Supporting Information

Detection of Tumor-Associated Glycopeptides by Lectins: The Peptide Context Modulates Carbohydrate Recognition

David Madariaga, Nuria Martínez-Sáez, Víctor J. Somovilla, Helena Coelho, Jessika Valero-González, Jorge Castro-López, Juan L. Asensio, Jesús Jiménez-Barbero, Jesús H. Busto, Alberto Avenoza, Filipa Marcelo, Ramón Hurtado-Guerrero, Francisco Corzana and Jesús M. Peregrina

GENERAL METHODS

2D NMR experiments. NMR experiments were performed on a 400 MHz spectrometer at 298 K. Magnitude-mode ge-2D COSY spectra were acquired with gradients by using the cosygpqf pulse program with a pulse width of 90º. Phase-sensitive ge-2D HSQC spectra were acquired by using z-filter and selection before t1 removing the decoupling during acquisition by use of the invigpndph pulse program with CNST2 (*J*HC)=145. 2D-NOESY experiments were recorded at 293 K and pH 6.0-6.5 in H_2O/D_2O (9:1). All The experiments were conducted by using phase-sensitive ge-2D NOESY with WATERGATE for H_2O/D_2O (9:1) spectra. NOEs intensities were normalized with respect to the diagonal peak at zero mixing time. Distances involving NH protons were semi-quantitatively determined by integrating the volume of the corresponding cross-peaks. The number of scans used was 16 and the mixing time was 500 ms.

Saturation-Transfer Difference (STD) NMR experiments. NMR experiments were recorded on a 600 MHz spectrometer equipped with a triple channel cryoprobe head. The molar ratio was adjusted between 4:1 to 30:1 depending on ligand:protein complex. STD-NMR experiments were recorded at 298K or 310K and with 2 s irradiation time using pulse sequences with or without water suppression. An excitation sculpting with gradients was employed to suppress the water proton signals. When necessary a spin lock filter (T1ρ) with a 4 kHz field and a length of 20 ms was applied to suppress protein background. STD NMR spectrum was acquired with 2048 transients in a matrix with 16K data points in t2 in a spectral window of 12019.23 Hz centered at 2825.27 Hz. Selective saturation of the protein resonances (on resonance spectrum) was performed by irradiating at −1.0 ppm using a series of Eburp2.1000-shaped 90° pulses (50 ms, 1 ms delay between pulses) for a total saturation time of 2.0 s. For the reference spectrum (off resonance), the samples were irradiated at 100 ppm. Proper control experiments were performed with the ligands in the presence and absence of SBA in order to optimize the frequency for protein saturation and to ensure that the ligand signals were not affected. Glycopeptide **1** showed residual STD intensities in the STD spectra that were taken into account (subtracted) when analyzing the STD spectra. In all cases, to accomplish the epitope mapping of each ligand, the STD intensities were normalized with respect to the highest response. Unfortunately, the signal of the anomeric proton as well as those corresponding to Hα protons of the amino acids of glycopeptides could not be analyzed in the STD spectra that were run applying water suppression pulse sequences, due to their close distance to the HDO signal.

Glycopeptide synthesis. All compounds were synthesized by stepwise solid-phase peptide synthesis using the Fmoc strategy on Rink Amide AM resin (0.1 mmol). Tn-threonine building block was obtained as described in the literature.[S1] This glycosylated amino acid building block was coupled manually using 2.0 equiv. and activated with HBTU, while the other Fmoc amino acids (10 equiv.) were coupled automatically on a peptide synthesizer using HBTU. The peptides were then released from the resin, and all acid sensitive side-chain protecting groups were simultaneously removed using TFA 95%, TIPS 2.5%, H_2O 2.5%, followed by precipitation with diethyl ether. The *O*-acetyl groups of $(AcO)_3$GalNAc moiety were deprotected in a mixture of NH_2NH_2/MeOH (7:3). Finally, all compounds were purified by HPLC.

Unrestrained Molecular Dynamics simulations. All molecular dynamics simulations were carried out on the Finis-Terrae cluster of the *Centro de Supercomputación de Galicia* (CESGA), Spain. Starting geometries for complexes were generated from the available data deposited in the Protein Data Bank (pdb code: 1SBF) and modified accordingly. Proton were added automatically using the Xleap module of AMBER 11.0 package.[S2] Each model complex was immersed in a 10 Å-sided cube with pre-equilibrated TIP3P water molecules. To equilibrate the system, we followed a protocol consisting of 10 steps. Firstly, only the water molecules are minimized, and then heated to 300 K. The water box, together with Na^+ Cl^- ions to neutralize the system, was then minimized, followed by a short MD simulation. At this point, the system was minimized in the four following steps with positional restraints imposed on the solute, decreasing the force constant step by step from 20 to 5 kcal mol^{-1}. Finally, a non-restraint minimization was performed. The production dynamics simulations were accomplished at a constant temperature of 300 K (by applying the Berendsen coupling algorithm for the temperature scaling) and constant pressure (1 bar). Particle Mesh Ewald Method, to introduce long-range electrostatic effects, and periodic boundary conditions were also used. SHAKE algorithm for hydrogen atoms, which allows using a 2 fs time step, was also employed. Finally, a 9Å cutoff was applied for the Lennard-Jones interactions. MD simulations were performed with the sander module of AMBER 11.0 package (parm99 force field),[S3] which was implemented with GLYCAM 06 parameters[S4] and phi and psi ff03 modifications.[S5] A simulation length of 100 ns and the trajectory coordinates were saved each 0.5 ps.

MD simulations with time-averaged restraints (MD-tar). MD-tar simulations were performed with AMBER 11 (parm99 force field), which was implemented with GLYCAM 06 and ff03 parameters. Distances derived from NOE cross-peaks were included as time-averaged distance restraints. A $\langle r^{-6} \rangle^{-1/6}$ average was used for the distances. Final trajectories were run using an exponential decay constant of 2000 ps and a simulation length of 20 ns in explicit TIP3P water molecules.

General procedure for the enzyme-linked lectin assay (ELLA). ELISA plate (*Amine-binding, Maleic anhydride 96-well plate*) was coated with 100 µL/well of a solution of the corresponding glycopeptide (0–150 nmol/well) in carbonate/bicarbonate buffer (0.2 M, pH 9.4) and incubated overnight at 25 ºC. Unbound sites were then blocked by adding 200 µL/well of blocking buffer. After 1 h at 25 ºC, the blocking buffer was removed and the plate wells were washed 3 × 200 µL/well with PBST (Phosphate-buffered saline –0.1 M sodium phosphate, 0.15 M sodium chloride, pH 7.2–, containing 0.05% Tween-20 detergent). As the next step, the wells were incubated with biotin-conjugated soybean lectin available from *EY laboratories* (100 µL, diluted 1/150 in PBST buffer) for 2 h. After washing with PBST (3 × 200 µL/well, 2 min/well), the wells were treated with conjugated streptavidin Horseradish Peroxidase (HRP) (100 µL, diluted 1/4000 in PBST buffer) for 1 h at 25 ºC. The wells were again washed first with PBST (3 × 200 µL/well, 2 min/well) and then with 350 µL of water. TMB was added (90 µL/well) and after incubation for 10 min, the reaction was terminated with the addition of 50 µL/well of stop solution (1 M H_2SO_4). Absorbance detection of the wells was immediately performed at 450 nm using an ELISA plate reader. Average absorbance intensities of three replicates were plotted against glycopeptides concentration.

Isothermal Titration Calorimetry (ITC). Binding studies were performed at 25 ºC in 10 mM phosphate at pH 7.2 by using a titration calorimeter with a reaction cell volume of 300 µL. Typically, SBA lectin solutions (40–120 µM, monomer concentration) were titrated in the reaction cell with solutions (2000–3000 µM) of **1**, **2** or **3** compounds contained in a 300 µL syringe. In all cases, buffer composition in the syringe and titration cell was identical. At least 30 consecutive injections of 5–10 µL were applied at 5 min intervals while the lectin solution was stirred at a constant speed of 300 rpm. Dilution heats of the ligands (**1**, **2** or **3**) were independently measured and subtracted from the heats of binding. Experimental curves were analyzed by using Origin, which was provided with the instrument. In all cases, thermodynamic parameters were derived from at least two independent experiments and then averaged.

Crystallization. Soybean agglutinin was dissolved in buffer A (25 mM TRIS-HCl, 150 mM NaCl pH 7.5). It was further purified by gel filtration chromatography in order to remove aggregates. Then, it was dialyzed in a 25 mM TRIS-HCl and 150 mM NaCl buffer (pH 7.5) without salt. The concentration of the protein was measured at 280 nm of absorbance using the theoretical extinction coefficient (ε) of 40450 mL mg^{-1} cm^{-1}. Crystals were grown by hanging drop diffusion at 18 ºC. The drops were prepared by mixing 1 µL of protein complex containing 8.5 mg mL^{-1} of soybean agglutinin and 10 mM glycopeptide **2** with 1 µL of solutions with 20-30% PEG 3350, 0.2 M NaBr and 0.5 µL of 20–30% PEG 400 as additive. The own mother solution was acted as cryoprotectant. The data were processed and scaled using the XDS package[S6] and CCP4[S7] software, relevant statistics are given in Table 4. The crystal structure was solved by molecular replacement using PHENIX[S8] and using the PDB entry 1SBF as the template. Initial phases were further improved by cycles of manual model building in Coot[S9] and refinement with REFMAC5.[S10] The final models were validated with PROCHECK,[S11] model statistics are given in Table S4 in Supporting Information. The asymmetric units of these crystals show 12 molecules of SBA. Coordinates and structure factors have been deposited in the Worldwide Protein Data Bank (wwPDB, code: 4d69).

Reagents and general procedures. Commercial reagents were used without further purification. Solvents were dried and redistilled prior to use in the usual way. All reactions were performed in oven-dried glassware with magnetic stirring under an inert atmosphere unless noted otherwise. Analytical thin layer chromatography (TLC) was performed on glass plates precoated with a 0.25 mm thickness of silica gel. The TLC plates were visualized with UV light and by staining with Hanessian solution (ceric sulfate and ammonium molybdate in aqueous sulfuric acid) or sulfuric acid-ethanol solution. Column chromatography was performed on silicagel (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). Spectra were assigned using COSY and HSQC. All NMR chemical shifts (δ) were recorded in ppm and coupling constants (J) were reported in Hz. The results of these experiments were processed with MestreNova software. 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.

Compound 1. Following SPPS methodology with Ala-Fmoc (311 mg, 1.0 mmol), His(Trt)-Fmoc (620 mg, 1.0 mmol), Gly-Fmoc (197 mg, 1.0 mmol), Val-Fmoc (339 mg, 1.0 mmol), Thr(α-D-GalNAc(AcO)$_3$)-Fmoc (318 mg, 0.474 mmol), Ser(^{t}Bu)-Fmoc (383 mg, 1.0 mmol) and Ala-Fmoc (311 mg, 1.0 mmol) glycopeptide **1** was obtained with a 94% yield after purification by reversed-phase HPLC and lyophilization. HPLC (*A*: MeCN; *B:* 0.1 % CF_3COOH/H_2O, A/B 2:98) t_R = 18.7 min. ^{1}H NMR (400 MHz, D_2O) δ = 0.95 (t, 6H, *J* = 6.9 Hz, $2CH_{3\ Val}$), 1.24 (d, 3H, *J* = 6.3 Hz, $CH_{3\ Thr}$), 1.39 (d, 3H, *J* = 7.2 Hz, $CH_{3\ Ala\text{-}7}$), 1.47 (d, 3H, *J* = 7.1 Hz, $CH_{3\ Ala\text{-}1}$), 1.99 – 2.14 (m, 4H, NHCOC$\underline{H}_3$, $H_{\beta\ Val}$), 3.15 – 3.36 (m, 2H, $2H_{\beta\ His}$), 3.65 – 3.77 (m, 2H, $2H_{6S}$), 3.79 – 3.89 (m, 3H, H_{3S}, $2H_{\beta\ Ser}$), 3.90 – 4.03 (m, 4H, H_{4S}, H_{5S}, $2H_{\alpha\ Gly}$), 4.03 – 4.12 (m, 2H, $H_{\alpha\ Val}$, H_{2S}), 4.19 – 4.37 (m, 3H, $H_{\alpha\ Ala\text{-}1}$, $H_{\beta\ Thr}$, $H_{\alpha\ Ala\text{-}7}$), 4.48 (t, 1H, *J* = 5.7 Hz, $H_{\alpha\ Ser}$), 4.61 (d, 1H, *J* = 1.9 Hz, $H_{\alpha\ Thr}$), 4.70 (t, 1H, *J* = 7.0 Hz, $H_{\alpha\ His}$), 4.91 (d, 1H, *J* = 3.7 Hz, H_{1S}), 7.32 (s, 1H, $H_{5\ arom\text{-}His}$), 8.62 (s, 1H, $H_{2\ arom\text{-}His}$). ^{1}H NMR [400 MHz, D_2O/H_2O (9:1)] δ = 7.04 (s, 1H, NH_2), 7.34 (s, 1H, NH_{His}), 7.63 (s, 1H, NH_2), 7.69 (d, 1H, *J* = 9.6 Hz, NHAc), 8.13 (d, 1H, *J* = 7.7 Hz, NH_{Val}), 8.42 (d, 1H, *J* = 7.5 Hz, NH_{Ser}), 8.51 (d, 1H, *J* = 6.0 Hz, $NH_{Ala\text{-}7}$), 8.54 (t, 1H, *J* = 5.2 Hz, NH_{Gly}), 8.61 (s, 1H, NH_{His}), 8.65 (d, 1H, *J* = 8.7 Hz, NH_{Thr}). ^{13}C NMR (100 MHz, D_2O) δ 16.4, 16.7 ($CH_{3\ Ala\text{-}1}$, $CH_{3\ Ala\text{-}7}$), 17.7 ($CH_{3\ Val}$), 18.1 ($CH_{3\ Thr}$), 18.5 ($CH_{3\ Val}$), 22.2 (NHCO$\underline{C}H_3$), 26.1 ($C_{\beta\ His}$), 30.3 ($C_{\beta\ Val}$), 42.3($C_{\alpha\ Gly}$), 48.8 ($C_{\alpha\ Val}$), 49.5, 49.7 ($C_{\alpha\ Ala\text{-}1}$, $C_{\alpha\ Ala\text{-}7}$), 52.7 ($C_{\alpha\ His}$), 54.8 ($C_{\alpha\ Ser}$), 57.2 ($C_{\alpha\ Thr}$), 59.3 (C_{2S}), 61.3, 61.5 ($C_{\beta\ Ser}$, C_{6S}), 68.1, 68.6 (C_{3S}, C_{4S}), 71.3 (C_{5S}), 76.2 ($C_{\beta\ Thr}$), 98.8 (C_{1S}), 117.4 ($C_{5\ arom\text{-}His}$), 128.2 ($C_{4\ arom\text{-}His}$), 133.6 ($C_{2\ arom\text{-}His}$), 170.8, 170.9, 170.9, 171.0 171.7, 173.8, 174.0, 177.6 (8 CO). HRMS (ESI$^+$): calcd for $C_{34}H_{58}N_{11}O_{14}$ (MH^+) 844.4159, found (MH^+) 844.4141.

Compound 2. Following SPPS methodology with Ala-Fmoc (311 mg, 1.0 mmol), Pro-Fmoc (337 mg, 1.0 mmol), Asp(O^{t}Bu)-Fmoc (412 mg, 1.0 mmol), Thr(α-D-GalNAc(AcO)$_3$)-Fmoc (318 mg, 0.474 mmol) and Arg(Pbf)-Fmoc (649 mg, 1.0 mmol), glycopeptide **2** was obtained with a 94% yield after purification by reversed-phase HPLC and lyophilization. HPLC (*A*: MeCN; *B:* 0.1 % CF_3COOH/H_2O, A/B 2:98) within 40 min; t_R = 15.5 min. ^{1}H NMR (400 MHz, D_2O) δ = 1.28 (d, 3H, *J* = 6.4 Hz, $CH_{3\ Thr}$), 1.54 (d, 3H, *J* = 7.0 Hz, $CH_{3\ Ala}$), 1.61 – 1.81 (m, 4H, $2H_{\beta\ Arg}$, $2H_{\gamma\ Arg}$), 1.83 – 1.97 (m, 1H, $H_{\beta\ Pro}$), 1.99 – 2.12 (m, 5H, NHCOC$\underline{H}_3$, $2H_{\gamma\ Pro}$), 2.31 – 2.40 (m, 1H, $H_{\beta\ Pro}$), 2.83 – 3.07 (m, 2H, $2H_{\beta\ Asp}$), 3.21 (t, 2H, *J* = 6.8 Hz, $2H_{\delta\ Arg}$), 3.57 – 3.68 (m, 1H, $H_{\delta\ Pro}$), 3.68 – 3.78 (m, 3H, $H_{\delta\ Pro}$, $2H_{6S}$), 3.88 (dd, 1H, *J* = 11.0 Hz, 3.2 Hz, H_{3S}), 3.94 – 3.98 (m, 1H, H_{4S}), 4.03 (t, 1H, *J* = 6.2 Hz, H_{5S}), 4.13 (dd, 1H, *J* = 11.0, 3.7 Hz, H_{2S}), 4.25 – 4.40 (m, 3H, $H_{\alpha\ Ala}$, $H_{\beta\ Thr}$, $H_{\alpha\ Arg}$), 4.47 – 4.56 (m, 2H, $H_{\alpha\ Thr}$, $H_{\alpha\ Pro}$), 4.83 – 4.93 (m, 2H, $H_{\alpha\ Asp}$, H_{1S}). ^{1}H NMR [400 MHz, D_2O/H_2O (9:1)] δ = 6.40-6.65 (br s, 2H, $NH_{2\ Arg}$), 6.94 (s, 1H, NH_2), 7.25 (t, 1H, *J* = 4.9 Hz, NH_{Arg}), 7.54 (s, 1H, NH_2), 7.59 (d, 1H, *J* = 9.6 Hz, NHAc), 8.27 (d, 1H, *J* = 6.8 Hz, NH_{Arg}), 8.42 (d, 1H, *J* = 8.7 Hz, NH_{Thr}), 8.48 (d, 1H, *J* = 6.7 Hz, NH_{Asp}). A second set of signals (in a small percentage) is observed, especially in the NH region. They correspond to the *cis* disposition of the amide bond of proline residues.[S12] However, in all the MD simulations, only the set of signals corresponding to *trans* disposition has been considered. ^{13}C NMR (100 MHz, D_2O) δ 15.1 ($CH_{3\ Ala}$), 18.4 ($CH_{3\ Thr}$), 22.4 (NHCO$\underline{C}H_3$), 24.5, 24.7 ($C_{\gamma\ Pro}$, $C_{\gamma\ Arg}$), 28.6 ($C_{\beta\ Arg}$), 29.4 ($C_{\beta\ Pro}$), 35.0 ($C_{\beta\ Asp}$), 40.6 ($C_{\delta\ Arg}$), 47.8 ($C_{\delta\ Pro}$), 48.7 ($C_{\alpha\ Ala}$), 49.8, 49.9 ($C_{\alpha\ Asp}$, C_{2S}), 53.6 ($C_{\beta\ Thr}$), 57.5 ($C_{\alpha\ Pro}$), 60.3 ($C_{\alpha\ Arg}$), 61.5 (C_{6S}), 68.1, 68.3, 71.8 (C_{3S}, C_{4S}, C_{5S}), 76.3 ($C_{\alpha\ Thr}$), 98.8 (C_{1S}), 156.7 (CNH_{Arg}) 168.1, 170.7, 172.6, 173.0, 173.9, 174.1, 175.6 (7 CO). HRMS (ESI+): calcd for $C_{30}H_{53}N_{10}O_{13}$ (MH+) 761.3788, found (MH+) 761.3792.

Compound 3. Following SPPS methodology with Ala-Fmoc (311 mg, 1.0 mmol), Pro-Fmoc (337 mg, 1.0 mmol), Gly-Fmoc (297 mg, 1.0 mmol), Ser(O^{t}Bu)-Fmoc (383 mg, 1.0 mmol), Thr(α-D-GalNAc(AcO)$_3$)-Fmoc (318 mg, 0.474 mmol), Ala-Fmoc (311 mg, 1.0 mmol) and Pro-Fmoc (337 mg, 1.0 mmol) glycopeptide **3** was obtained with a 94% yield after purification by reversed-phase HPLC and lyophilization. HPLC *(A*: MeCN; *B:* 0.1 % CF_3COOH/H_2O, A/B 2:98) within 40 min; t_R = 21.1 min. ^{1}H NMR (400 MHz, D_2O) δ 1.26 (d, 3H, *J* = 6.4 Hz, $CH_{3\ Thr}$), 1.36 (d, 3H, *J* = 7.1 Hz, $CH_{3\ Ala\text{-}6}$), 1.53 (d, 3H, *J* = 7.0 Hz, $CH_{3\ Ala\text{-}1}$), 1.90 – 2.13 (m, 9H, NHCOC$\underline{H}_3$, $H_{\beta\ Pro\text{-}2}$, $H_{\beta\ Pro\text{-}7}$, $2H_{\gamma\ Pro\text{-}2}$, $2H_{\gamma\ Pro\text{-}7}$), 2.26 – 2.40 (m, 2H, $H_{\beta\ Pro\text{-}2}$, $H_{\beta\ Pro\text{-}7}$), 3.60 – 3.80 (m, 6H, $2H_{\delta\ Pro\text{-}2}$, $2H_{\delta\ Pro\text{-}7}$, $2H_{6S}$), 3.83 – 4.05 (m, 7H, $2H_{\beta\ Ser}$, $2H_{\alpha\ Gly}$, H_{3S}, H_{4S}, H_{5S}), 4.09 (dd, 1H, *J* = 11.0, 3.8 Hz, H_{2S}), 4.32 – 4.41 (m, 3H, $H_{\alpha\ Ala\text{-}6}$, $H_{\beta\ Thr}$, $H_{\alpha\ Pro\text{-}2}$), 4.45 – 4.62 (m, 3H, $H_{\alpha\ Thr}$, $H_{\alpha\ Ala\text{-}1}$, $H_{\alpha\ Pro\text{-}7}$), 4.63 – 4.68 (m, 1H, $H_{\alpha\ Ser}$), 4.91 (d, 1H, *J* = 3.8 Hz, H_{1S}). ^{1}H NMR [400 MHz, D_2O/H_2O (9:1)] δ = 6.98 (s, 1H, NH_2), 7.65 (s, 1H, NH_2), 7.88 (d, 1H, *J* = 9.6 Hz, NHAc), 8.19 (d, 1H, *J* = 6.8 Hz, NH_{Ser}), 8.33 (d, 1H, *J* = 6.1 Hz, $NH_{Ala\text{-}6}$), 8.57 (t, 1H, *J* = 5.8 Hz, NH_{Gly}), 8.65 (d, 1H, *J* = 9.0 Hz, NH_{Thr}). A second set of signals (in a small percentage) is observed, especially in the NH region. They correspond to the *cis* disposition of the amide bond of proline residues.[S1] However, in all the MD simulations, only the set of signals corresponding to *trans* disposition has been considered. ^{13}C NMR (100 MHz, D_2O) δ 15.0 ($CH_{3\ Ala\text{-}1}$), 15.5 ($CH_{3\ Ala\text{-}6}$), 18.2 ($CH_{3\ Thr}$), 22.2 (NHCO$\underline{C}H_3$), 24.6, 24.7 ($C_{\gamma\ Pro\text{-}2}$, $C_{\gamma\ Pro\text{-}7}$), 29.3, 29.5 ($C_{\beta\ Pro\text{-}2}$, $C_{\beta\ Pro\text{-}7}$), 42.3 ($CH_{2\ Gly}$), 47.5, 47.6, 47.7,

48.1 ($C_{\delta\ Pro\text{-}2}$, $C_{\delta\ Pro\text{-}7}$, $C_{\alpha\ Ala\text{-}1}$, $C_{\alpha\ Ala\text{-}6}$), 49.7 (C_{2S}), 55.1 ($C_{\alpha\ Ser}$), 57.2 ($C_{\alpha\ Thr}$), 60.1 ($C_{\alpha\ Pro\text{-}7}$), 60.7 ($C_{\alpha\ Pro\text{-}2}$), 61.2, 61.3 (C_{6S}, $C_{\beta\ Ser}$), 68.1, 68.6, 71.4 (C_{3S}, C_{4S}, C_{5S}), 98.5 (C_{1S}), 169.4, 170.7, 171.2, 172.2, 172.5, 173.8, 174.5, 176.9 (8 CO). HRMS (ESI+): calcd for $C_{33}H_{56}N_9O_{14}$ (MH^+) 802.3941, found (MH^+) 802.3957.

Compound 4. Following SPPS methodology with Ala-Fmoc (311 mg, 1.0 mmol), Pro-Fmoc (337 mg, 1.0 mmol), Ala-Fmoc (311 mg, 1.0 mmol), Thr(α-D-GalNAc$(AcO)_3$)-Fmoc (318 mg, 0.474 mmol) and Ala-Fmoc (311 mg, 1.0 mmol) glycopeptide **4** is obtained with a 92% yield after purification by reversed-phase HPLC and lyophilization. HPLC *(A*: MeCN; *B:* 0.1 % CF_3COOH/H_2O, A/B 2:98) within 40 min; t_R = 15.3 min. ^{1}H NMR (400 MHz, D_2O) δ 1.32 (d, 3H, *J* = 6.3 Hz, $CH_{3\ Thr}$), 1.40 – 1.49 (m, 6H, $CH_{3\ Ala3}$, $CH_{3\ Ala5}$), 1.57 (d, 3H, *J* = 7.0 Hz, $CH_{3\ Ala1}$), 1.88 – 2.00 (m, 1H, $H_{\beta\ Pro}$), 2.00 – 2.16 (m, 5H, $NHCOCH_3$, $2H_{\gamma\ Pro}$), 2.33 – 2.42 (m, 1H, $H_{\beta\ Pro}$), 3.60 – 3.86 (m, 4H, $2H_{\delta\ Pro}$, $2H_{6S}$), 3.87 – 4.02 (m, 2H, H_{3S}, H_{4S}), 4.07 (t, 1H, *J* = 6.1 Hz, H_{5S}), 4.12 – 4.21 (m, 1H, H_{2S}), 4.25 – 4.46 (m, 3H, $H_{\alpha\ Ala3}$, $H_{\alpha\ Ala5}$, $H_{\beta\ Thr}$), 4.48 – 4.57 (m, 3H, $H_{\alpha\ Ala1}$, $H_{\alpha\ Pro}$, $H_{\alpha\ Thr}$), 4.96 (d, 1H, J = 3.7 Hz, H_{1S}). ^{1}H NMR [400 MHz, D_2O/H_2O (9:1)] δ = 6.97 (s, 1H, NH_2), 7.67 (s, 1H, NH_2), 7.75 (d, 1H, *J* = 9.6 Hz, NHAc), 8.38 (d, 1H, *J* = 6.1 Hz, $NH_{Ala\text{-}3}$), 8.47 (d, 1H, *J* = 8.7 Hz, NH_{Thr}), 8.57 (t, 1H, *J* = 5.7 Hz, $NH_{Ala\text{-}5}$). ^{13}C NMR (100 MHz, D_2O) δ 15.1 ($CH_{3\ Ala1}$), 16.5, 17.1 ($CH_{3\ Ala3}$, $CH_{3\ Ala5}$) 18.3 ($CH_{3\ Thr}$), 22.3 ($NHCOCH_3$), 24.8 ($C_{\gamma\ Pro}$), 29.5 ($C_{\beta\ Pro}$), 47.8 ($C_{\delta\ Pro}$), 48.1, 49.2, 49.5, 49.7, 49.9 ($C_{\alpha\ Ala1}$, $C_{\alpha\ Ala3}$, $C_{\alpha\ Ala5}$, $C_{\alpha\ Thr}$, C_{2S}), 57.1 ($C_{\alpha\ Pro}$), 61.3 (C_{6S}), 68.1, 68.6, 71.4 (C_{3S}, C_{4S}, C_{5S}), 76.3 ($C_{\beta\ Thr}$), 98.8 (C_{1S}), 169.1, 170.9, 173.5, 174.4, 175.3, 177.3 (6 CO). HRMS (ESI+): calcd for $C_{26}H_{46}N_7O_{11}$ (MH^+) 632.3250, found (MH^+) 632.3252.

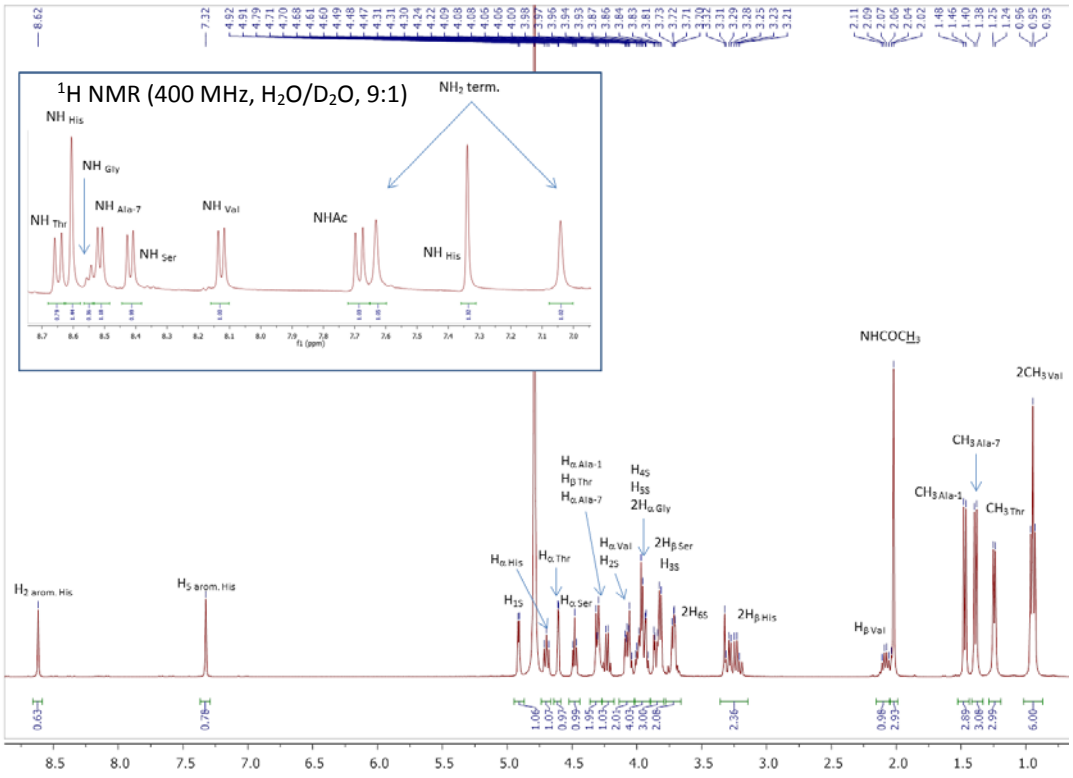
Compound m1. Following SPPS methodology with the adequately protected amino acids or glycoamino acids, compound **m1** was obtained and purified by semi-preparative HPLC: t_R = 24.85 min (Phenomenex Luna C18 (2), 21.20×250 mm, Grad: acetonitrile/water+0.1% TFA (10:90) → (18:82), 30 min, λ = 212 nm). HRMS (ESI+) m/z: calcd. for $[M+3H]^{3+}$: 720.6942 found: 720.6946.

Compound m1'. Following SPPS methodology with the adequately protected amino acids or glycoamino acids, compound **m1'** was obtained and purified by semi-preparative HPLC: t_R = 31.07 min (Phenomenex Luna C18 (2), 21.20×250 mm, Grad: acetonitrile/water+0.1% TFA (10:90) → (18:82), 30 min, λ = 212 nm). HRMS (ESI+) m/z: calcd. for $[M+3H]^{3+}$: 720.6942 found: 720.6938.

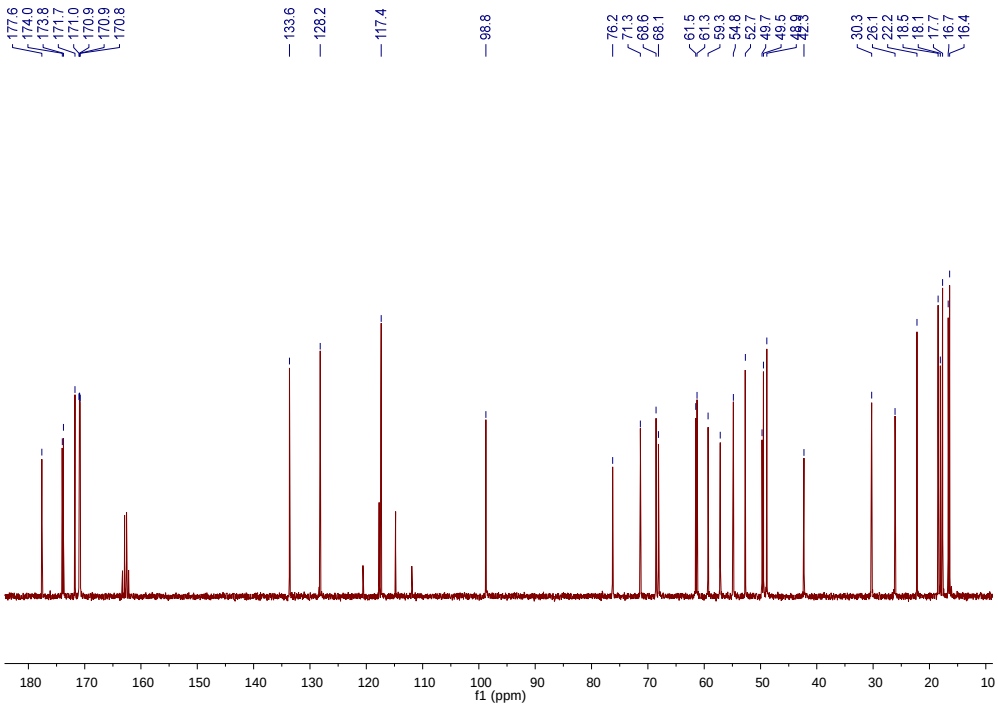
Compound m2. Following SPPS methodology with the adequately protected amino acids or glycoamino acids, compound **m2** was obtained and purified by semi-preparative HPLC: t_R = 17.40 min (Phenomenex Luna C18 (2), 21.20×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (18:82), 30 min, λ = 212 nm). HRMS (ESI+) m/z: calcd. for $[M+3H]^{3+}$: 720.6942 found: 720.6936.

Compound m3. Following SPPS methodology with the adequately protected amino acids or glycoamino acids, compound **m3** was obtained and purified by semi-preparative HPLC: t_R = 21.92 min (Phenomenex Luna C18 (2), 21.20×250 mm, Grad: acetonitrile/water+0.1% TFA (10:90) → (18:82), 40 min, λ = 212 nm). HRMS (ESI+) m/z: calcd. for $[M+3H]^{3+}$: 720.6942 found: 720.6948.

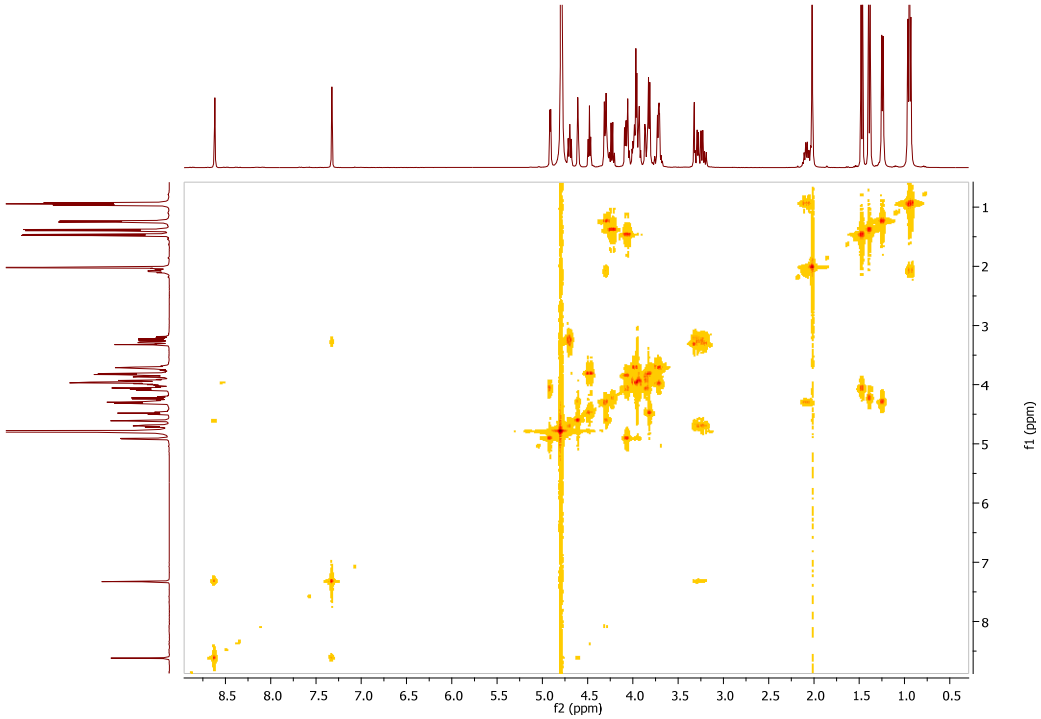
^{1}H NMR 400 MHz in D_2O registered at 298K for **compound 1**



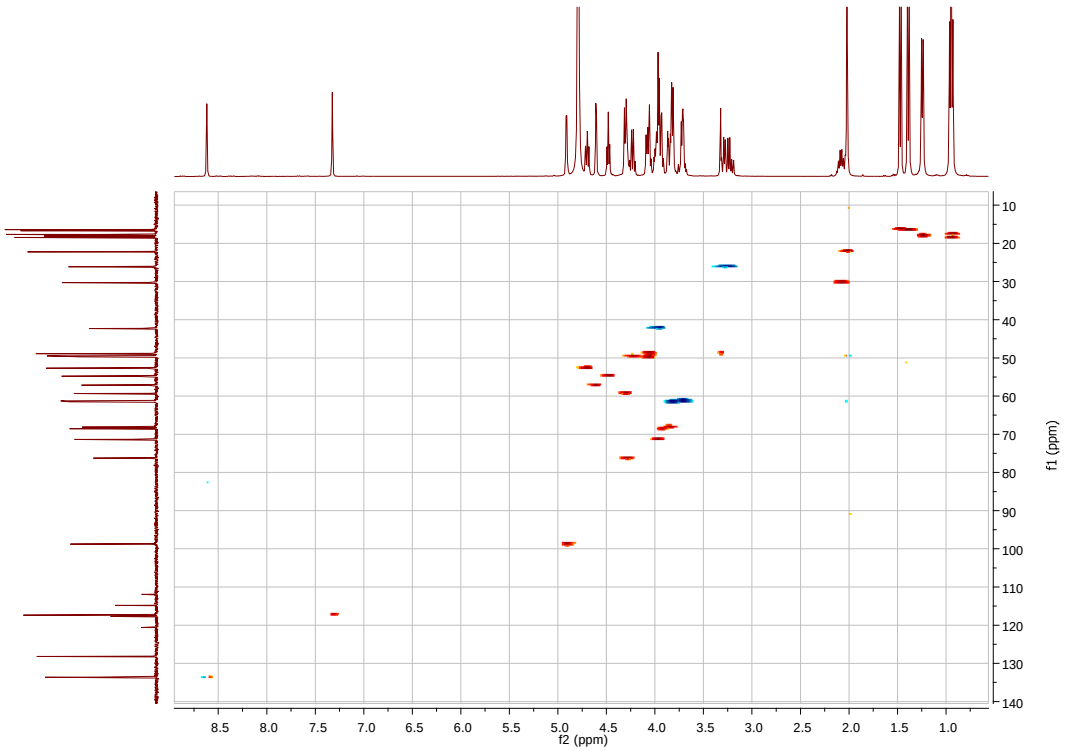
^{13}C NMR 100 MHz in D_2O registered at 298K for **compound 1**



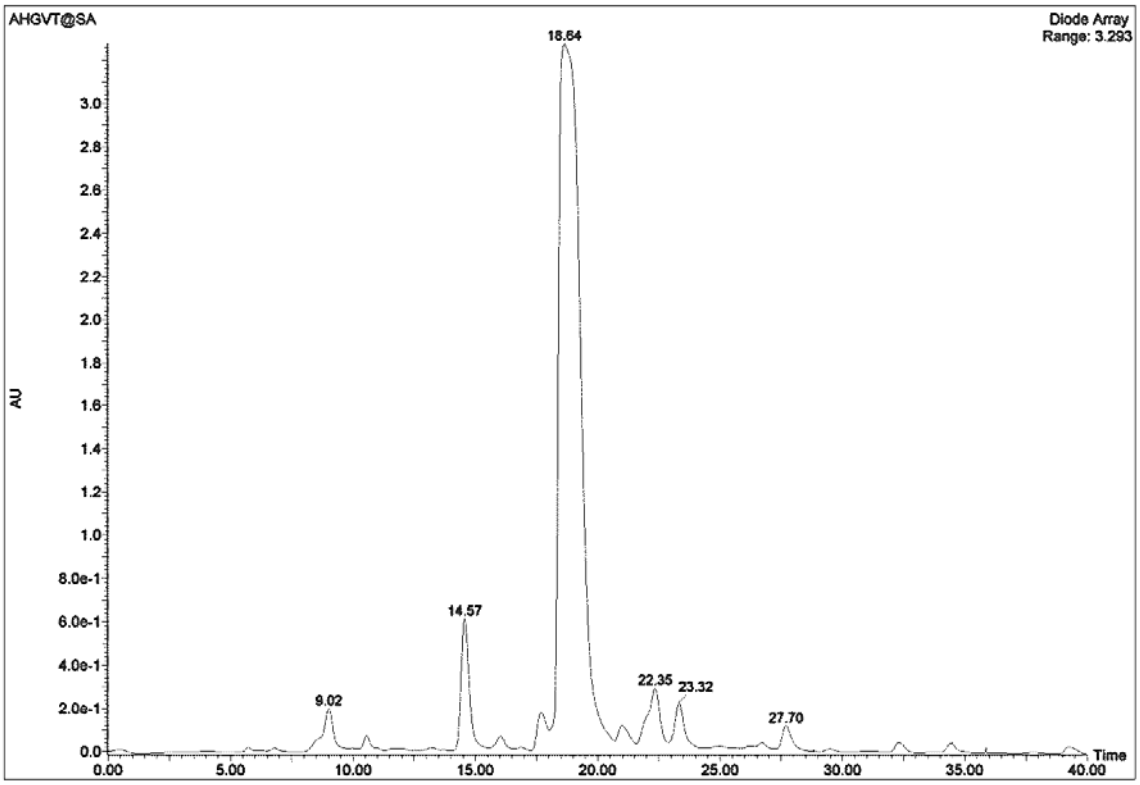
COSY in D_2O registered at 298K for **compound 1**



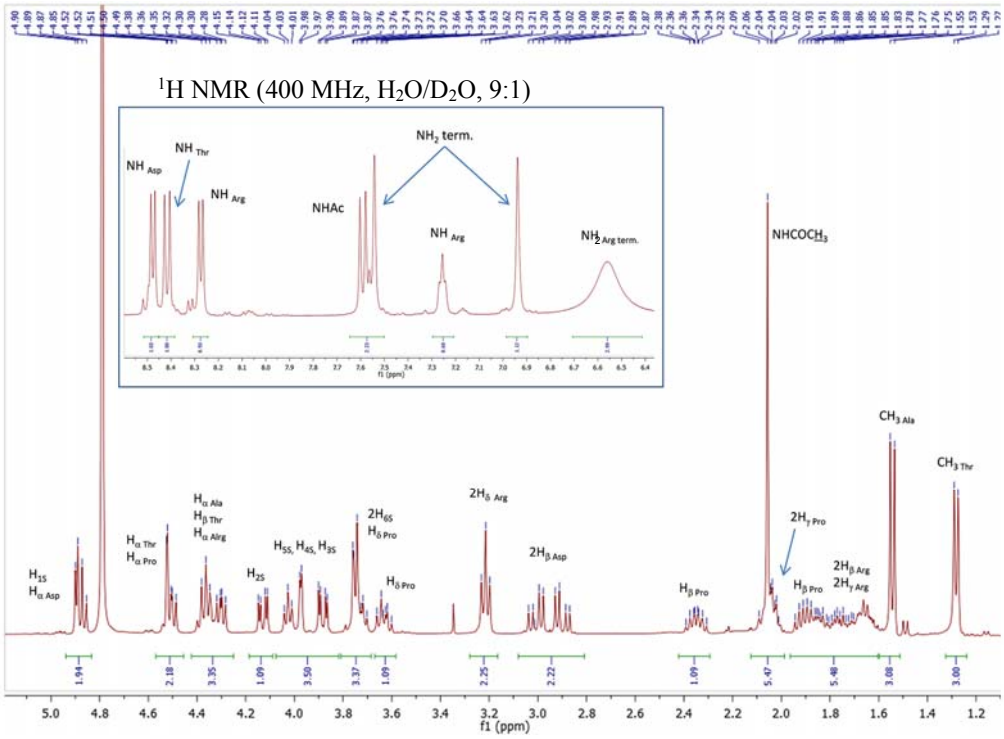
HSQC in D_2O registered at 298K for **compound 1**



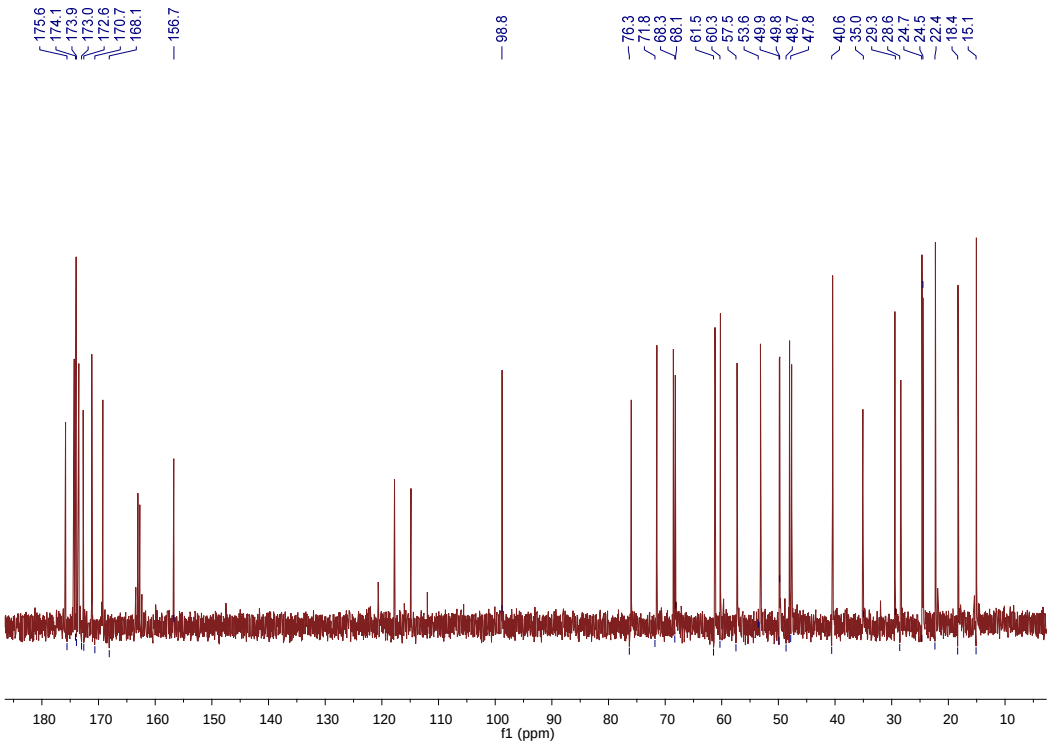
HPLC chromatogram obtained for **compound 1**



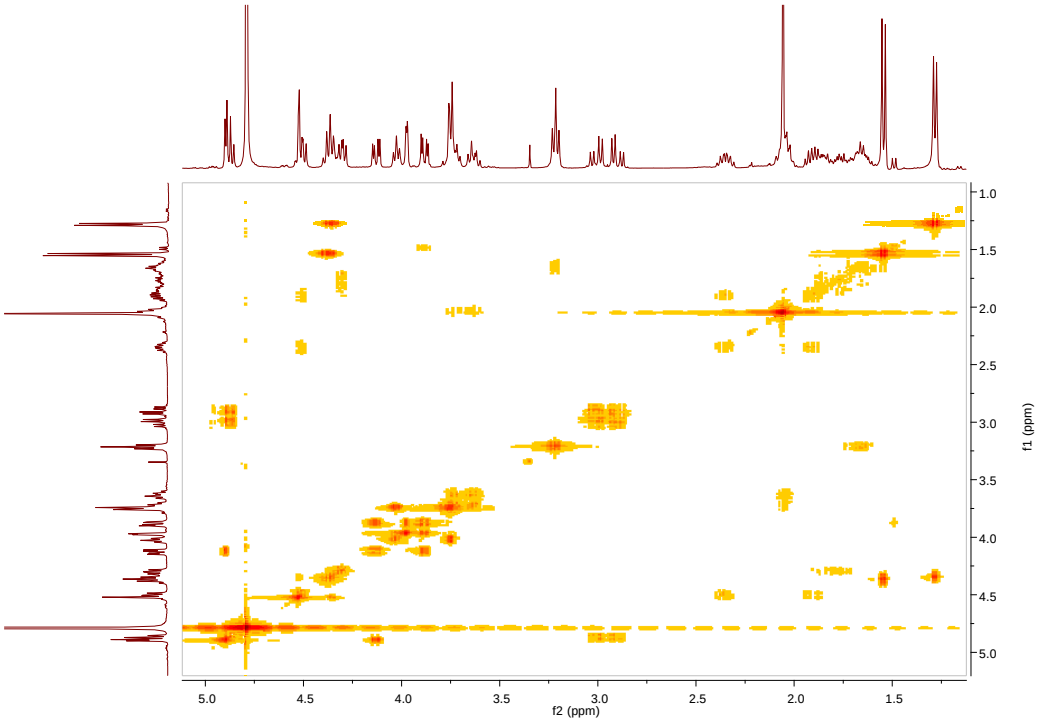
^{1}H NMR 400 MHz in D_2O registered at 298K for **compound 2**



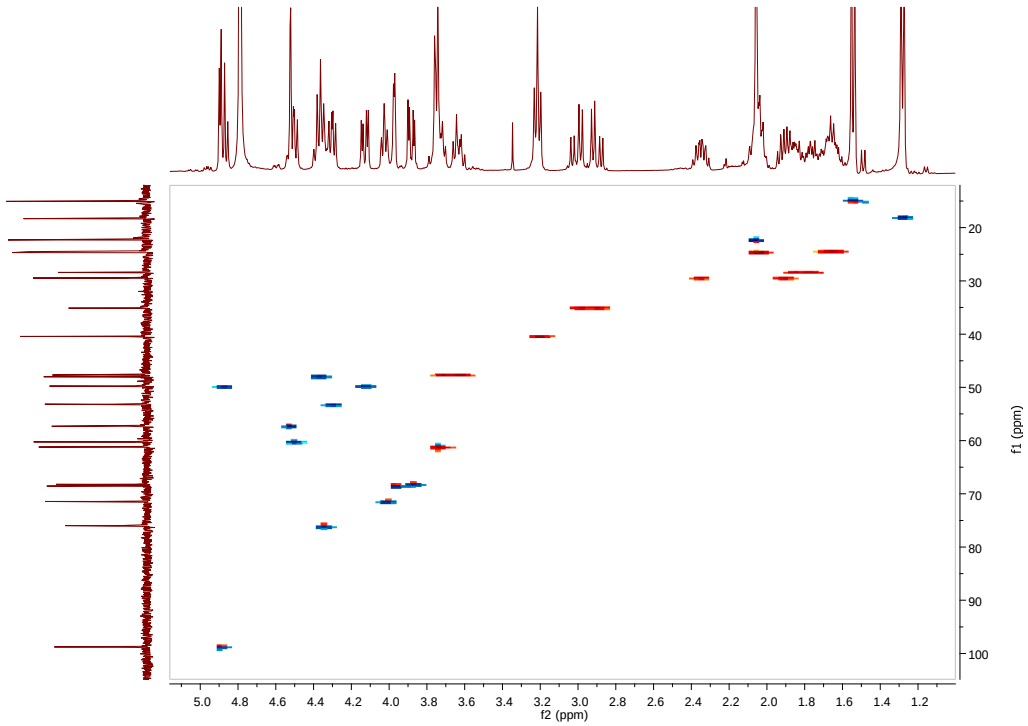
^{13}C NMR 100 MHz in D_2O registered at 298K for **compound 2**



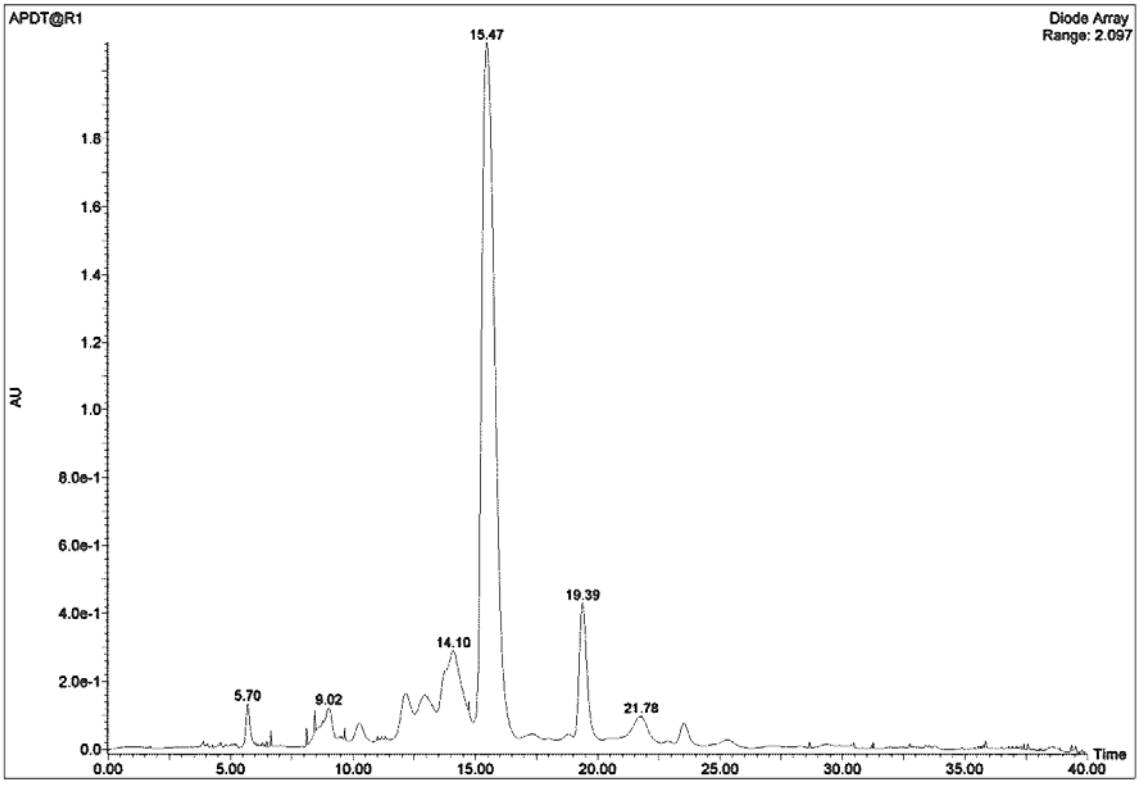
COSY in D_2O registered at 298K for **compound 2**



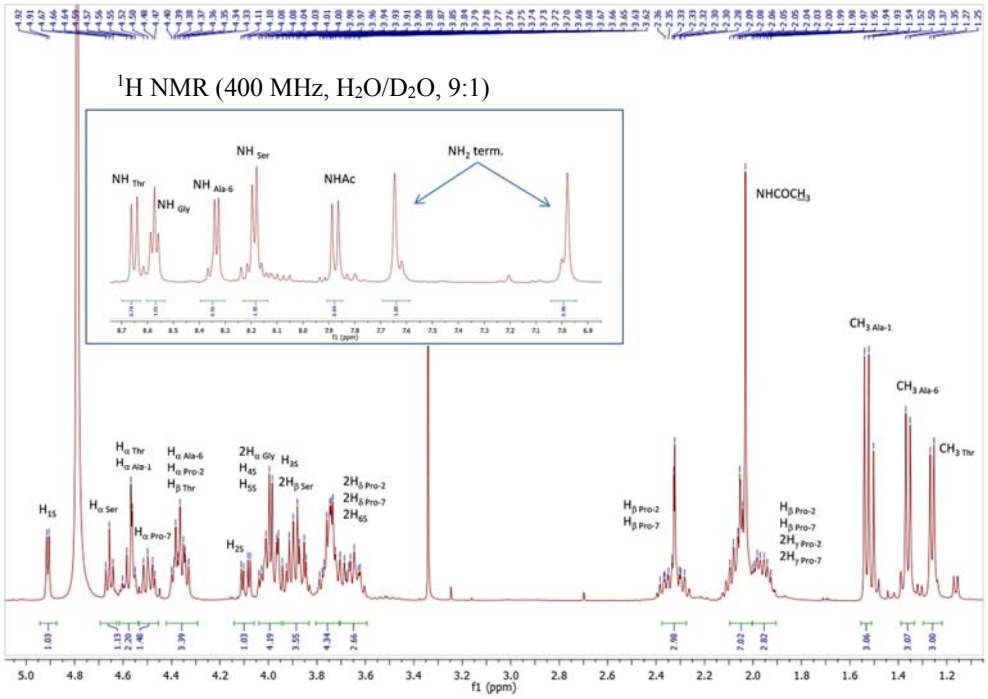
HSQC in D_2O registered at 298K for **compound 2**



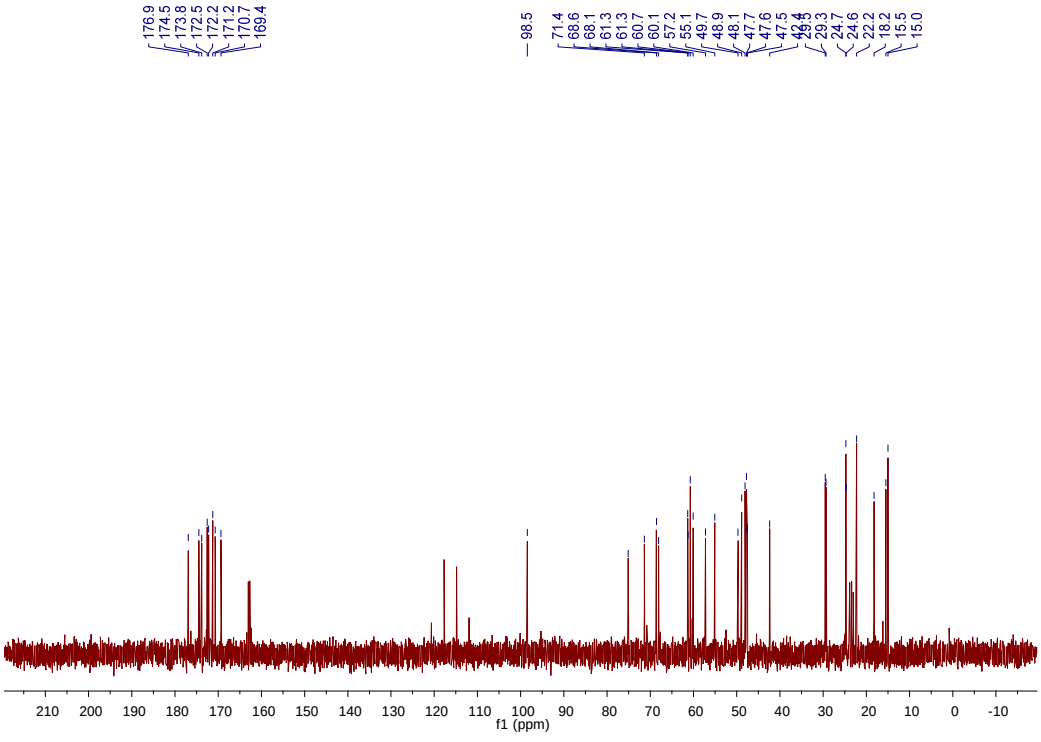
HPLC chromatogram for **compound 2**



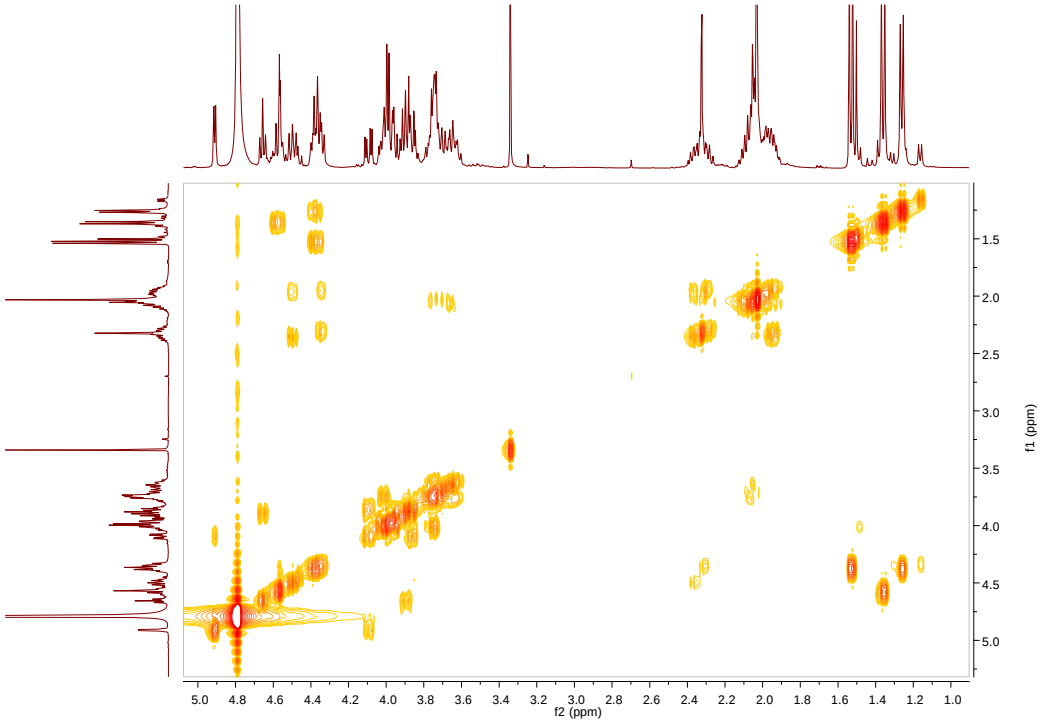
^{1}H NMR 400 MHz in D_2O registered at 298K for **compound 3**



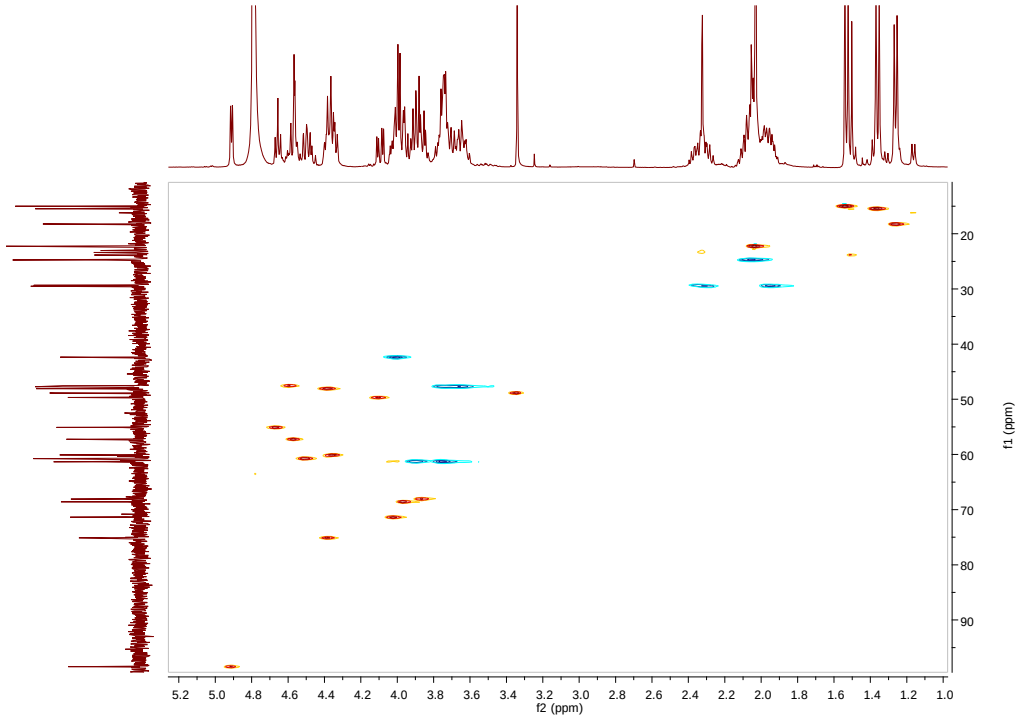
^{13}C NMR 100 MHz in D_2O registered at 298K for **compound 3**



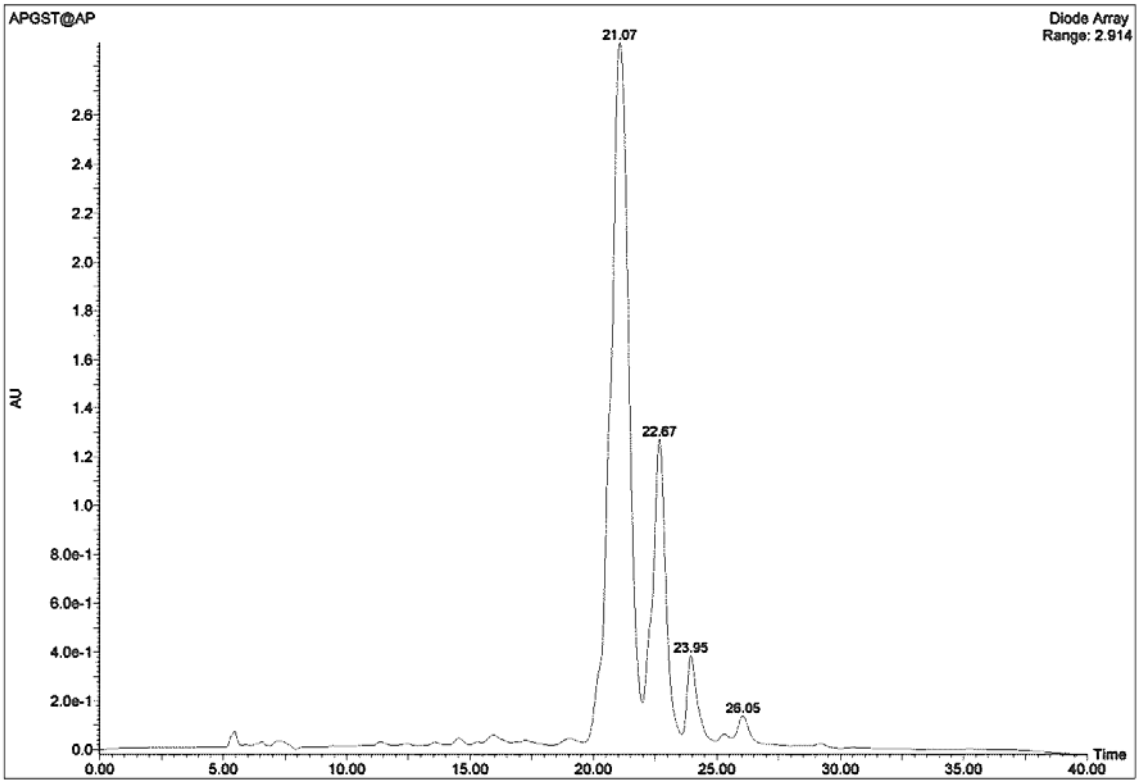
COSY in D_2O registered at 298K for **compound 3**



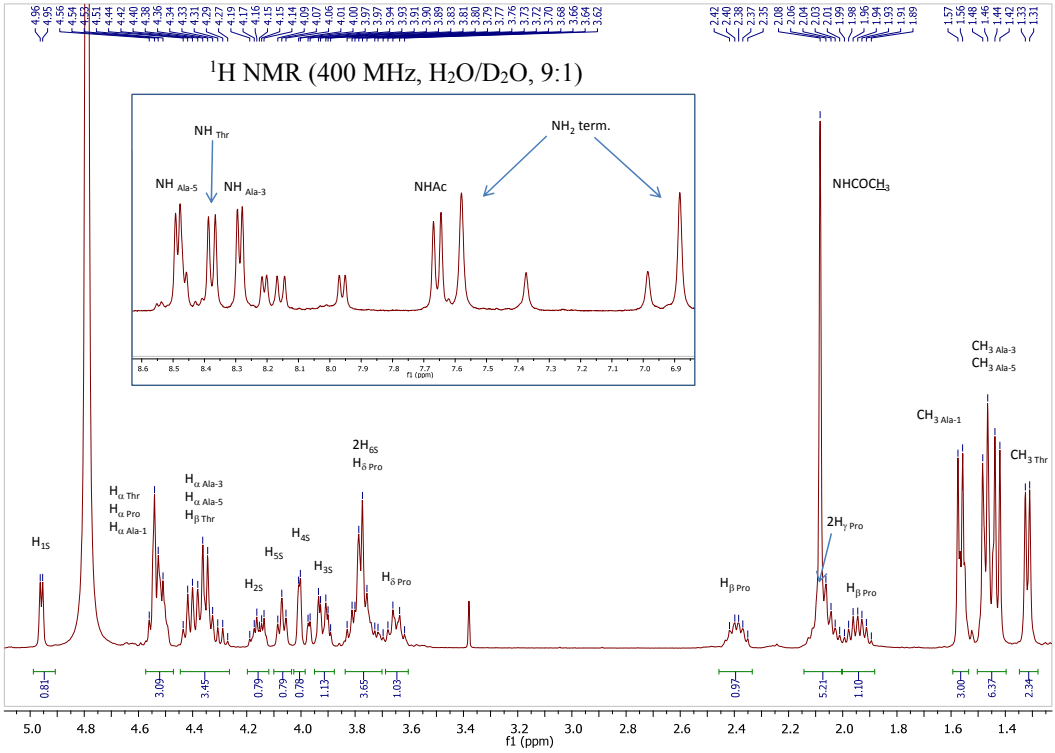
HSQC in D_2O registered at 298K for **compound 3**



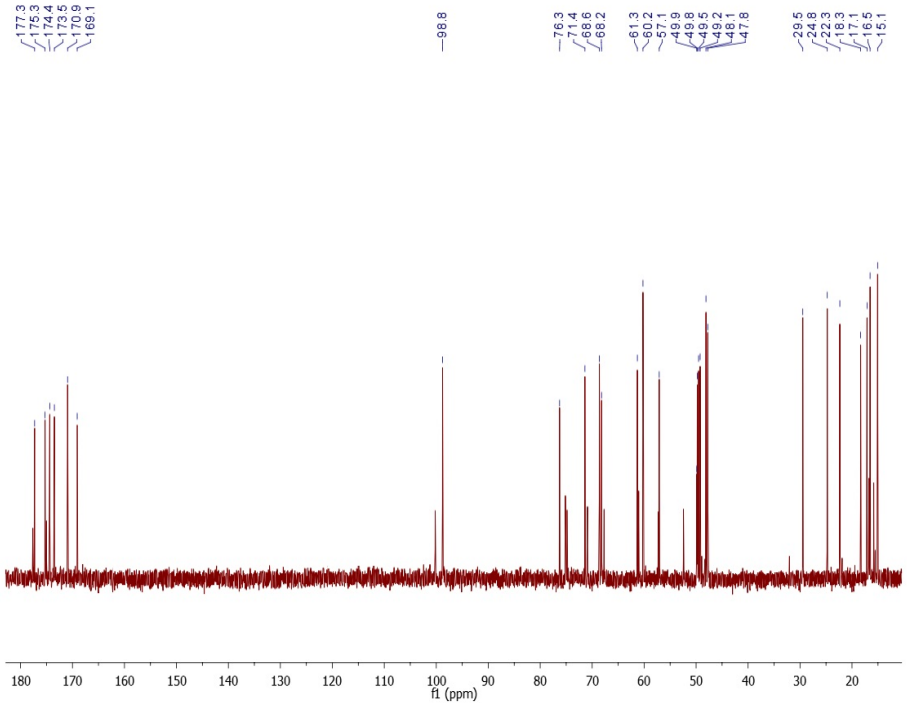
HPLC chromatogram for **compound 3**



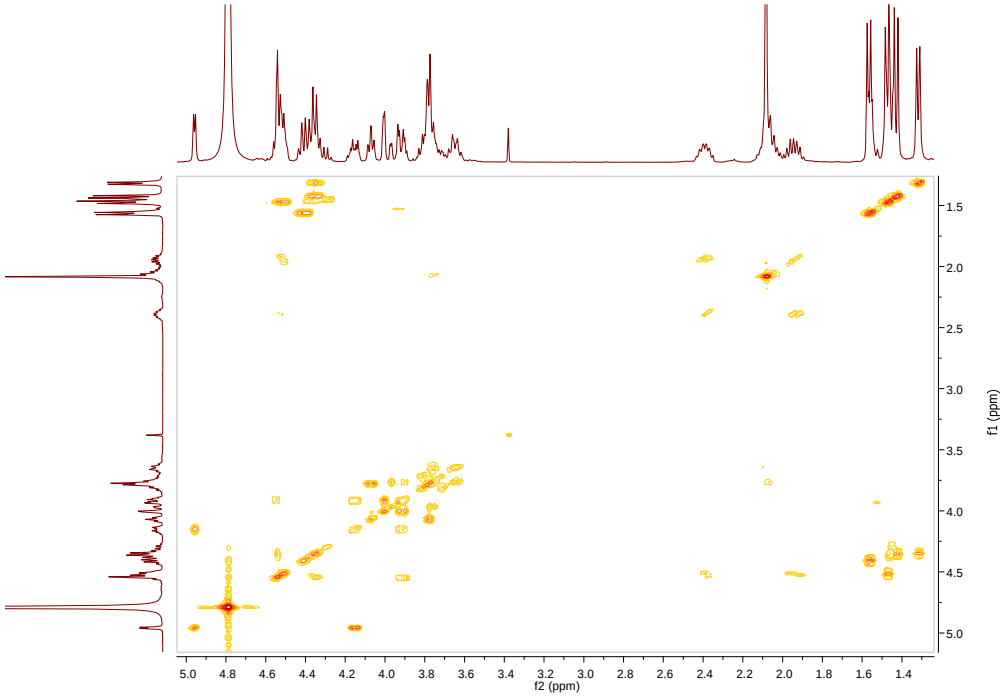
^{1}H NMR 400 MHz in D_2O registered at 298K for **compound 4**



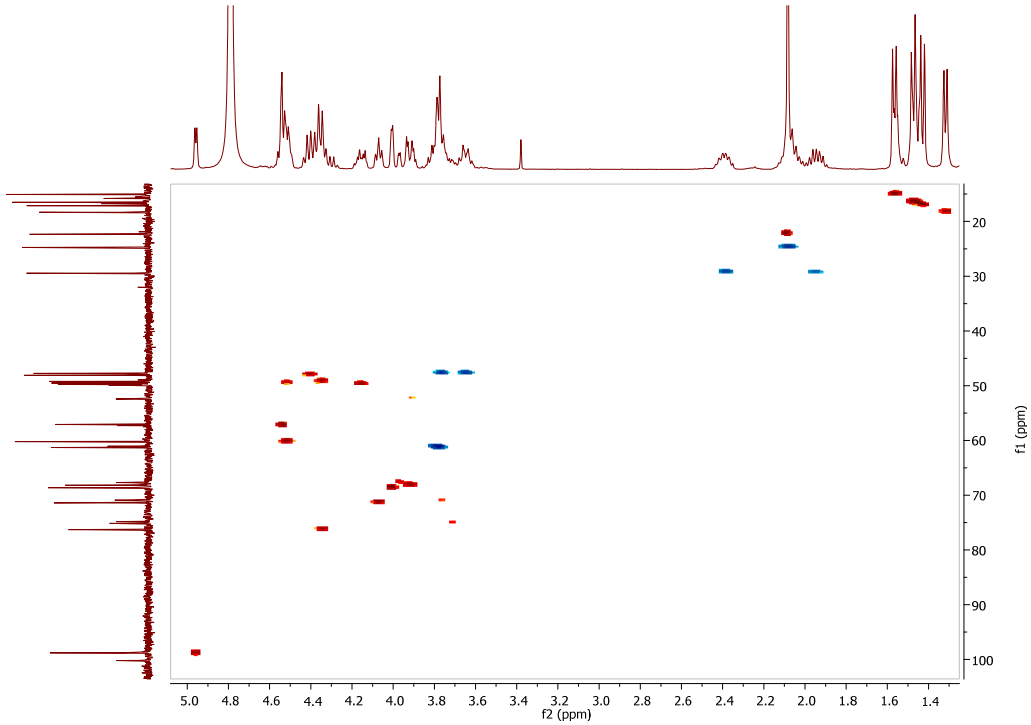
^{13}C NMR 100 MHz in D_2O registered at 298K for **compound 4**



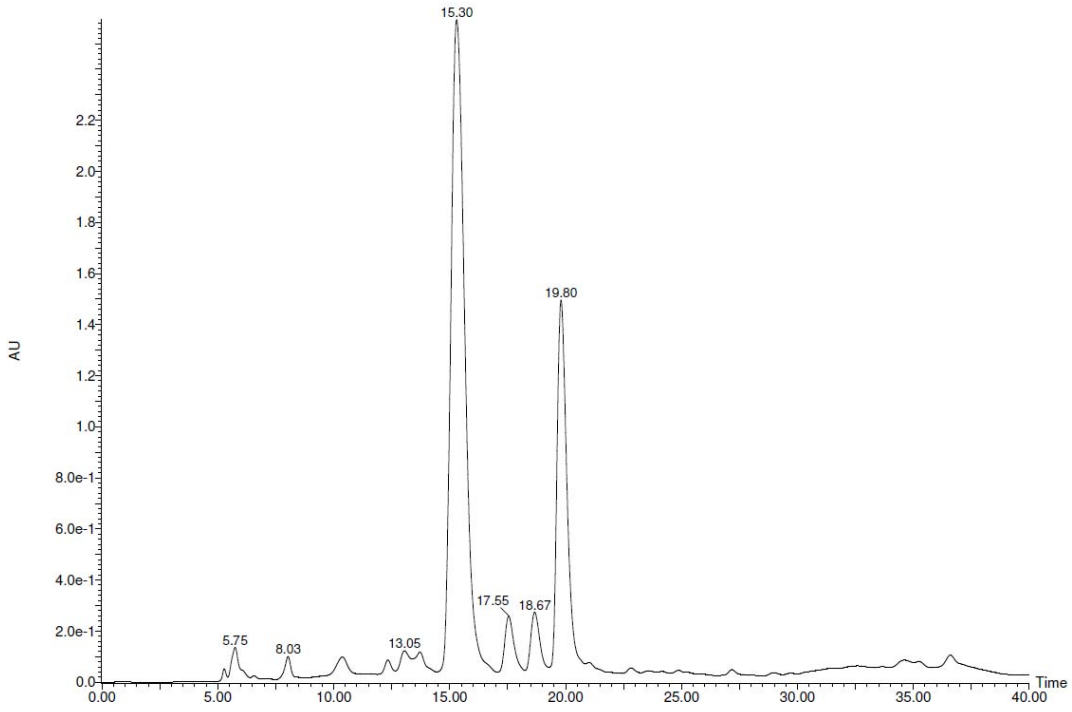
COSY in D_2O registered at 298K for **compound 4**



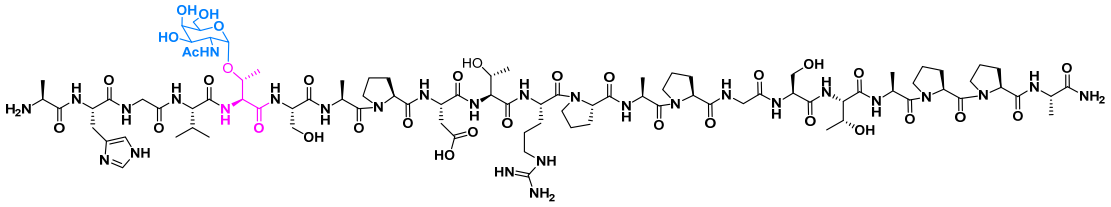
HSQC in D_2O registered at 298K for **compound 4**



HPLC chromatogram for **compound 4**



Compound m1

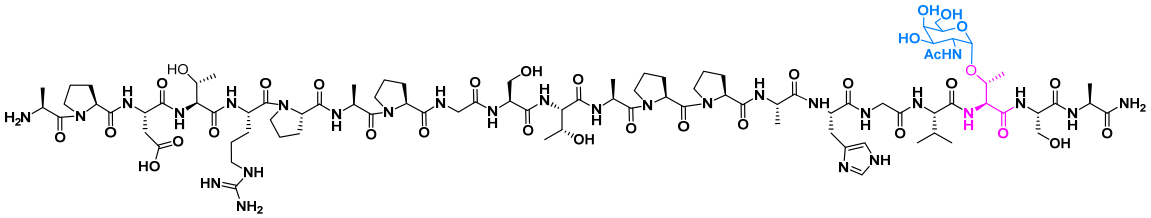


HRMS (ESI+) m/z: calcd. for $[M+3H]^{3+}$: 720.6942 found: 720.6946.

Semi-preparative HPLC:

R_t = 24.85 min (Phenomenex Luna C18 (2), 21.20×250mm, Grad: acetonitrile/water+0.1% TFA (10:90) → (18:82), 30 min, λ = 212nm).

Compound m1'

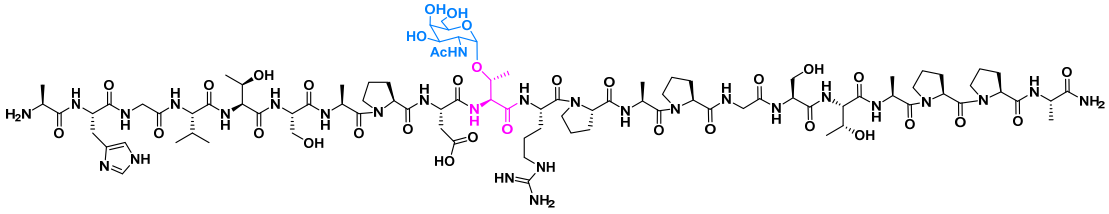


HRMS (ESI+) m/z: calcd. for $[M+3H]^{3+}$: 720.6942 found: 720.6938.

Semi-preparative HPLC:

R_t = 31.07 min (Phenomenex Luna C18 (2), 21.20×250mm, Grad: acetonitrile/water+0.1% TFA (10:90) → (18:82), 30 min, λ = 212nm).

Compound m2

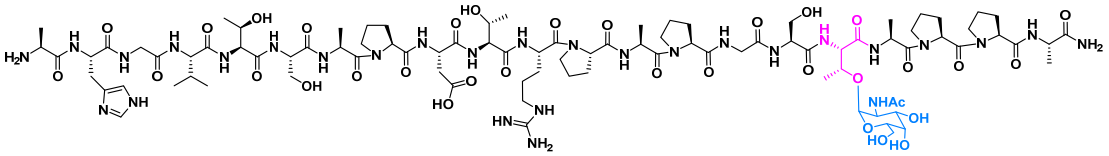


HRMS (ESI+) m/z: calcd. for $[M+3H]^{3+}$: 720.6942 found: 720.6936.

Semi-preparative HPLC:

R_t = 17.40 min (Phenomenex Luna C18 (2), 21.20×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (18:82), 30 min, λ = 212nm).

Compound m3



HRMS (ESI+) m/z: calcd. for $[M+3H]^{2+}$: 720.6942 found: 720.6948.

Semi-preparative HPLC:

R_t = 21.92 min (Phenomenex Luna C18 (2), 21.20×250mm, Grad: acetonitrile/water+0.1% TFA (10:90) → (18:82), 40 min, λ = 212nm).

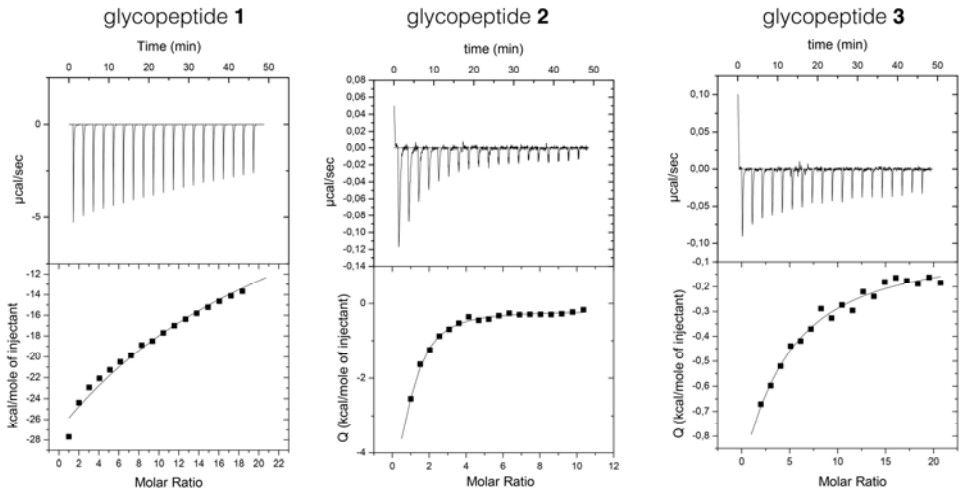


Figure S1. ITC profiles for glycopeptides **1**-**3** titration into SBA lectin solutions at 25 °C and pH 7.2. The solid line represents the least-squares fitting of the data to the simplest model (one binding site). See main text for more experimental details.

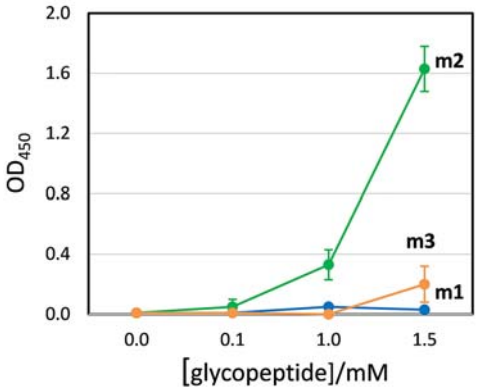


Figure S2. Binding curves for mucin derivatives **m1**, **m2** and **m3** with SBA lectin. Absorbance signals are the average of three replicate wells and the error bars show the standard deviations for these measurements.

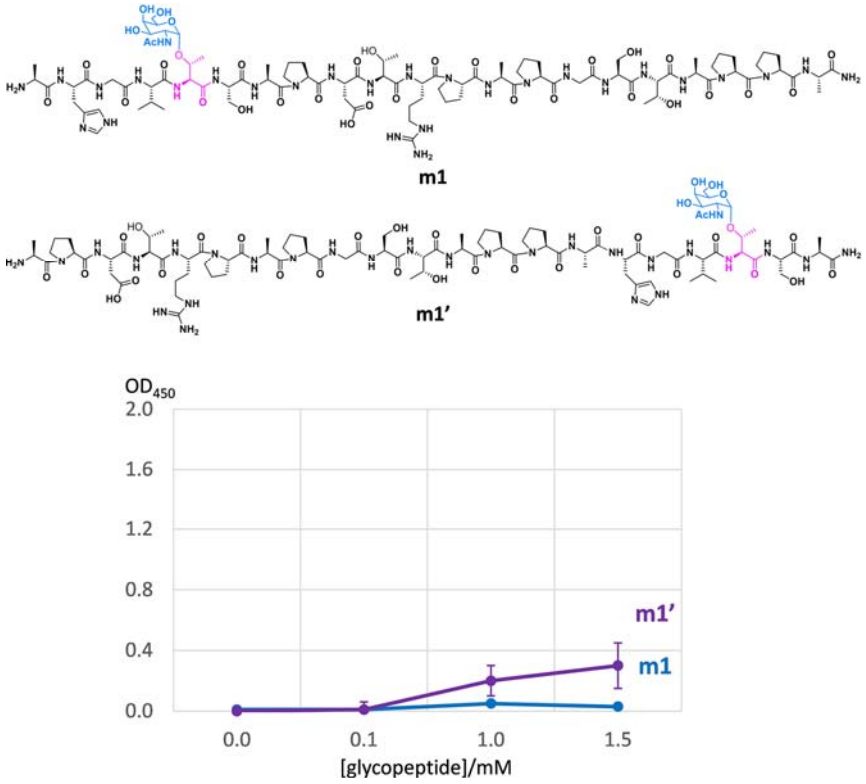


Figure S3. Binding curves for mucin derivatives **m1** and **m1'** with SBA lectin. Absorbance signals are the average of three replicate wells and the error bars show the standard deviations for these measurements. The comparable results obtained for **m1** and **m1'** ruled out the existence of accessibility difficulties in the recognition of glycopeptide **m1** by SBA lectin.

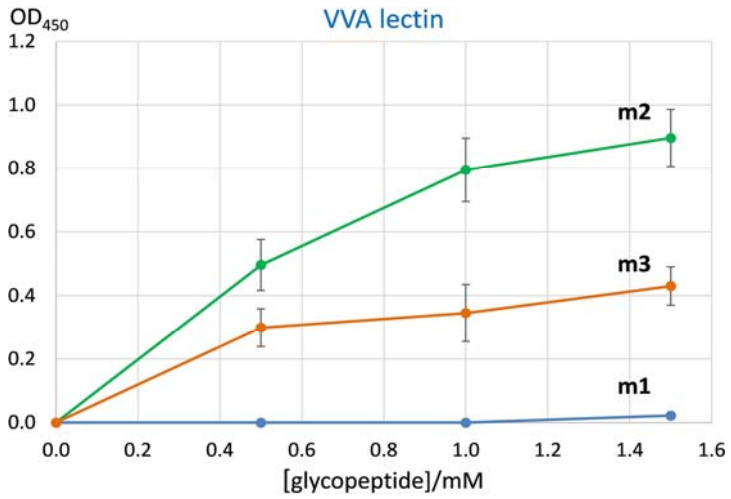


Figure S4. Binding curves for compounds **m1-m3** with VVA lectin. Absorbance signals are the average of three replicate wells and the error bars show the standard deviations for these measurements.

Table S1. Comparison of the experimental and deduced distances from 20 ns MD_{H2O}-tar simulations for glycopeptide **1**. Distances are given in Å.

[a] All distances are given in Å.

	glycopeptide 1	
distance	Exp.	MD_{H2O}-tar
NH_{Thr} – $H\alpha_{Thr}$	3.0	2.9
NH_{Gly} – $H\alpha_{His}$	2.6	2.4
NH_{Ala7} – $H\alpha_{Ala7}$	2.6	2.9
NH_{Ala7} – $H\alpha_{Ser}$	2.2	2.3
NH_{Ser} – $H\alpha_{Ser}$	2.7	2.9
NH_{Ser} – $H\alpha_{Thr}$	2.8	2.5
NH_{Val} – $H\alpha_{Val}$	2.8	2.9
NH_{Ser} – $H\beta_{Thr}$	2.3	2.3

Table S2. Comparison of the experimental and 20 ns MD-tar simulation (in explicit water) derived distances for glycopeptide **2**. Distances are given in Å.

	glycopeptide 2	
distance	Exp.	MD_{H2O}-tar
NH_{Arg} – $H\alpha_{Arg}$	2.7	2.9
NH_{Arg} – $H\alpha_{Thr}$	2.6	2.5
NH_{Thr} – $H\alpha_{Asp}$	2.4	2.6
NH_{Asp} – $H\alpha_{Pro}$	2.2	2.3

Table S3. Comparison of the experimental and 20 ns MD-tar simulation (in explicit water) derived distances for glycopeptide 3. Distances are given in Å.

	glycopeptide 3	
distance	Exp.	MD_{H2O}-tar
NH_{Thr} – $H\alpha_{Thr}$	2.9	2.6
NH_{Gly} – $H\alpha_{Pro2}$	2.2	2.3
NH_{Ala6} – $H\alpha_{Ala6}$	2.8	2.9
NH_{Ala6} – $H\alpha_{Thr}$	2.6	2.6
NH_{Ser} – $H\alpha_{Ser}$	3.0	2.9

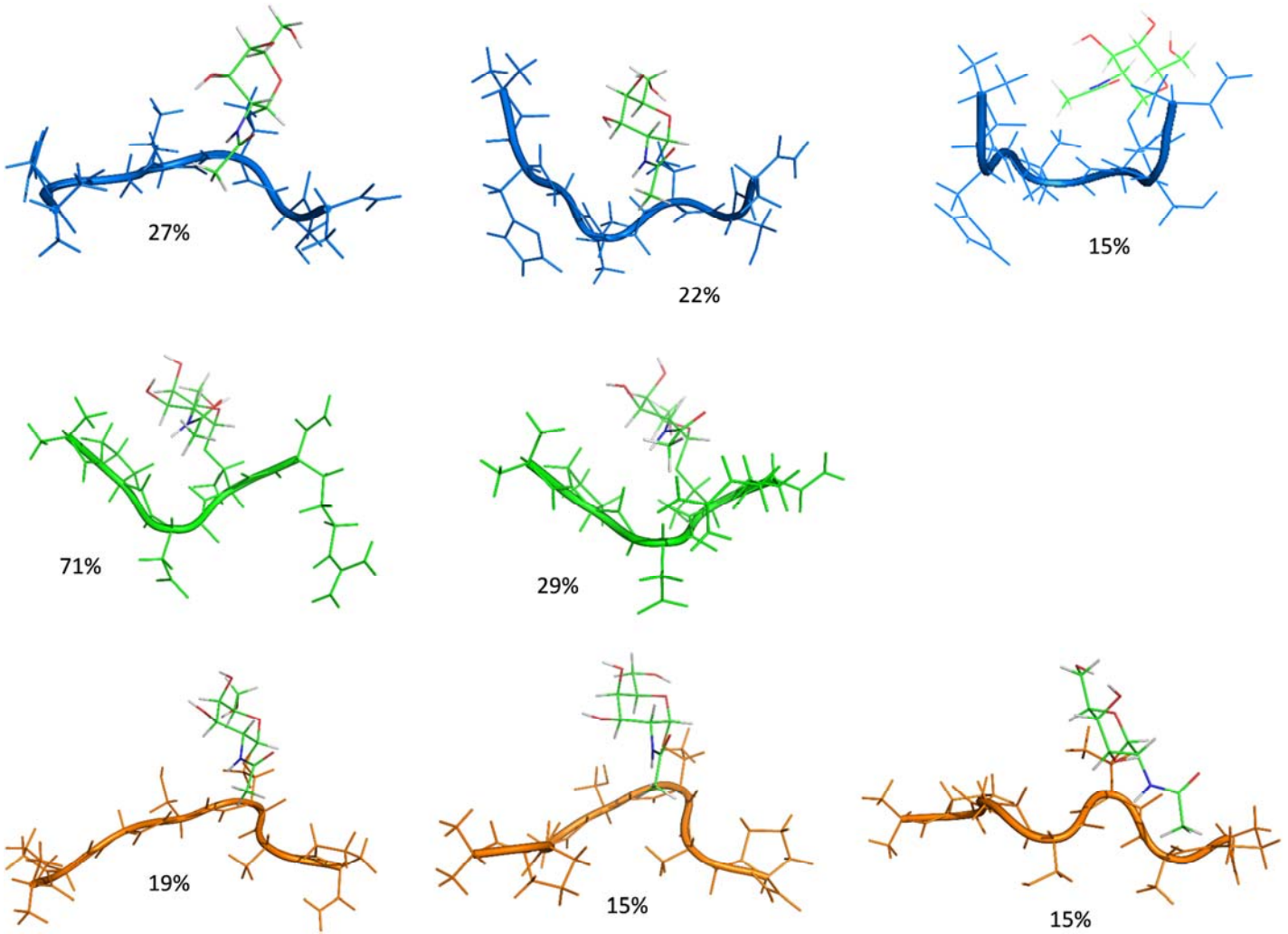


Figure S5. Main conformations in aqueous solution (free-state) for glycopeptides **1** (upper panel), **2** (middle panel) and **3** (lower panel) obtained from the 20 ns MD-tar in explicit water.

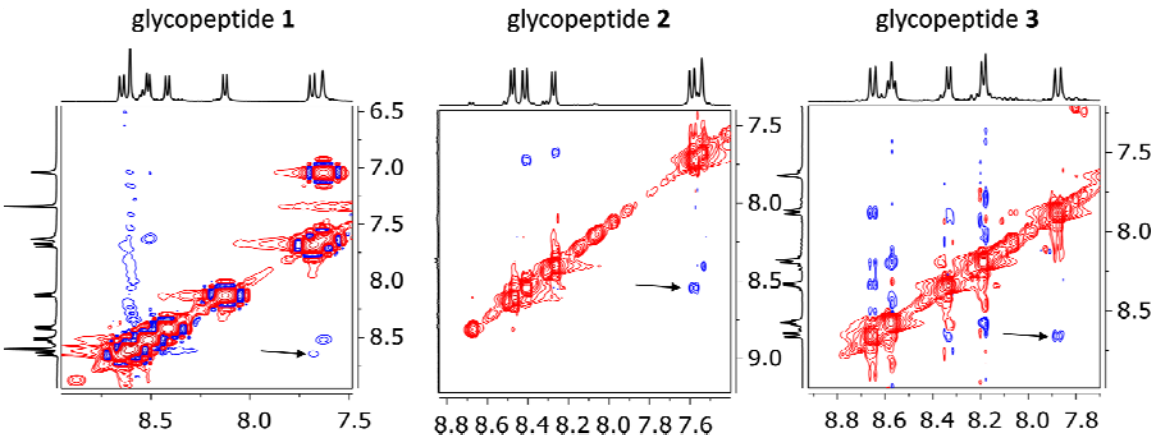


Figure S6. 2D-NOESY (H_2O/D_2O, 9:1, pH 6.5, 293 K) spectra of glycopeptides **1-3**, showing the cross-peak between NH of the GalNAc moiety and NH of the underlying Thr. This NOE is characteristic of an "eclipsed" conformation of the glycosidic linkage.[S13] Diagonal peaks and exchange cross-peaks connecting NH protons and water are negative (red color). The NOE contacts are represented as positive cross-peaks (blue color).

Table S4. Data collection and refinement statistics for SBA:glycopeptide **2** complex. Values in parentheses refer to the highest resolution shell. Ramachandran plot statistics were determined with PROCHECK.

	SBA:glycopeptide 2
Space group	$P3_2$
Wavelength (Å)	0.976
Resolution (Å)	20-2.70 (2.85-2.70)
Cell dimensions (Å)	*a =114.23* *b=114.23* *c* =202.89
Unique reflections	81148
Completeness	99.8 (100)
R_{sym}	0.122 (0.739)
$I/\sigma(I)$	10.8 (2.3)
Redundancy	4.8 (4.3)
R_{work} / R_{free}	0.230/0.274
RMSD from ideal geometry, bonds (Å)	0.011
RMSD from ideal geometry, angles (º)	1.590
<*B*> protein ($Å^2$)	55.76
<*B*> APD-T(GalNAc)-R	59.83
<*B*> solvent ($Å^2$)	26.91
<*B*> Mn^{+2} ($Å^2$)	58.25
Ramachandran plot: Most favoured (%) Additionally allowed (%) Disallowed (%)	 94.27 3.55 2.18

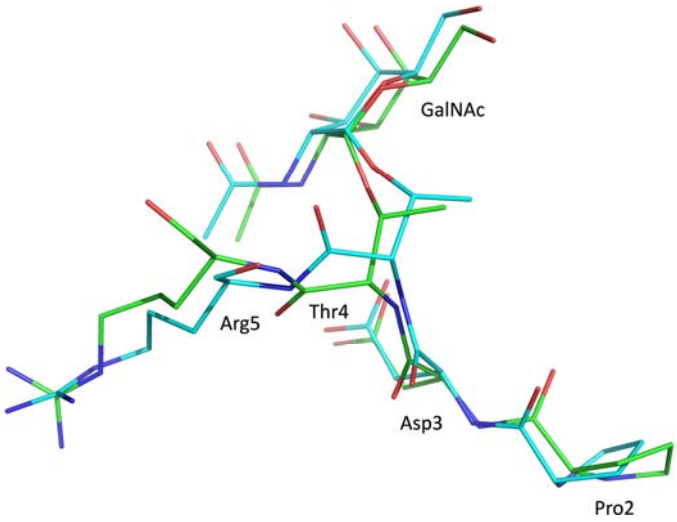


Figure S7. Superposition of the two different conformations found in the X-ray structure of SBA lectin in complex with glycopeptide **2**. The rmsd value for heavy atoms is 0.94 Å. The green structure corresponds to binding mode A and the cyan structure represents the binding mode B.

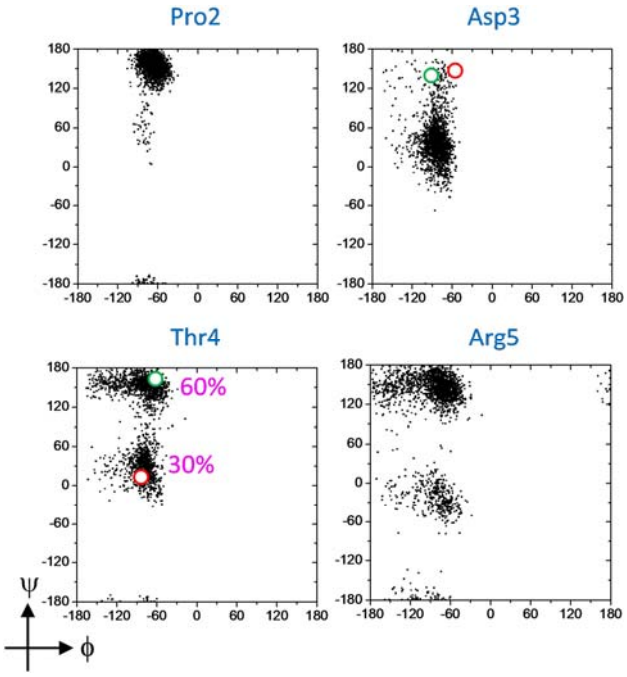


Figure S8. Dihedral angles (ϕ/ψ) distribution for the backbone of glycopeptide **2** obtained from the 20 ns MD-tar simulations performed in explicit water. The ϕ/ψ values found for Asp3 and Thr4 residues of this glycopeptide bound to SBA and found in the X-ray structure are shown as red (binding mode A) or green (binding mode B) circles.

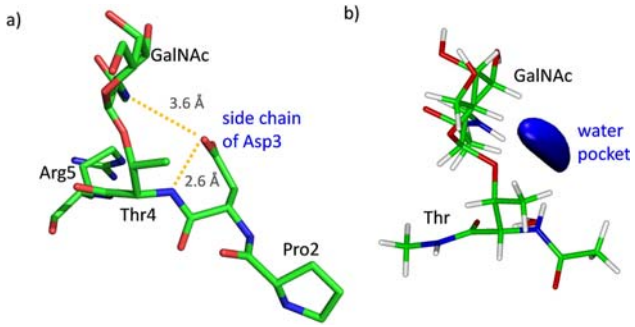


Figure S9. a) Hydrogen bond network found in the X-ray structure for Asp3 with sugar and peptide moieties in both binding modes. b) Inter-residue water pocket deduced from the MD-tar simulations on Ac-L-Thr(α-D-GalNAc)-NHMe.

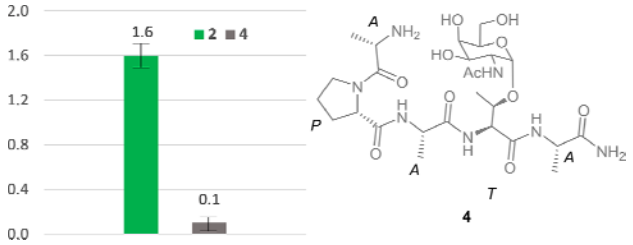


Figure S10. Binding studies of glycopeptides **2** and **4** with SBA lectin using ELLA tests. Absorbance signals are the average of three replicate wells and the error bars show the standard deviations for these measurements.

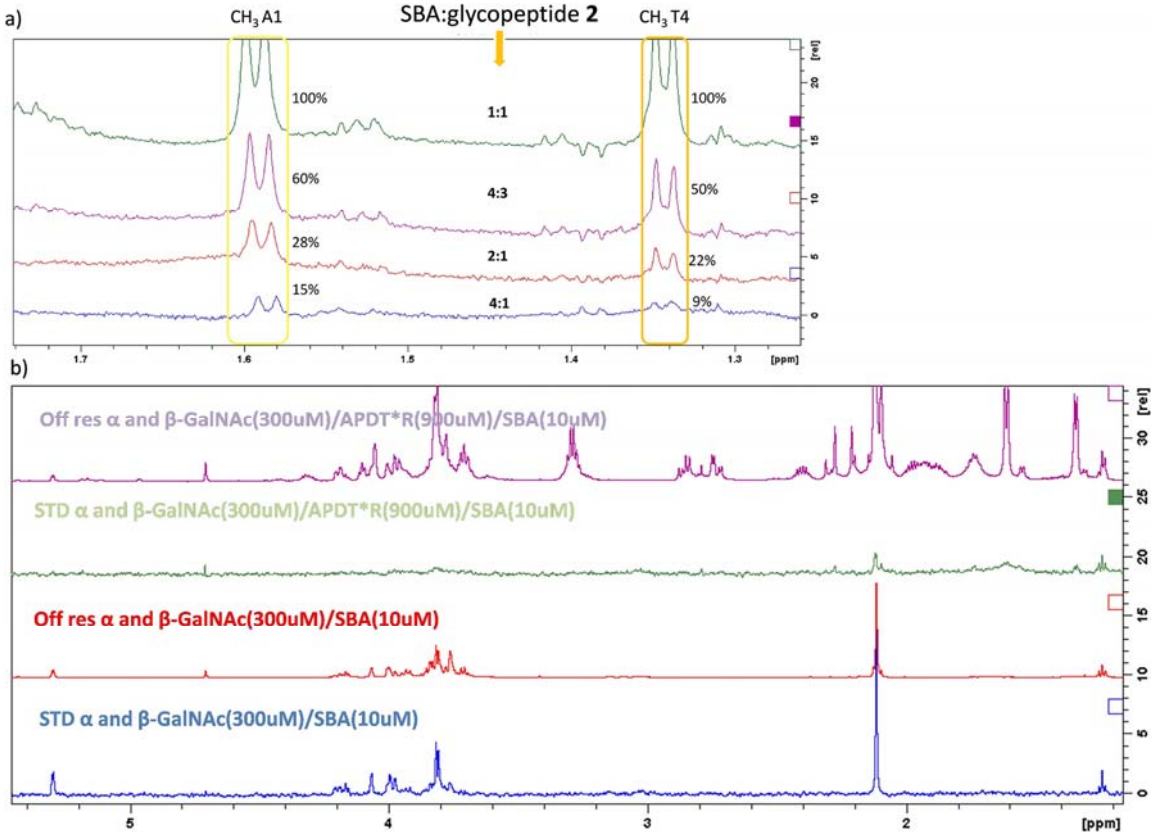


Figure S11. a) ^{1}H NMR spectra of titration experiments performed with a solution of SBA lectin (50 μM) and increasing amounts of glycopeptide **2**. The experiments were carried out at 310 K and at pH= 7.2 (phosphate buffer). The fact that the CH_3 protons of Thr4 do not increase as fast as CH_3 protons of Ala1 supports the idea that Thr4 exhibits a stronger interaction with the protein. b) STD competition binding experiments between α and β-GalNAc *vs*. glycopeptide **2** at 310 K.

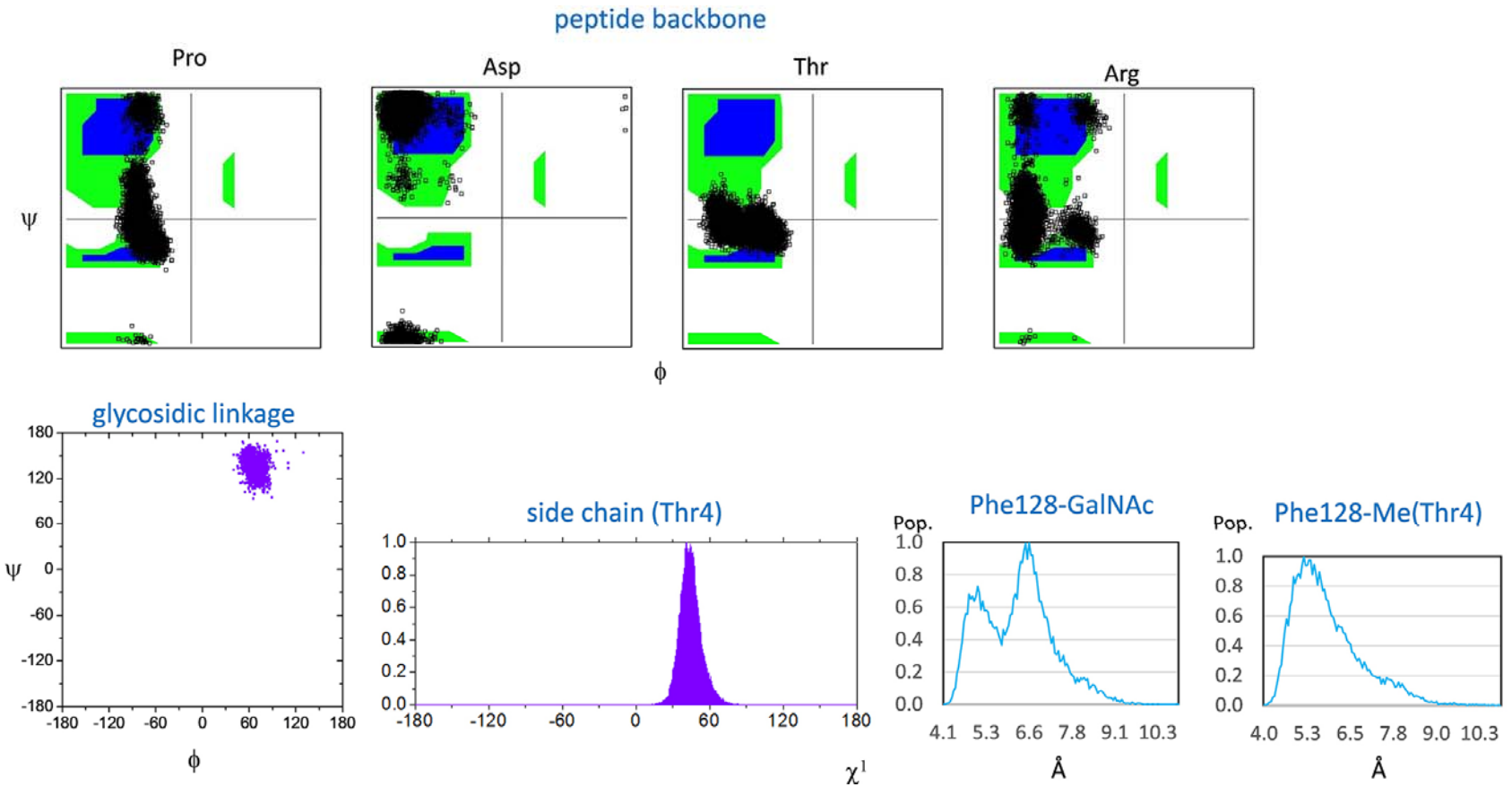


Figure S12. Dihedral angles (ϕ/ψ) distribution for the backbone, glycosidic linkage and side chain (χ^1) of glycopeptide **2** bound to SBA lectin obtained from the unrestrained 100 ns MD-tar simulations performed in explicit water. The binding mode A was used as starting structure in these simulations. Some relevance distances to monitor the interaction between the Phe128 residue of SBA and glycopeptide **2** are also shown. These distances were calculated from the center of the aromatic ring.

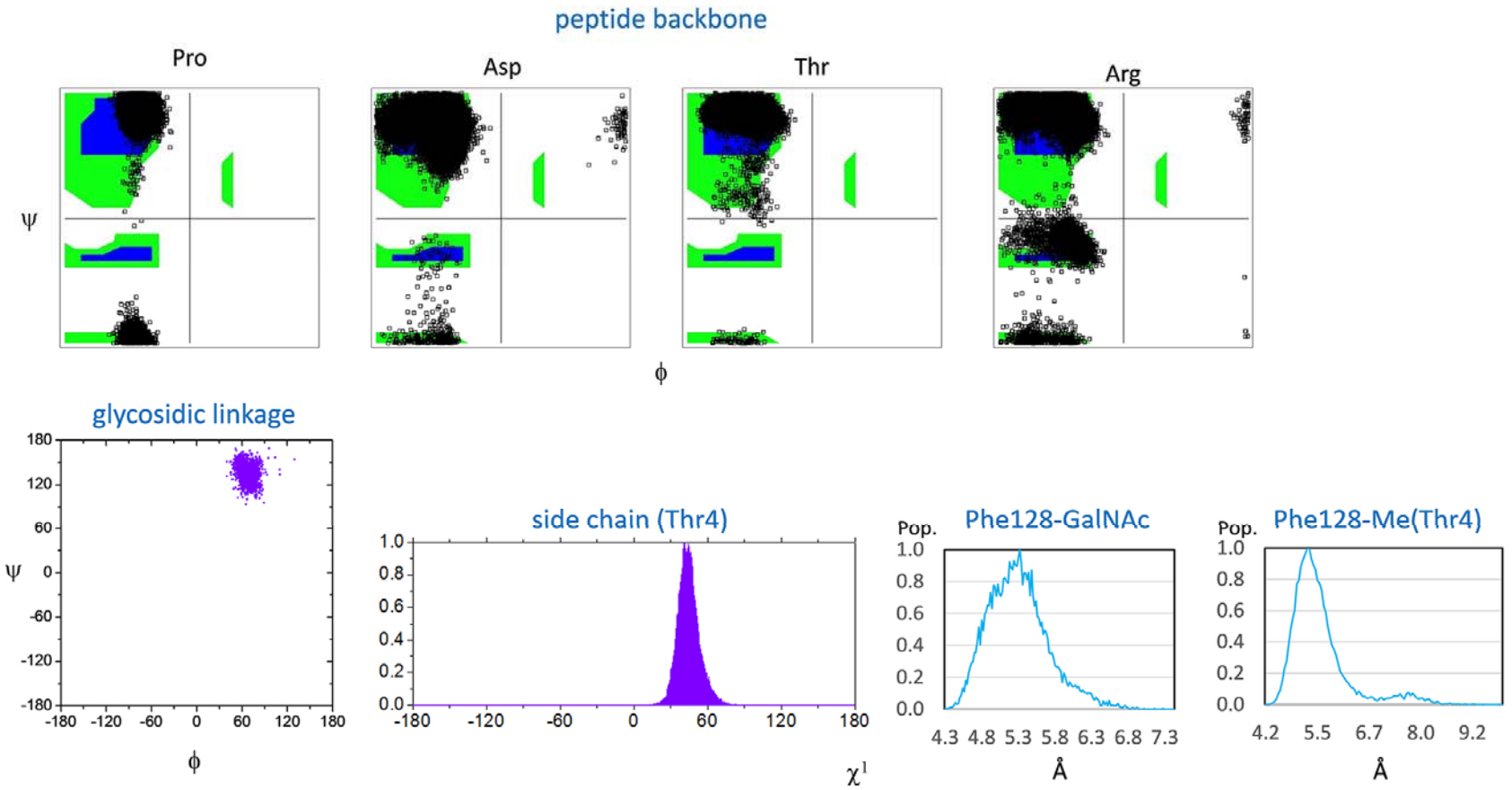


Figure S13. Dihedral angles (ϕ/ψ) distribution for the backbone, glycosidic linkage and side chain (χ^1) of glycopeptide **2** bound to SBA lectin obtained from the unrestrained 100 ns MD-tar simulations performed in explicit water. The binding mode B was used as starting structure in these simulations. Some relevance distances to monitor the interaction between the Phe128 residue of SBA and the glycopeptide **2** are also shown. These distances were calculated from the center of the aromatic ring.

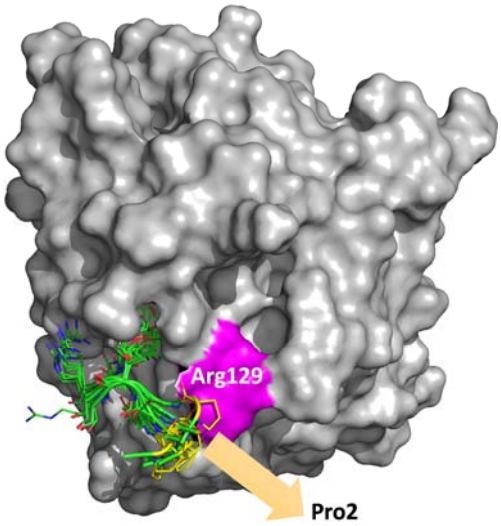


Figure S14. Ensembles obtained from the unrestrained 100 ns MD simulations carried out on SBA:glycopeptide **2** complex (starting from binding mode A). The MD simulations indicate that Pro2 establishes a hydrophobic interaction with the lateral chain of Arg129. This interaction, not observed in the X-ray structure, could be responsible for the STD effect found in Pro2 residue.

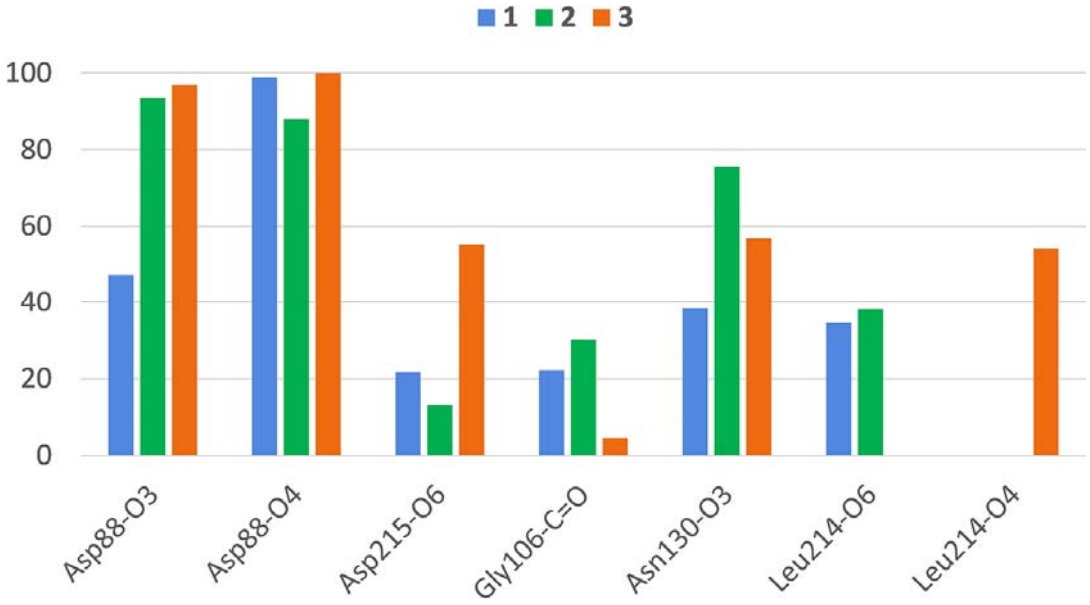


Figure S15. Population of the intramolecular hydrogen bonds found between GalNAc moiety of glycopeptides **1**-**3** complexed to SBA obtained from the unrestrained 100 ns MD simulations in explicit water.

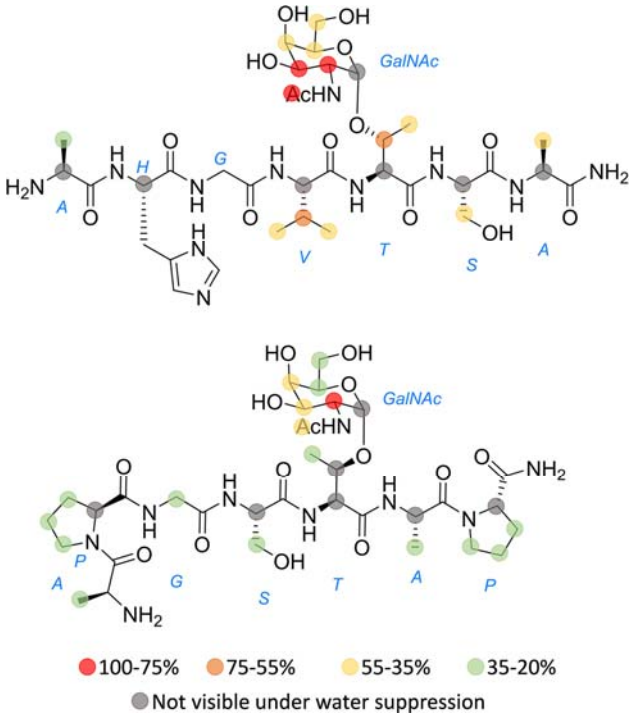


Figure S16. Epitope mapping of glycopeptides **1** (upper pannel) and **3** (lower pannel) complexed to SBA lectin deduced from STD experiments (see Methods in the manuscript).

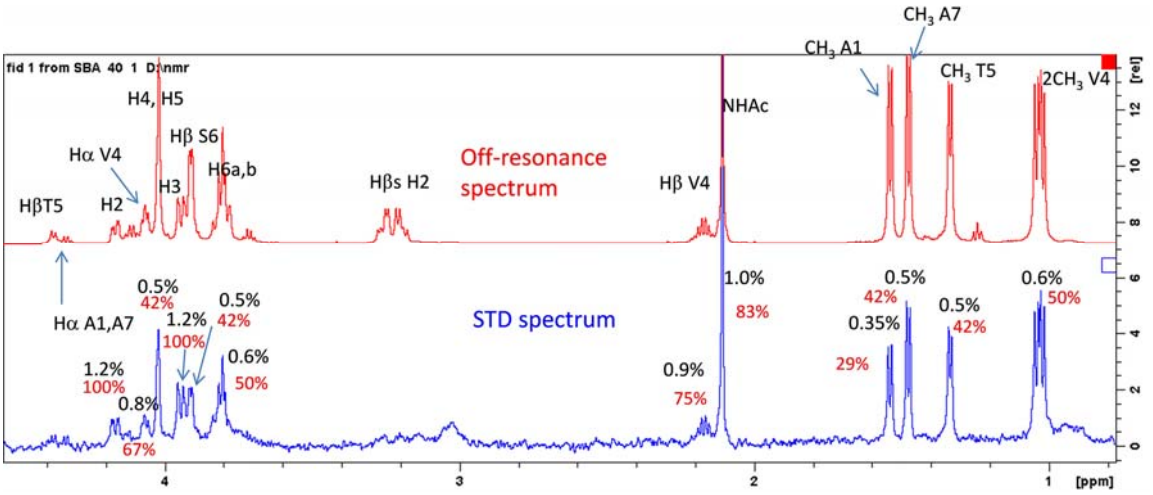


Figure S17. ^{1}H NMR 600 MHz spectra for glycopeptide **1** (450µM concentration) in the presence of SBA (15 µM) at 310 K. The reference spectrum (off res) is displayed in red and the STD spectrum (STD) in blue. The key proton resonances are showed in the figure. The red values represent the epitope mapping and the back ones are the absolute STD values.

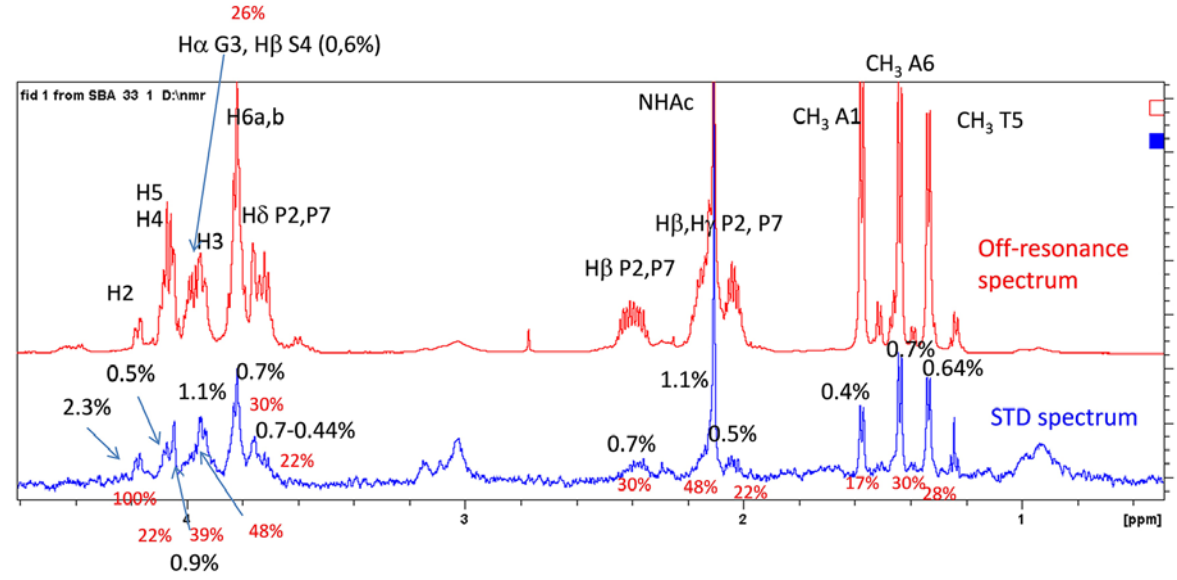


Figure S18. ^{1}H NMR 600 MHz spectra for glycopeptide **3** (450 µM concentration) in the presence of SBA (15 µM) at 310 K. The reference spectrum (off res) is displayed in red and the STD spectrum (STD) in blue. The key proton resonances are showed in the figure. The red values represent the epitope mapping and the back ones are the absolute STD values.

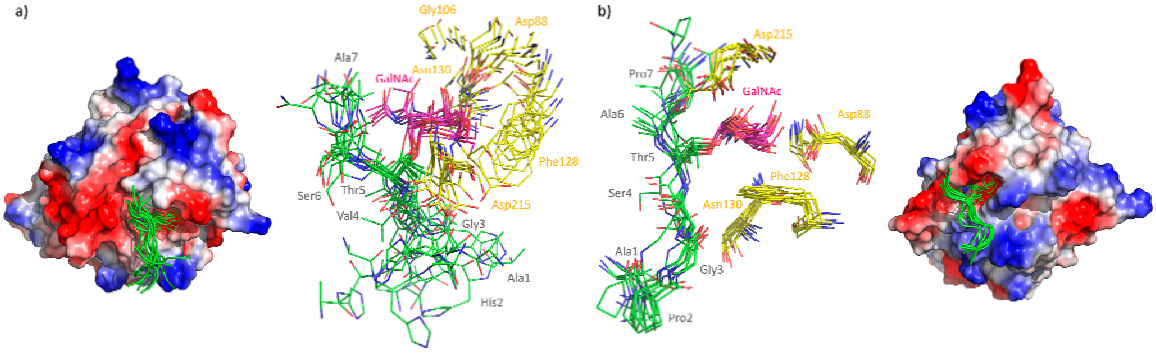


Figure S19. a) Ensembles obtained from unrestrained 100 ns MD simulations performed on SBA:**1** complex, showing the protein residues that interact with glycopeptide **1**. b) Ensembles obtained from unrestrained 100 ns MD simulations performed on SBA:**3** complex, showing the protein residues that interact with glycopeptide **3**. Carbon atoms of the protein are in yellow. Carbon atoms of GalNAc moiety are in hot pink and carbon atoms of the glycopeptides **1** and **3** are in green. Electrostatic potential surface of the protein is also represented for both complexes and was calculated using standard parameters of Pymol 1.3. Red color represents a net negative surface potential while blue color denotes a net positive surface potential. Glycopeptides are shown as a stick models colored according to atom type.

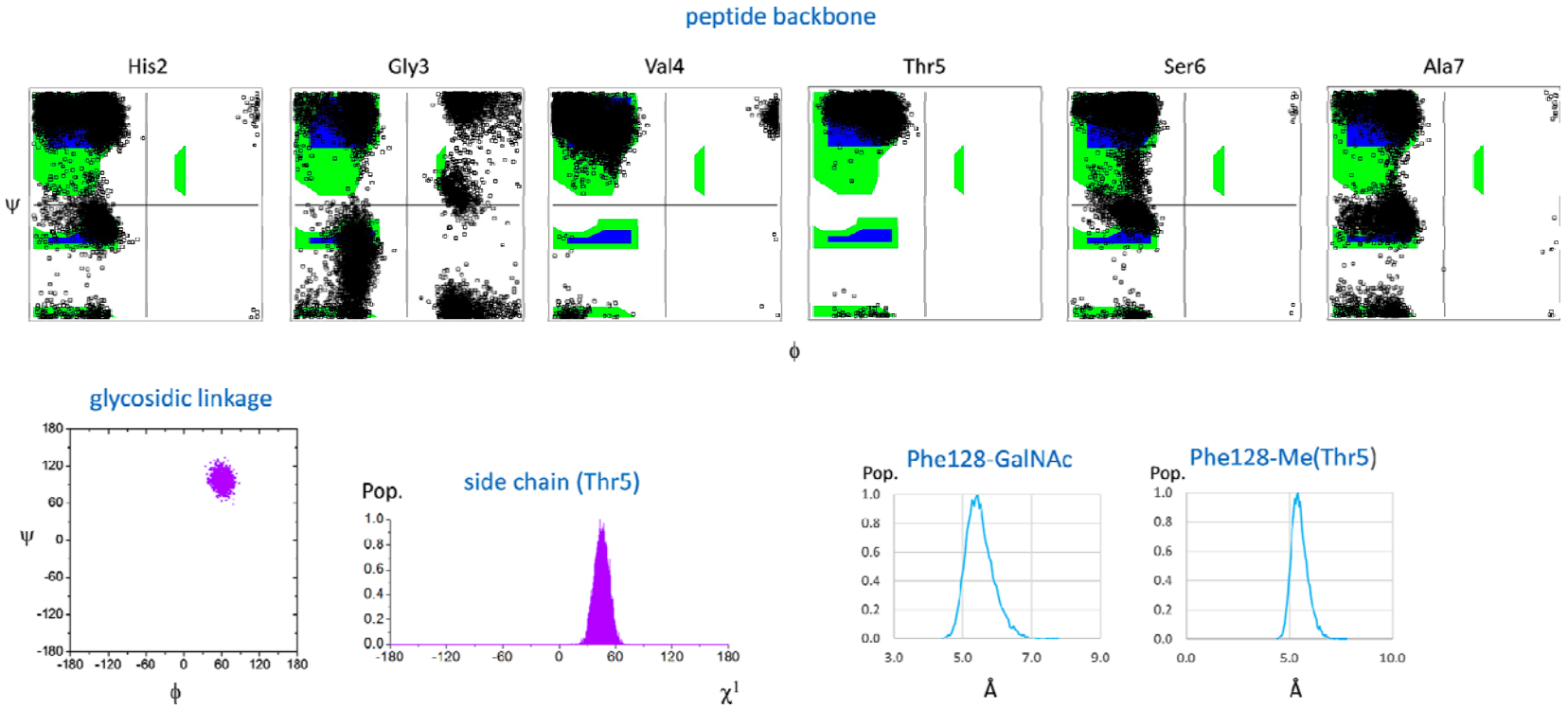


Figure S20. Dihedral angles (ϕ/ψ) distribution of the peptide backbone obtained from the 100 ns MD simulations in explicit water carried out on SBA:**1** complex, together with the geometry of the glycosidic linkage and the side chain of Thr5. Some relevance distances to monitor the interaction between the Phe128 residue of SBA and glycopeptide **1** are also shown. These distances were calculated from the center of the aromatic ring.

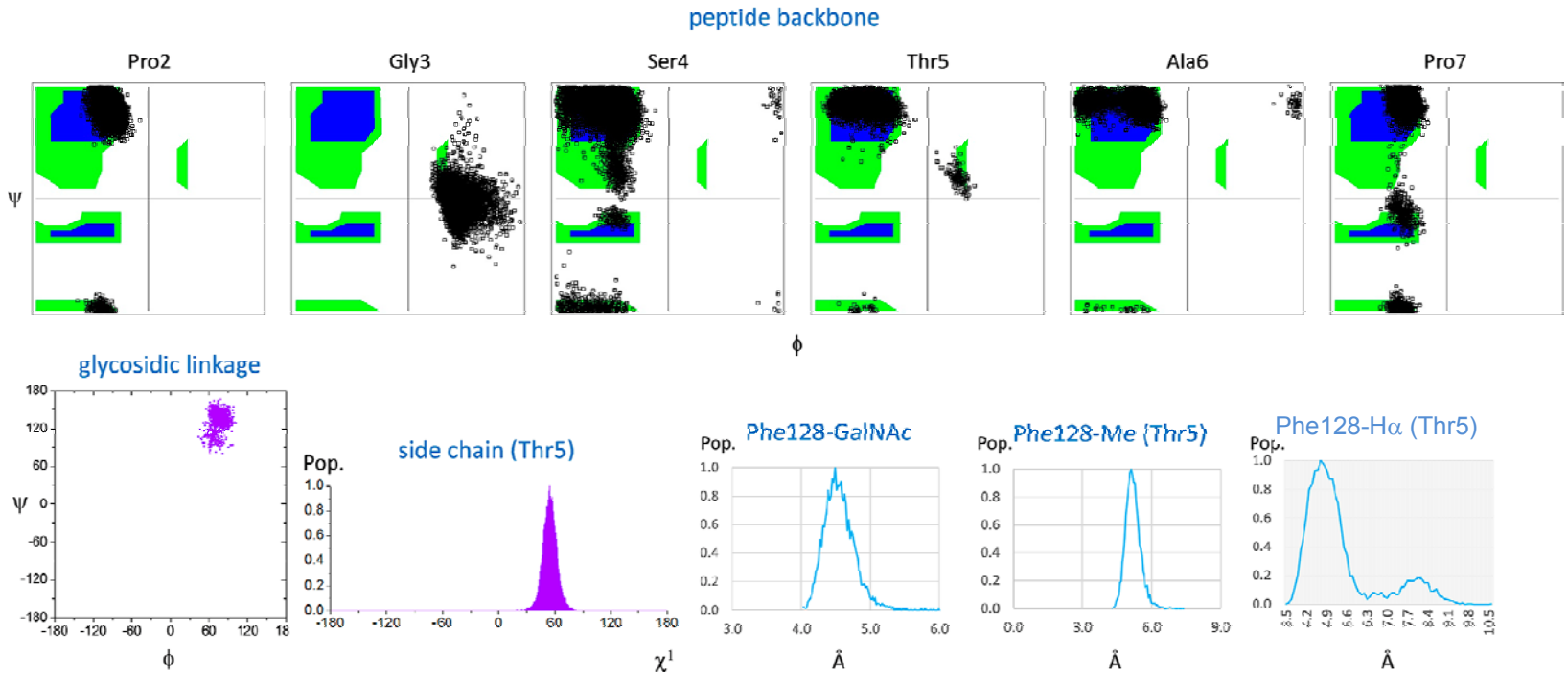


Figure S21. Dihedral angles (ϕ/ψ) distribution of the peptide backbone obtained from the 100 ns MD simulations in explicit water carried out on SBA:**3** complex, together with the geometry of the glycosidic linkage and the side chain of Thr4. Some relevance distances to monitor the interaction between the Phe128 residue of SBA and glycopeptide **3** are also shown. These distances were calculated from the center of the aromatic ring.

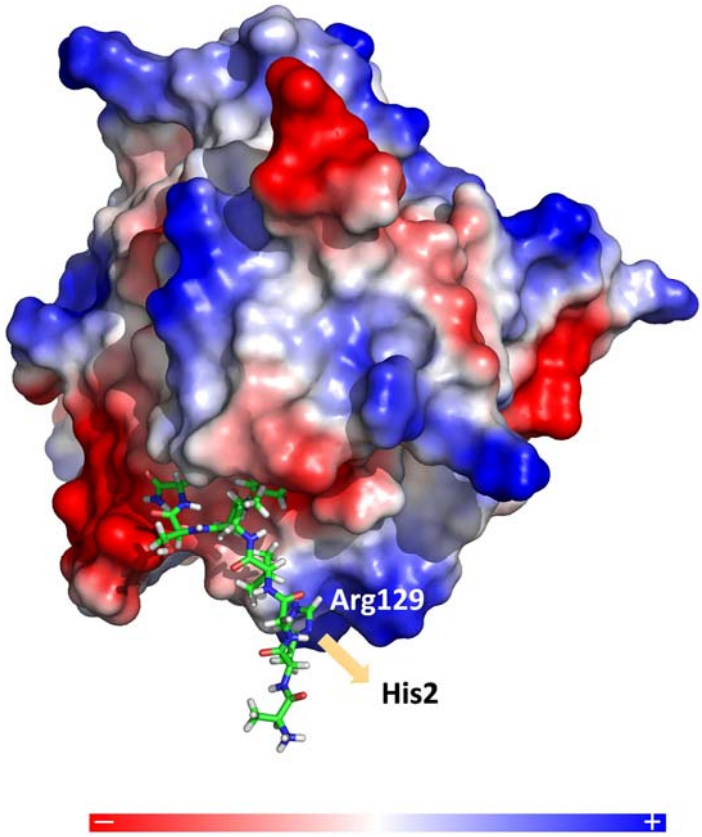


Figure S22. Representative frame obtained from the unrestrained 100 ns MD simulations on SBA:**1**. The surface potential was calculated using the standard parameters of Pymol 1.3. Red color represents a net negative surface potential while a blue color denotes a net positive surface potential. Glycopeptide **1** is shown as a stick model colored according to atom type. Positive charge or Arg129 is close to the positive charged amino acids (Ala1 and His2). This may be responsible (among other factors) for the low affinity observed between glycopeptide **1** and SBA lectin.

References

(S1) Plattner, C., Höfener, M., and Sewald, N. (2011) One-pot azidochlorination of glycals, *Org. Lett. 13*, 545–547.

(S2) Case, D., Darden, T. A., Cheatham, T. E., Simmerling, C., Wang, J., Duke, R.; Luo, R., Crowley, M., Walker, R., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wu, X., Brozell, S., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., and Kollman, P. Amber 11. University of California, San Francisco.

(S3) Wang, J., Cieplak, P., and Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?, *J. Comput. Chem. 21*, 1049–1074.

(S4) Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008). GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* 29, 622–655.

(S5) Duan, Y., Wu, C., Chowdhury, S., Lee, M.C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., and Lee, T. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem. 24*, 1999–2012.

(S6) Kabsch, W. (2014) Processing of X-ray snapshots from crystals in random orientations, *Acta Crystallogr. D Biol. Crystallogr. 70*, 2204–2216.

(S7) 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. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R.

J., Vagin, A., and Wilson, K. S. (2011) Overview of the *CCP*4 suite and current developments, *Acta Crystallogr. D Biol. Crystallogr. 67*, 235–242.

(S8) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) *PHENIX*: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr. D Biol. Crystallogr. 66*, 213–221.

(S9) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D Biol. Crystallogr. 60*, 2126–2132.

(S10) Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallogr. D Biol. Crystallogr. 67*, 355–367.

(S11) Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *PROCHECK*: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283–291.

(S12) Dziadek, S., Griesinger, C., Kunz, H., and Reinscheid, U. M. (2006) Synthesis and structural model of an α(2,6)-sialyl-T glycosylated MUC1 eicosapeptide under physiological Conditions, *Chem. Eur. J. 12*, 4981-4993.

(S13) Corzana, F., Busto, J. H., Jiménez-Osés, G., García de Luis, M., Asensio, J. L., Jiménez-Barbero, J., Peregrina, J. M., and Avenoza, A. (2007) Serine versus threonine glycosylation: The methyl group causes a drastic alteration on the carbohydrate orientation and on the surrounding water shell, *J. Am. Chem. Soc. 129*, 9458-9467.