Identification of 22 N-glycosites on spike glycoprotein of SARS-CoV-2 and accessible surface glycopeptide motifs: implications for vaccination and antibody therapeutics

Dapeng Zhou,[1,2*#¶]Xiaoxu Tian,[3#]RuibingQi,[4]Chao Peng,[3*]Wen Zhang[5,6*]

[1]*Tongji University School of Medicine, Shanghai, 200092, China;*

[2]*Shanghai Pudong New Area Mental Health Center affiliated with Tongji University School of Medicine, Shanghai, 200092, China*

[3]*National Facility for Protein Science in Shanghai, Zhangjiang Lab, Shanghai Advanced Research Institute, Chinese Academy of Science, Shanghai 201210, China;*

[4]*Innovation Team of Small Animal Infectious Disease, Shanghai Veterinary Research Institute, Chinese Academy of Agricultural Science, Shanghai 200241, China*

[5]*Fudan University Pudong Medical Center, Institutes of Biomedical Sciences;*

[6]*Department of Systems Biology for Medicine, Shanghai Medical College, Fudan University, 200032, China*

[*]Corresponding authors:

dapengzhoulab@tongji.edu.cn

pengchao@sari.ac.cn

wenz@fudan.edu.cn

[#] Contributed equally to this work

¶To whom proofs and reprints should be addressed; Tel/Fax: +86-2165983607

Key words: antibody/cryogenic electron microscopy structure/crystal structures/epitope prediction/glycopeptide/SARS-CoV-2

Running title: a model for paratopes interacting with glycopeptides

Supplementary data:5 figures and 5 tables

Abstract

Coronaviruses hijack human enzymes to assemble the sugar coat on their spike glycoproteins. The mechanisms by which human antibodies may recognize the antigenic viral peptide epitopes hidden by the sugar coat are unknown. Glycosylation by insect cells differs from the native form produced in human cells, but insect cell–derived influenza vaccines have been approved by the US Food and Drug Administration. In this study, we analyzed recombinant severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein secreted from BTI-Tn-5B1-4 insect cells, by trypsin and chymotrypsin digestion followed by mass spectrometry analysis. We acquired tandem mass spectrometry (MS/MS) spectrums for glycopeptides of all 22 predicted N-glycosylated sites. We further analyzed the surface accessibility of spike proteins according to cryogenic electron microscopy and homolog-modeled structures, and available antibodies that bind to SARS-CoV-1. All 22 N-glycosylated sites of SARS-CoV-2 are modified by high-mannose N-glycans. MS/MS fragmentation clearly established the glycopeptide identities. Electron densities of glycans cover most of the spike receptor-binding domain of SARS-CoV-2, except YQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQ, similar to a region FSPDGKPCTPPALNCYWPLNDYGFYTTTGIGYQ in SARS-CoV-1. Other surface-exposed domains include those located on central helix, connecting region, heptad repeats, and N-terminal domain. Because the majority of antibody paratopes bind to the peptide portion with or without sugar modification, we propose a snake-catching model for predicted paratopes: a minimal length of peptide is first clamped by a paratope, and sugar modifications close to the peptide either strengthen or do not hinder the binding.

Introduction

Spike proteins are located on the surface of coronaviruses and serve as entry proteins for infection (Xu et al. 2004). The spike molecules form trimers, which must be cleaved by cellular proteases so that the fusion peptide can facilitate the fusion of virus membrane with the infected cells. The proteases generate S1 and S2 subunits from spike molecules, and the S1 subunit contains the critical receptor-binding domain to bind the ACE2 receptor on host cells. The receptor-binding motif of the receptor-binding domain, rich in tyrosine, forms direct contacts with ACE2. Fusion of the virus with host cells involves several other critical structures of the spike protein, including central helix and heptad repeat 1 and 2 domains.

Spike glycoproteins are major targets for vaccine design and antibody-based therapies for coronaviruses, including Middle East respiratory syndrome coronavirus; severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1), which caused a multi-country outbreak in 2002–2003; and SARS-CoV-2, which is responsible for a pandemic beginning in 2019. Several antibodies targeting spike proteins of SARS-CoV-1 have shown promising efficacy in preclinical trials (Berry et al. 2010, Bian et al. 2009, Greenough et al. 2005, He et al. 2006, He et al. 2005, Hwang et al. 2006, Ishii et al. 2009, Miyoshi-Akiyama et al. 2011, Pak et al. 2009, Prabakaran et al. 2006, Rockx et al. 2008, Sui et al. 2014, Sui et al. 2005, ter Meulen et al. 2006, Traggiai et al. 2004, van den Brink et al. 2005, Zhu et al. 2007). Furthermore, structural studies suggest that domains other than the crucial receptor-binding domain are also potential targets for antibody binding; these include the fusion peptide, heptad repeat 1, and central helix domains (Yuan et al. 2017)(Table SI). In all coronaviruses, spike glycoproteins are densely glycosylated, with more than 20 predicted sites for N-glycosylation. The function of these glycans in immune evasion by the virus remains unknown.

In this study, we analyzed a recombinant SARS-CoV-2 spike protein expressed by insect cells. We acquired tandem mass spectrometry (MS/MS) spectrums for all glycopeptides generated by sequential digestion using trypsin and chymotrypsin. We further analyzed the cryogenic electron microscopy structures of the spike proteins to identify surface-exposed epitopes for antibody recognition and vaccine design.

Results

N-glycosylation sites for spike protein of coronaviruses

A total of 22 N-glycosylation sites were found in the recombinant spike protein of SARS-CoV-2secreted from BTI-Tn-5B1-4 insect cells (Figure 1). All 22 N-glycosites were confirmed by fragment ions of glycan moieties and characteristic b/y ions derived from peptide backbones (Supplemental Figure 1 and 2). Among them, eight are located in the N-terminal domain, two(N331 and N343) are located in the receptor-binding domain but are outside of the receptor-binding motif (limited to amino acids 438–506), and three are located in the rest of the S1 subunit. Nine are located in the S2 subunit. The glycosylation pattern of the spike protein is highly conserved in SARS-CoV-1, SARS-CoV-2, and Middle East respiratory syndrome coronavirus. The N-terminal and heptad repeat 2 domains are densely glycosylated. The fusion peptide domain is neighbored by N-glycosite N657. In contrast, the receptor-binding motif, central helix domain, and heptad repeat 1 domain are free of glycosylation. The majority of N-glycan moieties are a high-mannose type (Table SII), which is consistent with the glycosylation pathway of the BTI-Tn-5B1-4 insect cell line used to produce the recombinant spike protein. To evaluate the efficiency of glycosylation at every N-glycosite, we calculated the relative ion

abundance of glycopeptide and non-glycosylated sequences generated by trypsin and chymotrypsin digestion. All 22 N-glycosylation sites were occupied by N-glycans (Table SIII). For 18 of the 22 N-glycosites, more than 90% of ions were glycosylated peptides. For N-glycosites 331, 1074, 1158, and 1173, more than 80% of ions were glycosylated peptides. We also searched for the O-glycosylated glycopeptides (Table SIV). Preliminary analysis indicated that the ion abundance of O-glycopeptide sequences was less than 3% as compared with non-glycosylated sequences for all predicted O-glycosites (T323, T325, S678, S673, and S686).

By cryogenic electron microscopy structure modeling (Protein Data Bank [PDB]: 5X58) of the SARS-CoV-1 spike protein, 14 sites of N-glycosylation were observed. The Asn-GlcNAc groups were identified at the reducing end of the glycans at atomic resolution (PDB: 5X58, 3.2 Å), and the density maps of extending glycan chains were still visible although the density was relatively weak (Figure 2A, B, and C). The receptor-binding domain region of the SARS-CoV-1spike protein is densely covered by glycans except FSPDGKPCTPPALNCYWPLNDYGFYTTTGIGYQ, which overlaps with a previously identified “Achilles heel” (i.e., vulnerable spot) for antibody binding (Berry, et al. 2010).

The spike protein of SARS-CoV-2 contains 22 N-glycosylation sites (in yellow in Figure 2D). When trimer structures of the S protein of SARS-CoV-1 and SARS-CoV-2 are aligned (root-mean-square deviation~1.32 for single chain), the structures are very similar except for a few loops, such as those at the N-terminal of the N-terminal domain (Supplemental Figure 3). Sequence alignment and structure comparison revealed that the predicted glycosylation sites are highly conserved. Fourteen of 22 sites were observed by cryogenic electron microscopy for the

SARS-CoV-1 S protein, and most predicted sites of SARS-CoV-2 are located similarly to SARS-CoV-1 (Figure 2E). The receptor-binding domains were overall highly conserved with sequence identity (74.5%), structure (root-mean-square deviation~1.14Å), and two identical glycosylation sites near the N-terminal (Figure 2F), while the sequence specificity of epitopes remained unique in some regions (Tables I and II). A similar surfaced-exposed region, or Achilles heel, YQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQ, was identified in the receptor-binding domain of SARS-CoV-2. Interestingly, the Achilles heel regions for both SARS-CoV-1 and SARS-CoV-2 were also free of glycosylation, whereas its neighboring fragments were covered or interacting with glycosylation. This region free of glycosylation is favorable for ACE2 and other protein binding (Figure 2G).

Accessible surface area calculated according to electron density of glycans on spike proteins of SARS-CoV-1 and SARS-CoV-2

Accessible surface area profiling was used for predicting epitopes for monoclonal antibodies (MAbs) (Supplemental Figure 4). Candidate epitopes are listed in Table I and Figure 3. In addition to receptor-binding domains, multiple potential candidate epitopes were found from amino acid sequences at fusion peptide, heptad repeat 1, and central helix domains. Similar sites were found in receptor-binding domains and central helix domains of both viruses (Figure 3). However, unique sites were also found for each virus. For example, a unique epitope existing in SARS-CoV-2, but not in SARS-CoV-1, is the RARR (682–685) site for furin recognition (Supplemental Figure 5).

To evaluate the conservation of spike epitopes on a structural level, we further aligned the epitopes of SARS-CoV-1 and SARS-CoV-2 based on cryogenic electron microscopy structures.

Eleven predicted epitope pairs were found in receptor-binding domain, heptad repeat 1, and central helix (Table II, Figure 4, and Supplemental Figure 6). Two structurally conserved epitope pairs (AH1/ah1 and AH2/ah2) were predicted at the Achilles heel region which interacts with ACE2 (Table II). We also identified two conserved epitope pairs located on the surface of the receptor-binding domain but outside the ACE2-binding region (I/i and II/ii). Epitope pair II/ii has been proven to be a target for recognition by MAb S309 (Pinto et al. 2020), a potent neutralization antibody with half-maximal inhibitory concentration (IC50) at 69 ng/mL.

Discussion

Neutralizing antibodies toward spike proteins are critical for protective immunity. Traggiai et al. reported spike-specific MAbs isolated from a patient who recovered from SARS-CoV-1 infection, with in vitro neutralizing activity ranging from 10^{-8} to 10^{-11}M (Traggiai, et al. 2004). Several other groups have reported MAbs targeting spike (Berry, et al. 2010, Bian, et al. 2009, Greenough, et al. 2005, He, et al. 2006, He, et al. 2005, Ishii, et al. 2009, Miyoshi-Akiyama, et al. 2011, Rockx, et al. 2008, Sui, et al. 2014, Sui, et al. 2005, ter Meulen, et al. 2006, van den Brink, et al. 2005, Zhu, et al. 2007). Spike protein has also been the focus for vaccine development. For example, mice vaccinated with DNA or subunit vaccines composed of spike proteins (or receptor-binding domain of spike proteins) and adjuvants had high titers of immunoglobulin G antibodies and were protected from SARS-CoV-1 or Middle East respiratory syndrome coronavirus infection (Du et al. 2010, Du et al. 2007, Honda-Okubo et al. 2015, Iwata-Yoshikawa et al. 2014, Li et al. 2013, Lu et al. 2010, Sekimukai et al. 2020, Yang et al. 2004, Zhao et al. 2014). Toll-like receptor ligands, delta inulin, and monophosphoryl lipid A were reported as effective adjuvants to be combined with subunit vaccines. However, to avoid the use

of adjuvant, inactivated SARS-CoV-1 or recombinant adeno-associated virus encoding the receptor-binding domain of the SARS-CoV-1 spike protein has also been studied; these induced potent protective antibody responses against infection (Du et al. 2008, Okada et al. 2005, See et al. 2006, Spruth et al. 2006). The safety and efficacy of antibody therapeutics and vaccines in human clinical trials remain to be studied, as well as the mechanisms for specific vaccine components and formulations. For example, pulmonary pathology was reported when alum was used as an adjuvant for a spike protein subunit vaccine (Tseng et al. 2012). Antibody-induced lung injury was also reported in a macaque model of SARS-CoV-1 infection (Liu et al. 2019), which highlights the importance of avoiding antibody-mediated inflammation.

The receptor-binding domain has been a major focus for antibody and vaccine studies. Three antibodies,80R, m396, and F26G19, complexed with the receptor-binding domain of SARS-CoV-1 have been co-crystalized (Hwang, et al. 2006, Pak, et al. 2009, Prabakaran, et al. 2006). All three antibodies recognize non-continuous, conformational epitopes (Table SI). Several MAb clones that recognize linear continuous peptide sequences have also been reported (4D5, 17H9, F26G18, and 201), although co-crystal structures are not available yet.

In this study, we identified the accessible surface area profiling of the receptor-binding domain of SARS-CoV-2 and found a vulnerable region, YQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQ. Previously, the structural counterpart of this region was termed the Achilles heel of SARS-CoV-1 (Berry, et al. 2010). It mostly overlaps with the interface between ACE2 and S protein (Figure 2G). For SARS-CoV-1, multiple MAbs targeting its Achilles heel have been generated, including F26G18,

4D5,CR3006,m396,FM39,CR3014,F26G19,and 80R (Table SI). For antibody and vaccine development, ongoing studies are focusing on epitopes at the Achilles heel of SARS-CoV-2, especially ah1 and ah2 sites (listed in Table II and Figure 4), which directly interact with ACE2.

However, neutralizing antibodies which do not directly compete with ACE2 binding also exist in recovered SARS-CoV-1 and SARS-CoV-2 patients; for example, the CR3022 MAb neutralizes SARS-CoV-1 but not SARS-CoV-2 (Yuan et al. 2020). The S309 MAb isolated from a SARS-CoV-1 patient neutralizes both SARS-CoV-1 and SARS-CoV-2, and structural analysis revealed its epitope to be a glycopeptide sequence located on the N343 glycosite. Notably, this is exactly in structurally conserved regions as we predicted for the II/ii epitope pair (Table II). Clearly, other epitope pairs predicted in our study are candidate targets to isolate neutralizing antibodies as well.

It is well known that predicted epitopes of protein antigens may be masked by glycosylation. Complex datasets and algorithms, such as spatial epitope prediction for protein antigens (SEPPA) 3.0, have been developed which are based on training parameters related to interactions of glycans and surrounding amino acids (Kong et al. 2015). However, no experimental data are available on the effect of glycosylation sites on epitope surfaces. With the recent breakthrough by high-resolution cryogenic electron microscopy, many glycoproteins can be solved and modeled with glycosylation sites. Here we directly exploit experimental data of the SARS-CoV-1 spike protein from high-resolution cryogenic electron microscopy and screened epitopes for the SARS-CoV-2 spike protein by accessible surface area profiling based on homology-modeled structures. By this approach, we have identified an Achilles heel of SARS-CoV-2, as well as

multiple other surface-exposed epitopes within and outside the receptor-binding domain. For example, in the N-terminal domain (NTD) domain of the SARS-CoV-1spike protein, MAbs specific for linear epitopes have been reported (Table SI) (Greenough, et al. 2005). MAbs specific to other regions of the S1 and S2 subunits of SARS-CoV spike proteins were also reported (Miyoshi-Akiyama, et al. 2011). As summarized in Table I, promising antibody binding sites within and outside the receptor-binding domain have been identified for SARS-CoV-2; our future investigations will focus on vaccination studies to validate their function as neutralizing epitopes with preventive and therapeutic effects in virus challenge experiments.

Dense glycosylation of glycoproteins is a well-known strategy used by viruses to conceal surface peptide epitopes that would otherwise elicit antibody responses, as exemplified by the *Env* protein of human immunodeficiency virus 1. However, after decades of effort, MAbs which bind to conformational epitopes on the surface of the *Env* protein have been identified (Garces et al. 2015, Kong, et al. 2015, Kong et al. 2013). Most of these antibodies bind to the N-glycan portion neighboring the peptide epitopes, whereas some antibodies such as MAb 8ANC195 have evolved to recognize peptide epitopes with no dependence on glycan binding (Kong, et al. 2015). For antibodies specific to spike glycoproteins, there are no data available whether their recognition is hindered by the glycosylation of spike. However, antibodies that bind to both peptide and sugar portions of spike glycopeptides exist, such as MAb S309 which binds to a glycopeptide epitope on the N343 glycosylation site (Pinto, et al. 2020). We propose a “snake catching” model: A snake-like epitope is elusive and difficult for an antibody to "catch" because of the highly mobile, "wiggly" sugar chains that hide the peptide portion. Therefore, to overcome the sugar barrier, a minimum length of peptide portion, either conformational or linear

continuous, must first be clamped by a paratope. This clamping effect may either be strengthened by sugars close to the peptide epitope or not hindered by sugar modification. Clearly, surface-exposed glycopeptide motifs are critical for vaccine design.

In summary, our study clearly identified, by MS, all of the 22 N-glycosites of the SARS-CoV-2 spike protein. We have identified a list of linear surface-exposed candidate epitopes in the spike proteins of SARS-CoV-1 and SARS-CoV-2 and demonstrated the advantages of studying the effects of glycosylation with real cryogenic electron microscopy data. These candidate epitopes are critical for screening for MAb therapeutics to treat SARS-CoV-2, as well as mechanistic studies on vaccine development.

Methods

Prediction of glycosylation sites

Spike proteins for SARS-CoV-2(GenBank Accession Number: MN908947),SARS-CoV-1 (AB263618),and Middle East respiratory syndrome virus (KM027290) were predicted by NetNGlyc.

The sequence identity of the spike proteins between SARS-CoV-2 and SARS-CoV-1 is as high as 84%, which is sufficient to build an accurate homolog model. The sequence of MN908947 was submitted to SWISS-MODEL, and the structural model was built against all available homolog structures as templates. One stable conformation of trimer structure models for SARS-CoV-2 is very close to the spike protein structure from SARS-CoV-1(PDB: 5X58), and their root-mean-square deviation of a single protein chain is approximately 1.32Å after the two structures were superimposed and compared in PyMOL(Figure 2D and E).

Expression of a recombinant SARS-CoV-2 spike protein secreted by insect cells

Recombinant baculovirus was generated by a FastBac1 donor vector and DH10Bac *Escherichia coli* strain (Thermo-Fisher Scientific, San Jose, CA). The signal peptide and secretion signal of the spike protein (GenBank Accession Number: MN908947) were directly used in the recombinant protein. The cDNA sequence containing the encoding region of amino acids 1–1224, fused with a 9-histine tag at the C-terminal, was cloned into the pFastbac1 vector. The recombinant baculoviruses were generated by transposon-mediated recombination and used to infect BTI-Tn-5B1-4 insect cells. Recombinant protein was purified by affinity chromatography.

Protein digestion by trypsin and chymotrypsin

S protein was precipitated with trichloroacetic acid solution (6.1N). The protein pellet was subsequently dissolved in 8 M urea in 100mM Tris-HCl, pH 8.5. Tris(2-carboxyethyl)phosphine (5 mM) was added and incubated for 20 minutes at room temperature to reduce the protein, and iodoacetamide (10mM) was subsequently added and incubated for 15 minutes to alkylate the protein. The protein mixture was digested with chymotrypsin (Wako, Richmond, VA) at a 1:100 ratio at 25˚C, followed by trypsin (Promega) at 1:50 ratio (w/w) at 37˚C. The reaction was terminated by adding formic acid, and the peptide mixture was desalted with a mono-Spin C18 column (GL Sciences).

Liquid chromatography (LC) and MS/MS analyses

The desalted peptide mixture was loaded onto a homemade 30-cm analytical column (ReproSil-Pur C18-AQ 1.9-μm resin, Dr. Maisch GmbH, 360μm OD× 75μm ID) connected to an Easy-nLC

1000 system (Thermo Scientific, San Jose, CA) for MS analysis. The mobile phase and elution gradient used for peptide separation were set as follows: 0–1 min, 0%–2% B; 1–10 min, 2%–7% B; 10–90 min, 7%–27% B; 90–112 min, 27%–35% B; 112–115 min, 35%–95% B;115–125 min, 95% B; and 125–127 min, 95%–2% B (buffer A: 0.1% formic acid (FA) in water and buffer B: 0.1% FA in acetonitrile) at a flow rate of 300nL/min. Peptides eluted from the LC column were directly electro-sprayed into the mass spectrometer with the application of a distal 1.8-kV spray voltage. Survey full-scan MS spectra (from m/z 800–2000) were acquired in the orbitrap analyzer (Q Exactive mass spectrometer, Thermo Scientific), with resolution r =70,000 at m/z 400. The top 20 MS/MS events were sequentially generated from the full MS spectrum with a resolution of 35,000, stepped normalized collision energy (20, 30, 40), intensity threshold of 1.2×10^4, automatic gain control target 2×10^5, and maximum injection time of 250 ms of the ions, using an isolation window of 2.0 m/z.

MS data processing

All acquired MS/MS and MS data were interpreted and analyzed as described (Liu et al. 2017) by using pGlyco 2.0 (version 2019.01.01, http://pfind.ict.ac.cn/software/pGlyco/index.html) glycopeptide identification and by using Byologic v3.5 for quantification. Parameters for our database search of intact glycopeptides were as follows: mass tolerance for precursors and fragment ions were set as ± 7 and ±20 ppm, respectively. The enzymes were trypsin and chymotrypsin. Maximal missed cleavage was 2. Fixed modification was carbamidomethylation on all Cys residues (C +57.022 Da). Variable modifications contained oxidation on Met (M +15.995 Da). The N-glycosylation sequon (N-X-S/T, X ≠ P) was modified by changing “N” to “J” (the two shared the same mass). The glycan database was extracted from Glycome DB

(www.glycome-db.org). All identified spectra could be automatically annotated and displayed by the software tool gLabel embedded in pGlyco2.0, which facilitates manual verification. Parameter settings in Byonic were the same as that in pGlyco2.0 except that the built-in N-glycan database (N-glycan 38 insect glycan) was used for database searching. The O-glycan database was homemade according to previously reported glycan structures by Gaunitz et al.(Lindberg et al. 2013). The identified N-glycopeptides were further examined manually to verify the accuracy of identification. The glycopeptides were quantified by Byologic based on the extracted ion chromatogram area under the curve.

Calculation according to electron density of glycans on SARS-CoV-1spike protein

Glycosylation sites were solved and determined from high-resolution cryogenic electron microscopy density maps, and only N-Acetyl-D-glucosamine (Asn-GlcNAc) was determined to represent a whole glycan due to the glycan flexibility and disorder. TheSARS-CoV-1spike protein structure (PDB:5X58), together with the Asn-GlcNAc sites, were applied for molecular interface calculation with PISA (http://www.ccp4.ac.uk/pisa/). All the amino acids linking or interacting with Asn-GlcNAc were selected and excluded in epitope prediction. Besides the interaction between Asn-GlcNAc and amino acids, the effects of the larger structure of glycans extending from every Asn-GlcNAc may also need to be considered, as shown as in Figure 2C, although their electron densities are weak.

Calculation according to homology-modeled structure of SARS-CoV-2 protein

The aforementioned molecular interface calculation procedure was applied to calculate the accessible surface area and screen the corresponding antigen epitopes, except that the

glycosylation effect could not be measured because the structure is not yet available. Because most glycosylation sites are conserved due to the high similarity between these two spike proteins, we could predict the glycosylation site effects in the SARS-CoV-2 spike structure as well. When predicted epitopes coincided with the amino acid residues interacting with Asn-GlcNAc, they were removed from the candidates by cross-reference of the SARS-CoV-1 data.

Funding

This work was supported by National Natural Science Foundation of China grant 31870972 (DZ) and 11179012 (WZ), National Key Research and Development Plan grant 2017YFA0505901, Fundamental Research Funds for the Central Universities 22120200163, and the Outstanding Clinical Discipline Project of Shanghai Pudong (Grant No.: PWYgy2018–10). All these sponsors have no roles in the study design or the collection, analysis, and interpretation of data.

Abbreviations: LC, liquid chromatography; MAb, monoclonal antibody; MS, mass spectrometry; MS/MS, tandem mass spectrometry; PDB, Protein Data Bank; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Conflict of interest disclosures

The authors declare no conflict of interest.

Author contributions

Dapeng Zhou, Chao Peng, and Wen Zhang designed this study. Dapeng Zhou, Xiaoxu Tian, Ruibing Qi, Chao Peng, and Wen Zhang contributed to the collection, analysis, and

interpretation of data. Dapeng Zhou and Wen Zhang wrote the manuscript. All authors read and approved the final manuscript.

References

Berry JD, Hay K, Rini JM, Yu M, Wang LF, Plummer FA, Corbett CR, Andonov A. 2010. Neutralizing epitopes of the SARS-CoV S-protein cluster independent of repertoire, antigen structure or mAb technology. Mabs-Austin, 2:53-66.

Bian C, Zhang XQ, Cai XF, Zhang LQ, Chen ZW, Zha Y, Xu Y, Xu K, Lu W, Yan LC, et al. 2009. Conserved amino acids W423 and N424 in receptor-binding domain of SARS-CoV are potential targets for therapeutic monoclonal antibody. Virology, 383:39-46.

Du L, Zhao G, Lin Y, Sui H, Chan C, Ma S, He Y, Jiang S, Wu C, Yuen KY, et al. 2008. Intranasal vaccination of recombinant adeno-associated virus encoding receptor-binding domain of severe acute respiratory syndrome coronavirus (SARS-CoV) spike protein induces strong mucosal immune responses and provides long-term protection against SARS-CoV infection. J Immunol, 180:948-956.

Du LY, Zhao GY, Chan CCS, Li L, He YX, Zhou YS, Zheng BJ, Jiang SB. 2010. A 219-mer CHO-Expressing Receptor-Binding Domain of SARS-CoV S Protein Induces Potent Immune Responses and Protective Immunity. Viral Immunol, 23:211-219.

Du LY, Zhao GY, He YX, Guo Y, Zheng BJ, Jiang SB, Zhou YS. 2007. Receptor-binding domain of SARS-CoV spike protein induces long-term protective immunity in an animal model. Vaccine, 25:2832-2838.

Garces F, Lee JH, de Val N, de la Pena AT, Kong L, Puchades C, Hua YZ, Stanfield RL, Burton DR, Moore JP, et al. 2015. Affinity Maturation of a Potent Family of HIV Antibodies Is Primarily Focused on Accommodating or Avoiding Glycans. Immunity, 43:1053-1063.

Greenough TC, Babcock GJ, Roberts A, Hernandez HJ, Thomas WD, Coccia JA, Graziano RF, Srinivasan M, Lowy I, Finberg RW, et al. 2005. Development and characterization of a severe acute respiratory syndrome-associated coronavirus-neutralizing human monoclonal antibody that provides effective immunoprophylaxis in mice. J Infect Dis, 191:507-514.

He YX, Li JJ, Heck S, Lustigman S, Jiang SB. 2006. Antigenic and immunogenic characterization of recombinant baculovirus-expressed severe acute respiratory syndrome coronavirus spike protein: Implication for vaccine design. J Virol, 80:5757-5767.

He YX, Lu H, Siddiqui P, Zhou YS, Jiang SB. 2005. Receptor-binding domain of severe acute respiratory syndrome coronavirus spike protein contains multiple conformation-dependent epitopes that induce highly potent neutralizing antibodies. J Immunol, 174:4908-4915.

Honda-Okubo Y, Barnard D, Ong CH, Peng BH, Tseng CTK, Petrovsky N. 2015. Severe Acute Respiratory Syndrome-Associated Coronavirus Vaccines Formulated with Delta Inulin Adjuvants Provide Enhanced Protection while Ameliorating Lung Eosinophilic Immunopathology. J Virol, 89:2995-3007.

Hwang WC, Lin YQ, Santelli E, Sui JH, Jaroszewski L, Stec B, Farzan M, Marasco WA, Liddington RC. 2006. Structural basis of neutralization by a human anti-severe acute respiratory syndrome spike protein antibody, 80R. J Biol Chem, 281:34610-34616.

Ishii K, Hasegawa H, Nagata N, Ami Y, Fukushi S, Taguchi F, Tsunetsugu-Yokota Y. 2009. Neutralizing antibody against severe acute respiratory syndrome (SARS)-coronavirus spike is highly effective for the protection of mice in the murine SARS model. Microbiol Immunol, 53:75-82.

Iwata-Yoshikawa N, Uda A, Suzuki T, Tsunetsugu-Yokota Y, Sato Y, Morikawa S, Tashiro M, Sata T, Hasegawa H, Nagata N. 2014. Effects of Toll-Like Receptor Stimulation on Eosinophilic Infiltration in Lungs of BALB/c Mice Immunized with UV-Inactivated Severe Acute Respiratory Syndrome-Related Coronavirus Vaccine. J Virol, 88:8597-8614.

Kong L, de la Pena AT, Deller MC, Garces F, Sliepen K, Hua YZ, Stanfield RL, Sanders RW, Wilson IA. 2015. Complete epitopes for vaccine design derived from a crystal structure of the broadly neutralizing antibodies PGT128 and 8ANC195 in complex with an HIV-1 Env trimer. Acta Crystallogr D, 71:2099-2108.

Kong L, Lee JH, Doores KJ, Murin CD, Julien JP, McBride R, Liu Y, Marozsan A, Cupo A, Klasse PJ, et al. 2013. Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. Nat Struct Mol Biol, 20:796-803.

Li J, Ulitzky L, Silberstein E, Taylor DR, Viscidi R. 2013. Immunogenicity and Protection Efficacy of Monomeric and Trimeric Recombinant SARS Coronavirus Spike Protein Subunit Vaccine Candidates. Viral Immunol, 26:126-132.

Lindberg L, Liu JN, Gaunitz S, Nilsson A, Johansson T, Karlsson NG, Holgersson J. 2013. Mucin-type fusion proteins with blood group A or B determinants on defined O-glycan core chains produced in glycoengineered Chinese hamster ovary cells and their use as immunoaffinity matrices. Glycobiology, 23:720-735.

Liu L, Wei Q, Lin QQ, Fang J, Wang HB, Kwok H, Tang HY, Nishiura K, Peng J, Tan ZW, et al. 2019. Anti-spike IgG causes severe acute lung injury by skewing macrophage responses during acute SARS-CoV infection. Jci Insight, 4(4):e123158

Liu MQ, Zeng WF, Fang P, Cao WQ, Liu C, Yan GQ, Zhang Y, Peng C, Wu JQ, Zhang XJ, et al. 2017. pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. Nat Commun, 8. 438.

Lu BJ, Huang Y, Huang L, Li B, Zheng ZH, Chen Z, Chen JJ, Hu QX, Wang HZ. 2010. Effect of mucosal and systemic immunization with virus-like particles of severe acute respiratory syndrome coronavirus in mice. Immunology, 130:254-261.

Miyoshi-Akiyama T, Ishida I, Fukushi M, Yamaguchi K, Matsuoka Y, Ishihara T, Tsukahara M, Hatakeyama S, Itoh N, Morisawa A, et al. 2011. Fully Human Monoclonal Antibody Directed to Proteolytic Cleavage Site in Severe Acute Respiratory Syndrome (SARS) Coronavirus S Protein Neutralizes the Virus in a Rhesus Macaque SARS Model. J Infect Dis, 203:1574-1581.

Okada M, Takemoto Y, Okuno Y, Hashimoto S, Yoshida S, Fukunaga Y, Tanaka T, Kita Y, Kuwayama S, Muraki Y, et al. 2005. The development of vaccines against SARS corona virus in mice and SCID-PBL/hu mice. Vaccine, 23:2269-2272.

Pak JE, Sharon C, Satkunarajah M, Auperin TC, Cameron CM, Kelvin DJ, Seetharaman J, Cochrane A, Plummer FA, Berry JD, et al. 2009. Structural Insights into Immune Recognition of the Severe Acute Respiratory Syndrome Coronavirus S Protein Receptor Binding Domain. J Mol Biol, 388:815-823.

Pinto D, Park YJ, Beltramello M, Walls AC, Tortorici MA, Bianchi S, Jaconi S, Culap K, Zatta F, De Marco A, et al. 2020. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. Nature. online ahead of print.

Prabakaran P, Gan JH, Feng Y, Zhu ZY, Choudhry V, Xiao XD, Ji XH, Dimitrov DS. 2006. Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. J Biol Chem, 281:15829-15836.

Rockx B, Corti D, Donaldson E, Sheahan T, Stadler K, Lanzavecchia A, Baric R. 2008. Structural basis for potent cross-neutralizing human monoclonal antibody protection against lethal human and zoonotic severe acute respiratory syndrome coronavirus challenge. J Virol, 82:3220-3235.

See RH, Zakhartchouk AN, Petric M, Lawrence DJ, Mok CP, Hogan RJ, Rowe T, Zitzow LA, Karunakaran KP, Hitt MM, et al. 2006. Comparative evaluation of two severe acute respiratory syndrome (SARS) vaccine candidates in mice challenged with SARS coronavirus. The Journal of general virology, 87:641-650.

Sekimukai H, Iwata-Yoshikawa N, Fukushi S, Tani H, Kataoka M, Suzuki T, Hasegawa H, Niikura K, Arai K, Nagata N. 2020. Gold nanoparticle-adjuvanted S protein induces a strong antigen-specific IgG response against severe acute respiratory syndrome-related coronavirus infection, but fails to induce protective antibodies and limit eosinophilic infiltration in lungs. Microbiol Immunol, 64:33-51.

Spruth M, Kistner O, Savidis-Dacho H, Hitter E, Crowe B, Gerencer M, Bruhl P, Grillberger L, Reiter M, Tauer C, et al. 2006. A double-inactivated whole virus candidate SARS coronavirus vaccine stimulates neutralising and protective antibody responses. Vaccine, 24:652-661.

Sui JH, Deming M, Rockx B, Liddington RC, Zhu QK, Baric RS, Marasco WA. 2014. Effects of Human Anti-Spike Protein Receptor Binding Domain Antibodies on Severe Acute Respiratory Syndrome Coronavirus Neutralization Escape and Fitness. J Virol, 88:13769-13780.

Sui JH, Li WH, Roberts A, Matthews LJ, Murakami A, Vogel L, Wong SK, Subbarao K, Farzan M, Marasco WA. 2005. Evaluation of human monoclonal antibody 80R for immunoprophylaxis of severe acute respiratory syndrome by an animal study, epitope mapping, and analysis of spike variants. J Virol, 79:5900-5906.

ter Meulen J, van den Brink EN, Poon LLM, Marissen WE, Leung CSW, Cox F, Cheung CY, Bakker AQ, Bogaards JA, van Deventer E, et al. 2006. Human monoclonal antibody combination against SARS coronavirus: Synergy and coverage of escape mutants. Plos Med, 3:1071-1079.

Traggiai E, Becker S, Subbarao K, Kolesnikova L, Uematsu Y, Gismondo MR, Murphy BR, Rappuoli R, Lanzavecchia A. 2004. An efficient method to make human monoclonal antibodies from memory B cells: potent neutralization of SARS coronavirus. Nat Med, 10:871-875.

Tseng CT, Sbrana E, Iwata-Yoshikawa N, Newman PC, Garron T, Atmar RL, Peters CJ, Couch RB. 2012. Immunization with SARS Coronavirus Vaccines Leads to Pulmonary Immunopathology on Challenge with the SARS Virus. Plos One, 7.(4):e35421

van den Brink EN, ter Meulen J, Cox F, Jongeneelen MAC, Thijsse A, Throsby M, Marissen WE, Rood PML, Bakker ABH, Gelderblom HR, et al. 2005. Molecular and biological characterization of human monoclonal antibodies binding to the spike and nucleocapsid proteins of severe acute respiratory syndrome coronavirus. J Virol, 79:1635-1644.

Xu YH, Lou ZY, Liu YW, Pang H, Tien P, Gao GF, Rao ZH. 2004. Crystal structure of severe acute respiratory syndrome coronavirus spike protein fusion core. J Biol Chem, 279:49414-49419.

Yang ZY, Kong WP, Huang Y, Roberts A, Murphy BR, Subbarao K, Nabel GJ. 2004. A DNA vaccine induces SARS coronavirus neutralization and protective immunity in mice. Nature, 428:561-564.

Yuan M, Wu NC, Zhu XY, Lee CCD, So RTY, Lv HB, Mok CKP, Wilson IA. 2020. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. Science, 368:630-633.

Yuan Y, Cao DF, Zhang YF, Ma J, Qi JX, Wang QH, Lu GW, Wu Y, Yan JH, Shi Y, et al. 2017. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. Nat Commun, 8:15092

Zhao JC, Li K, Wohlford-Lenane C, Agnihothram SS, Fett C, Zhao JX, Gale MJ, Baric RS, Enjuanes L, Gallagher T, et al. 2014. Rapid generation of a mouse model for Middle East respiratory syndrome. P Natl Acad Sci USA, 111:4970-4975.

Zhu ZY, Chakraborti S, He Y, Roberts A, Sheahan T, Xiao XD, Hensley LE, Prabakaran P, Rockx B, Sidorov IA, et al. 2007. Potent cross-reactive neutralization of SARS coronavirus isolates by human monoclonal antibodies. P Natl Acad Sci USA, 104:12123-12128.

Figure legends

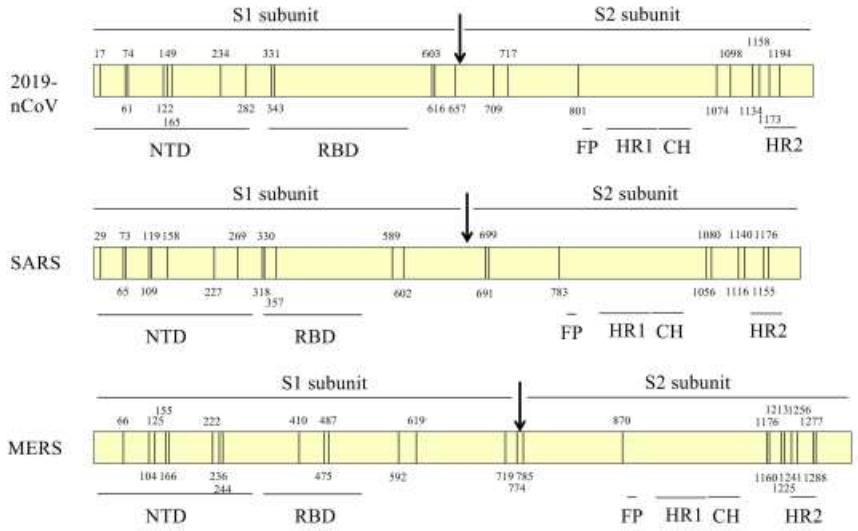

Figure 1. N-glycosylation sites of SARS-CoV-2 (2019-nCoV). The arrows represent the protease cleavage of S1 and S2 subunits. CH, central helix; FP, fusion peptide; HR, heptad repeat; MERS, Middle East respiratory syndrome coronavirus; NTD, N-terminal domain; RBD, receptor-binding domain.

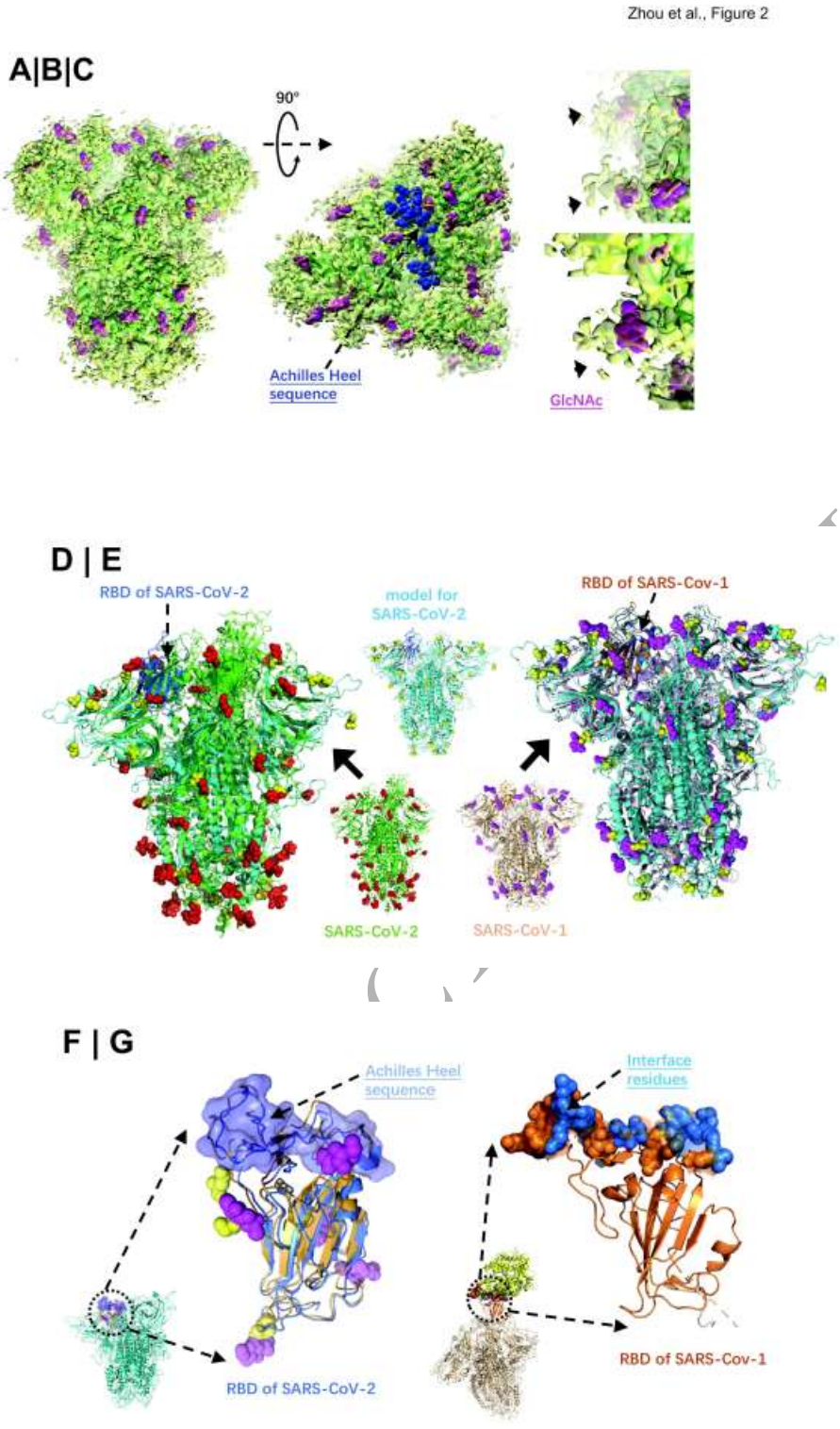

Figure 2. The spike structures of SARS-CoV-1 and SARS-CoV-2. (A) The SARS-CoV-1 spike protein structure(green, Protein Data Bank [PDB]:5X58) and its density map (yellow) with

Asn-linked GlcNAc (pink) from the solvent side view. **(B)**Top view with surface area of receptor-binding domain (RBD) (the Achilles heel, AH, blue) exposed in solvent. **(C)** The typical GlcNAc and its density map, indicated with arrows, extending to outside solvent or neighboring amino acids. **(D)**Structure comparison of S proteins between the SARS-CoV-2 (green, middle, PDB: 6VSB) and corresponding structure model (cyan, middle) with glycosylation sequons (residues not including GlcNAc, yellow spheres) and RBD highlighted (deep blue). GlcNAc for SARS-CoV-2(red spheres, middle) and GlcNAc for SARS-CoV-1 (pink sphere, middle) were highlighted. **(E)** Structure comparison between SARS-CoV-1 (middle, wheat color) and SARS-CoV-2 protein model (cyan); glycosylation sequons of SARS-CoV-2 (residues not including GlcNAc, yellow spheres); GlcNAc for SARS-CoV-1 (pink sphere) were highlighted. **(F)**The comparison of RBDs (dashed line circled on SARS-CoV-2 S protein) between SARS-CoV-1 S protein (RBD: orange) and SARS-CoV-2 protein (RBD: deep blue) with AH surface map (light blue). Note: the glycosylation sequons from SARS-CoV-1 and SARS-CoV-2 S proteins are surrounding the RBD. **(G)**AH fragment (sphere) of RBD (orange) in close-up view (dashed line circled part). The interface residues (spheres in deep blue) between SARS-CoV-1 S protein (wheat color, RBD highlighted as brown) and ACE2 (yellow) from the complex structure (PDB:6ACJ).Note: the interface is exactly located on the AH fragment (brown spheres) of the complex structure (4.2-Åcryogenic electron microscopy structure).

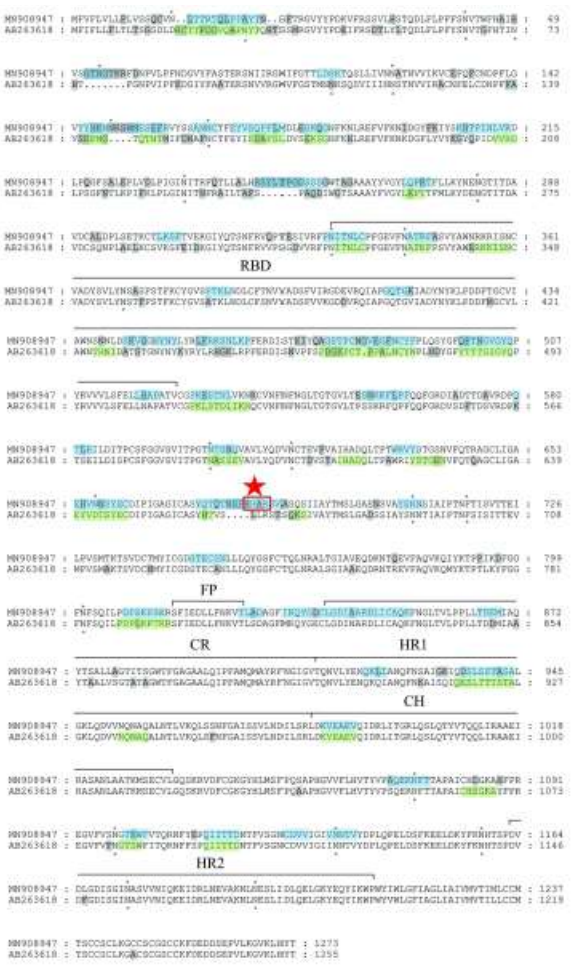

Figure 3. Surface-exposed amino acid sequences predicted by accessible surface area profiling and glycosylation effect with cryogenic electron microscopy structure. Furin site (red star), N-glycosylation sites(*), epitopes for SARS-CoV-1(green), and SARS-CoV-2 (cyan). CH, central helix; CR, connecting region; FP, fusion peptide; HR, heptad repeat; RBD, receptor-binding domain.

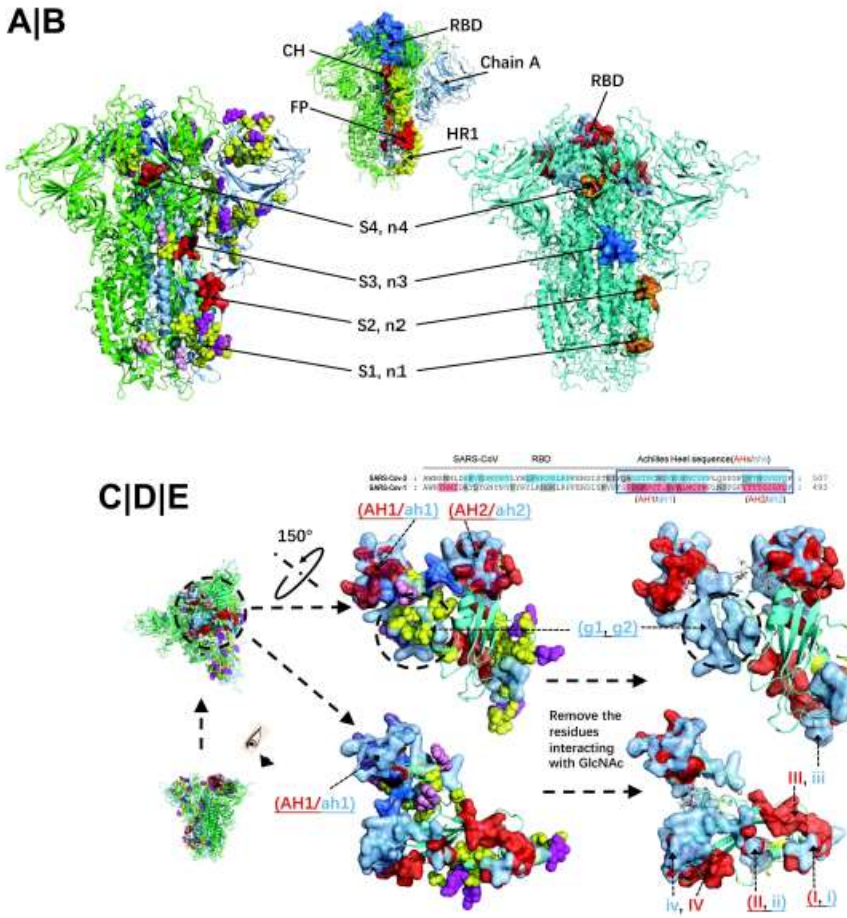

Figure 4. Alignment of 11 predicted epitope pairs on the spike protein structure of SARS-CoV-1 and SARS-CoV-2. In A and B, four epitope pairsS1/n1, S2/n2, S3/n3, and S4/n4 located on surface of spike proteins' heptad repeat (HR1) and central helix (CH) region were compared between SARS-CoV-1 (epitopes in red) and SARS-CoV-2 S proteins (epitopes in orange, or marine blue for site n3). In C, D, and E, seven epitope pairs I/i–IV/iv, AH1/ah1, AH2/ah2, and g1/g2 are compared. The amino acid sequence alignment around Achilles heel (AH) is shown as reference on the top of panel C/D/E (in the same settings as Figure 3). Amino acid sequences of all eleven epitope pairs are listed in Table II. **(A)** SARS-CoV-1 trimer (in green, chain A specifically in sky blue) with Asn-linked GlcNAc (pink at chain A, light pink from other chains) and their interacting amino acids (yellow); the detailed SARS-CoV-1 trimer is shown as

reference between panel A and B, highlighted with the surface map of fusion peptide (red), HR1(yellow), CH(brown), and receptor-binding domain (RBD; deep blue). **(B)** SARS-CoV-2 trimer (in cyan, with RBD of chain A marked as red). **(C)** Top solvent view of the RBD located at one side of trimer structure; the close-up view and the reversed view of RBD structure in dashed circle are in the bottom and top of panels D and E, respectively. **(D)** Comparison of epitopes in RBDs from SARS-CoV-1 (epitopes in red) and SARS-CoV-2 (epitopes in light blue); partially overlapping of AH/ah area is labeled as deep blue, and Asn-linked GlcNAc residues (pink) and their interacting amino acids (yellow) are shown as spheres. Note that the RBD shown here is a closed conformation before binding to ACE2. The AH1/ah1 epitope pair is seen from the top solvent view of RBD. The AH2/ah2 pair is seen from the reversed view, which will be flipped as open conformation of RBD after binding to ACE2. **(E)** View of epitopes in RBDs from SARS-CoV-1 (epitopes in red) and SARS-CoV-2 (epitopes in light blue), when GlcNAc residues and their interacting amino acids were removed.

Table I. Surface-exposed amino acid sequences of SARS-CoV-1 and SARS-CoV-2 (2019-nCoV)

Sites		Epitope details	Nearby N-glycosite	Monoclonal antibody clone	Ref
		SARS-CoV-2			
L18-29	18LTTRTQLPPAYT29		17NLT		
G72-75	72GTNG75		74NGT		
L110-13	110LDSK113		122NAT		
Y144-48	144YYHKN148		149NKS		
W152-58	152WMESEFR158		149NKS		
A163-66	163ANNC166		165NCT		
E169-77	169EYVSQPFLM177				
G181-84	181GKQG184				
K206-15	206KHTPINLVRD215				
R246-56	246RSYLTPGDSSS256		234NIT		
L270-74	270LQPRT274		282NGT		
L303-06	303LKSF306				
P330-36	330PNITNLC336	RBD	331NIT		
A344-47	344ATRF347	RBD	343NAT	S309	(Pinto, et al. 2020)
P384-87	384PTKL387	RBD			
G413-16	413GQTG416	RBD			
S443-51	443SKVG 446,448 NYNY451	RBD		4D5	(He, et al. 2005)
L455-463	455LFRKSNLKP463	RBD			
G476-490	476GSTPC 480,482 GVEGFNCYF490	RBD			
Q498-506	498QPTNGVGYQ506	RBD		201	(Greenough, et al. 2005)
L518-21	518LHAP521	RBD			
P527-33	527PKKSTNL533				
S555-62	555SNKKFLPF562				
Q580-83	580QTLE583				
N603-07	603NTSNQ607		603NTS,616NCT		
W633-36	633WRVY636		657NNS		
E654-62	654EHVNNSYEC662				
Y674-87	674YQTQTNSPRRARSV687				
Y707-71	707YSNN710		709NNS		
S746-51	746STECSN751				
D808-14	808DPSKPSK814		801NFS	5H10	(Miyoshi-Akiyama, et al. 2011)
T827-83	827TLAD830				
I834-54	834IKQYG 838,840 CLGDIAARDLICAQK854	CR			
T866-69	866TDEM869	CR			
Q920-23	920QKLI923	HR1			
D936-44	936DSLSSTASA944	HR1			
K986-91	986KVEAEV991	CH			
A1070-76	1070AQEKNFT1076		1074NFT		
T1100-03	1100THWF1103		1098NGT		
Q1113-18	1113QIITTD1118				
C1126-29	1126 CDVV1129		1134NNT		
V1133-37	1133VNNTV1137		1134NNT		
		SARS-CoV-1			
R18-31	18RCTTFDDVQAPNYT31		29NYT		
K142-15	142KPMG145,146QTHT150		158NCT	68	(Greenough, et al. 2005)
S165-17	165SDAFSL170		158NCT		
E174-77	174EKSG177				
V205-08	205VVRD208				
L257-26	257LKPT260		269NGT		
I319-23	319ITNLC323	RBD	318NIT		
A331-34	331ATKF334	RBD	330NAT	S309	(Pinto, et al. 2020)
R342-47	342RKKISN347	RBD	357NST		
T425-28	425TRNI428	RBD			
P462-76	462PDGKPCTPPALNCYW476	RBD		17H9,F26G18, 80R	(He, et al. 2005, Hwang, et al. 2006)
Y484-92	484YTTTGIGYQ492	RBD		F26G19,m396, 80R,201	(Greenough, et al. 2005, Hwang, et al. 2006, Pak, et al. 2009, Prabakaran, et al. 2006)
P513-22	513PKLSTDLIKN522				
N589-94	589NASSEV594		589NAS		
I610-14	610IHADQ614		602NCT	F26G8	(Berry, et al. 2010)
Y622-27	622YSTGNN627				
E640-48	640EHVDTSYEC648				
H661-73	**661HT662,672KS673**				
P789-97	789PDPLKPTKR797		783NFS	5H10	(Miyoshi-Akiyama, et al.

				2011)
Q917-26	917QESLTTTSTA926	HR1		
N935-39	935NQNAQ939	HR1		
K968-73	968KVEAEV973	CH		
C1064-69	1064CHEGKA1069		1056NFT	
G1081-84	1081GTSW1084		1080NGT	
Q1095-00	1095QIITTD1100			

CH, central helix; CR, connecting region; HR, heptad repeat; RBD, receptor-binding domain.

Table II. Alignment of epitopes on the spike protein structure of SARS-CoV-1 and SARS-CoV-2 (2019-nCov) based on cryogenic electron microscopy structure

RBD ofSARS-CoV-1				RBD of SARS-CoV-2			
I319-23	ITNLC	I	Similar site	P330-36	PNITNLC	i	Similar site
A331-34	ATKF	II	Similar site	A344-47	ATRF	ii	Similar site
R342-47	RKKISN	III	Unique (3AA short peptide in 2019-nCov)	P384-87	PTKL	iii	Inside trimer
Q401-05	Q(T)G(V)I	G1	Removed; discrete sequence, and glyco-interacting AA bracketed	G413-16	GQTG	g1	Glyco-interacting
T425-28	TRNI	IV	Unique (3AAs in 2019-nCov)	S443-51	SKVGNYNY	iv	New (discrete AA distribution on SARS-CoV)
Y442-50	Y(LRH)G(KL R)P	G2	Removed; discrete sequence, and glycol-interacting AAs bracketed	L455-463	LFRKSNLKP	g2	Glyco-interacting
P462-76	PDGKPCTPP ALNCYW	AH1	Similar site	G476-490	GSTPCGVEG FNCYF	ah1	Similar site
Y484-92	YTTTGIGYQ	AH2	Similar site	Q498-506	QPTNGVGYQ	ah2	Similar site
HR1 and CH of SARS-CoV-1				HR1 and CHof SARS-CoV-2			
E900-904	E(N)QK(Q)	S1	Same position, but the glyco-interacting AAs in bracket are removed	Q920-23	QKLI	n1	Glyco-masked
Q917-26	QESLTTTSTA	S2	Similar site	D936-44	DSLSSTASA	n2	Similar site
N935-39	NQNAQ	S3	Buried, exposed due to missing fragment in EM structure	I834-54 *	IKQYGCLGDI AARDLICAQ K	n3	CR (*connecting region, close to S3 in the structure)
K968-73	KVEAEV	S4	Same site	K986-91	KVEAEV	n4	Same site

AA, amino acid; AH, Achilles heel; CH, central helix; CR, connecting region; EM, electron microscopy; HR, heptad repeat; RBD, receptor-binding domain.