Electronic Supplementary Material (ESI) for Chemical Communications.


**Supporting Information**

# Structural Characterization of an Unprecedented Lectin-like Antitumoral anti-MUC1 Antibody

Javier Macías-León, Iris A. Bermejo, Alicia Asín, Ana García-García, Ismael Compañón, Ester Jiménez-Moreno, Helena Coelho, Vincenzo Mangini, Inês S. Albuquerque, Filipa Marcelo, Juan L. Asensio, Gonçalo J. L. Bernardes, Hiren J. Joshi, Roberto Fiammengo, Ola Blixt, Ramón Hurtado-Guerrero, and Francisco Corzana

**Reagents and general procedures.** Commercial reagents were used without further purification. Analytical thin layer chromatography (TLC) was performed on Macherey-Nagel precoated aluminium sheets with a 0.20 mm thickness of silica gel 60 with fluorescent indicator UV254. TLC plates were visualized with UV light and by staining with phosphomolybdic acid (PMA) solution (5 g of PMA in 100 mL of absolute ethanol) or sulfuric acid-ethanol solution (1:20). Column chromatography was performed on silica gel (230−400 mesh).

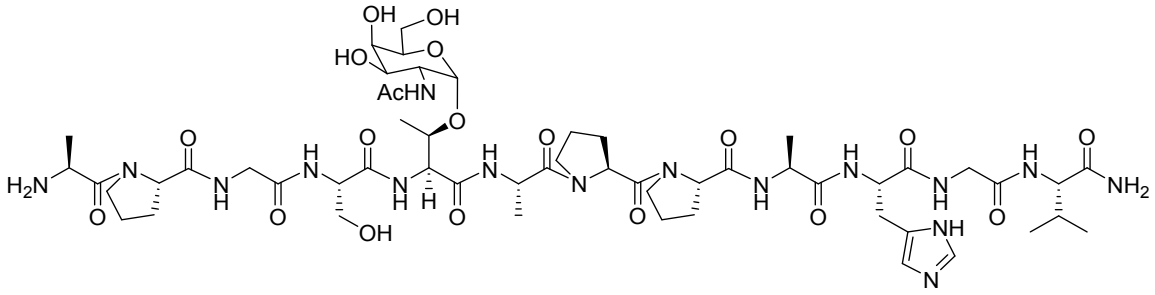
$^{1}H$ and $^{13}C$ NMR spectra were measured with a 400 MHz spectrometer with TMS as the internal standard. Multiplicities are quoted as singlet (s), broad singlet (br s), doublet (d), doublet of doublets (dd), triplet (t), or multiplet (m). Signals were assigned using COSY and HSQC experiments. NMR chemical shifts (δ) are reported in ppm and coupling constants (*J*) in Hz. High resolution electrospray mass (ESI) spectra were recorded on a microTOF spectrometer; accurate mass measurements were achieved by using sodium formate as an external reference.

**Solid-phase peptide synthesis (SPPS).** Glycopeptides were synthesized by stepwise microwave assisted solid-phase synthesis on a Liberty Blue synthesizer using the Fmoc strategy on Rink Amide MBHA resin (0.1 mmol). Fmoc-Thr[GalNAc(Ac)3-α-D]-OH, Fmoc-Ser[GalNAc(Ac)3-α-D]-OH and Neu5Ac(Ac)4(Bn)-α-(2-6)-GalNAc(Ac)3-α-1-*O*-Thr (2.0 equiv) were synthesized as described in the literature,[1,2] and manually coupled using HBTU [(2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate], 0.9 equiv., and 0.25 mL of DIPEA (2.0 M in NMP) dissolved in 1 mL of DMF, while all other Fmoc amino acids (5.0 equiv) were automatically coupled using oxyma pure/DIC (N,N'-diisopropylcarbodiimide). The O-acetyl groups of GalNAc moiety were removed in a mixture of $NH_2NH_2$/MeOH (7:3) for the Tn antigen. (Glyco)peptides were then released from the resin, and all acid sensitive protecting groups were simultaneously removed using TFA 95%, TIS (triisopropylsilane) 2.5% and $H_2O$ 2.5%, followed by precipitation with cold diethyl ether. In the case of the STn antigen derived glycopeptide, the derivative was first detached from the resin using the aforementioned cleavage cocktail, which was further evaporated, purified through cartridge and lyophilized. Afterwards, benzyl group of the sialic acid was removed through hydrogenolysis (Pd-C in MeOH) overnight. Once this reaction is completed, Pd-C is filtered and solvent evaporated. Finally, remaining AcO- groups are removed using a MeONa/MeOH solution until pH 9.5 is reached. After completion of the reaction, solution is neutralized and evaporated. The crude products were purified by HPLC on a Phenomenex Luna C18(2) column (10 μm, 250 mm × 21.2 mm) and a dual absorbance detector, with a flow rate of 10 or 20 mL/min.

**Glycopeptide 1**

Synthesis and characterization of glycopeptide **1'** has been previously described.[3]

**Glycopeptide 1'**



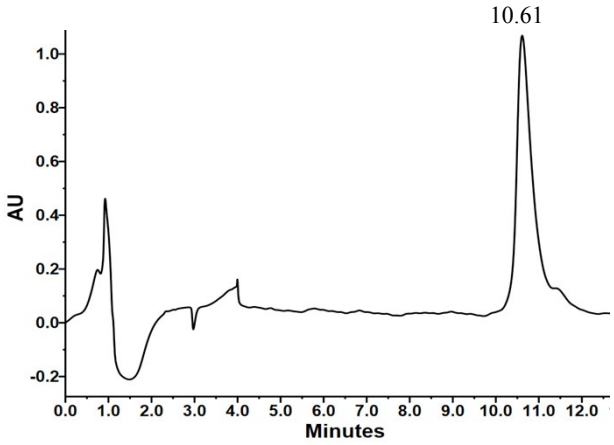
**$^{1}$H NMR (400 MHz, $D_2O$) δ (ppm):** 0.93 – 1.03 (m, 6H, $2CH_{3Val}$), 1.28 (d, 3H, *J* = 6.5 Hz, $CH_{3Thr}$), 1.32 – 1.40 (m, 6H, $CH_{3Ala6}$, $CH_{3Ala9}$), 1.55 (d, 3H, *J* = 7.0 Hz, $CH_{3Ala1}$), 1.85 – 2.16 (m, 13, $H\beta_{Pro2}$, $H\beta_{Pro7}$, $H\beta_{Pro8}$, $2H\gamma_{Pro2}$, $2H\gamma_{Pro7}$, $2H\gamma_{Pro8}$, $NHCH_3$, $H\beta_{Val}$), 2.25 – 2.42 (m, 3H, $H\beta_{Pro2}$, $H\beta_{Pro7}$, $H\beta_{Pro8}$), 3.19 – 3.34 (m, 2H, $2H\beta_{His}$), 3.63 – 4.05 (m, 17H, $2H\delta_{Pro7}$, $2H\delta_{Pro2}$, $2H\delta_{Pro9}$, $2H_{6S}$, $2H\beta_{Ser}$, $2H\alpha_{Gly3}$, $2H\alpha_{Gly11}$, $H_{3S}$, $H_{4S}$, $H_{5S}$), 4.11 (dd, 1H, *J* = 11.0, 3.8 Hz, $H_{2S}$), 4.15 (d, 1H, *J* = 6.8 Hz, $H\alpha_{Val}$), 4.25 (q, 1H, *J* = 7.2 Hz, $H\alpha_{Ala9}$), 4.33 – 4.43 (m, 3H, $H\alpha_{Pro7}$, $H\alpha_{Ala1}$, $H\beta_{Thr}$), 4.52 (t, 1H, $H\alpha_{Pro2}$), 4.54 – 4.63 (m, 2H, $H\alpha_{Ala6}$, $H\alpha_{Thr}$), 4.64 – 4.74 (m, 3H, $H\alpha_{Ser}$, $H\alpha_{Pro8}$, $H\alpha_{His}$), 4.95 (d, 1H, *J* = 3.9 Hz, $H_{1S}$), 7.34 (s, 1H, $H\delta2_{His}$), 8.63 (s, 1H, $H\varepsilon_{His}$).

**$^{13}$C NMR (100 MHz, $D_2O$) δ (ppm):** 15.0 ($CH_{3Ala1}$), 15.3, 16.4 ($CH_{3Ala6}$, $CH_{3Ala9}$), 17.3 ($CH_{3Val}$), 18.3 ($CH_{3Thr}$), 22.3 ($NHCOCH_3$), 24.7, 24.7, 24.8 ($C\gamma_{Pro2}$, $C\gamma_{Pro7}$, $C\gamma_{Pro8}$), 26.4 ($C\beta_{His}$), 28.0, 29.3, 29.3 ($C\beta_{Pro2}$, $C\beta_{Pro7}$, $C\beta_{Pro9}$), 29.8 ($C\beta_{Val}$), 42.3 ($C\alpha_{Gly}$), 47.5 ($C\alpha_{Ala6}$), 47.7, 47.8, 47.8 ($C\delta_{Pro7}$, $C\delta_{Pro2}$, $C\delta_{Pro8}$), 48.1 ($C\alpha_{Ala1}$), 49.6 ($C\alpha_{Ala9}$), 49.7 ($C_{2S}$), 52.4 ($C\alpha_{His}$), 55.2 ($C\alpha_{Ser}$), 57.1 ($C\alpha_{Thr}$), 58.4 ($C\alpha_{Pro8}$), 59.2 ($C\alpha_{Val}$), 60.1 ($C\alpha_{Pro7}$), 60.7 ($C\alpha_{Pro2}$), 61.3 ($C\beta_{Ser}$), 61.4 ($C_{6S}$), 68.2, 68.6, 71.3 ($C_{4S}$, $C_{5S}$, $C_{3S}$), 75.8 ($C\beta_{Thr}$), 98.7 ($C_{1S}$), 133.6, 162.8, 163.2 ($C\delta2_{His}$, $C\varepsilon1_{His}$, $C\gamma_{His}$), 169.3, 170.7, 171.1, 171.2, 172.0, 172.0, 172.0, 172.1, 173.6, 174.0, 174.4, 174.9, 176.1 (CON).

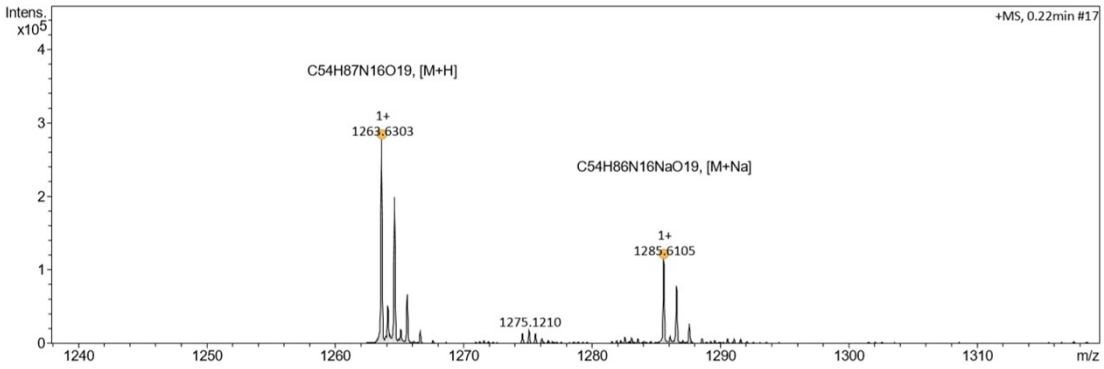
**Semi-preparative HPLC:** Rt = 13.0 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) →(12.5:87.5), 15 min, 20 mL/min, λ = 212 nm)

**Analytical HPLC:** Rt = 10.6 min (Phenomenex Luna 5 µm C18 (2), 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (7:93) →(13:87), 12 min, 1 mL/min, λ = 212 nm)

(glycopeptide **1'**)

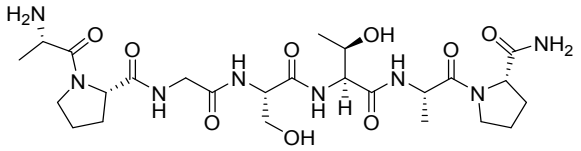



**HRMS ESI+ (m/z):** calcd for $C_{54}H_{87}N_{16}O_{19}[M+H]^+$: 1263.6328, found 1263.6303.
(glycopeptide **1'**)




| Ion Formula | m/z | z | \|err\| [ppm] | Mean err [ppm] | mSigma |
|---|---|---|---|---|---|
| C54H86N16NaO19 | 1285.614735 | 1+ | 3.3 | 397.4 | 14.5 |
| C54H87N16O19 | 1263.632791 | 1+ | 2.0 | 3.0 | 26.7 |

**Peptide 1-naked**

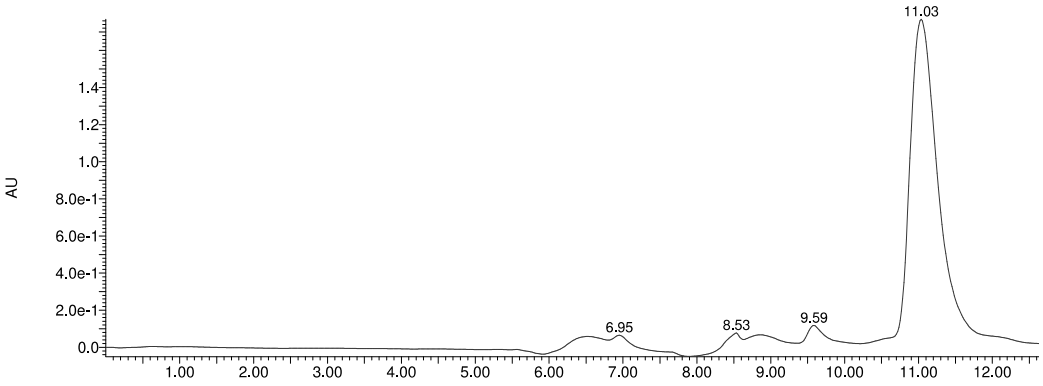


**$^1$H NMR (400 MHz, $D_2O$) δ (ppm):** 1.11 (d, 3H, $J$ = 6.4 Hz, $CH_{3Thr}$), 1.28 (d, 3H, $J$ = 7.1 Hz, $CH_{3Ala6}$), 1.43 (d, 3H, $J$ = 7.0 Hz, $CH_{3Ala1}$), 1.79 – 2.05 (m, 6H, $H\beta_{Pro2}$, $H\beta_{Pro7}$, $2H\gamma_{Pro2}$, $2H\gamma_{Pro7}$), 2.15 – 2.32 (m, 2H, $H\beta_{Pro2}$, $H\beta_{Pro7}$), 3.50 – 3.84 (m, 6H, $2H\beta_{Ser}$, $2H\delta_{Pro2}$, $2H\delta_{Pro7}$), 3.91 (s, 2H, $2H\alpha_{Gly}$), 4.09 – 4.17 (m, 1H, $H\beta_{Thr}$), 4.22 – 4.32 (m, 3H, $H\alpha_{Thr}$, $H\alpha_{Ala1}$, $H\alpha_{Pro7}$), 4.37 – 4.43 (m, 1H, $H\alpha_{Pro2}$), 4.45 (t, 1H, $J$ = 5.6 Hz, $H\alpha_{Ser}$), 4.52 (q, 1H, $J$ = 7.0 Hz, $H\alpha_{Ala6}$).

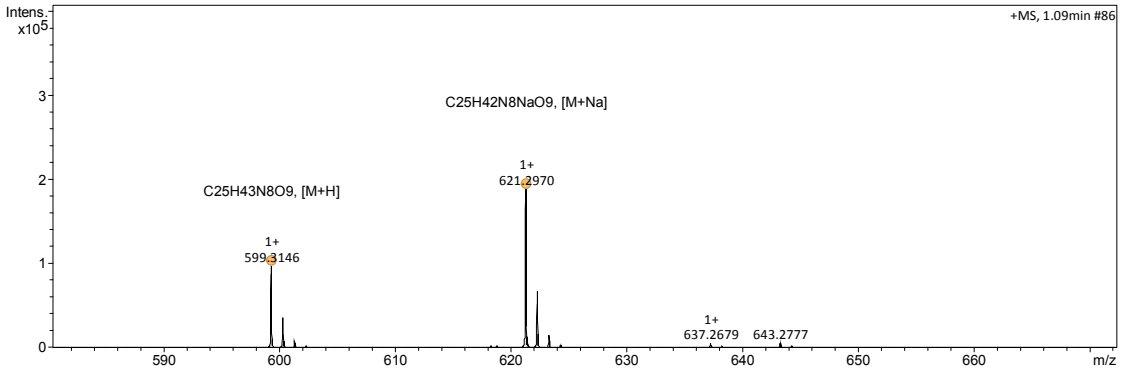
**$^1$H NMR (400 MHz, $D_2O/H_2O$, 1:9, amide region) δ (ppm):** 6.91 (s, $NH_{term}$), 7.53 (s, $NH_{term}$), 8.13 (d, $J$ = 6.8 Hz, $NH_{Ser}$), 8.19 (d, $J$ = 7.9 Hz, $NH_{Thr}$), 8.25 (d, $J$ = 5.8 Hz, $NH_{Ala6}$), 8.49 (t, $J$ = 5.8 Hz, $NH_{Gly}$).

**$^{13}$C NMR (100 MHz, $D_2O$) δ (ppm):** 14.9, 15.3 ($CH_{3Ala1}$, $CH_{3Ala6}$), 18.8 ($CH_{3Thr}$), 24.6, 24.7 ($C\gamma_{Pro2}$, $C\gamma_{Pro7}$), 29.3, 29.5 ($C\beta_{Pro2}$, $C\beta_{Pro7}$), 42.4 ($C\alpha_{Gly}$), 47.7 ($C\alpha_{Ala6}$), 47.8, 47.9 ($C\delta_{Pro7}$, $C\delta_{Pro2}$), 48.0 ($C\alpha_{Ala1}$), 55.4 ($C\alpha_{Ser}$), 58.8 ($C\alpha_{Thr}$), 60.3, 60.7 ($C\alpha_{Pro2}$, $C\alpha_{Pro7}$), 61.1 ($C\beta_{Ser}$), 67.0 ($C\beta_{Thr}$), 169.3, 171.2, 171.4, 172.0, 172.8, 174.5, 176.9 (CON).

**Semi-Semi-preparative HPLC:** Rt = 11.0 min (Phenomenex Luna C18 (2), 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (7:93) → (13:87), 12.0 min, 10 mL/min, λ = 212 nm).
(peptide **1-naked**)




**HRMS ESI+ (m/z)** calcd. for $C_{25}H_{43}N_8O_9$ $[M+H]^+$: 599.3147, found: 599.3146.
(peptide **1-naked**)

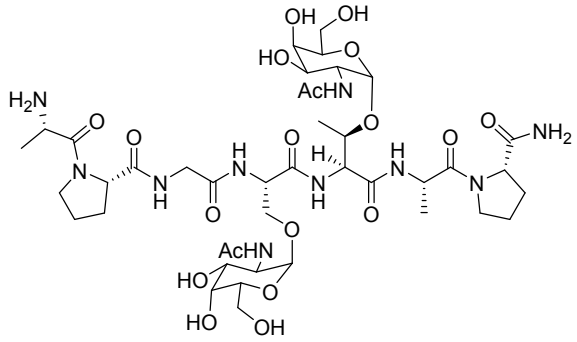



| Ion Formula | m/z | z | \|err\| [ppm] | Mean err [ppm] | mSigma |
|---|---|---|---|---|---|
| C25H42N8NaO9 | 621.296696 | 1+ | 0.4 | -0.3 | 23.1 |
| C25H43N8O9 | 599.314751 | 1+ | 0.3 | -0.2 | 28.1 |

## Glycopeptide 2

Synthesis and characterization of glycopeptide **2** has been previously described.[4]

**Glycopeptide 3**



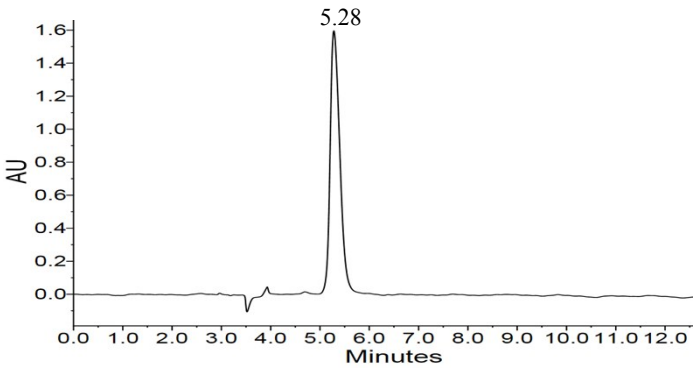
**$^{1}$H NMR (400 MHz, $D_2O$) δ (ppm):** 1.28 (d, 3H, *J* = 6.4 Hz, $CH_{3Thr}$), 1.39 (d, 3H, *J* = 7.2 Hz, $CH_{3Ala6}$), 1.55 (d, 3H, *J* = 7.0 Hz, $CH_{3Ala1}$), 1.93 – 2.15 (m, 12H, $H\beta_{Pro2}$, $H\beta_{Pro7}$, $2H\gamma_{Pro2}$, $2H\gamma_{Pro7}$, $NHC\underline{H}_{3SThr}$, $NHC\underline{H}_{3SSer}$), 2.28 – 2.41 (m, 2H, $H\beta_{Pro2}$, $H\beta_{Pro7}$), 3.62 – 3.80 (m, 8H, $2H\delta_{Pro7}$, $2H\delta_{Pro11}$, $2H_{6SThr}$, $2H_{6SSer}$), 3.85 – 4.16 (m, 12H, $2H\beta_{Ser}$, $2H\alpha_{Gly}$, $H_{3SThr}$, $H_{4SThr}$, $H_{5SThr}$, $H_{2SThr}$, $H_{3SSer}$, $H_{4SSer}$, $H_{5SSer}$, $H_{2SSer}$), 4.28 – 4.42 (m, 3H, $H\alpha_{Pro7}$, $H\alpha_{Ala1}$, $H\beta_{Thr}$), 4.49 – 4.57 (m, 1H, $H\alpha_{Pro2}$), 4.57 – 4.63 (m, 2H, $H\alpha_{Ala6}$, $H\alpha_{Thr}$), 4.83 (t, 1H, *J* = 4.7 Hz, $H\alpha_{Ser}$), 4.90 (d, 1H, *J* = 3.8 Hz, $H_{1SThr}$), 4.95 (d, 1H, *J* = 3.8 Hz, $H_{1SSer}$).

**$^{13}$C NMR (100 MHz, $D_2O$) δ (ppm):** 15.0 ($CH_{3Ala1}$), 15.6 ($CH_{3Ala6}$), 18.3 ($CH_{3Thr}$), 22.1, 22.4 ($NHCO\underline{C}H_{3SThr,}$ $NHCO\underline{C}H_{3SSer}$), 24.6, 24.7 ($C\gamma_{Pro2}$, $C\gamma_{Pro7}$), 29.4, 29.5 ($C\beta_{Pro2}$, $C\beta_{Pro7}$), 42.2 ($C\alpha_{Gly}$), 47.3 ($C\alpha_{Ala6}$), 47.6, 47.7 ($C\delta_{Pro7}$, $C\delta_{Pro2}$), 48.0 ($C\alpha_{Ala1}$), 49.6, 49.7 ($C_{2SThr}$, $C_{2SSer}$), 53.4 ($C\alpha_{Ser}$), 56.9 ($C\alpha_{Thr}$), 60.0 ($C\alpha_{Pro7}$), 60.5 ($C\alpha_{Pro2}$), 61.2, 61.3 ($C_{6SThr}$, $C_{6SSer}$), 67.7, 67.8, 68.2, 68.4, 68.6, 71.3 ($C_{4SThr}$, $C_{5SThr}$, $C_{3SThr}$, $C_{4SSer}$, $C_{5SSer}$, $C_{3SSer}$), 71.3 $C\beta_{Ser}$), 76.1 ($C\beta_{Thr}$), 98.1 ($C_{1SThr}$), 98.6 ($C_{1SSer}$), 169.3, 170.1, 171.0, 171.4, 172.4, 173.7, 174.1, 174.3, 176.9 (CON).

**Semi-preparative HPLC:** Rt = 16.0 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (0:100) →(10:90), 20 min, 20 mL/min, λ = 212 nm)

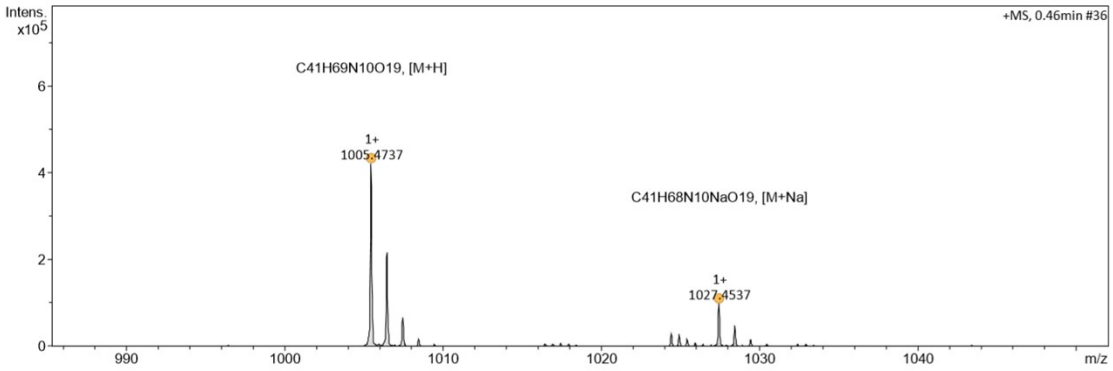
**Analytical HPLC:** Rt = 5.3 min (Phenomenex Luna 5 µm C18 (2), 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (7:93) →(13:87), 12 min, 1 mL/min, λ = 212 nm)

(glycopeptide **3**)



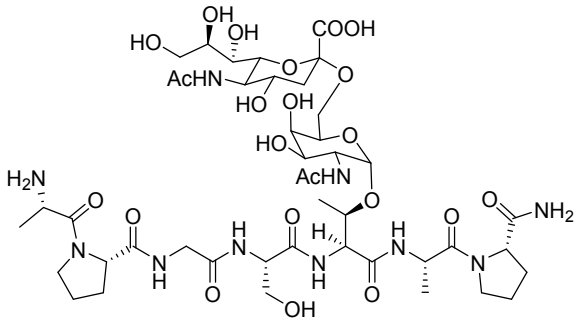
**HRMS ESI+ (m/z):** calcd for $C_{41}H_{69}N_{10}O_{19}[M+H]^+$: 1005.4735, found 1005.4737.

(glycopeptide **3**)




| Ion Formula | m/z | z | \|err\| [ppm] | Mean err [ppm] | mSigma |
|---|---|---|---|---|---|
| C41H68N10NaO19 | 1027.455441 | 1+ | 1.7 | 519.6 | 4.7 |
| C41H69N10O19 | 1005.473496 | 1+ | 0.2 | 529.6 | 8.4 |

**Glycopeptide 4**



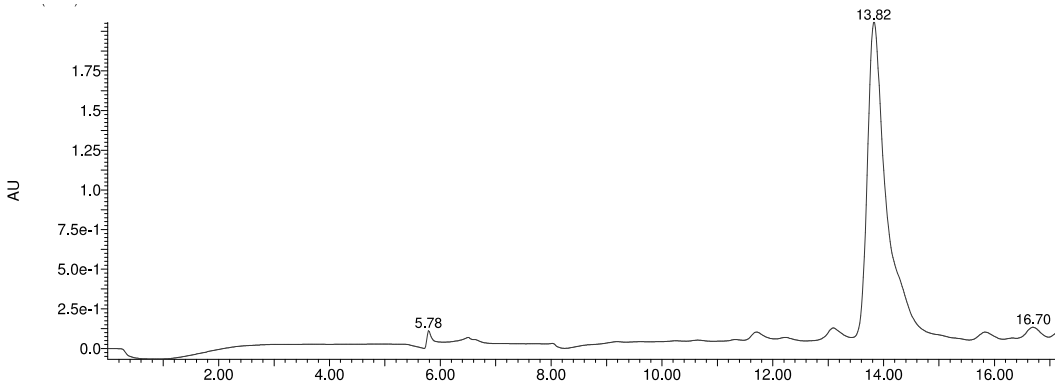
**$^1$H NMR (400 MHz, $D_2O$) δ:** 1.30 (d, 3H, *J* = 6.6 Hz, $CH_{3Thr}$), 1.38 (d, 3H, *J* = 7.0 Hz, $CH_{3Ala6}$), 1.55 (d, 3H, *J* = 7.0 Hz, $CH_{3Ala1}$), 1.75 (t, 1H, *J* = 12.2 Hz, $H_{S3}$'ax), 1.92 – 2.16 (m, 12H, NHCOC$\underline{H}_3$', NHCOC$\underline{H}_3$, Hβ$_{Pro2}$, Hβ$_{Pro7}$, 2Hγ$_{Pro2}$, 2Hγ$_{Pro7}$), 2.28 – 2.42 (m, 2H, Hβ$_{Pro2}$, Hβ$_{Pro7}$), 2.71 (dd, 1H, *J* = 12.7, 4.7 Hz, $H_{S3}$'ec), 3.55 – 4.15 (m, 21H, 2Hδ$_{Pro2}$, 2Hδ$_{Pro7}$, 2Hβ$_{Ser}$, 2Hα$_{Gly,}$ $H_{2S}$, $H_{3S}$, $H_{4S}$, $H_{5S}$, 2$H_{6S}$, $H_{4S}$', $H_{5S}$', $H_{6S}$', $H_{7S}$', $H_{8S}$', 2$H_{9S}$'), 4.33 – 4.43 (m, 3H, Hβ$_{Thr}$, Hα$_{Ala1}$, Hα$_{Pro7}$), 4.52 (t, 1H, *J* = 7.5, Hα$_{Pro2}$), 4.57 – 4.63 (m, 2H, Hα$_{Thr}$, Hα$_{Ala6}$), 4.68 (t, 1H, *J* = 6.0 Hz, Hα$_{Ser}$), 4.93 (d, 1H, *J* = 4.0 Hz, $H_{1S}$).

**$^1$H NMR (400 MHz, $D_2O$/$H_2O$, 1:9, amide region) δ (ppm):** 6.98 (s, $NH_{term}$), 7.66 (s, $NH_{term}$), 7.82 (d, *J* = 9.7 Hz, $NHAc_S$), 8.05 (d, *J* = 9.2 Hz, $NHAc_S$'), 8.20 (d, *J* = 6.9 Hz, $NH_{Ser}$), 8.35 (d, *J* = 5.9 Hz, $NH_{Ala6}$), 8.57 (t, *J* = 5.8 Hz, $NH_{Gly}$), 8.63 (d, *J* = 9.2 Hz, $NH_{Thr}$).

**$^{13}$C NMR (100 MHz, $D_2O$) δ (ppm):** 15.0 ($CH_{3Ala1}$), 15.4 ($CH_{3Ala6}$), 18.2 ($CH_{3Thr}$), 22.0, 22.2 (NHCO$\underline{C}H_3$', NHCO$\underline{C}H_3$), 24.6, 24.7 (Cγ$_{Pro2}$, Cγ$_{Pro7}$), 29.3, 29.5 (Cβ$_{Pro2}$, Cβ$_{Pro7}$), 39.6 ($C_{S3}$'), 42.4 (Cα$_{Gly}$), 47.5 (Cα$_{Ala6}$), 47.6, 47.7 (Cδ$_{Pro7}$, Cδ$_{Pro2}$), 48.0 (Cα$_{Ala1}$), 49.6 ($C_{2S}$), 51.8 ($C_{S5}$'), 55.1 (Cα$_{Ser}$), 57.2 (Cα$_{Thr}$), 60.0 (Cα$_{Pro7}$), 60.7 (Cα$_{Pro2}$), 61.2 (Cβ$_{Ser}$), 62.8, 67.8, 68.0, 68.3, 68.6, 69.9, 70.8, 71.3, 72.7 ($C_{3S}$, $C_{4S}$, $C_{5S}$, $C_{6S}$, $C_{4S}$', $C_{6S}$', $C_{7S}$', $C_{8S}$', $C_{9S}$'), 75.8 (Cβ$_{Thr}$), 98.7 ($C_{1S}$), 126.2 ($C_{2S}$'), 169.3, 170.6, 171.2, 172.1, 172.2, 172.5, 173.7, 174.4, 174.9 (CON), 176.9 (COO).

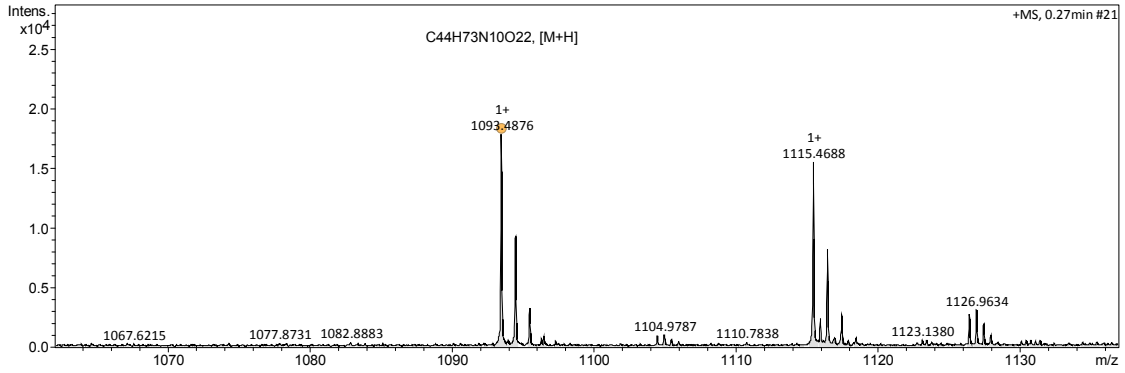
**Semi-Semi-preparative HPLC:** Rt = 13.8 min (Phenomenex Luna C18 (2), 10 μm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (13:87), 16.0 min, 10 mL/min, λ = 212 nm).

(glycopeptide **4**)



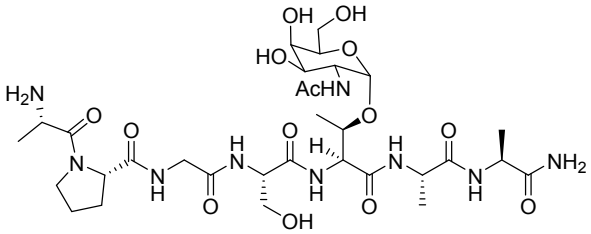

**HRMS ESI+ (m/z)** calcd. for $C_{44}H_{73}N_{10}O_{22}$ $[M+H]^+$: 1093.4895, found: 1093.4876.

(glycopeptide **4**)




| Ion Formula | m/z | z | \|err\| [ppm] | Mean err [ppm] | mSigma |
|---|---|---|---|---|---|
| C44H73N10O22 | 1093.489540 | 1+ | 1.8 | 1.6 | 5.9 |

**$A_1$PGST*$A_6A_7$**



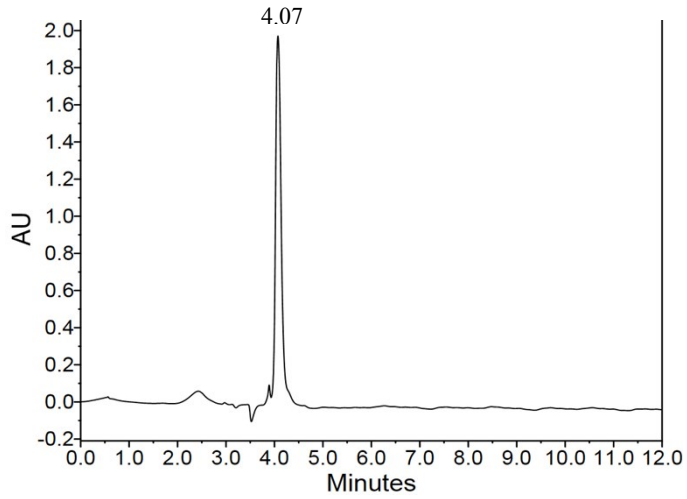
**$^1$H NMR (400 MHz, $D_2O$) δ (ppm):** 1.29 (d, 3H, *J* = 6.4 Hz, $CH_{3Thr}$), 1.37-1.43 (m, 6H, $CH_{3Ala6}$, $CH_{3Ala7}$), 1.55 (d, 3H, *J* = 7.0, $CH_{3Ala1}$), 1.94 – 2.15 (m, 6H, $H\beta_{Pro}$, $2H\gamma_{Pro}$, $NHC\underline{H}_3$), 2.32 – 2.42 (m, 1H, $H\beta_{Pro}$), 3.61 – 3.79 (m, 4H, $2H\delta_{Pro}$, $2H_{6S}$), 3.85 – 4.07 (m, 7H, $2H\beta_{Ser}$, $2H\alpha_{Gly}$, $H_{3S}$, $H_{4S}$, $H_{5S}$), 4.12 (dd, 1H, *J* = 11.0, 3.8 Hz, $H_{2S}$), 4.22 (q, 1H, *J* = 7.2 Hz, $H\alpha_{Ala6}$), 4.31– 4.43 (m, 3H, $H\alpha_{Ala7}$, $H\alpha_{Ala1}$, $H\beta_{Thr}$), 4.48 – 4.55 (m, 1H, $H\alpha_{Pro}$), 4.58 (d, 1H, *J* = 2.2 Hz, $H\alpha_{Thr}$), 4.67 (t, 1H, *J* = 6.0 Hz, $H\alpha_{Ser}$), 4.94 (d, 1H, *J* = 3.9 Hz, $H_{1S}$).

**$^{13}$C NMR (100 MHz, $D_2O$) δ (ppm):** 15.0 ($CH_{3Ala1}$), 16.6, 16.9 ($CH_{3Ala6}$, $CH_{3Ala7}$), 18.2 ($CH_{3Thr}$), 22.2 ($NHCO\underline{C}H_3$), 24.7 ($C\gamma_{Pro}$), 29.3 ($C\beta_{Pro}$), 42.4 ($C\alpha_{Gly}$), 47.7, 48.1 ($C\alpha_{Ala6}$, $C\alpha_{Ala7}$), 49.2 ($C\delta_{Pro}$), 49.3 ($C\alpha_{Ala1}$), 49.7 ($C_{2S}$), 55.1 ($C\alpha_{Ser}$), 57.4 ($C\alpha_{Thr}$), 60.7 ($C\alpha_{Pro}$), 61.2 ($C\beta_{Ser}$), 61.3 ($C_{6S}$), 68.2, 68.6, 71.4 ($C_{4S}$, $C_{5S}$, $C_{3S}$), 75.4 ($C\beta_{Thr}$), 98.6 ($C_{1S}$), 169.4, 170.9, 171.2, 172.3, 174.0, 174.3, 174.5, 177.7 (CON).

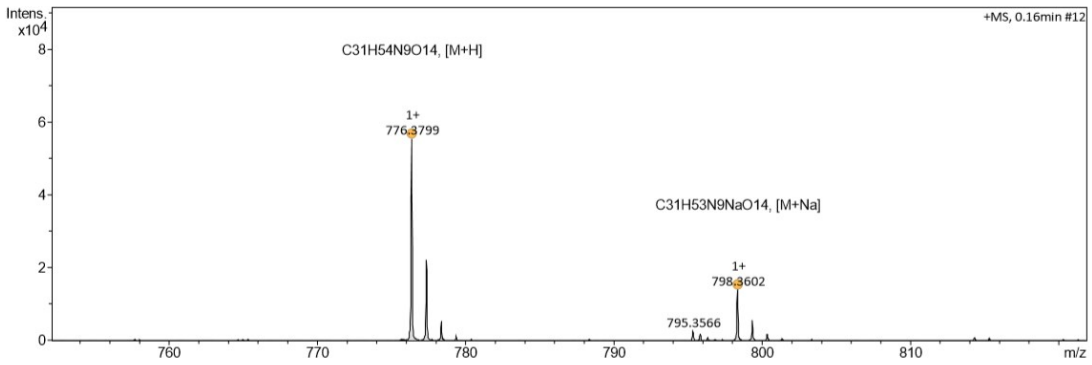
**Semi-preparative HPLC:** Rt = 13.7 min (Phenomenex Luna C18 (2), 10 μm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (0:100) →(10:90), 20 min, 20 mL/min, λ = 212 nm).

**Analytical HPLC:** Rt = 4.1 min (Phenomenex Luna 5 μm C18 (2), 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (7:93) →(13:87), 12 min, 1 mL/min, λ = 212 nm).

(APGST*AA)

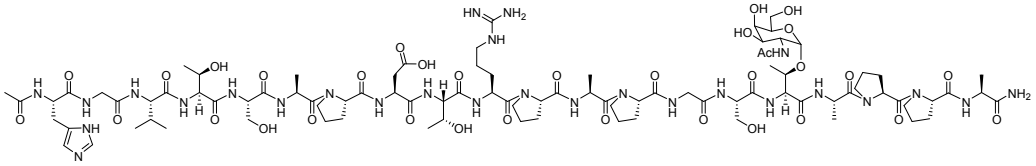


**HRMS ESI+ (m/z):** calcd for $C_{31}H_{54}N_9O_{14}[M+H]^+$: 776.3785, found 776.3799.
(APGST*AA)

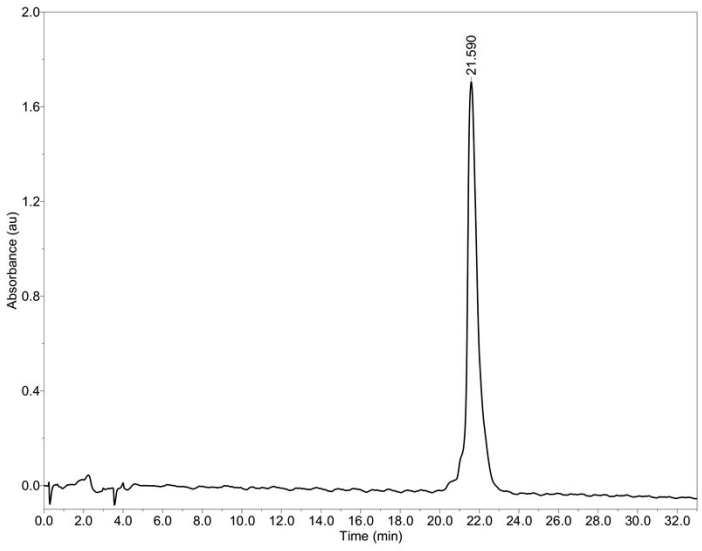



| Ion Formula | m/z | z | \|err\| [ppm] | Mean err [ppm] | mSigma |
|---|---|---|---|---|---|
| C31H53N9NaO14 | 798.360418 | 1+ | 0.3 | 3.0 | 14.2 |
| C31H54N9O14 | 776.378474 | 1+ | 1.8 | -1.4 | 10.6 |

**MUC1-Tn16**



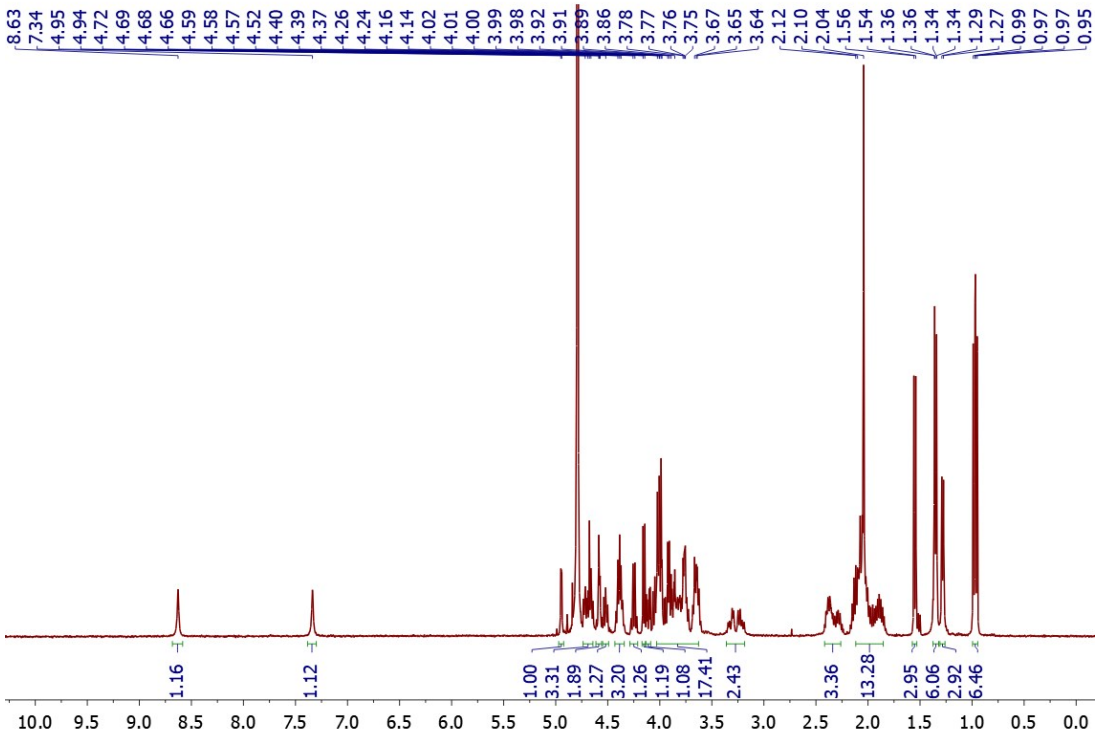
**Analytical HPLC:** Rt = 21.6 min (Phenomenex Luna C18 (2), 5 μ, 4.60x250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) →(20:80), 30 min, 1 mL/min, λ = 212 nm).



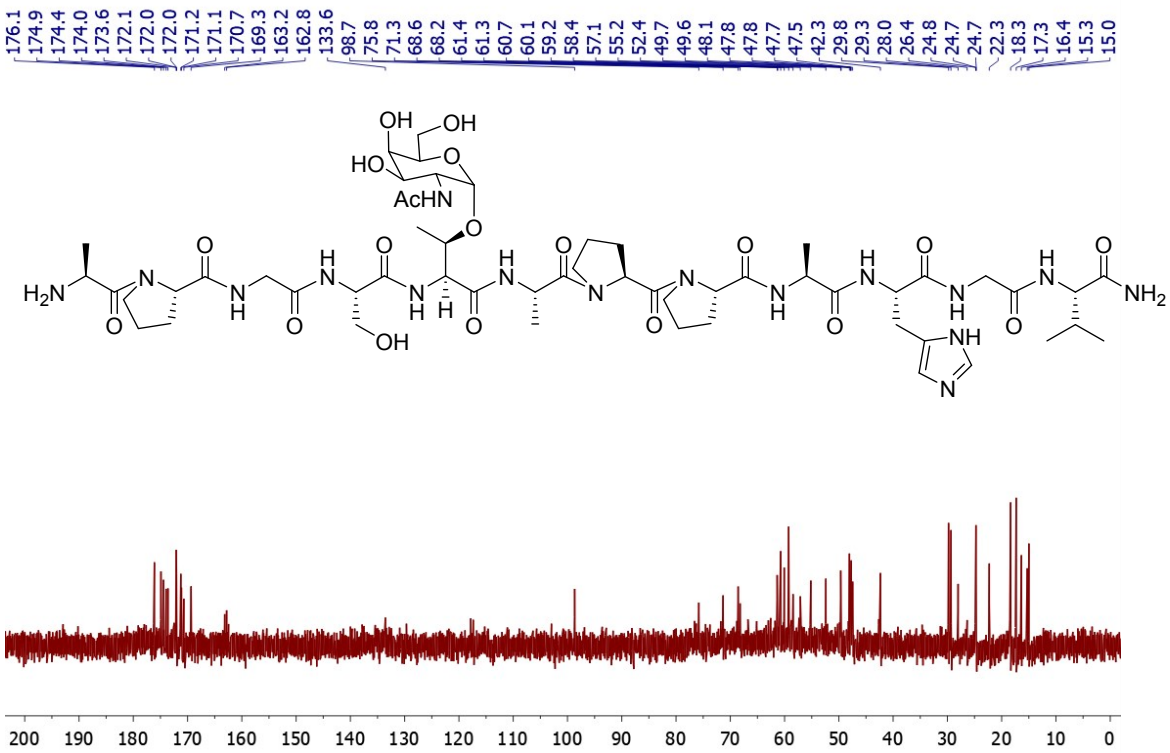

**HRMS ESI+ (m/z):** calcd for $C_{90}H_{145}N_{27}O_{33}$ $[M+2H]^{2+}$: 1066.0244, found 1066.0201.

**NMR spectra of the synthesized compounds**

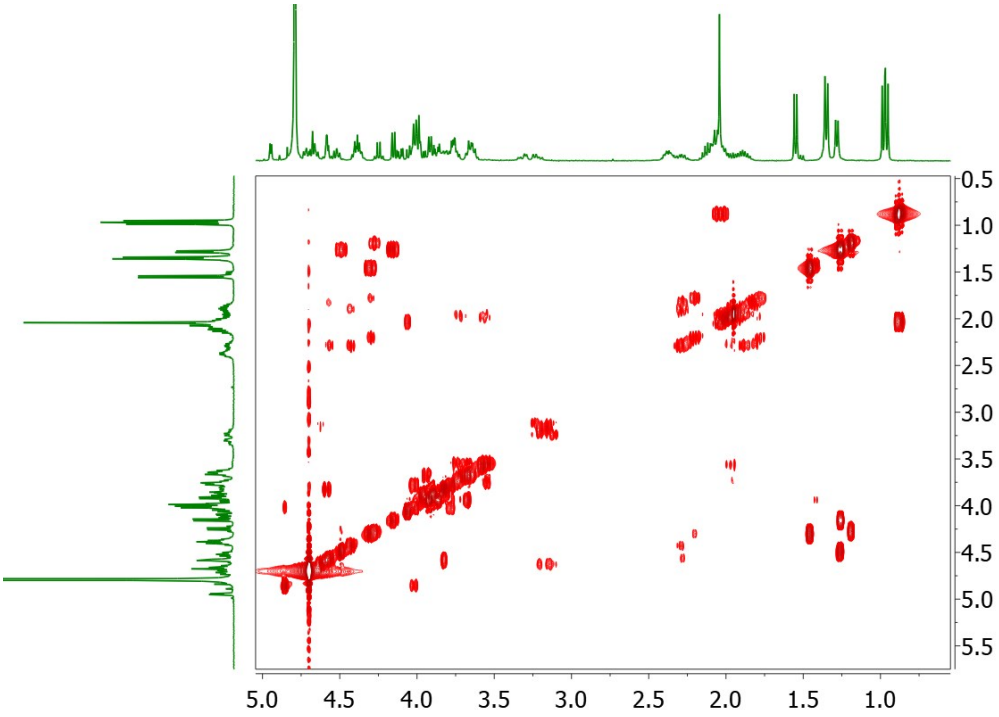
**$^{1}$H NMR 400 MHz in $D_2O$ registered at 298 K (glycopeptide 1')**



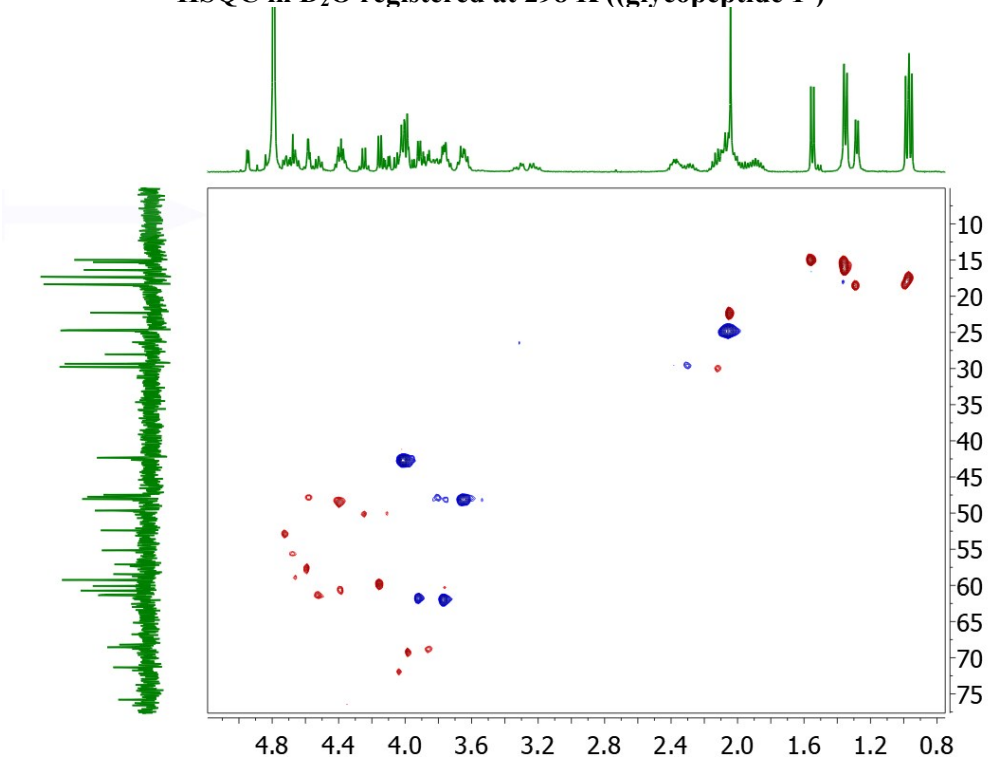

**$^{13}$C NMR 100 MHz in $D_2O$ registered at 298 K (glycopeptide 1')**



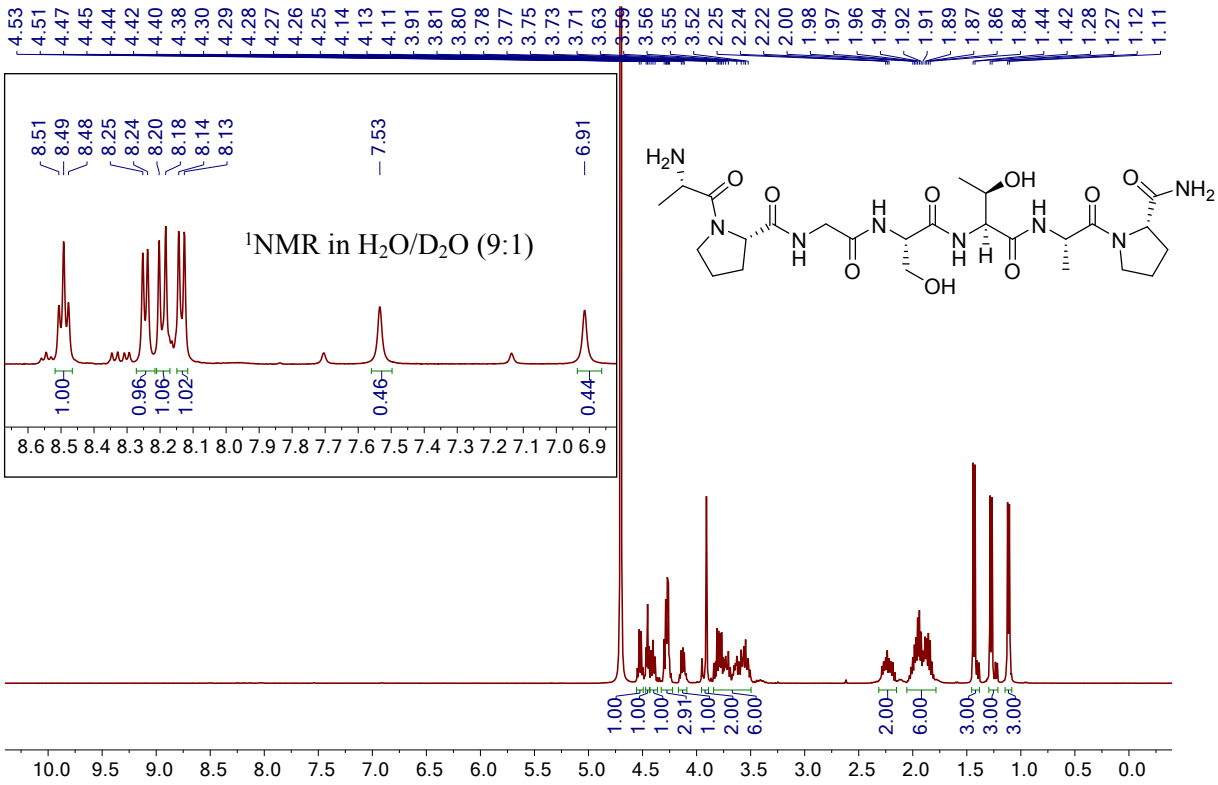
**COSY in $D_2O$ registered at 298 K (glycopeptide 1')**



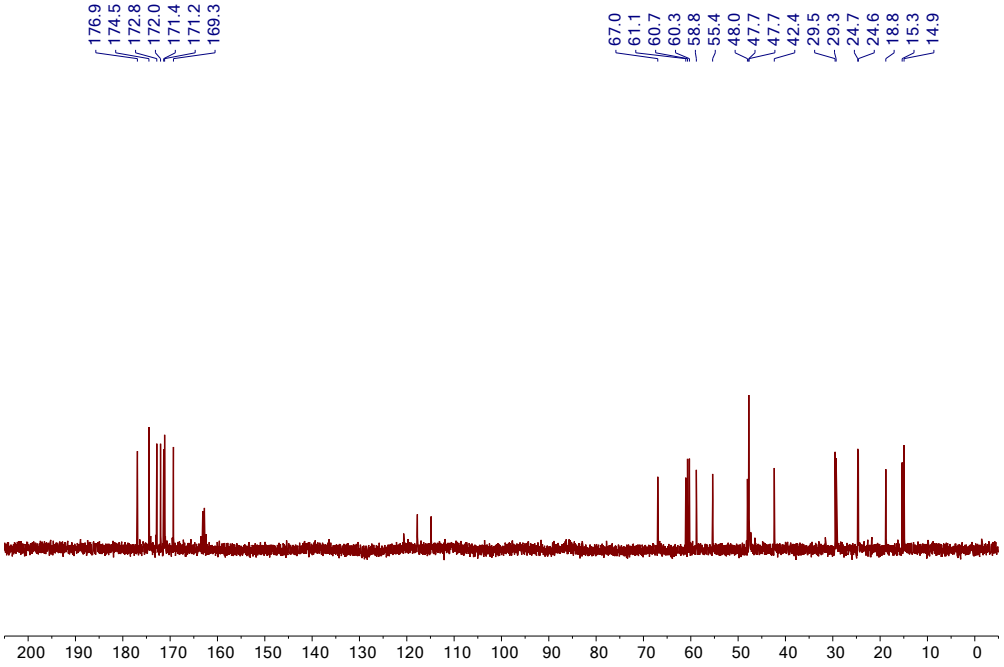

**HSQC in $D_2O$ registered at 298 K ((glycopeptide 1')**



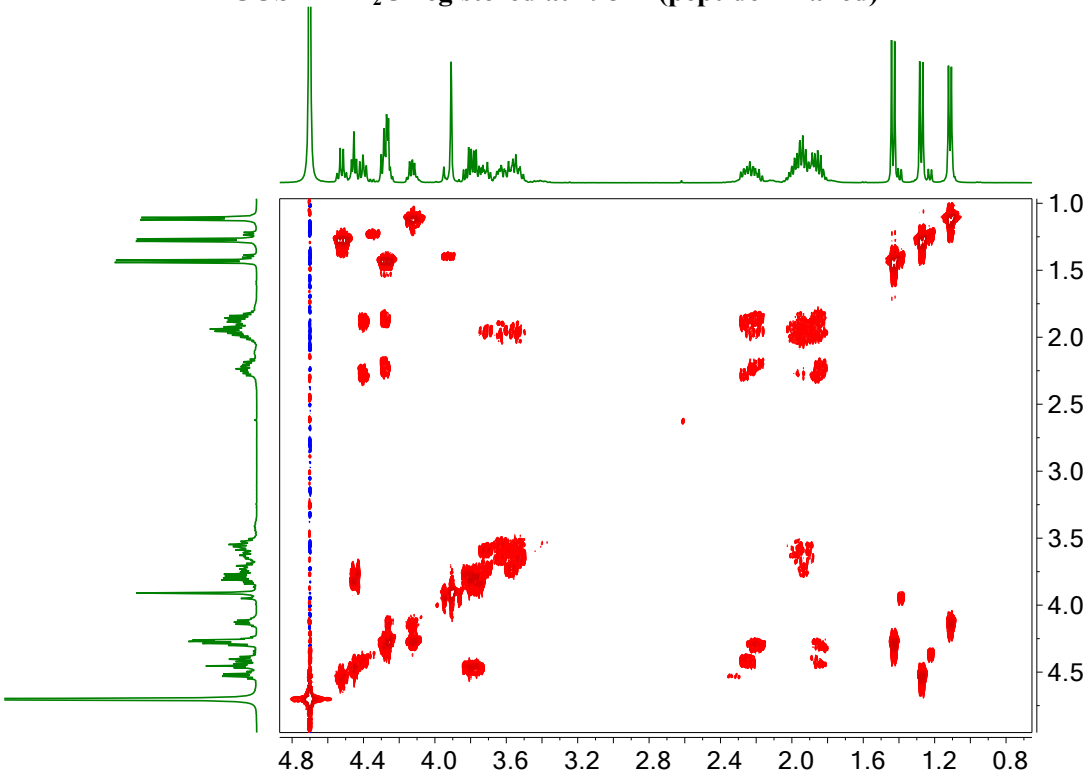
**$^{1}$H NMR 400 MHz in $D_2O$ registered at 298 K (peptide 1-naked).** A second set of signals (in a small percentage) is observed for this and other glycopeptides. They correspond to the *cis* disposition of the amide bond of proline residues.[5]



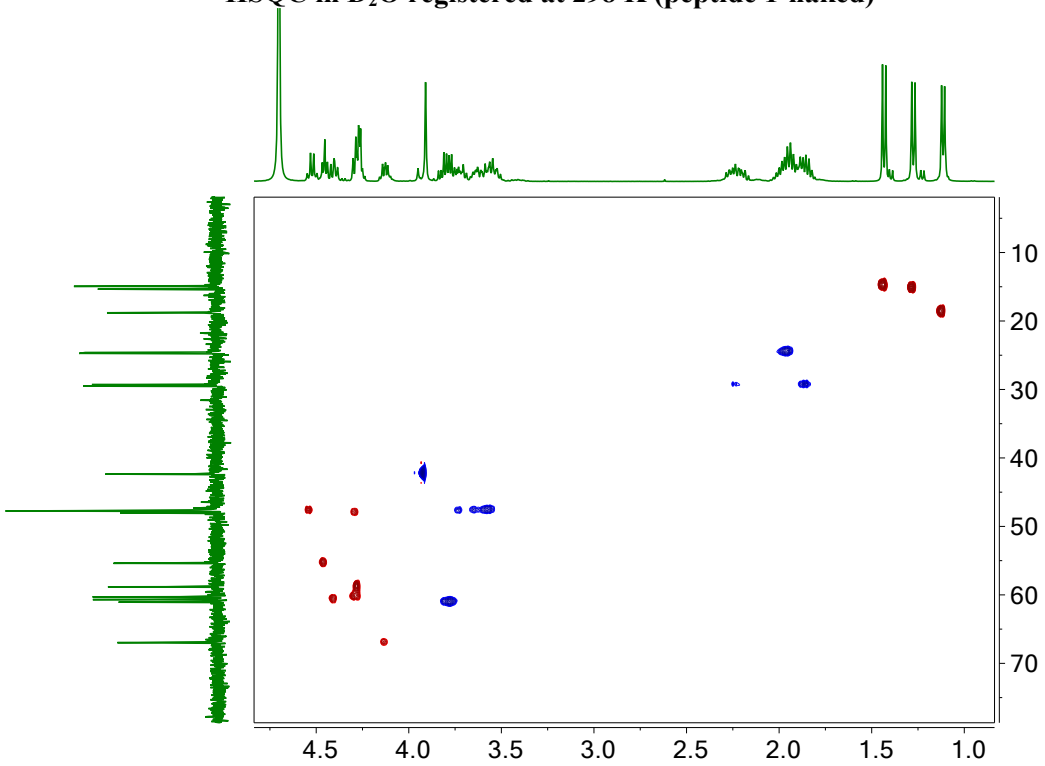

**$^{13}$C NMR 100 MHz in $D_2O$ registered at 298 K (peptide 1-naked)**



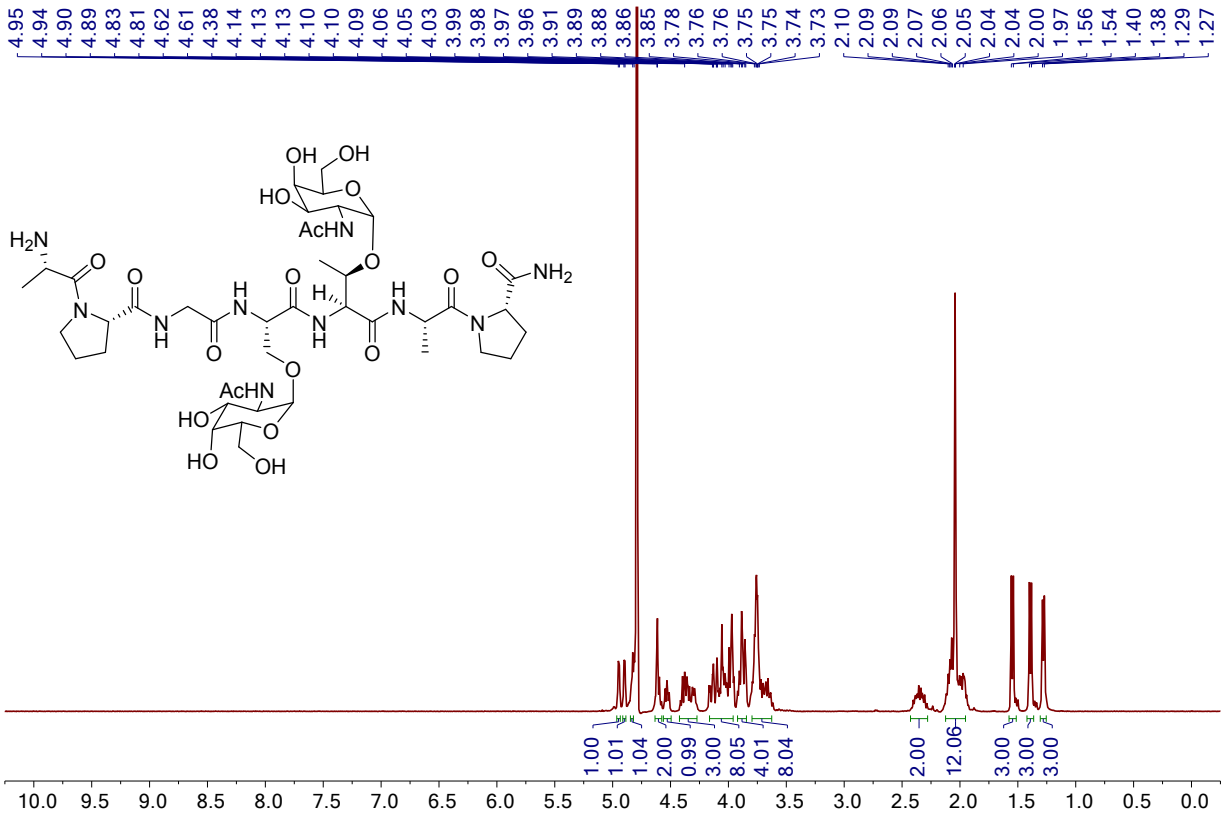
**COSY in $D_2O$ registered at 298 K (peptide 1-naked)**



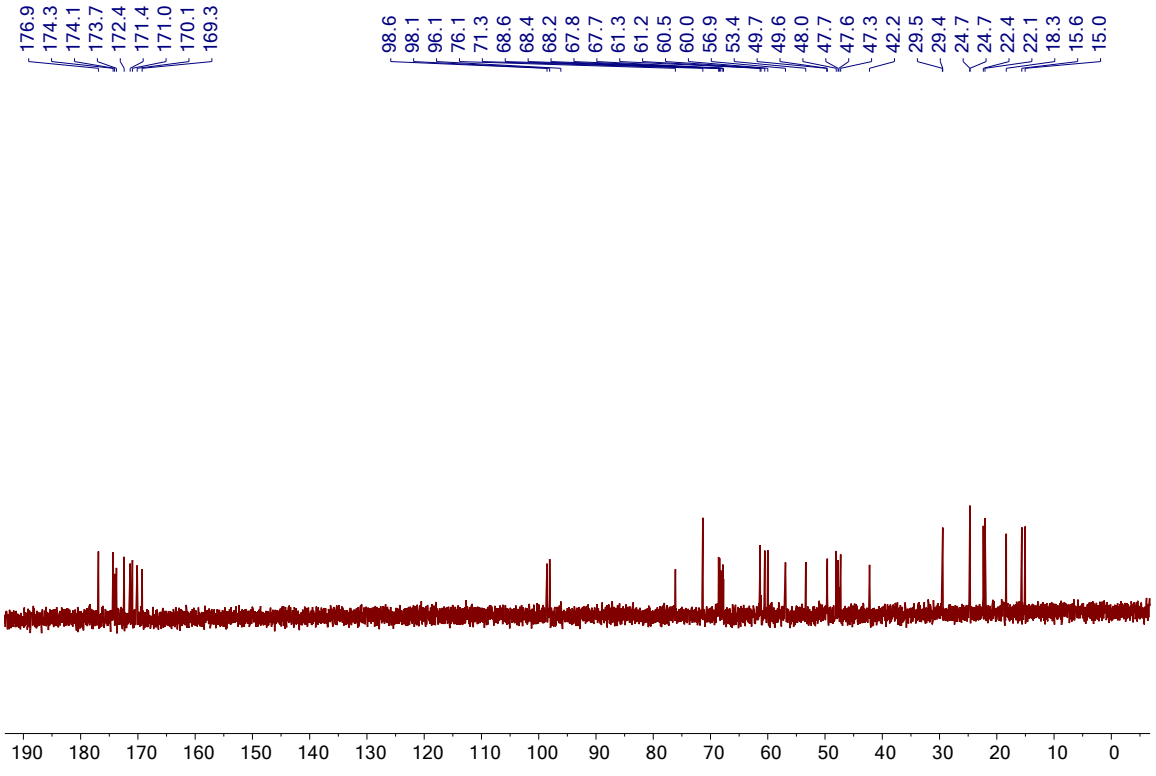

**HSQC in $D_2O$ registered at 298 K (peptide 1-naked)**



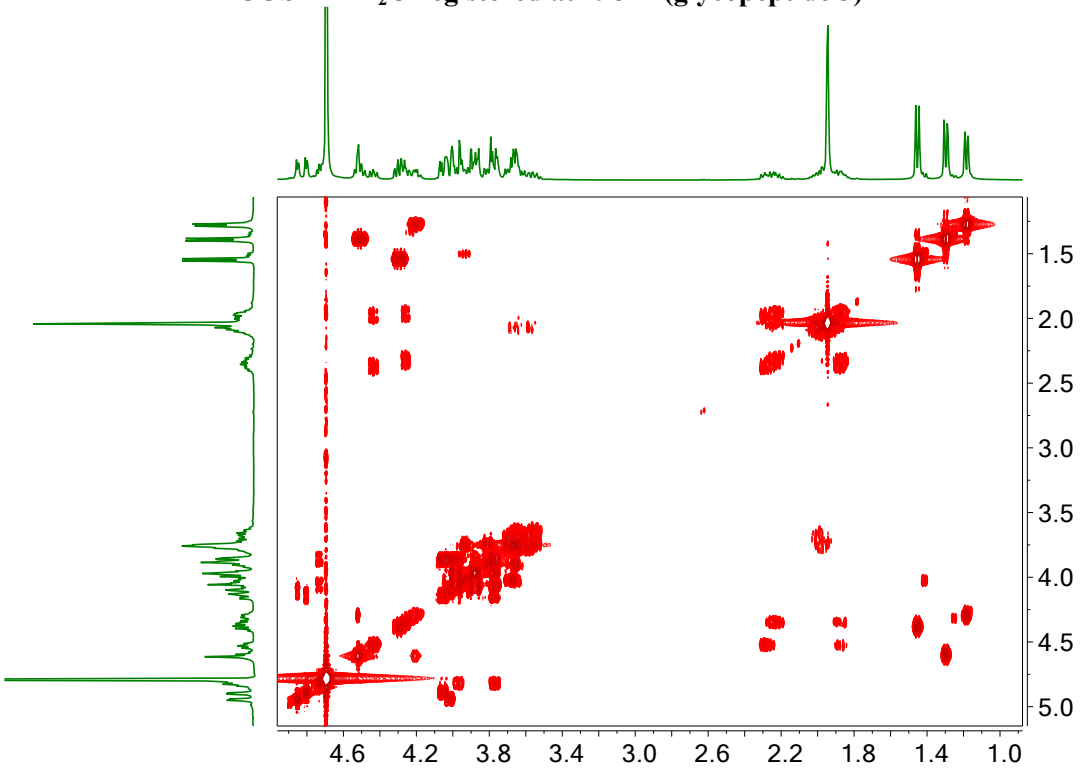
**$^{1}$H NMR 400 MHz in $D_2O$ registered at 298 K (glycopeptide 3)**



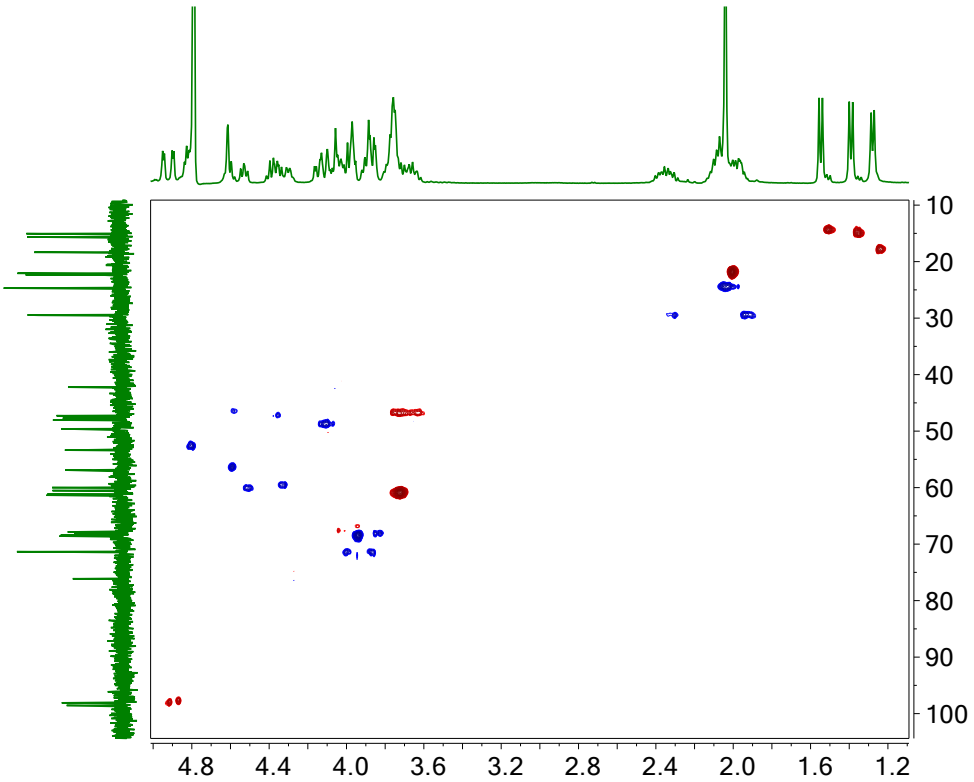

**$^{13}$C NMR 100 MHz in $D_2O$ registered at 298 K (glycopeptide 3)**



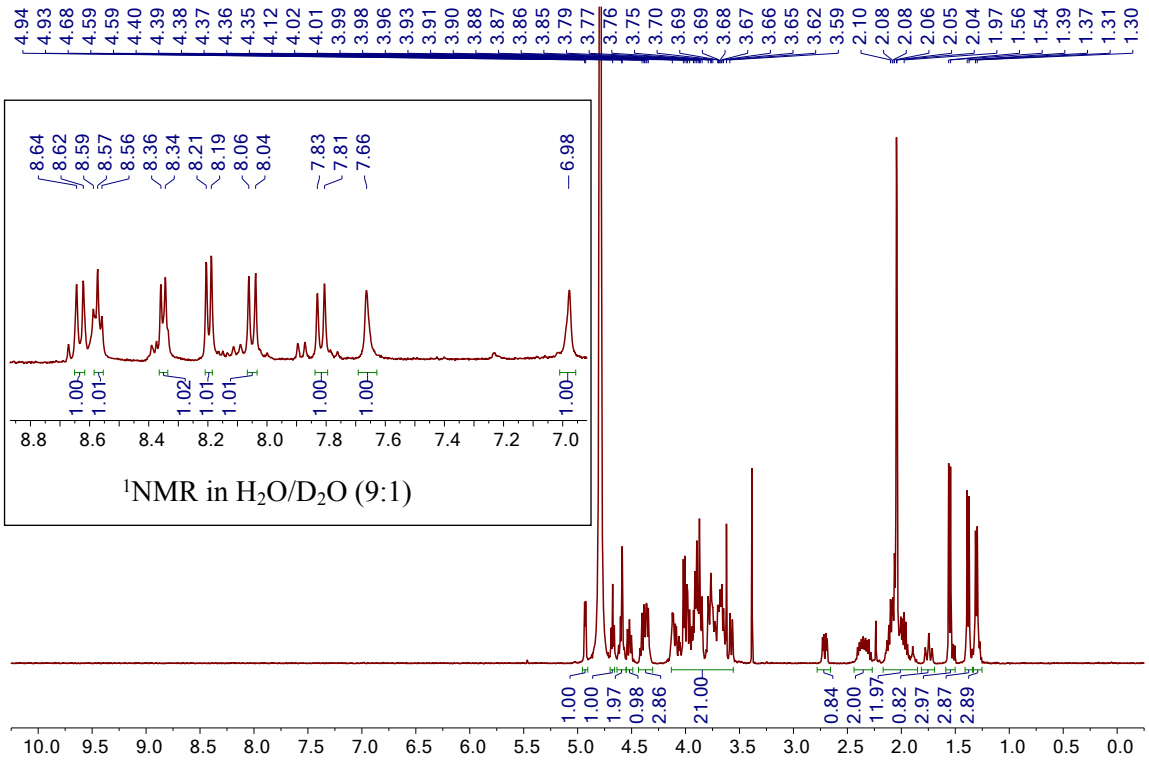
**COSY in $D_2O$ registered at 298 K (glycopeptide 3)**



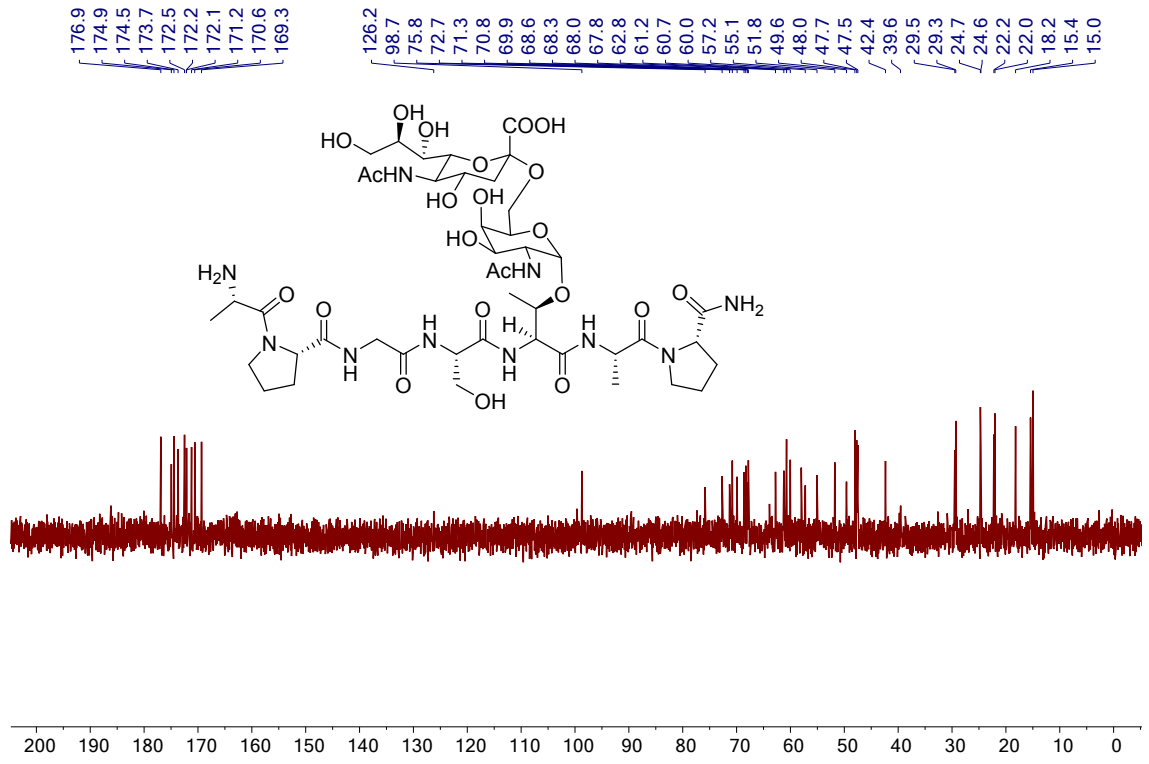

**HSQC in $D_2O$ registered at 298 K (glycopeptide 3)**



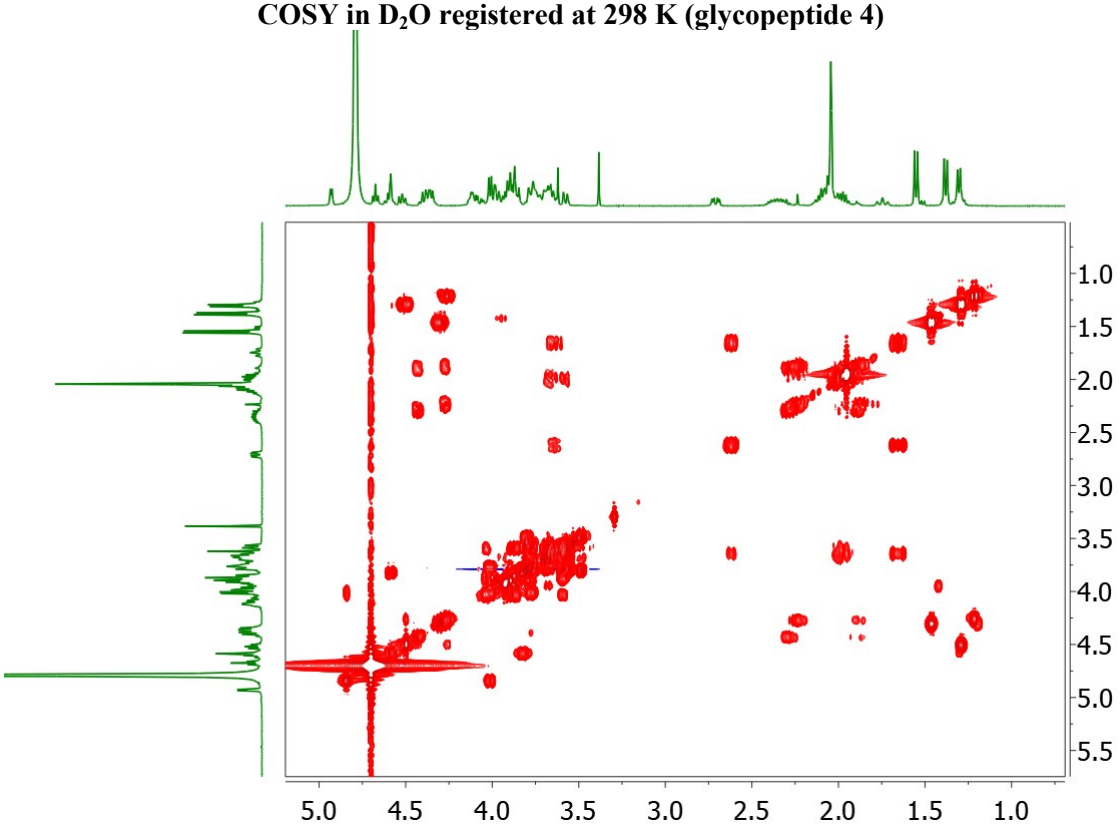
**$^{1}$H NMR 400 MHz in $D_2O$ registered at 298 K (glycopeptide 4)**



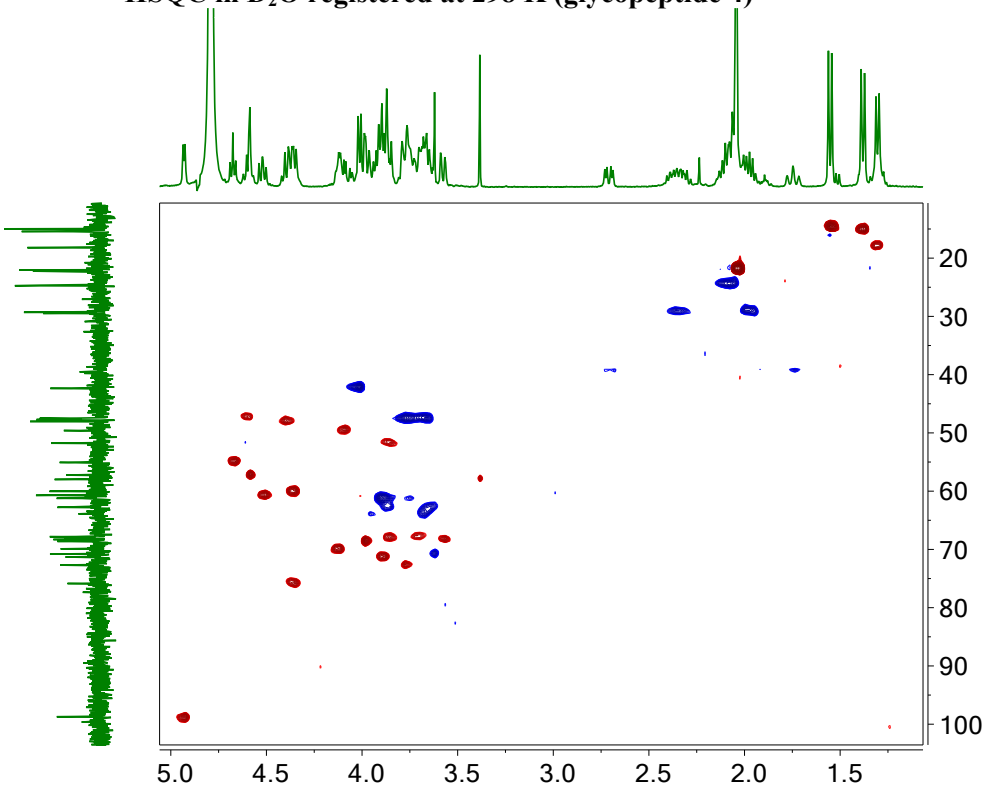

**$^{13}$C NMR 100 MHz in $D_2O$ registered at 298 K (glycopeptide 4)**



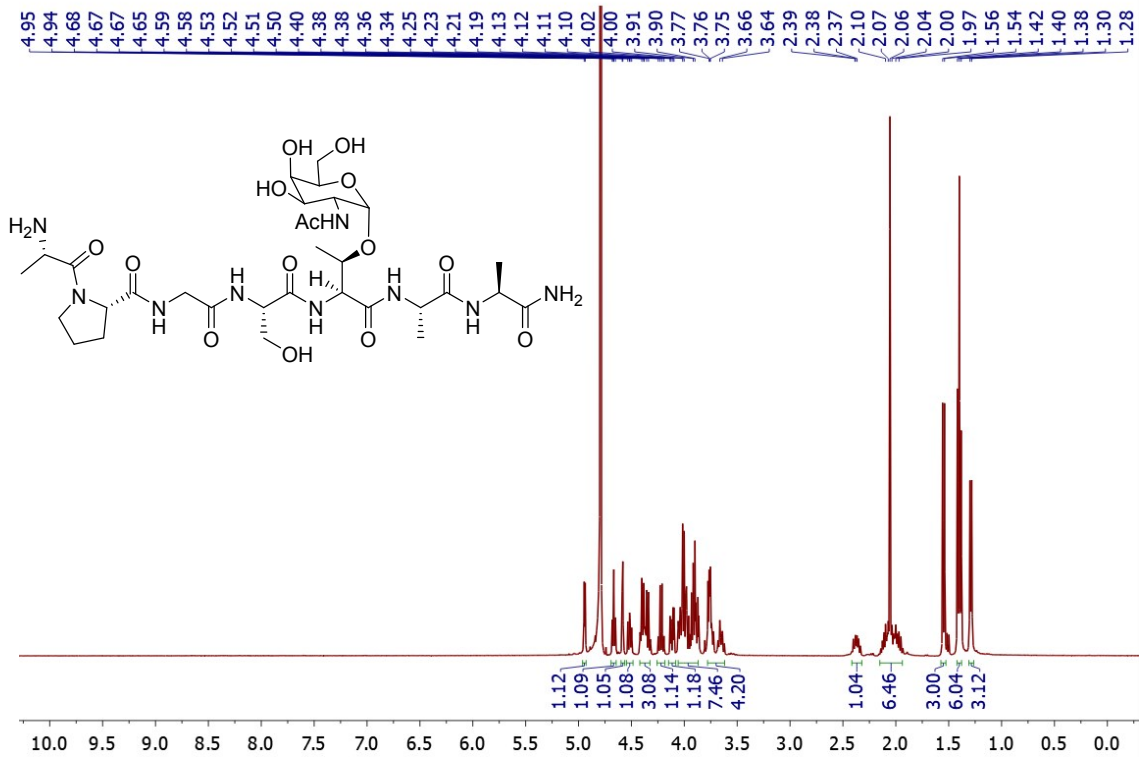
**COSY in $D_2O$ registered at 298 K (glycopeptide 4)**



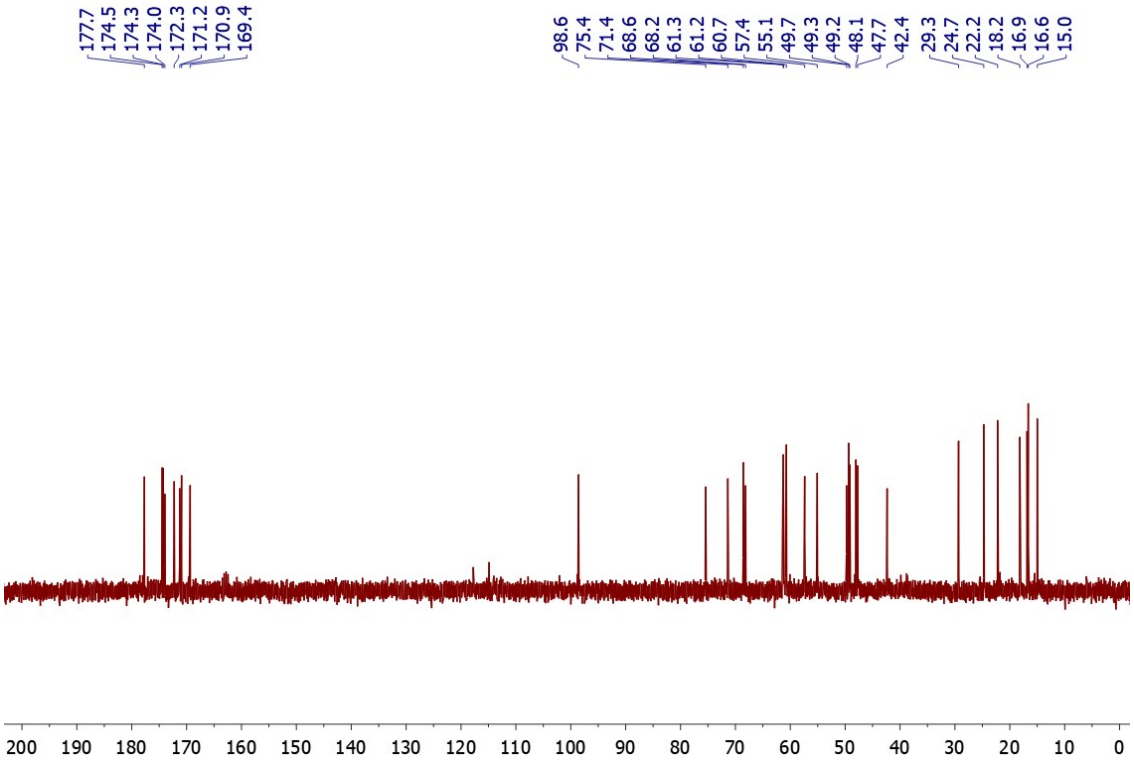

**HSQC in $D_2O$ registered at 298 K (glycopeptide 4)**



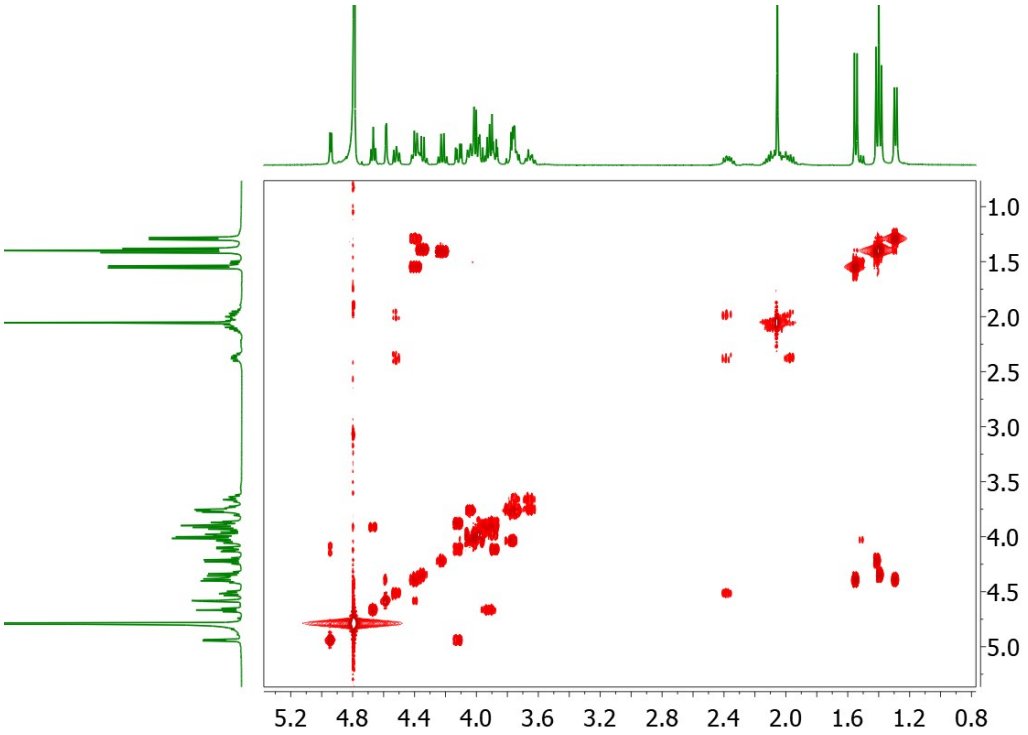
**$^{1}$H NMR 400 MHz in $D_2O$ registered at 298 K (APGST*AA)**



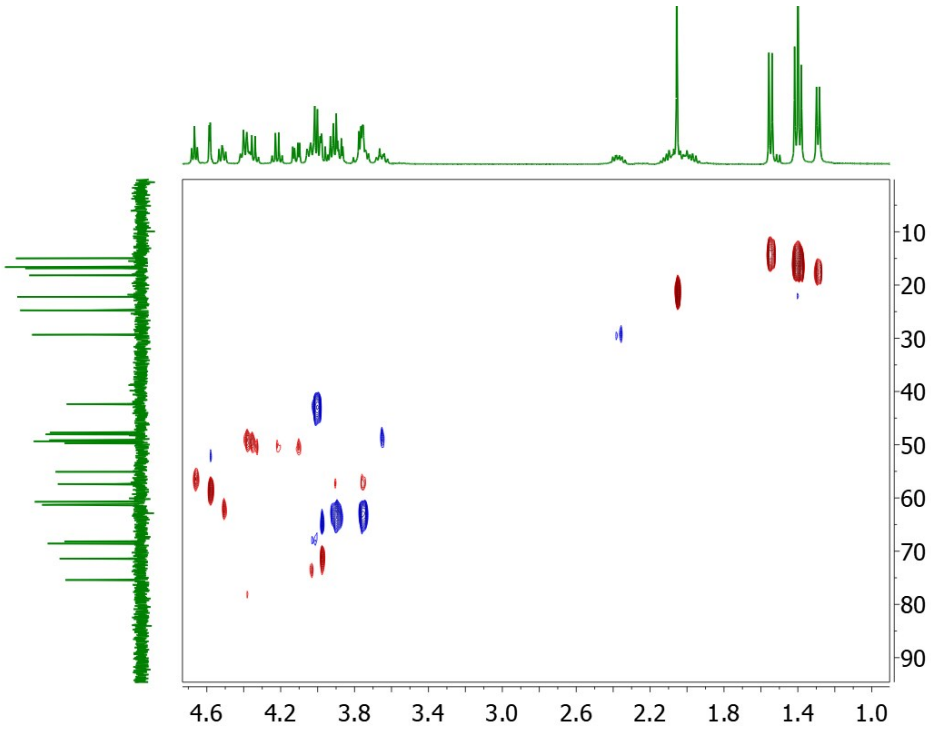

**$^{13}$C NMR 100 MHz in $D_2O$ registered at 298 K (APGST*AA)**



**COSY in $D_2O$ registered at 298 K (APGST*AA)**




**HSQC in $D_2O$ registered at 298 K (APGST*AA)**



**Expression and purification of scFv-5E5**

The DNA sequence encoding amino acid residues of the scFv-5E5 was codon optimized and synthesized by GenScript (USA) for expression in *E. coli* cells. The DNA, containing at the 5' end a recognition sequence for *Pst*I, and at the 3' end a 6×His tag, a recognition sequence for *BstE*II and a stop codon, was cloned into pHEN6c, rendering the vector pHEN6c-*scFv-5E5*-6His. scFv-5E5 mutants (H32A-H35A, Y100A and F102A) were generated by GenScript via site-directed mutagenesis using the vector pHEN6c-*ScFv-5E5*-6His.

All plasmids were transformed into WK6 cells and colonies were selected on LB/Agar plates containing 100 μg/ml of ampicillin. WK6 cells were grown at 37ºC in TB medium (0.23% (w/v) $KH_2PO_4$, 1.625% (w/v) $K_2HPO_4$, 1.2% (w/v) tryptone, 2.4% (w/v) yeast extract powder, 0.4% (v/v) glycerol, 0.1% (w/v) glucose and 2 mM $MgCl_2$) containing 100 μg/ml of ampicillin. Once the cells reached an optical density around 0.6, they were induced with the addition of 400 μM IPTG (isopropyl β-D-thiogalactoside) and the incubation time was prolonged for a further 18 h.

The cells were harvested by centrifugation at 10,000 rpm at 4 ºC for 10 min and resuspended in 4×TES buffer (0.5 M Sucrose, 0.2 M TRIS pH 8, 0.5 mM EDTA) in order to facilitate the breakage of the cell membrane. The cells were incubated for at least 6 h at 4 ºC on an orbital shaking platform. Then they were diluted in 1x TES buffer, and incubated overnight again on an orbital shaking platform. The resulting solution was centrifuged at 11500×g at 4 ºC for 30 min and the supernatant (containing the periplasmic extract) was dialyzed against buffer A (25 mM TRIS pH 7.5, 300 mM NaCl pH7.5), and loaded into a His-Trap Column (GE Healthcare). Protein was eluted with an imidazole gradient in buffer A from 10 mM up to 500 mM. The fractions containing the desired protein were dialyzed against buffer B (25 mM TRIS pH 9) and loaded into a HiTrap QFF Anion Exchange Column (GE Healthcare) and were eluted with a NaCl gradient in buffer B up to 1 M. Buffer exchange to Buffer C (25 mM TRIS pH 7.5, 50 mM NaCl) was carried out using a HiPrep 26/10 Desalting Column (GE Healthcare).

Quantification of protein was carried out by absorbance at 280 nm using his theoretical extinction coefficient ($\varepsilon^{280\ nm}$ (scFv-5E5) = 43890 $M^{-1}cm^{-1}$).

| | |
|---|---|
| ***ScFv-5E5*** | Q V Q L Q Q S D A E L V K P G S S V K I S C K A S G Y T F T D H A I H W V |
| ***ScFv-5E5*** | K Q K P E Q G L E W I G H F S P G N T D I K Y N D K F K G K A T L T V D R |
| ***ScFv-5E5*** | S S S T A Y M Q L N S L T S E D S A V Y F C K T S T F F F D Y W G Q G T T |
| ***ScFv-5E5*** | L T V S S S S G G G G S G G G G G S S G S S E L V M T Q S P S S L T V T A |
| ***ScFv-5E5*** | G E K V T M I C K S S Q S L L N S G D Q K N Y L T W Y Q Q K P G Q P P K L |
| ***ScFv-5E5*** | L I F W A S T R E S G V P D R F T G S G S G T D F T L T I S S V Q A E D L |
| ***ScFv-5E5*** | A V Y Y C Q N D Y S Y P L T F G A G T K L E L K G G H H H H H H |

**Figure S1.** Sequence of the construct of scFv-5E5 used in this work. The heavy chain is colored in brown and the light chain in blue. The linker and the histidine tag are indicated in yellow and dark blue, respectively.

**Isothermal titration microcalorimetry (ITC).** ITC was used to characterize the interaction of scFv-5E5 and its corresponding mutants with different ligands. All experiments were carried out in an Auto-iTC200 (Microcal, GE Healthcare) at 25ºC with scfv-5E5 and its mutants at 20-30 μM and ligands ranging from 300 μM to 2 mM in 25 mM TRIS pH 7.5 150 mM NaCl. The experiments were performed in duplicate. Data integration, correction and analysis were carried out in Origin 7 (Microcal). The data were fitted to a one-site equilibrium-binding model.

**Table S1.** Thermodynamic binding parameters obtained by ITC for scFv-5E5 and three mutants with several glycopeptides at 25 °C and pH 7.5.

| Protein | Ligand | $K_D$ (μM) | ΔG (cal/mol) | ΔH (cal/mol) | -T*ΔS (cal/mol) | n |
|---|---|---|---|---|---|---|
| scFv-5E5 | GalNAc-α-1-OMe | n.d.[a] | | | | |
| scFv-5E5 | APGSTAP (**1-naked**) | n.d.[a] | | | | |
| scFv-5E5 | APGST(Tn)AP (**1**) | 0.96 ±0.207 | -8205.5 | -726.1 ± 30.1 | -7479.4 | 1.2 ± 0.02 |
| scFv-5E5 | APGST(Tn)APPAHGV (**1'**) | 1.28 ±0.40 | -8035.2 | -5383.9 ± 276.1 | -2651.2 | 1.2 ± 0.03 |
| scFv-5E5 | APDT(Tn)RP (**2**) | 22.5 ± 3.3 | -6337.0 | -1885.1 ± 119.8 | -4451.9 | 1.2 ± 0.04 |
| scFv-5E5 | APGS(Tn)T(Tn)AP (**3**) | 4.24 ± 0.46 | -7325.9 | -10726.1 ± 376.1 | 3400.0 | 1.0 ± 0.02 |
| scFv-5E5 | APGST(STn)AP (**4**) | 19.0 ± 4.9 | -6438.0 | -467.4 ± 77.7 | -5970.6 | 0.7 ± 0.03 |
| scFv-5E5 | APGST(Tn)AA | 382.0 ± 44.7 | -4660.5 | -6542.5 ± 854.3 | 1881.9 | 1.0 ± 0.02 |
| scFv-5E5-H32A$^{H}$/H35$^{H}$A | APGST(Tn)AP (**1**) | 1275.4 ± 5 17.5 | -3946.6 | -16383.6 ± 8549.1 | 12436.4 | 1.2 ± 0.01 |
| scFv-5E5-Y100$^{L}$A | APGST(Tn)AP (**1**) | 757.6 ± 99.7 | -4255.1 | -1340.2 ± 220.0 | -2914.8 | 1.2 ± 0.08 |
| scFv-5E5- F102$^{H}$A | APGST(Tn)AP (**1**) | n.d.[a] | | | | |

[a] n.d. stands for binding not detected under our experimental conditions or an acceptable fitting could not be attained. T* = GalNAc-α-1-*O*-Thr and S* = GalNAc-α-1-*O*-Ser (Tn antigen) and T** = Neu5Ac-α-(2-6)-GalNAc-α-1-*O*-Thr (STn antigen).

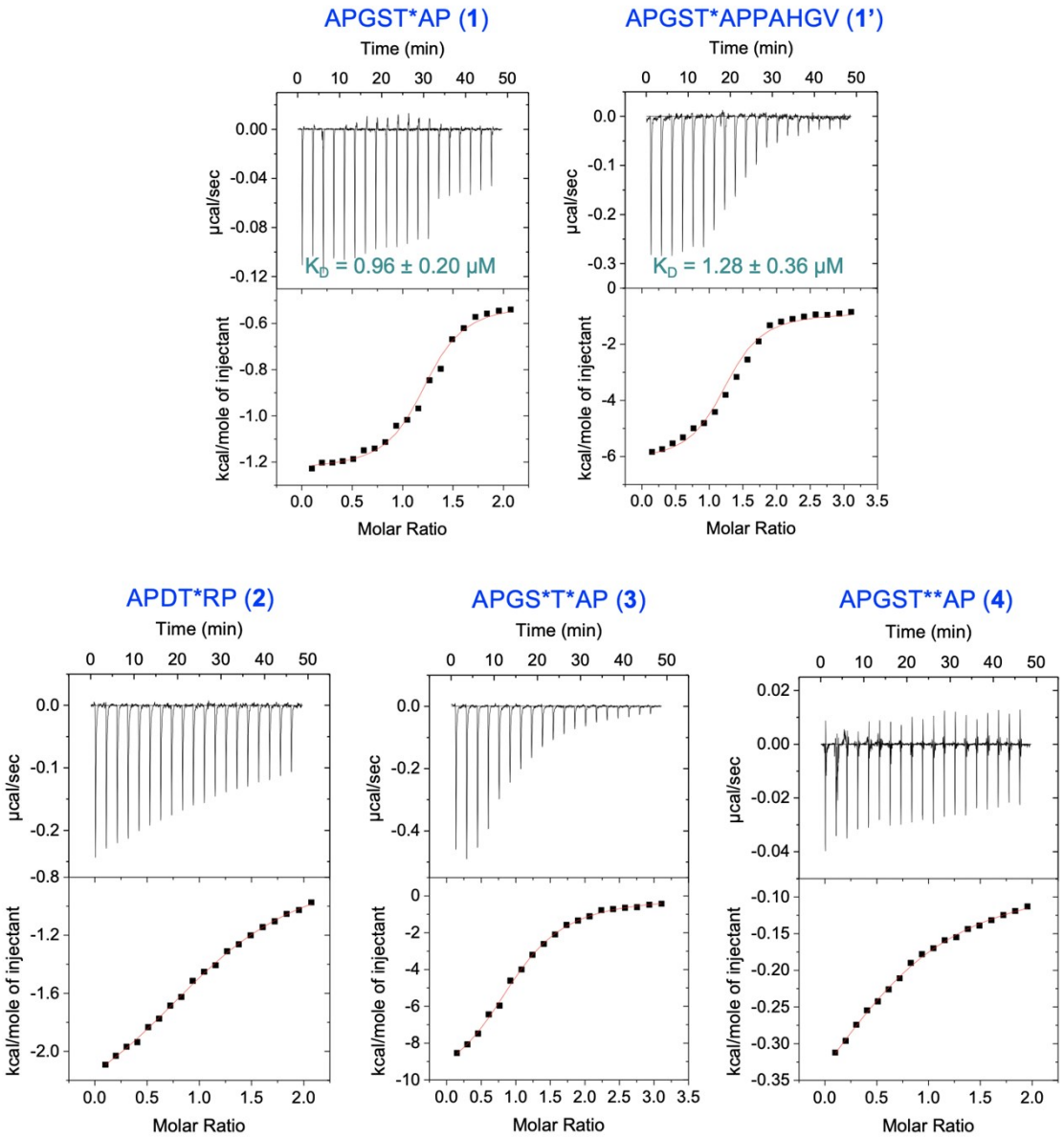



**Figure S2.** ITC profiles of glycopeptides **1-4** and **1'** with scFv-5E5 at 25 °C and pH 7.5. The solid line represents the least-squares fitting of the data to the simplest model (one binding site). The interaction of glycopeptide **4** with scFv-5E5 is weak and consequently the $K_D$ value derived from these assays should be considered in a qualitative manner ($K_D \geq 19$ µM).T* = GalNAc-α-1-*O*-Thr (Tn antigen) and T** = Neu5Ac-α-(2-6)-GalNAc-α-1-*O*-Thr (STn antigen).

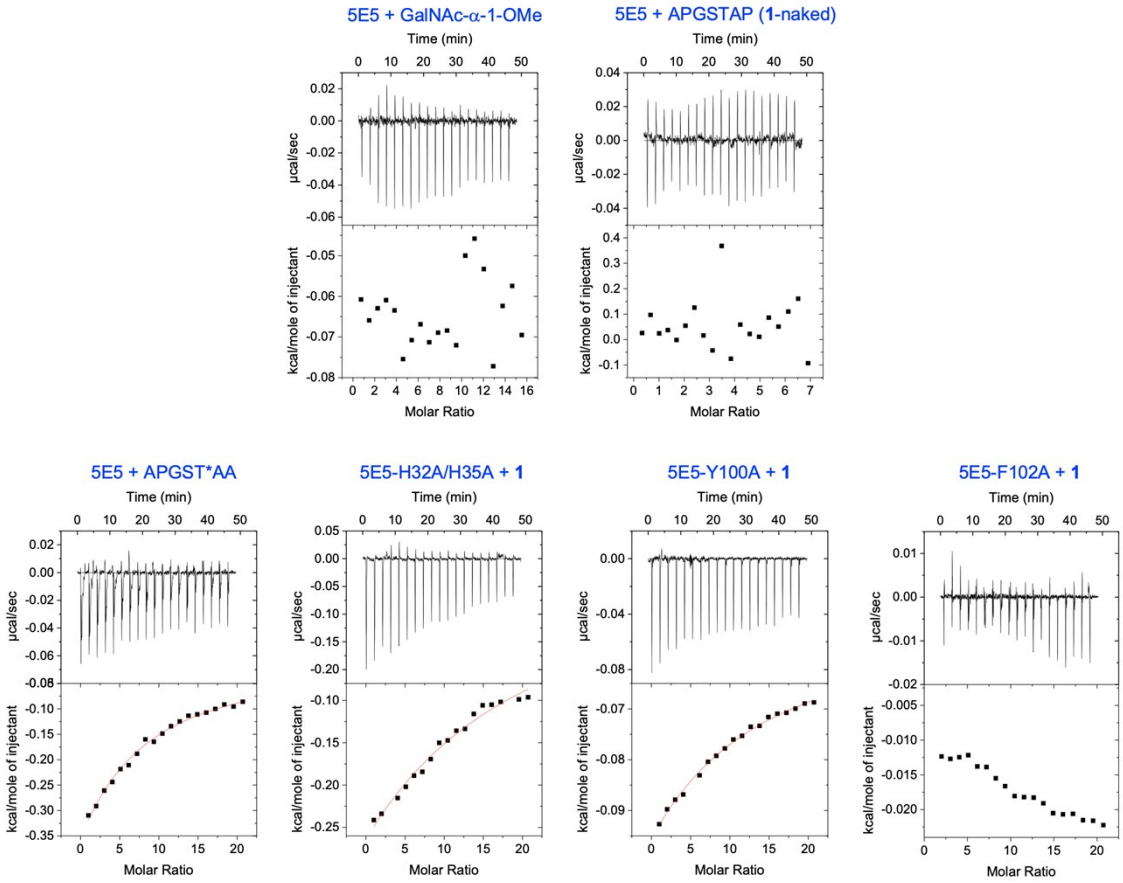



**Figure S3.** ITC profiles of GalNAc-α-1-OMe, peptide **1-naked**, and glycopeptide APGST*AA with scFv-5E5 at 25 °C and pH 7.5. The ITC profiles of glycopeptide **1** with several 5E5 mutants are also shown. The solid line represents the least-squares fitting of the data to the simplest model (one binding site). T* = GalNAc-α-1-*O*-Thr (Tn antigen). These antibody-ligand interactions shown here are weak and consequently the $K_D$ values derived from these assays should be considered in a qualitative manner.

**MicroScale Thermophoresis (MST).** Binding assays of antibody 5E5 to peptide **MUC1-Tn16** were studied with MST using a Monolith NT.115 Pico instrument (NanoTemper Technologies). Prior to the binding experiments, the antibody was labeled with reactive dyes using *N*-hydroxysuccinimide (NHS)–ester chemistry, which reacts efficiently with the primary amines of proteins to form highly stable dye–protein conjugates (Monolith NTTM Protein Labeling kit RED – NHS 2$^{nd}$ generation, NanoTemper Technologies).

For protein labeling, the proteins concentration was adjusted to 20 μM using the labeling buffer (supplied by the kit). Lower concentrations may result in loss of coupling efficiency. The solid fluorescent dye was dissolved in 100% DMSO at a concentration of about 600 μM and mixed thoroughly. Before mixing the protein and the dye, the concentration of the dye was adjusted to 3-fold concentration of the protein using the labeling buffer. Then, the protein and the fluorescent dye solutions were mixed in 1:1 ratio and incubated for 30 min at room temperature in the dark. Unreacted 'free' dye was eliminated by spin desalting columns (Zeba™ Spin Desalting Columns, ThermoFisher). Experiments were conducted in binding buffer (50 mM HEPES pH 7.5, 50 mM NaCl) using a constant concentration of the two proteins (5 nM) and serial dilutions of each ligand spanning from 10 mM to 150 nM of the glycopeptides. Duplicate measurements were performed in Premium capillaries at 25ºC using 40% laser power and 20% LED power in order to obtain optimized results. Dissociation constant ($K_D$) were calculated using MO Affinity analysis.

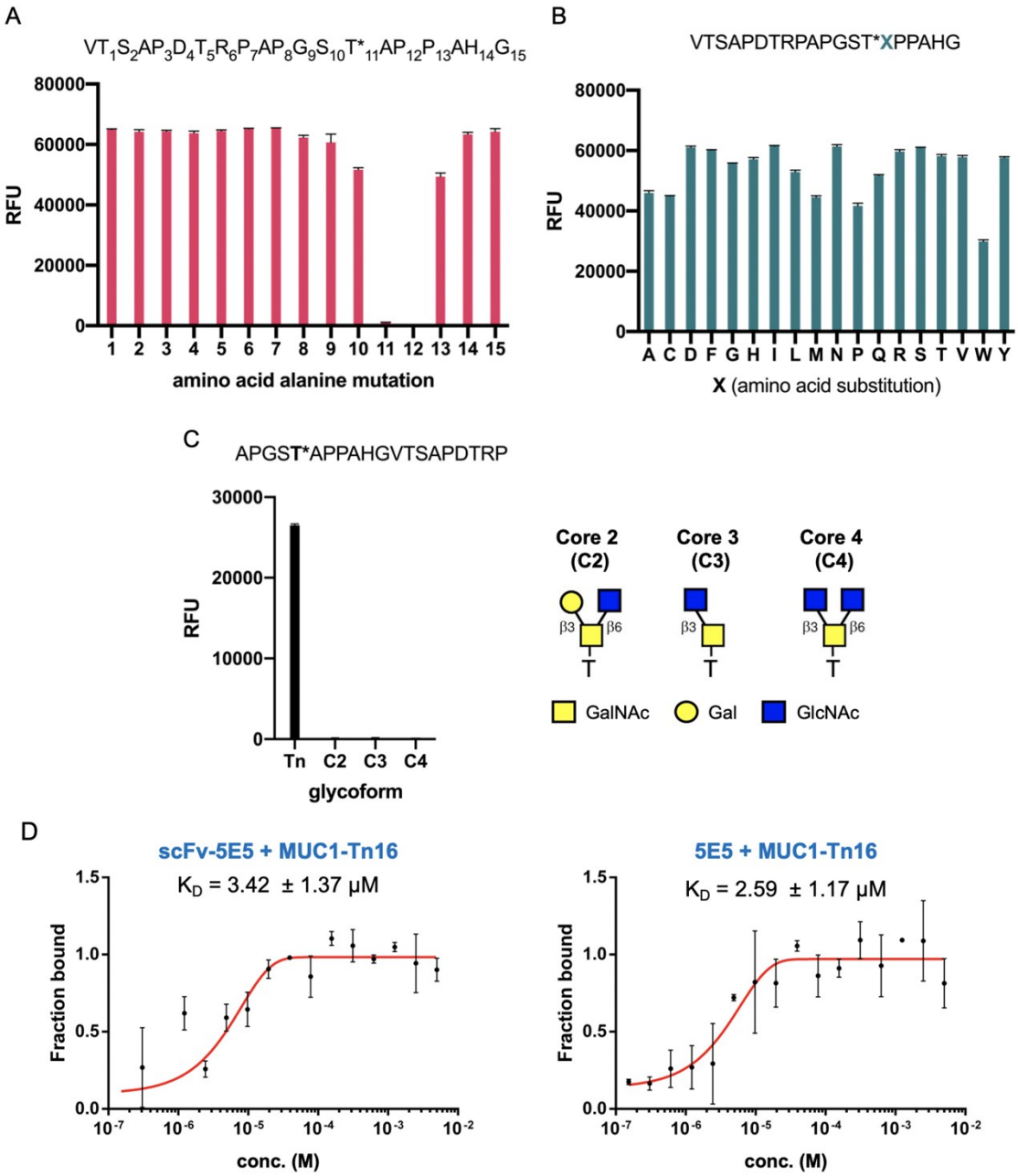



**Figure S4. (**A) Microarray data for an Ala mutation walk through the entire 20-mer MUC1-Tn glycopeptide. (B) Microarray data for the T*–X–P mutations in the GST*AP region. (C) Microarray data for the additional glycoform isomers **C2**, **C3** and **C4**. (D) MST binding curves used to determine $K_D$ for the scFv-5E5 (left panel) and complete antibody 5E5 (right panel) with glycopeptide **MUC1- Tn16**. Error represents SD between two separate experiments; $K_D$ values ± SD are shown.

**Crystallization of complex scFv-5E5/1 and data collection**. Crystals of the ScFv-5E5 were grown by sitting drop experiments at 18 ºC by mixing 0.4 µl of protein solution (6.7 mg/mL ScFv-5E5 and 5 mM glycopeptide **1** in buffer C) with an equal volume of a reservoir solution (0.2 M ammonium chloride, 0.1M HEPES 7.5, 25% v/v glycerol ethoxylate). The crystals were cryoprotected in mother liquor containing 20% glycerol and flash frozen in liquid nitrogen.

**Structure determination and refinement of complex scFv-5E5/1.** Diffraction data were collected on synchrotron beamline I03 of the Diamond Light Source (Harwell Science and Innovation Campus, Oxfordshire, UK) at a wavelength of 0.97 Å and a temperature of 100 K. Data were processed and scaled using XDS[6] and CCP4[7,8] software packages. Relevant statistics are given in Table S2. The crystal structure was solved by molecular replacement with Phaser[7,8] using the PDB entry 5YD5 as the template (note that this is a distant type of scFv that served us as template). Initial phases were further improved by cycles of manual model building in Coot[9] and refinement with REFMAC5.[10] Further rounds of Coot and refinement with REFMAC5 were performed to obtain the final structure. The final model was validated with PROCHECK, model statistics are given in Table S2. The asymmetric unit of the $P2_12_12_1$ crystal contained six molecules of ScFv-5E5. The Ramachandran plot for the complex scFv-5E5/**1** show that 86.10%, 12.90%, 0.60% and 0.30% of the amino acids are in most favoured, allowed, generously allowed and disallowed regions, respectively.

**Table S2.** Data collection and refinement statistics of complex **scFv-5E5/1**. Values in parentheses refer to the highest resolution shell. Ramachandran plot statistics were determined with PROCHECK.

| | **scFv-5E5/1** |
|---|---|
| Space group | $P2_12_12_1$ |
| Wavelength (Å) | 0.97 |
| Resolution (Å) | 20.00-3.00 (3.16-3.00) |
| Cell dimensions (Å) | $a$ = 86.57<br>$b$ = 113.19<br>$c$ = 151.20 |
| Unique reflections | 30356 |
| Completeness | 99.7 (100.0) |
| $R_{pim}$ | 0.072 (0.620) |
| Mn(I) half-set correlation CC(1/2*)* | 0.993 (0.503) |
| $I/\sigma(I)$ | 7.7 (2.1) |
| Redundancy | 6.8 (6.7) |
| $R_{work}$ / $R_{free}$ | 0.202/0.250 |
| RMSD from ideal geometry, bonds (Å) | 0.010 |
| RMSD from ideal geometry, angles (º) | 1.786 |
| <*B*> protein (Å$^2$) | 101.33 |
| <*B*> glycopeptide (Å$^2$) | 106.16 |
| <*B*> solvent (Å$^2$) | 65.11 |
| Ramachandran plot:<br>Most favoured (%)<br>Allowed (%)<br>Generously allowed (%)<br>Disallowed (%) | <br>86.80<br>13.00<br>0.60<br>0.50 |
| PDB ID | 6TNP |

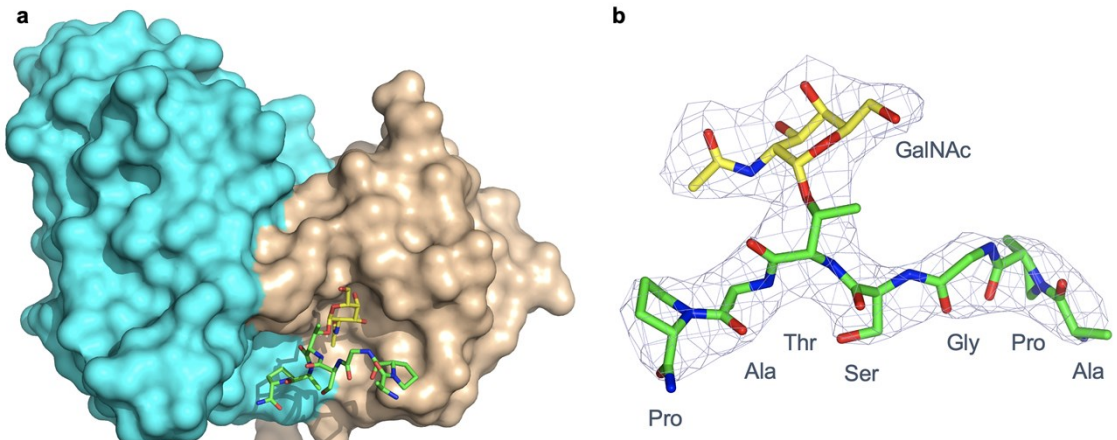



**Figure S5.** (A) Surface representation of scFv-5E5 complexed with glycopeptide **1**. The carbohydrate moiety lies in a deep solvent excluded pocket, whereas most of the peptide fragment is exposed to the solvent. (B) The $F_O$–$F_C$ electron density map (blue) is contoured at 2.2 σ for glycopeptide **1**.

**Saturation-Transfer Difference (STD) NMR:** These experiments were recorded on a Bruker Avance 600 MHz spectrometer equipped with a triple channel cryoprobe (Bruker pulse sequence: stddiffesgp) at 310K. We used a concentration of 830 μM for glycopeptide **1** and **1'** and 30 μM of scFv-5E5 in 25 mM perdeuterated TRIS (D11)-DCl (Sigma–Aldrich) pD 7.5 in $D_2O$ containing 150 mM NaCl and 0.1% sodium azide. Spectra were acquired with 2112 scans in a matrix with 64K data. An excitation sculpting module with gradients was used to suppress the water proton signals. Selective saturation of the protein resonances (on resonance spectrum) was performed by irradiating at 7 ppm (aromatic residues) using a series of 40 Eburp2.1000-shaped 90° pulses (50 ms) for a total saturation time of 2s and a relaxation delay of 3s (D1 on stddiffesgp bruker sequence). For the reference spectrum (off resonance), the samples were irradiated at 100 ppm. Control STD-NMR experiments were performed with glycopeptide **1** and **1'** at the same concentration and using the same experimental setup but in the absence of scFv-5E5. The STD spectra were obtained by subtracting the on-resonance spectrum to the off-resonance spectrum.

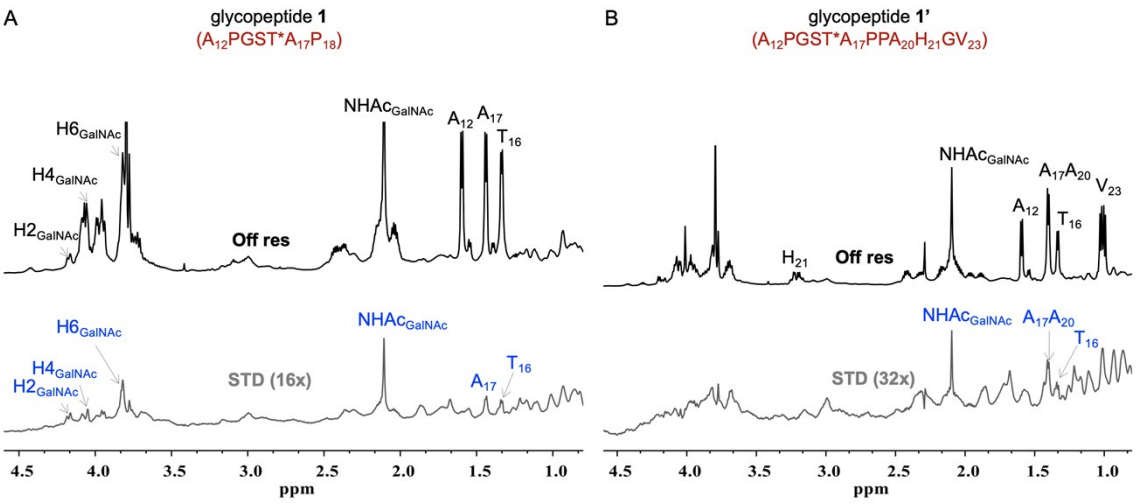



**Figure S6.** (A) STD-NMR spectrum of glycopeptide **1** at 830 μM in the presence of 30 μM scFv-5E5 obtained at 310 K. (B). STD-NMR spectra of glycopeptide **1'** at 830 μM in the presence of 30 μM scFv-5E5 obtained at 310 K. The off-resonance reference spectrum (labeled Off res) is displayed in black, and STD spectrum (labeled STD) is in dark grey. Key proton resonances are labeled in the STD spectra.

**Molecular Dynamics simulations.** The simulations were carried out with AMBER 18 package[11] implemented with ff14SB[12] and GLYCAM06j[13] force fields. The X-ray structure of complex scFv-5E5/**1** was used as the starting structure. This structure was conveniently modified with PyMOL to obtain the desired complex. In all cases, the complex was immersed in a water box with a 10 Å buffer of TIP3P water molecules.[14] The system was neutralized by adding explicit counter ions ($Cl^-$). A two-stage geometry optimization approach was performed. The first stage minimizes only the positions of solvent molecules and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems were then gently heated by incrementing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions. Harmonic restraints of 30 kcal·mol$^{-1}$ were applied to the solute, and the Andersen temperature-coupling scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Long-range electrostatic effects were modelled using the particle-mesh-Ewald method.[15] An 8 Å cut-off was applied to Lennard-Jones interactions. Each system was equilibrated for 2 ns with a 2-fs time step at a constant volume and temperature of 300 K. Production trajectories were then run for additional 0.5 μs under the same simulation conditions.

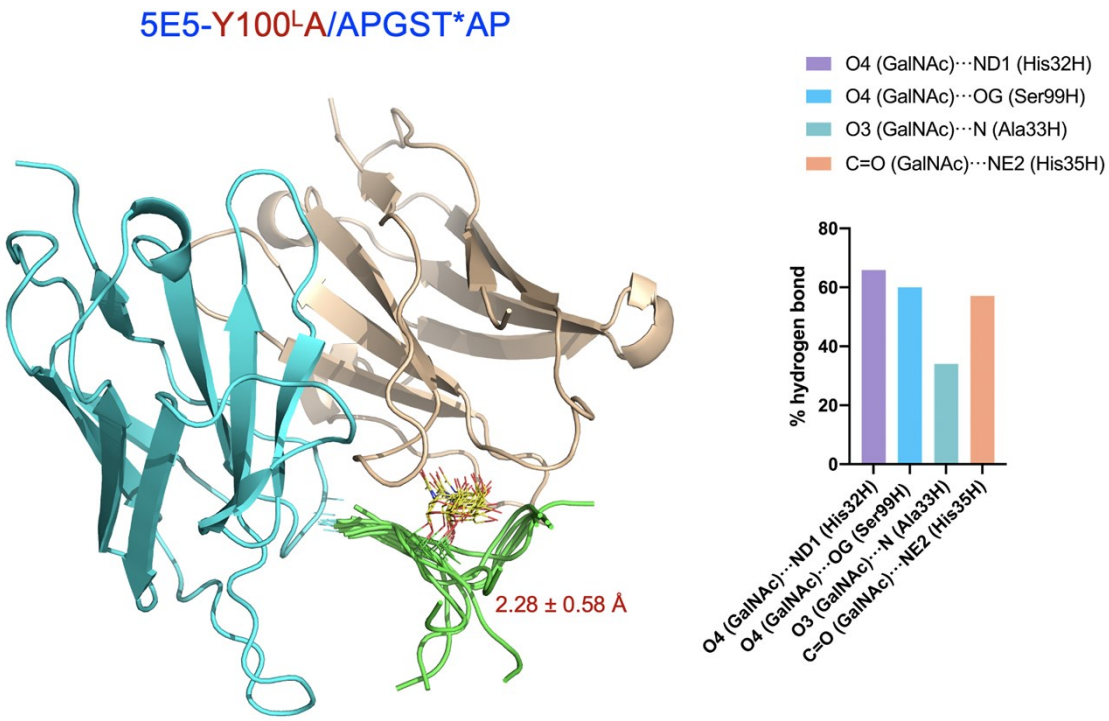



**Figure S7.** Overlay of 10 frames of glycopeptide **1** in complex to mutant scFv-5E5-Y100$^{L}$A sampled from 0.5 μs MD simulations performed in explicit water. The root-mean-square deviation (RMSD) value (± S.D.) of the peptide backbone is shown in red. Only the structure of the first frame of the antibody is shown for clarity. The antibody is shown as light blue (light chain) and brown (heavy chain) cartoon. The peptide backbone of the antigen is shown in green and the carbons of the GalNAc moiety in yellow. Population of the hydrogen bonds found between the GalNAc moiety of glycopeptide **1** is also shown.

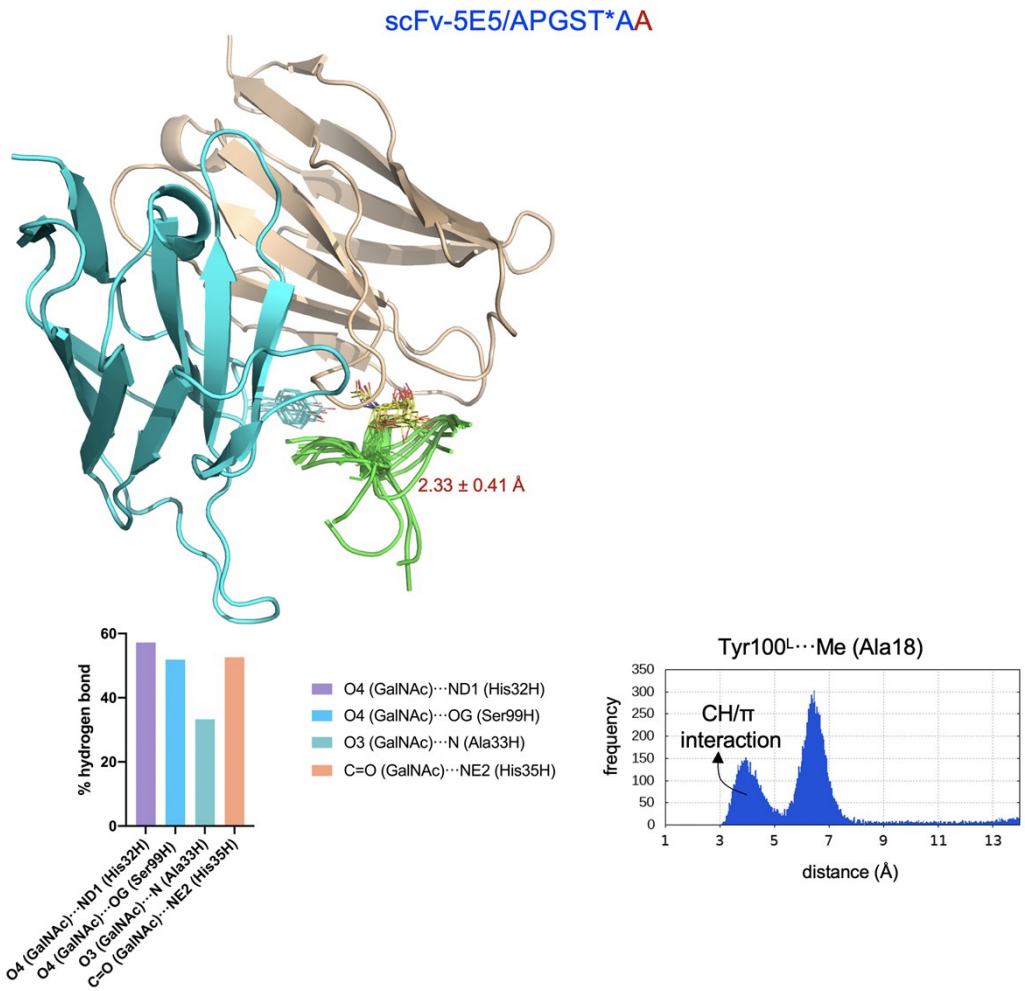



**Figure S8.** Overlay of 10 frames of glycopeptide APGST*AA$\mathbf{A_{18}}$ in complex to scFv-5E5 sampled from 0.5 μs MD simulations performed in explicit water. The root-mean-square deviation (RMSD) value (± S.D.) of the peptide backbone is shown in red. Only the structure of the first frame of the antibody is shown for clarity. The antibody is shown as light blue (light chain) and brown (heavy chain) cartoon. The peptide backbone of the antigen is shown in green and the carbons of the GalNAc moiety in yellow. Population of the hydrogen bonds found between the GalNAc moiety of glycopeptide APGST*AA$\mathbf{A_{18}}$ and the antibody derived from 0.5 μs MD simulations, as well as the distance distribution between the center of the mass of the π-electron system of $Y100^L$ and the methyl group of A18 are also shown.

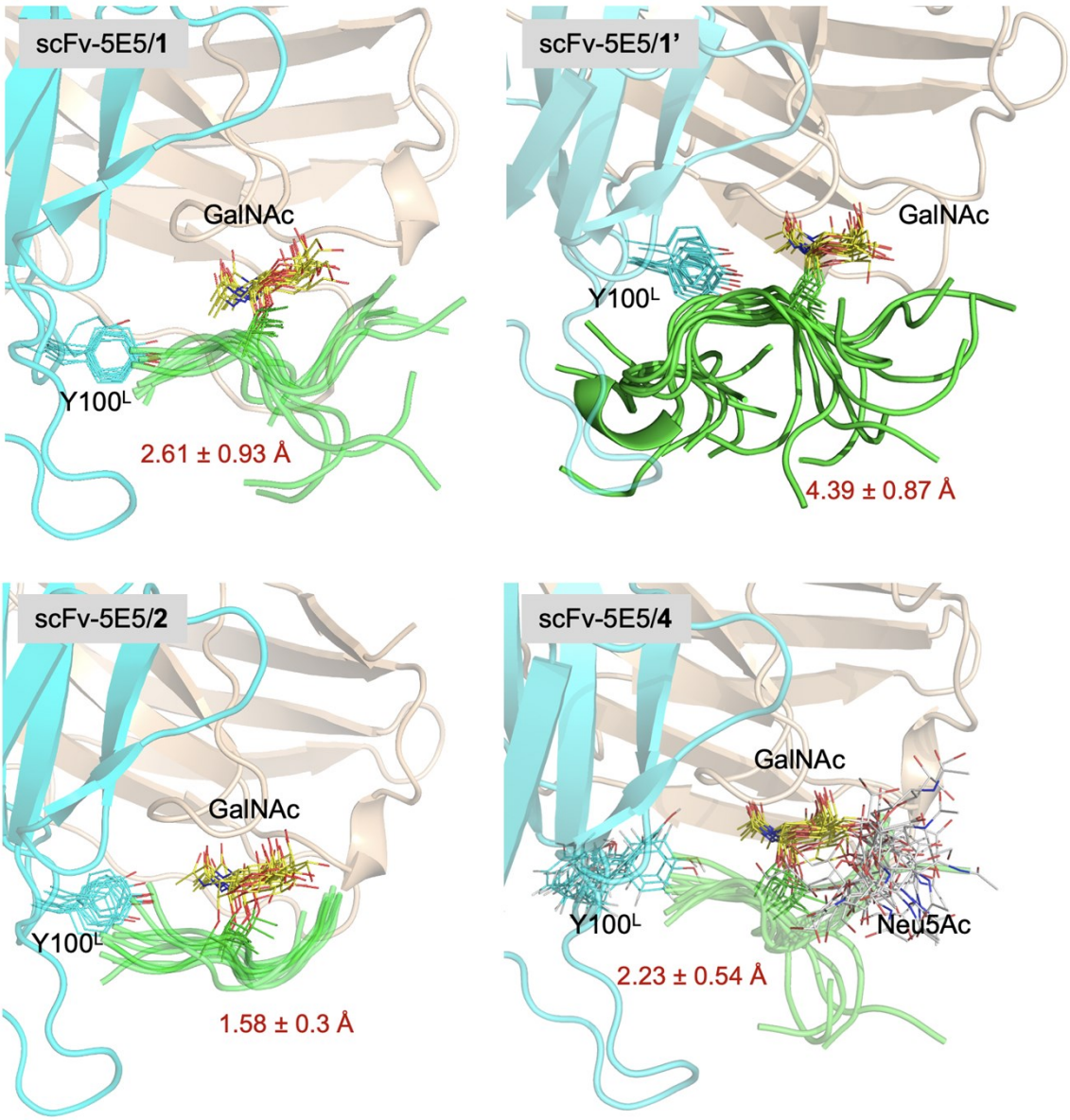



**Figure S9.** Overlay of 10 frames of glycopeptides **1**, **1'**, **2** and **4** in complex to scFv-5E5, sampled from 0.5 μs MD simulations performed in explicit water. Only the structure of the first frame of the antibody is shown for clarity. The antibody is shown as light blue (light chain) and brown (heavy chain) cartoon. The root-mean-square deviation (RMSD) values (± S.D.) of the peptide backbone are shown in red. The peptide backbone of the antigens is shown in green. The carbon atoms of GalNAc and Neu5Ac moieties are in yellow and grey, respectively.

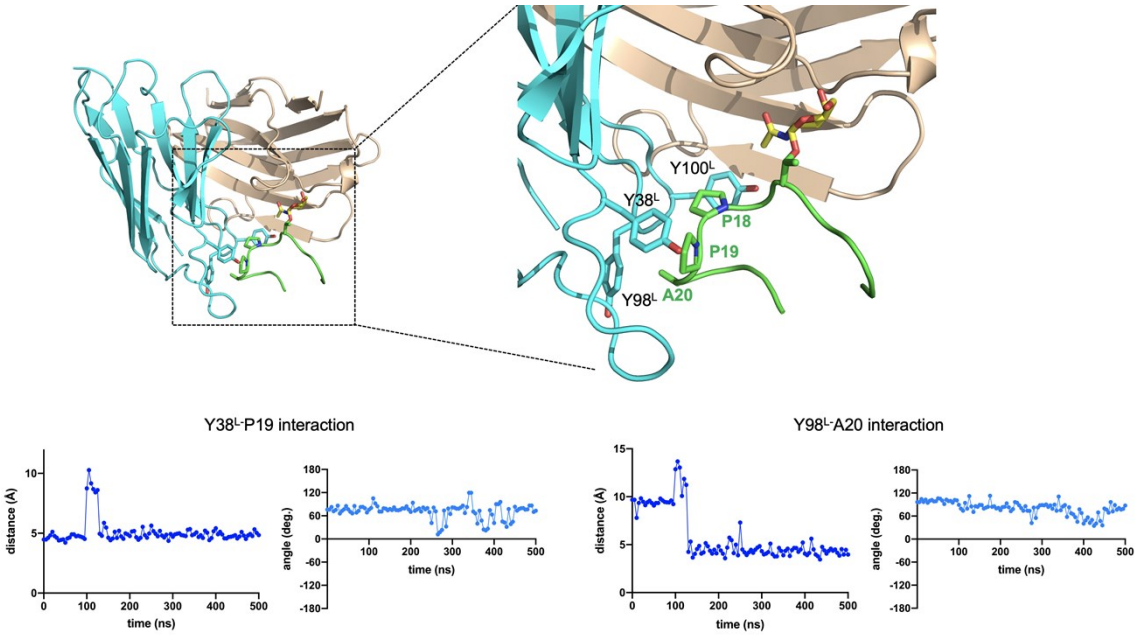



**Figure S10.** Representative snapshot of complex scFv-5E5/**1'** derived from 0.5 μs MD simulations in explicit solvent. The antibody is shown as light blue (light chain) and brown (heavy chain) cartoon. The peptide backbone of the antigen is shown in green and the carbons of the GalNAc moiety in yellow. Distance between the center of the mass of the π-electron system of $Y38^{L}$ and the center of the five-membered ring of Pro19 and distance between the center of the mass of the π-electron system of $Y98^{L}$ and the methyl group of A20 are shown, together with angle between the planes defined by the π-electron system of $Y38^{L}$ and the five-membered ring of this Pro19 and the angle between the planes defined by the π-electron system of $Y98^{L}$ and the methyl group of A20.

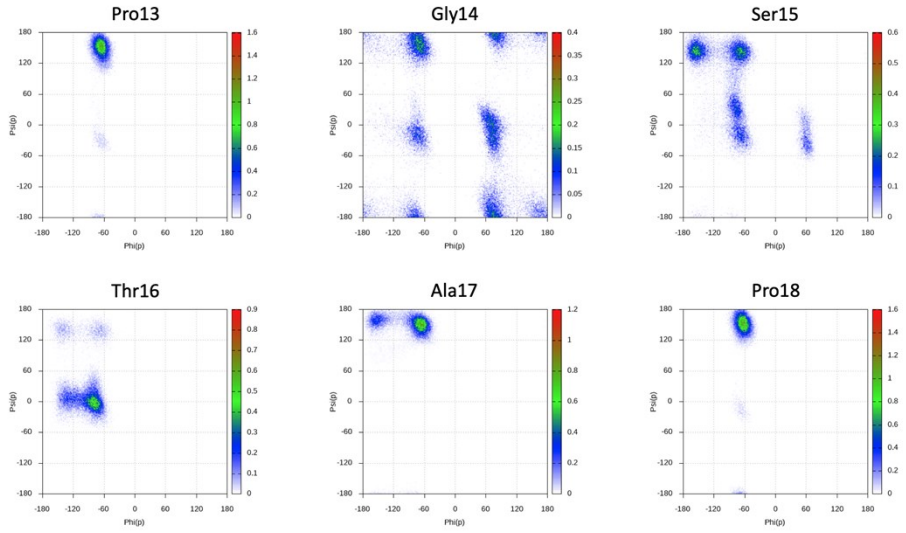



**Figure S11.** ϕ/ψ distributions of the peptide backbone of glycopeptide **1** obtained through 0.5 μs MD simulations of complex scFv-5E5/**1**.

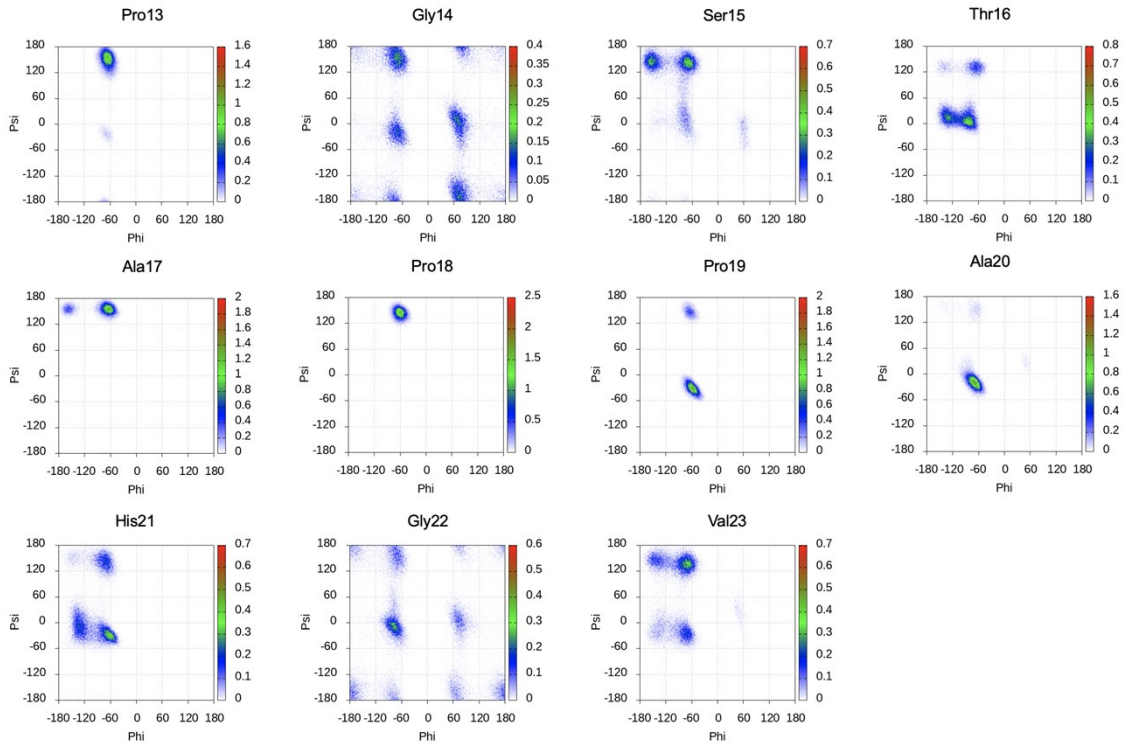



**Figure S12.** ϕ/ψ distributions of the peptide backbone of glycopeptide **1'** obtained through 0.5 μs MD simulations of complex scFv-5E5/**1'**.

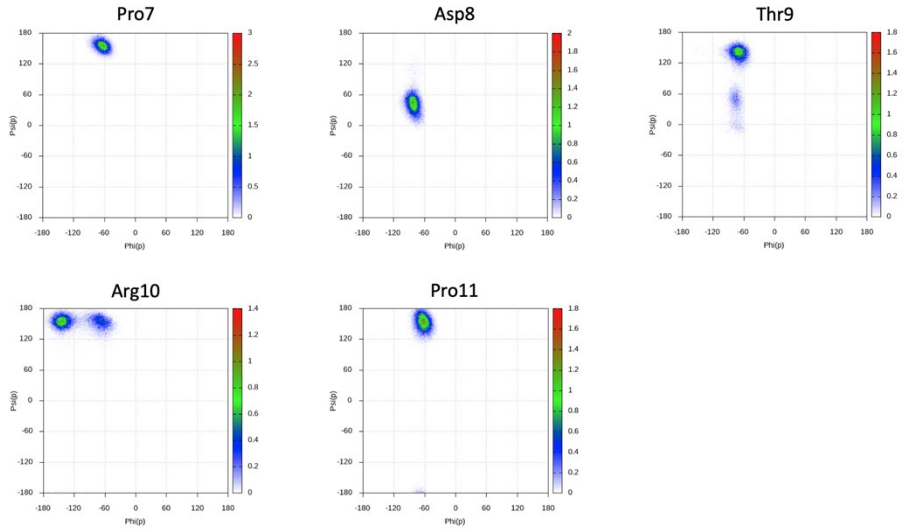



**Figure S13.** ϕ/ψ distributions of the peptide backbone of glycopeptide **2** obtained through 0.5 μs MD simulations of complex scFv-5E5/**2**.

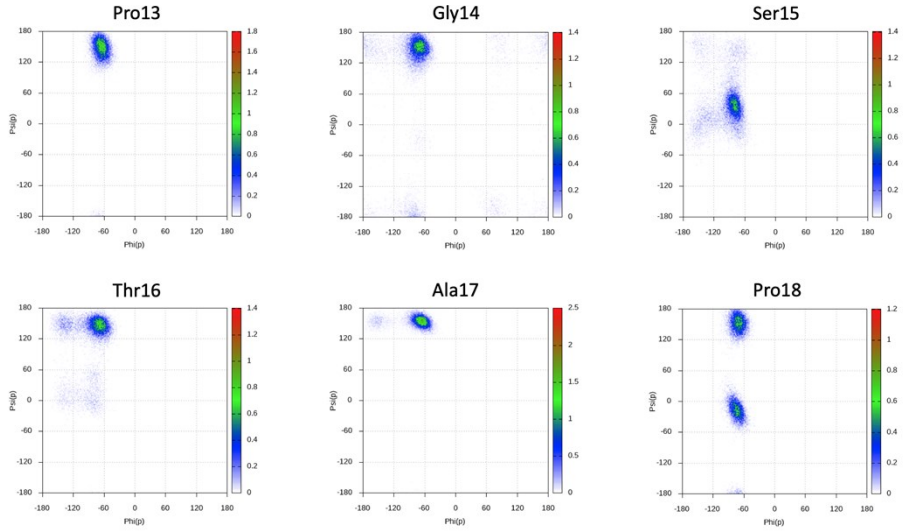



**Figure S14.** ϕ/ψ distributions of the peptide backbone of glycopeptide **3** obtained through 0.5 μs MD simulations of complex scFv-5E5/**3**.

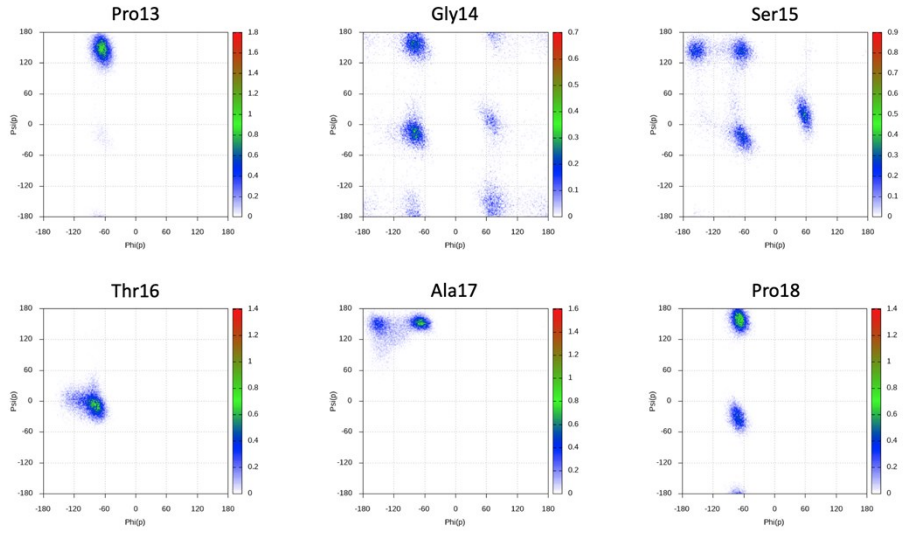



**Figure S15.** ϕ/ψ distributions of the peptide backbone of glycopeptide **4** obtained through 0.5 μs MD simulations of complex scFv-5E5/**4**.

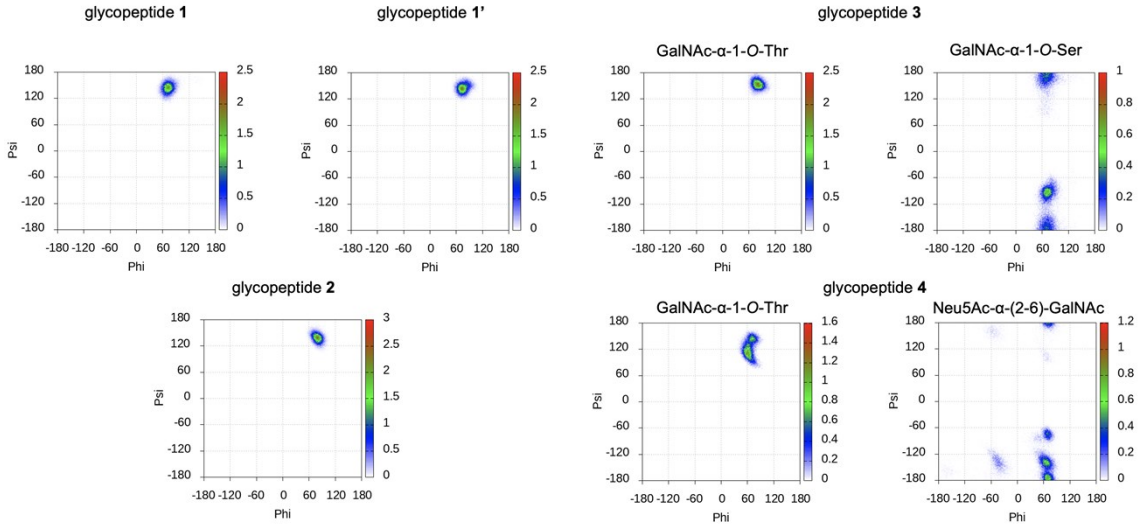



**Figure S16.** ϕ/ψ distributions derived from 0.5 μs MD simulations of the glycosidic linkage of glycopeptides **1-4** and **1'** bound to scFv-5E5. Torsion angle definition: ϕ = O5–C1–O1–Cβ, ψ = C1–O1–Cβ–Cα. In glycopeptide **4**, the torsional angles of the disaccharide Neu5Ac-α-(2-6)-GalNAc are defined as: ϕ = O6–C2–O6–C6, ψ = C2–O6–C6–C5.

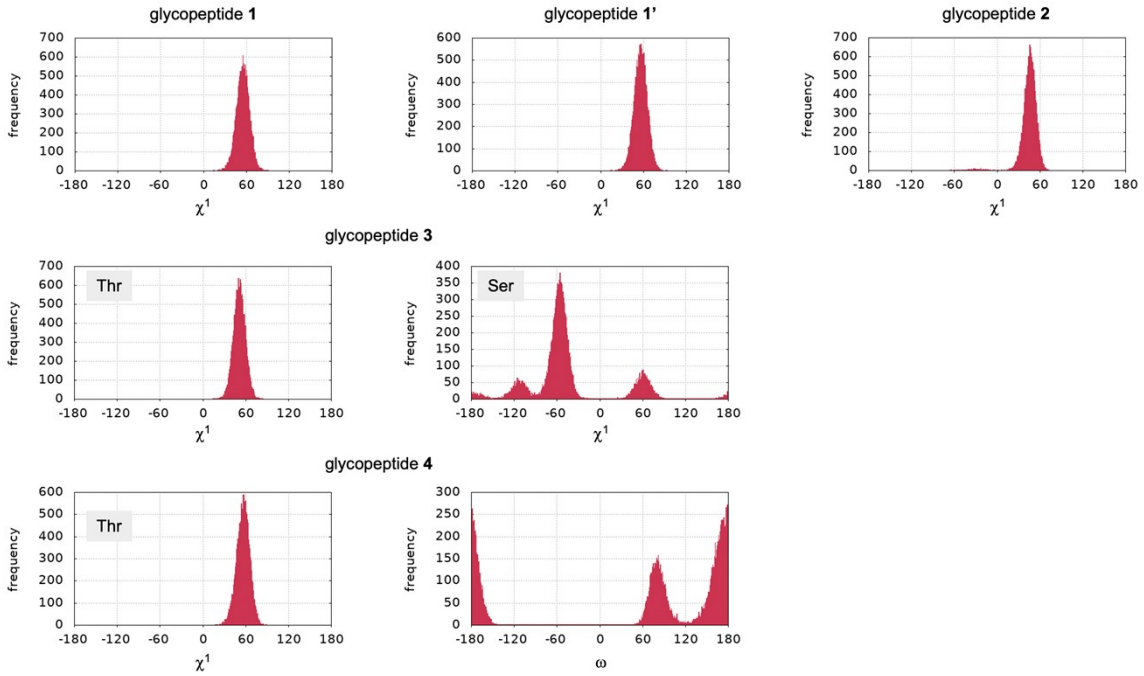



**Figure S17.** Distributions of the sidechain (torsion angle $\chi^1$) for glycopeptides **1-4** and **1'** bound to scFv-5E5 derived from 0.5 μs MD simulations. Torsion angle definition: $\chi^1$= O1–Cβ–Cα–N. ω = O6–C6–C5–O5.

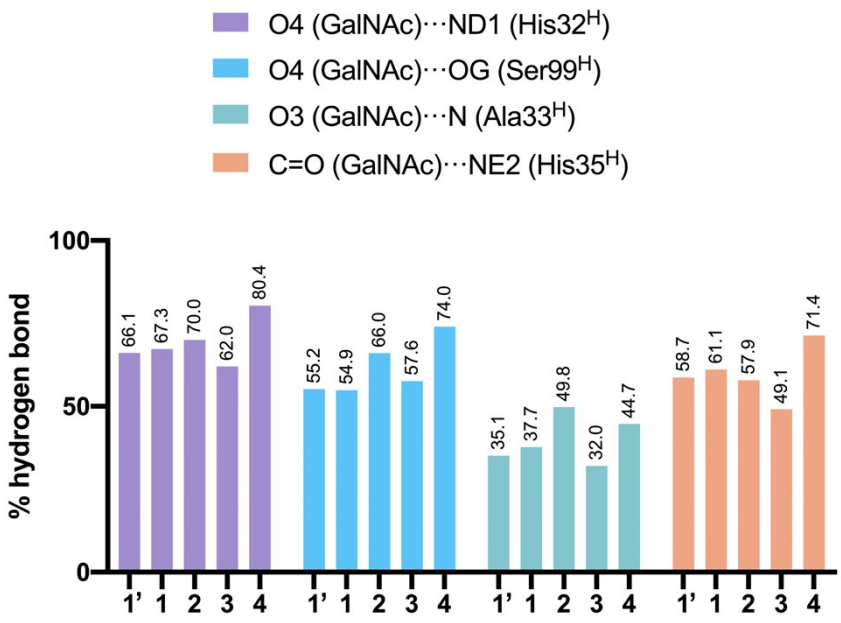



**Figure S18.** Population of hydrogen bonds found between the GalNAc moiety of glycopeptides **1-4** and **1'** and scFv-5E5 derived from 0.5 μs MD simulations. For glycopeptide **3**, only the sugar linked to Thr residue is considered.

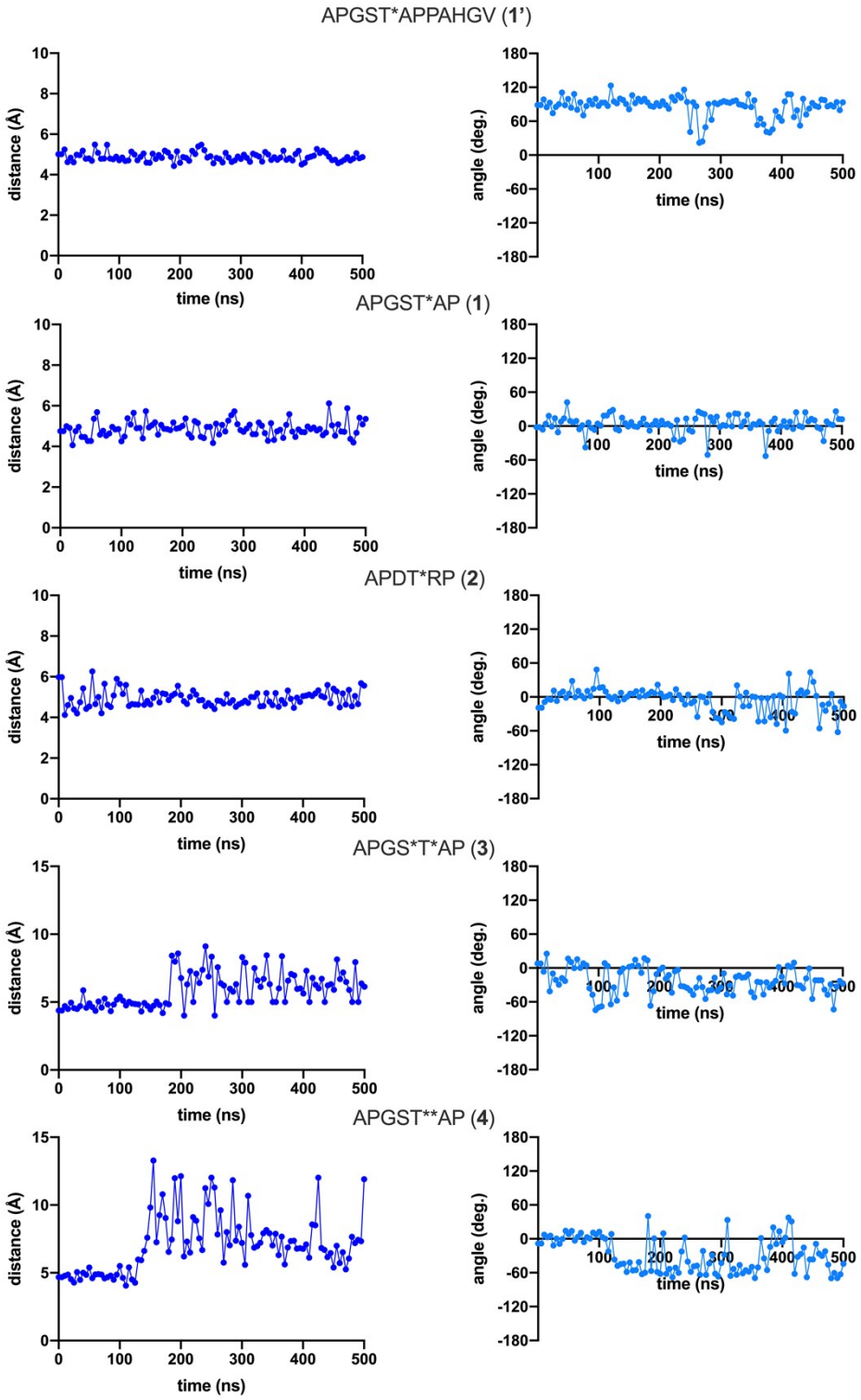



**Figure S19.** Distance between the center of the mass of the π-electron system of Y100 and the center of the five-membered ring of the C-terminal Pro of glycopeptides **1**-**4** and **1'** (left panel), together with the angle between the planes defined by the π-electron system of Y100 and the five-membered ring of this Pro (right panel) derived from 0.5 μs MD simulations.

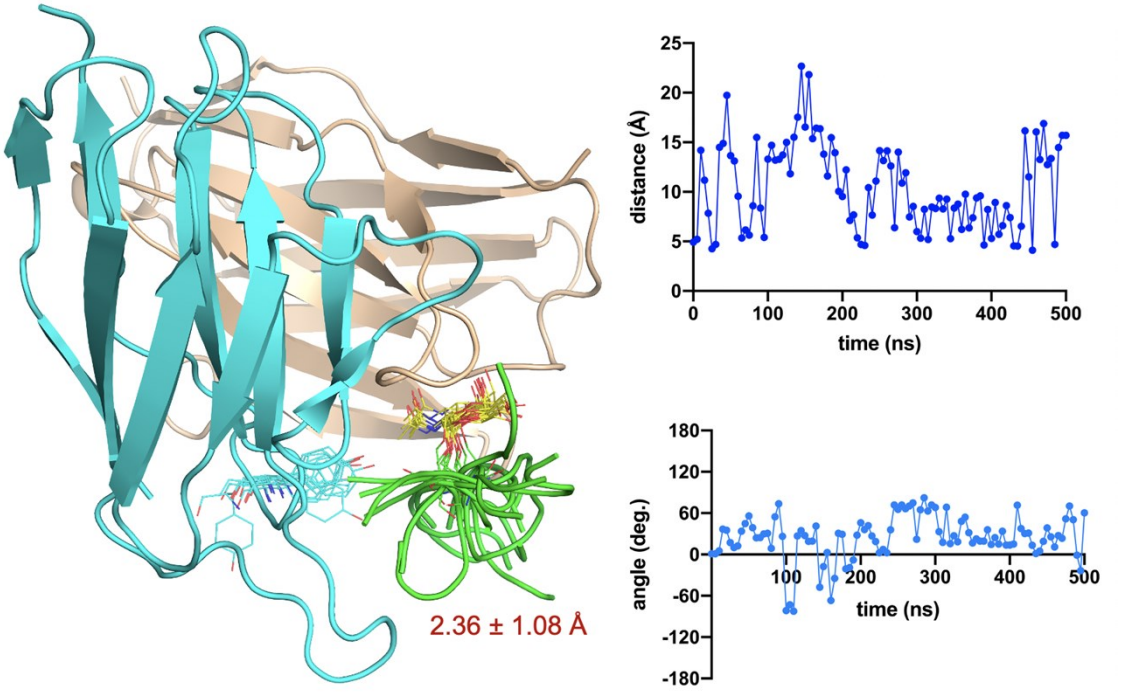



**Figure S20.** Overlay of 10 frames of glycopeptide APGSS*AP (**1-Ser**) in complex to scFv-5E5, sampled from 0.5 μs MD simulations performed in explicit water. The root-mean-square deviation (RMSD) values (± S.D.) of the peptide backbone are shown in red. Only the structure of the first frame of the antibody is shown for clarity. The antibody is shown as light blue (light chain) and brown (heavy chain) cartoon. The peptide backbone of the antigen is shown in green and the carbons of the GalNAc moiety in yellow. Distance between the center of the mass of the π-electron system of $Y100^{L}$ and the center of the five-membered ring of the C-terminal Pro of glycopeptide **1-Ser**, and angle between the planes defined by the π-electron system of $Y100^{L}$ and the five-membered ring of this Pro derived from 0.5 μs MD simulations are also shown.

**$^1$H,$^1$H-TOCSY and 2D-NOESY.** NMR samples of the glycopeptides (400-600 μM) were prepared in 20 mM sodium phosphate buffer (90:10 $H_2O$:$D_2O$) at pH 5.5. NMR spectra were acquired at 2 $^oC$ on a Bruker AMX-600 MHz spectrometer equipped with a cryoprobe and processed with MESTRE software. The TOCSY and NOESY experiments were performed in the phase-sensitive mode with the TPPI method for quadrature detection in F1. Typically, a data matrix of 2000 x 512 points was employed to digitize a spectral width of 6000 Hz. A total of 48 scans was used per increment with a relaxation delay of 1 s. Prior to Fourier transformation, zero filling was performed in F1 to expand the data to 2000 x 2000. Baseline correction was applied in both dimensions. The TOCSY spectra were recorded using MLEV-17 during the 60 ms of the isotropic mixing period. The NOESY experiments were performed with mixing times of 100, 200, and 300 ms. Water suppression was achieved by including a WATERGATE module in the pulse sequence prior to acquisition.

**MD simulations with time-averaged restraints on MUC1-Tn16.**[16] The protocol employed above for conventional MD simulations was used also in these calculations. For the final production trajectory, the ROESY derived distances were imposed as time-averaged constraint, applying an $r^{-6}$ averaging. The equilibrium distance range was set to $r_{exp} - 0.2$ Å $\leq r_{exp} \leq 0.2$ Å. Trajectories were run at 298 K, with a decay constant of 20 ns and a time step of 1 fs. The force constants *rk2* and *rk3* used in each case were 10 kcal·mol$^{-1}$·Å$^{-2}$. The overall simulation length was 200 ns. The coordinates were saved each 1 ps, obtaining MD trajectories of 200000 frames each.

The distance ranges for NOE constraints were 1.8-2.3, 1.8-3.0, 1.8-3.5, 1.8-4.0 Å for strong, medium, weak, and very weak respectively (see Table S4).

**Table S3.** $^{1}$H-NMR (600 MHz, $H_2O/D_2O$, 9:1, pH =5.5) assignments of glycopeptide **MUC1-Tn16** at 275K.

| Residue | Proton | ppm | Residue | Proton | ppm | Residue | Proton | ppm |
|---|---|---|---|---|---|---|---|---|
| **His1** | H | 8.30 | **Pro11** | Hδ2 | 3.25 | **Pro19** | Hδ2 | 3.32 |
| | Hβ2 | 2.80 | | Hβs | 1.91 | | Hβs | 1.96 |
| | Hβ3 | 2.95 | | Hγ3 | 1.64 | | Hγ3 | 1.69 |
| | Hα | 4.34 | | Hγ2 | 1.49 | | Hγ2 | 1.55 |
| **Gly2** | Hα2 | 3.63 | | Hδ3 | 3.48 | | Hδ3 | 3.46 |
| | H | 8.37 | | Hα | 4.05 | | Hα | 4.04 |
| | Hα3 | 3.58 | **Ala12** | H | 8.34 | **Ala20** | H | 8.23 |
| **Val3** | H | 7.93 | | Hα | 4.21 | | Hβs | 1.04 |
| | Hα | 3.87 | | Hβs | 1.02 | | Hα | 3.88 |
| | Hβ | 1.75 | **Pro13** | Hδ2 | 3.31 | | | |
| | Hγs | 0.59 | | Hβs | 1.97 | | | |
| **Thr4** | H | 8.18 | | Hγ2 | 1.69 | | | |
| | Hγs | 0.86 | | Hγ3 | 1.59 | | | |
| | Hα | 4.07 | | Hδ3 | 3.49 | | | |
| | Hβ | 3.87 | | Hα | 4.07 | | | |
| **Ser5** | H | 8.14 | **Gly14** | Hα2 | 3.66 | | | |
| | Hα | 4.10 | | H | 8.37 | | | |
| | Hβs | 3.49 | | Hα3 | 3.59 | | | |
| **Ala6** | H | 8.21 | **Ser15** | H | 7.92 | | | |
| | Hβs | 1.02 | | Hα | 4.29 | | | |
| | Hα | 4.26 | | Hβs | 3.58 | | | |
| **Pro7** | Hδ2 | 3.32 | **Thr16** | H | 8.51 | | | |
| | Hβs | 1.96 | | Hα | 4.24 | | | |
| | Hγ3 | 1.69 | | HB | 3.99 | | | |
| | Hγ2 | 1.57 | | Hγs | 0.91 | | | |
| | Hδ3 | 3.49 | **GalNAc** | HN2 | 7.58 | | | |
| | Hα | 4.05 | | H1 | 4.57 | | | |
| **Asp8** | H | 8.39 | | H8 | 1.68 | | | |
| | Hβ3 | 2.45 | | H2 | 3.73 | | | |
| | Hβ2 | 2.53 | | H3 | 3.53 | | | |
| | Hα | 4.34 | | H4 | 3.60 | | | |
| **Thr9** | H | 7.90 | | H62 | 3.39 | | | |
| | Hγs | 0.83 | | H5 | 3.66 | | | |
| | Hβ | 3.89 | | H61 | 3.42 | | | |
| | Hα | 3.96 | **Ala17** | H | 8.16 | | | |
| **Arg10** | H | 8.07 | | Hβs | 0.99 | | | |
| | Hα | 4.27 | | Hα | 4.20 | | | |
| | Hδs | 2.86 | **Pro18** | Hβs | 2.01 | | | |
| | Hγs | 1.35 | | Hδ3 | 3.44 | | | |
| | Hβ3 | 1.40 | | Hδ2 | 3.30 | | | |
| | Hβ2 | 1.49 | | Hα | 4.05 | | | |
| | | | | Hγ3 | 1.70 | | | |
| | | | | Hγ2 | 1.566 | | | |

**Table S4.** NOE interactions in glycopeptide **MUC1-Tn16** and used as experimental restraints in MD simulations with time averaged restraints (MD-tar).

| # | **Proton A** | **Proton B** | **Intensity** |
|---|---|---|---|
| 1 | NH-G2 | Hβ-V3 | weak |
| 2 | NH-G2 | Hα-H1 | strong |
| 3 | Hβ-V3 | Hα-T4 | medium |
| 4 | NH-V3 | Hα-G2 | strong |
| 5 | NH-T4 | Hγ-V3 | medium |
| 6 | NH-T4 | NH-V3 | medium |
| 7 | NH-T4 | Hβ-V3 | medium |
| 8 | NH-T4 | Hα-V3 | strong |
| 9 | NH-S5 | Hβ-T4 | strong |
| 10 | NH-S5 | Hα-T4 | strong |
| 11 | NH-S5 | Hγ-V3 | weak |
| 12 | NH-A6 | Hα-S5 | strong |
| 13 | NH-A6 | Hβ-S5 | strong |
| 14 | Hβ-A6 | Hδ2-P7 | medium |
| 15 | Hβ-A6 | Hα-A5 | medium |
| 16 | NH-D8 | Hα-P7 | strong |
| 17 | NH-D8 | Hδ2-P7 | medium |
| 18 | NH-D8 | Hβ-A6 | medium |
| 19 | NH-D8 | NH-T9 | strong |
| 20 | NH-T9 | Hα-D8 | strong |
| 21 | NH-T9 | NH-R10 | strong |
| 22 | NH-R10 | Hα-D8 | medium |
| 23 | NH-R10 | Hα-P11 | weak |
| 24 | NH-R10 | Hα-T9 | strong |
| 25 | NH-R10 | Hβ-T9 | strong |
| 26 | NH-R10 | Hδ3-P11 | medium |
| 27 | NH-R10 | Hδ2-P11 | medium |
| 28 | Hδ3-P11 | Hα-R10 | strong |
| 29 | Hα-P11 | Hβ-A12 | medium |
| 30 | NH-A12 | Hγ2-P11 | strong |
| 31 | NH-A12 | Hβ-P11 | strong |
| 32 | NH-A12 | Hδ2-P11 | medium |
| 33 | NH-A12 | Hδ3-P11 | weak |
| 34 | NH-A12 | Hα-P11 | strong |
| 35 | Hβ-A12 | Hα-P11 | medium |
| 36 | Hβ-A12 | Hα-G14 | weak |
| 37 | Hδ2-P13 | Hα-A12 | strong |
| 38 | NH-G14 | Hα-A12 | weak |
| 39 | NH-G14 | Hβ-P13 | medium |
| 40 | NH-G14 | Hα-P13 | strong |
| 41 | NH-G14 | Hδ3-P13 | weak |
| 42 | NH-S15 | NH-G14 | strong |
| 43 | NH-S15 | Hα-G14 | strong |
| 44 | NH-S15 | Hβ-A12 | weak |
| 45 | NH-T16 | Hβ-S15 | strong |
| **#** | **Proton A** | **Proton B** | **Intensity** |
| 46 | NH-T16 | Hα-S15 | strong |
| 47 | NH-T16 | H3-GalNAc- | weak |
| 48 | NH-T16 | H5-GalNAc-T16 | weak |
| 49 | NH-T16 | H1-GalNAc- | medium |
| 50 | NH-T16 | NH-GalNAc- | strong |
| 51 | NH-T16 | NH-A17 | strong |
| 52 | NH-T16 | NH-S15 | strong |
| 53 | NH-GalNAc | Hα-S15 | weak |
| 54 | NH-GalNAc | Hβ-T16 | very weak |
| 55 | NH-GalNAc | H2-GalNAc | strong |
| 56 | NH-GalNAc | H3-GalNAc | strong |
| 57 | NH-GalNAc | H4-GalNAc | very weak |
| 58 | NH-GalNAc | H5-GalNAc | very weak |
| 59 | NH-GalNAc | H8-GalNAc | strong |
| 60 | NH-GalNAc | Hγ-T16 | very weak |
| 61 | Hβ-A17 | Hδ3-P18 | medium |
| 62 | NH-A17 | Hβ -T16 | strong |
| 63 | NH-A17 | H1-GalNAc | medium |
| 64 | NH-A17 | Hα -T16 | strong |
| 65 | NH-A20 | Hα-P19 | Strong |
| 66 | NH-A20 | Hβ-P19 | medium |
| 67 | NH-A20 | Hγ3-P19 | weak |
| 68 | NH-A20 | Hγ2-P19 | medium |

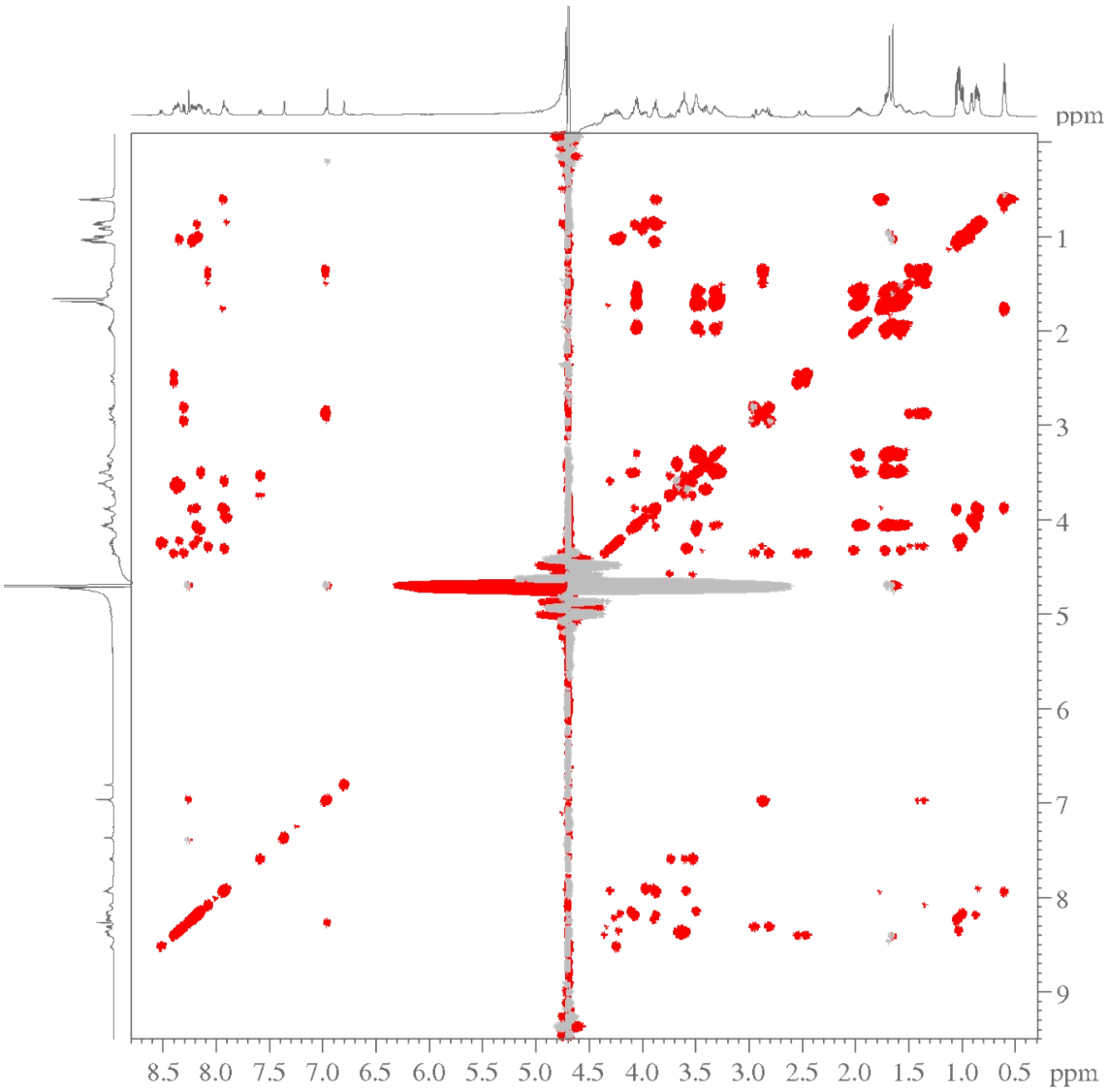



**Figure S21.** $^{1}H,^{1}H$-TOCSY spectrum (600 MHz, $H_2O/D_2O$, 9:1, pH =5.5) for free **MUC1-Tn16**.

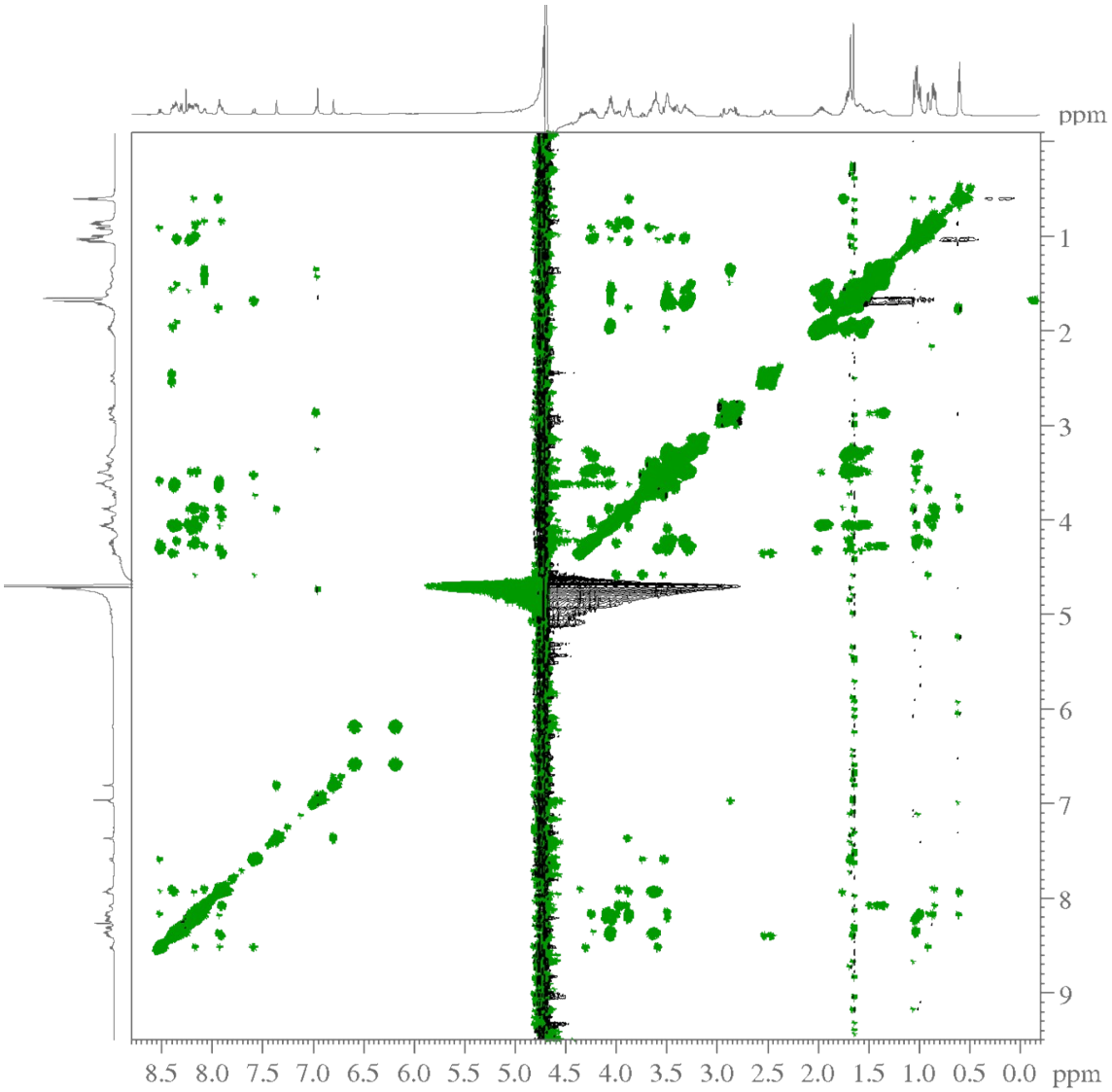



**Figure S22.** 2D-NOESY spectrum (600 MHz, $H_2O/D_2O$, 9:1, pH =5.5) for free **MUC1-Tn16**.

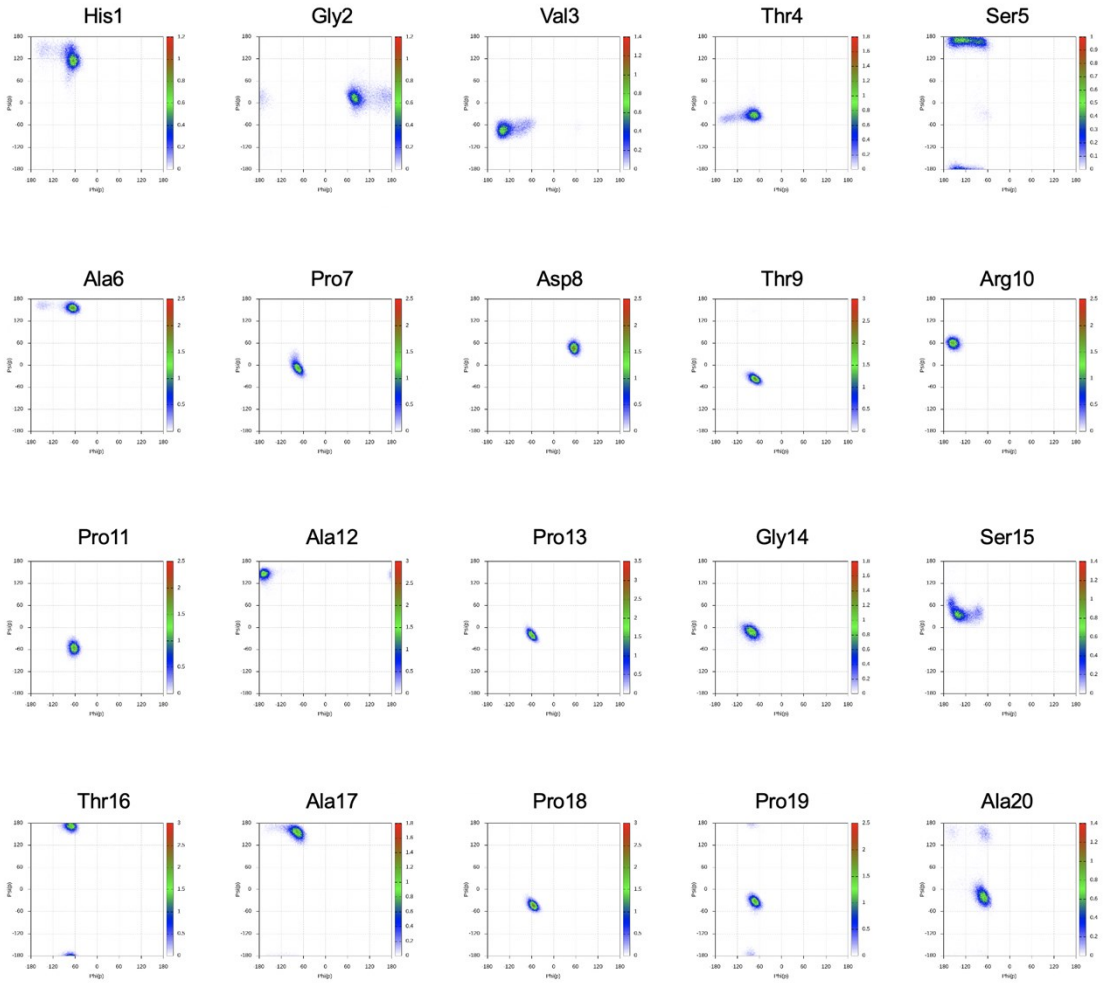



**Figure S23.** ϕ/ψ distributions of the peptide backbone of glycopeptide **MUC1-Tn16** obtained through 0.2 μs MD-tar simulations in solution**.**

## References

1. Plattner, C.; Höfener, M.; Sewald, N. One-Pot Azidochlorination of Glycals. *Org. Lett.* **2011**, *13*, 545–547.

2. Liebe, B.; Kunz, H. Synthesis of sialyl-Tn antigen. Regioselective sialylation of a galactosamine threonine conjugate unblocked in the carbohydrate portion. *Tetrahedron Lett.* **1994**, *35*, 8777–8778.

3. Madariaga, D.; Martínez-Sáez, N.; Somovilla, V. J.; Coelho, H.; Valero-González, J.; Castro-López, J.; Asensio, J. L.; Jiménez-Barbero, J.; Busto, J. H.; Avenoza, A.; Marcelo, F.; Hurtado-Guerrero, R.; Peregrina, J. M.; Corzana, F. Detection of Tumor-Associated Glycopeptides by Lectins: The Peptide Context Modulates Carbohydrate Recognition. *ACS Chem. Biol.* **2015**, *10*, 3, 747–756.

4. Martínez-Sáez, N.; Castro-López, J.; Valero-Gónzalez, J.; Madariaga, D.; Compañón, I.; Somovilla, V. J.; Salvadó, M.; Asensio, J. L.; Jiménez-Barbero, J.; Avenoza, A.; Busto, J. H.; Bernardes, G. J. L.; Peregrina, J. M.; Hurtado-Guerrero, R., Corzana, F. Deciphering the Non-Equivalence of Serine and Threonine O-Glycosylation Points: Implications for Molecular Recognition of the Tn Antigen by an anti-MUC1 Antibody. *Angew. Chem. Int. Ed.* **2015**, *54*, 9830–9834.

5. Dziadek, S.; Griesinger, C.; Kunz H., Reinscheid, U. M. Synthesis and Structural Model of an α(2,6)‐Sialyl‐T Glycosylated MUC1 Eicosapeptide under Physiological Conditions. *Chem. Eur. J.* **2006**, *12*, 4981–4993.

6. Kabsch, W. XDS. *Acta crystallogr. Section D Biol. Crystallogr.* **2010**, *66*, 125−132.

7. Winn, M.D.; Ballard, C.C.: Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.; McCoy, A.; McNicholas, S.J.; Murshudov, G.N.; Pannu, N.S.; Potterton, E.A.; Powell, H.R.; Read, R.J.; Vagin, A.; Wilson, K.S. Overview of the CCP4 suite and current developments. *Acta crystallogr. Section D Biol. Crystallogr.* **2011**, *67*, 235−242.

8. The CCP4 suite: programs for protein crystallography. *Acta crystallogr. Section D Biol. Crystallogr.* **1994**, *50,* 760−763.

9. Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126−2132.

10. Murshudov, G.N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355−367.

11. Case, D.A.; Brozell, S.R.; Cerutti, D.S.; Cheatham, III, T.E.; Cruzeiro, V.W.D.; Darden, T.A.; Duke, R.E.; Ghoreishi, D.; Gohlke, H.; Goetz, A.W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T.S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.J.; Merz, K.M.; Miao, Y.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.L.; Smith, J.; Swails, J.; Walker, R.C.; Wang, J.; Wei, H.; Wolf, R.M.; Wu, X.; Xiao, L.; York D.M.; Kollman P.A. (2018), AMBER 2018, University of California, San Francisco.

12. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.

13. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622−655.

14. Kiyohara, K.; Gubbins, K.; Panagiotopoulos, A. Phase coexistence properties of polarizable water models. *Mol. Phys.* **1998**, *94*, 803−808.

15. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
16. Compañón, I.; Guerreiro, A.; Mangini, V.; Castro-López, J.; Escudero-Casao, M.; Avenoza, A.; Busto, J. H.; Castillón, S.; Jiménez-Barbero, J.; Asensio, J. L.; Jiménez-Osés, G.; Boutureira, O.; Peregrina, J. M.; Hurtado-Guerrero, R.; Fiammengo, R.; Bernardes, G. J. L.; Corzana, F. Structure-based design of potent tumor-associated antigens: modulation of peptide presentation by single atom O/S or O/Se substitutions at the glycosidic linkage. *J. Am. Chem. Soc.* **2019**, *141*, 4063−4072.