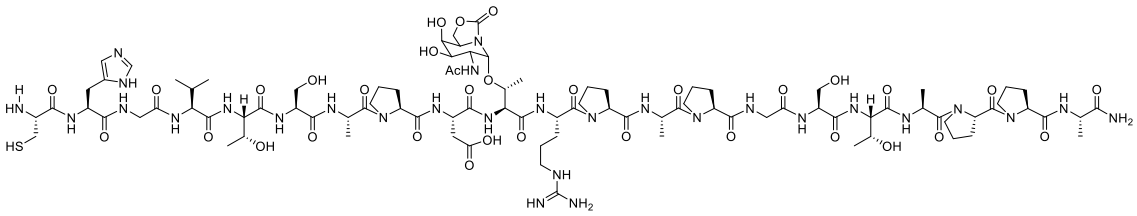
Supporting Information

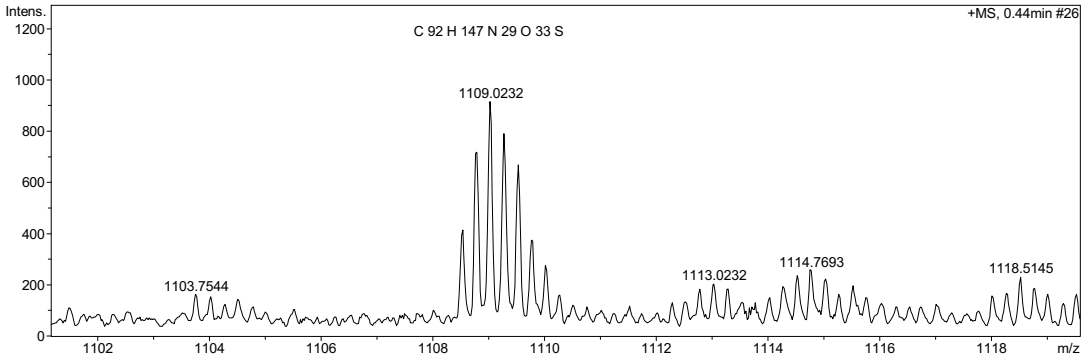
Synthesis, Conformational Analysis and *in vivo* Assays of an Anti-cancer Vaccine that features an Unnatural Antigen based on a sp^2-Iminosugar Fragment

Iris A. Bermejo, Claudio D. Navo, Jorge Castro-López, Ana Guerreiro, Ester Jiménez-Moreno, Elena M. Sánchez Fernández, Fayna García-Martín, Hiroshi Hinou, Shin-Ichiro Nishimura, José M. García Fernández, Carmen Ortiz Mellet, Alberto Avenoza, Jesús H. Busto, Gonçalo J. L. Bernardes, Ramón Hurtado-Guerrero, Jesús M. Peregrina, and Francisco Corzana

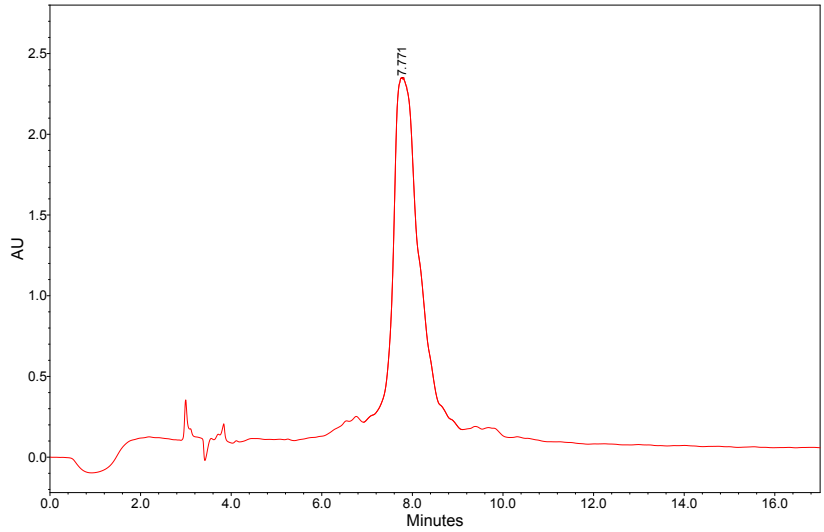
Synthesis and characterization of α-sp^2GalNAc-Thr **1a** and (glyco)peptides **2a**, **2b** and **2c** have being previously described.[S1]

Glycopeptide 3a: HRMS (m/z): $[M+2H]^{2+}$ calcd. for $C_{92}H_{147}N_{29}O_{33}S$, 1109.0213; found 1109.0232. Semipreparative HPLC on a Phenomenex Luna C18(2) column (10 µ, 250 mm x 21.2 mm), Rt = 8.50 min (Grad: water 0.1% TFA/acetonitrile (90:10) → (83:17), 17.0 min, λ = 212 nm). Flow rate: 10 mL/min. Analytical RP-HPLC: Rt = 7.771 min (C18, Grad: water 0.1% TFA/acetonitrile (90:10) → (83:17), 17.0 min, λ = 212 nm). Flow rate: 1 mL/min.

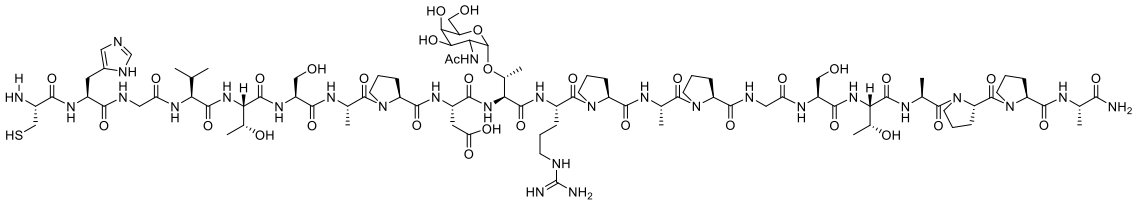


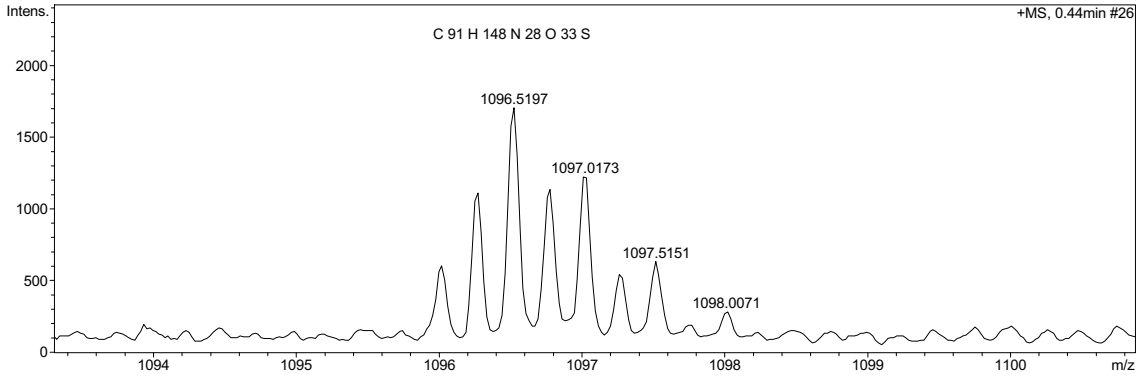


Formula	m/z	z	|err| [ppm]	Mean err [ppm]	mSigma
C 92 H 147 N 29 O 33 S	1109.0213	2+	1.7	-0.3	237.3

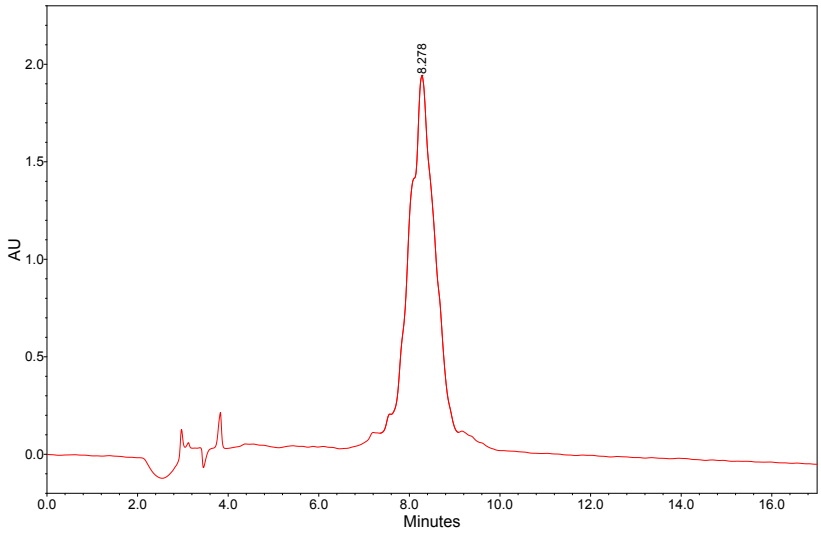


Glycopeptide 3b: HRMS (m/z): $[M+2H]^{2+}$ calcd. for $C_{91}H_{148}N_{28}O_{33}S$, 1096.5237; found 1096.5197. Semipreparative HPLC on a Phenomenex Luna C18(2) column (10 µ, 250 mm x 21.2 mm), Rt = 15.6 min (Grad: water 0.1% TFA/acetonitrile (90:10) → (83:17), 20.0 min, λ = 212 nm). Flow rate: 10 mL/min. Analytical RP-HPLC: Rt = 8.278 min (C18, Grad: water 0.1% TFA/acetonitrile (90:10) → (83:17), 17.0 min, λ = 212 nm). Flow rate: 1 mL/min.

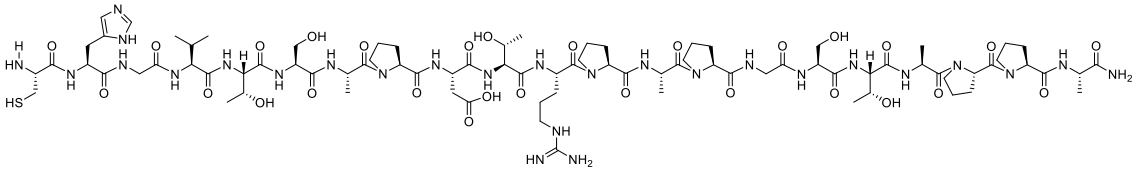


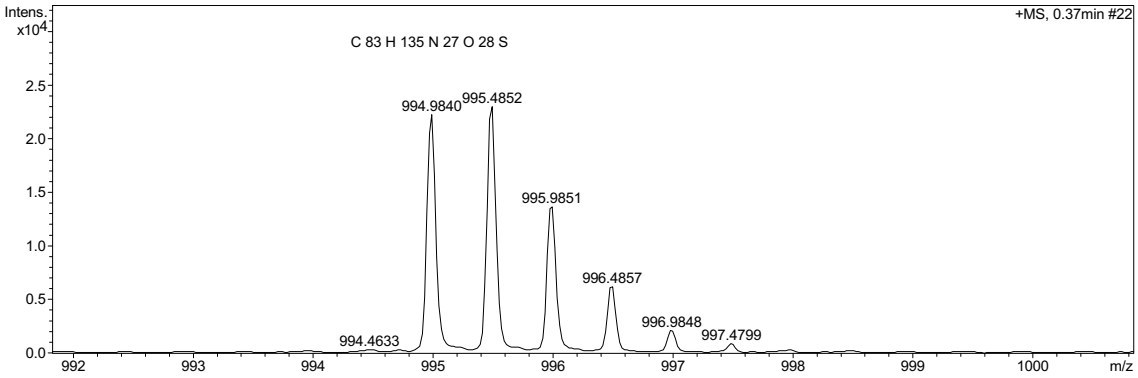


Formula	m/z	z	|err| [ppm]	Mean err [ppm]	mSigma
C 91 H 148 N 28 O 33 S	1096.5237	2+	3.6	7.1	187.8

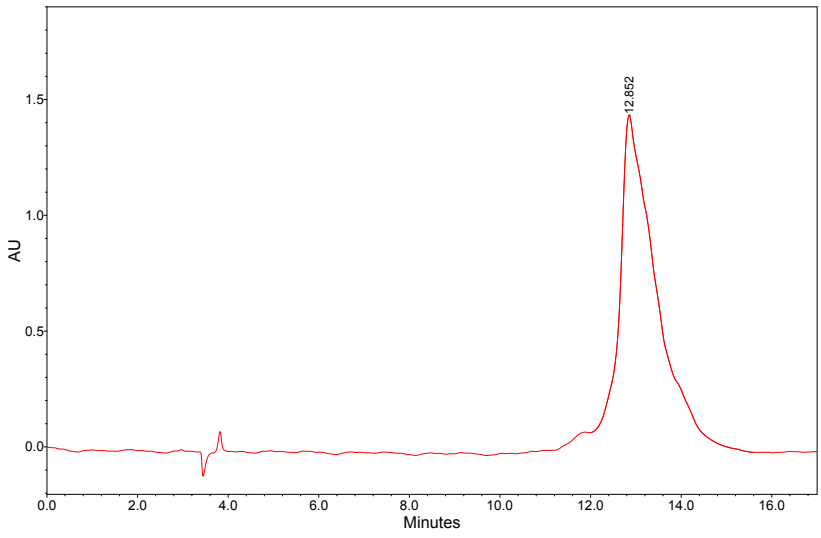


Peptide 3c: HRMS (m/z): $[M+2H]^{2+}$ calcd. for $C_{83}H_{135}N_{27}O_{28}S$, 994.9840; found 994.9840. Semipreparative HPLC on a Phenomenex Luna C18(2) column (10 µ, 250 mm x 21.2 mm), Rt = 15.4 min (Grad: water 0.1% TFA/acetonitrile (90:10) → (83:17), 10.0 min, λ = 212 nm). Flow rate: 10 mL/min. Analytical RP-HPLC: Rt = 12.852 min (C18, Grad: water 0.1% TFA/acetonitrile (90:10) → (83:17), 17.0 min, λ = 212 nm). Flow rate: 1 mL/min.





Formula	m/z	z	|err| [ppm]	Mean err [ppm]	mSigma
C 83 H 135 N 27 O 28 S	994.9840	2+	0.0	1.1	8.1



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 × 25 × 0.7 mm) were also obtained from Sumitomo Bakelite Co., Ltd. Anti-MUC1 mAb VU-11E2 (0.10 mg/mL) from Monosan (Uden, Netherlands). FluoroLinkTM CyTM3-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 on slides of naked peptide **2c** and glycopeptides **2a** and **2b** was done following the instructions of the maker and using the buffers of the microarray slides kit. (Glyco)peptides **2a**, **2b** and **2c** 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 S1, left panel). Each glycopeptide was printed at five different concentrations from 250 µM to 7.8 µM (Figure S1, right panel). Cy3 labeled BSA protein (25 µg/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 × 5 min) and dried by centrifugation and then used for further binding assay of mAb.

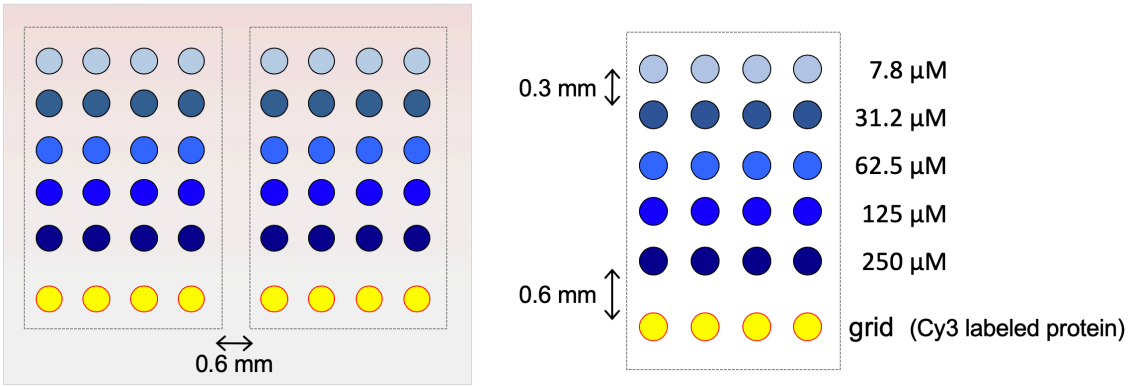


Figure S1. Microarray glycopeptides slides, schematic microarray printing on chamber slide (left panel) and printing pattern of each compound group (right panel).

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 $CaCl_2$, $MnCl_2$, $MgCl_2$, 0.05% Tween-20, 0.1% BSA, pH 7.4. Washing buffer: 50 mM Tris-HCl, 100 mM NaCl, 1 mM $CaCl_2$, $MnCl_2$, $MgCl_2$, 0.05% Triton X-100, pH 7.4. For the mAb incubation, 20 µL of mAb solution in buffer (mAb concentration: 50.0 µg/mL) was carefully added onto each chamber of

slides and they were kept at rt for 2h on humid conditions. Next, slides were washed with washing buffer (3 × 2 min) and dried up by centrifugation. For the analysis of the binding, secondary Ab (Cy3-labeled Ab) was diluted to 1 μg/mL in buffer and infused between hybridization covers and slides. After standing at rt for 1h at dark, slides were washed by: (1) washing buffer (3 × 2 min) and centrifugal drying; (2) followed by again washing buffer (2 × 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 (Figure S2).

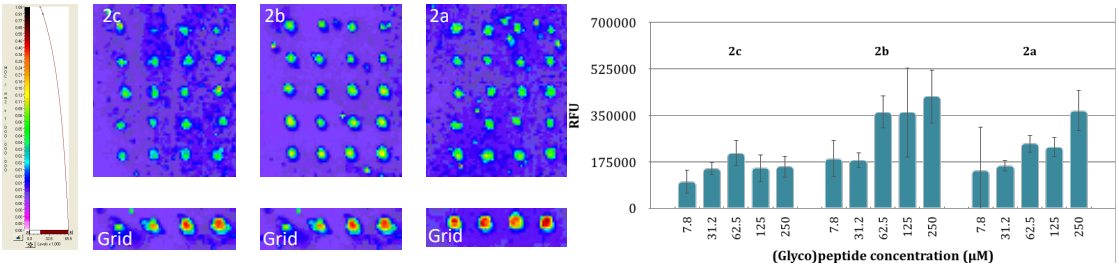


Figure S2. Binding studies of non-glycosylated APDTRP **2c** and glycopeptides **2b** and **2a** with the commercially available antibody VU-11E2, using a microarray platform. Compounds were printed onto an "Antibody chip" microarray in quadruplicate. Relative fluorescence units (RFU) due to the binding of the Cy3-labeled secondary antibody were measured and represented as mean values in a bar chart.

Table S1. Comparison of the experimental and experiment-guided MD simulations derived distances for glycopeptide **2a**.

Distance	Experimental (Å)	Experiment-guided MD (Å)
NH_{Asp}–$H\alpha_{Asp}$	2.8	2.9
NH_{Asp}–$H\alpha_{Pro}$	2.0	2.2
NH_{Thr}–$H\alpha_{Thr}$	2.8	2.9
NH_{Thr}–$H\alpha_{Asp}$	2.1	2.2
NH_{Arg}–$H\alpha_{Arg}$	2.8	2.9
NH_{Arg}–$H\alpha_{Thr}$	2.2	2.4
NH_{Thr}–$NH_{iminosugar}$	3.0	2.9

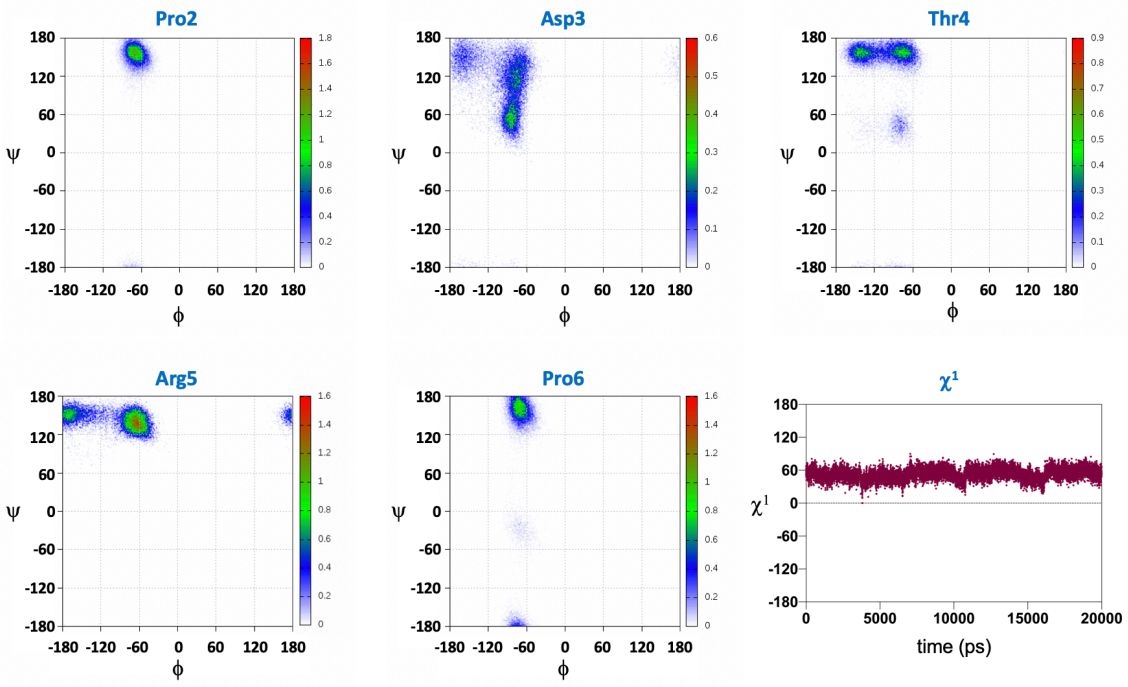


Figure S3. ϕ/ψ distribution of all amino acids in **2a** derived from 20 ns MD-tar simulations, together with the distribution of the side chain (χ^1) for Thr4.

Table S2. Data collection and refinement statistics. Values in parentheses refer to the highest resolution shell. Ramachandran plot statistics were determined with PROCHECK.

	Complex 2a/scFv-1SM3
Space group	$P2_12_12_1$
Wavelength (Å)	0.97
Resolution (Å)	20.00-1.998 (2.11-1.998)
Cell dimensions (Å)	a = 35.27 b = 68.69 c = 90.39
Unique reflections	15528
Completeness	99.9 (100)
R_{pim}	0.076 (0.342)
Mn(I) half-set correlation CC(1/2*)*	0.987 (0.758)
$I/\sigma(I)$	9.1 (4.8)
Redundancy	10.3 (10.3)
R_{work} / R_{free}	0.175/0.237
RMSD from ideal geometry, bonds (Å)	0.011
RMSD from ideal geometry, angles (º)	1.528
<*B*> protein (Å^2)	46.08
<*B*> glycopeptides (Å^2)	51.76
<*B*> solvent (Å^2)	32.03
<*B*> ethylenglycol (Å^2)	59.72
Ramachandran plot: Most favoured (%) Additionally allowed (%) Disallowed (%)	 95.07 3.14 1.79
PDB ID	6TGG

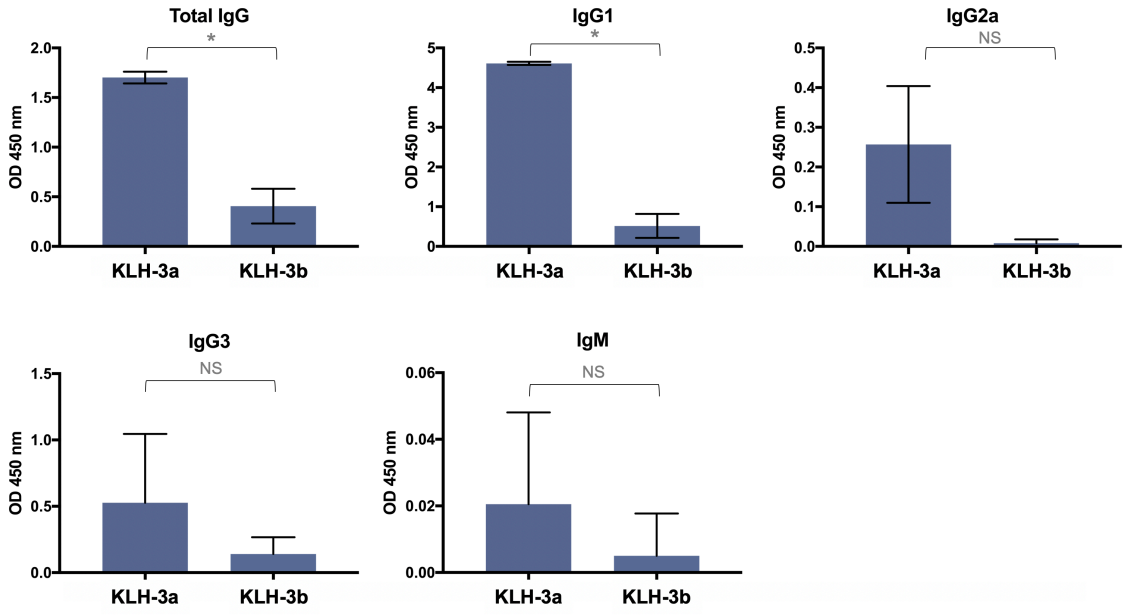


Figure S4. a) Total IgG and sub-typing (IgG1, IgG2a and IgG3) as well as IgM anti-MUC1 antibodies titrations after immunization with either vaccine **KLH-3a** or **KLH-3b**. ELISA plates were coated with natural MUC1-like glycopeptide **3b**. In all plots the horizontal lines indicate the mean for the group of mice (n=3). Asterisk indicates statistically significant difference ($P < 0.05$).

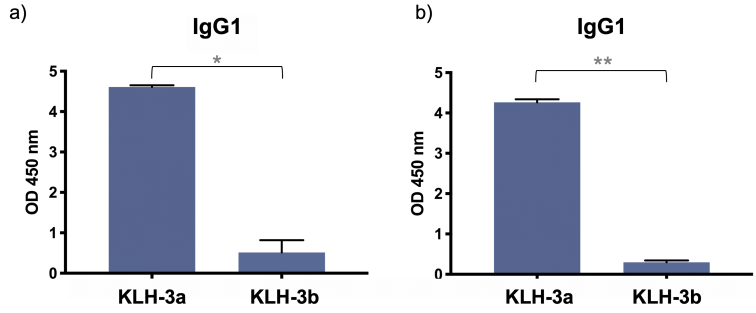


Figure S5. a,b) IgG1 sub-typing anti-MUC1 antibodies titrations after immunizations with either vaccine **KLH-3a** or **KLH-3b**. The ELISA plates were coated with **3b** (a) or naked peptide **3c** (b). This result indicates that the elicited antibodies target mainly the peptide and not the glycan moiety. Asterisk indicates statistically significant difference (** $P < 0.01$,* $P < 0.05$).

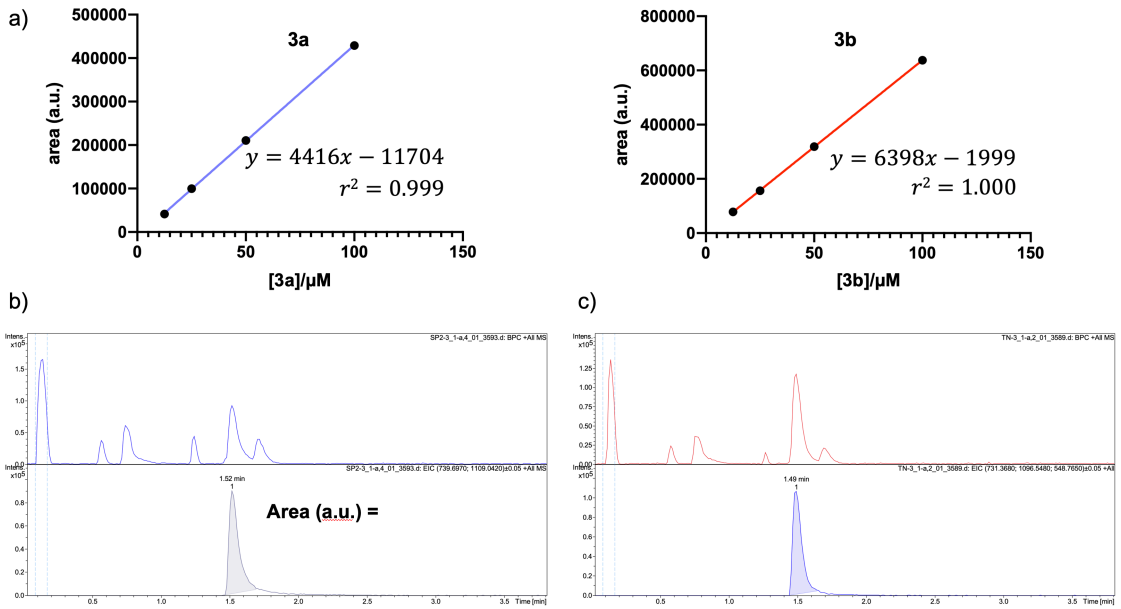


Figure S6. a) Calibration curves for glycopeptides **3a** and **3b** using UPLC/MS.[S2] A molecular weight for KLH of 8,000 KDa was used for the calculations.[S3] b) UPLC-MS/MS spectra of conjugation of KLH with glycopeptide **3a** (upper panel), together with the extracted ion chromatogram (glycopeptide **3a**, lower panel). c) UPLC-MS/MS spectra of conjugation of KLH with glycopeptide **3b** (upper panel), together with the extracted ion chromatogram (glycopeptide **3b**, lower panel). bioZen 1.7 µm Peptide XB-C18 column (100 mm x 2.1 mm), Grad: water/acetonitrile (99:1) → (0:100), 6.0 min. Flow rate: 0.35 mL/min.

Reference

[S1] E. M. Sánchez Fernández, C. D. Navo, N. Martínez-Sáez, R. Gonçalves-Pereira, V. J. Somovilla, A. Avenoza, J. H. Busto, G. J. L. Bernardes, G. Jiménez-Osés, F. Corzana, J. M. García Fernández, C. Ortiz Mellet and J. M. Peregrina, *Org. Lett.*, 2016, **18**, 3890–3893.

[S2] J. K. Gathuru, F. Koide, G. Ragupathi, J. L. Adams, R. T. Kerns, T. P. Coleman and P. O. Livingston, *Vaccine*, 2005, **23**, 4727–4733.

[S3] X. Wu, Z. Yin, C. McKay, C. Pett, J. Yu, M. Schorlemer, T. Gohl, S. Sungsuwan, S. Ramadan, C. Baniel, A. Allmon, R. Das, U. Westerlind, M. G. Finn and X. Huang, *J. Am. Chem. Soc.*, 2018, **140**, 16596–16609.