



## Simplifying glycan monitoring of complex antigens such as the SARS-CoV-2 spike to accelerate vaccine development

Janelle Sauvageau <sup>1✉</sup>, Izel Koyuturk<sup>2,3</sup>, Frank St. Michael<sup>1</sup>, Denis Brochu<sup>1</sup>, Marie-France Goneau<sup>1</sup>, Ian Schoenhofen<sup>1</sup>, Sylvie Perret<sup>3</sup>, Alexandra Star <sup>1</sup>, Anna Robotham<sup>1</sup>, Arsalan Haqqani<sup>1</sup>, John Kelly<sup>1</sup>, Michel Gilbert<sup>1</sup> & Yves Durocher<sup>2,3</sup>

Glycosylation is a key quality attribute that must be closely monitored for protein therapeutics. Established assays such as HILIC-Fld of released glycans and LC-MS of glycopeptides work well for glycoproteins with a few glycosylation sites but are less amenable for those with multiple glycosylation sites, resulting in complex datasets that are time consuming to generate and difficult to analyze. As part of efforts to improve preparedness for future pandemics, researchers are currently assessing where time can be saved in the vaccine development and production process. In this context, we evaluated if neutral and acidic monosaccharides analysis via HPAEC-PAD could be used as a rapid and robust alternative to LC-MS and HILIC-Fld for monitoring glycosylation between protein production batches. Using glycoengineered spike proteins we show that the HPAEC-PAD monosaccharide assays could quickly and reproducibly detect both major and minor glycosylation differences between batches. Moreover, the monosaccharide results aligned well with those obtained by HILIC-Fld and LC-MS.

<sup>1</sup>Human Health Therapeutics Research Centre, National Research Council of Canada, 100 Sussex Dr., Ottawa, ON K1A 0R6, Canada. <sup>2</sup>Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montréal H3C 3J7, Canada. <sup>3</sup>Human Health Therapeutics Research Centre, National Research Council of Canada, 6100 Avenue Royalmount, Montréal, QC H4P 2R2, Canada. ✉email: [Janelle.Sauvageau@nrc-cnrc.gc.ca](mailto:Janelle.Sauvageau@nrc-cnrc.gc.ca)

Since the beginning of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) pandemic, a range of vaccine products such as messenger RNA vaccines, virus-like particles, and protein subunits have been developed for the purpose of protecting populations against this deadly virus<sup>1</sup>. Even though messenger RNA vaccines have received the most attention due to their quick development and proven efficacy, recombinant protein antigens represent the majority of candidates at the preclinical and clinical stage and many have recently been approved (e.g. Sanofi, Medicago, Novavax)<sup>1</sup>. In addition to being potential vaccine antigen candidates, these products are also exploited in antibody serology assays to detect anti-SARS-CoV-2 Immunoglobulin G in patient sera for the purpose of identifying infected individuals and evaluating vaccine efficacy<sup>2</sup>.

The spike glycoprotein has up to 22 N-glycosylation sites per protomer (Supplementary Sequence 1) and each site can be modified with a range of glycans, resulting in a very complex glycosylation profile. Glycosylation varies from one site to the next (for example, high-mannose glycans are more abundant on certain sites) and is also dependent on the cell line in which the protein is produced<sup>3</sup>. Additionally, glycosylation can vary from batch-to-batch even when the same cell line and production conditions are followed. Importantly, major changes in glycosylation on the protein antigens can affect immunogenicity. For example, spike glycoprotein treated with endoglycosidase-H, resulting in the removal of high-mannose N-glycan and leaving an N-acetyl-D-glucosamine (GlcNAc) residue at each site, appears to elicit better protection from SARS-CoV-2 in mice than its fully glycosylated counterpart<sup>4</sup>. Sialylation was also shown to affect the immunogenicity of the H5N1 influenza virus hemagglutinin antigens<sup>5</sup>.

Despite this, the literature suggests that glycan analysis is not routinely performed on protein antigens and that glycosylation is assumed to be typical of that produced by the cells used to express them<sup>6–8</sup>. For example, it is known that the glycans on the surface of Human Embryonic Kidney (HEK-293) cells are mostly complex glycans containing sialic acid while insect cells express paucimannose glycans with no sialic acid. Also, glycosylation is impacted by the culture conditions and glycan structures may markedly vary between individual N-glycosylation sites. Without this information, it is not possible to assess the real impact of glycosylation on immunogenicity nor to compare it with data from other studies producing the same antigen.

Given the influence that glycans can have on critical product attributes in glycoprotein-based therapeutics such as stability and efficacy, it is important to have reliable assays for monitoring this post-translational modification. Regulatory agencies such as the U.S. Food and Drug Administration have issued guidelines related to glycan analysis for protein products<sup>9,10</sup>. There are well-established methods for assaying the relatively simple glycosylation found on biotherapeutic proteins such as monoclonal antibodies. For example, Hydrophilic Interaction Chromatography with fluorescence detection (HILIC-FLD) of N-glycans released with Peptide N-glycosidase F (PNGase F) works well in this context as it allows for the rapid identity of complete glycans when no co-elution occurs. The identity of glycan isomers by proteolytic digestion and glycopeptide analysis by Liquid Chromatography-Mass Spectrometry (LC-MS) is another powerful alternative and is especially useful to identify the location of the glycan and analyze glycan microheterogeneity on different sites, but requires a larger investment of time and resources to perform<sup>11</sup>. These methods are being used by the research community to characterize the extensive glycosylation on some SARS-CoV-2 spike protein constructs<sup>12–20</sup>. However, the HILIC-FLD profiles of spike protein N-glycans tend to be complex and so are challenging to interpret and quantify<sup>14</sup>. In addition, the huge

spike glycopeptide data sets generated by LC-MS are not easy to use for comparing glycosylation between clones or batches. As such, there remains a need for a fast, reliable assay for monitoring the glycosylation of complex biotherapeutics like the SARS-CoV-2 spike glycoprotein, one that is easy to implement and fast. The latter trait is important as notable efforts are being invested currently in finding ways to shorten the development time for protein antigens so as to improve response speed in future pandemics<sup>21</sup>.

High-Performance Anion Exchange Chromatography coupled with Pulsed Amperometric Detection (HPAEC-PAD) can be used to identify and quantitate the neutral and acidic monosaccharide content of glycoproteins<sup>22–24</sup>. Two separate analyses are performed: a neutral sugar analysis for the quantification of L-fucose (Fuc), N-acetyl-D-galactosamine (GalNAc), GlcNAc, D-galactose (Gal), D-glucose (Glc) and D-mannose (Man), and a sialic acid analysis for the quantification of N-acetylneuraminic acid (Neu5Ac) and N-glycolylneuraminic acid (Neu5Gc). These analyses are straightforward to perform, do not require labeling and the results are easy to analyze, there being only a handful of monosaccharide peaks to resolve and quantify. Other methods can be used for monosaccharide analysis, however, they require labeling or are less sensitive than HPAEC-PAD<sup>25</sup>.

HPAEC-PAD was used in the past to analyze glycoproteins. It waned in popularity due to the development of techniques such as HILIC-FLD which are better suited for identifying the simpler glycan populations found on mAbs. While monosaccharide analysis does not provide glycan identity, it does give information about the average composition of monosaccharide units. In addition, the robustness of the method relative to other assays was questioned<sup>26–28</sup>. However, recent technological advances such as the introduction of disposable electrodes to eliminate fouling as well as the use of calibration curves to normalize the response differences between monosaccharides, have addressed most of these criticisms. Moreover, it has been shown that HPAEC-PAD can detect monosaccharides with unusual modifications such as mannose-6-phosphate which has been found on some glycoproteins<sup>29</sup>. It has to be determined yet if HPAEC-PAD can identify sugars with other modifications such as sulfation, a modification observed in low levels in the literature for the spike protein in HEK-293 cells<sup>13,18</sup>. Nevertheless, when the glycan profile is very complex, such as is the case with the SARS-CoV-2 spike glycoprotein, HPAEC-PAD could fill the gap left by established assays for a quick, reproducible, and quantitative monitoring assay.

To explore whether HPAEC-PAD could be used to quickly monitor glycans on complex antigens, two cohorts of spike proteins, one with major and the other with minor differences in glycosylation, were used for analysis. For the first set, five spike proteins with different N-glycosylation patterns were produced in engineered Chinese hamster ovary (CHO) cell lines (Table 1). A wild-type CHO phenotype (WT CHO<sup>55E1</sup>) containing N-glycans composed of Fuc, GlcNAc, Gal, Man, and Neu5Ac<sup>30,31</sup>, was chosen as a control<sup>32</sup>. To assess the impact of changes in Fuc and Neu5Ac content, the spike protein was produced in three CHO knockout (KO) cell lines: F15 (for Fucosyltransferase 8 (*FUT8*) KO, absence of core Fuc), S9 (for Sialyltransferase 4 (*ST3GAL4*) KO, near abolition of 2,3-linked Neu5Ac on N-glycans) and dKO2 (for *FUT8* and *ST3GAL4* double KO). Finally, the spike protein was also expressed in WT CHO<sup>55E1</sup> supplemented with kifunensin, an inhibitor of mannosidase I, to produce only high-mannose N-glycans (mainly Man9 structures). The second set of spike proteins were produced from stable pools and were expected to display the minor differences in glycosylation that would be more likely to be encountered when comparing antigen production lots. Both spike protein cohorts were characterized using the three assays to determine if differences detected by HILIC-FLD and LC-MS were also

reflected in the simpler and faster HPAEC-PAD assays. The experimental plan is outlined in Fig. 1.

CHO cell line (spike lot number)	Genotype	Phenotype	Representative N-glycan structure
F15 (PRO1-468)	<i>FUT8</i> KO	-Fucosylation + $\alpha$ 2,3 sialylation	
S9 (PRO1-469)	<i>ST3GAL4</i> KO	+Fucosylation - $\alpha$ 2,3 sialylation	
dKO2 (PRO1-470)	<i>FUT8</i> KO and <i>ST3GAL4</i> KO	-Fucosylation - $\alpha$ 2,3 sialylation	
WT CHO <sup>55E1</sup> (PRO1-471)	wt	+Fucosylation + $\alpha$ 2,3 sialylation	
WT CHO <sup>55E1</sup> + Kifunensine (PRO1-472)	wt	+high-mannose structures -Fucosylation - $\alpha$ 2,3 sialylation	

Glycans are drawn using DrawGlycan-SNFG software<sup>47</sup>. Green circle: Man, blue square: GlcNAc, yellow circle: Gal, red triangle: Fuc, purple diamond: Neu5Ac.

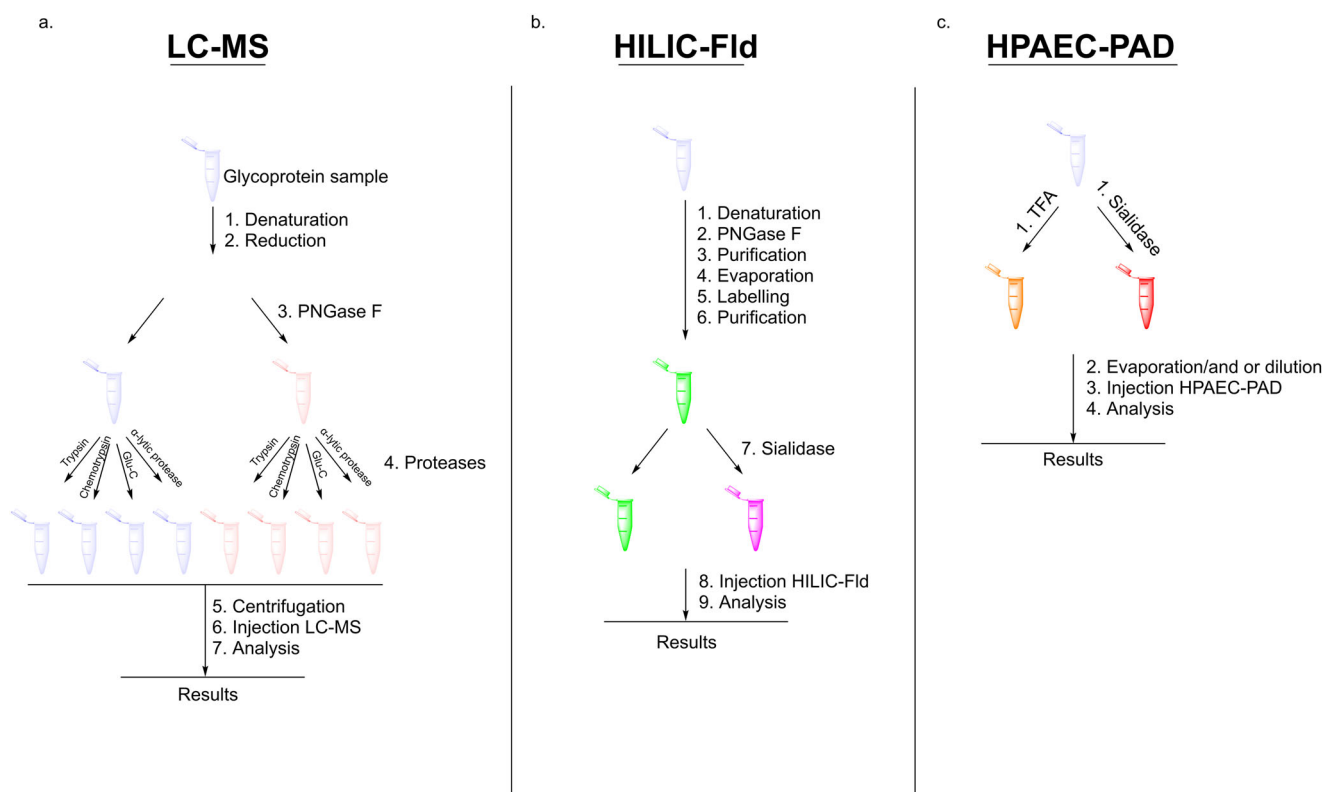
Our findings show that the major glycosylation differences in the spikes produced in the engineered CHO cell lines identified by HILIC-Fld and LC-MS could also be detected by HPAEC-PAD. The monosaccharide assays could also detect minor variations in the three batches obtained from the same stable pool and this also correlated with the results from the HILIC-Fld and LC-MS analyses. Thus, neutral sugar and sialic acid analyses using HPAEC-PAD can be an effective and robust alternative to the established HILIC-Fld and LC-MS assays for lot-to-lot monitoring of glycosylation of the SARS-CoV-2 glycoprotein and likely for other complex proteins also.

## Results

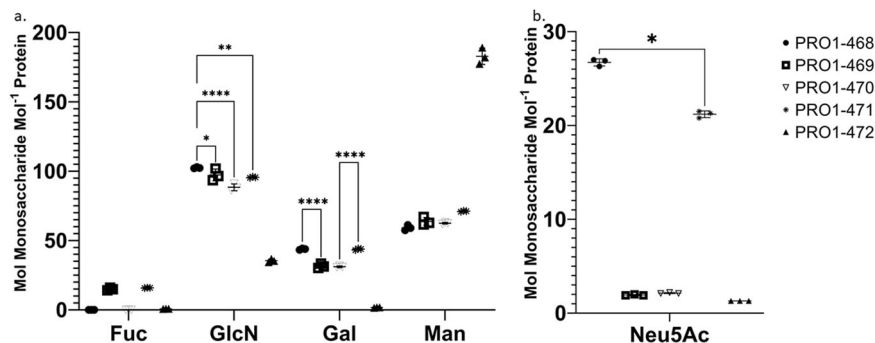
### Glycosylation analysis of spike protein from engineered CHO cell lines

**Monosaccharide analysis by HPAEC-PAD.** The neutral and acidic monosaccharide content of the spike proteins with various glycosylation patterns was determined using two separate assays. For neutral monosaccharide analysis, three aliquots of each spike protein were hydrolyzed with trifluoroacetic acid (TFA), injected in triplicate on an HPAEC-PAD system, and eluted with a sodium hydroxide (NaOH) gradient (Fig. 2a, and Supplementary Table 1). The average and standard deviation of the three hydrolysates and injections ( $n=9$ ) are reported in mol of monosaccharide mol<sup>-1</sup> of protein (protomer: 143,230.6 Da)

The monosaccharides Fuc, Gal, Man, and GlcN were detected in the hydrolysate of the WT spike protein (PRO1-471). D-Glucosamine (GlcN) is the N-deacetylation product of GlcNAc resulting from TFA treatment<sup>23</sup>. These monosaccharides are the main constituents of complex N-glycans and their presence is



**Fig. 1 Comparison of the sample preparation required for LC-MS, HILIC-Fld, and HPAEC-PAD.** **a** Illustrates the seven required steps to obtain the results via LC-MS, encompassing denaturation, reduction, PNGase F and protease digestion, centrifugation, injection, and analyses. **b** The nine necessary steps to acquire the results are outlined, involving denaturation, PNGase F digestion, purification, evaporation, labeling, purification, sialidase treatment, injection, and analyses. Lastly, **c** highlights the four essential steps prior to obtaining the results via HPAEC-PAD, encompassing TFA or sialidase cleavage, evaporation, injection, and analyses.



**Fig. 2 Monosaccharide analysis via HPAEC-PAD of glycovariants of the SARS-CoV-2 spike protein.** Results reported in mol monosaccharide mol<sup>-1</sup> monomer (protomer MW = 143,230.6 Da). Results are means of three triplicate injections and three triplicate reactions. Error bars are the standard deviation observed between the triplicate hydrolyses values. **a** Fuc, GlcN, Gal, and Man content obtained after TFA hydrolysis. The data were analyzed using two-way ANOVA, Tukey's multiple comparisons test, comparing the content of each monosaccharide type between variants. Not all comparisons shown. For full comparisons see Supplementary Table 2. **b** Neu5Ac content obtained after enzymatic hydrolysis. For a full comparison see Supplementary Table 3. *P* values: <0.03 (\*), <0.002 (\*\*), and <0.0001(\*\*\*\*).

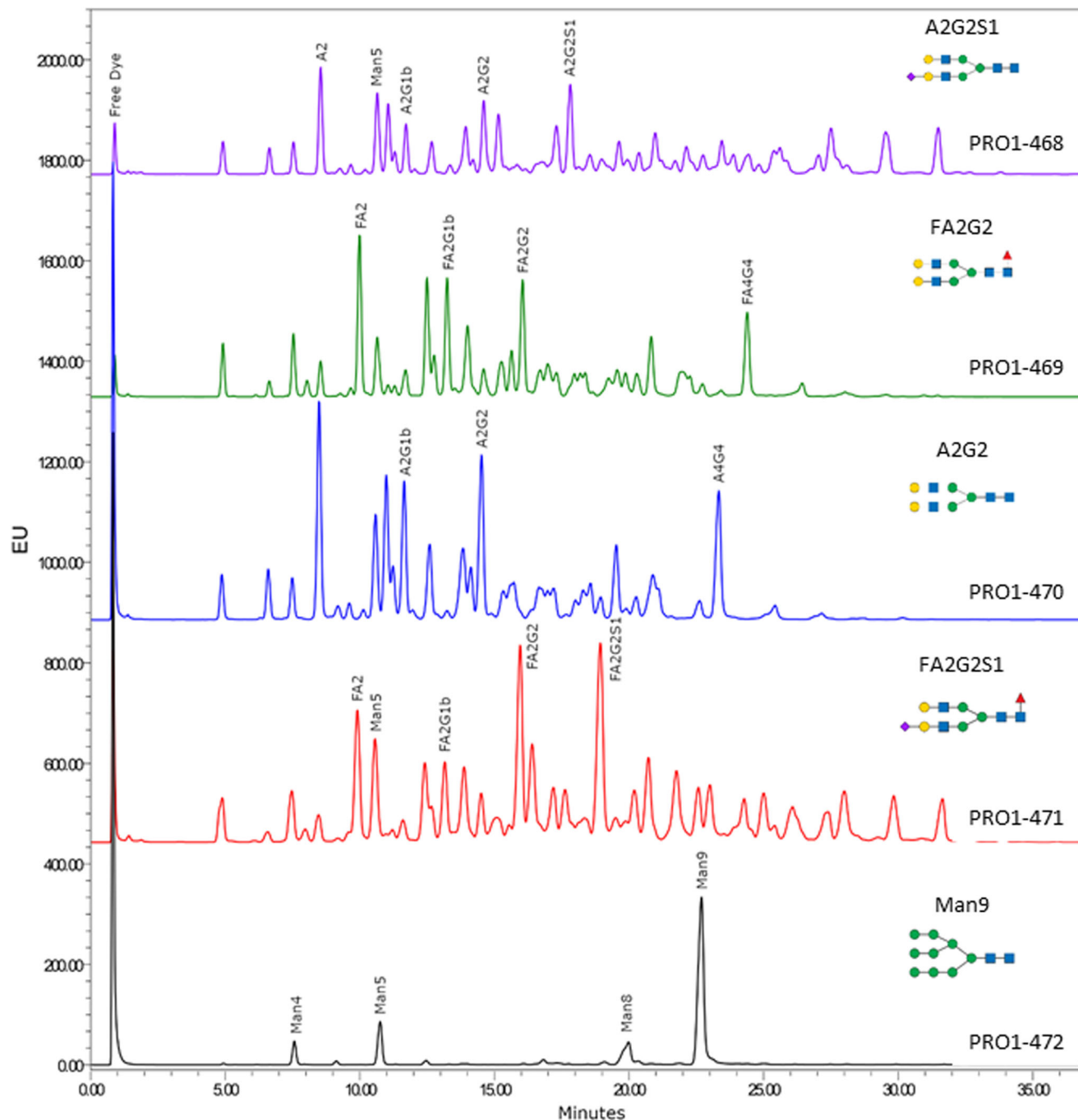
consistent with the previously published data for the spike protein<sup>12,13</sup>. These same monosaccharides were observed in spike produced from the *ST3GAL4* KO (PRO1-469 and PRO1-470), although Gal abundance was approximately 10 mol/mol lower (i.e., 27–29% reduction) than what was observed for WT (PRO1-471). Fuc was not detected in spikes from either of the *FUT8* KO (PRO1-468 and PRO1-470) indicating that their N-glycans are entirely afucosylated. In addition, HPAEC-PAD analysis detected a clear difference in neutral monosaccharide composition between the kifunensin-spike protein (PRO1-472) and the other variants. Man and GlcN were by far the most abundant monosaccharides in the PRO1-472 hydrolysate and the Man content was 2.6 times higher than what was observed for WT spike protein (PRO1-471). This result was expected given that kifunensin, a mannosidase I inhibitor, promotes the expression of high-mannose glycans at the expense of complex and hybrid N-glycans. GalN, the product of deacetylation of GalNAc after TFA hydrolysis, was not detected in any of the samples. GalNAc is a component of O-linked glycans that are attached to the side chains of serines and threonines<sup>33</sup>. The lack of detectable GalN in these hydrolysates suggests that O-linked glycans are either entirely absent from these proteins or are present at levels too low to be detected by this assay.

Next, the acidic monosaccharide content (i.e., Neu5Ac and Neu5Gc) was evaluated by treating the samples with Sialidase A in triplicate and injecting the released sialic acids on an HPAEC-PAD system with a NaOH and a sodium acetate linear-gradient (Fig. 2b, and Supplementary Table 1). As a control, the WT CHO (PRO1-471) exhibited 21.2 mol Neu5Ac mol<sup>-1</sup> protein. PRO1-468 (*FUT8* KO) had a similar Neu5Ac content (26.7 mol Neu5Ac mol<sup>-1</sup> protein). Unsurprisingly, PRO1-469 and PRO1-470 produced in *ST3GAL4* KO cell lines contained little Neu5Ac (2.0 and 2.1 mol Neu5Ac/mol protein, respectively). In these KO cells, the main sialyltransferase (*ST3GAL4*) that transfers 2,3-linked Neu5Ac onto *N*-acetylglucosamine branches is absent although other sialyltransferases can possibly contribute to the low sialylation observed<sup>34</sup>. Similarly, Neu5Ac content is low (1.3 mol Neu5Ac mol<sup>-1</sup> protein) in PRO1-472 (WT CHO supplemented with kifunensin). These glycans do not have terminal Gal which prevents the addition of Neu5Ac<sup>35</sup>. Traces of Neu5Gc was also detected in most samples (Supplementary Table 1) but at very low levels (0.1 mol Neu5Gc mol<sup>-1</sup> protein) confirming that in CHO cells Neu5Ac hydroxylase is present but not expressed as reported previously by Wang et al<sup>36</sup>.

**HILIC-Fld analysis of released N-glycans.** N-glycans were released from the spike proteins PRO1-468 to -472 with PNGase F, labeled with 2-aminobenzamide (2-AB), and separated on a HILIC-UPLC-Fld system (Fig. 3). The retention times were calibrated with a 2-AB labeled dextran ladder, converted to Glc units (GU) values and used to search the GlycoBase database for glycan identities<sup>37</sup>. Usually, glycan retention times are correlated to their polarity, with larger and negatively charged glycans appearing at higher GU values. This is usually a straightforward analysis for glycoproteins with simple N-glycan profiles, such as mAbs<sup>38</sup>. However, all the spike proteins analyzed here (except for PRO1-472), yielded very complex glycan profiles, making it difficult to assign an identity to all but a handful of the major peaks. In contrast to the other HILIC traces, only high-mannose glycans (Man4-9) were detected in PRO1-472 with Man9 being the most abundant. This glycan profile is expected for a glycoprotein expressed in CHO cells in the presence of kifunensin.

Despite challenges in identifying the majority of glycans detected in the HILIC traces, some inferences about the glycosylation of the different spike glycovariants can be made (Fig. 3). For example, the HILIC chromatograms for PRO1-468 and PRO1-471 contain glycan signals with high GU values that are characteristic of sialylated glycans, while these peaks are absent from PRO1-469 and PRO1-470 (*ST3GAL4* KO) and PRO1-472. Furthermore, prominent peaks corresponding to the fucosylated biantennary glycans, FA2, FA2G1, and FA2G2, are detected in PRO1-469 and PRO1-471 while peaks of similar intensity corresponding to afucosylated biantennary glycans, A2, A2G1, and A2G2, are present in PRO1-468 and PRO1-470, implying that fucosylation is impaired or absent in the *FUT8* KO. Note that it was not possible to determine the degree of fucosylation from the HILIC-Fld chromatograms of the N-glycans as most of the peaks could not be identified.

Additionally, it is also possible to draw some conclusions about the relative levels of galactosylation in the spike protein glycovariants. In Fig. 4 the HILIC-Fld chromatogram for *ST3GAL4* KO PRO1-469 and WT PRO1-471 after treatment with sialidase show that for PRO1-471, the biantennary galactosylated FA2G2 peak is higher than the biantennary non-galactosylated FA2 peak. While for PRO1-469 (*ST3GAL4* KO), the opposite is observed with the intensity of FA2 being higher than the intensity of FA2G2. This result aligns with the data from the monosaccharide analysis (Fig. 2a and Supplementary Table 1) showing higher galactosylation for PRO1-471 than PRO1-469. A similar difference in galactosylation is observed for PRO1-470

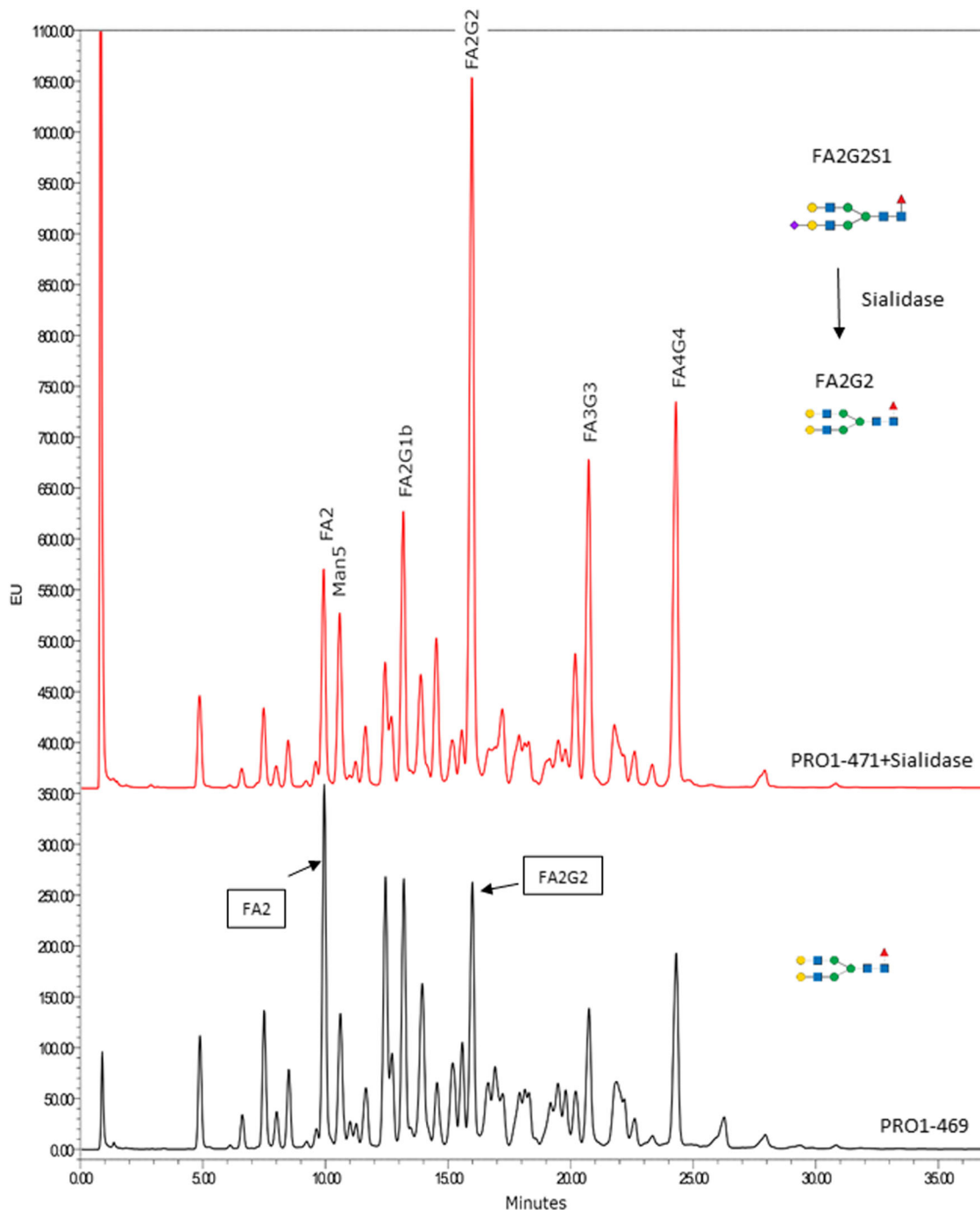


**Fig. 3** HILIC-UPLC-Flid chromatograms of PRO1-468 to PRO1-472 samples. The chromatograms of PRO1-468 (purple), PRO1-469 (green), PRO1-470 (blue), PRO1-471 (red), and PRO1-472 (black) are represented with no offset on the x axis, but offsets on the y-axis to facilitate comparison. An expected representative structure is depicted on the right. To note, this representative structure indicates only a single version of many different glycoforms present. Glycans are depicted by the symbolic nomenclature for glycans using DrawGlycan-SNFG software<sup>47</sup>. Green circle: Man, blue square: GlcNAc, yellow circle: Gal, red triangle: Fuc, purple diamond: Neu5Ac.

and PRO1-468 in the monosaccharide analysis (Supplementary Fig. 1) and the relative abundance of biantennary glycan A2G2 vs A2 in the de-sialylated HILIC-Flid results. Therefore, it can be concluded that the level of galactosylation of spike protein N-glycans is lower in the *ST3GAL4* KO than in those expressed with active *ST3GAL4*.

**LC-MS glycopeptide analysis.** Next, the spike proteins were analyzed by bottom-up LC-MS to characterize and compare the glycosylation at each of the N-glycosylation sites. In keeping with

the approaches taken by other groups<sup>12–15,19,39</sup> a protocol using four proteases (trypsin, chymotrypsin, alpha-lytic protease, and glutamyl endopeptidase Glu-C) was employed to ensure that each N-glycosylation site could be isolated on its own and on a peptide that was compatible with LC-MS. Glycopeptides with 2 or more N-glycosylation sites were ignored in this analysis. Additionally, a portion of each spike protein digest was treated with PNGase F to remove N-glycans and analyzed by LC-MS/MS. The data acquired for these deglycosylated samples helped to confirm the peptide sequences and to eliminate false positives. GlycoPIQ, an



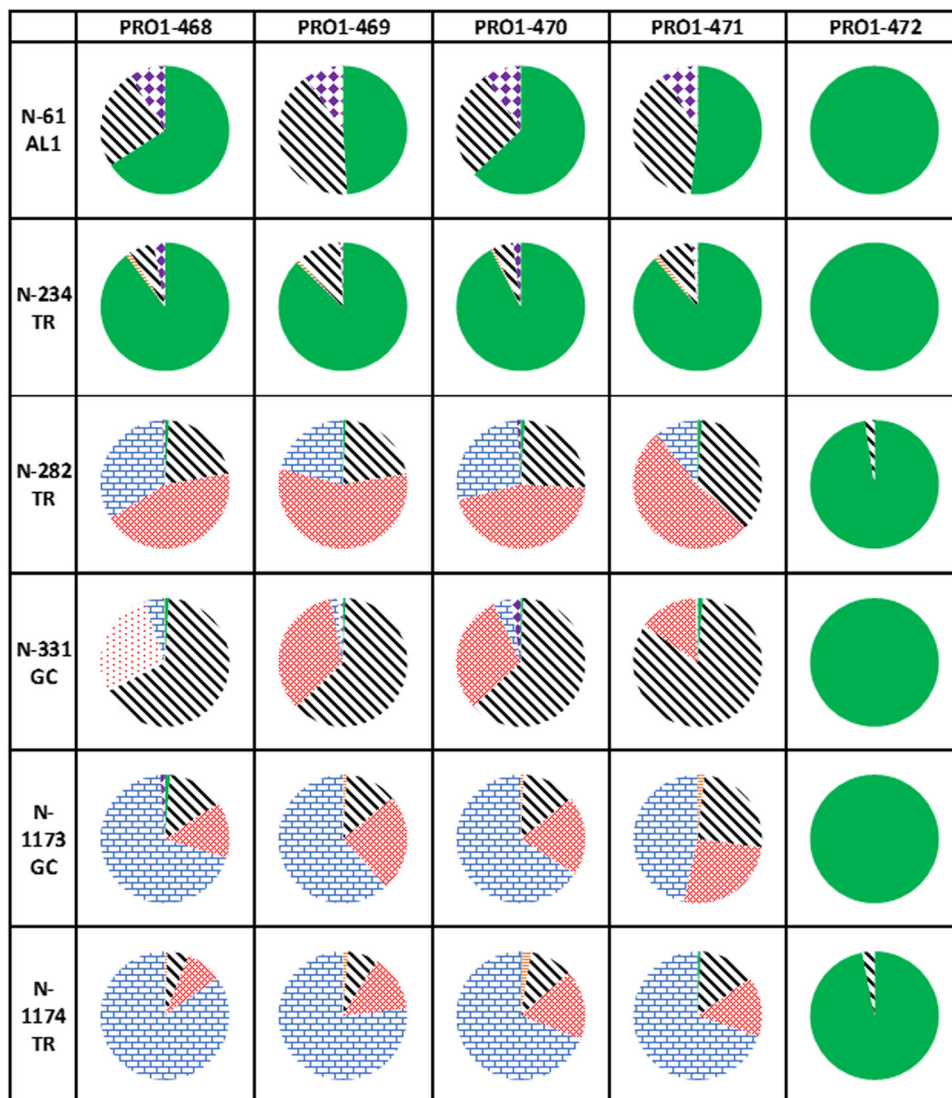
**Fig. 4** Comparison of HILIC-UPLC-FLD chromatograms with no offset on the x axis, but offsets on the y axis to facilitate comparison of PRO1-471 (red) after sialidase treatment and PRO1-469 (non-treated, black) with expected typical structure. Glycans depicted as by the symbolic nomenclature for glycans using DrawGlycan-SNFG software<sup>47</sup>. Green circle: Man, blue square: GlcNAc, yellow circle: Gal, red triangle: Fuc, purple diamond: Neu5Ac.

in-house algorithm, was used to identify the glycans at each site and to determine their relative abundances (Supplementary Data 1).

Glycopeptides that matched the above criteria were found for 21 of the 22 N-glycosylation sites (Supplementary Table 4). Unfortunately, due to poor intensities and interference from other ions, neither peptides nor glycopeptides covering N-glycosylation site N1134 could be detected with confidence in all of the spike sample digests. The remaining 21 N-glycosylation sites were modified with a range of high-mannose, complex, and/or hybrid glycans of differing relative abundances. Some glycosylation sites were observed partially

unoccupied. A tabulation of the glycans observed on the 21 N-glycosylation sites is presented in Supplementary Table 5. Also, as an example, the glycopeptide categories identified by GlycoPIQ for 6 of the 21 N-glycosylation sites are presented in Fig. 5. The size of each segment in the pie charts is determined by the sum of the relative intensities of the glycopeptides assigned to that category.

The overall glycosylation patterns observed for the WT spike protein (PRO1-471) aligned well with the average glycosylation for recombinant spike proteins reported in other studies<sup>12-14</sup>. High-mannose glycans were predominant on-site N234 and to a lesser extent on N61 and N717. The remainder of the sites were



**Fig. 5 Example of glycan distribution per site.** Pie charts showing how glycans are distributed for six different sites. AL:  $\alpha$ -lytic protease, GC: Glu-C, TR: trypsin. biantennary (black, wide diagonal stripes downward), triantennary (red, 30% dotted), tetra-antennary (blue, horizontal brick), hybrid (orange, narrow horizontal stripes), other glycans (includes glycans such as A1 and A1G, purple, solid diamond grid) and high-Man (green, solid fill).

occupied mostly with complex glycans. Interestingly, tri- and tetra-antennary glycans were more prevalent on sites nearer to the C-terminus (Supplementary Table 5). As expected, only trace amounts of sialylated glycopeptides were detected in the digests of the *ST3GAL4* KO spike proteins, PRO1-469 and PRO1-470. In contrast, most of the sites on both PRO1-468 and PRO1-471, with the exception of those dominated by high-mannose glycans, were occupied to some degree with sialylated glycans. Sialylation levels were greater on those sites with more multi-antennary glycans. Interestingly, PRO1-468 displayed a greater level of sialylation than PRO1-471. This is in keeping with the fact that PRO1-468 also expressed more of the larger, tetra-antennary glycans as was observed by HILIC-Fld and LC-MS (Supplementary Fig. 1 and Table 5).

Interestingly, the glycosylation differences between proteins observed at the individual site level by LC-MS, were also reflected in the overall monosaccharide and N-glycan analyses. For example, more sialylated glycans were consistently observed on PRO1-468 compared to PRO1-471 by LC-MS and corresponded to the greater overall abundance of Neu5Ac in the HPAEC-PAD results (26.7 vs 21.2 mol Neu5Ac mol<sup>-1</sup> monomer, respectively).

Additionally, the shift to glycans with a greater number of antennae that was observed by LC-MS in PRO1-468 relative to PRO1-471 was manifested in the HPAEC-PAD results as a greater GlcN concentration in PRO1-468 (102.5 vs 95.6 mol GlcN mol<sup>-1</sup> monomer for PRO1-468 and PRO1-471, respectively). This same trend was also observed by HILIC-Fld analysis once the complexity of the N-glycan profiles was reduced by sialidase treatment (Supplementary Fig. 2).

Similar to HPAEC-PAD and HILIC-Fld, glycopeptide LC-MS also detected the trend towards lower relative intensities of digalactosylated biantennary glycans relative to the corresponding non-galactosylated biantennary glycans for the *ST3GAL4* KO's (PRO1-469 and PRO1-470) when compared with the equivalent de-sialylated glycans from PRO1-468 and PRO1-471. The trend was less obvious due to the inherent complexity of the LC-MS results. Nevertheless, by averaging the proportion of Gal per site and summing for all sites (Supplementary Table 11, note that it is only an estimation and not a quantitative measure), it was apparent that the apparent galactosylation was lower for PRO1-469 and PRO1-470 (11.9% and 12.7%, respectively) than for PRO1-468 and PRO1-471 (15.0% and 14.0%, respectively). These

results are in remarkably good agreement with those obtained with HPAEC-PAD.

LC-MS analysis confirmed that fucosylated glycopeptides were absent from the digests of the spike proteins expressed in the *FUT8* KO cell lines (PRO1-468 and PRO1-470) as well as from the kifunensin-treated spike protein (PRO1-472). The complex-type glycans in the two remaining glycovariants (PRO1-469 and PRO1-471) were predominantly fucosylated. However, a noticeable proportion of the glycans on N1098 were afucosylated in both variants. The reason for this is not understood, although the same phenomenon has been observed by others<sup>12</sup>. It is notable that fucosylation levels were very similar for PRO1-469 and PRO1-471, implying that the presence or absence of sialylation does not impact fucosylation.

The same basic pattern of glycosylation was observed in all of the spike protein digests with the exception of PRO1-472. This suggests that knocking out *FUT8* (loss of core Fuc) and *ST3GAL4* (loss of 2–3-linked Neu5Ac) minimally disrupt glycan maturation and processing in the trans-golgi. As expected, all of the N-glycosylation sites on PRO1-472 (kifunensin-treated spike protein) were occupied with high-mannose glycans, with Man9 being the predominant glycan species.

**Glycosylation analysis of spike protein from independent production batches.** Three independent spike protein batches (PRO1-392, PRO1-394, and PRO1-412) produced from the same CHO pool were characterized using the neutral monosaccharide and sialic acid assays to determine if significant differences in monosaccharide content could be detected between them. The three batches were expressed in the same cell line (CHO<sup>55E1</sup>) and under the same conditions although harvesting was done on different days. As such, the glycan content and composition were expected to be relatively similar for all three batches, providing an opportunity to assess if the HPAEC-PAD monosaccharide assays could be sensitive enough to detect subtle variations in glycosylation. Here again, the monosaccharide results were compared with those from the HILIC-Fld and LC-MS assays.

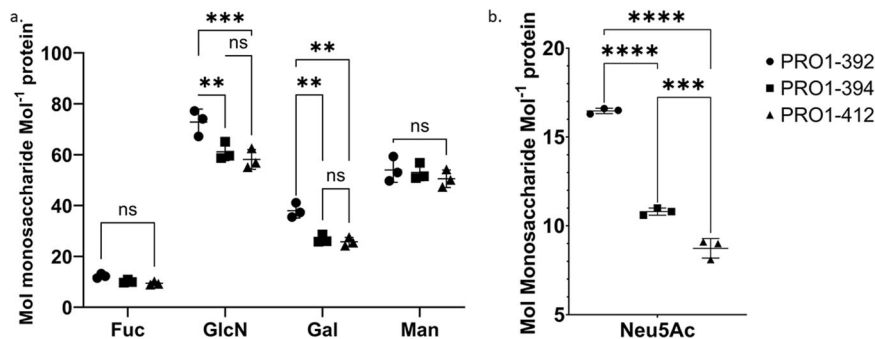
**Monosaccharide analysis by HPAEC-PAD.** Consistent levels of Fuc and Man were observed between all batches. However, GlcN and Gal content was higher in PRO1-392 compared to PRO1-394 and PRO1-412 (Fig. 6a, Supplementary Table 6). Also, Neu5Ac content was significantly higher for PRO1-392 (16.5 mol Neu5Ac mol<sup>-1</sup> protein) compared to PRO1-394 (10.8 mol Neu5Ac mol<sup>-1</sup> protein) and PRO1-412 (8.7 mol Neu5Ac mol<sup>-1</sup> protein) (Fig. 6b,

Supplementary Table 6). We looked at the production record for each batch to better understand the variation in Neu5Ac content, (Supplementary Table 12). A range of factors can affect sialylation during fed-batch processes, including cell viability and/or fed-batch duration<sup>40</sup>. Indeed, cell viability at harvest time for PRO1-392 (96.5%) was slightly higher compared to that of PRO1-394 (95.2%) while PRO1-412 had the lowest viability (94.3%) (Supplementary Table 12). A possible explanation for the higher Neu5Ac content of PRO1-392 could be that the culture was harvested earlier than for the other two (day 7 for PRO1-392, day 10 for PRO1-412, and day 11 for PRO1-394). It is possible that more sialidase was released from dead cells and/or had more time to act on the sialylated glycans over the longer culture durations.

**HILIC-Fld analysis of released N-glycans.** The HILIC-Fld chromatograms of PRO1-394 and PRO1-412 are very similar, although Man5 is the dominant peak (14%) in PRO1-394 while FA2G2 is the dominant peak (14.5%) in PRO1-412 (Fig. 7). The chromatogram of PRO1-392 is distinct in having a Man5 peak of much smaller relative intensity (5.7%) than for the other two samples (14% for PRO1-394 and 11.8% for PRO1-412, respectively). Interestingly, the HPAEC-PAD monosaccharide assay did not detect a significant difference in Man content between PRO1-392 and the other batches. This was anticipated as Man is a component of all N-glycans and so a reduction in high-mannose glycan content is offset, at least partially, by an increase in other N-glycan species that contain 3 Man residues each. Finally, the dominant peak in PRO1-392 is the mono-sialylated FA2G2S1 glycan (16.6%), which suggests the highest level of sialylation than to the other two batches. Here again, this is consistent with the data from the HPAEC-PAD monosaccharide assays.

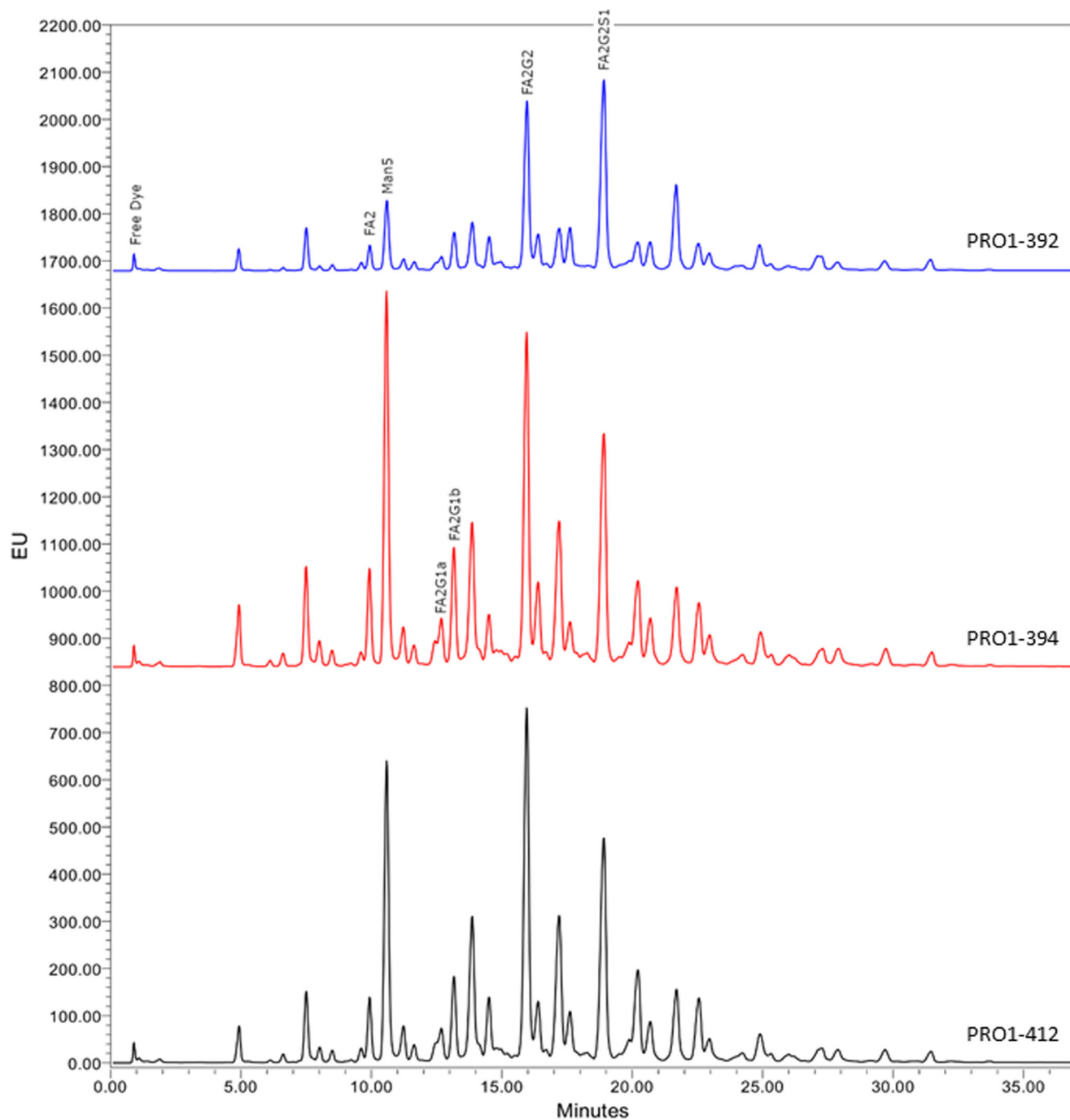
**LC-MS glycopeptide analysis.** Reliable LC-MS glycopeptide results were obtained for 20 of the 22 N-glycosylation sites for the three independent production batches (Supplementary Data 2, Supplementary Table 9). Here again, due to poor signal intensities and interference from other ions, glycopeptides could not be reliably identified and quantified for N709 or N1134 in all three batches.

Glycopeptide relative abundances revealed close similarity between PRO1-394 and -412. The PRO1-392 complex-type glycans were more sialylated and galactosylated than the other two batches and also possessed a larger proportion of tri- and tetra-antennary glycans (Supplementary Tables 9 and 10). This is



**Fig. 6 Monosaccharide analysis via HPAEC-PAD of three independent spike batches.** Results reported in mol monosaccharide mol<sup>-1</sup> monomer (MW = 143,230.6 Da). Results are the means of three triplicate injections and three triplicate reactions. Error bars are the standard deviation observed between the triplicate hydrolyses values. The data were analyzed using two-way ANOVA, Tukey's multiple comparisons test, comparing the content of each monosaccharide type between variants. Not all comparisons are shown. For full comparison see Supplementary Table 7. **a** Fuc, GlcN, Gal, and Man content obtained after TFA hydrolysis. **b** Neu5Ac content obtained after enzymatic hydrolysis. The data were analyzed using a one-way ANOVA, Tukey's multiple comparisons test, showing all comparisons. For exact *p* values see Supplementary Table 8. *P* values: <0.002 (\*\*), <0.0002 (\*\*\*) and <0.0001(\*\*\*\*).





**Fig. 7 Comparison of HILIC-UPLC-Fluorimetric Detection (Fld) chromatograms of 3 batches of the spike protein produced in CHO pools.** The three HILIC traces for samples PRO1-392 (blue), PRO1-394 (red), and PRO1-412 (black) with offsets on the y axis to facilitate comparison and no offsets on the x axis are depicted.

consistent with data generated by both HPAEC-PAD and HILIC-Fluorimetric Detection (Fld). In addition, the relative abundance of high-mannose glycopeptides was noticeably lower in PRO1-392 compared with the other two batches. As mentioned previously, the change in overall Man abundance in PRO1-392, as determined by monosaccharide analysis, is subtle due to the lower high-mannose glycan content being offset by the increase in complex and hybrid glycans, both of which contain mannose. This makes it difficult for the HPAEC-PAD monosaccharide assays to detect a significant change. On the other hand, increases in the Gal and GlcNAc content of PRO1-392 were detected by monosaccharide analysis and are indicative of the increased abundance of larger and/or more abundant complex-type glycans in PRO1-392.

## Discussion

For the reasons described previously, established assays struggle with the routine monitoring of the complicated and heterogeneous glycans found on highly glycosylated antigens such as the SARS-CoV-2 spike glycoprotein. This can lead to gaps in our ability to compare glycosylation across multiple lots and in our understanding of how variation in glycosylation affects immunogenicity. Furthermore, there is little established guidance available for characterizing highly complex glycoproteins, and early discussions with regulatory authorities are usually required in order to determine the analytical package required<sup>27</sup>.

The HPAEC-PAD monosaccharide assays are fast and sample preparation is simple, with no purification and derivatization steps

(Fig. 1). Furthermore, the results can be obtained in a few days without complex processing or calculations. Much effort has been put into addressing issues relating to the reliability of the HPAEC-PAD instrumentation<sup>27</sup>. In our experience, the technology has proven to be both reliable and robust. The goal of this investigation was to determine if batch-to-batch variations in spike protein glycosylation could be monitored effectively using the HPAEC-PAD monosaccharide assays. Two sets of engineered spike proteins, one with major and the other with minor differences in N-glycosylation, were analyzed by HPAEC-PAD and the results were compared with those obtained by glycan HILIC-Fld and glycopeptide LC-MS.

Glycan analysis by HILIC-Fld is regularly used for N-glycan batch monitoring of recombinant immunoglobulin Gs as their glycosylation profile is usually relatively simple. This simplicity stems from usually having just one N-glycosylation site on each heavy chain that is modified with biantennary glycans displaying limited variability. The HILIC-Fld assay requires the release, labeling, and purification of the glycan modifications, a process that is more elaborate than for HPAEC-PAD. Importantly, the HILIC-Fld data can be challenging to interpret for glycoproteins with multiple and heterogeneous N-glycans due to substantial peak overlap and the lack of suitable glycan standards for identification. As a result, it is often not possible to draw a precise picture of the glycosylation on these proteins. For example, many of the glycans released from the spike proteins used in this study could not be identified by HILIC-Fld and so it was not possible to determine their overall level of fucosylation or sialylation. In contrast, these were simple measurements to make via HPAEC-PAD.

The monosaccharide results generated via HPAEC-PAD were also compared with glycopeptide LC-MS data. Bottom-up LC-MS is typically used for the in-depth characterization of complex proteins and has the important advantage of being able to investigate glycosylation at individual sites. However, it is a slow, complex assay involving many processing and analysis steps (Fig. 1). Four different proteases were used to maximize the likelihood that each N-glycosylation site could be isolated on its own on a peptide that was compatible with LC-MS. As such it is unsuitable for the rapid and routine batch-to-batch monitoring of complex glycoproteins such as the SARS-CoV-2 spike. Furthermore, the resulting glycopeptide data sets are very complex which can often make it difficult to make important comparisons between clones or batches. For example, the lower galactose content of the non-sialylated spike variants was readily observed by the HPAEC-PAD monosaccharide analysis but was easy to overlook in the complex glycopeptide LC-MS data.

In conclusion, we demonstrate that it is possible to employ the HPAEC-PAD monosaccharide assays to monitor batch-to-batch changes in glycosylation for the SARS-CoV-2 spike glycoprotein antigens used for vaccine development. The results compared well with those obtained by glycan HILIC-Fld and glycopeptide LC-MS. In addition, we demonstrated that monosaccharide analysis could detect subtle glycosylation differences between batches of the same antigen expressed in the same cell line. Importantly, the monosaccharide assays were noticeably faster to perform and the results were quantitative, reproducible, and easy to interpret. As such, the monosaccharide assays have many of the features required for use in a QC environment. It can be very challenging to fully analyze glycan post-translational modifications and, certainly, HILIC-Fld and LC-MS are more suitable for the in-depth analysis of complex glycoproteins. Nevertheless, the HPAEC-PAD monosaccharide assays appear to be the better choice for monitoring batch-to-batch glycosylation of complex glycoproteins.

Ultimately, the strategic implementation of these monosaccharide assays could shorten the time it takes to develop a vaccine, an important consideration when responding to a future pandemic.

## Methods

**Production of the glycovariants (PRO1-468, PRO1-469, PRO1-470, PRO1-471, and PRO1-472).** Codon optimized DNA sequence encoding the ancestral (Wuhan), Genbank accession number: MN908947, SARS-CoV-2 spike protein<sup>32</sup>, was cloned into the pTT5<sup>®</sup> vector<sup>41</sup> by Genscript. The sequence of the newly designed construct was verified by Sanger DNA sequencing.

Different KO CHO cell lines, including afucosylated (F15), asialylated (S9) as well as afucosylated and asialylated (dKO2) cells, all derived from CHO<sup>55E1</sup> WT cells<sup>32</sup> using CRISPR/Cas9<sup>42</sup> (Supplementary Method 1), were used for protein expression. High-density transfection and post-transfection culture of cells were performed (Supplementary Method 2)<sup>43</sup>. Cells were cultured in a chemically-defined proprietary media formulation supplemented with 4 mM L-glutamine and incubated in shake flasks (Corning, NY, USA) under agitation (120 rpm) at 37 °C and with 5 % CO<sub>2</sub>. Two days prior to transfection, cells were seeded at  $1 \times 10^6$  cells mL<sup>-1</sup> in the same media to achieve a cell density of  $\sim 8 \times 10^6$  mL<sup>-1</sup> on the day of transfection. Right before transfection, cells were diluted with 25% fresh media, and dimethylacetamide was added to 0.083 % (v v<sup>-1</sup>). polyethylenimine -Max (Polysciences) was used to transfect cells at a DNA:polyethylenimine ratio of 1:7 (w:w) and plasmid DNA final concentration was 28 µg mL<sup>-1</sup> in cell culture media. The transfected DNA was a mix of 85% of the pTT5-spike construct, 10% pTT-Bcl-XL (anti-apoptotic effector), and 5% pTT-GFP were indicated, 10 µM (final) kifunensine (Toronto Research Chemicals, cat# K450000) was added to the culture 5 h post-transfection. At 24 h post-transfection, cells were moved to a 32 °C incubator and supplemented with Anti-Clumping Supplement (1:500 dilution) (Irvine Scientific) and Feed 4 (2.5% v v<sup>-1</sup>) (Irvine Scientific). At 5 days post-transfection, cultures were fed with 5% of Feed 4, and additional Glc was added when needed to maintain a minimal concentration of 17 mM. Cell supernatants were collected at 7 days post-transfection.

Recombinant spike glycovariants were purified using AVIPure-COV2S resins (Avidite, Lebanon, New Hampshire) according to the manufacturer's instructions. Briefly, 0.2 µm-filtered supernatants were loaded on columns equilibrated with calcium- and magnesium-free Dulbecco's PBS (DPBS). After the initial wash with 10 column volumes of DPBS, proteins were eluted using a solution containing 50 mM Bis-Tris and 1 M Arginine-HCl at pH 6.0. The fractions containing eluted proteins were pooled and the elution buffer was buffer-exchanged for DPBS, using desalting columns (GE Healthcare). Purified proteins were visualized using SDS-Page (Supplementary Fig. 3) and quantified by spectrophotometry (A<sub>280</sub>) using extinction coefficients calculated based on their amino acid composition.

**Production of spike from stable CHO pool PRO1-392, PRO1-394, and PRO1-412.** Batches of SARS-CoV-2 spike protein stably produced from a stable CHO pool (named PRO1-394, PRO1-392, and PRO1-412), were generated independently from a unique frozen cell vial (Supplementary Method 3)<sup>44,45</sup>. Culture conditions/characteristics for each batch are described in Supplementary Table 12.

**Analyses.** For analysis design, see Supplementary Method 4.

## HPAEC-PAD measurements

**Neutral sugar analysis.** The dried protein sample (25 µg) was heated, in triplicate with TFA (160 µL, 5 M) (Sigma, cat# 91707) in an oven at 100 °C for 2 h. The sample was allowed to cool and then was evaporated to dryness (Genevac EZ-2 Elite). The sample

was then re-suspended in water (125  $\mu\text{L}$ ), centrifuged at 14,000 rpm for 2 min, and then diluted with water 1:20 (v:v). 10  $\mu\text{L}$  injections were eluted with a stepwise gradient of NaOH and were repeated in triplicate on Dionex ICS-6000 system equipped with a CarboPac PA20 column (3  $\times$  150 mm, 060142), CarboPac PA20 guard column (3  $\times$  30 mm, 060144), a BorateTrap column (4  $\times$  50 mm, 047078) and a pulsed amperometric detector<sup>23</sup>.

**Sialic acid analysis.** The dried protein sample (25  $\mu\text{g}$ ) was treated in triplicate with 5 $\times$  Buffer B (10  $\mu\text{L}$ ) and Sialidase A (2  $\mu\text{L}$ ) (Agilent, GK80040) and adjusted to a final concentration of 0.48  $\mu\text{g}/\mu\text{L}$ . The sample was kept at 37  $^{\circ}\text{C}$  for 18 h. The sample were then spun (14,000 rpm, 2 min), diluted 1:20 water, and 25  $\mu\text{L}$  triplicate injections were eluted with a linear gradient of NaOH/sodium acetate buffer on a Dionex ICS-6000 system equipped with a CarboPac PA20 column (3  $\times$  150mm, 060142), CarboPac PA20 guard column (3  $\times$  30 mm, 060144) and a pulsed amperometric detector<sup>46</sup>.

Both neutral sugar and sialic acid data collection and analysis have been performed with Chromeleon 7 (ThermoFisher Scientific). The data was processed using GraphPad Prism 9.0.2.

**HILIC-Fld.** Glycans were released from the glycoproteins using Rapid™ PNGase F (New England BioLabs, cat# P0710) at 50  $^{\circ}\text{C}$  for 10 min, after a 2 min at 80  $^{\circ}\text{C}$  denaturation step. The PNGase F and protein were removed using a Discovery Glycan SPE 50 mg column (Millipore-Sigma, cat# 55465-U), and the glycans were evaporated to dryness under vacuum. The glycans were then labeled with 2-AB (Sigma-Aldrich, cat# PP0520)<sup>37</sup>. The excess dye was removed on PD MiniTrap G-10 (Cytiva, cat# 28-9180-10). The sample was divided into two aliquots. One aliquot was used as is for glycan analysis while the other one was treated with a recombinant sialidase (MNV-01, expressed and purified in-house) prior to analysis, in order to identify peaks containing sialic acid. Glycans were analyzed by HILIC with fluorescent detection using an Acquity UPLC wide-pore glycoprotein Ethylene Bridged Hybrid amide, 300  $\text{Å}$ , 1.7  $\mu\text{m}$ , 2.1  $\times$  150 mm column (Waters Corp., Cat# 176003702) heated to 60  $^{\circ}\text{C}$  and a flow rate of 0.5  $\text{mL min}^{-1}$ . Glycans were eluted using a gradient of 100 mM ammonium formate, pH 4.5 (mobile phase A), and 100% acetonitrile (mobile phase B), starting with an initial ratio of 22:78 and followed by a gradient of 44:56 over 43.5 min. The data was collected and analyzed using Xcalibur, version 4.4.16.14. Peaks were calibrated with a 2-AB labeled dextran ladder standard (Waters Corp., cat# 186006841) and compared to GU values in the database GlycoBase. NIBRT ([https://glycobase.nibrt.ie/glycobase/show\\_nibrt.action](https://glycobase.nibrt.ie/glycobase/show_nibrt.action)). A recombinant broad-specificity sialidase was used to identify peaks containing sialic acid. The traces resulted in a non-quantifiable fingerprint.

### LC-MS/MS analyses

**In-solution digestion.** Each of the variants (60  $\mu\text{g}$ ) was denatured with 0.1% Rapigest (Waters) in 150  $\mu\text{L}$  of 50 mM ammonium bicarbonate (Sigma). Disulfide bonds were reduced with dithiothreitol (9.1 mM) (Sigma) at 80  $^{\circ}\text{C}$  for 10 min then alkylated with iodoacetamide (34.1 mM) (Sigma) at room temperature for 30 min. Excess iodoacetamide was quenched with 15  $\mu\text{L}$  of 100 mM DTT (15 min at room temperature) to prevent over-alkylation. Samples were buffer-exchanged into 50 mM ammonium bicarbonate using 7 K Zeba spin-columns (Thermo Scientific), following product instructions, and then diluted to a concentration of 0.1  $\mu\text{g}/\mu\text{L}^{-1}$  with ammonium bicarbonate (50 mM). Half of each sample (30  $\mu\text{g}$  protein) was treated with PNGase F (3 U) (Sigma) and incubated at 37  $^{\circ}\text{C}$  overnight to deglycosylate. Both deglycosylated and untreated halves were

separately divided into four equal parts (7.5  $\mu\text{g}$  each) that were treated with either trypsin (0.5  $\mu\text{g}$ ) (Promega), Glu-C (0.5  $\mu\text{g}$ ) (Promega),  $\alpha$ -lytic protease (0.4  $\mu\text{g}$ ) (Sigma) or chymotrypsin (0.25  $\mu\text{g}$ ) (Sigma). The chymotrypsin digests were incubated overnight at 25  $^{\circ}\text{C}$  while the other digests were incubated overnight at 37  $^{\circ}\text{C}$ . All digests were acidified with a 1/20 volume of 10% TFA (Sigma), incubated at 37  $^{\circ}\text{C}$  for 40 min, and centrifuged at 13,000 rpm for 15 min to cleave and remove Rapigest. Each digest (2  $\mu\text{L}$ ) was diluted with 0.1% formic acid (10  $\mu\text{L}$ ) for injection. The trace

**LC-MS.** Peptide digests (10  $\mu\text{L}$ ,  $\sim$ 1.3 pmol) were injected on a C18 PepMap100 5  $\mu\text{m}$  trap (Thermo Scientific) installed before a nanoACQUITY Ethylene Bridged Hybrid 1.7  $\mu\text{m}$  column (100  $\mu\text{m}$   $\times$  100 mm C18, Waters) in an UltiMate 3000 nano-LC system (Thermo Scientific) and detected on an Orbitrap Eclipse Tribrid mass spectrometer equipped with an electrospray ionization source (Thermo Scientific). The peptides were eluted with a gradient of 0.1% formic acid in ddH<sub>2</sub>O (Mobile phase A) and 0.1% formic acid in acetonitrile (Mobile phase B). Peptides were eluted over a linear gradient of 0.2% to 40% mobile phase B over 89 min, 40% to 95% mobile phase B over 4 min, followed by re-equilibration, with a flow rate of 0.5  $\mu\text{L min}^{-1}$ . MS spectra were acquired on the Orbitrap from 350 to 2000  $m/z$  in positive electrospray ionization mode at a 120 K resolution. The most intense ions (threshold = 2e6, dynamic exclusion = 25 sec) were selected in the quadrupole for HCD-MS/MS (isolation window = 1.6  $m/z$ ) and a fixed HCD activation energy was selected based on peptide  $m/z$  and charge state using a decision tree. MS/MS spectra were acquired in the Orbitrap at 15 K resolution in profile mode with a 1 sec cycle time. The data was recorded using Xcalibur, version 4.4.16.14.

**Data analysis.** The data files were manually inspected to identify (glyco)peptides covering each putative N-glycosylation site (Supplementary Table 4). Peptides covering 2 or more N-glycosylation sites were rejected to ensure each site could be isolated on its own. Glycopeptide identity was confirmed by manual inspection of the HCD-MS/MS spectrum and expected retention time windows for related glycopeptides were defined from the manually identified glycopeptides. The data were processed using GlycoPIQ v2.1, an in-house software (see Supplementary Methods 5) that allows the identification and quantification of a large number glycopeptides within nano-LC-MS/MS data sets. The data was then visualized using Excel.

**Reporting summary.** Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The data are included in the paper and/ or its supplementary information files. The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files (Supplementary Data 1 and Data 2).

### Code availability

In-house software GlycoPIQ v2.1 is available upon request and is similar to other free data processing software such as FragPipe. [https://fragpipe.nesvilab.org/docs/tutorial\\_glyco.html](https://fragpipe.nesvilab.org/docs/tutorial_glyco.html).

Received: 3 December 2022; Accepted: 18 August 2023;  
Published online: 08 September 2023

## References

- Dai, L. & Gao, G. F. Viral targets for vaccines against COVID-19. *Nat. Rev. Immunol.* **21**, 73–82 (2021).
- Djaileb, A. et al. Cross-validation of ELISA and a portable surface plasmon resonance instrument for IgG antibody serology with SARS-CoV-2 positive individuals. *Analyst* **146**, 4905–4917 (2021).
- Croset, A. et al. Differences in the glycosylation of recombinant proteins expressed in HEK and CHO cells. *J. Biotechnol.* **161**, 336–348 (2012).
- Huang, H.-Y. et al. Vaccination with SARS-CoV-2 spike protein lacking glycan shields elicits enhanced protective responses in animal models. *Sci. Transl. Med.* **14**, eabm0899 (2022).
- Lin, S.-C. et al. Different immunity elicited by recombinant H5N1 hemagglutinin proteins containing pauci-mannose, high-mannose, or complex type N-glycans. *PLoS One* **8**, e66719 (2013).
- Deng, T. et al. Characterization and immunogenicity of SARS-CoV-2 spike proteins with varied glycosylation. *Vaccine* **40**, 6839–6848 (2022).
- Dowling, W. et al. Influences of glycosylation on antigenicity, immunogenicity, and protective efficacy of ebola virus GP DNA Vaccines. *J. Virol.* **81**, 1821–1837 (2007).
- Henderson, E. A. et al. Investigation of the immunogenicity of Zika glycan loop. *Virol. J.* **17**, 43 (2020).
- Administration, F. A. D. Development of therapeutic protein biosimilars: comparative analytical assessment and other quality-related considerations <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/development-therapeutic-protein-biosimilars-comparative-analytical-assessment-and-other-quality> (Food and Drug Administration, Rockville, 2019).
- Goh, J. B. & Ng, S. K. Impact of host cell line choice on glycan profile. *Crit. Rev. Biotechnol.* **38**, 851–867 (2018).
- Zhang, L., Luo, S. & Zhang, B. Glycan analysis of therapeutic glycoproteins. *MAbs* **8**, 205–215 (2016).
- Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **369**, 330–333 (2020).
- Zhao, P. et al. Virus-receptor interactions of glycosylated SARS-CoV-2 spike and human ACE2 receptor. *Cell Host Microbe* **28**, 586–601 (2020).
- Brun, J. et al. Assessing antigen structural integrity through glycosylation analysis of the SARS-CoV-2 viral spike. *ACS Cent. Sci.* **7**, 586–593 (2021).
- Zhou, D., Tian, X., Qi, R., Peng, C. & Zhang, W. Identification of 22 N-glycosites on spike glycoprotein of SARS-CoV-2 and accessible surface glycopeptide motifs: Implications for vaccination and antibody therapeutics. *Glycobiology* **31**, 69–80 (2021).
- Yao, H. et al. Molecular architecture of the SARS-CoV-2 virus. *Cell* **183**, 730–738 (2020).
- Bangaru, S. et al. Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate. *Science* **370**, 1089–1094 (2020).
- Wang, D. et al. Comprehensive analysis of the glycan complement of SARS-CoV-2 spike proteins using signature ions-triggered electron-transfer/higher-energy collisional dissociation (ET<sub>h</sub>CD) mass spectrometry. *Anal. Chem.* **92**, 14730–14739 (2020).
- Antonopoulos, A. et al. Site-specific characterization of SARS-CoV-2 spike glycoprotein receptor-binding domain. *Glycobiology* **31**, 181–187 (2021).
- Zong, G. et al. A facile chemoenzymatic synthesis of SARS-CoV-2 glycopeptides for probing glycosylation functions. *Chem. Commun.* **57**, 6804–6807 (2021).
- Vandeputte, J. et al. IABS/CEPI platform technology webinar: Is it possible to reduce the vaccine development time? *Biologicals* **71**, 55–60 (2021).
- Fu, D. T. & Oneill, R. A. Monosaccharide composition analysis of oligosaccharides and glycoproteins by high-performance liquid chromatography. *Anal. Biochem.* **227**, 377–384 (1995).
- Rohrer, J. S. Monosaccharide analysis of glycoproteins by high-performance anion-exchange chromatography with pulsed amperometric detection. *In: Applications of ion chromatography for pharmaceutical and biological products* (ed. L. Bhattacharyya, Rohrer, J. S.) 339–350 (Wiley, 2012).
- Hurum, D. C. & Rohrer, J. S. Five-minute glycoprotein sialic acid determination by high-performance anion exchange chromatography with pulsed amperometric detection. *Anal. Biochem.* **419**, 67–69 (2011).
- Harazono, A. et al. A comparative study of monosaccharide composition analysis as a carbohydrate test for biopharmaceuticals. *Biologicals* **39**, 171–180 (2011).
- Reusch, D. et al. Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles-part 1: separation-based methods. *MAbs* **7**, 167–179 (2015).
- Butré, C. I. & Delobel, A. HPLC- and CE-based methods for the characterization of therapeutic glycoproteins. *In: Carbohydrate analysis by modern liquid phase separation techniques* (2nd Edn) (ed. Ziad El Rassi) Ch. 17, 761–814 (Elsevier, 2021).
- Higgins, E. Carbohydrate analysis throughout the development of a protein therapeutic. *Glycoconj. J.* **27**, 211–225 (2010).
- Zhou, Q., Kyazike, J., Edmunds, T. & Higgins, E. Mannose 6-phosphate quantitation in glycoproteins using high-pH anion-exchange chromatography with pulsed amperometric detection. *Anal. Biochem.* **306**, 163–170 (2002).
- Miller, L. M. et al. Heterogeneity of glycan processing on trimeric SARS-CoV-2 spike protein revealed by charge detection mass spectrometry. *J. Am. Chem. Soc.* **143**, 3959–3966 (2021).
- Gstöttner, C. et al. Structural and functional characterization of SARS-CoV-2 RBD domains produced in mammalian cells. *Anal. Chem.* **93**, 6839–6847 (2021).
- Poulain, A. et al. Rapid protein production from stable CHO cell pools using plasmid vector and the cumate gene-switch. *J. Biotechnol.* **255**, 16–27 (2017).
- Darula, Z. & Medzihradsky, K. F. Analysis of mammalian O-glycopeptides—We have made a good start, but there is a long way to go. *Mol. Cell. Proteomics* **17**, 2–17 (2018).
- Kitagawa, H., Mattei, M.-G. & Paulson, J. C. Genomic organization and chromosomal mapping of the Gal $\beta$ 1,3GalNAc/Gal $\beta$ 1,4GlcNAc  $\alpha$ 2,3-sialyltransferase. *J. Biol. Chem.* **271**, 931–938 (1996).
- Elbein, A. D., Tropea, J. E., Mitchell, M. & Kaushal, G. P. Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. *J. Biol. Chem.* **265**, 15599–15605 (1990).
- Xu, X. et al. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* **29**, 735–741 (2011).
- Bigge, J. C. et al. Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid. *Anal. Biochem.* **230**, 229–238 (1995).
- Mellahi, K. et al. Process intensification for the production of rituximab by an inducible CHO cell line. *Bioprocess Biosyst. Eng.* **42**, 711–725 (2019).
- Shajahan, A., Supekar, N. T., Gleinich, A. S. & Azadi, P. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology* **30**, 981–988 (2020).
- Sethuraman, N. & Stadheim, T. A. Challenges in therapeutic glycoprotein production. *Curr. Opin. Biotechnol.* **17**, 341–346 (2006).
- Shi, C. et al. Purification and characterization of a recombinant G-protein-coupled receptor, *Saccharomyces cerevisiae* Ste2p, transiently expressed in HEK293 EBNA1 cells. *Biochemistry* **44**, 15705–15714 (2005).
- Koyuturk, I. et al. High-level production of wild-type and oxidation-resistant recombinant alpha-1-antitrypsin in glycoengineered CHO cells. *Biotechnol. Bioeng.* **119**, 2331–2344 (2022).
- Stuible, M. et al. Rapid, high-yield production of full-length SARS-CoV-2 spike ectodomain by transient gene expression in CHO cells. *J. Biotechnol.* **326**, 21–27 (2021).
- Colwill, K. et al. A scalable serology solution for profiling humoral immune responses to SARS-CoV-2 infection and vaccination. *Clin. Transl. Immunol.* **11**, e1380 (2022).
- Joubert, S. et al. A CHO stable pool production platform for rapid clinical development of trimeric SARS-CoV-2 spike subunit vaccine antigens. *Biotechnol. Bioeng.* **120**, 1746–1761 (2023).
- Rohrer, J. S. Analyzing sialic acids using high-performance anion-exchange chromatography with pulsed amperometric detection. *Anal. Biochem.* **283**, 3–9 (2000).
- Cheng, K., Zhou, Y. & Neelamegham, S. DrawGlycan-SNFG: a robust tool to render glycans and glycopeptides with fragmentation information. *Glycobiology* **27**, 200–205 (2016).

## Acknowledgements

We would like to thank Sam Williamson and Luc Tessier for the preparation and maintenance of the LC-MS instrumentation. A special thanks to Simon Lord-Dufour for the generation of the stable CHO pool and fed-batch productions. The expert contribution of Christian Gervais, Louis Bisson, Brian Cass, and Guillaume Arthus-Cartier for spike purification by IMAC is acknowledged. We would like to acknowledge Dr. Andrew Cox for his constant support and fruitful discussions during this project. We also acknowledge the NRC's Pandemic Response Challenge Program as a funding source for this work.

## Author contributions

J.S., M.G., J.K., and Y.D. secured funding for the work. J.S., M.G., and J.K. wrote the manuscript. J.S. and I.S. supported the development of the Neu5Ac assay. J.S. and F.S.M. developed the neutral sugar assay. F.S.M. performed the analyses on the HPAEC-PAD and J.S. performed the data analyses of neutral sugar and Neu5Ac analyses. I.K. and S.P. produced and purified the glycoproteins under Y.D.'s supervision. M.G., D.B., and M.F.G. performed the HILIC-FLD analyses. J.K., A.S., A.H., and A.R. performed the LC-MS analyses.

## Competing interests

The authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42004-023-00988-1>.

**Correspondence** and requests for materials should be addressed to Janelle Sauvageau.

**Peer review information** *Communications Chemistry* thanks Hyun Joo An and Jingzhong Guo for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© Crown 2023