Epitope-based peptide vaccine design against spike protein (S) of novel coronavirus (2019-nCoV): an immunoinformatics approach

Eman Ali Awadelkareem (eman9ali999@gmail.com)
Faculty of Veterinary Medicine, University of Khartoum

Nisreen Osman Mohammed
Ahfad Centre for Science and Technology, Ahfad University for Women

Bothina Bakor Mohammed Gaafar
Ministry of Animal Resources, South Darfur State, Nyala, Sudan

Zahra - Abdelmagid
School of Pharmacy, Ahfad University for Women, Omdurman, Sudan

Sumaia AwadElkariem Ali (somiahadlool@yahoo.com)
Sudan University of Science and Technology, College of Veterinary Medicine
https://orcid.org/0000-0003-2861-7670

Research

Keywords: spike protein, SARS-CoV-2, Wuhan sequences, vaccine design, coronavirus

DOI: https://doi.org/10.21203/rs.3.rs-30076/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

Recently the global pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has generated a significant need on identifying drugs or vaccines to prevent or reduce clinical infection of Coronavirus disease – 2019 (COVID-19). In this study, immuno-informatics tools were utilized to design a potential multi-epitopes vaccine against SARS-CoV-2 spike S protein. Structural analysis for SARS-CoV-2 spike S protein was also conducted.

Method:

SARS-CoV-2 spike S protein sequences were retrieved from the GeneBank of National Central Biotechnology Information (NCBI). Immune Epitope Database (IEDB) tools were used to predict B and T cell epitopes, to evaluate their allergenicity, toxicity and cross-reactivity and to calculate population coverage. Protparam sever was applied to determine protein characterization of spike protein and predicted epitopes. Molecular docking for the proposed MHCI epitopes were also achieved against Tall like Receptor (TLR8) receptors and HLA-B7 allele.

Result

Immuno-informatics analysis of S protein using IEDB identified only one B cell epitope 1054QSAPH1058 as linear, surface and antigenic. Although 1054QSAPH1058 was estimated as non-allergic and non-toxic, it showed protein instability. Moreover, around 45 discontinuous epitopes were also recognized as different exposed surface area. In MHCI methods, six conserved stable and safe epitopes (689FAMQMYARF906, 258WTAGAAAYY266 and 3FVFLVLLPL10-202KIYSKHTPF1210T7IIAPTNFT1120 and 1060YVFHVTTYV1068) were identified. These epitopes showed strong interaction when docked with TLR8 and HLA-B7 allele especially 1060YVFHVTTYV1068 and 3FVFLVLLPL10-202 epitopes. Three epitopes were also predicted (689FAMQMYARF906, 888FGAGAALQ896 and 342FNATRFASY260) using MHCII methods. Furthermore, the potential multi-epitopes were acquired by assessing allergenicity, toxicity and cross-reactivity to prevent autoimmunity.

Conclusion

The multi-epitopes vaccine was predicted based on Bioinformatics tools that may provide reliable results in a shorter time and at a lower cost. However, further in vivo and in vitro studies are required to validate their effectiveness.

Background

Recently, the World Health Organization announced the emergence of a new severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus as a major threat to human health because it causes a global pandemic of lower respiratory diseases and was known as New Coronary Pneumonia (NCP) by the Chinese government initially [1]. In a situational report 96 published on 23 of April 2020 reported that more than 2 million confirmed cases with SARS-CoV-2 (2.544.792) worldwide and in Eastern Mediterranean Region including Sudan were 144.450 [2]. First cases were reported by the Health Commission of Hubei province, China on December 2019 of unexplained pneumonia, latter on 9th of January 2020, was officially identified as the cause of the COVID-19 a SARS-CoV-2 outbreak in Wuhan,China [3, 4].

Coronaviruses (CoVs) are members of the family Coronaviridae, the enveloped viruses that possess extraordinarily large single-stranded RNA genomes ranging from 26 to 32 kilobases in length. SARS-CoV belongs to Beta coronaviruses which infect the mammals [4, 5]. SARS-CoV-2 causes flu-like symptoms, such as persistent coughing, fever, shortness of breath, and difficulty breathing, which are similar to the Severe Acute Respiratory Syndrome (SARS), and the Middle East Respiratory Syndrome (MERS) [6].

Structurally Coronaviruses have two types of proteins none structurally proteins proteases (nsp3 and nsp5) and RdRp (nsp12) and structurally proteins Nucleocapsid (N), Membrane glycoprotein (M), Envelope (E), and Spike (S). Spike protein is a part of virus that's bind to cell receptor and facilitate entering of this virus and is the main target for neutralization antibodies. Moreover, it is a trimeric protein present in outer surface of the virus. The molecular weight of spike protein is 180 kDa and contains two subunits S1 and S2, which they required cellular protease for the process of priming in to S1 and S2. These two subunits facilitate the virus attachment and membrane fusion [7, 8]. Spike S protein binds to specific cell receptor angiotensin-converting enzyme 2 (ACE2) and use the cellular serine protease TMPRSS2 for S protein priming [9, 10].

In last decade many vaccines have been proposed for SARS-CoV-2 including DNA vaccine, synthetic peptides and even in silico perdition peptides, however the DNA and synhtetics peptides elicits positive result against humeral and poor immunogenicity against T cell which need an adjuvant [11–13].

No specific anti-virus drugs or vaccines are available against SARS-CoV-2 lethal disease. It is reported that greater than 85% of SARS-CoV-2 patients in China have been receiving Traditional Chinese Medicine (TCM) treatment, and presented the clinical evidence showing the beneficial effect of TCM in the treatment of the patients [4]. However, no approved vaccine is designed for SARS-CoV-2, under circumstances that protection against virus is curicle, especially in African countries which have a poor economic, weak health systems, poor health-seeking behaviors and different cultural practices that's will delay detection of cases and transmission of virus [14, 15].
In this study spike S protein of SARS-CoV-2 was used to predict peptides that can stimulate humoral and cellular immunity using various immunoinformatics tools beside structural analysis of spike protein.

2. Materials And Methods

2.1. Protein Sequence Retrieval


2.2 Multiple Sequence Alignment and Epitope Conservancy Assessment

The conserve regions cross the Spike S protein were identified using ClustalW in BioEdit software version 7.2.5 [16]. Epitope conservancy analysis in Immune Epitope Database (IEDB) was used to detect potential epitope conservancy (http://tools.iedb.org/conservancy/) [17].

2.3 Phylogeny Analysis:

The retrieved sequences were subjected to MEGA7.0.26 (7170509) software using maximum likelihood parameter to determine the evolutionary relationship between retrieved sequences [18].

2.4 Protein Structural Analysis

Reference sequence of SARS-CoV-2 spike S protein was submitted to Protparam server to predict the physiochemical properties. Many characteristics were predicted include molecular weight, theoretical isoelectric point (pI), amino acid composition, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average of hydropathicity (GRAVY) [19].

The server SOPMA (https://npsa-prabi.ibcp.fr › NPSA › npsa sopma) was used to identify the spike protein secondary structure calculations.

Conserved Domains in Spike protein were predicted using CDD-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) [20-21] and PFAM (http://www.pfam.sanger.ac.uk/) [22]. The Ubiquitination sites were also identified via UbPred [23-25]. Amphiphilicity and Hydrophathy indices were calculated for the query protein sequence by SOSUI server which categories the protein nature into cytoplasmic or trans-membrane [26]. BLASTP in NCBI database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using a default parameter for conservation analysis was used to match homologous spike reference sequences of different coronaviruses in human and animals against SARS-CoV-2 spike protein sequence. Phylogenetic tree was also constructed based on constraint-based Multiple Alignment Tool (COBALY). (https://www.ncbi.nlm.nih.gov/blast/treeview/treeView.cgi) [27, 28].

2.5 Prediction of B and T cells

B and T cell epitopes were predicted using Immune Epitope Database (IEDB) (http://tools.iedb.org/mhci/) from reference sequences of Spike S protein [29]. Prediction of B-cell antigenic epitopes is important in designing vaccine components and immuno-diagnostic reagents. Generally, B-cell antigenic epitopes are classified as either continuous or discontinuous. The majority of available epitope prediction methods focus on continuous epitopes. Discontinuous epitopes dominate most antigenic epitope families [30].

To predict the continuous epitopes, BepiPred linear B-cell epitopes predicting method was used [31]. Then the predicted peptides were subjected to Emini surface accessibility prediction tool and Kolaskar and Tongaonkar antigenicity methods to determine the epitopes that located on the surface and the score of epitopes antigenicity respectively [32, 33].

The prediction of discontinuous epitopes was carried out using DiscoTope server [34]. Parameter was set at ≥ 0.5 which indicated 90% specificity and 23% sensitivity. This method based on surface accessibility and amino acid statistics in a collected form dataset of discontinuous epitopes found out by X-ray crystallography of antigen/antibody protein buildings. The position of predicted epitopes clusters on 3D structure of S protein was identified by Chimera [35].

The T cell epitopes were predicted for different alleles of major histocompatibility complex class I (MHC I) and class II (MHC II). Artificial neural networks and NN-align methods were used to predict the binding of proposed peptides with different MHC I and MHC II with binding affinity (IC50) less or equal to 300 and 1000 for MHC I and II respectively [36, 37].

2.6 Prediction of Antigenicity, Allergenicity and Toxicity for Proposed Epitopes:

The proposed epitopes were also subjected in VaxiJen v2.0 server to determine the antigenicity [38]. AllerTop server was used to identify allergenicity while Toxinpred server was used to estimate the safety of selected epitopes [39, 40].

2.7 Analysis for the Sequence Similarity with the Human Self-Epitopes:

To assess the possibility of autoimmune diseases for epitopes derived from Spike S protein. The selected epitopes were blasted against the non-redundant protein sequences of human [taxid: 9606] using NCBI Blastp suite program with default parameters (http://www.ncbi.nlm.nih.gov/BLAST/).

2.8 Population Coverage:
Immune Epitope Database (IEDB) was also used to calculate the population coverage for proposed epitopes for MHCI and II against whole population worldwide [41].

2.9 Homology Modeling

Raptor X structure prediction server (http://raptorx.uchicago.edu/StructurePrediction/predict/) was used to predict the 3D structure of reference sequence of spike S protein [42-46]. PEPFOLD server was used for homology modelling of MHCI epitopes (http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/) from amino acid sequences [47, 48]. 3D structure of TLR8 (PDB: 3W3G) (resolution 2.30 Å) was taken from Protein Data Bank (PDB) [49, 50]. Chimera software 1.8 was used as visualization tool [35].

2.10 Molecular Docking

Molecular docking was achieved via Patch Dock online autodock tools (https://bioinfo3d.cs.tau.ac.il/PatchDock/) [51, 52]. The 3D structures of MHCI epitopes were used as ligands and the 3D structures of (Toll-Like Receptors) TLR8 (PDB: 3W3G) and (Human Leucocytes Antigen) HLA-B7 (BDP: 3VCL) were used as receptors. FireDock (http://bioinfo3d.cs.tau.ac.il/FireDock/) was used to select the five top models/) [53]. Visualization of the results were performed using UCSF-Chimera software 1.8 [35].

3. Results

3.1 Retrieved Sequence Information:

Eight spike S protein sequences were retrieved from NCBI with their accession numbers, area and date of collection as shown in Table 1. All sequences are from China.

3.2 Multiple Sequence alignment and Epitopes Conservancy Assessment:

Multiple sequence alignment of the retrieved sequences was performed using ClustalW through BioEdit software showed high conservancy between the aligned sequences. The conserved regions were identified by identity and similarity of amino acid sequences (Fig.1).

3.3 Phylogeny Analysis

Evolutionary analyses were conducted in MEGA7.0.26 (7170509) software using maximum likelihood parameter. The analysis involved 8 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 1273 positions in the final dataset [18, 54] see Fig.2.

3.4 Structural Analysis:

The physiochemical properties of Spike S protein calculated by protparam server revealed that it contained 1273 amino acids (aa) with molecular weight of 141178.47 kDa, which reflects a good antigenic nature. Theoretical isoelectric point (pI) was 6.24 which indicate its negative in nature. An isoelectric point below 7 states a negatively charged protein, however the total number of negatively charged residues (Asp + Glu) was 110 aa and positively charged residues (Arg + Lys) was 103 aa. Protparam computed instability-index (II), which was a 33.01, this categories spike S protein as stable protein. Aliphatic-index was 84.67, which devotes a thought of proportional volume hold by aliphatic side chain and GRAVY value for protein sequence is 0.012. (Grand average of hydropathicity (GRAVY: -0.079). The half-life of protein described as the total time taken for its disappearing after it has been synthesized in cell, which was computed as 30 hour (h) for mammalian-reticulocytes, > 20 h for yeast, > 10 h for Escherichia coli. The N-terminal of the sequence considered is M (Met).

The component of secondary structure predicted by GOR IV server (https://npsa-prabi.ibcp.fr/cgi-bin/secpred_sopma.pl) revealed alpha helix (28.59%), Beta turn (3.38%), and random coil (44.78%) as in (Fig.3). The ambiguous states of the Spike S protein were predicted via UbPred server (Fig.4) it showed that there were six amino acids sites at the position 182, 776, 811, 947, 1255 and 1266 respectively with low condence ambiguity site (grey color). Moreover, the average of hydrophobicity predicted by SOSUI server was -0.079183. This server predicted two Trans-membrane regions as shown in Table 2. DiANNA1.1 tool calculated 20 disulphides bond (S–S) positions and assign them a score and it makes prediction based on trained neural system (see Additional le 1: Table S1).

Pfam server predicted 19 conserved domains (E-value cut-off to 1.0) in spike S protein (2 significant and 17 insignificant). The significant domains were Coronavirus S2 glycoprotein (corona-S2) and Spike receptor binding domain (Spike_rec.bind). The insignificant domains include Spike glycoprotein N-terminal domain, Baculovirus polyhedron envelope protein, Protein of unknown function (DUF2959), MuKF middle domain, Calcium-dependent calmodulin binding, Protein of unknown function (DUF1664), Tetramerisation domain of TRPM, Domain of unknown function (DUF4795), Retroviral envelope protein, SlyX protein and Biogenesis of lysosome-related organelles complex-1 subunit 2.

The significant conserved domains were sequenced by Conserved Domain (CDD) BLAST search. The results revealed that corona-S2 (pfam01601) is the only member of the superfamily cl20218 [55]. The top related sequences were Human coronavirus HKU1 (isolate N1), Bovine coronavirus, Porcine haemagglutinating, Human coronavirus HKU1, Bat SARS coronavirus HKU3-3, Murine coronavirus, SARS coronavirus ExoN1, SARS coronavirus ExoN1, Murine hepatitis virus strain A59.
Spike_rec_bind (pfam09408) is the only member of the superfamily cl09656 [56]. The top related sequences were Human coronavirus HKU1 (isolate N1), Bovine coronavirus, Equine coronavirus NC99, Porcine haemagglutinating encephalomyelitis virus, Human coronavirus HKU1, Human coronavirus HKU1 (isolate N2), Bat SARS coronavirus HKU3-3, Murine coronavirus RA59/R13, SARS coronavirus ExoN1 and Murine hepatitis virus strain A59.

The closest homologue obtained from BLASTP results was the severe acute respiratory syndrome-related coronavirus (75.96%) with E value 0.00 followed by Bat coronavirus BM48-31/BGR/2008 (71.96%) see Table 3 and Fig. 5 and 6.

3.5 Proposed B cell epitopes:

In B cell prediction methods, thirty two conserved epitopes were predicted using Bepipred Linear Epitope Prediction method. Among them only five epitopes were pass Emini surface accessibility prediction tool and kolasark and Tongaonkar antigenicity methods. These epitopes were (FVFLVLLPL 1054) (QSAPH1058-1068) (KAHFP 1090, and (1137 VYDPLQPELDSF 1148) Among these epitopes only one epitope 1054QSAPH1058 was found non-toxin and non-allergen when investigated by Allertop and ToxinPred servers (Table 4).

Unfortunately, the promising B cell epitope when subjected to Protparam server to determine its physiochemical properties, it was found unstable. The molecular weight is 538.56 kDa and the GRAVY value for protein sequence is -1.460.

However, Discotope 2.0 server was used to calculate surface availability in term of residue contact number and novel tendency amino acid score was utilized to predict the discontinuous epitopes. 3D structure of S protein (PDB ID: 6VSB) [57] was used for discontinuous epitopes prediction, 90% specificity, - 3.700 threshold and 22.000 Angstroms propensity score radius. Total 45 discontinuous epitopes were identified at different exposed surface areas (Table 5). Position of each predicted epitope on surface of 3D structure of S protein shown in Fig.7 were visualized using Chimera tool [35].

3.6 Proposed epitopes for MHCI and MHCII:

MHCI prediction tools outward 109 conserved epitopes of SARS-CoV-2 spike S protein. Of these 7 epitopes were identified as top MHCI epitopes based on the high antigenicity score and great linkase with MHCI alleles class A, B and C. These epitopes were (FVFLVLLPL 1054) (QSAPH1058-1068) (KAHFP 1090, and (1137 VYDPLQPELDSF 1148) Among these epitopes only one epitope 1054QSAPH1058 was found non-toxin and non-allergen when investigated by Allertop and ToxinPred servers (Table 4).

In MHCI prediction methods, many core sequences were predicted to interact with huge numbers of alleles as well as high antigenicity score. The core 898FAMQMAYRF 906, that predicted in MHCI methods was interacted with 101 MHCI alleles. 888FGAGAALOI 896 and 342FNATRFASV 350 epitopes were also interacted with 83 and 65 alleles in MHCII respectively (Additional file 2: Table S2).

3.7 Antigenicity, Allergenicity, Toxicity of MHCI and MHCII Epitopes:

The expected MHCI and MHCII epitopes were subjected to VaxiJen v2.0 server, AllerTop v2.0 and ToxiPred to predict the antigenity, allergenicity and toxicity of predicted epitopes respectively. The predicted MHCI and II epitopes were antigenic, but 1060VVLHHTTVV 1068 and 898FAMQMAYRF 906, epitopes displayed the higher scores (1.5122 and 1.0278 respectively). The epitopes were also free of causing allergenicity and toxicity (see Table. 6 and Additional file 2: Table S2).

3.8. Cross Reactivity with Human Epitopes:

The only one epitope "YIKWPYWYI" shared between MHCI and MHC II has been detected to have putative conserved domain identical to human peptide among all selected epitopes. Therefore, it was removed from the epitopes pool to avert triggering an autoimmune response.

3.9 Predicted Physicochemical Properties

The proposed epitopes for both MHCI and II were further subjected to Protparam server to determine their physicochemical properties. All predicted epitopes were stable except 718FTISVTTEI 726 (see Table 7 and 8 and Fig.8 and 9).

3.10 Population Coverage:

The proposed epitopes for MHCI revealed 95.74 coverage against whole population while the proposed epitopes for MHCII showed only 78.09 population coverage against whole population (Table 9).

3.11 Molecular Docking:

The epitope 1060VVLHHTTVV 1068 interacted strongly with TLR8 (global energy -84.58) followed by 2FVLVLLPL 10 (global energy -64.23) see Table 10 and Fig.10. Moreover, MHCI peptides were also docked with HLA-B7 (PDB ID: 3VCL). The epitopes were interacted strongly with HLA-B7, but the best one was 2FVLVLLPL 10 (global energy 78.81) followed by 1060VVLHHTTVV 1068 (global energy -63.20) see Table 11 and Fig.11.

Discussion

Recently, the World Health Organization announced the emergence of a new SARS-CoV-2 virus as a major risk to human health because it causes a global pandemic of lower respiratory diseases and was known as New Coronary Pneumonia (NCP) by the Chinese government initially [1]. The recent global
pandemic has placed a high priority on identifying drugs or vaccines to prevent or lessen clinical infection of Coronavirus disease – 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1, 58]. This study therefore focused on the in silico design and development of a potential multi-epitope vaccine against SARS-CoV-2 spike protein.

In the present study, the calculation of the physicochemical properties of spike S protein of severe acute respiratory syndrome (COVID-19) using protparam revealed that the protein has good antigenic property, negative in nature and stable [19].

The identification of epitopes from B cells is important in immuno-detection and immunotherapy applications since the epitope is a minimal immune unit strong enough to stimulate a strong humoral immune response with no harmful side effects to the human body [34].

In B cell prediction methods, five conserved epitopes were identified as linear, surface and antigenic based on Bepipred linear prediction methods, Emimi and Kolaskar and Tangonkar antigenicity measurement tools sequentially. Only one epitope (906,258 QSAPH406) was identified as non-allergic using AllerTop v. 2. Software and nontoxic using ToxinPred software. It was also free from provoking an autoimmune response; however it was found unstable as a protein when analyzed by protparam server.

The discontinuous epitopes are increasingly explicit and have higher dominant attributes over linear epitopes [59]. 3D structure of S protein was used for discontinuous epitope prediction using DiscoTope 2.0 server. The server uses a combination of amino acid statistics, spatial information, and surface exposure [60]. In this study, a total of forty five conserved discontinuous epitopes were identified at different exposed surface areas. These epitopes may have principal role in humoral immunity. However, it has been estimated that > 90% of B-cell epitopes are discontinuous, i.e., consist of segments that are distantly separated in the pathogen protein sequence and brought into proximity by the folding of the protein [60].

In MHCI methods, sex epitopes were expected to interact strongly with great numbers of HLA alleles (989FAMQMYARF906,238WTAGAAAY266, -202 202 KIYSKHTPI170 7124APTNTFT120 and 1060VFLHVTVY1068). These epitopes showed high antigenicity and safety.

MHCI prediction methods, predicted many epitopes such as (989FAMQMYARF906,888FGAGAAALIQ996 and 342FNATRFASV350) as they interacted with great numbers of HLA alleles as well as high antigenicity and safety. However, 989FAMQMYARF906 epitope that predicted in MHCI was also expected to interact with huge number of MHCI alleles. In a similar in silico study, five CTL epitopes, three sequential B cell epitopes and five discontinuous B cell epitopes were predicted from the viral surface glycoprotein of SARS-CoV-2 virus [61].

Physicochemical properties of MHCI and II epitopes using protparam server indicated that all epitopes were predicted to be stable except 718FTISVTTEI726. According to the server threshold, an instability index below 40 is indicative of protein stability, and a lower value demonstrates a more stable protein [62].

The molecular weights in all epitopes were slightly different ranging from 846.98 to 1164. Gravy values were also different. Gravy is a measure of hydrophobicity or hydrophilicity of the structures. Gravy value for all structures was positive, representing their slightly hydrophobic nature, except, 202 KIYSKHTPI170 showed negative GRAVY (hydrophilic) [62]. The theoretical pl values of epitopes were also varies in range of 9.70 to 4.00. In a vaccine designed for injection, pl is preferred closer to the normal blood pH, body fluids, or neutral pH [63].

The secondary structure predicted by GOR IV server indicated that the spike protein consisted of alpha helix (28.59%), Beta turn (3.38%), and random coil (44.78%). The ambiguous states predicted via UbPred server exhibited six amino acid site at position 182, 776, 811, 947, 1255 and 1266 respectively with low confidence ambiguity sites. Moreover, SOSUI server predicted two Trans-membrane regions while DiANNA1.1 tool calculated 20 disulphides bond (S–S) positions in SARS-CoV-2 spike protein.

Corona-S2 and Spike_rec_bind were identified as main motif in spike S protein. They were also sequenced by Conserved Domain (CDD) BLAST search [55]. The nearest homologue obtained from BLASTP results was the severe acute respiratory syndrome-related coronavirus (75.96%) with E value 0.00 followed by Bat coronavirus BM48:31/BGR/2008 (71.96%).

Furthermore, top MHC class I binding epitopes were subjected to PEPFOLD server for homology modeling. 3D structures of MHCI epitopes were docked with the TLR8 by Patch dock server. Firedock server identified five best models in the results. Previous studies have reported the involvement of TLR in immune protection against viral infection and other pathogens [64, 65].

To evaluate potential immune interaction between TLR8 and the 3D structure of predicted MHCI peptide, a protein-ligand docking analysis was performed. 1060VFLHVTVY1068 epitope interacted strongly with TLR8 that indicated by the lower global energy − 84.58 followed by -202 202 KIYSKHTPI170 (global energy − 64.23) (Table 10 and Fig. 10). In addition, docking with HLA-B7 exhibited strong association with HLA-B7 for all epitopes see Table 11 and Fig. 11. However, -202 202 VFLHVTVY1068 produced lower global energy 78.81 which indicates the strong binding affinity in comparison with other epitopes followed by 1060VFLHVTVY1068 (global energy − 63.20).

Furthermore, the proposed epitopes for MHCI revealed high coverage (95.74%) against whole population whereas the MHCI epitopes showed only 78.09% population coverage against whole population.

**Conclusion**

This study used various immuno-informatics tools to design a potential multi-epitopes vaccine coding for B-cell and T-cell (HTL and CTL) epitopes.
Immuno-informatics analyses of spike S protein generate a candidate vaccine that contain a number of high-affinity MHCI, and II, linear and conformational B-cell epitopes that lack the allergenicity, toxicity and autoimmune properties which support their potential as vaccine candidates. The effectiveness of the designed vaccine should be further confirmed in wet-lab experiments.

**Abbreviations**

Coronavirus disease-19 (COVID-19), Novel coronavirus, (2019-nCoV), Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), New Coronavirus Pneumonia (NCP), Coronaviruses (CoVs), Severe Acute respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS), spike glycoprotein (S), National Central Biotechnology Information (NCBI), Immune Epitope Database (IEDB) World health organization (WHO), multiple sequence alignment (MSA), Reference sequence (refseq), major histocompatibility complex (MHC), The half maximal inhibitory concentration (IC50), The human leukocyte antigen (HLA), Toll like Receptor (TLR), Conserved Domain (CDD), Traditional Chinese Medicine (TCM), Theoretical isoelectric point (PI), Grand average of hydropathicity (GRAVY), Molecular weight (MW), Instability index(Ii), Constructed based on constraint-based Multiple Alignment Tool (COBALY), Kilodaltons (kDa), Angiotensin-converting enzyme 2 (ACE2), Artificial neural networks (ANN), I neural network-based alignment (NN-align).

**Declarations**

**Acknowledgements**

Not applicable.

**Authors’ contributions**

All authors participating in designing the study, accomplished the experiments, analyze the results, interpreted the data and wrote the manuscript. All authors read and approved the final manuscript. The final revision done by Sumaia Awad-Elkariem Ali and Eman, Ali Awadelkareem

**Funding**

Not applicable.

**Availability of data and materials**

All the data supporting the findings are contained within the manuscript

**Supplementary information**

Additional S1 file1.

Additional S2 file2.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ details**

Eman, Ali Awadelkareem

Faculty of Veterinary Medicine, University of Khartoum, Khartoum, Sudan.

Nisreen Osman Mohammed

Ahfad Centre for Science and Technology. Ahfad University for Women Khartoum-Sudan.

Bothina Bakor Mohammed Gaafer

Ministry of Animal Resources, South Darfur State, Nyala, Sudan.

Zahra Abdelmagid

School of Pharmacy, Ahfad University for Women, Omdurman, Sudan.
References


Tables

Table 1. Accession numbers and area of collection of the retrieved sequences of Spike S protein sequences of SARS-CoV-2 from NCBI.* Reference sequence of Spike S protein.

<table>
<thead>
<tr>
<th>No.</th>
<th>Accession No.</th>
<th>Country</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>YP_009724390.1*</td>
<td>China</td>
<td>2020</td>
</tr>
<tr>
<td>2</td>
<td>QHR63280.2</td>
<td>China</td>
<td>2020</td>
</tr>
<tr>
<td>3</td>
<td>QHR63270.2</td>
<td>China</td>
<td>2020</td>
</tr>
<tr>
<td>4</td>
<td>QHR63260.2</td>
<td>China</td>
<td>2020</td>
</tr>
<tr>
<td>5</td>
<td>QHR63250.2</td>
<td>China</td>
<td>2020</td>
</tr>
<tr>
<td>6</td>
<td>QIC53213.1</td>
<td>China</td>
<td>2020</td>
</tr>
<tr>
<td>7</td>
<td>QIA20044.1</td>
<td>China</td>
<td>2020</td>
</tr>
</tbody>
</table>

Table 2: Transmembrane Region in Spike S protein of SARS-CoV-2

<table>
<thead>
<tr>
<th>No.</th>
<th>N-terminal</th>
<th>Transmembrane region</th>
<th>C-terminal</th>
<th>Type</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>MFVFLVLLPLVSSQCVNLTTRT</td>
<td>22</td>
<td>Secondary</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>1223</td>
<td>GLIAIVMVTIMLCCMTSCSKLCK</td>
<td>1245</td>
<td>Primary</td>
<td>23</td>
</tr>
</tbody>
</table>

Table 3: BLASTP similarity search of SARS-CoV-2 spike S protein against refseq of other coronaviruses spike S proteins in human and animals. BLASTP in NCBI was used as default parameter for conservation analysis to find homologous spike reference sequences of different coronaviruses in human and animals against SARS-CoV-2 spike protein sequence.

<table>
<thead>
<tr>
<th>NCBI Protein ID</th>
<th>Protein Name</th>
<th>E-value</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>YP_009724390.1</td>
<td>Severe acute respiratory syndrome-related coronavirus</td>
<td>0.0</td>
<td>75.96%</td>
</tr>
<tr>
<td>YP_003858584.1</td>
<td>Bat coronavirus BM4831/BGR/2008</td>
<td>0.0</td>
<td>71.96%</td>
</tr>
<tr>
<td>YP_009273005.1</td>
<td>Rousettus bat coronavirus</td>
<td>0.0</td>
<td>35.86%</td>
</tr>
<tr>
<td>YP_009047204.1</td>
<td>Middle East respiratory syndrome-related coronavirus</td>
<td>1e-176</td>
<td>35.10%</td>
</tr>
<tr>
<td>YP_009535241.1</td>
<td>Human coronavirus OC43</td>
<td>2e-145</td>
<td>37.63%</td>
</tr>
<tr>
<td>YP_209233.1</td>
<td>Murine hepatitis virus strain JHM</td>
<td>4e-142</td>
<td>36.65%</td>
</tr>
<tr>
<td>YP_009194639.1</td>
<td>Camel alphacoronavirus</td>
<td>7e-109</td>
<td>31.54%</td>
</tr>
<tr>
<td>YP_0037671.1</td>
<td>Human coronavirus NL63</td>
<td>4e-103</td>
<td>30.78%</td>
</tr>
<tr>
<td>YP_001941166.1</td>
<td>Turkey coronavirus</td>
<td>4e-103</td>
<td>36.92%</td>
</tr>
<tr>
<td>NP_040831.1</td>
<td>Infectious bronchitis virus</td>
<td>2e-101</td>
<td>35.91%</td>
</tr>
<tr>
<td>YP_004070194.1</td>
<td>Feline infectious peritonitis virus</td>
<td>9e-99</td>
<td>31.95%</td>
</tr>
<tr>
<td>NP_004244.1</td>
<td>Transmissible gastroenteritis virus</td>
<td>1e-98</td>
<td>31.98%</td>
</tr>
<tr>
<td>YP_009194242.1</td>
<td>Swine enteric coronavirus</td>
<td>2e-93</td>
<td>30.60%</td>
</tr>
<tr>
<td>NP_508310.1</td>
<td>Porcine epidemic diarrhoea virus</td>
<td>4e-91</td>
<td>30.44%</td>
</tr>
</tbody>
</table>

Table 4: B cell proposed epitopes of spike S protein of SARS-CoV-2 spike S protein.

*The score of Bepipred linear prediction methods is 0.350; + Emini is 1.00 and # Kolaskar and Tongaonkar antigenicity is 1.041. ** predicted epitope. Allergenicity and toxicity of epitopes were inspected by Allertop and ToxinPred servers.

<table>
<thead>
<tr>
<th>Epitope sequence</th>
<th>length</th>
<th>Start</th>
<th>End</th>
<th>+Emini</th>
<th>#Kolaskar</th>
<th>Allergenicity</th>
<th>Toxicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDSK</td>
<td>4</td>
<td>110</td>
<td>113</td>
<td>1.497</td>
<td>1.014</td>
<td>Allergen</td>
<td>Non-Toxin</td>
</tr>
<tr>
<td>RVYST</td>
<td>5</td>
<td>634</td>
<td>638</td>
<td>1.426</td>
<td>1.086</td>
<td>Allergen</td>
<td>Non-Toxin</td>
</tr>
<tr>
<td>*QSAPH</td>
<td>5</td>
<td>1054</td>
<td>1058</td>
<td>1.597</td>
<td>1.052</td>
<td>Non-Allergen</td>
<td>Non-Toxin</td>
</tr>
<tr>
<td>KAHFP</td>
<td>5</td>
<td>1086</td>
<td>1090</td>
<td>1.191</td>
<td>1.051</td>
<td>Allergen</td>
<td>Non-Toxin</td>
</tr>
<tr>
<td>VYDPQPELDSF</td>
<td>12</td>
<td>1137</td>
<td>1148</td>
<td>1.279</td>
<td>1.073</td>
<td>Allergen</td>
<td>Non-Toxin</td>
</tr>
</tbody>
</table>

Table 5: Discontinuous epitopes of SARS-CoV-2 spike S protein predicted through DISCOTOPE 2.0 Server. Parameter was set at ɛ 0.5 which indicated 90% specificity and 23% sensitivity. Residues are shown in three-letter code, and number of contacts shows the connection of amino acid with others.
Residue position | Residue Name | Contact Number | Propensity Score | Discotope Score
--- | --- | --- | --- | ---
281 | GLU | 0 | -3.366 | -2.979
282 | ASN | 7 | -2.664 | -3.162
415 | THR | 0 | -3.819 | -3.38
420 | ASP | 4 | -3.618 | -3.662
449 | TYR | 4 | -0.567 | -0.962
450 | ASN | 11 | -1.78 | -2.841
454 | ARG | 14 | -1.224 | -2.694
491 | PRO | 7 | -0.72 | -1.442
492 | LEU | 15 | -0.95 | -2.565
493 | GLN | 9 | -0.572 | -1.541
494 | SER | 7 | -0.846 | -1.553
496 | GLY | 3 | 0.041 | -0.309
498 | GLN | 4 | 0.68 | 0.142
499 | PRO | 5 | 0.178 | -0.417
500 | THR | 0 | 1.907 | 1.688
503 | VAL | 5 | -1.856 | -2.218
505 | TYR | 8 | -1.528 | -2.272
556 | ASN | 2 | -3.79 | -3.584
558 | LYS | 2 | -1.479 | -1.539
560 | LEU | 2 | -1.137 | -1.236
561 | PRO | 0 | -0.961 | -0.851
562 | PHE | 0 | -2.061 | -1.824
703 | ASN | 4 | -2.02 | -2.248
704 | SER | 3 | -1.469 | -1.645
705 | VAL | 10 | -2.821 | -3.646
793 | PRO | 1 | -2.278 | -2.131
794 | ILE | 1 | -2.5 | -2.327
809 | PRO | 4 | -2.691 | -2.841
914 | ASN | 7 | -1.117 | -1.794
917 | TYR | 9 | -2.702 | -3.426
918 | GLU | 13 | -2.285 | -3.517
1071 | GLN | 9 | -2.775 | -3.491
1099 | GLY | 1 | -3.789 | -3.468
1100 | THR | 0 | -3.877 | -3.431
1101 | HIS | 8 | -2.903 | -3.489
1111 | GLU | 10 | -1.693 | -3.684
1118 | ASP | 4 | -2.016 | -3.129
1140 | PRO | 7 | -0.961 | -1.656
1141 | LEU | 5 | -0.257 | -0.802
1142 | GLN | 7 | 0.318 | -0.523
1143 | PRO | 6 | 1.067 | 0.255
1144 | GLU | 6 | 0.716 | -0.056
1145 | LEU | 5 | 0.162 | -0.431
1146 | ASP | 5 | 0.731 | 0.072

Table 6: Top MHCI epitopes with interacted alleles and their antigenicity scores. The antigenicity of MHCI epitopes were predicted using VaxiJen v2.0 server. *VVFLHVTYV epitope showed high antigenicity score.

<table>
<thead>
<tr>
<th>Epitopes</th>
<th>Start</th>
<th>End</th>
<th>Antigenicity</th>
<th>Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAMOMAYRF</td>
<td>898</td>
<td>906</td>
<td>1.0278</td>
<td>HLA-A<em>02:06; HLA-A</em>23:01; HLA-A<em>24:02; HLA-A</em>25:02; HLA-B<em>08:01; HLA-B</em>15:01; HLA-B<em>35:01; HLA-B</em>53:01; HLA-B<em>58:01; HLA-C</em>03:03; HLA-C<em>05:01; HLA-C</em>12:03</td>
</tr>
<tr>
<td>WTAGAAYY</td>
<td>258</td>
<td>266</td>
<td>0.6306</td>
<td>HLA-A<em>01:01; HLA-A</em>26:01; HLA-A<em>28:02; HLA-A</em>30:02; HLA-A<em>46:01; HLA-A</em>47:01; HLA-A<em>49:01; HLA-A</em>50:01; HLA-B<em>15:01; HLA-B</em>35:01; HLA-B<em>53:01; HLA-B</em>58:01</td>
</tr>
<tr>
<td>FVFLLPLL</td>
<td>2</td>
<td>10</td>
<td>0.8601</td>
<td>HLA-A<em>02:01; HLA-A</em>02:06; HLA-A<em>36:02; HLA-B</em>35:01; HLA-B<em>39:01; HLA-C</em>03:03; HLA-C<em>12:03; HLA-C</em>14:02</td>
</tr>
<tr>
<td>KIVSKHTPI</td>
<td>202</td>
<td>210</td>
<td>0.7455</td>
<td>HLA-A<em>02:01; HLA-A</em>02:06; HLA-A<em>30:01; HLA-A</em>32:01; HLA-C<em>03:03; HLA-C</em>14:02; HLA-C*15:02</td>
</tr>
<tr>
<td>FTYSVTEI</td>
<td>718</td>
<td>726</td>
<td>0.8535</td>
<td>HLA-A<em>02:01; HLA-A</em>02:06; HLA-A<em>36:02; HLA-B</em>58:01; HLA-C<em>03:03; HLA-C</em>12:03; HLA-C*15:02</td>
</tr>
<tr>
<td>IAEPNTPTI</td>
<td>712</td>
<td>720</td>
<td>0.7052</td>
<td>HLA-A<em>02:06; HLA-A</em>23:01; HLA-B<em>53:01; HLA-B</em>58:01; HLA-C<em>03:03; HLA-C</em>12:03</td>
</tr>
<tr>
<td>*VVFLHVTYV</td>
<td>1060</td>
<td>1068</td>
<td>1.5122</td>
<td>HLA-A<em>02:01; HLA-A</em>02:06; HLA-A<em>46:01; HLA-C</em>07:01; HLA-C*12:03</td>
</tr>
</tbody>
</table>

Table 7: Physiochemical properties of top predicted MHCI peptides. MW* Molecular weight. II*Instability index. Ext. coefficient* Extinction coefficients. GRAVY*Grand average of hydropathicity.
### Table 8: Physiochemical properties of top predicted MHCII peptides.

<table>
<thead>
<tr>
<th>Epitopes</th>
<th>MW*</th>
<th>Theoretical pI</th>
<th>Estimated half-life</th>
<th>Formula</th>
<th>Ext. coefficient*</th>
<th>II*</th>
</tr>
</thead>
<tbody>
<tr>
<td>*FAMQMAYRF</td>
<td>1164.41</td>
<td>8.75</td>
<td>1.1 hours (mammalian reticulocytes, in vitro)</td>
<td>C_{46}H_{77}N_{13}O_{12}S</td>
<td>1490</td>
<td>stable</td>
</tr>
<tr>
<td>WTAGAAYY*</td>
<td>973.05</td>
<td>5.52</td>
<td>2.8 hours (mammalian reticulocytes, in vitro)</td>
<td>C_{49}H_{68}N_{10}O_{13}</td>
<td>8480</td>
<td>stable</td>
</tr>
<tr>
<td>FVFLVLLPL*</td>
<td>1060.39</td>
<td>5.52</td>
<td>1.1 hours (mammalian reticulocytes, in vitro)</td>
<td>C_{48}H_{60}N_{12}O_{12}</td>
<td>should not be visible by UV spectrophotometry.</td>
<td>stable</td>
</tr>
<tr>
<td>KIYSKHTPI</td>
<td>1086.30</td>
<td>9.70</td>
<td>1.3 hours (mammalian reticulocytes, in vitro)</td>
<td>C_{41}H_{64}N_{12}O_{12}</td>
<td>1490</td>
<td>stable</td>
</tr>
<tr>
<td>FTSYVZTEI</td>
<td>1010.15</td>
<td>4.00</td>
<td>1.1 hours (mammalian reticulocytes, in vitro)</td>
<td>C_{48}H_{60}N_{10}O_{12}</td>
<td>should not be visible by UV spectrophotometry.</td>
<td>unstable</td>
</tr>
<tr>
<td>IAIPTNFTI</td>
<td>989.18</td>
<td>5.52</td>
<td>20 hours (mammalian reticulocytes, in vitro)</td>
<td>C_{43}H_{58}N_{10}O_{12}</td>
<td>should not be visible by UV spectrophotometry.</td>
<td>stable</td>
</tr>
<tr>
<td>VVFLHVTVY</td>
<td>1076.30</td>
<td>6.71</td>
<td>100 hours (mammalian reticulocytes, in vitro)</td>
<td>C_{53}H_{81}N_{11}O_{12}</td>
<td>1490</td>
<td>stable</td>
</tr>
</tbody>
</table>

### Table 9: Population Coverage for Proposed MHC-I and II Epitopes.

<table>
<thead>
<tr>
<th>Epitopes</th>
<th>MHCI  Coverage</th>
<th>Epitopes</th>
<th>MHCII Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAMQMAYRF</td>
<td></td>
<td>FNATRFASV</td>
<td></td>
</tr>
<tr>
<td>FGAGAALQI</td>
<td>95.74</td>
<td>WTAGAAYY</td>
<td>78.09</td>
</tr>
<tr>
<td>KIYSKHTPI</td>
<td></td>
<td>FVFLVLLPL</td>
<td></td>
</tr>
<tr>
<td>IAIPTNFTI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VVFLHVTVY</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 10: Docking of MHC-I epitopes with TLR8.

Molecular docking was performed using PatchDock online autodock tools by submitting MHC-I predicted epitopes and 3D structures of TLR8 (PDB: 3W3G). VVFLHVTVY produced low global energy which indicates strong binding affinity.

<table>
<thead>
<tr>
<th>Epitopes</th>
<th>Global Energy</th>
<th>Attractive VdW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-49.69</td>
<td>-29.90</td>
</tr>
<tr>
<td>2</td>
<td>-37.73</td>
<td>-27.38</td>
</tr>
<tr>
<td>3</td>
<td>-64.23</td>
<td>-24.16</td>
</tr>
<tr>
<td>4</td>
<td>-35.61</td>
<td>-25.05</td>
</tr>
<tr>
<td>5</td>
<td>-56.95</td>
<td>-27.05</td>
</tr>
<tr>
<td>6</td>
<td>-84.58</td>
<td>-36.05</td>
</tr>
</tbody>
</table>

### Table 11: Docking of MHC-I epitopes with HLA-B7.

Molecular docking was achieved using PatchDock online autodock tools by submitting MHC-I predicted epitopes and 3D structures of HLA-B7 (PDB: 3VCL). FVFLVLLPL showed low global energy which indicates strong binding affinity.
### Table

<table>
<thead>
<tr>
<th>No</th>
<th>MHCI epitopes</th>
<th>Global Energy</th>
<th>Attractive VdW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FAMQMAYRF</td>
<td>-52.00</td>
<td>-26.51</td>
</tr>
<tr>
<td>2</td>
<td>WTAGAAYYT</td>
<td>-51.14</td>
<td>-23.00</td>
</tr>
<tr>
<td>3</td>
<td>FVFLVLLPL</td>
<td>-78.81</td>
<td>-33.08</td>
</tr>
<tr>
<td>4</td>
<td>KIYSKHTPI</td>
<td>-46.40</td>
<td>-23.32</td>
</tr>
<tr>
<td>5</td>
<td>IAIFNPFTI</td>
<td>-52.94</td>
<td>-26.75</td>
</tr>
<tr>
<td>6</td>
<td>VVFLHVTYTV</td>
<td>-63.20</td>
<td>-33.01</td>
</tr>
</tbody>
</table>

### Figures

**Figure 1**

Multiple sequence alignment (MSA) of the retrieved strains of spike S protein of SARS-CoV-2 using BioEdit software and ClustalW. Dots indicated the conservancy and letters in cubes showed the alteration in amino acid.
Figure 2
Phylogenetic tree of SARS-CoV-2 spike S protein using MEGA7 software using maximum likelihood parameter.

Figure 3
Prediction of Secondary Structure of SARS-CoV-2 using GOR IV server
Figure 4

Ubiquitination Sites in SARS-CoV-2 S Protein using UbPred server

<table>
<thead>
<tr>
<th>Sequence ID</th>
<th>Start</th>
<th>1</th>
<th>200</th>
<th>400</th>
<th>600</th>
<th>800</th>
<th>1000</th>
<th>1200</th>
<th>1400</th>
<th>1600</th>
<th>1700</th>
<th>1800</th>
<th>1900</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qc70_063</td>
<td>1</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>271</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qc70_064</td>
<td>2</td>
<td></td>
<td>225</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>295</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qc70_065</td>
<td>3</td>
<td></td>
<td>255</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>315</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qc70_066</td>
<td>4</td>
<td></td>
<td></td>
<td>355</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qc70_067</td>
<td>5</td>
<td></td>
<td></td>
<td>375</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qc70_068</td>
<td>6</td>
<td></td>
<td></td>
<td>375</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qc70_069</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>355</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Low confidence Grey
Medium confidence Blue
High confidence Red

Figure 5

Multiple sequence alignment (MSA) for SARS-CoV-2 spike S protein and different reference sequences of coronaviruses spike S proteins in human and animals using BLASTP

Figure 6

Low confidence Grey
Medium confidence Blue
High confidence Red
Phylogenetic tree for different reference sequences for SARS-CoV-2 spike S protein and different reference sequences of coronaviruses spike S proteins in human and animals based on constraint-based Multiple Alignment Tool (COBALY).

Figure 7

Position of B cells discontinues epitopes (coloured sphere) predicted through DISCOTOPE 2.0 Server on the crystal structure of SARS-CoV-2 spike S protein (green colour) highlighted with cartoon representation.

Figure 8
Position of MHCI proposed epitopes (coloured sphere) on 3D structure of SARS-CoV-2 spike S protein (green colour) using UCSF-Chimera software 1.8

Figure 9

Position of MHCII proposed epitopes (coloured sphere) on 3D structure of SARS-CoV-2 spike S protein (green colour) using UCSF-Chimera software 1.8
Graphical representation of interaction analyses between TLR8 (purple colour) and MHC class-I alleles binding epitopes (green colour). Visualization of the results were done using UCSF-Chimera software 1.8. The figure is in symmetry with the information provided in Table 10 and showing the interacting residues.
Figure 11

Graphical representation of interaction analyses between HLA-B7 (pink colour) and MHC class-I alleles binding epitopes (green colour). Visualization of the results were done using UCSF-Chimera software 1.8. The figure is in symmetry with the information provided in Table 11 and showing the interacting residues.

Supplementary Files

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

- supplement1.docx
- supplement2.docx