Convergent Synthesis of MUC1 Glycopeptides via Serine Ligation

Ci Xu, Hiu Yung Lam, Yinfeng Zhang and Xuechen Li*

Department of Chemistry, The University of Hong Kong, Hong Kong, P. R. China

Materials and Methods

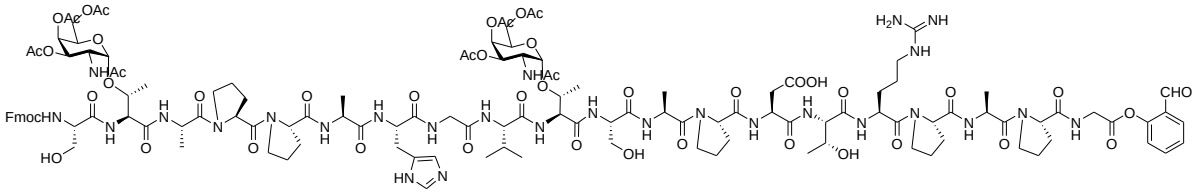
All commercial materials (Aldrich, Fluka and GL Biochem) were used without further purification. All solvents were reagent grade or HPLC grade (DUKSAN). Dry dichloromethane (DCM) was distilled from calcium hydride (CaH_2). All separations involved a mobile phase of 0.05% TFA (*v/v*) in acetonitrile (solvent A)/0.05% TFA (*v/v*) in water (Solvent B). HPLC separations were performed with a Waters HPLC system equipped with a photodiode array detector (Waters 2996) using a Vydac C18 column (5 µm, 300 Å, 4.6 x 150 mm) at a flow rate of 0.6 mL/min for analytical HPLC and Vydac Prep C18 column (10 µm, 300 Å, 22 x 250 mm) at a flow rate of 10 mL/min for preparative HPLC. Low-resolution mass spectral analyses were performed with a Waters 3100 mass spectrometer.

Solid-phase Peptide Synthesis According to Fmoc-strategy

The solid phase peptide synthesis of glycopeptides was carried out manually using H-Gly-2-Cl-Trityl resin (loading 0.5 mmol/g). The following Fmoc amino acids from GL Biochem were employed: Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asp(O*t*Bu)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Pro-OH, Fmoc-Ser(*t*Bu)-OH, Fmoc-Thr(*t*Bu)-OH, Fmoc-Val-OH. Fmoc removal was executed using a solution of 20% piperidine in dimethylformamide (DMF) at room temperature for 30 min. Coupling of Fmoc protected

amino acid units was carried out by activation with (*O*-(7-azabenzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate (HATU) using N,N-diisopropylethylamine (DIPEA) in DMF at room temperature for 40 min. The Fmoc amino acids (2.0 equiv), HATU (2.0 equiv) and DIPEA (5.0 equiv) were dissolved in DMF and subsequently mixed with the resin manually. This procedure was repeated twice for each coupling. Upon completion of synthesis, the peptide resin was subjected to a cleavage cocktail (TFA/*i*Pr_3SiH/H_2O, 15/0.9/0.9, *v/v/v*) for 2 h. The resin was filtered and the combined filtrates were blown off under a stream of condensed air. The crude product was triturated with cold diethyl ether to give a white suspension, which was centrifuged and the ether subsequently decanted. The remaining solid was ready for HPLC purification.

Synthesis of MUC1 Glycopeptide SAL-ester 5



5

Crude protected peptide sequence Fmoc-ST(GalNAc(Ac)$_3$-α-D)APPAHGVT(GalNAc(Ac)$_3$-α-D)SAPDTRPAPG-OH was firstly synthesized according to the general SPPS procedure, coupled with salicylaldehyde dimethyl acetal, benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) and DIPEA in anhydrous DCM overnight, then followed by treatment with TFA/Phenol/H_2O (95/2.5/2.5, *v/v/v*) for 2 h. The crude peptide salicylaldehyde ester was precipitated out by diethyl ether, followed by RP-HPLC purification (20-60% CH_3CN/H_2O containing 0.05% TFA over 30 min, C18 column, 190-400 nm) and lyophilization to give compound **5** as a white powder.

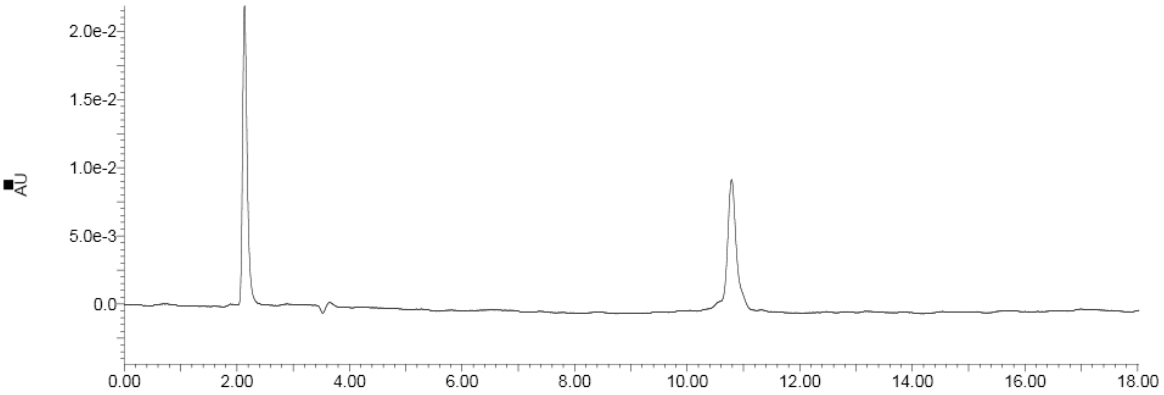


Figure S1: UV trace (190-400 nm) from LC-MS analysis of compound **5:** gradient 20-60% CH_3CN/H_2O containing 0.05% TFA over 15 min at a flow rate of 0.6 mL/min

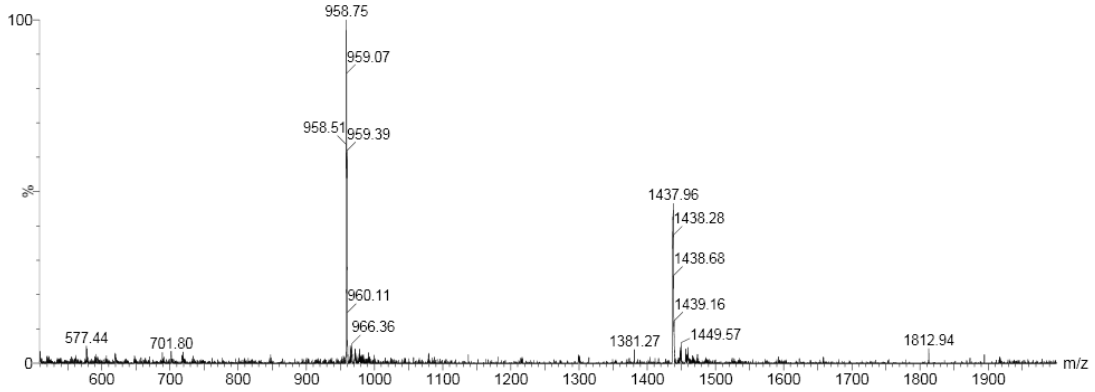
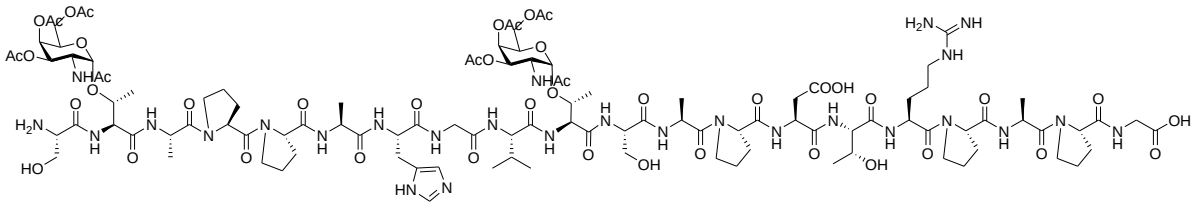


Figure S2: ESI-MS of compound **5**

ESI calcd for $C_{130}H_{179}N_{27}O_{47}$ $[M+2H]^{2+}$ *m/z* = 1436.98, $[M+3H]^{3+}$ *m/z* = 958.32, found: 1437.96, 958.75

Synthesis of MUC1 Glycopeptide 6



6

Compound **6** was synthesized according to the general SPPS procedure. Preparative HPLC

purification (10-30% CH_3CN/H_2O containing 0.05% TFA over 30 min, C18 column, 190-400 nm) followed by concentration at reduced pressure and lyophilization afforded compound **6** as a white powder.

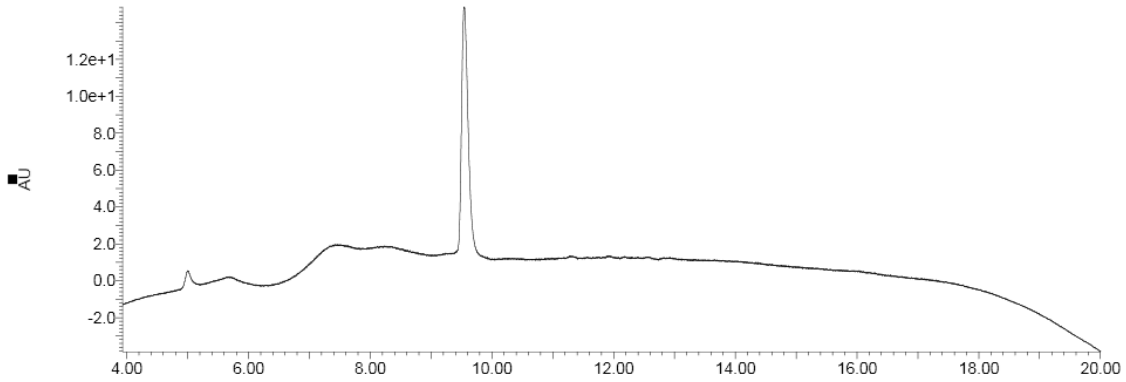


Figure S3: UV trace (190-400 nm) from LC-MS analysis of compound **6:** gradient 5-95% CH_3CN/H_2O containing 0.05% TFA over 15 min at a flow rate of 0.6 mL/min

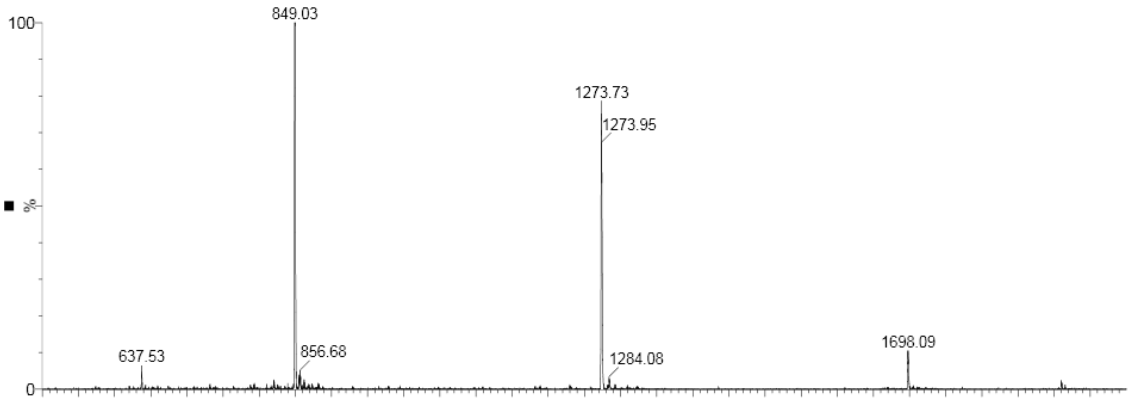
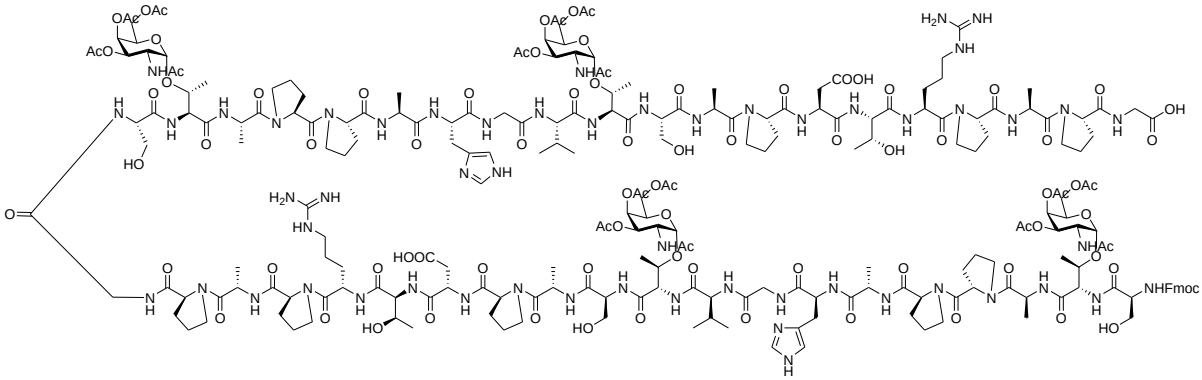


Figure S4: ESI-MS of compound **6**

ESI calcd for $C_{108}H_{165}N_{27}O_{44}$ $[M+2H]^{2+}$ *m/z* = 1273.81, $[M+3H]^{3+}$ *m/z* = 849.54, found: 1273.73, 849.03

Serine Ligation of Glycopeptide SAL-ester 5 and Glycopeptide 6



7

Compound **5** (16.6 mg, 1.0 equiv.) and compound **6** (17.7 mg, 1.2 equiv.) were dissolved in pyridine/acetic acid (1:1 mole/mole) at a concentration of 25 mM at room temperature. The reaction mixture was stirred at room temperature for 4 h. After completion of the reaction (5% of hydrolysis of compound **5** was detected), the solvent was blown off under a stream of condensed air. Then, the residue was treated with TFA/H_2O (95/5, *v/v*) for 10 min. The solvent was blown off under a stream of condensed air. Preparative HPLC purification (10-50% CH_3CN/H_2O containing 0.05% TFA over 30 min, C18 column, 190-400 nm) followed by concentration at reduced pressure and lyophilization afforded compound **7** (15.1 mg, 49% isolation yield) as a white powder.

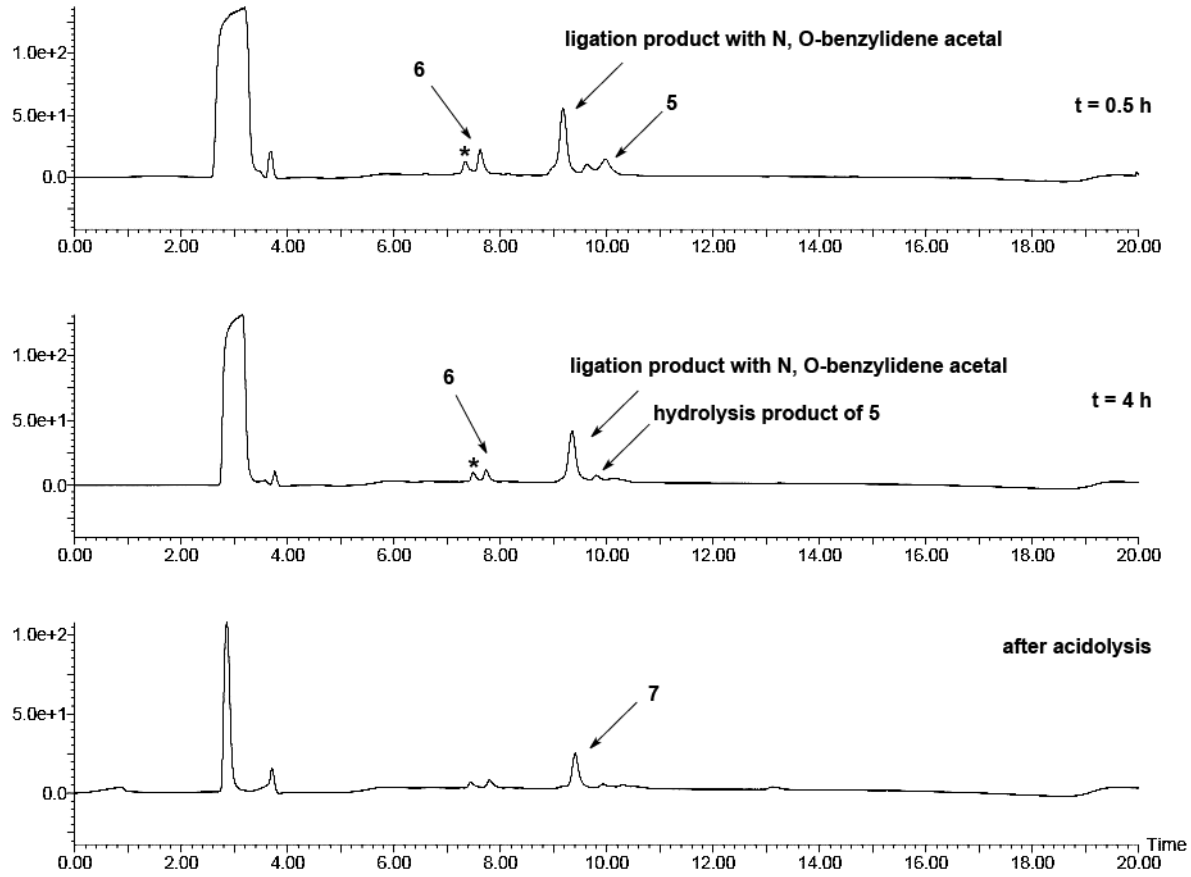


Figure S5: UV trace (190-400 nm) from LC-MS analysis of serine ligation at 0.5 h, 4 h and acidolysis: gradient 5-95% CH_3CN/H_2O containing 0.05% TFA over 15 min at a flow rate of 0.6 mL/min (* impurity)

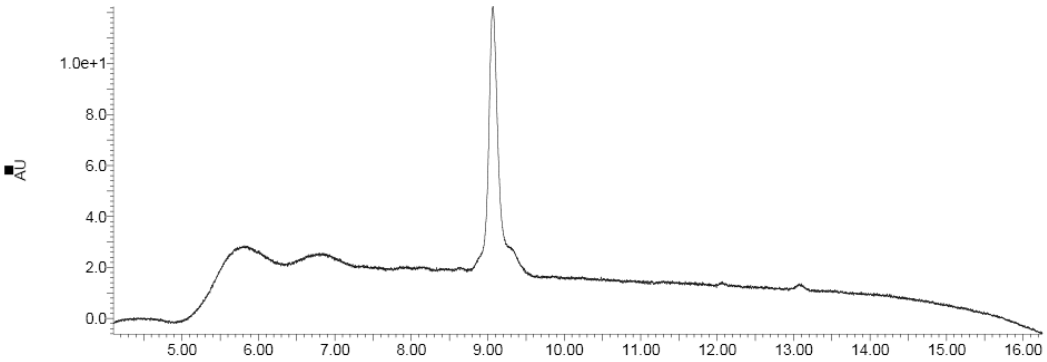


Figure S6: UV trace (190-400 nm) from LC-MS analysis of purified compound **7:** gradient 5-95% CH_3CN/H_2O containing 0.05% TFA over 15 min at a flow rate of 0.6 mL/min

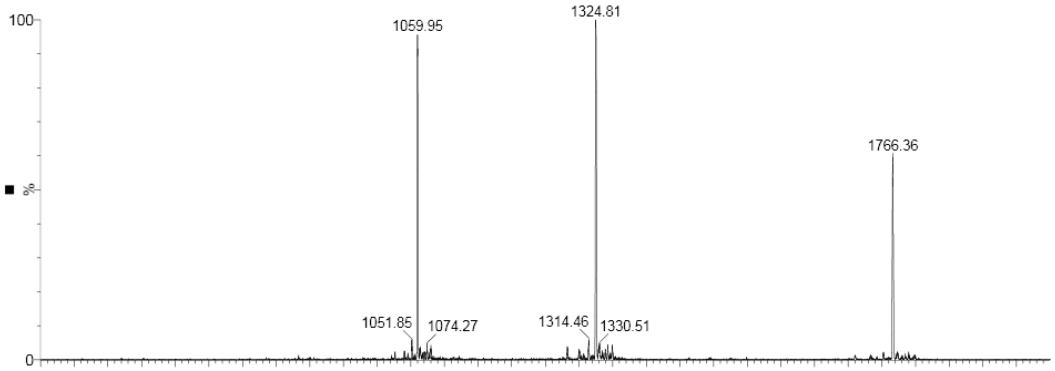
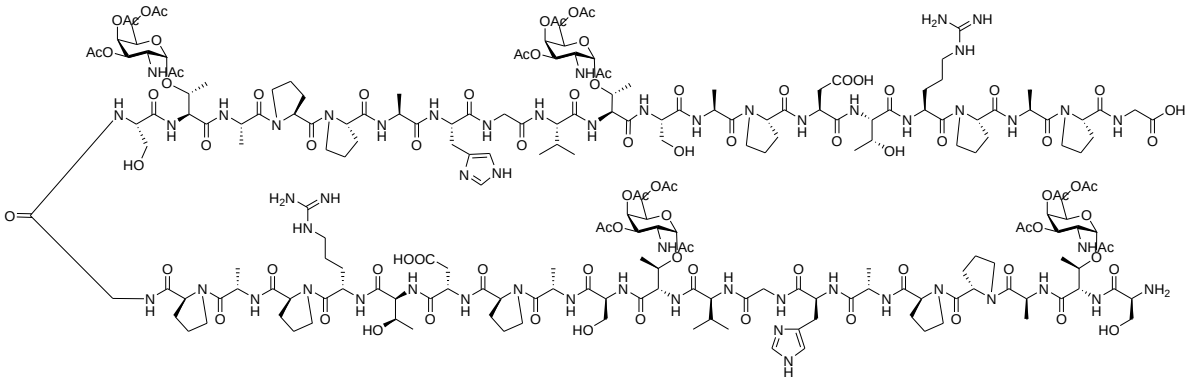


Figure S7: ESI calcd for $C_{231}H_{338}N_{54}O_{89}$ $[M+3H]^{3+}$ *m/z* = 1766.15, $[M+4H]^{4+}$ *m/z* = 1324.86, $[M+5H]^{5+}$ *m/z* = 1060.09, found: 1766.36, 1324.81, 1059.95

Glycopeptide 8



8

Compound **7** (1.7 mg) was treated with 0.3 mL diethylamine/DCM (1/2, *v/v*) for 1.5 h at room temperature to remove terminal Fmoc protecting group. The solvent was blown off under a stream of condensed air. Without further purification, the residue was lyophilized to afford compound **8** as a white powder.

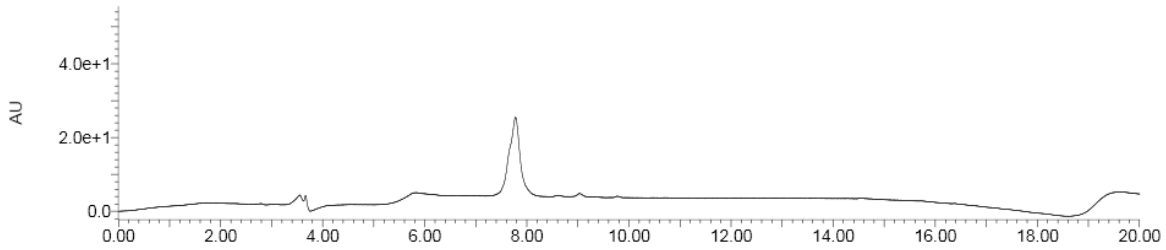


Figure S8: UV trace (190-400 nm) from LC-MS analysis of compound **8:** gradient 5-95% CH_3CN/H_2O containing 0.05% TFA over 15 min at a flow rate of 0.6 mL/min

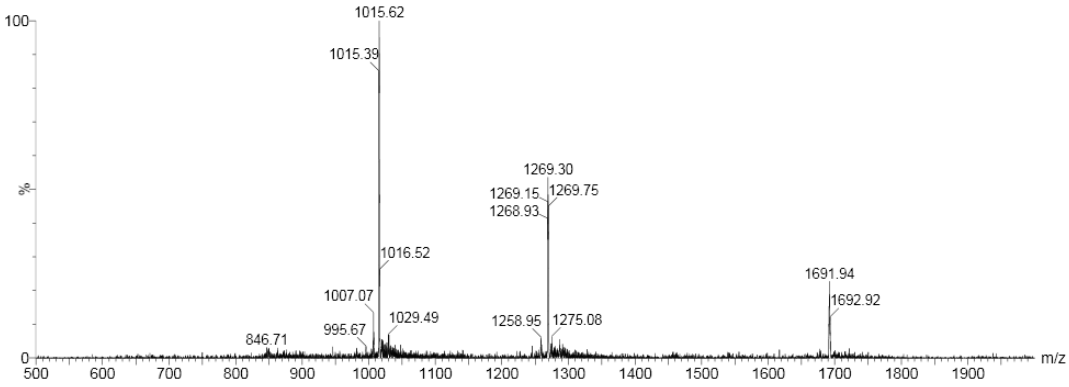
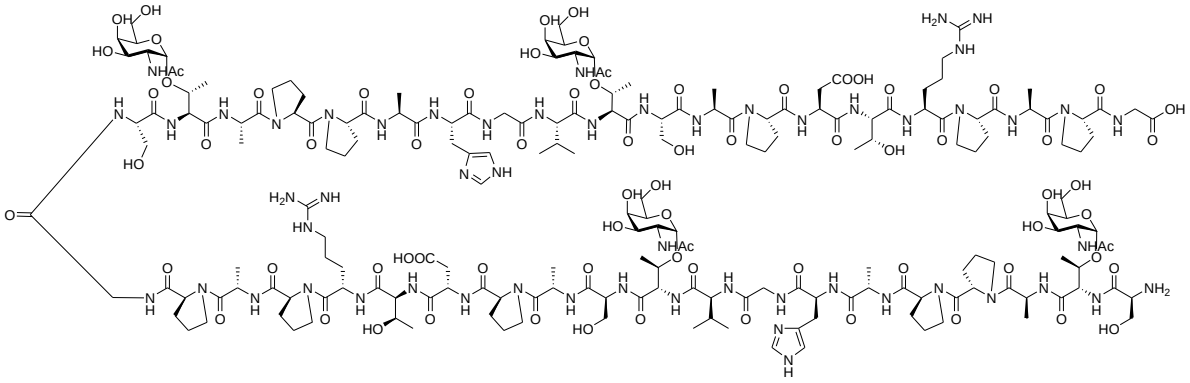


Figure S9: ESI calcd for $C_{216}H_{328}N_{54}O_{87}$ $[M+3H]^{3+}$ *m/z* = 1692.08, $[M+4H]^{4+}$ *m/z* = 1269.31, $[M+5H]^{5+}$ *m/z* = 1015.65, $[M+6H]^{6+}$ *m/z* = 846.54, found: 1691.94, 1269.30, 1015.62, 846.71

Glycopeptide 8'



8'

Compound **8** was incubated with 5% hydrazine in H_2O for 1 h at room temperature before the product was purified by ultrafiltration. The filtration was performed using Millipore Amicon Ultra-4 centrifugal filters (membrane 3 KDa cut-off). The reaction solution was diluted with HPLC grade water to 3.5 mL total volume in a Millipore centrifugal filter tube. The tube was centrifuged at 7,500 rpm (T15A 41/42 rotor, CF16R XII centrifuge, Hitachi) until residual volume was 0.25-0.5 mL. This procedure was repeated four times, followed by lyophilization delivered compound **8'** (1.0 mg, 68% recovery yield from compound **7**).

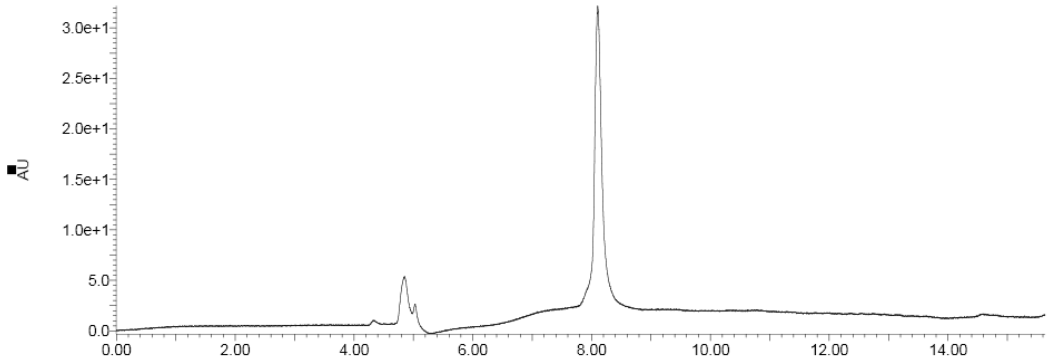


Figure S10: UV trace (190-400 nm) from LC-MS analysis of compound **8':** gradient 5-95% CH_3CN/H_2O containing 0.05% TFA over 15 min at a flow rate of 0.6 mL/min

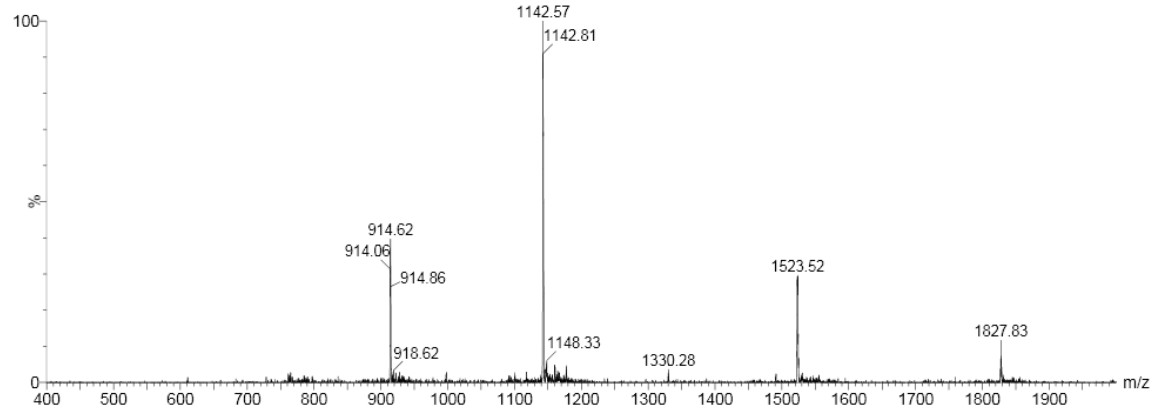
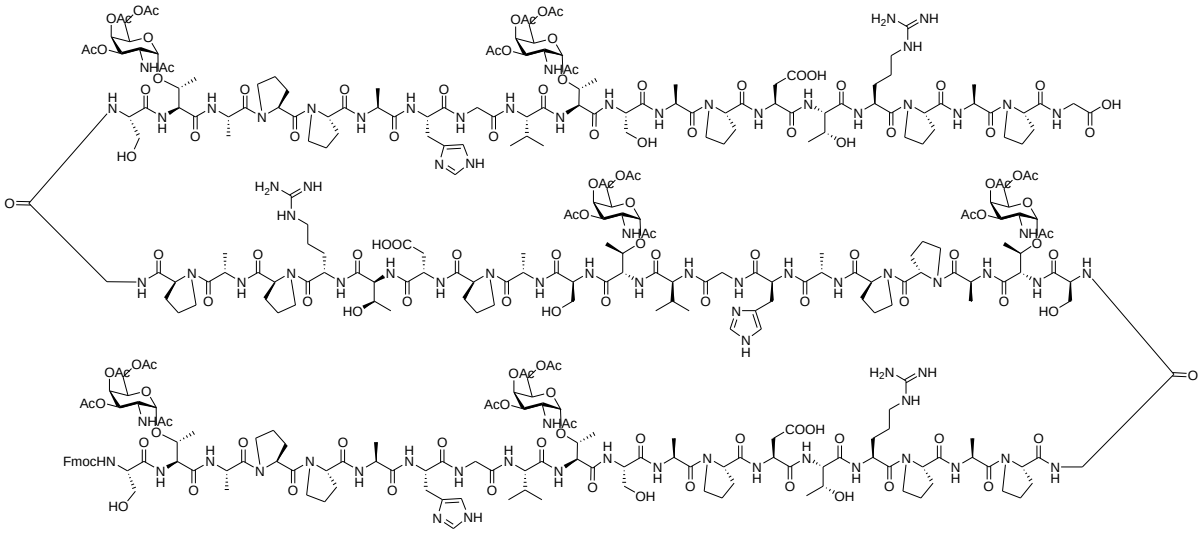


Figure S11: ESI calcd for $C_{192}H_{304}N_{54}O_{75}$ $[M+3H]^{3+}$ m/z = 1523.93, $[M+4H]^{4+}$ *m/z* = 1143.20, $[M+5H]^{5+}$ *m/z* = 914.76, found: 1523.52, 1142.57, 914.62

Serine Ligation of Glycopeptide SAL-ester 5 and Glycopeptide 8



9

Compound **5** (6.0 mg, 1.0 equiv.) and compound **8** (12.7 mg, 1.2 equiv.) were dissolved in pyridine/ acetic acid (1:1, mole/mole) at a concentration of 25 mM. The reaction was stirred at room temperature for 4 h. After completion of the reaction (10% of hydrolysis of compound **5** was detected), the solvent was blown off under a stream of condensed air. The residue was then treated with TFA/H_2O (95/5, *v/v*) for 10 min. The solvent was blown off by a stream of air before the product was purified by ultrafiltration. The filtration was performed using Millipore Amicon Ultra-4 centrifugal filters (membrane 3 KDa cut-off). The reaction solution was diluted with HPLC grade water to 3.5 mL total volume in a Millipore centrifugal filter tube. The tube was centrifuged at 7,500 rpm (T15A 41/42 rotor, CF16R XII centrifuge, Hitachi) until residual volume was 0.25-0.5 mL. This procedure was repeated four times, followed by lyophilization delivered compound **9** (14.2 mg, 87% recovery yield from compound **5**) as a white powder.

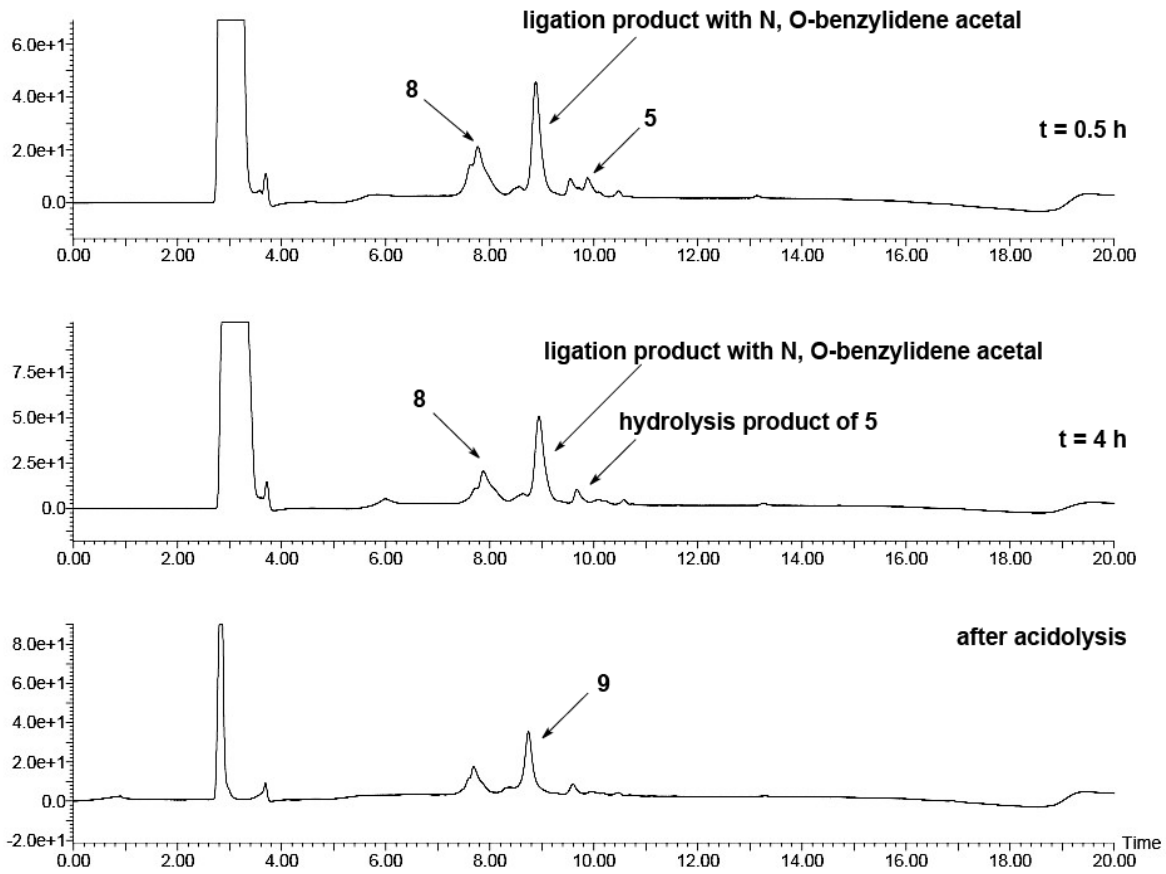


Figure S12: UV trace (190-400 nm) from LC-MS analysis of serine ligation at 0.5 h, 4 h and acidolysis: gradient 5-95% CH_3CN/H_2O containing 0.05% TFA over 15 min at a flow rate of 0.6 mL/min

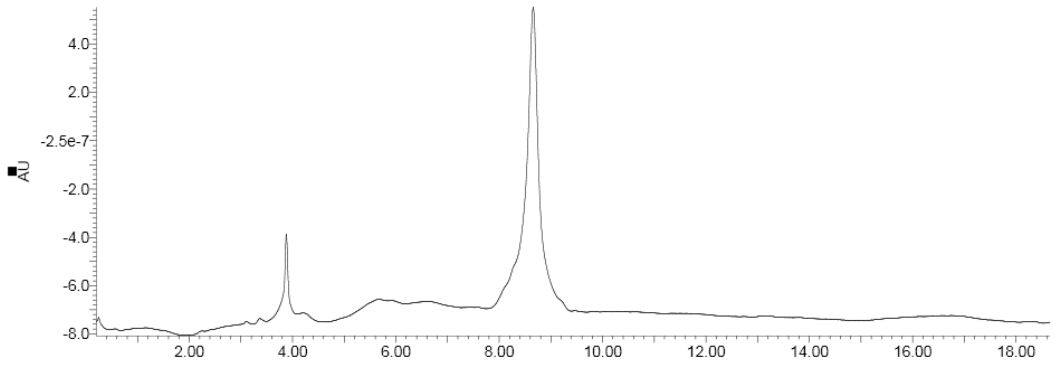


Figure S13: UV trace (190-400 nm) from LC-MS analysis of compound **9:** gradient 5-95% CH_3CN/H_2O containing 0.05% TFA over 15 min at a flow rate of 0.6 mL/min

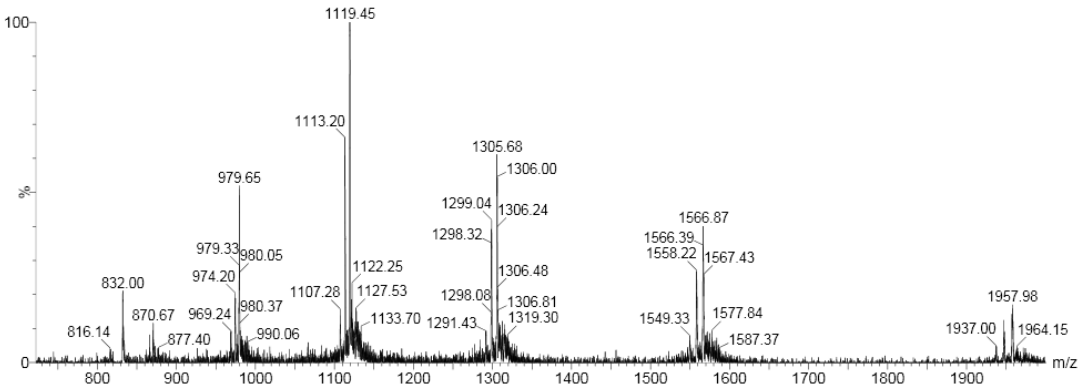
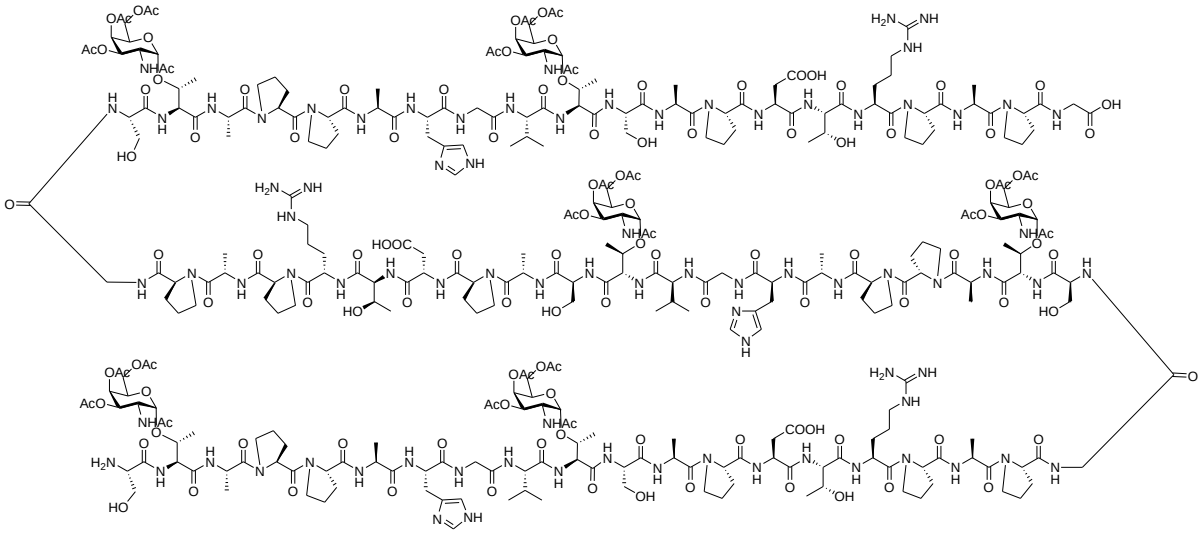


Figure S14: ESI calcd for $C_{339}H_{501}N_{81}O_{132}$ $[M+4H]^{4+}$ *m/z* = 1956.77, $[M+5H]^{5+}$ *m/z* = 1565.14, $[M+6H]^{6+}$ *m/z* = 1304.84, $[M+7H]^{7+}$ *m/z* = 1118.58, $[M+8H]^{8+}$ *m/z* = 978.88, $[M+9H]^{9+}$ *m/z* = 870.23, found: 1957.98, 1566.87, 1305.68, 1119.45, 979.65, 870.67; found: 1558.22, 1299.04, 1113.20, 974.20 (The mass observed corresponded to [M-Ac]); found: 1937.00, 1549.33, 1291.43, 1107.28, 969.24 (The mass observed corresponded to [M-2Ac])

Glycopeptide 10



10

Compound **9** was treated with 1 mL diethylamine/DCM (1/2, *v/v*) for 1.5 h at room temperature to remove terminal Fmoc protecting group. The solvent was blown off under a

stream of condensed air. Without further purification, the residue was lyophilized to afford compound **10** as a white powder.

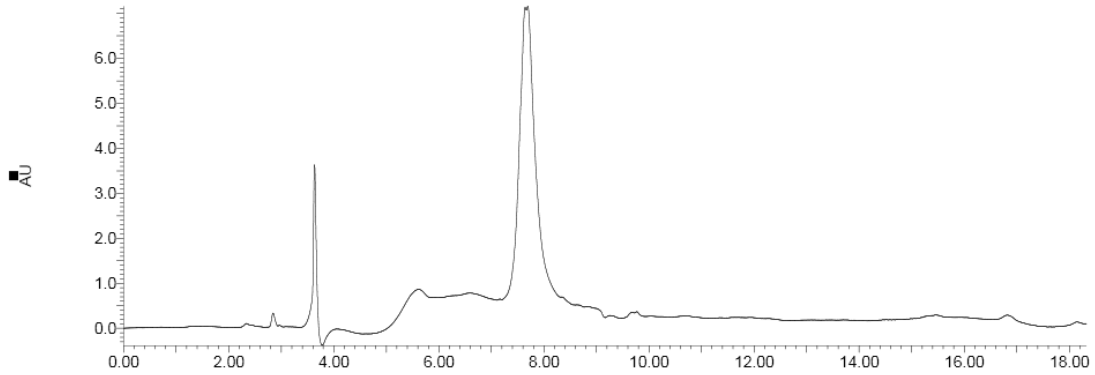


Figure S15: UV trace (190-400 nm) from LC-MS analysis of compound **10:** gradient 5-95% CH_3CN/H_2O containing 0.05% TFA over 15 min at a flow rate of 0.6 mL/min

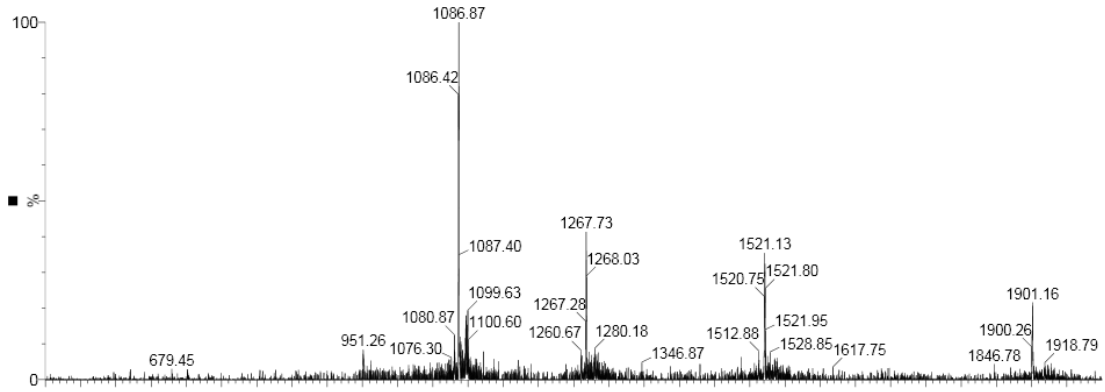
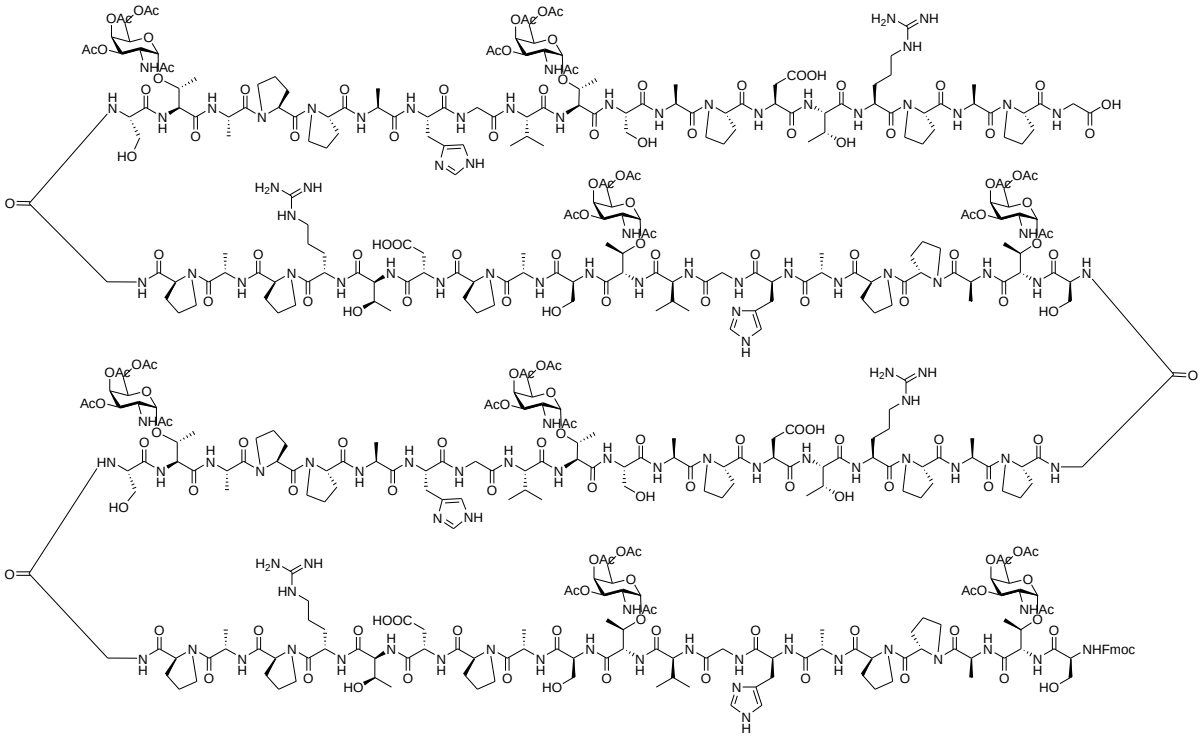


Figure S16: ESI calcd for $C_{324}H_{491}N_{81}O_{130}$ $[M+4H]^{4+}$ *m/z* = 1901.21, $[M+5H]^{5+}$ *m/z* = 1521.17, $[M+6H]^{6+}$ *m/z* = 1267.80, $[M+7H]^{7+}$ *m/z* = 1086.83, $[M+8H]^{8+}$ *m/z* = 951.10, found: 1901.16, 1521.13, 1267.73, 1086.87, 951.26

Serine Ligation of Glycopeptide SAL-ester 5 and Glycopeptide 10



11

Compound **5** (2.4 mg, 1.0 equiv.) and compound **10** (6.3 mg, 1.0 equiv.) were dissolved in pyridine/acetic acid (1:1, mole/mole) at a concentration of 25 mM. The reaction was stirred at room temperature for 4 h. After completion of the reaction（18% of hydrolysis of compound **5** was detected, the solvent was blown off under a stream of condensed air. The residue was then treated with TFA/H_2O (95/5, *v/v*) for 10 min. The solvent was blown off by a stream of air. Preparative HPLC purification (25-50% CH_3CN/H_2O containing 0.05% TFA over 30 min, C18 column, 190-400 nm) followed by concentration at reduced pressure and lyophilization afforded compound **11** (3.3 mg, 38% isolation yield) as a white powder.

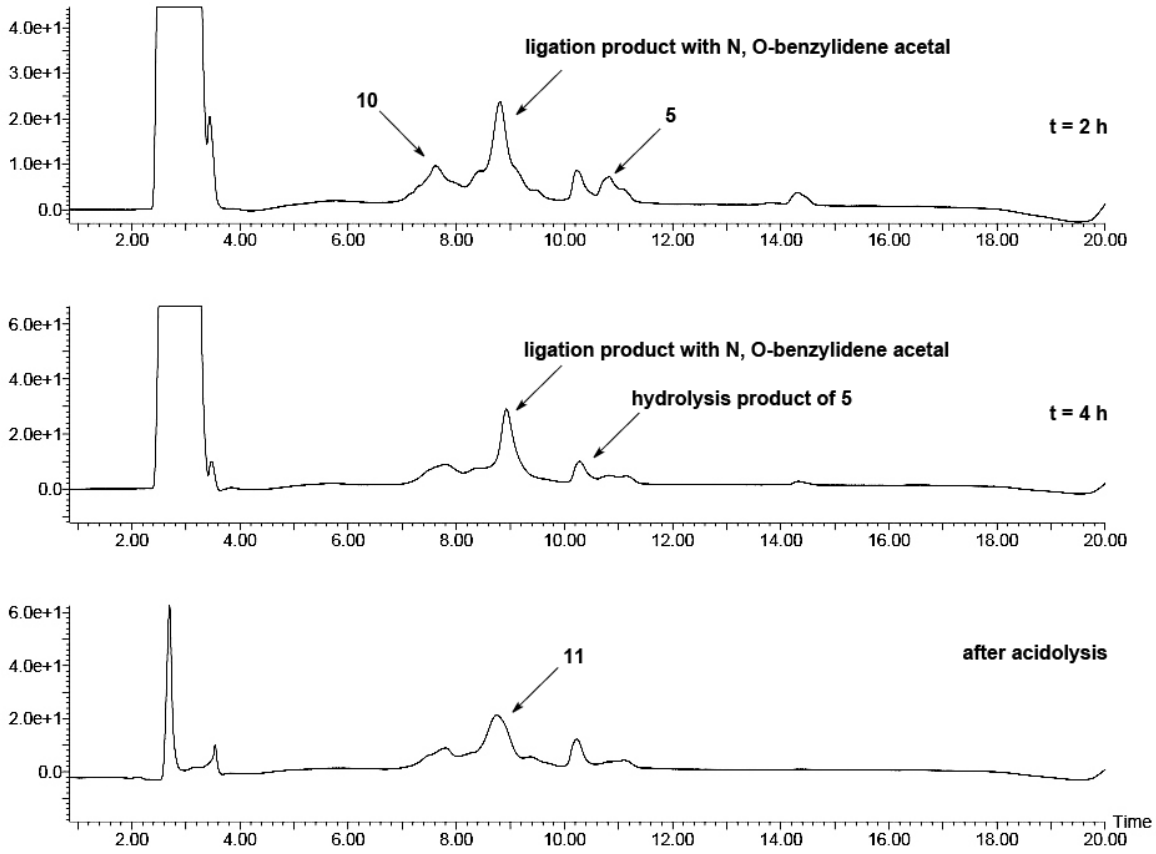


Figure S17: UV trace (190-400 nm) from LC-MS analysis of serine ligation at 2 h, 4 h and acidolysis: gradient 10-75% CH_3CN/H_2O containing 0.05% TFA over 15 min at a flow rate of 0.6 mL/min

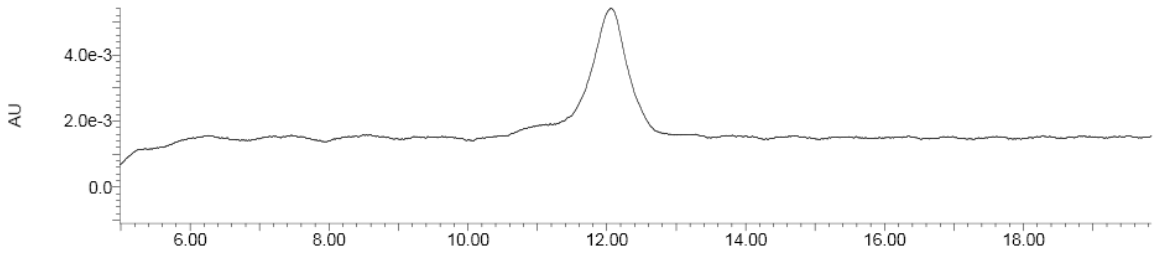


Figure S18: UV trace (190-400 nm) from LC-MS analysis of compound **11:** gradient 25-50% CH_3CN/H_2O containing 0.05% TFA over 15 min at a flow rate of 0.6 mL/min

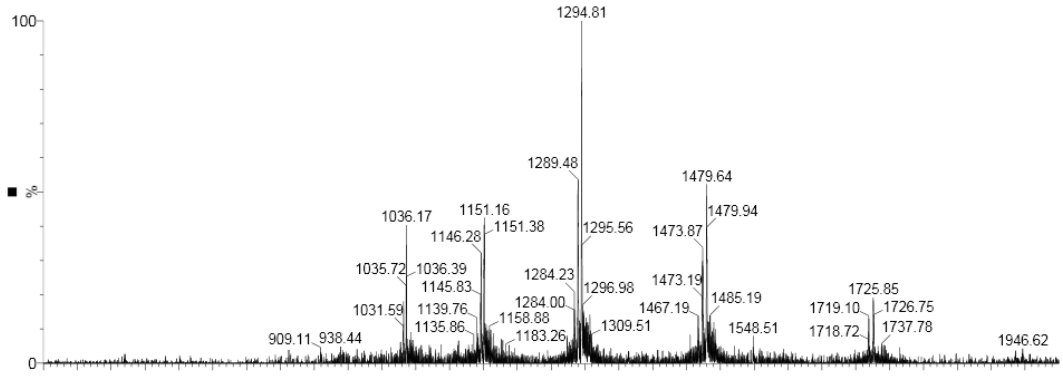
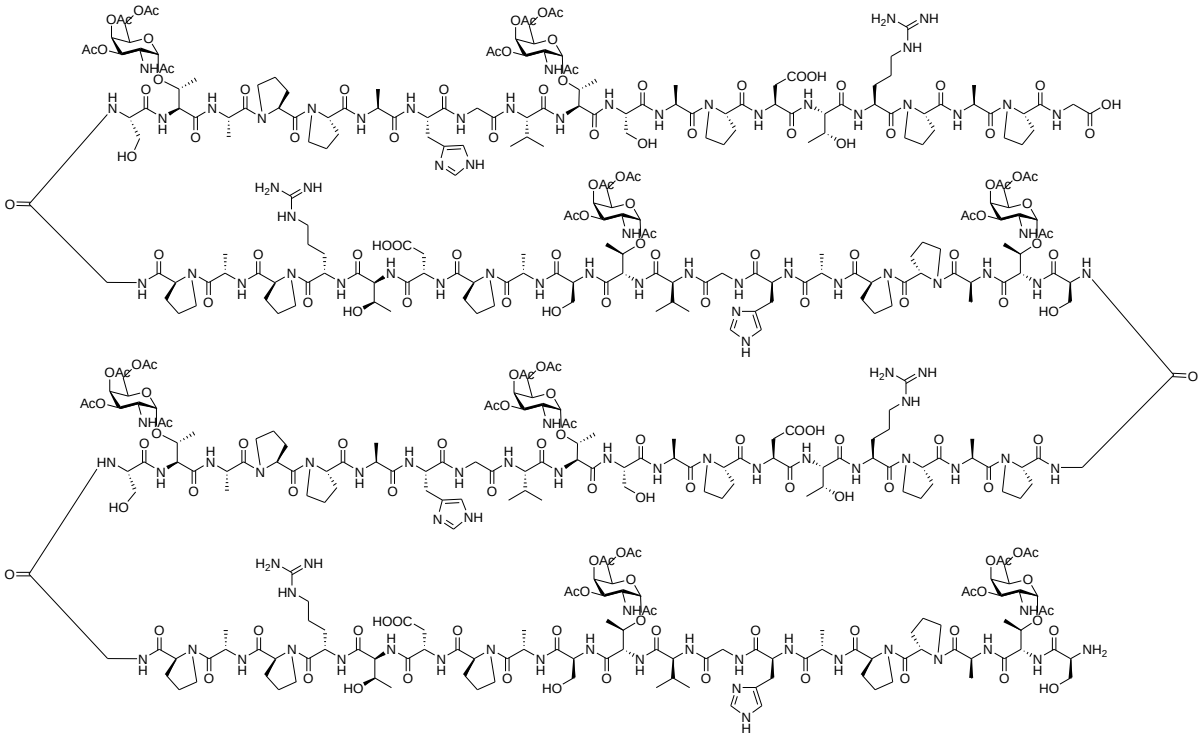


Figure S19: ESI calcd for $C_{447}H_{664}N_{108}O_{175}$ $[M+6H]^{6+}$ *m/z* = 1726.11, $[M+7H]^{7+}$ *m/z* = 1479.67, $[M+8H]^{8+}$ *m/z* = 1294.83, $[M+9H]^{9+}$ *m/z* = 1151.07, $[M+10H]^{10+}$ *m/z* = 1036.07, found: 1725.85, 1479.64, 1294.81, 1151.16, 1036.17; found: 1719.10, 1473.87, 1289.48, 1146.28, 1031.59 (The mass observed corresponded to [M-Ac]); found: 1467.19, 1284.00 (The mass observed corresponded to [M-2Ac])

Glycopeptide 12



12

Compound **11** (3.3 mg) was treated with 0.5 mL diethylamine/DCM (1/2, *v/v*) for 1.5 h at

room temperature to remove terminal Fmoc protecting group. The solvent was blown off under a stream of condensed air. Without further purification, the residue was lyophilized to afford compound **12** as a white powder.

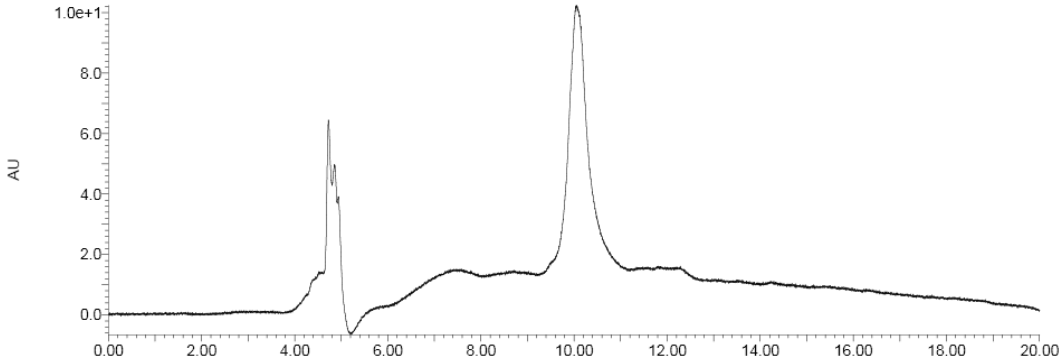


Figure S20: UV trace (190-400 nm) from LC-MS analysis of compound **12:** gradient 10-75% CH_3CN/H_2O containing 0.05% TFA over 15 min at a flow rate of 0.6 mL/min

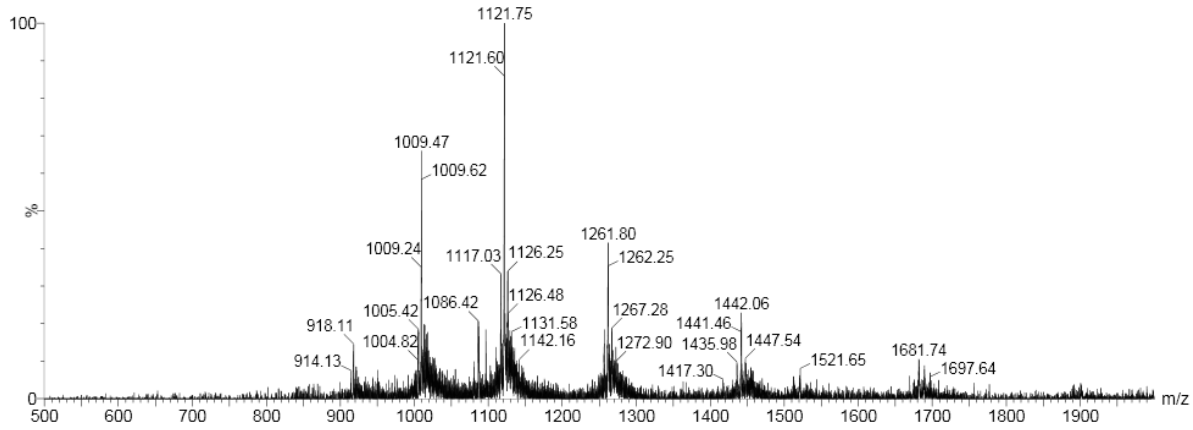
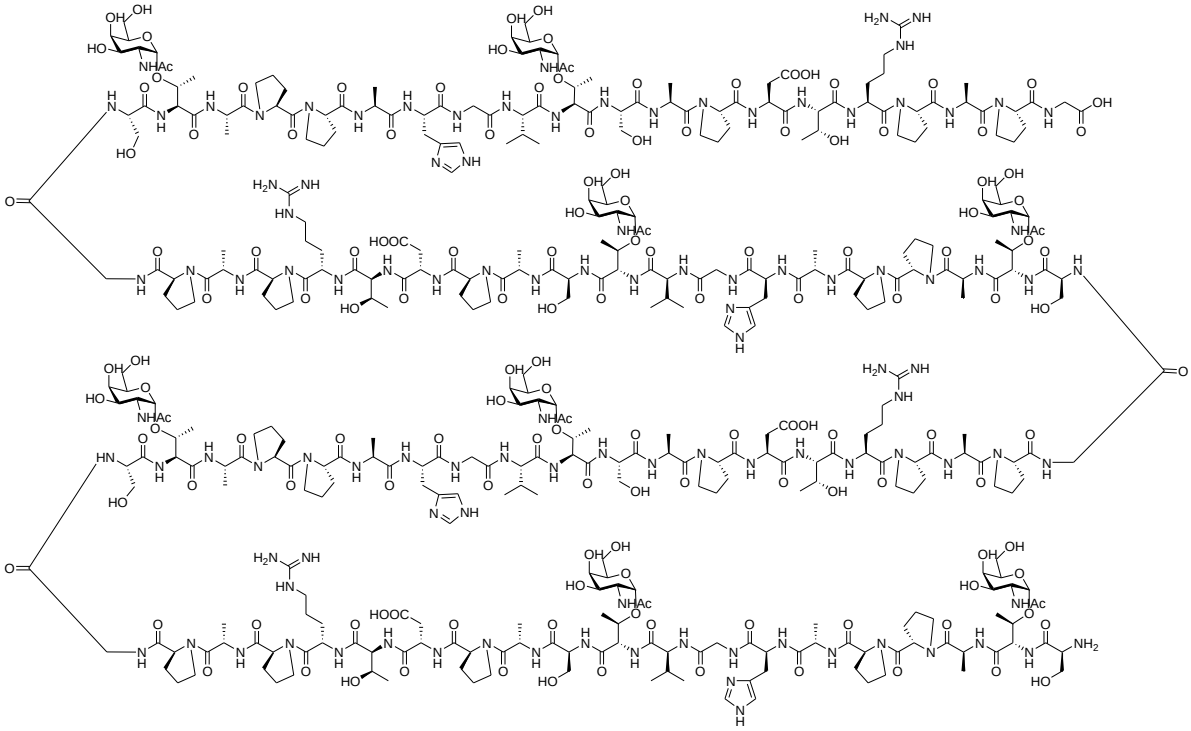


Figure S21: ESI calcd for $C_{432}H_{654}N_{108}O_{173}$ $[M+6H]^{6+}$ *m/z* = 1689.07, $[M+7H]^{7+}$ *m/z* = 1447.92, $[M+8H]^{8+}$ *m/z* = 1267.05, $[M+9H]^{9+}$ *m/z* = 1126.38, $[M+10H]^{10+}$ *m/z* = 1013.84, $[M+11H]^{11+}$ *m/z* = 921.77, found: 1447.54, 1267.28, 1126.48; found: 1681.74, 1442.06, 1261.80, 1121.75, 1009.47, 918.11 (The mass observed corresponded to [M-Ac]); found: 1435.98, 1117.03, 1005.42, 914.13 (The mass observed corresponded to [M-2Ac])

Glycopeptide 13



13

Compound **12** was incubated with 5% hydrazine in H_2O for 1 h at room temperature before the product was purified by ultrafiltration. The filtration was performed using Millipore Amicon Ultra-4 centrifugal filters (membrane 3 KDa cut-off). The reaction solution was diluted with HPLC grade water to 3.5 mL total volume in a Millipore centrifugal filter tube. The tube was centrifuged at 7,500 rpm (T15A 41/42 rotor, CF16R XII centrifuge, Hitachi) until residual volume was 0.25-0.5 mL. This procedure was repeated four times, followed by lyophilization delivered compound **13** (1.9 mg, 65% recovery yield from compound **11**) as a white powder.

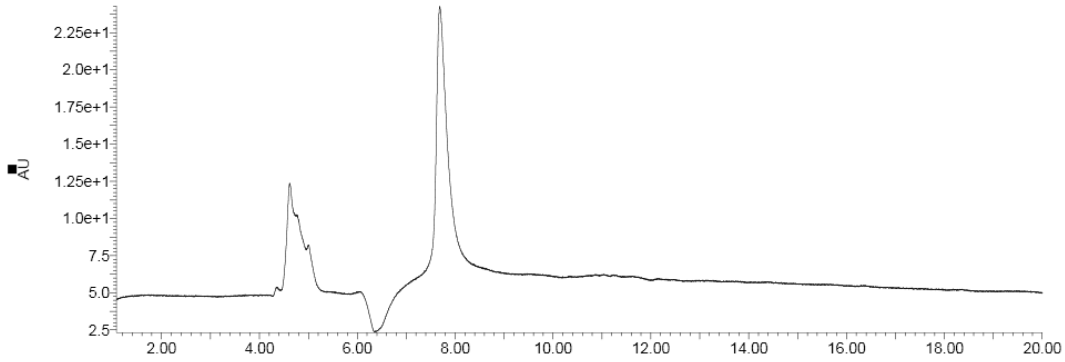


Figure S22: UV trace (190-400 nm) from LC-MS analysis of compound **13:** gradient 10-75% CH_3CN/H_2O containing 0.05% TFA over 15 min at a flow rate of 0.6 mL/min

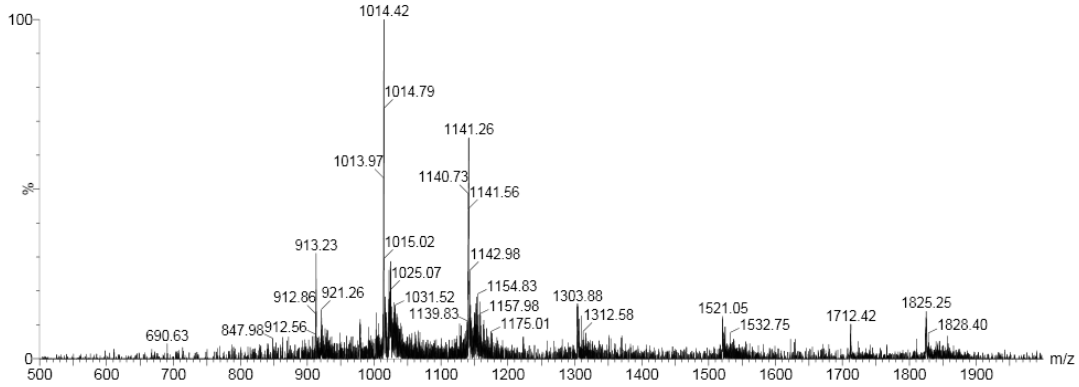


Figure S23: ESI calcd for $C_{384}H_{606}N_{108}O_{149}$ $[M+5H]^{5+}$ m/z = 1824.91, $[M+6H]^{6+}$ *m/z* = 1520.92, $[M+7H]^{7+}$ *m/z* =1303.79, $[M+8H]^{8+}$ *m/z* =1140.94, $[M+9H]^{9+}$ *m/z* = 1014.28, $[M+10H]^{10+}$ m/z = 912.96, found: 1825.25, 1521.05, 1303.88, 1141.26, 1014.42, 913.23