



# A Linkage-specific Sialic Acid Labeling Strategy Reveals Different Site-specific Glycosylation Patterns in SARS-CoV-2 Spike Protein Produced in CHO and HEK Cell Substrates

Qiong Wang<sup>1</sup>, Yan Wang<sup>2</sup>, Shuang Yang<sup>1,3</sup>, Changyi Lin<sup>4</sup>, Lateef Aliyu<sup>5</sup>, Yiqun Chen<sup>5</sup>, Lisa Parsons<sup>1</sup>, Yuan Tian<sup>1</sup>, Hongpeng Jia<sup>6</sup>, Andrew Pekosz<sup>7</sup>, Michael J. Betenbaugh<sup>5</sup> and John F. Cipollo<sup>1\*</sup>

<sup>1</sup>Laboratory of Bacterial Polysaccharides, Division of Bacterial, Parasitic and Allergenic Products, Center for Biologics Evaluation and Research, Food and Drug Administration, Baltimore, MD, United States, <sup>2</sup>Mass Spectrometry Facility, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, United States, <sup>3</sup>Center for Clinical Mass Spectrometry, School of Pharmaceutical Sciences, Soochow University, Jiangsu, China, <sup>4</sup>Facility for Biotechnology Resources, Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, MD, United States, <sup>5</sup>Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, United States, <sup>6</sup>Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, United States, <sup>7</sup>Department of Molecular Microbiology and Immunology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

## OPEN ACCESS

### Edited by:

Ganglong Yang,  
Jiangnan University, China

### Reviewed by:

Francisco Solano,  
University of Murcia, Spain  
Jeffery M. Sharp,  
Yale University, United States

### \*Correspondence:

John F. Cipollo  
john.cipollo@fda.hhs.gov

### Specialty section:

This article was submitted to  
Chemical Biology,  
a section of the journal  
Frontiers in Chemistry

Received: 02 July 2021

Accepted: 02 August 2021

Published: 24 September 2021

### Citation:

Wang Q, Wang Y, Yang S, Lin C, Aliyu L, Chen Y, Parsons L, Tian Y, Jia H, Pekosz A, Betenbaugh MJ and Cipollo JF (2021) A Linkage-specific Sialic Acid Labeling Strategy Reveals Different Site-specific Glycosylation Patterns in SARS-CoV-2 Spike Protein Produced in CHO and HEK Cell Substrates. *Front. Chem.* 9:735558. doi: 10.3389/fchem.2021.735558

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus utilizes the extensively glycosylated spike (S) protein protruding from the viral envelope to bind to angiotensin-converting enzyme-related carboxypeptidase (ACE2) as its primary receptor to mediate host-cell entry. Currently, the main recombinant S protein production hosts are Chinese hamster ovary (CHO) and human embryonic kidney (HEK) cells. In this study, a recombinant S protein truncated at the transmembrane domain and engineered to express a C-terminal trimerization motif was transiently produced in CHO and HEK cell suspensions. To further evaluate the sialic acid linkages presenting on S protein, a two-step amidation process, employing dimethylamine and ammonium hydroxide reactions in a solid support system, was developed to differentially modify the sialic acid linkages on the glycans and glycopeptides from the S protein. The process also adds a charge to Asp and Glu which aids in ionization. We used MALDI-TOF and LC-MS/MS with electron-transfer/higher-energy collision dissociation (ET<sub>h</sub>cD) fragmentation to determine global and site-specific N-linked glycosylation patterns. We identified 21 and 19 out of the 22 predicted N-glycosites of the SARS-CoV-2 S proteins produced in CHO and HEK, respectively. It was found that the N-glycosite at 1,158 position (N1158) and at 122, 282 and 1,158 positions (N122, N282 and N1158) were absent on S from CHO and HEK cells, respectively. The structural mapping of glycans of recombinant human S proteins reveals that CHO-Spike exhibits more complex and higher sialylation ( $\alpha$ 2,3-linked) content while HEK-Spike exhibits more high-mannose and a small amount of  $\alpha$ 2,3- and  $\alpha$ 2,6-linked sialic acids. The N74 site represents the most abundant glycosite on both spike proteins. The relatively higher amount of high-mannose abundant sites (N17, N234, N343, N616, N709, N717, N801, and N1134) on HEK-Spike suggests

that glycan-shielding may differ among the two constructs. HEK-Spike can also provide different host immune system interaction profiles based on known immune system active lectins. Collectively, these data underscore the importance of characterizing the site-specific glycosylation of recombinant human spike proteins from HEK and CHO cells in order to better understand the impact of the production host on this complex and important protein used in research, diagnostics and vaccines.

**Keywords:** sialic acid, amidation, SARS-CoV-2, glycoproteomics, site occupancy, N-glycosylation, glycan shield, cell substrate

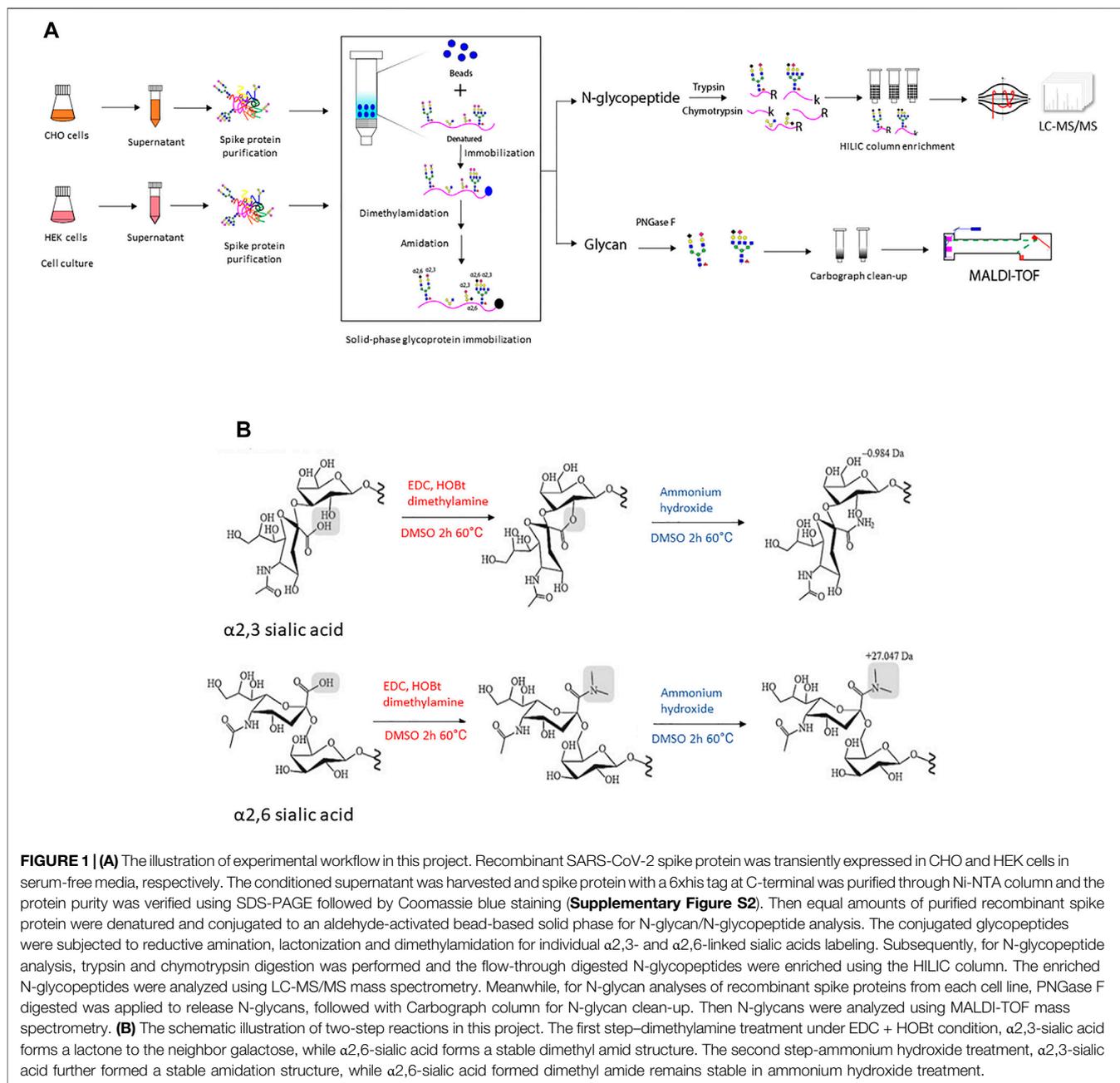
## INTRODUCTION

The ongoing outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) remains a major global pandemic affecting the lives of billions of people worldwide (Yuen et al., 2020). The pathogen of SARS-CoV-2 coronavirus causes fever and cough, shortness of breath, and hospitalization for pneumonia (Guo et al., 2020; Wiersinga et al., 2020). In view of its mortality, human-to-human and cross-species transmission ability, and virus mutation rate, there is a continuing need for effective diagnosis, efficient therapeutic treatment, and preventive means, including vaccines against this devastating disease (Wiersinga et al., 2020). As an enveloped virus, SARS-CoV-2 utilizes an extensively glycosylated spike (S) protein protruding from the viral envelope to bind to angiotensin-converting enzyme-related carboxypeptidase (ACE2) as its primary receptor (Bangaru et al., 2020; Wrapp et al., 2020). Current diagnostics, vaccine development, and neutralizing antibodies are all focused on S glycoprotein, which is the main humoral immune response target (Bangaru et al., 2020).

As a trimeric class I fusion protein, SARS-CoV-2 S has 1,273 amino acids (aa), consisting of a 13-aa signal peptide at the N-terminus, S1 subunit (14–685 aa) and membrane associated S2 subunit (686–1,273 aa) (Grant et al., 2020; Ghorbani et al., 2021). The S1 domain is responsible for binding to cell receptors, including the N-terminal domain (14–305 aa) and the receptor-binding domain (RBD, 319–541 aa); while the S2 domain mediates fusion between the virus and the cell membranes, which is comprised of the fusion peptide (FP) (788–806 aa), heptapeptide repeat sequence 1 (HR1) (912–984 aa), HR2 (1,163–1,213 aa), TM domain (1,213–1,237 aa), and cytoplasmic domain (1,237–1,273 aa) (Huang et al., 2020). The SARS-CoV-2 S binds to its receptor human ACE2 through the RBD region in its S1 subunit to initiate viral infection. Subsequently, human proteases proteolytically cleave the S protein at the S1/S2 and the S2' sites, which allows the virus to enter for membrane fusion (Belouzard et al., 2009; Glowacka et al., 2011; Hoffmann et al., 2020). The glycan profile of envelope glycoproteins can affect the pathobiological activities and characteristics of many viruses, such as glycoprotein folding, trafficking and stability, virus release, functional activity, and immune evasion including shielding of the peptide backbone by glycans (Watanabe et al., 2019). Similar to the SARS outbreak in 2003, the SARS-CoV-2 spike utilizes glycan shielding to thwart the host immune response. Here the surface of the virus envelope

is coated with host-derived glycans (Casalino et al., 2020), which hides the protein surface to avoid detection by the body fluids and cellular components of the immune system (Sommerstein et al., 2015; Pritchard et al., 2015; Watanabe et al., 2020a), thereby promoting pathogen immune evasion. Moreover, when the viral glycoprotein evolve to mask immunogenic epitopes with particularly dense host-derived glycan arrays, such as high abundance of high-mannose N-glycans, neutralizing antibodies may not always recognize the underlying viral protein surface (Grant et al., 2020). Recently, a research group discovered that SARS-COV-2 spike glycoprotein glycans shield approximately 40% of the protein surface, although the glycans only account for 17% of the total molecular weight of the S trimer when expressed from human embryonic kidney cells (Grant et al., 2020; Yang et al., 2020). Molecular dynamic simulation of the interaction of spike and ACE2 proteins revealed the role of spike glycans in sterically masking polypeptide epitopes and directly participates in spike-ACE2 interactions (Zhao et al., 2020). Thus, understanding the glycosylation of recombinant S protein is very important for studying virus biology and immune response, and also help to provide information for the application of recombinant spike glycoprotein in diagnosis and vaccines.

As one of the most complex post-translational modifications found on secreted proteins and membrane-bound proteins, glycosylation can affect numerous physiological and pathological cell functions. Compared with other post-translational processes events, glycosylation is unique in its structural heterogeneity. The glycan pattern can vary greatly among individual glycosites of a given protein, different organisms, and different production hosts. Therefore, the characterization of glycans has presented a special analytical challenge due to the extraordinary heterogeneity and complexity caused by the number of glycan processing enzymes residing in the ER and Golgi apparatus. However, the advent of powerful mass spectrometry tools and related analysis methods has revolutionized our ability to identify and characterize protein glycosylation (Nwosu et al., 2011; Han and Costello, 2013; Sun et al., 2016). After the outbreak of the pandemic, multiple research groups reported glycomic and glycoproteomic analysis of the SARS-CoV-2 spike protein (De Leoz et al., 2020; Shajahan et al., 2020; Watanabe et al., 2020b; Sanda et al., 2021). However, these studies have not yet compared the production of spike proteins in multiple mammalian production hosts or characterized their sialic acid linkages.



Given that the two most important potential commercial recombinant protein production hosts are Chinese hamster ovary (CHO) and human embryonic kidney (HEK) cells, comparing the glycosylation profiles of recombinant spikes produced by these two organisms is useful for understanding what host-specific differences in glycosylation are and the potential relevance of these differences. Indeed, the differences in glycosylation between hamster ovary and human kidney cells may result in distinct glycan profiles of the same recombinant protein. For example, a comparison of 12 glycoproteins expressed from HEK and CHO suspension cell cultures revealed distinctive glycan structures, and CHO cells tend to express higher levels of

sialic acid (Croset et al., 2012). In addition, CHO cells present exclusively  $\alpha_{2,3}$ -sialylation and trace amounts of Neu5Gc (Wang et al., 2015; Yehuda and Padler-Karavani, 2020). In contrast, HEK cells can simultaneously exhibit  $\alpha_{2,3}$ - or  $\alpha_{2,6}$ -linked sialylation. Sialic acids (Neu5Ac) are typically located at the terminus of oligosaccharides on glycoproteins, and these residues are mainly linked to the galactose residue by  $\alpha_{2,3}$ - or  $\alpha_{2,6}$ -linkages in humans (Varki et al., 2015; Zhou et al., 2020). In fact, differences in specific linkages can affect the function of viral protein such as the entry protein hemagglutinin of influenza A virus with an  $\alpha_{2,6}$ -linked sialic acid binding preference (Leung et al., 2012; Yang et al., 2017).

Therefore, in this study, recombinant full-size spike protein was transiently produced in suspension propagated CHO and HEK hosts. The resulting glycan and glycopeptide profiles were characterized in order to explore the different capabilities and characteristics of recombinant spike glycosylation. To further evaluate the presence of sialic acid linkages on the S glycans, a solid-phase method was used for two-step derivatization with dimethylamine and ammonium hydroxide to differentially modify the sialic acid linkages on glycans and glycopeptides from spike proteins secreted from the two hosts. **Figure 1** shows the schematic workflow of the glycan and glycoproteomic analytical process. We provided global and site-specific analysis of N-linked glycosylation on soluble full-size SARS-CoV-2 spike using MALDI-TOF and LC-MS/MS with electron-transfer/higher-energy collision dissociation (EThcD) fragmentation. It revealed extensive heterogeneity ranging from high-mannose type to complex type glycosylation profiles, and has detailed sialic acid linkage information. The structural mapping of glycans of recombinant human spike proteins shows that CHO-Spike is more complex and contains higher sialylation ( $\alpha$ 2,3-linked), while HEK-Spike has more high-mannose and a small amount of  $\alpha$ 2,3- and  $\alpha$ 2,6-linked sialylation. Some high-mannose abundant sites on HEK-Spike may indicate a comparatively different glycan shielding presentation and viral evasion ability compared with CHO-Spike, which interacts with high mannose specific lectins such as DC-SIGN and lung surfactant SP-D (Van Breedam et al., 2014). Overall, these data underscore the importance of characterizing glycosylation of recombinant human spike protein from HEK and CHO cells in order to better understand the impact of the production host on S protein used in research, diagnostics and vaccines.

## MATERIALS AND METHODS

### Cell Culture, Transfection and Protein Purification

The mammalian expression vector pCAGGS plasmid containing the stable soluble spike protein sequence was acquired from Dr. Andrew Pekosz of Bloomberg School of Public Health at Johns Hopkins University through the generosity of Dr. Florian Kramer of the Icahn School of Medicine at Mt. Sinai (Amanat et al., 2020; Stadlbauer et al., 2020). Substitutions at the furin cleavage sites and lysine (K) at 986 position and valine (V) at 987 position were replaced with prolines (P) in order to stabilize the quaternary structure for mammalian glycosylation processing (Watanabe et al., 2020b). This recombinant S protein was truncated at the transmembrane domain and engineered to express a C-terminal trimerization motif. A sequence comparison with the SARS-CoV-2 Wuhan strain is shown in **Supplementary Figure S1**. The CHOZN GS<sup>-/-</sup> cell line was acquired from MilliporeSigma. Suspension HEK 293 cells were acquired from the American Type Culture Collection (ATCC). For CHO cells, cell culture was maintained in EX-CELL CD CHO Fusion medium (MilliporeSigma) supplemented with 6 mM glutamine in 250 ml shake flasks with a working volume of 100 ml at 37°C and 8% CO<sub>2</sub>. Transfection and transient

expression of spike protein were performed using reagents and media from CHOgro<sup>®</sup> Expression System (Mirus Bio) following the corresponding protocol. The protein was harvested from 300 ml spent media 4 days after transfection. The protein was purified *via* the Ni-NTA purification approach (Spriestersbach et al., 2015). For HEK cells, cell culture was maintained in FreeStyle<sup>™</sup> F17 Expression Medium (Thermo Fisher) supplemented with 6 mM glutamine in 250 ml shake flasks with a working volume of 100 ml at 37°C and 8% CO<sub>2</sub>. Transfection was performed *via* a polyethyleneimine (PEI)-based approach (Longo et al., 2013). Specifically, 1 mg/ml of PEI in water was prepared and added to Opti-Pro SFM (Thermo Fisher) containing plasmid DNA in a 3.5:1 PEI to DNA ratio (w/w) to form the DNA-PEI complex. Spent media was collected 4 days after transfection and the protein was purified *via* the Ni-NTA purification approach. The purified spike protein with a 6xHis tag at the C terminal was evaluated using SDS-PAGE with Coomassie blue staining to check any impurities, as shown in **Supplementary Figure S2**. The protein concentration was measured by bicinchoninic acid assay (BCA assay).

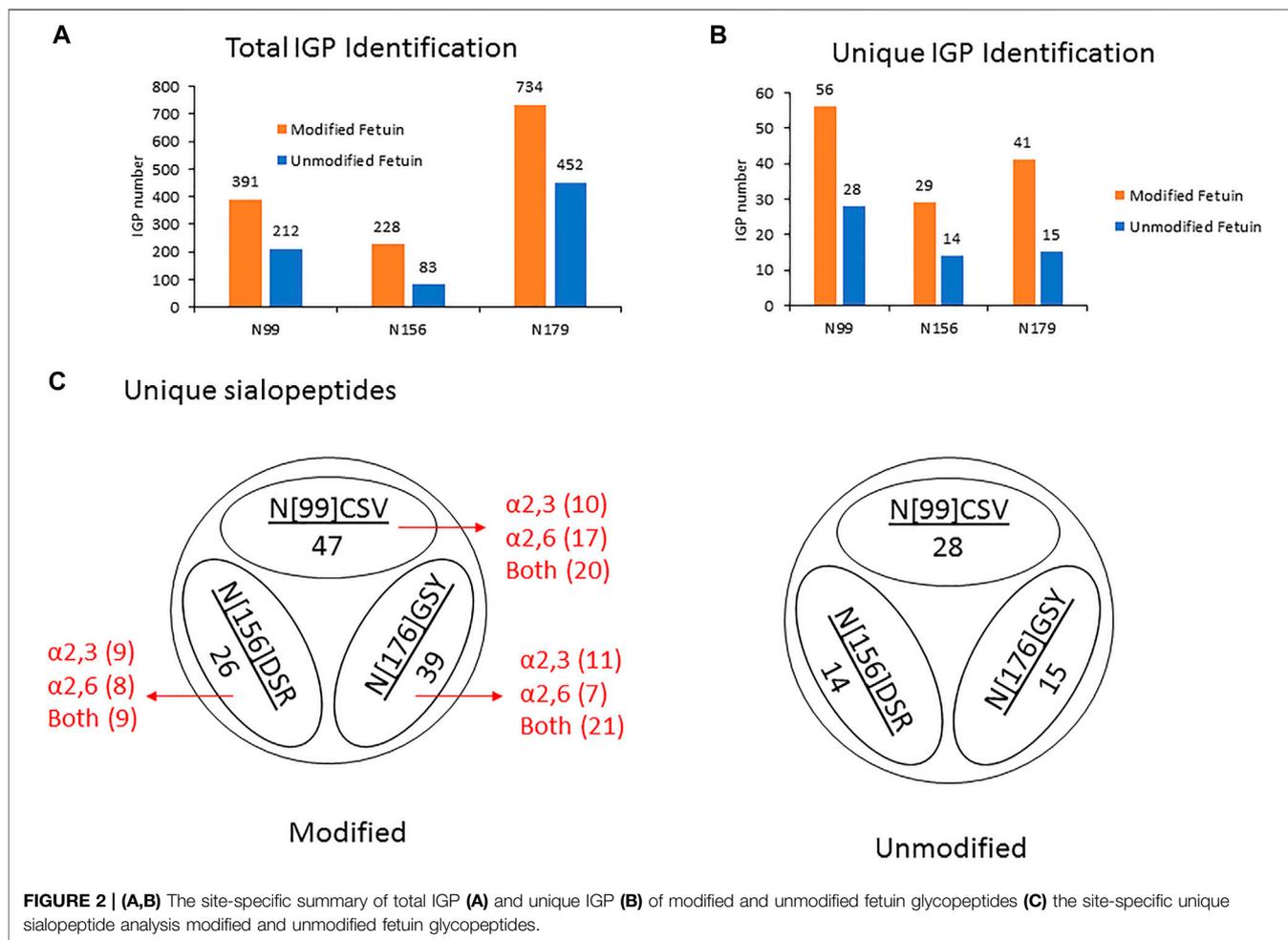
### Solid-phase Protein Conjugates and Modifications

Purified S protein, 400  $\mu$ g, was diluted in 500  $\mu$ L solution containing 1x binding buffer and 1x denaturing buffer (New England BioLabs, B1704S) and denatured at 95°C for 5 min. 1x binding buffer was prepared by dissolving 294 mg sodium citrate and 53 mg sodium carbonate in 10 ml HPLC water. All chemicals were purchased from Sigma-Aldrich unless otherwise noted. Meanwhile, Aminolink resin (Fisher Scientific) was washed by 500  $\mu$ L 1x binding buffer twice. Samples were then denatured, added to the pre-conditioned resin in a snap-cap spin column (SCSC, Fisher Scientific), and incubated overnight at room temperature. The 50 mM NaCNBH<sub>3</sub> reduction solution was added to the sample for another 4 h incubation. The resin was then washed with 1x PBS twice and further incubated with 50 mM NaCNBH<sub>3</sub> in PBS for another 4 h to block any active sites.

For sialic acid modification, samples were first treated with 200  $\mu$ L 0.25 mol/L dimethylamine, 0.25 mol/L 1-ethyl-3-(3-dimethylamino)propyl)carbodiimide (EDC) and 0.25 mol/L 1-hydroxybenzotriazole (HOBt) in DMSO solution at 60°C for 2 h. EDC and HOBt were used as a carboxylic acid activator and a catalyst individually. Next, an equal volume of ammonium hydroxide (pH10) was added into the samples with another 2 h incubation at 60°C. Samples were then washed sequentially with 10% formic acid, 10% acetonitrile, 1 M NaCl and H<sub>2</sub>O, each for three times. Finally, the sialic acid-modified protein conjugates were subjected to glycan and glycopeptide analysis individually.

### N-Glycan Analysis

After solid-phase protein conjugates and modifications, 3  $\mu$ L of PNGase F (500,000 U/ml, New England BioLabs, Ipswich, MA) in NEB Glycobuffer 2 was added to the bead mixture and incubated overnight at 37°C. The extracted glycans were subjected to Carbocolumn N-glycan clean-up and stored at 4°C for MALDI-ToF analysis. The purified glycan was analyzed using a Bruker AutoFlex Speed MALDI-ToF/ToF spectrometer in the



reflective-positive ion mode. The MALDI-ToF MS parameters were set as following: mass range 800–6,000 Da, laser 70% and 8,000 summed shots per sample. Search for predicted glycan structures-based compositions was performed using GlycoWorkBench software. The  $\alpha$ 2,3-linked sialic acids formed a stable amidation structure with a mass shift of  $-0.984$  Da, while the  $\alpha$ 2,6-linked sialic acids formed a stable dimethyl amidation structure in ammonium hydroxide with a mass shift of  $+27.047$  Da.

## On-Bead N-Glycopeptide Digestion and HILIC Enrichment

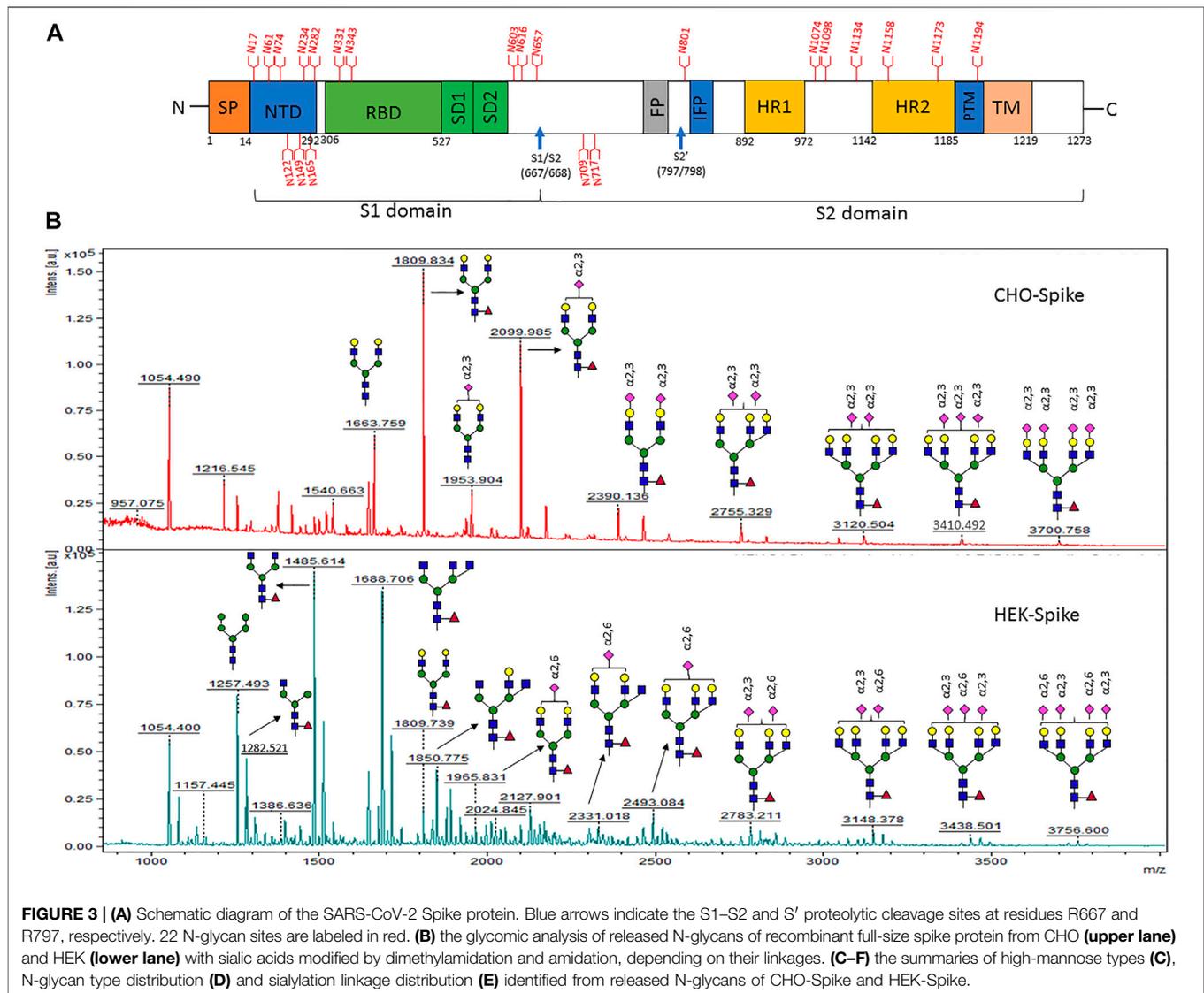
After protein conjugation and modification, the resin linked protein was incubated in 12 mM dithiothreitol (DTT) in 1M  $\text{NH}_4\text{HCO}_3$  and 8 M urea at  $37^\circ\text{C}$  for 1 h. Iodoacetamide was then added to a final concentration of 16 mM (1 h at room temperature in the dark) to alkylate the protein. After alkylation, samples were washed by 1M NaCl, HPLC water and 25 mM  $\text{NH}_4\text{HCO}_3$  sequentially (twice). For fetuin sample, sequence-grade trypsin digestion (protein: enzyme = 50:1, w/w) in 50 mM  $\text{NH}_4\text{HCO}_3$  was performed at  $37^\circ\text{C}$  overnight. For spike samples, sequence-grade trypsin and chymotrypsin (as a cocktail)

digestion (protein: enzyme = 50:1, w/w) in 50 mM  $\text{NH}_4\text{HCO}_3$  was performed at  $37^\circ\text{C}$  overnight. The sample digest was then eluted in 80% ACN and subjected to HILIC N-glycopeptide enrichment. HILIC SPE chromatography was prepared as follows: add empty SPE (solid-phase extraction) frits to Grace Alltech extract-clean empty reservoir (1.5 ml; Fisher Scientific), load 500  $\mu\text{l}$  TSKgel Amide-80 slurry in 50% ethanol (Sigma-Aldrich), and cap resin using empty SPE frits.

HILIC column enrichment was performed as follows: pre-condition HILIC column using 0.1% TFA and 60% ACN/0.1% TFA (1 ml, three times), load samples in 80% ACN/0.1% TFA (reload flow-through once), wash column by 80% ACN/0.1% TFA (1ml, twice), elute samples by 60% ACN/0.1% TFA, 40% ACN/0.1% TFA, and 0.1% TFA, and pool samples. The eluates were dried by rotary evaporation in a Speed-Vac instrument (Thermo Fisher) and re-suspended in 0.2% FA.

## Nano LC-MS/MS Analysis

The dried peptide sample was resuspended in 0.2% formic acid (FA) and measured peptide concentration using A280 Nanodrop. Then, 1  $\mu\text{g}$  of resuspended glycopeptide in 4  $\mu\text{l}$  FA buffer were analyzed by LC-MS/MS using a Thermo Orbitrap Fusion Lumos



mass spectrometer (Thermo Fisher Scientific) for each run. The analytical method was adapted from a previous publication (Wang et al., 2021). Briefly, glycopeptides were first loaded and de-salted with a trap column (Thermo Fisher PepMap™, C18, 3 μm, 100 Å, 75 μm × 2 cm) at 5 μL/min with 100% Solvent A (0.1% formic acid in HPLC water) for 5 min. Then, glycopeptides were separated by an Accalaim PepMap™ 100 nano column (3 μm, 100 Å, 75 μm × 250 mm) using a linear gradient of 2.5–37.5% solvent B (80% ACN, 0.1% formic acid) over 85 min, with a wash at 90% B for 5 min. Data-dependent analysis (DDA) was carried out with a duty cycle of 2 s. Precursor masses were detected in the Orbitrap at resolution (R) = 120,000 (at m/z 200) with internal calibration (Easy IC). Stepped HCD spectra (HCD energy at 15, 25, and 35%) were acquired for precursors with charges between two and eight and intensities over  $5.0 \times 10^4$  at R = 30,000. Dynamic exclusion was set at 20 s. If at least one of the three common glycan oxonium fragment ions (m/z 138.0545, 204.0867, and 366.1396 Da) was observed within

15 ppm mass accuracy, EThcD acquisition were performed in the orbitrap at R = 30,000. The electron-transfer dissociation (ETD) reagent target was  $2.0 \times 10^5$ , with supplemental collision energy at 15%. The ETD reaction time was dependent on the precursor charge state: 125 ms (ETD reaction time) for charge 2, 100 ms for 3, 75 ms for 4, and 50 ms for  $\geq 5$  (Wang et al., 2021).

## Data Processing and Bioinformatics Analysis

Peptide identification was performed using Byonic version 4.0 and glycopeptide quantification and characterization was performed by Byologic version 4.0 (Protein Metrics Inc., San Carlos, CA). The Byonic software parameters are listed in **Supplemental Table S1**. The precursor mass tolerance was 10 ppm, and the fragment mass tolerance was 15 ppm. The manual score cutoff was 50, the PEP2D score was less than

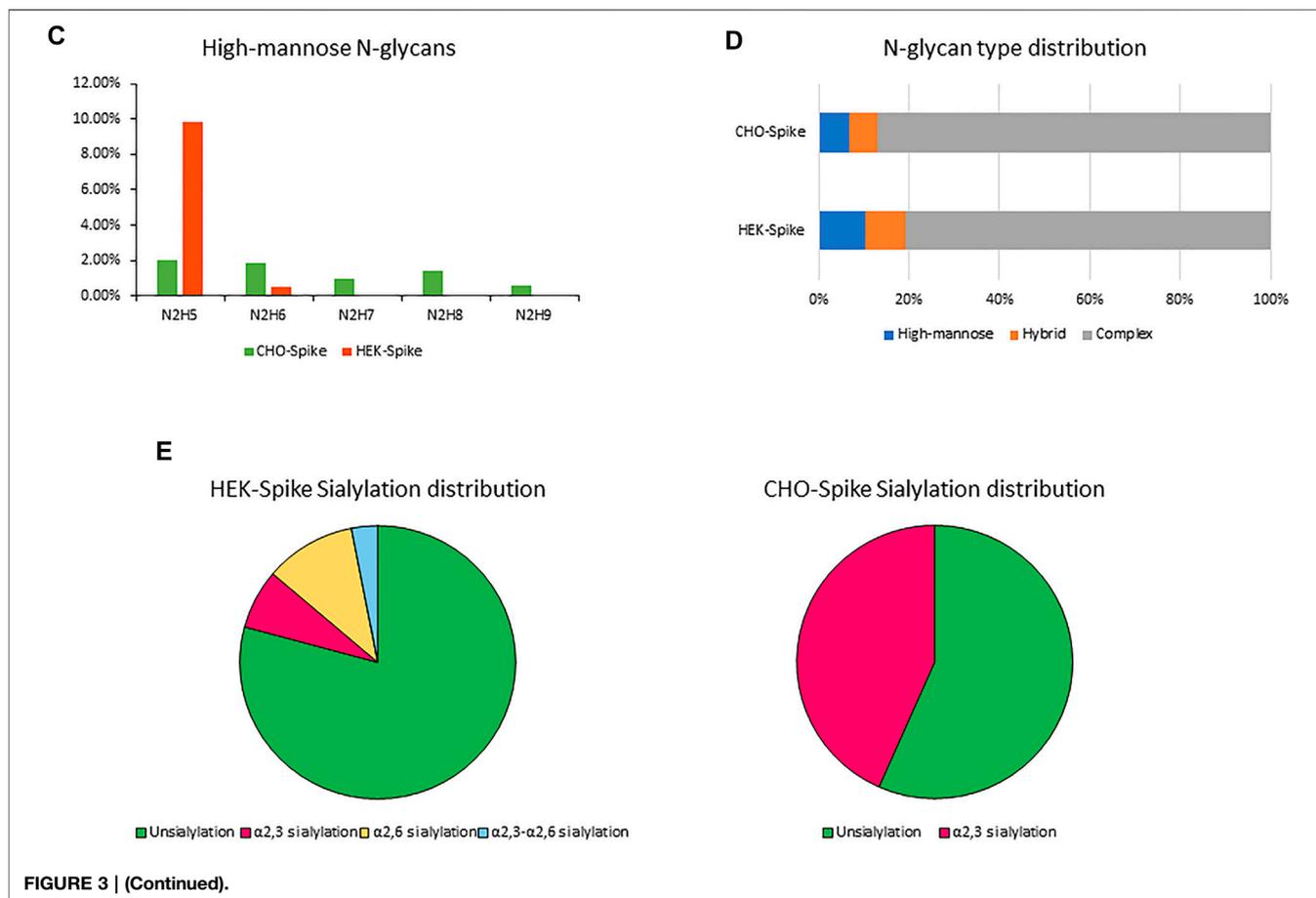


FIGURE 3 | (Continued).

0.5, and the protein false discovery rate (FDR) was 1%. To accommodate for amide-modified Neu5NAc and Neu5Gc masses, as incurred with use of dimethylamine and ammonium hydroxide, we built a mammalian N-glycan database to include these modifications. Dimethylamine can also modify the carboxyl groups on Aspartic acid (D), Glutamic acid (E) and the C-terminal of peptides. We noted this derivatization as a fixed modification in the Byonic search. We performed data searches using dimethylation only (Option 1) and both dimethylation and amidation (Option 2) on the same LC-MS/MS experimental dataset using modified fetuin. The search result from Option two showed that D and E are principally modified by dimethylation and only a trace amount was modified by amidation. According to these results, we use Option one in Byonic as the modification of D, E, and C-terminal of glycopeptides. Using Byonic search results as inputs, Byologic can compute extracted ion chromatograms (XICs) and relative abundances using a label-free quantification approach, a default setting in the Byologic software. The relative abundance of individual glycoforms on specific glycopeptides was calculated by XIC intensities over all charge states. The results of Byologic analysis were further processed and organized using the RStudio program software (Team, 2020). The mass spectrometry glycoproteomics data have been deposited to the ProteomeXchange Consortium *via* the PRIDE (Vizcaíno et al., 2016) partner repository with the dataset

identifier PXD027536 (Deutsch et al., 2017) at <http://www.ebi.ac.uk/pride>.

## RESULTS AND DISCUSSION

### Evaluation of Differential Sialic Acid Linkage Labeling of Fetuin Glycopeptides

Several derivatization methods have recently been developed to effectively stabilize sialic acid residues and distinguish the sialyl linkage isomers (Holst et al., 2016; Yang et al., 2018). For example, the application of dimethylamine ( $\text{NH}(\text{CH}_3)_2$ ) with carboxylic acid activator EDC and the catalyst HOBt in a DMSO solution can result in lactonization and dimethylamidation for individual  $\alpha$ 2,3- and  $\alpha$ 2,6-linked sialic acids, respectively (Figure 1B) (de Haan et al., 2015). Furthermore, the two-step derivatization by dimethylamine and ammonium hydroxide improved this dimethylamine derivation stability by converting  $\alpha$ 2,3-sialic acid lactonization into amidation in ammonium hydroxide solution while  $\alpha$ 2,6-sialic acid dimethylamidation remains stable (Holst et al., 2016; Zhou et al., 2019). This two-step method modified the sialic acid and made it easier to detect different linkages of sialic acids on glycans using MALDI-TOF MS (Zhou et al., 2019). In the current study, we applied this two-step sialic acid derivatization method to glycopeptide analysis

**TABLE 1** | Released N-glycans identified from CHO-Spike protein.

Experimental Mass [M + Na] <sup>+</sup>	Theoretical Mass [M + Na] <sup>+</sup>	Delta Mass	Composition	Relative percentage
1136.610	1136.40	0.209	N3H3	0.56%
1257.570	1257.43	0.142	N2H5	2.04%
1282.621	1282.46	0.161	N3H3F1	0.58%
1298.591	1298.45	0.137	N3H4	0.77%
1419.630	1419.48	0.149	N2H6	1.86%
1444.633	1444.51	0.121	N3H4F1	0.79%
1460.649	1460.51	0.142	N3H5	0.71%
1485.683	1485.54	0.145	N4H3F1	1.22%
1501.680	1501.53	0.146	N4H4	1.36%
1581.713	1581.53	0.179	N2H7	0.93%
1588.703	1588.57	0.137	S(2)1-N3H4	0.37%
1647.757	1647.59	0.165	N4H4F1	4.00%
1663.759	1663.59	0.172	N4H5	7.09%
1743.668	1743.59	0.082	N2H8	1.41%
1750.754	1750.62	0.135	S(2)1-N3H5	0.62%
1791.775	1791.65	0.129	S(2)1-N4H4	0.85%
1809.834	1809.64	0.190	N4H5F1	22.89%
1825.837	1825.64	0.198	N4H6	0.77%
1850.848	1850.67	0.177	N5H4F1	0.47%
1905.804	1905.64	0.165	N2H9	0.56%
1930.790	1930.67	0.120	N3H7F1	1.78%
1937.910	1937.70	0.207	S(2)1-N4H4F1	2.21%
1953.904	1953.70	0.206	S(2)1-N4H5	4.43%
2012.927	2012.72	0.203	N5H5F1	1.38%
2028.911	2028.72	0.193	N5H6	0.88%
2099.985	2099.76	0.229	S(2)1-N4H5F1	19.53%
2115.973	2115.75	0.222	S(2)1-N4H6	0.79%
2175.009	2174.78	0.232	N5H6F1	3.39%
2244.063	2243.81	0.253	S(2)2-N4H5	0.54%
2303.057	2302.84	0.222	S(2)1-N5H5F1	0.69%
2319.111	2318.83	0.281	S(2)1-N5H6	0.65%
2378.122	2377.86	0.266	N6H6F1	0.37%
2390.136	2389.87	0.268	S(2)2-N4H5F1	3.51%
2465.166	2464.89	0.278	S(2)1-N5H6F1	2.60%
2540.199	2539.91	0.290	N6H7F1	0.79%
2593.117	2592.95	0.170	S(2)2-N5H5F1	0.27%
2609.265	2608.94	0.323	S(2)2-N5H6	0.32%
2668.270	2667.97	0.303	S(2)1-N6H6F1	0.22%
2755.329	2755.00	0.329	S(2)2-N5H6F1	2.17%
2830.358	2830.02	0.338	S(2)1-N6H7F1	0.87%
3045.463	3045.11	0.352	S(2)3-N5H6F1	0.72%
3120.504	3120.13	0.372	S(2)2-N6H7F1	0.97%
3410.492	3410.24	0.249	S(2)3-N6H7F1	0.64%
3700.758	3700.35	0.403	S(2)4-N6H7F1	0.42%

using a solid support system. The specific sialic acid linkages on N-glycans of fetuin from fetal bovine serum were first evaluated as a model protein using a solid support that facilitates glycomic and glycoproteomic measurements. The differential sialic acid linkage labeling method used to generate the fetuin N-glycan profile is shown in **Supplementary Figure S3**. The resulting detailed released N-glycan quantification is listed in **Supplementary Table S2**, which is consistent with the fetuin sialylated N-glycan profiles reported in the literature (Nwosu et al., 2011; Yang and Zhang, 2014), and in agreement with reports that differentiate between terminal sialic acid linkage types identified by alternative labeling strategies (Yang et al., 2017) and high field NMR (Green et al., 1988; Hayase et al., 1992).

Next, this two-step differential sialic acid labeling method was performed on fetuin glycopeptides using the solid support system. To determine whether sialic acids on intact

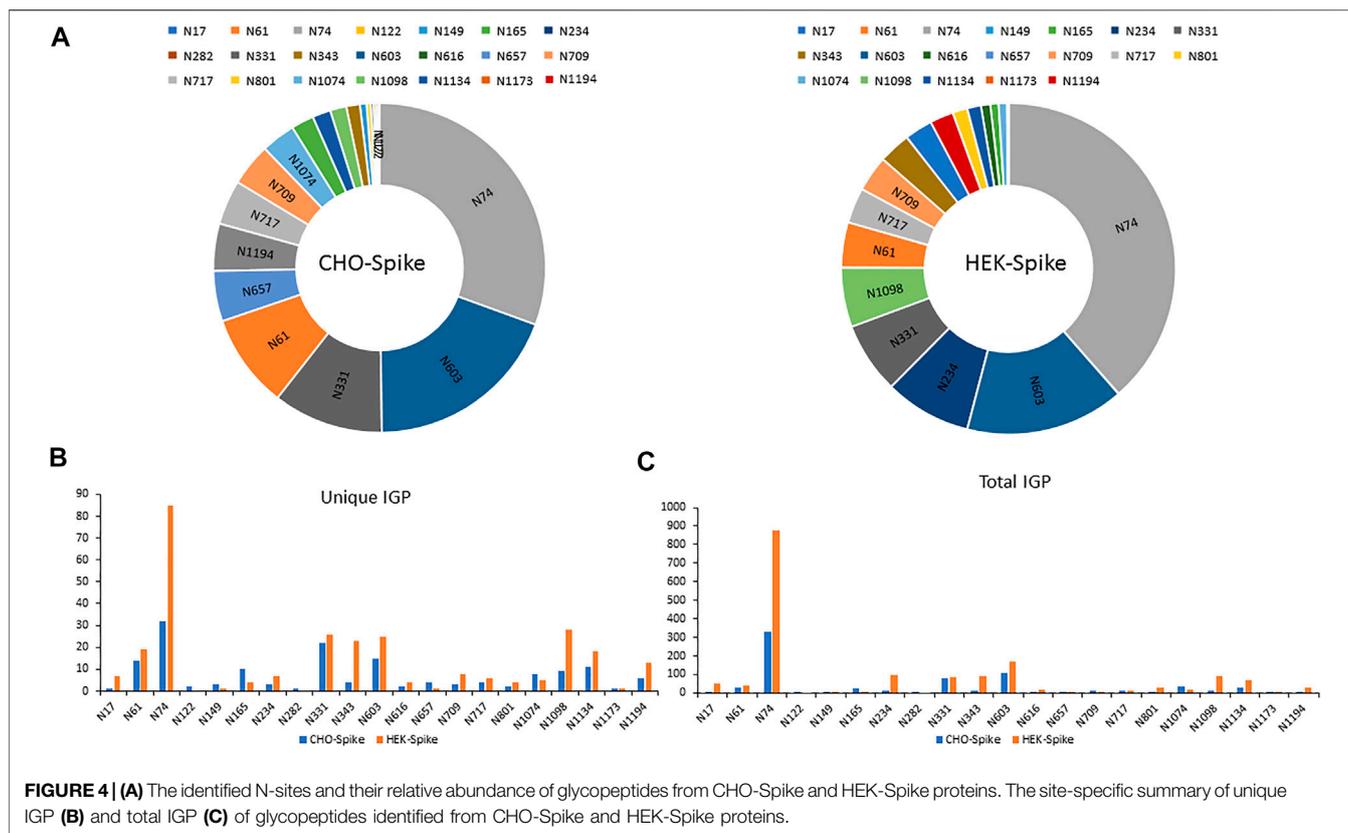
glycopeptides were successfully derivatized by dimethylamine/NH<sub>4</sub>OH, we inspected the fragment ions of tandem mass spectra. Representative glycopeptides KLC [+57] PD [+27.047]C [+57]PLLAPLN [2886.0791]DSR and VVHAVE [+27.0470]VALATFNAESN [3279.2423] GSYLQLVEISR were observed as shown in **Supplementary Figure S4** (upper lanes), which contain oxonium ions bearing mass shifts with both dimethylamine and NH<sub>4</sub>OH modified Neu5Ac ions. Other oxonium ions, C<sub>6</sub>H<sub>8</sub>NO<sub>2</sub>, C<sub>7</sub>H<sub>8</sub>NO<sub>2</sub>, HexNAc-H<sub>2</sub>O, HexNAc, HexNAcHex and HexNAxHex (2) were also observed. These two glycopeptides have only one N-glycan modified by dimethylamine and NH<sub>4</sub>OH. The fragment peaks are present at 308.289 Da for NeuAc-H<sub>2</sub>O + NH<sub>3</sub> (α<sub>2</sub>,3-linked) and 336.32 Da for NeuAc-H<sub>2</sub>O + NH(CH<sub>3</sub>)<sub>2</sub> (α<sub>2</sub>,6-linked) as shown in **Supplementary Figure S4**. These results verified that sialic acids are successfully

**TABLE 2** | Released N-glycans identified from HEK-Spike protein.

Experimental Mass [M + Na] <sup>+</sup>	Theoretical Mass [M + Na] <sup>+</sup>	Delta Mass	Composition	Relative percentage
1136.451	1136.402	0.050	N3H3	1.40%
1257.493	1257.428	0.065	N2H5	9.80%
1282.521	1282.459	0.062	N3H3F1	5.18%
1298.492	1298.454	0.037	N3H4	0.56%
1339.535	1339.481	0.054	N4H3	0.88%
1419.536	1419.481	0.056	N2H6	0.54%
1444.576	1444.512	0.064	N3H4F1	1.24%
1485.614	1485.539	0.075	N4H3F1	18.34%
1501.581	1501.534	0.047	N4H4	0.70%
1606.617	1606.565	0.052	N3H5F1	0.45%
1647.670	1647.592	0.078	N4H4F1	4.87%
1688.706	1688.618	0.088	N5H3F1	16.54%
1793.718	1793.65	0.068	N4H4F2	0.77%
1809.739	1809.644	0.094	N4H5F1	2.36%
1834.773	1834.676	0.097	N5H3F2	0.72%
1850.775	1850.671	0.103	N5H4F1	4.73%
1891.802	1891.698	0.104	N6H3F1	3.65%
1937.811	1937.703	0.108	S(23)1-N4H4F1	0.92%
1955.817	1955.702	0.115	N4H5F2	0.61%
1965.831	1965.734	0.097	S(26)1-N4H4F1	1.20%
1996.848	1996.729	0.119	N5H4F2	1.18%
2012.837	2012.724	0.113	N5H5F1	1.44%
2037.871	2037.756	0.115	N6H3F2	0.47%
2053.867	2053.75	0.117	N6H4F1	1.14%
2099.891	2099.756	0.136	S(23)1-N4H5F1	1.26%
2127.901	2127.787	0.114	S(26)1-N4H5F1	2.41%
2140.901	2140.782	0.118	S(23)1-N5H4F1	1.03%
2168.922	2168.813	0.109	S(26)1-N5H4F1	1.54%
2174.914	2174.777	0.137	N5H6F1	0.99%
2215.949	2215.803	0.145	N6H5F1	0.61%
2243.935	2243.809	0.126	S(23)2-N4H5	0.37%
2303.021	2302.835	0.186	S(23)1-N5H5F1	0.85%
2331.018	2330.866	0.152	S(26)1-N5H5F1	0.87%
2372.004	2371.893	0.111	S(26)1-N6H4F1	0.44%
2418.039	2417.898	0.141	S(23)1S(26)1-N4H5F1	0.39%
2446.047	2445.929	0.118	S(26)2-N4H5F1	0.35%
2465.069	2464.888	0.181	S(23)1-N5H6F1	0.91%
2493.084	2492.919	0.165	S(26)1-N5H6F1	1.51%
2506.124	2505.915	0.209	S(23)1-N6H5F1	0.39%
2534.118	2533.946	0.173	S(26)1-N6H5F1	0.60%
2668.099	2667.967	0.132	S(23)1-N6H6F1	0.29%
2755.228	2754.999	0.229	S(23)2-N5H6F1	0.46%
2783.211	2783.03	0.181	S(26)1S(23)1-N5H6F1	0.84%
2811.223	2811.061	0.161	S(26)2-N5H6F1	0.60%
2830.265	2830.02	0.245	S(23)1-N6H7F1	0.30%
2852.277	2852.088	0.189	S(26)2-N6H5F1	0.20%
2858.258	2858.051	0.207	S(26)1-N6H7F1	0.52%
2986.321	2986.11	0.211	S(23)1S(26)1-N6H6F1	0.22%
3101.333	3101.173	0.160	S(23)1S(26)2-N5H6F1	0.24%
3120.382	3120.132	0.250	S(23)2-N6H7F1	0.28%
3148.378	3148.163	0.216	S(26)1S(23)1-N6H7F1	0.59%
3176.373	3176.194	0.180	S(26)2-N6H7F1	0.47%
3438.501	3438.274	0.227	S(26)1S(23)2-N6H7F1	0.30%
3466.510	3466.305	0.205	S(26)2S(23)1-N6H7F1	0.27%
3756.600	3756.416	0.184	S(26)2S(23)2-N6H7F1	0.13%
3784.596	3784.447	0.148	S(26)3S(23)1-N6H7F1	0.09%

modified by dimethylamidation and amidation. Moreover, compared to the native fetuin glycopeptide (**Supplementary Figure S4** lower panels) under the same experimental condition, we observed that more oxonium ions were detected and at higher intensity when dimethylamine and

NH<sub>4</sub>OH modified Neu5Ac for the labeled fetuin glycopeptides compared to unmodified Neu5Ac of native fetuin glycopeptides. Additionally, we also noticed that the labeled glycopeptide has a higher mass charge state, which may be due to the dimethylamidation of the carboxyl groups of



Aspartic acid (D), Glutamic acid (E) and the C-terminal of peptides, where primary amines form  $-NH_3^+$  ions at a lower pH (i.e., 0.1% TFA).

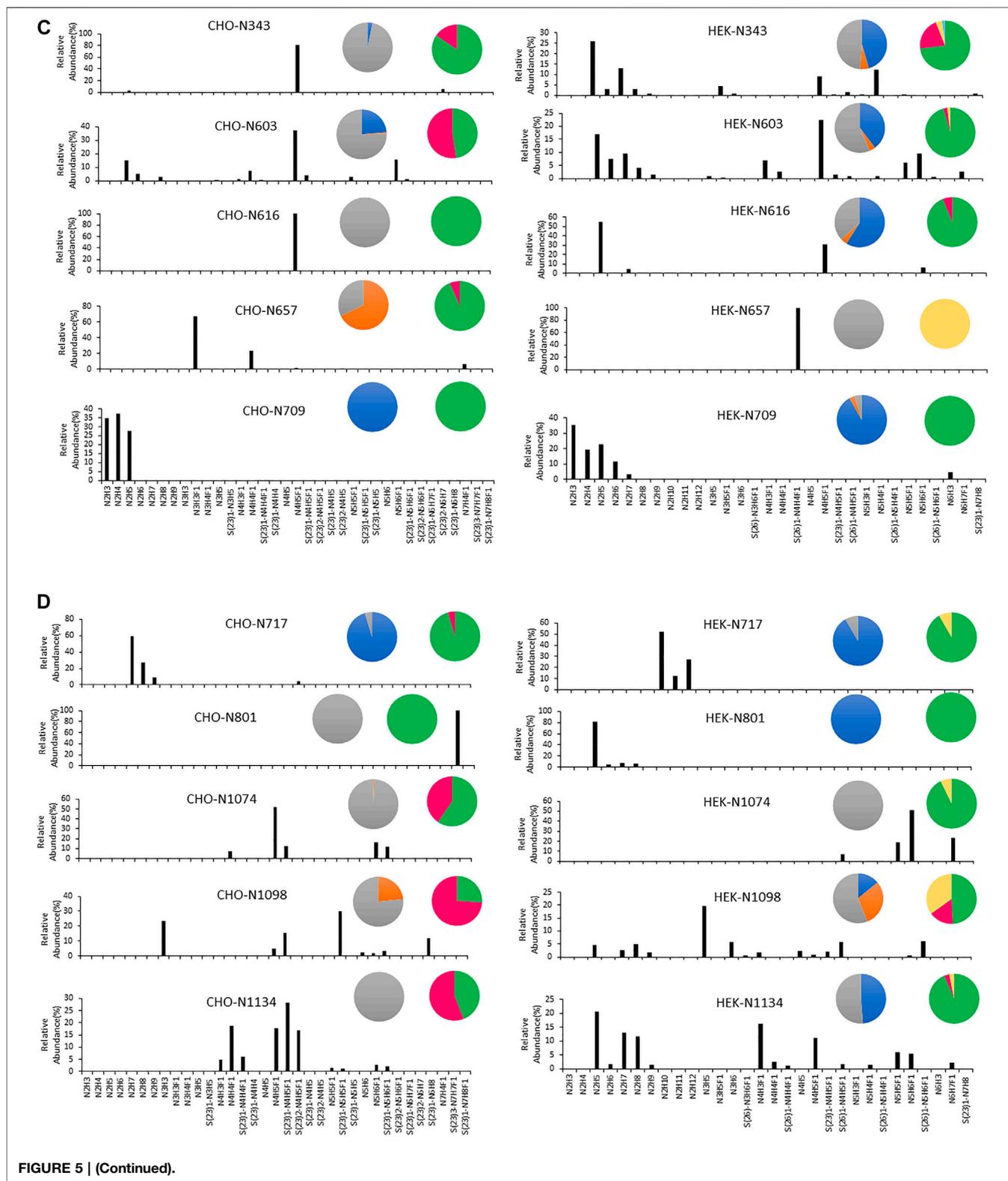
Thus, by using 3  $\mu$ g of the modified and unmodified intact glycopeptides, performed with three analytic repeats of fetuin, the total and unique intact glycopeptides (IGP) identified for the labeled samples were 1,353 and 126, compared to 747 and 57 in native sample separately. The site-specific total and unique IGP distribution are presented in **Figures 2A,B**, in which the sialic acid labeled glycopeptide showed more IGP identified compared to native glycopeptides at each N-glycan site. As mentioned above, derivatization can also occur on the carboxylic acids of D, E, or the C-terminus. **Supplementary Table S3** lists the relative abundance of glycopeptides containing D and/or E residues modified by dimethylamine in the analysis of fetuin. The glycopeptides at N99 and N179 sites contain a higher number of D and E residues, and the signal strength is higher. On the other hand, the glycopeptide at N156 site have fewer D and E residues, resulting in lower charge state and signal intensity in EThcD fragmentation. Moreover, the site-specific unique sialopeptide analysis (**Figure 2C**) identified more unique sialopeptides in modified glycopeptides versus unmodified fetuin owing to the linkage specific mass shifts imparted by the labeling strategy. Furthermore, the identified linkage specific unique sialopeptides account for the majority of the identified unique IGPs (**Figures 2B,C**), indicating that fetuin is a highly sialylated glycoprotein. Over 75 unique N-glycans were

identified in modified fetuin and 24 unique N-glycans without sialylation linkage information were identified for native fetuin with the detailed site-specific N-glycan profile tabulated in **Supplementary Table S4**.

### Characterization of Recombinant SARS-CoV-2 Spike Proteins Secreted From CHO or HEK Cell Culture

After evaluating the dimethylamine- $NH_4OH$  sialic acid linkage labeling method on fetuin, we next applied this method for the glycomic and glycoproteomic analysis of full-size recombinant SARS-CoV-2 spike protein expressed by either CHO-GS or HEK293 suspension cells. Currently, CHO and HEK cells are the two predominant production platforms for recombinant soluble S protein. Given that CHO and HEK have distinctive glycosylation processing pathways (Croset et al., 2012), the resulting soluble S proteins may produce different N-glycan profiles. Indeed, due to the lack of  $\alpha$ 2,6-sialyltransferase expression, CHO cells usually only have  $\alpha$ 2,3-sialylated glycans (Hossler et al., 2009), while HEK cells typically have both  $\alpha$ 2,3- and  $\alpha$ 2,6-linked sialic acids. CHO cells may also produce small amounts of Neu5Gc, which can be recognized as a foreign epitope and cause immunogenicity in humans (Borys et al., 2010). A schematic illustration of the different regions of SARS-CoV-2 S protein is presented in **Figure 3A**. According to literature, the SARS-CoV-2 S protein is heavily glycosylated and has 22 N-glycosites (Watanabe et al., 2020b) as designated in **Figure 3A**. Glycan





shielding provides a thick sugar-coated barrier against neutralizing antibody recognition. Consequently, global and site-specific deciphering of N-glycosylation of SARS-CoV-2 S

protein from different mammalian expression hosts will help characterize recombinant spike proteins, diagnose COVID-19, and develop protein vaccine.

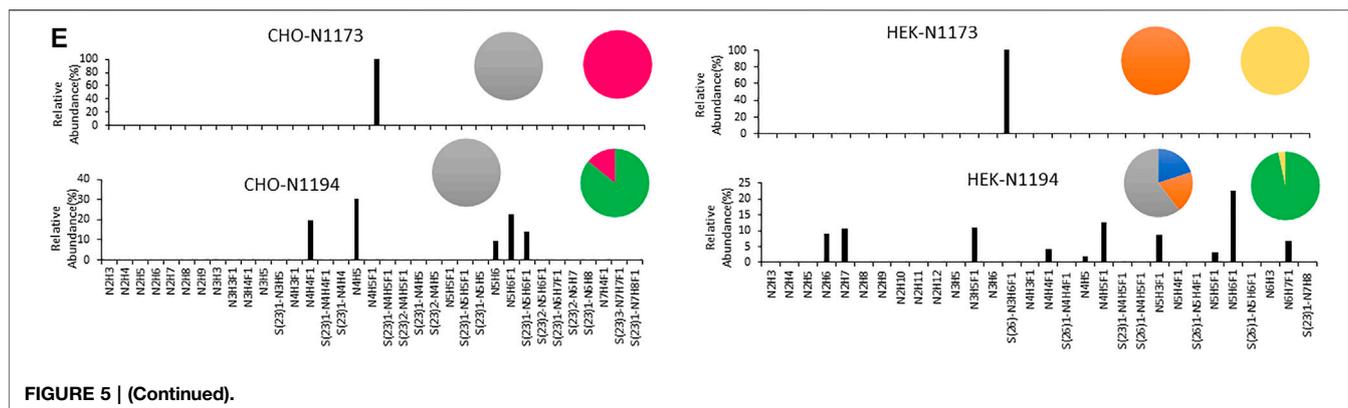


FIGURE 5 | (Continued).

The released N-glycan MS profile from three glycopeptide analytical replicates of the spike proteins from CHO and HEK hosts are shown in **Figure 3B**. The determined glycan compositions and percentage levels are listed in **Tables 1, 2**. In total, 44 and 56 N-glycan types were identified in the spike proteins from CHO and HEK cells, respectively. The distribution of high-mannose and N-glycan types is presented in **Figures 3C,D**. CHO-Spike mainly displays unsialylated and  $\alpha$ 2,3-monosialylated complex N-glycans, while HEK-Spike shows more high-mannose glycans, mainly  $\text{Man}_5\text{GlcNAc}_2$ , with less distribution between large and small MW glycans compared with CHO-Spike. HEK-Spike also mainly contains unsialylated bi- and tri-antennary N-glycans and a small amount of various  $\alpha$ 2,3- and  $\alpha$ 2,6-sialylated complex types. In accordance with our result in **Figures 3C,D**, another group also found that high oligomannose and highly fucosylated bi- and tri-antennary N-glycans have low sialylation content on the spikes of HEK cells. However, here we used solid-phase sialic acid linkage method to improve the content of the information (Zhao et al., 2020). Indeed, different expression systems and protein constructs may present different glycan patterns. The  $\alpha$ 2,3- and  $\alpha$ 2,6-sialylation distribution summary from glycomic analysis (**Figure 3E**) also reveals that HEK-Spike contained 7 and 11% exclusive  $\alpha$ 2,3- and  $\alpha$ 2,6-sialylation individually and 3% mixed sialylation, while CHO-Spike only bears 43%  $\alpha$ 2,3-sialylation. Meanwhile, the site-specific N-glycosylation analysis of S protein was performed under the co-digestion of trypsin and chymotrypsin. The representative MS/MS spectra are presented in **Supplementary Figure S5**. We identified the glycan compositions at 21 and 19 out of the 22 predicted N-glycosylation sites of SARS-CoV-2 S proteins in CHO and HEK, respectively. The determined N-glycosites and their relative abundance based on LC-MS signal intensity (percentage of total glycopeptide LC-MS intensity) are displayed in **Figure 4A**. We quantified the relative intensity of glycans at each site by comparing the area under the curve of each glycopeptide peak on the LC-MS chromatogram. Interestingly, in this study, the CHO derived spike N-glycosite at 1,158 position (N1158) was unoccupied while the N-glycosites at 122, 282 and 1,158 positions (N122, N282 and N1158) were found unoccupied in the HEK derived spike.

Another spike protein glycopeptide mapping analysis identified 17 out of 22 predicted N-sites from SARS-CoV-2 to be occupied when S1 and S2 protein fragments were expressed separately from HEK cells using a stepped HCD (higher-energy collisional dissociation) triggered CID (collision-induced dissociation) program (Shajahan et al., 2020) with 17, 603, 1,134, 1,158 and 1,173 positions unoccupied. We identified 19 out of 22 N-sites from full-size spike from HEK cells to be occupied. The identified differences may be attributed to the dissimilar LC-MS/MS fragmentation program applied, the sialic acids status as labeled or not, the size of proteins analyzed (full-size spike or individual S1 and S2 proteins), and the polypeptide processing in ER and Golgi apparatus. Alternatively, the N74 site represents the most abundantly occupied N-site of Spike from both CHO and HEK cells as displayed in **Figures 4B,C** for site-specific total and unique IGP numbers summed from three analytical repeats. Furthermore, the HEK-Spike protein analysis identified more unique IGPs at N74 site, indicating the site may be more heavily glycosylated on HEK-Spike than CHO-Spike.

In addition, the distribution of site-specific N-glycan on each identified N-glycosite from CHO and HEK cells is displayed in **Figure 5** and compared with the site-specific N-glycan compositions and percentages listed in **Supplementary Table S5**. Consistent with the global N-glycan analysis, we observed more N-glycan types and more high-mannose content on the S glycopeptides from HEK cells, especially in N17, N234, N343, N616, N709, N801, and N1134 sites. The higher abundance of high mannose glycans we observed at the listed sites may be due to differences in the inherent glycosylation propensities of the 2 cell substrates. It should be noted that the absence of complex N-glycans has been reported to reduce the entry of viruses into HEK-ACE2 expressing cells by more than 95% (Yang et al., 2020). Complex N-glycans mainly exist at the N74, N331, N1074, N1098 and N1194 sites on HEK-Spike. Most spike glycopeptides from HEK cells tend to be unsialylated, with a small amount of  $\alpha$ 2,3- and  $\alpha$ 2,6- or both  $\alpha$ 2,3 and  $\alpha$ 2,6-sialylation linkage types at the N17, N64 and N343 sites. The analysis of the glycans identified on the native-like SARS-CoV-2 virus revealed a high abundance of high-mannose type and complex N-glycans with low sialylation content (Casalino et al., 2020; Grant et al., 2020; Watanabe et al., 2021). Alternatively, CHO-Spike mainly presents complex

N-glycans as the principal glycan form on these and most other sites. Indeed, CHO-Spike is highly sialylated at N61, N74, N343, N1098 and N1134 sites, exclusively containing  $\alpha$ 2,3-sialylation. Additionally, the level of Neu5Gc detected at the N331 site (0.11% at this site) of spike from CHO cells was negligible. Furthermore, as a heavily glycosylated site on CHO-Spike and HEK-Spike, N74 presents highly  $\alpha$ 2,3-sialylated complex glycans on CHO-Spike and displays mainly unsialylated complex glycans with some high-mannose and hybrid N-glycans on HEK-Spike. Consequently, our results indicate that the N-glycosylation in full-length spike protein generated by HEK cells, is significantly different from other constructs, such as S1 and S2 independently expressed constructs as well as spike generated from CHO cells. However, separate serum ELISA assay using COVID antigens from either transient HEK and CHO lines revealed that there is no significant difference in the antibody binding response as a function of dilution level, so both antigens may be useful reagents for immunoassay detection (data not shown).

Finally, the S proteins analyzed here were engineered to contain a C-terminal trimerization domain as the transmembrane domain was removed. The trimerization domain was put in place to mimic trimerization in the engineered soluble form that would otherwise occur in the native transmembrane spike. Due to these changes glycosylation patterns may differ from those that would be observed in the TM containing native spike expressed in HEK and CHO cells.

## CONCLUSION

In this work, recombinant full-length SARS-CoV-2 spike was produced transiently in suspension CHO and HEK hosts. We then provided global and site-specific analysis of N-linked glycosylation on the soluble full-size S glycoprotein. To evaluate the sialic acid linkages present on the spike glycans, a two-step derivatization by dimethylamine and ammonium hydroxide was performed using a solid support system to differentially modify the sialic acid linkages on the resulting glycans and glycopeptides from the two hosts. We identified the glycan compositions at 21 and 19 out of the 22 predicted N-glycosylation sites of the SARS-CoV-2 S proteins in CHO and HEK, respectively. The N-glycan site at 1,158 position (N1158) was found unoccupied on spike from CHO and the N-glycan sites at 122, 282 and 1,158 positions (N122, N282 and N1158) were found unoccupied on spike secreted from HEK cells. Structural mapping of glycans of recombinant full-length human S proteins revealed that CHO-Spike presented more complex and higher sialylation ( $\alpha$ 2,3-linked only; ~40% of total) content while HEK-Spike had marginally more high-mannose glycan, almost exclusively as Man<sub>5</sub>GlcNAc<sub>2</sub>, and minor amounts of  $\alpha$ 2,3- and  $\alpha$ 2,6 linked sialylation (less than 15% total). The N74 site represents the most heavily and compositionally diverse N-glycosylated site on both spike proteins. The abundant high-mannose sites (N17, N234, N343, N616, N709, N717, N801 and N1134) on HEK-Spike may serve as glycan shields and offer viral evasion properties and/or provide targets for

mannose specific immune system lectins (Hsieh et al., 2018; Rahimi, 2020). Alternatively, the complex type N-glycans presenting at N74, N331, N1074, N1098 and N1194 sites on HEK-Spike may facilitate viral entry into ACE2 expressing cells (Yang et al., 2020) as well as serve as glycan shield. Collectively, these data underscore that certain N-glycan sites offer distinctive glycosylation patterns regardless of the host, while other sites exhibit site-specific differences in glycosylation in different mammalian hosts. These distinct site-specific N-glycan profiles may impact viral behavior *in vivo* and perhaps impact vaccine development, and these differences may have no problem or impact on specific diagnostic applications.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: ProteomeXchange Consortium *via* the PRIDE partner repository with the dataset identifier PXD027536.

## AUTHOR CONTRIBUTIONS

QW carried out experiments, provided scientific arguments, processed data and provided major contributions toward manuscript construction. YW designed and performed glycoproteomics experiments and data processing. SY provided expert advice in derivatization chemistries and solid phase design of amidations. CL performed mass spectrometry analysis and data processing. LA performed design, construction and expression of spike proteins YC performed design, construction and expression of spike proteins LP provided expert advice concerning spike structure including glycosylation TY provided expert advice concerning mass spectrometry experimental design. HJ provided expert advice concerning SARS CoV 2 expression and regulation. AP served as led for spike expression and serum testing MB provided expert advice concerning CHO and HEK293 expression systems, glycobiology, and manuscript construction. JC provided overall project focus, experimental design, development of amidation chemistries, project oversight and manuscript construction.

## FUNDING

The NIDCR Mass Spectrometry Facility is supported by the Division of Intramural Research, NIDCR/NIH (ZIA DE000751). This work was also performed under a Project Award Agreement from the National Institute for Innovation in Manufacturing Biopharmaceuticals (NIIMBL) and financial assistance award 70NANB20H037 from the Department of Commerce, National Institute of Standards and Technology, and by the United States Food and Drug Administration Program Number Z01 BJ 02044-13 LBP.

## ACKNOWLEDGMENTS

The authors would like to gratefully thank Michelle English and James Moore from Protein Metrics company for the analytical software support during the pandemic.

## REFERENCES

- Amanat, F., Stadlbauer, D., Strohmaier, S., Nguyen, T. H. O., Chromikova, V., McMahon, M., et al. (2020). A Serological Assay to Detect SARS-CoV-2 Seroconversion in Humans. *Nat. Med.* 26, 1033–1036. doi:10.1038/s41591-020-0913-5
- A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, et al. (2015). "Essentials of Glycobiology," in *The Consortium of Glycobiology Editors* (La Jolla, California(NY): Cold Spring Harbor Laboratory Press Copyright 2015-2017 All rights reserved., Cold Spring Harbor).
- Bangaru, S., Ozorowski, G., Turner, H. L., Antanasijevic, A., Huang, D., Wang, X., et al. (2020). Structural Analysis of Full-Length SARS-CoV-2 Spike Protein from an Advanced Vaccine Candidate. *Science* 370, 1089–1094. doi:10.1126/science.abe1502
- Belouzard, S., Chu, V. C., and Whittaker, G. R. (2009). Activation of the SARS Coronavirus Spike Protein via Sequential Proteolytic Cleavage at Two Distinct Sites. *Proc. Natl. Acad. Sci.* 106, 5871–5876. doi:10.1073/pnas.0809524106
- Borys, M. C., Dalal, N. G., Abu-Absi, N. R., Khattak, S. F., Jing, Y., Xing, Z., et al. (2010). Effects of Culture Conditions on N-Glycolylneuraminic Acid (Neu5Gc) Content of a Recombinant Fusion Protein Produced in CHO Cells. *Biotechnol. Bioeng.* 105, 1048–1057. doi:10.1002/bit.22644
- Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., et al. (2020). Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* 6, 1722–1734. doi:10.1021/acscentsci.0c01056
- Croset, A., Delafosse, L., Gaudry, J.-P., Arod, C., Glez, L., Losberger, C., et al. (2012). Differences in the Glycosylation of Recombinant Proteins Expressed in HEK and CHO Cells. *J. Biotechnol.* 161, 336–348. doi:10.1016/j.jbiotec.2012.06.038
- de Haan, N., Reiding, K. R., Habegger, M., Reusch, D., Falck, D., and Wührer, M. (2015). Linkage-Specific Sialic Acid Derivatization for MALDI-TOF-MS Profiling of IgG Glycopeptides. *Anal. Chem.* 87, 8284–8291. doi:10.1021/acs.analchem.5b02426
- De Leoz, M. L. A., Dwever, D. L., Fung, A., Liu, L., Yau, H. K., Potter, O., et al. (2020). NIST Interlaboratory Study on Glycosylation Analysis of Monoclonal Antibodies: Comparison of Results from Diverse Analytical Methods. *Mol. Cell Proteomics* 19, 11–30. doi:10.1074/mcp.ra119.001677
- Deutsch, E. W., Csordas, A., Sun, Z., Jarnuczak, A., Perez-Riverol, Y., Ternent, T., et al. (2017). The ProteomeXchange Consortium in 2017: Supporting the Cultural Change in Proteomics Public Data Deposition. *Nucleic Acids Res.* 45, D1100–d1106. doi:10.1093/nar/gkw936
- Ghorbani, M., Brooks, B. R., and Klauda, J. B. (2021). Exploring Dynamics and Network Analysis of Spike Glycoprotein of SARS-COV-2. *Biophysical J.* 120 (14), 2902–2913. doi:10.1016/j.bpj.2021.02.047
- Glowacka, I., Bertram, S., Müller, M. A., Allen, P., Soilleux, E., Pfeifferle, S., et al. (2011). Evidence that TMPRSS2 Activates the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. *J. Virol.* 85, 4122–4134. doi:10.1128/jvi.02232-10
- Grant, O. C., Montgomery, D., Ito, K., and Woods, R. J. (2020). Analysis of the SARS-CoV-2 Spike Protein Glycan Shield Reveals Implications for Immune Recognition. *Scientific Rep.* 10, 14991. doi:10.1038/s41598-020-71748-7
- Green, E. D., Adelt, G., Baenziger, J. U., Wilson, S., and Van Halbeek, H. (1988). The Asparagine-Linked Oligosaccharides on Bovine Fetuin. Structural Analysis of N-Glycanase-Released Oligosaccharides by 500-megahertz <sup>1</sup>H NMR Spectroscopy. *J. Biol. Chem.* 263, 18253–18268. doi:10.1016/s0021-9258(19)81354-6
- Guo, Y.-R., Cao, Q.-D., Hong, Z.-S., Tan, Y.-Y., Chen, S.-D., Jin, H.-J., et al. (2020). The Origin, Transmission and Clinical Therapies on Coronavirus Disease 2019

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.735558/full#supplementary-material>

- (COVID-19) Outbreak – an Update on the Status. *Mil. Med. Res.* 7, 11. doi:10.1186/s40779-020-00240-0
- Han, L., and Costello, C. E. (2013). Mass Spectrometry of Glycans. *Biochem. Mosc.* 78, 710–720. doi:10.1134/s0006297913070031
- Hayase, T., Rice, K. G., Dziegielewska, K. M., Kuhlenschmidt, M., Reilly, T., and Lee, Y. C. (1992). Comparison of N-Glycosides of Fetuins from Different Species and Human  $\alpha$ 2-HS-Glycoprotein. *Biochemistry* 31, 4915–4921. doi:10.1021/bi00135a024
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181, 271–280. doi:10.1016/j.cell.2020.02.052
- Holst, S., Heijs, B., de Haan, N., van Zeijl, R. J. M., Briaire-de Bruijn, I. H., van Pelt, G. W., et al. (2016). Linkage-Specific *In Situ* Sialic Acid Derivatization for N-Glycan Mass Spectrometry Imaging of Formalin-Fixed Paraffin-Embedded Tissues. *Anal. Chem.* 88, 5904–5913. doi:10.1021/acs.analchem.6b00819
- Hossler, P., Khattak, S. F., and Li, Z. J. (2009). Optimal and Consistent Protein Glycosylation in Mammalian Cell Culture. *Glycobiology* 19, 936–949. doi:10.1093/glycob/cwp079
- Hsieh, I. N., De Luna, X., White, M. R., and Hartshorn, K. L. (2018). The Role and Molecular Mechanism of Action of Surfactant Protein D in Innate Host Defense against Influenza A Virus. *Front. Immunol.* 9, 1368. doi:10.3389/fimmu.2018.01368
- Huang, Y., Yang, C., Xu, X.-f., Xu, W., and Liu, S.-w. (2020). Structural and Functional Properties of SARS-CoV-2 Spike Protein: Potential Antivirus Drug Development for COVID-19. *Acta Pharmacol. Sin* 41, 1141–1149. doi:10.1038/s41401-020-0485-4
- Leung, H. S. Y., Li, O. T. W., Chan, R. W. Y., Chan, M. C. W., Nicholls, J. M., and Poon, L. L. M. (2012). Entry of Influenza A Virus with a 2,6-Linked Sialic Acid Binding Preference Requires Host Fibronectin. *J. Virol.* 86, 10704–10713. doi:10.1128/jvi.01166-12
- Longo, P. A., Kavran, J. M., Kim, M.-S., and Leahy, D. J. (2013). Transient Mammalian Cell Transfection with Polyethylenimine (PEI). *Methods Enzymol.* 529, 227–240. doi:10.1016/b978-0-12-418687-3.00018-5
- Nwosu, C. C., Seipert, R. R., Strum, J. S., Hua, S. S., An, H. J., Zivkovic, A. M., et al. (2011). Simultaneous and Extensive Site-specific N- and O-Glycosylation Analysis in Protein Mixtures. *J. Proteome Res.* 10, 2612–2624. doi:10.1021/pr2001429
- Pritchard, L. K., Spencer, D. I., Royle, L., Bonomelli, C., Seabright, G. E., Behrens, A.-J., et al. (2015). Glycan Clustering Stabilizes the Mannose Patch of HIV-1 and Preserves Vulnerability to Broadly Neutralizing Antibodies. *Nat. Commun.* 6, 1–11. doi:10.1038/ncomms8479
- Rahimi, N. (2020). C-type Lectin CD209L/L-SIGN and CD209/DC-SIGN: Cell Adhesion Molecules Turned to Pathogen Recognition Receptors. *Biology (Basel)* 10. doi:10.3390/biology10010001
- Sanda, M., Morrison, L., and Goldman, R. (2021). N- and O-Glycosylation of the SARS-CoV-2 Spike Protein. *Anal. Chem.* 93, 2003–2009. doi:10.1021/acs.analchem.0c03173
- Shajahan, A., Supekar, N. T., Gleinich, A. S., and Azadi, P. (2020). Deducing the N- and O-Glycosylation Profile of the Spike Protein of Novel Coronavirus SARS-CoV-2. *Glycobiology* 30, 981–988. doi:10.1093/glycob/cwaa042
- Sommerstein, R., Flatz, L., Remy, M. M., Malinge, P., Magistrelli, G., Fischer, N., et al. (2015). Arenavirus Glycan Shield Promotes Neutralizing Antibody Evasion and Protracted Infection. *PLoS Pathog.* 11, e1005276. doi:10.1371/journal.ppat.1005276
- Spriestersbach, A., Kubicek, J., Schäfer, F., Block, H., and Maertens, B. (2015). Purification of His-Tagged Proteins. *Methods Enzymol.* 559, 1–15. doi:10.1016/bs.mie.2014.11.003
- Stadlbauer, D., Amanat, F., Chromikova, V., Jiang, K., Strohmaier, S., Arunkumar, G. A., et al. (2020). SARS-CoV-2 Seroconversion in Humans: A Detailed

- Protocol for a Serological Assay, Antigen Production, and Test Setup. *Curr. Protoc. Microbiol.* 57, e100. doi:10.1002/cpmc.100
- Sun, S., Shah, P., Eshghi, S. T., Yang, W., Trikannad, N., Yang, S., et al. (2016). Comprehensive Analysis of Protein Glycosylation by Solid-phase Extraction of N-Linked Glycans and Glycosite-Containing Peptides. *Nat. Biotechnol.* 34, 84–88. doi:10.1038/nbt.3403
- Team, R. (2020). *RStudio*. Boston, MA: Integrated Development for R.
- Van Breedam, W., Pöhlmann, S., Favoreel, H. W., de Groot, R. J., and Nauwynck, H. J. (2014). Bitter-sweet Symphony: Glycan-Lectin Interactions in Virus Biology. *FEMS Microbiol. Rev.* 38, 598–632. doi:10.1111/1574-6976.12052
- Vizcaíno, J. A., Csordas, A., del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., et al. (2016). 2016 Update of the PRIDE Database and its Related Tools. *Nucleic Acids Res.* 44, D447–D456. doi:10.1093/nar/gkv1145
- Wang, Q., Stuczynski, M., Gao, Y., and Betenbaugh, M. J. (2015). *Strategies for Engineering Protein N-Glycosylation Pathways in Mammalian Cells*. New York, NY: Humana Press, 287–305. doi:10.1007/978-1-4939-2760-9\_20
- Wang, Y., Wu, Z., Hu, W., Hao, P., and Yang, S. (2021). Impact of Expressing Cells on Glycosylation and Glycan of the SARS-CoV-2 Spike Glycoprotein. *ACS Omega* 6, 15988–15999. doi:10.1021/acsomega.1c01785
- Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., and Crispin, M. (2020). Site-specific Glycan Analysis of the SARS-CoV-2 Spike. *Science* 369, 330–333. doi:10.1126/science.abb9983
- Watanabe, Y., Berndsen, Z. T., Raghvani, J., Seabright, G. E., Allen, J. D., Pybus, O. G., et al. (2020). Vulnerabilities in Coronavirus Glycan Shields Despite Extensive Glycosylation. *Nat. Commun.* 11, 2688. doi:10.1038/s41467-020-16567-0
- Watanabe, Y., Bowden, T. A., Wilson, I. A., and Crispin, M. (2019). Exploitation of Glycosylation in Enveloped Virus Pathobiology. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1863, 1480–1497. doi:10.1016/j.bbagen.2019.05.012
- Watanabe, Y., Mendonça, L., Allen, E. R., Howe, A., Lee, M., Allen, J. D., et al. (2021). Native-like SARS-CoV-2 Spike Glycoprotein Expressed by ChAdOx1 nCoV-19/AZD1222 Vaccine. *ACS Cent. Sci.* 7, 594–602. doi:10.1021/acscentsci.1c00080
- Wiersinga, W. J., Rhodes, A., Cheng, A. C., Peacock, S. J., and Prescott, H. C. (2020). Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19). *JAMA* 324, 782–793. doi:10.1001/jama.2020.12839
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., et al. (2020). Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. *Science* 367, 1260–1263. doi:10.1126/science.abb2507
- Yang, Q., Hughes, T. A., Kelkar, A., Yu, X., Cheng, K., Park, S., et al. (2020). Inhibition of SARS-CoV-2 Viral Entry upon Blocking N-And O-Glycan Elaboration. *Elife* 9, e61552. doi:10.7554/elife.61552
- Yang, S., Jankowska, E., Kosikova, M., Xie, H., and Cipollo, J. (2017). Solid-Phase Chemical Modification for Sialic Acid Linkage Analysis: Application to Glycoproteins of Host Cells Used in Influenza Virus Propagation. *Anal. Chem.* 89, 9508–9517. doi:10.1021/acs.analchem.7b02514
- Yang, S., Wu, W. W., Shen, R.-F., Bern, M., and Cipollo, J. (2018). Identification of Sialic Acid Linkages on Intact Glycopeptides via Differential Chemical Modification Using IntactGIG-HILIC. *J. Am. Soc. Mass. Spectrom.* 29, 1273–1283. doi:10.1007/s13361-018-1931-0
- Yang, S., and Zhang, H. (2014). Glycomic Analysis of Glycans Released from Glycoproteins Using Chemical Immobilization and Mass Spectrometry. *Curr. Protoc. Chem. Biol.* 6, 191–208. doi:10.1002/9780470559277.ch140085
- Yehuda, S., and Padler-Karavani, V. (2020). Glycosylated Biotherapeutics: Immunological Effects of N-Glycolylneuraminic Acid. *Front. Immunol.* 11, 21. doi:10.3389/fimmu.2020.00021
- Yuen, K.-S., Ye, Z. W., Fung, S.-Y., Chan, C.-P., and Jin, D.-Y. (2020). SARS-CoV-2 and COVID-19: The Most Important Research Questions. *Cel Biosci.* 10, 40. doi:10.1186/s13578-020-00404-4
- Zhao, P., Praissman, J. L., Grant, O. C., Cai, Y., Xiao, T., Rosenbalm, K. E., et al. (2020). Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host & Microbe* 28, 586–601. doi:10.1016/j.chom.2020.08.004
- Zhou, X., Yang, G., and Guan, F. (2020). Biological Functions and Analytical Strategies of Sialic Acids in Tumor. *Cells* 9, 676–680. doi:10.3390/cells9020273
- Zhou, X., Yang, S., Yang, G., Tan, Z., and Guan, F. (2019). Two-step Derivatization and Mass Spectral Distinction of  $\alpha$ 2,3 and  $\alpha$ 2,6 Sialic Acid Linkages on N-Glycans by MALDI-TOF. *Chin. Chem. Lett.* 30, 676–680. doi:10.1016/j.ccllet.2018.12.016

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a past co-authorship with the authors (QW, MB, YC).

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wang, Wang, Yang, Lin, Aliyu, Chen, Parsons, Tian, Jia, Pekosz, Betenbaugh and Cipollo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.