Supplemental Information

Refining the N-termini of the SARS-CoV-2 Spike Protein and its Receptor Binding Domain

Robert A. D'Ippolito[1], Matthew R. Drew[1], Jennifer Mehalko[1], Kelly Snead[1], Vanessa Wall[1], Zoe Putman[1], Dominic Esposito[1], Caroline J. DeHart[1*]

[1]NCI RAS Initiative, Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA

*Corresponding Author: caroline.dehart@nih.gov

Summary

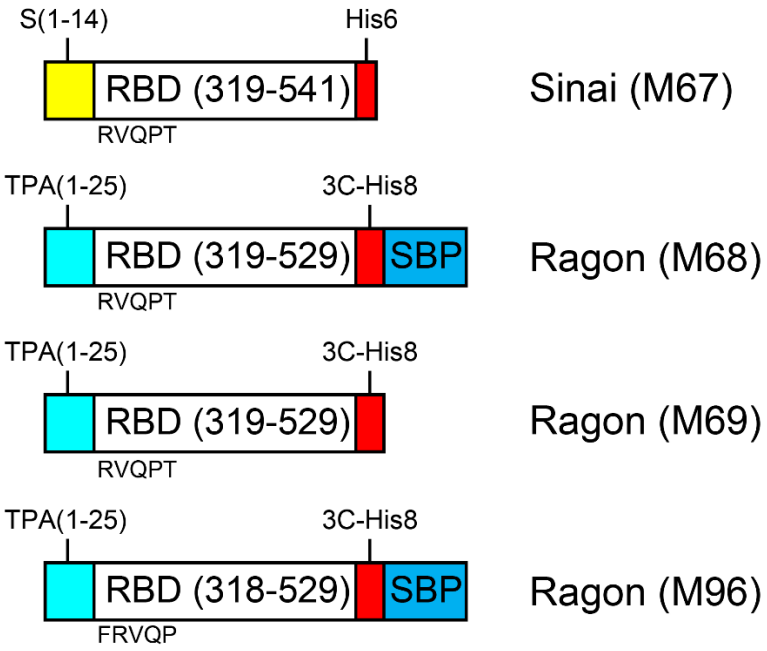

Figure S1. Structural features of the constructs analyzed. M67: S leader, Sinai 319-541, C-term His6 [Figure S3]. M68: TPA leader, Ragon 319-529, C-term 3C-His8-SBP [Figure S4]. M69: TPA leader, Ragon 319-529, C-term His8 [Figure S5]. M96: TPA leader, Ragon 318-529, C-term 3C-His8-SBP [Figure S8].

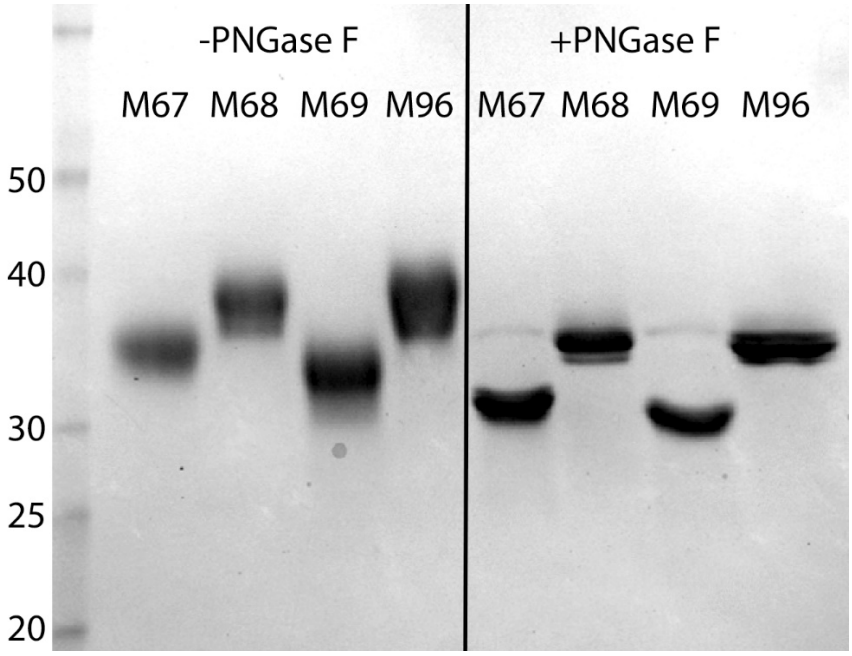

Figure S2. SDS-PAGE image of all constructs.

MFVFLVLLPL VSSQRVQPTE SIVRFPNITN LCPFGEVFNA TRFASVYAWN RKRISNCVAD YSVLYNSASF
STFKCYGVSP TKLNDLCFTN VYADSFVIRG DEVRQIAPGQ TGKIADYNYK LPDDFTGCVI AWNSNNLDSK
VGGNYNYLYR LFRKSNLKPF ERDISTEIYQ AGSTPCNGVE GFNCYFPLQS YGFQPTNGVG YQPYRVVVLS
FELLHAPATV CGPKKSTNLV KNKCVNFHHH HHH

Figure S3. The predicted Mt. Sinai RBD protein sequence from the construct DNA. Red: M1-Q14 of full-length spike signal sequence. Black: R319-F541; RBD of spike. Blue: His6 tag.

MDAMKRGLCC VLLLCGAVFV SPSASRVQPT ESIVRFPNIT NLCPFGEVFN ATRFASVYAW NRKRISNCVA
DYSVLYNSAS FSTFKCYGVS PTKLNDLCFT NVYADSFVIR GDEVRQIAPG QTGKIADYNY KLPDDFTGCV
IAWNSNNLDS KVGGNYNYLY RLFRKSNLKP FERDISTEIY QAGSTPCNGV EGFNCYFPLQ SYGFQPTNGV
GYQPYRVVVL SFELLHAPAT VCGPKKGAGS SLEVLFQGPG SGSSHHHHHH HHGGSGSSMD EKTTGWRGGH
VVEGLAGELE QLRARLEHHP QGQREP

Figure S4. The predicted Ragon RBD protein sequence from the construct DNA. Red: Tissue plasminogen activator (TPA) signal sequence. Black: R319-K529; RBD of spike. Blue: HRV3C protease cleavage site, His8, streptavidin-binding peptide combination tag.

MDAMKRGLCC VLLLCGAVFV SPSASVQPTE SIVRFPNITN LCPFGEVFNA TRFASVYAWN RKRISNCVAD
YSVLYNSASF STFKCYGVSP TKLNDLCFTN VYADSFVIRG DEVRQIAPGQ TGKIADYNYK LPDDFTGCVI
AWNSNNLDSK VGGNYNYLYR LFRKSNLKPF ERDISTEIYQ AGSTPCNGVE GFNCYFPLQS YGFQPTNGVG
YQPYRVVVLS FELLHAPATV CGPKKGHHHH HHHH

Figure S5. The predicted M69 RBD protein sequence from the construct DNA. Red: Tissue plasminogen activator (TPA) signal sequence. Black: V320-K529; RBD of spike. Blue: His8 tag.

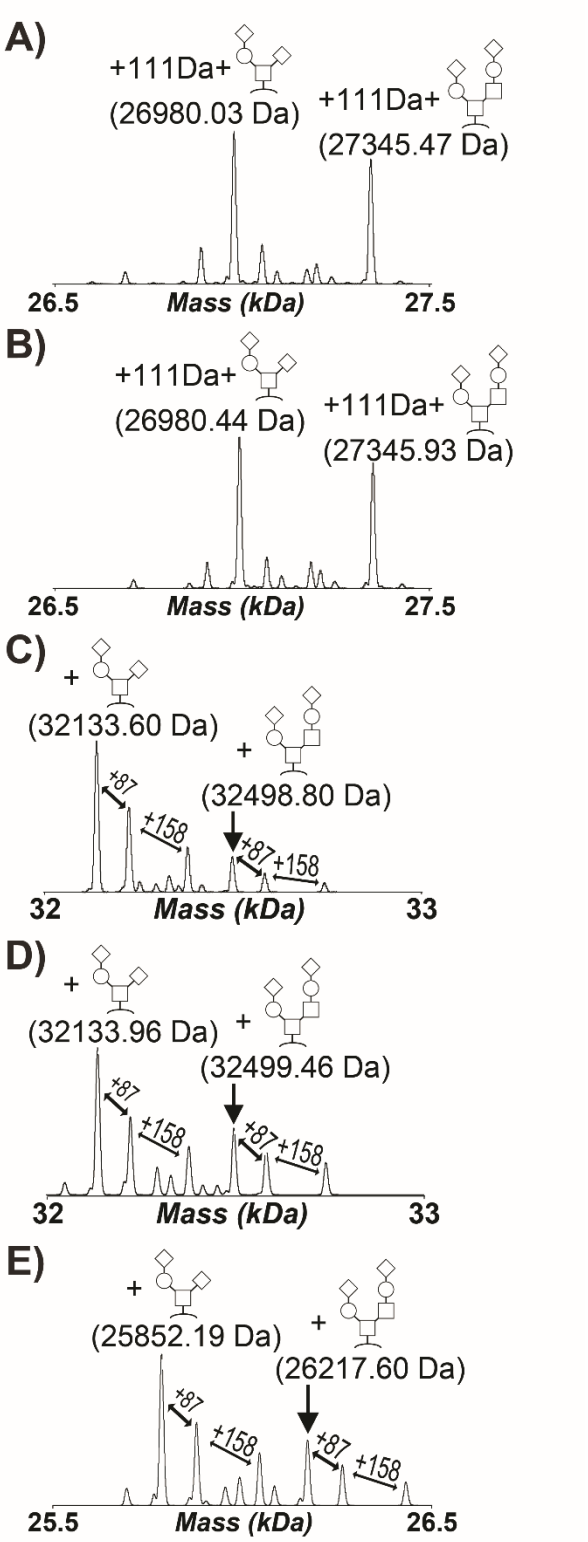

Figure S6. ReSpect™ deconvolution of initial RBD constructs. A) Original Mt. Sinai Construct. B) M67, the FNL vector optimized version of the Mt. Sinai construct. C) Original Ragon Construct. D) M68, the FNL vector optimized version of the Ragon Construct. E) M69, the FNL vector optimized version of the Ragon Construct with only a His8 C-terminal tag.

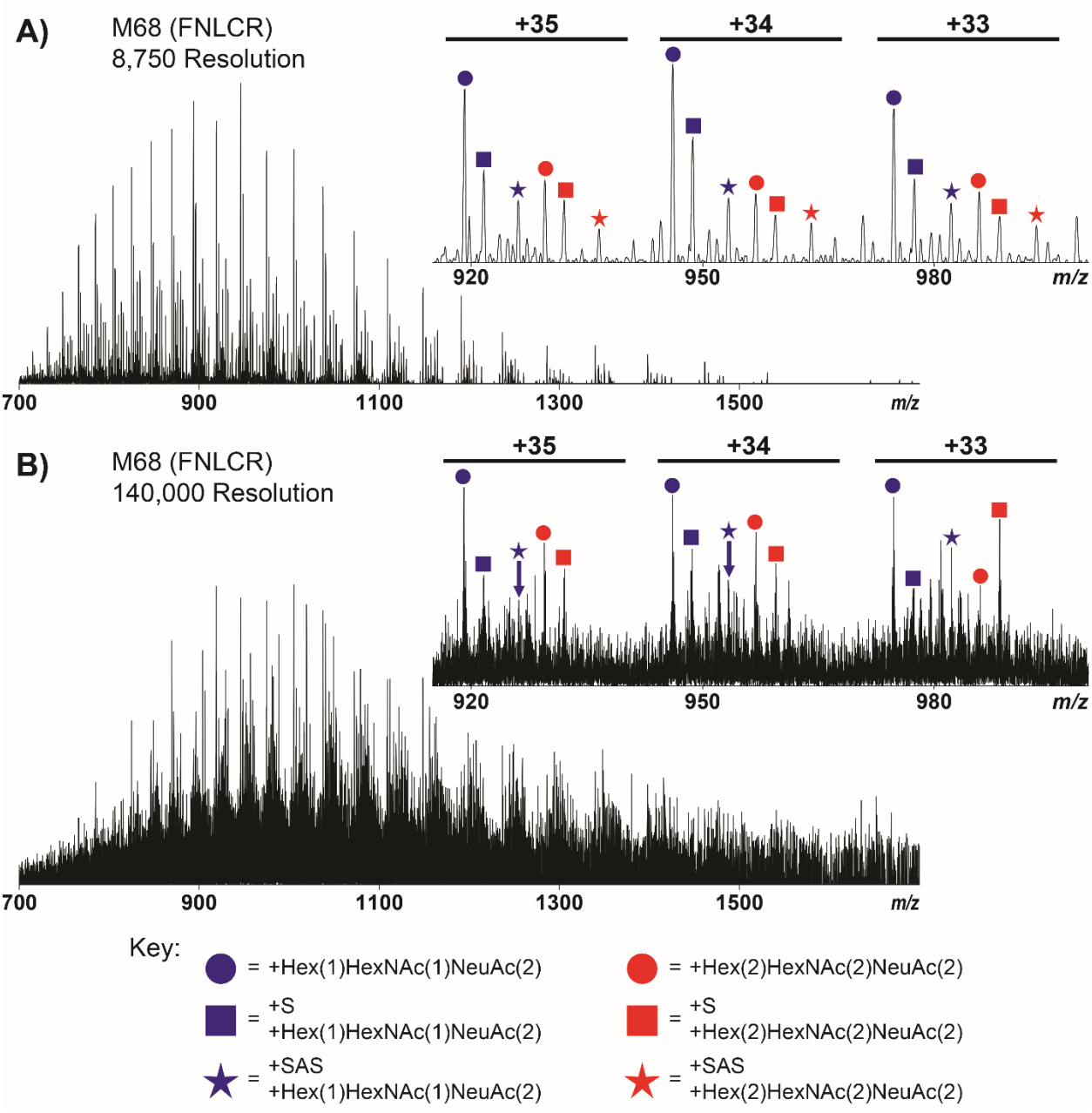

Figure S7: Intact Mass Resolution Comparison of M68. A) Intact mass spectrum of M68 following PNGaseF treatment with an FT resolution of 8,750. B) Intact mass spectrum of M68 following PNGaseF treatment with an FT resolution of 140,000. The insets contain an expansion of 915-1000 *m/z*, corresponding to charge states +33, +34, and +35, to highlight spectral complexity at the given FT resolution.

MDAMKRGLCC VLLLCGAVFV SPSASFRVQP TESIVRFPNI TNLCPFGEVF NATRFASVYA WNRKRISNCV
ADYSVLYNSA SFSTFKCYGV SPTKLNDLCF TNVYADSFVI RGDEVRQIAP GQTGKIADYN YKLPDDFTGC
VIAWNSNNLD SKVGGNYNYL YRLFRKSNLK PFERDISTEI YQAGSTPCNG VEGFNCYFPL QSYGFQPTNG
VGYQPYRVVV LSFELLHAPA TVCGPKKGAG SSLEVLFQGP GSGSSHHHHH HHHGGSGSSM DEKTTGWRGG
HVVEGLAGEL EQLRARLEHH PQGQREP

Figure S8. The predicted M96 RBD protein sequence from the construct DNA. Red: Tissue plasminogen activator (TPA) signal sequence. Black: F318-K529; RBD of spike. Blue: HRV3C protease cleavage site, His8, streptavidin-binding peptide combination tag.

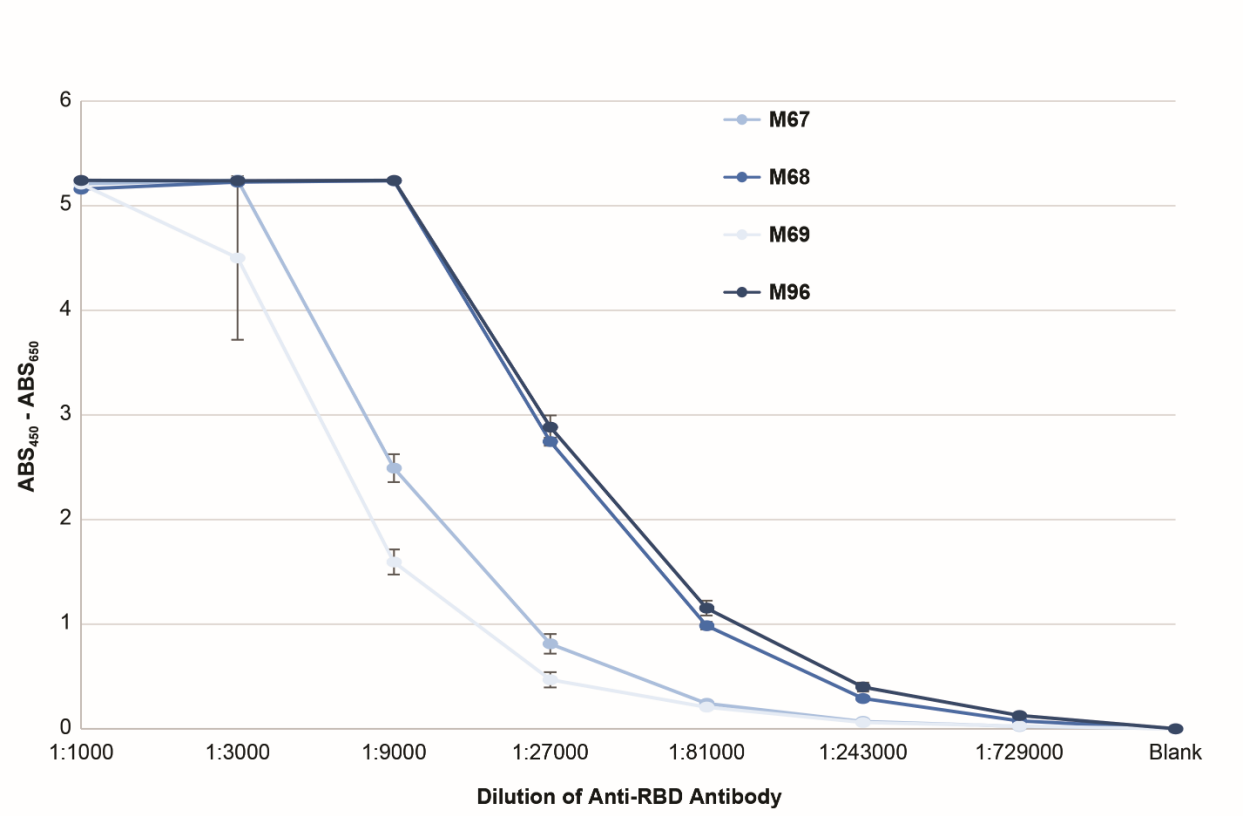

Figure S9. Comparative ELISA sensitivity of the four RBD constructs. ELISA was performed with the previous (M67, M68, and M69) and newest (M96) RBD constructs coating the ELSIA plates and treated with anti-RBD monoclonal antibodies at the reported dilutions (X-axis). Absorbance measurements were taken at 450 nm which were corrected by subtraction with the absorbance at 650 nm (Y-axis). Triplicate measurements were performed with error bars equal to the standard deviation.

Table S1. Intact mass analysis of the original Mt. Sinai construct.

Proteoform	Theoretical Average Mass (Da)	Observed Average Mass (Da)	Average Mass Error (Da)	Intensity	Relative Abundance	Fractional Abundance
+pyroGlu +Hex(1)HexNAc(1)NeuAc(2)	26980.19	26980.03	0.16	4.25E+08	100.00%	54.18%
+pyroGlu +Hex(2)HexNAc(2)NeuAc(2)	27345.52	27345.47	0.05	3.59E+08	84.59%	45.82%

Table S2. Intact mass analysis of M67.

Proteoform	Theoretical Monoisotopic Mass (Da)	Observed Monoisotopic Mass (Da)	Monoisotopic Mass Error (Da)	Theoretical Average Mass (Da)	Observed Average Mass (Da)	Average Mass Error (Da)	Intensity	Relative Abundance	Fractional Abundance
+pyroGlu +Hex(1)HexNAc(1)NeuAc(2)	26963.087	26964.117	1.030	26980.19	26980.44	0.25	5.63E+08	100.00%	55.86%
+pyroGlu +Hex(2)HexNAc(2)NeuAc(2)	27328.219	27329.218	0.999	27345.52	27345.93	0.41	4.45E+08	79.02%	44.14%

Table S3. Intact mass analysis of the original Ragon construct.

Proteoform	Theoretical Average Mass (Da)	Observed Average Mass (Da)	Average Mass Error (Da)	Intensity	Relative Abundance	Fractional Abundance
+Hex(1)HexNAc(1)NeuAc(2)	32133.64	32133.60	0.04	1.87E+08	100.00%	42.10%
+S +Hex(1)HexNAc(1)NeuAc(2)	32220.72	32220.73	0.01	1.10E+08	58.90%	24.80%
+SAS +Hex(1)HexNAc(1)NeuAc(2)	32378.88	32378.78	0.10	6.13E+07	32.81%	13.81%
+Hex(2)HexNAc(2)NeuAc(2)	32498.98	32498.80	0.18	4.28E+07	22.92%	9.65%
+S +Hex(2)HexNAc(2)NeuAc(2)	32586.06	32585.26	0.80	2.64E+07	14.13%	5.95%
+SAS +Hex(2)HexNAc(2)NeuAc(2)	32744.22	32744.87	0.65	1.64E+07	8.77%	3.69%

Table S4. Intact mass analysis of M68.

Proteoform	Theoretical Monoisotopic Mass (Da)	Observed Monoisotopic Mass (Da)	Monoisotopic Mass Error (Da)	Theoretical Average Mass (Da)	Observed Average Mass (Da)	Average Mass Error (Da)	Intensity	Relative Abundance	Fractional Abundance
+Hex(1)HexNAc(1)NeuAc(2)	32113.531	32114.634	1.103	32133.64	32133.96	0.32	2.37E+08	100.00%	34.13%
+S +Hex(1)HexNAc(1)NeuAc(2)	32200.563	32200.635	0.072	32220.72	32221.42	0.70	1.26E+08	53.32%	18.19%
+SAS +Hex(1)HexNAc(1)NeuAc(2)	32358.638	N/D‡	N/D	32378.88	32378.76	0.12	8.49E+07	35.86%	12.24%
+Hex(2)HexNAc(2)NeuAc(2)	32478.663	32478.711	0.048	32498.98	32499.46	0.48	1.09E+08	46.19%	15.76%
+S +Hex(2)HexNAc(2)NeuAc(2)	32565.695	N/D	N/D	32586.06	32586.02	0.04	8.20E+07	34.62%	11.81%
+SAS +Hex(2)HexNAc(2)NeuAc(2)	32723.764	N/D	N/D	32744.22	32745.24	1.02	5.46E+07	23.05%	7.87%

‡N/D = not detected in deconvolution

Table S5. Intact mass analysis of M69.

Proteoform	Theoretical Monoisotopic Mass (Da)	Observed Monoisotopic Mass (Da)	Monoisotopic Mass Error (Da)	Theoretical Average Mass (Da)	Observed Average Mass (Da)	Average Mass Error (Da)	Intensity	Relative Abundance	Fractional Abundance
+Hex(1)HexNAc(1)NeuAc(2)	25835.497	25836.543	1.046	25851.84	25852.19	0.35	2.47E+08	100.00%	35.60%
+S +Hex(1)HexNAc(1)NeuAc(2)	25922.529	25922.606	0.077	25938.92	25939.49	0.57	1.39E+08	56.34%	20.05%
+SAS +Hex(1)HexNAc(1)NeuAc(2)	26080.599	26080.615	0.016	26097.08	26097.34	0.26	8.58E+07	34.73%	12.36%
+Hex(2)HexNAc(2)NeuAc(2)	26200.630	26200.652	0.022	26217.18	26217.60	0.42	1.08E+08	43.77%	15.58%
+S +Hex(2)HexNAc(2)NeuAc(2)	26287.662	26287.675	0.013	26304.26	26304.88	0.62	6.74E+07	27.30%	9.72%
+SAS +Hex(2)HexNAc(2)NeuAc(2)	26245.731	N/D‡	N/D	26462.42	26463.21	0.79	4.64E+07	18.79%	6.69%

‡N/D = not detected in deconvolution

Table S6. Intact mass analysis of M98.

Proteoform	Theoretical Monoisotopic Mass (Da)	Observed Monoisotopic Mass (Da)	Monoisotopic Mass Error (Da)	Theoretical Average Mass (Da)	Observed Average Mass (Da)	Average Mass Error (Da)	Intensity	Relative Abundance	Fractional Abundance
Unmodified	31313.276	31313.312	0.036	31332.98	31332.53	0.45	2.61E+07	31.17%	15.16%
+Hex(1)HexNAc(1)NeuAc(2)	32260.599	32260.676	0.077	32280.82	32280.64	0.18	8.38E+07	100.00%	48.62%
+Hex(2)HexNAc(2)NeuAc(2)	32625.732	32626.786	1.054	32646.16	32646.06	0.10	6.24E+07	74.50%	36.22%