Design, synthesis and immunological evaluation of CRM_{197}-based immunogens functionalized with synthetic scaffolds displaying a tumor-associated MUC1 glycopeptide

Carlo Pifferi,*[a]‡ David Goyard,[d] Leire Aguinagalde,[a] Olivier Renaudet*[d], Juan Anguita*[bc] and Alberto Fernández-Tejada*[ac]

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

[b] *Inflammation and Macrophage Plasticity Laboratory, CIC bioGUNE, Basque Research and Technology Alliance (BRTA), Biscay Technology Park, Building 801A, 48160 Derio, Spain.*

[c] *Ikerbasque, Basque Foundation for Science, Bilbao, Spain.*

[d] *Univ. Grenoble Alpes, CNRS, DCM UMR 5250, F-38000 Grenoble, France.*

‡ *C. P. present address: Centre de Biophysique Moléculaire, CNRS UPR 4301, Rue Charles Sadron, 45071 Orléans Cedex 2, France.*

*Corresponding authors:
carlo.pifferi@cnrs-orleans.fr
olivier.renaudet@univ-grenoble-alpes.fr
janguita@cicbiogune.es
afernandeztejada@cicbiogune.es

Electronic Supplementary Information (ESI†)

Table of contents:

1- 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.

Reverse-phase (RP)-HPLC analyses (with or without tandem mass spectrometry) and purifications were carried out either 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; on a Waters Alliance 2695 s equipped with a Waters 2489 UV/visible detector (Solv. A = 0.1 % TFA in H_2O; Solv. B = 0.1 % TFA in 90 % aq. CH_3CN); on a Waters Acquity UPLC-MS equipped with a SQD2 detector (Solv. A = 0.1 % FA in H_2O; Solv. B = 0.1 % FA in 90 % aq. CH_3CN).

HR-MS spectra were recorded at ICMG (Institut de Chimie Moléculaire de Grenoble) mass spectrometry (MS) platform, on a Waters Xevo G2-S QTof. MALDI-TOF spectra were recorded by ICMG MS platform on an AutoFlex I Bruker after sample pre-treatment in an OligoR3 microcolumn (Applied Biosystems, USA) using a 2,5-dihydroxybenzoic acid matrix.

MALDI-TOF MS spectra were recorded at CIC biomaGUNE mass spectrometry platform, 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). 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.

2- General synthetic procedures

2-1 Solid-Phase Peptide Synthesis (SPPS)

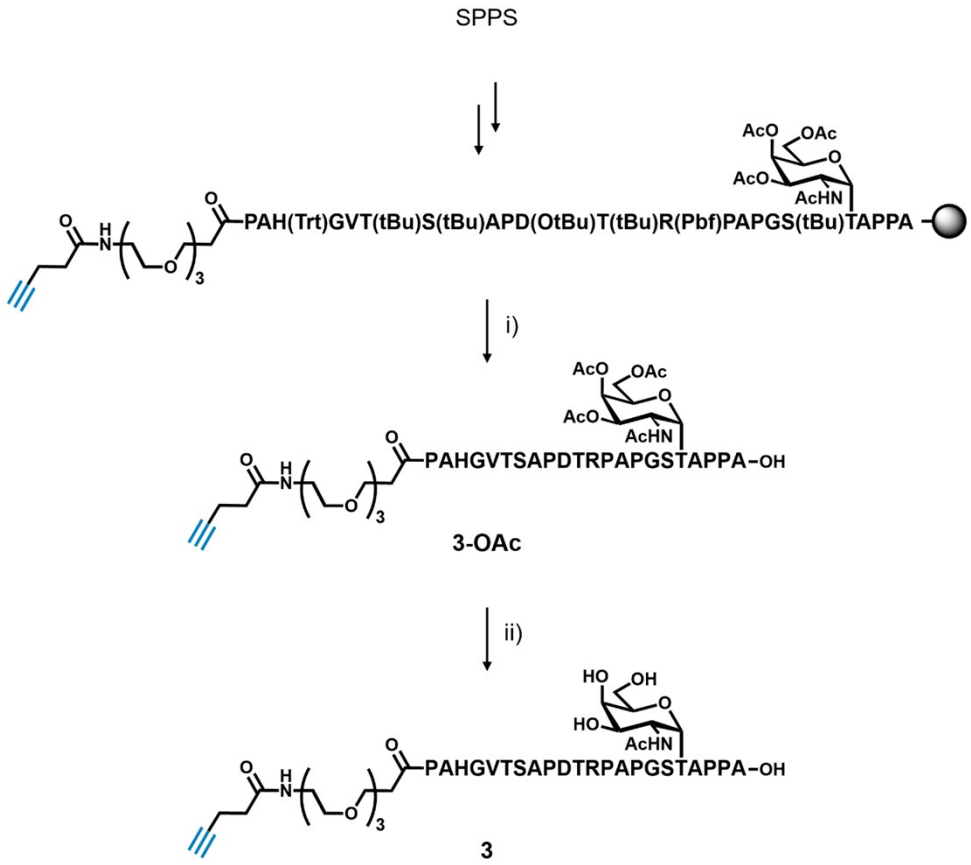
Peptide elongation was performed manually or in an automated fashion (Syro II automated parallel peptide synthesizer) by solid-phase peptide synthesis (SPPS) using the standard 9-fluorenylmethoxycarbonyl (Fmoc)/ *tert*-butyloxycarbonyl (*t*Bu) protection strategy. In manual SPPS, the device consisted of a polypropylene syringe-shaped reactor (150 mL volume) (#5147808 Grace SA, USA), equipped with a filter and a valve in its lower end. The side-chain protecting groups used were Arg(Pbf), Asp(O*t*Bu), Glu(O*t*Bu), His(Trt), Lys(Boc), Ser(*t*Bu), and Thr(*t*Bu).

Fmoc-deprotection consisted in treatments of 20 % piperidine in DMF (3 × 10 min), followed by several washes with DMF (5 × 10 min.). Couplings were performed using 1.5–2.0 equiv. of the $^{\alpha}$*N*-Fmoc-protected amino acids activated *in situ* with 1.5–2.0 equiv. of PyBOP® and 3.0–4.0 equiv. of DIPEA in DMF (10 mL per gram of resin) for 30 min at room temperature.

Automated SPPS protocols featured a 5-fold excess of reagents in the coupling step compared to the manual protocol, and included an additional capping step after each coupling.

3- Synthesis and characterization of antigen-presentation units 1 and 2

3-1 MUC1 glycopeptide 3



Scheme S1. Synthesis of alkyne-terminating MUC1 glycopeptide **3**.

(SPPS) The solid-supported peptide was synthesized according to general synthetic procedure 2-1-1, starting from 2-chlorotrityl resin (1.158 g, 0.338 mmol). The synthesis was performed manually for the first five coupling reactions (*N*-terminal amino acid αGalNAc-*O*-threonine). The rest of the synthesis was performed using automated peptide synthesis.

(i) The glycopeptide was cleaved from the resin using a solution of TFA/triisopropylsilane/H_2O (98:1:1), the resin was stirred for 1 h, filtered, and further washed twice with 5 mL of TFA for 5 min. The filtrate was concentrated under reduced pressure, and TFA co-evaporated twice with toluene. The crude product **3-OAc** was solubilized with water and purified by semi preparative RP-HPLC (0-30% Solv. B over 15 min.). Fractions containing the product were combined and lyophilized to afford pure **3OAc** as a white powder (496 mg, 0.186 mmol, 55%). HRMS (ESI$^+$-TOF) *m/z*: calcd. for $C_{116}H_{180}N_{29}O_{43}$ $[M+H]^+$ (monoisotopic): 2667.2790, found 2667.2776; RP-HPLC: R*t* = 7.32 min (C18, λ = 214 nm, 5-60% Solv. B over 15 min.)

(ii) To a solution of compound **3OAc** (50.0 mg, 18.7 µmol) in MeOH (5 mL), a solution of sodium methoxide 1 M in methanol (150 µL) was added. The mixture was stirred at r.t. for overnight, after which RP-UHPLC showed complete conversion into the desired, fully-deprotected compound. The mixture was neutralized with AcOH, diluted with H_2O (5 mL) and purified by semi preparative RP-HPLC

(0-34% Solv. B over 15 min.). Fractions containing the product were combined and lyophilized to afford pure **3** as a white powder (41.0 mg, 16.1 µmol, 86%). HRMS (ESI$^+$-TOF) *m/z*: calcd. for $C_{110}H_{174}N_{29}O_{40}$ $[M+H]^+$ (monoisotopic): 2541.2473, found 2541.2505; RP-HPLC: R*t* = 6.02 min (C18, λ = 214 nm, 5-60% B in 15 min.).

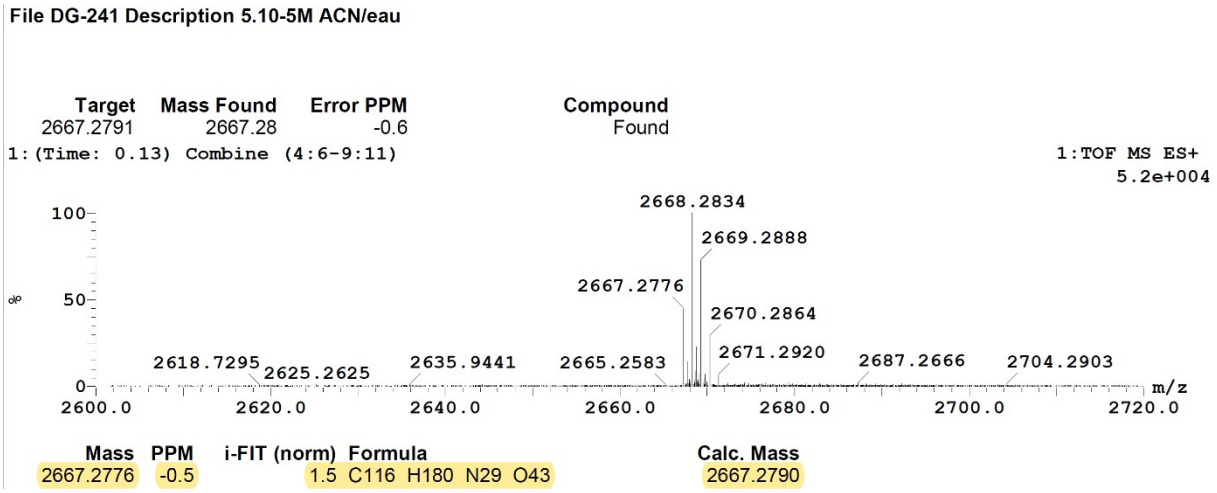


Figure S1. HRMS(+) spectrum of **3-OAc**.

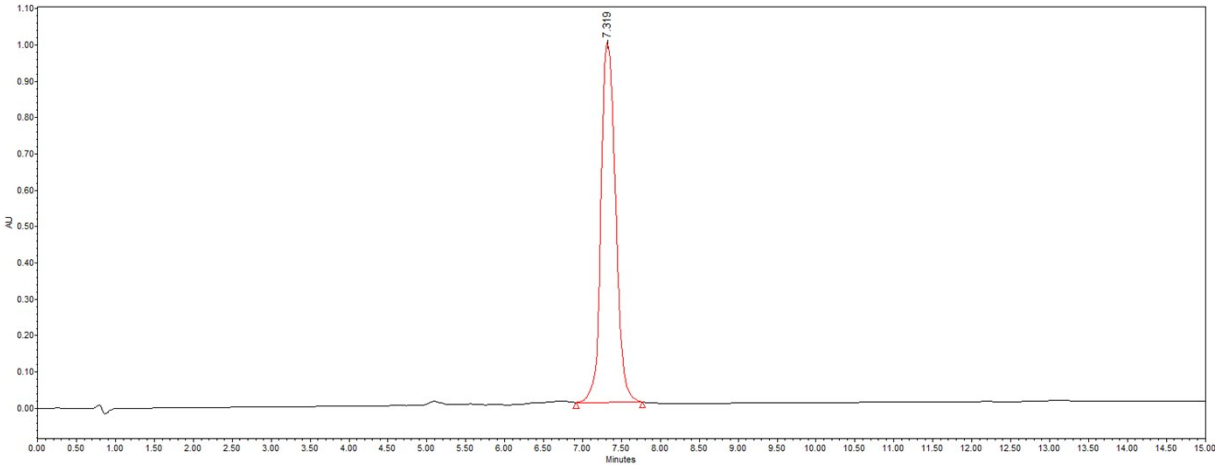


Figure S2. Analytical RP-HPLC of **3-OAc**.

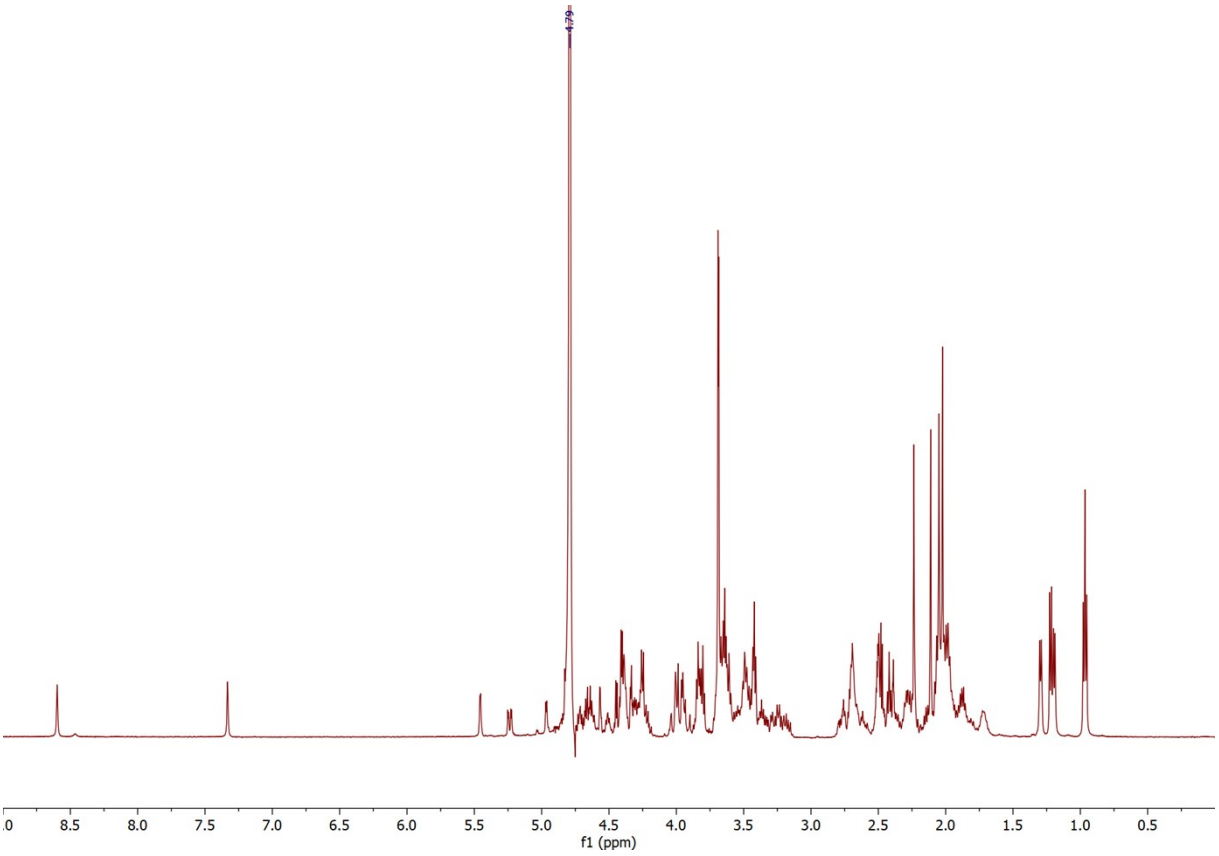


Figure S3. ^{1}H-NMR spectrum of **3-OAc** (400 MHz, D_2O).

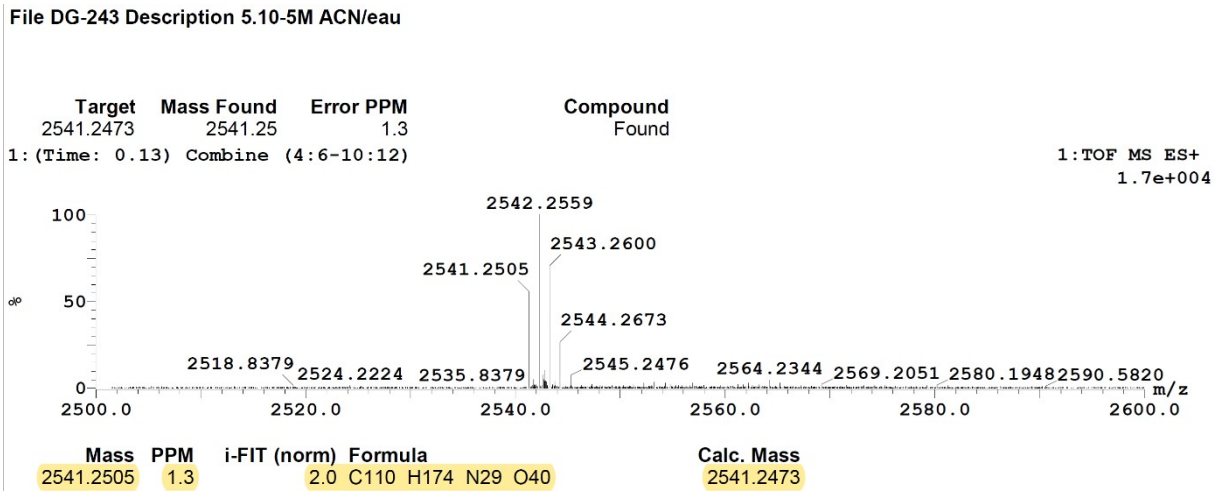


Figure S4. HRMS(+) spectrum of MUC1 glycopeptide **3**.

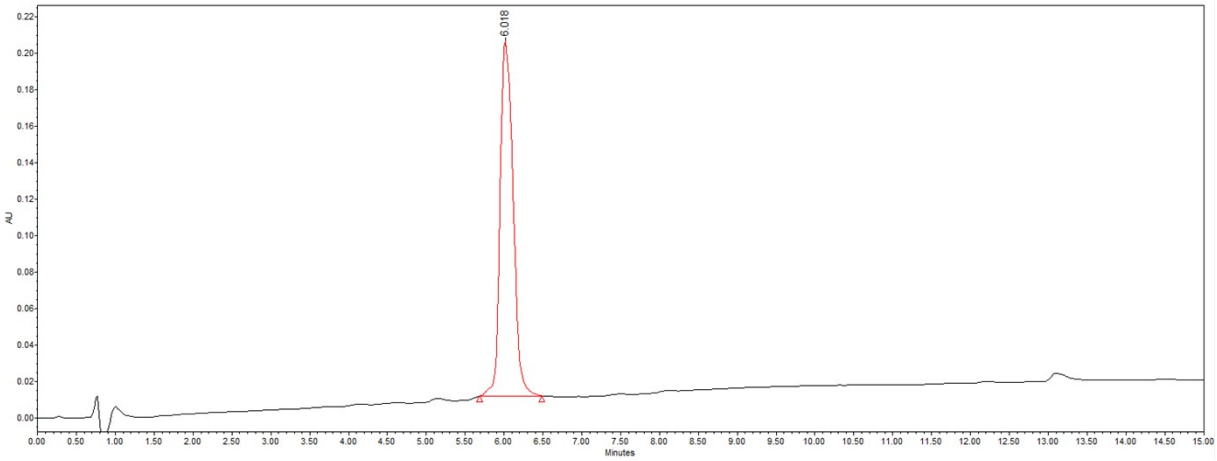


Figure S5. Analytical RP-HPLC of **3**.

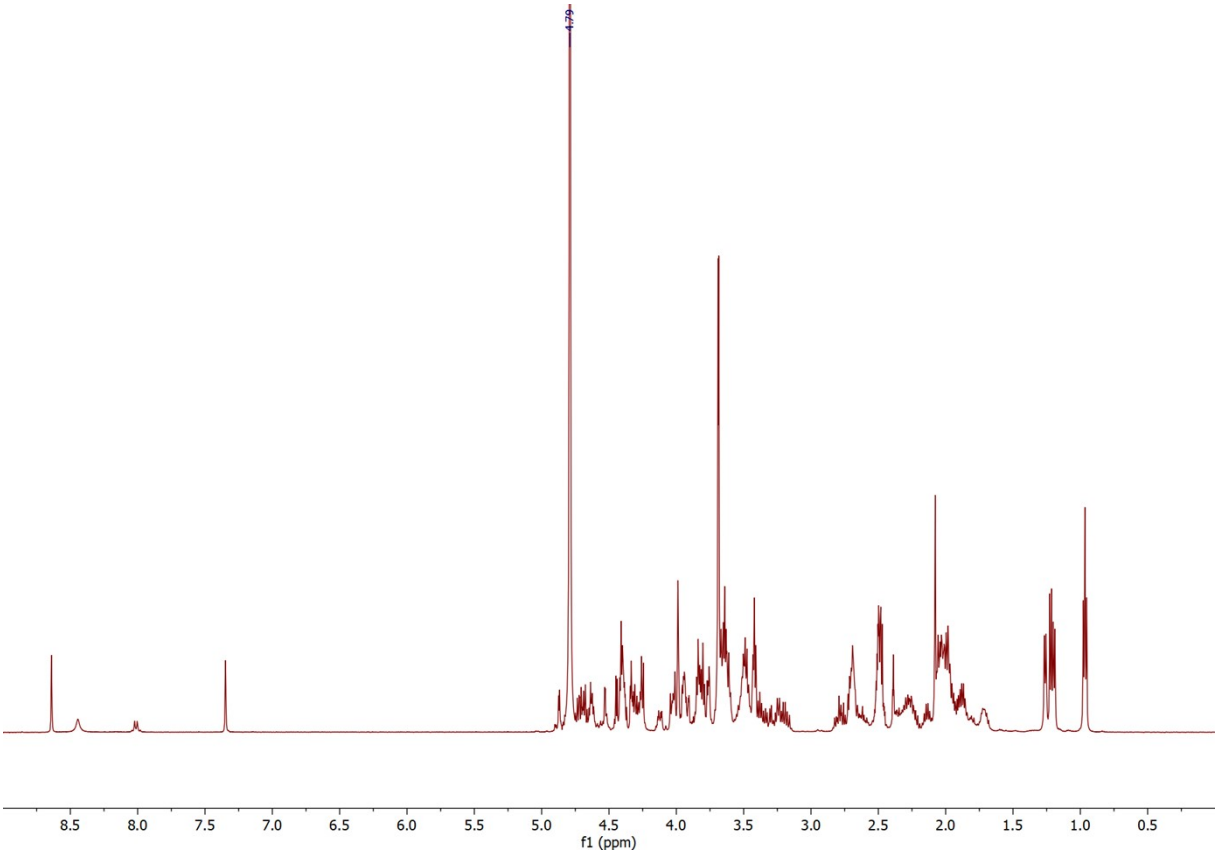
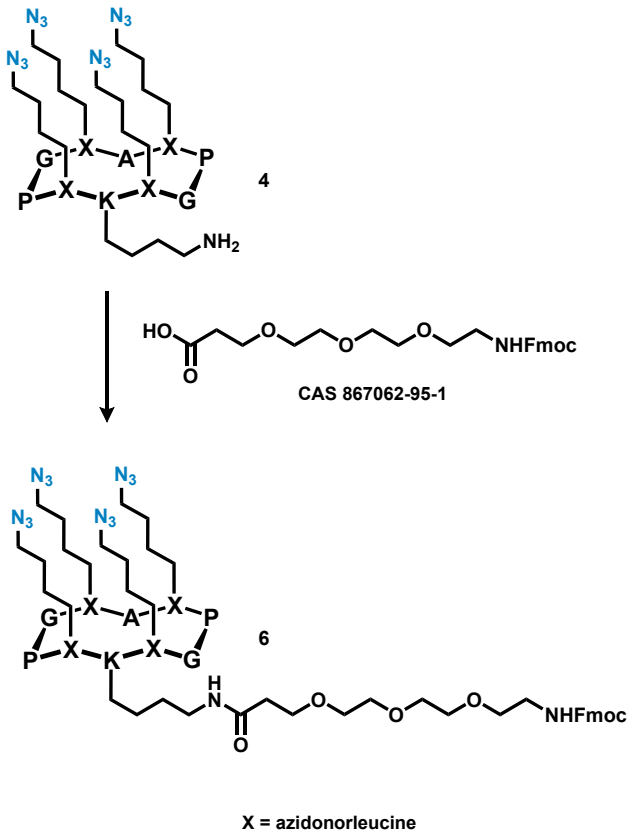


Figure S6. ^{1}H-NMR spectrum of **3** (400 MHz, D_2O).

3-2 Tetra-azido-(PEG)$_3$-NHFmoc scaffold 6



Scheme S2. Synthesis of spacer-containing scaffold **6**.

To a solution of compound **4** (25.0 mg, 22.2 µmol, 1 equiv.) in dry DMF (1 mL), 1-(9H-fluoren-9-yl)-3-oxo-2,8,11,14-tetraoxa-4-azaheptadecan-17-oic acid (CAS 867062-95-1, 14.8 mg, 33.4 µmol, 1.5 equiv.), PyBOP (17.4 mg, 33.4 µmol, 1.5 equiv.) and DIPEA (11.6 µL, 66.7 µmol, 3 equiv.) were added, and the resulting mixture was stirred at r.t. for 1 hour; RP-UHPLC analysis showed complete conversion. The mixture was diluted with H_2O containing 0.1% TFA (3 mL) and purified by semi preparative RP-HPLC (5-100% Solv. B over 15 min.). Fractions containing the product were combined and lyophilized to afford **6** as a white powder (27.9 mg, 18 µmol, 81%). UHPLC-MS (ESI$^+$) *m/z*: calcd for $C_{71}H_{105}N_{24}O_{16}$ $[M+H]^+$ (average): 1550.7, found 1550.3; RP-UHPLC: R*t* = 2.15 min (C18, λ = 214 nm, 5-100% Solv. B over 3.4 min.).

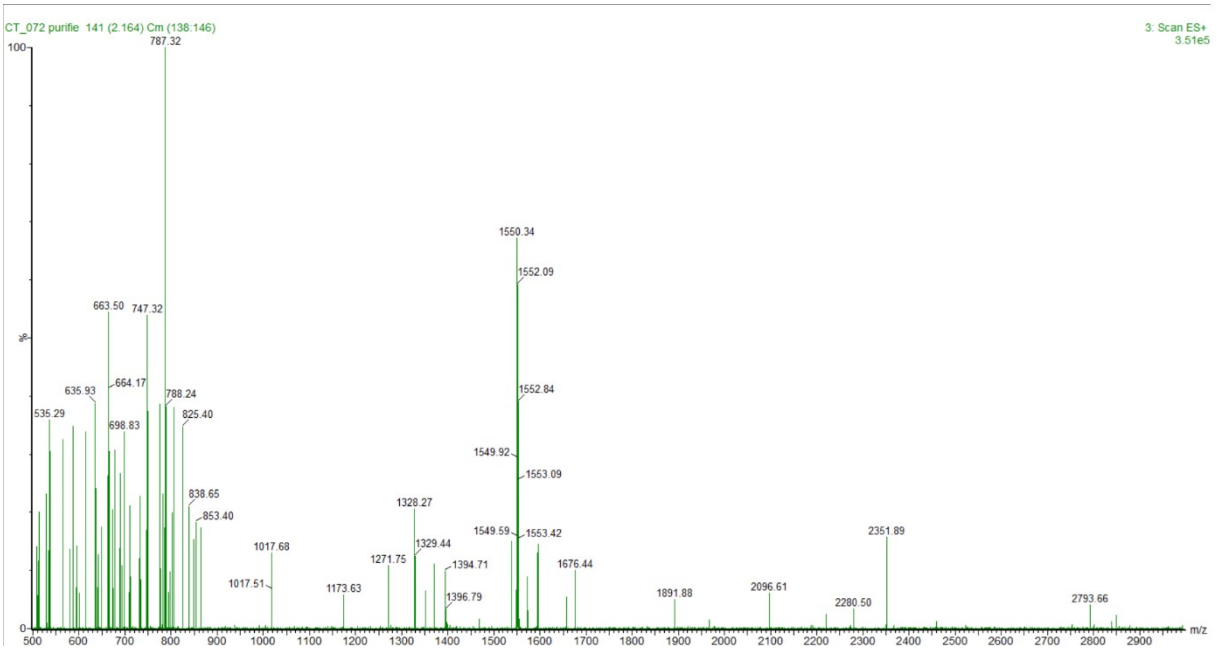


Figure S7. ESI-MS(+) spectrum of **6**.

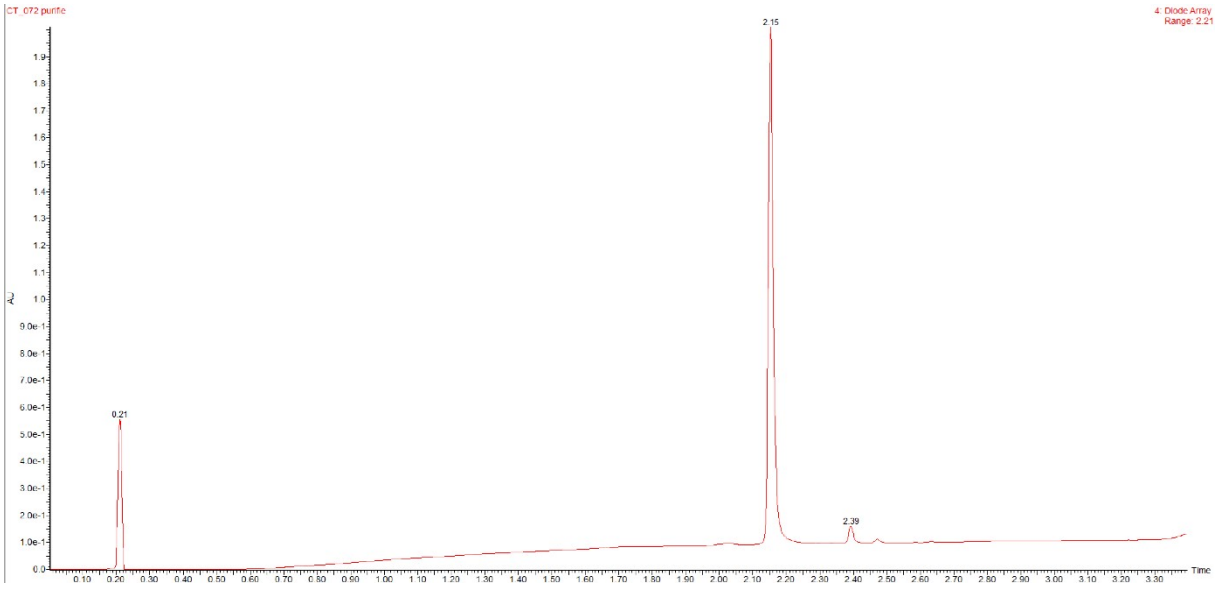
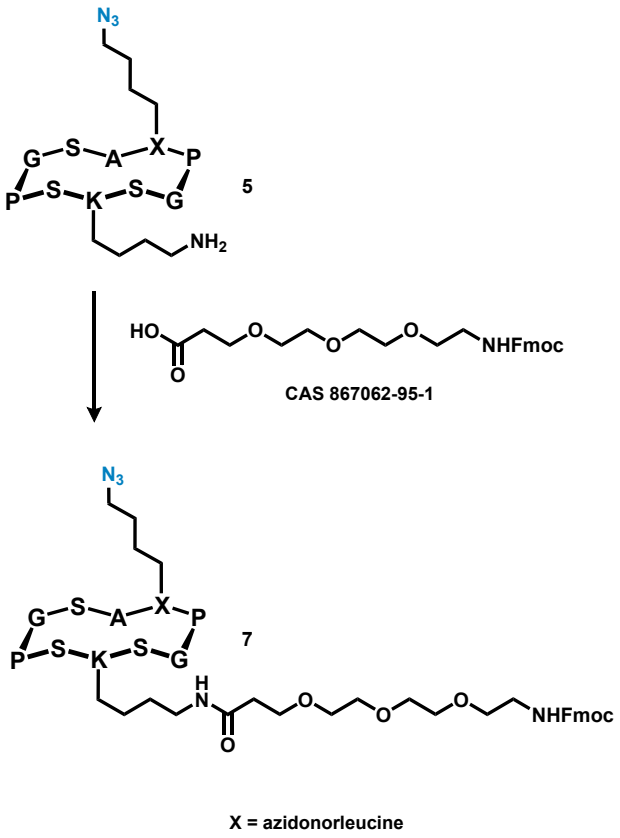


Figure S8. Analytical RP-UHPLC of **6**.

3-3 Mono-azido-$(PEG)_3$-NHFmoc scaffold 7



Scheme S3. Synthesis of spacer-containing scaffold **7**.

To a solution of compound **5** (19.7 mg, 21.3 µmol, 1 equiv.) in dry DMF (1 mL), 1-(9H-fluoren-9-yl)-3-oxo-2,8,11,14-tetraoxa-4-azaheptadecan-17-oic acid (CAS 867062-95-1, 14.2 mg, 32.0 µmol, 1.5 equiv.), PyBOP (16.7 mg, 32.0 µmol, 1.5 eq.) and DIPEA (11.1 µL, 66.7 µmol, 3 equiv.) were added, and the resulting mixture was stirred at r.t. for 1 hour; RP-UHPLC analysis showed complete conversion. The mixture was diluted with H_2O containing 0.1% TFA (3 mL) and then purified by semi preparative RP-HPLC (5-80% Solv. B over 15 min.). Fractions containing the product were combined and lyophilized to afford **7** as a white powder (19.9 mg, 17.7 µmol, 83%). HRMS (ESI$^+$-TOF) *m/z*: calcd for $C_{62}H_{90}N_{15}O_{19}$ $[M+H]^+$ (monoisotopic): 1348.6532, found 1348.6514; RP-HPLC: R*t* = 7.55 min (C18, λ = 214 nm, 5-100% Solv. B over 15 min.).

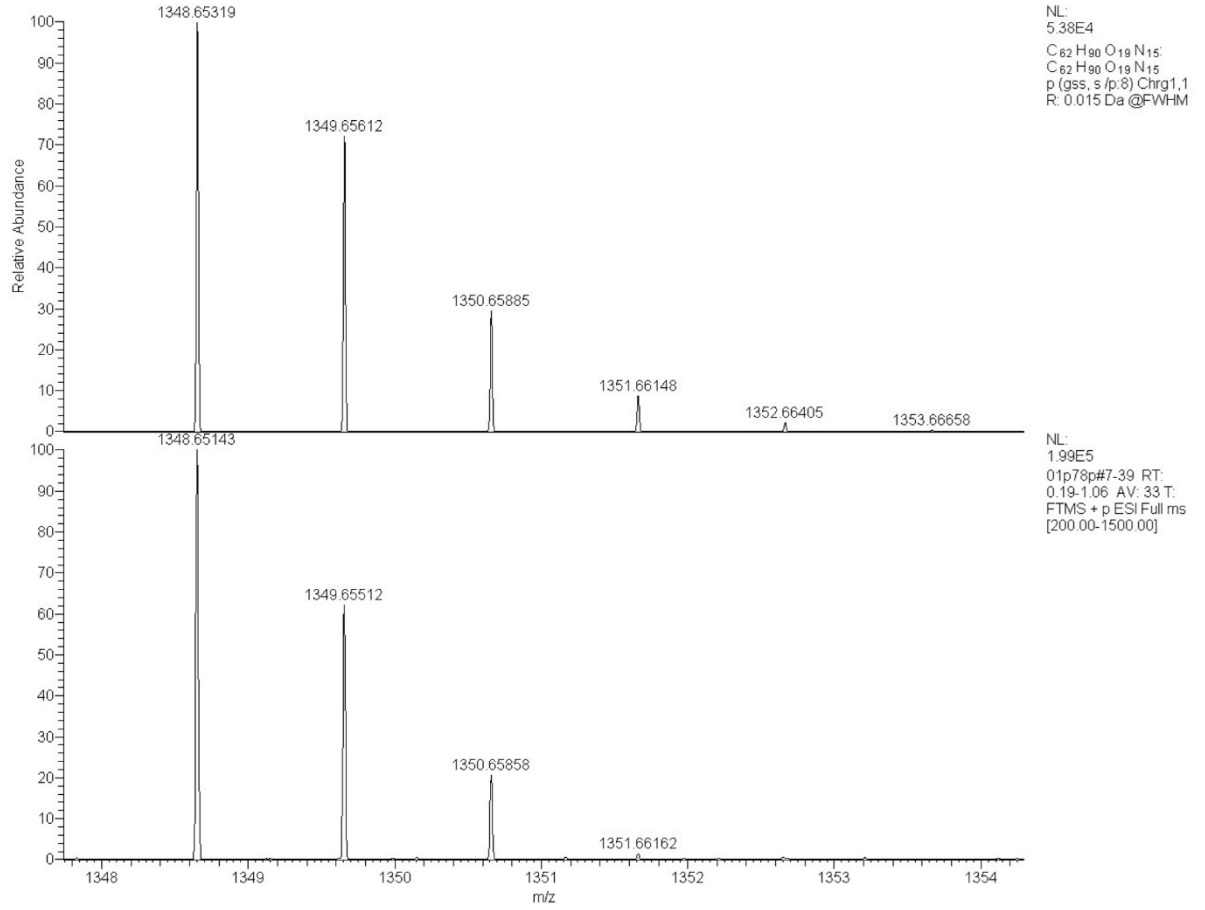


Figure S9. HRMS(+) spectrum of **7**.

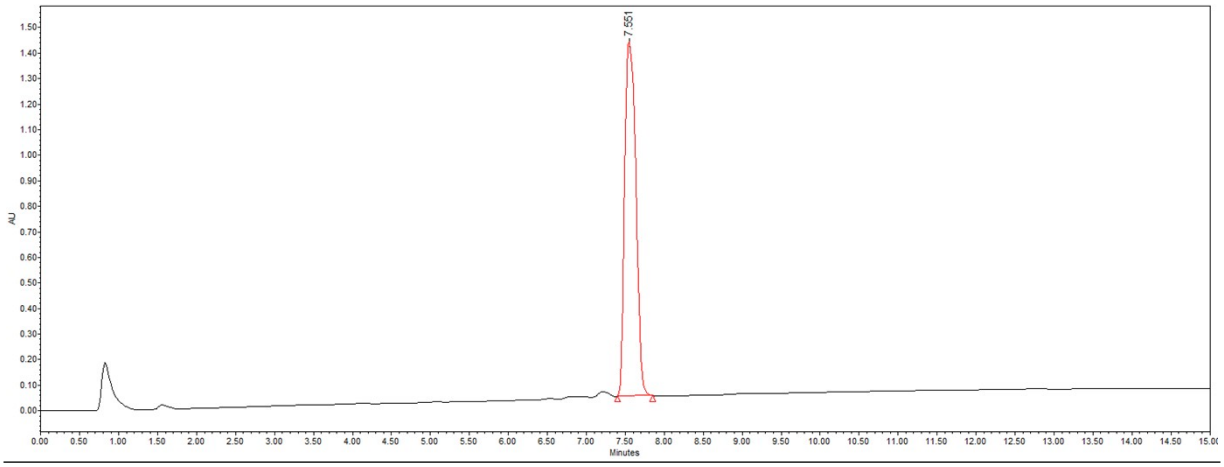
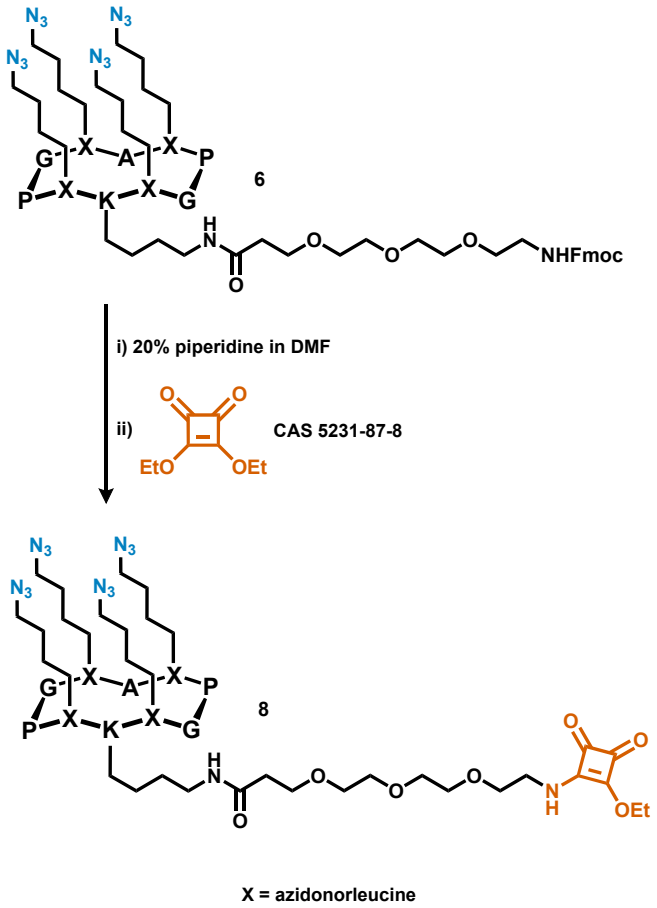


Figure S10. Analytical RP-HPLC of **7**.

3-4 Tetra-azido-(PEG)$_3$-squaramate 8



Scheme S4. Synthesis of activated scaffold **8**.

Compound **6** (15.0 mg, 9.68 µmol, 1 equiv.) was solubilized in DMF/piperidine (8:2, 2 mL) and the resulting mixture was stirred at r.t. for 1 hour, after which RP-UHPLC showed complete conversion of the starting material. The mixture was concentrated under reduced pressure and co-evaporated three times with toluene (10 mL). This crude mixture was solubilized in dry DMF (1 mL), DIPEA (3.4 µL, 19.36 µmol, 2 equiv.) and then, 3,4-diethoxycyclobut-3-ene-1,2-dione (CAS 5231-87-8, 2.1 µL, 14.52 µmol, 1.5 equiv.) was added. The reaction was stirred at r.t. for 2 hours, after which RP-UHPLC showed complete conversion into target compound **8**. The mixture was diluted with H_2O containing 0.1% TFA (3 mL) and purified by semi preparative RP-HPLC (5-100% B over 15 min.). Fractions containing the product were combined and lyophilized to afford activated scaffold **8** as a white powder (9.0 mg, 6.19 µmol, 68%). HRMS (ESI$^+$-TOF) *m/z*: calcd for $C_{62}H_{99}N_{24}O_{17}$ $[M+H]^+$ (monoisotopic): 1451.7615, found 1451.7610; RP-HPLC: R*t* = 7.07 min (C18, λ = 214 nm, 5-100% B over 15 min.)

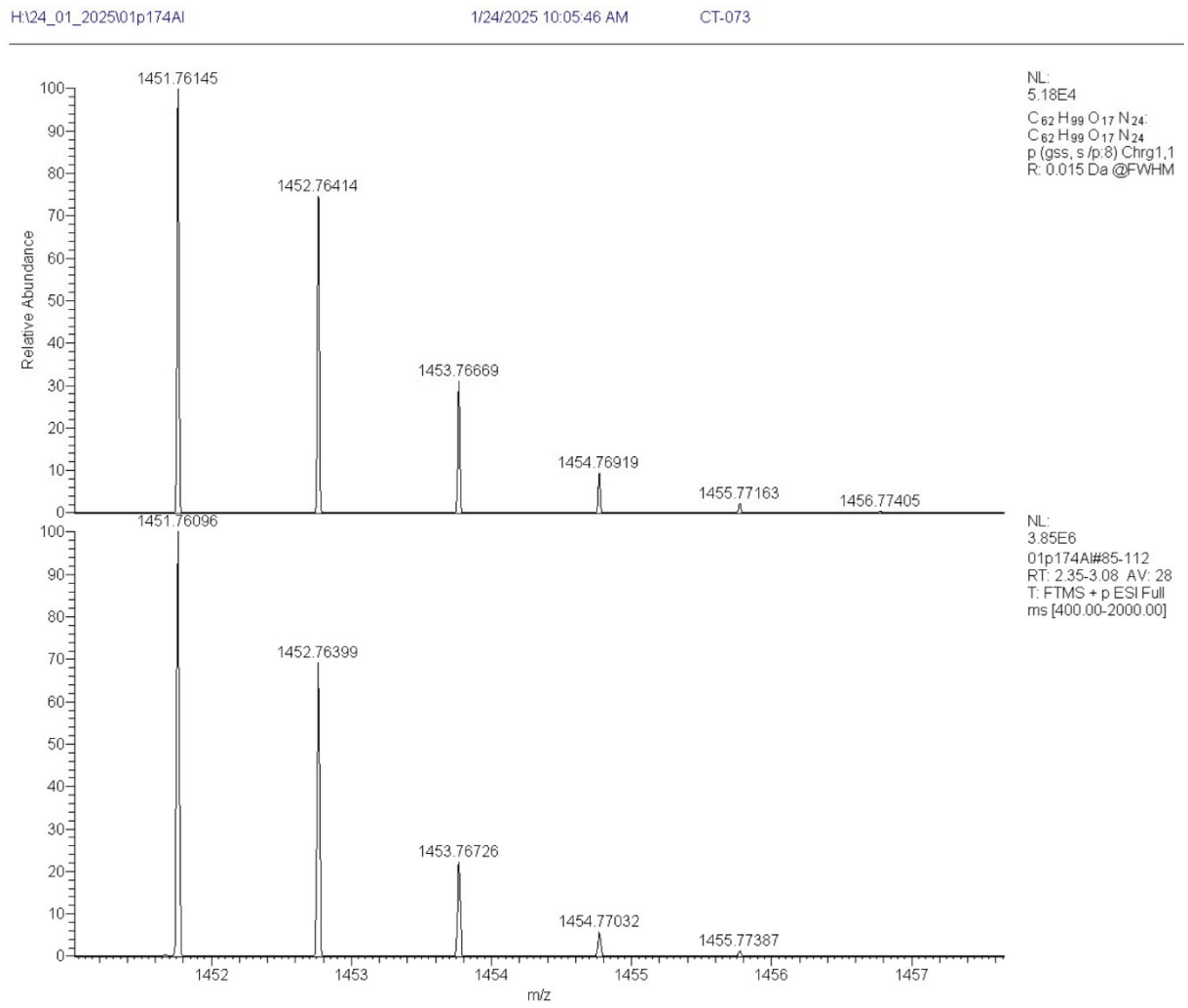


Figure S11. HRMS(+) spectrum of **8.**

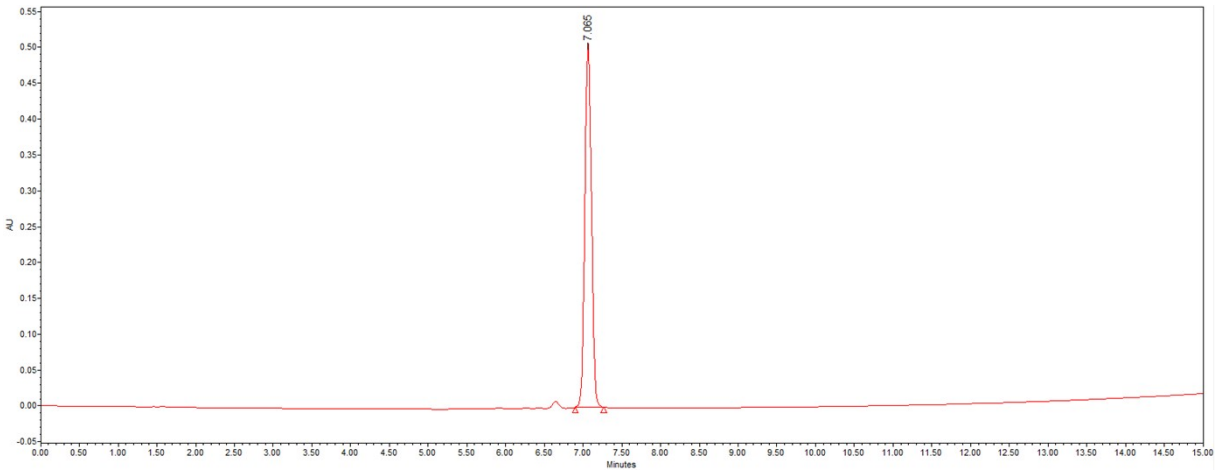


Figure S12. Analytical RP-HPLC of **8**.

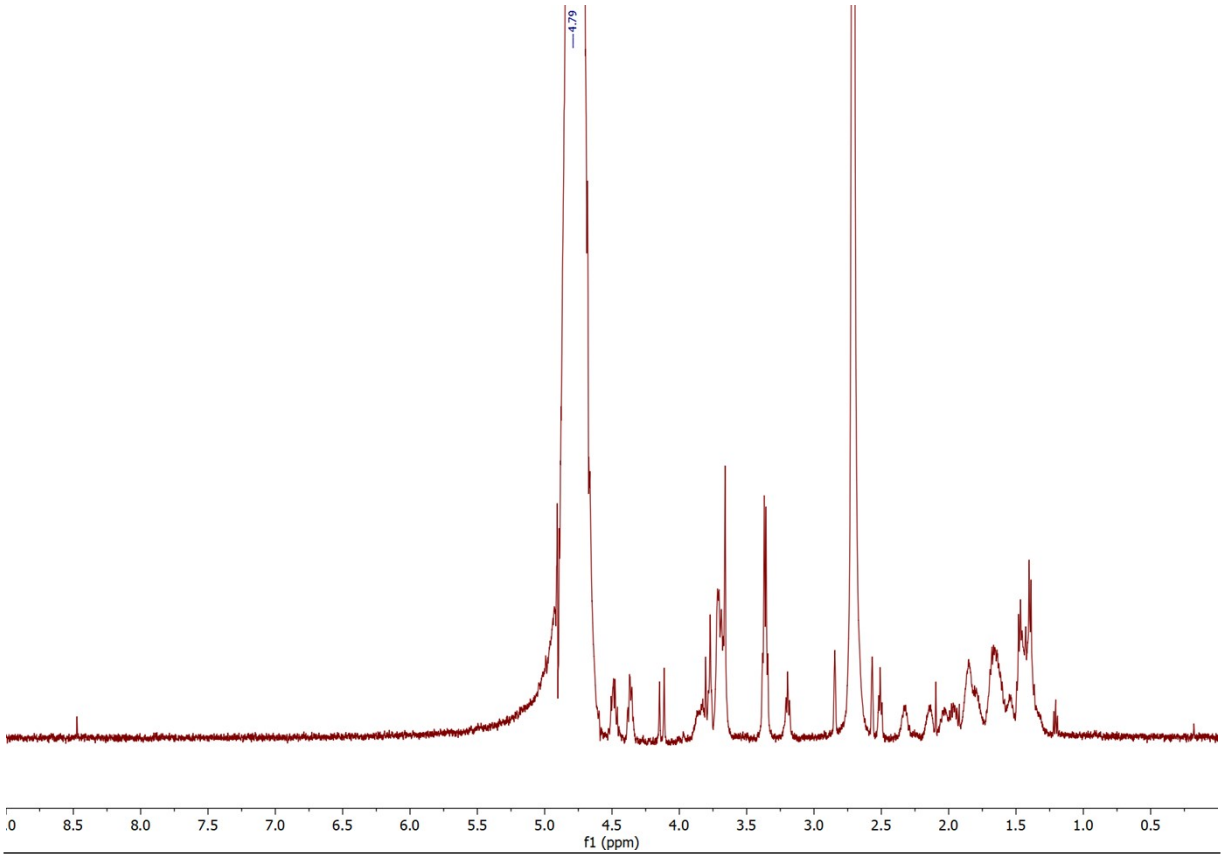
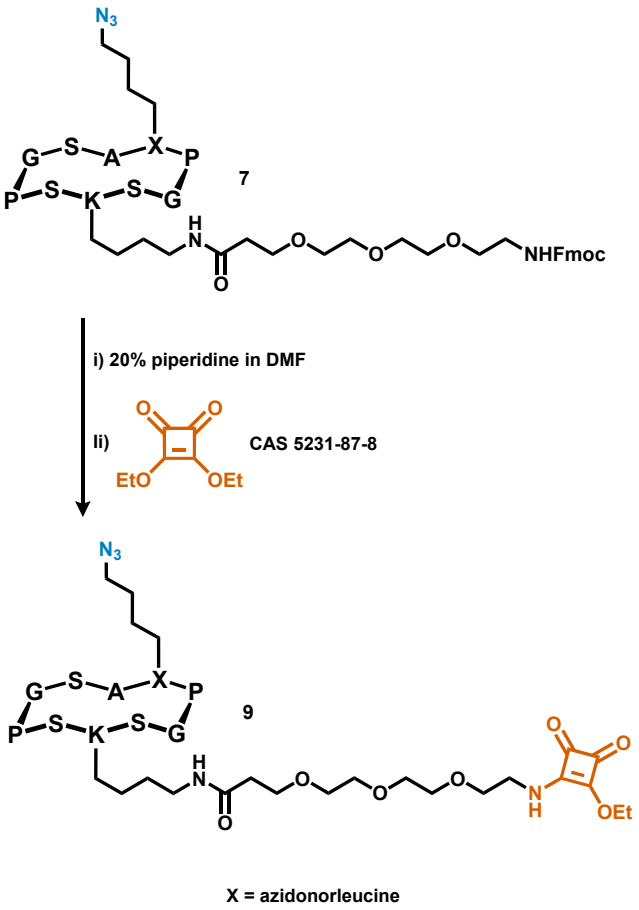


Figure S13. ^{1}H-NMR spectrum of **8** (500 MHz, D_2O).

3-5 Mono-azido-(PEG)$_3$-squaramate 9



Scheme S5. Synthesis of activated scaffold **9**.

Compound **7** (15.0 mg, 11.12 µmol, 1 equiv.) was solubilized in DMF/piperidine (8:2, 2 mL). The mixture was stirred at r.t. for 1 hour, after which RP-UHPLC showed complete conversion of the starting material. The mixture was concentrated under reduced pressure and co-evaporated three times with toluene (10 mL). This crude mixture was solubilized in dry DMF (1 mL), DIPEA (3.9 µL, 22.24 µmol, 2 equiv.) and then 3,4-diethoxycyclobut-3-ene-1,2-dione (CAS 5231-87-8, 2.5 µL, 16.69 µmol, 1.5 eq.) was added. The reaction was stirred at r.t. for 2 hours, after which RP-UHPLC showed complete conversion into target compound **9**. The mixture was diluted with H_2O containing 0.1% TFA (3 mL) and purified by semi preparative RP-HPLC (0-60%B in over min.). Fractions containing the product were combined and lyophilized to afford activated scaffold **9** as a white powder (9.9 mg, 7.90 µmol, 71%). HRMS (ESI$^+$-TOF) *m/z*: calcd for $C_{53}H_{85}N_{15}O_{20}$ $[M+2H]^{2+}$ (monoisotopic): 625.8042, found 625.8039; RP-HPLC: R*t* = 8.01 min (C18, λ = 214 nm, 0-60% B over 15 min.).

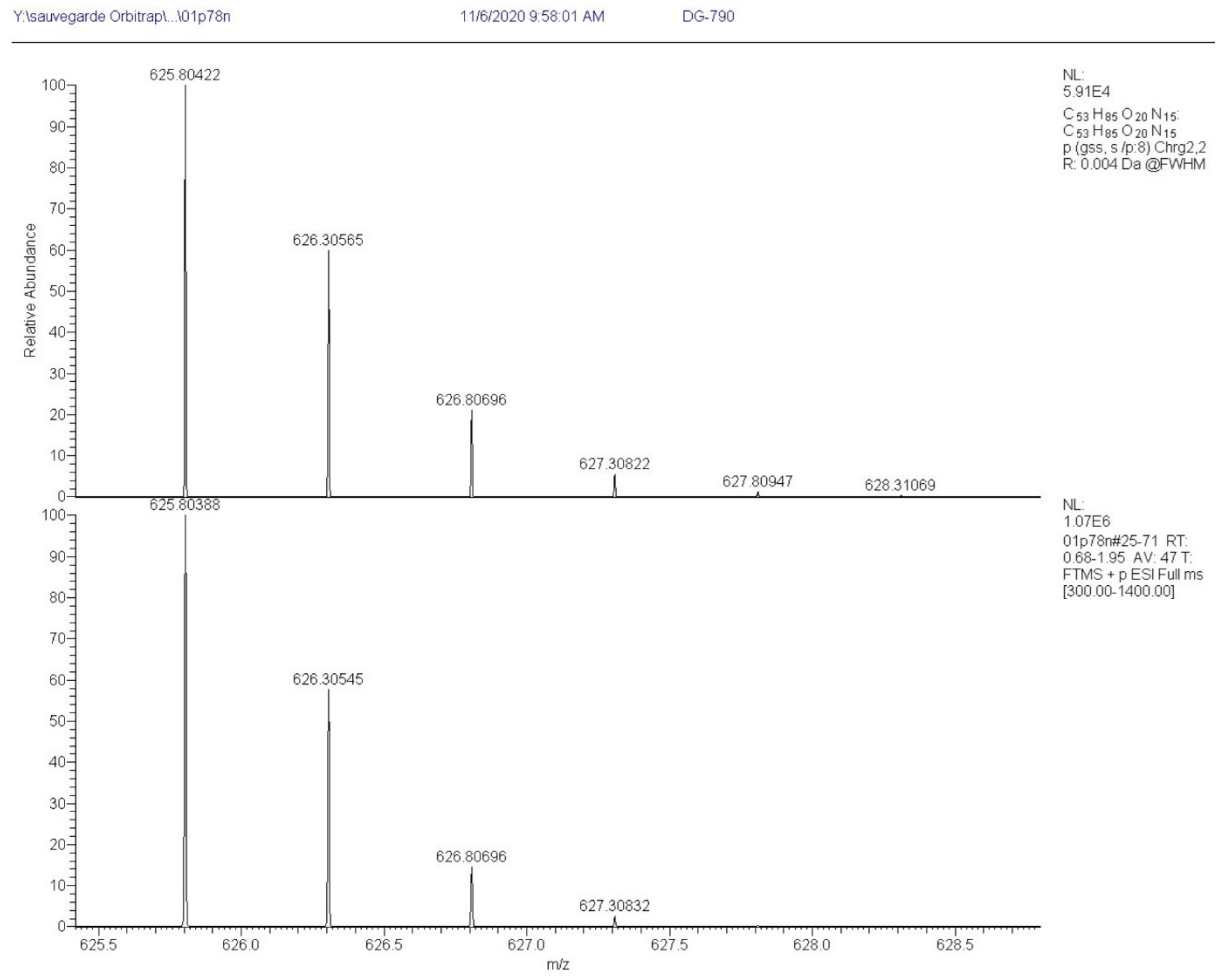


Figure S14. HRMS(+) spectrum of **9**.

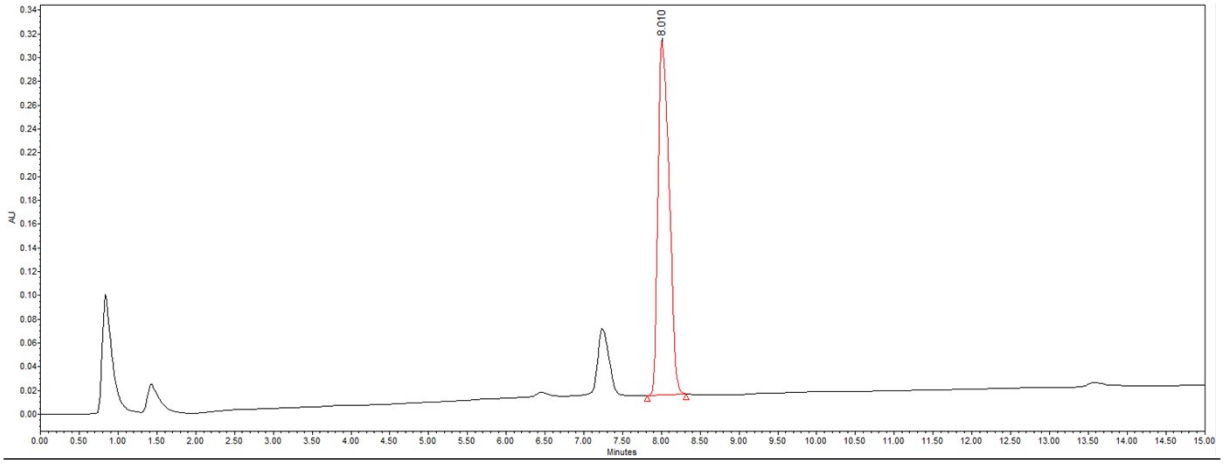


Figure S15. Analytical RP-HPLC of **9**.

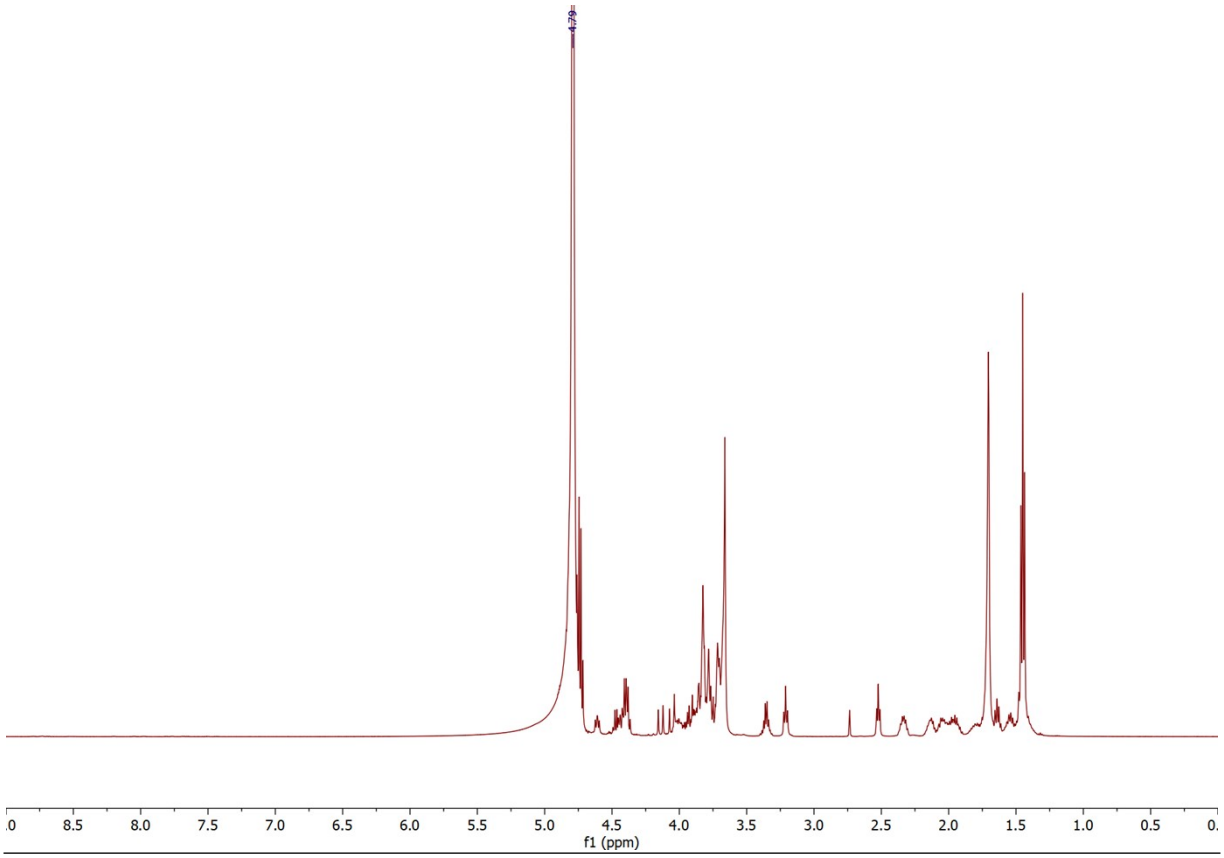
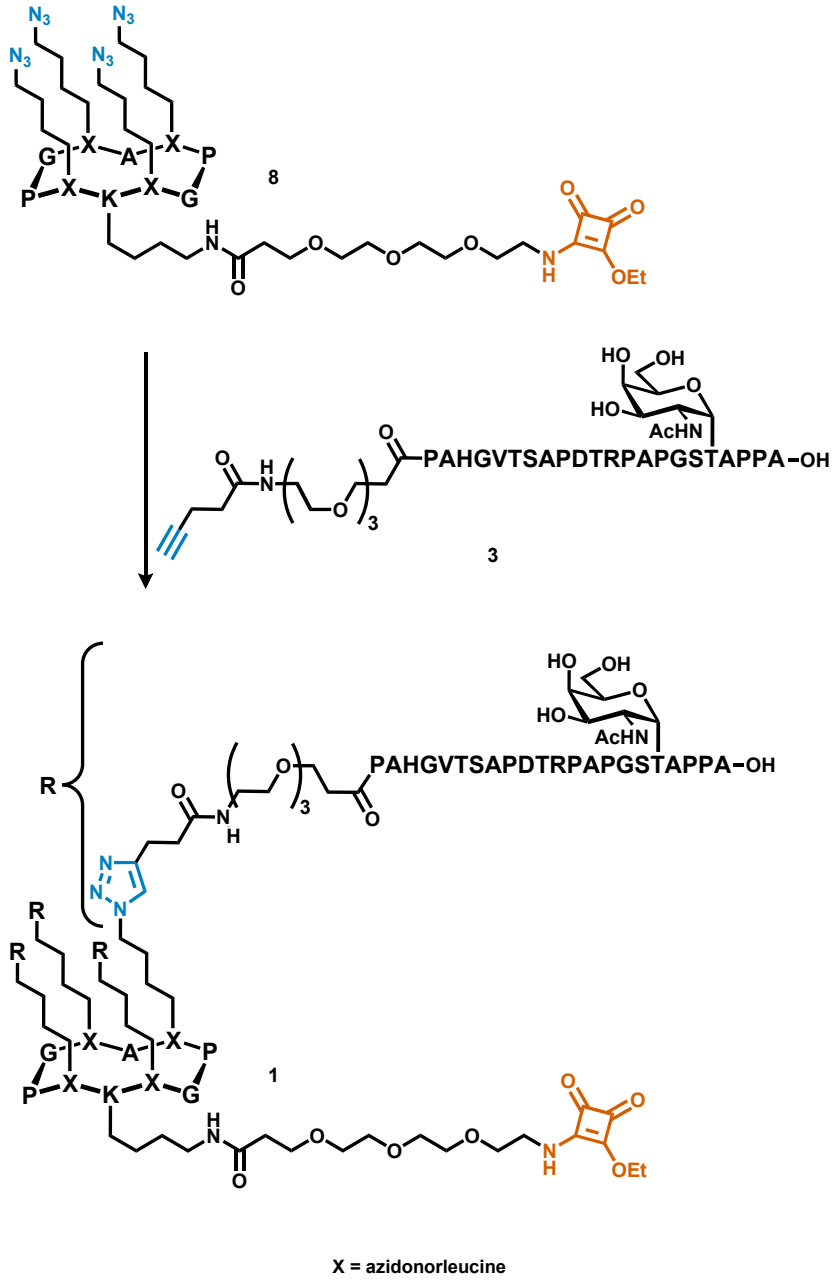


Figure S16. ^{1}H-NMR spectrum of **9** (500 MHz, D_2O).

3-6 TetraMUC1 antigen-presenting unit 1



Scheme S6. Synthesis of "tetraMUC1" antigen-presenting unit **1**.

To a solution of activated scaffold **8** (5.2 mg, 3.58 µmol, 1 equiv.) and alkyne-terminating MUC1 glycopeptide **3** (40.0 mg, 15.76 µmol, 4.4 equiv.) in a 1:1 mixture of DMF (0.5 mL) and PBS pH 7.5 (0.5 mL), a solution of $CuSO_4 \cdot 5H_2O$ (0.4 mg, 1.79 µmol, 0.5 equiv.) and THPTA (1.6 mg, 3.58 µmol, 1 equiv.) in PBS pH 7.5 (0.2 mL) and a solution of sodium ascorbate (2.1 mg, 10.75 µmol, 3 equiv.) in PBS pH 7.5 (0.2 mL) were added. The resulting mixture was degassed and stirred under argon atmosphere at r.t., until UHPLC showed completion of the reaction (ca. 2 hours.). Chelex® resin (50-100 mesh, CAS 11139-85-8) was added to the reaction, which was stirred for additional 45 min. at r.t. The resin was filtered off and rinsed with water (3 times, 1 mL) and the filtrate was purified by semi preparative RP-HPLC (5-60% B over 15 min.). Fractions containing the product were combined and lyophilized to afford

tetraMUC1 unit **1** as a white powder (29.6 mg, 2.55 µmol, 71%). MALDI-TOF MS(+) *m/z* calcd. for $C_{502}H_{791}N_{140}O_{177}$ $[M+H]^+$ (average): 11619.48, found 11620.98; RP-HPLC: R*t* = 8.27 min (C18, λ = 214 nm, 0-60% B over 15 min.)

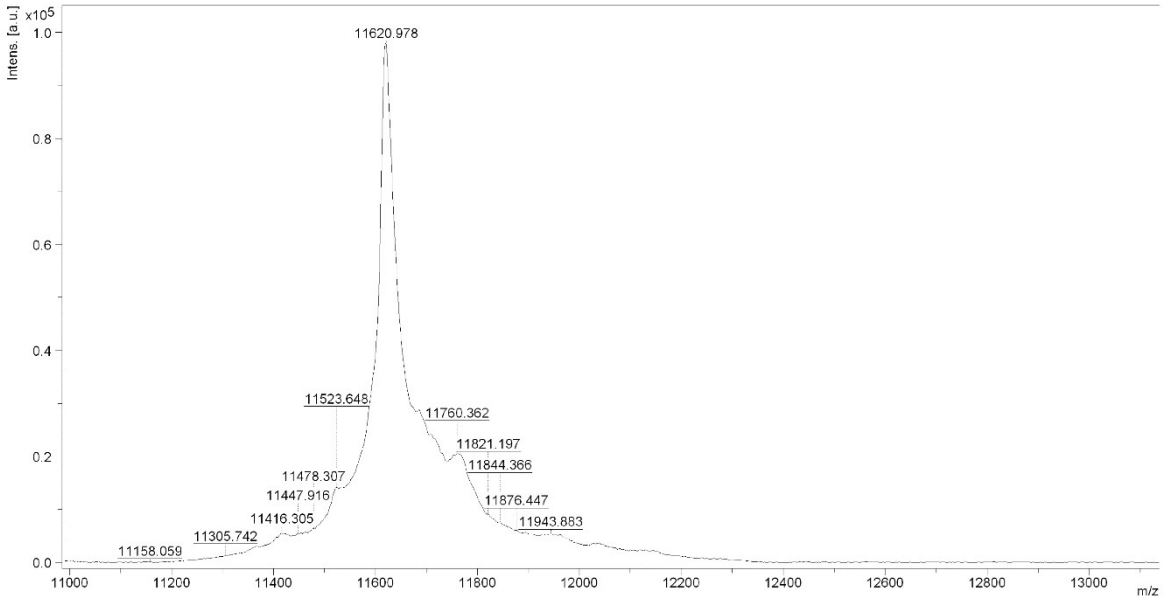


Figure S17. MALDI-TOF MS(+) spectrum of tetraMUC1 unit **1**.

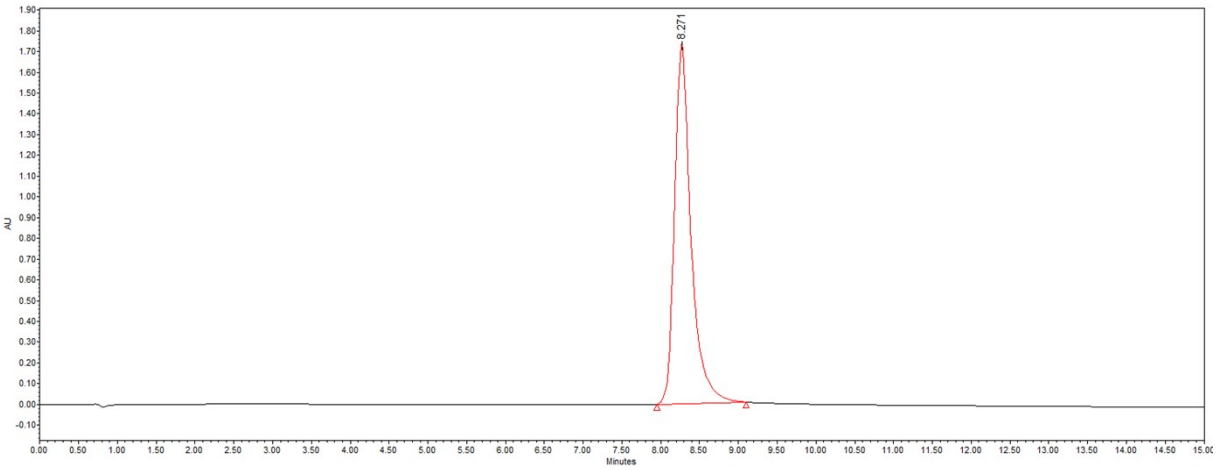


Figure S18. Analytical RP-HPLC of **1**.

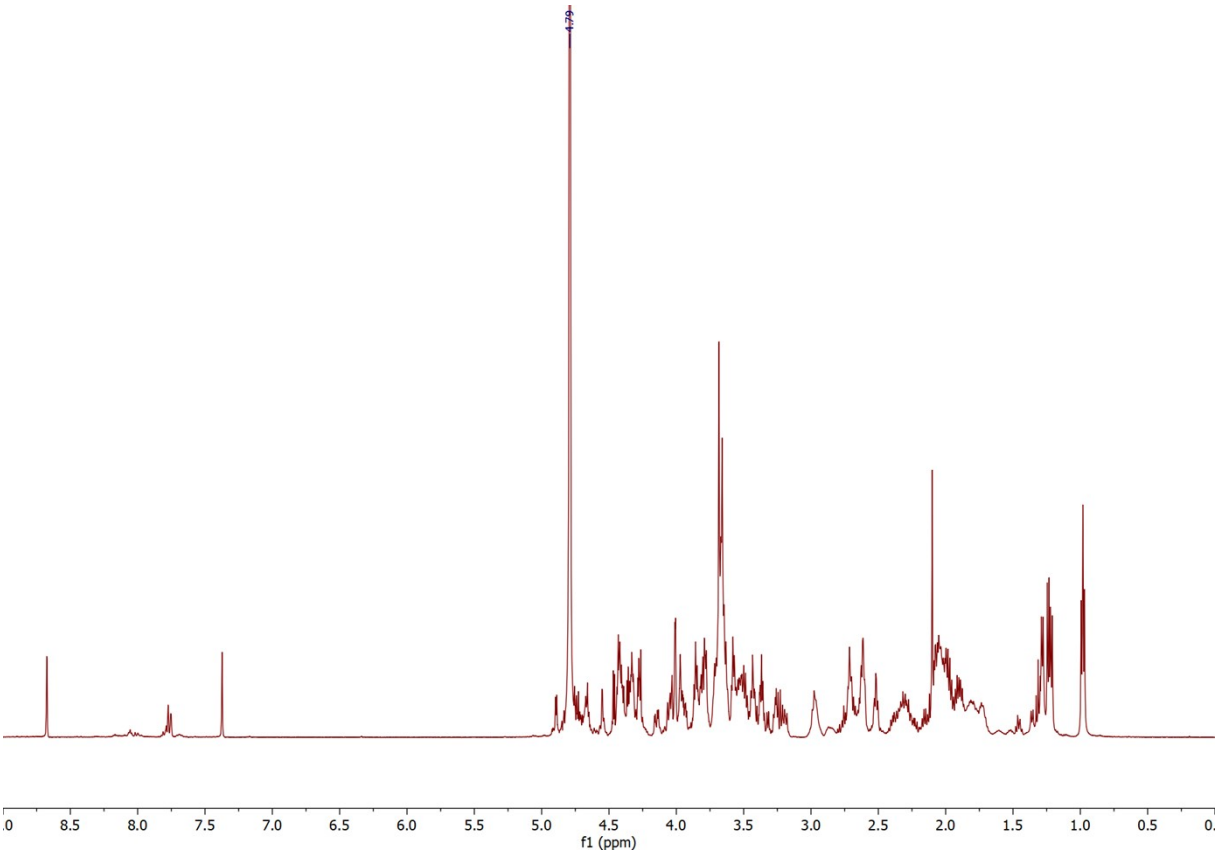
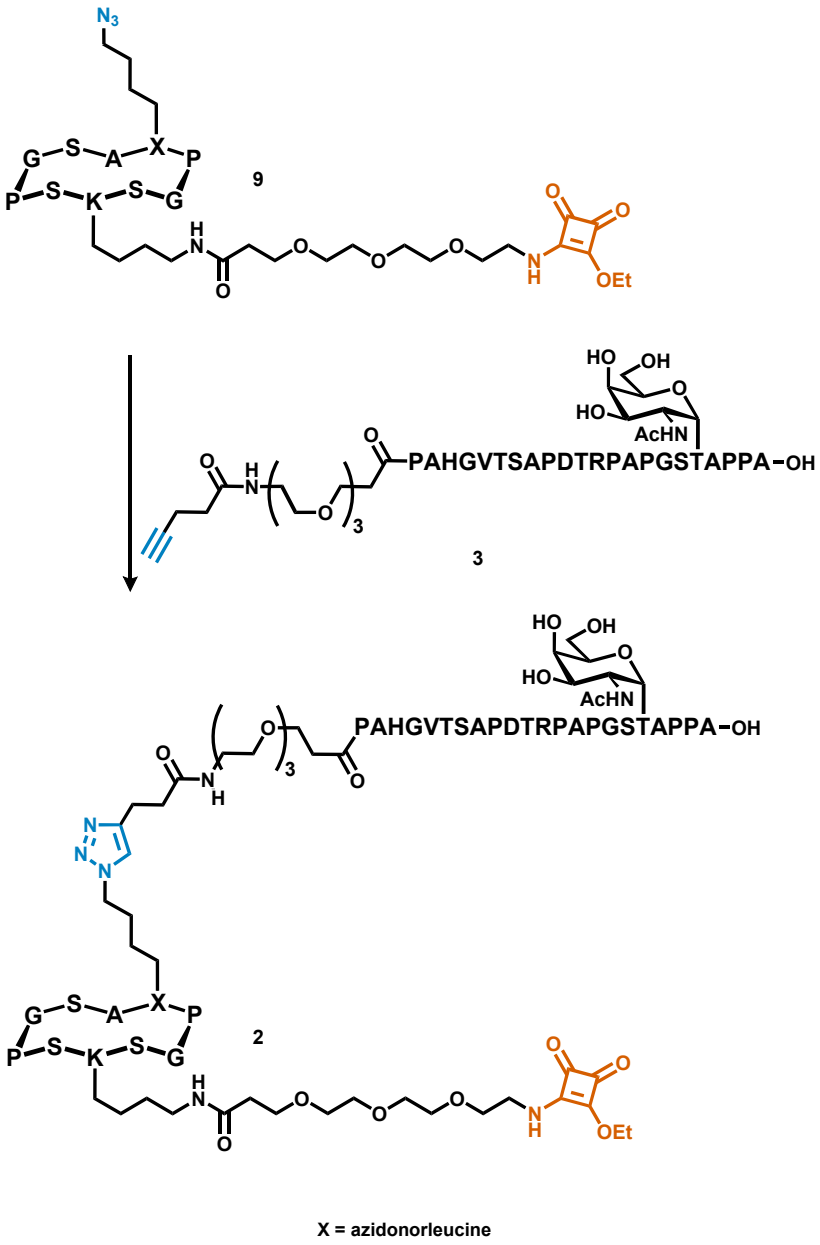


Figure S19. ^{1}H NMR spectrum of **1** (D_2O, 500 MHz).

3-7 MonoMUC1 antigen-presenting unit 2



Scheme S7. Synthesis of "monoMUC1" antigen-presenting unit **2**.

To a solution of scaffold **9** (8.3 mg, 6.62 µmol, 1 equiv.) and glycopeptide **3** (15.3 mg, 6.02 µmol, 4.4 equiv.) in a 1:1 mixture of DMF (0.5 mL) and PBS pH 7.5 (0.5 mL), a solution of $CuSO_4 \cdot 5H_2O$ (0.3 mg, 1.20 µmol, 0.2 equiv.) and THPTA (1.0 mg, 2.41 µmol, 0.4 equiv.) in PBS pH 7.5 (0.2 mL) and a solution of sodium ascorbate (1.2 mg, 6.02 µmol, 1 equiv.) in PBS pH 7.5 (0.2 mL) were added. The resulting mixture was degassed and stirred under argon atmosphere at r.t., until UHPLC showed completion of the reaction (ca. 2 hours.). Chelex® resin (50-100 mesh, CAS 11139-85-8) was added to the reaction, which was stirred for additional 45 min. at r.t. The resin was filtered off and rinsed with water (3 times, 1 mL) and the filtrate was purified by semi preparative RP-HPLC (0-40% B over 15 min.). Fractions containing the product were combined and lyophilized to afford monoMUC1 unit **2** as a white powder

(16.7 mg, 4.39 µmol, 74%). HRMS (ESI$^+$-TOF) *m/z* calcd. for $C_{163}H_{256}N_{44}O_{60}$ $[M+3H]^{3+}$ (monoisotopic): 1264.2851, found 1264.2853; RP-HPLC: R*t* = 11.90 min (C18, λ = 214 nm, 0-40% B over 15 min.)

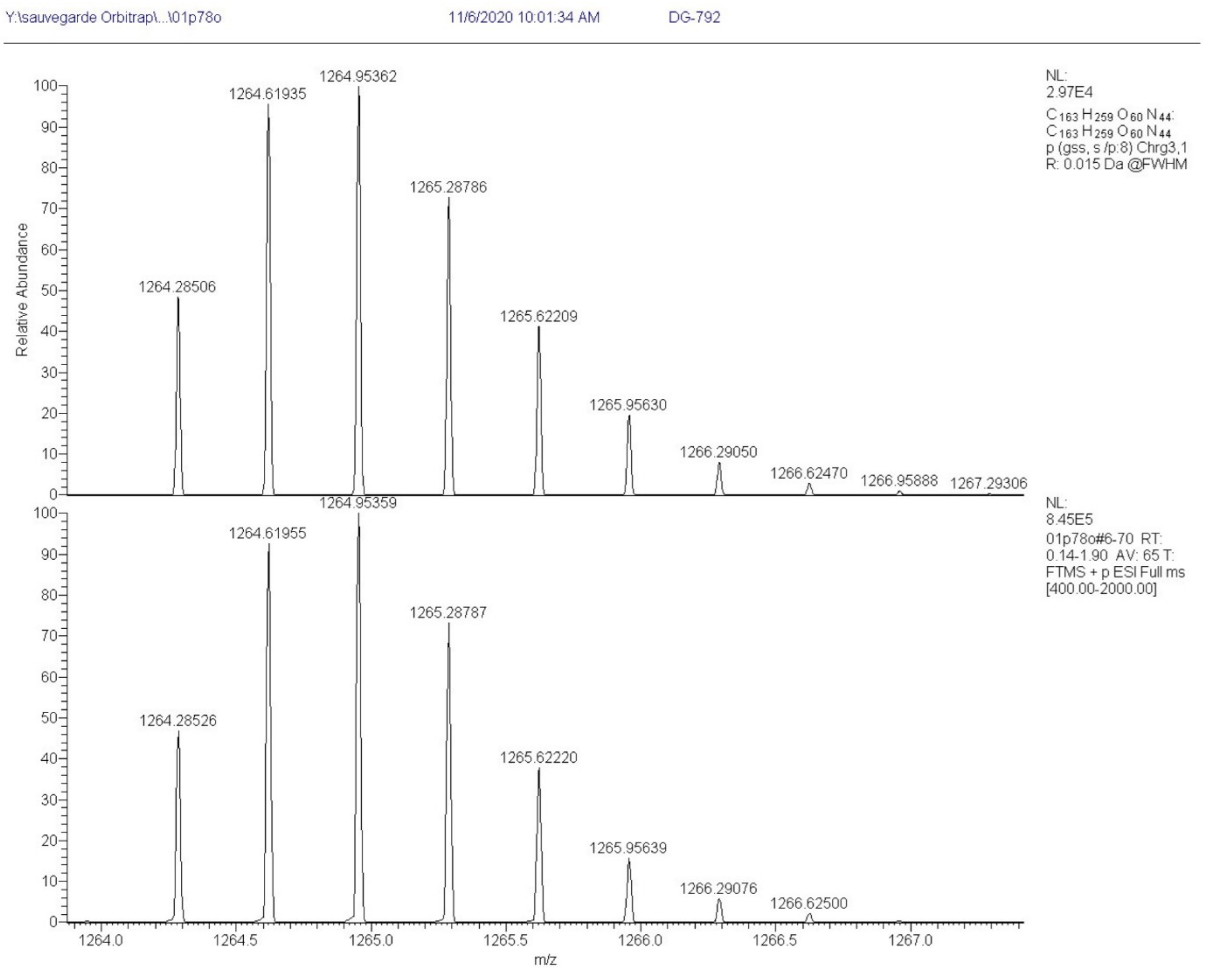


Figure S20. HRMS(+) spectrum of monoMUC1 unit **2**.

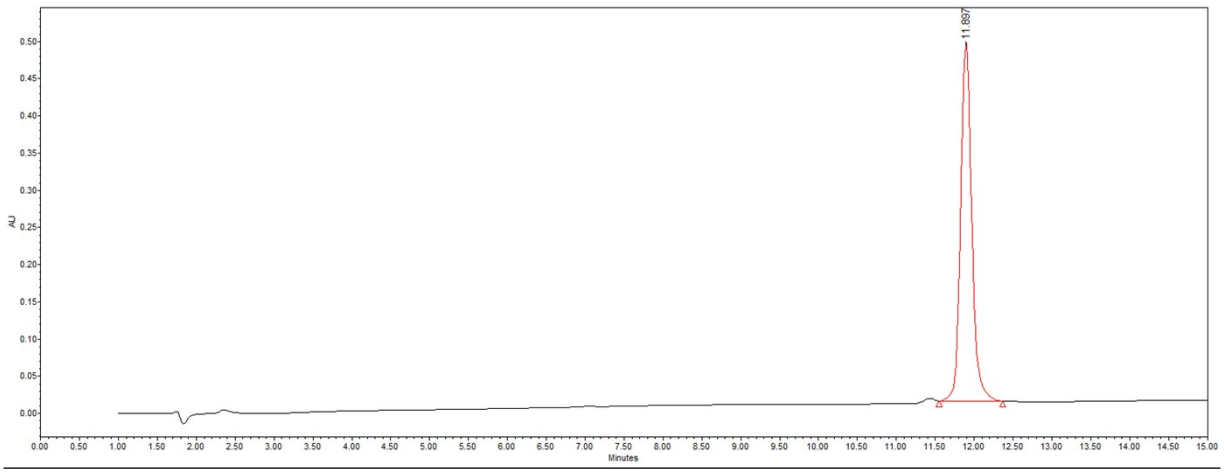


Figure S21. Analytical RP-HPLC of **2**.

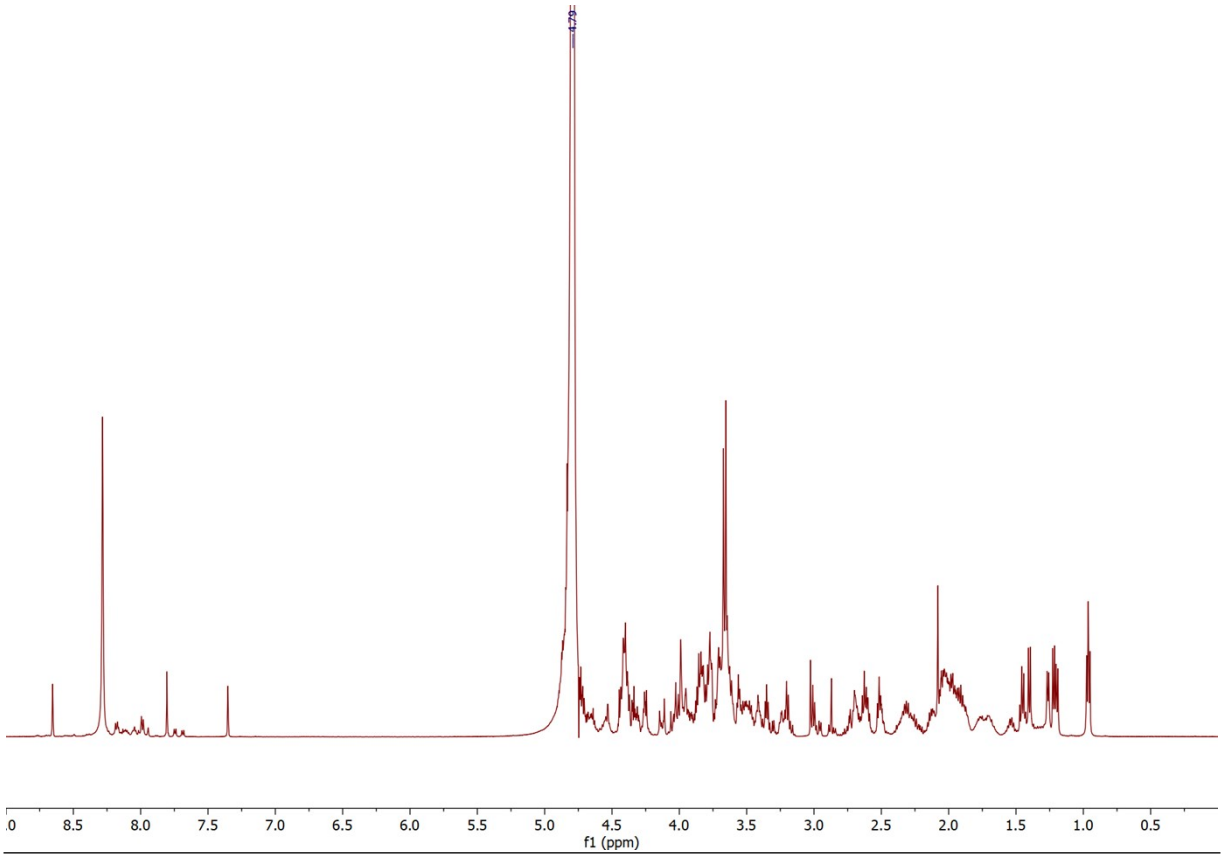
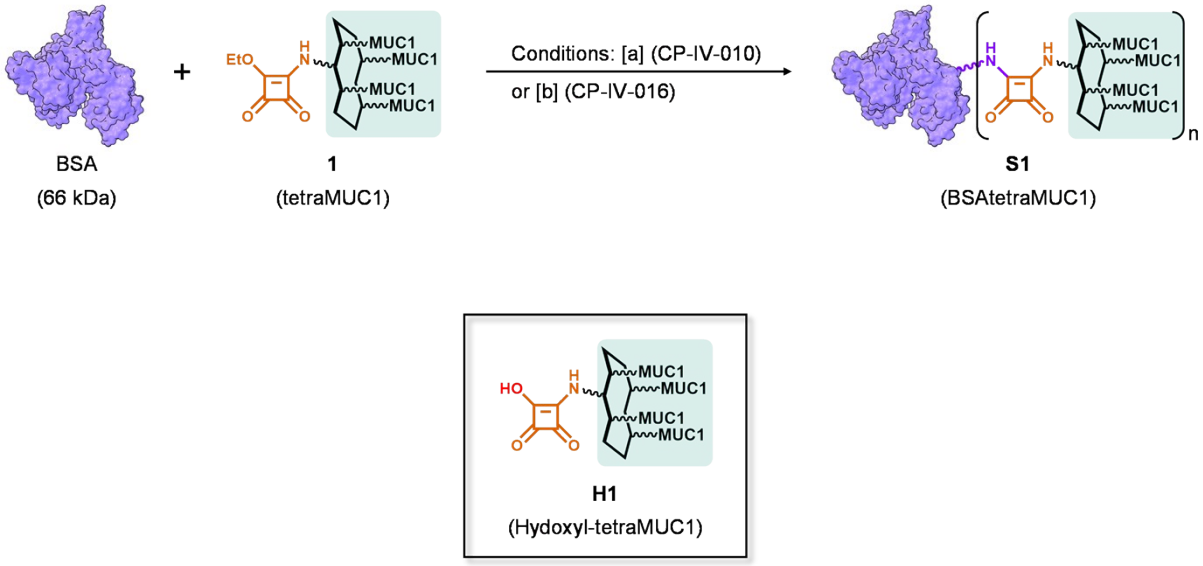


Figure S22. ^{1}H NMR spectrum of **2** (D_2O, 500 MHz).

4- Synthesis and characterization of BSA–tetraMUC1 conjugate (S1)



Scheme S8. General reaction scheme for BSA conjugation with tetraMUC1 **1**, showing also the hydrolyzed by-product **H1** (see Fig. S24 and S25).

Conditions [a] (CP-IV-010):

A mixture of bovine serum albumin (BSA, Fisher Scientific, CAS: 9048-46-8, 1.09 mg, 16 nmol) and **1** (3.80 mg, 327 nmol) in 1.0 mL of 0.1 M sodium carbonate buffer (pH 9.6) was heated at 40 °C under stirring in an Eppendorf ThermoMixer®. Reaction progress was monitored by LC-MS [Waters 1525 binary gradient system, Waters 2998 photodiode array detector, SQD2 single quadrupole] using a linear Solv. A / Solv. B gradient [1 ml/min., 10-80% Solv. B over 30 min.] on an analytical Nucleosil-C4 RP-HPLC column. After 48 h, the crude mixture was directly purified by RP-HPLC on a semi-preparative Nucleosil-C4 column [5 ml/min., 10% Solv. B for 5 min., then 10-80% Solv. B over 25 min.] and the collected fractions lyophilized, providing 0.90 mg of a white powder.

Conditions [b] (CP-IV-016):

A mixture of bovine serum albumin (BSA, Fisher Scientific, CAS: 9048-46-8, 1.42 mg, 21 nmol) and **1** (4.98 mg, 429 nmol) in 262 µL of a 0.07 M $Na_2B_4O_7$ / 0.035 M $KHCO_3$ buffer (pH 9.5) was stirred in an Eppendorf ThermoMixer® at room temperature. Reaction progress was monitored by LC-MS (*vide supra*). After 48 h, the crude mixture was directly purified by RP-HPLC on a C4, semi-preparative column; the collected fractions lyophilized, providing 1.45 mg of a white powder.

Observations:

Both protocols yielded target conjugate **S1**, but showing variable conjugation efficiency. Under conditions [b], Coomassie-stained SDS-PAGE analysis showed an extent of conjugation (n) ranging from 0-to-4 antigen-presenting units, while for conditions [a] "n" values were 0-to-3 (Fig. S23). In addition, LC-MS analysis showed that under conditions [a] (Fig. S24), starting compound **1** was completely

hydrolyzed at t = 24h, while under conditions [b] (Fig. S25) hydrolysis of **1** was < 50%. Finally, since BSA (66.4 kDa) migrated at a lower position than expected, we performed MALDI-TOF MS analysis of the starting carrier protein (Fig. S26A) and compared it to purified **S1** from conditions [b] (Fig. S26B), confirming our previous observations on the extent of conjugation.

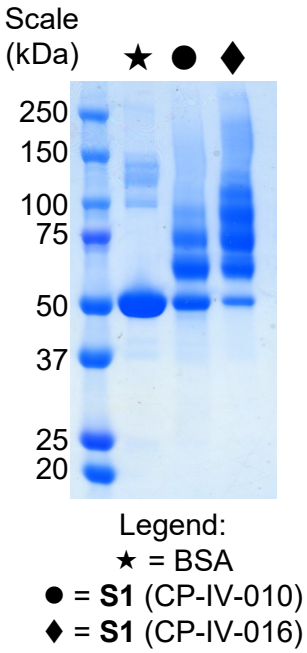


Figure S23. Coomassie-stained SDS-PAGE analysis of BSA, and purified BSA–tetraMUC1 conjugate **S1** obtained under conditions [a] (CP-IV-010) and conditions [b] (CP-IV-016).

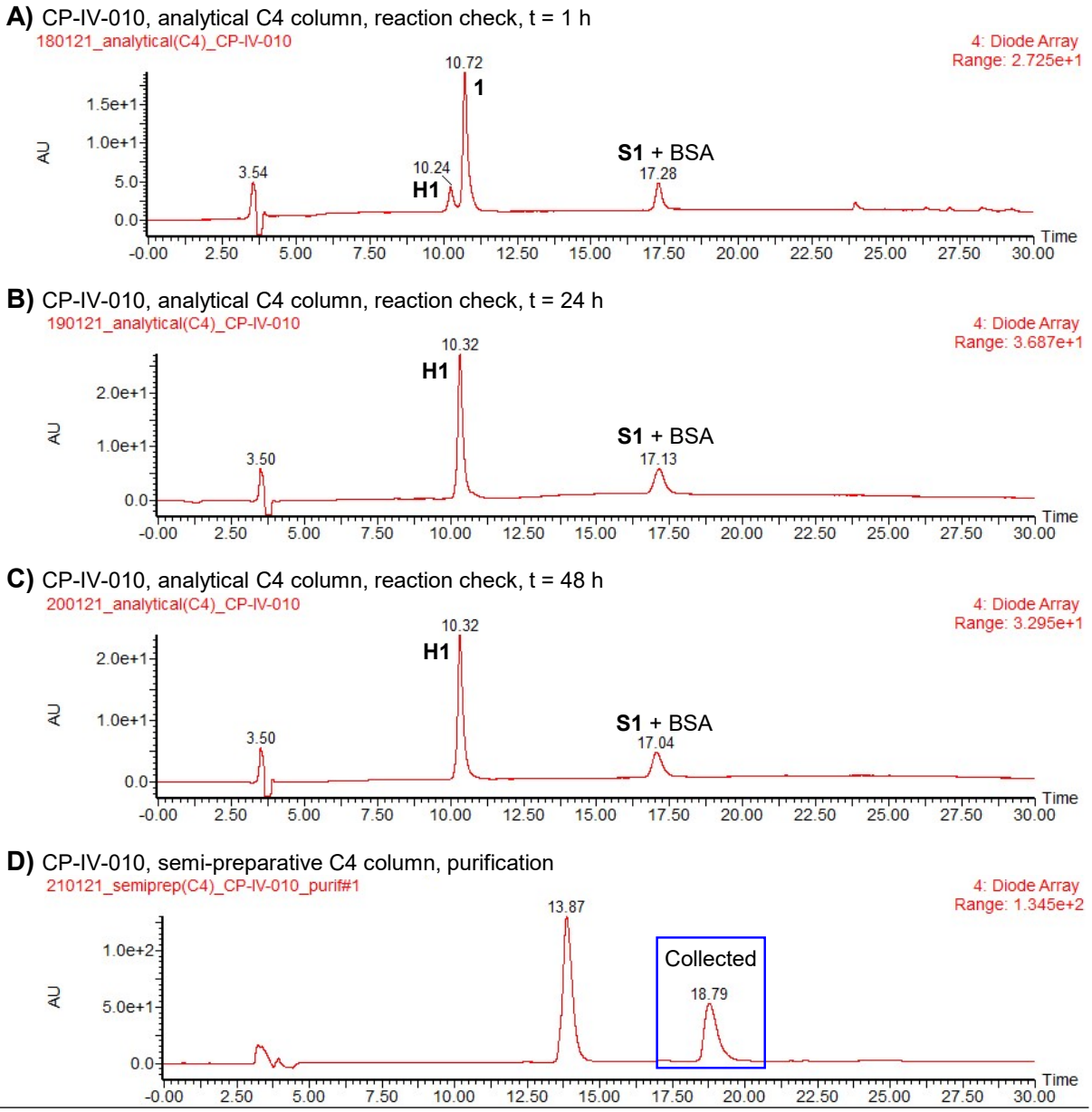


Figure S24. RP-HPLC traces for **S1** under conditions [a] (CP-IV-010). **A-C)** Analytical conditions: C4 column, 1 ml/min., 10-80% Solv. B over 30 minutes, λ = 190-400 nm. **D)** Semi-preparative conditions: C4 column, 5 ml/min., 10% Solv. B for 5 min., then 10-80% Solv. B over 25 min., λ = 190-400 nm.

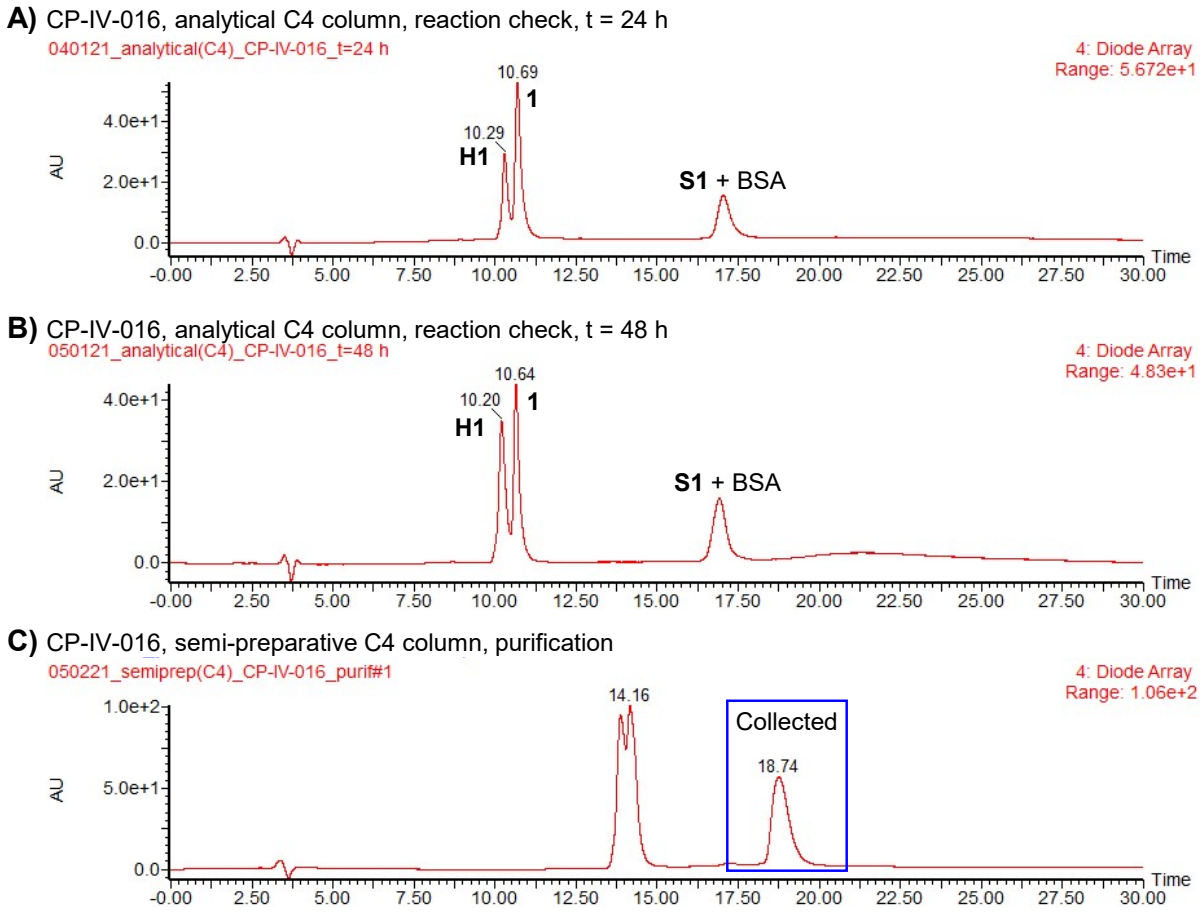


Figure S25. RP-HPLC traces for **S1** under conditions [b] (CP-IV-016). **A-B)** Analytical conditions: C4 column, 1 ml/min., 10-80% Solv. B over 30 minutes, λ = 190-400 nm. **C)** Semi-preparative conditions: C4 column, 5 ml/min., 10% Solv. B for 5 min., then 10-80% Solv. B over 25 min., λ = 190-400 nm.

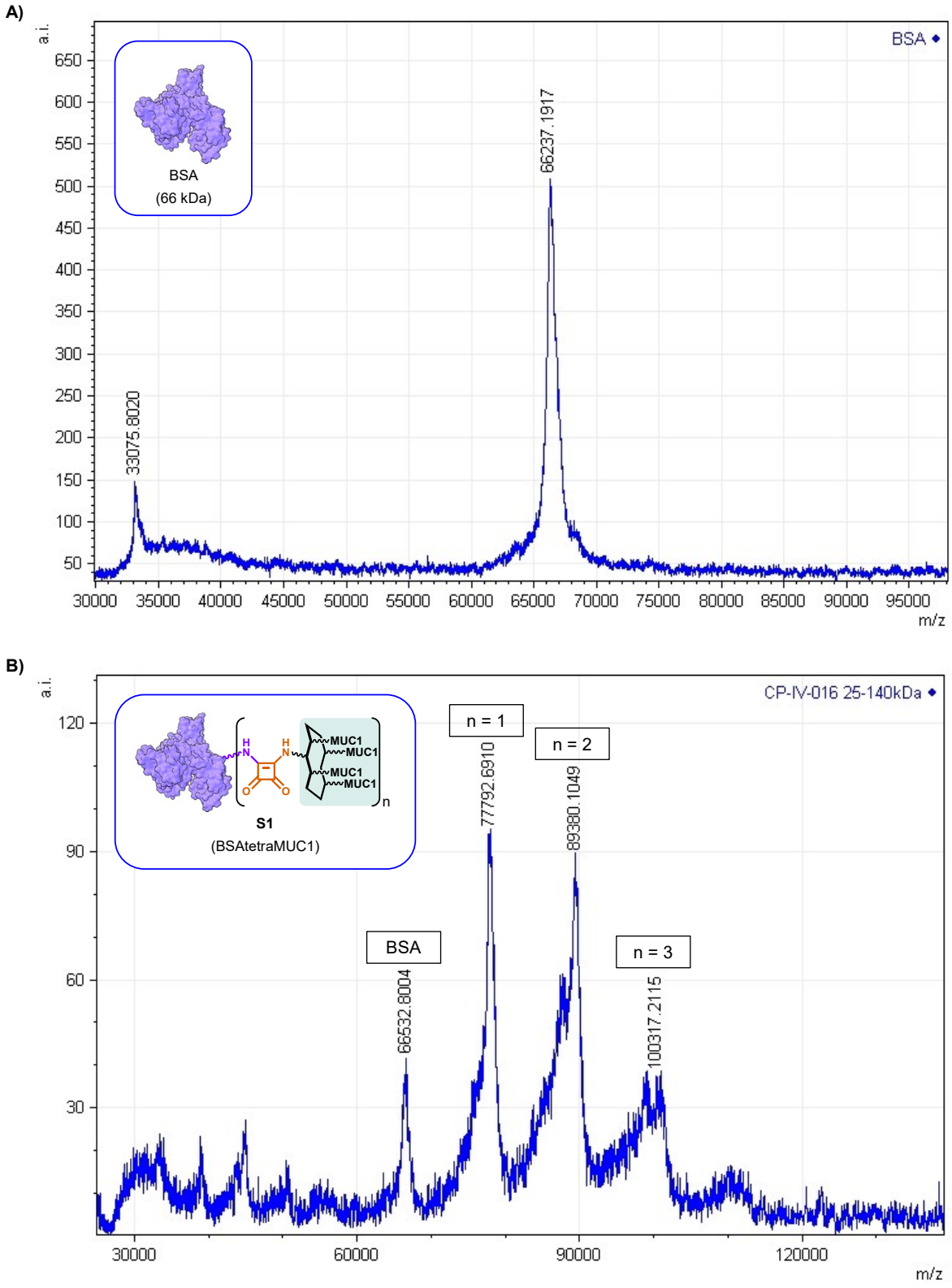


Figure S26. MALDI-TOF MS(+) spectra of **A)** BSA and **B) S1** obtained under conditions [b] (CP-IV-016).

N.B. Peak labels are centroid m/z values from linear-mode MALDI-TOF. Apparent masses may shift by hundreds of Daltons at 60–100 kDa due to Na^+/K^+/matrix adduction and centroiding of broad peaks; ladder spacing between n = 0–3 can therefore be slightly non-uniform even when the underlying chemical increment is fixed

5- Synthesis and characterization of CRM–tetraMUC1 (10) and CRM-monoMUC1 (11)

Synthetic procedure for CRM–tetraMUC1, **10** (CP-IV-034):

To an Eppendorf were 2.00 mg of CRM_{197} (34 nmol, Fina Biosolutions, LLC) and 7.02 mg of tetraMUC1 unit **1** (604 nmol) were previously lyophilized together, 300 µL of a 0.07 M $Na_2B_4O_7$ / 0.035 M $KHCO_3$ buffer (pH 9.5) were added. To the resulting suspension, 50 µL of EtOH were added, improving the overall solubility. The mixture was stirred in an Eppendorf ThermoMixer® at room temperature, purified by RP-HPLC on a semi-preparative Nucleosil-C4 column [5 ml/min., 30-100% over 30 min.] and the collected fractions lyophilized to provide 0.85 mg of a white powder.

Synthetic procedure for CRM–monoMUC1, **11** (CP-IV-036):

To an Eppendorf were 1.50 mg of CRM_{197} (26 nmol) and 2.05 mg of monoMUC1 unit **2** (541 nmol) were previously lyophilized together, 225 µL of a 0.07 M $Na_2B_4O_7$ / 0.035 M $KHCO_3$ buffer (pH 9.5) were added. In this case, the mixture was a clear solution, which was then stirred in an Eppendorf ThermoMixer® at room temperature, purified by RP-HPLC on a semi-preparative Nucleosil-C4 column [5 ml/min., 30-100% over 30 min.] and the collected fractions lyophilized to provide 1.64 mg of a white powder.

Observations: while Coomassie-stained SDS-PAGE analysis (Fig. S27) for the above reactions was consistent to what was observed for BSA conjugate **S1**, MALDI-TOF analysis of CRM conjugates **10** and **11** were performed under the same parameters used for **S1**, but without success (Fig. S28). We submitted our samples to another mass spectrometry facility, but only construct **11** provided appreciable signals (Fig. S29).

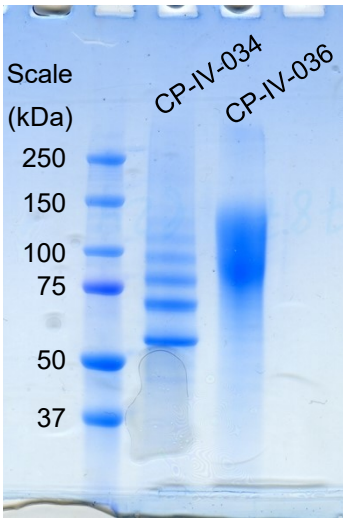


Figure S27. Coomassie-stained SDS-PAGE analysis of purified conjugates **10** (CP-IV-034) and **11** (CP-IV-036).

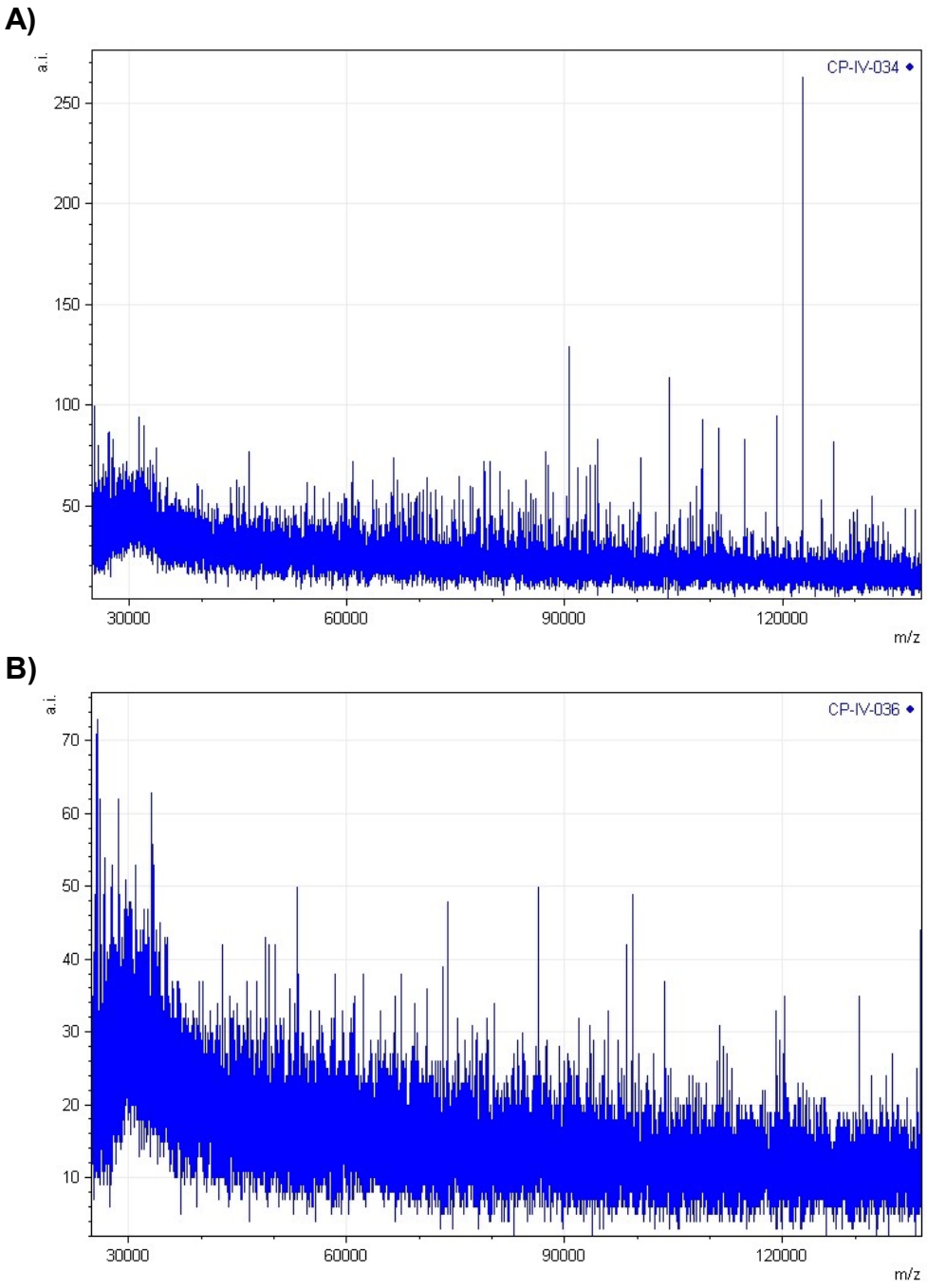


Figure S28. MALDI-TOF MS(+) spectra of **A) 10** (CP-IV-034) and **B) 11** (CP-IV-036).

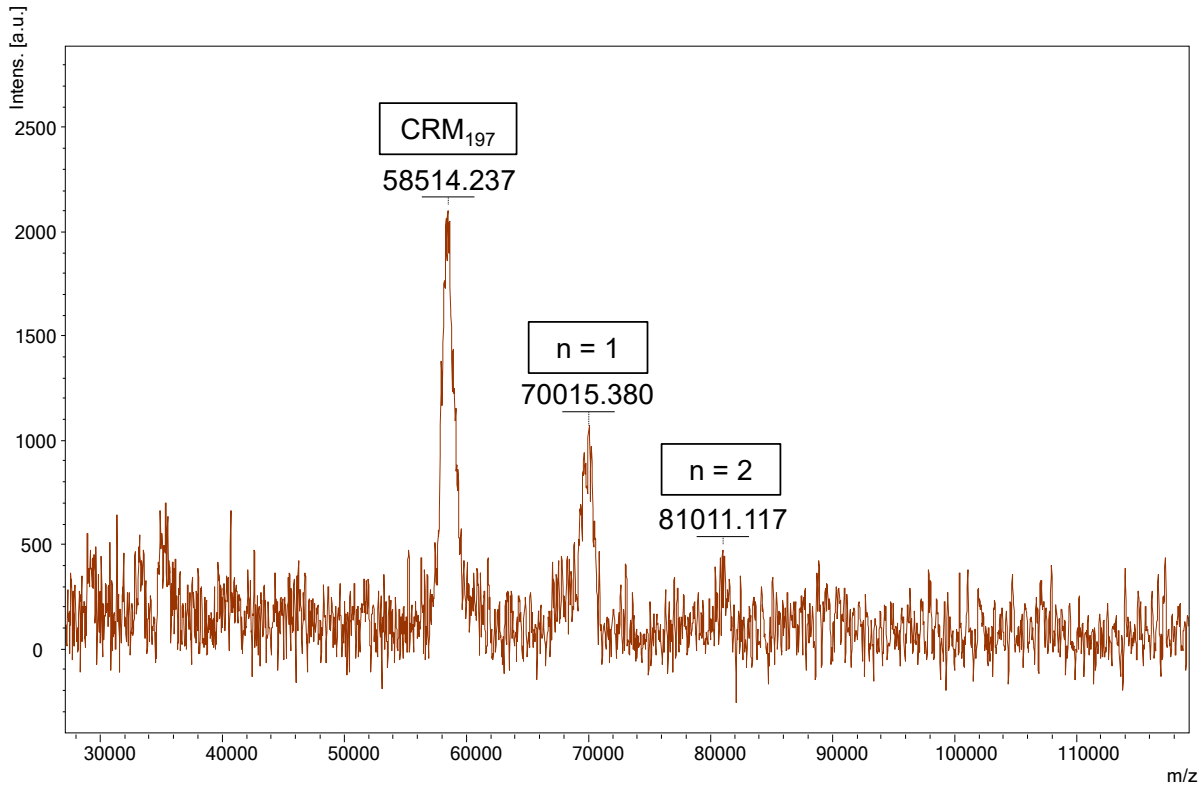


Figure S29. MALDI-TOF MS(+) spectrum of **11** (CP-IV-036).

N.B. Peak labels are centroid m/z values from linear-mode MALDI-TOF. Apparent masses may shift by hundreds of Daltons at 60–100 kDa due to Na^+/K^+/matrix adduction and centroiding of broad peaks; ladder spacing between n = 0–3 can therefore be slightly non-uniform even when the underlying chemical increment is fixed

6- Immunological evaluation

6-1 Mouse immunization

Groups of five mice (C57BL/6, female, 6-8 weeks old) were administered each immunization sample (Fig. 2A, see manuscript main text) diluted in PBS (10 mM, pH 7.4, 100 µL) via three subcutaneous injections. 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 time sera were harvested and stored at –20 °C until further analysis.

All groups were administered with 20 µg of immunogen. Aliquots were prepared from stock solutions in 10 mM PBS pH 7.4, obtained by weighing constructs **1**, **2**, **10** and **11** on a precision balance.

MUC1 glycopeptide antigen (PAHGVTSAPDTRPAPGST*APPA, T* = αGalNAc-*O*-Thr, M.W. 2258) doses corresponding to 20 µg of immunogen were calculated as follows.

- TetraMUC1 **1** (M.W. 11618): 20 µg = 0.0017 µmol. Glycopeptide content = 0.0017 µmol x 4 copies x 2258 = 15 µg
- MonoMUC1 **2** (M.W. 3792): 20 µg = 0.0053 µmol. Glycopeptide content = 0.0053 µmol x 2258 = 12 µg

For immunogens CRM–tetraMUC1 **10** (Groups A and E) and CRM–monoMUC1 **11** (Group B), doses of glycopeptide antigen per injection were roughly estimated from SDS-PAGE analysis (Figure S27).

- For the CRM–tetraMUC1 **10** immunogen, due to the presence of unreacted CRM_{197}, the average antigen loading was estimated as one tetraMUC1 copy per protein (M.W. 69976): 20 µg = 0.00029 µmol. Glycopeptide content = 0.00029 µmol x 4 copies x 2258 = 2.6 µg.
- For the CRM-monoMUC1 **11** immunogen, whereby all starting CRM_{197} was reacted, we estimated a mean M.W. of 100 kDa, corresponding to the middle point of the SDS-PAGE band. According to the above estimation, we considered that ≈11 copies of monoMUC1 per protein were conjugated (M.W. 99610): 20 µg = 0.00020 µmol. Glycopeptide content = 0.00020 x 11 copies x 2258 = 5.0 µg.

Table S1. Immunogen, MUC1 glycopeptide, and vaccine adjuvant doses administered per injection.

Group	Immunogen	Immunogen dose per injection (µg)	Glycopeptide antigen dose per injection (µg)	QS-21 adjuvant dose per injection (µg)
A	CRM–tetraMUC1 (**10**)	20	≈2.6	20
B	CRM–monoMUC1 (**11**)	20	≈5.0	20
C	tetraMUC1 (**1**)	20	15	20
D	monoMUC1 (**2**)	20	12	20
E	CRM–tetraMUC1 (**10**)	20	≈2.6	-

Ethical Statement. Animals 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 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 International.

Safety and Tolerability. No toxic side effects (*e.g.* local inflammation, systemic reactions, mouse weight loss or death) were observed over the course of the immunizations, indicating the non-toxicity of all vaccine constructs. While CRM_{197} is a clinically validated carrier, and our conjugates (**10** and **11**) as well as scaffolds (**1** and **2**) well tolerated *in vivo*, comprehensive GLP-compliant toxicology (repeat-dose, local tolerance, clinical pathology and histopathology) will still be required to support translational development.

6-2 Quantification of antibody production

Antibody titers were measured by an indirect enzyme-linked immunosorbent assay (ELISA). Briefly, ELISA plates (Thermo Scientific) were coated with BSA–tetraMUC1 conjugate **S1** (Scheme S8) at 0.5 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 at appropriate dilutions in blocking buffer and incubated for 1 h at room temperature (secondary antibody dilutions for total IgG, IgM and IgG3 = 1/1000; for IgG1 and IgG 2b = 1/4000; for IgG2c = 1/10000). KPL SureBlue Reserve™ commercial solution (100 µL/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.

ELISA plots were generated using directly measured OD values obtained from replicate samples at each dilution. For each experimental group and dilution, the mean OD was calculated, and variability was visualized as ± standard deviation (SD) around the mean. These values were plotted as line graphs with overlaid points for the mean, and shaded ribbons indicating the SD range.

Data analyses and visualizations were performed in R version 4.5.1 using RStudio version 2025.05.1 (Posit, PBC). The plots were created with the tidyverse suite (Wickham et al., (2019). Welcome to the Tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686) for data wrangling and the scales package for axis formatting.

The corresponding generic R script is reported below:

```
# 1. Packages
library(tidyverse)
library(scales)
# 2. Reusable plotting function
plot_elisa <- function(data, title_text = "ELISA Plot", file_name = NULL) {
  p <- ggplot(data, aes(x = dilution, y = mean_od, color = group, fill =
group)) +
```

```
    geom_ribbon(aes(ymin = mean_od - sd_od, ymax = mean_od + sd_od),
                alpha = 0.2, linetype = 0) +
    geom_line(linewidth = 1.2) +
    geom_point(size = 2) +
    scale_x_log10(
      breaks = 10^(2:9),
      minor_breaks = rep(1:9, each = 1) * rep(10^(2:8), each = 9),
      labels = trans_format("log10", math_format(10^.x))
    ) +
    coord_cartesian(ylim = c(0, 4)) +
    labs(
      title = title_text,
      x = "Dilution",
      y = "OD (450 nm)",
      color = "Groups",
      fill = "Groups"
    ) +
    theme_minimal(base_size = 14) +
    theme(
      panel.grid.major = element_blank(),
      panel.grid.minor = element_blank(),
      axis.line = element_line(color = "black", linewidth = 0.8),
      axis.text = element_text(color = "black"),
      axis.title = element_text(color = "black", face = "bold"),
      axis.ticks = element_line(color = "black"),
      axis.ticks.length = unit(0.2, "cm"),
      plot.title = element_text(hjust = 0.5, face = "bold"),
      legend.title = element_text(face = "bold")
    )

  if (!is.null(file_name)) {
    ggsave(file_name, plot = p, width = 6, height = 4, dpi = 300)
  }

  return(p)
}
# 3. Data preparation function
prepare_elisa_data <- function(df, group_name) {
  df %>%
    pivot_longer(-dilution, names_to = "mouse", values_to = "od") %>%
    group_by(dilution) %>%
    summarise(
      mean_od = mean(od, na.rm = TRUE),
      sd_od = sd(od, na.rm = TRUE),
      group = group_name,
      .groups = "drop"
    )
}
# 4. Example of usage
# group_X <- data.frame(
#   dilution = c(...),
#   mouse1 = c(...),
#   mouse2 = c(...),
#   ...
# )
# data_X <- prepare_elisa_data(group_X, "Group Label")
# all_data <- bind_rows(data_X, ...)
#   plot_elisa(all_data,   title_text   =   "Example   ELISA",   file_name   =
"example_elisa_plot.png")
```

The area under the curve (AUC) for IgG subclasses (IgG1, IgG2b, IgG2c, and IgG3) represented in Figure 2 was calculated with GraphPad Prism using the trapezoidal rule with baseline set to zero. For each group, the total area value along with the standard error is reported (Table S2). To assess differences between groups, ordinary one-way ANOVA followed by Tukey's multiple comparisons test was performed (Table S2).

Table S2. For each IgG subclass, the calculated AUC values (total area ± SEM) together with the results of pairwise comparisons (adjusted p-values and significance levels) are reported.

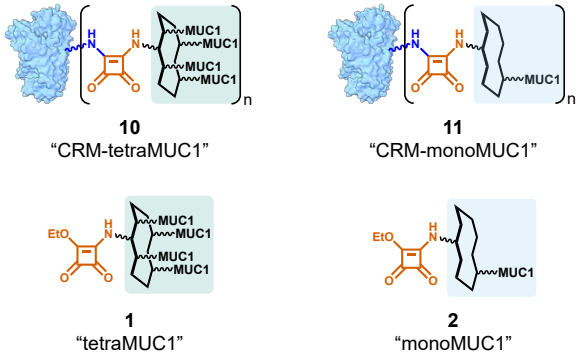
IgG1		
Group	Total area	Std. error
A	22227660	2545818
B	20202397	2482972
C	19111	623.1
D	13788	755.4
E	728333	42071
ANOVA Tukey multiple comparisons summary		
Comparison	Adjusted P Value	Significance
C vs. D	>0.9999	ns
C vs. E	0.9977	ns
C vs. A	<0.0001	****
C vs. B	<0.0001	****
D vs. E	0.9976	ns
D vs. A	<0.0001	****
D vs. B	<0.0001	****
E vs. A	<0.0001	****
E vs. B	<0.0001	****
A vs. B	0.8934	ns

IgG2b		
Group	Total area	Std. error
A	14336607	1186083
B	17220507	11522682
C	12271	492.9
D	16577	417.7
E	41931	6765
ANOVA Tukey multiple comparisons summary		
Comparison	Adjusted P Value	Significance
C vs. D	>0.9999	ns
C vs. E	>0.9999	ns
C vs. A	0.3223	ns
C vs. B	0.1709	ns
D vs. E	>0.9999	ns
D vs. A	0.3226	ns
D vs. B	0.171	ns
E vs. A	0.3243	ns
E vs. B	0.1721	ns
A vs. B	0.9945	ns

IgG2c		
Group	Total area	Std. error
A	11553753	591213
B	11928134	728888
C	13855	329.9
D	13509	1155
E	24061	7546
ANOVA Tukey multiple comparisons summary		
Comparison	Adjusted P Value	Significance
C vs. D	>0.9999	ns
C vs. E	>0.9999	ns
C vs. A	<0.0001	****
C vs. B	<0.0001	****
D vs. E	>0.9999	ns
D vs. A	<0.0001	****
D vs. B	<0.0001	****
E vs. A	<0.0001	****
E vs. B	<0.0001	****
A vs. B	0.9682	ns

IgG3		
Group	Total area	Std. error
A	56840	5181
B	38332	7154
C	11949	767.5
D	21997	5833
E	20045	2099
ANOVA Tukey multiple comparisons summary		
Comparison	Adjusted P Value	Significance
C vs. D	0.5932	ns
C vs. E	0.7606	ns
C vs. A	<0.0001	****
C vs. B	0.0078	**
D vs. E	0.9984	ns
D vs. A	0.0005	***
D vs. B	0.1597	ns
E vs. A	0.0003	***
E vs. B	0.0943	ns
A vs. B	0.0886	ns

Group	Formulation
A	CRM-tetraMUC1 (**10**) + QS21
B	CRM-monoMUC1 (**11**) + QS21
C	tetraMUC1 (**1**) + QS21
D	monoMUC1 (**2**) + QS21
E	CRM-tetraMUC1 (**10**)



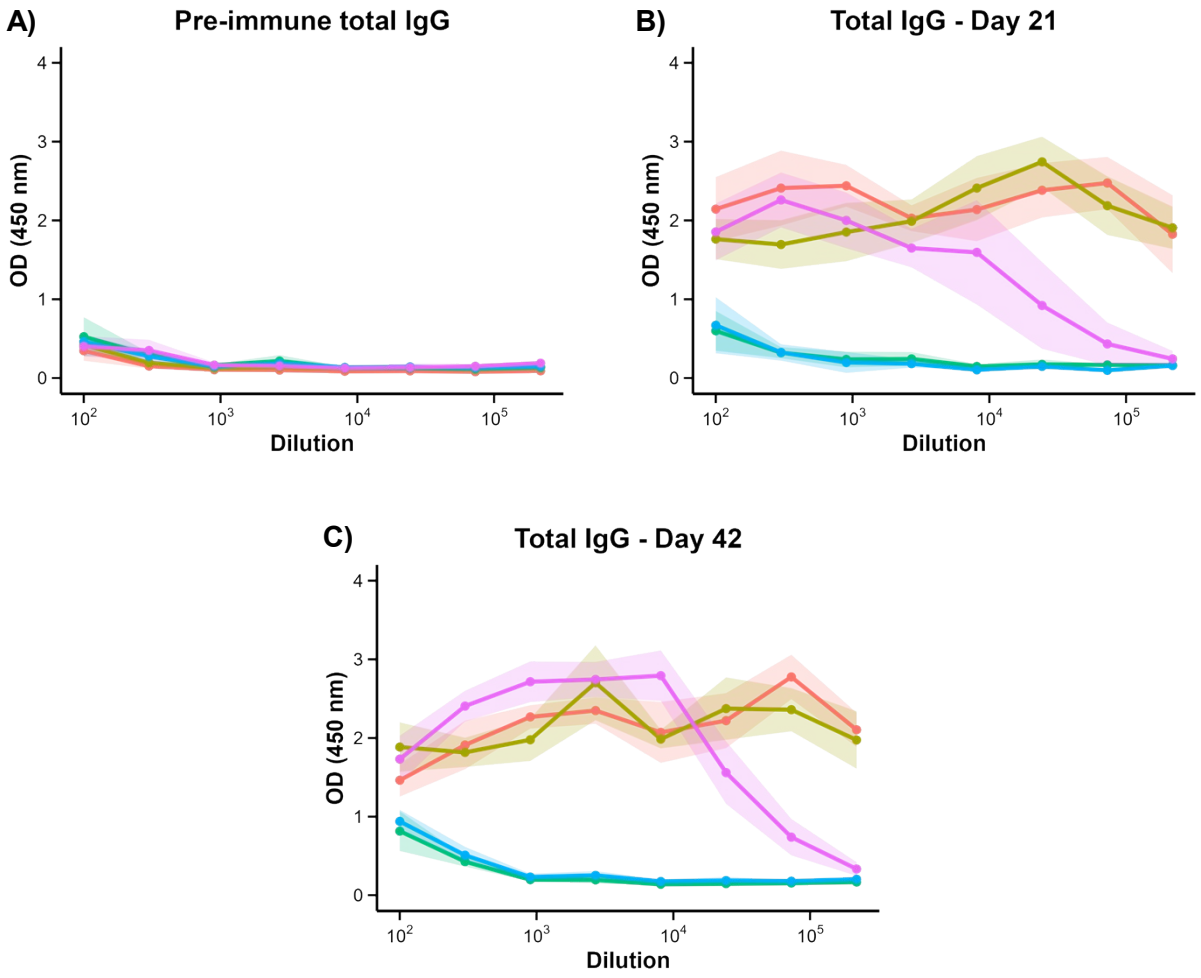
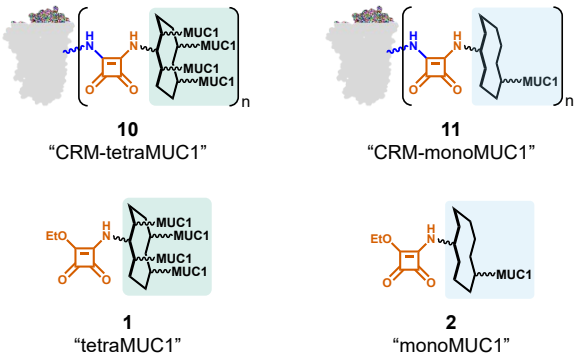


Figure S30. ELISA dilution curves for total IgG with curves representing mean OD_{450} values from five mice per group. **A)** Pre-immune sera showing negligible baseline reactivity. **B)** Day 21 sera. **C)** Day 42 sera. Sera from Groups A and B displayed a pronounced hook (prozone) effect across the 10^2–$2{\cdot}10^5$ dilution range.

Error visualization was achieved by plotting the mean as a line graph with overlaid points, while shaded ribbons represent inter-individual variability (± standard deviation, SD). ELISA plates were coated with the BSAtetraMUC1 construct **S1** (Scheme S8).

Group	Formulation
A	CRM-tetraMUC1 (**10**) + QS21
B	CRM-monoMUC1 (**11**) + QS21
C	tetraMUC1 (**1**) + QS21
D	monoMUC1 (**2**) + QS21
E	CRM-tetraMUC1 (**10**)



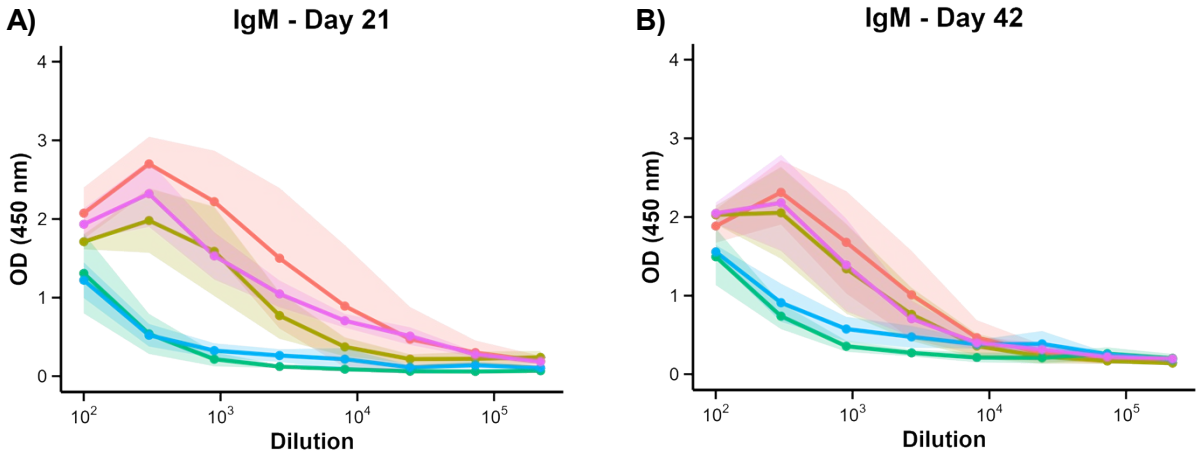
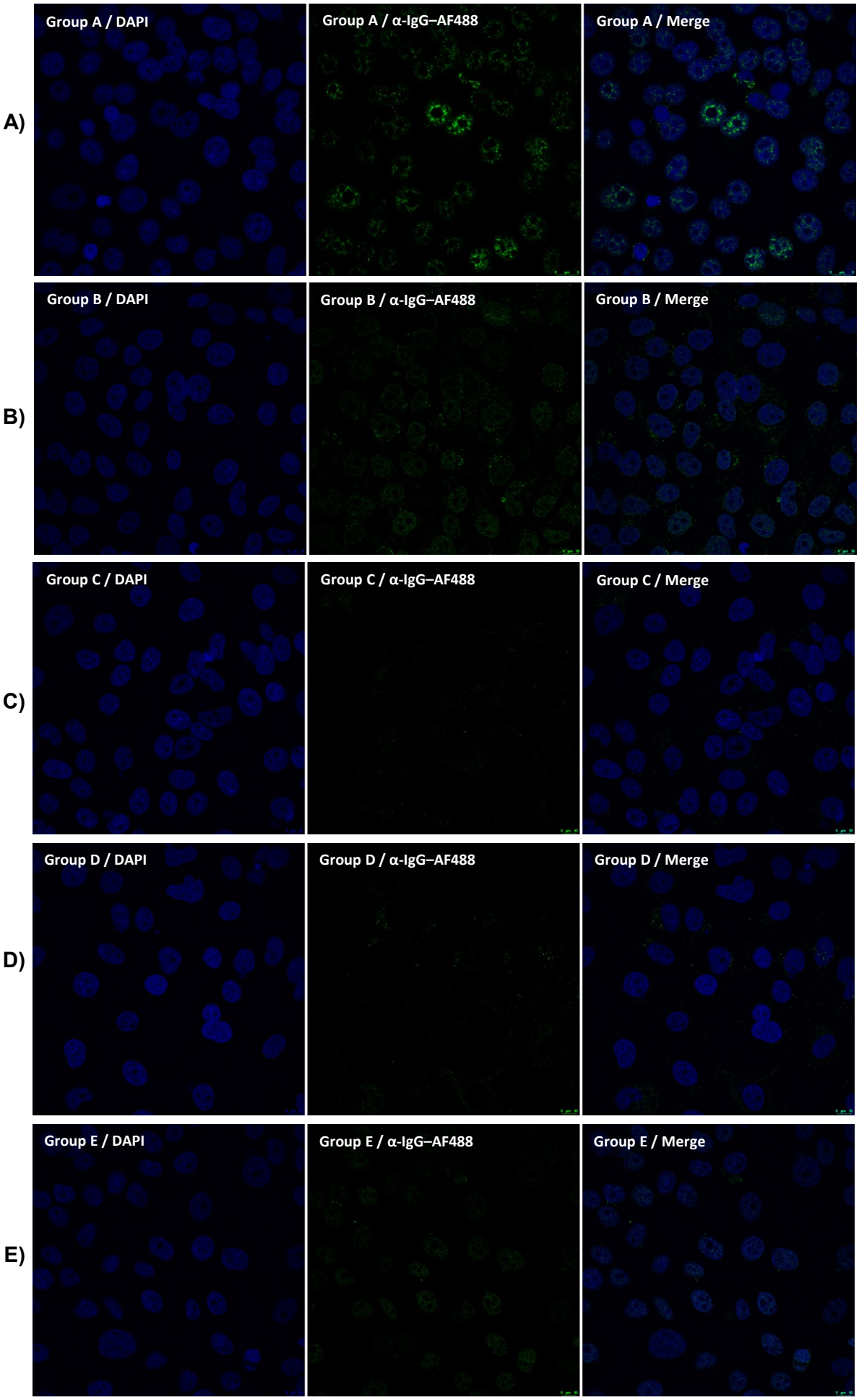


Figure S31. ELISA dilution curves for IgM with curves representing mean OD_{450} values from five mice per group. **A)** Day 21 sera. **B)** Day 42 sera, with curves representing mean OD_{450} values from five mice per group. Dose-response curves were plotted starting from the dilution at which signal decline was first observed.

Error visualization was achieved by plotting the mean as a line graph with overlaid points, while shaded ribbons represent inter-individual variability (± standard deviation, SD). ELISA plates were coated with the BSAtetraMUC1 construct **S1** (Scheme S8).

6-3 Cell-surface reactivity of antisera against MUC1-expressing MCF7 cells by immunofluorescence

MCF7 cells 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 ($2.5 \cdot 10^5$ 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 hour at room temperature, and then with sera from immunized mice at a 1/100 dilution in blocking buffer (500 µL/well) for 2 hours at 4 °C. Individual animal sera were used as representative examples of the respective group. The cells were extensively washed with PBS buffer and FITC-labeled anti-mouse IgG-AF488 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, they were washed thoroughly and fixed with 3.7 % formaldehyde in PBS for 20 minutes at 4 °C. Subsequently, the samples were incubated for 10 minutes at 4 °C with DAPI (Invitrogen, 0.3 µM, 0.105 µg/mL) for nuclear staining. Following extensive washing steps, the glass coverslips were carefully placed onto microscope slides using 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) and Image J.


A)
Group A / DAPI
Group A / α-IgG–AF488
Group A / Merge
B)
Group B / DAPI
Group B / α-IgG–AF488
Group B / Merge
C)
Group C / DAPI
Group C / α-IgG–AF488
Group C / Merge
D)
Group D / DAPI
Group D / α-IgG–AF488
Group D / Merge
E)
Group E / DAPI
Group E / α-IgG–AF488
Group E / Merge

F)

Figure S32. Confocal immunofluorescence microscopy of MCF7 cells incubated with sera from immunized mice. Cells were stained with sera collected at day 42 from the indicated groups (A–E) at 1:100 dilution, followed by FITC-conjugated anti-mouse IgG (green); nuclei were counterstained with DAPI (blue). Pre-immune sera were used as negative control.

6-4 *Ex vivo* cellular characterization

At sacrifice (day 42), spleens were aseptically harvested from immunized mice and processed into single-cell suspensions by mechanical dissociation through a 70 µm cell stainer in cold PBS supplemented with 1 % FBS. Red blood cells were removed by ACK lysis, and splenocytes were resuspended in TexMACS medium (Miltenyi Biotec) at a final density of 10^7 cells/mL. For *ex vivo* restimulation, $3{\cdot}10^6$ splenocytes per well were seeded in 48-well plates and cultured for 72 hours at 37 °C and 5 % CO_2 in the presence of the corresponding immunogen utilized for vaccination: CRM–tetraMUC1 **10** (Group A and Group E), CRM–monoMUC1 **11** (Group B), tetraMUC1 **1** (Group C), and monoMUC1 **2** (Group D). Following stimulation, cells were collected and stained for flow cytometric analysis using fluorophore-conjugated monoclonal antibodies against CD4, CD8, CD44, CD62L, and CD45RB. The following fluorophore–antibody combinations were employed: CD4–APC-Cy7, CD8–PE-Cy7, CD44–PerCP-Cy5.5, CD62L–APC, and CD45RB–PE (all from BioLegend). The frequencies of effector $CD4^+$ T cells ($CD4^{high}CD44^{high}CD45RB^{low}$) and effector $CD8^+$ T cells ($CD8^{high}CD44^{high}CD62L^{low}$) were quantified. Data are expressed as mean ± SEM (n = 5 mice/group). Statistical differences between Groups were determined by one-way ANOVA followed by Tukey's post hoc test (GraphPad Prism), with $P < 0.05$ (*), and $P < 0.01$ (**) considered significant.