Electronic Supplementary Material (ESI) for Chemical Science.

Supplementary Information

Membrane glycomics reveal heterogeneity and quantitative distribution of cell surface sialylation

Diane Dayoung Park[a,b,*], Gege Xu[a], Maurice Wong[a], Chatchai Phoomak[c], Mingqi Liu[d], Nathan E. Haigh[e], Sopit Wongkham[c], Pengyuan Yang[d], Emanual Maverakis[e], Carlito B. Lebrilla[a]

[a]Department of Chemistry, University of California, Davis, CA 95616, USA
[b]Department of Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02115, USA. E-mail: dpark4@bidmc.harvard.edu
[c]Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand
[d]Department of Chemistry, Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China
[e]Department of Dermatology, University of California, Davis School of Medicine, Sacramento, CA 95817, USA

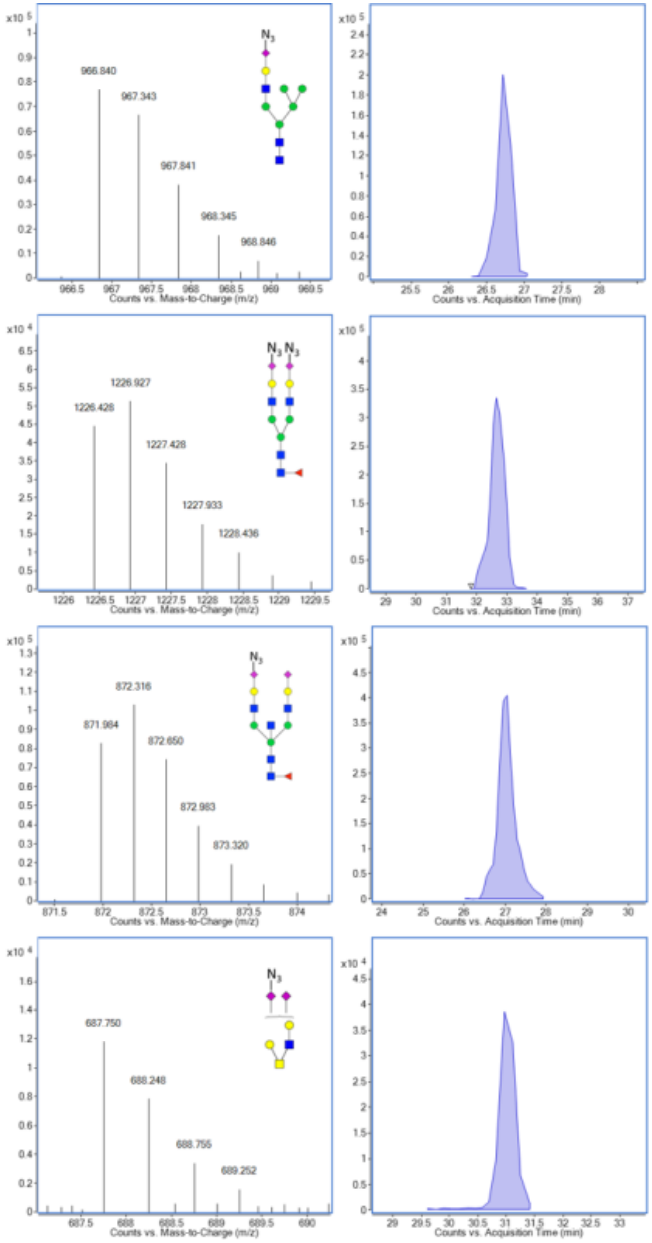
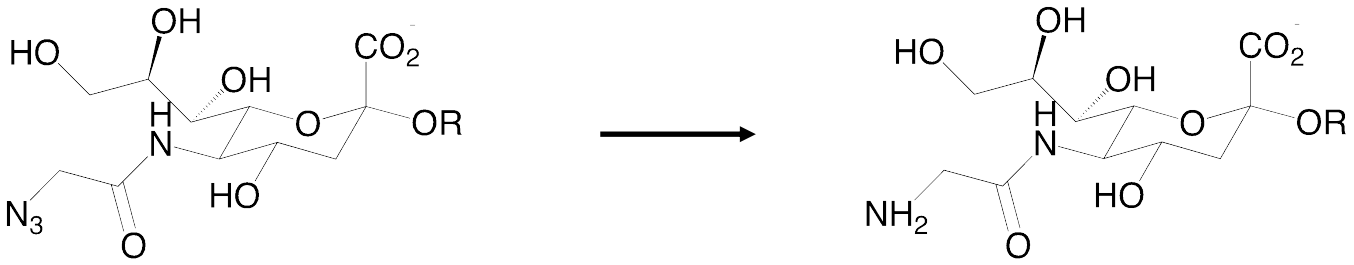


Figure S1. Identification and quantification of SiaNAz-containing glycans. Left panels: MS spectra showing isotopic distribution of representative SiaNAzylated N-glycans (hybrid, biantennary complex, and bisecting type) and O-glycans (disialylated core 2). Right panels: Chromatograms generated from the sum of ion counts of each specified glycan.



Azido sialic acid (SiaNAz)

Amino sialic acid (SiaNAm)

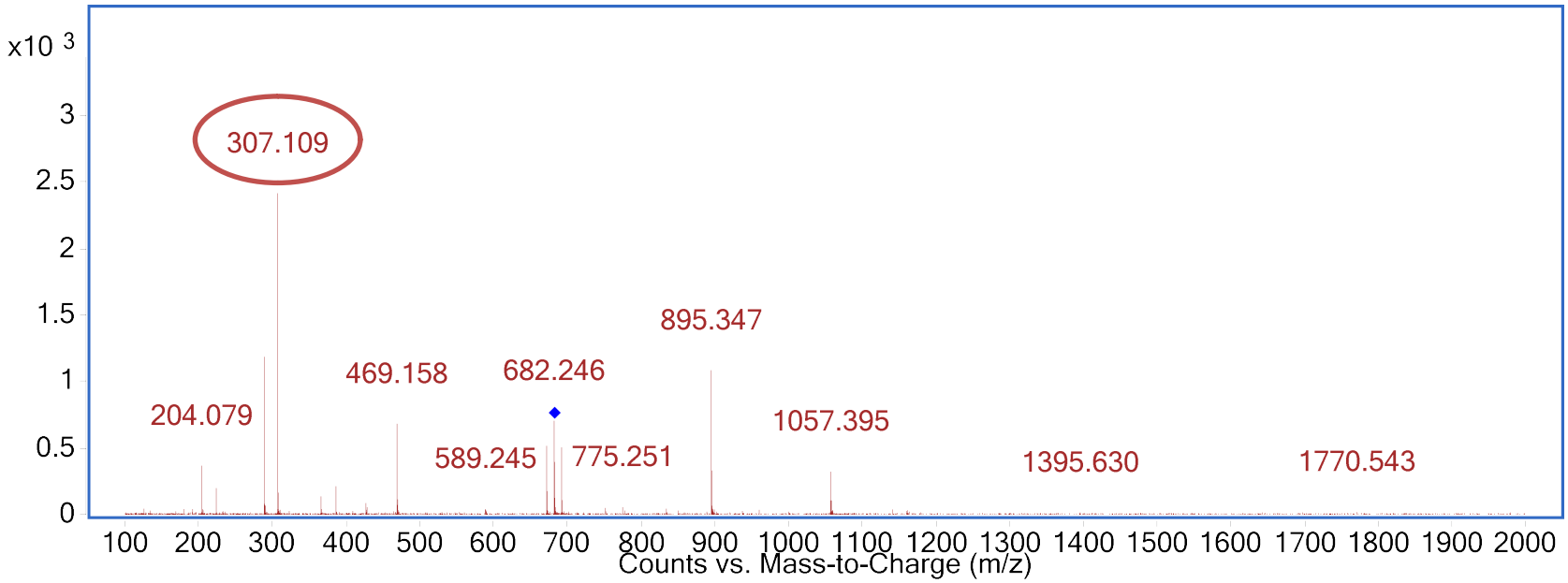


Figure S2. CID-MS/MS fragmentation spectrum of an O-glycan bearing a reduced SiaNAz residue. The red circle indicates the diagnostic ion representing an unnatural sialic acid that has undergone an azide to amine conversion. The blue diamond indicates the precursor ion (*m/z* 682.260) corresponding to a glycan with composition $Hex_2HexNAc_2SiaNAm_2$.

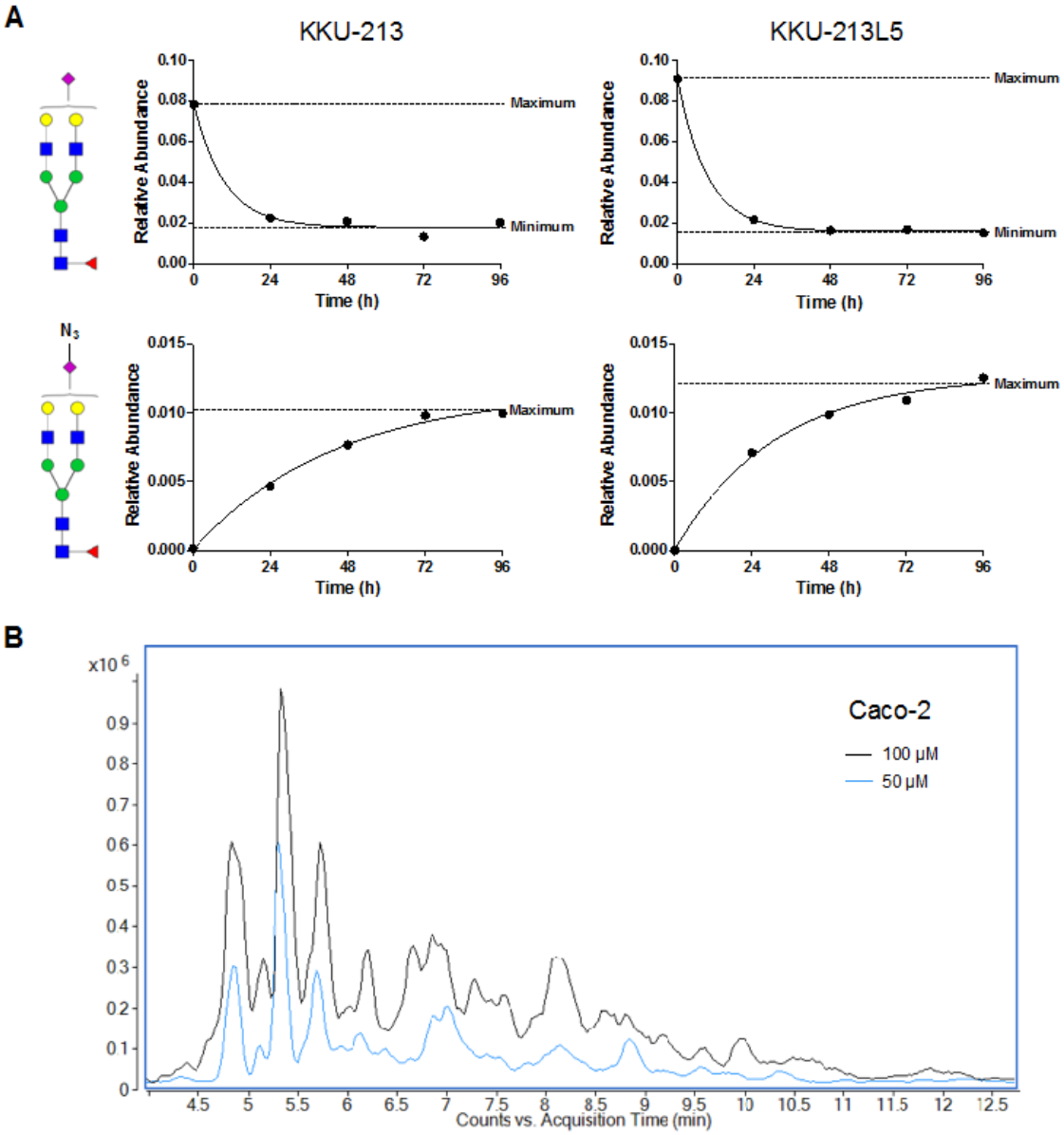


Figure S3. Optimization of ManNAz addition to cell lines. (A) Depletion of individual Sia-glycans and concurrent increase in the corresponding SiaNAz-glycans were monitored over time after ManNAz treatment. Relative abundances were calculated by normalizing ion counts to the total glycan abundances. (B) Increasing concentrations of ManNAz were added to cells until a maximal response was observed. An extracted compound chromatogram is shown per condition, in which the sum of all identified glycan compounds are outlined at their respective retention times.

	Cell Type	Tissue	Disease	Native Cell Surface Sialylation	Maximal [ManNAz] (μM)
Caco-2	Epithelial	Colon	Colorectal adenocarcinoma	78%	100
PNT2		Prostate	Normal	45%	50
KKU-213		Liver	Cholangiocarcinoma	22%	50
KKU-213L5		Liver	Metastatic cholangiocarcinoma	20%	50

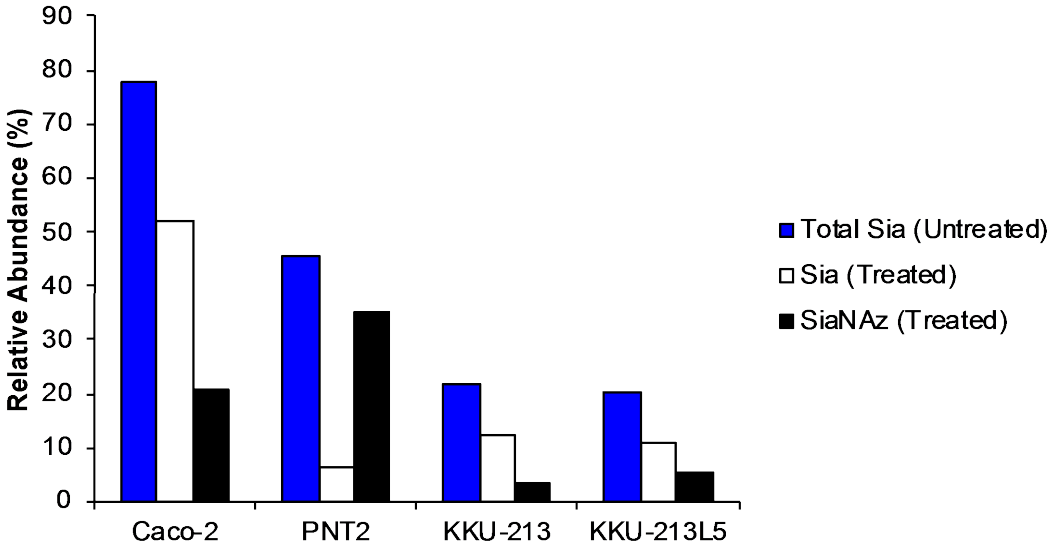


Figure S4. Measures of variability in SiaNAz incorporation onto cell surface N-glycans. The characteristics of enrolled cell lines are described in the table. Distribution of sialylated glycans are shown as unique to each cell. Relative abundances indicate the summed peak areas of specified glycan groups in untreated and ManNAz-treated cells. Conversion of ManNAz to SiaNAz and subsequent incorporation onto cell surface N-glycans are evident by the appearance of SiaNAzylated glycans in treated samples. Total sialylation levels between untreated and treated cells were comparable. Native sialylation levels did not dictate the extent of SiaNAz incorporation post-treatment. Fuc, fucosylated; Sia, sialylated (NeuAc); HM, high mannose.

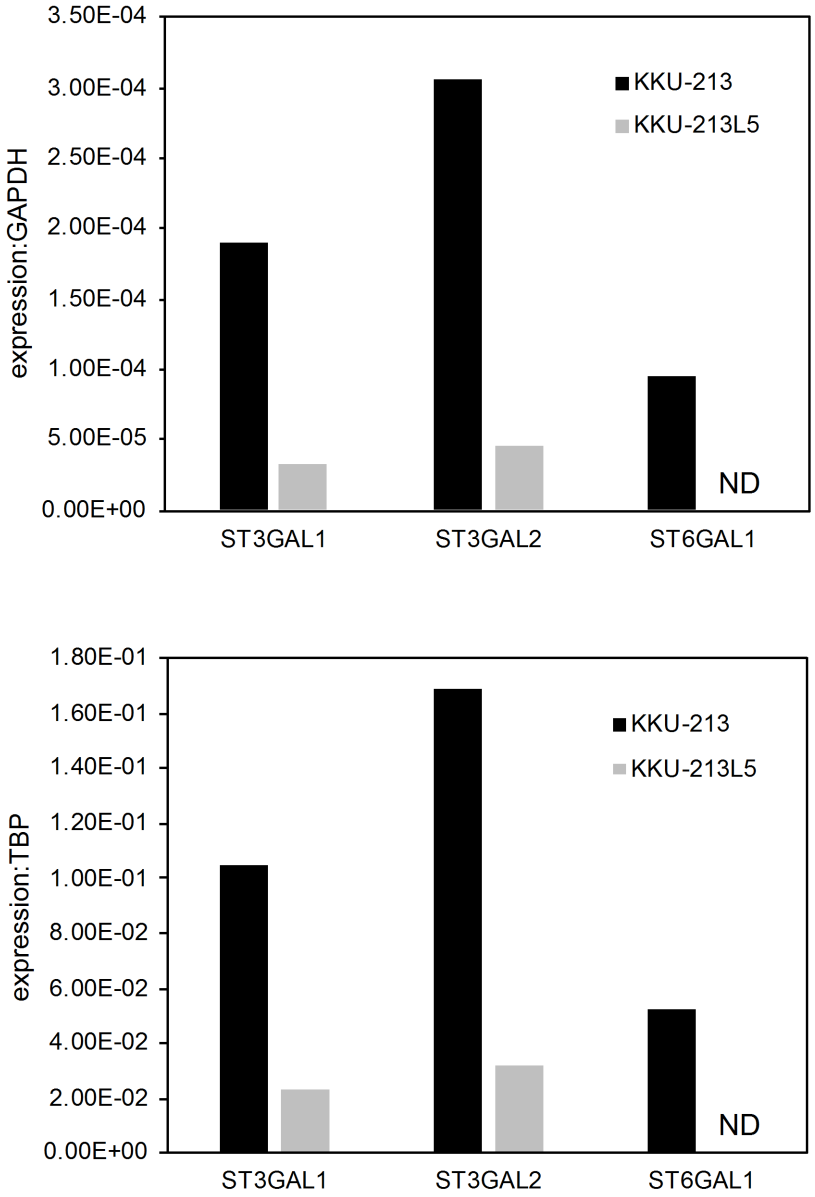


Figure S5. Q-RT-PCR of sialyltransferase genes in KKU-213 and KKU-213L5. Expression is normalized to GAPDH (top) and TBP (bottom). ND, not detectable.

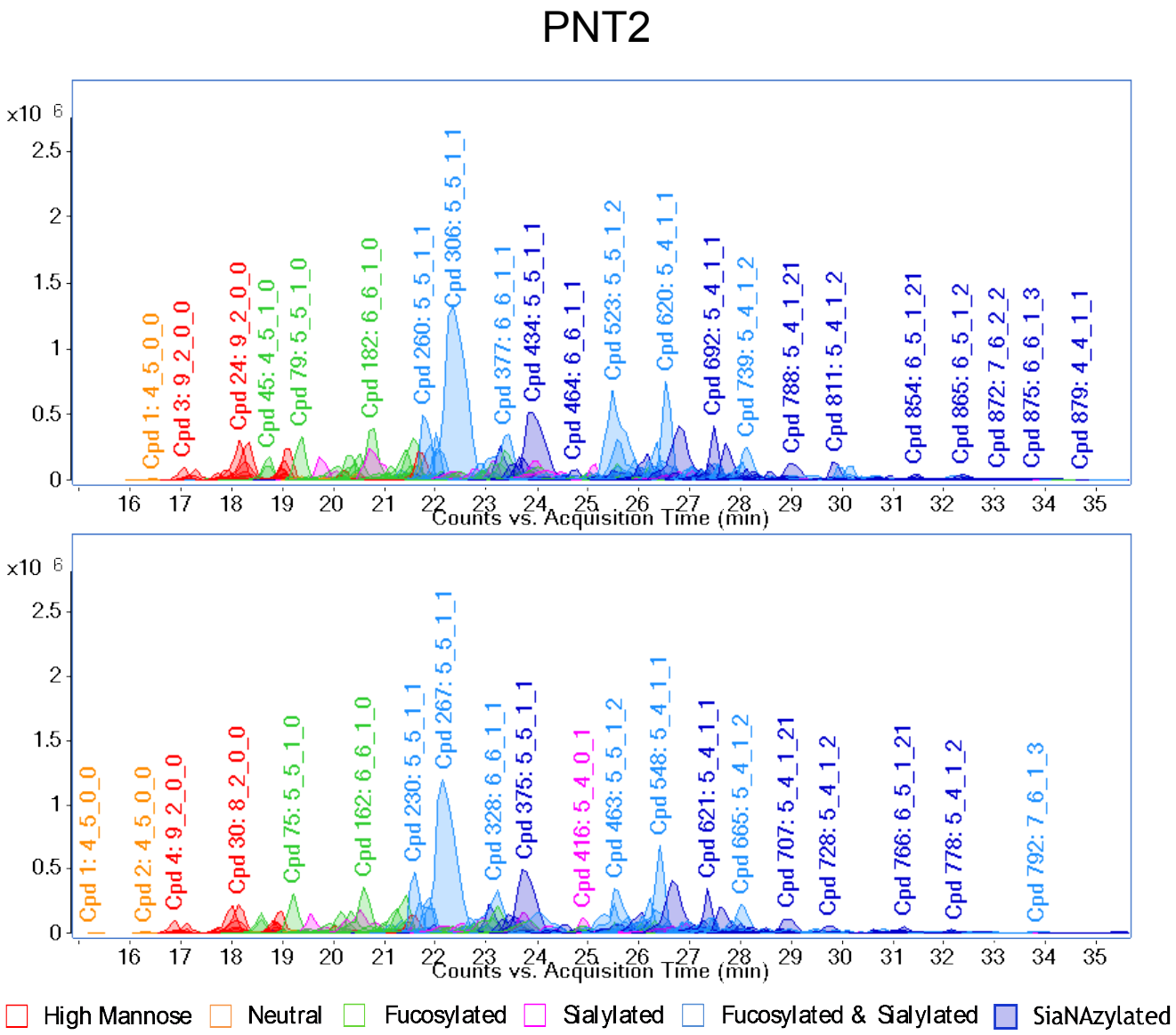


Figure S6. N-Glycans identified on PNT2 via PNGase F release from two independent experiments. Each peak in the chromatogram represents a unique glycan compound (Cpd). Peaks are colored according to glycan type. Notation is as follows: Hex_HexNAc_Fuc_SiaSiaNAz.

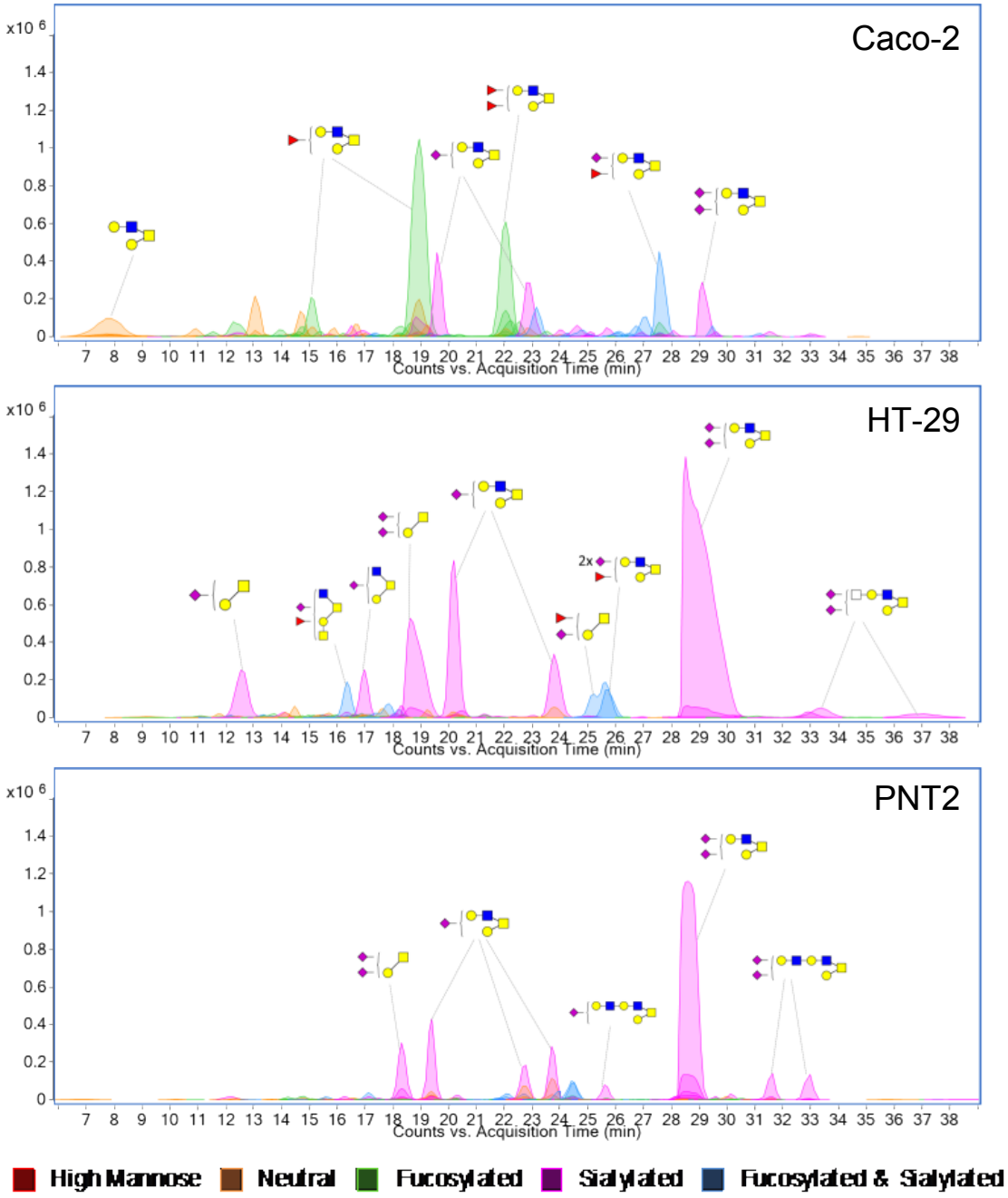


Figure S7. Comprehensive O-glycan profile of Caco-2, HT-29, and PNT2 membrane-associated proteins.

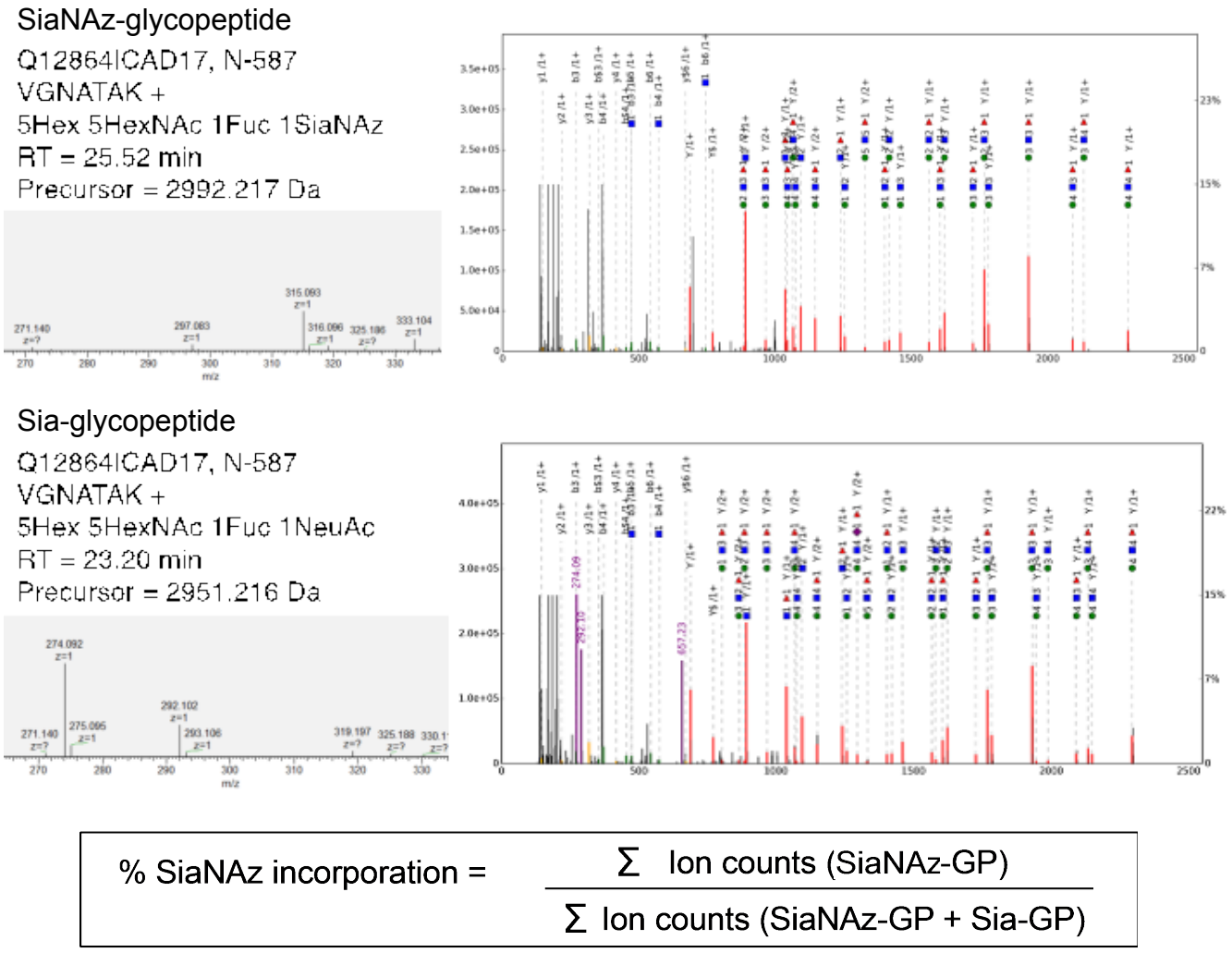


Figure S8. Differentiation and quantification of sialylated (Sia/SiaNAz) glycopeptides. Fragmentation pattern of a Cadherin -17 glycopeptide (VGNATAK, position 585 to 591) on Caco-2 bearing a SiaNAzylated N-glycan (top panel) versus the same glycopeptide bearing the Sia-N-glycan (bottom panel). Spectra were annotated using pGlyco. Glycopeptide signals are outlined in red. Peptide b- and y-ions are outlined in green and orange, respectively. Glycan signals are outlined in purple. SiaNAz (*m/z* 315.093 and 333.104) and Sia (*m/z* 274.092 and 292.102) were distinguished by exact mass (left panels). The sum of the ion counts of the SiaNAzylated glycopeptide was compared to that of the sialylated glycopeptide for quantification.

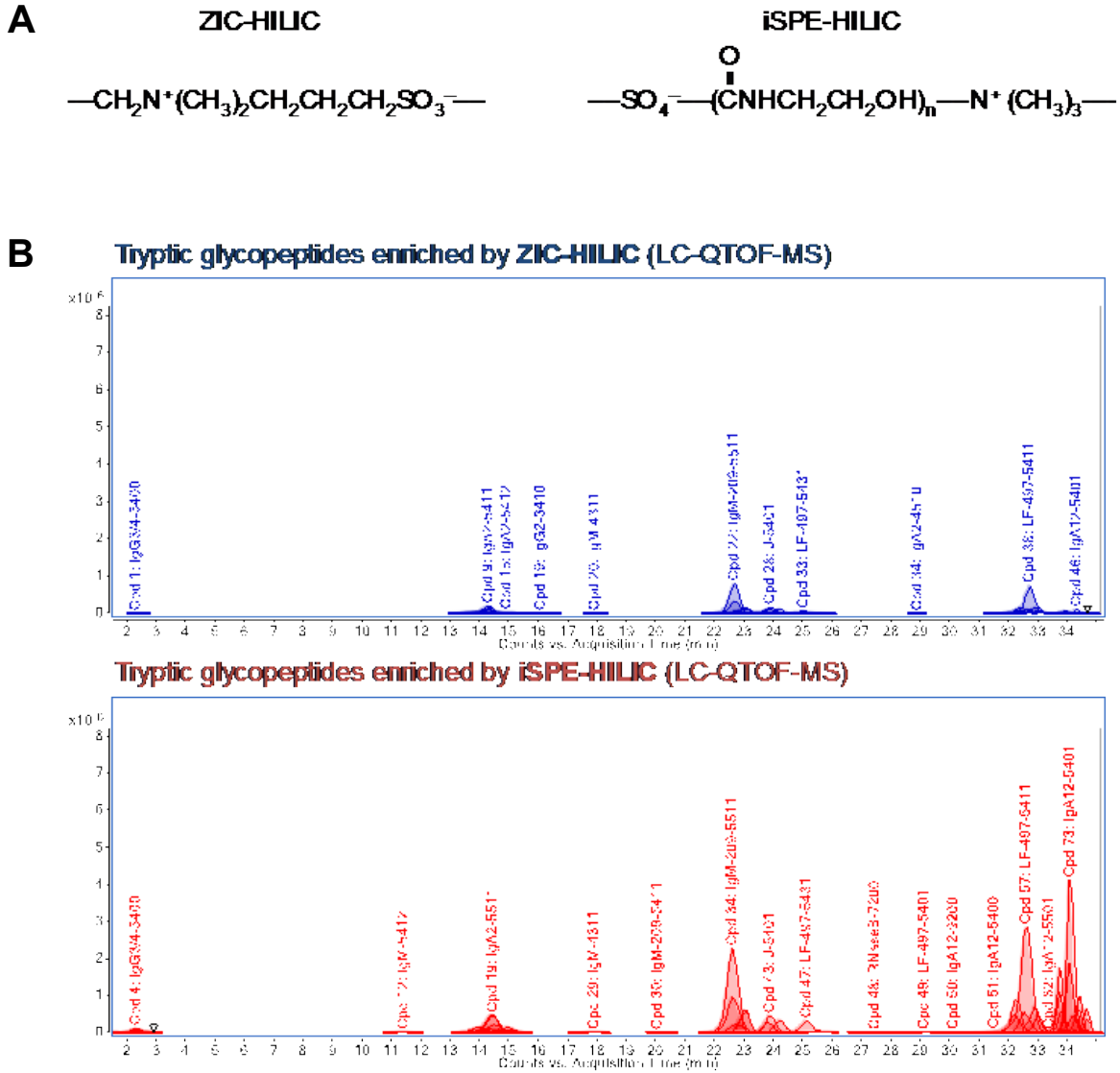
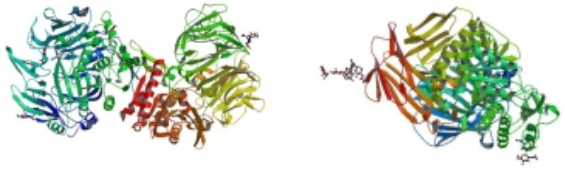


Figure S9. Glycopeptide enrichment strategy. (A) Comparison of conventional ZIC-HILIC and newly modified iSPE-HILIC. (B) Abundances and identities of glycopeptides prepared by trypsin digestion of a pool of known glycoprotein standards: Immunoglobulins (IgG, IgA, IgM, and J chain), lactoferrin (LF), and RNase B. Glycopeptides were enriched either with ZIC-HILIC (top panel) or iSPE-HILIC (bottom panel) and quantified using LC-QTOF-MS. Each peak represents a unique glycopeptide. Notation is as follows: Compound (Cpd) number: Protein Name-Site of Glycosylation-Glycan Composition (Hex HexNAc Fuc NeuAc).



Caco-2 membrane proteins

denature
reduce disulfide bonds
trypsin digestion

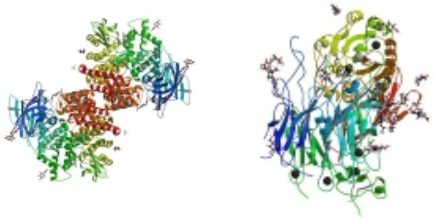
iSPE-HILIC
LC-MS/MS

Non-glycosylated

PAEAPMPAEPAPPGPASPGGAPEPPAAAR
DSQVPGPIGCNCDPQGSVSSQCDAAGQCQCK
PAEAPMPAEPAPPGPASPGGAPEPPAAAR
…

Glycosylated

IP**N**NTQWVTWSPVGHK
VFPYISVMVN**N**GSLSYDHSK
CVYEALC**N**TTSECPPPVITR
VLSLPL**N**SSAVVNCSVHGLPTPALR
IVY**N**ITLPLHPNQGIIEHR
G**N**LTYGYVTILNGSDIR
HIWSLEISNKP**N**VSEPEEPK
…



PNT2 membrane proteins

denature
reduce disulfide bonds
trypsin digestion

iSPE-HILIC
LC-MS/MS

Non-glycosylated

VCECNPNYTGSACDCSLDTSTCEASNGQICNGR
AYTNSVDVLDGEGWESQDAPSCSTHPCDSSLPCDSDL
…

Glycosylated

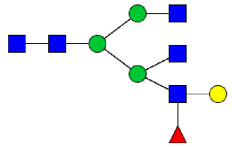
ILT**N**NSQTPILSPQEVVSCSQYAQGCEGGFPYLIAGK
AFS**N**SSYVLNPTTGELVFDPLSASDTGEYSCEAR
LSHDA**N**ETLPLHLYVK
HFK**N**CTSISGDLHILPVAFR
HRPTAGAF**N**HSDLDAELR
SLDAYPILNQAQALE**N**HTEVQFQK
…

	# Unique Peptides		
	Total	Glycopeptide	Enrichment
Caco-2	317587	307391	96.8%
PNT2	24786	26415	93.8%

Figure S10. Percent enrichment of glycopeptides over non-glycosylated peptides in Caco-2 and PNT2. Enriched membrane proteins were denatured and digested with trypsin in solution. Resulting peptides were enriched with iSPE-HILIC and sequenced using C18-MS/MS. The goodness of enrichment was calculated based on the number of unique peptides identified.

Protein: Cadherin-17

Glycan structure:



N-Glycosylation site 250

Peptide sequence:

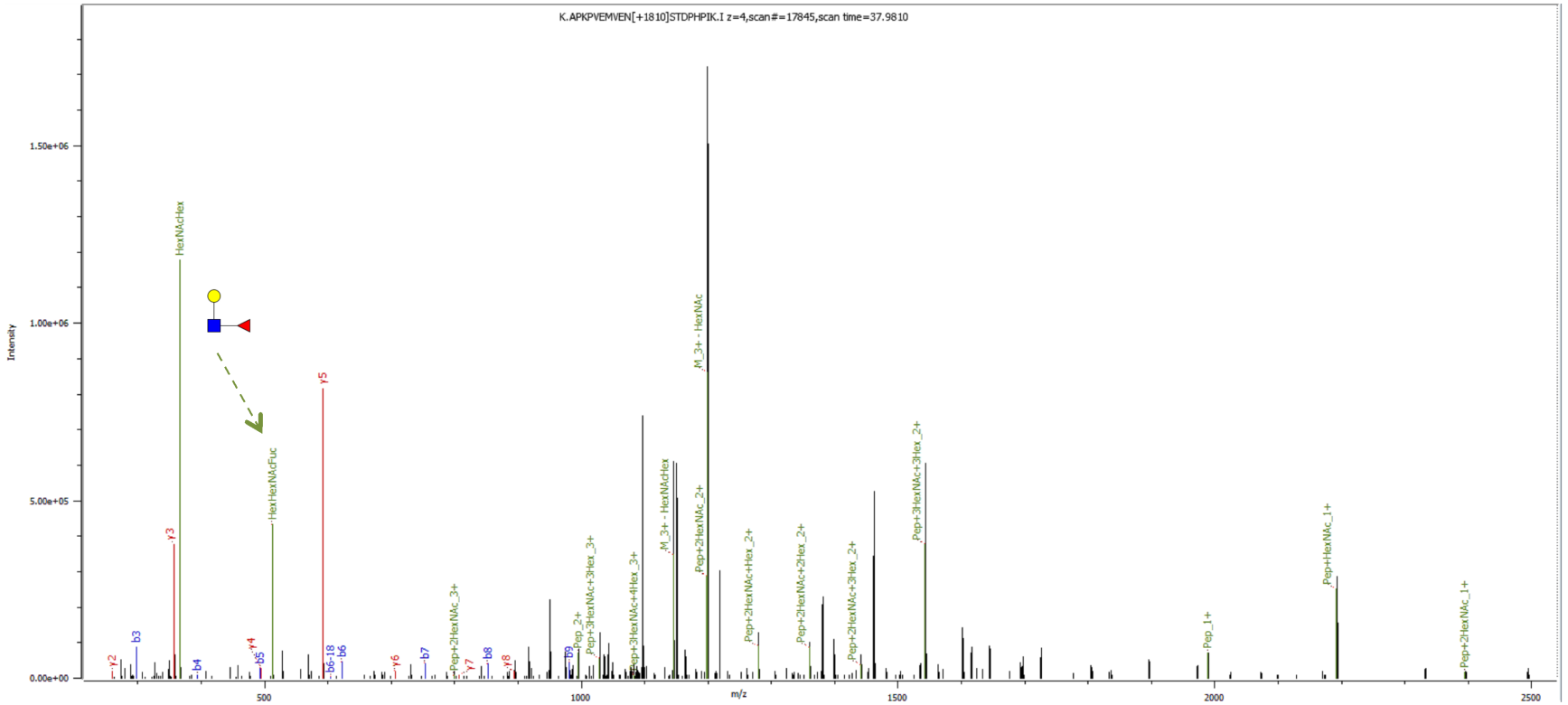


Figure S11. Informative glycopeptide fragmentation with stepped collision energy. Glycan ions (green) yield information about branching and the presence or absence of epitopes. Glycopeptide ions (green) facilitate localization of the glycosylation site. Peptide b-ions (blue) and y-ions (red) were used to sequence the peptide and search against the human proteome. Peptide sequence shows the site of glycosylation in bold and the b-/y-ions generated from the fragmentation event. Spectral searches were performed on Byonic (Protein Metrics, CA).

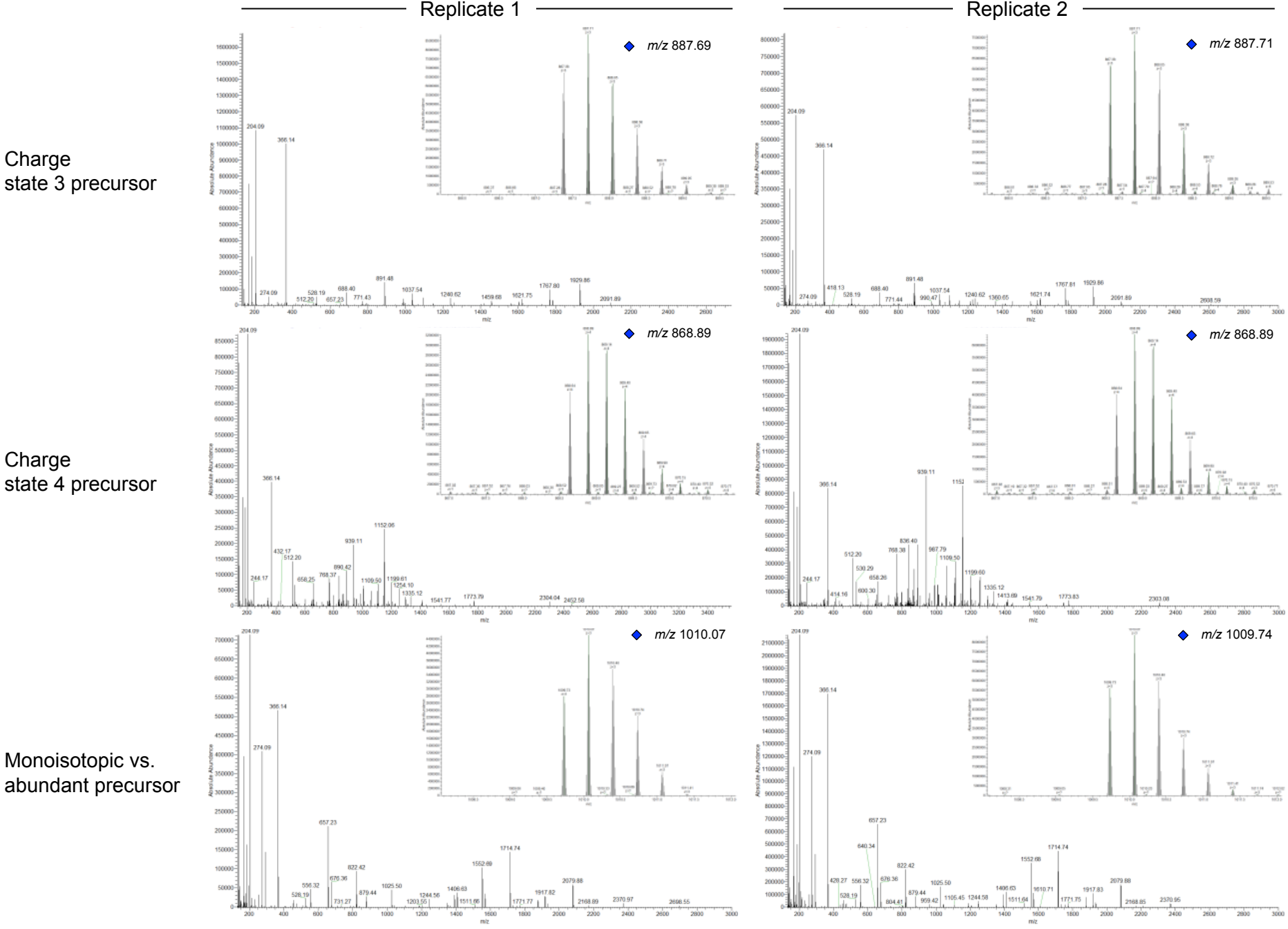


Figure S12. MS/MS spectra of representative glycopeptides from independent biological replicates. MS spectra are shown inset. Precursor ions are indicated by a blue diamond.

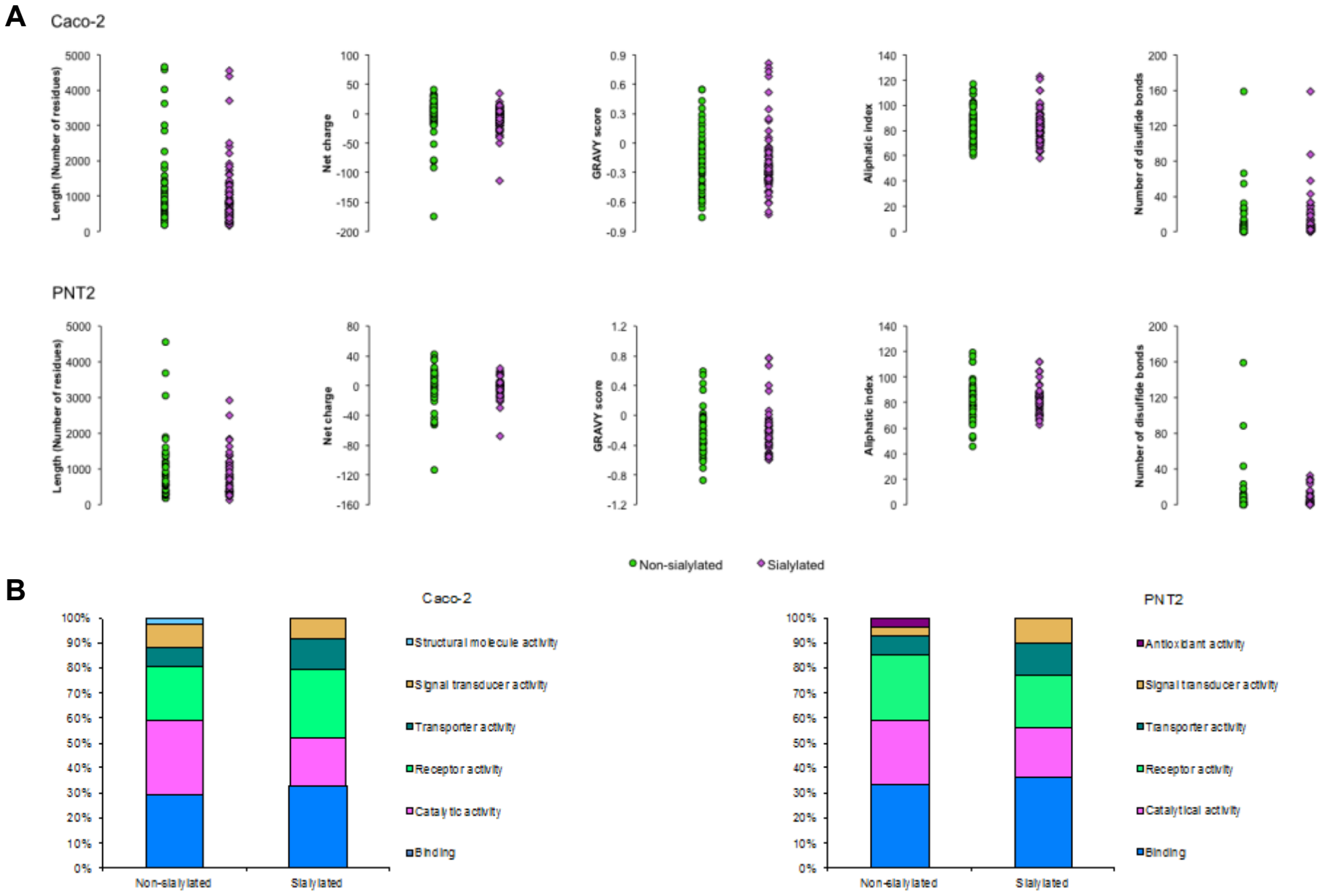


Figure S13. Classification of cell surface glycoproteins differentiated by the occurrence of sialylation. (A) Comparison of full-length sequence features between non-sialylated and sialylated cell surface glycoproteins identified on Caco-2 and PNT2. (B) Molecular functions of non-sialylated and sialylated Caco-2 and PNT2 cell surface glycoproteins.

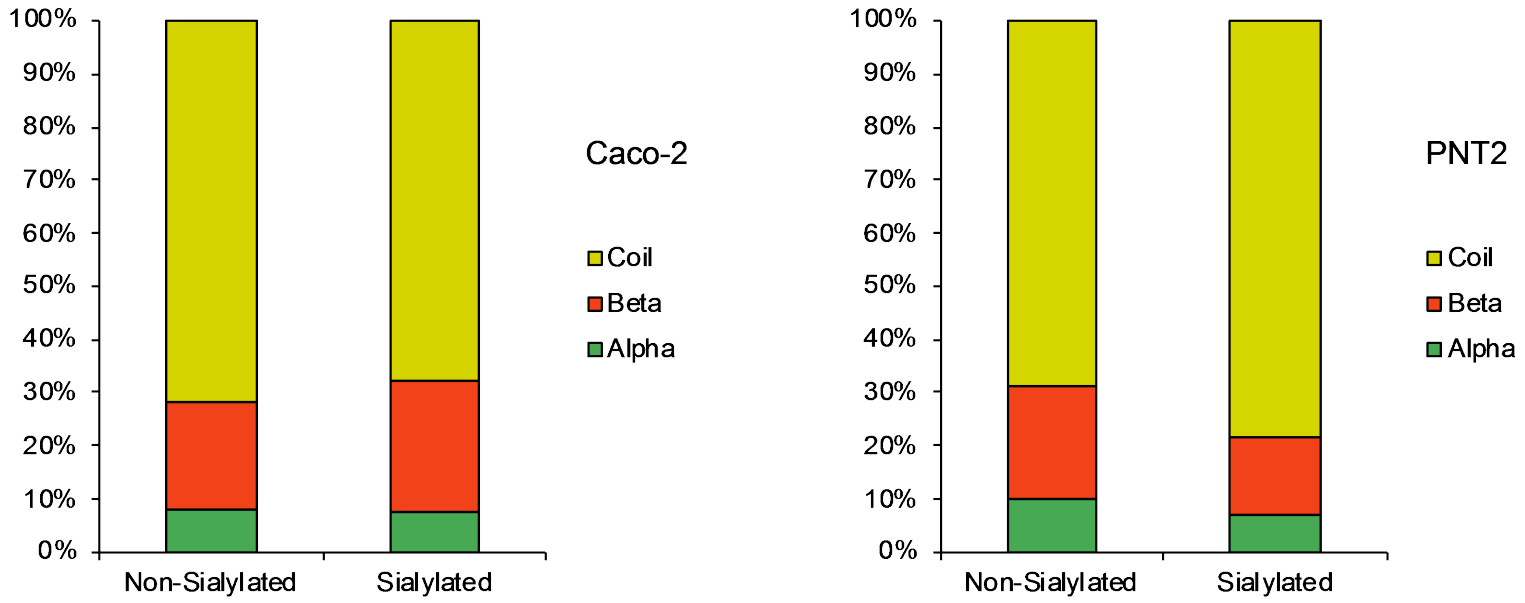


Figure S14. Secondary structure prediction of glycosylation sites that were identified as either non-sialylated or sialylated on the surface of Caco-2 and PNT2 cells.

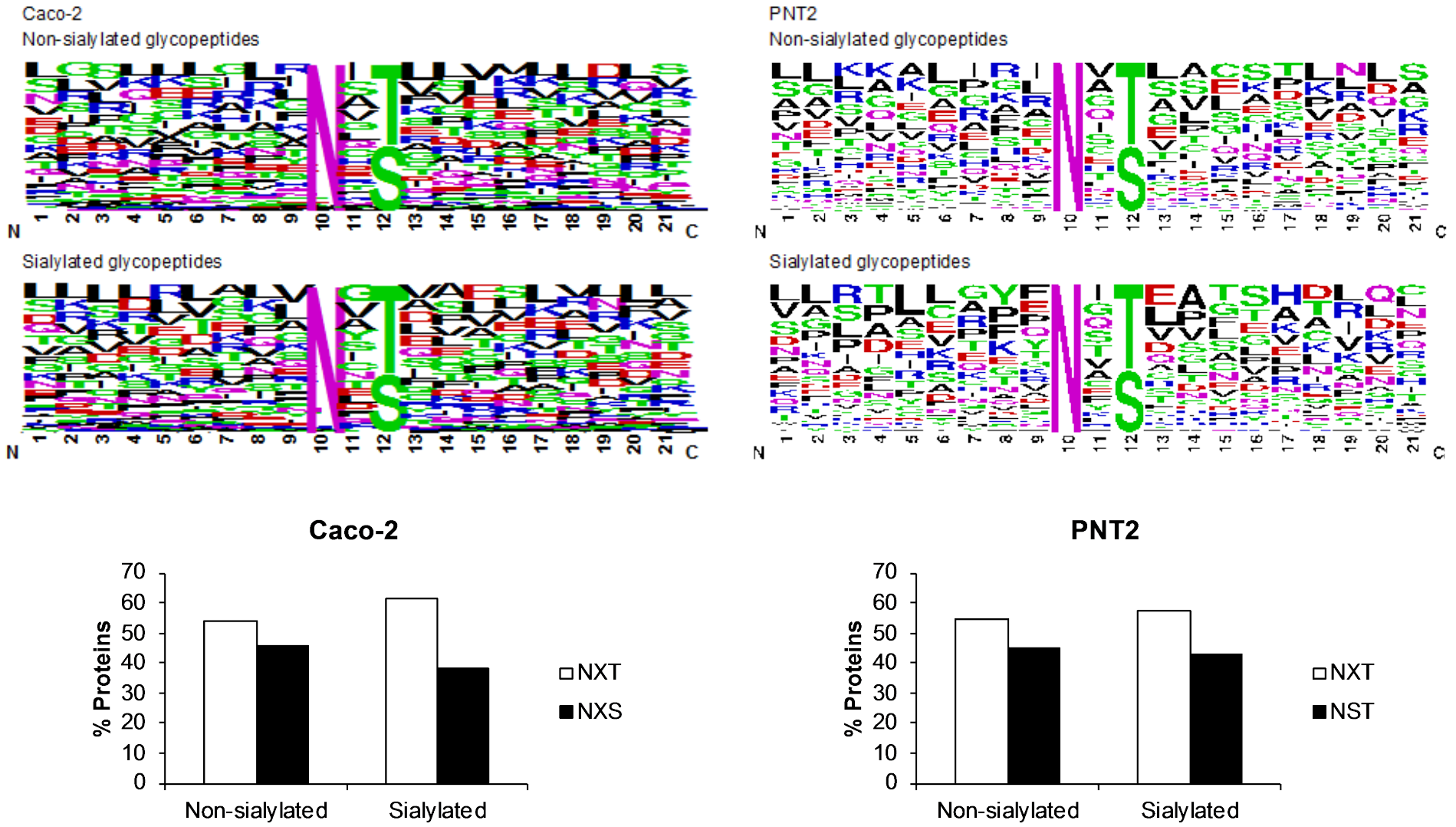


Figure S15. Frequency plots of amino acid residues ±9 positions from the conserved N-glycosylation consensus sequence, N-X-S/T. Residues are indicated as one letter symbols while frequency is proportional to the size of the letter. The plot is colored according to the charge of the amino acid: KRH, green; DE, blue; AVLIPWFM, red. Caco-2 (left) and PNT2 (right) cell surface glycopeptides are distinguished by the occurrence of sialylation on glycans attached to the Asn residue of the consensus sequence. Logos were generated using WebLogo version 2.8.2. Bar graphs show percent of non-sialylated and sialylated peptides that were glycosylated at NXT or NXS sequences, respectively.

A

P00533 | Epidermal growth factor receptor

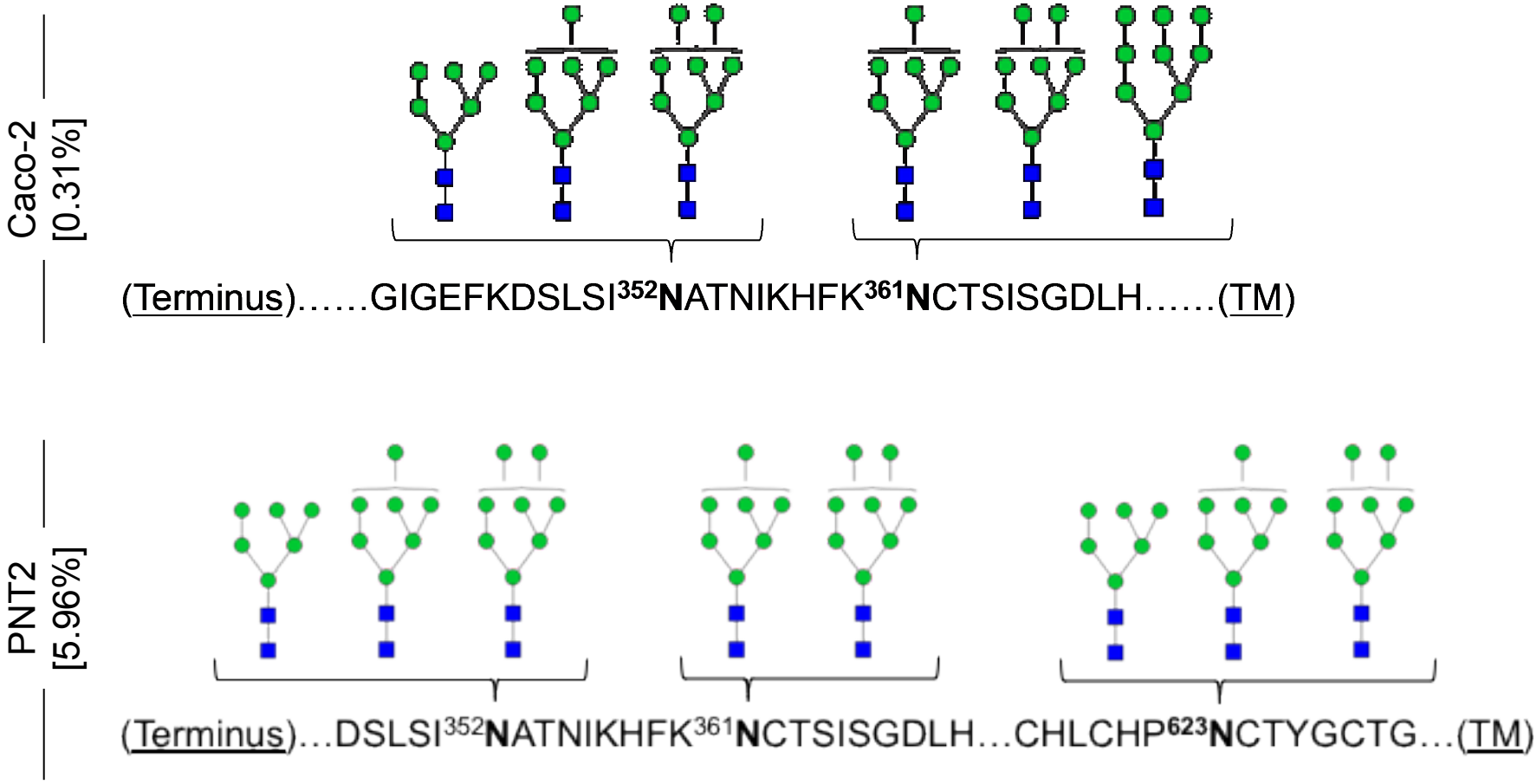
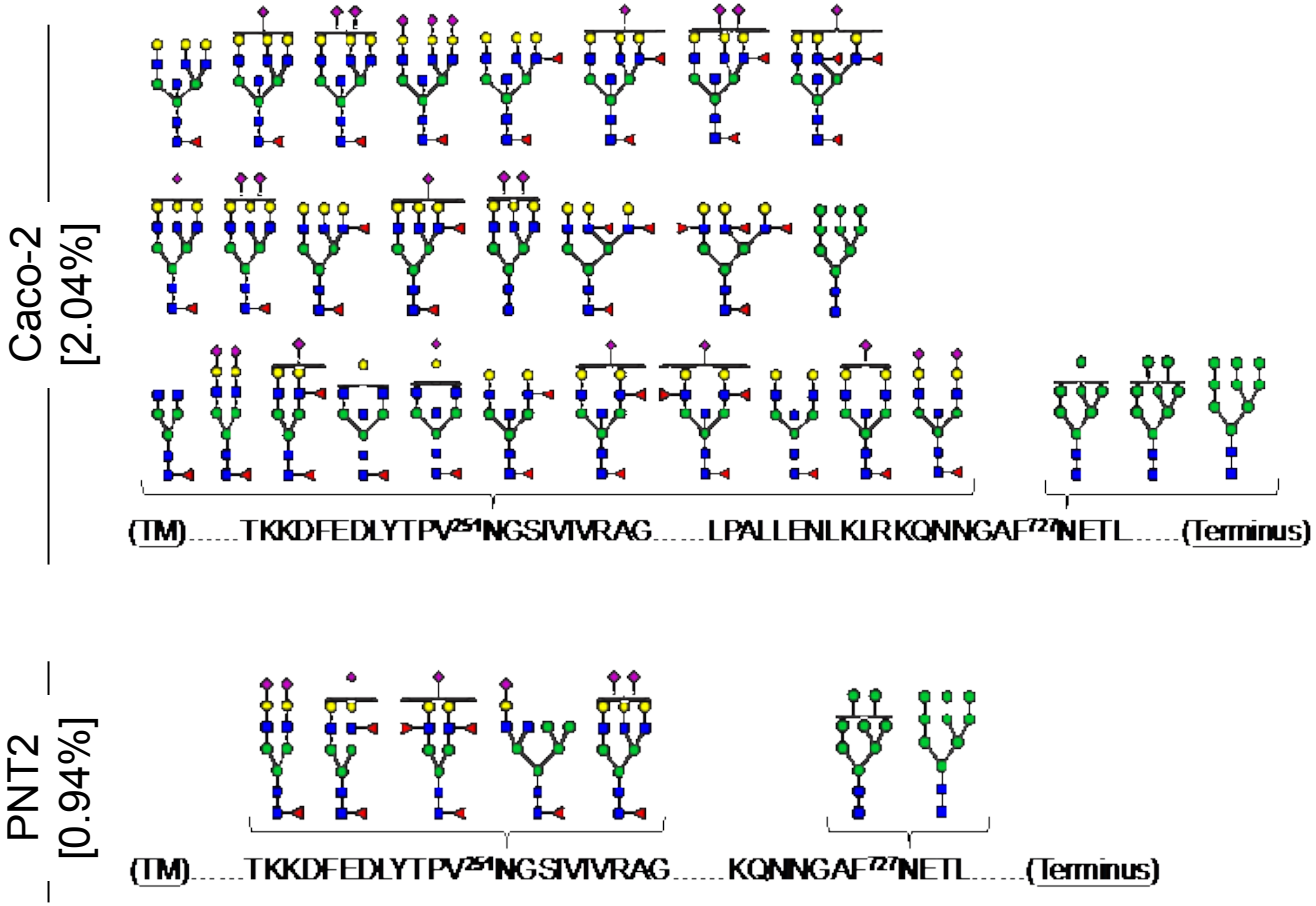


Figure S16. (A-K) Graphical representations of glycoforms associated with each occupied site on representative membrane proteins of Caco-2 and PNT2. Proteins are presented according to molecular function (i.e. receptor, hydrolase/peptidase, adhesion/signaling). Glycan structures are putative. Brackets indicate relative abundance, which was determined by normalization to the total membrane glycoprotein abundance per cell line using Byonic. The position and site of glycosylation are bolded within the sequence. The complete list of glycoforms is found in Tables S2-S5. TM, transmembrane.

Receptor

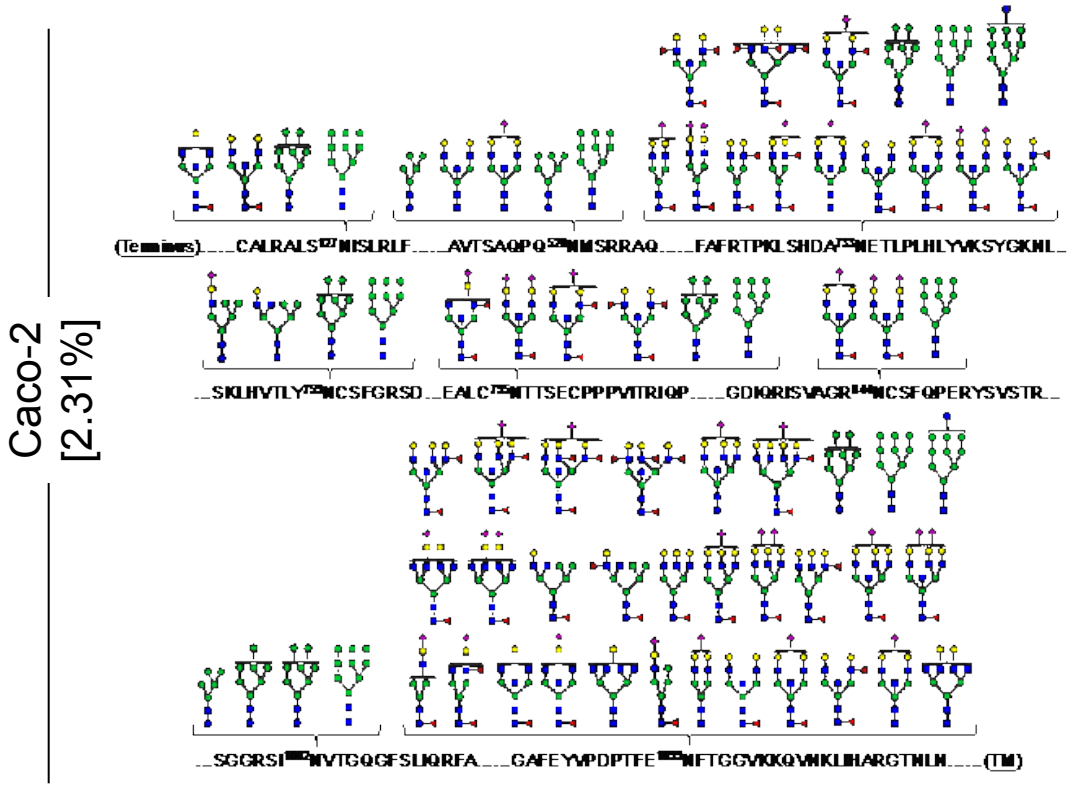
B

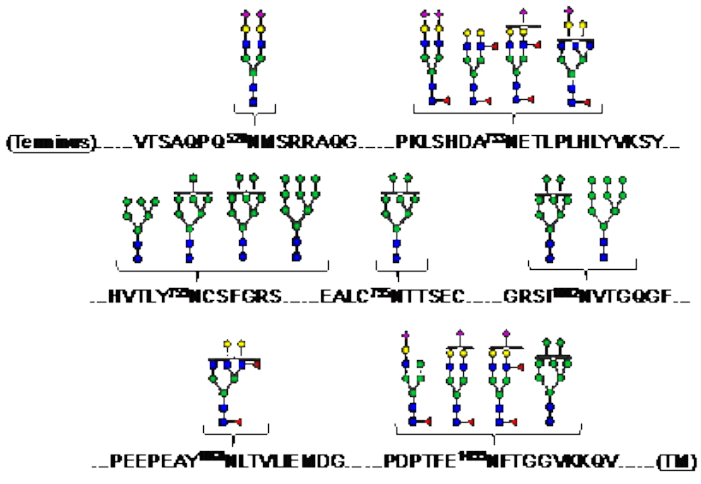
P02786 | Transferrin receptor protein 1



c

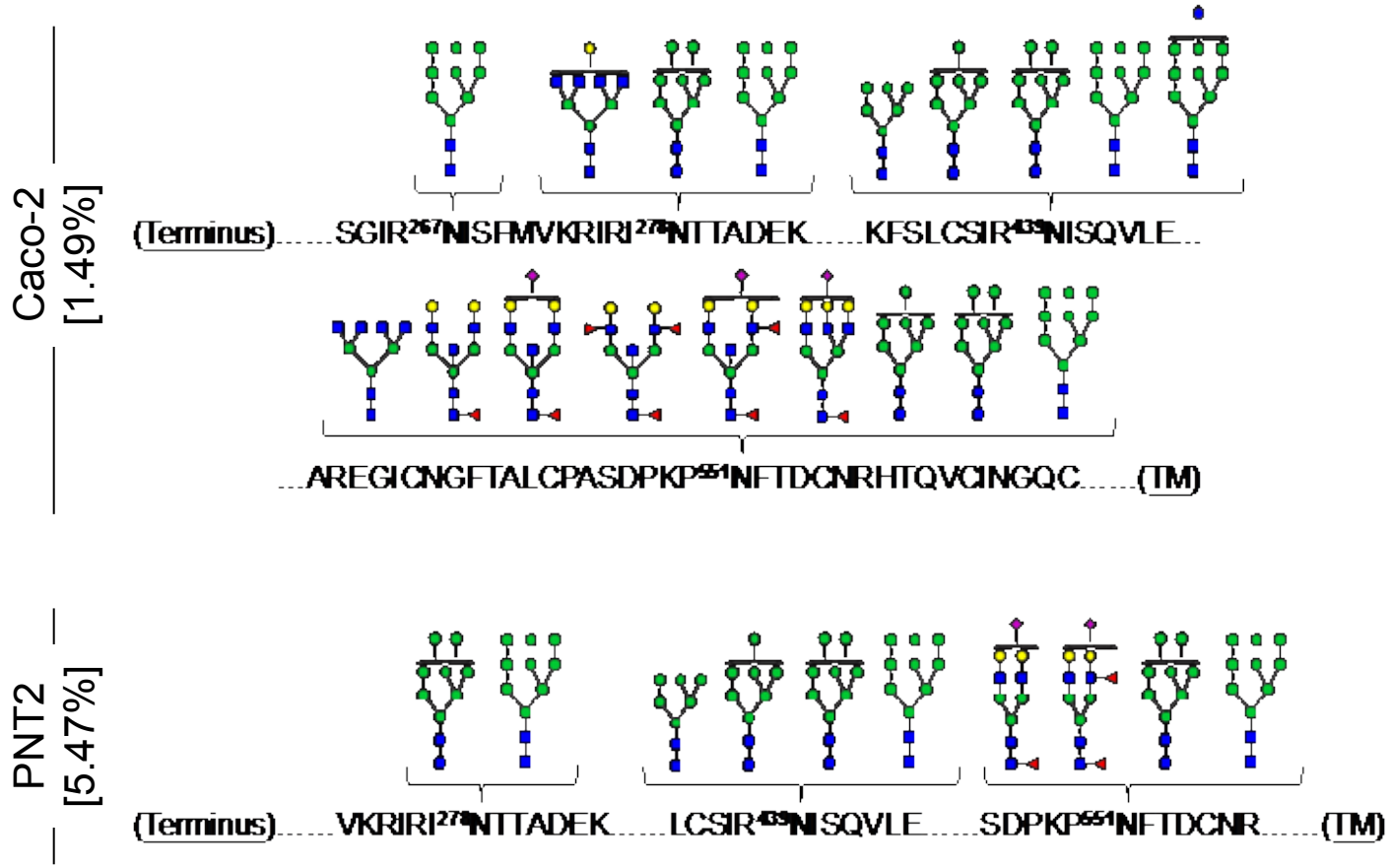
O15031 | Plexin-B2





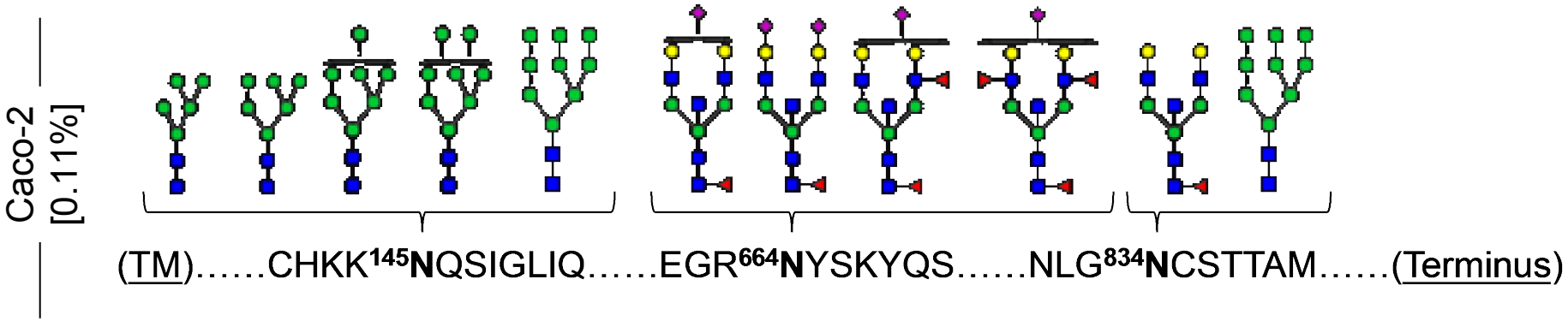
D

O14672 | Disintegrin and metalloproteinase domain-containing protein 10



E

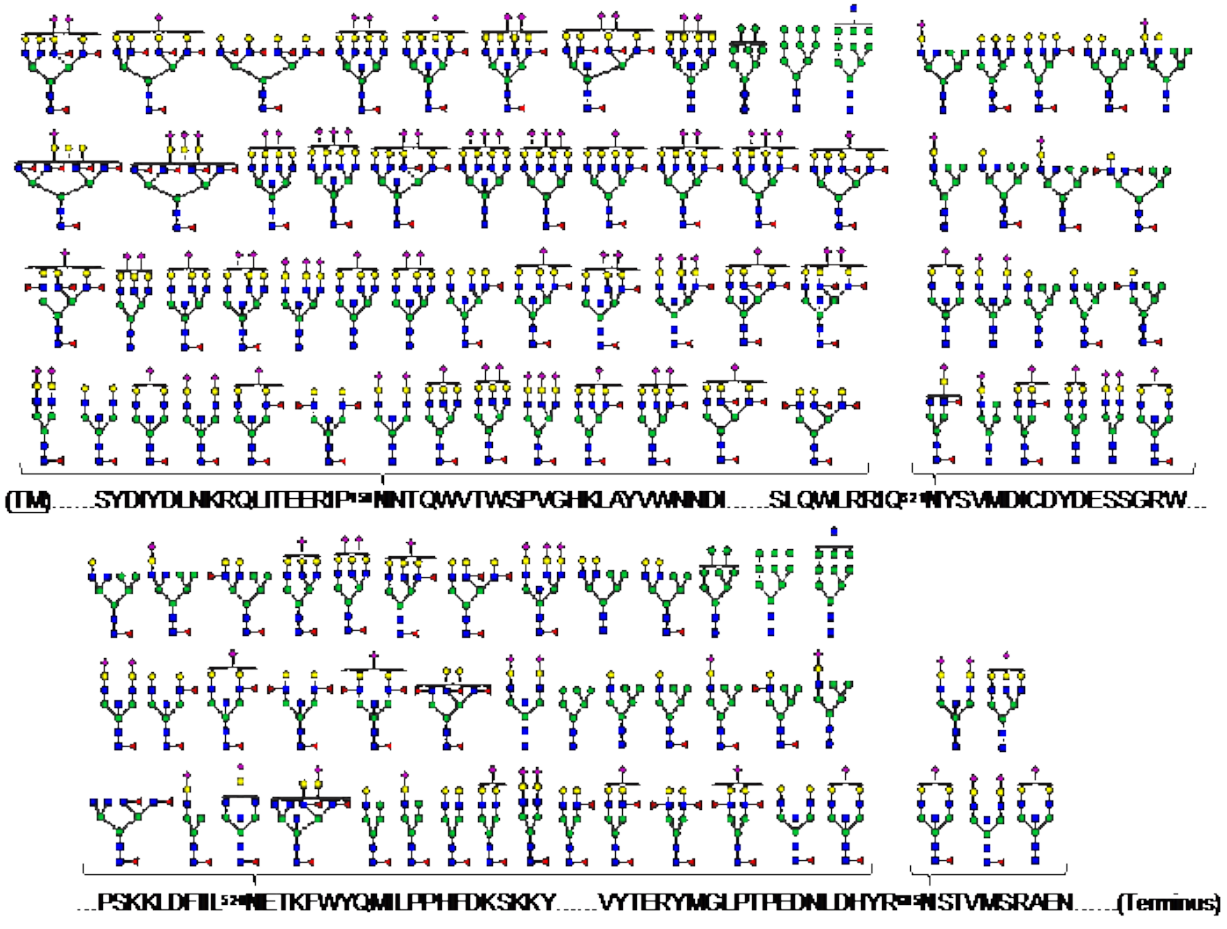
Q9UIQ6 | Leucyl-cystinyl aminopeptidase



F

P27487 | Dipeptidyl peptidase 4

Caco-2
[6.18%]

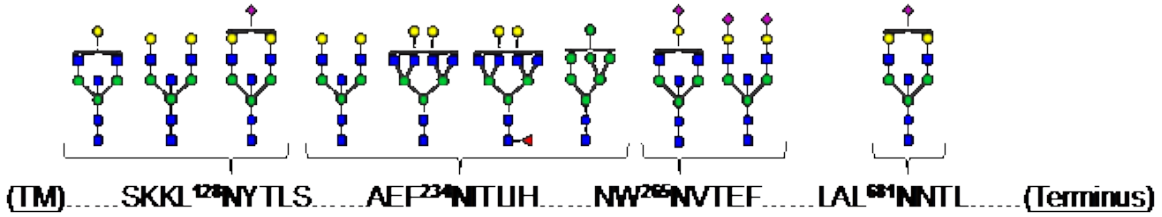


Hydrolase/Peptidase

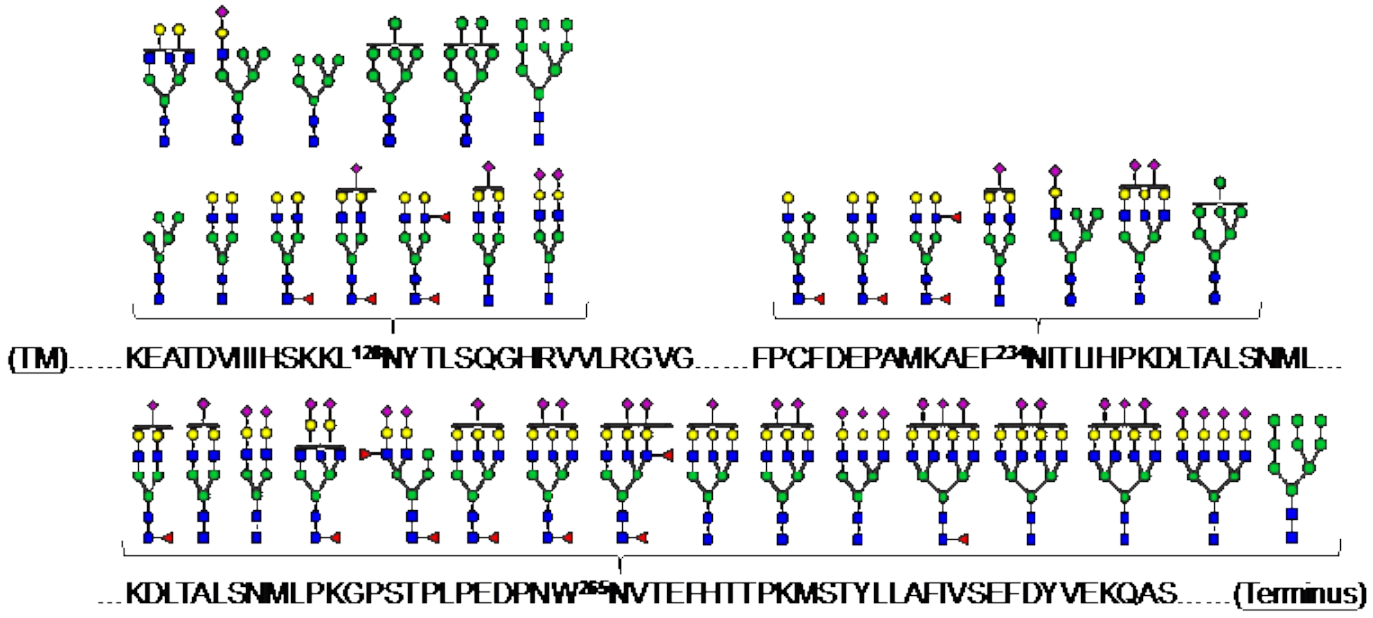
G

P15144 | Aminopeptidase N

Caco-2 [0.08%]



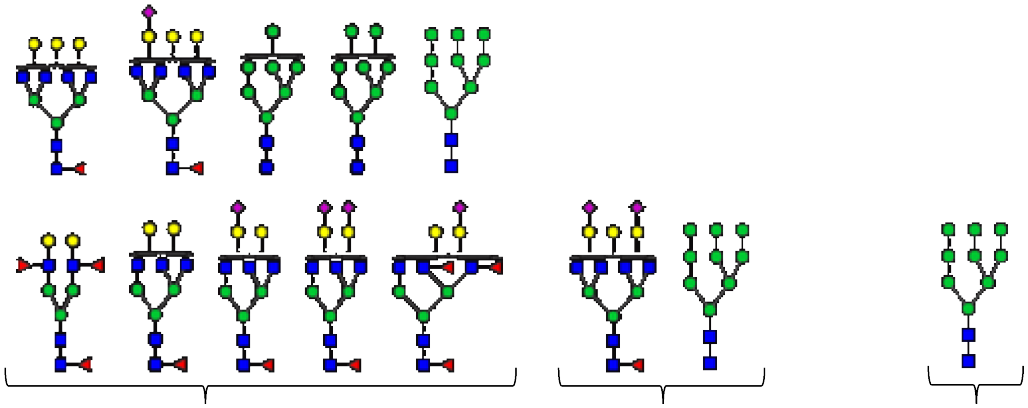
PNT2 [3.90%]



H

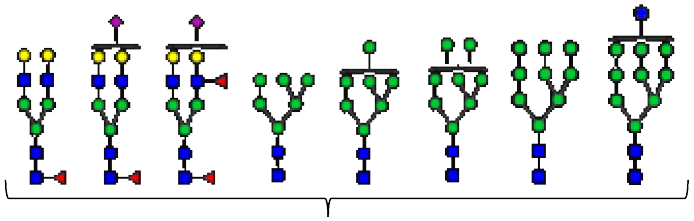
P42892 | Endothelin-converting enzyme 1

Caco-2 [0.14%]



(TM)……KHLLE^{166}**N**STASVSEAER……RLGGW^{210}**N**ITGPW……DYYL^{270}**N**KTENEK……(Terminus)

PNT2 [1.01%]



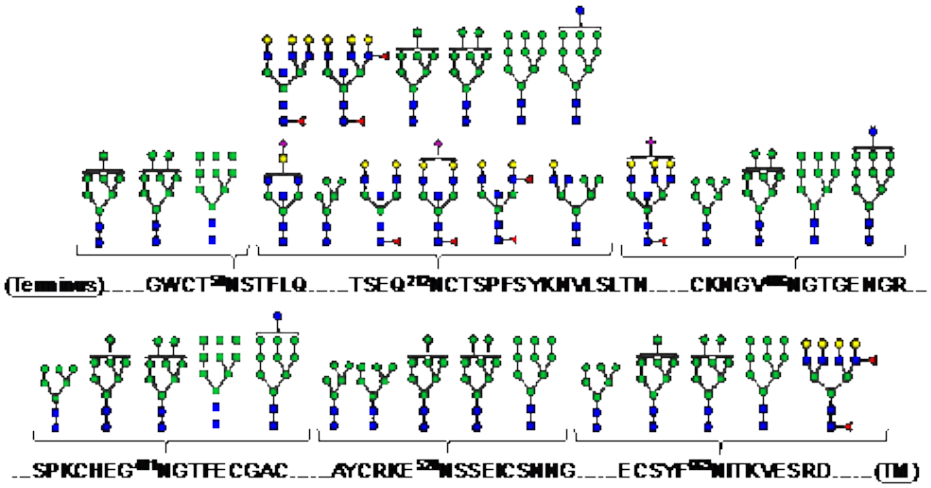
(TM)……NQAIIKHLLE^{166}**N**STASVSEAER……(Terminus)

Adhesion/Signaling

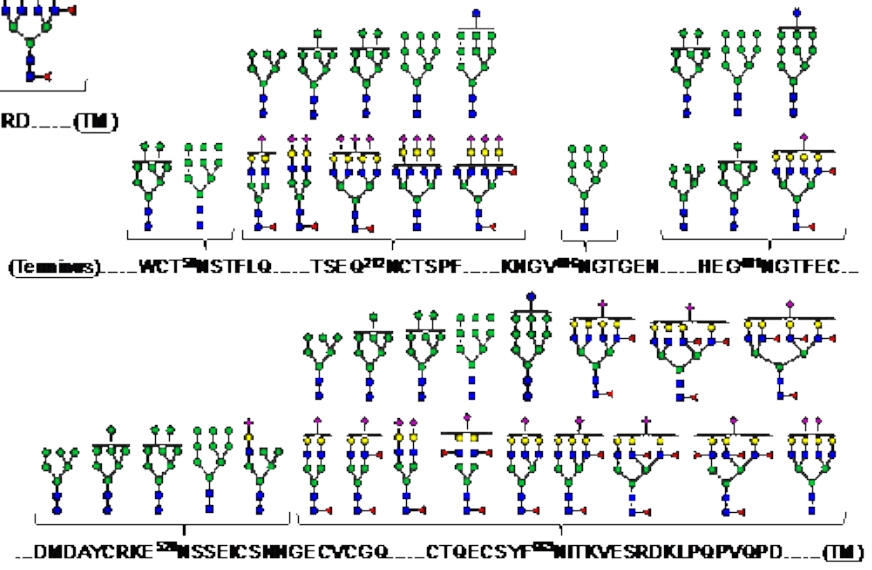
I

P05556 | Integrin beta-1

Caco-2 [2.11%]

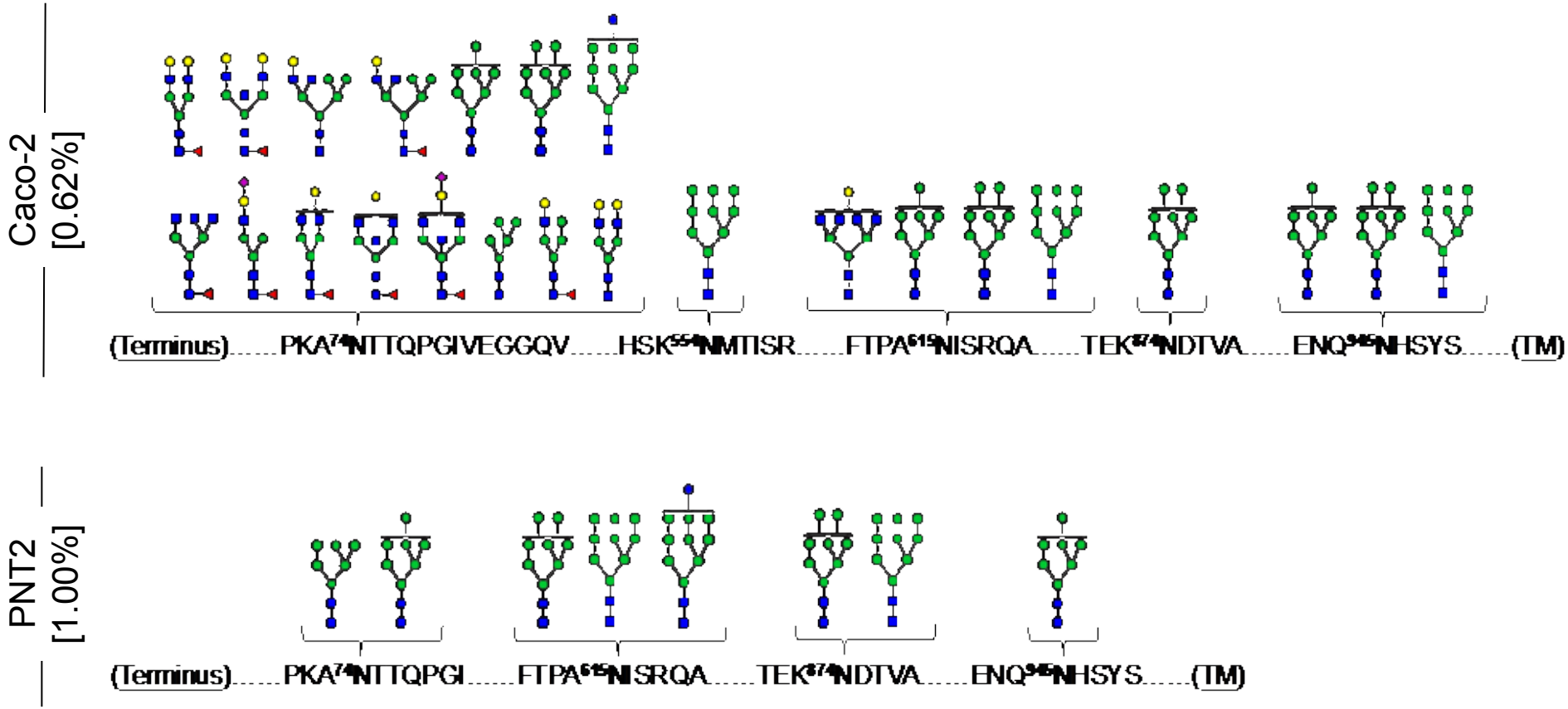


PNT2 [8.23%]



J

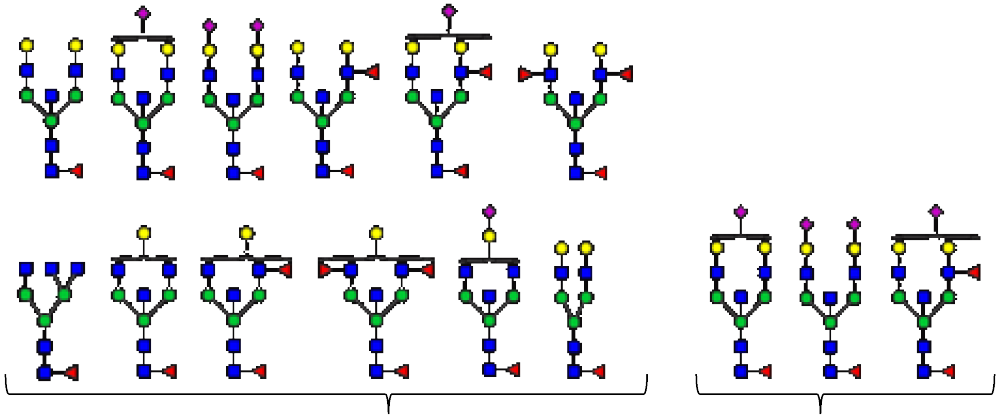
P06756 | Integrin alpha-V



K

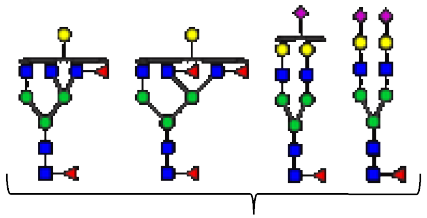
Q14126 | Desmoglein-2

Caco-2 [0.40%]



(Terminus)……TLVMKI^{182}**N**ATDADEPN……SRYVQ^{462}**N**GTYTVKIV……(TM)

PNT2 [1.45%]



(Terminus)……TLVMKI^{182}**N**ATDADEPN……(TM)

A

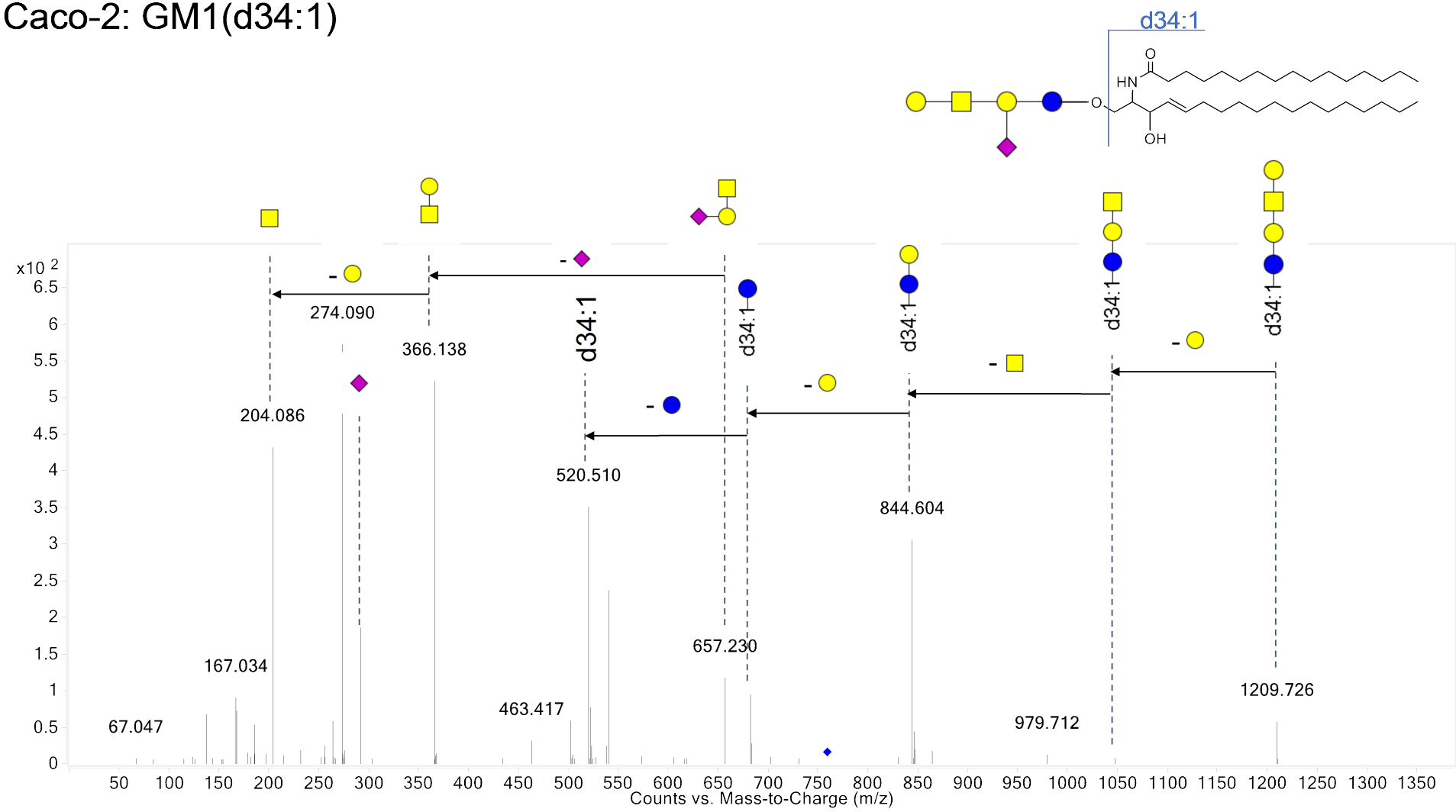
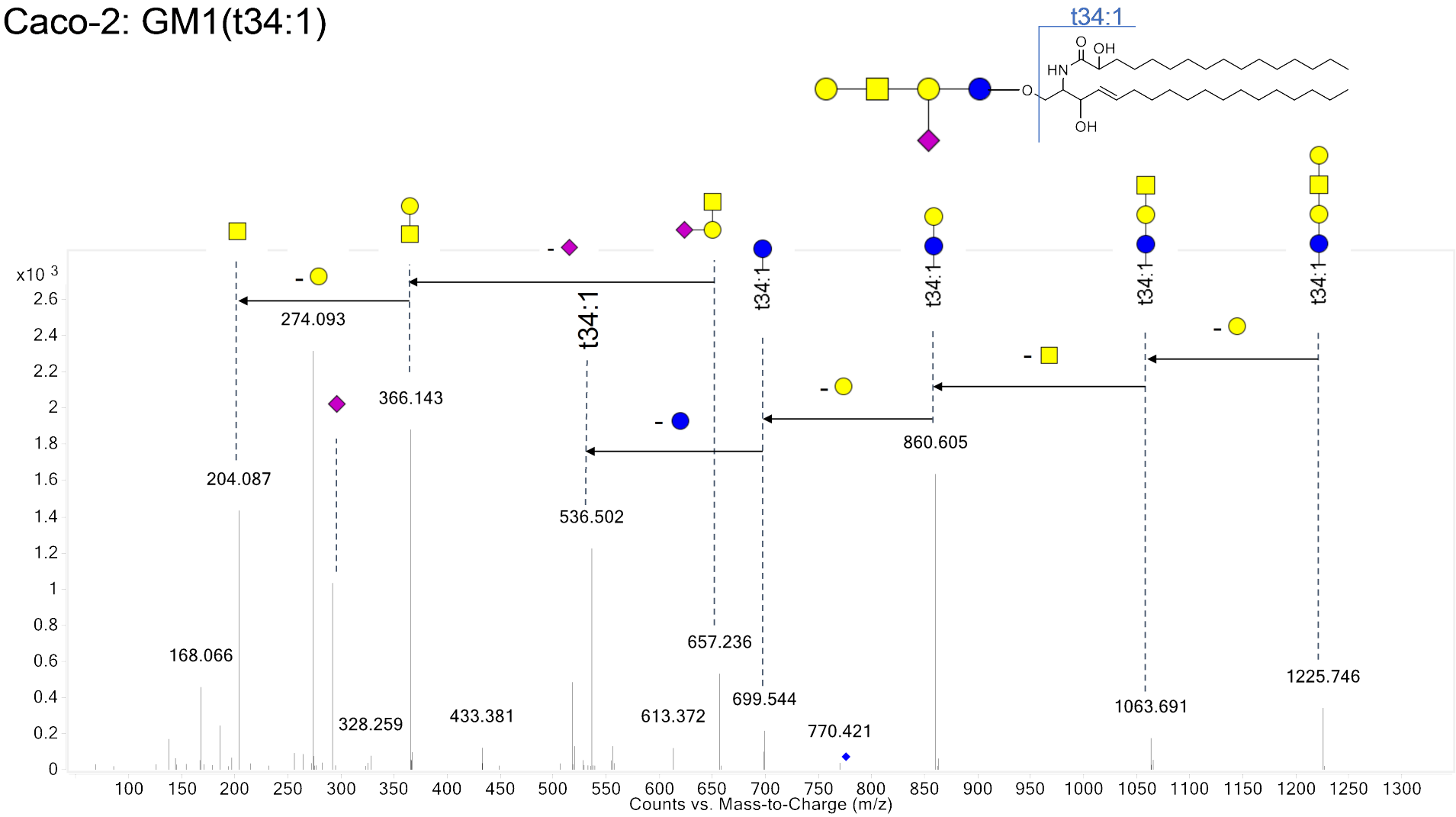


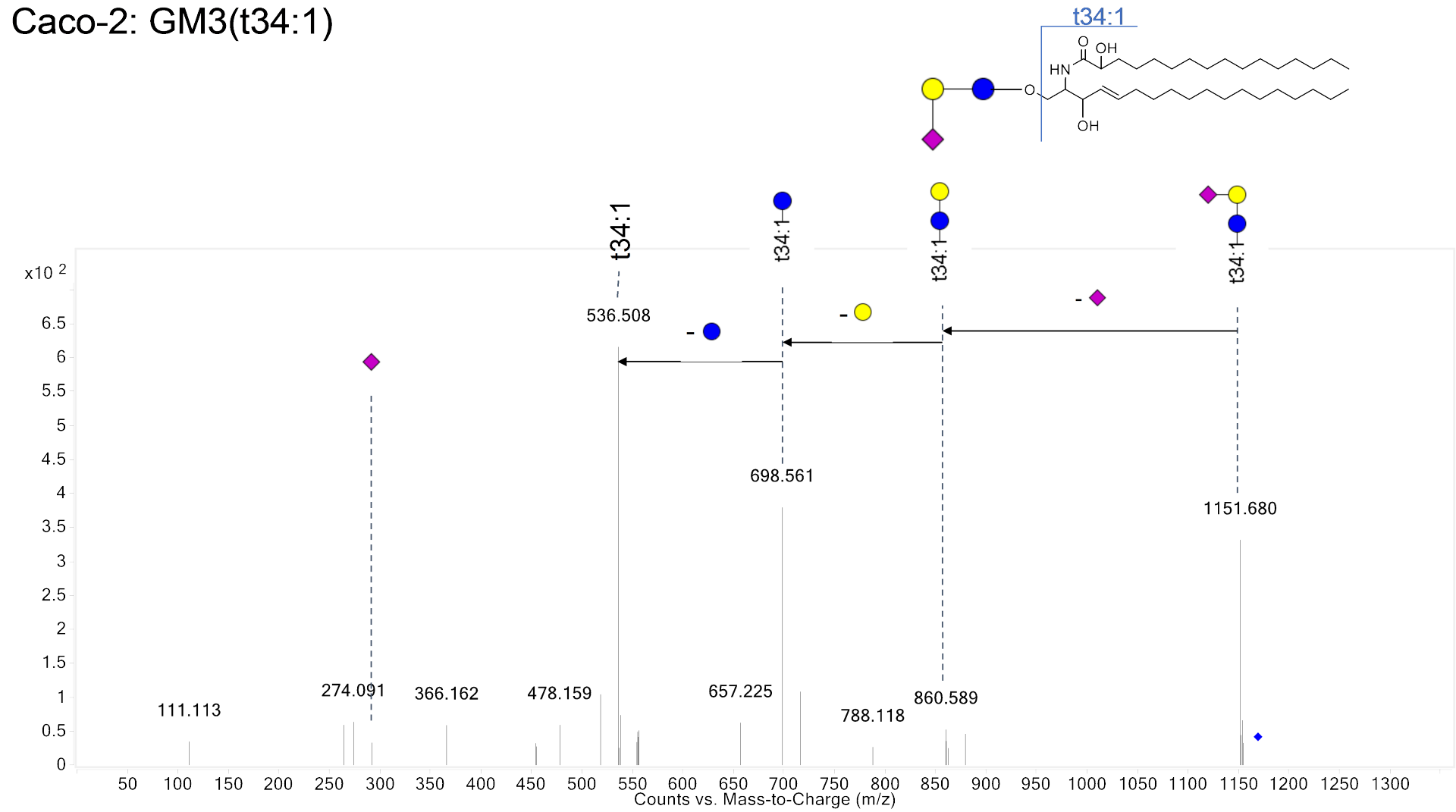
Figure S17. (A-D) CID fragmentation spectra of glycosphingolipid (GSL) structures derived from Caco-2. Each spectrum displays a fragmentation event as specified by *m/z* and symbolic representation. The precursor ion is indicated by a blue diamond. Sequential losses are annotated inset. Lc, lacto; Fuc, fucosylated.

B

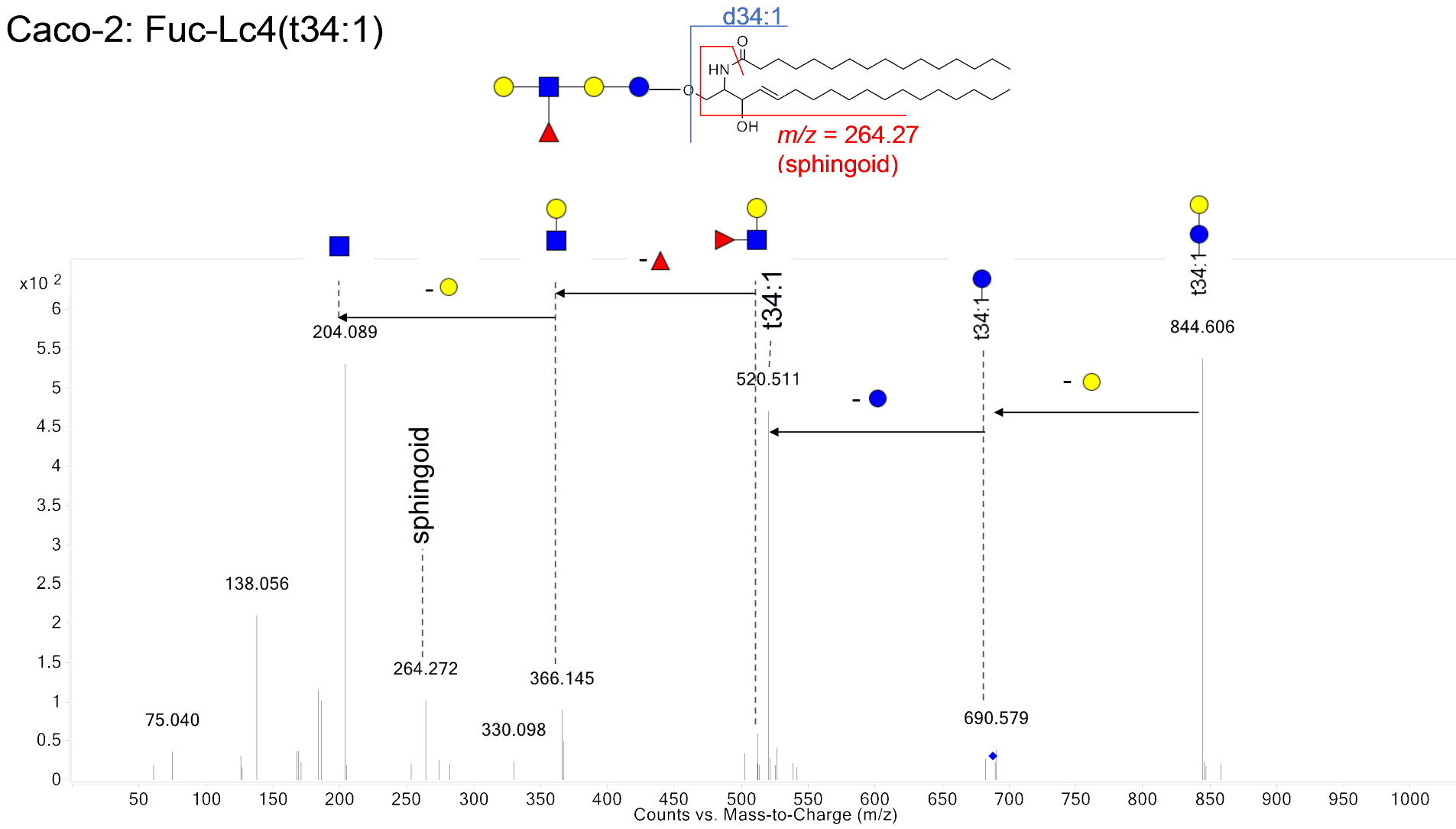

Caco-2: GM1(t34:1)
t34:1
HN
O
OH
OH
O
x10 3
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
274.093
366.143
204.087
t34:1
536.502
t34:1
657.236
t34:1
860.605
t34:1
1063.691
t34:1
1225.746
168.066
328.259
433.381
613.372
699.544
770.421
100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000 1050 1100 1150 1200 1250 1300
Counts vs. Mass-to-Charge (m/z)

c

Caco-2: GM3(t34:1)



D


Caco-2: Fuc-Lc4(t34:1)
d34:1
HN
O
OH
m/z = 264.27
(sphingoid)
x10 2
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
204.089
sphingoid
264.272
366.145
t34:1
520.511
t34:1
844.606
138.056
75.040
330.098
690.579
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
850
900
950
1000
Counts vs. Mass-to-Charge (m/z)

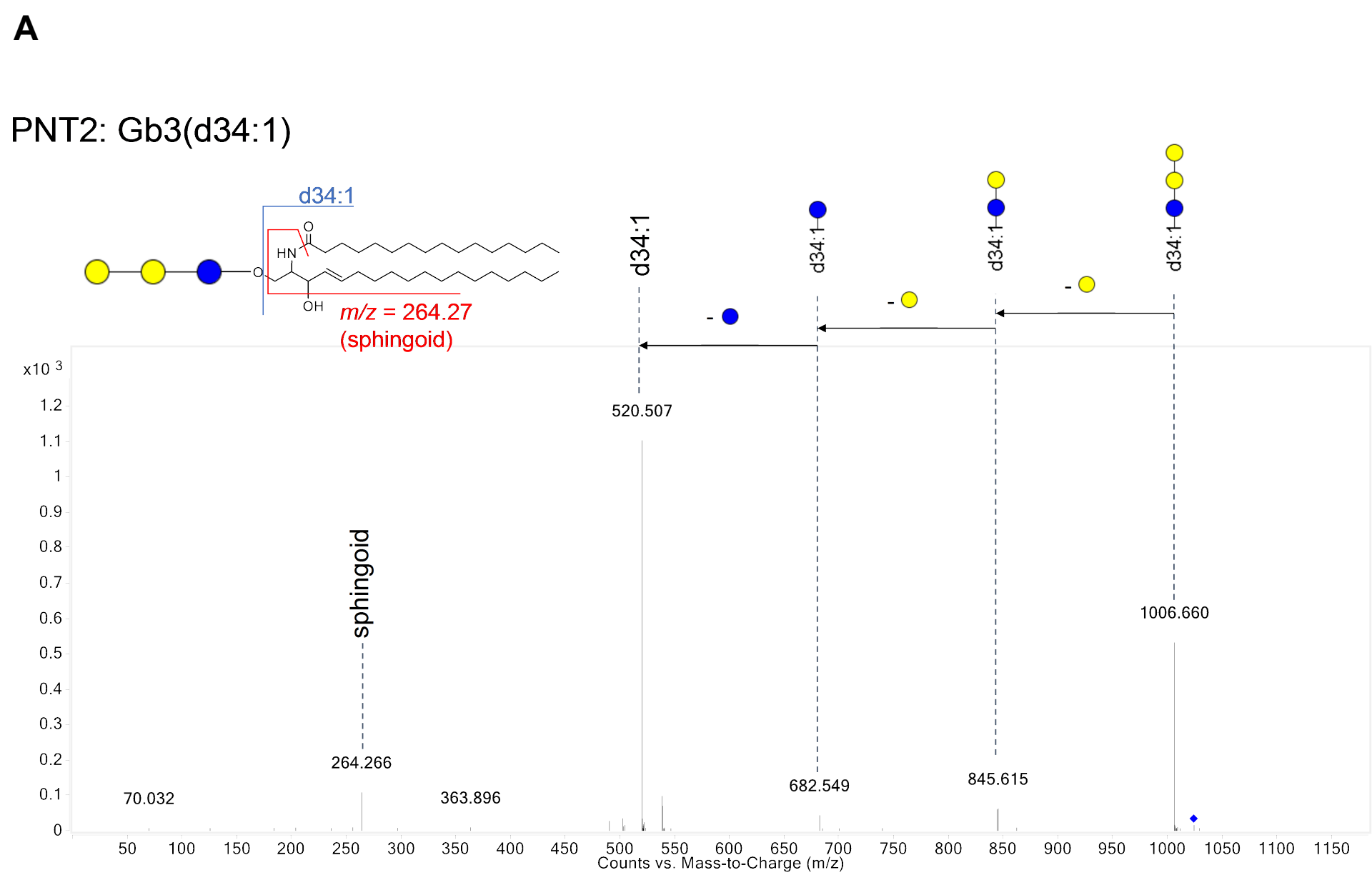
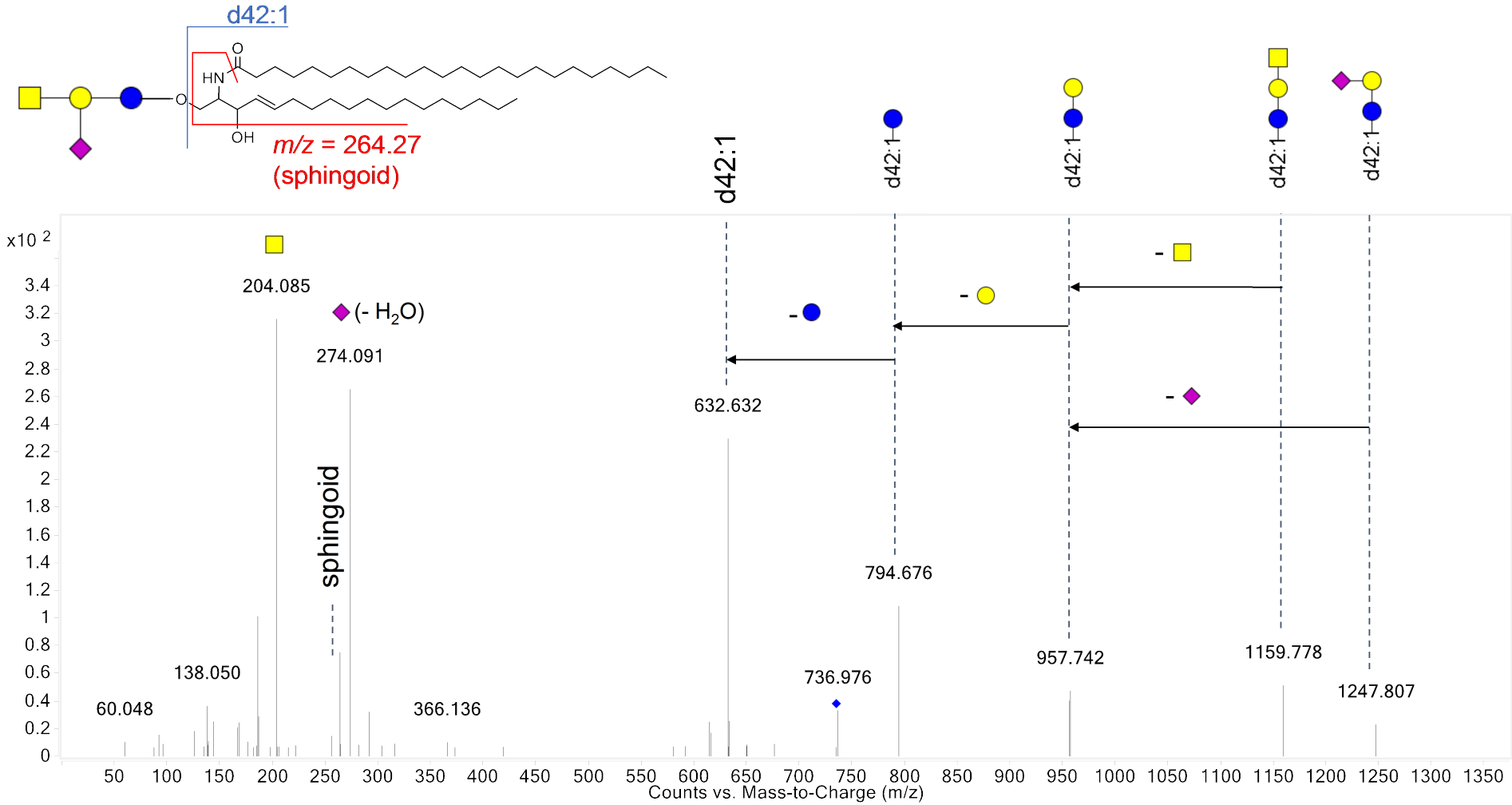


Figure S18. (A-E) CID fragmentation spectra of glycosphingolipid (GSL) structures derived from PNT2. Each spectrum displays a fragmentation event as specified by *m/z* and symbolic representation. The precursor ion is indicated by a blue diamond. Sequential losses are annotated inset. Gb, globo; Az, azido.

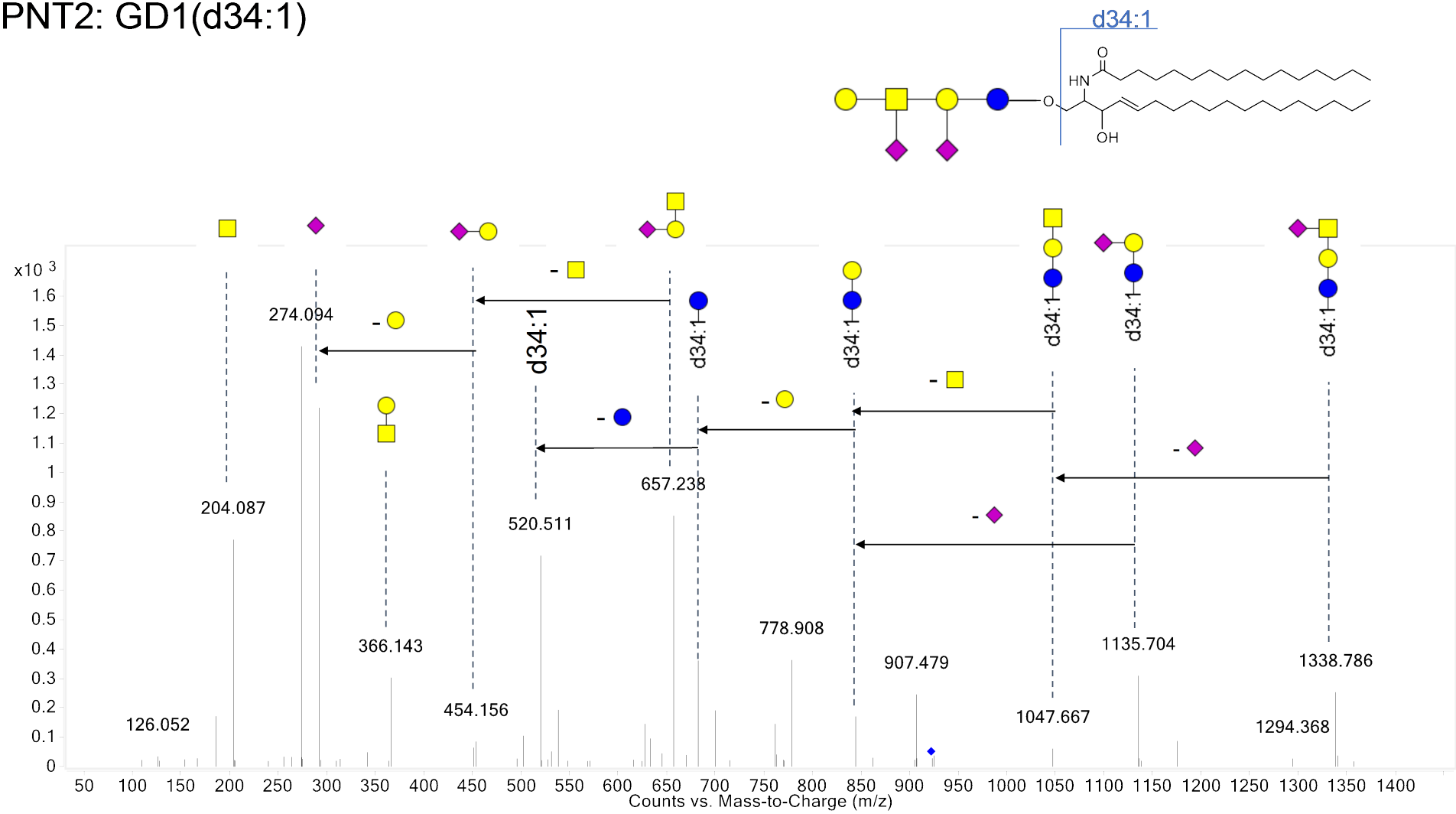
B

PNT2: GM2(d42:1)



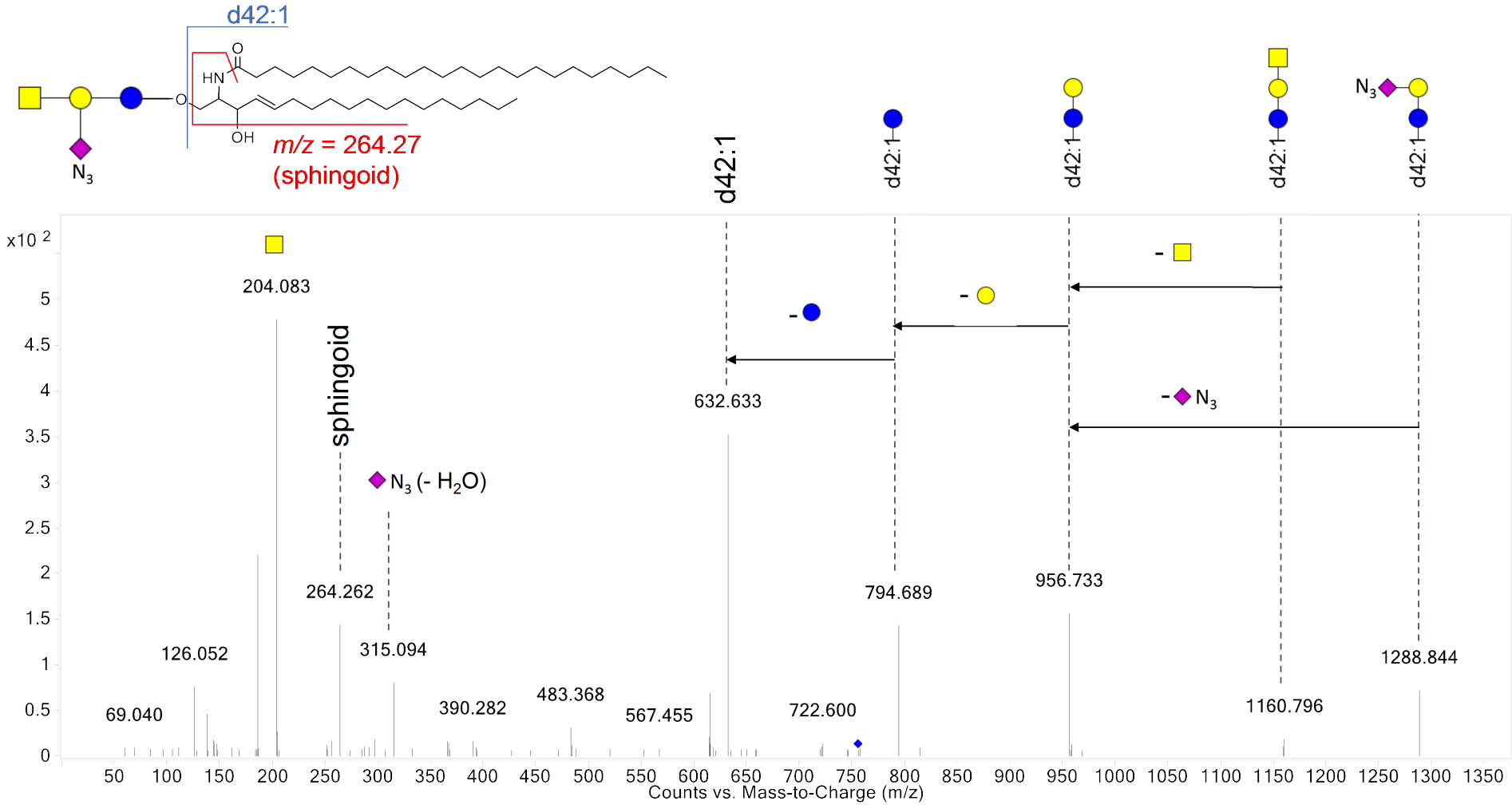
c

PNT2: GD1(d34:1)

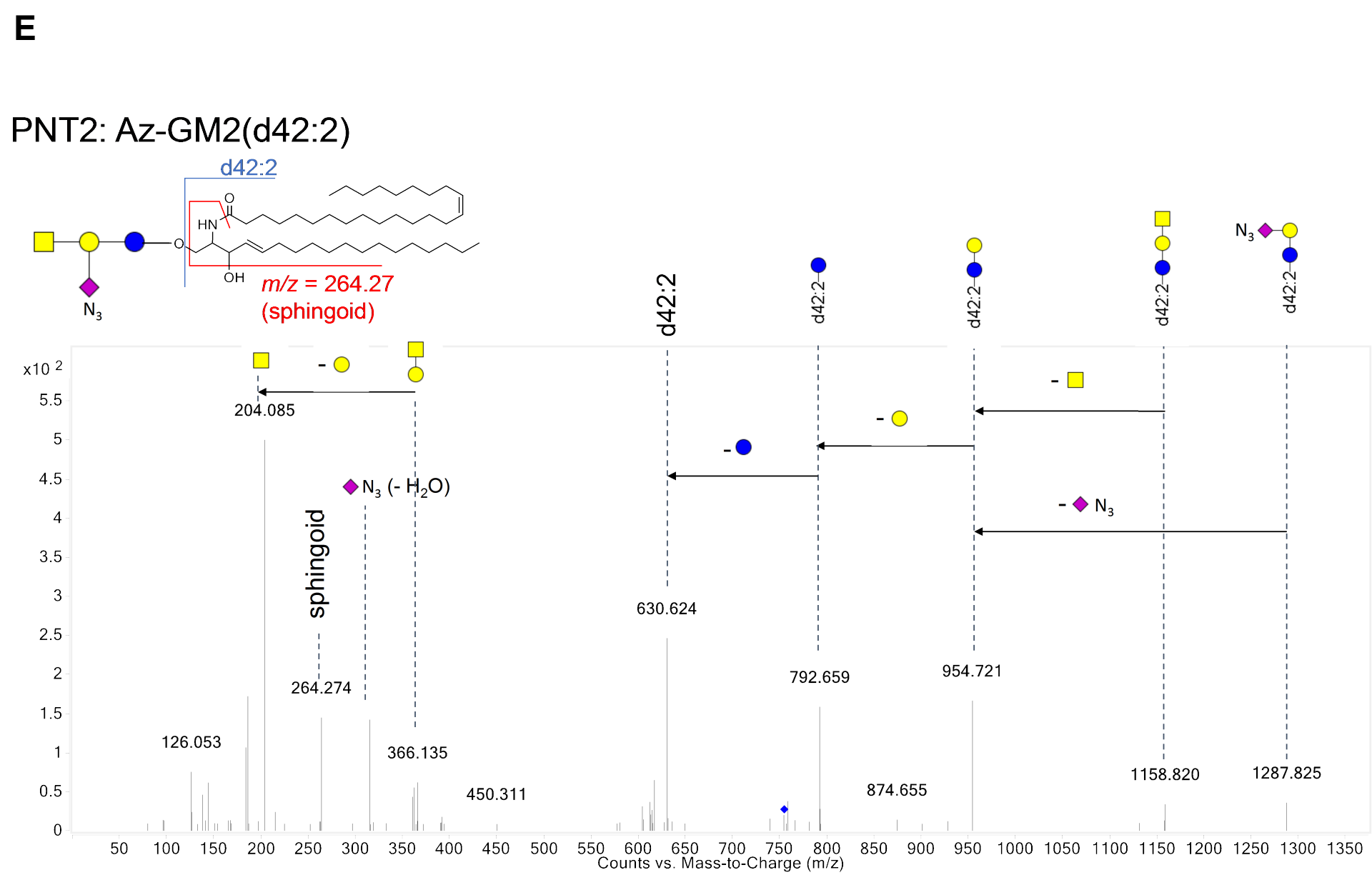


D

PNT2: Az-GM2(d42:1)



E


PNT2: Az-GM2(d42:2)
d42:2
HN
O
OH
O
N3
m/z = 264.27
(sphingoid)
d42:2
d42:2
d42:2
d42:2
d42:2
N3
x10 2
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
204.085
N3 (- H2O)
sphingoid
264.274
126.053
366.135
450.311
630.624
792.659
874.655
954.721
1158.820
1287.825
- N3
50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000 1050 1100 1150 1200 1250 1300 1350
Counts vs. Mass-to-Charge (m/z)