Impact of Glycosylation on SARS-CoV-2 Infection and Broadly Protective Vaccine Design

Hsin-Yu Liao[a,1], Han-Yi Huang[a,g,1], Xiaorui Chen[a,1], Cheng-Wei Cheng[a], Szu-Wen Wang [a], Md. Shahed-Al-Mahmud[a], Ting-Hua Chen[a], Jennifer M. Lo[a], Yo-Min Liu[a], Hsiu-Hua Ma[a], Yi-Hsuan Chang[a], Ching-Yen Tsai[c], Pau-Yi Huang[c], Sui-Yuan Chang[d,e], Tai-Ling Chao[d], Han-Chieh Kao[d], Ya-Min Tsai[d], Yen-Hui Chen[f], Chia-Yen Chen[a], Kuang-Cheng Lee[a], Chung-Yi Wu[a], Jia-Tsrong Jan[a], Kuo-I Lin[a], Ting-Jen Rachel Cheng[a], Che Ma[a,2] and Chi-Huey Wong[a,b,2]

[a]Genomics Research Center, Academia Sinica, Taipei, 11529, Taiwan

[b]Department of Chemistry, Scripps Research, La Jolla, CA 92037, USA

[c]Institute of Molecular Biology, Academia Sinica, Taiwan

[d]Department of Clinical Laboratory Sciences and Medical Biotechnology, National Taiwan University College of Medicine, Taipei, Taiwan

[e]Department of Laboratory Medicine, National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan

[f]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan

[g]GIP-TRIAD Master's Program, National Taiwan University College of Medicine, Taipei, Taiwan

[1]H.-Y.L., H.-Y.H. and X.C. contributed equally to this work.

[2]To whom correspondence should be addressed:

Che Ma (cma@gate.sinica.edu.tw), Genomics Research Center, Academia Sinica, 128 Academia Road, Section 2, Taipei 115, Taiwan, or Chi-Huey Wong (wong@scripps.edu), Department of Chemistry, Scripps Research, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.

Abstract

Coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has resulted in more than 167 million confirmed cases and over 3 million deaths so far[1-4]. This global pandemic has led to great efforts directed toward the study of this virus and its infection mechanism as well as development of effective means to control this devastating infectious disease[5-8]. Like many other viral surface proteins, the trimeric SARS-CoV-2 spike (S) protein is heavily glycosylated with 22 N- and 2 O-glycosites per monomer which are likely to influence S protein folding and evade host immune response[9]. More than one million S protein sequences with over 1,000 sites of mutation in its 1,273 amino acids have been reported to the GISAID database, including the highly transmissible variant strains found in the UK and South Africa[10]. This high frequency of transmission and mutation is a major challenge in the development of broadly protective vaccines to control the pandemic. We have studied the impact of glycosylation on receptor-ligand interaction through evaluation of ACE2 and S protein expressed in different cell lines. Of different S protein glycoforms, the one expressed from lung epithelial cells, the primary cells for infection, has more complex-type glycans and higher binding avidity to the receptor as compared with the S protein from HEK293T cells which have more high-mannose or hybrid-type glycoforms. We also found that most of the S protein glycosites are highly conserved and the glycosites at positions 801 and 1194 are essential for viral entry. In addition, the RBD of S1 and the HR regions of S2 contain most of highly conserved sequences, and removal of each glycosite on pseudotyped SARS-CoV-2 virus for evaluation of the impact on structure and function provides insights into the design of broadly protective vaccines. In an effort to develop such universal vaccines, we found that mice immunized with monoglycosylated S protein (S_{mg}) elicited better antibody responses capable of neutralizing not only the wild type but also the variants from the UK and South Africa than those with the fully-glycosylated S protein (S_{fg}), and strikingly, S_{mg} vaccination provides better survival for hACE2 transgenic mice when challenged with lethal dose of SARS-CoV-2. Moreover, using single B cell technology, we isolated a monoclonal antibody from S_{mg} immunized mice which was also able to neutralize the wild type and variants, suggesting that removal of unnecessary glycans from S protein to better expose the highly conserved sequences is an effective approach to developing broadly protective vaccines against SARS-CoV-2 and variants.

Introduction

SARS-CoV-2 S protein is a trimeric class I membrane fusion glycoprotein, composed of two functional subunits, S1 and S2, which are responsible for receptor binding and membrane fusion, respectively[11] (**Fig. 1a**). Each S protein monomer contains 22 N-linked glycosylation sites and at least two O-linked glycosylation sites, which are important for proper folding and modulation of host protease processing[12,13], and for evading immune response by shielding specific epitopes from antibody neutralization[9]. The glycan profile at all 22 N-glycosites on the S protein has been identified, with 14 sites harboring predominantly the complex-type or hybrid-type glycans while the other 8 sites mainly the high-mannose type[13]. Among them, 52% of the glycans contain the fucose residue and 15% bear the sialic acid[13]. A consistent profile reported later indicated that the three C-terminal sites with mainly complex-type glycans (N1158, N1173, and N1194) are often further processed into multiple antennae through GlcNAcylation/galactosylation and/or sialylation compared to the other sites[14], while their specific functions are unknown[14]. It was shown that a trimeric, complex-type glycosylated S protein is more efficient in receptor recognition and viral entry than the immaturely high-mannsose variants derived either from GnTI$^{-/-}$ HEK293[15,16] or insect cell[17,18]. We also found that the S protein with sialylated complex-type glycans has higher avidity to the receptor than that with high mannose- or hybrid-type or mono-glycan, and removal of the sialic acid from the complex type would reduce the binding significantly (**Extended Data Fig. 1**). However, the role of each glycosite in the virus life cycle still needs to be elucidated, which raises another critical question as whether the S protein glycosylation in human lung epithelial cells is different from that in HEK293 cells which have been widely used[13,14], and how this difference may contribute to our understanding of viral structure and function as well as antigen presentation and vaccine design.

Pseudotyped SARS-CoV-2 variants containing single glycosite mutation

To understand the differential influence of each glycosite of S protein on the infection and replication of SARS-CoV-2, we established a working system by generating a full panel of 24 lentivirus-based pseudotyped virus variants (22 N- and 2 O-glycosites) for comparing the viral entry efficiency in five different cell lines. To quantitate the pseudotyped virus infection efficiency, we employed Quantikine® HIV-1 Gag p24 Immunoassay to determine the titer of HIV-1-based lentiviral supernatant by measuring the color intensity which corresponds to the level of p24, an essential capsid protein of HIV-1, and plotting it into a p24 standard curve. We first compared the titer of viruses containing either the wild type S or nine mutants, each with different truncation at the cytoplasmic tail, and found that the C-terminal 19 aa deletion construct produced the highest titer than others assayed (**Extended Data Fig. 2**). Based on this C-terminal truncated construct, we mutated every N-glycosite Asn to Gln and each O-glycosite Thr/Ser to Ala to minimize the impact on the structure of S protein. We then measured the titer of each of these 24 pseudotyped virus variants (**Fig. 1b**) and tested the interaction between the S protein and its cell receptor, human ACE2 (hACE2), which was

transiently transfected into 5 different cells: 293T-hACE2, Vero-E6-hACE2, A549-hACE2, Calu-1-hACE2 and Calu-3-hACE2cells (**Fig. 1c**).

Results of the 24 pseudotyped viruses showed that disruption of S protein glycosylation led to an overall impaired infectivity, and among them the three glycosite mutants (N616Q, N657Q and N709Q) located near the S1/S2 furin cleavage site displayed the minimal impact in 293T cells, a pattern somewhat consistent in other four cell lines (**Fig. 1c**). Notably, the four mutations located within the receptor binding domain (RBD), including N331Q, N343Q, T323A and S325A, all showed lower virus infections, especially in Vero E cells. Although the RLU-reduction profile varied between different cell lines in most regions, we identified three mutants with the most drastically reduced infectivity in all five cells, which are N122Q in the S1 domain, N801Q near the N terminus of S2 domain, and N1194 at the very C terminus of S2 domain (**Fig. 1c**). This finding underlined the particular importance of glycans at these specific sites, especially the two sites at S2 domain, N801 and N1194. To further understand their impact on S protein, we purified the recombinant N801Q and N1194Q S protein mutants to evaluate their properties. As expected, both mutations resulted in severely decreased protein yield (**Extended Data Fig. 3a, b)**, and N801Q mutant showed additional degradation bands which were usually not seen in other mutants (**Extended Data Fig. 3c)**. The N1194Q mutant, on the other hand, did not cause such degradation but existed as a monomer on size-exclusion chromatography (**Extended Data Fig. 3c**). These results demonstrated that N-linked glycans at specific positions differentially affected S protein integrity and the subsequent viral infectivity.

Site-specific glycan analysis of SARS-CoV-2 S protein from BEAS-2B and HEK293 cells.

Until now most of the published studies used SARS-CoV-2 S protein expressed from HEK293 cells, which are not the primary host in SARS-CoV-2 infection. The glycosylation profile of the S protein from HEK293 cells could be different from lung epithelial cells[19]; therefore, the information on glycosylation mediated mechanism may not be accurate. The S protein was therefore expressed from human lung epithelial cell line BEAS-2B (ATCC CRL-9609) and from human kidney epithelial cell line HEK293T, and the glycan profile was analyzed by LC-MS/MS **(Fig. 2a)**.

The N-glycan profile of S protein from BEAS-2B displays more abundance of complex-type glycan structures containing more sialic acids compared to that from HEK293T cells **(Fig. 2a, b)**. There are two sites on BEAS-2B displaying principally the oligomannose-type (N234 and N717), and a mixture of oligomannose- and complex-type glycans can be found at sites N61, N122, N801 and N1074. It is worth noting that N234, N717 and N801 display an abundance of Man7-Man8 high-mannose structures, suggesting that the mannosidase processing is stalled when it enters the cis-Golgi **(Extended Data Fig. 4)**. The remaining 16 sites are highly processed complex-type glycan structures. In addition, almost all the sialylated complex-type glycans apart from N17 and N149 have high degree of fucosylation, indicating that N17 and N149 are not processed by fucosyltransferase 8 in the medial-Golgi. Furthermore, with the Oxford classifications our data demonstrated that N74, N149, N1098, N1158, and N1194, dominated by complex-type glycans, were more highly processed (i.e., multiple

antennae) and elaborated (i.e., galactosylation, fucosylation and sialylation) than other sites **(Extended Data Fig. 4)**. In general, we observed a high degree of sialylation for all 22 sites. Notably, there are more enriched sialic-acid-containing complex-type N-glycans from BEAS-2B (51%) than from HEK293T (35%) or HEK293E (26%) **(Extended Data Table 1, 2, 3).** The N-glycosites sites with predominantly highly processed sialylated complex-type glycans include N74, N149, N165, N282, N331, N343, N657, N1098, N1158 and N1194, while N1134 and N1173 have decreased sialic acids **(Fig. 2a, Extended Data Table 1, 2, 3)**. The glycan occupancy on the two N-glycosites (N331 and N343) of RBD are highly sialylated in Beas-2B, while the degree of sialylation in 293T is significantly lower. **(Fig. 2a)**

In contrast, although the complex type glycans still occupied dominantly (60%) on HEK293T N-glycosite, the N-glycan profiles showed that greater amount of hybrid type glycans was detected on the HEK293T than Beas-2B. The N-glycan profiles of HEK293T cells showed that two of the sites (N234, and N717) displayed more than 50% high-mannose glycans. We observed Man5GlcNAc2 structure as a predominant oligomannose-type glycan across the protein, and N234 and N717 were dominated by the high mannose structures Man8GlcNAc2 and Man7GlcNAc2, respectively **(Fig. 2a, Extended Data Fig. 5, Extended Data Table 2)**, consistent with the recent report [13]. Seven of the N-glycosites (N61, N122, N603, N709, N801, N1074 and N1098) contained a combination of high mannose, hybrid-type and complex-type glycans. There are 8 N-glycosites (N74, N165, N282, N331, N343, N1158, N1173, and N1194) with highly processed complex-type glycan structures, and only one site (N717) had 65% high mannose and 28% hybrid structures. Furthermore, the three C-terminal sites (N1158, N1173, and N1194) contain more complex-type glycans than other sites **(Fig. 2a, c, Extended Data Table 2)**. Overall, the data showed that the N-glycan profiles of HEK293T displayed more high mannose and hybrid-type N-glycans and underprocessed glycan antennae (i.e., under-GlcNAcylation/galactosylation and under-sialylation), and the S protein from Beas-2B contains more sialylated complex-type N-glycans that facilitate binding to the receptor. Since most of the published studies used the SARS-CoV-2 S proteins produced in human embryonic kidney cell lines (HEK293T or HEK293F)[9,13,14], but not the lung epithelial cells, which are the primary cell for SARS-CoV-2 infection, the different glycosylation profile of S protein from lung epithelial cells may have major impact on viral infection and antigen presentation in the immune system.

Analysis of S protein glycosylation and its influence on the conserved epitopes

The profile of S protein glycosylation further encouraged us to conduct the structural analysis to understand which residues are exposed at the protein surface but shielded by the glycans, and whether these residues are more prone to antigenic variation during SARS-CoV-2 evolution. Multiple alignment results using 1,117,474 of SARS-CoV-2 S protein sequences extracted from GISAID[10] (version: Apr. 18, 2021) revealed that the NTD domain has a much lower percentage of conserved residues (79.1%) than the RBD (93.3%) or the S2 region (93.81% on average), in which a mutation rate >0.1% is considered not conserved (**Fig. 2e, Extended Data Fig. 7, Extended Data Table 4**). Indeed, the RBD, Fusion Peptide (FP), Heptad Repeat 2 (HR2) contain higher percentage of

conserved epitopes (33.63%, 42.11% and 62.75%) in their exposed residues than other regions (**Extended Data Fig. 8, Extended Data Table 4**), and the conservation analysis of S protein glycosylation sites reinforces the idea that the sites at the RBD (N331, N343 and S325), FP (N801) and near HR2 (N1074, N1098, N1134, N1158, N1073, and N1194) are extremely conserved (**Extended Data Fig. 9a)**, all of which resulted in moderate to severe infection defects when mutated, as demonstrated in the pseudotyped virus studies (**Fig. 1c**). Mapping of glycan coverage and sequence conservation statistics on S protein structure displayed a consistent pattern of unburied clusters and highly conserved but glycan shielded regions, including the lower side of RBD, the non-complex type glycan belt nearby FP and the HR1 (**Fig. 2b-e**). However, when we focused only on the exposed residues, we found that while NTD may not be a good target for epitope study due to the overall high mutation rate, its conservation percentage raised a lot when only including the glycan shielded areas (82.9%) compared to unshielded areas (63.3%) (**Extended Data Fig. 9**). Moreover, an overall increase of conservation rate between glycan-shielded and unshielded residues for HR2 was observed, in which all its 14 shielded residues are highly conserved (100%) compared to only 81.8% for the unshielded ones (**Extended Data Fig. 9b**). These results again highlighted the significance of S protein glycosylation both functionally and evolutionarily, bringing forth an idea that removal of the S protein glycan shield may expose more highly conserved regions for eliciting neutralizing antibodies in a more efficient way.

Vaccination with monoglycosylated S protein (S_{mg}) provides better protection against SARS-CoV-2 infection *in vivo*

Based on the study of glycosylation in SARS-CoV-2 infection as described above, we decided to design S protein vaccines through glycoengineering to remove the glycans that may interfere with the presentation of conserved epitopes, with the hope to develop more broadly protective vaccines against SARS-CoV-2 variants. Our initial attempt to remove S protein glycans by combining glycosite mutations led to a dramatically reduced expression of folded S protein, probably due to the negative impact on protein folding through deletion of multiple glycosites (**Extended Data Fig. 3**). This problem however may not be encountered with the use of mRNA as vaccine, as when the glycoengineered RNA is inside the antigen presenting cells, the translated and unfolded S protein mutant may still be processed for presentation to T cells, and work is in progress to validate this approach. An alternative strategy was to generate S protein with all high-mannose glycans using $GnTI^{-/-}$ cells and perform the glycan trimming with enzymes after the S protein is fully folded, secreted, and purified[20,21]. Endoglycosidase H (Endo H) was used to cleave the high-mannose glycans to N-Acetylglucosamine (GlcNAc) at each N-glycosylation site (**Extended Data Fig. 10a, b, c**), generating the monoglycosylated form of S protein (S_{mg}) (**Fig. 3a**), which was further confirmed to be near 100% pure and fully soluble **(Extended Data Fig. 10 d, e**). This modified S_{mg} and the fully glycosylated S protein (S_{fg}) were used to immunize BALB/c mice (n=5) with two injections (20ug antigen with 20ug aluminum hydroxide) (**Fig. 3b, Extended Data Fig. 11**). S_{fg} is similar to the

immunogens used in many current COVID-19 vaccines which were either approved or in clinical trials, including the insect cell expressed S protein vaccines from companies such as Sanofi[22] and Novavax[23], the human cell expressed recombinant S-2p vaccine from Medigen[24], the adenovirus-based vaccines from AstraZeneca[25] and J&J[26], and mRNA vaccines developed by Pfizer-BioNTech[27] and Moderna[28].

The S_{mg} immunized mice induced superior humoral immune response than S_{fg}, with a 2.1-fold higher anti-S IgG avidity (~1.2 x 10^5) than that by S_{fg} (~5.6 x 10^4) in week 6 after the first immunization (**Fig. 3b, c**) and a stronger neutralizing antibody response in the neutralization test based on inhibition of cytopathic effect (CPE) assay (**Fig. 3d**). The protective efficacy by vaccination of S_{mg} vaccine against SARS-CoV-2 (hCoV-19/Taiwan/4/2020) in Syrian hamster model[29] (n = 5) was higher with less reduction in body weight as compared to the S_{fg} and PBS groups (**Fig. 3b, e**). The virus titers in the lungs of both S_{fg} and S_{mg} vaccinated hamsters were found to be 10^4-fold reduction compared with the PBS control, while no virus staining with anti-N antibody was detected in lung of both S_{fg} and S_{mg} vaccinated hamster (**Fig. 3f, i)** The pathological score also indicated that less lesion was found in S_{mg} immunized hamster lung. **(Fig. 3j, Extended Data Fig. 12**). Since hamsters only showed mild-to-moderate sickness upon SARS-CoV-2 infection, we also conducted vaccination experiments in the severe-disease model using highly susceptible CAG-hACE2 transgenic mice[30] (n = 3) (**Fig. 3b**). SARS-CoV-2 (1x10^3 TCID50) was used to challenge the mice intranasally, and the S_{fg} group failed to protect mice from severe disease, as a rapid decease of body weights (up to 20% loss of weight at 7 dpi for all) was observed, only slightly better than the adjuvant control (**Fig. 3k**). In contract, the S_{mg} group exhibited only mild weight loss (~ 5.75% at 7 dpi) with an 100% survival rate compared to the S_{fg} and adjuvant group (0% at 7 dpi) (**Fig. 3k, l**). Moreover, unlike such a large difference in disease progression between these groups, the neutralizing antibody titers of S_{fg} vaccinated hACE2 mice were only 2.8-fold less than that of the S_{mg} group, indicating that the humoral immune response may not be the only factor contributing to vaccine efficacy (**Fig. 3g, h**). The exciting findings that S_{mg} induced an all-improved immune response and protection again proved the importance of S protein glycosylation in vaccine design, and removal of its glycan shield has successfully generated an immunogen which exposes more conserved residues for the host cells to elicit efficient immune response.

In addition to the improved efficacy, this glyoengineered S_{mg} vaccine was also found to induce a broader protection against SARS-CoV-2 variants including the currently dominant D614G strain, the UK variant (B.1.1.7), and the South Africa variant (B.1.351)[31-35] (**Fig. 4a)**. The B.1.1.7 variant, which originated from UK and now is circulating dominantly in North America and Europe, showed 2-to-3-fold reduction in neutralizing activity in the serum of patients vaccinated with AstraZeneca, Pfizer-BioNTech and CoronaVac vaccines and resistance to some anti-RBD monoclonal antibodies[33,34,36-38]. The South Africa B.1.351 variant is highly resistant to several therapeutic antibodies and even the convalescent sera, and exhibited reduced protection by many approved vaccines[35-37,39-41]. The pseudotyped viruses of these variants were then tested for both S_{mg} and S_{fg} immunized sera, and the S_{mg} vaccine showed a superior neutralization ability towards the D614G,

B.1.1.7 and B.1.351 strains, with the pNT50 being 1.4-, 2.7- and 1.5-fold better than S_{fg} (**Fig. 4c, d**). Compared with S_{fg}, the S_{mg} vaccine also elicited better spike-specific IgG titer to not only spike protein from WT (Wuhan/WH01/2019), but also from the variants D614G, B.1.1.7, B.1.351, bat CoV RaTG13 and SARS-CoV-1 (**Fig. 4a, b**). Finally, in the SARS-CoV-2 real virus plague reduction neutralization assay, the S_{mg} vaccine again displayed a better neutralizing antibody response to D614G and B.1.1.7 variants as compared with S_{fg}, with 2.0- and 1.4-fold increase, respectively (**Fig. 4e, f**). All these data supported that removal of unnecessary glycans from S protein can effectively aid in the development of broadly protective vaccines and also increase the chance of identifying broadly neutralizing antibodies against SARS-CoV-2 infection.

Identification of a broadly neutralizing antibody from B cells of mice immunized with S_{mg}

We further examined the frequency of spike-specific B cells ($CD3^{-}CD19^{+}S^{+}$) from the spleen of mice immunized with S_{fg} or S_{mg} by fluorescence-activated cell sorting (FACS) analysis. Compared with S_{fg} immunized mice, S_{mg} immunized mice generated significantly higher frequency of S-specific B cells (**Fig. 5a, b**). We next sorted S-specific B cells from S_{mg} immunized mice for single B cell screening of monoclonal antibody (mAb) against S protein, which led to one resultant chimeric mAb, m31A7, with an excitingly high affinity ($10^{-10} \sim 10^{-11}$ M) in binding to HEK293T cells that express S protein from all the tested SARS-CoV-2 strains including the original Wuhan (Wuhan/WH01/2019), D614G, UK (B.1.1.7), and South Africa (B.1.351) variants, as measured by FACS (**Fig. 5c**). Additionally, we found that m31A7 binds to the RBD of S protein as ELISA results showed that m31A7 binds to full length, S1 and RBD, but not S2, of SARS-CoV-2 (WH01) S protein, indicating a potential interference with ACE2 recognition during virus infection (**Fig. 5d**). Indeed, m31A7 was successfully proven to neutralize both WH01 and various SARS-CoV-2 pseudotyped virus variants (D614G, B.1.1.7, and B.1.351) at sub-picomolar IC50 (**Fig. 5g**). The findings that S protein binding and virus neutralization by this mAb identified from S_{mg} immunized mice were not affected by S protein mutations of these variants reinforced the idea that S_{mg} immunization may increase the chance of eliciting broad-spectrum antibodies (**Fig. 5c, g**). Moreover, binding kinetics using bio-layer interferometry (BLI) showed that m31A7 bound to S protein and to RBD with a dissociation constant (K_D) value at sub-picomolar and nanomolar avidities, respectively. Compared to the binding with full-length S protein, m31A7 IgG bound to RBD with faster dissociation rate (k_{off}) ($6.39 \times 10^{-6} \pm 9.75 \times 10^{-7}$ s^{-1} to S and $4.56 \times 10^{-4} \pm 1.03 \times 10^{-6}$ s^{-1} to RBD, respectively) (**Fig. 5e**), while m31A7 Fab bound to both molecules with comparable binding affinities (K_D) ($2.19 \times 10^{-10} \pm 8.89 \times 10^{-12}$ M to S and $5.84 \times 10^{-10} \pm 2.91 \times 10^{-12}$ M to RBD) (**Fig. 5f**). These results have not only suggested that bivalent m31A7 can stabilize the interaction with S protein, thereby providing high avidity, but also highlighted its high affinity comparable to or higher than many other mAbs identified recently[6,42,43]. Taken together, the monoglycosylated S protein (S_{mg}) vaccine has the potential to induce broadly neutralizing antibodies with high affinities.

In conclusion, S protein glycosylation of SARS-CoV-2 is significant for virus infection and is complicated among different glycosites with differential glycan profile from different cell lines. The

two important glycosites identified in this paper, N801 located on the non-complex-type glycan belt in the S2 domain, and N1194 near the very C-terminus of the fusion-related HR2 region, have taken us one step further towards the mechanistic understanding of S protein. These findings together with the analysis of its glycoform abundance, glycan shielding, sequence variations and their mutual correlations further brought forth the idea of a monoglycosylated S protein vaccine, S_{mg}, from which essentially all glycan shields are removed, presenting a superior immunogen where all peptide sequence in S protein may be used as epitopes, and proven to provide better protection than conventional protein vaccines. The B cells from S_{mg} immunization paved the way for the identification of a high-avidity broad-spectrum monoclonal antibody capable of neutralizing both UK and South Africa virus variants, with many more to be explored. The potential impact of surface protein glycosylation on the viral infection and immune response to the current circulating SARS-CoV-2 and future variants shall never be underestimated and the collective knowledge of this virus will benefit the accelerated development of a broad-spectrum COVID-19 vaccine.

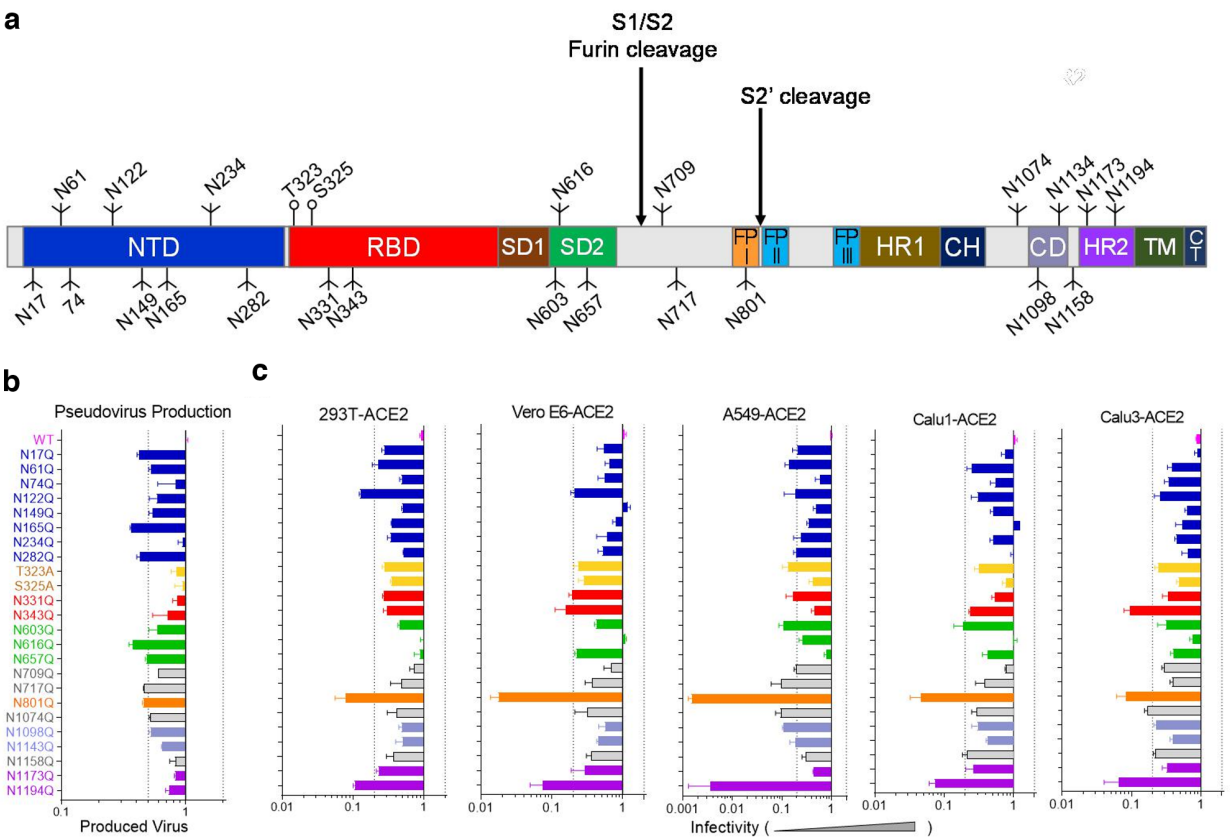


Fig. 1. **Impact of glycosylation of SARS-CoV-2 S protein and ACE2 receptor on virus packaging and virus entry. a,** Schematic overview of the full-length SARS-CoV-2 S protein primary structure colored by domain: N-terminal domain (NTD, 16–291), receptor binding domain (RBD, 319–528), SD1 (or CTD1, 529-591), SD2 (or CTD2, 192-685), furin cleavage site (S1/S2 at 685), fusion peptide I (FPI, 788-806), fusion peptide II (FP II or internal fusion peptide, 816-834), fusion peptide III (875-901), heptad repeat 1 (HR1, 902-986), central helix (CH, 987–1034), connecting domain (CD, 1080–1135), heptad repeat 2 (HR2, 1163–1210) domain, transmembrane domain (TM, 1214–1234), and cytoplasmic tail (CT, 1235–1273)[13,44]. Representative icons for N-glycan (Y) and O-glycan (I) are also marked with amino acid number. **b,** The amount of pseudovirus packaged by SARS-CoV-2 S proteins containing indicated glycosylation mutations. Individual sites were marked with the colors corresponding to the domains shown in **a** and O-glycosylation sites were marked in yellow. The virus titer was quantitated by measuring p24 levels in the pseudovirus packaged with wild-type SARS-CoV-2 spike or individual glycosylation mutants. The p24 levels measured with the pseudovirus packaged with the SARS-CoV-2 spike glycosylation mutants were normalized against the p24 level measured in the pseudovirus packaged with the wild-type SARS-CoV-2 spike protein (defined as 1). Values shown are mean ± standard deviation of the normalized p24 levels from three independent experiments. **c,** Impact of glycosylation on the infectivity of pseudovirus containing wild-type spike or spike with indicated glycosite mutations. Equal virus titers of the pseudovirus packaged by SARS-CoV-2 spike protein containing indicated glycosite mutations were used to infect the cells (HEK293T, Vero, and lung cells A549, Calu-1, Calu-3) expressing human ACE2 for 48 h and the infection efficiency was measured by luciferase activity. The relative luminescence unit (RLU) was measured

and normalized against the RLU measured in the cells infected with pseudovirus packages with wild-type SARS-CoV-2 spike. Values shown are mean ± SD.

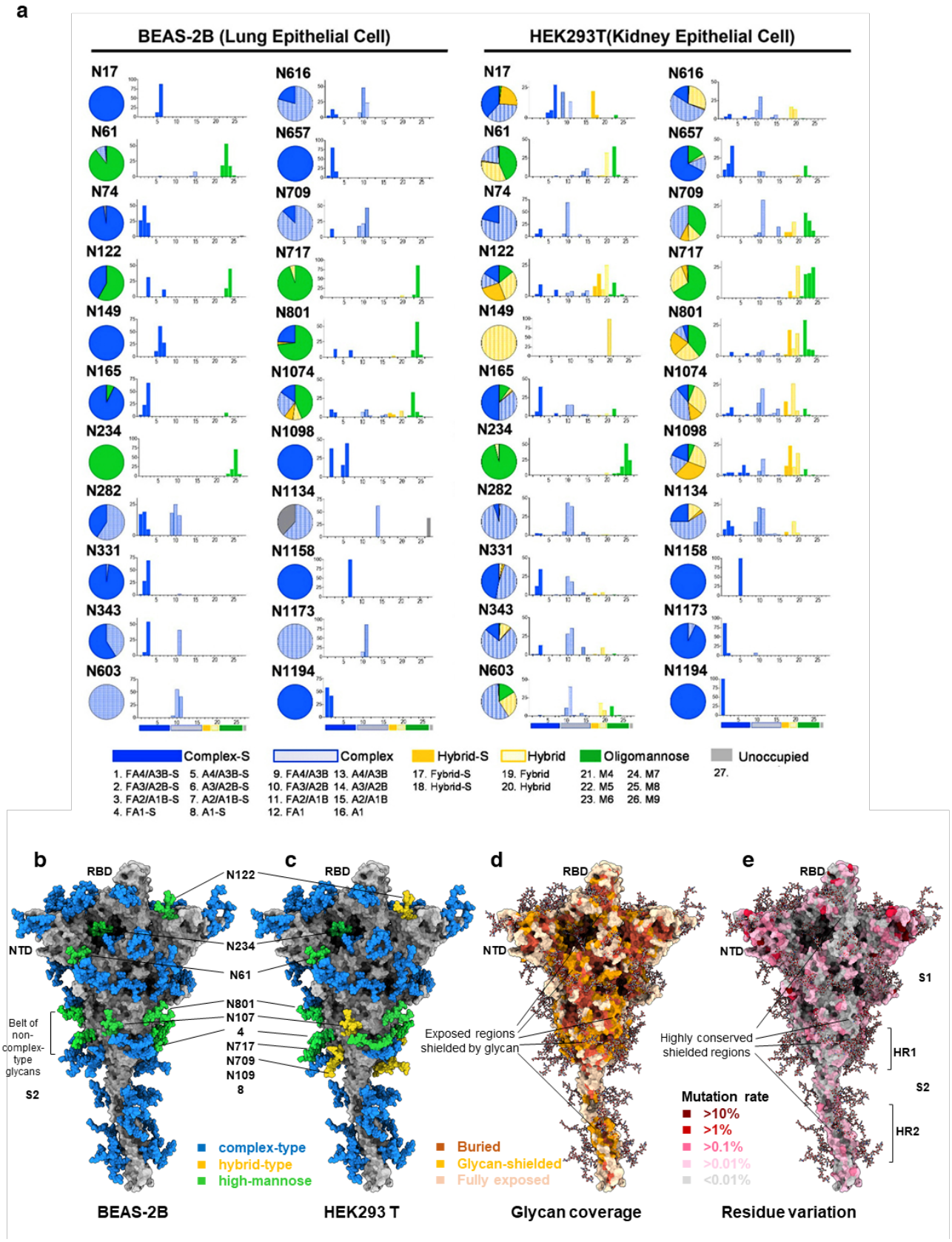


Fig. 2. **Site-specific glycan analysis of the SARS-CoV-2 S protein reveals different patterns in different cell lines. a**, N-glycosylation analysis of recombinant S protein purified from either BEAS-2B (lung epithelial cells, left) or HEK292T (human embryonic kidney cells, right) were measured by quantitative mass spectrometry analysis. The N-glycans are categorized into complex-S (complex-

type with sialic acid, solid blue), complex (complex-type with no sialic acid, light blue), hydrid-S (hybrid-type with sialic acid, solid yellow) and hybrid (hybrid type, light yellow), oligomannose in green and unoccupied in grey[13]. The pie chart highlights the overall population of these 6 categories in each glycosites while each bar chart presents the percentage (y axis) of different glycan types (x axis) at individual N-linked glycosylation sites. **b**, Mapping of glycan profile on S protein 3D structure (initial reference PDB 6VSB) modeled by CHARMM-GUI[2] and OpenMM[3] programs, with the glycans shown in spheres colored by the highest-abundance type according to BEAS-2B data. **c**, the same glycan profile mapping based on HEK293T data. Several N-glycosites of interest are labeled. A belt of non-complex-type glycans in the middle of S2 domain is also highlighted. **d**, Mapping of glycan coverage on S protein 3D structure was shown in the same view as in **b** and **c**, with dark orange for buried residues, light orange for exposed but shielded-by-glycans residues, and white for fully exposed residue. Glycan-shielded unburied regions are highlighted. **e**, Mapping of sequence variation coverage on S protein 3D structure in the same view as in **d**, colored by mutation rate >10% (crimson), >1% red, >0.1% pink, >0.01% light pink and <0.01% grey. Several highly conserved regions that overlap with glycan-shielded unburied regions are highlighted, including the lower side of RBD, the middle region of S2 and heptad repeat HR2. Images were made by ChimeraX[45]

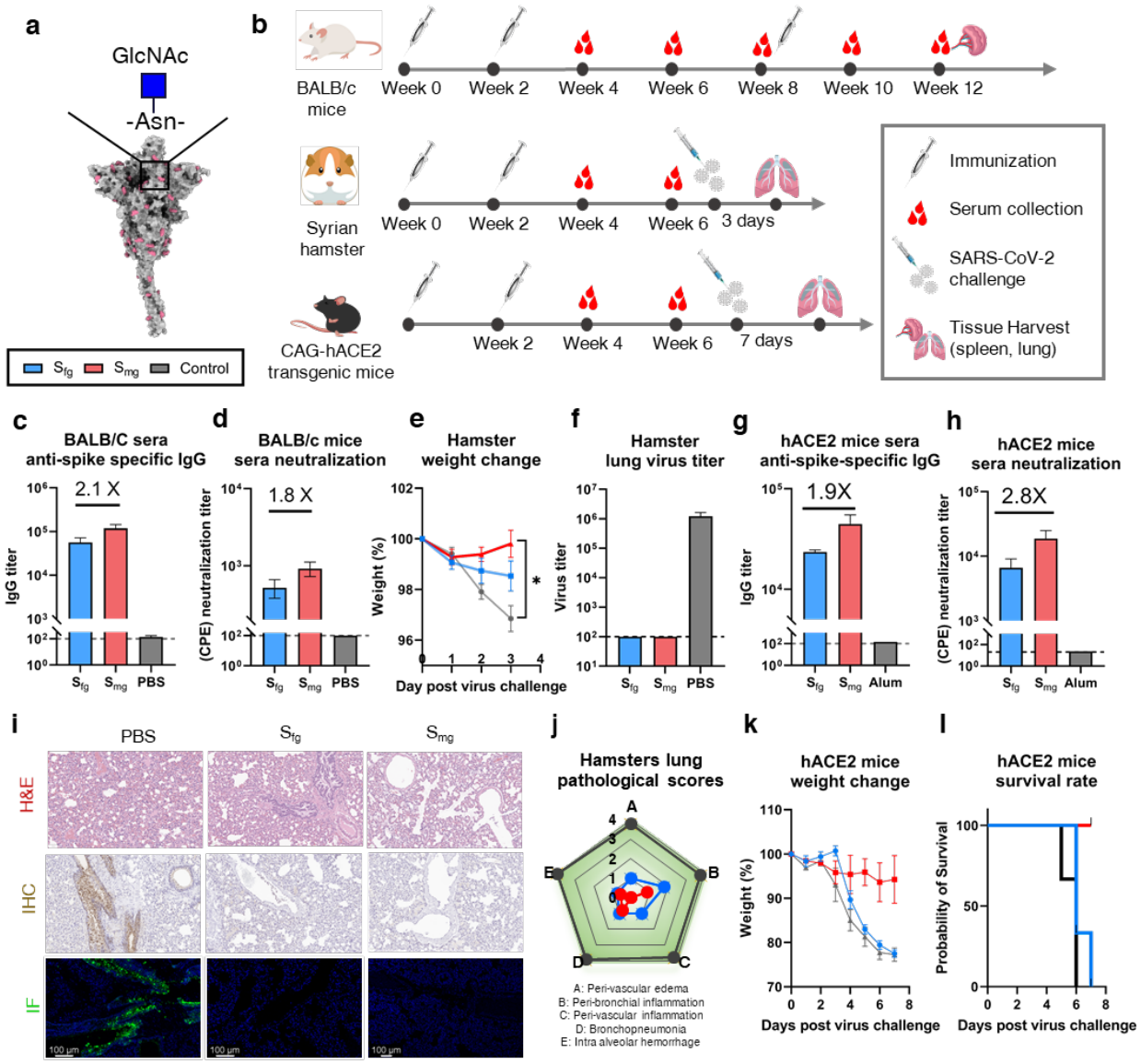


Fig. 3. S_{mg} vaccination provides better protection against SARS-CoV-2 infection *in vivo*. a, Structural model of S_{mg} generated in the same way as Fig. 2b, with single GlcNAc at each N-glycan site colored in pink. **b**, immunization and challenge schedule for three different animal models, BALB/c mice, Syrian hamster and hACE2 transgenic mice, using three groups of immunogen, S_{fg}(blue), S_{mg} (red) and control (grey). **c,** Hormonal immune response in BALB/c mice shown as sera of anti-S protein-specific IgG endpoint titer analyzed by ELISA (n = 5). **d,** Neutralization titer of BALB/c mice sera against wild type SARS-CoV-2 (Wuhan/WH01/2019) by CPE assay (n = 5). **e,** Weight change of Syrian hamsters (1-3 dpi) after challenge with wild type SARS-CoV-2 (n = 5). **f,** Virus titers in the lung of challenged hamsters as in (e) (n = 5). **g.** Sera from hACE2 mice at 6 weeks post initial vaccination were examined for anti-S protein specific IgG. **h,** Sera from hACE2 mice at 6 weeks post initial vaccination were examined for neutralizing antibody titer by CPE assay. **i,** Representative histopathology and immunostaining. First row: H&E staining of the lung in SARS-CoV-2-infected hamsters (3 dpi). Scale bar: 50 μm. Second row: immunohistochemistry (IHC) staining of lung tissues with SARS-CoV-2 N-specific polyclonal antibodies. Anti-N was used for

virus detection which indicated by brown dots. Scale bar: 50 μm. Third row: immunofluorescence staining of lung tissues with SARS-CoV-2 N-specific polyclonal antibody. Anti-N was used for virus detection which indicated by green dots. Blue dots: DAPI. White scale bar: 100 μm. **j,** Hamster lung pathological scores. A: Peri-vascular edema, B: Peri-bronchial inflammation, C: Peri-vascular inflammation, D: Bronchopneumonia, E: Intra alveolar hemorrhage. **k,** Weight change of CAG-hACE2 transgenic mice (1-7 dpi) after challenge with wild type SARS-CoV-2 (n = 3). **l,** Transgenic mice survival rate after challenge as in (k). (n = 3). Data are means ± SEM (standard error of the mean). Comparisons are performed by Student's t-test (unpaired, two tailed). The fold of increase in binding and neutralization of the antibody elicited by S_{mg} over S_{fg} vaccines is indicated above the bars. Blue dots, animals immunized with S_{fg}; Red squares, animals immunized with S_{mg}; Gray triangles, control group animals (PBS or adjuvant aluminum hydroxide (Alum) only group).

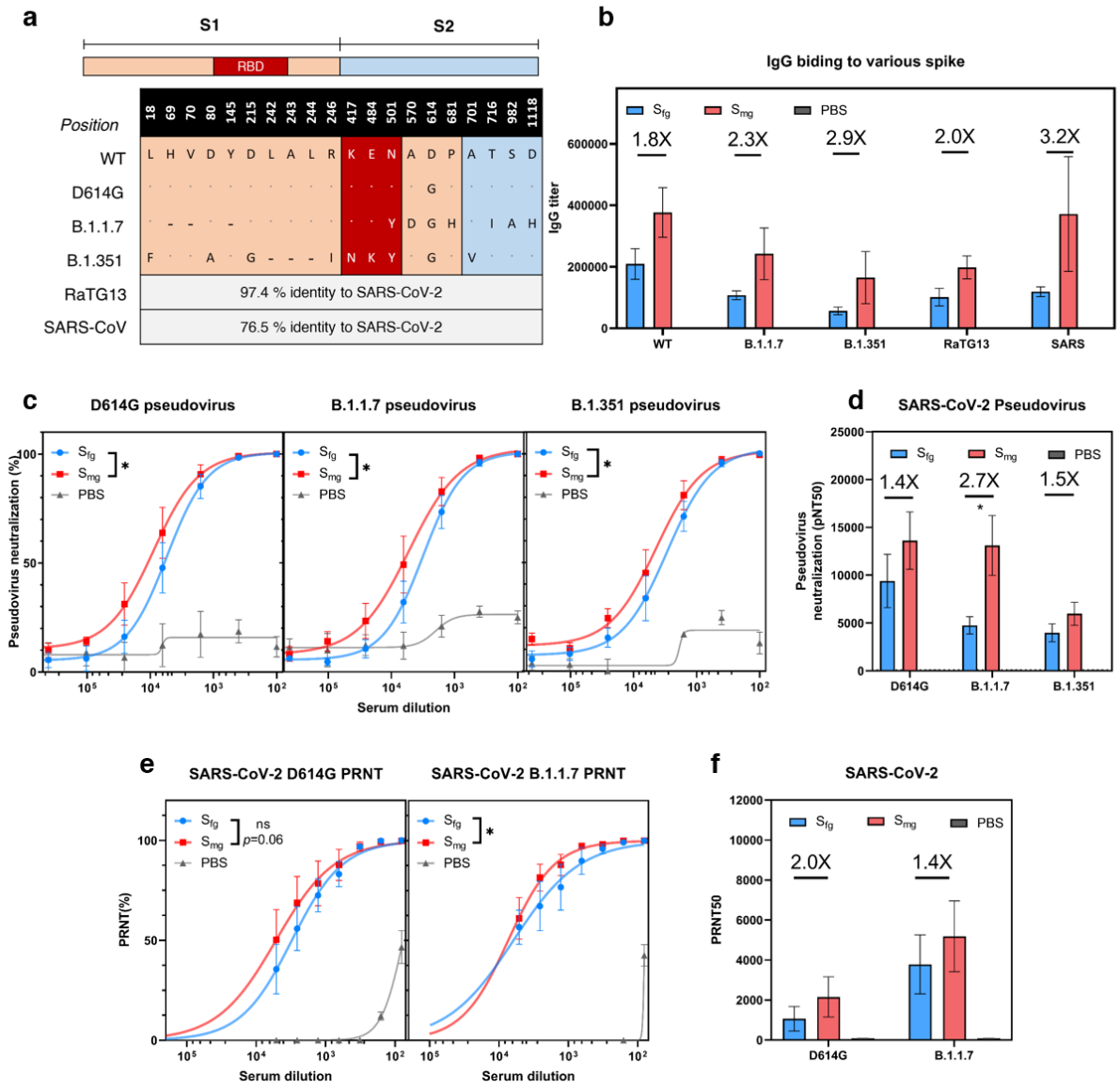


Fig. 4. Antibodies induced by S_{mg} vaccination in mice showed enhanced binding and neutralization breadth and potency against SARS-CoV-2 variants. a, Schematic SARS-CoV-2 S protein domain organization and the mutation landscape of variants used in this study are illustrated. RBD, receptor binding domain. B.1.17, the UK variant. B.1.351, the South African variant. RaTG13, bat coronavirus and SARS-CoV-1. In the mutation map, a dot (.) indicates the same amino acid in that position as wild type and a dash (–) indicates a deletion. **b,** The S protein -specific IgG antibody endpoint titer determined by ELISA. **c,** Neutralization curves of SARS-CoV-2 variants pseudovirus are shown. **d,** Titers that achieve 50% pseudovirus neutralization (pNT50) of SARS-CoV-2 variants are plotted. **e,** The neutralization curves of infectious SARS-CoV-2 variants determined by plaque reduction neutralization test (PRNT) are shown. **f,** Titers that achieve 50% neutralization (PRNT50) of SARS-CoV-2 variants are plotted. Data are means ± SEM (standard error of the mean). The curves were fit by nonlinear regression using GraphPrism 9.0 and comparisons are performed by multiple t-test (paired, two tailed). The comparisons of titers are performed by Student's t-test (unpaired, two tailed). The fold increase of titers is indicated by numbers above the bars. Blue dots, sera collected

from BALB/c mice immunized with S_{fg}; Red squares, sera collected from BALB/c mice immunized with S_{mg} ; Gray triangles, control group animals (PBS group).

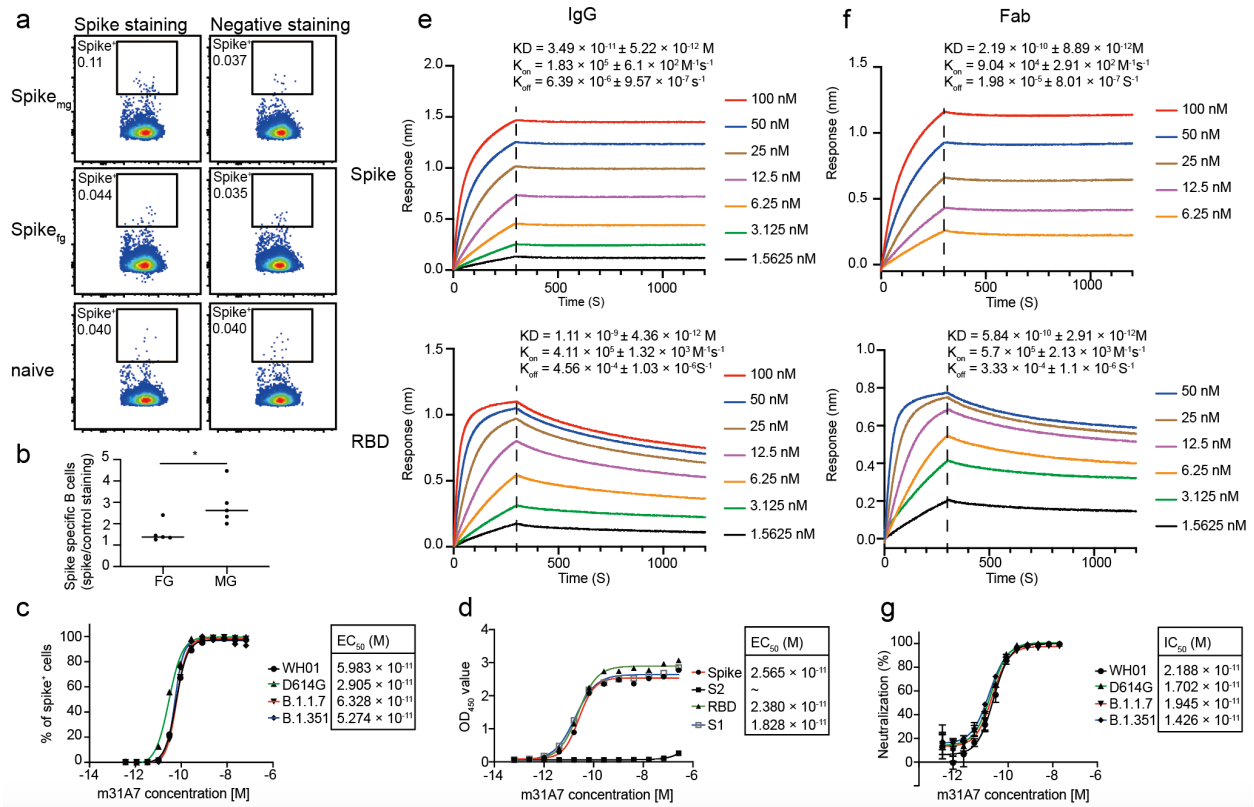
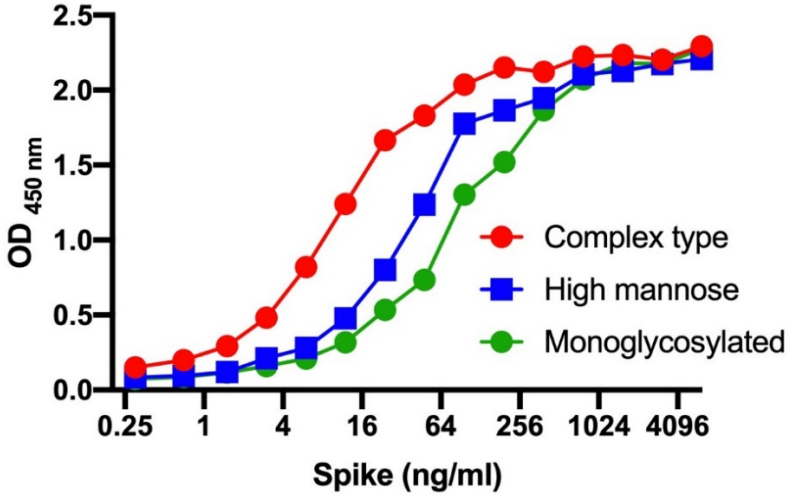
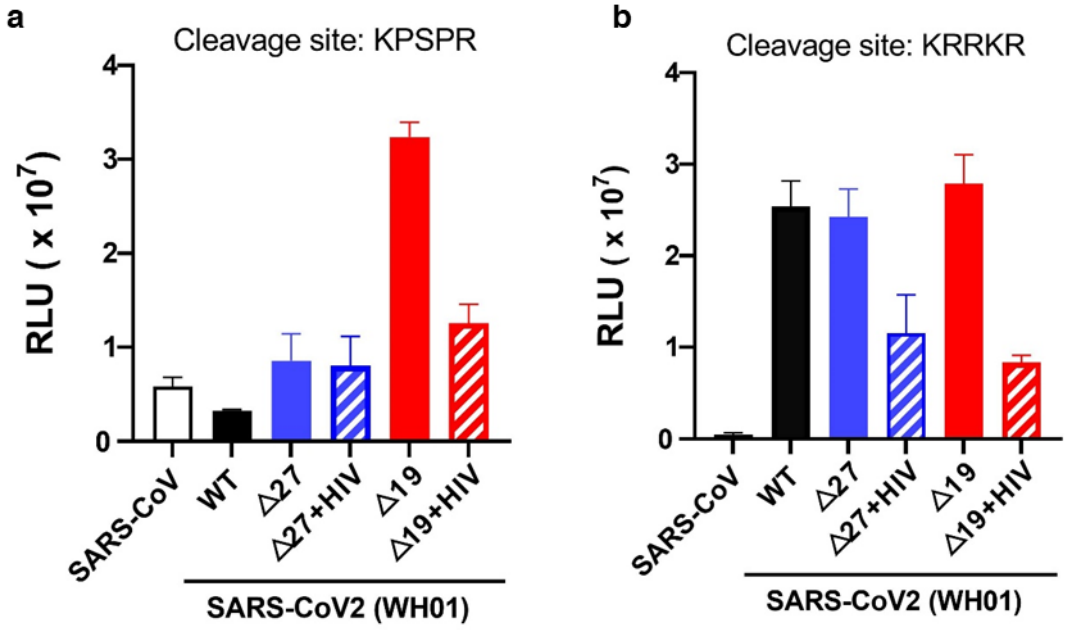


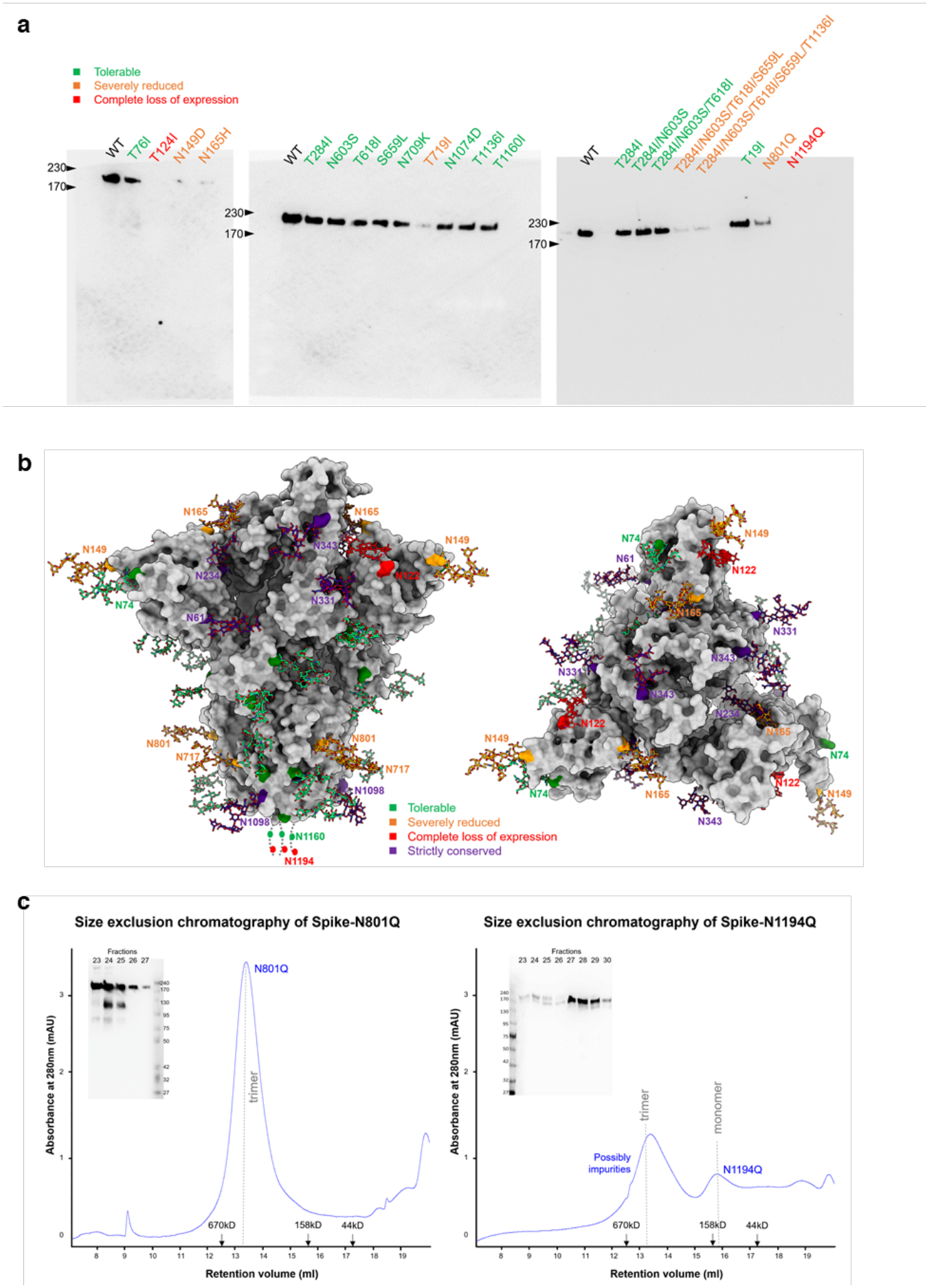
Fig. 5. Identification and characterization of a broadly neutralizing monoclonal antibody from S_{mg} immunized mice. a, FACS showing the percentage of spike-specific B cells from mice immunized with S_{fg} or S_{mg} at day 28 after the last immunization. **b**, The relative ratio of the percentage of S-specific B cells ($CD3^{-}CD19^{+}S^{+}$) vs. negative staining (without S) in the spleen of immunized mice or naïve mice (n=5). **c**, FACS analysis of m31A7 binding with 293T cells expressing S protein of WH01 and various variants. **d**, ELISA showing m31A7 binding to S1, S2, or RBD of the S protein. **e f**, Binding avidity and affinity of m31A7 IgG1 (**e**) and Fab (**f**) to full-length S protein (upper panel) and to RBD (lower panel) as measured by bio-layer interferometry. Results are mean ± s.d. **g**, Neutralization assay by m31A7 against SARS-CoV-2 pseudoviruses of WH01 and indicated variants.



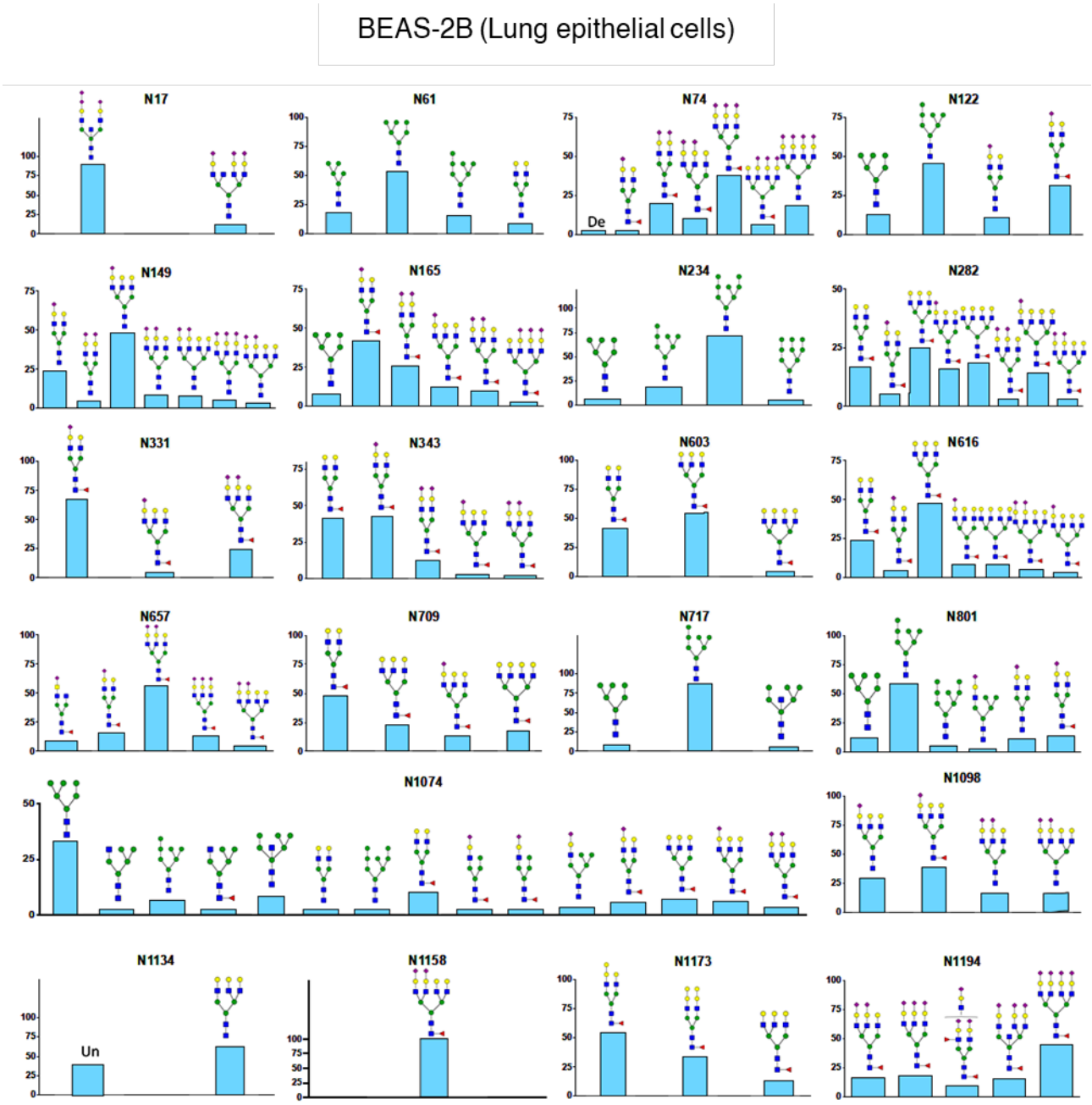
Extended Data Fig. 1. Impact of glycosylation on the binding of SARS-CoV-2 S protein to human ACE2. Wild type SARS-CoV-2 S protein was either produced from HEK 293T (complex type, red) or HEK293S (high mannose, blue) cells. Furthermore, S proteins from HEK293S were processed with Endo H to remove oligomannose N-glycan to produce monoglycosylated spike proteins (green). The ACE2 binding activities of S proteins containing various N-glycans were measured by ELISA. Shown are the absorbance at 450 nm.



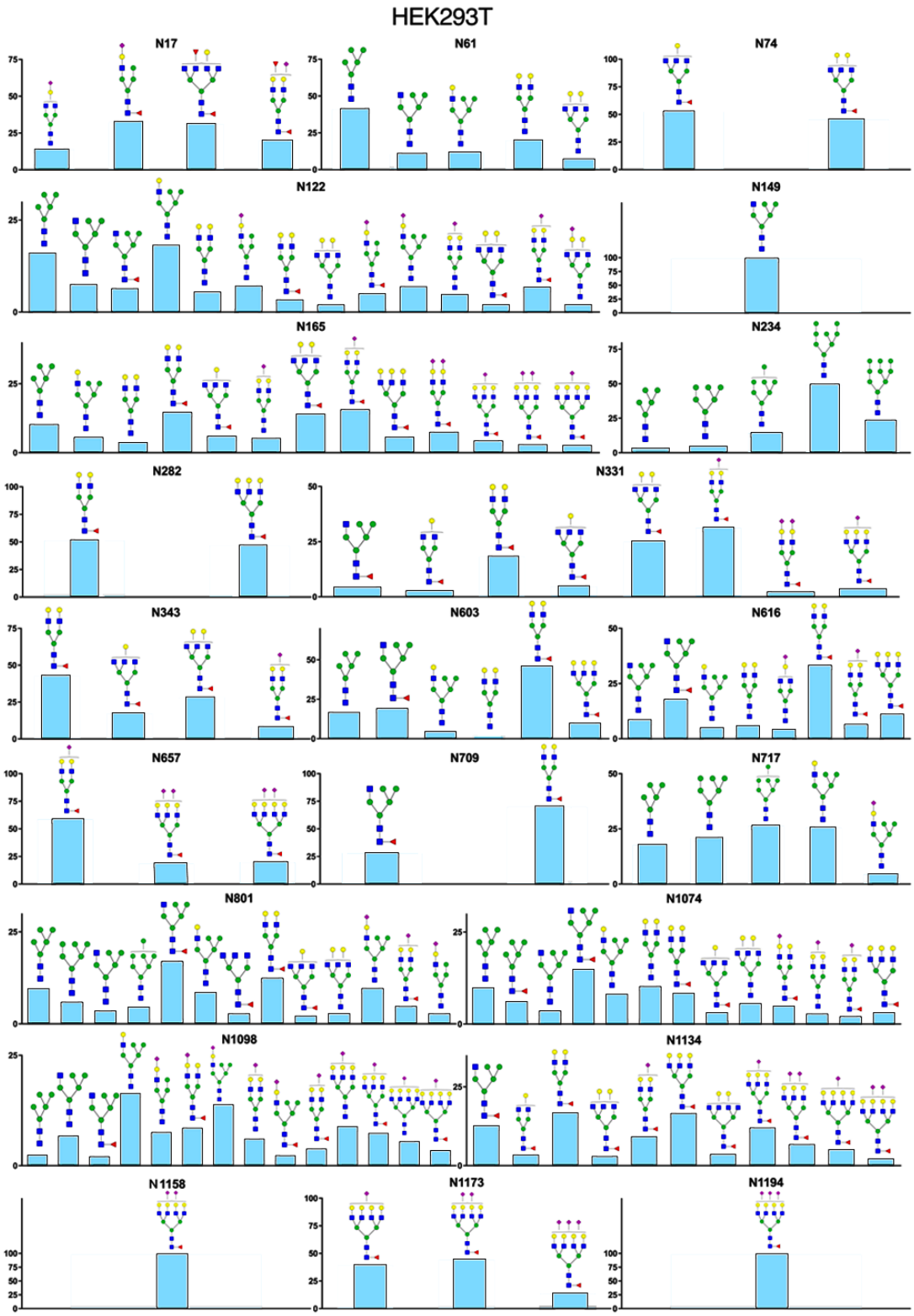
Extended Data Fig. 2. Impact of cleavage sites and C-terminal deletion of spike on pseudovirus production. a, Pseudoviruses were produced using HIV packaging system with spike of SARS-CoV (blank) or with spike of SARS-CoV-2. Wild-type spike (black), deletion of 27 amino acids from C-terminus (Δ27) (blue) or deletion of 19 amino acid from C-terminus (Δ19) (red) of spike protein were used for pseudovirus production. The addition of HIV sequence (NRVRQGYS) after deletions (striped blue and striped red) were also tested for pseudovirus production. Equal volumes of collected medium were used to infected 293-ACE2 cells for 48 h and the infection efficiency were measured by luciferase activity. The relative luminescence unit (RLU) was measured. Showed values are mean ± SD. **b,** The original furin cleavage site of spike proteins from SARS-CoV (blank) or SARS-CoV-2 (black) were changed to KRRKR. The impacts of cleavage sites of wild-type spike (black), spike with a deletion of 27 amino acids from C-terminus (Δ27) (blue) or spike with a deletion of 19 amino acid from C-terminus (Δ19) (red) were also tested. Equal volumes of collected medium were used to infect 293-ACE2 cells for 48 h and the infection efficiency was measured by luciferase activity. The relative luminescence unit (RLU) was measured. Values shown are mean ± SD.


a
Tolerable
Severely reduced
Complete loss of expression
WT
T76I
T124I
N149D
N165H
230
170
WT
T284I
N603S
T618I
S659L
N709K
T719I
N1074D
T1136I
T1160I
230
170
WT
T284I
T284I/N603S
T284I/N603S/T618I
T284I/N603S/T618I/S659L
T284I/N603S/T618I/S659L/T1136I
T19I
N801Q
N1194Q
230
170
b
N165
N165
N149
N149
N343
N234
N122
N331
N74
N61
N801
N801
N717
N717
N1098
N1098
N1160
N1194
N74
N61
N149
N122
N165
N331
N343
N331
N343
N122
N234
N165
N149
N74
N74
N165
N122
N149
N343
Tolerable
Severely reduced
Complete loss of expression
Strictly conserved
c
Size exclusion chromatography of Spike-N801Q
Fractions
23 24 25 26 27
N801Q
trimer
Absorbance at 280nm (mAU)
670kD
158kD
44kD
Retention volume (ml)
Size exclusion chromatography of Spike-N1194Q
Fractions
23 24 25 26 27 28 29 30
trimer
monomer
Possibly impurities
N1194Q
Absorbance at 280nm (mAU)
670kD
158kD
44kD
Retention volume (ml)

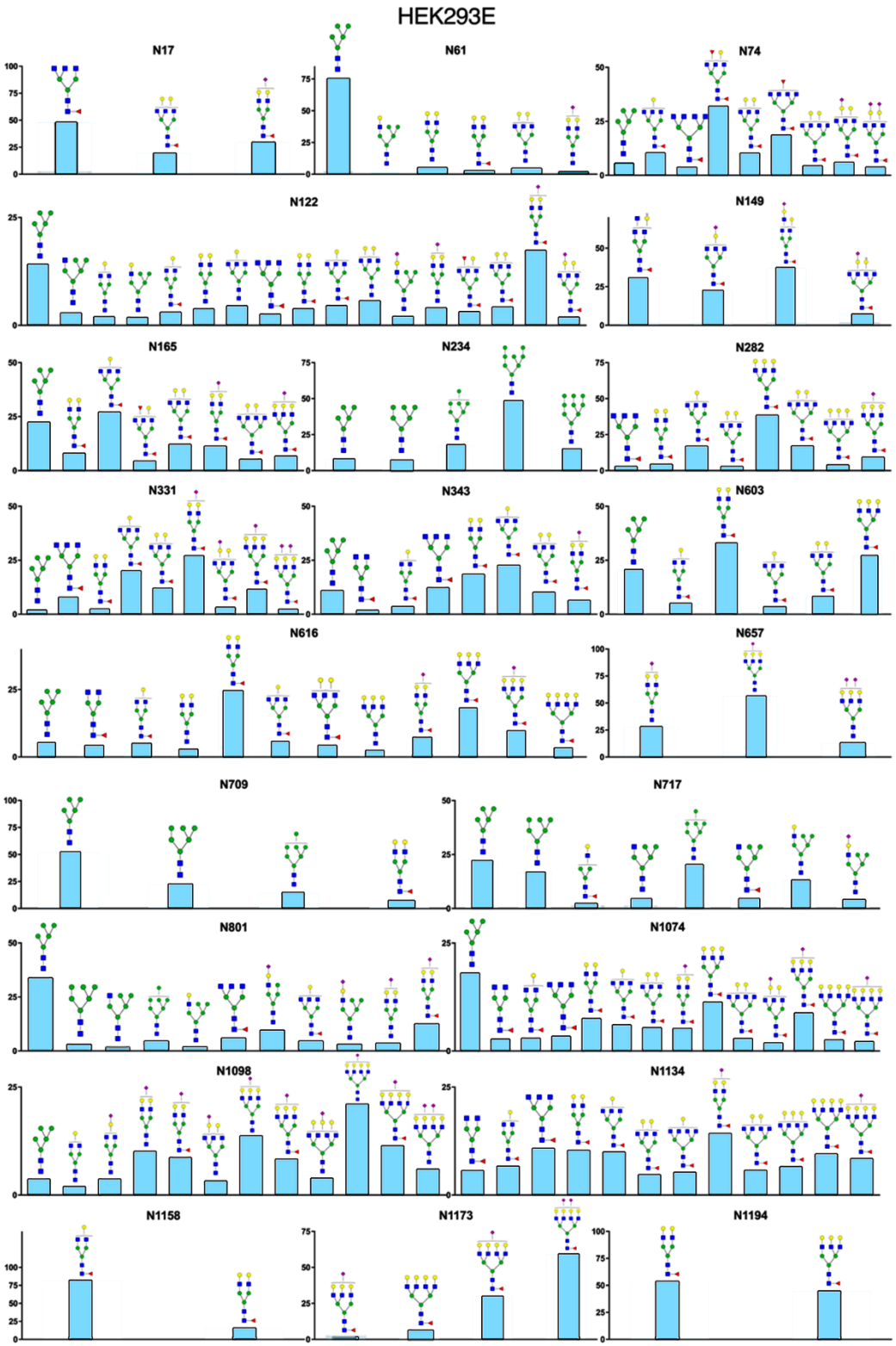
Extended Data Fig. 3. **Expression test of recombinant S protein with glycosite mutation. a,** Expression levels of S protein constructs carrying different mutations (as noted) in HEK293 Ebna suspension cell cultures tested by western blot using anti-penta-histidine monoclonal antibody. Expression tests were performed in three batches, each with the wild type as the positive control. Choice of amino acid change is based on S protein variation statistics from GISAID to maximize the similarity with the circulating virus variants. The reduced expression level could be due to the influence of mutation on either protein solubility or protein integrity during folding or after secretion. The combination of several tolerable glycosite mutations shows a threshold of four or more mutations which start to severely reduce the expression level while it is tolerable for each single mutation. **b,** Mapping the glycosite mutation tolerance on S protein 3D structure (PDB 7CN9) shown with glycan chains modeled in sticks on surface. N-glycan sites and adhering glycans were colored according to the results in **a**. Green indicates the tolerable sites, orange and red colors indicate severe reduction and complete loss of expression when mutated, and purple colors are the sites strictly conserved in which mutagenesis was not tried. Expression-critical sites are labeled with the residue number. **c,** Size-exclusion chromatography of two selective N-glycosite mutants, N801Q (left) and N1194Q (right), with their elution fractions individually checked with western blot using rabbit anti-S protein (14-1209) polyclonal antibodies to confirm protein identity at extremely low concentrations, as these mutations severely diminished the S protein expression level. The S-N801Q mutant has an additional degradation band (possibly S1) and S-N1194 mutant exhibits a big peak shift on SEC compared to the trimeric wild type, indicating these sites may be important for S protein integrity and/or trimerization.



Extended Data Fig. 4. **The N-linked glycan structures of SARS-CoV-2 S protein from BEAS-2B cells**. In-solution digestion of purified S protein purified from BEAS-2B cells and analysis with LC–MS/MS reveals the glycan identity (x axis) and distribution (y axis) of each N-glycosylation site, drawn by the symbolic representation of monosaccharides.



Extended Data Fig. 5. N-linked glycan structures of SARS-CoV-2 S protein from HEK 293T cells. In-solution digestion of purified S protein from HEK 293T cells and analysis with LC–MS/MS reveals the glycan identity (x axis) and distribution (y axis) of each N-glycosylation site, drawn by the symbolic representation of monosaccharides.



Extended Data Fig. 6. N-linked glycan structures of SARS-CoV-2 S protein from HEK 293E. In-solution digestion of purified S protein from HEK 293 Ebna cells and analysis with LC–MS/MS reveals the glycan identity (x axis) and distribution (y axis) of each N-glycosylation site, drawn by the symbolic representation of monosaccharides.

Extended Data Table 1. The abundance of N-linked glycans observed across the SARS-CoV-2 S protein from BEAS-2B cells.

The percentages of glycan compositions at each N-glycosite are shown in the upper table. The lower table was further categorized into oligomannose-, hybrid-, hybrid-S, Complex- and Complex-S as well as those carried at least one fucose or one sialic acid to highlight the difference in comparison with the glycan profile of S protein from other sources.

N glycan	N17	N61	N74	N122	N149	N165	N234	N282	N331	N343	N603	N616	N657	N709	N717	N801	N1074	N1098	N1134	N1158	N1173	N1194
FA4/A3B-S	0	0	26	0	0	3	0	17	1	0	0	3	5	0	0	0	0	0	0	0	0	58
FA3/A2B-S	0	0	50	0	0	23	0	19	28	5	0	13	80	13	0	0	10	38	0	0	0	42
FA2/A1B-S	0	0	22	31	0	67	0	5	69	54	0	5	16	0	0	13	6	0	0	0	0	0
FA1-S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A4/A3B-S	12	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0
A3/A2B-S	88	1	0	0	61	0	0	0	0	0	0	0	0	0	0	0	0	45	0	0	0	0
A2/A1B-S	0	0	0	11	28	0	0	0	0	0	0	0	0	0	0	11	0	0	0	100	0	0
A1-S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FA4/A3B	0	0	0	0	0	0	0	18	0	0	4	8	0	18	0	0	0	0	0	0	0	0
FA3/A2B	0	0	0	0	0	0	0	25	0	0	55	48	0	22	0	0	7	0	0	0	13	0
FA2/A1B	0	0	0	0	0	0	0	16	2	41	41	24	0	47	0	0	10	0	0	0	87	0
FA1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
A4/A3B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A3/A2B	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	62	0	0	0
A2/A1B	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
A1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
Fybrid-S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0
Hybrid-S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	0	0	0	0	0
Fybrid	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hybrid	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	8	0	0	0	0	0
M4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M5	0	18	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0
M6	0	53	0	13	0	7	6	0	0	0	0	0	0	0	7	11	33	0	0	0	0	0
M7	0	17	0	45	0	0	18	0	0	0	0	0	0	0	86	57	7	0	0	0	0	0
M8	0	2	0	0	0	0	71	0	0	0	0	0	0	0	1	4	2	0	0	0	0	0
M9	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
unoccupied	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	38	0	0	0

Total	
FA4/A3B-S	5
FA3/A2B-S	15
FA2/A1B-S	13
FA1-S	0
A4/A3B-S	2
A3/A2B-S	9
A2/A1B-S	7
A1-S	0
FA4/A3B	2
FA3/A2B	8
FA2/A1B	12
FA1	0
A4/A3B	0
A3/A2B	3
A2/A1B	0
A1	0
Fybrid-S	0
Hybrid-S	0
Fybrid	0
Hybrid	1
M4	0
M5	1
M6	6
M7	10
M8	4
M9	0
unoccupied	2

N glycan	N17	N61	N74	N122	N149	N165	N234	N282	N331	N343	N603	N616	N657	N709	N717	N801	N1074	N1098	N1134	N1158	N1173	N1194
Oligomannose	0	89	0	58	0	7	100	0	0	0	0	0	0	0	95	73	44	0	0	0	0	0
Hybrid	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	8	0	0	0	0	0
Hybrid-S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	8	0	0	0	0	0
Complex	0	9	0	0	0	0	0	59	2	41	100	79	0	87	0	0	25	0	62	0	100	0
Complex-S	100	1	98	42	100	93	0	41	98	59	0	21	100	13	0	24	15	100	0	100	0	100
Unoccupied	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	38	0	0	0
Fucosylation	0	0	98	31	0	93	0	100	100	100	100	100	100	100	0	13	40	38	0	0	100	100
Sialylation	100	1	98	42	100	93	0	41	98	59	0	21	100	13	0	26	23	100	0	100	0	100

Total	
oligomannose	21
Hybrid	1
Hybrid-S	0
Complex	26
Complex-S	50
unoccupied	2
Fucosylation	55
Sialylation	51

Extended Data Table 2. The abundance of N-linked glycans observed across the SARS-CoV-2 S protein from HEK 293T cells.

The percentages of glycan compositions at each N-glycosite are shown in the upper table. The lower table was further categorized into oligomannose-, hybrid-, hybrid-S, Complex- and Complex-S as well as those carried at least one fucose or one sialic acid to highlight the difference in comparison with the glycan profile of S protein from other sources.

N glycan	N17	N61	N74	N122	N149	N165	N234	N282	N331	N343	N603	N616	N657	N709	N717	N801	N1074	N1098	N1134	N1158	N1173	N1194
FA4/A3B-S	0	0	0	0	0	1	0	0	1	1	0	1	9	0	0	0	1	1	5	0	87	100
FA3/A2B-S	0	0	6	2	0	6	0	3	12	1	0	3	17	0	0	0	2	2	12	0	6	0
FA2/A1B-S	0	0	15	9	0	40	0	2	34	12	1	7	41	0	0	3	7	2	7	0	0	0
FA1-S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A4/A3B-S	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	100	0	0
A3/A2B-S	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0
A2/A1B-S	27	1	0	5	0	4	0	0	0	0	0	4	0	0	0	2	1	2	0	0	0	0
A1-S	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
FA4/A3B	21	0	6	0	0	1	0	5	0	0	1	2	0	1	0	0	1	0	11	0	7	0
FA3/A2B	0	0	69	3	0	15	0	43	25	28	11	12	7	3	1	3	10	4	23	0	0	0
FA2/A1B	14	2	0	4	0	15	0	39	18	36	40	31	7	30	0	4	22	13	22	0	0	0
FA1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0
A4/A3B	0	1	4	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0
A3/A2B	0	7	0	2	0	2	0	4	3	10	3	2	0	0	0	1	2	0	1	0	0	0
A2/A1B	0	11	0	5	0	2	0	1	0	0	1	5	0	8	1	2	5	1	1	0	0	0
A1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fybrid-S	22	0	0	8	0	0	0	0	2	1	0	1	0	4	0	1	12	8	3	0	0	0
Hybrid-S	2	1	0	18	0	1	0	0	0	0	0	0	0	4	5	21	1	24	0	0	0	0
Fybrid	0	2	0	5	0	0	0	0	3	9	17	16	1	12	2	6	26	7	11	0	0	0
Hybrid	0	32	0	25	100	3	4	1	0	0	8	13	2	0	26	18	4	18	2	0	0	0
M4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0
M5	0	40	0	11	0	10	2	0	0	1	13	1	14	19	19	29	5	5	0	0	0	0
M6	2	2	0	0	0	0	3	0	0	0	1	0	1	13	20	5	1	0	0	0	0	0
M7	0	0	0	1	0	0	14	0	0	0	1	0	0	5	25	4	0	1	0	0	0	0
M8	0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M9	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
unoccupied	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Total	
FA4/A3B-S	9
FA3/A2B-S	3
FA2/A1B-S	8
FA1-S	0
A4/A3B-S	5
A3/A2B-S	1
A2/A1B-S	2
A1-S	0
FA4/A3B	3
FA3/A2B	12
FA2/A1B	13
FA1	0
A4/A3B	0
A3/A2B	2
A2/A1B	2
A1	0
Fybrid-S	3
Hybrid-S	4
Fybrid	5
Hybrid	12
M4	0
M5	8
M6	2
M7	2
M8	2
M9	1
unoccupied	0

N glycan	N17	N61	N74	N122	N149	N165	N234	N282	N331	N343	N603	N616	N657	N709	N717	N801	N1074	N1098	N1134	N1158	N1173	N1194
Oligomannose	2	43	0	14	0	10	96	0	0	1	16	1	16	38	65	39	6	6	0	0	0	0
Hybrid	0	33	0	30	100	3	4	1	3	10	25	29	3	12	28	24	30	25	13	0	0	0
Hybrid-S	24	1	0	26	0	1	0	0	2	1	0	1	0	8	5	22	12	32	3	0	0	0
Complex	35	21	79	14	0	36	0	93	48	74	58	53	14	43	1	10	41	19	59	0	7	0
Complex-S	38	1	21	16	0	50	0	5	47	14	1	16	68	0	0	5	11	19	25	100	93	100
Unoccupied	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fucosylation	59	3	96	44	0	80	0	92	92	79	54	57	82	42	6	32	55	56	83	0	100	100
Sialylation	63	2	21	42	0	52	0	5	49	15	1	17	68	7	6	27	23	50	28	100	93	100

Total	
Oligomannose	16
Hybrid	17
Hybrid-S	6
Complex	32
Complex-S	29
Unoccupied	0
Fucosylation	55
Sialylation	35

Extended Data Table 3. The abundance of N-linked glycans observed across the SARS-CoV-2 S protein from HEK 293E cells.

The percentages of glycan compositions at each N-glycosite are shown in the upper table. The lower table was further categorized into oligomannose-, hybrid-, hybrid-S, Complex- and Complex-S as well as those carried at least one fucose or one sialic acid to highlight the difference in comparison with the glycan profile of S protein from other sources.

N glycan	N17	N61	N74	N122	N149	N165	N234	N282	N331	N343	N603	N616	N657	N709	N717	N801	N1074	N1098	N1134	N1158	N1173	N1194	Total	
FA4/A3B-S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	11	5	0	85	0	FA4/A3B-S	5
FA3/A2B-S	0	0	11	3	69	4	0	7	14	2	0	8	0	0	0	0	11	9	0	0	4	0	FA3/A2B-S	6
FA2/A1B-S	28	0	0	18	0	11	0	0	29	6	0	9	0	0	0	11	7	9	11	0	0	0	FA2/A1B-S	6
FA1-S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	FA1-S	0
A4/A3B-S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	0	A4/A3B-S	1
A3/A2B-S	0	0	0	0	0	0	0	0	0	0	0	0	65	0	0	0	0	14	0	0	0	0	A3/A2B-S	4
A2/A1B-S	0	1	0	6	0	0	0	0	0	0	0	0	35	0	0	7	0	19	0	0	0	0	A2/A1B-S	3
A1-S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A1-S	0
FA4/A3B	0	0	25	0	0	3	0	22	2	0	0	4	0	0	0	0	9	0	29	0	11	0	FA4/A3B	5
FA3/A2B	72	0	61	20	31	58	0	64	45	50	39	30	0	0	0	14	28	0	28	0	0	38	FA3/A2B	26
FA2/A1B	0	3	0	10	0	9	0	7	8	24	49	42	0	10	3	4	16	0	28	100	0	62	FA2/A1B	17
FA1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	3	0	1	0	0	0	0	0	FA1	0
A4/A3B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	A4/A3B	0
A3/A2B	0	7	0	10	0	0	0	0	0	0	0	1	0	0	0	5	0	0	0	0	0	0	A3/A2B	1
A2/A1B	0	6	0	7	0	0	0	0	0	0	0	3	0	0	0	3	1	5	0	0	0	0	A2/A1B	1
A1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	A1	0
Fybrid-S	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	Fybrid-S	0
Hybrid-S	0	0	0	3	0	0	0	0	0	0	0	0	0	0	7	14	0	0	0	0	0	0	Hybrid-S	1
Fybrid	0	0	0	2	0	0	0	0	0	1	0	0	0	0	4	0	2	0	0	0	0	0	Fybrid	0
Hybrid	0	5	0	6	0	0	0	0	0	1	0	0	0	0	19	4	3	2	0	0	0	0	Hybrid	2
M4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	M4	0
M5	0	76	3	11	0	16	4	0	2	14	12	3	0	58	19	30	15	3	0	0	0	0	M5	12
M6	0	2	0	0	0	0	6	0	0	0	0	0	0	19	16	3	2	0	0	0	0	0	M6	2
M7	0	0	0	2	0	0	27	0	0	0	0	0	0	14	27	5	0	0	0	0	0	0	M7	3
M8	0	0	0	0	0	0	54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	M8	2
M9	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	M9	0
unoccupied	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	unoccupied	0

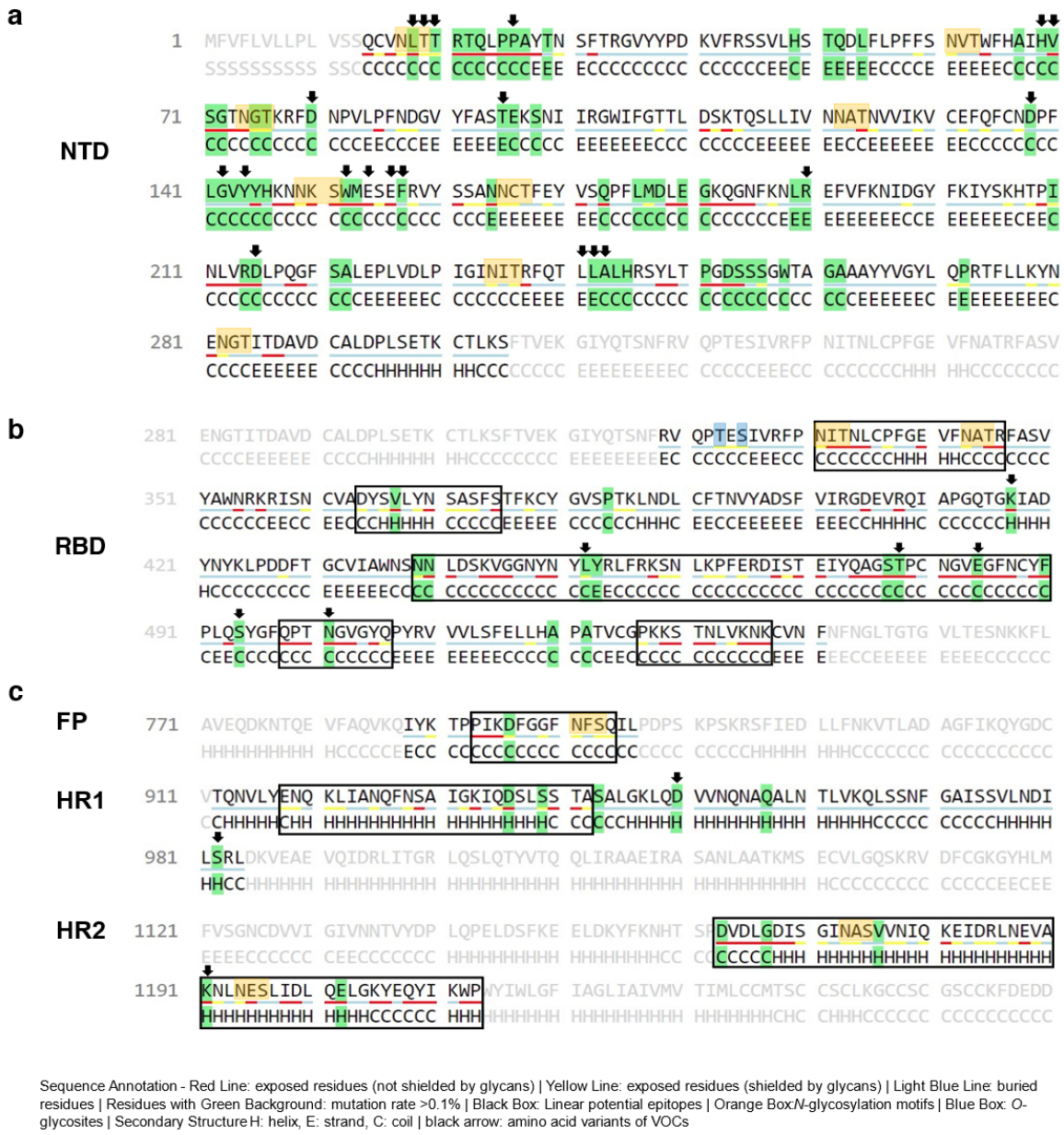
N glycan	N17	N61	N74	N122	N149	N165	N234	N282	N331	N343	N603	N616	N657	N709	N717	N801	N1074	N1098	N1134	N1158	N1173	N1194	Total	
Oligomannose	0	78	3	12	0	10	100	0	2	15	12	3	0	90	63	39	11	3	0	0	0	0	Oligomannose	20
Hybrid	0	5	0	8	0	0	0	0	0	2	0	0	0	0	23	4	4	2	0	0	15	0	Hybrid	3
Hybrid-S	0	0	0	3	0	0	0	0	0	0	0	0	0	0	8	14	1	0	0	0	0	0	Hybrid-S	1
Complex	72	15	86	49	31	71	0	93	55	74	88	80	0	10	7	25	62	5	89	100	15	100	Complex	51
Complex-S	28	1	11	28	69	19	0	7	43	9	0	17	100	0	0	18	22	90	11	0	70	0	Complex-S	25
Unoccupied	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Unoccupied	0
Fucosylation	100	3	97	55	0	84	0	100	98	84	88	92	0	10	11	29	78	29	100	100	100	100	Fucosylation	62
Sialylation	28	1	11	32	69	19	0	7	43	9	0	17	100	0	8	32	23	90	11	0	70	0	Sialylation	26

Extended Data Table 4. Summary of mutated and exposed residues in different regions of the S-protein.

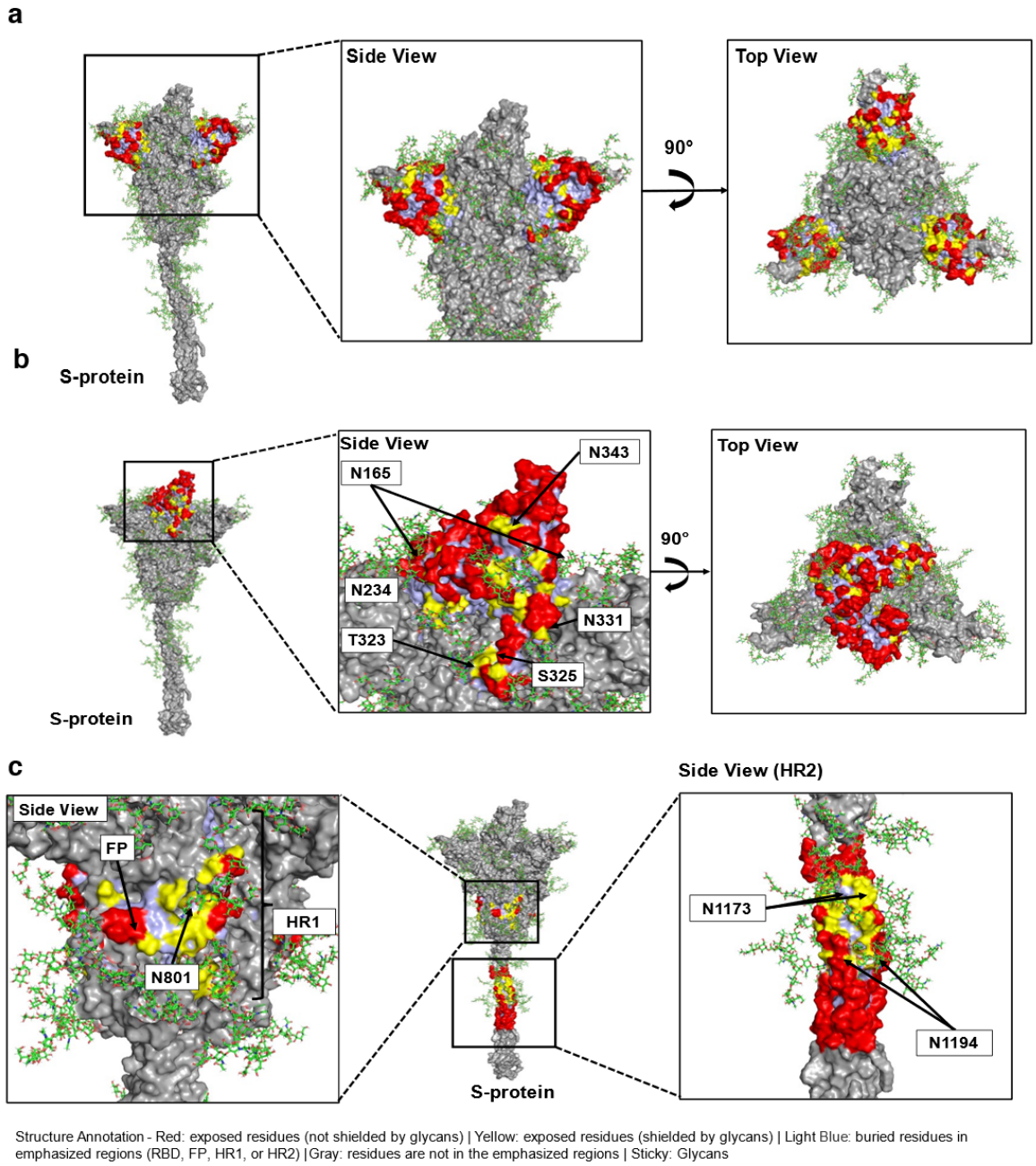
Region	**NTD**	**RBD**	**FP**	**HR1**	**HR2**	**S2***	**Entire S-protein#**
Start Position	14	319	788	912	1163	-	1
End Position	305	541	806	984	1213	-	1213
Length	292	223	19	73	51	404	1213
# of mutated residues with mutation rate >0.1%	61	15	1	6	5	25	131
Percentage	20.89%	6.73%	5.26%	8.22%	9.80%	6.19%	10.80%
Conserved percentage	**79.11%**	**93.27%**	94.74%	91.78%	90.20%	**93.81%**	89.20%
# of total exposed residues (# of exposed residues but shielded by glycans)	101 (41)	85 (29)	9 (6)	15 (10)	36 (14)	93 (55)	387 (172)
Percentage	34.59% (14.04%)	38.12% (13.00%)	47.37% (31.58%)	20.55% (13.70%)	70.59% (27.45%)	23.02% (13.61)	31.90% (14.18%)
# of exposed residues with mutation rate ≤0.1%	72	75	8	13	32	83	320
Percentage	24.66%	**33.63%**	**42.11%**	17.81%	**62.75%**	20.54%	26.38%
# of exposed residues but shielded by glycans with mutation rate ≤0.1%	34	27	5	9	14	51	154

* Exclude HR1, HR2, transmembrane region (TM, residue 1214-1237) and cytoplasmic domain (CP, residue 1238-1273).

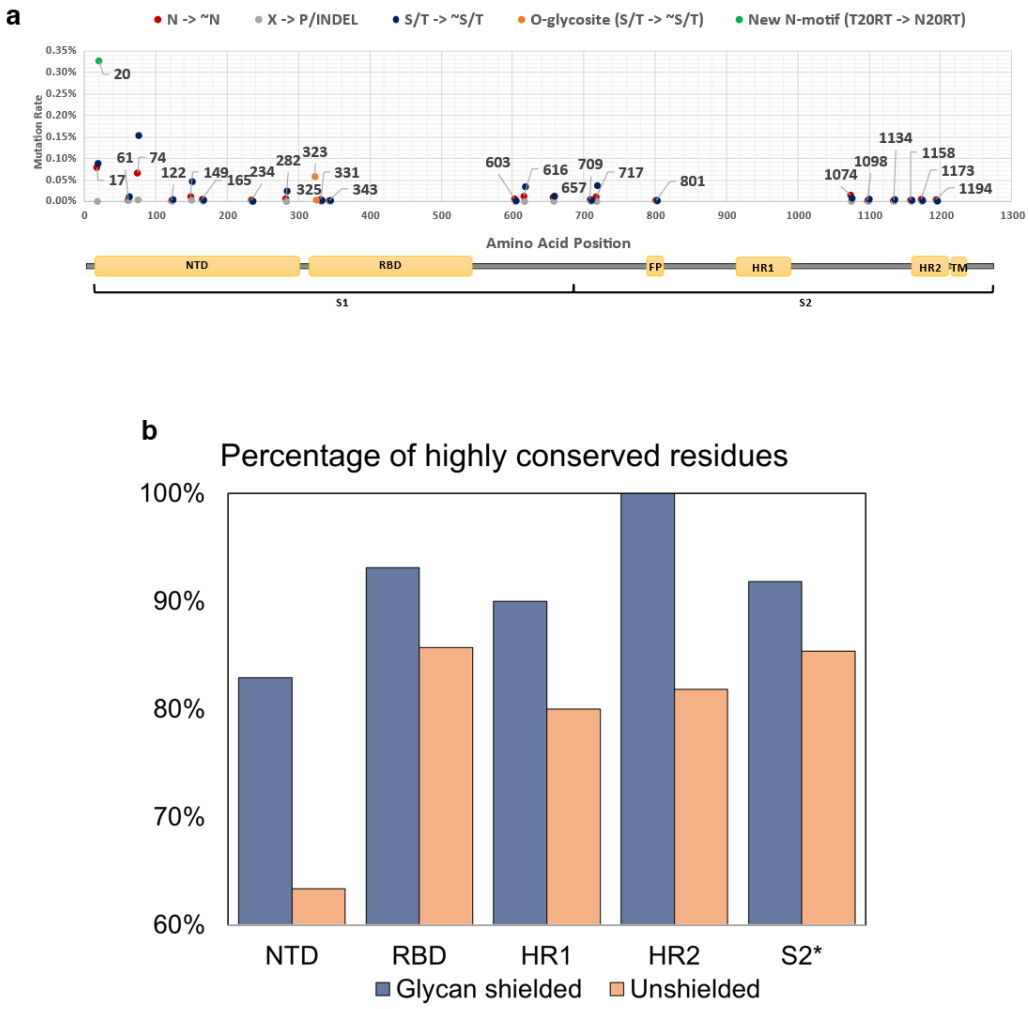
Exclude TM and CP.



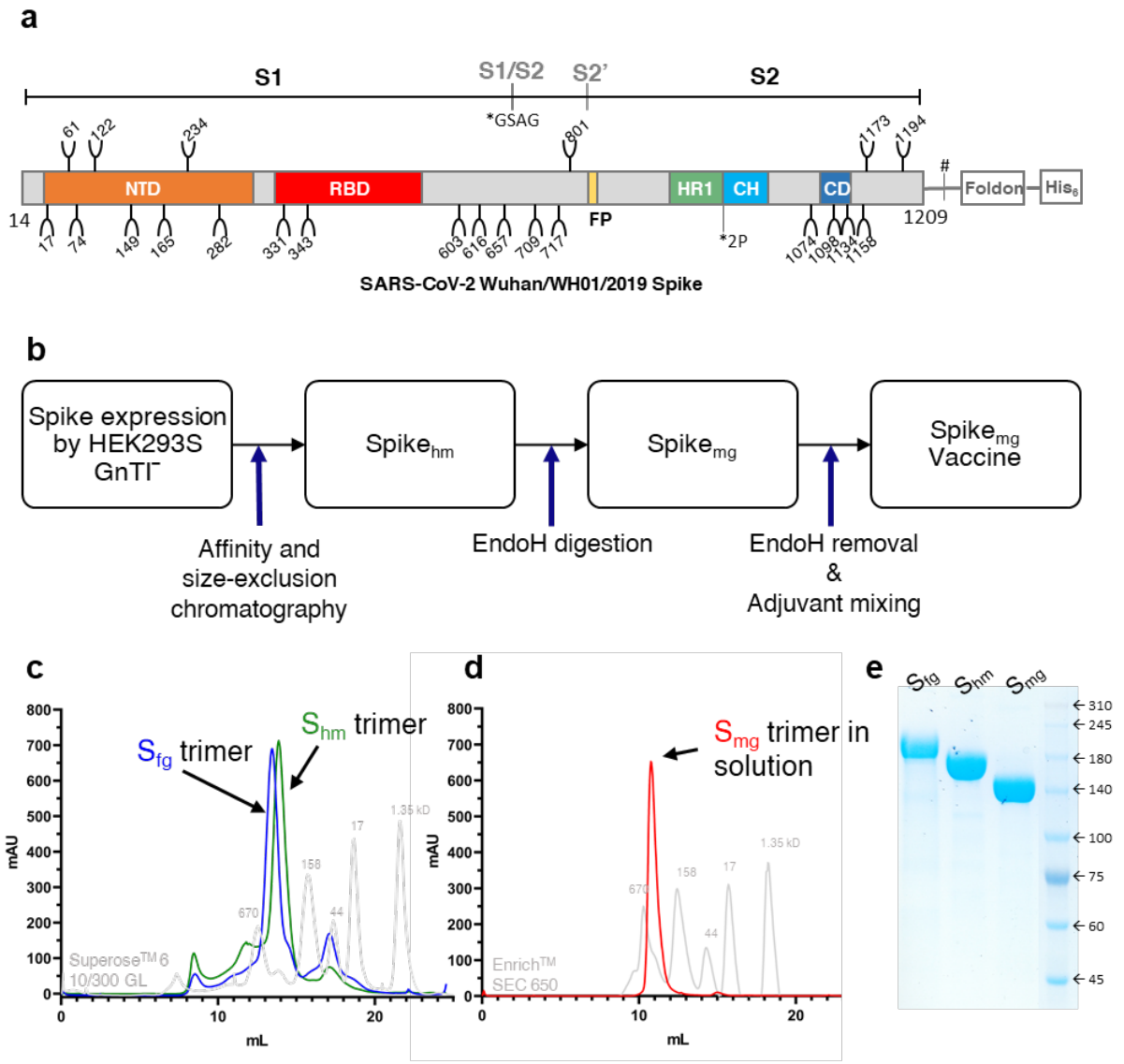
Extended Data Fig. 7. Conserved sequences of SARS-CoV-2 S protein and glycan shield. Sequence annotation. Red line: exposed residues (not shielded by glycans); yellow line: exposed residues (shielded by glycans); light blue line: buried residues; residues with green background: mutation rate >0.1%; black box: linear potential epitopes; orange box: *N*-glycosylation motif; blue box: *O*-glycosites; secondary structure H: helix, E: strand, C: coil, S: signal peptide; black arrow: VOC.



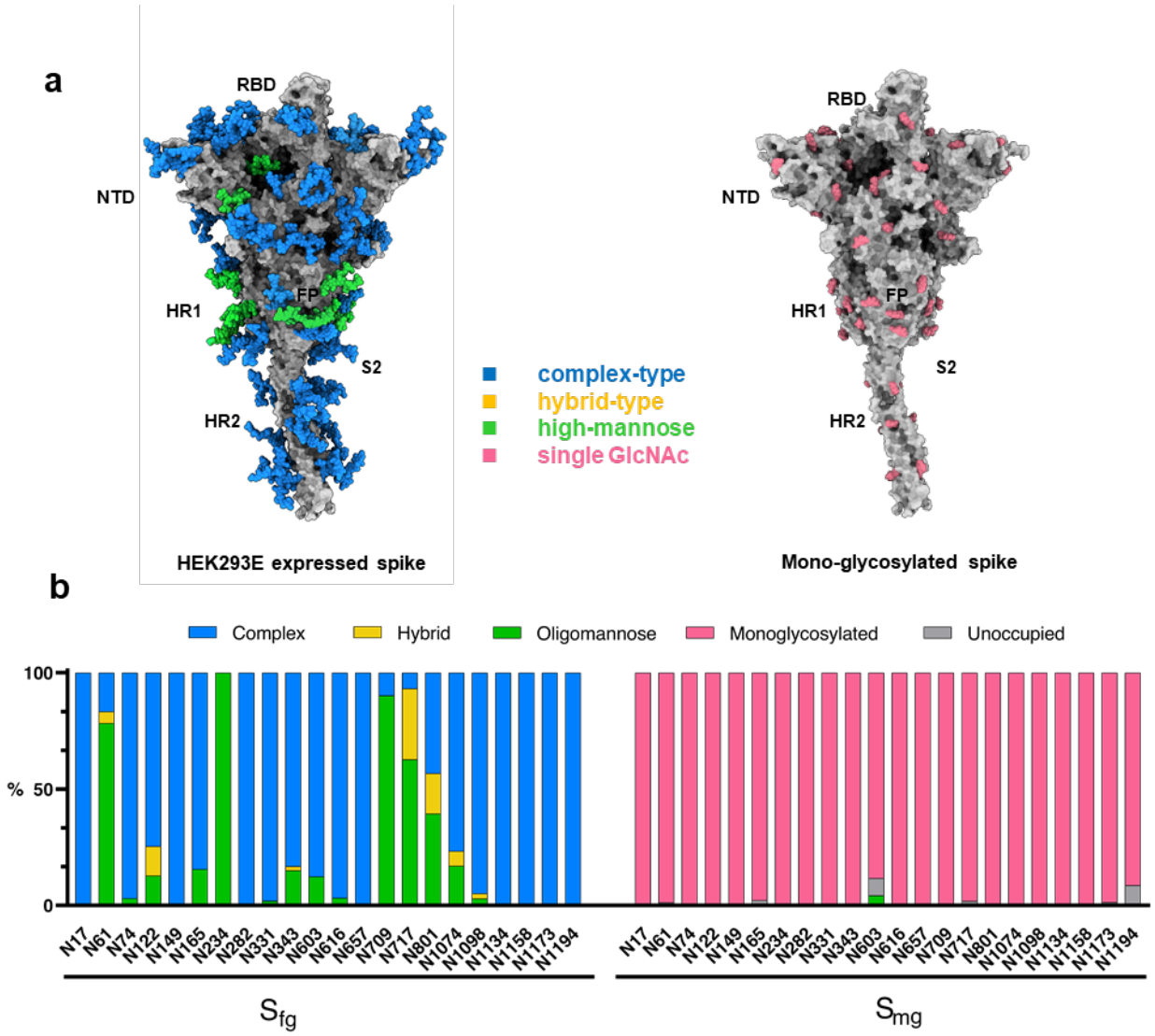
Extended Data Fig. 8. The S protein 3D structure of SARS-CoV-2 colored by glycan shield. (A) N-terminal Domain (NTD, residue 14-305). **(B)** Receptor-binding Domain (RBD, residue 319-541), and **(C)** Fusion Peptide (FP, residue 788-806), Heptad Repeat 1 (HR1, residue 912-984) and Heptad Repeat 2 (HR2, residue 1163-1213). Structure annotation. Red: exposed residues (not shielded by glycans); yellow: exposed residues (shielded by glycans); light blue: buried residues in the emphasized regions (RBD, FP, HR1, and HR2); gray: residues are not in the emphasized regions; sticky: glycans.



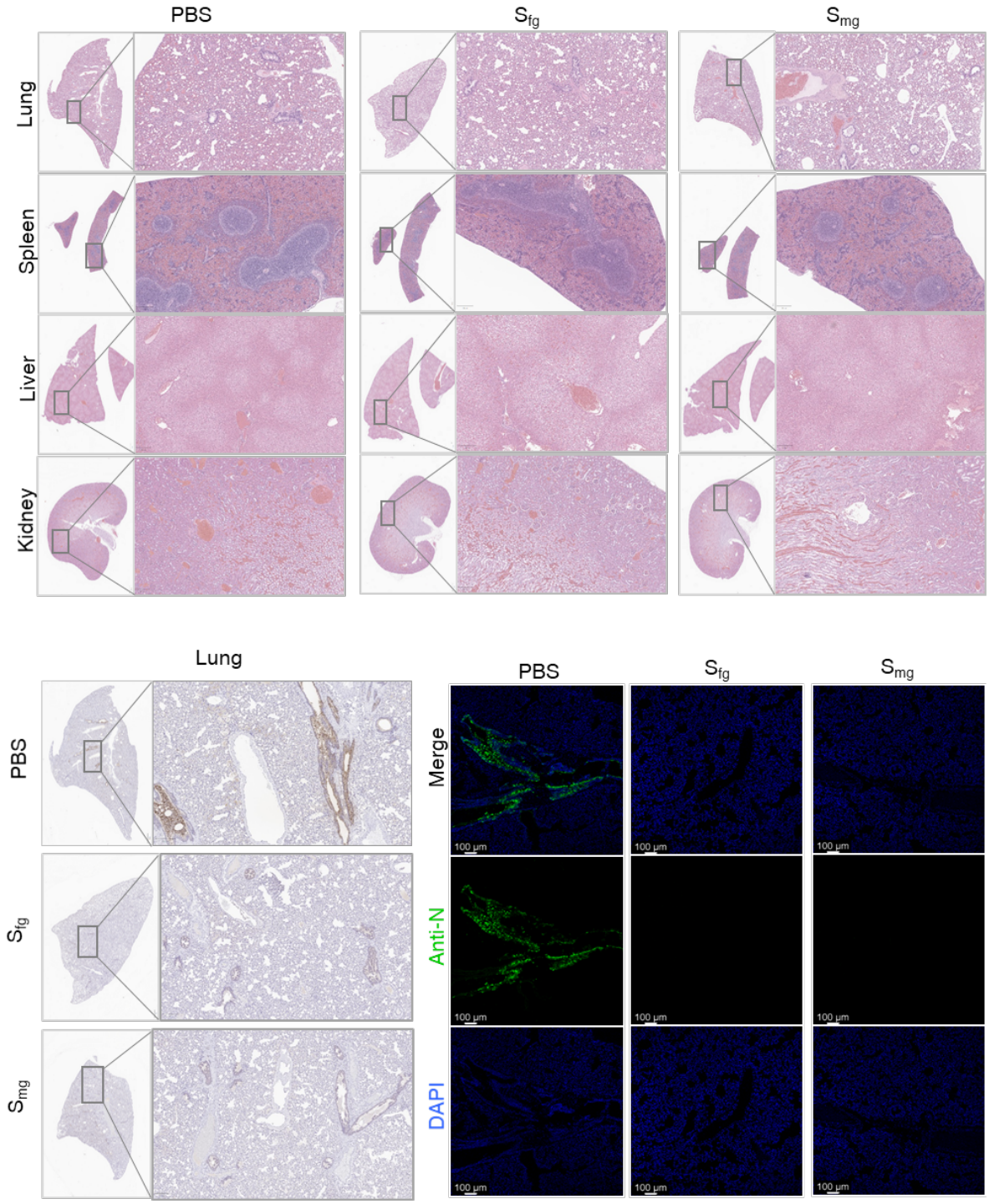
Extended Data Fig. 9. Conservation of S-protein glycosylation sites and comparison of sequence conservation between glycan shielded and unshielded residues in different regions of SARS-CoV-2 S protein. a, 22 *N*-motifs and 2 *O*-glycosites are highly conserved. Glycosylation sites, including S325, N331 and N343 in RBD, N801 in FP, and N1074, N1098, N1134, N1158, N1173 and N1194 closed to or in HR2, are extremely conserved. **b,** Percentage of highly conserved unburied residues (mutation rate <0.1%) are categorized in two groups, shielded and unshielded by glycans, based on the relative solvent accessibility (RSA) calculated by the FreeSASA program with RSA >5% regarded as exposed. Abbreviations include the receptor-binding domain (RBD), Fusion Peptide (FP), Heptad Repeat 1 (HR1), Heptad Repeat 2 (HR2) and the rest of S2 regions (S2*). Because FP is too short (only 19 aa) it is merged with S2* (the rest of S2 domain).



Extended Data Fig. 10. Design and characterization of glycoengineered monoglycosylated S protein (S_{mg}) vaccine. a, Schematic representation of the recombinant SARS-CoV-2 S glycoprotein construct. Protein domains are illustrated as N-terminal domain (NTD), receptor binding domain (RBD), fusion peptide (FP), heptad repeat 1 (HR1), central helix (CH), and connector domain (CD). The C-terminus of soluble S protein is attached with a T4 fibritin (foldon) sequence and a His-tag (His_6). The furin cleavage site is substituted by GSAG residues and two proline are mutated (K986P, and V987P) to fix the S protein in the prefusion state. The positions of N-linked glycosylation sequons (N-X-S/T, where X $\neq$ P) are shown as branches (N, Asn; X, any residue; S, Ser; T, Thr; P, Pro). The pound site represents the thrombin cleavage site. **b,** Schematic overview of S_{mg} vaccine production. S_{hm}, S protein with high mannose type N-glycans; S_{mg}, S protein with a single GlcNAc at each N-glycosylation site. **c,** Size-exclusion chromatography of S_{fg} (S protein with typical complex type N-glycans) and $S_{hm.}$ by Superose 6 Increase 10/300 GL (Cytiva) . Blue curve: $S_{fg,}$ Green curve: $S_{hm,}$ Gray curve: protein molecular weight markers. **d,** Size-exclusion chromatography of purified $Spike_{mg}$ by Enrich SEC 650 (Bio-Rad). Red curve: S_{mg}, Gray curve: protein molecular weight markers. **e,** SDS-PAGE analysis of purified S_{fg}, S_{hm}, and S_{mg}.



Extended Data Fig. 11. N-glycosylation analysis of S_{fg} and S_{mg} vaccines. a, Structure model of S_{fg} and S_{mg}. **b,** Mass Spectrometry analysis of the N-glycan compositions of the S_{fg} expressed by HEK293E and S_{mg}. N801 site of S_{fg} has almost equal abundances of oligomannose and complex-type glycans and is shown as oligomannose on the structural model.



Extended Data Fig. 12. Representative staining of the lungs from hamsters challenged with SARS-CoV-2. a, Representative histopathology (H&E staining) of multiple organs in SARS-CoV-2-infected hamster (3 dpi). Scale bar: 1mm & 200μm. **b,** Representative immunohistochemistry (IHC) staining of lung tissues with SARS-CoV-2 N-specific polyclonal antibodies. The presence of viruses is indicated by arrows. Scale bar: 1mm & 200μm. **c,** Representative immunostaining of lung tissues with SARS-CoV-2 N specific polyclonal antibody. White scale bar: 100μm

Material and Methods

Informatic analysis of SARS-CoV-2 S glycoprotein. The 1,117,474 spike protein sequences of SARS-CoV-2 and their variants information by multiple sequence alignment were extracted from the GISAID (Global Initiative on Sharing Avian Influenza Data) database[10] (version: Apr. 18, 2021). The S-protein 3D structure modeling of SARS-CoV-2 with representative glycan profile was constructed by CHARMM-GUI[46] and OpenMM[47] programs. The input of CHARMM-GUI includes the PDB file (model: 1_1_1)[4], the representative glycan profile, and parameter settings. The representative glycan structure of each *N*-glycosite is the most abundant glycan expressed in Beas-2B lung cells in this work. For O-glycans (T323 and S325), we used the Neu5Ac(α2,3)Gal(β1,3)GalNAc(α1) as representatives.[44] The definition of transmembrane region of S-protein is according to the record of Uniport (P0DTC2) and other parameters in CHARMM-GUI and were the same parameters used in the study by Woo, H. *et al.*[48]. The default scripts, parameters, and pre-optimized model generated by CHARMM-GUI were used as the input for the OpenMM program. Protein secondary structure was determined by majority voting of three protein chains given by the DSSP program[49,50] in 2Struc web service[51]. Relative solvent accessibility (RSA) of the S-protein was calculated by the FreeSASA program[52]. The final RSA of each residue was the average RSA of residues in three protein chains from all frames of 3D modeling. The probe radius in the FreeSASA program was set to 7.2Å to approximate the average size of the hypervariable loops in the CDR of an antibody[9]. Residues with RSA >5% are regarded as exposed residues; otherwise, they are regarded as buried residues. The protein 3D visualization in Figure 1 and Figure S1(A) were drawn by the PyMOL program[53].

Cell culture. HEK 293T (Homo sapiens, embryonic kidney), HEK293T-ACE2 (293T cells stably expressed hACE2), Vero E6 (Cercopithecus aethiops, kidney), Vero E6-ACE2 (Vero E6 cells stably expressed hACE2), A549 (Homo sapiens, lung) and A549-ACE2 cells (A549 cells stably expressed hACE2) were cultured in Dulbecco's modified Eagle medium (DMEM, high glucose; GIBCO, Cat#11995065). Calu1 (Homo sapiens, lung) and Calu1-ACE2 cells (Calu1 cells stably expressed hACE2) were cultured in DMEM/F-12 (GIBCO, Cat#11330032). Calu3 (Homo sapiens, lung) and Calu3-ACE2 cells (Calu3 cells stably expressed hACE2) were cultured in Minimum Essential Medium (MEM; GIBCO, Cat#11095080). BEAS-2B (Homo sapiens, lung) were cultured in RPMI 1640 (GIBCO, Cat#22400089). All the cells above were cultured in medium supplemented with 10% fetal bovine serum (FBS; GIBCO, Cat#10437028) and 1% Penicillin-Streptomycin solution (GIBCO, Cat#15140122). HEK293EBNA (ATCC number CRL-10852) and HEK293S GnTI^{-} (ATCC® CRL-3022™) cells were cultured in Freestyle 293 expression medium (Invitrogen) supplemented with 0.5% bovine calf serum.

Specific glycosylation site mutation and pseudotyped virus assay. For pseudoviruses construction, the full-length S gene was codon-optimized for human cells and several additional variants of S gene were created, including those modified with C-terminal 19 or 27 amino acids deletion, incorporated the eight most membrane-proximal residues of the HIV-1 envelope glycoprotein cytoplasmic domain (NRVRQGYS) (1) and amino acid positions 812 and 813 changed into arginine to make the

recognition sequence of furin-like protease, KRRKR (2) to facilitate efficient pseudotype formation. Pseudotyped viruses incorporated with spike protein from either full-length, variants or glycosylation mutants were constructed and cloned into expression plasmid pVax to generate the envelope recombinant plasmids.

For production of pseudotyped viruses and for the assay of virus entry, the day before transfection, HEK293T cells should be plated 18 to 24 h prior to transfection so that the cell density should be 50-80% confluent at the time of transfection. Luciferase-expressing HIV-1 genome plasmid (pNL4-3.luc.RE) and a plasmid expressing SARS-CoV-2 spike, other representative spike constructs in different variants, or glycosylation mutants were co-transfected according to the manufacturer's instructions. The vesicular stomatitis virus-G (VSV-G) protein was served as a control to evaluate entry efficiency. After overnight incubation, the transfected cells were washed with PBS and incubated with fresh culture medium for another 24 h. Viral supernatants were then collected, clarified to remove cell debris by centrifugation and filtered through 0.45-μm syringe filters, and stored at -80 °C until use. To transduce cells with pseudovirions, cells were seeded into Poly-D-Lysine 96-well plates and inoculated with 100 μl media containing pseudovirions. Media were changed 24 h post-infection. About 48 h post inoculation, cells were lysed with 60μl of Glo lysis buffer at room temperature for 5 min and mix 40μl of cell lysate with 100μl of Luciferase Assay System (Promega). The transduction efficiency was measured by quantification of the luciferase activity using a CLARIOstar plate-reader.

To construct SARS-CoV-2 spike glycosylation mutants, the codon-optimized S gene of SARS-CoV-2 was synthesized by Genscript and cloned into pcDNA to generate mammalian expression constructs or pVax to obtain pseudoviruses constructs, and the resulting plasmid was used as the template for site-directed mutagenesis. The N-glycosylation sites of SARS-CoV-2 S were predicted with The NetNGly server provided by Expasy (www.expasy.org). The putative N-X-S/T sequons were then mutated to Q-X-S/T using QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent Technologies). Reverse primers were designed for the target mutation sites. Following site-directed mutagenesis PCR, the amplification products were digested using DpnI restriction endonuclease. Afterward, the PCR product was directly used to transform XL10-Gold ultracompetent cells; single clones were selected and then confirmed by DNA sequencing. The specific mutation sites of the primers for the glycosylation mutants are listed in Extended Data Table 1.

For quantification of pseudotyped virus particles using HIV-1 Gag p24 ELISA, the titer of HIV-1-based pseudotyped virus particles was quantitatively determined using standard ELISA methods (R&D Systems). A monoclonal antibody specific for HIV-1 Gag p24 was pre-coated onto a microplate. Target pseudotyped viruses were diluted in 1000, 1500, 2000, 5000 and 10000-fold, while HIV-1 Gag p24 standard was prepared as 500, 250, 125, 62.5, 31.3, 15.6, 7.81, and 0 pg/ml as standards. Standards and samples are dispensing into appropriately labeled duplicate wells and any HIV-1 Gag p24 present is bound by the immobilized antibody. After washing away any unbound substances, a monoclonal antibody specific for human HIV-1 Gag p24 is added to the wells. Following a wash to remove any unbound reagent, a 200 μl of substrate solution is added to each well for color to develop in 7 min. The color development is stopped by adding 50 μl of stop solution and the intensity of the

color is measured. The absorbance values at 450 nm was read using SpectraMax M5 (Molecular Devices, Sunnyvale, CA, USA). A standard curve was obtained by plotting the absorbance versus the corresponding concentration of the standard. The HIV-1 Gag p24 concentration of the sample can be then calculated from the standard curve, and HIV-1 Gag p24 values can then be correlated to virus titer of packaging cell supernatants.

Plasmid constructions, expression and purification of SARS-CoV-2 S protein in HEK293T and BEAS-2B cells. SARS-CoV-2 S protein stable trimer with a HisTag is the ectodomain of SARS-CoV-2 S protein which contains residues 1-1208. Proline substitutions at residues 986 and 987 were used to stabilize the trimeric prefusion state and residues 682-685 replaced by a "GSAS" sequence was used to abolish the furin cleavage site. The transmembrane domain was replaced with additional residues from the bacteriophage T4 fibritin foldon trimerization motif (3), thrombin cleavage site and 6xHisTag at the C-terminus of the SARS-CoV-2 S protein was created and subsequently cloned into the mammalian expression vector pcDNA. This expression vector was used to transiently transfect human epithelial kidney (HEK) 293T cells or BEAS-2B (Homo sapiens, lung) using Mirus TransIT®-LT1 (Mirus Bio) transfection reagent. Prior to transfection, cells were replaced with the fresh corresponding medium supplemented. The TransIT®-LT1 reagent/DNA complex was added to the cells and incubated for 72 h at 37 ºC. Proteins was purified from cell supernatants using Ni-NTA affinity column (GE Healthcare). The purified SARS-CoV-2 S proteins were concentrated by Amicon Ultrafiltration Unit (MW100K cutoff) (Millipore) in PBS, pH 7.4. The purity was monitored by using SDS-PAGE and the proteins were confirmed using Western blot with anti-$(his)_6$ antibodies (Qiagen) or specific anti-SARS-CoV-2 S protein antibody and horseradish peroxidase-conjugated secondary antibody (PerkinElmer). Finally, the trimer form of SARS-CoV-2 S proteins were obtained by using size-exclusion column, Superose 6 Increase 10/300 GL gel filtration column (GE Healthcare).

Plasmid constructions, expression and purification of human ACE2 protein. The soluble ectodomain of human ACE2 protein with sequence (1-615 or 1-740) fused with a 8xHisTag at the C-terminus was synthesized by Genscript and cloned to the vector pcDNA. For Human ACE2 proteins, the expression construct was transfected in HEK 293T cells using Mirus TransIT®-LT1. Human ACE2 protein was purified from cell supernatants using Ni-NTA resin (GE Healthcare). The purified proteins were concentrated using Amicon (Millipore) and further purified by gel filtration chromatography on a Superdex 200 Increase 10/300 GL gel filtration column (GE Healthcare).

ELISA determination for SARS-CoV-2 S and human ACE2 protein binding. Ninety-six-well ELISA plates (Greiner bio-one, Frickenhausen, Germany) were coated with 100 µl of human ACE2 protein diluted in ELISA coating buffer, 100 mM sodium bicarbonate (pH 8.8), at a concentration of 2 µg/ml per well were covered with a plastic sealer at 4 °C for overnight. After blocked with 1% BSA in PBST (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.4 mM KH2PO4, 0.1% Tween 20, pH 7.4) containing 5% (w/v) skim milk at 37 °C for 1 h and washed 3 times with PBST, the plates were incubated with 100 µl of SARS-CoV-2 S protein in 2-fold serial dilutions at 37 ºC for 1 h. After washes with PBST, the plates were incubated with anti-SARS-CoV-2 S antibody (1:2000) at 37 °C

for 1 h. The plates were washed with PBST three times and then incubated with 200 µl of secondary HRP-conjugated goat anti-rabbit IgG (1:5000) (Jackson ImmunoResearch). After 1 h of incubation at 37 ºC, the plates were washed 3 times with PBST and developed with 50 µl of the 1-Step Ultra TMB substrate (Thermo Scientific) for 4 min. The reaction was stopped with addition of 50 µl of 1 M H2SO4. The absorbance of wells was measured at 450 nm using a SpectraMax M5 (Molecular Devices, Sunnyvale, CA, USA).

Plasmid constructions, expression and purification of recombinant S protein vaccine. The SARS-CoV-2 spike sequence was downloaded from GISAID database, (Wuhan/WH01/2019 strain) and was codon optimized for human cell expression. The furin cleavage site was replaced with GSAG and the 2P substitution was applied for protein to be stabilized in the prefusion state. The transmembrane domain was replaced with the thrombin cleavage site, foldon, and the histidine-tag at the C terminus. The modified S protein sequence was cloned into pTT vector for protein expression and purification.[20]

Expression and purification methods are modified from the previously published.[21] The plasmid that encodes the secreted form of SARS-CoV-2 S protein was transfected into the human embryonic kidney cell lines of either HEK293EBNA (ATCC number CRL-10852) or the HEK293S GnTI¯ cells using polyethyleneimine. The supernatant was collected 5-6 days after transfection and clarified by centrifugation. S_{fg} was purified as described earlier. S_{hm} protein was purified with Nickel-chelation chromatography and eluted fractions were concentrated by a Millipore Amicon Ultra Filter (100 kDa) and loaded onto a Superpose 6 gel filtration column (10/300 GL, Cytiva) preequilibrated with 20mM Tris/HCl, pH 8.0, 150mM NaCl, with the corresponding trimer fractions pooled and further concentrated. The purified S_{hm} was treated with Endo H (NEB) overnight at room temperature in a ratio of 1:100 (w/w) to produce S_{mg} protein with a single GlcNAc at each N-glycosylation site. For Endo H removal, S_{mg} was either purified through buffer exchange using Millipore Amicon Ultra Filter (100 kDa) for three times, or repurified by size-exclusion chromatography using ENrich SEC 650 (10 x 300 column; Bio-Rad) preequilibrated in the same buffer as above. Then, to check the protein purity, samples were mixed with SDS loading buffer, separated by 7.5% SDS-PAGE and stained by Coomassie Brilliant Blue-Plus (EBL).

N-Linked glycosylation profile on SARS-CoV-2 S protein by mass spectrometry. 20 µg of SARS-CoV-2 S protein, purified from three different cell lines, in three independent biological replicates, were incubated at 55°C for 1h in 50 mM triethylammonium bicarbonate buffer (pH 8.5) containing 10 mM of tris (2-carboxyethyl) phosphine. Next, the reduced S protein was alkylated by adding 18 mM iodoacetamide (IAA) and the mixture was incubated for 30 min in the dark at room temperature. The alkylated proteins were digested separately using different proteases and their combinations: chymotrypsin and alpha lytic protease at a ratio of 1:10 (w/w), or trypsin at a ratio of 1:20 (w/w). All proteases are in sequencing grade (Promega). After an overnight digestion at 37°C, the samples were acidified and processed for LC-MS/MS (liquid chromatography with tandem mass spectrometry) analysis. Glycans categorization follows previous studies according to the composition detected and visualized by Graphpad Prism 9.0.0.[13]

Animal vaccination and virus challenge. For mice vaccination study, female 6- to 8-week-old BALB/c mice (n= 5) were immunized intramuscularly with 20ug purified S_{fg} or S_{mg} proteins mixed with aluminum hydroxide (20ug) at day 0, day 14 and day 56. Blood was collected at week 4, 6, 8, 10 and 12 after first immunization, and serum samples were collected from each mouse.

For hamster vaccination study and virus challenge study, female 6- to 7-week-old Golden Syrian Hamsters (n= 5) were immunized intramuscularly with 25ug of purified S_{fg} or S_{mg} proteins mixed with aluminum hydroxide (250ug) at day 0 and day 14. Blood was collected 28 days and 42 days after first immunization, and serum samples were collected from each hamster. Hamsters were challenged at 4 weeks after second vaccination with 1 x 10^4 $TCID_{50}$ of SARS-CoV-2 TCDC#4 (hCoV-19/Taiwan/4/2020, GISAID accession ID: EPI_ISL_411927) intranasally in a volume of 100 µl per hamster. Body weight for each hamster was recorded daily after infection. On day 3 after challenge, hamsters were euthanized by carbon dioxide. The superior lobe of left lung was fixed in 10% paraformaldehyde for histopathological examination and the rest of lung was collected for viral load determination (TCID50 assay).

For transgenic mice vaccination and virus challenge study, male 8-week-old CAG-hACE2 transgenic mice (n= 3) were immunized intramuscularly with 10 ug of purified S_{fg} or S_{mg} proteins mixed with aluminum hydroxide (50 ug) at day 0 and day 14.[30] Blood was collected 28 days and 42 days after first immunization, and serum samples were collected from each transgenic mouse. Transgenic mice were challenged at 6 weeks after second vaccination with 1 x 10^3 $TCID_{50}$ of SARS-CoV-2 TCDC#4 (hCoV-19/Taiwan/4/2020, GISAID accession ID: EPI_ISL_411927) intranasally in a volume of 50 µL per mice. Body weight and body temperature for each transgenic mouse were recorded daily after infection. On day 7 after challenge, all transgenic mice were euthanized by carbon dioxide. The superior lobe of left lung was fixed in 10% paraformaldehyde for histopathological examination and the rest of lung was collected for viral load determination (TCID50 assay).

All animal experiments were evaluated and approved by the Institutional Animal Care and Use Committee of Academia Sinica.

Histopathology, immunohistochemistry (IHC) staining and immunofluorescence (IF) staining. Hamster organs at 3 dpi were immediately collected and placed in 10% neutral buffered formalin fixation for 24 h, then transferred into 70% ethanol for 72 h. Paraffin-embedded organs tissues were trimmed to the thickness of 5 mm. For histological staining, tissue was stained with hematoxylin and eosin (H & E), followed by microscopic examination. The histopathological scoring system was used for evaluating lung section cumulative histopathology score.[54] For immunohistochemistry (IHC) staining, the tissue sectioned were deparaffinized with xylene and rehydrated with ethanol gradient. Antigen retrieval was performed by heating the slides to 95°C for 10 min in 10 mM sodium citrate buffer (pH 6.0) in a microwave oven. After cooling at room temperature and washing with PBS, 3% H_2O_2 was pplied to eliminate endogenous peroxidase activity. The tissue sectioned were blocked with 5% normal goat serum and 1 % BSA in 1x PBST for 1 h, followed by incubation with rabbit anti-N primary antibody at 1:50 dilution (Anti-SARS-CoV-2 polyclonal antibody) overnight at 4°C. Then the tissue was incubated with goat anti-rabbit HRP secondary antibody at 1:500 dilutions for 1

h and visualized by incubation with 3,3-diaminobenzidine (DAB) substrate and counterstained with hematoxylin. For immunofluorescence staining, after antigen retrieval steps, tissue was permeabilized with Triton X-100 in PBS. The tissue sectioned were blocked with 5% normal goat serum and 1% BSA in 1x PBST for 1 h. Then, incubated with an autofluorescence quencher for 5 min. The samples were subsequently incubated with rabbit anti-N primary antibody at 1:50 dilution (Anti-SARS-CoV-2 polyclonal antibody) overnight at 4ºC, secondary antibody Alexa Fluor-488 (1:500, Thermo Fisher) for 1 h at room temperature, and 4,6-diamidino-2-phenylindole (DAPI), a nuclear dye for 3 min at room temperature. The coverslips were mounted on microscope slides and imaged under a Leica TCS SP8X confocal microscope with HC PL APO CS2 10x/0.40 lens (Leica AG, Wetzlar, Germany).

Sera antibody titer evaluation. Anti-S protein ELISA were used to determine sera IgG titer. Plates were first coated with 50 ng/well of S protein expressed by HEK293EBNA (S_{fg}) and then blocked with 5% skim milk. Mouse polyclonal anti-S protein primary antibody and HRP-conjugated secondary antibody were sequentially added. Peroxidase substrate solution (TMB) and 1M H_2SO_4 stop solution were used and absorbance (OD 450 nm) read by a microplate reader. Tested strains included SARS-CoV-2 variants (wild type, variants B.1.1.7 and B.1.135), RaTG13, and SARS-CoV-1.

Quantification of viral titer in lung tissue by cell culture infectious assay (TCID50). The middle, inferior, and post-caval lung lobes of hamsters were homogenized in 4 ml of DMEM with 2% FBS and 1% penicillin/streptomycin using a homogenizer. Tissue homogenate was centrifuged and the supernatant was collected for live virus titration. Briefly, 10-fold serial dilutions of each sample were added onto Vero E6 cell monolayer in duplicate and incubated for 4 days and cells were observed by microscope daily. The plates were washed with tap water and scored for infection. The fifty-percent tissue culture infectious dose (TCID50) / ml was calculated by the Reed and Muench method.

Pseudovirus neutralization assay for serum and antibody study. For production and purification of SARS-CoV-2 pseudotyped lentivirus, the pseudotyped lentivirus carrying SARS-CoV-2 S protein was generated by transiently transfecting HEK-293T cells with pCMV-ΔR8.91, pLAS2w.Fluc. Ppuro and pcDNA3.1-nCoV-SΔ18 (or pcDNA3.1-nCoV-SΔ18 D614G). HEK-293T cells were seeded one day before transfection, and indicated plasmids were delivered into cells by using TransITR-LT1 transfection reagent (Mirus). The culture medium was refreshed at 16 h and harvested at 48 h and 72h post-transfection. Cell debris were removed by centrifugation at 4,000 xg for 10 min, and the supernatant was passed through 0.45 μm syringe filter (Pall Corporation). The pseudotyped lentivirus was aliquot and then stored at -80°C.

To estimate the lentiviral titer by AlarmaBlue assay, the transduction unit (TU) of SARS-CoV-2 peudotyped lentivirus was estimated by using cell viability assay in responded to the limited dilution of lentivirus. In brief, HEK-293T cells stably expressing human ACE2 gene were plated on 96-well plate one day before lentivirus transduction. For determining the titer of pseudotyped lentivirus, different amounts of lentivirus were added into the culture medium containing polybrene (final

concentration 8 μg/ml). Spin infection was carried out at 1,100 xg in 96-well plate for 30 min at 37°C. After incubating cells at 37°C for 16 h, the culture medium containing virus and polybrene was removed and replaced with fresh complete DMEM containing 2.5 μg/ml puromycin. After treating puromycin for 48 h, the culture media was removed and the cell viability was detected by using 10% AlarmaBlue reagents according to manufacturer's instruction. The survival rate of uninfected cells (without puromycin treatment) was set as 100%. The virus titer (transduction units) was determined by plotting the survival cells versus diluted viral dose.

For performing the pseudotyped lentivirus neutralization assay, for the virus neutralization assay, heat-inactivated sera or antibodies were serially diluted with desired dilution and incubated with 1,000 TU of SARS-CoV-2 pseudotyped lentivirus in DMEM (supplemented with 1% FBS and 100 U/ml Penicillin/Streptomycin) for 1 h at 37°C. The mixture was then inoculated with 10,000 HEK-293T cells stably expressing human ACE2 gene in 96-well plate. The culture medium was replaced with fresh complete DMEM (supplemented with 10% FBS and 100 U/ml Penicillin/ Streptomycin) at 16 h post-infection and cells were continuously cultured for another 48 h before performing luciferase assay. For luciferase assay, the expression level of luciferase gene was determined by using Bright-Glo™ Luciferase Assay System (Promega). The relative light unit (RLU) was detected by Tecan i-control (Infinite 500). The percentage of inhibition was calculated as the ratio of RLU reduction in the presence of diluted serum to the RLU value of no serum control and the calculation formula was shown below:

$(RLU^{Control} - RLU^{Serum}) / RLU^{Control}$

Plaque reduction assay. Vero E6 cells were seeded into 24-well culture plates in DMEM with 10% FBS and antibiotics 1 day before infection. Sera were heated at 56°C for 30 min to inactivate complement. SARS-CoV-2 D614G variant (hCoV-19/Taiwan/NTU03/2020) or B.1.1.7 (hCoV-19/Taiwan/NTU49/2020) was incubated with antibodies for 1 h at 37°C before adding to the cell monolayer for another hour. Subsequently, virus-antibody mixtures were removed, and the cell monolayer was washed once with PBS before covering with media containing 1% methylcellulose for 5–7 days. The cells were fixed with 10% formaldehyde overnight. After removal of the overlay media, the cells were stained with 0.5% crystal violet, and the plaques were counted. The percentage of inhibition was calculated as $[1 - (VD/VC)] \times 100\%$, where VD and VC refer to the virus titers in the presence and absence of the sera, respectively.

CPE-based neutralization assay. For CPE-based neutralization assay, Vero E6 cells were plated onto a 6-well plate at $2x10^5$ cells/well overnight for 90% confluence. Sera were heated at 56°C for 30 min to inactivate complement and the mixture was diluted in DMEM with 2% FBS and 1% penicillin/streptomycin. The diluted sera were mixed with equal volume of 100 $TCID_{50}$ SARS-CoV-2 TCDC#4 (hCoV-19/Taiwan/4/2020, GISAID accession ID: EPI_ISL_411927) at 37°C for 1 h before adding onto the monolayer. The plate was then incubated at 37°C until cytopathic effects (CPE) were observed.

FACS analysis and sorting of S protein-specific B cells. Splenocytes isolated from S_{mg} immunized

mice were incubated with 2 μg/ml S protein at 4 ºC for 1 h, followed by washing and incubation with an antibody cocktail containing the following antibodies: PE-Cy7-conjugated anti-CD19 (Biolegend, 6D5), PE-conjugated anti-CD3 (Biolegend, 17A2), and APC-conjugated anti-His (Biolegend, J095G46), at 4℃ for 15 min. Propidium Iodide was used to exclude dead cells. Stained cells were washed, analyzed, and sorted for S protein-specific single B cells ($CD3^{-}CD19^{+}$) by BD FACSAria II into 96-well PCR plates (Applied Biosystems) containing 10 μl/well catch buffer (10 mM Tris-HCl pH 8, 5 U/μl RNasin (Promega)).

Single-B cell screening. Primers were designed based on a previous publication[55]. The reaction was then performed at 50 ºC for 30 min, 95 ºC for 15 min followed by 40 cycles at 94 ºC for 30 s, 50ºC for 30 s, 72ºC for 1 min, and final incubation at 72 ºC for 10 min. Semi-nested second round PCR was performed using KOD One PCR master mix (TOYOBO) with 1 μl of unpurified first round PCR product at 98 ˚C for 2 min followed by 45 cycles of 98 ºC for 10 s, 55 ºC for 10 s, 68 ºC for 10 s, and final incubation at 68 ºC for 1 min. PCR products were then analyzed on 1.5% agarose gels and sequencing. The Ig V and L genes were identified by searching on IMGT website (http://imgt.org/IMGT_vquest/input). The genes were then amplified from second round PCR product with single gene-specific V and L gene primers containing restriction sites for cloning into vectors containing human IgH or IgL expression backbone. The chimeric IgH and IgL expression constructions were co-transfected into Expi293 for antibody production.

Binding of antibody with S protein expressing 293T surface. 293T cells were transfected with pcDNA6/Spike-P2A-eGFP. Transfected cells were selected under 10 μg/ml of blasticidin for 2-3 weeks. Selected cells were then sorted by FACSAria II to obtain $eGFP^{+}$ expressing cells. These cells were maintained in DMEM containing 10% FBS and 10 μg/ml of blasticidin. 2-3x10^5 cells were incubated with serially diluted antibody in FACS buffer on ice for 1 h. Then, cells were washed with FACS buffer for 3 times, followed by staining in BV421 mouse anti-human IgG (BD Biosciences, 562581, 1:100) on ice for 20 min and washed with FACS buffer twice. The percentage of positive cells was quantified using FACS Canto II and the data were analyzed with FlowJo. The S protein variants used here are: WH01 spike: original S protein; D614G: D614G; B.1.1.7: 69-70 deletion, 144 deletion, N501Y, A570D, D614G, P681H, T716I, S982A and D1118H; B.1.351: L18F, D80A, D215G, 242-244 deletion, R246I, K417N, E484K, N501Y, D614G and A701V[35].

Affinity and avidity determination using Octet (bio-layer interferometry). Fab fragment was prepared by using Pierce Fab Micro Preparation Kit (ThermoFisher Scientific) according to the manufacturer's instructions. Briefly, 250 μg m31A7 IgG was digested by incubating with immobilized papain resin at 37℃ for 8 h. Fab was then purified by protein A column. Purified m31A7 IgG or Fab fragments were loaded at 10 or 6.7 μg/ml in kinetics buffer (0.01% endotoxin-free BSA, 0.002% Tween-20, 0.005% NaN_3 in PBS) onto Protein G or FAB2G biosensors (Molecular Devices, ForteBio), respectively. Association and dissociation of S protein or RBD, both from WH01, with IgG or Fab was performed in kinetics buffer at indicated concentration for 5 min and 15 min, respectively. Kd values were calculated using a 1:1 global fit model (Octet).

Statistical analysis. All of the data are expressed as the means ± standard errors of the means. For all

of the analyses, P values were obtained from Student's t-test (unpaired, two tailed) except for the curve comparison using Student's t-test (paired, two tailed) tests. ANOVA followed by the Tukey post-test was used for multiple comparisons. $p<0.05$ was considered to be statistically significant. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Acknowledgments:
We thank the National RNAi Core Facility at Academia Sinica in Taiwan for providing pseudovirus reagents and related services. We thank Chien-Hung Chen and Ya-Ping Lin for glycopeptide liquid chromatography MS/MS analysis and Tsung-Wei Su for SEC-MALS operation. Funding of this project was provided by Academia Sinica Genomics Research Center Summit Project AS-SUMMIT-109 and the Translational Medical Research Program AS-KPQ-109-BioMed as well as MOST grant 109-2113-M-001-009 (to C.M.). We thank the Transgenic Core Facility in IMB (AS-CFII-108–104), Academia Sinica for providing hACE2 transgenic mice. We thank Academia Sinica Biological Electron Microscopy Core Facility for EM technical support. The core facility is funded by the Academia Sinica Core Facility and Innovative Instrument Project (AS-CFII-108-119). We thank Chun-Kai Chang for helping the hamster vaccination experiment.

References

1 Jiang, S., Xia, S., Ying, T. & Lu, L. A novel coronavirus (2019-nCoV) causing pneumonia-associated respiratory syndrome. *Cellular & Molecular Immunology* **17**, 554-554, doi:10.1038/s41423-020-0372-4 (2020).

2 Xu, X. *et al.* Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Science China Life Sciences* **63**, 457-460, doi:10.1007/s11427-020-1637-5 (2020).

3 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273, doi:10.1038/s41586-020-2012-7 (2020).

4 Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* **395**, 497-506, doi:10.1016/s0140-6736(20)30183-5 (2020).

5 Tarighi, P. *et al.* A review of potential suggested drugs for coronavirus disease (COVID-19) treatment. *European Journal of Pharmacology* **895**, 173890, doi:10.1016/j.ejphar.2021.173890 (2021).

6 Taylor, P. C. *et al.* Neutralizing monoclonal antibodies for treatment of COVID-19. *Nature Reviews Immunology*, doi:10.1038/s41577-021-00542-x (2021).

7 Li, Y. *et al.* A Comprehensive Review of the Global Efforts on COVID-19 Vaccine Development. *ACS Central Science* **7**, 512-533, doi:10.1021/acscentsci.1c00120 (2021).

8 Ratre, Y. K., Kahar, N., Bhaskar, L. V. K. S., Bhattacharya, A. & Verma, H. K. Molecular mechanism, diagnosis, and potential treatment for novel coronavirus (COVID-19): a current literature review and perspective. *3 Biotech* **11**, doi:10.1007/s13205-021-02657-3 (2021).

9 Grant, O. C., Montgomery, D., Ito, K. & Woods, R. J. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Scientific reports* **10**, 14991-14991, doi:10.1038/s41598-020-71748-7 (2020).

10 Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges* **1**, 33-46, doi:10.1002/gch2.1018 (2017).

11 Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260, doi:10.1126/science.abb2507 (2020).

12 Watanabe, Y., Bowden, T. A., Wilson, I. A. & Crispin, M. Exploitation of glycosylation in enveloped virus pathobiology. *Biochim Biophys Acta Gen Subj* **1863**, 1480-1497, doi:10.1016/j.bbagen.2019.05.012 (2019).

13 Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **369**, 330-333, doi:10.1126/science.abb9983 (2020).

14 Zhao, P. *et al.* Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host Microbe* **28**, 586-601.e586, doi:10.1016/j.chom.2020.08.004 (2020).

15 Bouwman, K. M. *et al.* Multimerization- and glycosylation-dependent receptor binding of SARS-CoV-2 spike proteins. *PLoS Pathog* **17**, e1009282-e1009282, doi:10.1371/journal.ppat.1009282 (2021).

16 Yang, Q. *et al.* Inhibition of SARS-CoV-2 viral entry upon blocking N- and O-glycan elaboration. *Elife* **9**, e61552, doi:10.7554/eLife.61552 (2020).

17 Zhang, Y. *et al.* Site-specific N-glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins. *Mol Cell Proteomics* **20**, 100058, doi:10.1074/mcp.RA120.002295 (2020).

18 Amanat, F. *et al.* A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nature Medicine* **26**, 1033-1036, doi:10.1038/s41591-020-0913-5 (2020).

19 Goh, J. B. & Ng, S. K. Impact of host cell line choice on glycan profile. *Crit Rev Biotechnol* **38**, 851-867, doi:10.1080/07388551.2017.1416577 (2018).

20 Chen, J.-R. *et al.* Vaccination of monoglycosylated hemagglutinin induces cross-strain protection against influenza virus infections. *Proceedings of the National Academy of Sciences* **111**, 2476, doi:10.1073/pnas.1323954111 (2014).

21 Wang, C.-C. *et al.* Glycans on influenza hemagglutinin affect receptor binding and immune response. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 18137-18142, doi:10.1073/pnas.0909696106 (2009).

22 Francica, J. R. *et al.* Vaccination with SARS-CoV-2 Spike Protein and AS03 Adjuvant Induces Rapid Anamnestic Antibodies in the Lung and Protects Against Virus Challenge in Nonhuman Primates. *bioRxiv : the preprint server for biology*, 2021.2003.2002.433390, doi:10.1101/2021.03.02.433390 (2021).

23 Shinde, V. *et al.* Efficacy of NVX-CoV2373 Covid-19 Vaccine against the B.1.351 Variant. *New England Journal of Medicine* **384**, 1899-1909, doi:10.1056/nejmoa2103055 (2021).

24 Kuo, T.-Y. *et al.* Development of CpG-adjuvanted stable prefusion SARS-CoV-2 spike antigen as a subunit vaccine against COVID-19. *Scientific Reports* **10**, doi:10.1038/s41598-020-77077-z (2020).

25 van Doremalen, N. *et al.* ChAdOx1 nCoV-19 vaccine prevents SARS-CoV-2 pneumonia in rhesus macaques. *Nature* **586**, 578-582, doi:10.1038/s41586-020-2608-y (2020).

26 Mercado, N. B. *et al.* Single-shot Ad26 vaccine protects against SARS-CoV-2 in rhesus macaques. *Nature* **586**, 583-588, doi:10.1038/s41586-020-2607-z (2020).

27 Vogel, A. B. *et al.* BNT162b vaccines protect rhesus macaques from SARS-CoV-2. *Nature*

592, 283-289, doi:10.1038/s41586-021-03275-y (2021).

28 Corbett, K. S. *et al.* Evaluation of the mRNA-1273 Vaccine against SARS-CoV-2 in Nonhuman Primates. *New England Journal of Medicine* **383**, 1544-1555, doi:10.1056/NEJMoa2024671 (2020).

29 Muñoz-Fontela, C. *et al.* Animal models for COVID-19. *Nature* **586**, 509-515, doi:10.1038/s41586-020-2787-6 (2020).

30 Tsai, C.-Y. *et al.* Sex-biased response to and brain cell infection by SARS-CoV-2 in a highly susceptible human ACE2 transgenic model. *bioRxiv*, 2021.2005.2004.441029, doi:10.1101/2021.05.04.441029 (2021).

31 Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827.e819, doi:https://doi.org/10.1016/j.cell.2020.06.043 (2020).

32 Weissman, D. *et al.* D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization. *Cell Host Microbe* **29**, 23-31.e24, doi:10.1016/j.chom.2020.11.012 (2021).

33 Davies, N. G. *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055, doi:10.1126/science.abg3055 (2021).

34 Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology* **5**, 1403-1407, doi:10.1038/s41564-020-0770-5 (2020).

35 Tegally, H. *et al.* Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*, 2020.2012.2021.20248640, doi:10.1101/2020.12.21.20248640 (2020).

36 Wang, G. L. *et al.* Susceptibility of Circulating SARS-CoV-2 Variants to Neutralization. *N Engl J Med*, doi:10.1056/NEJMc2103022 (2021).

37 Wang, P. *et al.* Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130-135, doi:10.1038/s41586-021-03398-2 (2021).

38 Supasa, P. *et al.* Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. *Cell* **184**, 2201-2211.e2207, doi:https://doi.org/10.1016/j.cell.2021.02.033 (2021).

39 Garcia-Beltran, W. F. *et al.* Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2372-2383.e2379, doi:10.1016/j.cell.2021.03.013 (2021).

40 Shinde, V. *et al.* Efficacy of NVX-CoV2373 Covid-19 Vaccine against the B.1.351 Variant. *The New England journal of medicine* **384**, 1899-1909, doi:10.1056/NEJMoa2103055 (2021).

41 Zhou, D. *et al.* Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* **184**, 2348-2361.e2346, doi:10.1016/j.cell.2021.02.037 (2021).

42 Guharoy, R. & Krenzelok, E. P. US Food and Drug Administration (FDA) Emergency Use Authorization: Glass Half Full or Glass Half Empty? *Clinical Infectious Diseases*, doi:10.1093/cid/ciaa1653 (2020).

43 Hansen, J. *et al.* Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* **369**, 1010, doi:10.1126/science.abd0827 (2020).
44 Casalino, L. *et al.* Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Central Science* **6**, 1722-1734, doi:10.1021/acscentsci.0c01056 (2020).
45 Pettersen, E. F. *et al.* UCSF ChimeraX : Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70-82, doi:10.1002/pro.3943 (2020).
46 Park, S.-J. *et al.* CHARMM-GUIGlycan Modelerfor modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **29**, 320-331, doi:10.1093/glycob/cwz003 (2019).
47 Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology* **13**, e1005659, doi:10.1371/journal.pcbi.1005659 (2017).
48 Woo, H. *et al.* Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. *The Journal of Physical Chemistry B* **124**, 7128-7137, doi:10.1021/acs.jpcb.0c04553 (2020).
49 Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637, doi:10.1002/bip.360221211 (1983).
50 Touw, W. G. *et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Research* **43**, D364-D368, doi:10.1093/nar/gku1028 (2014).
51 Klose, D. P., Wallace, B. A. & Janes, R. W. 2Struc: the secondary structure server. *Bioinformatics* **26**, 2624-2625, doi:10.1093/bioinformatics/btq480 (2010).
52 Mitternacht, S. FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Research* **5**, 189, doi:10.12688/f1000research.7931.1 (2016).
53 Schrodinger, LLC. *The PyMOL Molecular Graphics System, Version 1.8* (2015).
54 Sanchez-Felipe, L. *et al.* A single-dose live-attenuated YF17D-vectored SARS-CoV-2 vaccine candidate. *Nature* **590**, 320-325, doi:10.1038/s41586-020-3035-9 (2021).
55 Tiller, T., Busse, C. E. & Wardemann, H. Cloning and expression of murine Ig genes from single B cells. *J Immunol Methods* **350**, 183-193, doi:10.1016/j.jim.2009.08.009 (2009).