SHORT REPORT   **Open Access**

Bioinformatics analysis of epitope-based vaccine design against the novel SARS-CoV-2

Hong-Zhi Chen[1*], Ling-Li Tang[2], Xin-Ling Yu[3], Jie Zhou[3], Yun-Feng Chang[4*] and Xiang Wu[5*]

Abstract

Background: An outbreak of infection caused by SARS-CoV-2 recently has brought a 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 pandemic. This study aimed to predict the protective epitopes with bioinformatics methods and resources for vaccine development.

Methods: The genome sequence and protein sequences of SARS-CoV-2 were retrieved from the National Center for Biotechnology Information (NCBI) database. ABCpred and BepiPred servers were utilized for sequential B-cell epitope analysis. Discontinuous B-cell epitopes were predicted via DiscoTope 2.0 program. 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 4 peptides (Spike$_{315-324}$, Spike$_{333-338}$, Spike$_{648-663}$, Spike$_{1064-1079}$) exhibited high antigenicity score and good surface accessibility. Ten residues within spike protein (Gly^{496}, Glu^{498}, Pro^{499}, Thr^{500}, Leu^{1141}, Gln^{1142}, Pro^{1143}, Glu^{1144}, Leu^{1145}, Asp^{1146}) are forecasted as components of discontinuous B-cell epitopes. The bioinformatics analysis of HLA binding peptides within nucleocapsid protein produced 81 and 64 peptides being able to bind MHC class I and MHC class II molecules respectively. The peptides (Nucleocapsid$_{66-75}$, Nucleocapsid$_{104-112}$) were predicted to bind a wide spectrum of both HLA-1 and HLA-2 molecules.

Conclusions: B-cell epitopes on spike protein and T-cell epitopes within nucleocapsid protein were identified and recommended for developing a protective vaccine against SARS-CoV-2.

Keywords: SARS-CoV-2, Epitope, Vaccine, Bioinformatics

* Correspondence: chenhongzhi2013@csu.edu.cn;
changyunfeng880@163.com; wxspring@126.com
[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 410011, Hunan, China
[4]Department of Forensic Medicine Science, Xiangya School of Basic
Medicine, Central South University, Changsha 410013, Hunan, China
[5]Department of Parasitology, Xiangya School of Basic Medicine, Central
South University, Changsha 410013, Hunan, China
Full list of author information is available at the end of the article

Background

An outbreak of infection caused by a novel coronavirus has spread worldwide rapidly [1, 2]. The World Health Organization (WHO) named the disease COVID-19, short for "coronavirus disease 2019". As of 27 May 2020, the WHO reported a total of 5 488 825 COVID-19 cases and 349 095 deaths globally [3], which brought a 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 as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, also called 2019-nCoV) by the Coronaviridae Study Group (CSG) of the International Committee on Taxonomy of Viruses [4]. Its genome sequence has been released and reported by Chinese scientists and submitted to the GenBank database on 12 January 2020 [5]. Like severe acute respiratory syndrome associated coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), the other two viruses that caused severe epidemic problems in recent years, SARS-CoV-2 also belongs to β-coronaviruses family. Bats are proved to be their natural host [6]. At present, insufficient knowledge of the latency and contagiosity of SARS-CoV-2 increased the uncertainty of virus persistency. Specific therapeutic agents targeting the virus are currently not available. Vaccination is still the most economic and effective approach to prevent virus infection. The selection and design of protective immunogens against pathogens is a major challenge in vaccine development, especially for the newly emerging
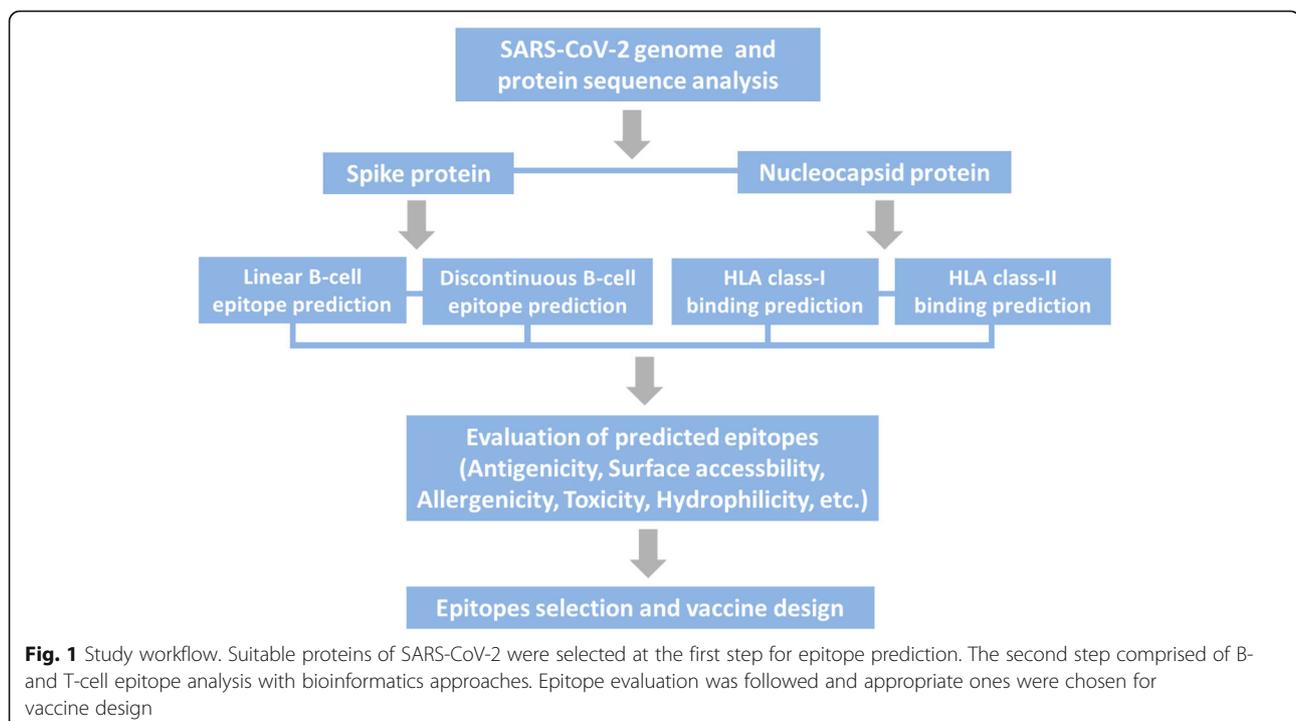
pathogens [7, 8]. Traditional methods based on lab experiments could not meet the needs of the pressing situation in the event of an outbreak [9]. Bioinformatics is an interdisciplinary field specialized in organizing, storing, and processing large amounts of data generated from biological experiments. Accumulation of large-scale immunological data gave rise to the field known as immunoinformatics, which provides insights into the mechanisms of immune function. As the genome and protein sequence information of SARS-CoV-2 is available, characteristics of the virus, as well as the epitopes presented in the pathogen, could be predicted by in silico analysis, which will greatly speed up the vaccine development [10–12].

The purpose of this study is to predict B-cell epitopes on spike protein and T-cell epitopes within nucleocapsid protein of SARS-CoV-2 by applying the bioinformatics methods and immunoinformatic tools. The step-by-step procedure of in silico analysis is depicted in Fig. 1. Epitopes information presented by this work may aid in developing a promising vaccine against SARS-CoV-2.

Methods

Data retrieval and sequence alignment

Protein sequences of SARS-CoV-2 were retrieved from the NCBI database (YP_009724390, YP_009724397). Clustal Omega is the current standard version of the Clustal family which was widely used for biological sequences alignment. The program uses seeded guide trees and Hidden Markov model engine to generate

Fig. 1 Study workflow. Suitable proteins of SARS-CoV-2 were selected at the first step for epitope prediction. The second step comprised of B- and T-cell epitope analysis with bioinformatics approaches. Epitope evaluation was followed and appropriate ones were chosen for vaccine design

alignments [13]. In this study, alignment of protein sequences was performed on the EMBL-EBI server with Clustal Omega program. Conserved domains within predicted polypeptides of SARS-CoV-2 were analyzed with CD-search on the NCBI website.

Linear B-cell epitope prediction

ABCpred [14] and BepiPred [15] servers were employed for B-cell epitope forecast. We used a threshold value of 0.85 to achieve a sensitivity between 95.5% and 99.5% for epitope prediction on ABCpred server [14]. The length of linear B-cell epitopes normally varies from 5 to 30 residues. In this study, we used the default window length of 16 to obtain the maximum accuracy of prediction [14]. Predicted epitopes were highlighted as sphere in SARS-CoV-2 spike protein structure viewed by the pymol molecular graphics system [16]. Surface accessibility of predicted peptides was evaluated with the recently resolved protein structure [17]. We utilized vaxijen2.0 server to analyze the antigenicity of chosen epitopes [18].

Discontinuous B-cell epitope prediction

Prediction of discontinuous epitopes on spike protein (PDB ID: 6VSB chain B) was conducted via DiscoTope 2.0 server [19]. The parameter was set at − 1.0 which indicates 85% specificity and 30% 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 disctope 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 recently resolved [17].

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 [20] or HLA-2 [21] molecule. A relatively small pool of HLA alleles covering the majority of the population, over 97 and 99% for class I and class II respectively, were chosen in the analysis [22, 23]. The sequences were given in plain format and the top 50% scoring peptides were retained for further analysis.

Profiling and evaluation of predicted T-cell epitopes

Key features including digestion, mutation, toxicity, allergenicity, hydro and physiochemical properties were analyzed via vaxijen 2.0 [18], protein digest server [12], AllerTOP v2.0 server [24], and ToxinPred server [25]. Immunogenicity of predicted HLA-1 binding peptides was assessed by the Class-I Immunogenicity service provided on IEDB.

Results

Protein coding features of SARS-CoV-2 genome

A map of the predicted open reading frames (ORFs) is depicted in Supplementary Figure 1 based on the genome sequence of the virus Wuhan-Hu-1 isolate (NCBI reference sequence number: NC_045512.2). The genomic structure of SARS-CoV-2 shares characteristics that are also found in other coronaviruses including SARS-CoV, MERS-CoV, and HCoV-NL63. All these coronaviruses contain recognizable ORFs including the replicase (ORF1ab polyprotein), surface glycoprotein (spike protein), envelope protein, membrane glycoprotein, nucleocapsid protein, and several non-structure proteins (NSP). The conserved domains of proteins encoded by the SARS-CoV-2 genome are summarized in Supplementary Table 1. Spike protein mediates the specific binding of the virion to the receptor on the host cell membrane. The overall structure of spike protein is outside the virus particle [17]. Thus, it is an ideal target for B-cell epitope screening. Compared to spike protein, nucleocapsid protein is more conserved in selected coronaviruses (Fig. 2). 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 (CTL)-mediated response [26, 27], which suggests 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. All protein sequences were downloaded from NCBI database with accession IDs shown in Fig. 2a. The total sequence identity and phylogenetic tree results were presented in Fig. 2b, c. Consistent with a recently published study [28], we found that both spike and nucleocapside proteins in SARS-CoV-2 are more closely related to that of SARS-CoV. The protein domains of spike and nucleocapsid proteins (Fig. 2d) were depicted based on previous studies on SARS-CoV [29, 30] and the protein alignment result in the current study (supplementary files S1 and S2). The amino acid sequence identity result confirmed a high similarity between SARS-CoV-2 and SARS-CoV. As anticipated, the nucleocapsid protein is more conserved among selected coronaviruses compared to spike protein.
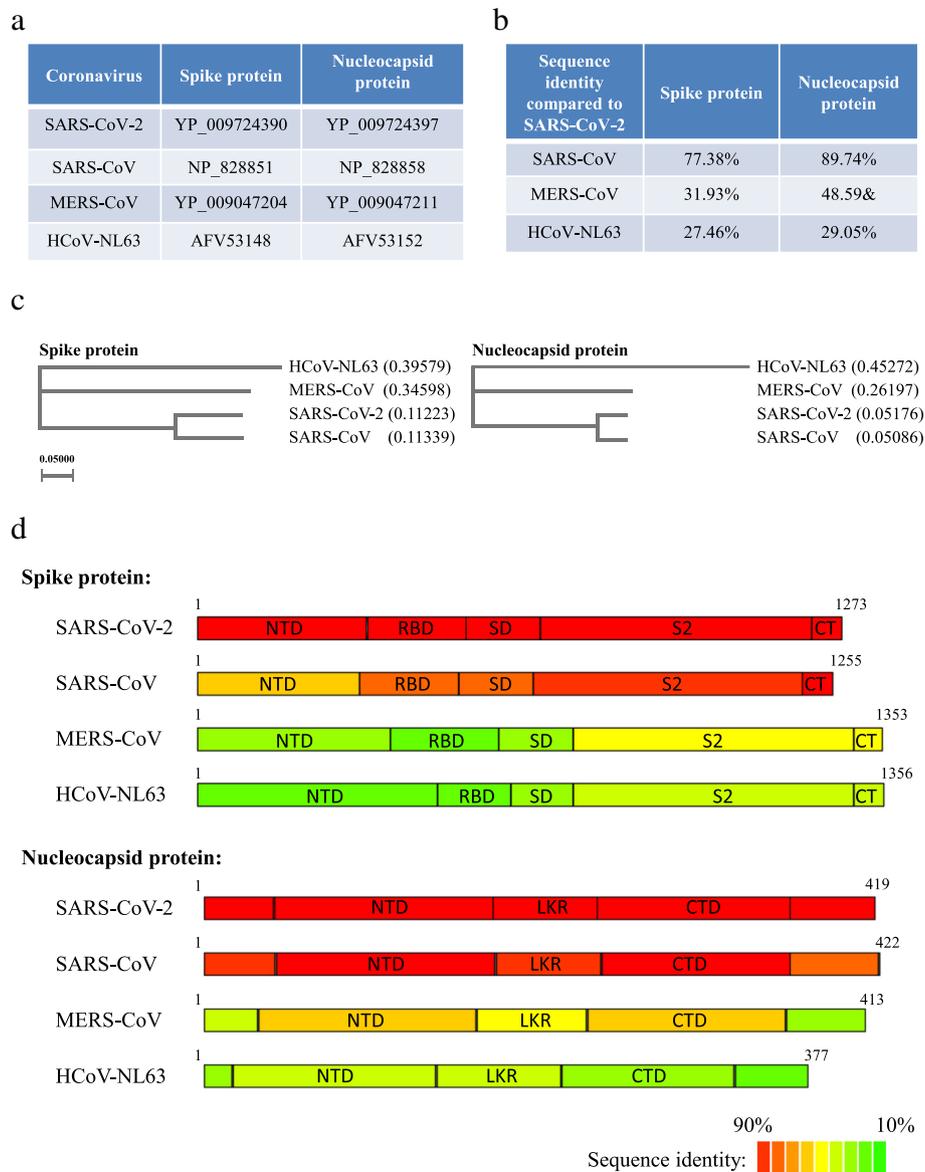
Fig. 2 Spike and nucleocapsid protein in selected coronaviruses. **a** Accession IDs of spike protein and nucleocapsid protein in selected coronavirus. **b** Sequence identity of spike protein and nucleocapsid protein among selected coronavirus. **c** Phylogenetic analysis of target proteins of selected coronaviruses. **d** Sequence identity of subdomains of spike protein and nucleocapsid protein reflected by color (red color indicates high sequence identity and green color indicates low sequence identity)

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

ABCpred	Position	Epitope sequence	Score	Antigenicity
Spike	406–421	EVRQIAPGQTGKIADY	0.85	1.3837
	648–663	GCLIGAEHVNNSYECD	0.90	0.8480
	898–913	FAMQMAYRFNGIGVTQ	0.88	1.3096
	1058–1073	HGVVFLHVTYVPAQEK	0.89	0.8847
	1064–1079	HVTYVPAQEKNFTTAP	0.88	0.8933
	1206–1221	YEQYIKWPWYIWLGFI	0.89	0.9510

Table 2 B-cell epitope predicted via BepiPred server are presented along with their position and antigenicity scores

| BepiPred	Position	Epitope sequence	Antigenicity
Spike	315–324	TSNFRVQPTE	1.3571
	333–338	TNLCPF	1.2508
	372–397	ASFSTFKCYGVSPTKLNDLCFTNVYA	1.2880
	406–426	EVRQIAPGQTGKIADYNYKLP	1.3005
	1037–1045	SKRVDFCGK	1.7321
	1204–1209	GKYEQY	1.2821

B-cell epitopes recognition

The full-length sequence of spike protein was scanned for putative sequential B-cell epitopes by two types of bioinformatics programs. A total of 28 non-overlapping peptides were identified by ABCpred server with the threshold set at 0.85 (Supplementary Table 2). For sequential B-cell epitopes prediction on BepiPred-2.0 server, a threshold value of 0.5 was applied and 35 peptides were predicted (Supplementary Table 3). Antigenicity was calculated by Vaxijen 2.0 server and peptides with the highest antigenicity scores were selected (Tables 1 and 2). The structure of SARS-CoV-2 spike protein was resolved recently with Cryo-electron microscopy (cryo-EM) [17], which could greatly facilitate the process of vaccine development. Predicted epitopes in Tables 1 and 2 were highlighted as sphere in monomer structure of spike protein viewed with pymol (Supplementary Figure 2). While most epitopes predicted were exposed on the surface of spike monomer, only epitopes $Spike_{315-324}$ (TSNFRVQPTE), $Spike_{333-338}$ (TNLCPF), $Spike_{648-663}$ (GCLIGAEHVNNSYECD), $Spike_{1064-1079}$ (HVTYVPAQEKNFTTAP) displayed good surface accessibility in spike trimer (Fig. 3 and Supplementary File: B-cell-epitope-animation.ppt), the pattern more likely exists in nature. Conformation-based B-cell epitopes were computed on DiscoTope 2.0 server [19]. A threshold value of − 1.0 was chosen for the computation, which corresponds to a specificity of 85% and a sensitivity of 30%. The contact number, propensity score, and disctope score for each amino acid that passed the threshold were presented in Table 3. The position of these residues was viewed with pymol and highlighted as sphere (Fig. 4). Processing with a combination of B-cell epitope scanning and peptide analysis forecasted 4 potent linear epitopes and 10 residues involved in discontinuous epitopes formation.

T-cell epitopes recognition

In our study, the 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 NetMHCII-pan methods for HLA-2 binding prediction).

For HLA-1 binding peptide prediction, the top 50% scoring peptides were 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 (Supplementary Table 4). Six peptides with the highest antigenicity scores 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 4). A similar strategy was applied for HLA-2 binding peptides prediction on the IEDB server and 64 peptides were identified as HLA-2 binding sequences (Supplementary Table 5). Six peptides with the highest antigenicity scores were selected for HLA-2 molecule screening and the result was presented in Table 5. In the selected peptides pool for HLA binding, $Nucleocapsid_{104-112}$ (LSPRWYFYY) was predicted as both HLA class-I and class II binding peptides.
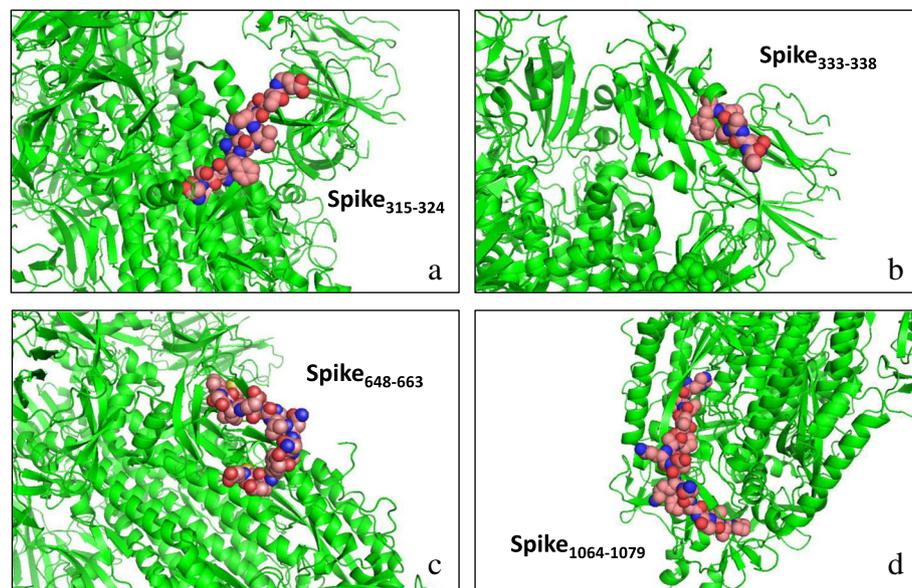
Fig. 3 Site of B cell epitopes predicted in SARS-CoV-2 spike protein trimer. $Spike_{315-324}$ (**a**), $spike_{333-338}$ (**b**), $spike_{648-663}$ (**c**), and $spike_{1064-1079}$ (**d**) were highlighted in sphere in the protein structure. Colors of elements presented in the sphere of protein structure: carbon, tint; hydrogen, gray; nitrogen, blue; oxygen, red; sulfur, yellow
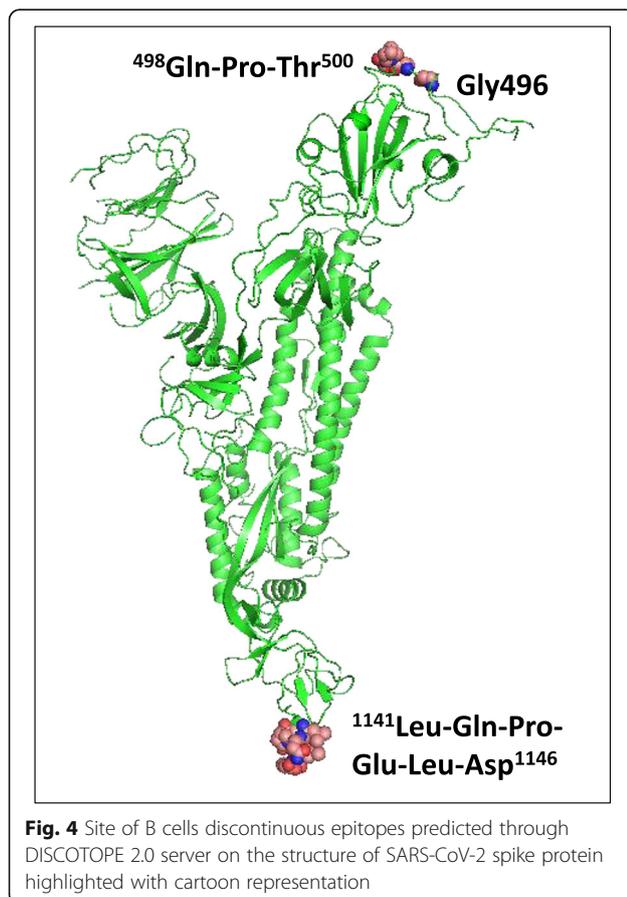
Table 3 Discontinuous B-cell epitopes predicted through DiscoTope 2.0 server

Residue position	Residues name	Contact number	Propensity score	Discotope score
496	GLY	1	−0.693	−0.728
498	GLN	7	1.188	0.246
499	PRO	5	0.294	−0.315
500	THR	1	2.231	1.860
1141	LEU	5	−0.017	−0.59
1142	GLN	7	0.372	−0.476
1143	PRO	6	0.629	−0.134
1144	GLU	4	0.704	0.163
1145	LEU	5	0.171	−0.424
1146	ASP	4	0.724	0.181

Additionally, this peptide may excel in the capability of binding to a large number of HLA molecules as shown in Tables 4 and 5. A partially overlapping region was found in the CTL epitope Nucleocapsid$_{66-74}$ (FPRG QGVPI) and the helper T-lymphocyte (Th) epitope Nucleocapsid$_{67-75}$ (PRGQGVPIN), which suggests the sequence containing these two epitopes may initiate both CD4$^+$ and CD8$^+$ dependent immune response.

Fig. 4 Site of B cells discontinuous epitopes predicted through DISCOTOPE 2.0 server on the structure of SARS-CoV-2 spike protein highlighted with cartoon representation

Selected T-cell epitopes feature profiling and evaluation

Peptide stability, mutation analysis, toxicity, allergenicity, hydro and physiochemical features were calculated and the results were presented in Supplementary Table 6. 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, the class I immunogenicity test was performed and the scores were presented in Table 6. A higher score indicates a higher potential of immune response induction.

Peptides selected and multi-epitope based vaccine design

To induce humoral and cellular immune response simultaneously, five peptides that contain four linear B-cell epitopes and three T-cell epitopes (Fig. 5a) were selected for vaccine development. To facilitate the processing of the T-cell epitopes, selected peptides from nucleocapsid protein were extended five amino acid residues at both N- and C-terminus as compared to the predicted epitopes. These peptides were scanned on the IEDB database. We found that five sequences presented in the vaccine construct were identical to the experimentally verified epitopes on SARS-CoV. These determined peptides displayed a strong or medium binding affinity to a series of MHC molecules (Supplementary Table 7). These epitopes likely possess cross-protective effects against SARS-CoV-2 as well. As shown in Fig. 5b, peptides selected in this study were joined together by using GPGPG and (GGGGS)$_2$ linkers. The Pan DR epitope (PADRE), a universal Th epitope that activates CD4$^+$ cells [31], was introduced at the N terminus of the vaccine construct to enhance helper T cell activity. The vaccine construct can be generated as previously reported [32, 33].

Discussion

Different from the less harmful human coronaviruses continuously circulating in the human population,

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

Peptide	Position	HLA class-I alleles	Vaxijen score
LSPRWYFYY	104–112	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
RSRNSSRNS	189–197	HLA-A*3001	1.1144
IGYYRRATR	84–92	HLA-A*3101, HLA-A*3301, HLA-A*6801, HLA-A*0301, HLA-A*3001	0.8880
FTALTQHGK	53–61	HLA-A*6801, HLA-A*1101, HLA-A*0301, HLA-A*0101, HLA-A*3001, HLA-A*3101	0.8510
KSAAEASKK	249–257	HLA-A*0301, HLA-A*1101, HLA-A*3001, HLA-A*6801, HLA-A*3101	0.7679
FPRGQGVPI	66–74	HLA-B*0702, HLA-B*5101, HLA-B*5301, HLA-B*0801, HLA-B*3501	0.7585

coronaviruses originated from animals could become fatal pathogens 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 pandemic.

Compared to traditional vaccine development, potent epitopes can be predicted via bioinformatics analysis, which makes the vaccine design straightforward and fast [34]. As the majority of spike protein is exposed outside the virion, it could be an ideal target to search for B-cell epitopes. Spike proteins in MERS-CoV and SARS-CoV have been shown to induce robust immune response [35, 36]. Baruah and Bose's study also indicated that epitopes on spike protein could be promising candidates for SARS-CoV-2 vaccine development [37]. In another study, the experimentally determined epitopes derived from SARS-CoV were screened and those harboring the same sequence in SARS-CoV-2 were identified [38]. Similar strategy was followed in Grifoni's study [39]. Though the SARS-CoV-2 spike protein displayed 77.38% sequence identity toward that of SARS-CoV, most of the antibodies against SARS-CoV spike protein showed poor cross-reactivity with that of SARS-CoV-2 [40], which

indicates the spike proteins have significantly variable structure. Thus, we applied distinct strategy and evaluated the surface accessibility of predicted B-cell epitopes to exclude those buried inside the protein. Though we utilized similar bioinformatics tools and resources, the peptides we selected for vaccine development are not alike. In our analysis of SARS-CoV-2 spike protein, four peptides identified from multi-step screening displayed excellent surface accessibility. The peptide $Spike_{333-338}$ produces a random coil structure and has a high antigenicity score. Noticeably, this peptide sits in the receptor-binding domain (RBD) of spike protein of SARS-CoV-2, which has been proved to mediate the binding to ACE2 on the epithelial membrane of human lungs [41]. Likely, antibodies recognizing this epitope could also neutralize the virus and prevent infection. The epitope $Spike_{648-663}$ predicted in this study is 12 amino acids upstream of the furin cleavage site which is critical for SARS-CoV-2 biogenesis [42]. We expect that an antibody recognizing the epitope $Spike_{648-663}$ could have expanded therapeutic function.

Besides B-cell epitope prediction on spike protein, we selected nucleocapsid protein as a target protein for T-cell epitope computation for the following reasons. First,

Table 5 HLA class-II alleles binding epitopes predicted by IEDB server

Peptides	Position	HLA class-II alleles	Vaxijen score
IKLDDKDPN	337–345	DRB1_0701, DRB1_0301, DRB3_0101, DRB1_0405, DPA1_0301/DPB1_0402, DQA1_0101/DQB1_0501, DRB1_0901, DRB1_1101, DRB5_0101, DRB4_0101, DRB1_0101, DRB1_1302, DQA1_0501/DQB1_0301, DPA1_0103/DPB1_0201,	2.3118
RSGARSKQR	32–40	DRB5_0101, DRB4_0101, DQA1_0501/DQB1_0301	1.7874
RIGMEVTPS	319–327	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, DRB5_0101, DQA1_0401/DQB1_0402, DRB1_0101, DRB1_0701, DQA1_0501/DQB1_0201, DRB1_1501, DPA1_0201/DPB1_0101, DPA1_0103/DPB1_0201	1.5314
RGTSPARMA	203–211	DQA1_0501/DQB1_0301, DRB1_0901, DQA1_0102/DQB1_0602	1.2953
LSPRWYFYY	104–112	DRB1_0405, DQA1_0101/DQB1_0501, DPA1_0103/DPB1_0201, DPA1_0201/DPB1_0501, DRB3_0101, DRB1_1201, DRB1_1302, DPA1_0201/DPB1_0101, DRB1_0901, DQA1_0501/DQB1_0201, DRB1_1501, DRB1_1101, DRB1_0101, DRB1_0401, DRB1_0701, DPA1_0301/DPB1_0402, DRB5_0101, DRB1_0301	1.2832
PRGQGVPIN	67–75	DQA1_0501/DQB1_0301, DRB1_1302, DRB1_0101, DRB1_0901, DRB1_0401, DRB1_0701, DRB1_1501, DRB1_0405	1.1707

Table 6 Class-I immunogenicity

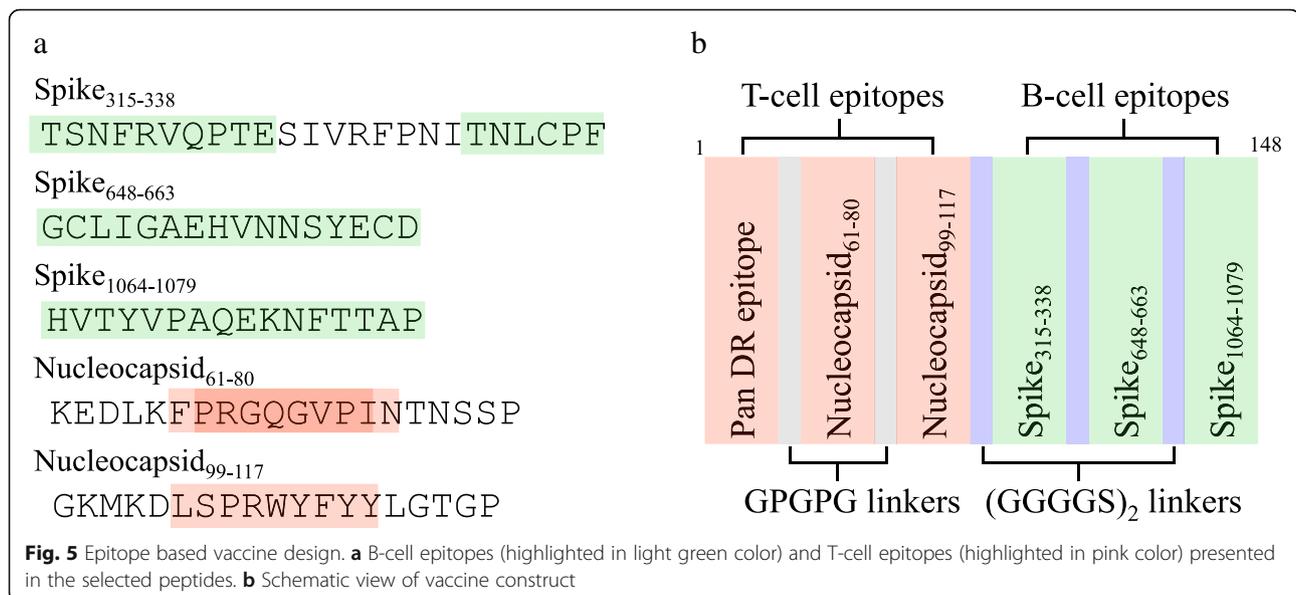| Peptide	Position	Class-I immunogenecity
LSPRWYFYY	104–112	0.35734
RSRNSSRNS	189–197	0.1499
IGYYRRATR	84–92	−0.00164
FTALTQHGK	53–61	−0.0226
KSAAEASKK	249–257	−0.07922
FPRGQGVPI	66–74	−0.26664

nucleocapsid protein in SARS-CoV and MERS-CoV has been experimentally tested as a robust immunogen to induce cytotoxic T lymphocyte response [26, 27]; Second, nucleocapsid protein is the predominant protein expressed in the virion during the early stage of infection [43]; Third, nucleocapsid protein was detected in the majority of SARS-CoV infected patients as early as day 1 of infection [44, 45], which suggests nucleocapsid protein-based vaccine may evoke T-cell dependent immune response timely. Eighty-one and sixty-four peptides within nucleocapsid protein were predicted to bind HLA-1 and HLA-2 molecules respectively. The peptide Nucleocapsid$_{104-112}$ showed an immunogenicity score of 0.36, suggesting a capability to initiate a strong immune response. Moreover, the peptides Nucleocapsid$_{104-112}$ and Nucleocapsid$_{66-75}$ harbor experimentally determined epitopes that bind to a number of HLA-1 molecules. In this study, Nucleocapsid$_{104-112}$ and Nucleocapsid$_{66-75}$ are forecasted as multi-epitopes that bind to both HLA-1 and HLA-2 molecules, suggesting a potency of both CTL and Th-mediated immune response initiation. 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 reactions during the pre-clinical and clinical trials. 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.

The final vaccine construct comprising of CTL, Th, and B-cell epitopes is predicted to initiate protective humoral and cellular immune response against SARS-CoV-2.

Conclusions

B-cell epitopes on spike protein and T-cell epitopes in the nucleocapsid protein were predicted and analyzed in the current study. A total of 63 linear B-cell epitopes on spike protein were forecasted by ABCpred and BepiPred servers. Ten residues within spike protein (Gly^{496}, Glu^{498}, Pro^{499}, Thr^{500}, Leu^{1141}, Gln^{1142}, Pro^{1143}, Glu^{1144}, Leu^{1145}, Asp^{1146}) were forecasted by DiscoTope 2.0 program as components of conformational B-cell epitopes. IEDB server was used for T-cell epitopes prediction, which gave rise to 81 and 64 peptides with binding capability to class-I and class-II molecule respectively. Four B-cell epitopes: Spike$_{315-324}$ (TSNFRVQPTE), Spike$_{648-663}$ (GCLIGAEHVNNSYECD), Spike$_{1064-1079}$ (HVTYVP AQEKNFTTAP), and Spike$_{333-338}$ (TNLCPF) were selected from the list based on their antigenicity score and surface accessibility. The T-cell epitopes Nucleocapsid$_{104-112}$ (LSPRWYFYY) and Nucleocapsid$_{66-75}$ (FPRG QGVPIN) on nucleocapsid protein could bind a wide spectrum of both HLA-1 and HLA-2 molecules. The final vaccine construct consists of CTL, Th, and B-cell epitopes that potentially protect individuals against SARS-CoV-2 by inducing both humoral and cellular immune response, which should be successively validated within both in vitro and in vivo models.

Fig. 5 Epitope based vaccine design. **a** B-cell epitopes (highlighted in light green color) and T-cell epitopes (highlighted in pink color) presented in the selected peptides. **b** Schematic view of vaccine construct

Chen *et al. Infectious Diseases of Poverty*   (2020) 9:88

Page 9 of 10

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s40249-020-00713-3.

Additional file 1. Supplementary file S1.

Additional file 2. Supplementary file S2.

Additional file 3. Supplementary figures and tables.

Additional file 4. B-cell-epitope-animation.

Abbreviations

SARS-CoV: Severe acute respiratory syndrome associated coronavirus; MERS-CoV: Middle East respiratory syndrome coronavirus; HCoV-NL63: Human Coronavirus NL63; HLA: Human leukocyte antigen; MHC: Major histocompatibility complex; WHO: World Health Organization; NSP: Non-structure protein; ORF: Open reading frame; CTL: Cytotoxic T-lymphocyte; Th: Helper T-lymphocyte

Authors' contributions

HC and XW performed bioinformatics analysis of epitope prediction. HC wrote the first draft of the manuscript. YC, LT, JZ, and XY revised the manuscript. XW was responsible for the final approval of the manuscript. The authors read and approved the final manuscript.

Availability of data and materials

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

Ethics approval and consent to participate

Not applicable.

Consent for publication

All presentation of the data have consent for publication.

Competing interests

The authors declare no conflict of interests.

Author details

[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 410011, Hunan, China. [2]Department of Laboratory Medicine, The Second Xiangya Hospital, Central South University, Changsha 410011, Hunan, China. [3]Hunan Institute of Parasite Disease, WHO Collaborating Center For Research and Control on Schistosomiasis in Lake Region, Yueyang 414000, Hunan, China. [4]Department of Forensic Medicine Science, Xiangya School of Basic Medicine, Central South University, Changsha 410013, Hunan, China. [5]Department of Parasitology, Xiangya School of Basic Medicine, Central South University, Changsha 410013, Hunan, China.

References

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395(10223):497–506.

2. Chakraborty C, Sharma AR, Bhattacharya M, Sharma G, Lee SS. The 2019 novel coronavirus disease (COVID-19) pandemic: a zoonotic prospective. Asian Pac J Trop Med. 2020;13:242–6.

3. WHO| Novel Coronavirus (2019-nCoV) Situation Report-128. Report 2020. Available from: https://www.whoint/docs/default-source/coronaviruse/situation-reports/20200527-covid-19-sitrep-128pdf.

4. Coronaviridae Study Group of the International Committee on Taxonomy of V. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020;5(4):536–44.

5. 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(8):727–33.

6. Ceraolo C, Giorgi FM. Genomic variance of the 2019-nCoV coronavirus. J Med Virol. 2020;92(5):522–8.

7. Qiu T, Yang Y, Qiu J, Huang Y, Xu T, Xiao H, et al. CE-BLAST makes it possible to compute antigenic similarity for newly emerging pathogens. Nat Commun. 2018;9(1):1772.

8. Chakraborty C, Sharma AR, Sharma G, Bhattacharya M, Lee SS. SARS-CoV-2 causing pneumonia-associated respiratory disorder (COVID-19): diagnostic and proposed therapeutic options. Eur Rev Med Pharmacol Sci. 2020;24(7):4016–26.

9. Rauch S, Jasny E, Schmidt KE, Petsch B. New vaccine technologies to combat outbreak situations. Front Immunol. 2018;9:1963.

10. Ul Qamar MT, Bari A, Adeel MM, Maryam A, Ashfaq UA, Du X, et al. Peptide vaccine against chikungunya virus: immuno-informatics combined with molecular docking approach. J Transl Med. 2018;16(1):298.

11. 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;132:243–53.

12. Ul Qamar MT, 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.

13. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019;47(W1):W636–41.

14. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins. 2006;65(1):40–8.

15. 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):W24–9.

16. DeLano WL. Pymol: An open-source molecular graphics tool. CCP4 Newsl Protein Crystallogr. 2002;40:82–92.

17. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020;367(6483):1260–3.

18. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007;8:4.

19. 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(12):e1002829.

20. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics. 2016;32(4):511–7.

21. 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.

22. Weiskopf D, Angelo MA, de Azeredo EL, Sidney J, Greenbaum JA, Fernando AN, 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):E2046–53.

23. 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):325–35.

24. Dimitrov I, Flower DR, Doytchinova I. AllerTOP--a server for in silico prediction of allergens. BMC Bioinformatics. 2013;14(Suppl 6):S4.

25. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP. Peptide toxicity prediction. Methods Mol Biol. 2015;1268:143–57.

26. Buchholz UJ, Bukreyev A, Yang L, Lamirande EW, Murphy BR, Subbarao K, et al. Contributions of the structural proteins of severe acute respiratory syndrome coronavirus to protective immunity. Proc Natl Acad Sci U S A. 2004;101(26):9804–9.

27. 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):718.

28. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, et al. 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):221–36.

29. Gui M, Song W, Zhou H, Xu J, Chen S, Xiang Y, et al. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. Cell Res. 2017;27(1): 119–29.

30. Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH. The SARS coronavirus nucleocapsid protein--forms and functions. Antivir Res. 2014;103:39–50.

31. Alexander J, del Guercio MF, Frame B, Maewal A, Sette A, Nahm MH, et al. Development of experimental carbohydrate-conjugate vaccines composed of Streptococcus pneumoniae capsular polysaccharides and the universal helper T-lymphocyte epitope (PADRE). Vaccine. 2004;22(19):2362–7.

32. Chauhan V, Rungta T, Goyal K, Singh MP. Designing a multi-epitope based vaccine to combat Kaposi sarcoma utilizing immunoinformatics approach. Sci Rep. 2019;9(1):2517.

33. Khan M, Khan S, Ali A, Akbar H, Sayaf AM, Khan A, et al. Immunoinformatics approaches to explore *Helicobacter Pylori* proteome (virulence factors) to design B and T cell multi-epitope subunit vaccine. Sci Rep. 2019;9(1):13321.

34. 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):469–81.

35. Zhou Y, Jiang S, Du L. Prospects for a MERS-CoV spike vaccine. Expert Rev Vaccines. 2018;17(8):677–86.

36. Du L, Zhao G, Lin Y, Sui H, Chan C, Ma S, 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):948–56.

37. Bhattacharya M, Sharma AR, Patra P, Ghosh P, Sharma G, Patra BC, et al. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. J Med Virol. 2020;92(6): 618–31.

38. Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses. 2020;12(3):254.

39. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. Cell Host Microbe. 2020;27(4): 671–680.e2.

40. 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.

41. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270–3.

42. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell. 2020; 181(2):281–292.e6.

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

44. Che XY, Hao W, Wang Y, Di B, Yin K, Xu YC, et al. Nucleocapsid protein as early diagnostic marker for SARS. Emerg Infect Dis. 2004;10(11):1947–9.

45. Li YH, Li J, Liu XE, Wang L, Li T, Zhou YH, et al. 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.