RESEARCH PAPER

Data-independent acquisition mass spectrometry for site-specific glycoproteomics characterization of SARS-CoV-2 spike protein

Deborah Chang [1] · Joshua A. Klein [2] · Mary Rachel Nalehua [2] · William E. Hackett [2] · Joseph Zaia [1,2]

Abstract

The spike protein of SARS-CoV-2, the virus responsible for the global pandemic of COVID-19, is an abundant, heavily glycosylated surface protein that plays a key role in receptor binding and host cell fusion, and is the focus of all current vaccine development efforts. Variants of concern are now circulating worldwide that exhibit mutations in the spike protein. Protein sequence and glycosylation variations of the spike may affect viral fitness, antigenicity, and immune evasion. Global surveillance of the virus currently involves genome sequencing, but tracking emerging variants should include quantitative measurement of changes in site-specific glycosylation as well. In this work, we used data-dependent acquisition (DDA) and data-independent acquisition (DIA) mass spectrometry to quantitatively characterize the five *N*-linked glycosylation sites of the glycoprotein standard alpha-1-acid glycoprotein (AGP), as well as the 22 sites of the SARS-CoV-2 spike protein. We found that DIA compared favorably to DDA in sensitivity, resulting in more assignments of low-abundance glycopeptides. However, the reproducibility across replicates of DIA-identified glycopeptides was lower than that of DDA, possibly due to the difficulty of reliably assigning low-abundance glycopeptides confidently. The differences in the data acquired between the two methods suggest that DIA outperforms DDA in terms of glycoprotein coverage but that overall performance is a balance of sensitivity, selectivity, and statistical confidence in glycoproteomics. We assert that these analytical and bioinformatics methods for assigning and quantifying glycoforms would benefit the process of tracking viral variants as well as for vaccine development.

Keywords Mass spectrometry · DIA · Glycoproteomics · Glycosylation · SARS-CoV-2 · Spike protein

Introduction

Variability in protein glycosylation is an important part of the immune evasion strategies of many enveloped viruses. Enveloped viruses respond to the pressure of the human immune system by changing the antigenicity of their surface spike glycoproteins. While amino acid substitutions are tracked from genetic sequences, there remains very little

information on changes to site-specific glycosylation of spike proteins as viruses evolve. In the example of influenza A virus, amino acid substitutions in the hemagglutinin protein (HA) result in changes to the three-dimensional conformation of HA. These conformational alterations may affect the accessibility of glycosylation machinery enzymes, such as ERmannosidase I and II, resulting in variations in glycoform compositions at certain glycosylation sites [1]. Even subtle changes in the protein sequence can cause measurable changes in glycosylation [2]. Influenza A virus has evolved to tolerate amino acid substitutions in the HA head domain regions that produce the dominant immune responses in humans [3, 4]. Some of these substitutions create or remove *N*-glycosylation sequons [5, 6]. Thus, the alteration of *N*-glycosylation sites is a mechanism whereby the virus balances the need for immune evasion while maintaining fitness [7, 8]. Furthermore, changes in glycosylation may shield antigenic sites on the protein, allowing the virus to escape detection and neutralization by adaptive immune molecules [9–13].

✉ Joseph Zaia
  jzaia@bu.edu

1 Department of Biochemistry, Center for Biomedical Mass Spectrometry, Boston University School of Medicine, Boston University Medical Campus, 670 Albany St., Rm. 509, Boston, MA 02118, USA

2 Boston University Bioinformatics Program, Boston University, Boston, MA, USA

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the novel coronavirus that caused the coronavirus disease 2019 (COVID-19) pandemic. The spike protein of SARS-CoV-2 is a trimeric protein studding the surface of the virus, which is necessary for receptor binding and viral entry, and is the major target for neutralizing antibodies. The virus is continually evolving, as demonstrated by the expanding number of widely circulating variants of interest [14–16], all of which exhibit mutations on the spike protein compared to the wild-type virus. The spike protein is potentially a key player in modulating viral fitness. Indeed, current variants of concern circulating in the USA (alpha, beta, gamma, and delta variants) exhibit spike protein substitutions associated with changes in viral characteristics, such as increased transmission, potentially increased severity, reduced susceptibility to some monoclonal antibody treatments, or less efficient neutralization by antibodies derived from post-vaccination sera or convalescent plasma from patients whose disease did not have these substitutions [15].

The Wuhan strain of the SARS-CoV-2 spike protein contains 22 N-glycosylation sequons and several potential sites of O-glycosylation. Comparisons of SARS-CoV-2 to the SARS-CoV strain that caused the SARS outbreak in 2003 show that there is greater variation in amino acid sequence and more differences in glycan sequons in the S1 subunit than in other parts of the spike protein [17]. The receptor binding domain and known epitopes for neutralizing antibodies are located in the spike S1 subunit, which suggests that glycosylation may shield specific areas of vulnerability [17]. Because the spike protein is the target for all vaccines currently available or in development, it is critical to understand how SARS-CoV-2 will use glycosylation as it evolves in the human population. Detailed, quantitative information on the changes in glycosylation due to strain variation is required to understand the impact of site-specific glycosylation on viral fitness, antigenicity, and immune evasion in the circulating variants of SARS-CoV-2. As such, the global surveillance of emerging SARS-CoV-2 variants should include all structural information of spike, including quantification of the changes in site-specific glycosylation, and not just prediction based on presence of N-glycosylation sequons.

As demonstrated by published data on SARS-CoV-2 spike protein glycosylation [18–21], mass spectrometry methods allow relatively facile assignment of glycosylation of complex glycoproteins. In the present work, we address the need for glycoproteomics methods with the sensitivity and selectivity to quantify glycosylation changes to such glycoproteins to inform rigorous statistical comparisons. The most commonly employed mass spectrometry method for glycoproteomics is data-dependent acquisition (DDA) because it is capable of discovery of new glycopeptide species without prior knowledge; the method is easy to set up and there exist several glycoproteomics bioinformatics tools to perform the analysis

[22–25]. In addition, when combined with various fragmentation methods, such as higher-energy collisional dissociation (HCD), electron transfer dissociation (ETD), or electron transfer/higher-energy collisional dissociation (EThcD), DDA can produce information-rich fragment ion peaks that can provide site localization as well as glycan composition. However, DDA selects for precursor ions in a stochastic manner with a bias toward high-intensity glycopeptides, often resulting in low reproducibility, a problem that is exacerbated as precursor ion abundance decreases [26]. Quantification of all glycopeptide glycoforms present is also difficult because separate glycopeptide glycoforms have overlapping reversed phase chromatographic elution profiles [27, 28]. As an alternative, data-independent acquisition (DIA) acquires tandem mass spectra in a mass window without the need to select precursor ions and takes advantage of fast scan speeds of modern mass spectrometers [29–31]. In SWATH-type DIA [32], the m/z range of interest is divided into windows, typically 25 m/z wide, and all precursor ions eluting in that window are fragmented together. The advantage of DIA is high-reproducibility quantification of peptides that are present in a compound library. For glycopeptides, DIA offers the capability of discovering and quantifying glycopeptides with greater sensitivity and reproducibility than possible using DDA.

With the goal of quantification of N-linked glycopeptides of the SARS-CoV-2 spike protein, we compared the depth of coverage produced by DDA versus DIA and the ability to assign identifications confidently. We demonstrate the advantages and tradeoffs of using DIA for glycoproteomics, taking into consideration the balance of sensitivity, specificity, and statistical power.

Materials and methods

Glycopeptide sample preparation

Recombinant spike protein from SARS-CoV-2 (BetaCoV/Wuhan/IVDC-HB-05/2019), purified from 293 cell culture, with the furin cleavage site removed, was purchased from Immune Technology Corp. (New York, NY; product # IT-002-032p). Alpha-1-acid glycoprotein (AGP), purified from human serum (Sigma-Aldrich, St. Louis, MO; product # G9885), was prepared in parallel with spike and used as a control.

Glycoprotein samples were denatured and reduced with 100 mM ammonium bicarbonate (J.T. Baker, Phillipsburg, NJ), 2-2-2 trifluoroethanol (TFE) (Sigma-Aldrich), and 200 mM dithiothreitol (DTT) (Sigma-Aldrich), for 1 h at 65 °C. We added 200 mM iodoacetamide (Bio-Rad, Hercules, CA), incubating for 1 h at room temperature in the dark to alkylate the cysteine residues. After alkylation, additional DTT was added to quench the iodoacetamide, 1 h at

room temperature in the dark. The amount of TFE in each tube was diluted to 5% by volume adding a 3:1 solution of water/ 100 mM ammonium bicarbonate. Three equal aliquots of the spike protein solution were digested each with alpha-lytic protease (aLP) (New England Biolabs, Ipswich, MA), sequencing-grade chymotrypsin, and trypsin (Promega, Madison, WI) at an enzyme/substrate ratio of 1:20. Two aliquots of AGP were digested in parallel with chymotrypsin and trypsin. Digestion solutions were incubated overnight at 37 °C, after which the enzymes were inactivated by heating the samples for 10 min at 95 °C.

A small aliquot from each spike protein digestion was reserved for deglycosylation, and the remainders were enriched using iSPE-HILIC solid-phase extraction columns (HILICON, Umeå, Sweden), as follows. (For AGP, the entirety of the tryptic and chymotryptic digestions were HILIC-enriched, and the aliquots meant for deglycosylation were reserved subsequent to the following HILIC enrichment steps.) Samples were evaporated to dryness, and the HILIC columns were conditioned with water, then acetonitrile, and equilibrated with 80% acetonitrile/1% trifluoroacetic acid (TFA). Immediately before loading onto HILIC SPE columns, samples were resuspended in loading buffer (80% acetonitrile/ 1% TFA). Columns were washed in sample loading buffer and then eluted with 1% acetonitrile/0.1% formic acid. Eluates were evaporated to dryness and then cleaned using PepClean C18 spin columns (Pierce Biotechnology, Rockford, IL). These enriched glycopeptide samples were used for glycoproteomics LC-MS/MS analysis.

The reserved aliquots (spike protein that was not enriched for glycopeptides and glycopeptide-enriched AGP) were incubated with 500 units of PNGase F (New England Biolabs) for every 20 μg of glycoprotein overnight at 37 °C to remove *N*-glycans. Deglycosylated samples were cleaned using PepClean C18 spin columns; the eluate containing the deglycosylated peptides were used for proteomics LC-MS/MS analysis.

Liquid chromatography–tandem mass spectrometry acquisition

All acquisitions were performed on a Q Exactive HF (Thermo-Fisher, San Jose, CA) mass spectrometer in positive mode, with a nanoAcquity UPLC system, nanoAcquity NPLC Symmetry C18 trap column and ACQUITY UPLC Peptide BEH C18 analytical column (Waters, Milford, MA), and a Triversa Nanomate source (Advion, Ithaca, NY). The column heater was set to 37 °C. Mobile phase A was 1% acetonitrile/0.1% formic acid, and mobile phase B was 99% acetonitrile/0.1% formic acid. Peptides were trapped for 4 min at 4 μL/min in A and then separated with the gradient 0–3 min: 2–5% B, 3–93 min: 5–40% B, 93–98 min: 40% B, 98–100 min: 40–98% B, 100–105 min: 98% B, 105–106 min: 98–2% B, and 106–120 min: 2% B, at flow rate 0.5 μL/min. We ran 3 DDA and 3 DIA replicates for all AGP samples. We ran 3 DDA replicates and 4 DIA replicates for all spike samples. We ran 2 replicates for all deglycosylated peptide samples. Replicates of all samples and acquisition methods were queued in randomized order to minimize any batch effect bias.

DDA acquisition was performed with the following instrument parameters. MS1 spectra were acquired at 60,000 resolution at *m/z* 400, scan range *m/z* 350–1800, 1 microscan per spectrum, AGC target of 3e6 (1e6 for deglycosylated peptides), and maximum injection time 64 ms. Tandem mass spectra were acquired at 30,000 resolution at *m/z* 400, 2 microscans per spectrum, AGC target 2e5 (1e6 for deglycosylated peptides), maximum injection time 64 ms, isolation window 2.0 *m/z*, isolation offset 0.4 *m/z*, fixed first mass of *m/z* 110, and NCE 35 (NCE 27 for deglycosylated peptides). Intensity threshold was 3.1e5, unassigned charge states and charge state 1 were excluded, and dynamic exclusion was 20 s. The 20 most abundant precursor ions per scan were fragmented. Profile data were recorded for both MS1 and MS2 scans.

DIA acquisition was performed with the following instrument parameters. MS1 spectra were acquired at 30,000 resolution at *m/z* 400, scan range *m/z* 350–1800, 1 microscan per spectrum, AGC target of 3e6, and maximum injection time 32 ms. For MS2 acquisition, the range of *m/z* 800–1600 was divided into 50 non-overlapping windows, each with an isolation width of *m/z* 16. Resolution was 30,000 at *m/z* 400, and we acquired 2 microscans per spectrum, AGC target 1e6, maximum injection time 32 ms, and NCE 35. Profile and centroid data were recorded for both MS1 and MS2 scans.

Data analysis

All raw files were first converted to mzML format [33] using MSConvert [34] from ProteoWizard version 3.0.11252 with no additional filters.

Proteomics data from deglycosylated samples were processed using the "Peaks PTM Search" function of Peaks Studio 8.5 software (Bioinformatics Solutions, Waterloo, ON) [35]. AGP was searched against the entire Uniprot Swiss-Prot [36] database (release 2017_04) with a valid protein count of 555,085; the species taxonomy was set to *Homo sapiens*. The database for SARS-CoV-2 spike protein was created by concatenating the spike protein sequence [37] (Fig. S1) with the human proteome (UP000005640 9606; release 2017_04). This database had a valid protein count of 21,043 entries. We searched technical replicates together. For tryptic and chymotryptic digestions, trypsin or chymotrypsin was specified from the available Peaks Studio enzyme menu as the proteolytic enzyme. For the aLP digestion, cleavage sites after Thr, Ala, Ser, and Val were specified. For all digestions, we allowed a

maximum of 3 missed cleavages and one non-specific cleavage. Cysteine carbamidomethylation and methionine oxidation were specified as fixed and variable modifications, respectively. The precursor ion (MS1) mass error tolerance was set to 10 ppm and fragment ion (MS/MS) error tolerance to 0.02 Da. We required a minimum of two unique peptides for protein identification. Default parameters were used for setting the threshold score for accepting individual spectra. The target-decoy false discovery threshold was set at 1% at the peptide level. Peaks Studio search results were exported in mzIdentML [38] format.

Both DDA and DIA raw files for all intact HILIC-enriched glycopeptides were processed using the GlycReSoft search engine, a complete open-source software package that assigns glycopeptides from tandem mass spectra [22].

Mixture spectrum splitting GlycReSoft preprocessing was accomplished in the following way for DIA files. Following deisotoping and charge state deconvolution, we limited MS2 spectra with co-isolating precursor ions based on their retention time patterns, using methodology described by DIA-Umpire [39]. Potential precursors were limited to those with an apex RT within 0.6 min of the product apex time, to account for RT shifts due to co-isolated precursors with shared product ions. These precursors were then correlated with the product ions using a Pearson correlation, and up to the top 20 co-isolating precursors were retained. For each precursor matching these criteria, we created a duplicate MS2 spectrum containing information for a single precursor, while retaining all of the product ion information. Default GlycReSoft parameters were used for preprocessing the DDA files.

Glycopeptide search space The glycopeptide search space was built from the proteomics mzIdentML file, exported from the Peaks Studio search results, and a glycan search space presented in Supplemental File 1, which was generated by glypy [40] and consists of biosynthetically feasible human glycan compositions and up to 1 sulfation. The same glycan hypothesis was used for both AGP and spike protein samples. Up to 1 ammonium adduct was considered. When there were two different glycosylation sites on the same peptide, GlycReSoft used a parsimony filter to weight and report the score toward the localization with the fewest missing ions.

Scoring Scoring was accomplished using default GlycReSoft parameters, with a false discovery rate threshold of 0.05 for positive detection. We used a peptide-centric scoring method to evaluate glycopeptide-spectrum matches, building on our previous coverage-weighted binomial model [41]. We masked all peaks matching theoretical peptide+Y ions, not allowing them to participate in scoring. Peptide+Y ions are fragment ions consisting of the peptide backbone with one or more monosaccharide residues attached. Due to the relatively high single collision energy used in these experiments and lack of calibration of collision energy for specific precursor m/z and charge state, we did not observe peptide+Y ions consistently for all glycopeptides, causing otherwise well-matched spectra to score poorly. In cases where more than one precursor ion existed in a precursor isolation window, GlycReSoft attributed the product ions to each of the co-isolated precursors separately.

Glycopeptides that passed the FDR scoring threshold of 0.05 were quantified by aggregating all associated MS1 peak areas deconvoluted by GlycReSoft. Site-specific glycoform identifications from all replicate runs of all enzymatic digestions of a particular sample were combined into one table, displaying abundances for each replicate. If a glycosite was identified in more than one enzymatic digestion, only that set with the highest average abundance across replicates was retained. The presence rate for each glycan composition was determined by dividing the number of replicates in which the glycoform was identified with a GlycReSoft FDR above 0.05 by the total number of replicates acquired. We required that all glycan compositions be identified in at least two replicates to be considered a positive assignment. When there were two different glycosylation sites on a single peptide backbone, GlycReSoft reported the one with more evidence, using a parsimony filter to weight the score toward the localization with the fewest missing ions.

For glycopeptides that were identified in at least two replicates by DDA but not by DIA, we found the DIA tandem mass spectra with the m/z window and retention time where we expected the glycopeptide precursor to have been isolated. We computed the approximate score that these tandem mass spectra would have been assigned by matching the product ion peaks from those spectra against the theoretical product ions for the glycopeptides, and computed the score identically to the way in which GlycReSoft would during a normal database search round.

Results

We sought to make an unbiased comparison of the performance of DDA versus DIA. We therefore made all glycopeptide assignments algorithmically without employing manual interpretation. We used the peptides identified in a proteomics search to calculate the theoretical glycopeptides present in the sample. Supplemental File 2 shows these deglycosylated peptides, including precursor charge, m/z, all modifications observed, peptide identification score, accession numbers, and protein groups corresponding to the identified peptides.

The abundances for each glycan composition, after aggregating the MS1 peak areas of the glycoforms and combining information from the different enzymatic digestions, as well as

the presence rate of each composition across replicates, are tabulated for each glycosylation site in Supplemental File 3. We required that a composition be identified in at least two replicates before being considered a positive assignment; compositions that were only identified in a single replicate are listed separately in Supplemental File 3.

DDA vs. DIA for AGP

We examined the glycoforms from the five glycosylation sites of the A1AG1 isoform of AGP, acquired by both DDA and DIA. In all sites, the number of glycoforms identified by DIA was greater than those by DDA (Table 1).

Figure 1 compares the AGP glycopeptide assignments for DDA versus DIA for all glycosylation sites. We indicated the composition and quantification for all glycoforms assigned by DDA and DIA in the bar plots. Glycan compositions for all glycoforms presented in the bar plots can be found in Supplemental File 4. The presence rate is the proportion of replicates in which a given glycoform was detected. For example, a presence rate of 0.67 means that the glycan composition was observed in two out of three total replicates. The violin plots in Fig. 1 show the presence rate across replicates, with the width indicating the number of identified glycopeptides for each presence rate. For all glycopeptides, DIA resulted in a larger number of assignments than DDA, but DDA had proportionally more assignments that had complete presence across replicates. We required glycan compositions to be identified in at least two replicates to be included in the bar plots; however, the violin presence plots represent all glycoforms including those only identified in a single replicate.

The aggregated MS1 abundance for all precursors from the A1AG1 isoform of AGP and from spike protein assigned in DDA and DIA replicates that met the FDR threshold, including precursors that were only identified in a single replicate, are plotted as a function of retention time in Fig. 2a,b, respectively. Most DDA precursors were within the range of 10^8 through 10^{10}. DIA precursors spanned a larger range of 10^5 through 10^{10}, consistent with the finding that DIA had greater numbers of assigned glycopeptides.

SARS-CoV-2 spike protein identified by DIA

For SARS-CoV-2 spike protein, we assigned glycoforms at all 22 glycosylation sites by DDA and at 20 sites by DIA. Table 2 shows the number of glycoforms identified by both methods.

Figure 3 displays spike protein data for the ten glycosylation sites for which DIA assigned as many or more glycoforms meeting 0.05 FDR threshold compared to DDA (sites 74-NGTK, 122-NATN, 234-NITR, 331-NITN, 343-NATN, 709-NNSI, 717-NFTI, 801-NFSQ, 1074-NFTT, and 1098-NGTH). The eight sites for which DIA underperformed DDA with respect to the number of assigned glycoforms (sites 165-NCTF, 282-NGTI, 603-NTSN, 616-NCTE, 657-NNSY, 1134-NNTV, 1173-NASV, and 1194-NESL) are shown in Fig. 4. Four sites had too few assignments passing FDR scoring requirements to make meaningful conclusions about the performance of DDA or DIA (sites 17-NLTT, 61-NVTW, 149-NKSW, and 1158-NHTS); these data can be found in Fig. S2.

The aggregated MS1 abundance for all precursors from SARS-CoV-2 spike protein in DDA and DIA replicates that met the FDR requirement, including precursors that were only identified in a single replicate, are plotted as a function of retention time in Fig. 2b. Similar to the AGP precursor abundances in Fig. 2a, DDA was able to make assignments for mainly the high-abundance glycopeptides, while DIA precursors spanned a much larger range of abundances. The time that the mass spectrometer spent per scan for DDA and DIA is shown in Fig. 2c. The RawTools package [42] was used to parse the DDA duty cycle, i.e., the time between MS1 scans, for a raw DDA file of a tryptic spike protein sample. The equivalent duty cycle information for DIA was obtained using the ms_deisotope library from GlycReSoft, also from a tryptic spike protein sample. The DDA cycle consisted of one MS1 full scan followed by up to 20 MS2 scans, and RawTools was used to determine the number of MS2 scans actually triggered in each cycle. To calculate the time spent per scan (Fig. 2c), the duty cycle time was divided by 1 (for the MS1 scan) plus the number of MS2 scans triggered. For DIA, there were a total of 51 scans per cycle (one MS1 full scan plus 50 MS2 scans), so the time spent per scan was the duty cycle time divided by 51.

Table 1 Number of glycoforms assigned by DDA and DIA for the A1AG1 isoform of alpha-1-acid glycoprotein (AGP) in at least two replicates. Values in parentheses represent the total number of glycoforms detected, including those that were identified in only a single replicate

A1AG1 glycosylation sites	Glycoforms identified by DDA	Glycoforms identified by DIA
33-NATL	17 (29)	39 (84)
56-NKSV	15 (18)	37 (100)
72-NKTE	17 (27)	59 (145)
93-NTTY	17 (30)	38 (91)
103-NGTI	7 (14)	10 (38)
Total	73 (118)	183 (458)

a 33-NATL

b 56-NKSV

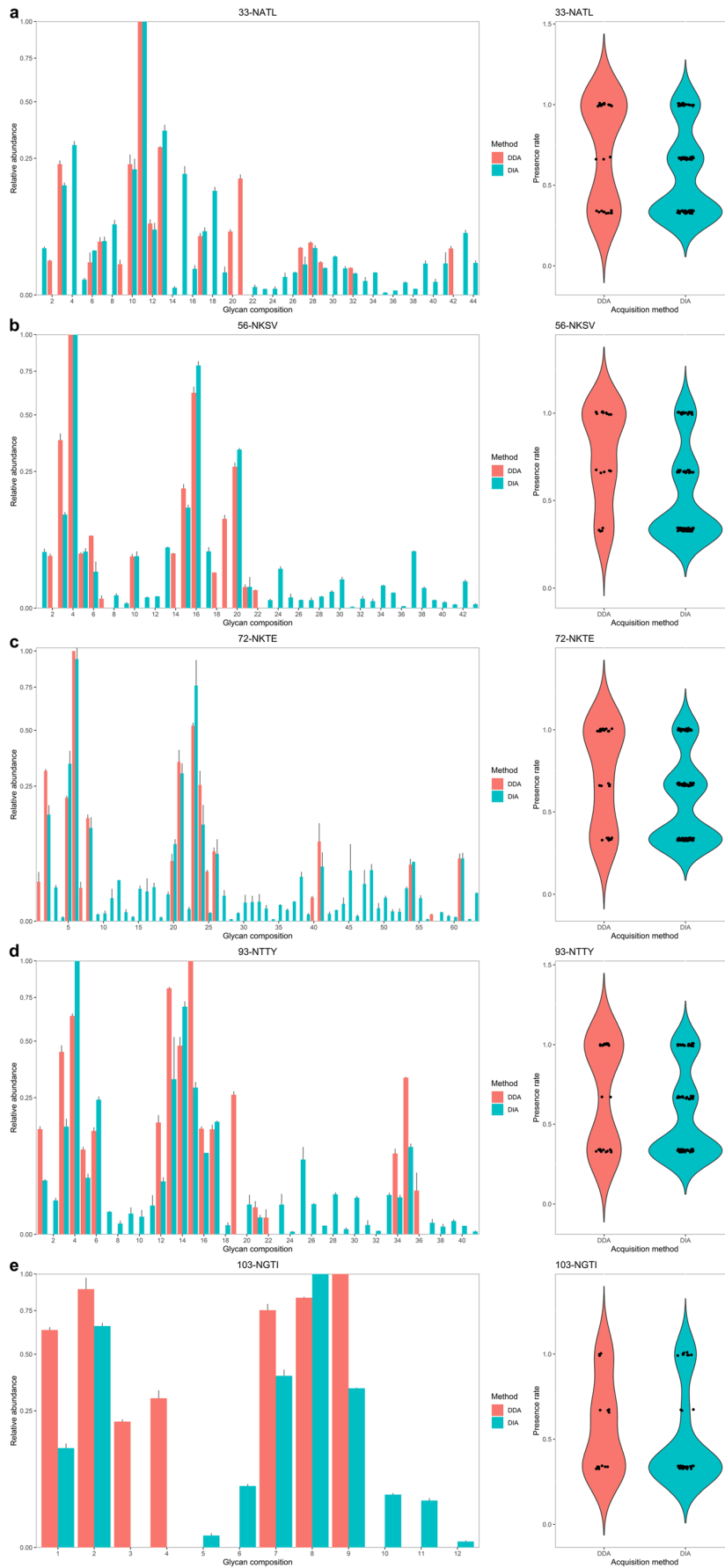c 72-NKTE

d 93-NTTY

e 103-NGTI

Fig. 1 Comparison of DDA and DIA results for the five N-glycosylation sites of the A1AG1 isoform of alpha-1-acid glycoprotein (AGP): **a** 33-NATL, **b** 56-NKSV, **c** 72-NKTE, **d** 93-NTTY, and **e** 103-NGTI. Bar plots indicate the mean relative abundance of glycoforms (± standard deviation) that met the GlycReSoft scoring threshold in at least two replicates. Glycan compositions for each glycoform can be found in Supplemental File 4. Violin plots display the presence rate across replicates, including those glycans only observed in a single replicate. For all sites, DIA produced more coverage but less reproducibility than DDA

The glycopeptides from spike protein that were assigned using DDA but not in DIA, along with the scan numbers and approximate scores for the tandem mass spectra they would have been found in, are tabulated in Supplemental File 3. The mirror plots in Fig. 5 and Supplemental File 5 show the annotated tandem mass spectra for the glycopeptides from this list, with the upper panels displaying spectra from DDA data that met the GlycReSoft FDR threshold and the lower panels displaying spectra from DIA that did not. Figure S3 shows b- and y-ions from the eight glycosites in which DIA underperformed DDA. Generally speaking, the DIA spectra that were outside of the FDR scoring threshold generated peptide backbone product ions that were fewer in number and/or lower in abundance than peptide backbone ions from DDA spectra that did pass the scoring threshold.

Discussion

We acquired site-specific glycopeptide data for a standard glycoprotein, AGP, as well as for recombinant SARS-CoV-2 spike protein using two mass spectrometry data acquisition methods. DDA has been used by many research groups and has proven to be successful in identifying site-specific glycopeptides. However, higher sensitivity and selectivity, as well as stronger replicate reproducibility, would better enable us to perform surveillance of evolving viral glycoproteins, such as the SARS-CoV-2 spike protein, for the purpose of designing effective vaccines and boosters against the pathogen. Thus, we evaluated the performance of DIA and used this method to characterize the site-specific glycopeptides of recombinant SARS-CoV-2 in a quantitative and unbiased approach.

For our standard glycoprotein, AGP, we compared the number of assignments obtained using the two mass spectrometry methods. The higher-abundance glycoforms that were assigned by both methods were in good agreement with previously published information about the extent and structure of glycosylation of AGP used as a standard glycoprotein [43]. DIA assigned more glycoforms at all five sites of the A1AG1 isoform of AGP than DDA did (Table 1). Looking at the bar plots in Fig. 1, the majority of DIA assignments that were missed by DDA were low-abundance compositions,

suggesting that DIA has improved sensitivity over DDA. However, the violin plots in Fig. 1 indicate a lower level of reproducibility across replicates for DIA. There was a greater number of compositions in DDA that were present in all replicates, with the exception of 103-NGTI(Fig. 1e), whereas for DIA, there were more compositions that had missing values. DIA had more single-replicate compositions than DDA as well, suggesting that the sensitivity gains of DIA were made at the cost of reproducibility.

The higher reproducibility of DDA may be due to the fact that DDA generally only detected the glycoforms with higher abundance (Fig. 2a). It stands to reason that higher-abundance precursors would be more easily selected in each replicate than lower abundance ones. DIA does not select precursors for fragmentation, but rather fragments all precursors within the specified m/z windows. Lower-abundance precursors would not be overlooked, as they would be with DDA; however, co-isolating precursors within the m/z windows may complicate the tandem mass spectra for these lower-abundance precursors, producing spectra that may not consistently pass GlycReSoft FDR scoring thresholds.

The situation with spike protein is more nuanced than for AGP. Just as with AGP, DIA data had lower presence rates across replicates (see violin plots in Figs. 3 and 4). There were ten glycosylation sites for which we assigned as many or more glycoforms by DIA than by DDA (Fig. 3). For these sites, the higher-abundance glycoforms were generally assigned by both methods. In total, DIA assigned more glycoforms among all glycosylation sites combined than DDA did (518 for DDA vs. 623 for DIA) (Table 2). However, there were eight sites where DDA outperformed DIA in assigning more glycan compositions (Fig. 4).

For all spike glycosites, the violin plots were widest in DDA generally when glycoforms had a presence rate of 1.0, i.e., most DDA-assigned glycoforms were highly reproducible across replicates. This trend can be attributed to the fact that the DDA precursors that produced tandem mass spectra that met thescoring threshold were higher-abundance glycopeptides (see Fig. 2b). The widest part of the DIA violins in Figs. 3 and 4 were when the presence rate was 0.25. That is to say, DIA exhibited high sensitivity, detecting a large number of glycopeptides with abundances ranging from 10^5 to 10^{10}(Fig. 2b); however, most of these glycan compositions did not count as positive assignments because they were only identified in a single replicate. The amount of time spent per spectrum may explain this trend. Across the entire retention time range, the duty cycle time for DIA was longer than for DDA because DIA was performing many more scans per cycle. However, the time spent per scan was shorter in DIA than in DDA (Fig. 2c); time spent per DIA scan was relatively constant, at approximately 155 ms throughout the retention time range, compared to the time spent per scan for DDA, which ranged from approximately 160 to 250 ms. Thus, for
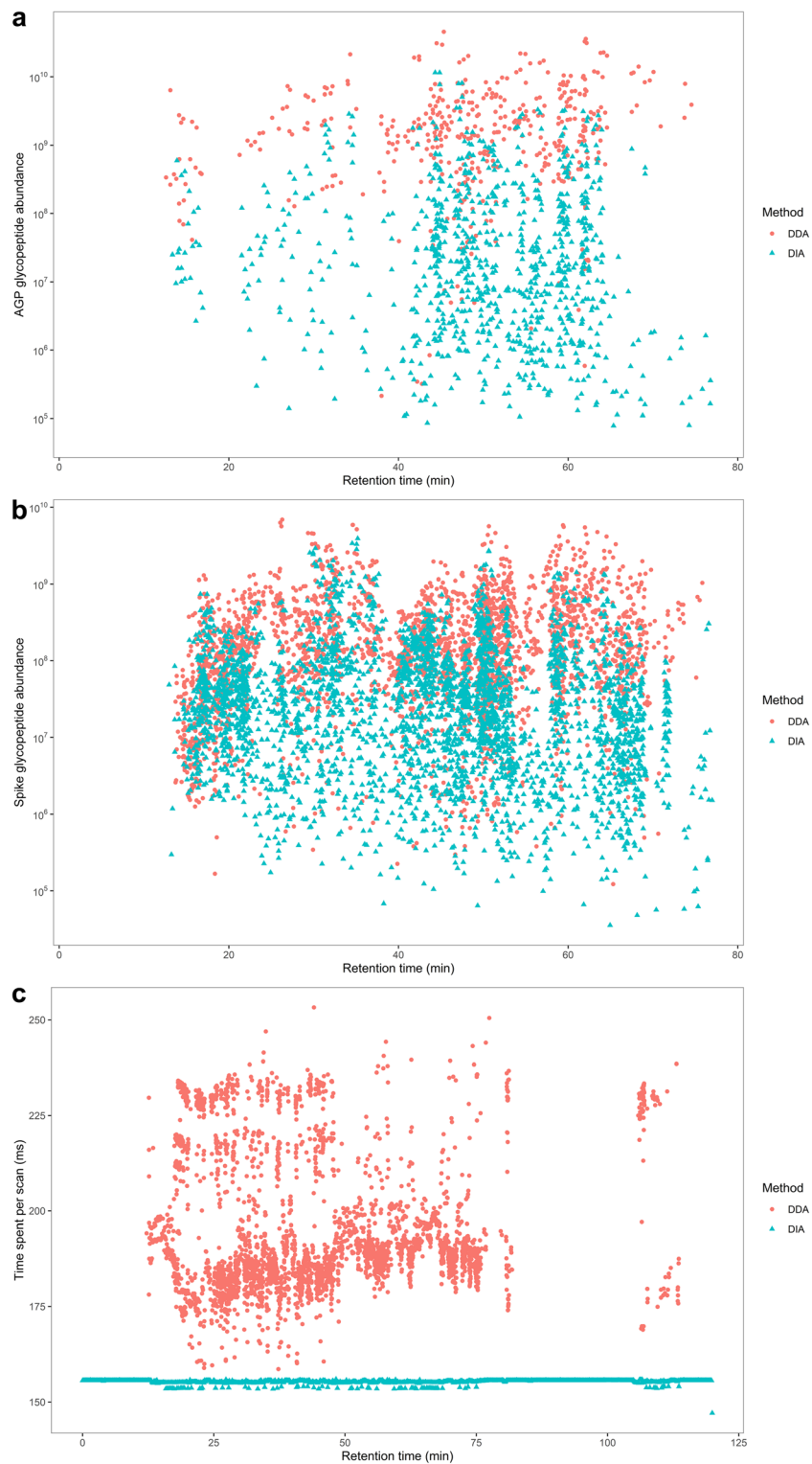
Fig. 2 The aggregated MS1 abundance for all precursors from the A1AG1 isoform of AGP (**a**) and SARS-CoV-2 spike protein (**b**) assigned in DDA and DIA replicates that met GlycReSoft scoring thresholds, including precursors that were only assigned in a single replicate, plotted as a function of retention time. **c** The time spent per scan was calculated by dividing the duty cycle time by 1 (for the MS1 scan) plus the number of MS2 scans triggered in each cycle

DIA, the mass spectrometer spent the same amount of time fragmenting lower- and higher-abundance precursors. The low-abundance precursors did not have extra accumulation time, resulting in fewer and/orlower-abundance peptide backbone fragment ions (Fig. S2). Our scoring algorithm relied solely on peptide fragment ions; thus, it makes sense that many of the lower-abundance DIA glycopeptides would not pass the FDR scoring threshold. Figure 5a shows the tandem mass spectra for the glycopeptide VLYQDVN(N-glycosylation)C(Carbamidomethyl)T + {Fuc:1; Hex:6; HexNAc:5; Neu5Ac:1} acquired by DDA (upper), which passed the scoring threshold, and DIA (lower), which did not meet the threshold and was therefore not assigned. The DIA tandem mass spectrum here is very sparse, dominated by non-specific oxonium ions (green) and only two low-abundance b-ions. Co-isolation of precursors in DIA is another potential cause of missed assignment. Figure 5b shows DDA and DIA tandem mass spectra for glycopeptide GIN(N-glycosylation)ASVV + {Fuc:1; Hex:7; HexNAc:6; Neu5Ac:1}. The DIA spectrum had sufficiently abundant b- and y-ions that matched the theoretical fragment ions for this glycopeptide, but there were other non-matching peaks (gray) that complicated the spectrum, producing a low expected MS2 score. Figure 5c for glycopeptide NLN(N-glycosylation)ESLIDLQELGK +

{Fuc:1; Hex:6; HexNAc:6; Neu5Ac:1} is an example in which the DIA spectrum had few peptide backbone fragment ions in addition to noise from a co-isolating precursor or precursors. An additional orthogonal method of separation, such as ion mobility, may improve the co-isolation problem, and hardware improvements to make a brighter source may be necessary to produce higher-scoring tandem mass spectra.

Conclusions

The COVID-19 pandemic has highlighted the need for rapid, global surveillance of viral pathogens, especially as viruses can rapidly produce variants with higher transmissibility or virulence. This surveillance should include quantitative glycoproteomics. In this work, we have demonstrated methods for performing this surveillance. It should be noted that manual interpretation of spectra not only introduces an unacceptable level of bias, but also would undermine the expediency required to support active public health measures. DDA is more unbiased than manual interpretation, and automated searching algorithms such as GlycReSoft are able to speed up DDA identifications; however, DDA still

| Table 2 Number of glycoforms assigned by DDA and DIA for SARS2-CoV-2 spike protein in at least two replicates. Values in parentheses represent the total number of glycoforms detected, including those that were identified in only a single replicate	Spike protein glycosylation sites	Glycoforms identified by DDA	Glycoforms identified by DIA
	17-NLTT	1 (2)	0 (0)
	61-NVTW	2 (4)	2 (5)
	74-NGTK	50 (67)	46 (94)
	122-NATN	57 (73)	68 (131)
	149-NKSW	2 (2)	1 (3)
	165-NCTF	19 (24)	5 (6)
	234-NITR	5 (9)	16 (34)
	282-NGTI	42 (63)	35 (61)
	331-NITN	22 (29)	69 (129)
	343-NATR	53 (90)	76 (169)
	603-NTSN	9 (14)	3 (14)
	616-NCTE	19 (29)	5 (19)
	657-NNSY	17 (20)	14 (36)
	709-NNSI	23 (36)	34 (80)
	717-NFTI	25 (54)	38 (96)
	801-NFSQ	34 (52)	35 (84)
	1074-NFTT	30 (42)	50 (82)
	1098-NGTH	79 (124)	114 (188)
	1134-NNTV	11 (16)	5 (13)
	1158-NHTS	1 (3)	0 (0)
	1173-NASV	6 (8)	4 (14)
	1194-NESL	11 (15)	3 (9)
	Total	73 (118)	183 (458)
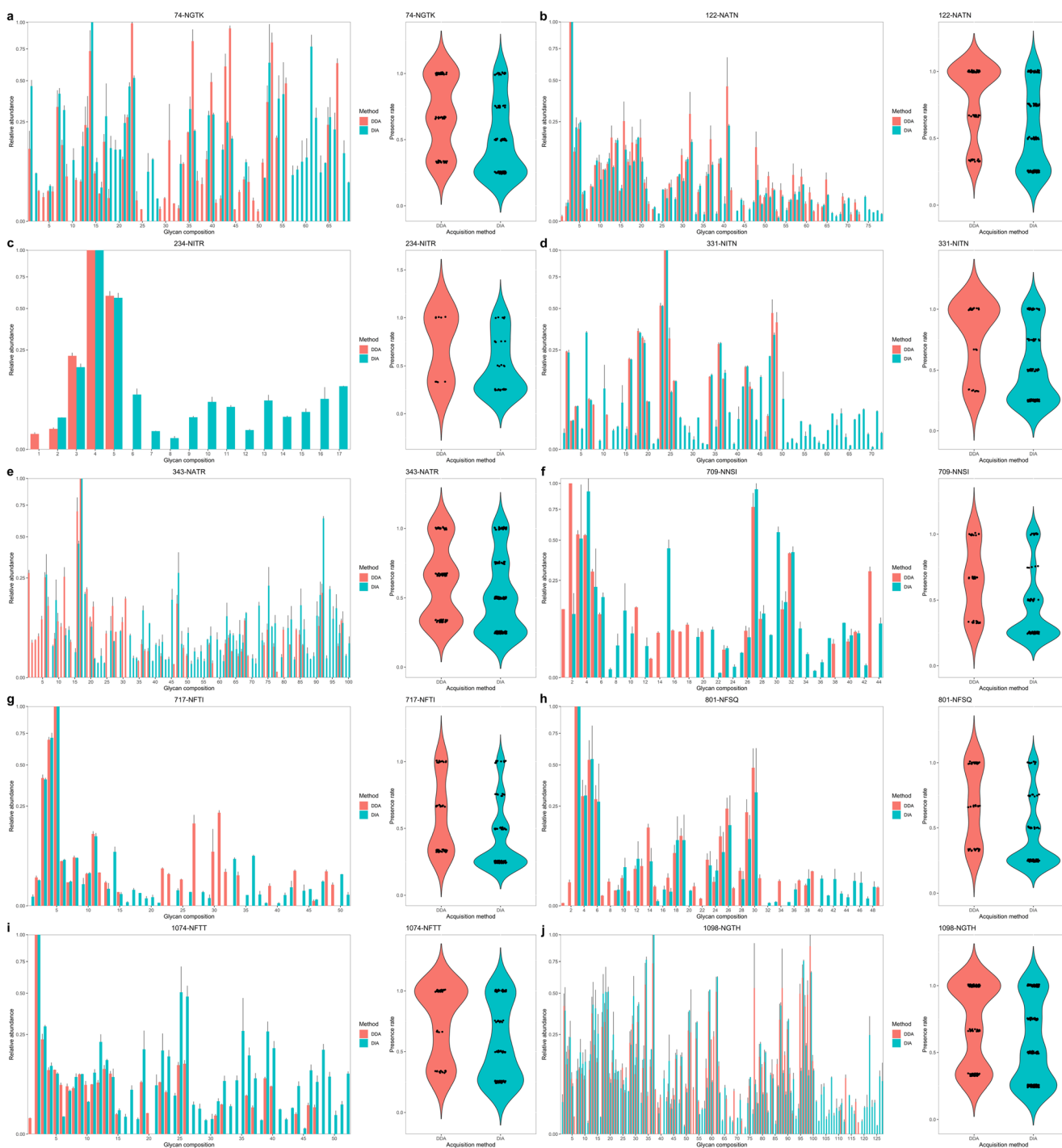
Fig. 3 Glycosylation sites for which as many or more glycoforms meeting scoring thresholds were observed in DIA than in DDA in SARS-CoV-2 spike protein: **a** 74-NGTK, **b** 122-NATN, **c** 234-NITR, **d** 331-NITN, **e** 343-NATR, **f** 709-NNSI, **g** 717-NFTI, **h** 801-NFSQ, **i** 1074-NFTT, and **j** 1098-NGTH. Bar plots indicate the mean relative abundance of glycoforms (± standard deviation) that met the GlycReSoft scoring threshold in at least two replicates. Glycan compositions for each glycoform can be found in Supplemental File 4. Violin plots display the presence rate across replicates, including those glycans only observed in a single replicate. For these sites, DIA produced comparable or increased coverage over DDA, but less reproducibility

suffers from a bias toward higher-abundance glycopeptides. We applied DIA to quantitative glycoproteomics to attempt to increase sensitivity while maintaining acceptable specificity relative to DDA. We used DIA-Umpire to employ retention time constraints during the preprocessing step to improve the confidence and selectivity of our identification. However, we found that while DIA succeeded in improving sensitivity in most
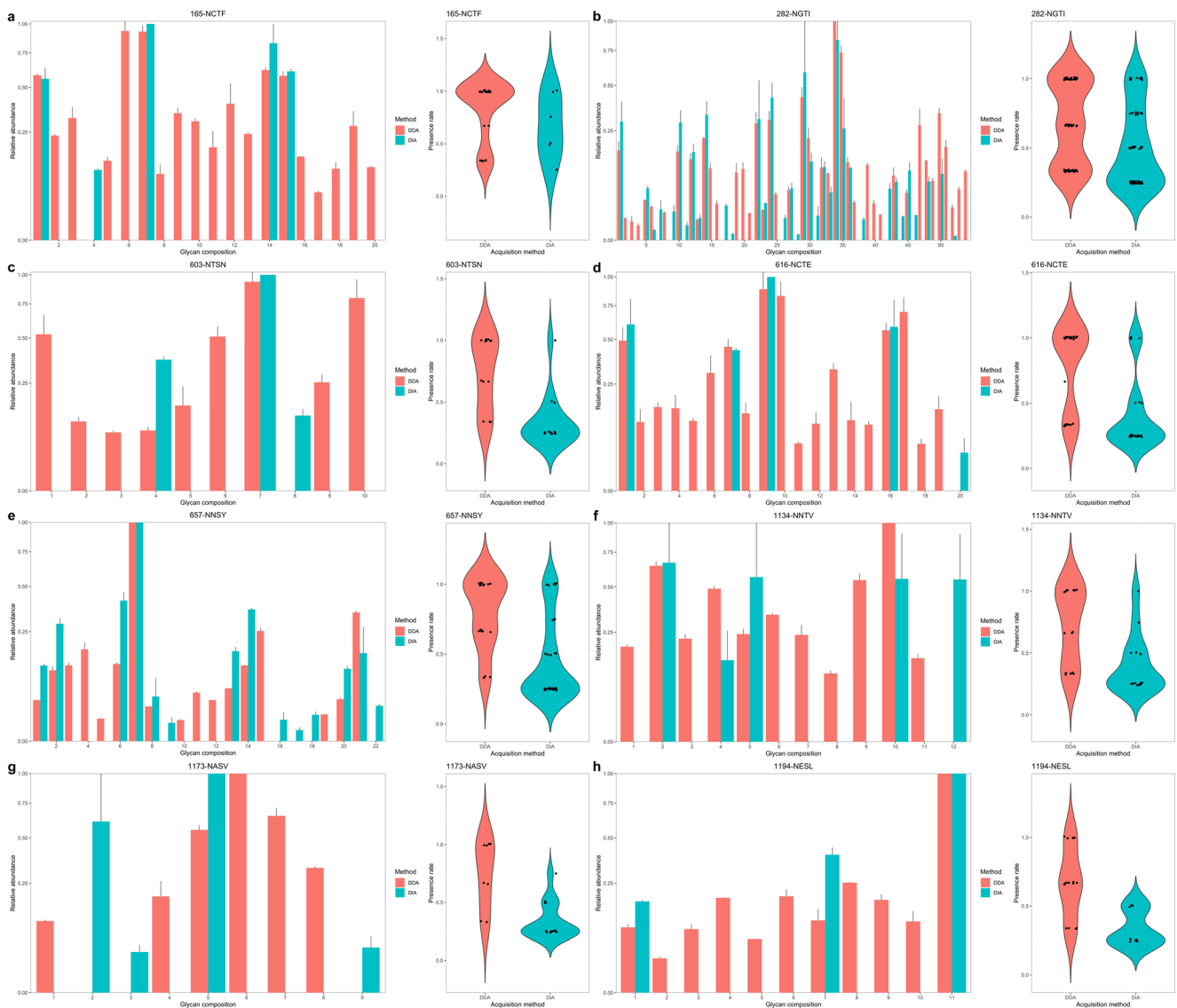
Fig. 4 Glycosylation sites for which fewer glycoforms meeting scoring thresholds were observed in DIA than in DDA in SARS-CoV-2 spike protein: **a** 165-NCTF, **b** 282-NGTI, **c** 603-NTSN, **d** 616-NCTE, **e** 657-NNSY, **f** 1134-NNTV, **g** 1173-NASV, and **h** 1194-NESL. Bar plots indicate the mean relative abundance of glycoforms (± standard deviation) that met the GlycReSoft scoring threshold in at least two replicates.

Glycan compositions for each glycoform can be found in Supplemental File 4. Violin plots display the presence rate across replicates, including those glycans only observed in a single replicate. For these sites, DIA produced less coverage and less reproducibility than DDA

cases and reducing bias, reliably observing and assigning low-abundance glycopeptides as well as deciphering complicated tandem mass spectra from co-isolated precursors remain a challenge.

Heterogeneity and low relative abundance of glycosylation means that we cannot simply import DIA methods developed for proteomics wholesale to do the work of glycoproteomics. Future efforts must include improving and standardizing methods for scoring tandem mass spectra of glycopeptides. There are a number of published datasets on site-specific SARS-CoV-2 spike glycoproteomics. Our results are in agreement with others

with respect to the general distribution of glycan types found at each site. For example, we and others [19, 44] found that sites 234-NITR, 709-NNSI, 717-NFTI, 801-NFSQ, and 1074-NFTT all had a high degree of high-mannose glycans, and we had agreement to which specific high-mannose composition was dominant. However, there is no consensus on glycopeptide identity and quantification of the spike protein [45]. This is because the available glycoproteomics identification software programs have different database search methods and, importantly, there is no standardized target-decoy analysis for calculating FDR [45]. The glycoproteomics community must
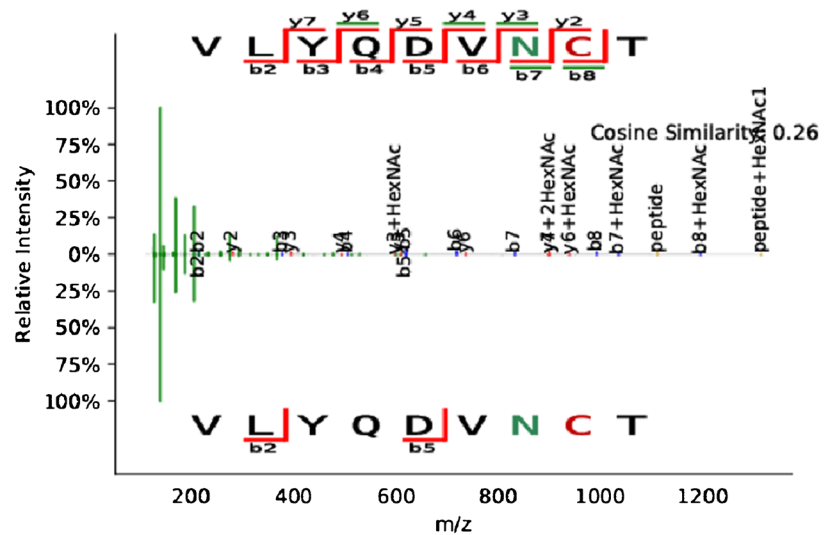
a

Glycan
{Fuc:1; Hex:6; HexNAc:5; Neu5Ac:1}
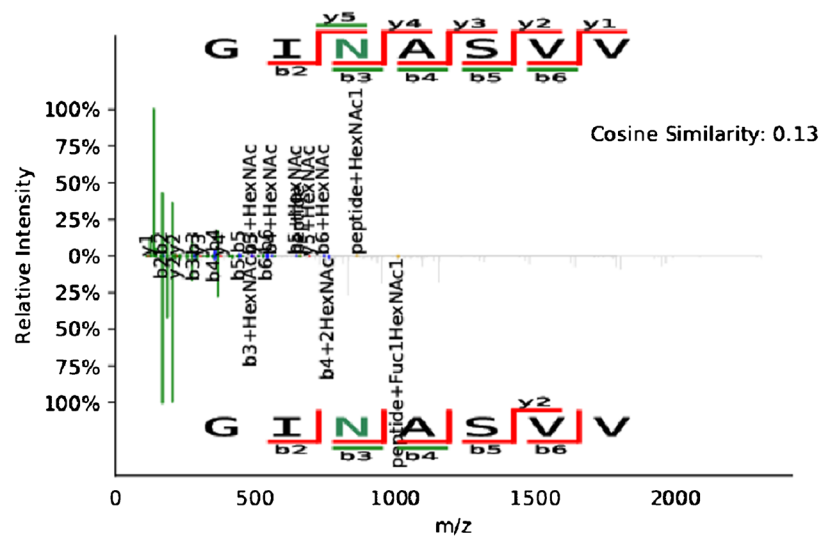
MS2 scores
DDA: 122.89 (upper)
DIA, expected: 0.20 (lower)

b

Glycan
{Fuc:1; Hex:7; HexNAc:6; Neu5Ac:1}

MS2 scores
DDA:148.30 (upper)
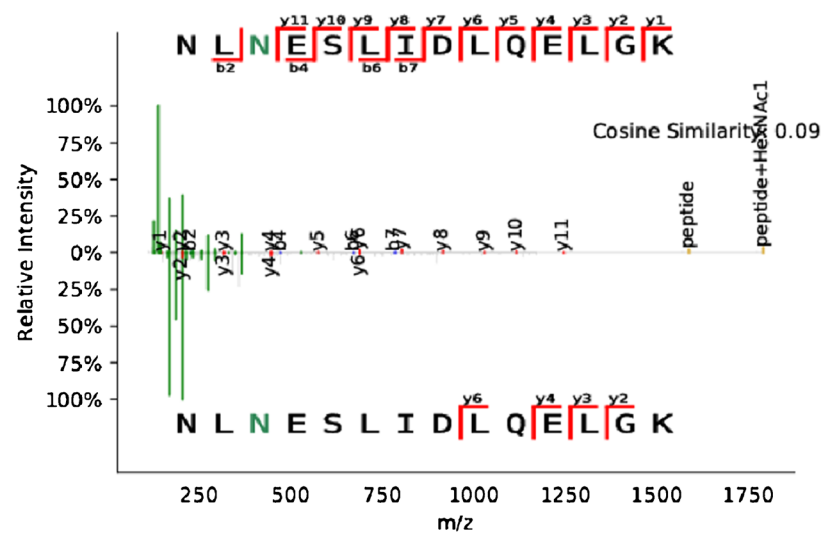DIA, expected: 13.59 (lower)

c

Glycan
{Fuc:1; Hex:6; HexNAc:6; Neu5Ac:1}

MS2 scores
DDA: 19.37 (upper)
DIA, expected: 1.03 (lower)

Fig. 5 Annotated tandem mass spectra for glycopeptides identified by DDA (upper panels) juxtaposed with tandem mass spectra from DIA data where the same glycopeptides were expected but did not meet scoring criteria by GlycReSoft (lower panels). **a** The DIA spectrum for VLYQDVN(N-Glycosylation)C(Carbamidomethyl)T + {Fuc:1; Hex:6; HexNAc:5; Neu5Ac:1} was very sparse, with very few fragment ions apart from oxonium ions. **b** The DIA spectrum for GIN(N-Glycosylation) + {Fuc:1; Hex:7; HexNAc:6; Neu5Ac:1} had more fragment ion matches than the previous case, but extraneous peaks not originating from this precursor ion complicated the interpretation of the spectrum. **c** The DIA spectrum for NLN(N-Glycosylation)ESLIDLQELGK + {Fuc:1; Hex:6; HexNAc:6; Neu5Ac:1} had fragment ion peaks, but very few matched the expected glycopeptide

collectively decide upon the best methods for doing so in order for glycosylation surveillance data to be at all meaningful.

Declarations

References

1. Mast SW, Moremen KW (2006) Family 47 α-mannosidases in N-glycan processing. In: Methods Enzymol. Elsevier, pp. 31–46.
2. Chang D, Hackett WE, Zhong L, Wan X-F, Zaia J. Measuring site-specific glycosylation similarity between influenza a virus variants with statistical certainty. Mol Cell Proteomics. 2020;19:1533–45. https://doi.org/10.1074/mcp.RA120.002031.
3. Wiley DC, Wilson IA, Skehel JJ. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature. 1981;289:373–8.
4. Koel BF, Mögling R, Chutinimitkul S, Fraaij PL, Burke DF, van der Vliet S, de Wit E, Bestebroer TM, Rimmelzwaan GF, Osterhaus ADME, Smith DJ, Fouchier RAM, de Graaf M. Identification of amino acid substitutions supporting antigenic change of influenza A(H1N1)pdm09 viruses. J Virol. 2015;89:3763–75. https://doi.org/10.1128/JVI.02962-14.
5. Blackburne BP, Hay AJ, Goldstein RA. Changing selective pressure during antigenic changes in human influenza H3. PLoS Pathog. 2008;4:e1000058. https://doi.org/10.1371/journal.ppat.1000058.
6. Cherry JL, Lipman DJ, Nikolskaya A, Wolf YI. Evolutionary dynamics of N-glycosylation sites of influenza virus hemagglutinin. PLoS Curr. 2009;1:RRN1001. https://doi.org/10.1371/currents.rrn1001.
7. Das SR, Hensley SE, David A, Schmidt L, Gibbs JS, Puigbo P, Ince WL, Bennink JR, Yewdell JW. Fitness costs limit influenza A virus hemagglutinin glycosylation as an immune evasion strategy. Proc Natl Acad Sci U S A. 2011;108:E1417–22. https://doi.org/10.1073/pnas.1108754108.
8. Kosik I, Ince WL, Gentles LE, Oler AJ, Kosikova M, Angel M, Magadán JG, Xie H, Brooke CB, Yewdell JW. Influenza A virus hemagglutinin glycosylation compensates for antibody escape fitness costs. PLoS Pathog. 2018;14:e1006796. https://doi.org/10.1371/journal.ppat.1006796.
9. Wanzeck K, Boyd KL, McCullers JA. Glycan shielding of the influenza virus hemagglutinin contributes to immunopathology in mice. Am J Respir Crit Care Med. 2011;183:767–73. https://doi.org/10.1164/rccm.201007-1184OC.
10. Iraqi M, Edri A, Greenshpan Y, Kundu K, Bolel P, Cahana A, Ottolenghi A, Gazit R, Lobel L, Braiman A, Porgador A. N-Glycans mediate the Ebola Virus-GP1 shielding of ligands to immune receptors and immune evasion. Front Cell Infect Microbiol. 2020;10:48. https://doi.org/10.3389/fcimb.2020.00048.
11. Re S, Mizuguchi K. Glycan cluster shielding and antibody epitopes on Lassa virus envelop protein. J Phys Chem B. 2021;125:2089–97. https://doi.org/10.1021/acs.jpcb.0c11516.
12. Lavie M, Hanoulle X, Dubuisson J. Glycan shielding and modulation of hepatitis C virus neutralizing antibodies. Front Immunol. 2018;9:910. https://doi.org/10.3389/fimmu.2018.00910.
13. Watanabe Y, Bowden TA, Wilson IA, Crispin M. Exploitation of glycosylation in enveloped virus pathobiology. Biochim Biophys Acta Gen Subj. 2019;1863:1480–97. https://doi.org/10.1016/j.bbagen.2019.05.012.
14. Mascola JR, Graham BS, Fauci AS. SARS-CoV-2 viral variants—tackling a moving target. JAMA. 2021;325:1261. https://doi.org/10.1001/jama.2021.2088.
15. CDC (2020) Coronavirus disease 2019 (COVID-19). In: Cent. Dis. Control Prev. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html. Accessed 6 Jul 2021.
16. Salvatore M, Bhattacharyya R, Purkayastha S, Zimmermann L, Ray D, Hazra A, Kleinsasser M, Mellan T, Whittaker C, Flaxman S, Bhatt S, Mishra S, Mukherjee B (2021) Resurgence of SARS-CoV-2 in India: potential role of the B.1.617.2 (Delta) variant and delayed interventions. medRxiv 2021.06.23.21259405 https://doi.org/10.1101/2021.06.23.21259405.
17. Watanabe Y, Berndsen ZT, Raghwani J, Seabright GE, Allen JD, Pybus OG, McLellan JS, Wilson IA, Bowden TA, Ward AB, Crispin M. Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. Nat Commun. 2020;11:2688. https://doi.org/10.1038/s41467-020-16567-0.
18. Sanda M, Morrison L, Goldman R. N- and O-glycosylation of the SARS-CoV-2 spike protein. Anal Chem. 2021;93:2003–9. https://doi.org/10.1021/acs.analchem.0c03173.

19. Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. Science. 2020;369:330–3. https://doi.org/10.1126/science.abb9983.

20. Zhang Y, Zhao W, Mao Y, Chen Y, Wang S, Zhong Y, Su T, Gong M, Du D, Lu X, Cheng J, Yang H. Site-specific N-glycosylation characterization of recombinant SARS-CoV-2 spike proteins. Mol Cell Proteomics. 2021;20:100058. https://doi.org/10.1074/mcp. RA120.002295.

21. Shajahan A, Supekar NT, Gleinich AS, Azadi P. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. Glycobiology. 2020;30:981–8. https://doi.org/ 10.1093/glycob/cwaa042.

22. Klein J, Carvalho L, Zaia J. Application of network smoothing to glycan LC-MS profiling. Bioinformatics. 2018;34:3511–8. https:// doi.org/10.1093/bioinformatics/bty397.

23. Polasky DA, Yu F, Teo GC, Nesvizhskii AI. Fast and comprehensive N- and O-glycoproteomics analysis with MSFragger-Glyco. Nat Methods. 2020;17:1125–32. https://doi.org/10.1038/s41592-020-0967-9.

24. Liu M-Q, Zeng W-F, Fang P, Cao W-Q, Liu C, Yan G-Q, Zhang Y, Peng C, Wu J-Q, Zhang X-J, Tu H-J, Chi H, Sun R-X, Cao Y, Dong M-Q, Jiang B-Y, Huang J-M, Shen H-L, Wong CCL, et al. pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. Nat Commun. 2017;8:438. https://doi.org/10.1038/ s41467-017-00535-2.

25. Bern M, Kil YJ, Becker C. Byonic: advanced peptide and protein identification software. Curr Protoc Bioinforma. 2012;40:13.20.1–13.20.14. https://doi.org/10.1002/0471250953.bi1320s40.

26. HUPO Test Sample Working Group, Bell AW, Deutsch EW, Au CE, Kearney RE, Beavis R, Sechi S, Nilsson T, JJM B. A HUPO test sample study reveals common problems in mass spectrometry–based proteomics. Nat Methods. 2009;6:423–30. https://doi.org/10. 1038/nmeth.1333.

27. Lin C-H, Krisp C, Packer NH, Molloy MP. Development of a data independent acquisition mass spectrometry workflow to enable glycopeptide analysis without predefined glycan compositional knowledge. J Proteome. 2018;172:68–75. https://doi.org/10.1016/j.jprot. 2017.10.011.

28. Pan K-T, Chen C-C, Urlaub H, Khoo K-H. Adapting data-independent acquisition for mass spectrometry-based protein site-specific N-glycosylation analysis. Anal Chem. 2017;89:4532–9. https://doi.org/10.1021/acs.analchem.6b04996.

29. Collins BC, Hunter CL, Liu Y, Schilling B, Rosenberger G, Bader SL, Chan DW, Gibson BW, Gingras A-C, Held JM, Hirayama-Kurogi M, Hou G, Krisp C, Larsen B, Lin L, Liu S, Molloy MP, Moritz RL, Ohtsuki S, et al. Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. Nat Commun. 2017;8:291. https://doi.org/10. 1038/s41467-017-00249-5.

30. Gillet LC, Leitner A, Aebersold R. Mass spectrometry applied to bottom-up proteomics: entering the high-throughput era for hypothesis testing. Annu Rev Anal Chem. 2016;9:449–72. https://doi.org/ 10.1146/annurev-anchem-071015-041535.

31. Bruderer R, Bernhardt OM, Gandhi T, Xuan Y, Sondermann J, Schmidt M, Gomez-Varela D, Reiter L. Optimization of experimental parameters in data-independent mass spectrometry significantly increases depth and reproducibility of results. Mol Cell Proteomics. 2017;16:2296–309. https://doi.org/10.1074/mcp. RA117.000314.

32. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for

consistent and accurate proteome analysis. Mol Cell Proteomics. 2012;11:O111.016717. https://doi.org/10.1074/mcp.O111.016717.

33. Martens L, Chambers M, Sturm M, Kessner D, Levander F, Shofstahl J, Tang WH, Römpp A, Neumann S, Pizarro AD, Montecchi-Palazzi L, Tasman N, Coleman M, Reisinger F, Souda P, Hermjakob H, Binz P-A, Deutsch EW. mzML—a community standard for mass spectrometry data. Mol Cell Proteomics. 2011;10:R110.000133. https://doi.org/10.1074/mcp.R110.000133.

34. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak M-Y, Paulse C, Creasy D, et al. A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol. 2012;30:918–20. https://doi.org/10.1038/nbt.2377.

35. Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, Zhang W, Zhang Z, Lajoie GA, Ma B. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics. 2012;11:M111.010587. https://doi.org/10. 1074/mcp.M111.010587.

36. Magrane M, Consortium U (2011) UniProt Knowledgebase: a hub of integrated protein data. Database 2011:bar009 . https://doi.org/ 10.1093/database/bar009

37. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, Graham BS, McLellan JS. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020;367:1260–3. https://doi.org/10.1126/science.abb2507.

38. Eisenacher M. mzIdentML: an open community-built standard format for the results of proteomics spectrum identification algorithms. Meth Mol Biol Clifton NJ. 2011;696:161–77. https://doi.org/10. 1007/978-1-60761-987-1_10.

39. Tsou C-C, Avtonomov D, Larsen B, Tucholska M, Choi H, Gingras A-C, Nesvizhskii AI. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. Nat Methods. 2015;12:258–64. https://doi.org/10.1038/nmeth.3255.

40. Klein J, Zaia J. Glypy: an open source glycoinformatics library. J Proteome Res. 2019;18:3532–7. https://doi.org/10.1021/acs. jproteome.9b00367.

41. Klein JA, Meng L, Zaia J. Deep sequencing of complex proteoglycans: a novel strategy for high coverage and site-specific identification of glycosaminoglycan-linked peptides. Mol Cell Proteomics. 2018;17:1578–90. https://doi.org/10.1074/mcp.RA118.000766.

42. Kovalchik KA, Colborne S, Spencer SE, Sorensen PH, Chen DDY, Morin GB, Hughes CS. RawTools: rapid and dynamic interrogation of Orbitrap data files for mass spectrometer system management. J Proteome Res. 2019;18:700–8. https://doi.org/10.1021/acs. jproteome.8b00721.

43. Khatri K, Klein JA, Zaia J. Use of an informed search space maximizes confidence of site-specific assignment of glycoprotein glycosylation. Anal Bioanal Chem. 2017;409:607–18. https://doi.org/ 10.1007/s00216-016-9970-5.

44. Zhao P, Praissman JL, Grant OC, Cai Y, Xiao T, Rosenbalm KE, Aoki K, Kellman BP, Bridger R, Barouch DH, Brindley MA, Lewis NE, Tiemeyer M, Chen B, Woods RJ, Wells L. Virus-receptor interactions of glycosylated SARS-CoV-2 spike and human ACE2 receptor. Cell host microbe. 2020;28:586–601.e6. https://doi.org/10.1016/j.chom.2020.08.004.

45. Hackett WE, Zaia J. The need for community standards to enable accurate comparison of glycoproteomics algorithm performance. Molecules. 2021;26:4757. https://doi.org/10.3390/ molecules26164757.