Supporting Information

Directed Evolution of Multivalent Glycopeptides Tightly Recognized by HIV Antibody 2G12

Satoru Horiya,[†] Jennifer K. Bailey,[†] J. Sebastian Temme,[†] Yollete V. Guillen Schlippe,[‡,§]* and Isaac J. Krauss[†]*

[†]Department of Chemistry, Brandeis University, Waltham, Massachusetts 02454-9110, United States
[‡]Department of Molecular Biology, and Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, United States
[§]Present address: 3B Pharmaceuticals GmbH, Magnusstrasse 11, 12489 Berlin, Germany

Table of contents

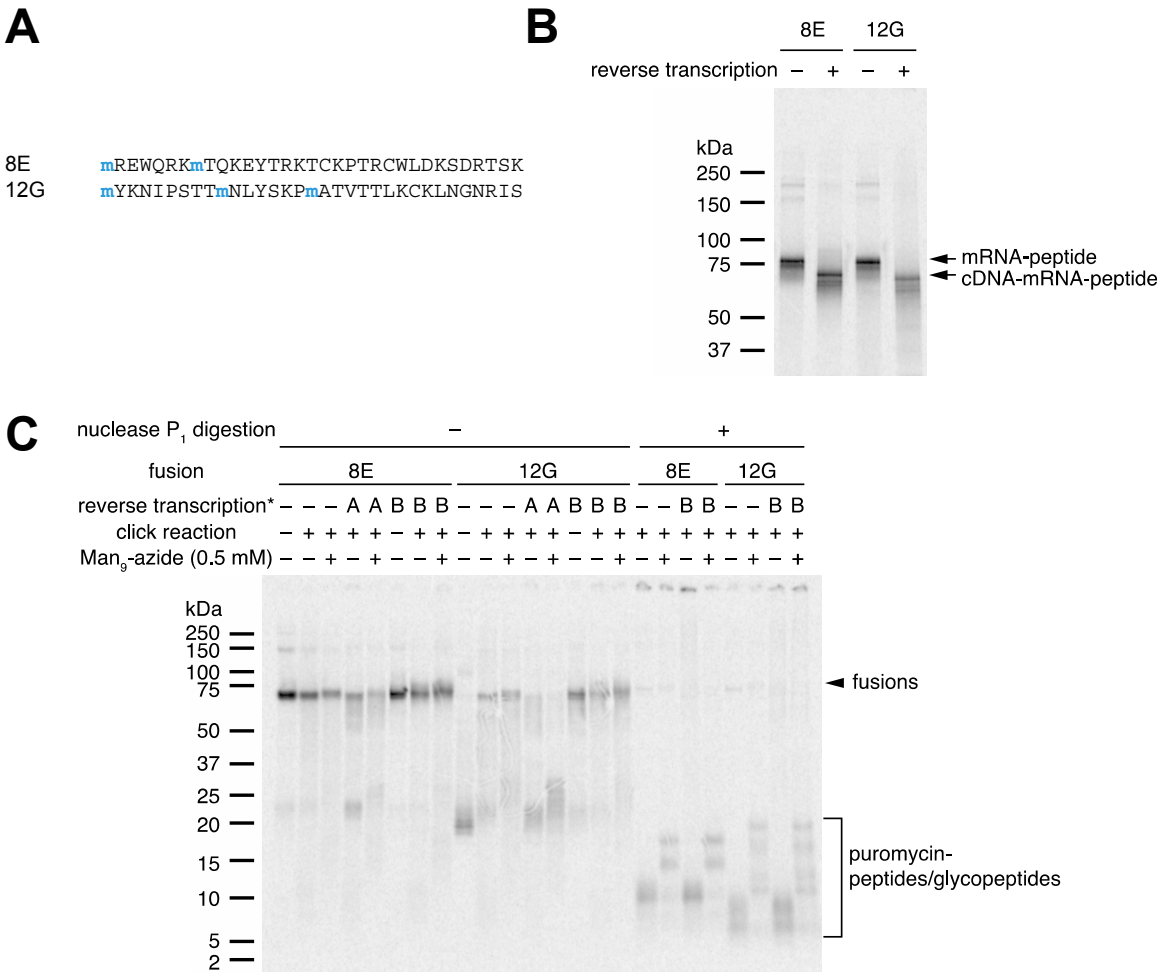


Figure S1. Examination of the integrity of "click" glycosylated fusions before or after reverse transcription. (**A**) The peptide sequences used in the experiments. The sequences were obtained by cloning non-selected library DNA. (**B**) SDS-PAGE analysis of the reverse transcribed mRNA-peptide fusions labeled with ^{35}S-cysteine using a 7.5 % precast gel (Bio-Rad). In this condition, cDNA-mRNA-peptide fusions migrate faster than mRNA-peptide fusions. (**C**) SDS-PAGE analysis of fusion integrity. mRNA-peptide fusions were reverse transcribed after (A) or before (B) click reaction. The click reaction was done using a slightly different condition described in Methods, in which 30-40 nM fusions shown in (**B**) were incubated with 100 mM HEPES-KOH (pH 7.6), 0.02 % Triton X-100 (v/v), 1 mM $CuSO_4$, 2 mM THPTA, 5 mM aminoguanidine hemisulfate and 5 mM sodium ascorbate in the presence (+) or absence (—) of 0.5 mM Man_9-azide, which is a lower concentration compared to the regular condition described in Online Methods, in argon-filled microtube at room temperature for 3 hours. The fusions without click reaction (—) were incubated with 100 mM HEPES-KOH (pH 7.6), 0.09 % Triton X-100 (v/v) at room temperature in argon-filled microtubes for 3 hours as well. The fusions with or without nuclease P_1 digestion were separated by 10% Criterion XT Bis-Tris precast gel (Bio-Rad) with XT MES Running Buffer (Bio-Rad). The gels were analyzed by autoradiography.

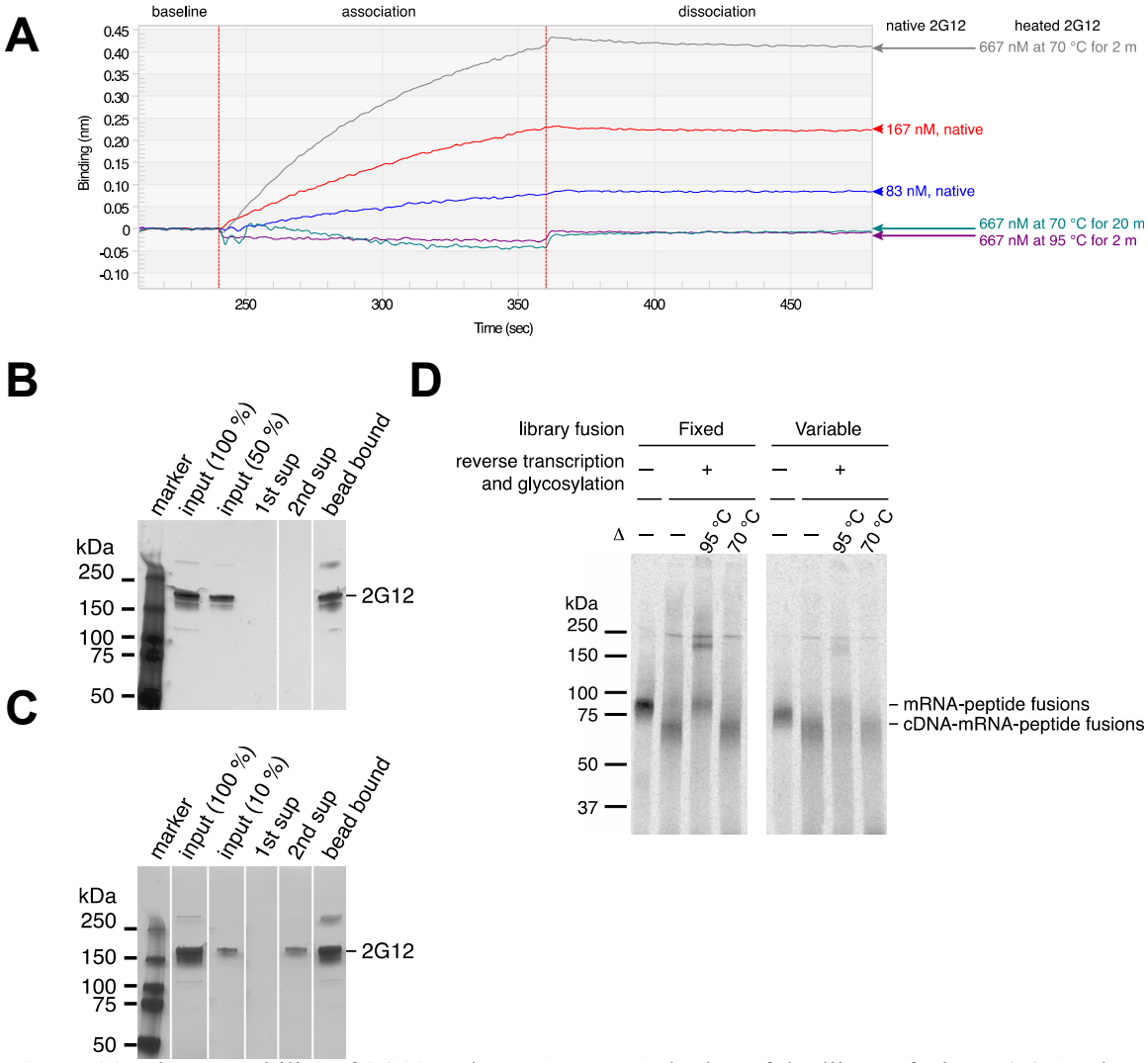


Figure S2. Thermostability of 2G12 and cDNA-mRNA duplex of the library fusions. (**A**) Analyses of the interaction between 2G12 and gp120 using BLitz with Dip and Read Ni-NTA Biosensors (ForteBio). 25 μg/ml His_6-tagged gp120 (HIV-1 JRFL) (Immune Technology) was loaded on equilibrated Ni-NTA sensors for 3 min and then the sensors were equilibrated with selection buffer for 30 s. The gp120-loaded sensors were used to associate with the native 2G12 or the 2G12 heated at 70 °C or 95 °C, chilled on ice for 5 min and incubated at room temperature in selection buffer before loading. The time of the 2G12 association and dissociation steps in selection buffer were 2 min. (**B**) 100 nM 2G12 was incubated with 6 mg/ml protein G magnetic beads in selection buffer for 1 hr and the supernatant was removed (1st sup). The beads were resupended in selection buffer and heated at 70 °C for 30 min, chilled on ice for 5 min, and incubated at room temperature for 10 min. Then, the supernatant (2nd sup) was removed and the 2G12 bound to the beads were eluted out by boil in Laemmli sample buffer (bead bound). The supernatants and bead bound fraction were analyzed with the controls of amounts of input to the beads using 4-20 % SDS-PAGE without addition of reducing agent. The gel was silver-stained. (**D**) Same experiment as (**B**) using 12 mg/ml protein A magnetic beads except that the incubation time at 70 °C was 20 min. (**D**) The ^{35}S-cysteine labeled library fusions (50 fmol) for selection round 1 were heated at 95 °C for 2 min or 70 °C for 20 min and chilled on ice before applied to 7.5 % SDS-PAGE. The gel was analyzed by autoradiography.

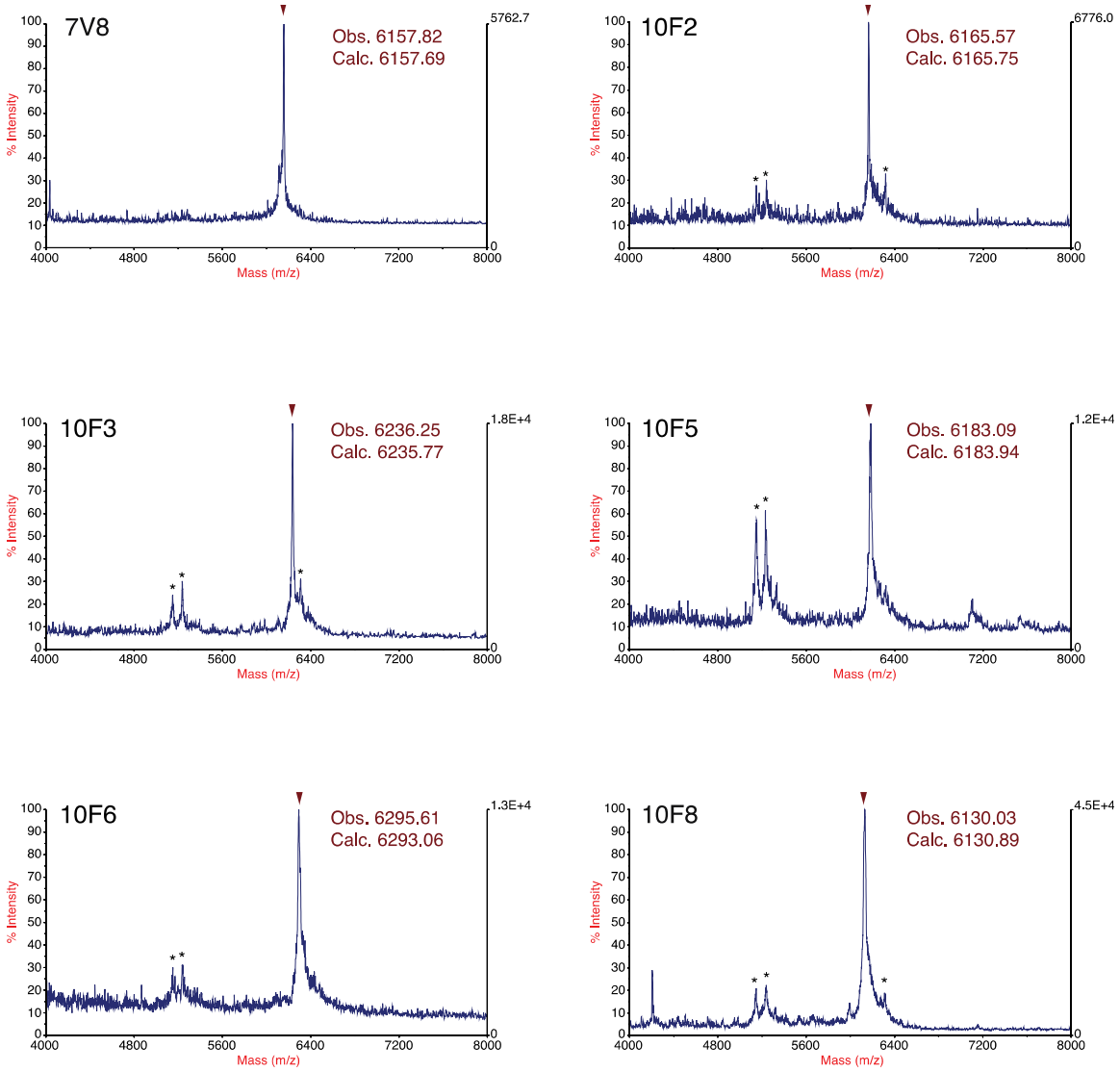


Figure S3. The MALDI-TOF-MS spectra of the translated peptides. An arrow head indicates the peak of the average mass of the peptide ion. The asterisks show the three representative impurities derived from the PURE system as observed in the "–RNA" control, in which translation reactions were done without mRNA and purified with Ni-NTA. The MALDI-TOF-MS was done in linear mode and calibrated with external standards. All observed masses (Obs.) corresponded to their calculated masses (Calc.) within a range of the potential error (0.05 % accuracy) allowed in this system according to the instruction manual.

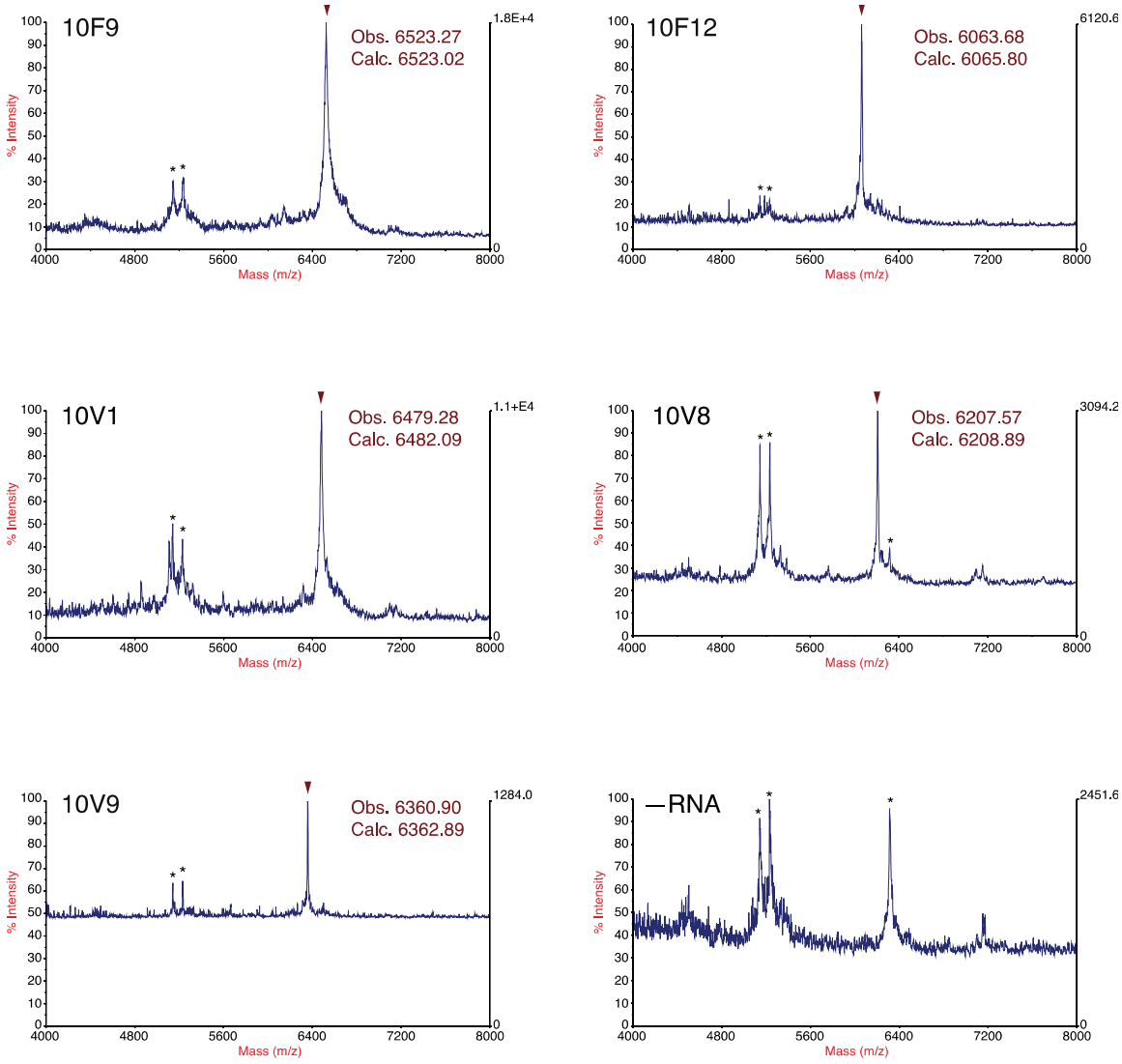


Figure S3 (*Continued*)

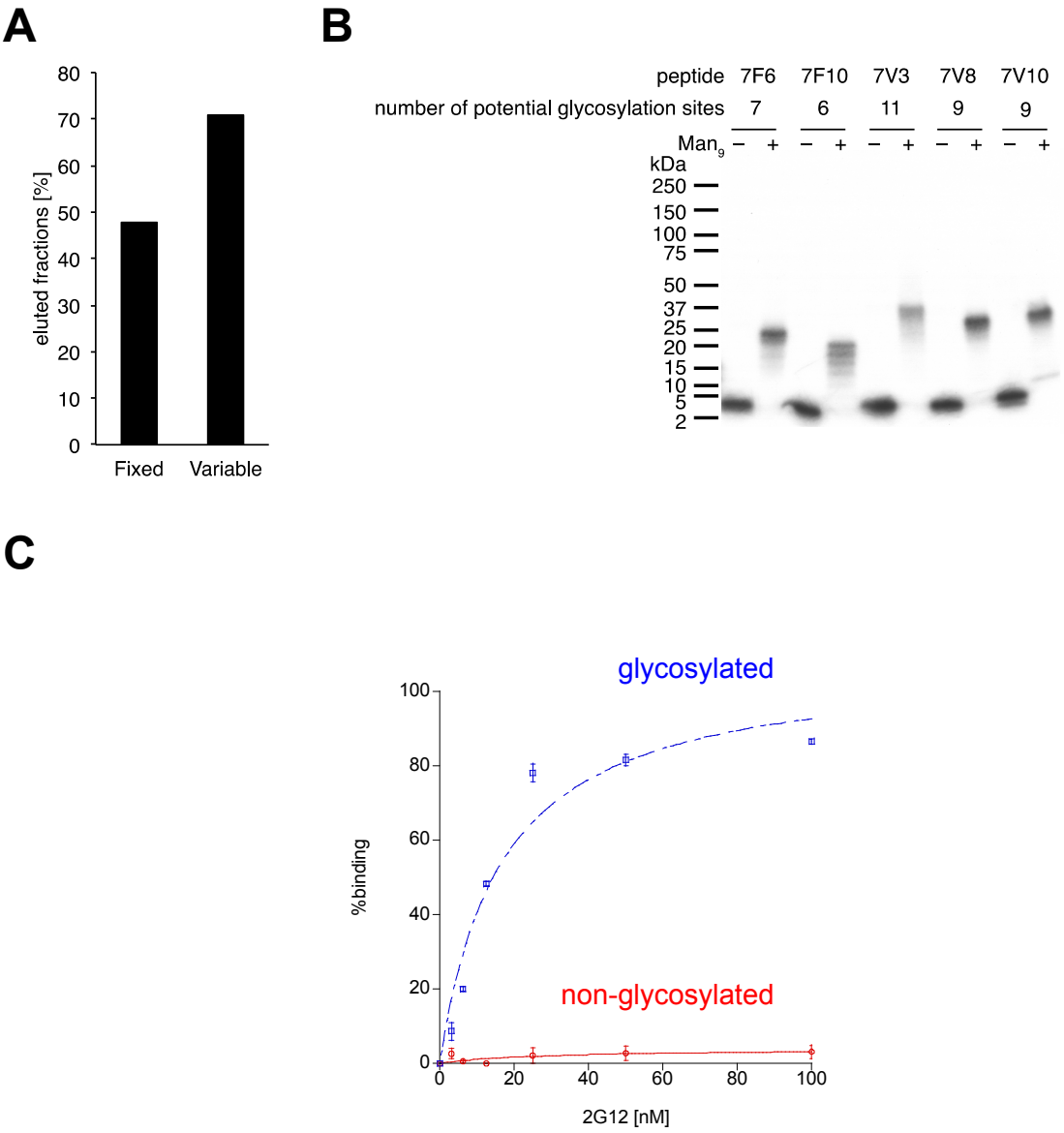


Figure S4. Analysis of the selected peptides from round 7. (**A**) Percentages of library fusions (0.18 pmol) for selection round 7, bound to protein G magnetic beads in the presence of 100 nM 2G12 without mannose and then heat eluted, which was the same condition as selection round 1. (**B**) SDS-PAGE analyses of click reaction of the peptides in the absence (–) or presence (+) of Man_9-azide. Peptides were separated in 4-20% gradient gel (Bio-Rad). The molecular marker was Precision Plus Protein Dual Xtra Standards (Bio-Rad). The peptides were labeled with ^{3}H-histidine and the bands were visualized by fluorography. (**C**) Binding curves of the interaction between 2G12 and glycosylated or non-glycosylated 7V8.

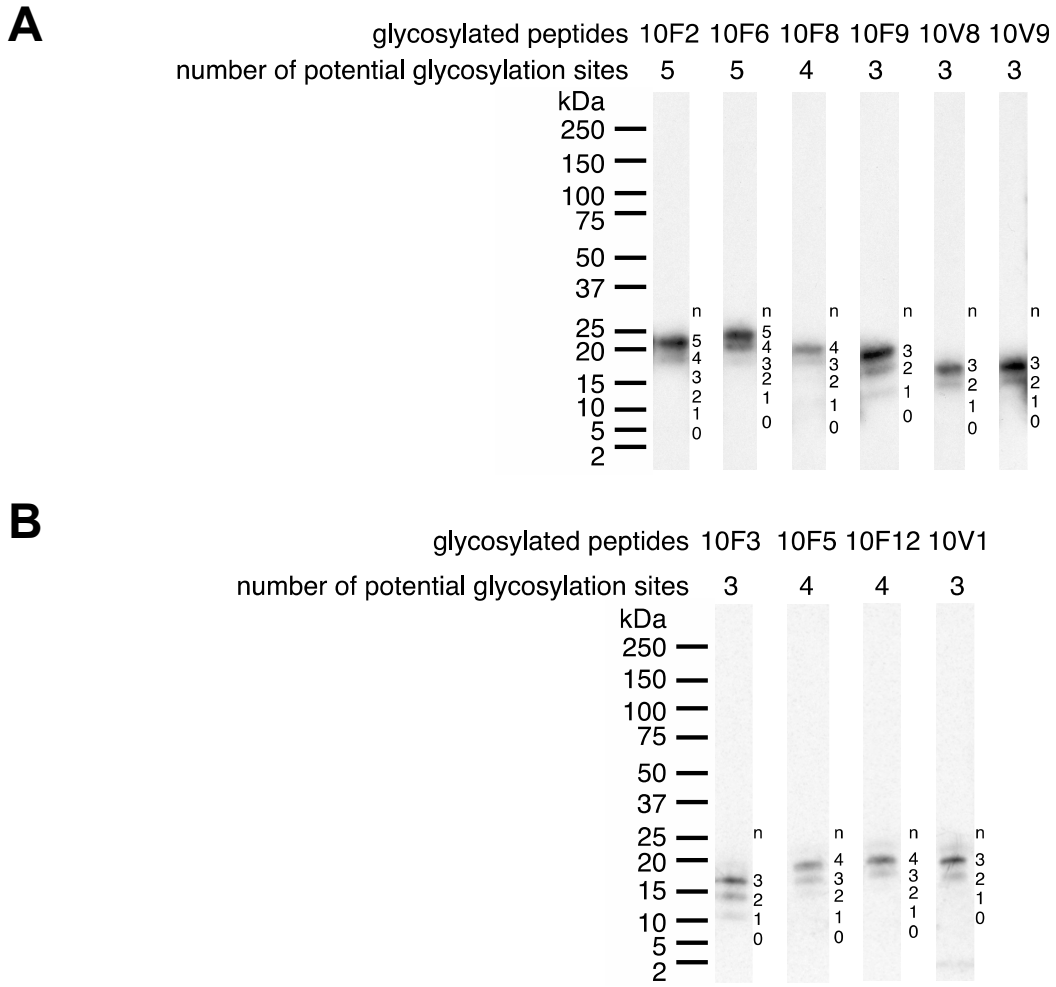


Figure S5. SDS-PAGE analyses of the glycosylated peptides selected in round 10. The numbers "n" on the left of the gel image indicate the putative number of glycans the peptides. (**A**) The peptides were labeled with ^{3}H-histidine and the bands were visualized by fluorography. (**B**) The peptides were labeled with ^{3}S-cysteine and the bands were visualized by autoradiography.

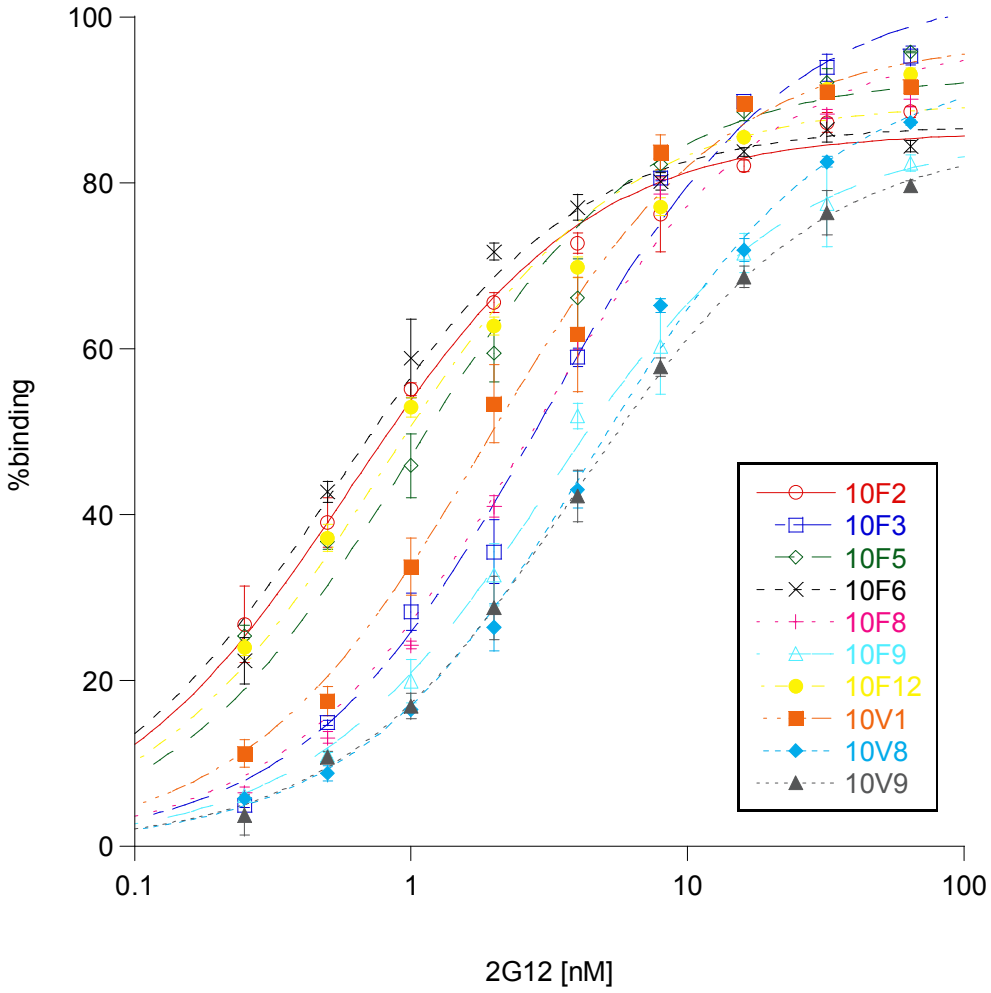


Figure S6. Binding curves of the interaction between 2G12 and selected glycopeptides. The error bars represent standard error.

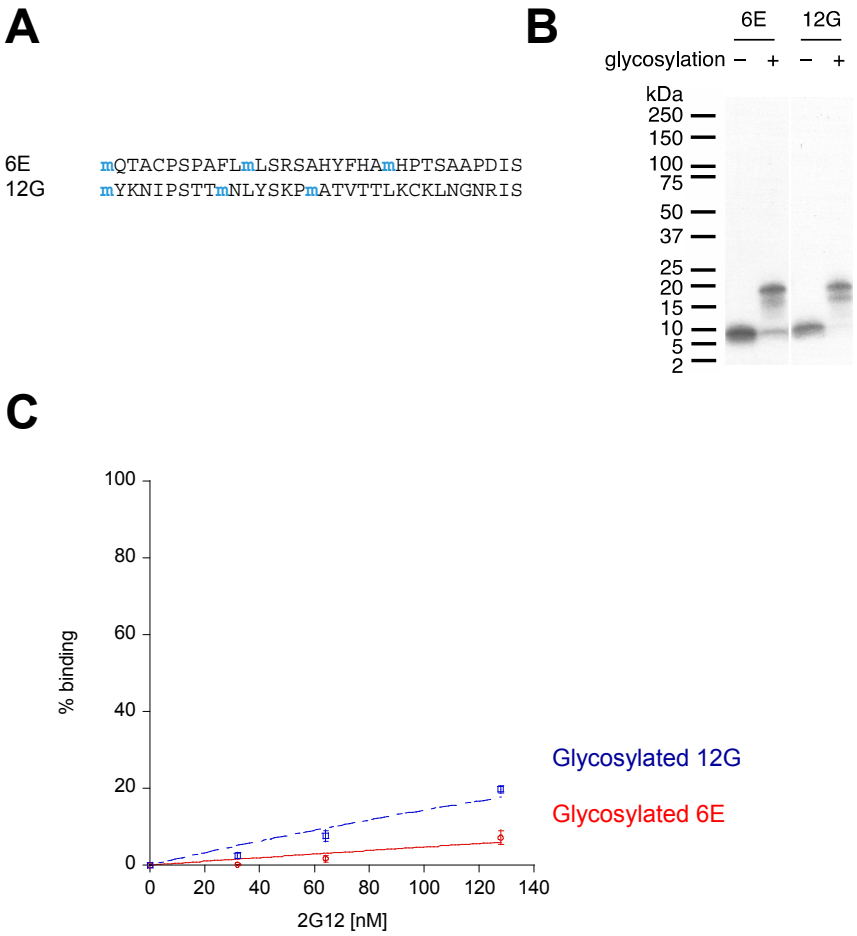
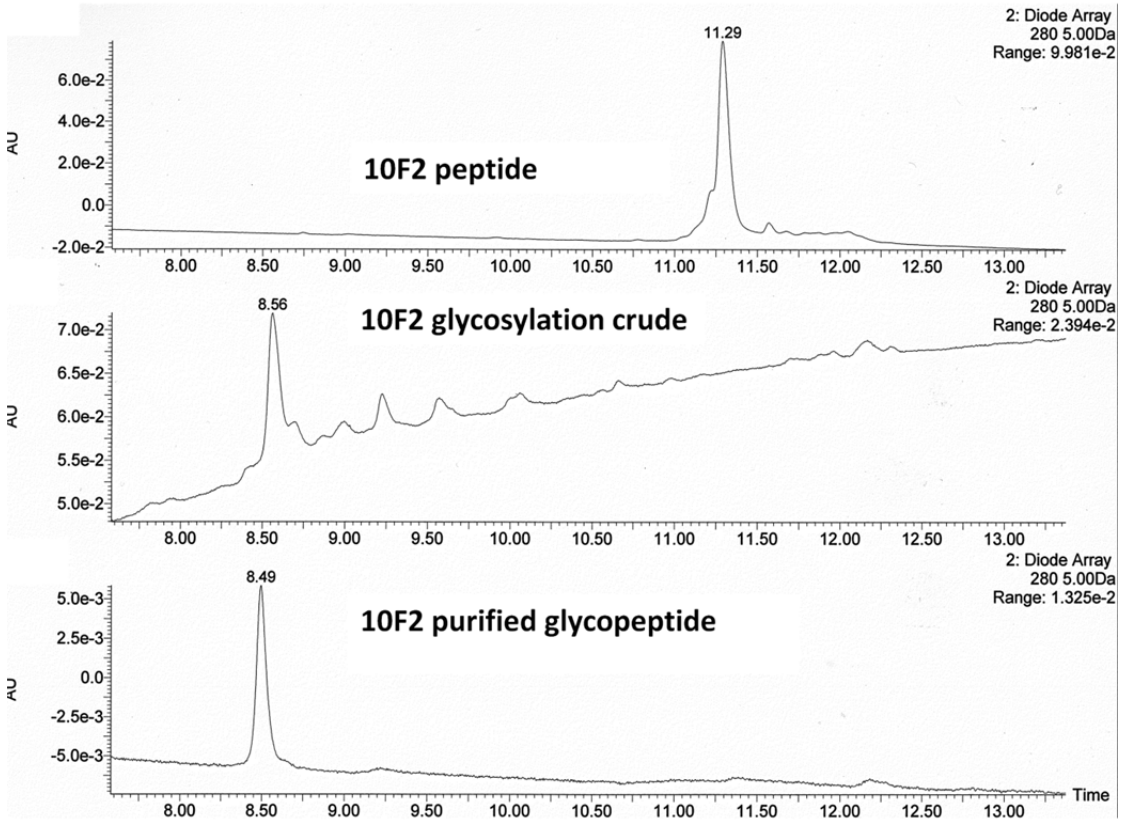


Figure S7. Binding study of individual glycopeptides obtained from the starting libraries before selection. (**A**) The peptide sequences used in the experiments. The sequences were obtained by cloning non-selected library DNA. Both peptides were followed by a flexible linker, a His_6-tag and a FLAG-tag (GSGSLGHHHHHHRDYKDDDDK) for purification and radiolabeling purposes. (**B**) SDS-PAGE analysis of the click reaction of the non-selected peptides. Both peptides were transcribed and translated from PCR-amplified DNA in Pure System that contains 0.02 mg/mL T7 RNA polymerase and additional 1 mM each NTP. The peptides were labeled with ^{3}H-histidine and the bands were visualized by fluorography. (**C**) The binding curves for clones 6E and 12G. Data were obtained as described in Materials and Methods except that the 2G12 concentration was 0, 32, 64 or 128 nM and the amount of protein G magnetic beads was doubled to accommodate the high 2G12 concentrations used.

Figure S8. LC/LRMS analysis of 10F2 glycosylation
UV chromatograms:



LRMS of unglycosylated 10F2 peptide:

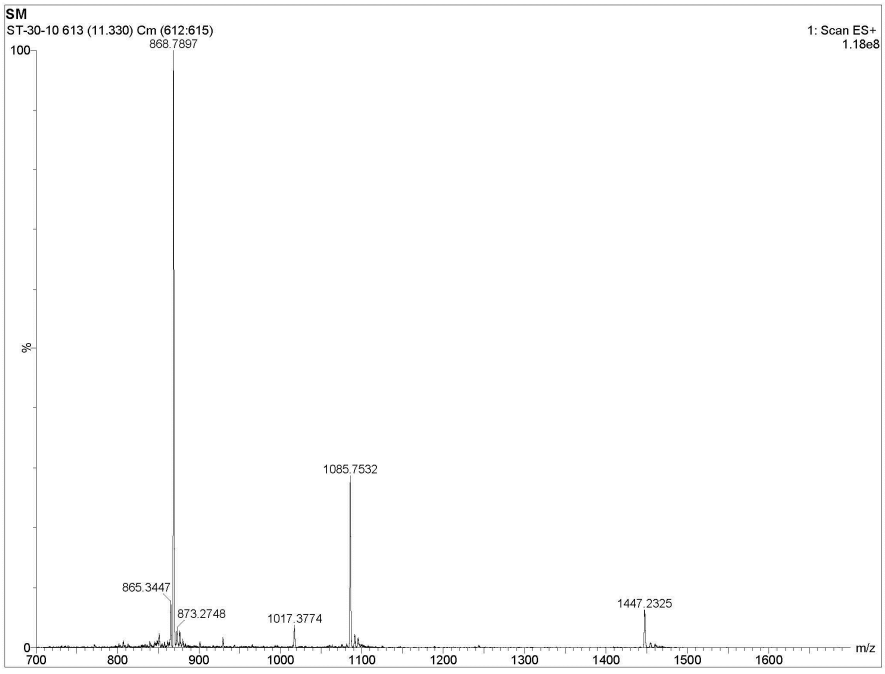
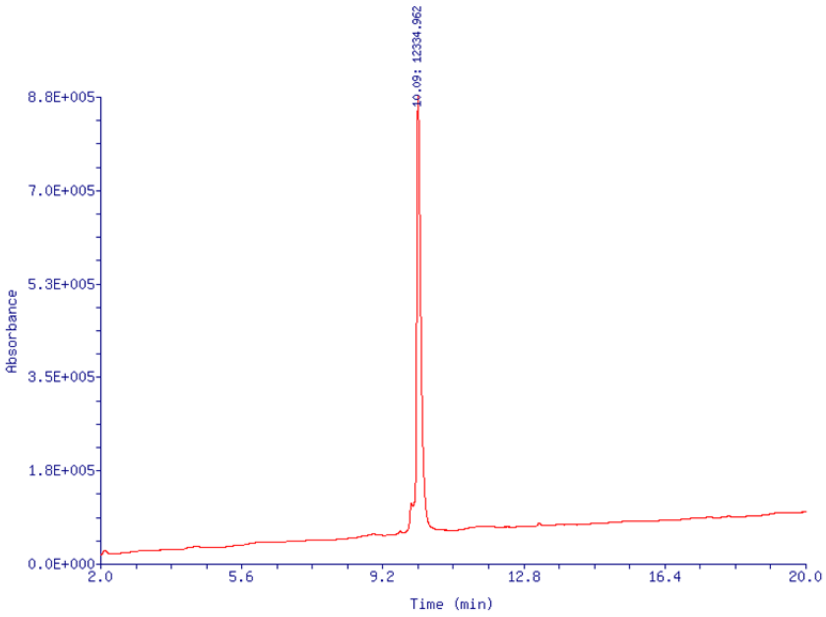
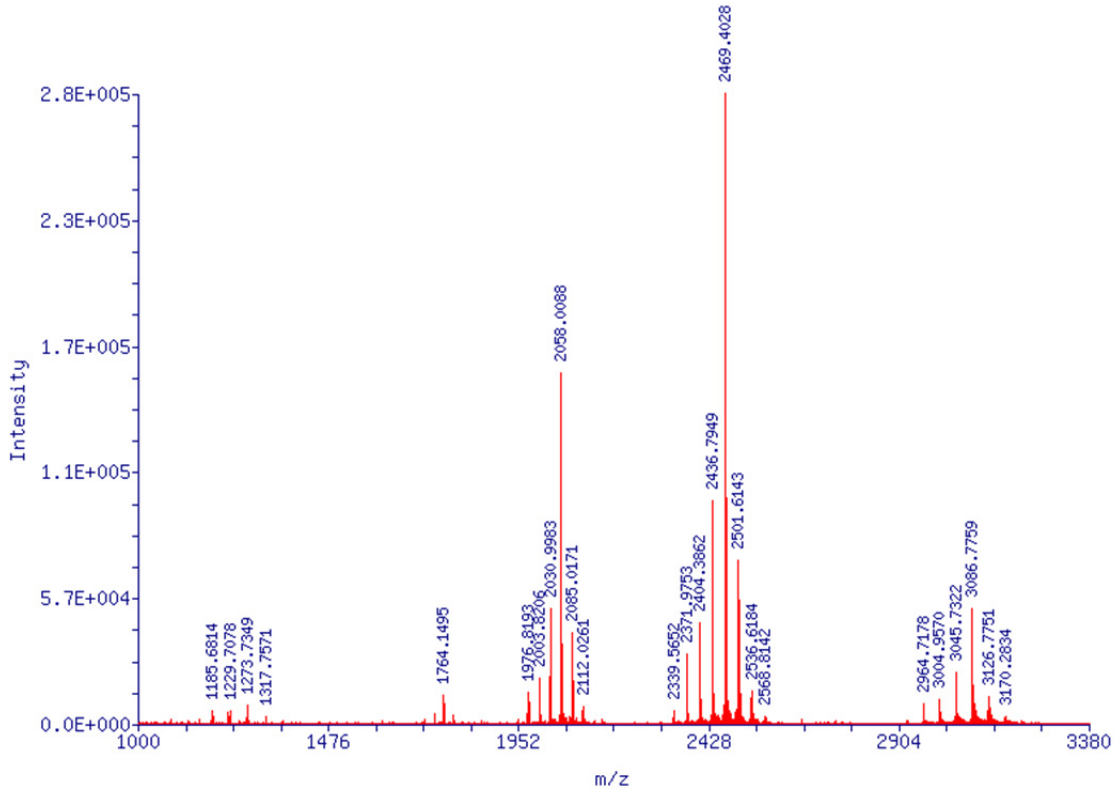


Figure S9. LC/HRMS Analysis of Glycopeptide 10F2
Analysis performed by Novatia, LLC. Method: 2.1x150mm Aeris C18 3.6u 5% B @ 0min, 5%B @ 1 min, 60% B @ 20 min, 60°C, A=0.05% TFA in water, B=0.05% TFA in ACN
UV chromatogram



Raw ESI-Orbitrap-MS of 10.09 min peak



Glycopeptide ion	Calculated Mass	Observed Mass
$[M+6H]^{6+}$	2058.0250 (avg.)	2058.0088
$[M+5H]^{5+}$	2469.4280 (avg.)	2469.4028
$[M+4H]^{4+}$	3086.5325 (avg.)	3086.7759
M (deconvolution)	12334.980	12334.962 ± 0.128

Deconvoluted MS

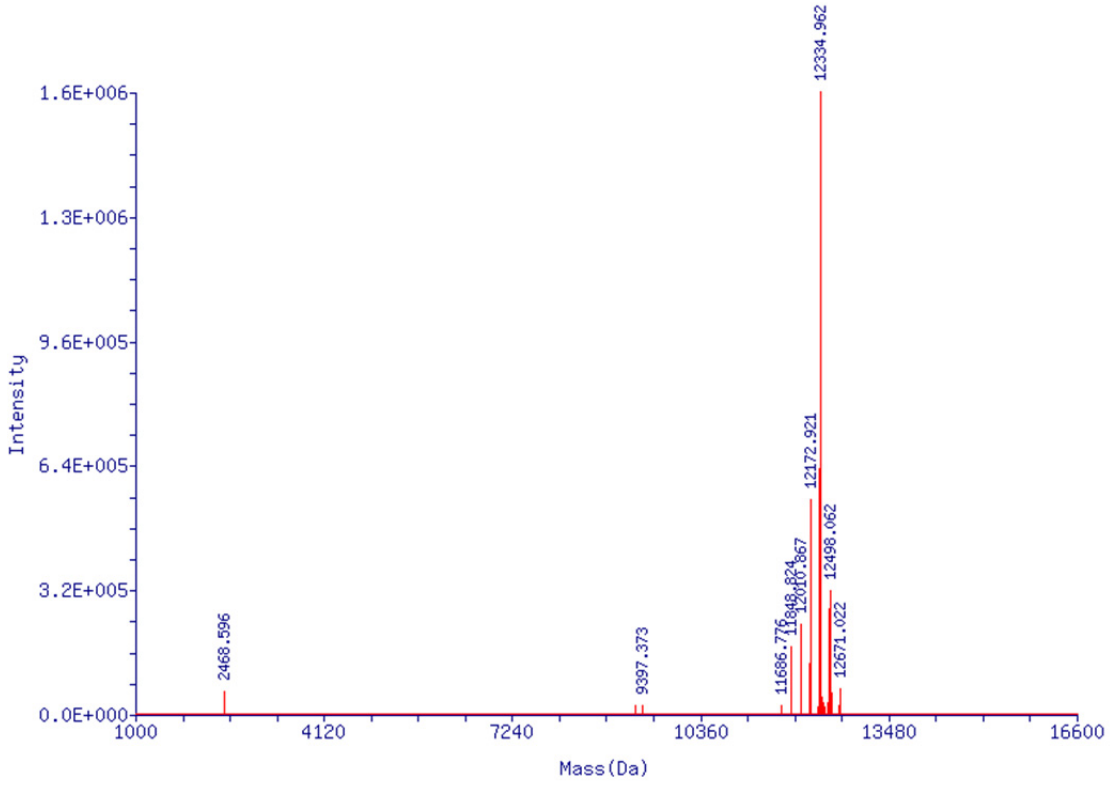


Table S1. Selected clones from round 7.

Library	Clone designation	Sequence	Number of potential glycosylation site
Fix	7F1	mYYLSVYPSYSmYFSSSYVVWPmPGHRLLIGLE	3
	7F10	mmEHKLTmLPLmSTDIFLVLLmmFGTTITQVSL	6
	7F6	mYLPDWmLKSLmLSKWRLPEmFmSPFmLELHmS	7
	7F8	mLTNITLQmSRmHLLWLHmHDLmmDLCRImLRS	7
	7F5	mVLTPTTKmmVmQSPmYFmRSNmLSKmYDYQRL	8
	7F11	mmImNSmRIDVmmSNFVHAKSTmVGQRHmGGVG	8
	7F12	mSmTmQFSHFWmRHmWESmNRWmLARTmDTPID	8
	7F16	mmCHCLPSHYmmLRFCPmTGSVmDmGLKRmVYH	8
	7F2	mAKFDEmmAmLNmSRmSSYLmmLmTGRTWPH	9
	7F15	mTFEmLPRSDSmRmLTmPmmHRmYmIYRGYSNR	9
	7F17	mSYSmSPRDPNmmIKFLmSRTmmRNPmNVIGSm	9
Variable	7V12	mHISTNCmPWRYWSIICmmPTWKTVHQmmKTKD	6
	7V6	mCSRKmACLSRANLmRmRSmmKRRmTmNTSFTm	9
	7V8	mIRmRTPTSRLmSTmRGmTmNmTSmITPRNDmI	9
	7V10	mTPFTmAYmTRRKPmmFPImHRmKSRTPLmmGK	9
	7V3	mKmNmRIWNPmmNNWSmDTASmLRLmSWmLNmm	11
	7V4	mTSImDNTmmLSVNmNRmKINRTLmmmmmHmSTm	12
	7V9	mCmKmYAPNmYDLmPmRmHWmPNVLmPLmSmRm	12

Table S2. Selected clones from 10.

Library	Representative clone designation	Sequence[a]	number of potential glycosylation site	number of appearance in 24 clones
Fix	10F5	mSPHLPVLLCKmVLNDGRRIVQmSCELPmVRRS	4	4/24
	10F2	mHPYNTSRTSAmmAALKmQVTDmYALALFHRIL	5	3/24
	10F12	mCYVTVIPAmNmPEARLGIVCHmPGIRRGKALY	4	2/24
	10F6	mLmFIRIYPTRmQYVYHAPLLTmVRmSPTGPLI	5	1/24
	10F16	mVRSAAVDTSPmTSSSQNAILLmFSYDVCLFDL	3	1/24
	10F20	mIALTSNCYLNmGPRIFRYDVGLTQLCQGRRRS	2	1/24
	10F3	mDTLHLKQIGGmPNCITQQDVRmTSIPYTYTWP	3	4/24
	10F23	mDTLHLKQIGVmPNCITQQDVRmTSIPYTYTWP	3	1/24
	10F8	mLLKmVDQSRLmPVPGIGVTLHmRSIPYSYLPI	4	4/24
	10F9	mRSTLNSLEYRmQYATEDPRIRmASIPYTYWWP	3	2/24
	10F18	mFSTANIYGAPmNTDmRLEHRQmKSIPYTYYWS	4	1/24
Variable	10V1	mATKTNCKREKTmDNHVTImRSIPWYTYRWLPN	3	14/24
	10V14	mATKTNCKREKTIDNHVTImRSIPWYTYRWLPN	2	2/24
	10V2	mATKTNFKREKTmDNHVTImRSIPWYTYRWLPN	3	1/24
	10V6	mATRTNCKREKTmDNHVTImRSIPWYTYRWLPN	3	1/24
	10V11	mATKTSCKREKTmDNHVTImRSIPWYTYRWLPN	3	1/24
	10V9	mTSIPYTYLNRSLWTNYRVNSWSmSKNVNVmPL	3	4/24
	10V8	mVLPTIISTNVNPFRmLSIPTYTYLmPITWGEI	3	1/24

[a]Consensus sequences are highlighted in yellow. The gray highlights in 10F23 and in 10V14, 10V2, 10V6 and 10V11 indicate the mutated amino acid from their potential parent sequences, 10F3 and 10V1, respectively.

Table S3. Peptide synthesis detailed conditions

Summary of coupling conditions:

	Coupling Reagent	AA/Coupling Conc.	mmol AA	Base Conc.	Coupling Flow Rate	Coupling Time
Most amino acids*	HATU	0.33 M	1	0.95 M	6 ml/min	30 sec
Cys(StBu)	HATU	0.33 M	0.15	0.86 M	N/A	10 min
His(Trt)	HATU	0.33 M	1	0.29 M	6 ml/min	30 sec
HPG	HATU	0.3 M	0.15	0.86 M	N/A	10 min
Gly	HBTU	0.33 M	1	0.95 M	6 ml/min	30 sec

*All amino acids were coupled according to this procedure, except Cys, His, HPG, and Gly

The general coupling procedure follows Pentulute's procedure, ref. 21, with a few modifications. After coupling and after Fmoc deprotection, the resin was washed with 20 ml of DMF at a flow rate of 10 ml/min. Fmoc deprotection was carried out at flow rate of 10 ml/min. Fmoc deprotection solution was 20% piperidine in DMF up until the coupling of aspartic acid, after which a solution of 19% piperidine/1% formic acid in DMF was used to prevent aspartimide formation.

Fmoc-Cys(StBu)-OH and Fmoc-HPG-OH were coupled outside of the reactor. Swelled resin was transferred to a 15 mL conical tube with a stir bar. 0.15 mmol amino acid and 0.15 mmol HATU were dissolved in 425 µL DMF, and 75 µL DIPEA was added just before adding to resin. The coupling reaction was allowed to take place under nitrogen for 10 minutes at 60 °C with stirring. After the reaction, the resin was transferred to the reactor for washing and Fmoc deprotection.
After peptide synthesis was complete, the N-terminus was formylated. The swelled resin was transferred to a 15 ml conical tube with a stir bar. 0.25 mmol 4-nitrophenylformate was dissolved in 632 µl DMF (0.33 M final). 125 µL DIPEA (0.86 M final) was added just before addition. Formylation was allowed to occur at 60 °C for 8 minutes while stirring under nitrogen. Next, the supernatant was removed, fresh reagents were added, and formylation was repeated. This was done again for a total of 3 – 8 minute periods. After the reaction, the resin was washed with DMF 5 x 10 ml and DCM 3 x 10 ml.

The peptide was cleaved from the resin with 10 ml of a cleavage cocktail B containing 87.5/5/5/2.5 TFA/phenol/water/TIPS. The resin and cocktail were tumbled at room temperature for 90 minutes. The resin was filtered and washed 3 x 4 ml DCM. The filtrate was concentrated by rotary evaporation and transferred to a 15 ml conical tube. The peptide was triturated with 5 x 10 ml cold ether to give 35 mg crude peptide.

4.5 mg of crude peptide was purified by HPLC on a Waters Symmetry300 C4 column (4.6x250mm, 5 µm particle size) following a 98% A/2% B to 58% A/42% B gradient over 60 minutes with a flow rate of 4 ml/min, where solvent A is water/0.1% formic acid and solvent B is acetonitrile/0.1% formic acid.

SI Table 4. BLI curve fit parameters

Conc (nM)	K_D(M)	ka ($M^{-1}s^{-1}$)	ka error	kd (s^{-1})	kd error	Rmax	Rmax error	Req
0.5	1.368e-9	1.106e5	3.989e2	1.513e-4	1.51e-6	1.797	0.2195	0.4810
1	1.368e-9	1.106e5	3.989e2	1.513e-4	1.51e-6	1.615	0.01391	0.6817
2	1.368e-9	1.106e5	3.989e2	1.513e-4	1.51e-6	1.662	0.01177	0.9866
4	1.368e-9	1.106e5	3.989e2	1.513e-4	1.51e-6	1.691	0.01057	1.26
8	1.368e-9	1.106e5	3.989e2	1.513e-4	1.51e-6	1.601	0.008136	1.367
16	1.368e-9	1.106e5	3.989e2	1.513e-4	1.51e-6	1.535	0.005597	1.414
32	1.368e-9	1.106e5	3.989e2	1.513e-4	1.51e-6	1.453	0.003017	1.394

Note S1

The Fixed library contains three potential Man_9-glycosylation sites encoded by AUG codons at the "fixed" positions 1, 12 and 33, and each site is followed by 10 random amino acid residues (X_{10}) encoded by NNN codon, where N is an equimolecular mixture of G, A, U and C, and in which AUG codon appears in a ratio of 1.6 % per position. The Variable library contains the N-terminal Man_9-glycosylation site encoded by the AUG start codon followed by 32 random amino acid residues (X_{32}). *X* is encoded by doped N'NS codons, where N' is a mixture of 40 % A, 20 % G, 20 % U and 20 % C, S is an equimolecular mixture of G and C, and in which AUG codon appear in a rate of 5 % per position.