

Supporting Information

for

The use of fluoroproline in MUC1 antigen enables efficient detection of antibodies in patients with prostate cancer

submitted to *The Journal of the American Chemical Society* by:

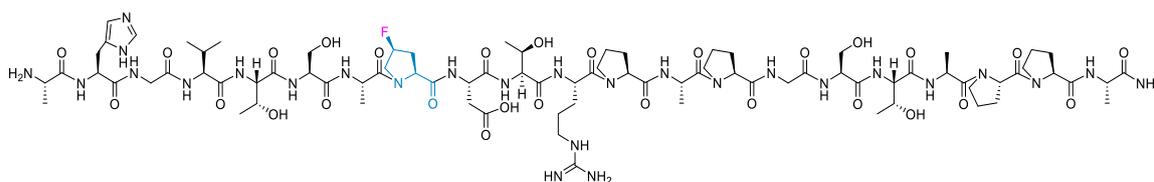
Víctor J. Somovilla, Iris A. Bermejo, Inês S. Albuquerque, Nuria Martínez-Sáez, Jorge Castro-López, Fayna García-Martín, Ismael Compañón, Hiroshi Hinou, Shin-Ichiro Nishimura, Jesús Jiménez-Barbero, Juan L. Asensio, Alberto Avenoz, Jesús H. Busto, Ramón Hurtado-Guerrero, Jesús M. Peregrina, Gonçalo J. L. Bernardes, and Francisco Corzana

Experimental

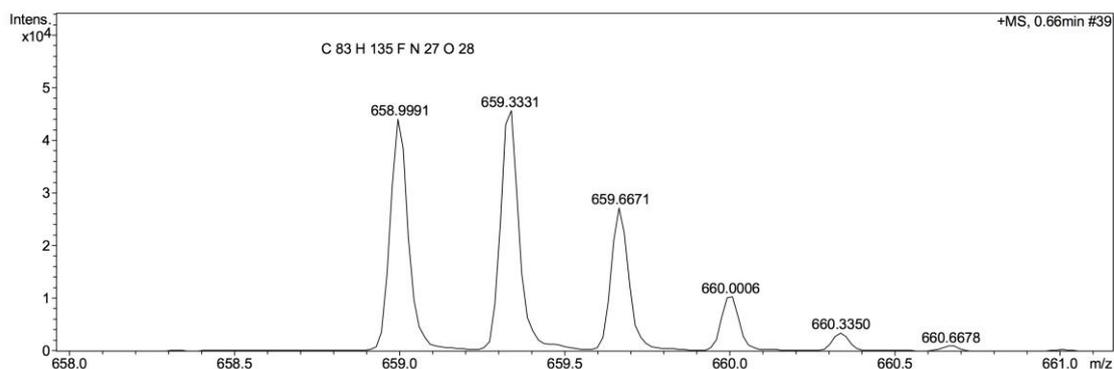
Reagents and general procedures. Commercial reagents were used without further purification. Analytical thin layer chromatography (TLC) was performed on Macherey-Nagel precoated aluminium sheets with a 0.20 mm thickness of silica gel 60 with fluorescent indicator UV254. TLC plates were visualized with 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 silicagel (230–400 mesh). ^1H and ^{13}C NMR spectra were measured with a 400 MHz or a 300 MHz spectrometer, using the H_2O residual as reference. 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 experiments. All NMR chemical shifts (δ) were recorded in ppm and coupling constants (J) were reported 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). All (glyco)peptides were synthesized by a stepwise micro-wave assisted solid-phase peptide synthesis on a Liberty Blue synthesizer using the Fmoc strategy on Rink Amide MBHA resin (0.1 mmol). The glycosylated amino acid building block (2.0 equiv) was synthesized as described in the literature^[S1] and manually coupled using HBTU [(2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate)], while the other Fmoc amino acids (5.0 equiv) were automatically coupled using oxyma pure/DIC (*N,N*-diisopropylcarbodiimide). The *O*-acetyl groups of $(\text{AcO})_3\text{GalNAc}$ moiety were removed in a mixture of $\text{NH}_2\text{NH}_2/\text{MeOH}$ (7:3). (Glyco)peptides were released from the resin, and all acid sensitive side-chain protecting groups were simultaneously removed using TFA 95%, TIS (triisopropylsilane) 2.5% and H_2O 2.5%, followed by precipitation with cold diethyl ether. The crude (glyco)peptides were purified by semi-preparative HPLC on a Phenomenex Luna C18(2) column (10 μ , 250 mm \times 21.2 mm) and a dual absorbance detector, with a flow rate of 10 or 20 mL/min.

Peptide fP

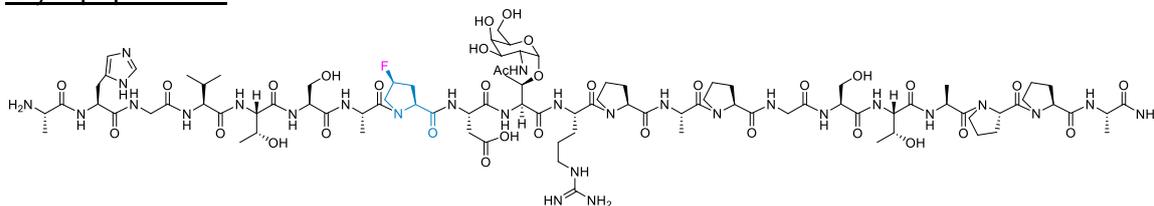


HRMS (ESI+) m/z: calcd. for C₈₃H₁₃₂FN₂₇O₂₈: [M+3H]³⁺: 658.9979 found: 658.9991.

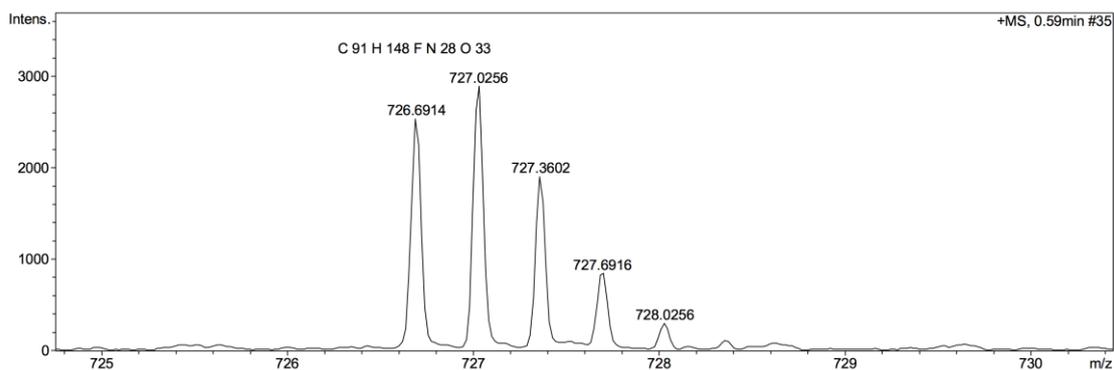


HPLC: R_t = 15.5 min, (Grad: acetonitrile/water+0.1% TFA (10:90) → (20:80), 18 min, 10 mL/min λ = 212 nm).

Glycopeptide fP*

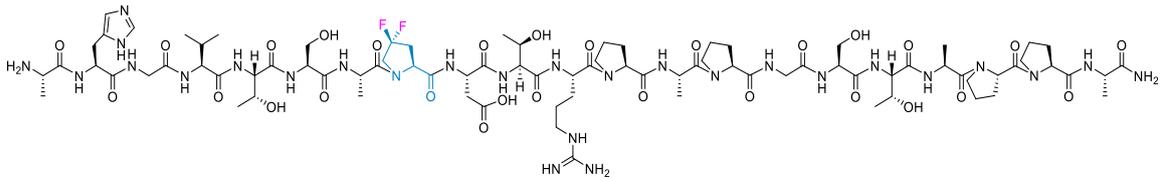


HRMS (ESI+) m/z: calcd. for C₉₁H₁₄₅FN₂₈O₃₃: [M+3H]³⁺: 726.6838 found: 726.6871.

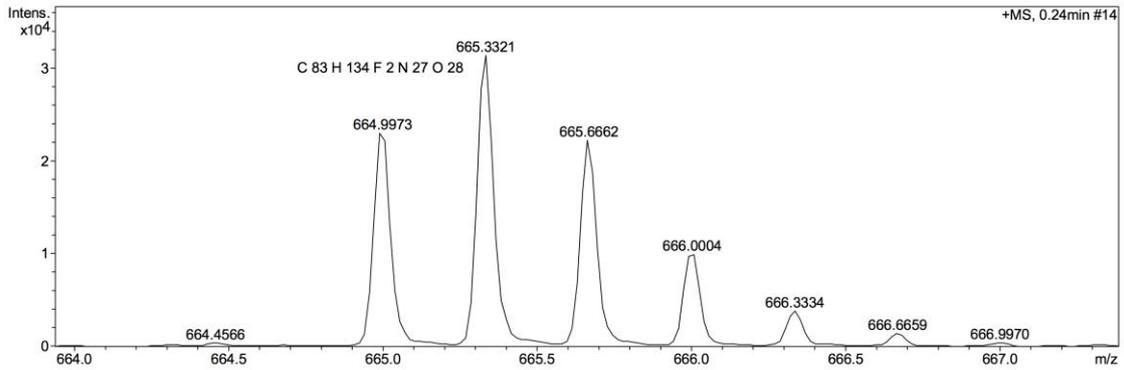


HPLC: R_t = 7.9 min (Grad: acetonitrile/water+0.1% TFA (11:89) → (15:85), 9 min, 20 mL/min, λ = 212 nm).

Peptide 2fP

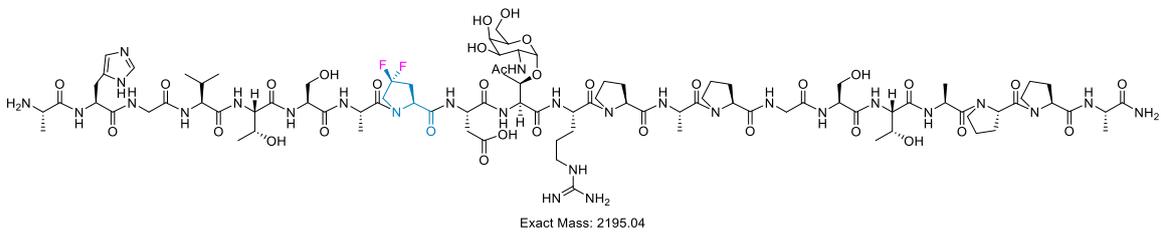


HRMS (ESI+) m/z : calcd. for $C_{83}H_{131}F_2N_{27}O_{28}$: $[M+3H]^{3+}$: 664.9948 found: 664.9973.

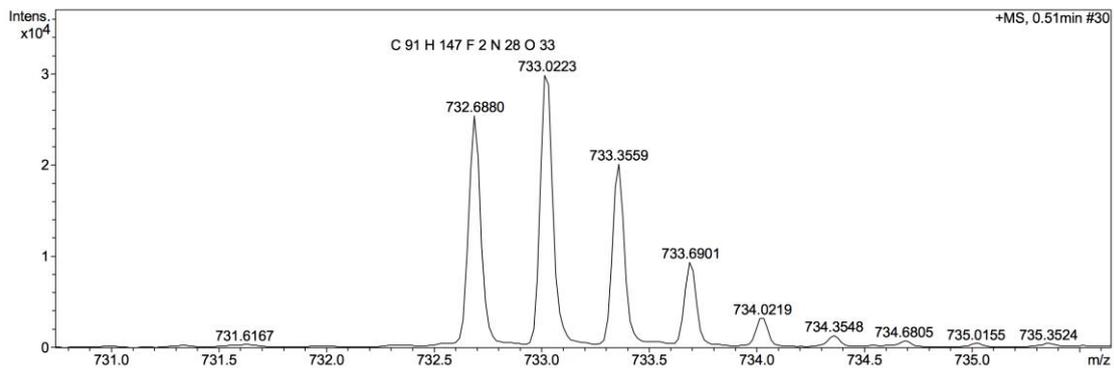


HPLC: $R_t = 6.6$ min (Grad: acetonitrile/water+0.1% TFA (13:87) \rightarrow (19:81), 8 min, 20 mL/min, $\lambda = 212$ nm).

Glycopeptide 2fP*

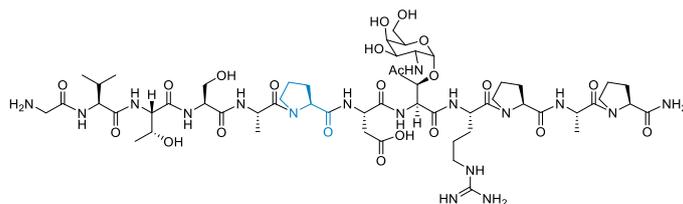


HRMS (ESI+) m/z : calcd. for $C_{91}H_{144}F_2N_{28}O_{33}$: $[M+3H]^{3+}$: 732.6806 found: 732.6880.

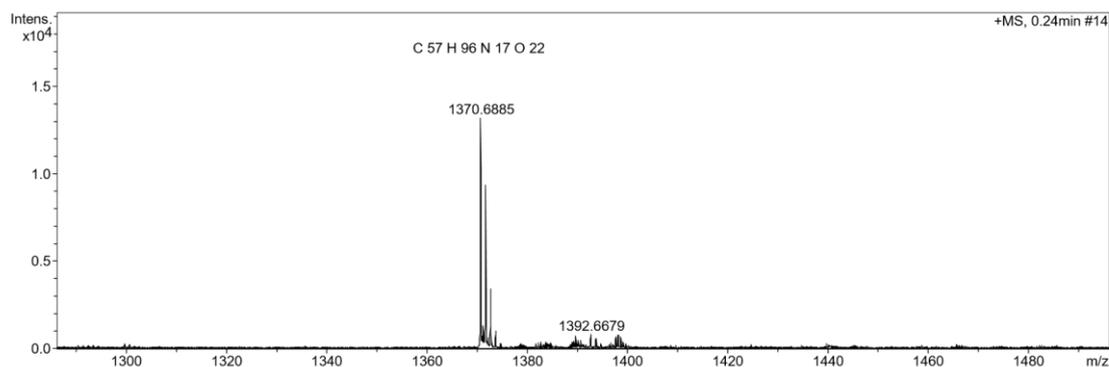


HPLC: $R_t = 7.9$ min (Grad: acetonitrile/water+0.1% TFA (12:88) \rightarrow (17:83), 9 min, 20 mL/min, $\lambda = 212$ nm).

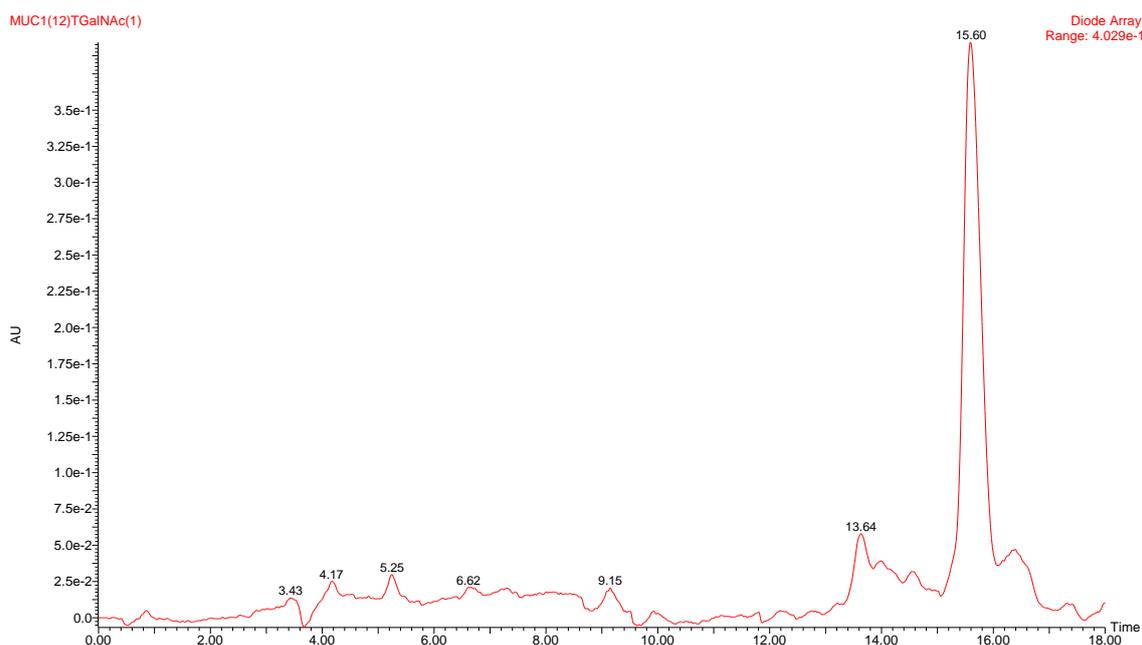
Glycopeptide P³



HRMS (ESI+) m/z: calcd. for C₅₇H₉₅N₁₇O₂₂: [M+H]: 1370.6838 found: 1370.6885.

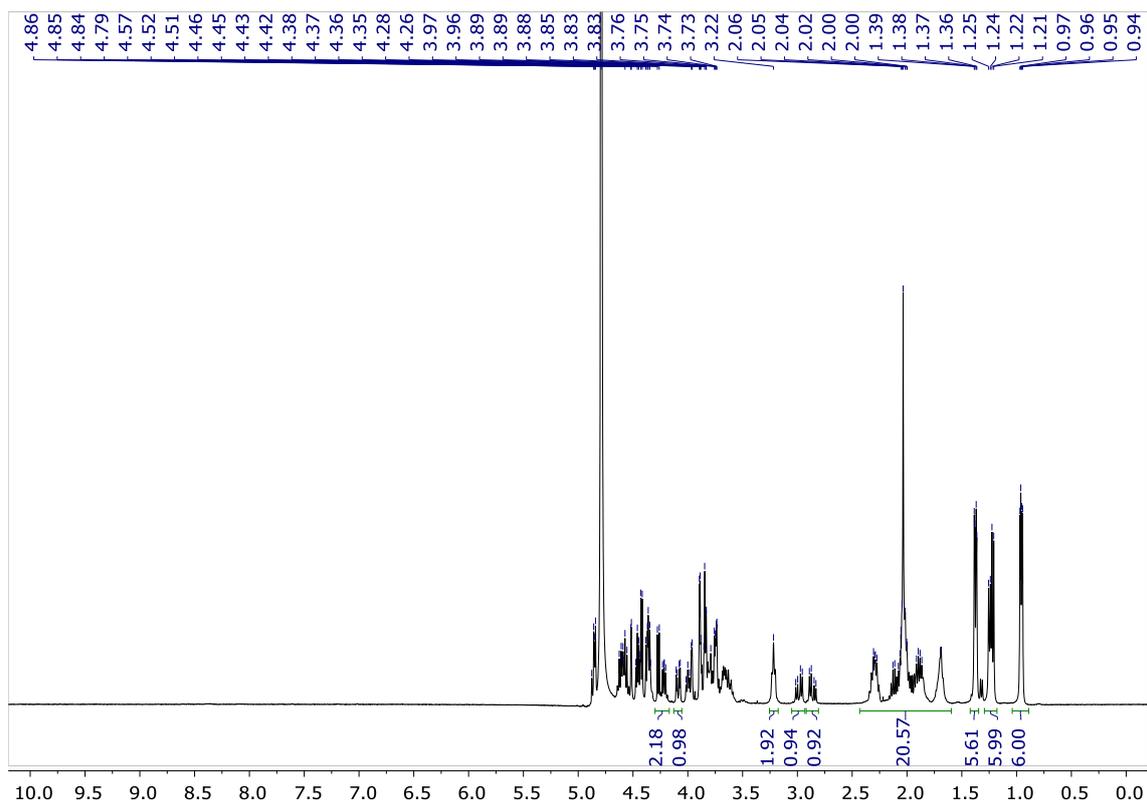


HPLC: R_t = 15.60 min (Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ = 212 nm).

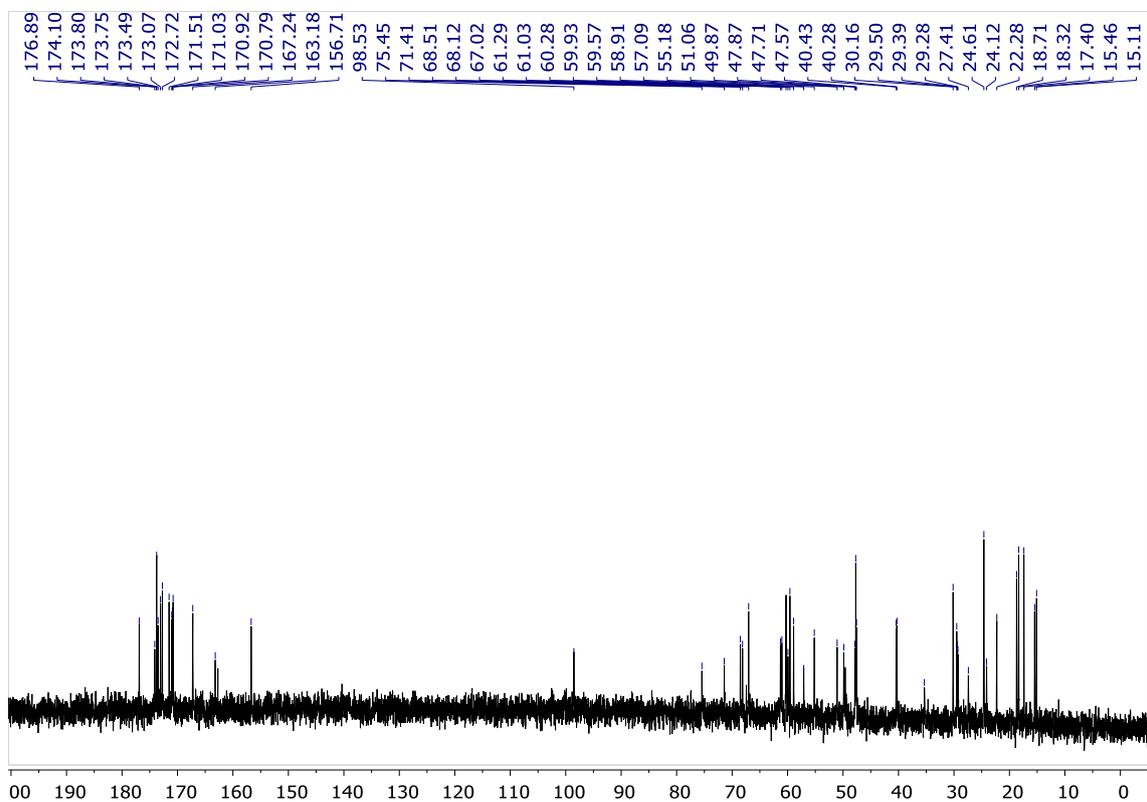


¹H NMR (400 MHz, D₂O) δ (ppm): 0.95-0.97 (m, 6H, Val²γ), 1.21-1.26 (m, 6H, Thr³γ, Thr⁸γ), 1.31-1.41 (m, 6H, Ala⁵β, Ala¹¹β), 1.69-2.34 (m, 20H, Ac, Pro⁶β, Pro¹⁰β, Pro¹²β, Pro⁶γ, Pro¹⁰γ, Pro¹²γ, Arg⁹β, Arg⁹γ, Val²β), 2.83-3.04 (m, 2H, Asp⁷β), 3.17-3.27 (m, 2H, Gly¹α).

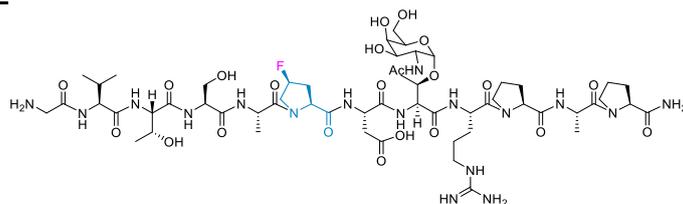
^1H NMR 400 MHz in D_2O registered at 298K



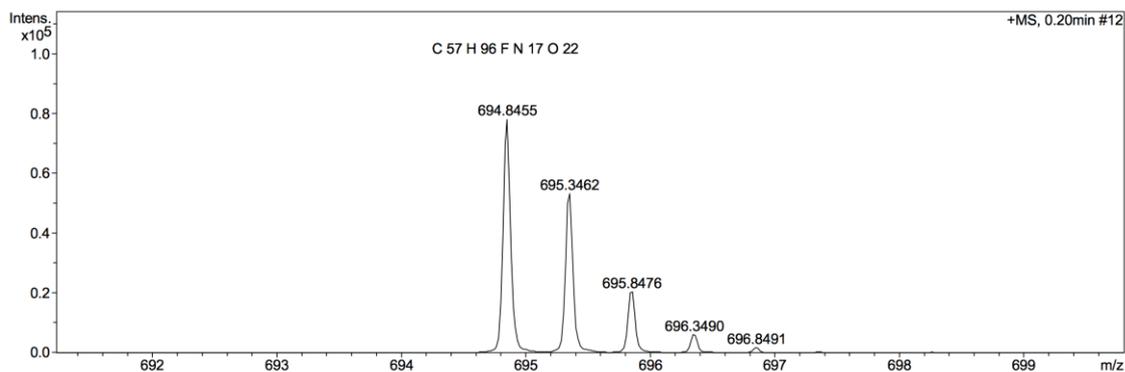
^{13}C NMR 75 MHz in D_2O registered at 298K



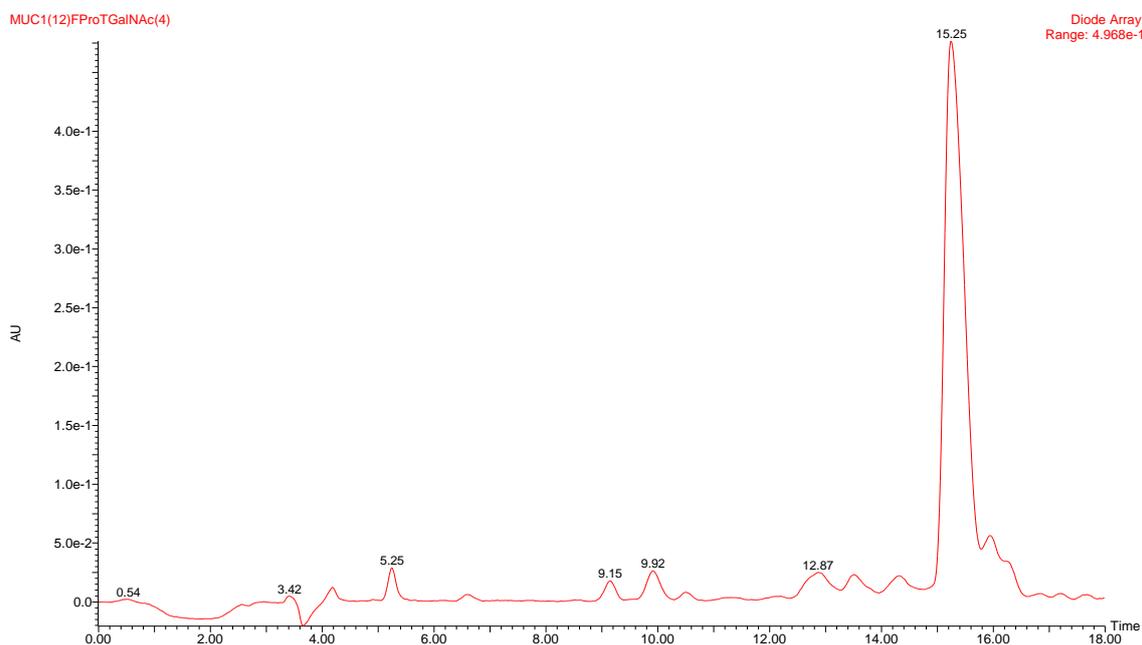
Glycopeptide fP*



HRMS (ESI+) m/z: calcd. for C₅₇H₉₄FN₁₇O₂₂: [M+2H]²⁺: 694.8444 found: 694.8455.



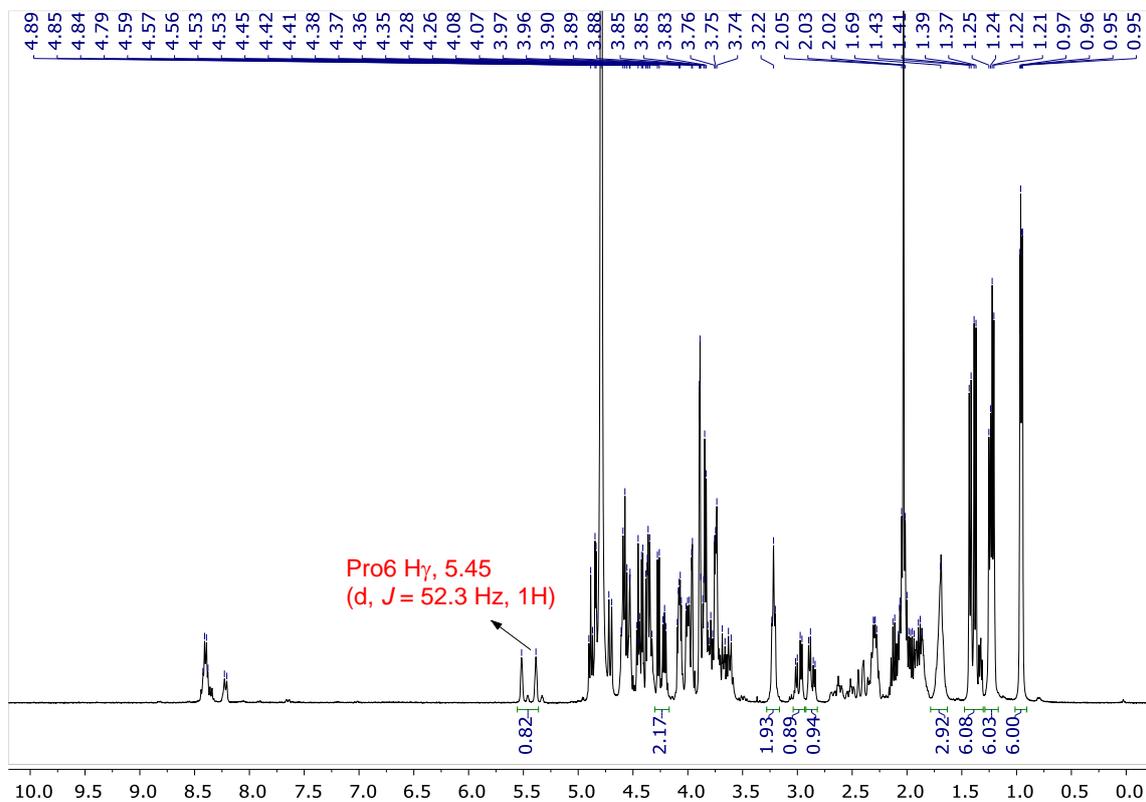
HPLC: R_t = 15.25 min (Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ = 212 nm).



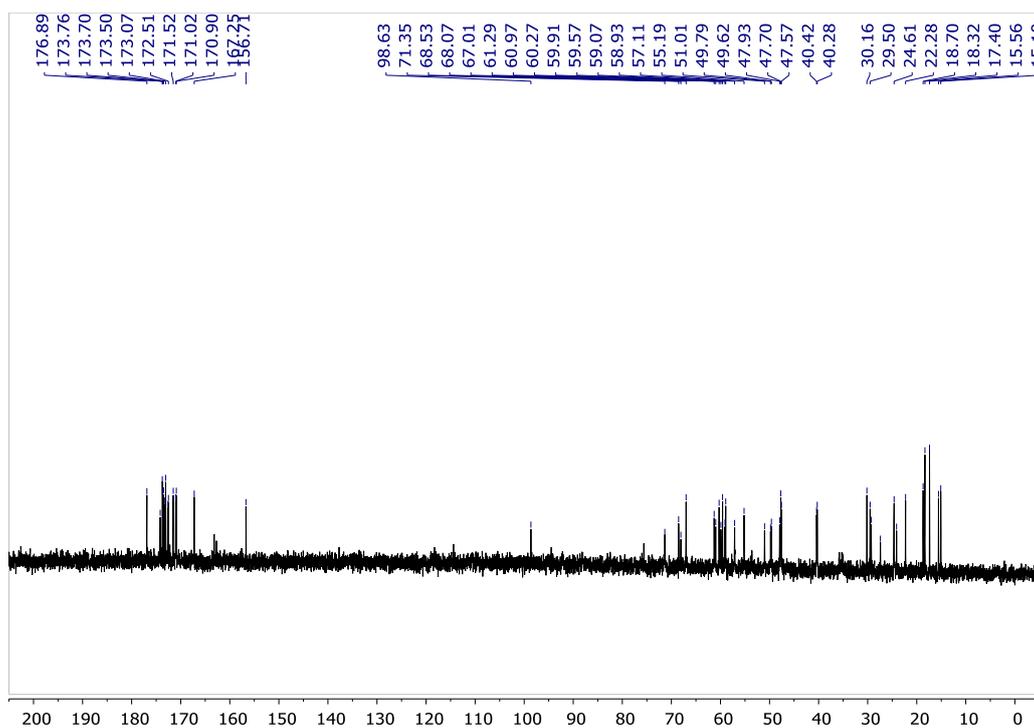
¹H NMR (400 MHz, D₂O) δ (ppm): 0.95-0.97 (m, 6H, Val2_γ), 1.21-1.25 (m, 6H, Thr3_γ, Thr8_γ), 1.31-1.43 (m, 6H, Ala5_β, Ala11_β), 1.69-2.69 (m, 18H, Ac, Pro6_β, Pro10_β, Pro12_β, Pro10_γ, Pro12_γ, Arg9_β, Arg9_γ, Val2_β), 2.84-3.07 (m, 2H, Asp7_β), 3.17-3.27 (m, 2H, Gly1_α), 5.33-5.52 (m, 1H, Pro6_γ).

^1H NMR 400 MHz in D_2O registered at 298K

A second set of signals (in a small percentage) is observed. They correspond to the *cis* disposition of the amide bond of proline residues (Dziadek, S.; Griesinger, C.; Kunz, H.; Reinscheid, U. M. *Chem. Eur. J.* **2006**, *12*, 4981–4993).

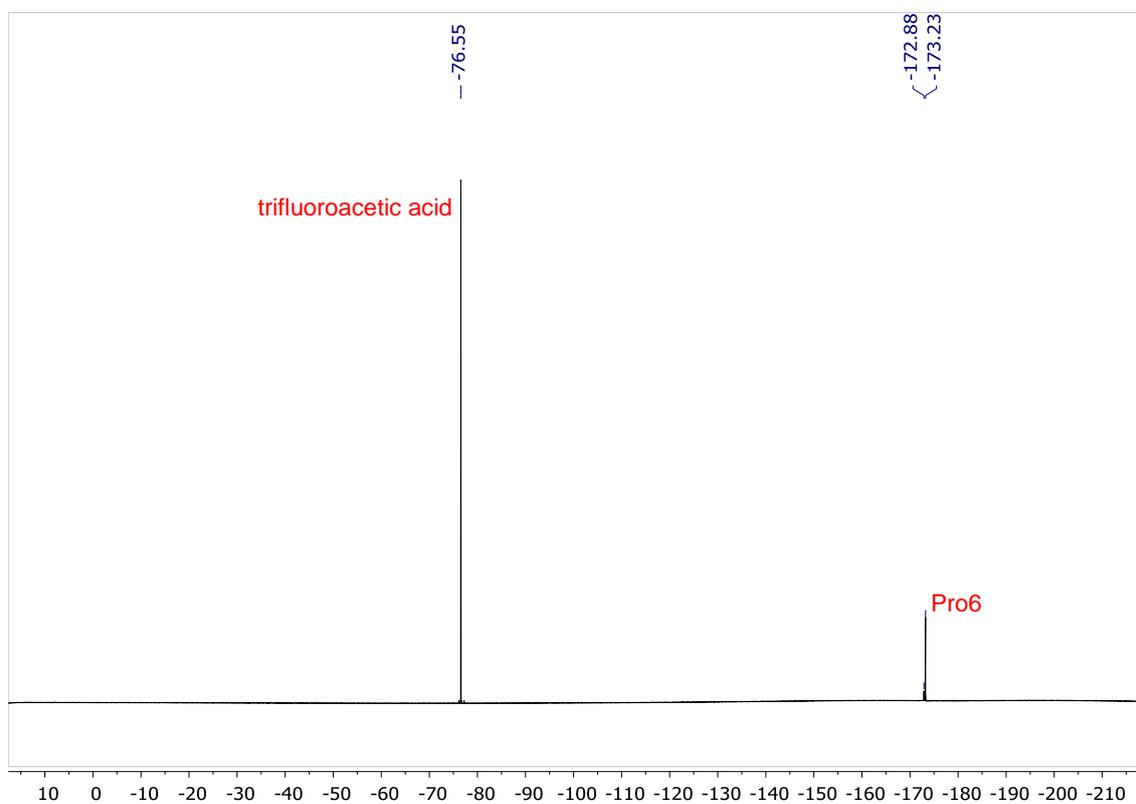


^{13}C NMR 75 MHz in D_2O registered at 298K

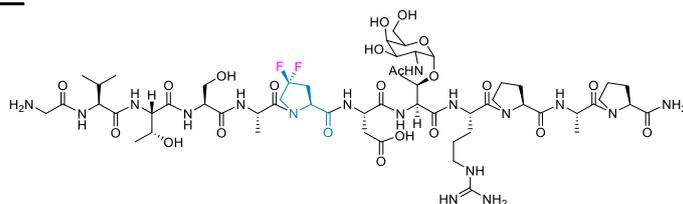


Decoupled $^{19}\text{F}\{^1\text{H}\}$ NMR (282 MHz, D_2O)

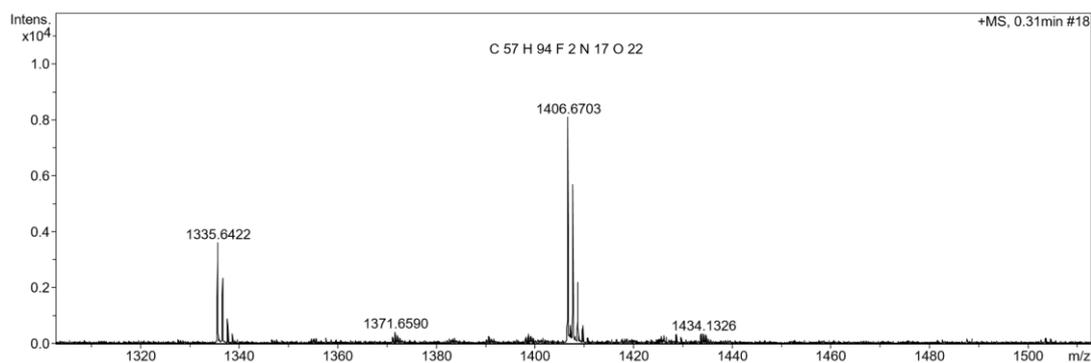
A second set of signals (in a small percentage) is observed. They correspond to the *cis* disposition of the amide bond of proline residues (Dziadek, S.; Griesinger, C.; Kunz, H.; Reinscheid, U. M. *Chem. Eur. J.* **2006**, 12, 4981–4993).



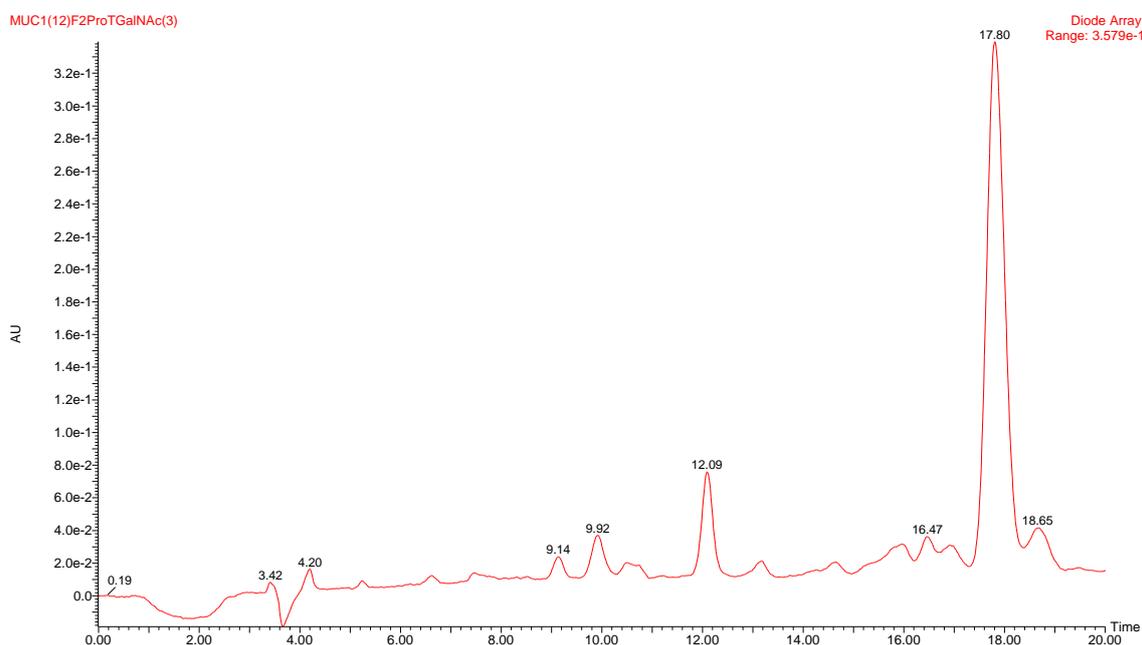
Glycopeptide 2fP**



HRMS (ESI+) m/z : calcd. for C₅₇H₉₄F₂N₁₇O₂₂: [M+H]⁺: 1406.6449 found: 1406.6703.

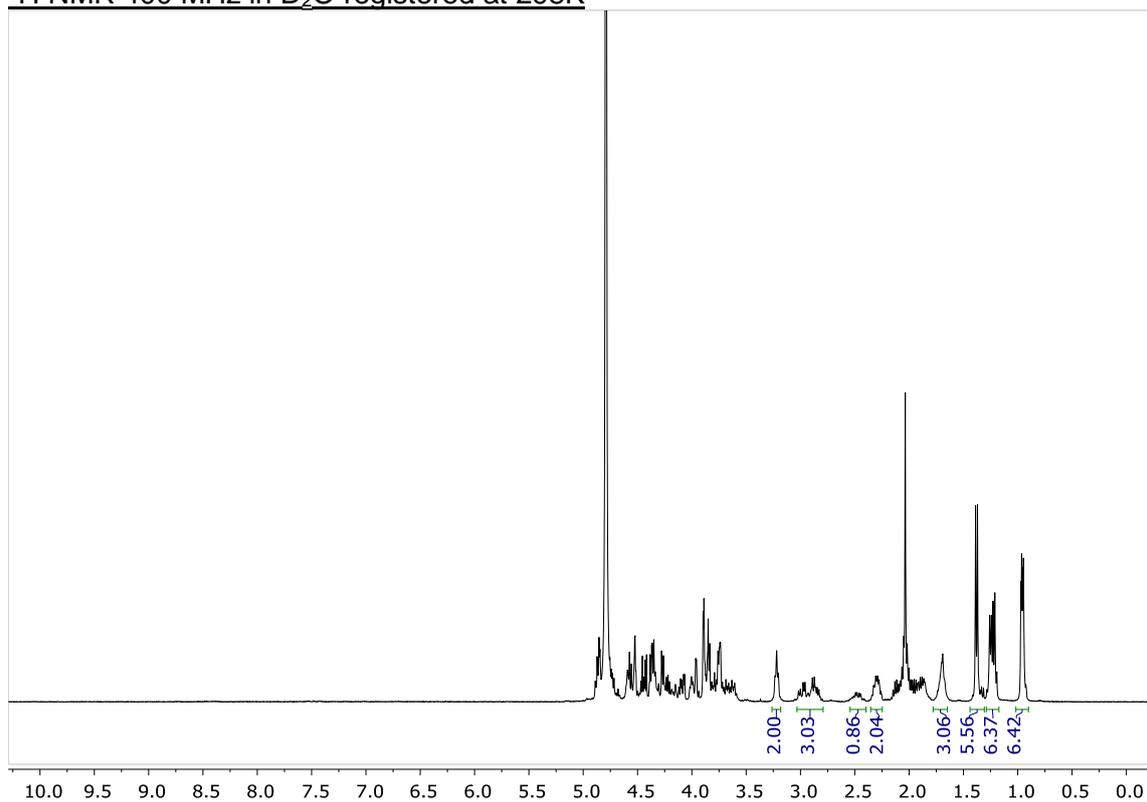


HPLC: R_t = 17.80 min (Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ = 212 nm).

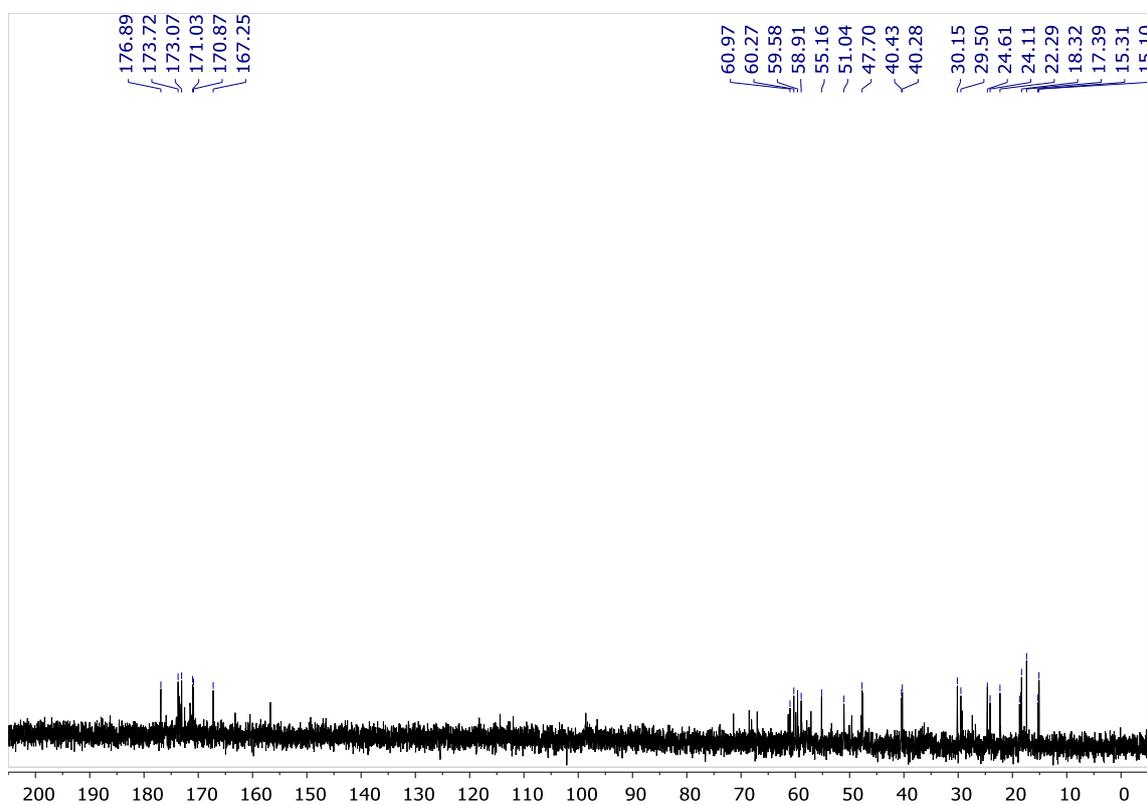


¹H NMR (400 MHz, D₂O) δ (ppm): 0.95-0.97 (m, 6H, Val2 γ), 1.19-1.28 (m, 6H, Thr3 γ , Thr8 γ), 1.31-1.42 (m, 6H, Ala5 β , Ala11 β), 1.69-3.05 (m, 20H, Ac, Pro6 β , Pro10 β , Pro12 β , Pro10 γ , Pro12 γ , Arg9 β , Arg9 γ , Val2 β , Asp7 β), 3.17-3.27 (m, 2H, Gly1 α).

^1H NMR 400 MHz in D_2O registered at 298K



^{13}C NMR 75 MHz in D_2O registered at 298K



Decoupled $^{19}\text{F}\{^1\text{H}\}$ NMR (282 MHz, D_2O)

A second set of signals (in a small percentage) is observed. They correspond to the *cis* disposition of the amide bond of proline residues (Dziadek, S.; Griesinger, C.; Kunz, H.; Reinscheid, U. M. *Chem. Eur. J.* **2006**, *12*, 4981–4993).

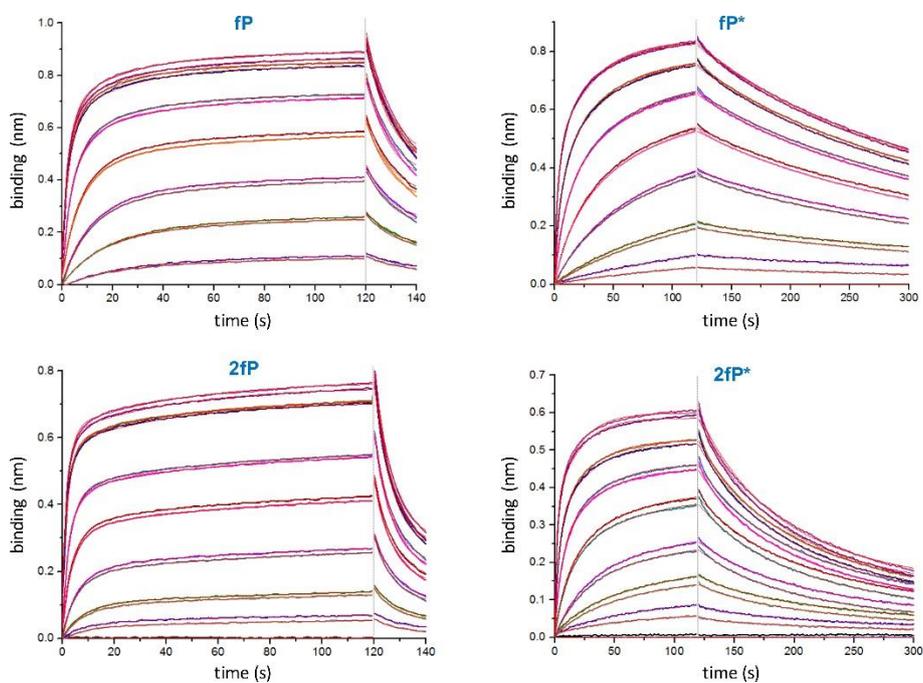
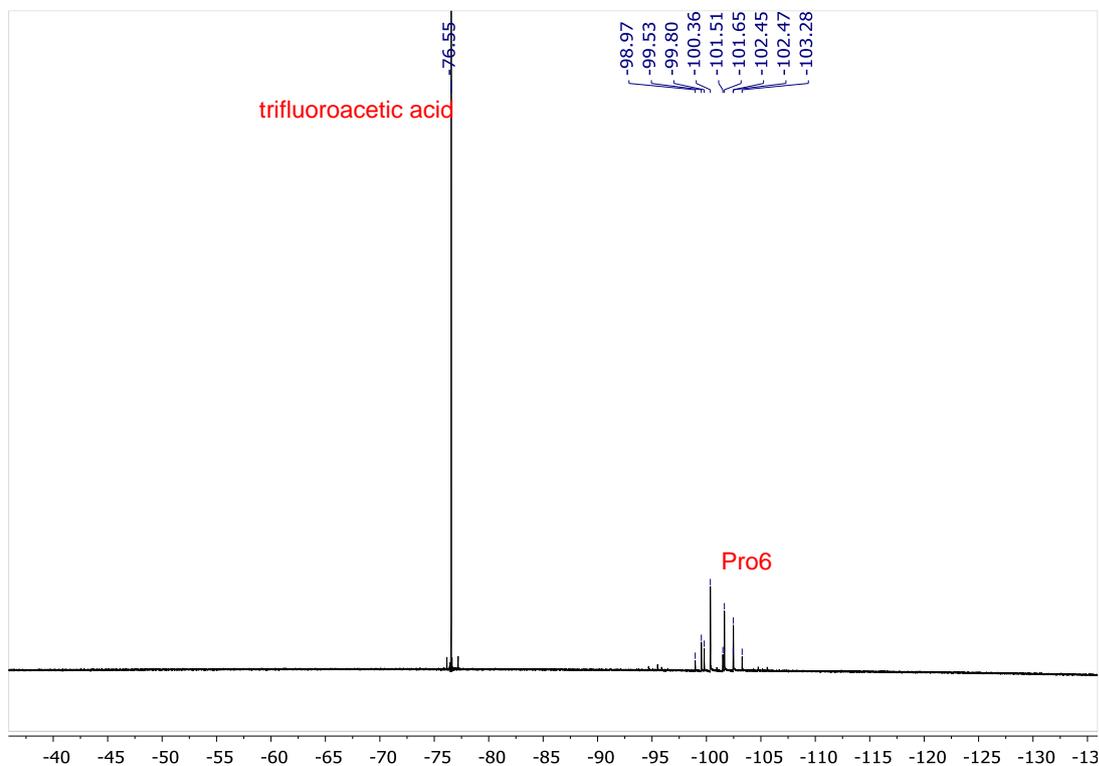


Figure S1. Bio-layer interferometry (BLI) curves obtained for the studied (glyco)peptides and antibody SM3.

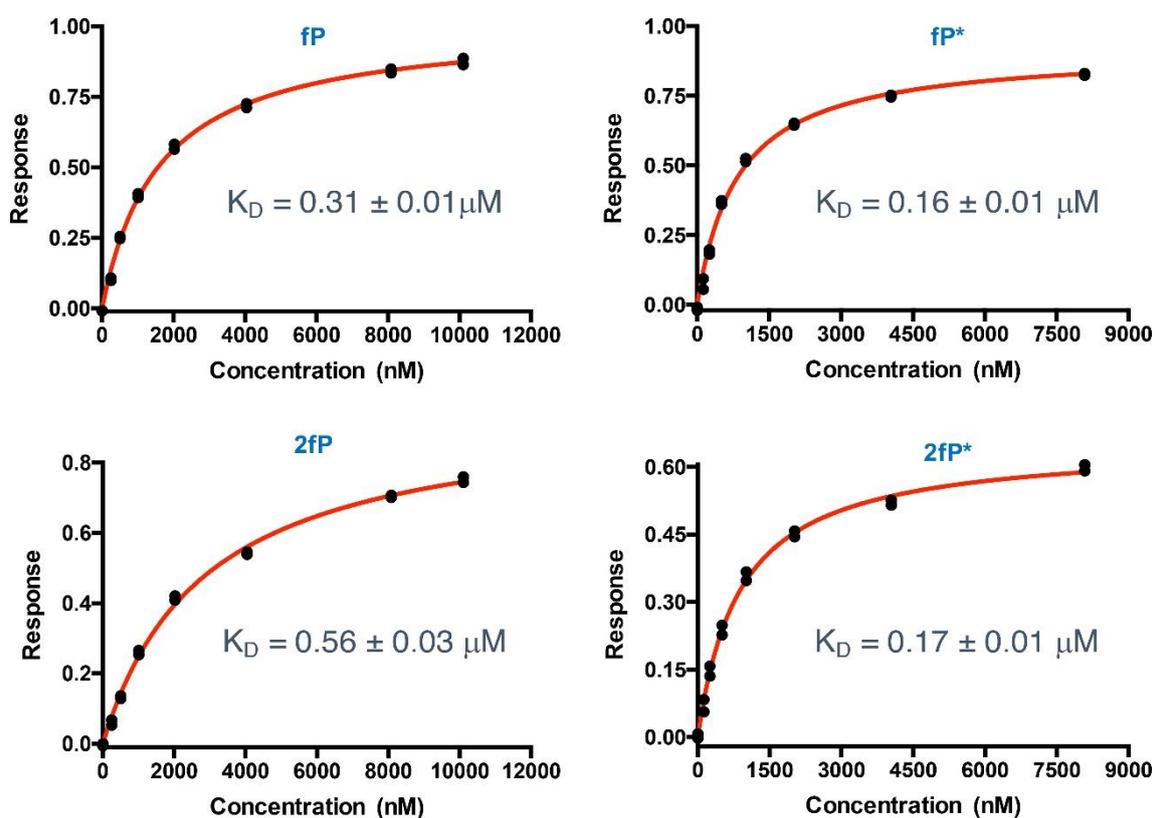


Figure S2. Bio-layer interferometry (BLI) fit obtained for the studied (glyco)peptides and antibody SM3, together with the K_D constants.

Microarrays. Microarrays slides called “Antibody chip” were obtained from Sumitomo Bakelite Co., Ltd. (Tokyo, Japan). The size of a slide is 75-mm long, 25-mm wide and 1-mm thick. Hybridization covers ($60 \times 25 \times 0.7$ mm) were also obtained from Sumitomo Bakelite Co., Ltd. Anti-MUC1 mouse mAbs clon SM3 (0.2 mg/mL) was purchased from Santa Cruz Biotechnology (TX, USA) and VU-3C6 (0.86 mg/mL) from Exalpha (MA, USA). FluoroLink™ Cy™3-labeled goat anti-mouse IgG was from Amersham Biosciences (Buckinghamshire, UK).

Microarray printing. We selected plastic “Antibody chip” (Sumitomo Bakelite, Japan) due to the non-fouling surface and selective covalent immobilization to the *N*-terminal amino group of the MUC1 (glyco)peptides library.

The printing/immobilization of the (glyco)peptides was done following the instructions and using the buffers of the microarray slides kit. (Glyco)peptides were spotted by MicroSys 5100 (Cartesian Technologies, CA, USA) with a 0.6-mm pitch using a Filgen solid spin (200 μm pin diameter). Each compound was printed in quadruplicate with 0.3 mm distance between spots of same compound and 0.6 mm gap among different

compounds (Figure S3, left panel). Each (glyco)peptide was printed at six different concentrations from 500 μ M to 15.6 μ M (Figure S3, right panel). Cy3 labeled BSA protein (25 μ g/mL) was used as grid. Spotting conditions were 23 °C and 60% of humidity. After printing, slides were incubated for overnight on dry conditions. Next, non-reacted groups were inactivated by blocking buffer at 37 °C for 1h under slow agitation. Finally, we rinsed the slides by washing buffer (3 x 5 min) and dried by centrifugation and then used for further binding assay of mAb.

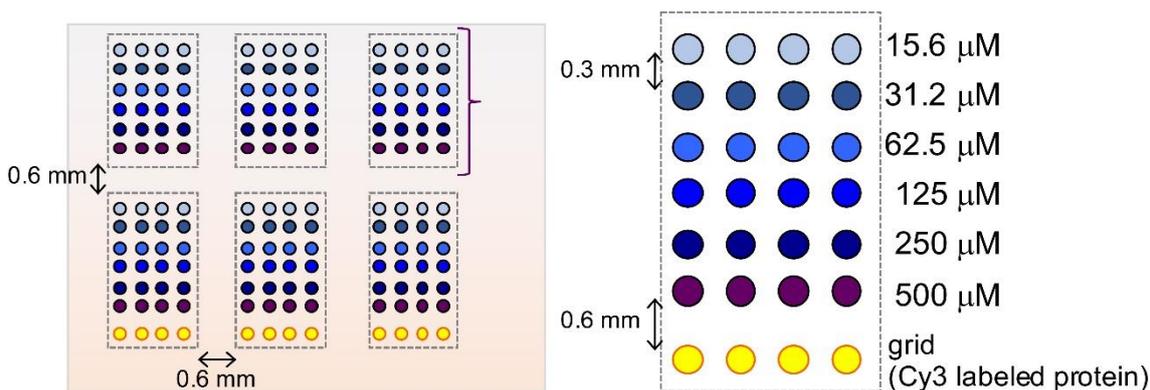


Figure S3. Microarray (glyco)peptides slides, schematic microarray printing on chamber slide (left panel) and printing pattern of each compound group (right panel).

Microarray mAb binding assay. The following buffers and solutions were used in this section: Buffer for the solution of mAb: 50 mM Tris-HCl, 100 mM NaCl, 1 mM CaCl₂, MnCl₂, MgCl₂, 0.05% Tween-20, 0.1% BSA, pH 7.4. Washing buffer: 50 mM Tris-HCl, 100 mM NaCl, 1 mM CaCl₂, MnCl₂, MgCl₂, 0.05% Triton X-100, pH 7.4. For the mAb incubation, 20 μ L of mAb solution in buffer (mAb concentration: 50.0 μ g/mL) was carefully added onto each chamber of slides and they were kept at r.t. for 2h on humid conditions. Next, slides were washed with washing buffer (3 x 2 min) and dried up by centrifugation. For the analysis of the binding, secondary Ab (Cy3-labeled Ab) was diluted to 1 μ g/mL in buffer and infused between hybridization covers and slides. After standing at r.t. for 1h at dark, slides were washed by different methods (stated for each case): (a) washing buffer (3 x 2 min) and dried up by centrifugation; (b) previous method (a) followed by washing buffer (2 x 2 min), water washing (2 x 2 min) and dried up by centrifugation. To storage the slides, they were degassed under vacuum and kept at 4 °C. Slides were subjected to fluorescent image scanning on a Tryphoon Trio Plus instrument (GE Healthcare). Array Vision software was used to quantify the fluorescence of each spot. The median value of relative fluorescence intensity (RFU) was used; spot intensities were determined by subtracting the average pixel intensity from the median pixel intensity of the local background within the spots. Fluorescence

of each spot is shown as the average of four replicate spots used to construct histograms showing the antibody-binding profile. As statistical analysis, Grubbs method was used to discriminate the outliers. Error bars are included showing the standard deviation for each peptide–mAb interaction. Results of the two anti-MUC1 mAbs at 50.0 $\mu\text{g}/\text{mL}$ are illustrated in Figures S4 and S5.

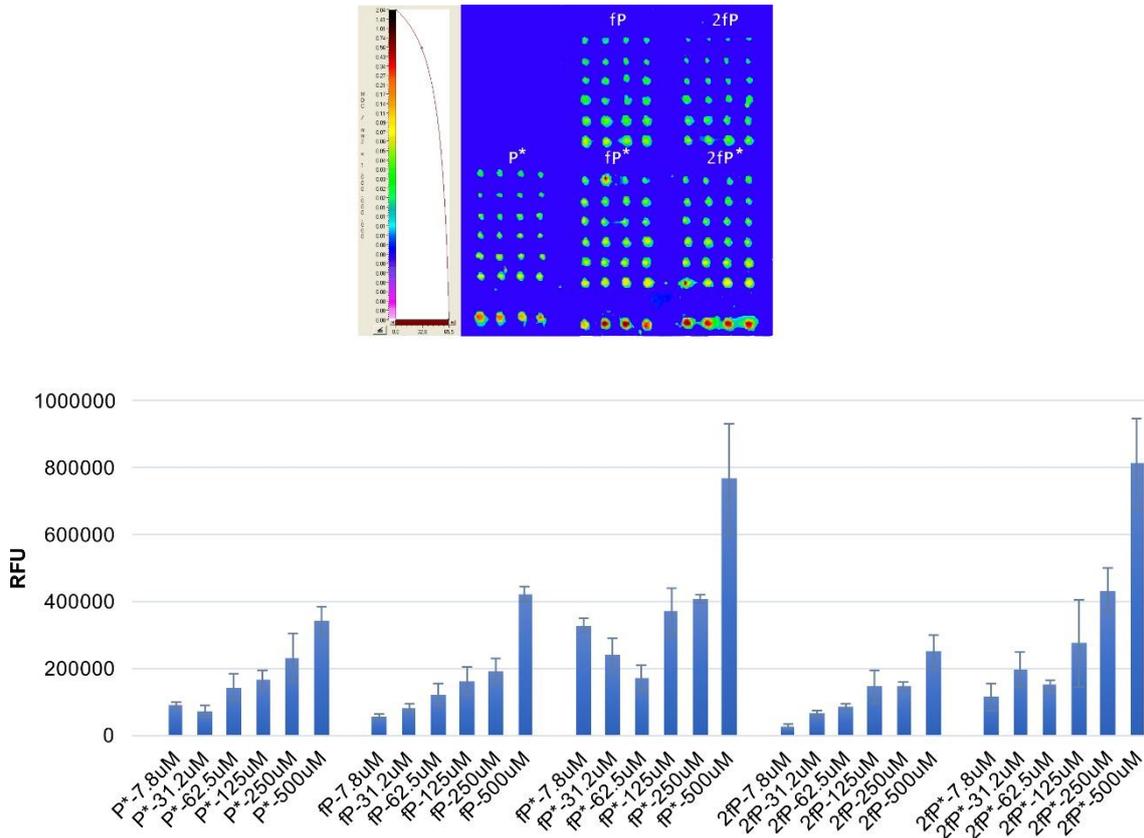


Figure S4. Binding assay with antibody SM3 (50.0 $\mu\text{g}/\text{mL}$). Washing method: a. Fluorescent image scan, together with fluorescent response graph is shown. RFU due to the binding of the Cy3-labeled secondary antibody were measured and represented as mean values in a bar chart.

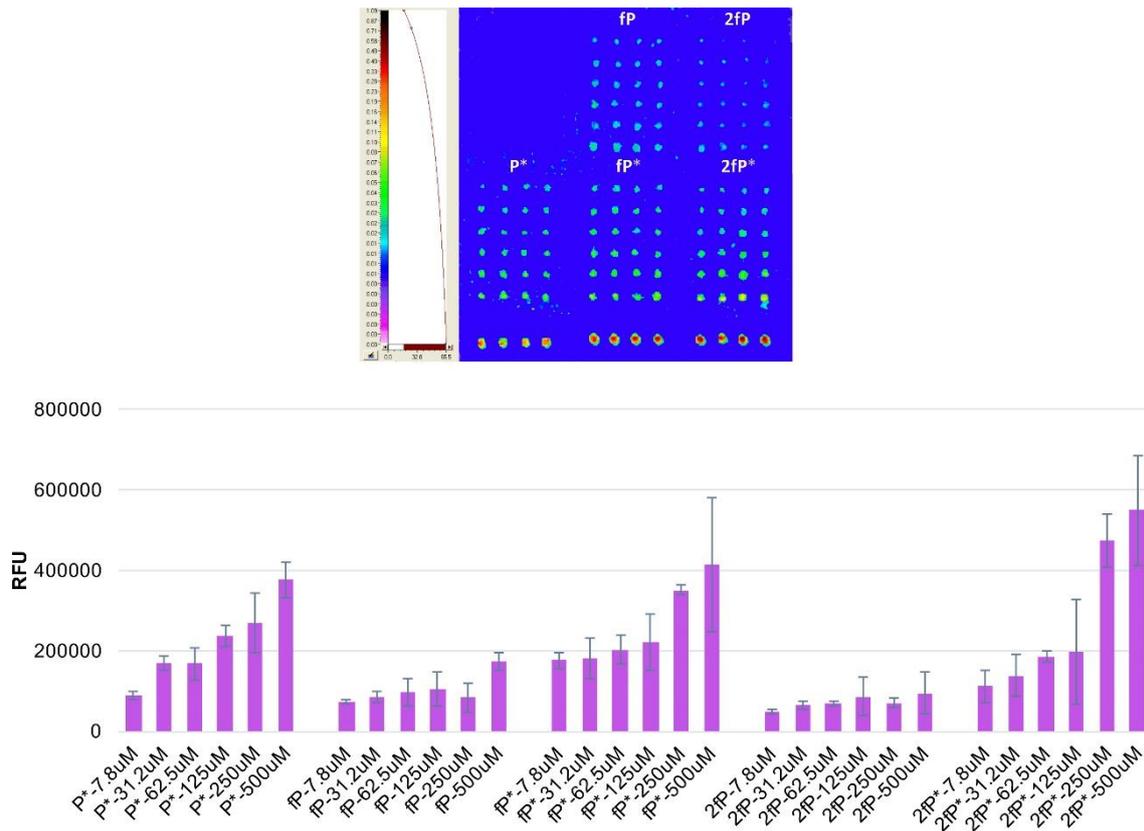


Figure S5. Binding assay with antibody VU-3C6 (50.0 ug/mL). Washing method: b. Fluorescent image scan, together with fluorescent response graph is shown. RFU due to the binding of the Cy3-labeled secondary antibody were measured and represented as mean values in a bar chart.

Crystallization. Expression and purification of scFv-SM3 has been described previously by us.^[S2] Crystals were grown by sitting drop diffusion at 18 °C. The drops were prepared by mixing 0.5 μ L of protein solution, containing 15 mg/mL of scFv-SM3 and 10 mM of **fp*** with 0.5 μ L of the mother liquor. Crystals of scFv-SM3 with the peptide above were grown in 20% PEG 3350, 0.2 M disodium hydrogen phosphate. The crystals were cryoprotected in mother liquor containing 15% ethylenglycol and frozen in a nitrogen gas stream cooled to 100 K.

Structure determination and refinement. The data was processed and scaled using the XDS package^[S3] and CCP4 software,^[S4] relevant statistics are given in Supplementary Table S1. The crystal structures were solved by molecular replacement with Phaser^[S5] and using the PDB entry 1SM3 as the template. Initial phases were further improved by cycles of manual model building in Coot63 and refinement with REFMAC5.^[S6] The final models were validated with PROCHECK.^[S7] Coordinates and structure factors have been deposited in the Worldwide Protein Data Bank (wwPDB). PDB id: 5OWP, see also Table S2.

Table S1. Data collection and refinement statistics. Values in parentheses refer to the highest resolution shell. Ramachandran plot statistics were determined with PROCHECK.

	1SM3 (scFv):fP**
Space group	P2 ₁ 2 ₁ 2 ₁
Wavelength (Å)	0.97
Resolution (Å)	20.00-1.85 (1.95-1.85)
Cell dimensions (Å)	<i>a</i> = 35.40 <i>b</i> = 68.15 <i>c</i> = 90.36
Unique reflections	19335
Completeness	99.7 (99.8)
R_{pim}	0.049 (0.394)
Mn(I) half-set correlation CC(1/2)	0.998 (0.685)
$I/\sigma(I)$	10.9 (1.9)
Redundancy	6.4 (6.3)
R_{work} / R_{free}	0.194/0.256
RMSD from ideal geometry, bonds (Å)	0.016
RMSD from ideal geometry, angles (°)	1.854
 protein (Å ²)	32.90
 glycopeptide (Å ²)	60.33
 solvent (Å ²)	42.13
Ramachandran plot: Most favoured (%)	94.59
Additionally allowed (%)	4.05
Disallowed (%)	1.35
PDB ID	5OWP

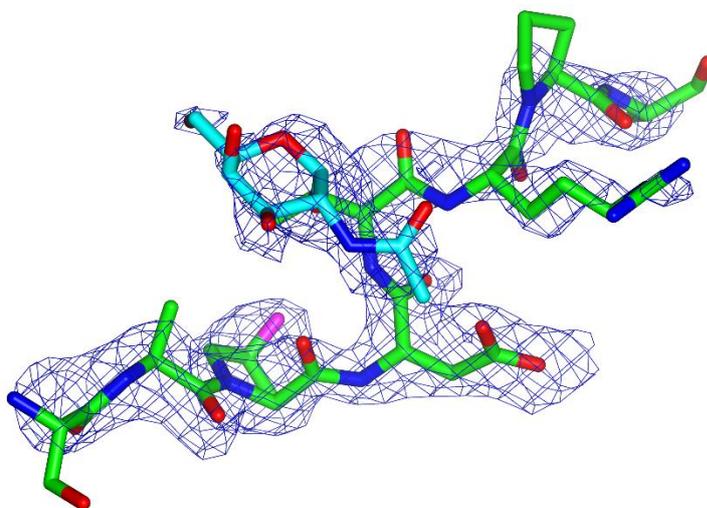


Figure S6. Electron density maps are $F_{\text{O}}-F_{\text{C}}$ syntheses (blue) contoured at 2.2σ for glycopeptide **fP***. The amino acid residues and the GalNAc moiety are colored in blue and green, respectively. The fluorine atom is in magenta. It is important to note that Pro at the C-terminal region could not be resolved.

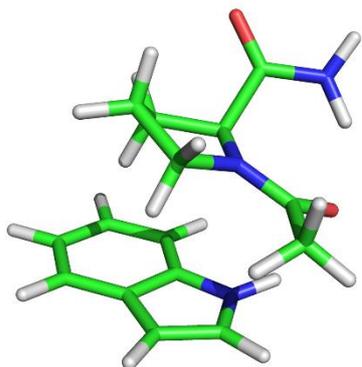
DFT calculations. Full geometry optimizations were carried out with Gaussian 09^[S8] using the M06-2X hybrid functional^[S9] and 6-31G(d,p) basis set. Bulk solvent effects were not considered in the calculations. Frequency analyses were carried out at the same level used in the geometry optimizations.

Table S2. Energies and lowest frequencies of the calculated structures.^a

Structure	E_{elec} (Hartree)	$E_{\text{elec}} + \text{ZPE}$ (Hartree)	Lowest freq. (cm-1)
Complex 1	-897.47018	-897.140361	27.5
Complex 2	-996.674061	-996.353785	24.1
Complex 3	-1095.897249	-1095.584985	33.1

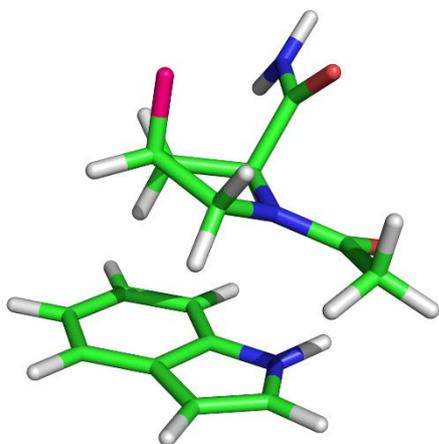
^a Energy values calculated at the M06-2X/6-31G(d,p) level. 1 Hartree = 627.51 Kcal mol⁻¹.

Cartesian coordinates of complex 1 calculated with M06-2X/6-31G(d,p)



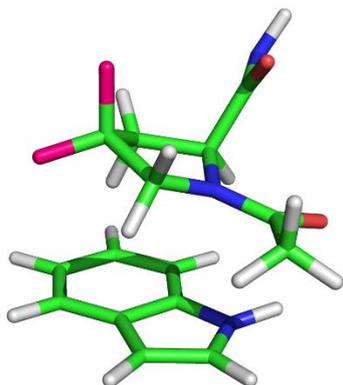
C	1.508039000000	1.628295000000	0.090457000000
C	1.223044000000	-0.826821000000	0.189882000000
C	2.621301000000	-1.371430000000	0.551535000000
C	0.496496000000	-1.730120000000	-0.801694000000
C	1.001147000000	-1.226774000000	-2.157907000000
C	1.033343000000	0.290928000000	-1.971608000000
N	1.337915000000	0.448968000000	-0.539326000000
N	3.288863000000	-0.632204000000	1.468889000000
O	1.627861000000	1.706259000000	1.322467000000
O	3.070559000000	-2.391301000000	0.052767000000
C	1.580928000000	2.859241000000	-0.785142000000
H	0.735445000000	2.910249000000	-1.476180000000
H	2.502669000000	2.840205000000	-1.375601000000
H	0.661329000000	-0.658670000000	1.114636000000
H	0.729641000000	-2.780599000000	-0.625620000000
H	-0.582457000000	-1.571379000000	-0.710903000000
H	1.791147000000	0.782936000000	-2.587682000000
H	0.051972000000	0.737420000000	-2.182242000000
H	4.230990000000	-0.907164000000	1.701884000000
H	2.906964000000	0.248279000000	1.797505000000
H	2.009862000000	-1.610137000000	-2.338106000000
H	0.358173000000	-1.519601000000	-2.990457000000
H	1.589926000000	3.736963000000	-0.139882000000
C	-2.260149000000	1.813580000000	-0.698040000000
C	-1.515497000000	2.351053000000	0.317359000000
C	-2.553131000000	0.457513000000	-0.321235000000
C	-1.954493000000	0.246479000000	0.946424000000
C	-3.258797000000	-0.598952000000	-0.924376000000
N	-1.342922000000	1.421811000000	1.316908000000
C	-2.042745000000	-0.977990000000	1.620664000000
C	-3.342343000000	-1.813927000000	-0.264791000000
C	-2.739275000000	-2.001080000000	0.998008000000
H	-1.094411000000	3.341153000000	0.423267000000
H	-3.730614000000	-0.461329000000	-1.892992000000
H	-0.649028000000	1.516971000000	2.046437000000
H	-1.584778000000	-1.118852000000	2.595518000000
H	-3.881552000000	-2.637894000000	-0.721015000000
H	-2.824965000000	-2.964514000000	1.490340000000
H	-2.581876000000	2.326307000000	-1.593301000000

Cartesian coordinates of complex 2 calculated with M06-2X/6-31G(d,p)



C	-1.195334000000	1.829065000000	-0.360328000000
C	-1.101222000000	-0.596692000000	-0.486837000000
C	-2.560231000000	-0.848453000000	-0.914301000000
C	-0.576015000000	-1.636478000000	0.514126000000
C	-1.059680000000	-1.107153000000	1.856859000000
C	-1.051052000000	0.415051000000	1.711329000000
N	-1.024475000000	0.637038000000	0.266108000000
N	-2.717776000000	-1.902696000000	-1.764738000000
O	-1.150552000000	1.918496000000	-1.586543000000
O	-3.487580000000	-0.183224000000	-0.500869000000
C	-1.440629000000	3.026366000000	0.526711000000
H	-2.463565000000	2.972274000000	0.914383000000
H	-1.335268000000	3.929666000000	-0.073401000000
H	-0.477144000000	-0.534934000000	-1.383809000000
H	0.518096000000	-1.638864000000	0.490484000000
H	-0.944439000000	-2.648721000000	0.333635000000
H	-1.952200000000	0.828553000000	2.175740000000
H	-0.156761000000	0.849797000000	2.174192000000
H	-3.641196000000	-2.075078000000	-2.133004000000
H	-1.930229000000	-2.332370000000	-2.222840000000
F	-2.368349000000	-1.538506000000	2.065824000000
H	-0.469648000000	-1.455907000000	2.707978000000
H	-0.753916000000	3.055692000000	1.377168000000
C	2.443201000000	1.598356000000	1.025539000000
C	1.881227000000	2.322250000000	0.007688000000
C	2.638340000000	0.265946000000	0.523796000000
C	2.171831000000	0.263383000000	-0.815756000000
C	3.166351000000	-0.922014000000	1.061770000000
N	1.737710000000	1.533248000000	-1.109495000000
C	2.216684000000	-0.882415000000	-1.620702000000
C	3.206096000000	-2.059527000000	0.271106000000
C	2.733667000000	-2.039518000000	-1.059326000000
H	1.576189000000	3.359288000000	-0.020142000000
H	3.541483000000	-0.944290000000	2.080920000000
H	1.123414000000	1.769042000000	-1.879641000000
H	1.869569000000	-0.858558000000	-2.649915000000
H	3.611840000000	-2.981700000000	0.674660000000
H	2.785415000000	-2.945168000000	-1.655568000000
H	2.710935000000	1.973526000000	2.002860000000

Cartesian coordinates of complex 3 calculated with M06-2X/6-31G(d,p)



C	1.162513000000	1.970495000000	-0.056786000000
C	1.087735000000	-0.307112000000	0.775004000000
C	2.551886000000	-0.390022000000	1.244385000000
C	0.596421000000	-1.614324000000	0.139770000000
C	1.027503000000	-1.477235000000	-1.310302000000
C	1.046859000000	0.020381000000	-1.628551000000
N	0.982956000000	0.646746000000	-0.313125000000
N	2.723665000000	-1.023218000000	2.438083000000
O	1.144970000000	2.397465000000	1.095171000000
O	3.471690000000	0.027003000000	0.569398000000
C	1.383458000000	2.864569000000	-1.252539000000
H	2.400233000000	2.707342000000	-1.628432000000
H	1.279813000000	3.901164000000	-0.933618000000
H	0.454248000000	0.003203000000	1.610641000000
H	-0.495670000000	-1.652881000000	0.174942000000
H	1.015936000000	-2.515692000000	0.590704000000
H	1.968686000000	0.267351000000	-2.164375000000
H	0.168700000000	0.280881000000	-2.228856000000
H	3.657306000000	-1.056393000000	2.820031000000
H	1.947824000000	-1.184800000000	3.060082000000
F	2.271859000000	-2.012139000000	-1.489756000000
F	0.204510000000	-2.154572000000	-2.160131000000
H	0.680422000000	2.646593000000	-2.061027000000
C	-2.388907000000	1.339683000000	-1.337140000000
C	-1.890699000000	2.337629000000	-0.542766000000
C	-2.589306000000	0.197857000000	-0.487574000000
C	-2.191106000000	0.585881000000	0.817246000000
C	-3.075401000000	-1.106439000000	-0.690983000000
N	-1.789768000000	1.898708000000	0.757259000000
C	-2.262327000000	-0.286959000000	1.910794000000
C	-3.146049000000	-1.973382000000	0.387151000000
C	-2.741319000000	-1.566864000000	1.677412000000
H	-1.605956000000	3.349721000000	-0.795267000000
H	-3.388277000000	-1.427837000000	-1.679844000000
H	-1.233553000000	2.365966000000	1.461327000000
H	-1.968114000000	0.034425000000	2.906035000000
H	-3.519370000000	-2.981958000000	0.242619000000
H	-2.816108000000	-2.267824000000	2.502954000000
H	-2.612259000000	1.414800000000	-2.391726000000

Molecular dynamics (MD) simulations on glycopeptides P^{}, fP^{**} and 2fP^{**} in complex with scFv-SM3.** The crystal structure of glycopeptide APD-TGalNAc-RP in complex with scFv-SM3 (PDB ID: 5A2K) was used in the MD simulations. Each complex was then immersed in a water box with a 10 Å buffer of TIP3P water molecules.^[S10] The simulations were carried out with AMBER 16 package^[S11] implemented with ff14SB,^[S12] GAFF^[S13] and GLYCAM06j^[S14] force fields. The parameters and charges for the unnatural amino acids were generated with the antechamber module of AMBER, using GAFF force field and AM1-BCC method for charges.^[S15] 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.^[S16] 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 200 ns under the same simulation conditions.

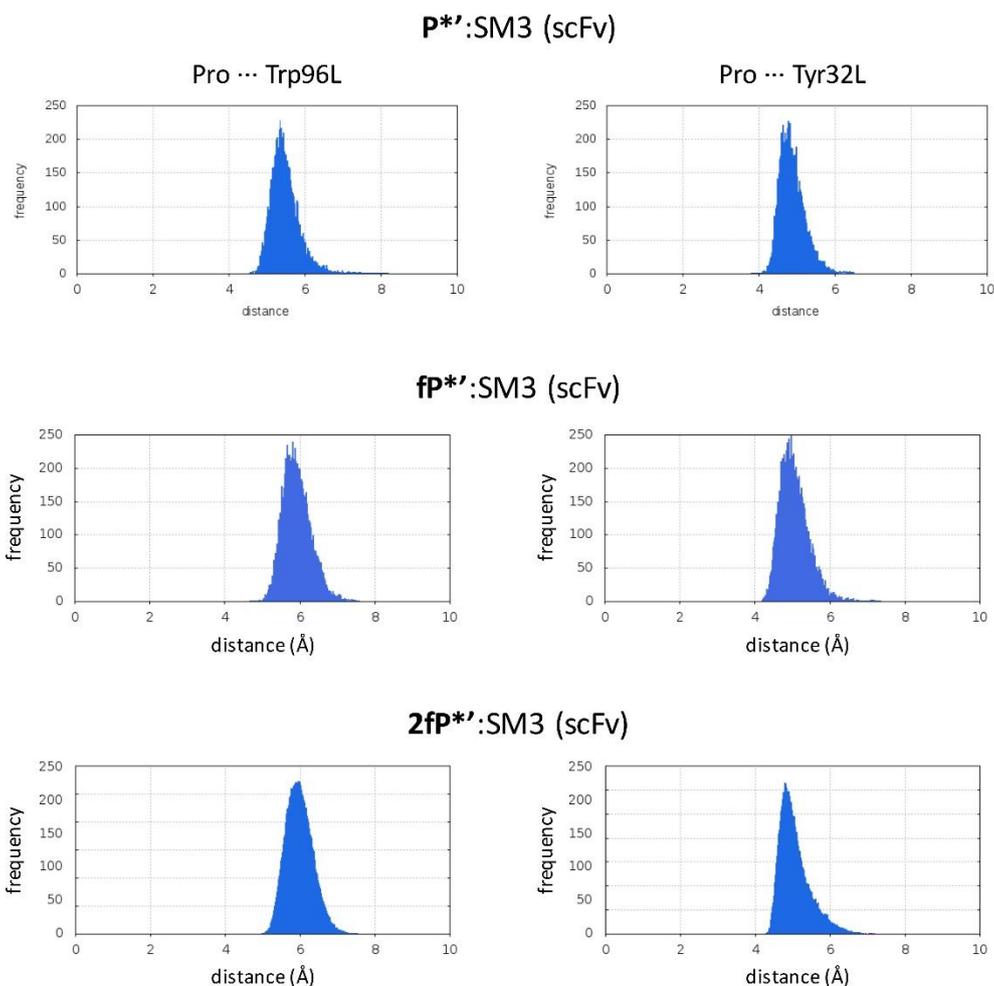


Figure S7. Distance distribution Pro-Trp96L and Pro-Tyr32L obtained from 200 ns MD simulations for glycopeptides **P^{*}**, **fP^{*}** and **2fP^{*}** bound to the antibody SM3.

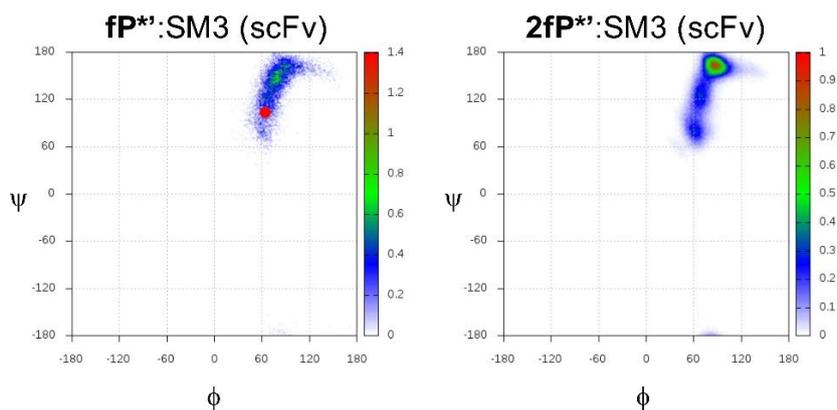


Figure S8. Monitoring of ϕ , ψ torsional angles for the glycosidic linkage (GalNAc-Thr) in glycopeptides **fP^{*}** and **2fP^{*}** bound to antibody SM3 through the 200 ns MD simulations. The conformational behavior of this glycosidic linkage is similar in both derivatives. The main conformer populated in solution for the glycosidic linkage differs from that observed in the X-ray structure of compound **fP^{*}** bound to SM3 (red circle, PDB ID: 5OWP).

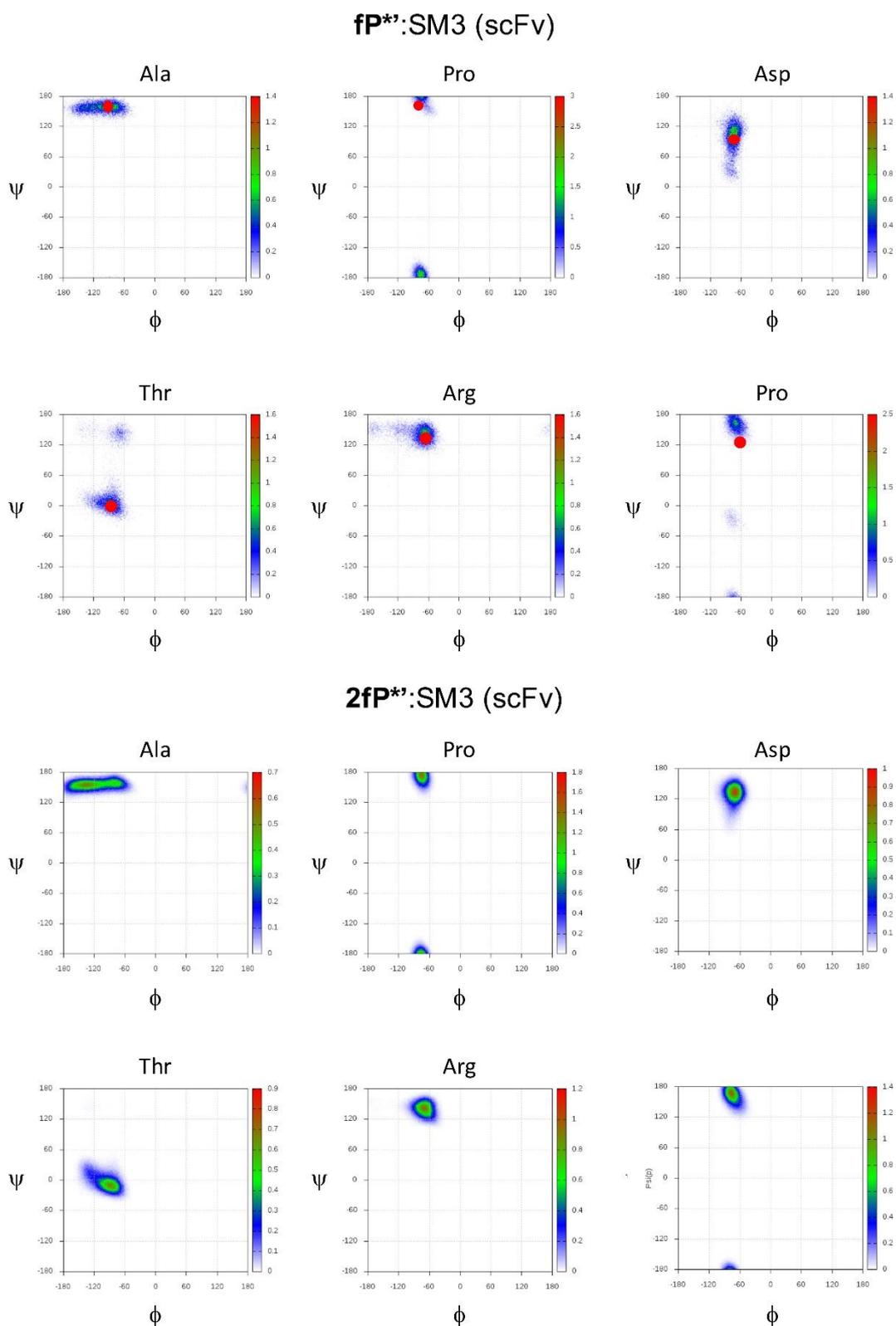


Figure S9. Monitoring of ϕ , ψ torsional angles for the backbone of glycopeptides **fP^{*}** and **2fP^{*}** bound to antibody SM3 through the 200 ns MD simulations. The red circle represents the ϕ/ψ values obtained for the peptide backbone of **fP^{*}** in complex with SM3 in the X-ray structure (PDB ID: 5OWP).

Human sera samples

Samples were obtained from Biobanco-iMM, Lisbon Academic Medical Center, Lisbon, Portugal.

Table S3. Age distribution of sera donors used for detection of human circulating antibodies against MUC-1 variants.

Diagnostic	Age distribution (years)	Number of samples
Prostatic adenocarcinoma	65.3 ± 6.0 (p=0.0947, n.s.)	9
Prostatic benign hyperplasia	71.6 ± 8.1 (p=0.0092, **)	11
Male controls	60.2 ± 2.5	5

Enzyme-linked Immunosorbent Assays (ELISA) for detection of human circulating antibodies against MUC-1 and synthetic variants

High-binding ELISA plates (JetBioFil, China) were coated with 100 µL of a 20 µM solution (in NaPi buffer, pH 7) of each MUC variant, and incubated for 2 h at 37 °C. Wells were thrice washed with wash buffer (0.05% Tween-20 in PBS 1x), and then blocked for 1 h at 37 °C with 100 µl of ELISA buffer (1% BSA in wash buffer). The wells were again washed 3 times with wash buffer, before incubation with sera samples (dilution 1:50 in ELISA buffer), performed at RT, for 90 min. The wells were again washed 3 times with wash buffer, and were incubated with donkey anti human IgG H&L antibody conjugated to HRP (Abcam, UK) for 1 h at RT (final concentration 300 ng/mL). Finally, the wells were washed, and development of the assay was performed. Briefly, 90 µL of 3,3',5,5'-Tetramethylbenzidine (TMB) 1x solution (eBiosciences, ThermoFisher Scientific, USA) was added to the wells, and after 10 min at RT, 50 µL of 2N H₂SO₄ was added to stop the reaction. Absorbance at 450 nm was read within the next 10 min, using a Tecan Infinite M200 plate-reader. Background absorbance values were subtracted, i.e. absorbance obtained for wells coated with the same MUC variant but incubated only with the secondary antibody. Values were normalized to those obtained for healthy controls. All groups were compared to **P*** using Wilcoxon matched-pairs signed rank test; Experimental groups were compared to equivalent **Healthy controls** using an unpaired T-test with Welch's correction (one-tailed); * p<0.05; ** p<0.02

Table S4. Raw data obtained for detection of circulating human antibodies against MUC-1. Data is shown as absorbance at 450 nm to which the background absorbance value was subtracted (see *Methods*). "Signal-to-noise" corresponds to the average of absorbance values of each group normalized to the values obtained for "healthy controls" (see Fig. 7 in the main text).

Diagnostic	Age	Absorbance at 450 nm (corrected for "blank")		"Signal-to-noise" ratio	
		P*	2fP*	P*	2fP*
Prostatic Adenocarcinoma	60	0.3171	0.1736	1.4806	2.4892
	61	0.1407	0.0431	0.6573	0.6178
	61	0.0385	-0.0026	0.1800	-0.0380
	62	0.2853	0.1971	1.3321	2.8254
	63	0.1354	0.0381	0.6323	0.5469
	64	2.2600	1.9428	10.5546	27.8491
	68	0.9091	0.5183	4.2456	7.4298
	71	2.3987	2.3477	11.2021	33.6532
	78	0.8835	0.3914	4.1260	5.6107
Average		0.8187 * (p=0.0424 vs Healthy)	0.6277 * (p=0.0484 vs Healthy)	3.8234	8.9982 * (p=0.0273 vs P*)
Prostatic Benign Hyperplasia	64	0.2351	0.1880	1.0982	2.6950
	65	0.2384	0.3711	1.1136	5.3197
	65	0.1864	0.2030	0.8707	2.9100
	65	0.1055	0.2215	0.4927	3.1759
	67	0.0781	0.1462	0.3650	2.0958
	68	0.3628	0.6662	1.6945	9.5492
	72	0.0511	-0.0236	0.2389	-0.3390
	72	0.5257	0.1613	2.4548	2.3115
	78	0.3893	0.2420	1.8178	3.4698
	84	0.0408	-0.0878	0.1908	-1.2593
	87	0.7009	0.4896	3.2735	7.0183
Average		0.2649 ^{N.S.} (p=0.2646 vs Healthy)	0.2343 * (p=0.00234 vs Healthy)	1.2373	3.3588 ** (p=0.0068 vs P*)
Healthy Male Controls	56	0.2068	0.0015	0.9658	0.0215
	60	0.2738	0.1211	1.2787	1.7360
	61	0.2408	0.1148	1.1248	1.6456
	62	0.3064	0.1542	1.4307	2.2097
	62	0.0428	-0.0427	0.2001	-0.6128
Average		0.2141	0.0698	1.0000	1.0000

References

- [S1] C. Plattner, M. Höfener,; N. Sewald, *Org. Lett.* 2011, **13**, 545-547.
- [S2] N. Martínez-Sáez, J. Castro-López, J. Valero-González, D. Madariaga, I. Compañón, V. J. Somovilla, M. Salvadó, J. L. Asensio, J. Jiménez-Barbero, A. Avenoza, J. H. Busto, G. J. L. Bernardes, J. M. Peregrina, R. Hurtado-Guerrero and F. Corzana, *Angew. Chem. Int. Ed.*, 2015, **54**, 9830–9834.
- [S3] W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.* 2010, **66**, 125-132.
- [S4] (a) M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, *Acta Crystallogr. D Biol. Crystallogr.* 2011, **67**, 235-242. (b) *Collaborative Computational Project N: The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallogr. D Biol. Crystallogr.* 1994, **50**, 760-763.
- [S5] P. Emsley, K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.* 2004, **60**, 2126-2132.
- [S6] G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallogr. D Biol. Crystallogr.* 2011, **67**, 355-367.
- [S7] R. A. Laskowski, M. W. Macarthur, D. S. Moss, J. M. Thornton, *J. Appl. Cryst.* 1993, **26**, 283-291.
- [S8] Gaussian 09, Revision A.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.
- [S9] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* 2008, **120**, 215–241.

- [S10] K. Kiyohara, K. Gubbins, A. Panagiotopoulos, *Mol. Phys.* 1998, **94**, 803–808.
- [S11] D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman (2016), AMBER 2016, University of California, San Francisco.
- [S12] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* 2015, **11**, 3696–3713.
- [S13] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* 2004, **25**, 1157–1174.
- [S14] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* 2008, **29**, 622–655.
- [S15] A. Jakalian, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* 2002, **23**, 1623–1641.
- [S16] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* 1993, **98**, 10089–10092.