Electronic Supplementary Material (ESI) for MedChemComm.

[Supporting Information]

Effects of the multiple *O*-glycosylation states on antibody recognition of the immunodominant motif in MUC1 extracellular tandem repeats

Shobith Rangappa,[a] Gerard Artigas,[a] Risho Miyoshi,[b] Yasuhiro Yokoi,[a] Shun Hayakawa,[a] Fayna Garcia-Martin,[a] Hiroshi Hinou,[a,b] and Shin-Ichiro Nishimura*[a,b]

[a]*Field of Drug Discovery Research, Faculty of Advanced Life Science, and Graduate School of Life Science, Hokkaido University, N21 W11, Sapporo, 001- 0021, Japan*

[b]*Medicinal Chemistry Pharmaceuticals, Co. Ltd, N21 W12, Sapporo 001- 0021, Japan*

*To whom correspondence should be addressed

E-mail: shin@sci.hokudai.ac.jp

Phone: +81-11-706-9043

Fax: +81-11-706-9042

Experimental section

General methods and materials. All commercially available chemicals and solvents were used without purification. NovaPEG Rink Amide resin (loading 0.37 mmol/g) and *N*α-[9-(fluorenylmethoxy)carbonyl (Fmoc)]-L-amino acids, except for glycosylated compounds, were purchased from Novabiochem Merck KG (Darmstadt, Germany). *N*α-Fmoc-amino acids having carbohydrate moiety, such as Fmoc-(Ac_3GalNAcα1→)Ser/Thr, Fmoc-(Ac_4Galβ1,3Ac_2GalNAcα1→)Ser/Thr and Fmoc-(Ac_4Galβ1,3[Ac_4GlcNAcβ1,6]Ac_2GalNAcα1→)Ser/Thr, were purchased from Medicinal Chemistry Pharmaceuticals, Co., Ltd. (Sapporo, Japan). 1-[Bis(dimethylamino)methyliumyl]-1H-benzotriazole-3-oxide hexafluorophosphate (HBTU), 1-hydroxybenzotriazole monohydrate (HOBt), (benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBOP), and *N,N*-diisopropylethylamine (DIEA) were purchased from Kokusan Chemical Co., Ltd (Tokyo, Japan). *N,N*-Dimethylformamide (DMF) and 20% piperidine in DMF were purchased from Watanabe Chemical Industries, Ltd. (Hiroshima, Japan). 1-Hydroxy-7-azabenzotriazole (HOAt), acetic anhydride, triisopropylsilane (TIS) and 2,2,2-trifluoroacetic acid (TFA) were purchased from Wako Pure Chemical Industries, Ltd. (Osaka, Japan). 5-Oxohexanoic acid was purchased from Tokyo Chemical Industry Co. Ltd. (Tokyo, Japan). All solid-phase reactions were performed manually in a polypropylene tube equipped with a filter (LibraTube, Hipep Laboratories, Kyoto, Japan). Cytidine-5'-monophospho-β-D-N-acetylneuraminic acid, disodium salt (CMP-Neu5Ac), and Uridine-5'-diphospho-D-galactose, disodium salt (UDP-Gal), were purchased from Yamasa Corporation (Chiba, Japan). Recombinant bovine D-glucose 4-β-D-galactosyltransferase (β1,4-GalT) and recombinant rat α2,3-(*O*)-sialyltransferase [α2,3-(*O*)-siaT] were purchased from Sigma Aldrich and Calbiochem Co. Ltd (Merck Millipore), respectively. Microwave-assisted reactions on the

resins were carried out on an EYELA microwave synthesizer Wave Magic (MWS-1000A, Tokyo Rikakikai Co., LTD., Tokyo, Japan) or on a Green Motif I microwave synthesis reactor (IDX Corp., Tochigi, Japan), in which a single-mode microwave was irradiated at 2450 MHz at 50ºC. Preparative HPLC purifications were performed on a Prominence Shimadzu HPLC system [(Shimadzu Corporation, Kyoto, Japan) equipped with two LC-6AD pumps, an SPD-20A UV/VIS detector at 220 nm for monitor, and Inertsil ODS-3 reversed-phase C-18 column (250 × 4.6 mm I.D., GL Sciences Inc., Tokyo, Japan). Analytical RP-HPLC was conducted using a Waters Acquity Ultra Performance LC system equipped with binary solvent delivery pump, an auto sampler and a UV detector and an Acquity UPLC BEN®C18 column (1.7 μm, 2.1 × 50 mm, Waters). High-resolution electrospray ionization mass spectra (ESI-HRMS) with JEOL JMS-700TZ, and amino acid analysis with a JEOL JLC-500/V equipped with ninhydrin detection system, were performed at the Center of Instrumental Analysis at Hokkaido University. MALDI-TOF mass spectra were measured on a Bruker Daltonics Ultraflex MALDI-TOF/TOF mass spectrometer using DHB as a matrix. Anti-MUC1 monoclonal antibodies, SM3 was purchased from Santa Cruz Biotechnology, Inc (Santa Cruz, CA, USA), DF3 mAb was from Covance Research Products, Inc. (Shirley, MA, USA), and anti-KL6 mAb labelled with alkaline phosphatase, Lumipulse® KL-6 Eisai (a reagent kit to determine serum KL6 antigen) was acquired from Sanko Junyaku Co., Ltd. (Tokyo, Japan). FluoroLinkTM CyTM3-labelled goat anti-mouse IgG (H+L) was from Amersham Biosciences (Buckinghamshire, UK). Microarray slides (75 x 25 x 1 mm) and hybridization covers (60 x 25 x 0.7 mm) were supplied from Sumitomo Bakelite Co., Ltd. (Tokyo, Japan). Fluorescence images of microarray slides were measured at 10 μm resolution on a Typhoon Trio plus variable mode imager (GE Healthcare) with a green laser (532 nm) and a 580 BP 30 filter at a PMT voltage of 600 V and normal sensitivity. The digital images of fluorescence responses were analysed using ArrayVisionTM software version 8.0 (GE Healthcare).

Synthesis of MUC1 glycopeptides.

MUC1 fragment as a model compound employed for the docking experiment in Figure 7 (an analogue of the compound **2** and **20** that firstly characterised as MUC1 glycopeptide involving a Tn-glycosylated PDTR motif showing high affinity with SM3 antibody in our previous paper, Matsushita, T., *et al*. *Biochim. Biophys. Acta* **2014**, *1840*, 1105-1116) was prepared by the present protocol used for the synthesis of abovementioned compound **2** and **20**.

General procedures for the preparative and analytical HPLC.

Preparative HPLC purifications were performed on a Prominence Shimadzu HPLC system [(Shimadzu Corporation, Kyoto, Japan) equipped with two LC-6AD pumps, an SPD-20A UV/VIS detector at 220 nm for monitor, and Inertsil ODS-3 reversed-phase C-18 column (250 × 4.6 mm I.D., GL Sciences Inc., Tokyo, Japan); flow rate, 5.0 mL/min; eluent A, H_2O with 0.1% TFA and eluent B, acetonitrile with 0.1% TFA for non-glycosylated peptides and glycopeptides having neutral sugar; or buffer A, containing 25 mM ammonium acetate, pH 5.5, and buffer B, acetonitrile containing 10% buffer A, for glycopeptides containing sialic acid residue; composition of the solvent, 0-60 min in a linear gradient flow from (A/B) = (98/2) to (70/30) or (A/B) = (90/10) to (55/45); detection, UV at 220 nm].

Analytical RP-HPLC was conducted using a Waters Acquity Ultra Performance LC system [equipped with binary solvent delivery pump, an auto sampler and a UV detector and an Acquity UPLC BEN®C18 column (1.7 μm, 2.1 × 50 mm, Waters); flow rate of 0.2 mL/min. the ratio of eluent B was linearly increased from 2 to 30% in 20.2 min, then the column was washed with 90% B and equilibrium with 2% B. Gradient condition A: eluent A, water with 0.1% TFA, and eluent B, acetonitrile with 0.1% TFA. Gradient condition B: eluent A, 50 mM ammonium acetate, pH 5.5, and eluent B, 10% eluent A in acetonitrile].

Structural Calculation. Three-dimensional structures of MUC1 glycopeptides were calculated using CNS 1.1[S1] according to the procedures reported previously (refs 17, 19, and 32 in the main text). Distance restraints for calculations were estimated from the cross peak intensities in the NOESY spectra having mixing time of 400 ms for all compounds used for structural studies after evaluating the results of the signal-to-noise ratio under the mixing time of 100, 200, and 400 ms. For the structural analysis, the restraints estimated were classified as strong, medium and weak signals and the upper limits of 2.6, 3.5, and 5.0 Å, respectively. For NOEs found only in the NOESY spectra with a mixing time of 400 m, the upper bond was set to 5.0 Å because no signal was obtained when the upper limit was set up to 6.0 Å. Dihedral angle restraints were based on $^{3}J_{HN\alpha}$ coupling constants which were measured by high resolution DQF-COSY. For $^{3}J_{HN\alpha}$ calculated more than 8.0 Hz, dihedral angle was constricted to -120 ± 30°. Complete RMSD values of secondary and tertiary structures were performed using PROCHECK[S2] and MOLMOL software.[S3]

References

S1. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., et al. Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D. Biol. Crystallogr*. **1998**, *54*, 905-921.

S2. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., Thornton, J. M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR, *J. Biomol. NMR* **1996**, *8*, 477-486.

S3. Koradi, R., Billeter, M., Wüthrich, K. MOLMOL: A Program for display and analysis of macromolecular structures, *J. Mol. Graph*. **1996**, *14*, 51-55.

Table S1. Summary of the dihedral angles for the MUC1 glycopeptides

Compounds		**8**	**14**	**17**	**18**	**19**
His1	ϕ	-127.8 ± 26.1	-123.2 ± 12.9	-117.0 ± 21.2	-121.9 ± 13.7	-128.8 ± 89.0
His1	χ1	-161.8 ± 85.8	-120.8 ± 98.3	-103.1 ± 62.3	-81.5 ± 85.9	-151.5 ± 54.4
His1	ψ	61.2 ± 84.4	81.9 ± 82.3	44.3 ± 80.1	55.2 ± 81.0	133.3 ± 98.8
Gly2	ϕ	-120.4 ± 11.4	-167.1 ± 67.3	-129.0 ± 17.2	-117.0 ± 13.8	-118.2 ± 14.2
Gly2	ψ	142.9 ± 89.0	116.9 ± 102.5	165.8 ± 95.4	136.2 ± 94.2	109.1 ± 42.5
Val3	ϕ	-113.6 ± 27.1	-113.0 ± 40.0	-122.5 ± 30.5	-124.6 ± 13.8	-94.6 ± 1.3
Val3	χ1	147.1 ± 89.1	145.1 ± 83.2	116.6 ± 79.4	-89.3 ± 91.3	-167.8 ± 0.4
Val3	ψ	137.8 ± 34.5	111.1 ± 95.1	25.2 ± 91.8	131.1 ± 92.3	18.5 ± 1.3
Thr4	ϕ	-116.0 ± 14.4	-110.7 ± 19.4	-103.5 ± 15.2	-141.2 ± 0.2	-124.6 ± 36.9
Thr4	χ1	14.3 ± 33.7	42.0 ± 43.3	8.7 ± 42.1	31.6 ± 76.5	48.5 ± 32.2
Thr4	ψ	50.0 ± 75.4	-141.4 ± 60.0	79.8 ± 79.2	-48.3 ± 44.1	-85.0 ± 88.5
Ser5	ϕ	-120.4 ± 14.8	-111.9 ± 28.8	-106.6 ± 21.9	-147.4 ± 33.5	-79.7 ± 0.1
Ser5	χ1	-30.0 ± 74.4	116.3 ± 92.0	173.7 ± 42.5	175.1 ± 85.5	-166.0 ± 5.8
Ser5	ψ	176.5 ± 81.2	151.1 ± 51.1	42.5 ± 29.8	142.4 ± 51.6	172.5 ± 10.5
Ala6	ϕ	-120.3 ± 14.6	-113.6 ± 28.4	-117.4 ± 26.8	-118.6 ± 16.8	-144.0 ± 34.7
Ala6	χ1	-3.1 ± 95.9	-22.9 ± 103.8	-142.3 ± 84.7	-133.7 ± 98.6	102.2 ± 102.6
Ala6	ψ	114.9 ± 66.1	115.2 ± 86.8	95.4 ± 51.1	150.0 ± 90.4	156.3 ± 80.9
Pro7	ψ	141.2 ± 93.8	146.8 ± 77.3	148.7 ± 91.6	-177.9 ± 6.7	133.7 ± 66.5
Asp8	ϕ	-88.0 ± 28.0	-129.4 ± 34.8	-137.4 ± 14.1	-133.2 ± 6.6	-123.1 ± 39.5
Asp8	χ1	-69.8 ± 73.2	-49.1 ± 84.3	-114.4 ± 90.9	111.5 ± 1.5	-126.2 ± 36.6
Asp8	ψ	-158.6 ± 45.8	157.4 ± 57.0	48.2 ± 78.0	-172.0 ± 1.1	-53.4 ± 15.3
`Thr9	ϕ	-127.1 ± 37.9	-97.6 ± 1.4	-133.2 ± 14.8	-119.0 ± 14.9	-134.1 ± 34.8
Thr9	χ1	-1.9 ± 71.8	-14.8 ± 25.7	2.9 ± 52.3	-49.9 ± 80.5	11.9 ± 89.4
Thr9	ψ	-35.0 ± 46.1	24.5 ± 79.4	134.8 ± 24.2	-60.6 ± 47.4	-24.2 ± 38.7
Arg10	ϕ	-132.8 ± 39.6	-119.4 ± 14.3	-138.9 ± 25.1	-122.3 ± 14.8	-53.6 ± 12.4
Arg10	χ1	-112.4 ± 73.1	-98.5 ± 65.3	-119.8 ± 16.5	-107.7 ± 75.9	-175.6 ± 33.3
Arg10	ψ	114.0 ± 86.4	110.2 ± 34.1	102.3 ± 17.4	118.5 ± 29.0	-112.4 ± 74.9
Pro11	ψ	123.4 ± 86.3	138.4 ± 73.2	8.6 ± 98.0	158.2 ± 65.1	142.2 ± 85.0
Ala12	ϕ	-130.4 ± 29.17	-136.0 ± 35.0	-131.3 ± 28.8	-138.7 ± 34.6	-119.1 ± 81.3
Ala12	χ1	33.3 ± 80.5	-48.0 ± 93.2	-66.1 ± 82.4	-107.4 ± 90.2	15.3 ± 92.6
Ala12	ψ	121.3 ± 85.0	-65.0 ± 2.3	121.1 ± 70.9	-63.1 ± 19.7	132.6 ± 91.6
Pro13	ψ	136.7 ± 68.6	109.7 ± 84.0	170.9 ± 103.4	99.8 ± 84.2	126.7 ± 69.1
Gly14	ϕ	-124.6 ± 13.8	-125.5 ± 25.3	-125.9 ± 18.6	-125.5 ± 30.5	-104.1 ± 55.7
Gly14	ψ	148.6 ± 86.7	-176.6 ± 61.8	-149.6 ± 71.3	175.0 ± 53.8	92.6 ± 83.1
Ser15	ϕ	-111.1 ± 22.3	-158.8 ± 3.1	-99.2 ± 16.2	-139.3 ± 3.9	-41.1 ± 34.0
Ser15	χ1	-38.1 ± 75.9	142.4 ± 65.3	43.3 ± 79.3	-160.1 ± 21.3	-53.7 ± 42.0
Ser15	ψ	152.4 ± 87.7	63.1 ± 53.7	-165.0 ± 37.9	72.5 ± 26.9	-130.2 ± 45.1
Thr16	ϕ	-113.0 ± 22.0	-106.6 ± 38.1	-120.5 ± 0.2	-105.8 ± 7.6	-96.6 ± 98.0
Thr16	χ1	-15.9 ± 43.0	2.1 ± 42.8	-50.1 ± 64.3	-34.7 ± 35.3	-68.7 ± 63.7
Thr16	ψ	-93.5 ± 94.0	-35.8 ± 22.9	159.5 ± 78.4	-49.5 ± 17.8	-32.3 ± 24.3
Ala17	ϕ	-116.6 ± 13.8	-138.1 ± 19.9	-136.1 ± 21.1	-126.2 ± 15.0	81.5 ± 94.0
Ala17	χ1	42 ± 80.9	0.4 ± 100.7	92.6 ± 106.8	-144 ± 89.6	-144.9 ± 100.9
Ala17	ψ	125.9 ± 62.3	-81.0 ± 58.0	1.5 ± 85.4	-73.5 ± 58.2	121.1 ± 69.9
Pro18	ψ	139.8 ± 55.4	138.8 ± 27.6	154.7 ± 87.0	132.3 ± 68.6	159.0 ± 77.1
Pro19	ψ	72.6 ± 75.6	94.2 ± 47.8	51.7 ± 0.4	130.5 ± 65.3	133.2 ± 87.1
Ala20	ϕ	-141.6 ± 72.9	-119.3 ± 24.4	-166.6 ± 1.3	-153.8 ± 49.0	-123.8 ± 24.6
Ala20	χ1	9.7 ± 80.6	169.1 ± 85.8	42.9 ± 97.0	-82.0 ± 94.3	31.5 ± 104.6

Table S2. Statistical analysis of NMR restraints for MUC1 glycopeptides

compound	**8**	**14**	**17**	**18**	**19**
no. of distance restraints					
total	163	280	238	214	134
intraresidue					
peptide	77	76	111	75	67
glycan	34	141	76	79	35
sequential ($\|i-j\|=1$)					
peptide	41	46	44	43	32
glycan	11	16	7	16	0
medium range ($2<\|i-j\|<4$)					
peptide	0	1	0	1	0
glycan	0	0	0	0	0
long range ($\|i-j\|>5$)					
peptide	0	0	0	0	0
glycan	0	0	0	0	0
peptide to glycan					
within the same glycosylated residue	33	32	28	27	16
glycans on other peptide residues	12	19	10	11	15
no. of dihedral restraints					
total	84	135	60	114	54
peptide backbone	14	15	14	14	14
glycan	70	120	46	100	40

compound	**8**	**14**	**17**	**18**	**19**
average potential energy (kcal/mol)[a]					
E_{total}	110.83 ± 10.09	140.76 ± 20.33	99.78 ± 4.24	140.45 ± 9.99	79.98 ± 4.03
E_{bond}	3.37 ± 0.41	7.58 ± 1.05	7.52 ± 0.33	7.11 ± 0.46	6.61 ± 0.41
E_{angle}	20.67 ± 2.79	46.57 ± 6.45	31.26 ± 1.31	46.43 ± 3.02	15.22 ± 0.89
E_{impr}[b]	6.79 ± 0.71	12.74 ± 1.02	4.49 ± 0.5	13.23 ± 0.48	2.51 ± 0.16
E_{VDW}[b]	12.3 ± 2.42	25.92 ± 7.49	7.82 ± 2.23	21.09 ± 5.16	8.96 ± 1.69
E_{NOE}[b]	18.83 ± 1.92	30.54 ± 5.89	42.89 ± 1.53	36.29 ± 2.35	42.3 ± 1.96
E_{cdih}[b]	0.06 ± 0.03	1.4 ± 1.13	0.14 ± 0.1	0.47 ± 0.24	0.06 ± 0.02
deviation from idealized geometry					
bond lengths (Å)	0.0026 ± 0.00021	0.0037 ± 0.0003	0.0044 ± 0.0001	0.0034 ± 0.00013	0.0040 ± 0.0001
bond angles (deg)	0.927 ± 0.051	1.099 ± 0.096	1.038 ± 0.299	1.108 ± 0.032	0.716 ± 0.015
impropers (deg)	0.557 ± 0.029	0.666 ± 0.022	0.474 ± 0.029	0.696 ± 0.012	0.372 ± 0.012
Average pairwise rmsd (Å)					
backbone atoms					
Val 3-Ser5	0.66 ± 0.35	0.68 ± 0.28	0.60 ± 0.29	0.57 ± 0.31	0.72 ± 0.35
Asp8-Thr9	0.36 ± 0.20	0.35 ± 0.26	0.44 ± 0.28	0.10 ± 0.07	0.42 ± 0.38
Pro11-Pro13	0.73 ± 0.29	0.46 ± 0.21	0.68 ± 0.25	0.45 ± 0.26	1.14 ± 0.63
Ser15-Ala17	0.76 ± 0.29	0.38 ± 0.18	0.64 ± 0.46	0.30 ± 0.14	1.14 ± 0.65
heavy atoms					
Val 3-Ser5	1.45 ± 0.52	1.46 ± 0.49	1.39 ± 0.60	1.39 ± 0.58	1.34 ± 0.60
Asp8-Thr 9	1.24 ± 0.34	1.15 ± 0.32	1.35 ± 0.42	0.81 ± 0.34	1.28 ± 0.49
Pro11-Pro13	1.40 ± 0.55	0.87 ± 0.35	1.32 ± 0.47	0.89 ± 0.46	1.76 ± 0.69
Ser15-Ala17	1.47 ± 0.46	0.77 ± 0.34	1.14 ± 0.56	0.59 ± 0.23	1.72 ± 0.69
Val 3-Ser5 and GalNAc4	1.72 ± 0.60	1.82 ± 0.89	2.07 ± 1.00	1.99 ± 1.04	1.33 ± 0.49
Val3-Ser5 and GalNAc5	2.45 ± 0.82	2.06 ± 0.61	1.93 ±0.87	2.30 ± 0.75	2.08 ± 0.68
Val3-Ser5, GalNAc4 and 5	2.75 ± 0.83	2.36 ± 0.78	2.59 ± 1.08	2.72 ± 0.95	2.38 ± 0.74
Asp8-Thr9 and GalNAc9	2.03 ± 0.61	1.93 ± 0.85	1.46 ± 0.49	1.65 ± 0.72	2.03 ± 0.83
Ser15-Ala17 and GalNAc15	2.53 ± 0.58	1.31 ± 0.52	1.70 ± 0.57	1.20 ± 0.53	2.39 ± 1.01
Ser15-Ala17 and GalNAc16	2.29 ± 0.69	1.07 ± 0.86	1.54 ± 0.84	1.04 ± 0.66	2.55 ± 0.91
Ser15-Ala17, GalNAc15 and 16	3.01 ± 0.65	1.55 ± 0.75	2.20 ± 0.76	1.44 ± 0.60	3.12 ± 1.08

[a]All energies and rmsd values were calculated using CNS1.1 and MOLMOL, respectively. [b]E_{impr}, E_{VDW}, E_{NOE} and E_{cdih} are the improper torsion angle energy, the van der Waals energy, the square-well NOE potential energy, and the dihedral potential energy, respectively. The force constants for the calculations of E_{VDW}, E_{NOE} and E_{cdih} were 4.0 kcal mol^{-1} Å^{-1}, 50 kcal mol^{-1} Å^{-1}, and 200 kcal mol^{-1} rad^{-2}, respectively. Dihedral angle restraints of the glycan moieties were not derived from the NMR measurements but from the individual numbers of the residual glycans fixed as a chair conformation.

Figure S1. Lowest-energy structure in the 30 calculated models for MUC1 glycopeptides **14** (top), **18** (middle), and **19** (bottom). Van der Waals surface of the peptide backbone (blue) and five carbohydrate branches indicated by wireframe representation of different colours.

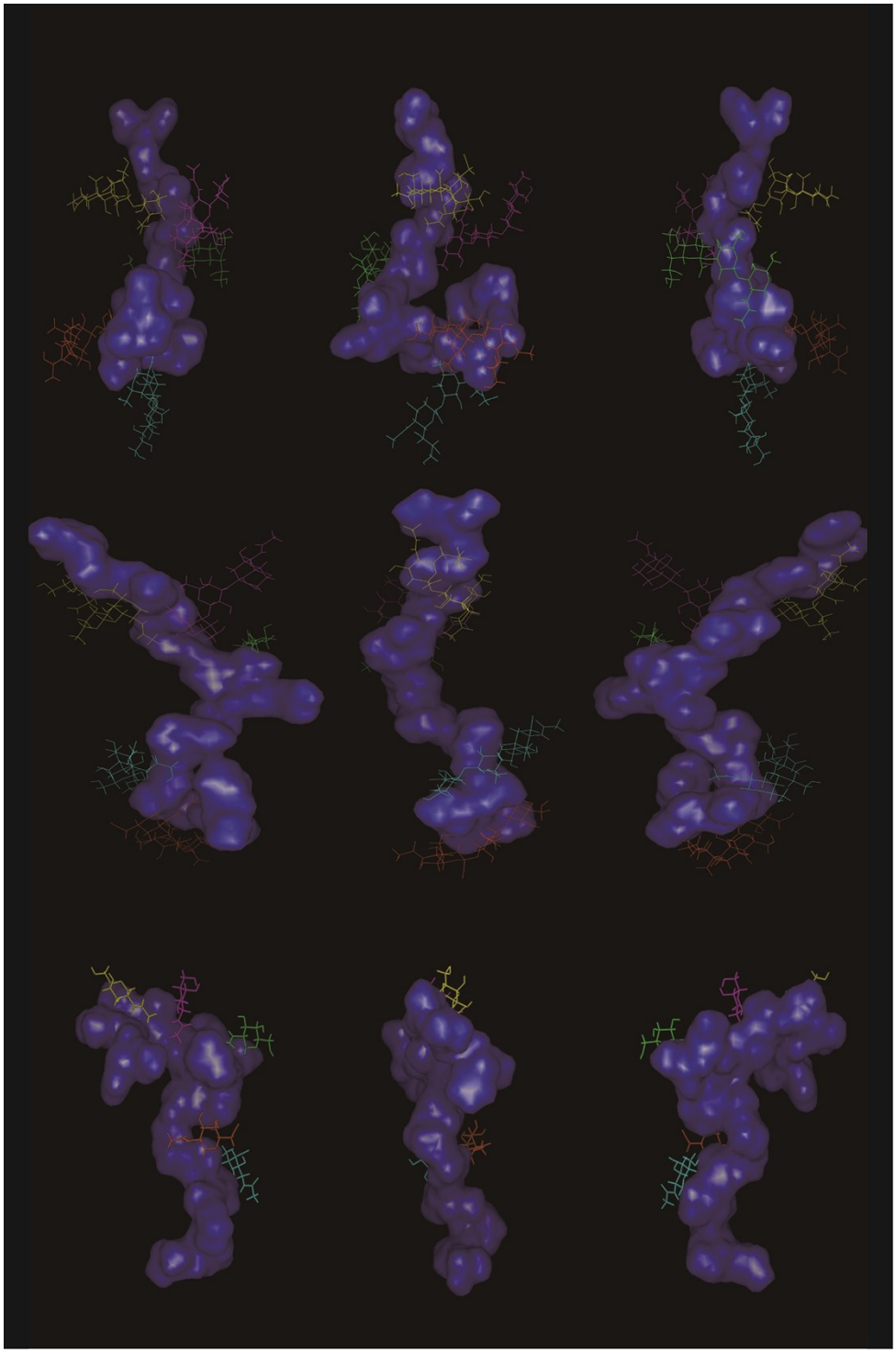
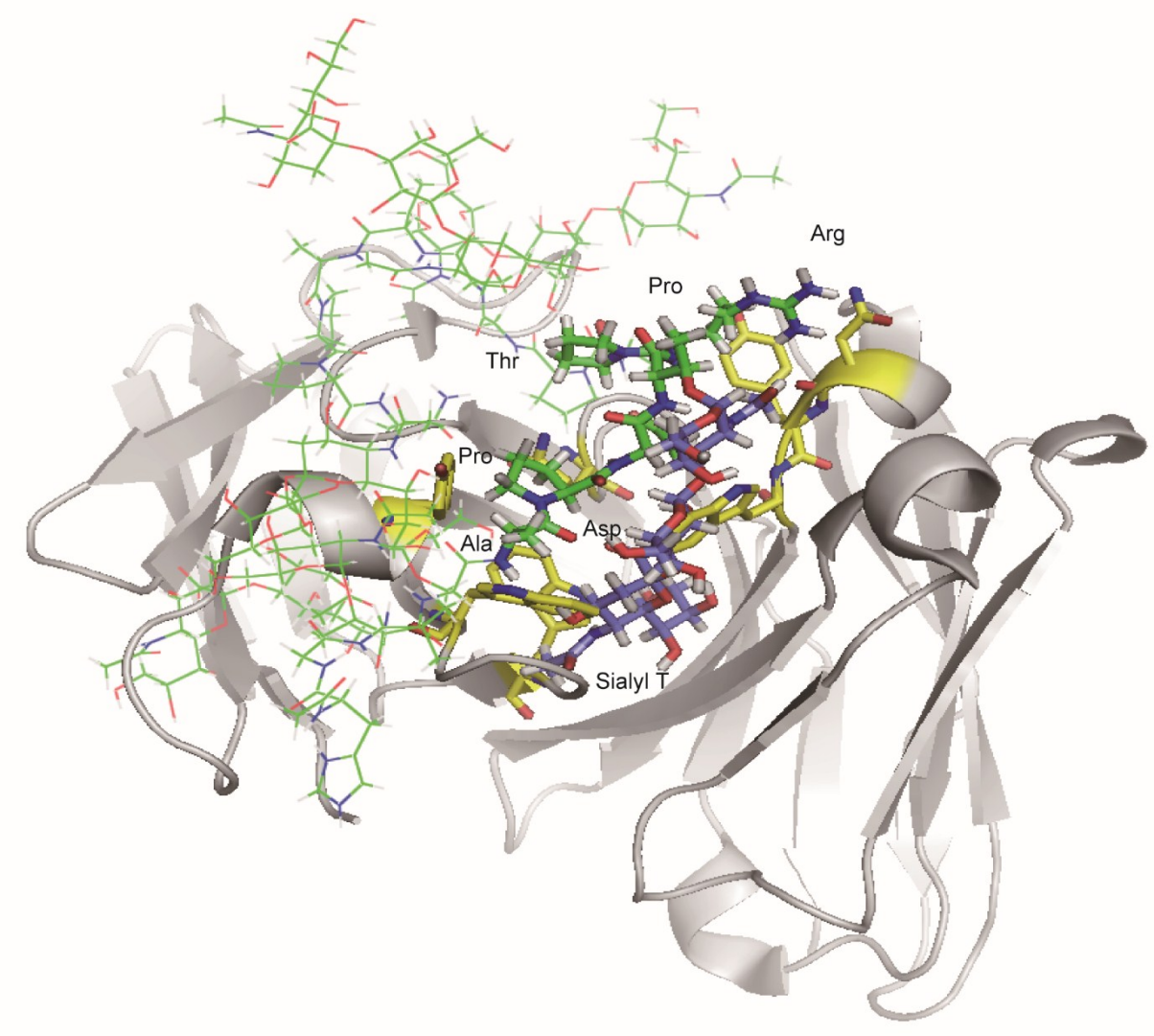
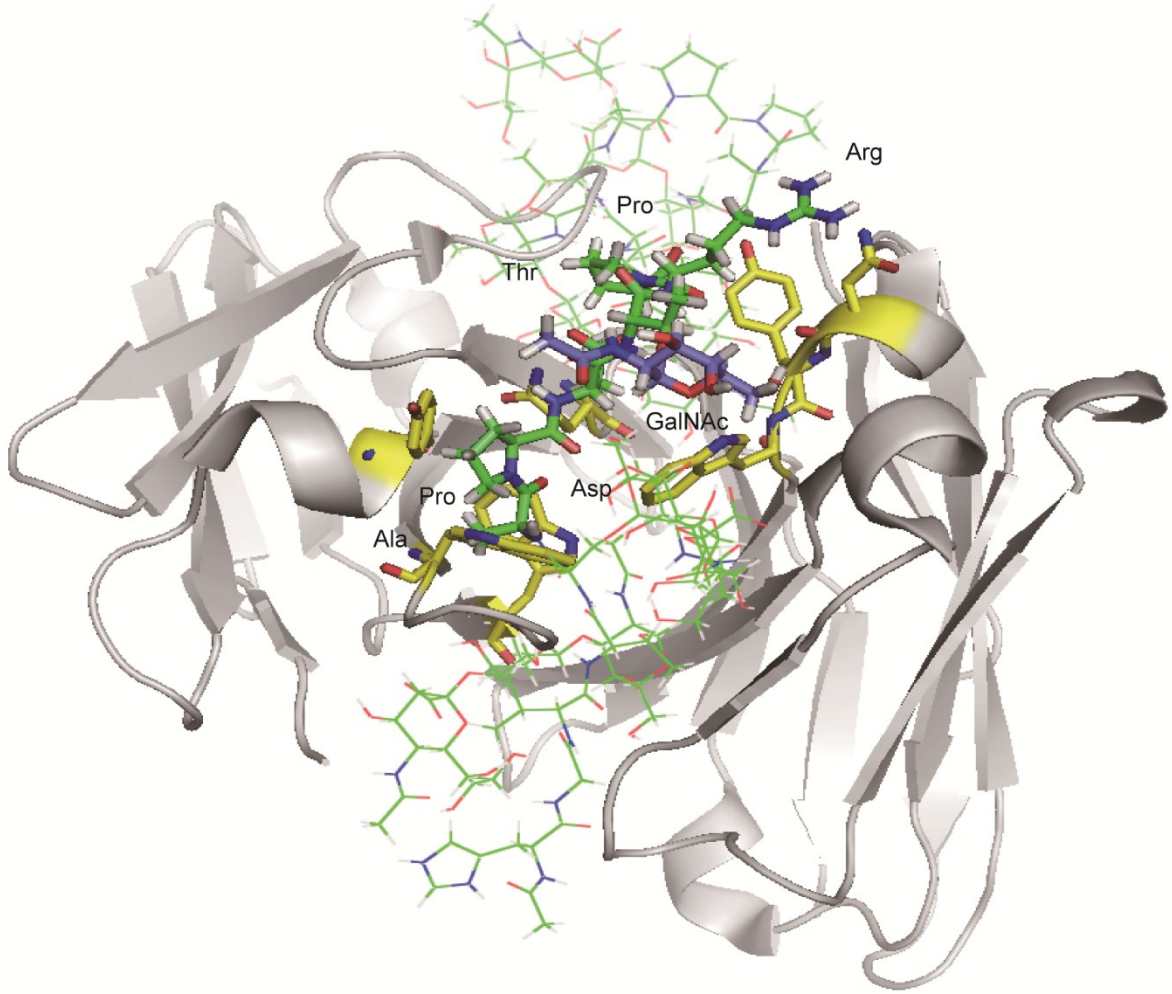


Figure S2. (A) Superposition of the NMR structure of compound **14** having five ST antigens at all potential *O*-glycosylation sites with the X-ray structure of the binding pocket in the SM3 derived from PDB data (PDB entry 5A2J). (B) Superposition of the NMR structure of compound **18** having the Tn antigen at PDTR motif and four ST antigens at other *O*-glycosylation sites with the X-ray structure of the binding pocket in the SM3 derived from PDB data (PDB entry 5A2J). (C) Superposition of the NMR structure of compound **19** having five Tn antigens at all potential *O*-glycosylation sites with the X-ray structure of the binding pocket in the SM3 derived from PDB data (PDB entry 5A2J).

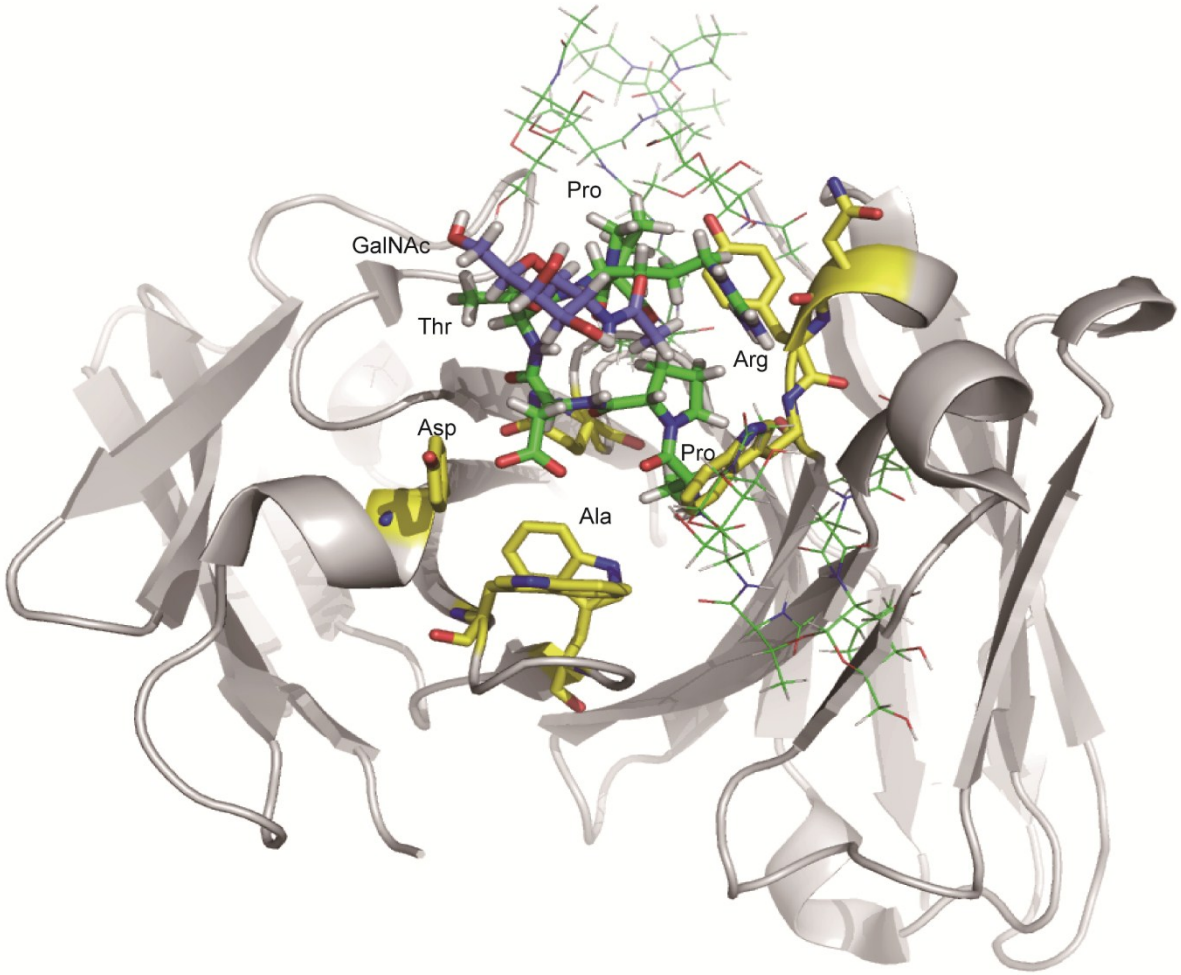
A



B

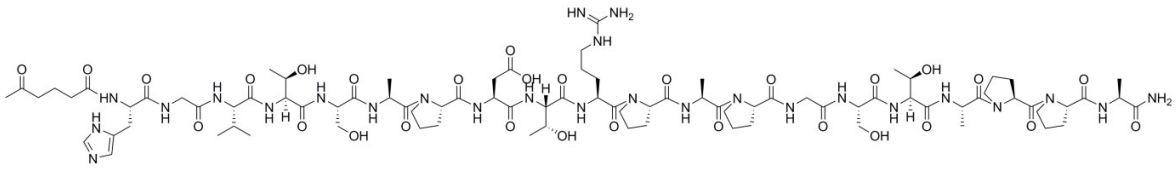


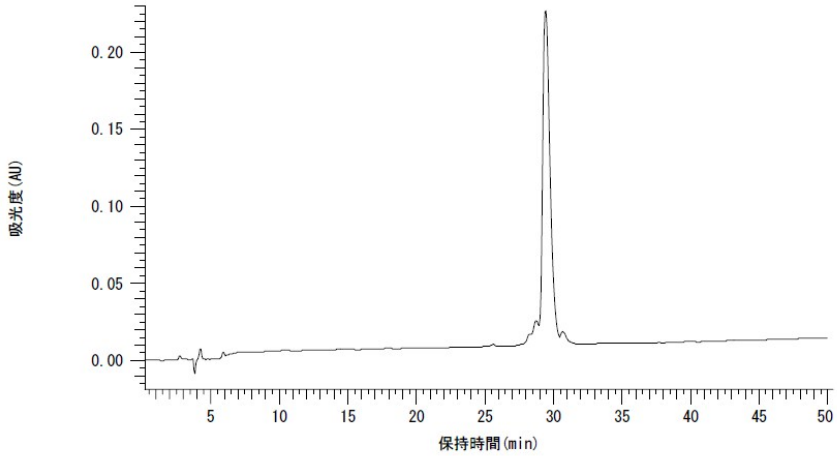
C

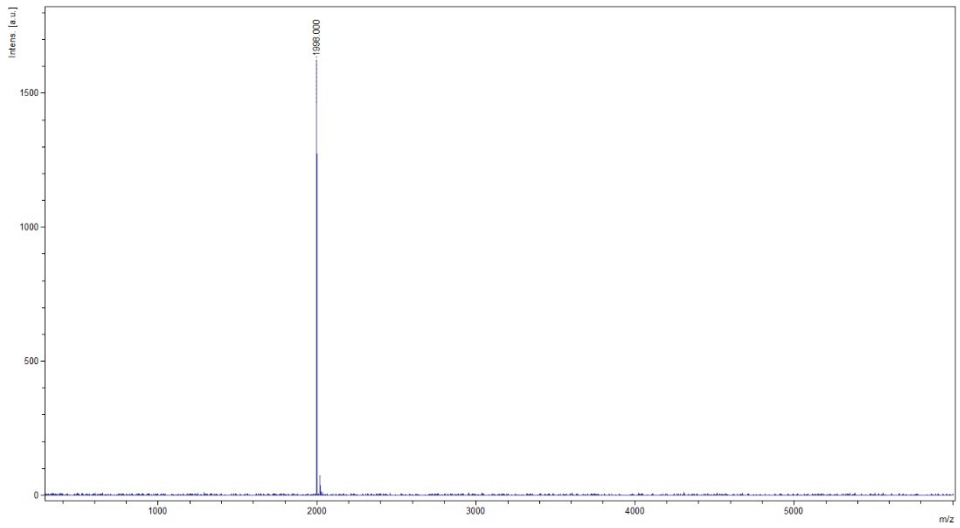


Characterisation of MUC1 glycopeptides.

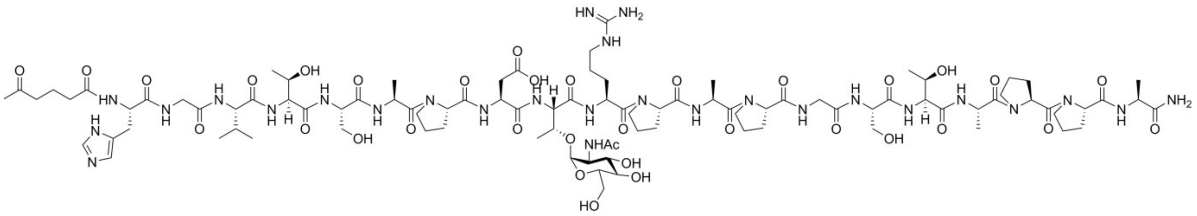
Compound **1**. 9.0 mg (4.5 μmole, 19% overall yield calculated from the resin employed). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-30% MALDI-TOFMS: $C_{86}H_{136}N_{26}O_{29}$, Theoretical mass: 1996.997, Observed mass: 1998.000 $[M+H]^+$

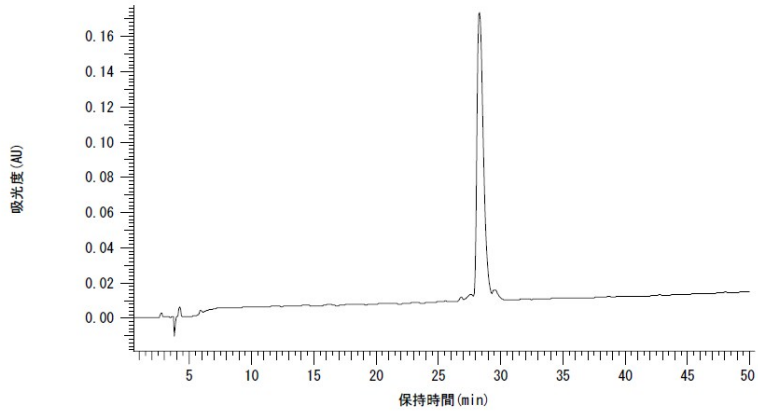


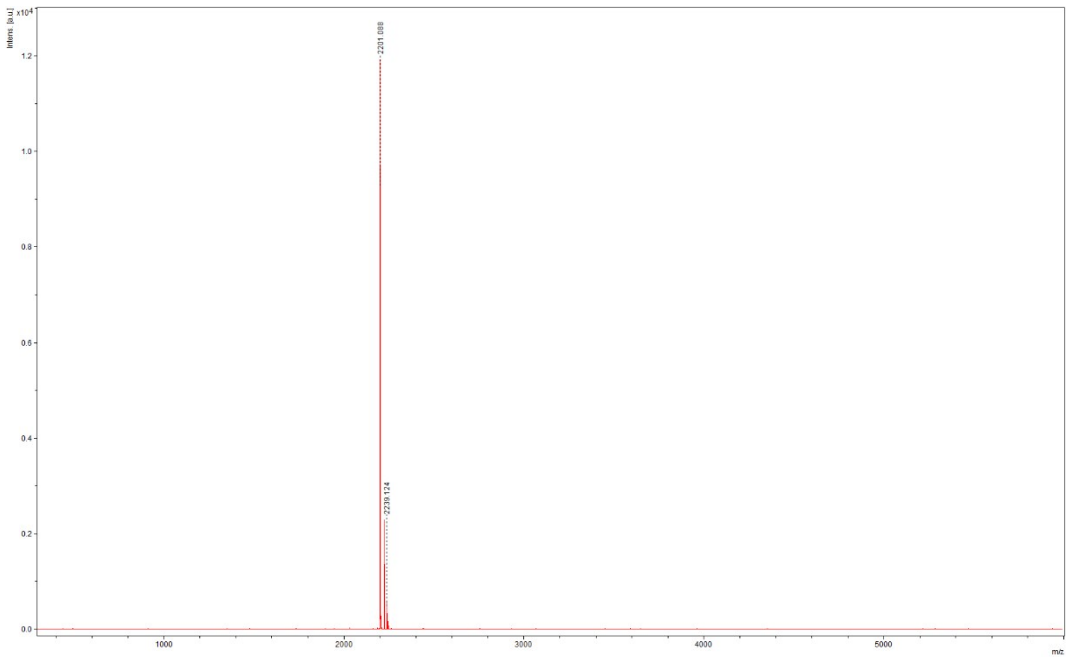




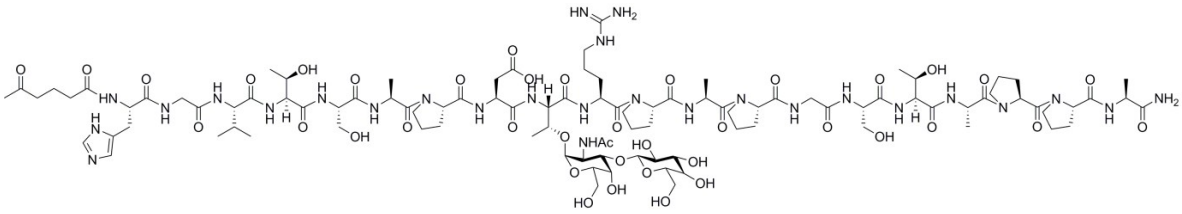
Compound **2**. 11.0 mg (5.0 μmole, 20% overall yield calculated from the resin employed). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-30% MALDI-TOFMS: $C_{94}H_{149}N_{27}O_{34}$, Theoretical mass 2200.076, Observed mass: 2201.088 $[M+H]^+$

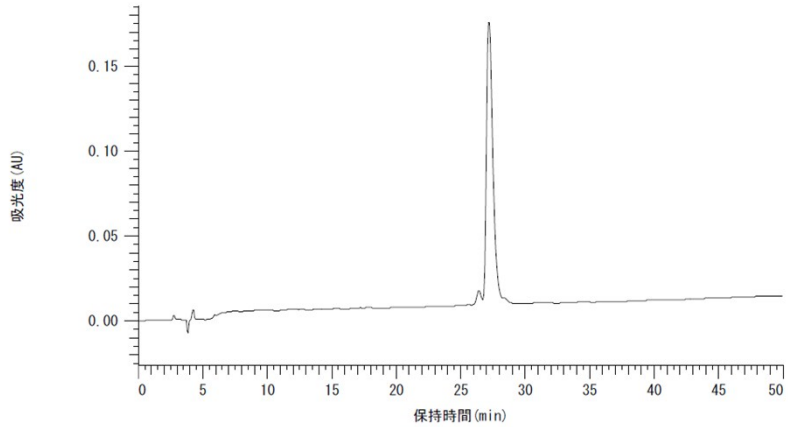


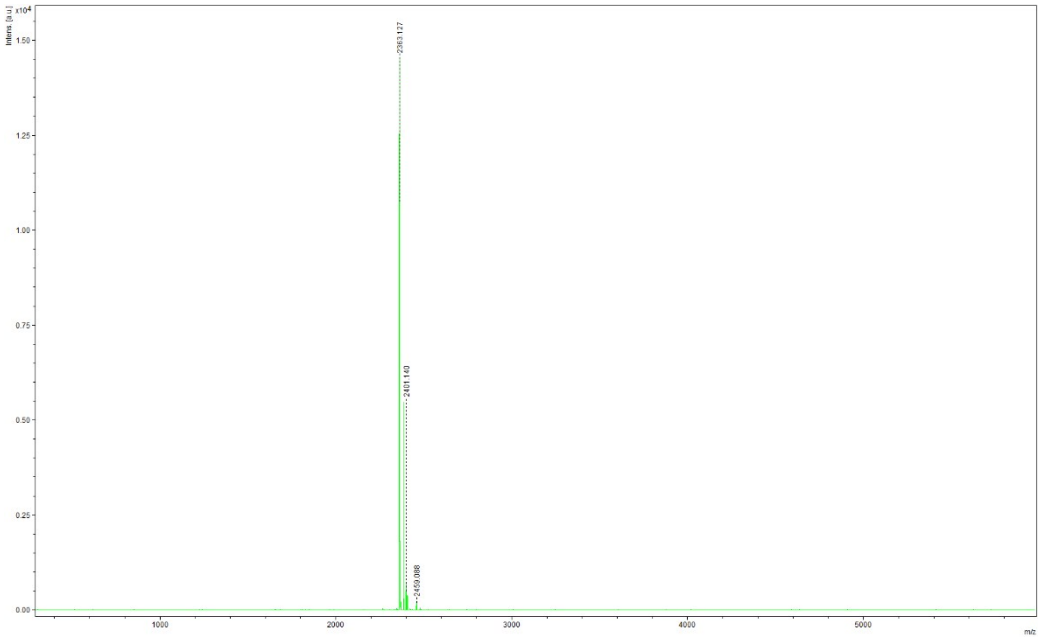




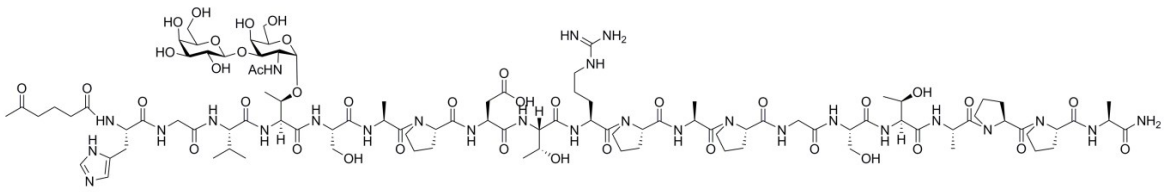
Compound **3**. 13.5 mg (5.7 μmole, 24% overall yield calculated from the resin employed). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-30% MALDI-TOFMS: $C_{100}H_{159}N_{27}O_{39}$ Theoretical mass: 2362.129, Observed mass: 2363.127 $[M+H]^+$

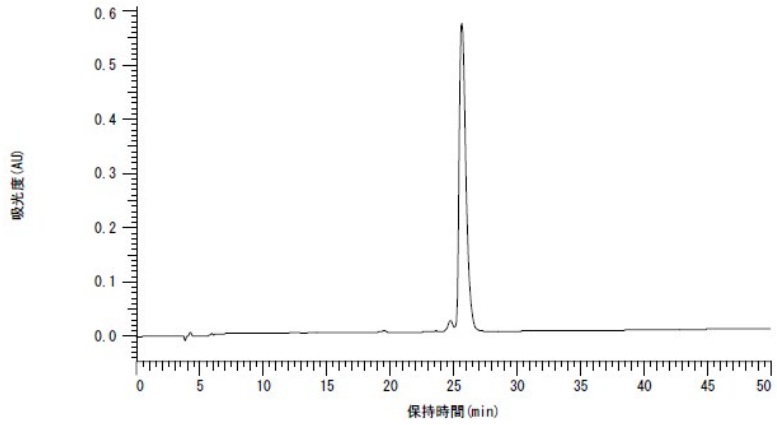


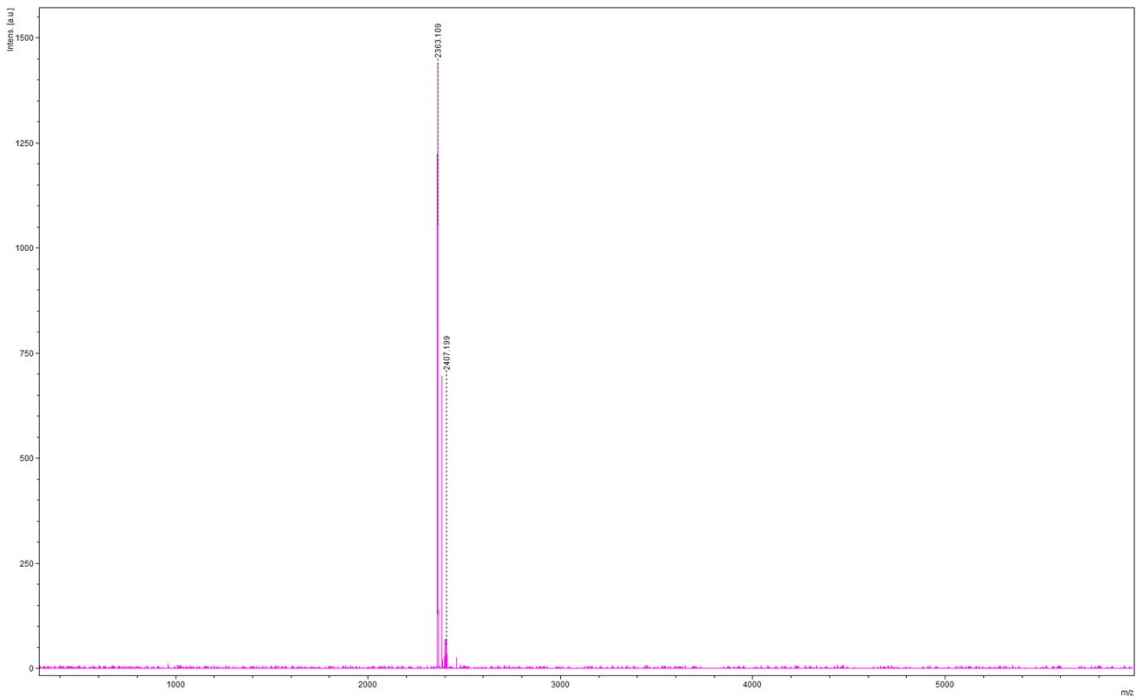




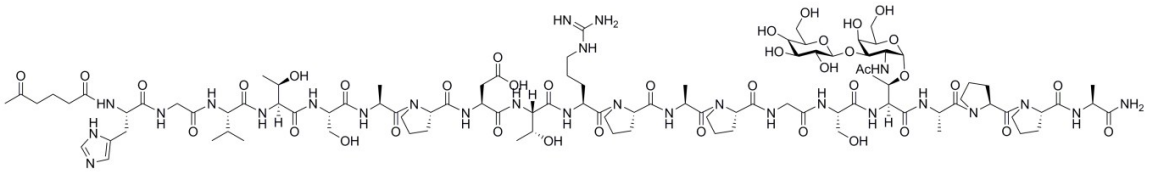
Compound **4**. 14.4 mg (6.1 μmole, 25% overall yield calculated from the resin employed). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-30% MALDI-TOFMS: $C_{100}H_{159}N_{27}O_{39}$, Theoretical mass: 2362.129, Observed mass: 2363.109 $[M+H]^+$

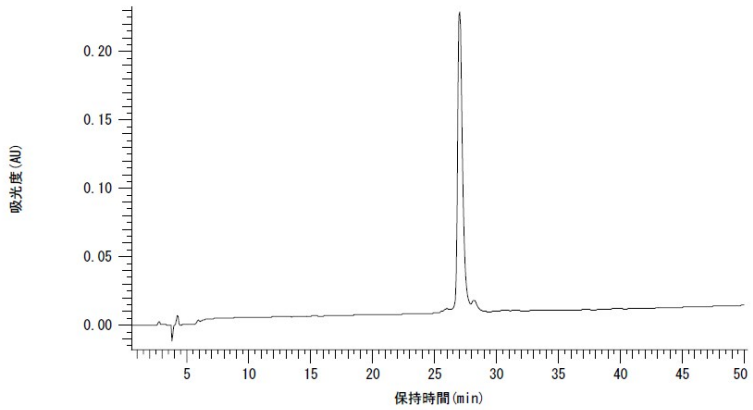


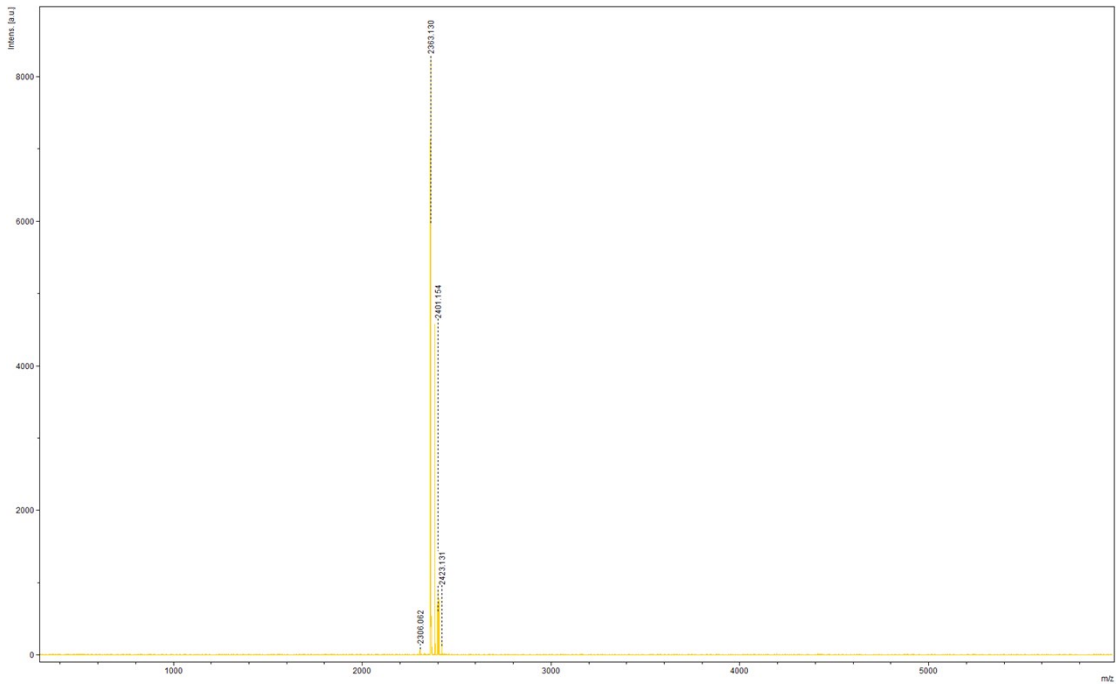




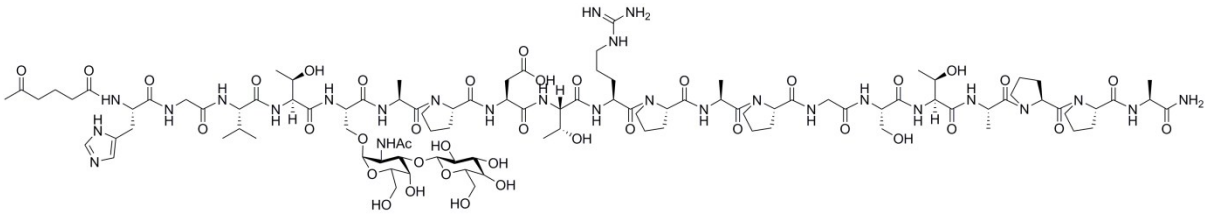
Compound **5**. 9.6 mg (4.1 μmole, 17% overall yield calculated from the resin employed, 50 mg, 24 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-30% MALDI-TOFMS: $C_{100}H_{159}N_{27}O_{39}$, Theoretical mass: 2362.129, Observed mass: 2363.130 $[M+H]^+$

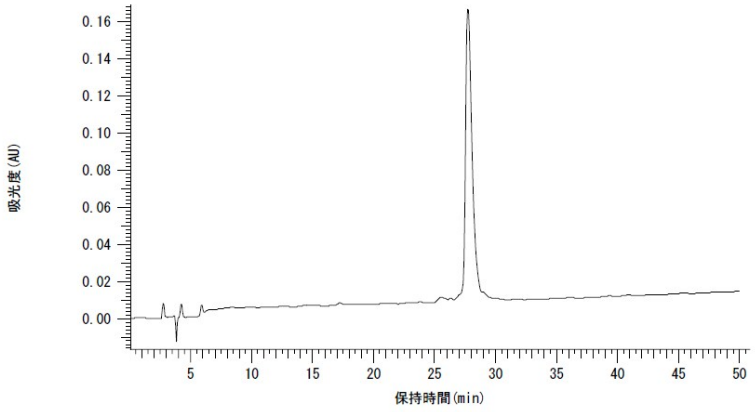


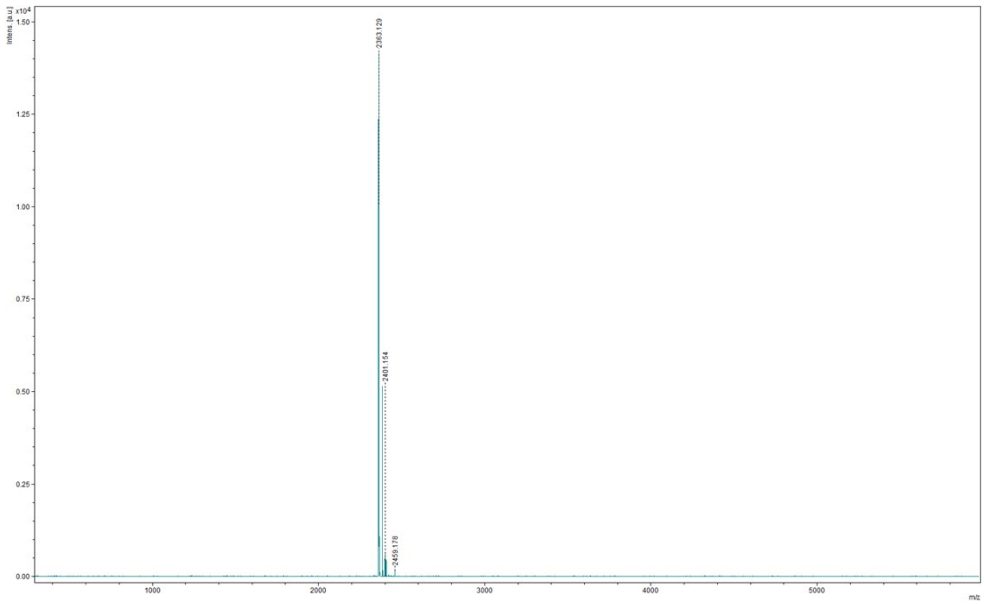




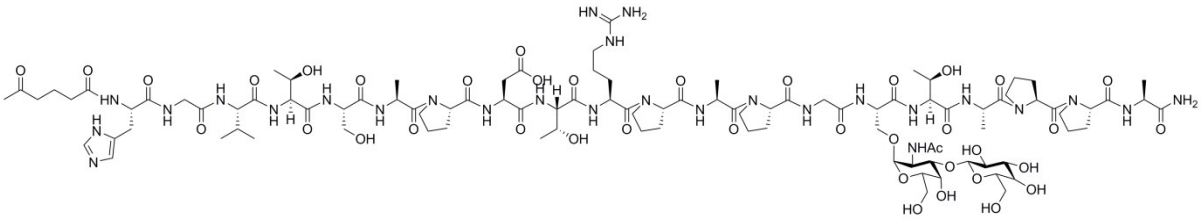
Compound **6**. 10.7 mg (4.5 μmole, 19% overall yield calculated from the resin employed, 50 mg, 24 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-30% MALDI-TOFMS: $C_{100}H_{159}N_{27}O_{39}$, Theoretical mass: 2362.129, Observed mass: 2363.129 $[M+H]^+$

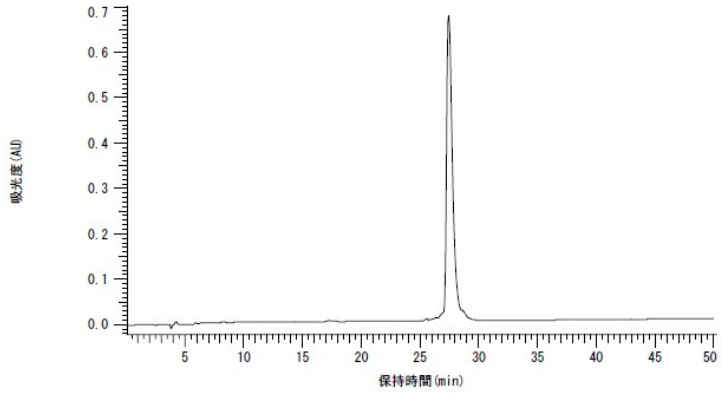


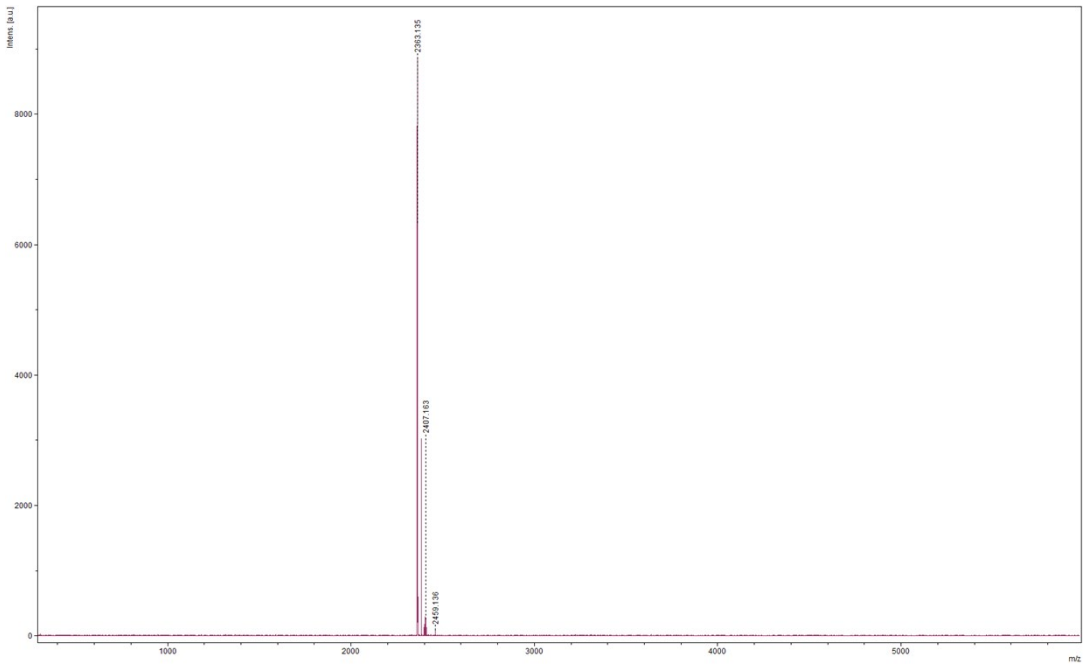




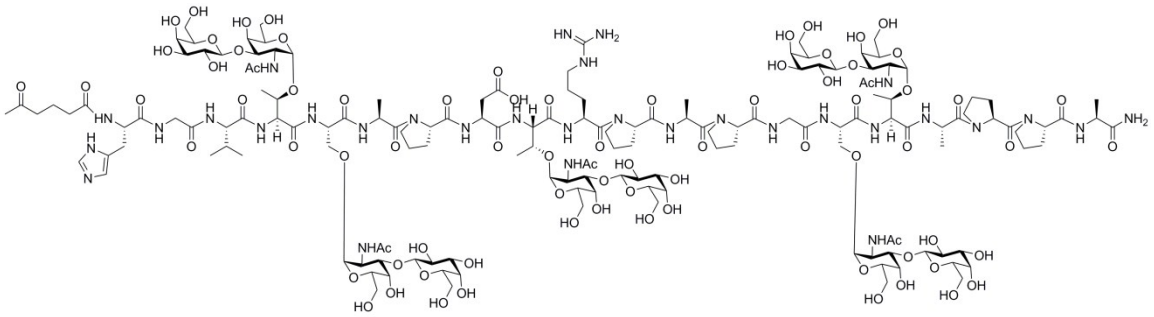
Compound **7**. 11.1 mg (4.7 μmole, 20% overall yield calculated from the resin employed, 50 mg, 24 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-30% MALDI-TOFMS: $C_{100}H_{159}N_{27}O_{39}$, Theoretical mass: 2362.129, Observed mass: 2363.135 $[M+H]^+$

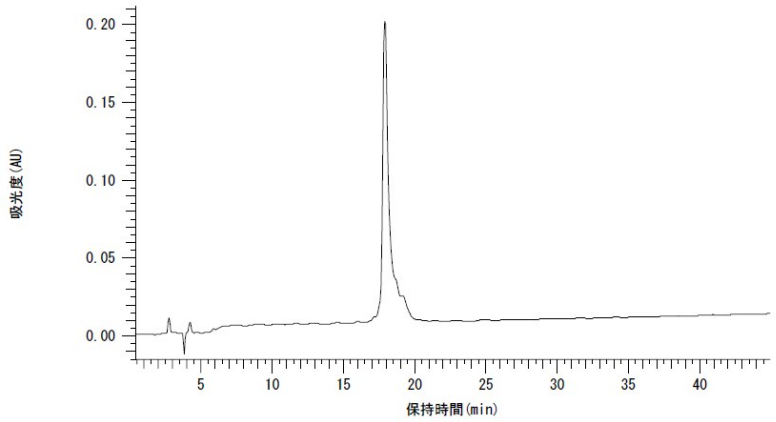


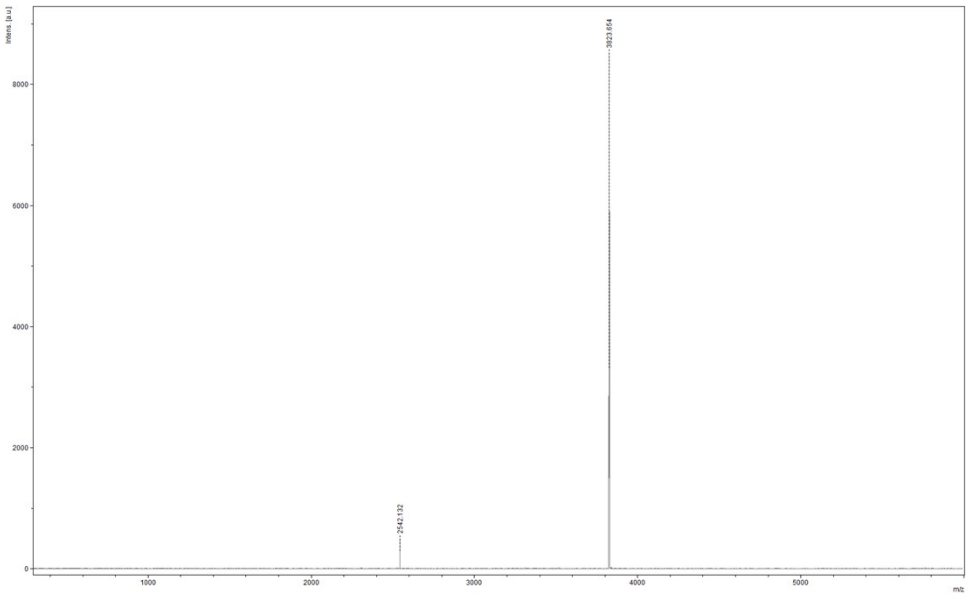




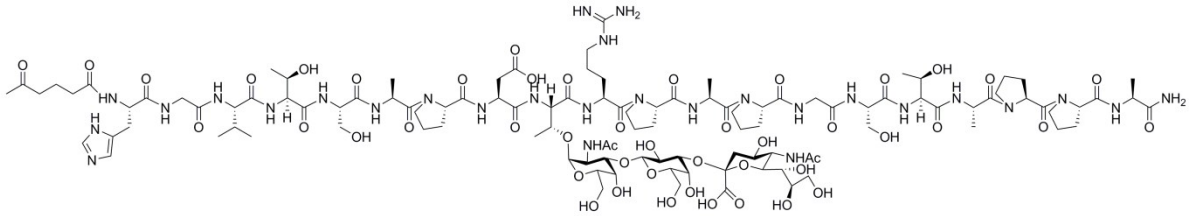
Compound **8**. 7.7 mg (2.0 μmole, 8% overall yield calculated from the resin employed, 50 mg, 24 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-30% MALDI-TOFMS: $C_{156}H_{251}N_{31}O_{79}$, Theoretical mass: 3822.658, Observed mass: 2654.218 $[M+H]^+$

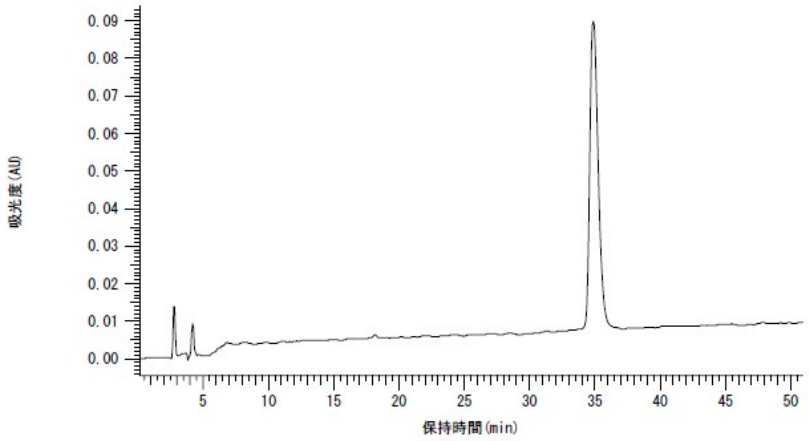


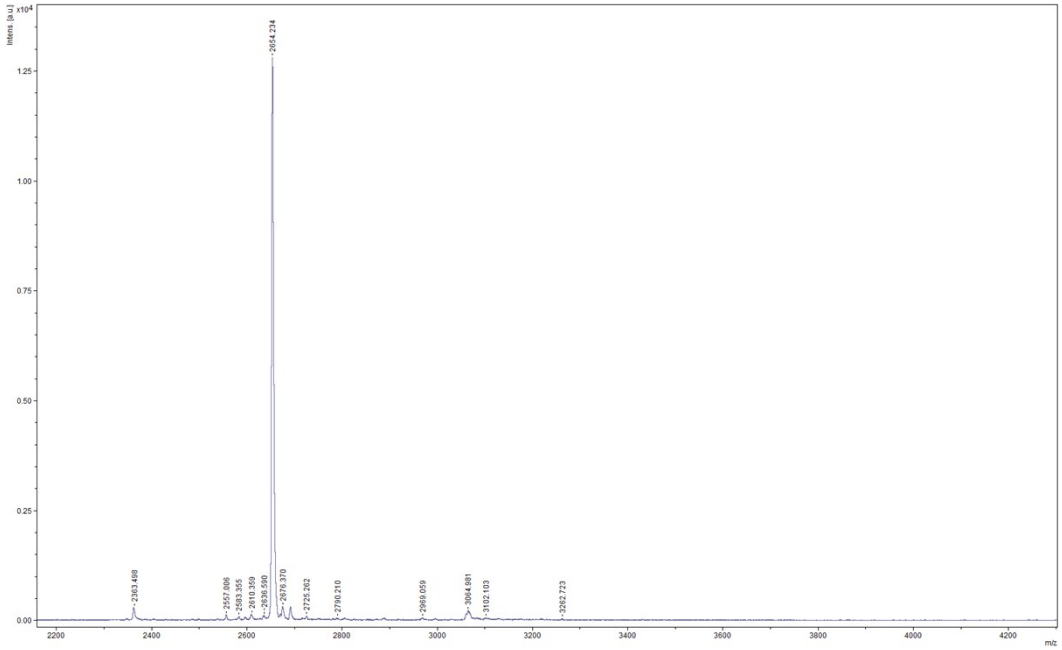




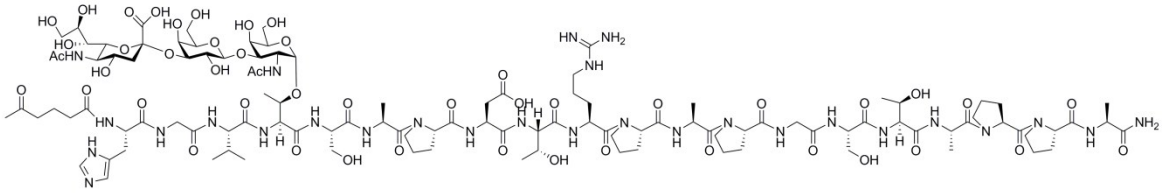
Compound **9**. 1.3 mg (0.49 μmole, 57% yield calculated from the starting compound **3**, 0.85 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-30% MALDI-TOFMS: $C_{111}H_{176}N_{28}O_{47}$, Theoretical mass: 2653.224, Observed mass: 2654.234 $[M+H]^+$

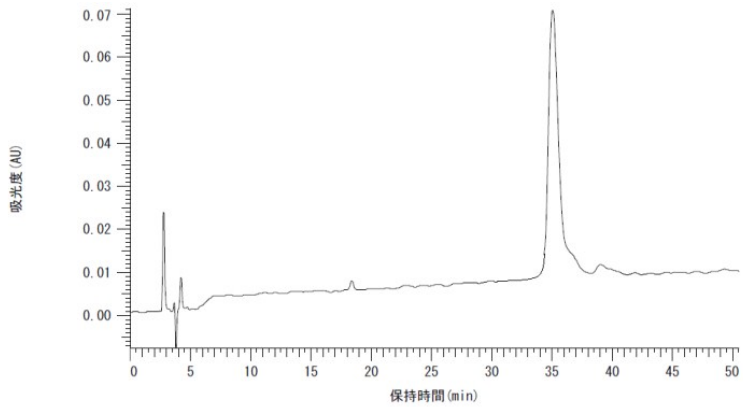


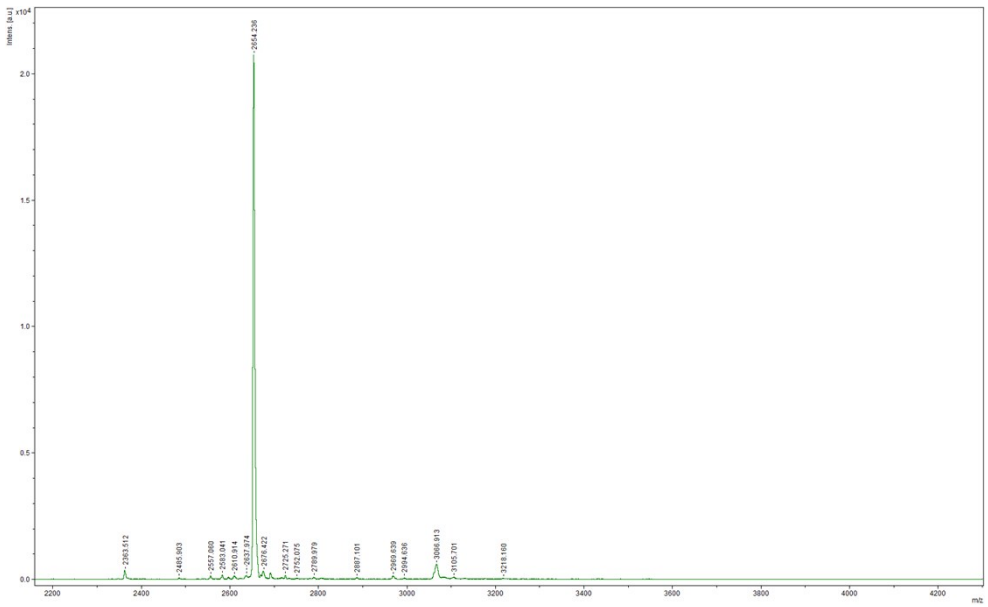




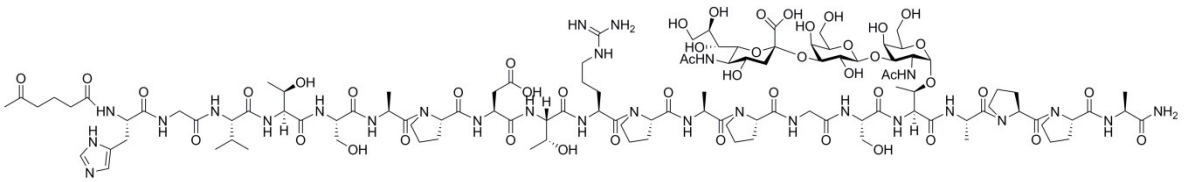
Compound **10**. 1.1 mg (0.49 μmole, 57% yield calculated from the starting compound **3**, 0.84 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-25% MALDI-TOFMS: $C_{111}H_{176}N_{28}O_{47}$, Theoretical mass: 2653.224, Observed mass: 2654.236 $[M+H]^+$

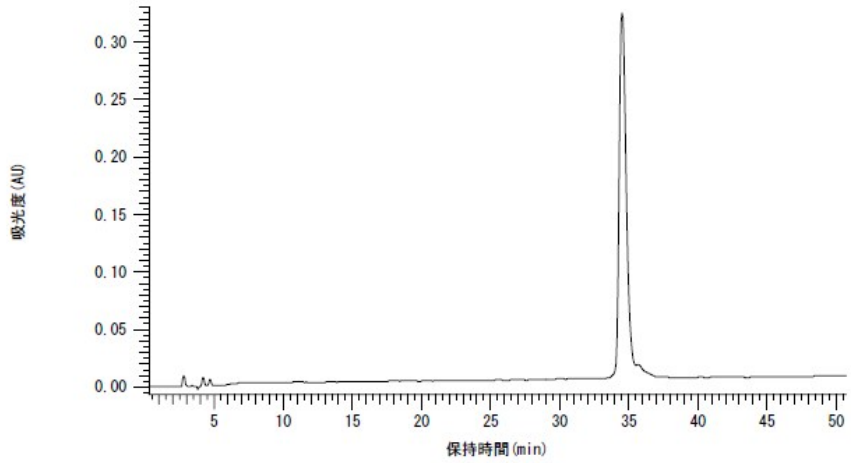


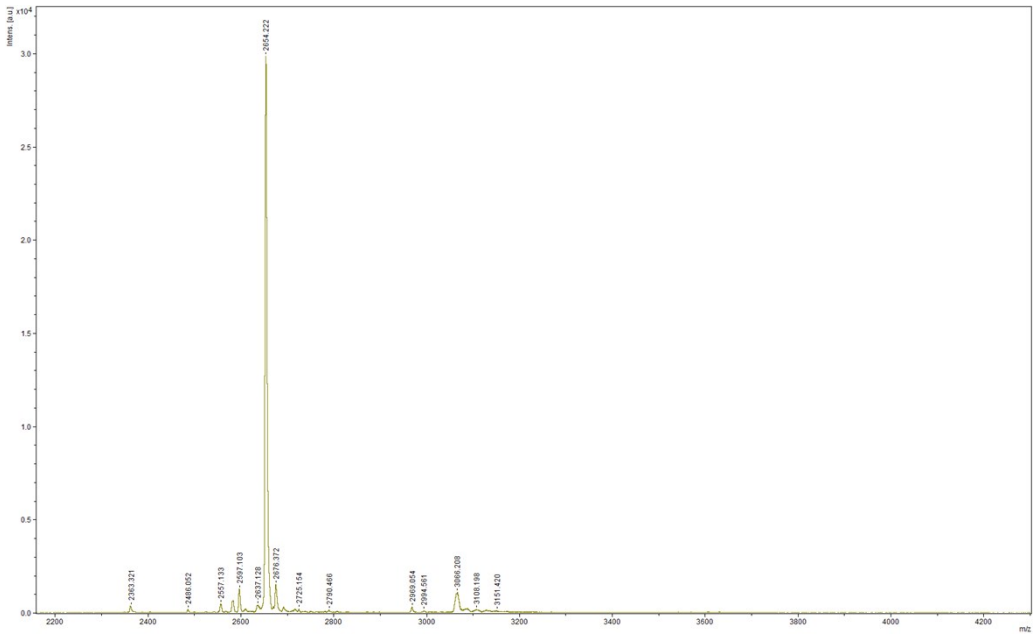




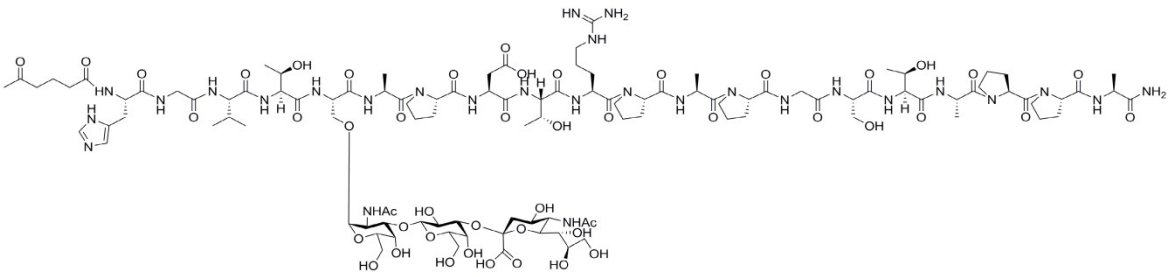
Compound **11**. 1.09 mg (0.41 μmole, 49% yield calculated from the starting compound **4**, 0.85 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-25% MALDI-TOFMS: $C_{111}H_{176}N_{28}O_{47}$, Theoretical mass: 2653.224, Observed mass: 2654.222 $[M+H]^+$

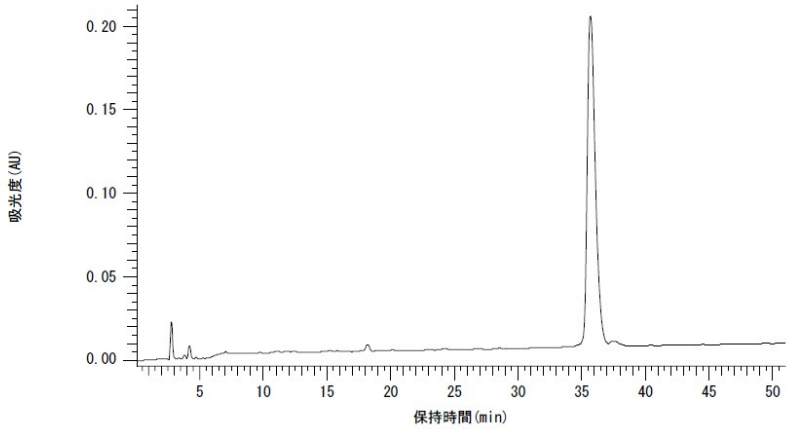


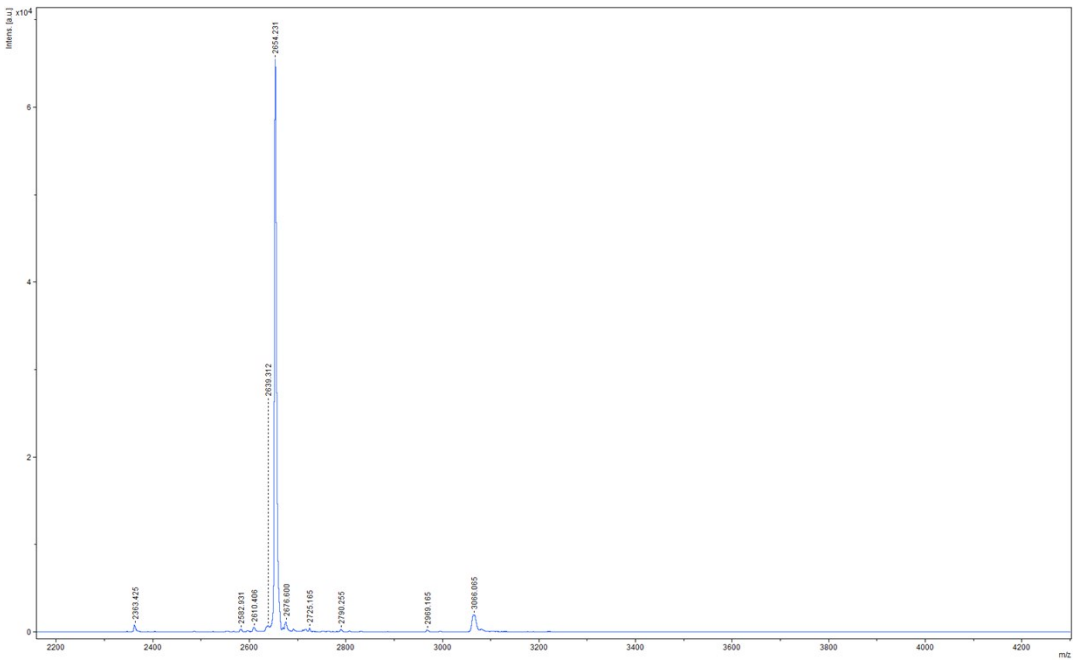




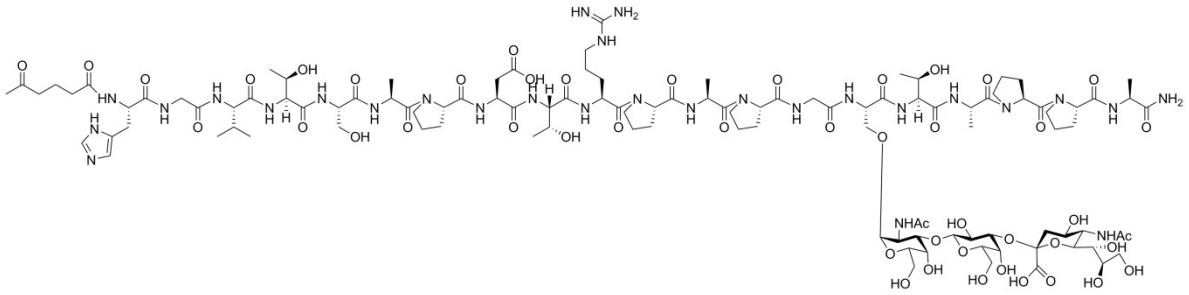
Compound **12**. 1.06 mg (0.41 μmole, 49% yield calculated from the starting compound **6**, 0.85 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-25% MALDI-TOFMS: $C_{111}H_{176}N_{28}O_{47}$, Theoretical mass: 2653.224, Observed mass: 2654.231 $[M+H]^+$

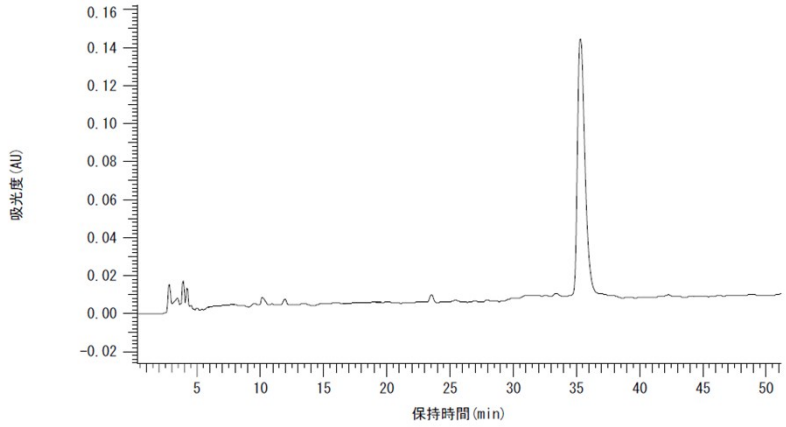


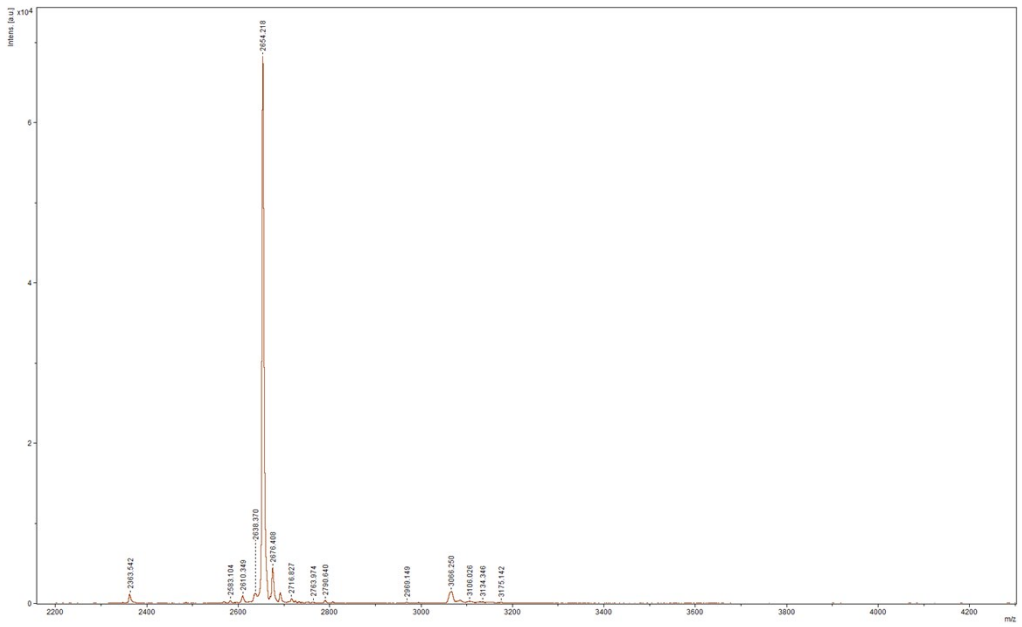




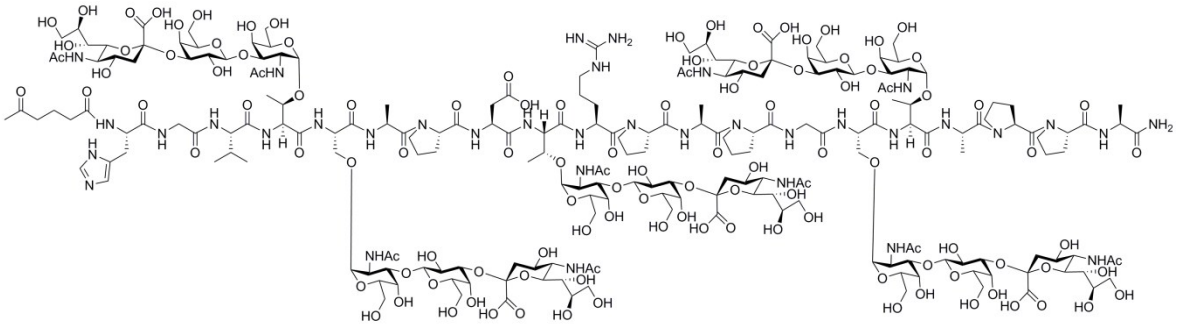
Compound **13**. 1.45 mg (0.55 μmole, 65% yield calculated from the starting compound **7**, 0.85 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-25% MALDI-TOFMS: $C_{111}H_{176}N_{28}O_{47}$, Theoretical: 2653.224, Observed mass: 2654.218 $[M+H]^+$

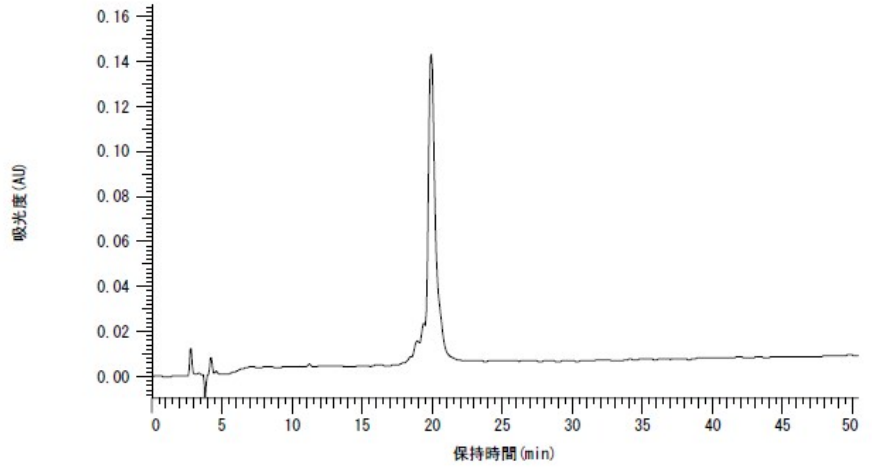


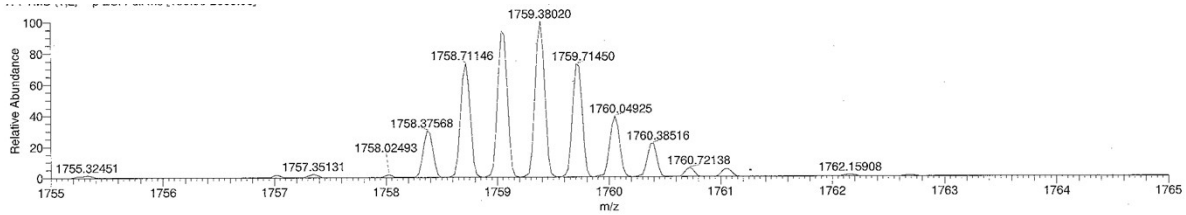




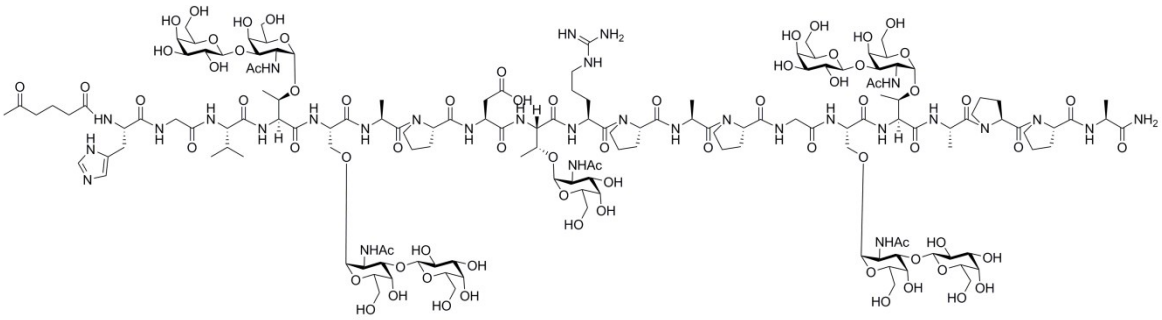
Compound **14**. 0.61 mg (0.115 μmole, 22% yield calculated from the starting compound **8**, 0.522 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-25% ESI-HRMS: $C_{211}H_{336}N_{36}O_{119}$, Theoretical mass: 5278.1347, *m/z* 1759.3782, Observed mass: *m/z* 1758.3756 $[M-3H]^{3-}$

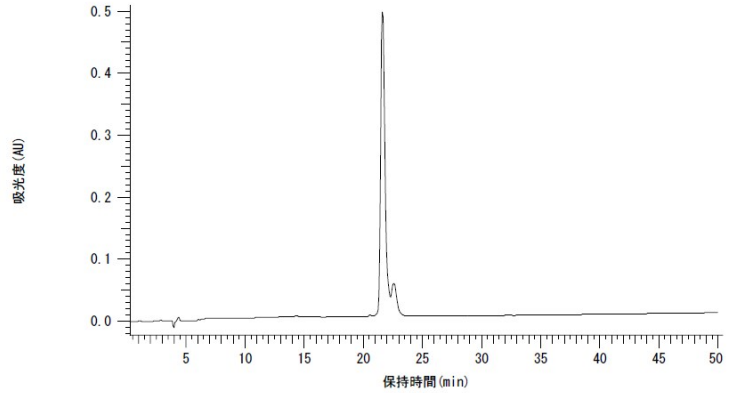


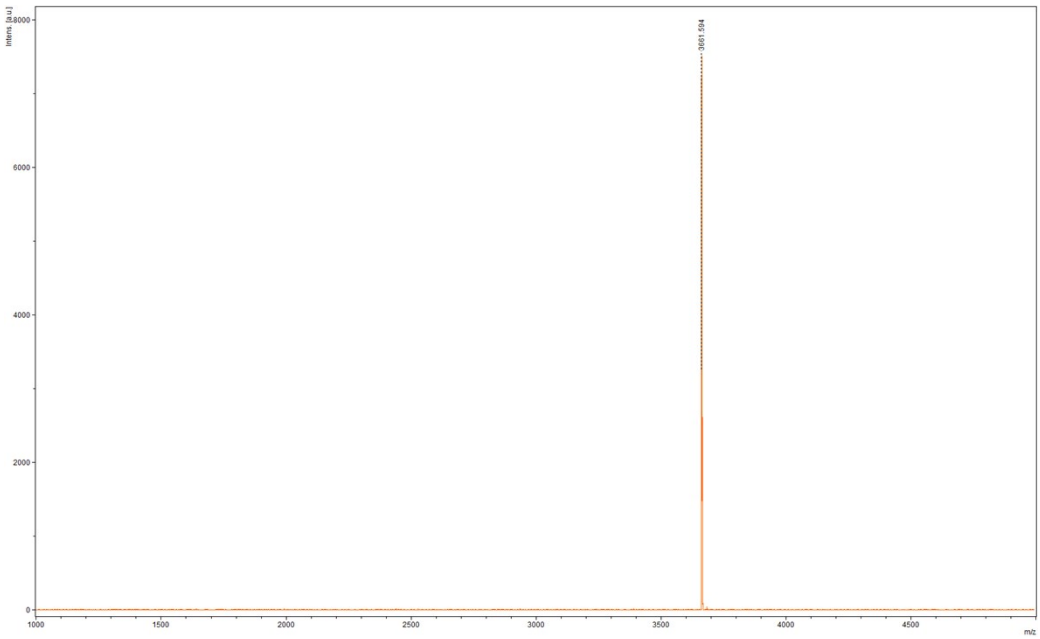




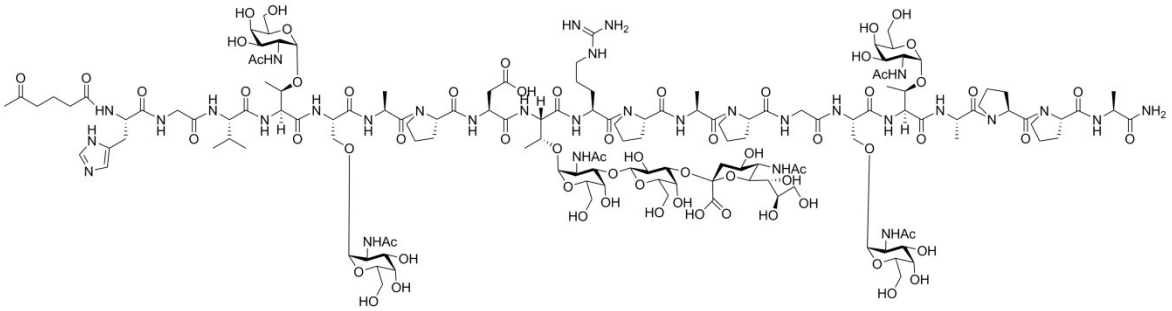
Compound **15**. 6.8 mg (1.9 μmole, 7.7% yield calculated from the starting resin, 50 mg, 24 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-30%
MALDI-TOFMS: $C_{150}H_{241}N_{31}O_{74}$, Theoretical mass: 3660.605, Observed mass: 3660.594 $[M+H]^+$

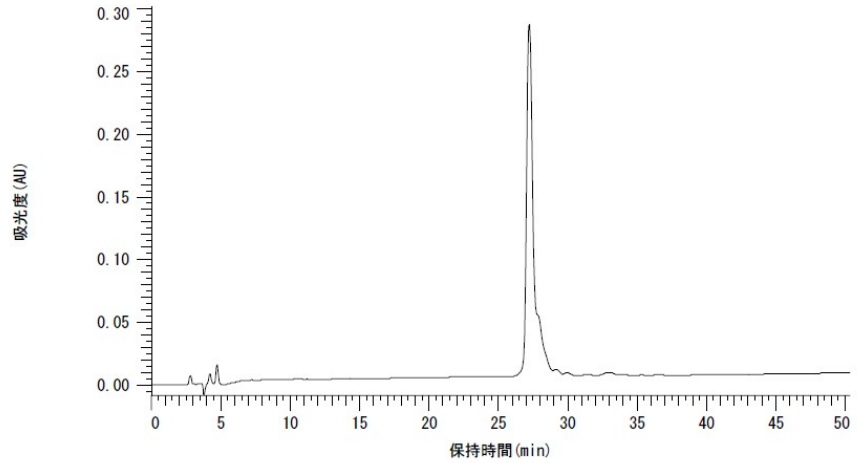


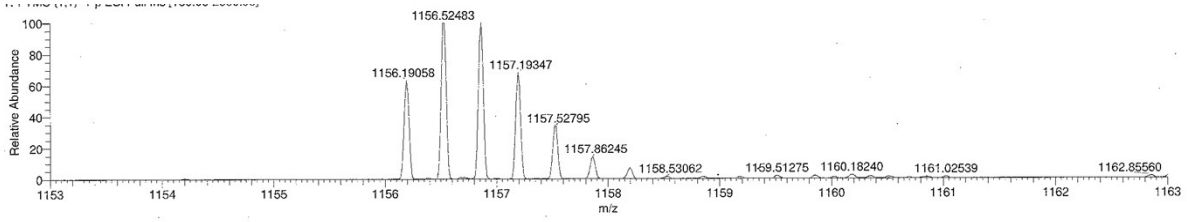




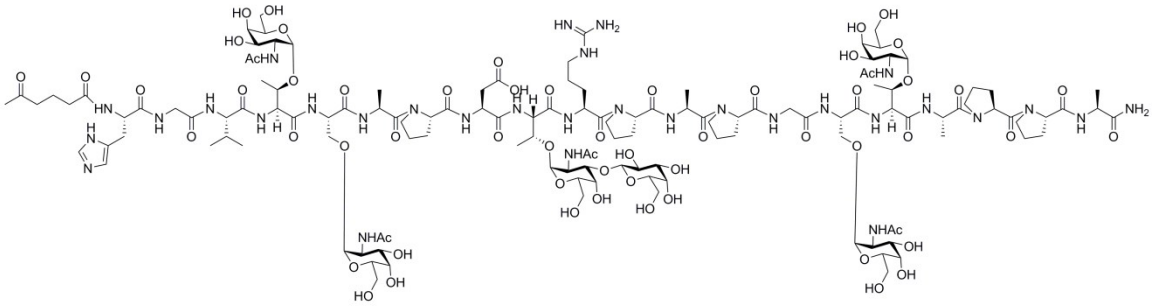
Compound **16**. 0.5 mg (0.144 µmole, 23% yield calculated from the starting compound **17**, 0.63 µmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-25% ESI-HRMS: $C_{143}H_{228}N_{32}O_{67}$, Theoretical mass: 3465.5417, *m/z* 1155.1805, Observed mass: *m/z* 1156.1905 $[M+3H]^{3+}$

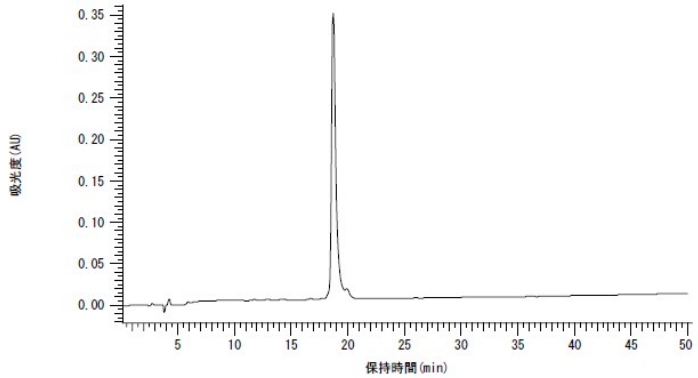


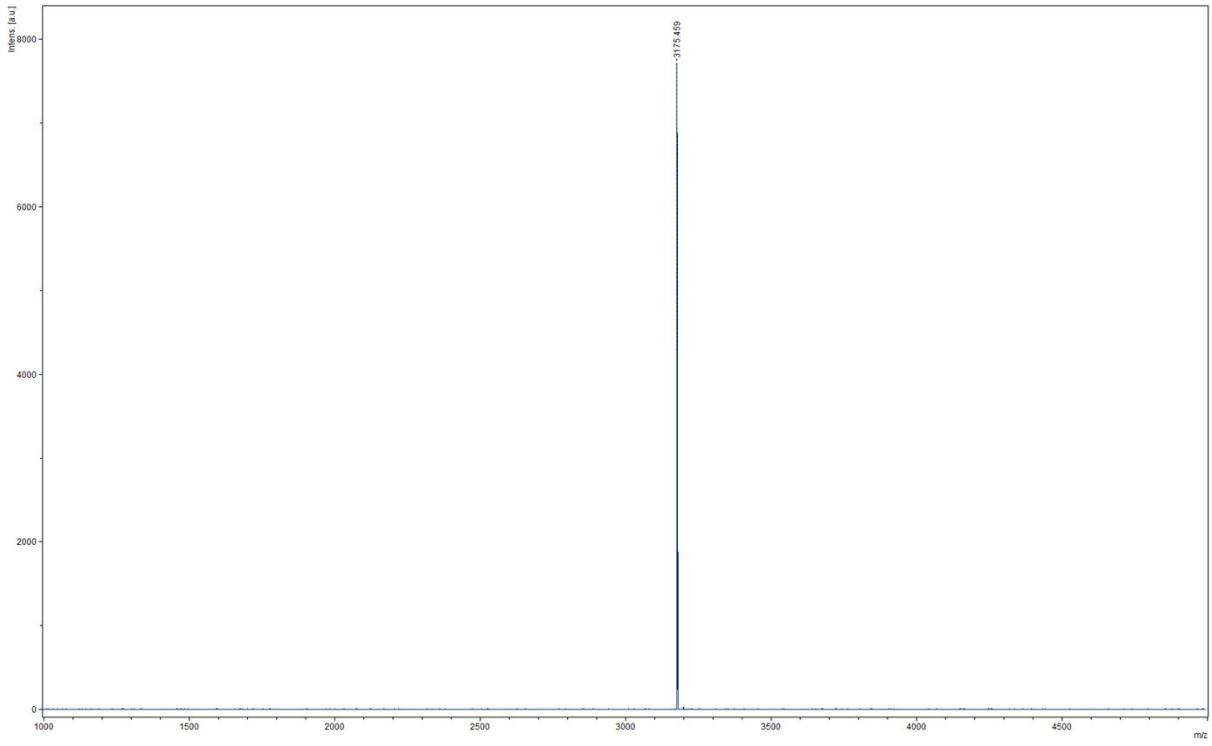




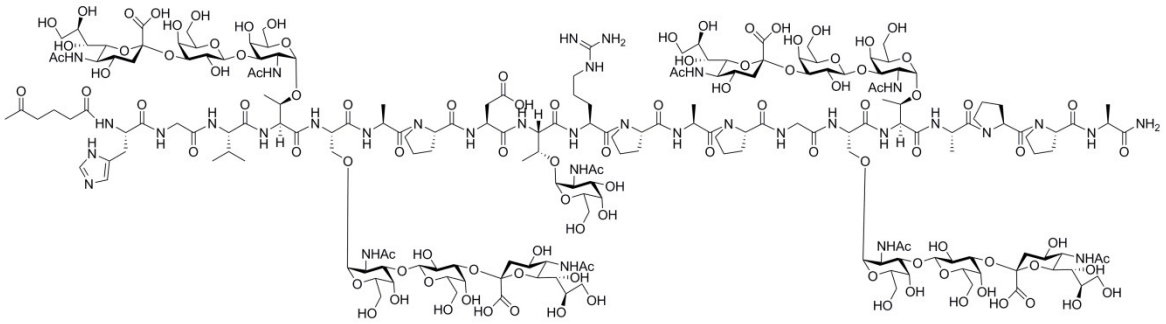
Compound **17**. 7.2 mg (2.3 μmole, 9.4% yield calculated from the starting resin, 50 mg, 24 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-30% MALDI-TOFMS: $C_{132}H_{211}N_{31}O_{59}$, Theoretical mass: 3174.446, Observed mass: 3175.459 $[M+H]^+$

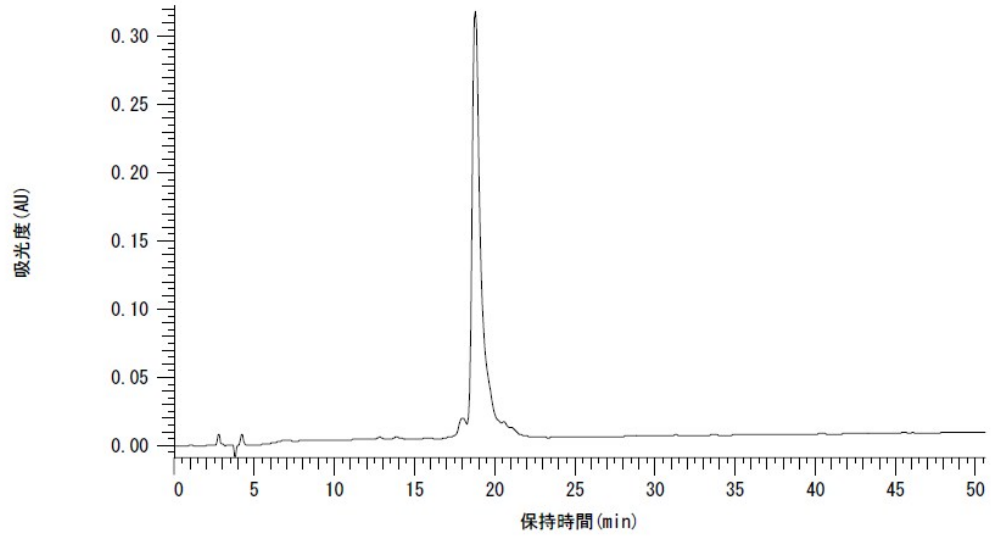


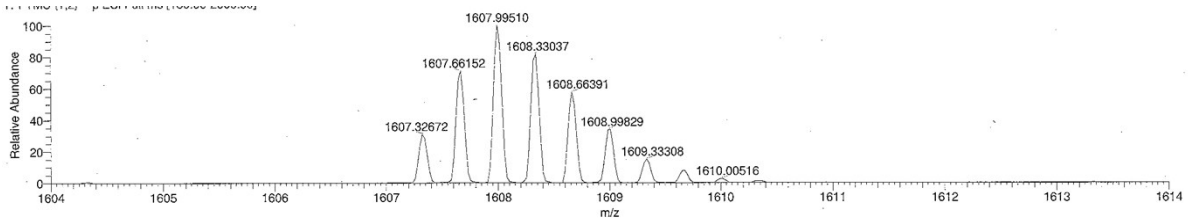




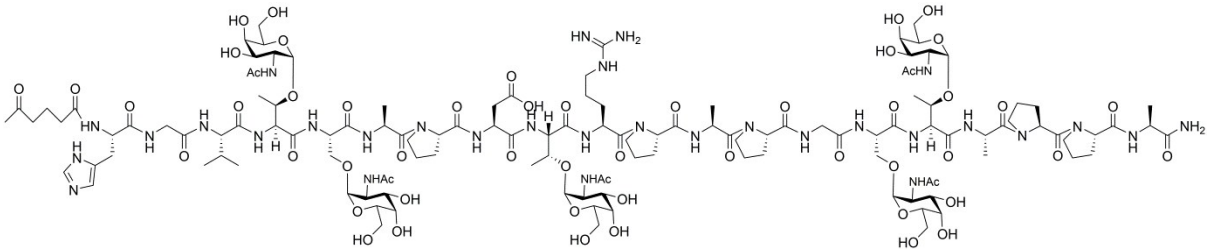
Compound **18**. 0.66 mg (0.136 μmole, 25% yield calculated from the starting compound **15**, 0.55 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-25% ESI-HRMS: $C_{194}H_{309}N_{35}O_{106}$, Theoretical mass: 4824.9864, *m/z* 1608.3288, Observed mass: *m/z* 1607.3267 $[M-H]^{3-}$

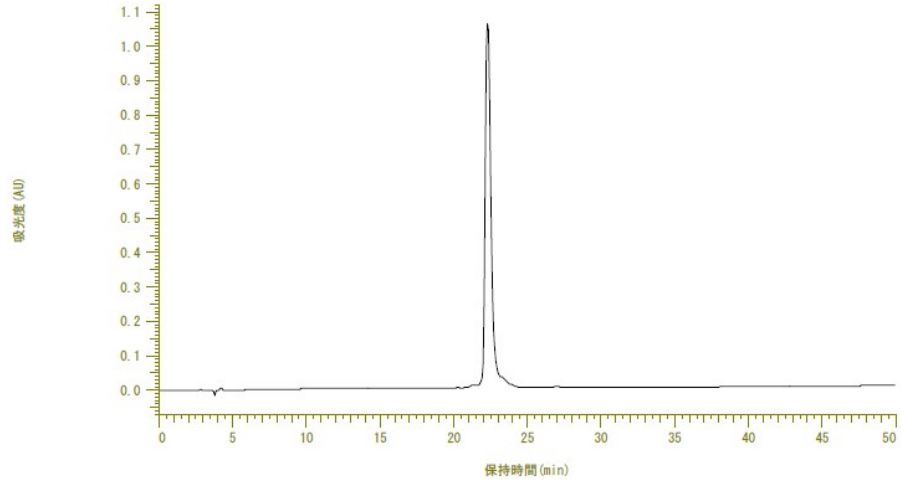


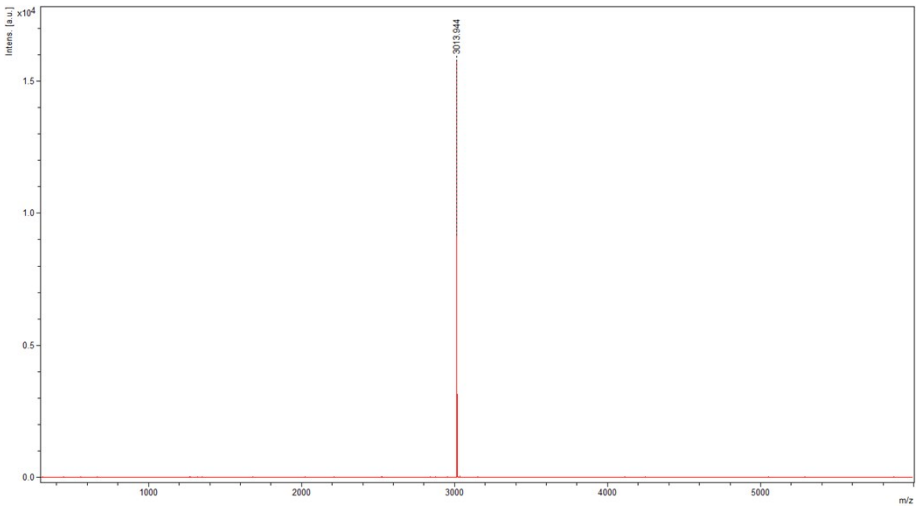




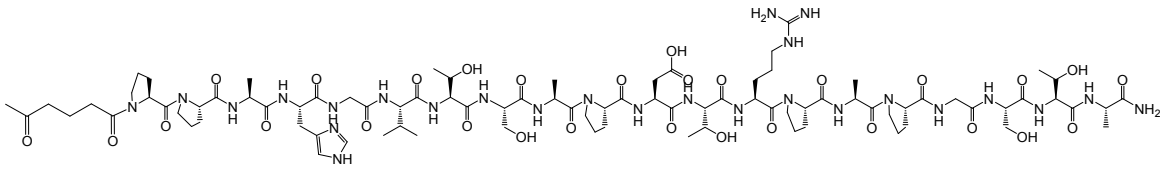
Compound **19**. 8.8 mg (2.9 μmole, 12% yield calculated from the starting resin, 50 mg, 24 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-30% MALDI-TOFMS: $C_{131}H_{211}N_{31}O_{49}$, Theoretical mass: 3012.3935: Observed mass: 3013.944 $[M+H]^+$

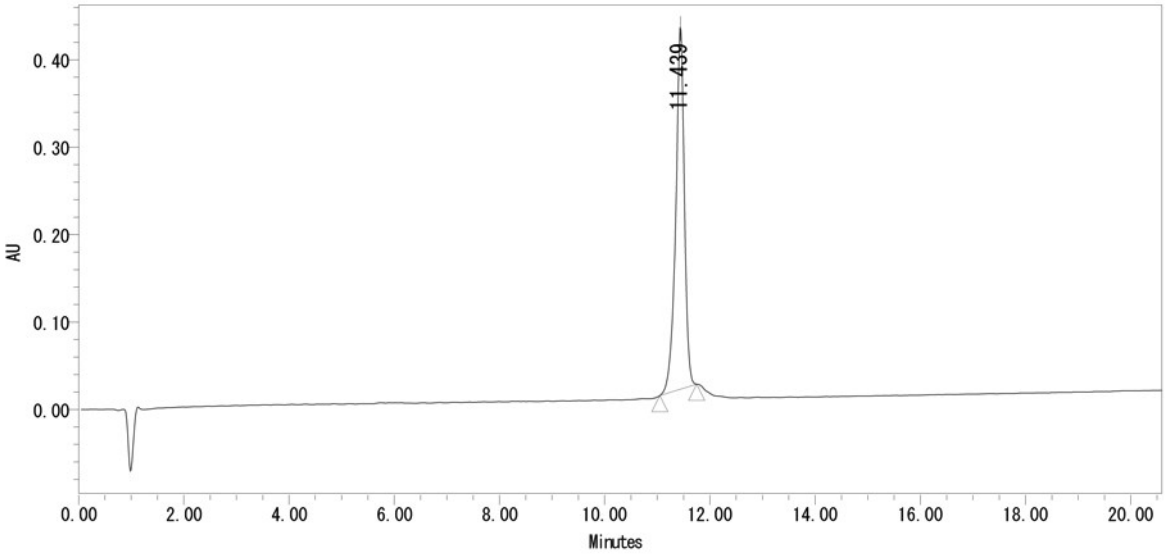


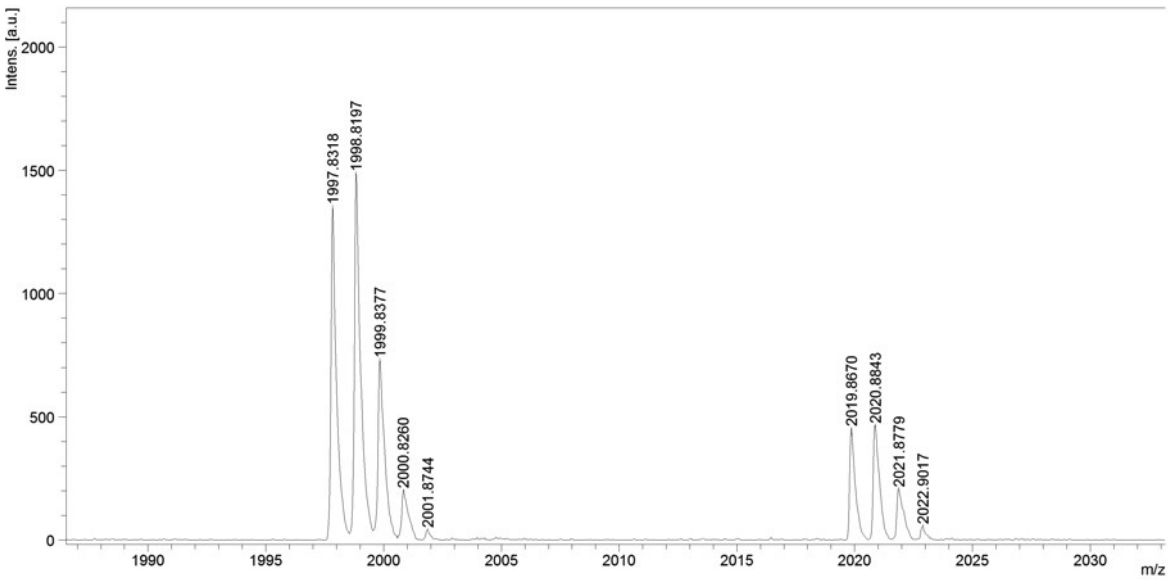




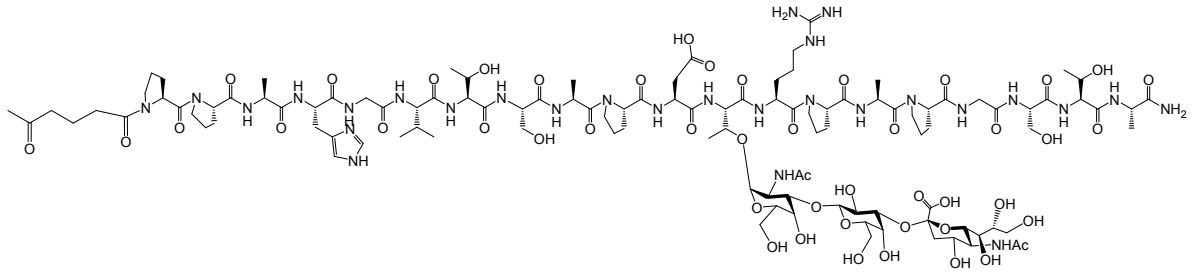
Compound **20**. 4.5 mg (15.3% based on resin capacity). Analytical UPLC (gradient condition A): t_R = 11.439 min, peak area ratio 100%. MALDI-TOFMS: $C_{86}H_{137}N_{26}O_{29}$ $[M+H]^+$ calcd (*m/z*) 1996.9967, found (*m/z*) 1997.8318..

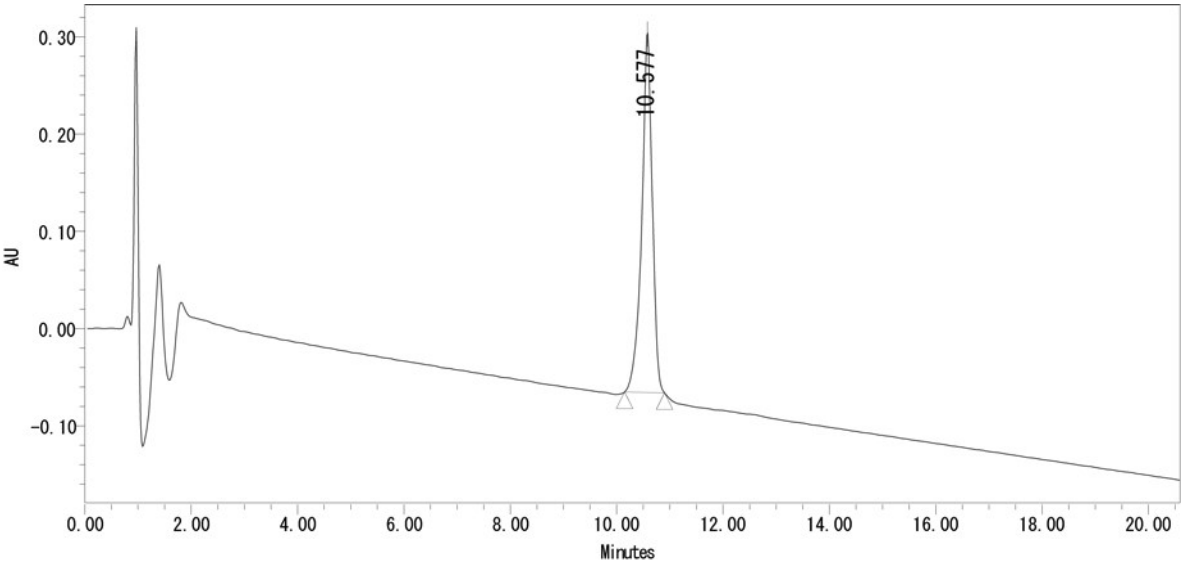


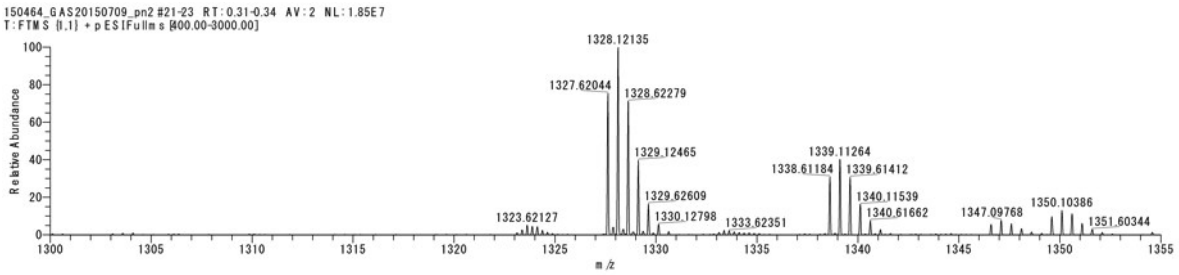




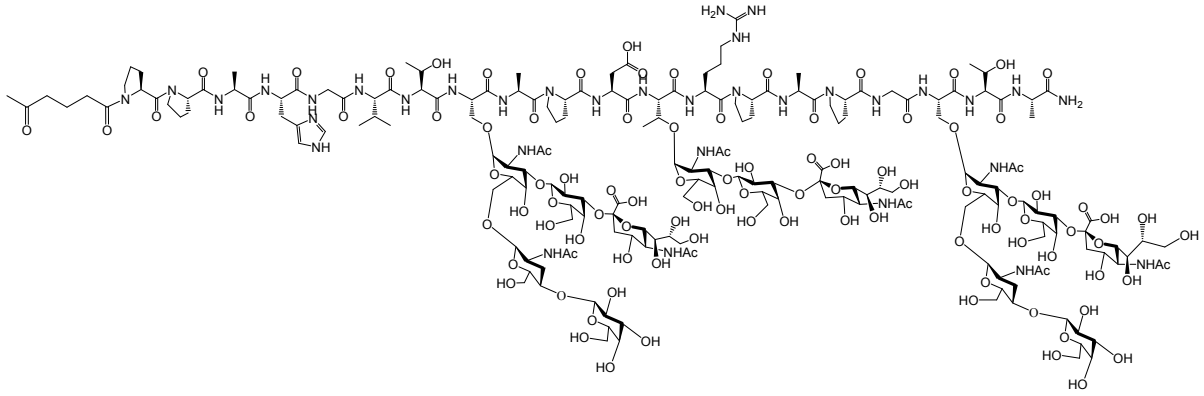
Compound **21**. 3.0 mg (7.7% isolated overall yield calculated from the resin capacity). Analytical UPLC (gradient condition B): t_R = 10.577 min, peak area ratio 100%. ESI-HRMS: $C_{111}H_{178}N_{28}O_{47}$ $[M-2H]^-$ calcd (*m/z*) 1327.6194, found (*m/z*) 1327.6204.

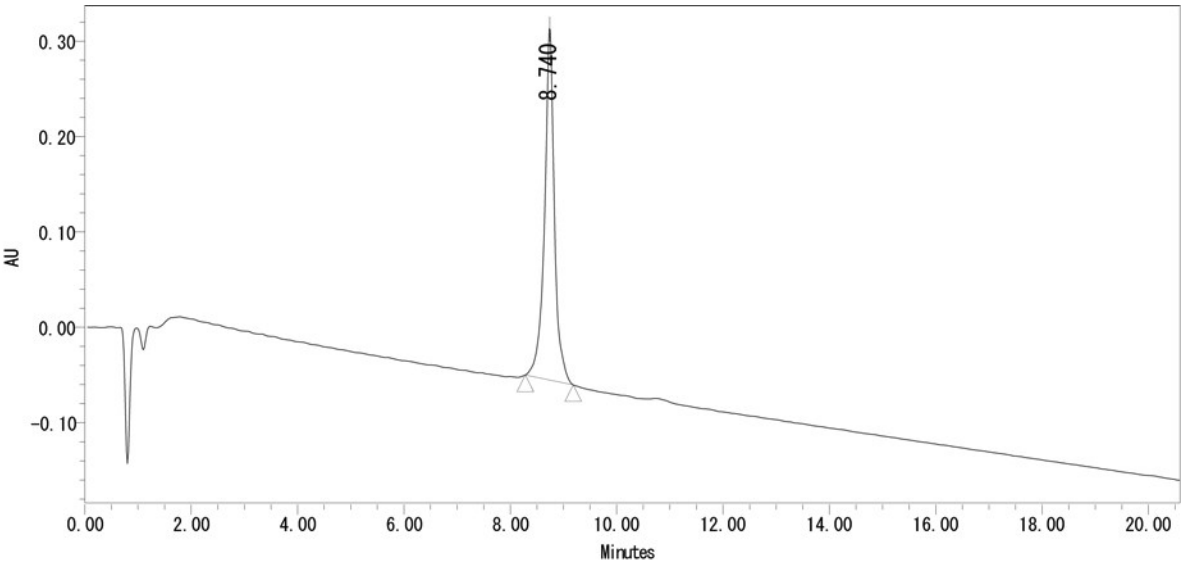


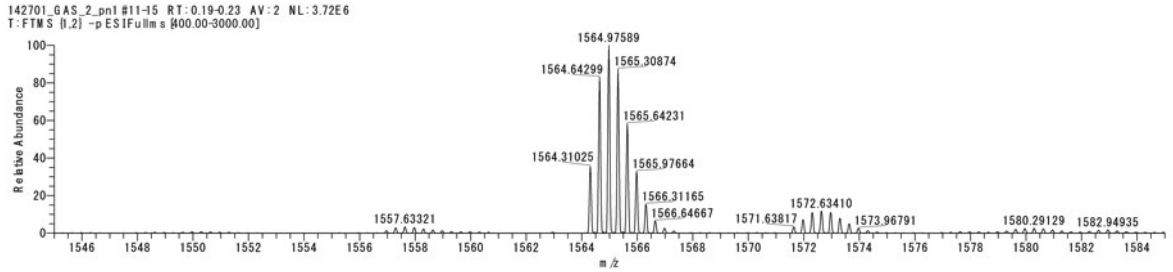




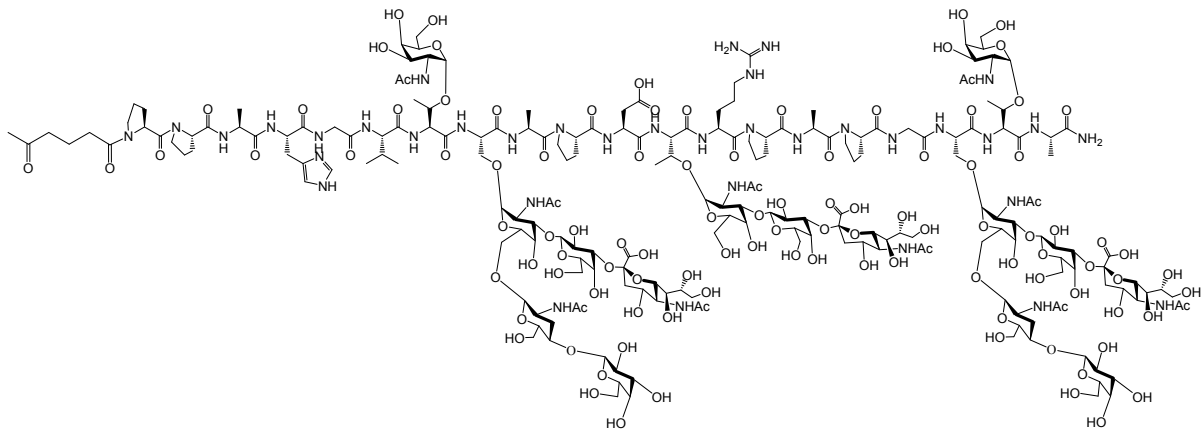
Compound **22**. 1.3 mg (4.9% overall yield calculated from the starting resin capacity, 5.5 μmole). Analytical UPLC (gradient condition B): t_R = 8.740 min, peak area ratio 100%. ESI-HRMS: $C_{189}H_{299}N_{34}O_{103}$ $[M-3H]^-$ calcd (*m/z*) 1564.3074, found (*m/z*) 1564.3103

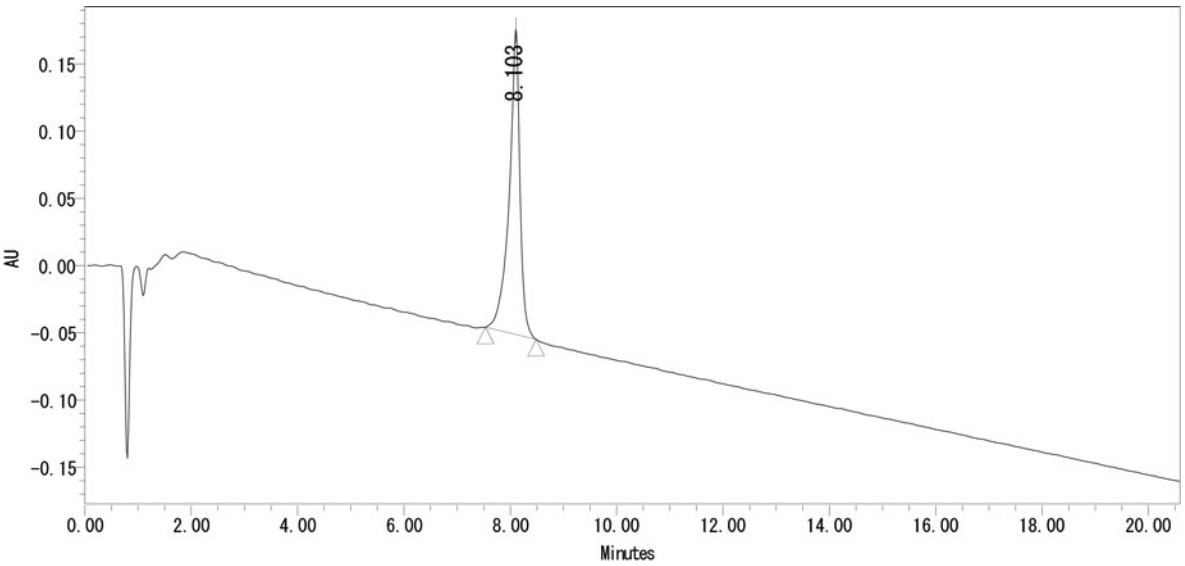


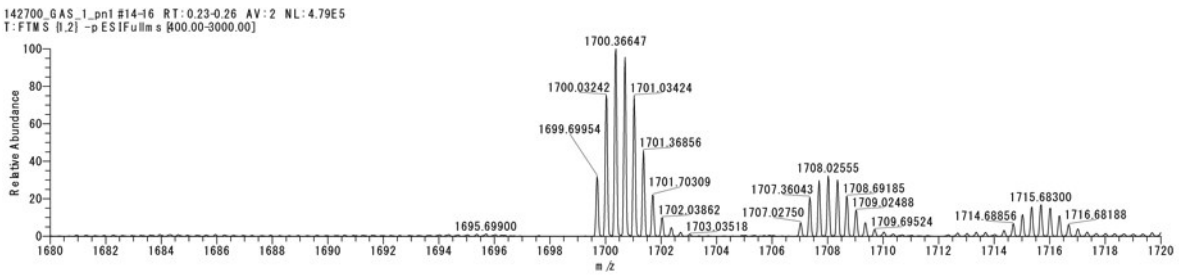




Compound **23**. 0.81 mg (2.9% overall yield calculated from the starting resin capacity, 0.77 μmole). Analytical UPLC (gradient condition B): t_R = 8.103 min, peak area ratio 100%. ESI-HRMS: $C_{205}H_{325}N_{36}O_{113}$ $[M-3H]^-$ calcd (*m/z*) 1699.6936, found (*m/z*) 1699.6995.

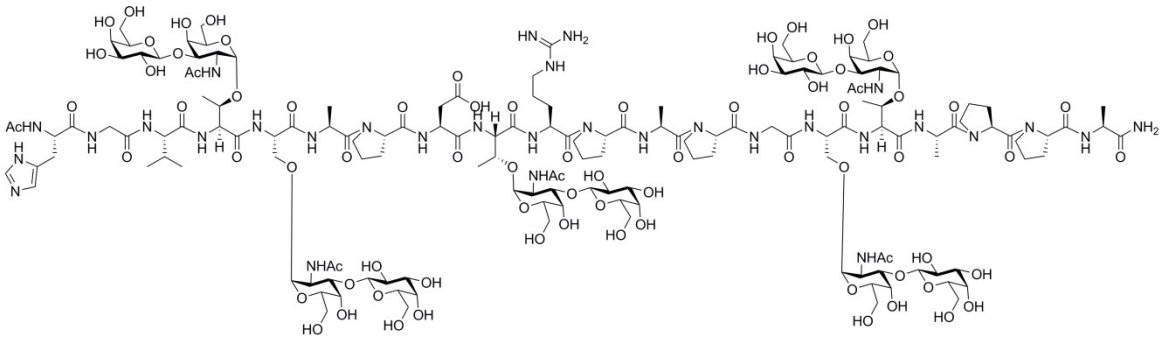


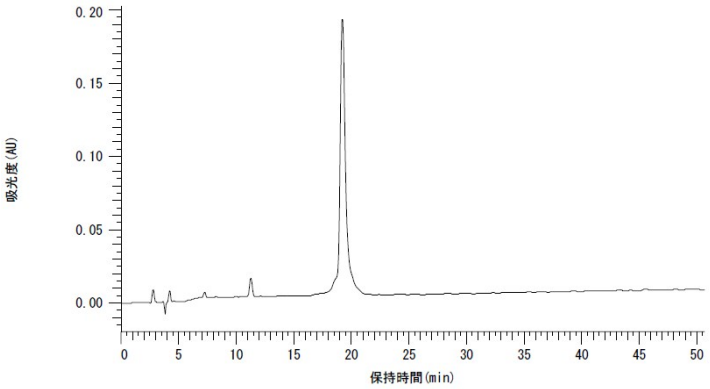


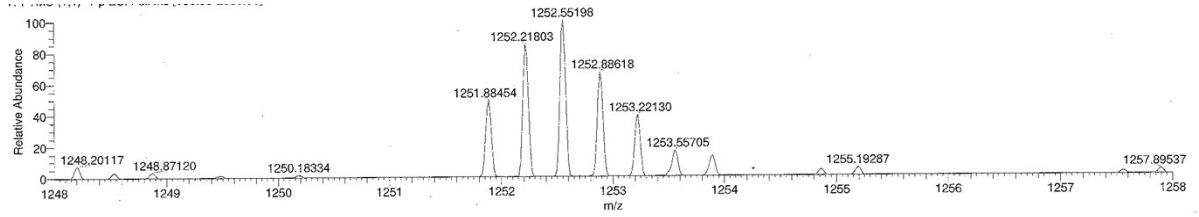


MUC1 glycopeptides for NMR study (*N*-terminus is acetylated)

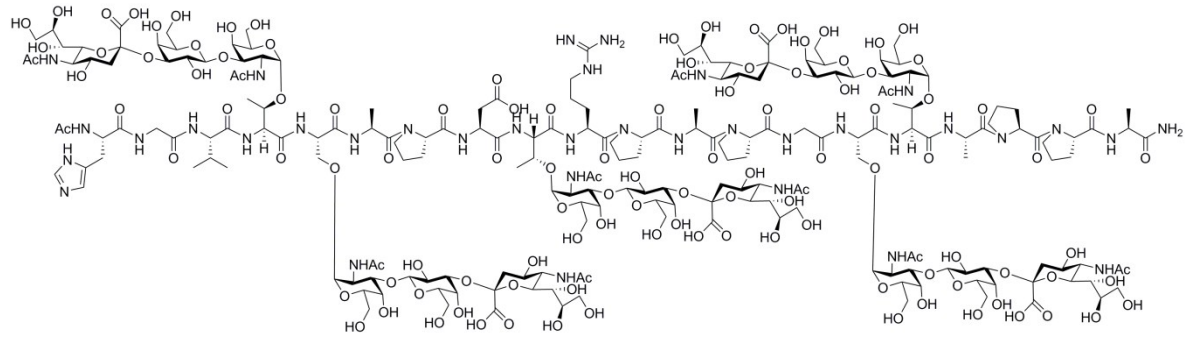
Compound **8**. 8.4 mg (2.23 μmole, 9.3% yield calculated from the starting resin, 50 mg, 24 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-30% ESI-HRMS: $C_{152}H_{245}N_{31}O_{78}$, Theoretical mass: 3752.6157, *m/z* 1250.8719, Observed mass: *m/z* 1251.8845 $[M+3H]^{3+}$

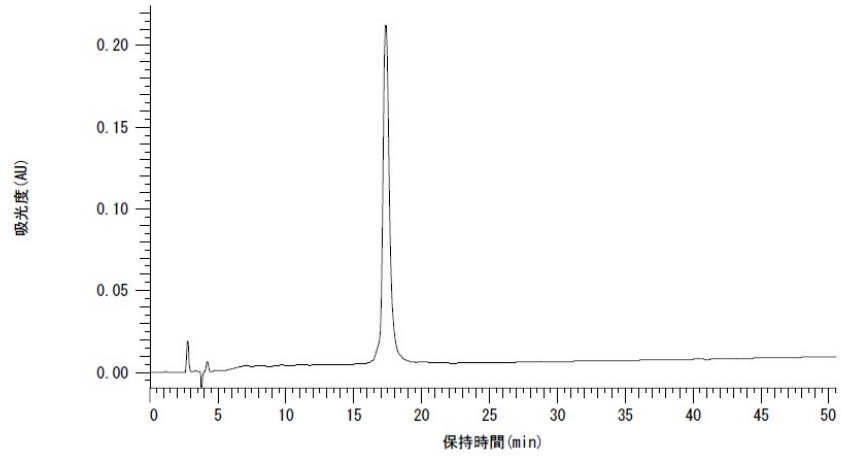


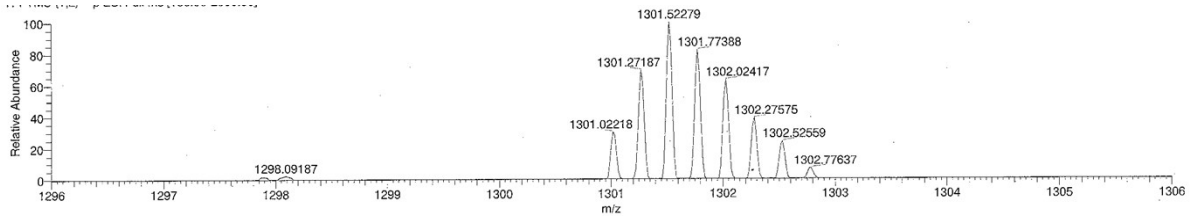




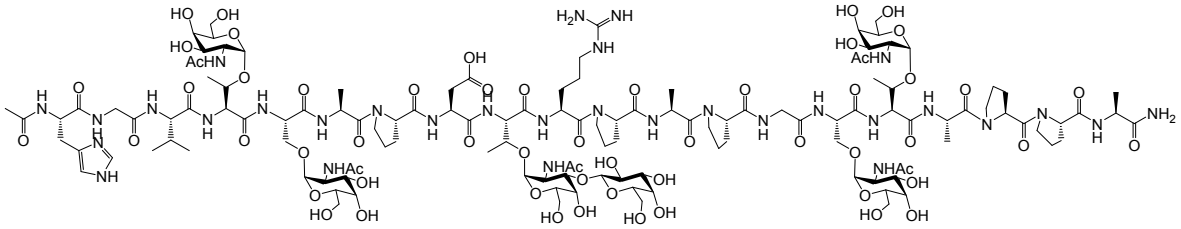
Compound **14**. 3.4 mg (0.65 μmole, 15.3% yield calculated from the compound **8**, 16.0 mg, 0.43 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-25% ESI-HRMS: $C_{207}H_{330}N_{36}O_{118}$, Theoretical mass: 5208.0928, *m/z* 1302.232, Observed mass: *m/z* 1301.0221 $[M-4H]^{4-}$

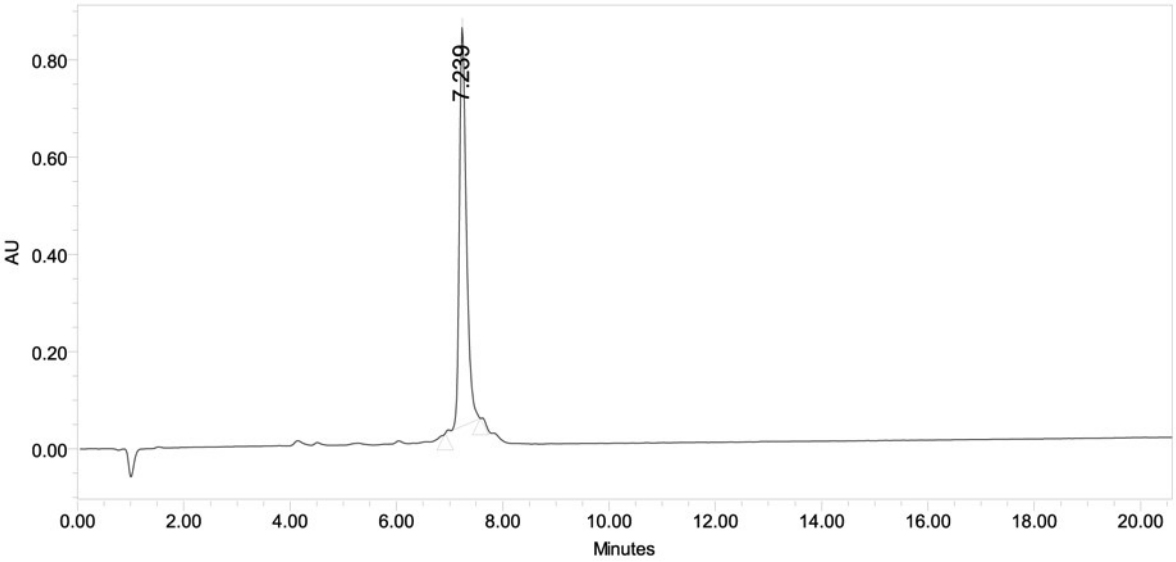


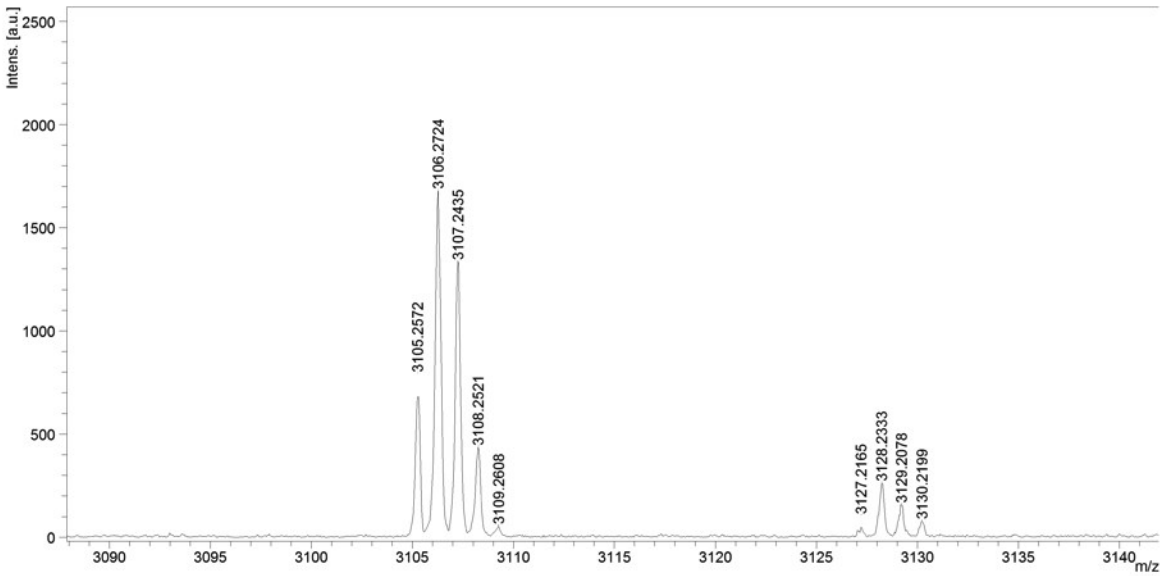




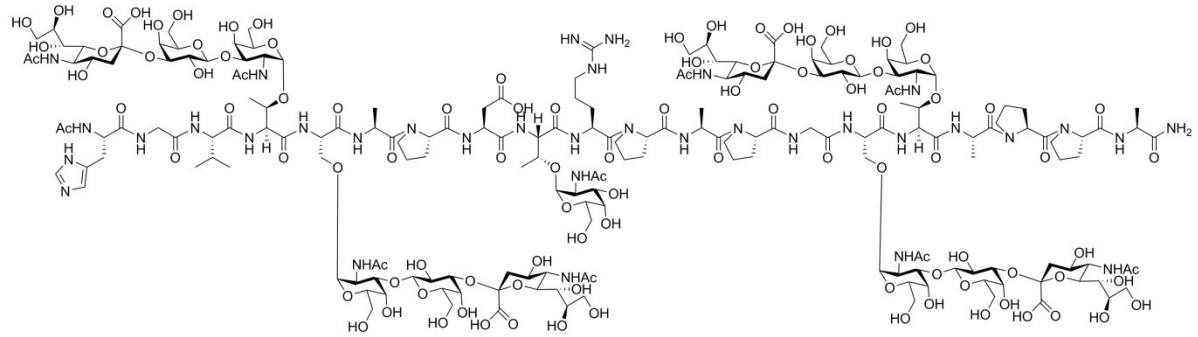
Compound 17. 9.2 mg (3.0 μmole, 12.3% yield calculated from the starting resin, 50 mg, 24 μmole). Analytical UPLC (gradient condition B): t_R = 7.239 min, peak area ratio 100%. MALDI-TOFMS: $C_{128}H_{242}N_{31}O_{76}$ $[M+H]^+$ 3105.2572 calcd (*m/z*), found (*m/z*) 3106.2724.

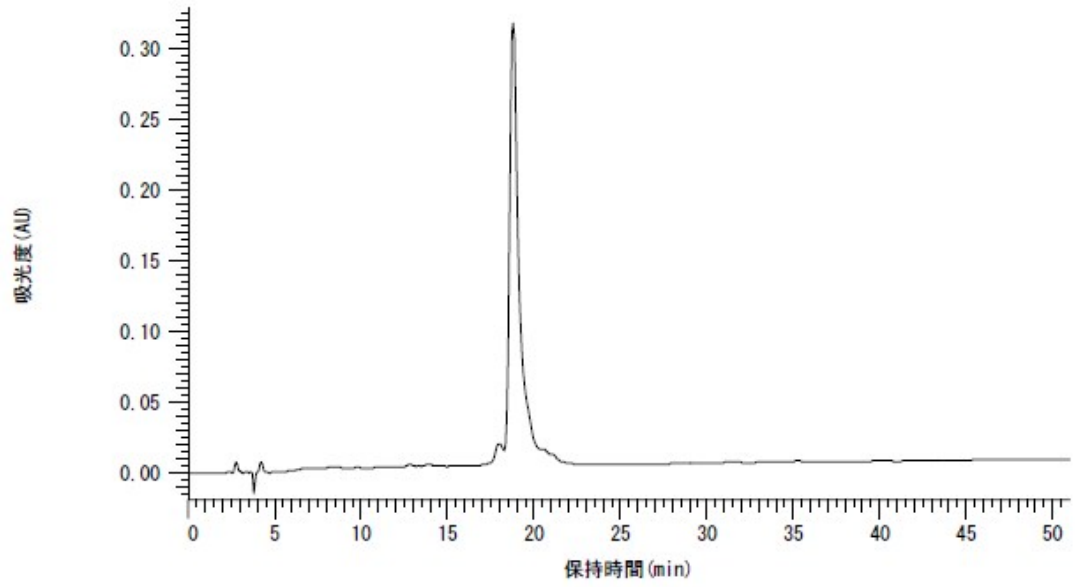


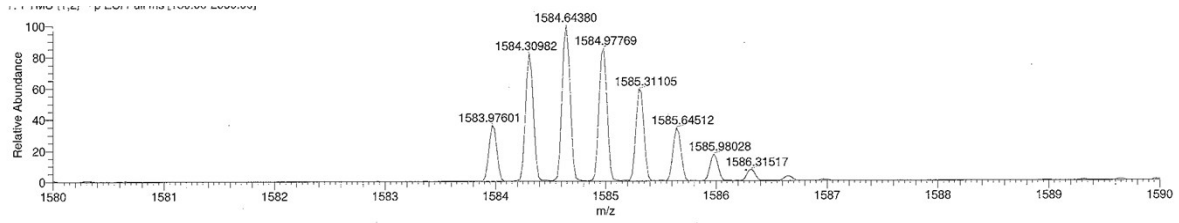




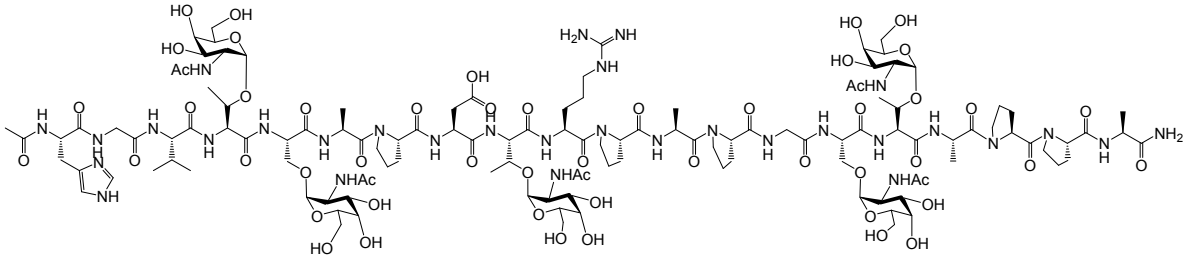
Compound **18**. 3.2 mg (0.67 μmole, 22% yield calculated from the compound **15**, 11.0 mg, 0.31 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-25% ESI-HRMS: $C_{190}H_{303}N_{35}O_{105}$, Theoretical mass: 4754.9446, *m/z* 1584.9815, Observed mass: *m/z* 1583.9760 $[M-3H]^{3-}$

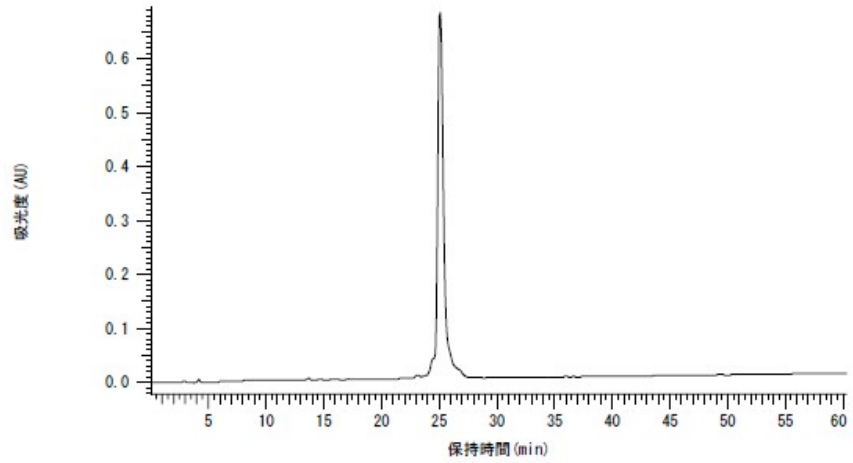


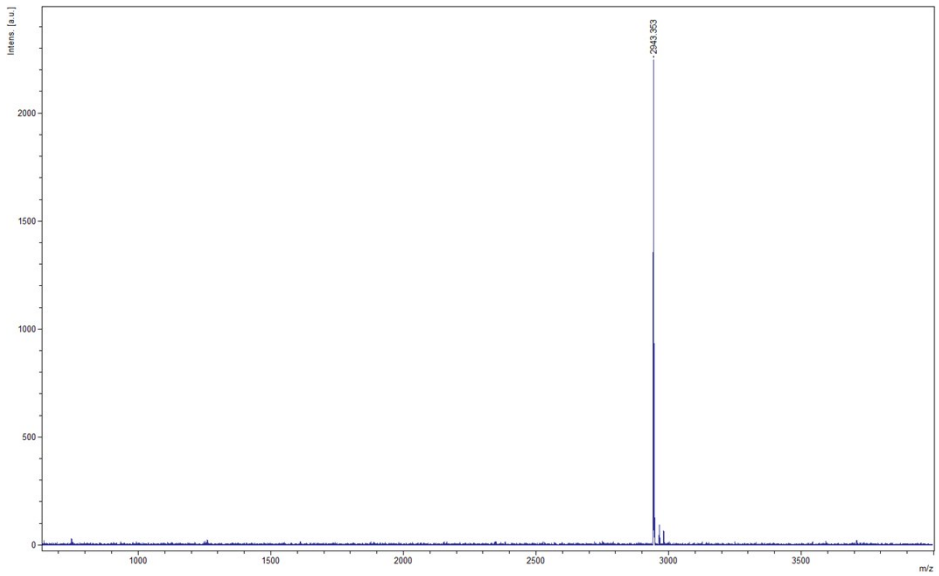




Compound **19**. 8.3 mg (2.8 μmole, 11.7% yield calculated from the starting resin, 50 mg, 24 μmole). Analytical HPLC: A : B (H_2O in 0.1%TFA : CH_3CN in 0.1%TFA) A/B gradient 0-50 min, 5-25%. MALDI-TOFMS: $C_{127}H_{205}N_{31}O_{48}$ $[M+H]^+$ 2943.358 calcd (m/z), found (m/z) 2943.353





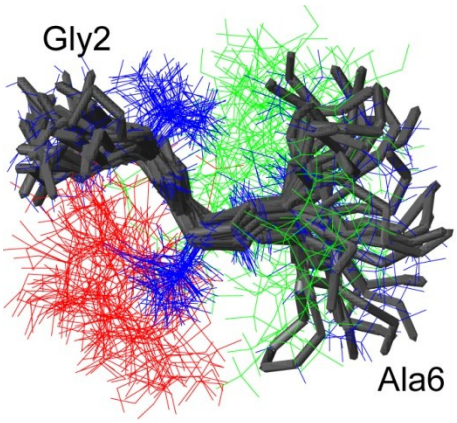


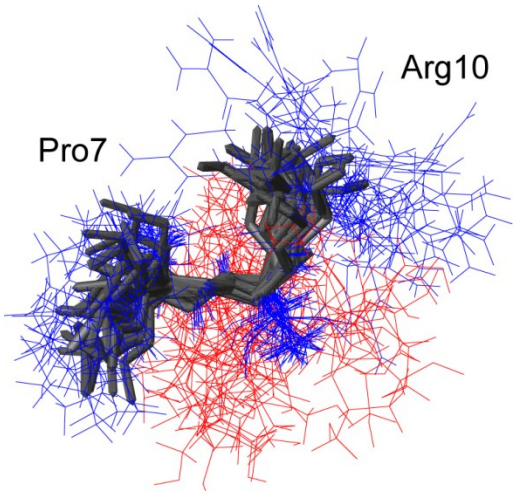
Supplementary NMR structures.

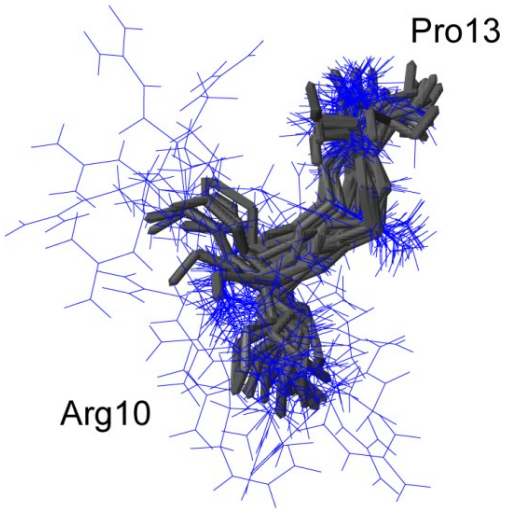
All NMR experiments by means of compounds **8**, **14**, **17**, **18**, **19** (20mer MUC1 glycopeptides), and 24mer MUC1 glycopeptide used for the docking experiment with SM3 antibody (Figure 7 in the main text) were performed on a Bruker AVANCE 800 MHz spectrometer at 298K under the conditions described in the experimental section in the main text. Three-dimensional structures of MUC1 glycopeptides were calculated and generated from the distance restraints and dihedral angle restraints according to the method reported previously (see, for example; Tachibana, Y., Fletcher, G. L., Fujitani, N., Tsuda, S., Monde, K., Nishimura, S. -I. Antifreeze glycoproteins: Elucidation of the structural motifs that are essential for antifreeze activity. *Angew. Chem. Int. Ed. Engl*. **2004**, *43*, 856-862. Matsushita, T., Ohyabu, N., Fujitani, N., Naruchi, K., Shimizu, H., Hinou, H., and Nishimura, S. I. Site-specific conformational alteration induced by sialylation of MUC1 tandem repeating glycopeptides at an epitope region for the anti-KL-6 monoclonal antibody. *Biochemistry* **2013**, *52*, 402–414. Hayakawa, S., Koide, R., Hinou, H., Nishimura, S. -I. Synthetic human NOTCH1 EGF modules unraveled molecular mechanisms for the structural and functional roles of calcium ions and *O*-glycans in the ligand-binding region. *Biochemistry* **2016**, *55*, 776-787).

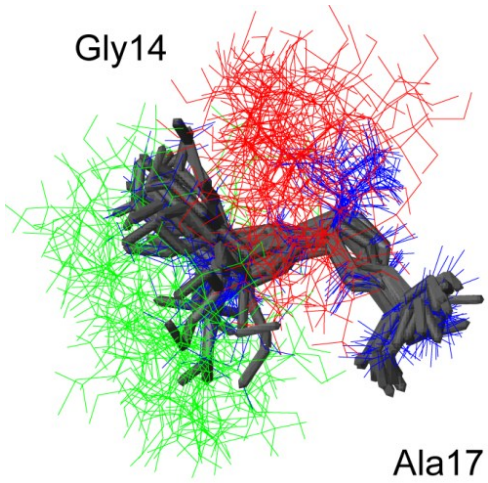
The following NMR stuructures were the 30 best calculated structures at four important segments including PDTR region of the 20mer MUC1 fragments **8**, **14**, **17**, **18**, **19** and 24mer MUC1 fragment bearing Tn antigen at PDTR motif. In the main text, only the conformation of the PDTR region was used as shown in Figure 5 and Figure 7). The GVTSA, PDTR, RPAP, GSTA regions are superimposed on backbone atoms VTS, DT (or DTR), PAP, STA, respectively. The main chain of the glycopeptide backbone is coloured in grey, the side chain is blue, and the GalNAc attached to Thr residue is red, and the GalNAc attached to Ser is green.

Compound **8**: TTTTT

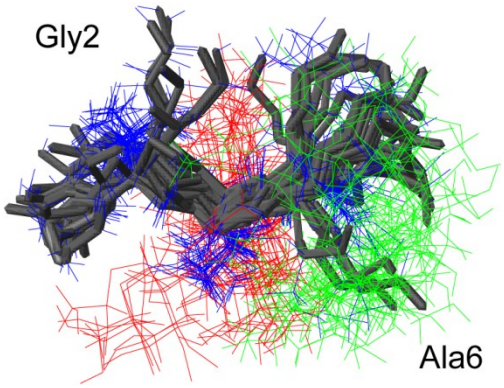


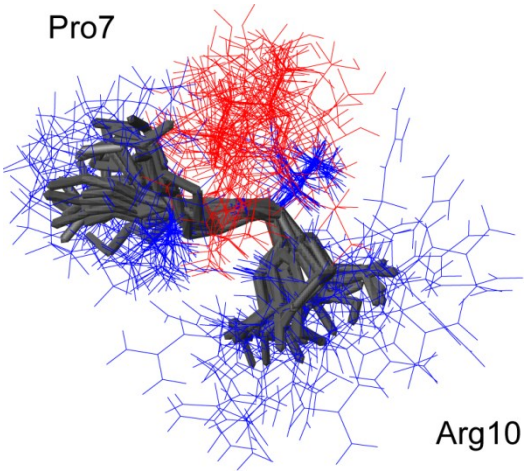


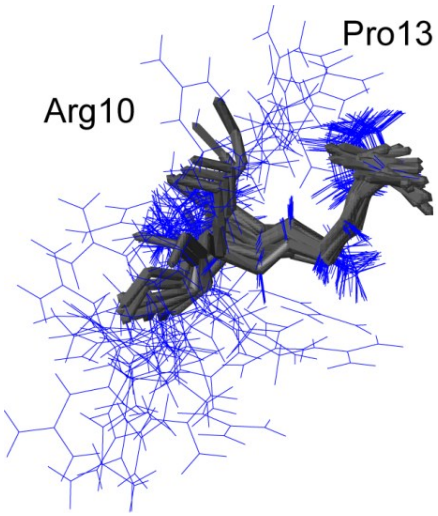


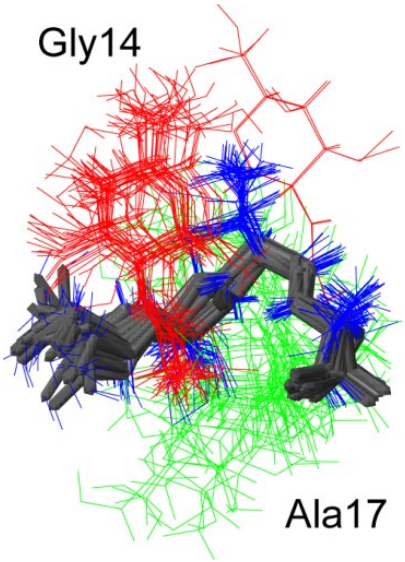


Compound **14**: STSTSTSTST

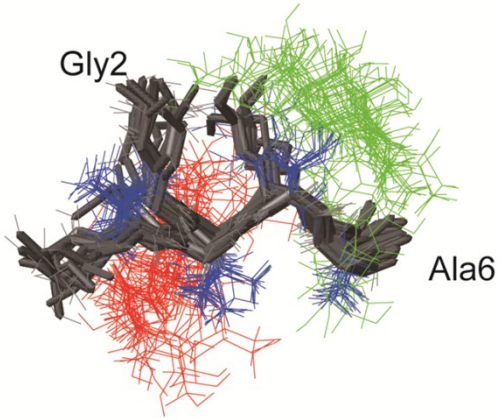


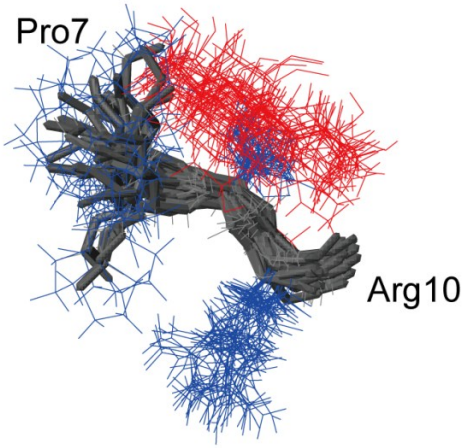


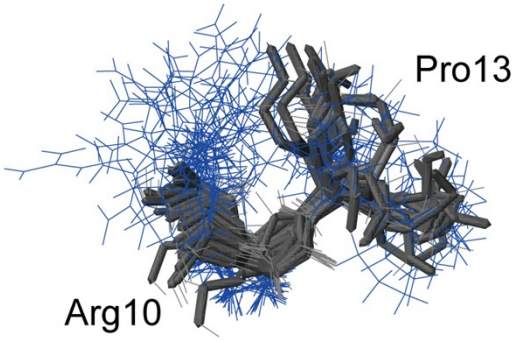


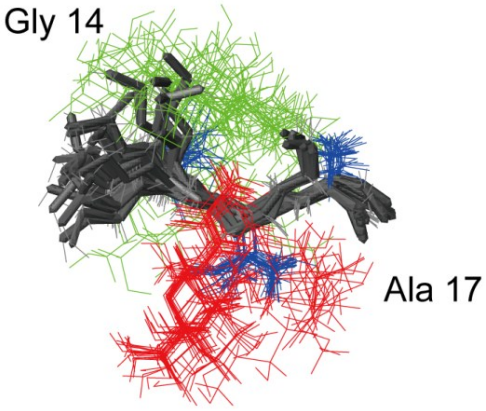


Compound**17**: TnTnTTnTn

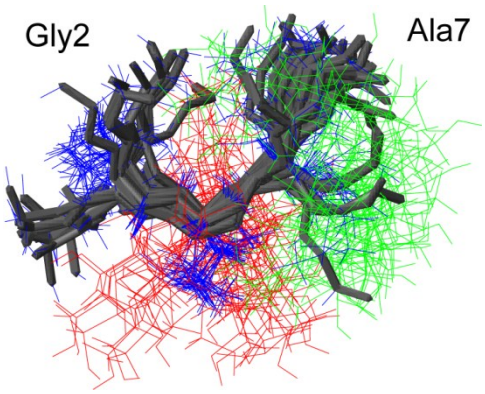


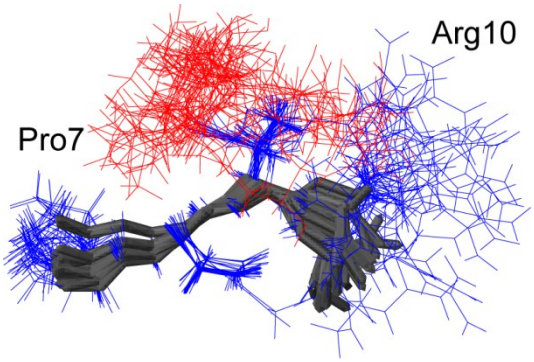


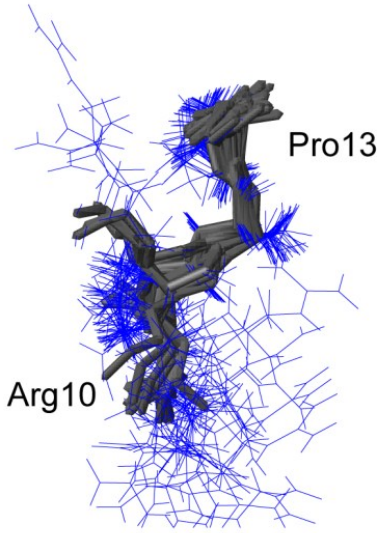


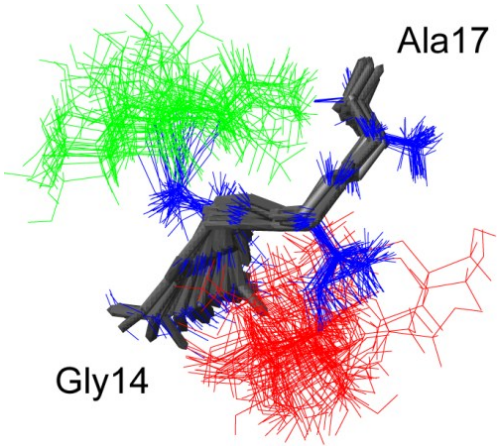


Compound **18**: STSTTnSTST

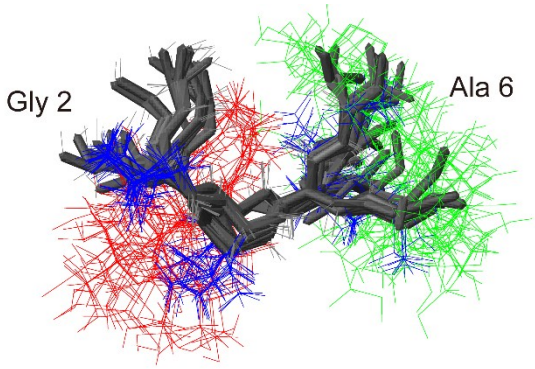


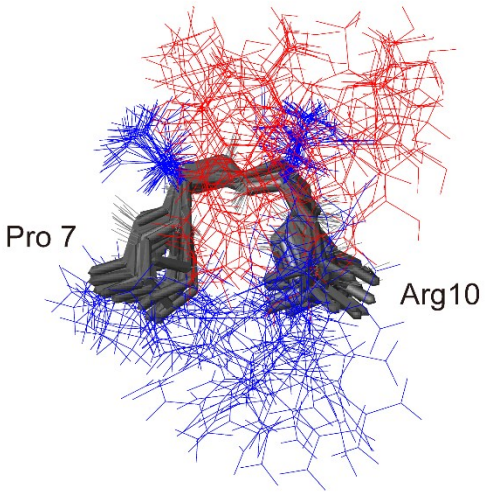


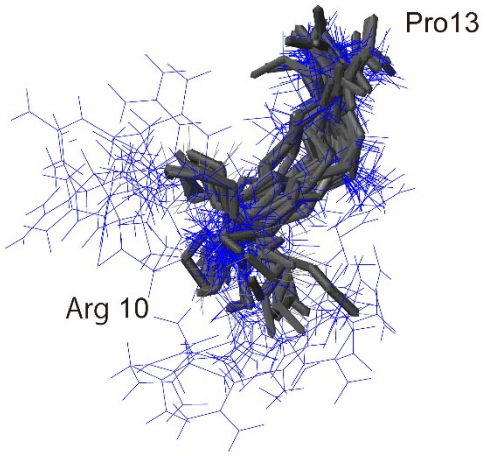


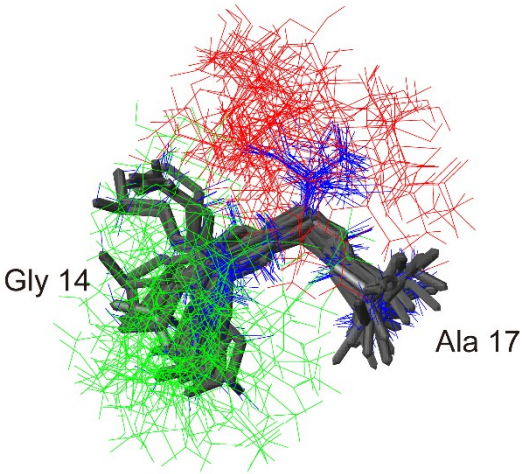


Compound **19**: TnTnTnTnTn









24mer MUC1 glycopeptide having a Tn antigen at Thr residue involved in the PDTR motif.