Bioinformatic prediction of immunodominant regions in spike protein for early diagnosis of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

Siqi Zhuang [1], **Yufeng Dai** [1], **Lingli Tang** [1], **Hongzhi Chen** [Corresp. 2]

[1] Department of Laboratory Medicine, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China

[2] Department of Metabolism & Endocrinology, Metabolic Syndrome Research Center, Key Laboratory of Diabetes Immunology, Ministry of Education, National Clinical Research Center for Metabolic Disease, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China

Corresponding Author: Hongzhi Chen
Email address: chenhongzhi2013@csu.edu.cn

Background. To contain the pandemics caused by SARS-CoV-2, early detection approaches with high accuracy and accessibility are critical. Generating an antigen-capture based detection system would be an ideal strategy complementing the current methods based on nucleic acids and antibody detection. The spike protein is found on the outside of virus particle and appropriate for antigen detection.

Methods. In this study, we utilized bioinformatics approaches to explore the immunodominant fragments on spike protein of SARS-CoV-2.

Results. The S1 subunit of spike protein was identified with higher sequence specificity. Additionally, glycosylation sites and high-frequency mutation sites on spike protein were circumvented in the antigen design. Three immunodominant fragments, $Spike_{56\text{-}94}$, $Spike_{199\text{-}264}$, and $Spike_{577\text{-}612}$, located at the S1 subunit were finally selected via bioinformatics analysis. All these fragments present qualified antigenicity, hydrophilicity, and surface accessibility. A recombinant antigen with a length of 194 amino acids (aa) consisting of the selected immunodominant fragments as well as a universal Th epitope was finally constructed.

Conclusion. The recombinant peptide encoded by the construct contains multiple immunodominant epitopes, which could stimulate strong immune response in mice and generate qualified antibodies for SARS-CoV-2 detection.

Bioinformatic prediction of immunodominant regions in spike protein for early diagnosis of the severe acute respiratory syndrome coronavirus 2（SARS-CoV-2）

Siqi Zhuang[2], Yufeng Dai[2], Lingli Tang[2], Hongzhi Chen[1]

1. Department of Metabolism & Endocrinology, Metabolic Syndrome Research Center, Key Laboratory of Diabetes Immunology, Ministry of Education, National Clinical Research Center for Metabolic Disease, The Second Xiangya Hospital, Central South University, Changsha, Hunan 410011, China

2. Department of Laboratory Medicine, The Second Xiangya Hospital, Central South University, Changsha, Hunan 410011, China.

Corresponding author:

Hongzhi Chen[1]

E-mail address: chenhongzhi2013@csu.edu.cn

Abstract

Background. To contain the pandemics caused by SARS-CoV-2, early detection approaches with high accuracy and accessibility are critical. Generating an antigen-capture based detection system would be an ideal strategy complementing the current methods based on nucleic acids and antibody detection. The spike protein is found on the outside of virus particle and appropriate for antigen detection.

Methods. In this study, we utilized bioinformatics approaches to explore the immunodominant fragments on spike protein of SARS-CoV-2.

Results. The S1 subunit of spike protein was identified with higher sequence specificity. Additionally, glycosylation sites and high-frequency mutation sites on spike protein were circumvented in the antigen design. Three immunodominant fragments, $Spike_{56-94}$, $Spike_{199-264}$, and $Spike_{577-612}$, located at the S1 subunit were finally selected via bioinformatics analysis. All these fragments present qualified antigenicity, hydrophilicity, and surface accessibility. A recombinant antigen with a length of 194 amino acids (aa) consisting of the selected immunodominant fragments as well as a universal Th epitope was finally constructed.

Conclusion. The recombinant peptide encoded by the construct contains multiple immunodominant epitopes, which could stimulate strong immune response in mice and generate qualified antibodies for SARS-CoV-2 detection.

Keywords: SARS-CoV-2, Spike protein, Antigen-capture, Immunodominant fragments, COVID-19

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is highly contagious and has caused more than 44 million infection cases and over 1 million deaths (https://www.who.int/), posing huge economic and social burden internationally [1,2]. The reports of SARS-CoV-2 reinfection cases suggest that stronger international efforts are required to prevent COVID-19 re-emergence in the future[3]. Nevertheless, we cannot exclude the possibility that SARS-CoV-2 become a seasonal epidemic [4]. Even worse, the large amount of asymptomatic infections greatly increases the difficulties of epidemic control [5]. To date, most vaccines against SARS-CoV-2 are still in the stages of clinical trials and their efficacy is uncertain (https://www.who.int/). Thus, early diagnosis of infected cases and population screening would still be the focus now [6].

The real-time reverse transcriptase polymerase chain reaction(RT-PCR) and antibody-capture serological tests are currently the main diagnostic methods for SARS-CoV-2 [7]. As the golden standard, RT-PCR is highly reliable [8,9]. However, the implementation costs and relatively cumbersome operation problems make it a big challenge for large population screening[10]. The antibody-capture serological test is convenient, but seroconversion generally occurs in the second or third week of illness. Therefore, it is not ideal for the early diagnosis of infection [11-13]. The antigen-capture test is an alternative diagnostic method that relies on the immunodetection of viral antigens in clinical samples. Accordingly, this method could be applied for the detection of early infection no matter if the patient was asymptomatic or not [14]. Compared with RT-PCR based detection method, it is relatively inexpensive and can be used at

the point-of-care.

Rapid viral antigen detection has been successfully used for diagnosing respiratory viruses such as influenza and respiratory syncytial viruses [15-23]. The sensitivity and specificity of the antigen-capture detection system depend highly on the antigen employed to generate antibodies [22]. The spike protein is one of the structural proteins of SARS-CoV-2, with the majority located on the outside surface of the viral particles [24-26]. It has a 76.4% homology with the spike protein of SARS-CoV. Sunwoo's study showed that the bi-specific spike protein derived monoclonal antibody system exhibited excellent sensitivity in SARS-CoV detection [27]. The virus infection is initiated by the interaction of spike protein receptor-binding domain (RBD) and angiotensin converting enzyme 2 (ACE2) on host cells. It is widely accepted that the spike protein is one of the earliest antigenic proteins recognized by the host immune system [28-32]. Nevertheless, the difficulties of using spike protein as antigen are also obvious. It is not easy to express and purify the full-length spike protein [33]. In addition, the spike protein is highly glycosylated [26] and prone to mutation[34], which would challenge the sensitivity of antigen-capture based detection method. Hence, it is critical to truncating the glycosylation and mutation sites on spike protein as much as possible in antigen design [33,35]. In fact, a study using the truncated spike protein to detect SARS-CoV achieved a diagnostic sensitivity of >99% and a specificity of 100% [36], which suggests that truncated spike protein of SARS-CoV-2 could also be an appropriate candidate for the early diagnostic testing and screening of SARS-CoV-2. In this study, we analyzed the spike protein via bioinformatics tools to obtain immunodominant fragments. The predicted sequences were joined together as a novel antigen to immunize mice

for the preparation of specific antibodies. The whole flowchart of our work is depicted in Fig. 1. Epitopes information presented by this work may aid in developing a promising antigen-capture based detection system in pandemic surveillance and containment.

Method

Data retrieval and sequence alignment

Human coronavirus (HCoV) includes α-coronaviruses and β-coronaviruses. HCoV-229E and HCoV-NL63 belong to the former, HCoV-HKU1, HCoV-OC43, the Middle East respiratory syndrome-related coronavirus (MERS-CoV), SARS-CoV, and the SARS-CoV-2 belong to the latter. We utilized the NCBI database to obtain sequences of all known human-related coronaviruses spike protein in this study (HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1, SARS-CoV, MERS-CoV, and SARS-CoV-2). Then, the Clustal Omega Server-Multiple Sequence Alignment was used to analyze the downloaded sequences. Clustal Omega is a new multiple sequence alignment program that uses HMM profile-profile techniques and seeds guide trees to generate alignments between three or more sequences. In this study, we set our parameters as default [37]. For further comparison between SARS-CoV-2 and SARS-CoV, we exerted the EMBOSS Needle Server-Pairwise Sequence Alignment. Needleman-Wunsch alignment algorithm supports this server to find the optimum alignment (including gaps) of two sequences by reading two input sequences and translating their optimal global sequence alignment to files [38]. It was also performed to compare major domains between SARS-CoV-2 and SARS-CoV.

Linear B-cell epitope prediction

Linear B-cell epitopes of the SARS-CoV-2 spike protein were calculated by ABCpred and Bepipred v2.0 server. The ABCpred server predicts B-cell epitopes based on artificial neural networks (machine-based technique). It grants a score for each predicted peptide. The peptides with high scores are more likely to be effective epitopes. For ABCpred, we set a threshold of 0.8 to achieve a specificity of 95.50% and an accuracy of 65.37% for prediction. The window length was set to 16 (the default window length) in this study [39]. The BepiPred v2.0 combines a hidden Markov model and a propensity scale method. The score threshold for the epitope was set to 0.5(the default value) to obtain a specificity of 57.16% and a sensitivity of 58.56% [40]. The residue with scores above 0.5 was predicted to be part of an epitope.

T-cell epitope prediction

The free online service TepiTool server, provided by the Immune Epitope Database (IEDB), was used to forecast epitopes binding to MHC-I and MHC-II molecules. TepiTool is a part of IEDB, providing some top MHC binding prediction algorithms for many species such as humans, mice, and so on [41]. For MHC-I binding epitopes, the mice were selected as host species. And we selected alleles including H-2-Db, H-2-Dd, H-2-Kb, H-2-Kd, H-2-Kk, and H-2-Ld for analysis. We took “IEDB recommended” as a prediction method and selected sections with predicted consensus percentile rank ≤1 as predicted epitope. As the default prediction method, “IEDB recommended” is based on the availability of predictors and previously observed predictive performance. The purpose of this method was to ensure the best possible method for a

given MHC-I molecule [42]. For MHC-II binding epitopes, we selected the mice as host species and adopted alleles including H2-IAb, H2-IAd, and H2-IEd for analysis. As same to MHC-I, we chose "IEDB recommended" as a prediction method, and peptides with predicted consensus percentile rank ≤10 were determined as epitopes. As above, this approach chose the best suitable method for a given MHC-II molecule according to evaluations by the IEDB team and bioinformatics community [43,44].

Profiling and evaluation of selected fragments

The secondary structure of the SARS-CoV-2 spike protein (PDB ID: 6VSB chain B) was calculated by the PyMOL molecular graphics system using the SSP algorithm. PyMOL (http://www.pymol.org) is a python-based tool, which is widely used when visualization of macromolecules is needed. Most tasks of representations could be achieved by PyMOL [45]. Vaxijen2.0 server was utilized to analyze the antigenicity of epitopes and selected fragments. VaxiJen was the first server for alignment-independent prediction of protective antigens, which is based on auto cross-covariance (ACC) transformation of protein sequences into uniform vectors of principal amino acid properties. A default threshold of 0.4 was set and the prediction accuracy of this server is 70% to 89 % [46]. The hydrophilicity of the selected fragment was analyzed by the online server ProtScale. ProtScale was one of the protein identification and analysis tools in the ExPASy server, which could calculate hydrophilicity based on amino acid scales [47]. Surface accessibility of predicted fragments was evaluated with the NetsurfP, an online server predicting the surface accessibility and secondary structure of amino sequence [48].

Critical features such as allergenicity and toxicity were evaluated by online server AllerTOP v2.0 [49] and ToxinPred [50]. PyMOL was used to visualize the selected fragments on the SARS-CoV-2 spike protein trimer. At last, we run a Protein BLAST search providing by NCBI database for the eventual construct, to check the possibility of cross-reactivity with other human and mouse proteins.

Results

Sequence alignment of spike protein in different coronaviruses

Coronaviruses had four genera composed of alpha-, beta-, gamma- and delta-coronaviruses. Among them, alpha- and beta- genera could infect humans. Including SARS-CoV-2, seven coronaviruses are known to infect humans (HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1, SARS-CoV, MERS-CoV, and SARS-CoV-2) [51,52]. Sequences of these viruses were obtained from the NCBI database, of which accession numbers were presented in Fig. 2A. We performed sequence alignment to determine the evolutionary relationships between SARS-CoV-2 and previously identified coronaviruses. According to the results of sequence alignment and phylogenetic tree (Fig. 2A, Fig. 2B), SARS-CoV is the closest virus to SARS-CoV-2 among the seven HCoVs, exhibiting a 77.46% sequence identity. To better understand the divergence of spike protein sequences between SARS-CoV-2 and SARS-CoV, we further analyzed the sequences of several major domains. Results showed that the S2 subunit was the most conserved domain with a 90.0% identity. While RBM and NTD domains, both located in the S1 subunit, only exhibited 49.3% and 50.0% identity respectively (Fig. 2C, Fig. 2D). Hence, we chose the

S1 subunit (amino acid 1-685) for the subsequent bioinformatics analysis as its potentially higher specificity.

Linear B-cell epitope prediction of S1 subunit in SARS-CoV-2 spike protein

The B-cell epitope is a surface accessible cluster of amino acids, which could be recognized by secreted antibodies or B-cell receptors. It could elicit humoral immune response [53]. The immunodominant fragments should contain high-quality linear B-cell epitopes to stimulate antibody production effectively. The sequence of SARS-CoV-2 S1 subunit was evaluated through ABCpred and BepiPred-2.0. The antigenicity was calculated via VaxiJen v2.0 with the given cutoff of $\geqslant$0.40. A total of 31 peptides were identified by the ABCpred algorithm (threshold set at 0.8, Table S1). For the Bepipred v2.0 server, we set a threshold of 0.5, and 14 epitopes were forecasted (Table S2). After antigenicity evaluation, 19 and 9 potential linear B-cell epitopes predicted by the ABCpred server and BepiPred v2.0 server were obtained respectively (Table 1). The area predicted by both bioinformatics programs is more likely to be an epitope recognized *in vivo*. After mapping the positions of peptides identified by these servers, 3 regions containing predicted epitopes were obtained. These regions could be preliminarily considered as candidates for immunodominant fragments (Fig. 3, Table 2).

Murine T-cell epitope prediction of S1 subunit in SARS-CoV-2 spike protein

Helper T cells (Th) recognize antigen peptides presented by MHC-II molecules and facilitate the humoral immune response [54,55]. Therefore, the immunodominant fragments had better contain MHC-II binding epitopes as well as linear B-cell epitopes. The S1 subunit was selected for the prediction of T-cell epitopes. We utilized the TepiTool server to forecast MHC-I and MHC-II binding epitopes. A total of 35 MHC-I binding epitopes was predicted (Table S3), and 27 peptides were identified as MHC-II binding epitopes (Table S4). The antigenicity of these peptides was calculated via Vaxijen 2.0 server (Table 3). Combined with the MHC-II epitopes prediction results, the candidate immunodominant fragments were adjusted (Fig.4). Compared with the preliminary candidate immunodominant fragments screened according to the linear B-cell epitope prediction, we added the $Spike_{14\text{-}34}$ fragment into consideration because it contains a linear B epitope and an MHC-II binding epitope, both of which had high antigenicity scores (Table 4).

Immunodominant fragments refinement according to the glycosylation site distribution, mutation site distribution, and secondary structure

A profile of 24 glycosylation sites of SARS-CoV-2 spike protein has been reported [56]. Since glycans could hinder the recognition of antigens by shielding the residues [57], which would affect the performance of antigen detection. Thus, when selected the immunodominant fragments, we circumvent glycosylation sites as much as possible. According to the study of Asif Shajahan et al, 15 glycosylation sites were located in the S1 subunit of the spike. Among them,

210 sites Asn^{17} and Asn^{603} were completely unoccupied[56]. Hence, the fragments in this study were
211 adjusted to $Spike_{14-34}$, $Spike_{49-101}$, $Spike_{199-261}$, and $Spike_{583-620}$. To retain some epitopes with high
212 antigenicity, the fragments inevitably contained 3 glycosylation sites.
213 Rapid transmission of COVID-19 provides the SARS-CoV-2 with substantial opportunities
214 for natural selection and mutations. To ensure the stability of the detection effect, the
215 immunodominant fragments were modified to avoid high-frequency mutation sites[58]. $Spike_{14-34}$
216 were excluded for containing four high-frequency mutation sites. Fragment $Spike_{49-101}$ was
217 adjusted to $Spike_{56-92}$, and fragment $Spike_{583-620}$ was adjusted to $Spike_{583-609}$. By adjusting the
218 fragments, we avoided in total of 8 high-frequency mutation sites. The adjusted fragments
219 contain none of the above high-frequency mutation sites, which could be beneficial to future
220 detection.
221 We use PyMOL to present the secondary structure of the spike protein （PDB ID: 6VSB
222 ）(Fig. S1). Attention was paid to maintaining the integrity of the secondary structure of the start
223 and end positions of fragments. Hence, the immunodominant fragments were finally adjusted to
224 $Spike_{56-94}$, $Spike_{199-264}$, and $Spike_{577-612}$. The epitopes and glycosylation sites contained in the
225 selected immunodominant fragments were displayed in Fig. 5.
226

227 Profiling, evaluation, and visualization of selected immunodominant
228 fragments

229 To further evaluate the antibody binding potentiality of these antigenic regions, several key
230 features of the selected fragments such as antigenicity, hydrophilicity, surface accessibility,

231 toxicity, and allergenicity were analyzed (Table 5). We used the full-length of subunit 1 spike
232 protein for computation of hydrophilicity and surface accessibility and analyzed antigenicity,
233 toxicity, and allergenicity using the sequence of selected fragments. Three fragments presented
234 relatively moderate hydrophilicity and surface accessibility. The proportion of hydrophilic amino
235 acids to the total length of the fragments $Spike_{56-94}$, $Spike_{199-264}$, and $Spike_{577-612}$ are 19/39, 30/66,
236 12/36 respectively. The surface accessibility of these fragments calculated by the online server
237 was shown in Table 5.
238 The antigenicity of the selected fragments was tested by the Vaxijen2.0 server. The toxicity
239 of the selected fragments was examined by ToxinPred, and no fragment was predicted to be
240 toxic. We assessed the allergenicity via the AllerTOP v2.0 server. Only fragment $Spike_{577-612}$ was
241 predicted to be a probable allergen. Attention should be paid to monitor potential allergic
242 reactions when injecting the recombinant protein into mice in further experiments. And the
243 selected fragments were presented as the sphere in the trimer structure (Fig. 6). Next, we
244 searched the sequence of the selected fragments and the epitopes in the IEDB server to determine
245 whether they were experimentally tested. None of the fragments contain experimentally tested
246 epitopes on the IEDB server, which also indicated that the fragments are highly specific for
247 SARS-CoV-2.
248

249 Immunodominant fragments based recombinant antigen design

250 Three immunodominant fragments embody several linear B-cell epitopes, MHC-I binding,
251 and MHC-II binding T-cell epitopes were selected. As a universal Th epitope, the PAN DR

252 epitope [PADRE(AKFVAAWTLKAAA)] can activate CD4+ cells, enhance helper T cell
253 activity, and assist activation of humoral immunity [59]. It was added into the construction
254 aiming to boost helper T cell activity [60,61]. $(GGGGS)_n$ is a wildly used flexible linker with the
255 function of segmenting protein fragments, maintaining biological activity, maintaining protein
256 conformation, and promoting protein expression [62]. Finally, we combined the fragments and a
257 PADRE epitope by linker peptide $(GGGGS)_2$ and $(GGGGS)_3$ [62](Fig. 7). The predicted
258 antigenicity of the final construct (194 aa) was 0.5690 (Table 6). A protein BLAST for the final
259 construct was conducted to evaluate the possibility of cross-reactivity. The blast results
260 suggested little similarity between the construct and any human or mice proteins (data not
261 shown).
262
263

264 Discussion

265 In this study, we explored the immunodominant fragments within the S1 subunit of the
266 SARS-CoV-2 spike protein. The final construct consists of three immunodominant fragments
267 $Spike_{56\text{-}94}$, $Spike_{199\text{-}264}$, $Spike_{577\text{-}612}$, and a PADRE epitope. After expression and purification, the
268 recombinant antigen will be used to immunize mice and qualified antibody generated could be
269 applied for developing an antigen-capture based detection system.
270 The antibody-based antigen capturing method is user-friendly, time-saving, and economical.
271 Thus, it is an ideal complementary detection strategy especially for early diagnosis and large
272 population screening. The monoclonal antibodies against SARS-CoV have been successfully

applied in the immunological antigen-detection of SARS-CoV [14]. Accordingly, we explored the immunodominant fragments on the spike protein of SARS-CoV-2, which would provide aid in developing an accurate and fast antigen-capture based early detection system for SARS-CoV-2.

We selected the S1 subunit for immunodominant fragments screening after divergence analysis. It had been reported that an S1 antigen-detected assay of SARS-CoV could detect the virus as soon as the infection occurs [27]. Jong-Hwan Lee *et al.* designed a method which captures and detects spike protein S1 subunit of SARS-CoV-2 using ACE2 receptor and S1-mAb[63], which suggests that it is appropriate to use the S1 subunit for a specific and early diagnosis of SARS-CoV-2. Three immunodominant fragments ($Spike_{56-94}$, $Spike_{199-264}$, and $Spike_{577-612}$) were obtained in our study. These sequences will be joined to construct recombinant peptides in the following step. The results of sequence BLAST suggested that these selected fragments have a high specificity. Instead of using inactivated full-length spike protein, we designed a novel recombinant protein construct which increased sequence specificity as well as circumvented mutation sites and glycosylation sites. As the antigen design is totally based on bioinformatics study, the exact ability of the selected fragments to produce qualified antibodies for virus detection has yet to be identified by experiments.

Noticeably, the spike protein of SARS-CoV-2 is heavily glycosylated. Glycans could shield epitopes during antibody recognition, and it may hinder the detection of viral proteins [56]. The spike protein possessed 22 potential N-glycosylation sites along with two O-glycosylation sites. Among 22 potential sites, about 17 N-glycosylation sites were found occupied [56]. We

deliberately circumnavigated glycosylation sites when selecting immunodominant fragments, but several sites are preserved because they are located on the predicted epitopes. The three selected fragments in this study still contain 4 glycosylation sites: Fragment $Spike_{577-612}$ does not contain glycosylation sites. N-glycosylation sites Asn^{61} and Asn^{74} located at $Spike_{56-94}$, and Asn^{234} located at $Spike_{199-264}$. To retain some epitopes with high antigenicity, the fragments inevitably contained 3 glycosylation sites. In case these glycosylation sites do affect the diagnostic performance, an additional deglycosylation step by N-glycanase could be applied for the test specimens [64], which is a simple and efficient method for deglycosylation and is widely used in multiple types of research [65-68]. Alternatively, a eukaryotic expressing system could be employed to mimic the antigen presented in human cells.

Coronaviruses have the ability to correct errors during replication and recombination [58]. while the SARS-CoV-2 genome still presents many mutations. Many of these mutations could spoil the proceeding effort in the development of diagnostic tests [69]. For diagnostic tests, the mutations on the antigen are roadblocks in developing effective diagnostic tests against COVID-19. In this study, we circumvented high-frequency mutation sites when selecting antigen fragments. In addition, our fragments also avoided RBD regions which are prone to mutation [69]. The construct finally built contained no high-frequency mutation.

To date, several studies using predictive algorithms to analyze SARS-CoV-2 have been reported. However, most of them intended to design vaccines and focused on the homology part of the virus sequence[70-72]. On the contrary, we aimed to detect the virus, more attention was paid to the sections with large variability. Therefore, our immunodominant fragments are more

specific. In addition, unlike vaccine researches using humans as host species, we predicted murine MHC-II binding T-cell epitopes to make sure the fragments could trigger a humoral immune response in mice. Nevertheless, our conclusion is based on *in silico* calculations. The efficiency needs to be evaluated by *in vitro* and *in vivo* experiments.

Conclusion

Through bioinformatics analysis, three immunodominant fragments were obtained in the present study. After connected by flexible linkers, we acquired a final recombinant peptide with a 194 aa length. It was predicted to have high antigenicity and possess specificity for SARS-CoV-2. Our next move is to express and purify the recombinant protein in a suitable expression system, followed by immunizing the mice with purified immunogen to obtain specific antibodies. The present study would provide aid in the development of an antigen-capture based detection system.

Ethics approval and consent to participate

Not applicable

Author Contributions:

Siqi Zhuang: Conception; Methodology; Software operation; Data analysis; Visualization; Writing of the manuscript; Review and modification of the manuscript.

Hongzhi Chen: Conception; Methodology; Visualization; Review and modification of the manuscript; Supervision and instruction; Project administration.

Lingli Tang: Conception; Methodology; Visualization; Review and modification of the manuscript; Supervision and instruction; Project administration; Funding acquisition.

Yufeng Dai: Visualization; Review and modification of the manuscript.

All authors have read and agreed to the published version of the manuscript.

Appendix A. Supplementary data

The following are the supplementary data to this article:

Fig. S1. The secondary structure presentation of SARS-CoV2 spike using PyMOL server

Table S1. Linear B-cell epitopes predicted by the ABCpred server along with their position, sequence, ABCpred prediction score, and antigenicity score.

Table S2. Linear B-cell epitopes predicted by the Bepipred v2.0 server along with their position, sequence, and antigenicity score.

Table S3. MHC-I binding epitopes predicted by TepiTool along with their position, sequence, allele, and antigenicity score.

Table S4. MHC-II binding epitopes predicted by TepiTool along with their position, sequence, allele, and antigenicity score.

Reference

[1] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, X. Wang, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor, Nature (2020). 10.1038/s41586-020-2180-5.
[2] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, F. Li, Structural basis of receptor recognition by SARS-CoV-2, Nature (2020). 10.1038/s41586-020-2179-y.
[3] S.H. Zhan, B.E. Deverman, Y.A. Chan, (2020). 10.1101/2020.05.01.073262.
[4] J. Shaman, M. Galanti, Will SARS-CoV-2 become endemic?, Science 370 (2020) 527-529. 10.1126/science.abe5960.
[5] C. Rothe, M. Schunk, P. Sothmann, G. Bretzel, G. Froeschl, C. Wallrauch, T. Zimmer, V. Thiel, C. Janke, W. Guggemos, M. Seilmaier, C. Drosten, P. Vollmar, K. Zwirglmaier, S. Zange, R. Wolfel, M. Hoelscher, Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany, N Engl J Med 382 (2020) 970-971. 10.1056/NEJMc2001468.
[6] Y. Yan, L. Chang, L. Wang, Laboratory testing of SARS-CoV, MERS-CoV, and SARS-CoV-2 (2019-nCoV): Current status, challenges, and countermeasures, Rev Med Virol 30 (2020) e2106. 10.1002/rmv.2106.
[7] T. Ishige, S. Murata, T. Taniguchi, A. Miyabe, K. Kitamura, K. Kawasaki, M. Nishimura, H. Igari, K. Matsushita, Highly sensitive detection of SARS-CoV-2 RNA by multiplex rRT-PCR for molecular diagnosis of COVID-19 by clinical laboratories, Clin Chim Acta 507 (2020) 139-142. 10.1016/j.cca.2020.04.023.
[8] S.A. Bustin, T. Nolan, RT-qPCR Testing of SARS-CoV-2: A Primer, Int J Mol Sci 21 (2020). 10.3390/ijms21083004.
[9] A. Padoan, C. Cosma, L. Sciacovelli, D. Faggian, M. Plebani, Analytical performances of a chemiluminescence immunoassay for SARS-CoV-2 IgM/IgG and antibody kinetics, Clin Chem Lab Med (2020). 10.1515/cclm-2020-0443.
[10] L. Thabet, S. Mhalla, H. Naija, M.A. Jaoua, N. Hannachi, L. Fki-Berrajah, A. Toumi, H. Karray-Hakim, SARS-CoV-2 infection virological diagnosis, La Tunisie medicale 98 (2020) 304-308.
[11] W. Liu, L. Liu, G. Kou, Y. Zheng, Y. Ding, W. Ni, Q. Wang, L. Tan, W. Wu, S. Tang, Z. Xiong, S. Zheng, Evaluation of Nucleocapsid and Spike Protein-based ELISAs for detecting antibodies against SARS-CoV-2, J Clin Microbiol (2020). 10.1128/JCM.00461-20.
[12] M.S. Tang, K.G. Hock, N.M. Logsdon, J.E. Hayes, A.M. Gronowski, N.W. Anderson, C.W. Farnsworth, Clinical Performance of Two SARS-CoV-2 Serologic Assays, Clin Chem (2020). 10.1093/clinchem/hvaa120.
[13] A. Hachim, N. Kavian, C.A. Cohen, A.W.H. Chin, D.K.W. Chu, C.K.P. Mok, O.T.Y. Tsang, Y.C. Yeung, R. Perera, L.L.M. Poon, J.S.M. Peiris, S.A. Valkenburg, ORF8 and ORF3b antibodies are accurate serological markers of early and late SARS-CoV-2 infection, Nat Immunol 21 (2020) 1293-1301. 10.1038/s41590-020-0773-7.
[14] K. Ohnishi, Establishment and characterization of monoclonal antibodies against SARS coronavirus, Methods Mol Biol 454 (2008) 191-203. 10.1007/978-1-59745-181-9_15.
[15] Q. He, Q. Du, S. Lau, I. Manopo, L. Lu, S.W. Chan, B.J. Fenner, J. Kwang, Characterization of monoclonal antibody against SARS coronavirus nucleocapsid antigen and development of an antigen capture ELISA, J Virol Methods 127 (2005) 46-53. 10.1016/j.jviromet.2005.03.004.
[16] L.W. Qiu, H.W. Tang, Y.D. Wang, J.E. Liao, W. Hao, K. Wen, X.M. He, X.Y. Che, [Development and application of triple antibodies-based sandwich ELISA for detecting nucleocapsid protein of SARS-associated coronavirus], Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi 26 (2005) 277-281.
[17] D.A. Muller, A.C. Depelsenaire, P.R. Young, Clinical and Laboratory Diagnosis of Dengue Virus Infection, The Journal of infectious diseases 215 (2017) S89-s95. 10.1093/infdis/jiw649.

[18] K. Ohnishi, M. Sakaguchi, T. Kaji, K. Akagawa, T. Taniyama, M. Kasai, Y. Tsunetsugu-Yokota, M. Oshima, K. Yamamoto, N. Takasuka, S. Hashimoto, M. Ato, H. Fujii, Y. Takahashi, S. Morikawa, K. Ishii, T. Sata, H. Takagi, S. Itamura, T. Odagiri, T. Miyamura, I. Kurane, M. Tashiro, T. Kurata, H. Yoshikura, T. Takemori, Immunological detection of severe acute respiratory syndrome coronavirus by monoclonal antibodies, Jpn J Infect Dis 58 (2005) 88-94.

[19] J. Fung, S.K.P. Lau, P.C.Y. Woo, Antigen Capture Enzyme-Linked Immunosorbent Assay for Detecting Middle East Respiratory Syndrome Coronavirus in Humans, Methods Mol Biol 2099 (2020) 89-97. 10.1007/978-1-0716-0211-9_7.

[20] Y. Ji, W. Guo, L. Zhao, H. Li, G. Lu, Z. Wang, G. Wang, C. Liu, W. Xiang, Development of an antigen-capture ELISA for the detection of equine influenza virus nucleoprotein, J Virol Methods 175 (2011) 120-124. 10.1016/j.jviromet.2011.04.016.

[21] Y.T. Chen, Z. Tsao, S.T. Chang, R.H. Juang, L.C. Wang, C.M. Chang, C.H. Wang, Development of an antigen-capture enzyme-linked immunosorbent assay using monoclonal antibodies for detecting H6 avian influenza viruses, J Microbiol Immunol Infect 45 (2012) 243-247. 10.1016/j.jmii.2011.11.012.

[22] K. Ohnishi, Y. Takahashi, N. Kono, N. Nakajima, F. Mizukoshi, S. Misawa, T. Yamamoto, Y.Y. Mitsuki, S. Fu, N. Hirayama, M. Ohshima, M. Ato, T. Kageyama, T. Odagiri, M. Tashiro, K. Kobayashi, S. Itamura, Y. Tsunetsugu-Yokota, Newly established monoclonal antibodies for immunological detection of H5N1 influenza virus, Jpn J Infect Dis 65 (2012) 19-27.

[23] L.H. Cazares, R. Chaerkady, S.H. Samuel Weng, C.C. Boo, R. Cimbro, H.E. Hsu, S. Rajan, W. Dall'Acqua, L. Clarke, K. Ren, P. McTamney, N. Kallewaard-LeLay, M. Ghaedi, Y. Ikeda, S. Hess, Development of a Parallel Reaction Monitoring Mass Spectrometry Assay for the Detection of SARS-CoV-2 Spike Glycoprotein and Nucleoprotein, Anal Chem 92 (2020) 13813-13821. 10.1021/acs.analchem.0c02288.

[24] P.C. Woo, S.K. Lau, B.H. Wong, H.W. Tsoi, A.M. Fung, R.Y. Kao, K.H. Chan, J.S. Peiris, K.Y. Yuen, Differential sensitivities of severe acute respiratory syndrome (SARS) coronavirus spike polypeptide enzyme-linked immunosorbent assay (ELISA) and SARS coronavirus nucleocapsid protein ELISA for serodiagnosis of SARS coronavirus pneumonia, J Clin Microbiol 43 (2005) 3054-3058. 10.1128/JCM.43.7.3054-3058.2005.

[25] A.R. Fehr, S. Perlman, Coronaviruses: an overview of their replication and pathogenesis, Methods Mol Biol 1282 (2015) 1-23. 10.1007/978-1-4939-2438-7_1.

[26] S. Kumar, V.K. Maurya, A.K. Prasad, M.L.B. Bhatt, S.K. Saxena, Structural, glycosylation and antigenic variation between 2019 novel coronavirus (2019-nCoV) and SARS coronavirus (SARS-CoV), Virusdisease 31 (2020) 13-21. 10.1007/s13337-020-00571-5.

[27] H.H. Sunwoo, A. Palaniyappan, A. Ganguly, P.K. Bhatnagar, D. Das, A.O. El-Kadi, M.R. Suresh, Quantitative and sensitive detection of the SARS-CoV spike protein using bispecific monoclonal antibody-based enzyme-linked immunoassay, J Virol Methods 187 (2013) 72-78. 10.1016/j.jviromet.2012.09.006.

[28] L. Lu, I. Manopo, B.P. Leung, H.H. Chng, A.E. Ling, L.L. Chee, E.E. Ooi, S.W. Chan, J. Kwang, Immunological characterization of the spike protein of the severe acute respiratory syndrome coronavirus, J Clin Microbiol 42 (2004) 1570-1576. 10.1128/jcm.42.4.1570-1576.2004.

[29] N. Gomez, C. Carrillo, J. Salinas, F. Parra, M.V. Borca, J.M. Escribano, Expression of immunogenic glycoprotein S polypeptides from transmissible gastroenteritis coronavirus in transgenic plants, Virology 249 (1998) 352-358. 10.1006/viro.1998.9315.

[30] J. Chen, H. Zhu, P.W. Horby, Q. Wang, J. Zhou, H. Jiang, L. Liu, T. Zhang, Y. Zhang, X. Chen, X. Deng, B. Nikolay,

W. Wang, S. Cauchemez, Y. Guan, T.M. Uyeki, H. Yu, Specificity, kinetics and longevity of antibody responses to avian influenza A(H7N9) virus infection in humans, J Infect 80 (2020) 310-319. 10.1016/j.jinf.2019.11.024.
[31] C.M. Sanchez, A. Izeta, J.M. Sanchez-Morgado, S. Alonso, I. Sola, M. Balasch, J. Plana-Duran, L. Enjuanes, Targeted recombination demonstrates that the spike gene of transmissible gastroenteritis coronavirus is a determinant of its enteric tropism and virulence, J Virol 73 (1999) 7607-7618.
[32] P. Callebaut, L. Enjuanes, M. Pensaert, An adenovirus recombinant expressing the spike glycoprotein of porcine respiratory coronavirus is immunogenic in swine, The Journal of general virology 77 (Pt 2) (1996) 309-313. 10.1099/0022-1317-77-2-309.
[33] Y.J. Tan, P.Y. Goh, B.C. Fielding, S. Shen, C.F. Chou, J.L. Fu, H.N. Leong, Y.S. Leo, E.E. Ooi, A.E. Ling, S.G. Lim, W. Hong, Profiles of antibody responses against severe acute respiratory syndrome coronavirus recombinant proteins and their potential use as diagnostic markers, Clinical and diagnostic laboratory immunology 11 (2004) 362-371. 10.1128/cdli.11.2.362-371.2004.
[34] R. Wang, Y. Hozumi, C. Yin, G.-W. Wei, Mutations on COVID-19 diagnostic targets, Genomics 112 (2020) 5204-5213. 10.1016/j.ygeno.2020.09.028.
[35] B. Meyer, C. Drosten, M.A. Muller, Serological assays for emerging coronaviruses: challenges and pitfalls, Virus Res 194 (2014) 175-183. 10.1016/j.virusres.2014.03.018.
[36] F. Mu, D. Niu, J. Mu, B. He, W. Han, B. Fan, S. Huang, Y. Qiu, B. You, W. Chen, The expression and antigenicity of a truncated spike-nucleocapsid fusion protein of severe acute respiratory syndrome-associated coronavirus, BMC microbiology 8 (2008) 207. 10.1186/1471-2180-8-207.
[37] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol Syst Biol 7 (2011) 539. 10.1038/msb.2011.75.
[38] S.B. Needleman, C.D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins, J Mol Biol 48 (1970) 443-453. 10.1016/0022-2836(70)90057-4.
[39] S. Saha, G.P. Raghava, Prediction of continuous B-cell epitopes in an antigen using recurrent neural network, Proteins 65 (2006) 40-48. 10.1002/prot.21078.
[40] M.C. Jespersen, B. Peters, M. Nielsen, P. Marcatili, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, Nucleic Acids Res 45 (2017) W24-W29. 10.1093/nar/gkx346.
[41] S. Paul, J. Sidney, A. Sette, B. Peters, TepiTool: A Pipeline for Computational Prediction of T Cell Epitope Candidates, Curr Protoc Immunol 114 (2016) 18 19 11-18 19 24. 10.1002/cpim.12.
[42] T. Trolle, I.G. Metushi, J.A. Greenbaum, Y. Kim, J. Sidney, O. Lund, A. Sette, B. Peters, M. Nielsen, Automated benchmarking of peptide-MHC class I binding predictions, Bioinformatics 31 (2015) 2174-2181. 10.1093/bioinformatics/btv123.
[43] P. Wang, J. Sidney, Y. Kim, A. Sette, O. Lund, M. Nielsen, B. Peters, Peptide binding predictions for HLA DR, DP and DQ molecules, BMC Bioinformatics 11 (2010) 568. 10.1186/1471-2105-11-568.
[44] L. Zhang, K. Udaka, H. Mamitsuka, S. Zhu, Toward more accurate pan-specific MHC-peptide binding prediction: a review of current methods and tools, Brief Bioinform 13 (2012) 350-364. 10.1093/bib/bbr060.
[45] S. Yuan, H.C.S. Chan, S. Filipek, H. Vogel, PyMOL and Inkscape Bridge the Data and the Data Visualization, Structure 24 (2016) 2041-2042. 10.1016/j.str.2016.11.012.
[46] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, BMC Bioinformatics 8 (2007) 4. 10.1186/1471-2105-8-4.

[47] M.R. Wilkins, E. Gasteiger, A. Bairoch, J.C. Sanchez, K.L. Williams, R.D. Appel, D.F. Hochstrasser, Protein identification and analysis tools in the ExPASy server, Methods Mol Biol 112 (1999) 531-552. 10.1385/1-59259-584-7:531.

[48] B. Petersen, T.N. Petersen, P. Andersen, M. Nielsen, C. Lundegaard, A generic method for assignment of reliability scores applied to solvent accessibility predictions, BMC Struct Biol 9 (2009) 51. 10.1186/1472-6807-9-51.

[49] I. Dimitrov, I. Bangov, D.R. Flower, I. Doytchinova, AllerTOP v.2--a server for in silico prediction of allergens, J Mol Model 20 (2014) 2278. 10.1007/s00894-014-2278-5.

[50] S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, C. Open Source Drug Discovery, G.P. Raghava, In silico approach for predicting toxicity of peptides and proteins, PLoS One 8 (2013) e73957. 10.1371/journal.pone.0073957.

[51] S. Su, G. Wong, W. Shi, J. Liu, A.C.K. Lai, J. Zhou, W. Liu, Y. Bi, G.F. Gao, Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses, Trends Microbiol 24 (2016) 490-502. 10.1016/j.tim.2016.03.003.

[52] N. Kin, F. Miszczak, W. Lin, M.A. Gouilh, A. Vabret, E. Consortium, Genomic Analysis of 15 Human Coronaviruses OC43 (HCoV-OC43s) Circulating in France from 2001 to 2013 Reveals a High Intra-Specific Diversity with New Recombinant Genotypes, Viruses 7 (2015) 2358-2377. 10.3390/v7052358.

[53] E.D. Getzoff, J.A. Tainer, R.A. Lerner, H.M. Geysen, The chemistry and mechanism of antibody binding to protein antigens, Advances in immunology 43 (1988) 1-98. 10.1016/s0065-2776(08)60363-6.

[54] S.H. Cho, A.L. Raybuck, J. Blagih, E. Kemboi, V.H. Haase, R.G. Jones, M.R. Boothby, Hypoxia-inducible factors in CD4(+) T cells promote metabolism, switch cytokine secretion, and T cell help in humoral immunity, Proceedings of the National Academy of Sciences of the United States of America 116 (2019) 8975-8984. 10.1073/pnas.1811702116.

[55] B.P. Mahon, K. Katrak, A. Nomoto, A.J. Macadam, P.D. Minor, K.H. Mills, Poliovirus-specific CD4+ Th1 clones with both cytotoxic and helper activity mediate protective humoral immunity against a lethal poliovirus infection in transgenic mice expressing the human poliovirus receptor, J Exp Med 181 (1995) 1285-1292. 10.1084/jem.181.4.1285.

[56] A. Shajahan, N.T. Supekar, A.S. Gleinich, P. Azadi, Deducing the N- and O- glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2, Glycobiology (2020). 10.1093/glycob/cwaa042.

[57] A.C. Walls, X. Xiong, Y.J. Park, M.A. Tortorici, J. Snijder, J. Quispe, E. Cameroni, R. Gopal, M. Dai, A. Lanzavecchia, M. Zambon, F.A. Rey, D. Corti, D. Veesler, Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion, Cell 176 (2019) 1026-1039.e1015. 10.1016/j.cell.2018.12.028.

[58] R. Wang, Y. Hozumi, C. Yin, G.W. Wei, Decoding SARS-CoV-2 Transmission and Evolution and Ramifications for COVID-19 Diagnosis, Vaccine, and Medicine, J Chem Inf Model (2020). 10.1021/acs.jcim.0c00501.

[59] V. Chauhan, T. Rungta, K. Goyal, M.P. Singh, Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach, Sci Rep 9 (2019) 2517. 10.1038/s41598-019-39299-8.

[60] H. Ghaffari-Nazari, J. Tavakkol-Afshari, M.R. Jaafari, S. Tahaghoghi-Hajghorbani, E. Masoumi, S.A. Jalali, Improving Multi-Epitope Long Peptide Vaccine Potency by Using a Strategy that Enhances CD4+ T Help in BALB/c Mice, PLoS One 10 (2015) e0142563. 10.1371/journal.pone.0142563.

[61] J. Alexander, M.F. del Guercio, A. Maewal, L. Qiao, J. Fikes, R.W. Chesnut, J. Paulson, D.R. Bundle, S. DeFrees, A. Sette, Linear PADRE T helper epitope and carbohydrate B cell epitope conjugates induce specific high titer IgG antibody responses, Journal of immunology (Baltimore, Md. : 1950) 164 (2000) 1625-1633. 10.4049/jimmunol.164.3.1625.

[62] X. Chen, J.L. Zaro, W.C. Shen, Fusion protein linkers: property, design and functionality, Adv Drug Deliv Rev 65 (2013) 1357-1369. 10.1016/j.addr.2012.09.039.
[63] J.H. Lee, M. Choi, Y. Jung, S.K. Lee, C.S. Lee, J. Kim, J. Kim, N.H. Kim, B.T. Kim, H.G. Kim, A novel rapid detection for SARS-CoV-2 spike 1 antigens using human angiotensin converting enzyme 2 (ACE2), Biosens Bioelectron 171 (2021) 112715. 10.1016/j.bios.2020.112715.
[64] F.K. Dermani, P. Samadi, G. Rahmani, A.K. Kohlan, R. Najafi, PD-1/PD-L1 immune checkpoint: Potential target for cancer therapy, J Cell Physiol 234 (2019) 1313-1325. 10.1002/jcp.27172.
[65] K. Zheng, C. Bantog, R. Bayer, The impact of glycosylation on monoclonal antibody conformation and stability, mAbs 3 (2011) 568-576. 10.4161/mabs.3.6.17922.
[66] E. Lattová, J. Bryant, J. Skřičková, Z. Zdráhal, M. Popovič, Efficient Procedure for N-Glycan Analyses and Detection of Endo H-Like Activity in Human Tumor Specimens, J Proteome Res 15 (2016) 2777-2786. 10.1021/acs.jproteome.6b00346.
[67] J. Huang, H. Wan, Y. Yao, J. Li, K. Cheng, J. Mao, J. Chen, Y. Wang, H. Qin, W. Zhang, M. Ye, H. Zou, Highly Efficient Release of Glycopeptides from Hydrazide Beads by Hydroxylamine Assisted PNGase F Deglycosylation for N-Glycoproteome Analysis, Anal Chem 87 (2015) 10199-10204. 10.1021/acs.analchem.5b02669.
[68] S. Hirani, R.J. Bernasconi, J.R. Rasmussen, Use of N-glycanase to release asparagine-linked oligosaccharides for structural analysis, Anal Biochem 162 (1987) 485-492. 10.1016/0003-2697(87)90424-6.
[69] J. Chen, R. Wang, M. Wang, G.W. Wei, Mutations Strengthened SARS-CoV-2 Infectivity, J Mol Biol 432 (2020) 5212-5226. 10.1016/j.jmb.2020.07.009.
[70] H.Z. Chen, L.L. Tang, X.L. Yu, J. Zhou, Y.F. Chang, X. Wu, Bioinformatics analysis of epitope-based vaccine design against the novel SARS-CoV-2, Infect Dis Poverty 9 (2020) 88. 10.1186/s40249-020-00713-3.
[71] B. Robson, Computers and viral diseases. Preliminary bioinformatics studies on the design of a synthetic vaccine and a preventative peptidomimetic antagonist against the SARS-CoV-2 (2019-nCoV, COVID-19) coronavirus, Comput Biol Med 119 (2020) 103670. 10.1016/j.compbiomed.2020.103670.
[72] M. Bhattacharya, A.R. Sharma, P. Patra, P. Ghosh, G. Sharma, B.C. Patra, S.S. Lee, C. Chakraborty, Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach, J Med Virol 92 (2020) 618-631. 10.1002/jmv.25736.

Figure 1

Work flow chart

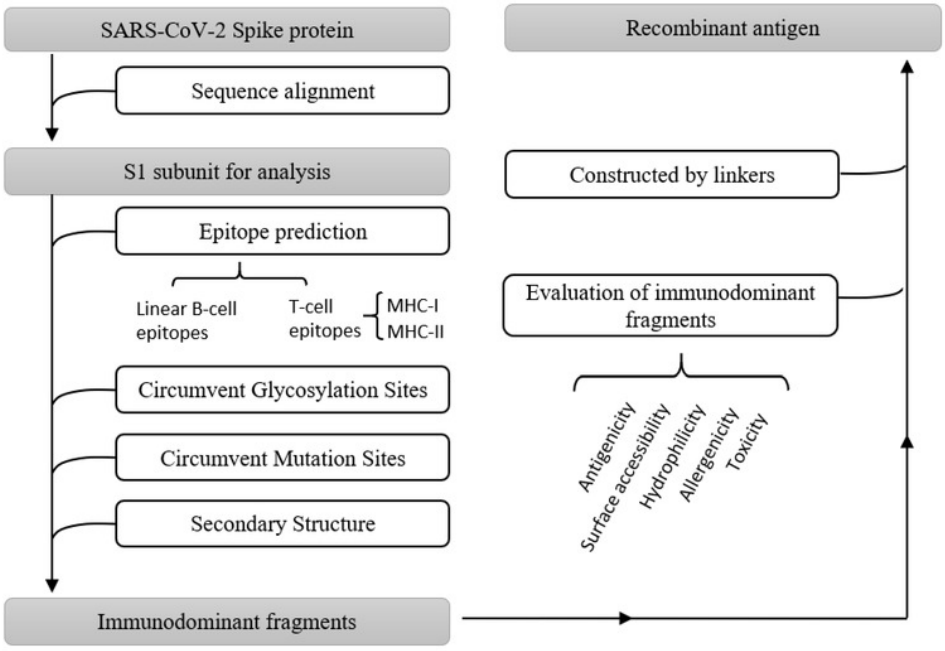


Fig. 1. Work flow chart

Figure 2

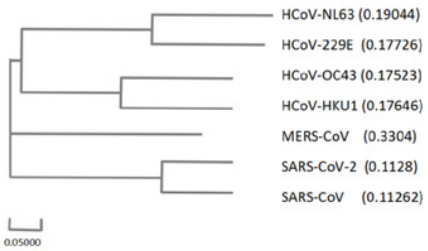
Sequence alignment results of spike protein

A. Accession IDs and sequence identities of selected coronavirus spike protein. B. Phylogenetic tree of spike proteins among selected coronavirus. C. Sequence identity of major domains in spike protein between SARS-CoV-2 and SARS-CoV. D. Sequence identity of domains in SARS-CoV-2 and SARS-CoV reflected by colors. From red to green, the color changing represents the sequence identity from high to low.

A

Virus	Accession ID	Identity compared to SARS-CoV-2(%)
SARS-CoV-2	QIC53213.1	100
SARS-COV	ABD72968.1	77.46
MERS-COV	QBM11748.1	32.04
HCoV-NL63	YP_003767.1	25.57
HCoV-229E	AWH62679.1	26.90
HCoV-OC43	AIX10763.1	30.92
HCoV-HKU1	AMN88694.1	29.92

B



C

Domains	Length	Identity (%)
Spike protein	1277	76.4
S1 subunit	68.9	64.6
NTD	304	50.0
RBM	71	49.3
RBD	222	73.0
S2 subunit	588	90.0

D

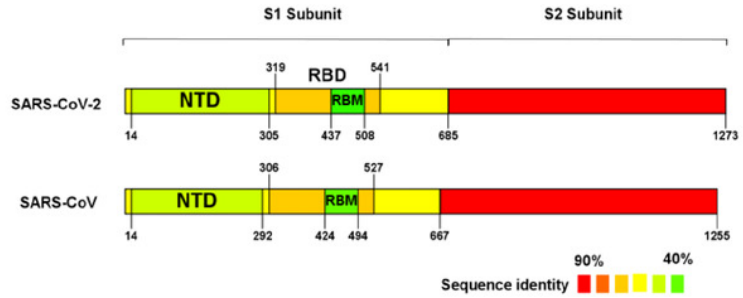


Fig. 2. Sequence alignment results of spike protein

A. Accession IDs and sequence identities of selected coronavirus spike protein.

B. Phylogenetic tree of spike proteins among selected coronavirus.

C. Sequence identity of major domains in spike protein between SARS-CoV-2 and SARS-CoV.

D. Sequence identity of domains in SARS-CoV-2 and SARS-CoV reflected by colors. From red to green, the color changing represents the sequence identity from high to low.

Figure 3

Preliminary immunodominant fragments based on B-cell epitope prediction results

The black squares represent epitopes predicted by ABCpred server, the black frames represent epitopes predicted by Bepipred v2.0 server, and the black lines with numbers on both ends represent the preliminary candidate immunodominant fragments.

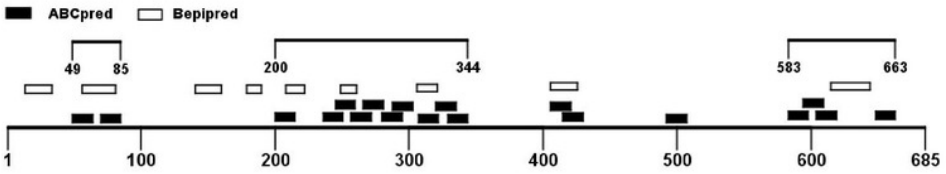


Fig. 3. Preliminary immunodominant fragments based on B-cell epitope prediction results
The black squares represent epitopes predicted by ABCpred server, the black frames represent epitopes predicted by Bepipred v2.0 server, and the black lines with numbers on both ends represent the preliminary candidate immunodominant fragments.

Figure 4

Adjusted candidate immunodominant fragments according to MHC-II T-cell epitope prediction results

The black squares represent epitopes predicted by ABCpred server, and the black frames represent epitopes predicted by Bepipred v2.0 server. The red frames denote MHC-II binding epitopes. The black lines with numbers on both ends represent the adjusted candidate fragments.

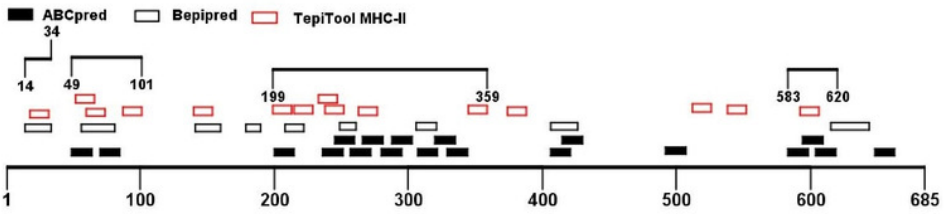


Fig. 4. Adjusted candidate immunodominant fragments according to MHC-II T-cell epitope prediction results

The black squares represent epitopes predicted by ABCpred server, and the black frames represent epitopes predicted by Bepipred v2.0 server. The red frames denote MHC-II binding epitopes. The black lines with numbers on both ends represent the adjusted candidate fragments.

Figure 5

The epitopes and glycosylation sites on the selected immunodominant fragments

The black squares represent epitopes predicted by ABCpred server, and the black frames represent epitopes predicted by Bepipred v2.0 server. The red squares represent MHC-I binding epitopes, and the red frames represent MHC-II binding epitopes. The grey squares means occupied glycosylation sites contained in the selected fragments.

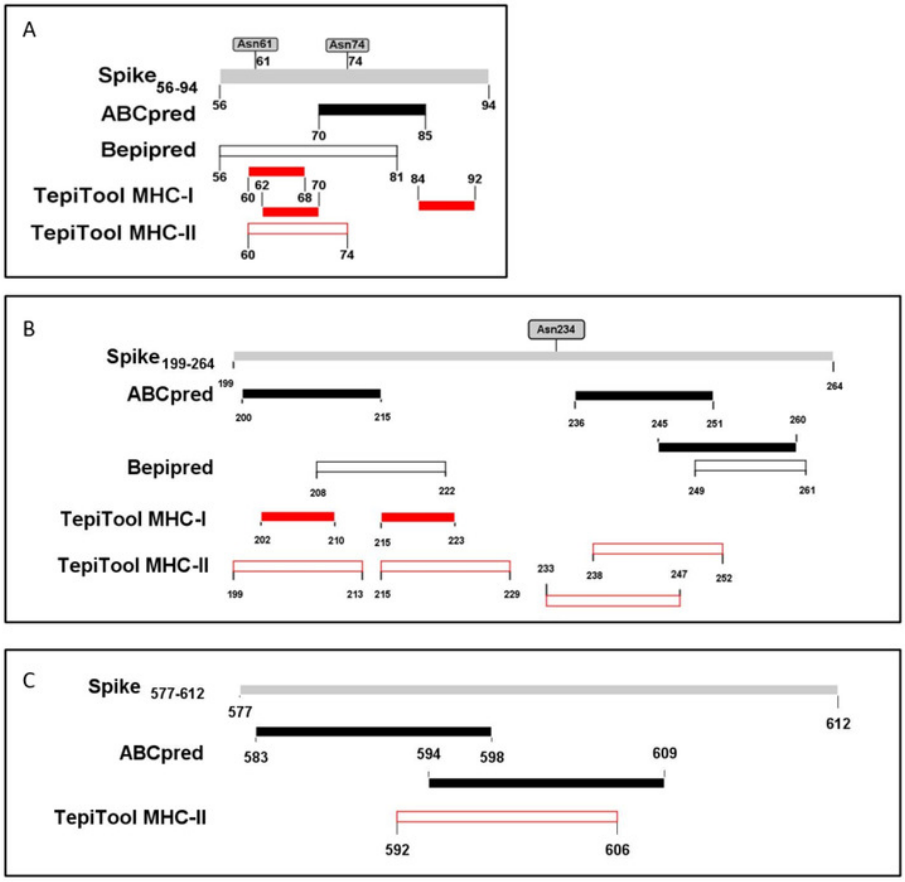


Fig. 5. The epitopes and glycosylation sites on the selected immunodominant fragments
The black squares represent epitopes predicted by ABCpred server, and the black frames represent epitopes predicted by Bepipred v2.0 server. The red squares represent MHC-I binding epitopes, and the red frames represent MHC-II binding epitopes. The grey squares means occupied glycosylation sites contained in the selected fragments.

Figure 6

Selected immunodominant fragments presented as spheres in the trimer structure of spike protein viewed by PyMOL

Selected fragments were presented as red spheres, green cartoons denote unselected sections. A, B, and C denote fragments Spike56-94, Spike199-264, and Spike577-612 respectively.

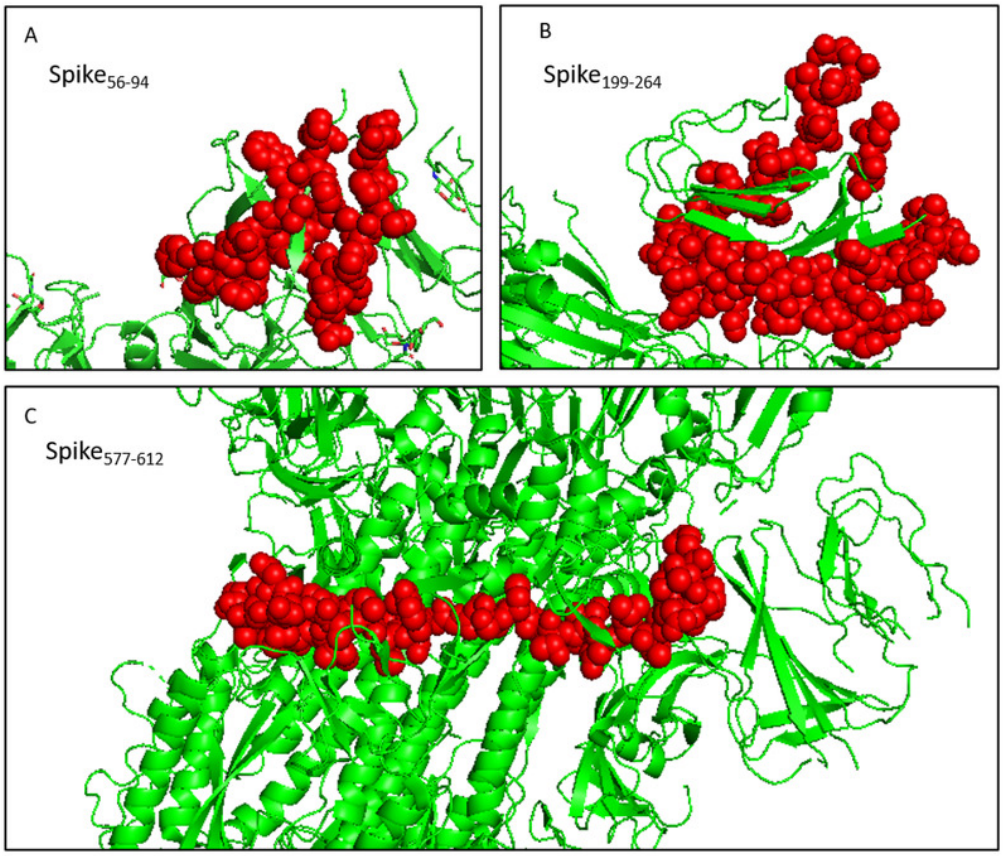


Fig. 6. Selected immunodominant fragments presented as spheres in the trimer structure of spike protein viewed by PyMOL

Selected fragments were presented as red spheres, green cartoons denote unselected sections. A, B, and C denote fragments $Spike_{56\text{-}94}$, $Spike_{199\text{-}264}$, and $Spike_{577\text{-}612}$ respectively.

Figure 7

A schematic diagram of recombinant peptide composed of selected fragments and a PADRE epitope.

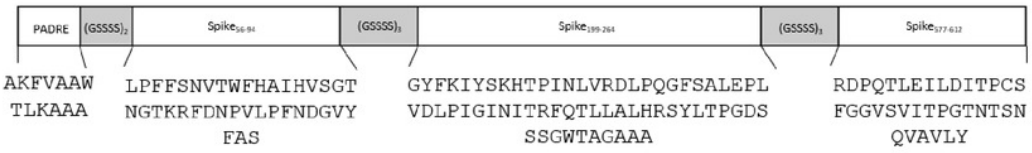


Fig. 7. A schematic diagram of recombinant peptide composed of selected fragments and a PADRE epitope.

Table 1(on next page)

Linear B-cell epitopes predicted by ABCpred and BepiPred v2.0 with antigenicity score exceed the threshold value

1

Tools	Position	Sequence	Length	Antigenicity (cut off≥0.4)
ABCpred	583-598	EILDITPCSFGGVSVI	16	1.3971
	406-421	EVRQIAPGQTGKIADY	16	1.3837
	415-430	TGKIADYNYKLPDDFT	16	0.9642
	648-663	GCLIGAEHVNNSYECD	16	0.848
	288-303	AVDCALDPLSETKCTL	16	0.7905
	604-619	TSNQVAVLYQDVNCTE	16	0.7593
	307-322	TVEKGIYQTSNFRVQP	16	0.6733
	200-215	YFKIYSKHTPINLVRD	16	0.657
	257-272	GWTAGAAAYYVGYLQP	16	0.621
	329-344	FPNITNLCPFGEVFNA	16	0.6058
	245-260	HRSYLTPGDSSSGWTA	16	0.6017
	280-295	NENGTITDAVDCALDP	16	0.5804
	49-64	HSTQDLFLPFFSNVTW	16	0.5305
	492-507	LQSYGFQPTNGVGYQP	16	0.5258
	70-85	VSGTNGTKRFDNPVLP	16	0.5162
	236-251	TRFQTLLALHRSYLTP	16	0.5115
	266-281	YVGYLQPRTFLLKYNE	16	0.5108
	594-609	GVSVITPGTNTSNQVA	16	0.4651
	320-335	VQPTESIVRFPNITNL	16	0.4454
Bepipred v2.0	179-190	LEGKQGNFKNLR	12	1.1188
	404-426	GDEVRQIAPGQTGKIADYNYKLP	23	1.1017
	14-34	QCVNLTTRTQLPPAYTNSFTR	21	0.7594
	56-81	LPFFSNVTWFHAIHVSGTNGTKRFDN	26	0.6041
	208-222	TPINLVRDLPQGFSA	15	0.5531
	141-160	LGVYYHKNNKSWMESEFRVY	20	0.5308
	249-261	LTPGDSSSGWTAG	13	0.495
	306-321	FTVEKGIYQTSNFRVQ	16	0.4361
	615-644	VNCTEVPVAIHADQLTPTWRVYSTGSNVFQ	30	0.4259

Table 2(on next page)

Details of epitopes in the preliminary immunodominant fragments selected according to linear B-cell epitope prediction results.

1

Regions	Epitope predicted by ABCpred			Epitope predicted by Bepipred v2.0		
	Position	Sequence	Antigenicity	Position	Sequence	Antigenicity
49-85	49-64	HSTQDLFLPFFSNVTW	0.5305	56-81	LPFFSNVTWFHAIHV SGTNGTKRFDN	0.6041
	70-85	VSGTNGTKRFDNPVLP	0.5162			
200-344	200-215	YFKIYSKHTPINLVRD	0.6570	208-222	TPINLVRDLPQGFSA	0.5531
	236-251	TRFQTLLALHRSYLTP	0.5115	249-261	LTPGDSSSGWTAG	0.4950
	245-260	HRSYLTPGDSSSGWTA	0.6017			
	257-272	GWTAGAAAYYVGYLQP	0.6210			
	266-281	YVGYLQPRTFLLKYNE	0.5108			
	280-295	NENGTITDAVDCALDP	0.5804			
	288-303	AVDCALDPLSETKCTL	0.7905	306-321	FTVEKGIYQTSNFRV Q	0.4361
	307-322	TVEKGIYQTSNFRVQP	0.6733			
	320-335	VQPTESIVRFPNITNL	0.4454			
	329-344	FPNITNLCPFGEVFNA	0.6058			
	415-430	TGKIADYNYKLPDDFT	0.9642			
583-663	583-598	EILDITPCSFGGVSVI	1.3971	615-644	VNCTEVPVAIHADQL TPTWRVYSTGSNVFQ	0.4259
	594-609	GVSVITPGTNTSNQVA	0.4651			
	604-619	TSNQVAVLYQDVNCTE	0.7593			
	648-663	GCLIGAEHVNNSYECD	0.8480			

Table 3(on next page)

MHC-II and MHC-I binding epitopes predicted by TepiTool server with antigenicity score exceed threshold value

Type	Position	Sequence	Length	Allele	Core (smm-align)	Core (nn-align)	Percentile Rank	Antigenicity (cut off≥0.4
MHC-II binding	538-552	CVNFNFNGLTGTGVL	15	H2-IAb	FNFNGLTGT	FNFNGLTGT	8.55	1.3281
	374-388	FSTFKCYGVSPTKLN	15	H2-IAb	FKCYGVSPT	YGVSPTKLN	6.45	1.0042
	199-213	GYFKIYSKHTPINLV	15	H2-Iab	KIYSKHTPI	YSKHTPINL	6.9	0.9278
	18-32	LTTRTQLPPAYTNSF	15	H2-IAb	TRTQLPPAY	TRTQLPPAY	9.9	0.79
	60-74	SNVTWFHAIHVSGTN	15	H2-IAb	VTWFHAIHV	TWFHAIHVS	9.1	0.7044
	263-277	AAYYVGYLQPRTFLL	15	H2-IAb	VGYLQPRTF	VGYLQPRTF	8.75	0.6073
	592-606	FGGVSVITPGTNTSN	15	H2-IAb	VITPGTNTS	VSVITPGTN	6	0.5825
	238-252	FQTLLALHRSYLTPG	15	H2-IEd	TLLALHRSY	TLLALHRSY	9.85	0.5789
	345-359	TRFASVYAWNRKRIS	15	H2-IAb	FASVYAWNR	YAWNRKRIS	7.45	0.4963
	215-229	DLPQGFSALEPLVDL	15	H2-IAb	FSALEPLVD	FSALEPLVD	6.05	0.4812
	140-154	FLGVYYHKNNKSWME	15	H2-IEd	GVYYHKNNK	YYHKNNKSW	6.4	0.4793
	512-526	VLSFELLHAPATVCG	15	H2-IAb	FELLHAPAT	FELLHAPAT	2.9	0.4784
	87-101	NDGVYFASTEKSNII	15	H2-Iab	YFASTEKSN	VYFASTEKS	6.85	0.4277
	52-66	QDLFLPFFSNVTWFH	15	H2-IAb	FLPFFSNVT	FLPFFSNVT	2.95	0.4159
	233-247	INITRFQTLLALHRS	15	H2-IAd	ITRFQTLLA	ITRFQTLLA	1.9	0.4118
MHC-I binding	643-651	FQTRAGCLI	9	H-2-Kk			0.6	1.7332
	612-620	YQDVNCTEV	9	H-2-Db			0.4	1.6172
	539-547	VNFNFNGLT	9	H-2-Kb			0.47	1.5069
	503-511	VGYQPYRVV	9	H-2-Kb			0.47	1.4383
	379-387	CYGVSPTKL	9	H-2-Kd			0.3	1.4263
	16-24	VNLTTRTQL	9	H-2-Kb			0.86	1.3468
	510-518	VVVLSFELL	9	H-2-Kb			0.43	1.0909
	202-210	KIYSKHTPI	9	H-2-Kb			0.27	0.7455
	168-176	FEYVSQPFL	9	H-2-Kk			0.5	0.6324
	268-276	GYLQPRTFL	9	H-2-Kd			0.2	0.6082
	505-513	YQPYRVVVL	9	H-2-Dd			0.3	0.5964
	488-496	CYFPLQSYG	9	H-2-Kd			0.64	0.578
	215-223	DLPQGFSAL	9	H-2-Dd			0.69	0.5622
	342-350	FNATRFASV	9	H-2-Kb			0.56	0.5609
	84-92	LPFNDGVYF	9	H-2-Ld			0.21	0.5593
	484-492	EGFNCYFPL	9	H-2-Kb			0.84	0.5453
	62-70	VTWFHAIHV	9	H-2-Kb			0.61	0.5426
	489-497	YFPLQSYGF	9	H-2-Dd			0.8	0.5107
	350-358	VYAWNRKRI	9	H-2-Kd			0.7	0.5003
	60-68	SNVTWFHAI	9	H-2-Kb			0.82	0.4892
	262-270	AAAYYVGYL	9	H-2-Kb			0.98	0.4605

1

Table 4(on next page)

Details of candidate immunodominant fragments adjusted according to the MHC-II binding T-cell epitopes prediction results.

1

Regions	Linear B-cell epitopes				MHC-II binding epitopes		
	Tools	Position	Sequence	Antigenicity	Position	Sequence	Antigenicity
14-34	Bepipred v2.0	14-34	QCVNLTTRTQLPPAYTN SFTR	0.7594	18-32	LTTRTQLPPAYTNSF	0.7900
49-101	Bepipred v2.0	56-81	LPFFSNVTWFHAIHVSG TNGTKRFDN	0.6041	52-66	QDLFLPFFSNVTWFH	0.4159
	ABCpred	49-64	HSTQDLFLPFFSNVTW	0.5305	60-74	SNVTWFHAIHVSGTN	0.7044
	ABCpred	70-85	VSGTNGTKRFDNPVLP	0.5162	87-101	NDGVYFASTEKSNII	0.4277
199-359	Bepipred v2.0	208-222	TPINLVRDLPQGFSA	0.5531	199-213	GYFKIYSKHTPINLV	0.9278
		249-261	LTPGDSSSGWTAG	0.4950			
		306-321	FTVEKGIYQTSNFRVQ	0.4361			
	ABCpred	200-215	YFKIYSKHTPINLVRD	0.6570			
		236-251	TRFQTLLALHRSYLTP	0.5115	215-229	DLPQGFSALEPLVDL	0.4812
		245-260	HRSYLTPGDSSSGWTA	0.6017			
		257-272	GWTAGAAAYYVGYLQP	0.6210	233-247	INITRFQTLLALHRS	0.4118
		266-281	YVGYLQPRTFLLKYNE	0.5108			
		280-295	NENGTITDAVDCALDP	0.5804	238-252	FQTLLALHRSYLTPG	0.5789
		288-303	AVDCALDPLSETKCTL	0.7905			
		307-322	TVEKGIYQTSNFRVQP	0.6733	263-277	AAYYVGYLQPRTFLL	0.6073
		320-335	VQPTESIVRFPNITNL	0.4454			
		329-344	FPNITNLCPFGEVFNA	0.6058			
					345-359	TRFASVYAWNRKRIS	0.4963
583-620	ABCpred	583-598	EILDITPCSFGGVSVI	1.3971	592-606	FGGVSVITPGTNTSN	0.5825
		594-609	GVSVITPGTNTSNQVA	0.4651			
		604-619	TSNQVAVLYQDVNCTE	0.7593			

Table 5(on next page)

Significant features of the selected immunodominant fragments. The sequences marked with gray shading in the table represent amino acids with hydrophilicity and surface accessibility respectively.

Fragments	$Spike_{56-94}$	$Spike_{199-264}$	$Spike_{577-612}$
Length(aa)	39	66	36
Sequence	LPFFSNVTWFHAIHVS GTNGTKRFDNPVLPF NDGVYFAS	GYFKIYSKHTPINLVRDLPQGFSALEPLVD LPIGINITRFQTLLALHRSYLTPGDSSSGW TAGAAA	RDPQTLEILDITPCSFG GVSVITPGTNTSNQVA VLY
Antigenicity	0.4590	0.5774	0.9127
Domain	S1(NTD)	S1(NTD)	S1
Hydrophilicity fragments	LPFFSNVTWFHAIHVS GTNGTKRFDNPVLPF NDGVYFAS	GYFKIYSKHTPINLVRDLPQGFSALEPLVD LPIGINITRFQTLLALHRSYLTPGDSSSGW TAGAAA	RDPQTLEILDITPCSFG GVSVITPGTNTSNQVA VLY
Surface Accessibility fragments	LPFFSNVTWFHAIHVS GTNGTKRFDNPVLPF NDGVYFAS	GYFKIYSKHTPINLVRDLPQGFSALEPLVD LPIGINITRFQTLLALHRSYLTPGDSSSGW TAGAAA	RDPQTLEILDITPCSFG GVSVITPGTNTSNQVA VLY
Toxicity	Non-toxin	Non-toxin	Non-toxin
Allergenicity	non-allergen	non-allergen	probable allergen

1

Table 6(on next page)

The structure and antigenicity of final recombinant peptides

Final construct	PAN DR + $(GGGGS)_2$ + $Spike_{56-94}$ +$(GGGGS)_3$ + $Spike_{199-264}$ +$(GGGGS)_3$ + $Spike_{577-612}$
Sequence	AKFVAAWTLKAAAGGGGSGGGGSLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASGGGGSGGGGS GGGGSGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAG GGGSGGGGSGGGGSRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLY
Antigenicity	0.5690

1