Identifying Sialylation Linkages at the Glycopeptide Level by Glycosyltransferase Labeling Assisted Mass Spectrometry (GLAMS)

He Zhu[†], Shuaishuai Wang[†], Ding Liu[†], Lang Ding[†], Congcong Chen[†], Yunpeng Liu[†], Zhigang Wu[†,#], Roni Bollag[‡], Kebin Liu[§], Jun Yin[†], Cheng Ma[†], Lei Li[*†] , Peng George Wang[*†,¶]

[†] Department of Chemistry, Georgia State University, Atlanta, GA 30303, United States

[‡] Georgia Cancer Center, Augusta University, Augusta, GA 30912, United States

[§] Department of Biochemistry and Molecular Biology, Augusta University, Augusta, GA 30912, United States

[#] Current address: College of Bioscience and Bioengineering, Hebei University of Science and Technology, Shijiazhuang, Hebei 050018, P. R. China

[¶] Current address: Department of Chemistry, Southern University of Science and Technology, Shenzhen, Guangdong 518055, P. R. China.

ABSTRACT

Precise assignment of sialylation linkages at the glycopeptide level is of importance in bottom-up glycoproteomics, and is also an indispensable step to understand the function of glycoproteins in pathogen-host interactions and cancer progression. Even though some efforts have been dedicated to the discrimination of α2,3/α2,6-sialylated isomers, unambiguous identification of sialoglycopeptide isomers is still needed. Herein, an innovative strategy of glycosyltransferase labeling assisted mass spectrometry (GLAMS) was developed. After specific enzymatic labeling, oxonium ions from higher-energy C-trap dissociation (HCD) fragmentation of α2,3-sailoglycopeptides generate unique reporters to distinctly differentiate those of α2,6-

24 sailoglycopeptide isomers. Using this strategy, a total of 1,236 linkage-specific
25 sialoglycopeptides were successfully identified from 161 glycoproteins in human serum.
26

27 **INTRODUCTION**

28 Glycosylation is one of the most important post-translational modifications (PTMs) of
29 proteins[1]. It plays an indispensable role in a wide variety of biological and physiological
30 processes including protein folding[2], cellular communication and signaling[3], immune response[4],
31 and pathogen-host interactions[5]. It was estimated that more than half of proteins in the human
32 proteome are glycosylated[6]. Glycosylation is generally classified into N-glycosylation and
33 mucin-type O-glycosylation. In N-glycosylation, the glycans are attached to asparagine residues
34 of proteins in the consensus motif Asn-X-Ser/Thr/Cys (X can be any amino acid except for
35 proline)[7], whereas the glycan moieties are linked to serine or threonine residues of proteins in O-
36 glycosylation[1]. N-acetylneuraminic acid (Neu5Ac) is most common form of diverse sialic acids
37 found on glycan chains in humans[8]. Neu5Ac is the frequently the terminal monosaccharide
38 residue that linked to galactose (Gal) via an α2,3 or an α2,6 linkage[9]. The two Neu5Ac linkage
39 isomers have important roles in pathogen binding. For example, avian influenza A viruses
40 preferentially bind to respiratory epithelial cells that contain α2,3-linked Neu5Ac moieties, while
41 human influenza viruses selectively bind to α2,6-linked Neu5Ac moieties[10]. In addition, altered
42 sialylation linkages is closely correlated with health status. In patients diagnosed with prostate
43 cancer, the structure of sialylated glycans released from the prostate-specific antigen (PSA) is
44 rich in α2,3 linkages, whereas the structure of sialylated glycans from healthy people's PSA is
45 predominantly via α2,6 linkages[11]. Without detailed information on sialylation linkages, it is
46 challenging to decipher the biological functions of sialoglycoproteins.

47 A variety of methods, including sialidase treatment[12], chemical derivatization[13-21], and
48 mass spectrometry (MS)[22-24], have been used to distinguish α2,3 and α2,6-sialylated glycans, but
49 the location of corresponding glycosylation sites is not evident. The emerging technology of MS-
50 based glycoproteomics[25-29] provides a high-throughput approach to analyze site-specific protein
51 glycosylation. Consequently, strategies for identifying sialylation linkages at the glycopeptide
52 level are required. Established chemical approaches[30, 31] modify not only sialic acids of the
53 glycan moiety but also acidic amino acid residues and the C-terminus of the peptide backbone,
54 which increases the complexity of spectra analysis, especially in relation to large scale
55 glycoproteomics data[32]. Pseudo MS^3 fragmentation[22] and ion mobility-mass spectrometry (IM-
56 MS)[33] have also been applied to distinguish sialoglycopeptide isomers but this technology has
57 not been **demonstrated (**?effective) for large-scale glycopeptides analysis. Even though the
58 fragmentation of high energy C-trap dissociation (HCD) can achieve the high-throughput
59 assignment of α2,3/α2,6-sialoglycopeptide isomers, there are still many ambiguous
60 glycopeptides without sialylation linkage information[34].

61 Enzymatic labeling has become a promising tool for probing proteins with specific
62 glycosylation because of its high specificity and mild reaction condition[35-41]. In our previous
63 research, a β1,4-N-acetylgalactosaminyltransferase from *Campylobacter jejuni* (CgtA) was
64 demonstrated to specifically recognize the Neu5Ac(α2,3)Gal epitope and transfer N-
65 acetylgalactosamine (GalNAc) or N-azidoacetylgalactosamine (GalNAz) to Gal residue via β1,4
66 linkage[42]. Taking advantage of the specificity of CgtA, we herein report a new strategy for
67 identification of α2,3 and α2,6 sialoglycopeptide isomers by glycosyltransferase labeling assisted
68 mass spectrometry (GLAMS). As shown in **Scheme 1**, sialoglycopeptides (monosialylation for
69 example) without linkages information were treated with GalNAz labeling by CgtA. After HCD

70 fragmentation, if only the oxonium ions of HexNAc (N-acetylhexosamine, m/z 204) and Neu5Ac
71 (m/z 292) are observed in the spectrum, this identifies an α2,6 linkage; if one more oxonium ion
72 m/z 245 from GalNAz is included, then it represents an α2,3 linkage. We demonstrated the
73 capacity of GLAMS to effectively differentiate α2,3/α2,6 isomers using synthetic
74 sialoglycopeptides, and applied it to profile sialylated N-glycopeptides digested from serum
75 samples, in which we identified 1,236 linkage specific sialoglycopeptides with an improved
76 version of pGlyco 2.0 software[27].
77

78 **EXPERIMENTAL SECTION**

79 **Chemicals and materials**

80 Amicon Ultra-0.5 Centrifugal Filter Unit with NMWL of 30 kD was obtained from
81 Millipore Sigma (St. Louis, MO). iSPE® HILIC SPE Cartridge was purchased from The Nest
82 Group (Southborough, MA). XAmide medium was provided by Dr. Xinmiao Liang's lab (Dalian
83 Institute of Chemical Physics, Chinese Academy of Sciences). Sequencing grade porcine trypsin
84 was purchased from Promega (Madison, WI); Deionized water was produced by a Milli-Q A10
85 system from Millipore (Bedford, MA). Optimal LC-MS grade quality acetonitrile (ACN), formic
86 acid (FA) and water were purchased from Thermo Fisher Scientific, Inc. (Waltham, MA).
87 Dithiothreitol (DTT), 2-Iodoacetamide (IAA) and trifluoroacetic acid (TFA) were bought from
88 Alfa Aesar (Ward Hill, MA). Ammonium bicarbonate (NH_4HCO_3) was purchased from
89 Millipore Sigma (St. Louis, MO). All other chemicals of analytical grade were commercially
90 available.

91 Sialylated oligosaccharide standards including 3'-sialyllactose (3SL) and 6'-sialyllactose
92 (6SL) were enzymatically synthesized using *Pasteurella multocida* α2,3-sialyltransferase 1 and
93 *Photobacterium damselae* α2,6-sialyltransferase as described before[43].
94 Serum samples from four healthy people, ten patients with colon cancer and four patients
95 with liver cancer were provided by Georgia Cancer Center at Augusta University and stored at -
96 80 °C until use. The protocol for serum sample preparation was approved by the Institutional
97 Review Board of Augusta University and was performed in accordance with the Helsinki
98 Declaration. All participants gave written informed consents.

99 **Protein digestion**

100 The protein digestion was performed using a filter aided sample preparation (FASP)[44]
101 method with minor modification. Approximately 1 mg of total proteins from human serum was
102 subjected to FASP procedures. Samples were denatured in 20 μL of lysis buffer (4% SDS, 100
103 mM DTT, 20 mM Tris-HCl, pH 7.6) at 95 °C for 10 min. When the solution was cooled down to
104 room temperature, 1 mL of UA solution (8 M urea, 20 mM Tris-HCl, pH 8.5) was added. The
105 solution was gradually transferred into a 30 kD filter and centrifugated at 15,000 rpm at 20 °C
106 for 15 min. Then 100 μL of IAA solution (50 mM IAA in UA solution) was added to the filter
107 and incubated for 30 min in the darkness followed by centrifugation at 15,000 rpm for 10 min.
108 After that, 100 μL of UA solution and 200 μL of digestion buffer (100 μL of 50 mM NH_4HCO_3
109 each time) were added and centrifuged at 15,000 rpm for 10 min, respectively. Finally, 20 μg of
110 trypsin was added into the filter and incubated at 37 °C for 12 h. The tryptic peptides were eluted
111 using 100 μL of 50 mM NH_4HCO_3 for six times. The concentration of the collected peptides was
112 measured based on absorbance at 280 nm with Nanodrop. Five hundred μg of the peptides were

113 taken and subsequently dried in Vacufuge at room temperature, then were lyophilized twice to
114 remove NH_4HCO_3. The dried samples were stored at −20 °C for further use.

115 **Enrichment of glycopeptides**

116 The enrichment of glycopeptides was performed using an iSPE HILIC cartridge by the
117 procedure of ZIC-HILIC[45]. Briefly, the HILIC cartridge was pre-washed with 300 μL of 0.1%
118 TFA and conditioning with 600 μL of 80% ACN containing 0.1% TFA. Peptide samples
119 dissolved in 400 μL of 80% ACN containing 0.1% TFA were loaded onto the cartridge. The
120 flow-through was reloaded onto the column for another two times. Then the column was washed
121 with 1.2 mL of 80% ACN containing 1% TFA. The glycopeptides were eluted using 750 μL of
122 0.1% TFA, 60 μL of H_2O, 60 μL of 25 mM NH_4HCO_3 and 60 μL of 50% ACN. The fractions
123 were combined followed by lyophilization and stored at -20 °C.

124 **Enzymatic labeling of glycopeptides**

125 The β1,4-N-acetylgalactosaminyltransferase from *C. jejuni* (CgtA) and sugar nucleotide
126 UDP-N-azidoacetylgalactosamine (UDP-GalNAz) were prepared as described previously[42]. For
127 enzymatic labeling, the 40 μL reaction mixture containing the enriched glycopeptides, 50 mM
128 Tris-HCl buffer (pH 7.5), 5 mM $MgCl_2$, 200 μM UDP-GalNAz and 10 μg CgtA was incubated
129 in a 37 °C water bath for 1 h. The reaction was quenched by adding 60 μL of ACN, 6 μL of 10%
130 TFA and 215 μL of ACN containing 1% TFA. Then the glycopeptides were purified with an in-
131 house packed Amide HILIC microcolumn (6 mg of XAmide HILIC medium in a 200 μL tip)
132 which was preconditioned with 100 μL of 0.1% TFA and 200 μL of 85% ACN containing 0.1%
133 TFA. After loading the reaction mixture onto the tip column, the column was washed with 100
134 μL of 85% ACN, 0.1% TFA to remove salts and proteins. The glycopeptides were collected
135 using 200 μL of 0.1% TFA and lyophilized.

136 **Nano LC-MS/MS analysis**

137 Nano RP HPLC-MS experiments were performed on an LTQ-Orbitrap Elite mass
138 spectrometer (Thermo Fisher) equipped with EASY-spray source and nano-LC UltiMate 3000
139 high-performance liquid chromatography system (Thermo Fisher). The separation was
140 performed on an EASY-Spray PepMap C18 Column (75 μm × 50 cm, 2 μm, ThermoFisher, US)
141 with a linear gradient from 3% to 40% mobile phase B for 120 min at a flow rate of 300 nL/min
142 (mobile phase A: 1.95% ACN, 97.95% H_2O, 0.1% FA; mobile phase B: 79.95% ACN, 19.95%
143 H_2O, 0.1% FA). The mass spectrometer was operated in the data-dependent mode. A full-scan
144 survey MS experiment (m/z range from 375 to 2000; automatic gain control target, 1,000,000
145 ions; resolution at 400 m/z, 60,000; maximum ion accumulation time, 50 ms) was acquired by
146 the Orbitrap mass spectrometer, and ten most intense ions were fragmented by higher-energy C-
147 trap dissociation (HCD) with an optimized stepped collision energy (SCE) of 30 ± 15% NCE.

148 **Data analysis**

149 The pGlyco 3.0 software (an improved version of pGlyco 2.0[27]) was used to identify N-
150 glycopeptides. In the glycan database, we defined an unnatural monosaccharide GalNAz (Gaz,
151 Z). We created all the possible "Gaz" containing human N-glycan sequences following the motif
152 H(A)(Z), where "H" represents Galactose, "A" represents Neu5Ac and "Z" represents GalNAz.
153 For the sialoglycopeptides identification, the raw data was first converted to an mgf file using
154 pParse[46]. The potential sialoglycopeptides spectra containing oxonium ions of both GlcNAc (m/z
155 204.09, m/z 186.08) and Neu5Ac (m/z 292.10, m/z 274.09) were extracted and merged into a
156 new mgf file using an in-house program. The new mgf file was searched against an Uniprot-
157 Swiss human protein database (20,417 reviewed entries on March 20, 2019) and a modified N-
158 glycan database. Other searching parameters were used including carbamidomethyl (Cys) as a

159 fixed modification, deamination (Asn) and oxidation (Met) as variable modifications, and trypsin
160 as the digestion enzyme with maximum two missed cleavage sites allowed. The mass tolerance
161 for the precursor ions and the fragment ions was set to 10 ppm and 20 ppm, respectively. A total
162 false discovery rate (FDR) of 1% was applied to all data sets at the glycopeptide level. The
163 identified glycoproteins were annotated using Uniprot ID mapping[47].
164

165 **RESULTS AND DISCUSSION**

166 **The specificity of CgtA for Sialoglycopeptides**

167 Our previous study showed that CgtA had a high specificity for (Neu5Ac-α2,3-Gal)-
168 containing glycans[42]. To test its selectivity for sialoglycopeptides, we synthesized 23SP and
169 26SP (synthesis and data not published) as shown in **Figure 1**. The activity of CgtA was
170 validated with sialolactose prior to the enzymatic reaction of sialoglycopeptides. After labeling
171 with GalNAz by CgtA, only α2,3-sialylated lactose showed a mass shift of 244 Da in the MS
172 spectrum and no substrate was obviously observed (**Figure S1**), which indicated that CgtA had
173 good catalytic activity and high specificity as well.

174 ESI-MS was used to monitor the GalNAz labeling reaction for sialoglycopeptides. For
175 the α2,3-sialoglycopeptide 23SP, the monoisotopic peak at m/z 1032.0808 is from the triply
176 charged glycopeptide (**Figure S2A**). After incubating with CgtA and UDP-GalNAz for 30 min, a
177 new dominant peak at m/z 1113.4422 with three positive charges was observed, which
178 correspond to the GalNAz labeled 23SP (**Figure S2C**). However, we could still detect minor
179 substrate 23SP. The α2,6-sialoglycopeptide 26SP has the same exact mass (**Figure S2B**) as 23SP.
180 After GalNAz labeling by CgtA, the strongest peak is located at the same exact position, and a

181 trace amount of product at m/z 1113.4422 was detected (**Figure S2D**). This may result from the
182 very weak activity of CgtA for 2,6-sialylated glycans[42].
183 The reaction time was further optimized to increase the conversion rate of GalNAz
184 labeling of 23SP. The reaction mixture at different time points was analyzed by LC-MS and the
185 reaction efficiency was calculated based on the peak area. For 23SP, in the first half an hour, the
186 reaction efficiency was dramatically increased. After that, the increase slowed down and the
187 reaction efficiency approached 100% at 60 min. Meanwhile, the conversion rate for the 26SP
188 remained unchanged (less than 1%) over time (**Figure 2**). Finally, we chose 60 min as the
189 reaction time for the subsequent GalNAz labeling reactions.
190
191 **Identifying Sialoglycopeptide Linkage Isomers by GalNAz Labeling and Tandem MS**
192 Although discrimination of α2,3/α2,6-sialoglycopeptide isomers can be achieved by the
193 mass shift in primary MS as shown above, we cannot accurately determine the glycopeptide
194 sequence using only molecular weight information. Tandem MS is an effective way for
195 sequencing glycopeptides, and HCD-MS/MS is commonly used to analyze glycopeptides
196 because of the high accuracy and detectable diagnostic oxonium ions of monosaccharides[26-28, 48].
197 It was also demonstrated that HCD with stepped collision energy (SCE) can simultaneously
198 determine the sequences of both glycans and peptides in a single run[27, 48]. To achieve a better
199 sequence coverage of glycopeptides on the Orbitrap we used, SCE was optimized with
200 glycopeptides enriched from human serum. The identification of N-glycopeptides was performed
201 using pGlyco 2.0 software[27]. A series of two-step SCEs and three-step SCEs were employed to
202 compare the number of identified glycopeptides. As shown in **Figure S3**, with a three-step SCE
203 of "15-30-45", more glycopeptides can be identified from the same sample.

204 The optimized SCE-HCD-MS/MS was applied for sequencing 23SP and 26SP. After
205 GalNAz labeling, 23SP (**Figure 3A**) showed diagnostic oxonium ions of GlcNAc and Neu5Ac at
206 m/z 204.09 and 292.10, which are common features of sialoglycopeptides. Apart from these
207 oxonium ions, 23SP also had a strong unknown peak at m/z 245.09. Compared with the spectrum
208 of 23SP before GalNAz labeling (**Figure S4**), we can obviously see that the peak with m/z
209 245.09 corresponds to the oxonium ion of GalNAz. Following the dissociation pathway of
210 GlcNAc[39, 49, 50] in **Scheme S1A**, a series of fragmentation ions of GalNAz in HCD were
211 illustrated in **Scheme S1B**. The peaks at m/z 245.09, m/z 227.08, m/z 209.07 and m/z 179.06,
212 which match the theoretical values very well, are diagnostic ions of GalNAz (**Figure S5**). For
213 26SP (**Figure 3B**), its spectrum showed that it shared similar fragmentation patterns with
214 GalNAz labeled 23SP. In the low m/z region, only diagnostic ions of GlcNAc and Neu5Ac were
215 observed, which clearly demonstrated that the GLAMS is an effective approach to distinguish
216 intact linkage specific sialoglycopeptides.
217
218 **Linkage Specific Sialoglycopeptides in Human Serum**
219 We then applied this method to analyze the N-linked α2,3/α2,6-sialoglycopeptides of
220 complex biological samples. Serum samples from healthy individuals (n = 4) and patients
221 diagnosed with liver cancer (n = 4) and colon cancer (n = 10) were used in this study. In the
222 beginning, we tried to perform the enzymatic labeling of the mixture of tryptic peptides and
223 glycopeptides. After drying from the FASP eluents, the mixture was not completely dissolved in
224 the Tris buffer (pH 7.5) for the enzymatic reaction, which may result from the existence of some
225 basic peptides. To solve the solubility issue, we adopted step-by-step glycopeptides enrichment
226 and enzymatic labeling as shown in the workflow **Figure 4**. It turned out that the enriched

227 glycopeptides were dissolved in the basic buffer very well. This may be attributed to that
228 conjugated glycans increase the hydrophilicity of peptides and some sialoglycans decrease the
229 isoelectric points of peptides according to the previous findings[51, 52].

230 All the serum samples were analyzed in three replicated LC-MS/MS runs for a large-
231 scale study of N-linked sialoglycopeptides. Using the oxonium ions of GalNAz as a tag to mark
232 α2,3-sialylation, a total of 22,819 sialoglycopeptide spectra were identified with a 1% total FDR,
233 corresponding to 1,236 unique site-specific sialoglycans on 261 N-glycosylation sites from 161
234 unique glycoproteins (**Table S1**), among of which 53% was also identified previously using a
235 deglycosylation method[53] as shown in **Figure S6**. For the distribution analysis of α2,3 or α2,6-
236 sialylation (**Figure 5A**), only 48 distinct site-specific sialoglycopeptides contained pure α2,3-
237 sialylation linkage, 236 sialoglycopeptides contained hybrid sialylation, and all the rest were
238 uniquely 2,6-sialylated, which accounts for more than 77% of the identified sialoglycopeptides.
239 This is consistent with the previous result that α2,6-sialylation is dominant in sialylated N-
240 glycans released from human serum samples[54]. Taking a closer look at the distribution of
241 sialylation levels, monosialylation is the major component in strictly α2,3 or α2,6-
242 sialoglycopeptides, and 74% of the hybrid sialoglycopeptides are disialylated.

243 The percentage of the overall overlapped glycosites (**Figure 5C**) and glycoproteins
244 (**Figure 5D**) in a specific type of serum is 72-80% and 65-78%, respectively. The percentage of
245 overlapping sialoglycopeptides (**Figure 5B**) in each kind of serum sample was significantly
246 lower than those of the glycosites and glycoproteins, reflecting the diversity of site-specific
247 glycosylation, especially for samples from colon cancer and liver cancer patients. A highly
248 glycosylated protein P01876 (IGHA1) was selected to demonstrate the diversity of site-specific
249 glycosylation in the three different sera. The previous study showed that IgA was α2,3 and α2,6-

250 sialylated on N-glycans[55]. Here, we identified 35 unique sialoglycopeptides including α2,3 and
251 α2,6 linkages on one N-glycosylation site of IGHA1 (**Table S2**). In addition, we found some
252 unique sialoglycopeptides of IGHA1 expressed in a serum-specific manner. For example, only
253 unique α2,6-sialoglycopeptides were identified in serum from liver cancer patients
254 (representative spectrum in **Figure S7**) and normal people (representative spectrum in **Figure**
255 **S8**), whereas unique α2,6-sialoglycopeptides and one hybrid sialoglycopeptide were detected
256 from colon cancer patients' sera (representative spectra in **Figure S9**).

257 **CONCLUSIONS**

258 Elucidating linkage-specific sialylation of glycoproteins is significant to understand
259 pathogen-host interactions and disease development. By integrating enzymatic labeling and MS-
260 based glycoproteomics, we developed a high-throughput glycosyltransferase labeling assisted
261 mass spectrometry (GLAMS) to identify sialoglycopeptide linkage isomers. Facilitated by the
262 acceptor specificity of the glycosyltransferase CgtA toward the Neu5Ac(α2,3)Gal epitope, we
263 have demonstrated the distinct diagnostic oxonium ions as tags to effectively discriminate α2,3-
264 and α2,6-sialoglycopeptide isomers using LC separation coupled with SCE-HCD-MS/MS
265 fragmentation. We also successfully applied this approach to a large-scale profiling of linkage-
266 specific sialoglycopeptides in human serum samples. The strategy of GLAMS should be of great
267 benefit in investigating differential changes in response to disease states from a glycoproteomics
268 perspective, and has potential to contribute to biomarker discovery in many diseases.

269

270

271

272

AUTHOR INFORMATION

Corresponding Authors

lli22@gsu.edu (L. L.) or pwang11@gsu.edu (P. G. W.)

Author Contributions

H.Z., L. L., and P. G. W. conceived the project; H.Z. designed and performed the experiments, analyzed the data, and wrote the manuscript; S. W. synthesized the sialoglycopeptides; D. L. and L. D. prepared the serum glycopeptides for collision energy optimization; C. C., Y. L. and Z. W. synthesized the UDP-GalNAz; R. B. and K. L. collected and prepared the human serum samples; C. M., J. Y., L. L., and P. G. W. edited the manuscript. All the authors revised and approved the manuscript.

ACKNOWLEDGEMENTS

We sincerely thank Dr. Wenfeng Zeng from the Key Lab of Intelligent Information Processing of Chinese Academy of Sciences for the help of parameter setting of pGlyco 3.0, Georgia Cancer Center Biorepository for providing human serum samples, and Georgia Research Alliance for financial support to purchase mass spectrometers. This work is supported by National Institute of Health (U01GM116263 and U54HL142019).

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

296 **REFERENCES**

297 1. Ohtsubo, K.; Marth, J. D., Glycosylation in cellular mechanisms of health and disease.
298 *Cell* **2006,** *126* (5), 855-867.
299 2. Banerjee, S.; Vishwanath, P.; Cui, J.; Kelleher, D. J.; Gilmore, R.; Robbins, P. W.;
300 Samuelson, J., The evolution of N-glycan-dependent endoplasmic reticulum quality control
301 factors for glycoprotein folding and degradation. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104* (28),
302 11676-11681.
303 3. Furukawa, K.; Ohkawa, Y.; Yamauchi, Y.; Hamamura, K.; Ohmi, Y.; Furukawa, K.,
304 Fine tuning of cell signals by glycosylation. *J. Biochem.* **2012,** *151* (6), 573-578.
305 4. Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A., Glycosylation and
306 the immune system. *Science* **2001,** *291* (5512), 2370-2376.
307 5. Karlsson, K. A., Pathogen-host protein-carbohydrate interactions as the basis of
308 important infections. *Molecular Immunology of Complex Carbohydrates-2* **2001,** *491*, 431-443.
309 6. Apweiler, R.; Hermjakob, H.; Sharon, N., On the frequency of protein glycosylation, as
310 deduced from analysis of the SWISS-PROT database. *BBA-Gen. Subjects* **1999,** *1473* (1), 4-8.
311 7. Zielinska, D. F.; Gnad, F.; Wisniewski, J. R.; Mann, M., Precision Mapping of an In
312 Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* **2010,** *141* (5),
313 897-907.
314 8. Varki, N. M.; Varki, A., Diversity in cell surface sialic acid presentations: implications
315 for biology and disease. *Lab. Invest.* **2007,** *87* (9), 851-857.
316 9. Varki, A.; Schauer, R., Sialic Acids. In *Essentials of Glycobiology*, nd; Varki, A.;
317 Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler,
318 M. E., Eds. Cold Spring Harbor (NY), 2009.
319 10. Stevens, J.; Blixt, O.; Paulson, J. C.; Wilson, I. A., Glycan microarray technologies:
320 tools to survey host specificity of influenza viruses. *Nat. Rev. Microbiol.* **2006,** *4* (11), 857-864.
321 11. Hatakeyama, S.; Yoneyama, T.; Tobisawa, Y.; Ohyama, C., Recent progress and
322 perspectives on prostate cancer biomarkers. *Int. J. Clin. Oncol.* **2017,** *22* (2), 214-221.
323 12. Mancera-Arteu, M.; Gimenez, E.; Barbosa, J.; Sanz-Nebot, V., Identification and
324 characterization of isomeric N-glycans of human alfa-acid-glycoprotein by stable isotope
325 labelling and ZIC-HILIC-MS in combination with exoglycosidase digestion. *Anal. Chim. Acta.*
326 **2016,** *940*, 92-103.
327 13. Mechref, Y.; Kang, P.; Novotny, M. V., Differentiating structural isomers of sialylated
328 glycans by matrix-assisted laser desorption/ionization time-of-flight/time-of-flight tandem mass
329 spectrometry. *Rapid Commun. Mass Spectrom.* **2006,** *20* (8), 1381-1389.
330 14. Alley, W. R., Jr.; Novotny, M. V., Glycomic analysis of sialic acid linkages in glycans
331 derived from blood serum glycoproteins. *J. Proteome. Res.* **2010,** *9* (6), 3062-3072.
332 15. Jiang, K.; Zhu, H.; Li, L.; Guo, Y. X.; Gashash, E.; Ma, C.; Sun, X. L.; Li, J.; Zhang,
333 L. W.; Wang, P. G., Sialic acid linkage-specific permethylation for improved profiling of protein
334 glycosylation by MALDI-TOF MS. *Anal. Chim. Acta.* **2017,** *981*, 53-61.
335 16. Wheeler, S. F.; Domann, P.; Harvey, D. J., Derivatization of sialic acids for stabilization
336 in matrix-assisted laser desorption/ionization mass spectrometry and concomitant differentiation
337 of α(2→3) and α(2→6)-isomers. *Rapid Commun. Mass Spectrom.* **2009,** *23* (2), 303-312.
338 17. Reiding, K. R.; Blank, D.; Kuijper, D. M.; Deelder, A. M.; Wuhrer, M., High-
339 Throughput Profiling of Protein N-Glycosylation by MALDI-TOF-MS Employing Linkage-
340 Specific Sialic Acid Esterification. *Anal. Chem.* **2014,** *86* (12), 5784-5793.

18. Holst, S.; Heijs, B.; de Haan, N.; van Zeijl, R. J. M.; Briaire-de Bruijn, I. H.; van Pelt, G. W.; Mehta, A. S.; Angel, P. M.; Mesker, W. E.; Tollenaar, R. A.; Drake, R. R.; Bovee, J. V. M. G.; McDonnell, L. A.; Wuhrer, M., Linkage-Specific in Situ Sialic Acid Derivatization for N-Glycan Mass Spectrometry Imaging of Formalin-Fixed Paraffin-Embedded Tissues. *Anal. Chem.* **2016,** *88* (11), 5904-5913.
19. Yang, S.; Jankowska, E.; Kosikova, M.; Xie, H.; Cipollo, J., Solid-Phase Chemical Modification for Sialic Acid Linkage Analysis: Application to Glycoproteins of Host Cells Used in Influenza Virus Propagation. *Anal. Chem.* **2017,** *89* (17), 9508-9517.
20. Li, H. H.; Gao, W. J.; Feng, X. J.; Liu, B. F.; Liu, X., MALDI-MS analysis of sialylated N-glycan linkage isomers using solid-phase two step derivatization method. *Anal. Chim. Acta.* **2016,** *924*, 77-85.
21. Nishikaze, T.; Tsurnoto, H.; Sekiya, S.; Iwamoto, S.; Miura, Y.; Tanaka, K., Differentiation of Sialyl Linkage Isomers by One-Pot Sialic Acid Derivatization for Mass Spectrometry-Based Glycan Profiling. *Anal. Chem.* **2017,** *89* (4), 2353-2360.
22. Chen, C. H.; Lin, Y. P.; Lin, J. L.; Li, S. T.; Ren, C. T.; Wu, C. Y.; Chen, C. H., Rapid Identification of Terminal Sialic Acid Linkage Isomers by Pseudo-MS3 Mass Spectrometry. *Isr. J. Chem.* **2015,** *55* (3-4), 412-422.
23. Guttman, M.; Lee, K. K., Site-Specific Mapping of Sialic Acid Linkage Isomers by Ion Mobility Spectrometry. *Anal. Chem.* **2016,** *88* (10), 5212-5217.
24. Depland, A. D.; Renois-Predelus, G.; Schindler, B.; Compagnon, I., Identification of sialic acid linkage isomers in glycans using coupled InfraRed Multiple Photon Dissociation (IRMPD) spectroscopy and mass spectrometry. *Int. J. Mass Spectrom.* **2018,** *434*, 65-69.
25. Woo, C. M.; Iavarone, A. T.; Spiciarich, D. R.; Palaniappan, K. K.; Bertozzi, C. R., Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat. Methods* **2015,** *12* (6), 561-567.
26. Sun, S. S.; Shah, P.; Eshghi, S. T.; Yang, W. M.; Trikannad, N.; Yang, S.; Chen, L. J.; Aiyetan, P.; Hoti, N.; Zhang, Z.; Chan, D. W.; Zhang, H., Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nat. Biotechnol.* **2016,** *34* (1), 84-88.
27. Liu, M. Q.; Zeng, W. F.; Fang, P.; Cao, W. Q.; Liu, C.; Yan, G. Q.; Zhang, Y.; Peng, C.; Wu, J. Q.; Zhang, X. J.; Tu, H. J.; Chi, H.; Sun, R. X.; Cao, Y.; Dong, M. Q.; Jiang, B. Y.; Huang, J. M.; Shen, H. L.; Wong, C. C. L.; He, S. M.; Yang, P. Y., pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nat Commun* **2017,** *8* (1), 438.
28. Stadlmann, J.; Taubenschmid, J.; Wenzel, D.; Gattinger, A.; Durnberger, G.; Dusberger, F.; Elling, U.; Mach, L.; Mechtler, K.; Penninger, J. M., Comparative glycoproteomics of stem cells identifies new players in ricin toxicity. *Nature* **2017,** *549* (7673), 538-542.
29. Riley, N. M.; Hebert, A. S.; Westphall, M. S.; Coon, J. J., Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nat Commun* **2019,** *10* (1), 1311.
30. Nishikaze, T.; Nakamura, T.; Jinmei, H.; Amano, J., Negative-ion MALDI-MS2 for discrimination of alpha2,3- and alpha2,6-sialylation on glycopeptides labeled with a pyrene derivative. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2011,** *879* (17-18), 1419-1428.
31. de Haan, N.; Reiding, K. R.; Haberger, M.; Reusch, D.; Falck, D.; Wuhrer, M., Linkage-Specific Sialic Acid Derivatization for MALDI-TOF-MS Profiling of IgG Glycopeptides. *Anal. Chem.* **2015,** *87* (16), 8284-8291.

387 32. Yang, S.; Wu, W. W.; Shen, R. F.; Bern, M.; Cipollo, J., Identification of Sialic Acid
388 Linkages on Intact Glycopeptides via Differential Chemical Modification Using IntactGIG-
389 HILIC. *J. Am. Soc. Mass Spectrom.* **2018,** *29* (6), 1273-1283.
390 33. Hinneburg, H.; Hofmann, J.; Struwe, W. B.; Thader, A.; Altmann, F.; Silva, D. V.;
391 Seeberger, P. H.; Pagel, K.; Kolarich, D., Distinguishing N-acetylneuraminic acid linkage
392 isomers on glycopeptides by ion mobility-mass spectrometry. *Chem. Commun.* **2016,** *52* (23),
393 4381-4384.
394 34. Pett, C.; Nasir, W.; Sihlbom, C.; Olsson, B. M.; Caixeta, V.; Schorlemer, M.; Zahedi,
395 R. P.; Larson, G.; Nilsson, J.; Westerlind, U., Effective Assignment of alpha 2,3/alpha 2,6-
396 Sialic Acid Isomers by LC-MS/MS-Based Glycoproteomics. *Angew Chem. Int. Ed. Engl.* **2018,**
397 *57* (30), 9320-9324.
398 35. Khidekel, N.; Arndt, S.; Lamarre-Vincent, N.; Lippert, A.; Poulin-Kerstien, K. G.;
399 Ramakrishnan, B.; Qasba, P. K.; Hsieh-Wilson, L. C., A chemoenzymatic approach toward the
400 rapid and sensitive detection of O-GlcNAc posttranslational modifications. *J. Am. Chem. Soc.*
401 **2003,** *125* (52), 16162-16163.
402 36. Chaubard, J. L.; Krishnamurthy, C.; Yi, W.; Smith, D. F.; Hsieh-Wilson, L. C.,
403 Chemoenzymatic probes for detecting and imaging fucose-alpha(1-2)-galactose glycan
404 biomarkers. *J. Am. Chem. Soc.* **2012,** *134* (10), 4489-4492.
405 37. Mbua, N. E.; Li, X.; Flanagan-Steet, H. R.; Meng, L.; Aoki, K.; Moremen, K. W.;
406 Wolfert, M. A.; Steet, R.; Boons, G. J., Selective exo-enzymatic labeling of N-glycans on the
407 surface of living cells by recombinant ST6Gal I. *Angew Chem. Int. Ed. Engl.* **2013,** *52* (49),
408 13012-13015.
409 38. Li, Q.; Li, Z.; Duan, X.; Yi, W., A tandem enzymatic approach for detecting and
410 imaging tumor-associated Thomsen-Friedenreich antigen disaccharide. *J. Am. Chem. Soc.* **2014,**
411 *136* (36), 12536-12539.
412 39. Zheng, J.; Xiao, H.; Wu, R., Specific Identification of Glycoproteins Bearing the Tn
413 Antigen in Human Cells. *Angew Chem. Int. Ed. Engl.* **2017,** *56* (25), 7107-7111.
414 40. Wen, L.; Liu, D.; Zheng, Y.; Huang, K.; Cao, X.; Song, J.; Wang, P. G., A One-Step
415 Chemoenzymatic Labeling Strategy for Probing Sialylated Thomsen-Friedenreich Antigen. *ACS*
416 *Cent. Sci.* **2018,** *4* (4), 451-457.
417 41. Wu, Z. L. L.; Person, A. D.; Anderson, M.; Burroughs, B.; Tatge, T.; Khatri, K.; Zou,
418 Y. L.; Wang, L. C.; Geders, T.; Zaia, J.; Sackstein, R., Imaging specific cellular glycan
419 structures using glycosyltransferases via click chemistry. *Glycobiology* **2018,** *28* (2), 69-79.
420 42. Wen, L. Q.; Zheng, Y.; Jiang, K.; Zhang, M. Z.; Kondengaden, S. M.; Li, S. S.;
421 Huang, K.; Li, J.; Song, J.; Wang, P. G., Two-Step Chemoenzymatic Detection of N-
422 Acetylneuraminic Acid-alpha(2-3)-Galactose Glycans. *J. Am. Chem. Soc.* **2016,** *138* (36), 11473-
423 11476.
424 43. Wu, Z.; Liu, Y.; Ma, C.; Li, L.; Bai, J.; Byrd-Leotis, L.; Lasanajak, Y.; Guo, Y.;
425 Wen, L.; Zhu, H.; Song, J.; Li, Y.; Steinhauer, D. A.; Smith, D. F.; Zhao, B.; Chen, X.;
426 Guan, W.; Wang, P. G., Identification of the binding roles of terminal and internal glycan
427 epitopes using enzymatically synthesized N-glycans containing tandem epitopes. *Org. Biomol.*
428 *Chem.* **2016,** *14* (47), 11106-11116.
429 44. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation
430 method for proteome analysis. *Nat. Methods* **2009,** *6* (5), 359-362.

431 45. Parker, B. L.; Thaysen-Andersen, M.; Solis, N.; Scott, N. E.; Larsen, M. R.; Graham,
432 M. E.; Packer, N. H.; Cordwell, S. J., Site-specific glycan-peptide analysis for determination of
433 N-glycoproteome heterogeneity. *J. Proteome. Res.* **2013,** *12* (12), 5791-5800.
434 46. Yuan, Z. F.; Liu, C.; Wang, H. P.; Sun, R. X.; Fu, Y.; Zhang, J. F.; Wang, L. H.; Chi,
435 H.; Li, Y.; Xiu, L. Y.; Wang, W. P.; He, S. M., pParse: a method for accurate determination of
436 monoisotopic peaks in high-resolution mass spectra. *Proteomics* **2012,** *12* (2), 226-235.
437 47. UniProt, C., The Universal Protein Resource (UniProt). *Nucleic Acids Res.* **2007,** *35*
438 (Database issue), D193-197.
439 48. Pioch, M.; Hoffmann, M.; Pralow, A.; Reichl, U.; Rapp, E., glyXtool(MS): An Open-
440 Source Pipeline for Semiautomated Analysis of Glycopeptide Mass Spectrometry Data. *Anal.*
441 *Chem.* **2018,** *90* (20), 11908-11916.
442 49. Zhao, P.; Viner, R.; Teo, C. F.; Boons, G. J.; Horn, D.; Wells, L., Combining high-
443 energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification
444 site assignment. *J. Proteome. Res.* **2011,** *10* (9), 4088-4104.
445 50. Halim, A.; Westerlind, U.; Pett, C.; Schorlemer, M.; Ruetschi, U.; Brinkmalm, G.;
446 Sihlbom, C.; Lengqvist, J.; Larson, G.; Nilsson, J., Assignment of saccharide identities through
447 analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. *J. Proteome.*
448 *Res.* **2014,** *13* (12), 6024-6032.
449 51. Willey, K. P., An elusive role for glycosylation in the structure and function of
450 reproductive hormones. *Hum. Reprod. Update* **1999,** *5* (4), 330-355.
451 52. Hayakawa, S.; Nakai, S., Relationships of Hydrophobicity and Net Charge to the
452 Solubility of Milk and Soy Proteins. *J. Food Sci.* **1985,** *50* (2), 486-491.
453 53. Zhu, H.; Li, X.; Qu, J.; Xiao, C.; Jiang, K.; Gashash, E.; Liu, D.; Song, J.; Cheng, J.;
454 Ma, C.; Wang, P. G., Diethylaminoethyl Sepharose (DEAE-Sepharose) microcolumn for
455 enrichment of glycopeptides. *Anal. Bioanal. Chem.* **2017,** *409* (2), 511-518.
456 54. Jiang, K.; Zhu, H.; Li, L.; Guo, Y.; Gashash, E.; Ma, C.; Sun, X.; Li, J.; Zhang, L.;
457 Wang, P. G., Sialic acid linkage-specific permethylation for improved profiling of protein
458 glycosylation by MALDI-TOF MS. *Anal. Chim. Acta.* **2017,** *981*, 53-61.
459 55. Maurer, M. A.; Meyer, L.; Bianchi, M.; Turner, H. L.; Le, N. P. L.; Steck, M.;
460 Wyrzucki, A.; Orlowski, V.; Ward, A. B.; Crispin, M.; Hangartner, L., Glycosylation of
461 Human IgA Directly Inhibits Influenza A and Other Sialic-Acid-Binding Viruses. *Cell Rep.* **2018,**
462 *23* (1), 90-99.
463

464

465

466

467

468 **Captions for Figures, and Scheme**

469

470 **Scheme 1**. Strategy for distinguishing α2,3/α2,6-sialoglycopeptides by glycosyltransferase
471 labeling assisted mass spectrometry.

472

473 **Figure 1**. Structure of the synthesized α2,3-sialoglycopeptides (23SP) and α2,6-
474 sialoglycopeptides (26SP).

475

476 **Figure 2**. Time optimization for CgtA labeling of 23SP and 26SP.

477

478 **Figure 3**. HCD-MS/MS spectra of 23SP (A) and 26SP (B) after enzymatic labeling.

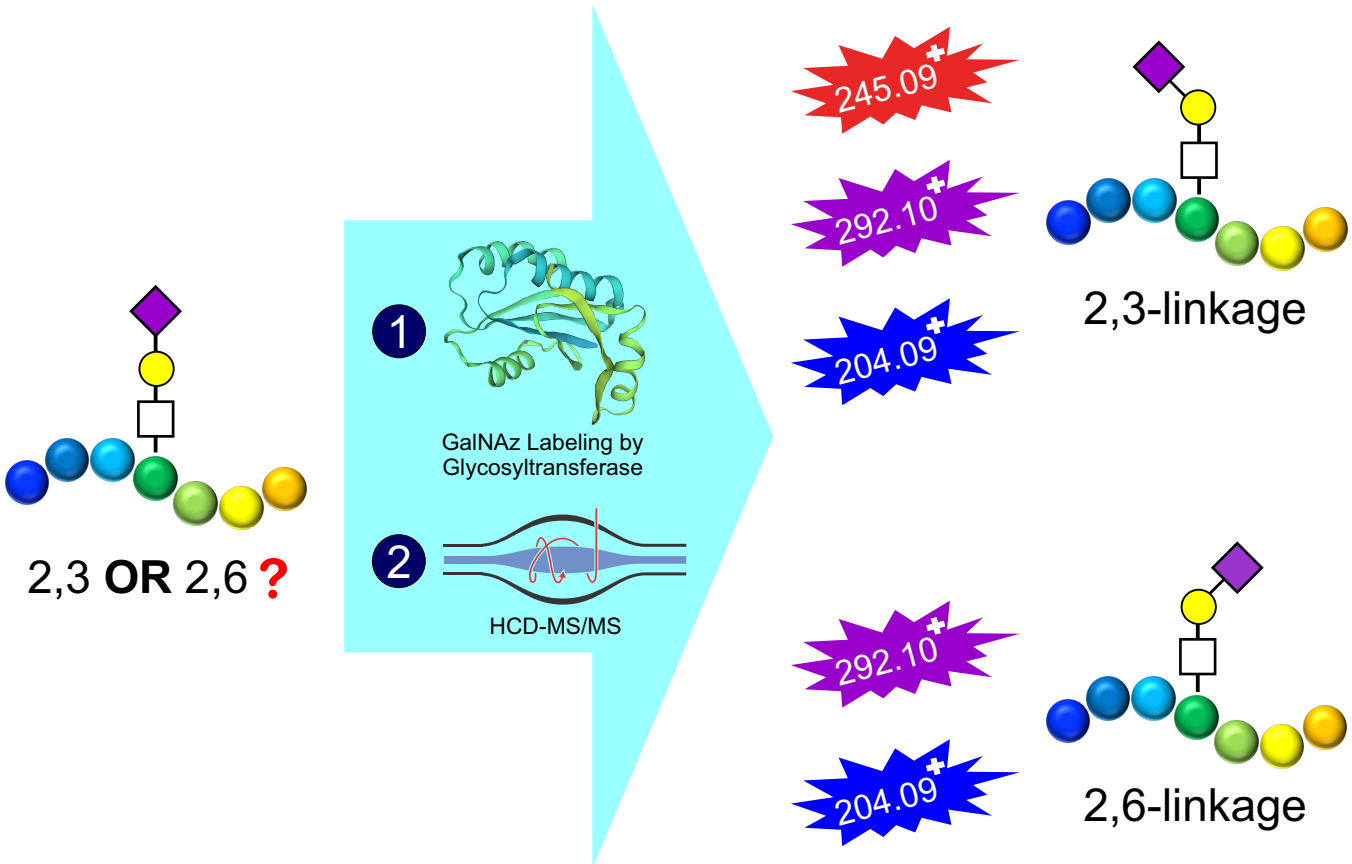
479

480 **Figure 4**. Workflow for identification of linkage specific sialoglycopeptides from human serum.

481

482 **Figure 5**. (A) Distribution of the sialoglycopeptides from human serum; Venn diagrams of
483 identified glycopeptides (B), N-glycosylation sites (C), and glycoproteins (D) in serum samples
484 from colon cancer patients, liver cancer patients and normal people.

Abstract


2,3 OR 2,6 ?
1
GalNAz Labeling by
Glycosyltransferase
2
HCD-MS/MS
245.09+
292.10+
204.09+
2,3-linkage
292.10+
204.09+
2,6-linkage

Scheme 1

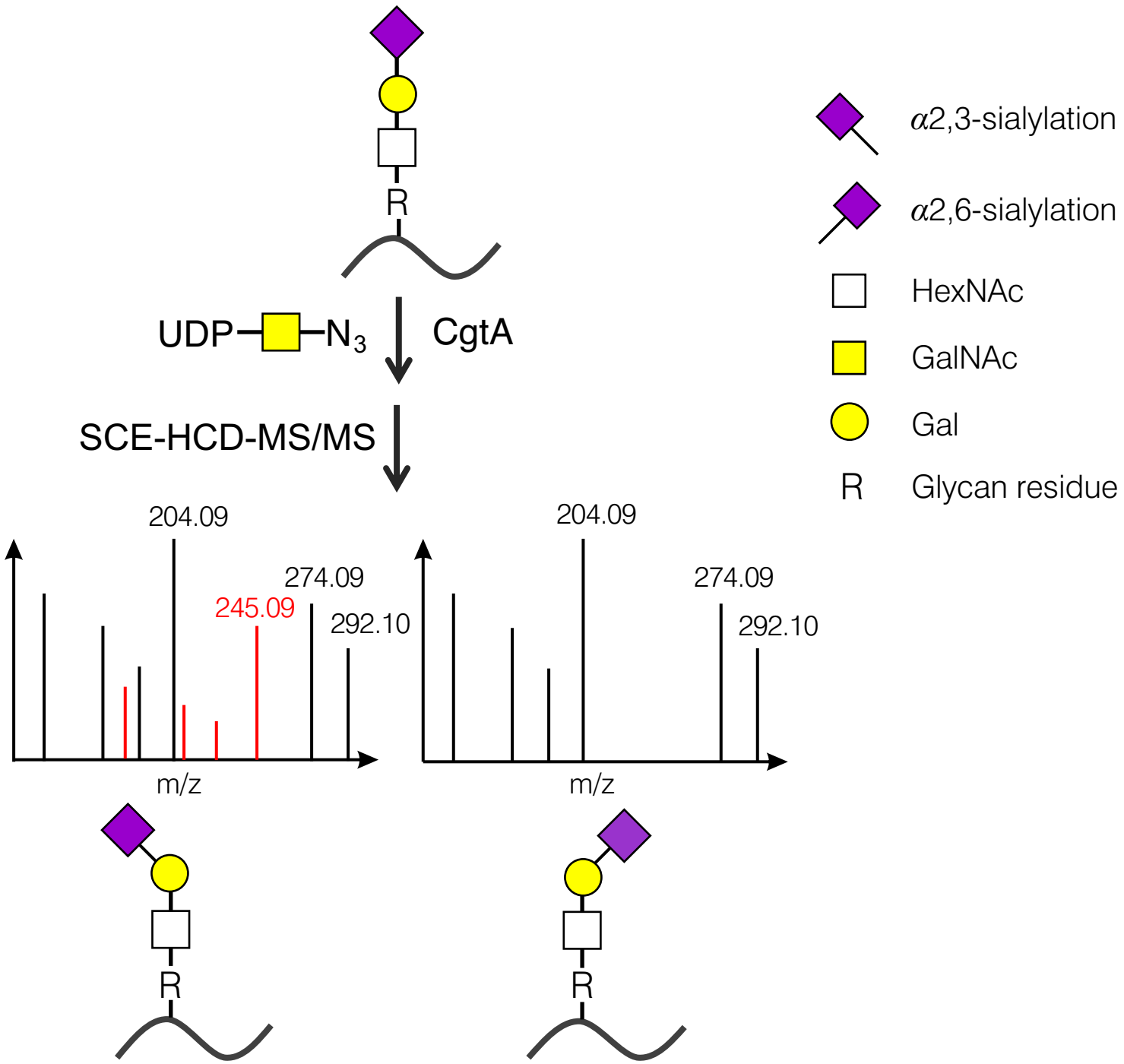

R
UDP
N3
CgtA
SCE-HCD-MS/MS
204.09
274.09
245.09
292.10
m/z
204.09
274.09
292.10
m/z
R
R
α2,3-sialylation
α2,6-sialylation
HexNAc
GalNAc
Gal
R
Glycan residue

Figure 1

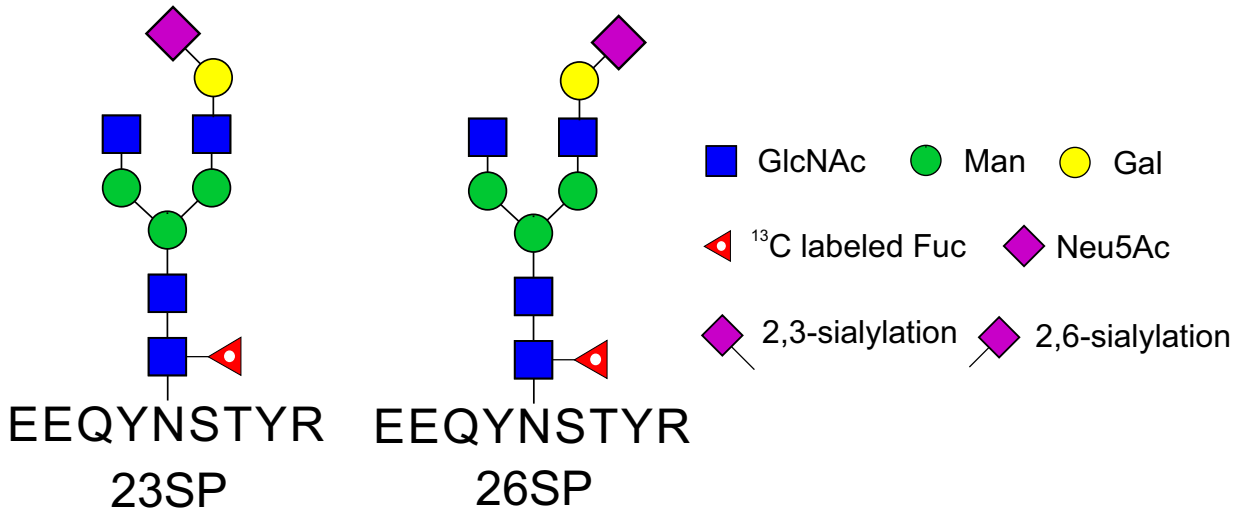

EEQYNSTYR
23SP
EEQYNSTYR
26SP
GlcNAc
Man
Gal
13C labeled Fuc
Neu5Ac
2,3-sialylation
2,6-sialylation

Figure 2

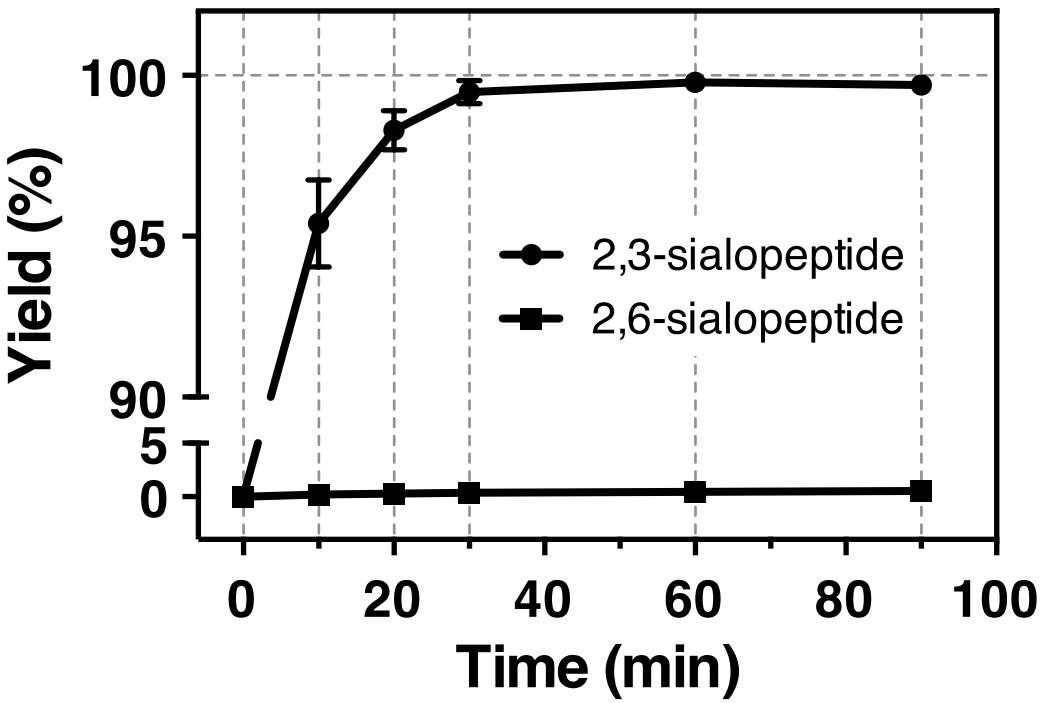
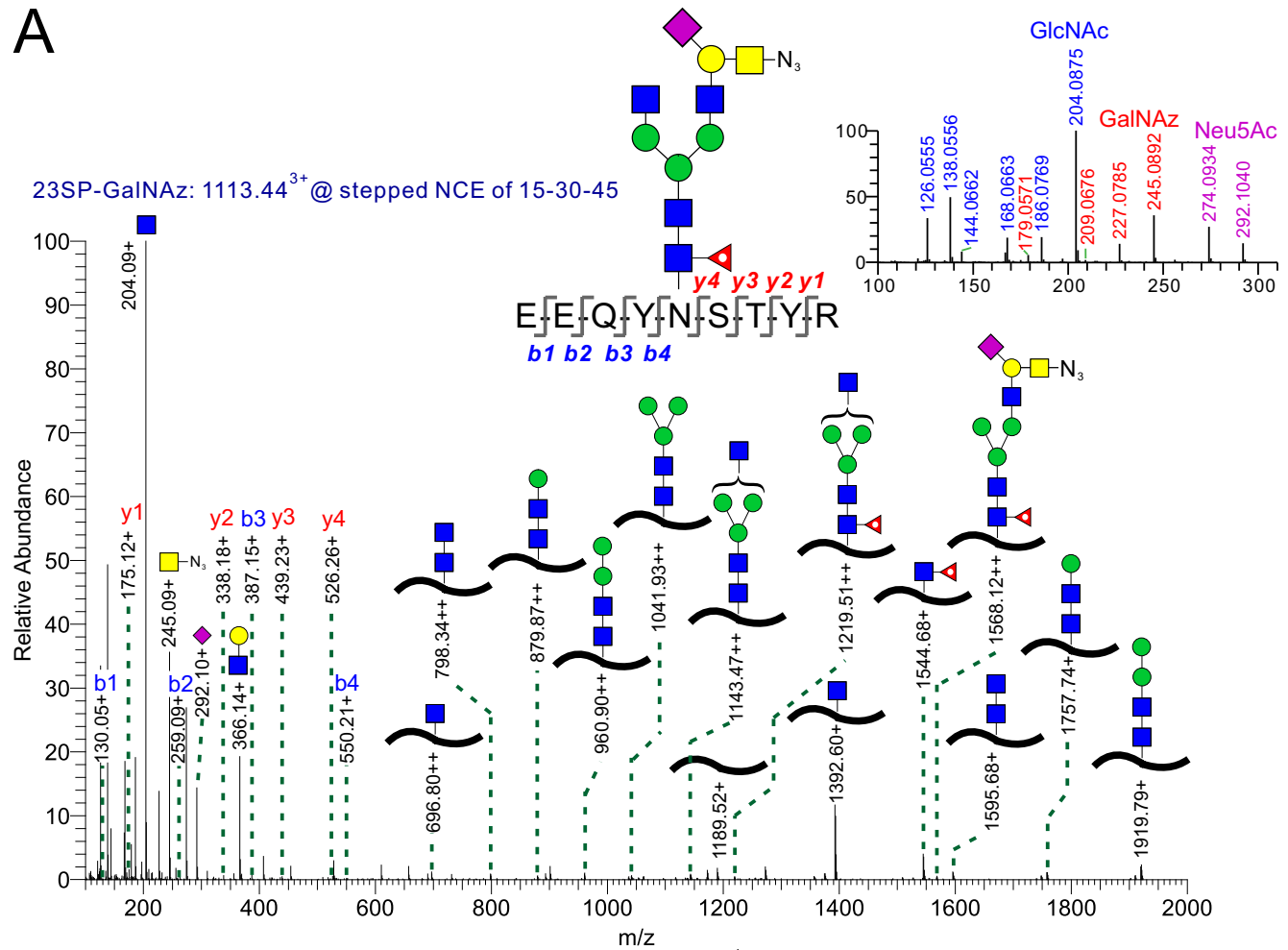

100
95
90
5
0
Yield (%)
2,3-sialopeptide
2,6-sialopeptide
0
20
40
60
80
100
Time (min)

Figure 3


A
23SP-GalNAz: 1113.44 3+ @ stepped NCE of 15-30-45
GlcNAc
GalNAz
Neu5Ac
204.0875
126.0555
138.0556
144.0662
168.0663
179.0571
186.0769
209.0676
227.0785
245.0892
274.0934
292.1040
y4 y3 y2 y1
E-E-Q-Y-N-S-T-Y-R
b1 b2 b3 b4
Relative Abundance
m/z
204.09+
y1
175.12+
245.09+
y2 b3 y3
338.18+
387.15+
439.23+
y4
526.26+
b1
130.05+
b2
259.09+
292.10+
366.14+
b4
550.21+
798.34++
879.87++
960.90++
1041.93++
1143.47++
1219.51++
1544.68+
1568.12++
1757.74+
696.80++
1189.52+
1392.60+
1595.68+
1919.79+

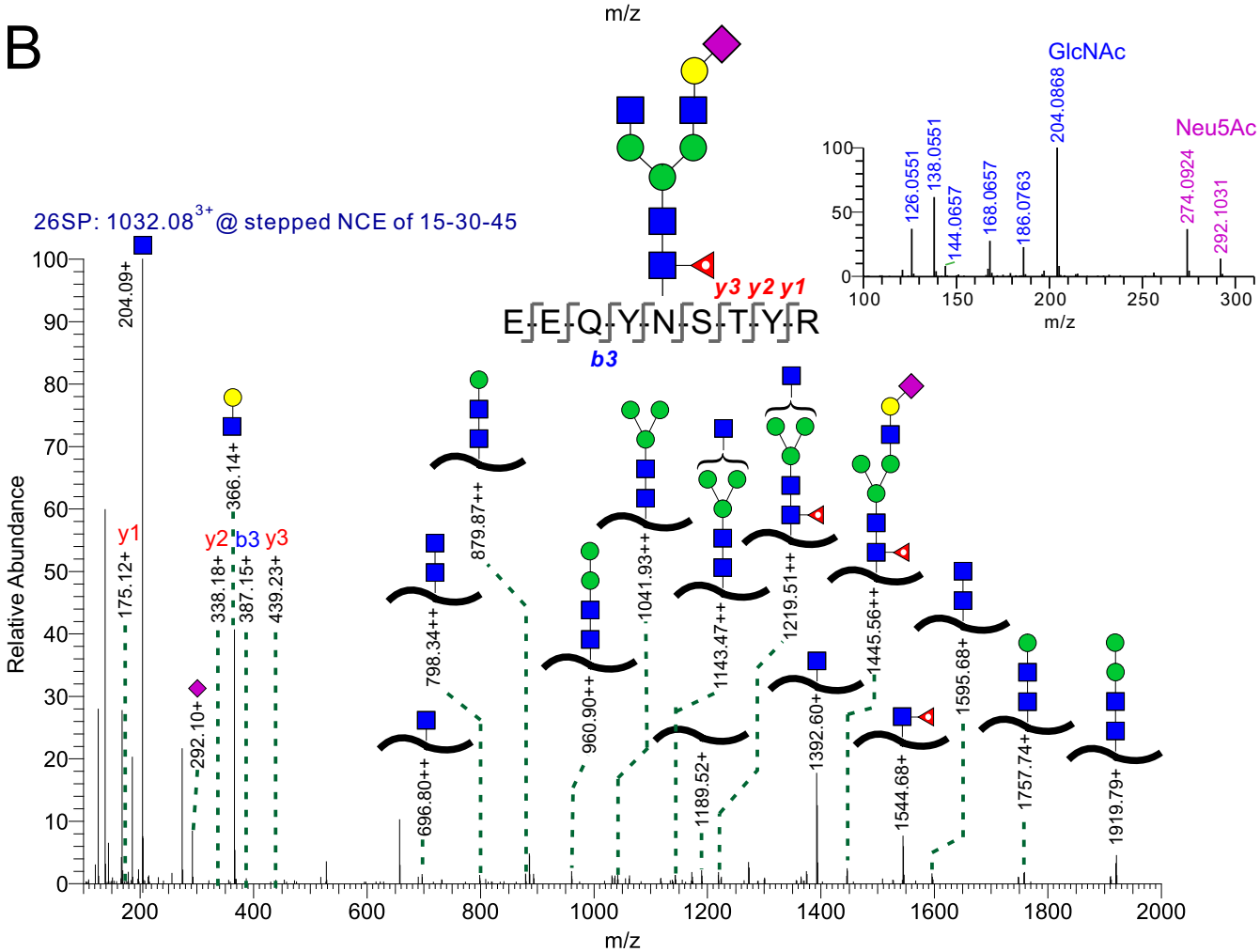

B
26SP: 1032.08 3+ @ stepped NCE of 15-30-45
GlcNAc
Neu5Ac
204.0868
126.0551
138.0551
144.0657
168.0657
186.0763
274.0924
292.1031
y3 y2 y1
E-E-Q-Y-N-S-T-Y-R
b3
Relative Abundance
m/z
204.09+
y1
175.12+
292.10+
366.14+
y2 b3 y3
338.18+
387.15+
439.23+
879.87++
798.34++
960.90++
1041.93++
1143.47++
1219.51++
1445.56++
696.80++
1189.52+
1392.60+
1544.68+
1595.68+
1757.74+
1919.79+

Figure 4

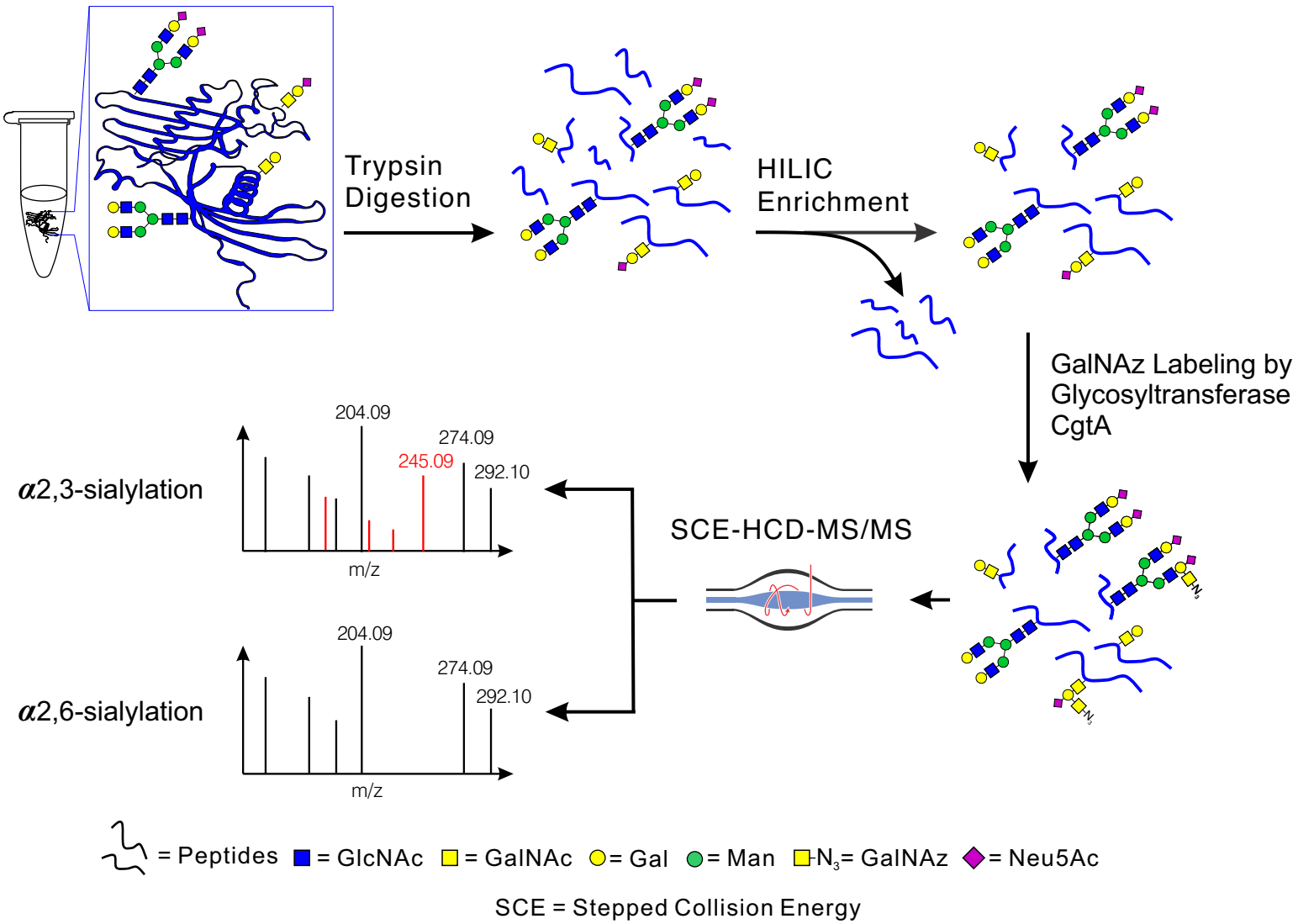
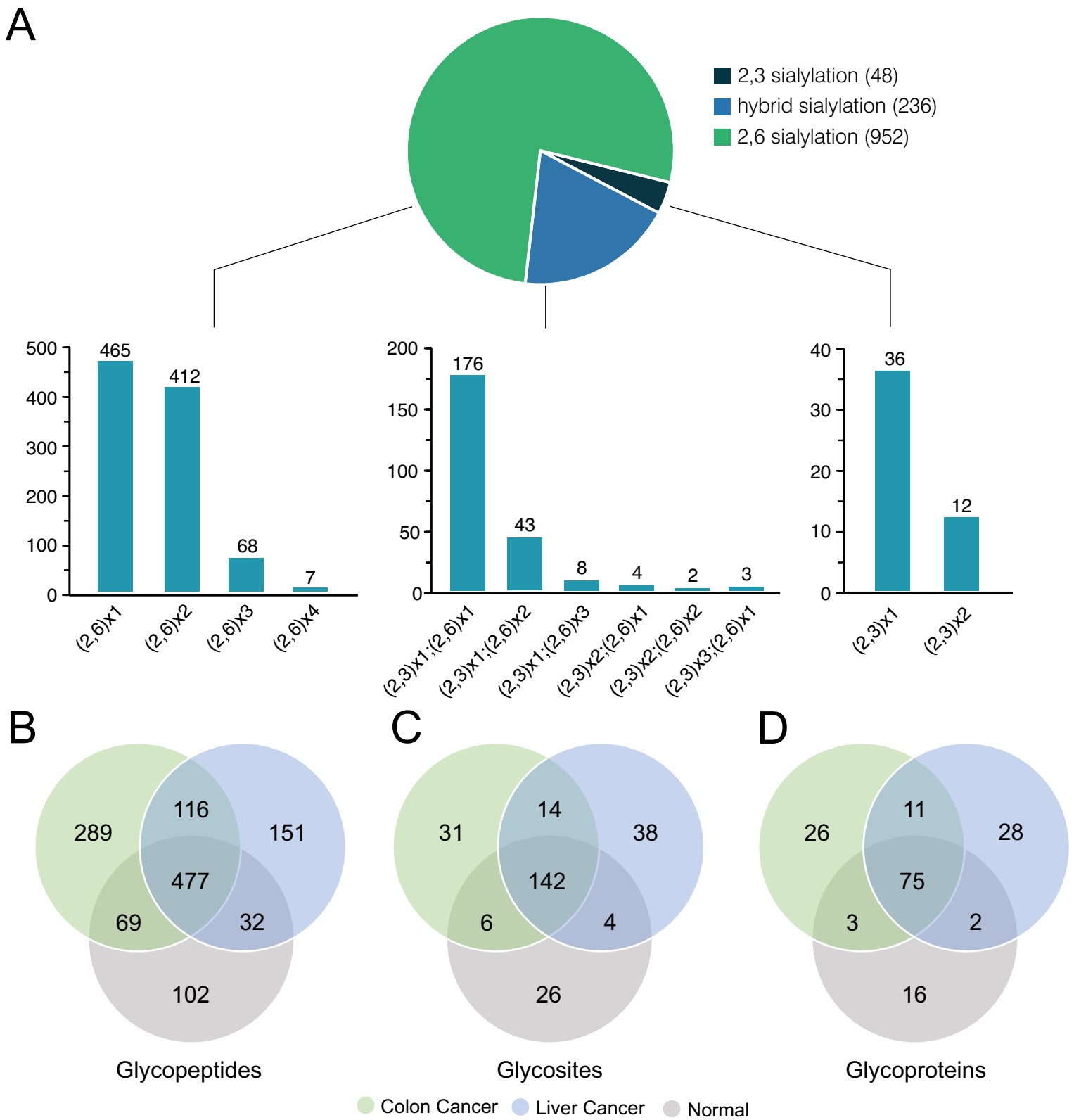

Trypsin
Digestion
HILIC
Enrichment
GalNAz Labeling by
Glycosyltransferase
CgtA
SCE-HCD-MS/MS
α2,3-sialylation
204.09
274.09
245.09
292.10
m/z
α2,6-sialylation
204.09
274.09
292.10
m/z
= Peptides
= GlcNAc
= GalNAc
= Gal
= Man
= GalNAz
= Neu5Ac
SCE = Stepped Collision Energy

Figure 5


A
2,3 sialylation (48)
hybrid sialylation (236)
2,6 sialylation (952)
500
400
300
200
100
0
465
412
68
7
(2,6)x1
(2,6)x2
(2,6)x3
(2,6)x4
200
150
100
50
0
176
43
8
4
2
3
(2,3)x1;(2,6)x1
(2,3)x1;(2,6)x2
(2,3)x1;(2,6)x3
(2,3)x2;(2,6)x1
(2,3)x2;(2,6)x2
(2,3)x3;(2,6)x1
40
30
20
10
0
36
12
(2,3)x1
(2,3)x2
B
289
116
151
477
69
32
102
Glycopeptides
C
31
14
38
142
6
4
26
Glycosites
D
26
11
28
75
3
2
16
Glycoproteins
Colon Cancer
Liver Cancer
Normal