

Non-covalent dimer formation in LC-MS analysis

Katalin F. Medzihradzky, Department of Pharmaceutical Chemistry, School of Pharmacy,
University of California San Francisco, San Francisco, CA; & BRC, Institute of
Biochemistry, Szeged, Hungary

Katalin F. Medzihradzky
Department of Pharmaceutical Chemistry
University of California, San Francisco
600 16th Street
Genentech Hall, N474A, , Box 2240
San Francisco, CA 94158-2517
e-mail: folk1@cgl.ucsf.edu

My observations are based on information presented in the article “Novel LC-MS2 Product Dependent Parallel Data Acquisition Function and Data Analysis Workflow for Sequencing and Identification of Intact Glycopeptides.” written by Sz-Wei Wu, Tsung-Hsien Pu, Rosa Viner, and Kay-Hooi Khoo, published in Anal Chem. 2014 Jun 3;86(11):5478-86. doi: 10.1021/ac500945m, and raw data associated with it.

Supplementary Figures S1-S22

Figure S1

Copied from

“Novel LC-MS2 Product Dependent Parallel Data Acquisition Function and Data Analysis Workflow for Sequencing and Identification of Intact Glycopeptides.” written by Sz-Wei Wu, Tsung-Hsien Pu, Rosa Viner, and Kay-Hooi Khoo.

Anal Chem. 2014 Jun 3;86(11):5478-86. doi: 10.1021/ac500945m.

Table S1-1. Subset of Incorrect Glycopeptide Identification by Direct Database Search Despite Giving Positive ID by Y1-based Search

Sequence	Byonic search results			Y1-based Mascot search			Y1-based Byonic search			Byonic-assigned glycan composition [-H3N2]*					Correct Y1		Glycan Composition [-H3N2]*									
	PSMs	n	Modification Type(s)	m/z	z	Score	Score	Identified peptide	Score	Identified peptide	F	H	N	Ac	Gc	Na	Total mass (Da)#	Y1 (m/z)	Z	F	H	N	Ac	Gc	+ Extra Mod	Total mass (Da)#
DIVSSDFLSNM SMD FQNHLGSCQK	6	2	C[+57], M[+16]*2	1214.49	4	569.88	34.45	same	518.98	same	3	2	2	0	0	0	2060.76	1498.41	2	1	2	2	1	0		2059.73
	1	1	C[+57], M[+16]*2	1227.47	4	434.29	27.26	same	315.71	same	0	3	2	0	1	1	2114.71	1498.7	2	1	2	3	1	0		2116.76
	2	2	C[+57], M[+16]*2	1251.01	4	497.95	29.65	same	450.78	same	4	2	2	0	0	0	2206.81	1498.39	2	2	2	2	1	0		2205.79
DLSLSINATNIKHFKNCT S ISGDLHILPVAFR	1	1	C[+57], M[+16]	1141.47	4	350.06			301.9	same	0	3	2	0	0	0	1784.63	1490.61	2	1	2	2	0	0	17	1785.64
	4	2	C[+57]	1321.19	5	905.72	34.85	NCT S ISGDLHILPVAFR	464.9	NCT S ISGDLHILPVAFR	2	7	4	0	0	0	3131.12	1052.62	2	0	3	0	0	0	3322.47	4700.95
	2	2	C[+57]	1345	5	948.61	43.5	NCT S ISGDLHILPVAFR	480.24	NCT S ISGDLHILPVAFR	1	2	5	3	0	0	3251.16	1052.72	2	0	4	0	0	0	3277.90	4819.00
EITGFLLIQAWPENR	2	2		1137.86	3	260.38			508.24	same	0	2	2	0	0	0	1622.58	996.15	2	1	1	2	0	0	17.00	1623.59
HRKNCT S ISGDLHILPVAFR	7	2	C[+57]	893.67	4	1448.27	40	same	658.47	same	0	1	1	0	0	0	1257.45	839.29	3	0	2	0	0	0	42	1258.42
NCT S ISGDLHILPVAFR	2	2	C[+57]	804.37	5	1262.09	22.45	same	523.67	same	0	0	4	0	0	0	1704.63	839.49	3	0	5	0	0	0		1702.58
	2	2	C[+57]	777.55	5	1376.8	28.12	same	564.39	same	0	1	1	1	0	1	1571.53	629.82	4	0	4	0	0	0	28	1568.53
	1	1	C[+57]	1205.27	4	417.92			212.72	NCT S ISGDLHILPVAFR	0	3	4	1	0	1	2504.88	1052.56	2	2	5	3	1	0	Na	2917.01
	8	2	C[+57]	1161.84	3	864.42	74.2	same	607.01	same	0	3	1	0	0	0	1581.56	1052.64	2	0	4	0	0	0	42	1582.53
	2	2	C[+57]	1202.2	3	1094.91	67.53	same	616.81	same	0	0	4	0	0	0	1704.63	1052.42	2	0	5	0	0	0		1702.58
TKQHGGFSLAVVSLNITSLGLR	2	2	C[+57]	1202.53	3	1077.61	66.63	same	619.22	same	0	0	4	0	0	0	1704.63	1052.65	2	0	5	0	0	0		1702.58
	2	2	C[+57]	1157.51	3	1082.28	68.58	same	629.01	same	0	1	1	1	0	1	1571.53	1052.59	2	0	4	0	0	0	28	1568.53
	2	2		1082.52	4	663.12	19.79	same	629.01	same	2	1	3	0	0	0	1955.72	858.88	3	0	1	3	1	0		1954.70
VCNIGIGIFK DLSLSINATNIKHFKNCT S ISGDLHILPVAFR	2	2	C[+57]*2	1385.82	5	654.03	27.74	NCT S ISGDLHILPVAFR	369.68	NCT S ISGDLHILPVAFR	0	2	2	0	2	2	2282.74	1052.63	2	0	4	0	0	0	3483.02	5023.10

n = number of redundancies

same : the peptide identified by Y1-based search is the same as the peptide backbone identified by direct Byonic search

*Glycan composition in terms of Fuc (F), Hex (H), HexNAc (N), NeuAc (Ac), NeuGc (Gc), and Na is given without counting the trimannosyl core (H3N2)

#Total mass refers to the entire glycan moiety plus any extra modification (Mod)

Figure S2

Copied from

“Novel LC-MS2 Product Dependent Parallel Data Acquisition Function and Data Analysis Workflow for Sequencing and Identification of Intact Glycopeptides.” written by Sz-Wei Wu, Tsung-Hsien Pu, Rosa Viner, and Kay-Hooi Khoo.

Anal Chem. 2014 Jun 3;86(11):5478-86.
doi: 10.1021/ac500945m.

I believe this precursor was a heterodimer of Man₆ & Man₇:
2+ ions at m/z 1640 and 1721
See Table 1, Figures 1 and 2

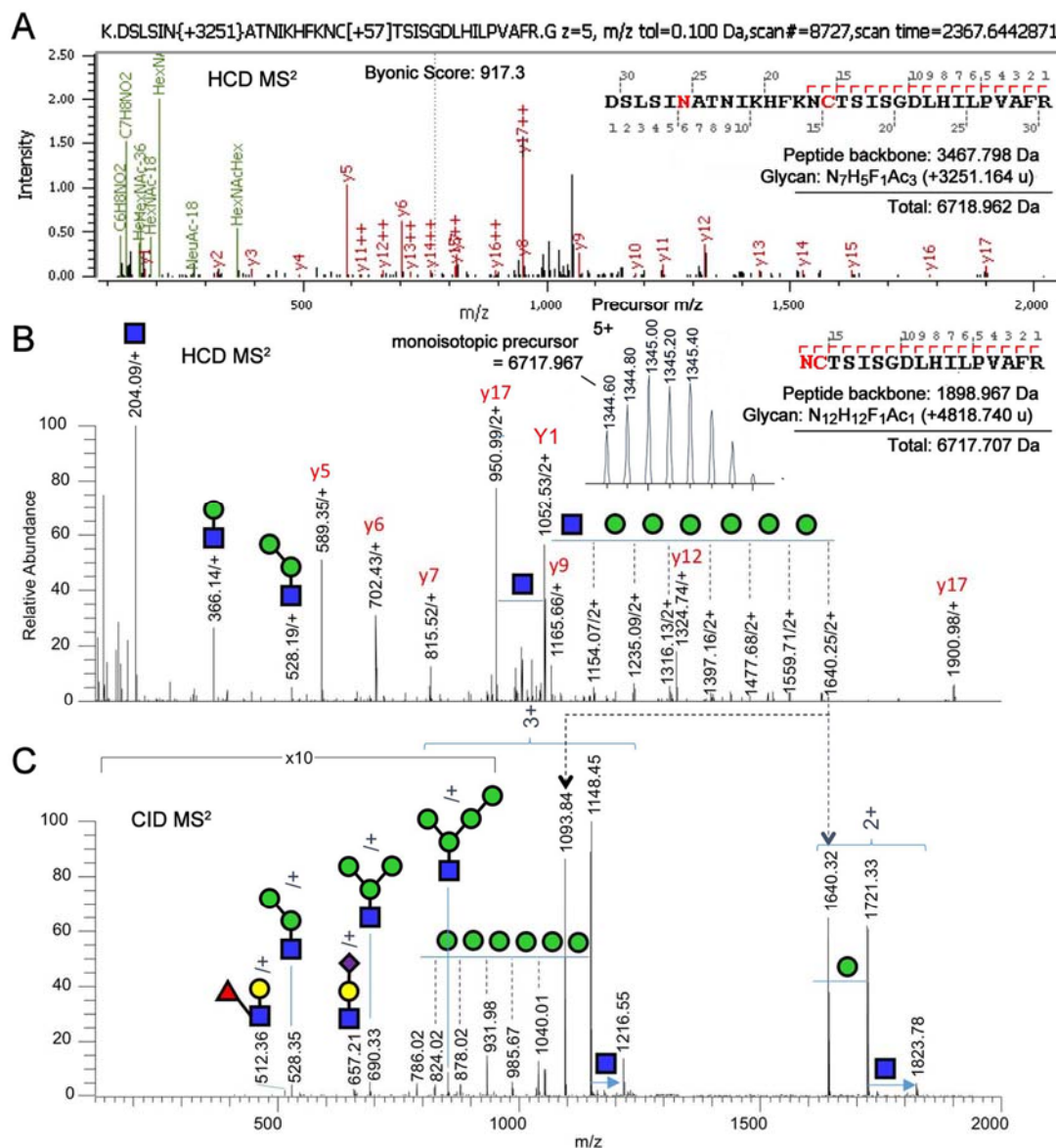


Figure 3. Exemplary false positive due to incorrect peptide backbone identification and glycan assignment. The extra masses assigned by Byonic to a stretch of amino acids extending from tryptic miscleavage sites (A) can be alternatively attributed to larger glycan moiety (B), consistent with the Y1 ion (m/z 1052.53) identified by a Y1-based search, and the sequential glycosyl losses evident in both manually assigned HCD (B) and CID (C) MS² spectra. The 5+ monoisotopic precursor at m/z 1344.6012 corresponds to a glycopeptide of 6717.967 Da (B, inset).

Figure S3

Calculated isotope distribution
for the Man₆-Man₇ heterodimer,
Shown in Figures 1 and 2, and in Figure S2

Isotope Number	m/z	Percent Total	Percent Maximum
0	1344.39502	2.33	12.79
1	1344.59563	7.96	43.67
2	1344.79620	14.40	79.00
3	1344.99673	18.23	100.00
4	1345.19724	18.05	99.01
5	1345.39773	14.83	81.38
6	1345.59821	10.50	57.62
7	1345.79867	6.57	36.03
8	1345.99913	3.69	20.26
9	1346.19959	1.89	10.38
10	1346.40003	0.89	4.90
11	1346.60048	0.39	2.15
12	1346.80093	0.16	0.88
13	1347.00138	0.06	0.34
14	1347.20183	0.02	0.12
15	1347.40232	0.01	0.04
16	1347.60283	0.00	0.01
17	1347.80333	0.00	0.00
18	1348.00344	0.00	0.00

C276 H450 N52 O135 S2 (+5H⁺)

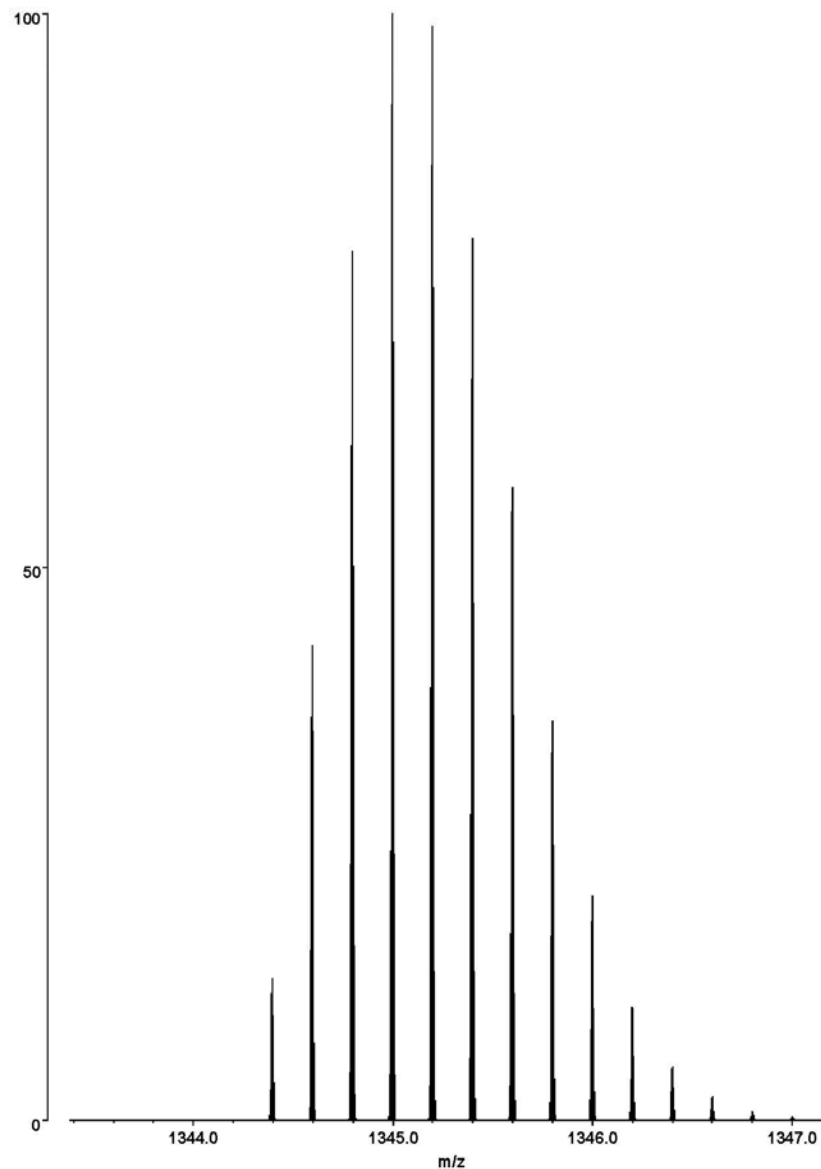


Figure S4

Copied from

“Novel LC-MS2 Product Dependent Parallel Data Acquisition Function and Data Analysis Workflow for Sequencing and Identification of Intact Glycopeptides.” written by Sz-Wei Wu, Tsung-Hsien Pu, Rosa Viner, and Kay-Hooi Khoo.

Anal Chem. 2014 Jun 3;86(11):5478-86.
doi: 10.1021/ac500945m.

I believe this precursor was a mixture of different heterodimers:
Man₆ & Man₇Man(42):
3+ ions at m/z 1040 and 1216;
and Man₇ & Man₆Man(42): 3+ ions at m/z 1148 and 1161
See Table 1

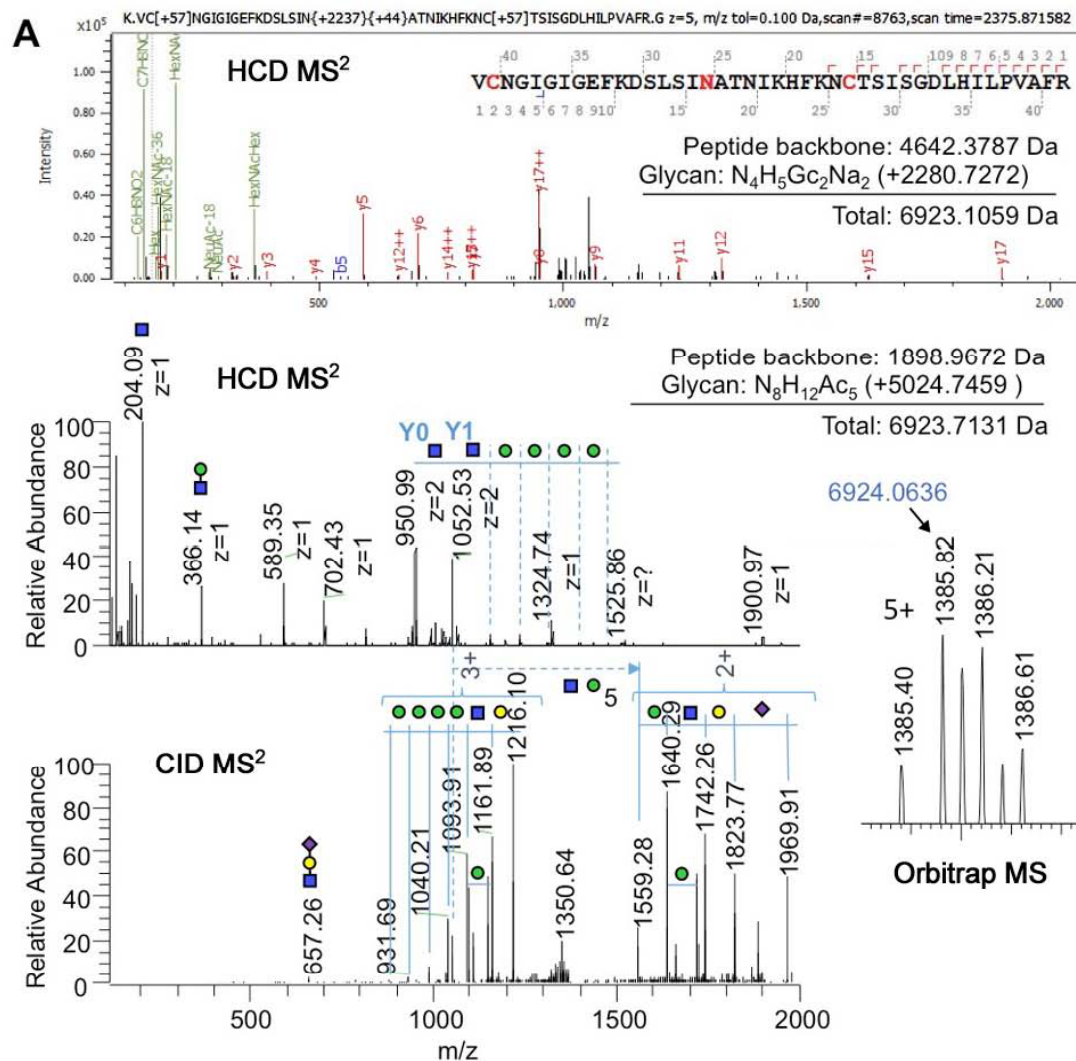


Figure S1. More examples of false positives due to incorrect peptide backbone identification and glycan assignment. In each of the 3 cases shown here (A-C), the extra masses assigned by Byonic to a stretch of amino acids extending from tryptic mis-cleavage sites can be alternatively attributed to larger glycan moiety, consistent with the Y1 ion (m/z 1052.53) most readily identified in the HCD MS² data (upper panel), and the sequential glycosyl losses evident in manually assigned CID MS² spectra (lower panel). The zoomed in mass region of the precursors shows that the most likely monoisotopic precursor within the isotopic cluster cannot always be determined unambiguously, particularly for signals of low intensity corresponding to high molecular weight glycopeptides.

Figure S5

Zoom in on the upper mass region of the previous CID spectrum

Erik_sEGFR_HCDpdETDCID_130912194806 #8766 RT: 39.61 AV: 1 NL: 5.36E3
T: ITMS + c NSI d Full ms2 1385.82 @cid30.00 [376.00-2000.00]

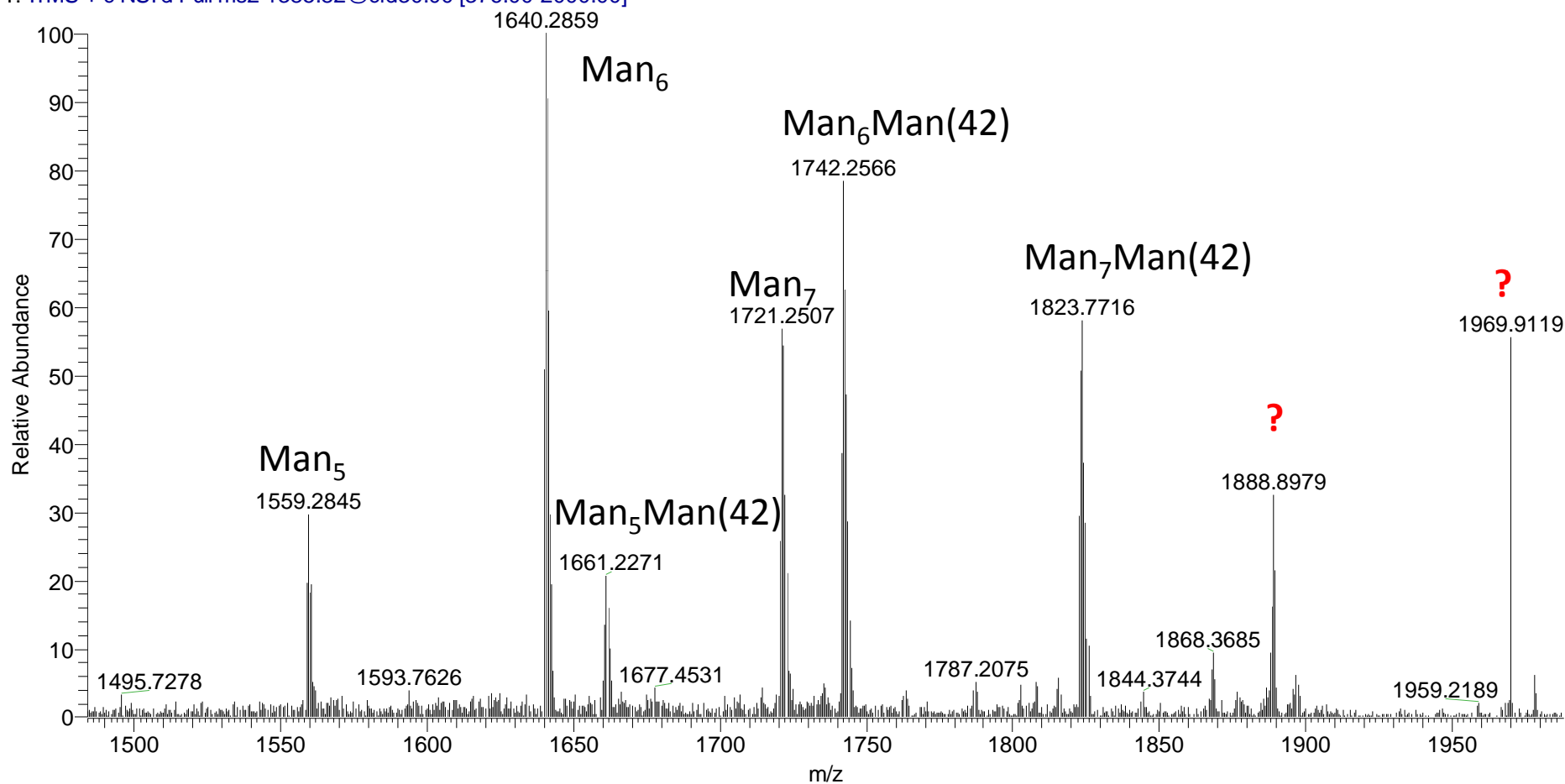
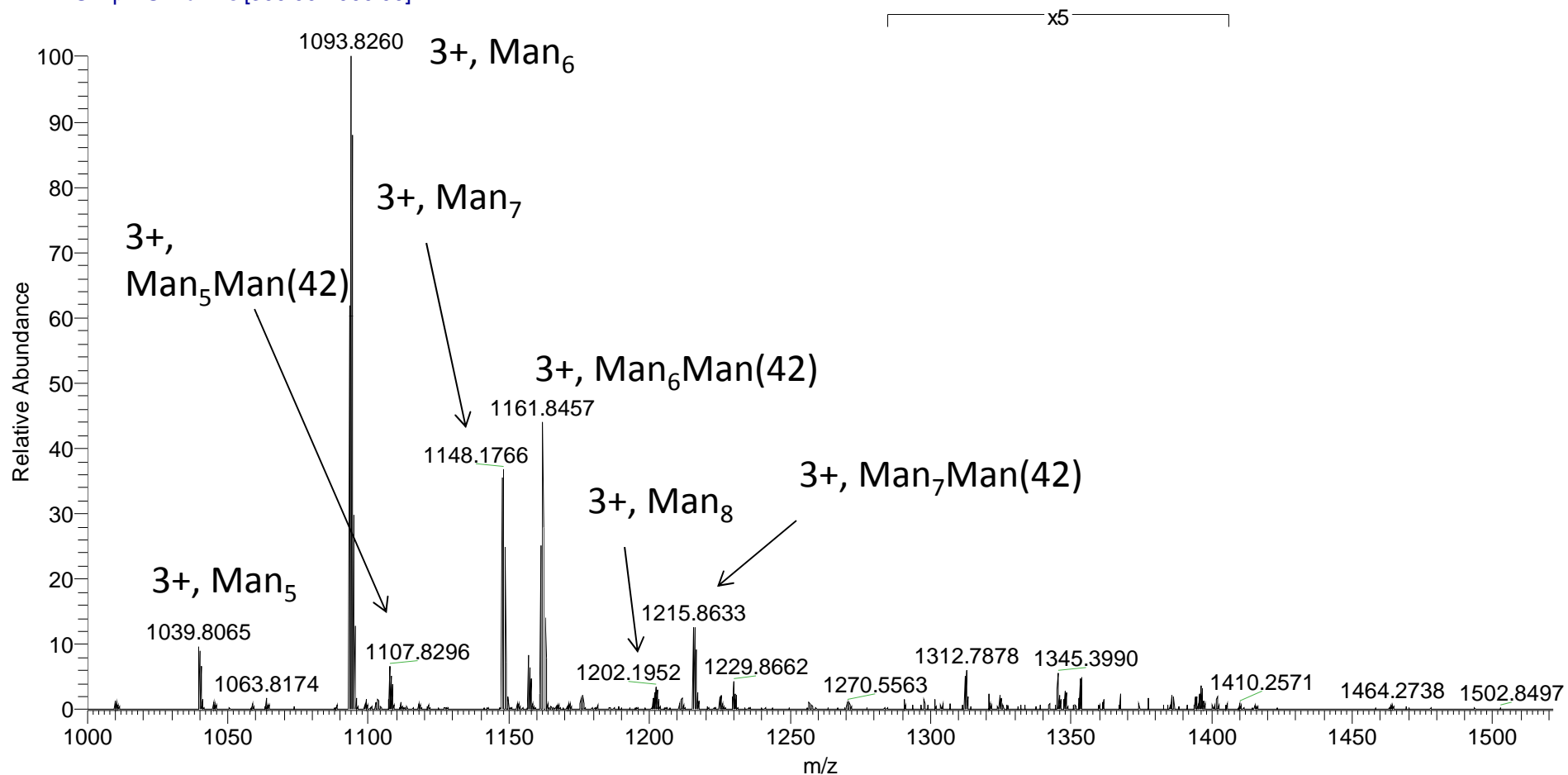


Figure S6

Erik_sEGFR_HCDpdETDCID_130912194806 #8752 RT: 39.55 AV: 1 NL: 1.00E8
T: FTMS + p NSI Full ms [300.00-2000.00]

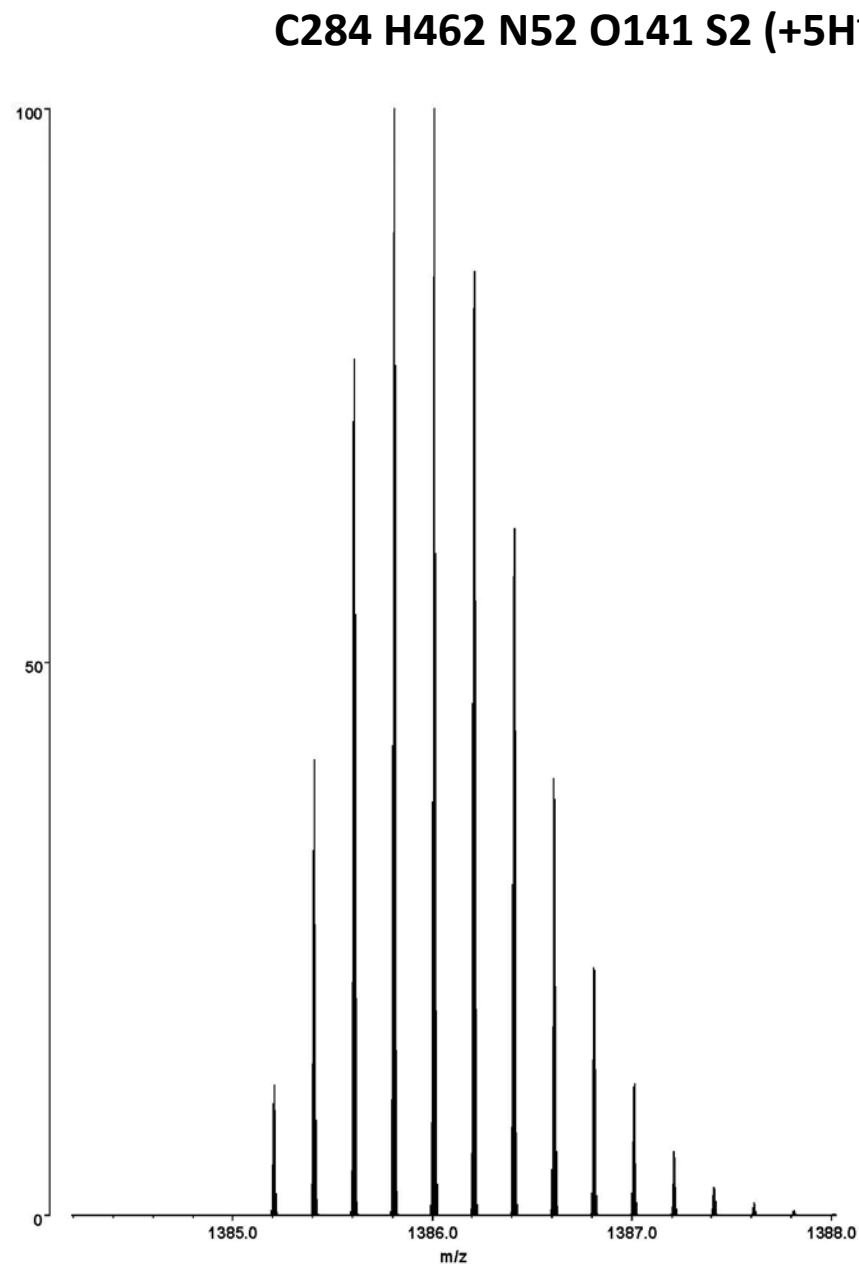


MS survey from which m/z 1385 was selected for MS/MS analysis – all major components were detected in the CID spectrum doubly and triply charged

Figure S7

Calculated* isotope distribution for
the Man6-Man7Man(42)
& Man7-Man6Man(42) heterodimers

Isotope Number	m/z	Percent Total	Percent Maximum
0	1385.20770	2.10	11.66
1	1385.40831	7.36	40.91
2	1385.60888	13.66	75.93
3	1385.80942	17.74	98.56
4	1386.00993	18.00	100.00
5	1386.21043	15.15	84.20
6	1386.41091	10.99	61.05
7	1386.61138	7.03	39.08
8	1386.81184	4.05	22.49
9	1387.01230	2.12	11.80
10	1387.21275	1.03	5.70
11	1387.41321	0.46	2.55
12	1387.61365	0.19	1.07
13	1387.81411	0.08	0.42
14	1388.01457	0.03	0.15
15	1388.21505	0.01	0.05
16	1388.41557	0.00	0.02
17	1388.61617	0.00	0.00
18	1388.81654	0.00	0.00



*based on the monoisotopic masses and CID data I considered the +42 Da modification as acetylation of Man residues

Figure S8

Copied from

“Novel LC-MS2 Product Dependent Parallel Data Acquisition Function and Data Analysis Workflow for Sequencing and Identification of Intact Glycopeptides.” written by Sz-Wei Wu, Tsung-Hsien Pu, Rosa Viner, and Kay-Hooi Khoo.

Anal Chem. 2014 Jun 3;86(11):5478-86.
doi: 10.1021/ac500945m.

I believe this precursor was a heterodimer:
Man₅ & Man₆Man(42):
2+ ions at m/z 1559 and 1742

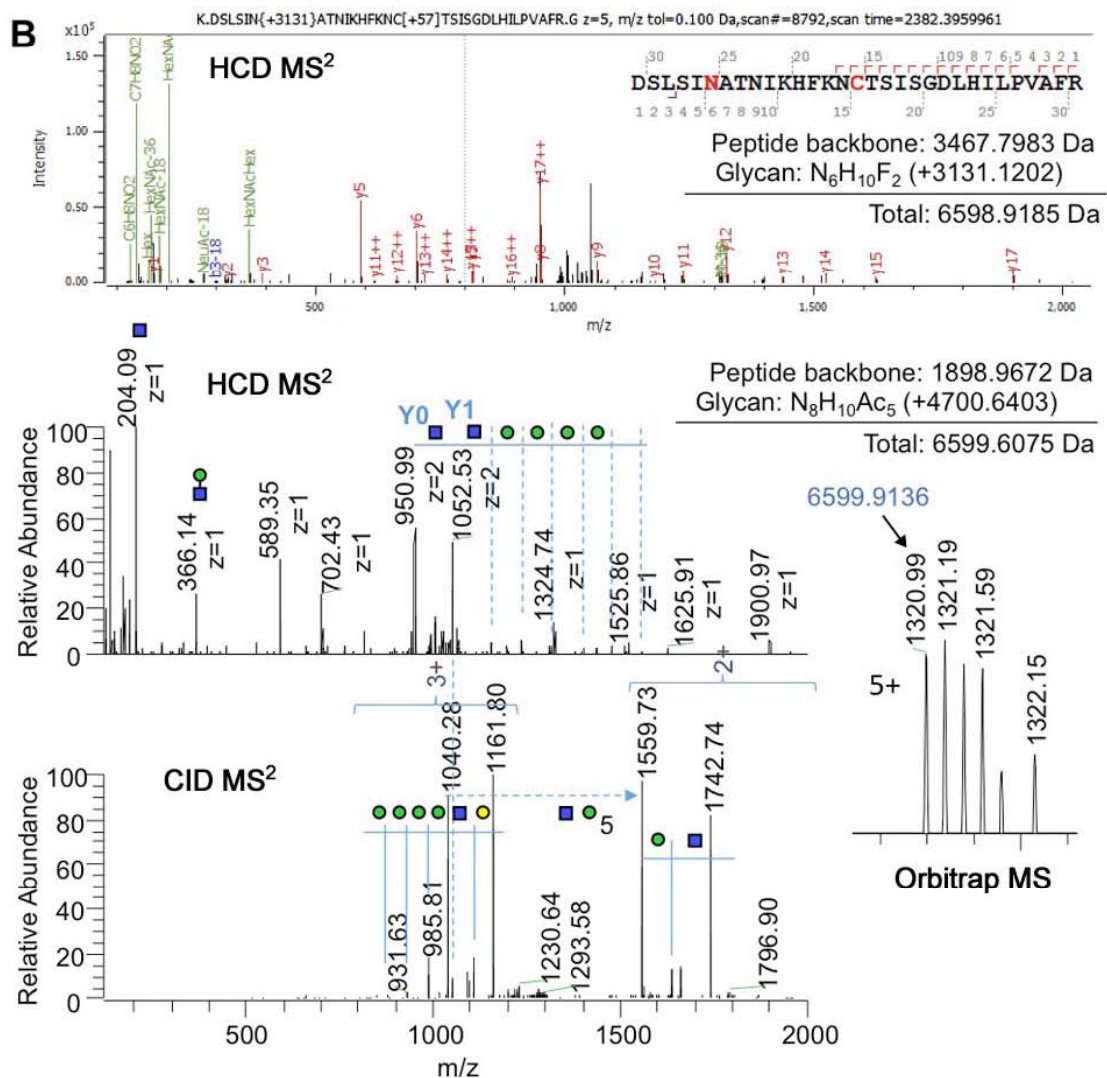
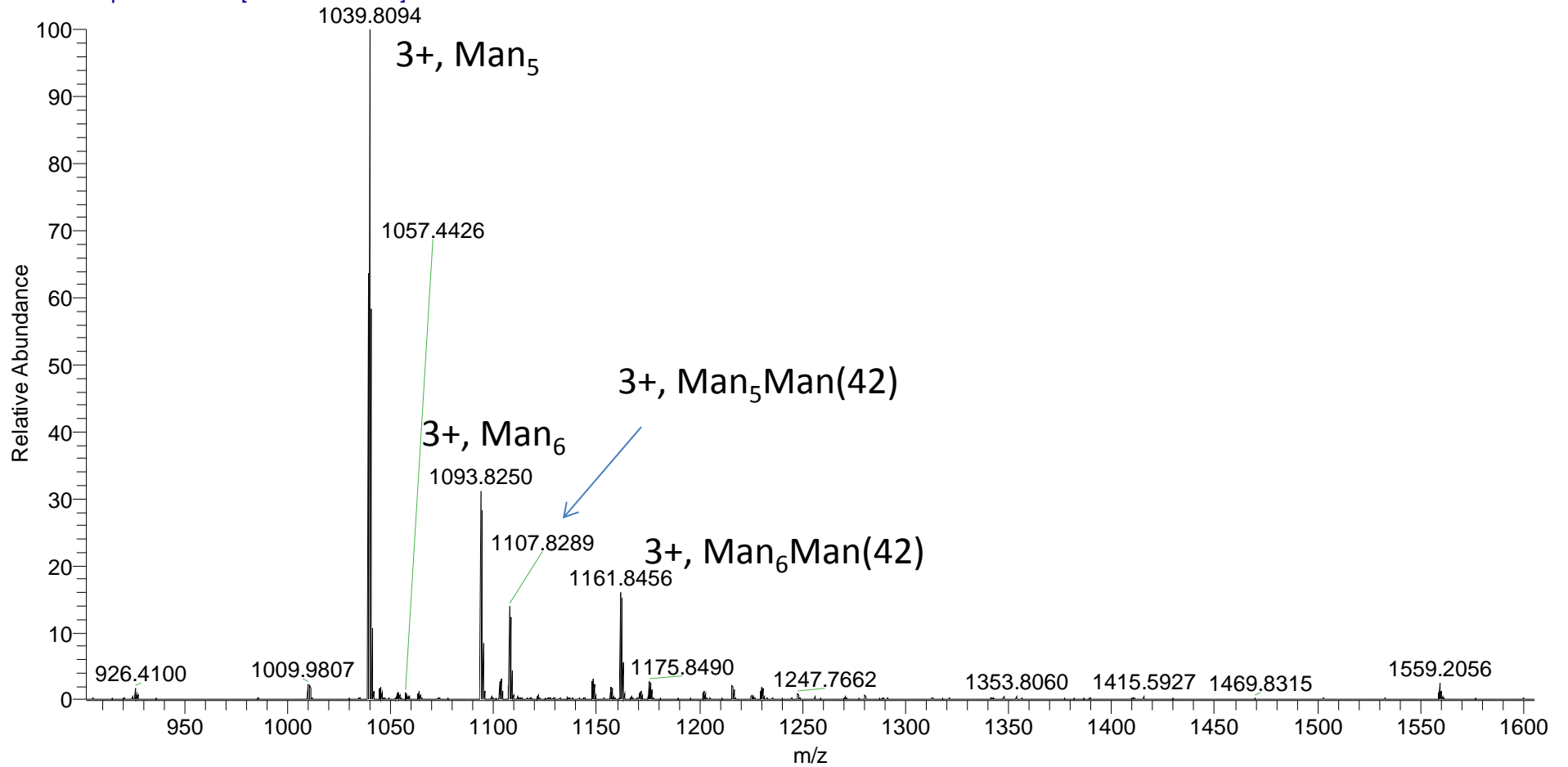


Figure S1B. More examples of false positives due to incorrect peptide backbone identification and glycan assignment (continued from previous page).

Figure S9

Erik_sEGFR_HCDpdETDCID_130912194806 #8788 RT: 39.69 AV: 1 NL: 1.69E8
T: FTMS + p NSI Full ms [300.00-2000.00]

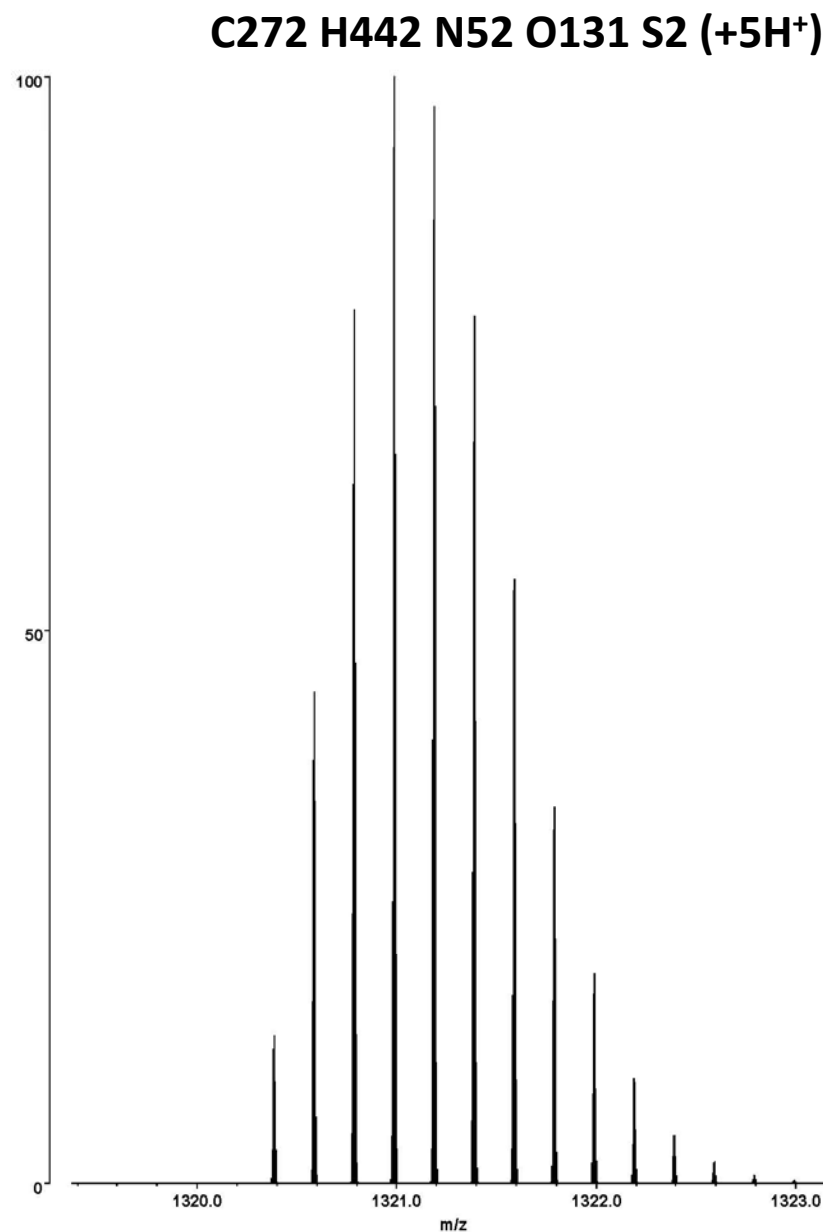


MS survey from which m/z 1320 was selected for MS/MS

Figure S10

Calculated* isotope distribution for the Man₅ & Man₆Man(42) dimer

Isotope Number	m/z	Percent Total	Percent Maximum
0	1320.38657	2.46	13.33
1	1320.58718	8.30	44.88
2	1320.78774	14.80	80.08
3	1320.98827	18.49	100.00
4	1321.18878	18.06	97.69
5	1321.38927	14.65	79.23
6	1321.58974	10.23	55.36
7	1321.79021	6.32	34.16
8	1321.99066	3.50	18.96
9	1322.19111	1.77	9.59
10	1322.39156	0.83	4.47
11	1322.59200	0.36	1.93
12	1322.79245	0.14	0.78
13	1322.99290	0.05	0.29
14	1323.19335	0.02	0.10
15	1323.39384	0.01	0.03
16	1323.59436	0.00	0.01
17	1323.79500	0.00	0.00
18	1323.99499	0.00	0.00



*based on the monoisotopic masses and CID data I considered the +42 Da modification as acetylation of Man residues

Figure S11

Copied from

“Novel LC-MS2 Product Dependent Parallel Data Acquisition Function and Data Analysis Workflow for Sequencing and Identification of Intact Glycopeptides.” written by Sz-Wei Wu, Tsung-Hsien Pu, Rosa Viner, and Kay-Hooi Khoo.

Anal Chem. 2014 Jun 3;86(11):5478-86. doi: 10.1021/ac500945m.

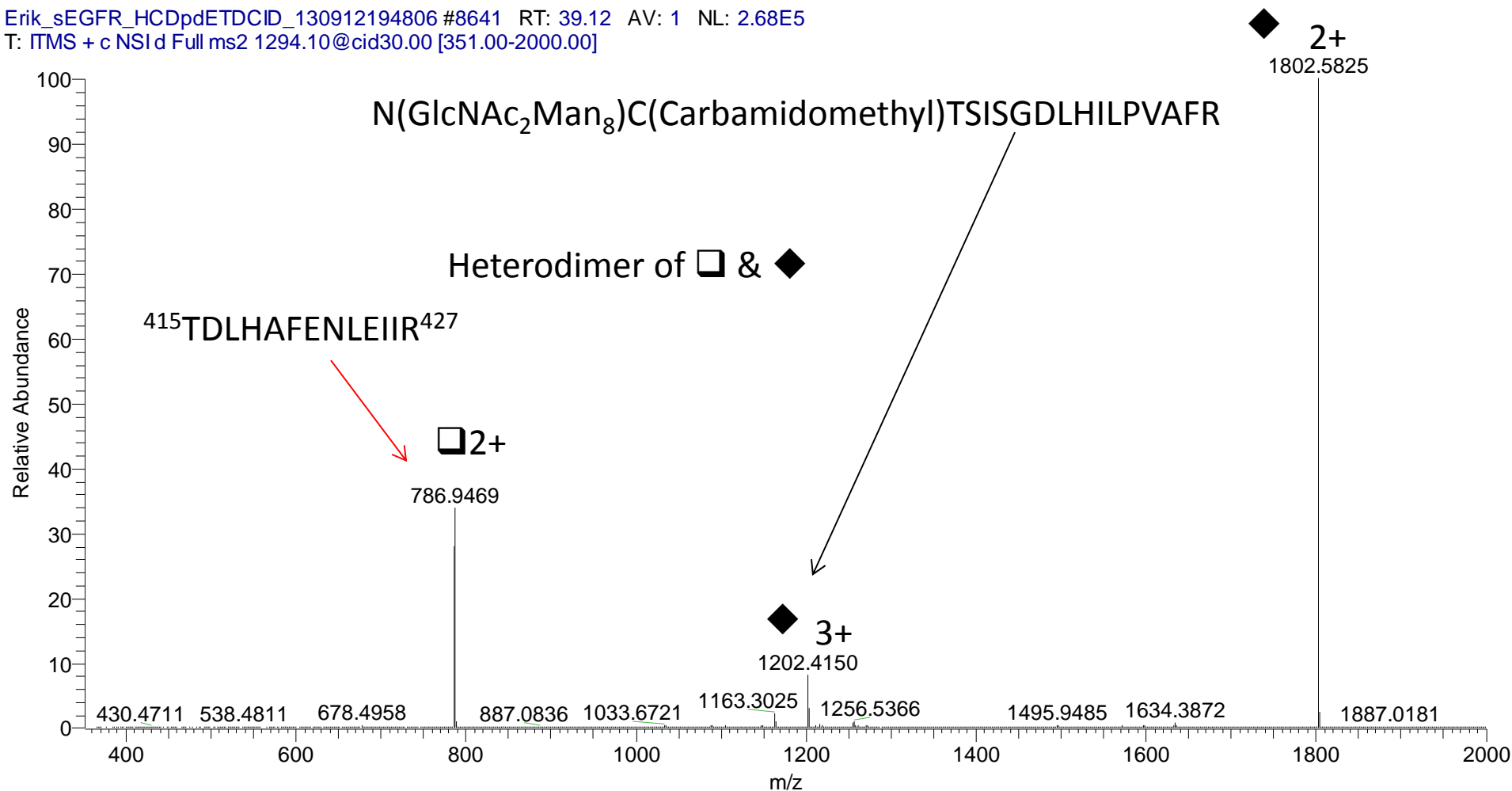
Table S1-2. Subset of Incorrect Glycopeptide Identification by Direct Database Search Without Supporting Positive ID by Y1-based Search

Byonic search results							Assigned glycan composition (-H3N2)*						Correct Y1		Glycan Composition (-H3N2)*							
Sequence	PSMs	n	Modification(s)	m/z	z	Score	F	H	N	Ac	Gc	Na	Total mass (Da)#	Y1 (m/z)	Z	F	H	N	Ac	Gc	Extra Mod	Total mass(Da)#
DIVSSDFLSNMSMDFQNHLSGCQK	11	4	C[+57], M[+16]	1111.21	4	581.76	0	1	3	0	0	0	1663.61	1466.62	2	2	2	1	0	0		1711.62
		1	C[+57]	1147.73	4	353.77	0	2	3	0	0	0	1825.66	1458.7	2	2	3	1	0	0		1873.67
		2	C[+57], M[+16]	1151.72	4	481.49	0	2	3	0	0	0	1825.66	1466.65	2	2	3	1	0	0		1873.67
		2	C[+57], M[+16]*2	1161.98	4	505.07	1	0	4	0	0	0	1850.69	978.25	3	2	2	2	0	0		1914.70
		2	C[+57]	1188.24	4	438.07	0	3	3	0	0	0	1987.71	978.27	3	3	3	1	0	0		2019.73
DLSINATNIKHFKNCTISGDLHILPVAFR	7	2	C[+57]	1294.1	4	428.93	0	5	0	0	0	0	1702.58	1052.03	2	0	5	0	0	0	1569.8328	3272.41
		2	C[+57]	1383.81	5	670.11	0	4	5	2	1	0	3445.21	1060.03	2	0	5	0	0	0	3294.9784	4996.92
		2	C[+57]	1448.42	5	659.16	0	5	4	4	0	4	3771.24	1060.03	2	0	6	0	0	0	3456.0684	5320.07
		1	C[+57]	1469.91	4	373.59	0	2	3	2	0	0	2407.85	1052.03	2	0	2	0	0	0	2759.0256	3974.65
		2	C[+57]*4, M[+16]	1135.47	4	468.36	1	0	1	0	0	0	1257.45	1168.15	3	0	2	0	0	0	+Na	1239.42
EFVNSECIQCHPECLPQAMNITCTGR	2	2	C[+57]*4, M[+16]	1135.47	4	468.36	1	0	1	0	0	0	1257.45	1168.15	3	0	2	0	0	0	+Na	1239.42
HFKNCTISGDLHILPVAFR	16	2	C[+57]	893.41	4	412.69	0	1	1	0	0	0	1257.45	848.47	3	0	2	0	0	0	42	1258.42

Figure S12

CID of m/z 1294.1 listed in Table S1-2 – explanation given with a 1569.8324 Da ‘defect’

Erik_sEGFR_HCDpdETDCID_130912194806 #8641 RT: 39.12 AV: 1 NL: 2.68E5
T: ITMS + c NSI d Full ms2 1294.10@cid30.00 [351.00-2000.00]



Sequence positions are given according to P00533 of SwissProt data base

Figure S13

MS survey from which the m/z 1294 ion was selected for MS/MS; precursor ion in the insert

Erik_sEGFR_HCDpdETDCID_130912194806 #8632 RT: 39.07 AV: 1 NL: 1.53E8
T: FTMS + p NSI Full ms [300.00-2000.00]

Erik_sEGFR_HCDpdETDCID_130912194806 #8632 RT: 39.07 AV: 1 NL: 5.89E5
T: FTMS + p NSI Full ms [300.00-2000.00]

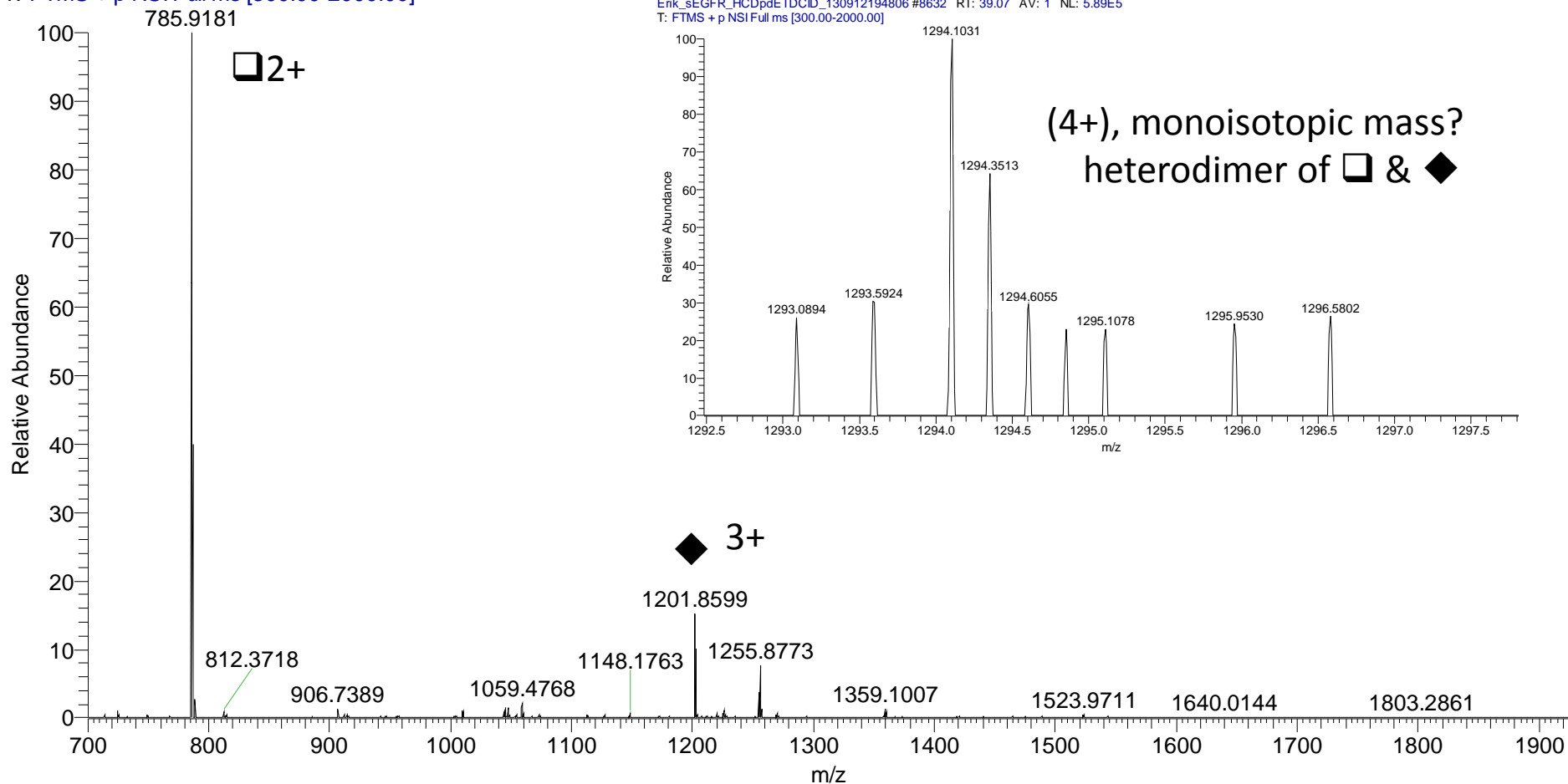


Figure S14

Calculated isotope distribution for the peptide-glycopeptide (Man₈) heterodimer

Isotope Number	m/z	Percent Total	Percent Maximum
0	1293.84820	5.40	25.88
1	1294.09896	14.54	69.65
2	1294.34966	20.80	99.65
3	1294.60033	20.87	100.00
4	1294.85097	16.41	78.62
5	1295.10159	10.72	51.38
6	1295.35220	6.04	28.96
7	1295.60280	3.01	14.42
8	1295.85340	1.35	6.46
9	1296.10398	0.55	2.64
10	1296.35457	0.21	0.99
11	1296.60515	0.07	0.35
12	1296.85575	0.02	0.11
13	1297.10637	0.01	0.03
14	1297.35702	0.00	0.01
15	1297.60764	0.00	0.00
16	1297.85977	0.00	0.00

C217 H351 N45 O97 S1 (+4H⁺)

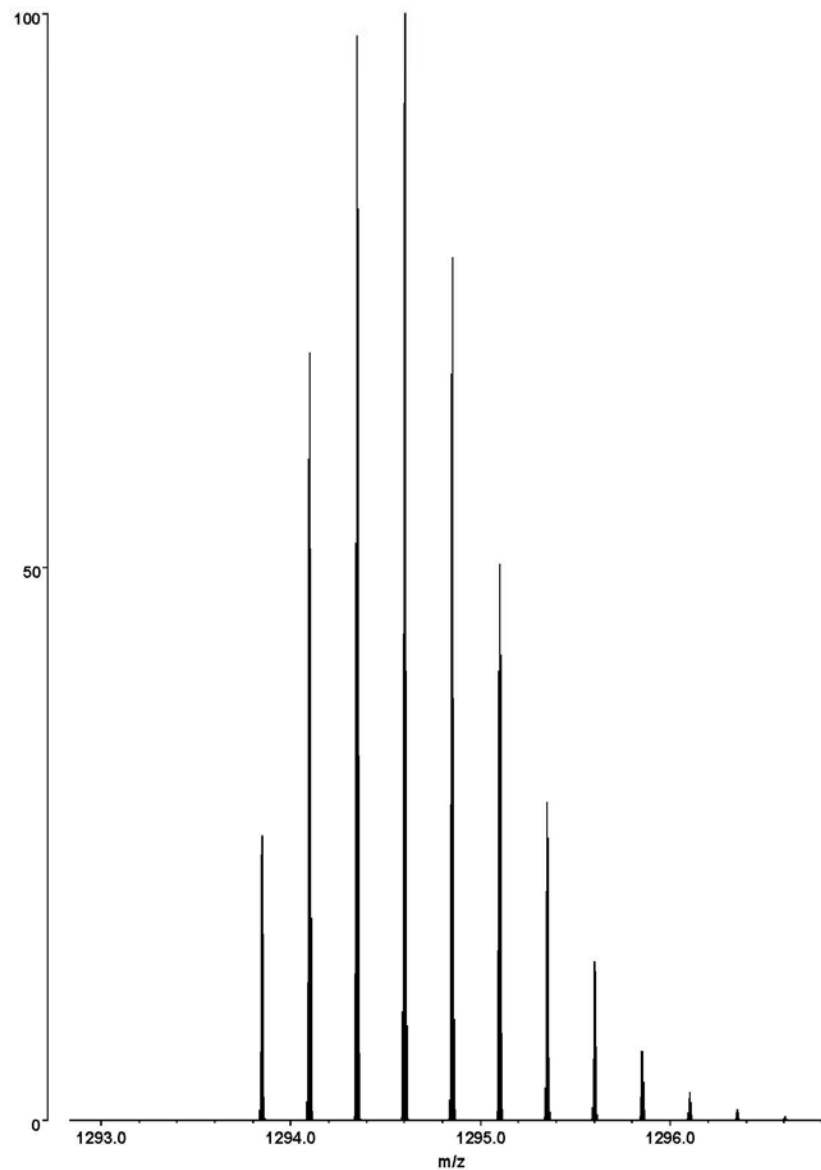


Figure S15

HCD spectrum of 1294.1(4+)



415**TDLHAFENLEIIR**427 □

The fragments clearly indicate the presence of both peptides

Erik_sEGFR_HCDpdETDCID_130912194806 #8640 RT: 39.12 AV: 1 NL: 3.13E5
T: FTMS + p NSI d Full ms2 1294.10@hcd35.00 [120.00-2000.00]

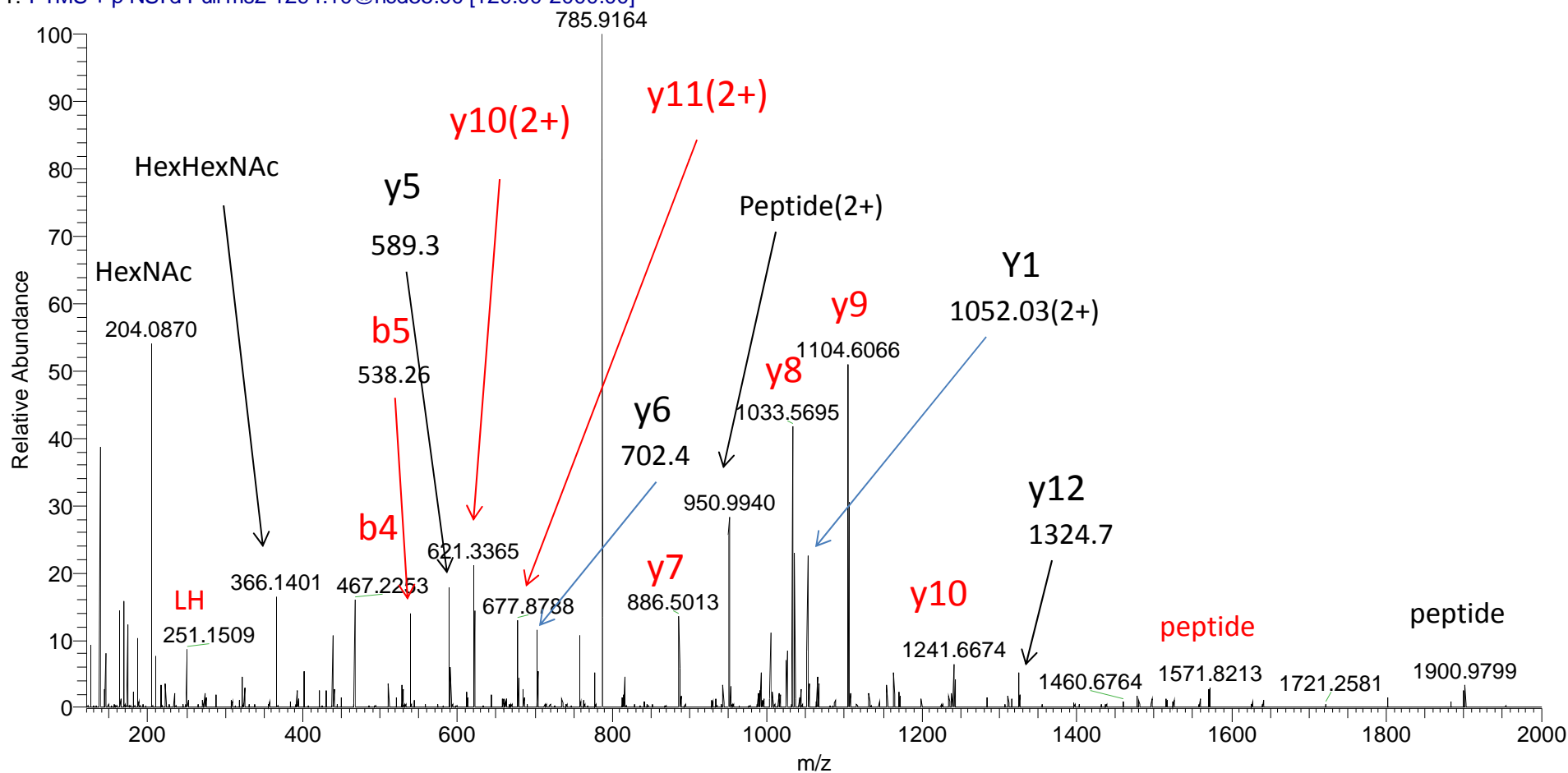
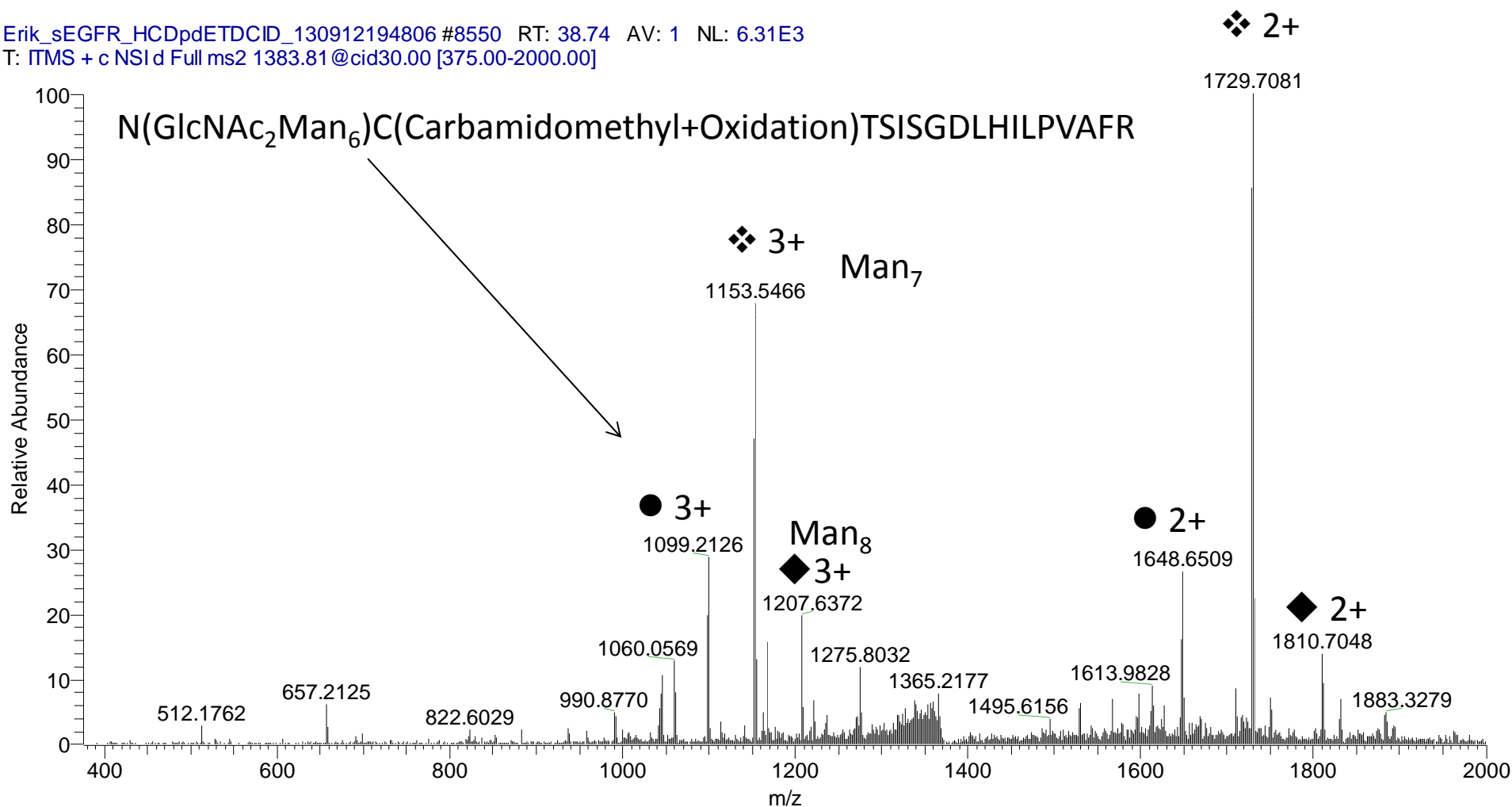


Figure S16

CID of m/z 1383.81 listed in Table S1-2 – explanation given with a 3294.9784 Da mass 'defect'

Erik_sEGFR_HCDpdETDCID_130912194806 #8550 RT: 38.74 AV: 1 NL: 6.31E3
T: ITMS + c NSI d Full ms2 1383.81 @cid30.00 [375.00-2000.00]



Heterodimer of ● & ◆ AND homodimer of ❖

Figure S17

MS survey from which the m/z 1383 ion was selected for MS/MS; precursor ion in the insert

Erik_sEGFR_HCDpdETDCID_130912194806 #8544 RT: 38.71 AV: 1 NL: 5.87E7
T: FTMS + p NSI Full ms [300.00-2000.00]

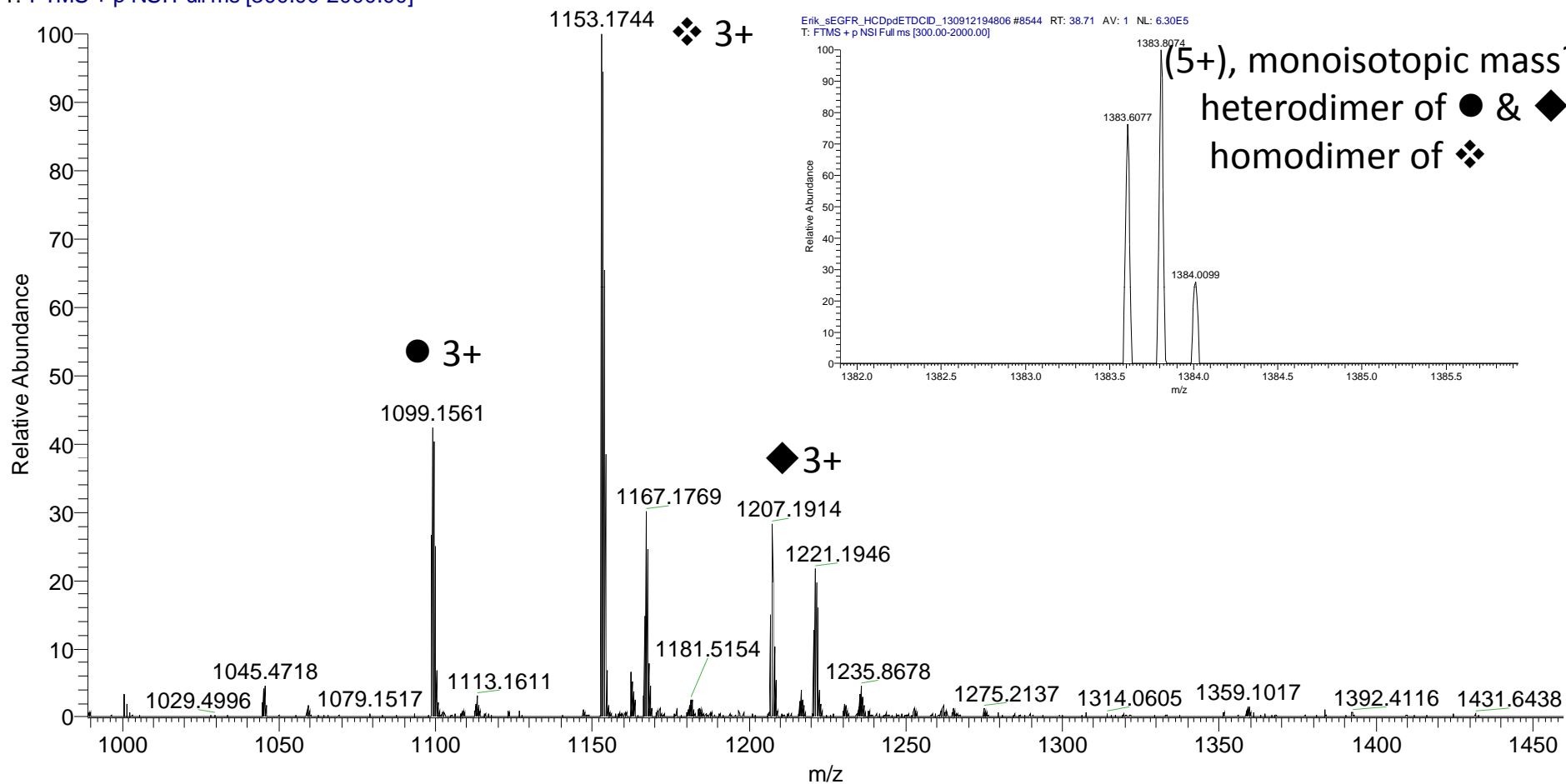
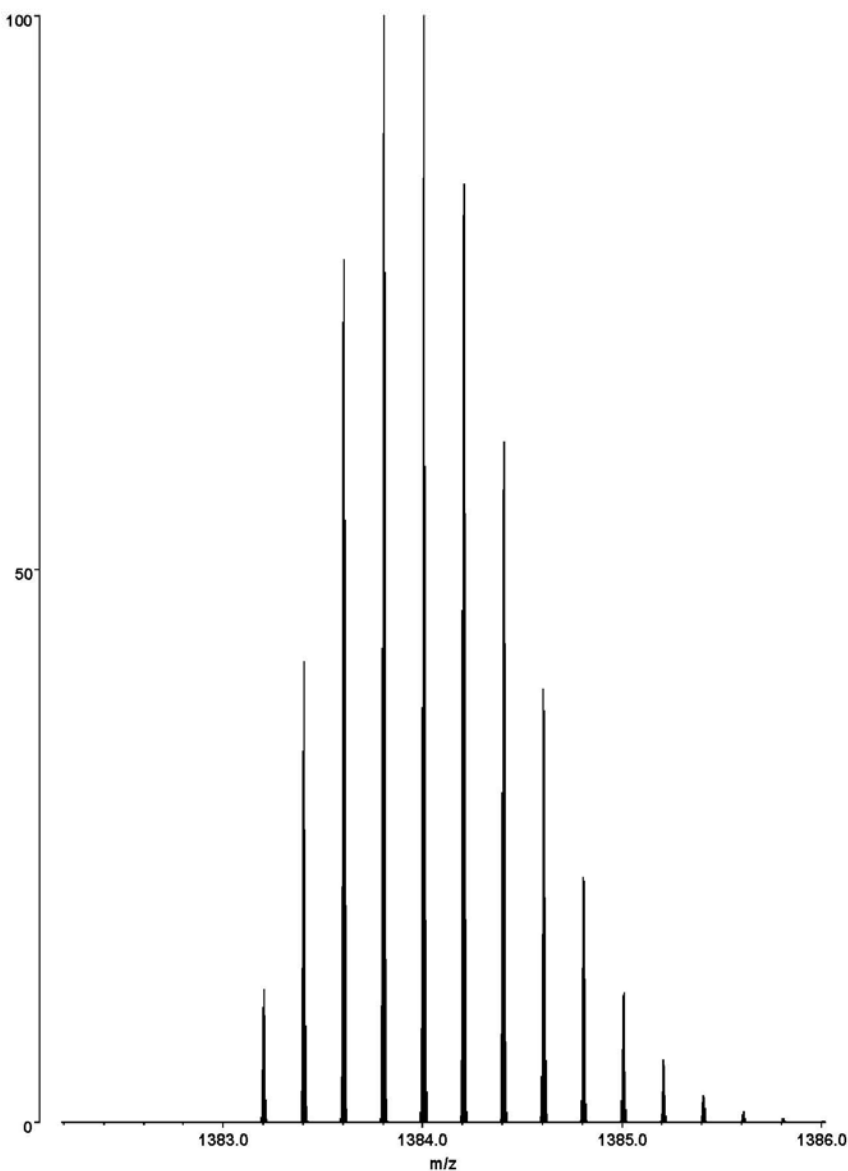


Figure S18

C282 H460 N52 O142 S2 (+5H⁺)

Calculated* isotope distribution for the Man₇ homodimer or Man₆-Man₈ heterodimer

Isotope Number	m/z	Percent Total	Percent Maximum
0	1383.20355	2.14	11.90
1	1383.40416	7.46	41.48
2	1383.60473	13.78	76.59
3	1383.80527	17.80	98.96
4	1384.00578	17.99	100.00
5	1384.20627	15.09	83.89
6	1384.40675	10.91	60.63
7	1384.60723	6.96	38.70
8	1384.80769	4.00	22.21
9	1385.00815	2.09	11.62
10	1385.20860	1.01	5.60
11	1385.40905	0.45	2.51
12	1385.60949	0.19	1.05
13	1385.80995	0.07	0.41
14	1386.01041	0.03	0.15
15	1386.21088	0.01	0.05
16	1386.41141	0.00	0.01
17	1386.61195	0.00	0.00
18	1386.81252	0.00	0.00



*the peptide is oxidized in these glycoforms

Figure S19

CID of m/z 1448.42 listed in Table S1-2 – explanation given with a 3456.0684 Da mass 'defect'

Erik_sEGFR_HCDpdETDCID_130912194806 #8529 RT: 38.65 AV: 1 NL: 6.19E3

F: ITMS + c NSI d Full ms2 1448.42@cid30.00 [393.00-2000.00]

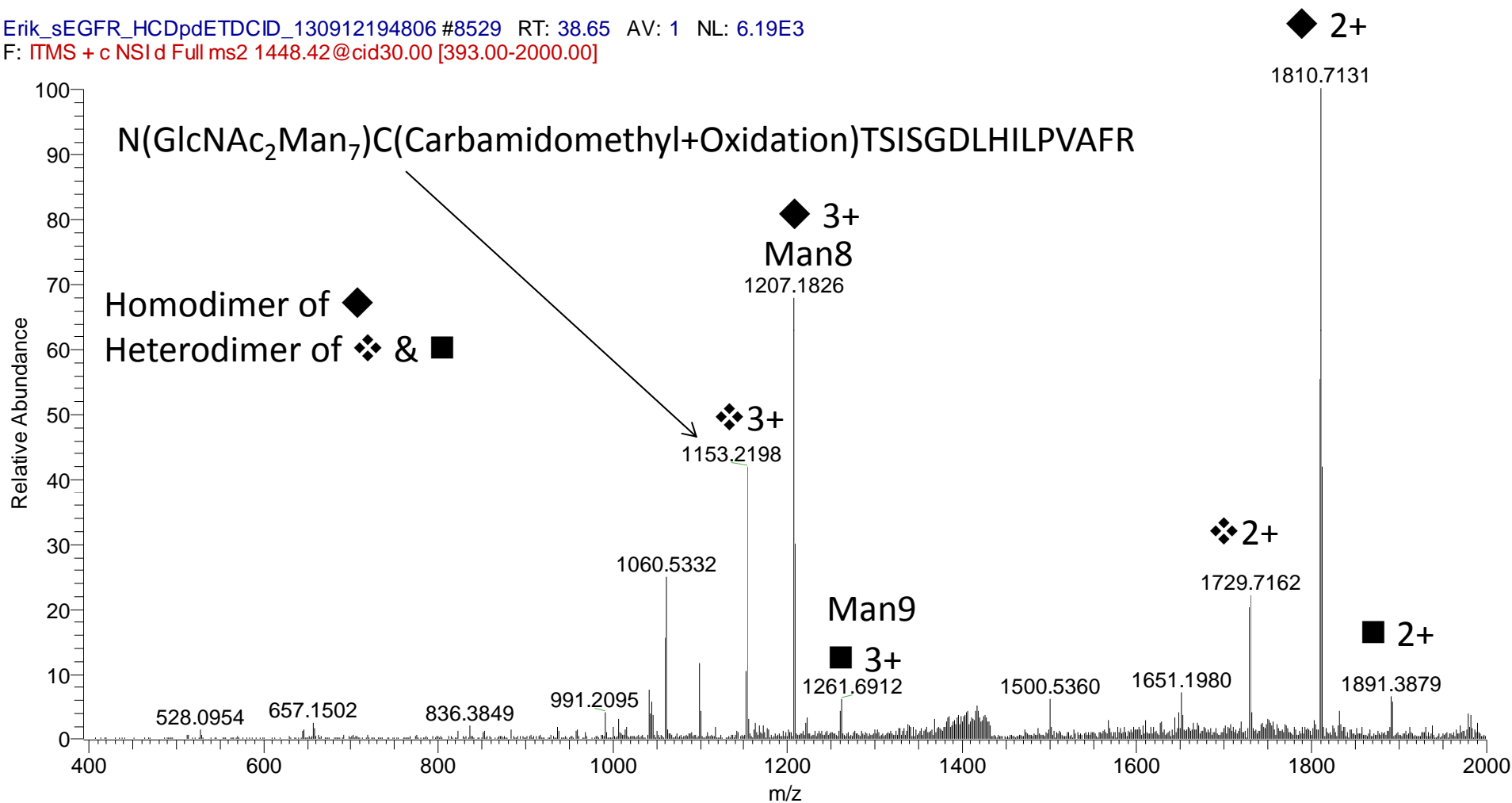


Figure S20

MS survey from which the m/z 4448 ion was selected for MS/MS; precursor ion in the insert

Erik_sEGFR_HCDpdETDCID_130912194806 #8520 RT: 38.60 AV: 1 NL: 3.95E7
T: FTMS + p NSI Full ms [300.00-2000.00]

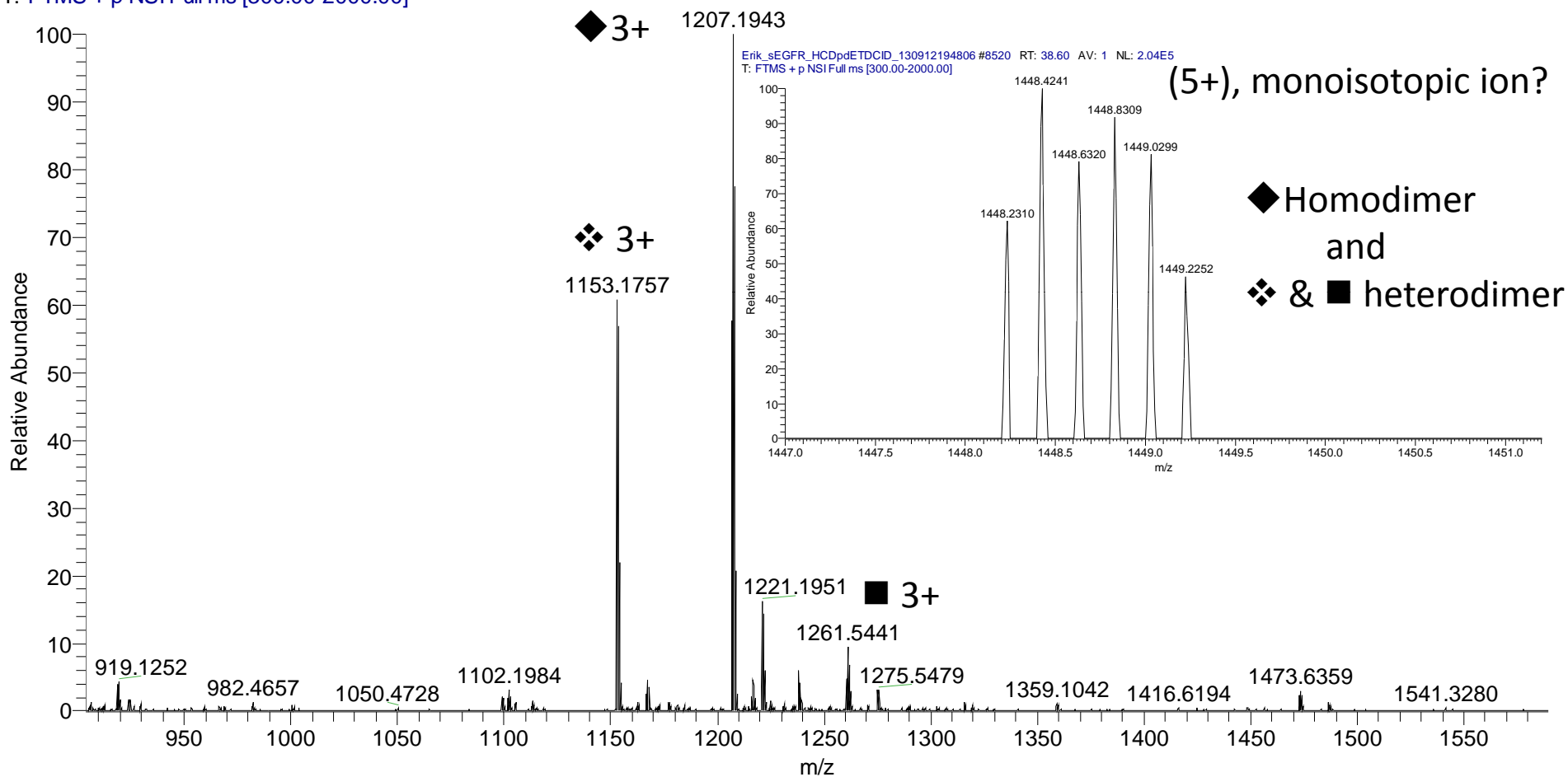
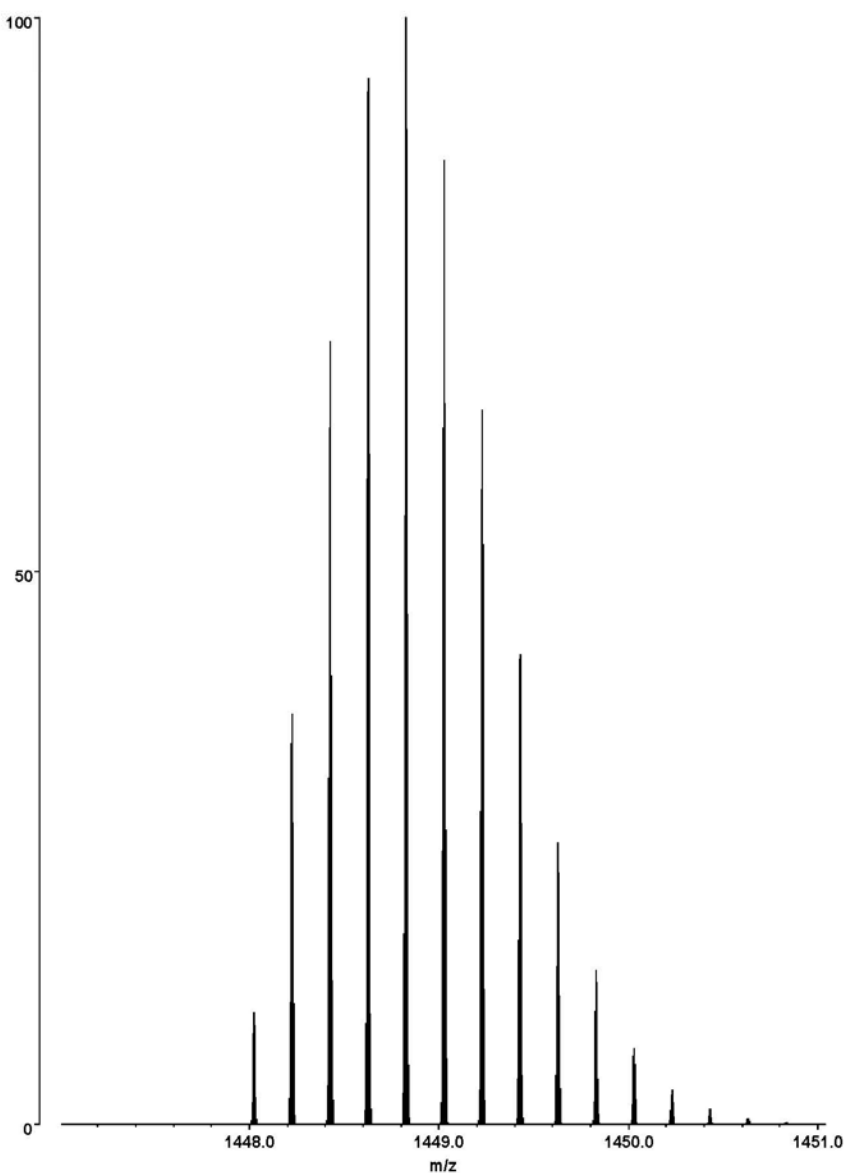


Figure S21

Calculated* isotope distribution for the Man₈ homodimer and Man₇-Man₉ heterodimer

Isotope Number	m/z	Percent Total	Percent Maximum
0	1448.02468	1.82	10.23
1	1448.22530	6.61	37.09
2	1448.42587	12.69	71.15
3	1448.62641	17.01	95.41
4	1448.82693	17.83	100.00
5	1449.02743	15.50	86.96
6	1449.22792	11.61	65.12
7	1449.42840	7.68	43.05
8	1449.62887	4.56	25.59
9	1449.82933	2.47	13.86
10	1450.02979	1.23	6.92
11	1450.23024	0.57	3.20
12	1450.43070	0.25	1.39
13	1450.63115	0.10	0.56
14	1450.83161	0.04	0.21
15	1451.03209	0.01	0.08
16	1451.23261	0.00	0.02
17	1451.43312	0.00	0.01
18	1451.63379	0.00	0.00

C294 H480 N52 O152 S2 (+5H⁺)



*the peptide is oxidized in these glycoforms