

Site-specific *N*-glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins

Authors

Yong Zhang, Wanjun Zhao, Yonghong Mao, Yaohui Chen, Shisheng Wang, Yi Zhong, Tao Su, Meng Gong, Dan Du, Xiaofeng Lu, Jingqiu Cheng, and Hao Yang

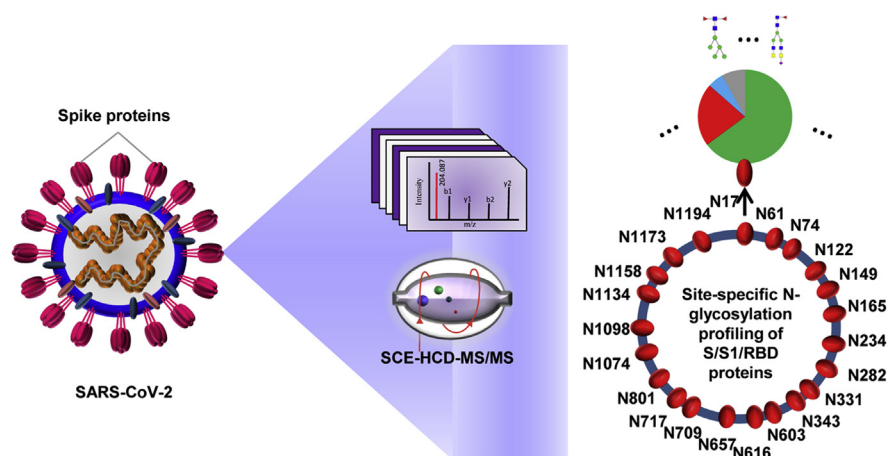
Correspondence

jqcheng@scu.edu.cn; yanghao@scu.edu.cn

Graphical Abstract

In Brief

Site-specific *N*-glycosylation of SARS-CoV-2 spike proteins expressed in human and insect cells has been decoded *via* enriched glycopeptides using a stepped collision energy HCD scan. The regular *N*-glycan occupancy on preserved *N*-glycosites has been revealed, despite the heterogeneity in *N*-glycan compositions. The complex *N*-glycans on the human cell-expressed spike protein reflect the relatively accessible antigens and epitopes for immune recognition. The insect cell-expressed spike protein subunits decorated with immature high-mannose *N*-glycans might become promising recombinant vaccine candidates.



Highlights

- Decoding *N*-glycosites and site-specific *N*-glycans on spike protein
- The variation of *N*-glycosylation sequons among 753 SARS-CoV-2 variants
- *N*-glycosylation comparison among different lengths of spike proteins from different hosts
- Prospective insights on host immunity and vaccine development

Site-specific *N*-glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins

Yong Zhang^{1,2,‡}, Wanjun Zhao^{3,‡}, Yonghong Mao⁴, Yaohui Chen⁴, Shisheng Wang^{1,2}, Yi Zhong¹, Tao Su¹, Meng Gong^{1,2}, Dan Du¹, Xiaofeng Lu^{1,2}, Jingqiu Cheng^{1,2,*}, and Hao Yang^{1,2,*}

The glycoprotein spike (S) on the surface of severe acute respiratory syndrome coronavirus (SARS-CoV-2) is a determinant for viral invasion and host immune response. Herein, we characterized the site-specific *N*-glycosylation of S protein at the level of intact glycopeptides. All 22 potential *N*-glycosites were identified in the S-protein protomer and were found to be preserved among the 753 SARS-CoV-2 genome sequences. The glycosites exhibited glycoform heterogeneity as expected for a human cell-expressed protein subunit. We identified masses that correspond to 157 *N*-glycans, primarily of the complex type. In contrast, the insect cell-expressed S protein contained 38 *N*-glycans, completely of the high-mannose type. Our results revealed that the glycan types were highly determined by the differential processing of *N*-glycans among human and insect cells, regardless of the glycosites' location. Moreover, the *N*-glycan compositions were conserved among different sizes of subunits. Our study indicates that the S protein *N*-glycosylation occurs regularly at each site, albeit the occupied *N*-glycans were diverse and heterogenous. This *N*-glycosylation landscape and the differential *N*-glycan patterns among distinct host cells are expected to shed light on the infection mechanism and present a positive view for the development of vaccines and targeted drugs.

The spread of a novel severe acute respiratory syndrome coronavirus (SARS-CoV-2) has caused a pandemic of coronavirus disease 2019 (COVID-19) worldwide. Distinguished from severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), SARS-CoV-2 transmits more rapidly and efficiently from infected individuals, even those without symptoms, to healthy humans, frequently leading to severe or lethal respiratory symptoms (1, 2). The World Health Organization has declared the spread of SARS-CoV-2 a Public Health Emergency of International Concern. As of August 11, 2020, SARS-CoV-2 has led to over 20 million confirmed cases. From

SARS-CoV to SARS-CoV-2, the periodic outbreak of highly pathogenic coronavirus infections in humans urgently calls for strong prevention and intervention measures. However, there are no approved vaccines or effective antiviral drugs for either SARS-CoV or SARS-CoV-2.

Human coronaviruses, including HCoV 229E, NL63, OC43, and HKU1, are responsible for 10 to 30% of all upper respiratory tract infections in adults. SARS-CoV-2 can actively replicate in the throat; however, this virus predominately infects the lower respiratory tract *via* the envelope spike (S) protein (2, 3). Due to its high exposure on the viral surface, the S protein can prime a protective humoral and cellular immune response, thus commonly serving as the main target for antibodies, entry inhibitors, and vaccines (4–6). It has been found that human sera from recovered COVID-19 patients can neutralize SARS-CoV-2 S protein-overexpressed pseudovirions effectively (7). The use of convalescent sera in the clinic is actively undergoing a comprehensive evaluation. However, the passive antibody therapy with convalescent sera would be a stopgap measure and may not provide a protective immunity owing to their limited cross reactions (7, 8). Therefore, specific neutralizing antibodies and vaccines against SARS-CoV-2 are in rapid development for providing potent and long-lasting immune protection (9–11).

A mature SARS-CoV-2 has four structural proteins, including the S protein, envelope (E) protein, membrane (M) protein, and nucleocapsid (N) protein (1). Given its indispensable role in viral entry and infectivity, the S protein is probably the most promising immunogen, especially given the comprehensive understanding of the structure and function provided by recent studies (4, 12–14). The S protein is comprised of an ectodomain, a transmembrane anchor, and a short C-terminal intracellular tail (15). The ectodomain consists of a receptor-binding S1 subunit and a membrane-fusion S2 subunit. Following attachment to the host cell surface *via* S1, the S protein is cleaved at multiple sites by host cellular

From the ¹Key Laboratory of Transplant Engineering and Immunology, Ministry of Health, Institutes for Systems Genetics, Frontiers Science Center for Disease-Related Molecular Network, ²Regenerative Medical Research Center, ³Department of Thyroid Surgery, ⁴Department of Thoracic Surgery, West China Hospital, Sichuan University, Chengdu, China

[‡]These authors contributed equally to this work.

*For correspondence: Hao Yang, yanghao@scu.edu.cn; Jingqiu Cheng, jqcheng@scu.edu.cn.

proteases, consequently mediating membrane fusion and making way for the viral genetic materials to enter the host cell (1, 4, 6). The S protein can bind to the angiotensin-converting enzyme II (ACE2) receptor on host cells (13, 16). The recognition of the S protein to the ACE2 receptor primarily involves extensive polar residue interactions between the receptor-binding domain (RBD) and the peptidase domain of ACE2 (13, 14). The S protein RBD is located in the S1 subunit and undergoes a hinge-like dynamic movement to capture the receptor through three grouped residue clusters. Consequently, the S protein of SARS-CoV-2 displays an up to 10- to 20-fold higher affinity for the human ACE2 receptor than that of SARS-CoV, supporting the higher transmissibility of this new virus (13, 14).

Apart from the structural information at the residue level, the trimeric S protein is highly glycosylated, possessing 22 potentially N-linked glycosylation motifs (N-X-S/T, X/=P) in each protomer (17–19). The N-glycans on the S protein play a pivotal role in proper protein folding and protein priming by host proteases. Importantly, glycosylation is an underlying mechanism for coronavirus to evade both innate and adaptive immune responses of their hosts, as the glycans might shield the amino acid residues of viral epitopes from cell and antibody recognition (4, 5, 20). Cryo-EM has revealed the N-glycosylation on 14 to 16 of 22 potential sites in the SARS-CoV-2 S protein protomer (4, 13). However, these glycosites and their glycan occupancies need to be experimentally identified in detail. Glycosylation analysis *via* glycopeptides can provide insight into the N-glycan microheterogeneity of a specific site (21). Therefore, further identification of site-specific N-glycosylation information of the SARS-CoV-2 S protein, including that regarding intact N-glycopeptides, glycosites, glycan compositions, and the site-specific number of glycans, could be meaningful to obtain a deeper understanding of the mechanism of the viral invasion and provide guidance for vaccine design and antiviral therapeutics development (4, 22). Thus, Watanabe *et al.* (2020) (17) first identified the 22 potential N-glycosites and their linked N-glycans on the human cell-expressed recombinant S protein variant.

In this study, we characterized the site-specific N-glycosylation of both human and insect cell-produced SARS-CoV-2 S proteins by comparative analysis of the intact glycopeptides using tandem mass spectrometry (MS/MS). Based on an integrated method (23), we identified 22 potential N-glycosites and their corresponding N-glycans from the recombinant S protein. All of these glycosites were found to be highly conserved among SARS-CoV-2 genome sequences. The glycosite-specific occupancy by different glycoforms was resolved and compared among S protein subunits expressed in human cells and insect cells. These detailed glycosylation profiles decoded from MS/MS analysis are expected to facilitate the development of vaccines and therapeutic drugs against SARS-CoV-2.

EXPERIMENTAL PROCEDURES

Experimental Design and Statistical Rationale

Recombinant SARS-CoV-2 spike proteins or the subunits (100 µg) expressed in insect and human cells were digested using trypsin, Glu-C, and a combination of trypsin and Glu-C. Digestion products were enriched by zwitterionic hydrophilic interaction liquid chromatography (Zic-HILIC) and digested using PNGase F. Finally, the intact N-glycopeptides before enrichment, the intact N-glycopeptides after enrichment, and deglycosylated peptides were analyzed by stepped collision energy–higher-energy collisional dissociation (SCE-HCD)-MS/MS. Data analysis was performed using the Byonic software (version 3.6.0, Protein Metrics, Inc) and was verified manually. Three technical replicates were performed. The number of intact N-glycopeptides and N-glycans identified from triplicate replicates was analyzed by Student's *t*-test for statistical comparison between the two groups, before and after enrichment. Data were presented as mean ± SD. A *p*-value < 0.05 was considered significant.

Materials and Chemicals

Dithiothreitol (DTT), iodoacetamide (IAA), formic acid (FA), trifluoroacetic acid (TFA), TRIS base, and urea were purchased from Sigma. Acetonitrile (ACN) was purchased from Merck. The Zic-HILIC materials were obtained from Fresh Bioscience. The recombinant SARS-CoV-2 S protein (S1+S2 ECD and RBD) expressed in insect cells (High Five) and S protein (S1 and RBD) expressed in human embryonic kidney cells (HEK293) were purchased from Sino Biological. Codon-optimized DNA sequences encoding the SARS-CoV-2 S protein subunits were cloned into pCMV3-C-His and a baculovirus vector with a polyhistidine tag at the C terminus for recombinant expression of these proteins in human and insect cells, respectively. Sequencing-grade trypsin and Glu-C were obtained from Enzyme & Spectrum. The quantitative colorimetric peptide assay kit was purchased from Thermo Fisher Scientific. Deionized water was prepared using a Milli-Q system (Millipore). All other chemicals and reagents of the best available grade were purchased from Sigma-Aldrich or Thermo Fisher Scientific.

Protein Digestion

The recombinant S proteins were proteolyzed using an in-solution protease digestion protocol. In brief, 50 µg of protein in a tube was denatured for 10 min at 95 °C. After reduction by DTT (20 mM) for 45 min at 56 °C and alkylating with IAA (50 mM) for 1 h at 25 °C in the dark, 2 µg of protease (trypsin, Glu-C, and a combination of trypsin and Glu-C) was added to the tube and incubated for 16 h at 37 °C. After desalting using a pipette tip packed with a C18 membrane, the peptide concentration was determined using a peptide assay kit based on the absorbance measured at 480 nm. The peptide mixtures (intact N-glycopeptides before enrichment) were freeze-dried for further analysis.

Enrichment of Intact N-glycopeptides

Intact N-glycopeptides were enriched with Zic-HILIC (Fresh Bioscience). Specifically, 20 µg of peptides was resuspended in 100 µl of 80% ACN/0.2% TFA solution, and 2 mg of processed Zic-HILIC was added to the peptide solution and rotated for 2 h at 37 °C. Finally, the mixture was transferred to a 200-µl pipette tip packed with a C8 membrane and washed twice with 80% ACN/0.2% TFA. After enrichment, intact N-glycopeptides were eluted three times with 70 µl of 0.1% TFA and dried using a SpeedVac for further analysis.

Deglycosylation

Enriched intact N-glycopeptides were digested using 1 U PNGase F dissolved in 50 µl of 50 mM NH₄HCO₃ for 2 h at 37 °C. The reaction

was terminated by the addition of 0.1% formic acid (FA). The deglycosylated peptides were dried using a SpeedVac for further analysis.

Liquid Chromatography–MS/MS Analysis

All samples were analyzed using SCE-HCD-MS/MS with an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific). In brief, intact *N*-glycopeptides before or after enrichment and deglycosylated peptides were dissolved in 0.1% FA and separated on a column (ReproSil-Pur C18-AQ, 1.9 μ m, 75 μ m inner diameter, length 20 cm; Dr Maisch) over a 78-min gradient (buffer A, 0.1% FA in water; buffer B, 0.1% FA in 80% ACN) at a flow rate of 300 nl/min. MS1 was analyzed using a scan range (*m/z*) of 800 to 2000 (intact *N*-glycopeptides before or after enrichment) or 350 to 1550 (deglycosylated peptides) at an Orbitrap resolution of 120,000. The RF lens, AGC target, maximum injection time, and exclusion duration were 30%, 2.0 e^4 , 100 ms, and 15 s, respectively. MS2 was analyzed with an isolation window (*m/z*) of 2 at an Orbitrap resolution of 15,000. The AGC target, maximum injection time, and the HCD type were standard, 250 ms, and 30%, respectively. The stepped collision mode was turned on with an energy difference of $\pm 10\%$ (20–30–40%).

Data Analysis

The raw data files were searched against the SARS-CoV-2 S protein sequence using Byonic software (version 3.6.0, Protein Metrics, Inc), with the mass tolerance for precursors and fragment ions set at ± 10 ppm and ± 20 ppm, respectively. Two missed cleavage sites were allowed for trypsin or/and Glu-C digestion. The fixed modification was carbamidomethyl (C), and variable modifications included oxidation (M), acetyl (protein N-term), and deamidation (N). In addition, 38 insect *N*-glycans or 182 human *N*-glycans were specified as *N*-glycan modifications for intact *N*-glycopeptides before or after enrichment. We then checked the protein database options, including the decoy database. All other parameters were set at the default values, and protein groups were filtered to a 1% false discovery rate based on the number of hits obtained for searches against these databases. Stricter quality control methods for intact *N*-glycopeptides and peptide identification were implemented, requiring a score of no less than 200 and identification of at least six amino acids.

Furthermore, all peptide spectrum matches (PSMs) and glycopeptide-spectrum matches (GPSMs) were examined manually. Quantitative analysis of the intact *N*-glycopeptide was performed as described previously (24). Quantification information (MS1 peak intensity) of *N*-glycopeptides spectra was acquired from the “allpeptide.txt” file in MaxQuant (Max Planck Gesellschaft) based on their unique MS/MS scan numbers from Byonic results”. *N*-glycosite conservation analysis was performed using R software packages. Model building based on the Cryo-EM structure (PDB: 6VSB) of SARS-CoV-2 S protein was performed using PyMOL.

RESULTS

Strategy for Site-specific *N*-glycosylation Characterization

Previous studies have revealed that the glycosylated coronavirus S protein plays a critical role in the induction of neutralizing antibodies and protective immunity. However, the glycans on S protein might also surround the protein surface and form an immunologically inert “self” glycan shield for virus evasion from the immune system (5, 20, 25). Herein, we aimed to decode the detailed site-specific *N*-glycosylation profile of the SARS-CoV-2 S protein. The native S protein ectodomain expressed by a baculovirus expression vector in insect cells with no amino acid substitution was first used to analyze the glycosylation landscape, since this expression system can

express the S protein without resulting in splicing *via* host proteases. Moreover, insect cells can mimic the process of mammalian cell glycosylation. Insect and human cells can yield a common intermediate *N*-glycan in the *N*-glycosylation processing pathway. The intermediate product is usually elongated by further adding various oligosaccharides to form the complex end product in human cells while often being trimmed without elongation, forming paucimannose in insect cells (4, 26). The S protein of the human coronavirus NL63 (HCoV-NL63) has been successfully expressed in insect cells for the resolution of its protein structure and glycan shield (20). The recombinant SARS-CoV-2 S ectodomain contains 1209 amino acids (residues 16–1213) that are translated from a complete genome sequence (GenBank: MN908947.3) (27) and includes 22 putative *N*-glycosylation sequons (motif N-X-S/T, X/=P). Theoretical analysis of the enzymatic sites showed that trypsin alone did not produce peptides of sufficiently appropriate length of peptides to cover all potential *N*-glycosites (supplemental Fig. S1A). The missing potential *N*-glycosites were found back by introducing the endoprotease Glu-C (supplemental Fig. S1B). Hence, we took advantage of this complementary trypsin and Glu-C digestion approach by using either a single or a dual enzyme (supplemental Fig. S1C). Meanwhile, the recombinant SARS-CoV-2 S protein S1 subunit expressed in human cells was obtained for analysis of the site-specific *N*-glycans, as the *N*-glycan compositions in insect cells would be different from those in native human host cells (26). The S1 subunit contains 681 amino acids (residues 16–685) and 13 potential *N*-glycosites. Trypsin alone or dual digestion could cover all potential *N*-glycosites (supplemental Fig. S1C). By doing so, each *N*-glycosylation sequon of the different recombinant proteins was covered by glycopeptides of a suitable length for achieving good ionization and fragmentation.

To raise the abundance of intact glycopeptides, Zic-HILIC materials were used to enrich glycopeptides. Concurrently, the enrichment of intact *N*-glycopeptides can reduce signal suppression from unglycosylated peptides. However, there are no materials available that can capture all glycopeptides without preference. For these reasons, site-specific glycosylation was determined based on a combined analysis of the intact *N*-glycopeptides before and after enrichment. Furthermore, the deglycosylated peptides following enrichment were used to confirm or retrieve the *N*-glycosites by removing potential interferences from glycans. In brief, the integration of complementary digestion and *N*-glycoproteomic analysis at three levels (before and after enrichment and at the deglycopeptides levels) is a promising approach to comprehensively and confidently profile the site-specific *N*-glycosylation of recombinant SARS-CoV-2 S proteins (Fig. 1).

N-glycosite Landscape of Recombinant SARS-CoV-2 S Proteins

The S protein contains 22 potential *N*-glycosites. Using our integrated analysis method, 21 glycosites were assigned

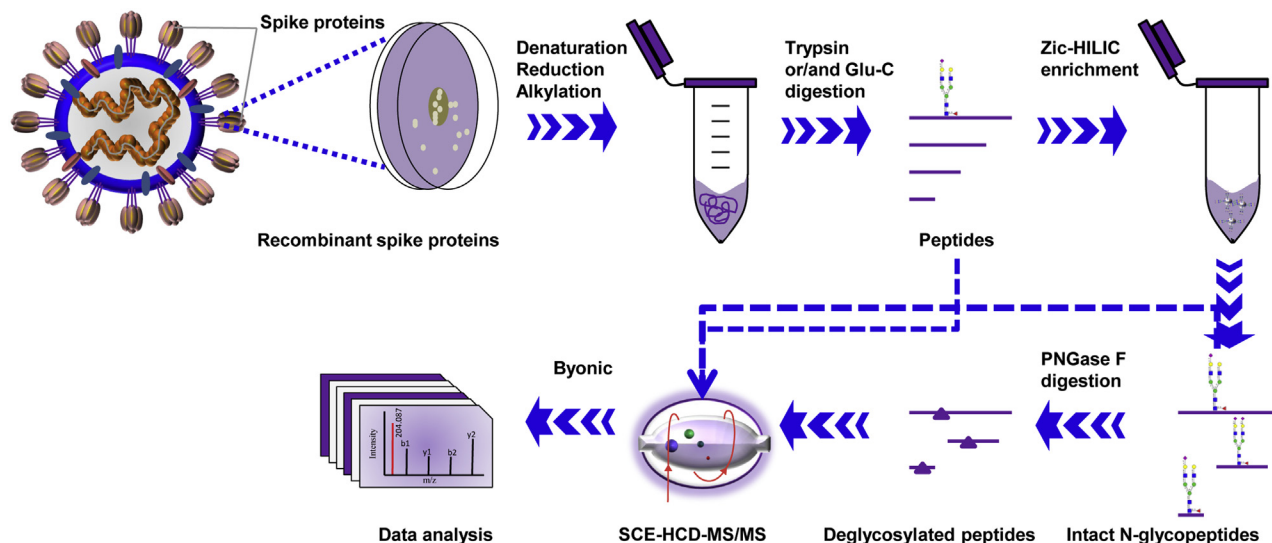


FIG. 1. Workflow for site-specific *N*-glycosylation characterization of recombinant SARS-CoV-2 S proteins using two complementary proteases for digestion and an integrated *N*-glycoproteomic analysis.

unambiguously with high-quality spectral evidence (Fig. 2A and supplemental Table S1). One glycosite, N1134, was ambiguously assigned with relatively lower spectral scores (score <200) (supplemental Fig. S2). Nevertheless, the N1134 glycosite has been observed in the Cryo-EM structure of the SARS-CoV-2 S protein (4). The relatively low spectral evidence of this glycosite indicates that a low-frequency glycosylation may occur, because our integrated methods, including glycopeptide enrichment and deglycosylation, failed to improve the spectra. Apart from the canonical *N*-glycosylation sequons, three noncanonical motifs of *N*-glycosites (N164, N334, and N536) involving *N*-X-C sequons were not identified as glycosylated. Before enrichment, an average of 15 *N*-glycosites from trypsin-digested peptides and 13 *N*-glycosites from Glu-C-digested peptides were assigned. In contrast, hydrophilic enrichment resulted in a significant increase of these glycosites to 18 and 16, respectively (supplemental Table S1). The representative spectra of one intact *N*-glycopeptide (N149) before and after enrichment are shown in supplemental Fig. S3. Complementary digestion with trypsin and Glu-C promoted the confident identification of four *N*-glycosites (N603, N616, N709, and N717) on two intact *N*-glycopeptides (supplemental Table S1 and supplemental Fig. S1C). The introduction of Glu-C digestion resulted in the production of two short intact *N*-glycopeptides containing 23 and 36 amino acids, respectively. These peptides are more suitable for achieving better ionization and fragmentation than the long peptide of 48 and 57 amino acids obtained from trypsin digestion (supplemental Fig. S4). Deglycopeptides are suitable for verifying glycosylation sites (supplemental Fig. S5). Unexpectedly, deglycopeptide peptides led to the loss of a few glycosites, presumably because of peptide loss during deglycosylation procedures. However, almost all

glycosites were confidently confirmed using trypsin and Glu-C dual digestion (supplemental Table S1).

For the recombinant protein S1 subunit expressed in human cells, all 13 *N*-glycosites were assigned unambiguously (supplemental Table S2). Finally, we profiled all 22 potential *N*-glycosites of S protein (supplemental Tables S3 and S4). These sites were preferentially distributed in the S1 subunit of the N terminus and the S2 subunit of the C terminus, including two sites in the RBD (Fig. 2, A and B). To visualize the *N*-glycosylation on the protein structure, all of the experimentally determined *N*-glycosites were hand-marked on the surface of the trimeric S protein following refinement of the recently reported SARS-CoV-2 S protein Cryo-EM structure (PDB: 6VSB) (Fig. 2C) (13).

Based on these findings, we further analyzed the conservation of glycosites among 753 SARS-CoV-2 genome sequences from the Global Initiative on Sharing All Influenza Data (GISAID) database. After the removal of redundant sequences of the S protein at the amino acid residue level, we refined 145 S protein variants. A very low frequency of alterations in 38 residue sites was found, uniformly spanning the full length of the S protein among all S variants, except for the variant D614G, which was identified at a high frequency in 98 variants (supplemental Table S5). However, nearly all of the 22 *N*-glycosylated sequons were conserved in the S protein, except for the loss of the N717 glycosite due to the T719A substitution in only one S variant. Following further comparison with the closely related SARS-CoV S protein (5, 28), 18 of the 22 *N*-glycosites were identified as conserved in the SARS-CoV-2 S protein, indicating the importance of glycosylation of the virus. Four newly arisen *N*-glycosites (N17, N74, N149, and N657) are located in the SARS-CoV-2 S protein S1 subunit away from the RBD. Moreover, four previously confirmed

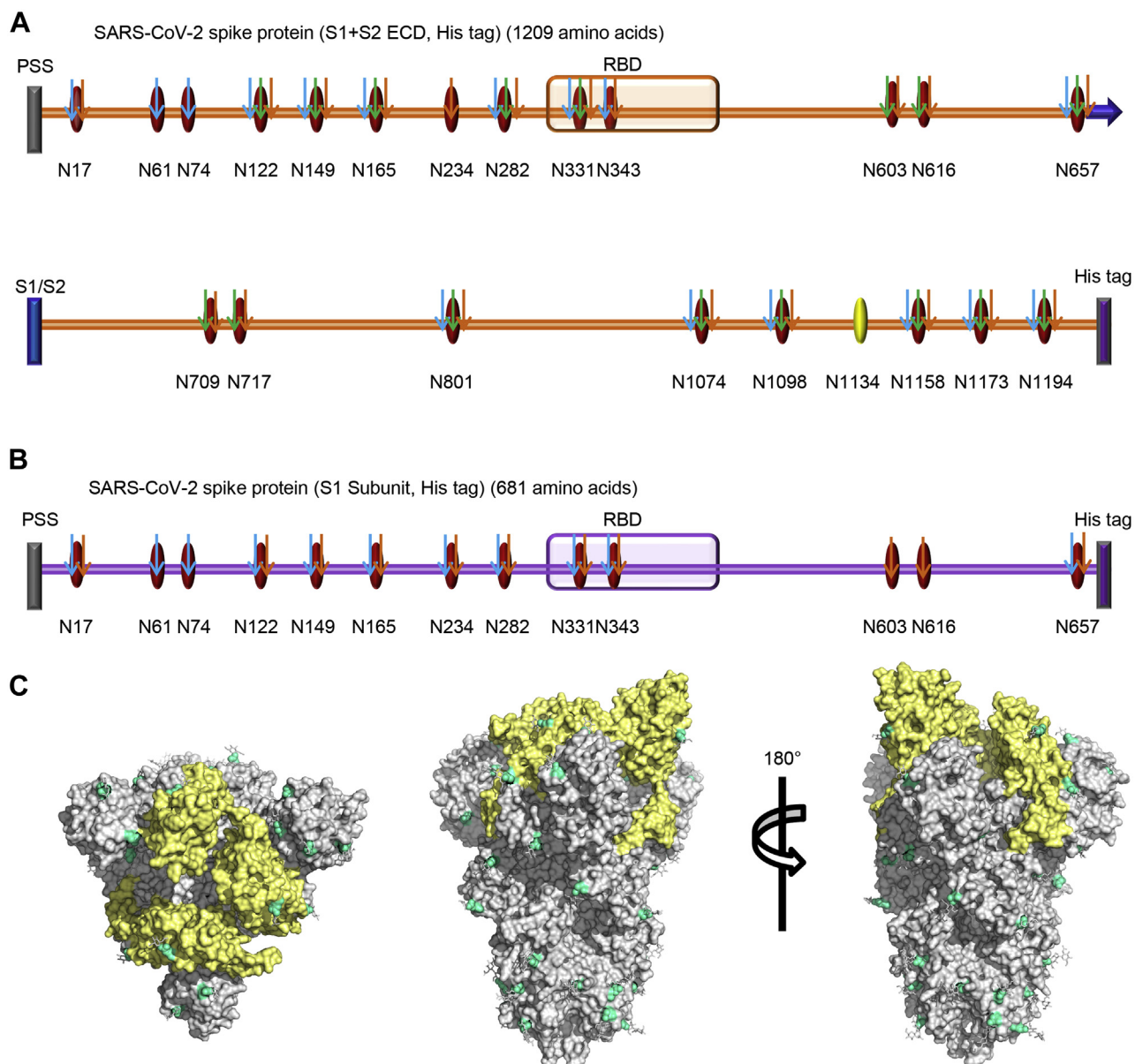


FIG. 2. N-glycosites characterization of SARS-CoV-2 S proteins. A and B, N-glycosites of the recombinant SARS-CoV-2 S protein or subunits expressed in insect cells (A) and human cells (B). PSS, putative signal sequence; RBD, receptor-binding domain; S1/S2, S1/S2 protease cleavage site; *Oval*, potential N-glycosite; *Yellow oval*, ambiguously assigned N-glycosite; *Red oval*, unambiguously assigned N-glycosite; *Blue arrow*, unambiguously assigned N-glycosite using trypsin digestion; *Green arrow*, unambiguously assigned N-glycosite using Glu-C digestion; *Yellow arrow*, unambiguously assigned N-glycosite using the combination of trypsin and Glu-C digestion. The unambiguously glycosite was determined by at least twice identification within each digestion list in supplemental Tables S1 and S2. C, N-glycosites were demonstrated in the three-dimensional structure of the SARS-CoV-2 S protein trimers (PDB code: 6VSB). RBDs, *yellow*; N-glycosites, *blue*.

N-glycosites (N29, N73, N109, and N357) in the SARS-CoV S protein were missing in SARS-CoV-2 S, one of which (N357) lies in the RBD (supplemental Fig. S6). Additionally, two N-glycosites (N1158 and N1173) identified in SARS-CoV-2 S in this study were not detected in SARS-CoV S (N1140 and N1155) in previous studies (5, 28). Our results suggest that the preferential change of the glycosylation landscape of the S1 subunit tends to change the distribution of glycan

shield, especially in the N terminal half of S1 (supplemental Fig. S6).

Site-specific N-glycan Occupancy of Recombinant SARS-CoV-2 S Proteins

Intact N-glycopeptide analysis can provide N-glycoproteomic information, including the composition and number of N-glycans decorating a specific N-glycopeptide or

N-glycosite. The potential *N*-glycopeptides in the S protein sequence are shown in [supplemental Fig. S1](#). Hundreds of nonredundant intact *N*-glycopeptides were identified from the recombinant S ectodomain ([supplemental Table S3](#)) and S1 subunit ([supplemental Table S4](#)). Representative and high-quality spectra of intact *N*-glycopeptides are shown in [supplemental Fig. S7](#). Following glycopeptide enrichment, the number of intact *N*-glycopeptides and *N*-glycans significantly increased ($p < 0.05$) ([Fig. 3, A and B](#)). Further, the intact *N*-glycopeptides were further quantified, and the top five *N*-glycan compositions on each site are shown in [supplemental Figs. S8 and S9](#), and [supplemental Table S9](#). Although not all identified intact *N*-glycopeptides can be quantified, those highly *N*-glycosylated sites and attached *N*-glycans can be highlighted.

S protein expressed in insect cells had smaller and fewer complex *N*-glycans attached to intact *N*-glycopeptides than S1 subunit produced in human cells. Both recombinant products contained the common *N*-acetylglucosamine (HexNAc) as a canonical *N*-glycan characteristic ([Fig. 3, C and D](#)). S protein expressed in insect cells was decorated with 38 *N*-glycans, with the majority preferentially containing oligomannose (Hex) and fucose (Fuc) ([Fig. 3C and supplemental Table S3](#)). By contrast, the S1 subunit expressed in human cells was attached to up to 157 *N*-glycans, mainly containing extra *N*-acetylglucosamine (HexNAc) and galactose (Hex), variably terminating with sialic acid (NeuAc) ([Fig. 3D and supplemental Table S4](#)). Returning to the glycosite level, most of the *N*-glycosites in the S protein were modified with 17 to 35 types of *N*-glycans and classified into a high proportion of high-mannose *N*-glycans (~65%) and a lower proportion (~23%) of hybrid *N*-glycans. Almost all *N*-glycosites contained no more than 10% of complex *N*-glycans ([Fig. 3E](#)). For the S1 subunit expressed in human cells, the occupancy of *N*-glycans on each *N*-glycosite was quite nonuniform. Surprisingly, six *N*-glycosites (N122, N165, N282, N331, N343, and N657) were decorated with markedly heterogeneous *N*-glycans of up to 139 types. The average occupancies of all glycosites presented as an overwhelming proportion (~75%) of complex *N*-glycans and a small proportion of hybrid (~13%) or high-mannose (~12%) *N*-glycans ([Fig. 3F](#)).

The glycan occupancy on two *N*-glycosites (N331 and N343) of RBD was identified ([Fig. 3, E and F](#)). The high occupancy of RBD glycosites by various *N*-glycan compositions implies that *N*-glycosylation might be associated with the recognition of RBD to ACE2 receptor, since the interaction between RBD and ACE2 mainly depends on polar residue interactions (14). Our results suggest that S proteins expressed in different cells display distinct *N*-glycosylation patterns. In particular, the glycosylation of the S protein in human cells exhibits remarkable heterogeneity on *N*-glycosites. However, the *N*-glycan types on each glycosite are primarily determined by the host cells rather than the location of different glycosites ([Fig. 3, E and F, supplemental Fig. S10](#)).

Site-specific *N*-glycan Occupancy of Recombinant SARS-CoV-2 S Protein RBD

To confirm site-specific *N*-glycan occupancy and exploit the potential impact on *N*-glycosylation by different protein sizes, recombinant RBDs (residues 319–541) from both human and insect cells were further analyzed ([supplemental Table S6](#)). The representative glycan compositions and deduced structures are shown on each site ([Fig. 4A](#)). Intriguingly, the number of glycan compositions and their types on each glycosite ([Fig. 4, B and C](#)) are very close to those found in the S ectodomain and S1 subunit ([Fig. 3, E and F](#)). The human cell-produced RBDs displayed more *N*-glycan compositions and complex glycan types, compared with insect cell-expressed proteins ([Fig. 4, B and C](#)). Moreover, more than 80% of glycan compositions are identical at each site among different lengths of insect cell-expressed proteins ([Fig. 4D](#)). Similarly, over 75% of the glycan compositions at each site were found to be shared by human cell-produced products ([Fig. 4E](#)). The *N*-glycosylation of RBDs was verified by SDS-PAGE ([supplemental Fig. S11](#)). These results suggest that the *N*-glycan compositions are conserved among different sizes of RBD proteins. Thus, our data reveal a regular *N*-glycan occupancy on S protein, despite the heterogeneity in *N*-glycan compositions. Intriguingly, the *N*-glycan types on S protein subunits are predominantly determined by host cells, regardless of the location of glycosites.

DISCUSSION

The global outbreak and rapid spread of COVID-19 caused by SARS-CoV-2 urgently call for specific prevention and intervention measures (29). The development of preventative vaccines and neutralizing antibodies remains a chief goal in the efforts to control viral spread and stockpile candidates for future use. However, this work greatly relies on the understanding of the antigen structure and state of glycosylation for the rational determination of accessible epitopes. The S protein is posited to be the main or even the only antigen on viral surfaces for priming the immune system to produce an effective response (20, 25). Previous studies have revealed the structural information of the SARS-CoV-2 S protein and found the coverage of *N*-glycans (4, 13, 17). In this study, we profiled and compared the site-specific *N*-glycosylation of the recombinant SARS-CoV-2 S protein expressed in insect and human cells. All 22 potential *N*-glycosites on the insect cell-expressed S protein were identified experimentally. Furthermore, 13 out of the 22 *N*-glycosites were verified on the RBD-containing S1 subunit expressed in human cells. The *N*-glycan compositions and types on S protein subunits were revealed among different host cells. Our data provide a large-scale *N*-glycosylation information of the S protein and present a promising prospect for developing a vaccine and therapeutic antibodies.

N-glycosylation is a common feature of the viral envelope proteins, including those of HIV-1, Lassa virus, hepatitis C

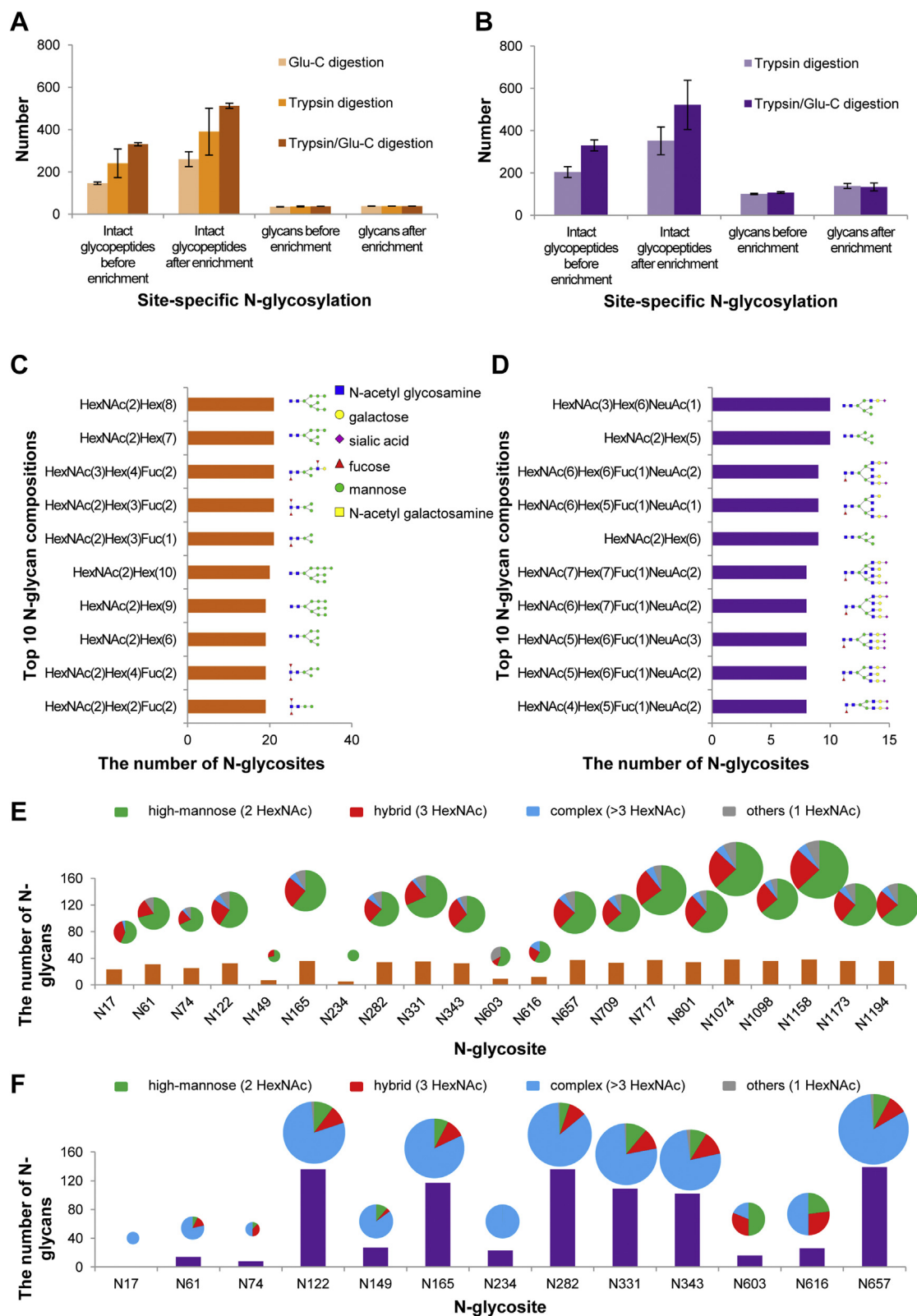


FIG. 3. Site-specific N-glycosylation of recombinant SARS-CoV-2 S proteins. A and B, the number of intact N-glycopeptides and N-glycans in recombinant SARS-CoV-2 S proteins expressed in insect cells (A) or human cells (B). C and D, the numbers of the N-glycosites containing one representative N-glycan and its deduced structure from the recombinant SARS-CoV-2 S protein or subunit expressed in insect cells (C) and human cells (D). E and F, different types and numbers of N-glycan compositions on each N-glycosite of the recombinant SARS-CoV-2 S protein or subunit expressed in insect cells (E) or human cells (F). The size of the pies represents the number of N-glycan compositions.

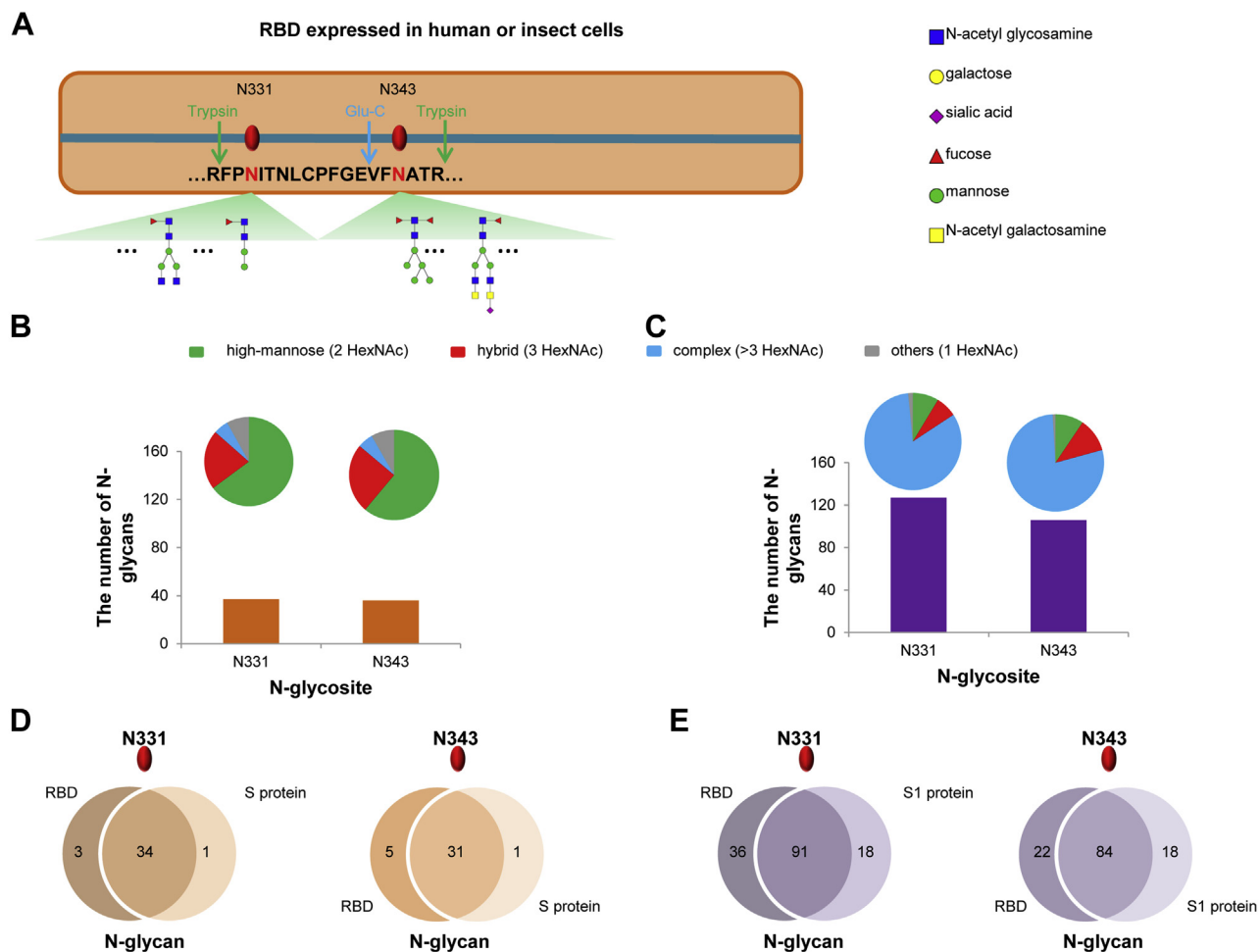


FIG. 4. Site-specific N-glycosylation profile of RBDs expressed in human and insect cells. A, the glycosites and deduced representative N-glycans on N331 and N343 of RBD. B and C, different types and numbers of N-glycans on N331 and N343 of RBD expressed in insect cells (B) or human cells (C). The size of the pies represents the number of N-glycan compositions. D, comparison of N-glycans on N331 and N343 between RBD and S ectodomain expressed in insect cells. E, comparison of N-glycans on N331 and N343 between RBD and S1 subunit expressed in human cells.

virus, Epstein–Barr virus, and influenza A virus. Glycosylation promotes proper glycoprotein folding; however, the glycans obstruct receptor binding and proteolytic processing during antigen presentation (20, 22, 30). Characterization of the landscape of N-glycans on the SARS-CoV-2 S protein is crucial for promoting immunogen design and prevention of the potential viral evasion of intervention measures (5, 20, 31). Precise characterization of intact N-glycopeptides can reveal the occupancy of each glycosite by different glycoforms (32, 33). In this study, all 22 N-glycosites of the S protein were identified (Fig. 2, A and B). By comparison, the alteration of N-glycosites among SARS-CoV-2 and SARS-CoV S proteins focused on glycosites in the S1 subunit (supplemental Fig. S6). The N-glycosites located in the S2 subunit are completely conserved, and seven out of nine sites, along with glycans, have been disclosed by previous studies (4, 5). Moreover, we found that N-glycosites in the S protein were

highly preserved among 145 SARS-CoV-2 S protein variants (supplemental Table S5), which is advantageous for circumventing potential viral immune evasion from the vaccines and neutralizing antibodies currently being developed.

Glycosylation of proteins is intricately processed by various enzymes coordinated in the endoplasmic reticulum and Golgi apparatus. The glycan composition and structure decoration at specific sites occurred in a nontemplated manner governed by host cells, thus frequently resulting in a heterogeneous glycan occupancy on each glycosite (34). Glycosylation processing in insect and human cells can yield a common intermediate N-glycan, which will further elongate in human cells but is only trimmed in insect cells to form end products (26). Consequently, the S protein expressed in human cells displayed a larger size and a much higher proportion of complex N-glycans than that expressed in insect cells, owing to the additional elongation of the glycan backbone with multiple

oligosaccharides (Fig. 3, C–F, supplemental Figs. S8 and S9, supplemental Table S7). Our findings are in line with those of previous studies that demonstrated the predominant complex of *N*-glycans attached to MERS-CoV and SARS-CoV proteins (4, 5, 35). Moreover, two recent studies by Watanabe *et al.* and Shajahan *et al.* have revealed that the complex *N*-glycans dominate the glycosites on human cell-expressed SARS-CoV-2 S protein (17, 18). In contrast, S protein expression in insect cells led to a high ratio of high-mannose *N*-glycans (Fig. 3, C and E, supplemental Figs. S8 and S9, supplemental Table S7), which has also been found in the insect cell-produced HCoV-NL63 S protein, despite expression in a different insect cell, the *Drosophila* S2 cell (20).

In accordance with our study, Watanabe *et al.* (2020) also identified 22 *N*-glycosites and 110 *N*-glycans in human-cell-produced S protein variants. However, eight sites (61, 122, 234, 603, 709, 717, 801, and 1074) out of 22 were mainly modified by the oligomannose-type *N*-glycan. Moreover, other sites were occupied by the complex-type *N*-glycan. In this study, we found that the human cell-expressed native S1 subunit (16–685) contained 157 *N*-glycans, primarily of the complex-type. Only one glycosite (603) was of the high-mannose-type. Our study is consistent with the previous study on SARS-CoV S protein, in which Walls *et al.* (5) revealed that the complex *N*-glycans rather than oligomannose-types were linked to the preserved glycosites (61, 122, 603, 709, 801, and 1074) between SARS-CoV S and SARS-CoV-2. In contrast, we found that the insect cell-expressed S protein contained 38 *N*-glycans, comprising entirely the high-mannose-type. The differences between the two studies may be attributed to the different protein sequence used, sample processing method (*i.e.*, glycopeptide enrichment), LC-MS/MS analysis, and data analysis method. Similarly, Shajahan *et al.* (18) also identified 17 *N*-glycosites (61, 74, 122, 149, 165, 234, 282, 331, 343, 616, 657, 709, 717, 801, 1074, 1098, 1194) by using human cell-expressed S1 and S2 subunits. These glycosites were mainly modified by both high mannose- and complex-type *N*-glycans. Moreover, two O-glycosites (323 and 325) were identified and predominantly modified by Core-1 and Core-2 O-glycans. A detailed comparison between these studies on S protein *N*-glycosylation is shown in supplemental Table S8. Intriguingly, during the review of our manuscript, the native glycans were identified by analyzing the virus sample using mass spectrometer. Almost all 22 glycosites were modified with the complex-type *N*-glycans, which is highly consistent with the findings in our study (36). Furthermore, in another study, we have revealed that the SARS-CoV-2 S protein is also a mucin-type glycoprotein (37).

Despite the occupancy by the high-abundant complex type *N*-glycans, a low ratio (~12%) of high-mannose glycans ubiquitously exist on human cell-expressed S protein (Fig. 3F and supplemental Fig. S9). The HIV envelope glycoprotein gp120 is heavily decorated with the immature intermediate, high-mannose glycans. The high-density glycans surrounding

HIV glycoproteins limit the accessibility of glycan biosynthetic processing enzymes, terminating the synthesis of more complex end products (22, 31). By contrast, the high ratio of complex *N*-glycans in the SARS-CoV-2 S protein were successfully processed on most glycosites by the enzymes, without extensive obstruction by the ongoing synthesis of glycan shields (Figs. 3F and 4C). Therefore, we posit that the glycan coverage on the SARS-CoV-2 S protein could leave relatively accessible antigens and epitopes, although the complex *N*-glycans might mask some surface immunogens. These features may provide a promising landscape on the SARS-CoV-2 S protein for immune recognition. This potential is bolstered by the findings that the convalescent sera from COVID-19 patients contain antibodies against the SARS-CoV-2 S protein (7, 10). A previous study on SARS-CoV has revealed that the oligomannose on the S protein can be recognized by mannose-binding lectin (MBL) and may interfere with viral entry into host cells by inhibition of S protein function (38). Besides the direct neutralization effect, MBL, as a serum complement protein, can initiate the complement cascade. Complement hyperactivation in lung tissues of COVID-19 patients has been revealed by a recent preprint study (39). However, it remains unknown whether MBL can bind to oligomannose on SARS-CoV-2 S protein to impact the viral spread or initiate complement activation in patients. Glycans also play crucial and multifaceted roles in B-cell and T-cell differentiation *via* cell-surface or secreted proteins, including selectins, galectins, and siglecs, which can further connect SARS-CoV-2 to immune response and immune regulation (34). In particular, the complex *N*-glycans are ligands for galectins, which are able to engage different glycoproteins to regulate immune cell infiltration and activation upon virus infection (40, 41). These mechanisms underlying SARS-CoV-2 infection and spread are worth further clarification based on a detailed analysis of clinical characteristics in humoral and cellular immunity.

The remarkable heterogeneity of *N*-glycosylation in the S protein subunit expressed in human cells was revealed in our study (Fig. 3F and supplemental Fig. S9) and other studies (17, 18). By contrast, the *N*-glycosylation of the S protein subunit in insect cells showed less heterogeneity and complexity than that of human cell-derived proteins (Fig. 3E). Moreover, the site-specific glycan occupancy tended to be identical in the same host cell, regardless of protein length (Fig. 4, D and E). These results indicate that the *N*-glycan compositions and types on S protein largely attribute to different host cells with the differential processing pathways of glycosylation. We can expect that the native *N*-glycosylation profile of the SARS-CoV-2 S protein in humans tends to be consistent with that of the recombinant protein expressed in human cells, unless the virus buds off early in the glycosylation processing pathway and produces immature glycans (26, 34, 42). Intriguingly, the immature *N*-glycans such as high mannose are regarded as non-self-like glycans when occurring in dense

patches (42, 43). Therefore, the insect cell-expressed recombinant antigens predominantly decorated with high mannose may be more immunogenic than those produced in human cells (44, 45). To prime a robust humoral immune reaction upon vaccination against SARS-CoV-2, insect cell-produced antigens could be promising candidates for vaccine development (46). Apart from amino acid epitopes, the glycopeptide can be presented by major histocompatibility complex (MHC) and recognized by a CD4+ T-cell population to help B cells produce antibodies against glycans. The glycoconjugate has been used to boost the immune response against infections (30, 31). The insect cell-produced glycopeptides as antigens might prime immune response against glycans, in cases that immature N-glycans occur on the native envelope proteins of SARS-CoV-2, which seems to have emerged during SARS-CoV replication (42). In contrast, the human cell-expressed S protein subunits as vaccines mimic the “self” glycans in humans, which could impair immune response to the antigens (47). However, the remaining accessible and nonglycosylated regions of S protein can serve as the antigens to produce vaccines in human cells (11, 48). The rational design of antigens to prime potent and broad immune responses against accessible epitopes on SARS-CoV-2 S protein is essential and promising.

The RBD-containing subunit is an ideal immunogen since antibodies against the receptor-binding motif within RBD could directly block the engagement of S protein to the receptor and inhibit viral infections of host cells. Vaccination with SARS-CoV-2 RBD has been demonstrated to induce protective immunity against SARS-CoV-2 (46). Meanwhile, the subunit vaccines are posited to minimize the potentially undesired immunopotentialization of the full-length S protein, which might induce severe acute injury in the lungs (49). Intriguingly, SARS-CoV-2 is missing one N-glycosite in RBD compared with SARS-CoV. The remaining two N-glycosites were outside of the motifs essential for direct interaction with the ACE2 receptor (14) (Fig. 2C). The glycan compositions of RBD are highly identical in the same host cell, regardless of the length of the RBD-containing proteins (Fig. 4, B–E). These features of the RBD, along with its highly exposed structure, endow more antigens and accessible epitopes for vaccine design and immune recognition. The RBD-containing proteins, especially the insect cell-expressed products, could become promising candidates for SARS-CoV-2 vaccine development. However, drug discovery related to glycosylation inhibition is supposed to be performed based on human cell-expressed products.

In this study, we decoded the site-specific profile of N-glycosylation on SARS-CoV-2 S proteins expressed in insect and human cells, revealing a regular N-glycan site occupancy on S protein, despite the heterogeneity of N-glycan compositions on each site. All glycosites were conserved among the 753 public SARS-CoV-2 genome sequences. In conclusion, our data indicate that differential N-glycan occupancies

among distinct host cells might help elucidate the infection mechanism and develop an effective vaccine and targeted drugs. Nevertheless, the implication of S protein site-specific N-glycosylation in immunogenicity, receptor binding, and viral infectivity should be investigated further.

DATA AVAILABILITY

The raw MS data have been deposited to the ProteomeXchange Consortium *via* the PRIDE partner repository with the data set identifier PXD018506. The annotated spectra for all identified glycans and glycopeptides from the Byonic search results have been deposited in the online repository Figshare (https://figshare.com/articles/dataset/SARS-CoV-2_rar/13023095).

Supplemental data—This article contains [supplemental data](#).

Acknowledgments—We would like to express our special thanks to the National Center for Protein Sciences Beijing, Professor Catherine E. Costello in Boston University School of Medicine, and the COVID-19 Mass Spectrometry Coalition for sharing knowledge on glycopeptide analysis. Thanks to Beijing Sino Biological Inc for providing high-quality recombinant proteins for this project.

Funding and additional information—This work was funded by grants from the National Natural Science Foundation of China (grant number 31901038), the 1.3.5 Project for Disciplines of Excellence, West China Hospital, Sichuan University (ZYGD18014, CJQ), the Chengdu Science and Technology Department Foundation (grant number 2020-YF05-00240-SN), and the Science and Technology Department of Sichuan Province (2020YFH0029).

Author contributions—H. Y., M. G., D. D., X. L., Y. C., and J. C. directed and designed the research; Y. Z. and W. Z. directed and performed the analyses of mass spectrometry data; Y. Z., H. Y., and S. W. adapted algorithms and software for data analysis; Y. Z. and T. S. coordinated the acquisition, distribution, and quality evaluation of samples; Y. Z. and H. Y. wrote the article.

Conflict of interest—The authors declare that they have no competing interests.

Abbreviations—The abbreviations used are: ACE2, angiotensin-converting enzyme II; ACN, Acetonitrile; Cryo-EM, cryoelectron microscopy; DTT, Dithiothreitol; E, envelope protein; FA, formic acid; GPSMs, glycopeptide spectrum matches; HCD, higher-energy collisional dissociation; HCoV-NL63, human coronavirus NL63; IAA, iodoacetamide; M, membrane protein; MBL, mannose-binding lectin; MERS-CoV, Middle East respiratory syndrome coronavirus; MS, mass spectrometry; N, nucleocapsid protein; PSMs, peptide

spectrum matches; RBD, receptor-binding domain; S, spike protein; SARS-CoV-2, severe acute respiratory syndrome coronavirus; SCE, stepped collision energy; TFA, trifluoroacetic acid; Zic-HILIC, zwitterionic hydrophilic interaction liquid chromatography.

Published, MCPRO Papers in Press, February 11, 2021, <https://doi.org/10.1074/mcp.RA120.002295>

REFERENCES

- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. D., *et al.* (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273
- Wolfel, R., Corman, V. M., Guggemos, W., Seilmaier, M., Zange, S., Muller, M. A., Niemeyer, D., Jones, T. C., Vollmar, P., Rothe, C., Hoelscher, M., Bleicker, T., Brunink, S., Schneider, J., Ehmann, R., *et al.* (2020) Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465–469
- Paules, C. I., Marston, H. D., and Fauci, A. S. (2020) Coronavirus infections—more than just the common cold. *JAMA* **323**, 707–708
- Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020) Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281–292.e286
- Walls, A. C., Xiong, X., Park, Y. J., Tortorici, M. A., Snijder, J., Quispe, J., Cameron, E., Gopal, R., Dai, M., Lanzavecchia, A., Zamboni, M., Rey, F. A., Corti, D., and Veesler, D. (2019) Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell* **176**, 1026–1039.e1015
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Muller, M. A., Drosten, C., and Pohlmann, S. (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.e278
- Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, R., Chen, T., Hu, J., Xiang, Z., Mu, Z., Chen, X., Chen, J., Hu, K., *et al.* (2020) Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* **11**, 1620
- Casadevall, A., and Pirofski, L. A. (2020) The convalescent sera option for containing COVID-19. *J. Clin. Invest.* **130**, 1545–1548
- Cao, Y., Su, B., Guo, X., Sun, W., Deng, Y., Bao, L., Zhu, Q., Zhang, X., Zheng, Y., Geng, C., Chai, X., He, R., Li, X., Lv, Q., Zhu, H., *et al.* (2020) Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* **182**, 73–84.e16
- Chi, X., Yan, R., Zhang, J., Zhang, G., Zhang, Y., Hao, M., Zhang, Z., Fan, P., Dong, Y., Yang, Y., Chen, Z., Guo, Y., Zhang, J., Li, Y., Song, X., *et al.* (2020) A neutralizing human antibody binds to the N-terminal domain of the spike protein of SARS-CoV-2. *Science* **369**, 650–655
- Zhu, F. C., Li, Y. H., Guan, X. H., Hou, L. H., Wang, W. J., Li, J. X., Wu, S. P., Wang, B. S., Wang, Z., Wang, L., Jia, S. Y., Jiang, H. D., Wang, L., Jiang, T., Hu, Y., *et al.* (2020) Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored COVID-19 vaccine: A dose-escalation, open-label, non-randomised, first-in-human trial. *Lancet* **395**, 1845–1854
- Tai, W., He, L., Zhang, X., Pu, J., Voronin, D., Jiang, S., Zhou, Y., and Du, L. (2020) Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: Implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell. Mol. Immunol.* **17**, 613–620
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., Graham, B. S., and McLellan, J. S. (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263
- Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020) Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448
- Li, F. (2016) Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* **3**, 237–261
- Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H., and Farzan, M. (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454
- 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
- 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
- Shajahan, A., Archer-hartmann, S. A., Supekar, N. T., Gleinich, A. S., Heiss, C., and Azadi, P. (2020) Comprehensive characterization of N-and O-glycosylation of SARS-CoV-2 human receptor angiotensin converting enzyme 2. *bioRxiv*. <https://doi.org/10.1101/2020.05.01.071688>
- Walls, A. C., Tortorici, M. A., Frenz, B., Snijder, J., Li, W., Rey, F. A., DiMaio, F., Bosch, B. J., and Veesler, D. (2016) Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* **23**, 899–905
- Clerc, F., Reiding, K. R., Jansen, B. C., Kammeijer, G. S., Bondt, A., and Wuhrer, M. (2016) Human plasma protein N-glycosylation. *Glycoconj. J.* **33**, 309–343
- Chang, D., and Zaia, J. (2019) Why glycosylation matters in building a better flu vaccine. *Mol. Cell. Proteomics* **18**, 2348–2358
- Zhang, Y., Zhao, W., Zhao, Y., Mao, Y., Su, T., Zhong, Y., Wang, S., Zhai, R., Cheng, J., Fang, X., Zhu, J., and Yang, H. (2020) Comparative glycoproteomic profiling of human body fluid between healthy controls and patients with papillary thyroid carcinoma. *J. Proteome Res.* **19**, 2539–2552
- Zhang, Y., Lin, T., Zhao, Y., Mao, Y., Tao, Y., Huang, Y., Wang, S., Hu, L., Cheng, J., and Yang, H. (2020) Characterization of N-linked intact glycopeptide signatures of plasma IgGs from patients with prostate carcinoma and benign prostatic hyperplasia for diagnosis pre-stratification. *Analyst* **145**, 5353–5362
- Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J., and Jiang, S. (2009) The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nat. Rev. Microbiol.* **7**, 226–236
- Kost, T. A., Condreay, J. P., and Jarvis, D. L. (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat. Biotechnol.* **23**, 567–575
- Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., *et al.* (2020) A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269
- Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., Lu, G., Wu, Y., Yan, J., Shi, Y., Zhang, X., and Gao, G. F. (2017) Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* **8**, 15092
- Wang, M., Zhou, Y., Zong, Z., Liang, Z., Cao, Y., Tang, H., Song, B., Huang, Z., Kang, Y., Feng, P., Ying, B., and Li, W. (2020) A precision medicine approach to managing 2019 novel coronavirus pneumonia. *Precis. Clin. Med.* **3**, 14–21
- Wolfert, M. A., and Boons, G. J. (2013) Adaptive immune activation: Glycosylation does matter. *Nat. Chem. Biol.* **9**, 776–784
- Seabright, G. E., Doores, K. J., Burton, D. R., and Crispin, M. (2019) Protein and glycan mimicry in HIV vaccine design. *J. Mol. Biol.* **431**, 2223–2247
- Riley, N. M., Hebert, A. S., Westphall, M. S., and Coon, J. J. (2019) Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nat. Commun.* **10**, 1311
- Sun, S., Hu, Y., Jia, L., Eshghi, S. T., Liu, Y., Shah, P., and Zhang, H. (2018) Site-specific profiling of serum glycoproteins using N-linked glycan and glycosite analysis revealing atypical N-glycosylation sites on albumin and alpha-1B-glycoprotein. *Anal. Chem.* **90**, 6292–6299
- Reilly, C., Stewart, T. J., Renfrow, M. B., and Novak, J. (2019) Glycosylation in health and disease. *Nat. Rev. Nephrol.* **15**, 346–366
- Ritchie, G., Harvey, D. J., Feldmann, F., Stroehrer, U., Feldmann, H., Royle, L., Dwek, R. A., and Rudd, P. M. (2010) Identification of N-linked carbohydrates from severe acute respiratory syndrome (SARS) spike glycoprotein. *Virology* **399**, 257–269
- Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Weng, T., Zhang, Z., Wu, Z., Cheng, L., Shi, D., Lu, X., Lei, J., Crispin, M., *et al.* (2020) Molecular architecture of the SARS-CoV-2 virus. *Cell* **183**, 730–738.e13
- Zhang, Y., Zhao, W., Mao, Y., Chen, Y., Zhu, J., Hu, L., Gong, M., Cheng, J., and Yang, H. (2020) Mucin-type O-glycosylation landscapes of SARS-CoV-2 spike proteins. *bioRxiv*. <https://doi.org/10.1101/2020.07.29.227785>

38. Zhou, Y., Lu, K., Pfeifferle, S., Bertram, S., Glowacka, I., Drosten, C., Pohlmann, S., and Simmons, G. (2010) A single asparagine-linked glycosylation site of the severe acute respiratory syndrome coronavirus spike glycoprotein facilitates inhibition by mannose-binding lectin through multiple mechanisms. *J. Virol.* **84**, 8753–8764
39. Gao, T., Hu, M., Zhang, X., Li, H., Zhu, L., Liu, H., Dong, Q., Zhang, Z., Wang, Z., Hu, Y., Fu, Y., Jin, Y., Li, K., Zhao, S., Xiao, Y., *et al.* (2020) Highly pathogenic coronavirus N protein aggravates lung injury by MASP-2-mediated complement over-activation. *medRxiv*. <https://doi.org/10.1101/2020.03.29.20041962>
40. Robinson, B. S., Arthur, C. M., Evavold, B., Roback, E., Kamili, N. A., Stowell, C. S., Vallecillo-Zuniga, M. L., Van Ry, P. M., Dias-Baruffi, M., Cummings, R. D., and Stowell, S. R. (2019) The sweet-side of leukocytes: Galectins as master regulators of neutrophil function. *Front. Immunol.* **10**, 1762
41. Wang, W. H., Lin, C. Y., Chang, M. R., Urbina, A. N., Assavalapsakul, W., Thitithyanont, A., Chen, Y. H., Liu, F. T., and Wang, S. F. (2019) The role of galectins in virus infection - a systemic literature review. *J. Microbiol. Immunol. Infect.* **53**, 925–935
42. Watanabe, Y., Bowden, T. A., Wilson, I. A., and Crispin, M. (2019) Exploitation of glycosylation in enveloped virus pathobiology. *Biochim. Biophys. Acta Gen. Subj.* **1863**, 1480–1497
43. Pritchard, L. K., Spencer, D. I., Royle, L., Bonomelli, C., Seabright, G. E., Behrens, A. J., Kulp, D. W., Menis, S., Krumm, S. A., Dunlop, D. C., Crispin, D. J., Bowden, T. A., Scanlan, C. N., Ward, A. B., Schief, W. R., *et al.* (2015) Glycan clustering stabilizes the mannose patch of HIV-1 and preserves vulnerability to broadly neutralizing antibodies. *Nat. Commun.* **6**, 7479
44. Li, D., von Schaewen, M., Wang, X., Tao, W., Zhang, Y., Li, L., Heller, B., Hrebikova, G., Deng, Q., Ploss, A., Zhong, J., and Huang, Z. (2016) Altered glycosylation patterns increase immunogenicity of a subunit hepatitis C virus vaccine, inducing neutralizing antibodies which confer protection in mice. *J. Virol.* **90**, 10486–10498
45. Urbanowicz, R. A., Wang, R., Schiel, J. E., Keck, Z. Y., Kerzic, M. C., Lau, P., Rangarajan, S., Garagusi, K. J., Tan, L., Guest, J. D., Ball, J. K., Pierce, B. G., Mariuzza, R. A., Fong, S. K. H., and Fuerst, T. R. (2019) Antigenicity and immunogenicity of differentially glycosylated hepatitis C virus E2 envelope proteins expressed in mammalian and insect cells. *J. Virol.* **93**, e01403-18
46. Yang, J., Wang, W., Chen, Z., Lu, S., Yang, F., Bi, Z., Bao, L., Mo, F., Li, X., Huang, Y., Hong, W., Yang, Y., Zhao, Y., Ye, F., Lin, S., *et al.* (2020) A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity. *Nature* **586**, 572–577
47. Ozdilek, A., Paschall, A. V., Dookwah, M., Tiemeyer, M., and Avci, F. Y. (2020) Host protein glycosylation in nucleic acid vaccines as a potential hurdle in vaccine design for nonviral pathogens. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 1280–1282
48. Jackson, L. A., Anderson, E. J., Roupael, N. G., Roberts, P. C., Makhene, M., Coler, R. N., McCullough, M. P., Chappell, J. D., Denison, M. R., Stevens, L. J., Pruijssers, A. J., McDermott, A., Flach, B., Doria-Rose, N. A., Corbett, K. S., *et al.* (2020) An mRNA vaccine against SARS-CoV-2 - preliminary report. *N. Engl. J. Med.* **383**, 1920–1931
49. Liu, L., Wei, Q., Lin, Q., Fang, J., Wang, H., Kwok, H., Tang, H., Nishiura, K., Peng, J., Tan, Z., Wu, T., Cheung, K. W., Chan, K. H., Alvarez, X., Qin, C., *et al.* (2019) Anti-spike IgG causes severe acute lung injury by skewing macrophage responses during acute SARS-CoV infection. *JCI Insight* **4**, e123158