Bioinformatic analysis of epitope-based vaccine design against the novel SARS-COV-2

Hongzhi Chen (✉ chenhongzhi2013@csu.edu.cn)

Central South University

Xiang Wu (✉ wxspring@126.com)

Central South University

Short Report

Keywords: 2019-nCoV, Epitope, Vaccine, Bioinformatics

Posted Date: March 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-16262/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Infectious Diseases of Poverty on July 10th, 2020. See the published version at <https://doi.org/10.1186/s40249-020-00713-3>.

Abstract

Background: An outbreak of infection caused by SARS-COV-2 recently has brought great challenge to public health. Rapid identification of immune epitopes would be an efficient way to screen the candidates for vaccine development at the time of an impending pandemic. This study aimed to predict the protective epitopes with bioinformatic methods and resources for vaccine development.

Methods: ABCpred and BepiPred servers were utilized for sequential B-cell epitope analysis. Discontinuous B-cell epitopes were predicted via DISCTOPE server. IEDB server was utilized for HLA-1 and HLA-2 binding peptides computation. Surface accessibility, antigenicity, and other important features of forecasted epitopes were characterized for immunogen potential evaluation.

Results: A total of 63 sequential B-cell epitopes on spike protein were predicted and 3 peptides (333-338, 648-663, 1064-1079) exhibited high antigenicity score and good surface accessibility in our modeling. 8 residues within spike protein (Gly¹⁸⁴, Gly⁵⁰², Lys⁵⁵⁸, Pro⁵⁶¹, Pro⁸⁰⁹, Ser⁸¹⁰, Lys⁸¹¹, Pro⁸¹²) are forecasted as components of discontinuous B-cell epitopes. The bioinformatic analysis of HLA binding peptides within nucleocapsid protein produced 81 and 64 peptides being able to bind class-I and class-II molecule respectively. The peptide (104-112) has a high immunogenicity score and was predicted to bind a wide spectrum of both HLA-1 and HLA-2 molecules.

Conclusions: Linear B-cell epitopes (333-338, 648-663, 1064-1079) and discontinuous B-cell epitopes (Gly¹⁸⁴, Gly⁵⁰², Lys⁵⁵⁸, Pro⁵⁶¹, Pro⁸⁰⁹, Ser⁸¹⁰, Lys⁸¹¹, Pro⁸¹²) on spike protein, T-cell epitope (104-112) within nucleocapsid protein were identified and recommended for developing vaccine against SARS-COV-2.

Background

Since the first new pneumonia patient been diagnosed last December in China[1], an outbreak of infection caused by a novel coronavirus has spread rapidly. To date, 75,569 cases in China and 1,200 cases in 26 countries outside of China have been confirmed (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>), which brought great challenge to public health worldwide. Therefore, it is imminent to prevent and control this infectious disease.

The pathogen causing the new type of pneumonia was named 2019-nCoV (recently with its name changed to SARS-COV-2) by WHO(<https://www.who.int/publications-detail/surveillance-case-definitions-for-human-infection>). Its genome sequence has been released and reported by Chinese scientists and submitted to GenBank database on 12 January 2020[2]. Like SARS-COV and MERS-COV, the other two virus that caused severe epidemic problem in recent years, SARS-COV-2 also belongs to β -coronaviruses family. Bats are proved to be their natural host[3]. At present, insufficient knowledge on the latency and contagiousity of SARS-COV-2 increased the uncertainty of virus persistency. Specific therapeutic agents targeting the virus are currently not available[4]. Vaccine is still the most expecting and effective approach of prevention. Whereas selection and design of protective immunogens against pathogens rapidly is a major challenge on vaccine development, especially for the newly emerging pathogens[5]. Traditional methods of vaccine development could not meet the needs of the pressing situation in the event of an outbreak[6]. Bioinformatic analysis combined with Big Data of immunoinformatic resources makes the mission possible[7–9].

The purpose of this research is to predict B-cell eiptopes on spike protein and T-cell epitopes within nucleocapsid phosphoprotein of SARS-COV-2 by using bioinformatics methods and immunoinformatic resources. Epitopes information presented by this work may aid in developing promising vaccine against SARS-COV-2.

Methods

Data retrieval and sequence alignment

Protein sequences of SARS-COV-2 were retrieved from NCBI database. Alignment was performed on EMBL-EBI server with Clustal Omega method[10]. Conserved domains within predicted polypeptides of SARS-COV-2 were analyzed with CD-search in NCBI.

Structure modeling

Experimentally revealed 3D structure of SARS-COV Spike protein was retrieved from Protein Data Bank (PDB ID: 6ACD)[11]. Structure modeling of SARS-COV-2 Spike protein was performed with SWISS-MODEL Homology Modelling server (<https://swissmodel.expasy.org/>) using SARS-COV Spike protein as template.

Linear B-cell epitope prediction

ABCpred[12] and BepiPred[13] were used for B-cell epitope forecast. Threshold of 0.85 was set for the prediction on ABCpred server and 16 residues lengthy epitopes were presented as adequate to induce defensive immune response. Predicted epitopes were highlighted as sphere in SARS-COV-2

spike protein viewed by pymol. We utilized vaxijen2.0 server to analyze the antigenicity of chosen epitopes[14]. Surface accessibility of predicted peptides were evaluated with the protein 3D structure built with SWISS-MODEL Homology Modelling server.

Discontinuous B-cell epitope prediction

Prediction of discontinuous epitopes were conducted via DiscoTope 1.1 server[15]. Parameter was set at -3.1 which indicate 95% specificity and 16% sensitivity. This method is based on surface accessibility, residue statistics, and spatial information in a compiled data set of discontinuous epitopes discovered by X-ray crystallography of antigen/antibody complex structure. The contact number, propensity score, and discotope score for each amino acid are provided for conformation-based epitope prediction. Pymol was employed to illustrate the position of predicted epitopes on the 3D structure of SARS-COV-2 spike protein.

T-cell epitope prediction

We used the free online service provided by IEDB to forecast T-cell epitopes within nucleocapsid protein binding to HLA-1[16] or HLA-2[17] molecule. A relatively small pool of HLA alleles covering majority of the population, over 97% and 99% for class-I and class-II respectively, were chosen in the analysis[18, 19]. The sequences were given in plain format and top 50 rank percentile results were retained for further analysis.

Profiling and evaluation of predicted T-cell epitopes

Key features including digestion, mutation, toxicity, allergenicity, hydro and physiochemical were analyzed via vaxijen 2.0, protein digest server(<http://db.systemsbio.net:8080/teomicsToolkit/proteinDigest.html>), AllerTOP v2.0 server[20], and ToxinPred server[21]. Immunogenicity of predicted HLA-1 binding peptides were assessed by Class I Immunogenicity service provided on IEDB.

Results

Protein coding features of SARS-COV-2 genome

A map of the predicted ORFs is depicted in Fig. 1 based on the genome sequence of virus of Wuhan-Hu-1 isolate. The genomic structure of SARS-COV-2 shares characteristics that also found in other coronaviruses including SARS-COV, MERS-COV, and Human coronavirus NL63. All these coronaviruses contain recognizable ORFs including the replicase (orf1ab polyprotein), surface glycoprotein (spike protein), envelope protein, membrane glycoprotein, nucleocapsid phosphoprotein, and a number of non-structure proteins (nsp). The conserved domains of proteins encoded by SARS-COV-2 genome are summarized in Supplementary Table 1. Spike protein consists of 1,273 amino acids and the majority part exposed outside of the virion[22]. Thus, it is an ideal target for B-cell epitope screening. Nucleocapsid protein is more conserved in coronaviruses (Fig. 2B-D). Though unable to induce humoral immunity, nucleocapsid protein in SARS-COV and MERS-COV has been experimentally tested as a robust immunogen to induce cytotoxic T lymphocyte response[23, 24], which suggest nucleocapsid protein in SARS-COV-2 could be a good candidate for T-cell epitope prediction.

Sequence analysis of spike protein and nucleocapsid protein in selected coronaviruses

To better understand the characteristic of SARS-COV-2, we compared its protein sequences with other selected coronaviruses. The total sequence identity and guide tree result was presented in Fig. 2B and Fig. 2C respectively. Consistent with recently published study[25], we found that both spike (S protein) and nucleocapsid (N protein) in SARS-COV-2 are more closely related to that of SARS-COV. The protein domains of spike and nucleocapsid protein (Fig. 2D) were depicted based on previous studies on SARS-COV[22, 26] and the protein alignment result in the current study (supplementary files: N align.clustal_num; S align.clustal_num). The amino acid sequence identity result confirmed a highly similarity between SARS-COV-2 and SARS-COV. As anticipated, the nucleocapsid protein is more conserved among selected coronaviruses compared to spike protein.

B-cell epitopes recognition

Full length sequence of spike protein was scanned via ABCpred[12] and BepiPred-2.0 server[13] for putative sequential B-cell epitopes. A threshold value of 0.85 in ABCpred server was set when search for linear B-cell epitope with 16 amino acids in length. A total of 28 non-overlapping peptides were identified by ABCpred server. For sequential B-cell epitopes prediction on BepiPred-2.0 server, a threshold value of 0.5 was applied and 35 peptides were predicted. Antigenicity was calculated by Vaxijen 2.0 server and peptides with highest antigenicity score were selected (Table 1). 3D structure prediction of SARS-COV-2 spike protein was conducted as described in the methods. The new PDB file reflecting SARS-COV-2 spike protein structure could be download from supplementary file SPIK.pdb. Predicted epitopes in Table 1 were highlighted as sphere in pymol (Supplementary Figure 1). While all peptides predicted were exposed on the surface of spike monomer, only seq#2'GCLIGAEHVNNSYECD', seq#5'HVTYVPAQEKNFTTAP', and seq#8'TNLCPF' displayed good surface accessibility in spike trimer (Fig. 3 and supplementary file: B-cell-epitope-on-trimer.gif), the pattern more likely exist in nature. Among these three peptides, seq#8 'TNLCPF' predicted at position 333 possesses a random coil

structure and presented highest antigenicity score. Conformation-based B-cell epitopes were computed on Discotope 1.1 server[15]. A threshold value of -3.1 was chosen for the computation, which corresponds to a specificity of 95% and a sensitivity of 16%. The contact number, propensity score, and disotope score for each amino acid which passed the threshold were presented in Table 2. The position of these residues were viewed with pymol and highlighted as sphere (Fig. 4). Processing with a combination of B-cell epitope scanning and 3D structure modeling forecasted 3 potent linear epitopes and 8 residues involved in discontinuous epitopes formation.

T-cell epitopes recognition

In our study, IEDB server was utilized following prediction methods recommended (a combination of ANN, SMM, CombLib, and NetMHCpan EL methods for HLA-1 binding prediction, and a combination of NN-align, SMM-align, CombLib, Sturniolo, and NetMHCIIpan methods for HLA-2 binding prediction).

For HLA-1 binding peptide prediction, the top 50 rank percentile was retained for further analysis. A total of 81 nonrepetitive peptides with ANN_IC50 value not higher than 500, indicative of stronger than medium binding affinity, were identified. 6 peptides with highest antigenicity score by vaxijen 2.0 were chosen for next step processing. In this step, we screened all HLA-1 molecules being able to bind these peptides (Table 3). Similar strategy was applied for HLA-2 binding peptides prediction on IEDB server and 64 peptides were identified as HLA-2 binding sequences. SMM_IC50 value of 500 and 5000 were indicative of medium and weak binding affinity respectively. 6 peptides with highest antigenicity score were selected for HLA-2 molecule screening and the outcome was presented in Table 4. In the selected peptides pool for HLA binding, 'LSPRWYFYY' was predicted as both HLA class-I and class-II binding peptide. Additionally, this peptide may excel in capability of binding to a large number of HLA molecules as shown in Table 3&4. Partially overlapping region was found in sequence 'FPRGQGVPI' in Table 3 and sequence 'PRGQGVPI' in Table 4, which suggests the overlapping region, like the peptide 'LSPRWYFYY' being able to bind both HLA-1 and HLA-2, may initiate both CD4+ and CD8+ dependent immune response.

Selected T-cell epitopes features profiling and evaluation

Peptide stability, mutation analysis, toxicity, allergenicity, hydro and physiochemical features were calculated as indicated in methods and the results were presented in Supplementary Table 2. While no peptide listed is toxic, a majority of them are potentially allergenic. To forecast the probability of an immune response induced by the predicted HLA-1 binding peptides, Class I Immunogenicity test was performed and the scores were presented in Table 5. A higher score indicates higher potential of immune response induction.

Discussion

Different from the less harmful human coronaviruses continuously circulating in human population, coronaviruses originated from animals could become fatal pathogen by crossing species barriers. SARS-COV in 2003, MERS-COV since 2012, and SARS-COV-2 right now all have caused large-scale epidemic problems. Effective and economic preventive approaches are in need urgently at the current situation of PHEIC (Public Health Emergency of International Concern).

Compared to traditional vaccine development, potent epitopes can be predicted via bioinformatic analysis, which makes the vaccine design straightforward and fast[27]. As the majority of spike protein is exposed outside the virion, it could be an ideal target to search for B-cell epitope. In fact, spike proteins in MERS-COV and SARS-COV have been shown to induce robust immune response[28, 29]. In our analysis of SARS-COV-2 spike protein, 3 peptides (Fig. 3) identified from multi-step screening displayed excellent surface accessibility in the 3D structure modeling. The peptide 'TNLCPF' at position 333 produces a random coil structure and has high antigenicity score. Noticeably, this peptide sits in the receptor binding domain (RBD) of spike in SARS-COV-2, which has been proved to mediate the binding to ACE2 on the epithelial membrane of human lungs[30]. It is likely that antibody recognizing this epitope could also neutralize the virus and prevent infection. We noticed a recent published study of epitope prediction[31]. Though they utilized the same approach in B-cell epitopes prediction, their results are different with the current study which could be due to different parameter setting and strategy. In addition, surface accessibility of potent B-cell epitopes were further evaluated with structure modeling in the current study to exclude B-cell epitopes buried inside.

Besides B-cell epitope prediction on spike protein, we selected nucleocapsid protein as target protein for T-cell epitope computation for the following reasons: First, nucleocapsid protein in SARS-COV and MERS-COV has been experimentally tested as a robust immunogen to induce cytotoxic T lymphocyte response[23, 24]; Second, nucleocapsid protein is the predominant protein expressed in the virion during early stage of infection[32]; Third, nucleocapsid protein was detected in the majority of SARS-COV infected patients as early as day 1 of infection[33, 34], which suggests nucleocapsid protein-based vaccine may evoke T-cell dependent immune response timely. 81 and 64 peptides within nucleocapsid protein were predicted to bind HLA-1 and HLA-2 molecules respectively. We selected 6 peptides (Table 3&4) with the highest antigenicity score as potent T-cell epitopes. The peptide 'LSPRWYFYY' at position 104 showed an immunogenicity score of 0.36, suggesting a relatively strong capability to initiate immune response. Interestingly, this peptide was forecasted to bind to both HLA-1 and HLA-2 molecules, suggesting a priority of this peptide in vaccine design. Though these peptides were predicted to be non-toxic, we noticed that a large number of these peptides could be allergens. Thus, special attention should be

paid to potential allergic reaction during pre-clinical and clinical trial. As nucleocapsid protein is highly conserved between SARS-COV-2 and SARS-COV, available information on nucleocapsid protein-based vaccine against SARS-COV could be helpful.

Conclusions

In summary, specific functional linear B-cell epitopes (648'GCLIGAEHVNNSEYCD', 1064'HVTYVPAQEKNFTTAP', and 333'TNLCPF'), and discontinuous B-cell epitopes (Gly¹⁸⁴, Gly⁵⁰², Lys⁵⁵⁸, Pro⁵⁶¹, Pro⁸⁰⁹, Ser⁸¹⁰, Lys⁸¹¹, Pro⁸¹²) on spike protein and T-cell epitope (104'LSRWYFYF') on nucleocapsid protein of SARS-COV-2 were predicted to be promising in our study, which would aid in the development of epitope-based vaccine against SARS-COV-2.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

All presentation of the data have consent for publication.

Availability of data and materials

All analysed data of this study are included with the manuscript and its supplementary information files.

Competing interests

The authors declare no conflict of interests

Funding

This work was funded by the National Science Foundation of China [No.81371834, No. 81171597, No. 81828004, and No. 81970746], and Hunan Provincial Natural Sciences Foundation [2018JJ3740].

Acknowledgements

We appreciate the generous technical support from Dr. Hezhi Fang and Dr. Ning Zhang.

Authors' contributions

HC and XW performed bioinformatic analysis of epitope prediction. HC wrote the first draft of the manuscript. XW revised the manuscript and was responsible for the final approval of the manuscript.

References

- 1.Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X *et al*: *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet* 2020, *395*(10223):497–506.
- 2.Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R *et al*: *A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med* 2020, *382*(8):727–733.
- 3.Ceraolo C, Giorgi FM: *Genomic variance of the 2019-nCoV coronavirus. J Med Virol* 2020.
- 4.Wu YC, Chen CS, Chan YJ: *Overview of The 2019 Novel Coronavirus (2019-nCoV): The Pathogen of Severe Specific Contagious Pneumonia (SSCP). J Chin Med Assoc* 2020.
- 5.Qiu T, Yang Y, Qiu J, Huang Y, Xu T, Xiao H, Wu D, Zhang Q, Zhou C, Zhang X *et al*: *CE-BLAST makes it possible to compute antigenic similarity for newly emerging pathogens. Nat Commun* 2018, *9*(1):1772.
- 6.Rauch S, Jasny E, Schmidt KE, Petsch B: *New Vaccine Technologies to Combat Outbreak Situations. Front Immunol* 2018, *9*:1963.
- 7.Tahir Ul Qamar M, Bari A, Adeel MM, Maryam A, Ashfaq UA, Du X, Muneer I, Ahmad HI, Wang J: *Peptide vaccine against chikungunya virus: immunoinformatics combined with molecular docking approach. J Transl Med* 2018, *16*(1):298.

- 8.Ahmad B, Ashfaq UA, Rahman MU, Masoud MS, Yousaf MZ: *Conserved B and T cell epitopes prediction of ebola virus glycoprotein for vaccine development: An immuno-informatics approach. Microb Pathog* 2019, 132:243–253.
- 9.Tahir Ul Qamar M, Saleem S, Ashfaq UA, Bari A, Anwar F, Alqahtani S: *Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. J Transl Med* 2019, 17(1):362.
- 10.Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD *et al*: *The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res* 2019, 47(W1):W636-W641.
- 11.Song W, Gui M, Wang X, Xiang Y: *Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. PLoS Pathog* 2018, 14(8):e1007236.
- 12.Saha S, Raghava GP: *Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins* 2006, 65(1):40–48.
- 13.Jespersen MC, Peters B, Nielsen M, Marcatili P: *BepiPred–2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res* 2017, 45(W1):W24-W29.
- 14.Doytchinova IA, Flower DR: *VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics* 2007, 8:4.
- 15.Haste Andersen P, Nielsen M, Lund O: *Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Sci* 2006, 15(11):2558–2567.
- 16.Andreatta M, Nielsen M: *Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics* 2016, 32(4):511–517.
- 17.Nielsen M, Lundegaard C, Lund O: *Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinformatics* 2007, 8:238.
- 18.Weiskopf D, Angelo MA, de Azeredo EL, Sidney J, Greenbaum JA, Fernando AN, Broadwater A, Kolla RV, De Silva AD, de Silva AM *et al*: *Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. Proc Natl Acad Sci U S A* 2013, 110(22):E2046–2053.
- 19.Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A: *Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. Immunogenetics* 2011, 63(6):325–335.
- 20.Dimitrov I, Flower DR, Doytchinova I: *AllerTOP—a server for in silico prediction of allergens. BMC Bioinformatics* 2013, 14 Suppl 6:S4.
- 21.Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP: *Peptide toxicity prediction. Methods Mol Biol* 2015, 1268:143–157.
- 22.Gui M, Song W, Zhou H, Xu J, Chen S, Xiang Y, Wang X: *Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. Cell Res* 2017, 27(1):119–129.
- 23.Buchholz UJ, Bukreyev A, Yang L, Lamirande EW, Murphy BR, Subbarao K, Collins PL: *Contributions of the structural proteins of severe acute respiratory syndrome coronavirus to protective immunity. Proc Natl Acad Sci U S A* 2004, 101(26):9804–9809.
- 24.Veit S, Jany S, Fux R, Sutter G, Volz A: *CD8+ T Cells Responding to the Middle East Respiratory Syndrome Coronavirus Nucleocapsid Protein Delivered by Vaccinia Virus MVA in Mice. Viruses* 2018, 10(12).
- 25.Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, Yuen KY: *Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect* 2020, 9(1):221–236.
- 26.Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH: *The SARS coronavirus nucleocapsid protein—forms and functions. Antiviral Res* 2014, 103:39–50.
- 27.Rappuoli R, Bottomley MJ, D’Oro U, Finco O, De Gregorio E: *Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. J Exp Med* 2016, 213(4):469–481.
- 28.Zhou Y, Jiang S, Du L: *Prospects for a MERS-CoV spike vaccine. Expert Rev Vaccines* 2018, 17(8):677–686.
- 29.Du L, Zhao G, Lin Y, Sui H, Chan C, Ma S, He Y, Jiang S, Wu C, Yuen KY *et al*: *Intranasal vaccination of recombinant adeno-associated virus encoding receptor-binding domain of severe acute respiratory syndrome coronavirus (SARS-CoV) spike protein induces strong mucosal immune responses and provides long-term protection against SARS-CoV infection. J Immunol* 2008, 180(2):948–956.

30.Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL *et al: A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature* 2020.

31.Baruah V, Bose S: *Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. J Med Virol* 2020.

32.Surjit M, Lal SK: *The SARS-CoV nucleocapsid protein: a protein with multifarious activities. Infect Genet Evol* 2008, 8(4):397–405.

33.Che XY, Hao W, Wang Y, Di B, Yin K, Xu YC, Feng CS, Wan ZY, Cheng VC, Yuen KY: *Nucleocapsid protein as early diagnostic marker for SARS. Emerg Infect Dis* 2004, 10(11):1947–1949.

34.Li YH, Li J, Liu XE, Wang L, Li T, Zhou YH, Zhuang H: *Detection of the nucleocapsid protein of severe acute respiratory syndrome coronavirus in serum: comparison with results of other viral markers. J Virol Methods* 2005, 130(1–2):45–50.

Abbreviations

SARS-COV, severe acute respiratory syndrome associated coronavirus; MERS-COV, Middle East respiratory syndrome coronavirus; HLA, human leukocyte antigen; MHC, major histocompatibility complex; WHO, world health organization

Tables

Table 1 B-cell epitope predicted via ABCpred and BepiPred server are presented along with their position and antigenicity scores

ABCpred	Seq#	Position	Epitope sequence	Score	Antigenicity
Spike	1	406-421	EVRQIAPGQTGKIADY	0.85	1.3837
	2	648-663	GCLIGAEHVNNSEYCD	0.90	0.8480
	3	898-913	FAMQMAYRFNGIGVTQ	0.88	1.3096
	4	1058-1073	HGVVFLHVTYVPAQEK	0.89	0.8847
	5	1064-1079	HVTYVPAQEKNF TTAP	0.88	0.8933
	6	1206-1221	YEQYIKWPWYIWLGFI	0.89	0.9510
BepiPred	seq#	Position	Epitope sequence	Antigenicity	
Spike	7	315-324	TSNFRVQPTE	1.3571	
	8	333-338	TNLCPF	1.2508	
	9	372-397	ASFSTFKCYGVSP TKLNDLCFTNVYA	1.2880	
	10	406-426	EVRQIAPGQTGKIADYNYKLP	1.3005	
	11	1037-1045	SKRVDFCGK	1.7321	
	12	1204-1209	GKYEQY	1.2821	

Table 2 Discontinuous B-cell epitopes predicted through DISCOTOPE server

Residue position	Residues name	Contact number	Propensity score	Discotope score
184	GLY	9	1.641	-2.859
502	GLY	10	1.981	-3.019
558	LYS	7	0.513	-2.987
561	PRO	7	0.591	-2.909
809	PRO	12	4.78	-1.22
810	SER	8	3.646	-0.354
811	LYS	11	4.151	-1.349
812	PRO	9	3.958	-0.542

Table 3 HLA class-I alleles binding epitopes predicted by IEDB server

No.	Peptides	HLA class-I alleles	Vaxijen score
1	LSPRWYFYF	HLA-A*0101,HLA-A*3002,HLA-B*5701,HLA-A*1101,HLA-A*2601,HLA-B*5801,HLA-A*2402,HLA-A*3101,HLA-B*3501,HLA-B*1501,HLA-A*0301,HLA-A*6801	1.2832
2	RSRNSSRNS	HLA-A*3001	1.1144
3	IGYYRRATR	HLA-A*3101,HLA-A*3301,HLA-A*6801,HLA-A*0301,HLA-A*3001	0.888
4	FTALTQHGK	HLA-A*6801,HLA-A*1101,HLA-A*0301,HLA-A*0101,HLA-A*3001,HLA-A*3101	0.851
5	KSAAEASKK	HLA-A*0301,HLA-A*1101,HLA-A*3001,HLA-A*6801,HLA-A*3101	0.7679
6	FPRGQGVPI	HLA-B*0702,HLA-B*5101,HLA-B*5301,HLA-B*0801,HLA-B*3501	0.7585

Table 4 HLA class-II alleles binding epitopes predicted by IEDB server

No.	Peptides	HLA class-II alleles	Vaxijen score
1	IKLDDKDPN	DRB1_0701,DRB1_0301,DRB3_0101,DRB1_0405,DRB1_1101,DPA1_0301/DPB1_0402,DQA1_0101/DQB1_0501,DPA1_0103/DPB1_0201,DRB5_0101,DRB4_0101,DRB1_0101,DRB1_1302,DQA1_0501/DQB1_0301,DRB1_0901	2.3118
2	RSGARSKQR	DRB5_0101,DRB4_0101,DQA1_0501/DQB1_0301	1.7874
3	RIGMEVTPS	DRB1_1101,DRB1_0401,DRB1_0405,DQA1_0102/DQB1_0602,DRB1_0802,DRB1_0301,DPA1_0301/DPB1_0402,DRB4_0101,DQA1_0501/DQB1_0301,DRB1_0901,DQA1_0401/DQB1_0402,DRB1_0101,DRB1_0701,DQA1_0501/DQB1_0201,DPA1_0201/DPB1_0101,DRB5_0101,DRB1_1501,DPA1_0103/DPB1_0201	1.5314
4	RGTSPARMA	DQA1_0501/DQB1_0301,DQA1_0102/DQB1_0602,DRB1_0901	1.2953
5	LSPRWYFYF	DRB1_0405,DQA1_0101/DQB1_0501,DPA1_0103/DPB1_0201,DPA1_0201/DPB1_0501,DRB3_0101,DRB1_1201,DPA1_0201/DPB1_0101,DRB1_0901,DQA1_0501/DQB1_0201,DRB1_1501,DRB1_1101,DRB1_0101,DRB1_0401,DPA1_0301/DPB1_0402,DRB5_0101,DRB1_0301,DRB1_0701,DRB1_1302	1.2832
6	PRGQGVPI	DQA1_0501/DQB1_0301,DRB1_1302,DRB1_0101,DRB1_0901,DRB1_0401,DRB1_0701,DRB1_1501,DRB1_0405	1.1707

Table 5 Class-I immunogenicity

No.	Peptides	Class-I Immunogenicity
1	LSPRWYFYF	0.35734
2	RSRNSSRNS	0.1499
3	IGYYRRATR	-0.00164
4	FTALTQHGK	-0.0226
5	KSAAEASKK	-0.07922
6	FPRGQGVPI	-0.26664

Figures

Figure 1

Map of the predicted ORFs in the SARS-CoV-2 genome sequence.

Figure 2

Spike and nucleocapsid protein in selected coronaviruses A. Accession number of spike protein and nucleocapsid protein in selected coronavirus B. Sequence identity of spike protein and nucleocapsid protein among selected coronavirus C. Guide tree from protein sequences alignment result D. Sequence identity of subdomains of spike protein and nucleocapsid protein reflected by color (Red indicates high identity and green indicates low identity)

Figure 3

Site of B cell epitopes predicted in SARS-CoV2 spike protein trimer A. Seq#2'GCLIGAEHVNNSYECD' highlighted in sphere in the protein structure modeling B. Seq#5 'HVTYVPAQEKNFTTAP' highlighted in sphere in the protein structure modeling C. Seq#8 'TNLCPF' highlighted in sphere in the protein structure modeling. Colors of elements presented in the sphere of protein structure modeling: carbon, tint; hydrogen, gray; nitrogen, blue; oxygen, red; sulfur, yellow

Figure 4

Site of B cells discontinuous epitopes predicted through DISCOTOPE 1.1 server on the predicted structure of SARS-CoV2 spike protein highlighted with cartoon representation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Salign.clustalnum](#)
- [Nalign.clustalnum](#)
- [SPIK.pdb](#)
- [Bcellepitopeontrimer.gif](#)
- [SupTab1.PNG](#)
- [SupTab2.PNG](#)
- [SupFig1.PNG](#)