SUPPORTING INFORMATION

A tandem mass spectrometry sequence database search method for identification of O-fucosylated proteins by mass spectrometry.

Kristian E. Swearingen[1], Jimmy K. Eng[2], David Shteynberg[1], Vladimir Vigdorovich[3], Timothy A. Springer[4], Luis Mendoza[1], Noah D. Sather[3], Eric W. Deutsch[1], Stefan H. I. Kappe[3,] , Robert L. Moritz[1*]*

[1] Institute for Systems Biology, Seattle, WA, USA

[2] Proteomics Resource, University of Washington, Seattle, WA, USA

[3] Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, WA, USA

[4]Harvard Medical School and Children's Hospital, Boston, MA, USA

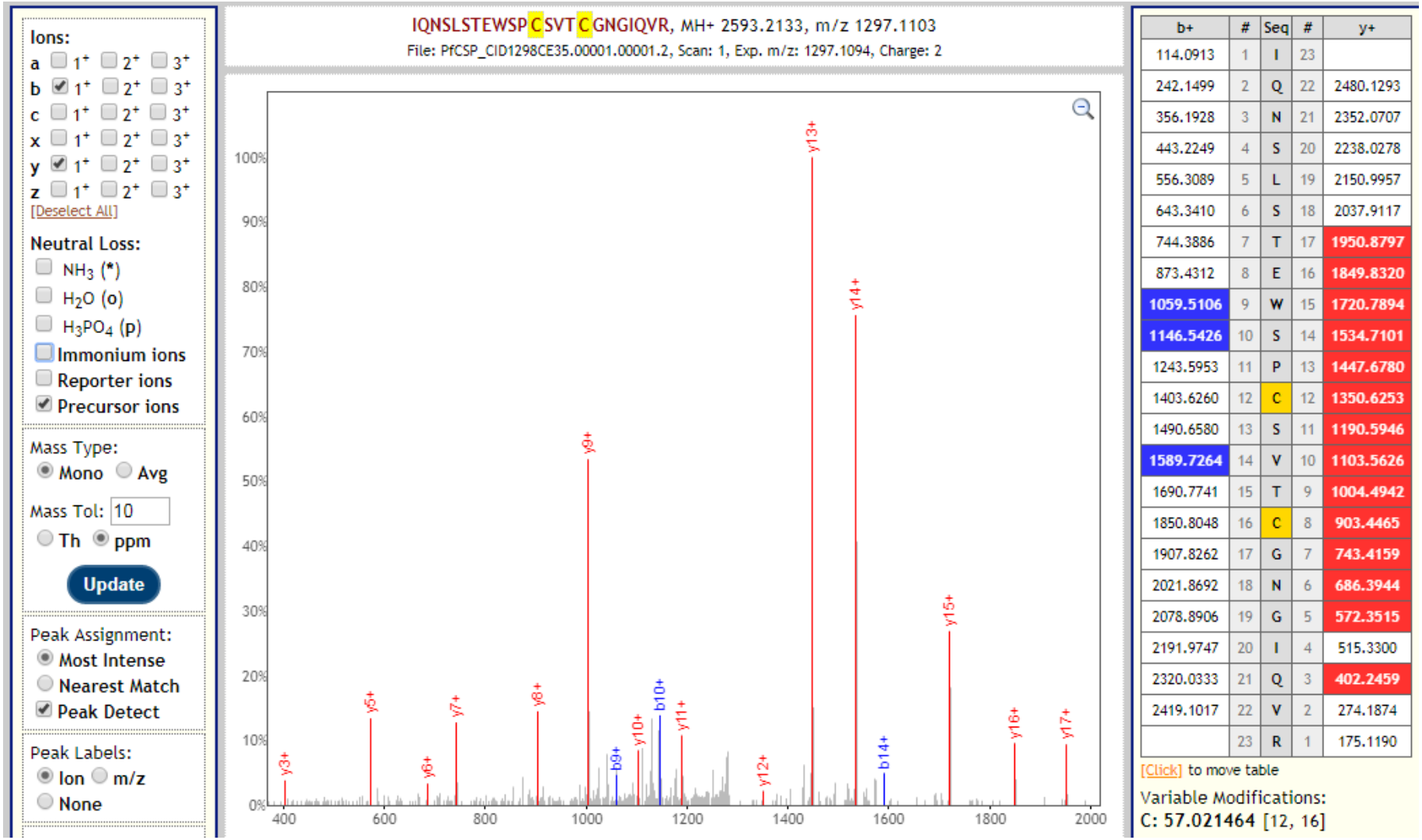
*To whom correspondence should be addressed.

kristian.swearingen@systemsbiology.org

robert.moritz@systemsbiology.org

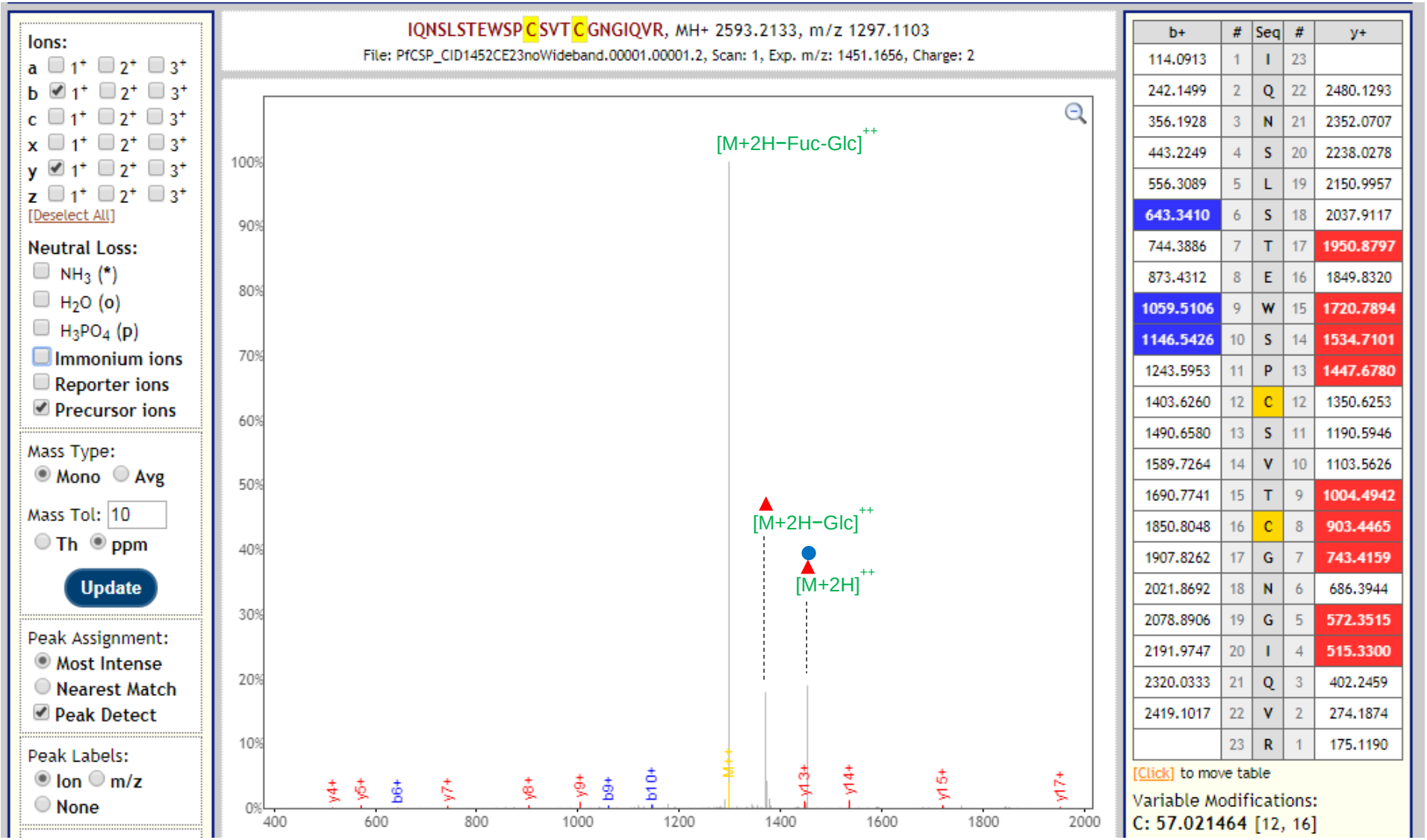
TABLE OF CONTENTS

Figure S1. MS^n confirming the identity of fragment spectra arising from neutral loss of O-linked glycans. A tryptic digest of recombinant *P. falciparum* circumsporozoite protein (CSP) TSR domain was infused into an LTQ-Orbitrap Elite by nanospray. Precursor ions for the unmodified or O-glucosylfucosylated peptide were isolated and fragmented. All spectra are collision-induced dissociation (CID) collected at high resolution in the Orbitrap. CE = normalized collision energy. Annotated fragment spectra were visualized using a development version of Lorikeet as implemented in the Trans-Proteomic Pipeline.



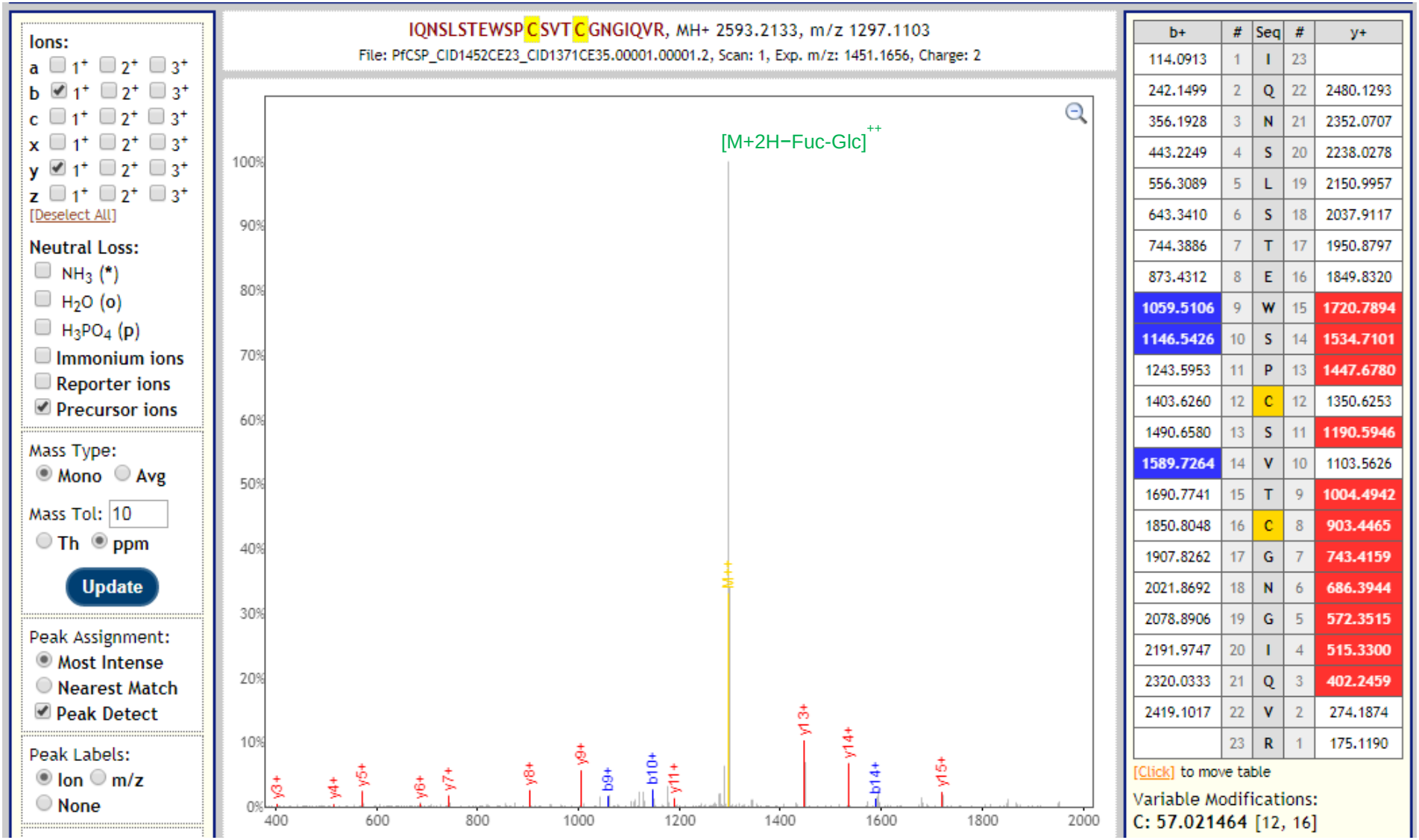
b+	#	Seq	#	y+
114.0913	1	I	23	
242.1499	2	Q	22	2480.1293
356.1928	3	N	21	2352.0707
443.2249	4	S	20	2238.0278
556.3089	5	L	19	2150.9957
643.3410	6	S	18	2037.9117
744.3886	7	T	17	1950.8797
873.4312	8	E	16	1849.8320
1059.5106	9	W	15	1720.7894
1146.5426	10	S	14	1534.7101
1243.5953	11	P	13	1447.6780
1403.6260	12	C	12	1350.6253
1490.6580	13	S	11	1190.5946
1589.7264	14	V	10	1103.5626
1690.7741	15	T	9	1004.4942
1850.8048	16	C	8	903.4465
1907.8262	17	G	7	743.4159
2021.8692	18	N	6	686.3944
2078.8906	19	G	5	572.3515
2191.9747	20	I	4	515.3300
2320.0333	21	Q	3	402.2459
2419.1017	22	V	2	274.1874
	23	R	1	175.1190

S1a. CID of the unmodified peptide, $[M+2H]^{++}$ = 1297 *m/z*, CE = 35 %. Note that the dominant peaks are y_{13} and y_{14}, the y-ion peaks containing the Pro residue. The y_{13} and y_{14} peaks are the only species visible in the spectra in Figure 1 still bearing the O-linked disaccharide.



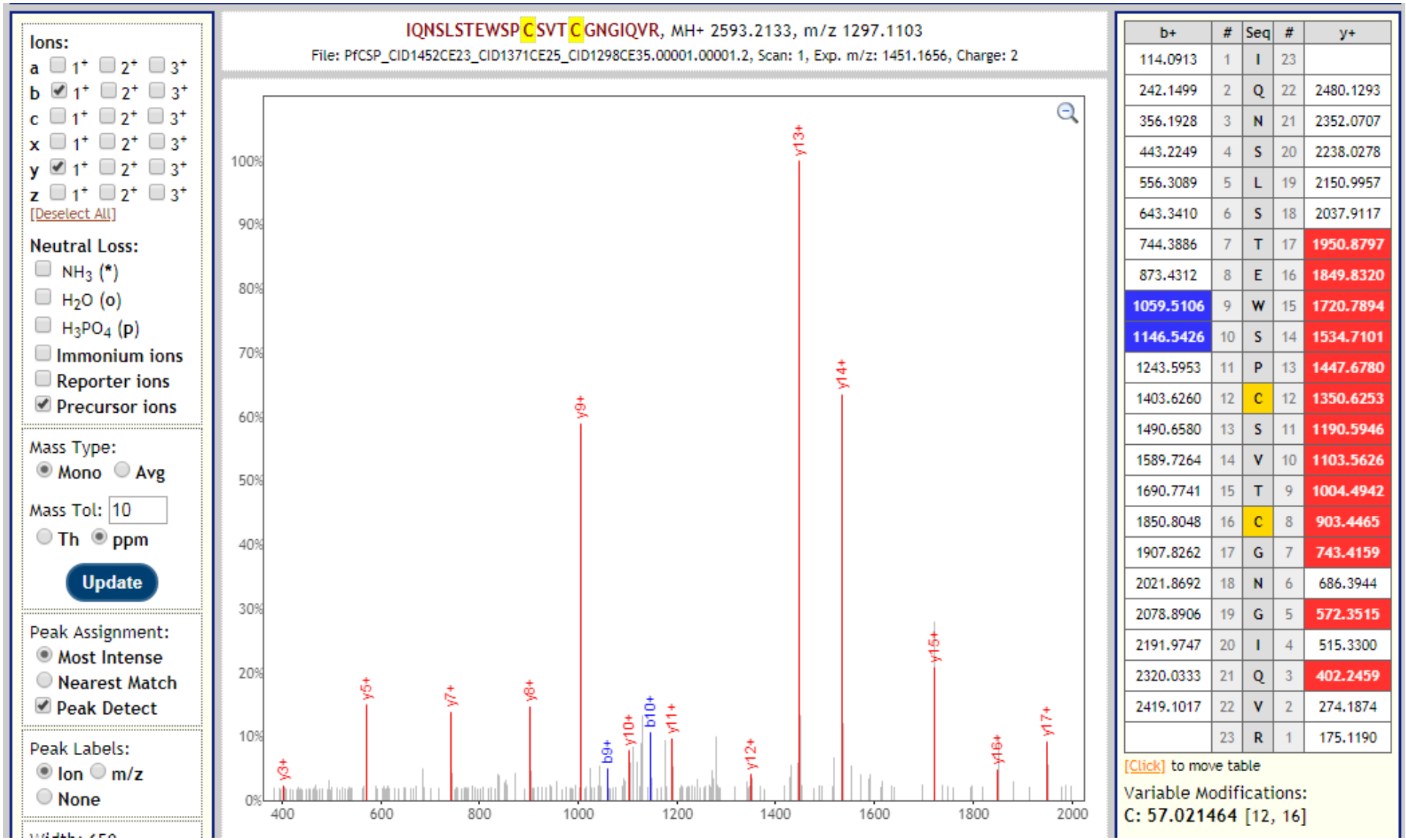
b+	#	Seq	#	y+
114.0913	1	I	23	
242.1499	2	Q	22	2480.1293
356.1928	3	N	21	2352.0707
443.2249	4	S	20	2238.0278
556.3089	5	L	19	2150.9957
643.3410	6	S	18	2037.9117
744.3886	7	T	17	1950.8797
873.4312	8	E	16	1849.8320
1059.5106	9	W	15	1720.7894
1146.5426	10	S	14	1534.7101
1243.5953	11	P	13	1447.6780
1403.6260	12	C	12	1350.6253
1490.6580	13	S	11	1190.5946
1589.7264	14	V	10	1103.5626
1690.7741	15	T	9	1004.4942
1850.8048	16	C	8	903.4465
1907.8262	17	G	7	743.4159
2021.8692	18	N	6	686.3944
2078.8906	19	G	5	572.3515
2191.9747	20	I	4	515.3300
2320.0333	21	Q	3	402.2459
2419.1017	22	V	2	274.1874
	23	R	1	175.1190

Figure S1b. CID of the peptide modified with O-Fuc-Glc, $[M+2H]^{++}$ = 1451.17 *m/z*, CE = 23 %. At this CE, some of the precursor is still intact ($[M+2H]^{++}$, red triangle = O-Fuc, blue circle = β1,3-Glc). Neutral loss of Glc can be seen as a peak at 1370.14 *m/z* ($[M+2H-Glc]^{++}$, red triangle = O-Fuc). The dominant peak at 1297.11 *m/z* is the precursor with neutral loss of the O-Fuc-Glc disaccharide ($[M+2H-Fuc-Glc]^{++}$). Peptide fragments are present at low abundance, identifying the sequence of the unmodified peptide.



b+	#	Seq	#	y+
114.0913	1	I	23	
242.1499	2	Q	22	2480.1293
356.1928	3	N	21	2352.0707
443.2249	4	S	20	2238.0278
556.3089	5	L	19	2150.9957
643.3410	6	S	18	2037.9117
744.3886	7	T	17	1950.8797
873.4312	8	E	16	1849.8320
1059.5106	9	W	15	**1720.7894**
1146.5426	10	S	14	**1534.7101**
1243.5953	11	P	13	**1447.6780**
1403.6260	12	C	12	1350.6253
1490.6580	13	S	11	**1190.5946**
1589.7264	14	V	10	1103.5626
1690.7741	15	T	9	**1004.4942**
1850.8048	16	C	8	**903.4465**
1907.8262	17	G	7	**743.4159**
2021.8692	18	N	6	**686.3944**
2078.8906	19	G	5	**572.3515**
2191.9747	20	I	4	**515.3300**
2320.0333	21	Q	3	**402.2459**
2419.1017	22	V	2	274.1874
	23	R	1	175.1190

Figure S1c. MS^3 (CE = 35 %) of the 1370.14 *m/z* fragment seen in Figure S1b produces a dominant peak at 1297.11 *m/z* matching the mass of the unmodified peptide. Many unmodified peptide fragment peaks confirm the sequence of the peptide. This spectrum confirms that the 1370.14 *m/z* peak seen in Figure S1b arises from neutral loss of Glc from the O-Fuc-Glc dissacharide.



b+	#	Seq	#	y+
114.0913	1	I	23	
242.1499	2	Q	22	2480.1293
356.1928	3	N	21	2352.0707
443.2249	4	S	20	2238.0278
556.3089	5	L	19	2150.9957
643.3410	6	S	18	2037.9117
744.3886	7	T	17	1950.8797
873.4312	8	E	16	1849.8320
1059.5106	9	W	15	1720.7894
1146.5426	10	S	14	1534.7101
1243.5953	11	P	13	1447.6780
1403.6260	12	C	12	1350.6253
1490.6580	13	S	11	1190.5946
1589.7264	14	V	10	1103.5626
1690.7741	15	T	9	1004.4942
1850.8048	16	C	8	903.4465
1907.8262	17	G	7	743.4159
2021.8692	18	N	6	686.3944
2078.8906	19	G	5	572.3515
2191.9747	20	I	4	515.3300
2320.0333	21	Q	3	402.2459
2419.1017	22	V	2	274.1874
	23	R	1	175.1190

Figure S1d. MS^4 of the 1298.11 *m/z* peak seen in Figure S1c (MS^2 CE = 23 %, MS^3 CE = 25 %, MS^4 CE = 35 %). This peak arises from neutral loss of O-Fuc after neutral loss of Glc from the O-Fuc-Glc disaccharide. Fragment ions confirm the sequence of peptide. This spectrum further confirms that the 1370.14 *m/z* peak seen in Figure S1b arises from neutral loss of Glc from the O-Fuc-Glc dissacharide.

Figure S2. Neutral loss of O-linked and C-linked glycans at increasing collision energies.
A tryptic digest of recombinant *P. falciparum* thrombospondin related anonymous protein (TRAP) was analyzed by LC-MS/MS with an LTQ-Orbitrap Elite. Precursor ions for the TSR peptide modified with C-Man and O-Fuc-Glc were isolated and fragmented at sequentially increasing normalized collision energies (CE). All spectra are collision-induced dissociation (CID) collected at high resolution in the Orbitrap. Annotated fragment spectra were visualized using a development version of Lorikeet as implemented in the Trans-Proteomic Pipeline.

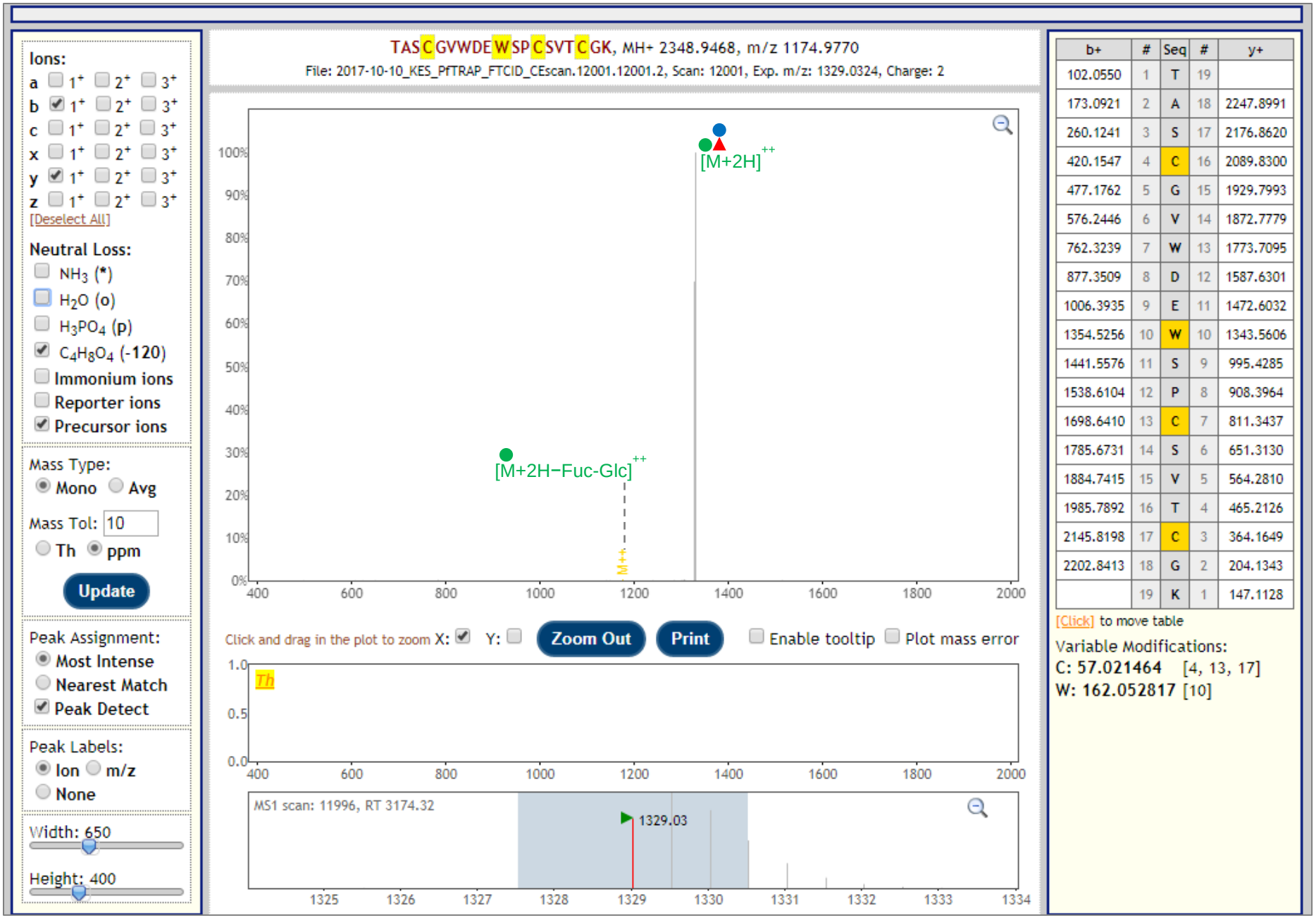


Figure S2a. CID at 1329.03 *m/z*, CE = 20 %. The precursor peak is largely unreacted ([M+2H]$^{++}$, green circle = C-Man, red triangle = O-Fuc, blue circle = β1,3-Glc), though a small peak (~0.3% relative abundance) matches the mass of the peptide after neutral loss of the O-linked disaccharide (1174.98 *m/z,* yellow, automatically annotated by Lorikeet as the precursor peak M++).

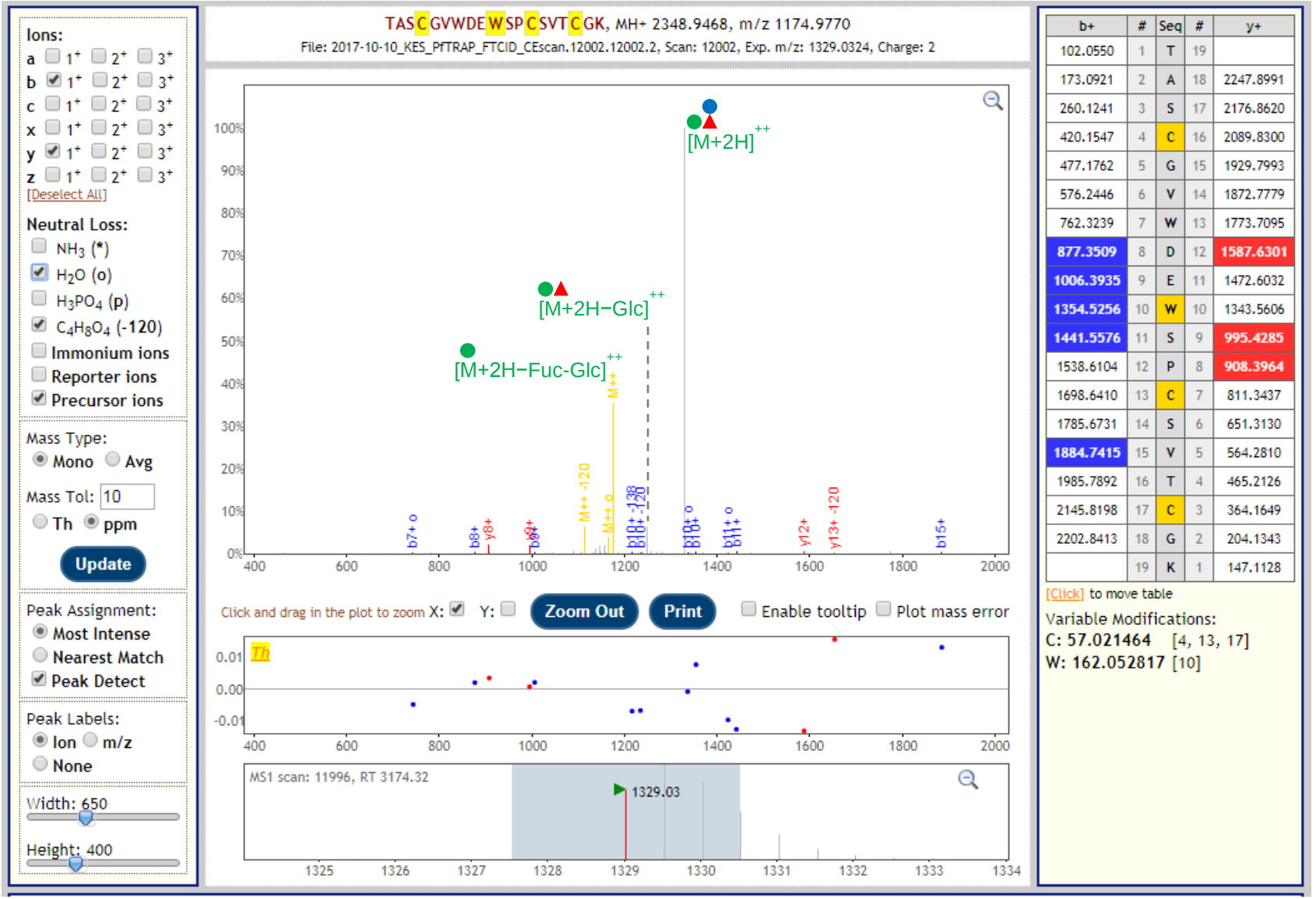


Figure S2b. CID at 1329.03 *m/z*, CE = 25 %. The dominant peak is still unreacted precursor ($[M+2H]^{++}$, green circle = C-Man, red triangle = O-Fuc, blue circle = β1,3-Glc). A smaller peak at 1248.01 *m/z* matches the mass of the precursor after neutral loss of Glc from the O-Fuc-Glc disaccharide ($[M+2H-Glc]^{++}$). A third peak at 1174.98 *m/z* (yellow, automatically annotated by Lorikeet as the precursor peak M++) matches the mass of the peptide precursor ion after neutral loss of the O-Fuc-Glc disaccharide ($[M+2H-Fuc-Glc]^{++}$). Note that peptide fragment ions appearing at this CE identify the sequence of the peptide with C-Man intact, positively localizing the glycan to the C-terminal Trp residue. Neutral loss of 120.04 Da from cross-ring cleavage is seen on the precursor and fragment peaks.

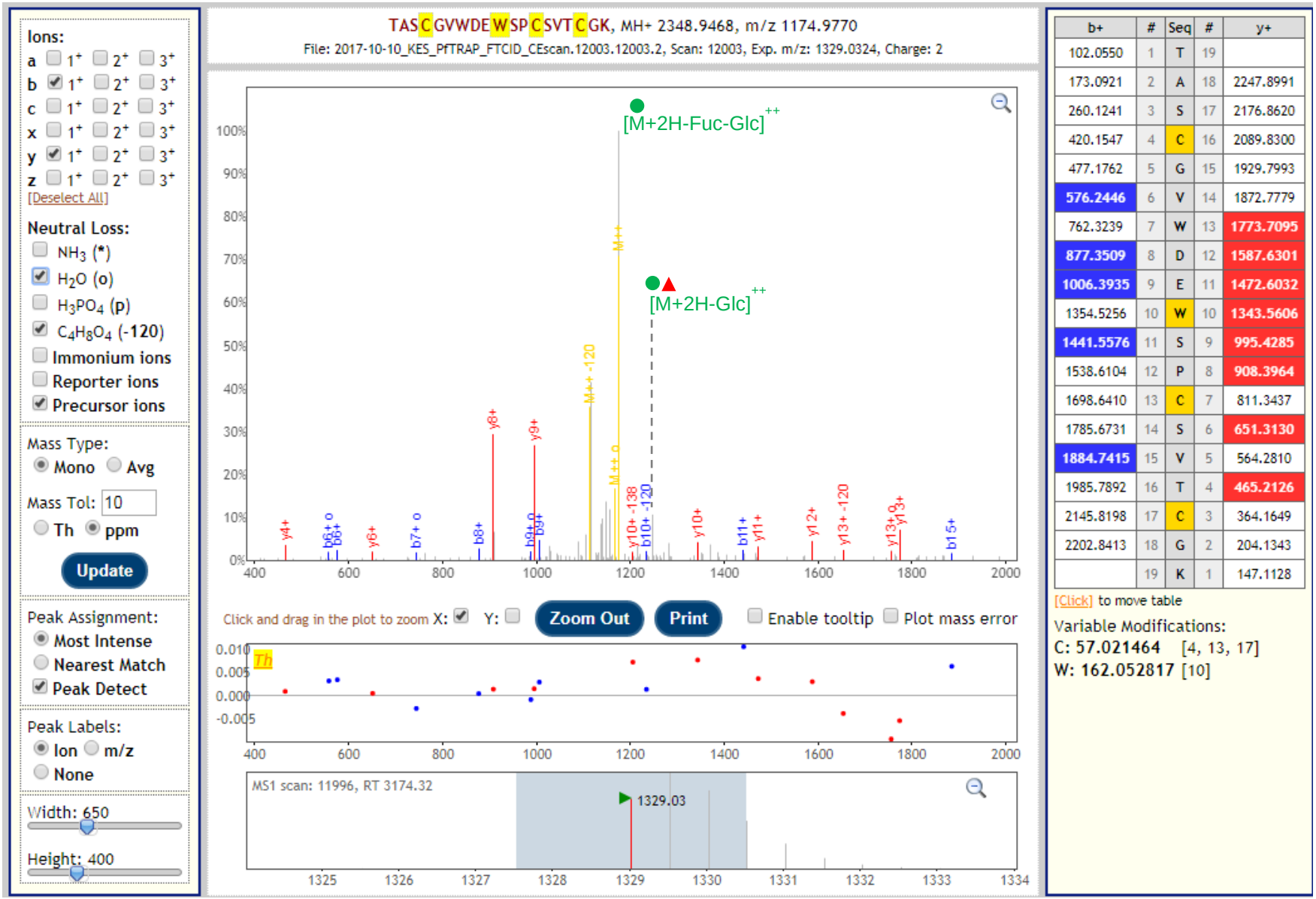
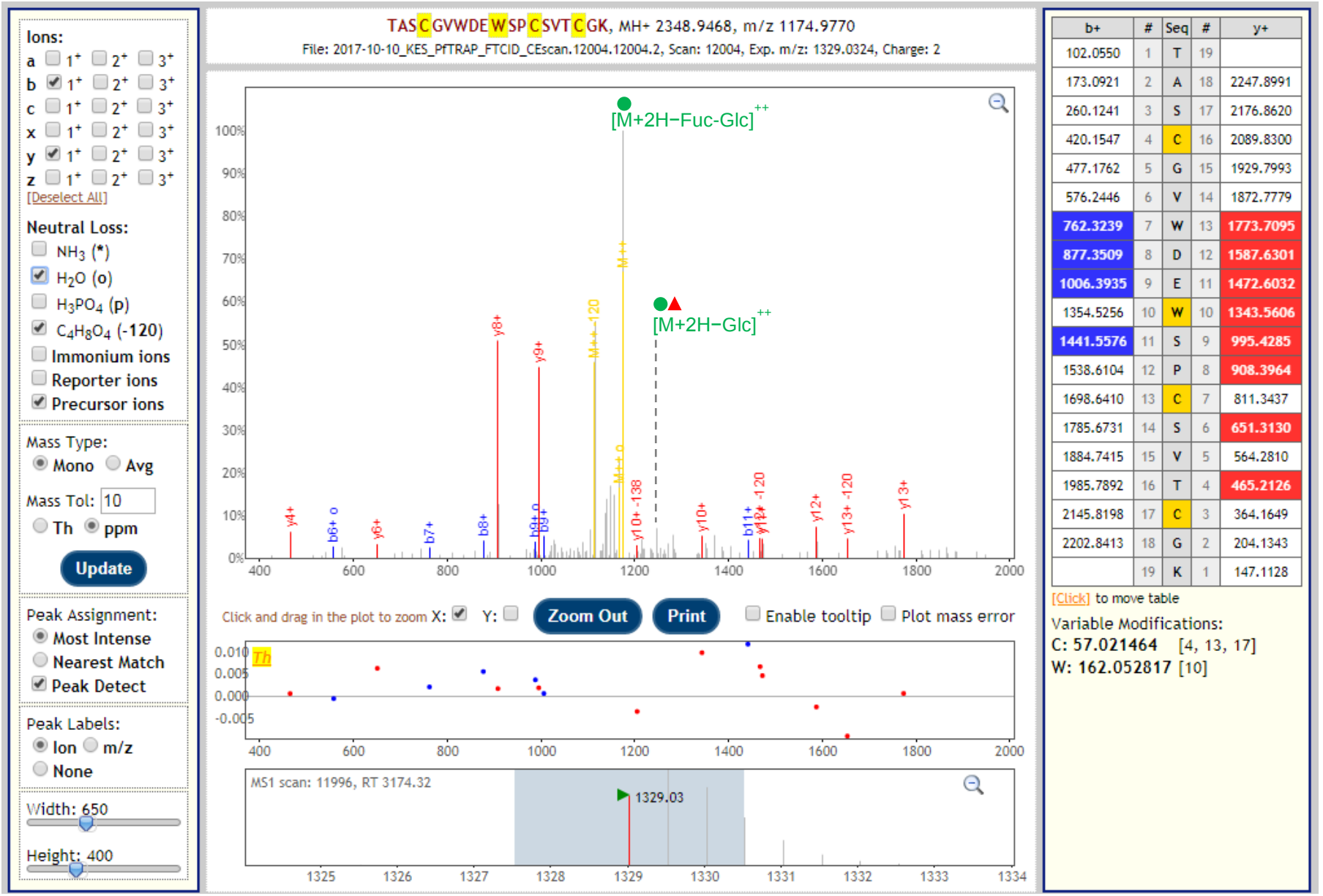


Figure S2c. CID at 1329.03 *m/z*, CE = 30 %. At this CE all parent precursor ion has been fragmented. The small peak at 1248.01 *m/z* remains, the mass of the precursor after neutral loss of Glc from the O-Fuc-Glc disaccharide ([M+2H−Glc]$^{++}$). The dominant peak is now 1174.98 *m/z* (yellow M++), the peptide precursor ion after neutral loss of the O-Fuc-Glc disaccharide ([M+2H−Fuc-Glc]$^{++}$). Further peptide fragment ions confidently identify the sequence of the peptide with C-Man intact.



b+	#	Seq	#	y+
102.0550	1	T	19	
173.0921	2	A	18	2247.8991
260.1241	3	S	17	2176.8620
420.1547	4	C	16	2089.8300
477.1762	5	G	15	1929.7993
576.2446	6	V	14	1872.7779
762.3239	7	W	13	1773.7095
877.3509	8	D	12	1587.6301
1006.3935	9	E	11	1472.6032
1354.5256	10	W	10	1343.5606
1441.5576	11	S	9	995.4285
1538.6104	12	P	8	908.3964
1698.6410	13	C	7	811.3437
1785.6731	14	S	6	651.3130
1884.7415	15	V	5	564.2810
1985.7892	16	T	4	465.2126
2145.8198	17	C	3	364.1649
2202.8413	18	G	2	204.1343
	19	K	1	147.1128

[Click] to move table

Variable Modifications:
C: 57.021464 [4, 13, 17]
W: 162.052817 [10]

Figure S2d. CID at 1329.03 *m/z*, CE = 35 %. This is the CE typically used in shotgun proteomics methods. The C-mannosylated peptide precursor lacking the O-linked glycan remains the dominant peak, but peptide fragment ions are abundant enough to confidently sequence the peptide. Unlike the recombinant *Pf*CSP spectra in Figure 1 and supplemental Figure S1b, no peptide fragment ions with the O-linked glycan intact are visible at any CE. This is likely a signal-to-noise issue; the spectra in Figure 1 and Figure S1 were acquired by direct infusion of the peptide, allowing signal and ion fill time to be maximized, whereas these spectra were acquired on-the-fly by LC-MS/MS.

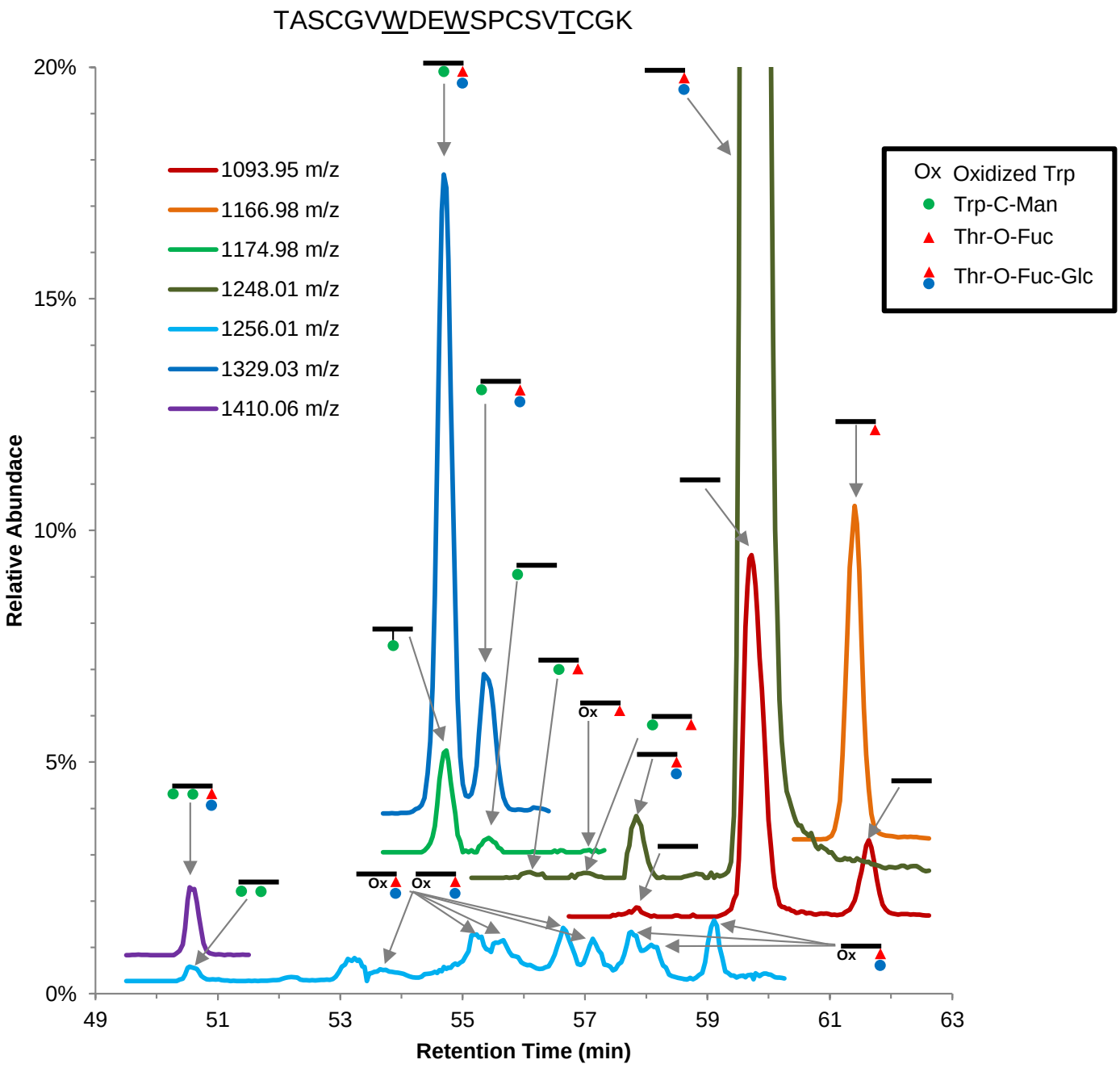
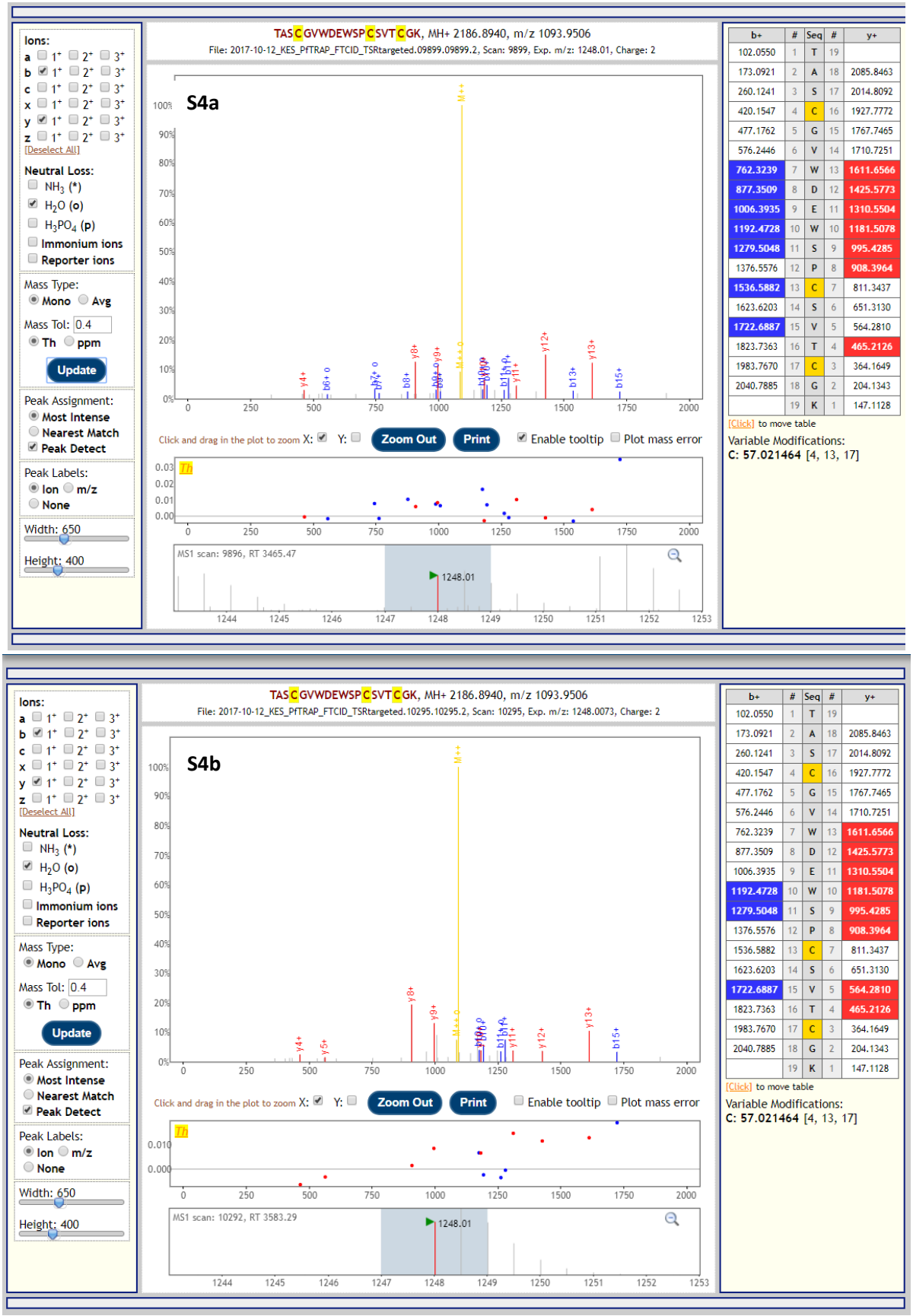


Figure S3. Glycoforms identified by automated sequence database search of data from LC-MS/MS analysis of recombinant *P. falciparum* TRAP. The variably C-mannosylated or oxidized Trp residues and the variably O-fucosylated Thr are underlined. Traces are offset for clarity. Peaks indicated with arrows were positively identified by PSMs. Both positional isoforms of the oxidized peptide were identified at several co-eluting retention times. See table S2.

Figure S4. Indistinguishable isobaric glycopeptides from recombinant *P. falciparum* TRAP.



b+	#	Seq	#	y+
102.0550	1	T	19	
173.0921	2	A	18	2085.8463
260.1241	3	S	17	2014.8092
420.1547	4	C	16	1927.7772
477.1762	5	G	15	1767.7465
576.2446	6	V	14	1710.7251
762.3239	7	W	13	1611.6566
877.3509	8	D	12	1425.5773
1006.3935	9	E	11	1310.5504
1192.4728	10	W	10	1181.5078
1279.5048	11	S	9	995.4285
1376.5576	12	P	8	908.3964
1536.5882	13	C	7	811.3437
1623.6203	14	S	6	651.3130
1722.6887	15	V	5	564.2810
1823.7363	16	T	4	465.2126
1983.7670	17	C	3	364.1649
2040.7885	18	G	2	204.1343
	19	K	1	147.1128

[Click] to move table

b+	#	Seq	#	y+
102.0550	1	T	19	
173.0921	2	A	18	2085.8463
260.1241	3	S	17	2014.8092
420.1547	4	C	16	1927.7772
477.1762	5	G	15	1767.7465
576.2446	6	V	14	1710.7251
762.3239	7	W	13	1611.6566
877.3509	8	D	12	1425.5773
1006.3935	9	E	11	1310.5504
1192.4728	10	W	10	1181.5078
1279.5048	11	S	9	995.4285
1376.5576	12	P	8	908.3964
1536.5882	13	C	7	811.3437
1623.6203	14	S	6	651.3130
1722.6887	15	V	5	564.2810
1823.7363	16	T	4	465.2126
1983.7670	17	C	3	364.1649
2040.7885	18	G	2	204.1343
	19	K	1	147.1128

[Click] to move table

Figure S4a,b. Representative MS^2 of the two 1248.01 *m/z* peaks eluting at **a)** 57.8 min and **b)** 58.9 min. The MS^2 from both species appear to identify the same species, i.e. the peptide modified with a gas-phase-labile moiety having a mass matching that of O-Fuc-Glc. The dominant species of the MS^2 spectra, 1093.95 *m/z*, matches the mass of the unmodified peptide. Additional PSMs are listed in Table S2.

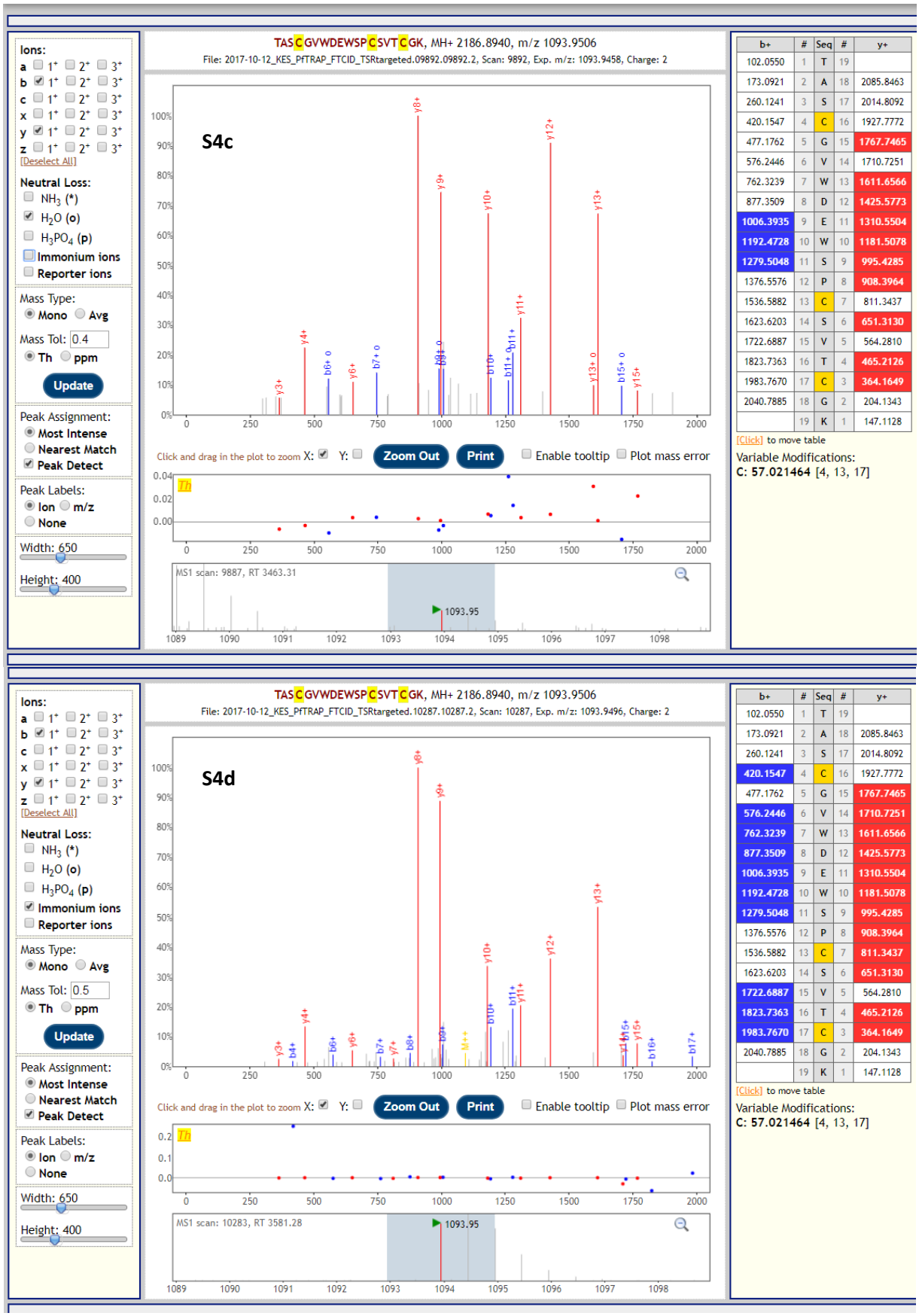


Figure S4c,d. Representative MS^2 of the two 1093.5 *m*/z peaks eluting at **c)** 57.8 min and **d)** 58.9 min. The MS^2 from both species appear to identify the unmodified peptide. Both species co-elute with the 1248.01 m/z species identified in Fig. S4a and b, consistent with loss of the gas-phase-labile O-Fuc-Glc from in-source fragmentation. Additional PSMs are listed in Table S2.

S5. Misidentification of a glycoform in LC-MS/MS of recombinant *P. falciparum* TRAP

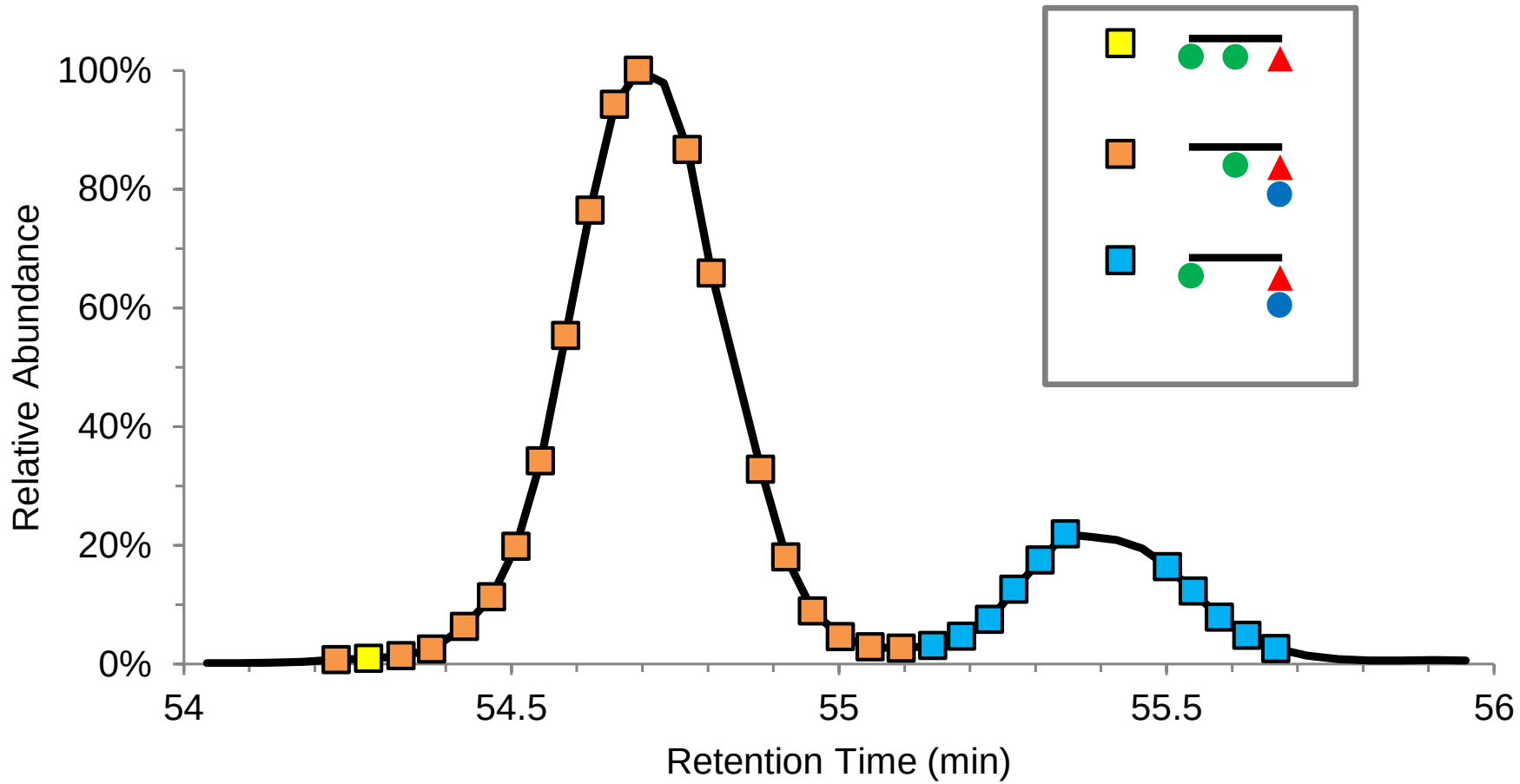
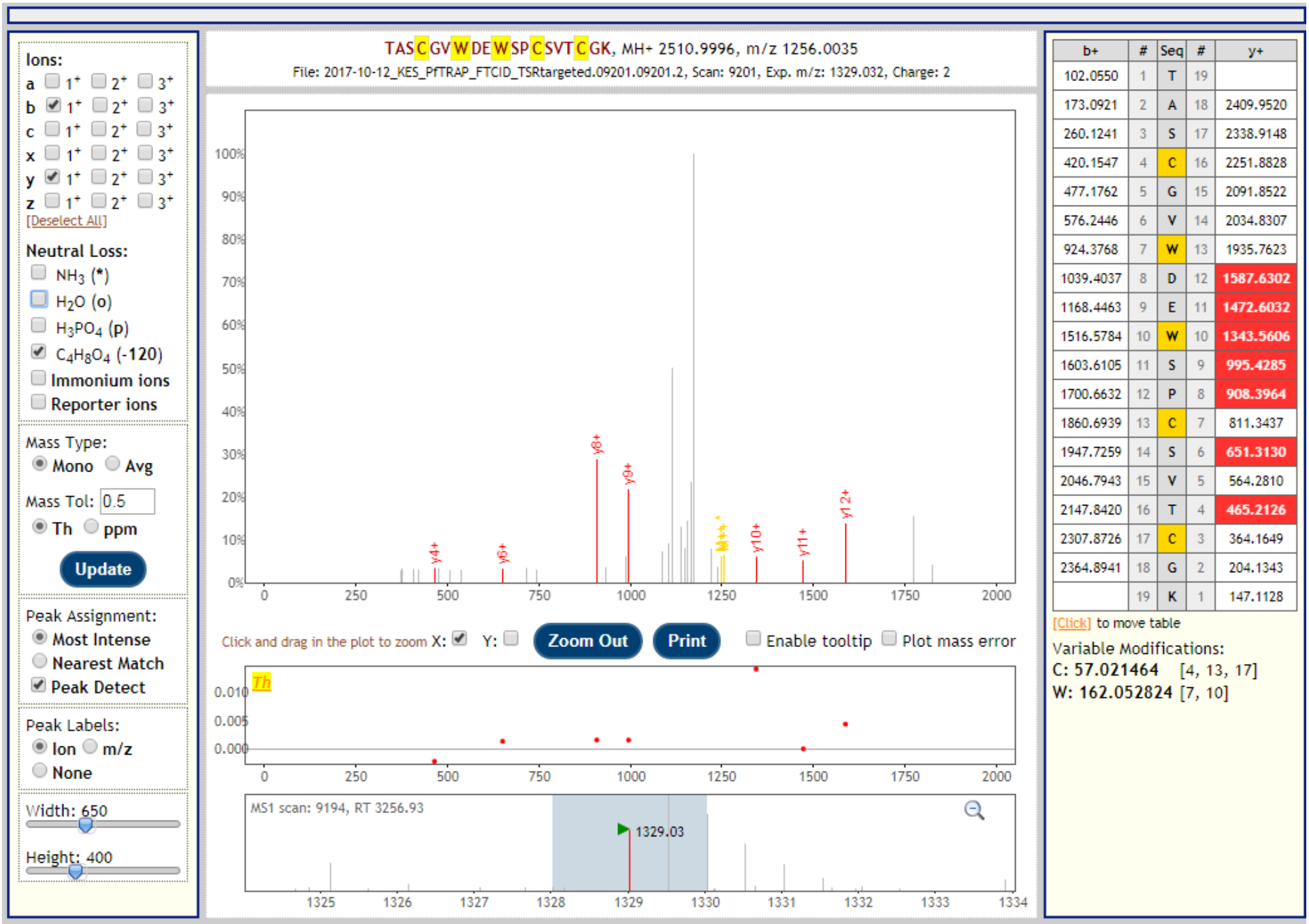


Figure S5a. Extracted Ion Chromatogram (XIC) of the precursor ion matching multiple isobaric glycoforms. The XIC of 1329.03 *m/z* (black line) is overlaid with squares showing points where MS^2 resulted in high-quality peptide spectrum matches (PSMs) identifying glycoforms of the recombinant *Pf*TRAP TSR peptide TASCGVWDEWSPCSVTCGK modified with combinations of C-linked mannose at tryptophan (green circle), O-linked fucose at the C-terminal threonine (red triangle) and β1,3-linked glucose (blue circle) added to the fucose. See Table S2.



b+	#	Seq	#	y+
102.0550	1	T	19	
173.0921	2	A	18	2409.9520
260.1241	3	S	17	2338.9148
420.1547	4	C	16	2251.8828
477.1762	5	G	15	2091.8522
576.2446	6	V	14	2034.8307
924.3768	7	W	13	1935.7623
1039.4037	8	D	12	**1587.6302**
1168.4463	9	E	11	**1472.6032**
1516.5784	10	W	10	**1343.5606**
1603.6105	11	S	9	**995.4285**
1700.6632	12	P	8	**908.3964**
1860.6939	13	C	7	811.3437
1947.7259	14	S	6	**651.3130**
2046.7943	15	V	5	564.2810
2147.8420	16	T	4	**465.2126**
2307.8726	17	C	3	364.1649
2364.8941	18	G	2	204.1343
	19	K	1	147.1128

[Click] to move table

Variable Modifications:
C: 57.021464 [4, 13, 17]
W: 162.052824 [7, 10]

Figure S5b A single PSM (the yellow square in Figure S5a) was incorrectly identified as the glycoform of the peptide featuring double mannosylation and an O-Fuc, with the MS^2 annotated as shown here.

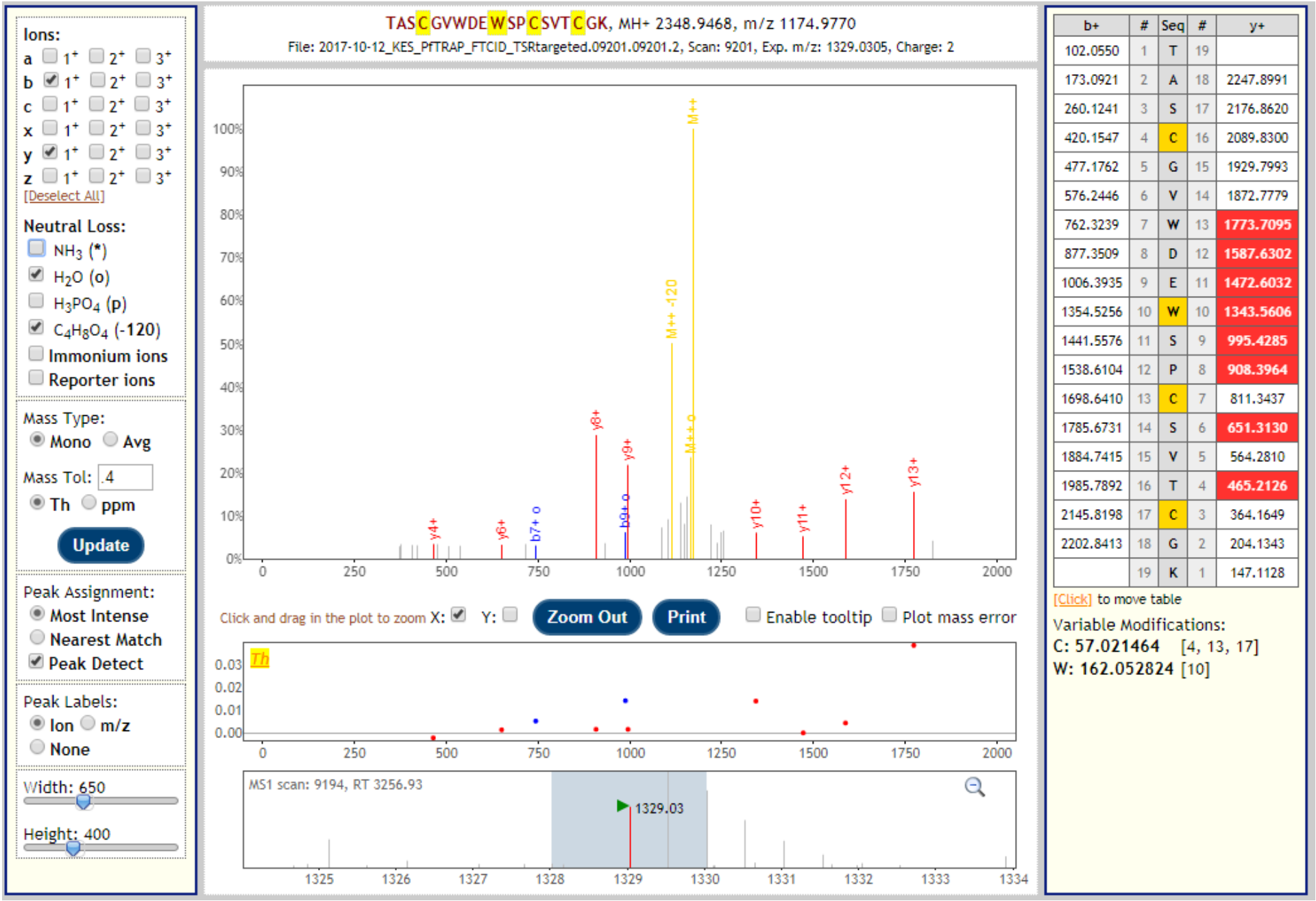


Figure S5c The same MS^2 from Figure S5b, re-annotated with the correct glycoform. While most of the fragment ions are the same as for the incorrect annotation (except y_{13}), the precursor fragment exhibiting neutral loss of O-Fuc-Glc (the yellow M++ peak) as well as cross-ring cleavage of C-man (yellow M++-120 peak) confirm that this is the correct annotation.

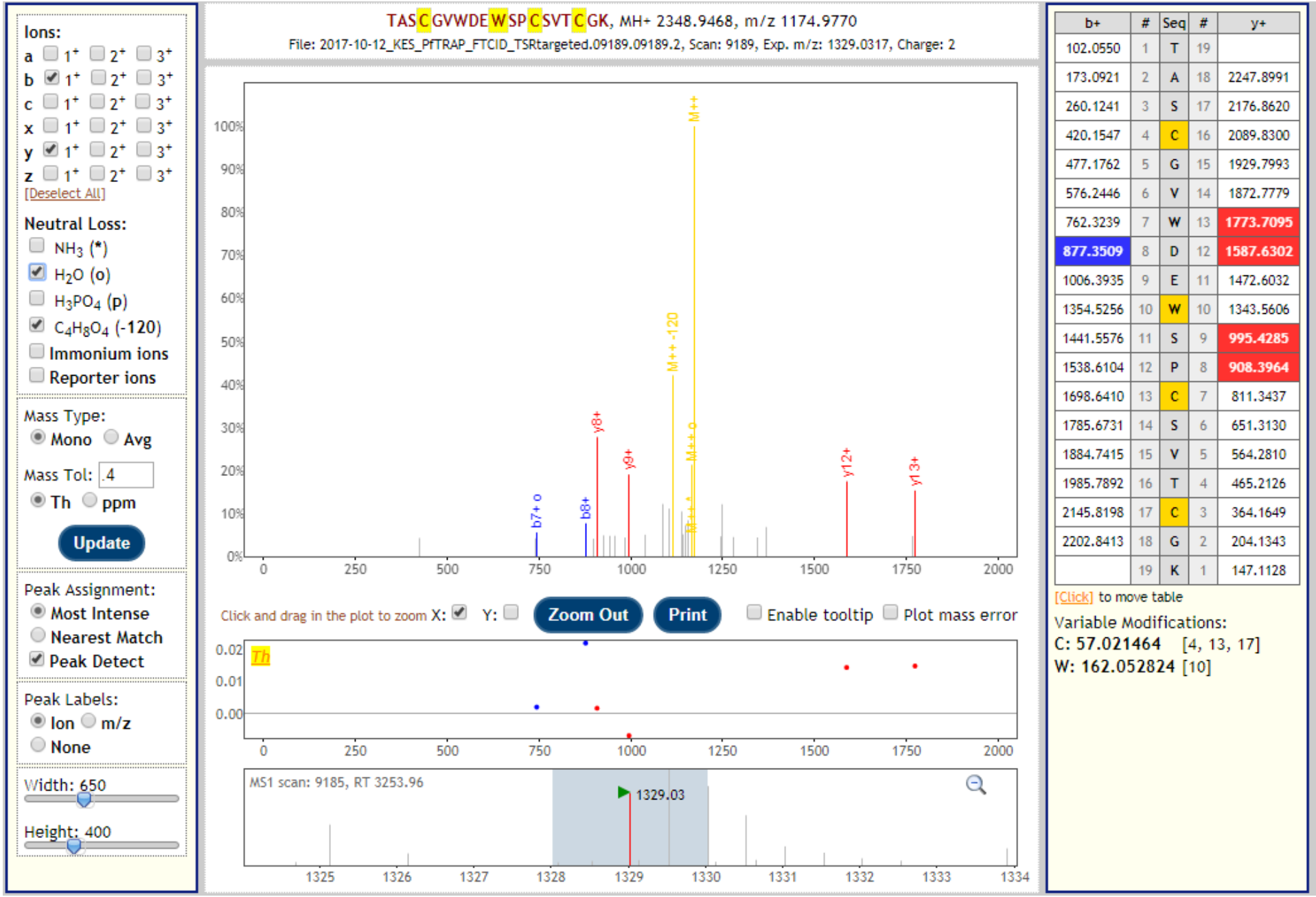


Figure S5d. The MS^2 for the same parent ion that immediately precedes the incorrect PSM, though sparse, identifies the correct glycoform.

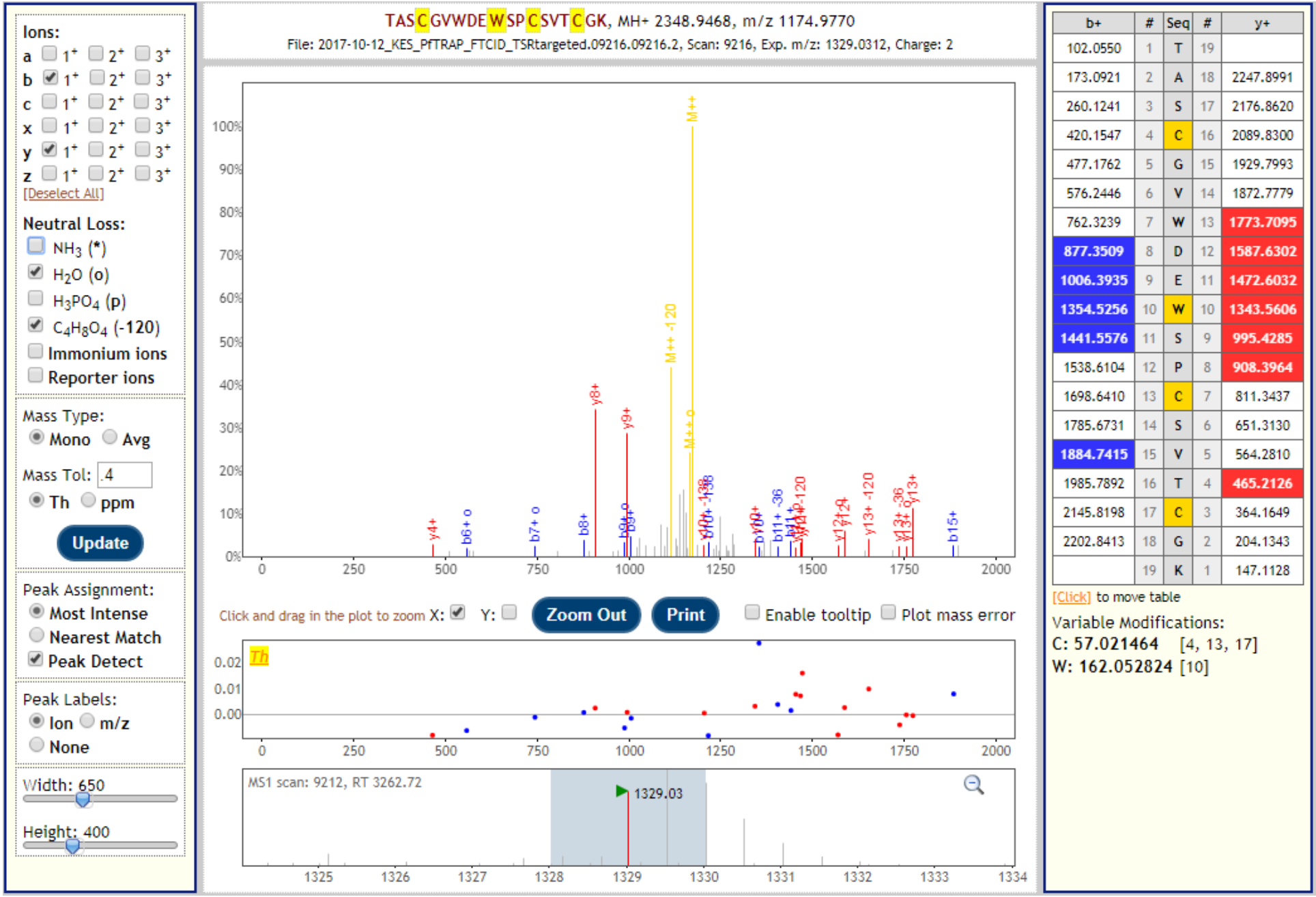


Figure S5e. The MS^2 for the same parent ion immediately following the incorrect PSM identifies the correct glycoform, as do the remaining PSMs for the parent ion in the same chromatographic peak.

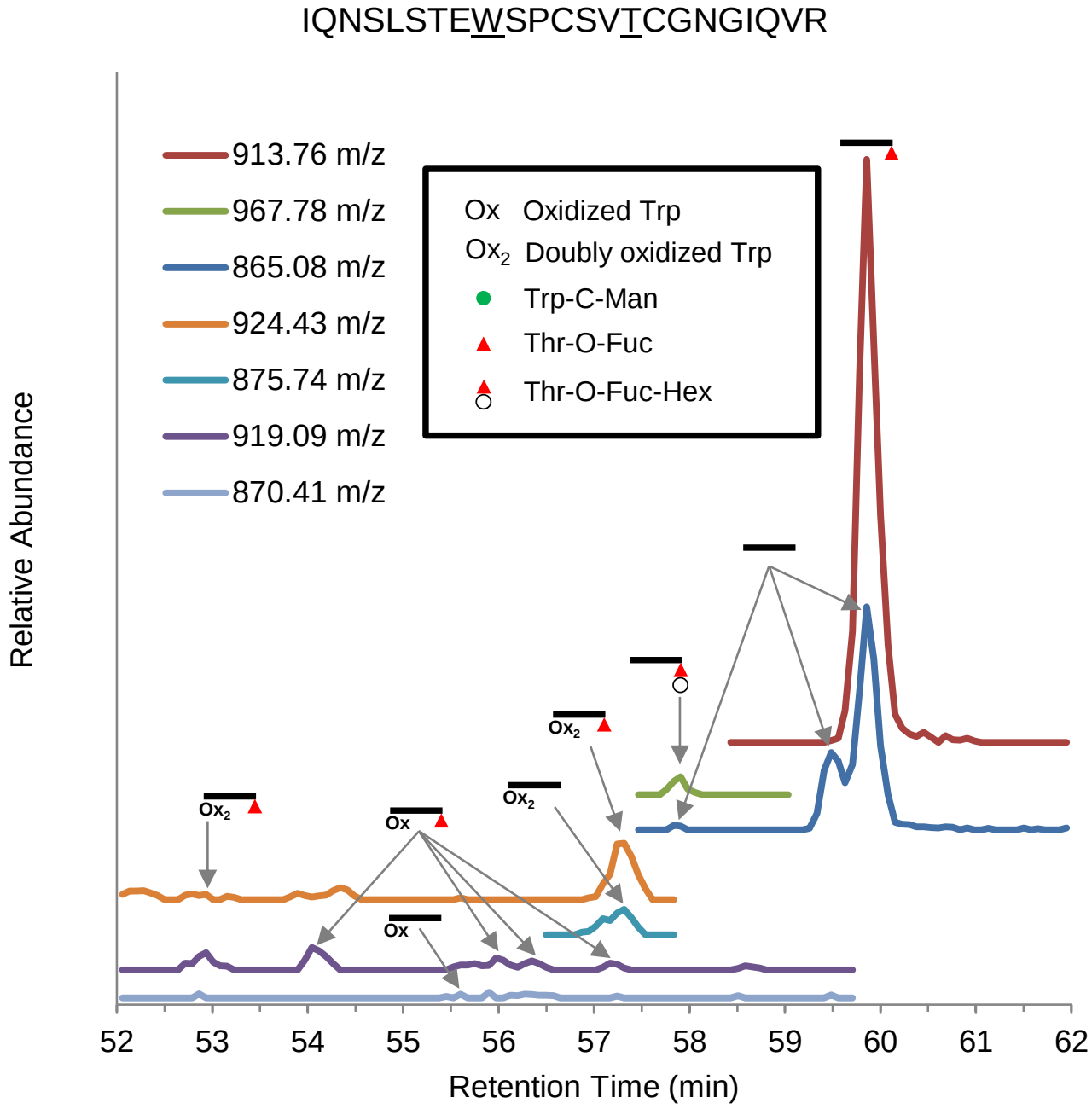


Figure S6. Representative XIC of glycosylated CSP from *P. falciparum* salivary gland sporozoites. The variably oxidized Trp and the variably O-fucosylated Thr are underlined. Traces are offset for clarity. Peaks indicated with arrows were positively identified by PSMs. The Hex of the O-Fuc-Hex disaccharide is indicated as a generic hexose because the identity of the glycan has not been confirmed.

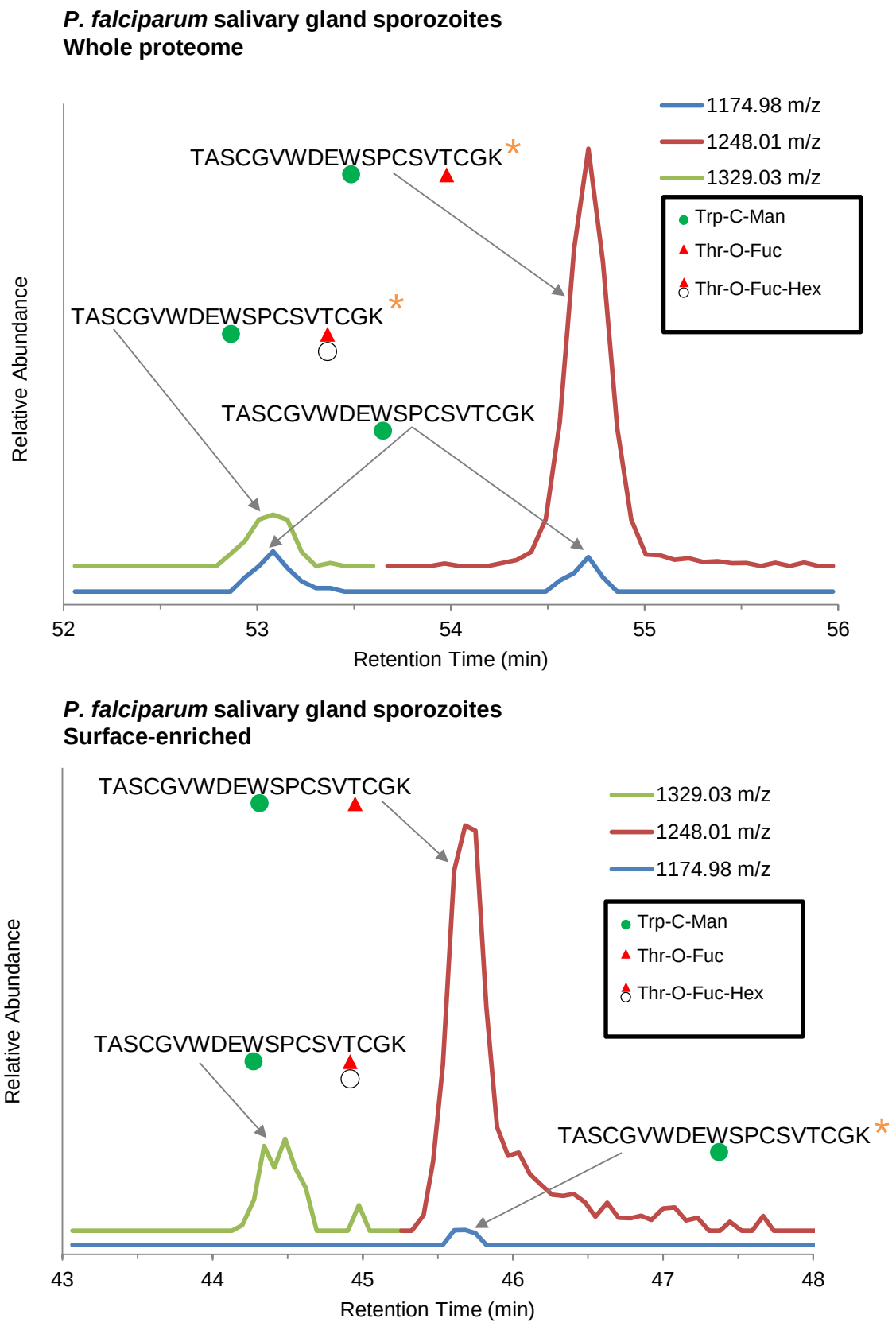


Figure S7. Representative XIC of glycosylated TRAP from *P. falciparum* salivary gland sporozoites. Traces are offset for clarity. Peaks marked with an orange asterisk (*) were not identified by PSMs because the precursor was not selected for MS^2, but their retention time and mass correspond with the indicted glycoform. The Hex of the O-Fuc-Hex disaccharide is indicated as a generic hexose because the identity of the glycan has not been confirmed.

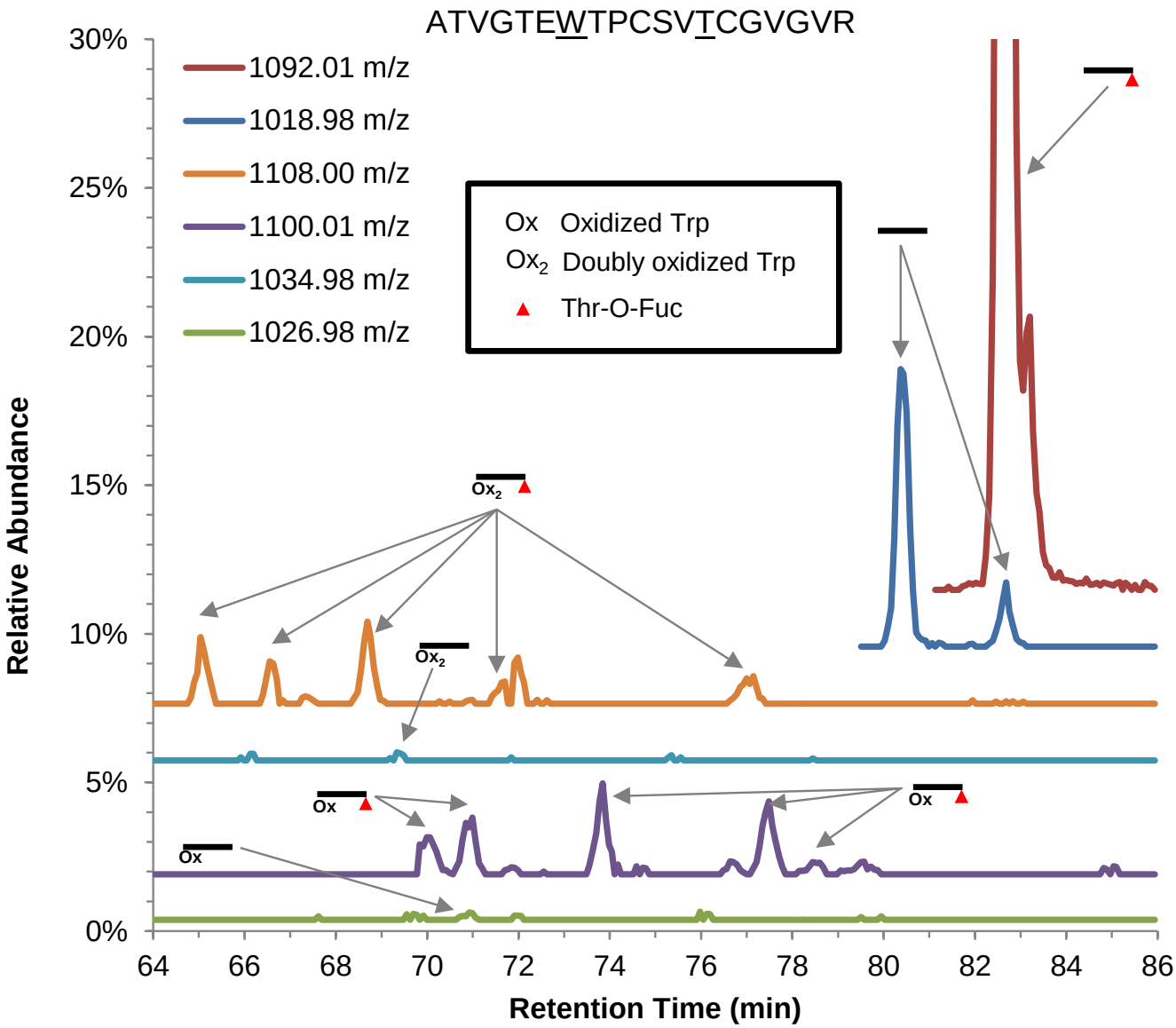


Figure S8. Representative XIC of glycosylated CSP from *P. vivax* VK210 salivary gland sporozoites. The variably oxidized Trp and the variably O-fucosylated Thr are underlined. Traces are offset for clarity. Peaks indicated with arrows were positively identified by PSMs.

VANCGPWDPWTACSVTCGR

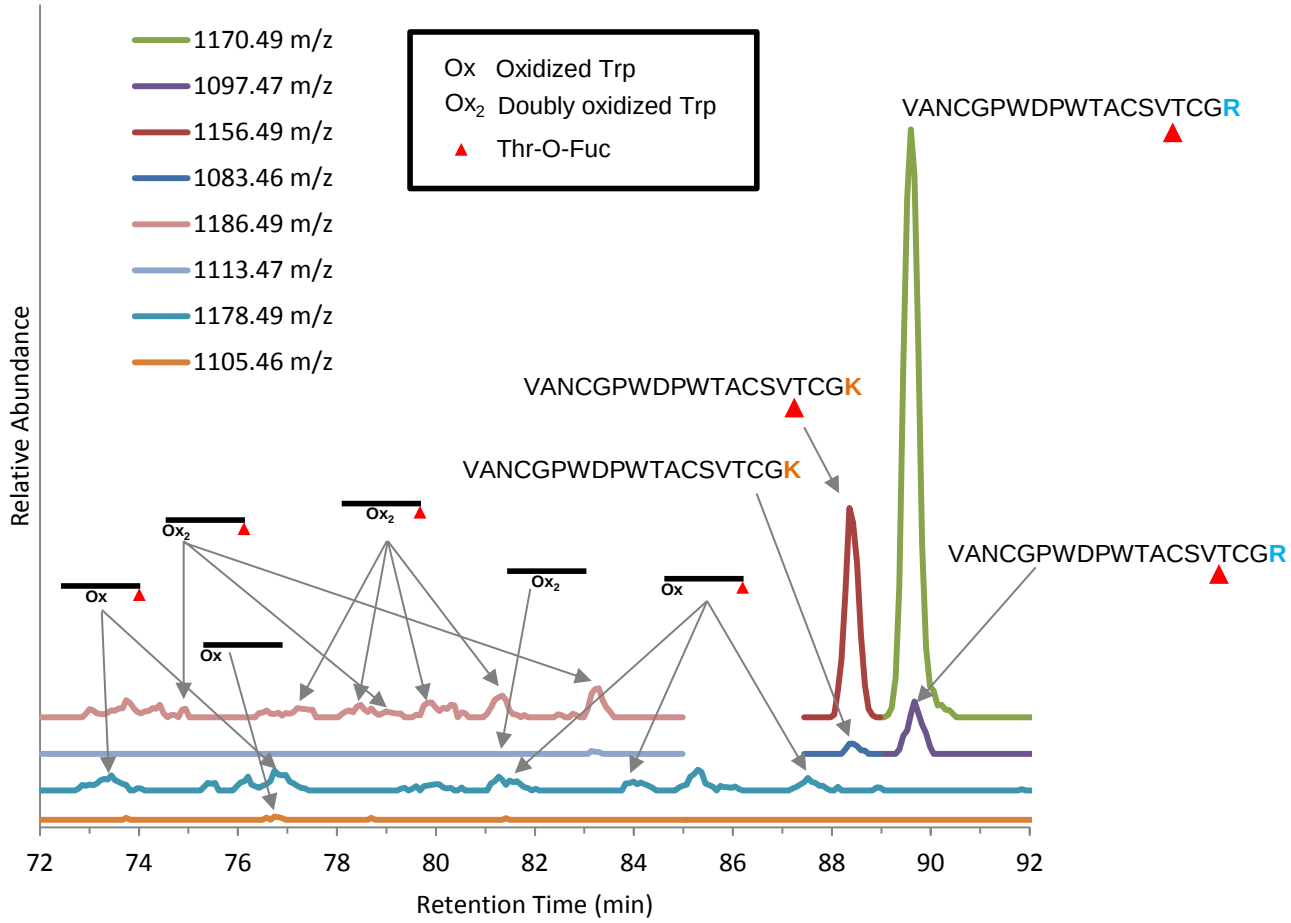


Figure S9. Representative XIC of glycosylated TRAP from *P. vivax* VK210 salivary gland sporozoites. The variably oxidized Trp residues and the variably O-fucosylated Thr are underlined. Traces are offset for clarity. Peaks indicated with arrows were positively identified by PSMs. Note that this sample contained a mixture of two field isolates, each of which carried a different version of the TRAP gene with either a Lys or Arg at the C-terminus of the glycopeptide.

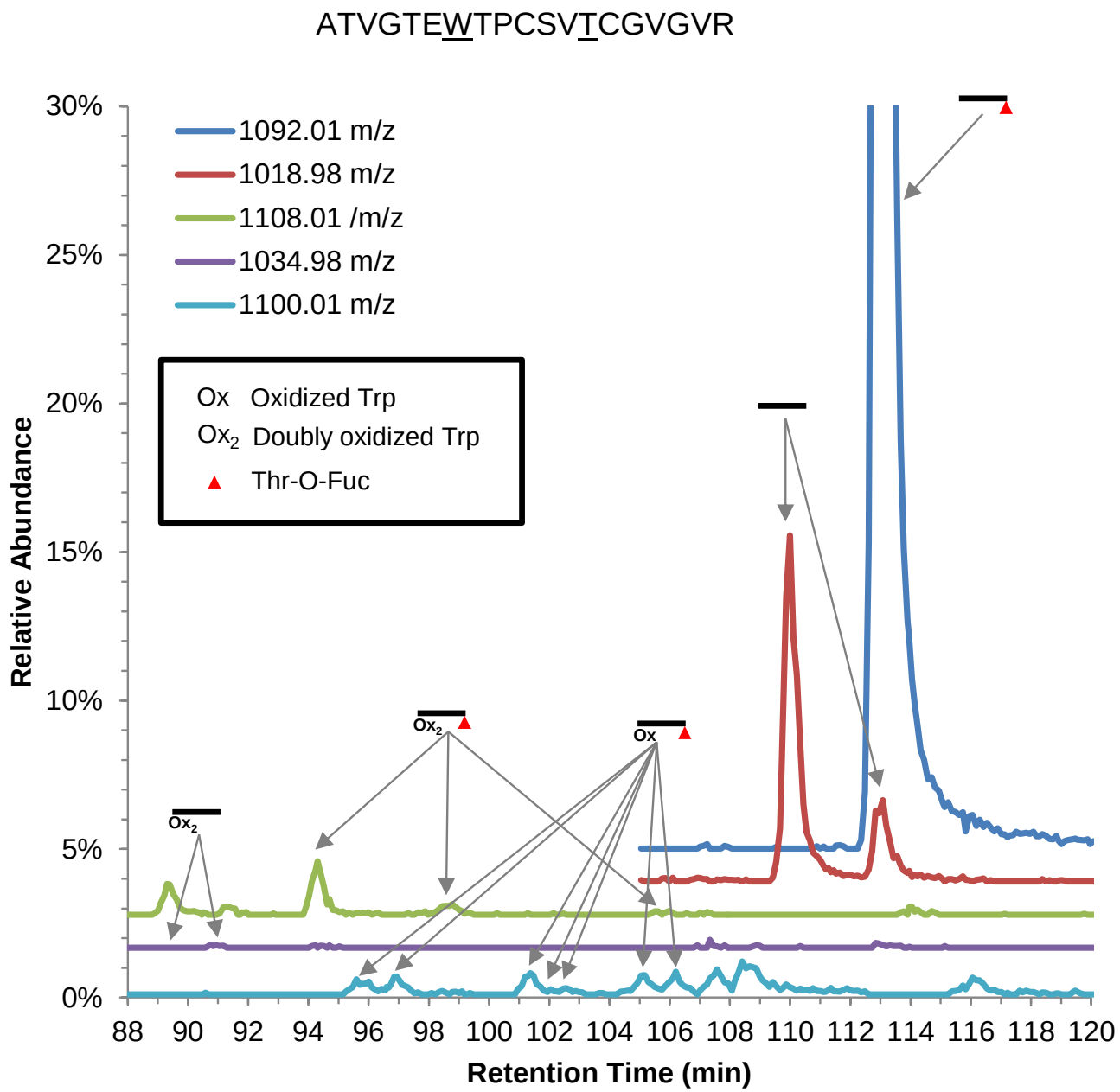


Figure S10. Representative XIC of glycosylated CSP from *P. vivax* VK247 salivary gland sporozoites. The variably oxidized Trp and the variably O-fucosylated Thr are underlined. Traces are offset for clarity. Peaks indicated with arrows were positively identified by PSMs.

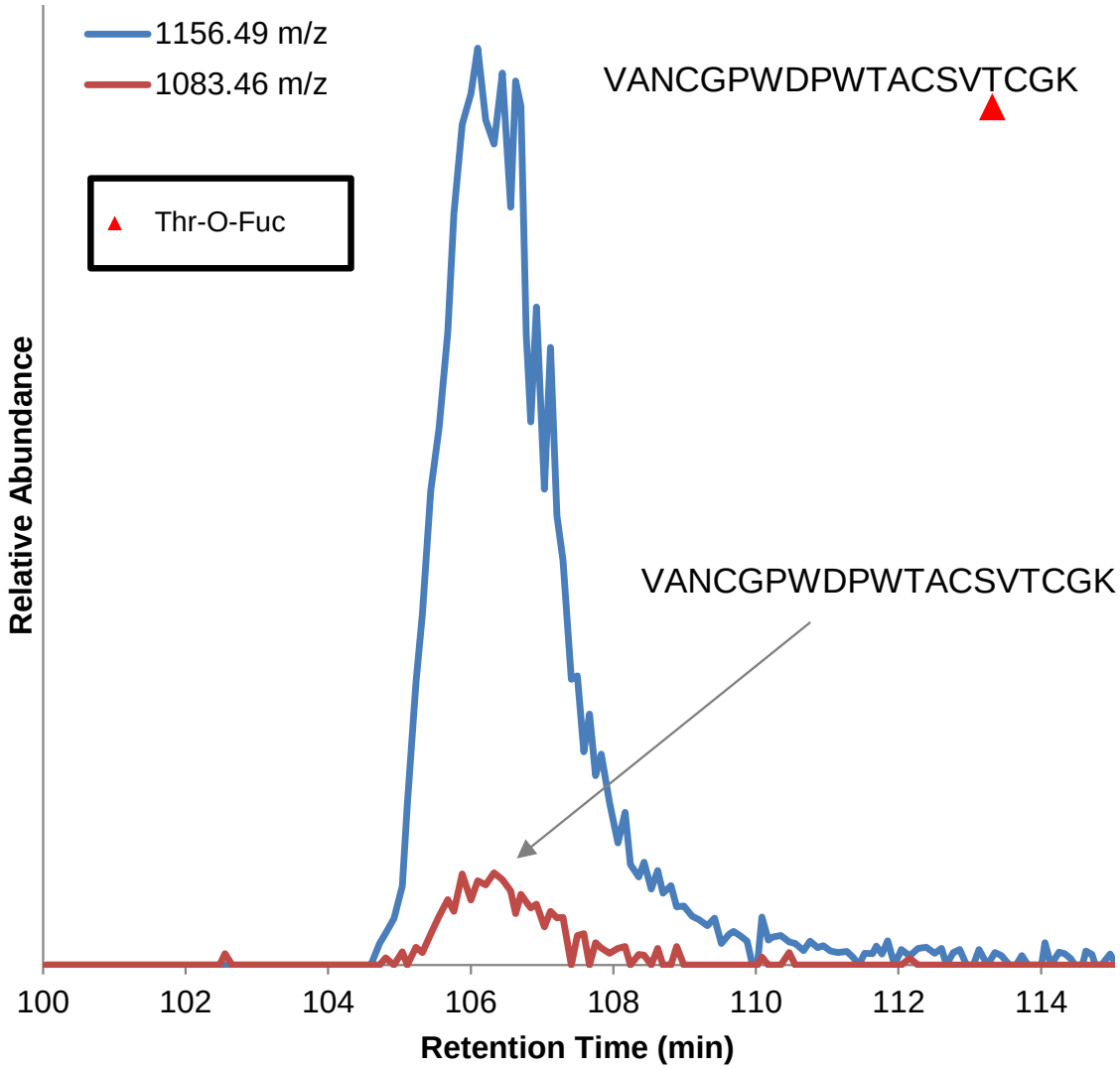


Figure S11. Representative XIC of glycosylated TRAP from *P. vivax* VK247 salivary gland sporozoites.

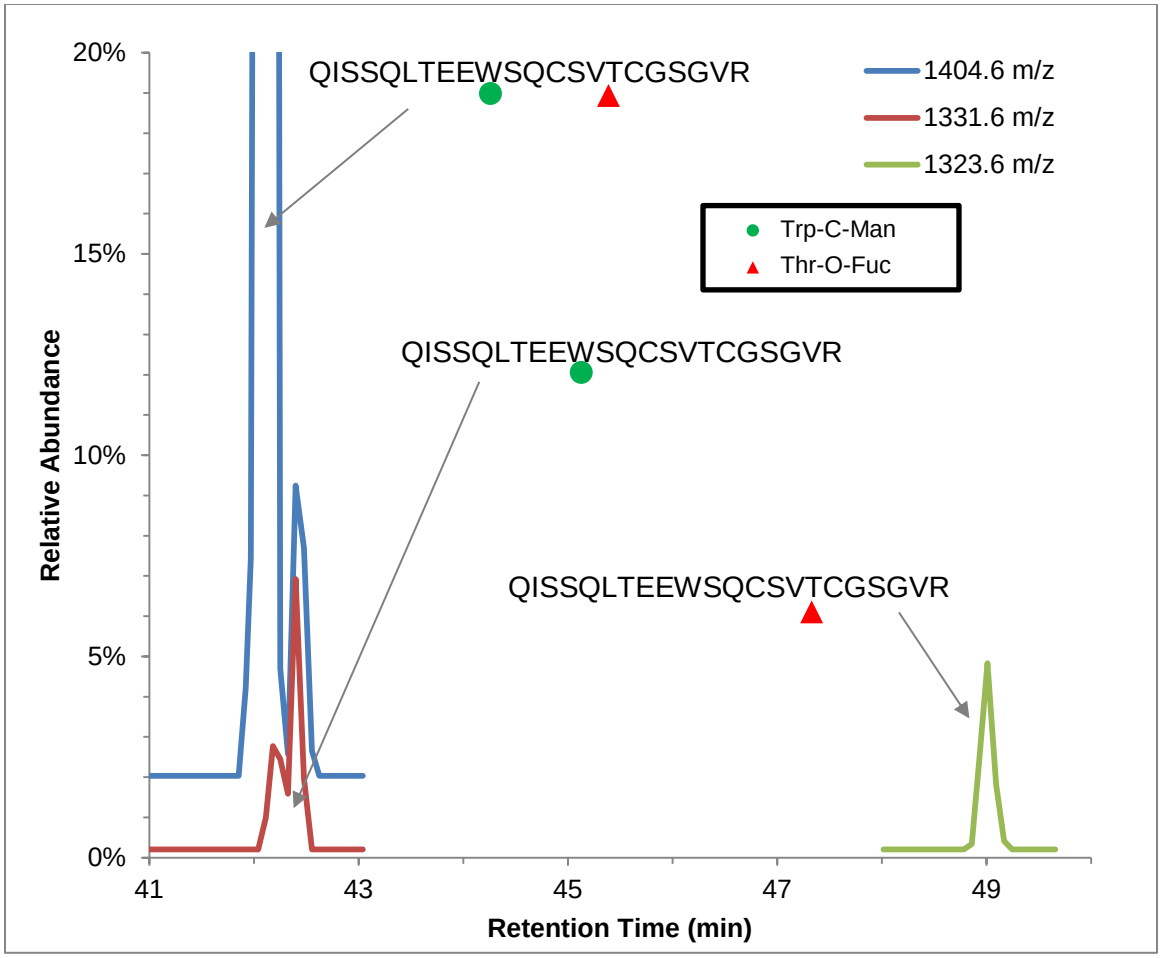


Figure S12. Representative XIC of glycosylated CSP from *P. yoelii* salivary gland sporozoites. Traces are offset for clarity. The reason for the doublet shape of the 1404.6 *m/z* and co-eluting 1331.6 *m/z* peaks is unknown, and may reflect spray instability or closely eluting isobaric species. Only the indicated glycopeptides were identified from MS^2 of these masses collected during the elution window of the peaks.

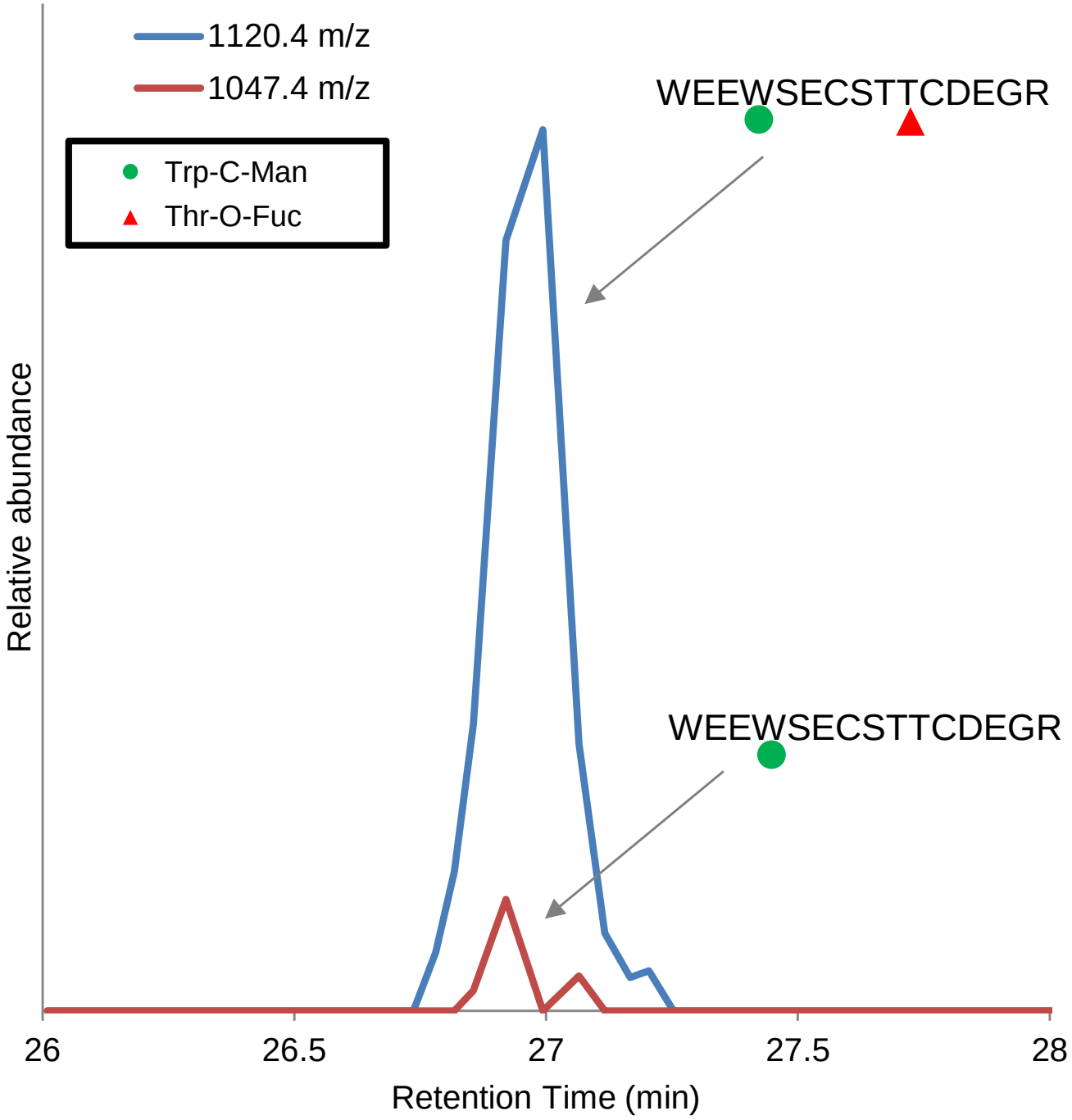


Figure S13. Representative XIC of glycosylated TRAP from *P. yoelii* salivary gland sporozoites.

Ions:
a ☐ 1+ ☐ 2+ ☐ 3+
b ☑ 1+ ☐ 2+ ☐ 3+
c ☐ 1+ ☐ 2+ ☐ 3+
x ☐ 1+ ☐ 2+ ☐ 3+
y ☑ 1+ ☐ 2+ ☐ 3+
z ☐ 1+ ☐ 2+ ☐ 3+
[Deselect All]

Neutral Loss:
☐ NH_3 (*)
☑ H_2O (o)
☐ H_3PO_4 (p)
☑ $C_4H_8O_4$ (-120)
☐ Immonium ions
☐ Reporter ions
☑ Precursor ions

Mass Type:
◉ Mono ○ Avg

Mass Tol: 0.5
◉ Th ○ ppm

Update

Peak Assignment:
◉ Most Intense
○ Nearest Match
☑ Peak Detect

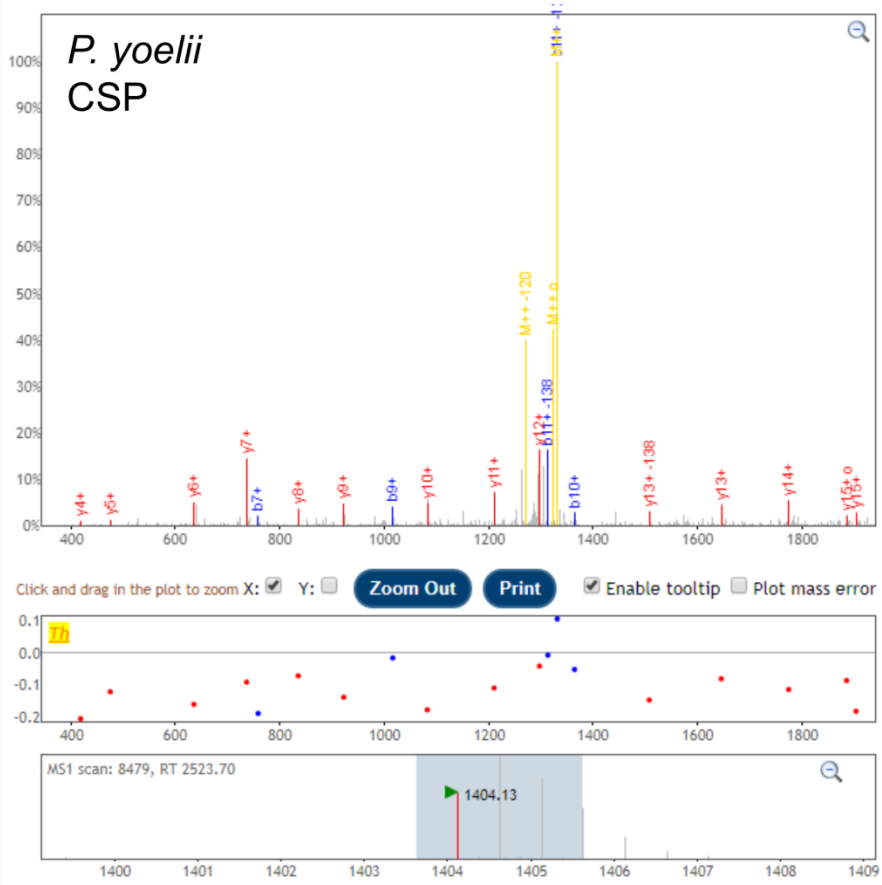
Peak Labels:
◉ Ion ○ m/z
○ None

Width: 650

Height: 400

QISSQLTEEWSQCSVTCGSGVR, MH+ 2661.1767, m/z 1331.0920

File: V2_2011_1111_KES_Pyoelii_11_2.08482.08482.2, Scan: 8482, Exp. m/z: 1404.1322, Charge: 2



Click and drag in the plot to zoom X: ☑ Y: ☐ Zoom Out Print ☑ Enable tooltip ☐ Plot mass error

b+	#	Seq	#	y+
129.0659	1	Q	22	
242.1499	2	I	21	2533.1181
329.1819	3	S	20	2420.0340
416.2140	4	S	19	2333.0020
544.2726	5	Q	18	2245.9700
657.3566	6	L	17	2117.9114
758.4043	7	T	16	2004.8273
887.4469	8	E	15	1903.7797
1016.4895	9	E	14	1774.7371
1364.6216	10	W	13	1645.6945
1451.6536	11	S	12	1297.5623
1579.7122	12	Q	11	1210.5303
1739.7429	13	C	10	1082.4717
1826.7749	14	S	9	922.4411
1925.8433	15	V	8	835.4091
2026.8910	16	T	7	736.3406
2186.9216	17	C	6	635.2930
2243.9431	18	G	5	475.2623
2330.9751	19	S	4	418.2409
2387.9966	20	G	3	331.2088
2487.0650	21	V	2	274.1874
	22	R	1	175.1190

[Click] to move table

Variable Modifications:
C: 57.021464 [13, 17]
W: 162.052824 [10]

Ions:
a ☐ 1+ ☐ 2+ ☐ 3+
b ☑ 1+ ☐ 2+ ☐ 3+
c ☐ 1+ ☐ 2+ ☐ 3+
x ☐ 1+ ☐ 2+ ☐ 3+
y ☑ 1+ ☐ 2+ ☐ 3+
z ☐ 1+ ☐ 2+ ☐ 3+
[Deselect All]

Neutral Loss:
☐ NH_3 (*)
☑ H_2O (o)
☐ H_3PO_4 (p)
☑ $C_4H_8O_4$ (-120)
☐ Immonium ions
☐ Reporter ions
☑ Precursor ions

Mass Type:
◉ Mono ○ Avg

Mass Tol: .4
◉ Th ○ ppm

Update

Peak Assignment:
◉ Most Intense
○ Nearest Match
☑ Peak Detect

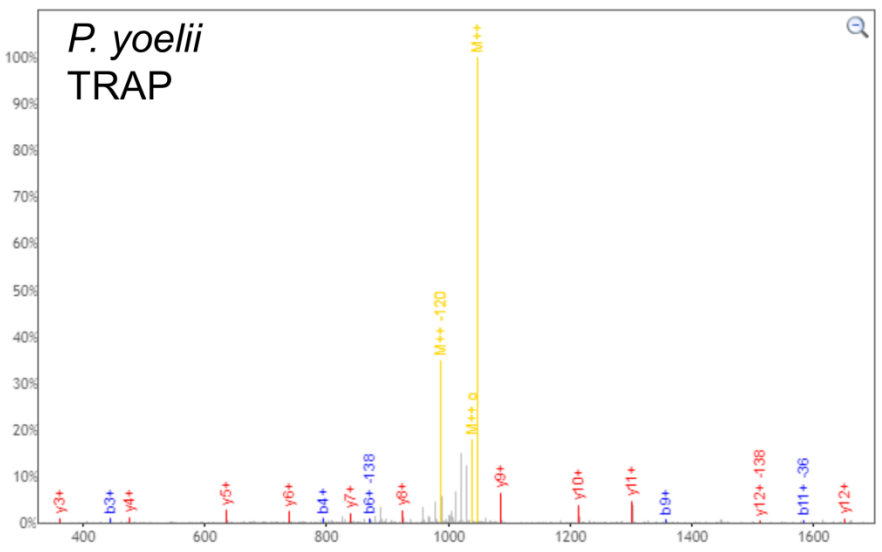
Peak Labels:
◉ Ion ○ m/z
○ None

Width: 650

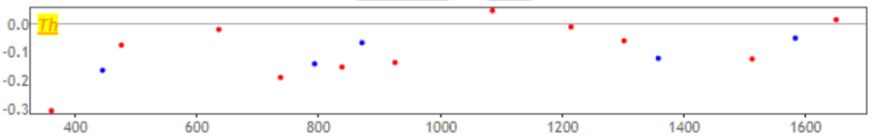
Height: 400

WEEWSECSTTCDEGR, MH+ 2093.7699, m/z 1047.3886

File: V2_2011_1111_KES_Pyoelii_07_2.05953.05953.2, Scan: 5953, Exp. m/z: 1120.4241, Charge: 2



Click and drag in the plot to zoom X: ☑ Y: ☐ Zoom Out Print ☐ Enable tooltip ☐ Plot mass error



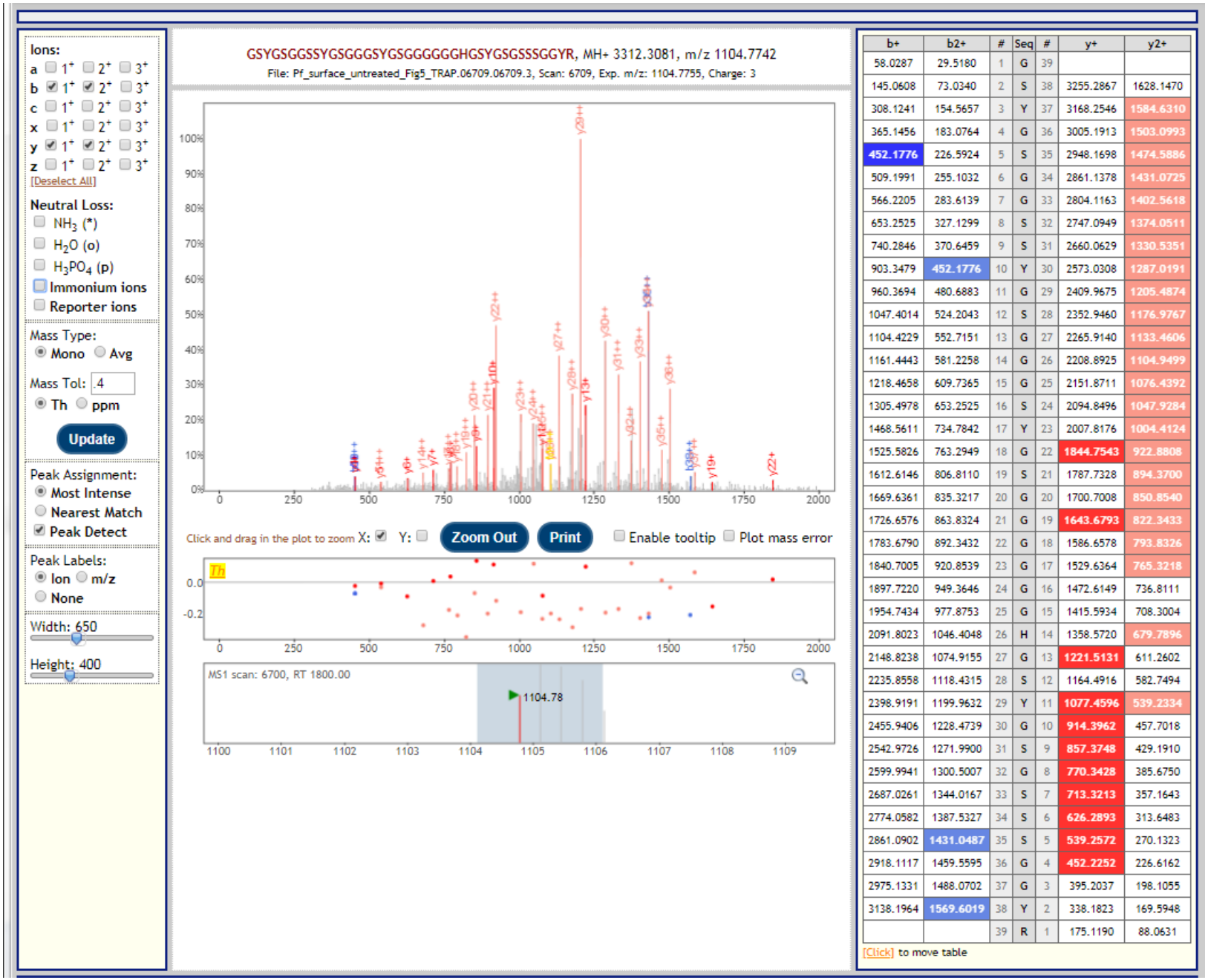
b+	#	Seq	#	y+
187.0866	1	W	15	
316.1292	2	E	14	1907.6906
445.1718	3	E	13	1778.6480
793.3039	4	W	12	1649.6054
880.3359	5	S	11	1301.4733
1009.3785	6	E	10	1214.4412
1169.4092	7	C	9	1085.3986
1256.4412	8	S	8	925.3680
1357.4889	9	T	7	838.3360
1458.5366	10	T	6	737.2883
1618.5672	11	C	5	636.2406
1733.5942	12	D	4	476.2100
1862.6368	13	E	3	361.1830
1919.6582	14	G	2	232.1404
	15	R	1	175.1190

[Click] to move table

Variable Modifications:
C: 57.021464 [7, 11]
W: 162.052824 [4]

Figure S14. Automated identification of TSR glycosylation in *P. yoelii* salivary gland sporozoites. A previously published proteomic analysis of salivary gland sporozoites of the rodent-infective *Plasmodium* species *P. yoelii* (Lindner and Swearingen *et al*. 2013 PMID:23325771) was re-analyzed with the mass offset approach described here, providing the first reported evidence of O-fucosylation and C-mannosylation of TSR domains in that species. Sample MS^2 spectra for the CSP peptide (top) and the TRAP peptide (bottom) are shown as automatically annotated by the Lorikeet spectrum viewer incorporated into the TPP. Both examples show a mass difference of -146.06 Da between the observed precursor *m/z* and that of the matched peptide, corresponding to neutral loss of O-Fuc. The C-mannosylated peptide precursor after neutral loss of O-Fuc is the dominant peak in both spectra (yellow peak annotated M++. In the CSP spectrum, this peak is isobaric with the b11-120 Da peak). Neutral loss of 120.04 Da from the M++ peak due to cross-ring cleavage of C-Man is also indicated.

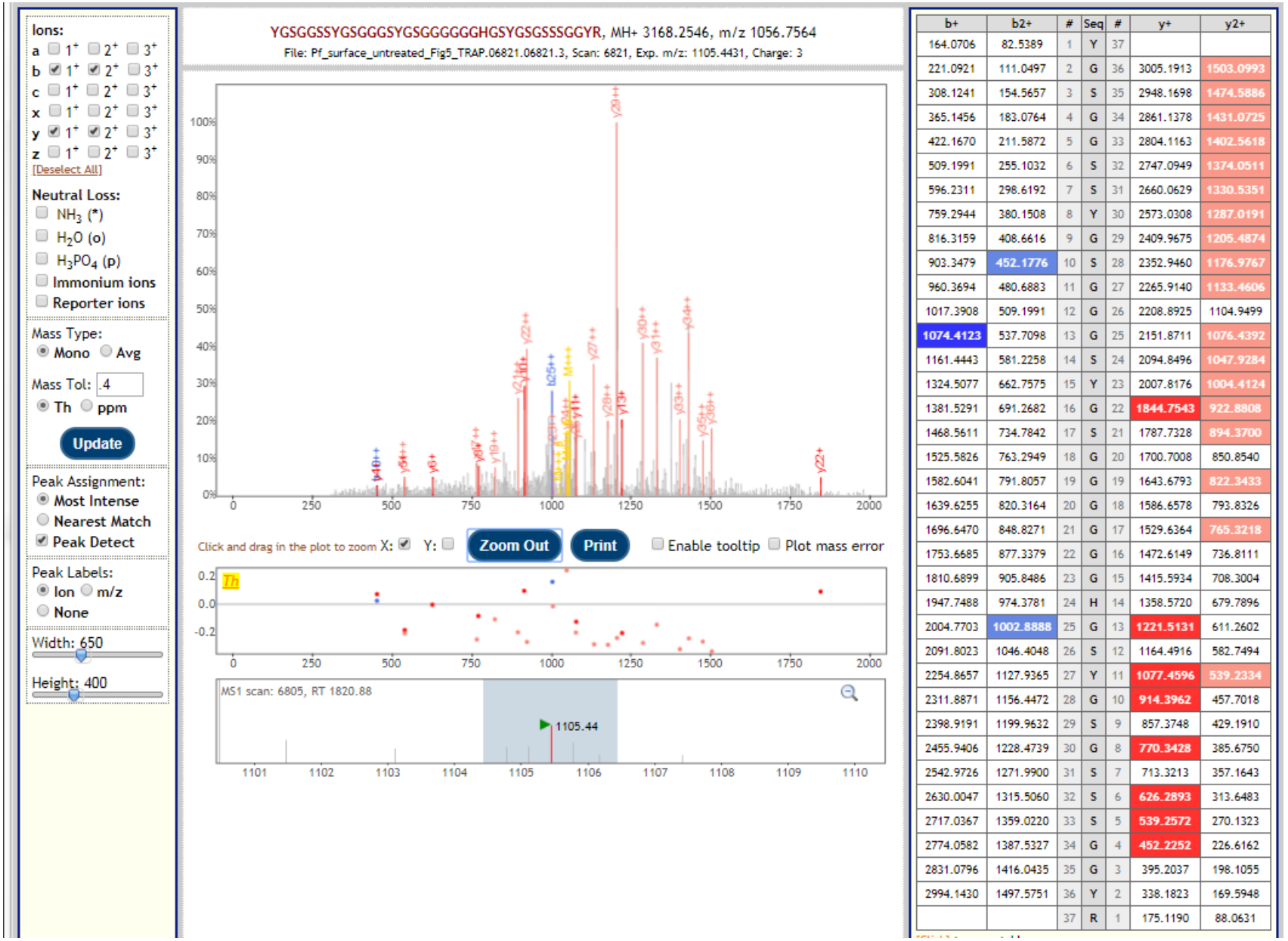
Figure S15. Example false positive peptides showing neutral loss of O-linked glycans.



b+	b2+	#	Seq	#	y+	y2+
58.0287	29.5180	1	G	39		
145.0608	73.0340	2	S	38	3255.2867	1628.1470
308.1241	154.5657	3	Y	37	3168.2546	1584.6310
365.1456	183.0764	4	G	36	3005.1913	1503.0993
452.1776	226.5924	5	S	35	2948.1698	1474.5886
509.1991	255.1032	6	G	34	2861.1378	1431.0725
566.2205	283.6139	7	G	33	2804.1163	1402.5618
653.2525	327.1299	8	S	32	2747.0949	1374.0511
740.2846	370.6459	9	S	31	2660.0629	1330.5351
903.3479	452.1776	10	Y	30	2573.0308	1287.0191
960.3694	480.6883	11	G	29	2409.9675	1205.4874
1047.4014	524.2043	12	S	28	2352.9460	1176.9767
1104.4229	552.7151	13	G	27	2265.9140	1133.4606
1161.4443	581.2258	14	G	26	2208.8925	1104.9499
1218.4658	609.7365	15	G	25	2151.8711	1076.4392
1305.4978	653.2525	16	S	24	2094.8496	1047.9284
1468.5611	734.7842	17	Y	23	2007.8176	1004.4124
1525.5826	763.2949	18	G	22	1844.7543	922.8808
1612.6146	806.8110	19	S	21	1787.7328	894.3700
1669.6361	835.3217	20	G	20	1700.7008	850.8540
1726.6576	863.8324	21	G	19	1643.6793	822.3433
1783.6790	892.3432	22	G	18	1586.6578	793.8326
1840.7005	920.8539	23	G	17	1529.6364	765.3218
1897.7220	949.3646	24	G	16	1472.6149	736.8111
1954.7434	977.8753	25	G	15	1415.5934	708.3004
2091.8023	1046.4048	26	H	14	1358.5720	679.7896
2148.8238	1074.9155	27	G	13	1221.5131	611.2602
2235.8558	1118.4315	28	S	12	1164.4916	582.7494
2398.9191	1199.9632	29	Y	11	1077.4596	539.2334
2455.9406	1228.4739	30	G	10	914.3962	457.7018
2542.9726	1271.9900	31	S	9	857.3748	429.1910
2599.9941	1300.5007	32	G	8	770.3428	385.6750
2687.0261	1344.0167	33	S	7	713.3213	357.1643
2774.0582	1387.5327	34	S	6	626.2893	313.6483
2861.0902	1431.0487	35	S	5	539.2572	270.1323
2918.1117	1459.5595	36	G	4	452.2252	226.6162
2975.1331	1488.0702	37	G	3	395.2037	198.1055
3138.1964	1569.6019	38	Y	2	338.1823	169.5948
		39	R	1	175.1190	88.0631

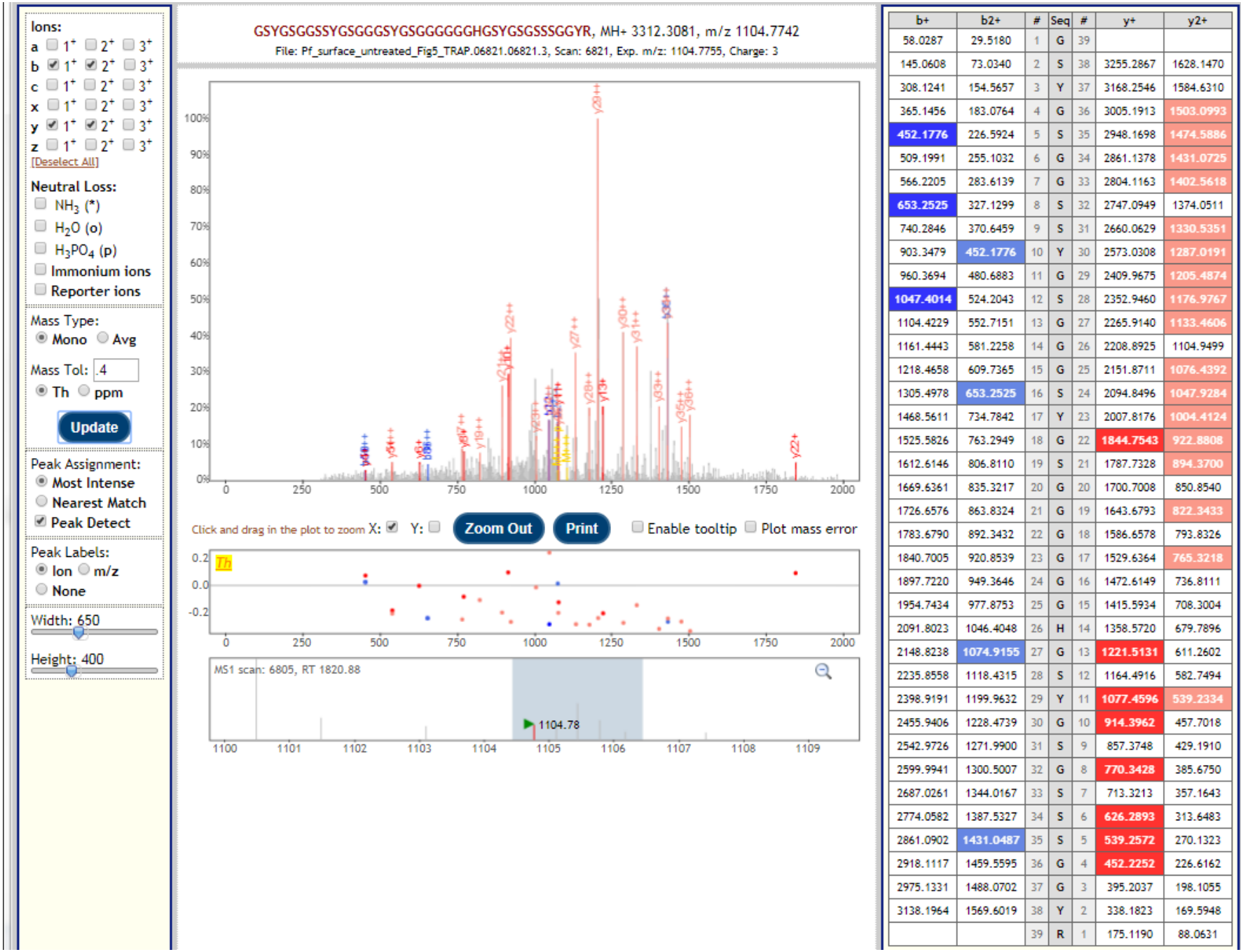
[Click] to move table

Figure S15a. A sample high-quality PSM identified the fully tryptic human keratin peptide GSYGSGGSSYGSGGGSYGSGGGGGGHGSYGSGSSSGGYR.



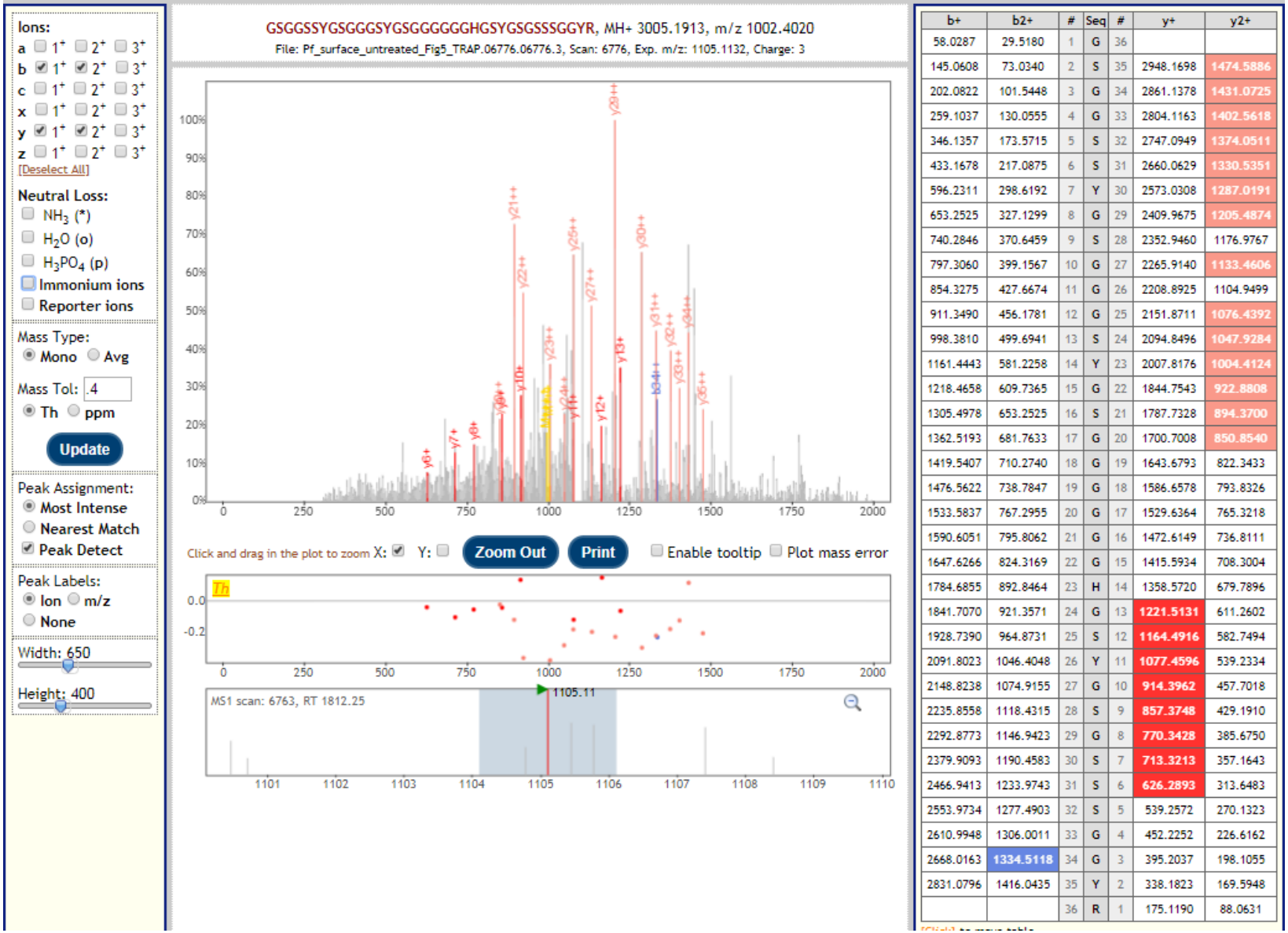
b+	b2+	#	Seq	#	y+	y2+
164.0706	82.5389	1	Y	37		
221.0921	111.0497	2	G	36	3005.1913	1503.0993
308.1241	154.5657	3	S	35	2948.1698	1474.5886
365.1456	183.0764	4	G	34	2861.1378	1431.0725
422.1670	211.5872	5	G	33	2804.1163	1402.5618
509.1991	255.1032	6	S	32	2747.0949	1374.0511
596.2311	298.6192	7	S	31	2660.0629	1330.5351
759.2944	380.1508	8	Y	30	2573.0308	1287.0191
816.3159	408.6616	9	G	29	2409.9675	1205.4874
903.3479	452.1776	10	S	28	2352.9460	1176.9767
960.3694	480.6883	11	G	27	2265.9140	1133.4606
1017.3908	509.1991	12	G	26	2208.8925	1104.9499
1074.4123	537.7098	13	G	25	2151.8711	1076.4392
1161.4443	581.2258	14	S	24	2094.8496	1047.9284
1324.5077	662.7575	15	Y	23	2007.8176	1004.4124
1381.5291	691.2682	16	G	22	1844.7543	922.8808
1468.5611	734.7842	17	S	21	1787.7328	894.3700
1525.5826	763.2949	18	G	20	1700.7008	850.8540
1582.6041	791.8057	19	G	19	1643.6793	822.3433
1639.6255	820.3164	20	G	18	1586.6578	793.8326
1696.6470	848.8271	21	G	17	1529.6364	765.3218
1753.6685	877.3379	22	G	16	1472.6149	736.8111
1810.6899	905.8486	23	G	15	1415.5934	708.3004
1947.7488	974.3781	24	H	14	1358.5720	679.7896
2004.7703	1002.8888	25	G	13	1221.5131	611.2602
2091.8023	1046.4048	26	S	12	1164.4916	582.7494
2254.8657	1127.9365	27	Y	11	1077.4596	539.2334
2311.8871	1156.4472	28	G	10	914.3962	457.7018
2398.9191	1199.9632	29	S	9	857.3748	429.1910
2455.9406	1228.4739	30	G	8	770.3428	385.6750
2542.9726	1271.9900	31	S	7	713.3213	357.1643
2630.0047	1315.5060	32	S	6	626.2893	313.6483
2717.0367	1359.0220	33	S	5	539.2572	270.1323
2774.0582	1387.5327	34	G	4	452.2252	226.6162
2831.0796	1416.0435	35	G	3	395.2037	198.1055
2994.1430	1497.5751	36	Y	2	338.1823	169.5948
		37	R	1	175.1190	88.0631

Figure S15b. A representative PSM identifying the same keratin peptide seen in Figure S15a, but spuriously identifying it as a semi-tryptic fragment with neutral loss of 146.06. The semi-tryptic fragment GS.YGSGGSSYGSGGGSYGSGGGGGGHGSYGSGSSSGGYR has a mass loss of Gly-Ser = 144.05 Da. In addition, the +2H isotope peak 1105.44 *m/z* was incorrectly selected as the monoisotopic peak for this peptide. The total mass loss is then 144.05 +2*1.0078 = 146.07, which is within the matching tolerance of the searched-for neutral loss of 146.06 Da for putative loss of O-Fuc. Because of the extensive y-ion coverage, the PSM still received a high score. The more likely correct annotation is shown in Figure S15c.



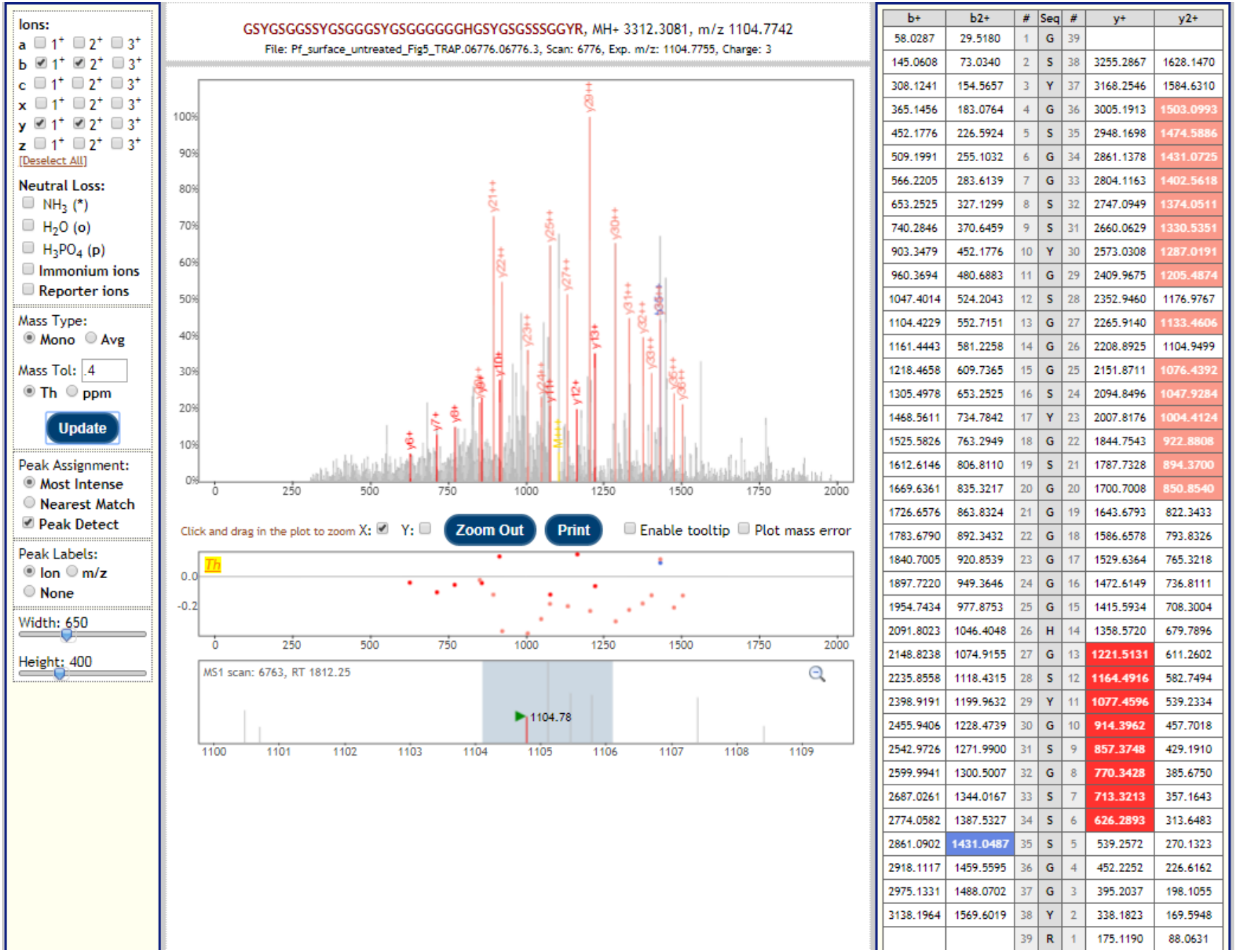
b+	b2+	#	Seq	#	y+	y2+
58.0287	29.5180	1	G	39		
145.0608	73.0340	2	S	38	3255.2867	1628.1470
308.1241	154.5657	3	Y	37	3168.2546	1584.6310
365.1456	183.0764	4	G	36	3005.1913	1503.0993
452.1776	226.5924	5	S	35	2948.1698	1474.5886
509.1991	255.1032	6	G	34	2861.1378	1431.0725
566.2205	283.6139	7	G	33	2804.1163	1402.5618
653.2525	327.1299	8	S	32	2747.0949	1374.0511
740.2846	370.6459	9	S	31	2660.0629	1330.5351
903.3479	452.1776	10	Y	30	2573.0308	1287.0191
960.3694	480.6883	11	G	29	2409.9675	1205.4874
1047.4014	524.2043	12	S	28	2352.9460	1176.9767
1104.4229	552.7151	13	G	27	2265.9140	1133.4606
1161.4443	581.2258	14	G	26	2208.8925	1104.9499
1218.4658	609.7365	15	G	25	2151.8711	1076.4392
1305.4978	653.2525	16	S	24	2094.8496	1047.9284
1468.5611	734.7842	17	Y	23	2007.8176	1004.4124
1525.5826	763.2949	18	G	22	1844.7543	922.8808
1612.6146	806.8110	19	S	21	1787.7328	894.3700
1669.6361	835.3217	20	G	20	1700.7008	850.8540
1726.6576	863.8324	21	G	19	1643.6793	822.3433
1783.6790	892.3432	22	G	18	1586.6578	793.8326
1840.7005	920.8539	23	G	17	1529.6364	765.3218
1897.7220	949.3646	24	G	16	1472.6149	736.8111
1954.7434	977.8753	25	G	15	1415.5934	708.3004
2091.8023	1046.4048	26	H	14	1358.5720	679.7896
2148.8238	1074.9155	27	G	13	1221.5131	611.2602
2235.8558	1118.4315	28	S	12	1164.4916	582.7494
2398.9191	1199.9632	29	Y	11	1077.4596	539.2334
2455.9406	1228.4739	30	G	10	914.3962	457.7018
2542.9726	1271.9900	31	S	9	857.3748	429.1910
2599.9941	1300.5007	32	G	8	770.3428	385.6750
2687.0261	1344.0167	33	S	7	713.3213	357.1643
2774.0582	1387.5327	34	S	6	626.2893	313.6483
2861.0902	1431.0487	35	S	5	539.2572	270.1323
2918.1117	1459.5595	36	G	4	452.2252	226.6162
2975.1331	1488.0702	37	G	3	395.2037	198.1055
3138.1964	1569.6019	38	Y	2	338.1823	169.5948
		39	R	1	175.1190	88.0631

Figure S15c. The more likely correct annotation of the peptide from Figure S15b, matched to the fully tryptic peptide with no neutral loss.



b+	b2+	#	Seq	#	y+	y2+
58.0287	29.5180	1	G	36		
145.0608	73.0340	2	S	35	2948.1698	1474.5886
202.0822	101.5448	3	G	34	2861.1378	1431.0725
259.1037	130.0555	4	G	33	2804.1163	1402.5618
346.1357	173.5715	5	S	32	2747.0949	1374.0511
433.1678	217.0875	6	S	31	2660.0629	1330.5351
596.2311	298.6192	7	Y	30	2573.0308	1287.0191
653.2525	327.1299	8	G	29	2409.9675	1205.4874
740.2846	370.6459	9	S	28	2352.9460	1176.9767
797.3060	399.1567	10	G	27	2265.9140	1133.4606
854.3275	427.6674	11	G	26	2208.8925	1104.9499
911.3490	456.1781	12	G	25	2151.8711	1076.4392
998.3810	499.6941	13	S	24	2094.8496	1047.9284
1161.4443	581.2258	14	Y	23	2007.8176	1004.4124
1218.4658	609.7365	15	G	22	1844.7543	922.8808
1305.4978	653.2525	16	S	21	1787.7328	894.3700
1362.5193	681.7633	17	G	20	1700.7008	850.8540
1419.5407	710.2740	18	G	19	1643.6793	822.3433
1476.5622	738.7847	19	G	18	1586.6578	793.8326
1533.5837	767.2955	20	G	17	1529.6364	765.3218
1590.6051	795.8062	21	G	16	1472.6149	736.8111
1647.6266	824.3169	22	G	15	1415.5934	708.3004
1784.6855	892.8464	23	H	14	1358.5720	679.7896
1841.7070	921.3571	24	G	13	1221.5131	611.2602
1928.7390	964.8731	25	S	12	1164.4916	582.7494
2091.8023	1046.4048	26	Y	11	1077.4596	539.2334
2148.8238	1074.9155	27	G	10	914.3962	457.7018
2235.8558	1118.4315	28	S	9	857.3748	429.1910
2292.8773	1146.9423	29	G	8	770.3428	385.6750
2379.9093	1190.4583	30	S	7	713.3213	357.1643
2466.9413	1233.9743	31	S	6	626.2893	313.6483
2553.9734	1277.4903	32	S	5	539.2572	270.1323
2610.9948	1306.0011	33	G	4	452.2252	226.6162
2668.0163	1334.5118	34	G	3	395.2037	198.1055
2831.0796	1416.0435	35	Y	2	338.1823	169.5948
		36	R	1	175.1190	88.0631

Figure S15d. A representative PSM identifying the same keratin peptide seen in Figure S15a, but spuriously identifying it as a semi-tryptic fragment with neutral loss of 308.11. The semi-tryptic fragment GSY.GSGGSSYGSGGGSYGSGGGGGGHGSYGSGSSSGGYR has a mass loss of Gly-Ser-Tyr = 307.12 Da. In addition, the +1H isotope peak 1105.11 *m/z* was incorrectly selected as the monoisotopic peak for this peptide. The total mass loss is then 307.12 +1.0078 = 308.12, which is within the matching tolerance of the searched-for neutral loss of 308.11 Da for putative loss of O-Fuc-Glc. Because of the extensive y-ion coverage, the PSM still received a high score. The correct annotation is shown in Figure S15e.



b+	b2+	#	Seq	#	y+	y2+
58.0287	29.5180	1	G	39		
145.0608	73.0340	2	S	38	3255.2867	1628.1470
308.1241	154.5657	3	Y	37	3168.2546	1584.6310
365.1456	183.0764	4	G	36	3005.1913	1503.0993
452.1776	226.5924	5	S	35	2948.1698	1474.5886
509.1991	255.1032	6	G	34	2861.1378	1431.0725
566.2205	283.6139	7	G	33	2804.1163	1402.5618
653.2525	327.1299	8	S	32	2747.0949	1374.0511
740.2846	370.6459	9	S	31	2660.0629	1330.5351
903.3479	452.1776	10	Y	30	2573.0308	1287.0191
960.3694	480.6883	11	G	29	2409.9675	1205.4874
1047.4014	524.2043	12	S	28	2352.9460	1176.9767
1104.4229	552.7151	13	G	27	2265.9140	1133.4606
1161.4443	581.2258	14	G	26	2208.8925	1104.9499
1218.4658	609.7365	15	G	25	2151.8711	1076.4392
1305.4978	653.2525	16	S	24	2094.8496	1047.9284
1468.5611	734.7842	17	Y	23	2007.8176	1004.4124
1525.5826	763.2949	18	G	22	1844.7543	922.8808
1612.6146	806.8110	19	S	21	1787.7328	894.3700
1669.6361	835.3217	20	G	20	1700.7008	850.8540
1726.6576	863.8324	21	G	19	1643.6793	822.3433
1783.6790	892.3432	22	G	18	1586.6578	793.8326
1840.7005	920.8539	23	G	17	1529.6364	765.3218
1897.7220	949.3646	24	G	16	1472.6149	736.8111
1954.7434	977.8753	25	G	15	1415.5934	708.3004
2091.8023	1046.4048	26	H	14	1358.5720	679.7896
2148.8238	1074.9155	27	G	13	1221.5131	611.2602
2235.8558	1118.4315	28	S	12	1164.4916	582.7494
2398.9191	1199.9632	29	Y	11	1077.4596	539.2334
2455.9406	1228.4739	30	G	10	914.3962	457.7018
2542.9726	1271.9900	31	S	9	857.3748	429.1910
2599.9941	1300.5007	32	G	8	770.3428	385.6750
2687.0261	1344.0167	33	S	7	713.3213	357.1643
2774.0582	1387.5327	34	S	6	626.2893	313.6483
2861.0902	1431.0487	35	S	5	539.2572	270.1323
2918.1117	1459.5595	36	G	4	452.2252	226.6162
2975.1331	1488.0702	37	G	3	395.2037	198.1055
3138.1964	1569.6019	38	Y	2	338.1823	169.5948
		39	R	1	175.1190	88.0631

Figure S15e. The more likely correct annotation of the peptide from Figure S15d, matched to the fully tryptic peptide with no neutral loss