UC San Diego
UC San Diego Previously Published Works

Title

Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor.

Permalink

Journal

Authors

Zhao, Peng
Praissman, Jeremy
Grant, Oliver
et al.

Publication Date

DOI

Peer reviewed

Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor

Graphical Abstract
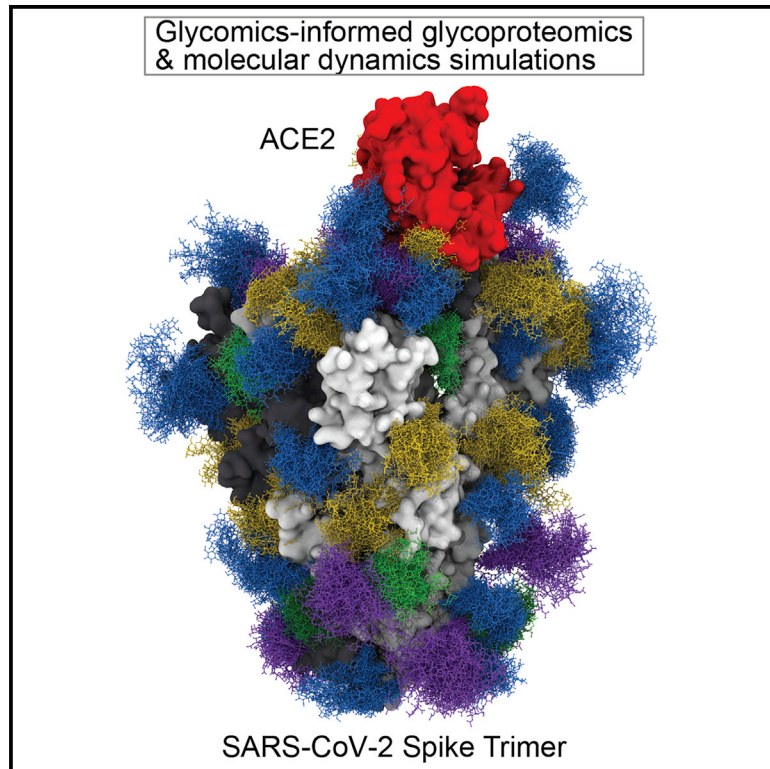
Authors

Peng Zhao, Jeremy L. Praissman, Oliver C. Grant, ..., Bing Chen, Robert J. Woods, Lance Wells

Correspondence

rwoods@ccrc.uga.edu (R.J.W.),
lwells@ccrc.uga.edu (L.W.)

In Brief

Combining glycomics-informed glycoproteomics and bioinformatic analyses of variants with molecular dynamics simulations, Zhao et al. detail a role for glycan-protein and glycan-glycan interactions in the SARS-CoV-2 viral Spike protein-ACE2 human receptor complex.

Highlights

- Site-specific N-linked microheterogeneity is defined at 22 sites of SARS-CoV-2 Spike

- Six sites of N-linked microheterogeneity of human ACE2 receptor are described

- Molecular dynamics simulations of Spike and ACE2 show essential roles for glycosylation

- We uncover roles for variants in protein-glycan and glycan-glycan interactions

CellPress

Article

Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor

Peng Zhao,[1,7] Jeremy L. Praissman,[1,7] Oliver C. Grant,[1,7] Yongfei Cai,[2] Tianshu Xiao,[2] Katelyn E. Rosenbalm,[1] Kazuhiro Aoki,[1] Benjamin P. Kellman,[3] Robert Bridger,[1] Dan H. Barouch,[4] Melinda A. Brindley,[5] Nathan E. Lewis,[3,6] Michael Tiemeyer,[1] Bing Chen,[2] Robert J. Woods,[1,*] and Lance Wells[1,8,*]

[1]Complex Carbohydrate Research Center, Department of Biochemistry and Molecular Biology, and Department of Chemistry, University of Georgia, Athens, GA 30602, USA
[2]Division of Molecular Medicine, Children's Hospital and Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA
[3]Departments of Pediatrics and Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA
[4]Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA
[5]Department of Infectious Diseases, Department of Population Health, Center for Vaccines and Immunology, College of Veterinary Medicine, University of Georgia, Athens, GA 30602, USA
[6]Novo Nordisk Foundation Center for Biosustainability at UC San Diego, La Jolla, CA 92093, USA
[7]These authors contributed equally
[8]Lead Contact
*Correspondence: rwoods@ccrc.uga.edu (R.J.W.), lwells@ccrc.uga.edu (L.W.)
https://doi.org/10.1016/j.chom.2020.08.004

SUMMARY

The SARS-CoV-2 betacoronavirus uses its highly glycosylated trimeric Spike protein to bind to the cell surface receptor angiotensin converting enzyme 2 (ACE2) glycoprotein and facilitate host cell entry. We utilized glycomics-informed glycoproteomics to characterize site-specific microheterogeneity of glycosylation for a recombinant trimer Spike mimetic immunogen and for a soluble version of human ACE2. We combined this information with bioinformatics analyses of natural variants and with existing 3D structures of both glycoproteins to generate molecular dynamics simulations of each glycoprotein both alone and interacting with one another. Our results highlight roles for glycans in sterically masking polypeptide epitopes and directly modulating Spike-ACE2 interactions. Furthermore, our results illustrate the impact of viral evolution and divergence on Spike glycosylation, as well as the influence of natural variants on ACE2 receptor glycosylation. Taken together, these data can facilitate immunogen design to achieve antibody neutralization and inform therapeutic strategies to inhibit viral infection.

INTRODUCTION

The SARS-CoV-2 coronavirus, a positive-sense single-stranded RNA virus, is responsible for the severe acute respiratory syndrome referred to as COVID-19 that was first reported in China in December 2019 (Zhou et al., 2020). In approximately six months, this betacoronavirus has spread globally, with more than 14 million people testing positive worldwide resulting in greater than 600,000 deaths as of July 20, 2020 (https://coronavirus.jhu.edu/map.html). The SARS-CoV-2 coronavirus is highly similar (nearly 80% identical at the genomic level) to SARS-CoV-1, which was responsible for the severe acute respiratory syndrome outbreak that began in 2002 (Lu et al., 2020; Zhong et al., 2003). Furthermore, human SARS-CoV-2 at the whole-genome level is >95% identical to a bat coronavirus (RaTG13), the natural reservoir host for multiple coronaviruses (Xia, 2020; Zhang et al., 2020; Zhou et al., 2020). Given the rapid appearance and spread of this virus, there is no current validated vaccine or SARS-CoV-2-specific targeting therapy that is clinically approved, although statins, heparin, and steroids look

promising for lowering fatality rates, and antivirals likely reduce the duration of symptomatic disease presentation (Alijotas-Reig et al., 2020; Beigel et al., 2020; Beun et al., 2020; Dashti-Khavidaki and Khalili, 2020; Fedson et al., 2020; Shi et al., 2020; Tang et al., 2020).

SARS-CoV-2, like SARS-CoV-1, utilizes the host angiotensin-converting enzyme 2 (ACE2) for binding and entry into host cells (Hoffmann et al., 2020; Li et al., 2003). Like many viruses, SARS-CoV-2 utilizes a Spike glycoprotein trimer for recognition and binding to the host cell entry receptor and for membrane fusion (Watanabe et al., 2019). Given the importance of viral Spike proteins for targeting and entry into host cells along with their location on the viral surface, Spike proteins are often used as immunogens for vaccines to generate neutralizing antibodies and frequently targeted for inhibition by small molecules that might block host receptor binding and/or membrane fusion (Li, 2016; Watanabe et al., 2019). In similar fashion, wild-type or catalytically impaired ACE2 has also been investigated as a potential therapeutic biologic that might interfere with the infection cycle of ACE2-targeting coronaviruses (Lei et al., 2020; Monteil

et al., 2020). Thus, a detailed understanding of SARS-CoV-2 Spike binding to ACE2 is critical for elucidating mechanisms of viral binding and entry, as well as for undertaking the rational design of effective therapeutics.

The SARS-CoV-2 Spike glycoprotein consists of two subunits, a receptor binding subunit (S1) and a membrane fusion subunit (S2) (Lu et al., 2020; Zhou et al., 2020). The Spike glycoprotein assembles into stable homotrimers that together possess 66 canonical sequons for N-linked glycosylation (N-X-S/T, where X is any amino acid except P) as well as a number of potential O-linked glycosylation sites (Watanabe et al., 2020a; Watanabe et al., 2020b). Interestingly, coronaviruses virions bud into the lumen of the endoplasmic reticulum-Golgi intermediate compartment, ERGIC, raising unanswered questions regarding the precise mechanisms by which viral surface glycoproteins are processed as they traverse the secretory pathway (Stertz et al., 2007; Ujike and Taguchi, 2015). Although this and similar studies (Shajahan et al., 2020; Watanabe et al., 2020a) analyze recombinant proteins, a previous study on SARS-CoV-1 suggested that glycosylation of the Spike can be impacted by this intracellular budding, and this remains to be investigated in SARS-CoV-2 (Ritchie et al., 2010). Nonetheless, it has been proposed that this virus, and others, acquires a glycan coat sufficient and similar enough to endogenous host protein glycosylation that it serves as a glycan shield, facilitating immune evasion by masking non-self viral peptides with self-glycans (Stertz et al., 2007; Ujike and Taguchi, 2015; Watanabe et al., 2020b; Watanabe et al., 2019). In parallel with their potential masking functions, glycan-dependent epitopes can elicit specific, even neutralizing, antibody responses, as has been described for HIV-1 (Duan et al., 2018; Escolano et al., 2019; Pinto et al., 2020; Seabright et al., 2020; Watanabe et al., 2019; Yu et al., 2018; https://www.biorxiv.org/content/10.1101/2020.06.30.178897v1). Thus, understanding the glycosylation of the viral Spike trimer is fundamental for the development of efficacious vaccines, neutralizing antibodies, and therapeutic inhibitors of infection.

ACE2 is an integral membrane metalloproteinase that regulates the renin-angiotensin system (Tikellis et al., 2011). Both SARS-CoV-1 and SARS-CoV-2 have co-opted ACE2 to function as the receptor by which these viruses attach and fuse with host cells (Hoffmann et al., 2020; Li et al., 2003). ACE2 is cleavable by ADAM proteases at the cell surface (Lambert et al., 2005), resulting in the shedding of a soluble ectodomain that can be detected in apical secretions of various epithelial layers (gastric, airway, etc.) and in serum (Epelman et al., 2009). The N-terminal extracellular domain of ACE2 contains six canonical sequons for N-linked glycosylation and several potential O-linked sites. Several nonsynonymous single-nucleotide polymorphisms (SNPs) in the ACE2 gene have been identified in the human population and could potentially alter ACE2 glycosylation and/or affinity of the receptor for the viral Spike protein (Li et al., 2005). Given that glycosylation can affect the half-life of circulating glycoproteins in addition to modulating the affinity of their interactions with receptors and immune/inflammatory signaling pathways (Marth and Grewal, 2008; Varki, 2017), understanding the impact of glycosylation of ACE2 with respect to its binding of SARS-CoV-2 Spike glycoprotein is of high importance. The proposed use of soluble extracellular domains of ACE2 as decoy,
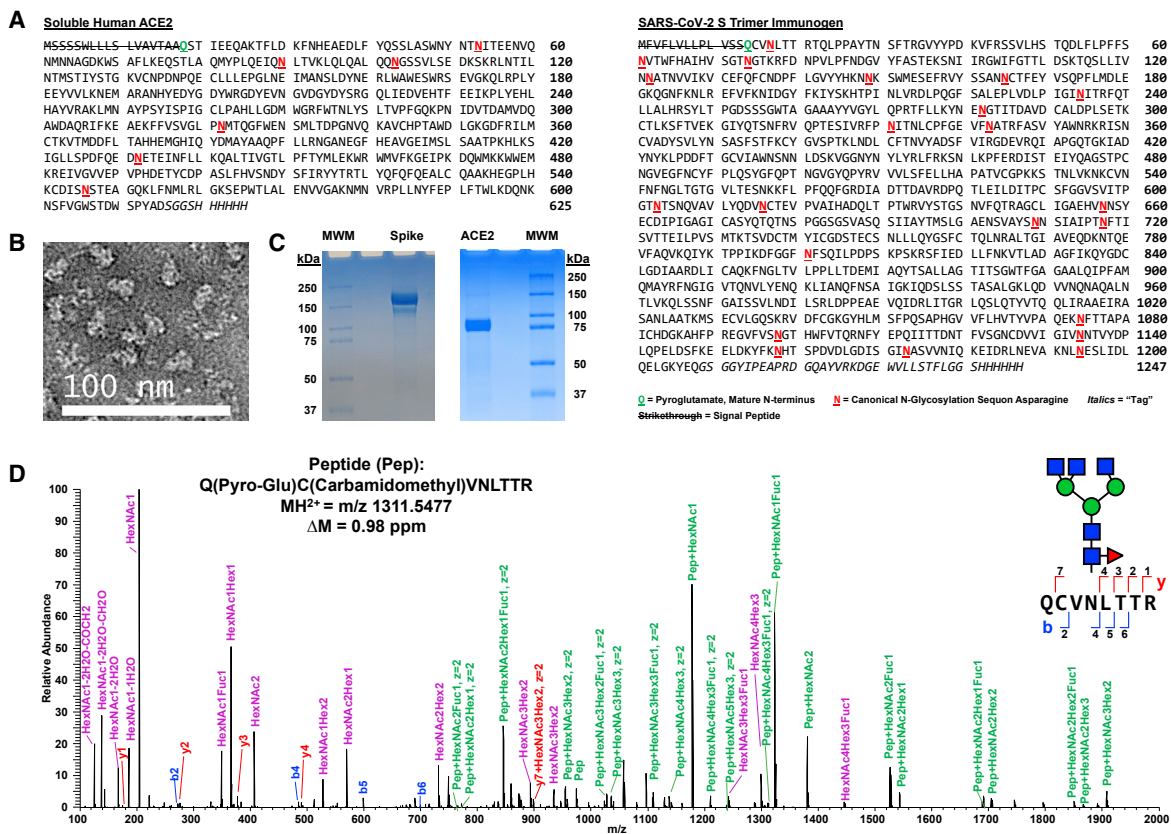
competitive inhibitors for SARS-CoV-2 infection emphasizes the critical need for understanding the glycosylation profile of ACE2 so that optimally active biologics can be produced (Lei et al., 2020; Monteil et al., 2020).

To accomplish the task of characterizing site-specific glycosylation of the trimer Spike of SARS-CoV-2 and the host receptor ACE2, we began by expressing and purifying a stabilized, soluble trimer Spike glycoprotein mimetic immunogen (that we define here and forward as S, [Yu et al., 2020]) and a soluble version of the ACE2 glycoprotein from a human cell line. We utilized multiple mass-spectrometry-based approaches, including glycomic and glycoproteomic approaches, to determine occupancy and site-specific heterogeneity of N-linked glycans. Occupancy (i.e., the percent of any given residue being modified by a glycan) is an important consideration when developing neutralizing antibodies against a glycan-dependent epitope. We also identified sites of O-linked glycosylation and the heterogeneity of the O-linked glycans on S and ACE2. We leveraged this rich dataset, along with existing 3D-structures of both glycoproteins, to generate static and molecular dynamics (MD) models of S alone, and in complex with the glycosylated, soluble ACE2 receptor. By combining bioinformatics characterization of viral evolution and variants of S and ACE2 with MD simulations of the glycosylated S-ACE2 interaction, we identified important roles for glycans in multiple processes, including receptor-viral binding and glycan shielding of S. Our rich characterization of the recombinant, glycosylated S trimer mimetic immunogen of SARS-CoV-2 in complex with the soluble human ACE2 receptor provides a detailed platform for guiding rational vaccine, antibody, and inhibitor design.

RESULTS

Expression, Purification, and Characterization of SARS-CoV-2 Spike Glycoprotein Trimer and Soluble Human ACE2

A trimer-stabilized, soluble variant of the SARS-CoV-2 S that contains 22 canonical N-linked glycosylation sequons per protomer and a soluble version of human ACE2 that contains six, lacking the most C-terminal seventh, canonical N-linked glycosylation sequons (Figure 1A) were purified from the media of transfected HEK293 cells, and the quaternary structure confirmed by negative EM staining for the S trimer (Figure 1B) and purity examined by SDS-PAGE Coomassie G-250 stained gels for both (Figure 1C). In addition, proteolytic digestions followed by proteomic analyses confirmed that the proteins were highly purified (Table S12). Finally, the N terminus of both the mature S and the soluble mature ACE2 were empirically determined via proteolytic digestions and liquid chromatography-tandem mass spectrometry (LC-MS/MS) analyses. These results confirmed that both the secreted, mature forms of S protein and ACE2 begin with an N-terminal glutamine that has undergone condensation to form pyroglutamine at residues 14 and 18, respectively (Figures 1D and S1). The N-terminal peptide observed for S also contains a glycan at Asn-0017 (Figure 1D), and mass spectrometry analysis of non-reducing proteolytic digestions confirmed that Cys-0015 of S is in a disulfide linkage with Cys-0136 (Figure S2; Table S2). Given that SignalP (Almagro Armenteros et al., 2019) predicts signal sequence cleavage between Cys-0015 and Val-0016 but we observed cleavage

A Soluble Human ACE2

```
MSSSSWLLLS LVAVTAAQST IEEQAKTFLD KFNHEAEDLF YQSSLASWNY NTNITEENVQ    60
NMNNAGDKWS AFLKEQSTLA QMYPLQEIQN LTVKLQLQAL QQNGSSVLSE DKSKRLNTIL   120
NTMSTIYSTG KVCNPDNPQE CLLLEPGLNE IMANSLDYNE RLWAWESWRS EVGKQLRPLY   180
EEYVVLKNEM ARANHYEDYG DYWRGDYEVN GVDGYDYSRG QLIEDVEHTF EEIKPLYEHL   240
HAYVRAKLMN AYPSYISPIG CLPAHLLGDM WGRFWTNLYS LTVPFGQKPN IDVTDAMVDQ   300
AWDAQRIFKE AEKFFVSVGL PNMTQGFWEN SMLTDPGNVQ KAVCHPTAWD LGKGDFRILM   360
CTKVTMDDFL TAHHEMGHIQ YDMAYAAQPF LLRNGANEGF HEAVGEIMSL SAATPKHLKS   420
IGLLSPDFQE DNETEINFLL KQALTIVGTL PFTYMLEKWR WMVFKGEIPK DQWMKKWWEM   480
KREIVGVVEP VPHDETYCDP ASLFHVSNDY SFIRYYTRTL YQFQFQEALC QAAKHEGPLH   540
KCDISNSTEA GQKLFNMLRL GKSEPWTLAL ENVVGAKNMN VRPLLNYFEP LFTWLKDQNK   600
NSFVGWSTDW SPYADSGGSH HHHHH                                        625
```

SARS-CoV-2 S Trimer Immunogen

```
MFVFLVLLPL VSSQCVNLTT RTQLPPAYTN SFTRGVYYPD KVFRSSVLHS TQDLFLPFFS    60
NVTWFHAIHV SGTNGTKRFD NPVLPFNDGV YFASTEKSNI IRGWIFGTTL DSKTQSLLIV   120
NNATNVVIKV CEFQFCNDPF LGVYYHKNNK SWMESEFRVY SSANNCTFEY VSQPFLMDLE   180
GKQGNFKNLR EFVFKNIDGY FKIYSKHTPI NLVRDLPQGF SALEPLVDLP IGINITRFQT   240
LLALHRSYLT PGDSSSGWTA GAAAYYVGYL QPRTFLLKYN ENGTITDAVD CALDPLSETK   300
CTLKSFTVEK GIYQTSNFRV QPTESIVRFP NITNLCPFGE VFNATRFASV YAWNRKRISN   360
CVADYSVLYN SASFSTFKCY GVSPTKLNDL CFTNVYADSF VIRGDEVRQI APGQTGKIAD   420
YNYKLPDDFT GCVIAWNSNN LDSKVGGNYN YLYRLFRKSN LKPFERDIST EIYQAGSTPC   480
NGVEGFNCYF PLQSYGFQPT NGVGYQPYRV VVLSFELLHA PATVCGPKKS TNLVKNKCVN   540
FNFNGLTGTG VLTESNKKFL PFQQFGRDIA DTTDAVRDPQ TLEILDITPC SFGGVSVITP   600
GTNTSNQVAV LYQDVNCTEV PVAIHADQLT PTWRVYSTGS NVFQTRAGCL IGAEHVNNSY   660
ECDIPIGAGI CASYQTQTNS PGGSGSVASQ SIIAYTMSLG AENSVAYSNN SIAIPTNFTI   720
SVTTEILPVS MTKTSVDCTM YICGDSTECS NLLLQYGSFC TQLNRALTGI AVEQDKNTQE   780
VFAQVKQIYK TPPIKDFGGF NFSQILPDPS KPSKRSFIED LLFNKVTLAD AGFIKQYGDC   840
LGDIAARDLI CAQKFNGLTV LPPLLTDEMI AQYTSALLAG TITSGWTFGA GAALQIPFAM   900
QMAYRFNGIG VTQNVLYENQ KLIANQFNSA IGKIQDSLSS TASALGKLQD VVNQNAQALN   960
TLVKQLSSNF GAISSVLNDI LSRLDPPEAE VQIDRLITGR LQSLQTYVTQ QLIRAAEIRA  1020
SANLAATKMS ECVLGQSKRV DFCGKGYHLM SFPQSAPHGV VFLHVTYVPA QEKNFTTAPA  1080
ICHDGKAHFP REGVFVSNGT HWFVTQRNFY EPQIITTDNT FVSGNCDVVI GIVNNTVYDP  1140
LQPELDSFKE ELDKYFKNHT SPDVDLGDIS GINASVVNIQ KEIDRLNEVA KNLNESLIDL  1200
QELGKYEQGS GGYIPEAPRD GQAYVRKDGE WVLLSTFLGG SHHHHHH                1247
```

Q = Pyroglutamate, Mature N-terminus   N = Canonical N-Glycosylation Sequon Asparagine   *Italics* = "Tag"
~~Strikethrough~~ = Signal Peptide

B

C

D Peptide (Pep):
Q(Pyro-Glu)C(Carbamidomethyl)VNLTTR
MH^{2+} = m/z 1311.5477
ΔM = 0.98 ppm

Figure 1. Expression and Characterization of SARS-CoV-2 Spike Glycoprotein Trimer Immunogen and Soluble Human ACE2

(A) Sequences of SARS-CoV-2 S immunogen and soluble human ACE2. The N-terminal pyroglutamines for both mature protein monomers are bolded, underlined, and shown in green. The canonical N-linked glycosylation sequons are bolded, underlined, and shown in red.

(B and C) Negative stain electron microscopy of the purified trimer (B) and Coomassie G-250-stained reducing SDS-PAGE gels (C) confirmed purity of the SARS-CoV-2 S protein trimer and of the soluble human ACE2. MWM, molecular weight markers.

(D) A representative Step-HCD fragmentation spectrum from mass-spectrometry analysis of a tryptic digest of S annotated manually based on search results from pGlyco 2.2. This spectrum defines the N terminus of the mature protein monomer as (pyro-)glutamine 0014. A representative N-glycan consistent with this annotation and our glycomics data (Figure 2) is overlaid by using the Symbol Nomenclature For Glycans (SNFG) code. This complex glycan occurs at N0017. Note, that as expected, the cysteine is carbamidomethylated, and the mass accuracy of the assigned peptide is 0.98 ppm. On the sequence of the N-terminal peptide and in the spectrum, the assigned b (blue) and y (red) ions are shown. In the spectrum, purple highlights glycan oxonium ions and green marks intact peptide fragment ions with various partial glycan sequences still attached. Note that the green-labeled ions allow for limited topology to be extracted including defining that the fucose is on the core and not the antennae of the glycopeptide.

between Ser-0013 and Gln-0014, we examined the possibility that an in-frame upstream methionine to the proposed start methionine (Figure 1A) might be used to initiate translation (Figure S3). If one examines the predicted signal sequence cleavage using the in-frame Met that is encoded nine amino acids upstream, SignalP now predicts cleavage between the Ser and Gln that we observed in our studies (Figure S3). To examine whether this impacted S expression, we expressed constructs that contained or did not contain the upstream 27 nucleotides in a pseudovirus (VSV) system expressing SARS-CoV-2 S (Figure S4) and in our HEK293 system (data not shown). Both expression systems produced a similar amount of S regardless of which expression construct was utilized (Figure S4). Thus, while the translation initiation start site has still not been fully defined, allowing for earlier translation in expression construct design did not have a significant impact on the generation of S.

Glycomics-Informed Glycoproteomics Reveals Site-Specific Microheterogeneity of SARS-CoV-2 S Glycosylation

We utilized multiple approaches to examine glycosylation of the SARS-CoV-2 S trimer. First, the portfolio of glycans linked to SARS-CoV-2 S trimer immunogen was analyzed after their release from the polypeptide backbone. N-glycans were released from protein by treatment with PNGase F- and O-glycans were subsequently released by beta-elimination. After permethylation to enhance detection sensitivity and structural characterization, released glycans were analyzed by multi-stage mass spectrometry (MS^n) (Aoki et al., 2007; Aoki et al., 2008). Mass spectra were processed by GRITS Toolbox, and the resulting annotations were validated manually (Weatherly et al., 2019). Glycan assignments were grouped by type and by additional structural features for relative quantification of profile characteristics (Figure 2A; Table S3). This analysis
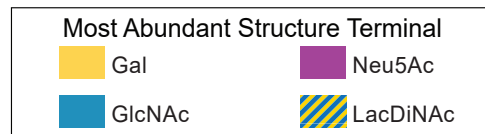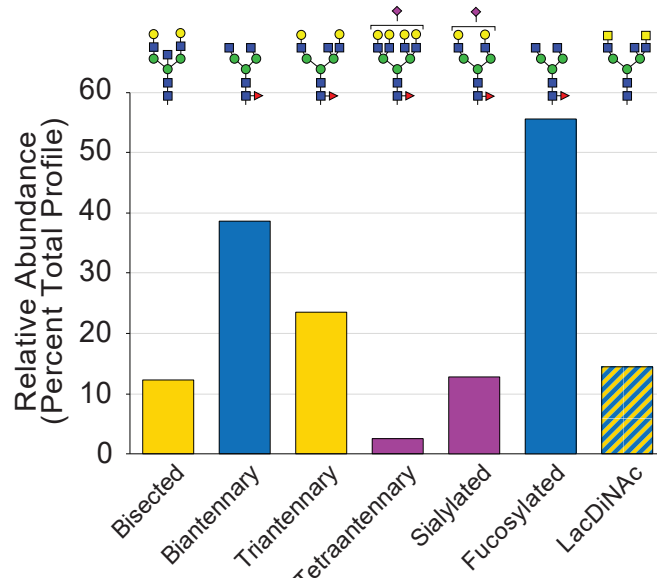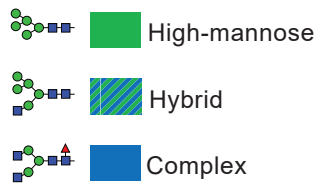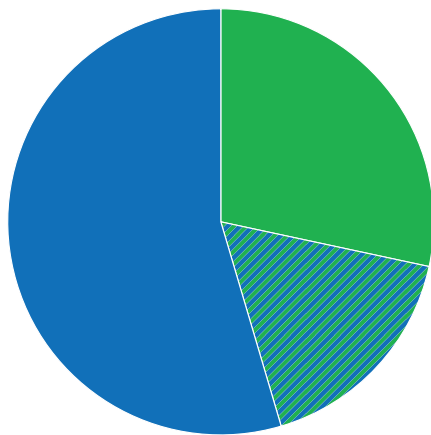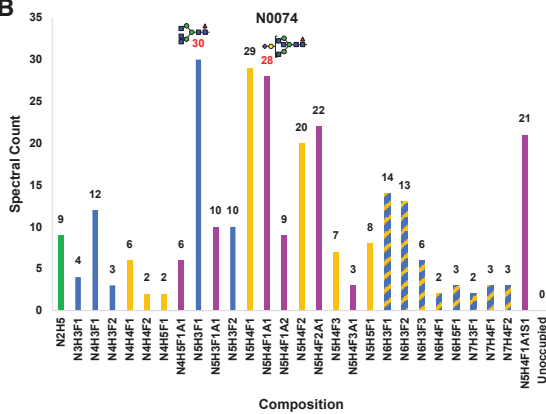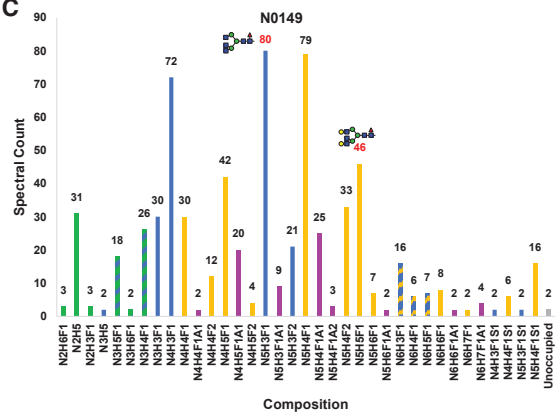
quantified 49 N-glycans and revealed that 55% of the total glycan abundance was of the complex type, 17% was of the hybrid type, and 28% was high mannose. Among the complex and hybrid N-glycans, we observed a high degree of core fucosylation and significant abundance of bisected and LacDiNAc structures. We also observed sulfated N-linked glycans by using negative mode MS^n analyses (Table S13), although signal intensity was too low in positive ion mode (at least 10-fold lower than any of the non-sulfated glycans) for accurate quantification. In addition, we detected 15 O-glycans released from the S trimer (Figure S5; Table S4).

To determine occupancy of N-linked glycans at each site, we employed a sequential deglycoslyation approach by using Endoglycosidase H and PNGase F in the presence of ^{18}O-H$_2$O after tryptic digestion of S (Wang et al., 2020; Yu et al., 2018). After LC-MS/MS analyses, the resulting data confirmed that 19 of the canonical sequons had occupancies greater than 95% (Table S5). One canonical sequence, N0149, had insufficient spectral counts for quantification by this method, but subsequent analyses described below suggested high occupancy. The two most C-terminal N-linked sites, N1173 and N1194, had reduced occupancy, 52% and 82% respectively. Reduced occupancy at these sites could reflect hindered en bloc transfer by the oligosaccharyltransferase (OST) due to primary amino acid sequences at or near the N-linked sequon. Alternatively, this could reflect these two sites being post-translationally modified after release of the protein by the ribosome by a less efficient STT3B-containing OST, either due to activity or initial folding of the polypeptide, as opposed to co-translationally modified by the STT3A-containing OST (Ruiz-Canada et al., 2009). None of the non-canonical sequons (three N-X-C sites and four N-G-L/I/V sites; Zielinska et al., 2010) showed significant occupancy (>5%), except for N0501, which showed moderate (19%) conversion to ^{18}O-Asp that could be due to deamidation that is facilitated by glycine at the +1 position (Table S5) (Palmisano et al., 2012). Further analysis of this site (see below) by direct glycopeptide analyses allowed us to determine that N0501 undergoes deamidation but is not glycosylated. Thus, all, and only the, 22 canonical sequences for N-linked glycosylation (N-X-S/T) are utilized, with only N1173 and N1194 demonstrating occupancies below 95%.

Next, we applied three different proteolytic digestion strategies to the SARS-CoV-2 S immunogen to maximize glycopeptide coverage by subsequent LC-MS/MS analyses. Extended gradient nanoflow reverse-phase LC-MS/MS was carried out

on a ThermoFisher Lumos Tribrid instrument using Step-HCD fragmentation on each of the samples (see STAR Methods for details, as well as Duan et al., 2018; Escolano et al., 2019; Wang et al., 2020; Yu et al., 2018; Zhou et al., 2017). After data analyses using pGlyco 2.2.2 (Liu et al., 2017), Byonic (Bern et al., 2012), and manual validation of glycan compositions against our released glycomics findings (Figure 2A; Tables S3 and S13), we were able to determine the microheterogeneity at each of the 22 canonical sites (Figures 2B–2E; Table S6). Notably, none of the non-canonical consensus sequences, including N0501, displayed any quantifiable glycans. The N-glycosites N0074 (Figure 2B) and N0149 (Figure 2C) are highly processed and display a typical mammalian N-glycan profile. N0149 is, however, modified with several hybrid N-glycan structures, whereas N0074 is not. N0234 (Figure 2D) and N0801 (Figure 2E) have N-glycan profiles more similar to those found on other viruses such as HIV (Watanabe et al., 2019) that are dominated by high-mannose structures. N0234 (Figure 2D) displays an abundance of Man7-Man9 high-mannose structures, suggesting stalled processing by early-acting ER and *cis*-Golgi mannosidases. In contrast, N0801 (Figure 2E) is processed more efficiently to Man5 high-mannose and hybrid structures, suggesting that access to the glycan at this site by MGAT1 and α-Mannosidase II is hindered. In general, for all 22 sites (Figures 2B–2E; Table S6), we observed underprocessing of complex glycan antennae (i.e., under-galactosylation and under-sialylation) and a high degree of core fucosylation in agreement with released glycan analyses (Figure 2A; Table S3). We also observed a small percent of sulfated N-linked glycans at several sites (Tables S6 and S8). Based on the assignments and the spectral counts for each topology, we were able to determine the percent of total N-linked glycan types (high-mannose, hybrid, or complex) present at each site (Figure 3; Table S7). Notably, three of the sites (N0234, N0709, and N0717) displayed more than 50% high-mannose glycans, whereas 11 other sites (N0017, N0074, N0149, N0165, N0282, N0331, N0657, N1134, N1158, N1173, and N1194) were more than 90% complex when occupied. The other eight sites were distributed between these two extremes. Notably, only one site (N0717 at 45%), which also had greater than 50% high mannose (55%), had greater than 33% hybrid structures. To further evaluate the heterogeneity, we grouped all the topologies into the 20 classes recently described by the Crispin laboratory, adding two categories (sulfated and unoccupied) that we refer

Figure 2. Glycomics-Informed Glycoproteomics Reveals Substantial Site-Specific Microheterogeneity of N-linked Glycosylation on SARS-CoV-2 S

(A) Glycans released from SARS-CoV-2 S protein trimer immunogen were permethylated and analyzed by MS^n. Structures were assigned and grouped by type and structural features, and prevalence was determined based on ion current. The pie chart shows basic division by broad N-glycan type. The bar graph provides additional detail about the glycans detected. The most abundant structure with a unique categorization by glycomics for each N-glycan type in the pie chart, or above each feature category in the bar graph, is indicated.

(B–E) Glycopeptides were prepared from SARS-CoV-2 S protein trimer immunogen by using multiple combinations of proteases, analyzed by LC-MS^n, and the resulting data were searched by using several different software packages. Four representative sites of N-linked glycosylation with specific features of interest were chosen and are presented here. N0074 (B) and N0149 (C) are shown that occur in variable insert regions of S compared to SARS-CoV and other related coronaviruses, and there are emerging variants of SARS-CoV-2 that disrupt these two sites of glycosylation in S. N0234 (D) contains the most high-mannose N-linked glycans. N0801 (D) is an example of glycosylation in the S2 region of the immunogen and displays a high degree of hybrid glycosylation compared to other sites. The abundance of each composition is graphed in terms of assigned spectral counts. Representative glycans (as determined by glycomics analysis) for several abundant compositions are shown in SNFG format. The abbreviations used here and throughout the manuscript are as follows: N, HexNAc; H, hexose; F, fucose; A, Neu5Ac; S, sulfation. Note that the graphs for the other 18 sites and other graphs grouping the microheterogeneity observed by other properties are presented in Supplemental Information.
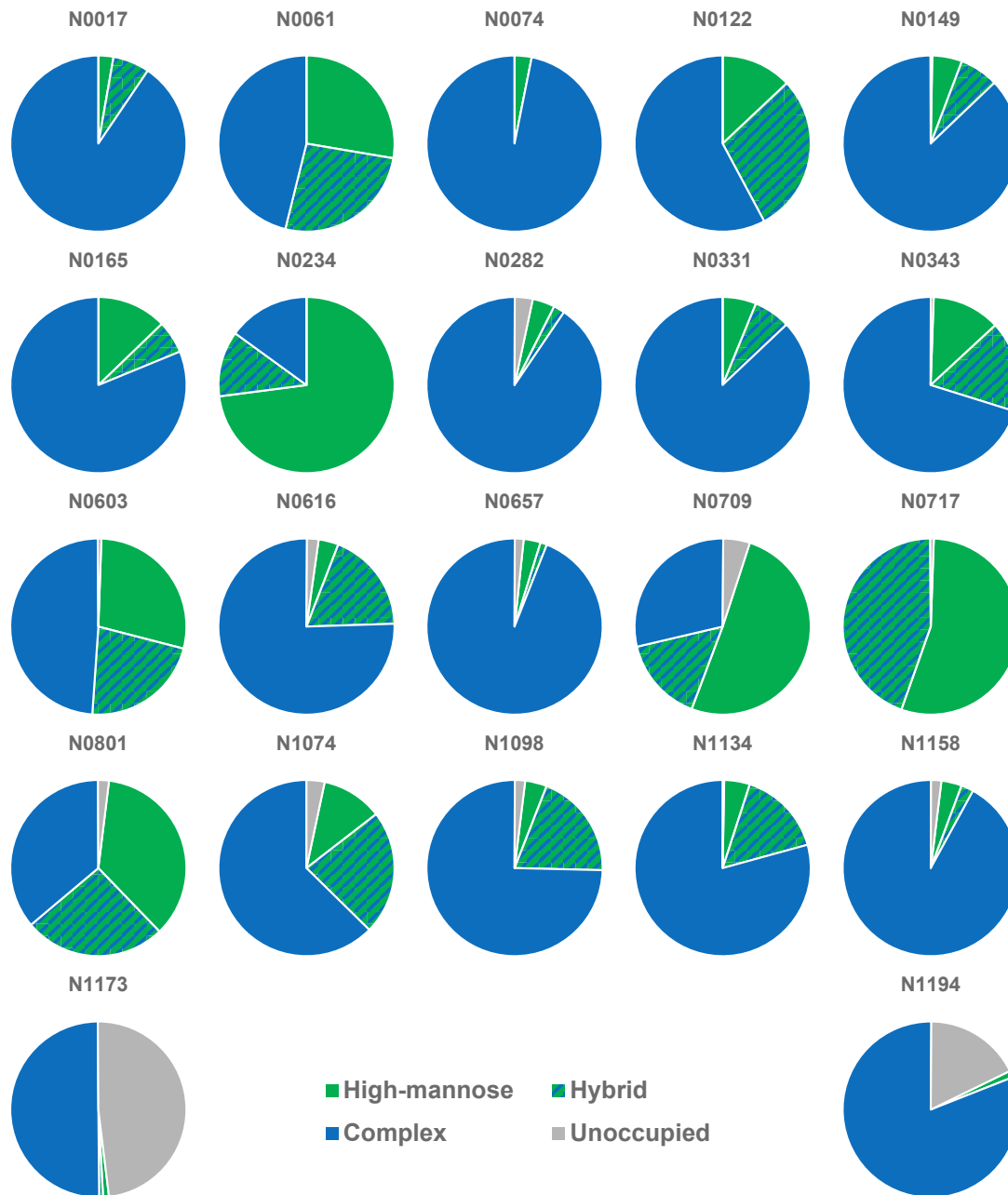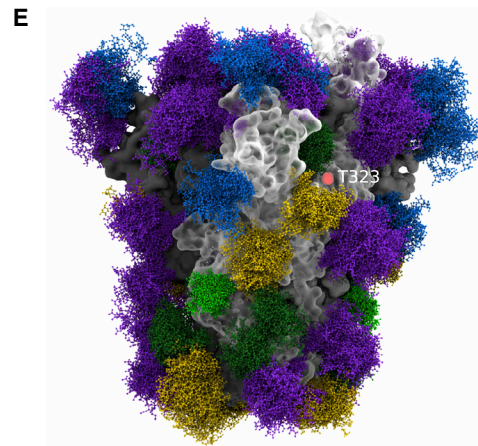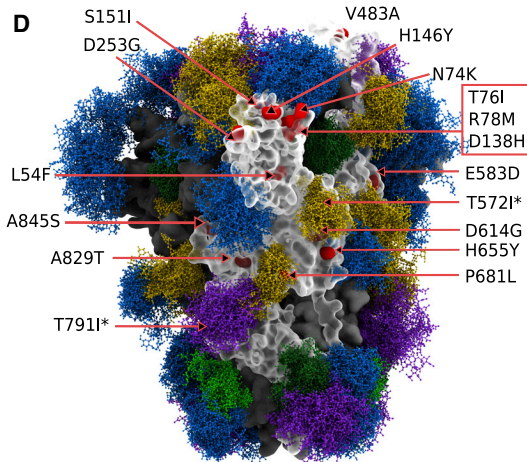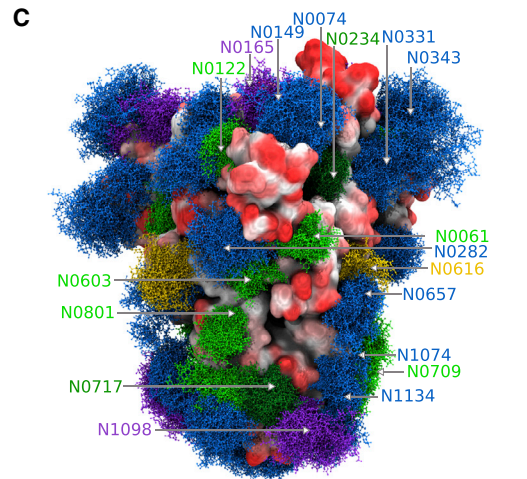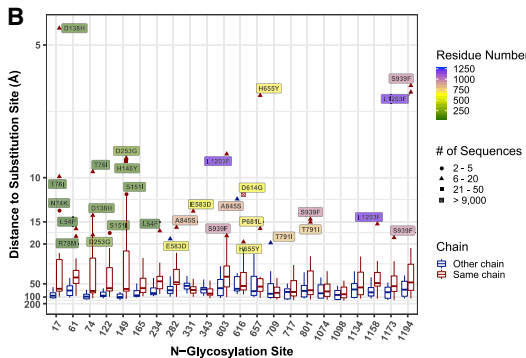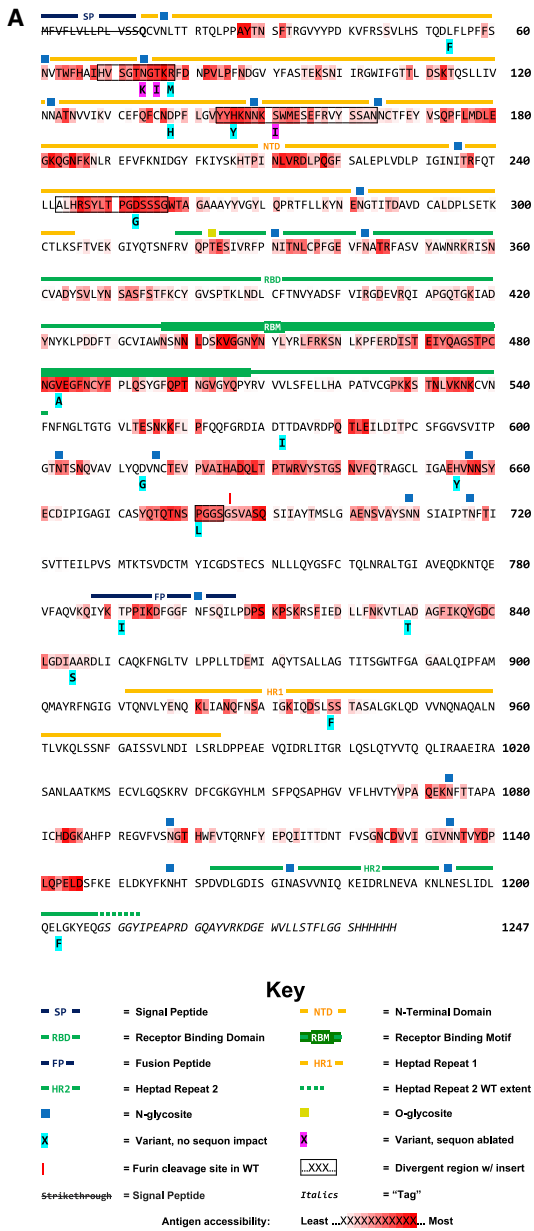
Figure 3. SARS-CoV-2 S Immunogen N-glycan Sites Are Predominantly Modified by Complex N-glycans

N-glycan topologies were assigned to all 22 sites of the S protomer and the spectral counts for each of the three types of N-glycans (high-mannose, hybrid, and complex), as well as the unoccupied peptide spectral match counts at each site, were summed and visualized as pie charts. Note that only N1173 and N1194 show an appreciable amount of the unoccupied amino acid.

to here as the Oxford classification (Table S8) (Watanabe et al., 2020a). Among other features observed, this classification allowed us to observe that although most sites with high-mannose structures were dominated by the Man5GlcNAc2 structure, N0234 and N0717 were dominated by the higher Man structures of Man8GlcNAc2 and Man7GlcNAc2, respectively (Figure S7; Table S8). Limited processing at N0234 is in agreement with a recent report suggesting that high-mannose structures at this site help to stabilize the receptor-

binding domain of S (www.biorxiv.org/content/10.1101/2020.06.11.146522v1). Furthermore, applying the Oxford classifications to our dataset clearly demonstrates that the three most C-terminal sites (N1158, N1173, and N1194), dominated by complex-type glycans, were more often further processed (i.e., multiple antennae) and elaborated (i.e., galactosylation and sialylation) than other sites (Table S8).

We also analyzed our generated mass spectrometry data for the presence of O-linked glycans based on our glycomic findings

A

MFVFLVLLPL VSSQCVNLTT RTQLPPAYTN SFTRGVYYPD KVFRSSVLHS TQDLFLPFFS 60

NVTWFHAIHV SGTNGTKRFD NPVLPFNDGV YFASTEKSNI IRGWIFGTTL DSKTQSLLIV 120

NNATNVVIKV CEFQFCNDPF LGVYYHKNNK SWMESEFRVY SSANNCTFEY VSQPFLMDLE 180

GKQGNFKNLR EFVFKNIDGY FKIYSKHTPI NLVRDLPQGF SALEPLVDLP IGINITRFQT 240

LLALHRSYLT PGDSSSGWTA GAAAYYVGYL QPRTFLLKYN ENGTITDAVD CALDPLSETK 300

CTLKSFTVEK GIYQTSNFRV QPTESIVRFP NITNLCPFGE VFNATRFASV YAWNRKRISN 360

CVADYSVLYN SASFSTFKCY GVSPTKLNDL CFTNVYADSF VIRGDEVRQI APGQTGKIAD 420

YNYKLPDDFT GCVIAWNSNN LDSKVGGNYN YLYRLFRKSN LKPFERDIST EIYQAGSTPC 480

NGVEGFNCYF PLQSYGFQPT NGVGYQPYRV VVLSFELLHA PATVCGPKKS TNLVKNKCVN 540

FNFNGLTGTG VLTESNKKFL PFQQFGRDIA DTTDAVRDPQ TLEILDITPC SFGGVSVITP 600

GTNTSNQVAV LYQDVNCTEV PVAIHADQLT PTWRVYSTGS NVFQTRAGCL IGAEHVNNSY 660

ECDIPIGAGI CASYQTQTNS PGGSGSVASQ SIIAYTMSLG AENSVAYSNN SIAIPTNFTI 720

SVTTEILPVS MTKTSVDCTM YICGDSTECS NLLLQYGSFC TQLNRALTGI AVEQDKNTQE 780

VFAQVKQIYK TPPIKDFGGF NFSQILPDPS KPSKRSFIED LLFNKVTLAD AGFIKQYGDC 840

LGDIAARDLI CAQKFNGLTV LPPLLTDEMI AQYTSALLAG TITSGWTFGA GAALQIPFAM 900

QMAYRFNGIG VTQNVLYENQ KLIANQFNSA IGKIQDSLSS TASALGKLQD VVNQNAQALN 960

TLVKQLSSNF GAISSVLNDI LSRLDPPEAE VQIDRLITGR LQSLQTYVTQ QLIRAAEIRA 1020

SANLAATKMS ECVLGQSKRV DFCGKGYHLM SFPQSAPHGV VFLHVTYVPA QEKNFTTAPA 1080

ICHDGKAHFP REGVFVSNGT HWFVTQRNFY EPQIITTDNT FVSGNCDVVI GIVNNTVYDP 1140

LQPELDSFKE ELDKYFKNHT SPDVDLGDIS GINASVVNIQ KEIDRLNEVA KNLNESLIDL 1200

QELGKYEQGS GGYIPEAPRD GQAYVRKDGE WVLLSTFLGG SHHHHHH 1247

Key

SP	= Signal Peptide	NTD	= N-Terminal Domain
RBD	= Receptor Binding Domain	RBM	= Receptor Binding Motif
FP	= Fusion Peptide	HR1	= Heptad Repeat 1
HR2	= Heptad Repeat 2		= Heptad Repeat 2 WT extent
	= N-glycosite		= O-glycosite
X	= Variant, no sequon impact	X	= Variant, sequon ablated
	= Furin cleavage site in WT	..XXX..	= Divergent region w/ insert
~~Strikethrough~~	= Signal Peptide	*Italics*	= "Tag"

Antigen accessibility: Least ...XXXXXXXXXX... Most

B

Distance to Substitution Site (Å) vs N−Glycosylation Site

Residue Number: 1250 / 1000 / 750 / 500 / 250

# of Sequences: 2 - 5 / 6 - 20 / 21 - 50 / > 9,000

Chain: Other chain / Same chain

C

N0074, N0149, N0165, N0122, N0234, N0331, N0343, N0061, N0282, N0616, N0603, N0657, N0801, N1074, N0709, N0717, N1134, N1098

D

S151I, D253G, V483A, H146Y, N74K, T76I, R78M, D138H, L54F, E583D, T572I*, A845S, D614G, A829T, H655Y, P681L, T791I*

E

T323

(legend on next page)

(Figure S5; Table S4) and a recent manuscript suggesting significant levels of O-glycosylation of S1 and S2 when expressed independently (Shajahan et al., 2020). We were able to confirm sites of O-glycan modification with microheterogeneity observed for the vast majority of these sites (Table S9). However, occupancy at each site, determined by spectral counts, was observed to be very low (below 4%), except for Thr0323, which had a modestly higher but still low 11% occupancy (Figure S6; Table S10).

3D Structural Modeling of Glycosylated SARS-CoV-2 Trimer Immunogen Enables Predictions of Epitope Accessibility and Other Key Features

A 3D structure of the S trimer was generated by using a homology model of the S trimer described previously (based on PDB: 6VSB; Wrapp et al., 2020). Onto this 3D structure, we installed explicitly defined glycans at each glycosylated sequon based on one of three separate sets of criteria, thereby generating three different glycoform models for comparison that we denote as "Abundance," "Oxford Class," and "Processed" models (STAR Methods; Table S1). These criteria were chosen in order to generate glycoform models that represent reasonable expectations for glycosylation microheterogeneity and integrate cross-validating glycomic and glycoproteomic characterization of S and ACE2.

The three glycoform models were subjected to multiple all-atom MD simulations with explicit water. Information from analyses of these structures is presented in Figure 4A along with the sequence of the SARS-CoV-2 S protomer. We also determined variants in S that are emerging in the virus that have been sequenced to date (Table S11). The inter-residue distances were measured between the most α-carbon-distal atoms of the N-glycan sites and Spike glycoprotein population variant sites in 3D space (Figure 4B). Notable from this analysis, there are several variants that don't ablate the N-linked sequon but are sufficiently close in 3D space to N-glycosites, such as D138H, H655Y, S939F, and L1203F, to warrant further investigation.

The percentage of simulation time that each S protein residue is accessible to a probe that approximates the size of an antibody variable domain was calculated for a model of the S trimer by using the Abundance glycoforms (Table S1) (Ferreira et al., 2018). The predicted antibody accessibility is visualized across the sequence, as well as mapped onto the 3D surface, via color shading (Figures 4A and 4C; Table S13; Video S1). Additionally, the Oxford Class glycoforms model (Table S1), which is arguably

the most encompassing means for representing glycan microheterogeneity because it captures abundant structural topologies (Table S8), is shown with the sequence variant information (Figure 4D; Table S11). A substantial number of these variants occur (directly by comparison to Figure 4A or visually by comparison to Figure 4C) in regions of high calculated epitope accessibility (e.g., N74K, T76I, R78M, D138H, H146Y, S151I, D253G, V483A, etc.; Table S14), suggesting potential selective pressure to avoid host immune response. Also, it is interesting to note that three of the emerging variants would eliminate N-linked sequons in S; N74K and T76I would eliminate N-glycosylation of N74 (found in the insert variable region 1 of CoV-2 S compared to CoV-1 S), and S151I eliminates N-glycosylation of N149 (found in the insert variable region 2) (Figures 4A and S7; Table S11). Lastly, the SARS-CoV-2 S Processed glycoform model is shown (Table S1), along with marking amino acid T0323 that has a modest (11% occupancy, Figure S6; Table S10) amount of O-glycosylation to represent the most heavily glycosylated form of S (Figure 4E).

Glycomics-Informed Glycoproteomics Reveals Complex N-linked Glycosylation of ACE2

We also analyzed ACE2 glycosylation utilizing the same glycomic and glycoproteomic approaches described for S protein. Glycomic analyses of released N-linked glycans (Figure 5A; Table S3) revealed that the majority of glycans on ACE2 are complex with limited high-mannose and hybrid glycans, and we were unable to detect sulfated N-linked glycans. Glycoproteomic analyses revealed that occupancy was high (> 75%) at all six sites, and significant microheterogeneity dominated by complex N-glycans was observed for each site (Figures 5B–5G; Tables S5–S8). We also observed, consistent with the O-glycomics (Figure S5; Table S4), that Ser 155 and several S/T residues at the C terminus of ACE2 outside of the peptidase domain were O-glycosylated, but stoichiometry was extremely low (less than 2%; Tables S9 and S10).

3D Structural Modeling of Glycosylated, Soluble, ACE2-Highlighting Glycosylation and Variants
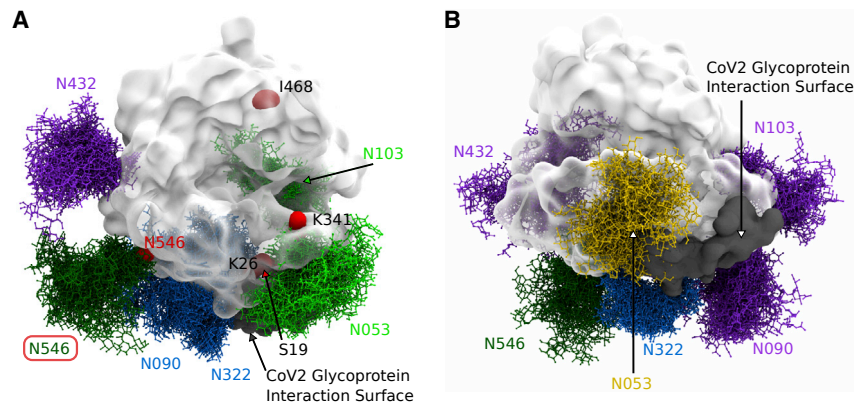
We integrated our glycomics, glycoproteomics, and population variant analyses results with a 3D model of Ace 2 (based on PDB: 6M0J (Lan et al., 2020; see STAR Methods for details) to generate two versions of the soluble glycosylated ACE2 for visualization and MD simulations. We visualized the ACE2 glycoprotein with the Abundance glycoform model simulated at each site

---

Figure 4. 3D Structural Modeling of Glycosylated SARS-CoV-2 Spike Trimer Immunogen Reveals Predictions for Antigen Accessibility and Other Key Features

Results from glycomics and glycoproteomics experiments were combined with results from bioinformatics analyses and used to model several versions of glycosylated SARS-CoV-2 S trimer immunogen.

(A) Sequence of the SARS-CoV-2 S immunogen displaying computed antigen accessibility and other information. Antigen accessibility is indicated by red shading across the amino acid sequence.

(B) Emerging variants confirmed by independent sequencing experiments were analyzed based on the 3D structure of SARS-CoV-2 S to generate a proximity chart to the determined N-linked glycosylation sites.

(C) SARS-CoV-2 S trimer immunogen model from MD simulation displaying abundance glycoforms and antigen accessibility shaded in red for most accessible, white for partial, and black for inaccessible (see Video S1).

(D) SARS-CoV-2 S trimer immunogen model from MD simulation displaying Oxford Class glycoforms and sequence variants. Asterisk indicates not visible, whereas the box represents three amino acid variants that are clustered together in 3D space.

(E) SARS-CoV-2 S trimer immunogen model from MD simulation displaying processed glycoforms plus shading of Thr-323 that has O-glycosylation at low stoichiometry in yellow.

Figure 5. Glycomics-Informed Glycoproteomics of Soluble Human ACE2 Reveals High Occupancy, Complex N-linked Glycosylation

(A) Glycans released from soluble, purified ACE2 were permethylated and analyzed by MS^n. Structures were assigned, grouped by type and structural features, and prevalence was determined based on ion current. The pie chart shows basic division by broad N-glycan type. The bar graph provides additional detail about

(legend continued on next page)

Figure 6. 3D Structural Modeling of Glycosylated Soluble Human ACE2

Results from glycomics and glycoproteomics experiments were combined with results from bioinformatics analyses and used to model several versions of glycosylated soluble human ACE2.

(A) Soluble human ACE2 model from MD simulations displaying abundance glycoforms, interaction surface with S, and sequence variants. N546 variant is boxed that would remove N-linked glycosylation at that site (see Video S2).

(B) Soluble human ACE2 model from MD simulations displaying processed glycoforms and interaction surface with S.

as well as highlighting the naturally occurring variants observed in the human population (Figure 6A; Video S2; Table S11). Note, that the Abundance glycoform model and the Oxford Class glycoform model for ACE2 are identical (Tables S1 and S8). Notably, one site of N-linked glycosylation (N546) is predicted to not be present in three out of 10,000 humans based on naturally occurring variation in the human population (Table S11). We also modeled ACE2 using the Processed glycoform model (Figure 6B). In both models, the interaction domain with S is defined (Figures 6A and 6B; Video S2) .

MD Simulation of the Glycosylated Trimer Spike of SARS-CoV-2 in Complex with Glycosylated, Soluble, Human Ace 2 Reveals Protein and Glycan Interactions

MD simulations were performed to examine the co-complex (generated from a crystal structure of the ACE2-RBD co-complex, PDB: 6M0J; Lan et al., 2020) of glycosylated S with glycosylated ACE2 with the three different glycoforms models (Abundance, Oxford Class, and Processed; Table S1; Videos S5, S6, and S7). Information from these analyses is laid out along the primary structure (sequence) of the SARS-CoV-2 S protomer and ACE2 highlighting regions of glycan-protein interaction observed in the MD simulations (Table S14; Videos S5, S6, and S7). Interestingly, two glycans on ACE2 (at N090 and N322), which are highlighted in Figure 7A and shown in a more close-up view in Figure 7B, are predicted to form interactions with the S protein (Table S15). The N322 glycan interaction with the S trimer is outside of the receptor-binding domain, and the interaction is observed across multiple simulations and throughout each simulation (Figures 7A and 7B; Video S5, S6, and S7). The ACE2 glycan at N090 is close enough to the S trimer surface to repeatedly form interactions; however, the glycan arms interact

with multiple regions of the surface over the course of the simulations, reflecting the relatively high degree of glycan dynamics (Figures 7A and 7B; Video S3). Inter-molecule glycan-glycan interactions are also observed repeatedly between the glycan at N546 of ACE2 and those in the S protein at residues N0074 and N0165 (Figure 7D; Table S16). Finally, a full view of the ACE2-S complex with Oxford class glycoforms on ACE2 illustrates the extensive glycosylation at the interface of the complex (Figure 7C; Video S4).

DISCUSSION

We have defined the glycomics-informed, site-specific microheterogeneity of 22 sites of N-linked glycosylation per monomer on a SARS-CoV-2 trimer and the six sites of N-linked glycosylation on a soluble version of its human ACE2 receptor by using a combination of mass spectrometry approaches coupled with evolutionary and variant sequence analyses to provide a detailed understanding of the glycosylation states of these glycoproteins (Figures 1, 2, 3, 4, 5, and 6). Our results suggest essential roles for glycosylation in mediating receptor binding, antigenic shielding, and potentially the evolution/divergence of these glycoproteins.

The highly glycosylated SARS-CoV-2 Spike protein, unlike several other viral proteins including HIV-1 (Watanabe et al., 2019) but in agreement with another recent report (Watanabe et al., 2020a), presents significantly more processing of N-glycans toward complex glycosylation, suggesting that steric hindrance to processing enzymes is not a major factor at most sites (Figures 2 and 3). However, the N-glycans still provide considerable shielding of the peptide backbone (Figure 4). Our glycomics-guided glycoproteomic data are generally in strong

the glycans detected. The most abundant structure with a unique categorization by glycomics for each N-glycan type in the pie chart, or above each feature category in the bar graph, is indicated.

(B–G) Glycopeptides were prepared from soluble human ACE2 by using multiple combinations of proteases, analyzed by LC-MSⁿ, and the resulting data were searched by using several different software packages. All six sites of N-linked glycosylation are presented here. Displayed in the bar graphs are the individual compositions observed graphed in terms of assigned spectral counts. Representative glycans (as determined by glycomics analysis) for several abundant compositions are shown in SNFG format. The pie chart (analogous to Figure 3 for SARS-CoV-2 S) for each site is displayed in the upper corner of each panel.
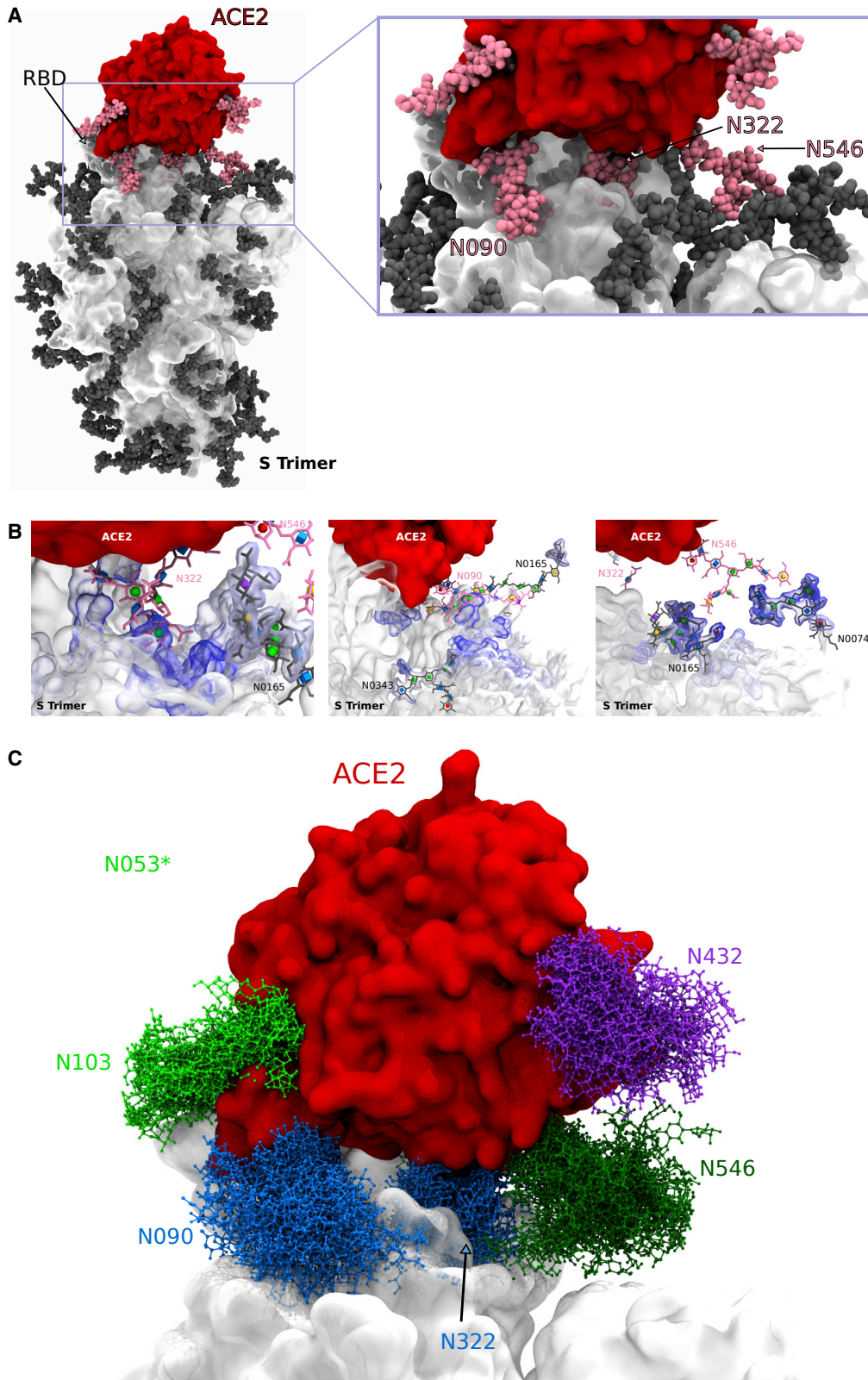
(B) N053.

(C) N090.

(D) N103.

(E) N322.

(F) N432.

(G) N546, a site that does not exist in three in 10,000 people.

agreement with the trimer immunogen data recently published by Crispin (Watanabe et al., 2020a), although we also observed sulfated N-linked glycans; were able to differentiate branching, bisected, and diLacNAc containing structures by glycomics; and observed less occupancy on the two most C-terminal N-linked sites by using a different approach. Our detection of sulfated N-linked glycans at multiple sites on S is in agreement with a recent manuscript re-analyzing the Crispin data (https://www.biorxiv.org/content/10.1101/2020.05.31.125302v1). Sulfated N-linked glycans could potentially play key roles in immune regulation and receptor binding as in other viruses (Wang et al., 2009). This result is especially significant in that sulfated N-glycans were not observed when we performed glycomics on ACE2. At each individual site, the glycans we observed on our immunogen appear to be slightly more processed, but the overlap between our analysis and the Crispin's group results (Watanabe et al., 2020a) at each site in terms of major features are nearly superimposable. This agreement differs substantially when comparing our and Crispin's data (Watanabe et al., 2020a) to that of the Azadi group (Shajahan et al., 2020), which analyzed S1 and S2 that had been expressed individually. When expressed as two separate polypeptides and not purified for trimers, several unoccupied sites of N-linked glycosylation were observed and processing at several sites was significantly different (Shajahan et al., 2020) than we and others (Watanabe et al., 2020a) observed. Although O-glycosylation has recently been reported for individually expressed S1 and S2 domains of the Spike glycoprotein (Shajahan et al., 2020), in trimeric form the level of O-glycosylation is extremely low, with the highest level of occupancy we observed being 11% at T0323 (Figure 4E). The low level of O-linked occupancy we observed is in agreement with the Crispin group's analysis of a Spike Trimer immunogen (Watanabe et al., 2020a) but differs significantly from the Azadi group's analyses of individually expressed S1 and S2 (Shajahan et al., 2020). Thus, the context in which the Spike protein is expressed and purified before analysis significantly alters the glycosylation of the protomer that is reminiscent of previous studies looking at expression of the HIV-1 envelope Spike (Behrens et al., 2017; Watanabe et al., 2019). The soluble ACE2 protein examined here contains six highly utilized sites of N-linked glycosylation dominated by complex type N-linked glycans (Figure 5). O-glycans were also present on this glycoprotein but at very low levels of occupancy at all sites (<2%).

Our glycomics-informed glycoproteomics allowed us to assign defined sets of glycans to specific glycosylation sites on 3D-structures of S and ACE2 glycoproteins based on experimental evidence (Figures 4 and 6). Similar to almost all glycoproteins, microheterogeneity is evident at most glycosylation sites of S and ACE2; each glycosylation site can be modified with one of several glycan structures, generating site-specific glyco-

sylation portfolios. For modeling purposes, however, explicit structures must be placed at each glycosylation site. In order to capture the impact of microheterogeneity on S and ACE2 MD we chose to generate glycoforms for modeling that represented reasonable portfolios of glycan types. Using three glycoform models for S (Abundance, Oxford Class, and Processed) and two models for ACE2 (Abundance, which was equivalent to Oxford Class, and Processed), we generated three MD simulations of the co-complexes of these two glycoproteins (Figure 7; Videos S5, S6, and S7). The observed interactions over time allowed us to evaluate glycan-protein contacts between the two proteins and examine potential glycan-glycan interactions (Figure 7).

We observed glycan-mediated interactions between the S trimer and glycans at N090, N322, and N546 of ACE2. Thus, variations in glycan occupancy or processing at these sites could alter the affinity of the SARS-CoV-2–ACE2 interaction and modulate infectivity. It is well established that glycosylation states vary depending on tissue and cell type as well as in the case of humans, on age (Krištić et al., 2014), underlying disease (Pavić et al., 2018; Rudman et al., 2019), and ethnicity (Gebrehiwot et al., 2018). Thus, glycosylation portfolios could in part be responsible for tissue tropism and individual susceptibility to infection. The importance of glycosylation for S binding to ACE2 is even more emphatically demonstrated by the direct glycan-glycan interactions observed (Figure 7) between S glycans (at N0074 and N0165) and an ACE2 receptor glycan (at N546), adding an additional layer of complexity for interpreting the impact of glycosylation on individual susceptibility.

Several emerging variants of the virus appear to be altering N-linked glycosylation occupancy by disrupting N-linked sequons. Interestingly, the two N-linked sequons in SARS-CoV-2 S directly impacted by variants, N0074 and N0149, are in divergent insert regions 1 and 2, respectively, of SARS-CoV-2 S in comparison with SARS-CoV-1 S (Figure 4A). The N0074, in particular, is one of the S glycans that interact directly with ACE2 glycan (at N546; Figure 7), suggesting that glycan-glycan interactions could contribute to the unique infectivity differences between SARS-CoV-2 and SARS-CoV-1. These sequon variants will also be important to examine in terms of glycan shielding that could influence immunogenicity and efficacy of neutralizing antibodies, as well as interactions with the host cell receptor ACE2. Naturally occurring amino acid-changing SNPs in the ACE2 gene generate a number of variants including one variant, with a frequency of three in 10,000 humans, that eliminates a site of N-linked glycosylation at N546 (Figure 6). Understanding the impact of ACE2 variants on glycosylation and more importantly on S binding, especially for N546S, which impacts the glycan-glycan interaction between S and ACE2 (Figure 7), should be prioritized in light of efforts to

---

Figure 7. Interactions of Glycosylated Soluble Human ACE2 and Glycosylated SARS-CoV-2 S Trimer Immunogen Revealed By 3D-Structural Modeling and MD Simulations

(A) MD simulation of glycosylated soluble human ACE2 and glycosylated SARS-CoV-2 S trimer immunogen interaction (see Videos S5, S6, and S7). ACE2 (top) is colored red with glycans in pink, whereas S is colored white with glycans in dark gray. Highlighted are ACE2 glycans that interact with S that are magnified to the right.

(B) Magnification of ACE2-S interface highlighting ACE2 glycan interactions by using 3D-SNFG icons (Thieker et al., 2016) with S protein (pink) as well as ACE2-S glycan-glycan interactions.

(C) Magnification of dynamics trajectory of glycans at the interface of soluble human ACE2 and S (see Videos S3 and S4).

develop ACE2 as a potential decoy therapeutic. Intelligent manipulation of ACE2 glycosylation could lead to more potent biologics capable of acting as better competitive inhibitors of S binding.

The data presented here, and related similar recent findings (Casalino et al., 2020; Watanabe et al., 2020a; Wrobel et al., 2020), provide a framework to facilitate the production of immunogens, vaccines, antibodies, and inhibitors as well as additional information regarding mechanisms by which glycan microheterogeneity is achieved. However, considerable efforts still remain in order to fully understand the role of glycans in SARS-CoV-2 infection and pathogenicity. Although HEK-expressed S and ACE2 provide a useful window for understanding human glycosylation of these proteins, glycoproteomic characterization after expression in cell lines of more direct relevance to disease and target tissue is sorely needed. Although site occupancy could change depending on presentation and cell type (Struwe et al., 2018), processing of N-linked glycans will almost certainly be altered in a cell-type-dependent fashion. Thus, analyses of the Spike trimer extracted from pseudoviruses, virion-like particles, and ultimately from infectious SARS-CoV-2 virions harvested from airway cells or patients will provide the most accurate view of how trimer immunogens reflect the true glycosylation pattern of the virus. Detailed analyses of the impact of emerging variants in S and natural and designed-for-biologics variants of ACE2 on glycosylation and binding properties are important next steps for developing therapeutics. Finally, it will be important to monitor the slow evolution of the virus to determine if existing sites of glycosylation are lost or new sites emerge with selective pressure that might alter the efficacy of vaccines, neutralizing antibodies, and/or inhibitors.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

REFERENCES

Alijotas-Reig, J., Esteve-Valverde, E., Belizna, C., Selva-O'Callaghan, A., Pardos-Gea, J., Quintana, A., Mekinian, A., Anunciacion-Llunell, A., and Miró-Mur, F. (2020). Immunomodulatory therapy for the management of severe COVID-19. Beyond the anti-viral therapy: A comprehensive review. Autoimmun. Rev. 19, 102569.

Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37, 420–423.

Anumula, K.R., and Taylor, P.B. (1992). A comprehensive procedure for preparation of partially methylated alditol acetates from glycoprotein carbohydrates. Anal. Biochem. 203, 101–108.

Aoki, K., Perlman, M., Lim, J.M., Cantu, R., Wells, L., and Tiemeyer, M. (2007). Dynamic developmental elaboration of N-linked glycan complexity in the Drosophila melanogaster embryo. J. Biol. Chem. 282, 9127–9142.

Aoki, K., Porterfield, M., Lee, S.S., Dong, B., Nguyen, K., McGlamry, K.H., and Tiemeyer, M. (2008). The diversity of O-linked glycans expressed during Drosophila melanogaster development reflects stage- and tissue-specific requirements for cell signaling. J. Biol. Chem. 283, 30385–30400.

Aoki-Kinoshita, K., Agravat, S., Aoki, N.P., Arpinar, S., Cummings, R.D., Fujita, A., Fujita, N., Hart, G.M., Haslam, S.M., Kawasaki, T., et al. (2016). GlyTouCan 1.0–The international glycan structure repository. Nucleic Acids Res. 44 (D1), D1237–D1242.

Behrens, A.J., Harvey, D.J., Milne, E., Cupo, A., Kumar, A., Zitzmann, N., Struwe, W.B., Moore, J.P., and Crispin, M. (2017). Molecular Architecture of the Cleavage-Dependent Mannose Patch on a Soluble HIV-1 Envelope Glycoprotein Trimer. J. Virol. 91, 91.

Beigel, J.H., Tomashek, K.M., Dodd, L.E., Mehta, A.K., Zingman, B.S., Kalil, A.C., Hohmann, E., Chu, H.Y., Luetkemeyer, A., Kline, S., et al. (2020). Remdesivir for the Treatment of Covid-19 - Preliminary Report. N Engl J Med.

Bern, M., Kil, Y.J., and Becker, C. (2012). Byonic: advanced peptide and protein identification software. Curr. Protoc. Bioinformatics Chapter 13, 20, https://doi.org/10.1002/0471250953.bi1320s40.

Beun, R., Kusadasi, N., Sikma, M., Westerink, J., and Huisman, A. (2020). Thromboembolic events and apparent heparin resistance in patients infected with SARS-CoV-2. Int. J. Lab. Hematol. 42 (Suppl 1), 19–20.

Brindley, M.A., and Plemper, R.K. (2010). Blue native PAGE and biomolecular complementation reveal a tetrameric or higher-order oligomer organization of the physiological measles virus attachment protein H. J. Virol. 84, 12174–12184.

Brindley, M.A., Plattet, P., and Plemper, R.K. (2014). Efficient replication of a paramyxovirus independent of full zippering of the fusion protein six-helix bundle domain. Proc. Natl. Acad. Sci. USA 111, E3795–E3804.

Casalino, L., Gaieb, Z., Dommer, A.C., Harbison, A.M., Fogarty, C.A., Barros, E.P., Taylor, B.C., Fadda, E., and Amaro, R.E. (2020). Shielding and Beyond: The Roles of Glycans in SARS-CoV-2 Spike Protein. bioRxiv, 2020.06.11.146522.

Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M.J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25, 1422–1423.

Darden, T.A., and Pedersen, L.G. (1993). Molecular modeling: an experimental tool. Environ. Health Perspect. 101, 410–412.

Dashti-Khavidaki, S., and Khalili, H. (2020). Considerations for Statin Therapy in Patients with COVID-19. Pharmacotherapy 40, 484–486.

Duan, H., Chen, X., Boyington, J.C., Cheng, C., Zhang, Y., Jafari, A.J., Stephens, T., Tsybovsky, Y., Kalyuzhniy, O., Zhao, P., et al. (2018). Glycan Masking Focuses Immune Responses to the HIV-1 CD4-Binding Site and Enhances Elicitation of VRC01-Class Precursor Antibodies. Immunity 49, 301–311.e5.

Epelman, S., Shrestha, K., Troughton, R.W., Francis, G.S., Sen, S., Klein, A.L., and Tang, W.H. (2009). Soluble angiotensin-converting enzyme 2 in human heart failure: relation with myocardial function and clinical outcomes. J. Card. Fail. 15, 565–571.

Escolano, A., Gristick, H.B., Abernathy, M.E., Merkenschlager, J., Gautam, R., Oliveira, T.Y., Pai, J., West, A.P., Jr., Barnes, C.O., Cohen, A.A., et al. (2019). Immunization expands B cells specific to HIV-1 V3 glycan in mice and macaques. Nature 570, 468–473.

Fedson, D.S., Opal, S.M., and Rordam, O.M. (2020). Hiding in Plain Sight: an Approach to Treating Patients with Severe COVID-19 Infection. MBio 11, 11.

Ferreira, R.C., Grant, O.C., Moyo, T., Dorfman, J.R., Woods, R.J., Travers, S.A., and Wood, N.T. (2018). Structural Rearrangements Maintain the Glycan Shield of an HIV-1 Envelope Trimer After the Loss of a Glycan. Sci. Rep. 8, 15031.

Gebrehiwot, A.G., Melka, D.S., Kassaye, Y.M., Rehan, I.F., Rangappa, S., Hinou, H., Kamiyama, T., and Nishimura, S.I. (2018). Healthy human serum N-glycan profiling reveals the influence of ethnic variation on the identified cancer-relevant glycan biomarkers. PLoS One 13, e0209515.

Götz, A.W., Williamson, M.J., Xu, D., Poole, D., Le Grand, S., and Walker, R.C. (2012). Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. J. Chem. Theory Comput. 8, 1542–1555.

Grant, O.C., Xue, X., Ra, D., Khatamian, A., Foley, B.L., and Woods, R.J. (2016). Gly-Spec: a webtool for predicting glycan specificity by integrating glycan array screening data and 3D structure. Glycobiology 26, 1027–1028.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 181, 271–280.e8.

Kahsay, R., Vora, J., Navelkar, R., Mousavi, R., Fochtman, B.C., Holmes, X., Pattabiraman, N., Ranzinger, R., Mahadik, R., Williamson, T., et al. (2020). GlyGen data model and processing workflow. Bioinformatics 36, 3941–3943.

Kirschner, K.N., Yongye, A.B., Tschampel, S.M., González-Outeiriño, J., Daniels, C.R., Foley, B.L., and Woods, R.J. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J. Comput. Chem. 29, 622–655.

Krištić, J., Vučković, F., Menni, C., Klarić, L., Keser, T., Beceheli, I., Pučić-Baković, M., Novokmet, M., Mangino, M., Thaqi, K., et al. (2014). Glycans are a novel biomarker of chronological and biological ages. J. Gerontol. A Biol. Sci. Med. Sci. 69, 779–789.

Lambert, D.W., Yarski, M., Warner, F.J., Thornhill, P., Parkin, E.T., Smith, A.I., Hooper, N.M., and Turner, A.J. (2005). Tumor necrosis factor-alpha convertase (ADAM17) mediates regulated ectodomain shedding of the severe-acute respiratory syndrome-coronavirus (SARS-CoV) receptor, angiotensin-converting enzyme-2 (ACE2). J. Biol. Chem. 280, 30113–30119.

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., and Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215–220.

Lei, C., Qian, K., Li, T., Zhang, S., Fu, W., Ding, M., and Hu, S. (2020). Neutralization of SARS-CoV-2 spike pseudotyped virus by recombinant ACE2-Ig. Nat. Commun. 11, 2070.

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., et al. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 426, 450–454.

Li, W., Zhang, C., Sui, J., Kuhn, J.H., Moore, M.J., Luo, S., Wong, S.K., Huang, I.C., Xu, K., Vasilieva, N., et al. (2005). Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J. 24, 1634–1643.

Liu, M.Q., Zeng, W.F., Fang, P., Cao, W.Q., Liu, C., Yan, G.Q., Zhang, Y., Peng, C., Wu, J.Q., Zhang, X.J., et al. (2017). pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. Nat. Commun. 8, 438.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395, 565–574.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 47 (W1), W636–W641.

Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J. Chem. Theory Comput. 11, 3696–3713.

Marth, J.D., and Grewal, P.K. (2008). Mammalian glycosylation in immunity. Nat. Rev. Immunol. 8, 874–887.

Monteil, V., Kwon, H., Prado, P., Hagelkrüys, A., Wimmer, R.A., Stahl, M., Leopoldi, A., Garreta, E., Hurtado Del Pozo, C., Prosper, F., et al. (2020). Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. Cell 181, 905–913.e7, e907.

Nivedha, A.K., Makeneni, S., Foley, B.L., Tessier, M.B., and Woods, R.J. (2014). Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. J. Comput. Chem. 35, 526–539.

Paal, T., Brindley, M.A., St Clair, C., Prussia, A., Gaus, D., Krumm, S.A., Snyder, J.P., and Plemper, R.K. (2009). Probing the spatial organization of measles virus fusion complexes. J. Virol. 83, 10480–10493.

Palmisano, G., Melo-Braga, M.N., Engholm-Keller, K., Parker, B.L., and Larsen, M.R. (2012). Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. J. Proteome Res. 11, 1949–1957.

Pavić, T., Dilber, D., Kifer, D., Selak, N., Keser, T., Ljubičić, Đ., Vukić Dugac, A., Lauc, G., Rumora, L., and Gornik, O. (2018). N-glycosylation patterns of plasma proteins and immunoglobulin G in chronic obstructive pulmonary disease. J. Transl. Med. 16, 323.

Peng, W., de Vries, R.P., Grant, O.C., Thompson, A.J., McBride, R., Tsogtbaatar, B., Lee, P.S., Razi, N., Wilson, I.A., Woods, R.J., and Paulson, J.C. (2017). Recent H3N2 Viruses Have Evolved Specificity for Extended, Branched Human-type Receptors, Conferring Potential for Increased Avidity. Cell Host Microbe 21, 23–34.

Pinto, D., Park, Y.J., Beltramello, M., Walls, A.C., Tortorici, M.A., Bianchi, S., Jaconi, S., Culap, K., Zatta, F., De Marco, A., et al. (2020). Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. Nature 583, 290–295.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277.

Ritchie, G., Harvey, D.J., Feldmann, F., Stroeher, 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.

Rudman, N., Gornik, O., and Lauc, G. (2019). Altered N-glycosylation profiles as potential biomarkers and drug targets in diabetes. FEBS Lett. 593, 1598–1615.

Ruiz-Canada, C., Kelleher, D.J., and Gilmore, R. (2009). Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. Cell 136, 272–283.

Salomon-Ferrer, R., Götz, A.W., Poole, D., Le Grand, S., and Walker, R.C. (2013). Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J. Chem. Theory Comput. 9, 3878–3888.

Seabright, G.E., Cottrell, C.A., van Gils, M.J., D'addabbo, A., Harvey, D.J., Behrens, A.J., Allen, J.D., Watanabe, Y., Scaringi, N., Polveroni, T.M., et al. (2020). Networks of HIV-1 Envelope Glycans Maintain Antibody Epitopes in the Face of Glycan Additions and Deletions. Structure 28, 897–909.e6.

Shaik, M.M., Peng, H., Lu, J., Rits-Volloch, S., Xu, C., Liao, M., and Chen, B. (2019). Structural basis of coreceptor recognition by HIV-1 envelope spike. Nature 565, 318–323.

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, cwaa042.

Shi, Q., Zhou, Q., Wang, X., Liao, J., Yu, Y., Wang, Z., Lu, S., Ma, Y., Xun, Y., Luo, X., et al.; COVID-19 Evidence and Recommendations Working Group (2020). Potential effectiveness and safety of antiviral agents in children with coronavirus disease 2019: a rapid review and meta-analysis. Ann. Transl. Med. 8, 624.

Stertz, S., Reichelt, M., Spiegel, M., Kuri, T., Martínez-Sobrido, L., García-Sastre, A., Weber, F., and Kochs, G. (2007). The intracellular sites of early replication and budding of SARS-coronavirus. Virology 361, 304–315.

Struwe, W.B., Chertova, E., Allen, J.D., Seabright, G.E., Watanabe, Y., Harvey, D.J., Medina-Ramirez, M., Roser, J.D., Smith, R., Westcott, D., et al. (2018). Site-Specific Glycosylation of Virion-Derived HIV-1 Env Is Mimicked by a Soluble Trimeric Immunogen. Cell Rep. 24, 1958–1966.e5, e1955.

Tang, N., Bai, H., Chen, X., Gong, J., Li, D., and Sun, Z. (2020). Anticoagulant treatment is associated with decreased mortality in severe coronavirus

disease 2019 patients with coagulopathy. J. Thromb. Haemost. 18, 1094–1099.

Thieker, D.F., Hadden, J.A., Schulten, K., and Woods, R.J. (2016). 3D implementation of the symbol nomenclature for graphical representation of glycans. Glycobiology 26, 786–787.

Tiemeyer, M., Aoki, K., Paulson, J., Cummings, R.D., York, W.S., Karlsson, N.G., Lisacek, F., Packer, N.H., Campbell, M.P., Aoki, N.P., et al. (2017). GlyTouCan: an accessible glycan structure repository. Glycobiology 27, 915–919.

Tikellis, C., Bernardi, S., and Burns, W.C. (2011). Angiotensin-converting enzyme 2 is a key modulator of the renin-angiotensin system in cardiovascular and renal disease. Curr. Opin. Nephrol. Hypertens. 20, 62–68.

Ujike, M., and Taguchi, F. (2015). Incorporation of spike and membrane glycoproteins into coronavirus virions. Viruses 7, 1700–1725.

Varki, A. (2017). Biological roles of glycans. Glycobiology 27, 3–49.

Varki, A., Cummings, R.D., Aebi, M., Packer, N.H., Seeberger, P.H., Esko, J.D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., et al. (2015). Symbol Nomenclature for Graphical Representations of Glycans. Glycobiology 25, 1323–1324.

Wang, C.C., Chen, J.R., Tseng, Y.C., Hsu, C.H., Hung, Y.F., Chen, S.W., Chen, C.M., Khoo, K.H., Cheng, T.J., Cheng, Y.S., et al. (2009). Glycans on influenza hemagglutinin affect receptor binding and immune response. Proc. Natl. Acad. Sci. USA 106, 18137–18142.

Wang, S., Voronin, Y., Zhao, P., Ishihara, M., Mehta, N., Porterfield, M., Chen, Y., Bartley, C., Hu, G., Han, D., et al. (2020). Glycan Profiles of gp120 Protein Vaccines from Four Major HIV-1 Subtypes Produced from Different Host Cell Lines under Non-GMP or GMP Conditions. J. Virol. 94, 94.

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.

Watanabe, Y., Allen, J.D., Wrapp, D., McLellan, J.S., and Crispin, M. (2020a). Site-specific glycan analysis of the SARS-CoV-2 spike. Science 369, 330–333.

Watanabe, Y., Berndsen, Z.T., Raghwani, J., Seabright, G.E., Allen, J.D., Pybus, O.G., McLellan, J.S., Wilson, I.A., Bowden, T.A., Ward, A.B., and Crispin, M. (2020b). Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. Nat. Commun. 11, 2688.

Weatherly, D.B., Arpinar, F.S., Porterfield, M., Tiemeyer, M., York, W.S., and Ranzinger, R. (2019). GRITS Toolbox-a freely available software for processing, annotating and archiving glycomics mass spectrometry data. Glycobiology 29, 452–460.

Wells, L., and Hart, G.W.; Athens Guidelines for the Publication of Glycomics Data (2013). Glycomics: building upon proteomics to advance glycosciences. Mol. Cell. Proteomics 12, 833–835.

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.

Wrobel, A.G., Benton, D.J., Xu, P., Roustan, C., Martin, S.R., Rosenthal, P.B., Skehel, J.J., and Gamblin, S.J. (2020). SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. Nat. Struct. Mol. Biol. 27, 763–767.

Xia, X. (2020). Extreme genomic CpG deficiency in SARS-CoV-2 and evasion of host antiviral defense. Mol. Biol. Evol. msaa094.

York, W.S., Mazumder, R., Ranzinger, R., Edwards, N., Kahsay, R., Aoki-Kinoshita, K.F., Campbell, M.P., Cummings, R.D., Feizi, T., Martin, M., et al. (2020). GlyGen: Computational and Informatics Resources for Glycoscience. Glycobiology 30, 72–73.

Yu, W.H., Zhao, P., Draghi, M., Arevalo, C., Karsten, C.B., Suscovich, T.J., Gunn, B., Streeck, H., Brass, A.L., Tiemeyer, M., et al. (2018). Exploiting glycan topography for computational design of Env glycoprotein antigenicity. PLoS Comput. Biol. 14, e1006093.

Yu, J., Tostanoski, L.H., Peter, L., Mercado, N.B., McMahan, K., Mahrokhian, S.H., Nkolola, J.P., Liu, J., Li, Z., Chandrashekar, A., et al. (2020). DNA vaccine protection against SARS-CoV-2 in rhesus macaques. Science, eabc6284.

Zhang, T., Wu, Q., and Zhang, Z. (2020). Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. Curr. Biol. *30*, 1346–1351.e2.

Zhong, N.S., Zheng, B.J., Li, Y.M., Poon, Xie, Z.H., Chan, K.H., Li, P.H., Tan, S.Y., Chang, Q., Xie, J.P., et al. (2003). Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. Lancet *362*, 1353–1358.

Zhou, T., Doria-Rose, N.A., Cheng, C., Stewart-Jones, G.B.E., Chuang, G.Y., Chambers, M., Druz, A., Geng, H., McKee, K., Kwon, Y.D., et al. (2017). Quantification of the Impact of the HIV-1-Glycan Shield on Antibody Elicitation. Cell Rep. *19*, 719–732.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature *579*, 270–273.

Zielinska, D.F., Gnad, F., Wiśniewski, J.R., and Mann, M. (2010). Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. Cell *141*, 897–907.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
SARS-CoV-2 S protein	This Study	N/A
Human ACE2 protein	This Study	N/A
2x Laemmli sample buffer	Bio-Rad	Cat#161-0737
Invitrogen NuPAGE 4 to 12%, Bis-Tris, Mini Protein Gel	Thermo Fisher Scientific	Cat#NP0321PK2
Coomassie Brilliant Blue G-250 Dye	Thermo Fisher Scientific	Cat#20279
Dithiothreitol	Sigma Aldrich	Cat#43815
Iodoacetamide	Sigma Aldrich	Cat#I1149
Trypsin	Promega	Cat#V5111
Lys-C	Promega	Cat#V1671
Arg-C	Promega	Cat#V1881
Glu-C	Promega	Cat#V1651
Asp-N	Promega	Cat#VA1160
Endoglycosidase H	Promega	Cat#V4871
PNGaseF	Promega	Cat#V4831
Chymotrypsin	Athens Research and Technology	Cat#16-19-030820
Alpha lytic protease	New England BioLabs	Cat#P8113
18O water	Cambridge Isotope Laboratories	OLM-782-10-1
O-protease OpeRATOR	Genovis	Cat#G1-OP1-020
Deposited Data		
MS data for site-specific N-linked glycopeptides for SARS-Cov-2 S and human ACE2	This Study	PXD019937
MS data for site-specific O-linked glycopeptides for SARS-Cov-2 S and human ACE2	This Study	PXD019940
MS data for deglycosylated N-linked glycopeptides for SARS-Cov-2 S and human ACE2	This Study	PXD019938
MS data for disulfide bond analysis for SARS-Cov-2 S	This Study	PXD019939
MS data for N-linked glycomics deposited at GlycoPost	This Study	GPST000120
MS data for O-linked glycomics deposited at GlycoPost	This Study	GPST000121
Experimental Models: Cell Lines		
293-F Cells	GIBCO	Cat#R79007
Vero-6 Cells	ATCC	CRL1586
Experimental Models: Organisms/Strains		
VSV(G)-Pseudoviruses	This Study	N/A
Software and Algorithms		
pGlyco v2.2.2	Liu et al., 2017	http://pfind.ict.ac.cn/software/pGlyco/index.html
Proteome Discoverer v1.4	Thermo Fisher Scientific	CAT#OPTON-30945
Byonic v3.8.13	Protein Metrics Inc.	https://www.proteinmetrics.com/products/byonic/
ProteoIQ v2.7	Premier Biosoft (Bern et al., 2012)	http://www.premierbiosoft.com/protein_quantification_software/index.html
GRITS Toolbox V1.1	Weatherly et al., 2019	http://www.grits-toolbox.org/
EMBOSS needle v6.6.0	Rice et al., 2000	https://www.ebi.ac.uk/Tools/psa/emboss_needle/
Biopython v1.76	Cock et al., 2009	https://biopython.org/

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Rpdb v2.3	Julien Ide	https://rdrr.io/cran/Rpdb/
SignalP V5.0	Almagro Armenteros et al., 2019	http://www.cbs.dtu.dk/services/SignalP/
LibreOFFICE Writer v6.4.4.2	The Document Foundation	https://www.libreoffice.org/download/download/
GlyGen V1.5	York et al., 2020	https://www.glygen.org
GNOme V1.5.5	OBO Foundry	https://github.com/glygen-glycan-data/GNOme/blob/master/README.md
GlyTouCan V3.1.0	Aoki-Kinoshita et al., 2016	https://glytoucan.org
Inkscape V1.0	Inkscape project contributors	https://inkscape.org/release/inkscape-1.0/
ffmpeg V3.4	The FFmpeg developers	https://ffmpeg.org/
Cygwin V3.1.5	Cygwin developers	https://www.cygwin.com/

RESOURCE AVAILABILITY

Lead Contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Lance Wells (lwells@ccrc.uga.edu) or alternatively by Peng Zhao (pengzhao@uga.edu).

Materials Availability
This study did not generate new unique reagents.

Data and Code Availability
The mass spectrometry proteomics data are available via ProteomeXchange with identifiers PXD019937, PXD019940, PXD019938, and PXD019939. The mass spectrometry glycomics data are available via GlycoPost with identifiers GPST000120 and GPST000121.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

HEK293-F cells (GIBCO) were maintained and passaged in FreeStyle Media (GIBCO) containing 1% Pen Strep (GIBCO). Vero-6 cells (ATCC) were maintained and passaged in DMEM medium supplemented with 10% fetal bovine serum and 1% Pen Strep (GIBCO) and amphotericin B antibiotics. All cells were maintained at 37°C with 5% CO_2 before and after transfection.

METHOD DETAILS

Expression, Purification, and Characterization of SARS-CoV-2 S and Human ACE2 Proteins
To express a stabilized ectodomain of Spike protein, a synthetic gene encoding residues 1−1208 of SARS-CoV-2 Spike with the furin cleavage site (residues 682–685) replaced by a "GGSG" sequence, proline substitutions at residues 986 and 987, and a foldon trimerization motif followed by a C-terminal 6xHisTag was created and cloned into the mammalian expression vector pCMV-IRES-puro (Codex BioSolutions, Inc, Gaithersburg, MD). The expression construct was transiently transfected in HEK293F cells using polyethylenimine (Polysciences, Inc, Warrington, PA). Protein was purified from cell supernatants using Ni-NTA resin (QIAGEN, Germany), the eluted fractions containing S protein were pooled, concentrated, and further purified by gel filtration chromatography on a Superose 6 column (GE Healthcare). Negative stain electron microscopy (EM) analysis was performed as described (Shaik et al., 2019). Briefly, analysis was performed at room temperature with a magnification of 52,000x and a defocus value of 1.5 μm following low-dose procedures, using a Philips Tecnai F20 electron microscope (Thermo Fisher Scientific) equipped with a Gatan US4000 CCD camera and operated at voltage of 200 kV. The DNA fragment encoding human ACE2 (1-615) with a 6xHis tag at C terminus was synthesized by Genscript and cloned to the vector pCMV-IRES-puro. The expression construct was transfected in HEK293F cells using polyethylenimine. The medium was discarded and replaced with FreeStyle 293 medium after 6-8 h. After incubation in 37°C with 5.5% CO2 for 5 days, the supernatant was collected and loaded to Ni-NTA resin for purification. The elution was concentrated and further purified by a Superdex 200 column.

In-Gel Analysis of SARS-CoV-2 S and Human ACE2 Proteins
A 3.5-μg aliquot of SARS-CoV-2 S protein as well as a 2-μg aliquot of human ACE2 were combined with Laemmli sample buffer, analyzed on a 4%–12% Invitrogen NuPage Bis-Tris gel using the MES pH 6.5 running buffer, and stained with Coomassie Brilliant Blue G-250.

Analysis of N-linked and O-linked Glycans Released from SARS-Cov-2 S and Human ACE2 Proteins

Aliquots of approximately 25-50 μg of S or ACE2 protein were processed for glycan analysis as previously described (Aoki et al., 2007; Aoki et al., 2008). For N-linked glycan analysis, the proteins were digested with trypsin. Following trypsinization, glycopeptides were enriched by C18 Sep-Pak and subjected to PNGaseF digestion to release N-linked glycans. Following PNGaseF digestion, released glycans were separated from residual glycosylated peptides bearing O-linked glycans by C18 Sep-Pak. O-glycosylated peptides were eluted from the Sep-Pak and subjected to reductive β-elimination to release the O-glycans. Another 25-50 μg aliquot of each protein was denatured with SDS and digested with PNGaseF to remove N-linked glycans. The de-N-glycosylated, intact protein was precipitated with cold ethanol and then subjected to reductive β-elimination to release O-glycans. The profiles of O-glycans released from peptides or from intact protein were found to be comparable. N- and O-linked glycans released from glycoproteins were permethylated with methyliodide according to the method of Anumula and Taylor prior to MS analysis (Anumula and Taylor, 1992). Glycan structural analysis was performed using an LTQ-Orbitrap instrument (Orbitrap Discovery, Thermo Fisher Scientific). Detection and relative quantification of the prevalence of individual glycans was accomplished using the total ion mapping (TIM) and neutral loss scan (NL scan) functionality of the Xcalibur software package version 2.0 (Thermo Fisher Scientific) as previously described (Aoki et al., 2007; Aoki et al., 2008). Mass accuracy and detector response was tuned with a permethylated oligosaccharide standard in positive ion mode. For fragmentation by collision-induced dissociation (CID in MS^2 and MS^n), normalized collision energy of 45% was applied. Most permethylated glycans were identified as singly or doubly charged, sodiated species in positive mode. Sulfated N-glycans were detected as singly or doubly charged, deprotonated species in negative ion mode. Peaks for all charge states were deconvoluted by the charge state and summed for quantification. All spectra were manually interpreted and annotated. The explicit identities of individual monosaccharide residues have been assigned based on known human biosynthetic pathways. Graphical representations of monosaccharide residues are consistent with the Symbol Nomenclature for Glycans (SNFG), which has been broadly adopted by the glycomics community (Varki et al., 2015). The MS-based glycomics data generated in these analyses and the associated annotations are presented in accordance with the MIRAGE standards and the Athens Guidelines (Wells et al., 2013). Data annotation and assignment of glycan accession identifiers were facilitated by GRITS Toolbox, GlyTouCan, GNOme, and GlyGen (Kahsay et al., 2020; Tiemeyer et al., 2017; Weatherly et al., 2019; York et al., 2020).

Analysis of Disulfide Bonds for SARS-Cov-2 S Protein by LC-MS

Two 10-μg aliquots of SARS-CoV-2 S protein were denatured by incubating with 20% acetonitrile at room temperature and alkylated by 13.75 mM of iodoacetamide at room temperature in dark. The two aliquots of proteins were then digested respectively using alpha lytic protease, or a combination of trypsin, Lys-C and Glu-C. Following digestion, the proteins were deglycosylated by PNGaseF treatment. The resulting peptides were separated on an Acclaim PepMap RSLC C18 column (75 μm x 15 cm) and eluted into the nano-electrospray ion source of an Orbitrap Fusion Lumos Tribrid mass spectrometer at a flow rate of 200 nL/min. The elution gradient consists of 1%–40% acetonitrile in 0.1% formic acid over 370 min followed by 10 min of 80% acetonitrile in 0.1% formic acid. The spray voltage was set to 2.2 kV and the temperature of the heated capillary was set to 280°C. Full MS scans were acquired from m/z 200 to 2000 at 60k resolution, and MS/MS scans following electron transfer dissociation (ETD) were collected in the Orbitrap at 15k resolution. The raw spectra were analyzed by Byonic (v3.8.13, Protein Metrics Inc.) with mass tolerance set as 20 ppm for both precursors and fragments. The search output was filtered at 1% false discovery rate and 10 ppm mass error. The spectra assigned as cross-linked peptides were manually evaluated for Cys0015.

Analysis of Site-Specific N-linked Glycopeptides for SARS-Cov-2 S and Human ACE2 Proteins by LC-MS

Four 3.5-μg aliquots of SARS-CoV-2 S protein were reduced by incubating with 10 mM of dithiothreitol at 56°C and alkylated by 27.5 mM of iodoacetamide at room temperature in dark. The four aliquots of proteins were then digested respectively using alpha lytic protease, chymotrypsin, a combination of trypsin and Glu-C, or a combination of Glu-C and AspN. Three 10-μg aliquots of ACE2 protein were reduced by incubating with 5 mM of dithiothreitol at 56°C and alkylated by 13.75 mM of iodoacetamide at room temperature in dark. The three aliquots of proteins were then digested respectively using alpha lytic protease, chymotrypsin, or a combination of trypsin and Lys-C. The resulting peptides were separated on an Acclaim PepMap RSLC C18 column (75 μm x 15 cm) and eluted into the nano-electrospray ion source of an Orbitrap Fusion Lumos Tribrid mass spectrometer at a flow rate of 200 nL/min. The elution gradient consists of 1%–40% acetonitrile in 0.1% formic acid over 370 min followed by 10 min of 80% acetonitrile in 0.1% formic acid. The spray voltage was set to 2.2 kV and the temperature of the heated capillary was set to 280°C. Full MS scans were acquired from m/z 200 to 2000 at 60k resolution, and MS/MS scans following higher-energy collisional dissociation (HCD) with stepped collision energy (15%, 25%, 35%) were collected in the Orbitrap at 15k resolution. pGlyco v2.2.2 (Liu et al., 2017) was used for database searches with mass tolerance set as 20 ppm for both precursors and fragments. The database search output was filtered to reach a 1% false discovery rate for glycans and 10% for peptides. Quantitation was performed by calculating spectral counts for each glycan composition at each site. Any N-linked glycan compositions identified by only one spectra were removed from quantitation. N-linked glycan compositions were categorized into 22 classes (including Unoccupied): HexNAc(2)Hex(9∼5)Fuc(0∼1) was classified as M9 to M5 respectively; HexNAc(2)Hex(4∼1)Fuc(0∼1) was classified as M1-M4; HexNAc(3∼6)Hex(5∼9)Fuc(0)NeuAc(0∼1) was classified as Hybrid with HexNAc(3∼6)Hex(5∼9)Fuc(1∼2)NeuAc(0∼1) classified as F-Hybrid; Complex-type glycans are classified based on the number of antenna, fucosylation, and sulfation: HexNAc(3)Hex(3∼4)Fuc(0)NeuAc(0∼1) is assigned as A1 with HexNAc(3)Hex(3∼4)Fuc(1∼2)NeuAc(0∼1) assigned as F-A1; HexNAc(4)Hex(3∼5)Fuc(0)NeuAc(0∼2) is assigned as A2/A1B with HexNAc(4)Hex(3∼5)Fuc(1∼5)NeuAc(0∼2) assigned as F-A2/A1B; HexNAc(5)Hex(3∼6)Fuc(0)NeuAc(0∼3) is

assigned as A3/A2B with HexNAc(5)Hex(3∼6)Fuc(1∼3)NeuAc(0∼3) assigned as F-A3/A2B; HexNAc(6)Hex(3∼7)Fuc(0)NeuAc(0∼4) is assigned as A4/A3B with HexNAc(6)Hex(3∼7)Fuc(1∼3)NeuAc(0∼4) assigned as F-A4/A3B; HexNAc(7)Hex(3∼8)Fuc(0)NeuAc(0∼1) is assigned as A5/A4B with HexNAc(7)Hex(3∼8)Fuc(1∼3)NeuAc(0∼1) as F-A5/A4B; HexNAc(8)Hex(3∼9)Fuc(0) is assigned as A6/A5B with HexNAc(8)Hex(3∼9)Fuc(1) assigned as F-A6/A5B; any glycans identified with a sulfate are assigned as Sulfated.

Analysis of Deglycosylated SARS-Cov-2 S and Human ACE2 Proteins by LC-MS

Three 3.5-μg aliquots of SARS-CoV-2 S protein were reduced by incubating with 10 mM of dithiothreitol at 56°C and alkylated by 27.5 mM of iodoacetamide at room temperature in dark. The three aliquots were then digested respectively using chymotrypsin, Asp-N, or a combination of trypsin and Glu-C. Two 10-μg aliquots of ACE2 protein were reduced by incubating with 5 mM of dithiothreitol at 56°C and alkylated by 13.75 mM of iodoacetamide at room temperature in dark. The two aliquots were then digested respectively using chymotrypsin, or a combination of trypsin and Lys-C. Following digestion, the proteins were deglycosylated by Endoglycosidase H followed by PNGaseF treatment in the presence of 18O water. The resulting peptides were separated on an Acclaim PepMap RSLC C18 column (75 μm x 15 cm) and eluted into the nano-electrospray ion source of an Orbitrap Fusion Lumos Tribrid mass spectrometer at a flow rate of 200 nL/min. The elution gradient consists of 1%–40% acetonitrile in 0.1% formic acid over 370 min followed by 10 min of 80% acetonitrile in 0.1% formic acid. The spray voltage was set to 2.2 kV and the temperature of the heated capillary was set to 280°C. Full MS scans were acquired from m/z 200 to 2000 at 60k resolution, and MS/MS scans following collision-induced dissociation (CID) at 38% collision energy were collected in the ion trap. The spectra were analyzed using SEQUEST (Proteome Discoverer 1.4) with mass tolerance set as 20 ppm for precursors and 0.5 Da for fragments. The search output was filtered using ProteoIQ (v2.7) to reach a 1% false discovery rate at protein level and 10% at peptide level. Occupancy of each N-linked glycosylation site was calculated using spectral counts assigned to the 18O-Asp-containing (PNGaseF-cleaved) and/or HexNAc-modified (EndoH-cleaved) peptides and their unmodified counterparts.

Analysis of Site-Specific O-linked Glycopeptides for SARS-Cov-2 S and Human ACE2 Proteins by LC-MS

Three 10-μg aliquots of SARS-CoV-2 S protein and one 10-μg aliquot of ACE2 protein were reduced by incubating with 5 mM of dithiothreitol at 56°C and alkylated by 13.75 mM of iodoacetamide at room temperature in dark. The four aliquots were then digested respectively using trypsin, Lys-C, Arg-C, or a combination of trypsin and Lys-C. Following digestion, the proteins were deglycosylated by PNGaseF treatment and then digested with O-protease OpeRATOR®. The resulting peptides were separated on an Acclaim PepMap RSLC C18 column (75 μm x 15 cm) and eluted into the nano-electrospray ion source of an Orbitrap Fusion Lumos Tribrid mass spectrometer at a flow rate of 200 nL/min. The elution gradient consists of 1%–40% acetonitrile in 0.1% formic acid over 370 min followed by 10 min of 80% acetonitrile in 0.1% formic acid. The spray voltage was set to 2.2 kV and the temperature of the heated capillary was set to 280°C. Full MS scans were acquired from m/z 200 to 2000 at 60k resolution, and MS/MS scans following higher-energy collisional dissociation (HCD) with stepped collision energy (15%, 25%, 35%) or electron transfer dissociation (ETD) were collected in the Orbitrap at 15k resolution. The raw spectra were analyzed by Byonic (v3.8.13) with mass tolerance set as 20 ppm for both precursors and fragments. MS/MS filtering was applied to only allow for spectra where the oxonium ions of HexNAc were observed. The search output was filtered at 1% false discovery rate and 10 ppm mass error. The spectra assigned as O-linked glycopeptides were manually evaluated. Quantitation was performed by calculating spectral counts for each glycan composition at each site. Any O-linked glycan compositions identified by only one spectra were removed from quantitation. Occupancy of each O-linked glycosylation site was calculated using spectral counts assigned to any glycosylated peptides and their unmodified counterparts from searches without MS/MS filtering.

Sequence Analysis of SARS-CoV-2 S and Human ACE2 Proteins

The genomes of SARS-CoV as well as bat and pangolin coronavirus sequences reported to be closely related to SARS-CoV-2 were downloaded from NCBI. The S protein sequences from all of those genomes were aligned using EMBOSS needle v6.6.0 (Rice et al., 2000) via the EMBL-EBI provided web service (Madeira et al., 2019). Manual analysis was performed in the regions containing canonical N-glycosylation sequons (N-X-S/T). For further sequence analysis of SARS-CoV-2 S variants, the genomes of SARS-CoV-2 were downloaded from NCBI and GISAID and further processed using Biopython 1.76 to extract all sequences annotated as ''surface glycoprotein'' and to remove any incomplete sequence as well as any sequence containing unassigned amino acids. For sequence analysis of human ACE2 variants, the single nucleotide polymorphisms (SNPs) of ACE2 were extracted from the NCBI dbSNP database and filtered for missense mutation entries with a reported minor allele frequency. Manual analysis was performed on both SARS-CoV-2 S and human ACE2 variants to further examine the regions containing canonical N-glycosylation sequons (N-X-S/T). LibreOffice Writer and its macro capabilities was used to shade regions on the linear sequence of S and ACE2.

3D Structural Modeling and Molecular Dynamics Simulation of Glycosylated SARS-CoV-2 S and Human ACE2 Proteins

SARS-CoV-2 Spike (S) protein structure and ACE2 co-complex – A 3D structure of the prefusion form of the S protein (RefSeq: YP_009724390.1, UniProt: P0DTC2 SPIKE_SARS2), based on a Cryo-EM structure (PDB code 6VSB) (Wrapp et al., 2020), was obtained from the SWISS-MODEL server (swissmodel.expasy.org). The model has 95% coverage (residues 27 to 1146) of the S protein. The receptor binding domain (RBD) in the ''open'' conformation was replaced with the RBD from an ACE2 co-complex (PDB code 6M0J) by grafting residues C336 to V524.

Glycoform generation – Glycans (detected by glycomics) were selected for installation on glycosylated S and ACE2 sequons (detected by glycoproteomics) based on three sets of criteria designed to reasonably capture different aspects of glycosylation microheterogeneity. We denote the first of these glycoform models as "Abundance." The glycans selected for installation to generate the Abundance model were chosen because they were identified as the most abundant glycan structure (detected by glycomics) that matched the most abundant glycan composition (detected by glycoproteomics) at each individual site. We denote the second glycoform model as "Oxford Class." The glycans selected for installation to generate the Oxford Class model were chosen because they were the most abundant glycan structure, (detected by glycomics) that was contained within the most highly represented Oxford classification group (detected by glycoproteomics) at each individual site (Figure S7; Tables S1 and S8). Finally, we denote the third glycoform model as "Processed." The glycans selected for installation to generate the Processed model were chosen because they were the most highly trimmed, elaborated, or terminally decorated structure (detected by glycomics) that corresponded to a composition (detected by glycoproteomics) which was present at $\geq 1/3^{rd}$ of the abundance of the most highly represented composition at each site (Table S1). 3D structures of the three glycoforms (Abundance, Oxford Class, Processed) were generated for the SARS-CoV-2 S protein alone, and in complex with the glycosylated ACE2 protein. The glycoprotein builder available at GLYCAM-Web (www.glycam.org) was employed together with an in-house program that adjusts the asparagine side chain torsion angles and glycosidic linkages within known low-energy ranges (Nivedha et al., 2014) to relieve any atomic overlaps with the core protein, as described previously (Grant et al., 2016; Peng et al., 2017).

Energy minimization and Molecular dynamics (MD) simulations – Each glycosylated structure was placed in a periodic box of TIP3P water molecules with a 10 Å buffer between the solute and the box edge. Energy minimization of all atoms was performed for 20,000 steps (10,000 steepest decent, followed by 10,000 conjugant gradient) under constant pressure (1 atm) and temperature (300 K) conditions. All MD simulations were performed under nPT conditions with the CUDA implementation of the PMEMD (Götz et al., 2012; Salomon-Ferrer et al., 2013) simulation code, as present in the Amber14 software suite (University of California, San Diego). The GLYCAM06j force field (Kirschner et al., 2008) and Amber14SB force field (Maier et al., 2015) were employed for the carbohydrate and protein moieties, respectively. A Berendsen barostat with a time constant of 1 ps was employed for pressure regulation, while a Langevin thermostat with a collision frequency of 2 ps⁻¹ was employed for temperature regulation. A nonbonded interaction cut-off of 8 Å was employed. Long-range electrostatics were treated with the particle-mesh Ewald (PME) method (Darden and Pedersen, 1993). Covalent bonds involving hydrogen were constrained with the SHAKE algorithm, allowing an integration time step of 2 fs to be employed. The energy minimized coordinates were equilibrated at 300K over 400 ps with restraints on the solute heavy atoms. Each system was then equilibrated with restraints on the Ca atoms of the protein for 1ns, prior to initiating 4 independent 250 ns production MD simulations with random starting seeds for a total time of 1 μs per system, with no restraints applied.

Antigenic surface analysis. A series of 3D structure snapshots of the simulation were taken at 1 ns intervals and analyzed in terms of their ability to interact with a spherical probe based on the average size of hypervariable loops present in an antibody complementarity determining region (CDR), as described recently (https://www.biorxiv.org/content/10.1101/2020.04.07.030445v2). The percentage of simulation time each residue was exposed to the AbASA probe was calculated and plotted onto both the 3D structure and primary sequence.

Analysis of SARS-CoV-2 Spike VSV Pseudoparticles (ppVSV-SARS-2-S)

293T cells were transfected with an expression plasmid encoding SARS-CoV-2 Spike (pcDNAintron-SARS-2-SΔ19). To increase cell surface expression, the last 19 amino acids containing the Golgi retention signal were removed. Two SΔ19 constructs were compared, one started with Met1 and the other with Met2. Twenty-four h following transfection, cells were transduced with ppVSVΔG-VSV-G (particles that were pseudotyped with VSV-G in *trans*). One h following transduction cells were extensively washed and media was replaced. Supernatant containing particles were collected 12-24 h following transduction and cleared through centrifugation. Cleared supernatant was frozen at −80°C for future use. Target cells Vero E6 were seeded in 24-well plates (5x10⁵ cells/mL) at a density of 80% coverage. The following day, ppVSV-SARS-2-S/GFP particles were transduced into target cells for 60 min, particles pseudotyped with VSV-G, Lassa virus GP, or no glycoprotein were included as controls. 24 h following transduction, transduced cells were released from the plate with trypsin, fixed with 4% formaldehyde, and GFP-positive virus-transduced cells were quantified using flow cytometry (Bectin Dickson BD-LSRII). To quantify the ability of various SARS-CoV-2 S mutants to mediate fusion, effector cells (HEK293T) were transiently transfected with the indicated pcDNAintron-SARS-2-S expression vector or measles virus H and F (Brindley et al., 2014). Effector cells were infected with MVA-T7 four h following transduction to produce the T7 polymerase (Paal et al., 2009). Target cells naturally expressing the receptor ACE2 (Vero) or ACE2 negative cells (HEK293T) were transfected with pTM1-luciferase, which encodes for firefly luciferase under the control of a T7 promoter (Brindley and Plemper, 2010). 24 h following transfection, the target cells were lifted and added to the effector cells at a 1:1 ratio. 4 h following co-cultivation, cells were washed, lysed and luciferase levels were quantified using Promega's Steady-Glo substrate. To visualize cell-to-cell fusion, Vero cells were co-transfected with pGFP and the pcDNAintron-SARS-2-S constructs. 24 h following transfection, syncytia was visualized by fluorescence microscopy.

QUANTIFICATION AND STATISTICAL ANALYSIS

Raw glycoproteomic data from the mass spectrometers was searched using Proteome Discoverer v1.4 (SEQUEST), Protein Metrics Inc. Byonic v3.8.13, and pGlyco v2.2.2. For data searches using Proteome Discoverer, the results were processed to apply false

discovery rate filtering using ProteoIQ v2.7. For the deglycosylated protein work, search results from SEQUEST were filtered in ProteoIQ with a 1% false discovery rate at the protein level and 10% at the peptide level. For N-linked glycopeptide analysis, pGlyco was used with false discovery rate of 1% at the glycan level and 10% at the peptide level. For disulfide bond analysis and O-glycopeptide searches, Byonic was used and the false discovery rate was set to 1%. All mass spectrometry results were manually curated.

Antigen accessibility simulations were carried out as described in the Method Details section and the mean of four simulations (three of length 350ns, one of length 200ns; amounting to 1.25 μs of total molecular dynamics simulation time) were utilized. Glycan-glycan and glycan-peptide interactions were also calculated based on simulations as a percentage of time residues were in contact and averaged (mean) to produce the corresponding supplemental (colored) sequence figures with the raw numbers for coloring present also in each corresponding supplemental table tab. 3D distances were computed using Rpdb as described in more detail in the Method Details section. This data is presented using box & whisker plots with all underlying statistics calculated in R.