

Altered N-/O-Glycosylation Sites On The Receptor-Binding Domain (RBD) in COVID-19 Are Related To Coronavirus Infection And Pathogenesis.

Shiwei Wang

Capital Medical University Affiliated Beijing Ditan Hospital

Fan Xiao

Capital Medical University Affiliated Beijing Ditan Hospital

Lingling He

Capital Medical University Affiliated Beijing Ditan Hospital

Junru Yang

Capital Medical University Affiliated Beijing Ditan Hospital

Hongshan Wei (✉ drwei@ccmu.edu.cn)

University Ditan Teaching Hospital <https://orcid.org/0000-0001-8893-653X>

Short report

Keywords: COVID-19, coronavirus, RBD, glycosylation

Posted Date: April 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-407527/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Altered N-/O-glycosylation sites on the receptor-binding domain(RBD) in COVID-19 are related to coronavirus infection and pathogenesis.

Shiwei Wang¹, Fan Xiao², Lingling He¹, Junru Yang¹, Hongshan Wei^{1,3,*}

¹ Department of Gastroenterology, Beijing Ditan Hospital, Capital Medical University, Beijing, China;

² Institute of Infectious Diseases, Beijing Ditan Hospital, Capital Medical University, Beijing, China;

³ Department of Gastroenterology, Peking University Ditan Teaching Hospital, Beijing, China;

*Corresponding authors:

Hongshan Wei, Ph. D.

Department of Gastroenterology, Beijing Ditan Hospital, Capital Medical University, No.8, Jingshun East Street, Chaoyang District, Beijing, 100015, China.

Email: drwei@ccmu.edu.cn

Abstract:

Objective: To discuss the possible significances of N-Linked and O-Linked glycosylation on the RBD of COVID-19 **Methods:** Amino acid sequences multiple alignments of RBD used Clustal Omega (v.1.2.4) using default parameters. Prediction of potential N-linked glycosylation sites by NetNGlyc 1.0 server. Prediction of potential O-linked glycosylation sites by NetOGlyc 4.0 Server. **Result:** COVID-19 Spike glycoprotein has 22 potential N-linked glycosylation sites and 3 O-linked glycosylation sites. 2 of 22 N-linked glycosylation sites distributed in RBD. None of the 3 O-linked glycosylation sites distributed in RBD, which is markedly different from SARS and other bat coronavirus using ACE2 as a receptor. Comparing with its close coronavirus, but which can't use ACE2 as a receptor, the COVID-19 has little N- and O-linked glycosylation sites. **Conclusion:** we show the obvious differences in glycosylation sites in RBD between COVID-19 and other coronaviruses. We speculate that altered N-/O-glycosylation sites on RBD in COVID-19 are related to its infection and pathogenesis.

Keywords: COVID-19, coronavirus, RBD, glycosylation,

Intrudction:

A novel coronavirus disease (COVID-19) created epidemics in December 2019 and subsequently developing into a global epidemic, threatening human health seriously[1, 2]. This is the third outbreak of coronavirus in just about 20 years. It's only been seven years since the last outbreak that the Middle East respiratory syndrome coronavirus (MERS-CoV) outbreak in the Arabian peninsula in 2012. First is the severe acute respiratory syndrome coronavirus (SARS-CoV), which outbreak in 2002 in China. Nevertheless, so far, there are no effective treatments and vaccines yet. So the study of coronavirus is urgently necessary to prevent a direct health threat to humans again.

Coronavirus spike (S) glycoprotein that is the trimeric transmembrane glycoprotein distributed around the viral envelope, mediates viral entry into host cells and affects its stability and infectivity. The four genera of coronavirus(Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus) all contain envelope-anchored spike glycoproteins. It is also the primary target of neutralizing antibodies and the vaccine design[3-5]. The S glycoprotein consists of three segments: a large ectodomain, a single-pass transmembrane anchor, and a short intracellular tail. The ectodomain consists of two basic functional subunits S1 and S2. The S1 domain drives receptor association, whereas the S2 domain mediates membrane fusion[6, 7]. The receptor-binding domain(RBD), located on the carboxy (C)-terminal domain (CTD) on S1, displays a concave surface during interaction with the receptor for SARS-CoV and MERS-CoV[8]. The entire receptor-binding loop, known as the receptor-binding motif (RBM), is located on the RBD and is responsible for complete contact with ACE2[9].

Protein glycosylation is one of the most abundant post-translational modifications, which affects almost every aspect of protein from synthesis to playing physiological function[10]. According to the amino acid atom where sugar moiety attached to, the glycosylation is classified as different types including N-, O- and C-linked glycosylation[11]. Which also extensively exists in mang virus including HIV-1, Lassa virus, Hepatitis C virus, and Epstein Barr virus[12-15]. It has long been known that glycosylation is critical for the coronavirus spike glycoprotein synthesis[16]. And even, the inhibition of glycans processing could prevent SARS infection[17]. In the infectious bronchitis virus(IBV), glycosylation is necessary for RBD interacting with appropriate receptors[18], and N-linked glycans at different positions may also differentially affect S protein-mediated fusion including guide ligand binding and viral infectivity and replication[19]. Sialic acid, a 9-carbon monosaccharide, on human HKU1 glycoproteins have significant effects on receptor determinants [20]. Interestingly, carbohydrate-carbohydrate interactions between virus and host have also been observed[16].

Besides, more importantly, significant effects of glycosylation on S glycoproteins also have been confirmed in MERS-Cov, SARS-Cov and other coronaviruses that infect humans, including human CoV NL63(HCoV-NL63) that generally causes human mild respiratory disease, human coronavirus 229E (HCoV-229E) belong to

Alphacoronavirus, and HCoV-OC43, HKU1 belong to Betacoronavirus[5]. Firstly, by cryoEM reconstructions, Walls, A.C. et al. found that N-linked glycans decorated the S trimer on the surface for both MERS CoV and SARS CoV[4]. Yanchen Zhou et al. have demonstrated that glycans were critical for SARS-CoV S protein which affects the interaction between DC-SIGN that is the alternative receptors for SARS-CoV and SARS-CoV [21]. HCoV-NL63 S trimer is covered by an extensive glycan shield consisting of 102 N-linked oligosaccharides obstructing the protein surface. Meanwhile, they put forward that coronavirus S glycans are masking the protein surface to limit access to neutralizing antibodies and thwart the humoral immune response[22]. Swatantra Kumar. et al analysis that some new glycosylation sites were introduced in the new coronavirus spike glycoproteins, which is related to receptors binding, transmission, and pathogenesis[23]. The difference of amino acid substitutions between the new CoVs and SARS CoV summarily reported in the article[24]. Some of those sites were involved in glycosylation.

In sum, glycosylation has critical roles for coronavirus. While the significance of glycosylation on S protein is still unclear. In this article, we analyze the number and location of N-linked glycosylation and O-linked glycosylation on S glycoprotein, especially the RBD that interact directly with corresponding receptors and affect virus entry and pathogenesis, and further discuss its possible influences on coronavirus.

Materials and methods:

Sequences used in the study:

The S and RBD amino acid sequence of all coronavirus strain collected from NCBI GenBank. COVID-19(YP_009724390), SARS coronavirus GZ02(AY390556), SARS coronavirus BJ01 (AY278488), Tor2(AY274119), SARS-like coronavirus WIV16(ALK02457), Bat SARS-like coronavirus WIV1(AGZ48828), Bat SARS-like coronavirus RsSHC014(AGZ48806), bat-SL-CoVZXC21(AVP78042), bat-SL-CoVZC45(AVP78031). More details were shown in Table 3.

Differences in glycosylation distribution:

Amino acid sequences multiple alignments of RBD used Clustal Omega (v.1.2.4) using default parameters. Prediction of potential N-linked glycosylation sites by NetNGlyc 1.0 server (<http://www.cbs.dtu.dk/services/NetNGlyc/>). Prediction of potential O-linked glycosylation sites(mucin type GalNAc O-glycosylation sites) by NetOGlyc 4.0 Server (<http://www.cbs.dtu.dk/services/NetOGlyc/>).

Sequence analysis of COVID-19

1, Position of potential N-and O-linked glycosylation sites in COVID-19 Spike glycoprotein.

COVID-19 Spike glycoprotein has 1273 amino acids with 22 potential N-linked glycosylation sites (Table 1) and 3 O-linked glycosylation sites(Table 2). According to S glycoprotein which had been annotated fully in Swiss Prot, N-linked glycosylation sites are almost all located on the extracellular domain of S glycoprotein. 2 of 22 distributed in RBD that are less than SARS and MERS(table 3). All three O-linked glycosylation sites are concentrated nearly(table 2), Which may locate around the polybasic cleavage site of the new coronavirus[25].

2, Variation in glycosylation sites in RBD among COVID-19, SARS-CoV and previous report Bat SARS-like coronavirus that uses human ACE2 as a receptor

To discuss the effect of glycosylation on receptor binding, the author analyzed the difference RBD amino acid sequence, which all use human ACE2 as a receptor. According to the previous investigation from China, we hold that human SARS-CoV GZ02, BJ01, and Tor2 represent the early, middle, and late phases of the 2002/2003 epidemic, respectively[26]. COVID-19 and SARS-CoV are all maybe come from bat [27]. So we collect some closely related Bat SARS-like coronavirus. Bat SARS-like coronavirus WIV16(SL-CoV- WIV16) represented the closest relatives to the epidemic SARS-CoV strains. SHC014 and WIV1 are also closely related. Meanwhile, they are all able to bind with human ACE2[28-30].

The most obvious feature of COVID-19 is that the N-linked glycosylation site at N370 and the O-linked glycosylation site at S349 are removed(figure 1). So that the COVID-19 glycosylation site number in RBD is less than SARS-CoV and bat SARS-like coronavirus. Especially that there isn't an O-linked glycosylation site on COVID-19 RBD. The N-linked glycosylation sites and motif (Asn-X-Ser/Thr, where X is any amino acid except for proline) among humans and bat are conserved(figure 1). The affinity binding ACE2 of COVID-19 is higher evidently(10 to 20 fold)than SARS-Cov[31]. Recently, another study also experimentally documented that[32]. This may be involved in a higher infection rate and widespread. Meanwhile, COVID-19 can use all ACE2 including humans, Chinese horseshoe bats, civet, pig, but mouse[27]. This may indicate that removing an N-linked glycan at N370 and an O-linked glycan at S349 is beneficial to the new coronavirus transmission. Whilst, this is necessary for efficient receptor binding. Recently, a result of another team also supports this[33]. More details were shown in table 3.

3, Variation in glycosylation sites in RBD among COVID-19, ZC45, and ZXC21.

To discuss the effect of glycosylation on cross-species transmission. The author analyzed the different RBD amino acid sequences, which was close, but use different receptors. The bat SL-CoV ZC45 and ZXC21, which founded in 2005 in Zhoushan, China, and from Chinese horseshoe bats(*Rhinolophus sinicus*)[34], was considered closely related to COVID-19 at the whole-genome level, but they can't interact with human ACE2[35]. The most remarkable difference in glycosylation between them is that the new coronavirus has less an N-linked glycan at N370 and an O-linked glycan at S349 than ZC45 and ZXC21(figure 2). This may indicate that glycosylation involved in cross-species transmission.

Discussion

Currently, quickly, COVID-19 has become a global public health crisis and is devastating spreading globally[36]. the novel CoVs has a greater capability for the human to human transmission, comparing with SARS and MERS[37]. This may be a potential cause of a large global pandemic. So far, the main measurements had adopted by preventing human-to-human transmission, because of the lack of effective medicine[38].

Coronavirus S glycoprotein is highly glycosylated by host-cell, generally containing 23-38 N-linked glycosylation sites on each spike trimer. The number of O-linked glycosylation sites on the spike trimer isn't fully clear. In this study, there are three potential O-linked glycosylation sites on the S glycoprotein of COVID-19. Although the length of the novel CoVs spike glycoprotein is longer than the bat SARS-like coronaviruses, SARS-CoV, and MERS-CoV[39], the number of glycosylation site is fewer. The current analysis shows that the N-linked glycosylation sites and motif on the novel CoVs RBD are conserved and located in the N-terminus. The result is consistent with a previous study[40]. However, it is striking that O-linked glycosylation sites can't be predicted on the novel coronavirus RBD, which may be an important mutation.

It's seemly that a higher density of N-linked glycans is beneficial for virus immune escape. N-linked glycans on spike glycoprotein help to shield certain epitopes[41]. The RBD of SARS and MERS are further divided into a core subdomain and a receptor-binding subdomain[42]. The N-linked glycans mainly located on the core subdomain[43]. In the NL63, from α genera coronavirus, the S2 subunits with higher N-linked glycan density than S1[22]. The difference shielding way on S glycoprotein may involve in different host immune[44]. Which all may be beneficial for escaping the host immune system. It's amazing that comparing to HIV, influenza with stronger evasion, coronavirus has less sparsity of the glycan shield[41]. Watanabe et al also suggested that there is a strong correlation between immune escape and glycan density[45].

The deletion of the N-linked glycosylation site(N370) on novel coronavirus RBD may help to induce neutralizing antibodies. Spike glycoprotein is the main target of coronavirus vaccines. Besides, RBD is also a great vaccine candidate, due to S glycoproteins major neutralizing epitopes locate on which[9]. The monoclonal antibodies of RBD could effectively against the virus entry[46]. Exploring the structure and distribution of glycans on the surface of S glycoprotein help to immunogen design and treatment measures[4]. Yuan, M. et al. supposed that the N-glycosylation site at residue N370 on COVID-19 results in the difference of antigenicity between COVID-19 and SARS-CoV[33]. Indeed, glycosylation may affect the antigenicity of RBD, which helpful for vaccine production[47]. This article also observed the differences in glycosylation on different coronavirus RBD.

N-linked glycosylation may be involved in S glycoprotein binding with host receptors. In the current analysis, we seemly find that a higher density of glycosylation is harmful to binding affinity. Visually, in Bat coronavirus, the glycans site generally at the apex of the spike trimer, oriented towards target cells[8]. So it's understandable that glycans on S or receptor sterically hinder binding between RBD and corresponding receptor[47, 48]. Similarly, glycans on human ACE2 also against the binding. The hot spot on human ACE2 regularly on the region without glycosylation, for easily accessible to viruses[49]. Proteins, in general, have advantages over sugars as viral receptors by providing higher affinity and specificity for viral attachment [56]. This makes us more convinced of the current opinion.

How did coronavirus cross the species barrier is still an intriguing puzzle.

Nevertheless understanding the mechanism of which is critical to combat outbreaks of coronavirus and develop new drugs. Tissue-specific glycosylation has a significant influence on viral transmission[50]. Two mutations that are S746R and N762A on HKU4 spike glycoprotein can let the HKU4 get the ability to enter into human cells. It is probably because that the mutations disrupted the N-linked glycosylation sites in the human protease motif in HKU4 [51]. The glycoprotein sites N227 and N699 on S protein, outside the RBD, from whether human or animal, are beneficial for coronavirus transmission[52]. In this article, we speculate that removing the glycosylation site on RBD benefit to coronavirus cross-species transmission.

The studies about ACE2 that is also the receptor of COVID-19 further support this hypothesis. If introducing the residues 90–93 of civet ACE2 into the human receptor, the binding capability of human ACE2 mediated by spike glycoproteins would be remarkably increased. This phenomenon probably because of removing of a glycosylation site at position 90 in human ACE2[7]. Another study also showed that rats are resistant to SARS-CoV because of introducing a glycosylation site at rat ACE2 position 82 residues, which interfere with the binding between SARS-Cov and rat ACE2[27]. Similar to the ACE2, dipeptidyl peptidase 4 (DPP4), MERS-CoV uses which as an entry receptor, is affected by glycosylation. Two N-linked glycans, N410 and N487, distributed on the core and RBM, respectively[28]. Disruption of the glycosylation sites in human DPP4 significantly regulates the affinity, which makes human DPP4 act as a receptor by MERS-CoV[29]. So they put forward that glycosylation is an important barrier relate to virus transmission. Aminopeptidase N (APN) serves as a receptor for HCoV-229E. While HCoV-229E can't interact with by substituting an N-glycosylation site at amino acids 291 to 293 in human APN. This means that glycosylation at amino acid 291 of human APN blocks infection by HCoV-229E[30].

The pattern of N-linked glycans on SARS-CoV S glycoprotein is mainly high-mannose(30%), hybrid(28%), and bi-, tri- and tetra-antennary complex glycans (42%)[17]. Meanwhile, 8 of the 22 sites on the novel coronavirus S protein are oligomannose-type glycans, and the remaining are complex-type glycans[53]. On the HKU1 S glycoproteins, the N-linked glycans are mainly high-mannose glycans[16]. Maybe, the different patterns of glycans also play an important role in the utilization of ACE2 proteins of different species by SARS-CoVs[54]. It is supposed that, in HIV, high-mannose type N-glycans may be more beneficial for antigen presentation and virus degradation[55]. The complex-type glycans are more involved in immunogen engineering in coronavirus[53]. And yet, the pattern of N-/O-linked glycans of other coronavirus remains unclear.

However, the researches about O-linked glycosylation in coronavirus are very little, and more details about O-glycosylation are still unclear. According to previous studies, the O-glycans on envelope virus are also important for receptor binding and entry[11]. It seems to be that the functions of O-linked glycans are similar to N-glycans in virus biology. We are surprised to find that there isn't a predicted O-linked glycosylation site on COVID-19 RBD. And the all three O-linked glycosylation sites (residues S673, T678, S686) on the novel CoVs S glycoprotein are

nearly around the S1/S2 boundary. Which may be involved in the polybasic cleavage site that is critical to viral infectivity and host range[25]. Our analysis suggests that removing O-linked glycans in RBD is beneficial to coronavirus transmission.

Conclusion

In conclusion, the number and location of glycosylation sites on RBD may influence the immune escape, infinity, host range, and pathogenesis of coronavirus, although without strong evidence currently. Much more experimental work is needed to do on it. The study about the effects of glycosylation on RBD will provide clues for further study of coronavirus. The functions of glycosylation are remaining huge challenges, due to the heterogeneity and complexity of glycosyltransferases and glycosylation sites[41]. This article analysis just focuses on primary structures. The function of glycans or glycosylation sites on secondary, and tertiary structures are also not yet fully understood. We need to note that another scenario is that the glycans don't have a direct affection while affecting the spike glycoprotein conformational stability. Fortunately, mass spectrometric, chromatographic, X-ray crystallography, cryo-electron microscopy(cryo-EM), and other techniques will be beneficial for studies of glycosylation.

Abbreviations: ACE2, angiotensin-converting enzyme 2; APN ,Aminopeptidase N ;RBD,receptor-binding domain; MERS-CoV Middle East respiratory syndrome coronavirus; SARS-CoV, severe acute respiratory syndrome coronavirus; S,Coronavirus spike;

Authors' contributions : Shiwei Wang, Fan Xiao: Conceptualization, Methodology, Data curation, Writing- Original draft preparation.Lingling He,Junru Yang: Visualization, Software, Validation.Hongshan wei:Writing- Reviewing and Editing.

Acknowledgements: “Not applicable”

Funding: This study was funded by The Digestive Medical Coordinated Development Center of Beijing Municipal Administration of Hospital(XXZ0404), Capital Foundation for Clinical Characteristic Applied Research Projects (Z181100001718084) and Beijing Natural Science Foundation (7202071).

Conflict of Interest: All authors declare that he/she has no conflict of interest.

Consent for publication: Written informed consent for publication was obtained from all participants.

Availability of data and materials: The authors confirm that the data supporting the findings of this study are available within the article.

Ethical approval: “Not applicable” .This article does not contain any studies with human participants or animals performed by any of the authors.

Reference

1. Wu, F., et al., A new coronavirus associated with human respiratory disease in China. *Nature*, 2020.
2. Sohrabi, C., et al., World Health Organization declares Global Emergency: A review of the 2019 Novel Coronavirus (COVID-19). *International Journal of Surgery*, 2020.
3. Tian, X., et al., Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerg Microbes Infect*, 2020. **9**(1): p. 382-385.
4. Walls, A.C., et al., Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. *Cell*, 2019. **176**(5): p. 1026-1039 e15.
5. Perlman, S. and J. Netland, Coronaviruses post-SARS: update on replication and pathogenesis. *Nature Reviews Microbiology*, 2009. **7**(6): p. 439-450.
6. Li, F., Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol*, 2016. **3**(1): p. 237-261.
7. Li, W., et al., Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J*, 2005. **24**(8): p. 1634-43.
8. Kirchdoerfer, R.N., et al., Pre-fusion structure of a human coronavirus spike protein. *Nature*, 2016. **531**(7592): p. 118-21.
9. Jiang, S., Y. He, and S. Liu, SARS vaccine development. *Emerg Infect Dis*, 2005. **11**(7): p. 1016-20.
10. Taherzadeh, G., et al., SPRINT-Gly: predicting N- and O-linked glycosylation sites of human and mouse proteins by using sequence and predicted structural properties. *Bioinformatics*, 2019. **35**(20): p. 4140-4146.
11. Bagdonaite, I. and H.H. Wandall, Global aspects of viral glycosylation. *Glycobiology*, 2018. **28**(7): p. 443-467.
12. Stewart-Jones, G.B., et al., Trimeric HIV-1-Env Structures Define Glycan Shields from Clades A, B, and G. *Cell*, 2016. **165**(4): p. 813-26.
13. Sommerstein, R., et al., Arenavirus Glycan Shield Promotes Neutralizing Antibody Evasion and Protracted Infection. *PLoS Pathog*, 2015. **11**(11): p. e1005276.
14. Falkowska, E., et al., Hepatitis C virus envelope glycoprotein E2 glycans modulate entry, CD81 binding, and neutralization. *J Virol*, 2007. **81**(15): p. 8072-9.
15. Szakonyi, G., et al., Structure of the Epstein-Barr virus major envelope glycoprotein. *Nat Struct Mol Biol*, 2006. **13**(11): p. 996-1001.
16. Chan, C.M., et al., Spike protein, S, of human coronavirus HKU1: role in viral life cycle and

- application in antibody detection. *Exp Biol Med* (Maywood), 2008. **233**(12): p. 1527-36.
17. Ritchie, G., et al., Identification of N-linked carbohydrates from severe acute respiratory syndrome (SARS) spike glycoprotein. *Virology*, 2010. **399**(2): p. 257-69.
18. Wickramasinghe, I.N., et al., Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity. *J Virol*, 2011. **85**(17): p. 8903-12.
19. Zheng, J., et al., Identification of N-linked glycosylation sites in the spike protein and their functional impact on the replication and infectivity of coronavirus infectious bronchitis virus in cell culture. *Virology*, 2018. **513**: p. 65-74.
20. Huang, X., et al., Human Coronavirus HKU1 Spike Protein Uses O-Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme. *J Virol*, 2015. **89**(14): p. 7202-13.
21. Zhou, Y., et al., A single asparagine-linked glycosylation site of the severe acute respiratory syndrome coronavirus spike glycoprotein facilitates inhibition by mannose-binding lectin through multiple mechanisms. *J Virol*, 2010. **84**(17): p. 8753-64.
22. Walls, A.C., et al., Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat Struct Mol Biol*, 2016. **23**(10): p. 899-905.
23. Kumar, S., et al., Structural, glycosylation and antigenic variation between 2019 novel coronavirus (2019-nCoV) and SARS coronavirus (SARS-CoV). *Virusdisease*, 2020. **31**(1): p. 13-21.
24. Wu, A., et al., Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe*, 2020.
25. Andersen, K.G., et al., The proximal origin of SARS-CoV-2. *Nat Med*, 2020. **26**(4): p. 450-452.
26. Chinese, S.M.E.C., Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science*, 2004. **303**(5664): p. 1666-9.
27. Zhou, P., et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 2020.
28. Yang, X.L., et al., Isolation and Characterization of a Novel Bat Coronavirus Closely Related to the Direct Progenitor of Severe Acute Respiratory Syndrome Coronavirus. *J Virol*, 2015. **90**(6): p. 3253-6.
29. Menachery, V.D., et al., A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat Med*, 2015. **21**(12): p. 1508-13.
30. Ge, X.-Y., et al., Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*, 2013. **503**(7477): p. 535-538.
31. Wrapp, D., et al., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 2020.
32. Tai, W., et al., Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol*, 2020.
33. Yuan, M., et al., A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science*, 2020.
34. Wong, A.C.P., et al., Global Epidemiology of Bat Coronaviruses. *Viruses*, 2019. **11**(2).
35. Ren, L.L., et al., Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. *Chin Med J (Engl)*, 2020.

36. Benvenuto, D., et al., The global spread of 2019-nCoV: a molecular evolutionary analysis. *Pathog Glob Health*, 2020: p. 1-4.
37. Ralph, R., et al., 2019-nCoV (Wuhan virus), a novel Coronavirus: human-to-human transmission, travel-related cases, and vaccine readiness. *J Infect Dev Ctries*, 2020. **14**(1): p. 3-17.
38. Lai, C.C., et al., Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int J Antimicrob Agents*, 2020. **55**(3): p. 105924.
39. Lu, R., et al., Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, 2020. **395**(10224): p. 565-574.
40. Parsons, L.M., et al., Glycosylation of the viral attachment protein of avian coronavirus is essential for host cell and receptor binding. *Journal of Biological Chemistry*, 2019. **294**(19): p. 7797-7809.
41. Watanabe, Y., et al., Exploitation of glycosylation in enveloped virus pathobiology. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 2019. **1863**(10): p. 1480-1497.
42. Wang, N., et al., Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res*, 2013. **23**(8): p. 986-93.
43. Li, F., Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *J Virol*, 2012. **86**(5): p. 2856-8.
44. Vankadari, N. and J.A. Wilce, Emerging WuHan (COVID-19) coronavirus: glycan shield and structure prediction of spike glycoprotein and its interaction with human CD26. *Emerg Microbes Infect*, 2020. **9**(1): p. 601-604.
45. Watanabe, Y., et al., Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. 2020.
46. Shanmugaraj, B., et al., Perspectives on monoclonal antibody therapy as potential therapeutic intervention for Coronavirus disease-19 (COVID-19). *Asian Pac J Allergy Immunol*, 2020. **38**(1): p. 10-18.
47. Li, F., et al., Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science*, 2005. **309**(5742): p. 1864-8.
48. Holmes, K.V., Structural biology. Adaptation of SARS coronavirus to humans. *Science*, 2005. **309**(5742): p. 1822-3.
49. Wu, K., et al., A virus-binding hot spot on human angiotensin-converting enzyme 2 is critical for binding of two different coronaviruses. *J Virol*, 2011. **85**(11): p. 5331-7.
50. Seymour, R.M., et al., Evolution of the human ABO polymorphism by two complementary selective pressures. *Proc Biol Sci*, 2004. **271**(1543): p. 1065-72.
51. Yang, Y., et al., Two Mutations Were Critical for Bat-to-Human Transmission of Middle East Respiratory Syndrome Coronavirus. *J Virol*, 2015. **89**(17): p. 9119-23.
52. Han, D.P., M. Lohani, and M.W. Cho, Specific asparagine-linked glycosylation sites are critical for DC-SIGN- and L-SIGN-mediated severe acute respiratory syndrome coronavirus entry. *J Virol*, 2007. **81**(21): p. 12029-39.
53. Watanabe, Y., et al., Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*, 2020.
54. Ren, W., et al., Difference in receptor usage between severe acute respiratory syndrome (SARS) coronavirus and SARS-like coronavirus of bat origin. *J Virol*, 2008. **82**(4): p.

1899-907.

55. van Montfort, T., et al., HIV-1 N-glycan composition governs a balance between dendritic cell-mediated viral transmission and antigen presentation. *J Immunol*, 2011. **187**(9): p. 4676-85.

Tables

Position	Sequence	Potential	July Agreement	N-Glyc result
17	NLTT	0.6606	(8/9)	+
61	NVTW	0.7820	(9/9)	+++
74	NGTK	0.7192	(9/9)	++
122	NATN	0.6781	(8/9)	+
149	NKSW	0.6318	(7/9)	+
165	NCTF	0.6220	(8/9)	+
234	NITR	0.7613	(9/9)	+++
282	NGTI	0.7378	(9/9)	++
331	NITN	0.5970	(7/9)	+
343	NATR	0.5671	(8/9)	+
603	NTSN	0.5783	(6/9)	+
616	NCTE	0.7163	(9/9)	++
657	NNSY	0.4724	(6/9)	-
709	NNSI	0.3528	(9/9)	--
717	NFTI	0.6426	(9/9)	++
801	NFSQ	0.6146	(8/9)	+
1074	NFTT	0.4084	(7/9)	-
1098	NGTH	0.5496	(5/9)	+
1134	NNTV	0.5800	(6/9)	+
1158	NHTS	0.3730	(9/9)	--
1173	NASV	0.3998	(8/9)	-
1194	NESL	0.6791	(9/9)	++

Table1: The potential N-linked glycosylation sites of COVID-19 S glycosylation. Prediction of potential N-linked glycosylation sites by NetNGlyc 1.0 server (<http://www.cbs.dtu.dk/services/NetNGlyc/>)

Position	score	comment
673	0.589454	POSITIVE
678	0.631365	POSITIVE
686	0.577497	POSITIVE

Table2:The potential O-linked glycosylation sites of COVID-19 S glycosylation. Prediction of potential O-linked glycosylation sites by NetOGlyc 4.0 Server(<http://www.cbs.dtu.dk/services/NetOGlyc/>)

	COVID-19	GZ02	BJ01	TOR2	WIV16	WIV1	ZC45	ZXC-21
Assession number	YP_009724390	AY390556	AY278488	AY274119	ALK02457	AGZ48828	AVP78031	AVP78042
Size aa (S)	1273	1255	1255	1255	1255	1256	1246	1245
N-glyc sites (S)	22	23	23	23	21	21	26	25
O-glyc sites (S)	3	1	2	2	3	2	3	2
Location(RBD)	331-524	317-569				318-570	327-560	326-559
Size aa(RBD)	194	253				253	234	234
N-glyc sites On RBD (position)	2 (331, 343)	3 (318, 330, 357)				3 (319, 331, 358)	3(327,339,366)	3(326,338,365)
N-glyc sites on RBD (position)	#	#	1(336)	1(336)	1(336)	2(337)	1(345)	2(344,445)

Table 3: potential N-/O-linked glycosylation sites on RBD among different coronavirus. S,spike glycoprotein; RBD, receptor-binding domain; #:There isn't O-linked glycosylation site on

Figures

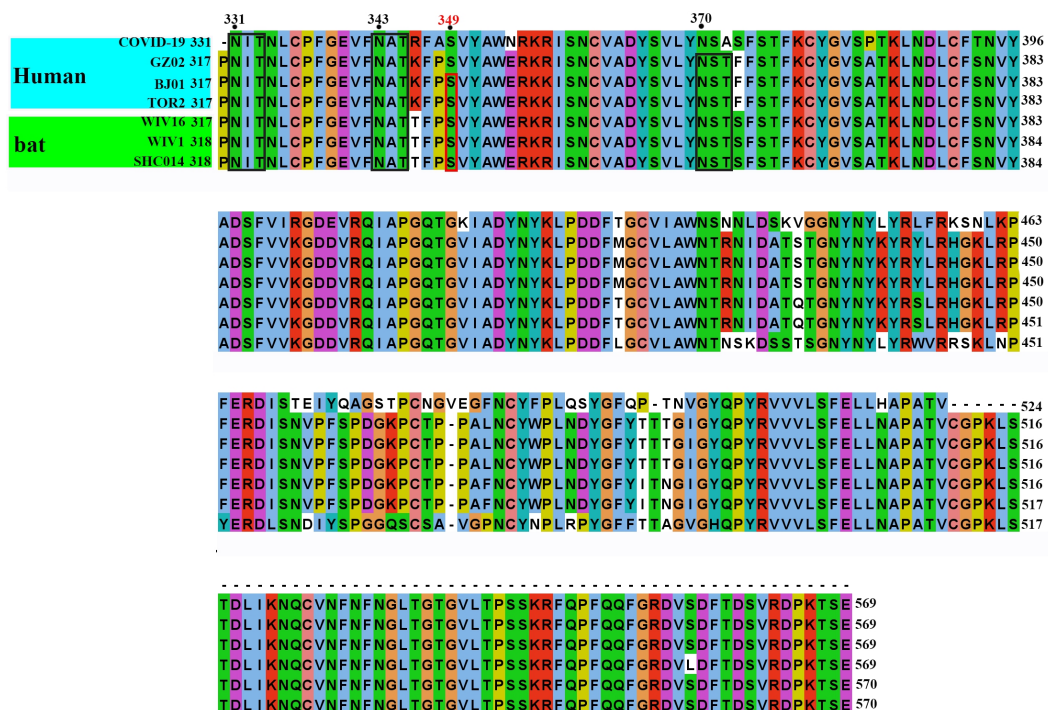


Figure 1: Amino acid sequence alignment of RBD among COVID-19, SARS-CoV and previous report Bat SARS-like coronavirus that uses human ACE2 as a receptor. The black squares and number are marked as N-Linked glycosylation motif (Asn-X-Ser/Thr, where X is any amino acid except for proline) and site respectively. The red squares and number are marked as O-Linked glycosylation site. Black spots represents residues that correspond to N/O-linked glycosylation sites. Outside the squares means the corresponding

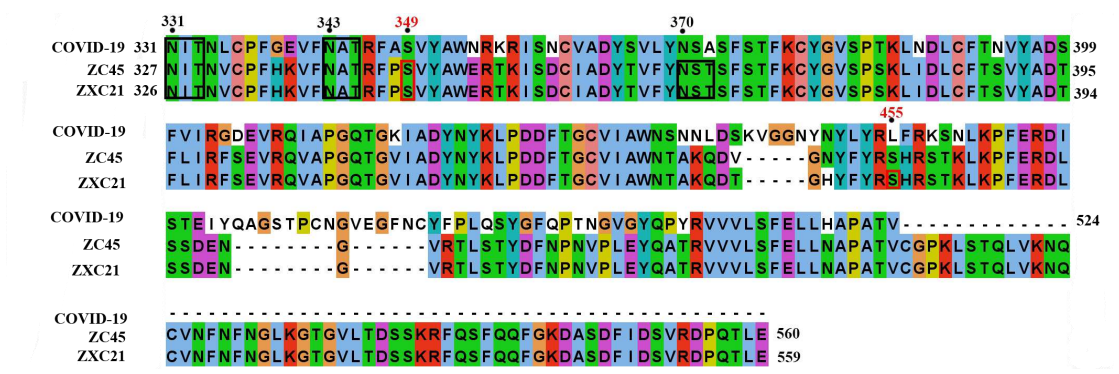


Figure 2, Alignment of RBD between SARS-CoV and related bat SARS-like coronavirus. The black squares and number are marked as N-Linked glycosylation motif (Asn-X-Ser/Thr, where X is any amino acid except for proline) and site respectively; The black numbers are marked as N-Linked sites. The red squares and number are marked as O-Linked glycosylation site. Black spots represent residues that correspond to N/O-linked glycosylation sites. Outside the squares means the corresponding glycosylation sites were

Figures

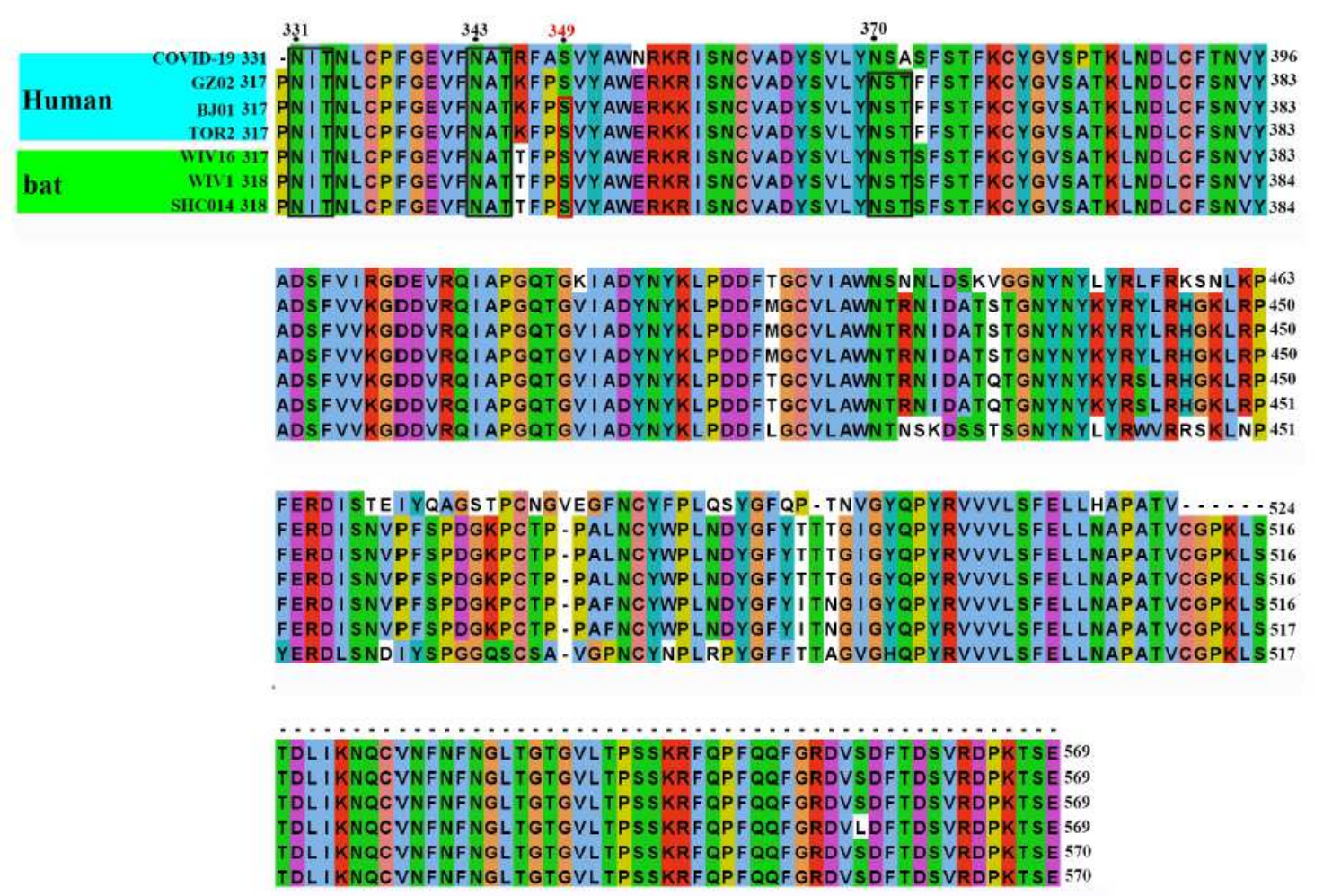


Figure 1

Amino acid sequence alignment of RBD among COVID-19, SARS-CoV and previous report Bat SARS-like coronavirus that uses human ACE2 as a receptor. The black squares and number are marked as N-Linked glycosylation motif (Asn-X-Ser/Thr, where X is any amino acid except for proline) and site respectively. The red squares and number are marked as O-Linked glycosylation site. Black spots represents residues that correspond to N-/O-linked glycosylation sites. Outside the squares means the corresponding

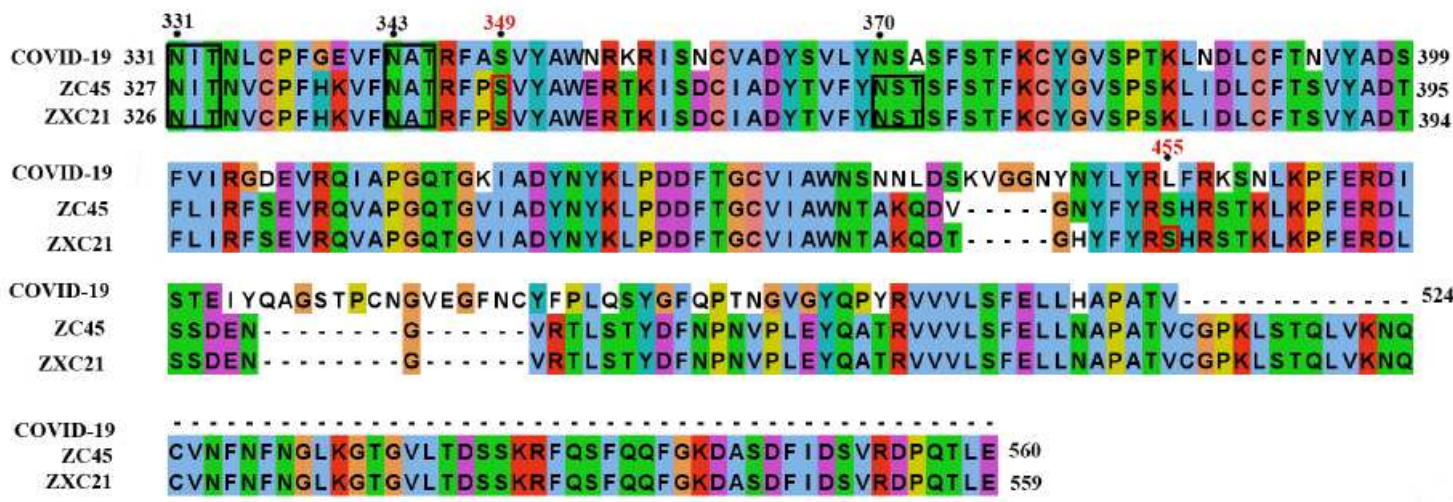


Figure 2

Alignment of RBD between SARS-CoV and related bat SARS-like coronavirus. The black squares and number are marked as N-Linked glycosylation motif (Asn-X-Ser/Thr, where X is any amino acid except for proline) and site respectively; The black numbers are marked as N-Linked sites. The red squares and number are marked as O-Linked glycosylation site. Black spots represents residues that correspond to N-/O-linked glycosylation sites. Outside the squares means the corresponding glycosylation sites were