ELECTRONIC SUPPLEMENTARY INFORMATION

Development of synthetic, self-adjuvanting, and self-assembling anticancer vaccines based on a minimal saponin adjuvant and the tumor-associated MUC1 antigen

Carlo Pifferi,[a] Leire Aguinagalde,[a] Ane Ruiz-de-Angulo,[a] Priscila Tonon Baschirotto,[a] Nagore Sacristán,[a] Ana Poveda,[c] Jesús Jiménez-Barbero,[cdef] Juan Anguita,*[bd] and Alberto Fernández-Tejada*[ad]

[a]Chemical Immunology Laboratory, [b]Inflammation and Macrophage Plasticity Laboratory, [c]Chemical Glycobiology Laboratory, Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Biscay Technology Park, Building 801A, 48160 Derio, Spain.

[d]Ikerbasque, Basque Foundation for Science, Maria Diaz de Haro 13, 48009 Bilbao, Spain.

[e]Department of Organic Chemistry II, Faculty of Science & Technology, University of the Basque Country, 48940 Leioa, Spain. [f]Centro de Investigación Biomédica En Red de Enfermedades Respiratorias, Madrid, Spain

E-mail: janguita@cicbiogune.es
E-mail: afernandeztejada@cicbiogune.es

TABLE OF CONTENTS

I. GENERAL INFORMATION

All commercially available materials were used without further purification. All manipulations with air-sensitive reagents and chemical reactions were carried out under a dry argon atmosphere using standard Schlenk techniques. Air- and moisture-sensitive liquids and solutions were transferred via syringe. When anhydrous conditions were required, the appropriate reagents were dried via azeotropic removal of water with dry toluene. Molecular sieves were activated at 350 °C and were crushed immediately prior to use, then dried under vacuum. Organic solutions were concentrated under reduced pressure by rotary evaporation below 40 °C. Column chromatography was performed employing 230–400 mesh silica gel. Thin-layer chromatography (TLC) was performed using aluminum-backed sheets pre-coated with 230–400 mesh silica gel 60 containing fluorescent indicator (F254). TLC plates were visualized under UV light (254 nm) and by staining with cerium ammonium molybdate (CAM), phosphomolybidic acid (PMA), or 5% sulfuric acid in ethanol solutions.

Nuclear magnetic resonance (NMR). ^{1}H and ^{13}C spectra for characterization purposes were recorded on a Bruker Avance III instrument (^{1}H NMR at 600 MHz and ^{13}C NMR at 151 MHz). Chemical shifts are expressed in parts per million (δ scale) downfield from tetramethylsilane and are referenced to residual proton in the NMR solvent ($CDCl_3$: δ 7.26 for ^{1}H NMR, δ 77.00 for ^{13}C NMR; methanol-d_4: δ 3.31 for ^{1}H NMR, δ 49.15 for ^{13}C NMR). Data are presented as follows: chemical shift, multiplicity (s = singlet, br s = broad singlet, d = doublet, t = triplet, q = quartet, m = multiplet and/or multiple resonances), coupling constant in Hertz (Hz), integration, assignment. The ^{1}H-NMR and DOSY experiments were recorded on a Bruker Avance III instrument (^{1}H NMR at 800 MHz) equipped with a cryogenic probe in aqueous conditions (PBS/D_2O; chemical shifts referenced to D_2O residual proton: δ 4.79).

RP-HPLC purification and LC-MS. All reverse-phase RP-HPLC analyses/purifications were carried out on a Waters 1525 binary gradient system (Solv. A = 0.05% TFA in H_2O; Solv. B = 0.05% TFA in CH_3CN) equipped with a Waters 2998 photodiode array detector (PDA), and combined with a low-resolution single quadrupole (SQD2, Waters Corporation) mass spectrometer. Absorbances were monitored at wavelengths of 190–400 nm.

HR-ESI-MS. High resolution electrospray ionization mass spectrometry (HR-ESI-MS) analyses of compounds **8**, **10**, **13**, **S3** and **S6** were performed on a Waters LCT Premier XE (Waters, Milford, MA, USA) in W-optics positive ionization scan mode. Mass spectrometry parameters were optimized to achieve the best signal-to-noise ratio: capillary voltage 1 kV, sample cone voltage 100 V, desolvation gas flow 600 Lh^{-1}, cone gas flow 50 Lh^{-1}, desolvation temperature 350 ^{0}C, source temperature 150 ^{0}C . The instrument was calibrated over the range m/z 200-2000 before measurement using a standard NaI solution (1 μM). In order to minimize the accuracy in the measurements, Leucine-Enkephalin was used as lockmass reference [2M+Na], m/z 1111,5459. Data analysis was performed with Masslynx software version 4.1 (Waters, Milford, MA, USA).

Characterization by MS was corroborated after comparing the experimental isotopic pattern with the theoretical one.

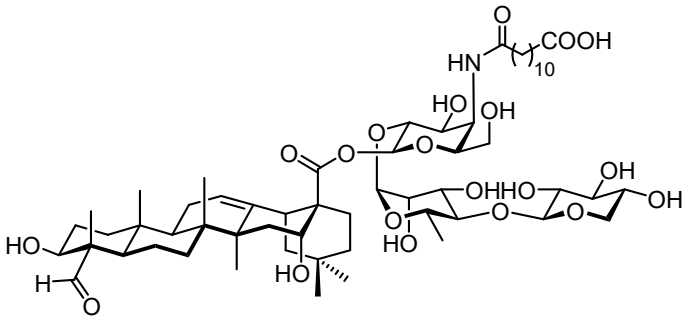
MALDI-TOF-HR-MS. High resolution MALDI-TOF mass spectra analyses of compounds **2**, **4–7**, **9**, **11**, **12**, **15** and **PV** peptide were performed on an UltrafleXtreme III MALDI-time-of-flight (TOF) mass spectrometer equipped with a pulsed Nd:YAG laser (355 nm) and controlled by FlexControl 3.3 software (Bruker Daltonics, Bremen, Germany). The acquisitions were carried out in positive reflector ion mode with pulse duration of 50 ns. Laser intensity was set marginally above the threshold of ionization to avoid fragmentation. The *m/z* range was chosen according to the mass of the sample. The acquired data was processed using the mMass software.

MALDI-TOF analysis of BSA conjugates: BSA–MUC1 and BSA–TnMUC1 conjugates were desalted using ZipTip® C4 micro-columns (Millipore) with elution using 0.5µL SA (sinapinic acid, 10 mg/ml in [70:30] acetonitrile:trifluoroacetic acid 0.1%) matrix onto a GroundSteel massive 384 target (Bruker Daltonics). Matrix-assisted laser desorption/ionization-time of flight high resolution mass spectrometry (MALDI-TOF-HR-MS) analyses were carried out on an Autoflex III Samrtbeam MALDI-TOF spectrometer (Bruker Daltonics) which was used in linear mode with the following settings: 8.000–100.000 Th window, linear positive mode, ion source 1: 20 kV, ion source 2: 18.5 kV, lens: 9 kV, pulsed ion extraction of 120 ns, high gating ion suppression up to 1000 Mr. Mass calibration was performed externally with protein 1 standard calibration mixture (Bruker Daltonics) in the same range as the samples. Data acquisition was performed using FlexControl 3.0 software (Bruker Daltonics), and peak peaking and subsequent spectra analysis was performed using FlexAnalysis 3.0 software (Bruker Daltonics).

Solid-Phase Peptide Synthesis (SPPS). All (glyco)peptides were synthesized using the standard 9-fluorenylmethoxycarbonyl (Fmoc)/tert-butyloxycarbonyl (tBu) protection strategy. Automated, microwave (MW)-assisted SPPS was carried out on a CEM Liberty Blue™ MW synthesizer using default cycles and commercially available amino acids (Novabiochem®). Fmoc-deprotection cycles were performed with 20% piperidine in DMF, under microwave irradiation at 85-90 ºC for 1 minute. Amino acid (AA) couplings involved reaction of 5 eq. of AA (0.2 M), in the presence of *N,N′*-diisopropylcarbodiimide (DIC) and ethyl cyano(hydroxyimino)acetate (Oxyma Pure), under MW irradiation at 85–90 ºC for 2 minutes. Upon completion of the sequence, the resin was transferred in a solid-phase peptide synthesis vessel provided with a sintered glass filter, and thoroughly washed with MeOH prior treatment with the cleavage cocktail. For the synthesis of the TnMUC1 glycopeptide, glyco-amino acid building block Fmoc-Thr(O-α-Ac_3-GalNAc)-OH (**S3**, see Scheme S2) was coupled manually, by suspending the resin in the solid-phase peptide synthesis vessel, following addition of the reagents and gentle mixing of the resulting suspension via a N_2 stream.

II. SYNTHESES AND CHARACTERIZATION DATA

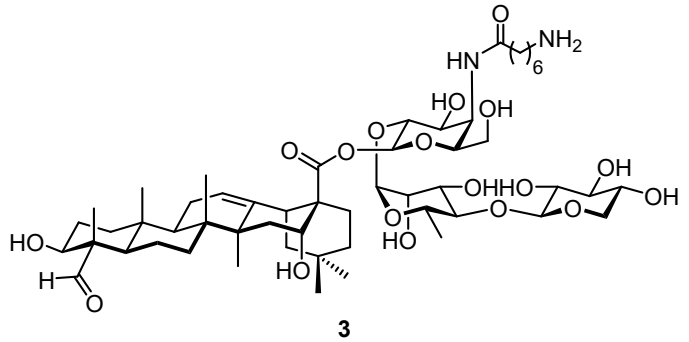
Synthetic saponin adjuvant 2



2

Data for **synthetic saponin 2** [CP-III-066] ($C_{59}H_{95}NO_{20}$) was in accordance with that of the previous report by Ghirardello *et al*.[1] **HPLC:** t_R = 17.57 min, gradient = 20-100% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+H]^+$ 1138.65, found 1138.72. **HRMS (MALDI-TOF):** *m/z* (monoisotopic) calcd for $[M+Na]^+$ 1160.6336, found 1160.6293. **^{1}H-NMR** (600 MHz, CD$_3$OD) characteristic resonances: δ 9.31 (s, 1H), 5.38 (d, *J* = 1.9 Hz, 1H), 5.34 (d, *J* = 7.9 Hz, 1H), 5.31 (t, *J* = 3.5, 1H), 4.50-4.47 (m, *2*H), 4.34-4.30 (m, 1H), 3.95 – 3.90 (m, 3H), 3.87 – 3.75 (m, 4H), 3.69 (td, *J* = 6.6, 1.7 Hz, 1H), 3.57 – 3.49 (m, 2H), 3.46 (ddd, *J* = 10.4, 8.9, 5.4 Hz, 1H), 3.41 (dd, *J* = 11.6, 6.9 Hz, 1H), 3.23 – 3.17 (m, 2H), 2.94 (dd, *J* = 14.3, 4.6 Hz, 1H), 1.41 (s, 3H), 1.02 (s, 3H), 1.01 (s, 3H), 0.96 (s, 3H), 0.88 (s, 3H), 0.78 (s, 3H).

Synthetic saponin scaffold 3

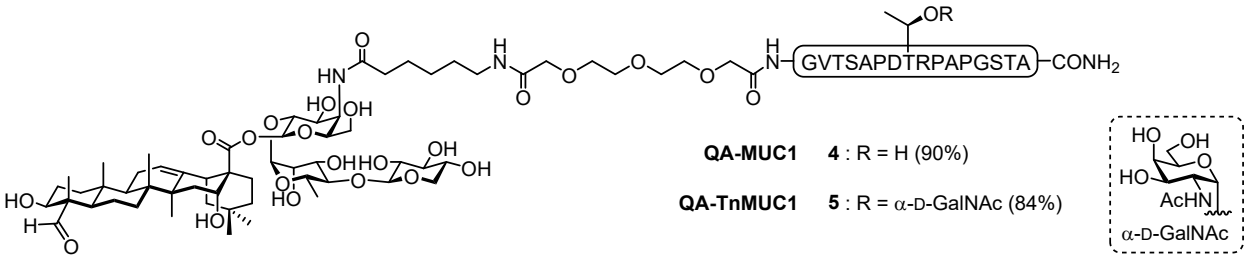


3

Data for **synthetic saponin 3** [CP-I-100] ($C_{53}H_{86}N_2O_{18}$) was in accordance with that of the previous report by Fernández-Tejada *et al*.[2] **HPLC:** t_R = 12.64 min, gradient = 20-100% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+H]^+$ 1039.60, found 1039.68. **HRMS (MALDI-TOF):** *m/z* (monoisotopic) calcd for $[M+Na]^+$ 1061.5764, found

1061.5720. **^{1}H-NMR** (600 MHz, CD_3OD) characteristic resonances: δ 9.31 (s, 1H), 5.42 (d, *J* = 1.8 Hz, 1H), 5.34 (t, *J* = 3.5 Hz, 1H), 5.31 (d, *J* = 3.6 Hz, 1H), 4.51 – 4.48 (m, 2H), 4.37 (dd, *J* = 4.7, 1.7 Hz, 1H), 3.96 (dd, *J* = 9.5, 4.7 Hz, 1H), 3.91 – 3.80 (m, 5H), 3.79 – 3.76 (m, 1H), 3.72 (td, *J* = 6.5, 1.8 Hz, 1H), 3.59 – 3.51 (m, 2H), 3.49 – 3.41 (m, 2H), 3.24 – 3.17 (m, 2H), 2.94 – 2.91 (m, 3H), 2.43 – 2.27 (m, 3H), 2.02 – 1.90 (m, 4H), 1.41 (s, 3H), 1.34 (d, *J* = 6.1 Hz, 3H), 1.08 – 1.04 (m, 1H), 1.02 (s, 3H), 1.01 (s, 3H), 0.95 (s, 3H), 0.88 (s, 3H), 0.77 (s, 3H).

Di-component saponin–Tn/MUC1 conjugates **4** and **5**



MUC1- and TnMUC1-containing, di-component constructs (4 and 5) [CP-I-128, CP-III-052, CP-III-098, CP-III-124, CP-III-128]. To a vial containing saponin amine **3**[1] (1.0 eq.) and *para*-nitrophenyl-activated MUC1 (glyco)peptide **11** or **12** (1.5 eq.), a solution of dry DMF containing DIPEA (3.0 eq.) was added (final conc. = 2.5 mM, with respect to **3**). The reaction mixture rapidly turned to a yellow solution, which was stirred at room temperature (r.t.) for 1 hour. The reaction was quenched by slowly adding a 0.05% TFA solution in H_2O/CH_3CN (1:1) until disappearance of the yellow coloration, then directly purified via RP-HPLC (20% solv. B for 5 min, then 20-100% over 25 min.). The collected fractions were lyophilized, affording saponin conjugates QA–MUC1 **4** and QA–TnMUC1 **5** as a white foam.

Data for QA–MUC1 **4** ($C_{122}H_{198}N_{22}O_{46}$): prepared on a 3.18 μmol scale, yield: 7.72 mg, 90%. **HPLC:** t_R = 11.98 min, gradient = 20-100% solv. B over 30 min, λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+3H]^{3+}$ 903.47, found 903.45; calcd for $[M+2H]^{2+}$ 1354.70, found 1355.00; calcd for $[2M+3H]^{3+}$ 1805.93, found 1805.96; calcd for $[3M+4H]^{4+}$ 2031.55, found 2031.06; calcd for $[4M+5H]^{5+}$ 2166.91, found 2166.90; calcd for $[5M+6H]^{6+}$ 2257.16, found 2257.63; calcd for $[M+H]^+$ 2708.39, found 2708.56. **HRMS (MALDI-TOF):** *m/z* (monoisotopic) calcd for $[M+H]^+$ 2708.3900, found 2708.3990.

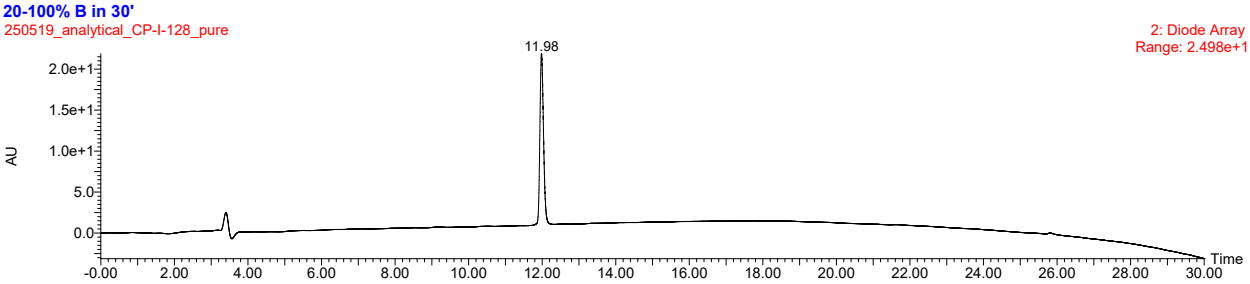


Fig. S1 RP-HPLC chromatogram of **4**.

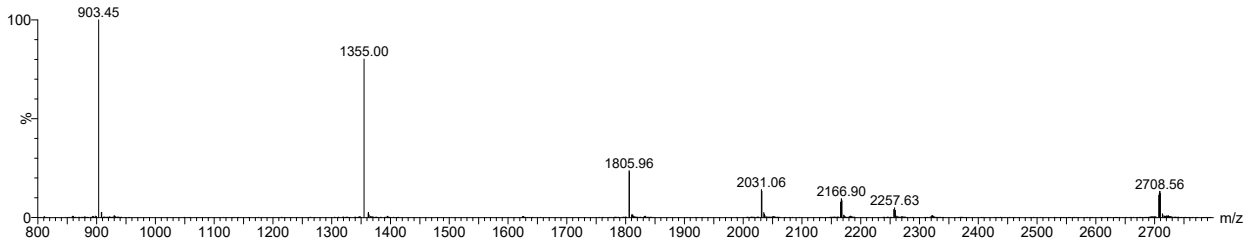


Fig. S2 LRMS (ESI$^+$-MS) spectrum of **4**.

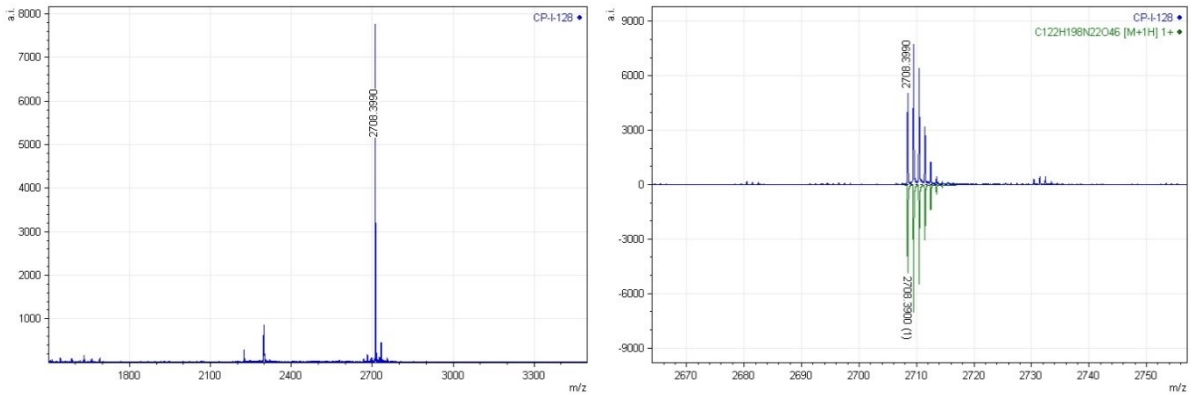


Fig. S3 HRMS (MALDI-TOF) spectrum of **4**.

Data for QA–TnMUC1 **5** ($C_{130}H_{211}N_{23}O_{51}$): prepared on a 3.59 µmol scale, yield: 8.75 mg, 84%. **HPLC:** t_R = 12.00 min, gradient = 20-100% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+3H]^{3+}$ 971.16, found 971.02; calcd for $[3M+5H]^{5+}$ 1747.29, found 1746.49; calcd for $[2M+3H]^{3+}$ 1941.31, found 1940.99; calcd for $[3M+4H]^{4+}$ 2183.85, found 2184.16; calcd for $[4M+5H]^{5+}$ 2329.38, found 2329.22; calcd for $[5M+6H]^{6+}$ 2426.39, found 2426.70; calcd for $[M+H]^+$ 2911.47, found 2911.94. **HRMS (MALDI-TOF):** *m/z* (monoisotopic) calcd for $[M+H]^+$ 2911.4693, found 2911.4576.

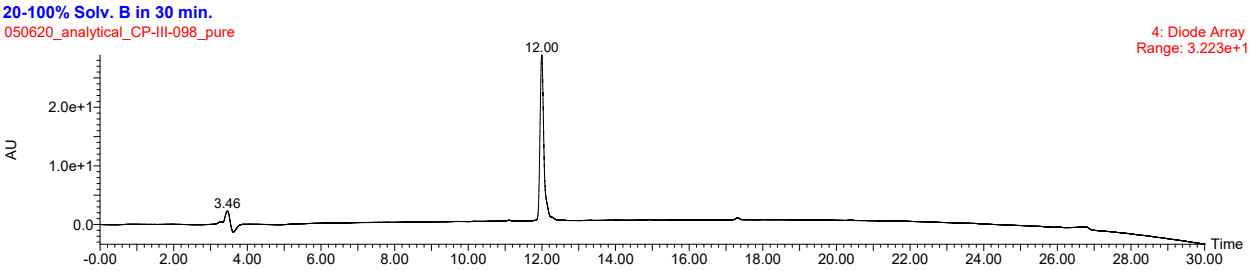


Fig. S4 RP-HPLC chromatogram of **5**.

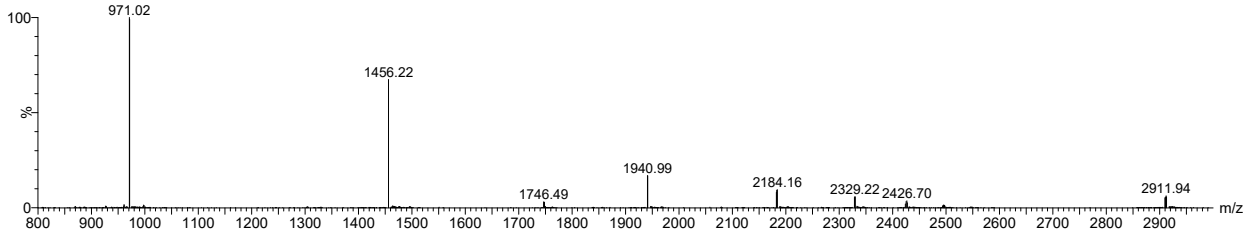


Fig. S5 LRMS (ESI^+-MS) spectrum of **5**.

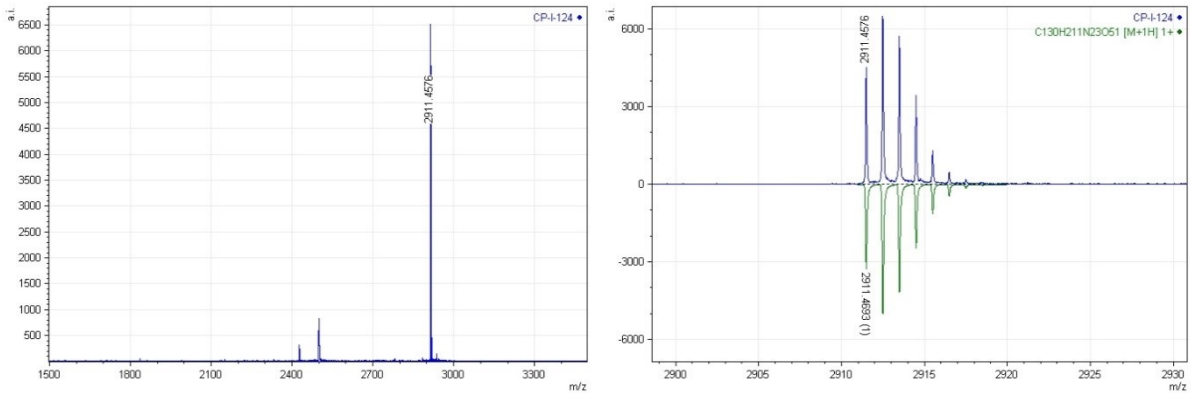
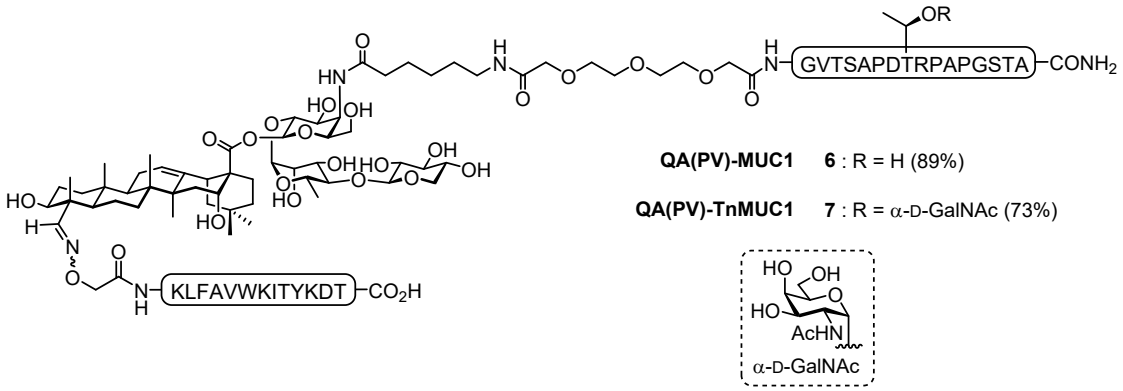


Fig. S6 HRMS (MALDI-TOF) spectrum of **5**.

Tri-component saponin(PV)–Tn/MUC1 conjugates **6** and **7**



MUC1- and TnMUC1-containing, tri-component vaccines (6 and 7) [CP-III-040, CP-III-058, CP-III-112, CP-III-126]. A stirred solution of QA–MUC1 **4** or QA–TnMUC1 **5** (1.0 eq, 1.5 mM) and aminooxy-functionalized $PV_{103\text{-}115}$ peptide **13** (2.0 eq.) in a H_2O/CH_3CN (2:1) mixture containing 0.05% TFA was heated at 40 ºC for 24 h. The reaction mixture was directly purified via RP-HPLC (30% solv. B for 5 min, then 30-80% over 25 min.). The collected fractions were lyophilized, affording the saponin conjugates QA(PV)–MUC1 **6** and QA(PV)–TnMUC1 **7** as a white foam.

Data for QA(PV)–MUC1 **6** ($C_{203}H_{320}N_{40}O_{66}$): prepared on a 1.14 µmol scale, yield: 4.46 mg, 89%. **HPLC:** t_R = 12.90 min, gradient = 20-100% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+6H]^{6+}$ 730.06, found 730.32; calcd for $[M+5H]^{5+}$ 875.87, found 876.04; calcd for $[M+4H]^{4+}$ 1094.58, found 1094.78; calcd for $[M+3H]^{3+}$ 1459.10, found 1459.34; calcd for $[2M+5H]^{5+}$ 1750.72, found 1750.70; calcd for $[M+2H]^{2+}$ 2188.15, found 2188.30. **HRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+H]^+$ 4375.2982, found 4375.2856.

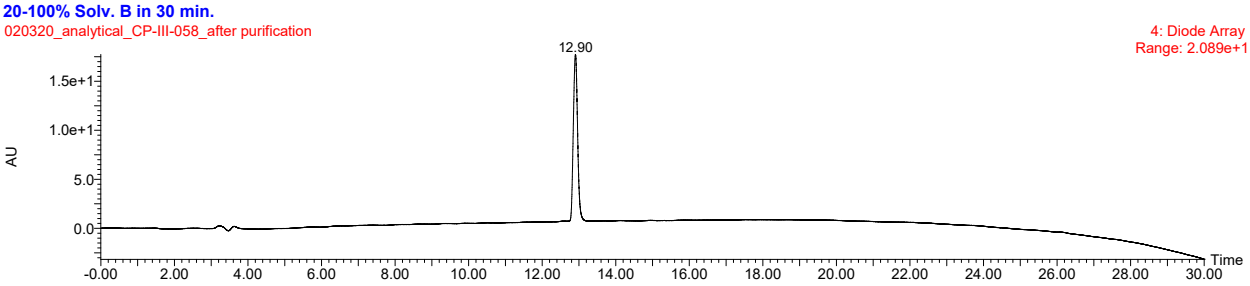


Fig. S7 RP-HPLC chromatogram of **6**.

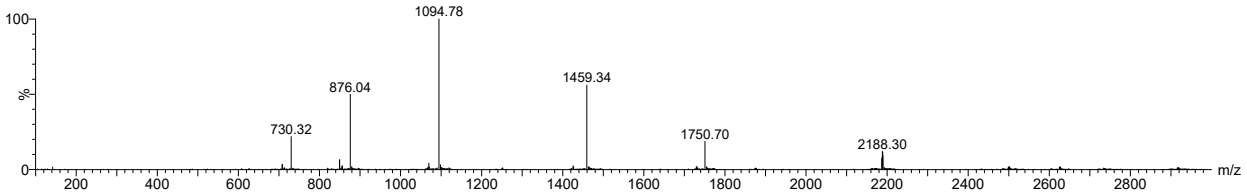


Fig. S8 LRMS (ESI$^+$-MS) spectrum of **6**.

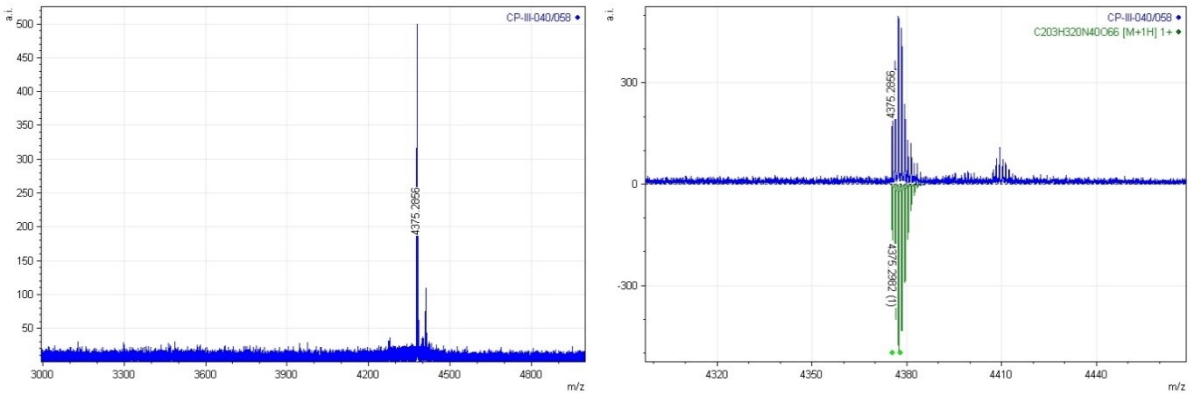


Fig. S9 HRMS (MALDI-TOF) spectrum of **6**.

Data for QA(PV)–TnMUC1 **7** ($C_{211}H_{333}N_{41}O_{71}$): prepared on a 0.86 µmol scale, yield: 2.87 mg, 73%. **HPLC:** t_R = 12.59 min, gradient = 20-100% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+6H]^{6+}$ 763.90, found 764.32; calcd for $[M+5H]^{5+}$ 916.48, found 916.86; calcd for $[M+4H]^{4+}$ 1145.35, found 1145.46; calcd for $[M+3H]^{3+}$ 1526.80, found 1527.14; calcd for $[2M+5H]^{5+}$ 1831.96, found 1832.14; $[M+2H]^{2+}$ 2289.69, found 2289.18. **HRMS (MALDI-TOF):** *m/z* (monoisotopic) calcd for $[M+H]^+$ 4578.3778, found 4578.3799.

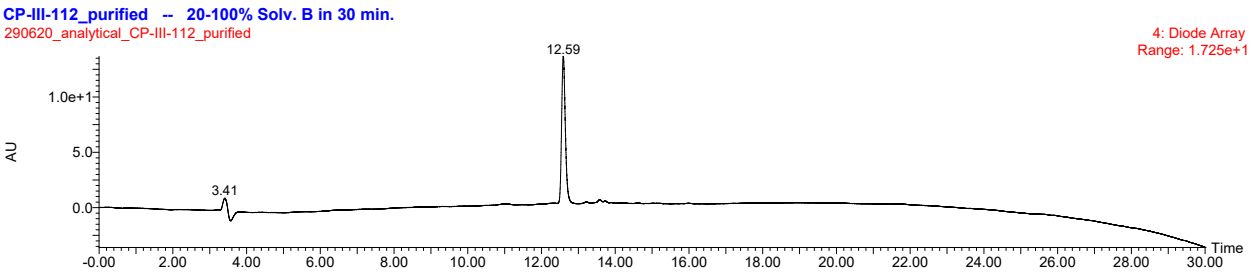


Fig. S10 RP-HPLC chromatogram of **7**.

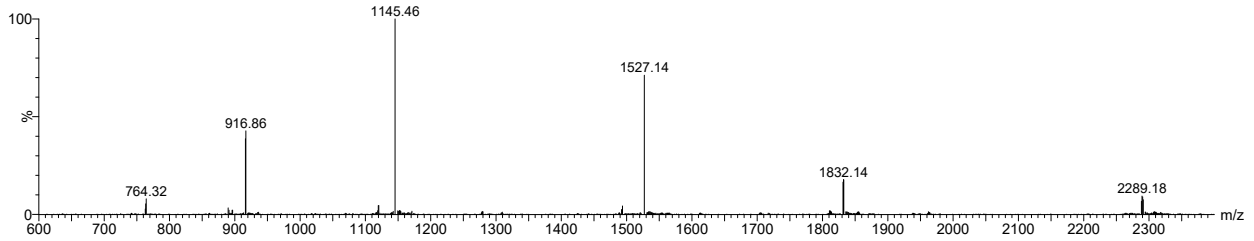


Fig. S11 LRMS (ESI^+-MS) spectrum of **7**.

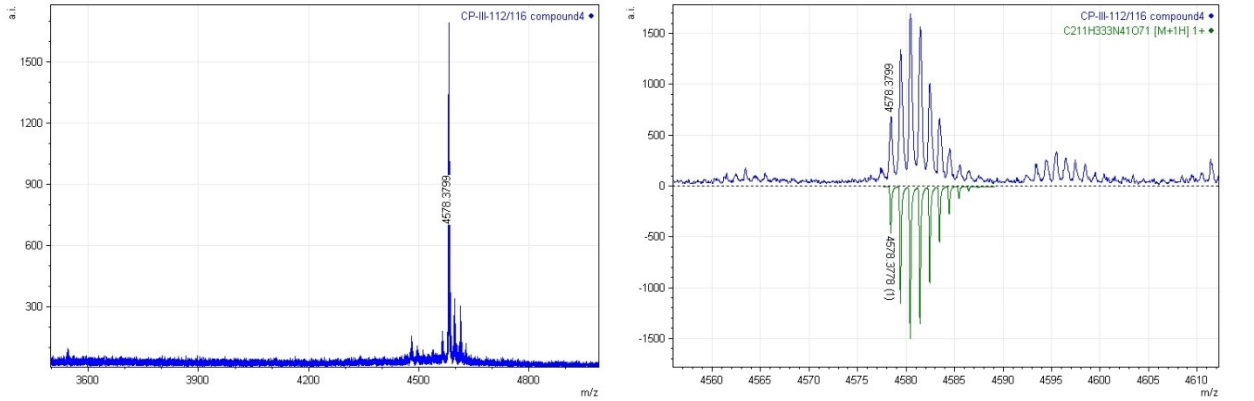
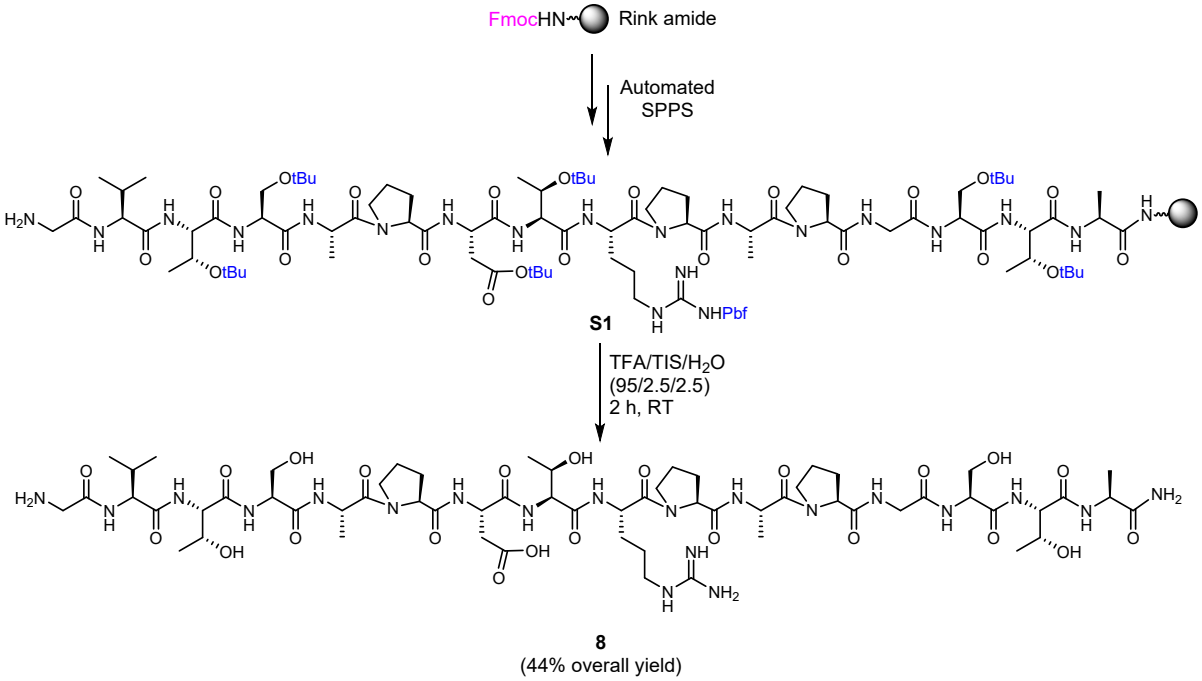


Fig. S12 HRMS (MALDI-TOF) spectrum of **7**.

MUC1 peptide 8



Scheme S1 Synthesis of MUC1 peptide fragment **8**, consisting of part of the MUC1 tandem repeat domain.

MUC1 peptide "GVTSAPDTRPAPGSTA" (8) [CP-I-018, CP-I-052]. Resin-bound sequence **S1** was synthesized on 192 mg of Rink Amide MBHA resin (loading = 0.520 mmol/g) via MW-assisted automated SPPS, using commercially available amino acids (Novabiochem®). Upon sequence completion, the resin was washed with MeOH (5 × 10 mL), transferred in a 50 mL falcon tube and treated with 10 mL of a TFA/TIS/H_2O (95/2.5/2.5) cocktail. After for 2 hours of gentle stirring (rocker platform) at r.t., the filtrate was added to ice-cold Et_2O to induce precipitation of crude peptide **8**. Following three cycles of (i) centrifugation (3000 rpm at r.t.), (ii) removal of the supernatant, and (iii) resuspension in ice-cold Et_2O, the precipitate was dried under vacuum to provide 150 mg of crude **8** as a white powder. Purification of the crude was performed via RP-HPLC (5% solv. B for 5 min, then 5-15% over 25 min.), the collected fractions were lyophilized, affording 65.28 mg of MUC1 peptide **8** ($C_{61}H_{102}N_{20}O_{23}$) as a white foam in a 44% overall yield.

HPLC: t_R = 13.03 min, gradient = 5-40% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+2H]^{2+}$ 742.38, found 742.41; calcd for $[M+H]^+$ 1483.75, found 1483.61. **HRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+Na]^+$ 1505.7324, found 1505.7296.

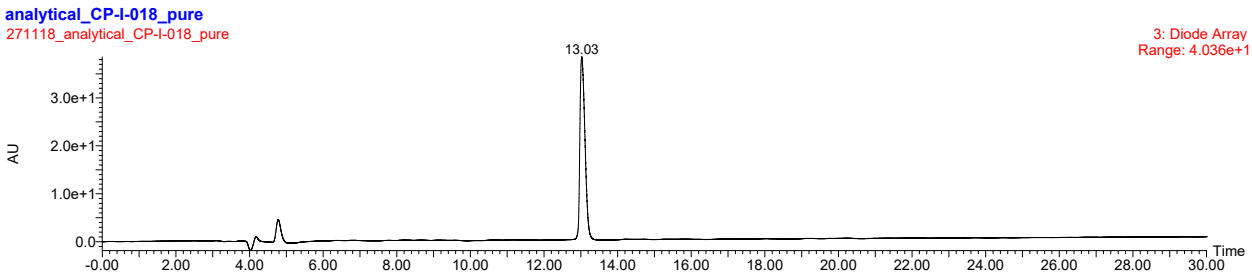


Fig. S13 RP-HPLC chromatogram of **8**.

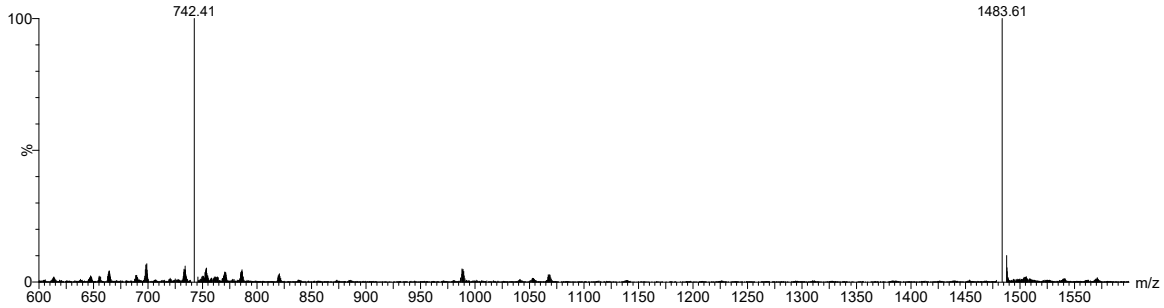


Fig. S14 LRMS (ESI$^+$-MS) spectrum of **8**.

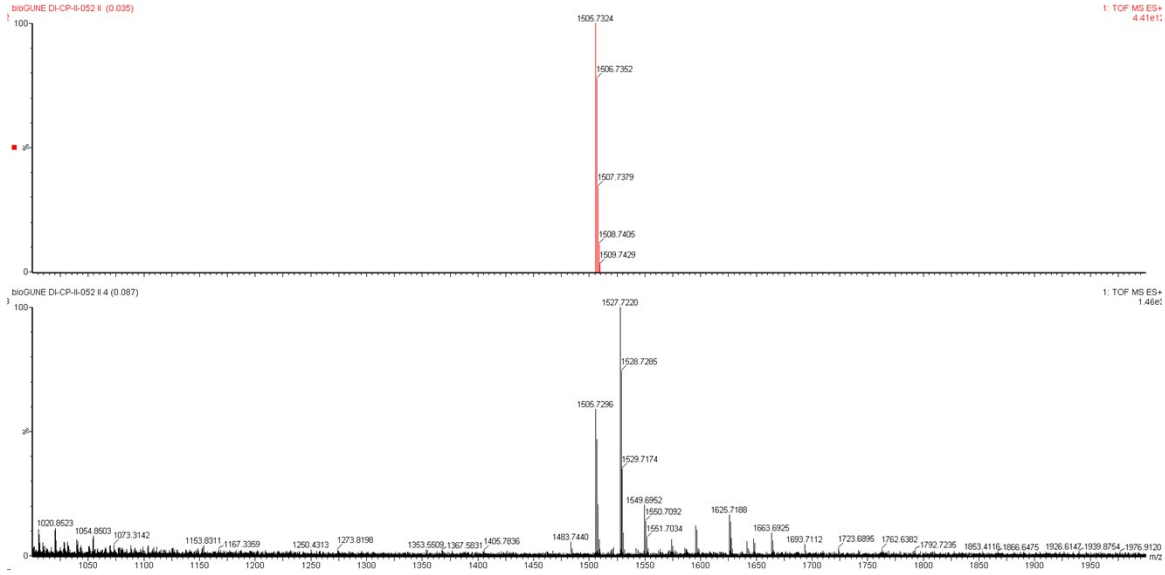


Fig. S15 HRMS (ESI$^+$-MS) spectrum of **8**.

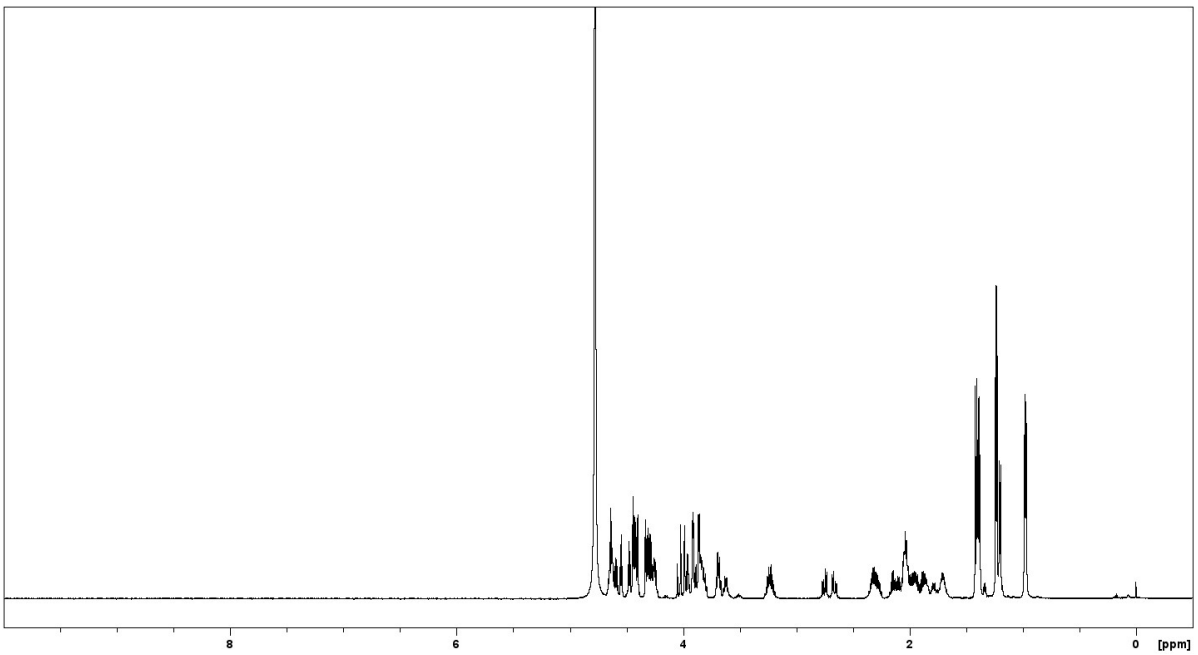
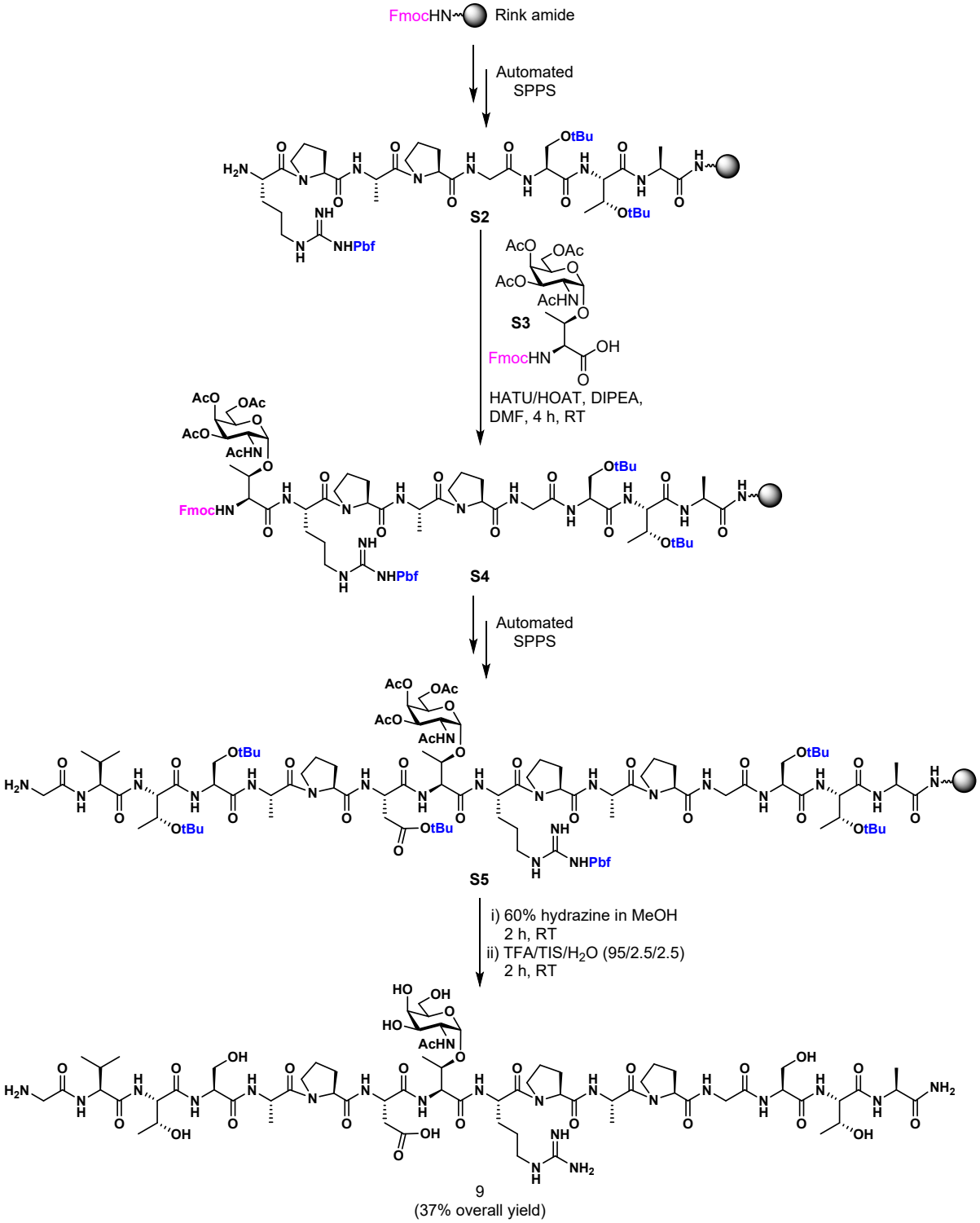


Fig. S16 ^{1}H-NMR spectrum of **8** (7 mM, PBS/D_2O, 298 K, 600 MHz).

TnMUC1 glycopeptide 9



Scheme S2 Synthesis of TnMUC1 glycopeptide fragment **9**.

TnMUC1 glycopeptide "GVTSAPDT(αGalNAc)RPAPGSTA" (9) [CP-I-050, CP-I-056, CP-I-058, CP-I-062, CP-I-068]. Resin-bound sequence **S2** was synthesized on 192 mg of Rink Amide MBHA resin (loading = 0.520 mmol/g) via MW-assisted automated SPPS,. The functionalized resin was then transferred in a solid-phase peptide synthesis vessel provided with a sintered glass filter to manually perform the coupling of glyco-amino acid **S3**[3] (see characterization data on pages S16-S17). After two cycles of swelling (10 min.)/solvent filtration, first with CH_2Cl_2, then with DMF, resin **S2** was suspended in 2.0 mL of DMF. In a separate flask, a mixture of DIPEA (80 µL, 0.46 mmol, 4.6 eq.), HATU (60.0 mg, 0.158 mmol, 1.6 eq.), HOAt (21.0 mg, 0.154 mmol, 1.5 eq.) and **S3** (112.65 mg, 0.168 mmol, 1.7 eq.) in 5.0 mL DMF was stirred at r.t. for 10 min., then added to the reactor. The suspension was stirred using a stream of N_2, after 4 hours at r.t. the solvent was removed through the filter and the resin washed with CH_2Cl_2 (5 x 10 mL) and DMF (5 x 10 mL). Resin **S4** was then transferred back to the peptide synthesizer, and sequence **S5** was completed via automated SPPS. The resin was transferred in the glass reactor, washed with MeOH (5 x 10 mL), and treated with a 60% hydrazine in MeOH for 2 hours at r.t. under a stream of N_2 to remove the acetyl protecting groups. After washing with MeOH (5 × 10 mL) and CH_2Cl_2 (5 × 10 mL), the resin was lastly transferred in a 50 mL falcon tube and treated with 10 mL of a TFA/TIS/H_2O (95/2.5/2.5) cocktail. After for 2 hours of gentle stirring (rocker platform) at r.t., the filtrate was added to ice-cold Et_2O to induce precipitation of crude glycopeptide **9**. Following three cycles of (i) centrifugation (3000 rpm at r.t.), (ii) removal of the supernatant, and (iii) resuspension in ice-cold Et_2O, the precipitate was dried under vacuum to provide 170 mg of crude **9** as a white powder. Purification of the crude peptide was performed via RP-HPLC (5% solv. B for 5 min, then 5-15% over 25 min.), the collected fractions were lyophilized, affording 62.41 mg of TnMUC1 glycopeptide **9** ($C_{69}H_{115}N_{21}O_{28}$) as a white foam in a 37% overall yield.

HPLC: t_R = 12.74 min, gradient = 5-40% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+3H]^{3+}$ 562.95, found 563.07; calcd for $[M+2H]^{2+}$ 843.92, found 844.15; calcd for $[M+H]^+$ 1686.83, found 1686.30. **HRMS (MALDI-TOF):** *m/z* (monoisotopic) calcd for $[M+H]^+$ 1686.8289, found 1686.8384.

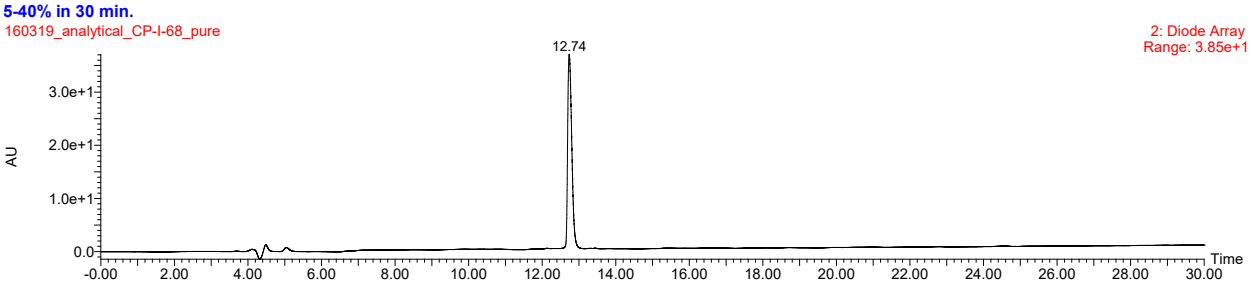


Fig. S17 RP-HPLC chromatogram of **9**.

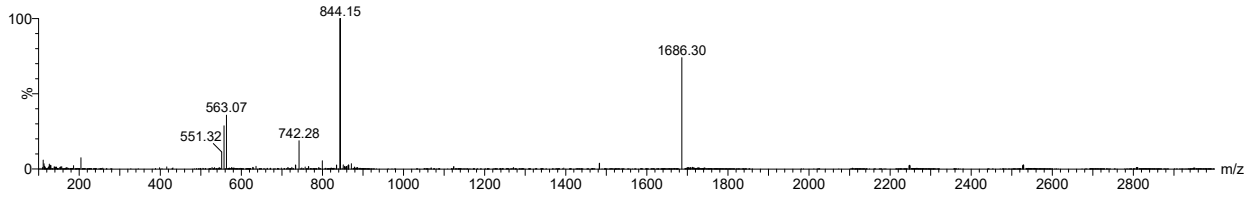


Fig. S18 LRMS (ESI$^+$-MS) spectrum of **9**.

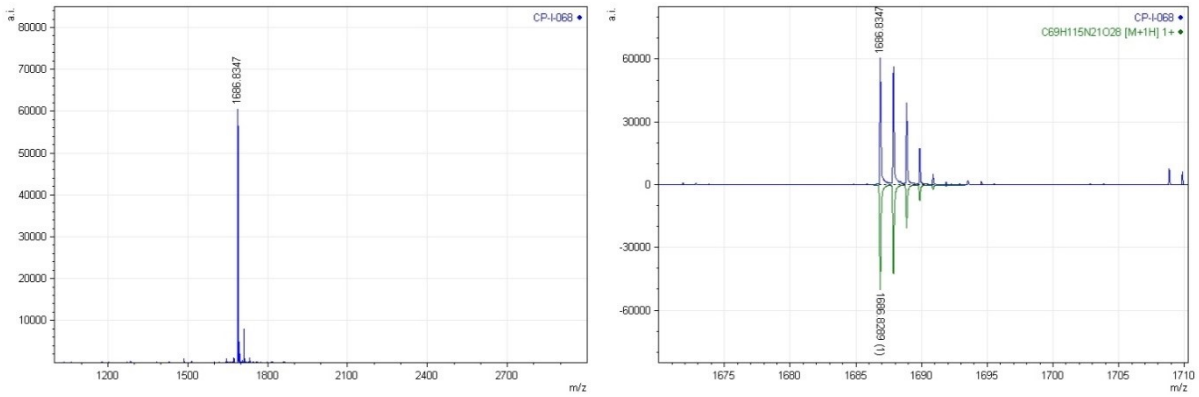


Fig. S19 HRMS (MALDI-TOF) spectrum of **9**.

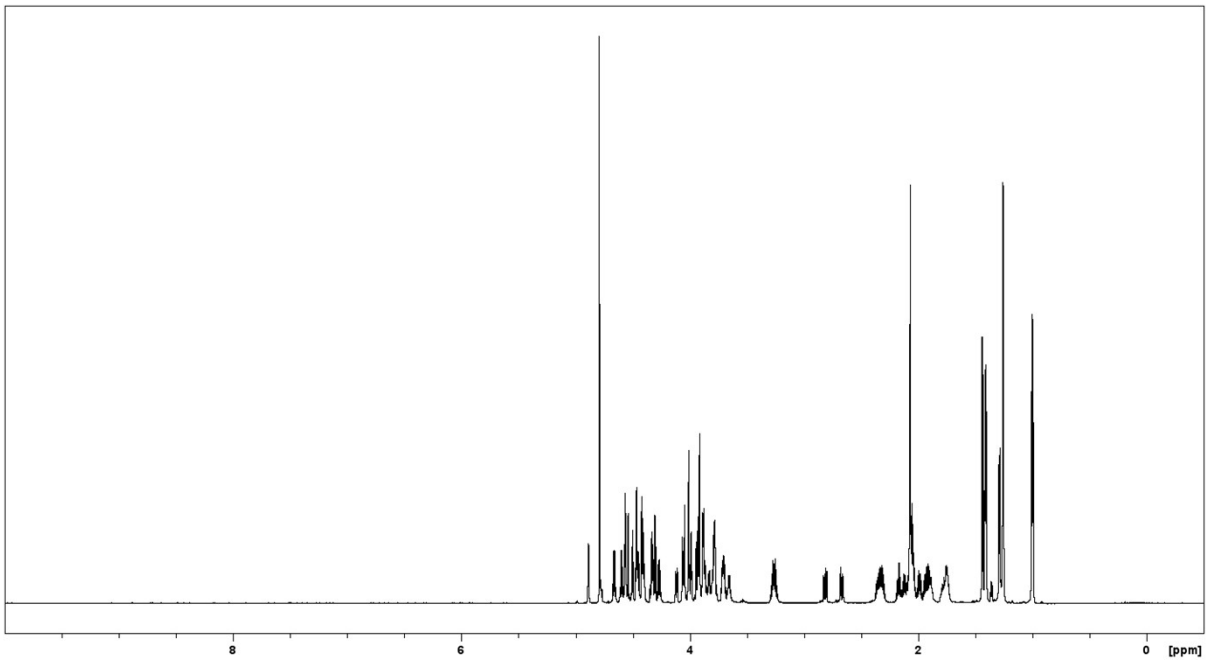
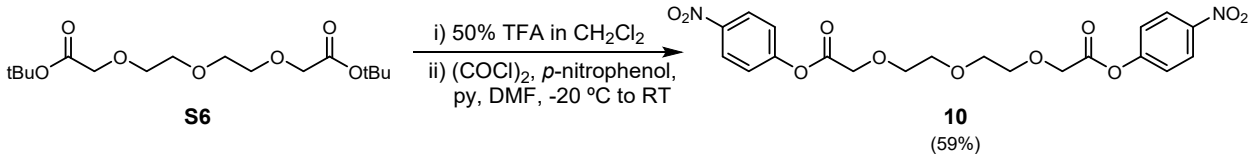


Fig. S20 ^{1}H-NMR spectrum of **9** (1.6 mM, PBS/D_2O, 298 K, 600 MHz).

Fmoc-Thr[αGalNAc(OAc)$_3$]-OH (S3) [CP-I-024, CP-I-054] ($C_{33}H_{38}N_2O_{13}$) was prepared following previously reported synthetic procedures.[4,5] **HPLC:** t_R = 18.27 min, gradient = 20-100% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+H]^+$ 671.25, found 671.31. **HRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+Na]^+$ 693.2272, found 693.2315. **^{1}H-NMR** data in accordance with Payne *et al*.[6] (600 MHz, $CDCl_3$/CD_3OD 1:1 v/v, major rotamer reported) δ 7.76 (d, *J* = 7.6 Hz, 2H), 7.66 – 7.62 (m, 2H), 7.39 – 7.35 (m, 2H), 7.32 – 7.28 (m, 2H), 5.37 – 5.36 (m, 1H), 5.03 (dd, *J* = 11.5, 3.3 Hz, 1H), 4.94 (d, *J* = 3.9 Hz, 1H), 4.54 – 4.35 (m, 4H), 4.30 – 4.03 (m, 6H), 2.14 (s, 3H), 2.03 (s, 3H), 1.96 (s, 3H), 1.92 (s, 3H), 1.24 (d, *J* = 6.5 Hz, 3H).

Linker 10



Scheme S3 Synthesis of *p*-nitrophenyl activated linker **10**.

***Para*-nitrophenyl-activated homobifunctional linker (10)** [CP-III-032]. A solution of 3,6,9-trioxaundecane-1,11-dioic acid di-*tert*-butyl ester **S6**[7] (1.038 g, 3.07 mmol, 1.0 eq. – see characterization data on pages S21-S23) in a TFA/CH_2Cl_2 (1:1, 3.0 mL) mixture was stirred at r.t. for 4 hours. The solvent mixture was removed under high vacuum trough a Schlenk line, and co-evaporated with dry toluene (3 × 1 mL). The resulting di-acid was solubilized in dry CH_2Cl_2 (7.0 mL), and the resulting solution cooled to 0 ºC. Oxalyl chloride (1.6 mL, 18.63 mmol, 6.1 eq.) and a catalytic amount of dry DMF (60 µL) were added, and the mixture was stirred at 0 ºC for 10 minutes before allowing it to passively warm to r.t. for 1 hour. To remove the volatiles, the reaction mixture was concentrated under reduced pressure (via Schlenk line), then dry CH_2Cl_2 (7.0 mL) was added. This operation was repeated two times, then the reaction mixture was cooled again to 0 ºC. In a separated flask, azeotropically dried (dry toluene: 3 × 1 mL) *para*-nitrohpenol (940 mg, 6.76 mmol, 2.2 eq.) was solubilized in a CH_2Cl_2/THF (1:1, 14 mL) mixture, and the resulting solution was cooled to –20 ºC. To this solution, the mixture containing the activated di-acid was added via cannula, followed by dry pyridine (550 µL, 6.94 mmol, 2.3 eq.). The reaction mixture was stirred at –20 ºC for 15 minutes and then allowed to passively warm to r.t.. After 2 hours, the reaction mixture was diluted with CH_2Cl_2 (120 mL) and washed with 1% aq. AcOH (50 mL). The organic phase was dried over $MgSO_4$, filtered, and the solvent removed under vacuum to afford 1.54 g of brown solid as crude. The residue was purified via silica gel flash chromatography (eluent: 0 to 20% AcOEt in toluene), affording 836 mg of doubly activated linker **10** ($C_{20}H_{20}N2O_{11}$) as white solid in a 59% yield over two steps.

HPLC: t_R = 23.56 min, gradient = 20-100% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+H]^+$ 465.11, found 465.12. **HRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+Na]^+$ 487.0965, found 487.0955. **TLC**: R_f = 0.32 (toluene/AcOEt, 7:3, v/v). **^{1}H-NMR** (600 MHz, $CDCl_3$) δ 8.28 – 8.25 (m, 4H, Ph), 7.33 – 7.29 (m, 4H, Ph), 4.45 (s, 4H, $COCH_2O$) 3.86 – 3.83 (m, 4H, O-CH_2CH_2-O), 3.78 – 3.76 (m, 4H, O-CH_2CH_2-O). **^{13}C NMR** (151 MHz, $CDCl_3$): δ = 168.26 (CO), 154.98 (Ph), 145.72 (Ph), 125.41 (Ph), 122.36 (Ph), 71.38 (O-CH_2CH_2-O), 70.87(O-CH_2CH_2-O), 68.72 ($COCH_2$).

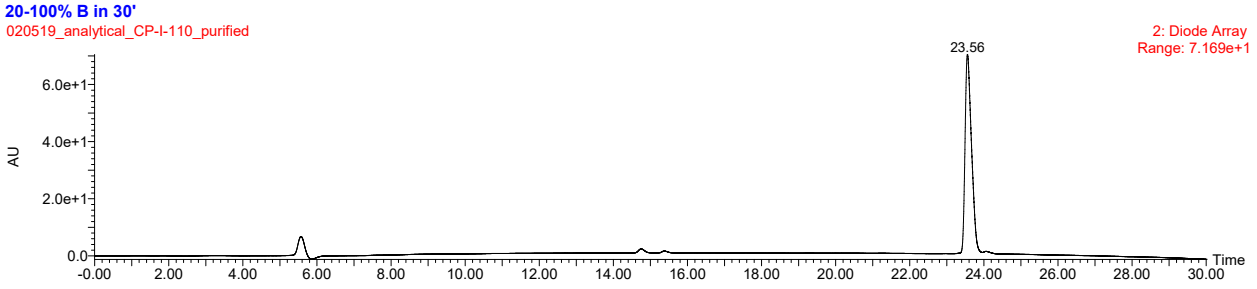


Fig. S21 RP-HPLC chromatogram of **10**.

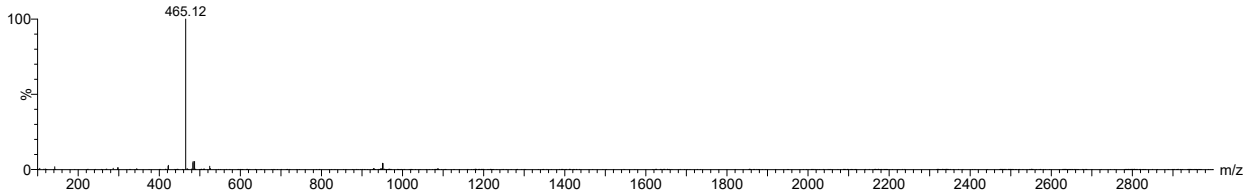


Fig. S22 LRMS (ESI^+-MS) spectrum of **10**.

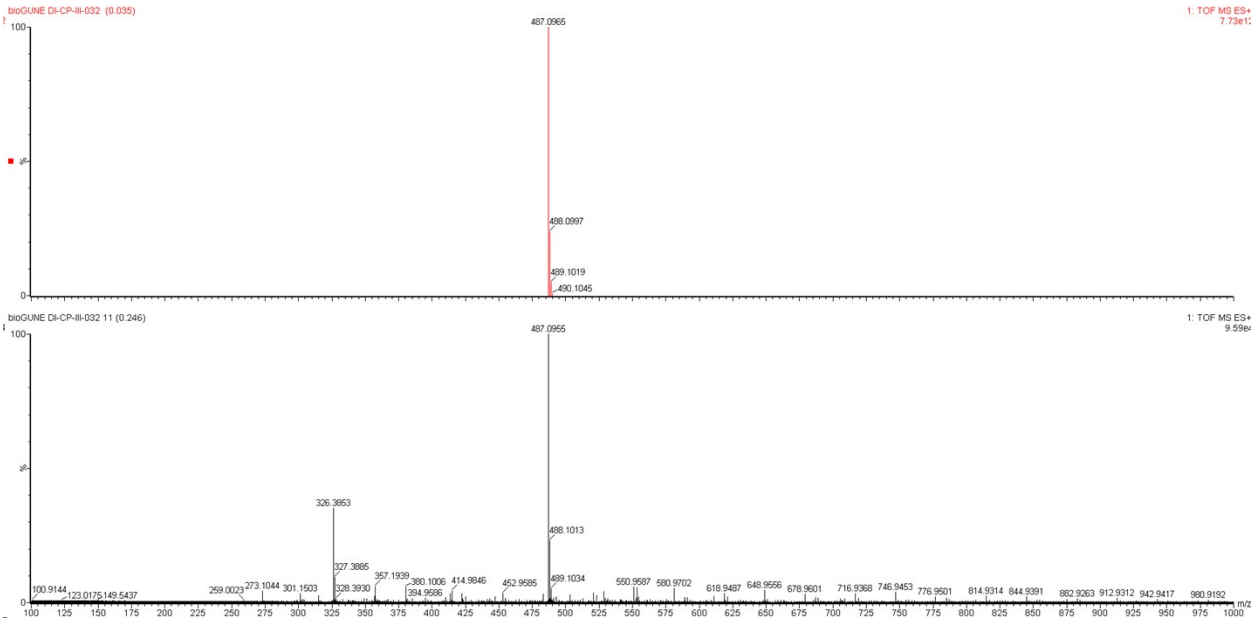


Fig. S23 HRMS (ESI^+-MS) spectrum of **10**.

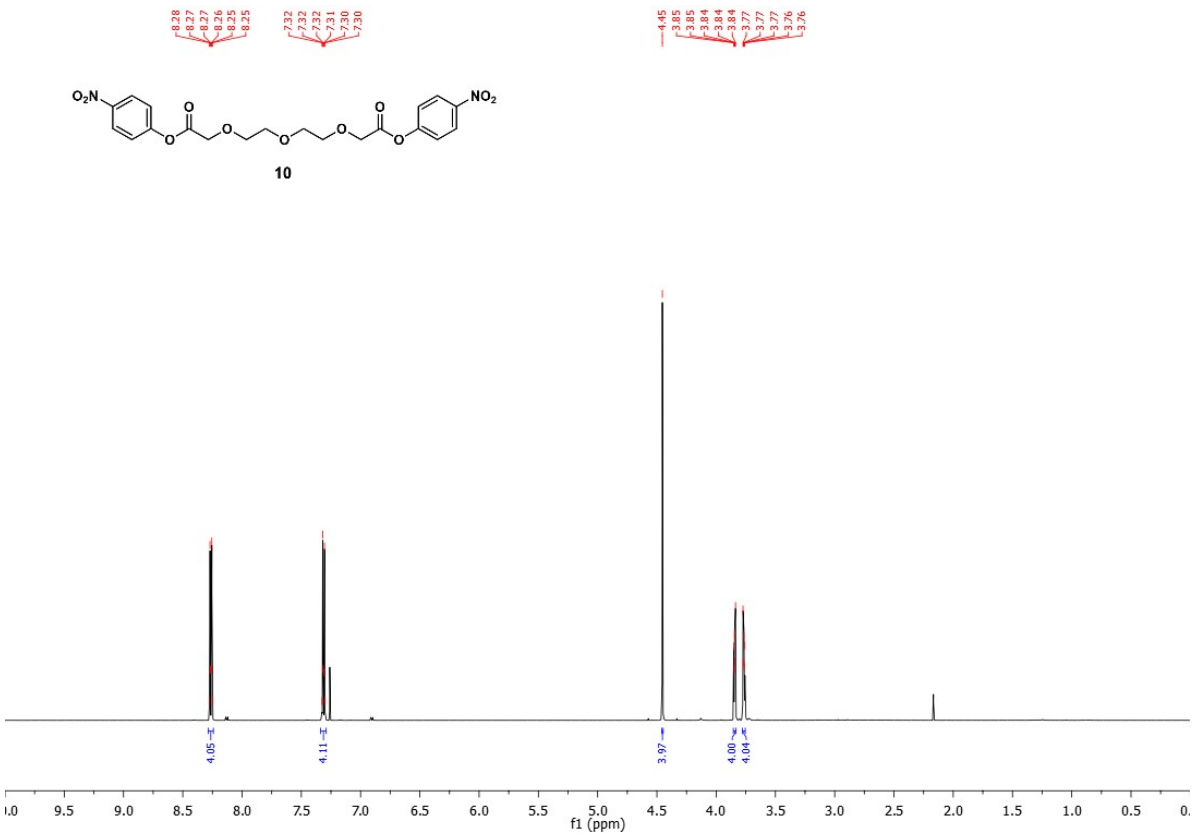


Fig. S24 ^{1}H-NMR spectrum of **10**

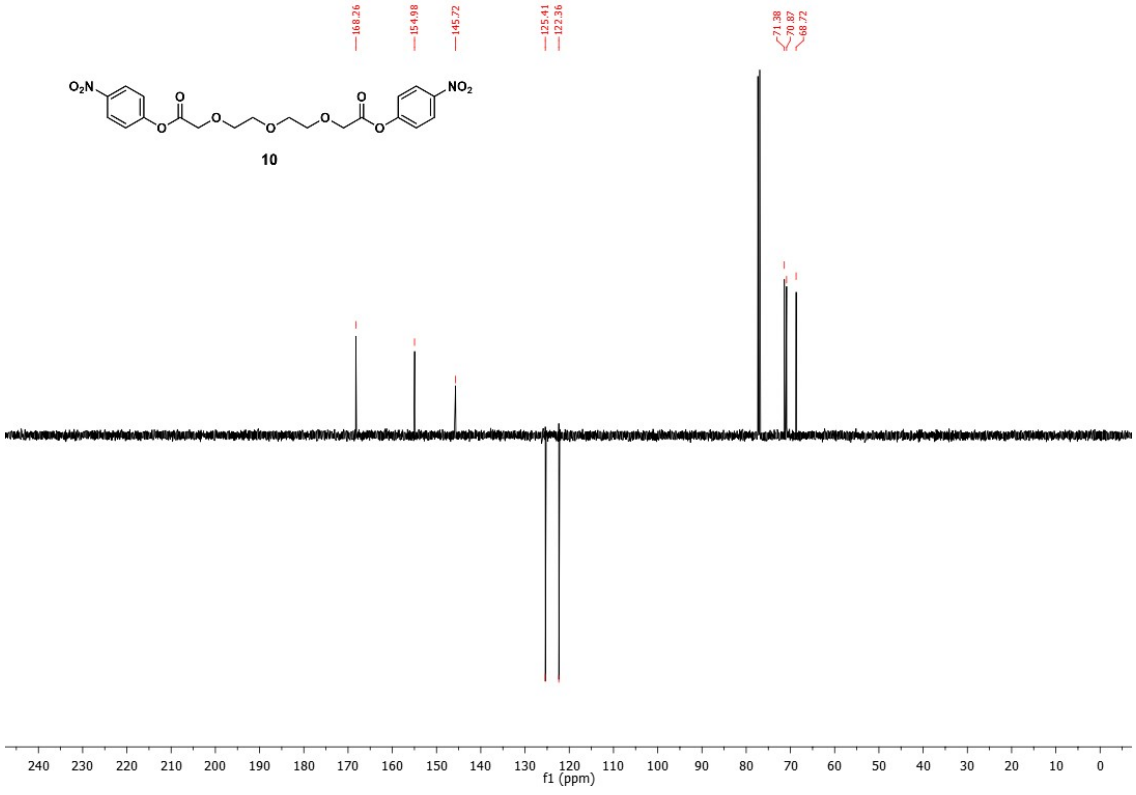
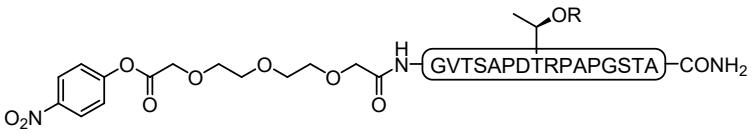


Fig. S25 ^{13}C-NMR spectrum of **10**.

Data for **3,6,9-trioxaundecane-1,11-dioic acid di-*tert*-butyl ester (S6)** [CP-III-030] ($C_{16}H_{30}O_7$) was in accordance with that of the previous report by Wittman *et al*.[7]

HPLC: t_R = 18.42 min, gradient = 20-100% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+H]^+$ 335.21, found 335.27, calcd for $[M+Na]^+$ 357.19, found 357.31. **HRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+Na]^+$ 357.1889, found 357.1925. **^{1}H-NMR** (600 MHz, $CDCl_3$) δ 4.01 (s, 4H), 3.75 – 3.66 (m, 8H), 1.46 (s, 18H).

Linker-containing (Tn)MUC1 (glyco)peptides **11** and **12**



11 : R = H (73%)

12 : R =α-D-GalNAc (70%)

***Para*-nitrophenyl-activated, linker-containing, MUC1- and TnMUC1 sequences (11 and 12)** [CP-I-118, CP-III-016, CP-III-046, CP-III-090, CP-III-092]. To a stirred solution of MUC1 peptide **8** or TnMUC1 glycopeptide **9** (1.0 eq.) in dry DMF containing DIPEA (2.5 eq.), a solution of di-activated linker **10** (5.0 eq.) in dry DMF was added (final conc. = 32 mM with respect to **10**). The reaction mixture readily turned yellow and was kept under stirring at r.t. After 1 hour, the reaction was quenched by slowly adding a 0.05% TFA solution in H_2O/CH_3CN (1:1) until disappearance of the yellow coloration, then directly purified via RP-HPLC (15% solv. B for 5 min, then 15-50% over 25 min.). The collected fractions were lyophilized affording the linker–MUC1 peptide **11** and linker-TnMUC1 glycopeptide **12** as a white foam (73% and 70% yield, respectively).

Data for **11** ($C_{75}H_{117}N_{21}O_{31}$): prepared on a 9.62 µmol scale, yield: 12.71 mg, 73%. **HPLC:** t_R = 17.10 min, gradient = 5-70% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+3H]^{3+}$ 603.62, found 603.52; calcd for $[M+2H]^{2+}$ 904.92, found 904.68; calcd for $[M+H]^+$ 1808.83, found 1808.49. **HRMS (MALDI-TOF):** *m/z* (monoisotopic) calcd for $[M+H]^+$ 1808.8293, found 1808.8181.

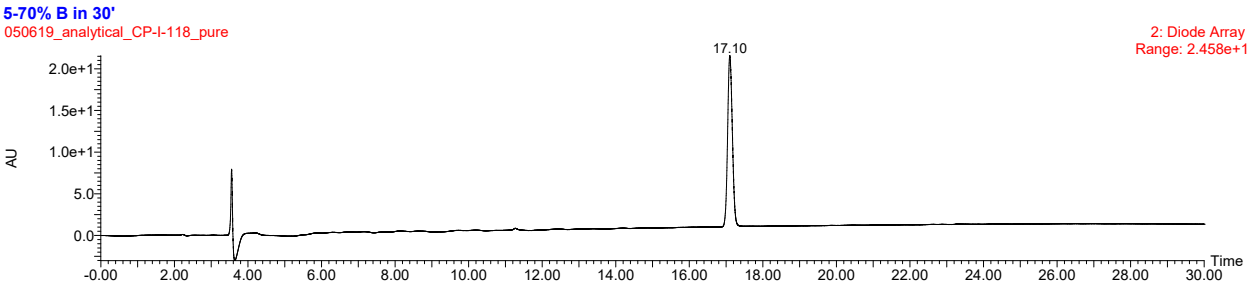


Fig. S26 RP-HPLC chromatogram of **11**.

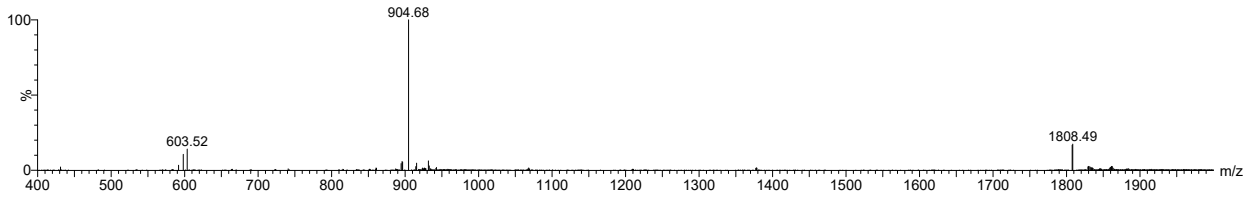


Fig. S27 LRMS (ESI$^+$-MS) spectrum of **11**.

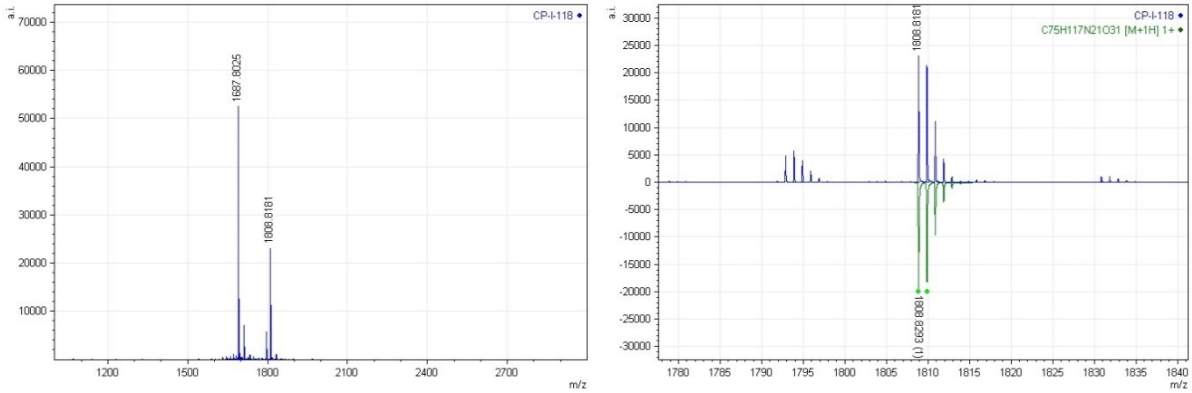


Fig. S28 HRMS (MALDI-TOF) spectrum of **11**.

Data for **12** ($C_{83}H_{130}N_{22}O_{36}$): prepared on a 9.47 µmol scale, yield: 13.40 mg, 70%. **HPLC:** t_R = 16.73 min, gradient = 5-70% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+3H]^{3+}$ 671.31, found 671.38; calcd for $[M+2H]^{2+}$ 1006.46, found 1006.47; calcd for $[2M+3H]^{3+}$ 1341.61, found 1341.51; calcd for $[3M+4H]^{4+}$ 1509.18, found 1508.87; calcd for $[M+H]^+$ 2011.91, found 2012.21. **HRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+H]^+$ 2011.9087, found 2011.9045.

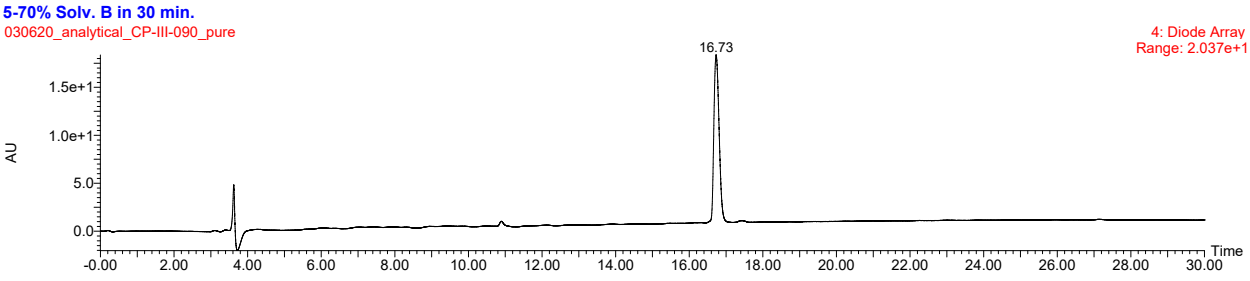


Fig. S29 RP-HPLC chromatogram of **12**.

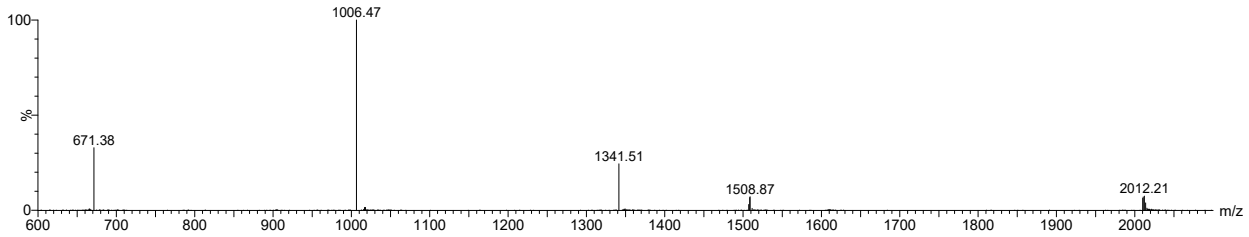


Fig. S30 LRMS (ESI$^+$-MS) spectrum of **12**.

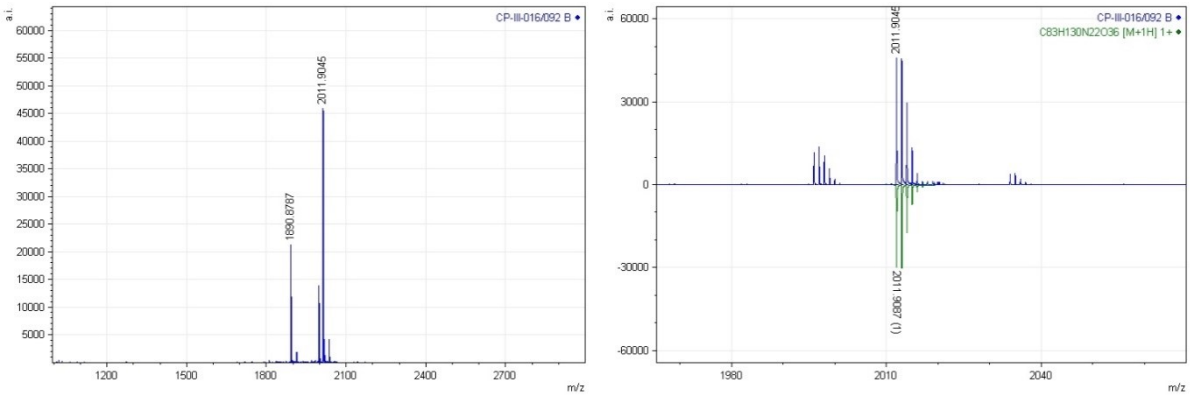
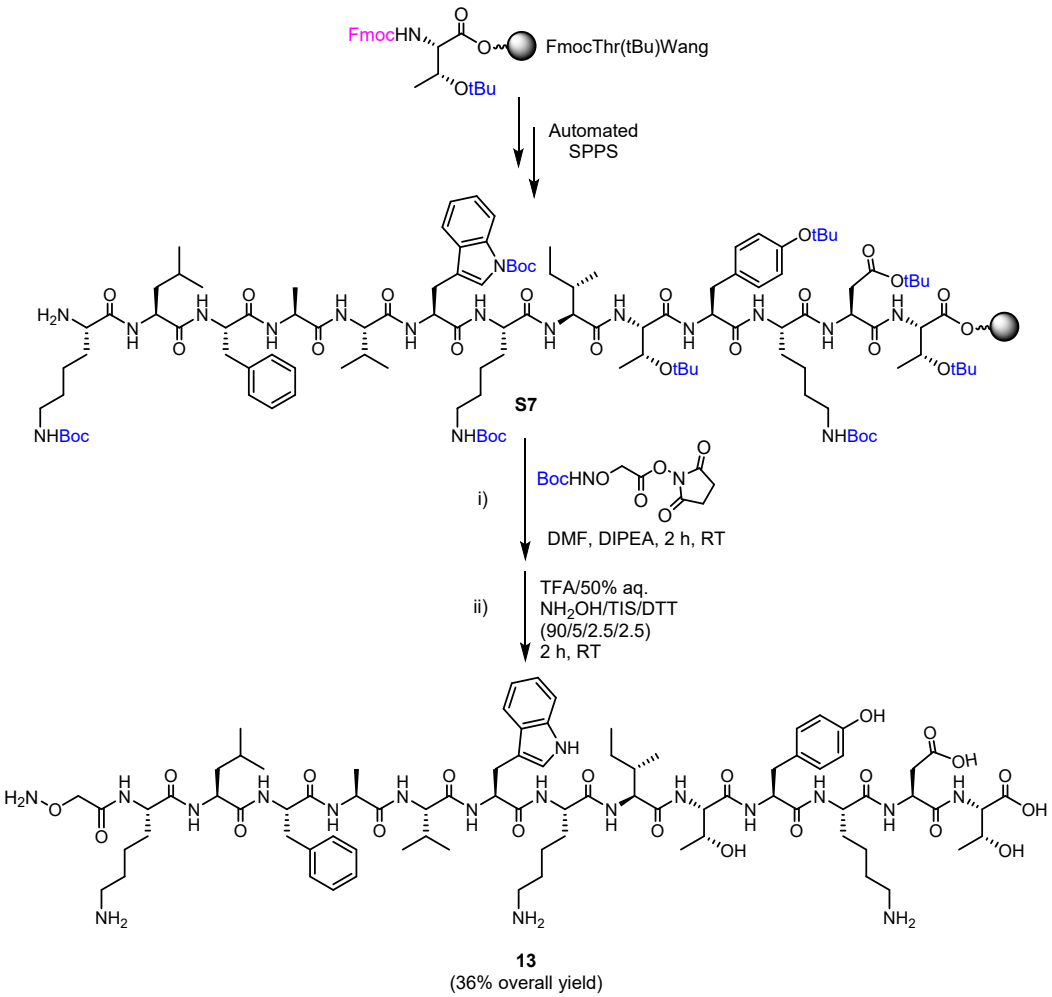


Fig. S31 HRMS (MALDI-TOF) spectrum of **12**.

Aminooxy-functionalized PV peptide 13



13
(36% overall yield)

Scheme S4 Synthesis of aminooxy-functionalized PV peptide **13**.

Aminooxy-functionalized $PV_{103-115}$ peptide (13) [CP-III-034]. Resin-bound sequence **S7** was synthesized on 94 mg of FmocThr(*t*Bu)Wang resin (loading = 0.530 mmol/g) via automated MW-SPPS. The functionalized resin was then transferred in a solid-phase peptide synthesis vessel provided with a sintered glass filter, washed with MeOH (5 × 5 mL), and two cycles of swelling (10 min.)/solvent filtration, first with CH_2Cl_2, then with DMF, were performed. Resin-peptide **S7** was suspended in 2.0 mL of DMF, then DIPEA (8.75 μL, 0.05 mmol, 1.0 eq.) and (Boc-aminooxy)acetic acid *N*-hydroxysuccinimidyl ester (57.65 mg, 0.20 mmol, 4.0 eq.) were added, and the suspension was stirred using a stream of N_2. After 2 hours at r.t., the resin was washed with DMF (5 × 5 mL), MeOH (5 × 5 mL) and CH_2Cl_2 (5 × 5 mL), and then transferred to a 50 mL falcon tube and treated with 5 mL of a TFA/50% aq. NH_2OH/TIS/DTT H_2O (90/5/2.5/2.5) cocktail. After for 2 hours of gentle stirring (rocker platform) at r.t., the filtrate was added to ice-cold Et_2O to induce precipitation of crude peptide **13**. Following three cycles of (i) centrifugation (3000 rpm at r.t.), (ii) removal of the supernatant, and (iii) resuspension in ice-cold Et_2O, the precipitate was dried under vacuum, re-suspended in 10 mL of 1% aq. AcOH, and lyophilized to provide 103 mg of crude **13** as a white powder. Purification of the crude peptide was performed via RP-HPLC (5% solv. B for 5 min, then 5-65% over 25 min.), the collected fractions were lyophilized, affording 30.35 mg of aminooxy-functionalized PV peptide **13** ($C_{81}H_{124}N_{18}O_{21}$) as a white foam in a 36% overall yield.

HPLC: t_R = 10.49 min, gradient = 20-100% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+3H]^{3+}$ 562.65, found 562.78; calcd for $[M+2H]^{2+}$ 843.47, found 843.57; calcd for $[M+H]^+$ 1685.93, found 1685.94. **HRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+H]^+$ 1685.9266, found 1685.9308.

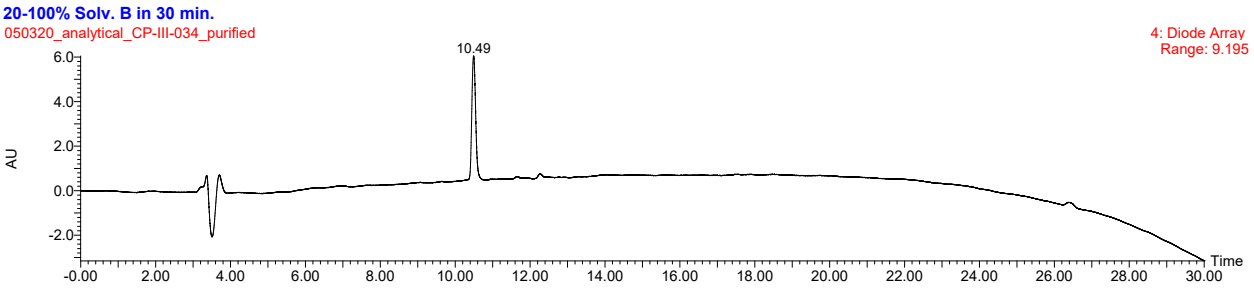


Fig. S32 RP-HPLC chromatogram of **13**.

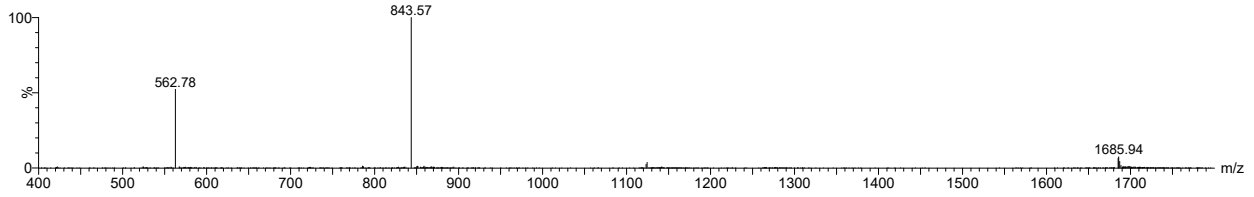


Fig. S33 LRMS (ESI$^+$-MS) spectrum of **13**.

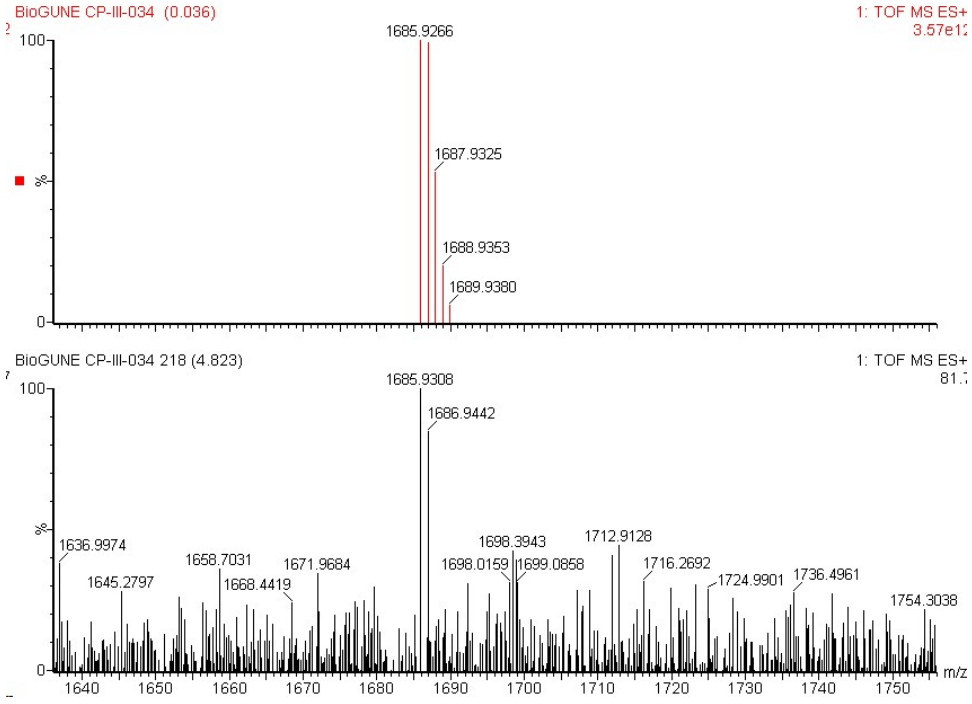
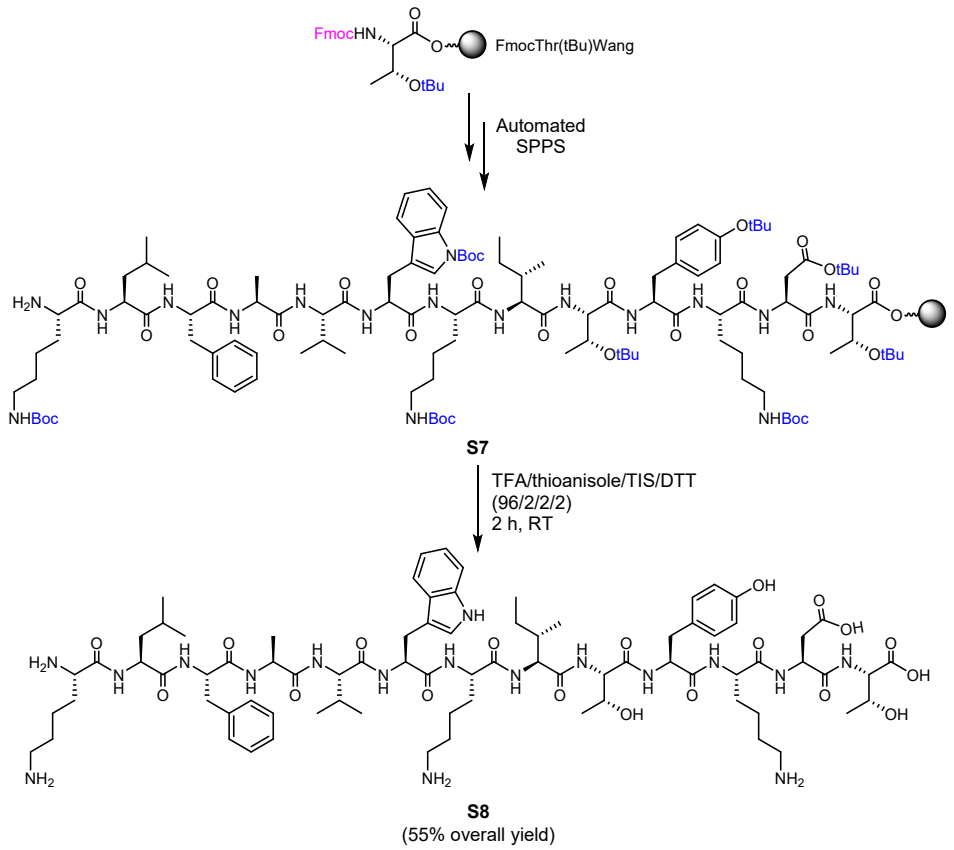


Fig. S34 HRMS (ESI$^+$-MS) spectrum of **13**.

PV peptide S8



Scheme S5 Synthesis of PV peptide **S8**.

$PV_{103-115}$ peptide (S8) [CP-III-042]. Resin-bound sequence **S7** was synthesized on 94 mg of FmocThr(*t*Bu)Wang resin (loading = 0.530 mmol/g) via automated MW-SPPS. Upon sequence completion, the resin was washed with MeOH (5 × 5 mL), transferred in a 50 mL falcon tube and treated with 5 mL of a TFA/thioanisole/TIS/DTT (96/2/2/2) cocktail. After for 2 hours of gentle stirring (rocker platform) at room temperature, the filtrate was added to ice-cold Et_2O to induce precipitation of crude peptide **15**. Following three cycles of (i) centrifugation (3000 rpm at room temperature), (ii) removal of the supernatant, and (iii) resuspension in ice-cold Et_2O, the precipitate was dried under vacuum, re-suspended in 10 mL of 1% aq. AcOH and lyophilized to provide 101 mg of crude peptide as a white powder. Purification of the crude was performed via RP-HPLC (15% solv. B for 5 min, then 15-85% over 25 min.), the collected fractions were lyophilized, affording 44.35 mg of PV peptide **S8** ($C_{79}H_{121}N_{17}O_{19}$) as a white foam in a 55% overall yield.

HPLC: t_R = 9.61 min, gradient = 20-100% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI^+-MS):** *m/z* (monoisotopic) calcd for $[M+3H]^{3+}$ 538.31, found 538.42; calcd for $[M+2H]^{2+}$ 806.96, found 806.81; calcd for $[2M+3H]^{3+}$ 1075.61, found 1075.93; calcd for $[M+H]^+$ 1612.91, found 1612.78. **HRMS (MALDI-TOF):** *m/z* (monoisotopic) calcd for $[M+H]^+$ 1612.9095, found 1612.9175.

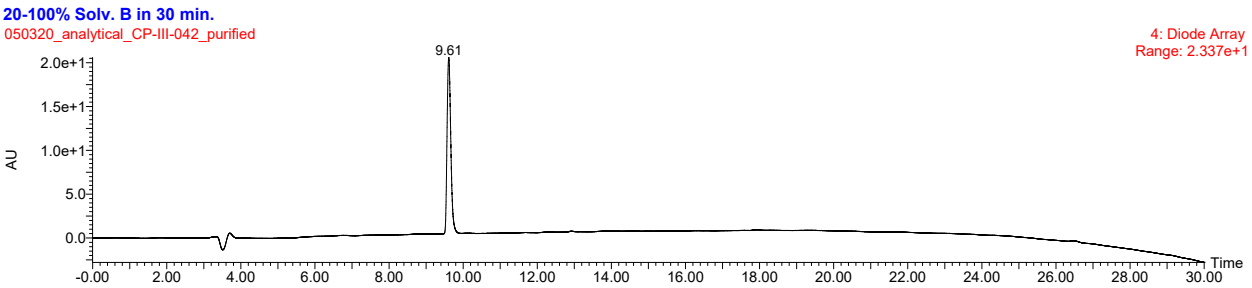


Fig. S35 RP-HPLC chromatogram of **S8**.

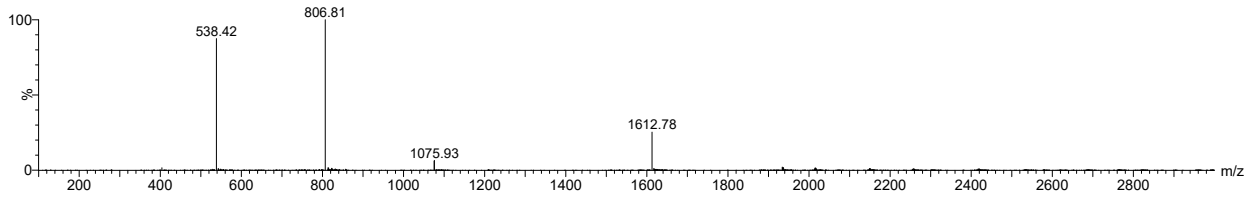


Fig. S36 LRMS (ESI$^+$-MS) spectrum of **S8**.

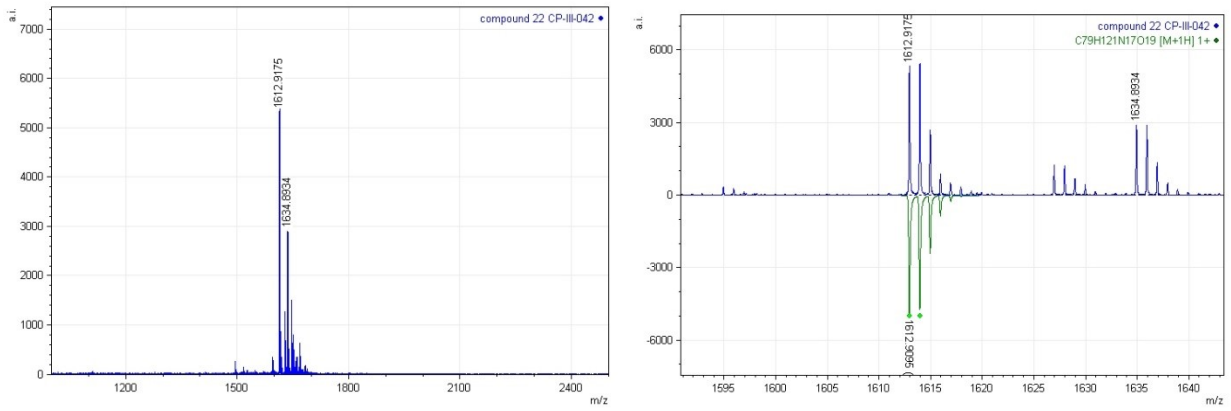


Fig. S37 HRMS (MALDI-TOF) spectrum of **S8**.

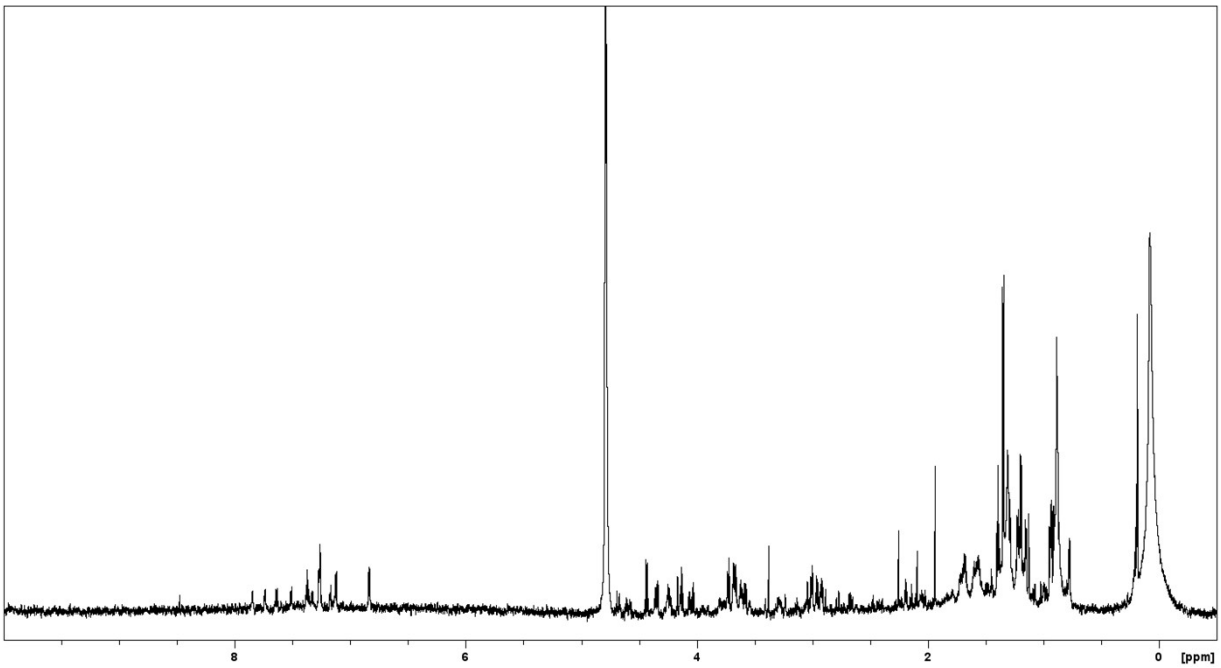
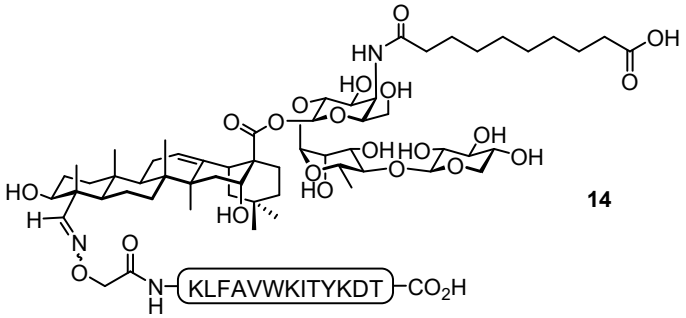


Fig. S38 ^{1}H-NMR spectrum of PV peptide **S8** (30 μM, PBS/D_2O, 298 K, 800 MHz).

Saponin–PV conjugate [QA(PV)] 14



14

Synthetic saponin–$PV_{103\text{-}115}$ conjugate (14) [CP-III-136]. A stirred solution of saponin variant **2** (2.07 mg, 1.82 μmol, 1.0 eq.) and aminooxy-functionalized $PV_{103\text{-}115}$ peptide **13** (6.48 mg, 3.84 μmol, 2.1 eq.) in a H_2O/CH_3CN (1:2) mixture containing 0.05% TFA (final conc. = 1.5 mM with respect to **2**) was heated at 40 ºC for 24 h. The reaction mixture was directly purified via RP-HPLC (20% solv. B for 5 min, then 20-100% over 25 min.). The collected fractions were lyophilized, affording 2.72 mg of saponin–PV conjugate **14** ($C_{140}H_{217}N_{19}O_{40}$) as a white foam in a 53% yield.

HPLC: t_R = 15.80 min, gradient = 20-100% solv. B over 30 min λ_{max} = 200 nm. **LRMS (ESI$^+$-MS):** *m/z* (monoisotopic) calcd for $[M+3H]^{3+}$ 935.86, found 935.64; calcd for $[M+2H]^{2+}$ 1403.28, found 1403.58; calcd for $[2M+3H]^{3+}$ 1870.71, found 1871.08; calcd for $[3M+4H]^{4+}$ 2104.42, found 2105.25; calcd for $[4M+5H]^{5+}$ 2244.65, found 2245.01; calcd for $[M+H]^{+}$ 2805.56, found 2805.39; (most intense peak of

isotopic cluster) calcd for $[M+H]^+$ 2806.56, found 2806.55. **HRMS (MALDI-TOF):** *m/z* (monoisotopic) calcd for $[M+Na]^+$ 2827.5419, found 2827.5482.

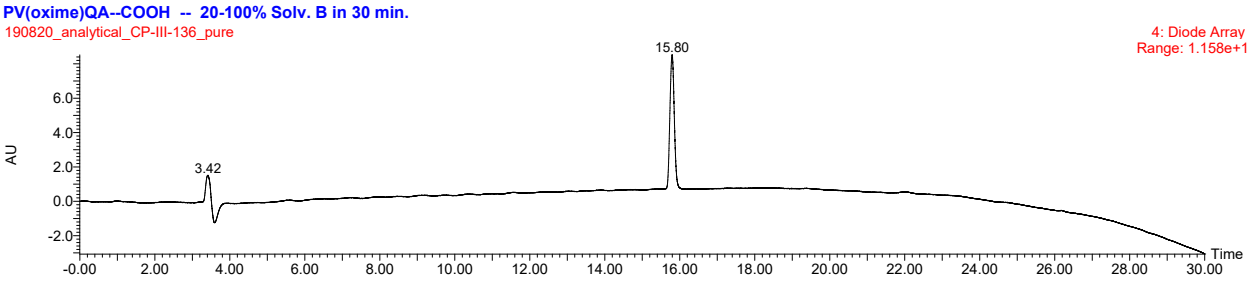


Fig. S39 RP-HPLC chromatogram of **14**.

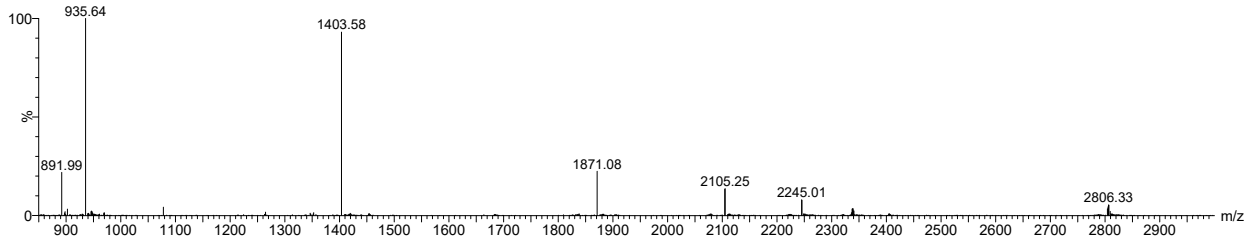


Fig. S40 LRMS (ESI^+-MS) spectrum of **14**.

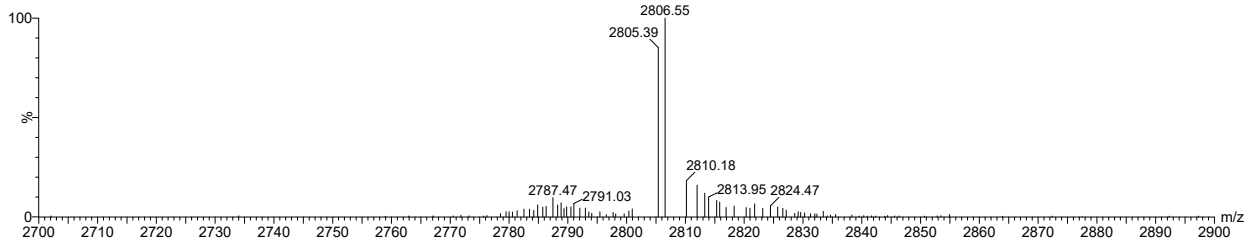


Fig. S41 LRMS (ESI^+-MS) spectrum of **14** zoomed in between 2700 and 2900 *m/z* units.

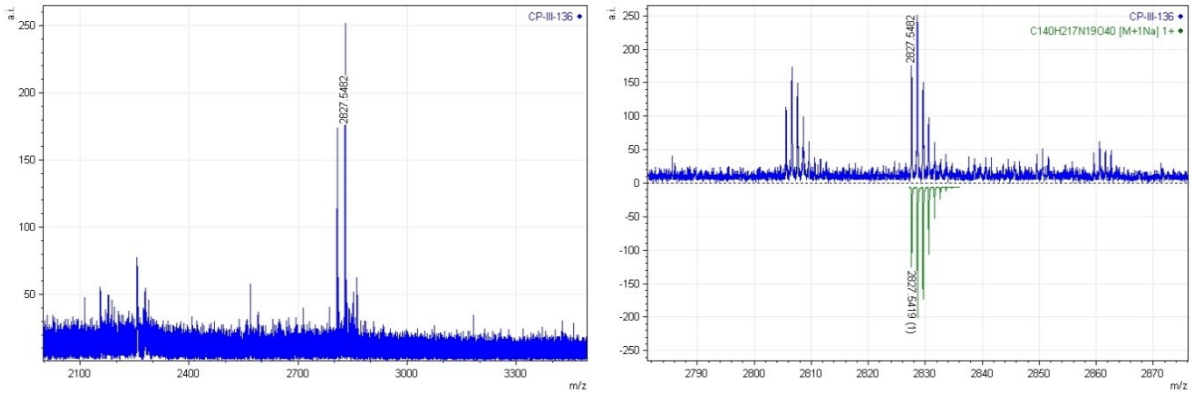
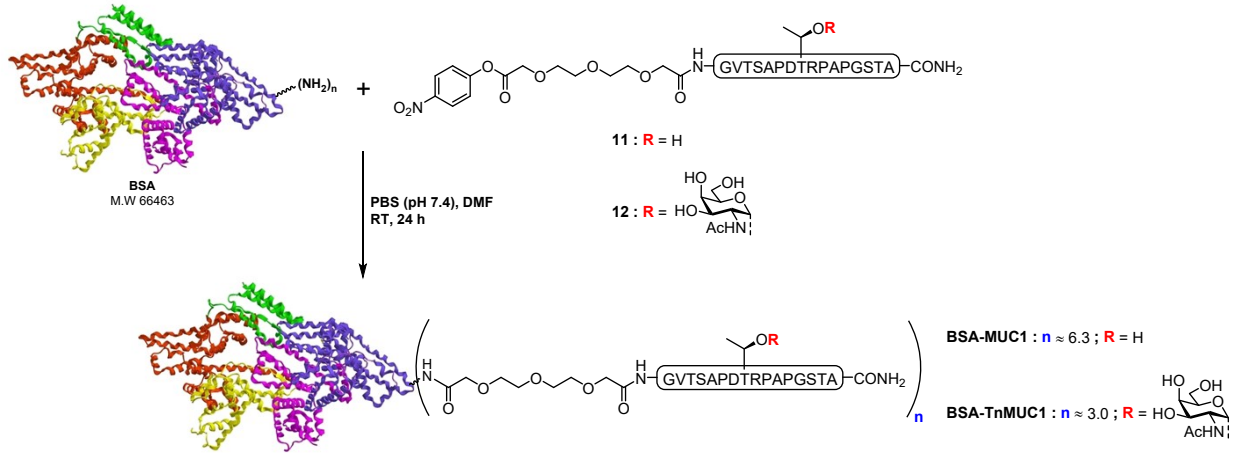


Fig. S42 HRMS (MALDI-TOF) spectrum of **14**.

BSA-conjugates BSA–MUC1 and BSA–TnMUC1



Scheme S6 Synthesis of BSA–MUC1 and BSA–TnMUC1.

BSA–MUC1 and BSA–TnMUC1 conjugates [CP-III-018-020-068]. To a stirred solution of BSA (Fisher Scientific, CAS: 9048-46-8, 1.0 eq.) in PBS (10 mM, pH 7.4) at a 37.5 µM concentration, a solution of activated linker-containing MUC1 peptide **11** or -TnMUC1 glycopeptide **12** (30 eq.) in DMF (200 µL) was added. The reaction mixture was stirred at r.t. and progressively turned yellow. After 24 hours, the reaction was quenched by slowly adding a 0.05% TFA solution in H_2O until disappearance of the yellow coloration, then it was directly purified via RP-HPLC (5% solv. B for 5 min, then 5-70% over 25 min.). The collected fractions were lyophilized, affording the corresponding BSA-conjugates as white foams.

Data for BSA–MUC1: prepared on a 0.030 µmol scale of BSA, yield: 1.85 mg, n ≈ 6.3.

Data for BSA–TnMUC1: prepared on a 0.045 µmol scale of BSA, yield: 2.86 mg, n ≈ 3.0.

"n" was calculated as follows: $n = \frac{MALDI - BSA}{EPITOPE}$; where "MALDI" is the most intense peak observed via MALDI-MS, "BSA" corresponds to the molecular weight of the native protein (66463 Da), and "EPITOPE" corresponds to mass increase of the conjugate for 1 unit of either MUC1 (1669 Da) or TnMUC1 (1872 Da) epitopes.

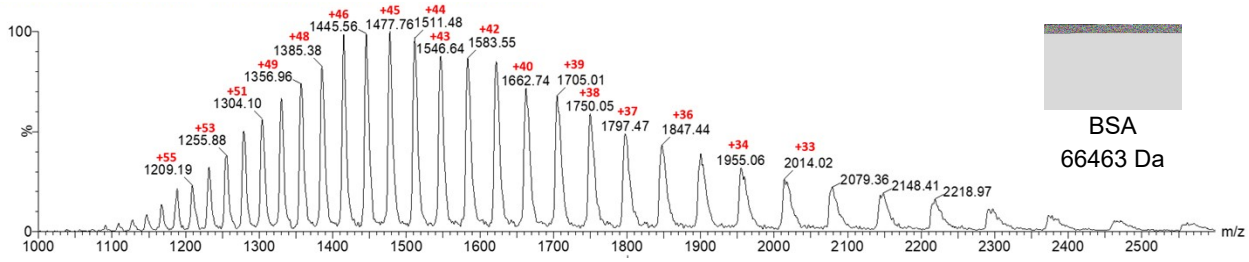


Fig. S43 LRMS (ESI^+-MS) spectrum of commercial BSA.

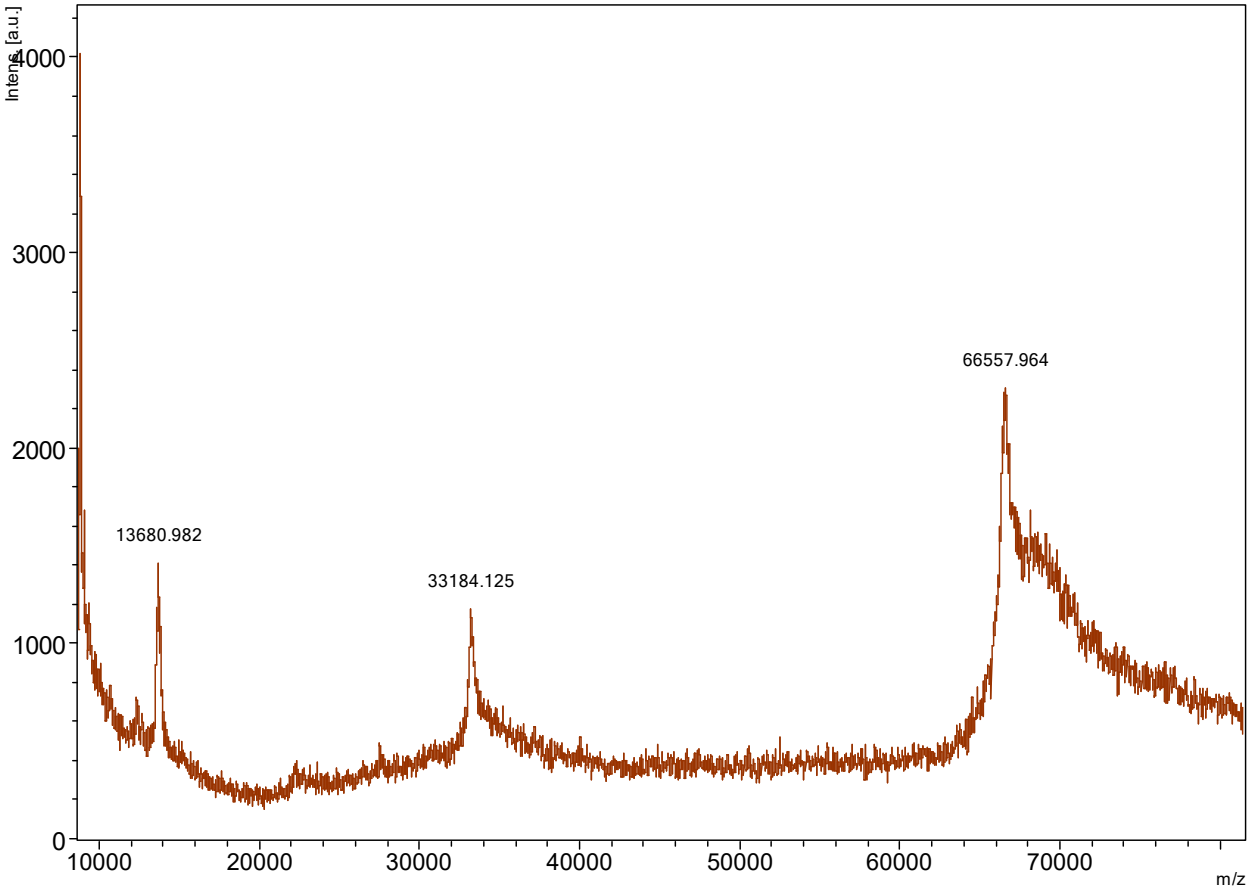


Fig. S44 MALDI-TOF spectrum of commercial BSA.

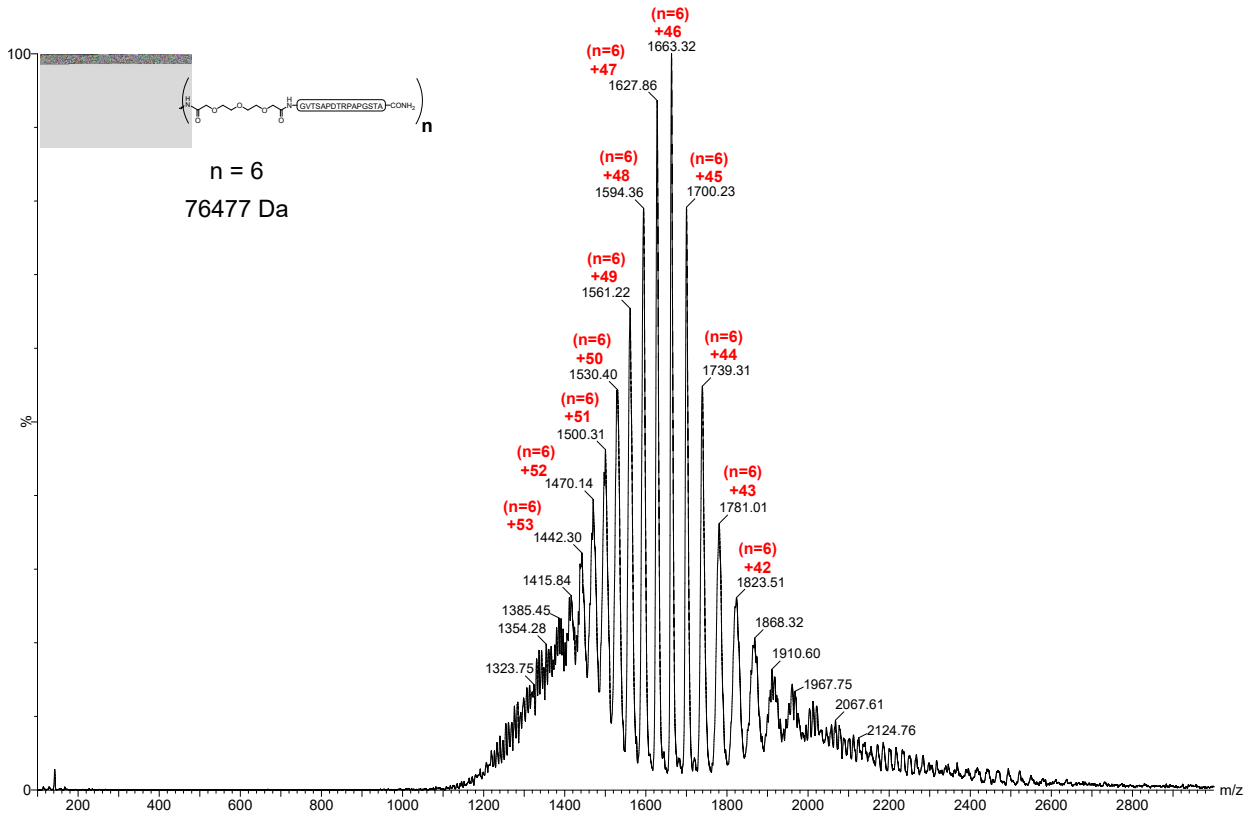


Fig. S45 LRMS (ESI⁺-MS) spectrum of BSA–MUC1.

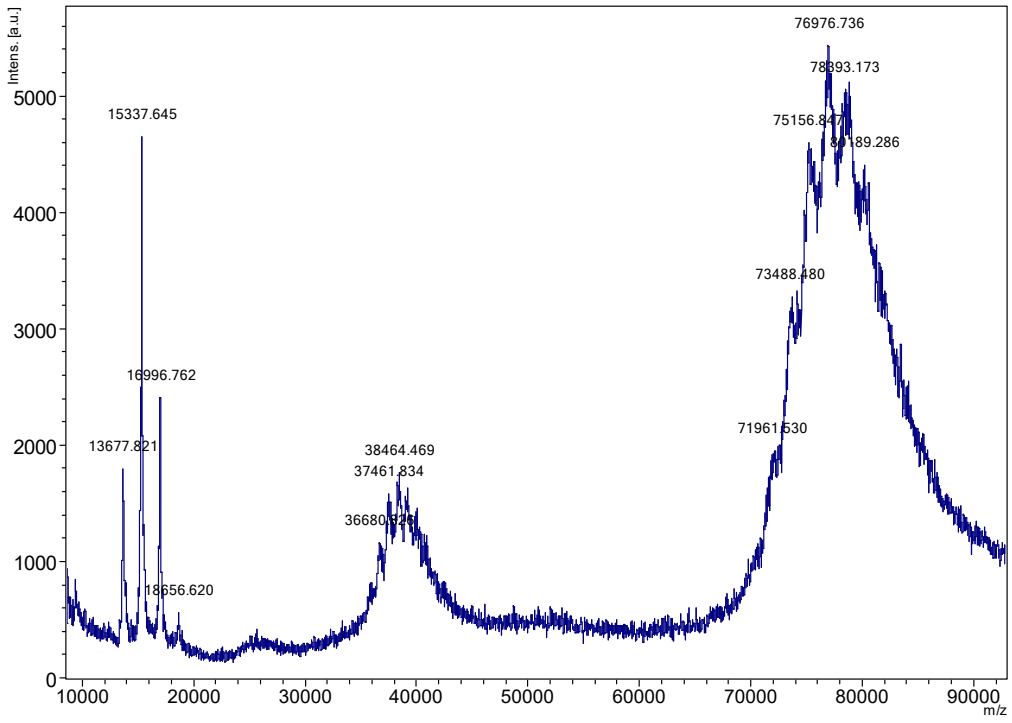


Fig. S46 MALDI-TOF spectrum of BSA–MUC1.

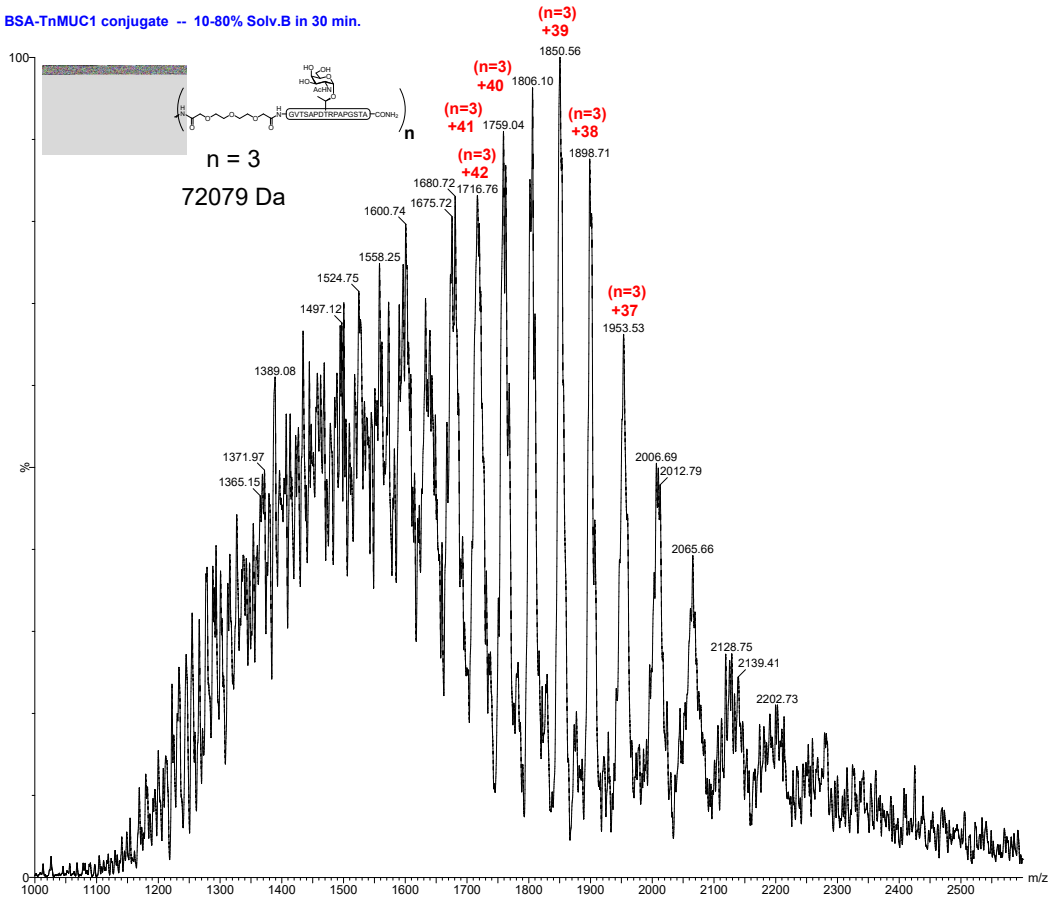


Fig. S47 LRMS (ESI$^+$-MS) spectrum of BSA–TnMUC1.

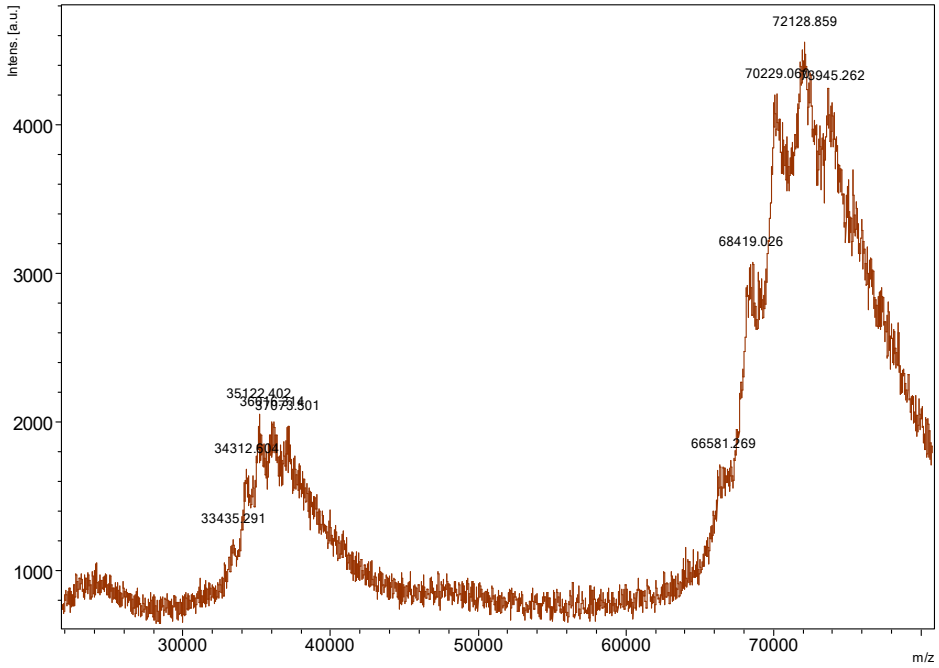


Fig. S48 MALDI-TOF spectrum of BSA–TnMUC1.

III. *IN VIVO* IMMUNOLOGICAL EVALUATION IN MICE

Animals. Mice were cared for and handled in compliance with the Guidelines for Accommodation and Care of Animals (European Convention for the Protection of Vertebrate Animals Used for Experimental and Other Scientific Purposes) and internal guidelines. Mice were housed in standard cages with an automatic water system and fed on a standard diet ad libitum. All the experimental procedures were approved by the appropriate local authorities. The CIC bioGUNE animal facility is fully accredited by AAALAC Intl.

Sample preparation. Samples for each injection were prepared prior to immunization as follows: 1.5-2.5 mg of HPLC-purified, lyophilized compounds (**2**, **4**-**9**, **14**, **S8**) were weighed in a precision balance (Sartorius MSA125P-100-DI Cubis Semi-Micro Balance, S/N 35401101) to prepare stock solutions at ≈ 2 mg/mL conc. by adding sterile PBS (pH 7.4, 10 mM). For each immunization group, formulation aliquots of 600 µL volume per immunization (*i.e.* 3 aliquots for the MUC1 immunizations assay, 2 aliquots for TnMUC1 the immunization assay, according to the number of injections), containing the required amount of the corresponding molecule per mouse (**Table S1-S2**) in 100 µL (= injection volume per mouse), were prepared. The 600 µL aliquots for the prime injections were used the same day, while the boost aliquots were kept at –20 ºC. Prior to boost injections, frozen aliquots were thawed and allowed to passively reach r.t., then mixed vigorously (vortex) before being administered.

- For the MUC1-series immunization assay (see **Fig. 3** in the main text), each mouse was administered 100 µL PBS solutions of equimolar amounts of antigen/construct corresponding to 50 µg of MUC1 peptide **8** [molecular weight (M.W.) = 1483.58, 33.7 nmol]. Accordingly, the amount (dose) of each molecule that was administered is: 54.36 µg PV (**S8**) (M.W. = 1612.91), 91.29 µg QA–MUC1 (**4**) (M.W. = 2708.99), and 147.51 µg QA(PV)–MUC1 (**6**) (M.W. = 4376.94) (**Table S1**).

Table S1. Description of injected samples for MUC1 immunization assay. Amounts of compounds (µg) per mouse, administered in PBS solution (100 µL), are reported.

		Injection volume per mouse = 100 µL			
Group	Description	MUC1 (**8**) (µg)	PV (**S8**) (µg)	QA–MUC1 (**4**) (µg)	QA(PV)–MUC1 (**6**) (µg)
A	MUC1	50.0	-	-	-
B	MUC1 + PV	50.0	54.4	-	-
C	QA–MUC1 + PV	-	54.4	91.3	-
D	QA(PV)–MUC1	-	-	-	147.5

- For the TnMUC1-series immunization assay (see **Fig. 4** in the main text), each mouse was administered 100 µL PBS solutions of equimolar amounts of antigen/construct corresponding to 50 µg of TnMUC1 glycopeptide **9** (M.W. = 1686.78, 29.6 nmol). Accordingly, the amount (dose) of each molecule that was administered is: 47.74 µg PV (**S7**) (M.W. = 1612.91), 78.34 µg QA–TnMUC1 (**5**) (M.W. = 2912.19), 83.07 µg QA(PV) (**14**) (M.W. = 2806.32), 135.57 µg QA(PV) – TnMUC1 (**7**) (M.W. = 4580.13) (**Table S2**).
- The amount of saponin adjuvant "QA" (compound **2**) administered per mouse in groups B and C corresponds to 50 µg, as reported by Ghirardello *et al*. (**Table S2**).[1]

Table S2. Description of injected samples for TnMUC1 immunization assay. Amounts of compounds (μg) per mouse, administered in PBS solution (100 μL), are reported.

		Injection volume per mouse = 100 μL					
Group	Construct(s)	TnMUC1 (**9**) (μg)	QA (**2**) (μg)	PV (**S8**) (μg)	QA–TnMUC1 (**5**) (μg)	QA(PV) (**14**) (μg)	QA(PV)–TnMUC1 (**7**) (μg)
A	PBS	-	-	-	-	-	-
B	QA + TnMUC1	50.0	50.0	-	-	-	-
C	QA + TnMUC1 + PV	50.0	50.0	47.7	-	-	-
D	QA-TnMUC1	-	-	-	78.3	-	-
E	QA-TnMUC1 + PV	-	-	47.7	78.3	-	-
F	QA(PV) + TnMUC1	50.0	-	-	-	83.1	-
G	QA(PV)-TnMUC1	-	-	-	-	-	135.6

Mouse immunization. Groups of five mice (C57BL/6, female, 6-8 weeks old) were administered each immunization sample diluted in PBS (10 mM, 100 μL) via subcutaneous injections following two different vaccination schedules. For the MUC1-series, three times every two weeks (days 0, 14, and 28), and for the TnMUC1-series, two times every two weeks (days 0 and 14). To analyze antibody production over time, mice were bled via the submandibular vein at the indicated pre- and post-vaccination time points, and by cardiac puncture at the time of sacrifice (experimental endpoint). Blood was collected in BD Microtainer® tubes (Clot Activator/SST™ Gel) and centrifuged at 7500g for 10 min, after which the sera were harvested and stored at –20 °C until further analysis.

Quantification of antibody production. Antibody titers against MUC1 or TnMUC1 peptides were measured by an indirect enzyme-linked immunosorbent assay (ELISA). Briefly, ELISA plates (Thermo Scientific) were coated with BSA–MUC1 and BSA–TnMUC1 conjugates or PV peptide **S8** at 0.05 μg/well in carbonate buffer (pH 9.5) and the plates were incubated overnight at 4 °C. After washing the wells (PBS, 10 mM, containing 0.05% Tween 20), plates were blocked with 10% of fetal calf serum (FCS, Biowest) in PBS buffer for 1 h. Serial dilutions of mouse sera in blocking buffer (10% FCS in PBS buffer) were added to wells with appropriate controls and incubated for 1h at room temperature. After washing, goat anti-mouse total IgG (Jackson Immuno Research) antibodies conjugated to horseradish peroxidase (HRP) were added diluted 1/1000 in blocking buffer and incubated for 1 h at room temperature. KPL SureBlue reserve™ commercial solution (100μL/well, SeraCare) containing 3,3',5,5'-tetramethylbenzidine (TMB) was added as peroxidase substrate and after incubation for 10 min, the reaction was stopped with 2N H_2SO_4 (100 μL/well). For absorbance measurements, optical density (OD) at 450 nm was immediately determined using a BioTek® Epoch Microplate Spectrophotometer.

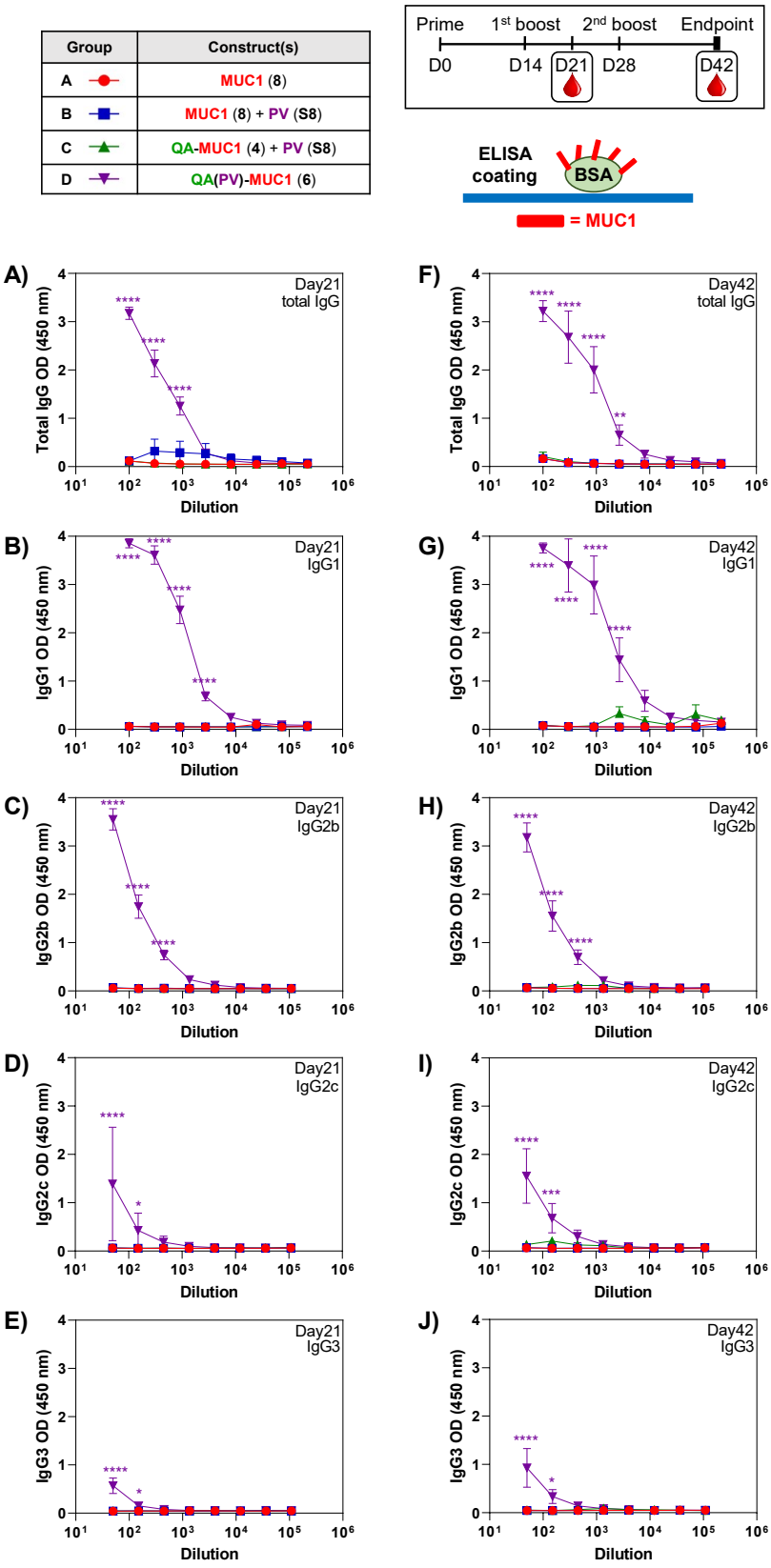


Fig. S49 IgG subtyping of MUC1-series. Day 21 (**A**) total IgG, (**B**) IgG1, (**C**) IgG2b, (**D**) IgG2c, (**E**) IgG3; Day 42 (**F**) total IgG, (**G**) IgG1, (**H**) IgG2b, (**I**) IgG2c, (**J**) IgG3.

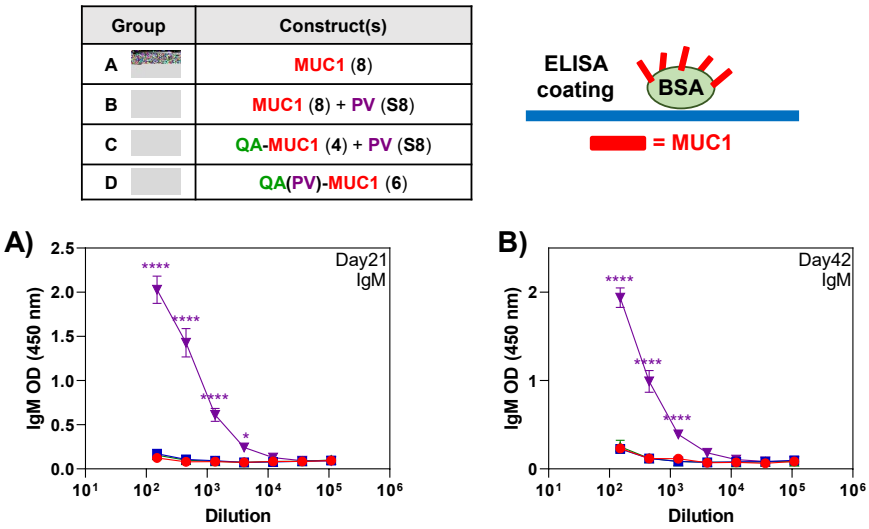


Fig. S50 IgM analysis of MUC1-series. **A)** Day 21 IgM; **B)** Day 42 IgM.

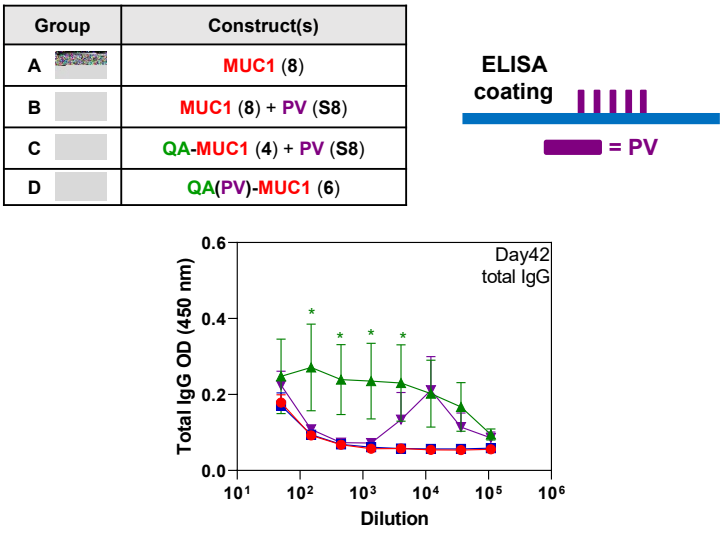


Fig. S51 Day 42 total IgG levels in sera of groups A–D against PV peptide **S8**.

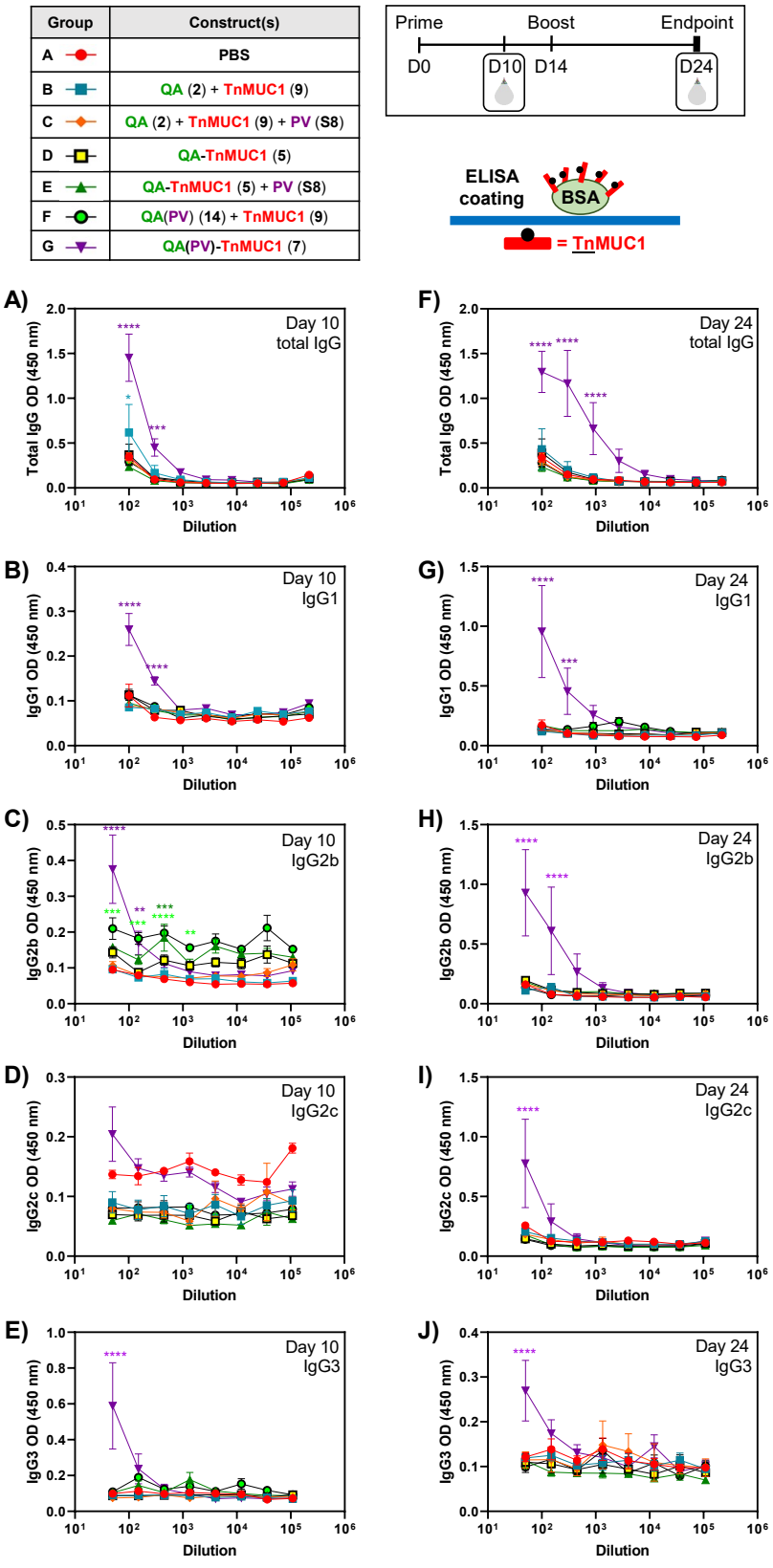


Fig. S52 IgG subtyping of TnMUC1-series. Day 21 total IgG (**A**), IgG1 (**B**), IgG2b (**C**), IgG2c (**D**), IgG3 (**E**); Day 42 total IgG (**F**), IgG1 (**G**), IgG2b (**H**), IgG2c (**I**), IgG3 (**J**).

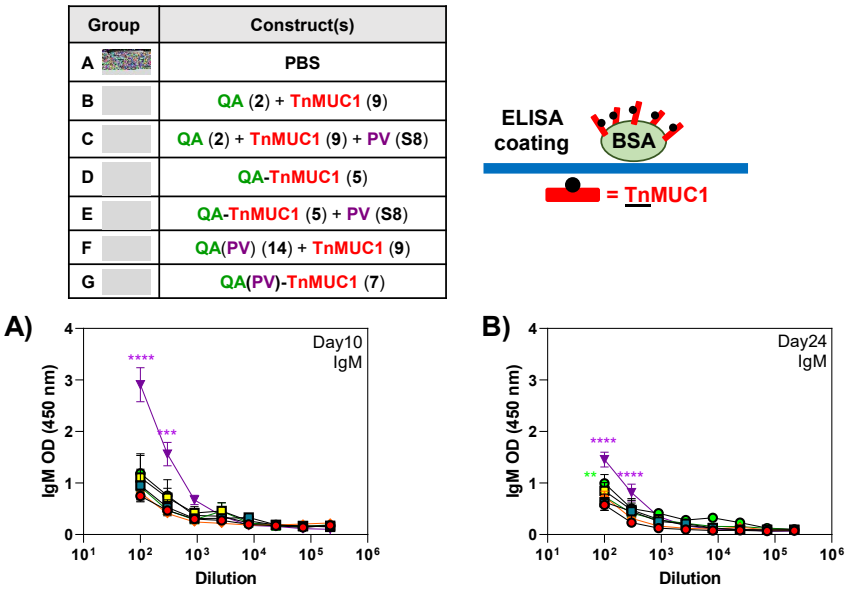


Fig. S53 IgM analysis of TnMUC1-series. **A)** Day 10 IgM; **B)** Day 24 IgM.

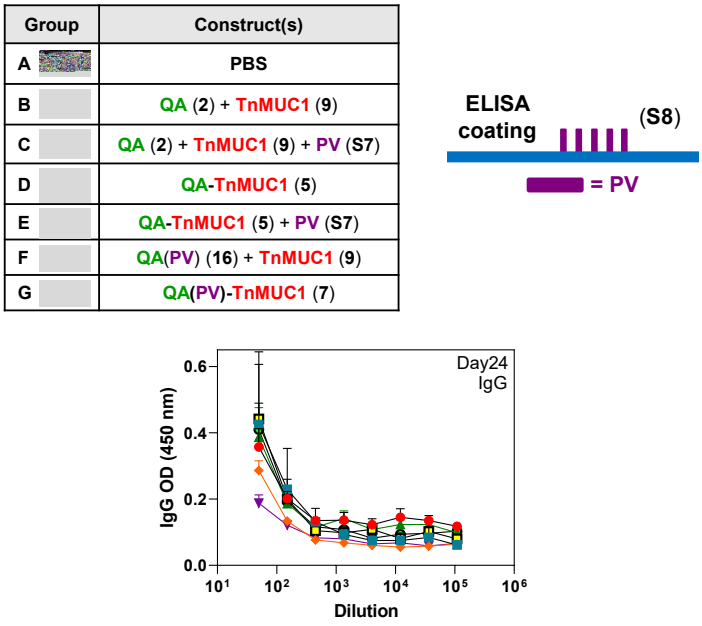


Fig. S54 Day 24 total IgG levels in sera of groups A-G against PV peptide **S8**.

Statistics. ELISA dilution curves were plotted including post-immunization sera from each individual mouse and represented as mean values of five mice; the respective error bars indicate variability between those mice for each group. ELISA plates were coated with a BSA–MUC1/TnMUC1 conjugates (Scheme S6) or the PV peptide **S8** (Scheme S5). For graphs reported in Fig. 3 (article main text), and Fig. S49-S51, statistical significance across the different antibody response curves (OD) at the various dilutions was assessed by comparing to the group administered with the MUC1 construct (**8**) (Group A) using two-way ANOVA analysis with Dunnett's multiple comparisons test. *$p \leq 0.05$; ***$p \leq 0.001$; ****$p \leq 0.0001$. For graphs reported in in Fig. 4 (article main text) and Fig. S52-S54, the same statistical analysis was applied, by comparing to the group administered with PBS (Group A).

Cell-surface reactivity of antisera against MUC1-expressing tumour cells by immunofluorescence. MCF7 cells (and HEK293T cells as negative control) were cultured using DMEM (Dulbecco's Modified Eagle Medium) supplemented with 10% of FBS (Gibco) and 1% of penicillin/streptomycin antibiotic solution (P/S, Gibco). For microscopy assays, MCF7 cells were collected by trypsinization (TrypLE Express, Gibco), seeded on top of sterile round cover glasses in 24-well plates ($2 \cdot 10^5$ cells/well) and allowed to adhere overnight. After washing with PBS (10 mM), the cells were incubated first with blocking buffer (10% FBS in PBS) for 1 h at 37 °C and then with sera from immunized mice at a 1/100 dilution in blocking buffer (500 µL/well) for 2 h at 4 ºC. The cells were extensively washed with PBS buffer and FITC-labeled anti-mouse IgG secondary antibody (Sigma-Aldrich) diluted 1/2000 in blocking buffer was then added (500 µL/well). After incubating the cells for 2 h at 4 °C, the cells were washed again and fixed using a formaldehyde solution (3.7% in PBS, VWR) for 20 min at 4 °C. After this time, DAPI (Invitrogen) was added (5 µg/mL) to stain the nuclei and cells were incubated for 10 min at 4 °C. Finally, after extensive washing, glass coverslips were carefully mounted onto a microscope slide containing the ProLong™ Gold Antifade Mounting reagent (Invitrogen). Slides were kept at 4 °C in the darkness until further analysis. The stained cells were analyzed using a TCS SP8 confocal system (Leica Microsystems). Image analysis was performed using the Leica Application Suite X software (version 3.6).

Cell-surface reactivity of antisera against MUC1-expressing tumour cells by flow cytometry. The reactivity of the antibodies elicited by MUC1 towards breast cancer cell lines was determined by staining the cells with the antisera followed by flow cytometry analysis. MCF7 cells (with high expression of TA-MUC1) and HEK293T cells (control, with no expression of TA-MUC1) were incubated in blocking buffer (10% FBS in PBS) for 1h at room temperature, followed by the mouse sera (1:100 dilution) for 2 h at 4 ºC. After washing, the cells were incubated with an Alexa Fluor 488-conjugated goat anti-mouse IgG H&L secondary antibody (Thermo Fisher, 1:2000) for 2 h at 4 ºC. Acquisition was performed in an Attune NxT flow cytometer (Thermo Fisher). The data were analyzed using FlowJo software (version 10, BD).

IV. STRUCTURAL STUDIES BY NMR: ^{1}H-NMR AND DOSY EXPERIMENTS

^{1}H-NMR experiments. Synthetic di-component and tri-component saponin conjugates **4**–**7** were weighed and transferred to Eppendorfs in which 600 µL of PBS (pH 7.4, 10 mM) had been previously lyophilized. 600 µL of a PBS/D_2O (9:1) mixture were added to each sample, providing the corresponding concentrations of immunization: C_{IMM} = 0.337 mM (for **4** and **6**) [0.55 mg QA–MUC1 (**4**) (M.W. = 2708.99); 0.89 mg QA(PV)–MUC1 (**6**) (M.W. = 4376.94)] and 0.296 mM (for **5** and **7**) [0.47 mg QA–TnMUC1 (**5**) (M.W. = 2912.19); 0.81 mg QA(PV) –TnMUC1 (**7**) (M.W. = 4580.13)]. For each solution at C_{IMM}, an aliquot of 60 µL was taken and then diluted to 600 µL with PBS/D_2O (9:1) to obtain the corresponding $C_{IMM/10}$ sample.

Samples of conjugates **4** –**7** (Figs. S55–S58), MUC1 peptide **8** (Fig. S16), TnMUC1 glycopeptide **9** (Fig. S20) and PV peptide **S8** (Fig. S38) were analyzed in PBS/D_2O (9:1) (500 µL) at 298 K, using residual deuterated solvent signal as chemical shift reference. NMR spectra were recorded on a Bruker AVIII-600 MHz spectrometer equipped with a 5mm PATXI 1 H/D- 13 C/15 N XYZ-GRD probe or on a AVIII-800 MHz spectrometer equipped with a 5mm CPTCI 1 H- 13 C/15 N/D Z-GRD cryoprobe.

DOSY experiments. The diffusion coefficients of the different saponin-peptide solutions were measured in PBS/D_2O solutions, at the indicated concentrations at 298 K, in a Bruker AV-III 800 MHz NMR spectrometer equipped with a cryoprobe with Z gradients. The standard Bruker pulse sequence ledbpgp2s, with longitudinal eddy current delay (5ms) and bipolar gradient pulses, was applied, using 500 ms of diffusion time Δ (d20) and 4 ms of diffusion gradient length δ (2 x p30). The experiments were acquired with 24 gradient increments with increasing gradient strength. The data were analyzed with TOPSPIN 4.1.4 software to obtain both the 2D-DOSY spectra (see Fig.s 7 and S71) and the diffusion coefficient values using the T1/T2 module included in the software. The final value indicated in the table S3 for each compound was calculated from the mean of the values obtained for different peaks along the NMR spectra.

From the Stokes-Einstein Equation (1), it is possible to correlate the diffusion coefficient to the hydrodynamic radii of two molecules, and thus with the corresponding hydrodynamic volumes:

$$D = \frac{k_B.T}{6\pi\eta R} \quad (1)$$

$$\frac{D_d}{D} = \frac{R}{R_d} \qquad \Rightarrow \qquad \frac{V}{V_d} = \frac{R^3}{R_d^3}$$

Since these molecules form concentration-dependent aggregates in size that are also in dynamic equilibrium with monomers or smaller size aggregates, the experimentally measured diffusion coefficients can be described as the sum of the value of the coefficient for every aggregate modulated by their molar fraction. These averaged values and the derived averaged hydrodynamic parameters provide an idea of the averaged global size of the molecules present in each solution, which can be used for comparison purposes among them.

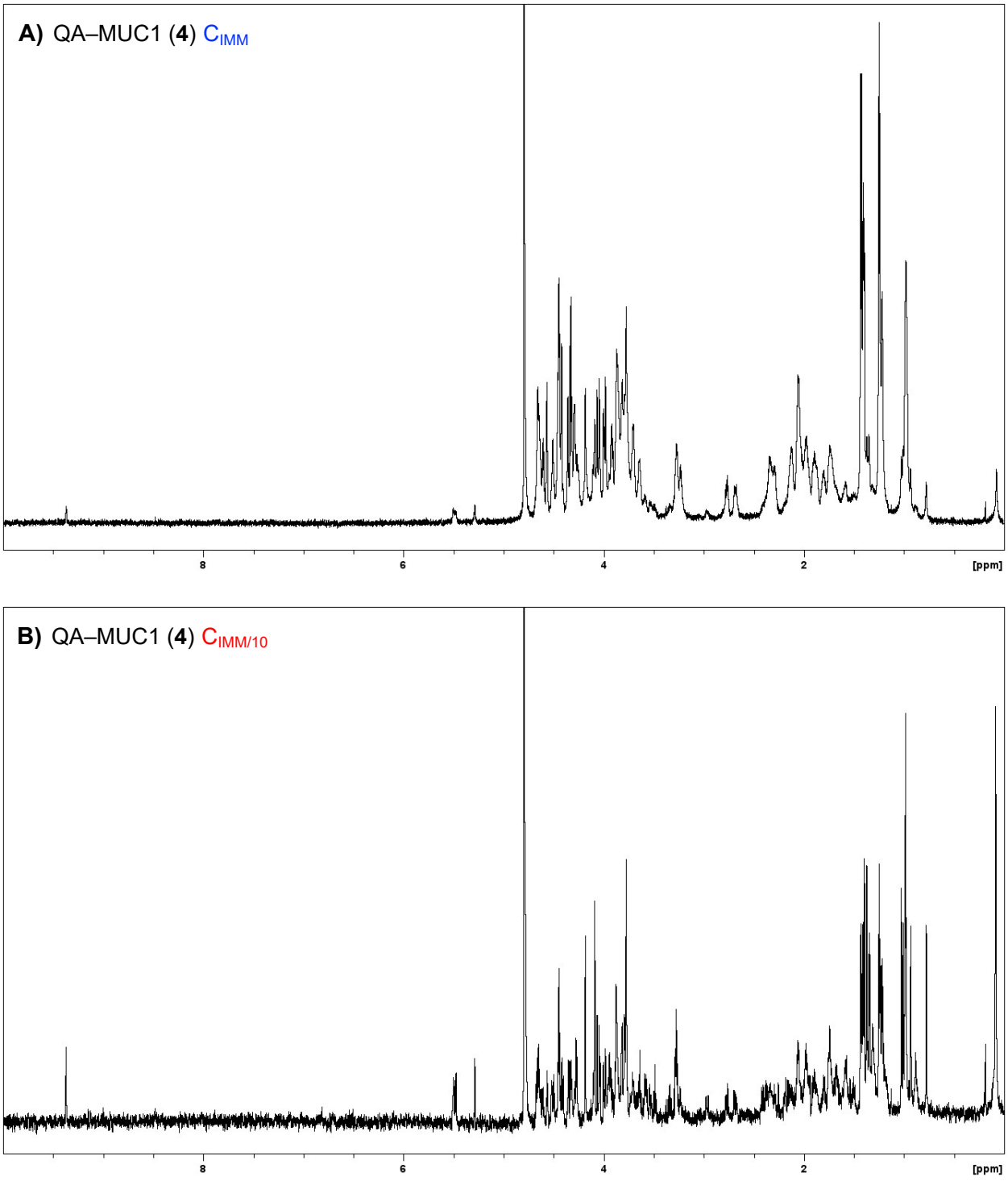


Fig. S55 ^{1}H-NMR spectra (800 MHz) of QA–MUC1 **4** (PBS/D_2O) at (**A**) the immunization concentration (C_{IMM} = 0.337 mM) and (**B**) ten-fold diluted sample ($C_{IMM/10}$ = 0.0337 mM).

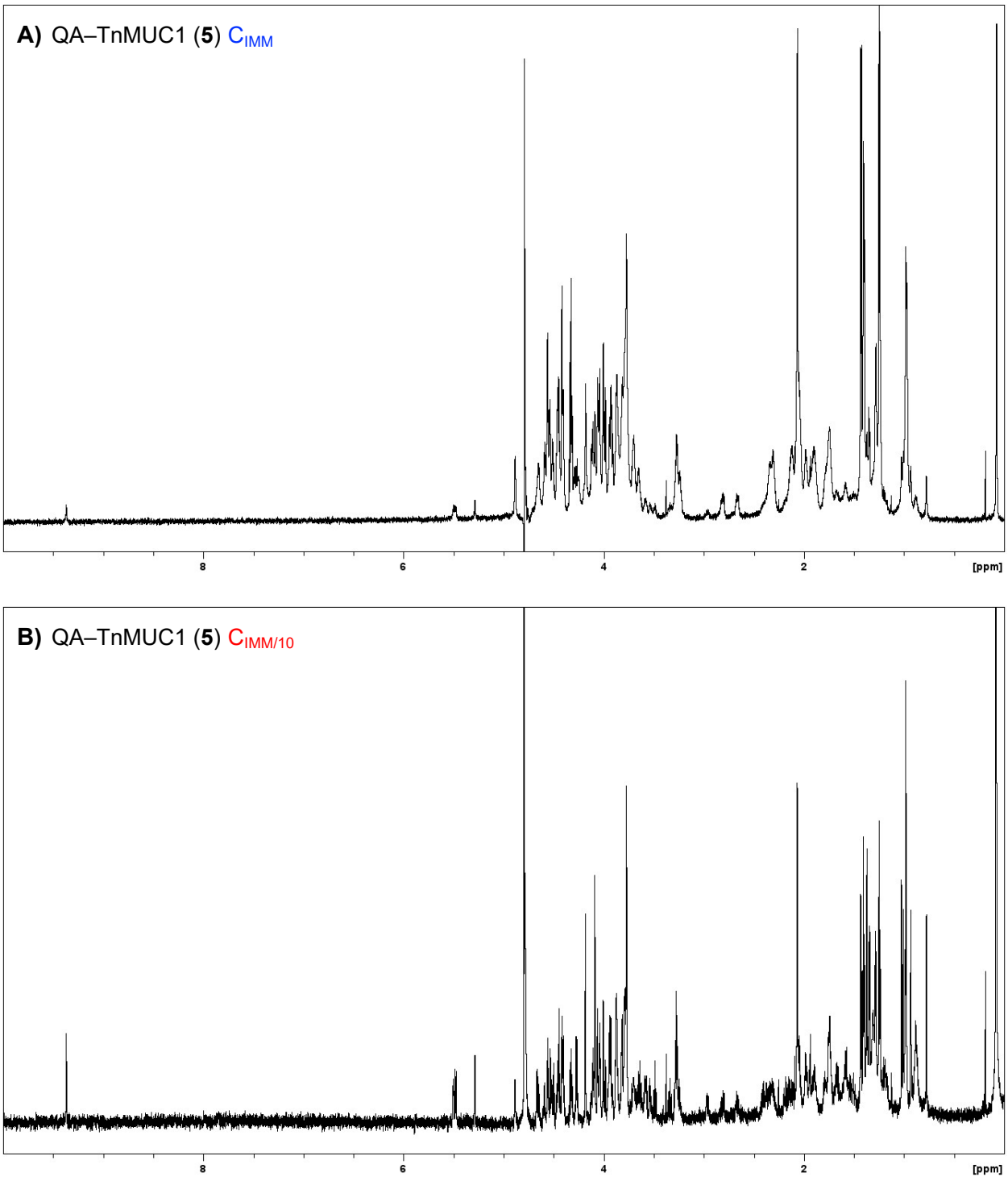


Fig. S56 ^{1}H-NMR spectra (800 MHz) of QA–TnMUC1 **5** (PBS/D_2O) at (**A**) the immunization concentration (C_{IMM} = 0.296 mM) and (**B**) ten-fold diluted sample ($C_{IMM/10}$ = 0.0296 mM).

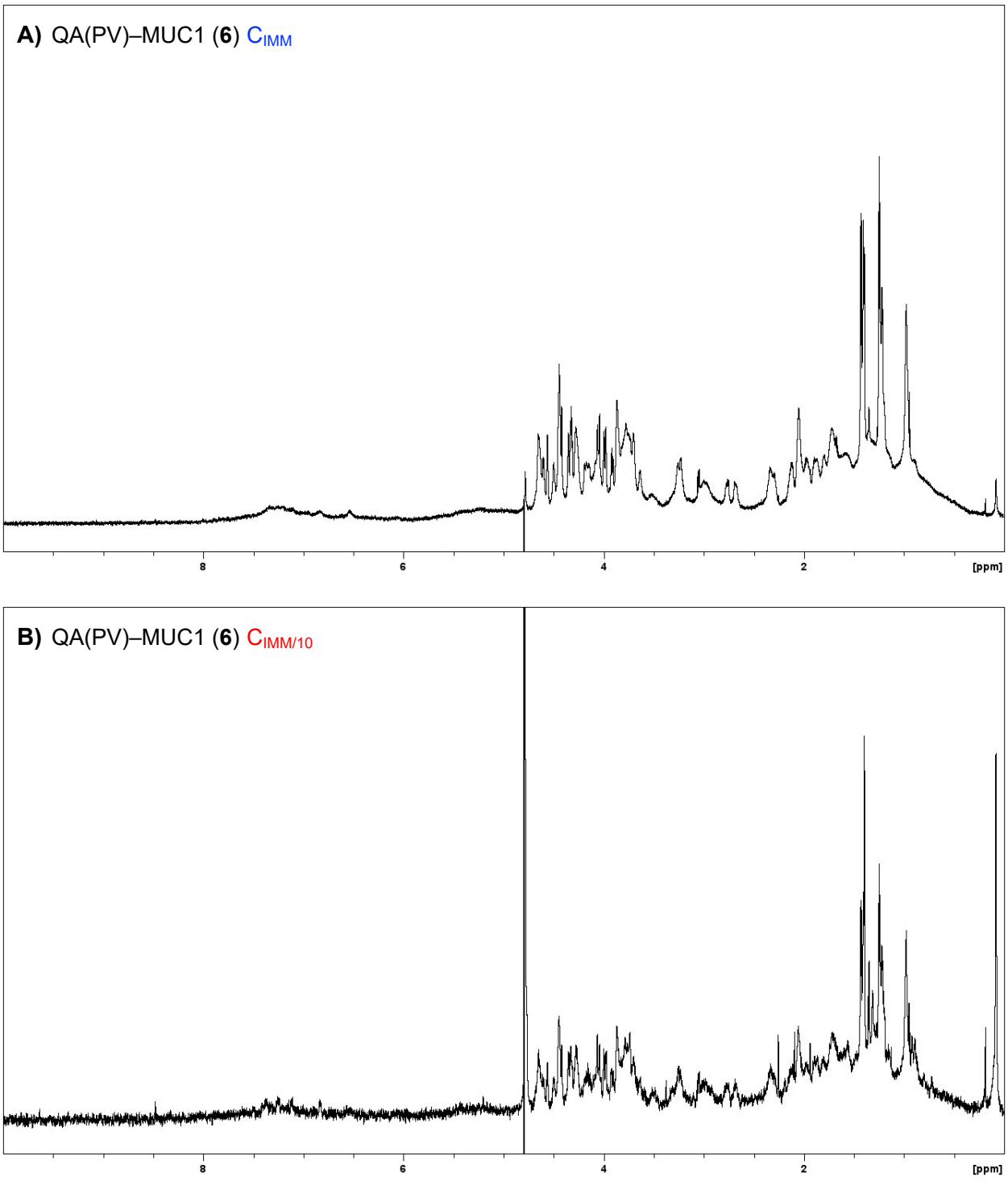


Fig. S57 ^{1}H-NMR spectra (800 MHz) of QA(PV)–MUC1 **6** (PBS/D_2O) at (**A**) the immunization concentration (C_{IMM} = 0.337 mM) and (**B**) ten-fold diluted sample ($C_{IMM/10}$ = 0.0337 mM).

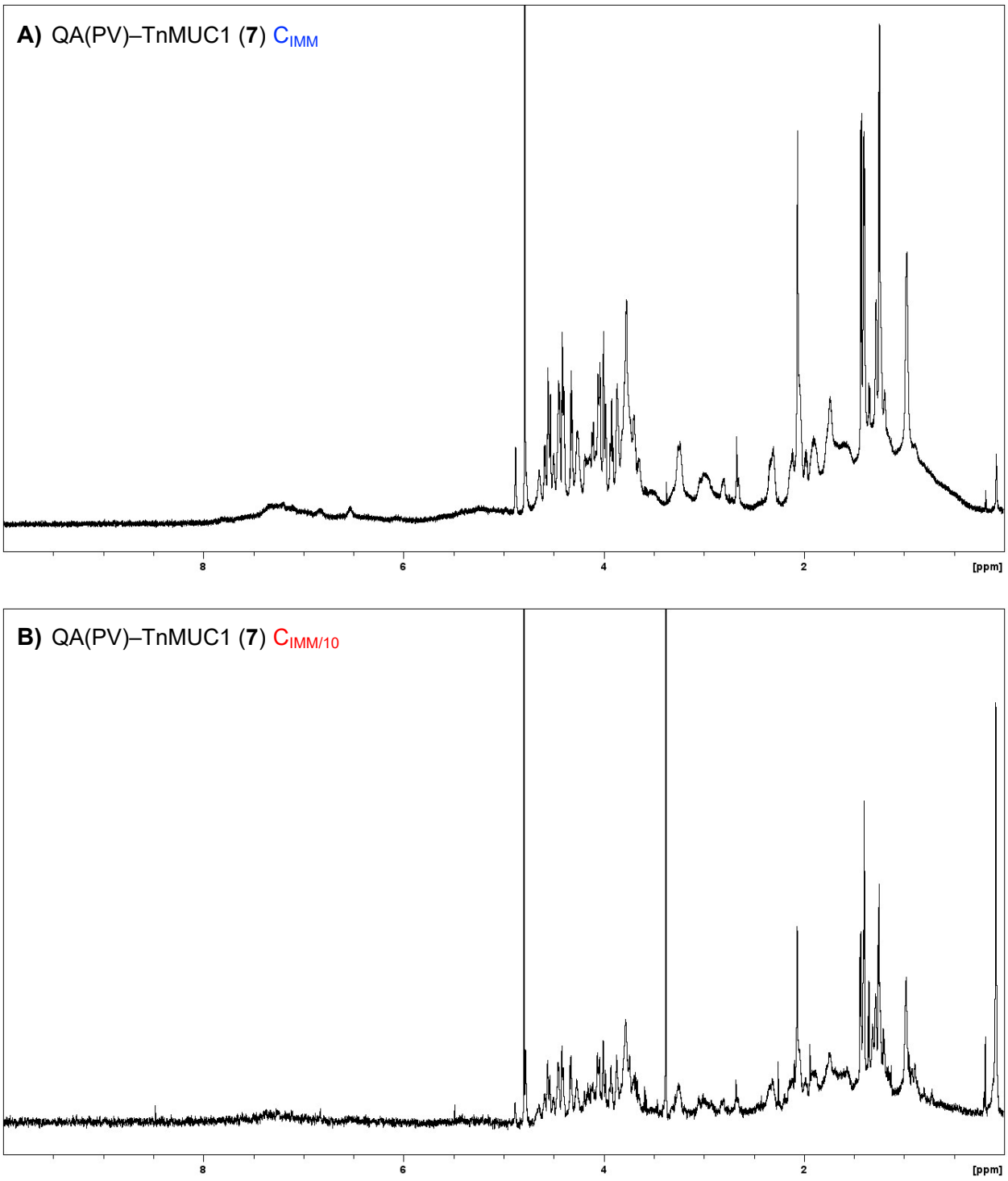


Fig. S58 ^{1}H-NMR spectra (800 MHz) of QA(PV)–TnMUC1 **7** (PBS/D_2O) at (**A**) the immunization concentration (C_{IMM} = 0.296 mM) and (**B**) ten-fold diluted sample ($C_{IMM/10}$ = 0.0296 mM).

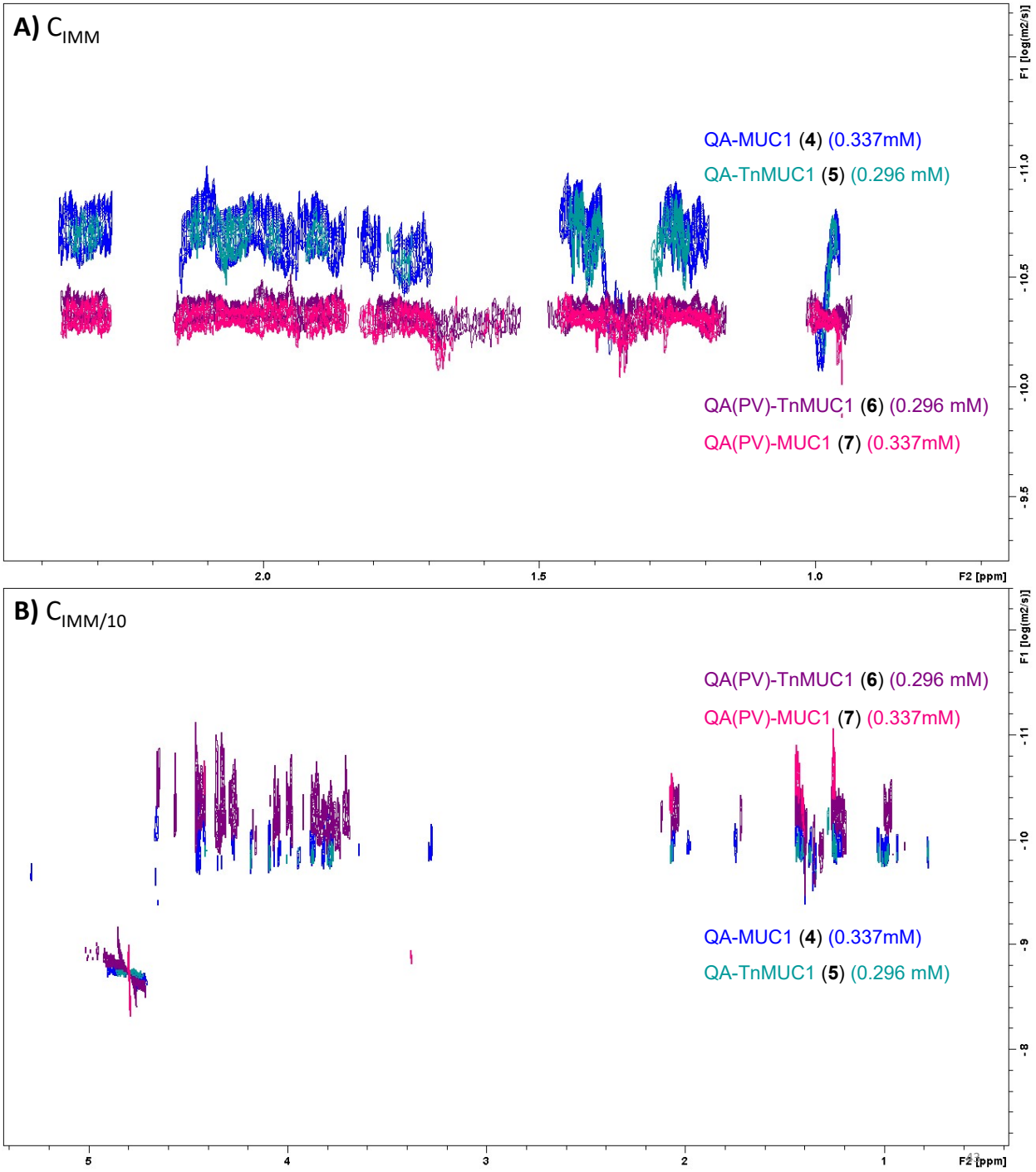


Fig. S59 Superimposition of the DOSY-NMR spectra acquired for compounds QA–MUC1 **4**, QA–TnMUC1 **5,** QA(PV)–MUC1 **6** and QA(PV)–TnMUC1 **7** at (**A**) the immunization concentration, C_{IMM} and (**B**) ten-fold diluted sample, $C_{IMM/10}$.

Table S3. Diffusion coefficients **(a)** and hydrodynamic parameters of the different saponin conjugate solutions estimated from DOSY measurements at the two concentrations, i.e. concentration of immunization and 1/10 dilution (≈0.3 mM, and ≈0.03 mM, respectively) (**Fig. S59**). The rows marked with "_d" refer to the diluted conditions. **(b)** Ratios of the averaged hydrodynamic radius derived from the diffusion coefficients values D, according to the Stokes-Einstein equation. These values are related to the relative size of the aggregates that are formed at the higher concentrations compared to those found at lower concentrations. **(c)** The hydrodynamic volumes are related to the hydrodynamic radius through a cubic relationship. The corresponding ratios, related to the relative volumes, are given here as well. For instance, for QA–TnMUC1 (**5**), the relative average hydrodynamic volume between the immunization concentration and the 10-fold diluted conditions is 274. **(d)** Ratio of the averaged hydrodynamic radius related to that measured for ***QA-TnMUC1_d*** (diluted), which shows the highest diffusion coefficient, and thus the smaller averaged size. **(e)** Ratios for the corresponding averaged hydrodynamic volumes. From these data, it is possible to deduce that the tricomponent molecules QA(PV)–MUC1 **6** and QA(PV)–TnMUC1 **7** are basically always aggregated in this range of concentrations. On the other hand, the di-component compounds QA–MUC1 **4** and QA–TnMUC1 **5** generate larger size aggregates at high concentrations that disaggregate upon 10-fold dilution.

Entry	D (m2/s) (a)	r /r_d (b)	V /V_d (c)	r /$r_{QA\text{-}TnMUC1\_d}$ (d)	V /$V_{QA\text{-}TnMUC1\_d}$ (e)
QA-MUC1 (**4**)	2,01E-11			6,5	276,7
QA-MUC1_d	7,97E-11	4.0	64	1,6	4,4
QA-TnMUC1 (**5**)	2,02E-11			6,5	274
QA-TnMUC1_d	***1,31E-10***	***6.5***	***274***	***1,0***	***1,0***
QA(PV)-MUC1 (**6**)	4,70E-11			2,8	21,6
QA(PV)-MUC1_d	4,97E-11	1.1	1.2	2,6	18,3
QA(PV)-TnMUC1 (**7**)	4,43E-11			3,0	25,9
QA(PV)-TnMUC1_d	5,10E-11	1.2	1,5	2,6	17,0

V. SUPPORTING INFORMATION REFERENCES

1 M. Ghirardello, A. Ruiz-De-Angulo, N. Sacristan, D. Barriales, J. Jiménez-Barbero, A. Poveda, F. Corzana, J. Anguita and A. Fernández-Tejada, *Chem. Commun.*, 2020, **56**, 719–722.

2 A. Fernández-Tejada, E. K. Chea, C. George, N. Pillarsetty, J. R. Gardner, P. O. Livingston, G. Ragupathi, J. S. Lewis, D. S. Tan and D. Y. Gin, *Nat. Chem.*, 2014, **6**, 635–643.

3 Z. Wu, X. Guo and Z. Guo, *Chem. Commun.*, 2010, **46**, 5773–5774.

4 S. Ingale, M. A. Wolfert, J. Gaekwad, T. Buskas and G. J. Boons, *Nat. Chem. Biol.*, 2007, **3**, 663–667.

5 Z. Guo and Q. Wang, *Curr. Opin. Chem. Biol.*, 2009, **13**, 608–617.

6 R. J. Payne, S. Ficht, S. Tang, A. Brik, Y. Y. Yang, D. A. Case and C. H. Wong, *J. Am. Chem. Soc.*, 2007, **129**, 13527–13536.

7 V. Wittmann, S. Takayama, K. W. Gong, G. Weitz-Schmidt and C. H. Wong, *J. Org. Chem.*, 1998, **63**, 5137–5143.