

# Comprehensive Analysis of the Glycan Complement of SARS-CoV-2 Spike Proteins Using Signature Ions-Triggered Electron-Transfer/Higher-Energy Collisional Dissociation (ETHcD) Mass Spectrometry

Dongxia Wang,\* Jakub Baudys, Jonathan L. Bundy, Maria Solano, Theodore Keppel, and John R. Barr\*



Cite This: *Anal. Chem.* 2020, 92, 14730–14739



Read Online

ACCESS |



Metrics & More

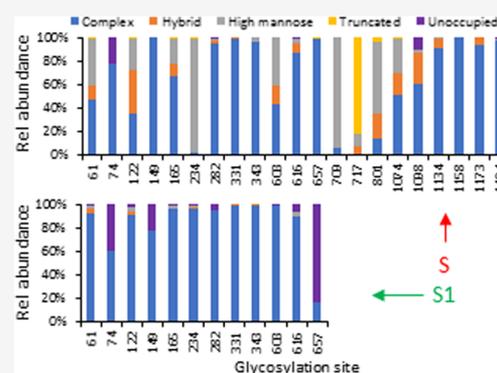


Article Recommendations



Supporting Information

**ABSTRACT:** Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has led to a global pandemic of coronavirus disease 2019 (COVID-19). The spike protein expressed on the surface of this virus is highly glycosylated and plays an essential role during the process of infection. We conducted a comprehensive mass spectrometric analysis of the N-glycosylation profiles of the SARS-CoV-2 spike proteins using signature ions-triggered electron-transfer/higher-energy collision dissociation (ETHcD) mass spectrometry. The patterns of N-glycosylation within the recombinant ectodomain and S1 subunit of the SARS-CoV-2 spike protein were characterized using this approach. Significant variations were observed in the distribution of glycan types as well as the specific individual glycans on the modification sites of the ectodomain and subunit proteins. The relative abundance of sialylated glycans in the S1 subunit compared to the full-length protein could indicate differences in the global structure and function of these two species. In addition, we compared N-glycan profiles of the recombinant spike proteins produced from different expression systems, including human embryonic kidney (HEK 293) cells and *Spodoptera frugiperda* (SF9) insect cells. These results provide useful information for the study of the interactions of SARS-CoV-2 viral proteins and for the development of effective vaccines and therapeutics.



## INTRODUCTION

The global pandemic of coronavirus disease 2019 (COVID-19) caused by the severe acute respiratory syndrome-2 coronavirus (SARS-CoV-2) emerged in late 2019 and has led to considerable economic and social disruption throughout the globe.<sup>1–3</sup> The disease has also led to considerable morbidity and mortality; according to the data compiled by the World Health Organization (September 7th Weekly Epidemiological Update) as of early September 2020, there are approximately 27.4 million diagnosed cases of infection and 895 000 estimated deaths from the disease.<sup>4</sup> Analyzing molecular mechanisms of viral cellular entry and infectivity will help guide research and development of therapeutic countermeasures and treatment and treatment.

The SARS-CoV-2 viral genome encodes four structural proteins present in the mature virion—spike protein (S), envelope protein (E), membrane protein (M), and nucleoprotein (N).<sup>5</sup> The spike protein of SARS-CoV-2 is a class I viral fusion protein, employed by the virus for host cell entry, binding through interaction with angiotensin-converting enzyme 2 (ACE2) protein present on the cell surface.<sup>6,7</sup> The spike protein is similar to the structures of the spike protein in other coronaviruses, such as those that cause SARS (SARS-CoV-1) and Middle East respiratory syndrome (MERS);<sup>8–11</sup> recent cryo-electron microscopy (cryoEM) studies<sup>12</sup> indicate

that the S protein is a homotrimer, each monomer consisting of two covalently linked functional subunits, an S1 subunit, that contains the receptor-binding domain (RBD), and an S2 subunit, which is responsible for membrane fusion, with a furin-like cleavage site at the boundary between the two domains. The S1 and S2 subunits combined have a monomeric, unmodified molecular weight of ca. 142 kDa. Because the spike protein is one of the primary targets for the development of vaccines and therapeutics against COVID-19, a more thorough understanding of its structure and function will be vital to mitigating the impact of SARS-CoV-2 on global health.

Viral proteins used in cell entry are often extensively glycosylated (often described as a “glycan shield”) for several reasons: to assist in protein folding, provide stability, and most importantly, shield the virus from immune recognition by its host, as has been reported for other coronavirus species.<sup>10,11</sup> The SARS-CoV-2 S protein is a prime example, having 22

Received: August 4, 2020

Accepted: October 6, 2020

Published: October 16, 2020



potential sites of glycosylation per protein monomer, as predicted from the primary sequence.<sup>13</sup> Development of vaccines and therapeutics for SARS-CoV-2 requires an understanding of the overall composition of the spike protein “glycan shield”, both in naturally occurring isolates and vaccine formulations containing this protein. Observed variances in glycan site occupation and speciation might affect antigenicity as well as vaccine safety and efficacy.<sup>14,15</sup>

In the past several decades, mass spectrometry has become one of the primary tools for the analysis of protein glycosylation,<sup>16–18</sup> both on a proteome-wide scale<sup>19–21</sup> and in targeted analyses of single protein therapeutics, most notably monoclonal antibodies.<sup>22</sup> Mass spectrometry has also been employed extensively for viral protein characterization. Seasonal and pandemic influenza, with its substantial morbidity and mortality, has been a primary focus of many of these efforts, for antigen quantification,<sup>23–28</sup> vaccine potency determination,<sup>29,30</sup> and analysis of protein glycosylation.<sup>31–34</sup> As the COVID-19 pandemic has progressed, several reports on the use of mass spectrometry for the analysis of SARS-CoV-2 have been made. These investigations have focused on both targeted quantitative analysis for potential diagnostic development<sup>35</sup> and structural/post-translational modification analyses of the virus protein complement.<sup>13,36–40</sup>

Electron-transfer dissociation (ETD), developed by Hunt and co-workers,<sup>41–44</sup> and subsequently introduced hybrid ETD and higher-energy collisional dissociation (HCD) fragmentation regimes such as electron-transfer/higher-energy collisional dissociation (ETHCD), developed by Heck et al.,<sup>45</sup> have been extensively employed for mass spectrometric post-translational modification analysis of phosphorylated<sup>46</sup> and glycosylated<sup>47,48</sup> proteins and peptides. A key advantage of ETD-based approaches is that they provide more comprehensive fragment ion coverage than conventional collision-induced dissociation (CID) for post-translational modification of larger peptides due to the less biased nature of the fragmentation mechanism. This allows the sites of peptide modification to be unambiguously determined in many cases.<sup>49</sup>

The glycan shield on the SARS-CoV-2 S protein has been recently characterized using mass spectrometry by several laboratories.<sup>13,37,38,50,51</sup> Their results show significant differences in the glycan occupancy and abundance of some sequons. For instance, all 22 potential glycosylation sites were determined to be occupied on a stabilized extracellular domain (ectodomain) S protein expressed in human HEK cells in one report by Crispin and co-workers,<sup>13</sup> whereas in another report, only 17 sites were found occupied based on the analysis of individual HEK cell expressed S1 and S2 subunits.<sup>50</sup> A third report analyzed an insect cell expressed full-length SARS-CoV-2 protein, with unambiguous determination of 21 of 22 glycosylation sites on this construct. A most recent investigation<sup>38</sup> analyzed a full-length ectodomain of SARS-CoV-2 protein of similar composition to the Crispin report. They also observed occupation of all 22 potential sites, with some significant differences. Interestingly, O-glycosylation and/or N-linked sulfated glycans on the S protein and subunits were also observed by several groups,<sup>13,39,50,52</sup> albeit at relatively low levels of site occupation. These variations could be attributed to differences in protein structure, expression systems, sample preparation, instrumentation, acquisition methods, or data analysis. The use of manual inspection of the mass spectra for the confirmation of identities could be subjective as well.

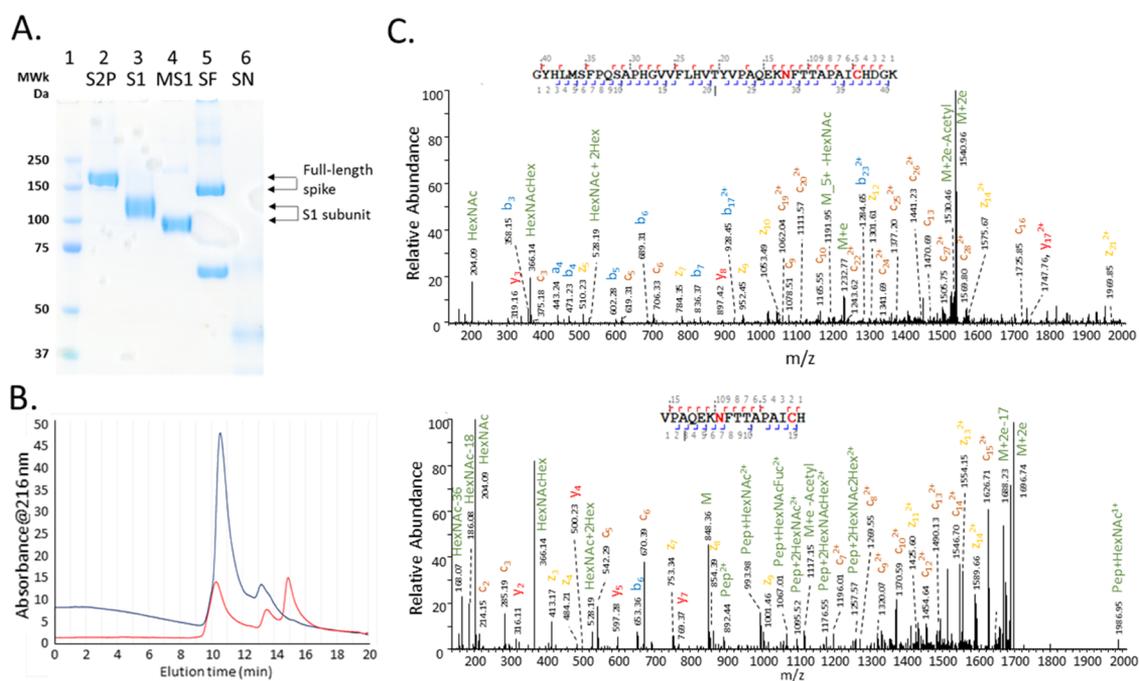
In this report, we describe a comprehensive, high-fidelity mass spectrometric approach to the glycosylation analysis of SARS-CoV-2 S protein, applied to multiple recombinant spike protein sources and constructs (S1 domain and ectodomain) with identical analytical methods. In contrast to other reports on SARS-CoV-2 S protein glycoproteomic analysis that used conventional CID, HCD,<sup>13,37,50</sup> or stepped collision energy HCD (sceHCD),<sup>38</sup> we employed an approach based on glycan reporter ion-triggered ETHCD, which allowed the sites of glycosylation to be unambiguously determined with a greater proportion of fragment ions observed, which increased the degree of confidence in the observed search results.

## ■ EXPERIMENTAL SECTION

**Materials.** All chemicals were obtained from Sigma-Aldrich (St. Louis, MO) except where otherwise indicated. Endoproteases including trypsin, chymotrypsin, Asp-N, and Lys-C were purchased from Promega (Madison, WI). Recombinant SARS-CoV-2 prefusion stabilized spike protein was provided by Dr. Jason McLellan at University of Texas at Austin. The protein was expressed in human FreeStyle293F cells.<sup>12</sup> Recombinant SARS-CoV-2 spike protein expressed in baculovirus insect cells (S1 + S2 ECD, Cat. no. 40589-V08B1), SARS-CoV-2 spike protein S1 subunit (Cat. no. 40591-V08H), and MERS-CoV spike protein S1 subunit, both expressed in human-derived HEK cells (Cat. no. 40069-V08H) were purchased from Sino Biological (Wayne, PA). SARS-CoV-1 spike protein (Cat. no. NR-686) prepared in insect cells was obtained from BEI Resources (Manassas, VA).

**One-Dimensional (1D) Gel Electrophoresis.** Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed on a NuPAGE Novex Bis-Tris gel following manufacturer's instructions (Invitrogen). Protein solution or pellets from acetone precipitation were mixed with 4× sample buffer and deionized water (1:3 v/v) and heated at 80 °C for 10 min. The proteins were loaded on a 4–12% gradient gel, and the gel was run in 3-(*N*-morpholino)-propanesulfonic acid (MOPS) buffer at 200 V for 45 min. The gel was stained with GelCode Blue Safe Protein Stain (Cat. no. 243596, Thermo Scientific, Waltham, MA).

**Size-Exclusion High-Performance Liquid Chromatography (HPLC) Analysis.** Size-exclusion HPLC of recombinant spike protein samples was performed using an Agilent Technologies (Santa Clara, CA) 1200 HPLC System equipped with a quaternary pump, diode array detector, and fraction collector, coupled to an Agilent BioSEC-3 size-exclusion column (4.6 mm ID × 150 mm OD). The column flow rate was 200  $\mu$ L/min with a sample injection size of 15  $\mu$ L. The sample was eluted isocratically over 45 min using a commercially prepared Dulbecco's phosphate-buffered saline mobile phase (Thermo/Life Technologies). Detection was performed by monitoring the UV wavelengths of 216 and 280 nm using a diode array detector, and 1 min fractions were collected for subsequent mass spectrometric identification. The column was calibrated using commercially available size-exclusion chromatography (SEC) standards (Acquity BEH 450 Protein Standards) obtained from Waters Corporation (Milford, MA) that spanned the molecular weight range of 112–700 kDa. A plot of log MW vs retention time of the standards was used to determine a calibration curve fit to a quadratic model, which was employed to calculate the estimated relative molecular weight of the spike protein samples. The UV traces were baseline-corrected.



**Figure 1.** (A) SDS-PAGE gel of the unfractionated S proteins (full-length and S1 subunit). Lane 1: molecular markers; lanes 2–6: S2P, S1, MS1, SF, and SN, respectively. (B) Size-exclusion chromatogram (UV detection at 216 nm) of the recombinant SARS-CoV-2 S proteins expressed in human (S2P, blue trace) and insect cells (SF, red trace). (C) Representative MS/MS spectra of the glycopeptides with a complex glycan, HexNAc<sub>4</sub>Hex<sub>3</sub>Fuc, at amino acid position 1074. The two peptides were cleaved from S2P by trypsin (top) and Asp-N/chymotrypsin (bottom), respectively.

**Digestion of Spike Proteins.** The aliquots of full-length spike protein or S1 subunit (0.4–0.7  $\mu$ g) were denatured and reduced at 90 °C for 20 min in a solution containing 50 mM ammonium bicarbonate (pH 7.8), 0.015% RapiGest SF surfactant (Waters Corporation) and 1.5 mM dithiothreitol (DTT). Samples were alkylated using iodoacetamide for 30 min in the dark at room temperature with gentle mixing. Nine identical aliquots of each protein were prepared for three digestion methods, and each digestion was performed with three replicates. The digestions with trypsin were conducted at 37 °C overnight at an enzyme-to-protein ratio of 1/20 (w/w). For the samples digested sequentially by two enzymes, the first digestion was performed at 52 °C for 60 min with Asp-N or Lys-C and the second digestion was conducted at 37 °C overnight using chymotrypsin at an enzyme/substrate ratio of 1/5 (w/w). The proteolytic reactions were quenched, and the RapiGest was precipitated by adding 5% trifluoroacetic acid to decrease the pH to below 3. The mixture was then incubated at 37 °C for 30 min. The solutions were centrifuged at 4000 rpm for 10 min, and the supernatants (12  $\mu$ L) were transferred into new sample vials.

**Mass Spectrometry Analysis.** Nanoflow liquid chromatography coupled to electrospray ionization tandem mass spectrometry (LC-MS/MS) analysis was performed on an Orbitrap Eclipse Tribrid mass spectrometer connected to an UltiMate 3000 RSLCnano chromatography system (Thermo Scientific). The protein digest was separated on an integrated separation column/nanospray device (Thermo Scientific EASY-Spray PepMap RSLC C<sub>18</sub>, 75  $\mu$ m ID  $\times$  15 cm length, 3  $\mu$ m 100 Å particles) coupled to an EASY-Spray ion source. The mobile phase was 0.1% formic acid in water (mobile phase A) and 0.1% formic acid in 80% acetonitrile/20% water (mobile phase B) using the following gradient: 4% B (0–8 min); 4–10% B (8–10 min); 10–35% B (10–43 min); 35–

60% B (43–45 min); 60–95% B (45–46 min); 95% B (46–53 min); 95–4% B (53–53 min); 4% B (53–63 min). The flow rate was 300 nL/min, and 9  $\mu$ L of samples was injected. The spray voltage was set to 1.8 kV, and the temperature of the integrated column/nanospray device was set at 55 °C. The temperature of the ion transfer tube was set at 275 °C.

Mass spectrometric data acquisition was performed using a signature ion-triggered EThcD method.<sup>53</sup> MS precursor scans were acquired by the orbitrap at a resolution of 120 000 (measured at  $m/z$  200), from  $m/z$  375 to 2000 with the automatic gain control (AGC) target setting as “standard” and the maximum injection time as “auto”. An initial data-dependent MS/MS scan was acquired using HCD at a resolution of 30 000, mass range of  $m/z$  120–2000, and a normalized collision energy (NCE) of 28%. Signature ions representing glycan oxonium fragments were used to trigger the ETD fragmentation. If one of three common glycan signature ions ( $m/z$  204.0867 (HexNAc), 138.0545 (HexNAc fragment), or 366.1396 (HexNAcHex)) was detected in the HCD spectrum within 15 ppm mass accuracy, additional precursor isolation and EThcD acquisition were performed at a resolution of 50 000 (measured at  $m/z$  200), scan range of  $m/z$  150–2000 with normalized AGC target = 500% and maximum injection time = 150 ms. The supplemental activation NCE was set to 35%.

**Data Analysis.** MS/MS data were processed using PMi-Byonic (version 3.7) and PMi-Byologic (version 3.7, Protein Metrics, Inc.). Data were searched using the Protein Metrics 182 human N-glycan library (for proteins expressed in HEK cells) or 38 insect N-glycan library (for proteins expressed in insect cells). The search parameters for enzyme digestion were set to fully specific, three allowed missed cleavage sites, and 6 and 20 ppm mass tolerance for precursors and fragment ions, respectively. Carbamidomethylation of cysteine was set as a

**Table 1. Relative Abundance of the Glycans on the Glycosylation Sites of the Prefusion Spike Protein (S2P) of the SARS-CoV-2 Virus Determined by LC/MS Analysis of Protein Digests under Various Experimental Conditions<sup>a</sup>**

glyco-site	Lys-C/chymotrypsin digestion				Asp-N/chymotrypsin digestion				trypsin digestion						
	complex	high mannose	hybrid	truncated	unoccupied	complex	high mannose	hybrid	truncated	unoccupied	complex	high mannose	hybrid	truncated	unoccupied
61	0.47	0.41	0.11	0.01	0.47	0.37	0.15	0.01	0.22	0.38	0.24	0.38			
74					0.78										
122	0.31	0.30	0.39		1.00										
149	1.00			0.01	0.73	0.23	0.02	0.01		0.69		0.31			
165	0.76	0.21	0.01		0.01	0.97	0.01	0.01		0.01	0.98				
234	0.01	0.98	0.01		0.99		0.01	0.01		0.97		0.02			
282	0.93	0.02	0.04	0.01	0.98		0.02								
331	0.99		0.01		0.96										
343	0.96	0.04			0.40	0.04	0.18								
603	0.45	0.40	0.14		0.81	0.42	0.14		0.01						
616	0.94	0.03	0.03		0.99	0.04									
657	1.00							0.01							
709	0.06	0.94				1.00									
717		0.11		0.89		0.12		0.88							
801	0.15	0.60	0.21	0.04	0.07	0.70	0.20	0.02		0.21	0.54	0.21	0.04		
1074	0.45	0.35	0.20		0.66	0.21	0.13			0.41	0.36	0.23			
1098	0.72	0.02	0.27		0.57		0.31		0.12	0.69	0.03	0.28			
1134	1.00				0.91	0.01	0.09								
1158					1.00										
1173	1.00				0.93		0.07								
1194	1.00		0.00		1.00					0.98	0.02				0.01

<sup>a</sup>The values were determined by summing the mean area ratio of individual glycans of three biological replicates then normalizing at each glycosylation site.

fixed modification with variable modifications set to include deamidation at Asn and Gln and oxidation of Met. Tandem mass spectra of identified glycopeptides with a Byonic score<sup>54</sup> of higher than 300 were considered valid identifications and were examined by manual inspection for further validation based on the presence of predicted peptide fragments and diagnostic N-glycan fragment ions. The precursor ion peak areas of identified glycopeptides and unoccupied peptides were obtained through the Byologic algorithm, with the relative abundance of each glycan at each site calculated as the area ratio of the peptides bearing a particular glycan over the total peptides of the same peptide sequence. The glycan abundance was represented as the mean of either two or three replicates, depending upon the number of observations, along with standard deviation of the mean.

## RESULTS AND DISCUSSION

To compare the glycan profiles of SARS-CoV-2 S protein and its subunits prepared from different expression systems, we examined three SARS-CoV-2 recombinant S protein constructs, including SARS-CoV-2 prefusion stabilized spike protein (S2P), unmutated spike protein (SF), and the S1 subunit (S1), using bottom-up LC-MS/MS techniques on a Thermo Orbitrap Eclipse mass spectrometer. In addition, SARS-CoV-1 spike protein (SN) and MERS-CoV spike protein S1 subunit (MS1) were analyzed for comparison. Among these five proteins, S2P, S1, and MS1 were expressed in human-derived HEK cells, and SF and SN were prepared in insect cells using a baculovirus-SF9 expression system. To generate high-quality MS/MS spectra for potential glycopeptides, HCD-triggered EThcD was used because it focused more data acquisition time on analyzing glycan bearing peptide ions. The EThcD fragmentation regime favored the formation of both glycan and peptide backbone fragment ions. A recent report by Bertozzi and co-workers<sup>55</sup> compared the analytical figures of merit of ETD, EThcD, HCD, and stepped collision energy HCD (sceHCD) for analysis of N- and O-linked glycopeptides. Our results corroborate their findings that ETD-based methods provide better spectral quality—superior peptide and glycan sequence coverage to HCD methods alone. It is important to note that they observed sceHCD methods provided a greater number of peptide spectral matches (PSMs) with similar peptide and glycan sequence coverages than EThcD from a standard glycopeptide mixture at a compromise in spectral quality.

The heterogeneity of the five samples was demonstrated by SDS-PAGE gel (Figure 1A); a single band of S2P, S1, and MS1 revealed the high purity of these proteins, whereas an additional band observed in the SF sample indicated a spike protein fragment or contaminating protein was present in the sample. A band was barely visible in the lane loaded with SN, suggesting that protein had degraded; this reagent was manufactured in 2005 as a vaccine candidate for SARS-CoV-1 and held in long-term storage. The size-exclusion chromatogram obtained under native conditions, shown in Figure 1B, illustrated the difference in structure between the stabilized prefusion protein (S2P) and wild-type version of the S protein (SF), where the former was observed to be predominantly a trimer of apparent molecular mass 700 kDa, peak eluting at 10.5 min, with a relatively lower amount of monomer eluting at 13 min. The SF sample was observed to be present as a mixture of trimer (700 kDa, 10.4 min) and monomer (230 kDa, 13.5 min) with a degradation product (100 kDa) at 15

min that presumably corresponds to the species observed in SDS-PAGE at this approximate mass. The identities of the chromatographic peaks observed in the SEC peaks were confirmed to be SARS-CoV-2 S protein by LC-MS/MS analysis of collected fractions.

The unambiguous characterization of glycoproteins by mass spectrometric analysis of proteolytically cleaved glycopeptides presents several significant analytical challenges. Compared to the typical molecular mass of unmodified peptides (1–5 kDa), the size of the glycan moiety (typically >1 kDa) can lead to a higher prevalence of false-positive spectrum matches even at high mass measurement accuracies. In addition, a lower population of fragment ions containing the glycan moiety is often observed in MS/MS spectra of glycopeptides.<sup>16</sup> To address these issues, the samples were digested using three protease/protease combinations (trypsin, Lys-C + chymotrypsin, and Asp-N + chymotrypsin) with each digestion condition conducted in triplicate. The identification of the same glycan at a sequon by two or more peptides derived from the cleavage by a different enzyme or enzyme combination can cross-validate each other. For example, a complex glycan, HexNAc<sub>4</sub>Hex<sub>4</sub>Fuc, modified at N1074 of S2P, was determined with high confidence by two glycopeptides of different lengths, one was S2P(1046-1086) derived from trypsin proteolysis and another was S2P(1068-1087) generated from the digestion by Asp-N/chymotrypsin (Figure 1C).

To obtain accurate quantitative results, we processed the data of specific proteolytic digestion experiments separately. This resulted in a majority of the peptides bearing a specific sequon to have identical amino acid lengths and sequences because they were produced from the digestion by identical enzyme(s). This is particularly important because the LC-MS/MS peak intensity or integrated peak area derived from a given peptide is dependent on its amino acid composition and sequence, assuming that it is not affected by the associated glycans. Summing the areas of a group of peptides with the same glycan site but different amino acid sequences could lead to an inaccurate calculation of relative abundance of individual glycans on a specific modification site. In addition, a 6 ppm precursor ion tolerance was employed in database searching, and the criteria for positive identification of a glycopeptide were a Byonic score of  $\geq 300$  and validation of MS/MS data by manual inspection. Moreover, any glycopeptides that were detected in only one of the three replicates were not included for the quantification of glycan distribution, even if their identities were unambiguously determined by the criteria described previously.

Of the 22 possible N-linked glycosylation sites on the S2P protein, 21 sites were detected as glycosylated by this comprehensive approach (Data S-1). No peptides containing the sequon N17 were detected from any of data meeting the selection criteria described above. In addition to the categories of high mannose (HexNAc<sub>2</sub>Hex<sub>3-4</sub>X), hybrid containing three HexNAc, and complex glycans with more than three HexNAc, we also added a truncated species type including paucimannose and other small glycans consisting of only one HexNAc or two HexNAc but less than five Hex groups in their compositions, due to the presence of large amounts of such glycans in some samples. The relative abundances of different types of N-glycosylation on the prefusion S2P proteins are listed in Table 1. The results generated from various experimental conditions showed high consistency within two or three datasets, demonstrating the reproducibility of this

approach. Although the glycosylation on the N74 and N1158 sequons was not detected from the experiment using Lys-C/chymotrypsin digestion, their modifications were determined from Asp-N/chymotrypsin digests, revealing that complementary outcomes can be gained by the combination of multiple enzyme digestions.

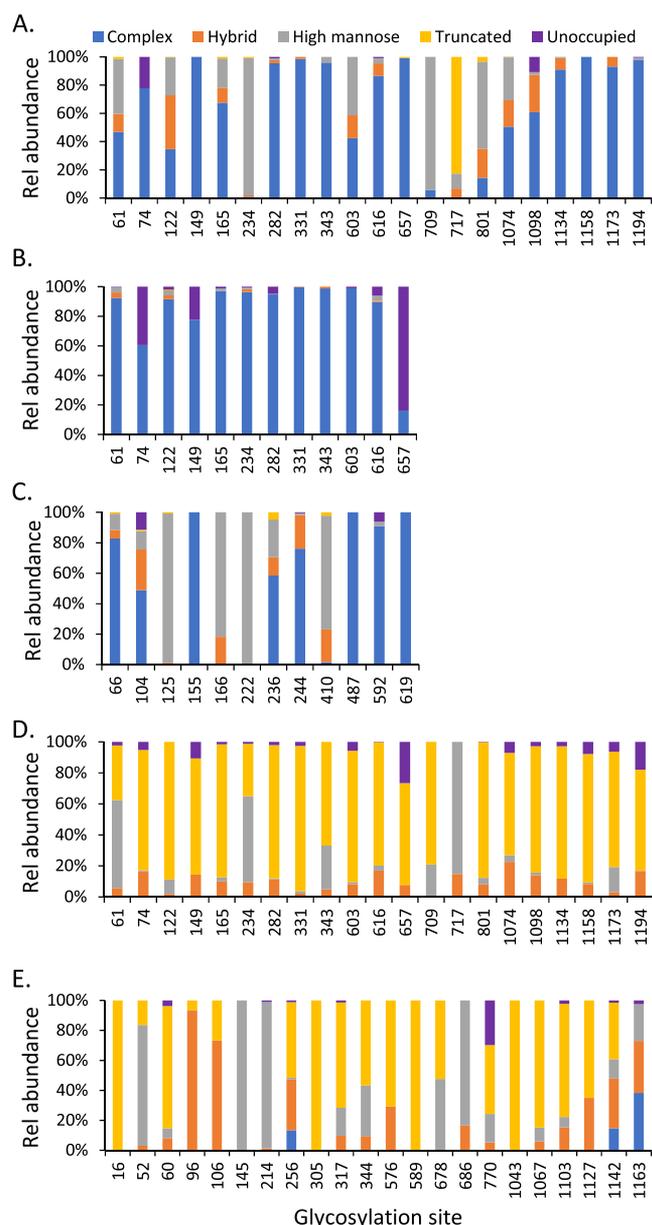
There were 19 of 21 detected sequons on the prefusion stabilized ectodomain of the SARS-CoV-2 S protein (S2P) that were fully occupied by various glycans (Table 1 and Figure 2A). The unoccupied fraction of sequons N74 and N1098 was relatively low (10–20%). This is consistent with the results

reported for other glycoproteins. Among four types of glycans, complex glycans were the most frequently observed within most of the individual N-glycosylation sites. There were 17 sites that contained more than 40% complex glycans, of which 11 sites contained 80% or higher relative abundance of complex glycans including N74, N149, N282, N331, N343, N616, N657, N1134, N1158, N1173, and N1194, suggesting this protein was modified with extensively processed N-glycans. This observation is similar to a recent report on a full-length S protein<sup>13</sup> but different from that reported on two subunits of the spike protein expressed separately, where a larger proportion of high-mannose glycans was determined.<sup>50</sup>

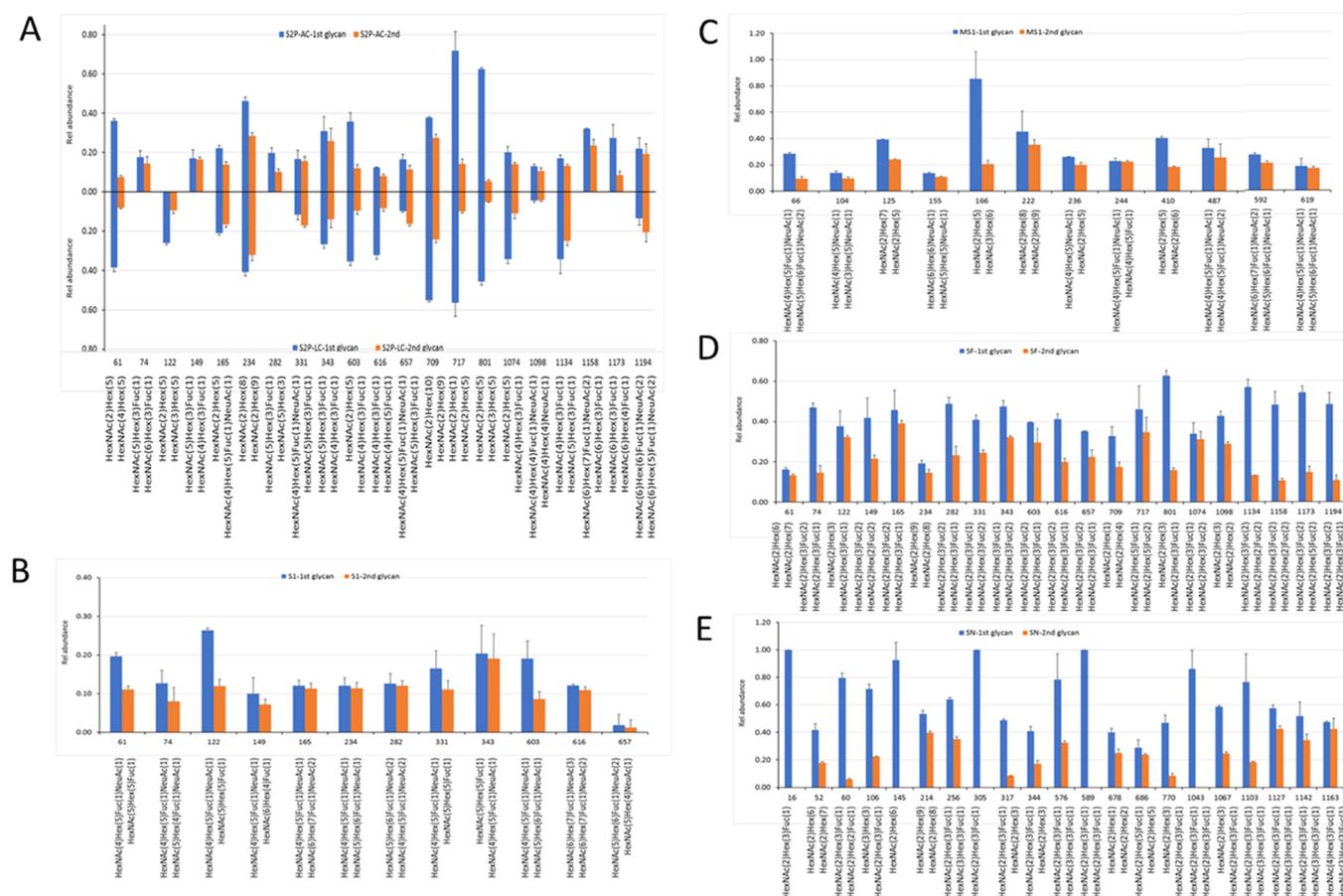
Two glycosylation sites (N234 and N709) were almost fully occupied by high-mannose glycans, while the sites of N61, N122, N603, N801, and N1074 were also modified by both high-mannose and complex glycans in comparable abundance. It was not surprising that this trend agreed well with the results reported on the same full-length S protein construct.<sup>13</sup> A significant difference between the two studies lies on the glycans attached to N717. The glycans detected at the N717 site were paucimannosidic or truncated small N-glycans in this report, but mainly high-mannose glycans in the other study. This inconsistency could be attributed to the differences in sample preparation, data acquisition, or data processing.

The difference in the N-glycosylation profiles on the SARS-CoV-2 S proteins reported recently by several laboratories<sup>13,37,38,50,51</sup> led us to investigate the effect of the S protein structure (full-length vs S1 subunit alone) on the glycan complement. We examined the N-glycosylation pattern of the SARS-CoV-2 S1 subunit prepared from HEK cells. The major glycan type in the available sequons of the S1 protein was still complex (Figure 2B). In comparison to the complete S2P construct, significant variations occurred. For instance, site N234 was almost fully occupied by complex glycans in S1 while the same site contained high-mannose glycans in S2P. This suggests that N234 may represent a location of a trimer-associated mannose patch, a phenomenon observed for HIV-1 envelope glycoprotein where underprocessed oligomannose-type glycans dominate some sites on the trimeric protein to stabilize its conformation.<sup>56</sup> In addition, N657 was solely modified with complex glycans on S2P, but less than 20% of the sites were occupied by complex glycans on S1 with the remainder of sites left unoccupied. Four sequons including N61, N122, N165, and N603 on the full-length spike contained a mixture of complex, hybrid, and high-mannose glycans, while complex glycans almost exclusively occupied the sites observed on the S1 subunit.

The composition of individual glycans of the full-length version and the S1 subunit of SARS-CoV-2 is displayed in Figure 3, illustrating the quantitative occupancy of the glycans with the highest and the second highest relative abundance on each glycosylation sites (referred to as the primary glycans thereafter). The reproducibility of our approach was demonstrated by the results obtained from two independent sample preparation methods, Asp-N/chymotrypsin digestion (Figure 3A upward) and Lys-C/chymotrypsin (Figure 3A, downward). As shown in the representative plots for S2P protein, the speciation and the relative abundance of the top 2 occupied N-glycans were comparable for majority of the modification sites. Although heterogeneity is a common phenomenon of N-linked glycosylation, our data show that many of the sites on SARS-CoV-2 S protein were mainly occupied by only one or two primary glycans. For example,



**Figure 2.** N-Glycosylation profile of recombinant S proteins including (A) stabilized SARS-CoV-2 S protein expressed in human cells (S2P); (B) SARS-CoV-2 S protein S1 subunit expressed in human cells; (C) MERS-CoV S protein S1 subunit expressed in human cells; (D) SARS-CoV-2 S protein expressed in insect cells; and (E) SARS-CoV-1 S protein expressed in insect cells. The data represent the sum of the mean values of three replicates for each type of glycans. The majority of the glycan data was obtained with Asp-N/chymotrypsin digestion.



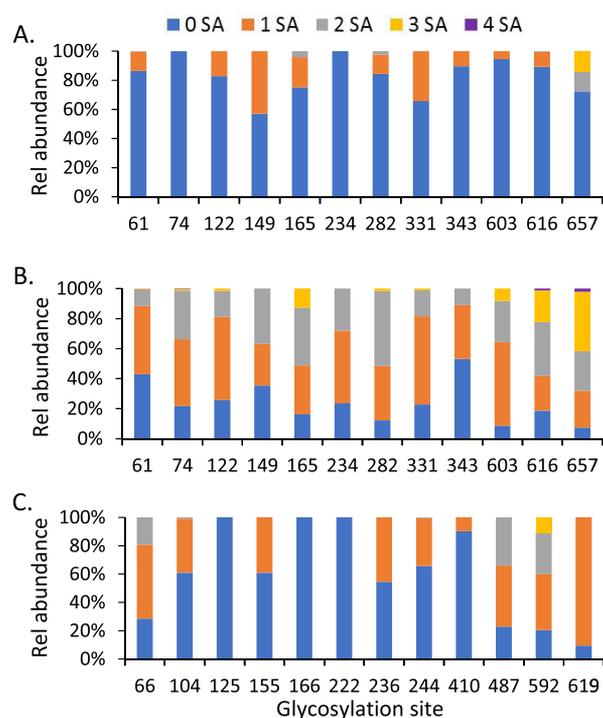
**Figure 3.** Relative abundance of the top 2 N-glycans on individual glycosylation sites of (A) SARS-CoV-2 prefusion ectodomain spike protein (S2P) derived from the digestion by Asp-N/chymotrypsin (upward) and Lys-C/chymotrypsin (downward). Note that the reflected *y*-axis represents positive values as per the labeled *y* scale; (B) SARS-CoV-2 S protein S1 subunit (S1); (C) MERS-CoV S protein S1 subunit (MS1); (D) SARS-CoV-2 S protein (SF); and (E) SARS-CoV S protein (SN). The data represent the average values of three replicates, and the error bars represent the standard deviation of the mean. Most of the data were obtained from the Asp-N/chymotrypsin digests.

more than half of each of the sites including N61, N234, N343, N603, N709, N717, N801, N1074, and N1134 were modified primarily by one or two glycans. Another interesting observation was some of the glycan species were predominant in multiple sites. For example, the biantennary complex glycan with the composition of HexNAc<sub>5</sub>Hex<sub>3</sub>Fuc occurred as the highest or the second highest abundant N-glycan on N74, N149, N282, N331, N343, N657, and N1134. Among those sites, N331 and N343 are located within the receptor-binding domain (RBD) of the SARS-CoV-2 S protein. Meanwhile, the high-mannose glycan Man<sub>5</sub>HexNAc<sub>2</sub>HexNAc<sub>2</sub>Hex<sub>5</sub> appeared on six modification sites, as one of the two primary modifiers. These observations revealed that only a few primary glycan species might play an important role in shielding the virus' immunogenic epitopes. In other words, the glycan density of viral spike protein was dominated by a few major glycan species although hundreds of distinct glycan species were identified on this protein.

The variation between the full-length and S1 subunit recombinant proteins was further demonstrated by the distribution of primary glycans (Figure 3A,B). In contrast to the ectodomain protein, S2P, the combined abundance of the top two glycans did not exceed 50% at any site of the S1 subunit, revealing a higher degree of microheterogeneity (different glycans at the same site) in the S1 protein than in S2P. In addition, the predominantly occurring glycans in these

two proteins were not the same. On the S1 protein, a possible biantennary complex glycan with the composition of HexNAc<sub>4</sub>Hex<sub>3</sub>FucNeuAc containing one fucose and one sialic acid modified 9 of 12 identified sites (N61, N74, N122, N149, N165, N234, N331, N343, and N603) including two sites on the RBD domain. In contrast, only three N-glycan sites (N165, N331, and N657) on the S2P were dominated by the complex type of glycans.

Another significant difference between the S2P and S1 proteins was the content of sialic acid glycans. Among 24 of the top 2 glycans detected on the 12 sequons in the sequence of the S1 subunit, 19 were sialylated glycans for S1 protein while only three N-glycans on the full-length S2P had sialic acid groups (Figure 4A,B). Increased sialylation on S1 subunit protein was further confirmed by comparing the relative ratio of sialylated glycans to nonsialylated ones on the two proteins (Figure 4). On each of the 12 sites of S2P protein, there were approximately 60% or higher total detected glycans containing no sialic acids; no sialylated glycans at all were detected on N74 and N234 (Figure 4A). In contrast, 9 of 12 sites on the S1 protein were sialylated by greater than 75%. Less than 50% of the glycans on the other three sites (N61, N149, N343) were unsialylated. In addition, there was a range in the degree of sialylation of S1 protein, including mono- and disialylated forms with some tri- and tetrasialylated glycans. The sialylation on the glycans of S2P was predominately the monosialylated



**Figure 4.** Distribution of sialylated N-glycans on the glycosylation sites within the sequence range of the S1 subunit of (A) the SARS-CoV-2 ectodomain S protein and (B) its S1 subunit recombinant protein, (C) and the S1 subunit of the MERS-CoV spike protein, MS1. “0 SA” indicates glycans without sialic acid attached, and “1–4 SA” represents mono-, bi-, tri-, and tetrasialylated glycans, respectively.

glycoforms. For the sequons located in the RBD domain of the spike protein, for N331 on the S1 protein, a total of 80% of glycans were sialylated (60% were monosialylated). In contrast, on the S2P protein, only 40% of the glycans were sialylated at sequon N331. On the other hand, only ~10% of the glycans on N343 of S2P were sialylated but approximately half of the glycoforms on the same residue of S1-contained terminal sialic acids.

Sialylated N-glycans play an important role in the immune system, pathogen recognition, protein–protein interactions, and cancer.<sup>57</sup> The substantial variation in sialic acid content between ectodomain and S1 subunit of SARS-CoV-2 S protein observed in our results suggests careful consideration in data interpretation is needed when a construct derived from a partial protein sequence is used to study ligand–receptor binding, antibody recognition, or other structure/function relationships. The existence of a large quantity of negatively charged branch-terminal sialic acids, particularly in a smaller protein, could affect the isoelectric point of this protein, and this may cause differences in the three-dimensional structure or conformation between full-length and reduced-size proteins.

To understand the effect of the primary sequence of the protein on the glycosylation profile and the formation of sialylated N-glycans, we examined the glycan profile of the S1 subunit (MS1) of the spike protein of MERS coronavirus (MERS-CoV), which belongs to the same coronavirus family as SARS-CoV-2. The recombinant MS1 has a similar molecular weight to the S1 protein (Figure 1B) and was prepared in HEK cells by the same manufacturer that produced the SARS-CoV-2 S1 recombinant protein examined here. The two proteins, S1

and MS1, differed significantly in the distribution of N-glycan types (Figure 2C) and also displayed clear variation in the composition of the primary glycans on each site (Figure 3C), presumably due to their variation in primary amino acid sequences. One-third of the N-glycosylation sites of MS1 (N125, N166, N222, N410) were high-mannose glycans with high occupancy (>80%), and 10–20% of sites N66, N104, and N236 were occupied by this glycan type. While complex glycans were still the major N-glycan type on the remaining sites, relatively low abundant hybrid glycans were also observed on some of these sites. Although the same biantennary complex glycan (HexNAc<sub>4</sub>Hex<sub>5</sub>FucNeuAc) occupied multiple sites in MS1 and 14 sialylated glycans were among 24 of the most frequently observed glycans, the high preference for nonsialylated glycans on most of the sequons implies that reduced size of a recombinant protein might not necessarily cause enhanced sialylation (Figure 4C). The amino acid sequence of the S1 subunit of SARS-CoV-2 S protein may play a role in its elevated sialic acid content when expressed as a recombinant protein. Further study is needed to confirm this speculation and its biological significance.

The baculovirus insect cell system has been widely utilized to produce functional, post-translationally modified recombinant proteins.<sup>58</sup> Since the beginning of the COVID-19 pandemic, this system has been used by different manufacturers and laboratories to produce many SARS-CoV-2 recombinant proteins for research and therapeutic purposes. A recent preprint has reported the N-glycosylation mapping on the surface of SARS-CoV-2 S protein expressed in insect cells, but no quantitative results regarding the relative abundance of glycans are included.<sup>37</sup> It is well known that expression hosts affect the glycosylation of a recombinant protein. Expression in baculoviral-insect cells is expected to form smaller or truncated glycans including paucimannosidic and other smaller species because the glycosylation pathway in insect cells is far simpler than in higher eukaryotes.<sup>59</sup> Characterizing the glycan profile of this protein under identical analytical conditions as the spike proteins expressed in the HEK expression systems was useful. When this protein construct (SF) was analyzed, not surprisingly, a vastly different N-glycosylation profile was observed (Figure 2D). Most of the N-glycosylation sequons were modified by truncated glycans with high abundances. For example, 15 Asn residues were modified with 80% or higher truncated glycans. There were, however, a few exceptions including N61, N234, and N717, where the main glycan type was high mannose and their relative abundances were comparable to those that were measured in human expressed S2P protein. A similar pattern was observed on a SARS-CoV S protein (SN) that was prepared in insect cells, although some sites such as N256, N1142, and 1163 were modified by complex glycans of comparable densities (Figure 2E).

O-glycosylation has been reported in SARS-CoV-2 full-length S protein constructs<sup>38,52</sup> and subunits.<sup>50</sup> In all cases, the relative degree of occupation for these observed O-glycosylation sites was relatively low—less than 11% in the report on full-length construct. In addition, Wells and colleagues have reported the observation of sulfonated glycans on a recombinant full-length S protein.<sup>38</sup> Prompted by these reports, we also searched our data for O-linked glycans and sulfoglycans in Byonic using a similar approach and searched identification criteria as we employed for N-glycosylation. Given our relatively strict and conservative search criteria (Byonic score  $\geq 300$ ), we observed only a few peptides

modified by sulfoglycans with a lower variety and diversity of modification sites than Wells and colleagues (Table S-1). On the other hand, we did not observe any occupied O-glycosylation sites that passed our filters. The relatively low amount (<1  $\mu\text{g}$ ) of proteins used in this study might hinder the detection of high-quality tandem mass spectra for low-abundance O-glycan or sulfoglycan peptides. In the future, we will more thoroughly characterize the O-linked and sulfonated glycan complement on SARS-CoV-2 proteins.

## CONCLUSIONS

We conducted a comprehensive characterization of the N-glycosylation profiles of several SARS-CoV-2 S protein constructs and contrasted them with the glycan profiles from SARS-CoV-1 and MERS. The combination of multiple enzyme digestions, glycan signature-triggered EThcD analysis, and rigorous data processing parameters allowed the detection and relative quantification of N-glycan distribution on the proteins with high confidence and reproducibility. The patterns of the N-glycosylation between two versions of the recombinant spike proteins, ectodomain and S1 subunit, were carefully examined, and significant variations were observed on the distribution of glycan types and specific individual glycans on various sequons of entire and partially expressed proteins. Our data demonstrate that the relative abundance of sialylated glycans was significantly elevated in the reduced-size S1 subunit than in the full-length protein. This observation could possibly point to differences in the structure and function of these two species. In addition, we compared N-glycan profiles of the recombinant S proteins produced from different expression systems including human cells and insect cells. These results should provide useful information for the study of the interactions of SARS-CoV-2 viral proteins and the development of effective vaccines and therapeutics.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.0c03301>.

List of the peptides/glycopeptides containing potential glycosylation sites (Data S-1) (XLSX)

List of the peptides containing sulfated glycans (Table S-1) (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

**Dongxia Wang** – Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, Georgia 30341, United States;

orcid.org/0000-0002-0993-1018; Phone: 770-488-0446; Email: [dov2@cdc.gov](mailto:dov2@cdc.gov); Fax: 770-488-0509

**John R. Barr** – Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, Georgia 30341, United States; Phone: 770-488-7848; Email: [jbarr@cdc.gov](mailto:jbarr@cdc.gov); Fax: 770-488-0509

### Authors

**Jakub Baudys** – Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, Georgia 30341, United States

**Jonathan L. Bundy** – Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, Georgia 30341, United States

**Maria Solano** – Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, Georgia 30341, United States;

orcid.org/0000-0002-7853-3523

**Theodore Keppel** – Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, Georgia 30341, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.analchem.0c03301>

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors thank Daniel Wrapp, Nianshuang Wang, and Jason McLellan for providing recombinant SARS-CoV-2 prefusion stabilized spike protein. They also thank Eric Wang for processing some data using tools developed in R. The following reagent was obtained through BEI Resources, NIAID, NIH: SARS-CoV spike (S) protein with C-terminal Histidine Tag, Recombinant from Baculovirus, NR-686. References in this paper to any specific commercial products, process, service, manufacturer, or company do not constitute an endorsement or a recommendation by the U.S. Government or the Centers for Disease Control and Prevention. The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

## REFERENCES

- (1) Bassetti, M.; Vena, A.; Giacobbe, D. R. *Eur. J. Clin. Invest.* **2020**, *50*, No. e13209.
- (2) Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; Cheng, Z.; Yu, T.; Xia, J.; Wei, Y.; Wu, W.; Xie, X.; Yin, W.; Li, H.; Liu, M.; Xiao, Y.; et al. *Lancet* **2020**, *395*, 497–506.
- (3) Yang, X.; Yu, Y.; Xu, J.; Shu, H.; Xia, J.; Liu, H.; Wu, Y.; Zhang, L.; Yu, Z.; Fang, M.; Yu, T.; Wang, Y.; Pan, S.; Zou, X.; Yuan, S.; Shang, Y. *Lancet Respir. Med.* **2020**, *8*, 475–481.
- (4) COVID-19 Weekly Epidemiological Update; World Health Organization: Geneva, 2020 (Sept 7, 2020).
- (5) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veelsler, D. *Cell* **2020**, *181*, 281.e6–292.e6.
- (6) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. *Nature* **2020**, *581*, 215–220.
- (7) Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. *Nature* **2020**, *581*, 221–224.
- (8) da Costa, V. G.; Moreli, M. L.; Saivish, M. V. *Arch. Virol.* **2020**, *165*, 1517–1526.
- (9) Kaul, D. *Curr. Med. Res. Pract.* **2020**, *10*, 54–64.
- (10) Walls, A. C.; Tortorici, M. A.; Bosch, B. J.; Frenz, B.; Rottier, P. J. M.; DiMaio, F.; Rey, F. A.; Veelsler, D. *Nature* **2016**, *531*, 114–117.
- (11) Walls, A. C.; Tortorici, M. A.; Frenz, B.; Snijder, J.; Li, W.; Rey, F. A.; DiMaio, F.; Bosch, B. J.; Veelsler, D. *Nat. Struct. Mol. Biol.* **2016**, *23*, 899–905.
- (12) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C. L.; Abiona, O.; Graham, B. S.; McLellan, J. S. *Science* **2020**, *367*, 1260–1263.
- (13) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. *Science* **2020**, *369*, 330–333.
- (14) Lin, B.; Qing, X.; Liao, J.; Zhuo, K. *Cells* **2020**, *9*, No. 1022.

- (15) Watanabe, Y.; Bowden, T. A.; Wilson, I. A.; Crispin, M. *Biochim. Biophys. Acta, Gen. Subj.* **2019**, *1863*, 1480–1497.
- (16) Hu, H.; Khatri, K.; Klein, J.; Leymarie, N.; Zaia, J. *Glycoconjugate J.* **2016**, *33*, 285–296.
- (17) Nilsson, J. *Glycoconjugate J.* **2016**, *33*, 261–272.
- (18) Hever, H.; Darula, Z.; Medzihradsky, K. F. *Methods Mol. Biol.* **2019**, *1934*, 93–125.
- (19) Hoffmann, M.; Marx, K.; Reichl, U.; Wuhler, M.; Rapp, E. *Mol. Cell. Proteomics* **2016**, *15*, 624–641.
- (20) Neubert, P.; Halim, A.; Zauser, M.; Essig, A.; Joshi, H. J.; Zatorska, E.; Larsen, I. S.; Loibl, M.; Castells-Ballester, J.; Aebi, M.; Clausen, H.; Strahl, S. *Mol. Cell. Proteomics* **2016**, *15*, 1323–1337.
- (21) Leung, K. K.; Wilson, G. M.; Kirkemo, L. L.; Riley, N. M.; Coon, J. J.; Wells, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 7764–7775.
- (22) Yang, L.; Sun, Z.; Zhang, L.; Cai, Y.; Peng, Y.; Cao, T.; Zhang, Y.; Lu, H. *Chem. Sci.* **2019**, *10*, 9302–9307.
- (23) Williams, T. L.; Luna, L.; Guo, Z.; Cox, N. J.; Pirkle, J. L.; Donis, R. O.; Barr, J. R. *Vaccine* **2008**, *26*, 2510–2520.
- (24) Pierce, C. L.; Williams, T. L.; Moura, H.; Pirkle, J. L.; Cox, N. J.; Stevens, J.; Donis, R. O.; Barr, J. R. *Anal. Chem.* **2011**, *83*, 4729–4737.
- (25) Williams, T. L.; Pirkle, J. L.; Barr, J. R. *Vaccine* **2012**, *30*, 2475–2482.
- (26) Santana, W. I.; Williams, T. L.; Winne, E. K.; Pirkle, J. L.; Barr, J. R. *Anal. Chem.* **2014**, *86*, 4088–4095.
- (27) Johnson, A.; Chen, L. M.; Winne, E.; Santana, W.; Metcalfe, M. G.; Mateu-Petit, G.; Ridenour, C.; Hossain, M. J.; Villanueva, J.; Zaki, S. R.; Williams, T. L.; Cox, N. J.; Barr, J. R.; Donis, R. O. *PLoS One* **2015**, *10*, No. e0128982.
- (28) Ridenour, C.; Johnson, A.; Winne, E.; Hossain, J.; Mateu-Petit, G.; Balish, A.; Santana, W.; Kim, T.; Davis, C.; Cox, N. J.; Barr, J. R.; Donis, R. O.; Villanueva, J.; Williams, T. L.; Chen, L. M. *Influenza Other Respir. Viruses* **2015**, *9*, 263–270.
- (29) Morgenstern, K.; Xie, Y.; Palladino, G.; Barr, J. R.; Settembre, E. C.; Williams, T. L.; Wen, Y. *Vaccine* **2018**, *36*, 6144–6151.
- (30) Pierce, C. L.; Williams, T. L.; Santana, W. I.; Levine, M.; Chen, L. M.; Cooper, H. C.; Solano, M. I.; Woolfitt, A. R.; Marasco, W. A.; Fang, H.; Donis, R. O.; Barr, J. R. *Vaccine* **2017**, *35*, 5011–5018.
- (31) Blake, T. A.; Williams, T. L.; Pirkle, J. L.; Barr, J. R. *Anal. Chem.* **2009**, *81*, 3109–3118.
- (32) Chang, D.; Zaia, J. *Mol. Cell. Proteomics* **2019**, *18*, 2348–2358.
- (33) Cipollo, J. F.; Parsons, L. M. *Mass Spectrom. Rev.* **2020**, *39*, 371–409.
- (34) Hargett, A. A.; Renfrow, M. B. *Curr. Opin. Virol.* **2019**, *36*, 56–66.
- (35) Bezstarosti, K.; Lamers, M. M.; Haagmans, B. L.; Demmers, J. A. A. *bioRxiv* **2020**, 1–20.
- (36) Davidson, A. D.; Williamson, M. K.; Lewis, S.; Shoemark, D.; Carroll, M. W.; Heesom, K.; Zambon, M.; Ellis, J.; Lewis, P. A.; Hiscox, J. A.; Matthews, D. A. *bioRxiv* **2020**, 1–39.
- (37) Zhang, Y.; Zhao, W.; Mao, Y.; Wang, S.; Zhong, Y.; Su, T.; Gong, M.; Du, D.; Lu, X.; Cheng, J.; Yang, H. *bioRxiv* **2020**, 1–37.
- (38) Zhao, P.; Praissman, J. L.; Grant, O. C.; Cai, Y.; Xiao, T.; Rosenbalm, K. E.; Aoki, K.; Kellman, B. P.; Bridger, R.; Barouch, D. H.; Brindley, M. A.; Lewis, N. E.; Tiemeyer, M.; Chen, B.; Woods, R. J.; Wells, L. *Cell Host Microbe* **2020**, *28*, 586.e6–601.e6.
- (39) Klein, J. A.; Zaia, J. *bioRxiv* **2020**, 1–39.
- (40) Chalk, R.; Greenland, W.; Moreira, T.; Coker, J.; Mukhopadhyay, S. M. M.; Williams, E.; Manning, C.; Bohstedt, T.; McCrorie, R.; Fernandez-Cid, A.; Burgess-Brown, N. A. *bioRxiv* **2020**, 1–13.
- (41) Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528–9533.
- (42) Coon, J. J.; Ueberheide, B.; Syka, J. E.; Dryhurst, D. D.; Ausio, J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 9463–9468.
- (43) Garcia, B. A.; Shabanowitz, J.; Hunt, D. F. *Methods* **2005**, *35*, 256–264.
- (44) Mikesch, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E.; Shabanowitz, J.; Hunt, D. F. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764*, 1811–1822.
- (45) Frese, C. K.; Altelaar, A. F.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J.; Mohammed, S. *Anal. Chem.* **2012**, *84*, 9668–9673.
- (46) Frese, C. K.; Zhou, H.; Taus, T.; Altelaar, A. F.; Mechtler, K.; Heck, A. J.; Mohammed, S. *J. Proteome Res.* **2013**, *12*, 1520–1525.
- (47) Yu, Q.; Wang, B.; Chen, Z.; Urabe, G.; Glover, M. S.; Shi, X.; Guo, L. W.; Kent, K. C.; Li, L. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1751–1764.
- (48) Čavala, T.; Zhu, J.; Heck, A. J. R. *Anal. Chem.* **2019**, *91*, 10401–10406.
- (49) Lermyte, F.; Valkenburg, D.; Loo, J. A.; Sobott, F. *Mass Spectrom. Rev.* **2018**, *37*, 750–771.
- (50) Shajahan, A.; Supekar, N. T.; Gleinich, A. S.; Azadi, P. *Glycobiology* **2020**, No. cwaa042.
- (51) Watanabe, Y.; Berndsen, Z. T.; Raghwan, J.; Seabright, G. E.; Allen, J. D.; Pybus, O. G.; McLellan, J. S.; Wilson, I. A.; Bowden, T. A.; Ward, A. B.; Crispin, M. *Nat. Commun.* **2020**, *11*, No. 2688.
- (52) Sanda, M.; Morrison, L.; Goldman, R. *bioRxiv* **2020**, 1–6.
- (53) Saba, J.; Dutta, S.; Hemenway, E.; Viner, R. *Int. J. Proteomics* **2012**, *2012*, No. 560391.
- (54) Bern, M.; Cai, Y.; Goldberg, D. *Anal. Chem.* **2007**, *79*, 1393–1400.
- (55) Riley, N. M.; Malaker, S. A.; Driessen, M. D.; Bertozzi, C. R. *J. Proteome Res.* **2020**, *19*, 3286–3301.
- (56) Behrens, A. J.; Harvey, D. J.; Milne, E.; Cupo, A.; Kumar, A.; Zitzmann, N.; Struwe, W. B.; Moore, J. P.; Crispin, M. *J. Virol.* **2017**, *91*, No. e01894-16.
- (57) Bhide, G. P.; Colley, K. J. *Histochem. Cell Biol.* **2017**, *147*, 149–174.
- (58) Mishra, V. *Protein Pept. Lett.* **2020**, *27*, 529–537.
- (59) Shi, X.; Jarvis, D. L. *Curr. Drug Targets* **2007**, *8*, 1116–1125.