Supporting Information

Figure S1. Comparison between the extracted ion chromatogram at m/z 204.086 and the total ion chromatogram of the HCD-alone LC-MS/MS experiment on the enriched HEK293T sample.

The extracted ion chromatogram at m/z 204.086 (bottom panel) delineates the retention time profile of O-GlcNAc modified peptides eluted over the entire gradient. In comparison with the total ion chromatogram, the normalized peak height of the extracted ion chromatogram indicates the substoichiometric nature of O-GlcNAc modification and therefore suggests the importance of selective fragmentation on O-GlcNAc peptides.

Figure S2. Comparison between the total ion chromatogram and the respective extracted ion chromatogram at m/z 204.086, m/z 186.076, m/z 168.065, m/z 144.065, m/z 138.054, and m/z 126.054 of the HCD/ETD LC-MS/MS experiment on the enriched HEK293T sample.

In comparison to the total ion chromatogram, the respective extracted ion chromatograms at the m/z values of 6 signature ions of HexNAc illustrates the retention time profiles of O-GlcNAc modified peptides and further indicates the substoichiometry of O-GlcNAc modification. The difference in the normalized peak heights of each extracted ion chromatogram suggests the prevalence of respective signature ions.

Figure S3. Multiple-engine database search strategy for analysis on the enriched HEK293T sample.

Raw spectra were searched twice using three search engines: SEQUEST, Mascot and ProteinProspector. The first search was performed against the entire human database without allowing for HexNAc modification, and the output was filtered at 5% peptide-level FDR and exported as the limited protein database; the second search was performed against the exported database allowing for HexNAc modification, and the output from each search engine was filtered at 1% peptide-level FDR, manually verified and combined with each other to produce the final list of O-GlcNAc modified peptides.

Figure S4. HCD and ETD spectra of O-Mannose and O-GalNAc modified peptides.

(A) HCD spectrum of peptide Ac-IRt(Man)t(Man)t(GalNAc)SGVPR-NH_2 at low m/z range; (B) HCD spectrum of peptide Ac-PTTt(GalNAc)PLK-NH_2 at low m/z range; (C) HCD spectrum of peptide Ac-RIRTTt(Man)SGVPR-NH_2 at low m/z range; (D) ETD spectrum of peptide Ac-IRt(Man)t(Man)t(GalNAc)SGVPR-NH_2; (E) ETD spectrum of peptide Ac-PTTt(GalNAc)PLK-NH_2; (F) ETD spectrum of peptide Ac-RIRTTt(Man)SGVPR-NH_2.

Table S1. Fragments of HexNAc and Hexose oxonium ions.

Table S2. List of proteins identified in the enriched HEK293T sample by HCD/ETD approach.

186 proteins were identified at 1% pepFDR from the multi-antibody enriched HEK293T cell extract. In comparison with the recently published work from our group[11] that used the same sample, 67% (124/186) of the identified proteins were consistent with the result from previous experiment. Of the

33% (62/186) of the dataset that was only observed in our current study, 17 proteins were supported in the literature[6, 12-17] as being O-GlcNAc modified and 45 were identified as novel O-GlcNAc proteins.

Table S3. List of validated O-GlcNAc modified peptides identified in the enriched HEK293T sample by HCD/ETD approach.

83 O-GlcNAc sites on 172 glycopeptides from 13 proteins were validated. By comparing to literature[1-10], 13 of the 83 sites have been previously identified and 70 O-GlcNAc sites were newly discovered in our study; and 11 of the 13 characterized proteins were also observed but not site-mapped in the previously work from our group[11].

Table S4. List of validated O-GlcNAc modified peptides identified in the enriched HEK293T sample by HCD-alone approach.

7 O-GlcNAc sites were identified on 9 glycopeptides from 3 proteins after manual validation. 5 of the glycopeptides, corresponding to 2 O-GlcNAc sites, were also observed in the HCD/ETD analysis. Most validated peptides in the table present with low Xcorr scores due the low ratio of ion match.

Table S5. List of validated O-GlcNAc modified peptides identified in the enriched HEK293T sample by HCD/ETD approach using multiple-engine database searches.

165 O-GlcNAc sites on 178 glycopeptides from 40 proteins were identified using multiple search engines. By comparing to literature[1-10], 29 of the 165 sites have been previously assigned leaving 136 novel O-GlcNAc sites; and 22 of the 40 characterized proteins were also observed in the previously work from our group[11], whereas 18 of the 40 proteins are novel.

Supporting References

1. Wang, Z.; Udeshi, N. D.; Slawson, C.; Compton, P. D.; Sakabe, K.; Cheung, W. D.; Shabanowitz, J.; Hunt, D. F.; Hart, G. W., Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci Signal* **2010,** 3, (104), ra2.
2. Chalkley, R. J.; Thalhammer, A.; Schoepfer, R.; Burlingame, A. L., Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc Natl Acad Sci U S A* **2009,** 106, (22), 8894-9.
3. Zachara, N. E.; Molina, H.; Wong, K. Y.; Pandey, A.; Hart, G. W., The dynamic stress-induced "O-GlcNAc-ome" highlights functions for O-GlcNAc in regulating DNA damage/repair and other cellular pathways. *Amino Acids* **2011,** 40, (3), 793-808.
4. Zeidan, Q.; Wang, Z.; De Maio, A.; Hart, G. W., O-GlcNAc cycling enzymes associate with the translational machinery and modify core ribosomal proteins. *Mol Biol Cell* **2010,** 21, (12), 1922-36.
5. Zachara, N. E.; Cheung, W. D.; Hart, G. W., Nucleocytoplasmic glycosylation, O-GlcNAc: identification and site mapping. *Methods Mol Biol* **2004,** 284, 175-94.
6. Vosseller, K.; Trinidad, J. C.; Chalkley, R. J.; Specht, C. G.; Thalhammer, A.; Lynn, A. J.; Snedecor, J. O.; Guan, S.; Medzihradszky, K. F.; Maltby, D. A.; Schoepfer, R.; Burlingame, A. L., O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol Cell Proteomics* **2006,** 5, (5), 923-34.

7. Wells, L.; Vosseller, K.; Cole, R. N.; Cronshaw, J. M.; Matunis, M. J.; Hart, G. W., Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics* **2002,** 1, (10), 791-804.
8. Wang, Z.; Gucek, M.; Hart, G. W., Cross-talk between GlcNAcylation and phosphorylation: site-specific phosphorylation dynamics in response to globally elevated O-GlcNAc. *Proc Natl Acad Sci U S A* **2008,** 105, (37), 13793-8.
9. Wang, Z.; Park, K.; Comer, F.; Hsieh-Wilson, L. C.; Saudek, C. D.; Hart, G. W., Site-specific GlcNAcylation of human erythrocyte proteins: potential biomarker(s) for diabetes. *Diabetes* **2009,** 58, (2), 309-17.
10. Hu, P.; Shimoji, S.; Hart, G. W., Site-specific interplay between O-GlcNAcylation and phosphorylation in cellular regulation. *FEBS Lett* **2010,** 584, (12), 2526-38.
11. Teo, C. F.; Ingale, S.; Wolfert, M. A.; Elsayed, G. A.; Not, L. G.; Chatham, J. C.; Wells, L.; Boons, G. J., Glycopeptide-specific monoclonal antibodies suggest new roles for O-GlcNAc. *Nat Chem Biol* **2010,** 6, (5), 338-43.
12. Zachara, N. E.; Hart, G. W., O-GlcNAc a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular function in response to nutrition and stress. *Biochim Biophys Acta* **2004,** 1673, (1-2), 13-28.
13. Zachara, N. E.; Hart, G. W., O-GlcNAc modification: a nutritional sensor that modulates proteasome function. *Trends Cell Biol* **2004,** 14, (5), 218-21.
14. Dias, W. B.; Hart, G. W., O-GlcNAc modification in diabetes and Alzheimer's disease. *Mol Biosyst* **2007,** 3, (11), 766-72.
15. Hart, G. W.; Housley, M. P.; Slawson, C., Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **2007,** 446, (7139), 1017-22.
16. Copeland, R. J.; Bullen, J. W.; Hart, G. W., Cross-talk between GlcNAcylation and phosphorylation: roles in insulin resistance and glucose toxicity. *Am J Physiol Endocrinol Metab* **2008,** 295, (1), E17-28.
17. Laczy, B.; Hill, B. G.; Wang, K.; Paterson, A. J.; White, C. R.; Xing, D.; Chen, Y. F.; Darley-Usmar, V.; Oparil, S.; Chatham, J. C., Protein O-GlcNAcylation: a new signaling paradigm for the cardiovascular system. *Am J Physiol Heart Circ Physiol* **2009,** 296, (1), H13-28.

Fig. S1

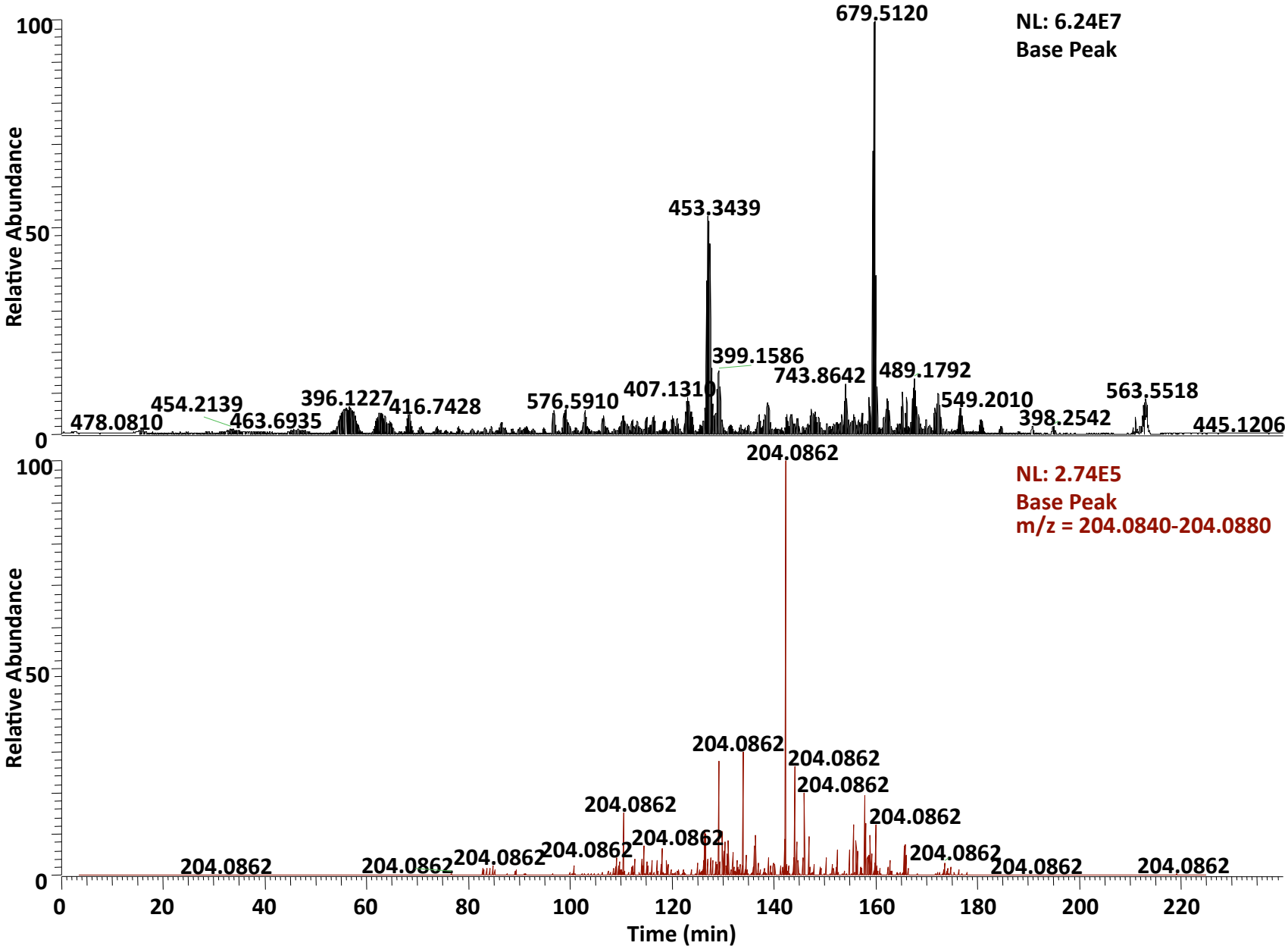

679.5120
NL: 6.24E7
Base Peak
453.3439
399.1586
743.8642
489.1792
407.1310
576.5910
396.1227
416.7428
454.2139
463.6935
478.0810
549.2010
563.5518
398.2542
445.1206
Relative Abundance
204.0862
NL: 2.74E5
Base Peak
m/z = 204.0840-204.0880
Relative Abundance
Time (min)

Fig. S2

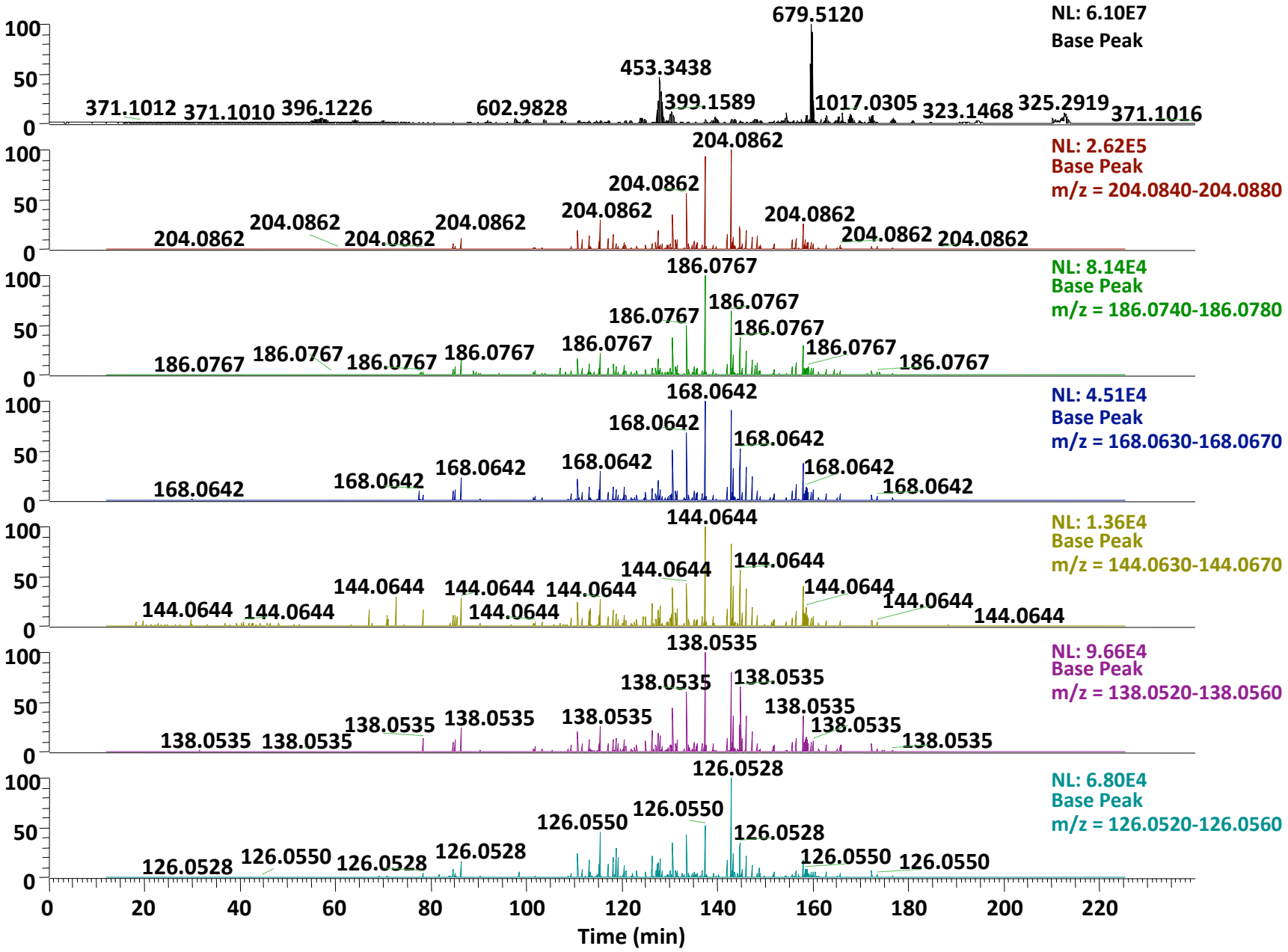

679.5120
NL: 6.10E7
Base Peak
453.3438
371.1012 371.1010 396.1226 602.9828 399.1589 1017.0305 323.1468 325.2919 371.1016
204.0862
NL: 2.62E5
Base Peak
m/z = 204.0840-204.0880
186.0767
NL: 8.14E4
Base Peak
m/z = 186.0740-186.0780
168.0642
NL: 4.51E4
Base Peak
m/z = 168.0630-168.0670
144.0644
NL: 1.36E4
Base Peak
m/z = 144.0630-144.0670
138.0535
NL: 9.66E4
Base Peak
m/z = 138.0520-138.0560
126.0528
126.0550
NL: 6.80E4
Base Peak
m/z = 126.0520-126.0560
0 20 40 60 80 100 120 140 160 180 200 220
Time (min)

Fig. S3

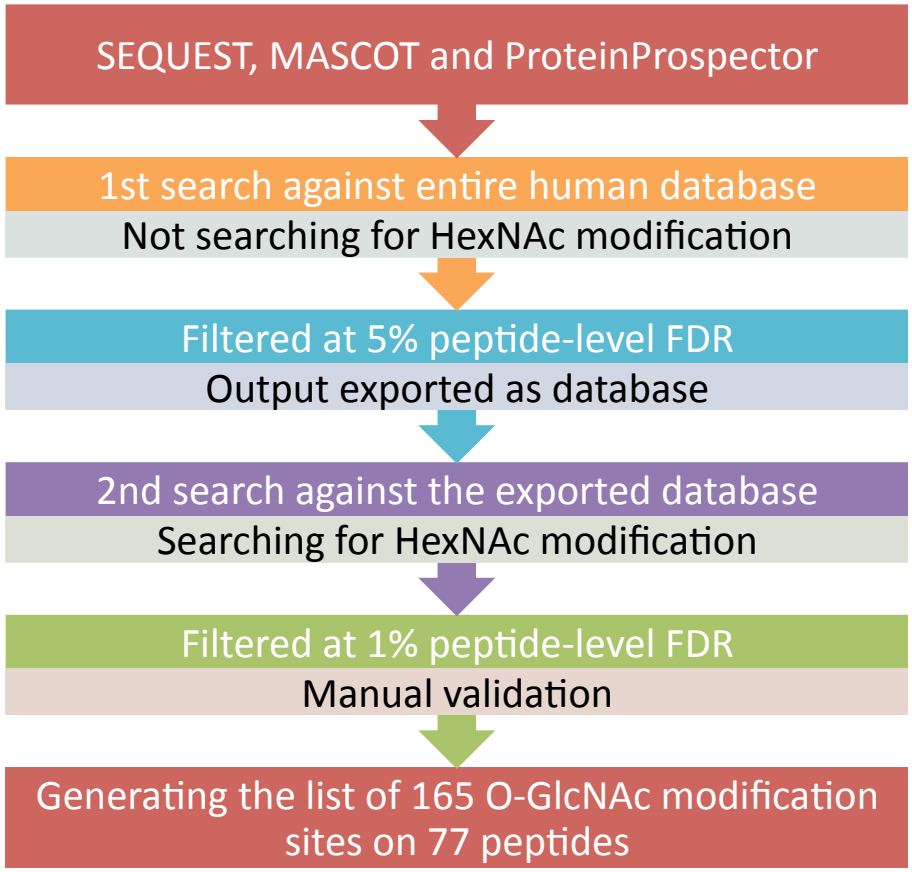

SEQUEST, MASCOT and ProteinProspector
1st search against entire human database
Not searching for HexNAc modification
Filtered at 5% peptide-level FDR
Output exported as database
2nd search against the exported database
Searching for HexNAc modification
Filtered at 1% peptide-level FDR
Manual validation
Generating the list of 165 O-GlcNAc modification sites on 77 peptides

Fig. S4

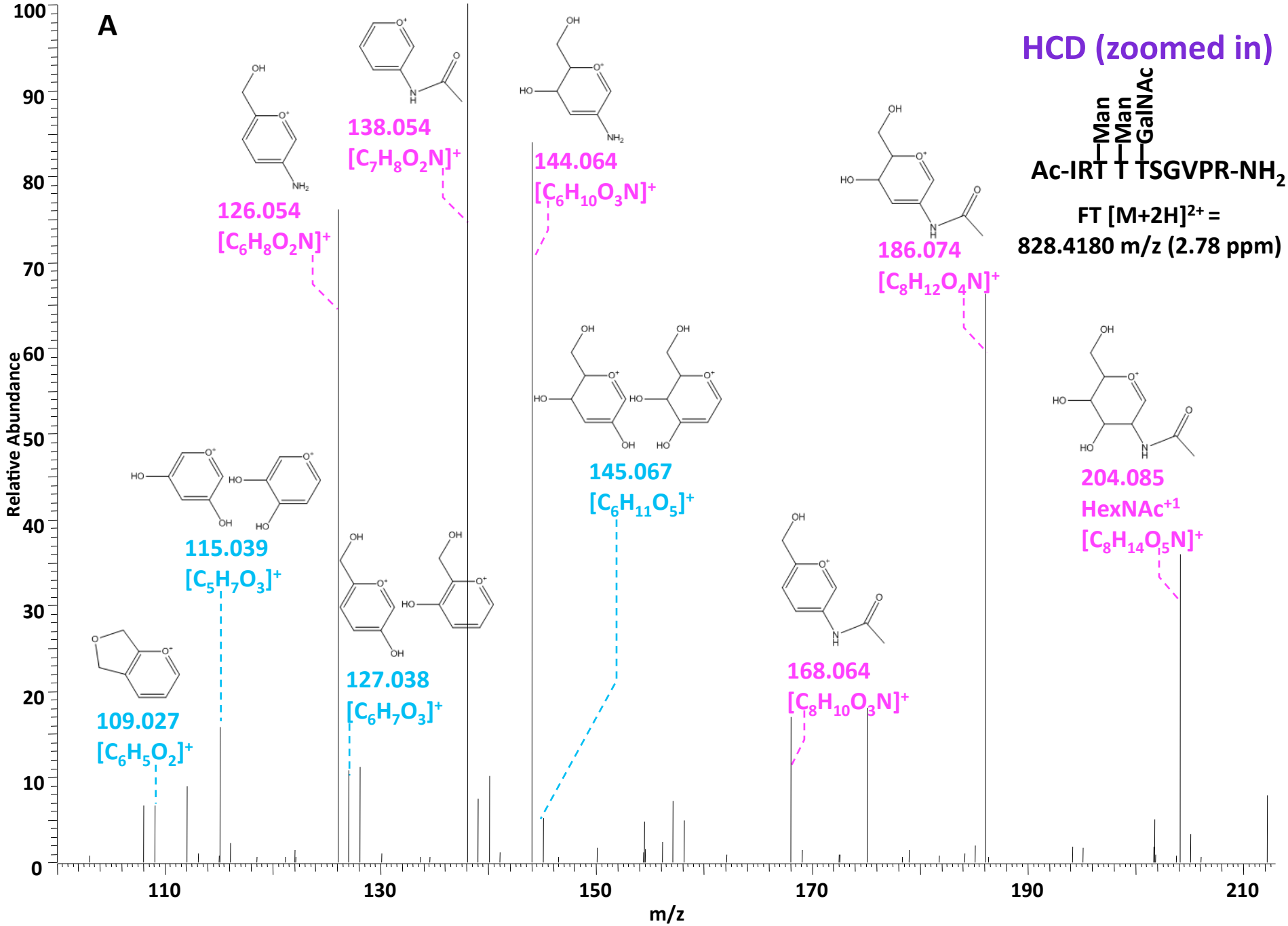

A
HCD (zoomed in)
Man
Man
GalNAc
Ac-IRT T TSGVPR-NH2
FT [M+2H]2+ =
828.4180 m/z (2.78 ppm)
126.054
[C6H8O2N]+
138.054
[C7H8O2N]+
144.064
[C6H10O3N]+
186.074
[C8H12O4N]+
204.085
HexNAc+1
[C8H14O5N]+
145.067
[C6H11O5]+
115.039
[C5H7O3]+
127.038
[C6H7O3]+
109.027
[C6H5O2]+
168.064
[C8H10O3N]+
Relative Abundance
m/z
0
10
20
30
40
50
60
70
80
90
100
110
130
150
170
190
210

Fig. S4

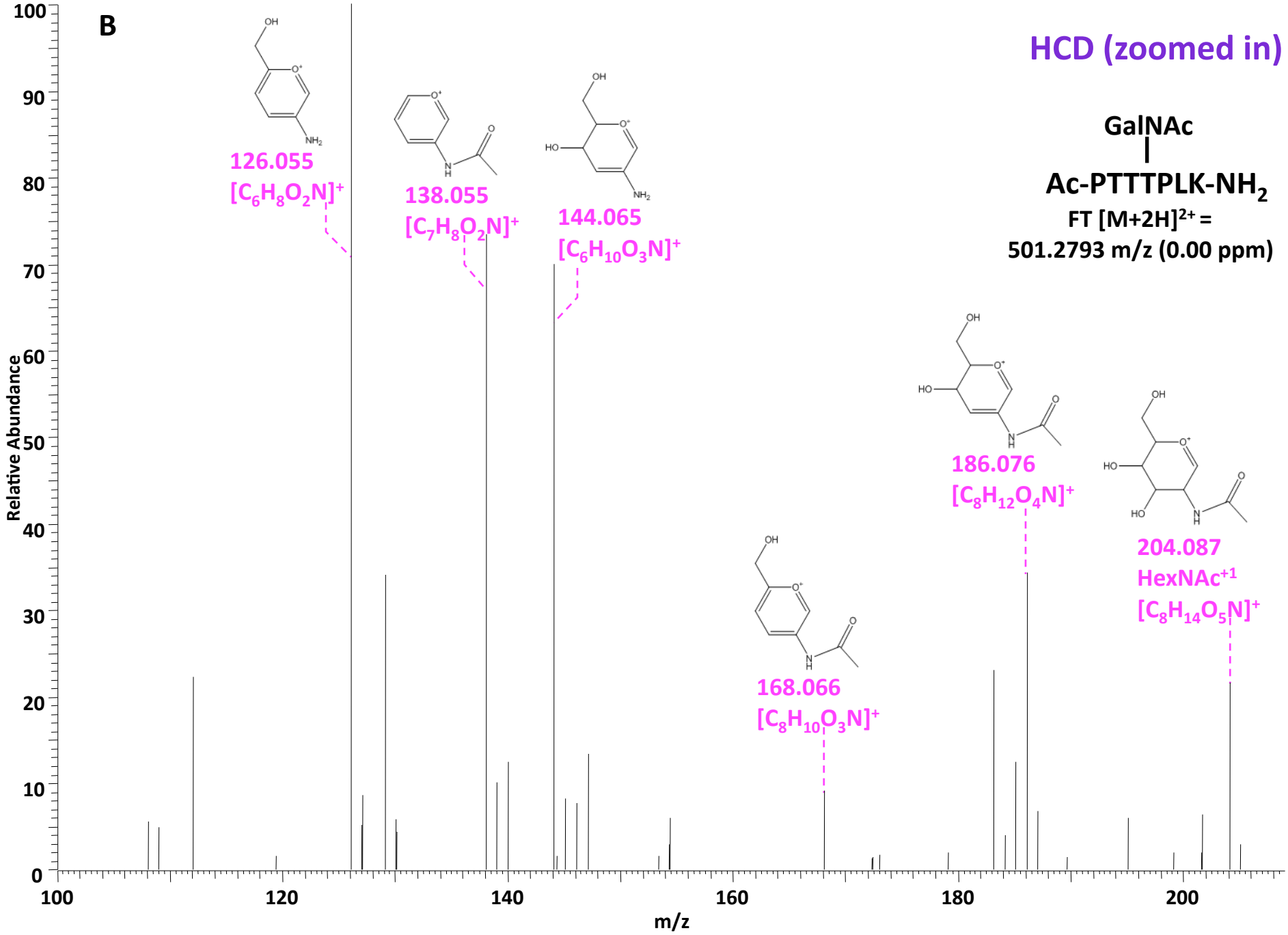

B
HCD (zoomed in)
GalNAc
Ac-PTTTPLK-NH2
FT [M+2H]2+ =
501.2793 m/z (0.00 ppm)
126.055
[C6H8O2N]+
138.055
[C7H8O2N]+
144.065
[C6H10O3N]+
168.066
[C8H10O3N]+
186.076
[C8H12O4N]+
204.087
HexNAc+1
[C8H14O5N]+
Relative Abundance
m/z
100
90
80
70
60
50
40
30
20
10
0
100
120
140
160
180
200

Fig. S4

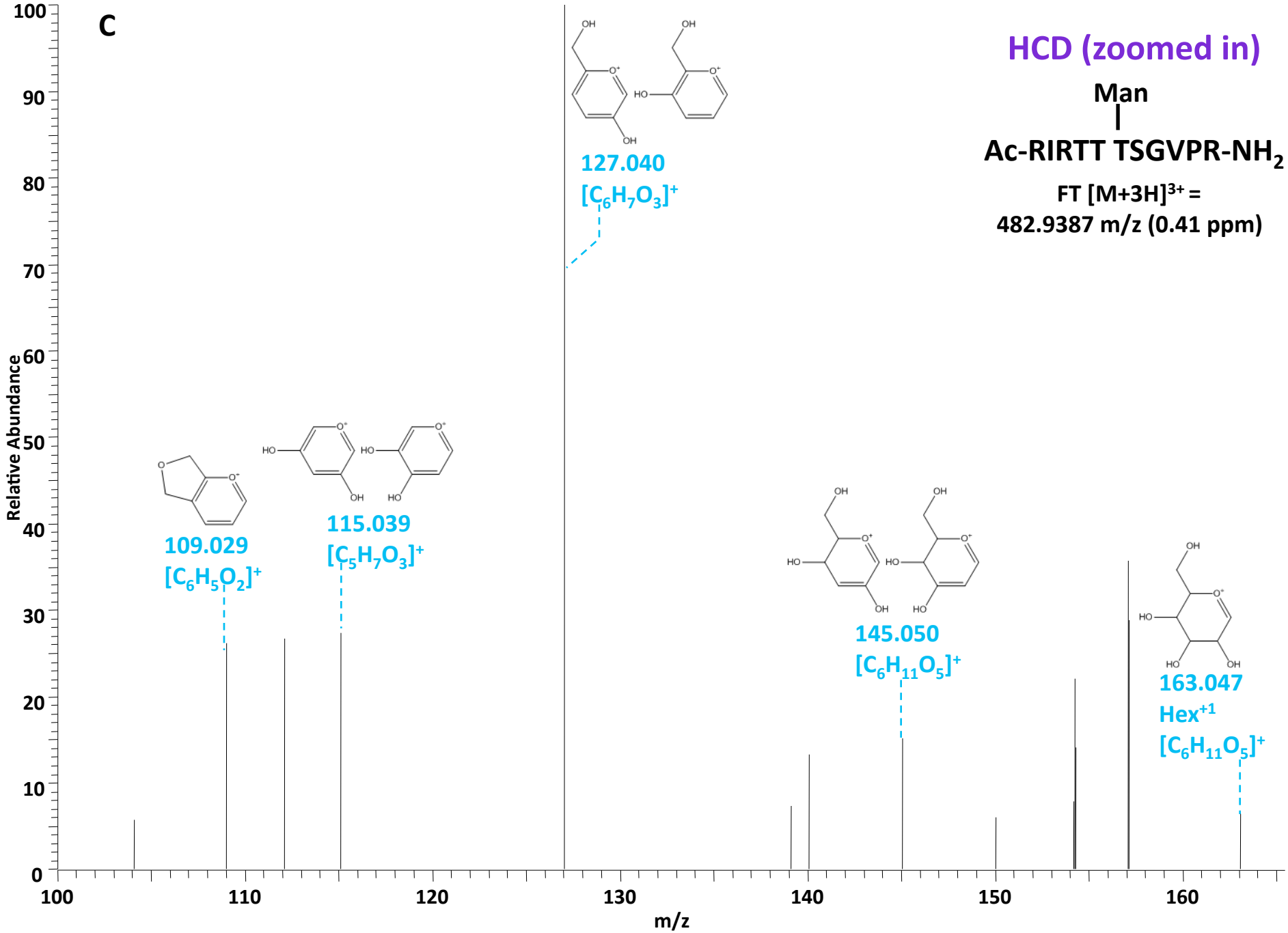

C
HCD (zoomed in)
Man
Ac-RIRTT TSGVPR-NH2
FT [M+3H]3+ =
482.9387 m/z (0.41 ppm)
127.040
[C6H7O3]+
109.029
[C6H5O2]+
115.039
[C5H7O3]+
145.050
[C6H11O5]+
163.047
Hex+1
[C6H11O5]+
Relative Abundance
m/z
100
90
80
70
60
50
40
30
20
10
0
100
110
120
130
140
150
160

Fig. S4

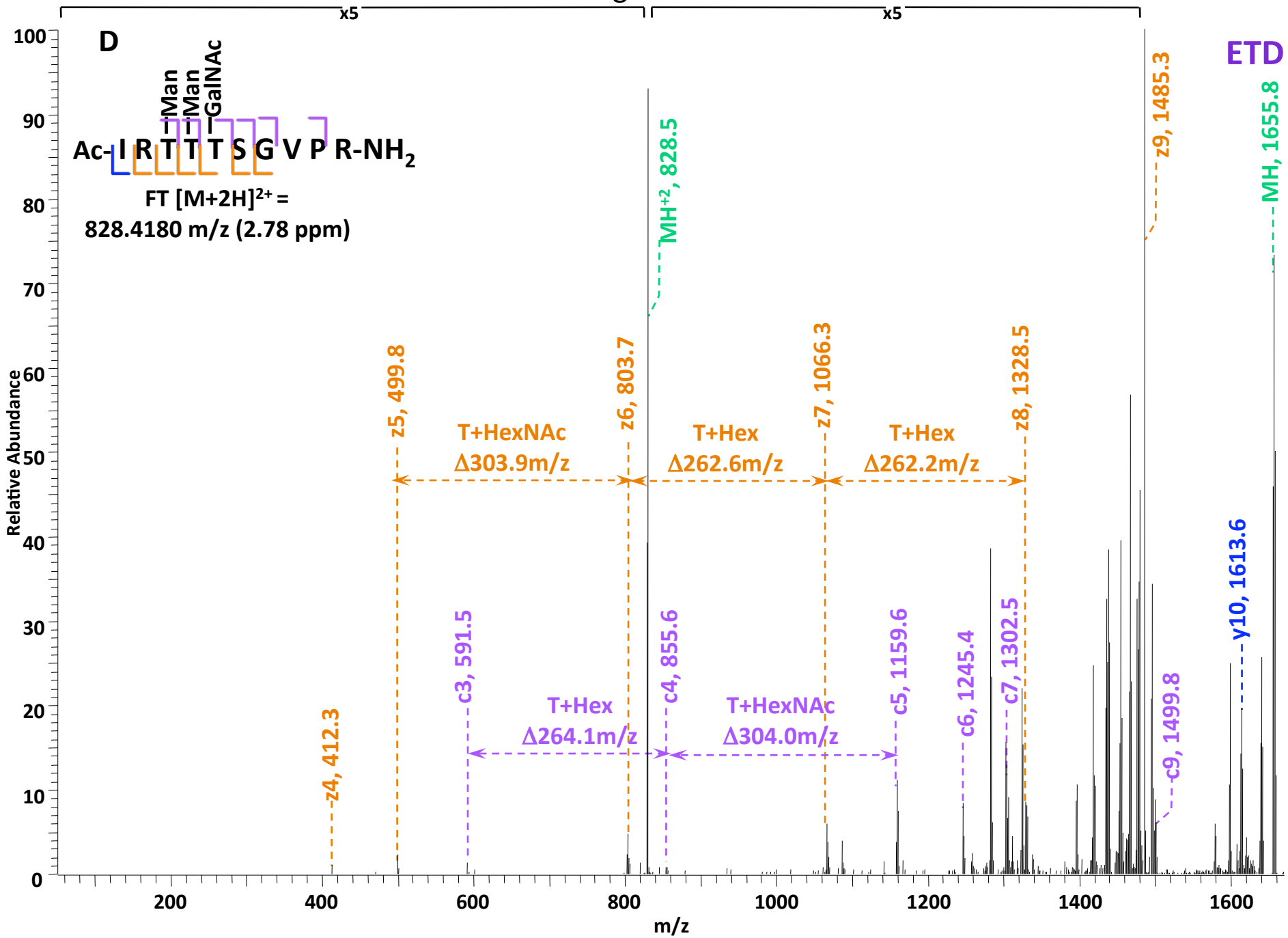

x5
x5
D
Ac-I R T T T S G V P R-NH2
Man
Man
GalNAc
FT [M+2H]2+ =
828.4180 m/z (2.78 ppm)
ETD
Relative Abundance
m/z
MH+2, 828.5
MH, 1655.8
z9, 1485.3
z8, 1328.5
z7, 1066.3
z6, 803.7
z5, 499.8
z4, 412.3
T+HexNAc
Δ303.9m/z
T+Hex
Δ262.6m/z
T+Hex
Δ262.2m/z
c3, 591.5
c4, 855.6
c5, 1159.6
c6, 1245.4
c7, 1302.5
c9, 1499.8
y10, 1613.6
T+Hex
Δ264.1m/z
T+HexNAc
Δ304.0m/z

Fig. S4

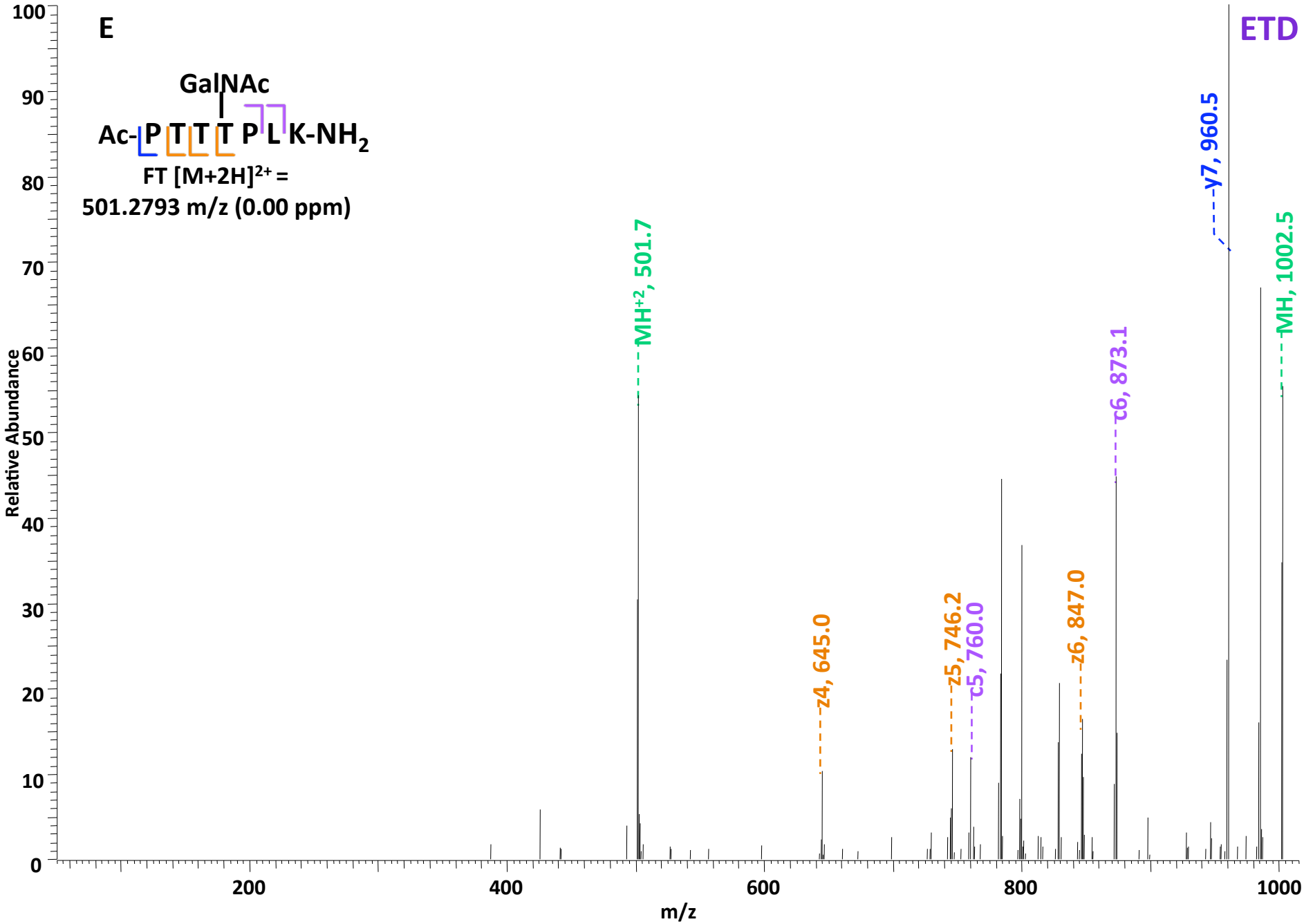
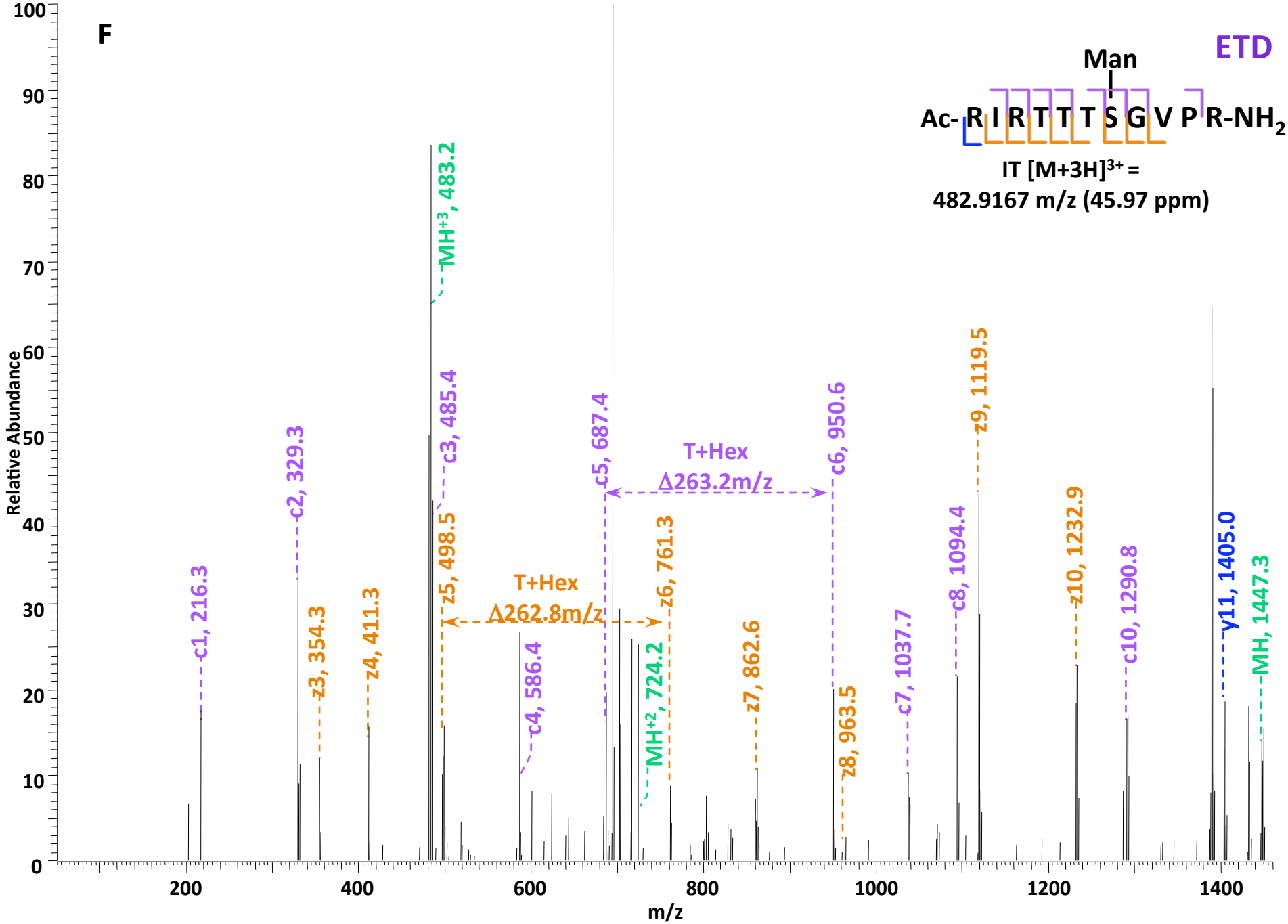

E
GalNAc
Ac-P T T T P L K-NH2
FT [M+2H]2+ =
501.2793 m/z (0.00 ppm)
ETD
MH+2, 501.7
z4, 645.0
z5, 746.2
c5, 760.0
z6, 847.0
c6, 873.1
y7, 960.5
MH, 1002.5
Relative Abundance
m/z
0
10
20
30
40
50
60
70
80
90
100
200
400
600
800
1000

Fig. S4


F
ETD
Man
Ac-R I R T T T S G V P R-NH2
IT [M+3H]3+ =
482.9167 m/z (45.97 ppm)
Relative Abundance
m/z
c1, 216.3
c2, 329.3
z3, 354.3
z4, 411.3
MH+3, 483.2
c3, 485.4
z5, 498.5
T+Hex
Δ262.8m/z
c4, 586.4
c5, 687.4
T+Hex
Δ263.2m/z
MH+2, 724.2
z6, 761.3
z7, 862.6
z8, 963.5
c6, 950.6
c7, 1037.7
c8, 1094.4
z9, 1119.5
z10, 1232.9
c10, 1290.8
y11, 1405.0
MH, 1447.3