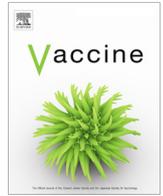




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



The biological characteristics of SARS-CoV-2 spike protein Pro330-Leu650



Qiu-dong Su^a, Yao Yi^a, Ye-ning Zou^b, Zhi-yuan Jia^a, Feng Qiu^a, Feng Wang^a, Wen-jiao Yin^a, Wen-ting Zhou^a, Shuang Zhang^a, Peng-cheng Yu^a, Sheng-li Bi^{a,*}, Li-ping Shen^{a,*}, Gui-zhen Wu^{a,*}

^a National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

^b Sinovac Biotech Co., Ltd, China

ARTICLE INFO

Article history:

Received 19 March 2020
Received in revised form 23 April 2020
Accepted 27 April 2020
Available online 30 April 2020

Keywords:

Biological characteristics
Spike protein Pro330-Leu650
SARS-CoV-2

ABSTRACT

SARS-CoV-2 is the cause of the worldwide outbreak of COVID-19 that has been characterized as a pandemic by the WHO. Since the first report of COVID-19 on December 31, 2019, 179,111 cases were confirmed in 160 countries/regions with 7426 deaths as of March 17, 2020. However, there have been no vaccines approved in the world to date. In this study, we analyzed the biological characteristics of the SARS-CoV-2 Spike protein, Pro330-Leu650 (SARS-CoV-2-SPL), using biostatistical methods. SARS-CoV-2-SPL possesses a receptor-binding region (RBD) and important B (Ser438-Gln506, Thr553-Glu583, Gly404-Aps427, Thr345-Ala352, and Lys529-Lys535) and T (9 CD4 and 11 CD8 T cell antigenic determinants) cell epitopes. High homology in this region between SARS-CoV-2 and SARS-CoV amounted to 87.7%, after taking the biological similarity of the amino acids into account and eliminating the receptor-binding motif (RBM). The overall topology indicated that the complete structure of SARS-CoV-2-SPL was with RBM as the head, and RBD as the trunk and the tail region. SARS-CoV-2-SPL was found to have the potential to elicit effective B and T cell responses. Our findings may provide meaningful guidance for SARS-CoV-2 vaccine design.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

The family *Coronaviridae* has historically been known to cause common colds and diarrheal illnesses in humans; however, SARS-CoV emerged as the causative agent of Severe Acute Respiratory Syndrome (SARS) in 2002 [1]. Seventeen years after SARS, the current outbreak of coronavirus disease (COVID-19) caused by SARS-CoV-2, which has a high similarity with SARS-CoV, has shocked the world [2]. As of March 17, 2020, SARS-CoV-2 has infected 179,111 people, involving 160 countries/regions worldwide, whereas China bears a significant burden, with approximately 45.3% of cases and 3231 deaths [3]. However, the evolution of SARS-CoV-2 has continued, with the identification of two major strains, type L (approximately 70%) and type S (approximately 30%) [4].

SARS-CoV-2 belongs to the beta-coronavirus genus, which includes SARS-CoV, (Middle East respiratory syndrome) MERS-CoV, bat SARSr-CoV, and others [5]. SARS-CoV-2 has a positive,

single-strand RNA genome that is over 29 kilobases in length [6]. Moreover, SARS-CoV-2 encodes four major structural proteins, the spike (S), membrane (M), envelope (E), and nucleocapsid (N) proteins [7]. The S protein is the most likely target of neutralizing antibodies as it is the main trans-membrane glycoprotein responsible for receptor-binding and virion entry [7]. Viral mutations occur in regions with much stronger MHC-I binding ability, while no other mutations are located near the S receptor-binding domain (RBD, Cys336-Glu516) [8].

Humoral immunity mediated by antibodies produced by B cells is critical for effective vaccines, whereas cellular immunity mediated by T cells (CD4 and CD8 T cells) is also considered to be essential [9]. The CD4 T cell response is critical to both antibody production and the killing of infected cells mediated by CD8+CTLs. In addition, only neutralizing antibodies can fully block viral entry into host cells; however, the general location of the binding of other antibodies would strongly affect the body's ability to produce neutralizing antibodies [7]. To date, no coronavirus vaccine has been approved [10]. Inactivated SARS-CoV has been demonstrated to offer incomplete protection, whereas purified coronavirus spike protein nanoparticles have been found to induce coronavirus neutralizing antibodies in mice [10]. Therefore, full exposure of

* Corresponding authors.

E-mail addresses: bisl@ivdc.chinacdc.cn (S.-l. Bi), shenlp@ivdc.chinacdc.cn (L.-p. Shen), wugz@ivdc.chinacdc.cn (G.-z. Wu).

neutralizing epitopes with other B cell epitopes are likely to represent important factors in coronavirus vaccine design.

The high homology of the S protein between SARS-CoV-2 and SARS-CoV resulted in the design of a subunit vaccine for SARS-CoV-2, which is focused on the RBD region and its vicinity. For the purpose of vaccine design, we analyzed the biological characteristics (e.g., homology analysis, T and B cell epitope prediction, and advanced structural analysis) by focusing on Pro330-Leu650 of the SARS-CoV-2 S protein (SARS-CoV-2-SPL) to evaluate its potential as a vaccine antigen of Pro330-Leu650.

2. Methods

2.1. Antigen sequences and alignment

The spike protein sequences of SARS-CoV-2 and SARS were obtained from GenBank (QHD43416.1 and AAP41037.1). Both of the protein sequences were aligned by MEGA X. The correspondence between SARS spike Pro317-Leu636 (SARS-CoV-SPL) and SARS-CoV-2 spike Pro330-Leu650 (SARS-CoV-2-SPL) was determined.

2.2. T and B cell epitope predictions of SARS-CoV-2-SPL

The B cell epitopes of SARS-CoV-2-SPL were predicted online with DiscoTope2 [11] and IEDB [12]. A DiscoTope score peak chart was subsequently constructed using GraphPad Prism 8.0. The MHC-I (9 mer) and MHC-II (15 mer) presentation scores were predicted by netMHCpan4 [13] and MARIA [14], respectively. Thirty-two MHC alleles were included in the analysis of CD4 and CD8 T cell epitopes with a more stringent 99.5% threshold for both netMHCpan4 and MARIA. An oligopeptide was selected if it was presented by more than one of three common alleles [8].

2.3. Homology modeling of SARS-CoV-2-SPL

The approximate tertiary structures of the monomer SARS-CoV-2-SPL (PDB: 6nb6.1.C) and SARS-CoV-SPL (PDB: 6cix.1.C), as well as the homotrimer SARS-CoV-2-SPL (PDB: 6vsb.1.A) were constructed with SWISS-MODEL using homology modeling [15].

3. Results

3.1. The Pro330-Leu650 region possesses RBD and important B and T cell epitopes.

Angiotensin-converting enzyme 2 (ACE2) was identified as the SARS-CoV-2 receptor, with its receptor binding domain (RBD) located on the spike protein [5]. The RBD region was referred to as the Arg319-Phe541; however, Cys336-Phe515 was sufficient for maintaining the RBD structure [5]. The SARS-CoV-2-SPL contains an RBD region (Pro330-Phe541) and tail region (Asn542-Leu650).

According to the DiscoTope scores of each amino acid, the potential B cell epitope motif distribution of SARS-CoV-2-SPL and SARS-CoV-SPL was created using GraphPad Prism software (Fig. 1). There were five B cell epitope motifs in the SARS-CoV-2-SPL, Ser438-Gln506 (Gly446, 3.263; Thr500, 3.290), Thr553-Glu583 (Lys558, -1.298), Gly404-Asp427 (Lys417, -2.402), Thr345-Ala352 (Ala352, -7.612), and Lys529-Lys535 (Asn532, -7.715) (Fig. 1A). Among these motifs, three were located in the RBD region, whereas two were located in the tail region. As expected, there were also five motifs in the SARS-CoV-SPL, Thr425-Gln492 (Thr433, 2.403; Thr486, 0.943), Leu538-Leu571 (Arg544, 1.618), Val389-Asp414 (Val404, -1.604), Thr332-Trp340

(Ala339, -5.031), and Thr517-Lys521 (Asp518, -8.507) (Fig. 1B). Furthermore, the two DiscoTope score peaks were almost mirror images (Fig. 1).

According to the MARIA percentiles for all possible 15 mer peptide sequences and NetMHCpan4 percentiles for all possible 9 mer peptide sequences for SARS-CoV-2-SPL, broad coverage was defined as being presented by 1/3 of MHC alleles [8]. We found that 15 CD4 epitopes and 17 CD8 T cell epitopes met this requirement. After overlapping integration, there were nine potential CD4 and 11 potential CD8 T cell antigenic determinants (Table 1). The RBD region possessed seven CD4 epitopes and nine CD8 T cell epitopes, whereas the tail region has two CD4 epitopes and two CD8 T cell epitopes. Moreover, when comparing these epitopes with verified CD4 and CD8 T cell epitopes of SARS-CoV-SPL [16], a one-to-one correspondence was observed (e.g., Ser516-Asn530 [SARS-CoV] to Ser530-Asn544 [SARS-CoV-2]) for the CD4 T cell epitope and Lys411-Val420 (SARS-CoV) to Lys424-Val433 (SARS-CoV-2).

3.2. High homology between SARS-CoV-2-SPL and SARS-CoV-SPL

The amino acid identity was 76.25% between SARS-CoV-2-SPL and SARS-CoV-SPL (Fig. 2). There were 76 amino acid differences between them, 36 of which were located in the receptor-binding motif (RMB) (Ser438-Gln506), accounting for 47.4% (36/76). Empirically, the biological characteristics of some amino acids (e.g. Arg-Lys, Thr-Ser, Glu-Asp, and Ile-Leu) were similar. After taking the biological similarity of amino acids into account and eliminating the RBM, the amino acid identity was approximately 87.7%.

According to the overall topology of the SARS-CoV-2-SPL monomer (Fig. 3A), there were six disulfide bonds (Cys336-Cys361, Cys379-Cys432, Cys480-Cys488, Cys391-Cys525, Cys538-Cys590, and Cys617-Cys649). Correspondingly, the SARS-CoV-SPL monomer also had six disulfide bonds (Cys323-Cys348, Cys366-Cys419, Cys467-Cys474, Cys378-Cys511, Cys524-Cys576, and Cys603-Cys635) (Fig. 3B). Furthermore, the six pairs of cysteines from the two coronaviruses were located at the same position in both the primary (Fig. 2) and tertiary (Fig. 3) structures. When artificially divided into three parts, the region contained a head (receptor-binding motif, RBM), trunk (RBD except RBM) and tail (remainders). The structural differences of the trunk, especially the head, were obvious (e.g., the head and trunk of SARS-CoV-2-SPL appeared to be more compact than those of SARS-CoV-SPL), whereas the tail region was highly similar (Fig. 3). Homotrimer SARS-CoV-2-SPL was arranged in a barrel with three RBMs situated on the top (Fig. 3C).

3.3. The acquisition of SARS-related antigen provides the basis for SARS-CoV-2 vaccine candidate preparation.

In our previous study regarding SARS-related antigens from 2003 [17], we prepared an insoluble SARS-related antigen with high antigenicity using prokaryotic expression, chromatography purification, and an in-house dialysis procedure. Moreover, the immunogenicity of this SARS-related antigen was also preliminarily confirmed (unpublished). Compared with inactivated virions and other subunit antigens prepared in cell culture, antigens prepared using our method have clear advantages, including decreased time and labor, as well as enhanced convenience.

As mentioned above, the presence of a complex structure and numerous cysteines in SARS-CoV-2-SPL is likely to seriously hinder the correct folding of the protein, which is crucial for the vaccine to generate effective immunity. Therefore, the successful preparation of SARS-related antigens may increase the speed at which a SARS-CoV-2 vaccine can be developed and thereby control the COVID-2019 outbreak.

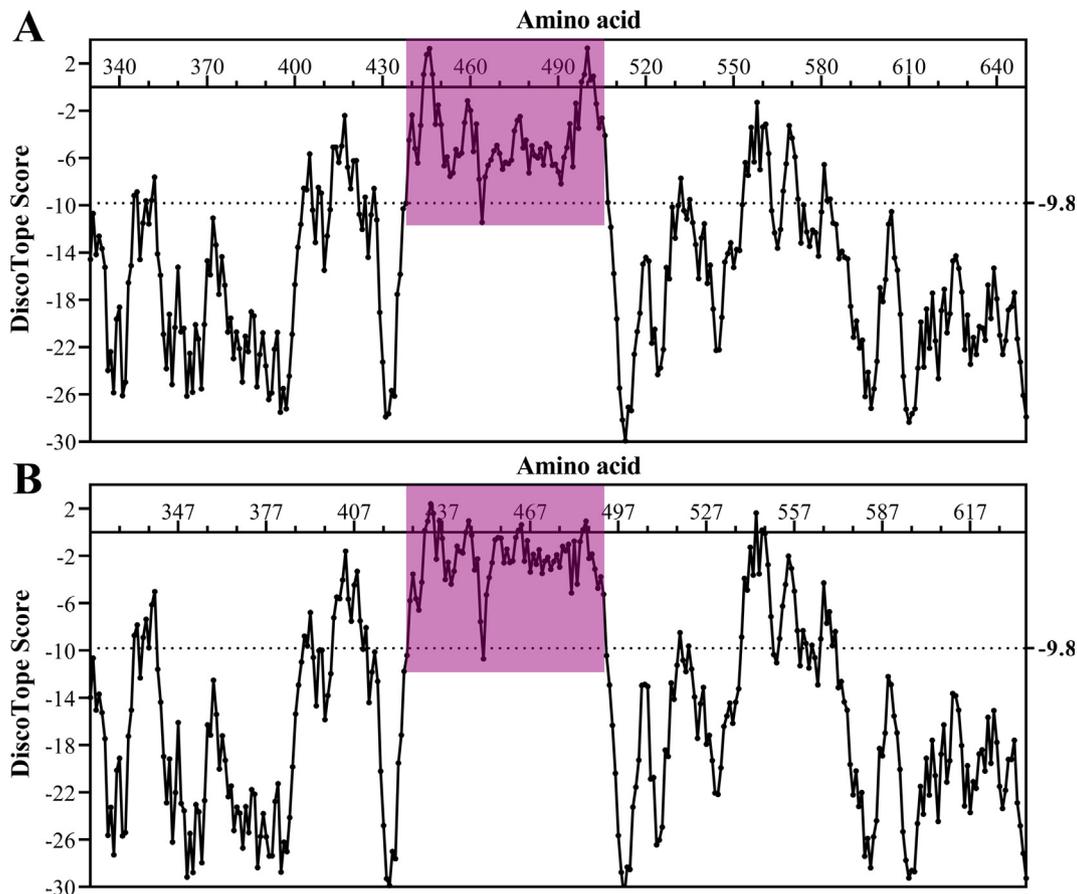


Fig. 1. The potential B cell epitope motifs in SARS-CoV-2-SPL (A) and SARS-CoV-SPL (B) (RBM in red-purple).

Table 1
Potential CD4 and CD8 T cell epitopes in the SARS-CoV-2-SPL.

Initial site (aa)	CD4 T cell epitope	Initial site (aa)	CD8 T cell epitope
347	FASVYAWNKRKISNC	345	TRFASVYAW
397	ADSFVIRGDEVQRQIAPGQTG	366	SVLYNSASFTF
430	TGCVIAWNSNNLDSKVGG	392	FTNVYADSFVI
468	ISTEIQAGSTPCNG	417	KIADYNYKL
494	SYGFQPTNGVGYQPY	433	VIAWNSNNL
512	VLSFELLHAPATVCG	444	KVGGNYNYL
531	TNLVKNKCVNFNFG	464	FERDISTEI
565	FGRDIADTTDAVRDPQ	495	YGFQPTNGV
629	LTPTWRVYSTGSNVFQ	503	VGYQPYRVVLSFEL
		576	VRDPQTLEILDITPCSF
		625	ADQLTPTWRVYSTGSNV

4. Discussion

Based on the knowledge gained from SARS-CoV vaccine development, a SARS-CoV-2 vaccine should target the structural proteins, particularly the RBD region [18]. The RBD of SARS-CoV-2 and SARS-CoV possesses a similar structure, despite amino acid variations at some key residues [6]. Therefore, many SARS-related studies may provide a valuable reference for SARS-CoV-2 vaccine development.

Amino acid sequence variations between the RBD of the SARS-CoV-2 spike protein RBD and other coronaviruses have been elucidated completely [6,19]. Despite the high sequence homology with the SARS-CoV spike protein (approx. 76%), SARS-CoV RBD-specific antibodies could not bind to SARS-CoV-2 [7], and vice versa

[20,21]. Therefore, exploiting SARS-CoV-related antigens to develop a SARS-CoV-2 vaccine is not practical. In this study, our focus was centered on SARS-CoV-2-SPL, which was found to contain the RBD and skeleton supporting its conformation. Furthermore, we analyzed the protein skeleton comparisons instead of performing simple sequence comparisons between SARS-CoV and SARS-CoV-2. The similar protein skeleton was suggestive of a similar protein structure and procedure of prokaryotic expression and chromatography purification.

Burial of the neutralization epitope and an unsustainable T cell response might result in the failure of an inactivated SARS vaccine. Moreover, the inclusion of effective B and T cell epitopes is essential for inducing specific humoral and cellular immunity in an ideal vaccine [22]. The recognition of effective B cell epitopes is dependent on antigenicity, accessibility of the surface and assistance of T helper (CD4 T) cells. CD4 T cells are the main components of the immune system found to impart immunological 'memory' to the new strains. Pathogen-specific T cell responses are also required for protection against infection [23]. In addition, a specific cellular response is considered to be equally important for sustainable protection and pathogen elimination, and CD4 T cells assist in promoting the production of antibodies and CTL reactions. The SARS-CoV-2-SPL was found to possess 5 B cell epitopes, 15 CD4 T cell epitopes, and 17 CD8 T cell epitopes, respectively. Moreover, from the overall topology (Fig. 3A), we found that most of these epitopes were exposed on the surface, which is an advantage of a small molecular protein. In addition, the RBD region of the SARS-CoV-2-SPL appeared to be compact, which might explain why SARS-CoV-2 RBD exhibited a stronger affinity with ACE [5]. The preparation of SARS-CoV-2-SPL was successfully performed in our laboratory. A high level of antibodies to SARS-CoV-2-SPL

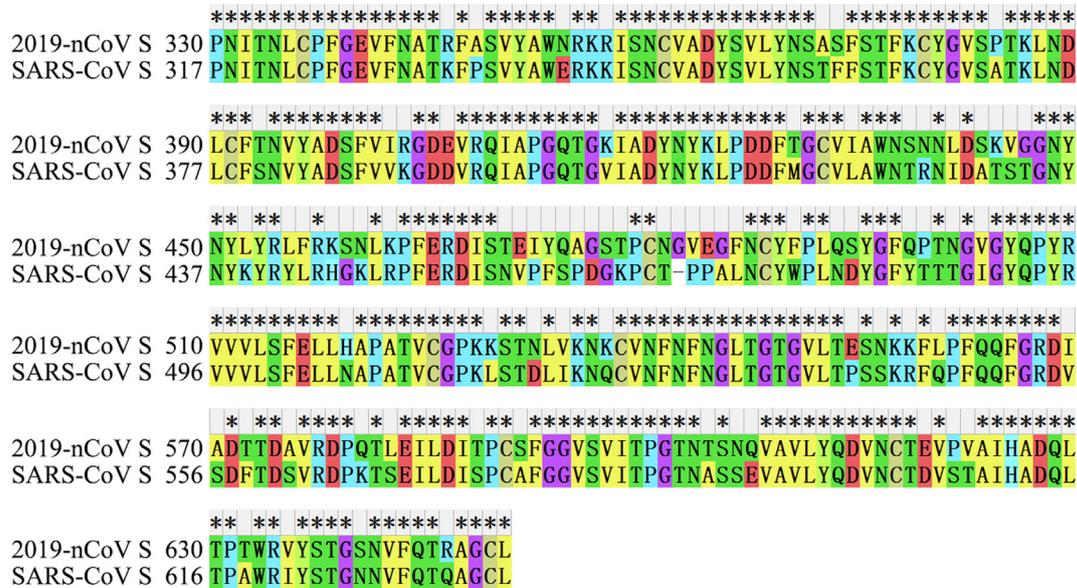


Fig. 2. Sequence alignment of SARS-CoV-2-SPL and SARS-CoV-SPL.

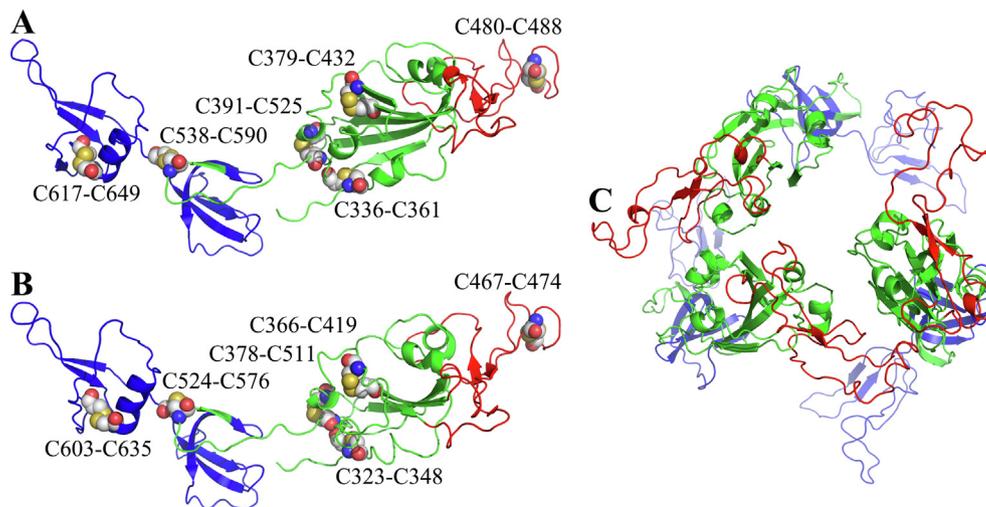


Fig. 3. Overall topology of SARS-CoV-2-SPL monomer (A), SARS-CoV-SPL monomer (B) and SARS-CoV-2-SPL homotrimer (C).

appeared in the Balb/C mice after immunization. However, further neutralization experiments demonstrated that antibodies other than neutralizing antibodies were predominantly generated. In this circumstance, antibody-dependent enhancement (ADE) may exist, which would enhance viral invasion. Thus, a potential vaccine antigen should not contain too many epitopes, especially B cell epitopes. Deleting unnecessary B cell epitopes or inserting neutralizing epitopes to virus-like particles may be a practical means of achieving this reduction.

Empirically, the structure of the three distinct parts (head, trunk, and tail) is critical for protein stability. Moreover, the topology of SARS-CoV-2-SPL and SARS-CoV-SPL appear similar to a goldfish, with RMB as the head, RBD as the trunk, and remainders as the tail (Fig. 3A and 3B). The RMB (head; critical for receptor binding) of SARS-CoV-2-SPL was completely exposed on the surface (Fig. 3A), which is also crucial for the production of neutralizing antibodies. Homotrimer SARS-CoV-2-SPL was arranged in a barrel with three RBMs on the top (Fig. 3A), which might represent the natural state of the SARS-CoV-2-SPL structure.

5. Conclusions

In this study, we analyzed the biological characteristics of SARS-CoV-2-SPL, and found it possessed a B cell-epitope-rich head and trunk and T cell-epitope-rich tail, giving it an ideal antigenic structure. Thus, these findings may provide meaningful guidance for future SARS-CoV-2 vaccine design.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by a grant from Chinese Center of Disease Control and Prevention. The funders had no role in study design, data collection and analysis, decision to publish, or prepa-

ration of the manuscript. We also thanked International Science Editing (<http://www.internationalscienceediting.com>) very much for polishing this manuscript freely.

References

- [1] Benvenuto D, Giovanetti M, Ciccozzi A, Spoto S, Angeletti S, Ciccozzi M. The 2019-new coronavirus epidemic: evidence for virus evolution. *J Med Virol* 2020;92:455–9. <https://doi.org/10.1002/jmv.25688>.
- [2] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel Coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020;382:727–33. <https://doi.org/10.1056/NEJMoa2001017>.
- [3] WHO. Coronavirus disease 2019 (COVID-19) Situation Report - 57. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200317-sitrep-57-covid-19.pdf?sfvrsn=a26922f2_4;17 March 2020.
- [4] Tang X, Wu C, Li X, Song Y, Yao X, Wu X, et al. On the origin and continuing evolution of SARS-CoV-2. *Nat Sci Rev* 2020:nwaa036. <https://doi.org/10.1093/nsr/nwaa036>.
- [5] Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Crystal structure of the 2019-nCoV spike receptor-binding domain bound with the ACE2 receptor. *J BioRxiv*. 2020:2020.02.19.956235. <http://doi.org/10.1101/2020.02.19.956235> %J bioRxiv.
- [6] Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 2020;395:565–74. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8).
- [7] Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020;367(6483):1260–3. <https://doi.org/10.1126/science.abb2507> % Science.
- [8] Fast E, Chen B. Potential T-cell and B-cell Epitopes of 2019-nCoV. *bioRxiv*. 2020:2020.02.19.955484. <http://doi.org/10.1101/2020.02.19.955484> %J bioRxiv.
- [9] Rappuoli R, Black S, Bloom DE. Vaccines and global health: In search of a sustainable model for vaccine development and delivery. *Sci Transl Med* 2019;11. <https://doi.org/10.1126/scitranslmed.aaw2888>.
- [10] Coleman CM, Liu YV, Mu H, Taylor JK, Massare M, Flyer DC, et al. Purified coronavirus spike protein nanoparticles induce coronavirus neutralizing antibodies in mice. *Vaccine* 2014;32:3169–74. <https://doi.org/10.1016/j.vaccine.2014.04.016>.
- [11] Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput Biol* 2012;8:. <https://doi.org/10.1371/journal.pcbi.1002829>.
- [12] Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The Immune Epitope Database (IEDB): 2018 update. *Nucl Acids Res* 2019;47: D339–43. <https://doi.org/10.1093/nar/gky1006>.
- [13] Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan- 4.0: Improved peptide-MHC Class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol* 2017;199:3360–8. <https://doi.org/10.4049/jimmunol.1700893>.
- [14] Chen B, Khodadoust MS, Olsson N, Wagar LE, Fast E, Liu CL, et al. Predicting HLA class II antigen presentation through integrated deep learning. *Nat Biotechnol* 2019;37:1332–43. <https://doi.org/10.1038/s41587-019-0280-2>.
- [15] Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 2011;27:343–50. <https://doi.org/10.1093/bioinformatics/btq662>.
- [16] Ng OW, Chia A, Tan AT, Jadi RS, Leong HN, Bertolotti A, et al. Memory T cell responses targeting the SARS coronavirus persist up to 11 years post-infection. *Vaccine* 2016;34:2008–14. <https://doi.org/10.1016/j.vaccine.2016.02.063>.
- [17] Yi Y, Duan SM, Zhang MC, Zhou YD, Gao Y, Yang GH, et al. Cloning and expression of SARS coronavirus spike gene fragment and its application. *Chin J Virol* 2003;19:267–8.
- [18] Tian X, Li C, Huang A, Xia S, Lu S, Shi Z, et al. Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerg Microbes Infect* 2020;9(1):382–5. <https://doi.org/10.1080/22221751.2020.1729069>.
- [19] Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat Med* 2020. <https://doi.org/10.1038/s41591-020-0820-9>.
- [20] Ju B, Zhang Q, Ge X, Wang R, Yu J, Shan S, et al. Potent human neutralizing antibodies elicited by SARS-CoV-2 infection. 2020:2020.03.21.990770. <http://doi.org/10.1101/2020.03.21.990770> %J bioRxiv.
- [21] Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun* 2020;11:1620. <https://doi.org/10.1038/s41467-020-15562-9>.
- [22] Dudek NL, Perlmutter P, Aguilar MI, Croft NP, Purcell AW. Epitope discovery and their use in peptide based vaccines. *Curr Pharm Des* 2010;16:3149–57. <https://doi.org/10.2174/138161210793292447>.
- [23] Feng Y, Qiu M, Zou S, Li Y, Luo K, Chen R, et al. Multi-epitope vaccine design using an immunoinformatics approach for 2019 novel coronavirus in China (SARS-CoV-2). 2020:2020.03.03.962332. <http://doi.org/10.1101/2020.03.03.962332> %J bioRxiv.