GRable version 1.0: A software tool for site-specific glycoform analysis using the improved Glyco-RIDGE method with parallel clustering and 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

[*] Correspondence

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

RUNNING TITLE

GRable software for glycoproteomics using Glyco-RIDGE

This article contains supplemental data.

KEYWORDS

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

ABBREVIATIONS

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

ABSTRACT

High-throughput intact glycopeptide analysis is crucial to elucidate the physiological and pathological status of glycans attached to each glycoprotein. Mass spectrometry-based glycoproteomic methods have challenges due to 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 chromatographical properties of glycopeptides and then searches each combination of core peptide and glycan composition 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 algorisms, including "parallel clustering." This unique function improved the comprehensiveness of glycopeptide detection and allowed analysis to focus on specific glycan structures, such as pauci-mannose. The other notable improvement is evaluating the "confidence level" of the GRable results using MS2 information. This function facilitated reduced misassignment of core peptide and glycan composition and improved interpretation of the results. 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 of glycans to obtain biological and clinical insights into protein glycosylations. GRable can run online freely using a web browser via the GlyCosmos Portal (https://glycosmos.org/grable).

INTRODUCTION

Protein glycosylation is a common and complex post-translational modification in eukaryotes (1) and plays a pivotal role in various biological processes (2). This post-translational modification regulates the glycoprotein functions and localizations by modulating their tertiary structure and interaction with other molecules (2). However, revealing the glycan structure-function relationship is challenging because glycan structures are highly diverse and heterogeneous and differ in organisms, cells, proteins, and the attached sites (3). Furthermore, post-translational modification alters the cell state change through differentiation, carcinogenesis, infection, nutrition, medicine, and stimulus (3). Because glycans are naturally heterogeneous, elucidating the function of glycans by uncovering the state and changes of glycans with heterogeneities of each level is important. The heterogeneous levels include micro (glycan variation of one glycosite), macro (with or without glycosylation of one glycosite), and meta (glycan variation of the entire glycoprotein molecule) (4).

Although current improvements in mass spectrometry (MS) made it possible for in-depth structural analysis of glycopeptide peptide backbones and glycomes, elucidating the glycan compositions of specific sites for proteins with multiple glycosylated sites (i.e., micro-heterogeneity) is challenging (5). First, the glycan compositions of each site must be determined to estimate the site-specific glycoforms. For this purpose, glycomes and deglycosylated peptide analysis is insufficient, and thus intact glycopeptides must be analyzed directly. Current standard approaches rely on MS2 spectra to identify site-specific glycoforms (6). For MS2-based glycoproteomics, a commercial software named Byonic (7) is often used, and several in-house

software programs have been developed (8–21). However, these MS2-based methods have a critical problem that they require glycopeptide fragmentation, leading to lower detection sensitivity than the corresponding naked peptides. The decreased detection sensitivity is due to the lower ionization efficiency of glycopeptides, which is notable, especially for sialylated glycopeptides, owing to the negative property. In addition, glycan heterogeneity decreases the abundance of each glycopeptide. Therefore, in some MS1 spectra, many signals are not fragmented in data-dependent acquisition mode, and some MS2 spectra may be insufficient for glycopeptide identification.

To overcome this limitation, we developed an MS1-based identification approach named "Glycan heterogeneity-based Relational IDentification of Glycopeptide signals on Elution profile" (Glyco-RIDGE) (22). MS1 signals derived from glycopeptides with the same core peptide are assigned as a cluster in liquid chromatography-mass spectrometry (LC-MS) data based on their mass difference and closeness of retention times (RTs). The combination of the core peptide and glycan composition for these glycopeptides is annotated by matching the observed accurate glycopeptide mass with the sum of the masses of pre-identified core peptide candidates and presumed glycans (22). Accordingly, this MS1-based glycopeptide identification method is expected to achieve a more high-sensitive and comprehensive analysis of protein glycosylations compared with current fragmentation-dependent methods. Among the MS2 spectra assigned by the Glyco-RIDGE method and Byonic search, the ratio of the same assignment was quite high (349/353 = 98.9%) (23), demonstrating that the reliability of this method is similar to that of Byonic. Furthermore, the feasibility and usefulness of this method have been demonstrated using in-house software by applying it to analyze complex glycoprotein

mixtures (23, 24). Now this method can be used for site-specific glycoform analysis of intact (i.e., sialylated) glycopeptides (25–28).

However, prototype software has several limitations, especially in the comprehensiveness of glycan heterogeneity, including sialylated glycans and presenting the confidence level for each site-specific glycoform. Therefore, the present study introduces a novel software named "GRable" to execute the Glyco-RIDGE analysis semi-automatically. This software has some improvements in the monoisotopic mass selection and Glyco-RIDGE procedure compared with our previous in-house software to more accurately and comprehensively analyze intact glycopeptides. We analyzed MS data obtained with intact glycopeptides prepared from human serum α1-acid glycoprotein (hAGP) as a glyco-biomarker candidate for liver fibrosis to demonstrate the feasibility and utility of this software (29). In addition, we used previous data for glycopeptides obtained from human promyelocytic leukemia-60 (HL-60) cells as a crude sample model (23).

EXPERIMENTAL PROCEDURES

Principle of the Glyco-RIDGE method

The principle of the Glyco-RIDGE method was reported previously (22). This method first identifies glycopeptide signals based on the chromatographic properties of glycopeptides and mass differences due to the glycan heterogeneity. Glycopeptides with the same core peptide but different glycans elute within a narrow RT range; therefore, glycopeptide signals with similar RT and mass differences corresponding to glycan units can be assigned as a cluster without MS2 information. In parallel, core peptides in the glycopeptide sample are identified through isotope-coded glycosylation site-specific tagging (IGOT)-LC/MS/MS (30, 31). Considering that the mass value of the glycopeptide is the sum of the core peptide and glycan masses, the peptide and glycan combination for each glycopeptide is searched from these three lists (the lists of glycopeptides and core peptides containing masses and RTs, and a glycan composition list). The mass of a putative glycopeptide can be calculated using the following equation: observed M(glycopeptide) = calculated M(core peptide identified) + M(Hex)*i + M(HexNAc)*j + M(dHex)*k + M(NeuAc)*l (where M is a mass value, and i, j, k, and l are integers).

Glyco-RIDGE analysis using GRable

Site-specific glycoform analysis of hAGP as a model glycoprotein was performed (Figure 1). Briefly, hAGP was reduced, alkylated, and digested with trypsin and Lys-C endopeptidase. The digest was used in hydrophilic interaction chromatography (HILIC) on an

Amide-80 column to capture glycopeptides. A small aliquot of the glycopeptide fraction was treated with peptide-*N*-glycosidase F (PNGase F) in ^{18}O-labeled water to remove *N*-glycans and to label deglycosylated Asn as ^{18}O-labeled Asp (IGOT treatment) (30, 31). The IGOT-treated deglycopeptides were analyzed using LC/MS/MS, and the MS2 data were used for database search using Mascot to prepare a core peptide list. In parallel, another aliquot of the glycopeptide fraction was analyzed through LC/MS/MS. MS1 data of the glycopeptide analysis underwent GRable analysis. GRable used the core peptide (Data S1) and glycan composition (Data S2) lists to assign each glycopeptide signal, whereas MS2 data are used for adding the confidence level to the GRable results. Glycopeptide analysis MS2 data was also used for glycopeptide estimation using Byonic (7). The resulting glycopeptide lists of GRable and Byonic analyses were compared. The detailed procedures of each step are described below.

Preparation of hAGP glycopeptide samples

To prepare large-scale analytes for repeated verification, 5 mg of freeze-dried hAGP (Sigma-Aldrich, St.Louis, MO, USA) was dissolved in Milli-Q water (500 μL) and the aliquot was denatured with 0.1% Rapigest (Waters, Milford, MA, USA) in 50 mM Tris-HCl, reduced with dithiothreitol (Fujifilm Wako, Osaka, Japan), alkylated with iodoacetamide (Fujifilm Wako), and digested with Lys-C (1/200 weight of protein; Fujifilm Wako) and sequencing grade modified trypsin (1/100 weight of protein; Promega, Madison, WI, USA) at 37 °C overnight in 0.1 % Rapigest. After digestion, Rapigest was cleaved by acidification with 0.1% trifluoroacetic acid (TFA; Fujifilm Wako). Glycopeptides of hAGP tryptic digest were then captured using an

Amide-80 column (TSK-gel Amide-80, 2 × 50 mm; TOSOH, Tokyo, Japan) equilibrated with 75% acetonitrile (Merck, Darmstadt, Germany)/0.1% TFA and eluted isocratically with 50 % acetonitrile/0.1% TFA after washing.

LC/MS/MS analysis of hAGP glycopeptide samples

Glycopeptides were analyzed using an LC-electrospray ionization-MS system equipped with nanoflow LC (UltiMate-3000; Thermo Fisher Scientific, Waltham, MA, USA) and a tandem mass spectrometer (Orbitrap Fusion Tribrid mass spectrometer; Thermo Fisher Scientific). Glycopeptides were trapped on a C18 cartridge column (Acclaim PepMap100 C18, 0.3 mm I.D. × 5 mm; Thermo Fisher Scientific) and separated on a C18 tip column (75 μm I.D. × 15 cm, 3 μm particle; Nikkyo Technos, Tokyo, Japan) using 2–36% acetonitrile/0.1% formic acid (Fujifilm Wako) gradient (90 min) at a flow rate of 300 nL/min. The ionization voltage was 2.0 kV (positive), and the temperature of the ion transfer tube was 275 °C. Data was acquired using the data-dependent mode (cycle time: 2.5 s) with internal mass calibration (Lock mass) at 445.12003. The mass range was 377–2,000 m/z (MS1) and 135–2,000 m/z (MS2). The mass resolution was 120,000 (MS1, orbitrap) and 15,000 (MS2, orbitrap). The fragmentation mode was high-energy collision-induced dissociation with a stepped collision energy of 25, 30, and 35%. The acquired raw data file was deconvoluted with Xtract of Proteome Discoverer (ver. 2.4, Thermo Fisher Scientific) and exported as an mzML file that was used for GRable analysis.

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

For core peptide identification, IGOT-LC/MS analysis was performed as reported previously (30, 31). Glycopeptides were treated with PNGase F (Takara Bio, Shiga, Japan) in 50 mM Tris-HCl, pH 8.5, prepared with $H_2{}^{18}O$ (isotope purity: 95%; Taiyo Nippon Sanso, Tokyo, Japan) at 37 °C overnight. The deglycosylated peptides (IGOT peptides) were analyzed before the glycopeptide analysis through LC/MS/MS under the same conditions as the glycopeptide sample.

Preparation of a core peptide list

Obtaining RT information is essential when the signal intensity of the precursor is the highest when preparing a core peptide list. In addition, the core peptides should be sorted in descending signal intensity order (total ion current of MS2 of precursor ions recorded in a Mascot generic format (mgf) file). The Glyco-RIDGE analysis identifies glycopeptides and core peptides combinations by calculating the mass differences within the limited RT range. Therefore, the fact that the core peptides are present in the sample is crucial and should be accurately identified. Currently, using stable isotope labeling is recommended but not essential. A list of the expected core peptides can be applied if the sample protein is highly purified. The RT of the core peptide is important for selection but is unavailable when core peptides are predicted. Presently, the RT of peptides can be predicted, which will be useful for selection. The format of the list required by the software is fixed and must be created similarly with the list used in this study (Data S1).

For hAGP, acquired raw data was converted to mgf using Mascot Distiller (ver. 2.7.0; Matrix Science, Boston, MA, USA) and searched using Mascot (ver. 2.5.1; Matrix Science) using the human protein sequence file of SwissProt_UniProtKB_isoform (downloaded on April, 2019; entry: 42,431). The search conditions were as follows: enzyme: trypsin (full or semi); missed cleavage: 2; mass tolerance: 7 ppm (MS1) and 0.02 Da (MS2); fixed modification: carbamidomethyl (C); variable modifications: Gln to pyro-Glu (peptide N-term, Q), ammonia-loss (peptide N-term, carbamidomethyl C), oxidation (M), and Delta: H(-1)N(-1)18O(1)(N); and target false discovery rate: 1%. Mascot search results with a peptide rank of 1 and peptide expectation value of <0.05 were selected. Matched sequences containing IGOT modification at Asn on the consensus sequence for *N*-glycosylation (Asn-Xaa-[Ser/Thr]; Xaa is not any Pro) obtained with trypsin (full) and trypsin (semi) were combined. The resultant glycopeptide list was used to create a core peptide list for Grable analysis. hAGP peptide sequences that were not identified but presumed were manually added to the list (Data S1); four core peptides with predicted hAGP sequences were added to the list, and three core peptides were matched with glycopeptides as clusters.

Preparation of a glycan composition list

A glycan point list required for the Matching step was prepared as described below. Glycan compositions used for matching are determined by the number of each component in the setting. The default setting is Hex: 0–12, HexNAc: 1–12, dHex: 0–4, and NeuAc: 0–4. Among the possible compositions, many ones are considered unlikely based on the biosynthetic pathways.

If they matched, the combination is considered incorrect. In the glycan list, unusual compositions are predetermined in advance, and such wrong matches are marked in the match results, which will be considered in the selection. The approximate mass of ionizable glycopeptides determines the maximum number of Hex and HexNAc. The maximum number of NeuAc is limited to four because our sample preparation conditions did not allow the remaining oligo- or polysialic acids. The number of dHex (fucose) is also limited to four or less because the mass of dHex(5) is similar to that of Hex(2)HexNAc(2) (difference = 0.025), corresponding to approximately 5 ppm of 5,000 Da. If the measurement is performed with an accuracy higher than 2 ppm, it may be distinguishable for five or more fucoses. The format of the glycan point lists is also fixed (Data S2). The list used in this study can be used to analyze human cell-derived protein samples and is not considered for NeuGc. In this list, 139 glycan compositions were given 1 point if the composition is matched. The list contained glycan compositions in *N*-glycan major biosynthetic pathways, from the glycan that was produced by oligosaccharide transferase to the glycan that was processed by many glycosyltransferases and glycosidases. The largest one is Hex(7)HexNAc(6)dHex(4)NeuAc(4) corresponding to tetra-antennary glycan attaching four Fuc and four NeuAc, and the smallest is HexNAc(1), a breakdown product remaining in GlcNAc on Asn. Unusual compositions are determined considering *N*-glycan biosynthetic pathways.

This study considered four glycan units, including Hex, HexNAc, dHex, and NeuAc, as glycan components attached to hAGP. *N*-glycan compositions were expressed as "the number of Hex (-3), HexNAc (-2), dHex – NeuAc" after subtracting the composition corresponding to the trimannosyl core of *N*-glycans (e.g., 440-0 means Hex=6, HexNAc=7, dHex=0, and NeuAc=0). A glycan composition list for hAGP was almost identical to the human *N*-glycome list supplied

by Byonic, except for one composition (440-0) (Table S1), which was manually added because the composition was considered common for human *N*-glycome. In the glycan composition list, unusual compositions were defined based on *N*-glycan biosynthetic pathways, but negative points were not given for these unusual compositions in this study.

Byonic search for hAGP data

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

GRable analysis of HL-60 cell data

The LC/MS/MS data of HL-60 cell lysates were obtained in a previous study (Figure S1) (23). A protease digest of released HL-60 cell lysate was prepared similarly to hAGP. An

aliquot of the digest was acidified and heated to remove sialic acid, decreasing glycan heterogeneity and increasing the relative abundance of each glycopeptide. Notably, Glyco-RIDGE analysis after desialylation is still effective. Removing sialic acids helps detect minor glycopeptides because sialylation increases glycan heterogeneity and decreases their ionization efficiency, lowering the overall sensitivity of sialylated glycopeptides. An aliquot of the desialylated glycopeptide fractions was subjected to HILIC to collect four glycopeptide fractions and then analyzed using IGOT-LC/MS/MS and LC/MS/MS to obtain deglycopeptide and glycopeptide MS data, respectively. The glycans obtained in the IGOT procedure were analyzed using matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) MS to prepare a glycan composition list. Since the HL-60 cell data is for the desialylated glycopeptides, only three glycan units, including Hex, HexNAc, and dHex, were considered glycan components. The definition of unusual compositions was the same as for hAGP. The resulting core peptide list (Data S3) and glycan point list (Data S4) for Fraction 2 were used for GRable analysis.

Software implementation

We implemented GRable as a web application. The user interface was written mostly in JavaScript, data management in Java, and scientific calculations in Python. PostgreSQL was used in the background for user and data file management. The software was developed and tested on Ubuntu Linux 22.04 LTS.

RESULTS AND DISCUSSION

GRable is a web application that users can access through a web browser. This software proceeds along seven steps (Figure 2). Each step except for Step 2 is executed step-by-step after uploading the required data and setting appropriate parameters. The results of Steps 3 to 5 can be visually confirmed using a viewer in the main window of the software. Detailed Steps 4 to 7 results can be exported as an Excel file with each setting, as shown for hAGP (Data S5 and S6) and HL-60 lysates (Data S7 and S8). A detailed instruction manual for GRable is described in Document S1. The data processing, conditions, and results of each step are described below, primarily focusing on improving the prototype.

Monoisotopic peak picking (Step 4) with improved processing speed and accuracy

In a previous in-house prototype, the search for monoisotopic peaks started from the highest signal, which was time-consuming. Therefore, GRable used a new algorism to improve data processing speed. First, a filter of five scans × 5 Da was used to search a local peak for each "island" comprising a single ion. When the highest signal is centered in the filter, the signal is recorded as a local peak and other signals within the filter are excluded from the local peak candidates. The local peaks were used as a starting point for finding corresponding islands comprising the same ion. The spectrum of a local peak signal was integrated with the same spectra of the ion before and after scans, and the resulting spectrum was used to find the monoisotopic signal. This algorism change greatly speeds up the monoisotopic peak-picking process.

In addition, an algorism for finding monoisotopic signals was also modified to improve the accuracy of the monoisotopic assignment in GRable. The prototype used a monoisotopic signal assignment method using criteria based on the relative intensity of isotope signals; however, the miss-assignment of monoisotopic ions was frequently observed. Accordingly, GRable has a function that evaluates the picked monoisotopic peaks using the multinomial distribution based on abundance ratios of four isotopes, including C, N, O, and S. By fitting the observed spectrum with the calculated spectrum, the need for correcting the monoisotopic signal obtained by the preceding step is suggested. This correction function facilitated the increased number of glycopeptide group members matched compared to the GRable results without using this function. For hAGP, 5,369 monoisotopic signals were detected, of which 2,140 signals were automatically corrected by the correction function (Data S6).

Clustering (Step 5) for efficiently and comprehensively detecting glycopeptide signals

The prototype could set a single RT range to find the glycopeptide group as a cluster because the time difference was considered for extending neutral monosaccharides, namely Hex, HexNAc, and dHex. Glycopeptide groups with different numbers of acidic saccharides, such as NeuAc, were found as clusters (25) because adding neutral saccharides slightly shortens the RT, whereas adding acidic saccharides delayed the RT. With this previous RT setting, as the acidic shift increases in proportion to the number of sialic acid units, four glycopeptide clusters having asialo-, monosialo-, disialo-, and trisialo-glycans, respectively, were found separately in a representative hAGP cluster of (Figure 3A). Therefore, glycopeptide groups having the same core

but different numbers of acidic units could not be combined. To improve this point, GRable was constructed to allow setting the RT difference separately for each acidic and neutral saccharide and any motifs such as Hex + HexNAc (LacNAc). In the current version, the four clusters with the same core peptide were successfully connected into a single cluster, since a range of RT shifts for each unit could be set separately (Figure 3B). This improvement facilitates a better understanding of the micro-heterogeneity of one glycosite by visualizing the signal intensity of each glycopeptide plot. Notably, a new member with tetrasialo-glycans could also be detected. Such highly sialylated members could not be detected as a cluster because the variety of the glycan stem portion with four sialylations is limited. Accordingly, this improvement facilitates the increased detection number of glycan micro-heterogeneity for one glycosite.

In addition, GRable can integrate multiple clustering results obtained in parallel under different conditions for a more comprehensive detection of glycopeptide clusters. Since the clustering results of each processing step are combined into a single cluster, each condition setting is independent of the order of processing. Using this "parallel clustering" function, the previous setting (Condition 1; Figure 3A) and current setting (Condition 2; Figure 3B) results could be obtained simultaneously. In addition, another setting can be set up to five conditions at one execution. We often use a condition that considers only the number of Hex to find clusters, aiming to detect glycopeptide clusters only with high- or oligo-mannosylated glycans. Many possible compositions of complex- and hybrid-type glycans exist; however, many oligo-mannosylated glycans are limited to M9–M5. Thus, the minimum cluster number was lowered to three in this setting to find oligo-mannosylated glycan carriers. In the hAGP samples, the abundance of oligo-mannosylated glycan is quite low, resulting in no detection with this setting (Condition 3; Figure

3C). This finding is consistent with previous reports that glycan structures attached to hAGP are branched, sialylated, fucosylated at the branches, and slightly extended with polylactosamine (32–35).

The individual time setting for each type of glycan unit and the parallel clustering with different settings effectively increase the detection sensitivity, and multiple clusters with the same core but with different sialic acid numbers can now be combined into one cluster. With parallel clustering of these three conditions for hAGP, 65 clusters comprising 1,422 types of glycopeptides were detected (Figure 3D). The average number of members in a cluster was 19.1. Among the detected monoisotopic signals, 2,309 signals were assigned in clusters.

To evaluate the usefulness of parallel clustering function to detect glycopeptides with a specific glycan structure, such as oligo-mannosylated glycopeptides, we used a data set of HL-60 cell lysates (HILIC fraction 2) as a complex mixture of unknown proteins containing glycopeptides decorated with oligo-mannose (23). Since the sample glycopeptides used were desialylated, parallel clustering for this sample was performed under two conditions, i.e., using Condition 1 (previous setting) and Condition 3 (oligo-mannosylated glycans) for the hAGP sample. Figure 4A shows clusters detected at Condition 1, where the minimum cluster number was set to four. When the Condition 2 setting (only Hex was considered a glycan unit, and the minimum number of a cluster was three) was added to the Condition 1 search, the number of detected clusters increased (Figure 4B). For clarity, only increased clusters (No. of members: 3) are visualized (Figure 4C), demonstrating the presence of oligo-mannosylated glycopeptides. Thus, GRable applies to large-scale and in-depth analyses of glycoproteins, in which the

comprehensiveness of glycan heterogeneity was improved, and the glycan structure-specific search was allowed.

Matching (Step 6) with improved evaluation of analysis

The previous Glyco-RIDGE analysis had the limitations of a low match rate. The mass calculation assumes that peptides are ionized with protons such as MH^+. Some clusters with no appropriate match were clusters of non-proton adducts. Accordingly, GRable was equipped with an inter-cluster analysis function in the Matching step to detect clusters comprising glycopeptides ionized with a non-proton cation such as iron (Fe^{3+}) and ammonium (NH^{4+}). In the matching results for hAGP using the lists of 56 core peptides (Data S1) and 134 glycan compositions (Data S2), 65 clusters were detected. Among them, 46 clusters were matched; however, 19 were 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, and thus, reducing them during sample preparation and analysis is essential. The remaining unmatched clusters may be attributed to the lack of core peptides in the list, and therefore, the inter-cluster analysis function provides a clue for how the list should be improved.

The result sheet shows information on the three matched components (i.e., glycopeptides, peptides, and glycans) and differences from the calculated mass values and RT difference for each cluster. All matched results that meet the criteria can be provided for a single

cluster in this step. Therefore, the most plausible matching is selected and evaluated in the subsequent step. For this selection, positive points (and negative points for unusual compositions as an option) are given to the glycan compositions matched with the list, and the total score is calculated for each cluster as the sum of the points of all members.

Selection (Step 7) with improved results interpretation by using MS2 information

This last step is crucial because selecting the most plausible matching is essential in guaranteeing the certainty of the Glyco-RIDGE analysis results, considering that the false discovery rate cannot be defined. Since this selection step has been conducted with manual inspections in our previous procedure, using this step with clear criteria in GRable is a large technical improvement. As described above, a crucial criterion is the total score of each matched cluster. Accordingly, this step sets the total score threshold, and only the matchings with the threshold are shown in the Selection result sheet. When the Clustering and Matching steps are performed correctly, all members will match any compositions, and many will get points based on the glycan point list. In addition to the total score, additional criteria are the delta mass and delta RT between the glycopeptide and the corresponding core peptide. Thus, all the members within these criteria are highlighted in the result sheet, visually confirming the certainty of the results. A commercial hAGP glycoprotein specimen used here had serum glycoproteins contamination, such as haptoglobin (HPT), in the core peptide list (Data 1). In this situation, for cluster 1, two candidate core peptides of hAGP isoform 1 (A1AG1) and HPT were used. The total scores of A1AG1 and HPT are 28 and 11, respectively. However, the delta RTs of all members for HPT

are over 15 min, whereas the delta RT of A1AG1 members is within -6.5–5.1 min, meeting the defined criteria. Accordingly, A1AG1 can be selected as the core peptide for cluster 1. Similarly, the most plausible matching can be easily selected for all the matched clusters.

As the other important improved point, this selection step allows evaluating the "confidence level" of the selected results by using MS2 information. This scrutinizing and curating step for reducing misassignment and ambiguousness is critical in current N-glycoproteomics (36). Therefore, as with many MS2-based glycopeptide identification methods (13), GRable first searches glycan fragment ions called "diagnostic ions" in MS2 spectra for cluster members to evaluate whether the spectrum is attributed to a glycopeptide. Dagnostic ions such as fragments of HexNAc and HexNAc + Hex are searched as defaults. The variation of diagnostic ions can be set without restricting the number. In addition, glycopeptide fragments such as Y0, Y1, and Y2 are searched to estimate the mass of Y0, namely peptide moiety. The matched peptide is correct if the mass value coincides with that of matched peptides. This MS2 information is summarized for the selected clusters in the "MS2 info for Clusters" sheet, and all the searching results is available in the "All MS2 info" sheet of an exported file (Data S6). The presence of such MS2 information is visible in the "Selection results" sheet and accessible directly via an intra-file link for all glycopeptides in the sheet. In the case of hAGP cluster 1, 45 peaks were accompanied with MS2 information among 78 monoisotopic peaks, and its predicted peptide sequence is a A1AG1-derived peptide (WFYIASAFRNEEYNK). Therefore, the selection result for this cluster was confirmed.with this MS2 information.

In this context, the Glyco-RIDGE results were presented with the "confidence level" using GRable. Thus, there are ranks of the probability/confidence of the assignment of core peptides; namely, 1) "High"; clusters having any member(s) showing Y0-related ions corresponding to the presumed core, 2) "High"; clusters having any member(s) identified as the same core using the MS2-based search engine, 3) "Medium"; clusters having any member(s) showing glycan diagnostic ions, 4) "Low"”; clusters without any MS2 support. Originally, GRable was a tool designed based on how to estimate signals that MS2 spectra cannot identify. Therefore, cluster members without MS2 support and clusters with no MS2 data for all members are notable features of GRable. We do not believe that the confidence of the assignment in the “Low” rank is high when their mass and RT differences fit under the threshold and the glycan point is the highest in the cluster. Considering the percentage of agreement between the identification results from MS2 and the GR assignments, it is thought that the certainty of the assignment of points without MS2 data is no different from that with MS2; however, individual validation is required. Therefore, developing an evaluation method for the certainty of low-rank results is critical in further sophistication of the GRable-assisted Glyco-RIDGE analysis.

Improved in-depth glycoproteomics results using GRable

Two isoforms exist for hAGP: isoform 1 (A1AG1) and isoform 2 (A1AG2), of which A1AG1 is the major component. According to the improved GRable procedures, five glycoproteins, including HPT, plasma protease C1 inhibitor, and lymphatic vessel endothelial hyaluronic acid receptor 1 as serum-derived contaminants, were assigned for hAGP samples

(Table S2). Among 14 assigned clusters, seven (1, 4, 13, 23, 26, 43, and 48) and six (1, 3, 6, 11, 19, and 23) were assigned to A1AG1 and A1AG2, respectively. These clusters covered all five *N*-glycosylation potential sites of A1AG1 and A1AG2 (the two sites are common in A1AG1), except for site 93 of A1AG2; *N*-glycosylation impairment at this site is consistent with a previous report (35). The estimated core peptides for each cluster were the same, estimated using the MS2 information utilization function of GRable and Byonic, confirming the certainty of the results. Each cluster strictly corresponded to one core peptide, and thus, their clustering view facilitates effective glycan macro-heterogeneity visualization for each glycosite (Figure 5). Notably, two *N*-glycosites (sites 33 and 103) were detected as almost the same core peptide except for one residue between the two isoforms, but their site-specific glycoforms could be visualized separately. This result highlighted the usefulness of GRable for elucidating differences in the glycosylation status of target glycoproteins, even with other contaminated glycoproteins.

To evaluate the certainty of the Glyco-RIDGE results obtained using GRable, the results were compared with those of a representative MS2-based software, Byonic. Using GRable, 123 and 108 site-specific glycoforms were estimated for A1AG1 and A1AG2, respectively (Table 1 and Table S3). Notably, GRable could estimate over four-fold more site-specific glycoforms than Byonic and covered most of the site-specific glycoforms estimated by Byonic. Only two of the site-specific glycoforms were estimated by Byonic but not by GRable (0(-1)0-0 and 000-0 at site 56) due to one component being absent (010-0 or 000-1). As presumed, the MS2-based estimation by Byonic succeeded on higher intensity glycopeptides (Data S9 and S10). hAGP had *N*-glycosylations with branched, highly sialylated, fucosylated at branches, and extended with polylactosamine (34, 35). Similar results were obtained with GRable (Table S3); for example,

tetrasialo-glycans were estimated at all five A1AG1 sites, indicating its high sialylation. Glycans with #dHex ≥ 2 were estimated at four A1AG1 sites, indicating its fucosylation at both its core and branches. In addition, glycan components for polylactosamine (i.e., #Hex + #HexNAc ≥ 10) were estimated at four sites, indicating the presence of such extended glycans. Notably, such highly sialylated and polylactosamine-attached glycopeptides failed to be detected by Byonic (Table S3), which may be due to their low ionization efficiency. Thus, these results confirmed the reliability of the Glyco-RIDGE results obtained by GRable, highlighting the usefulness of this method as a complementary role of current gold standard MS2-based glycoproteomics.

CONCLUSION

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

REFERENCES

1. Pascovici, D., Wu, J. X., McKay, M. J., Joseph, C., Noor, Z., Kamath, K., et al. (2018) Clinically relevant post-translational modification analyses—maturing workflows and bioinformatics Tools. *Int J Mol Sci.* **20**, 16

2. Varki, A. (2017) Biological roles of glycans. *Glycobiology* **27**, 3–49

3. Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., et al. eds. (2022) *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press) 4th ed.

4. Čaval, T., Heck, A. J. R., and Reiding, K. R. (2021) Meta-heterogeneity: Evaluating and describing the diversity in glycosylation between sites on the same glycoprotein. *Mol Cell Proteomics* **20**, 100010

5. Fang, P., Ji, Y., Oellerich, T., Urlaub, H., and Pan, K. T. (2022) Strategies for proteome-wide quantification of glycosylation macro- and micro-heterogeneity. *Int J Mol Sci.* **23**, 1609

6. Kawahara, R., Chernykh, A., Alagesan, K., Bern, M., Cao, W., Chalkley, R. J., et al. (2021) Community evaluation of glycoproteomics informatics solutions reveals high-performance search strategies for serum glycopeptide analysis. *Nat Methods* **18**, 1304–1316

7. Bern, M., Kil, Y. J., and Becker, C. (2012) in *Current Protocols in Bioinformatics* (John Wiley & Sons, Inc., Hoboken, NJ, USA), p Unit13.20.

8. Fang, Z., Qin, H., Mao, J., Wang, Z., Zhang, N., Wang, Y., et al. (2022) Glyco-Decipher enables glycan database-independent peptide matching and in-depth characterization of site-specific N-glycosylation. *Nat. Commun.* **13**, 1900

9. He, S.-M., Yang, P.-Y., Zeng, W.-F., Cao, W.-Q., Liu, M.-Q., and Yang, P.-Y. (2021) Precise, fast and comprehensive analysis of intact glycopeptides and modified glycans with pGlyco3. *Nat. Methods* **2021**, 1–9

10. Polasky, D. A., Yu, F., Teo, G. C., and Nesvizhskii, A. I. (2020) Fast and comprehensive N- and O-glycoproteomics analysis with MSFragger-Glyco. *Nat. Methods* **17**, 1125–1132

11. Shen, J., Jia, L., Dang, L., Su, Y., Zhang, J., Xu, et al. (2021) StrucGP: de novo structural sequencing of site-specific N-glycan on glycoproteins using a modularization strategy. *Nat Methods* **18**, 921–929

12. Choo, M. S., Wan, C., Rudd, P. M., and Nguyen-Khuong, T. (2019) GlycopeptideGraphMS: Improved glycopeptide detection and identification by exploiting graph theoretical patterns in mass and retention time. *Anal Chem.* **91**, 7236–7244

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

14. Baker, P. R., Trinidad, J. C., and Chalkley, R. J. (2011) Modification site localization

scoring integrated into a search engine. *Mol Cell Proteomics* **10**, M111.008078

15. Cheng, K., Chen, R., Seebun, D., Ye, M., Figeys, D., and Zou, H. (2014) Large-scale characterization of intact N-glycopeptides using an automated glycoproteomic method. *J Proteomics* **110**, 145–154

16. Toghi Eshghi, S., Shah, P., Yang, W., Li, X., and Zhang, H. (2015) GPQuest: A spectral library matching algorithm for site-specific assignment of tandem mass spectra to intact N-glycopeptides. *Anal Chem.* **87**, 5181–5188

17. Park, G. W., Kim, J. H. J. Y., Hwang, H., Lee, J. Y., Ahn, Y. H., Lee, H. K., et al. (2016) Integrated GlycoProteome Analyzer (I-GPA) for automated identification and quantitation of site-specific N-glycosylation. *Sci Rep.* **6**, 21175

18. Nasir, W., Toledo, A. G., Noborn, F., Nilsson, J., Wang, M., Bandeira, N., and Larson, G. (2016) SweetNET: A bioinformatics workflow for glycopeptide MS/MS spectral analysis. *J Proteome Res.* **15**, 2826–40

19. Pioch, M., Hoffmann, M., Pralow, A., Reichl, U., and Rapp, E. (2018) glyXtool MS: An open-source pipeline for semiautomated analysis of glycopeptide mass spectrometry data. *Anal Chem.* **90**, 11908–11916

20. An, Z., Shu, Q., Lv, H., Shu, L., Wang, J., Yang, F., and Fu, Y. (2018) N-linked glycopeptide identification based on open mass spectral library search. *Biomed Res. Int.* **2018**, 1564136

21. Devabhaktuni, A., Lin, S., Zhang, L., Swaminathan, K., Gonzalez, C. G., Olsson, N., et al. (2019) TagGraph reveals vast protein modification landscapes from large tandem mass spectrometry datasets. *Nat Biotechnol.* **37**, 469–479

22. Narimatsu, H., Kaji, H., Vakhrushev, S. Y., Clausen, H., Zhang, H., Noro, E., et al. (2018) Current technologies for complex glycoproteomics and their applications to biology/disease-driven glycoproteomics. *J Proteome Res.* **17**, 4097–4112

23. Togayachi, A., Tomioka, A., Fujita, M., Sukegawa, M., Noro, E., Takakura, D., et al. (2018) Identification of poly-N-acetyllactosamine-carrying glycoproteins from HL-60 human promyelocytic leukemia cells using a site-specific glycome analysis method, Glyco-RIDGE. *J Am Soc Mass Spectrom.* **29**, 1138–1152

24. Noro, E., Togayachi, A., Sato, T., Tomioka, A., Fujita, M., Sukegawa, M., et al. (2015) Large-scale identification of N-glycan glycoproteins carrying Lewis x and site-specific N-glycan alterations in Fut9 knockout mice. *J Proteome Res.* **14**, 3823–3834

25. Noro, E., Matsuda, A., Kyoutou, T., Sato, T., Tomioka, A., Nagai, M., et al. (2021) N-glycan structures of *Wisteria floribunda* agglutinin-positive Mac2 binding protein in the serum of patients with liver fibrosis. *Glycobiology* **31**, 1268–1278

26. Mitani, Y., Yasuno, R., Kihira, K., Chung, K. M., Mitsuda, N., Kanie, S., et al. (2022) Host-dependent producibility of recombinant *Cypridina noctiluca* luciferase with glycosylation defects. *Front Bioeng Biotechnol.* **10**, 92

27. Sogabe, M., Kojima, S., Kaya, T., Tomioka, A., Kaji, H., Sato, T., et al. (2022) A

sensitive new assay system for serum *Wisteria floribunda* agglutinin-reactive ceruloplasmin that distinguishes ovarian clear cell carcinoma from endometrioma. *Anal Chem.* **94**, 2476–2484

28. Hiono, T., Tomioka, A., Hiroyuki, K., Sasaki, M., Orba, Y., Sawa, H., and Kuno, A. (2021) Combinatorial approach with mass spectrometry and lectin microarray dissected glycoproteomic features of virion-derived spike protein of SARS-CoV-2. *bioRxiv*, doi: https://doi.org/10.1101/2021.04.10.439300

29. Kuno, A., Ikehara, Y., Tanaka, Y., Angata, T., Unno, S., Sogabe, M., et al. (2011) Multilectin assay for detecting fibrosis-specific glyco-alteration by means of lectin microarray. *Clin Chem* **57**, 48–56

30. Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., et al. (2003) Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol.* **21**, 667–72

31. Kaji, H., Yamauchi, Y., Takahashi, N., and Isobe, T. (2006) Mass spectrometric identification of N-linked glycopeptides using lectin-mediated affinity capture and glycosylation site-specific stable isotope tagging. *Nat Protoc.* **1**, 3019–27

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

33. Hevér, H., Nagy, K., Xue, A., Sugár, S., Komka, K., Vékey, K., et al. (2022) Diversity matters: Optimal collision energies for tandem mass spectrometric analysis of a large set of N-glycopeptides. *J Proteome Res.* **21**, 2743–2753

34. Wong, T. L., Mooney, B. P., Cavallero, G. J., Guan, M., Li, L., Zaia, J., et al. (2023) Glycoproteomic analyses of influenza A viruses using timsTOF Pro MS. *J Proteome Res.* **22**, 62–77

35. Keser, T., Tijardović, M., Gornik, I., Lukić, E., Lauc, G., Gornik, O., et al. (2021) High-throughput and site-specific N-glycosylation analysis of human alpha-1-acid glycoprotein offers a great potential for new biomarker discovery. *Mol Cell Proteomics* **20**, 100044

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

TABLE

Table 1. Summary of 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	36 (32)	6 (2)	36 (32)	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)
		123 (98)	27 (2)	108 (84)	26 (2)

[a] *N*-glycosylation sites are indicated in bold. Different residues between two isoforms are underlined. Asterisks indicate the presence of other peptide sequences with different lengths.

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

FIGURES

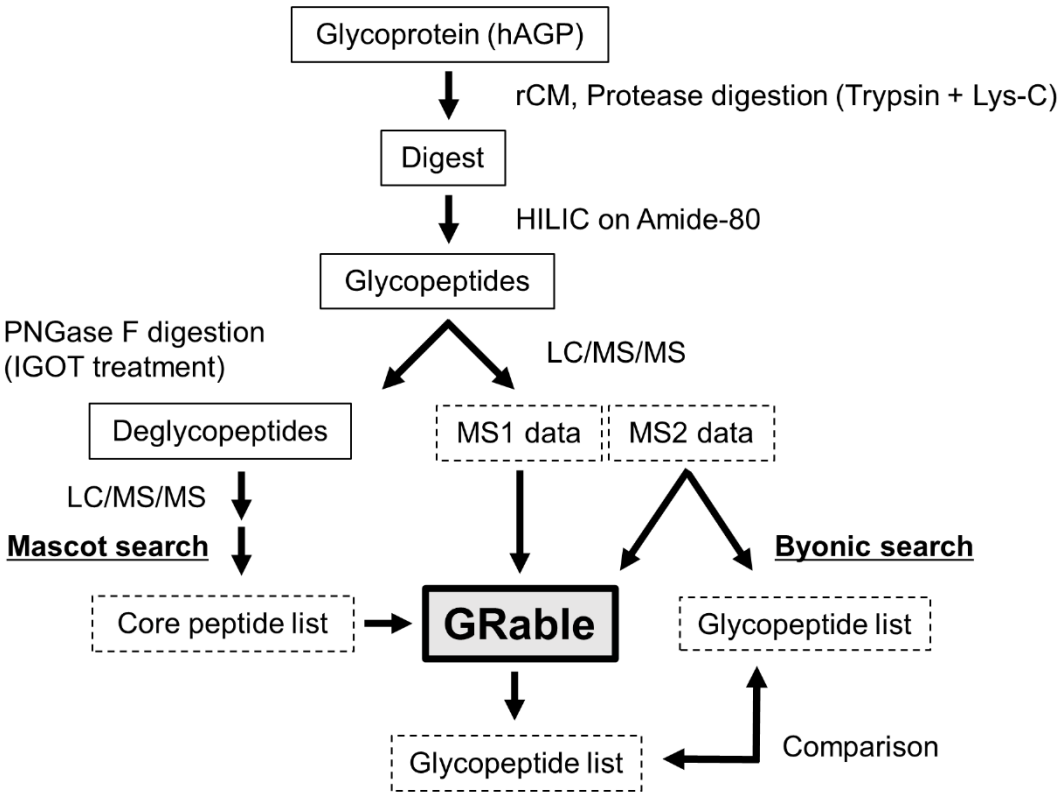

Figure 1. Workflow of glycoproteomic analysis for hAGP using GRable. rCM, reduction and carbamidomethylation; HILIC, hydrophilic interaction liquid chromatography; IGOT, isotope-coded glycosylation site-specific tagging; LC/MS/MS, liquid chromatography-tandem mass spectrometry.

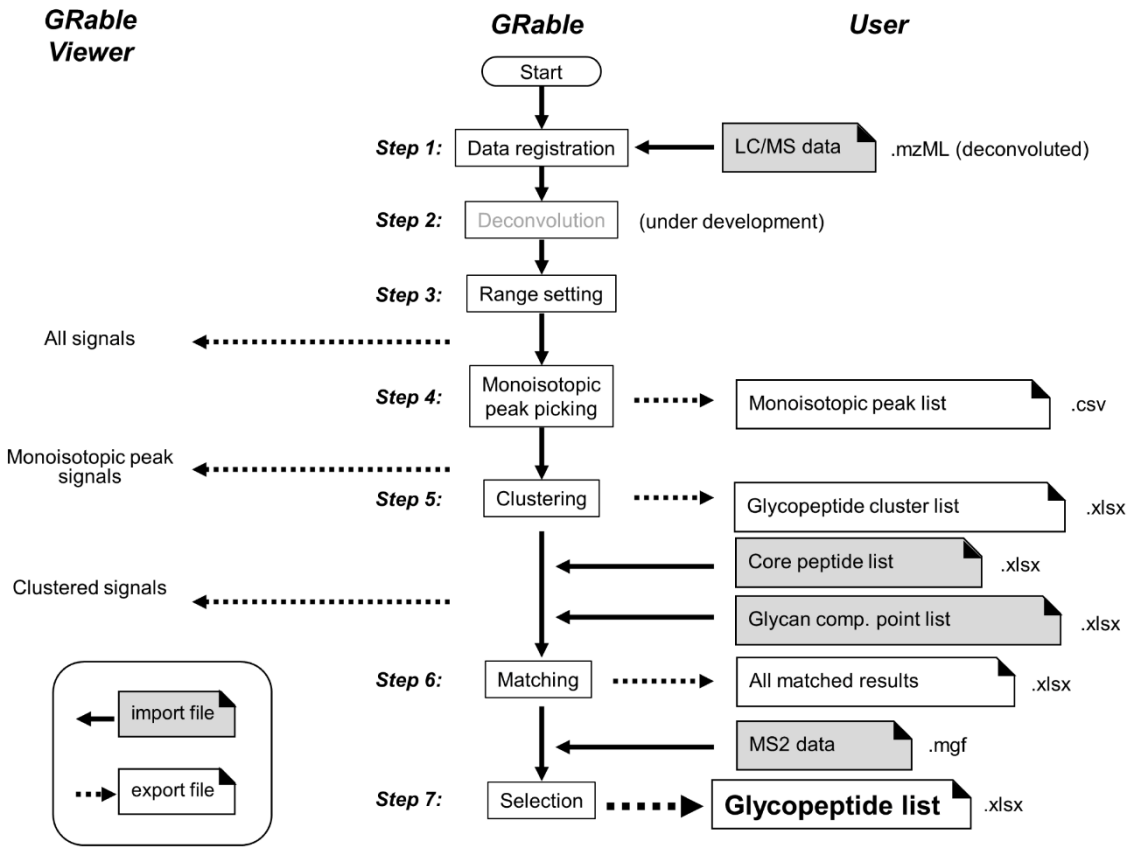

Figure 2. Overview of data processing by GRable

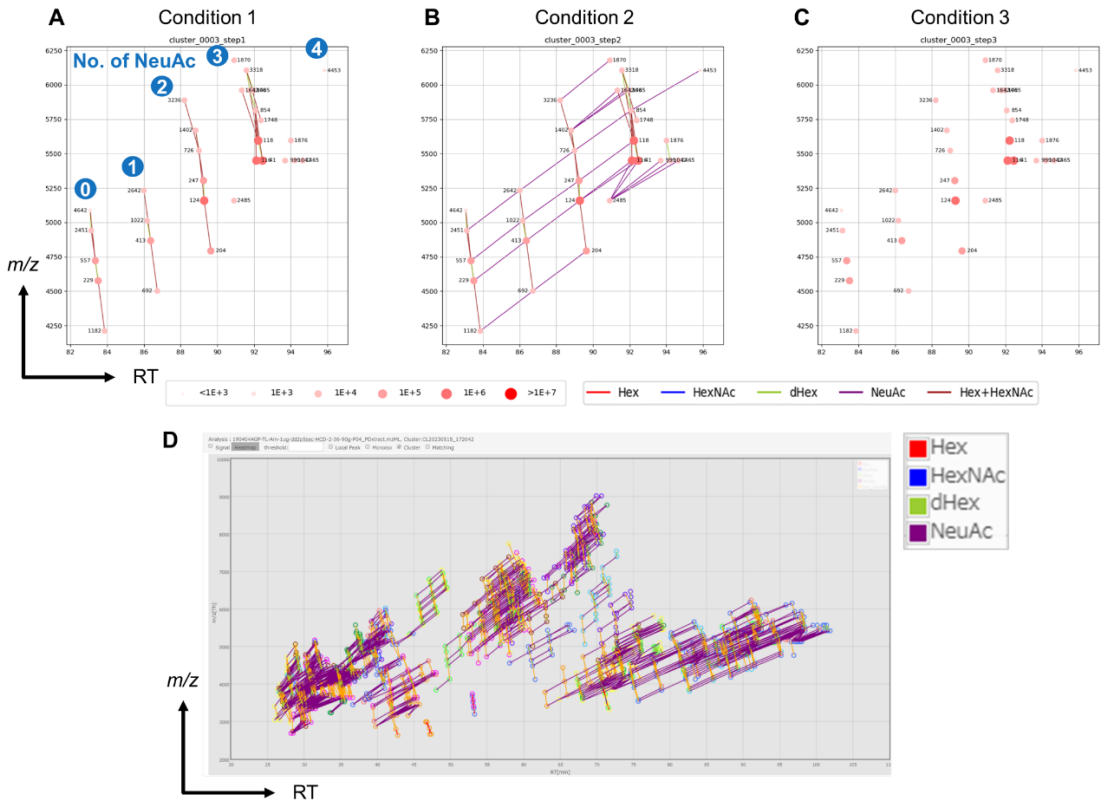

Figure 3. hAGP clustering results. Representative results (Cluster 3) of parallel clustering with three conditions: **(A)** Condition 1 (a setting in our previous prototype), **(B)** Condition 2 (an extra setting for detecting sialylated glycopeptides), and **(C)** Condition 3 (an extra setting for detecting oligo-mannosylated glycopeptides). The conditions are as follows: for Condition 1, neutral saccharides (Hex, HexNAc, dHex, Hex+HexNAc): -1–0 (min), and the minimum number of members: 4; for Condition 2, neutral saccharide: -1–0 (min), acidic saccharide (NeuAc for human): 1–4 (min), and minimum number of members: 10; and for Condition 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 indicate 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.

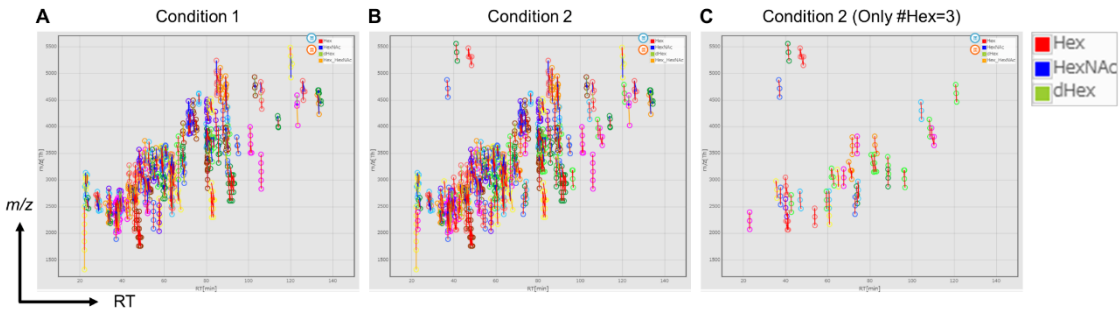

Figure 4. Clustering view of HL-60 cell lysate samples. All the clusters for HL-60 cell lysates (HILIC Fr.2) shown in Data S8 can be visualized in the viewer of the GRable main window. Parallel clustering was performed under two conditions: **(A)** Condition 1 (a setting of the prototype) and (**B**)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

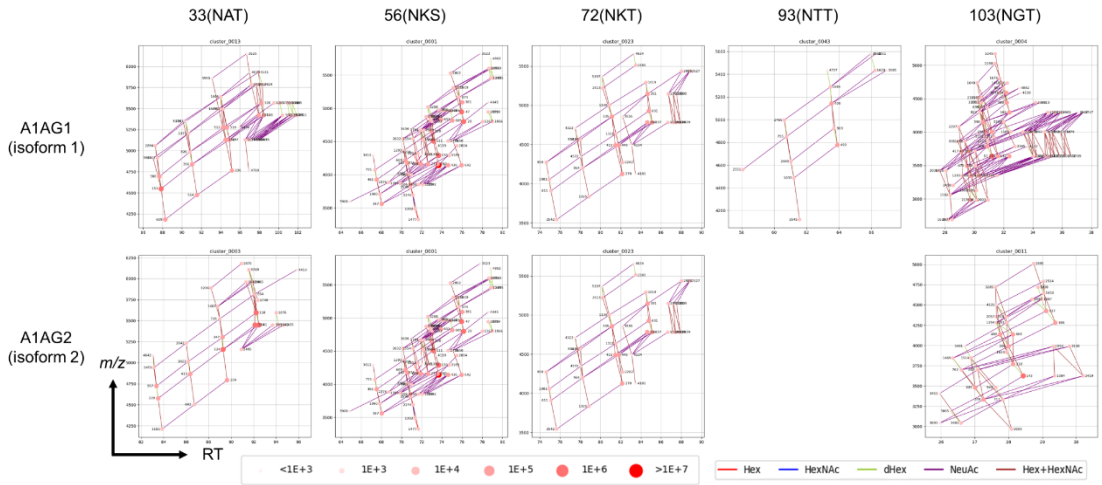

Figure 5. Glycopeptide clusters assigned for 5 *N*-glycosylation sites of hAGP isoforms. Corresponding core peptide sequences for each cluster are indicated in Table 1. Since the core peptide of two isoforms is identical for sites 56 and 72, the same clusters are indicated for both isoforms.