Contents lists available at ScienceDirect

Journal of Genetics and Genomics

Letter to the editor

Identification of potential cross-protective epitope between a new type of coronavirus (2019-nCoV) and severe acute respiratory syndrome virus

Recently, a new type of unknown virus causing severe respiratory infection was reported in Wuhan city, Hubei province, China (WHO, 2020b). Infection of this virus was first reported in December 2019, and origin of the virus was traced back to a large seafood/wide animal market in Wuhan city. The serious clinical symptoms of the viral infection, including fever, dry cough, dyspnea, and pneumonia, may result in progressive respiratory failure and even death. Moreover, the quick spread of the virus has caused an epidemic in China, as well as infection cases worldwide. The whole-genome sequence of Wuhan new virus (WH-Human_1) was first released on January 10, 2020 (Zhang, 2020), followed by additional ones released in Global Initiative on Sharing All Influenza Data (GISAID) (Shu and McCauley, 2017). Later, this new virus was determined and announced as a new type of coronavirus (CoV; 2019-nCoV) by the World Health Organization (WHO, 2020a).

CoVs are single-stranded RNA viruses that belong to the order *Nidovirales*, family *Coronaviridae*, and subfamily *Coronavirinae* (Schoeman and Fielding, 2019) and have been classified into four major groups: α-CoVs, β-CoVs, γ-CoVs, and δ-CoVs with 17 subtypes (Saminathan et al., 2014). CoVs primarily infect wild animals including mammals and birds. They also infect humans and cause various diseases such as upper and lower respiratory tract infections and respiratory syndromes. Among them, severe acute respiratory syndrome (SARS) CoV and Middle Eastern respiratory syndrome (MERS) CoV can cause serious respiratory syndrome in humans (Schoeman and Fielding, 2019). For instance, the outbreak of SARS in 2003 led to a pandemic with 8906 infected cases and 774 deaths reported worldwide (WHO, 2003). Meanwhile, the outbreak of MERS confirmed 2229 cases globally, including 791 associated deaths (WHO, 2018).

The coronaviral genome normally encodes four structural proteins including spike (S) protein, nucleocapsid (N) protein, membrane (M2) protein, and envelope (E) protein (Schoeman and Fielding, 2019). S protein contains the receptor-binding domain (RBD) and mediates the attachment of viruses to the surface receptors in host cells, as well as subsequent fusion between the viral and host cell membranes, to facilitate viral entry into host cells (Kirchdoerfer et al., 2016). Multiple binding and neutralization epitopes have been identified in the S proteins of CoVs (Hwang et al., 2006; Prabakaran et al., 2006; Reguera et al., 2012), which makes S protein an essential antigen for vaccine design.

Latest bioinformatic analysis indicated that 2019-nCoV is phylogenetically close to SARS-CoV and bat CoV (BCoV) (Xu et al., 2020). The genomes of 2019-nCoV and SARS-CoV share more than 79% sequence similarity on average (Zhou et al., 2020), and their S proteins share 76.47% identity (Xu et al., 2020). Yet the antigenicity similarity between them remains unknown and is urgently needed for vaccine design. Cross-reactive epitopes (CREs) are shared or similar epitope regions on the antigen surface among viruses that can be bound or neutralized by the same antibodies. Desirably, if any CREs were identified, previous antibodies for other CoVs might be reused to facilitate 2019-nCoV intervention.

Previously, we have developed an algorithm, namely, Conformational Epitope (CE)-BLAST, which enables antigenic similarity computation for new emerging pathogens, and have used it to successfully identify CREs between the dengue and Zika viruses (Qiu et al., 2018). In this study, we investigated the antigenic similarity of 2019-nCoV to other CoVs based on their spike antigens.

Sequences of S protein were downloaded for known CoVs from UniProt, and 2019-nCoV sequences were obtained from Shanghai Public Health Clinical Center and GISAID (Supplementary data). After data processing, a total of 53 unique S proteins were selected and structure modeled which represent different subtypes of CoVs, including 2019-nCoV (WH-Human_1), 3 SARS strains, 2 BCoV strains, and 47 strains from other CoVs (Table S1). S proteins of 2019-nCoV and SARS strains share high structural similarity with a root-mean-square deviation of 1.21 Å according to modeling structures. Individual epitope residues were derived from immune complexes of CoV S protein and further merged into 6 epitope regions (Table S2). Mapping the epitope regions to the 3D structure of S protein of 2019-nCoV demonstrated that five epitope regions are located in the RBD (Fig. S1). For each region, CE-BLAST calculates the similarity score of antigenicity between CoV pairs by comparing the physicochemical difference in 3D adjacent structural regions (Qiu et al., 2018). The higher the score, the better the potential for cross-reaction between the paired antigens.

The antigenic clustering of the 53 CoV strains was made according to the score matrix for each epitope region (Figs. 1 and S2). It can be seen that, although the clustering results may vary slightly for different regions, the antigenicity of CoV S proteins can be generally divided into two major groups. Detailed results for the 6th epitope region shown in Fig. 1A and B displayed the detailed clustering of the 2019-nCoV group, as well as the similarity score between them. Antigenically, 2019-nCoV is most similar to SARS-CoV, followed by BCoV (Fig. 1B). Similarity scores higher than 0.80 were detected between 2019-nCoV and SARS-CoV strains. Considering the default cutoff of 0.75 as antigenically similar (Qiu et al., 2018), these results indicate the existence of potential CREs between 2019-nCoV and SARS virus.
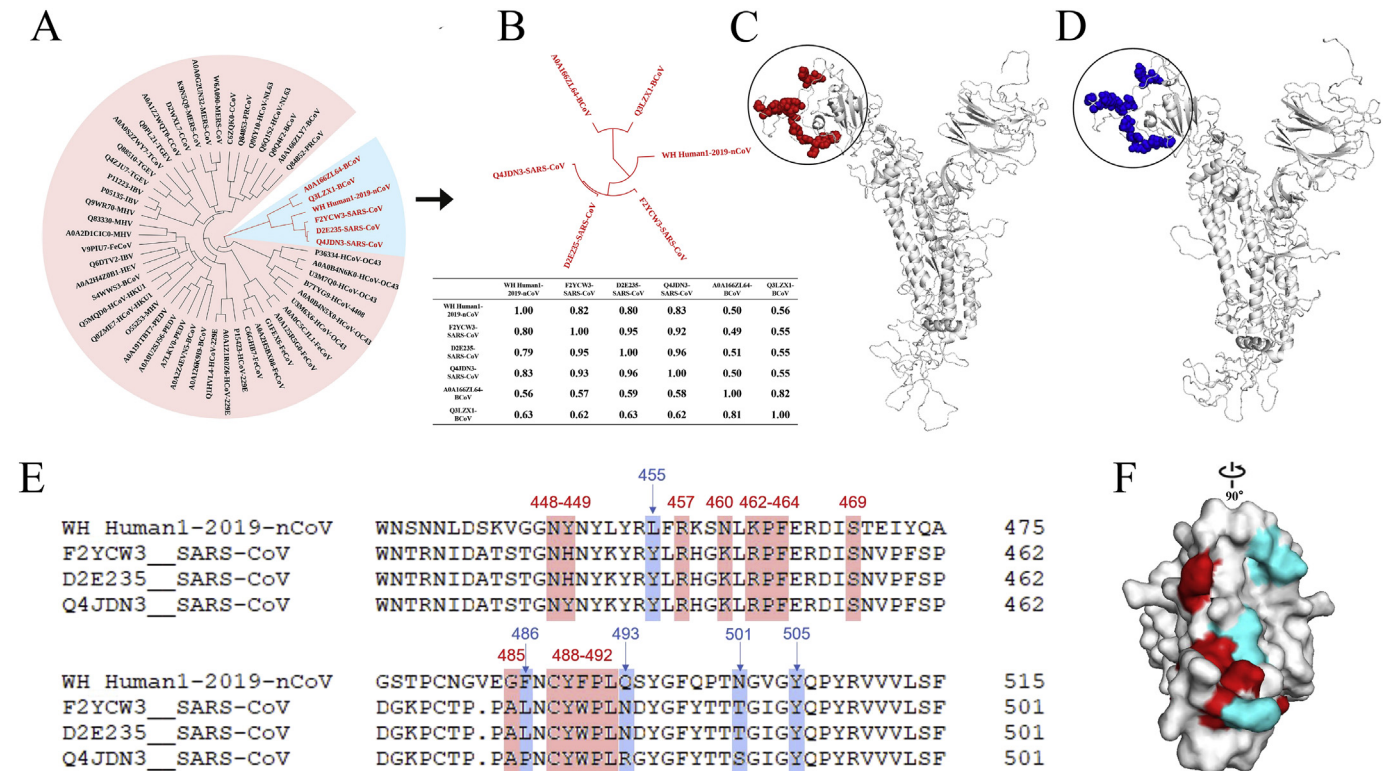
Fig. 1. Antigenic clustering and potential CREs between 2019-nCoV and SARS. **A**: Antigenic clustering of S proteins for 53 representative CoVs based on the 6th epitope regions. One cluster is marked in cyan, and the other is marked in pink. The 2019-nCoV is marked as WH Human1-2019-nCoV here. **B**: Details in the 2019-nCoV cluster with score matrix of antigenic similarity. **C**: Position mapping of potential CREs on the modeled structure of 2019-nCoV (WH-Human1-2019-nCoV). **D**: Position mapping of potential CREs on the modeled structure of SARS-CoV B039 (Q4JDN3). **E**: Sequence alignment of S proteins for 2019-nCoV and SARS-CoVs. Epitope regions are highlighted in red, and five essential human ACE2-binding sites are highlighted in blue. Sites are numbered based on the S protein sequence of 2019-nCoV. Dot in the sequence of SARS-CoV represents gap insertion. **F**: Surface view of CREs and ACE2-binding sites. CREs are marked in red, and ACE2-binding sites are marked in cyan. CRE, cross-reactive epitope; CoV, coronavirus; SARS, severe acute respiratory syndrome.

Further structural mapping showed the potential CREs in the S protein of 2019-nCoV (Fig. 1C) and SARS virus (Fig. 1D). Fig. 1E shows the multiple sequence alignment (MSA) for 2019-nCoV and SARS strains. The MSA results of full sequences can be found in Fig. S3. The potential CREs are conformational epitopes, in which the component residues are close in 3D space but disconnected in protein sequence (Fig. 1E). Compared with SARS virus, most mutations at CRE sites are residual substitutions of amino acids with similar properties, such as changes of alkaline residue Lys (K) to Arg (R) and aromatic residue Phe (F) to Trp (W).

Recently, the binding sites of S protein to human ACE2 were identified as residues 455, 486, 493, 501, and 505 (numbered according to the 2019-nCoV S protein sequence), which are located near the potential CRE positions (Fig. 1E). Interestingly, although no direct overlapping, the potential CRE residues are highly adjacent to essential ACE2-binding sites in the 3D structure (Zhou et al., 2020; Xu et al., 2020), suggesting that neutralization antibodies targeting CRE may also interfere or even block the interaction between the new CoV and ACE2 receptor because of steric hindrance (Fig. 1F). As such, the CRE region recommended here may likely become the cross-protective epitope between 2019-nCoV and SARS-CoV.

In this study, we used the computational method CE-BLAST to identify the potential CREs between 2019-nCoV and SARS-CoV. CE-BLAST requires epitope structures as the input file. Epitopes are usually referred to as highly specific and continuous areas on the surface of an antigen which can be recognized and bound by corresponding antibodies. Previous analysis on antibody-

antigen complexes found that more than 90% of epitope residues appeared conformational (Van Regenmortel, 1996). As conformational epitopes can be influenced by various factors such as mutation/insertion/deletion, structure change, or neighboring mutations, examining sequence conservation alone is usually not enough to evaluate the similarity of epitope regions among viruses.

In addition to the epitope similarity between 2019-nCoV and SARS-CoV, we also examined the location and residual distribution for each epitope region in S protein of 2019-nCoV. As Fig. S1 indicates, epitope region 3 is located in S1 domain, which is far from the important ACE2-binding site. For the rest 5 regions, although located in the RBD, some epitope residues become spatially scattered, rather than maintaining the epitope conformation, such as the case in region 1. Thus, region 6 was recommended here for CRE candidate because of its high score, surface continuity, and proximity to essential binding sites of ACE2. Possibility of an additional CRE may also exist in other regions of S protein or other antigens between 2019-nCoV and SARS-CoV. It is noted that these potential CREs are highly conformational. Following antibody design and screening is suggested to consider the whole domain as antibodies derived from linear epitopes might have difficulty in fully recognizing the CRE structures.

In summary, a highly similar epitope was identified computationally between the 2019-nCoV and SARS virus, in the region of the binding site of the S proteins to the human ACE2 receptor. This timely work may shed light on the vaccine intervention for the emergent 2019-nCoV.

Acknowledgments

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgg.2020.01.003.

References

Hwang, W.C., Lin, Y., Santelli, E., Sui, J., Jaroszewski, L., Stec, B., Farzan, M., Marasco, W.A., Liddington, R.C., 2006. Structural basis of neutralization by a human anti-severe acute respiratory syndrome spike protein antibody, 80R. J. Biol. Chem. 281, 34610–34616.

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., Ward, A.B., 2016. Pre-fusion structure of a human coronavirus spike protein. Nature 531, 118–121.

Prabakaran, P., Gan, J., Feng, Y., Zhu, Z., Choudhry, V., Xiao, X., Ji, X., Dimitrov, D.S., 2006. Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. J. Biol. Chem. 281, 15829–15836.

Qiu, T., Yang, Y., Qiu, J., Huang, Y., Xu, T., Xiao, H., Wu, D., Zhang, Q., Zhou, C., Zhang, X., Tang, K., Xu, J., Cao, Z., 2018. CE-BLAST makes it possible to compute antigenic similarity for newly emerging pathogens. Nat. Commun. 9, 1772.

Reguera, J., Santiago, C., Mudgal, G., Ordoño, D., Enjuanes, L., Casasnovas, J.M., 2012. Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. PLoS Pathog. 8, e1002859.

Saminathan, M., Chakraborty, S., Tiwari, R., Dhama, K., Verma, A., 2014. Coronavirus infection in equines: a review. Asian J. Anim. Vet. Adv. 9, 164–176.

Schoeman, D., Fielding, B.C., 2019. Coronavirus envelope protein: current knowledge. Virol. J. 16, 69.

Shu, Y., McCauley, J., 2017. GISAID: global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 22, 30494.

Van Regenmortel, M.H.V., 1996. Mapping epitope structure and activity: from one-dimensional prediction to four-dimensional description of antigenic specificity. Methods 9, 465–472.

WHO, 2020a. Novel Coronavirus-China. https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/.

WHO, 2020b. Pneumonia of Unknown Cause in China. https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/.

WHO, 2003. Summary of Probable SARS Cases with Onset of Illness from 1 November 2002 to 31 July 2003. http://www.who.int/csr/sars/country/table2004_04_21/en/index.html.

WHO, 2018. WHO MERS-CoV Global Summary and Assessment of Risk, August 2018 (WHO/MERS/RA/August18). http://www.who.int/csr/disease/coronavirus_infections/riskassessment-august-2018.pdf?ua=1.

Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X., Zhong, W., Hao, P., 2020. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. Sci. China Life Sci. 1–4. https://doi.org/10.1007/s11427-020-1637-5.

Zhang, Y.Z., 2020. Initial Genome Release of Novel Coronavirus. http://virological.org/t/initial-genome-release-of-novel-coronavirus/319?from=groupmessage.

Zhou, P., Yang, X., Wang, X., et al., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. https://doi.org/10.1038/s41586-020-2012-7.

Tianyi Qiu[1]
Shanghai Public Health Clinical Center, Fudan University, Shanghai, 200433, China

Tiantian Mao[1], Yuan Wang, Mengdi Zhou, Jingxuan Qiu
Department of Gastroenterology, Shanghai 10th People's Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, 200092, China

Jianwei Wang
NHC Key Laboratory of Systems Biology of Pathogens and Christophe Mérieux Laboratory, IPB, CAMS-Fondation Mérieux, Institute of Pathogen Biology (IPB), Chinese Academy of Medical Science & Peking Union Medical College, Beijing, 100730, China

Jianqing Xu[**]
Shanghai Public Health Clinical Center, Fudan University, Shanghai, 200433, China

Zhiwei Cao[*]
Department of Gastroenterology, Shanghai 10th People's Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, 200092, China

[**] Corresponding author.

[*] Corresponding author.
E-mail address: xujianqing@shphc.org.cn (J. Xu).
E-mail address: zwcao@tongji.edu.cn (Z. Cao).

23 January 2020
Available online 30 January 2020

---

[1] These authors contributed equally to this work.