GRable version 1.0: A software tool for site-specific glycoform analysis with improved MS1-based glycopeptide detection with parallel clustering and confidence evaluation with MS2 information

Chiaki Nagai-Okatani,[*,†,‡] *Daisuke Tominaga,*[†,‡,‖] *Azusa Tomioka,*[†,‡] *Hiroaki Sakaue,*[‡] *Norio Goda,*[§,**] *Shigeru Ko,*[§,**] *Atsushi Kuno,*[‡] *Hiroyuki Kaji* [*,‡,¶]

[‡] Molecular and Cellular Glycoproteomics Research Group, Cellular and Molecular Biotechnology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8565, Japan

[§] Department of Systems Medicine, Keio University School of Medicine, Shinjuku, Tokyo 160-8582, Japan

[¶]Institute for Glyco-core Research (iGCORE), Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi 464-8601, Japan

[‖]Present address: Education and Research Center for Pharmacy, Meiji Pharmaceutical University, Kiyose, Tokyo 204-8588, Japan

** Present address: Center for Preventive Medicine, Keio University, Shinjuku, Tokyo 160-8582, Japan

19 *Correspondence

20 E-mail: chiaki-okatani@aist.go.jp (C.N.-O.); kaji.hiroyuki.j8@f.mail.nagoya-u.ac.jp (H.K.)

21

22 **RUNNING TITLE**

23 GRable software for glycoproteomics using Glyco-RIDGE

24

25 This article contains supplemental data.

26

27 **KEYWORDS**

28 glycoproteomics, site-specific glycoform, glycan heterogeneity, retention time, software

29

30 **ABBREVIATIONS**

31 Glyco-RIDGE, Glycan heterogeneity-based Relational IDentification of Glycopeptide signals on Elution
32 profile; hAGP, serum α1-acid glycoprotein; HILIC, hydrophilic interaction chromatography; HPT,
33 haptoglobin; HL-60, human promyelocytic leukemia-60; IGOT, isotope-coded glycosylation site-specific
34 tagging; LC/MS, liquid chromatography/mass spectrometry; MALDI-TOF, matrix-assisted laser
35 desorption/ionization time-of-flight; mgf, Mascot generic format; MS, mass spectrometry; PNGase F,
36 peptide-N-glycosidase F; rCM, reduction and carbamidomethylation; RT, retention time; TFA,
37 trifluoroacetic acid

38

ABSTRACT

High-throughput intact glycopeptide analysis is crucial for elucidating the physiological and pathological status of the glycans attached to each glycoprotein. Mass spectrometry-based glycoproteomic methods are challenging because of the diversity and heterogeneity of glycan structures. Therefore, we have developed an MS1-based site-specific glycoform analysis method named "Glycan heterogeneity-based Relational IDentification of Glycopeptide signals on Elution profile (Glyco-RIDGE)" for a more comprehensive analysis. This method detects glycopeptide signals as a cluster based on the mass and chromatographic properties of glycopeptides and then searches for each combination of core peptides and glycan compositions by matching their mass and retention time differences. Here we developed a novel browser-based software named GRable for semi-automated Glyco-RIDGE analysis with significant improvements in glycopeptide detection algorithms, including "parallel clustering." This unique function improved the comprehensiveness of glycopeptide detection and allowed the analysis to focus on specific glycan structures, such as pauci-mannose. The other notable improvement is evaluating the "confidence level" of the GRable results, especially using MS2 information. This function facilitated reduced misassignment of the core peptide and glycan composition and improved the interpretation of the results. Additional improved points are: "correction function" for accurate monoisotopic peak picking; one-to-one correspondence of clusters and core peptides even for multiply sialylated glycopeptides; and "inter-cluster analysis" function for understanding the reason for detected but unmatched clusters. The significance of these improvements was demonstrated using purified and crude glycoprotein samples, showing that GRable allowed site-specific glycoform analysis of intact sialylated glycoproteins on a large scale and in depth. Therefore, this software will help to analyze the status and changes in glycans to obtain biological and clinical insights into protein

60 glycosylation by complementing the comprehensiveness of MS2-based glycoproteomics. GRable can run

61 freely online using a web browser via the GlyCosmos Portal (https://glycosmos.org/grable).

62 **INTRODUCTION**

63 Protein glycosylation is a common and complex post-translational modification in eukaryotes (1) that plays
64 pivotal roles in various biological processes (2). This post-translational modification regulates the
65 glycoprotein function and localization by modulating their tertiary structures and interactions with other
66 molecules (2). However, revealing the glycan structure-function relationship is challenging because glycan
67 structures are highly diverse and heterogeneous and differ in organisms, cells, proteins, and the attached sites
68 (3). Furthermore, the post-translational modification alters the cell state change through differentiation,
69 carcinogenesis, infection, nutrition, medicine, and stimuli (3). Glycans are naturally heterogeneous, and
70 elucidating their function of glycans by uncovering the state and changes in glycans with heterogeneities at
71 each level is important. Heterogeneous levels include the micro (glycan variation of one glycosite), macro
72 (with or without glycosylation of one glycosite), and meta (glycan variation of the entire glycoprotein
73 molecule) levels, which are important features responsible for the presence of different proteoforms (4).

74 Although current improvements in mass spectrometry (MS) have made it possible to perform an
75 in-depth structural analysis of glycopeptide peptide backbones and glycomes, elucidating the glycan
76 compositions of specific sites for proteins with multiple glycosylated sites (i.e., micro-heterogeneity) is
77 challenging (5). First, the glycan compositions of each site must be determined to estimate the site-specific
78 glycoforms. For this purpose, analysis of only glycomes and deglycosylated peptides is insufficient; thus,
79 intact glycopeptides must be analyzed directly. Current standard glycoproteomic approaches rely on MS2
80 spectra to identify site-specific glycoforms (6). For MS2-based glycoproteomics, the commercial software
81 Byonic (7) is often used (6) and many in-house software programs have been developed (8–20). However,
82 these MS2-based methods have a critical problem in that they require glycopeptide fragmentation, leading
83 to lower detection sensitivity than the corresponding naked peptides. The decreased detection sensitivity

84 results from the lower ionization efficiency of glycopeptides, which is notable, especially for sialylated
85 glycopeptides because of their negative properties. In addition, glycan heterogeneity decreases the
86 abundance of each glycopeptide. Therefore, in some MS1 spectra, many signals are not selected for MS2
87 analysis, and frequently, MS2 spectra do not contain sufficient information for confident glycopeptide
88 identification.

89 To overcome this limitation, we developed an MS1-based glycoproteomic approach named
90 "Glycan heterogeneity-based Relational IDentification of Glycopeptide signals on Elution profile" (Glyco-
91 RIDGE) (21), which was first introduced in 2015 (22). One limitation of MS1-based approaches is to
92 "estimate" combinations of core peptides and glycans, having a possibility of misassignment because of
93 isobaric combination issues caused by adducts and core peptide modifications observed even in MS2-based
94 approaches (23, 24). To minimize such errors, a general workflow of the Glyco-RIDGE analysis using
95 liquid chromatography/mass spectrometry (LC/MS) data of glycopeptide samples requires information on
96 core peptides present in the analyte, as well as presumed glycans (Figure 1); these sequences and *N*-
97 glycosites can be identified by isotope-coded glycosylation site-specific tagging (IGOT)-LC/MS/MS (25,
98 26), in which N-glycosites are specifically labeled by peptide-*N*-glycosidase F (PNGase F)-mediated
99 incorporation of a stable isotope tag (^{18}O). In the Glyco-RIDGE workflow, MS1 signals derived from
100 glycopeptides with the same core peptide are assigned as a cluster in the LC/MS data based on their mass
101 differences and closeness of retention times (RTs). Then, the combination of the core peptide and glycan
102 composition for these glycopeptides is annotated by matching the observed accurate glycopeptide mass
103 with the sum of the masses of the pre-identified core peptide candidates and presumed glycans. A notable
104 feature of Glyco-RIDGE is the detection of glycopeptide signals based only on MS1 data, in contrast to
105 other MS1-based approaches that employ MS2 data to identify at least one glycopeptide for each core

106 peptide (27–31). With this feature, the Glyco-RIDGE method is expected to achieve a more high-sensitive
107 and comprehensive analysis of protein glycosylation than current fragmentation-dependent identification
108 methods. Notably, a previous study (32) has demonstrated that among the MS2 spectra assigned using the
109 Glyco-RIDGE method and Byonic search, the ratio of the same assignment was quite high (349/353 =
110 98.9%), demonstrating that the reliability of this method is similar to that of Byonic.

111 The feasibility and usefulness of the Glyco-RIDGE method have been demonstrated using in-house
112 software by applying it to the analysis of specific glycan motifs (i.e., Lewis X and polylactosamine)-
113 containing glycopeptides in complex glycoprotein mixtures after the removal of sialic acids (22, 32). Then,
114 this method has been used also for site-specific glycoform analysis of intact (i.e., sialylated) glycopeptides
115 (33–36), revealing site-specific "glycostem" and "glycoleaf" features (36). However, the prototype software
116 has several limitations. One limitation was the misfinding of monoisotopic signals from MS1 data, causing
117 a decreased number of detected glycopeptide signals. Regarding clustering, glycopeptides with the same
118 core peptide and different numbers of sialic acid residues were detected as discrete clusters, impairing the
119 detection and interpretation of microheterogeneity. Another issue was the presence of glycopeptide clusters
120 that were detected but not assigned, which might result from adducts or peptide modifications; however, the
121 reason is unclear and thereby remains to be improved. In addition, selecting the most plausible match for
122 each glycopeptide cluster as the final assignment and evaluating the assignments for each glycopeptide were
123 performed manually by searching the MS2 spectra, which is time-consuming. Accordingly, it is also
124 difficult to optimize search parameter settings and acceptance criteria while balancing the coverage and
125 accuracy of the analysis, which is an important consideration also in MS2-based approaches (37).

126 The present study introduces a novel software named "GRable" to semi-automatically execute the
127 Glyco-RIDGE analysis with some technical improvements for the above issues. The notable features are
128 "parallel clustering" for efficient and comprehensive finding of glycopeptide clusters and adding a Selection
129 step for selecting and evaluating assignments using MS2 information as a "confidence level." Additional
130 improved points are: "correction function" for accurate monoisotopic peak picking; one-to-one
131 correspondence of clusters and core peptides even for multiply sialylated glycopeptides; and "inter-cluster
132 analysis" function for understanding the reason for detected but unmatched clusters; To demonstrate the
133 feasibility and utility of this software, we analyzed MS data obtained with intact glycopeptides prepared
134 from human serum α1-acid glycoprotein (hAGP), a glyco-biomarker candidate for liver fibrosis (38), as a
135 model of targeted analysis. In addition, we used data on glycopeptides obtained from human promyelocytic
136 leukemia-60 (HL-60) cell lysates (32), as a crude sample model.
137

138 **EXPERIMENTAL PROCEDURES**

139 GRable version 1.0 requires the following four files: LC/MS data of glycopeptides after
140 deconvolution (.mzML); a core peptide list (.xlsx); a glycan point list (.xlsx); and LC/MS/MS data for
141 glycopeptides (.mgf) (Figure 2). Details of the input files and their preparation are described below and in
142 the user manual (Document S1). Note that this version of GRable is guaranteed only for Thermo Fisher
143 Scientific data.

144

145 **Preparation of hAGP glycopeptide samples**

146 To prepare large-scale analytes for repeated verification, 5 mg of freeze-dried hAGP (Sigma-
147 Aldrich, St.Louis, MO, USA) was dissolved in Milli-Q water (500 μL) and the aliquot was denatured with
148 0.1% Rapigest (Waters, Milford, MA, USA) in 50 mM Tris-HCl, reduced with dithiothreitol (Fujifilm
149 Wako, Osaka, Japan), alkylated with iodoacetamide (Fujifilm Wako), and digested with Lys-C (1/200
150 weight of protein; Fujifilm Wako) and sequencing grade modified trypsin (1/100 weight of protein;
151 Promega, Madison, WI, USA) at 37 °C overnight in 0.1 % Rapigest. After digestion, Rapigest was cleaved
152 by acidification with 0.1% trifluoroacetic acid (TFA; Fujifilm Wako). Glycopeptides from hAGP tryptic
153 digests were then captured using an Amide-80 column (TSK-gel Amide-80, 2 × 50 mm; TOSOH, Tokyo,
154 Japan) equilibrated with 75% acetonitrile (Merck, Darmstadt, Germany)/0.1% TFA, and eluted
155 isocratically with 50 % acetonitrile/0.1% TFA after washing.

156

157 **LC/MS/MS analysis of hAGP glycopeptide samples**

158 Glycopeptides were analyzed using an LC-electrospray ionization-MS system equipped with a
159 nanoflow LC system (UltiMate-3000; Thermo Fisher Scientific, Waltham, MA, USA) and a tandem mass

160 spectrometer (Orbitrap Fusion Tribrid mass spectrometer; Thermo Fisher Scientific). Glycopeptides were
161 trapped on a C18 cartridge column (Acclaim PepMap100 C18, 0.3 mm I.D. × 5 mm; Thermo Fisher
162 Scientific) and separated on a C18 tip column (75 μm I.D. × 15 cm, 3 μm particle; Nikkyo Technos, Tokyo,
163 Japan) using 2–36% acetonitrile/0.1% formic acid (Fujifilm Wako) gradient (90 min) at a flow rate of 300
164 nL/min. The ionization voltage was 2.0 kV (positive), and the temperature of the ion transfer tube was
165 275 °C. Data were acquired using the data-dependent mode (cycle time: 2.5 s) with internal mass calibration
166 (lock mass) at 445.12003. The mass range was 377–2,000 m/z (MS1) and 135–2,000 m/z (MS2). The mass
167 resolutions were 120,000 (MS1, Orbitrap) and 15,000 (MS2, Orbitrap). The fragmentation mode was high-
168 energy collision-induced dissociation with stepped collision energies of 25, 30, and 35%. The acquired raw
169 data file was deconvoluted using the Xtract of Proteome Discoverer (ver. 2.4, Thermo Fisher Scientific),
170 and exported as an mzML file for GRable analysis.

171

172 **IGOT-LC/MS/MS analysis of hAGP glycopeptide samples**

173 IGOT-LC/MS analysis was performed for core peptide identification as reported previously (25,
174 26). Glycopeptides were treated with PNGase F (Takara Bio, Shiga, Japan) in 50 mM Tris-HCl, pH 8.5,
175 prepared with $H_2{}^{18}O$ (isotope purity: 95%; Taiyo Nippon Sanso, Tokyo, Japan) at 37 °C overnight.
176 Deglycosylated peptides (IGOT peptides) were analyzed before glycopeptide analysis using LC/MS/MS
177 under the same conditions as for the glycopeptide sample.

178

179 **Preparation of a core peptide list**

180 The format of the core peptide list required by the software was fixed and created in a manner
181 similar to that used in this study (Data S1). The details of the preparation are documented in the manual

182 (Document S1, Section 4). Items minimally required for preparing a core peptide list are the calculated mass,
183 retention time of each peptide and at least one peptide identifier such as CP (core peptide) number. These
184 information are used to detect the combinations of glycopeptides and core peptides by calculating the mass
185 differences within the limited RT range in the Matching step. For the identification of core peptides, IGOT-
186 LC/MS analysis, as described above, is recommended to confirm their *N*-glycosylations, but it is possible
187 to list the peptides identified by proteomic analysis. Confirmation of the presence of core peptides for
188 glycoproteins is crucial for crude samples. For the analysis of a highly purified glycoprotein, it is also possible
189 to list the peptides predicted based on protease digestion patterns.

190 For hAGP, the raw data acquired by IGOT-LC/MS/MS were converted to mgf using Mascot
191 Distiller (ver. 2.7.0; Matrix Science, Boston, MA, USA), and searched using Mascot (ver. 2.5.1; Matrix
192 Science) using a human protein sequence file (SwissProt_UniProtKB_isoform; downloaded in April 2019;
193 entry: 42,431). The search parameter settings were as follows: enzyme: trypsin (full or semi); missed
194 cleavage: 2; mass tolerance: 7 ppm (MS1) and 0.02 Da (MS2); fixed modification: carbamidomethyl (C);
195 variable modifications: Gln to pyro-Glu (peptide N-term, Q), ammonia-loss (peptide N-term,
196 carbamidomethyl C), oxidation (M), and Delta: H(-1)N(-1)18O(1)(N); and target false discovery rate: 1%.
197 Mascot search results with a peptide rank of 1 and peptide expectation value of <0.05 were selected.
198 Matched sequences containing IGOT modifications at Asn on the consensus sequence for *N*-glycosylation
199 (Asn-Xaa-[Ser/Thr]; Xaa is not Pro) obtained using trypsin (full) and trypsin (semi) were combined. The
200 resulting glycopeptide list was used to create a core peptide list for GRable analysis. To ensure the
201 comprehensive analysis, hAGP peptide sequences that were not identified in the data used here but found
202 in other analyses were manually added to the list with observed or predicted RTs (Data S1). That is, five

203 core peptides with predicted hAGP sequences were added to the list, and three core peptides were matched
204 with glycopeptides as clusters.

205 It is noteworthy that considering oxidized Met in the deglycopeptide search for improving
206 coverage may cause incidental mismatching for one glycopeptide; for example, oxidation of the peptide and
207 HexNAc and Hex and carbamidomethylation of the peptide (23). However, in the Glyco-RIDGE method,
208 such a misassignment is removed at the selection step because the core peptide of each glycopeptide is
209 evaluated as a glycopeptide cluster. Similarly, carbamidomethylation of the Met within a core peptide,
210 which was not considered in the peptide list preparation and has the possibility of mismatching (23), should
211 be omitted from the selection results. It should also be noted that the Mascot search for *O*-glycosylation of
212 IGOT-treated *N*-deglycopeptides may be useful for evaluating the possibility of misinterpretation of the
213 GRable results of site-specific N-glycoform analysis because this software is intended to assign
214 glycopeptides with single glycosite.

215

216 **Preparation of a glycan composition list**

217 The format of the core peptide list required by the software was fixed and created in a manner
218 similar to that used in this study (Data S2). The details of the preparation are documented in the manual
219 (Document S1). In the Matching setting, users can define the permissible glycan compositions (i.e., the type
220 and number of each glycan unit). The default settings were as follows: Hex: 0–12, HexNAc: 1–12, dHex:
221 0–4, and NeuAc: 0–4. The maximum numbers of Hex and HexNAc were determined using the
222 approximate masses of the ionizable glycopeptides. The maximum number of NeuAc was limited to four
223 because our sample preparation conditions did not allow the remaining oligo- or polysialic acids. The
224 number of dHex (fucose) was also limited to four or fewer because the mass of dHex(5) is similar to that of

225 Hex(2)HexNAc(2) (difference = 0.025), corresponding to approximately 5 ppm of 5,000 Da. If the
226 measurement is performed with an accuracy greater than 2 ppm, five or more fucoses may be
227 distinguishable. Within the range of glycan compositions, users can provide arbitrary points for likely
228 compositions in the glycan list. In contrast, among the possible compositions, many are considered unlikely
229 based on biosynthetic pathways, and matches with such unusual compositions can be considered incorrect.
230 Accordingly, users can also define unusual compositions in the glycan list and such incorrect matches are
231 marked in the match results by giving negative points, and thus, will be considered in the selection.

232 The format of the glycan point list was fixed (Data S2). The list used in this study can be used to
233 analyze human cell-derived protein samples and was not considered for NeuGc. In this list, 139 glycan
234 compositions were assigned 1 point if the composition is matched. The list contained glycan compositions
235 in *N*-glycan major biosynthetic pathways, from the glycan produced by oligosaccharide transferase to the
236 glycan processed by many glycosyltransferases and glycosidases. The largest is
237 Hex(7)HexNAc(6)dHex(4)NeuAc(4), which corresponds to a tetra-antennary glycan with four Fuc and
238 four NeuAc. The smallest is HexNAc(1), a breakdown product remaining in GlcNAc on Asn. Unusual
239 compositions were determined by considering *N*-glycan biosynthetic pathways.

240 This study considered four glycan units (Hex, HexNAc, dHex, and NeuAc) as the glycan
241 components attached to hAGP. *N*-glycan compositions were expressed as the generic monosaccharide
242 composition as "Hex(*)HexNAc(*)dHex(*)NeuAc(*)", where * represents the number of individual
243 monosaccharide residues. The glycan composition list for hAGP was almost identical to the human *N*-
244 glycome list supplied by Byonic, except for one composition (Hex(7)HexNAc(6)) (Table S1), which was
245 manually added because this composition is considered common in the human *N*-glycome. In the glycan

246 composition list, unusual compositions were defined based on *N*-glycan biosynthetic pathways; however,
247 negative points were not given for these unusual compositions in this study.

248

249 **Byonic search for hAGP data**

250 Glycopeptide MS2 data were searched using the Byonic search engine ver. 2.15.7 (Protein
251 Metrics, Cupertino, CA, USA) using the human protein sequence file from SwissProt (downloaded on May
252 19, 2020; entry: 42,296) and the glycan composition list (Table S1). Search parameter settings were as
253 follows: enzyme: Trypsin_KR (Full or Semi), max missed cleavages: 2, static modifications:
254 carbamidomethyl (C), dynamic modifications: ammonia-loss (N-term C), Gln to pyro-Glu (N-term Q),
255 oxidation (M), peptide mass tolerance: ± 3 ppm, fragment mass tolerance: 0.02 Da. The search results used
256 estimated glycopeptides with Confidence = High and Byonic Score ≥ 200 to compare GRable results. The
257 two results obtained with Trypsin_KR (full) and Trypsin_KR (semi) were combined, and the resultant
258 glycopeptide list was compared with the GRable results.

259

260 **GRable analysis of HL-60 cell data**

261 LC/MS/MS data for HL-60 cell lysates were obtained from a previous study (Figure S1) (32). A
262 protease digest of the HL-60 cell lysate was prepared in a manner similar to that for hAGP. An aliquot of
263 the digest was acidified and heated to remove sialic acid, decreasing glycan heterogeneity and increasing the
264 relative abundance of each glycopeptide. An aliquot of the desialylated glycopeptide fraction was subjected
265 to HILIC to collect four glycopeptide fractions, which were then analyzed using LC/MS/MS and IGOT-
266 LC/MS/MS to obtain glycopeptide and deglycopeptide MS data, respectively. The glycans obtained from
267 the IGOT procedure were analyzed using matrix-assisted laser desorption/ionization time-of-flight

268 (MALDI-TOF) MS to prepare a glycan composition list. Because the HL-60 cell data were for desialylated
269 glycopeptides, only three glycan units (Hex, HexNAc, and dHex) were considered glycan components.
270 The definition of unusual composition was the same as that used for the hAGP. The resulting core peptide
271 list (Data S3) and glycan point list (Data S4) for fraction 2 were used for the GRable analysis.

272

273 **Software implementation**

274 GRable was implemented as a web application by Hitachi Solutions Technology Ltd. The user
275 interface was mostly written in JavaScript, data management in Java, and scientific calculations in Python.
276 PostgreSQL was used as the background for the user and data file management. The software was
277 developed and tested on Ubuntu Linux 22.04 LTS.

278

RESULTS AND DISCUSSION

GRable is a web application that users can access through web browsers. It should be noted that the deconvoluted LC/MS data must be uploaded in the current version. For data processing (Figure 2), all steps were executed step-by-step after uploading the required data and setting the appropriate parameters. A detailed instruction manual for GRable is provided in Document S1. The data processing, search parameter settings, and results of each step are described below, primarily focusing on improved points compared with the prototype.

Monoisotopic peak picking (Step 4) with improved accuracy using the correction function

After the ranges of RT and mass were set in the Range setting (Step 3), monoisotopic peaks were selected using the original algorithm in Step 4. In a previous in-house prototype, the search for monoisotopic peaks started from the highest signal, which is time-consuming. Therefore, GRable uses a new algorithm to improve data processing speed. First, a filter of five scans × 5 Da was used to search for a local peak in each group comprising a single ion. When the highest signal was centered in the filter, it was recorded as a local peak, and the other signals within the filter were excluded from the local peak candidates. The local peaks were used as the starting points for identifying the corresponding groups comprising the same ion. The local peak signal spectrum was integrated with the same ion spectra before and after the scans, and the resulting spectrum was used to determine the monoisotopic signal. This algorithm change greatly speeds up the monoisotopic peak-picking process.

In addition, the algorithms for finding monoisotopic signals were also modified to improve the accuracy of the monoisotopic assignment in GRable. The prototype used a monoisotopic signal assignment method with criteria based on the relative intensity of the isotope signals; however, the misdetection of

301 monoisotopic ions was frequently observed. The accuracy of the monoisotopic ions is important for
302 accurate glycopeptide assignment. Accordingly, GRable has a function that evaluates the selected
303 monoisotopic peaks using a multinomial distribution based on averagine for a peptide as the abundance ratio
304 of five isotopes, including C, H, N, O, and S. By fitting the observed spectrum with the calculated spectrum,
305 the need to correct the monoisotopic signal obtained in the preceding step is suggested. This correction
306 function facilitated an increased number of matched glycopeptide group members compared to the GRable
307 results obtained without using this function. Users can confirm corrected monoisotopic peaks by the update
308 flag=1 in a "Monoiso Peak List" sheet and in other results sheets of subsequent steps, where the peak no.
309 was identical to that indicated in the monoisotopic peak list. For hAGP, 5,369 monoisotopic signals were
310 detected, of which 2,140 were automatically corrected by the correction function (Data S6), demonstrating
311 its utility for the accurate selection of monoisotopic signals.

312

313 **Clustering (Step 5) for efficient and comprehensive detection of glycopeptide signals using parallel**
314 **clustering function**

315 In this step, glycopeptide signals were detected from the monoisotopic peaks selected in Step 4
316 as a cluster, based on the relationship between the masses and retention times. The prototype could set a
317 single RT range to find the glycopeptide group as a cluster because the time difference was considered for
318 extending the neutral monosaccharides, namely Hex, HexNAc, and dHex. Glycopeptide groups with
319 different numbers of acidic saccharides, such as NeuAc, were found as clusters (33), because adding neutral
320 saccharides slightly shortens the RT, whereas adding acidic sugar units increases it (39). In the previous RT
321 setting, as the acidic shift increased in proportion to the number of sialic acid units, four glycopeptide clusters
322 with asialo-, monosialo-, disialo-, and trisialo-glycans were found separately in a representative hAGP

323 cluster (Figure 3A). Therefore, glycopeptide groups with the same core but different numbers of acidic units
324 cannot be combined. To improve this, GRable was constructed to allow setting the RT difference separately
325 for each acidic and neutral saccharide and any motif, such as Hex + HexNAc (LacNAc). In the current
326 version, four clusters with the same core peptide were successfully connected into a single cluster, because
327 the range of RT shifts for each unit could be set separately (Figure 3B). This improvement facilitated a better
328 understanding of the microheterogeneity of a glycosite by visualizing the signal intensity of each
329 glycopeptide plot. Notably, a new tetrasialo-glycan was also detected. Such highly sialylated members
330 could not be detected as a cluster because of the limited variety of the glycan stem portion with the four
331 sialylations. Accordingly, this improvement facilitated the increased detection of glycan microheterogeneity
332 for one glycosite. In addition, considering the RT difference separately for each glycan unit is also useful for
333 reducing misdetection caused by isobaric combinations, especially neutral and acidic combinations, such as
334 NeuAc + $NH4^+$ and Hex + dHex (24).

335 As is the case with MS2-based approaches, optimization of search parameter settings is crucial
336 for obtaining reliable results in glycopeptide identification (40); thus, optimization is needed for the purpose
337 of analysis. For this need, GRable can integrate multiple clustering results obtained in parallel under different
338 (up to five) search parameter settings for more comprehensive detection of glycopeptide clusters. As the
339 clustering results of each processing step were combined into a single cluster, each search parameter setting
340 was independent of the processing order. This "parallel clustering" function facilitates optimization for the
341 minimum No. of members for clustering, which is the most important parameter that should be optimized
342 according to the purpose of the analysis; a smaller No. of members is expected to result in an increase in the
343 detected clusters; however, the ratio of clusters without supportive information for evaluating certainty will
344 also increase. For example, using this function, the clusters obtained using the previous setting with the No.

345 of members =4 (setting 1; Figure 3A), and the current setting with No. of members =10 (setting 2; Figure
346 3B) can be easily compared. When the results with No. of members = 2, 3, and 4 were compared for hAGP,
347 the No. of members = 2 resulted in an increased number of detected clusters, but most of the increased
348 clusters were excluded in the subsequent selection step (Table S2). In addition, manual inspection after the
349 selection step was required with No. of members = 2. In contrast, with No. of members = 3, more
350 assignments were obtained without manual inspection after selection compared to No. of members = 4,
351 demonstrating that the No. of members = 3 is the most suitable for hAGP analysis.

352 We often use a setting that considers only the number of Hex to find clusters, with the aim of
353 detecting glycopeptide clusters with only high- or oligo-mannosylated glycans. There are many possible
354 compositions of complex- and hybrid-type glycans; however, many oligo-mannosylated glycans are
355 limited to M9–M5. Thus, the minimum cluster number was lowered to three to estimate oligo-
356 mannosylated glycan carriers. In the hAGP samples, the abundance of oligo-mannosylated glycans was
357 low, resulting in no detection in this setting (Setting 3; Figure 3C). This finding is consistent with previous
358 reports that glycan structures attached to hAGP are branched, sialylated and fucosylated at the branches, and
359 slightly extended by polylactosamine (41–44).

360 The individual time settings for each type of glycan unit and parallel clustering with different
361 settings effectively increased the detection sensitivity, and multiple clusters with the same core but different
362 sialic acid numbers could now be combined into one cluster. With parallel clustering of these three search
363 parameter settings for hAGP, 65 clusters comprising 1,422 types of glycopeptides were detected (Figure
364 3D). Notably, among these clustered monoisotopic peaks of glycopeptides, 532 (37%) were corrected by
365 correction functions, demonstrating their utility in improving glycopeptide detection efficiency. The average

366 number of members in each cluster was 19.1. Among the detected monoisotopic signals, 2,309 signals were
367 assigned to glycopeptide clusters.

368 To evaluate the usefulness of the parallel clustering function for detecting glycopeptides with a
369 specific glycan structure, such as oligo-mannosylated glycopeptides, we used a dataset of HL-60 cell lysates
370 (HILIC fraction 2) as a complex mixture of unknown proteins containing glycopeptides decorated with
371 oligo-mannose (32). Because the sample glycopeptides used were desialylated, parallel clustering was
372 performed under two search parameter settings: setting 1 (previous setting) and setting 3 (oligo-
373 mannosylated glycans) for the hAGP sample. Figure 4A shows the clusters detected in setting 1, in which
374 the minimum number of clusters was set to four. When setting 2 (only Hex was considered a glycan unit
375 and the minimum number of clusters was three) was added to the setting 1 search, the number of detected
376 clusters increased (Figure 4B). For clarity, only the increased numbers of clusters (No. of members: 3) were
377 visualized (Figure 4C), demonstrating the presence of oligo-mannosylated glycopeptides. Thus, GRable
378 can be applied to large-scale and in-depth analyses of glycoproteins, in which the comprehensiveness of
379 glycan heterogeneity is improved, and glycan structure-specific search is allowed.

380

381 **Matching (Step 6) with improved evaluation of the analysis using inter-clustering analysis function**

382 In this step, GRable comprehensively searches all combinations of core peptide and glycan
383 compositions for each glycopeptide cluster by consistency in their masses, based on the lists of possible
384 peptide and glycan compositions provided by users. Accordingly, the results sheet shows information on
385 the three matched components (glycopeptides, peptides, and glycans) and the differences from the
386 calculated mass values for each cluster. Note that this step considers matching by mass alone; thus, all
387 matched results that meet the criteria can be provided for a single cluster. In the subsequent step, the most

plausible match was selected from among the matching candidates for each cluster, and the accuracy of the selected match was evaluated. The likelihood of the assigned glycan composition was crucial for this selection. Accordingly, positive points for likely compositions (and negative points for unusual compositions as an option), which can be defined in the glycan composition list, were assigned to the glycan compositions that matched with the matching result list, and the total score was calculated for each cluster as the sum of the points of all members.

Regarding this step, the previous Glyco-RIDGE analysis was limited by a low match rate. Mass calculations assume that peptides are ionized with protons such as MH^+. Some clusters with no appropriate match were non-proton adduct clusters. Accordingly, GRable was equipped with an inter-cluster analysis function in the Matching step to detect clusters comprising glycopeptides ionized with non-proton cations such as iron (Fe^{3+}) and ammonium (NH^{4+}). In this inter-cluster analysis, the presence of such adducts was judged for each unit of the glycopeptide cluster and was unlikely to include unintended misdetection caused by isobaric combinations (e.g., NeuAc + NH_4^+ and Hex + dHex (24)). In the matching results for hAGP, using lists of 56 core peptides (Data S1) and 134 glycan compositions (Data S2), 65 clusters were identified. Among them, 46 clusters matched; however, 19 did not (Data S6). Over half of the clusters (i.e., 44 clusters) were related to the corresponding iron or ammonium adducts with the inter-cluster analysis function, as shown in the "relation between clusters" sheet (Data S6). This information facilitates technical evaluation because non-proton adducts cause sensitivity loss; thus, reducing these adducts during sample preparation and analysis is essential. The remaining unmatched clusters may be attributed to the lack of core peptides in the list; therefore, the inter-cluster analysis function provides clues as to how the list should be improved.

409 **Selection (Step 7) with improved selecting and evaluating of assignments using MS2 information**

410 This last step is crucial because selecting the most plausible match is essential for guaranteeing
411 the certainty of the Glyco-RIDGE analysis results, considering that the false discovery rate cannot be defined.
412 Because this selection step was conducted with manual inspections in our previous procedure, using this
413 step with clear criteria in GRable represents a large technical improvement. As previously described, a
414 crucial criterion is the total score of each matched cluster. Accordingly, this step sets the total score threshold,
415 and only matches within the threshold are shown in the selection result sheet. When the Clustering and
416 Matching steps are performed correctly, all members match any composition, and many points are obtained
417 based on the glycan point list. In addition to the total score, additional criteria were the delta mass and delta
418 RT between the glycopeptide and the corresponding core peptide, albeit unused for automated selection by
419 GRable. The usefulness of relative RT information has also been demonstrated in MS2-based *N*-
420 glycoproteomics (24). Thus, all members within these criteria are highlighted in the results sheet, visually
421 confirming the certainty of the results. Notably, identical results were obtained for hAGP analysis, even
422 without the RT of the core peptide (i.e., RT=0 for all peptides in the core peptide list), suggesting that it is
423 possible to use predicted peptide lists when an analyte is a purified glycoprotein.

424 The commercial hAGP glycoprotein specimen used here also contained serum glycoproteins
425 such as haptoglobin (HPT) in the core peptide list (Data 1). For cluster 1, two candidate core peptides, hAGP
426 isoform 1 (A1AG1) and HPT, were used. The total scores for A1AG1 and HPT were 28 and 11,
427 respectively. However, the delta RTs of all members for HPT are over 15 min, whereas the delta RT of
428 A1AG1 members is within -6.5–5.1 min, meeting the defined criteria. Accordingly, A1AG1 was selected
429 as the core peptide for cluster 1. Similarly, the most plausible matching can be easily selected for all matched
430 clusters.

431 As the other important improved point, this selection step allows evaluating the "confidence level"
432 of the selected results by using MS2 information. This scrutiny and curation step to reduce misassignment
433 and ambiguity is critical in current *N*-glycoproteomics (37). Therefore, as with many MS2-based
434 glycopeptide identification methods (12), GRable first searches glycan fragment ions called "diagnostic
435 ions" in MS2 spectra for cluster members to evaluate whether the spectrum is attributed to a glycopeptide.
436 Diagnostic ions, such as fragments of HexNAc and HexNAc + Hex, were searched as defaults. The
437 variation in the diagnostic ions can be set without restricting the number of ions. These diagnostic ions are
438 not used for selection and are only added to the selection result list, allowing users to utilize this information
439 for different purposes. In addition, glycopeptide fragments such as Y0, Y1, and Y2 were searched to
440 estimate the mass of Y0, namely, the peptide moiety. Because the Y0 signals are often too weak to be
441 detected, GRable is equipped with a function for predicting the Y0 mass based on the masses of other related
442 signals if multiple signals are present. A matched peptide was considered correct if the mass value coincided
443 with that of the matched peptide. This MS2 information is summarized for the selected clusters in the "MS2
444 info for Clusters" sheet, and all the search results are available in the "All MS2 info" sheet of an exported file
445 (Data S6). The presence of such MS2 information is visible in the "Selection results" sheet and accessible
446 directly via an intra-file link for all glycopeptides in the sheet.

447 For hAGP, all assigned glycopeptides showed MS2 spectra with HexNAc(204) and HexNAc +
448 Hex(366) as diagnostic ions characteristic of glycopeptides, supporting the certainty of assignment. In
449 addition, 45 of the 78 monoisotopic peaks in cluster 1 were accompanied by MS2 information, and the
450 predicted peptide sequence was an A1AG1-derived peptide (WFYIASAFRNEEYNK). Therefore, the
451 selection result for this cluster was confirmed using the MS2 information.

452 In this context, the Glyco-RIDGE results were presented with the "confidence level" using
453 GRable. The probability/confidence of the assignment for core peptides is ranked as follows: 1) "High";
454 clusters having any member(s) showing Y0-related ions corresponding to the presumed core, 2) "High";
455 clusters having any member(s) identified as the same core using the MS2-based search engine, 3)
456 "Medium"; clusters having any member(s) showing glycan diagnostic ions, 4) "Low"; clusters without any
457 MS2 support. We believe that the confidence of the assignment is high even in the "Low" rank when their
458 mass and RT differences fit under the threshold and the glycan point is the highest in the cluster. This was
459 also supported when the assigned core peptide for a low-rank cluster was found in the core peptide list based
460 on confident deglycopeptide identification for the identical samples with 1% FDR. Originally, GRable was
461 designed to estimate signals that MS2 spectra could not identify; thus, cluster members without MS2
462 support and clusters with no MS2 data for all members are notable features of GRable. Such a glycopeptide
463 detection methodology allows the evaluation of the certainty of MS2 information-lacking assignments
464 based on their connected members in the same cluster. In this evaluation, visualization of the two-
465 dimensional (i.e., RT and mass) relationship of each cluster member in the cluster sheet allows manual
466 inspection of whether an unexpected glycan composition is likely; with this manual evaluation,
467 unconvincing assignments can be excluded from final results.

468

469 **Improved in-depth glycoproteomics results using GRable**

470 Two isoforms of hAGP exist, isoform 1 (A1AG1) and isoform 2 (A1AG2), of which A1AG1
471 is the major component. Based on the improved GRable procedure, five glycoproteins, including HPT,
472 plasma protease C1 inhibitor, and lymphatic vessel endothelial hyaluronic acid receptor 1 as serum-derived

473 contaminants, were assigned to the hAGP samples (Table S3). Among the 14 assigned clusters, seven (1,
474 4, 13, 23, 26, 43, and 48) and six (1, 3, 6, 11, 19, and 23) were assigned as A1AG1 and A1AG2, respectively.
475 These clusters covered all five *N*-glycosylation potential sites of A1AG1 and A1AG2 (the two sites are
476 common in A1AG1), except for site 93 of A1AG2; *N*-glycosylation impairment at this site is consistent
477 with a previous report (44). The estimated core peptides for each cluster were the same as those estimated
478 using the MS2 information utilization function of GRable and Byonic, increasing the confidence of the
479 results. Each cluster strictly corresponded to one core peptide; thus, the clustering view facilitated effective
480 glycan macroheterogeneity visualization for each glycosite (Figure 5). Notably, two *N*-glycosites (sites 33
481 and 103) were detected in almost the same core peptide, except for one residue between the two isoforms;
482 however, their site-specific glycoforms could be visualized separately. This result highlights the usefulness
483 of GRable for elucidating differences in the glycosylation status of target glycoproteins, even with other
484 contaminated glycoproteins.

485 Using GRable, 123 and 108 site-specific glycoforms of A1AG1 and A1AG2 were estimated,
486 respectively (Table 1 and Table S4). For some *N*-glycosites, the same glycan compositions were estimated
487 for multiple core peptides, supporting the certainty of the results. In the case of site 103 of A1AG1, 24 of 29
488 compositions assigned for ENGTISR were also detected for the other two peptide sequences. hAGP had
489 *N*-glycosylations with highly sialylated and fucosylated branches and extended with polylactosamine (43,
490 44). Similar results were obtained with GRable (Table S4); tetrasialo-glycans were estimated at all five
491 A1AG1 sites, indicating high sialylation. Glycans with dHex $\geq$ 2 were estimated at four A1AG1 sites,
492 indicating its fucosylation at both its core and branches. In addition, glycan components for polylactosamine
493 (i.e., Hex + HexNAc $\geq$ 10) were estimated at four sites, indicating the presence of such extended glycans.

494

495 **Evaluation of the certainty of glycoproteomic results using GRable**

496 To evaluate the suitability of the search parameter settings used here for accurate glycopeptide
497 detection, we performed a GRable analysis of the IGOT-LC/MS data (i.e., *N*-deglycoproteomic data) used
498 for preparing the core peptide list using the same search parameter settings as those for the *N*-glycoproteomic
499 data. As expected, the detected clusters for *N*-deglycoproteomic data significantly decreased for both hAGP
500 (from 65 to 4 clusters) and HL-60 (from 207 to 32 clusters), and all assigned glycopeptides were also
501 observed for *N*-glycoproteomic data. These results indicate that the settings used for GRable analysis were
502 appropriate.

503 In the hAGP analysis, the confidence level of all assigned clusters was High or Medium,
504 indicating that at least one member had MS2 information for every glycopeptide cluster. To validate the
505 certainty of the obtained results for hAGP analysis, 110 MS2 spectra of all the assigned glycopeptides with
506 MS2 information were manually checked (Figure S2), confirming that the core peptides were correctly
507 assigned. It is noteworthy that NeuAc-derived diagnostic ions were observed for several glycopeptides
508 without NeuAc owing to insufficient isolation, highlighting the risk of evaluating the accuracy of
509 glycoforms based only on diagnostic ions.

510 In addition, to evaluate the certainty and significance of the Glyco-RIDGE results obtained using
511 GRable, the results were compared with those obtained using the representative MS2-based software
512 Byonic. Notably, GRable could estimate more than four-fold more site-specific glycoforms than Byonic
513 and covered most of the site-specific glycoforms estimated by Byonic. Only two of the site-specific
514 glycoforms (Hex(2)HexNAc(1) and Hex(3)HexNAc(2)) were estimated by Byonic but not by GRable at

site 56) because one component was absent (Hex(3)HexNAc(3) or Hex(3)HexNAc(2)NeuAc(1)). As expected, the MS2-based estimation by Byonic succeeded on higher-intensity glycopeptides (Data S9 and S10). Notably, such highly sialylated and polylactosamine-attached glycopeptides were not detected by Byonic (Table S4), and the median number of NeuAc for 38 glycopeptides assigned by GRable but not by Byonic was 3, whereas the median number of 27 glycopeptides assigned by Byonic was 2. Accordingly, it is likely that MS2 spectra was too poor to assign a glycopeptide, partly due to their low ionization efficiency. Thus, these results confirm the reliability of the Glyco-RIDGE results obtained by GRable, highlighting the usefulness of this method as a complementary method to the current gold standard, MS2-based glycoproteomics.

CONCLUSION

We developed a novel software program, GRable, to enable semiautomated Glyco-RIDGE analysis. GRable version 1.0 can run online freely with a demo or user data using a web browser via the GlyCosmos Portal (https://glycosmos.org/grable). Note that the current version is guaranteed only for Thermo Fisher Scientific data. Some algorithms of the existing Glyco-RIDGE method were improved during the implementation of GRable version 1.0, which can be applied to in-depth site-specific glycoform analysis of intact sialylated glycopeptides derived from purified and crude glycoproteins. Thus, this software will help analyze the status and changes in glycans to obtain biological and clinical insights into protein glycosylation. The novel parallel clustering function enabled a targeted search focusing on multiple glycan structural layers, including the stem, branching, and terminal moieties, such as Lewis epitopes. This software also allows evaluating of a "confidence level" especially using MS2 information, for each glycopeptide obtained by MS1-based detection methodology. Using the MS2 utilization function opened doors to all glycoproteomics researchers who considered MS2-based glycoproteomics as the gold standard, allowing for the community evaluation of GRable by comparing it with other glycoproteomics software. The combined use of MS1-based glycoproteomics will be useful for expanding glycopeptide detection and providing supporting evidence for MS2-based glycoproteomics. Thus, the Glyco-RIDGE method with its unique MS1-based glycopeptide detection principle has complementary roles to MS2-based glycoproteomic methods, and GRable will be a powerful glycoproteomics tool when combined with currently developed MS2-based software. Accordingly, GRable will be updated by continuously improving the Glyco-RIDGE methodology and software while comparing the upcoming MS2-based methods and software.

ACKNOWLEDGMENTS

This study was funded by a project for utilizing glycans in the development of innovative drug discovery technologies (grant number JP20ae0101021) from the Japan Agency for Medical Research and Development (AMED). We thank Dr. Mika Fujita and Ms. Masako Sukegawa for their technical assistance. We also appreciate Dr. Kiyoko F Aoki-Kinoshita and Dr. Akihiro Fujita of Soka University for releasing GRable at the GlyCosmos Portal (https://glycosmos.org/).

DATA AVAILABILITY

The MS data set for hAGP presented in this study has been deposited in the ProteomeXchange Consortium via the jPOST partner repository with the dataset identifiers PXD046226 (ProteomeXchange) and JPST001876 (jPOST).

AUTHOR INFORMATION

Author Contributions

All authors contributed to writing the manuscript. All authors have approved the final version of the manuscript. †C.N.-O., D.T., and A.T. contributed equally.

Notes

The authors declare no competing financial interest.

565 **REFERENCES**

566 1. Pascovici, D., Wu, J. X., McKay, M. J., Joseph, C., Noor, Z., Kamath, K., et al. (2018) Clinically
567 relevant post-translational modification analyses—maturing workflows and bioinformatics Tools.
568 *Int. J. Mol. Sci.* **20**, 16
569 2. Varki, A. (2017) Biological roles of glycans. *Glycobiology* **27**, 3–49
570 3. Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita,
571 T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. eds. (2022) *Essentials of*
572 *Glycobiology* (Cold Spring Harbor Laboratory Press) 4th ed.
573 4. Čaval, T., Heck, A. J. R., and Reiding, K. R. (2021) Meta-heterogeneity: Evaluating and describing
574 the diversity in glycosylation between sites on the same glycoprotein. *Mol. Cell. Proteomics.* **20**,
575 100010
576 5. Fang, P., Ji, Y., Oellerich, T., Urlaub, H., and Pan, K. T. (2022) Strategies for proteome-wide
577 quantification of glycosylation macro- and micro-heterogeneity. *Int J Mol Sci.* **23**, 1609
578 6. Kawahara, R., Chernykh, A., Alagesan, K., Bern, M., Cao, W., Chalkley, R. J., Cheng, K., Choo,
579 M. S., Edwards, N., Goldman, R., Hoffmann, M., Hu, Y., Huang, Y., Kim, J. Y., Kletter, D., Liquet,
580 B., Liu, M., Mechref, Y., Meng, B., Neelamegham, S., Nguyen-Khuong, T., Nilsson, J., Pap, A.,
581 Park, G. W., Parker, B. L., Pegg, C. L., Penninger, J. M., Phung, T. K., Pioch, M., Rapp, E., Sakalli,
582 E., Sanda, M., Schulz, B. L., Scott, N. E., Sofronov, G., Stadlmann, J., Vakhrushev, S. Y., Woo, C.
583 M., Wu, H. Y., Yang, P., Ying, W., Zhang, H., Zhang, Y., Zhao, J., Zaia, J., Haslam, S. M.,
584 Palmisano, G., Yoo, J. S., Larson, G., Khoo, K. H., Medzihradszky, K. F., Kolarich, D., Packer, N.
585 H., and Thaysen-Andersen, M. (2021) Community evaluation of glycoproteomics informatics
586 solutions reveals high-performance search strategies for serum glycopeptide analysis. *Nat. Methods.*
587 **18**, 1304–1316

588 7. Bern, M., Kil, Y. J., and Becker, C. (2012) in *Current Protocols in Bioinformatics* (John Wiley &
589 Sons, Inc., Hoboken, NJ, USA), p Unit13.20.
590 8. Fang, Z., Qin, H., Mao, J., Wang, Z., Zhang, N., Wang, Y., Liu, L., Nie, Y., Dong, M., and Ye, M.
591 (2022) Glyco-Decipher enables glycan database-independent peptide matching and in-depth
592 characterization of site-specific N-glycosylation. *Nat. Commun.* **13**, 1900

593 9. He, S.-M., Yang, P.-Y. ✉, Zeng, W.-F., Cao, W.-Q., Liu, M.-Q., and Yang, P.-Y. (2021) Precise,
594 fast and comprehensive analysis of intact glycopeptides and modified glycans with pGlyco3. *Nat.*
595 *Methods.* **2021**, 1–9
596 10. Polasky, D. A., Yu, F., Teo, G. C., and Nesvizhskii, A. I. (2020) Fast and comprehensive N- and O-
597 glycoproteomics analysis with MSFragger-Glyco. *Nat. Methods.* **17**, 1125–1132
598 11. Shen, J., Jia, L., Dang, L., Su, Y., Zhang, J., Xu, Y., Zhu, B., Chen, Z., Wu, J., Lan, R., Hao, Z.,
599 Ma, C., Zhao, T., Gao, N., Bai, J., Zhi, Y., Li, J., Zhang, J., and Sun, S. (2021) StrucGP: de novo
600 structural sequencing of site-specific N-glycan on glycoproteins using a modularization strategy.
601 *Nat. Methods.* **18**, 921–929

602 12. Cao, W., Liu, M., Kong, S., Wu, M., Zhang, Y., and Yang, P. (2021) Recent advances in software
603 tools for more generic and precise intact glycopeptide analysis. *Mol. Cell. Proteomics* **20**, 100060

604 13. Baker, P. R., Trinidad, J. C., and Chalkley, R. J. (2011) Modification site localization scoring
605 integrated into a search engine. *Mol. Cell. Proteomics.* **10**, M111.008078
606 14. Cheng, K., Chen, R., Seebun, D., Ye, M., Figeys, D., and Zou, H. (2014) Large-scale
607 characterization of intact N-glycopeptides using an automated glycoproteomic method. *J.*
608 *Proteomics.* **110**, 145–154

609 15. Toghi Eshghi, S., Shah, P., Yang, W., Li, X., and Zhang, H. (2015) GPQuest: A spectral library
610 matching algorithm for site-specific assignment of tandem mass spectra to intact N-glycopeptides.
611 *Anal. Chem.* **87**, 5181–5188
612 16. Park, G. W., Kim, J. H. J. Y., Hwang, H., Lee, J. Y., Ahn, Y. H., Lee, H. K., Ji, E. S., Kim, K. H.,
613 Jeong, H. K., Yun, K. N., Kim, Y. S., Ko, J. H., An, H. J., Kim, J. H. J. Y., Paik, Y. K., and Yoo, J.
614 S. (2016) Integrated GlycoProteome Analyzer (I-GPA) for automated identification and
615 quantitation of site-specific N-glycosylation. *Sci. Rep.* **6**, 21175
616 17. Nasir, W., Toledo, A. G., Noborn, F., Nilsson, J., Wang, M., Bandeira, N., and Larson, G. (2016)
617 SweetNET: A bioinformatics workflow for glycopeptide MS/MS spectral analysis. *J. Proteome*
618 *Res.* **15**, 2826–40

619 18. Pioch, M., Hoffmann, M., Pralow, A., Reichl, U., and Rapp, E. (2018) glyXtool MS: An open-
620 source pipeline for semiautomated analysis of glycopeptide mass spectrometry data. *Anal. Chem.*
621 **90**, 11908–11916
622 19. An, Z., Shu, Q., Lv, H., Shu, L., Wang, J., Yang, F., and Fu, Y. (2018) N-linked glycopeptide
623 identification based on open mass spectral library search. *Biomed Res. Int.* **2018**, 1564136
624 20. Devabhaktuni, A., Lin, S., Zhang, L., Swaminathan, K., Gonzalez, C. G., Olsson, N., Pearlman, S.
625 M., Rawson, K., and Elias, J. E. (2019) TagGraph reveals vast protein modification landscapes from
626 large tandem mass spectrometry datasets. *Nat Biotechnol.* **37**, 469–479
627 21. Narimatsu, H., Kaji, H., Vakhrushev, S. Y., Clausen, H., Zhang, H., Noro, E., Togayachi, A., Nagai-
628 Okatani, C., Kuno, A., Zou, X., Cheng, L., Tao, S.-C., Sun, Y. (2018) Current technologies for
629 complex glycoproteomics and their applications to biology/disease-driven glycoproteomics. *J*
630 *Proteome Res.* **17**, 4097–4112

631 22. Noro, E., Togayachi, A., Sato, T., Tomioka, A., Fujita, M., Sukegawa, M., Suzuki, N., Kaji, H., and
632 Narimatsu, H. (2015) Large-scale identification of N-glycan glycoproteins carrying Lewis x and
633 site-specific N-glycan alterations in Fut9 knockout mice. *J Proteome Res.* **14**, 3823–3834
634 23. Darula, Z., and Medzihradszky, K. F. (2015) Carbamidomethylation side reactions may lead to
635 glycan misassignments in glycopeptide analysis. *Anal. Chem.* **87**, 6297–6302
636 24. Klein, J., and Zaia, J. (2020) Relative retention time estimation improves N-glycopeptide
637 identifications by LC-MS/MS. *J. Proteome Res.* **19**, 2113–2121
638 25. Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi,
639 N., and Isobe, T. (2003) Lectin affinity capture, isotope-coded tagging and mass spectrometry to
640 identify N-linked glycoproteins. *Nat. Biotechnol.* **21**, 667–72
641 26. Kaji, H., Yamauchi, Y., Takahashi, N., and Isobe, T. (2006) Mass spectrometric identification of N-
642 linked glycopeptides using lectin-mediated affinity capture and glycosylation site-specific stable
643 isotope tagging. *Nat. Protoc.* **1**, 3019–27
644 27. Goldberg, D., Bern, M., Parry, S., Sutton-Smith, M., Panico, M., Morris, H. R., and Dell, A. (2007)
645 Automated N-glycopeptide identification using a combination of single- and tandem-MS. *J.*
646 *Proteome Res.* **6**, 3995–4005
647 28. Deshpande, N., Jensen, P. H., Packer, N. H., and Kolarich, D. (2010) GlycoSpectrumScan: fishing
648 glycopeptides from MS spectra of protease digests of human colostrum sIgA. *J. Proteome Res.* **9**,
649 1063–1075
650 29. Choo, M. S., Wan, C., Rudd, P. M., and Nguyen-Khuong, T. (2019) GlycopeptideGraphMS:
651 Improved glycopeptide detection and identification by exploiting graph theoretical patterns in mass
652 and retention time. *Anal. Chem.* **91**, 7236–7244

653 30. Trinidad, J. C., Schoepfer, R., Burlingame., A. L., and Medzihradszky, K. F. (2013) N- and O-
654 glycosylation in the murine synaptosome. *Mol. Cell. Proteomics* **12**, 3474–3488
655 31. Chalkley, R. J., Medzihradszky, K. F., Darula, Z., Pap, A., and Baker, P. R. (2020) The effectiveness
656 of filtering glycopeptide peak list files for Y ions. *Mol. Omics.* **16**, 147–155
657 32. Togayachi, A., Tomioka, A., Fujita, M., Sukegawa, M., Noro, E., Takakura, D., Miyazaki, M.,
658 Shikanai, T., Narimatsu, H., and Kaji, H. (2018) Identification of poly-N-acetyllactosamine-
659 carrying glycoproteins from HL-60 human promyelocytic leukemia cells using a site-specific
660 glycome analysis method, Glyco-RIDGE. *J. Am. Soc. Mass Spectrom.* **29**, 1138–1152
661 33. Noro, E., Matsuda, A., Kyoutou, T., Sato, T., Tomioka, A., Nagai, M., Sogabe, M., Tsuruno, C.,
662 Takahama, Y., Kuno, A., Tanaka, Y., Kaji, H., and Narimatsu, H. (2021) N-glycan structures of
663 *Wisteria floribunda* agglutinin-positive Mac2 binding protein in the serum of patients with liver
664 fibrosis. *Glycobiology* **31**, 1268–1278
665 34. Mitani, Y., Yasuno, R., Kihira, K., Chung, K. M., Mitsuda, N., Kanie, S., Tomioka, A., Kaji, H.,
666 and Ohmiya, Y. (2022) Host-dependent producibility of recombinant *Cypridina noctiluca*
667 luciferase with glycosylation defects. *Front. Bioeng. Biotechnol.* **10**, 92
668 35. Sogabe, M., Kojima, S., Kaya, T., Tomioka, A., Kaji, H., Sato, T., Chiba, Y., Shimizu, A.,
669 Tanaka, N., Suzuki, N., Hayashi, I., Mikami, M., Togayachi, A., and Narimatsu, H. (2022) A
670 sensitive new assay system for serum *Wisteria floribunda* agglutinin-reactive ceruloplasmin that
671 distinguishes ovarian clear cell carcinoma from endometrioma. *Anal. Chem.* **94**, 2476–2484

672 36. Hiono, T., Sakaue, H., Tomioka, A., Kaji, H., Sasaki, M., Orba, Y., Sawa, H., and Kuno, A. (2024)
673 Combinatorial approach with mass spectrometry and lectin microarray dissected site-specific
674 glycostem and glycoleaf features of the virion-derived spike protein of ancestral and γ variant

675 SARS-CoV-2 Strains. *J. Proteome Res.* **23**, 1408–1419

676 37. Chau, T. H., Chernykh, A., Kawahara, R., and Thaysen-Andersen, M. (2023) Critical
677 considerations in N-glycoproteomics. *Curr Opin Chem. Biol.* **73**, 102272

678 38. Kuno, A., Ikehara, Y., Tanaka, Y., Angata, T., Unno, S., Sogabe, M., Ozaki, H., Ito, K., Hirabayashi,
679 J., Mizokami, M., and Narimatsu, H. (2011) Multilectin assay for detecting fibrosis-specific glyco-
680 alteration by means of lectin microarray. *Clin. Chem.* **57**, 48–56

681 39. Medzihradszky, K. F. (2005) Characterization of protein N-glycosylation. *Methods Enzymol.* **405**,
682 116–138

683 40. Go, E. P., Zhang, S., Ding, H., Kappes, J. C., Sodroski, J., and Desaire, H. (2021) The opportunity
684 cost of automated glycopeptide analysis: case study profiling the SARS-CoV-2 S glycoprotein.
685 *Anal. Bioanal. Chem.* **413**, 7215–7227

686 41. Nakano, M., Kakehi, K., Tsai, M. H., and Lee, Y. C. (2004) Detailed structural features of glycan
687 chains derived from alpha1-acid glycoproteins of several different animals: the presence of
688 hypersialylated, O-acetylated sialic acids but not disialyl residues. *Glycobiology* **14**, 431–441

689 42. Hevér, H., Nagy, K., Xue, A., Sugár, S., Komka, K., Vékey, K., Drahos, L., and Révész, Á. (2022)
690 Diversity matters: Optimal collision energies for tandem mass spectrometric analysis of a large set
691 of N-glycopeptides. *J. Proteome Res.* **21**, 2743–2753

692 43. Wong, T. L., Mooney, B. P., Cavallero, G. J., Guan, M., Li, L., Zaia, J., and Wan, X. F. (2023)
693 Glycoproteomic analyses of influenza A viruses using timsTOF Pro MS. *J. Proteome Res.* **22**, 62–
694 77

695 44. Keser, T., Tijardović, M., Gornik, I., Lukić, E., Lauc, G., Gornik, O., and Novokmet, M. (2021)
696 High-throughput and site-specific N-glycosylation analysis of human alpha-1-acid glycoprotein

697 offers a great potential for new biomarker discovery. *Mol. Cell Proteomics.* **20**, 100044

698

699 **TABLE**

700 **Table 1. Site-specific hAGP glycoforms estimated using GRable and Byonic.**

		No. of site-specific glycoforms [b]			
		A1AG1		A1AG2	
Site	Sequence [a]	GRable	Byonic	GRable	Byonic
33(NAT)	A1AG1: QIPLCANLVPVPITNATLDQITGK A1AG2*: QIPLCANLVPVPITNATLDRITGK	21 (14)	7 (0)	23 (13)	10 (0)
56(NKS)	WFYIASAFRNEEYNK	35 (31)	6 (2)	35 (31)	6 (2)
72(NKT)	SVQEIQATFFYFTPNK	24 (18)	6 (0)	24 (18)	6 (0)
93(NTT)	A1AG1: QDQCIYNTTYLNVQR A1AG2: QNQCFYNSSYLNVQR	13 (11)	2 (0)	0 (0)	0 (0)
103(NGT)	A1AG1*: ENGTISR A1AG2*: ENGTVSR	29 (23)	6 (0)	25 (21)	4 (0)
		122 (97)	27 (2)	107 (83)	26 (2)

701

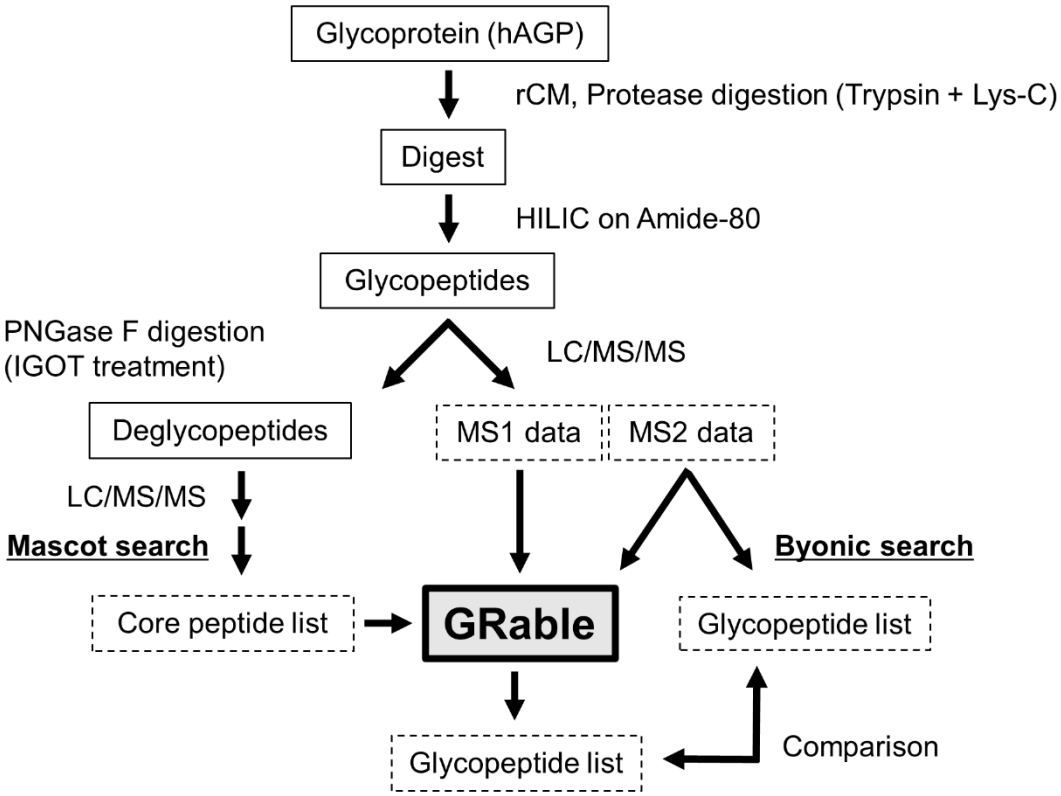
702 [a] *N*-glycosylation sites are indicated in bold. Different residues in the two isoforms are underlined. The

703 asterisks indicate the presence of other peptide sequences of different lengths.

704 [b] Numbers within parentheses indicate the number of glycoforms estimated using only GRable or Byonic.

705

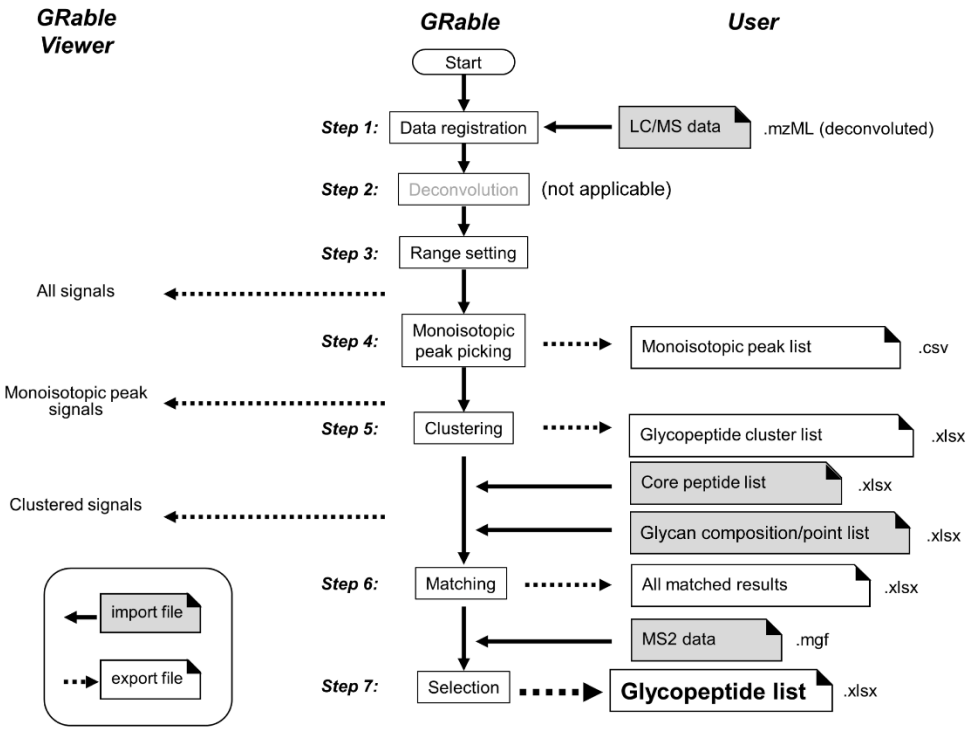
706 **FIGURES**



707

708 **Figure 1. Workflow of glycoproteomic analysis for hAGP using GRable.** hAGP was subjected to
709 reduction and carbamidomethylation (rCM) using dithiothreitol and iodoacetamide, respectively, and
710 digested with trypsin and Lys-C endopeptidase. The digest was used in hydrophilic interaction
711 chromatography (HILIC) on an Amide-80 column to capture glycopeptides. A small aliquot of the
712 glycopeptide fraction was treated with peptide-*N*-glycosidase F (PNGase F) in ^{18}O-labeled water to remove
713 *N*-glycans and to label deglycosylated Asn as ^{18}O-labeled Asp (isotope-coded glycosylation site-specific
714 tagging; IGOT) (25, 26). The IGOT-treated deglycopeptides were analyzed using liquid chromatography-
715 tandem mass spectrometry (LC/MS/MS), and the MS2 data were used for database search using Mascot
716 to prepare a core peptide list. In parallel, another aliquot of the glycopeptide fraction was analyzed through
717 LC/MS/MS. MS1 data of the glycopeptide analysis underwent GRable analysis. GRable used the core
718 peptide (Data S1) and glycan composition (Data S2) lists to assign each glycopeptide signal, whereas MS2
719 data are used for adding the confidence level to the GRable results. Glycopeptide analysis MS2 data was
720 also used for glycopeptide estimation using Byonic. The resulting glycopeptide lists of GRable and Byonic
721 analyses were compared.

722



723

724 **Figure 2. Overview of data processing by GRable.** GRable proceeds in seven steps using the four files
725 indicated in gray font. In Step 1, GRable allows the registration of the LC/MS data of glycopeptides in the
726 mzML format, which ensures great versatility and ease of data processing. In the current version, the
727 deconvolution function (Step 2) is not applicable although it is visible in the user interface for future
728 implementation. Step 3 was designed to set the RT range, mass range, and minimum signal intensity
729 (threshold) over which the analysis was performed. Step 4 was intended to find and group signals of an
730 identical ion based on three parameters: time (scan), mass (MH+), and intensity, and to obtain the
731 monoisotopic mass of each signal group at the peak time, using our unique algorithms. In Step 5, a series of
732 signals from the glycopeptide group were found as a cluster based on the elution behavior and mass
733 difference of its members. In Step 6, GRable searches for a combination of core peptides and glycan
734 compositions that match the mass of the putative glycopeptide within the allowed mass error (user setting)
735 according to the following equation: Observed M(glycopeptide) = calculated M(core peptide identified) +
736 M(Hex)*i + M(HexNAc)*j + M(dHex)*k + M(NeuAc)*l (M is a mass value, and i, j, k, and l are integers).
737 In Step 7, the most plausible combination among multiple candidate combinations suggested for one
738 glycopeptide cluster was selected. Then, their reliability at the cluster level as well as the single glycopeptide
739 level was evaluated using information from the results and additional MS2 information. The results of Steps
740 3–5 were visually confirmed using a viewer in the main window of the software. The detailed results of
741 Steps 4–7 can be exported as an Excel file with each setting, as shown for hAGP (Data S5 and S6) and HL-
742 60 cell lysates (Data S7 and S8).

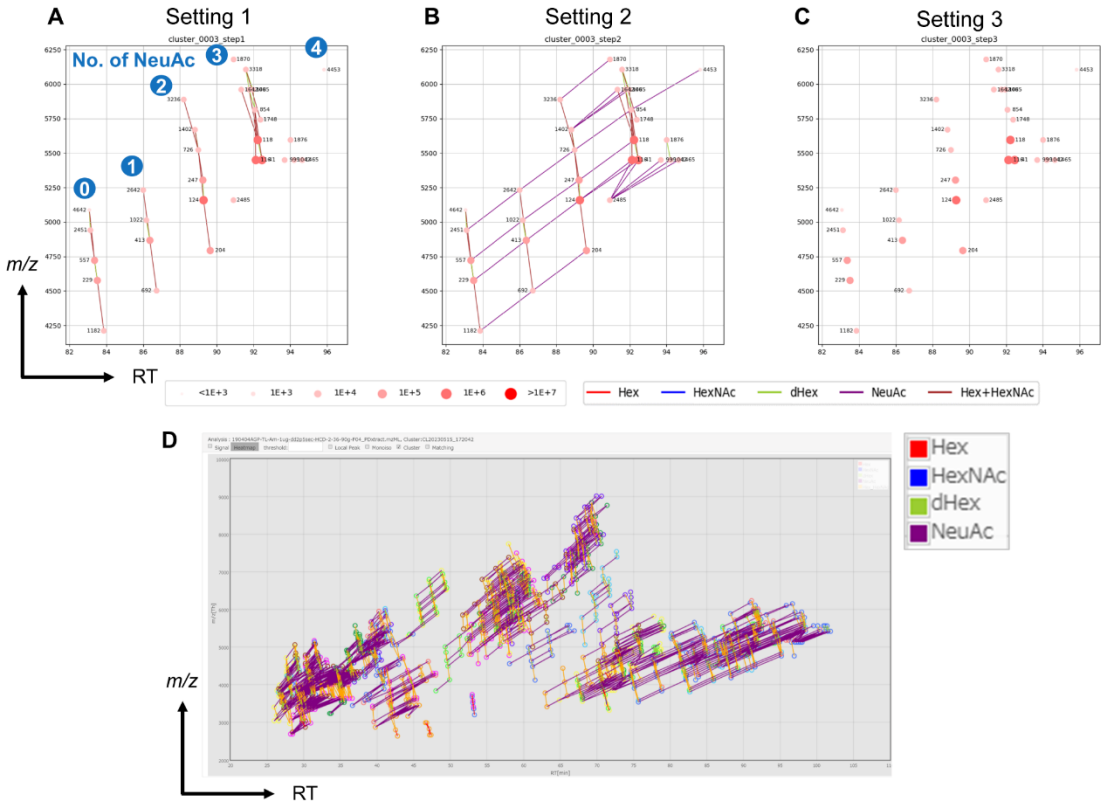


Figure 3. hAGP clustering results. Representative results (Cluster 3) of parallel clustering with three search parameter settings: **(A)** Setting 1 (a setting in our previous prototype), **(B)** Setting 2 (an extra setting for detecting sialylated glycopeptides), and **(C)** Setting 3 (an extra setting for detecting oligo-mannosylated glycopeptides). The detailed settings are as follows: for Setting 1, neutral saccharides (Hex, HexNAc, dHex, Hex+HexNAc): -1–0 (min), and the minimum number of members: 4 (setting for maximizing No. of detected clusters with minimized misdetection); for Setting 2, neutral saccharide: -1–0 (min), acidic saccharide (NeuAc for human): 1–4 (min), and minimum number of members: 10 (setting for guaranteeing the accuracy of assignment); and for Setting 3, neutral saccharide (Hex only): -1–0 (min), and minimum number of members: 3. Dots indicate the monoisotopic masses of the glycopeptide signals at each RT peak, and the size and color depth of the dot indicates the signal intensity. Deep red and larger dots indicate high intensity, and small pale red dots indicate low intensity. All the clusters for hAGP shown in Data S6 can be visualized in the viewer of the GRable main window **(D)**. The colors of lines that connect the signals indicate the type of glycan unit between the signals: red: Hex, blue: HexNAc, green: dHex, Hex + HexNAc: brown, and NeuAc: purple.

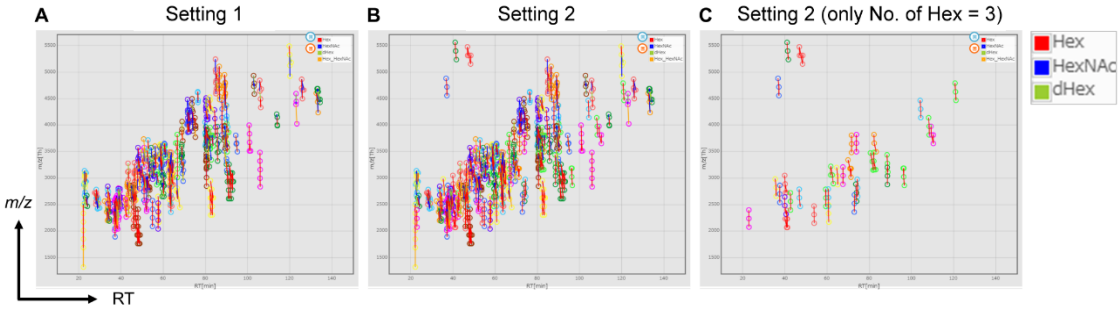


Figure 4. Clustering view of HL-60 cell lysate samples. All the clusters for HL-60 cell lysates (HILIC Fr.2) shown in Data S8 can be visualized in the viewer of the GRable main window. Parallel clustering was performed under two search parameter settings: **(A)** Setting 1 (a setting of the prototype) and **(B)** Setting 2 (a setting for detecting oligo-mannosylated glycopeptides). The detailed settings are as follows: for Setting 1, neutral saccharides (Hex, HexNAc, dHex, Hex+HexNAc): -1–0 (min), and minimum number of members: 4; for Setting 2, neutral saccharide (Hex only): -1–0 (min), and minimum number of members: 3. Among the detected clusters in Setting 2, only clusters with three members having oligo-mannosylated glycans is presented **(C)**.

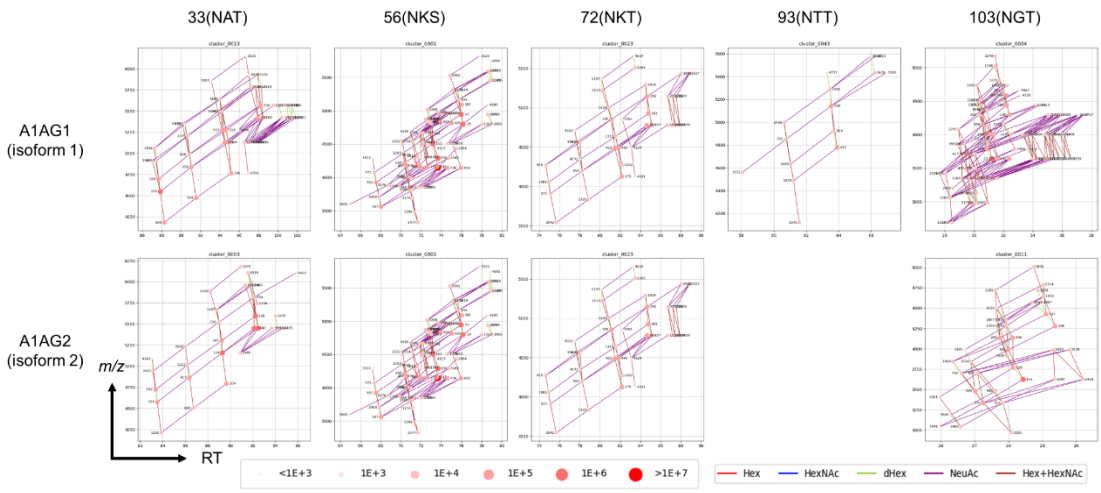
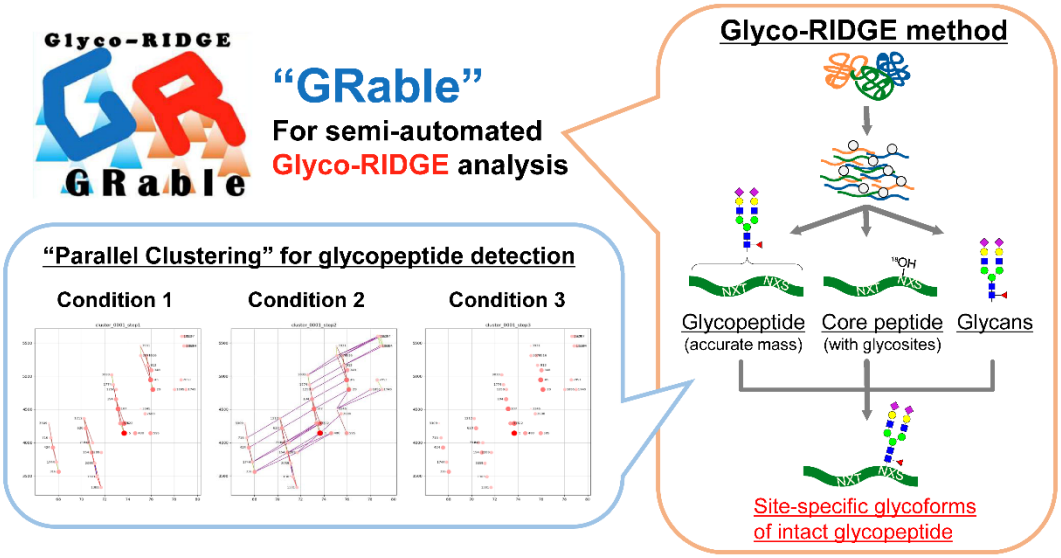


Figure 5. Glycopeptide clusters assigned for 5 *N*-glycosylation sites of hAGP isoforms. Corresponding core peptide sequences for each cluster are indicated in Table 1. Since the core peptide of two isoforms is identical for sites 56 and 72, the same clusters are indicated for both isoforms.

Graphical Abstract