8519
805.9305
907.4215
1451.0818
1028.4823
1042.9080
1217.2284
1025.2039
1045.0597
805.8191
808.3720
609.2778
910.0465
609.1321
710.8231
813.3688
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
m/z



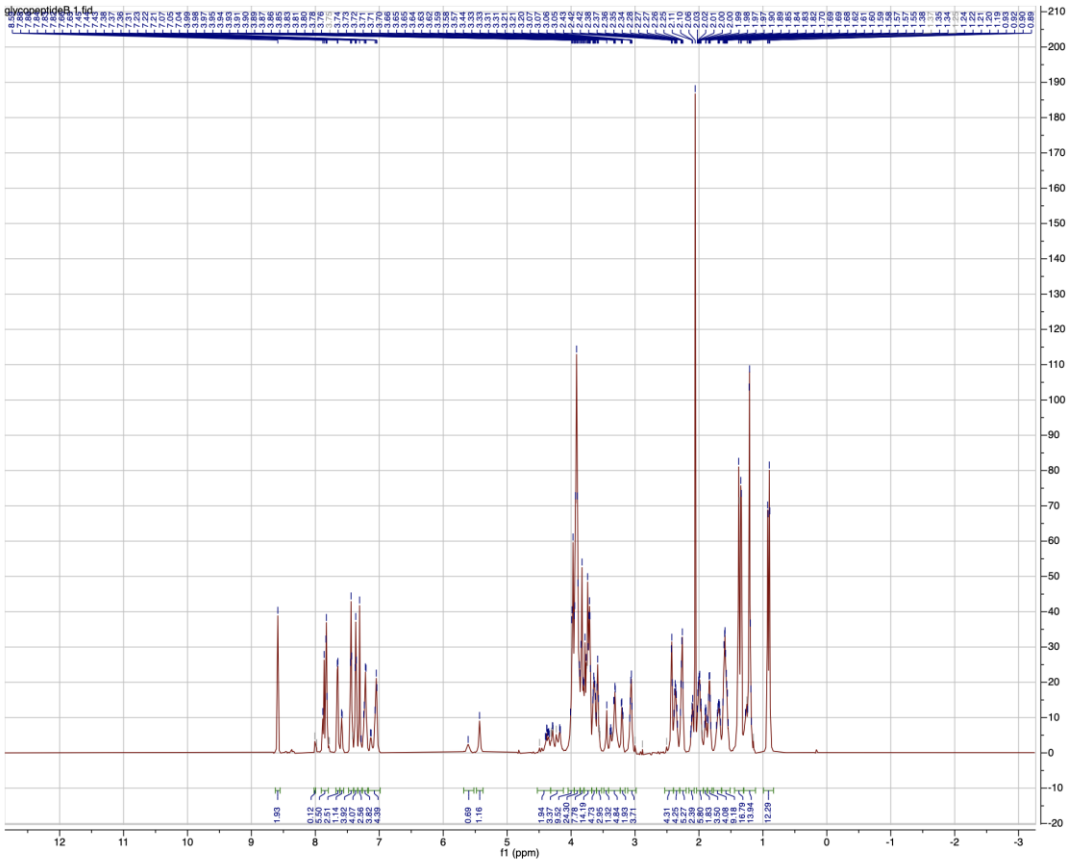
2

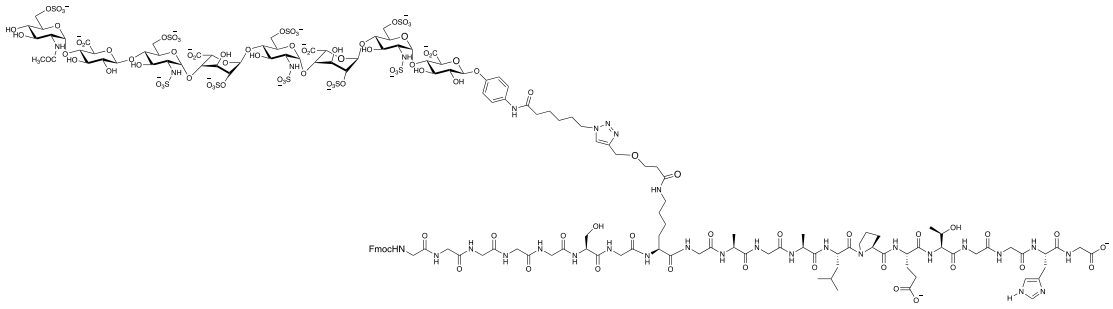
The purity of glycopeptide **2** was verified with analytical C-18 HPLC (5-100% acetonitrile/water; 0.1% trifluoroacetic acid). ^{1}H-NMR (900 MHz, D_2O), δ 0.91 (12 H, m), 1.12 – 1.30 (14 H, m), 1.30 – 1.44 (17 H, m), 1.51 – 1.64 (9 H, m), 1.65 – 1.77 (4 H, m), 1.84 (3 H, m), 1.89 (2 H, m), 2.00 (6 H, m), 2.11 (2 H, m), 2.27 (5 H, m), 2.36 (4 H, m), 2.40 – 2.53 (4 H, m), 3.06 (4 H, m), 3.20 (2 H, m), 3.23 – 3.41 (5 H, m), 3.44 (1 H, m), 3.58 (3 H, m), 3.64 (5 H,m), 3.67 – 3.80 (14 H, m), 3.84 (8 H, m), 3.92 (24 H, m), 3.95 – 4.05 (10 H, m), 4.12 – 4.32 (3 H, m), 4.37 (2 H, m), 5.43 (1 H, m), 5.61 (1 H, m), 6.99 – 7.17 (4 H, m), 7.23 (4 H, m), 7.31 (3 H, m), 7.37 (4 H, m), 7.44 (4 H, m), 7.59 (1 H, m), 7.65 (3 H, m), 7.80 – 7.91 (6 H, m), 8.59 (2 H, m). ^{13}C-NMR (225 MHz, D_2O), δ 16.3, 16.6, 18.7, 20.5, 21.9, 22.3, 22.4, 24.3, 24.4, 24.6, 24.8, 26.6, 26.9, 27.7, 28.8, 29.2, 30.3, 32.3, 35.8, 36.0, 38.9, 39.0, 42.3, 42.3, 42.5, 43.2, 43.6, 46.7, 47.7, 49.3, 49.9, 50.1, 50.2, 52.2, 53.5, 53.6, 53.9, 55.8, 57.5, 57.9, 58.0, 59.0, 60.4, 61.0, 62.8, 65.8, 66.1, 66.2, 66.8, 67.0, 68.8, 69.0, 69.3, 70.0, 70.5, 72.4, 73.4, 75.8, 76.2, 76.3, 77.0, 77.3, 97.0, 97.6, 100.2, 101.8, 116.9, 117.5, 120.1, 123.4, 124.5, 124.7, 125.0, 127.4, 128.0, 128.1, 131.8, 133.5, 140.8, 143.6, 143.8, 153.8, 170.8, 171.1, 171.3, 171.8, 172.7, 174.0, 174.2, 174.6. ESI-MS: $C_{146}H_{207}N_{31}O_{99}S_9$ $[M+11H]^{4-}$ calcd: 1066.4901, obsd: 1066.4874 (2.51 ppm).



2

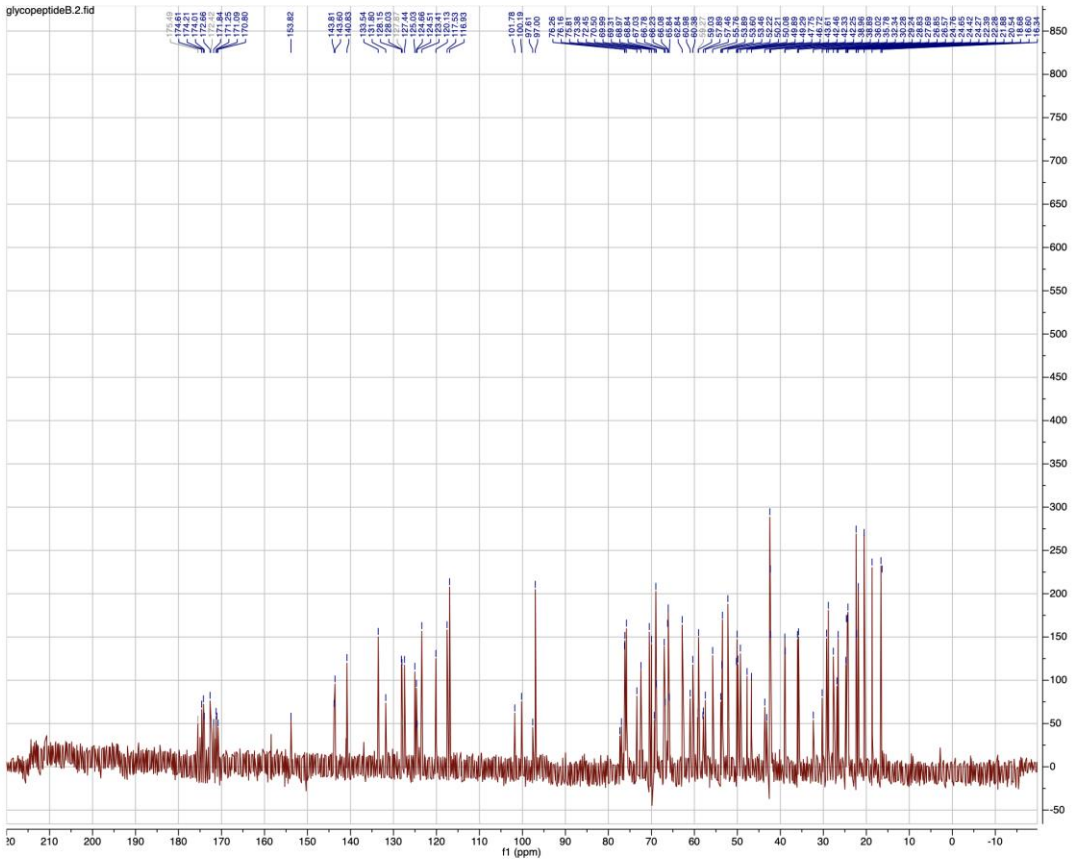
^{1}H-NMR of **2** (900 MHz D_2O)

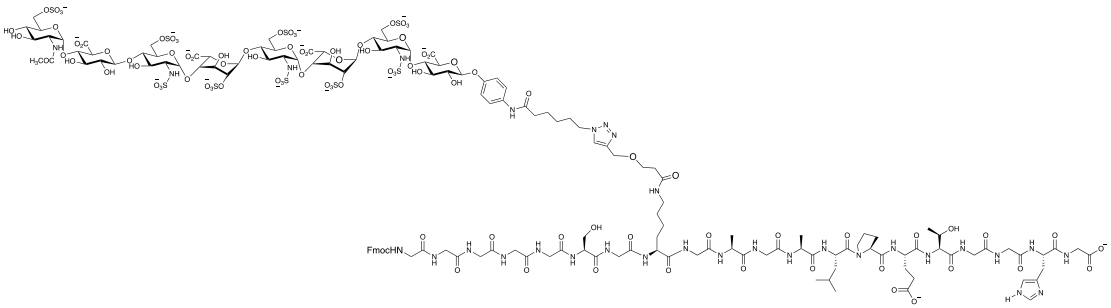




2

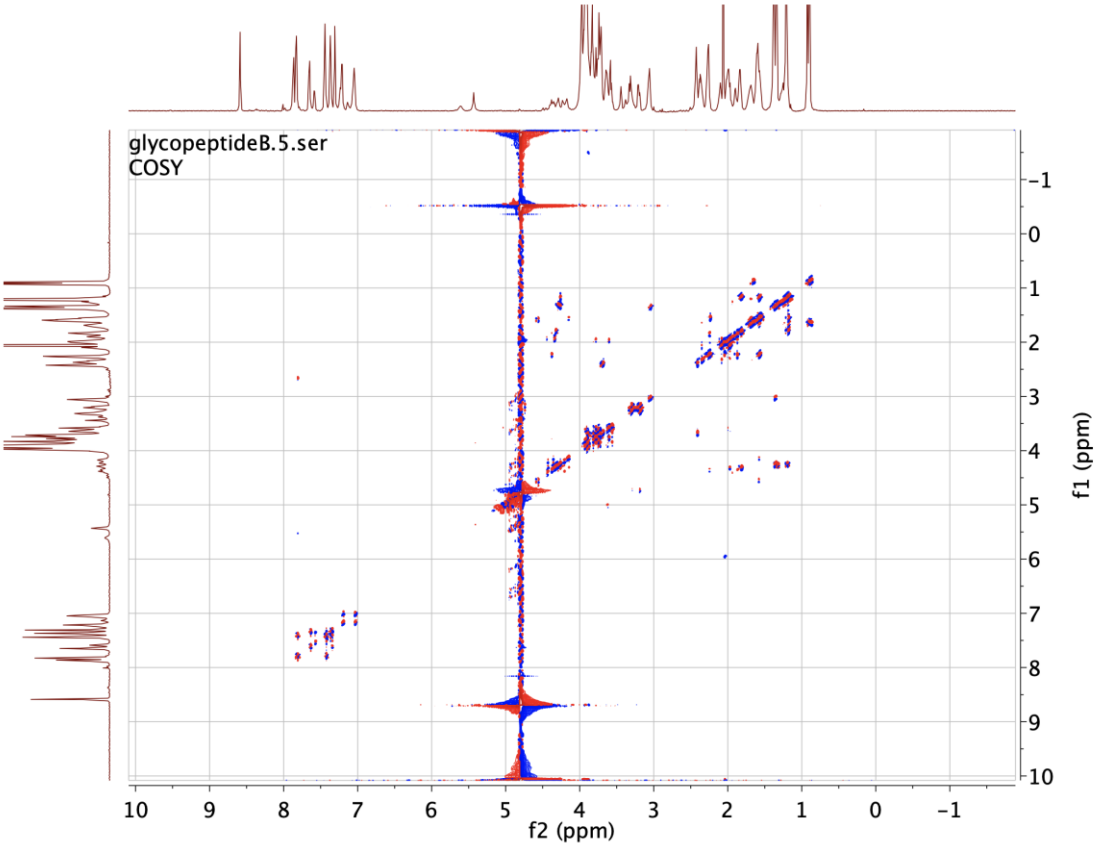
^{13}C-NMR of **2** (225 MHz D_2O)

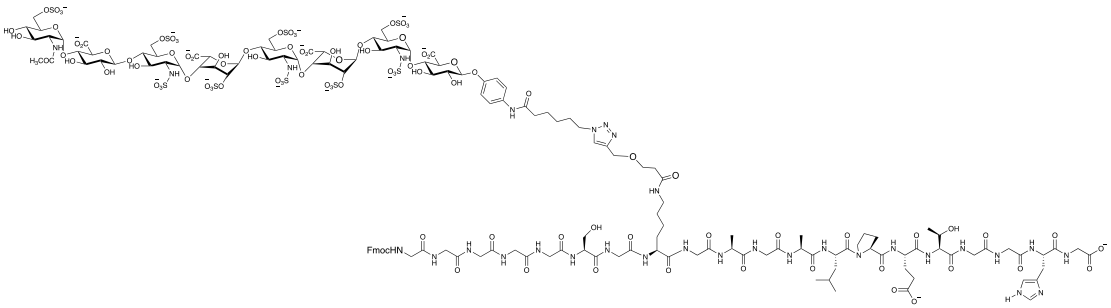




2

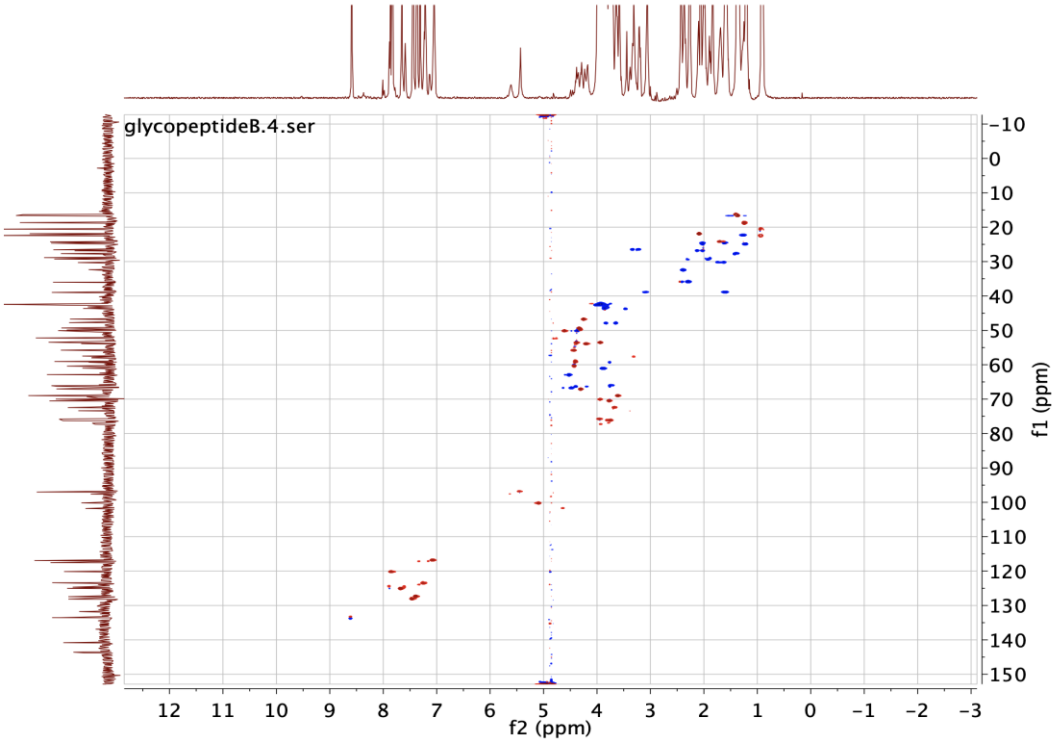
^{1}H-^{1}H gCOSY of **2** (900 MHz D_2O)

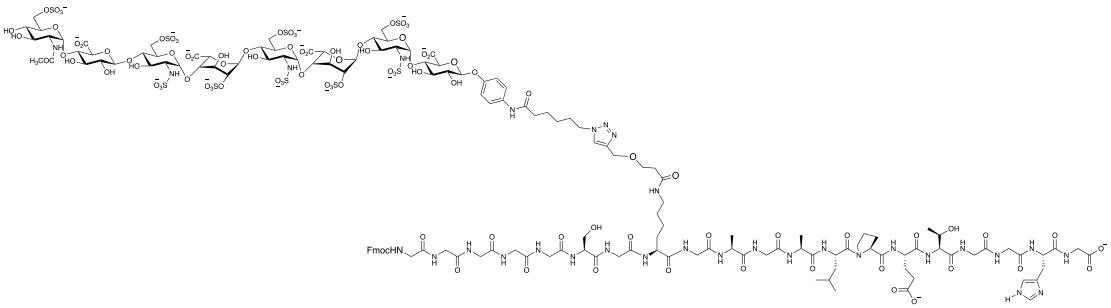




2

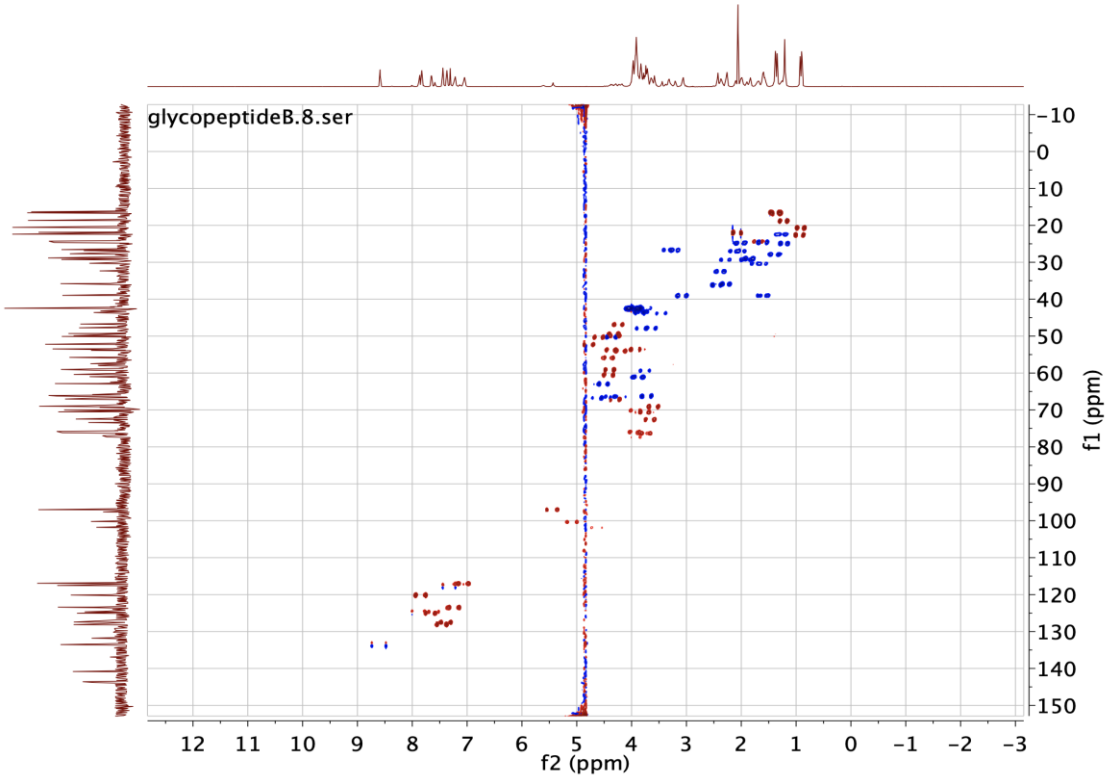
^{1}H-^{13}C gHSQCAD of **2** (900 MHz D_2O)

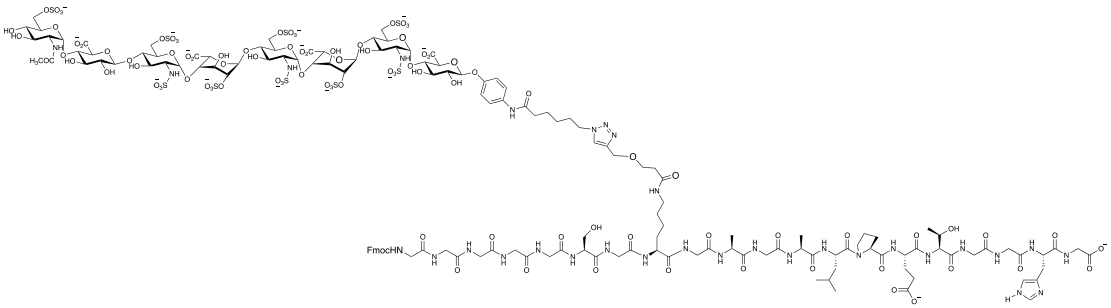




2

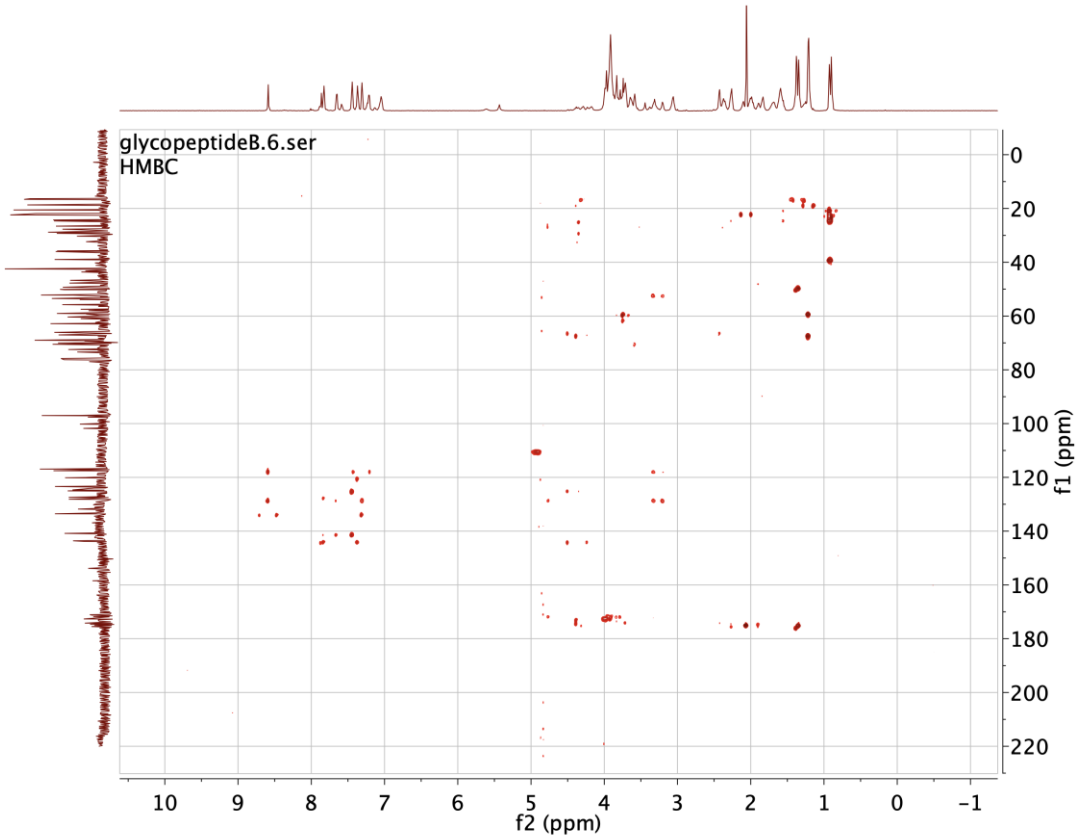
^{1}H-^{13}C Coupled gHSQCAD of **2** (900 MHz D_2O)

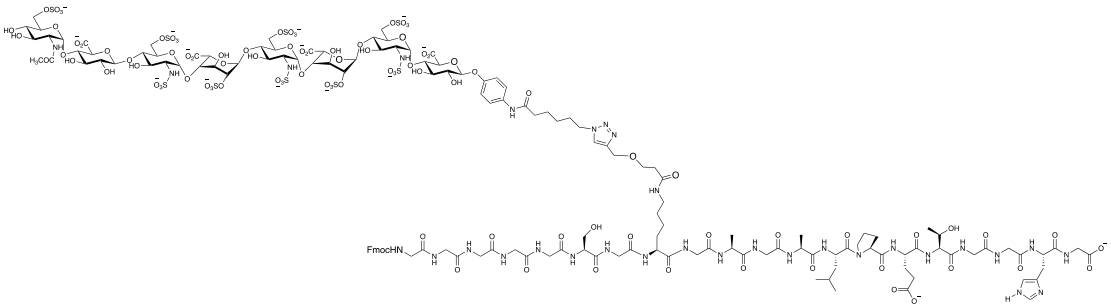




2

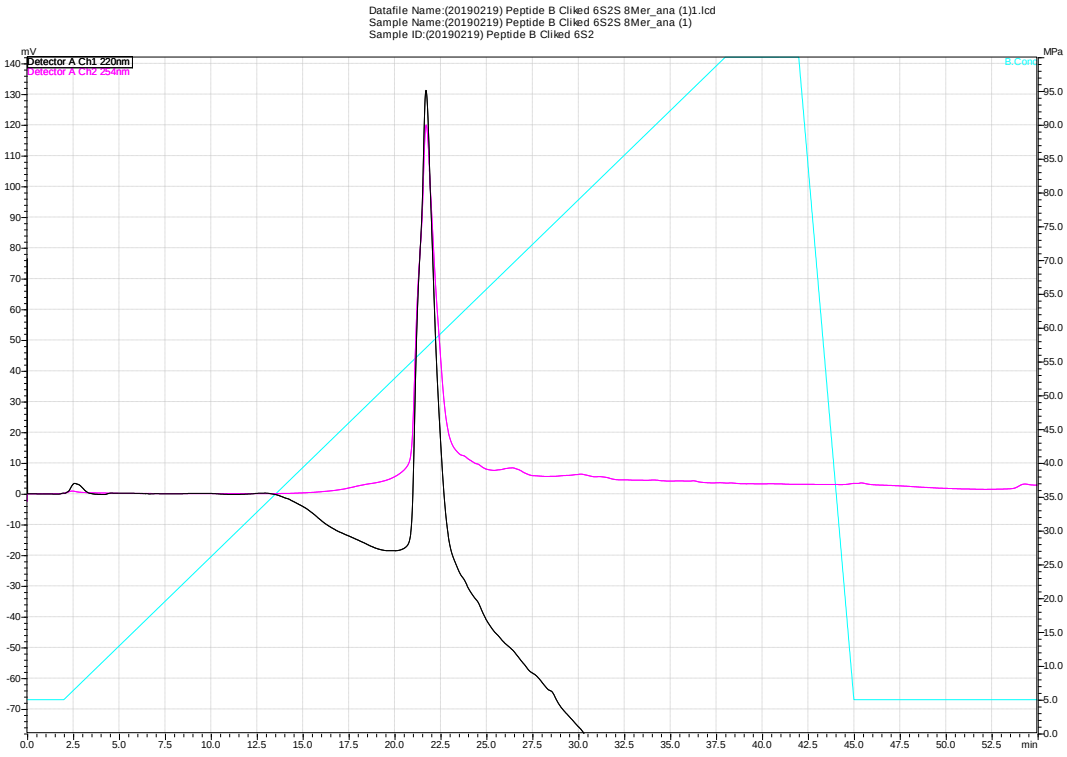
^{1}H-^{13}C gHMBC of **2** (900 MHz D_2O)

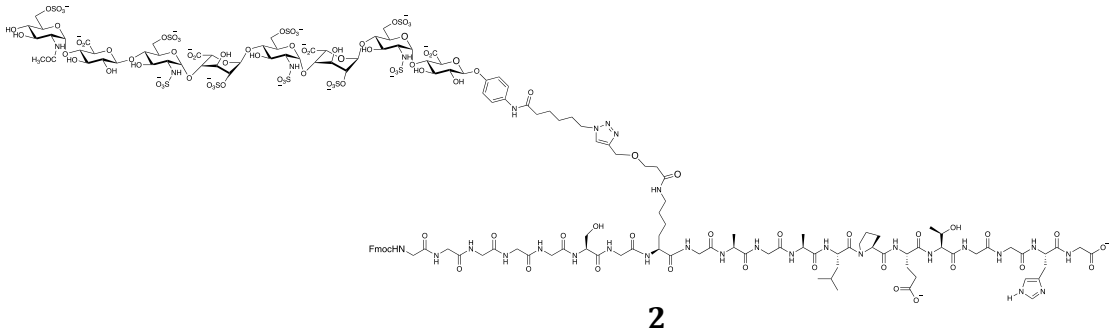




2

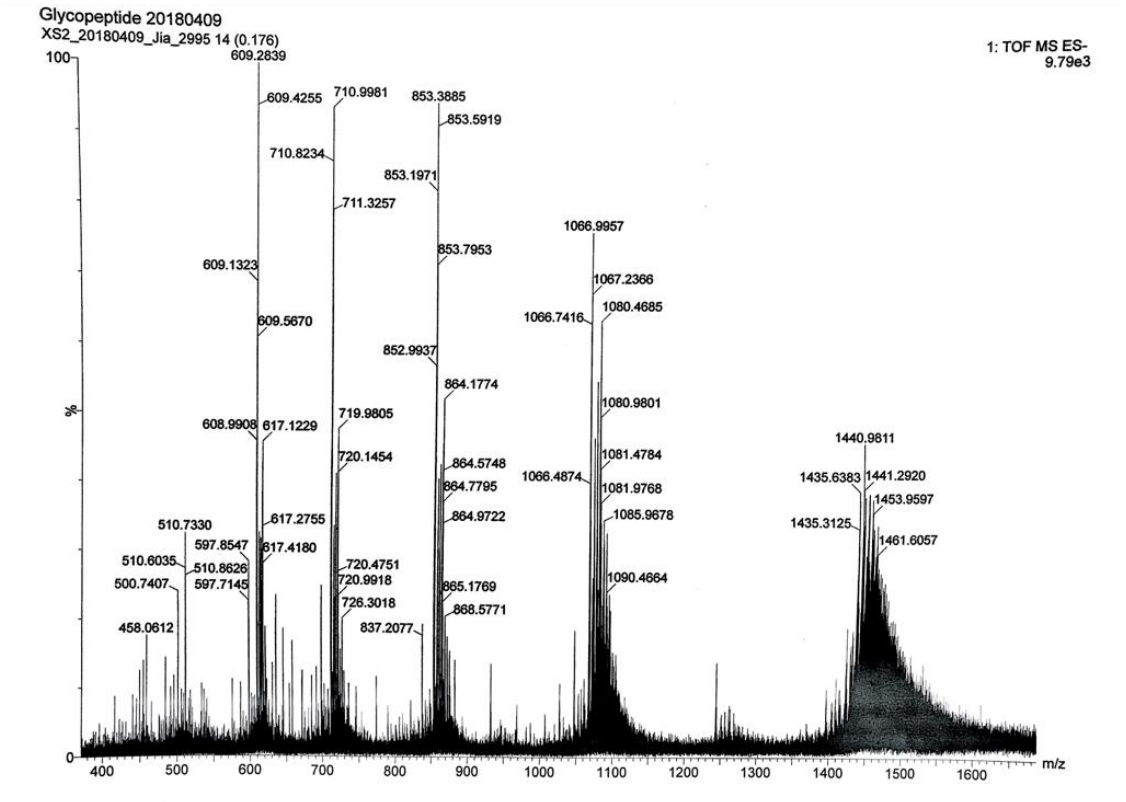
HPLC Chromatogram of **2**

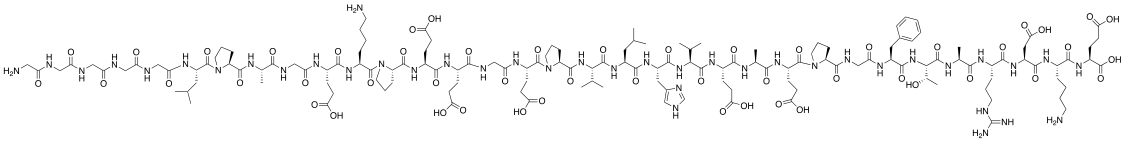




2

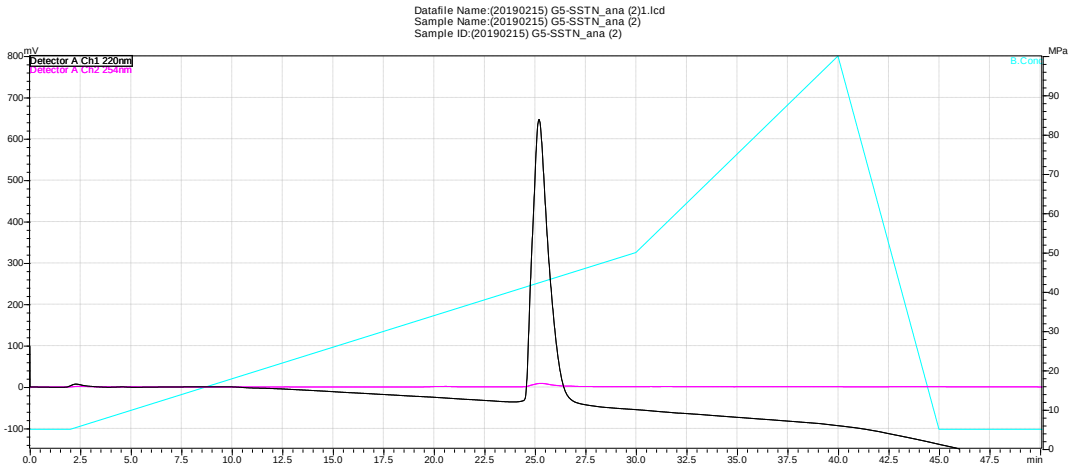
ESI-MS of **2**


Glycopeptide 20180409
XS2_20180409_Jia_2995 14 (0.176)
1: TOF MS ES-
9.79e3
100
%
0
609.2839
609.4255
710.9981
853.3885
853.5919
710.8234
853.1971
711.3257
1066.9957
853.7953
609.1323
1067.2366
1080.4685
1066.7416
609.5670
852.9937
864.1774
1080.9801
719.9805
608.9908
617.1229
1440.9811
720.1454
1081.4784
1066.4874
1435.6383
1441.2920
864.5748
864.7795
1081.9768
1453.9597
510.7330
617.2755
1085.9678
1435.3125
597.8547
617.4180
1461.6057
510.6035
510.8626
720.4751
720.9918
1090.4664
500.7407
597.7145
865.1769
726.3018
868.5771
458.0612
837.2077
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
m/z

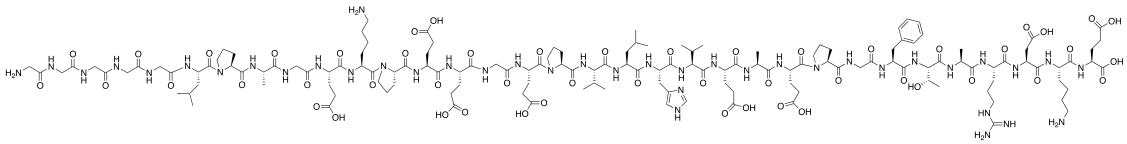


3

HPLC Chromatogram of **3**

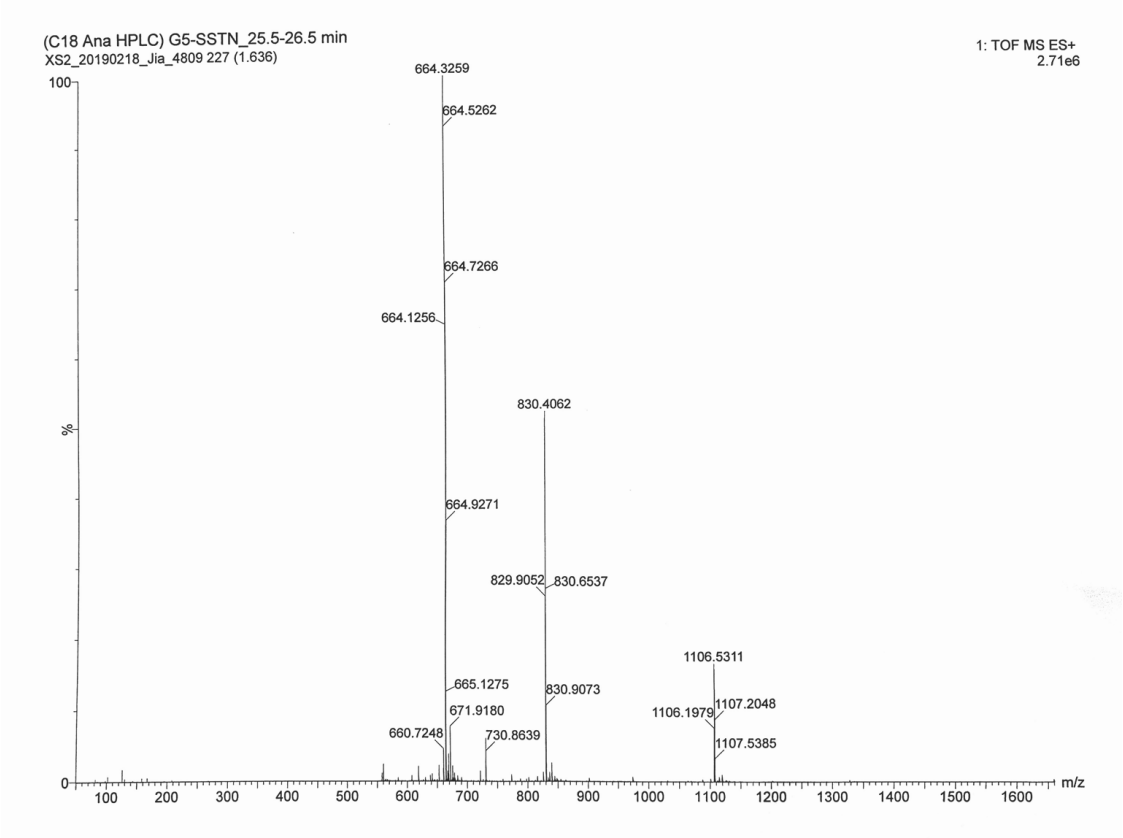


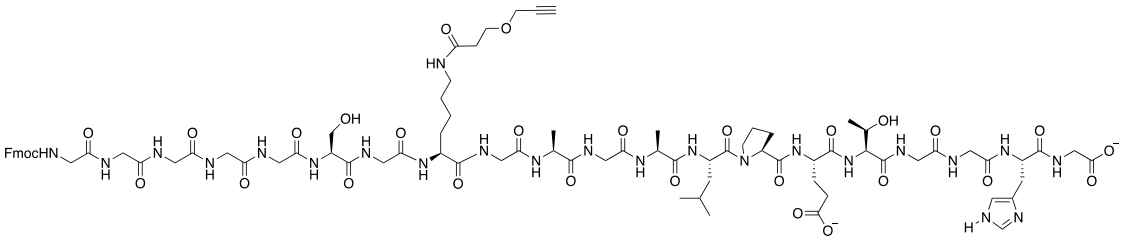
The purity of peptide **3** was verified with analytical C-18 HPLC (5-100% acetonitrile/water; 0.1% trifluoroacetic acid). ESI-MS: $C_{143}H_{226}N_{40}O_{51}$ $[M+4H]^{4+}$ calcd: 829.9075, obsd: 829.9052 (2.77 ppm).



3

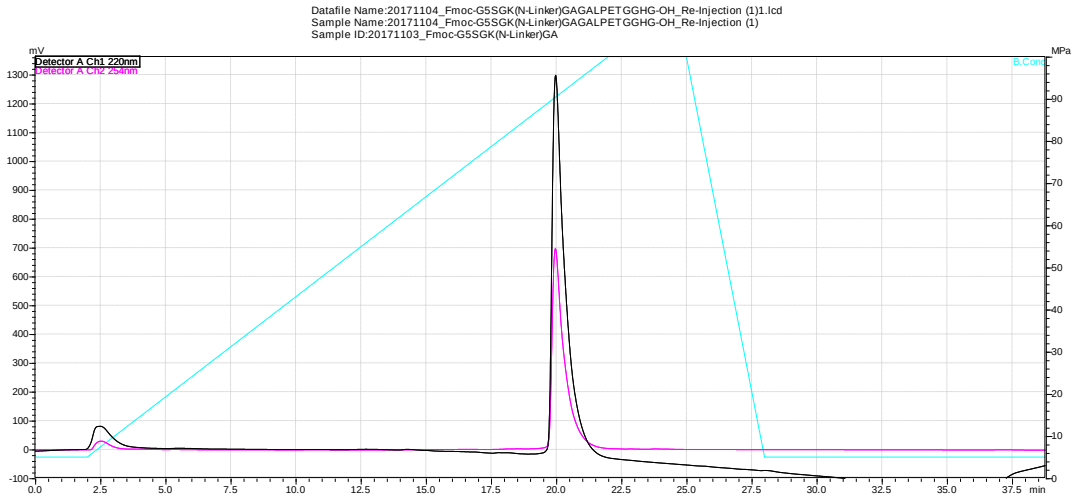
ESI-MS of **3**


(C18 Ana HPLC) G5-SSTN_25.5-26.5 min
XS2_20190218_Jia_4809 227 (1.636)
1: TOF MS ES+
2.71e6
100
%
0
664.3259
664.5262
664.7266
664.1256
830.4062
664.9271
829.9052
830.6537
1106.5311
665.1275
830.9073
671.9180
1106.1979
1107.2048
660.7248
730.8639
1107.5385
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
m/z

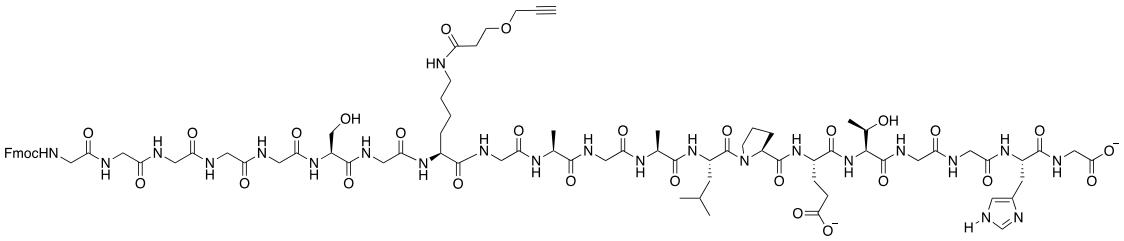


5

HPLC Chromatogram of **5**

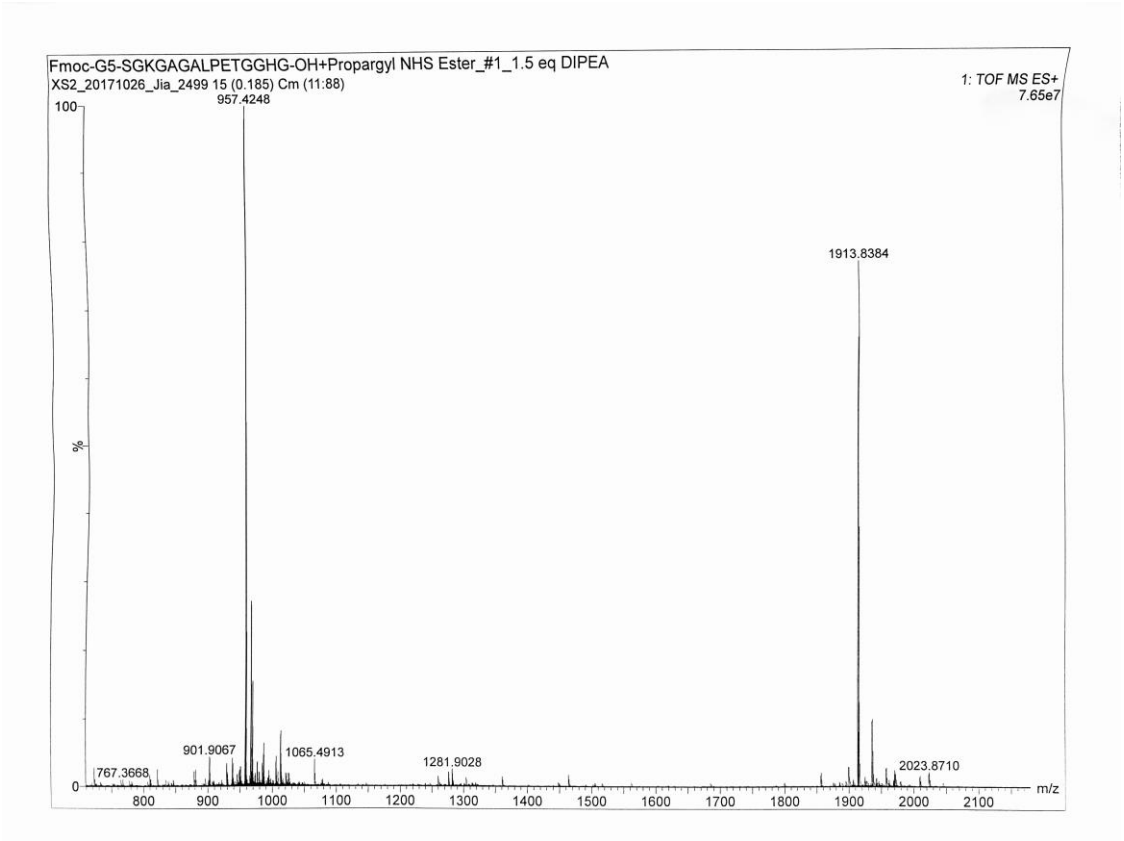


The purity of peptide **5** was verified with analytical C-18 HPLC (5-100% acetonitrile/water; 0.1% trifluoroacetic acid). ESI-MS: $C_{84}H_{118}N_{23}O_{29}$ $[M+4H]^{4+}$ calcd: 1912.8461, obsd: 1912.8388 (3.82ppm).



5

ESI-MS of **5**


Fmoc-G5-SGKGAGALPETGGHG-OH+Propargyl NHS Ester_#1_1.5 eq DIPEA
XS2_20171026_Jia_2499 15 (0.185) Cm (11:88)
1: TOF MS ES+
7.65e7
100
%
0
957.4248
1913.8384
901.9067
1065.4913
1281.9028
767.3668
2023.8710
m/z
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100

References

1 A. Saladin, J. Rey, P. Thevenet, M. Zacharias, G. Moroy and P. Tuffery, *Nucleic Acids Res.*, 2014, **42**, W221.

2 D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov and S. Vajda, *Nat. Protoc.*, 2017, **12**, 255.

3 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605.

4 Woods Group. (2005-2021) GLYCAM Web. Complex Carbohydrate Research Center, University of Georgia, Athens, GA. (http://glycam.org)

5 D. Xu and Y. Zhang, *Proteins*, 2012, **80**, 1715.

6 Schrodinger Release 2020-4: Maestro, Schrodinger, LLC, New York, NY, 2020.