Supporting Information
for
"Comprehensive Characterization of Glycosylation and Hydroxylation of Basement Membrane Collagen IV by High-Resolution Mass Spectrometry"

Trayambak Basak[a, b], Lorenzo Vega-Montoto[a,b,#], Lisa J. Zimmerman[c], David L. Tabb[c, d], Billy G. Hudson[a, b, c] and Roberto M. Vanacore[a, b,*]

[a] Department of Medicine, Division of Nephrology and Hypertension, [b] Center for Matrix Biology, [c]Department of Biochemistry, [d]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37232

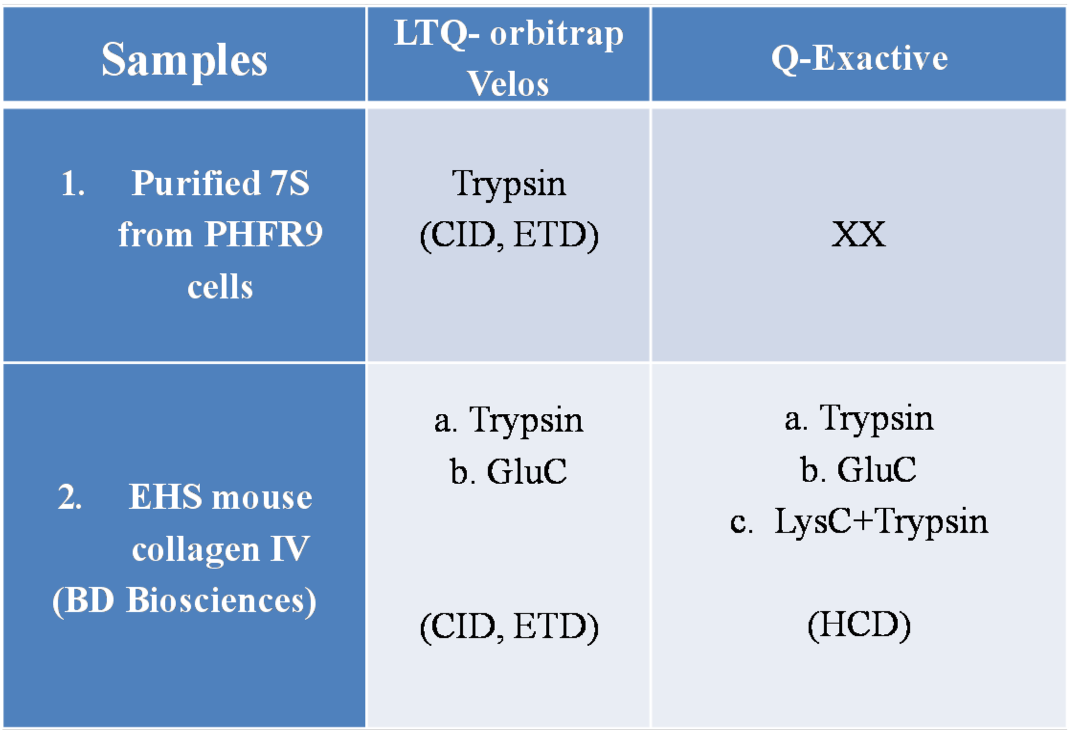
Running Title: Mass spectrometry analysis of collagen IV

To whom correspondence should be addressed: Department of Medicine, Division of Nephrology and Hypertension, Vanderbilt University Medical Center, 1161 21st Avenue South, B-3113 Medical Center North, Nashville, TN 37232, USA, Tel.: (615) 322-8323; Fax: (615) 343-7156; E-mail: roberto.vanacore@vanderbilt.edu

[#]Current address: LVM is at Idaho National Laboratory, Idaho Falls, ID.
*Corresponding author

Table of Content

Supporting table S1: Detail of samples analyzed with multiple nzymes and different fragmentation method using LTQ-orbitrap Velos and Q-Exactive



Samples	LTQ- orbitrap Velos	Q-Exactive
1. Purified 7S from PHFR9 cells	Trypsin (CID, ETD)	XX
2. EHS mouse collagen IV (BD Biosciences)	a. Trypsin b. GluC (CID, ETD)	a. Trypsin b. GluC c. LysC+Trypsin (HCD)

Supplementary table S2 Comparative analysis of O-glycosylation sites detected in mouse and human col4a1

O-glycosylation sites	Mouse Collagen IV	Human Collagen IV
$\mathbf{K^{45}}$	Glucosyl-galactosyl - hydroxylysine	Not detected
$\mathbf{K^{48}}$	Glucosyl-galactosyl - hydroxylysine	Not detected
$\mathbf{K^{78}}$	Galactosyl / glucosyl-galactosyl-hydroxylysine	Galactosyl / glucosyl-galactosyl-hydroxylysine
$\mathbf{K^{90}}$	Galactosyl / glucosyl-galactosyl-hydroxylysine	Galactosyl / glucosyl-galactosyl-hydroxylysine
$\mathbf{K^{155}}$	No glycosylation detected	Galactosyl - hydroxylysine
$\mathbf{K^{217}}$	Glucosyl-galactosyl - hydroxylysine	Not detected
$\mathbf{K^{228}}$	Glucosyl-galactosyl - hydroxylysine	Not detected
$\mathbf{K^{231}}$	Glucosyl-galactosyl - hydroxylysine	Not detected
$\mathbf{K^{295}}$	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{298}}$	Not conserved	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{322}}$	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{343}}$	No glycosylation detected	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{361}}$	Glucosyl-galactosyl - hydroxylysine	Galactosyl / glucosyl-galactosyl-hydroxylysine
$\mathbf{K^{393}}$	Glucosyl-galactosyl - hydroxylysine	Not detected
$\mathbf{K^{460}}$	Glucosyl-galactosyl - hydroxylysine	Galactosyl - hydroxylysine
$\mathbf{K^{463}}$	Galactosyl / glucosyl-galactosyl-hydroxylysine	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{497}}$	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{527}}$	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{573}}$	Galactosyl / glucosyl-galactosyl-hydroxylysine	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{582}}$	Glucosyl-galactosyl - hydroxylysine	Hydroxylysine
$\mathbf{K^{617}}$	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{635}}$	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{681}}$	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{698}}$	Hydroxylysine	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{731}}$	No glycosylation detected	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{742}}$	Hydroxylysine	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{757}}$	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{825}}$	Glucosyl-galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{828}}$	Glucosyl-galactosyl - hydroxylysine	Hydroxylysine
$\mathbf{K^{1025}}$	Galactosyl - hydroxylysine	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{1043}}$	Galactosyl / glucosyl-galactosyl-hydroxylysine	Hydroxylysine
$\mathbf{K^{1046}}$	Galactosyl / glucosyl-galactosyl-hydroxylysine	Galactosyl - hydroxylysine
$\mathbf{K^{1049}}$	Hydroxylysine	Galactosyl / glucosyl-galactosyl-hydroxylysine
$\mathbf{K^{1066}}$	Not detected	Galactosyl / glucosyl-galactosyl-hydroxylysine
$\mathbf{K^{1081}}$	Galactosyl / glucosyl-galactosyl-hydroxylysine	Hydroxylysine
$\mathbf{K^{1117}}$	Galactosyl - hydroxylysine	Galactosyl - hydroxylysine
$\mathbf{K^{1120}}$	Glucosyl-galactosyl - hydroxylysine	Hydroxylysine
$\mathbf{K^{1132}}$	Galactosyl - hydroxylysine	Galactosyl / glucosyl-galactosyl-hydroxylysine
$\mathbf{K^{1150}}$	No glycosylation detected	Galactosyl - hydroxylysine
$\mathbf{K^{1165}}$	Galactosyl - hydroxylysine	Hydroxylysine
$\mathbf{K^{1185}}$	Galactosyl – hydroxylysine	Glucosyl-galactosyl – hydroxylysine
$\mathbf{K^{1188}}$	Hydroxylysine	Galactosyl / glucosyl-galactosyl-hydroxylysine
$\mathbf{K^{1206}}$	Glucosyl-galactosyl - hydroxylysine	Hydroxylysine
$\mathbf{K^{1268}}$	Galactosyl - hydroxylysine	Hydroxylysine
$\mathbf{K^{1283}}$	Galactosyl - hydroxylysine	Hydroxylysine
$\mathbf{K^{1304}}$	Galactosyl - hydroxylysine	Galactosyl / glucosyl-galactosyl-hydroxylysine
$\mathbf{K^{1319}}$	Not detected	Galactosyl - hydroxylysine
$\mathbf{K^{1328}}$	Not detected	Glucosyl-galactosyl - hydroxylysine
$\mathbf{K^{1340}}$	Not detected	Galactosyl / glucosyl-galactosyl-hydroxylysine
$\mathbf{K^{1371}}$	Galactosyl - hydroxylysine	No glycosylation detected
$\mathbf{K^{1380}}$	Glucosyl-galactosyl - hydroxylysine	No glycosylation detected
$\mathbf{K^{1407}}$	Glucosyl-galactosyl - hydroxylysine	Not detected

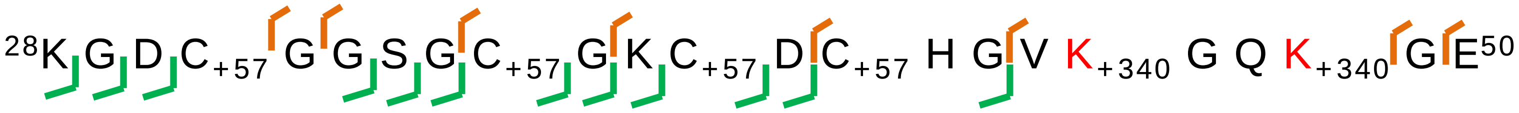
Supporting figure S1: A representative diagram showing improved sequence coverage of mouse col4a1 by optimizing the number of allowed posttranslational modifications and miscleavages in the database search using MyriMatch. A. A MS data file from a trypsin digested EHS collagen sample was searched with a standard database search using carbamidomethylation as static modification, oxidation of methionine as dynamic modification and allowing for only 2 dynamic modifications per peptide. B. The same MS data file searched with MyriMatch, but increasing the number of miscleavages (up to 4) and dynamic modifications (up to 10) using the hydroxylation and O-linked glycosylation motifs described in Table 1.

Supporting figure S2: Peptide spectrum matches (PSMs) of 35 O-glycosylation,1 hydroxyproline site at Xaa position of Gly-HyP-Gln motif and 7 3-hydroxyproline sites in mouse col4a1 chain. PSMs were annotated using pLABEL[1] or IDPicker utility software Ion Matcher[2].

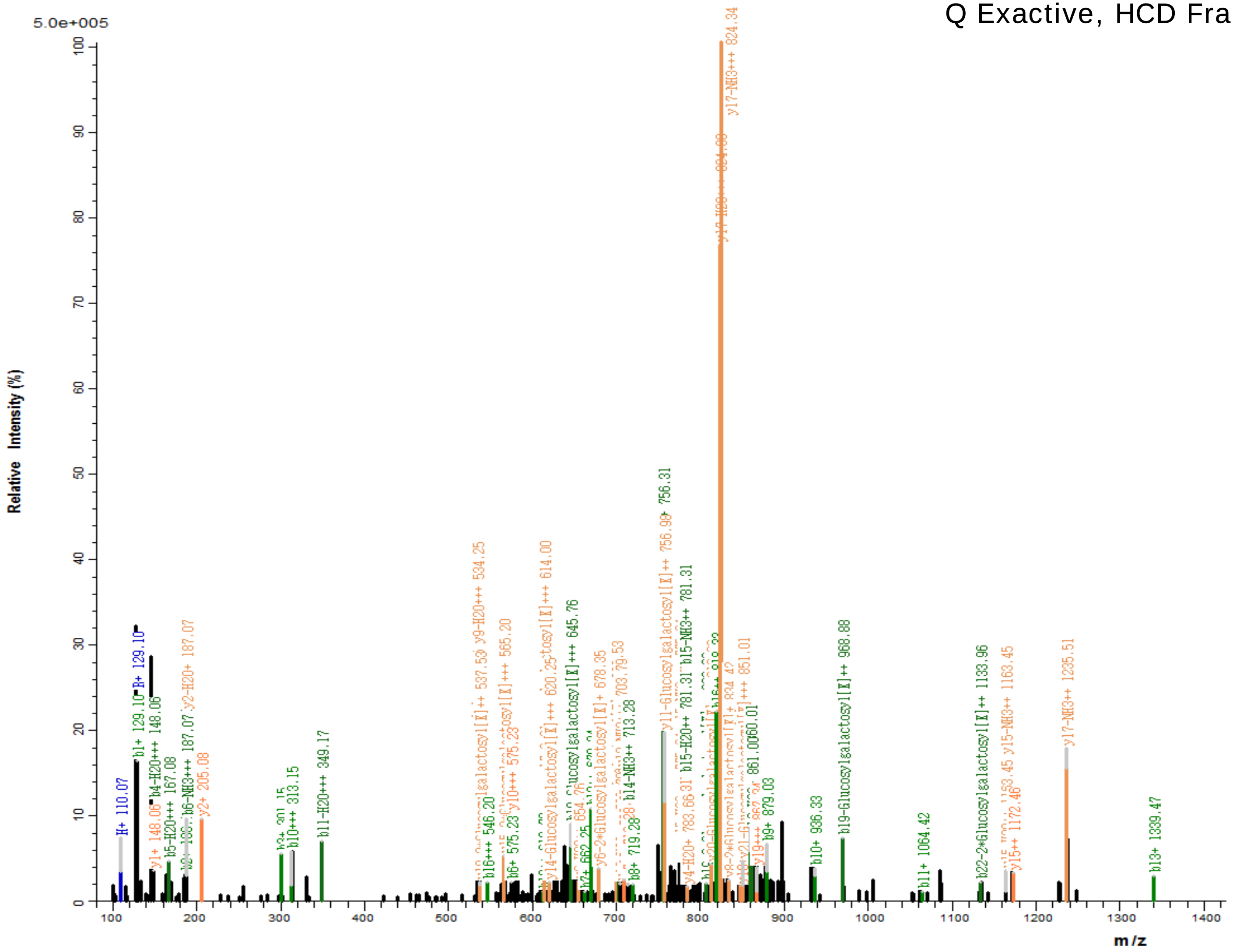
m/z= 780.3089 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK[45,48]

Mouse col4a1 chain


$^{28}K\ G\ D\ C_{+57}\ G\ G\ S\ G\ C_{+57}\ G\ K\ C_{+57}\ D\ C_{+57}\ H\ G\ V\ K_{+340}\ G\ Q\ K_{+340}\ G\ E^{50}$

Q Exactive, HCD Fragmentation



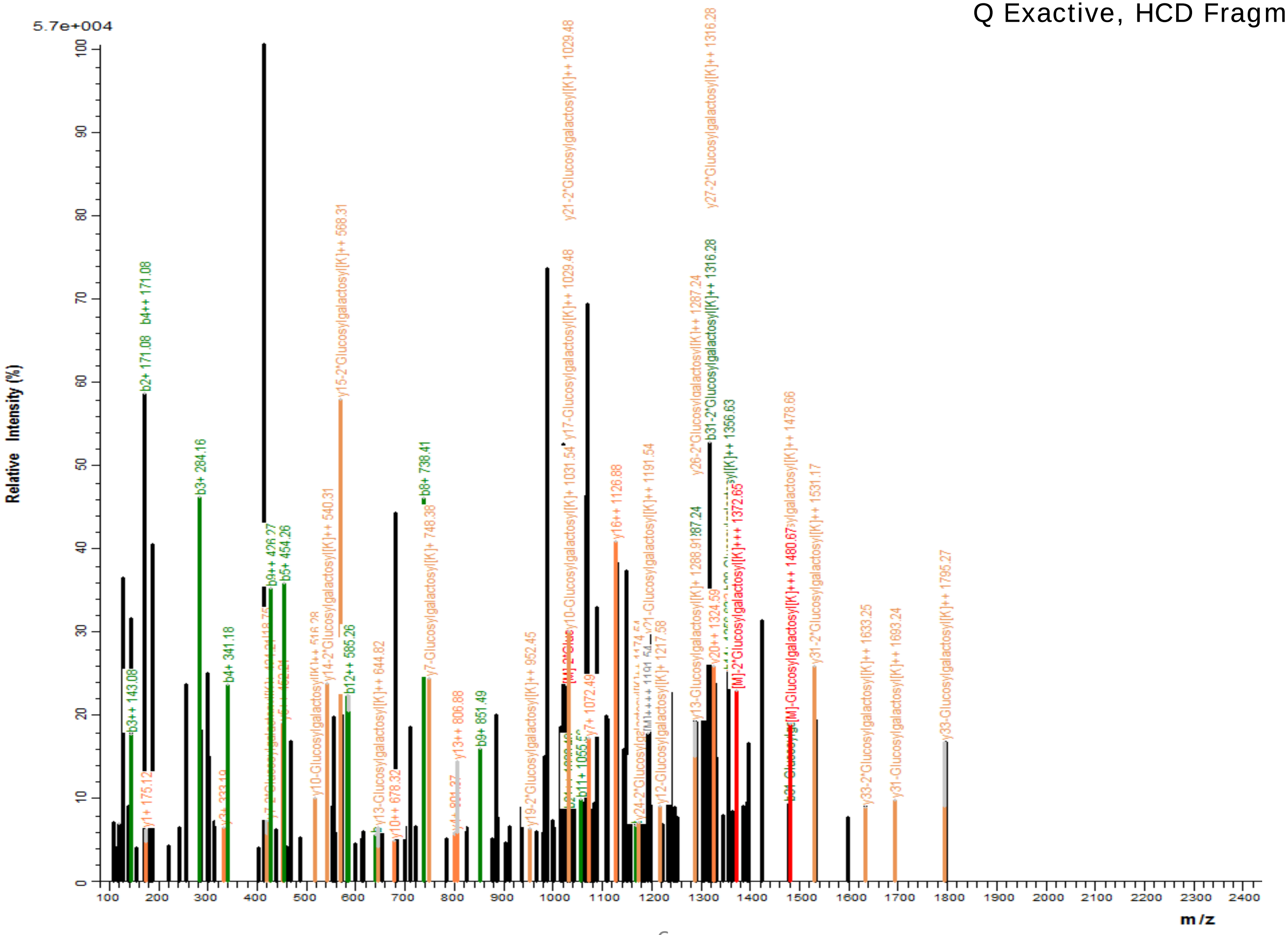
m/z= 1191.5593 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{78,90}

Mouse col4a1 chain

^{52}G L P$_{+16}$ G L Q G V I G F P$_{+16}$ G M Q G P E G P H G P P$_{+16}$ G Q K$_{+340}$ G D A G E P$_{+16}$ G L P$_{+16}$ G T K$_{+340}$ G T R^{93}

Q Exactive, HCD Fragmentation



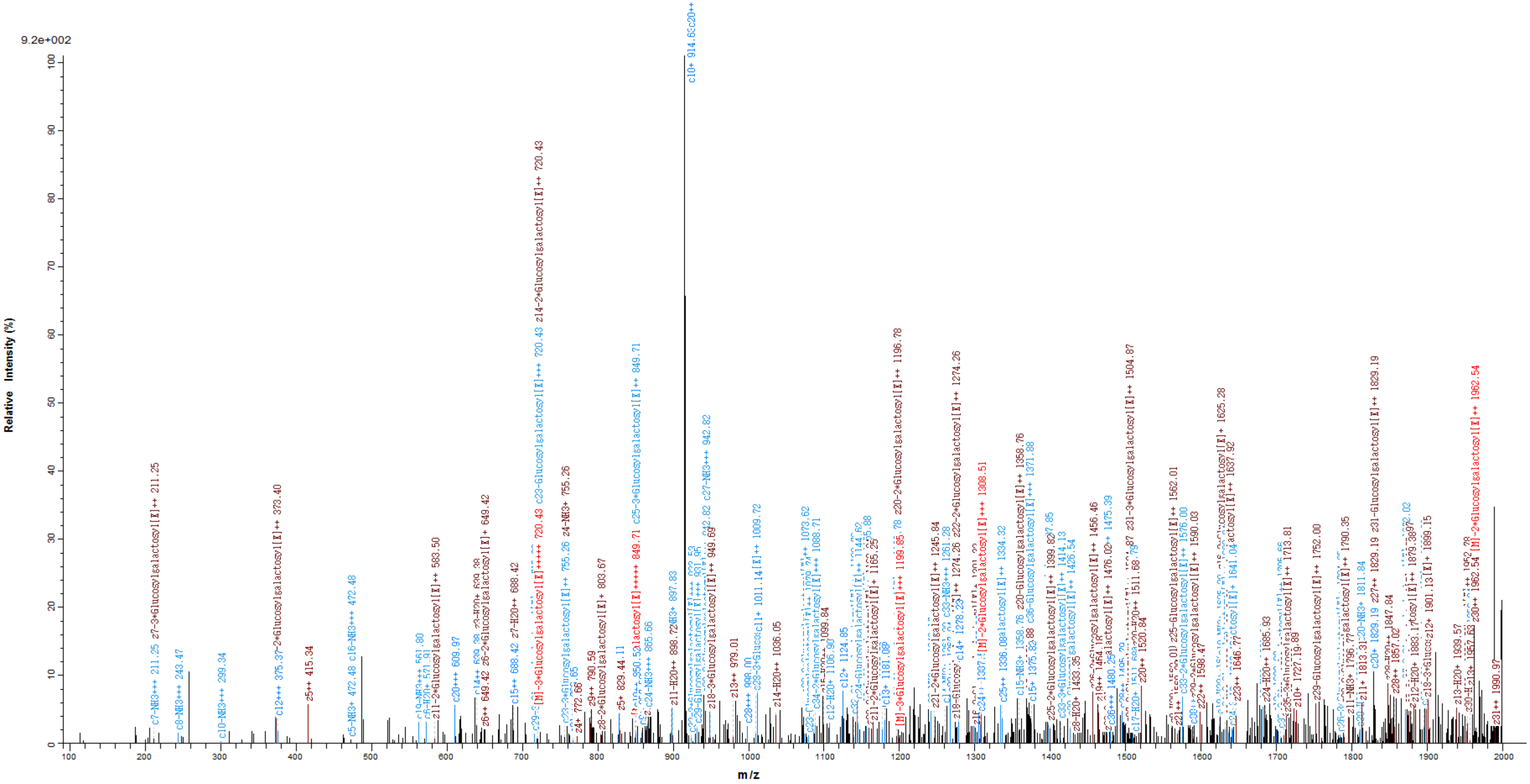
m/z= 915.5043,
charge=+5

PSM showing the identification of Galactosyl-HyK[217,228,231]

Mouse col4a1 chain


^{197}G P P$_{+16}$ G F T G P P$_{+16}$ G P P$_{+16}$ G P P G P P$_{+16}$ G E K$_{+340}$ G Q M G S S F Q G P K$_{+340}$ G D K$_{+340}$ G E^{233}

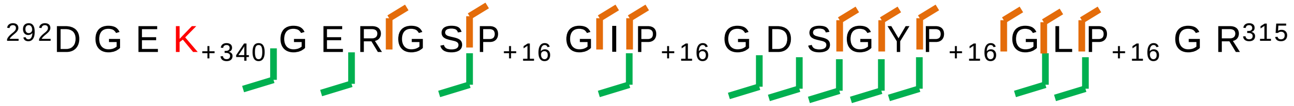
Orbitrap-Velos, ETD Fragmentation



m/z= 690.5574,
charge=+4

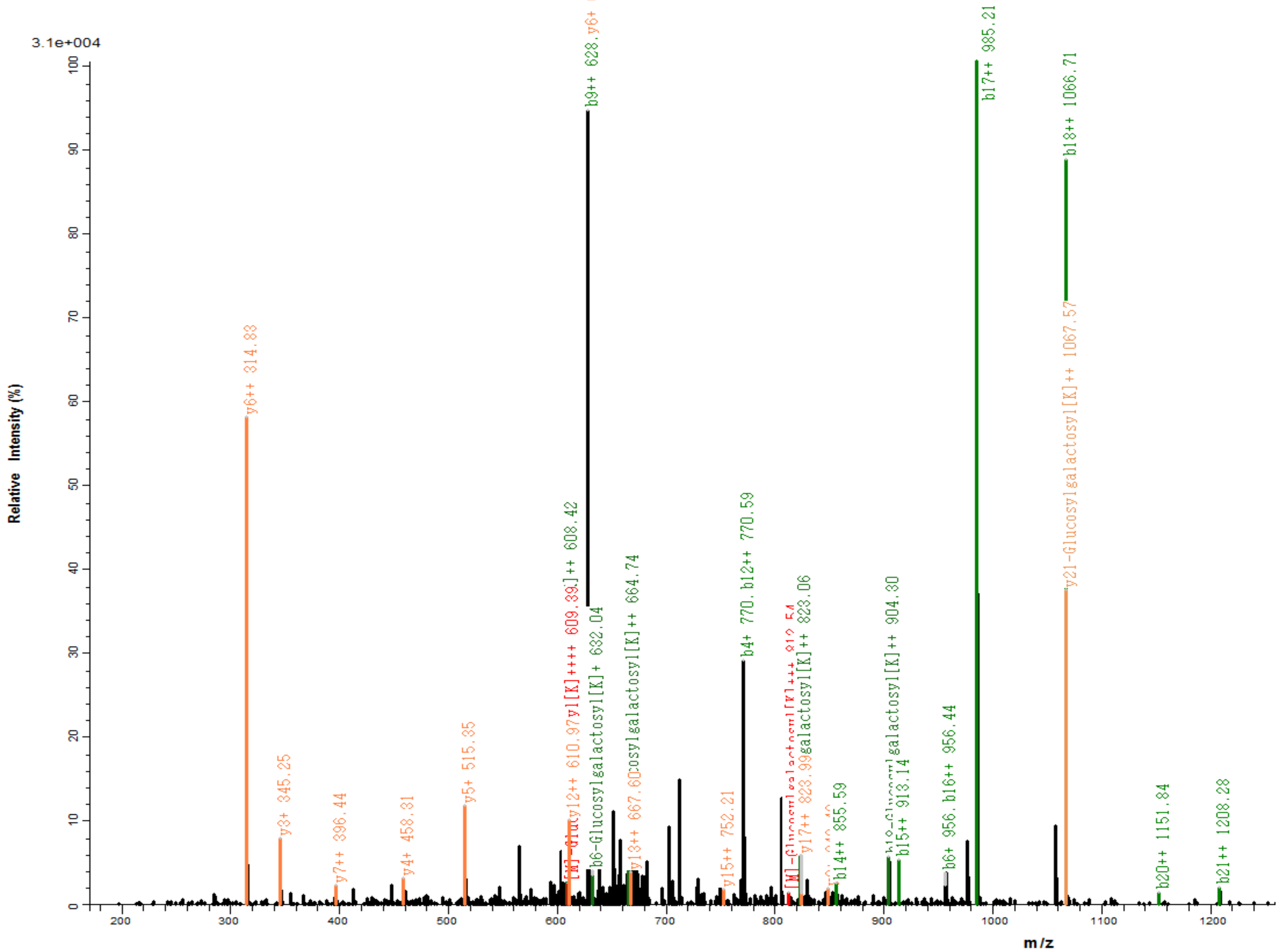
PSM showing the identification of Glucosylgalactosyl-HyK^{95}

Mouse col4a1 chain



^{292}D G E K$_{+340}$ G E R G S P$_{+16}$ G I P$_{+16}$ G D S G Y P$_{+16}$ G L P$_{+16}$ G R^{315}

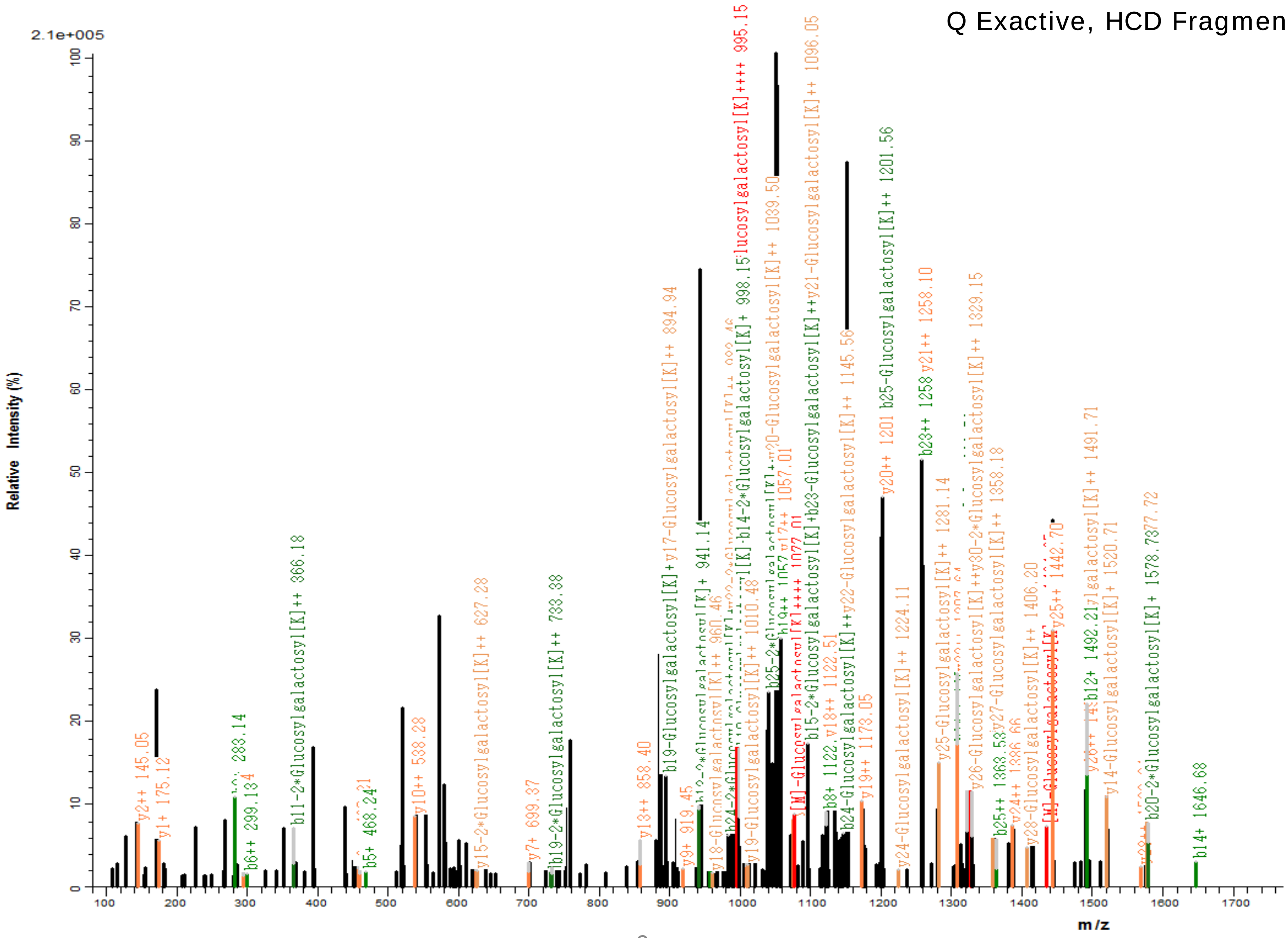
Q Exactive, HCD Fragmentation



m/z= 1157.5483 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK[322,343]

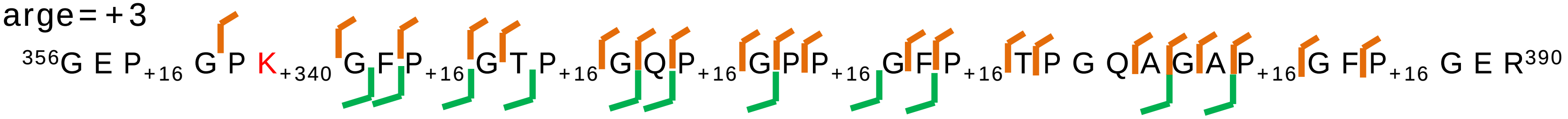
Mouse col4a1 chain


Q Exactive, HCD Fragmentation

m/z= 1255.8875,
charge=+3

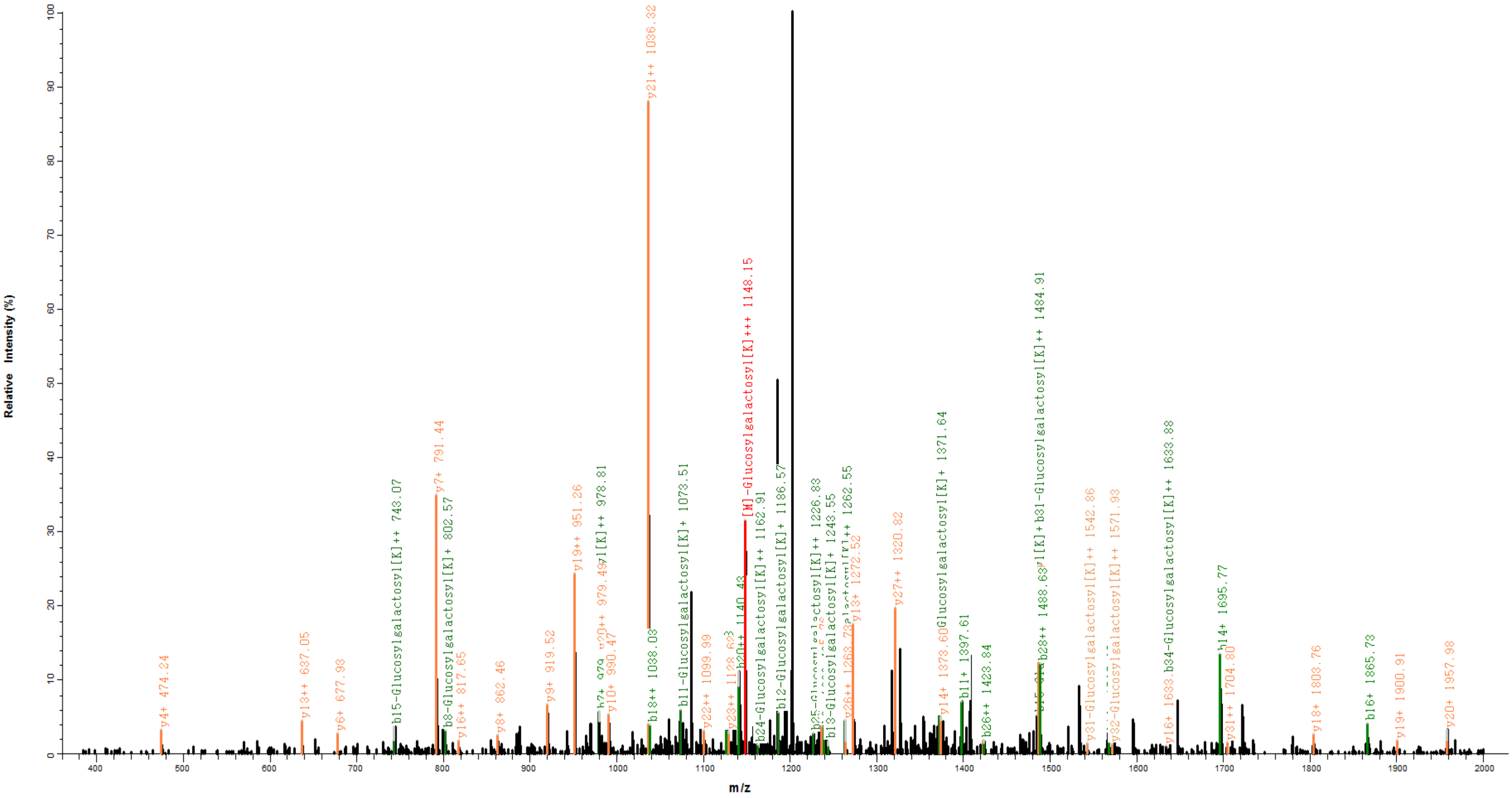
PSM showing the identification of Glucosylgalactosyl-HyK^{361}

Mouse col4a1 chain



^{356}G E P$_{+16}$ G P K$_{+340}$ G F P$_{+16}$ G T P$_{+16}$ G Q P$_{+16}$ G P P$_{+16}$ G F P$_{+16}$ T P G Q A G A P$_{+16}$ G F P$_{+16}$ G E R^{390}

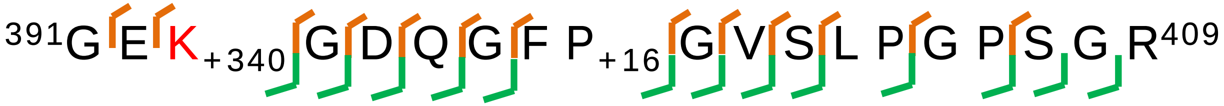
Q Exactive, HCD Fragmentation



m/z= 733.34 charge=+3

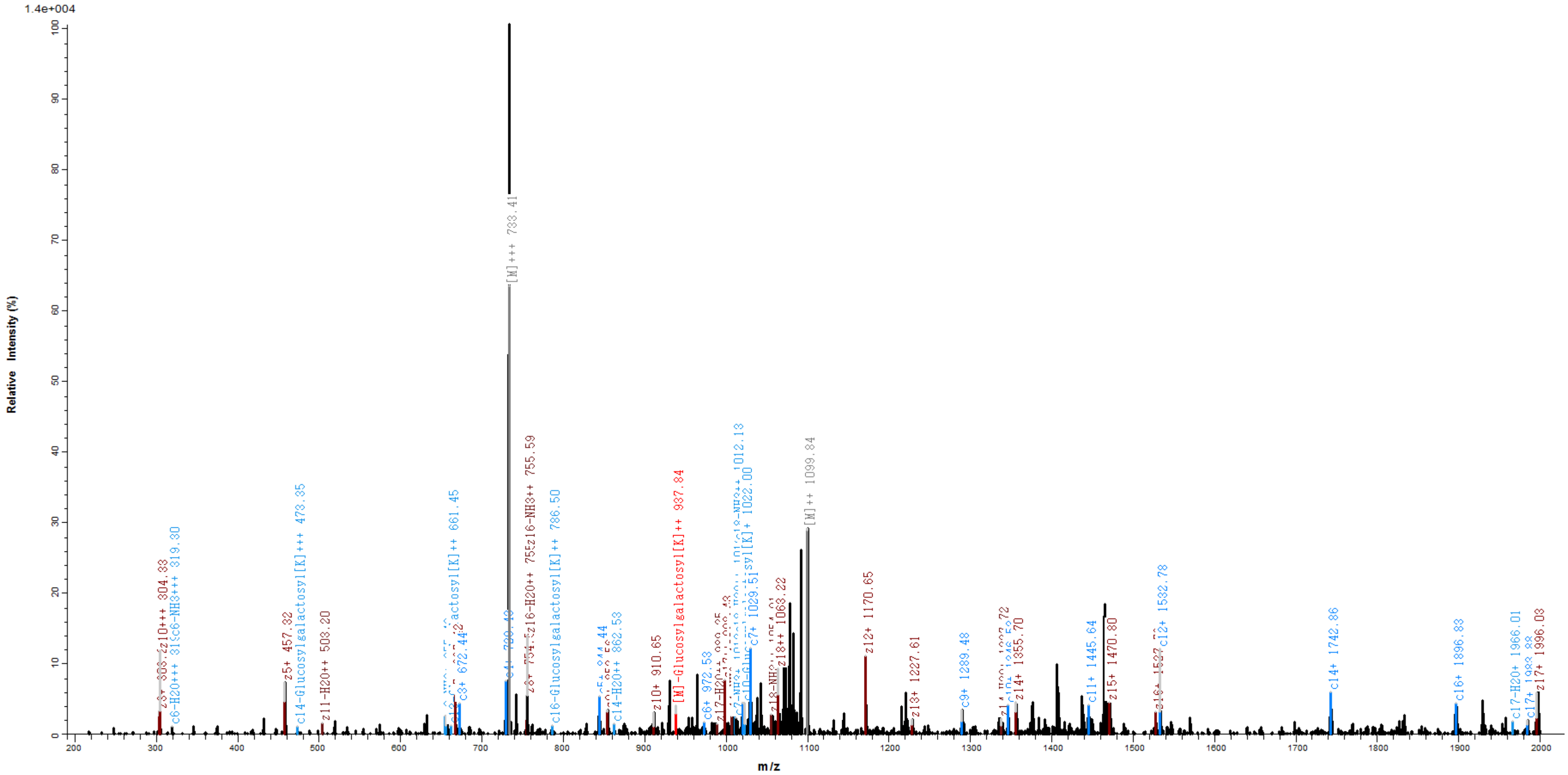
PSM showing the identification of Glucosylgalactosyl-HyK^{393}

Mouse col4a1 chain



^{391}G E K$_{+340}$ G D Q G F P$_{+16}$ G V S L P G P S G R^{409}

Orbitrap-Velos, ETD Fragmentation



11

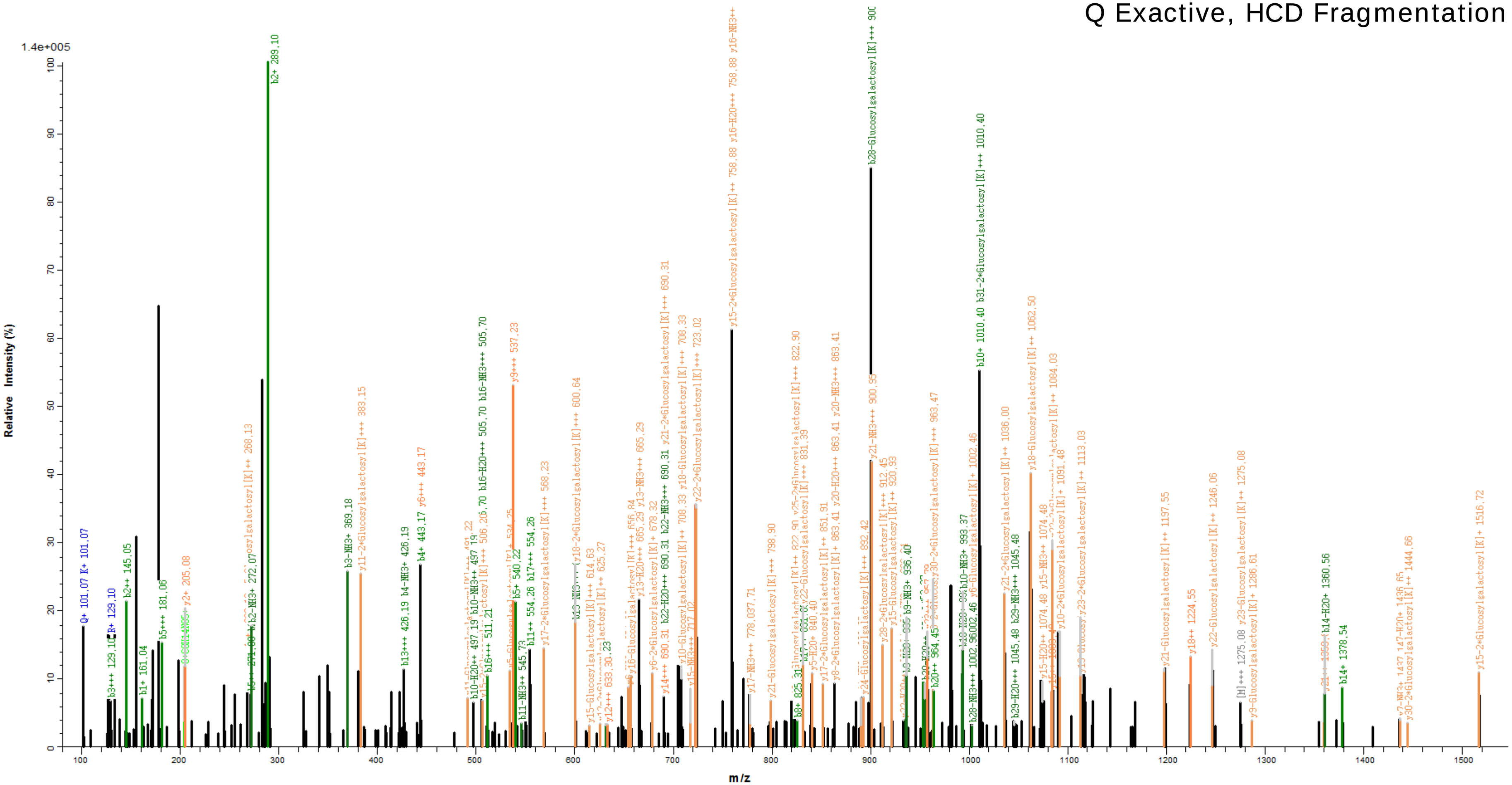
m/z= 956.9242 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{460,463}

Mouse col4a1 chain

$^{434}C_{+57}$ Q P G P P_{+16} G D Q G P P_{+16} G T P G Q P G L T G E V G Q K_{+340} G Q K_{+340} G E^{465}

Q Exactive, HCD Fragmentation



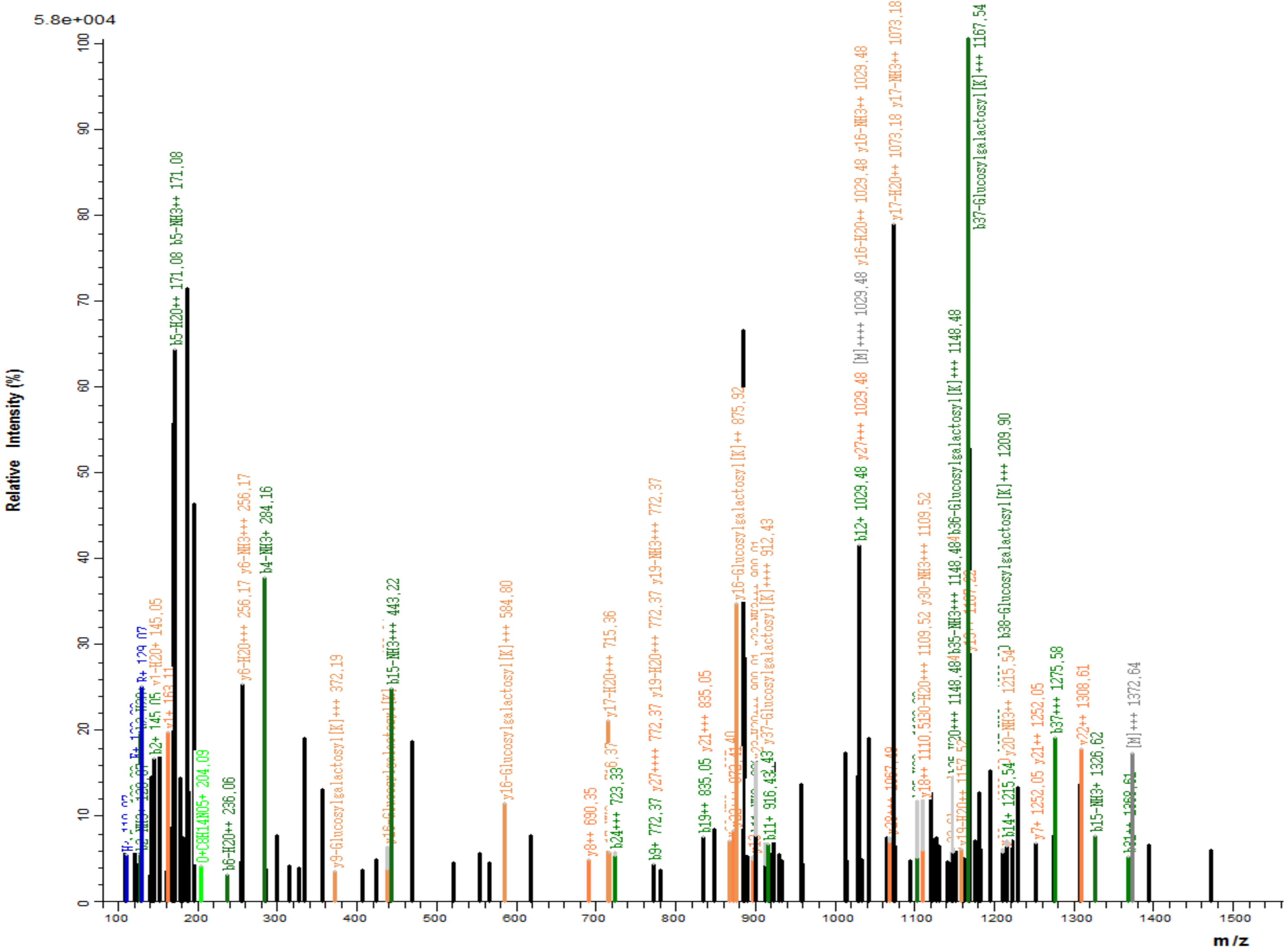
m/z= 823.7868 charge=+5

PSM showing the identification of Glucosylgalactosyl-HyK^{481}

Mouse col4a1 chain

^{1011}G S V G G M G L P$_{+16}$ G S P$_{+16}$ G E K G V P$_{+16}$ G I P$_{+16}$ G S Q G V P$_{+16}$ G S P G E K$_{+340}$ G A K$_{+178}$ G E K$_{+16}$^{1049}

Q Exactive, HCD Fragmentation



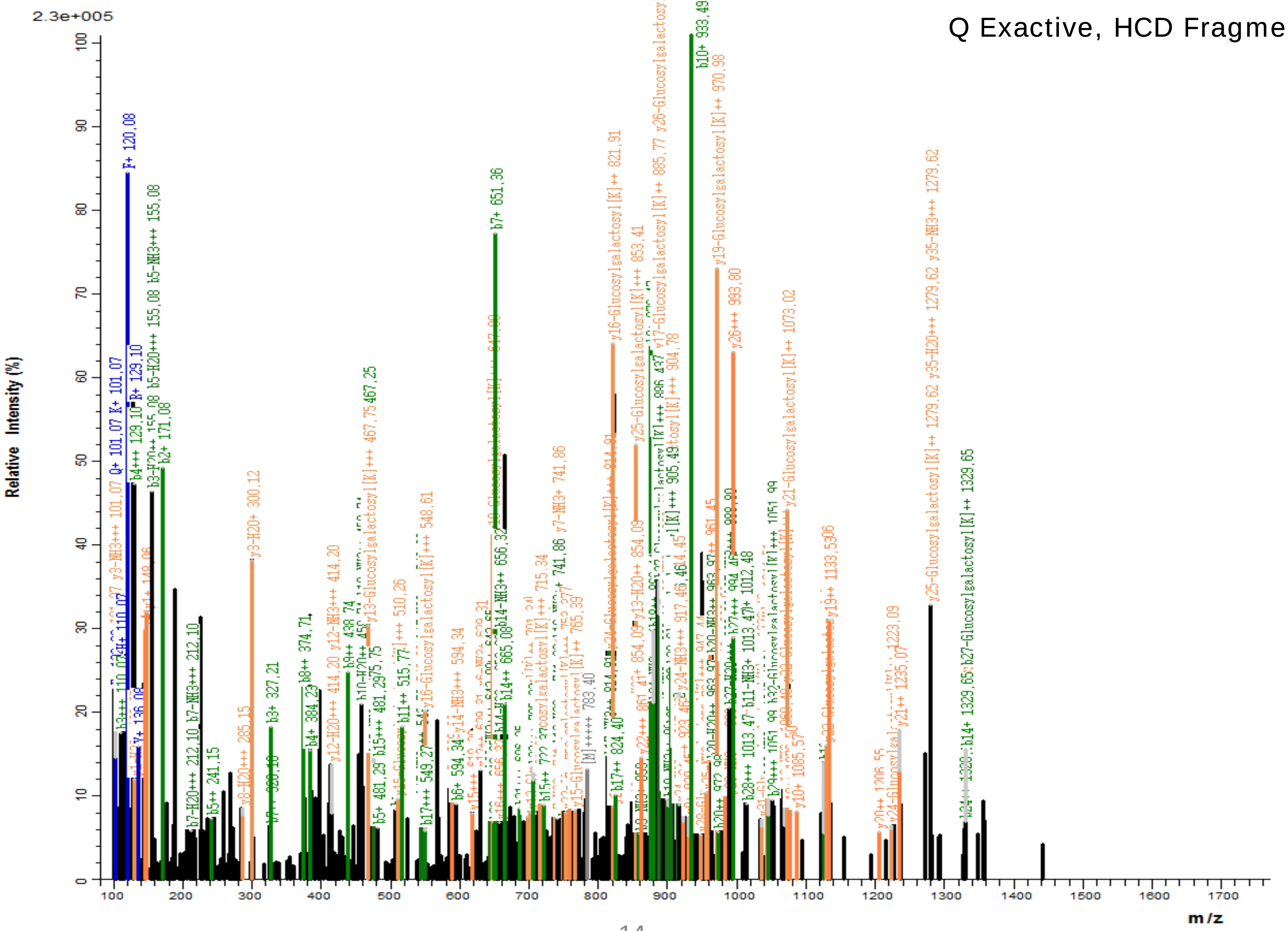
m/z= 782.9878 charge=+5

PSM showing the identification of Glucosylgalactosyl-HyK^{497}

Mouse col4a1 chain

^{474}G L R G P P$_{+16}$ G P Q G P P$_{+16}$ G E I G F P$_{+16}$ G Q P$_{+16}$ G A K$_{+340}$ G D R G L P$_{+16}$ G R D G L E^{509}

Q Exactive, HCD Fragmentation



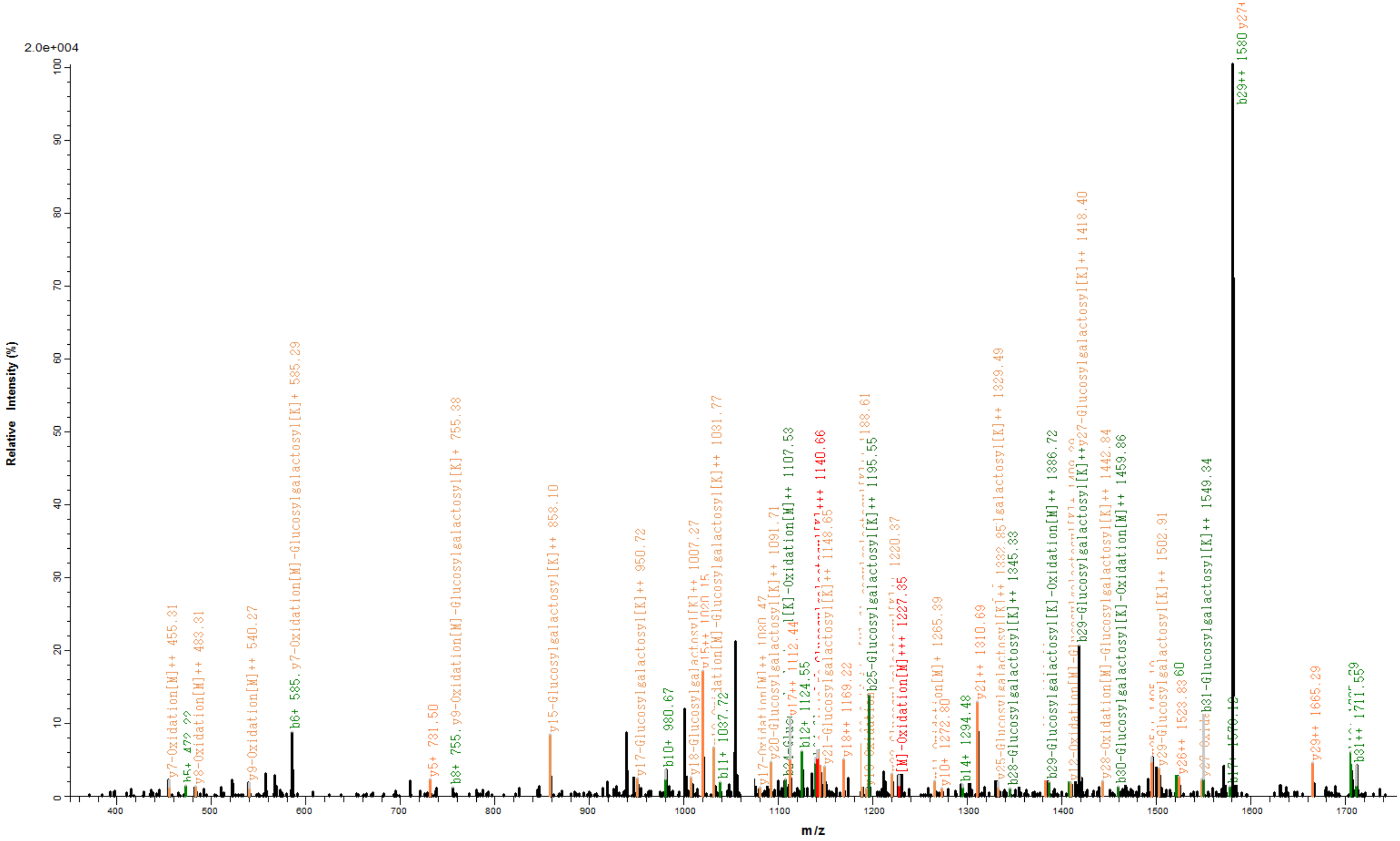
m/z= 1248.57,
charge=+3

PSM showing the identification of Glucosylgalactosyl-HyK^{527}

Mouse col4a1 chain

^{506}D G L E G L P$_{+16}$ G P Q G S P$_{+16}$ G L I G Q P$_{+16}$ G A K$_{+340}$ G E P$_{+16}$ G E I F F D M$_{+16}$ R^{538}

Q Exactive, HCD Fragmentation



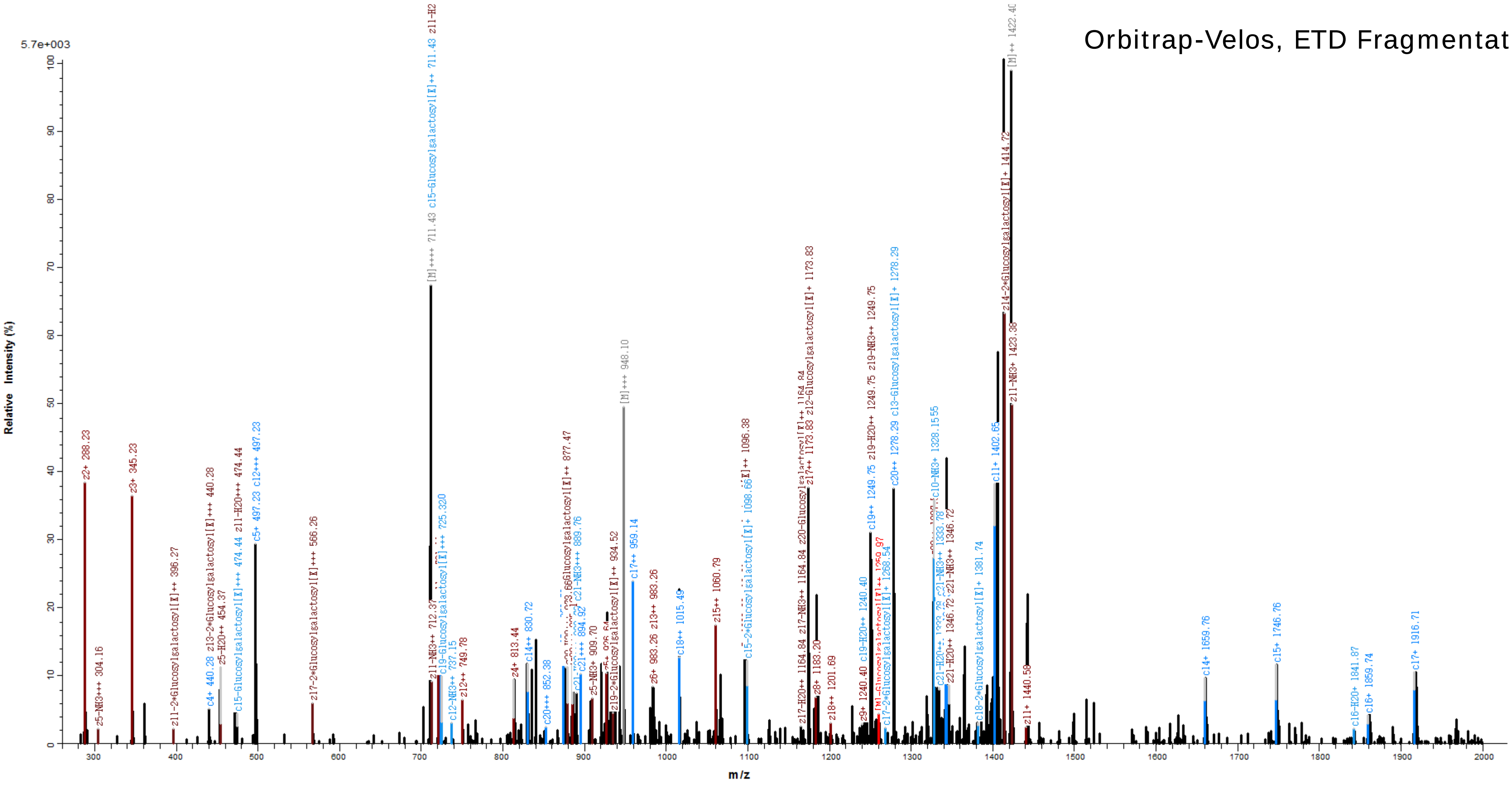
m/z= 1248.5738 ,
charge=+3

PSM showing the identification of Glucosylgalactosyl-HyK^{573,582}

Mouse col4a1 chain

^{564}D G H P$_{+16}$ G L P$_{+16}$ G P K$_{+340}$ G S P$_{+16}$ G S I G L K$_{+340}$ G E R^{585}

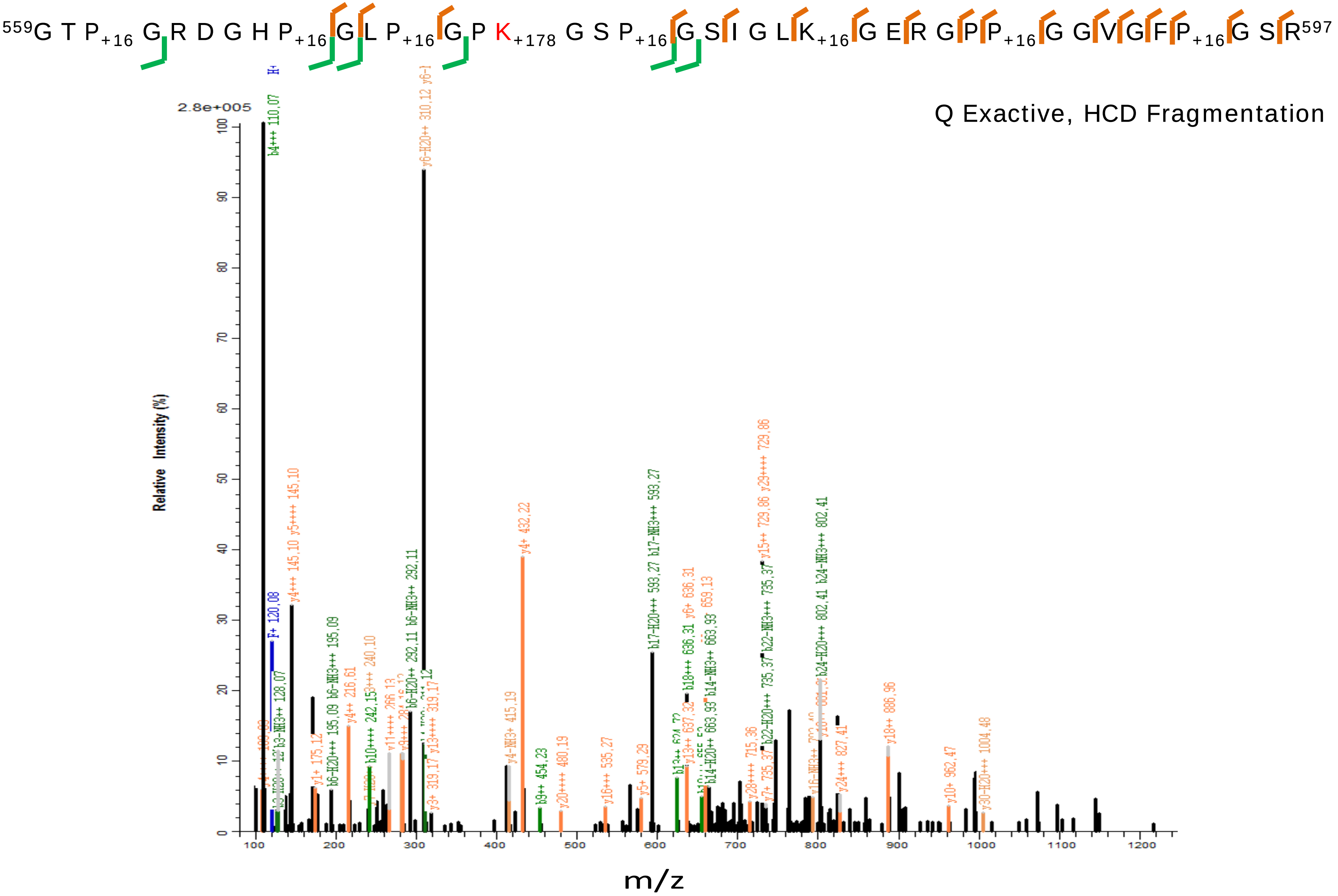
Orbitrap-Velos, ETD Fragmentation



m/z= 655.5620 charge=+4

PSM showing the identification of Galactosyl-HyK^{573}

Mouse col4a1 chain



^{559}G T P$_{+16}$ G R D G H P$_{+16}$ G L P$_{+16}$ G P K$_{+178}$ G S P$_{+16}$ G S I G L K$_{+16}$ G E R G P P$_{+16}$ G G V G F P$_{+16}$ G S R^{597}

Q Exactive, HCD Fragmentation

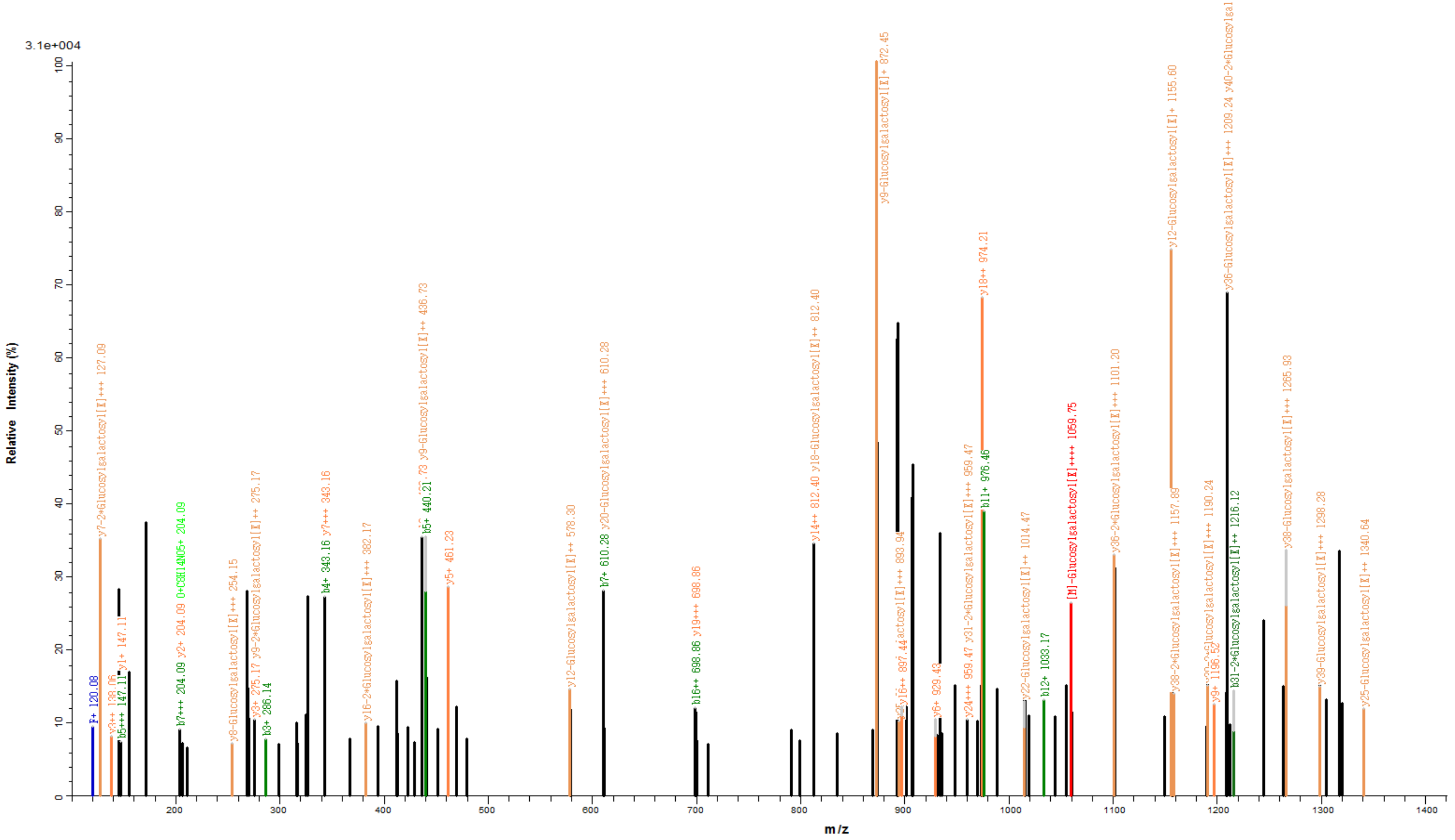
m/z= 1140.2943 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{617,635}

Mouse col4a1 chain

^{598}G D I G P P$_{+16}$ G P P$_{+16}$ G V G P I G P V G E K$_{+340}$ G Q A G F P G G P$_{+16}$ G S P$_{+16}$ G L P$_{+16}$ G P K$_{+340}$ G E A G K^{640}

Q Exactive, HCD Fragmentation



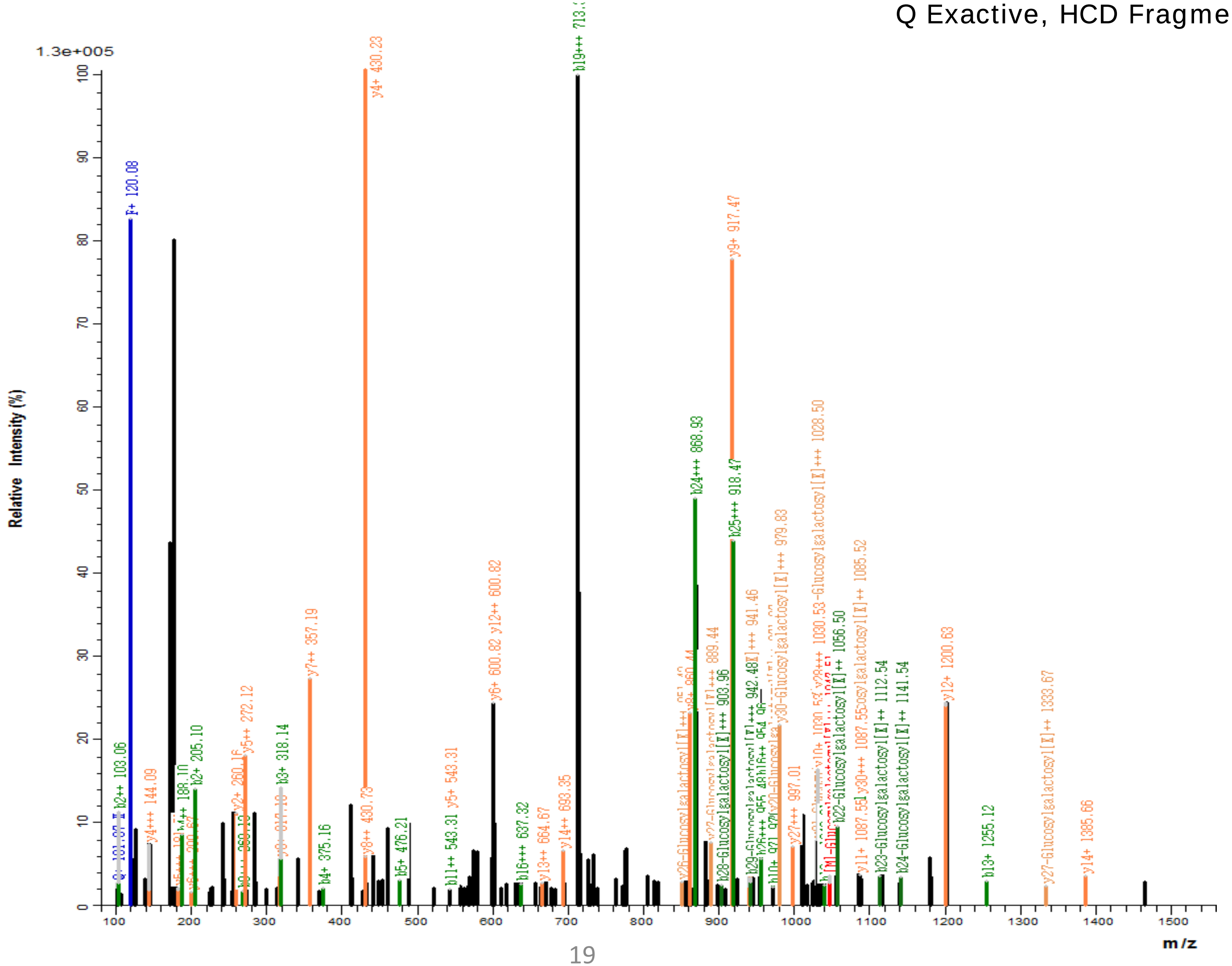
m/z= 866.9266charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{681}

Mouse col4a1 chain

^{667}G F P$_{+16}$ G T P$_{+16}$ G R P$_{+16}$ G I P$_{+16}$ G E K$_{+340}$ G A V G Q P$_{+16}$ G I G F P$_{+16}$ G L P$_{+16}$ G P K$_{+16}$ ^{698}

Q Exactive, HCD Fragmentation



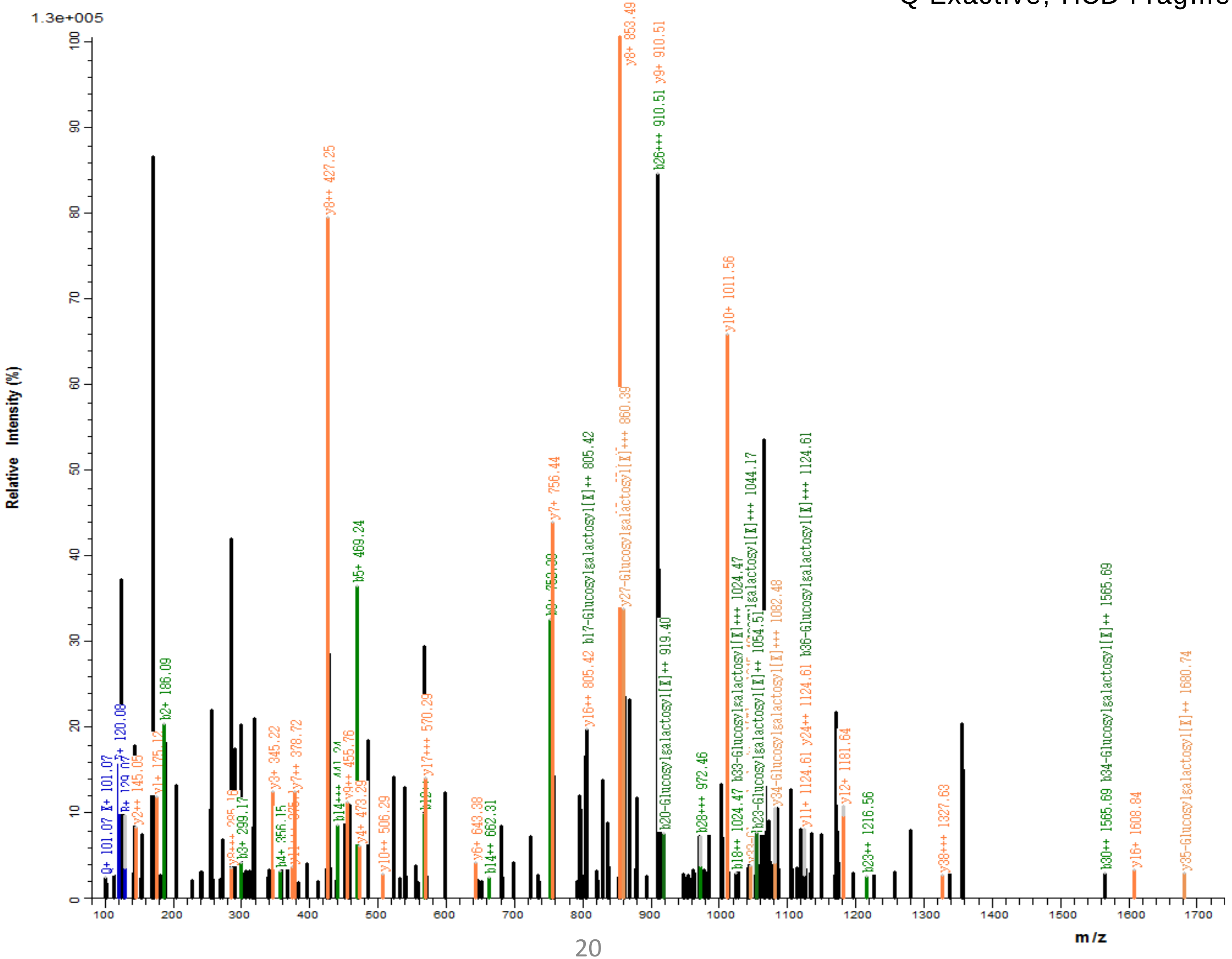
m/z= 1010.4997 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{757}

Mouse col4a1 chain

^{743}G Q P$_{+16}$ G L P$_{+16}$ G I P$_{+16}$ G T P$_{+16}$ G E K$_{+340}$ G S I G G P$_{+16}$ G V P$_{+16}$ G E Q G L T G P P$_{+16}$ G L Q G I R^{781}

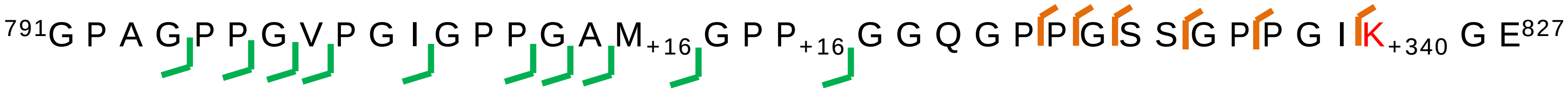
Q Exactive, HCD Fragmentation



m/z= 502.5 charge=+7

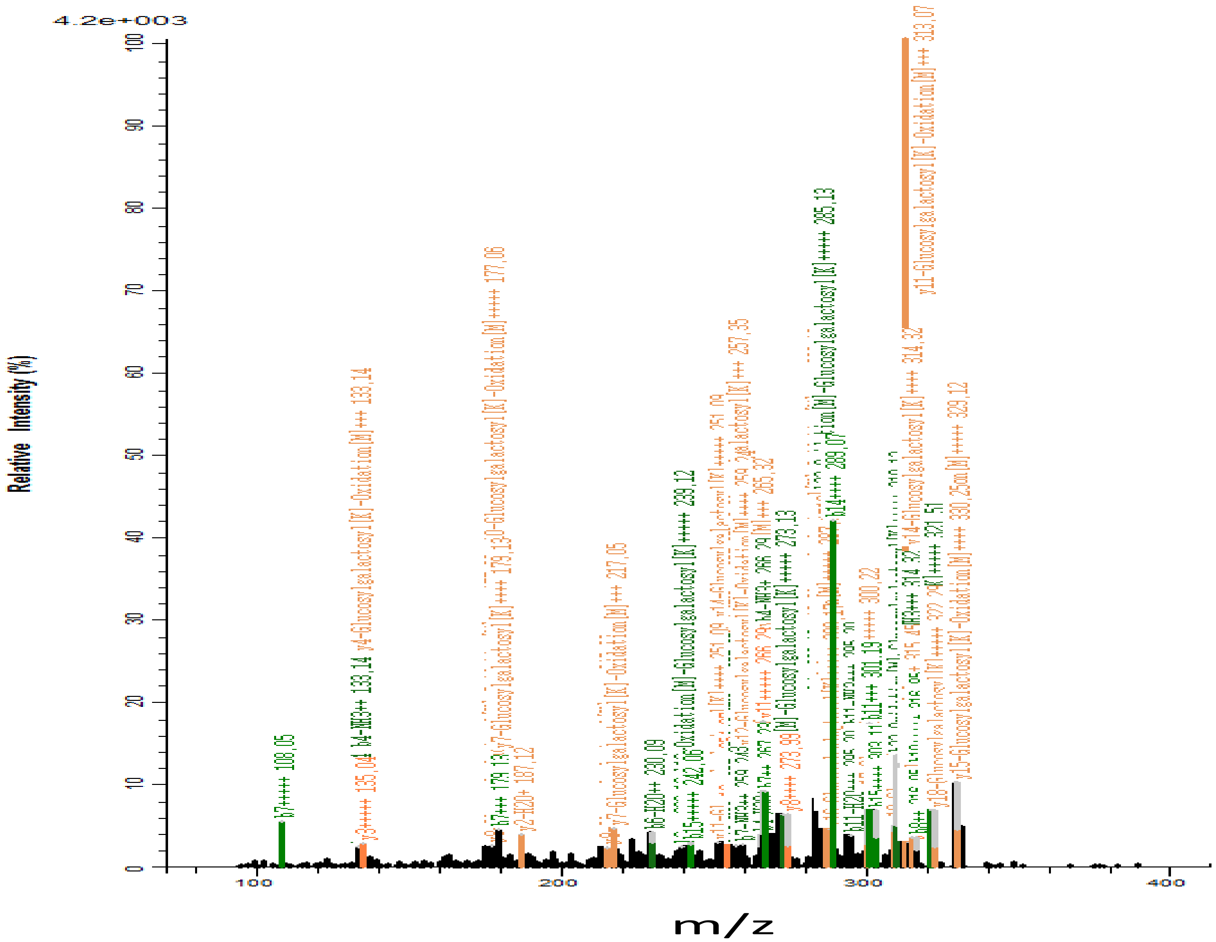
PSM showing the identification of Glucosylgalactosyl-HyK^{825}

Mouse col4a1 chain



^{791}G P A G P P G V P G I G P P G A M$_{+16}$ G P P$_{+16}$ G G Q G P P G S S G P P G I K$_{+340}$ G E^{827}

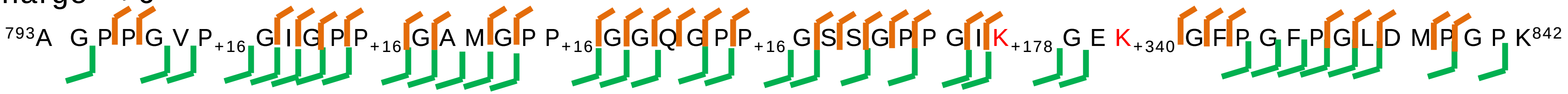
Q Exactive, HCD Fragmentation



m/z= 849.7198,
charge=+6

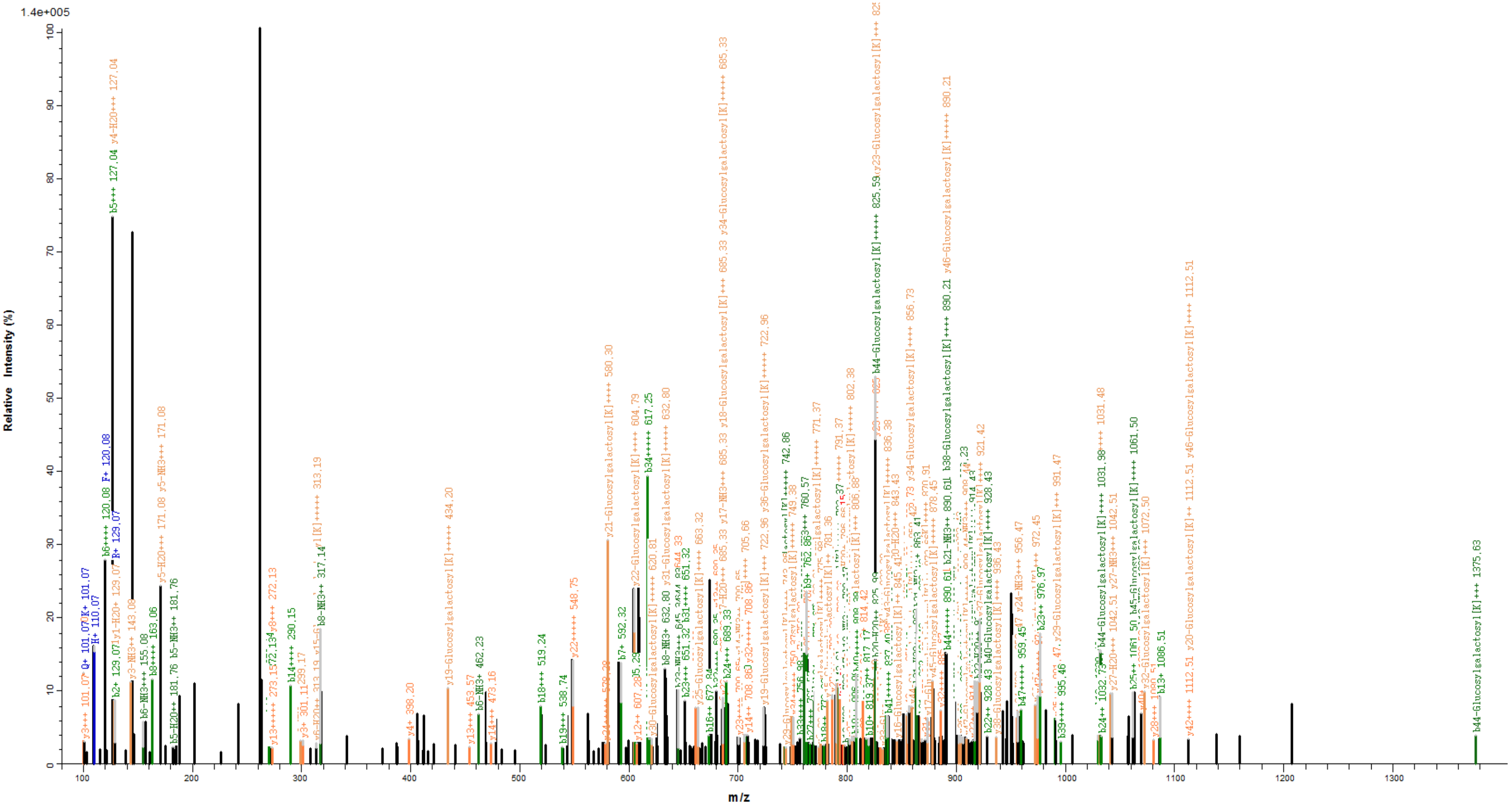
PSM showing the identification of Galactosyl-HyK^{825} & Glucosylgalactosyl-HyK^{828}

Mouse col4a1 chain



^{793}A G P P G V P$_{+16}$ G I G P P$_{+16}$ G A M G P P$_{+16}$ G G Q G P P$_{+16}$ G S S G P P G I K$_{+178}$ G E K$_{+340}$ G F P G F P G L D M P G P K^{842}

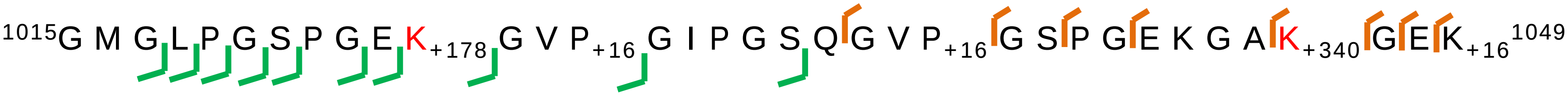
Q Exactive, HCD Fragmentation



m/z= 627.9313 charge=+6

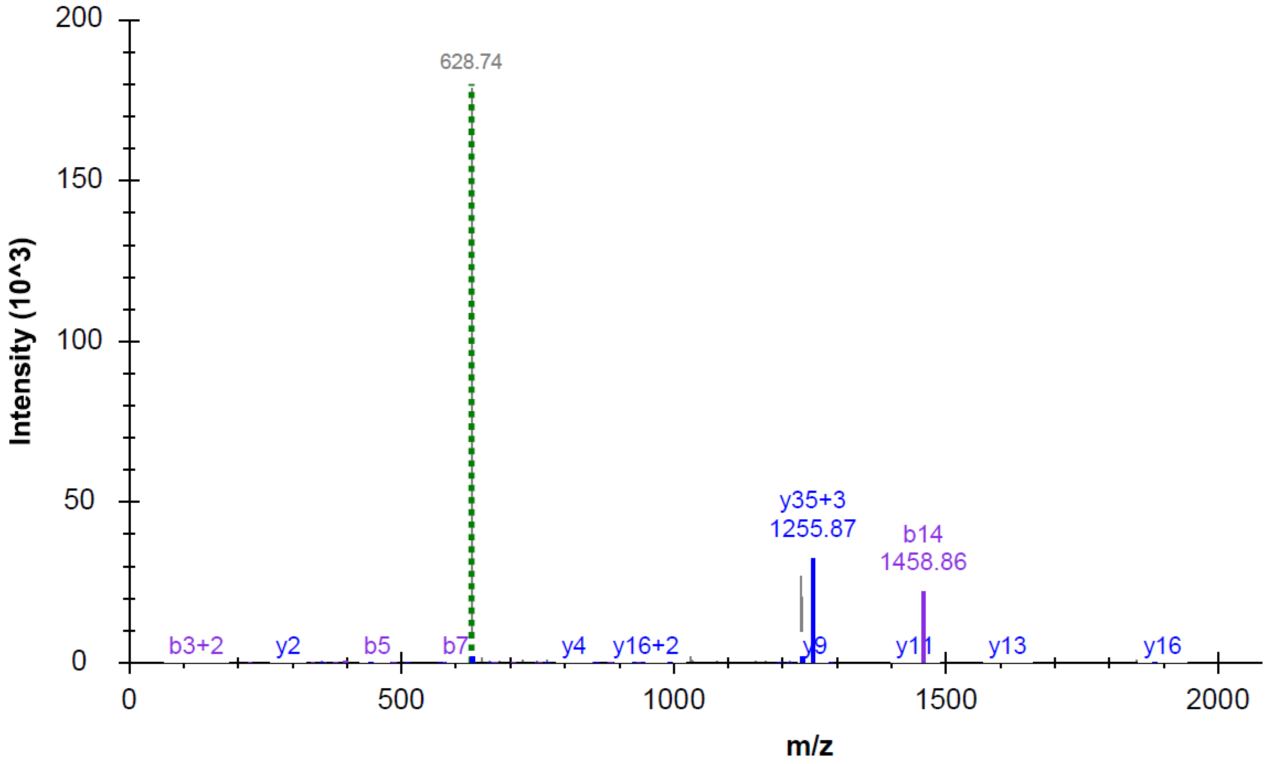
PSM showing the identification of Glucosylgalactosyl-HyK^{1025}

Mouse col4a1 chain



^{1015}G M G L P G S P G E K$_{+178}$ G V P$_{+16}$ G I P G S Q G V P$_{+16}$ G S P G E K G A K$_{+340}$ G E K$_{+16}$^{1049}

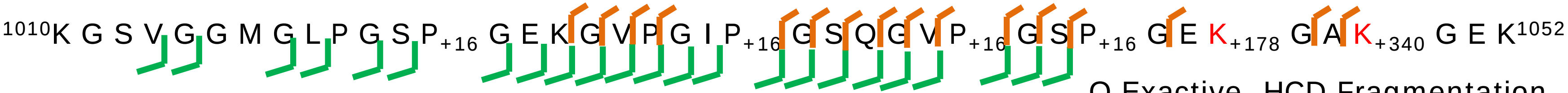
Orbitrap-Velos, CID Fragmentation



m/z= 702.6838
charge=+6

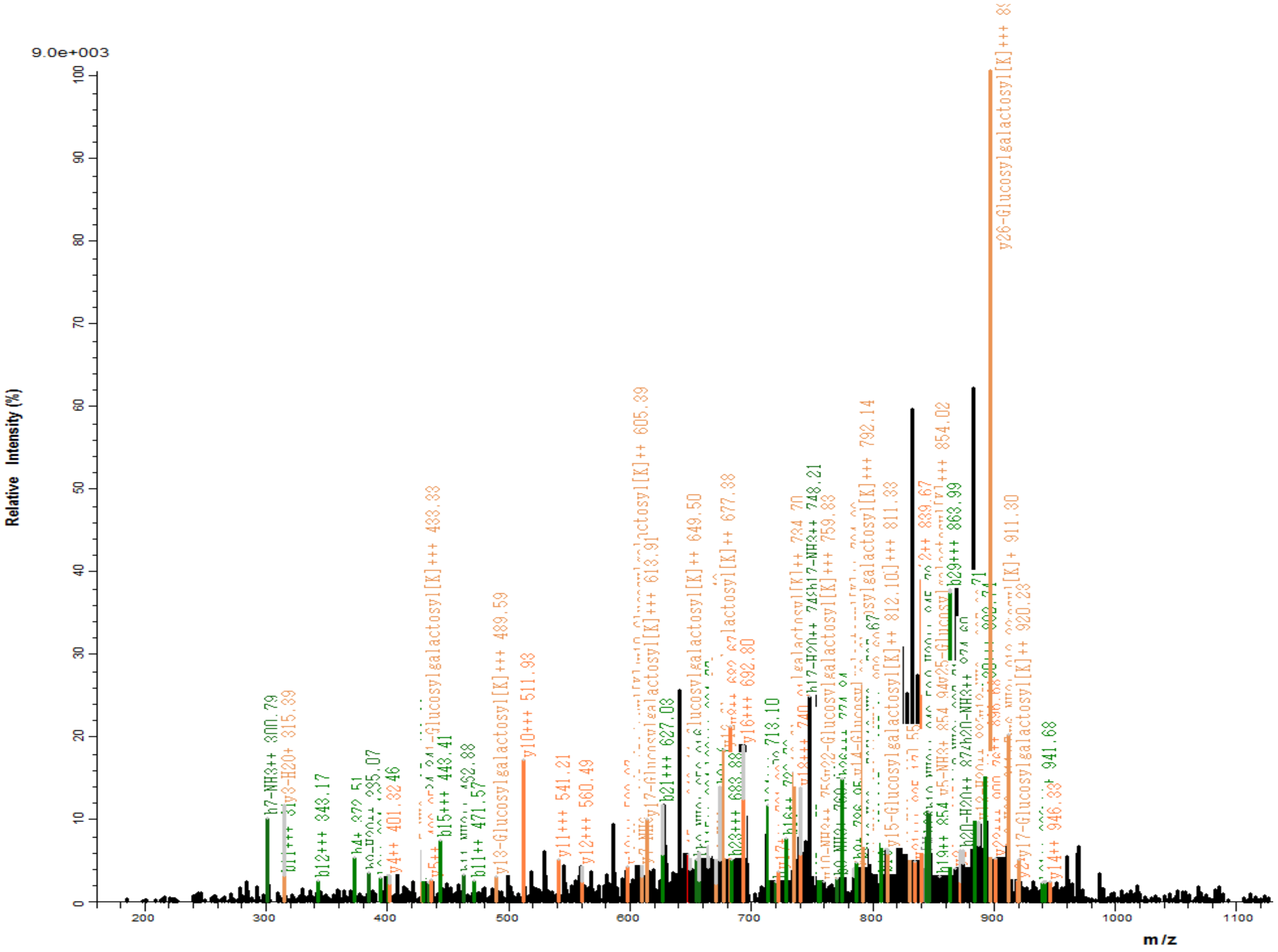
PSM showing the identification of Galactosyl-HyK^{1046} & Glucosylgalactosyl-HyK^{1049}

Mouse col4a1 chain



^{1010}K G S V G G M G L P G S P$_{+16}$ G E K G V P G I P$_{+16}$ G S Q G V P$_{+16}$ G S P$_{+16}$ G E K$_{+178}$ G A K$_{+340}$ G E K^{1052}

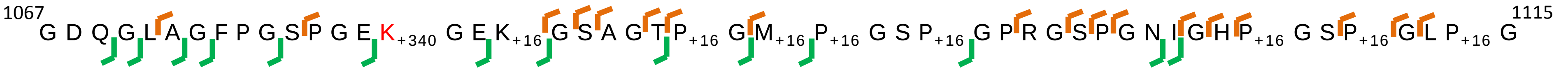
Q Exactive, HCD Fragmentation



m/z= 791.01 charge=+6

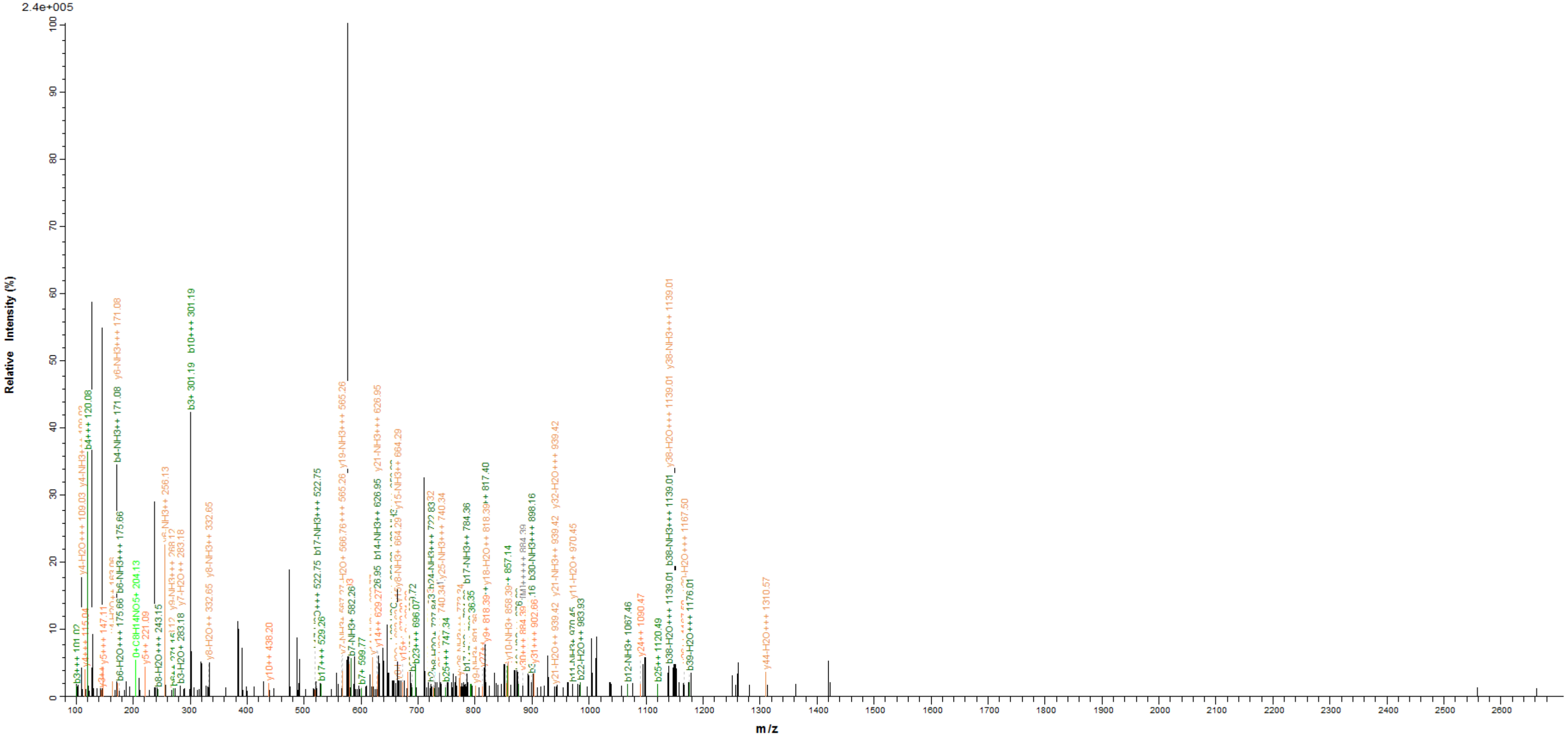
PSM showing the identification of Galactosyl-HyK[1081]

Mouse col4a1 chain



1067 G D Q G L A G F P G S P G E K_{+340} G E K_{+16} G S A G T P_{+16} G M_{+16} P_{+16} G S P_{+16} G P R G S P G N I G H P_{+16} G S P_{+16} G L P_{+16} G 1115

Q Exactive, HCD Fragmentation



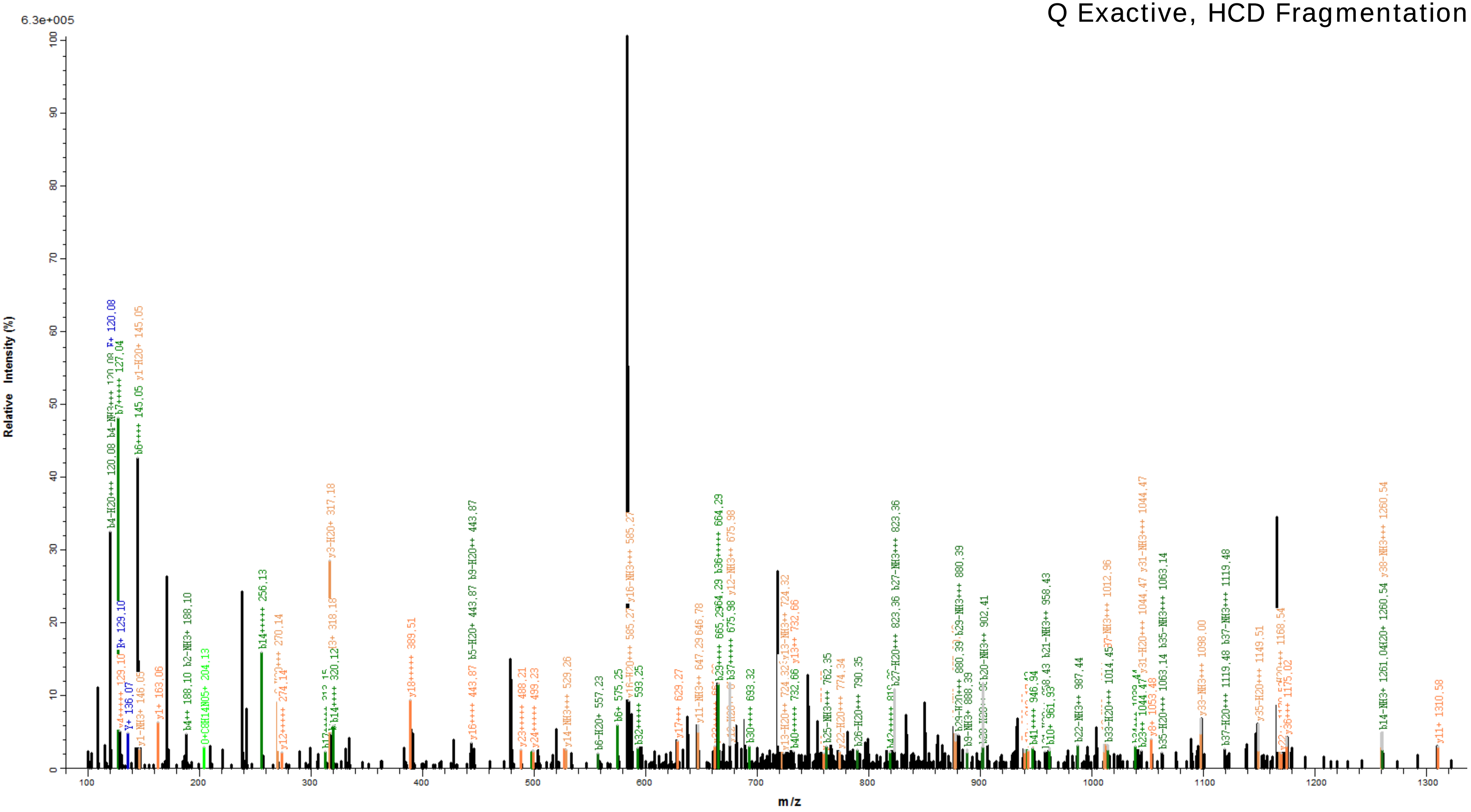
m/z= 793.6757 charge=+6

PSM showing the identification of Galactosyl-HyK^{1117}

Mouse col4a1 chain

^{1073}G F P$_{+16}$ G S P$_{+16}$ G E K$_{+16}$ G E K$_{+16}$ G S A G T P G M P$_{+16}$ G S P$_{+16}$ G P R G S P G N I G H P G S P$_{+16}$ G L P$_{+16}$ G E K$_{+178}$ G D K$_{+16}$ ^{1120}

Q Exactive, HCD Fragmentation



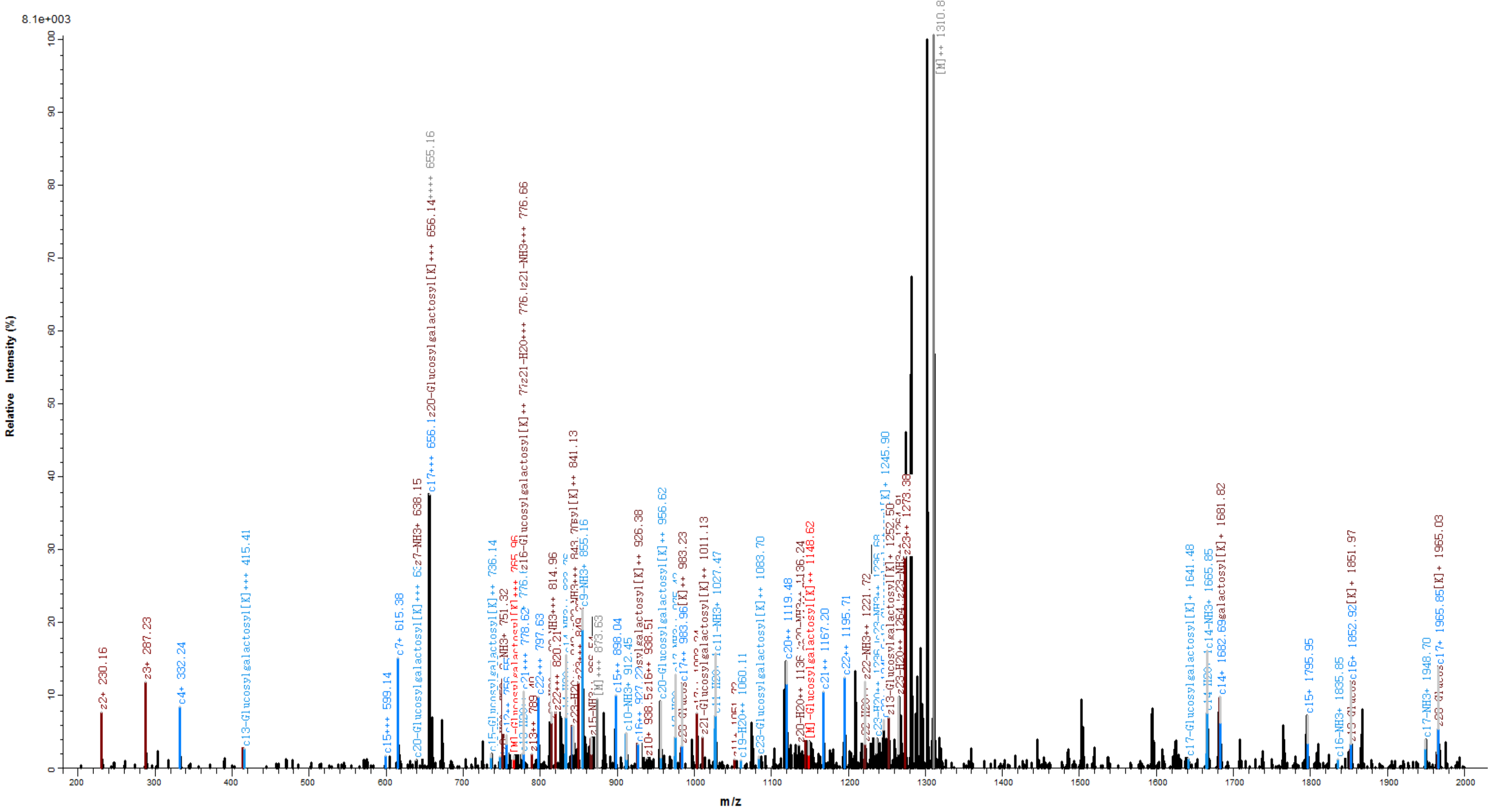
m/z= 655.5620 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{1120}

Mouse col4a1 chain

^{1109}G S P$_{+16}$ G L P$_{+16}$ G E K G D K$_{+340}$ G L P$_{+16}$ G L D G V P G V K^{1132}

Orbitrap-Velos, ETD Fragmentation



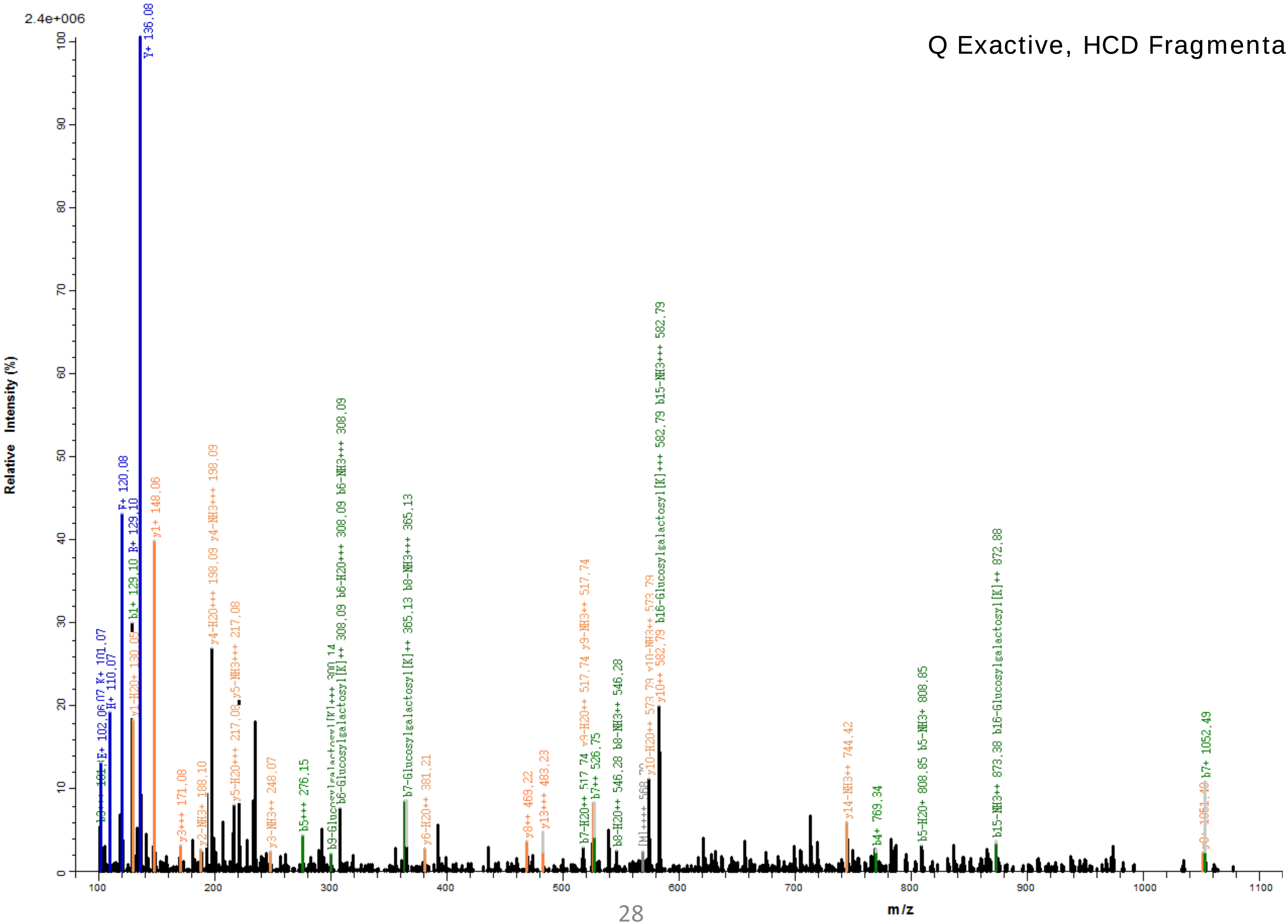
m/z= 569.0098 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK[1120,1132]

Mouse col4a1 chain

^{1117}K G D K$_{+340}$ G L P$_{+16}$ G L D G V P$_{+16}$ G V K$_{+178}$ G E^{1134}

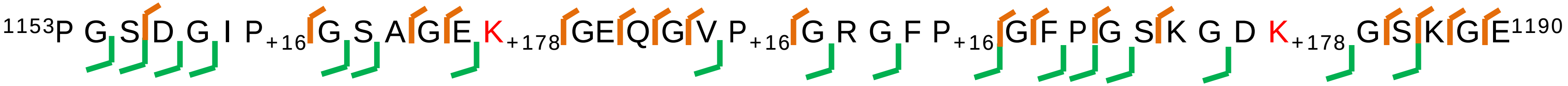
Q Exactive, HCD Fragmentation



m/z= 997.968 charge=+4

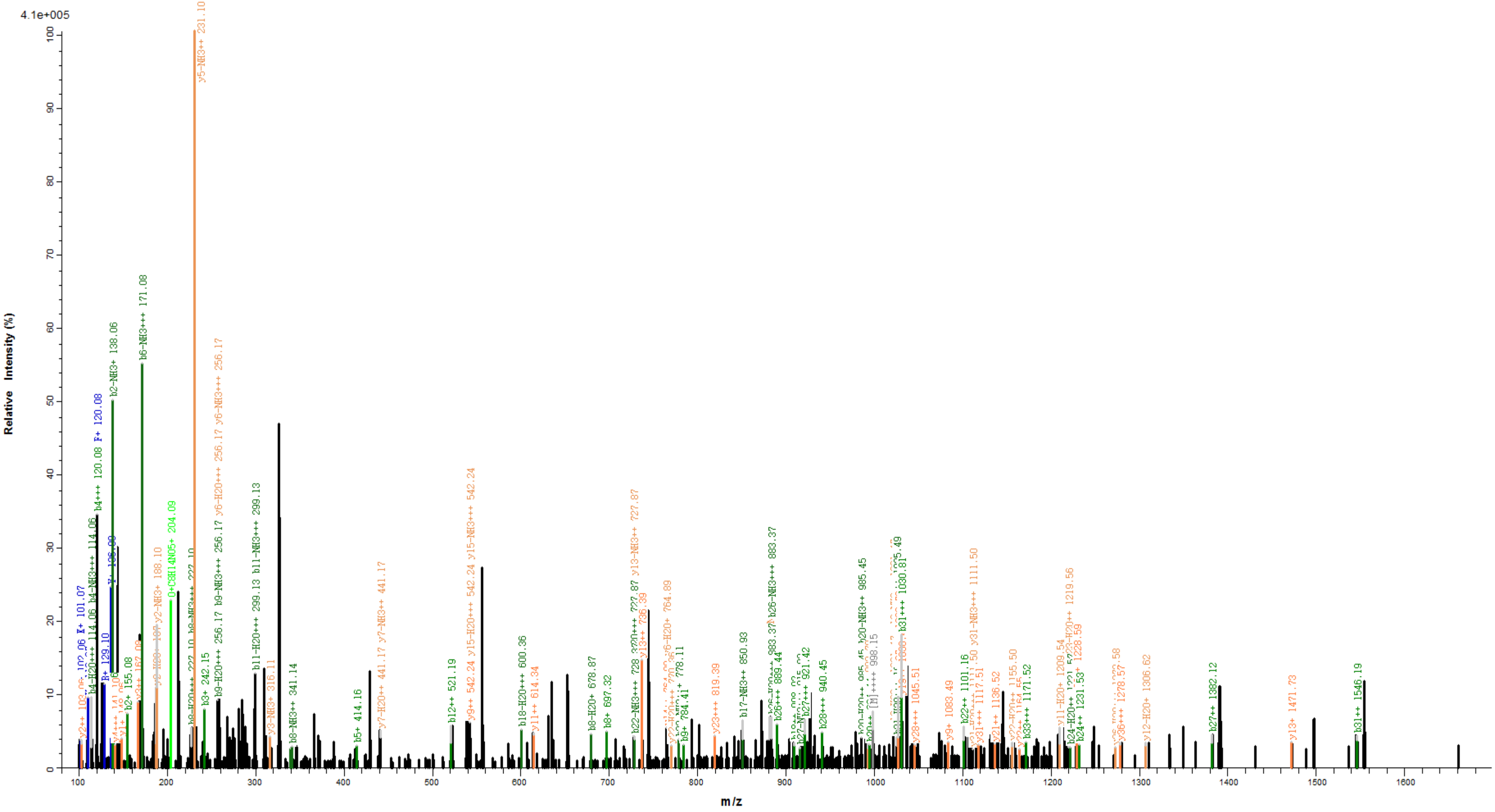
PSM showing the identification of Glucosylgalactosyl-HyK^{1165,1185}

Mouse col4a1 chain



^{1153}P G S D G I P$_{+16}$ G S A G E K$_{+178}$ G E Q G V P$_{+16}$ G R G F P$_{+16}$ G F P G S K G D K$_{+178}$ G S K G E^{1190}

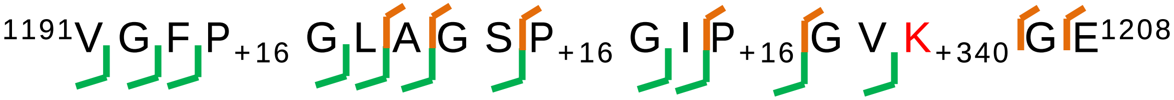
Q Exactive, HCD Fragmentation



m/z= 1014.0001 charge=+2

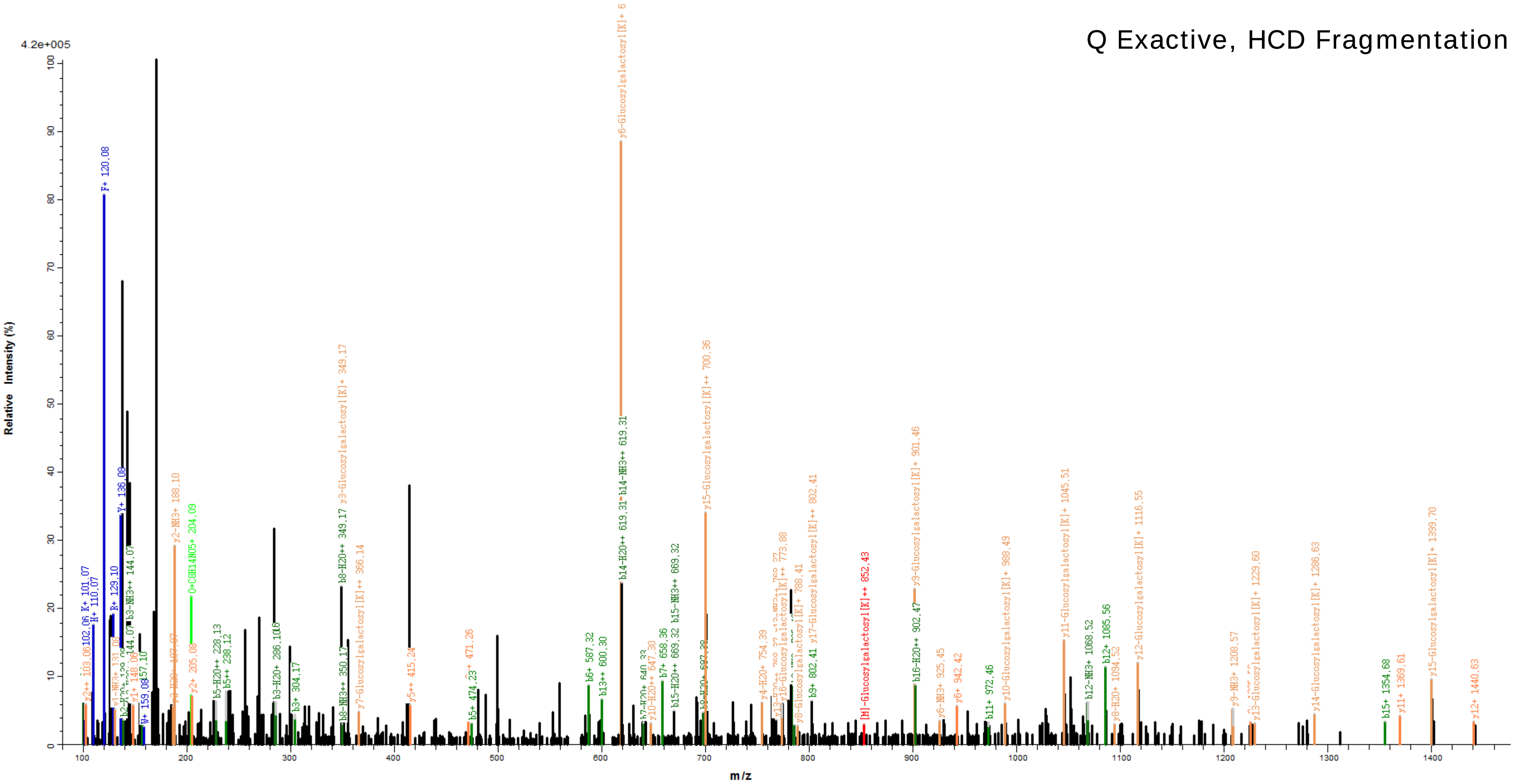
PSM showing the identification of Glucosylgalactosyl-HyK^{1206}

Mouse col4a1 chain



^{1191}V G F P$_{+16}$ G L A G S P$_{+16}$ G I P$_{+16}$ G V K$_{+340}$ G E^{1208}

Q Exactive, HCD Fragmentation



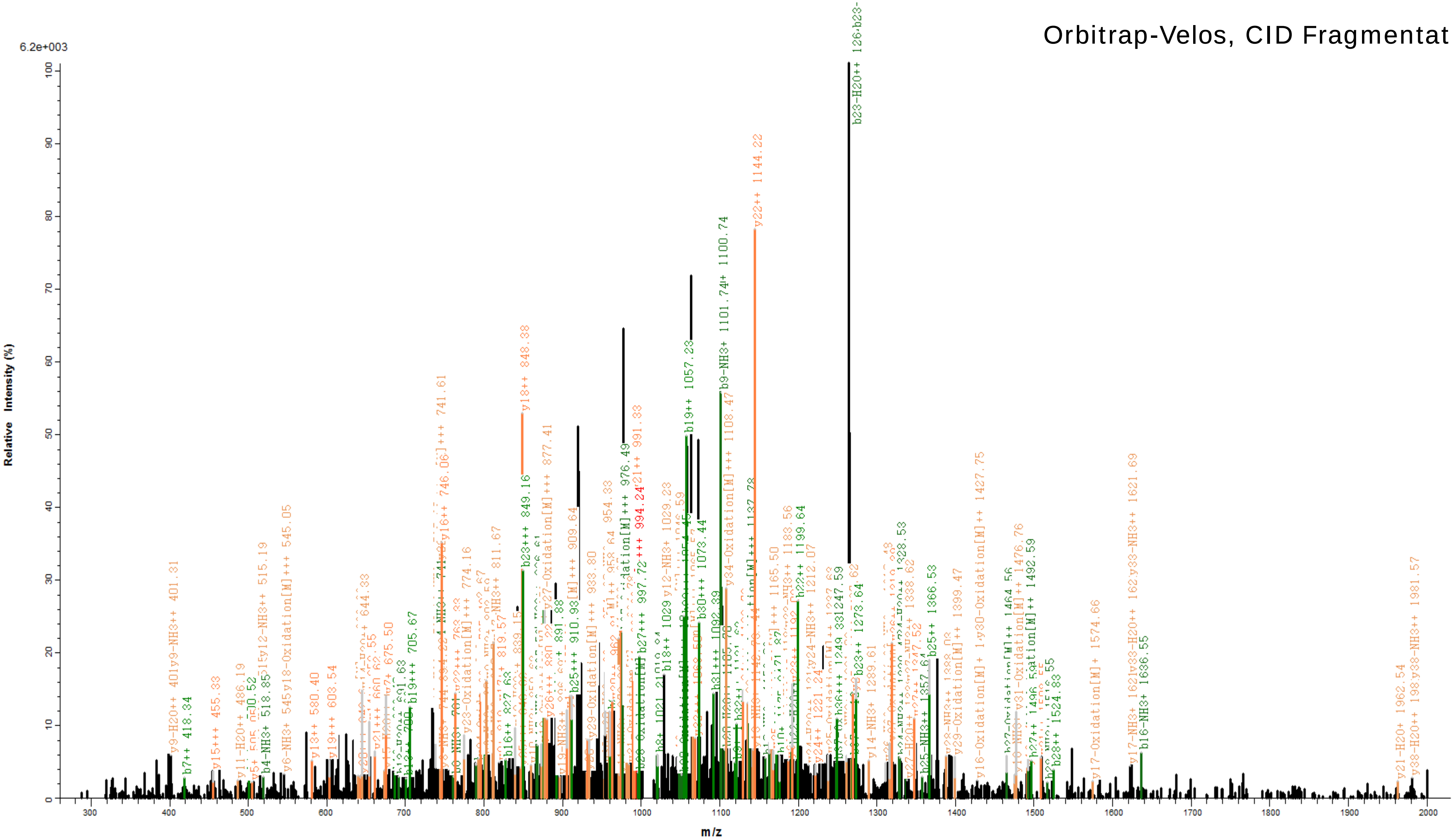
m/z= 655.5620 charge=+4

PSM showing the identification of Galactosyl-HyK^{1268,1283}

Mouse col4a1 chain

^{1266}G D K$_{+178}$ G N Q G W P G A P G V P G P K$_{+178}$ G D P$_{+16}$ G F Q G M$_{+16}$ P$_{+16}$ G I G G S P$_{+16}$ G I T G S K^{1304}

Orbitrap-Velos, CID Fragmentation



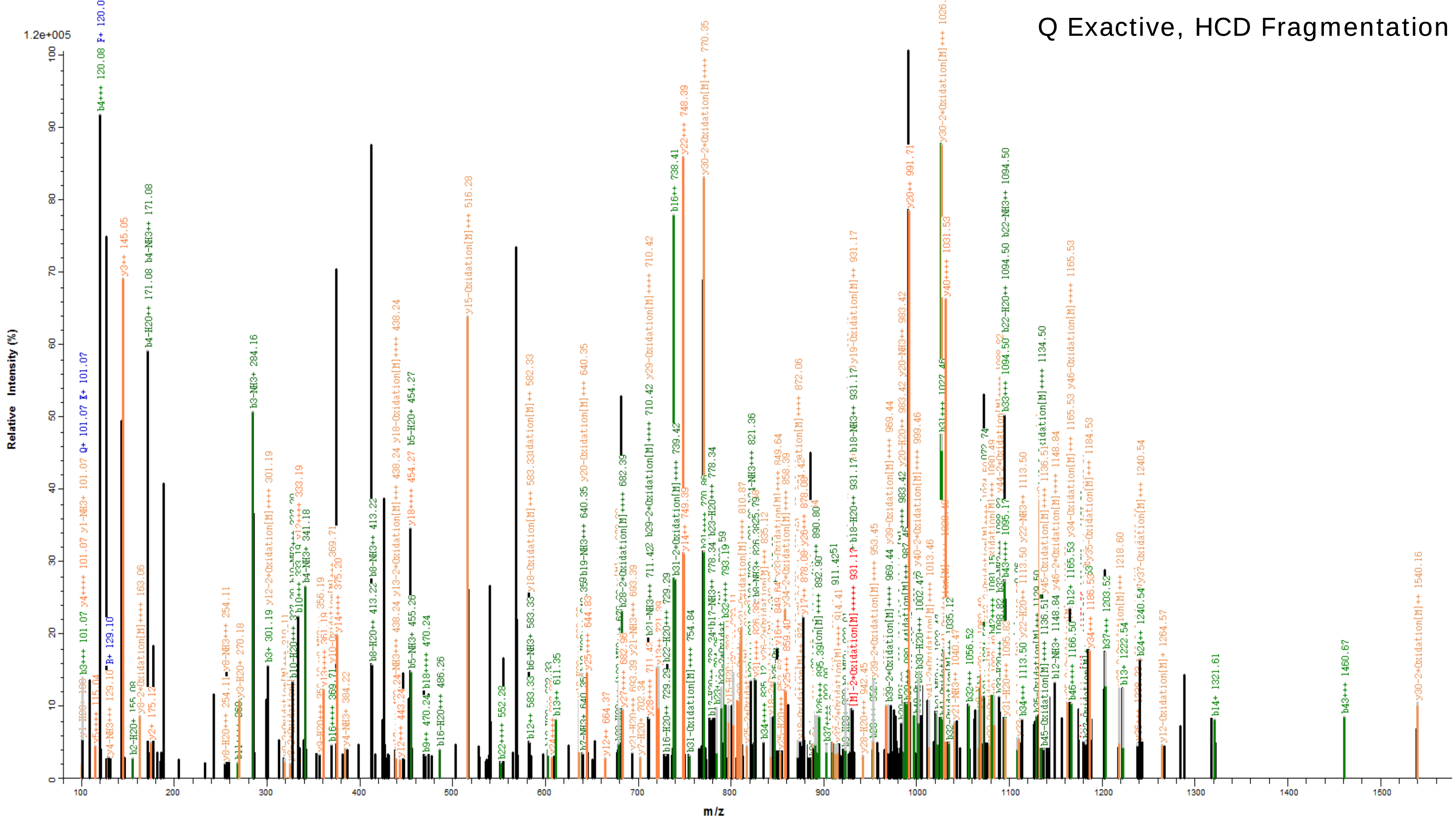
m/z= 956.6473,
charge=+5

PSM showing the identification of Galactosyl-HyK^{1283,1304}

Mouse col4a1 chain

^{1266}G D K G N Q G W P G A P G V P G P K$_{+178}$ G D P G F Q G M$_{+16}$ P$_{+16}$ G I G G S P$_{+16}$ G I T G S K$_{+178}$ G D M$_{+16}$ G L P$_{+16}$ G V^{1312}

Q Exactive, HCD Fragmentation



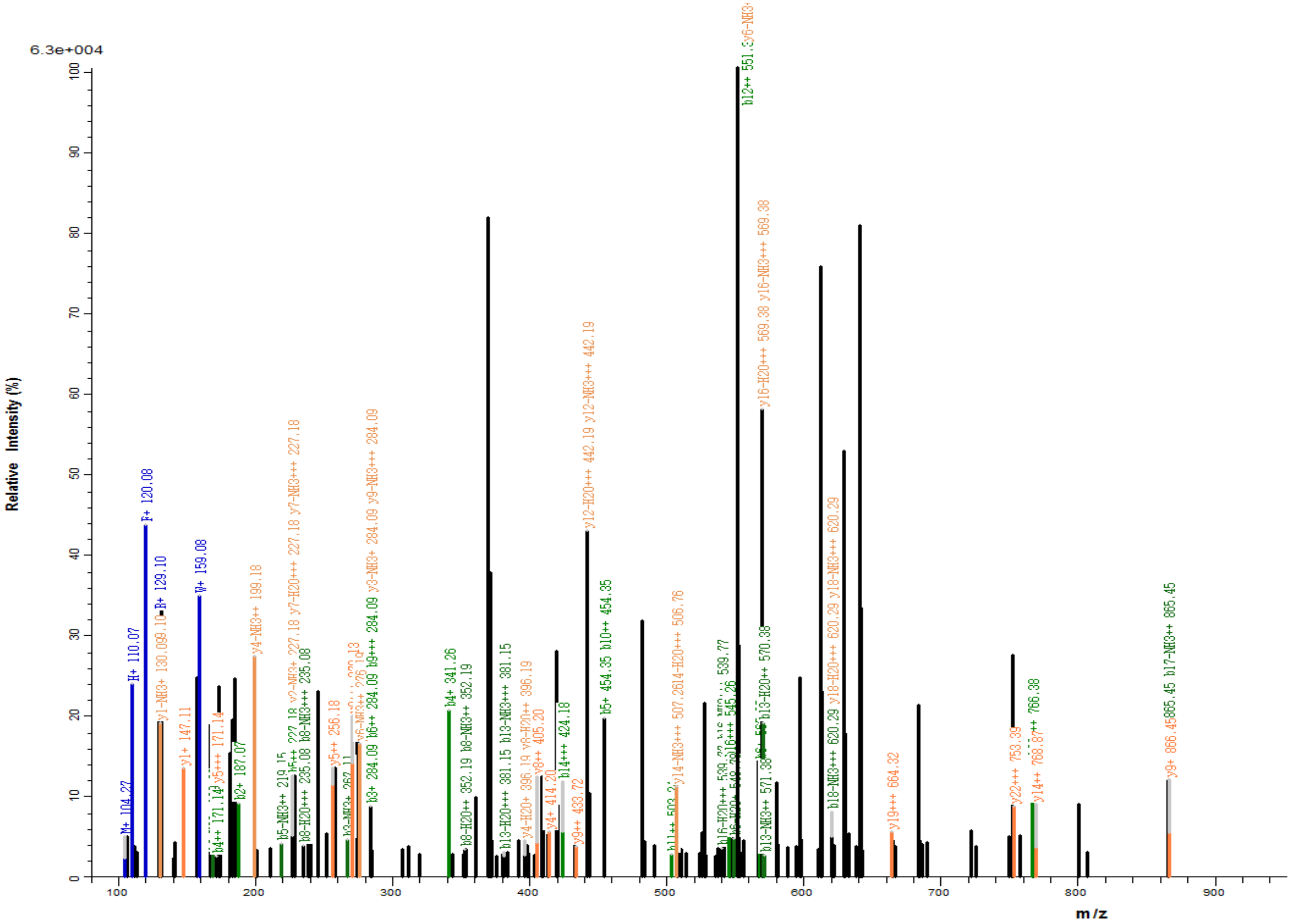
m/z= 489.4524,
charge=+4

PSM showing the identification of Galactosyl-HyK^{1371}

Mouse col4a1 chain

^{1357}G E P G L P$_{+16}$ G P E G P P G L K$_{+178}$ G L Q G P P$_{+16}$ G P K^{1380}

Q Exactive, HCD Fragmentation



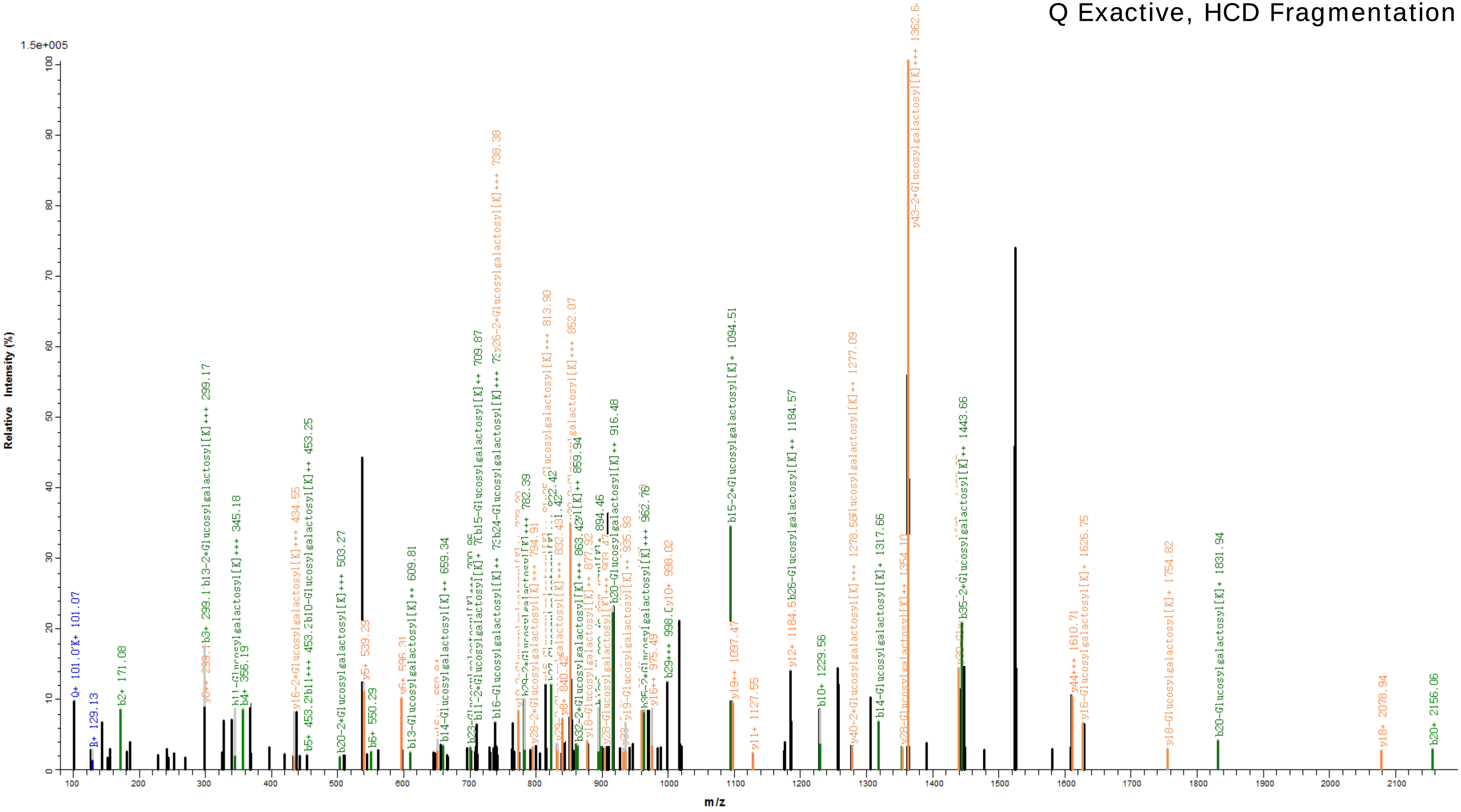
m/z= 1301.3679 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{1380,1407}

Mouse col4a1 chain

^{1372}G L Q G P P G P K$_{+340}$ G Q Q G V T G S V G L P$_{+16}$ G P P$_{+16}$ G V P$_{+16}$ G F D G A P$_{+16}$ G Q K$_{+340}$ G E T G P F G P P$_{+16}$ G P R^{1419}

Q Exactive, HCD Fragmentation



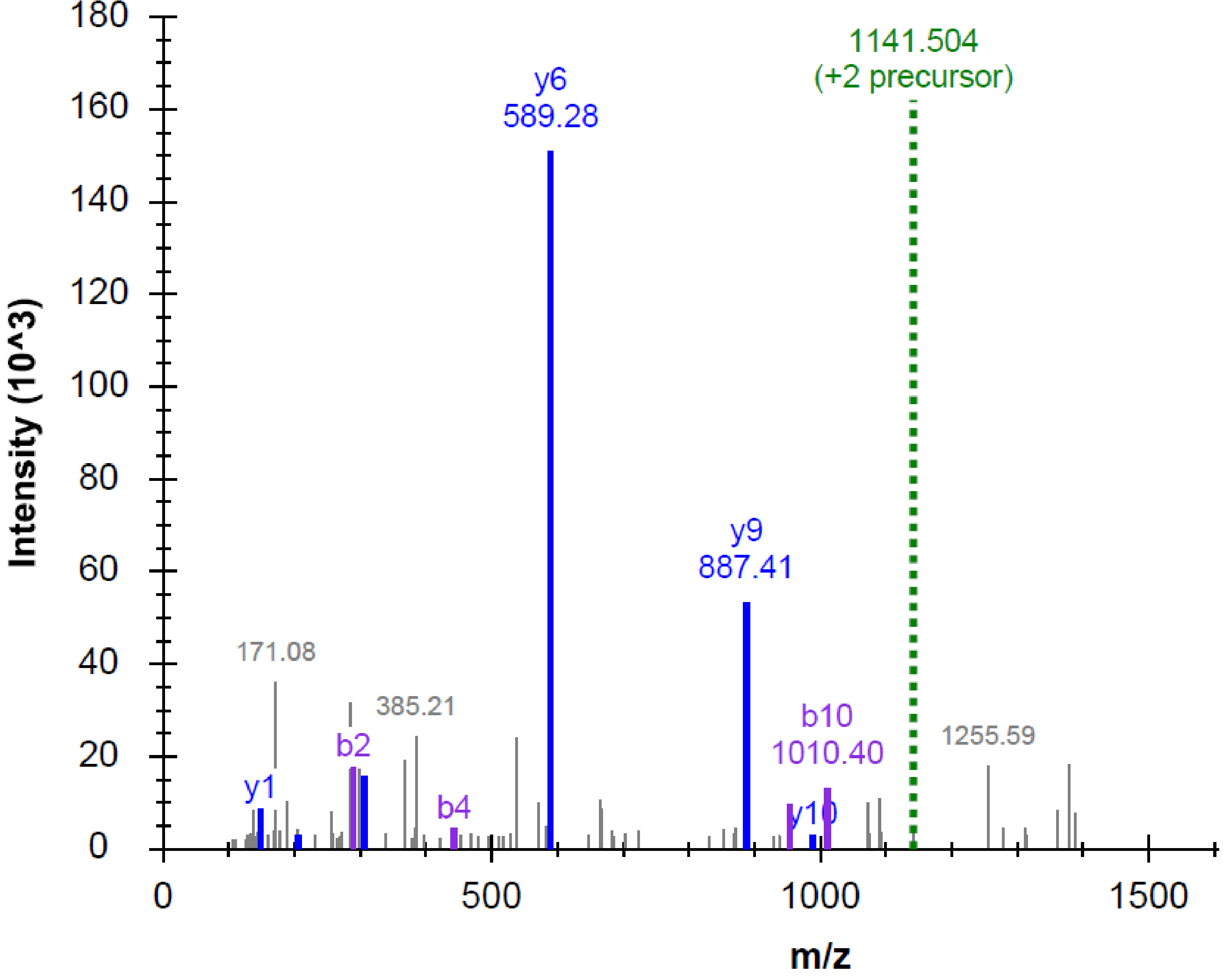
m/z= 1141.5 (+2)

PSM showing the identification of 3-HyP^{444}

Mouse col4a1 chain

$^{434}C_{+57}$ Q P G P_{+16} P G D Q G P$_{+16}$ P_{+16} G T P_{+16} G Q P_{+16} G L T G E^{456}

Q Exactive, HCD Fragmentation



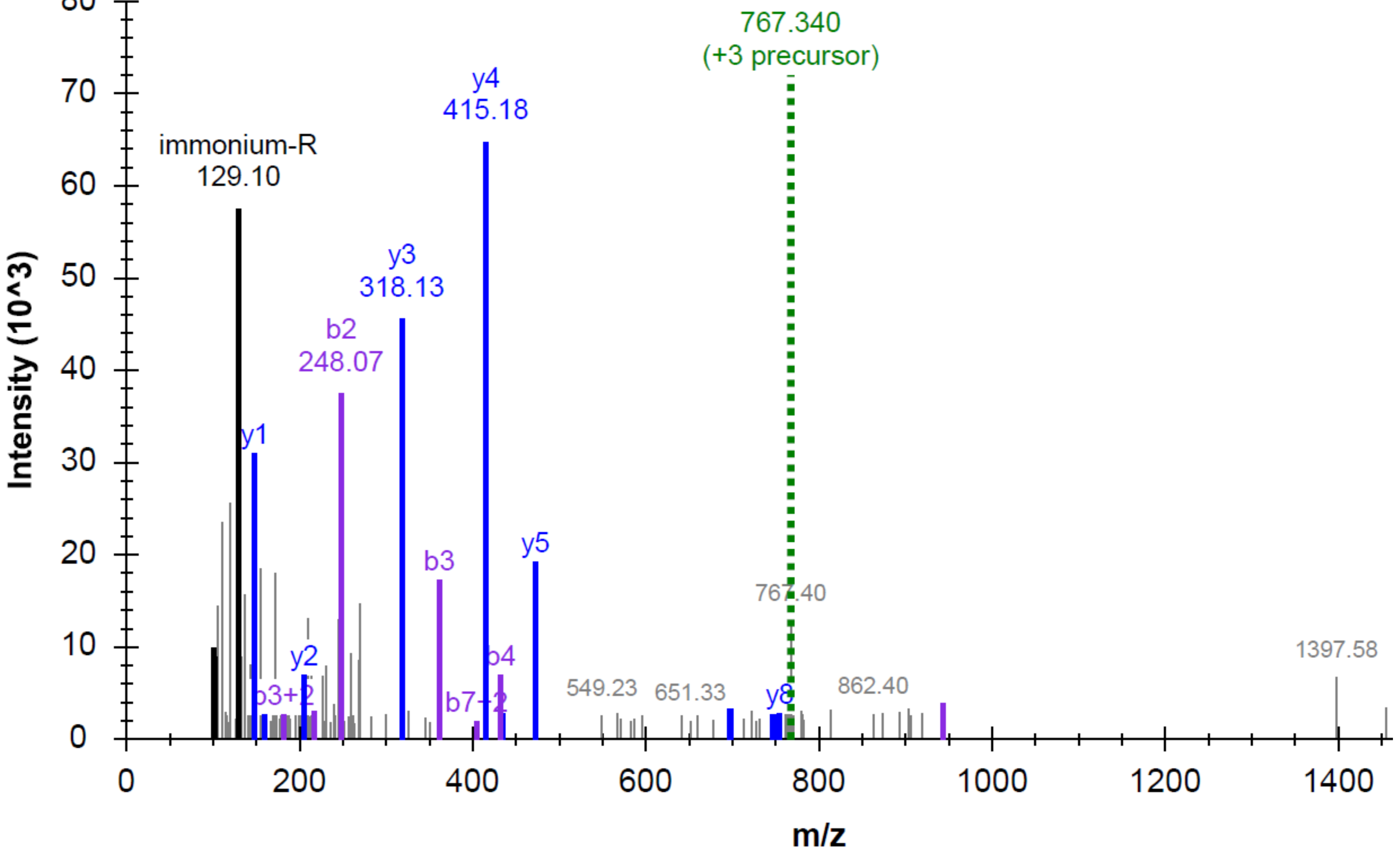
m/z= 767.34 (+3)

PSM showing the identification of 3-HyP^{478}

Mouse col4a1 chain

^{466}S C$_{+57}$ L A C$_{+57}$ D T E G L R G P$_{+16}$ P$_{+16}$ G P Q G P P$_{+16}$ G E^{487}

Q Exactive, HCD Fragmentation



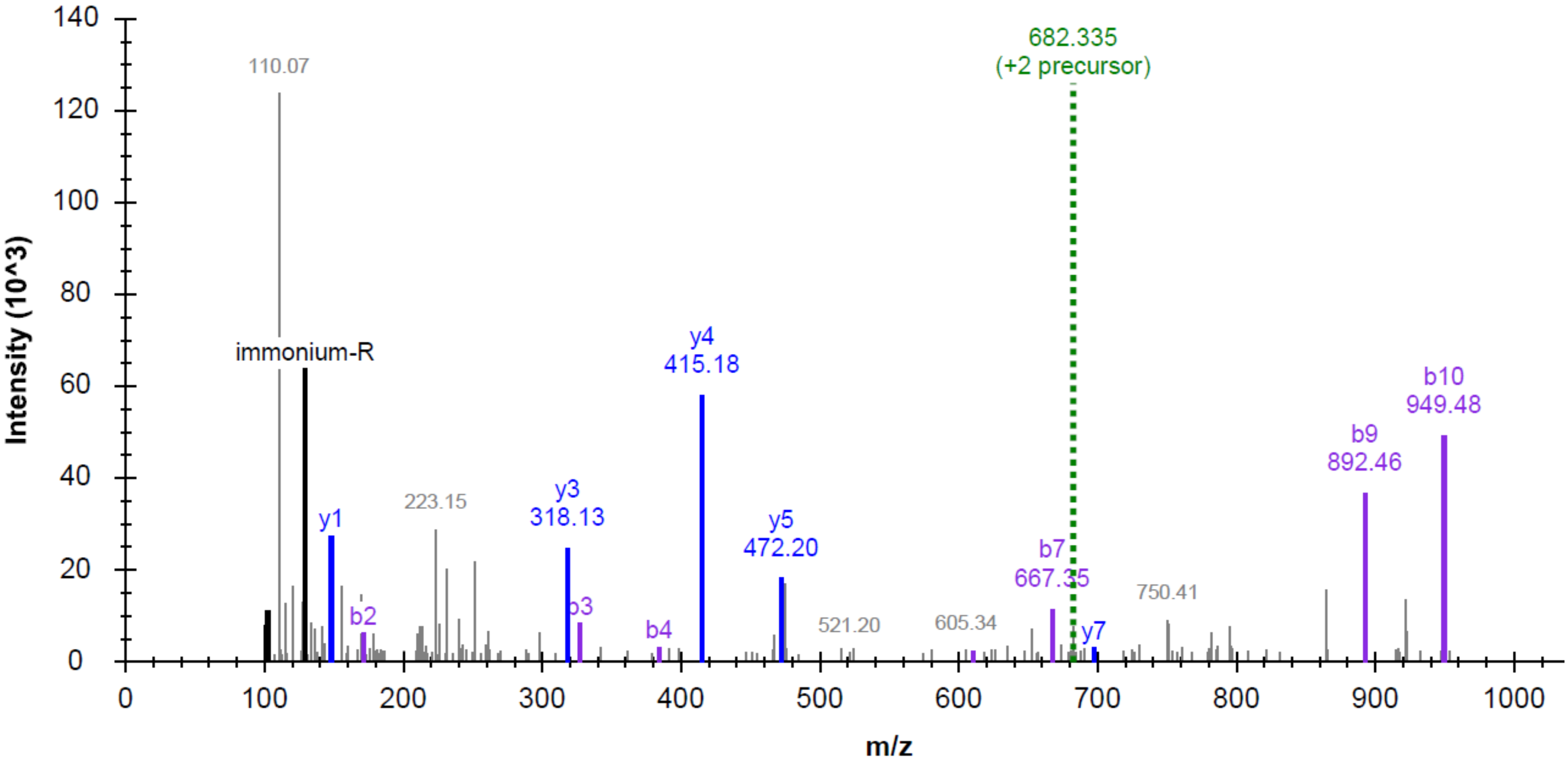
m/z= 682.33 (+2)

PSM showing the identification of HyP^{481}

Mouse col4a1 chain

^{474}G L R G P$_{+16}$ P G P$_{+16}$ Q G P P$_{+16}$ G E^{487}

Orbitrap-Velos, CID Fragmentation



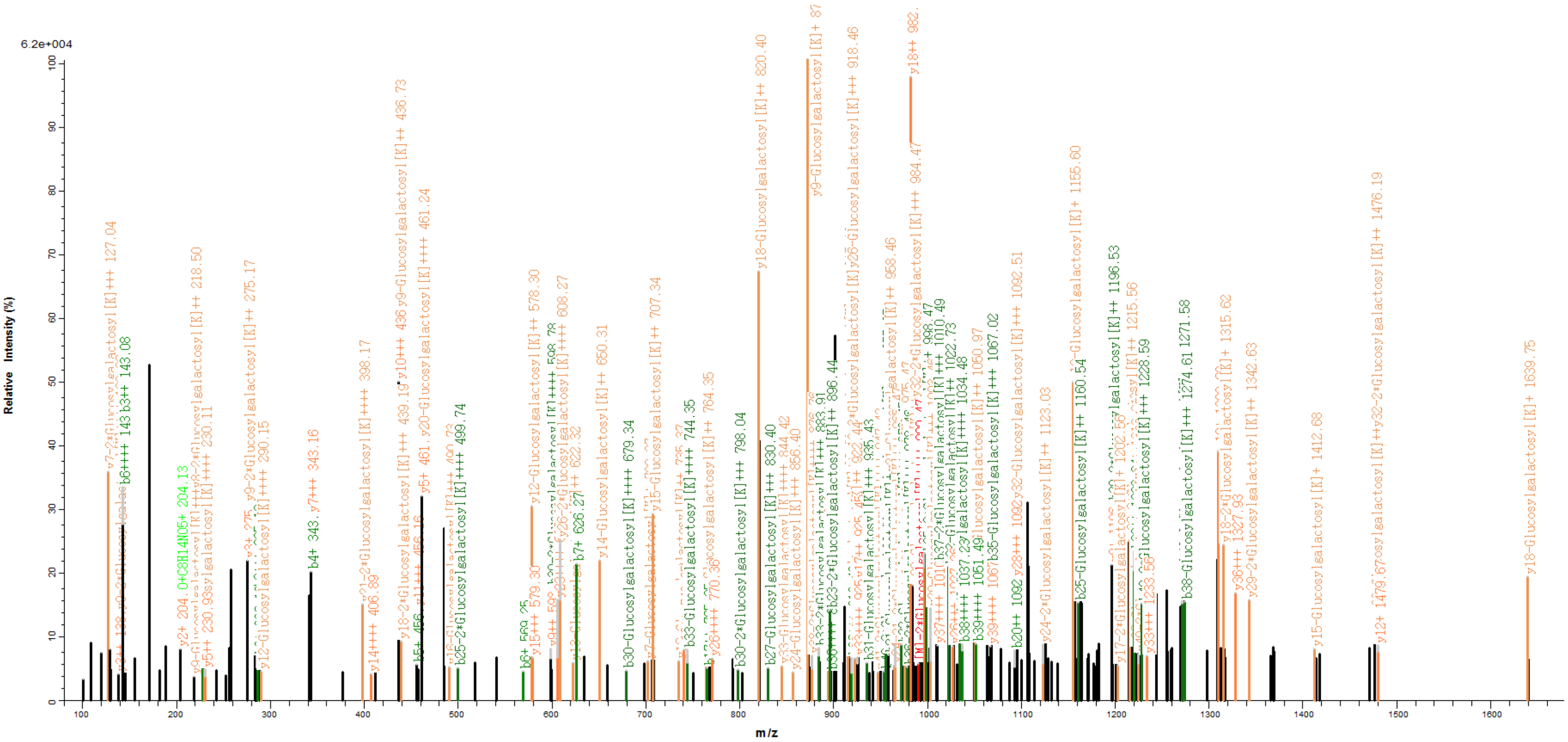
m/z= 1148.28 (+4)

PSM showing the identification of 3-HyP^{602,605}

Mouse col4a1 chain

^{598}G D I G P$_{+16}$ P$_{+16}$ G P$_{+16}$ P$_{+16}$ G V G P I G P V G E K$_{+340}$ G Q A G F P$_{+16}$ G G P$_{+16}$ G S P$_{+16}$ G L P G P K$_{+340}$ G E A G K^{640}

Q Exactive, HCD Fragmentation

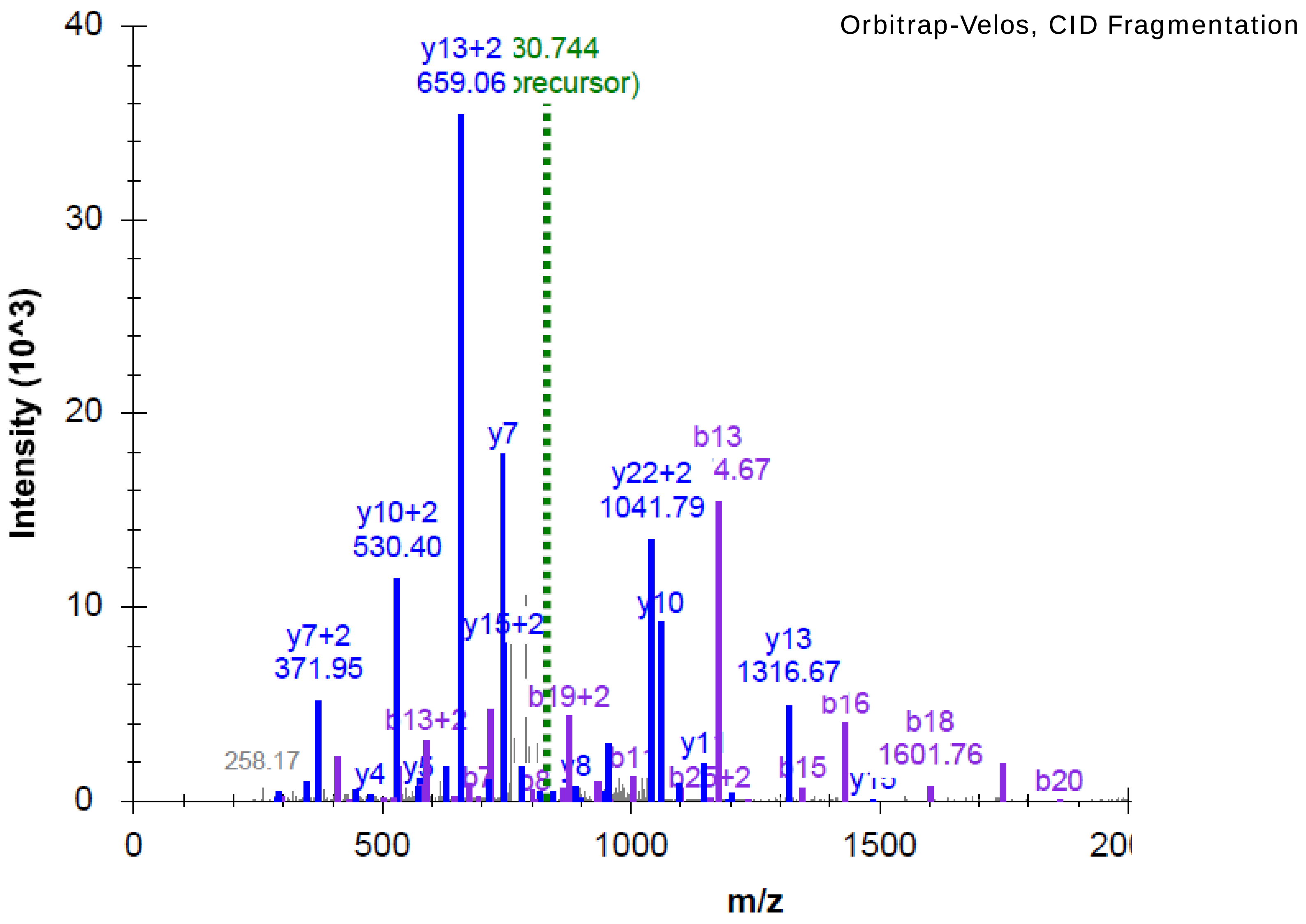


m/z= 830.744 (+3)

PSM showing the identification of 3-HyP^{647}

Mouse col4a1 chain

^{641}V V P L P$_{+16}$ G P$_{+16}$ P G A A G L P$_{+16}$ G S P$_{+16}$ G F P$_{+16}$ G P Q G D R^{666}



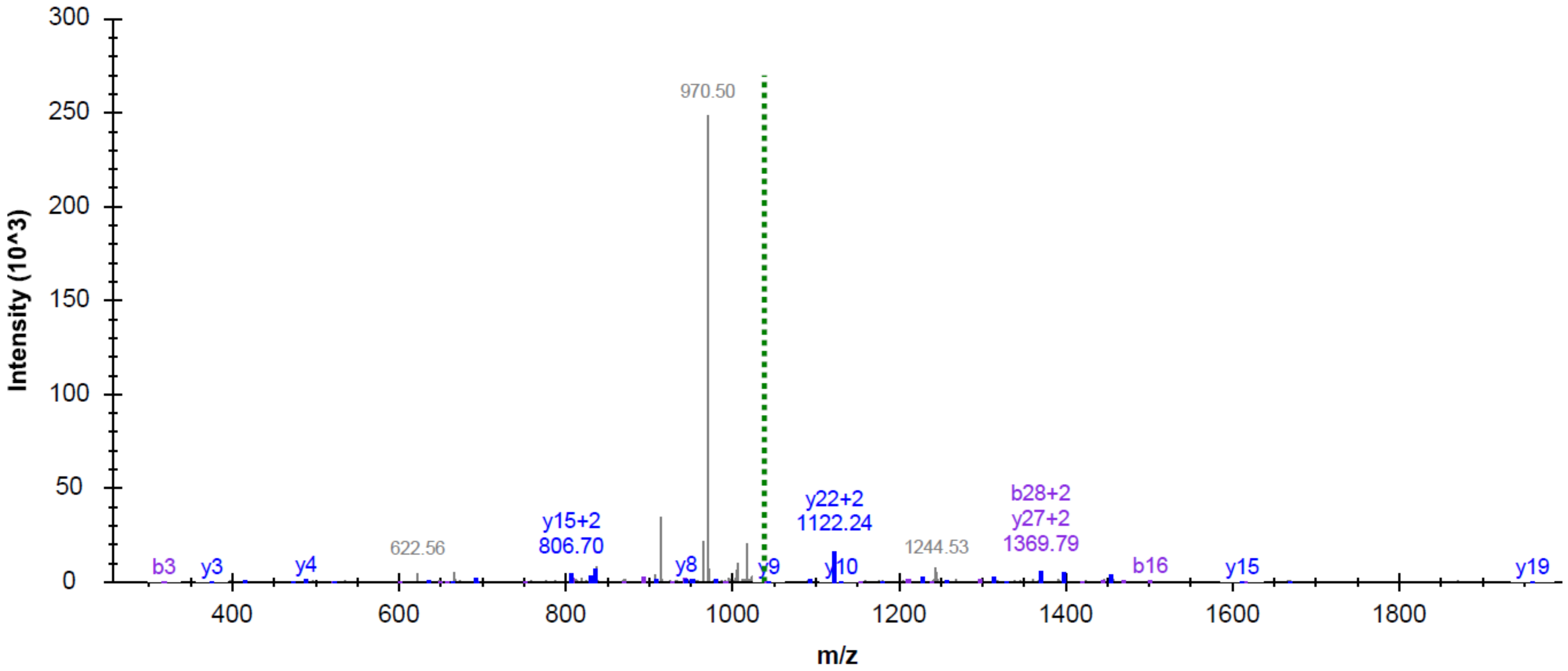
m/z= 1038.1483 (+3)

PSM showing the identification of 3-HyP^{1424}, 3-HyP^{1436}

Mouse col4a1 chain

^{1420}GFP$_{+16}$ G P$_{+16}$ P$_{+16}$ G P D G L P$_{+16}$ G S M$_{+16}$ G P$_{+16}$ P$_{+16}$ G T P S V D H G F L V T R^{1450}

Q Exactive, HCD Fragmentation



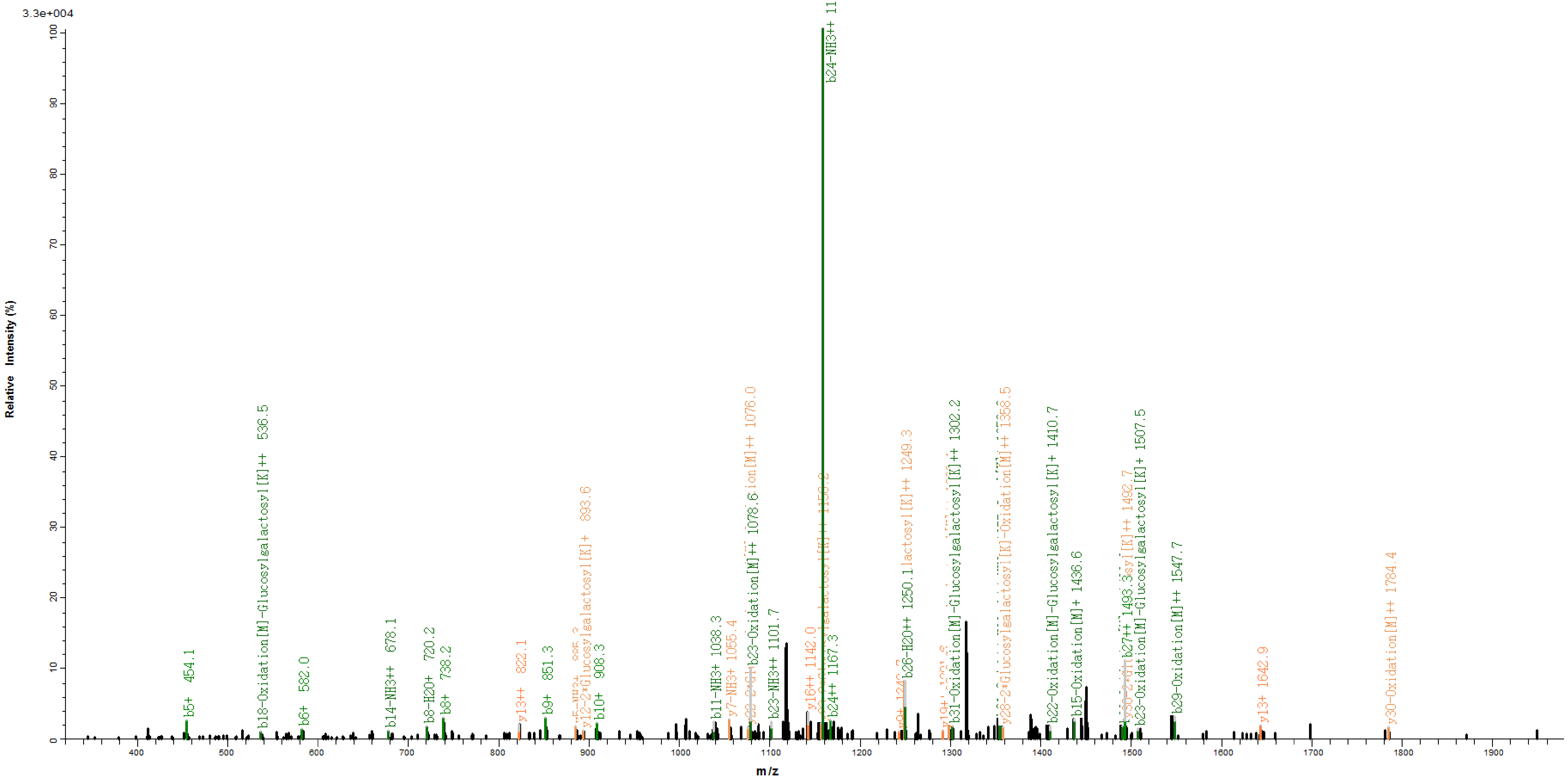
Supporting figure S3: Peptide spectrum matches (PSMs) of 39 O-glycosylation, 14 3-hydroxyproline and 3 hydroxyproline sites at Xaa position of Gly-HyP-Gln motif in human col4a1 chain. PSMs were annotated using pLABEL[1] or IDPicker utility software Ion Matcher[2].

m/z= 1200.8137; charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK[78,90]

Human col4a1 chain

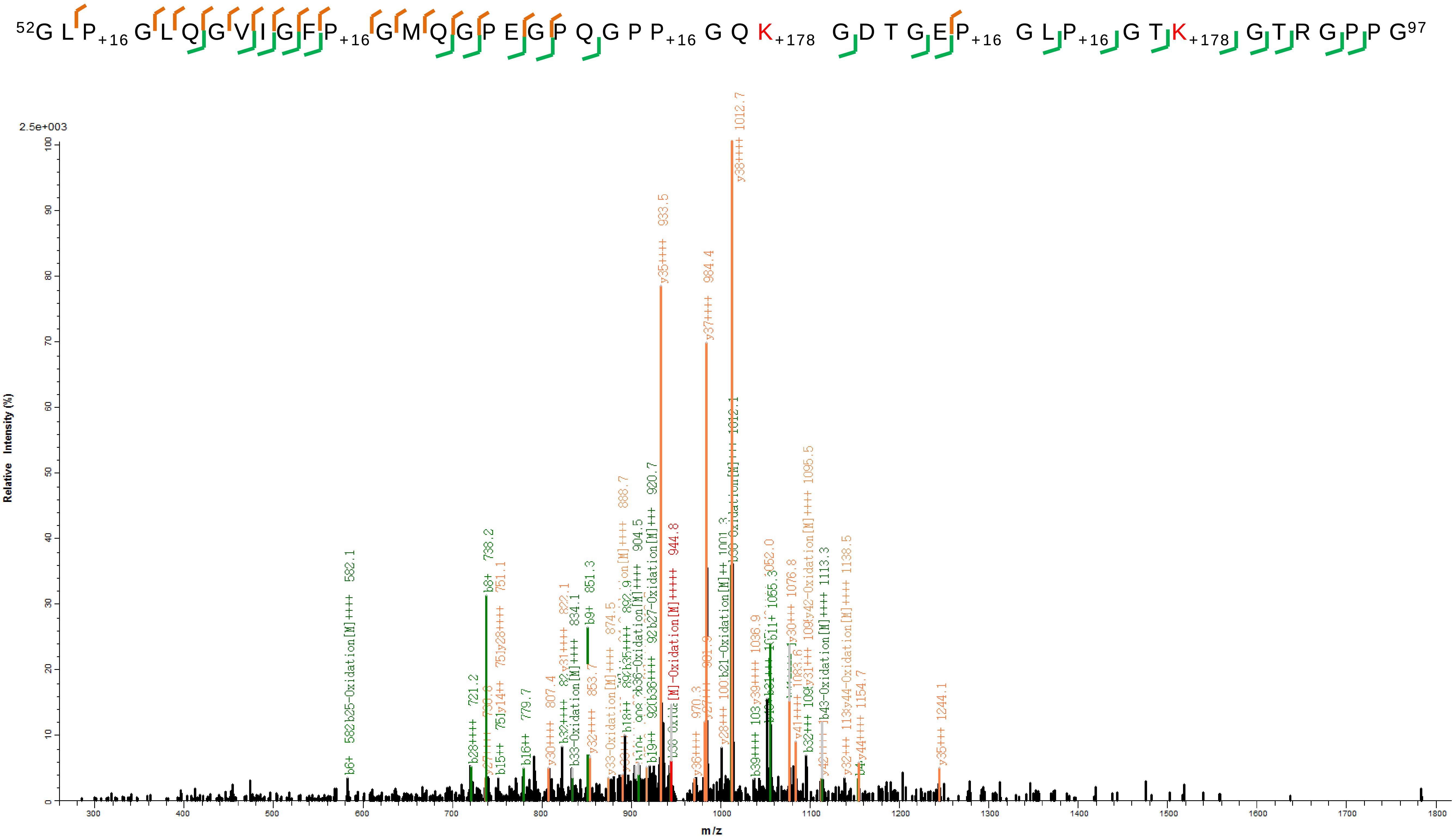
^{52}G L P$_{+16}$ G L Q G V I G F P$_{+16}$ G M Q G P E G P Q G P P$_{+16}$ G Q K$_{+340}$ G D T G E P$_{+16}$ G L P$_{+16}$ G T K$_{+340}$ G T R^{93}



m/z= 957.6457; charge=+5

PSM showing the identification of Galactosyl-HyK^{78,90}

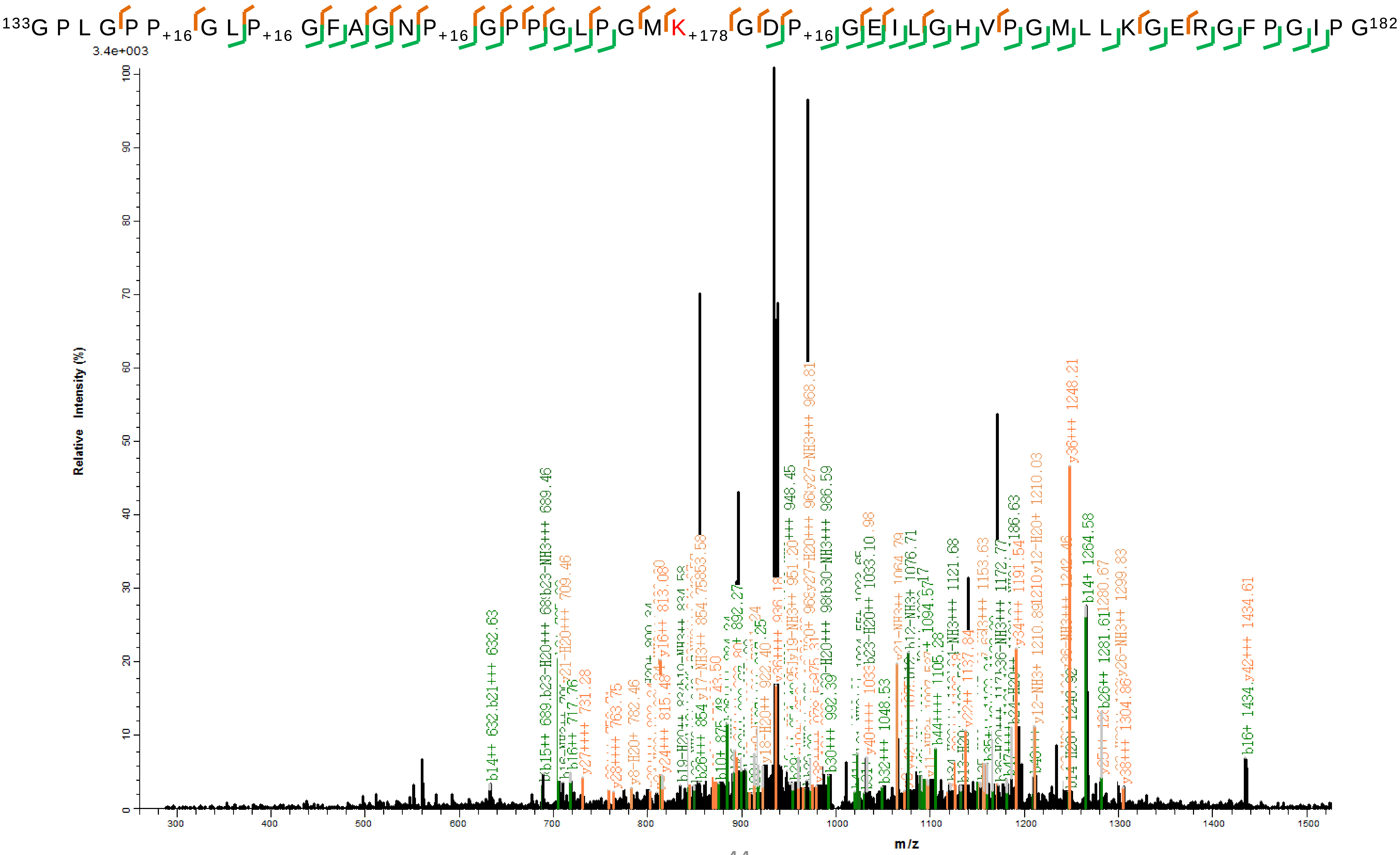
Human col4a1 chain



m/z= 1001.8837; charge=+5

PSM showing the identification of Glucosylgalactosyl-HyK^{156}

Human col4a1 chain

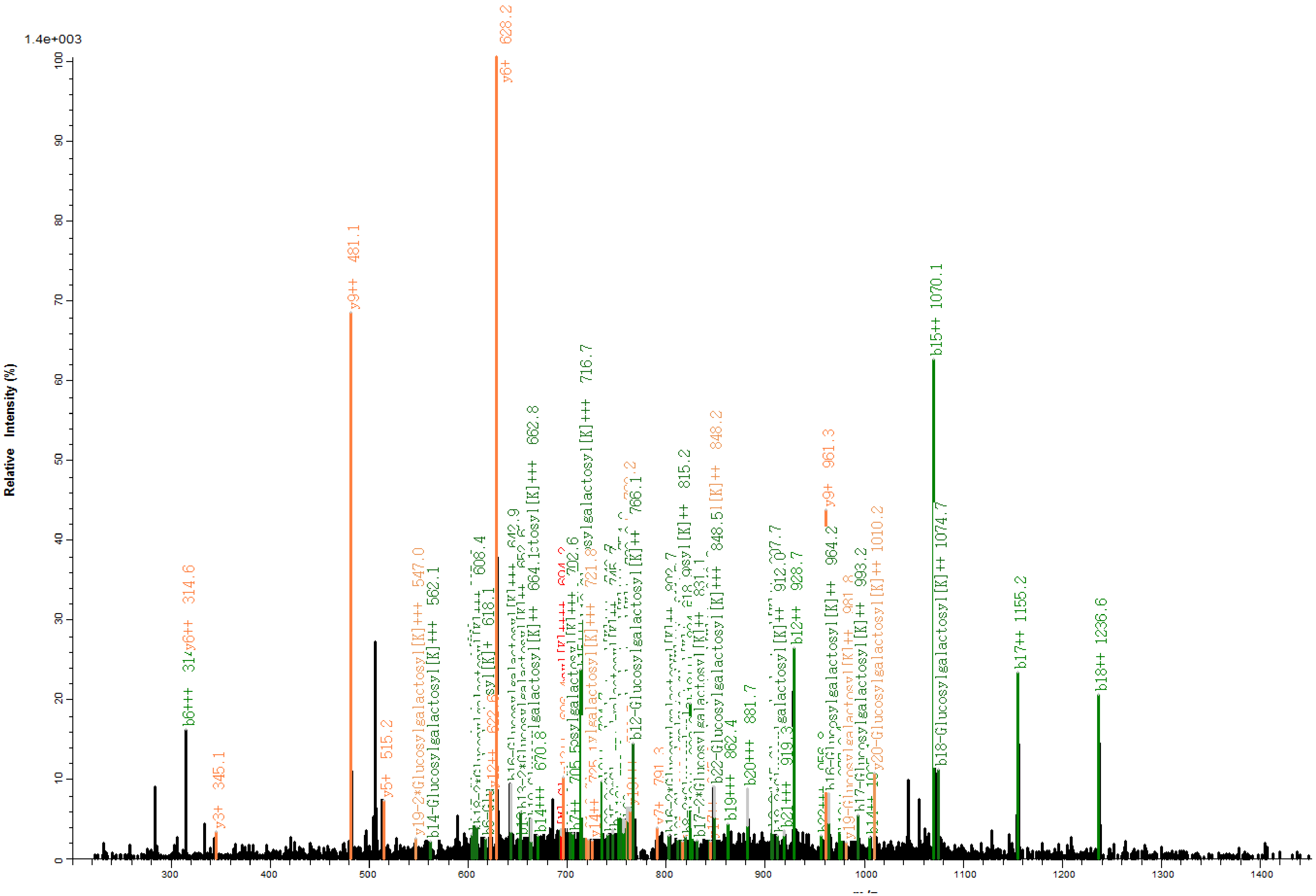


m/z= 775.5952; charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{295,298}

Human col4a1 chain

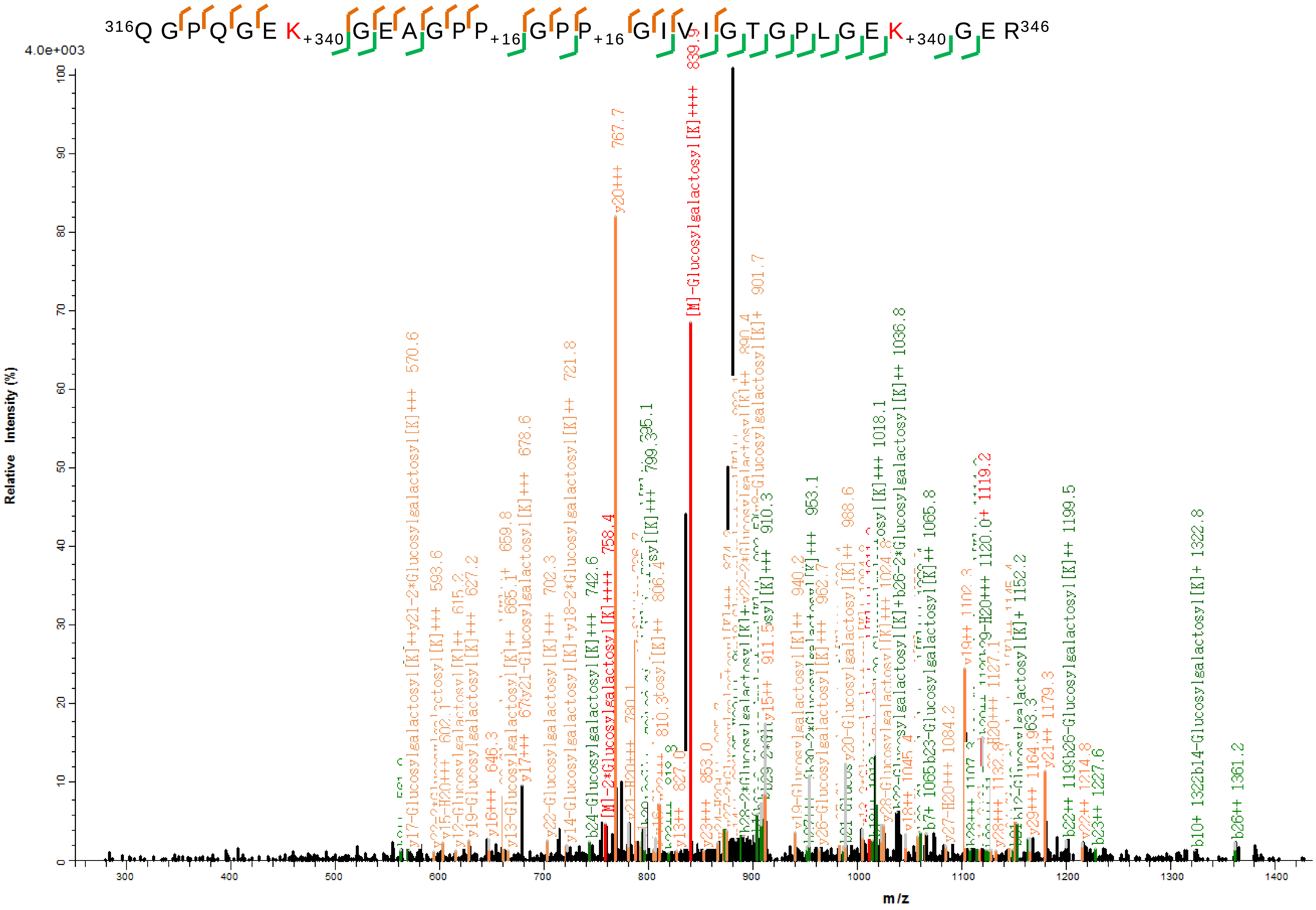
^{292}D G D K$_{+340}$ G E K$_{+340}$ G S P G F P G E P$_{+16}$ G Y P$_{+16}$ G L I G R^{315}



m/z= 920.4349; charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{321,343}

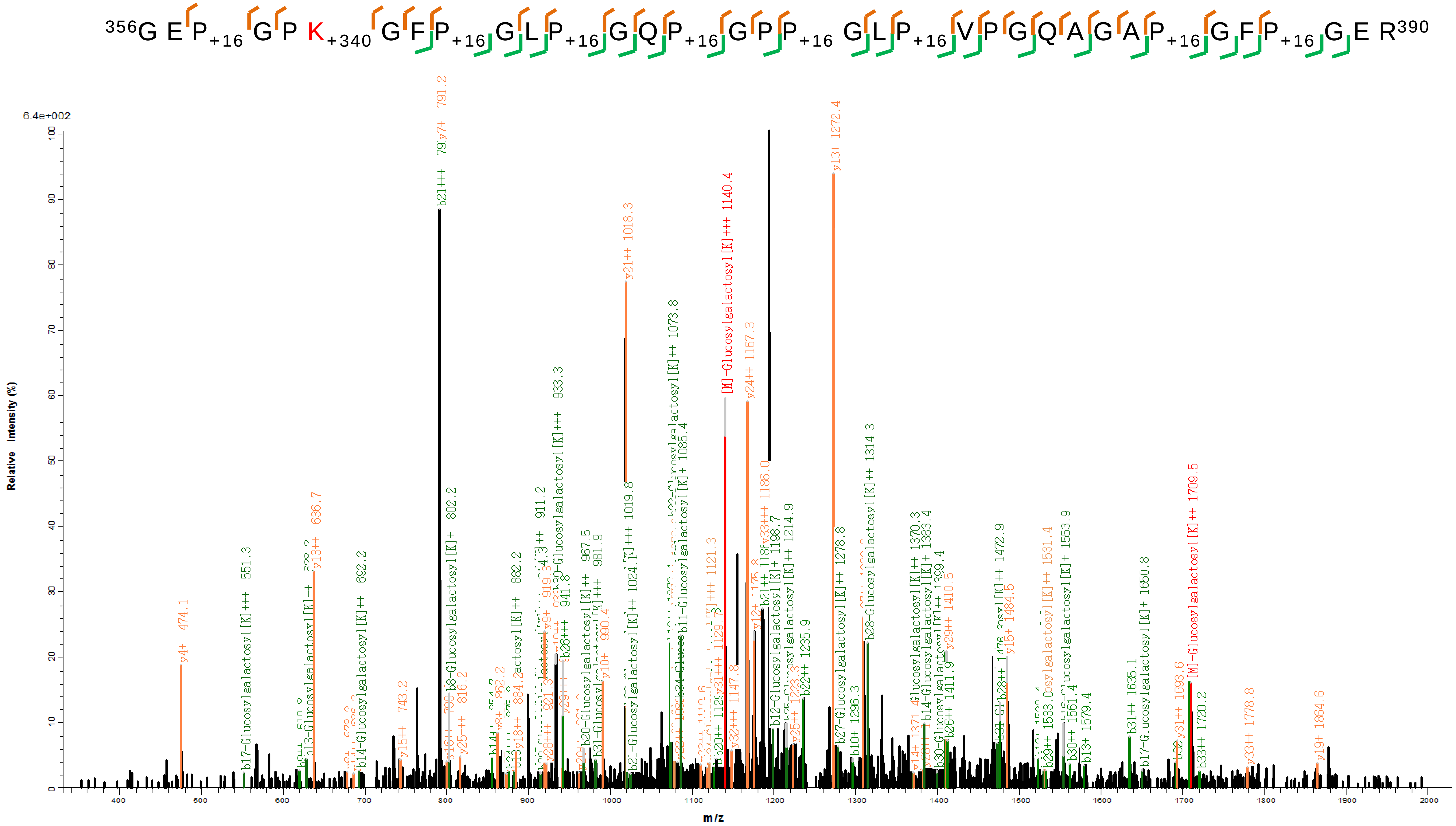
Human col4a1 chain



m/z= 1247.923; charge=+3

PSM showing the identification of Glucosylgalactosyl-HyK^{361}

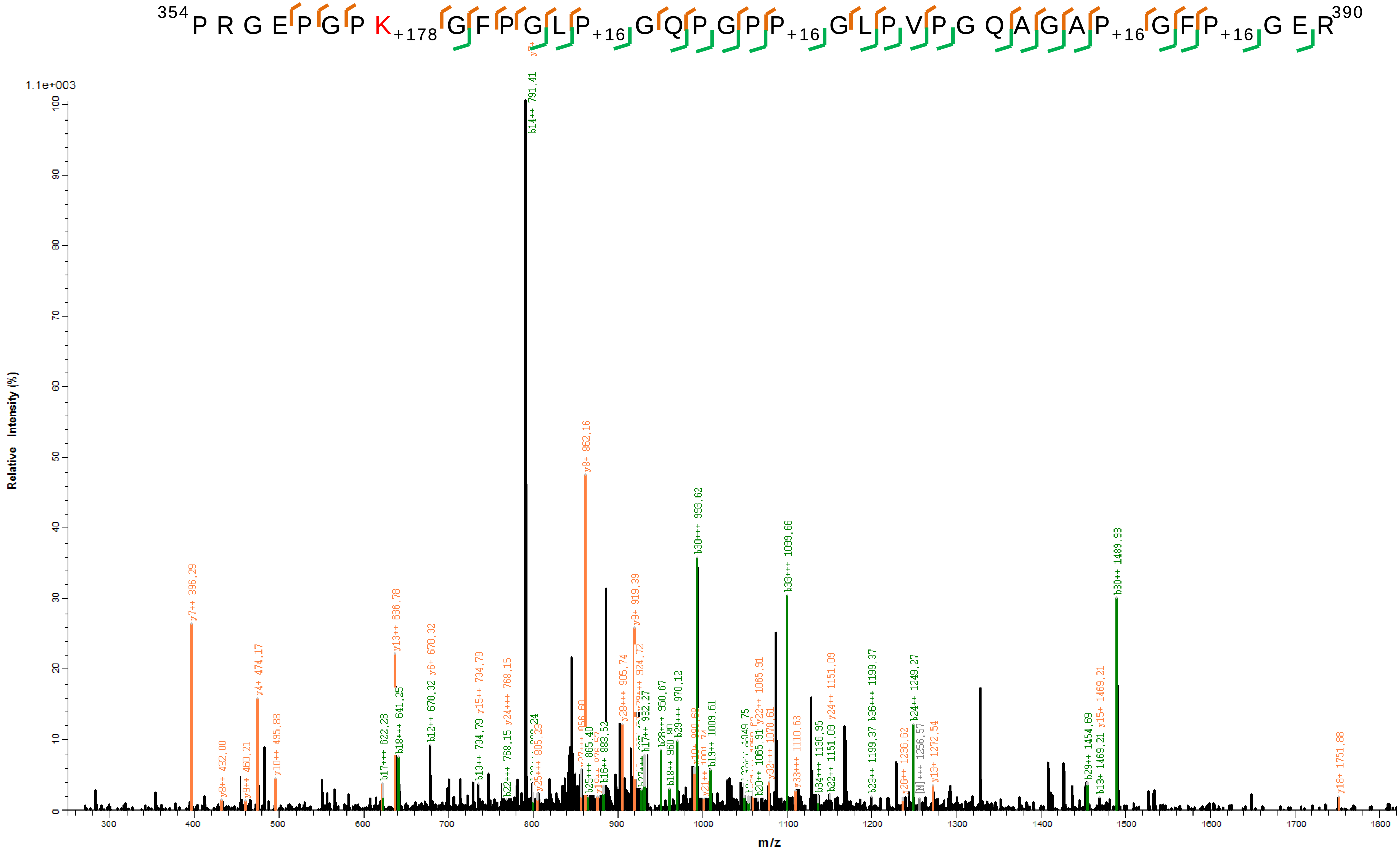
Human col4a1 chain



m/z= 942.94; charge=+4

PSM showing the identification of Galactosyl-HyK^{361}

Human col4a1 chain

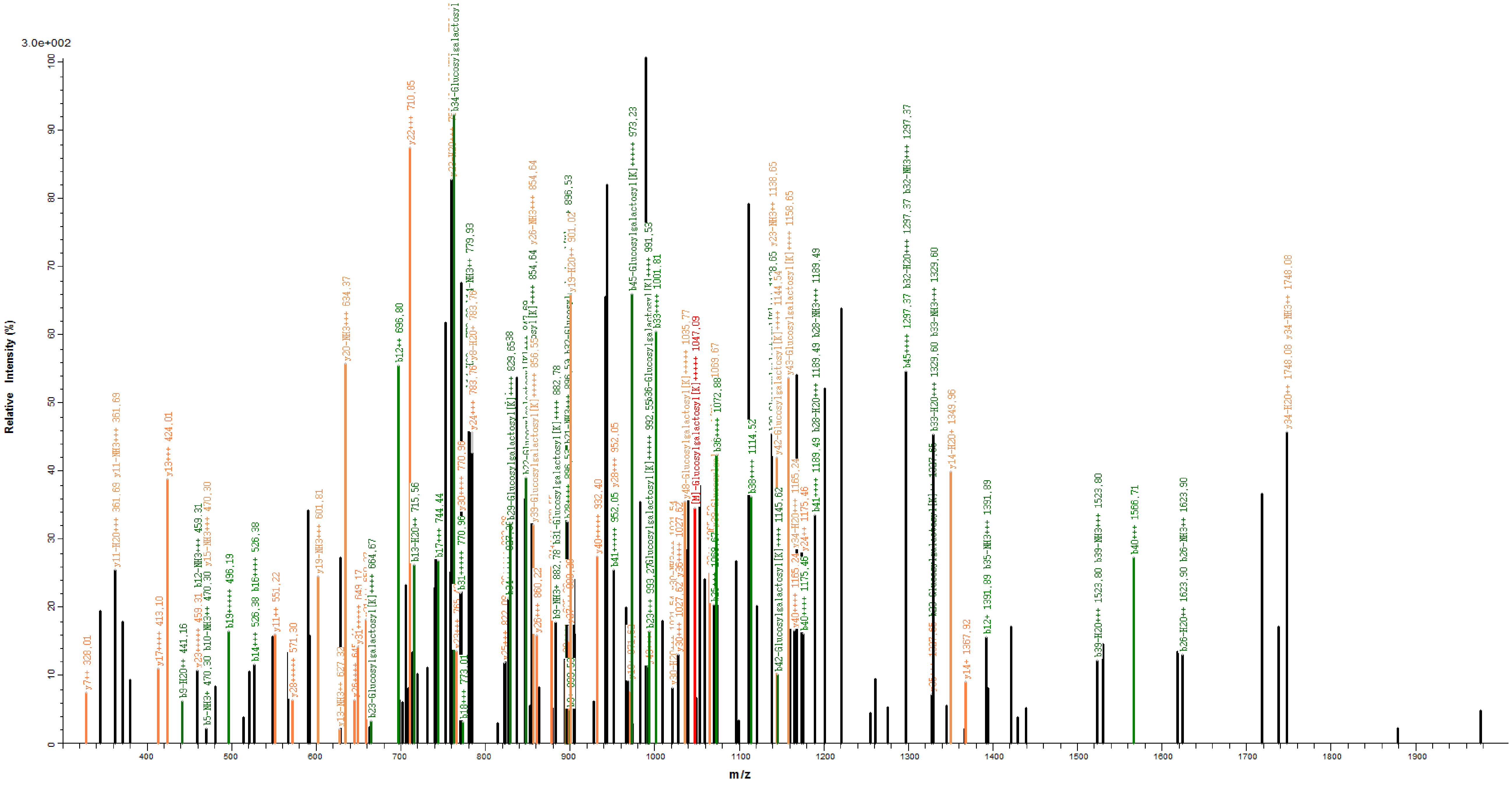


m/z= 926.9843 charge=+6

PSM showing the identification of Galactosyl-HyK^{460}

Human col4a1 chain

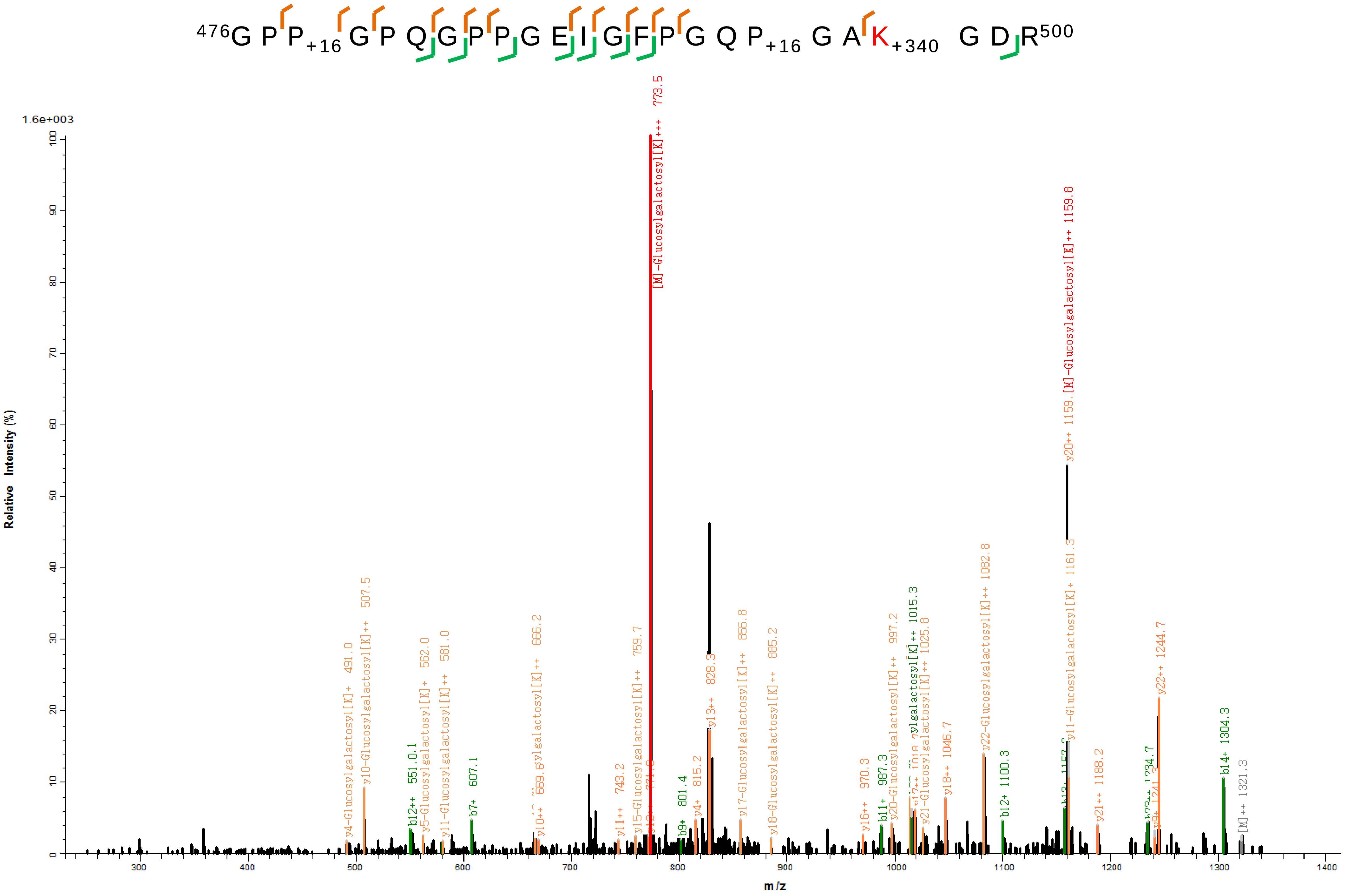
^{449}G Q P G F I G E I G E K$_{+178}$ G Q K$_{+340}$ G E S C$_{+57}$ L I C$_{+57}$ D I D G Y R G P P$_{+16}$ G P Q G P P$_{+16}$ G E I G F P G Q P G A K^{497}



m/z= 881.4156; charge=+3

PSM showing the identification of Glucosylgalactosyl-HyK^{497}

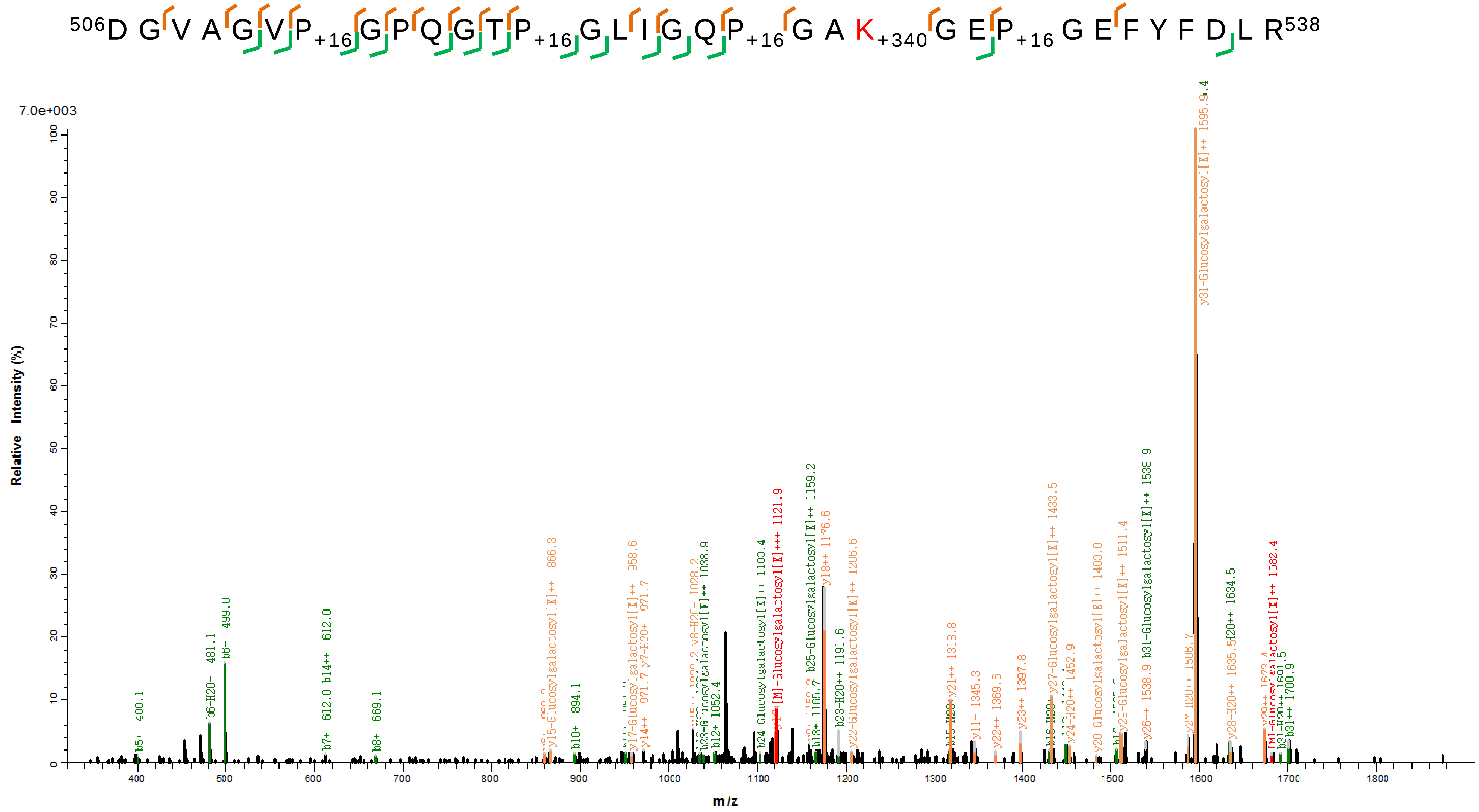
Human col4a1 chain



m/z= 1229.9143; charge=+3

PSM showing the identification of Glucosylgalactosyl-HyK^{527}

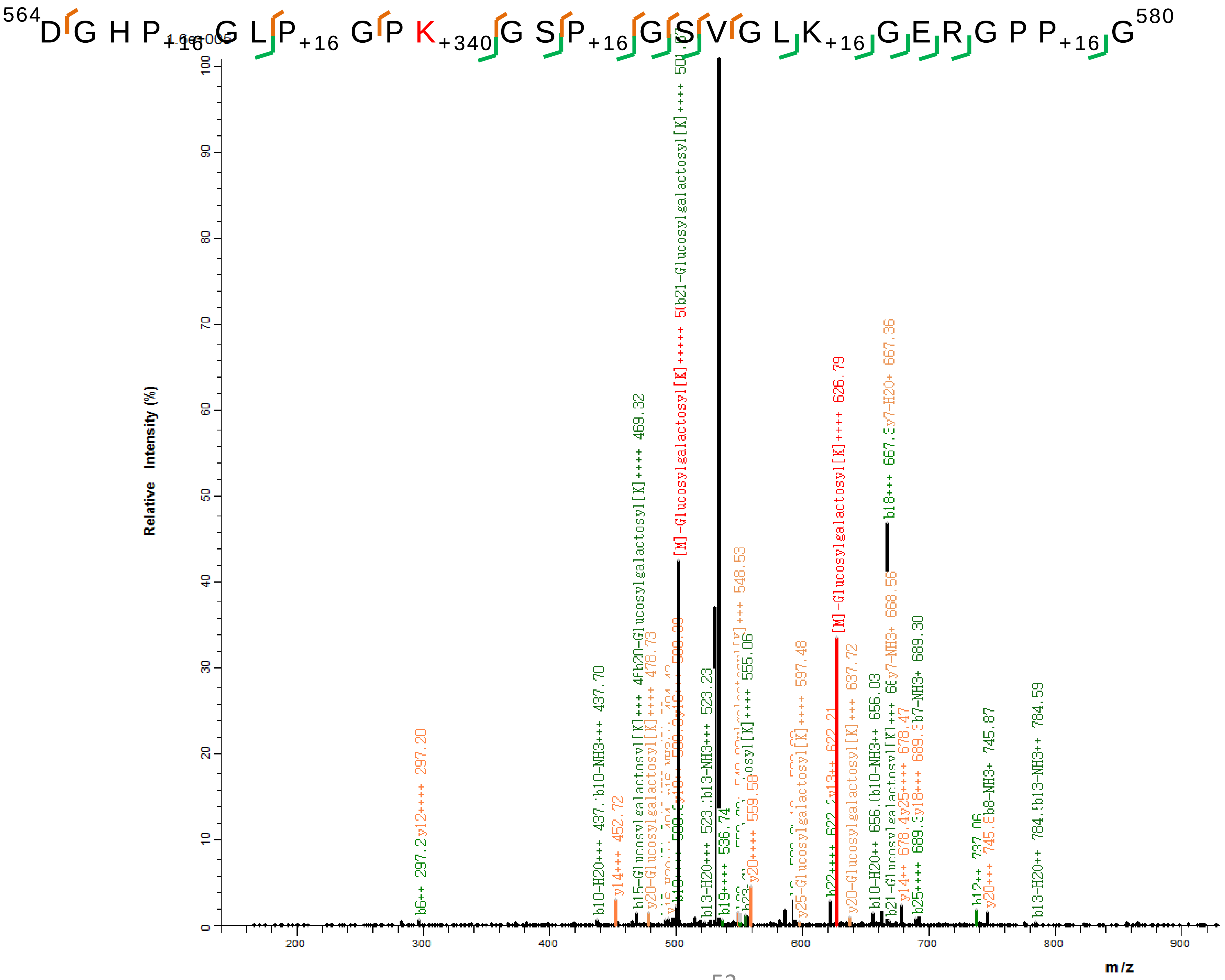
Human col4a1 chain



m/z= 566.2649; charge=+5

PSM showing the identification of Glucosylgalactosyl-HyK^{573}

Human col4a1 chain

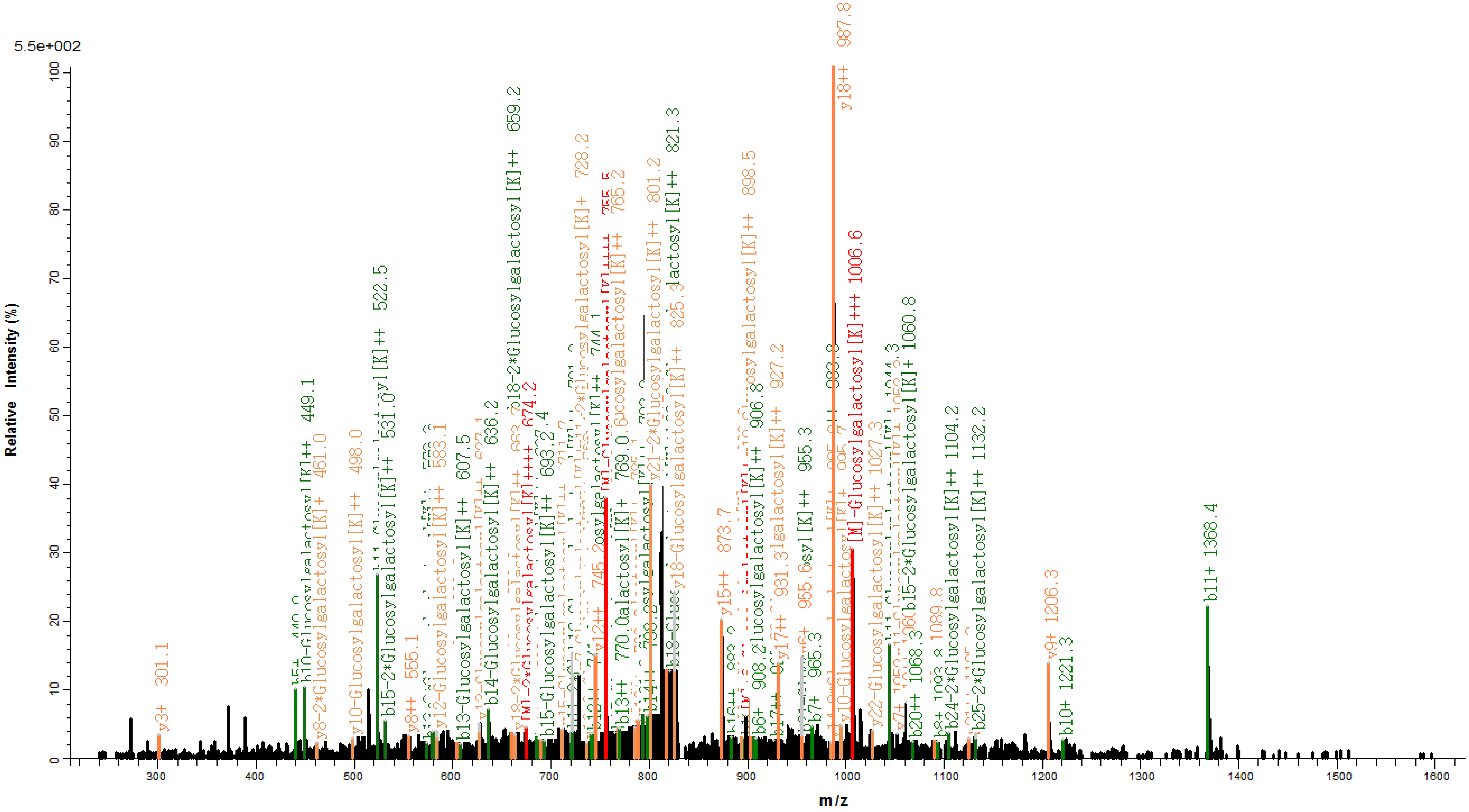


m/z= 836.1401charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{617,635}

Human col4a1 chain

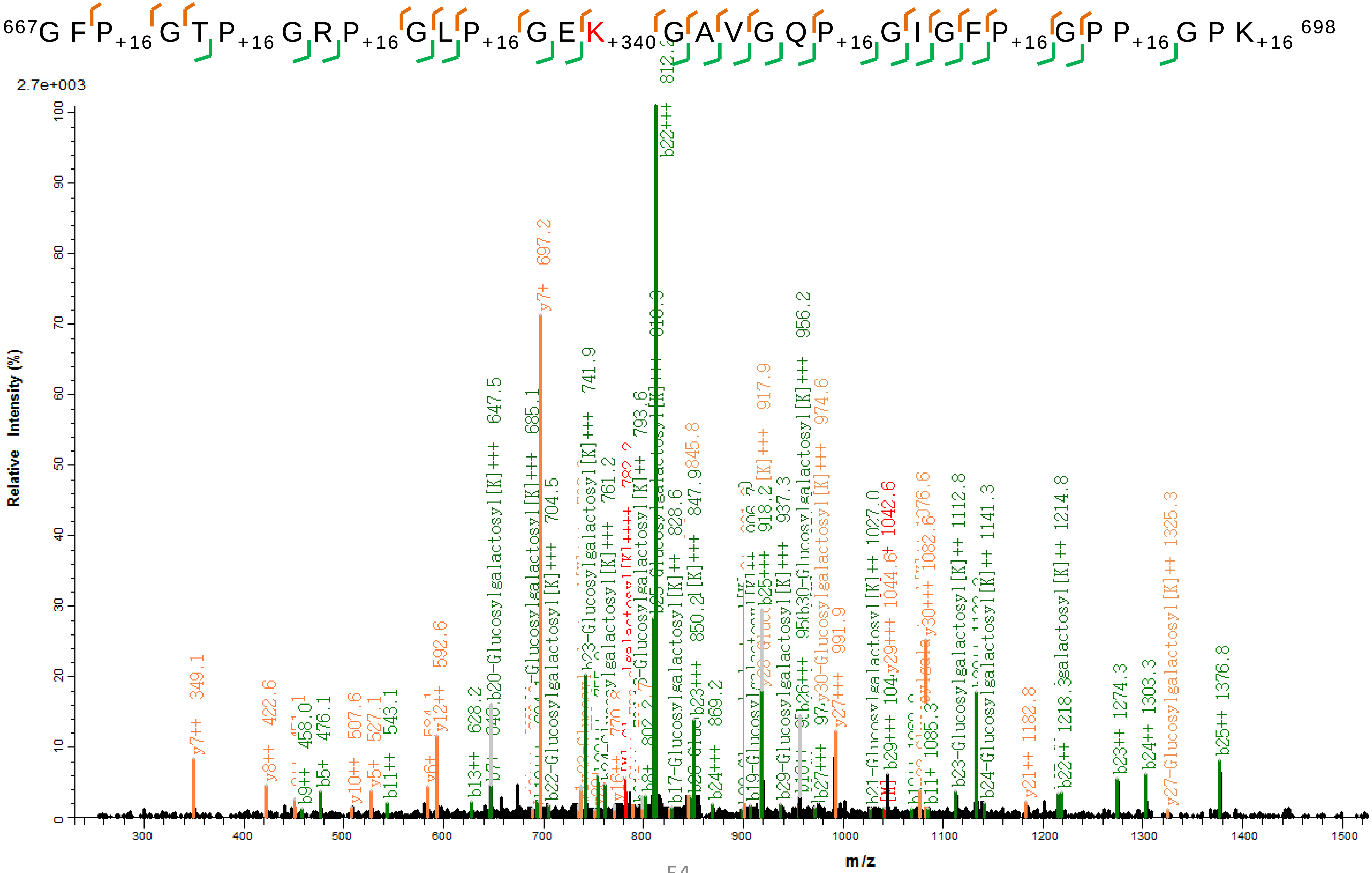
^{612}G P I G D K$_{+340}$ G Q A G F P$_{+16}$ G G P$_{+16}$ G S P$_{+16}$ G L P G P K$_{+340}$ G E P G K^{640}



m/z= 862.9186 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{681}

Human col4a1 chain

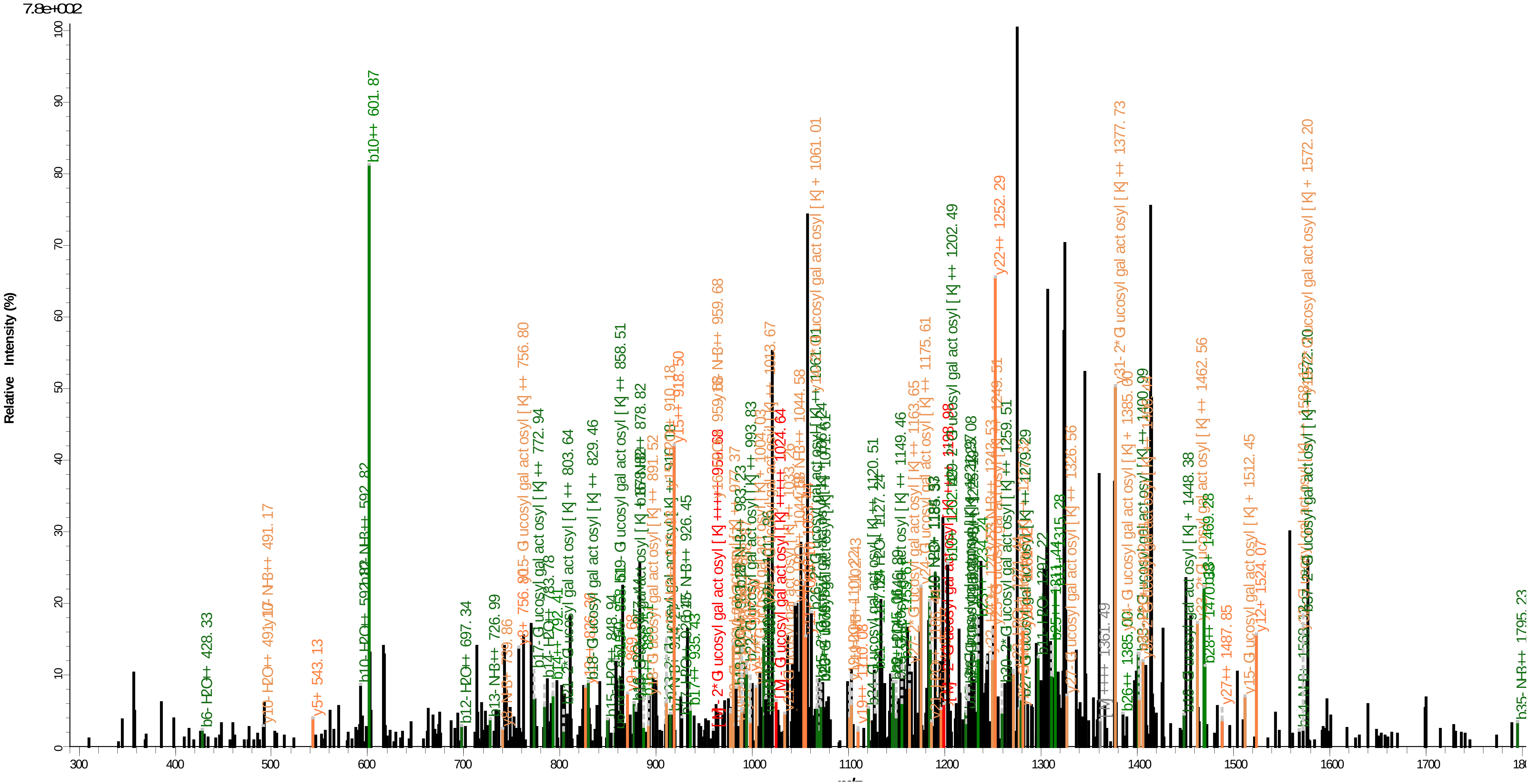


m/z= 1089.1052 charge=+5

PSM showing the identification of Glucosylgalactosyl-HyK^{698}

Human col4a1 chain

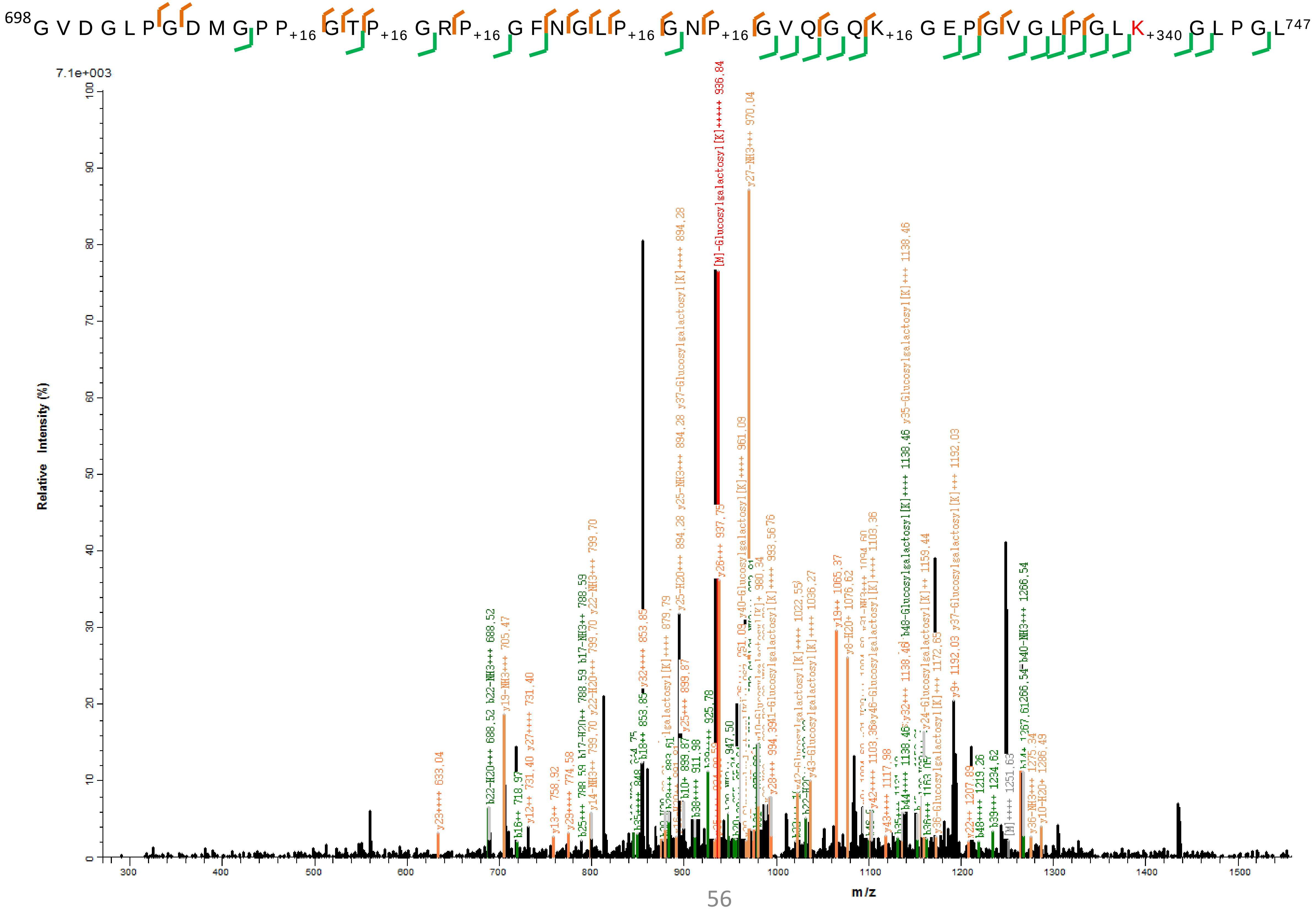
^{693}G P P G P K$_{+340}$ G V D G L P G D M G P P$_{+16}$ G T P$_{+16}$ G R P G F N G L P$_{+16}$ G N P$_{+16}$ G V Q G Q K$_{+340}$ G E P$_{+16}$ G V G L P$_{+16}$ G L K



m/z= 1001.8720 charge=+5

PSM showing the identification of Glucosylgalactosyl-HyK^{742}

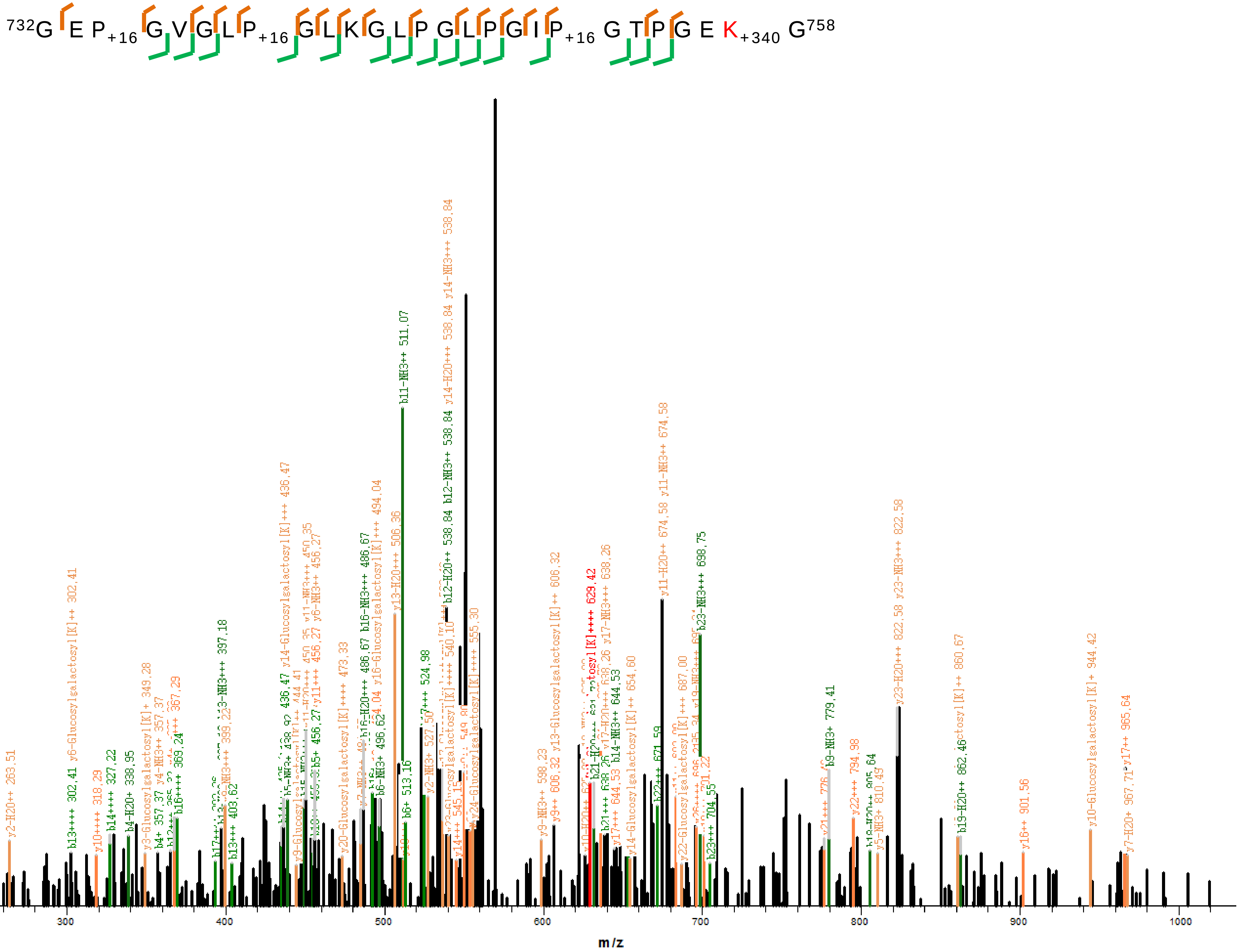
Human col4a1 chain



m/z= 568.6699 charge=+5

PSM showing the identification of Glucosylgalactosyl-HyK^{757}

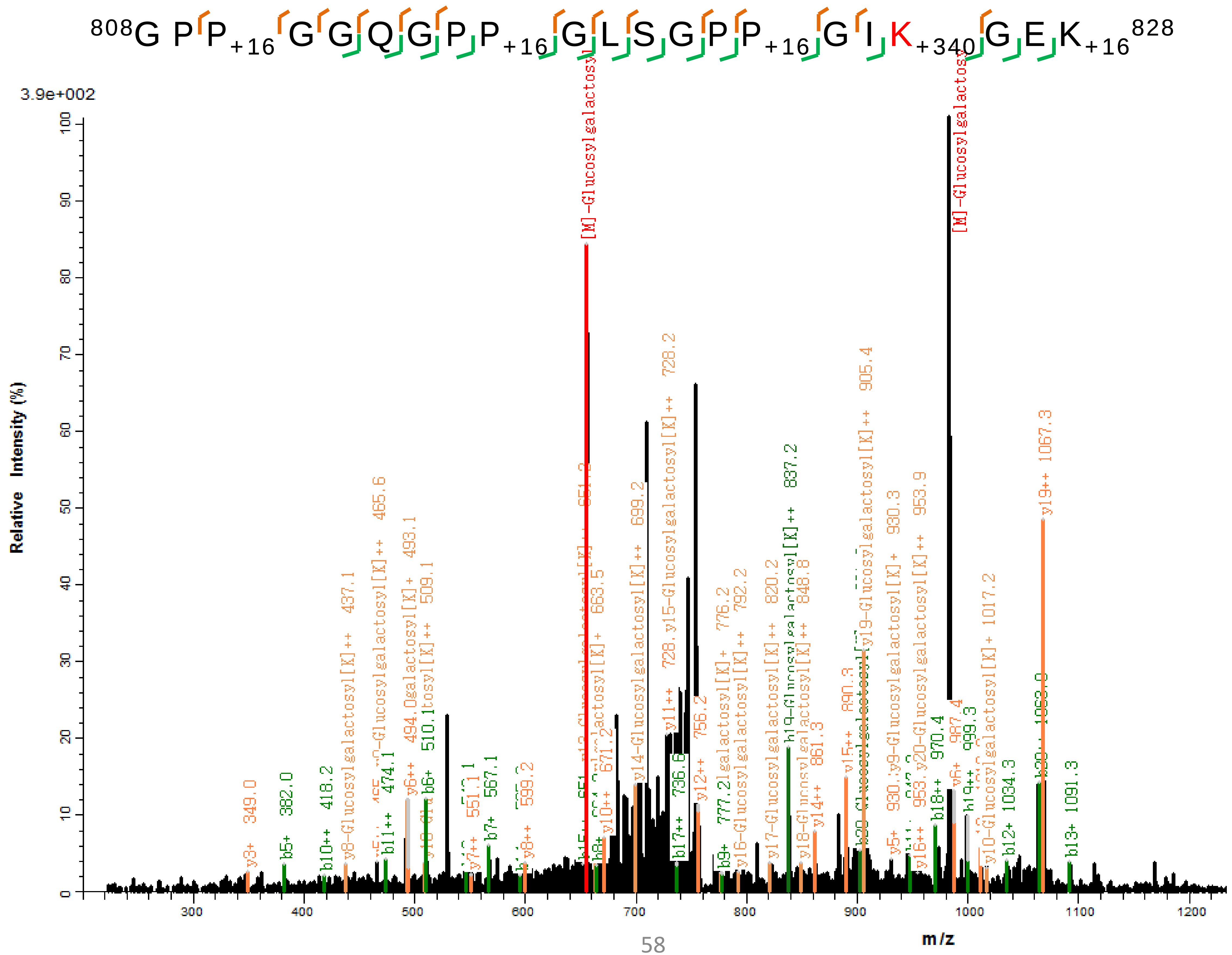
Human col4a1 chain



m/z= 763.3694 charge=+3

PSM showing the identification of Glucosylgalactosyl-HyK[825]

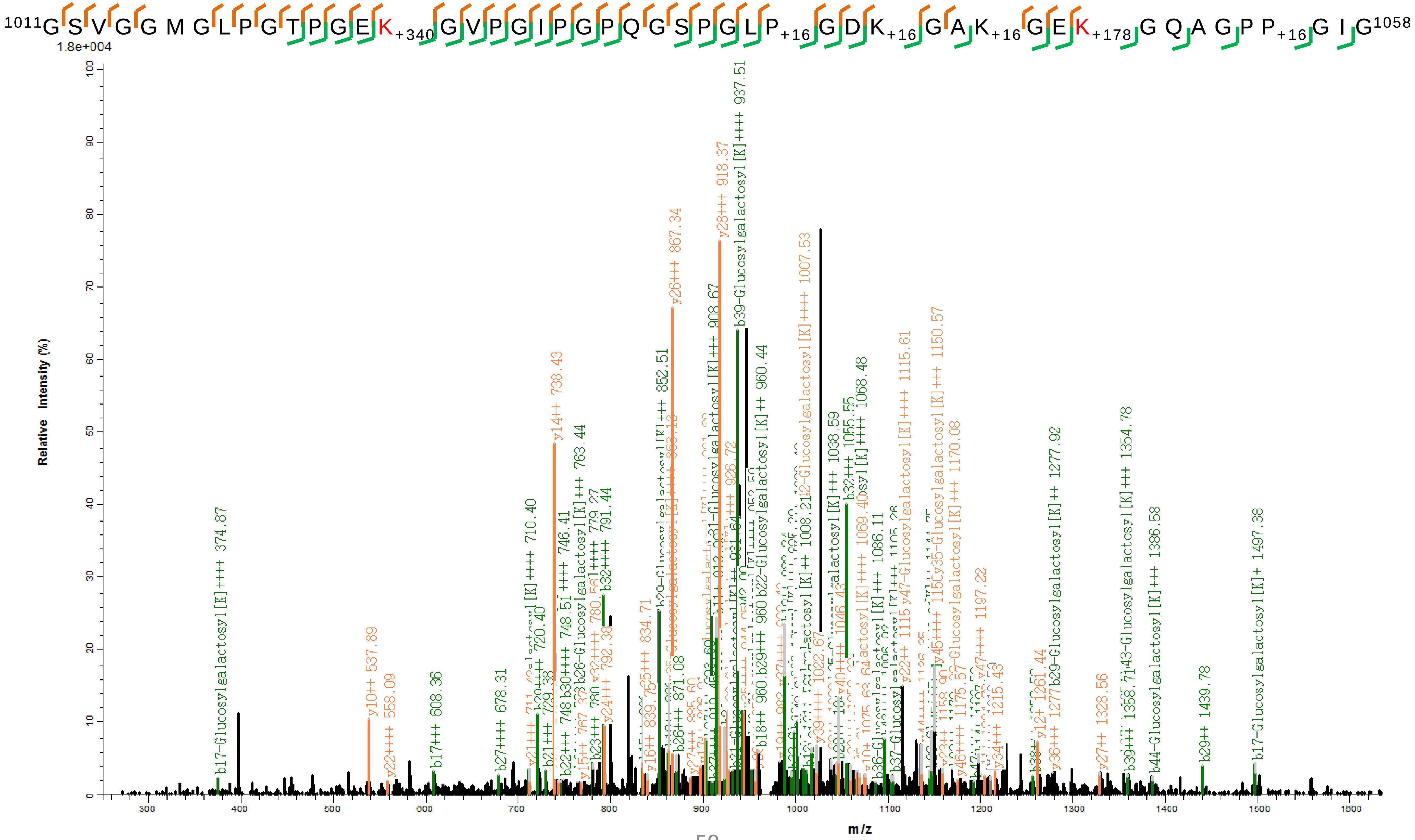
Human col4a1 chain



m/z= 969.05 charge=+3

PSM showing the identification of Glucosylgalactosyl-HyK^{1025}

Human col4a1 chain

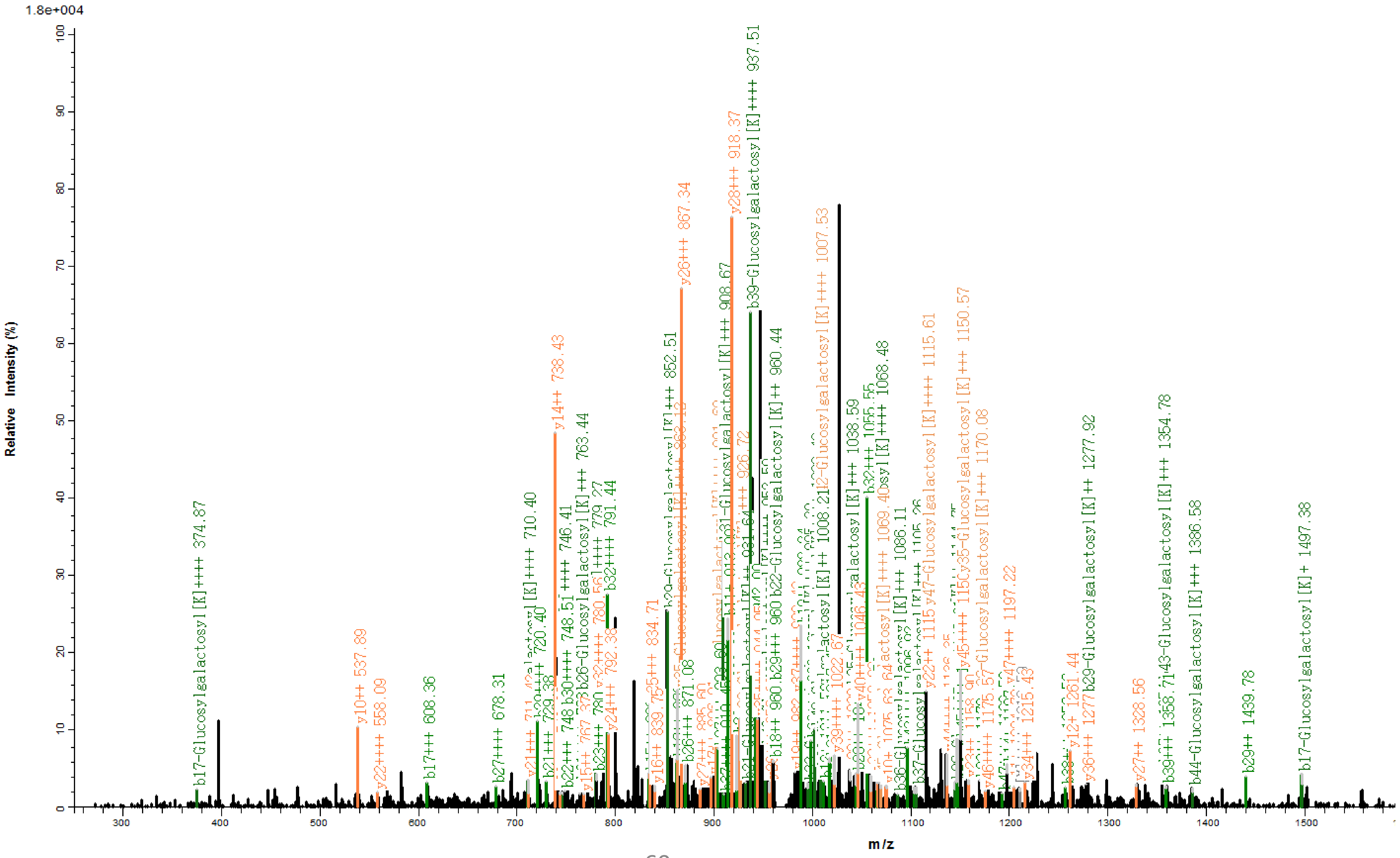


m/z= 906.4286 charge=+4

PSM showing the identification of Galactosyl-HyK^{1046}

Human col4a1 chain

^{1011}G S V G G M G L P G T P G E K G V P$_{+16}$ G I P$_{+16}$ G P Q G S P G L P$_{+16}$ G D K G A K$_{+178}$ G E^{1048}

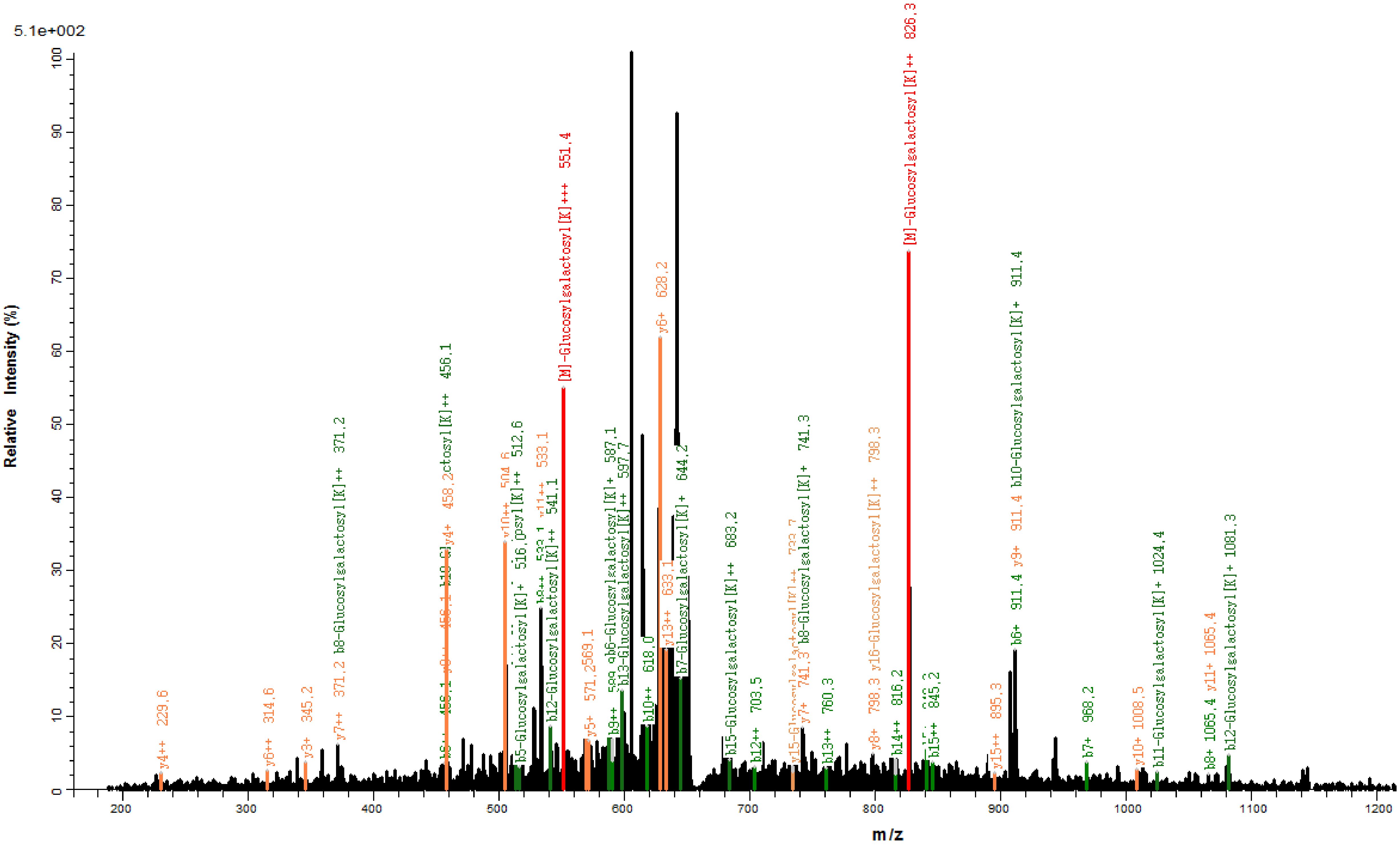


m/z= 659.3373 charge=+3

PSM showing the identification of Glucosylgalactosyl-HyK^{1049}

Human col4a1 chain

^{1047}G E K$_{+340}$ G Q A G P P$_{+16}$ G I G I P$_{+16}$ G L R^{1063}

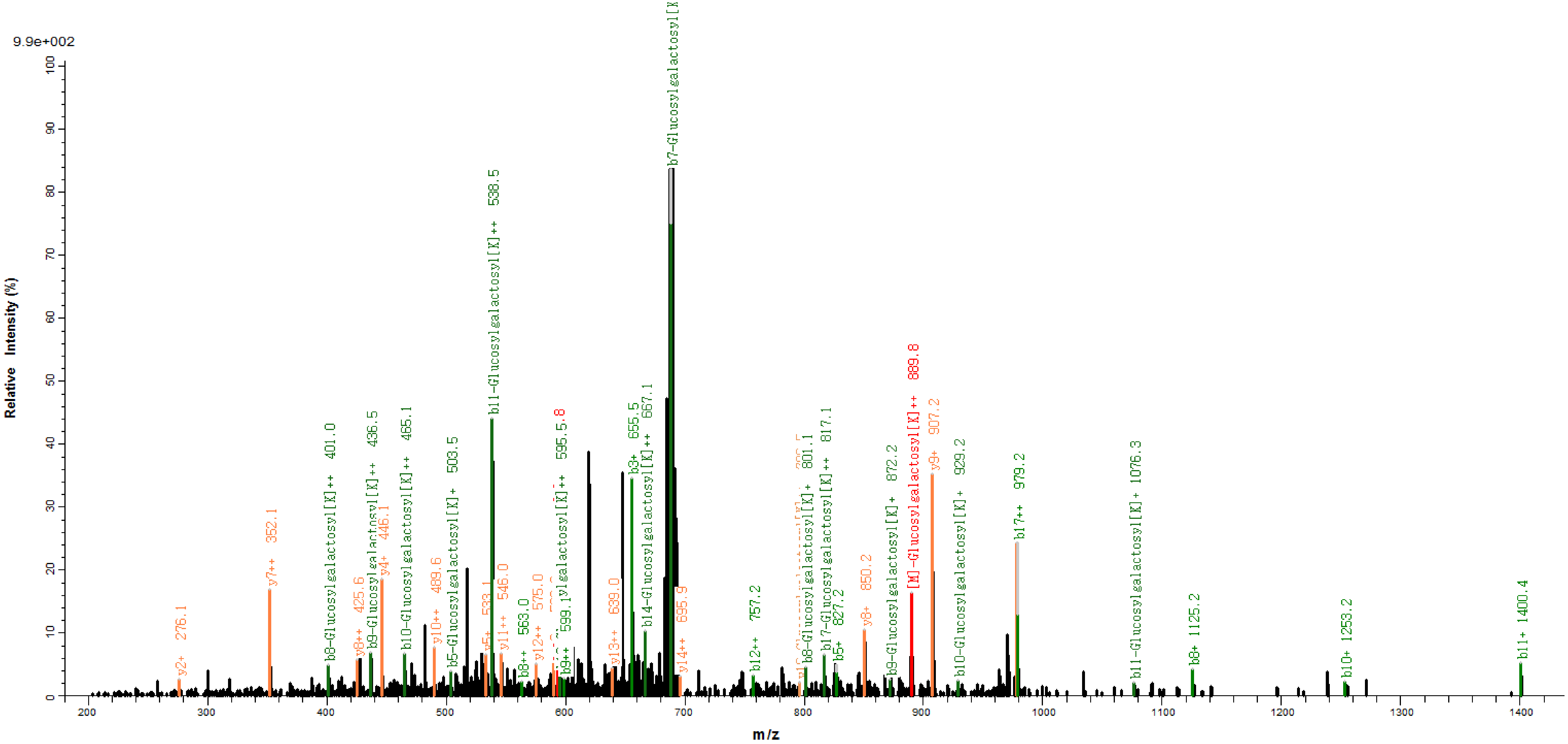


m/z= 701.6459 charge=+3

PSM showing the identification of Glucosylgalactosyl-HyK^{1066}

Human col4a1 chain

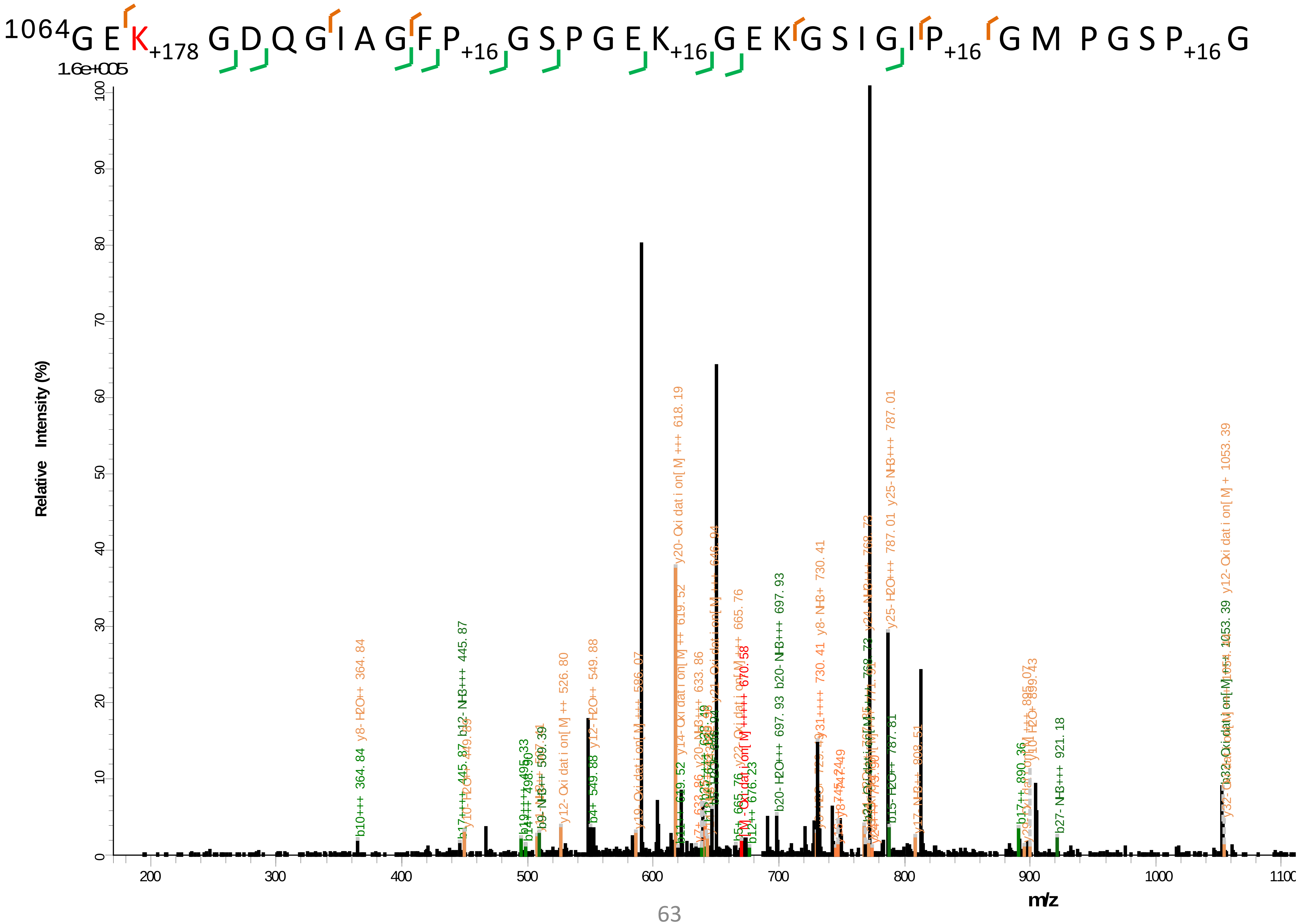
^{1064}G E K$_{+340}$ G D Q G I A G F P$_{+16}$ G S P$_{+16}$ G E K ^{1081}



m/z= 682.9234 charge=+5

PSM showing the identification of Galactosyl-HyK^{1066}

Human col4a1 chain

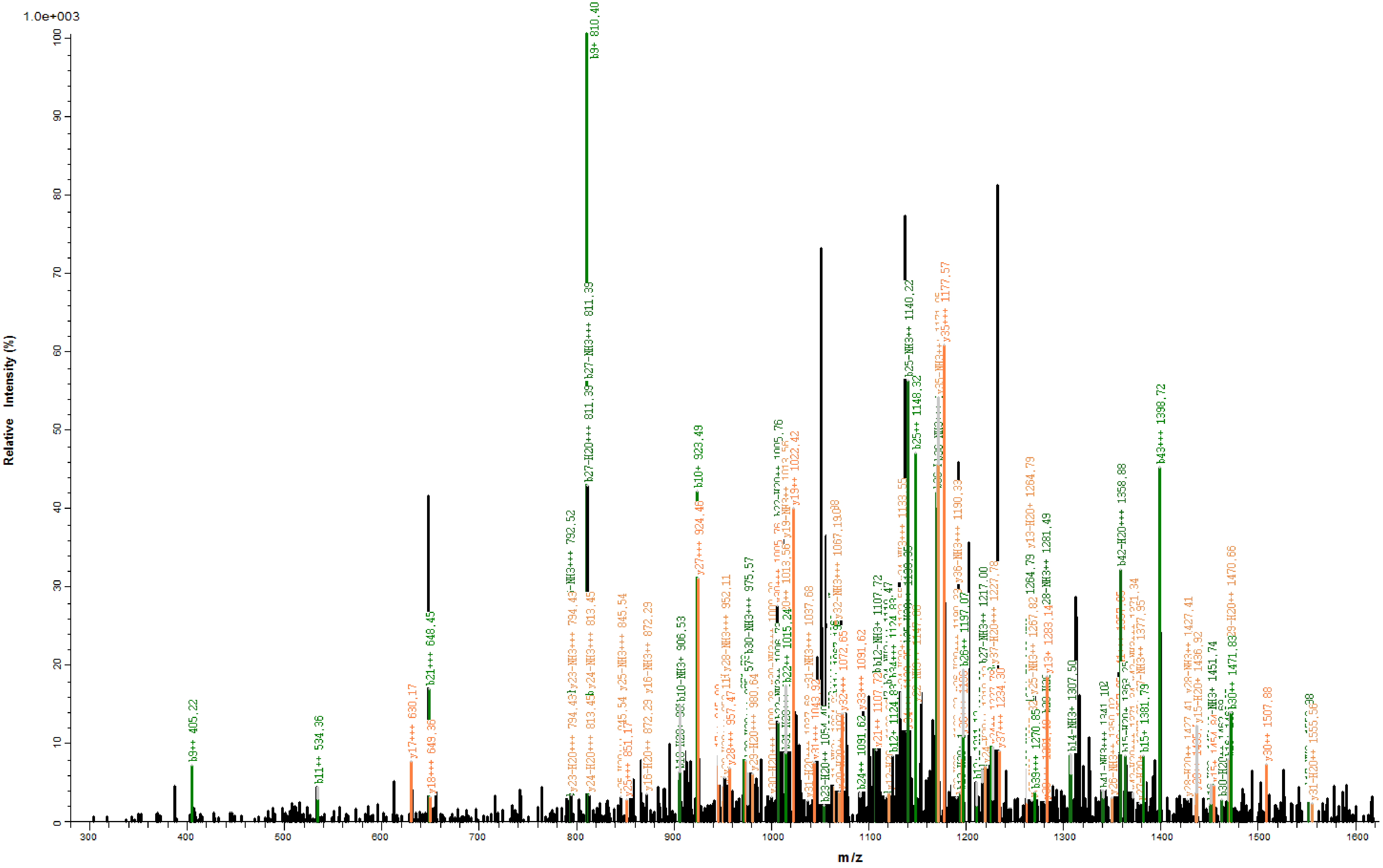


m/z= 1085.5333 charge=+4

PSM showing the identification of Galactosyl-HyK^{1117}

Human col4a1 chain

^{1089}I P G M P G S P$_{+16}$ G L K$_{+16}$ G S P$_{+16}$ G S V G Y P G S P G L P G E K$_{+178}$ G D K$_{+16}$ G L P G L D G I P$_{+16}$ G V K^{1132}

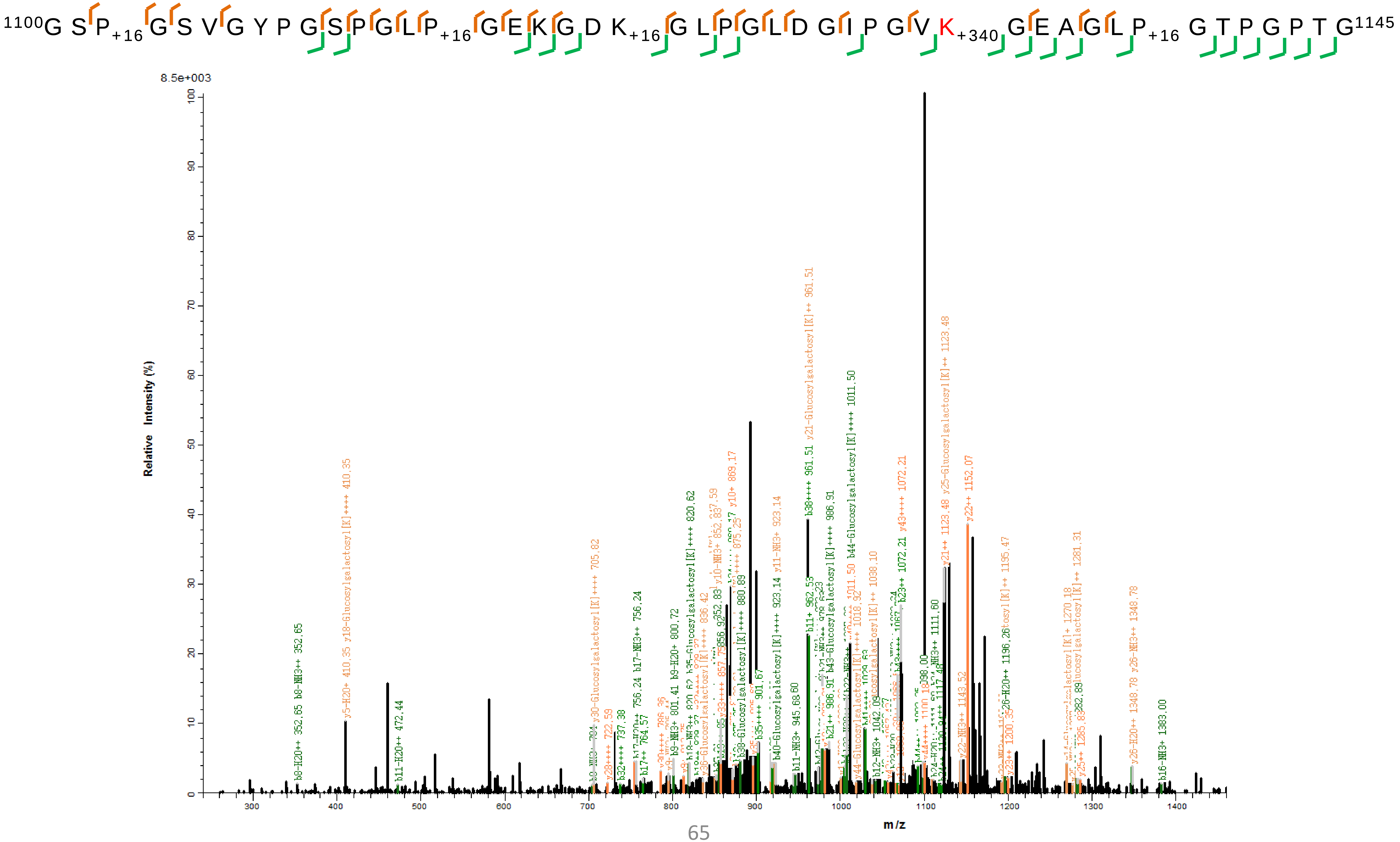


m/z= 909.2668 charge=+5

PSM showing the identification of Glucosylgalactosyl-HyK^{1132}

Human col4a1 chain

^{1100}G S P$_{+16}$ G S V G Y P G S P G L P$_{+16}$ G E K G D K$_{+16}$ G L P G L D G I P G V K$_{+340}$ G E A G L P$_{+16}$ G T P G P T G^{1145}

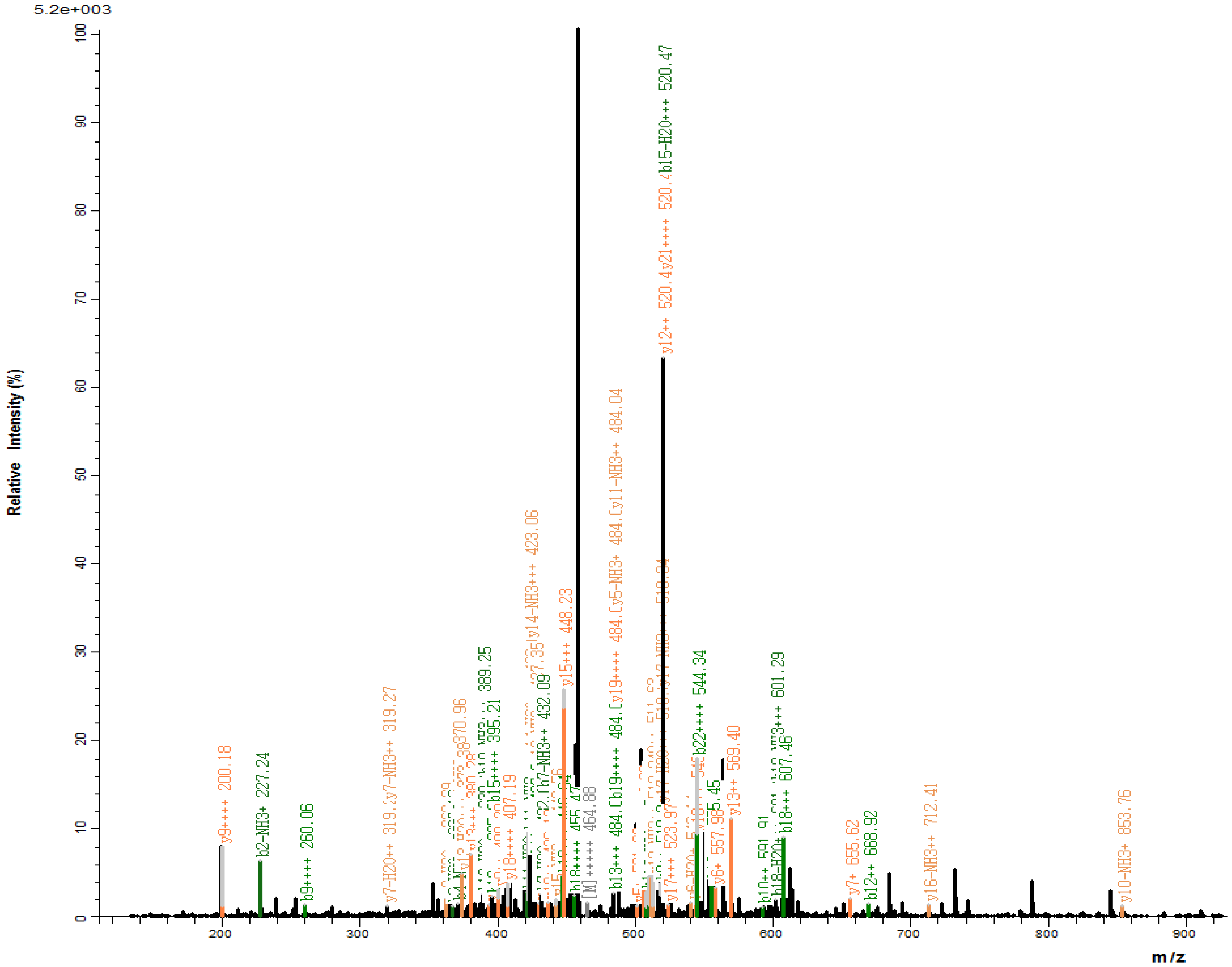


m/z= 464.8250 charge=+5

PSM showing the identification of Galactosyl-HyK^{1188}

Human col4a1 chain

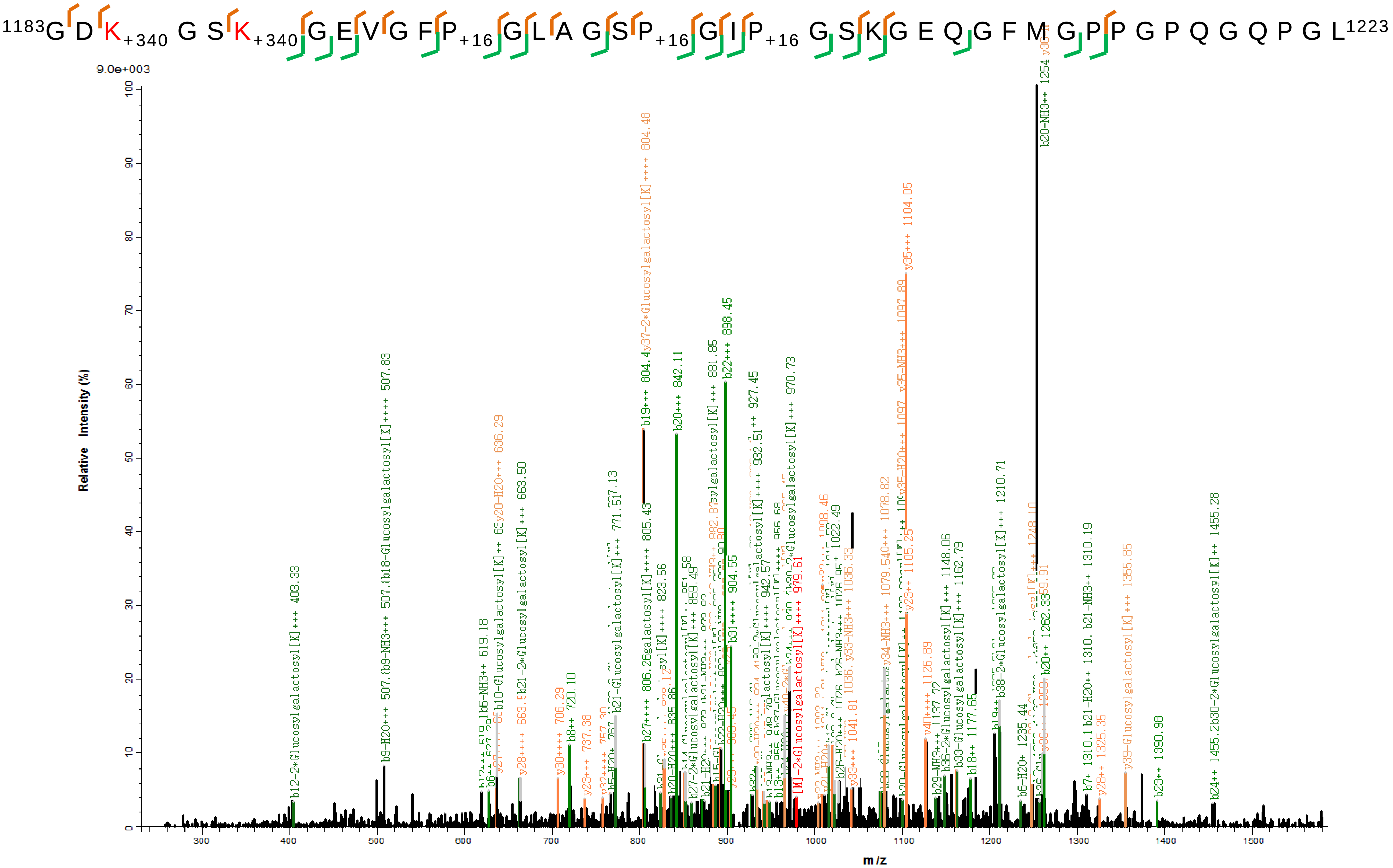
^{1184}D K G S K$_{+178}$ G E V G F P G L A G S P G I P G S K^{1206}



m/z= 913.2631 charge=+5

PSM showing the identification of Glucosylgalactosyl-HyK^{1185,1188}

Human col4a1 chain

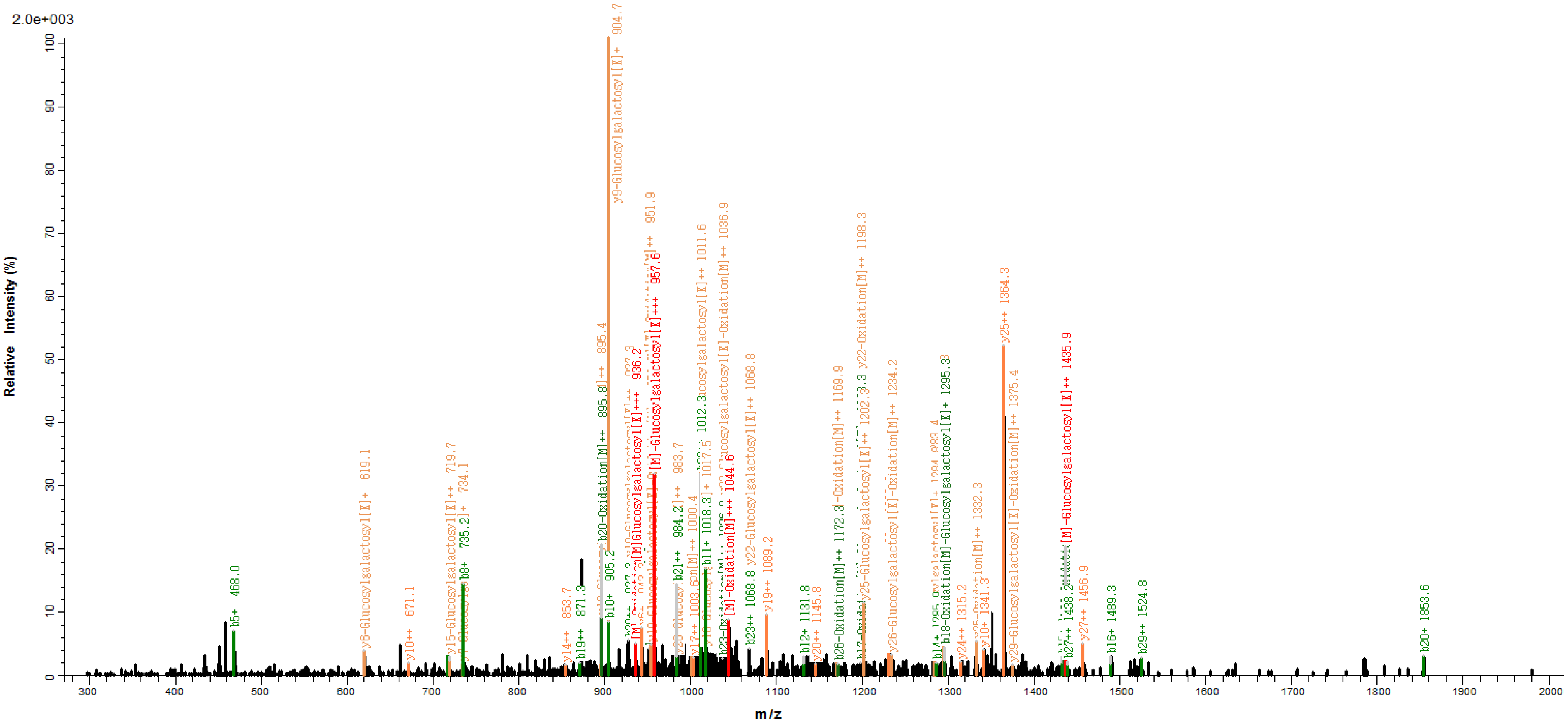


m/z= 1065.5275 charge=+3

PSM showing the identification of Glucosylgalactosyl-HyK[1265]

Human col4a1 chain

^{1239}G P Q G Q P G L P$_{+16}$ G L P$_{+16}$ G P M G P P G L P$_{+16}$ G I D G V K$_{+340}$ G D K^{1268}

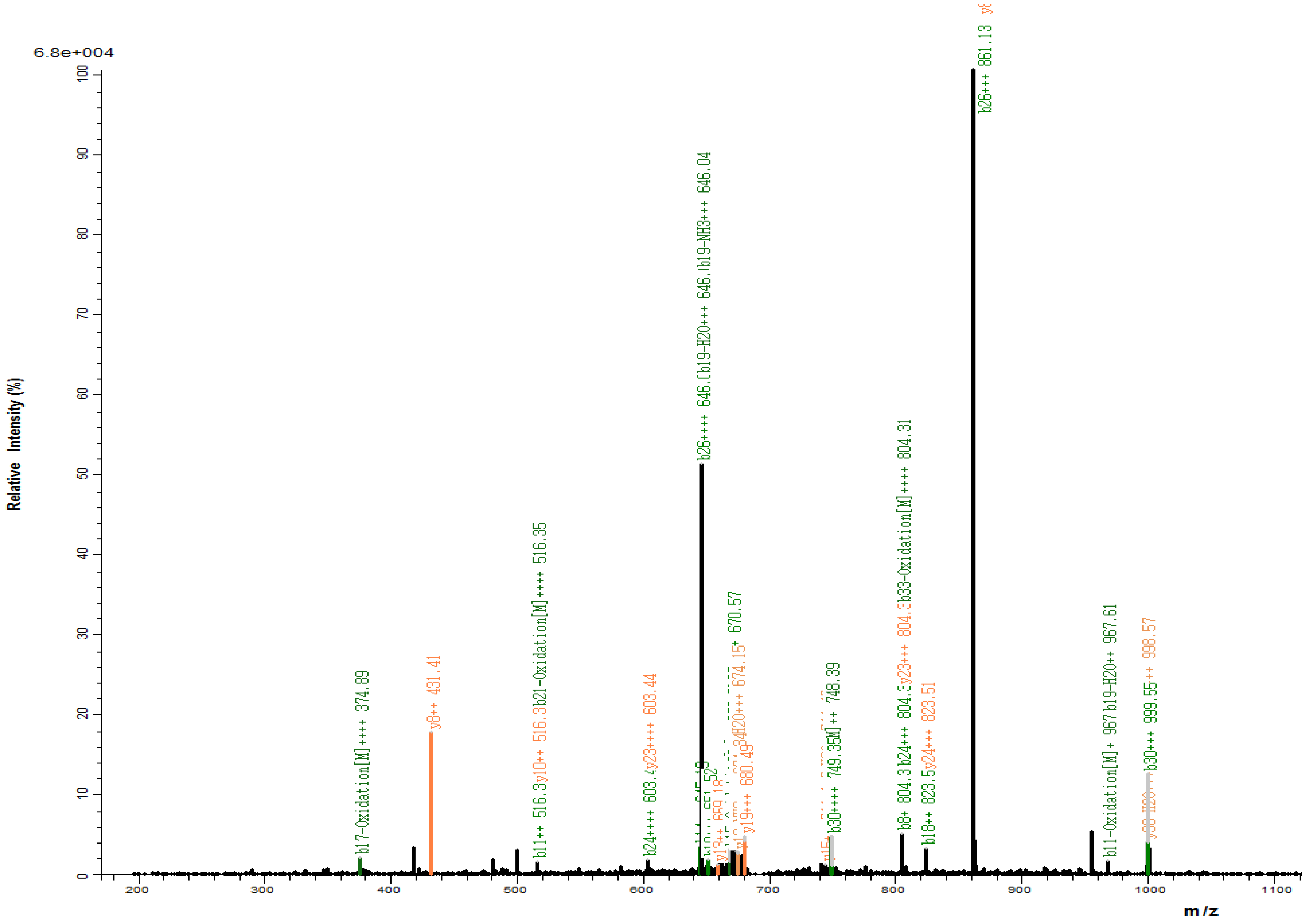


m/z= 1688.9413 charge=+5

PSM showing the identification of Galactosyl-HyK^{1304}

Human col4a1 chain

^{1286}P G F Q G M P$_{+16}$ G I G G S P$_{+16}$ G I T G S K$_{+178}$ G D M G P P$_{+16}$ G V P$_{+16}$ G F Q G P K$_{+16}$^{1319}

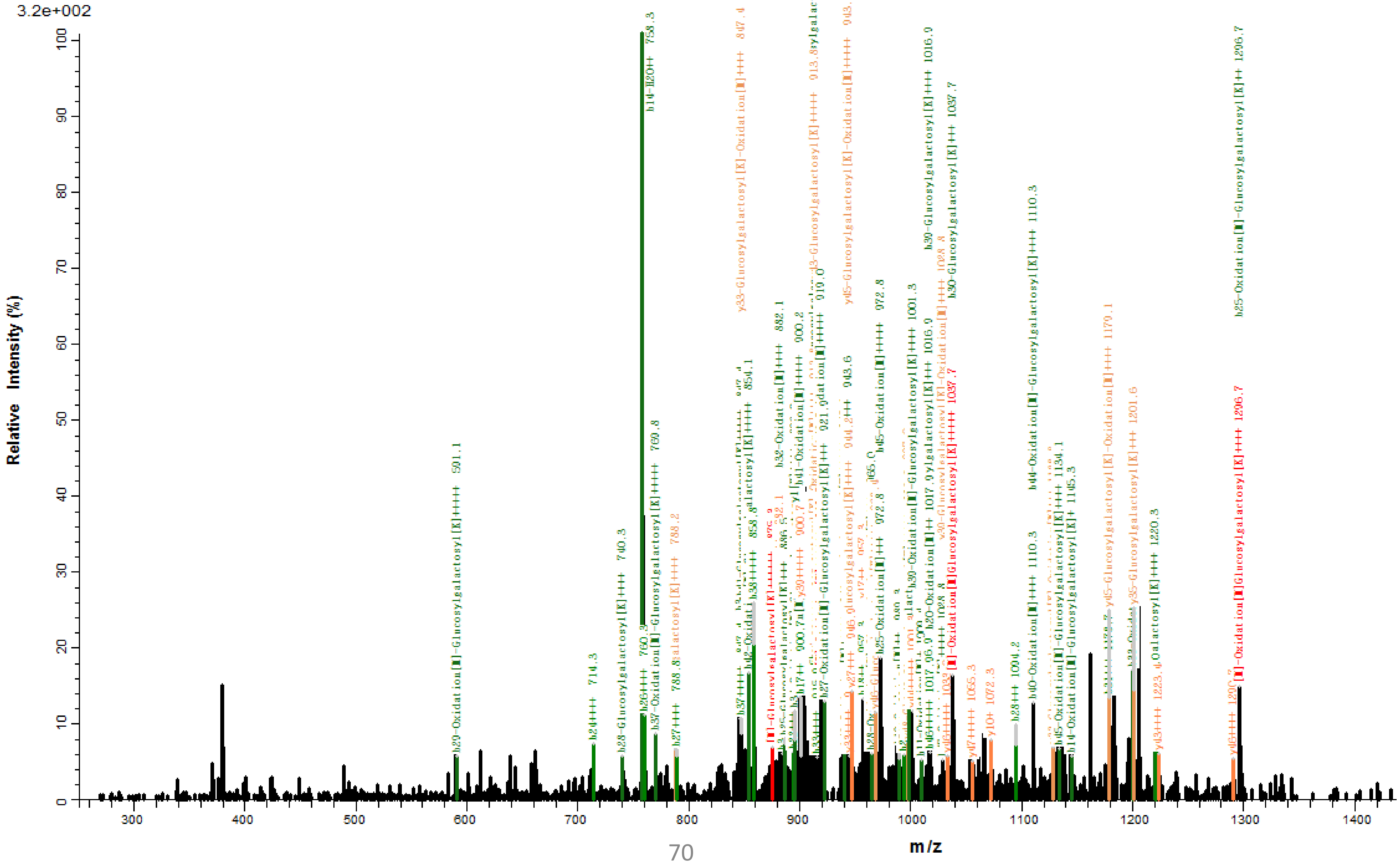


m/z= 929.8210 charge=+6

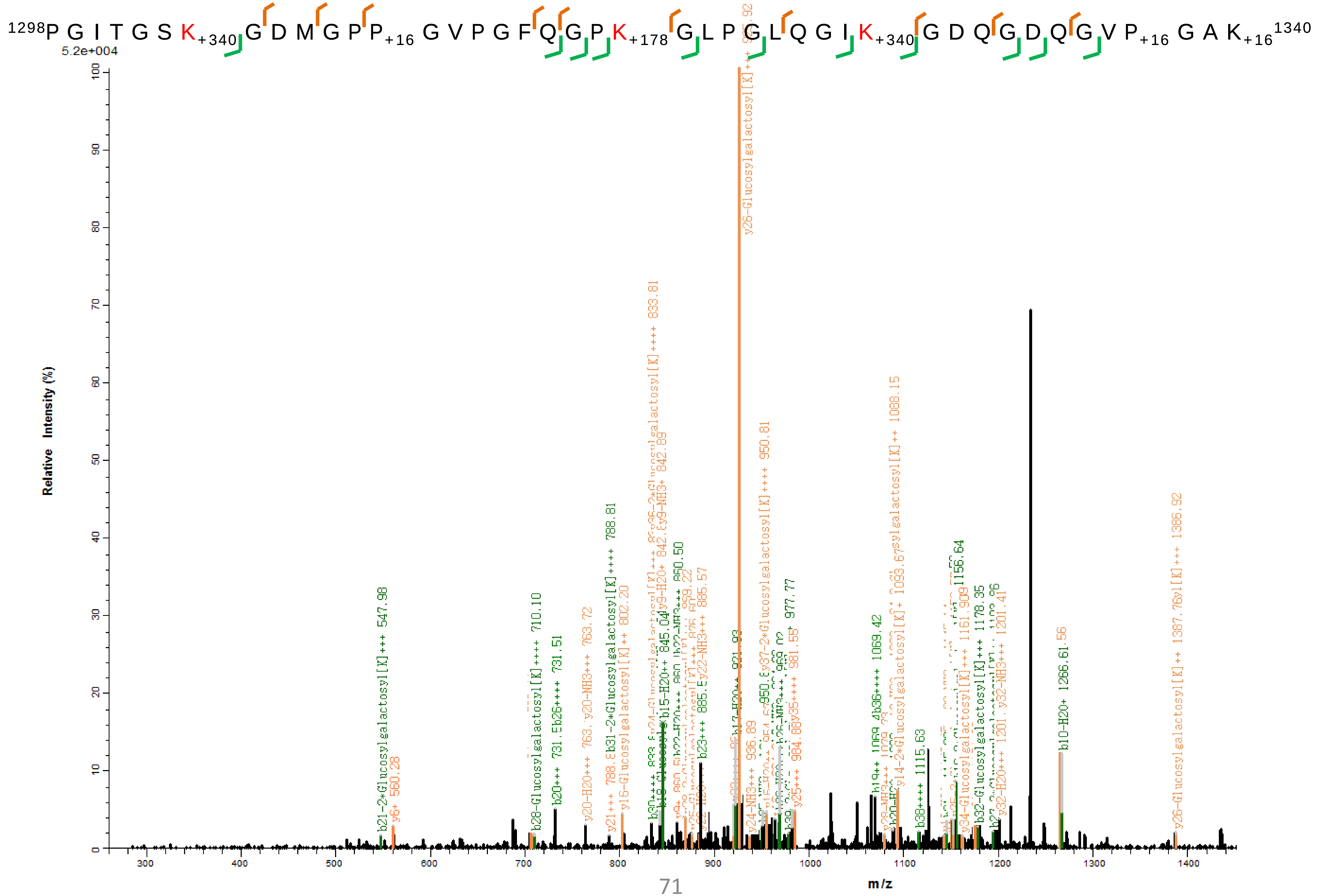
PSM showing the identification of Galactosyl-HyK^{1319,1328}

Human col4a1 chain

^{1307}M G P P$_{+16}$ G V P$_{+16}$ G F Q G P K$_{+178}$ G L P G L Q G I K$_{+340}$ G D Q G D Q G V P G A K$_{+178}$ G L P G P P$_{+16}$ G P P$_{+16}$ G P Y D I I K^{1356}



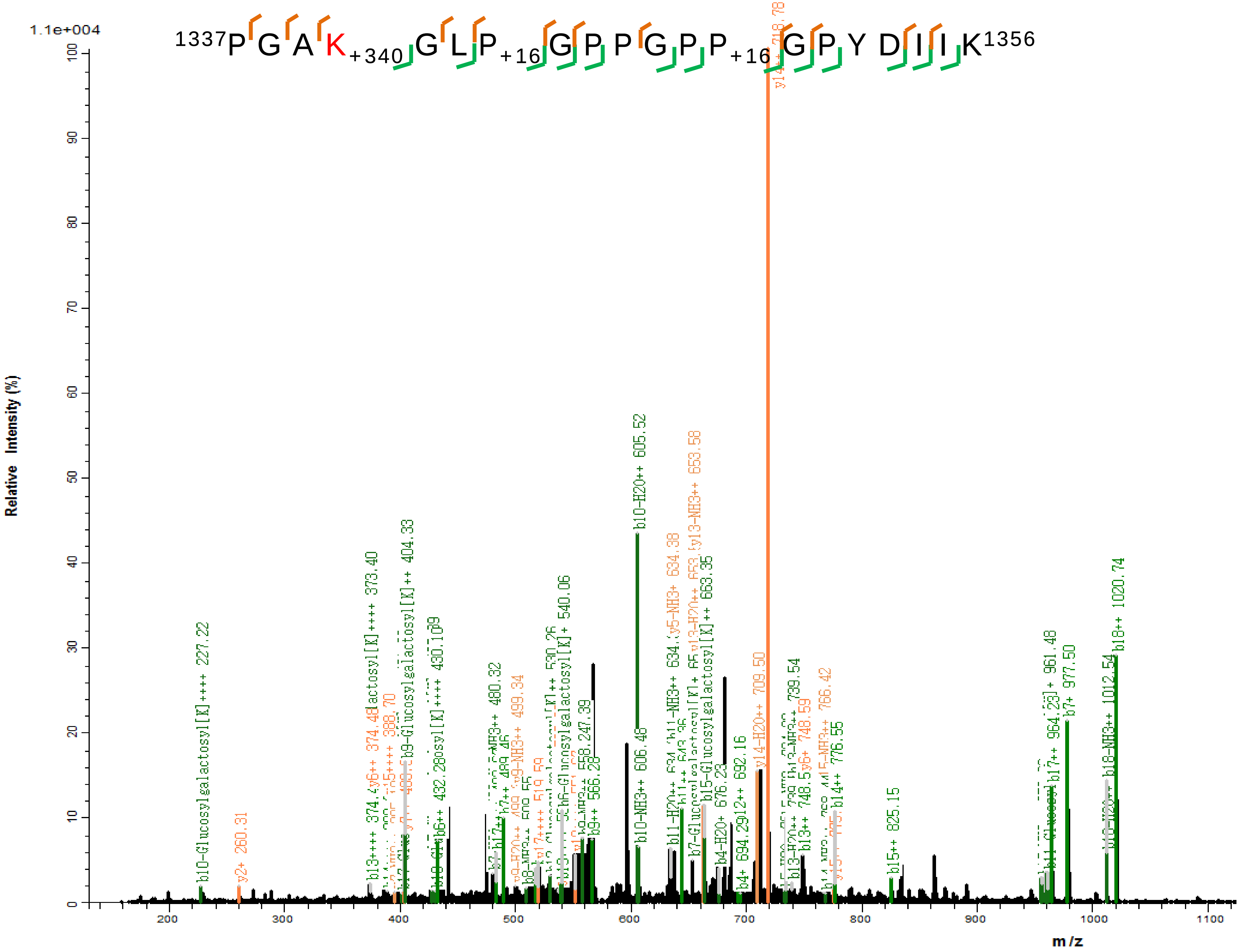
m/z= 992.8703 charge=+5 PSM showing the identification of Glucosylgalactosyl-HyK^{1304,1328} and Galactosyl-HyK^{1319} Human col4a1 chain



m/z= 575.8043 charge=+4

PSM showing the identification of Glucosylgalactosyl-HyK^{1340}

Human col4a1 chain

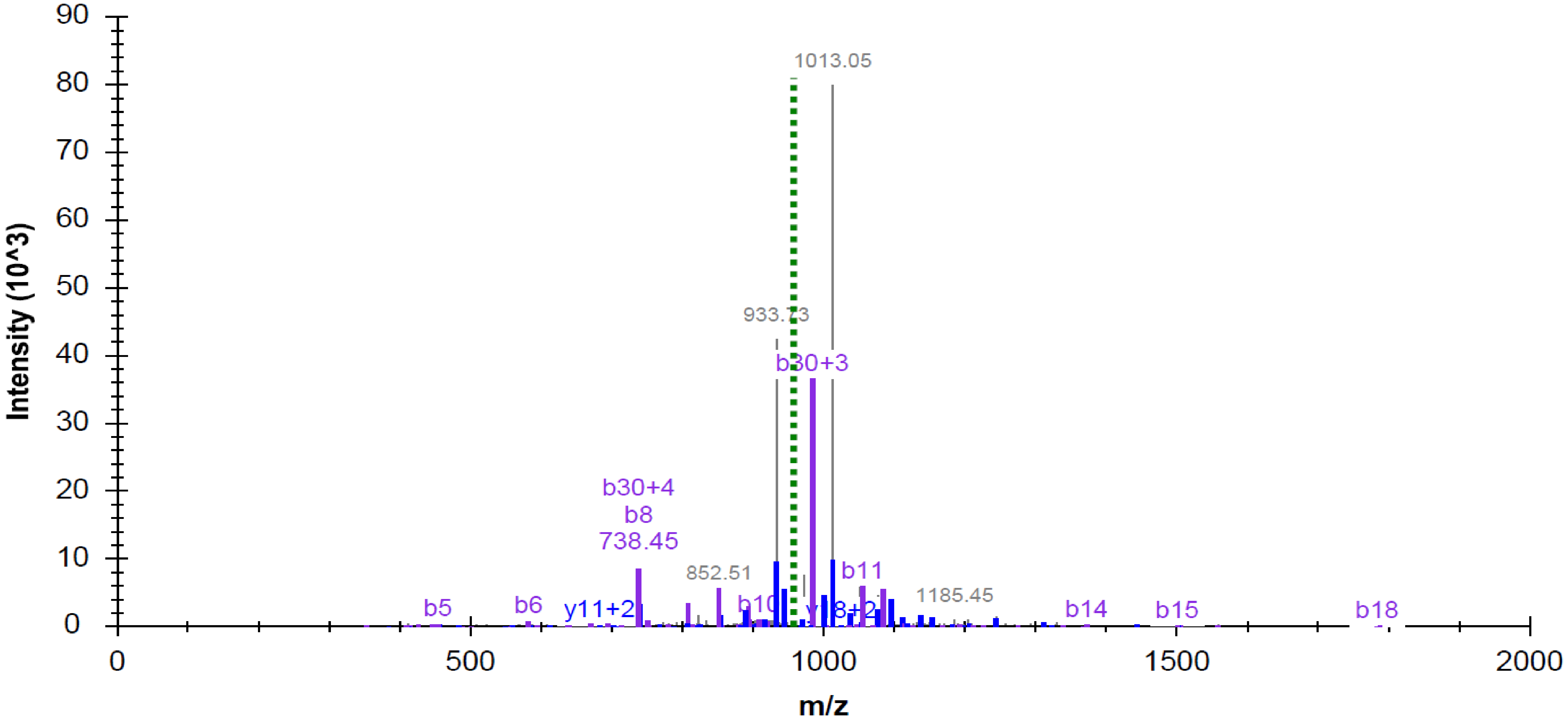


m / z = 957.6498(+5)

PSM showing the identification of HyP^{71} and 3-HyP^{74}

Human col4a1 chain

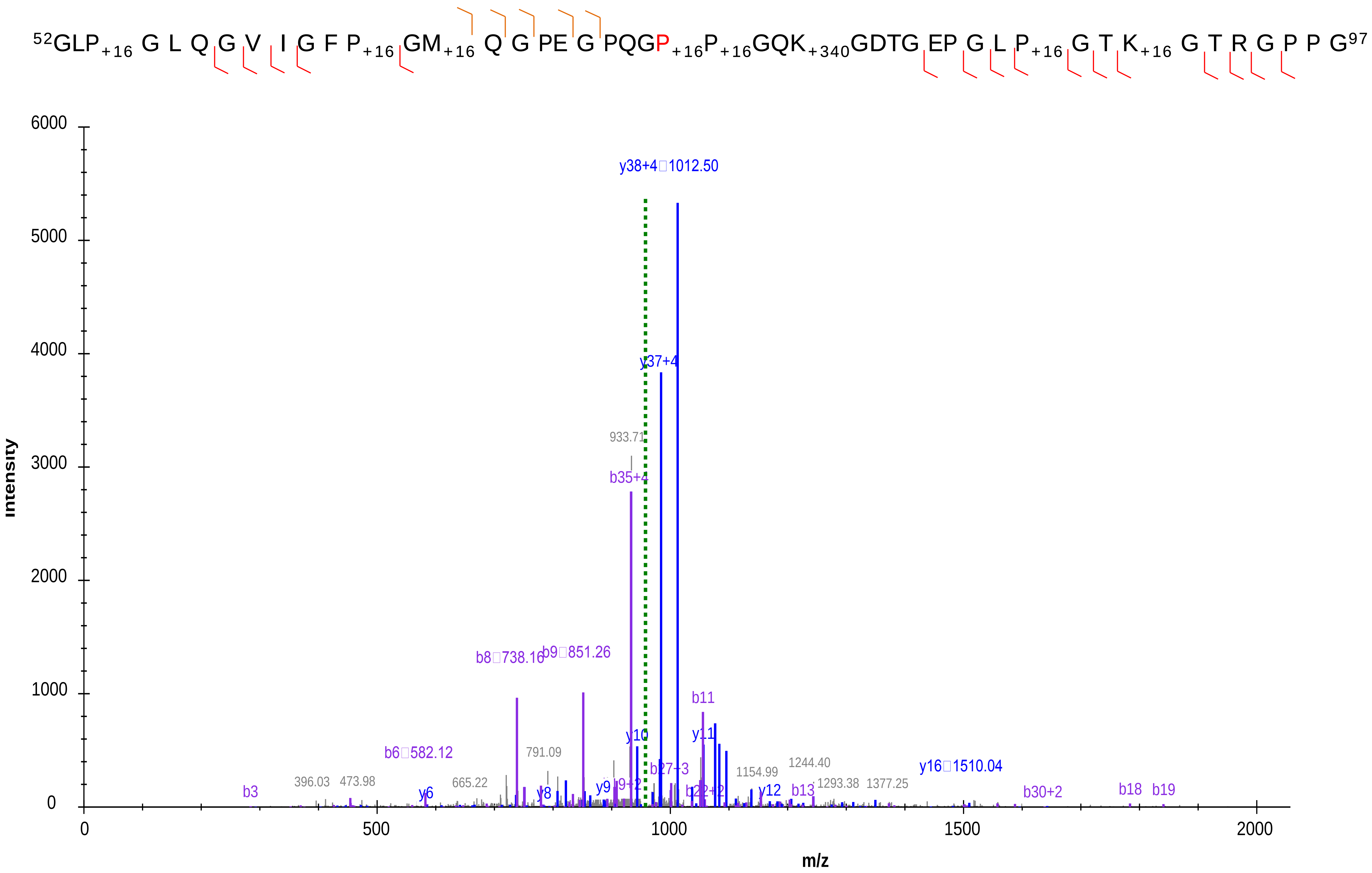
^{52}G L P$_{+16}$ G L Q G V I G F P$_{+16}$ G M Q G P E G P$_{+16}$ Q G P$_{+16}$ P$_{+16}$ G Q K G D T G E P$_{+16}$ G L P G T K$_{+340}$ G T R G P P G^{97}



m / z = 957.63(+6)

PSM showing the identification of 3-HyP^{74}

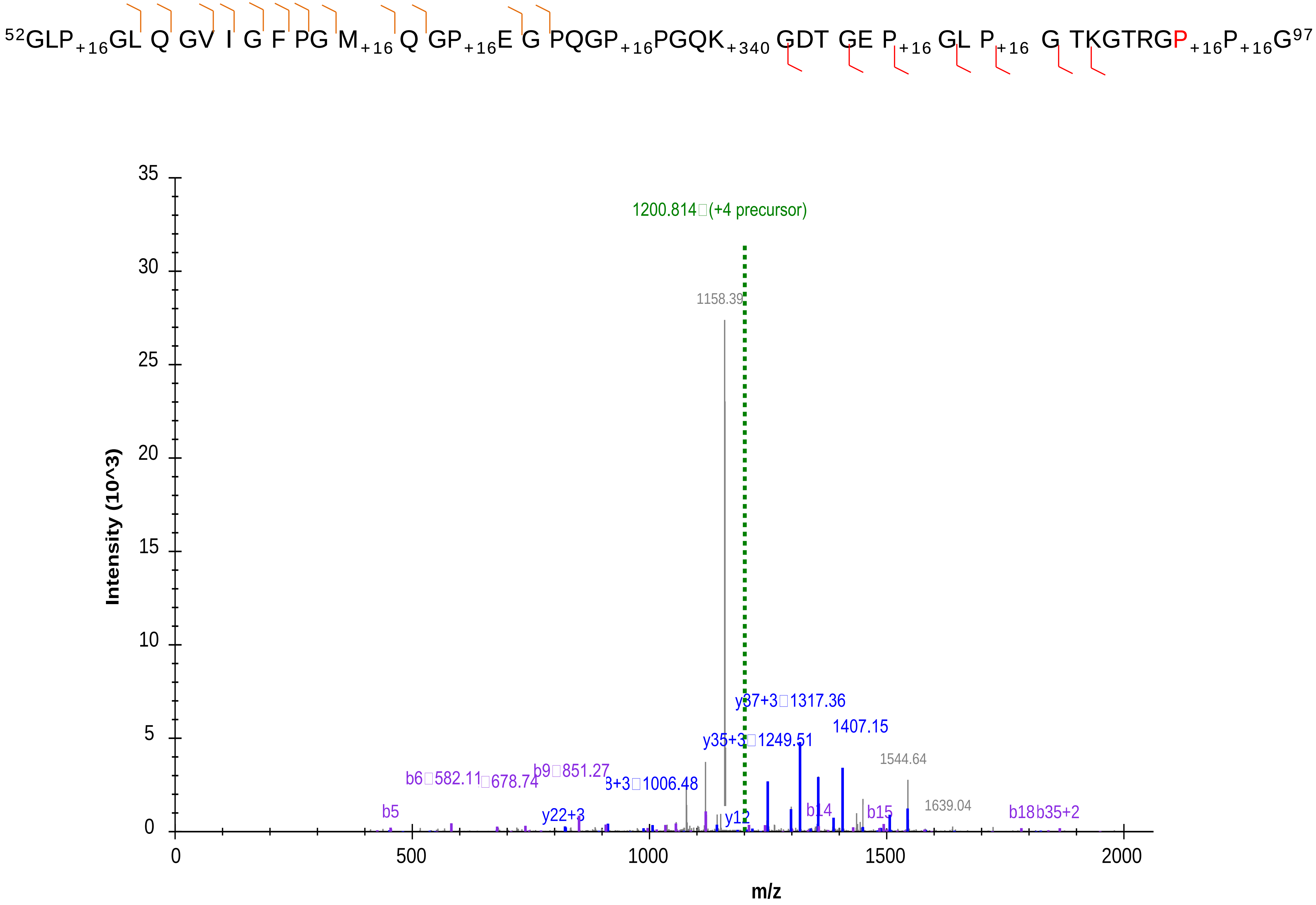
Human col4a1 chain



m/z= 1200.8142 (+4)

PSM showing the identification of 3-HyP^{95}

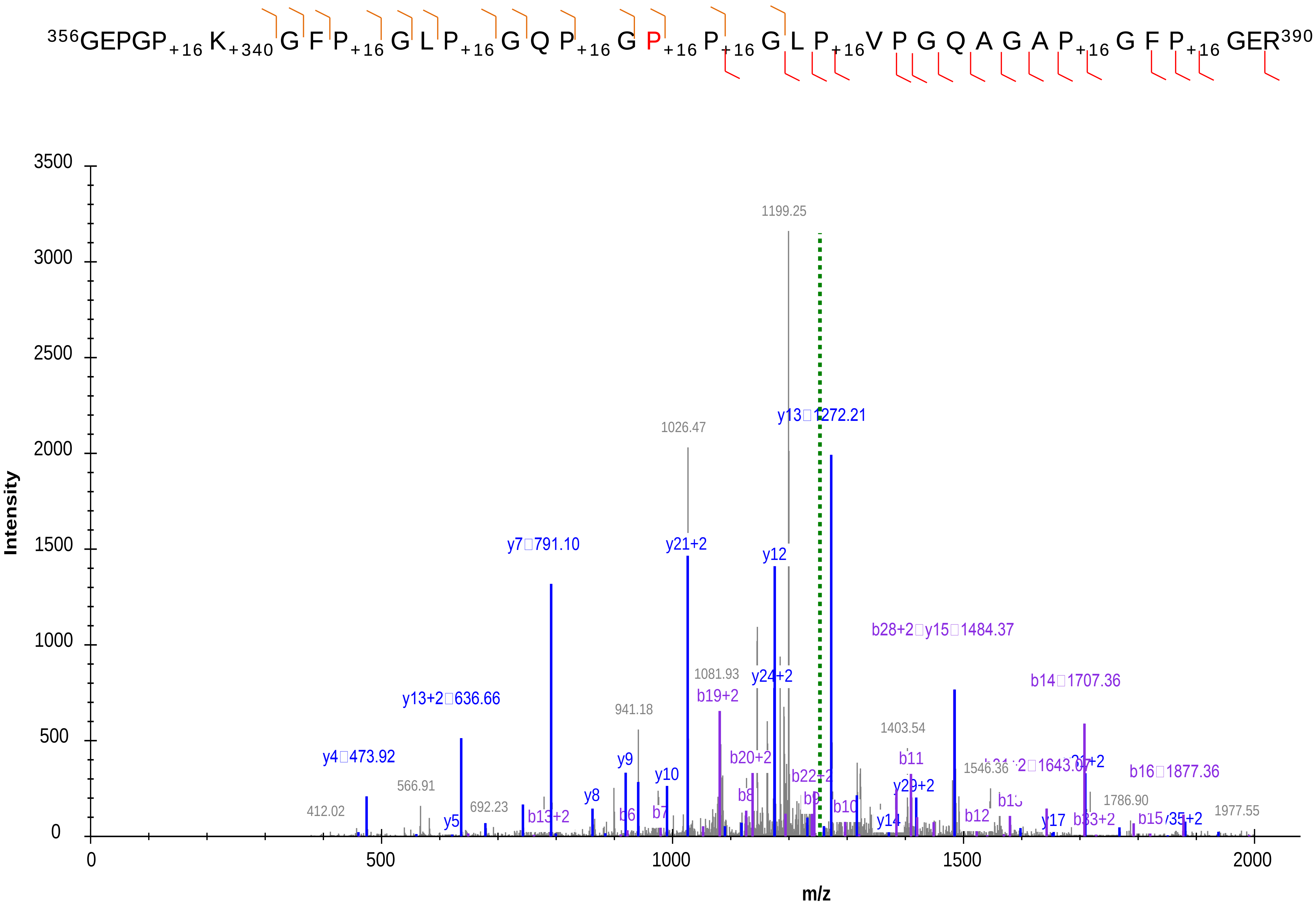
Human col4a1 chain



m/z= 1253.2531 (+3)

PSM showing the identification of 3-HyP^{372}

Human col4a1 chain

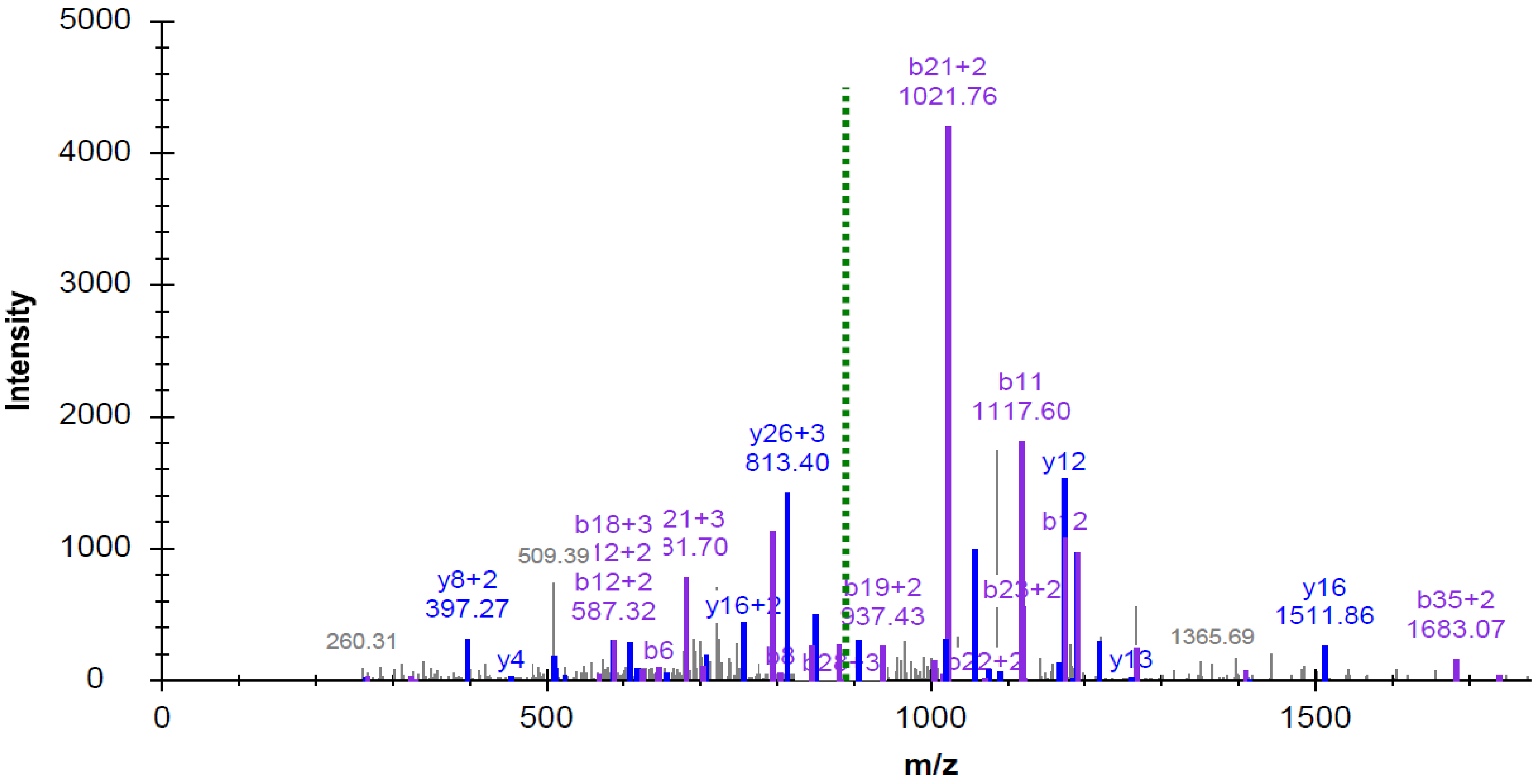


m/z= 889.1826(+4)

PSM showing the identification of 3-HyP^{421}

Human col4a1 chain

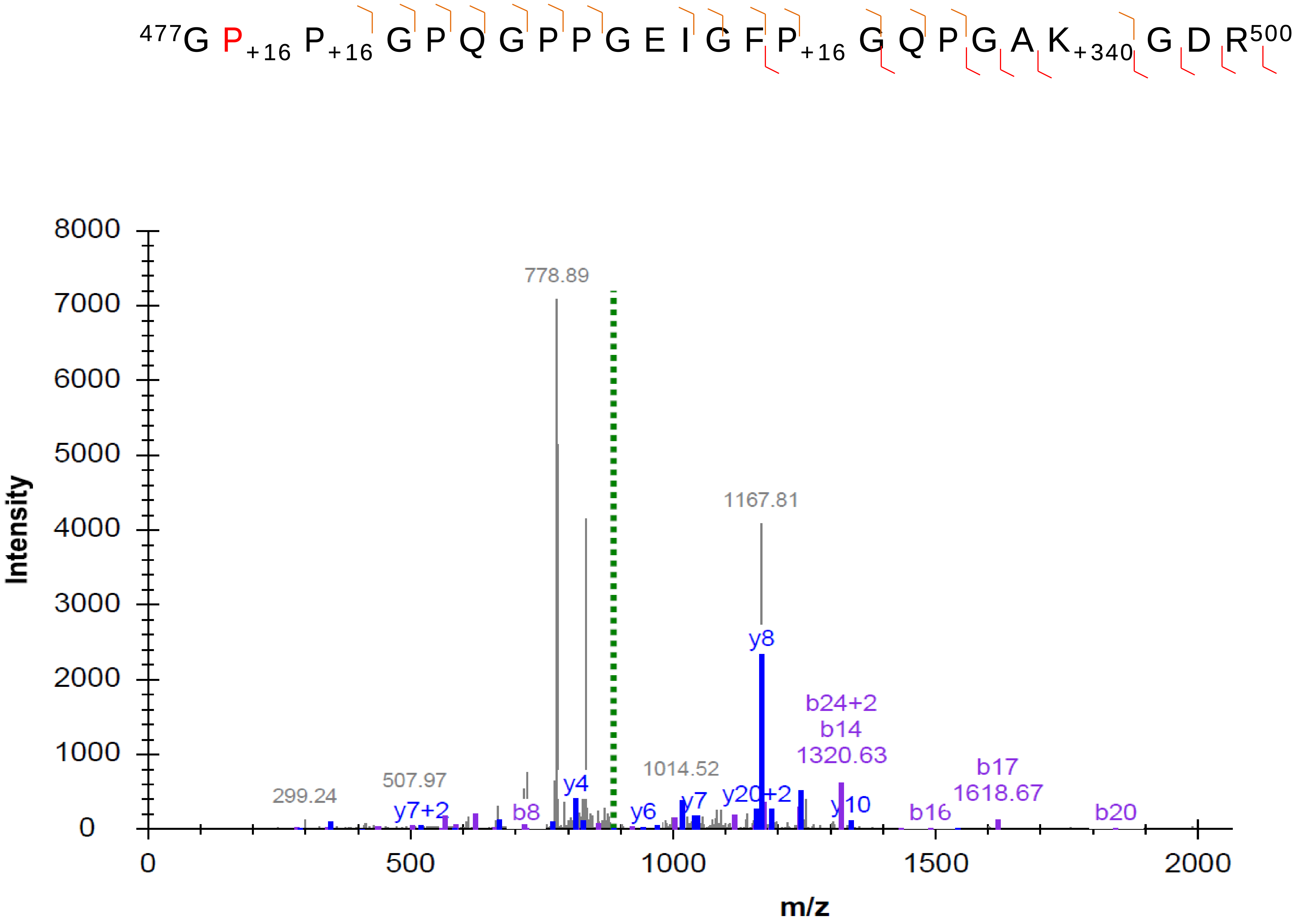
^{394}G D R G F P$_{+16}$ G T S L P$_{+16}$ G P$_{+16}$ S G R D G L P$_{+16}$ G P P G S P G P$_{+16}$ P$_{+16}$ G Q P G Y T N G^{430}



m/z= 886.748 (+3)

PSM showing the identification of 3-HyP478

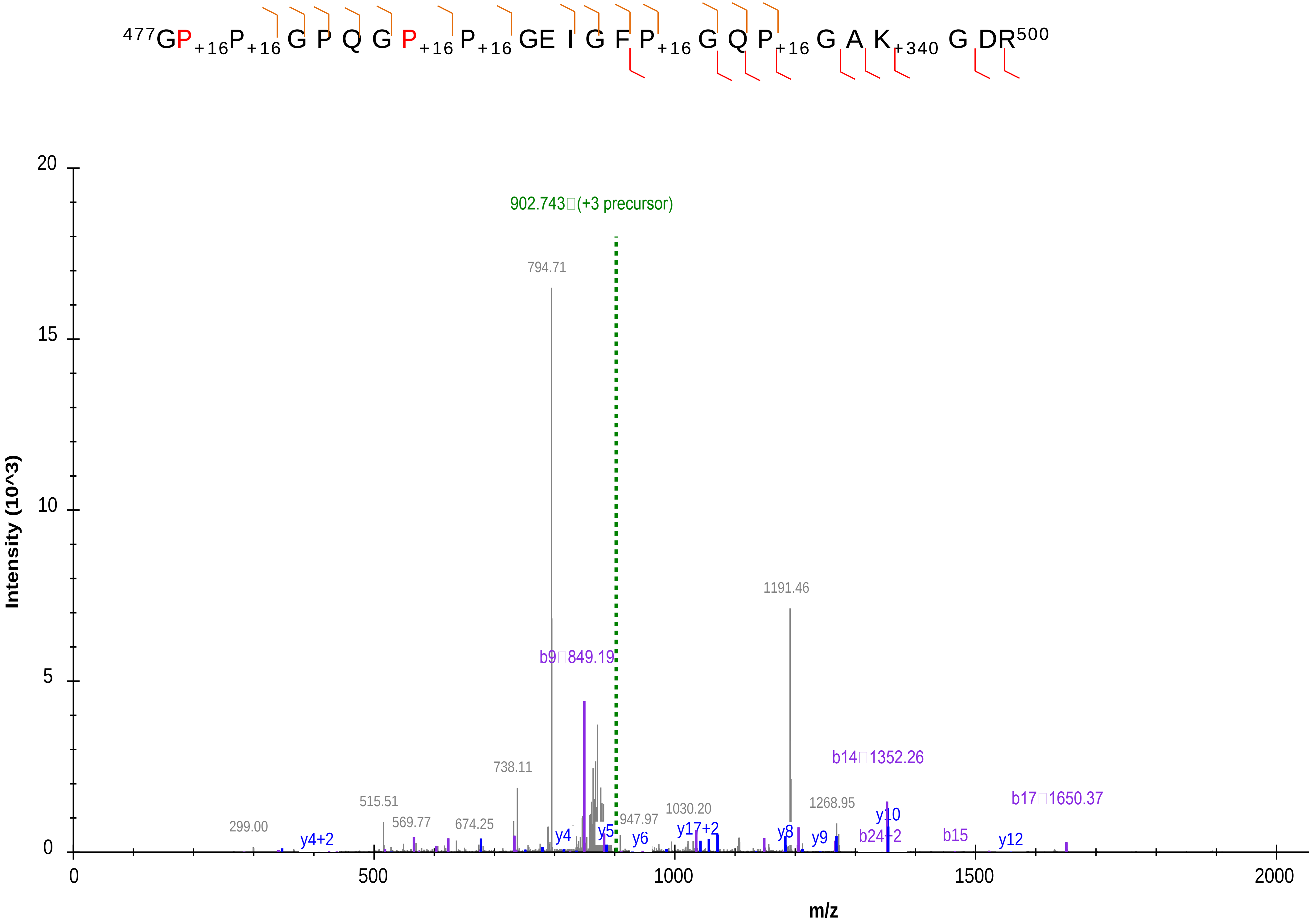
Human col4a1 chain



m/z= 902.7432 (+3)

PSM showing the identification of 3-HyP^{478,484}

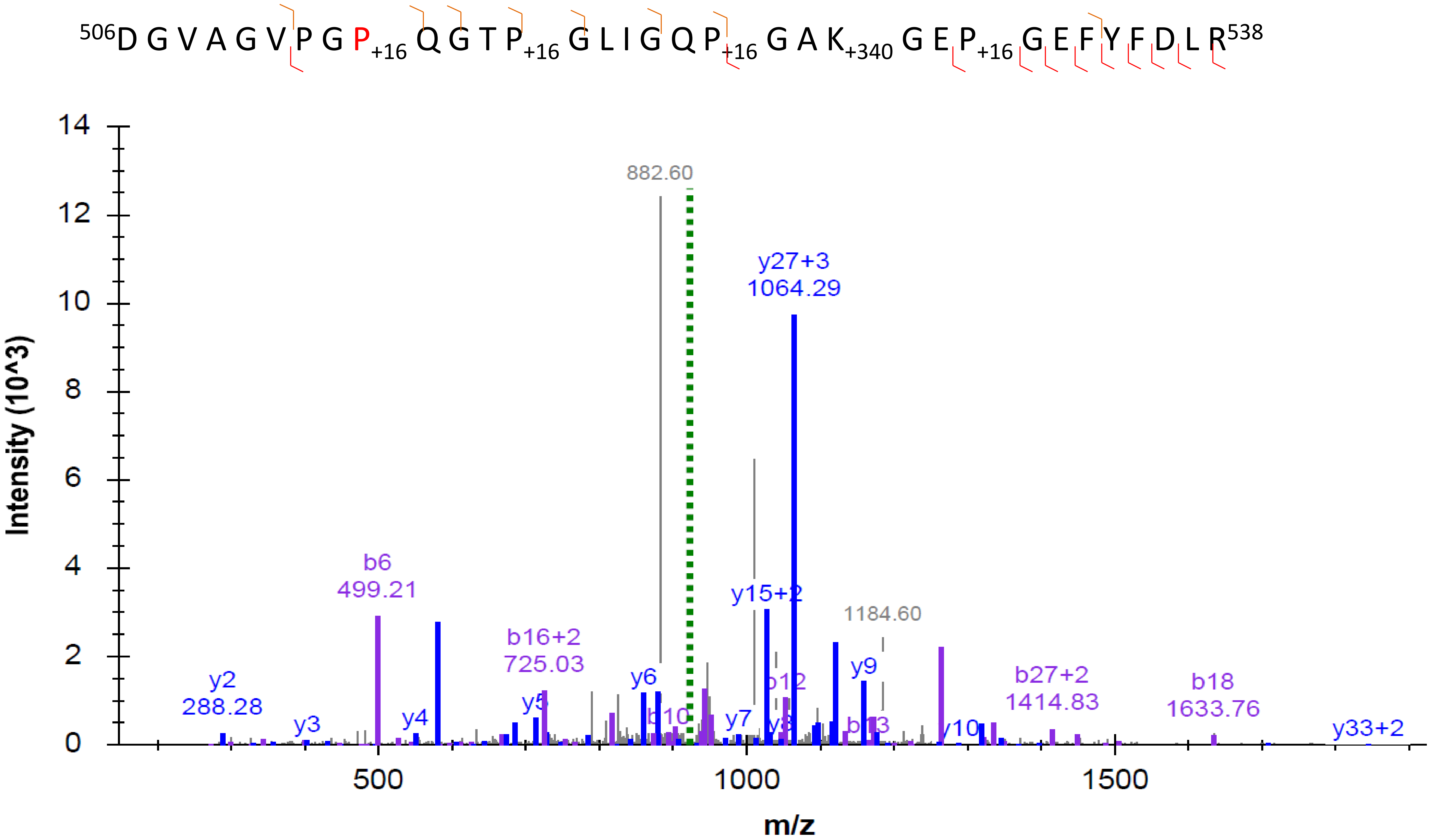
Human col4a1 chain



m/z= 922.687 (+4)

PSM showing the identification of HyP^{514}

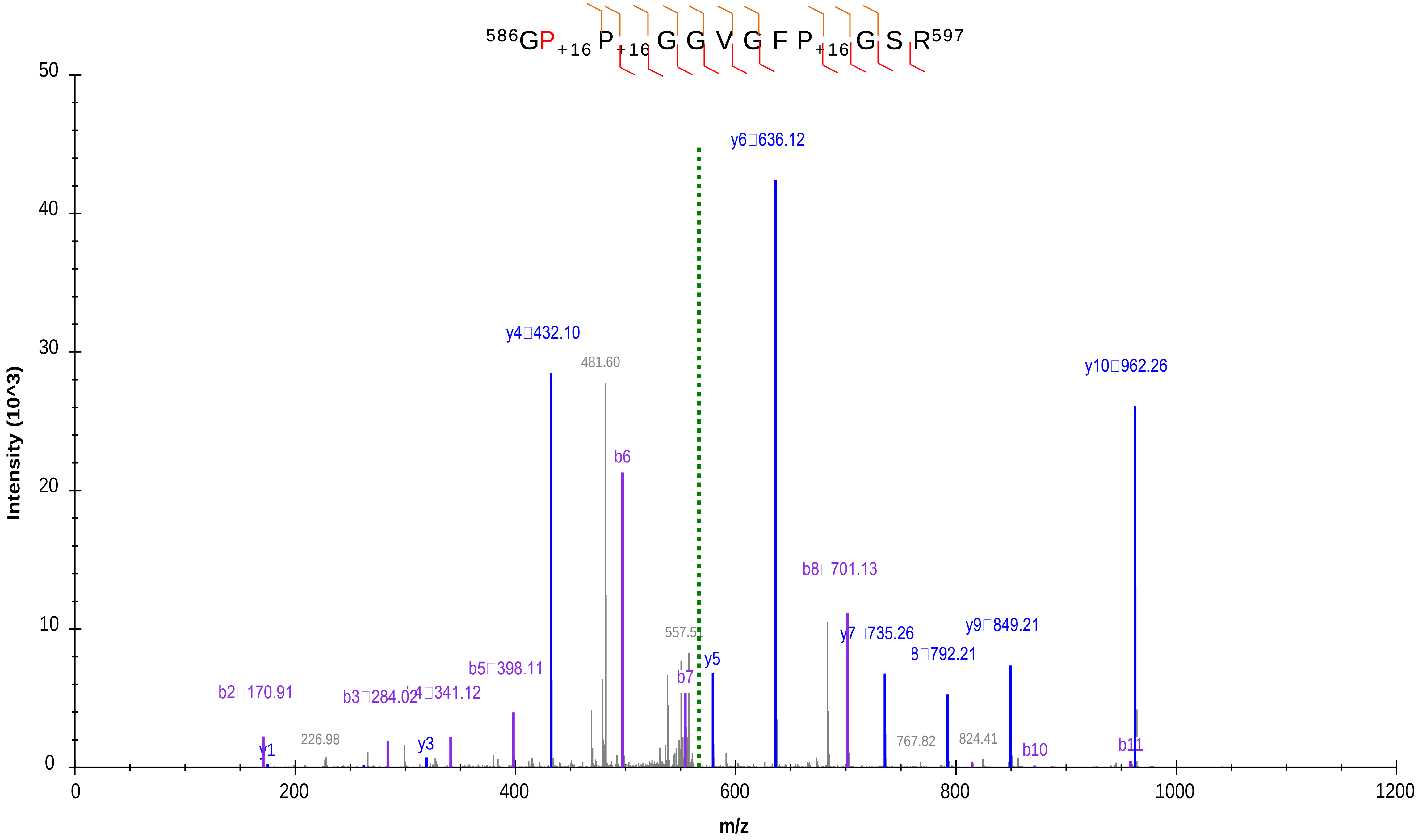
Human col4a1 chain



m/z= 566.7737 (+2)

PSM showing the identification of 3-HyP^{587}

Human col4a1 chain

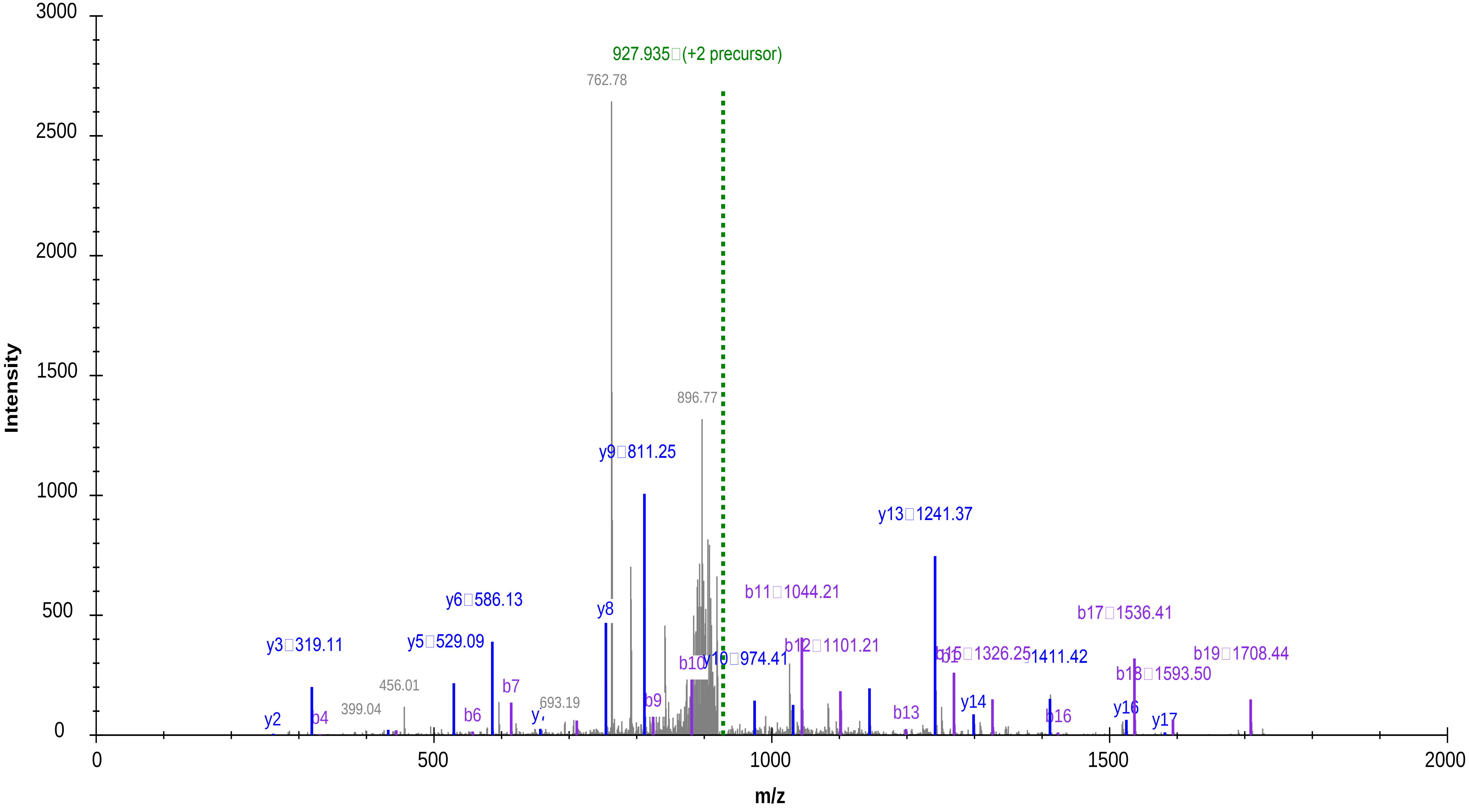


m/z= 927.9346(+2)

PSM showing the identification of 3-HyP^{602}

Human col4a1 chain

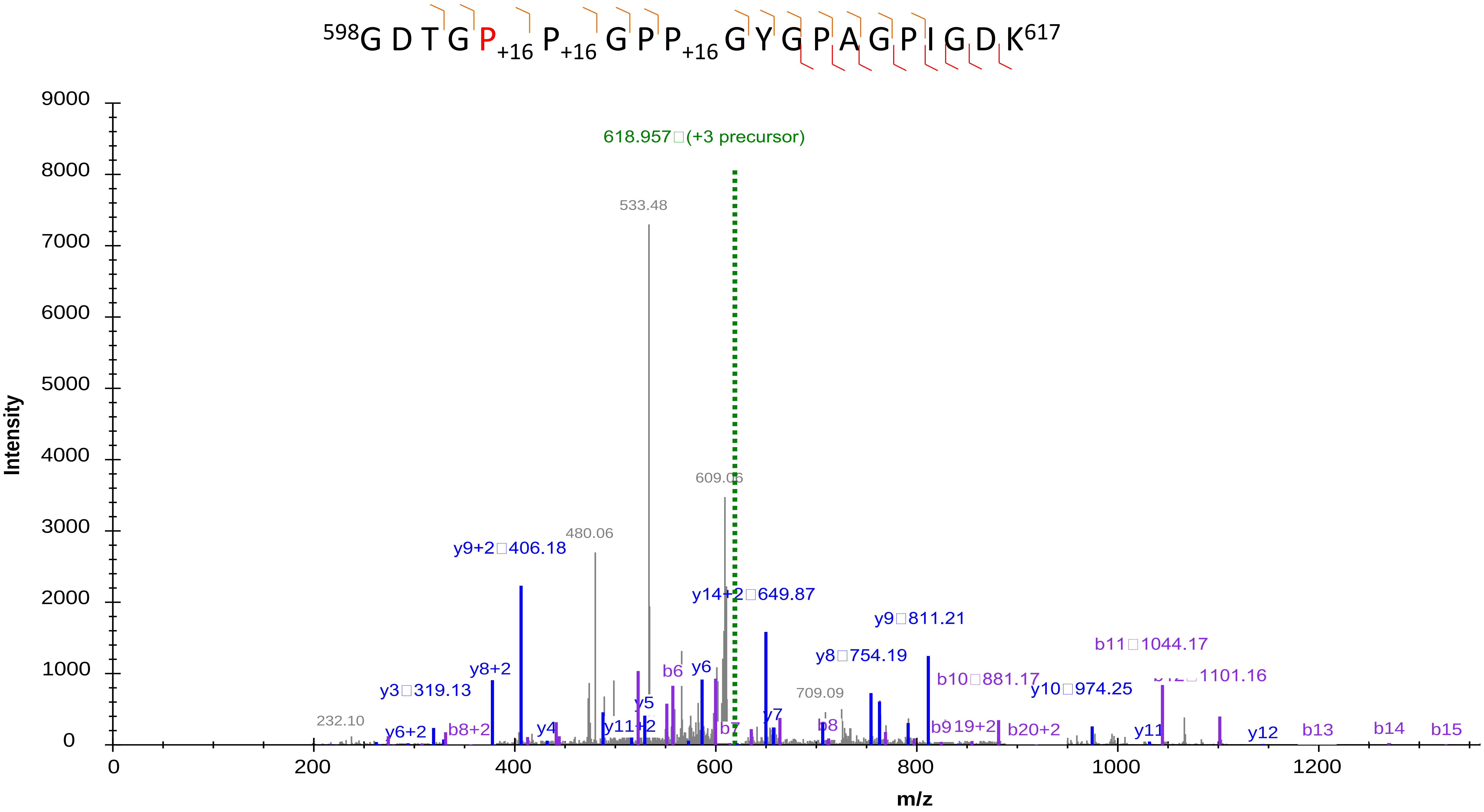
^{598}GDTG P$_{+16}$ P$_{+16}$ G P P$_{+16}$ G Y G P A G P I G D K^{617}



m/z= 618.9572(+3)

PSM showing the identification of 3-HyP^{602}

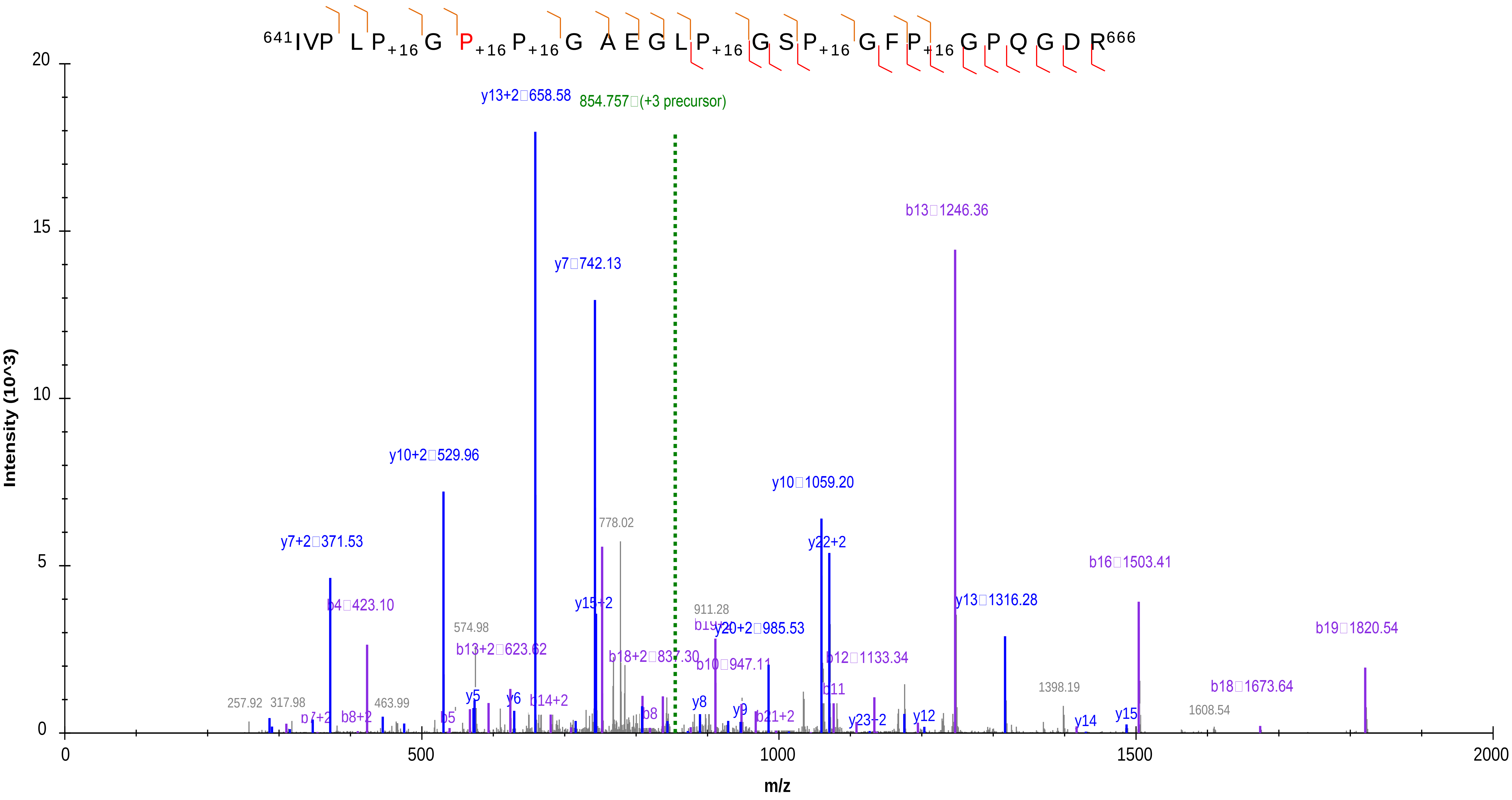
Human col4a1 chain



m/z= 854.7567 (+3)

PSM showing the identification of 3-HyP^{647}

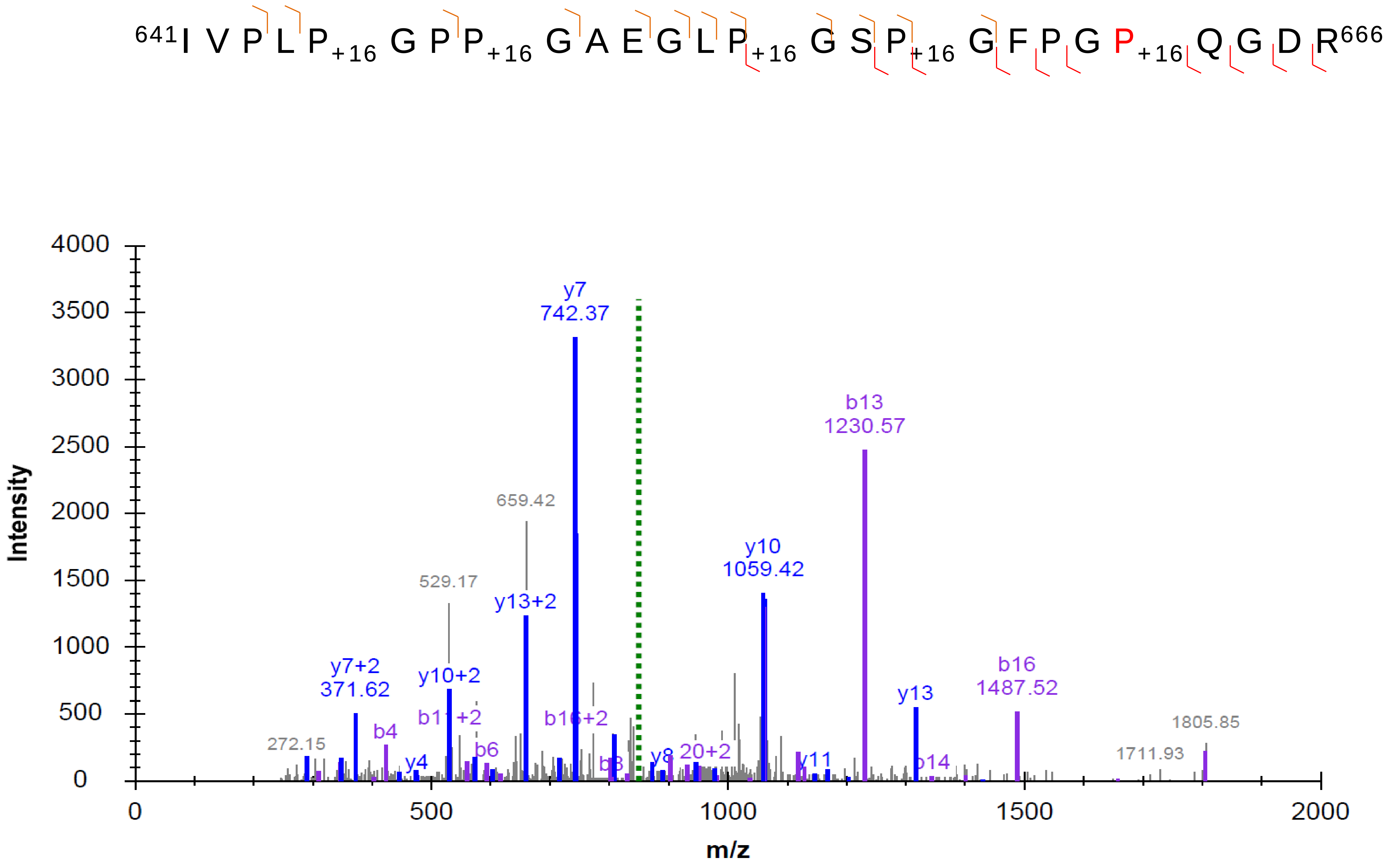
Human col4a1 chain



m/z= 849.429 (+3)

PSM showing the identification of HyP^{662}

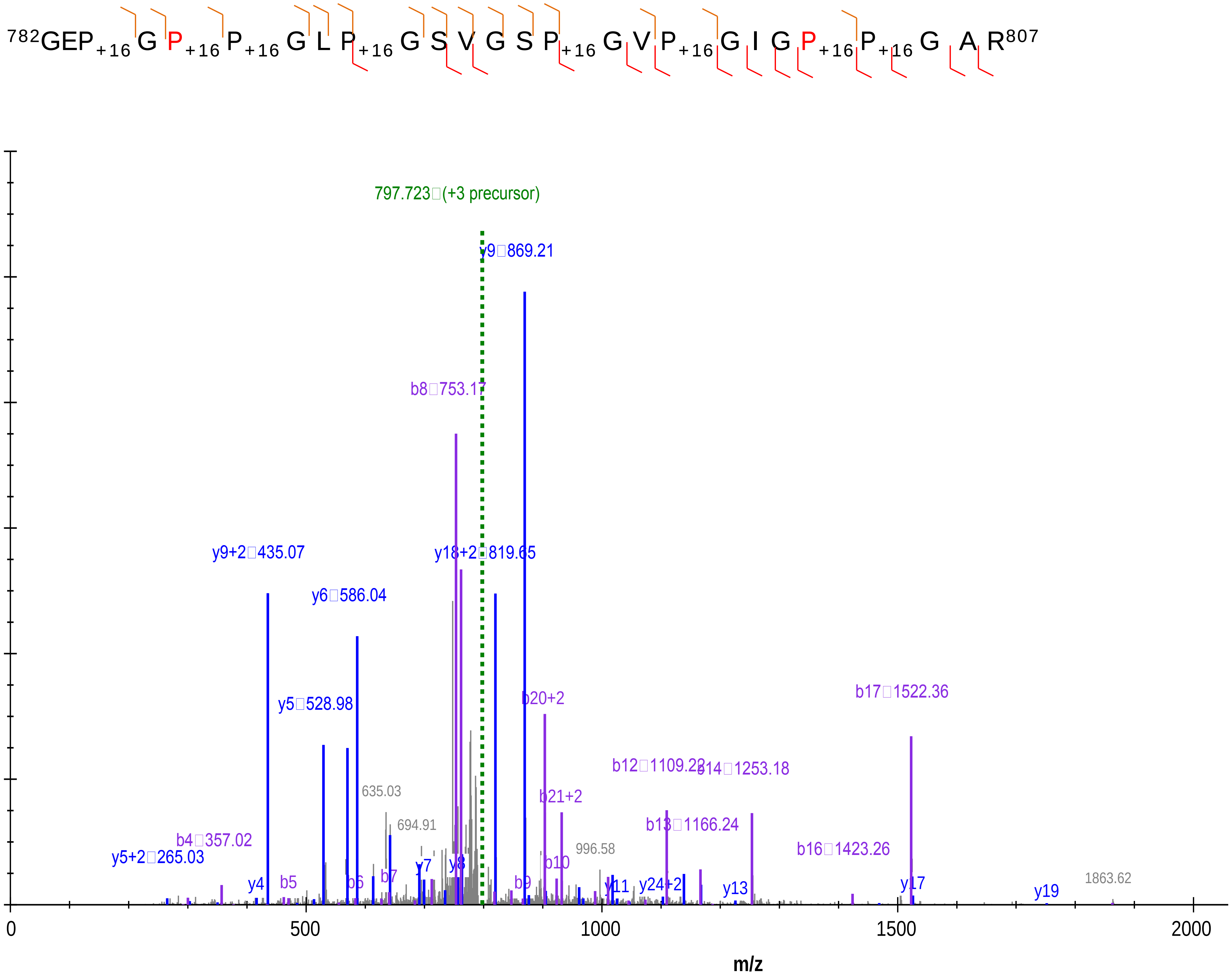
Human col4a1 chain



m/z= 797.7227 (+3)

PSM showing the identification of 3-HyP^{786,803}

Human col4a1 chain

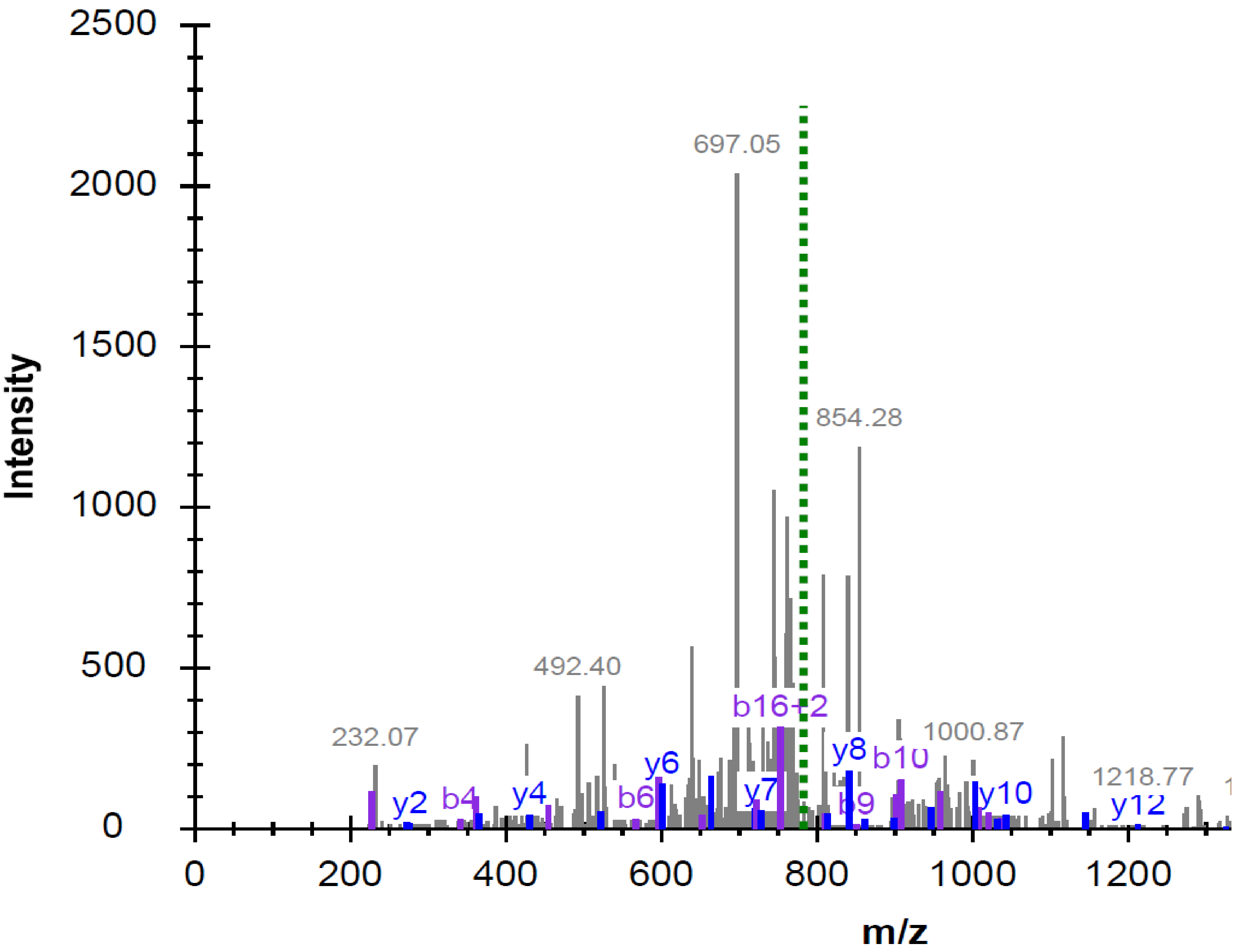


m/z= 782.3984 (+3)

PSM showing the identification of 3-HyP^{1367}

Human col4a1 chain

^{1357}G E P G L P$_{+16}$ G P E G P$_{+16}$ P$_{+16}$ G L K$_{+16}$ G L Q G L P G P$_{+16}$ K$_{+16}$^{1380}

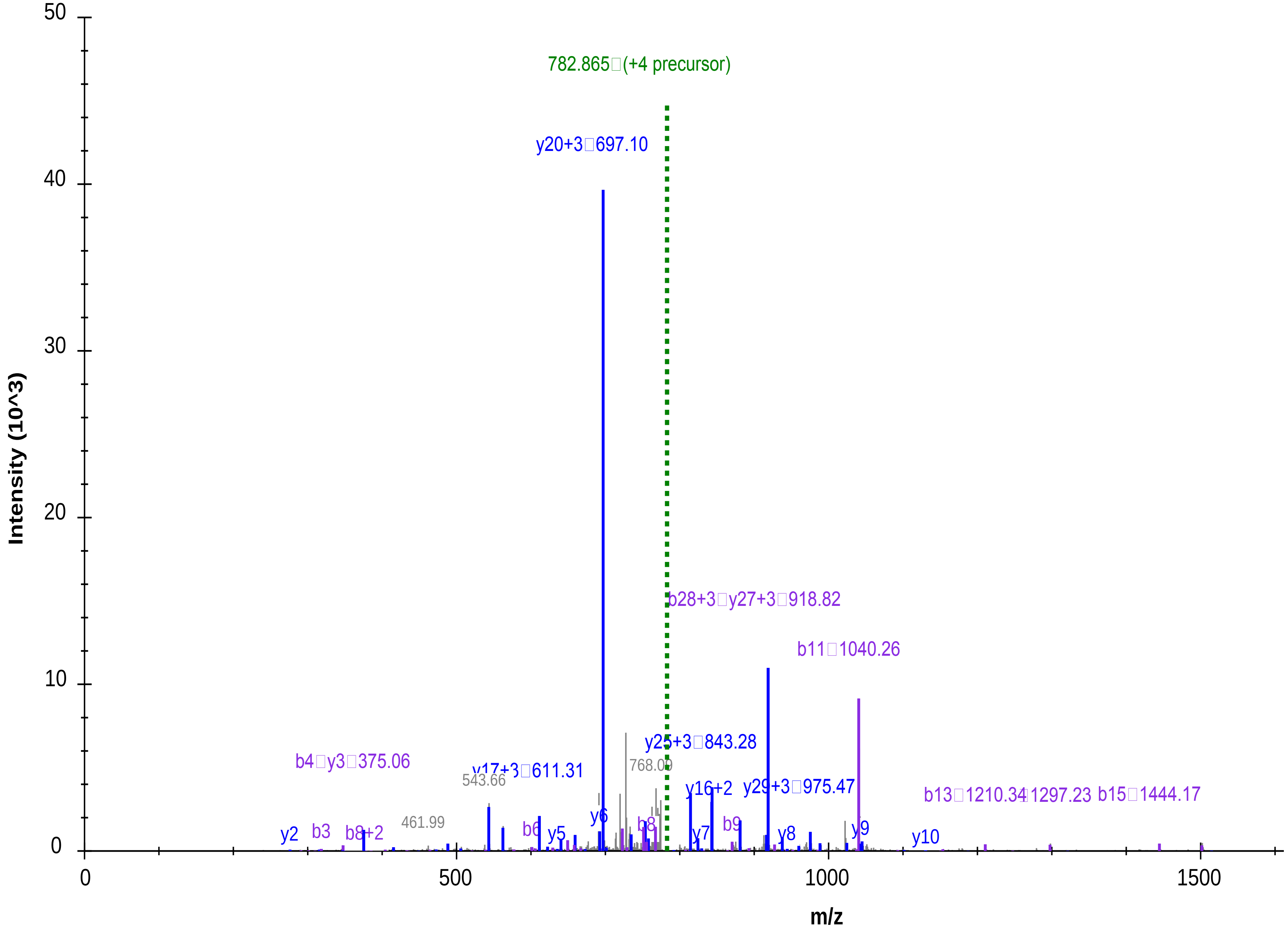


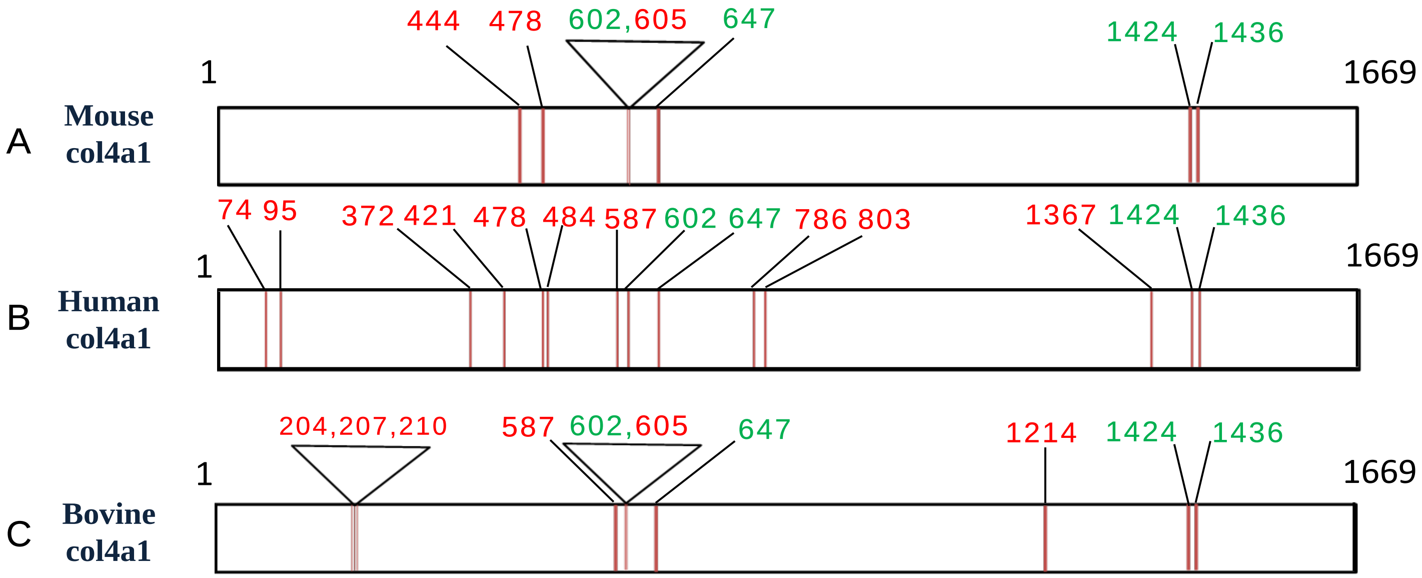
m / z = 782.8649 (+4)

PSM showing the identification of 3-HyP^{1424,1436}

Human col4a1 chain

^{1420}GFP$_{+16}$ G P$_{+16}$ P$_{+16}$ G P D G L P$_{+16}$ G S M$_{+16}$ G P$_{+16}$ P$_{+16}$ GTP$_{+16}$ S V D H G F L V T R^{1450}





Supporting figure S4: Comparison of 3-HyP sites identified in col4a1 from mouse, human and bovine. The horizontal box represents the full length col4a1 sequence and the vertical red lines indicate the corresponding 3-HyP sites. The information for 3 HyP sites from bovine lens capsule (C) were summarized from Pokidysheva et al[3] and Hudson et al[4]. The 3-HyP sites marked with green represents the conserved sites among mouse, human and bovine.

1	MFSFVDLRLL	LLLGATALLT	HGQEDIPEVS	CIHNGLRVPN	GETWKPEVCL	ICICHNGTAV
61	CDDVQCNEEL	DCPNPQRREG	ECCAFCPEEY	VSPNSEDVGV	EGPKGDPGPQ	GPRGPVGPPG
121	RDGIPGQPGL	PGPPGPPGPP	GPPGLGGNFA	SQMSYGYDEK	SAGVSVPGPM	GPSGPRGLPG
181	PPGAPGPQGF	QGPPGEPGEP	GGSGPMGPRG	PPGPPGKNGD	DGEAGKPGRP	GERGPPGPQG
		1	2	3		
241	ARGLPGTAGL	PGM**K**GHRGFS	GLDGA**K**GDAG	PAGPKGEPGS	PGENGAPGQM	GPRGLPGERG
					4	
301	RPGPPGTAGA	RGNDGAVGAA	GPPGPTGPTG	PPGFPGAVGA	**K**GEAGPQGAR	GSEGPQGVRG
			5			6
361	EPGPPGPAGA	AGPAGNPGAD	GQPGAKGANG	APGIAGAPGF	PGARGPSGPQ	GPSGPPGPKG
		7 8				
421	NSGEPGAPGN	KGDTGAKGEP	GATGVQGPPG	PAGEEGKRGA	RGEPGPSGLP	GPPGERGGPG
		9	10		11	
481	SRGFPGADGV	AGPKGPSGER	GAPGPAGPKG	SPGEAGRPGE	AGLPGAKGLT	GSPGSPGPDG
541	KTGPPGPAGQ	DGRPGPAGPP	GARGQAGVMG	FPGPKGTAGE	PGKAGERGLP	GPPGAVGPAG
601	KDGEAGAQGA	PGPAGPAGER	GEQGPAGSPG	FQGLPGPAGP	PGEAGKPGEQ	GVPGDLGAPG
				12		
661	PSGARGERGF	PGERGVQGPP	GPAGPRGNNG	APGNDGAKGD	TGAPGAPGSQ	GAPGLQGMPG
		13 14			15	
721	ERGAAGLPGP	**K**GDRGDAGPK	GADGSPGKDG	ARGLTGPIGP	PGPAGAPGD**K**	GEAGPSGPPG
				16	17	
781	PTGARGAPGD	RGEAGPPGPA	GFAGPPGADG	QPGAKGEPGD	TGVKGDAGPP	GPAGPAGPPG
		18				19
841	PIGNVGAPGP	**K**GPRGAAGPP	GATGFPGAAG	RVG**P**PGPSGN	AGPPGPPGPV	GKEGGKGPRG
			20			
901	ETGPAGRPGE	VGPPGPPGPA	GEKGSPGADG	PAGSPGTPGP	QGIAGQRGVV	GLPGQRGERG
961	FPGLPGPSGE	PGKQGPSGSS	GERGPPGPMG	PPGLAGPPGE	SGREGSPGAE	GSPGRDGAPG
	21					
1021	AKGDRGETGP	AGPPGAPGAP	GAPGPVGPAG	KNGDRGETGP	AGPAGPIGPA	GARGPAGPQG
	22	23				
1081	PRGDKGETGE	QGDRGIKGHR	GFSGLQGPPG	SPGSPGEQGP	SGASGPAGPR	GPPGSAGSPG
1141	KDGLNGLPGP	IG**P**PGPRGRT	GDSGPAG**P**PG	**P**PG**P**PGPPGP	PSGGYDFSFL	PQPPQEKSQD
1201	GGRYYRADDA	NVVRDRDLEV	DTTLKSLSQQ	IENIRSPEGS	RKNPARTCRD	LKMCHSDWKS
1261	GEYWIDPNQG	CNLDAIKVYC	NMETGQTCVF	PTQPSVPQKN	WYISPNPKEK	KHVWFGESMT
1321	DGFPFEYGSE	GSDPADVAIQ	LTFLRLMSTE	ASQNITYHCK	NSVAYMDQQT	GNLKKALLLQ
1381	GSNEIELRGE	GNSRFTYSTL	VDGCTSHTGT	WGKTVIEYKT	TKTSRLPIID	VAPLDIGAPD
1441	QEFGLDIGPA					

Supporting figure S5: Summary of prolyl 3-hydroxylation and glycosylation sites in mouse col1a1[4-7]. The matured col1a1 sequences are shown in black. The 23 potential lysine sites for O-glycosylation are numbered. Bold '**K**' indicates O-glycosylated residue in the Yaa position of Gly-Xaa-Yaa motif. Bold '**P**' indicates 3 hydroxyproline in the Xaa position of Gly-Xaa-HyP.

References:

1. Sun, R.X.; Dong, M.Q.; Song, C.Q.; Chi, H.; Yang, B.; Xiu, L.Y.; Tao, L.; Jing, Z.Y.; Liu, C.; Wang, L.H.; Fu, Y.; He, S.M. Improved peptide identification for proteomic analysis based on comprehensive characterization of electron transfer dissociation spectra. J Proteome Res 2010, 9, 6354–6367.
2. Ma, Z. Q.; Chambers, M. C.; Ham, A.J.; Cheek, K.L.; Whitewell, C.W.; Aerni, H.R.; Schilling, B.; Miller, A.W.; Caprioli, R.M.; Tabb, D. L. ScanRanker: Quality assessment of tandem mass spectra via sequence tagging. J Proteome Res 2011, 10, 2896-2904.
3. Pokidysheva, E.; Boudko, S.; Vranka, J.; Zientek, K.; Maddox, K.; Moser, M.; Fassler, R.; Ware, J.; Bachinger, H. P.: Biological role of prolyl 3-hydroxylation in type IV collagen. Proc Natl Acad Sci U S A 2014, 111, 161-6.
4. Hudson, D. M.; Joeng, K. S.; Werther, R.; Rajagopal, A.; Weis, M.; Lee, B. H.; Eyre, D. R. Post-translationally abnormal collagens of prolyl 3-hydroxylase-2 null mice offer a pathobiological mechanism for the high myopia linked to human LEPREL1 mutations. J Biol Chem 2015, 290, 8613-8622.
5. Pokidysheva, E.; Zientek, K. D.; Ishikawa, Y.; Mizuno, K.; Vranka, J. A.; Montgomery, N. T.; Keene, D. R.; Kawaguchi, T.; Okuyama, K.; Bachinger, H. P. Posttranslational modifications in type I collagen from different tissues extracted from wild type and prolyl 3-hydroxylase 1 null mice. J Biol Chem 2013, 288, 24742-24752.
6. Cabral, W. A.; Perdivara, I.; Weis, M.; Terajima, M.; Blissett, A. R.; Chang, W.; Perosky, J. E.; Makareeva, E. N.; Mertz, E. L.; Leikin, S.; Tomer, K. B.; Kozloff, K. M.; Eyre, D. R.; Yamauchi, M.; Marini, J. C.: Abnormal type I collagen post-translational modification and crosslinking in a cyclophilin B KO mouse model of recessive osteogenesis imperfecta. PLoS genetics 2014, 10, e1004465.
7. Perdivara, I.; Yamauchi, M.; Tomer, K. B. Molecular Characterization of Collagen Hydroxylysine -Glycosylation by Mass Spectrometry: Current Status. Aust J Chem 2013, 66, 760-769.