Receptor recognition by novel coronavirus from Wuhan:

An analysis based on decade-long structural studies of SARS

Yushun Wan [1,*], Jian Shang [1,*], Rachel Graham [2], Ralph S. Baric [2], Fang Li [1,#]

[1] Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, Saint Paul, MN, USA

[2] Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA

* These authors contributed equally to this work.

Correspondence:

Fang Li (lifang@umn.edu)

Keywords: Wuhan coronavirus, SARS coronavirus, angiotensin-converting enzyme 2, animal reservoir, cross-species transmission, human-to-human transmission

Running title: Analyzing receptor usage by Wuhan coronavirus

Abstract

Recently a novel coronavirus (2019-nCoV) has emerged from Wuhan, China, causing symptoms in humans similar to those caused by SARS coronavirus (SARS-CoV). Since SARS-CoV outbreak in 2002, extensive structural analyses have revealed key atomic-level interactions between SARS-CoV spike protein receptor-binding domain (RBD) and its host receptor angiotensin-converting enzyme 2 (ACE2), which regulate both the cross-species and human-to-human transmissions of SARS-CoV. Here we analyzed the potential receptor usage by 2019-nCoV, based on the rich knowledge about SARS-CoV and the newly released sequence of 2019-nCoV. First, the sequence of 2019-nCoV RBD, including its receptor-binding motif (RBM) that directly contacts ACE2, is similar to that of SARS-CoV, strongly suggesting that 2019-nCoV uses ACE2 as its receptor. Second, several critical residues in 2019-nCoV RBM (particularly Gln493) provide favorable interactions with human ACE2, consistent with 2019-nCoV's capacity for human cell infection. Third, several other critical residues in 2019-nCoV RBM (particularly Asn501) are compatible with, but not ideal for, binding human ACE2, suggesting that 2019-nCoV has acquired some capacity for human-to-human transmission. Last, while phylogenetic analysis indicates a bat origin of 2019-nCoV, 2019-nCoV also potentially recognizes ACE2 from a diversity of animal species (except mice and rats), implicating these animal species as possible intermediate hosts or animal models for 2019-nCoV infections. These analyses provide insights into the receptor usage, cell entry, host cell infectivity and animal origin of 2019-nCoV, and may help epidemic surveillance and preventive measures against 2019-nCoV.

48 **Significance**

49 The recent emergence of Wuhan coronavirus (2019-nCoV) puts the world on
50 alert. 2019-nCoV is reminiscent of the SARS-CoV outbreak in 2002-2003. Our decade-
51 long structural studies on the receptor recognition by SARS-CoV have identified key
52 interactions between SARS-CoV spike protein and its host receptor angiotensin-
53 converting enzyme 2 (ACE2), which regulate both the cross-species and human-to-
54 human transmissions of SARS-CoV. One of the goals of SARS-CoV research was to
55 build an atomic-level iterative framework of virus-receptor interactions to facilitate
56 epidemic surveillance, predict species-specific receptor usage, and identify potential
57 animal hosts and animal models of viruses. Based on the sequence of 2019-nCoV spike
58 protein, we apply this predictive framework to provide novel insights into the receptor
59 usage and likely host range of 2019-nCoV. This study provides a robust test of this
60 reiterative framework, providing the basic, translational and public health research
61 communities with predictive insights that may help study and battle this novel 2019-
62 nCoV.
63

Introduction

A novel coronavirus (2019-nCoV) from Wuhan, China has recently caused over 500 confirmed cases of human infections and at least 17 deaths in China (https://www.cdc.gov/coronavirus/novel-coronavirus-2019.html). There are also numerous confirmed cases of 2019-nCoV infections in other countries including USA. Many of the symptoms caused by 2019-nCoV, such as acute respiratory syndrome, are similar to those caused by SARS coronavirus (SARS-CoV). SARS-CoV emerged in 2002-2003 and transmitted among humans, causing over 8000 confirmed cases of human infections and about 800 deaths (1-4). It briefly re-emerged in 2003-2004, with 4 confirmed cases of mild human infections and no human-to-human transmission (5-7). SARS-CoV has also been isolated from animals and been adapted to lab cell culture (5, 8-11). It is believed that bats and palm civets were the natural and intermediate reservoirs for SARS-CoV, respectively, and that SARS-CV transmitted from palm civets to humans in an animal market in Southern China (12-14). It has been reported that 2019-nCoV also infected humans in an animal market in Wuhan, although the animal source of the outbreak is currently unknown. Moreover, it has been confirmed that 2019-nCoV has the capacity to transmit from human to human.

Coronaviruses are a large family of single-stranded enveloped RNA viruses and can be divided into four major genera (15). Both SARS-CoV and 2019-nCoV belong to the β-genus. An envelope-anchored spike protein mediates coronavirus entry into host cells by first binding to a host receptor and then fusing viral and host membranes (16). A defined receptor-binding domain (RBD) of SARS-CoV spike specifically recognizes its

host receptor angiotensin-converting enzyme 2 (ACE2) (17, 18). Different lines of research have shown that which host is susceptible to SARS-CoV infection is primarily determined by the affinity between the viral RBD and host ACE2 in the initial viral attachment step (19-23). In a span of about 10 years, we determined a series of crystal structures of SARS-CoV RBD complexed with ACE2; the RBDs were from SARS-CoV strains isolated from different hosts species in different years and the ACE2 receptor orthologues were derived from different animal species (18, 24-26). These structures showed that SARS-CoV RBD contains a core structure and a receptor-binding motif (RBM), and that the RBM binds to the outer surface of the claw-like structure of ACE2 (Fig. 1A) (25). Importantly, we identified two virus-binding hotspots on human ACE2 (24, 26). A number of naturally selected RBM mutations occurred near these two virus-binding hotspot and these residues largely determined the host range of SARS-CoV (Fig. 1B, 1C). Furthermore, we discovered specific amino acids at 442, 472, 479, 480 and 487 positions that enhance viral binding to human ACE2, and some other amino acids at these same positions that enhance viral binding to civet ACE2 (Fig. 1C). Importantly, when all human-ACE2-favoring residues were combined into one RBD, this RBD binds to human ACE2 with super affinity and the corresponding spike protein mediates viral entry into human cells with super efficiency (Fig. 1C) (26). An RBD with super affinity for civet ACE2 was also designed and empirically confirmed (Fig. 1C) (26). These gain-of-function data provided strong supporting evidence for the accuracy of our structural predictions. A long-term goal of these earlier studies is to establish a structure-function predictive framework for improved epidemic surveillance. More specifically, we aim to predict the receptor usage and host cell infectivity of future SARS-CoV or SARS-like

viral strains and identify their possible animal origins and animal models, based on the sequences of their spike proteins and the known atomic structures of original SARS-CoV RBD/ACE2 complex. Here, based on the newly released sequence of 2019-nCoV RBD, we reiteratively apply this predictive framework to provide novel insights into the receptor usage and likely host range of 2019-nCoV.

Results

The 2019-nCoV spike phylogeny is firmly rooted among other β-genus lineage b bat SARS-like coronaviruses (Fig. 2), but is ancestral to both human SARS-CoV (epidemic strain isolated in year 2002) and bat SARS-CoV strains that use ACE2 receptor to enter and infect primary host lung cells (11, 17). The overall sequence similarities between 2019-nCoV spike and SARS-CoV spike (isolated from human, civet or bat) are around 76%-78% for the whole protein, around 73%-76% for the RBD, and 50%-53% for the RBM (Fig. 3A, 3B). In comparison, human coronavirus MERS coronavirus (MERS-CoV) and bat MERS-like coronavirus HKU4 share lower sequence similarities in their spikes, RBDs or RBM (Fig. 3C), and yet they recognize the same receptor dipeptidyl peptidase 4 (DPP4) (27, 28). Thus, sequence similarities between 2019-nCoV and SARS-CoV spikes suggest the possibility for them to share the same receptor ACE2. Importantly, compared to SARS-CoV RBM, 2019-nCoV RBM does not contain any deletion or insertion (except for a one-residue insertion on a loop away from the ACE2-binding region) (Fig. 3A), providing additional evidence that 2019-nCoV uses ACE2 as its receptor. Furthermore, among the 14 ACE2-contacting residues in the RBD, 9 are fully conserved and 4 are partially conserved among 2019-nCoV and SARS-CoV

from human, civet and bat (Fig. 3A). A final piece of strong evidence supporting ACE2 as the receptor for 2019-nCoV surrounds the five residues in 2019-nCoV RBM that underwent natural selections in SARS-CoV and played critical roles in the cross-species transmission of SARS-CoV (i.e., residue 442, 472, 479, 480, 487 in SARS-CoV RBD) (Fig. 1B). We discuss these residues in more details below.

First, residue 493 in 2019-nCoV RBD (corresponding to residue 479 in SARS-CoV) is a glutamine (Fig. 1B, 1D). A previously designed SARS-CoV RBD is optimal for binding to human ACE2 (Fig. 1B, 1C) (26). Residue 479 in SARS-CoV RBD is located near virus-binding hotspot Lys31 (i.e., hotspot-31) on human ACE2 (Fig. 1C). Hotspot-31 consists of a salt bridge between Lys31 and Glu35 buried in a hydrophobic environment. In civet SARS-CoV RBD (year 2002), residue 479 is a lysine, which imposes steric and electrostatic interference with hotspot-31. In human SARS-CoV RBD (year 2002), residue 479 becomes an asparagine. The K479N mutation removes the unfavorable interaction at the RBD/human ACE2 interface, enhances viral binding to human ACE2, and played a critical role in the civet-to-human transmission of SARS-CoV (Fig. 1C) (24-26). Importantly, Gln493 in 2019-nCoV RBD is compatible with hotspot-31, suggesting that 2019-nCoV is capable of recognizing human ACE2 and infecting human cells.

Second, residue 501 in 2019-nCoV RBD (corresponding to residue 487 in SARS-CoV) is an asparagine (Fig. 1B, 1D). Based on our previous structural analysis, residue 487 in SARS-CoV is located near virus-binding hotspot Lys353 (i.e., hotspot-353) on human ACE2 (Fig. 1C) (26). Hotspot-353 consists of a salt bridge between Lys353 and Asp38 also buried in a hydrophobic environment. In civet SARS-CoV RBD (year 2002),

155 residue 487 is a serine, which cannot provide favorable support for hotspot-353. In
156 human SARS-CoV isolated in year 2002, residue 487 is a threonine, which strengthens
157 the structural stability of hotspot-353. The S487T mutation adds the favorable interaction
158 at the RBD/human ACE2 interface, enhances viral binding to human ACE2, and played a
159 critical role in the human-to-human transmission of SARS-CoV (24-26). In human
160 SARS-CoV isolated in year 2003, residue 487 is a serine and there was no human-to-
161 human transmission for this SARS-CoV strain. Asn501 in 2019-nCoV RBD provides
162 more support to hotspot-353 than Ser487, but less than Thr487. This analysis suggests
163 that 2019-nCoV recognizes human ACE2 less efficiently than human SARS-CoV (year
164 2002), but more efficiently than human SARS-CoV (year 2003). Hence, at least when
165 considering the ACE2-RBD interactions, 2019-nCoV has gained some capability to
166 transmit from human and human.

167 Third, residues 455, 486 and 494 are leucine, phenylalanine and serine in 2019-
168 nCoV RBD, respectively (corresponding to residues 442, 472 and 480 in SARS-CoV,
169 respectively) (Fig. 1B, 1C, 1D). Based on our previous structural analysis, these three
170 residues in SARS-CoV RBD play significant roles, albeit not as dramatic as residues 479
171 and 487, in ACE2 binding (24-26). More specifically, Tyr442 of human and civet SARS-
172 CoV RBDs provides unfavorable interactions with hotspot-31 on human ACE2 (this
173 residue has been mutated to Phe442 in the optimized RBD); Leu455 of 2019-nCoV RBD
174 provides favorable interactions with hotspot-31, hence enhancing viral binding to human
175 ACE2. Leu472 of human and civet SARS-CoV RBDs provides favorable support for
176 hotspot-31 on human ACE2 through hydrophobic interactions with ACE2 residue Met82
177 and several other hydrophobic residues (this residue has been mutated to Phe472 in the

optimized RBD); Phe486 of 2019-nCoV RBD provides even more support for hotspot-31, hence also enhancing viral binding to human ACE2. Asp480 of human and civet SARS-CoV RBDs provides favorable support for hotspot-353 on human ACE2 through a neighboring tyrosine (this residue remains as an aspartate in the optimized RBD); Ser494 in 2019-nCoV RBD still provides positive support for hotspot-353, but the support is not as favorable as provided by Asp480. Overall, Leu455, Phe486 and Ser494 of 2019-nCoV RBD support that 2019-nCoV recognizes human ACE2 and infects human cells.

Last, having analyzed the interactions between 2019-nCoV RBD and human ACE2, how does 2019-nCoV RBD interact with putative ACE2 receptor orthologues from other animal species? Compared to human ACE2, both hotspot-31 and hotspot-353 on civet ACE2 have changed significantly (Fig. 4A). Specifically, residue 31 of civet ACE2 becomes a threonine, which can no longer form a salt bridge with Glu35; residue 38 of civet ACE2 becomes a glutamate, which forms a strong bifurcated salt bridge with Lys353 and no longer needs strong support from neighboring residues. A previously designed SARS-CoV RBD is optimal for binding to civet ACE2 (Fig. 1B, 4B) (26). In this designed RBD, Tyr442 forms a hydrogen bond with Thr31 of civet ACE2, and Arg479 forms a strong bifurcated salt bridge with Glu35 of civet ACE2. Moreover, in the designed RBD, Pro472 avoids unfavorable interactions with Thr82 of civet ACE2, and Gly480 does not provide unneeded support for hotspot-353. Furthermore, in the designed RBD, Thr487 provides limited but helpful support for hotspot-353. Here we constructed a structural model for the complex of 2019-nCoV RBD and civet ACE2 (Fig. 4C). Based on this model, Phe486 of 2019-nCoV RBD forms moderately unfavorable interaction with the polar side chain of Thr82 of civet ACE2, and Leu455 and Gln493 would lose

favorable interactions with civet ACE2 but they would be still compatible with civet ACE2. Thus, 2019-nCoV likely still uses civet ACE2 as its receptor, although it appears that 2019-nCoV RBD has not evolved adaptively for civet ACE2 binding. Moreover, 2019-nCoV likely does not use mouse or rat ACE2 as its receptor because mouse or rat ACE2 contains a histidine at the 353 position, which does not fit into the virus/receptor interact as well as a lysine does (Fig. 3A). 2019-nCoV RBD likely recognizes ACE2 from pigs, ferrets, cats, orangutans, monkeys and humans with similar efficiency, because these ACE2 molecules are identical or similar in the critical virus-binding residues. The situation involving bat ACE2 is complex because of the diversity of bat species (29). Based on the sequence of ACE2 from *Rhinolophus sinicus* bats (which can be recognized by bat SARS-CoV strain Rs3367), 2019-nCoV RBD likely also recognizes bat ACE2 as its receptor. Overall, 2019-nCoV likely recognizes ACE2 orthologues from a diversity of species, except for mouse and rat ACE2 (which should be poor receptors for 2019-nCoV).

Discussion

Atomic level resolution of complex virus-receptor interactions provides new opportunities for predictive biology. In this instance, we used prior knowledge gleamed from multiple SARS-CoV strains (isolated from different hosts in different years) and ACE2 receptors (from different animal species) to model predictions for novel 2019-nCoV. Our structural analyses confidently predict that 2019-nCoV uses ACE2 as its host receptor, consistent to two other new publications (30, 31). Compared to previously isolated SARS-CoV strains, 2019-nCoV likely uses human ACE2 less efficiently than

human SARS-CoV (year 2002), but more efficiently than human SARS-CoV (year 2003). Because ACE2-binding affinity has been shown to be one of the most important determinants of SARS-CoV infectivity, 2019-nCoV has evolved the capability to infect humans and some capability to transmit among humans. Alarmingly, our data predict that a single N501T mutation (corresponding to the S487T mutation in SARS-CoV) may significantly enhance the binding affinity between 2019-nCoV RBD and human ACE2. Thus, 2019-nCoV evolution in patients should be closely monitored for the emergency of novel mutations at the 501 position (to a lesser extent, also the 494 position).

What is the source of 2019-nCoV and did a key intermediate host play an important role in the current 2019-nCoV outbreak? Similar to SARS-CoV, 2019-nCoV most likely has originated from bats, given its close phylogenetic relationship with other β-genus lineage b bat SARS-CoV (Fig. 2). Moreover, 2019-nCoV likely recognizes ACE2 from a diversity of animal species, including palm civets, as its receptor. In the case of SARS-CoV, some of its critical RBM residues were adapted to human ACE2, while some others were adapted to civet ACE2 (26); this type of partial viral adaptations to two host species promoted virus replication and cross-species transmission between the two host species. In the case of 2019-nCoV, however, there is no strong evidence for adaptive mutations in its critical RBM residues that specifically promote viral binding to civet ACE2. Hence, either palm civets were not intermediate hosts for 2019-nCoV, or they passed 2019-nCoV to humans quickly before 2019-nCoV had any chance to adapt to civet ACE2. Like SARS-CoV, 2019-nCoV will likely replicate inefficiently in mice and rats, ruling them out as intermediate hosts for 2019-nCoV. Moreover, we predict that either 2019-nCoV or laboratory mice and rats would need to be genetically engineered

247 before a robust mouse or rat model for 2019-nCoV would become available. Pigs, ferrets,
248 cats and non-human primates contain largely favorable 2019-nCoV-contacting residues
249 in their ACE2, and hence may serve as animal models or intermediate hosts for 2019-
250 nCoV. It is worth noting that SARS-CoV was isolated in wild palm civets near Wuhan in
251 2005 (9), and its RBD had already been well adapted to civet ACE2 (except for residue
252 487). Thus, bats and other wild animals in and near Wuhan should be screened for both
253 SARS-CoV and 2019-nCoV.

254 These above analyses are based on the modeling of 2019-nCoV RBD/ACE2
255 interactions, heavily grounded in a series of atomic level structures of SARS-CoV
256 isolated from different hosts in different years (18, 24-26). There are certainly other
257 factors that affect the infectivity and pathogenesis of 2019-nCoV and will need to be
258 investigated. Nevertheless, our decade-long structural studies on SARS-CoV have firmly
259 shown that receptor recognition by SARS-CoV is one of the most important determinants
260 of its cross-species transmission and human-to-human transmission, a conclusion that has
261 been confirmed by different lines of research (13, 14). One of the long-term goals of our
262 previous structural studies on SARS-CoV was to build an atomic-level iterative
263 framework of virus-receptor interactions that facilitate epidemic surveillance, predict
264 species-specific receptor usage, and identify potential animal hosts and likely animal
265 models of human diseases. This study provides a robust test of this reiterative framework,
266 providing the basic, translational and public health research communities with predictive
267 insights that may help study and battle this novel 2019-nCoV.
268

Materials and Methods

Structural analysis. Software Coot was used for introducing mutations to structural models (32). Software PyMol was used for preparing structural figures (33).

Phylogenetic analysis. Consensus radial phylograms were generated in Geneious Prime (v.2020.0.3), with the Jukes-Cantor genetic distance model, the Neighbor-Joining build method, and no outgroup, with 100 bootstrap replicates. Phylograms were rendered for publication in Adobe Illustrator CC 2020.

Sequence alignment. Protein sequence alignments were done using Clustal Omega (34).

Acknowledgements

This work was supported by NIH grants R01AI089728 and R01AI110700 (to F.L. and R.S.B.). We would like to thank Shanghai Public Health Clinical Center and School of Public Health, Central Hospital of Wuhan, Huazhong University of Science and Technology, Wuhan Center for Disease Control and Prevention, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control, and University of Sydney Australia for releasing the sequence of 2019-nCoV genome.

References

1. **Lee N, Hui D, Wu A, Chan P, Cameron P, Joynt GM, Ahuja A, Yung MY, Leung CB, To KF, Lui SF, Szeto CC, Chung S, Sung JJY.** 2003. A major outbreak of severe acute respiratory syndrome in Hong Kong. New England Journal of Medicine **348:**1986-1994.

2. **Yu ITS, Li YG, Wong TW, Tam W, Chan AT, Lee JHW, Leung DYC, Ho T.** 2004. Evidence of airborne transmission of the severe acute respiratory syndrome virus. New England Journal of Medicine **350:**1731-1739.

3. **Marra MA, Jones SJM, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YSN, Khattra J, Asano JK, Barber SA, Chan SY, Cloutier A, Coughlin SM, Freeman D, Girn N, Griffith OL, Leach SR, Mayo M, McDonald H, Montgomery SB, Pandoh PK, Petrescu AS, Robertson AG, Schein JE, Siddiqui A, Smailus DE, Stott JE, Yang GS, Plummer F, Andonov A, Artsob H, Bastien N, Bernard K, Booth TF, Bowness D, Czub M, Drebot M, Fernando L, Flick R, Garbutt M, Gray M, Grolla A, Jones S, Feldmann H, Meyers A, Kabani A, Li Y, Normand S, Stroher U, Tipples GA, Tyler S, et al.** 2003. The genome sequence of the SARS-associated coronavirus. Science **300:**1399-1404.

4. **Peiris JSM, Lai ST, Poon LLM, Guan Y, Yam LYC, Lim W, Nicholls J, Yee WKS, Yan WW, Cheung MT, Cheng VCC, Chan KH, Tsang DNC, Yung RWH, Ng TK, Yuen KY.** 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet **361:**1319-1325.

5. **Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, Butt KM, Wong KL, Chan KW, Lim W, Shortridge KF, Yuen KY, Peiris JSM, Poon LLM.** 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in Southern China. Science **302:**276-278.

6. **Liang GD, Chen QX, Xu JG, Liu YF, Lim W, Peiris JSM, Anderson LJ, Ruan L, Li H, Kan B, Di B, Cheng P, Chan KH, Erdman DD, Gu SY, Yan XG, Liang WL, Zhou DH, Haynes L, Duan SM, Zhang X, Zheng H, Gao Y, Tong SX, Li DX, Fang L, Qin PZ, Xu WB.** 2004. Laboratory diagnosis of four recent sporadic cases of community-acquired SARS, Guangdong Province, China. Emerging Infectious Diseases **10:**1774-1781.

7. **Song HD, Tu CC, Zhang GW, Wang SY, Zheng K, Lei LC, Chen QX, Gao YW, Zhou HQ, Xiang H, Zheng HJ, Chern SWW, Cheng F, Pan CM, Xuan H, Chen SJ, Luo HM, Zhou DH, Liu YF, He JF, Qin PZ, Li LH, Ren YQ, Liang WJ, Yu YD, Anderson L, Wang M, Xu RH, Wu XW, Zheng HY, Chen**

JD, Liang GD, Gao Y, Liao M, Fang L, Jiang LY, Li H, Chen F, Di B, He LJ, Lin JY, Tong SX, Kong XG, Du L, Hao P, Tang H, Bernini A, Yu XJ, Spiga O, Guo ZM, et al. 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. Proceedings of the National Academy of Sciences of the United States of America **102:**2430-2435.

8. **Sheahan T, Rockx B, Donaldson E, Sims A, Pickles R, Corti D, Baric R.** 2008. Mechanisms of zoonotic severe acute respiratory syndrome coronavirus host range expansion in human airway epithelium. Journal of Virology **82:**2274-2285.

9. **Liu L, Fang Q, Deng F, Wang HZ, Yi CE, Ba L, Yu WJ, Lin RD, Li TS, Hu ZH, Ho DD, Zhang LQ, Chen ZW.** 2007. Natural mutations in the receptor binding domain of spike glycoprotein determine the reactivity of cross-neutralization between palm civet coronavirus and severe acute respiratory syndrome coronavirus. Journal of Virology **81:**4694-4700.

10. **Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, Li B, Xie JZ, Shen XR, Zhang YZ, Wang N, Luo DS, Zheng XS, Wang MN, Daszak P, Wang LF, Cui J, Shi ZL.** 2017. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS Pathog **13:**e1006698.

11. **Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, Mazet JK, Hu B, Zhang W, Peng C, Zhang YJ, Luo CM, Tan B, Wang N, Zhu Y, Crameri G, Zhang SY, Wang LF, Daszak P, Shi ZL.** 2013. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature **503:**535-538.

12. **Cui J, Li F, Shi ZL.** 2019. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol **17:**181-192.

13. **Li F.** 2013. Receptor recognition and cross-species infections of SARS coronavirus. Antiviral Res **100:**246-254.

14. **Li WH, Wong SK, Li F, Kuhn JH, Huang IC, Choe H, Farzan M.** 2006. Animal origins of the severe acute respiratory syndrome coronavirus: Insight from ACE2-S-protein interactions. Journal of Virology **80:**4211-4219.

15. **Perlman S, Netland J.** 2009. Coronaviruses post-SARS: update on replication and pathogenesis. Nature Reviews Microbiology **7:**439-450.

16. **Li F.** 2016. Structure, Function, and Evolution of Coronavirus Spike Proteins. Annu Rev Virol **3:**237-261.

17. **Li WH, Moore MJ, Vasilieva N, Sui JH, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC, Choe H, Farzan M.** 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature **426:**450-454.

18. **Li F.** 2015. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. J Virol **89:**1954-1964.

19. **Li WH, Greenough TC, Moore MJ, Vasilieva N, Somasundaran M, Sullivan JL, Farzan M, Choe H.** 2004. Efficient replication of severe acute respiratory syndrome coronavirus in mouse cells is limited by murine angiotensin-converting enzyme 2. Journal of Virology **78:**11429-11433.

20. **Li WH, Zhang CS, Sui JH, Kuhn JH, Moore MJ, Luo SW, Wong SK, Huang IC, Xu KM, Vasilieva N, Murakami A, He YQ, Marasco WA, Guan Y, Choe HY, Farzan M.** 2005. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. Embo Journal **24:**1634-1643.

21. **McCray PB, Pewe L, Wohlford-Lenane C, Hickey M, Manzel L, Shi L, Netland J, Jia HP, Halabi C, Sigmund CD, Meyerholz DK, Kirby P, Look DC, Perlman S.** 2007. Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus. Journal of Virology **81:**813-821.

22. **Moore MJ, Dorfman T, Li WH, Wong SK, Li YH, Kuhn JH, Coderre J, Vasilieva N, Han ZC, Greenough TC, Farzan M, Choe H.** 2004. Retroviruses pseudotyped with the severe acute respiratory syndrome coronavirus spike protein efficiently infect cells expressing angiotensin-converting enzyme 2. Journal of Virology **78:**10628-10635.

23. **Qu XX, Hao P, Song XJ, Jiang SM, Liu YX, Wang PG, Rao X, Song HD, Wang SY, Zuo Y, Zheng AH, Luo M, Wang HL, Deng F, Wang HZ, Hu ZH, Ding MX, Zhao GP, Deng HK.** 2005. Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. Journal of Biological Chemistry **280:**29588-29595.

24. **Li F.** 2008. Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections. Journal of Virology **82:**6984-6991.

25. **Li F, Li WH, Farzan M, Harrison SC.** 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science **309:**1864-1868.

26. **Wu KL, Peng GQ, Wilken M, Geraghty RJ, Li F.** 2012. Mechanisms of Host Receptor Adaptation by Severe Acute Respiratory Syndrome Coronavirus. Journal of Biological Chemistry **287:**8904-8911.

27. **Yang Y, Du L, Liu C, Wang L, Ma C, Tang J, Baric RS, Jiang S, Li F.** 2014. Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus. Proc Natl Acad Sci U S A **111:**12516-12521.

28. **Raj VS, Mou HH, Smits SL, Dekkers DHW, Muller MA, Dijkman R, Muth D, Demmers JAA, Zaki A, Fouchier RAM, Thiel V, Drosten C, Rottier PJM, Osterhaus A, Bosch BJ, Haagmans BL.** 2013. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature **495:**251-254.

29. **Hou YX, Peng C, Yu M, Li Y, Han ZG, Li F, Wang LF, Shi ZL.** 2010. Angiotensin-converting enzyme 2 (ACE2) proteins of different bat species confer variable susceptibility to SARS-CoV entry. Archives of Virology **155:**1563-1569.

30. **Michael C Letko VM.** 2020. Functional assessment of cell entry and receptor usage for lineage B β -coronaviruses, including 2019-nCoV. wwwbiorxivorg.

31. **Zheng-Li Shi PZ, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, Hui-Dong Chen, Jing Chen, Yun Luo, Hua Guo, Ren-Di Jiang, Mei-Qin Liu, Ying Chen, Xu-Rui Shen, Xi Wang, Xiao-Shuang Zheng, Kai Zhao, Quan-Jiao Chen, Fei Deng, Lin-Lin Liu, Bing Yan, Fa-Xian Zhan, Yan-Yi Wang, Gengfu Xiao.** 2020. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. wwwbiorxivorg.

32. **Emsley P, Cowtan K.** 2004. Coot: model-building tools for molecular graphics. Acta Crystallographica Section D-Biological Crystallography **60:**2126-2132.

33. **Anonymous.** The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.

34. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG.** 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol **7:**539.

Figure legends:

Figure 1: Structural analysis of human ACE2 recognition by 2019-nCoV and SARS-CoV. (A) Overall structure of human SARS-CoV RBD (year 2002) complexed with human ACE2. PDB ID is 2AJF. ACE2 is in green, the core of RBD (receptor-binding domain) is in cyan, and RBM (receptor-binding motif) is in magenta. (B) Critical residue changes in the RBMs of SARS-CoV and 2019-nCoV. All these five residues in SARS-CoV underwent natural selections and were shown to be critical for ACE2 recognition, cell entry, and host range of SARS-CoV. The residue numbers are shown as in SARS-CoV RBD, with the corresponding residue numbers in 2019-nCoV shown in parentheses. For viral adaption to ACE2, > means “is more adapted” and = means “is similarly adapted”. (C) Experimentally determined structure of the interface between a designed SARS-CoV RBD (optimized for human ACE2 recognition) and human ACE2. PDB ID is 3SCI. (D) Modeled structure of the interface between 2019-nCoV RBD and human ACE2. Here mutations were introduced to the RBD region in panel (C) based on sequence differences between SARS-CoV and 2019-nCoV. GenBank accession numbers are: MN908947.1 for 2019-nCoV Spike; NC_004718.3 for human SARS -CoV Spike (year 2002; strain Tor2); AGZ48818.1 for bat SARS-CoV Spike (year 2013; strain Rs3367); AY304486.1 for civet SARS-CoV spike (year 2002; SZ3); AY525636 for human/civet SARS-CoV spike (year 2003; strain GD03). References for the other sequences are: civet SARS-CoV spike (year 2005) (9); human SARS-CoV spike (year 2008) (8).

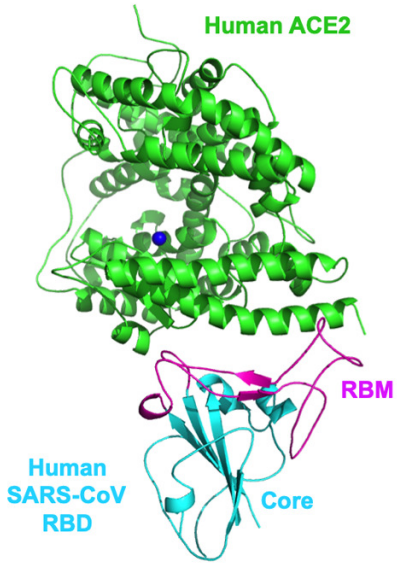
Figure 2. Spike phylogeny of representative β-genus lineage b coronaviruses. The Spike protein sequences of selected β-genus lineage b coronaviruses were aligned and phylogenetically compared. Sequences were aligned using free end gaps with the Blosum62 cost matrix in Geneious Prime. The tree was constructed using the neighbor-joining method based on the multiple sequence alignment, also in Geneious Prime. Numbers following the underscores in each sequence correspond to the GenBank accession number. The radial phylogram was exported from Geneious and then rendered for publication using EvolView (evolgenius.info) and Adobe Illustrator CC 2020.

Figure 3: Sequence comparison of 2019-nCoV and SARS-CoV. (A) Sequence alignment of SARS-CoV and 2019-nCoV RBDs. RBM residues are in magenta. The five critical residues in Fig. 1B are in blue. ACE2-contacting residues are shaded. Asterisks indicate positions that have a single, fully conserved residue. Colons indicate positions that have strongly conserved residues. Periods indicate positions that have weakly conserved residues. (B) Sequence similarities of SARS-CoV and 2019-nCoV in the spike protein, RBD and RBM, respectively. (C) Sequence similarities of MERS-CoV and HKU4 virus in the spike protein, RBD and RBM, respectively. GenBank accession numbers are: JX869059.2 for human MERS-CoV Spike; NC_009019.1 for bat HKU4-CoV Spike.

Figure 4: Structural analysis of animal ACE2 recognition by 2019-nCoV and SARS-CoV. (A) Critical changes in virus-contacting residues of ACE2 from different host species. GenBank accession numbers for ACE2 are as follows: NM_001371415.1

(human), AAX63775.1 (civet), KC881004.1 (bat), NP_001123985.1 (mouse), AY881244 (rat), NP_001116542.1 (pig), AB208708 (ferret), NM_001039456 (cat), Q5RFN1 (orangutan), and AY996037 (monkey). (B) Experimentally determined structure of the interface between a designed SARS-CoV RBD (optimized for civet ACE2 recognition) and civet ACE2. PDB ID is 3SCK. (C) Modeled structure of the interface between 2019-nCoV RBD and civet ACE2. Here mutations were introduced to the RBD region in panel (B) based on sequence differences between SARS-CoV and 2019-nCoV.

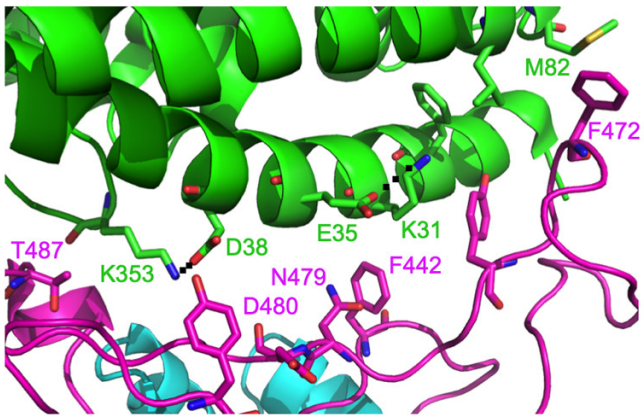
A



B

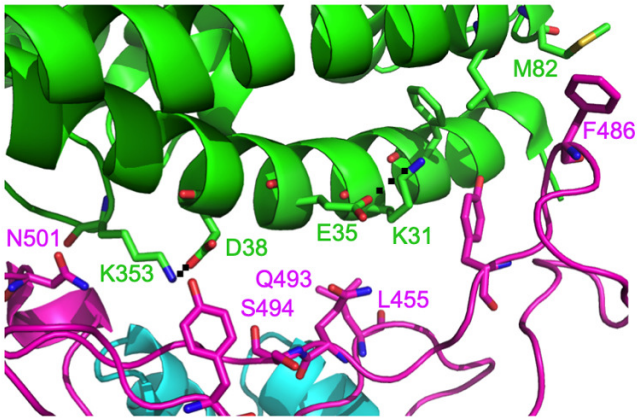
Virus	Year	442	472	479	480	487
SARS - human	2002	Y	L	**N**	D	**T**
SARS - civet	2002	Y	L	**K**	D	**S**
SARS - human/civet	2003	Y	P	**N**	G	**S**
SARS - civet	2005	Y	P	**R**	G	**S**
SARS - human	2008	F	F	**N**	D	**S**
Viral adaption to human ACE2		F > Y	F > L > P	**N = R >>> K**	D > G	**T >>> S**
Optimized - human	In vitro design	F	F	**N**	D	**T**
Viral adaptation to civet ACE2		Y > F	P = L > F	**R > K = N**	G > D	**T > S**
Optimized - civet	In vitro design	Y	P	**R**	G	**T**
SARS - bat	2013	S	F	**N**	D	**N**
2019-nCoV – human	2019	L (455)	F (486)	**Q (493)**	S (494)	**N (501)**

C



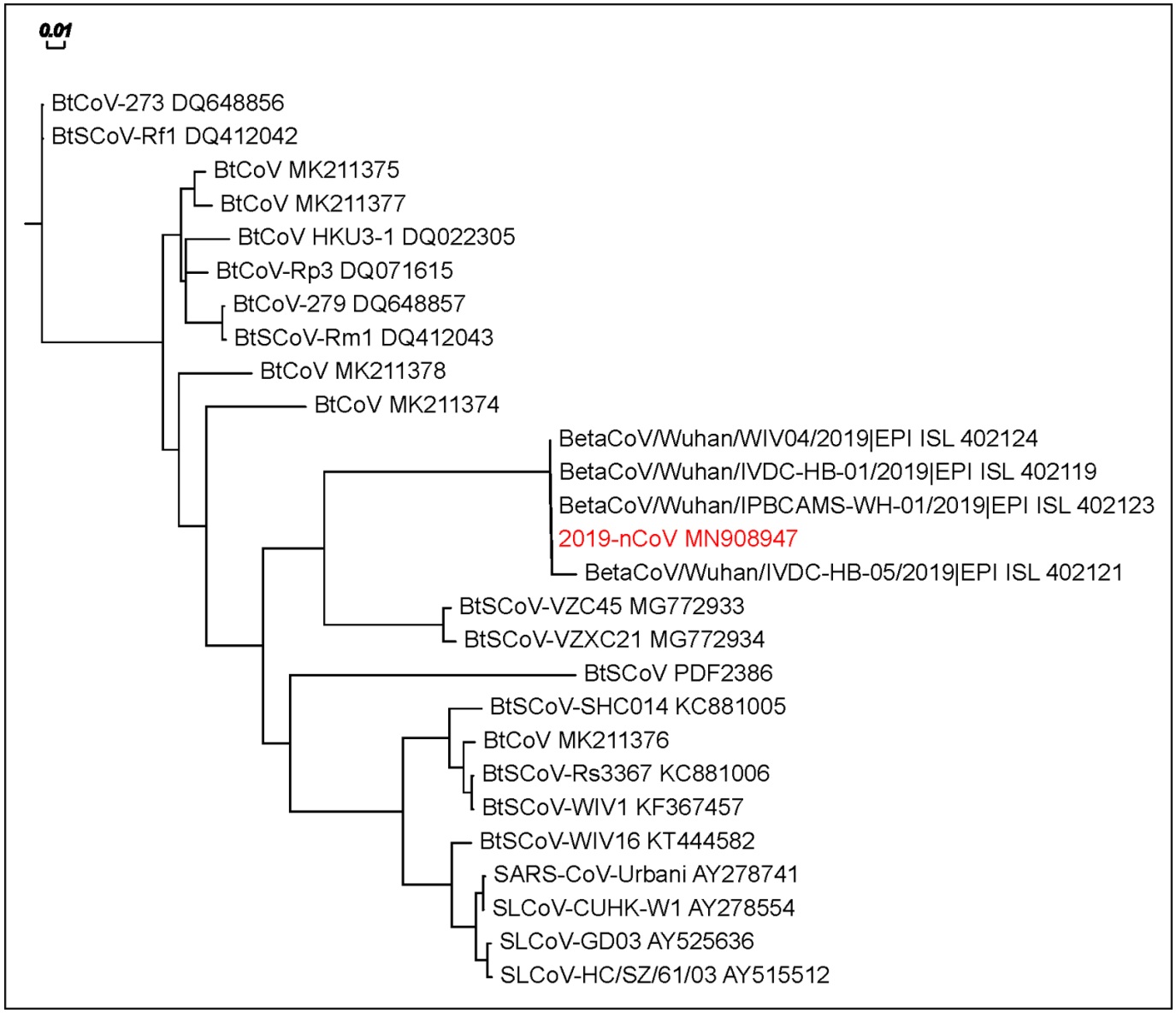
Human SARS-CoV-optimized RBD

D



Human 2019-nCoV RBD (model)

Journal of Virology



A

```
Human-SARS-2002    306 RVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF 360
Civet-SARS-2002    319 RVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKRISNCVA DYSVLYNSTS 373
Bat-SARS-2013      319 RVAPS KEVVRFPNIT NLCPFGEVFN ATTFPSVYAW ERKRISNCVA DYSVLYNSTS 373
2019-nCoV          319 RVQPT ESIVRFPNIT NLCPFGEVFN ATRFASVYAW NRKRISNCVA DYSVLYNSAS 373
                       ** *:  .:******* ********** ** * ***** :**:****** ********:

Human-SARS-2002  FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
Civet-SARS-2002  FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 433
Bat-SARS-2013    FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFTGCV 433
2019-nCoV        FSTFKCYGVS PTKLNDLCFT NVYADSFVIR GDEVRQIAPG QTGKIADYNY KLPDDFTGCV 433
                 **********  ********: ********:: **:******* *** ****** ****** ***

Human-SARS-2002  LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTP-P ALNCYWPLND 480
Civet-SARS-2002  LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTP-P ALNCYWPLKD 493
Bat-SARS-2013    LAWNTRNIDA TQTGNYNYKY RSLRHGKLRP FERDISNVPF SPDGKPCTP-P AFNCYWPLND 493
2019-nCoV        IAWNSNNLDS KVGGNYNYLY RLFRKSNLKP FERDISTEIY QAGSTPCNGVE GFNCYFPLQS 494
                 :***:.*:*: .  ***** * * :*:.:*:* ******.  : . ...**.    .:***:**:.

Human-SARS-2002  YGFYTTTGIG YQPYRVVVLS FELLNAPATV CGPKL 515
Civet-SARS-2002  YGFYTTSGIG YQPYRVVVLS FELLNAPATV CGPKL 528
Bat-SARS-2013    YGFYITNGIG YQPYRVVVLS FELLNAPATV CGPKL 528
2019-nCoV        YGFQPTNGVG YQPYRVVVLS FELLHAPATV CGPKK 529
                 ***  *.*:* ********** ****:***** ****
```

B

Spike / RBD / RBM	SARS-human	SARS-civet	SARS-bat	2019-nCoV
SARS-human	100% / 100% / 100%			
SARS-civet	98.12% / 98.10% / 97.18%	100% / 100% / 100%		
SARS-bat	92.33% / 94.29% / 92.96%	92.75% / 94.76% / 91.55%	100% / 100% / 100%	
2019-nCoV	76.04% / 73.33% / 50.00%	76.78% / 74.29% / 50.00%	77.50% / 75.71% / 52.78%	100% / 100% / 100%

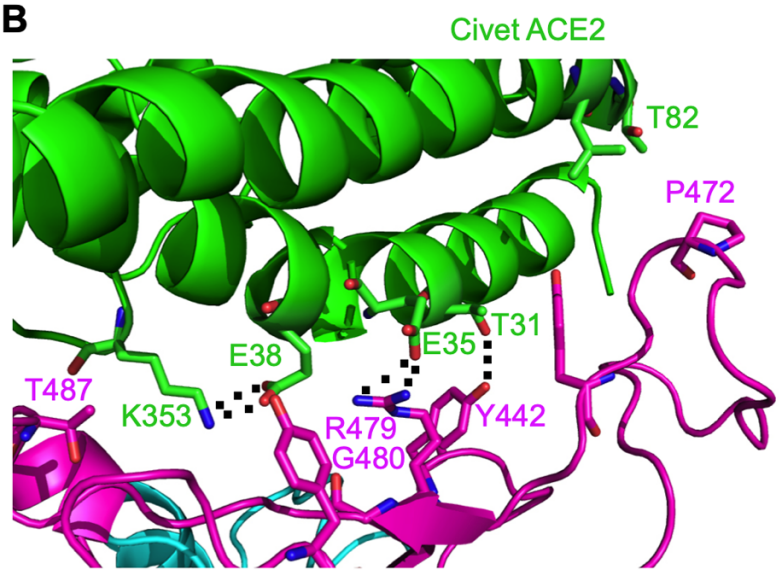
C

Spike /RBD /RBM	MERS-human
HKU4-bat	67.04% /57.69% /40.79%

A

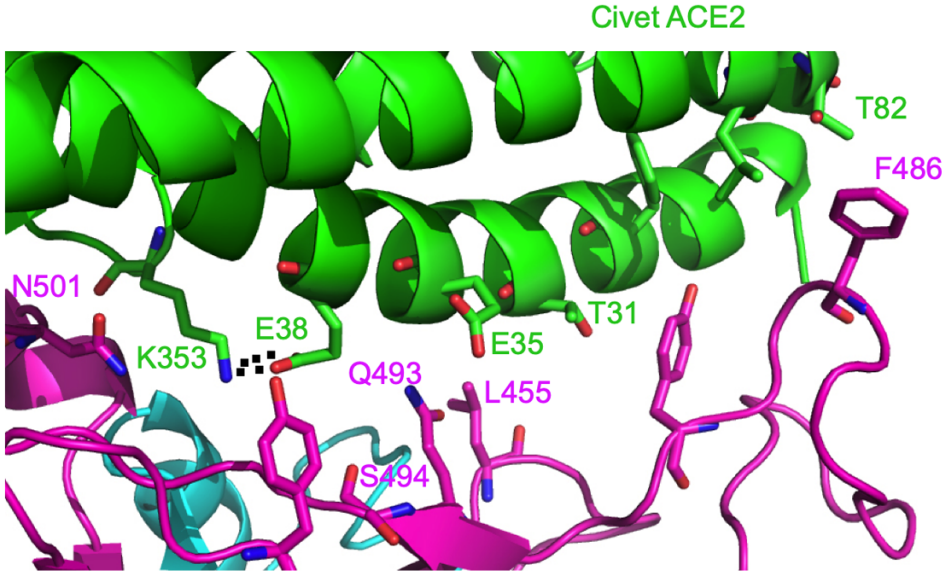
ACE2	31	35	38	82	353
Human	K	E	D	M	K
Civet	T	E	E	T	K
Bat	K	K	D	N	K
Mouse	N	E	D	S	H
Rat	K	E	D	N	H
Pig	K	E	D	T	K
Ferret	K	E	E	T	K
Cat	K	E	E	T	K
Orangutan	K	E	D	M	K
Monkey	K	E	D	M	K

B



Civet SARS-CoV-optimized RBD

C



Human 2019-nCoV RBD (model)