 doi:10.20944/preprints202002.0381.v2

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

Dapeng Zhou[1*#], Xiaoxu Tian[2#], Ruibing Qi[3], Chao Peng[2*], Wen Zhang[4,5 *]

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

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

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

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

 doi:10.20944/preprints202002.0381.v2

Abstract

Corona viruses hijack human enzymes to assembly sugar coat on Spike glycoproteins. The mechanism that human antibodies may uncover the antigenic viral peptide epitopes hidden by sugar coat are unknown. In this study, we analyzed recombinant SARS-CoV-2 Spike protein secreted from BTI-Tn-5B1-4 cells, by trypsin and chymotrypsin digestion followed by mass spectrometry analysis. We acquired MS/MS spectrums for glycopeptides of all 22 predicted N-glycosylated sites. We further analyzed the surface accessibility of Spike proteins according to Cryo-EM and homolog-modeled structures, and available antibodies that bind to SARS-CoV-1. The results showed that all 22 N-glycosylated sites of SARS-CoV-2 are modified by high-mannose type of 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 included those located on Central Helix, between amino acids 967 and 1016 of SARS-CoV-1, and 985 to 1034 of SARS-CoV-2 Spike protein. As the majority of antibody paratopes bind to peptide portion with or without sugar modification, we propose a snake-catcher model that a minimal length of peptide is first clamped by a paratope, and the binding is either strengthened by sugars close to peptide, or not interfered by sugar modification.

Key words: SARS-CoV2; corona virus; glycopeptide; N-linked glycans; mass spectrometry; antibody; cryo-EM structure; crystal structures; epitope prediction

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

Introduction

Spike proteins are located on the surface of corona viruses and serve as entry proteins for infection (1). The Spike molecule forms 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 molecule, and the S1 subunit contains the critical receptor binding domain (RBD) to bind ACE2 of host cells. The receptor binding motif (RBM) of the receptor binding domain, rich in tyrosine, forms direct contacts with ACE2. The fusion of the virus with the host cells involves several other critical structures of the Spike protein, including Central Helix (CH) and heptad repeat 1 and 2 (HR1 and HR2) domains.

Spike glycoproteins are major targets for vaccine design and antibody-based therapies for corona viruses. Several antibodies targeting Spike proteins of SARS-CoV showed promising efficacy in preclinical trials (2-18). Besides the crucial RBD, structural studies suggest that other domains including fusion peptide, HR1 and Central Helix are also potential targets for antibody binding (19). In all corona viruses, Spike glycoproteins are densely glycosylated, with more than 20 predicted sites for N-glycosylation. The function of these glycans in immune escape of virus remain unknown.

In this study, we analyzed recombinant SARS-CoV-2 Spike protein expressed by insect cells. We acquired MS/MS spectrum for all glycopeptides generated by sequential digestion using trypsin and chymotrypsin. We further analyzed the cryo-EM structure of Spike proteins, to identify surface-exposed epitopes for antibody recognition as well as vaccine design.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

Results

N-glycosylation sites for coronal viruses

A total of 22 N-glycosylation sites were found in the recombinant Spike protein of SARS-CoV-2 secreted from BTI-Tn-5B1-4 cells (Figure 1). All 22 N-glyco-sites were confirmed by fragment ions of glycan moieties and characteristic b/y ions derived from peptide backbones (Supplemental Figure 1). Among them 8 are located in N-terminal domain (NTD), 2 are located in receptor-binding domain (RBD), 3 are located in the rest of S1 subunit. 9 are located in the S2 subunit. The glycosylation pattern of Spike protein is highly conserved in SARS-CoV-1, MERS, and SARS-CoV-2 corona viruses. The NTD and HR2 domains are densely glycosylated. The fusion peptide (FP) domain is neighbored by N-glyco-site N657. In contrast, the receptor binding motif, the CH domain and the HR1 domain are free of glycosylation. The majority of N-glycan moieties are high-mannose type (Supplemental Table 1&2), which is consistent with the glycosylation pathway of the BTI-Tn-5B1-4 insect cell line used to produce recombinant Spike protein.

By Cryo-EM structure modeling (PDB: 5X58), 14 sites of N-glycosylation were observed. The GlcNAc (NAG) groups were identified at the reducing end of glycans, and the density map of extending glycan chains are still visible although the density is relatively weak (Figure 2A, B, and C). The RBD region of SARS-CoV Spike protein is covered by glycan density except FSPDGKPCTPPALNCYWPLNDYGFYTTTGIGYQ, which overlaps with an “Achilles heel” for antibody binding as pointed out by Berry et al (9).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

The Spike protein of SARS-CoV-2 contain 22 N-glycosylation sites (displayed in yellow in Figure 2D). When trimer structures of S protein of SARS-CoV-1 and SARS-CoV-2 are aligned (RMSD~1.32 for single chain), the structures are very similar except few loops, such as those at the N-terminal of NTD (Supplemental Figure 2). The predicted glycosylation sites are most conserved by sequence alignment and structure comparison. Fourteen of 22 sites are observed by Cryo-EM for SARS-CoV-1 S protein, and most predicted sites of SARS-CoV-2 are located similarly to SARS-CoV (Figure 2E). The RBD domain are overall highly conserved with sequence identity (74.5%), structure (RMSD~1.14Å), and two identical glycosylation-sites near the N terminal (Figure 2F), while the sequence specificity of epitopes remains unique in some region (Tables 1&2). A similar surfaced exposed region, or "Achilles heel", YQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQ, was identified in RBD of SARS-CoV-2. Interestingly, the "Achilles heel" for both SARS-CoV-1 and SARS-CoV-2 is also free of glycosylation, while its neighbor fragments are covered or interacted by glycosylation. This region free of glycosylation is favorable for ACE2 and other protein binding (Figure 2G).

Accessible surface area (ASA) calculated according to electron density of glycans on Spike proteins of SARS-CoV-1 and SARS-CoV-2

The ASA profiling was used for mAb epitopes prediction (Supplemental Figure 3). Candidate epitopes were listed in Table 1 and Figure 3. In addition to RBD domains, multiple potential candidate epitopes from amino acid sequences at FP, HR1 and CH domains. Figure 4 shows the alignment of epitopes of Spike proteins of SARS-CoV-1 and SARS-CoV-2. Similar sites were found in RBD domains and CH domains of both viruses. However, unique sites were also found for each virus (Table 2 and Supplemental Figure 4). A unique epitope only existing in SARS-

 doi:10.20944/preprints202002.0381.v2

CoV-2, but not in SARS-CoV-1, is the RARR (682-685) site for furin recognition (Supplemental Figure 5).

Discussion

Neutralizing antibodies toward Spike proteins are critical for protective immunity. Traggiai et al. reported Spike-specific monoclonal antibodies isolated from a patient who recovered from SARS-CoV infection, with in vitro neutralizing activity ranging from 10^{-8} M to 10^{-11} M (2). Several other groups reported monoclonal antibodies targeting Spike (3-15). Spike protein has also been the focus for vaccine development (20). High titers of IgG antibodies were reported to protect mice from SARS-CoV or MERS-CoV viral infection in mice vaccinated by DNA or subunit vaccines composed by Spike proteins (or RBD of Spike proteins) and adjuvants (21-29). TLR ligands, delta inulin, monophosphoryl lipid A were reported as effective adjuvants to be combined with subunit vaccines. To avoid the use of adjuvant, inactivated SARS-CoV viruses or recombinant adeno-associated virus encoding RBD of SARS-CoV spike protein have been studied, which induced potent protective antibody responses against infection (30-33). The safety and efficacy of antibody therapeutics and vaccines in human clinical trials remain to be studied, as well as the mechanism for specific vaccine component and formulation. For example, pulmonary pathology was reported when alum was used as adjuvant for Spike protein subunit vaccine (34). Antibody-induced lung injury was also reported in macaque model of SARS-CoV infection (35), which highlights the importance to avoid antibody-medicated inflammation.

RBD domain has been a main focus for antibody and vaccine studies. Three antibodies complexed with RBD of SARS-CoV has been co-crystalized, including 80R, m396, F26G19

 doi:10.20944/preprints202002.0381.v2

(16-18). All three antibodies recognize non-continuous, conformational epitopes (Supplemental Table 3). Several mAb clones that recognize linear continuous peptide sequences have been reported (4D5, 17H9, F26G18, and 201), although co-crystal structures are not available yet.

In this study, we have identified the ASA profiling of RBD of SARS-CoV-2, and found a vulnerable region, YQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQ. Previously, the structural counterpart of this region is termed as “the Achilles heel” of SARS-CoV (9). It is mostly overlapped with the interface between ACE2 and S protein (Figure 1G). For SARS-CoV, multiple mAbs targeting the “the Achilles heel” of SARS-CoV have been generated, including F26G18, 4D5, CR3006, m396, FM39, CR3014, F26G19 and 80R (Supplemental Table 3). Ongoing studies are being focused on the epitopes at “the Achilles heel” of SARS-CoV-2 for antibody and vaccine development.

In the past, it is well known that the predicted epitopes of protein antigens may be masked by glycosylation. Complex dataset and algorithm have been developed, which are based on training parameters related to interactions of glycans and surrounding amino acids, such as SEPPA 3.0 (36). However, no experimental data is available on the effect of glycosylation sites on epitope surface. With the recent breakthrough by high-resolution Cryo-EM, many glycoproteins can be solved and modeled with glycosylation sites. Here we directly exploit experiment data of SARS-CoV Spike protein from high resolution Cryo-EM, and screened epitopes for SARS-CoV2 Spike protein by ASA profiling based on homology-modeled structure. By this approach, we have identified the “Achilles heel” of SARS-CoV-2 virus, as well as multiple other surface-exposed epitopes within and outside of RBD. For example, in NBD domain of SARS-CoV-1 Spike

 doi:10.20944/preprints202002.0381.v2

protein, mAbs specific for linear epitopes have been reported (3, Supplemental Table 3). MAbs specific to other regions of S1 subunit and S2 subunits of SARS-CoV Spike protein were also reported (6). As summarized in Table 1, promising antibody binding sites within RBD and outside of RBD have been identified for SARS-CoV-2, future studies will be focused 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 which elicit antibody responses, as exemplified by *Env* protein of HIV-1 virus. However, after decades of effort, monoclonal antibodies which bind to conformational epitopes on surface of the *Env* protein have been identified (36-38). Most of these antibodies bind to N-glycan portion neighboring the peptide epitopes, while some antibodies such as mAb 8ANC195 have evolved to recognize peptide epitope with no dependence on glycan binding (36). For antibodies specific to Spike glycoproteins, there is no data available whether their recognition is interfered by the glycosylation of Spike. We propose a "snake catcher" model that a minimum length of peptide portion, either linear continuous, or conformational, must first be first clamped by a paratope. This clamping effect may either be strengthened by sugars close to the peptide epitope, or not interfered by sugar modification. Clearly, the availability of surface-exposed glycopeptide motifs are critical for inducing antibody responses.

In summary, our study clearly identified all of the 22 N-glyco-sites of SARS-CoV2 Spike protein by mass spectrometry. We have identified a list of linear surface exposed epitopes in Spike proteins of SARS-CoV-1 and SARS-CoV-2, and demonstrated the advantages to study

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

glycosylation effect with real Cryo-EM data. These epitopes are critical for screening of monoclonal antibody therapeutics to treat SARS-CoV-2 viruses, 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), MERS (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 and the structure model was built against all available homolog structures as templates by SWISS-MODEL. One stable conformation of trimer structure models for SARS-CoV-2 is very close to Spike protein structure from SARS-CoV-1 (PDB: 5X58), and their RMSD of single protein chain is about 1.32 Å after two structures are super-imposed and compared in PyMol (Figure 2D&E).

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

Recombinant baculovirus was generated as by a Fastbac.1 donor vector and DH10Bac *E. coli* strain. The signal peptide and secretion signal of Spike protein (GenBank Accession Number: MN908947) were directly used in recombinant protein. The cDNA sequence containing the encoding region of aa1 to 1224, fused with a 9-histine tag at C-terminal, was cloned into pFastbac.1 vector. The recombinant baculoviruses were generated by transposon-mediated

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

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. TCEP (tris(2-carboxyethyl)phosphine, 5 mM) was added and incubated for 20 minutes at room temperature to reduce the protein, and iodoacetamide (10mM) were subsequently added and incubated for 15 minutes to alkylate the protein. The protein mixture was digested with chymotrypsin (Wako) at 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 mono-Spin C18 column (GL Sciences).

LC/MS/MS analysis

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 mass spectrometry 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;125-127 min, 95-2% (buffer A: 0.1% FA in water and buffer B: 0.1% FA in Acetonitrile) at a flow rate of 300 nL/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

 doi:10.20944/preprints202002.0381.v2

Exactive mass spectrometer, Thermo Scientific, San Jose, CA), with resolution r =70,000 at m/z 400. Top 20 MS/MS events were sequentially generated from the full MS spectrum with a resolution of 35, 000, step-NCE (20, 30, 40), intensity threshold of 1.2 x 10^4, AGC target 2 x 10^5 and maximum injection time 250 ms of the ions, using an isolation window of 2.0 m/z.

Mass spectrometry data processing

All acquired MS/MS and MS data were interpreted and analyzed as described (39) by the pGlyco 2.0 (version 2019.01.01, http://pfind.ict.ac.cn/software/pGlyco/index.html) glycopeptide identification, and by Byologic v3.5 for quantification . Parameters for database search of intact glycopeptide were as follows: mass tolerance for precursors and fragment ions were set as ± 7 and ± 20 ppm, respectively. The enzyme 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 GlycomeDB (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. Parameters setting in Byonic were same as that in pGlyco2.0 except the built-in N-glycan database (N-glycan 38 insect glycan) was used for database searching. The identified N-glycopeptides were further examined manually to verify the accuracy of identification. The glycopeptides were quantified by Byologic based on XIC AUC (the extracted ion chromatogram area under the curve).

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

Glycosylation sites were solved and determined from high-resolution Cryo-EM density map, while only N-Acetyl-D-glucosamine (NAG, GlcNAc) is determined to represent a whole glycan due to the glycan flexibility and disorder. The SARS spike protein structure (PDB:5X58), together with the NAG (GlcNAc) sites, were applied for molecular interface calculation with PISA (http://www.ccp4.ac.uk/pisa/). All the amino acids linking or interacting with NAG (GlcNAc) were selected and excluded in epitope prediction. Besides the interaction between NAG (GlcNAc at reducing end) and amino acids, the effects of larger structure of glycans extending from every NAG (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 same molecular interface calculation procedure described above was applied to calculate the ASA and screen the corresponding antigen epitopes, except the glycosylation effect could not be measured due to structure unavailable so far. As most glycosylation sites are conserved due to high similarity of these two spike proteins, we could predict the glycosylation site effects in SARS-CoV-2 spike structure as well. When predicted epitopes collide with the amino acid residues interacting with NAG (GlcNAc), they were removed from the candidates by cross-reference of the SARS-CoV data.

Acknowledgement

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,

and Fundamental Research Funds for the Central Universities 22120180201. All these sponsors have no roles in the study design, or the collection, analysis, and interpretation of data.

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.

Reference:

1. Xu, Y. et al. Crystal structure of severe acute respiratory syndrome coronavirus spike protein fusion core. J. Biol. Chem. 279, 49414–49419 (2004).

2. Traggiai E, Becker S, Subbarao K, Kolesnikova L, Uematsu Y, Gismondo MR, Murphy BR, Rappuoli R, Lanzavecchia A. An efficient method to make human monoclonal antibodies from memory B cells: potent neutralization of SARS coronavirus. Nat Med. 2004 Aug;10(8):871-5.

3. Greenough TC, Babcock GJ, Roberts A, Hernandez HJ, Thomas WD Jr, Coccia JA, Graziano RF, Srinivasan M, Lowy I, Finberg RW, Subbarao K, Vogel L, Somasundaran M, Luzuriaga K, Sullivan JL, Ambrosino DM. Development and characterization of a severe acute respiratory syndrome-associated coronavirus-neutralizing human monoclonal antibody that provides effective immunoprophylaxis in mice. J Infect Dis. 2005 Feb 15;191(4):507-14.

4. Sui J, Li W, Roberts A, Matthews LJ, Murakami A, Vogel L, Wong SK, Subbarao K, Farzan M, Marasco WA. 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. 2005 May;79(10):5900-6.

5. Ishii K, Hasegawa H, Nagata N, Ami Y, Fukushi S, Taguchi F, Tsunetsugu-Yokota Y. 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. 2009 Feb;53(2):75-82.

6. Miyoshi-Akiyama T, Ishida I, Fukushi M, Yamaguchi K, Matsuoka Y, Ishihara T, Tsukahara M, Hatakeyama S, Itoh N, Morisawa A, Yoshinaka Y, Yamamoto N, Lianfeng Z, Chuan Q, Kirikae T, Sasazuki T. 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. 2011 Jun 1;203(11):1574-81.

7. Bian C, Zhang X, Cai X, Zhang L, Chen Z, Zha Y, Xu Y, Xu K, Lu W, Yan L, Yuan J, Feng J, Hao P, Wang Q, Zhao G, Liu G, Zhu X, Shen H, Zheng B, Shen B, Sun B. Conserved amino acids W423 and N424 in receptor-binding domain of SARS-CoV are potential targets for therapeutic monoclonal antibody. Virology. 2009 Jan 5;383(1):39-46.

8. He Y, Lu H, Siddiqui P, Zhou Y, Jiang S. Receptor-binding domain of severe acute respiratory syndrome coronavirus spike protein contains multiple conformation-dependent epitopes that induce highly potent neutralizing antibodies. J Immunol. 2005 Apr 15;174(8):4908-15.

9. Berry JD, Hay K, Rini JM, Yu M, Wang L, Plummer FA, Corbett CR, Andonov A. Neutralizing epitopes of the SARS-CoV S-protein cluster independent of repertoire, antigen structure or mAb technology. MAbs. 2010 Jan-Feb;2(1):53-66. Epub 2010 Jan 27.

10. van den Brink EN, Ter Meulen J, Cox F, Jongeneelen MA, Thijsse A, Throsby M, Marissen WE, Rood PM, Bakker AB, Gelderblom HR, Martina BE, Osterhaus AD, Preiser W, Doerr HW, de Kruif J, Goudsmit J. Molecular and biological characterization of human monoclonal antibodies binding to the spike and nucleocapsid proteins of severe acute respiratory syndrome coronavirus. J Virol. 2005 Feb;79(3):1635-44.

11. Zhu Z, Chakraborti S, He Y, Roberts A, Sheahan T, Xiao X, Hensley LE, Prabakaran P, Rockx B, Sidorov IA, Corti D, Vogel L, Feng Y, Kim JO, Wang LF, Baric R, Lanzavecchia A, Curtis KM, Nabel GJ,

Subbarao K, Jiang S, Dimitrov DS. Potent cross-reactive neutralization of SARS coronavirus isolates by human monoclonal antibodies. Proc Natl Acad Sci U S A. 2007 Jul 17;104(29):12123-8.

12. Sui J, Deming M, Rockx B, Liddington RC, Zhu QK, Baric RS, Marasco WA. Effects of human anti-spike protein receptor binding domain antibodies on severe acute respiratory syndrome coronavirus neutralization escape and fitness. J Virol. 2014 Dec;88(23):13769-80.

13. ter Meulen J, van den Brink EN, Poon LL, Marissen WE, Leung CS, Cox F, Cheung CY, Bakker AQ, Bogaards JA, van Deventer E, Preiser W, Doerr HW, Chow VT, de Kruif J, Peiris JS, Goudsmit J. Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. PLoS Med. 2006 Jul;3(7):e237.

14. He Y, Li J, Heck S, Lustigman S, Jiang S. Antigenic and immunogenic characterization of recombinant baculovirus-expressed severe acute respiratory syndrome coronavirus spike protein: implication for vaccine design. J Virol. 2006 Jun;80(12):5757-67.

15. Rockx B, Corti D, Donaldson E, Sheahan T, Stadler K, Lanzavecchia A, Baric R. Structural basis for potent cross-neutralizing human monoclonal antibody protection against lethal human and zoonotic severe acute respiratory syndrome coronavirus challenge. J Virol. 2008 Apr;82(7):3220-35.

16. Pak JE, Sharon C, Satkunarajah M, Auperin TC, Cameron CM, Kelvin DJ, Seetharaman J, Cochrane A, Plummer FA, Berry JD, Rini JM. Structural insights into immune recognition of the severe acute respiratory syndrome coronavirus S protein receptor binding domain. J Mol Biol. 2009 May 15;388(4):815-23.

17. Prabakaran P, Gan J, Feng Y, Zhu Z, Choudhry V, Xiao X, Ji X, Dimitrov DS. Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. J Biol Chem. 2006 Jun 9;281(23):15829-36.

18. Hwang WC, Lin Y, Santelli E, Sui J, Jaroszewski L, Stec B, Farzan M, Marasco WA, Liddington RC. Structural basis of neutralization by a human anti-severe acute respiratory syndrome spike protein antibody, 80R. J Biol Chem. 2006 Nov 10;281(45):34610-6.

19. Yuan Y, Cao D, Zhang Y, Ma J, Qi J, Wang Q, Lu G, Wu Y, Yan J, Shi Y, Zhang X, Gao GF. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. Nat Commun. 2017 Apr 10;8:15092.

20. He Y, Zhou Y, Wu H, Luo B, Chen J, Li W, Jiang S. Identification of immunodominant sites on the spike protein of severe acute respiratory syndrome (SARS) coronavirus: implication for developing SARS diagnostics and vaccines. J Immunol. 2004 Sep 15;173(6):4050-7.

21. Yang ZY, Kong WP, Huang Y, Roberts A, Murphy BR, Subbarao K, Nabel GJ. A DNA vaccine induces SARS coronavirus neutralization and protective immunity in mice. Nature. 2004 Apr 1;428(6982):561-4.

22. Du L, Zhao G, He Y, Guo Y, Zheng BJ, Jiang S, Zhou Y. Receptor-binding domain of SARS-CoV spike protein induces long-term protective immunity in an animal model. Vaccine. 2007 Apr 12;25(15):2832-8.

23. Lu B, Huang Y, Huang L, Li B, Zheng Z, Chen Z, Chen J, Hu Q, Wang H. Effect of mucosal and systemic immunization with virus-like particles of severe acute respiratory syndrome coronavirus in mice. Immunology. 2010 Jun;130(2):254-61.

24. Du L, Zhao G, Chan CC, Li L, He Y, Zhou Y, Zheng BJ, Jiang S. A 219-mer CHO-expressing receptor-binding domain of SARS-CoV S protein induces potent immune responses and protective immunity. Viral Immunol. 2010 Apr;23(2):211-9.

25. Li J, Ulitzky L, Silberstein E, Taylor DR, Viscidi R. Immunogenicity and protection efficacy of monomeric and trimeric recombinant SARS coronavirus spike protein subunit vaccine candidates. Viral Immunol. 2013 Apr;26(2):126-32.

26. Zhao J, Li K, Wohlford-Lenane C, Agnihothram SS, Fett C, Zhao J, Gale MJ Jr, Baric RS, Enjuanes L, Gallagher T, McCray PB Jr, Perlman S. Rapid generation of a mouse model for Middle East respiratory syndrome. Proc Natl Acad Sci U S A. 2014 Apr 1;111(13):4970-5.

27. Iwata-Yoshikawa N, Uda A, Suzuki T, Tsunetsugu-Yokota Y, Sato Y, Morikawa S, Tashiro M, Sata T, Hasegawa H, Nagata N. 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. 2014 Aug;88(15):8597-614.

28. Honda-Okubo Y, Barnard D, Ong CH, Peng BH, Tseng CT, Petrovsky N. Severe acute respiratory syndrome-associated coronavirus vaccines formulated with delta inulin adjuvants provide enhanced protection while ameliorating lung eosinophilic immunopathology. J Virol. 2015 Mar;89(6):2995-3007.

29. Sekimukai H, Iwata-Yoshikawa N, Fukushi S, Tani H, Kataoka M, Suzuki T, Hasegawa H, Niikura K, Arai K, Nagata N. 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. 2020 Jan;64(1):33-51.

30. Du L, Zhao G, Lin Y, Sui H, Chan C, Ma S, He Y, Jiang S, Wu C, Yuen KY, Jin DY, Zhou Y, Zheng BJ. 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. 2008 Jan 15;180(2):948-56.

31. See RH, Zakhartchouk AN, Petric M, Lawrence DJ, Mok CP, Hogan RJ, Rowe T, Zitzow LA, Karunakaran KP, Hitt MM, Graham FL, Prevec L, Mahony JB, Sharon C, Auperin TC, Rini JM, Tingle AJ, Scheifele DW, Skowronski DM, Patrick DM, Voss TG, Babiuk LA, Gauldie J, Roper RL, Brunham RC, Finlay BB. Comparative evaluation of two severe acute respiratory syndrome (SARS) vaccine candidates in mice challenged with SARS coronavirus. J Gen Virol. 2006 Mar;87(Pt 3):641-50.

32. Spruth M, Kistner O, Savidis-Dacho H, Hitter E, Crowe B, Gerencer M, Brühl P, Grillberger L, Reiter M, Tauer C, Mundt W, Barrett PN. A double-inactivated whole virus candidate SARS coronavirus vaccine stimulates neutralising and protective antibody responses. Vaccine. 2006 Jan 30;24(5):652-61.

33. Okada M, Takemoto Y, Okuno Y, Hashimoto S, Yoshida S, Fukunaga Y, Tanaka T, Kita Y, Kuwayama S, Muraki Y, Kanamaru N, Takai H, Okada C, Sakaguchi Y, Furukawa I, Yamada K, Matsumoto M, Kase T, Demello DE, Peiris JS, Chen PJ, Yamamoto N, Yoshinaka Y, Nomura T, Ishida I, Morikawa S, Tashiro M, Sakatani M. The development of vaccines against SARS corona virus in mice and SCID-PBL/hu mice. Vaccine. 2005 Mar 18;23(17-18):2269-72.

34. Tseng CT, Sbrana E, Iwata-Yoshikawa N, Newman PC, Garron T, Atmar RL, Peters CJ, Couch RB. Immunization with SARS coronavirus vaccines leads to pulmonary immunopathology on challenge with the SARS virus. PLoS One. 2012;7(4):e35421.

35. Liu L, Wei Q, Lin Q, Fang J, Wang H, Kwok H, Tang H, Nishiura K, Peng J, Tan Z, Wu T, Cheung KW, Chan KH, Alvarez X, Qin C, Lackner A, Perlman S, Yuen KY, Chen Z. Anti-spike IgG causes severe

acute lung injury by skewing macrophage responses during acute SARS-CoV infection. JCI Insight. 2019 Feb 21;4(4). . pii: 123158.

36. Kong L, Torrents de la Peña A, Deller MC, Garces F, Sliepen K, Hua Y, Stanfield RL, Sanders RW, Wilson IA. 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 Biol Crystallogr. 2015 Oct;71(Pt 10):2099-108.

37. Kong L, Lee JH, Doores KJ, Murin CD, Julien JP, McBride R, Liu Y, Marozsan A, Cupo A, Klasse PJ, Hoffenberg S, Caulfield M, King CR, Hua Y, Le KM, Khayat R, Deller MC, Clayton T, Tien H, Feizi T, Sanders RW, Paulson JC, Moore JP, Stanfield RL, Burton DR, Ward AB, Wilson IA. Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. Nat Struct Mol Biol. 2013 Jul;20(7):796-803.

38. Garces F, Lee JH, de Val N, de la Pena AT, Kong L, Puchades C, Hua Y, Stanfield RL, Burton DR, Moore JP, Sanders RW, Ward AB, Wilson IA. Affinity Maturation of a Potent Family of HIV Antibodies Is Primarily Focused on Accommodating or Avoiding Glycans. Immunity. 2015 Dec 15;43(6):1053-63.

39. Liu MQ, Zeng WF, Fang P, Cao WQ, Liu C, Yan GQ, Zhang Y, Peng C, Wu JQ, Zhang XJ, Tu HJ, Chi H, Sun RX, Cao Y, Dong MQ, Jiang BY, Huang JM, Shen HL, Wong CCL, He SM, Yang PY. pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. Nature Communications, 8, 438, 2017.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

Figure legends:

Figure 1. N-glycosylation sites of SARS-CoV-2. NTD, N-terminal domain; RBD, receptor binding domain; FP, fusion peptide; HR1, helix region 1; CH, central helix; HR2, helix region 2.

Figure 2. The spike structures of SARS-CoV-1 and SARS-CoV-2

A. The SARS-CoV spike protein structure (green, PDB:5X58) and its density map (yellow) with glycosylation (pink) from the solvent side view;

B. Bottom view with surface area of RBD (the "Achilles Heel", AH, blue) exposed in solvent; **C.** The typical NAG and its density map, indicated with arrows, extending to outside solvent or neighbor amino acids;

D. The SARS-CoV-2 spike protein structure (cyan) with glycosylation amino acids (yellow) and RBD highlighted;

E. Structure comparison between The SARS-CoV-1 (middle) and SARS-CoV-2 protein;

F. The comparison of RBD domains (dash 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 (blue); notes: the glycosylation sites from SARS-CoV-1 and SARS-CoV-2 S protein are surrounding the RBD domain;

G. AH fragment (sphere) of RBD domain (orange) in closeup view (dash line circled part); The interface (blue) between SARS -CoV S protein (wheat) and ACE2 (yellow) from the complex structure (PDB:6ACJ);notes: the interface is exactly located on the AH fragment of the complex structure (4.2 angstroms Cryo-EM structure).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

Figure 3. Surface-exposed amino acid sequences predicted by ASA profiling and glycosylation effect with Cryo-EM structure. Furin site (Red pentagram), N-glycosylation sites (*); epitopes for SARS-CoV-1 (green) and SARS-CoV-2 (cyan).

Figure 4. Alignment of epitopes on the spike protein structure of SRAS-CoV-1 and SARS-CoV-2.

A. The comparison of the protein chain A between SARS-CoV-1 trimer (in green, chain A specifically in sky blue) and SARS-CoV-2 trimer (cyan), with glycosylation sites (pink at chain A, light pink from other Chains) and their interacting amino acids (yellow) for Chain A of SARS-CoV-1;

B. Four epitope pairs S1/n1, S2/n2, S3/n3, and S4/n4 compared between SARS-CoV-1 (epitopes in red) and SARS-CoV2 S protein (epitopes in grey or light blue for site n3), and SARS-CoV-2 S protein cartoon shown individually on right panel; the conserved fragments at FP (red), HR1 (yellow) and CH (orange) shown by small cartoon of SARS-CoV trimer (grey) in the middle. The epitopes pairs are listed in the Table 2.

C. Bottom solvent view of the RBD domain located at one side of trimer structure bottom;

D. Comparison of epitopes in RBD domains from SARS-CoV-1 (epitopes in red) and SARS-CoV-2 trimer (epitopes in light blue, RBD cartoon in cyan), together shown with AH (dark blue for whole AH, partially overlapping with AH/ah for epitopes predicted), glycosylation sites (pink) and their interacting amino acids (yellow).

E. The epitopes pairs I/i~IV/iv, AH/ah and g1/g2 are compared and listed in the Table 2.

 doi:10.20944/preprints202002.0381.v2

Supplemental Online Materials

Supplemental Table 1: Peptide and Glycopeptide mixtures of trypsin and chymotrypsin-digested SARS-CoV-2 Spike protein identified by LC-MS. Red bars indicate molecular ions identified by the software in pGlyco2.0. A total of 22 N-glyco-sites were identified and further confirmed by MS-MS analysis (Supplemental Figure 1).

Supplemental Table 2: List of trypsin and chymotrypsin-digested glycopeptides of SARS-CoV-2 Spike protein identified by LC-MS. The majority of N-glycans are high mannose type.

Supplemental Table 3: List of monoclonal antibodies for Spike protein of SARS-CoV-1

Supplemental Figure 1: MS-MS spectrum of glycan moieties and b/y ions for glycopeptides of 22 N-glycosites. MS^2 spectrum was automatically annotated and displayed by the software tool gLabel embedded in pGlyco2.0. "J" indicates the N-glycosylation site "N". Colored circles indicate the number of sugars in attached glycan: green circle, hexose (H); blue square, N-acetylglucosamine (N); red triangle, fucose (F). Colored peaks in MS^2 spectrum include: green peaks representing the fragment ions of a glycan moiety; blue peaks representing a diagnostic glycan ion; red peaks representing the Y ions from glycan fragmentation; and yellow/cyan peaks representing the b/y ions from peptide backbone fragmentation. Mass deviations of the annotated peaks are shown in the lower box.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

Supplemental Figure 2: Structure-based alignment of SARS-CoV-2 (2019-nCoV) and SARS-CoV-1 Spike proteins. The sequences are directly extracted from PDB 5X58 and 2019-nCoV homology model, and the sequence alignment was based on above two structures by ENDscript and ESPRIPT with default settings (http://espript.ibcp.fr/ESPript/ENDscript/index.php).

Supplemental Figure 3: Accessible surface area profiling of Spike proteins of SARS-CoV-2 (2019-nCoV) and SARS-CoV-1. A) The epitopes predicted on the S protein structure for SARS-CoV, Epi (yellow) denotes the epitopes screened by simple ASA profiling (the same for nCoV), and EpiS (red) denotes the epitopes were calculated by excluding the glycosylation sites and the glyco-interacting amino acids; B) The epitopes predicted for nCoV. The values of Y axis means nm^2 of ASA.

Supplemental Figure 4: Connecting region (CR) of SARS-CoV-2 (2019-nCoV) and SARS-CoV-1 Spike proteins.

Supplemental Figure 5: Furin recognition site of SARS-CoV-2 (2019-nCoV) Spike protein.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

Table 1. Surface exposed amino acid sequences of SARS-CoV-1 and SARS-CoV2 (2019-nCoV)

Sites	Epitope details	Nearby N-glycosite	Mab clone	Ref
2019-nCoV				
L18-29	18 LTTRTQLPPAYT 29	17NLT		
G72-75	72 GTNG 75	74NGT		
L110-13	110 LDSK 113	122NAT		
Y144-48	144 YYHKN 148	149NKS		
W152-58	152 WMESEFR 158	149NKS		
A163-66	163 ANNC 166	165NCT		
E169-77	169 EYVSQPFLM 177			
G181-84	181 GKQG 184			
K206-15	206 KHTPINLVRD 215			
R246-56	246 RSYLTPGDSSS 256	234NIT		
L270-74	270 LQPRT 274	282NGT		
L303-06	303 LKSF 306			
P330-36	330 PNITNLC 336	RBD 331NIT		
A344-47	344 ATRF 347	RBD 343NAT		
P384-87	384 PTKL 387	RBD		
G413-16	413 GQTG 416	RBD		
S443-51	443 SKVG 446,448 NYNY 451	RBD	4D5	8
L455-463	455 LFRKSNLKP 463	RBD		
G476-490	476 GSTPC 480,482 GVEGFNCYF 490	RBD		
Q498-506	498 QPTNGVGYQ 506	RBD	201	3
L518-21	518 LHAP 521	RBD		
P527-33	527 PKKSTNL 533			
S555-62	555 SNKKFLPF 562			
Q580-83	580 QTLE 583			
N603-07	603 NTSNQ 607	603NTS,616NCT		
W633-36	633 WRVY 636	657NNS		
E654-62	654 EHVNNSYEC 662			
Y674-87	674 YQTQTNSPRRARSV 687			
Y707-71	707 YSNN 710	709NNS		
S746-51	746 STECSN 751			
D808-14	808 DPSKPSK 814	801NFS	5H10	6
T827-83	827 TLAD 830			
I834-54	834 IKQYG 838,840 CLGDIAARDLICAQK 854	CR		
T866-69	866 TDEM 869	CR		
Q920-23	920 QKLI 923	HR1		
D936-44	936 DSLSSTASA 944	HR1		
K986-91	986 KVEAEV 991	CH		
A1070-76	1070 AQEKNFT 1076	1074NFT		
T1100-03	1100 THWF 1103	1098NGT		
Q1113-18	1113 QIITTD 1118			
C1126-29	1126 CDVV 1129	1134NNT		
V1133-37	1133 VNNTV 1137	1134NNT		
SARS-CoV				
R18-31	18 RCTTFDDVQAPNYT 31	29NYT		
K142-15	142 KPMG 145,146 QTHT 150	158NCT	68	3
S165-17	165 SDAFSL 170	158NCT		
E174-77	174 EKSG 177			
V205-08	205 VVRD 208			
L257-26	257 LKPT 260	269NGT		
I319-23	319 ITNLC 323	RBD 318NIT		
A331-34	331 ATKF 334	RBD 330NAT		
R342-47	342 RKKISN 347	RBD 357NST		
T425-28	425 TRNI 428	RBD		
P462-76	462 PDGKPCTPPALNCYW 476	RBD	17H9, F26G18,80R	8,18
Y484-92	484 YTTTGIGYQ 492	RBD	F26G19, m396, 80R,201	3,16,17,18
P513-22	513 PKLSTDLIKN 522			
N589-94	589 NASSEV 594	589NAS		
I610-14	610 IHADQ 614	602NCT	F26G8	9
Y622-27	622 YSTGNN 627			
E640-48	640 EHVDTSYEC 648			
H661-73	**661 HT 662,672 KS 673**			
P789-97	789 PDPLKPTKR 797	783NFS	5H10	6
Q917-26	917 QESLTTTSTA 926	HR1		
N935-39	935 NQNAQ 939	HR1		
K968-73	968 KVEAEV 973	CH		
C1064-69	1064 CHEGKA 1069	1056NFT		
G1081-84	1081 GTSW 1084	1080NGT		
Q1095-00	1095 QIITTD 1100			

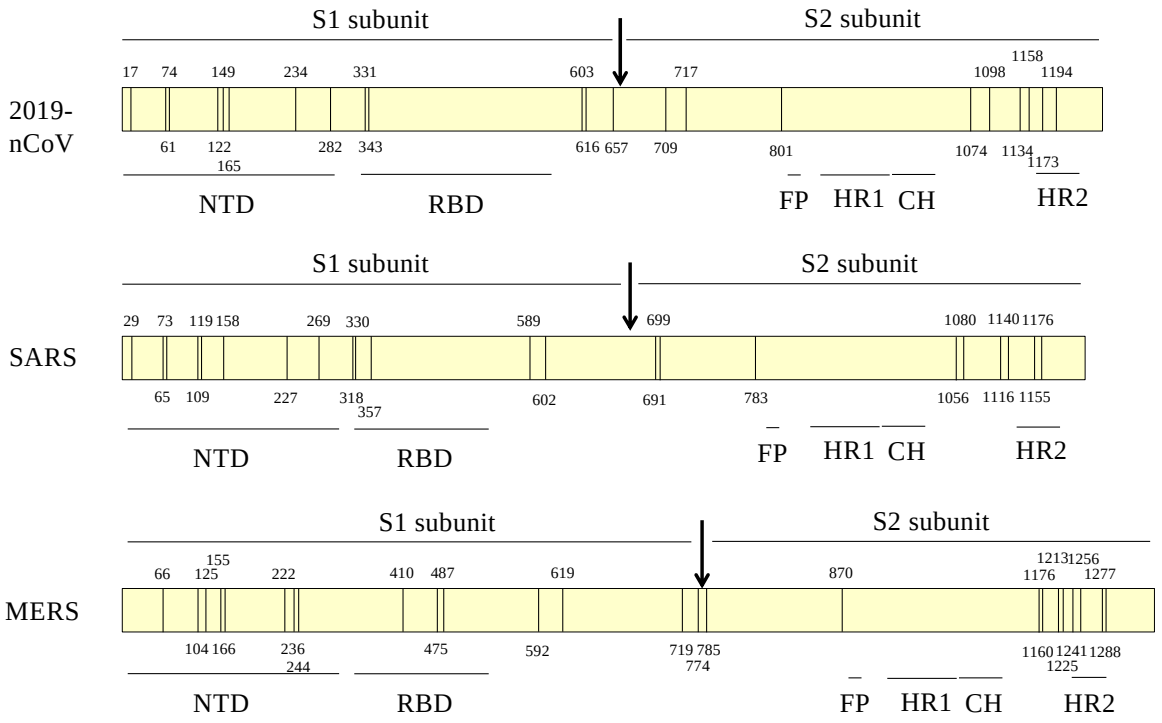
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

Table 2. Alignment of epitopes on the spike protein structure of SRAS and 2019-nCov based on Cryo-EM structure

HR1 and CH of SARS-CoV				HR1 and CH of 2019-nCoV			
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
RBD of SARS-CoV				RBD of 2019-nCoV			
I319-23	ITNLC	I	Similar site	P330-36	PNITNLC	i	Similar site
A331-34	ATKF	II	Same site	A344-47	ATRF	ii	Same 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

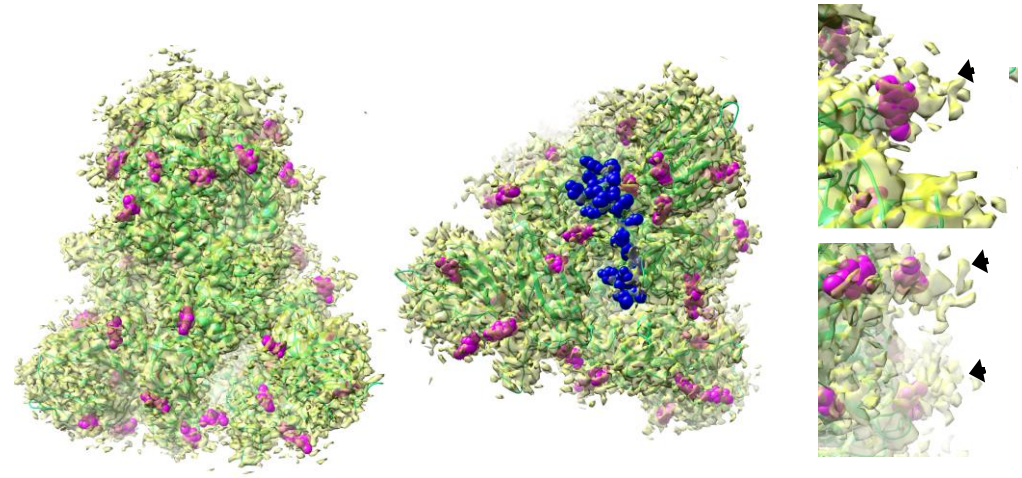
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

Zhou et al., Figure 1



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

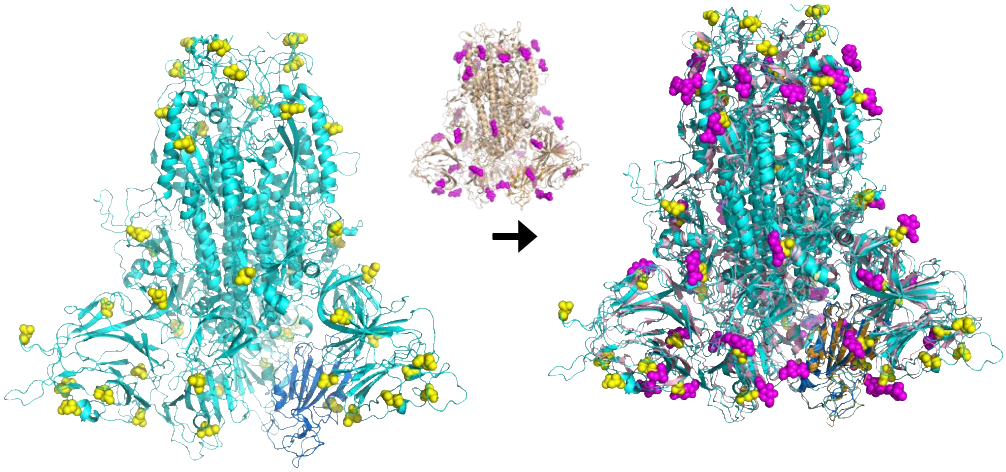
Zhou et al., Figure 2



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

Zhou et al., Figure 2

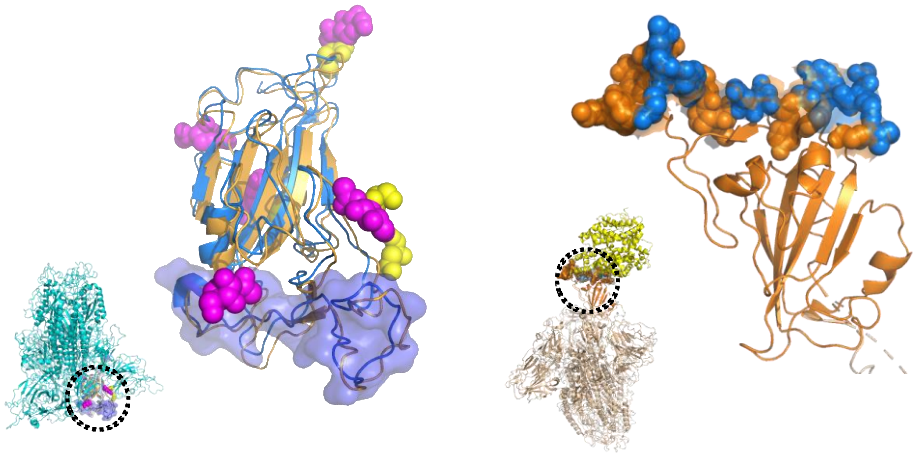
D | E



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

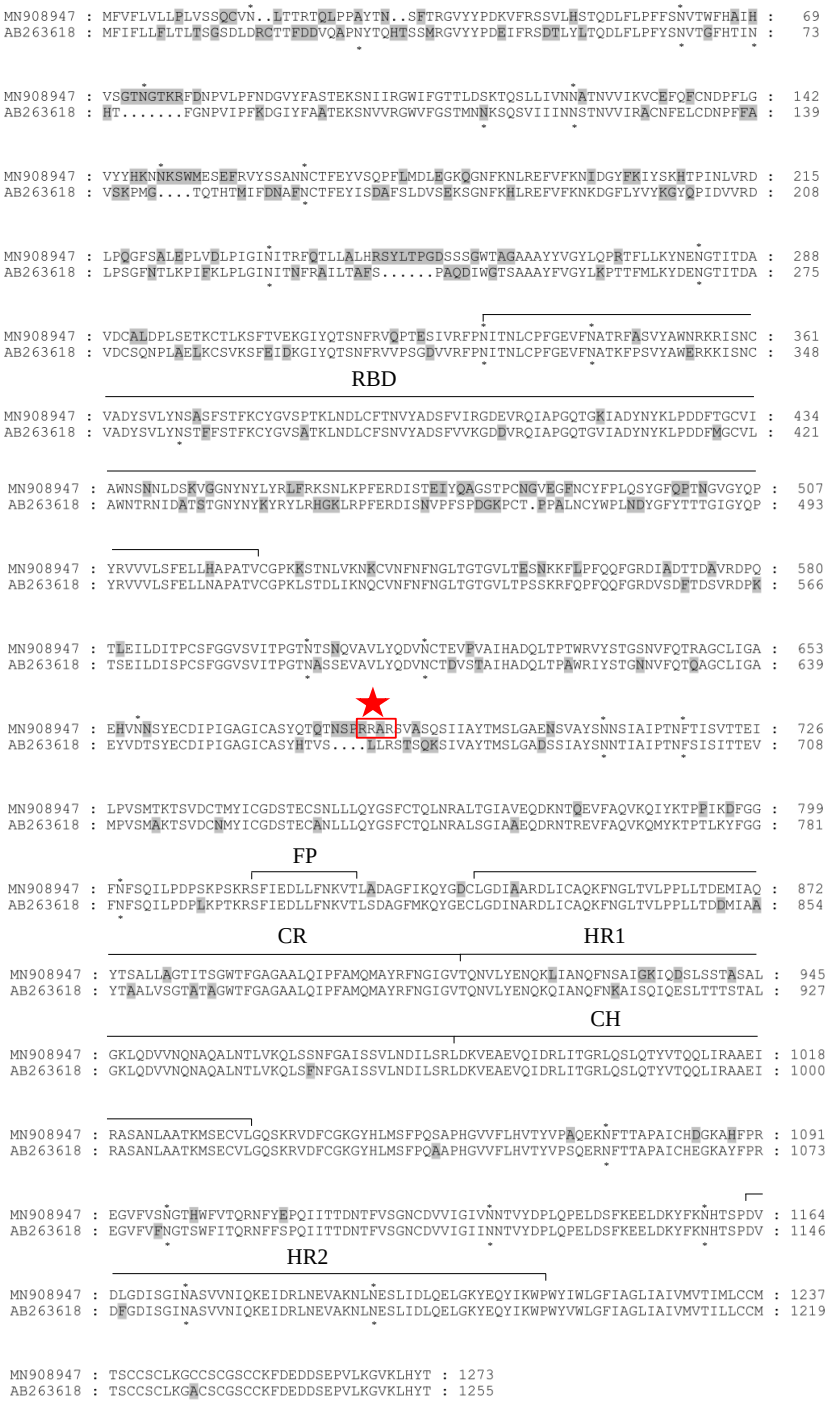
Zhou et al., Figure 2

F | G



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

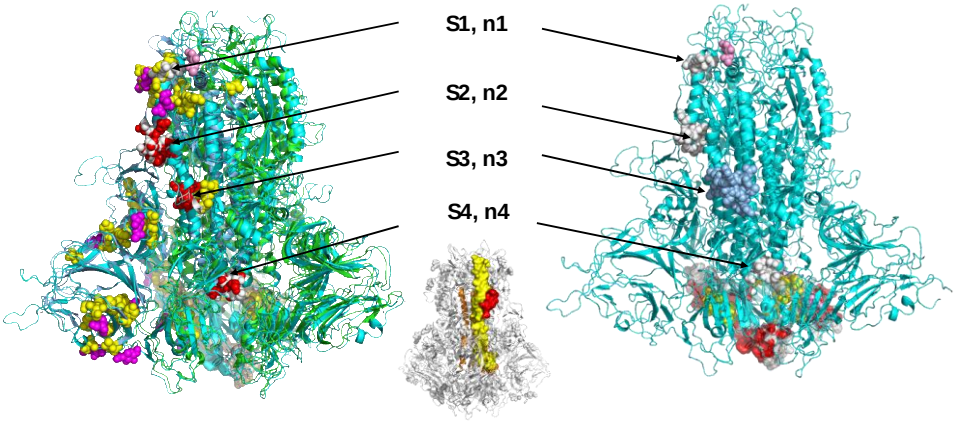
Zhou et al., Figure 3



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

Zhou et al., Figure 4

A|B



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2020 doi:10.20944/preprints202002.0381.v2

Zhou et al., Figure 4