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

Peng Zhao[1,#], Jeremy L. Praissman[1,#], Oliver C. Grant[1,#], 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*]

[1] Complex Carbohydrate Research Center, Department of Biochemistry and Molecular Biology, and Department of Chemistry, University of Georgia, Athens, Georgia, 30602, USA

[2] Division of Molecular Medicine, Children's Hospital and Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, 02115, USA

[3] Departments of Pediatrics and Bioengineering, University of California, San Diego, La Jolla, California, 92093, USA

[4] Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, 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, California, 92093, USA

[#] Authors contributed equally

* To whom correspondence should be addressed: Robert J. Woods, rwoods@ccrc.uga.edu , and/or Lance Wells, lwells@ccrc.uga.edu

SUMMARY

The current COVID-19 pandemic is caused by the SARS-CoV-2 betacoronavirus, which utilizes its highly glycosylated trimeric Spike protein to bind to the cell surface receptor 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 bioinformatic analyses of natural variants and with existing 3D-structures of both glycoproteins to generate molecular dynamics simulations of each glycoprotein 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 that, taken together, can facilitate immunogen design to achieve antibody neutralization and inform therapeutic strategies to inhibit viral infection.

Keywords: SARS-CoV-2; COVID-19; Spike protein; coronavirus; ACE2; glycoprotein; glycosylation; mass spectrometry; molecular dynamics; 3D-modeling

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 of 2019 (1). 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 20th, 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 (2,3). 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 (1,4,5). Given the rapid appearance and spread of this virus, there is no current validated vaccine or SARS-CoV-2-specific targeting therapy clinically approved although statins, heparin, and steroids look promising for lowering fatality rates and antivirals likely reduce the duration of symptomatic disease presentation (6-12).

SARS-CoV-2, like SARS-CoV-1, utilizes the host angiotensin converting enzyme II (ACE2) for binding and entry into host cells (13,14). 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 (15). 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 (15,16). In similar fashion, wildtype or catalytically-impaired ACE2 has also been investigated as a potential therapeutic biologic that might interfere with the infection cycle of ACE2 targeting coronaviruses (17,18). 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.

82 The SARS-CoV-2 Spike glycoprotein consists of two subunits, a receptor binding subunit (S1)
83 and a membrane fusion subunit (S2) (1,2). The Spike glycoprotein assembles into stable
84 homotrimers that together possess 66 canonical sequons for N-linked glycosylation (N-X-S/T,
85 where X is any amino acid except P) as well as a number of potential O-linked glycosylation sites
86 (19,20). Interestingly, coronaviruses virions bud into the lumen of the endoplasmic reticulum-
87 Golgi intermediate compartment, ERGIC, raising unanswered questions regarding the precise
88 mechanisms by which viral surface glycoproteins are processed as they traverse the secretory
89 pathway (21,22). While this and similar studies (19,23) analyze recombinant proteins, a previous
90 study on SARS-CoV-1 suggest that glycosylation of the Spike can be impacted by this
91 intracellular budding and remains to be investigated in SARS-CoV-2 (24). Nonetheless, it has
92 been proposed that this virus, and others, acquires a glycan coat sufficient and similar enough to
93 endogenous host protein glycosylation that it serves as a glycan shield, facilitating immune
94 evasion by masking non-self viral peptides with self-glycans (15,20-22). In parallel with their
95 potential masking functions, glycan-dependent epitopes can elicit specific, even neutralizing,
96 antibody responses, as has been described for HIV-1 ((15,25-29),
97 https://www.biorxiv.org/content/10.1101/2020.06.30.178897v1). Thus, understanding the
98 glycosylation of the viral Spike trimer is fundamental for the development of efficacious vaccines,
99 neutralizing antibodies, and therapeutic inhibitors of infection.
100 ACE2 is an integral membrane metalloproteinase that regulates the renin-angiotensin system
101 (30). Both SARS-CoV-1 and SARS-CoV-2 have co-opted ACE2 to function as the receptor by
102 which these viruses attach and fuse with host cells (13,14). ACE2 is cleavable by ADAM
103 proteases at the cell surface (31), resulting in the shedding of a soluble ectodomain which can be
104 detected in apical secretions of various epithelial layers (gastric, airway, etc.) and in serum (32).
105 The N-terminal extracellular domain of ACE2 contains 6 canonical sequons for N-linked
106 glycosylation and several potential O-linked sites. Several nonsynonymous single-nucleotide
107 polymorphisms (SNPs) in the ACE2 gene have been identified in the human population and could
108 potentially alter ACE2 glycosylation and/or affinity of the receptor for the viral Spike protein (33).

109 Given that glycosylation can affect the half-life of circulating glycoproteins in addition to
110 modulating the affinity of their interactions with receptors and immune/inflammatory signaling
111 pathways (34,35), understanding the impact of glycosylation of ACE2 with respect to its binding
112 of SARS-CoV-2 Spike glycoprotein is of high importance. The proposed use of soluble
113 extracellular domains of ACE2 as decoy, competitive inhibitors for SARS-CoV-2 infection
114 emphasizes the critical need for understanding the glycosylation profile of ACE2 so that optimally
115 active biologics can be produced (17,18).
116 To accomplish the task of characterizing site-specific glycosylation of the trimer Spike of SARS-
117 CoV-2 and the host receptor ACE2, we began by expressing and purifying a stabilized, soluble
118 trimer Spike glycoprotein mimetic immunogen (that we define here and forward as S, (36)) and a
119 soluble version of the ACE2 glycoprotein from a human cell line. We utilized multiple mass
120 spectrometry-based approaches, including glycomic and glycoproteomic approaches, to
121 determine occupancy and site-specific heterogeneity of N-linked glycans. Occupancy (i.e. the
122 percent of any given residue being modified by a glycan) is an important consideration when
123 developing neutralizing antibodies against a glycan-dependent epitope. We also identified sites
124 of O-linked glycosylation and the heterogeneity of the O-linked glycans on S and ACE2. We
125 leveraged this rich dataset, along with existing 3D-structures of both glycoproteins, to generate
126 static and molecular dynamics models of S alone, and in complex with the glycosylated, soluble
127 ACE 2 receptor. By combining bioinformatic characterization of viral evolution and variants of the
128 Spike and ACE2 with molecular dynamics simulations of the glycosylated Spike-ACE2 interaction,
129 we identified important roles for glycans in multiple processes, including receptor-viral binding
130 and glycan-shielding of the Spike. Our rich characterization of the recombinant, glycosylated
131 Spike trimer mimetic immunogen of SARS-CoV-2 in complex with the soluble human ACE2
132 receptor provides a detailed platform for guiding rational vaccine, antibody, and inhibitor design.
133
134 **RESULTS**

135 *Expression, Purification, and Characterization of SARS-CoV-2 Spike Glycoprotein Trimer and*
136 *Soluble Human ACE2*
137 A trimer-stabilized, soluble variant of the SARS-CoV-2 Spike protein (S) that contains 22
138 canonical N-linked glycosylation sequons per protomer and a soluble version of human ACE2
139 that contains 6, lacking the most C-terminal 7th, canonical N-linked glycosylation sequons (**Fig.**
140 **1A**) were purified from the media of transfected HEK293 cells and the quaternary structure
141 confirmed by negative EM staining for the S trimer (**Fig. 1B**) and purity examined by SDS-PAGE
142 Coomassie G-250 stained gels for both (**Fig. 1C**). In addition, proteolytic digestions followed by
143 proteomic analyses confirmed that the proteins were highly purified (**Supplemental Table, Tab**
144 **12**). Finally, the N-terminus of both the mature S and the soluble mature ACE2 were empirically
145 determined via proteolytic digestions and LC-MS/MS analyses. These results confirmed that both
146 the secreted, mature forms of S protein and ACE2 begin with an N-terminal glutamine that has
147 undergone condensation to form pyroglutamine at residue 14 and 18, respectively (**Figs. 1D and**
148 **S1**). The N-terminal peptide observed for S also contains a glycan at Asn-0017 (**Fig. 1D**) and
149 mass spectrometry analysis of non-reducing proteolytic digestions confirmed that Cys-0015 of S
150 is in a disulfide linkage with Cys-0136 (**Fig. S2, Supplemental Table, Tab 2**). Given that SignalP
151 (37) predicts signal sequence cleavage between Cys-0015 and Val-0016 but we observed
152 cleavage between Ser-0013 and Gln-0014, we examined the possibility that an in-frame upstream
153 Methionine to the proposed start Methionine (**Fig. 1A**) might be used to initiate translation (**Fig.**
154 **S3**). If one examines the predicted signal sequence cleavage using the in-frame Met that is
155 encoded 9 amino acids upstream, SignalP now predicts cleavage between the Ser and Gln that
156 we observed in our studies (**Fig. S3**). To examine whether this impacted S expression, we
157 expressed constructs that contained or did not contain the upstream 27 nucleotides in a
158 pseudovirus (VSV) system expressing SARS-CoV2 S (**Fig. S4**) and in our HEK293 system (data
159 not shown). Both expression systems produced a similar amount of S regardless of which
160 expression construct was utilized (**Fig. S4**). Thus, while the translation initiation start site has still

161 not been fully defined, allowing for earlier translation in expression construct design did not have
162 a significant impact on the generation of S.

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

165 We utilized multiple approaches to examine glycosylation of the SARS-CoV-2 S trimer. First, the
166 portfolio of glycans linked to SARS-CoV-2 S trimer immunogen was analyzed following their
167 release from the polypeptide backbone. N-glycans were released from protein by treatment with
168 PNGase F and O-glycans were subsequently released by beta-elimination. Following
169 permethylation to enhance detection sensitivity and structural characterization, released glycans
170 were analyzed by multi-stage mass spectrometry (MS^n) (38,39). Mass spectra were processed
171 by GRITS Toolbox and the resulting annotations were validated manually (40). Glycan
172 assignments were grouped by type and by additional structural features for relative quantification
173 of profile characteristics (**Fig. 2A**, **Supplemental Table, Tab 3**). This analysis quantified 49 N-
174 glycans and revealed that 55% of the total glycan abundance was of the complex type, 17% was
175 of the hybrid type, and 28% was high-mannose. Among the complex and hybrid N-glycans, we
176 observed a high degree of core fucosylation and significant abundance of bisected and LacDiNAc
177 structures. We also observed sulfated N-linked glycans using negative mode MS^n analyses
178 (**Supplemental Table, Tab 13**) though signal intensity was too low in positive ion mode (at least
179 10-fold lower than any of the non-sulfated glycans) for accurate quantification. In addition, we
180 detected 15 O-glycans released from the S trimer (**Fig. S5, Supplemental Table, Tab 4**).

181 To determine occupancy of N-linked glycans at each site, we employed a sequential
182 deglycoslyation approach using Endoglycosidase H and PNGase F in the presence of ^{18}O-H_2O
183 following tryptic digestion of S (28,41). Following LC-MS/MS analyses, the resulting data
184 confirmed that 19 of the canonical sequons had occupancies greater than 95% (**Supplemental**
185 **Table, Tab 5**). One canonical sequence, N0149, had insufficient spectral counts for quantification
186 by this method but subsequent analyses described below suggested high occupancy. The 2 most
187 C-terminal N-linked sites, N1173 and N1194, had reduced occupancy, 52% and 82% respectively.

188 Reduced occupancy at these sites may reflect hindered en-bloc transfer by the
189 oligosaccharyltransferase (OST) due to primary amino acid sequences at or near the N-linked
190 sequon. Alternatively, this may reflect these two sites being post-translationally modified after
191 release of the protein by the ribosome by a less efficient STT3B-containing OST, either due to
192 activity or initial folding of the polypeptide, as opposed to co-translationally modified by the
193 STT3A-containing OST (42). None of the non-canoncial sequons (3 N-X-C sites and 4 N-G-L/I/V
194 sites, (43)) showed significant occupancy (>5%) except for N0501 that showed moderate (19%)
195 conversion to ^{18}O-Asp that could be due to deamidation that is facilitated by glycine at the +1
196 position (**Supplemental Table, Tab 5,** (44)). Further analysis of this site (see below) by direct
197 glycopeptide analyses allowed us to determine that N0501 undergoes deamidation but is not
198 glycosylated. Thus, all, and only the, 22 canonical sequences for N-linked glycosylation (N-X-S/T)
199 are utilized with only N1173 and N1194 demonstrating occupancies below 95%.
200 Next, we applied 3 different proteolytic digestion strategies to the SARS-CoV-2 S immunogen to
201 maximize glycopeptide coverage by subsequent LC-MS/MS analyses. Extended gradient
202 nanoflow reverse-phase LC-MS/MS was carried out on a ThermoFisher Lumos™ Tribrid™
203 instrument using Step-HCD fragmentation on each of the samples (see STAR methods for details,
204 (25,26,28,41,45)). Following data analyses using pGlyco 2.2.2 (46), Byonic (47), and manual
205 validation of glycan compositions against our released glycomics findings (**Fig. 2A**,
206 **Supplemental Table, Tab 3 and 13**), we were able to determine the microheterogeneity at each
207 of the 22 canonical sites (**Fig. 2B-2E, Supplemental Table, Tab 6**). Notably, none of the non-
208 canonical consensus sequences, including N0501, displayed any quantifiable glycans. The N-
209 glycosites N0074 (**Fig. 2B**) and N0149 (**Fig. 2C**) are highly processed and display a typical
210 mammalian N-glycan profile. N0149 is, however, modified with several hybrid N-glycan structures
211 while N0074 is not. N0234 (**Fig. 2D**) and N0801 (**Fig. 2E**) have N-glycan profiles more similar to
212 those found on other viruses such as HIV (15) that are dominated by high-mannose structures.
213 N0234 (**Fig. 2D**) displays an abundance of Man7 - Man9 high-mannose structures suggesting
214 stalled processing by early acting ER and cis-Golgi mannosidases. In contrast, N0801 (**Fig. 2E**)

215 is processed more efficiently to Man5 high-mannose and hybrid structures suggesting that access
216 to the glycan at this site by MGAT1 and α-Mannosidase II is hindered. In general, for all 22 sites
217 (**Fig. 2B-2E, Supplemental Table, Tab 6**), we observed under processing of complex glycan
218 antennae (i.e. under-galactosylation and under-sialylation) and a high degree of core fucosylation
219 in agreement with released glycan analyses (**Fig. 2A**, **Supplemental Table, Tab 3**). We also
220 observed a small percent of sulfated N-linked glycans at several sites (**Supplemental Table, Tab**
221 **6 and 8**). Based on the assignments and the spectral counts for each topology, we were able to
222 determine the percent of total N-linked glycan types (high-mannose, hybrid, or complex) present
223 at each site (**Figure 3, Supplemental Table, Tab 7**). Notably, 3 of the sites (N0234, N0709, and
224 N0717) displayed more than 50% high-mannose glycans while 11 other sites (N0017, N0074,
225 N0149, N0165, N0282, N0331, N0657, N1134, N1158, N1173, and N1194) were more than 90%
226 complex when occupied. The other 8 sites were distributed between these 2 extremes. Notably,
227 only 1 site (N0717 at 45%), which also had greater than 50% high-mannose (55%), had greater
228 than 33% hybrid structures. To further evaluate the heterogeneity, we grouped all the topologies
229 into the 20 classes recently described by the Crispin laboratory with adding 2 categories (sulfated
230 and unoccupied) that we refer to here as the Oxford classification (**Supplemental Table, Tab 8,**
231 (19)). Among other features observed, this classification allowed us to observe that while most
232 sites with high mannose structures were dominated by the Man5GlcNAc2 structure, N0234 and
233 N0717 were dominated by the higher Man structures of Man8GlcNAc2 and Man7GlcNAc2,
234 respectively (**Fig. S7, Supplemental Table, Tab 8**). Limited processing at N0234 is in agreement
235 with a recent report suggesting that high mannose structures at this site help to stabilize the
236 receptor-binding domain of S (www.biorxiv.org/content/10.1101/2020.06.11.146522v1).
237 Furthermore, applying the Oxford classifications to our dataset clearly demonstrates that the 3
238 most C-terminal sites (N1158, N1173, and N1194), dominated by complex type glycans, were
239 more often further processed (i.e. multiple antennae) and elaborated (i.e. galactosylation and
240 sialylation) than other sites (**Supplemental Table, Tab 8**).

241 We also analyzed our generated mass spectrometry data for the presence of O-linked glycans
242 based on our glycomic findings (**Fig. S5, Supplemental Table, Tab 4**) and a recent manuscript
243 suggesting significant levels of O-glycosylation of S1 and S2 when expressed independently (23).
244 We were able to confirm sites of O-glycan modification with microheterogeneity observed for the
245 vast majority of these sites (**Supplemental Table, Tab 9**). However, occupancy at each site,
246 determined by spectral counts, was observed to be very low (below 4%) except for Thr0323 that
247 had a modestly higher but still low 11% occupancy (**Fig. S6, Supplemental Table, Tab 10**).

248 *3D Structural Modeling of Glycosylated SARS-CoV-2 Trimer Immunogen Enables Predictions of*
249 *Epitope Accessibility and Other Key Features*

250 A 3D structure of the S trimer was generated using a homology model of the S trimer described
251 previously (based on PDB code 6VSB, (48)). Onto this 3D structure, we installed explicitly defined
252 glycans at each glycosylated sequon based on one of three separate sets of criteria, thereby
253 generating three different glycoform models for comparison that we denote as "Abundance,"
254 "Oxford Class," and "Processed" models (**see Methods and Supplemental Table, Tab 1**). These
255 criteria were chosen in order to generate glycoform models that represent reasonable
256 expectations for glycosylation microheterogeneity and integrate cross-validating glycomic and
257 glycoproteomic characterization of S and ACE2.

258 The three glycoform models were subjected to multiple all atom MD simulations with explicit
259 water. Information from analyses of these structures is presented in **Figure 4A** along with the
260 sequence of the SARS-CoV-2 S protomer. We also determined variants in S that are emerging in
261 the virus that have been sequenced to date (**Supplemental Table, Tab 11**). The inter-residue
262 distances were measured between the most α-carbon-distal atoms of the N-glycan sites and
263 Spike glycoprotein population variant sites in 3D space (**Figure 4B**). Notable from this analysis,
264 there are several variants that don't ablate the N-linked sequon, but that are sufficiently close in
265 3-dimensional space to N-glycosites, such as D138H, H655Y, S939F, and L1203F, to warrant
266 further investigation.

267 The percentage of simulation time that each S protein residue is accessible to a probe that
268 approximates the size of an antibody variable domain was calculated for a model of the S trimer
269 using the Abundance glycoforms (**Supplemental Table, Tab 1,** (49)). The predicted antibody
270 accessibility is visualized across the sequence, as well as mapped onto the 3D surface, via color
271 shading **(Figure 4A, 4C, Supplemental Table, Tab 13, and Supplemental Movie A)**.
272 Additionally, the Oxford Class glycoforms model (**Supplemental Table, Tab 1)**, which is arguably
273 the most encompassing means for representing glycan microheterogeneity since it captures
274 abundant structural topologies (**Supplemental Table, Tab 8**), is shown with the sequence variant
275 information (Figure 4D, **Supplemental Table, Tab 11**). A substantial number of these variants
276 occur (directly by comparison to **Figure 4A** or visually by comparison to **Figure 4C**) in regions of
277 high calculated epitope accessibility (e.g. N74K, T76I, R78M, D138H, H146Y, S151I, D253G,
278 V483A, etc., **Supplemental Table, Tab 14**) suggesting potential selective pressure to avoid host
279 immune response. Also, it is interesting to note that 3 of the emerging variants would eliminate
280 N-linked sequons in S; N74K and T76I would eliminate N-glycosylation of N74 (found in the insert
281 variable region 1 of CoV-2 S compared to CoV-1 S), and S151I eliminates N-glycosylation of
282 N149 (found in the insert variable region 2) (**Fig. 4A, S7, Supplemental Table, Tab 11)**. Lastly,
283 the SARS-CoV-2 S Processed glycoform model is shown (**Supplemental Table, Tab 1**), along
284 with marking amino acid T0323 that has a modest (11% occupancy, **Fig. S6, Supplemental**
285 **Table, Tab 10**) amount of O-glycosylation to represent the most heavily glycosylated form of S
286 (**Figure 4E**).

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

288 We also analyzed ACE2 glycosylation utilizing the same glycomic and glycoproteomic
289 approaches described for S protein. Glycomic analyses of released N-linked glycans (**Fig. 5A,**
290 **Supplemental Table, Tab 3**) revealed that the majority of glycans on ACE2 are complex with
291 limited high-mannose and hybrid glycans and we were unable to detect sulfated N-linked glycans.
292 Glycoproteomic analyses revealed that occupancy was high (>75%) at all 6 sites and significant
293 microheterogeneity dominated by complex N-glycans was observed for each site (**Fig. 5B-5G,**

294 **Supplemental Table, Tabs 5-8**). We also observed, consistent with the O-glycomics (**Fig. S5,**
295 **Supplemental Table, Tab 4**), that Ser 155 and several S/T residues at the C-terminus of ACE2
296 outside of the peptidase domain were O-glycosylated but stoichiometry was extremely low (less
297 than 2%, **Supplemental Table, Tab 9 and 10**).

298 *3D Structural Modeling of Glycosylated, Soluble, ACE2 Highlighting Glycosylation and Variants.*

299 We integrated our glycomics, glycoproteomics, and population variant analyses results with a 3D
300 model of Ace 2 (based on PDB code 6M0J (50), see methods for details) to generate two versions
301 of the soluble glycosylated ACE2 for visualization and molecular dynamics simulations. We
302 visualized the ACE2 glycoprotein with the Abundance glycoform model simulated at each site as
303 well as highlighting the naturally occurring variants observed in the human population (**Fig. 6A,**
304 **Supplemental Movie B, Supplemental Table, Tab 11**). Note, that the Abundance glycoform
305 model and the Oxford Class glycoform model for ACE2 are identical (**Supplemental Table, Tabs**
306 **1 and 8**). Notably, one site of N-linked glycosylation (N546) is predicted to not be present in 3
307 out of 10,000 humans based on naturally occurring variation in the human population
308 (**Supplemental Table, Tab 11**). We also modeled ACE2 using the Processed glycoform model
309 (**Fig. 6B**). In both models, the interaction domain with S is defined (**Fig. 6A-B, Supplemental**
310 **Movie B**) .

311 *Molecular Dynamics Simulation of the Glycosylated Trimer Spike of SARS-CoV-2 in Complex*
312 *with Glycosylated, Soluble, Human Ace 2 Reveals Protein and Glycan Interactions*

313 Molecular dynamics simulations were performed to examine the co-complex (generated from a
314 crystal structure of the ACE2-RBD co-complex, PDB code 6M0J, (50)) of glycosylated S with
315 glycosylated ACE2 with the 3 different glycoforms models (Abundance, Oxford Class, and
316 Processed, **Supplemental Table, Tab 1, Supplemental Simulations 1-3**). Information from
317 these analyses is laid out along the primary structure (sequence) of the SARS-CoV-2 S protomer
318 and ACE2 highlighting regions of glycan-protein interaction observed in the MD simulations
319 (**Supplemental Table, Tab 14**, **Supplemental Simulations 1-3**). Interestingly, two glycans on
320 ACE2 (at N090 and N322), that are highlighted in **Figure 7A** and shown in a more close-up view

321 in **Figure 7B**, are predicted to form interactions with the S protein (**Supplemental Table, Tab 15**).
322 The N322 glycan interaction with the S trimer is outside of the receptor binding domain, and the
323 interaction is observed across multiple simulations and throughout each simulation (**Fig. 7A-B**,
324 **Supplemental Simulations 1-3**). The ACE2 glycan at N090 is close enough to the S trimer
325 surface to repeatedly form interactions, however the glycan arms interact with multiple regions of
326 the surface over the course of the simulations, reflecting the relatively high degree of glycan
327 dynamics (**Fig. 7A-B, Supplemental Movie C**). Inter-molecule glycan-glycan interactions are
328 also observed repeatedly between the glycan at N546 of ACE2 and those in the S protein at
329 residues N0074 and N0165 (**Fig. 7D, Supplemental Table, Tab 16**). Finally, a full view of the
330 ACE2-S complex with Oxford class glycoforms on both proteins illustrates the extensive
331 glycosylation at the interface of the complex (**Fig. 7C, Supplemental Movie D**).
332

333 **DISCUSSION**

334 We have defined the glycomics-informed, site-specific microheterogeneity of 22 sites of N-linked
335 glycosylation per monomer on a SARS-CoV-2 trimer and the 6 sites of N-linked glycosylation on
336 a soluble version of its human ACE2 receptor using a combination of mass spectrometry
337 approaches coupled with evolutionary and variant sequence analyses to provide a detailed
338 understanding of the glycosylation states of these glycoproteins (**Figs. 1-6**). Our results suggest
339 essential roles for glycosylation in mediating receptor binding, antigenic shielding, and potentially
340 the evolution/divergence of these glycoproteins.
341 The highly glycosylated SARS-CoV-2 Spike protein, unlike several other viral proteins including
342 HIV-1 (15) but in agreement with another recent report (19), presents significantly more
343 processing of N-glycans towards complex glycosylation, suggesting that steric hindrance to
344 processing enzymes is not a major factor at most sites (**Figs. 2-3**). However, the N-glycans still
345 provide considerable shielding of the peptide backbone (**Fig. 4**). Our glycomics-guided
346 glycoproteomic data is in general in strong agreement with the trimer immunogen data recently
347 published by Crispin (19) though we also observed sulfated N-linked glycans, were able to

348 differentiate branching, bisected, and diLacNAc containing structures by glycomics, and observed
349 less occupancy on the 2 most C-terminal N-linked sites using a different approach. Our detection
350 of sulfated N-linked glycans at multiple sites on S is in agreement with a recent manuscript re-
351 analyzing the Crispin data (https://www.biorxiv.org/content/10.1101/2020.05.31.125302v1).
352 Sulfated N-linked glycans could potentially play key roles in immune regulation and receptor
353 binding as in other viruses (51). This result is especially significant in that sulfated N-glycans
354 were not observed when we performed glycomics on ACE2. At each individual site, the glycans
355 we observed on our immunogen appear to be slightly more processed but the overlap between
356 our analysis and the Crispin's group results (19) at each site in terms of major features are nearly
357 superimposable. This agreement differs substantially when comparing our and Crispin's data (19)
358 to that of the Azadi group (23) that analyzed S1 and S2 that had been expressed individually.
359 When expressed as 2 separate polypeptides and not purified for trimers, several unoccupied
360 sites of N-linked glycosylation were observed and processing at several sites was significantly
361 different (23) than we and others (19) observed. Although O-glycosylation has recently been
362 reported for individually-expressed S1 and S2 domains of the Spike glycoprotein (23), in trimeric
363 form the level of O-glycosylation is extremely low, with the highest level of occupancy we observed
364 being 11% at T0323 (**Fig. 4E**). The low level of O-linked occupancy we observed is in agreement
365 with Crispin's analysis of a Spike Trimer immunogen (19) but differs significantly from Azadi's
366 analyses of individually expressed S1 and S2 (23). Thus, the context in which the Spike protein
367 is expressed and purified before analyses significantly alters the glycosylation of the protomer
368 that is reminiscent of previous studies looking at expression of the HIV-1 envelope Spike (15,52).
369 The soluble ACE2 protein examined here contains 6 highly utilized sites of N-linked glycosylation
370 dominated by complex type N-linked glycans (**Fig. 5**). O-glycans were also present on this
371 glycoprotein but at very low levels of occupancy at all sites (<2%).
372 Our glycomics-informed glycoproteomics allowed us to assign defined sets of glycans to specific
373 glycosylation sites on 3D-structures of S and ACE2 glycoproteins based on experimental
374 evidence (**Figs. 4, 6**). Similar to almost all glycoproteins, microheterogeneity is evident at most

375 glycosylation sites of S and ACE2; each glycosylation site can be modified with one of several
376 glycan structures, generating site-specific glycosylation portfolios. For modeling purposes,
377 however, explicit structures must be placed at each glycosylation site. In order to capture the
378 impact of microheterogenity on S and ACE2 molecular dynamics we chose to generate
379 glycoforms for modeling that represented reasonable portfolios of glycan types. Using 3
380 glycoform models for S (Abundance, Oxford Class, and Processed) and 2 models for ACE2
381 (Abundance, which was equivalent to Oxford Class, and Processed), we generated 3 molecular
382 dynamics simulations of the co-complexes of these 2 glycoproteins (**Fig. 7 and Supplemental**
383 **Simulations 1-3**). The observed interactions over time allowed us to evaluate glycan-protein
384 contacts between the 2 proteins as well as examine potential glycan-glycan interactions (**Fig. 7**).
385 We observed glycan-mediated interactions between the S trimer and glycans at N090, N322 and
386 N546 of ACE2. Thus, variations in glycan occupancy or processing at these sites, could alter the
387 affinity of the SARS-CoV-2 – ACE2 interaction and modulate infectivity. It is well established that
388 glycosylation states vary depending on tissue and cell type as well as in the case of humans, on
389 age (53), underlying disease (54,55) and ethnicity (56). Thus glycosylation portfolios may in part
390 be responsible for tissue tropism and individual susceptibility to infection. The importance of
391 glycosylation for S binding to ACE2 is even more emphatically demonstrated by the direct glycan-
392 glycan interactions observed (**Fig. 7**) between S glycans (at N0074 and N0165) and an ACE2
393 receptor glycan (at N546), adding an additional layer of complexity for interpreting the impact of
394 glycosylation on individual susceptibility.
395 Several emerging variants of the virus appear to be altering N-linked glycosylation occupancy by
396 disrupting N-linked sequons. Interestingly, the 2 N-linked sequons in SARS-CoV-2 S directly
397 impacted by variants, N0074 and N0149, are in divergent insert regions 1 and 2, respectively, of
398 SARS-CoV-2 S compared to SARS-CoV-1 S (**Fig. 4A**). The N0074, in particular, is one of the S
399 glycans that interact directly with ACE2 glycan (at N546, **Fig. 7**), suggesting that glycan-glycan
400 interactions may contribute to the unique infectivity differences between SARS-CoV-2 and SARS-
401 CoV-1. These sequon variants will also be important to examine in terms of glycan shielding that

402 could influence immunogenicity and efficacy of neutralizing antibodies, as well as interactions with
403 the host cell receptor ACE2. Naturally-occurring amino acid-changing SNPs in the ACE2 gene
404 generate a number of variants including 1 variant, with a frequency of 3 in 10,000 humans, that
405 eliminates a site of N-linked glycosylation at N546 (**Fig. 6**). Understanding the impact of ACE2
406 variants on glycosylation and more importantly on S binding, especially for N546S which impacts
407 the glycan-glycan interaction between S and ACE2 (**Fig. 7**), should be prioritized in light of efforts
408 to develop ACE2 as a potential decoy therapeutic. Intelligent manipulation of ACE2 glycosylation
409 may lead to more potent biologics capable of acting as better competitive inhibitors of S binding.
410 The data presented here, and related similar recent findings (19,57,58), provide a framework to
411 facilitate the production of immunogens, vaccines, antibodies, and inhibitors as well as providing
412 additional information regarding mechanisms by which glycan microheterogeneity is achieved.
413 However, considerable efforts still remain in order to fully understand the role of glycans in SARS-
414 CoV-2 infection and pathogenicity. While HEK-expressed S and ACE2 provide a useful window
415 for understanding human glycosylation of these proteins, glycoproteomic characterization
416 following expression in cell lines of more direct relevance to disease and target tissue is sorely
417 needed. While site occupancy may change depending on presentation and cell type (59),
418 processing of N-linked glycans will almost certainly be altered in a cell-type dependent fashion.
419 Thus, analyses of the Spike trimer extracted from pseudoviruses, virion-like particles, and
420 ultimately from infectious SARS-CoV-2 virions harvested from airway cells or patients will provide
421 the most accurate view of how trimer immunogens reflect the true glycosylation pattern of the
422 virus. Detailed analyses of the impact of emerging variants in S and natural and designed-for-
423 biologics variants of ACE2 on glycosylation and binding properties are important next steps for
424 developing therapeutics. Finally, it will be important to monitor the slow evolution of the virus to
425 determine if existing sites of glycosylation are lost or new sites emerge with selective pressure
426 that might alter the efficacy of vaccines, neutralizing antibodies, and/or inhibitors.
427

428 **ACKNOWLEDGEMENTS**

The authors would like to thank Protein Metrics for providing licenses for their software used here and the developers of pGlyco for productive discussions regarding their software. We would also like to thank Galit Alter of the Ragon Institute for facilitating this collaborative effort. This effort was facilitated by the ThermoFisher Scientific appointed Center of Excellence in Glycoproteomics at the Complex Carbohydrate Research Center at the University of Georgia (co-directed by MT and LW). This research is supported in part by R35GM119850 (NEL), NNF10CC1016517 (NEL), R01AI139238 (MAB), R01AI147884-01A1S1 (BC), Massachusetts Consortium on Pathogen Readiness (BC), U01CA207824 (R.J.W), P41GM103390 (R.J.W.), P41GM103490 (MT and LW), U01GM125267 (MT), and R01GM130915 (LW). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

AUTHOR CONTRIBUTIONS

Conceptualization: M.T., B.C., R.J.W. and L.W.; Methodology, Software, Validation, Formal Analysis, Investigation, Resources, and Data Curation: P.Z., J.L.P., O.C.G., Y.C., T.X., K.E.R., K.A., B.P.K., R.B., D.H.B., M.A.B., N.E.L., M.T., B.C., R.J.W., and L.W.; Writing-Original Draft: P.Z., J.L.P., and L.W.; Writing-Review & Editing: All authors; Visualization: P.Z., J.L.P., O.C.G, Y.C., M.T., B.C., R.J.W., and L.W.; Supervision, Project Administration, and Funding Acquisition: D.H.B., M.A.B., N.E.L., M.T., B.C., R.J.W., and L.W..

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

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

456 bolded, underlined, and shown in green. The canonical N-linked glycosylation sequons are
457 bolded, underlined, and shown in red. Negative stain electron microscopy of the purified trimer
458 (B) and Coomassie G-250 stained reducing SDS-PAGE gels (C) confirmed purity of the SARS-
459 CoV-2 S protein trimer and of the soluble human ACE2. MWM = molecular weight markers. D)
460 A representative Step-HCD fragmentation spectrum from mass spectrometry analysis of a tryptic
461 digest of S annotated manually based on search results from pGlyco 2.2. This spectrum defines
462 the N-terminus of the mature protein monomer as (pyro-)glutamine 0014. A representative N-
463 glycan consistent with this annotation and our glycomics data (Fig. 2) is overlaid using the Symbol
464 Nomenclature For Glycans (SNFG) code. This complex glycan occurs at N0017. Note, that as
465 expected, the cysteine is carbamidomethylated and the mass accuracy of the assigned peptide
466 is 0.98 ppm. On the sequence of the N-terminal peptide and in the spectrum, the assigned b (blue)
467 and y (red) ions are shown. In the spectrum, purple highlights glycan oxonium ions and green
468 marks intact peptide fragment ions with various partial glycan sequences still attached. Note that
469 the green-labeled ions allow for limited topology to be extracted including defining that the fucose
470 is on the core and not the antennae of the glycopeptide.
471
472 **Figure 2. Glycomics Informed Glycoproteomics Reveals Substantial Site-Specific**
473 **Microheterogeneity of N-linked Glycosylation on SARS-CoV-2 S.** A) Glycans released from
474 SARS-CoV-2 S protein trimer immunogen were permethylated and analyzed by MSn. Structures
475 were assigned, grouped by type and structural features, and prevalence was determined based
476 on ion current. The pie chart shows basic division by broad N-glycan type. The bar graph provides
477 additional detail about the glycans detected. The most abundant structure with a unique
478 categorization by glycomics for each N-glycan type in the pie chart, or above each feature
479 category in the bar graph, is indicated. B – E) Glycopeptides were prepared from SARS-CoV-2
480 S protein trimer immunogen using multiple combinations of proteases, analyzed by LC-MSn, and
481 the resulting data was searched using several different software packages. Four representative
482 sites of N-linked glycosylation with specific features of interest were chosen and are presented

483 here. N0074 (B) and N0149 (C) are shown that occur in variable insert regions of S compared to
484 SARS-CoV and other related coronaviruses, and there are emerging variants of SARS-CoV-2
485 that disrupt these two sites of glycosylation in S. N0234 (D) contains the most high-mannose N-
486 linked glycans. N0801 (D) is an example of glycosylation in the S2 region of the immunogen and
487 displays a high degree of hybrid glycosylation compared to other sites. The abundance of each
488 composition is graphed in terms of assigned spectral counts. Representative glycans (as
489 determined by glycomics analysis) for several abundant compositions are shown in SNFG format.
490 The abbreviations used here and throughout the manuscript are N for HexNAc, H for Hexose, F
491 for Fucose, A for Neu5Ac, and S for Sulfation. Note that the graphs for the other 18 sites and
492 other graphs grouping the microheterogeneity observed by other properties are presented in
493 Supplemental Information.
494
495 **Figure 3. SARS-CoV-2 S Immunogen N-glycan Sites are Predominantly Modified by**
496 **Complex N-glycans.** N-glycan topologies were assigned to all 22 sites of the S protomer and
497 the spectral counts for each of the 3 types of N-glycans (high-mannose, hybrid, and complex) as
498 well as the unoccupied peptide spectral match counts at each site were summed and visualized
499 as pie charts. Note that only N1173 and N1194 show an appreciable amount of the unoccupied
500 amino acid.
501
502 **Figure 4. 3D Structural Modeling of Glycosylated SARS-CoV-2 Spike Trimer Immunogen**
503 **Reveals Predictions for Antigen Accessibility and Other Key Features.** Results from
504 glycomics and glycoproteomics experiments were combined with results from bioinformatics
505 analyses and used to model several versions of glycosylated SARS-CoV-2 S trimer immunogen.
506 A) Sequence of the SARS-CoV-2 S immunogen displaying computed antigen accessibility and
507 other information. Antigen accessibility is indicated by red shading across the amino acid
508 sequence. B) Emerging variants confirmed by independent sequencing experiments were
509 analyzed based on the 3D structure of SARS-CoV-2 S to generate a proximity chart to the

510 determined N-linked glycosylation sites. C) SARS-CoV-2 S trimer immunogen model from MD
511 simulation displaying abundance glycoforms and antigen accessibility shaded in red for most
512 accessible, white for partial, and black for inaccessible (see **Supplemental movie A**). D) SARS-
513 CoV-2 S trimer immunogen model from MD simulation displaying oxford class glycoforms and
514 sequence variants. * indicates not visible while the box represents 3 amino acid variants that are
515 clustered together in 3D space. E) SARS-CoV-2 S trimer immunogen model from MD simulation
516 displaying processed glycoforms plus shading of Thr-323 that has O-glycoslyation at low
517 stoichiometry in yellow.
518
519 **Figure 5: Glycomics Informed Glycoproteomics of Soluble Human ACE2 Reveals High**
520 **Occupancy, Complex N-linked Glycosylation.** A) Glycans released from soluble, purified
521 ACE2 were permethylated and analyzed by MSn. Structures were assigned, grouped by type and
522 structural features, and prevalence was determined based on ion current. The pie chart shows
523 basic division by broad N-glycan type. The bar graph provides additional detail about the glycans
524 detected. The most abundant structure with a unique categorization by glycomics for each N-
525 glycan type in the pie chart, or above each feature category in the bar graph, is indicated. B – G)
526 Glycopeptides were prepared from soluble human ACE2 using multiple combinations of
527 proteases, analyzed by LC-MSn, and the resulting data was searched using several different
528 software packages. All six sites of N-linked glycosylation are presented here. Displayed in the bar
529 graphs are the individual compositions observed graphed in terms of assigned spectral counts.
530 Representative glycans (as determined by glycomics analysis) for several abundant compositions
531 are shown in SNFG format. The abbreviations used here and throughout the manuscript are N
532 for HexNAc, H for Hexose, F for Fucose, and A for Neu5Ac. The pie chart (analogous to Figure
533 3 for SARS-CoV-2 S) for each site is displayed in the upper corner of each panel. B) N053. C)
534 N090. D) N103. E) N322. F) N432. G) N546, a site that does not exist in 3 in 10,000 people.
535

536 **Figure 6: 3D Structural Modeling of Glycosylated Soluble Human ACE2.** Results from
537 glycomics and glycoproteomics experiments were combined with results from bioinformatics
538 analyses and used to model several versions of glycosylated soluble human ACE2. A) Soluble
539 human ACE2 model from MD simulations displaying abundance glycoforms, interaction surface
540 with S, and sequence variants. N546 variant is boxed that would remove N-linked glycosylation
541 at that site (see **Supplemental movie B**). B) Soluble human ACE2 model from MD simulations
542 displaying processed glycoforms and interaction surface with S.
543
544 **Figure 7: Interactions of Glycosylated Soluble Human ACE2 and Glycosylated SARS-CoV-**
545 **2 S Trimer Immunogen Revealed By 3D-Structural Modeling and Molecular Dynamics**
546 **Simulations.** A) Molecular dynamics simulation of glycosylated soluble human ACE2 and
547 glycosylated SARS-CoV-2 S trimer immunogen interaction (see **Supplemental simulations 1-**
548 **3**). ACE2 (top) is colored red with glycans in pink while S is colored white with glycans in dark
549 grey. Highlighted are ACE2 glycans that interacts with S that are zoomed in on to the right. B)
550 Zoom in of ACE2-S interface highlighting ACE2 glycan interactions using 3D-SNFG icons (60)
551 with S protein (pink) as well as ACE2-S glycan-glycan interactions. C) Zoom in of dynamics
552 trajectory of glycans at the interface of soluble human ACE2 and S (see **Supplemental movies**
553 **C and D**).
554
555 **STAR METHODS**
556 **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
^{18}O 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
Experimental Models: Organisms/Strains		
VSV(G)-Pseudoviruses	This Study	
Software and Algorithms		
pGlyco	(43)	v2.2.2
Proteome Discoverer	Thermo Fisher Scientific	v1.4
Byonic	Protein Metrics Inc. (44)	v3.8.13
ProteoIQ	Premier Biosoft	v2.7
GRITS Toolbox	http://www.grits-toolbox.org (40)	v1.1
EMBOSS needle	(61)	v6.6.0
Biopython	https://biopython.org/	v1.76
Rpdb	https://rdrr.io/cran/Rpdb/	v2.3
SignalP	http://www.cbs.dtu.dk/services/SignalP/ (34)	v5.0
LibreOFFICE Writer	The Document Foundation	V6.4.4.2
GlyGen	https://www.glygen.org	V1.5

GNOme	https://github.com/glygen-glycan-data/GNOme/blob/master/README.md	V1.5.5
GlyTouCan	https://glytoucan.org	V3.1.0
Inkscape	Inkscape Developers	V1.0

557

558 **LEAD CONTACT**

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

562

563 **METHOD DETAILS**

564 **Expression, Purification, and Characterization of SARS-CoV-2 S and Human ACE2**
565 **Proteins**

566 To express a stabilized ectodomain of Spike protein, a synthetic gene encoding residues
567 1−1208 of SARS-CoV-2 Spike with the furin cleavage site (residues 682–685) replaced by a
568 "GGSG" sequence, proline substitutions at residues 986 and 987, and a foldon trimerization
569 motif followed by a C-terminal 6xHisTag was created and cloned into the mammalian
570 expression vector pCMV-IRES-puro (Codex BioSolutions, Inc, Gaithersburg, MD). The
571 expression construct was transiently transfected in HEK 293F cells using polyethylenimine
572 (Polysciences, Inc, Warrington, PA). Protein was purified from cell supernatants using Ni-NTA
573 resin (Qiagen, Germany), the eluted fractions containing S protein were pooled, concentrated,
574 and further purified by gel filtration chromatography on a Superose 6 column (GE Healthcare).
575 Negative stain electron microscopy (EM) analysis was performed as described (62). Briefly,
576 analysis was performed at room temperature with a magnification of 52,000x and a defocus
577 value of 1.5 μm following low-dose procedures, using a Philips Tecnai F20 electron microscope
578 (Thermo Fisher Scientific) equipped with a Gatan US4000 CCD camera and operated at voltage
579 of 200 kV. The DNA fragment encoding human ACE2 (1-615) with a 6xHis tag at C terminus

580 was synthesized by Genscript and cloned to the vector pCMV-IRES-puro. The expression
581 construct was transfected in HEK293F cells using polyethylenimine. The medium was discarded
582 and replaced with FreeStyle 293 medium after 6-8 hours. After incubation in 37 °C with 5.5%
583 CO2 for 5 days, the supernatant was collected and loaded to Ni-NTA resin for purification. The
584 elution was concentrated and further purified by a Superdex 200 column.

585 **In-Gel Analysis of SARS-CoV-2 S and Human ACE2 Proteins**

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

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

591 Aliquots of approximately 25-50 μg of S or ACE2 protein were processed for glycan analysis as
592 previously described (38,39). For N-linked glycan analysis, the proteins were digested with
593 trypsin. Following trypsinization, glycopeptides were enriched by C18 Sep-Pak and subjected to
594 PNGaseF digestion to release N-linked glycans. Following PNGaseF digestion, released
595 glycans were separated from residual glycosylated peptides bearing O-linked glycans by C18
596 Sep-Pak. O-glycosylated peptides were eluted from the Sep-Pak and subjected to reductive β-
597 elimination to release the O-glycans. Another 25-50 μg aliquot of each protein was denatured
598 with SDS and digested with PNGaseF to remove N-linked glycans. The de-N-glycosylated,
599 intact protein was precipitated with cold ethanol and then subjected to reductive β-elimination to
600 release O-glycans. The profiles of O-glycans released from peptides or from intact protein were
601 found to be comparable. N- and O-linked glycans released from glycoproteins were
602 permethylated with methyliodide according to the method of Anumula and Taylor prior to MS
603 analysis (63). Glycan structural analysis was performed using an LTQ-Orbitrap instrument
604 (Orbitrap Discovery, ThermoFisher). Detection and relative quantification of the prevalence of

605 individual glycans was accomplished using the total ion mapping (TIM) and neutral loss scan
606 (NL scan) functionality of the Xcalibur software package version 2.0 (Thermo Fisher Scientific)
607 as previously described (38,39). Mass accuracy and detector response was tuned with a
608 permethylated oligosaccharide standard in positive ion mode. For fragmentation by collision-
609 induced dissociation (CID in MS^2 and MSn), normalized collision energy of 45% was applied.
610 Most permethylated glycans were identified as singly or doubly charged, sodiated species in
611 positive mode. Sulfated N-glycans were detected as singly or doubly charged, deprotonated
612 species in negative ion mode. Peaks for all charge states were deconvoluted by the charge
613 state and summed for quantification. All spectra were manually interpreted and annotated. The
614 explicit identities of individual monosaccharide residues have been assigned based on known
615 human biosynthetic pathways. Graphical representations of monosaccharide residues are
616 consistent with the Symbol Nomenclature for Glycans (SNFG), which has been broadly adopted
617 by the glycomics community (64). The MS-based glycomics data generated in these analyses
618 and the associated annotations are presented in accordance with the MIRAGE standards and
619 the Athens Guidelines (65). Data annotation and assignment of glycan accession identifiers
620 were facilitated by GRITS Toolbox, GlyTouCan, GNOme, and GlyGen (40,66-68).

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

622 Two 10-µg aliquots of SARS-CoV-2 S protein were denatured by incubating with 20%
623 acetonitrile at room temperature and alkylated by 13.75 mM of iodoacetamide at room
624 temperature in dark. The two aliquots of proteins were then digested respectively using alpha
625 lytic protease, or a combination of trypsin, Lys-C and Glu-C. Following digestion, the proteins
626 were deglycosylated by PNGaseF treatment. The resulting peptides were separated on an
627 Acclaim PepMap RSLC C18 column (75 µm x 15 cm) and eluted into the nano-electrospray ion
628 source of an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer at a flow rate of 200
629 nL/min. The elution gradient consists of 1-40% acetonitrile in 0.1% formic acid over 370 minutes
630 followed by 10 minutes of 80% acetonitrile in 0.1% formic acid. The spray voltage was set to 2.2

631 kV and the temperature of the heated capillary was set to 280 °C. Full MS scans were acquired
632 from m/z 200 to 2000 at 60k resolution, and MS/MS scans following electron transfer
633 dissociation (ETD) were collected in the Orbitrap at 15k resolution. The raw spectra were
634 analyzed by Byonic (v3.8.13, Protein Metrics Inc.) with mass tolerance set as 20 ppm for both
635 precursors and fragments. The search output was filtered at 0.1% false discovery rate and 10
636 ppm mass error. The spectra assigned as cross-linked peptides were manually evaluated for
637 Cys0015 and Cys0136.

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

640 Four 3.5-µg aliquots of SARS-CoV-2 S protein were reduced by incubating with 10 mM of
641 dithiothreitol at 56 °C and alkylated by 27.5 mM of iodoacetamide at room temperature in dark.
642 The four aliquots of proteins were then digested respectively using alpha lytic protease,
643 chymotrypsin, a combination of trypsin and Glu-C, or a combination of Glu-C and AspN. Three
644 10-µg aliquots of ACE2 protein were reduced by incubating with 5 mM of dithiothreitol at 56 °C
645 and alkylated by 13.75 mM of iodoacetamide at room temperature in dark. The three aliquots of
646 proteins were then digested respectively using alpha lytic protease, chymotrypsin, or a
647 combination of trypsin and Lys-C. The resulting peptides were separated on an Acclaim
648 PepMap RSLC C18 column (75 µm x 15 cm) and eluted into the nano-electrospray ion source
649 of an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer at a flow rate of 200 nL/min. The
650 elution gradient consists of 1-40% acetonitrile in 0.1% formic acid over 370 minutes followed by
651 10 minutes of 80% acetonitrile in 0.1% formic acid. The spray voltage was set to 2.2 kV and the
652 temperature of the heated capillary was set to 280 °C. Full MS scans were acquired from m/z
653 200 to 2000 at 60k resolution, and MS/MS scans following higher-energy collisional dissociation
654 (HCD) with stepped collision energy (15%, 25%, 35%) were collected in the Orbitrap at 15k
655 resolution. pGlyco v2.2.2 (43) was used for database searches with mass tolerance set as 20
656 ppm for both precursors and fragments. The database search output was filtered to reach a 1%

657 false discovery rate for glycans and 10% for peptides. Quantitation was performed by
658 calculating spectral counts for each glycan composition at each site. Any N-linked glycan
659 compositions identified by only one spectra were removed from quantitation. N-linked glycan
660 compositions were categorized into 22 classes (including Unoccupied):
661 HexNAc(2)Hex(9~5)Fuc(0~1) was classified as M9 to M5 respectively;
662 HexNAc(2)Hex(4~1)Fuc(0~1) was classified as M1-M4;
663 HexNAc(3~6)Hex(5~9)Fuc(0)NeuAc(0~1) was classified as Hybrid with
664 HexNAc(3~6)Hex(5~9)Fuc(1~2)NeuAc(0~1) classified as F-Hybrid; Complex-type glycans are
665 classified based on the number of antenna, fucosylation, and sulfation:
666 HexNAc(3)Hex(3~4)Fuc(0)NeuAc(0~1) is assigned as A1 with
667 HexNAc(3)Hex(3~4)Fuc(1~2)NeuAc(0~1) assigned as F-A1;
668 HexNAc(4)Hex(3~5)Fuc(0)NeuAc(0~2) is assigned as A2/A1B with
669 HexNAc(4)Hex(3~5)Fuc(1~5)NeuAc(0~2) assigned as F-A2/A1B;
670 HexNAc(5)Hex(3~6)Fuc(0)NeuAc(0~3) is assigned as A3/A2B with
671 HexNAc(5)Hex(3~6)Fuc(1~3)NeuAc(0~3) assigned as F-A3/A2B;
672 HexNAc(6)Hex(3~7)Fuc(0)NeuAc(0~4) is assigned as A4/A3B with
673 HexNAc(6)Hex(3~7)Fuc(1~3)NeuAc(0~4) assigned as F-A4/A3B;
674 HexNAc(7)Hex(3~8)Fuc(0)NeuAc(0~1) is assigned as A5/A4B with
675 HexNAc(7)Hex(3~8)Fuc(1~3)NeuAc(0~1) as F-A5/A4B; HexNAc(8)Hex(3~9)Fuc(0) is assigned
676 as A6/A5B with HexNAc(8)Hex(3~9)Fuc(1) assigned as F-A6/A5B; any glycans identified with a
677 sulfate are assigned as Sulfated.

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

679 Three 3.5-μg aliquots of SARS-CoV-2 S protein were reduced by incubating with 10 mM of
680 dithiothreitol at 56 °C and alkylated by 27.5 mM of iodoacetamide at room temperature in dark.
681 The three aliquots were then digested respectively using chymotrypsin, Asp-N, or a combination
682 of trypsin and Glu-C. Two 10-μg aliquots of ACE2 protein were reduced by incubating with 5

683 mM of dithiothreitol at 56 °C and alkylated by 13.75 mM of iodoacetamide at room temperature
684 in dark. The two aliquots were then digested respectively using chymotrypsin, or a combination
685 of trypsin and Lys-C. Following digestion, the proteins were deglycosylated by Endoglycosidase
686 H followed by PNGaseF treatment in the presence of ^{18}O water. The resulting peptides were
687 separated on an Acclaim PepMap RSLC C18 column (75 µm x 15 cm) and eluted into the nano-
688 electrospray ion source of an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer at a flow
689 rate of 200 nL/min. The elution gradient consists of 1-40% acetonitrile in 0.1% formic acid over
690 370 minutes followed by 10 minutes of 80% acetonitrile in 0.1% formic acid. The spray voltage
691 was set to 2.2 kV and the temperature of the heated capillary was set to 280 °C. Full MS scans
692 were acquired from m/z 200 to 2000 at 60k resolution, and MS/MS scans following collision-
693 induced dissociation (CID) at 38% collision energy were collected in the ion trap. The spectra
694 were analyzed using SEQUEST (Proteome Discoverer 1.4) with mass tolerance set as 20 ppm
695 for precursors and 0.5 Da for fragments. The search output was filtered using ProteoIQ (v2.7) to
696 reach a 1% false discovery rate at protein level and 10% at peptide level. Occupancy of each N-
697 linked glycosylation site was calculated using spectral counts assigned to the ^{18}O-Asp-
698 containing (PNGaseF-cleaved) and/or HexNAc-modified (EndoH-cleaved) peptides and their
699 unmodified counterparts.

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

702 Three 10-µg aliquots of SARS-CoV-2 S protein and one 10-µg aliquot of ACE2 protein were
703 reduced by incubating with 5 mM of dithiothreitol at 56 °C and alkylated by 13.75 mM of
704 iodoacetamide at room temperature in dark. The four aliquots were then digested respectively
705 using trypsin, Lys-C, Arg-C, or a combination of trypsin and Lys-C. Following digestion, the
706 proteins were deglycosylated by PNGaseF treatment and then digested with O-protease
707 OpeRATOR®. The resulting peptides were separated on an Acclaim PepMap RSLC C18
708 column (75 µm x 15 cm) and eluted into the nano-electrospray ion source of an Orbitrap

709 Fusion™ Lumos™ Tribrid™ mass spectrometer at a flow rate of 200 nL/min. The elution
710 gradient consists of 1-40% acetonitrile in 0.1% formic acid over 370 minutes followed by 10
711 minutes of 80% acetonitrile in 0.1% formic acid. The spray voltage was set to 2.2 kV and the
712 temperature of the heated capillary was set to 280 °C. Full MS scans were acquired from m/z
713 200 to 2000 at 60k resolution, and MS/MS scans following higher-energy collisional dissociation
714 (HCD) with stepped collision energy (15%, 25%, 35%) or electron transfer dissociation (ETD)
715 were collected in the Orbitrap at 15k resolution. The raw spectra were analyzed by Byonic
716 (v3.8.13) with mass tolerance set as 20 ppm for both precursors and fragments. MS/MS filtering
717 was applied to only allow for spectra where the oxonium ions of HexNAc were observed. The
718 search output was filtered at 0.1% false discovery rate and 10 ppm mass error. The spectra
719 assigned as O-linked glycopeptides were manually evaluated. Quantitation was performed by
720 calculating spectral counts for each glycan composition at each site. Any O-linked glycan
721 compositions identified by only one spectra were removed from quantitation. Occupancy of each
722 O-linked glycosylation site was calculated using spectral counts assigned to any glycosylated
723 peptides and their unmodified counterparts from searches without MS/MS filtering.

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

725 The genomes of SARS-CoV as well as bat and pangolin coronavirus sequences reported to be
726 closely related to SARS-CoV-2 were downloaded from NCBI. The S protein sequences from all
727 of those genomes were aligned using EMBOSS needle v6.6.0 (61) via the EMBL-EBI provided
728 web service (69). Manual analysis was performed in the regions containing canonical N-
729 glycosylation sequons (N-X-S/T). For further sequence analysis of SARS-CoV-2 S variants, the
730 genomes of SARS-CoV-2 were downloaded from NCBI and GISAID and further processed
731 using Biopython 1.76 to extract all sequences annotated as “surface glycoprotein” and to
732 remove any incomplete sequence as well as any sequence containing unassigned amino acids.
733 For sequence analysis of human ACE2 variants, the single nucleotide polymorphisms (SNPs) of
734 ACE2 were extracted from the NCBI dbSNP database and filtered for missense mutation entries

735 with a reported minor allele frequency. Manual analysis was performed on both SARS-CoV-2 S
736 and human ACE2 variants to further examine the regions containing canonical N-glycosylation
737 sequons (N-X-S/T). LibreOffice Writer was used to shade regions on the linear sequence of S
738 and ACE2.

739 **3D Structural Modeling and Molecular Dynamics Simulation of Glycosylated SARS-CoV-2**
740 **S and Human ACE2 Proteins**

741 *SARS-CoV2 Spike (S) protein structure and ACE2 co-complex* – A 3D structure of the prefusion
742 form of the S protein (RefSeq: YP_009724390.1, UniProt: P0DTC2 SPIKE_SARS2), based on a
743 Cryo-EM structure (PDB code 6VSB) (48), was obtained from the SWISS-MODEL server
744 (swissmodel.expasy.org). The model has 95% coverage (residues 27 to 1146) of the S protein.
745 The receptor binding domain (RBD) in the "open" conformation was replaced with the RBD from
746 an ACE2 co-complex (PDB code 6M0J) by grafting residues C336 to V524.

747 *Glycoform generation* – Glycans (detected by glycomics) were selected for installation on
748 glycosylated S and ACE2 sequons (detected by glycoproteomics) based on three sets of criteria
749 designed to reasonably capture different aspects of glycosylation microheterogeneity. We
750 denote the first of these glycoform models as "Abundance." The glycans selected for
751 installation to generate the Abundance model were chosen because they were identified as the
752 most abundant glycan structure (detected by glycomics) that matched the most abundant glycan
753 composition (detected by glycoproteomics) at each individual site. We denote the second
754 glycoform model as "Oxford Class." The glycans selected for installation to generate the Oxford
755 Class model were chosen because they were the most abundant glycan structure, (detected by
756 glycomics) that was contained within the most highly represented Oxford classification group
757 (detected by glycoproteomics) at each individual site (**Fig. S7, Supplemental Table, Tabs 1,8).**
758 Finally, we denote the third glycoform model as "Processed." The glycans selected for
759 installation to generate the Processed model were chosen because they were the most highly
760 trimmed, elaborated, or terminally decorated structure (detected by glycomics) that

761 corresponded to a composition (detected by glycoproteomics) which was present at ≥ 1/3rd of
762 the abundance of the most highly represented composition at each site (**Supplemental Table,**
763 **Tab 1**). 3D structures of the three glycoforms (Abundance, Oxford Class, Processed) were
764 generated for the SARS-CoV2 S protein alone, and in complex with the glycosylated ACE2
765 protein. The glycoprotein builder available at GLYCAM-Web (www.glycam.org) was employed
766 together with an in-house program that adjusts the asparagine side chain torsion angles and
767 glycosidic linkages within known low-energy ranges (70) to relieve any atomic overlaps with the
768 core protein, as described previously (71,72).
769 *Energy minimization and Molecular dynamics (MD) simulations* **–** Each glycosylated structure
770 was placed in a periodic box of TIP3P water molecules with a 10 Å buffer between the solute
771 and the box edge. Energy minimization of all atoms was performed for 20,000 steps (10,000
772 steepest decent, followed by 10,000 conjugant gradient) under constant pressure (1 atm) and
773 temperature (300 K) conditions. All MD simulations were performed under nPT conditions with
774 the CUDA implementation of the PMEMD (73,74) simulation code, as present in the Amber14
775 software suite (University of California, San Diego). The GLYCAM06j force field (75) and
776 Amber14SB force field (76) were employed for the carbohydrate and protein moieties,
777 respectively. A Berendsen barostat with a time constant of 1 ps was employed for pressure
778 regulation, while a Langevin thermostat with a collision frequency of 2 ps^{-1} was employed for
779 temperature regulation. A nonbonded interaction cut-off of 8 Å was employed. Long-range
780 electrostatics were treated with the particle-mesh Ewald (PME) method (77). Covalent bonds
781 involving hydrogen were constrained with the SHAKE algorithm, allowing an integration time
782 step of 2 fs to be employed. The energy minimized coordinates were equilibrated at 300K over
783 400 ps with restraints on the solute heavy atoms. Each system was then equilibrated with
784 restraints on the Cα atoms of the protein for 1ns, prior to initiating 4 independent 250 ns
785 production MD simulations with random starting seeds for a total time of 1 μs per system, with
786 no restraints applied.

787 *Antigenic surface analysis.* A series of 3D structure snapshots of the simulation were taken at 1
788 ns intervals and analysed in terms of their ability to interact with a spherical probe based on the
789 average size of hypervariable loops present in an antibody complementarity determining region
790 (CDR), as described recently (https://www.biorxiv.org/content/10.1101/2020.04.07.030445v2).
791 The percentage of simulation time each residue was exposed to the AbASA probe was
792 calculated and plotted onto both the 3D structure and primary sequence.
793
794 **Analysis of SARS-CoV-2 Spike VSV pseudoparticles (ppVSV-SARS-2-S)**
795 293T cells were transfected with an expression plasmid encoding SARS-CoV-2 Spike
796 (pcDNAintron-SARS-2-SΔ19). To increase cell surface expression, the last 19 amino acids
797 containing the Golgi retention signal were removed. Two SΔ19 constructs were compared, one
798 started with Met1 and the other with Met2. Twenty-four hours following transfection, cells were
799 transduced with ppVSVΔG-VSV-G (particles that were pseudotyped with VSV-G in trans). One
800 hour following transduction cells were extensively washed and media was replaced.
801 Supernatant containing particles were collected 12-24 hour following transduction and cleared
802 through centrifugation. Cleared supernatant was frozen at -80°C for future use. Target cells
803 VeroE6 were seeded in 24-well plates ($5x10^5$ cells/mL) at a density of 80% coverage. The
804 following day, ppVSV-SARS-2-S/GFP particles were transduced into target cells for 60 minutes,
805 particles pseudotyped with VSV-G, Lassa virus GP, or no glycoprotein were included as
806 controls. 24 hours following transduction, transduced cells were released from the plate with
807 trypsin, fixed with 4% formaldehyde, and GFP-positive virus-transduced cells were quantified
808 using flow cytometry (Bectin Dickson BD-LSRII). To quantify the ability of various SARS-CoV-2
809 S mutants to mediate fusion, effector cells (HEK293T) were transiently transfected with the
810 indicated pcDNAintron-SARS-2-S expression vector or measles virus H and F (78). Effector
811 cells were infected with MVA-T7 four hours following transduction to produce the T7 polymerase
812 (79). Target cells naturally expressing the receptor ACE2 (Vero) or ACE2 negative cells

813 (HEK293T) were transfected with pTM1-luciferase, which encodes for firefly luciferase under the
814 control of a T7 promoter (80). 24 hours following transfection, the target cells were lifted and
815 added to the effector cells at a 1:1 ratio. 4 hours following co-cultivation, cells were washed,
816 lysed and luciferase levels were quantified using Promega's Steady-Glo substrate. To visualize
817 cell-to-cell fusion, Vero cells were co-transfected with pGFP and the pcDNAintron-SARS-2-S
818 constructs. 24 hours following transfection, syncytia was visualized by fluorescence microscopy.

819 **DATA AVAILABILITY**

820 The mass spectrometry proteomics data are available via ProteomeXchange with identifiers
821 listed in the KEY RESOURCES TABLE.

822 **SUPPLEMENTAL INFORMATION**

823 Tables (1, 16 tabs), Figures (7), Movies (4), and Simulations (3).

824 **SUPPLEMENTAL LEGEND:**

825 **Supplemental Table Tab 1**. Glycans modeled as Abundance, Oxford Class, and Processed.
826 **Supplemental Table Tab 2.** Cys0015-Cys0136 Disulfide Linked Peptide for SARS-CoV-2 S.
827 **Supplemental Table Tab 3**. Detection of N-linked glycans released from SARS-CoV-2 S and
828 human ACE2. Relative abundance (prevalence) of each species is calculated based on peak
829 intensity in full MS.
830 **Supplemental Table Tab 4**. Detection of O-linked glycans released from SARS-CoV-2 S and
831 human ACE2. Relative abundance (prevalence) of each species is calculated based on peak
832 intensity in full MS.
833 **Supplemental Table Tab 5**. N-linked glycan occupancy at each site of SARS-CoV-2 S and
834 human ACE2. Occupancy is calculated using spectral counts assigned to the 18O-Asp-
835 containing (PNGaseF-cleaved) and/or HexNAc-modified (EndoH-cleaved) peptides and their
836 unmodified counterparts. Sequon refers to the Asn-x-Ser/Thr/Cys, Asn-Gly-x sequences.
837 **Supplemental Table Tab 6**. N-linked glycan compositions identified at each site of SARS-CoV-
838 2 S and human ACE2. Asn(N)# indicates the numbers of asparagines in protein sequences. In

839 compositions: N=HexNAc, H=Hexose (Hex), F=Fucose (Fuc), A=Neu5Ac, S=Sulfate. In
840 fucosylation: NoFuc=No Fuc identified; 1Core=One Fuc identified at core position; 1Term=One
841 Fuc identified at terminal position; 1Core and 1Term=One Fuc identified as a mixture of core
842 and terminal positions; 1Core1Term=Two Fuc identified and one is at core and the other is at
843 terminal; 2Term=Two Fuc identified at terminal positions; 1Core1Term and 2Term=Two Fuc
844 identified as a mixture of core and terminal positions; 1Core2Term=Three Fuc identified and
845 one is at core and the others are at terminal; 3Term=Three Fuc identified at terminal positions;
846 1Core2Term and 3Term=Three Fuc identified as a mixture of core and terminal positions;
847 1Core3Term=Four Fuc identified and one is at core and the others are at terminal; 4Term=Four
848 Fuc identified at terminal positions; 1Core3Term and 4Term=Four Fuc identified as a mixture of
849 core and terminal positions; 1Core4Term=Five Fuc identified and one is at core and the others
850 are at terminal.
851 **Supplemental Table Tab 7**. N-linked glycan types identified at each site of SARS-CoV-2 S and
852 human ACE2. All N-linked glycans are categorized into 3 types: high-mannose, hybrid and
853 complex.
854 **Supplemental Table Tab 8**. N-linked glycan oxford classes identified at each site of SARS-
855 CoV-2 S and human ACE2. All N-linked glycan compositions are categorized into 22 classes:
856 M9 to M5 respectively is defined as HexNAc(2)Hex(9~5)Fuc(0~1); M1-M4 is defined as
857 HexNAc(2)Hex(4~1)Fuc(0~1); Hybrid is defined as HexNAc(3~6)Hex(5~9)Fuc(0)NeuAc(0~1)
858 and F-Hybrid is defined as HexNAc(3~6)Hex(5~9)Fuc(1~2)NeuAc(0~1). Complex-type glycans
859 are classified based on the number of antenna, fucosylation, and sulfation:
860 HexNAc(3)Hex(3~4)Fuc(0)NeuAc(0~1) is assigned as A1 with
861 HexNAc(3)Hex(3~4)Fuc(1~2)NeuAc(0~1) assigned as F-A1;
862 HexNAc(4)Hex(3~5)Fuc(0)NeuAc(0~2) is assigned as A2/A1B with
863 HexNAc(4)Hex(3~5)Fuc(1~5)NeuAc(0~2) assigned as F-A2/A1B;
864 HexNAc(5)Hex(3~6)Fuc(0)NeuAc(0~3) is assigned as A3/A2B with

865 HexNAc(5)Hex(3~6)Fuc(1~3)NeuAc(0~3) assigned as F-A3/A2B;

866 HexNAc(6)Hex(3~7)Fuc(0)NeuAc(0~4) is assigned as A4/A3B with

867 HexNAc(6)Hex(3~7)Fuc(1~3)NeuAc(0~4) assigned as F-A4/A3B;

868 HexNAc(7)Hex(3~8)Fuc(0)NeuAc(0~1) is assigned as A5/A4B with

869 HexNAc(7)Hex(3~8)Fuc(1~3)NeuAc(0~1) assigned as F-A5/A4B; HexNAc(8)Hex(3~9)Fuc(0) is

870 assigned as A6/A5B with HexNAc(8)Hex(3~9)Fuc(1) assigned as F-A6/A5B; any glycans

871 identified with a sulfate are assigned as Sulfated.

872 **Supplemental Table Tab 9**. O-linked glycan compositions identified at each site of SARS-CoV-

873 2 S and human ACE2. Ser/Thr# indicates the numbers of serines or threonines in protein

874 sequences. In compositions: N=HexNAc, H=Hexose (Hex), F=Fucose (Fuc), and A=Neu5Ac.

875 **Supplemental Table Tab 10**. O-linked glycan occupancy at each site of SARS-CoV-2 S and

876 human ACE2. Occupancy is calculated using spectral counts assigned to the glycosylated

877 peptides and their unmodified counterparts.

878 **Supplemental Table Tab 11**. SARS-CoV-2 S and human ACE2 variants.

879 **Supplemental Table Tab 12**. Proteomic Analyses of purified S and ACE2.

880 **Supplemental Table Tab 13.** Sulfated N-linked glycans released from SARS-CoV-2 S.

881 Following permethylation, almost all of the sulfated hybrid and complex N-glycans are recovered

882 in the organic phase despite their anionic charge. Organic phase permethylated glycans were

883 analyzed by mass spectrometry using negative ion mode. The indicated glycan structures are

884 consistent with the compositions detected at the m/z values shown.

885 **Supplemental Table Tab 14.** Surface Antigen Exposure of Abundance Glycosylated S. The

886 scale used is 0 (not accessible) to 1.0 (fully accessible).

887 **Supplemental Table Tab 15.** ACE2-Glycan-S-Peptide Interactions.

888 The scale used is 0 (no interaction) to 1.0 (interacted throughout entire simulation).

889 **Supplemental Table Tab 16.** S-Glycan-ACE2-Peptide Interactions.

890 The scale used is 0 (no interaction) to 1.0 (interacted throughout entire simulation).

891

892 **Supplemental Figure S1**. Defining N-terminus of ACE2 as pyro-glutamine at site Q0018.
893 Representative HCD MS2 spectrum shown.
894 **Supplemental Figure S2**. Disulfide bond formed between Cysteines 0015 and 0136 of SARS-
895 CoV-2 S. Representative EThcD MS2 spectrum shown.
896 **Supplemental Figure S3**. Signal P prediction of two different start methionines for SARS-CoV-
897 2 S.
898 **Supplemental Figure S4**. Functional characterization of various S constructs in Pseudovirus.
899 (A) Syncytia produced by SARS-CoV-2 S constructs in VeroE6 cells co-transfected with a GFP
900 plasmid to visualize cell-to-cell fusion. Quantification of fusion using a luciferase
901 complementation assay in 293T (B) or VeroE6 cells (C). (D) Transduction efficiency in Vero E6
902 cells of ppVSV-GFP particles coated in the indicated glycoprotein. Results suggest that start
903 methionine does not alter fusion or efficiency.
904 **Supplemental Figure S5**. Detection of O-linked glycans released from SARS-CoV-2 S and
905 human ACE2. The detected O-glycans were categorized based on their structures and types.
906 Relative abundance (prevalence) of each species is calculated based on peak intensity in full
907 MS.
908 **Supplemental Figure S6**. O-linked glycans detected at site T0323 of SARS-CoV-2 S.
909 Representative Step-HCD spectra shown for 6 glycoforms.
910 **Supplemental Figure S7**. Sequence alignments of SARS-CoV-1 and SARS-CoV-2 S variants
911 as well as alignment of multiple S proteins from related coronaviruses.
912 **Supplemental Movie A:** Linked to Figure 4C, Glycosylated S antigen accessibility
913 **Supplemental Movie B:** Linked to Figure 6A, Glycosylated ACE2 with variants
914 **Supplemental Movie C:** Linked to Figure 7C, Interface of ACE2-S Complex
915 **Supplemental Movie D:** Linked to Figure 7C, the glycosylated ACE2-S Complex
916 **Simulation 1:** Linked to Figure 7A, Abundance glycoforms on ACE2-S Complex

917 **Simulation 2:** Linked to Figure 7A, Oxford class glycoforms on ACE2-S Complex

918 **Simulation 3:** Linked to Figure 7A, Processed glycoforms on ACE2-S Complex

919 **REFERENCES**

920

921 1. Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B.,
922 Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. D., Liu, M. Q., Chen, Y., Shen,
923 X. R., Wang, X., Zheng, X. S., Zhao, K., Chen, Q. J., Deng, F., Liu, L. L., Yan, B., Zhan, F. X.,
924 Wang, Y. Y., Xiao, G. F., and Shi, Z. L. (2020) A pneumonia outbreak associated with a
925 new coronavirus of probable bat origin. *Nature* **579**, 270-273
926 2. Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi,
927 Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., Chen, J., Meng,
928 Y., Wang, J., Lin, Y., Yuan, J., Xie, Z., Ma, J., Liu, W. J., Wang, D., Xu, W., Holmes, E. C.,
929 Gao, G. F., Wu, G., Chen, W., Shi, W., and Tan, W. (2020) Genomic characterisation
930 and epidemiology of 2019 novel coronavirus: implications for virus origins and
931 receptor binding. *Lancet* **395**, 565-574
932 3. Zhong, N. S., Zheng, B. J., Li, Y. M., Poon, Xie, Z. H., Chan, K. H., Li, P. H., Tan, S. Y.,
933 Chang, Q., Xie, J. P., Liu, X. Q., Xu, J., Li, D. X., Yuen, K. Y., Peiris, and Guan, Y. (2003)
934 Epidemiology and cause of severe acute respiratory syndrome (SARS) in
935 Guangdong, People's Republic of China, in February, 2003. *Lancet* **362**, 1353-1358
936 4. Xia, X. (2020) Extreme genomic CpG deficiency in SARS-CoV-2 and evasion of host
937 antiviral defense. *Mol Biol Evol*
938 5. Zhang, T., Wu, Q., and Zhang, Z. (2020) Probable Pangolin Origin of SARS-CoV-2
939 Associated with the COVID-19 Outbreak. *Curr Biol* **30**, 1346-1351 e1342
940 6. Alijotas-Reig, J., Esteve-Valverde, E., Belizna, C., Selva-O'Callaghan, A., Pardos-Gea, J.,
941 Quintana, A., Mekinian, A., Anunciacion-Llunell, A., and Miro-Mur, F. (2020)
942 Immunomodulatory therapy for the management of severe COVID-19. Beyond the
943 anti-viral therapy: A comprehensive review. *Autoimmun Rev* **19**, 102569
944 7. Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C.,
945 Hohmann, E., Chu, H. Y., Luetkemeyer, A., Kline, S., Lopez de Castilla, D., Finberg, R.
946 W., Dierberg, K., Tapson, V., Hsieh, L., Patterson, T. F., Paredes, R., Sweeney, D. A.,
947 Short, W. R., Touloumi, G., Lye, D. C., Ohmagari, N., Oh, M. D., Ruiz-Palacios, G. M.,
948 Benfield, T., Fatkenheuer, G., Kortepeter, M. G., Atmar, R. L., Creech, C. B., Lundgren,
949 J., Babiker, A. G., Pett, S., Neaton, J. D., Burgess, T. H., Bonnett, T., Green, M.,
950 Makowski, M., Osinusi, A., Nayak, S., Lane, H. C., and Members, A.-S. G. (2020)
951 Remdesivir for the Treatment of Covid-19 - Preliminary Report. *N Engl J Med*
952 8. Beun, R., Kusadasi, N., Sikma, M., Westerink, J., and Huisman, A. (2020)
953 Thromboembolic events and apparent heparin resistance in patients infected with
954 SARS-CoV-2. *Int J Lab Hematol* **42 Suppl 1**, 19-20
955 9. Dashti-Khavidaki, S., and Khalili, H. (2020) Considerations for Statin Therapy in
956 Patients with COVID-19. *Pharmacotherapy* **40**, 484-486
957 10. Fedson, D. S., Opal, S. M., and Rordam, O. M. (2020) Hiding in Plain Sight: an
958 Approach to Treating Patients with Severe COVID-19 Infection. *mBio* **11**

11. Shi, Q., Zhou, Q., Wang, X., Liao, J., Yu, Y., Wang, Z., Lu, S., Ma, Y., Xun, Y., Luo, X., Li, W., Fukuoka, T., Ahn, H. S., Lee, M. S., Luo, Z., Liu, E., Chen, Y., Li, Q., Yang, K., Guan, Q., Evidence, C.-., and Recommendations Working, G. (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
12. 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
13. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Muller, M. A., Drosten, C., and Pohlmann, S. (2020) SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280 e278
14. Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H., and Farzan, M. (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450-454
15. 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
16. Li, F. (2016) Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* **3**, 237-261
17. 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
18. Monteil, V., Kwon, H., Prado, P., Hagelkruys, A., Wimmer, R. A., Stahl, M., Leopoldi, A., Garreta, E., Hurtado Del Pozo, C., Prosper, F., Romero, J. P., Wirnsberger, G., Zhang, H., Slutsky, A. S., Conder, R., Montserrat, N., Mirazimi, A., and Penninger, J. M. (2020) Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. *Cell* **181**, 905-913 e907
19. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., and Crispin, M. (2020) Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*
20. 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. (2020) Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat Commun* **11**, 2688
21. Stertz, S., Reichelt, M., Spiegel, M., Kuri, T., Martinez-Sobrido, L., Garcia-Sastre, A., Weber, F., and Kochs, G. (2007) The intracellular sites of early replication and budding of SARS-coronavirus. *Virology* **361**, 304-315
22. Ujike, M., and Taguchi, F. (2015) Incorporation of spike and membrane glycoproteins into coronavirus virions. *Viruses* **7**, 1700-1725
23. 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*
24. 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

1005 25. Duan, H., Chen, X., Boyington, J. C., Cheng, C., Zhang, Y., Jafari, A. J., Stephens, T.,
1006 Tsybovsky, Y., Kalyuzhniy, O., Zhao, P., Menis, S., Nason, M. C., Normandin, E.,
1007 Mukhamedova, M., DeKosky, B. J., Wells, L., Schief, W. R., Tian, M., Alt, F. W., Kwong,
1008 P. D., and Mascola, J. R. (2018) Glycan Masking Focuses Immune Responses to the
1009 HIV-1 CD4-Binding Site and Enhances Elicitation of VRC01-Class Precursor
1010 Antibodies. *Immunity* **49**, 301-311 e305
1011 26. Escolano, A., Gristick, H. B., Abernathy, M. E., Merkenschlager, J., Gautam, R., Oliveira,
1012 T. Y., Pai, J., West, A. P., Jr., Barnes, C. O., Cohen, A. A., Wang, H., Golijanin, J., Yost, D.,
1013 Keeffe, J. R., Wang, Z., Zhao, P., Yao, K. H., Bauer, J., Nogueira, L., Gao, H., Voll, A. V.,
1014 Montefiori, D. C., Seaman, M. S., Gazumyan, A., Silva, M., McGuire, A. T., Stamatatos, L.,
1015 Irvine, D. J., Wells, L., Martin, M. A., Bjorkman, P. J., and Nussenzweig, M. C. (2019)
1016 Immunization expands B cells specific to HIV-1 V3 glycan in mice and macaques.
1017 *Nature* **570**, 468-473
1018 27. Seabright, G. E., Cottrell, C. A., van Gils, M. J., D'Addabbo, A., Harvey, D. J., Behrens, A.
1019 J., Allen, J. D., Watanabe, Y., Scaringi, N., Polveroni, T. M., Maker, A., Vasiljevic, S., de
1020 Val, N., Sanders, R. W., Ward, A. B., and Crispin, M. (2020) Networks of HIV-1
1021 Envelope Glycans Maintain Antibody Epitopes in the Face of Glycan Additions and
1022 Deletions. *Structure*
1023 28. Yu, W. H., Zhao, P., Draghi, M., Arevalo, C., Karsten, C. B., Suscovich, T. J., Gunn, B.,
1024 Streeck, H., Brass, A. L., Tiemeyer, M., Seaman, M., Mascola, J. R., Wells, L.,
1025 Lauffenburger, D. A., and Alter, G. (2018) Exploiting glycan topography for
1026 computational design of Env glycoprotein antigenicity. *PLoS Comput Biol* **14**,
1027 e1006093
1028 29. Pinto, D., Park, Y. J., Beltramello, M., Walls, A. C., Tortorici, M. A., Bianchi, S., Jaconi, S.,
1029 Culap, K., Zatta, F., De Marco, A., Peter, A., Guarino, B., Spreafico, R., Cameroni, E.,
1030 Case, J. B., Chen, R. E., Havenar-Daughton, C., Snell, G., Telenti, A., Virgin, H. W.,
1031 Lanzavecchia, A., Diamond, M. S., Fink, K., Veesler, D., and Corti, D. (2020) Cross-
1032 neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature*
1033 **583**, 290-295
1034 30. Tikellis, C., Bernardi, S., and Burns, W. C. (2011) Angiotensin-converting enzyme 2 is
1035 a key modulator of the renin-angiotensin system in cardiovascular and renal
1036 disease. *Curr Opin Nephrol Hypertens* **20**, 62-68
1037 31. Lambert, D. W., Yarski, M., Warner, F. J., Thornhill, P., Parkin, E. T., Smith, A. I.,
1038 Hooper, N. M., and Turner, A. J. (2005) Tumor necrosis factor-alpha convertase
1039 (ADAM17) mediates regulated ectodomain shedding of the severe-acute respiratory
1040 syndrome-coronavirus (SARS-CoV) receptor, angiotensin-converting enzyme-2
1041 (ACE2). *J Biol Chem* **280**, 30113-30119
1042 32. Epelman, S., Shrestha, K., Troughton, R. W., Francis, G. S., Sen, S., Klein, A. L., and
1043 Tang, W. H. (2009) Soluble angiotensin-converting enzyme 2 in human heart failure:
1044 relation with myocardial function and clinical outcomes. *J Card Fail* **15**, 565-571
1045 33. Li, W., Zhang, C., Sui, J., Kuhn, J. H., Moore, M. J., Luo, S., Wong, S. K., Huang, I. C., Xu, K.,
1046 Vasilieva, N., Murakami, A., He, Y., Marasco, W. A., Guan, Y., Choe, H., and Farzan, M.
1047 (2005) Receptor and viral determinants of SARS-coronavirus adaptation to human
1048 ACE2. *EMBO J* **24**, 1634-1643
1049 34. Marth, J. D., and Grewal, P. K. (2008) Mammalian glycosylation in immunity. *Nat Rev*
1050 *Immunol* **8**, 874-887

1051 35. Varki, A. (2017) Biological roles of glycans. *Glycobiology* **27**, 3-49
1052 36. Yu, J., Tostanoski, L. H., Peter, L., Mercado, N. B., McMahan, K., Mahrokhian, S. H.,
1053 Nkolola, J. P., Liu, J., Li, Z., Chandrashekar, A., Martinez, D. R., Loos, C., Atyeo, C.,
1054 Fischinger, S., Burke, J. S., Slein, M. D., Chen, Y., Zuiani, A., FJ, N. L., Travers, M., Habibi,
1055 S., Pessaint, L., Van Ry, A., Blade, K., Brown, R., Cook, A., Finneyfrock, B., Dodson, A.,
1056 Teow, E., Velasco, J., Zahn, R., Wegmann, F., Bondzie, E. A., Dagotto, G., Gebre, M. S.,
1057 He, X., Jacob-Dolan, C., Kirilova, M., Kordana, N., Lin, Z., Maxfield, L. F., Nampanya, F.,
1058 Nityanandam, R., Ventura, J. D., Wan, H., Cai, Y., Chen, B., Schmidt, A. G., Wesemann,
1059 D. R., Baric, R. S., Alter, G., Andersen, H., Lewis, M. G., and Barouch, D. H. (2020) DNA
1060 vaccine protection against SARS-CoV-2 in rhesus macaques. *Science*
1061 37. Almagro Armenteros, J. J., Tsirigos, K. D., Sonderby, C. K., Petersen, T. N., Winther, O.,
1062 Brunak, S., von Heijne, G., and Nielsen, H. (2019) SignalP 5.0 improves signal peptide
1063 predictions using deep neural networks. *Nat Biotechnol* **37**, 420-423
1064 38. Aoki, K., Perlman, M., Lim, J. M., Cantu, R., Wells, L., and Tiemeyer, M. (2007)
1065 Dynamic developmental elaboration of N-linked glycan complexity in the Drosophila
1066 melanogaster embryo. *J Biol Chem* **282**, 9127-9142
1067 39. Aoki, K., Porterfield, M., Lee, S. S., Dong, B., Nguyen, K., McGlamry, K. H., and
1068 Tiemeyer, M. (2008) The diversity of O-linked glycans expressed during Drosophila
1069 melanogaster development reflects stage- and tissue-specific requirements for cell
1070 signaling. *J Biol Chem* **283**, 30385-30400
1071 40. Weatherly, D. B., Arpinar, F. S., Porterfield, M., Tiemeyer, M., York, W. S., and
1072 Ranzinger, R. (2019) GRITS Toolbox-a freely available software for processing,
1073 annotating and archiving glycomics mass spectrometry data. *Glycobiology* **29**, 452-
1074 460
1075 41. Wang, S., Voronin, Y., Zhao, P., Ishihara, M., Mehta, N., Porterfield, M., Chen, Y.,
1076 Bartley, C., Hu, G., Han, D., Wells, L., Tiemeyer, M., and Lu, S. (2020) Glycan Profiles of
1077 gp120 Protein Vaccines from Four Major HIV-1 Subtypes Produced from Different
1078 Host Cell Lines under Non-GMP or GMP Conditions. *J Virol* **94**
1079 42. Ruiz-Canada, C., Kelleher, D. J., and Gilmore, R. (2009) Cotranslational and
1080 posttranslational N-glycosylation of polypeptides by distinct mammalian OST
1081 isoforms. *Cell* **136**, 272-283
1082 43. Zielinska, D. F., Gnad, F., Wisniewski, J. R., and Mann, M. (2010) Precision mapping of
1083 an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell*
1084 **141**, 897-907
1085 44. Palmisano, G., Melo-Braga, M. N., Engholm-Keller, K., Parker, B. L., and Larsen, M. R.
1086 (2012) Chemical deamidation: a common pitfall in large-scale N-linked
1087 glycoproteomic mass spectrometry-based analyses. *J Proteome Res* **11**, 1949-1957
1088 45. Zhou, T., Doria-Rose, N. A., Cheng, C., Stewart-Jones, G. B. E., Chuang, G. Y., Chambers,
1089 M., Druz, A., Geng, H., McKee, K., Kwon, Y. D., O'Dell, S., Sastry, M., Schmidt, S. D., Xu,
1090 K., Chen, L., Chen, R. E., Louder, M. K., Pancera, M., Wanninger, T. G., Zhang, B., Zheng,
1091 A., Farney, S. K., Foulds, K. E., Georgiev, I. S., Joyce, M. G., Lemmin, T., Narpala, S.,
1092 Rawi, R., Soto, C., Todd, J. P., Shen, C. H., Tsybovsky, Y., Yang, Y., Zhao, P., Haynes, B. F.,
1093 Stamatatos, L., Tiemeyer, M., Wells, L., Scorpio, D. G., Shapiro, L., McDermott, A. B.,
1094 Mascola, J. R., and Kwong, P. D. (2017) Quantification of the Impact of the HIV-1-
1095 Glycan Shield on Antibody Elicitation. *Cell Rep* **19**, 719-732

1096 46. Liu, M. Q., Zeng, W. F., Fang, P., Cao, W. Q., Liu, C., Yan, G. Q., Zhang, Y., Peng, C., Wu, J.
1097 Q., Zhang, X. J., Tu, H. J., Chi, H., Sun, R. X., Cao, Y., Dong, M. Q., Jiang, B. Y., Huang, J. M.,
1098 Shen, H. L., Wong, C. C. L., He, S. M., and Yang, P. Y. (2017) pGlyco 2.0 enables
1099 precision N-glycoproteomics with comprehensive quality control and one-step mass
1100 spectrometry for intact glycopeptide identification. *Nat Commun* **8**, 438
1101 47. Bern, M., Kil, Y. J., and Becker, C. (2012) Byonic: advanced peptide and protein
1102 identification software. *Curr Protoc Bioinformatics* **Chapter 13**, Unit13 20
1103 48. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., Graham, B.
1104 S., and McLellan, J. S. (2020) Cryo-EM structure of the 2019-nCoV spike in the
1105 prefusion conformation. *Science* **367**, 1260-1263
1106 49. Ferreira, R. C., Grant, O. C., Moyo, T., Dorfman, J. R., Woods, R. J., Travers, S. A., and
1107 Wood, N. T. (2018) Structural Rearrangements Maintain the Glycan Shield of an HIV-
1108 1 Envelope Trimer After the Loss of a Glycan. *Sci Rep* **8**, 15031
1109 50. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., and
1110 Wang, X. (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound
1111 to the ACE2 receptor. *Nature* **581**, 215-220
1112 51. Wang, C. C., Chen, J. R., Tseng, Y. C., Hsu, C. H., Hung, Y. F., Chen, S. W., Chen, C. M.,
1113 Khoo, K. H., Cheng, T. J., Cheng, Y. S., Jan, J. T., Wu, C. Y., Ma, C., and Wong, C. H. (2009)
1114 Glycans on influenza hemagglutinin affect receptor binding and immune response.
1115 *Proc Natl Acad Sci U S A* **106**, 18137-18142
1116 52. Behrens, A. J., Harvey, D. J., Milne, E., Cupo, A., Kumar, A., Zitzmann, N., Struwe, W. B.,
1117 Moore, J. P., and Crispin, M. (2017) Molecular Architecture of the Cleavage-
1118 Dependent Mannose Patch on a Soluble HIV-1 Envelope Glycoprotein Trimer. *J Virol*
1119 **91**
1120 53. Kristic, J., Vuckovic, F., Menni, C., Klaric, L., Keser, T., Beceheli, I., Pucic-Bakovic, M.,
1121 Novokmet, M., Mangino, M., Thaqi, K., Rudan, P., Novokmet, N., Sarac, J., Missoni, S.,
1122 Kolcic, I., Polasek, O., Rudan, I., Campbell, H., Hayward, C., Aulchenko, Y., Valdes, A.,
1123 Wilson, J. F., Gornik, O., Primorac, D., Zoldos, V., Spector, T., and Lauc, G. (2014)
1124 Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol*
1125 *Sci Med Sci* **69**, 779-789
1126 54. Pavic, T., Dilber, D., Kifer, D., Selak, N., Keser, T., Ljubicic, D., Vukic Dugac, A., Lauc, G.,
1127 Rumora, L., and Gornik, O. (2018) N-glycosylation patterns of plasma proteins and
1128 immunoglobulin G in chronic obstructive pulmonary disease. *J Transl Med* **16**, 323
1129 55. Rudman, N., Gornik, O., and Lauc, G. (2019) Altered N-glycosylation profiles as
1130 potential biomarkers and drug targets in diabetes. *FEBS Lett* **593**, 1598-1615
1131 56. Gebrehiwot, A. G., Melka, D. S., Kassaye, Y. M., Rehan, I. F., Rangappa, S., Hinou, H.,
1132 Kamiyama, T., and Nishimura, S. I. (2018) Healthy human serum N-glycan profiling
1133 reveals the influence of ethnic variation on the identified cancer-relevant glycan
1134 biomarkers. *PLoS One* **13**, e0209515
1135 57. Wrobel, A. G., Benton, D. J., Xu, P., Roustan, C., Martin, S. R., Rosenthal, P. B., Skehel, J.
1136 J., and Gamblin, S. J. (2020) SARS-CoV-2 and bat RaTG13 spike glycoprotein
1137 structures inform on virus evolution and furin-cleavage effects. *Nat Struct Mol Biol*
1138 58. Casalino, L., Gaieb, Z., Dommer, A. C., Harbison, A. M., Fogarty, C. A., Barros, E. P.,
1139 Taylor, B. C., Fadda, E., and Amaro, R. E. (2020) Shielding and Beyond: The Roles of
1140 Glycans in SARS-CoV-2 Spike Protein. *bioRxiv*

1141 59. Struwe, W. B., Chertova, E., Allen, J. D., Seabright, G. E., Watanabe, Y., Harvey, D. J.,
1142 Medina-Ramirez, M., Roser, J. D., Smith, R., Westcott, D., Keele, B. F., Bess, J. W., Jr.,
1143 Sanders, R. W., Lifson, J. D., Moore, J. P., and Crispin, M. (2018) Site-Specific
1144 Glycosylation of Virion-Derived HIV-1 Env Is Mimicked by a Soluble Trimeric
1145 Immunogen. *Cell Rep* **24**, 1958-1966 e1955
1146 60. Thieker, D. F., Hadden, J. A., Schulten, K., and Woods, R. J. (2016) 3D implementation
1147 of the symbol nomenclature for graphical representation of glycans. *Glycobiology*
1148 **26**, 786-787
1149 61. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular
1150 Biology Open Software Suite. *Trends Genet* **16**, 276-277
1151 62. Shaik, M. M., Peng, H., Lu, J., Rits-Volloch, S., Xu, C., Liao, M., and Chen, B. (2019)
1152 Structural basis of coreceptor recognition by HIV-1 envelope spike. *Nature* **565**,
1153 318-323
1154 63. Anumula, K. R., and Taylor, P. B. (1992) A comprehensive procedure for preparation
1155 of partially methylated alditol acetates from glycoprotein carbohydrates. *Anal*
1156 *Biochem* **203**, 101-108
1157 64. Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley,
1158 P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J. J., Schnaar, R. L., Freeze, H. H.,
1159 Marth, J. D., Bertozzi, C. R., Etzler, M. E., Frank, M., Vliegenthart, J. F., Lutteke, T.,
1160 Perez, S., Bolton, E., Rudd, P., Paulson, J., Kanehisa, M., Toukach, P., Aoki-Kinoshita, K.
1161 F., Dell, A., Narimatsu, H., York, W., Taniguchi, N., and Kornfeld, S. (2015) Symbol
1162 Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25**, 1323-1324
1163 65. Wells, L., Hart, G. W., and Athens Guidelines for the Publication of Glycomics, D.
1164 (2013) Glycomics: building upon proteomics to advance glycosciences. *Mol Cell*
1165 *Proteomics* **12**, 833-835
1166 66. Kahsay, R., Vora, J., Navelkar, R., Mousavi, R., Fochtman, B. C., Holmes, X.,
1167 Pattabiraman, N., Ranzinger, R., Mahadik, R., Williamson, T., Kulkarni, S., Agarwal, G.,
1168 Martin, M., Vasudev, P., Garcia, L., Edwards, N., Zhang, W., Natale, D. A., Ross, K., Aoki-
1169 Kinoshita, K. F., Campbell, M. P., York, W. S., and Mazumder, R. (2020) GlyGen data
1170 model and processing workflow. *Bioinformatics* **36**, 3941-3943
1171 67. Tiemeyer, M., Aoki, K., Paulson, J., Cummings, R. D., York, W. S., Karlsson, N. G.,
1172 Lisacek, F., Packer, N. H., Campbell, M. P., Aoki, N. P., Fujita, A., Matsubara, M.,
1173 Shinmachi, D., Tsuchiya, S., Yamada, I., Pierce, M., Ranzinger, R., Narimatsu, H., and
1174 Aoki-Kinoshita, K. F. (2017) GlyTouCan: an accessible glycan structure repository.
1175 *Glycobiology* **27**, 915-919
1176 68. York, W. S., Mazumder, R., Ranzinger, R., Edwards, N., Kahsay, R., Aoki-Kinoshita, K.
1177 F., Campbell, M. P., Cummings, R. D., Feizi, T., Martin, M., Natale, D. A., Packer, N. H.,
1178 Woods, R. J., Agarwal, G., Arpinar, S., Bhat, S., Blake, J., Castro, L. J. G., Fochtman, B.,
1179 Gildersleeve, J., Goldman, R., Holmes, X., Jain, V., Kulkarni, S., Mahadik, R., Mehta, A.,
1180 Mousavi, R., Nakarakommula, S., Navelkar, R., Pattabiraman, N., Pierce, M. J., Ross, K.,
1181 Vasudev, P., Vora, J., Williamson, T., and Zhang, W. (2020) GlyGen: Computational
1182 and Informatics Resources for Glycoscience. *Glycobiology* **30**, 72-73
1183 69. Madeira, F., Madhusoodanan, N., Lee, J., Tivey, A. R. N., and Lopez, R. (2019) Using
1184 EMBL-EBI Services via Web Interface and Programmatically via Web Services. *Curr*
1185 *Protoc Bioinformatics* **66**, e74

1186 70. Nivedha, A. K., Makeneni, S., Foley, B. L., Tessier, M. B., and Woods, R. J. (2014)
1187 Importance of ligand conformational energies in carbohydrate docking: Sorting the
1188 wheat from the chaff. *J Comput Chem* **35**, 526-539
1189 71. Grant, O. C., Xue, X., Ra, D., Khatamian, A., Foley, B. L., and Woods, R. J. (2016) Gly-
1190 Spec: a webtool for predicting glycan specificity by integrating glycan array
1191 screening data and 3D structure. *Glycobiology* **26**, 1027-1028
1192 72. Peng, W., de Vries, R. P., Grant, O. C., Thompson, A. J., McBride, R., Tsogtbaatar, B.,
1193 Lee, P. S., Razi, N., Wilson, I. A., Woods, R. J., and Paulson, J. C. (2017) Recent H3N2
1194 Viruses Have Evolved Specificity for Extended, Branched Human-type Receptors,
1195 Conferring Potential for Increased Avidity. *Cell Host Microbe* **21**, 23-34
1196 73. Gotz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012)
1197 Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1.
1198 Generalized Born. *J Chem Theory Comput* **8**, 1542-1555
1199 74. Salomon-Ferrer, R., Gotz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013)
1200 Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2.
1201 Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* **9**, 3878-3888
1202 75. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R.,
1203 Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force
1204 field. Carbohydrates. *J Comput Chem* **29**, 622-655
1205 76. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and
1206 Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and
1207 Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696-3713
1208 77. Darden, T. A., and Pedersen, L. G. (1993) Molecular modeling: an experimental tool.
1209 *Environ Health Perspect* **101**, 410-412
1210 78. Brindley, M. A., Chaudhury, S., and Plemper, R. K. (2015) Measles virus glycoprotein
1211 complexes preassemble intracellularly and relax during transport to the cell surface
1212 in preparation for fusion. *J Virol* **89**, 1230-1241
1213 79. Paal, M., Heel, T., Schneider, R., and Auer, B. (2009) A novel Ecotin-Ubiquitin-Tag
1214 (ECUT) for efficient, soluble peptide production in the periplasm of Escherichia coli.
1215 *Microb Cell Fact* **8**, 7
1216 80. Brindley, M. A., and Plemper, R. K. (2010) Blue native PAGE and biomolecular
1217 complementation reveal a tetrameric or higher-order oligomer organization of the
1218 physiological measles virus attachment protein H. *J Virol* **84**, 12174-12184

1219

A

Soluble Human ACE2

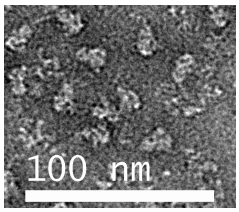
```
MSSSSWLLLS LVAVTAAQST IEEQAKTFLD KFNHEAEDLF YQSSLASWNY NTNNITEENVQ   60
NMNNAGDKWS AFLKEQSTLA QMYPLQEIQN LTVKLQLQAL QQNGSSVLSE DKSKRLNTIL  120
NTMSTIYSTG KVCNPDNPQE CLLLEPGLNE IMANSLDYNE RLWAWESWRS EVGKQLRPLY  180
EEYVVLKNEM ARANHYEDYG DYWRGDYEVN GVDGYDYSRG QLIEDVEHTF EEIKPLYEHL  240
HAYVRAKLMN AYPSYISPIG 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

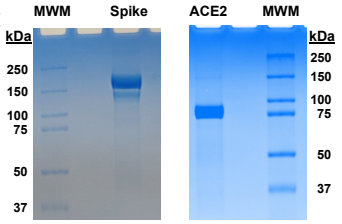
```
MFVFLVLLPL VSSQCVNLTT RTQLPPAYTN SFTRGVYYPD KVFRSSVLHS TQDLFLPFFS   60
NVTWFHAIHV SGTNGTKRFD NPVLPFNDGV YFASTEKSNI IRGWIFGTTL DSKTQSLLIV  120
NNATNVVIKV CEFQFCNDPF LGVYYHKNNK SWMESEFRVY SSANNCTFEY VSQPFLMDLE  180
GKQGNFKNLR EFVFKNIDGY FKIYSKHTPI NLVRDLPQGF SALEPLVDLP IGINITRFQT  240
LLALHRSYLT PGDSSSGWTA 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

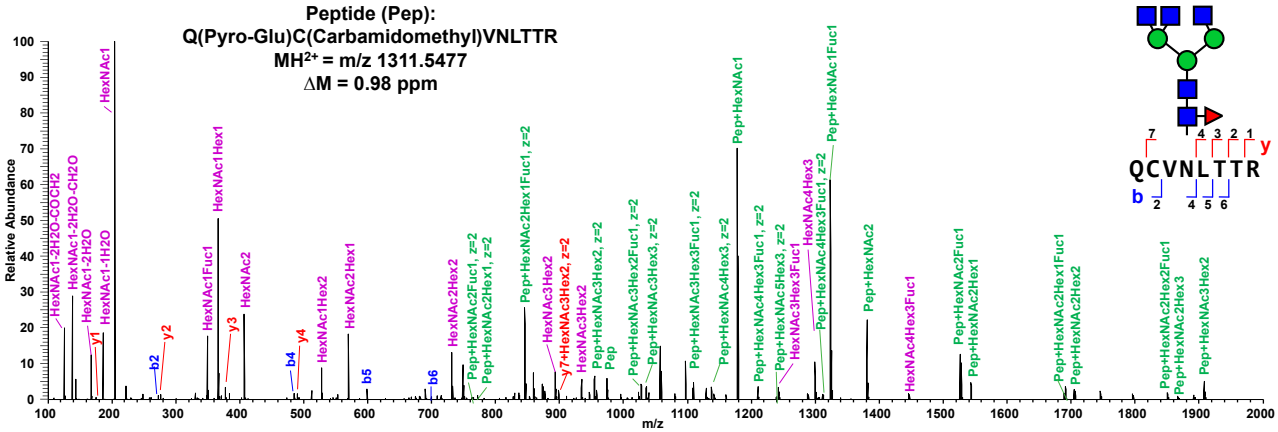
B

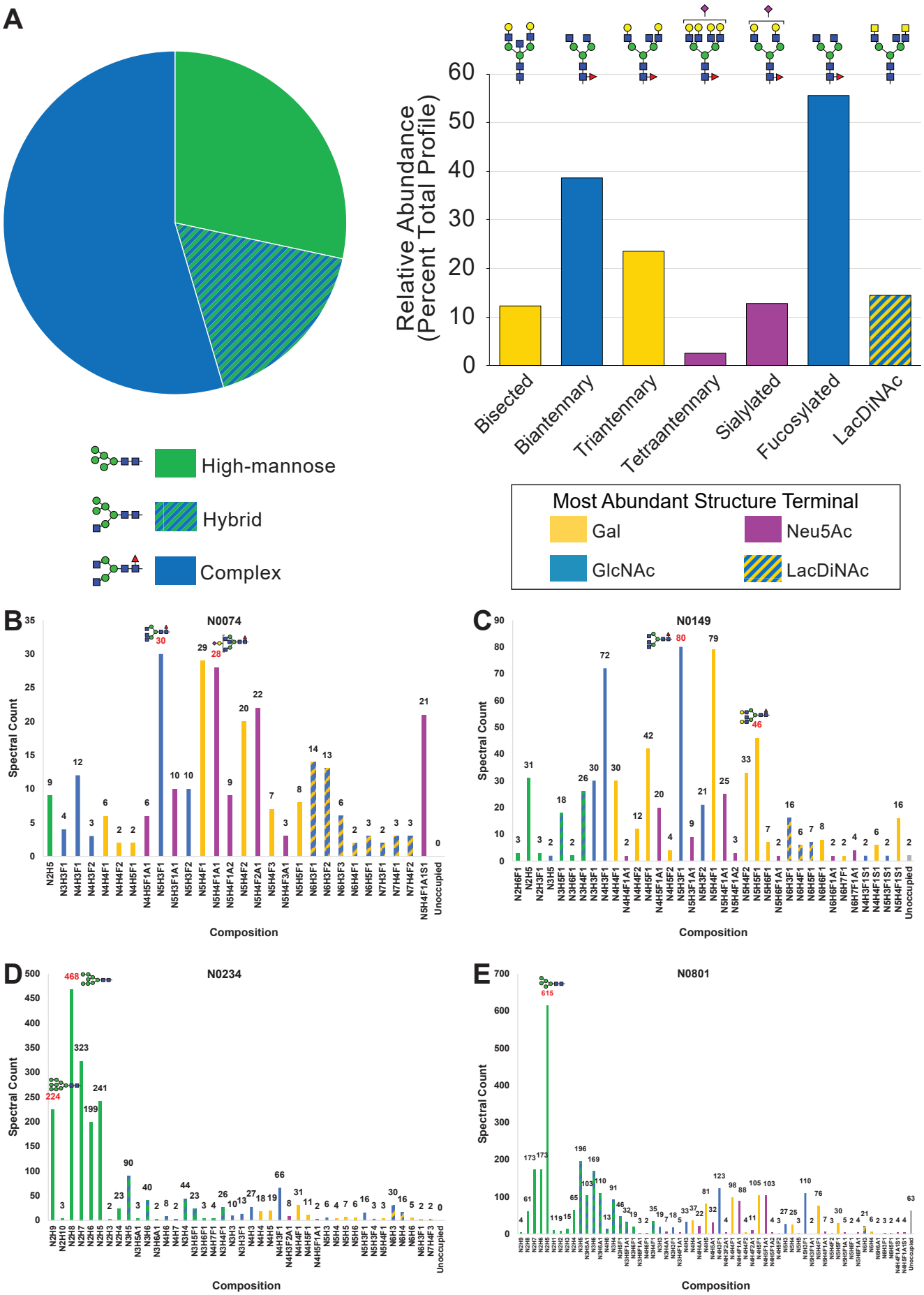


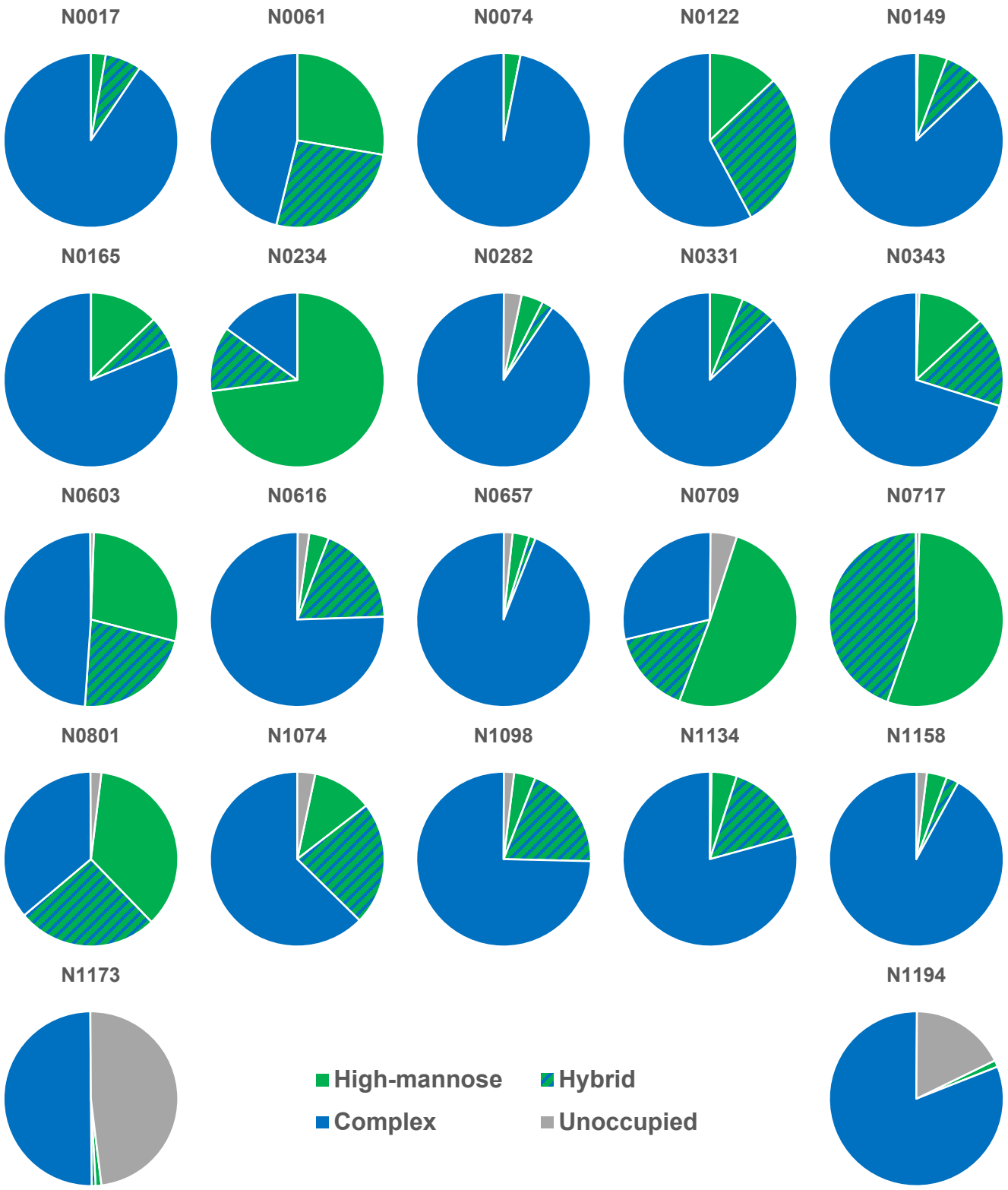
C

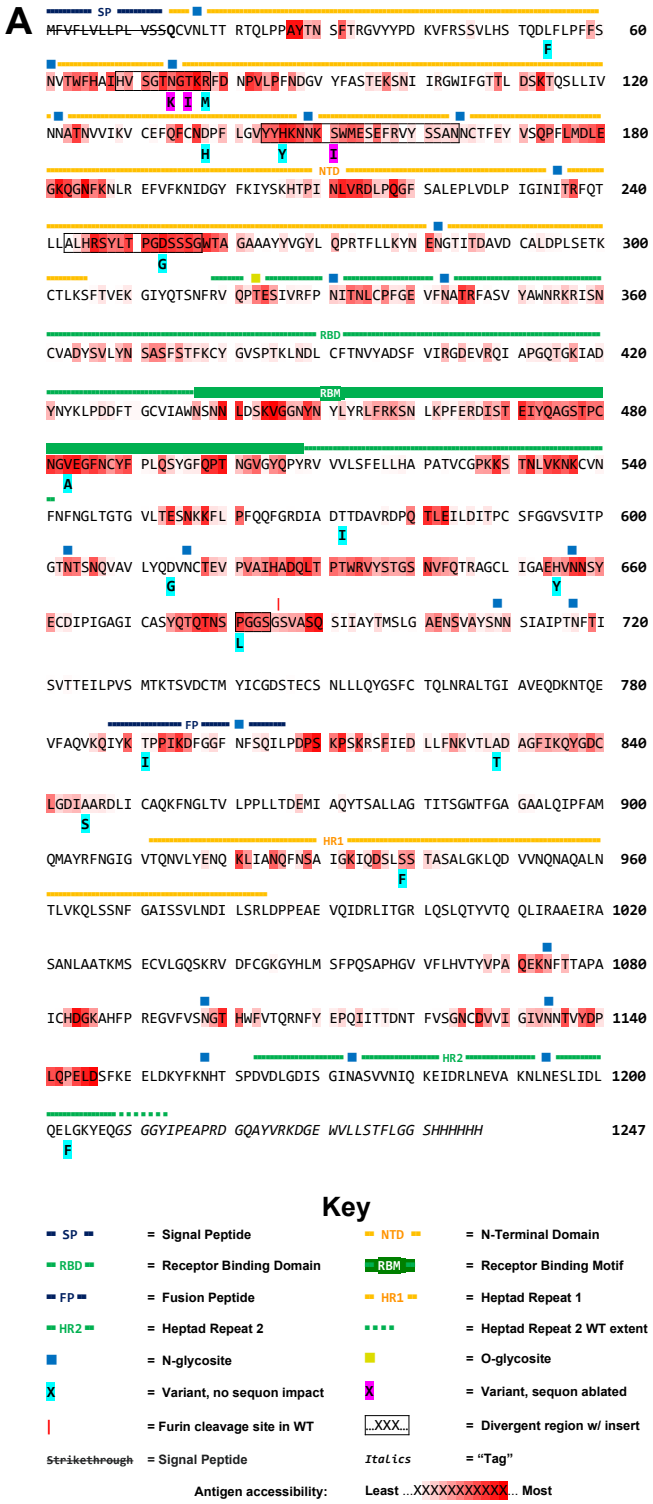


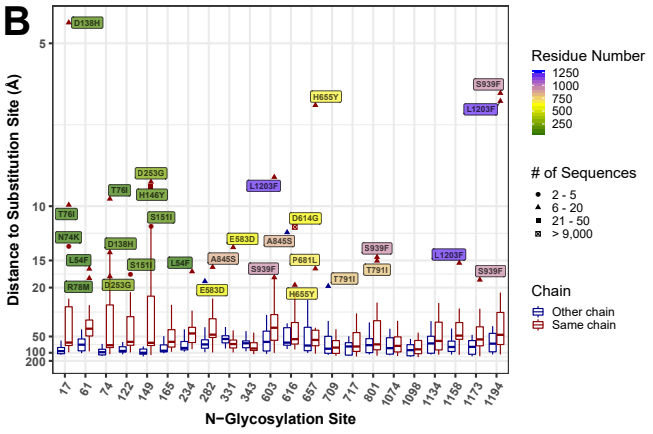
D

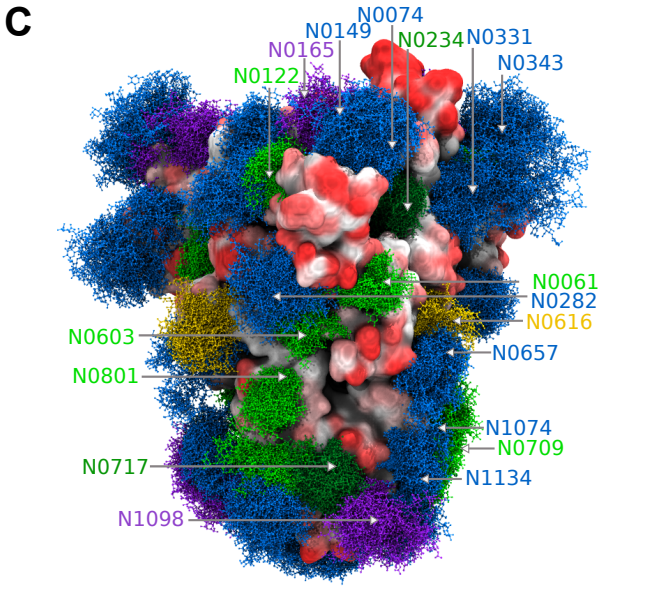


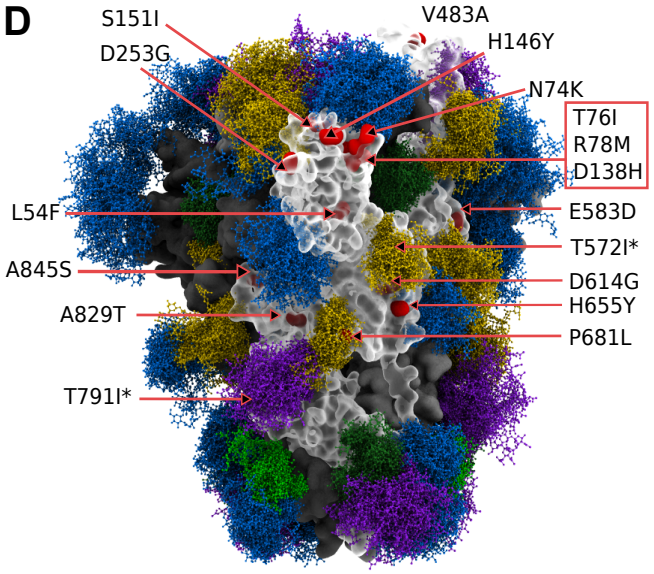

A
High-mannose
Hybrid
Complex
Relative Abundance (Percent Total Profile)
Bisected
Biantennary
Triantennary
Tetraantennary
Sialylated
Fucosylated
LacDiNAc
Most Abundant Structure Terminal
Gal
Neu5Ac
GlcNAc
LacDiNAc
B
N0074
Spectral Count
Composition
C
N0149
Spectral Count
Composition
D
N0234
Spectral Count
Composition
E
N0801
Spectral Count
Composition

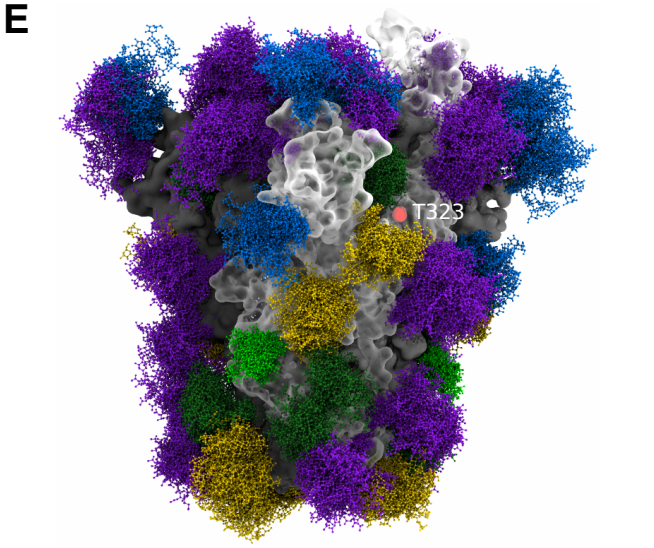

N0017
N0061
N0074
N0122
N0149
N0165
N0234
N0282
N0331
N0343
N0603
N0616
N0657
N0709
N0717
N0801
N1074
N1098
N1134
N1158
N1173
N1194
High-mannose
Hybrid
Complex
Unoccupied

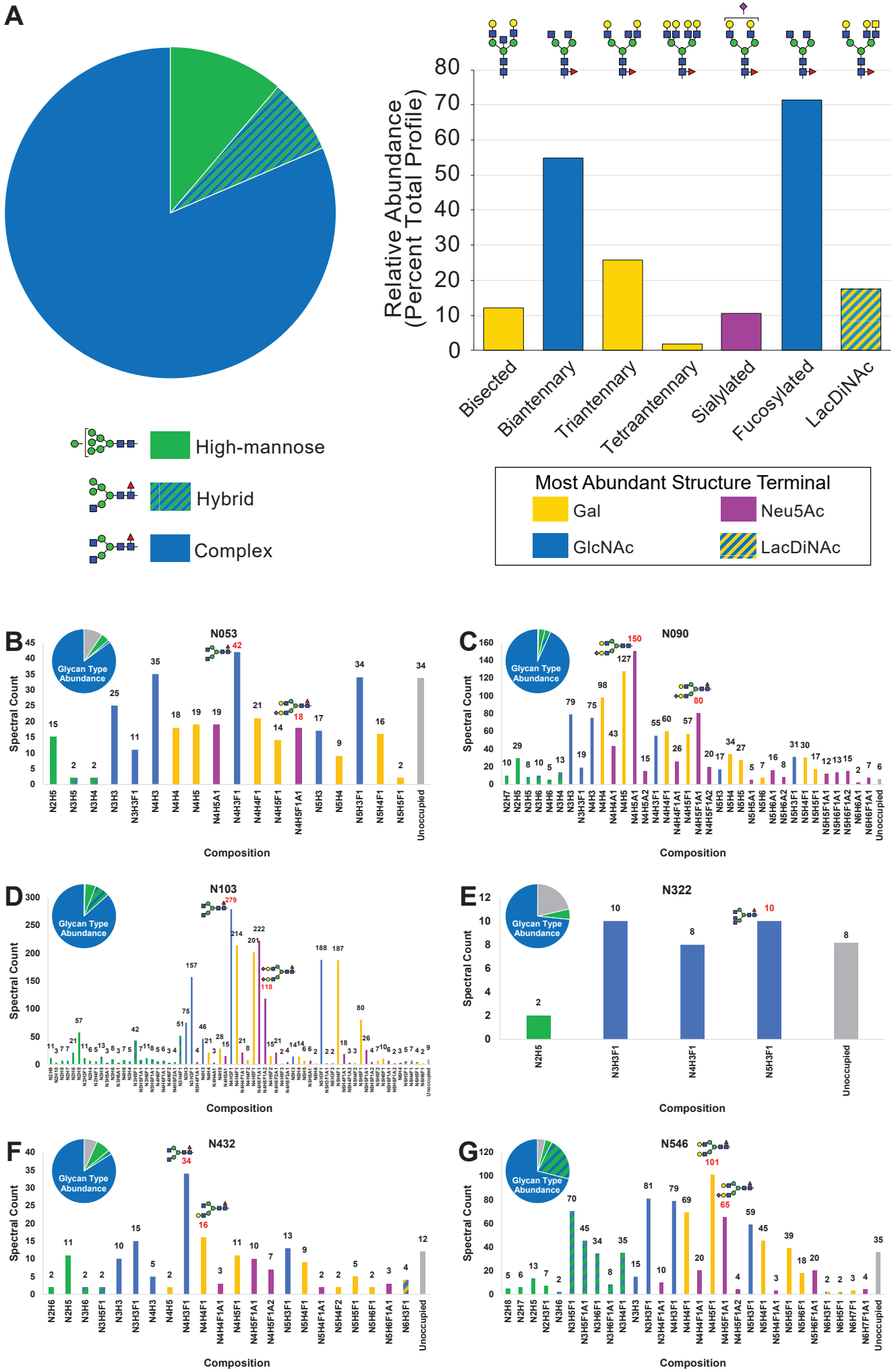

A
SP
NTD
RBD
RBM
FP
HR1
HR2
Key
SP = Signal Peptide
NTD = N-Terminal Domain
RBD = Receptor Binding Domain
RBM = Receptor Binding Motif
FP = Fusion Peptide
HR1 = Heptad Repeat 1
HR2 = Heptad Repeat 2
= Heptad Repeat 2 WT extent
= N-glycosite
= O-glycosite
= Variant, no sequon impact
= Variant, sequon ablated
= Furin cleavage site in WT
= Divergent region w/ insert
Strikethrough = Signal Peptide
Italics = "Tag"
Antigen accessibility:
Least ...XXXXXXXXXXX... Most

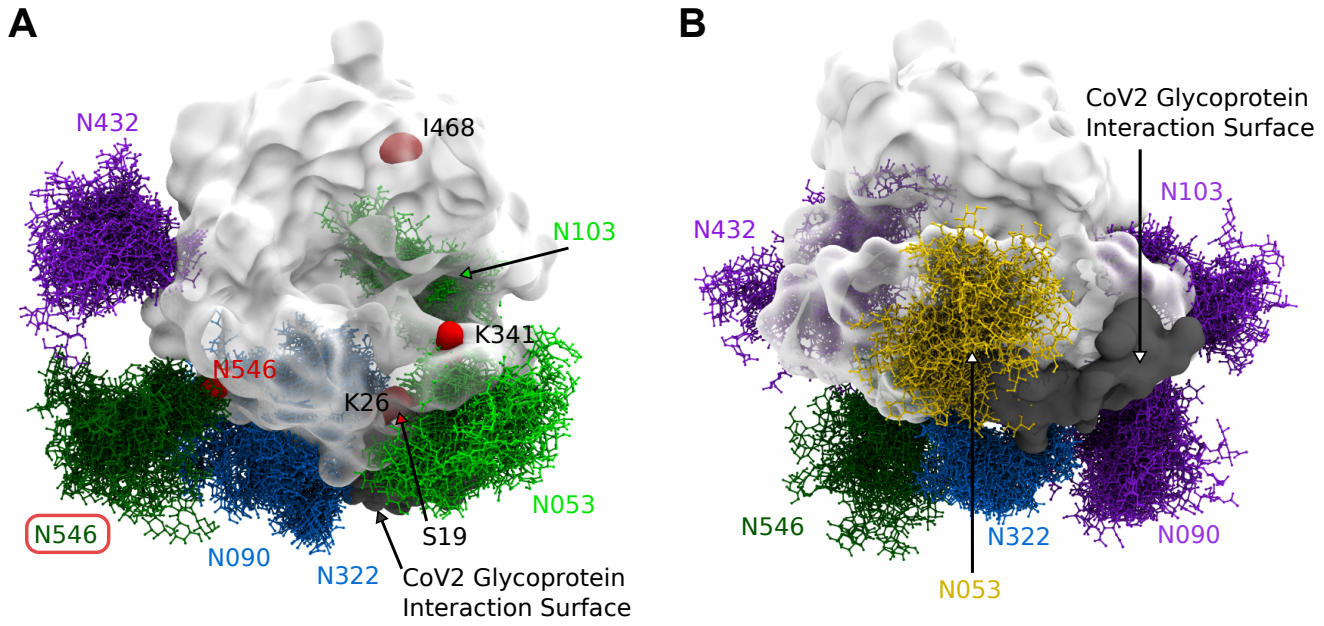

B
Distance to Substitution Site (A)
N-Glycosylation Site
Residue Number
1250
1000
750
500
250
# of Sequences
2 - 5
6 - 20
21 - 50
> 9,000
Chain
Other chain
Same chain

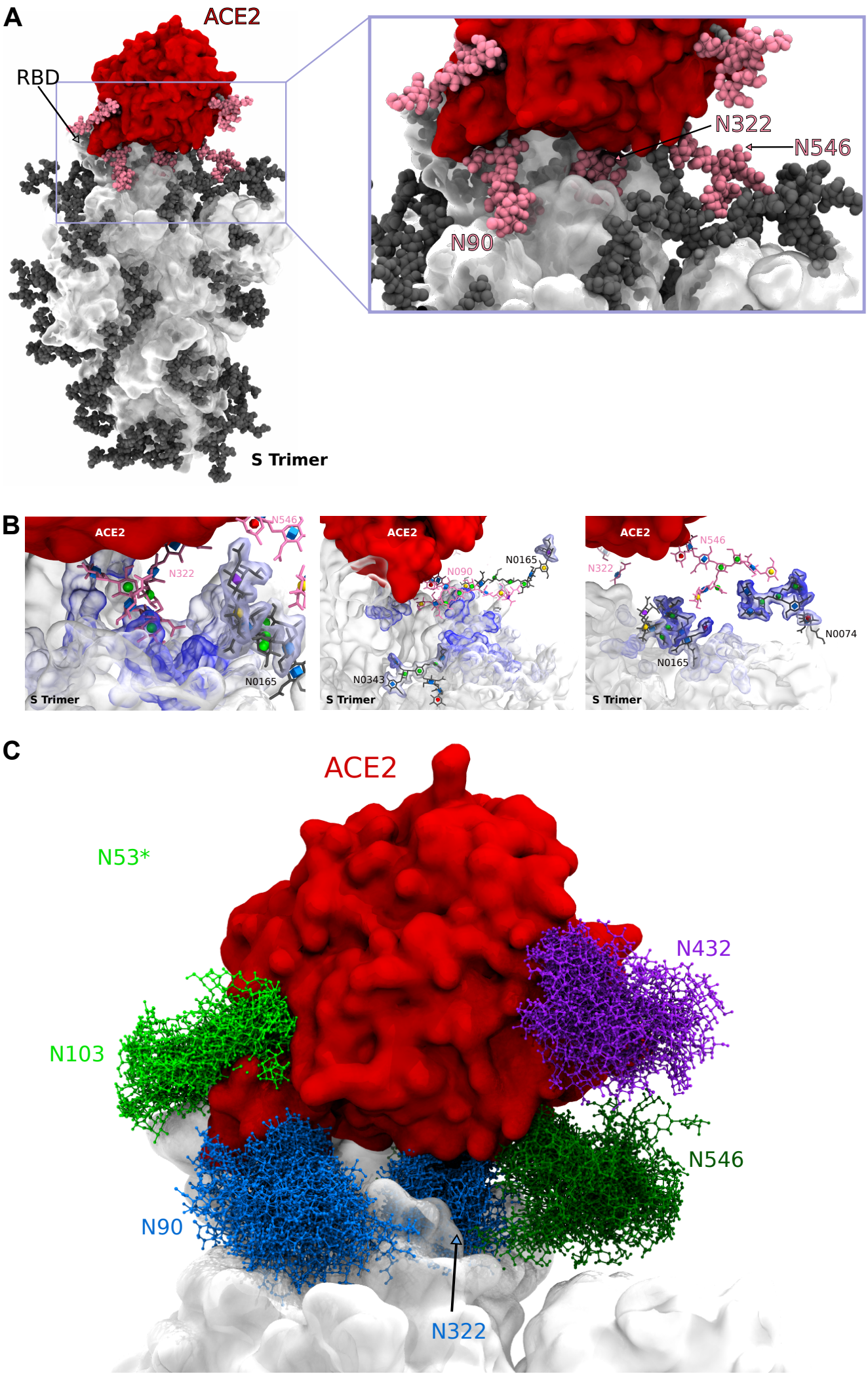

C
N0074
N0149
N0234
N0331
N0165
N0343
N0122
N0061
N0282
N0616
N0603
N0657
N0801
N1074
N0709
N0717
N1134
N1098


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


E
T323


A
High-mannose
Hybrid
Complex
Relative Abundance (Percent Total Profile)
Bisected
Biantennary
Triantennary
Tetraantennary
Sialylated
Fucosylated
LacDiNAc
Most Abundant Structure Terminal
Gal
Neu5Ac
GlcNAc
LacDiNAc
B
N053
Glycan Type Abundance
Spectral Count
Composition
C
N090
D
N103
E
N322
F
N432
G
N546


A
N432
I468
N103
K341
N546
K26
N053
N546
S19
N090
N322
CoV2 Glycoprotein
Interaction Surface
B
CoV2 Glycoprotein
Interaction Surface
N103
N432
N546
N322
N090
N053


A
ACE2
RBD
N322
N546
N90
S Trimer
B
ACE2
N546
N322
N0165
S Trimer
ACE2
N090
N0165
N0343
S Trimer
ACE2
N546
N322
N0074
N0165
S Trimer
C
ACE2
N53*
N432
N103
N546
N90
N322