ChemBioChem

Supporting Information

Chemoenzymatic Synthesis of Glycopeptides to Explore the Role of Mucin 1 Glycosylation in Cell Adhesion

Claudia Bello,* Erica Pranzini, Emanuele Piemontese, Maximilian Schrems, Maria Letizia Taddei, Lisa Giovannelli, Mario Schubert, Christian F. W. Becker, Paolo Rovero, and Anna Maria Papini

Abstract

Posttranslational modifications affect protein biology in physiological and pathological conditions. Efficient methods for the preparation of peptides and proteins carrying defined, homogeneous modifications are fundamental tools for investigating these functions. In the case of Mucin 1 (MUC1) an altered glycosylation pattern is observed in carcinogenesis. To better understand the role of MUC1 glycosylation in interaction and adhesion of cancer cells, we prepared a panel of homogeneously *O*-glycosylated MUC1 peptides using a quantitative chemoenzymatic approach. Cell adhesion experiments with MCF-7 cancer cells on surfaces carrying up to 6 differently glycosylated MUC1 peptides demonstrated that different glycans significantly impact adhesion. This finding suggests a distinct role of MUC1 glycosylation patterns in cancer cell migration and/or invasion. To decipher the molecular mechanism for the observed adhesion we investigated conformation of the glycosylated MUC1 peptides by NMR. These experiments revealed only minor differences in peptide structure, therefore clearly relating the adhesion behaviour to the type and number of glycans linked to MUC1.

Table of content

Materials and methods

Fmoc-His(Trt) TentaGel® R PHB resin (0.19 mmol/g) was purchased from Rapp Polymere GmbH (Tübingen, Germany). 9-Fluorenylmethoxycarbonyl (Fmoc)-protected amino acids (Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Leu-OH, Fmoc-Phe-OH, Fmoc-Ser(tBu)-OH, Fmoc-Pro-OH, Fmoc-Thr(tBu)-OH, Fmoc-Tyr(tBu)-OH, Fmoc-Val-OH) were purchased from NovaBiochem (Darmstadt, Germany), Iris Biotech (Marktredwitz, Germany), CBL (Patras, Greece), and Carbosynth (Berkshire, UK). Fmoc-Thr[GalNAc($OAc)_3$]-OH and Fmoc-Ser[GalNAc($OAc)_3$]-OH were purchased from Sussex Research (Ottawa, Canada) and from Bachem (Basel, Switzerland). Fmoc-NH-$(PEG)_{27}$-COOH was purchased from Polypure AS (Oslo, Norway). N,N,N',N'-Tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate (HBTU) was purchased from Advanced Biotech Italia (Milan, Italy) and 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU) from Carbosynth (Berkshire, UK). Diisopropylethylamine (DIEA) was purchased from Acros Organics (Geel, Belgium). N,N-dimethylformamide, piperidine and trifluoroacetic acid were purchased from Carlo Erba (Milan, Italy). Hydrazine monohydrate was purchased from TCI Europe (Zwijndrecht, Belgium). Dichloromethane, acetonitrile, methanol and diethyl ether were purchased from Sigma Aldrich (Darmstadt, Germany). Primo® Cell Culture multiwell plates 24 wells (cat ET3024) were purchased from Euroclone (Milan, Italy).

Purification methods: Semipreparative Waters Separation Module 2695, detector diode array 2996 column Sepax Bio-C18. Biotage Isolera One equipped with Snap Cartridge, KP-Sil 10g. In both cases mixtures of eluent A (0.1% (v/v) TFA in H_2O) and eluent B (84% (v/v) CH_3CN in H_2O +0.1% (v/v) TFA) were used as mobile phases. Waters 2545 Quarternary Gradient Module with a Waters 2489 UV/Vis detector and a Waters Fraction Collector III, Semiprep Kromasil 300-10-C4 column 10x250 mm, mixtures of eluent A (0.1% (v/v) TFA in H_2O) and solution B (0.08% (v/v) TFA in CH_3CN) were used as mobile phases.

Characterization methods: HPLC system Waters Alliance 2695 Separations Module, column Bioshell: A160 Peptide C18 (10 cm x 3 mm x 2.7 µm), coupled to a UV 2996 PDA detector and to an ESI-MS single quadrupole Micromass ZQ detector; Dionex Ultimate 3000 with CAD detector; Waters AutoPurification System (used for direct injection MS analyses); Dionex Ultimate 3000, column Waters X-select CSH C18 XP (3 mm x 75 mm x 2.5 µm) coupled to Thermo Scientific MSQ Plus single quadrupole ESI-MS detector (mobile phases: mixture of eleuent A (0.1% (v/v) formic acid in H_2O) and eluent B (0.08% (v/v) formic acid CH_3CN)); Thermo scientific Vanquish Flex HPLC with UV/Vis detector, column Macherey-Nagel 150/4.6 Nucleodur 300-5 C4ec. In all cases, except in the one specified above, mixtures of eluent A (0.1% (v/v) TFA in H_2O) and eluent B (0.08% (v/v) TFA in CH_3CN) were used as mobile phases.

Synthesis of peptides 1 - 4

Peptides **1**-**4** (GVTSAPDTRPAPGSTAPPAH (**1**), GVTSAPDTRPAPGST(GalNAc)APPAH (**2**), GVTSAPDT(GalNAc)RPAPGST(GalNAc)APPAH (**3**), GVTS(GalNAc)APDT(GalNAc)RPAPGST(GalNAc)APPAH (**4**)) were synthesized following the general procedure for solid phase peptide synthesis (SPPS) reported in the main text. Fmoc-His(Trt) TentaGel® R PHB resin (0.19 mmol/g, 0.1 mmol) was elongated using the above mentioned procedure, to get the non-glycosylated peptidyl resin. The three modified peptidyl resins carrying one, two and, respectively, three GalNAc monosaccharides, were prepared starting from the same batch of Fmoc-His(Trt) TentaGel® R PHB resin (0.19 mmol/g, 0.3 mmol). The peptide was elongated till Ala16 and deprotected at the N-terminus, then Fmoc-Thr[GalNAc($OAc)_3$]-OH was coupled (1.5 equiv amino acid, 1.35 equiv HATU (0.3M solution in DMF), 2.5 equiv DIEA, 40 min; double coupling) to introduce the glycosylated amino acid. The peptide was elongated till Arg9, then 1/3 of the resin was separated and elongated using Fmoc-protected amino acids (see Materials and methods above) till completion of the sequence. Fmoc-

Thr[GalNAc(OAc)$_3$]-OH was coupled (1.5 equiv amino acid, 1.35 equiv HATU (0.3M solution in DMF), 2.5 equiv DIEA, 40 min; double coupling) to the remaining 2/3 of initial peptidyl resin. Elongation with Fmoc-protected amino acids (see Materials and methods above) was performed till Ala5. The resin was split in two, and Fmoc-Ser[GalNAc(OAc)$_3$]-OH was coupled to one half of the resin (1.5 equiv amino acid, 1.35 equiv HATU (0.3M solution in DMF), 2.5 equiv DIEA, 40 min; double coupling), followed by elongation with Fmoc-protected amino acids till completion of the sequence. Fmoc-Ser(tBu)-OH and then Fmoc-protected amino acids were coupled to the other half of the peptidyl resin, producing the peptidyl resin with three GalNAc monosaccharides protected as acetate on the hydroxyl groups. The resins were dried under vacuum. Final Fmoc removal was performed on 50 mg of peptidyl resin GVT(tBu)S(tBu)APD(tBu)T(tBu)R(Pbf)PAPGS(tBu)T(tBu)APPAH(Trt)-TentaGel® R PHB , 50 mg of peptidyl resin GVT(tBu)S(tBu)APD(tBu)T(tBu)R(Pbf)PAPGS(tBu)T[GalNAc(OAc)$_3$]APPAH(Trt)-TentaGel® R PHB, 54 mg of peptidyl resin GVT(tBu)S(tBu)APD(tBu)T[GalNAc(OAc)$_3$]R(Pbf)PAPGS(tBu)T[GalNAc(OAc)$_3$]APPAH(Trt)-TentaGel® R PHB, and 53 mg of peptidyl resin GVT(tBu)S[GalNAc(OAc)$_3$]APD(tBu)T[GalNAc(OAc)$_3$]R(Pbf)P-APGS(tBu)T[GalNAc(OAc)$_3$]APPAH(Trt)-TentaGel® R PHB), followed by deprotection and cleavage (general procedure in the main manuscript) to get crude peptides **1**, **2**, **3** and **4**.

Each crude (glyco)peptide (GVTSAPDTRPAPGSTAPPAH (**1**): 9 mg, 4.8 µmol; GVTSAPDTRPAPGST(GalNAc)APPAH (**2**): 12 mg, 5.7 µmol; GVTSAPDT(GalNAc)RPAPGST(GalNAc)APPAH (**3**): 12 mg, 5.2 µmol; GVTS(GalNAc)APDT(GalNAc)RPAPGST(GalNAc)APPAH (**4**): 15.1 mg, 6.0 µmol) was purified via semipreparative RP-HPLC, using a gradient of 5-40% eluent B in A in 30 min (Eluents: A = 0.1% (v/v) TFA in H_2O, B = 84% (v/v) CH_3CN in H_2O +0.1% (v/v) TFA). For each peptide, fractions containing the desired product were collected and freeze-dried giving peptide **1** in 15% yield (1.3 mg, 0.7 µmol) (Figure S1), peptide **2** in 19% yield (2.3 mg, 1.1 µmol) (Figure S2), peptide **3** in 27% yield (3.1 mg, 1.4 µmol) (Figure S3) and peptide **4** in 35% yield (5.4 mg, 2.1 µmol), yields are based on crude peptides. Fractions containing the product with impurities were collected separately, freeze-dried and stored at -20°C.

RP-HPLC analysis of GVTSAPDTRPAPGSTAPPAH (**1**): Rt: 3.3 min. Bioshell: A160 Peptide C18 column (10 cm x 3 mm x 2.7 µm), temperature 35°C; flow 0.5 mL/min; eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B); gradient 5-60% B in A, in 5 min. ESI-MS (m/z): $[M+H]^+$ 1887.6 (found), $[M+H]^+$ 1887.0 (calcd); $[M+2H]^{2+}$ 944.2 (found), $[M+2H]^{2+}$ 944.5 (calcd); $[M+3H]^{3+}$ 629.9 (found), 630.0 (calcd).

RP-HPLC analysis of GVTSAPDTRPAPGST(GalNAc)APPAH (**2**): Rt: 3.0 min. Bioshell: A160 Peptide C18 column (10 cm x 3 mm x 2.7 µm), temperature 35°C; flow 0.5 mL/min; eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B); gradient 5-60% B in A, in 5 min. ESI-MS (m/z): $[M+2H]^{2+}$ 1046.0 (found), $[M+2H]^{2+}$ 1046.1 (calcd); $[M+3H]^{3+}$ 697.6 (found), 697.7 (calcd); $[M+4H]^{4+}$ 523.4 (found), $[M+4H]^{4+}$ 523.6 (calcd).

RP-HPLC analysis of GVTSAPDT(GalNAc)RPAPGST(GalNAc)APPAH (**3**): Rt: 3.1 min. Bioshell: A160 Peptide C18 column (10 cm x 3 mm x 2.7 µm), temperature 35°C; flow 0.5 mL/min; eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B); gradient 5-60% B in A, in 5 min. ESI-MS (m/z): $[M+2H]^{2+}$ 1147.6 (found), $[M+2H]^{2+}$, 1147.7 (calcd); $[M+3H]^{3+}$ 765.4 (found), 765.5 (calcd); $[M+4H]^{4+}$ 574.2 (found), $[M+4H]^{4+}$ 574.5 (calcd).

RP-HPLC analysis of GVTS(GalNAc)APDT(GalNAc)RPAPGST(GalNAc)APPAH (**4**): Rt: 3.1 min. Bioshell: A160 Peptide C18 column (10 cm x 3 mm x 2.7 µm), temperature 35°C; flow 0.5 mL/min; eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B); gradient 5-60% B in A, in 5 min. ESI-MS (m/z): $[M+2H]^{2+}$ 1249.2 (found), $[M+2H]^{2+}$ 1249.3 (calcd); $[M+3H]^{3+}$ 833.1 (found), 833.2 (calcd); $[M+4H]^{4+}$ 625.1 (found), $[M+4H]^{4+}$ 625.2 (calcd).

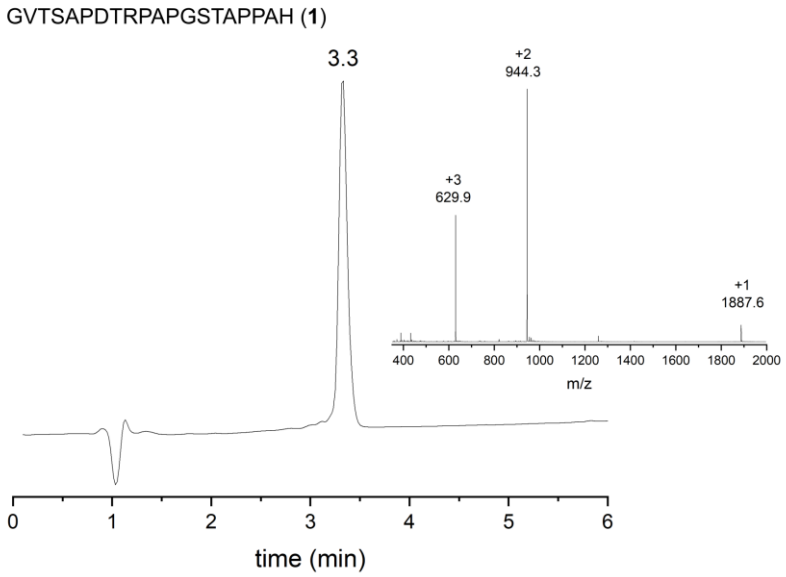

Figure S1. RP-HPLC-ESI-MS analysis of peptide **1**. Eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B). Gradient 5-60% B in A, in 5 min. $[M+H]^+$ 1887.6 (found), $[M+H]^+$ 1887.0 (calcd); $[M+2H]^{2+}$ 944.2 (found), $[M+2H]^{2+}$ 944.5 (calcd); $[M+3H]^{3+}$ 629.9 (found), 630.0 (calcd).

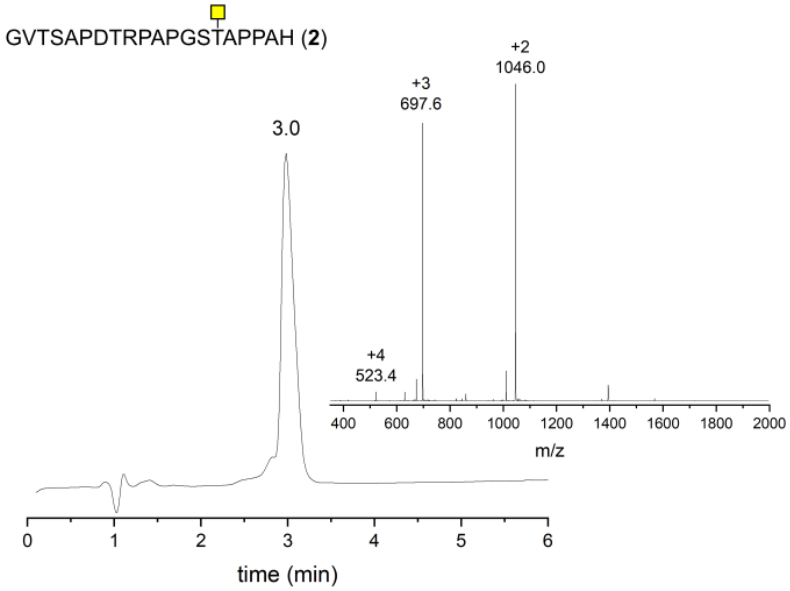

Figure S2. RP-HPLC-ESI-MS analysis of glycopeptide **2**. Eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B). Gradient 5-60% B in A, in 5 min. $[M+2H]^{2+}$ 1046.0 (found), $[M+2H]^{2+}$ 1046.1 (calcd); $[M+3H]^{3+}$ 697.6 (found), 697.7 (calcd); $[M+4H]^{4+}$ 523.4 (found), $[M+4H]^{4+}$ 523.6 (calcd).

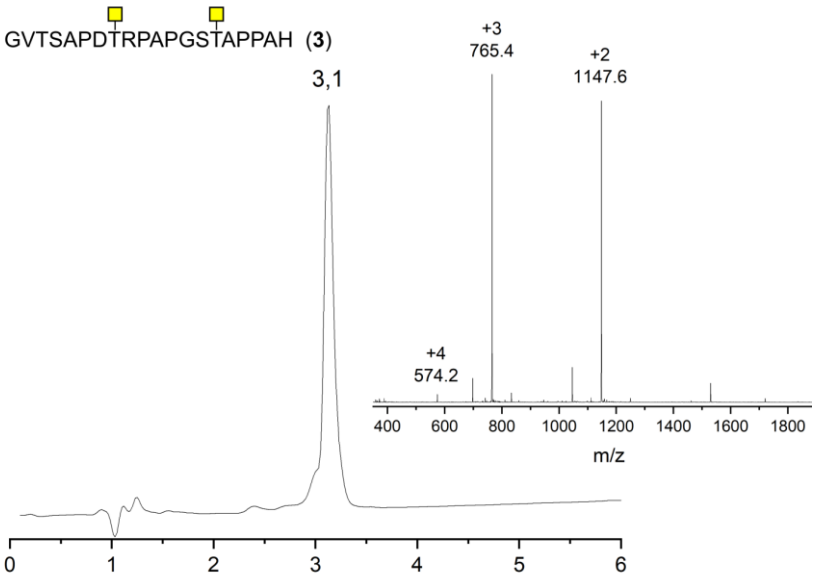

Figure S3. RP-HPLC-ESI-MS analysis of glycopeptide **3**. Eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B). Gradient 5-60% B in A, in 5 min $[M+2H]^{2+}$ 1147.6 (found), $[M+2H]^{2+}$ 1147.7 (calcd); $[M+3H]^{3+}$ 765.4 (found), 765.5 (calcd); $[M+4H]^{4+}$ 574.2 (found), $[M+4H]^{4+}$ 574.5 (calcd).

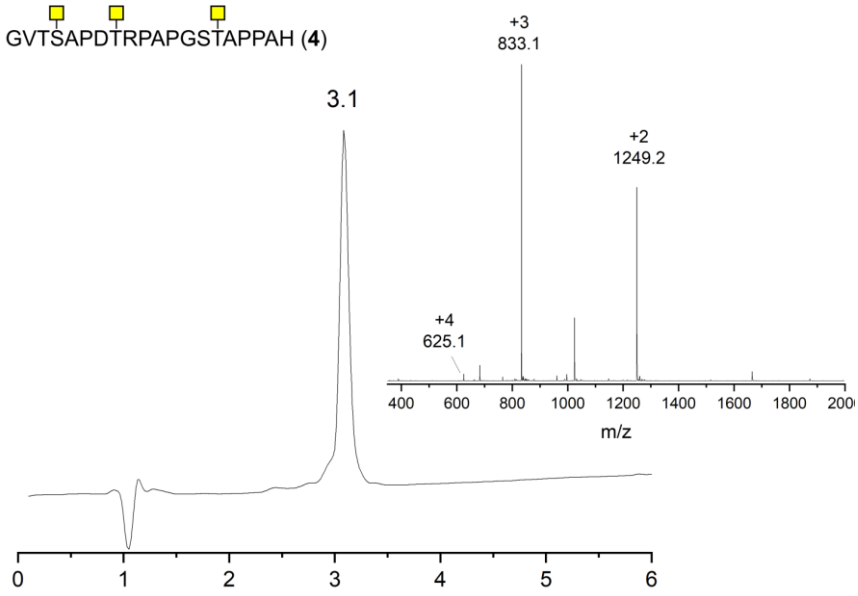

Figure S4. RP-HPLC-ESI-MS analysis of glycopeptide **4**. Eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B). Gradient 5-60% B in A, in 5 min. $[M+2H]^{2+}$ 1249.2 (found), $[M+2H]^{2+}$ 1249.3 (calcd); $[M+3H]^{3+}$ 833.1 (found), 833.2 (calcd); $[M+4H]^{4+}$ 625.1 (found), $[M+4H]^{4+}$ 625.2 (calcd).

Synthesis of glycopeptide conjugates 5 and 8

The peptidyl resins GVT(tBu)S(tBu)APD(OtBu)T(tBu)R(Pbf)PAPGS(tBu)T[GalNAc(OAc)$_3$]APPAH(Trt)-TentaGel® R PHB (214 mg) and GVT(tBu)S[GalNAc(OAc)$_3$]APD(OtBu)T[GalNAc(OAc)$_3$]R(Pbf)PAPGS(tBu)-T[GalNAc(OAc)$_3$]APPAH(Trt)-TentaGel® R PHB (210 mg) were swelled in DMF. After deprotection of the N-terminus by treatment with a 20% v/v solution of piperidine in DMF (3 min + 7 min incubation, with short

wash with DMF in between), each peptidyl resin was conjugated to monodispersed Fmoc-NH-PEG_{27}-OH. Crude glycopeptide conjugates were cleaved from the resin and deprotected on the amino acid side chains and on the hydroxyl groups of the sugars, following the general procedures described for peptide conjugation with monodispersed PEG and for deprotection, cleavage, and purification of peptides reported in the main text. Each crude glycopeptide conjugate, before the deprotection of the acetate on the hydroxyl functions of the GalNAc, was freeze-dried and weighted (PEG_{27}-GVTSAPDTRPAPGST[GalNAc$(OAc)_3$]APPAH: 63 mg, 17.9 µmol; PEG_{27}-GVTS[GalNAc$(OAc)_3$]-APDT[GalNAc$(OAc)_3$]RPAPGST[GalNAc$(OAc)_3$]APPAH: 78 mg, 20.5 µmol). Each crude glycopeptide conjugate, after deprotection of the hydroxyl functions of GalNAc, was recovered as light yellow oil.

Crude glycopeptide conjugate PEG_{27}-GVTSAPDTRPAPGST(GalNAc)APPAH (**5**) was purified via flash chromatography (Biotage Isolera One, equipped with a Snap Cartridge, KP-Sil 10g) using a gradient 5-80% eluent B in A in 25 min (Eluents: A = 0.1% (v/v) TFA in H_2O, B = 84% (v/v) CH_3CN in H_2O + 0.1% (v/v) TFA). Homogenous fractions containing the desired conjugate, were collected and freeze-dried giving the glycopeptide conjugate **5** in 29% yield (17.6 mg, 5.2 µmol) (Figure S5).

RP-HPLC ESI-MS analysis of PEG_{27}-GVTSAPDTRPAPGST(GalNAc)APPAH (**5**): Rt: 3.6 min. Bioshell: A160 Peptide C18 column (10 cm x 3 mm x 2.7 µm), temperature 35°C; flow 0.5 mL/min; eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B); gradient 10-70% B in A, in 5 min. ESI-MS (m/z): $[M+2H]^{2+}$ 1698.4 (found), $[M+2H]^{2+}$ 1698.6 (calcd); $[M+3H]^{3+}$ 1132.7 (found), 1132.7 (calcd); $[M+4H]^{4+}$ 849.6 (found), $[M+4H]^{4+}$ 849.8 (calcd); $[M+5H]^{5+}$ 679.9 (found), $[M+5H]^{5+}$ 680.0 (calcd); $[M+6H]^{6+}$ 566.6 (found), $[M+6H]^{6+}$ 566.8 (calcd).

Crude glycopeptide conjugate PEG_{27}-GVTS(GalNAc)APDT(GalNAc)RPAPGST(GalNAc)APPAH (**8**) was purified via flash chromatography (Biotage Isolera One, equipped with a Snap Cartridge, KP-Sil 10g) using a gradient 3-30% eluent B in A in 10 min, then 30-60% B in A in 20 minutes (Eluents: A = 0.1% (v/v) TFA in H_2O, B = 84% (v/v) CH_3CN in H_2O + 0.1% (v/v) TFA). Homogenous fractions containing the desired conjugate, were collected and freeze-dried giving the glycopeptide conjugate **8** in 15% yield (11.6 mg, 3.1 µmol) (Figure S6).

RP-HPLC-ESI-MS analysis of PEG_{27}-GVTS(GalNAc)APDT(GalNAc)RPAPGST(GalNAc)APPAH (**8**): Rt: 4.4 min, Bioshell: A160 Peptide C18 column (10 cm x 3 mm x 2.7 µm), temperature 35°C; flow 0.5 mL/min; eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B); gradient 10-70% B in A, in 5 min. ESI-MS (m/z): $[M+2H]^{2+}$ 1901.0 (found), $[M+2H]^{2+}$ 1901.8 (calcd); $[M+3H]^{3+}$ 1267.8 (found), 1268.1 (calcd); $[M+4H]^{4+}$ 951.1 (found), $[M+4H]^{4+}$ 951.4 (calcd); $[M+5H]^{5+}$ 761.1 (found), $[M+5H]^{5+}$ 761.3 (calcd); $[M+6H]^{6+}$ 634.4 (found), $[M+6H]^{6+}$ 634.6 (calcd); $[M+7H]^{7+}$ 544.0 (found), $[M+7H]^{7+}$ 544.1 (calcd).

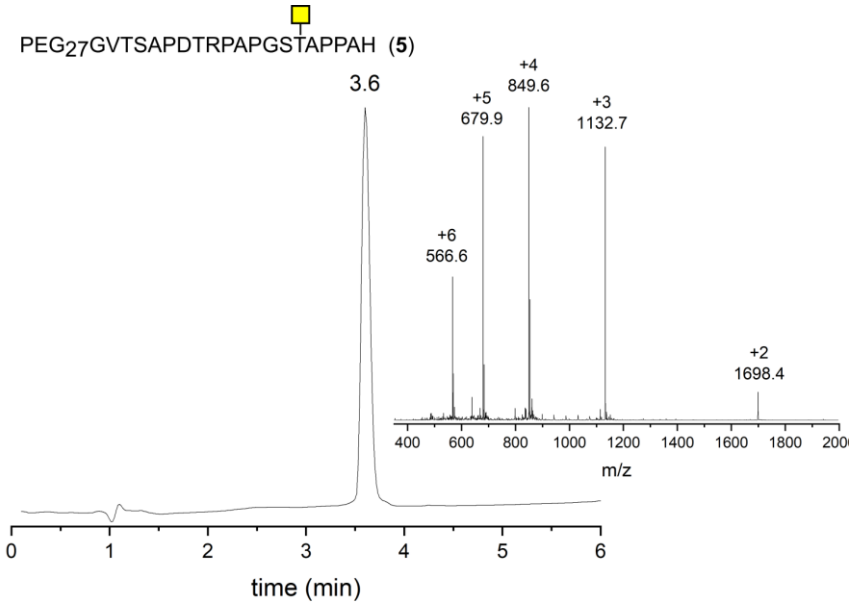

Figure S5. RP-HPLC-ESI-MS analysis of glycopeptide conjugate **5**. Eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN. Gradient 10-70% B in A, in 5 min. $[M+2H]^{2+}$ 1698.4 (found), $[M+2H]^{2+}$ 1698.6 (calcd); $[M+3H]^{3+}$ 1132.7 (found), 1132.7 (calcd); $[M+4H]^{4+}$ 849.6 (found), $[M+4H]^{4+}$ 849.8 (calcd); $[M+5H]^{5+}$ 679.9 (found), $[M+5H]^{5+}$ 680.0 (calcd); $[M+6H]^{6+}$ 566.6 (found), $[M+6H]^{6+}$ 566.8 (calcd).

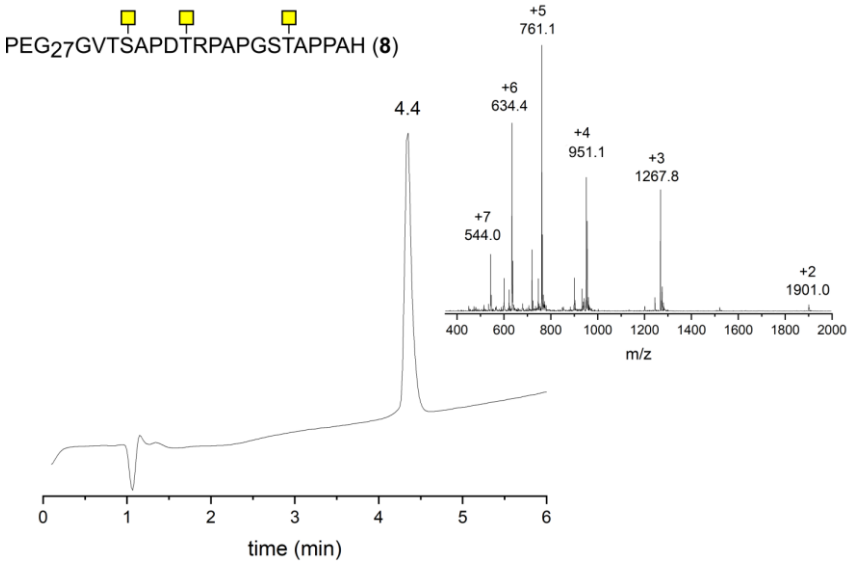

Figure S6. RP-HPLC-ESI-MS analysis of glycopeptide conjugate **8**. Eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B). Gradient 10-70% B in A, in 5 min. $[M+2H]^{2+}$ 1901.0 (found), $[M+2H]^{2+}$ 1901.8 (calcd); $[M+3H]^{3+}$ 1267.8 (found), 1268.1 (calcd); $[M+4H]^{4+}$ 951.1 (found), $[M+4H]^{4+}$ 951.4 (calcd); $[M+5H]^{5+}$ 761.1 (found), $[M+5H]^{5+}$ 761.3 (calcd); $[M+6H]^{6+}$ 634.4 (found), $[M+6H]^{6+}$ 634.6 (calcd); $[M+7H]^{7+}$ 544.0 (found), $[M+7H]^{7+}$ 544.1 (calcd).

Synthesis of MUC1 glycopeptides **6** and **7**

Glycopeptide conjugate PEG_{27}-GVTSAPDTRPAPGST(GalNAc)APPAH (**5**) (6.3 mg, 1.9 µmol) was dissolved in Milli-Q water (560 µL). The solution was diluted in a galactosylation buffer (30 mM TRIS (pH = 7.5), 0.06% Triton X-100, 10 mM $MnCl_2$, 2 mM UDP-Gal, 0.01 mg/mL C1GalT1, final volume: 5.16 mL) and shaken at 37°C. The reaction was monitored by RP-HPLC and finished after 3 h. The reaction mixture was added with ethanol (48 mL) and then with diethyl ether (108 mL) and stored at -80°C overnight. The precipitate was recovered after centrifugation (20 min, 5000 rpm). The quantitative recovery was verified by HPLC (comparison of the area under the curve of the peaks at t = 0 and after recovery). Half of the recovered pellet was desalted via RP-HPLC (Eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B). Gradient: 15-60% B in A in 8 min), obtaining 2.1 mg (0.6 µmol) of glycopeptide conjugate PEG_{27}-GVTSAPDTRPAPGST(Galβ1,3GalNAc)APPAH (**6**) carrying the Galβ1,3GalNAcα disaccharide. The second half of the recovered pellet was dissolved in a sialylation buffer (5 mM $MgCl_2$, 1 mM CMP-Neu5Ac, 0.01% ST3Gal1 transferase, final volume: 3.02 ml) and shaken at 37°C. The reaction was monitored by RP-HPLC and finished after 4 hours. The reaction mixture was poured in ethanol (24 ml). Then diethyl ether was added (54 ml) and the suspension was stored at -80°C overnight. The precipitate was recovered after centrifugation (20 min, 5000 rpm). The quantitative recovery was verified by HPLC (comparison of the area under the curve of the peaks at t = 0 and after recovery). The pellet was redissolved in water, freeze-dried and then desalted via RP-HPLC (Eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B). Gradient: 15-60% B in A in 8 min), obtaining 2.6 mg (0.7 µmol) of glycopeptide conjugate **7** carrying the Neu5Acα2-3Galβ1-3GalNAcα-trisaccharide.

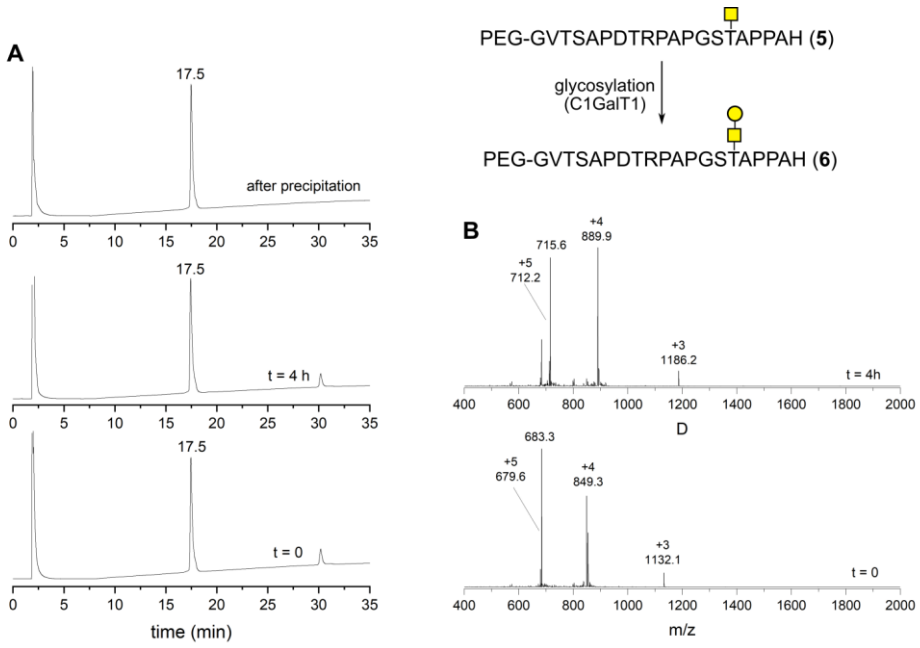

Figure S7. Galactosylation reaction of glycopeptide conjugate **5**. Panel A: RP-HPLC analysis (Eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN. Gradient: 5-65% B in A in 30 min) of the reaction mixture at t = 0, t = 4 h, and of the pellet recovered after precipitation. Panel B: Mass spectra corresponding to the peak Rt = 17.5 min in the chromatogram (t = 0) ($[M+3H]^{3+}$ 1132.1 (found), $[M+3H]^{3+}$ 1132.7 (calcd); $[M+4H]^{4+}$ 849.3 (found), $[M+4H]^{4+}$ 849.8 (calcd); $[M+5H]^{5+}$ 679.6 (found), $[M+5H]^{5+}$ 680.0 (calcd)) and to the peak Rt = 17.5 min (t = 4h) ($[M+3H]^{3+}$ 1186.2 (found), $[M+3H]^{3+}$ 1186.7 (calcd); $[M+4H]^{4+}$ 889.9 (found), $[M+4H]^{4+}$ 890.3 (calcd); $[M+5H]^{5+}$ 712.2 (found), $[M+5H]^{5+}$ 712.4 (calcd)).

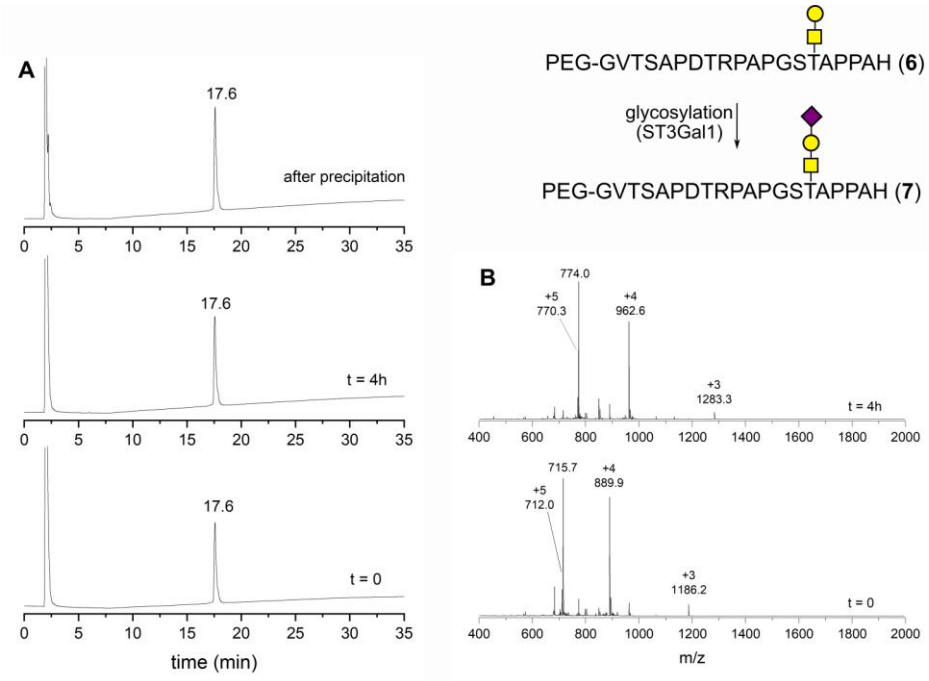

Figure S8. Sialylation reaction of glycopeptide conjugate **6**. Panel A: RP-HPLC analysis (Eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B). Gradient: 5-65% B in A in 30 min) of the reaction mixture at t = 0, t = 4 h, and of the pellet recovered after precipitation. Panel B: Mass spectra corresponding to the peak at Rt = 17.6 min in the chromatogram (t = 0) ($[M+3H]^{3+}$ 1186.2 (found), $[M+3H]^{3+}$ 1186.7 (calcd); $[M+4H]^{4+}$ 889.9 (found), $[M+4H]^{4+}$ 890.3 (calcd); $[M+5H]^{5+}$ 712.0 (found), $[M+5H]^{5+}$ 712.4 (calcd)) and Rt = 17.6 min (t = 4h) ($[M+3H]^{3+}$ 1283.3 (found), $[M+3H]^{3+}$ 1283.3 (calcd); $[M+4H]^{4+}$ 962.6 (found), $[M+4H]^{4+}$ 963.1 (calcd); $[M+5H]^{5+}$ 770.3 (found), $[M+5H]^{5+}$ 770.7 (calcd)).

Synthesis of glycopeptide conjugate 9

Peptidyl resin GVT(tBu)S(tBu)APD(OtBu)T(tBu)R(Pbf)PAPGS(tBu)T[GalNAc$(OAc)_3$]APPAH(Trt)-TentaGel® R PHB (170 mg) was elongated with the TEV protease recognition sequence (GDENLYFQ), following the general procedure for solid-phase peptide synthesis (see main text). After deprotection of the N-terminus, Fmoc-$(PEG)_{27}$-OH (71 mg, 46 µmol) was coupled (see general procedure for peptide conjugation with monodispersed PEG in the main text), followed by final Fmoc deprotection. The glycopeptide conjugate was cleaved from the resin and deprotected (see general procedure for deprotection and cleavage) and the crude glycopeptide conjugate, still protected on the hydroxyl functions of the sugar, was precipitated with diethyl ether, recovered after centrifugation and freeze-dried, giving crude **9** still protected as acetate on the hydroxyl functions of GalNAc (60.5 mg, 13.5 µmol). After deprotection of the hydroxyl functions with 5% v/v hydrazine monohydrate in methanol in 1.5 h, the solvent was evaporated and the crude directly dissolved in water and purified via flash chromatography (Biotage Isolera One, equipped with a Snap Cartridge, KP-Sil 10g, eluents A: 0.1% (v/v) TFA in H_2O and B: 0.1% (v/v) TFA in CH_3CN, gradient: 15-60% B in A in 8 min). Homogenous fractions containing the desired product were collected and freeze-dried giving the glycopeptide conjugate PEG_{27}-TEV-GVTSAPDTRPAPGST(GalNAc)APPAH (**9**) in 27% yield (16.2 mg, 3.7 µmol) (Figure S9).

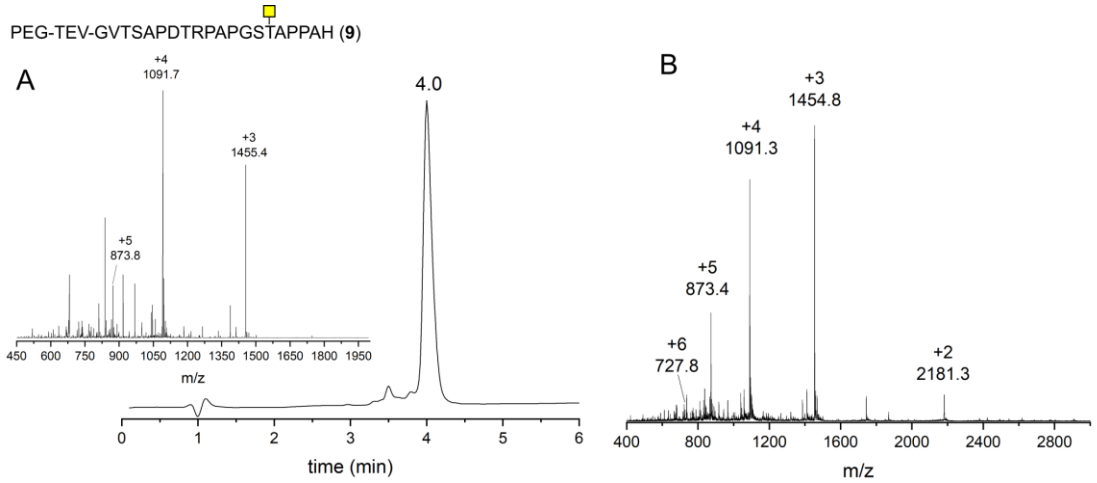

Figure S9. Panel **A**: RP-HPLC-ESI-MS analysis of glycopeptide conjugate **9**. Eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B). Gradient 10-70% B in A, in 5 min. $[M+3H]^{3+}$ 1455.4 (found), $[M+3H]^{3+}$ 1454.8 (calcd); $[M+4H]^{4+}$ 1091.7 (found), $[M+4H]^{4+}$ 1091.4 (calcd); $[M+5H]^{5+}$ 873.8 (found), $[M+5H]^{5+}$ 873.3 (calcd). The other peaks visible in the mass spectrum are due to a column contamination with detergents. Panel **B**: Mass spectrum of peptide conjugate **9** measured with an Arc HPLC system, column Waters XBridge C4 column, 3.5µm, 300A, 2.1x100mm, coupled to a Waters SQD2 mass spectrometer detector (Eluents: 0.1% (v/v) TFA in H_2O (A) and 0.1% (v/v) TFA in CH_3CN (B). Gradient 5-65% B in A in 7 min, flow: 0.5 mL/min). $[M+2H]^{2+}$ 2181.3 (found), $[M+2H]^{2+}$ 2181.7 (calcd); $[M+3H]^{3+}$ 1454.8 (found), $[M+3H]^{3+}$ 1454.8 (calcd); $[M+4H]^{4+}$ 1091.3 (found), $[M+4H]^{4+}$ 1091.4 (calcd); $[M+5H]^{5+}$ 873.4 (found), $[M+5H]^{5+}$ 873.3 (calcd); $[M+6H]^{6+}$ 727.8 (found), $[M+6H]^{6+}$ 727.9 (calcd)

Synthesis of glycopeptides **11** and **12**

Glycopeptide conjugate PEG_{27}-TEV-GVTSAPDTRPAPGST(GalNAc)APPAH (**9**) (9.6 mg, 2.2 µmol) was dissolved in Milli-Q water (10 mg/mL). The solution was diluted in a glycosylation buffer (30 mM TRIS (pH = 7.5), 0.06% Triton X-100, 10 mM $MnCl_2$, 2 mM UDP-Gal, 15 µg/mL C1GalT1, final volume: 8.75 mL) and shaken at 37°C. The reaction was monitored by RP-HPLC and finished after 3 h, giving PEG_{27}-TEV-GVTSAPDTRPAPGST(Galβ1,3GalNAc)APPAH (**10**) (Figure S10A and B). The reaction mixture was split in two. One part was added with DTT (1mM) followed by TEV protease (1:16 w/w peptide/enzyme), to enzymatically remove the PEG_{27}-TEV moiety, and the reaction was shaken at 4°C overnight. Glycopeptide GVTSAPDTRPAPGST(Galβ1,3GalNAc)APPAH (**11**) (1.9 mg, 0.9 µmol, 64% yield over two steps) was obtained (Scheme 2, main text) after purification via RP-HPLC (Eluents:0.1% (v/v) TFA in H_2O (A) and 0.08% (v/v) TFA in CH_3CN (B), gradient 5-45% B in A, in 30 min). The remaining reaction solution was diluted with ethanol (18 mL) and diethyl ether (80 mL) and stored at -80°C overnight. PEG_{27}-TEV-GVTSAPDTRPAPGST(Galβ1,3GalNAc)APPAH (**10**) precipitate was recovered after centrifugation (20 min, 5000 rpm). The pellet dissolved in Milli-Q water and $MgCl_2$ (final concentration 5 mM) was added with CMP-Neu5Ac (2 mM) and 25 µg/mL ST3Gal1 transferase (final volume: 4.3 mL) and shaken at 37°C to obtain glycopeptide conjugate PEG_{27}-TEV-GVTSAPDTRPAPGST(Neu5Acα2-3Galβ1,3GalNAc)APPAH (**10'**) (Figure S10C and D). The reaction was monitored by RP-HPLC. After 3.5 h the reagents for TEV cleavage were added (DTT, final concentration 1mM, TEV protease 1:16 w/w peptide/enzyme) and the reaction was shaken at 4°C overnight. Glycopeptide GVTSAPDTRPAPGST(Neu5Acα2-3Galβ1,3GalNAc)APPAH (**12**) (1.8 mg, 0.7 µmol, 77% yield over two steps) was obtained (Scheme 2, main text) after purification via RP-HPLC (Eluents:0.1% (v/v) TFA in H_2O (A) and 0.08% (v/v) TFA in CH_3CN (B), gradient 5-45% B in A, in 30 min).

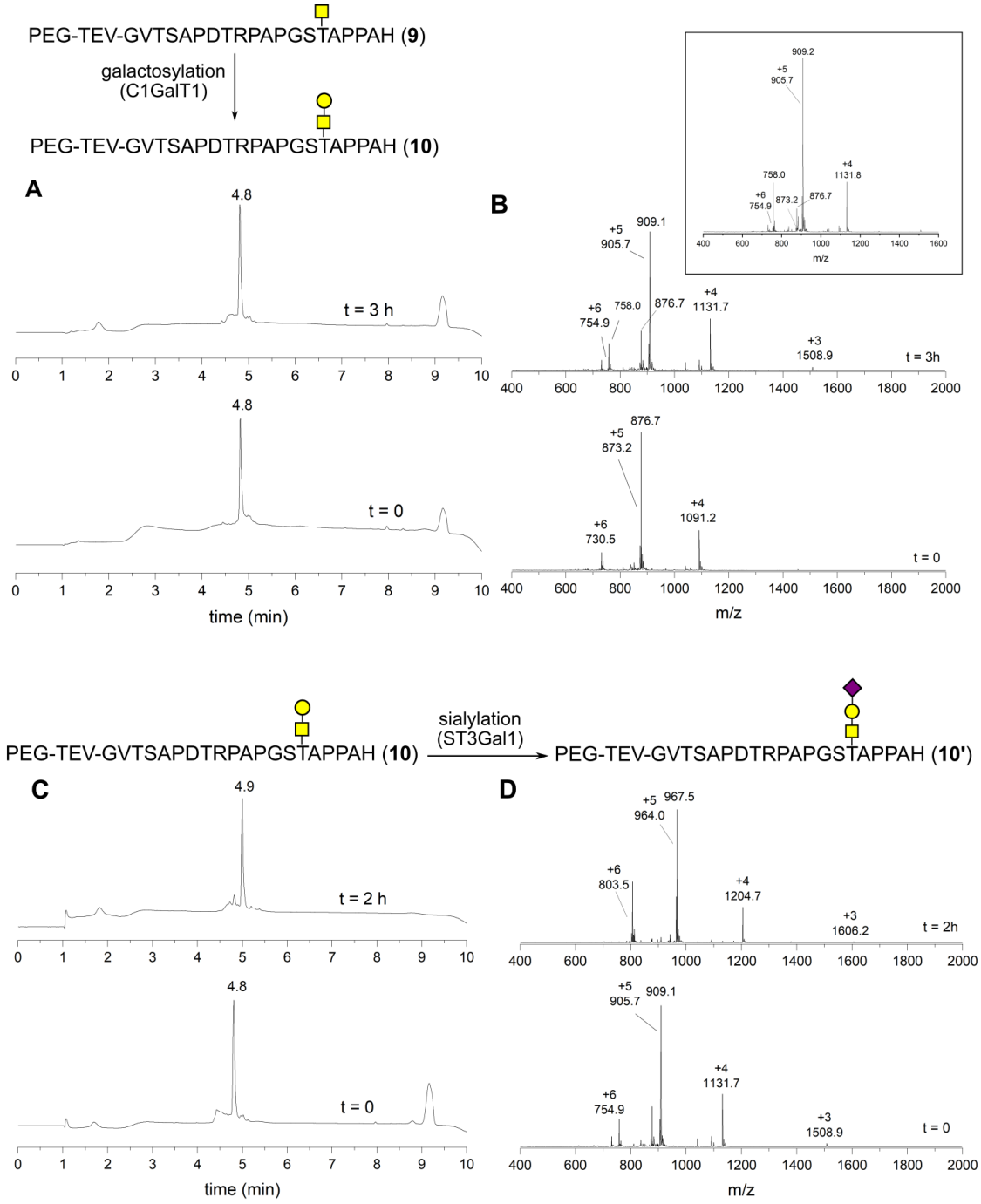

Figure S10. RP-HPLC and Mass analyses monitoring of further glycosylation reactions on glycopeptide conjugate **9** to obtain glycopeptide conjugates **10** and **10'** precursors of glycopeptides **11** and **12**, respectively. A: HPLC analysis of the galactosylation reaction mixture at t =0 and t = 2h. Eluents: 0.1% (v/v) HCOOH in H_2O (A) and 0.08% (v/v) HCOOH in CH_3CN. Gradient: 1-61% B in A in 30 min in 7.5 min. Panel B): mass spectra corresponding to the peaks at 4.8 min in the two chromatograms in panel A: t = 0 ($[M+4H]^{4+}$ 1091.2 (found), $[M+4H]^{4+}$ 1091.4 (calcd); $[M+5H]^{5+}$ 873.2 (found), $[M+5H]^{5+}$ 873.3 (calcd); $[M+6H]^{6+}$ 730.5 (found), $[M+6H]^{6+}$ 727.9 (calcd). The peak at m/z = 876.7 corresponds to the $[M+3H]^{3+}$ ion of a species with Δm = 18 compared to the mass of the desired product. We often observe the presence of such species in the spectra of PEGylated peptides, which are not detected after PEG removal, t = 3 h ($[M+3H]^{3+}$ 1508.9 (found), $[M+3H]^{3+}$ 1509.2 (calcd); $[M+4H]^{4+}$ 1131.7 (found), $[M+4H]^{4+}$ 1132.2 (calcd), $[M+5H]^{5+}$ 905.7 (found), $[M+5H]^{5+}$ 905.9; $[M+6H]^{6+}$ 754.9 (found), $[M+6H]^{6+}$ 755.1. The peaks at m/z = 876.7 corresponds to the $[M+3H]^{3+}$ ion of a species with Δm = 18 compared to the mass of the glycopeptide without the galactose (the correct mass of the galactosylated glycopeptide is visible as a small peak next to it), the presence of this mass can be explained by fragmentation of the sugar during ionization. Change of the ionization settings (lower cone voltage) let the intensity of this peak clearly

decreasing (see insert in panel B). Panel C: HPLC analysis of the sialylation reaction mixture at t =0 and t = 2h. Eluents: 0.1% (v/v) HCOOH in H_2O (A) and 0.08% (v/v) HCOOH in CH_3CN. Gradient: 1-61% B in A in 30 min in 7.5 min. Panel D: mass spectra corresponding to the most intense peak in each of the chromatograms in panel C (Rt: 4.8 min for t = 0, $[M+3H]^{3+}$ 1508.9 (found), $[M+3H]^{3+}$ 1509.2 (calcd); $[M+4H]^{4+}$ 1131.7 (found), $[M+4H]^{4+}$ 1132.2 (calcd); $[M+5H]^{5+}$ 905.7 (found), $[M+5H]^{5+}$ 905.9 (calcd); $[M+6H]^{6+}$ 754.9 (found), $[M+6H]^{6+}$ 755.1 (calcd), and Rt: 4.9 min for t = 2h, $[M+3H]^{3+}$ 1606.2 (found), $[M+3H]^{3+}$ 1606.6 (calcd), $[M+4H]^{4+}$ 1204.7 (found), $[M+4H]^{4+}$ 1205.2; $[M+5H]^{5+}$ 964.0 (found), $[M+5H]^{5+}$ 964.4 (calcd); $[M+6H]^{6+}$ 803.5 (found), $[M+6H]^{6+}$ 803.8 (calcd).

NMR spectroscopy

(Glyco)peptide samples were measured on a 600 MHz Bruker Avance III HD spectrometer equipped with a QXI ($^{1}H/^{13}C/^{15}N/^{31}P$)-quadruple resonance probe at 298 K. Samples were either dissolved in a buffer of 20 mM KH_2PO_4/K_2HPO4 pH 7.0 containing 10% D_2O resulting in a (glyco)peptide concentration of approx. 1 mM. For measurements in D_2O, dissolved samples were lyophilized and dissolved in the same volume D_2O. A sample volume 275 µL was used with 5 mm Shigemi-tubes. Standard 2D ^{1}H-^{1}H TOCSY spectra were measured using mixing times of either 120 ms or 12 ms, 4 scans, a recycle delay of 1 sec, 2048×512 points and spectral widths of 12.9 ppm × 8.3 ppm with a duration of 50 min. 2D ^{1}H-^{1}H ROESY spectra were acquired with the pulse sequence roesygpph19 (Bruker library) using 96 scans, a recycle delay of 1 sec, 2048×600 points and spectral widths of 12.9 ppm × 8.3 ppm with a duration of 22 hours. Standard 2D ^{1}H-^{15}N HSQC spectra were recorded with 384 scans, a recycle delay of 1 sec, 1024×128 points and spectral widths of 13.9 ppm × 36.5 ppm with a duration of 15 h 30 min. Standard 2D ^{1}H-^{13}C HSQC spectra were recorded with 192 scans, a recycle delay of 1.2 sec, 1024×270 points and spectral widths of 13.9 ppm × 77.1 ppm with a duration of 18 hours 30 min. 2D ^{1}H-^{13}C HMBC spectra were recorded using the pulse sequence hmbcgpndqf (Bruker library) with 128 scans, a recycle delay of 1.2 sec, 4096×512 points and spectral widths of 20.02 ppm × 77.1 ppm with a duration of 26 hours 42 min. 2D ^{1}H-^{13}C HMBC spectra optimized for correlations with carbonyl resonances using shaped pulses[1] were recorded with 256 scans, a recycle delay of 1.2 sec, 2048×136 points and spectral widths of 13.9 ppm × 16.0 ppm with a duration of 15 hours 30 min. 2D ^{1}H-^{13}C HSQC-TOCSY spectra were acquired using the pulse sequence hsqcdietgpsisp.2 (Bruker library) with 192 scans, a recycle delay of 1.2 sec, 1024×270 points, a mixing times of either 80 or 12 ms and spectral widths of 16.0 ppm × 77.1 ppm with a duration of 18 hours 42 min. The combined chemical shift deviation Δδ was calculated as a combination of ^{1}H and ^{15}N chemical shift changes according to:

$$\Delta\delta = \sqrt{(\Delta\delta^{1}\mathrm{H})^2 + \left(\frac{\Delta\delta^{15}\mathrm{N}}{4}\right)^2}$$

where $\Delta\delta^{1}H$ and $\Delta\delta^{15}N$ denote the chemical shift differences in parts per million (p.p.m.) of amide hydrogen and nitrogen atoms, respectively.

Spectra were processed with Topspin 3.2 (Bruker Biospin), referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) with an external sample of 2 mM sucrose and 0.5 mM DSS (Bruker Biospin) and analyzed using the software Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco, USA). ^{13}C and ^{15}N axis were referenced indirectly according to IUPAC-IUBMB-IUPAB[2] using the chemical shift referencing ratios of 0.251449530 and 0.101329118.

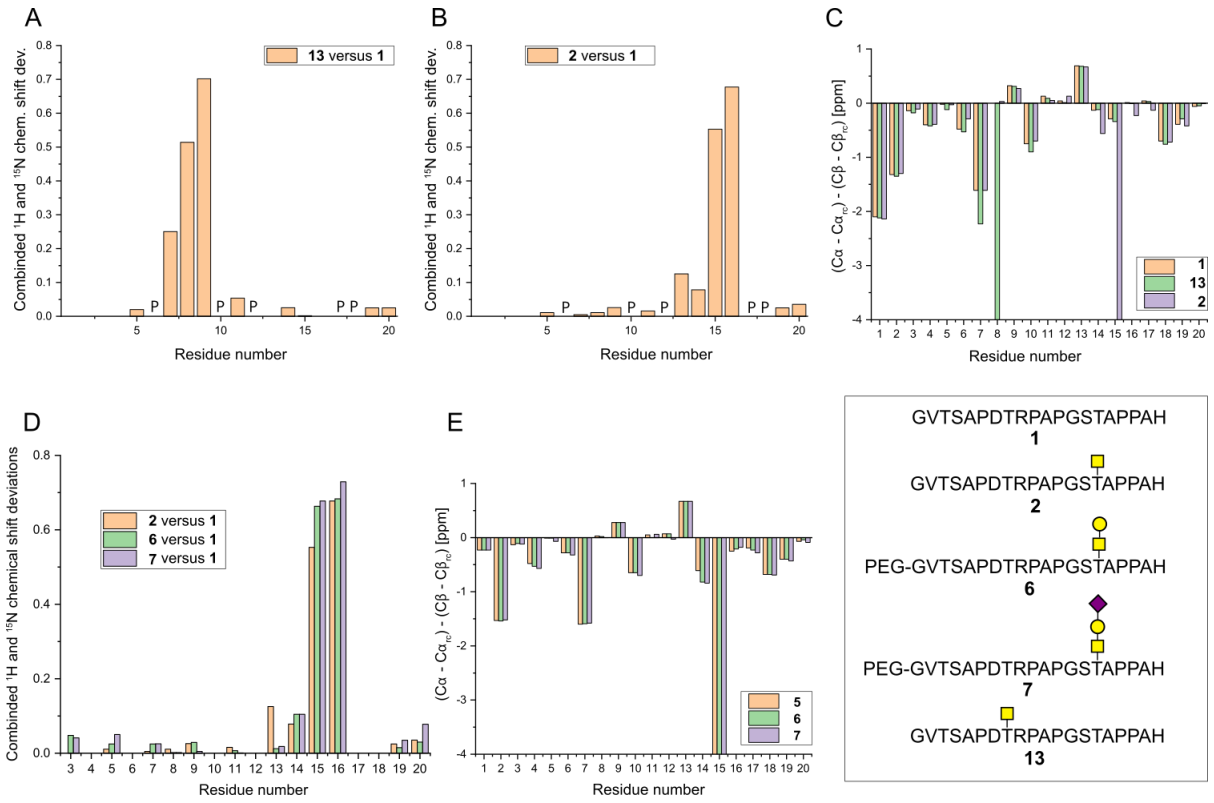

Figure S11. Effect of the position and of the length of the *O*-glycans on the peptide backbone sensed by NMR spectroscopy. Panel A) Combined ^{1}H and ^{15}N chemical shift deviations from the comparison between glycopeptides **13** and **1**. B) Combined ^{1}H and ^{15}N chemical shift deviations from the comparison between (glycol)peptides **2** and **1**. Panel C) ^{13}C chemical shift deviations from random coil values for glycopeptides **5**, **6** and **18**. Panel D) Combined chemical shift deviations of glycopeptide **2**, **6** and **7** compared with the non-glycosylated peptide **1**. Panel E) Effect of the glycan length on the backbone conformation. Shown are ^{13}C chemical shift deviations from random coil values for glycopeptides **5**, **6** and **7**.

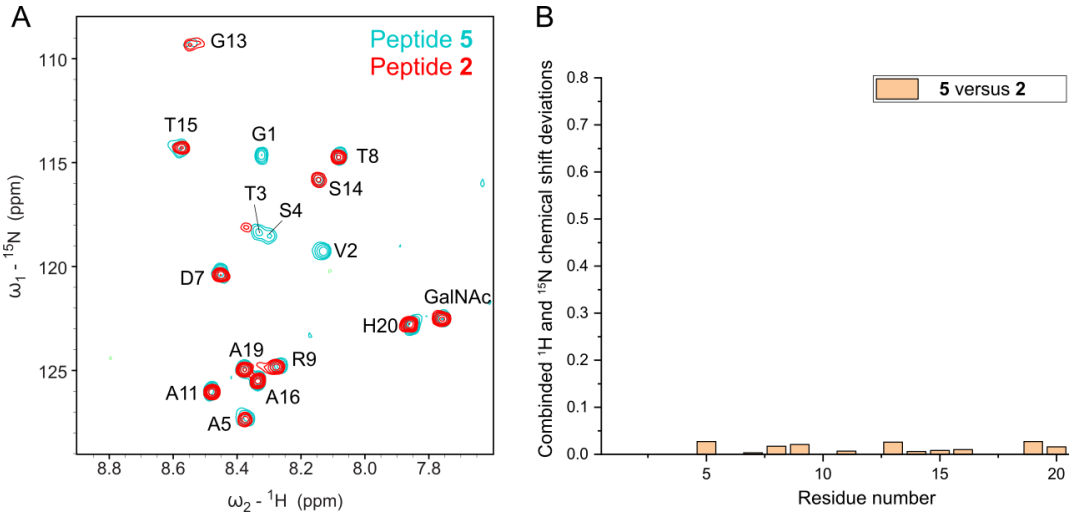

Figure S12: Effect of PEGylation on the ^{1}H-^{15}N HSQC spectra**.** Panel A) Comparison between the PEGylated glycopeptide **5** with the non-modified glycopeptide **2**. Panel B) Combined ^{1}H and ^{15}N chemical shift deviations between both spectra with the same scale as in Fig. S11 A, B and D.

	Peptide **5**		Peptide **11**		Peptide **12**[a]	
NOE observed	category	distance	category	distance	category	distance
Gal H1- GalNAc H3	—	—	s	2.28	s	2.43
Gal H1- GalNAc H^N	—	—	m	3.08	w	3.55
GalNAc H1 - Thr15 HB	s	2.57	s	2.47	n.d.	
GalNAc H6 - Thr15 HG2	m	4.61	m	4.6	w	5.12
GalNAc H5 - Thr15 HG2	s	2.99	s	2.84	s	3.15
GalNAc H5 - Thr15 HB	n.d. / ov.		w	3.56	n.d.	
GalNAc H8 - Ser14 HA	n.d. / ov.		w	4.32	n.d.	
GalNAc H8 - Ser14 HB	w	4.79	w	4.73	n.d.	
GalNAc H8 - Ala16 HA	m	3.61	m	3.56	m	3.76
GalNAc H8 - Ala16 HB	m	4.93	m	4.66	w	5.12
GalNAc H8 - Pro17 HA	m	3.35	m	3.35	m	3.76
GalNAc H^N - Thr15 H^N	w	4.61	w	4.15	n.d.	
GalNAc H^N - Pro17 HA	w	4.20	w	4.36	n.d.	
Thr15 HG2 - Thr15 H^N	w	4.93	w	4.73	w	5.12

[a] the ROESY spectrum of 12 had lower signal to noise ratios compared to the spectra of the other two peptides, therefore the distances are less reliable

Table 1 Observed NOE cross-peaks between the O-glycan and the peptide backbone. Peptide **5** = PEG_{27}-GVTSAPDTRPAPGST(GalNAc)APPAH; peptide **11** = GVTSAPDTRPAPGST(Galβ1,3GalNAc)APPAH; peptide **12** = GVTSAPDTRPAPGST(Neu5Acα23Galβ1,3GalNAc)APPAH. The signal intensities were categorized into weak (w), medium (m) and strong (s). In addition distances were estimated based on signal-to-noise ratios of the NOEs and the known distance between GalNAc H1-H5 of 3.7 Å considering also the number of protons involved as described in Zierke et al.[3] n.d. stands for not detected and ov. for overlap.

Cell Viability Assay

20.000 MCF-7 cells were seeded in 24-multiwell plates and treated with or without increasing MUC1 (glyco)peptide concentrations for 48h. Cell viability was evaluated by the addition of crystal violet solution (0.5% in 20% methanol). After 5 min of staining, the fixed cells were washed with phosphate-buffered saline (PBS) and solubilized with 200 µL/well of SDS 2%. The absorbance at 595 nm was evaluated using a microplate reader (BioRad, Hercules, CA, USA).

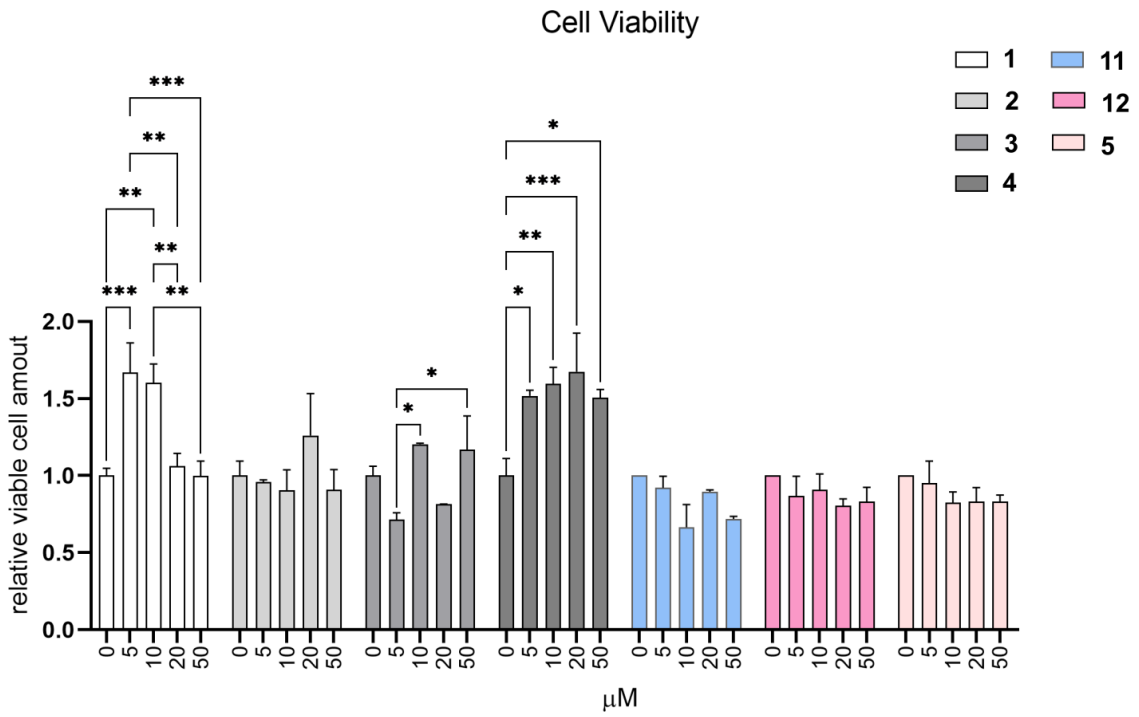

Figure S13. Viability assays of peptides on MCF-7 cells. None of the synthetic (glyco)peptides is toxic for MCF-7 cells at the tested concentrations. Two-way Anova with Tukey's multiple comparisons test. *p<0.05, ** p<0.01, ***p<0.001

Adhesion assays

Plates were coated by adding 100 µL of a solution containing 10 µg/mL MUC1 peptides in PBS and let to adhere for 16h at RT. To estimate the quantity of the peptide adhered on the surface, we performed an HPLC analysis, comparing the area under the peak of the solution before coating with the area under the peak of the solution after overnight incubation in the well. The area under the peak decreased of 4% after overnight incubation, in a reproducible manner. From these data we calculated that ca. 40 ng peptide (0.02 nmol) adhered on the well. The analysis was performed for peptides **3** and **11** with very similar outcome. Our results were in agreement to a previous study demonstrating adhesion of fibronectin-derived peptides by radiolabelling.[4] The coating of MUC1 (glyco)peptides on wells has not been performed via oriented immobilization, thus all possible conformations/orientations on the surface are represented. Therefore hydrophilic glycans relevant for biological activity are predominantly exposed. 30.000 MCF-7 cells were seeded onto 24-multiwell coated-plates for 30 min. Adherent cells were stained with crystal violet solution (0.5% in 20% methanol). After 5 min of staining, photos of the wells were taken. Cells were counted in randomly chosen fields using the ImageJ software (NIH, USA). The test was repeated several times, performing both technical and biological replicates of the experiment, to confirm that modification of cell adhesion ability depends on the nature and abundance of the glycans on the peptide.

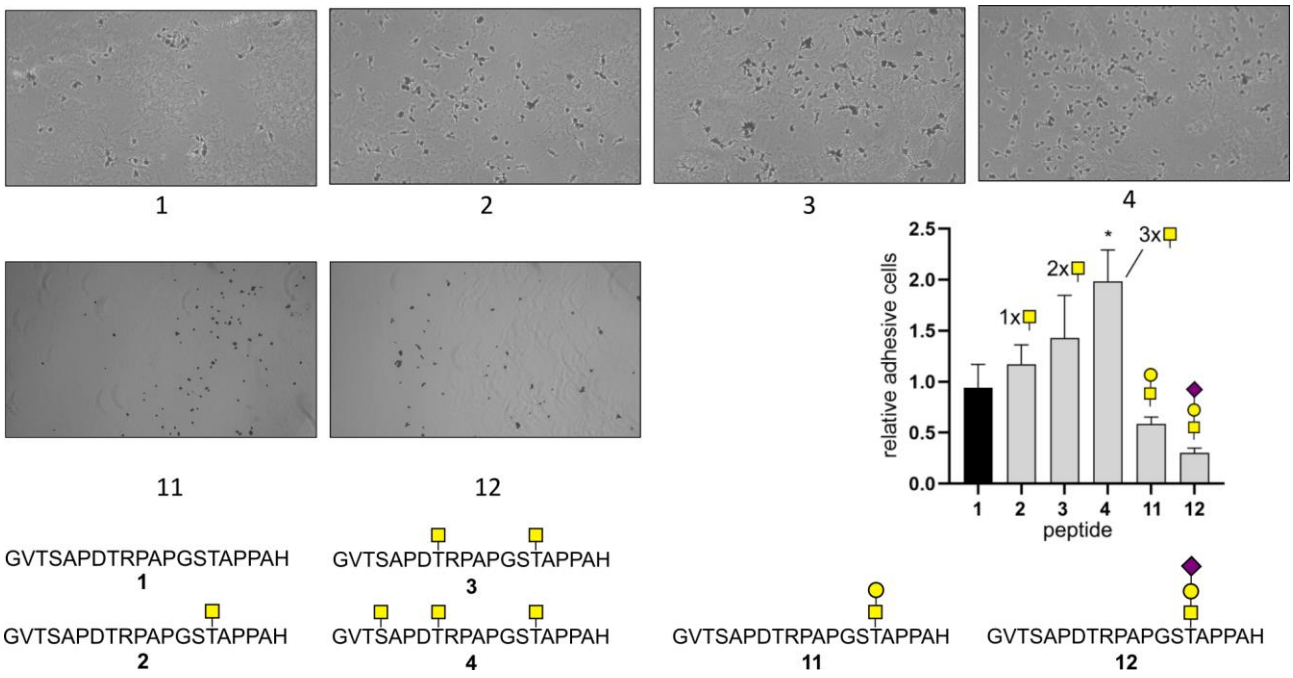

Figure S14. Adhesion experiments. MCF-7 cells were seeded for 30 min on cell culture dishes pre-coated with MUC1 glycopeptides with different glycosylation patterns (the number corresponding to each peptide is reported on the x axis). Images of the adherent cells are presented. Results in the bar graph are shown as relative to the number of adherent cells on uncoated plate. Data are represented as mean ± SEM of three replicates. *p < 0.05 for 4 vs 1; p < 0.5 for 12 vs 1 (One-way Anova with Dunnett's post hoc test).

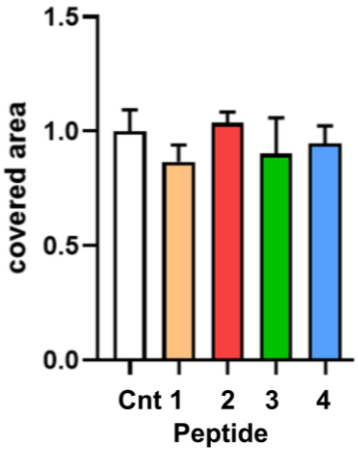

Figure S15. Adhesion assay performed co-incubating MUC1 (glyco)peptides with cells during adhesion assay on collagen-coated plates. Cell adhesion was not affected by the presence of the peptide in solution. Cnt = Control.

[1] L. Grassi, C. Regl, S. Wildner, G. Gadermaier, C. G. Huber, C. Cabrele, and M. Schubert *Analytical Chemistry* **2017** *89 (22)*, 11962-11970. DOI: 10.1021/acs.analchem.7b01645

[2] J.L. Markley, A. Bax, Y. Arata, C.W. Hilbers, R. Kaptein, B.D. Sykes, P.E. Wright, K. Wüthrich *Journal of Biomolecular NMR* **1998**, *12(1)*, 1-23. DOI: 10.1023/a:1008290618449

[3] Zierke, M. Smieško ,M. Rabbani, S. Aeschbacher, T. Cutting, B. Allain, F.H. Schubert, M. Ernst, B. *Journal of the American Chemical Society* **2013**, 135(36),13464-72. DOI: 10.1021/ja4054702

[4] Our results were in agreement to a previous study demonstrating adhesion of fibronectin-derived peptides by radiolabelling. In particular, the fibronectin peptide was radioiodinated and used for coating (overnight incubation in well), the amount of labelled peptide adhered to the well was calculated from the specific radioactivity (Pierschbacher M. et al PNAS, 1980, 80, 1224).