IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

COMPARATIVE GENOMICS OF RECEPTOR BINDING DOMAINS OF S PROTEINS AND HOST RECEPTOR INTERACTION IN COVID-19 PATIENT

[1]Rimjhim Dasgupta
Founder & Director, 4NBIO, Mumbai, India

The outbreak of viral pneumonia in the city of Wuhan, China, in Dec, 2019 was caused by a novel coronavirus designated 2019-nCoV, as determined by sequencing the viral RNA genome. It was thought that SARS-related coronaviruses (SARSr-CoV) is mainly found in bats. Previous studies have shown that some bat SARSr-CoVs have the potential to infect humans. Among its genome S protein is surface-exposed and mediates entry into host cells. Currently it is one of the main targets for designing antibodies (Abs), therapeutic and vaccine. Earlier studies stated that ACE2 (angiotensin converting enzyme 2) could facilitate S protein mediated entry for this newly emerged coronavirus. Here we have taken an attempt to compare the genetic structure of receptor binding domain within S protein of highly pathogenic human coronaviruses (special reference to 2019-nCoV) with Bat coronavirus RaTG13. We have compared 2019-nCov receptor binding domain (RBD) with other pathogenic human coronaviruses (MERS and SARSr-CoV) and Bat coronavirus RaTG13. We found that it is closest to RaTG13 RBD than MERS and SARSr-CoV. Our study shows that 2019-nCov RBD also has significant identity with pangolin S protein RBD. We have also predicted the amino acid residues within RDB those may play important role for ACE2 receptor interaction. We identified unique signature for furin cleavage in 2019-nCov S protein but not in of other pathogenic human coronaviruses (tested here), bat coronavirus RaTG13 or pangolin.

Keywords - Coronavirus, sequence alignment, S protein, receptor binding domain, RBD, Covid-19

I. INTRODUCTION

The Corona Virus Disease 2019 (COVID-19) caused by a novel coronavirus (CoV) named ‘‘2019 novel coronavirus’’ or ‘‘2019-nCoV’’ is responsible for the recent pneumonia outbreak that started in early December, 2019 in Wuhan City, China.

Coronaviruses mainly cause respiratory and gastrointestinal tract infections and are genetically classified into four major genera: Alphacoronavirus, Beta coronavirus, Gamma coronavirus, and Delta coronavirus. The former two genera primarily infect mammals, whereas the latter two predominantly infect birds. Three highly pathogenic human coronaviruses (CoVs) have been identified so far, including Middle East respiratory syndrome coronavirus (MERS-CoV), severe acute respiratory syndrome (SARS) coronavirus (SARS-CoV), 2019-nCoV [1, 2, 3]. A large number of studies have proved that the pathogen COVID-19 is a novel coronavirus, which belongs to the *Coronavirus* family, *Betacoronavirus* genus and *Sarbecovirus* subgenus, with a linear single-stranded positive-strand RNA genome of about 30 kb [4, 5, 6]. Coronavirus entry into host cells is mediated by the transmembrane spike (S) protein that forms homotrimer protruding from the viral surface [2]. S comprises two functional subunits N-terminal S1 subunit and a membrane-embedded C-terminal S2 region [7]. S1 specializes in recognizing host-cell receptors and is normally more variable in sequence among different CoVs than the S2 region [8, 9]. For many CoVs, S is cleaved at the boundary between the S1 and S2 subunits, which remain non-covalently bound in the prefusion conformation [10, 11, 12, 13, 14]. The distal S1 subunit comprises the receptor-binding domain and contributes to stabilization of the prefusion state of the membrane-anchored S2 subunit that contains the fusion machinery [6, 13, 15, 16, 17, 18, 19].

Two discrete domains that can fold independently are located in the S1 N- and C-terminal portions, both of which can be used for receptor engagement [20]. The N-terminal domain (NTD), functioning as the entity involved in receptor recognition. In the S1 subunit, the receptor binding domain (RBD, also called the C terminal domain, CTD) is localized in the C-terminal region, spanning 200 amino acids, and structural studies have revealed that the RBD consists of two subdomains: the core and external subdomains [21, 22, 23, 24]. In the S2 subunit, the heptad repeat (HR) regions are also well characterized [25, 26, 27], and as expected, the HR1 and HR2 of MERS-CoV fold into an intra-hairpin helical structure that can assemble trimerically into a six-helix bundle (a trimer of theHR1/HR2 heterodimer), demonstrating a classical type I membrane fusion process (10). Peptide inhibitors have been designed targeting these HR regions and been proven to be effective in vitro and in vivo [25, 28, 29, 30, 31]. These studies have provided insight about the characteristics of overall S protein structures. We have taken an attempt to further analyse the RBD domain and compare 2019-nCov RBD with two highly pathogenic human coronaviruses (MERS and SARSr-CoV) and RaTG13. We have tried to predict the residues in RBD those are engaged in host receptor interaction. This may enhance our understanding of S protein function and subsequent design of broadly neutralizing antibodies and vaccine.

I. RESEARCH METHODOLOGY

Wuhan isolate, SARS-CoV-2 sequence NC_045512.2 (length 29903 nt) was used as a reference sequence and for sequence comparisons. In the present report we have focused on sequence alignments, we have used NCBI BLAST, and CLUSTAL OMEGA.

IV. RESULTS AND DISCUSSION

We have characterised 2019-nCov spike protein RBD by conducting multiple sequence alignment between 2019-nCov, SARS-CoV, MERS-CoV and RaTG13. The alignment was performed using Clustal Omega. It shows that they are having significant identifies in this domain (Figure 1, yellow highlighted). However, RaTG13 RBD is closest (97% identity) to 2019-nCov. Our result is consistent with recent report of Zhou et al. that states that 2019-nCov is most closely related to the bat RaTG13, with which it forms a distinct lineage from other SARSr-CoVs, and that their S glycoproteins share 97% amino acid sequence identity [32].

Receptor recognition is the first step of viral infection and is a key determinant of host cell and tissue tropism. Previous structural work identified 14 positions for binding of SARS-CoV to human ACE2. Those are T402, R426, Y436, Y440, Y442, L472, N473, Y475, N479, Y484, T486, T487, G488, and Y491 [33]. Our result shows that 9 out of these 14 positions are strictly conserved in 2019-nCov, whereas the other 5 positions are semi-conservative R426/N, Y442/L, L472/F, N479/Q, Y484/Q (Figure 1, in red box). The conservation of key contact residues could explain the similar binding affinities of 2019-nCov to human ACE2. These probably suggest that 2019-nCov is well adapted to the ACE2 ortholog as the 2002–2003 epidemic strains of SARSCoV. This also explain the efficient transduction efficiency mediated by their respective S glycoproteins and the current rapid transmission in human.

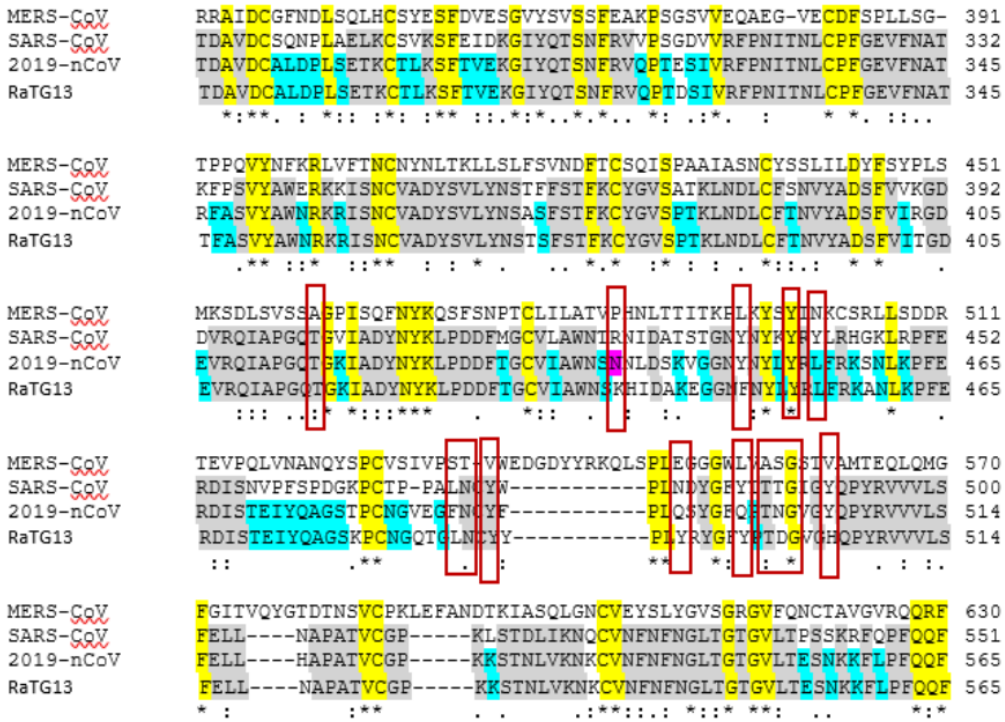

Figure 1: Multiple sequence alignment of RBDs of 2019-nCov , SARS-CoV, MERS-CoV and Bat spike (S)proteins. And GenBank accession numbers are QHR63250.1 (2019-nCov S), AY278488.2 (SARS-CoV S), AFS88936.1 (MERS-CoV S) and Bat spike protein QHR63300.2. Asterisks represent fully conserved residues, colons represent highly conserved residues, and periods represent lowly conserved residues. Conserved residues among 2019-nCov, SARS-CoV, MERS-CoV and Bat are

highlighted in yellow. Identical residues between 2019-nCov, SARS-CoV, and Bat are highlighted in cyan. Identical residues between 2019-nCov and Bat are highlighted in aqua colour. Human ACE2 interacting residues are in red box. The alignment was performed using Clustal Omega.

We have identified unique signature (681 to 684 residues) in 2019-nCov of S protein (highlighted with green in figure 2) at the boundary between the S1 and S2 subunits (figure 3). This region was reported as furin cleavage site [29]. We noticed that this is conserved among other 2019-nCov isolates (data not shown).

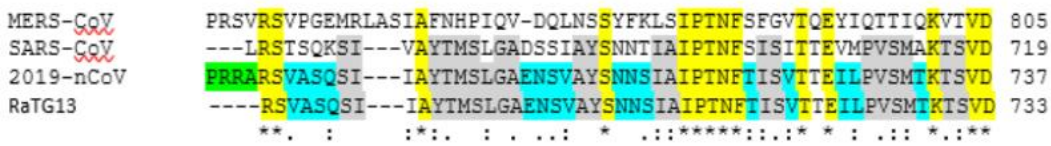

Figure 2: Multiple sequence alignment of C terminal end of S1 subunit of 2019-nCov , SARS-CoV , MERS-CoV and Bat spike (S)proteins. And GenBank accession numbers are QHR63250.1 (2019-nCov S), AY278488.2 (SARS-CoV S), AFS88936.1 (MERS-CoV S) and Bat spike protein QHR63300.2. Asterisks represent fully conserved residues, colons represent highly conserved residues, and periods represent lowly conserved residues. Conserved residues among 2019-nCov, SARS-CoV, MERS-CoV and Bat are highlighted in yellow. Identical residues between 2019-nCov, SARS-CoV, and Bat are highlighted in cyan. Identical residues between 2019-nCov and Bat are highlighted in aqua colour. Unique motif of 2019-nCov is highlighted in green. . The alignment was performed using Clustal Omega.

As this region was absent in SARS-CoV (figure 2) it probably indicates that S1/S2 cleavage during S biosynthesis was not necessary for S-mediated entry into the host cell. This polybasic cleavage site in S protein of 2019-nCov could putatively expand its tropism and/or enhance its transmissibility, compared with SARS-CoV. Earlier mutation study revealed that the detection of a polybasic cleavage site in the fusion glycoprotein of SARS-CoV-2 could putatively expand its tropism and/or enhance its transmissibility, compared with SARS-CoV and SARSr-CoV isolates, due to the near-ubiquitous distribution of furin-like proteases and their reported effects on other viruses [34].

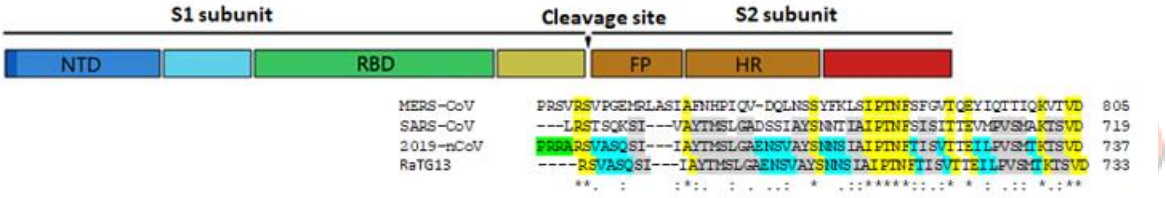

Figure 3: Structural diagram of 2019-CoV S protein (up). It contains S1 subunit and S2 subunit, which were divided by the S cleavage sites. FP, fusion peptide; HR, heptad repeat; RBD, receptor-binding domain, contains core binding motif in the external subdomain; signal peptide. Below: Sequence alignment of 2019-nCov, SARS-CoV, MERS-CoV and Bat spike (S)proteins displaying common S cleavage site

We have conducted sequence alignment for RBD domain of S protein for 2019-CoV, bat and pangolin. It showed that they have higher identity in this region compare to rest of the genome. 2019-CoV RBD sequence from 320 to 540 possesses 93% identity with pangolin whereas it is ~86% while considering complete spike protein. This is consistent with earlier study [35]. If we are focusing on only the spike RBD, pangolin has probability to cross host barriers and infect humans. Recent report indicates that pangolin-associated coronaviruses that belong to two sub-lineages of SARS-CoV-2-related coronaviruses, including one that exhibits strong similarity to SARS-CoV-2 in the receptor-binding domain [36]. The discovery of multiple lineages of pangolin coronavirus and their similarity to 2019-nCov suggests that pangolins can be considered as possible hosts in the emergence of novel coronaviruses.

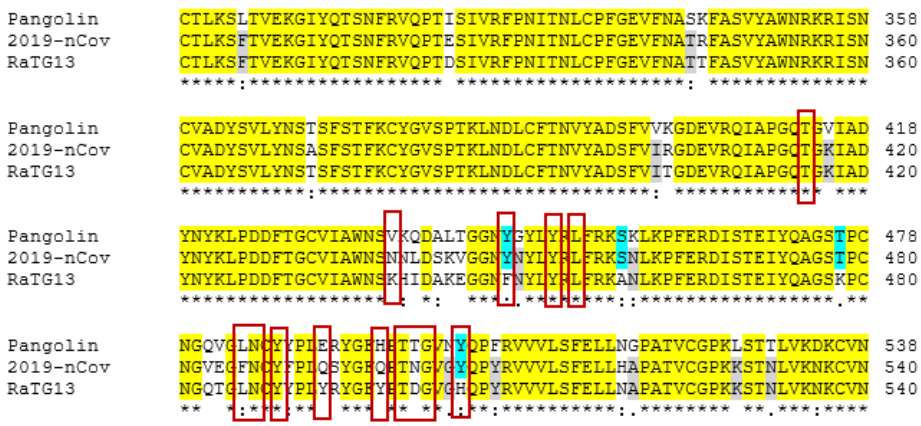

Figure 4: Multiple sequence alignment of RBDs of 2019-nCov (QHR63250.2), Pangolin (QIA48632.1)

Bat QHR63300.2. Asterisks represent fully conserved residues, colons represent highly conserved residues, and periods represent lowly conserved residues. Conserved residues among 2019-nCov), Pangolin, and Bat are highlighted in yellow. Identical residues between 2019-nCov and Bat are highlighted in cyan. Identical residues between 2019-nCov and Pangolin are highlighted in aqua. The alignment was performed using Clustal Omega

As mentioned earlier previous structural work identified 14 positions for binding of SARS-CoV to human ACE2 [37]. Those are T402, R426, Y436, Y440, Y442, L472, N473, Y475, N479, Y484, T486, T487, G488, and Y491 (marked red box in Figure: 1). Our result showed that 9 out of these 14 positions are strictly conserved in 2019-nCov, whereas the other 5 positions are semi-conservative R426/N, Y442/L, L472/F, N479/Q, Y484/Q (Figure 1, in red box). We have marked the corresponding residues of 2019-nCov in figure 4 (in red box) along with respective pangolin and BaTG13 residues. Our alignment result suggests that pangolin is more related to 2019-CoV RBD than RatG13 with respect to ACE2 binding residues. Moreover, the ability to engage ACE2 from different animal species appears to reflect host susceptibility to SARS-CoV infection and facilitated the jump of the virus from animals to humans [38, 39]. It was reported [40] that SARS-CoV-2 uses hACE2 as an entry receptor and recognizes it with a similar affinity to the 2002–2003 SARS-CoV isolates. This suggests that it can spread efficiently in humans, in agreement with the numerous SARS-CoV-2 human-to-human transmission events reported to date.

```
Pangolin        ECDIPVGAGICASYHSMS----SFRSVNQRSIIAYTMSLGAENSVAYSNNSIAIPTNFTI 714
2019-nCov       ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI 720
RaTG13          ECDIPIGAGICASYQTQTNS----RSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI 716
```

Figure 5: Multiple sequence alignment of C terminal of S1 subunit 2019-nCov (QHR63250.2), Pangolin (QIA48632.1)
Bat QHR63300.2. Asterisks represent fully conserved residues, colons represent highly conserved residues, and periods represent lowly conserved residues. Conserved residues among 2019-nCov, Pangolin, and Bat are highlighted in yellow. Identical residues between 2019-nCov and Bat are highlighted in cyan. Unique motif of 2019-nCov is highlighted in green. The alignment was performed using Clustal Omega.

Another recent study showed that SARS-CoV-2 RBD protein exhibits strong binding to its cell-associated and soluble ACE2 receptors with human and bat origin [41]. This RBD domain also demonstrated significantly higher binding affinity to ACE2 than SARS-CoV RBD. SARS-CoV-2 RBD protein could block S protein mediated SARS-CoV-2 pseudovirus and SARS-CoV pseudo virus entry into their respective ACE2 receptor expressing target cells, suggesting the potential of SARS-CoV-2 RBD protein as a viral attachment or entry inhibitor against SARS-CoV-2 and SARS-CoV [41].

Conclusion

All together our results provide a structural analysis to identify conserved regions in RBD across S proteins that will support ongoing research and vaccine design efforts. We identified the amino acid residues within RDB which may play important role for binding ACE2 receptor. It suggests that 2019-nCov can spread efficiently in humans, in agreement with the numerous 2019-nCov human-to-human transmission events reported to date. It underscores the importance of continued surveillance of coronaviruses at the sequence and functional levels to better prepare for the future.

DECLARATION OF INTERESTS

The author declares no competing financial interests

REFERENCES

1. Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J. & Jiang, S. 2009, "The spike protein of SARSCoV-a target for vaccine and therapeutic development" *Nat. Rev. Microbio*l. Vol.7,226–236.
2. Tortorici, M.A., and Veesler, D. 2019. "Structural insights into coronavirus entry" Adv. Virus Res. 105, 93–116.
3. Weinstein, R. A. 2004, Planning for epidemics–the lessons of SARS. *N. Engl. J. Med*.Vol. 350, pp. 2332–2334.
4. Ceraolo C, Giorgi FM. 2020, "Genomic variance of the 2019-nCoV coronavirus"*J Med Virol*. Vol. 92. Pp.522–528.
5. Li, W., Greenough, T.C., Moore, M.J., Vasilieva, N., Somasundaran, M., Sullivan,J.L., Farzan, M., and Choe, H. 2004 "Efficient replication of severe acute respiratory syndrome coronavirus in mouse cells is limited by murine angiotensin-converting enzyme 2". *J. Virol.* Vol. 78, Issue. 20, pp. 11429–11433.
6. Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., Lu, G., Wu, Y., Yan, J., Shi, Y.2017, "Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains.*" Nat. Commun.*
7. Wang Lu, G., , Q. & Gao, G. F. 2015, "Bat-to-human: spike features determining 'hostjump' of coronaviruses SARS-CoV, MERS-CoV, and beyond" *Trends Microbiol*, Vol. 23, pp. 468–478.
8. Wang, N. et al. 2013, "Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4". *Cell. Res*. Vol. 23, pp. 986–993.
9. WHO.. 2020, Naming the coronavirus disease (COVID-19) and the virus that causes it. https://www.who.int/emergencies/diseases/novelcoronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that -causes-it
10. Belouzard, S., Chu, V.C., and Whittaker, G.R. 2009, "Activation of the SARScoronavirus spike protein via sequential proteolytic cleavage at two distinct sites" *Proc. Natl. Acad. Sci. USA* Vol.106, Issue. 14, pp. 5871–5876.
11. Bosch, B.J., van der Zee, R., de Haan, C.A., and Rottier, P.J. 2003, "The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex". *J. Virol*. Vol. 77, Issue. 16, pp. 8801–8811.
12. Burkard, C., Verheije, M.H., Wicht, O., van Kasteren, S.I., van Kuppeveld, F.J. Haagmans, B.L., Pelkmans, L., Rottier, P.J., Bosch, B.J., and de Haan, C.A.2014, "Coronavirus cell entry occurs through the endo-/lysosomal pathwayin a proteolysis-dependent manner" *PLoS Pathog*. Vol. 10, e1004502.
13. Kirchdoerfer, R.N., Cottrell, C.A., Wang, N., Pallesen, J., Yassine, H.M.,Turner, H.L., Corbett, K.S., Graham, B.S., McLellan, J.S., and Ward, A.B. 2016, "Pre-fusion structure of a human coronavirus spike protein". Nature 531, 118–121.

14. Millet, J.K., and Whittaker, G.R. 2014, Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc. Natl. Acad. Sci*. USA Vol. 111, pp. 15214–15219.
15. Gui, M., Song, W., Zhou, H., Xu, J., Chen, S., Xiang, Y., and Wang, X. 2017 "Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding" *Cell Res.* Vol. 27, pp. 119–129.
16. Pallesen, J., Wang, N., Corbett, K.S., Wrapp, D., Kirchdoerfer, R.N., Turner, H.L., Cottrell, C.A., Becker, M.M., Wang, L., Shi, W., et al. 2017, "Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen" *Proc. Natl. Acad. Sci. USA* Vol. 114, E7348–E7357.
17. Song, W., Gui, M., Wang, X., and Xiang, Y. 2018, "Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2" *PLoS Pathog*. Vol. 14, e1007236.
18. Walls, A.C., Tortorici, M.A., Bosch, B.J., Frenz, B., Rottier, P.J.M., DiMaio, F., Rey, F.A., and Veesler, D. 2016, "Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer" *Nature* Vol. 531, pp.114–117.
19. Walls, A.C., Tortorici, M.A., Snijder, J., Xiong, X., Bosch, B.-J., Rey, F.A., and Veesler, D. 2017, "Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion", *Proc. Natl. Acad. Sci*. USA Vol. 114, pp. 11157–11162,
20. Millet, J.K., and Whittaker, G.R. 2014, "Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein." *Proc. Natl. Acad. Sci*. USA Vol. 111, pp. 15214–15219.
21. Chen, Y. et al. 2013, "Crystal structure of the receptor-binding domain from newlyemerged Middle East respiratory syndrome coronavirus". *J. Virol.* Vol. 87, pp. 10777–10783.
22. Lu et al. 2015, "Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond", *Trends in Microbiology* Vol. 23, Issue. 8, pp. 468-478.
23. Modjarrad, K. et al. 2016, "A roadmap for MERS-CoV research and productdevelopment: report from a World Health Organization consultation." *Nat.Med*. Vol. 22, pp. 701–705.
24. Wang, N., Shang, J., Jiang, S. & Du, L. 2020, "Subunit vaccines against emerging pathogenic human coronaviruses". *Front. Microbiol*. https://doi.org/10.3389/fmicb.2020.00298
25. Gao, J. et al. 2013, "Structure of the fusion core and inhibition of fusion by a heptadrepeat peptide derived from the S protein of Middle East respiratory syndrome coronavirus" *J. Virol*. Vol. 87, pp. 13134–13140.
26. Lu, G. et al. 2013, "Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26." *Nature*, 500, pp. 227–231.
27. WHO. Coronavirus Infections: Disease Outbreak News 2016, (WHO, http://www.who.int/csr/don/25-july-2016-mers-saudi-arabia/en/.
28. Channappanavar, R. et al. 2015, "Protective effect of intranasal regimens containing peptidic middle east respiratory syndrome coronavirus fusion inhibitor against MERS-CoV infection" *J. Infect. Dis*. Vol. 212, Issue. 12, pp. 1894–1903.
29. Liu Z, Xiao X, Wei X, et al. 2020 "Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. *J Med Virol*, pp. 1-7.
30. Lu, G. et al. 2013, "Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26". *Nature*, 500, pp. 227–231.
31. Xu, Y. et al. 2004, Crystal structure of severe acute respiratory syndrome coronavirus spike protein fusion core. *J. Biol. Chem*. Vol. 279, pp. 49414–49419.
32. Zhou et al. 2020, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, Vol 579, https://doi.org/10.1038/s41586-020-2012-7
33. Li, F., Li, W., Farzan, M., and Harrison, S.C. 2005a, "Structure of SARS coronavirus spike receptor-binding domain complexed with receptor". Science 309, 1864–1868.
34. Walls et al. 2020, "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein" *Cell* Vol. 180, pp. 281–292. https://doi.org/10.1016/j.cell.2020.02.05827.
35. Liu Z, Xiao X, Wei X, et al. 2020, " Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2", *J Med Virol*, pp. 1-7.
36. Lam et al. 2020, "Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature*. https://doi.org/10.1038/s41586-020-2169-0
37. Li, F., Li, W., Farzan, M., and Harrison, S.C. 2005a, "Structure of SARS coronavirus spike receptor-binding domain complexed with receptor", Science 309, 1864–1868.
38. Li, F. 2008, "Structural analysis of major species barriers between humans andpalm civets for severe acute respiratory syndrome coronavirus infections" .J. Virol. 82, 6984–6991.
39. Li, W., Greenough, T.C., Moore, M.J., Vasilieva, N., Somasundaran, M., Sullivan,J.L., Farzan, M., and Choe, H. 2004, " Efficient replication of severe acute respiratory syndrome coronavirus in mouse cells is limited by murine angiotensin-converting enzyme 2". *J. Virol.* Vol. 78, Issue. 20, pp. 11429–11433.
40. Walls et al. 2020 "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein"; *Cell* Vol. 180, pp. 281–292,. https://doi.org/10.1016/j.cell.2020.02.05827.
41. Tai et al. 2020, "Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine". *Cellular & Molecular Immunology*, https://doi.org/10.1038/s41423-020-0400-4