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 TechnologyAhfad 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

Posted Date: June 9th, 2020

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

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. Protparm 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 $_{1054}QSAPH_{1058}$ as linear, surface and antigenic. Although $_{1054}QSAPH_{1058}$ 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 ($_{898}FAMQMAYRF_{906,}$ $_{258}WTAGAAAYY_{266}$ *and* $_{2}FVFLVLLPL_{10,}$ $_{202}KIYSKHTPI_{210,}$ $_{712}IAIPTNFTI_{720}$ and $_{1060}VVFLHVTYV_{1068}$) were identified. These epitopes showed strong interaction when docked with TLR8 and HLA-B7 allele especially $_{1060}VVFLHVTYV_{1068}$ and $_{2}FVFLVLLPL_{10}$ epitopes. Three epitopes were also predicted ($_{898}FAMQMAYRF_{906,}$ $_{888}FGAGAALQI_{896}$ and $_{342}FNATRFASV_{350}$) 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 including DNA vaccine, synthetic peptides and even in silico perdition peptides, however the DNA and synthetics 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 curricle, 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 humeral and cellular immunity using various immunoinformatics tools beside structural analysis of spike protein.

2. Materials And Methods

2.1. Protein Sequence Retrieval

Spike S protein sequences of SARS-CoV-2 virulent strains were retrieved in FASTA format from the GeneBank of National Central Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/protein/) database in April 2020.

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 Hydropathy 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 (MHCI) and class II (MHCII). 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 A) 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). Total number of atoms was 19710. The total number of Carbon (C), Oxygen (O), Nitrogen (N), Hydrogen (H) and Sulfur (S) were entitled by Formula: $C_{6336}H_{9770}N_{1656}O_{1894}S_{54}$.

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 confidence 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 file 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 kolaskar and Tongaonkar antigenicity methods. These epitopes were ($_{110}LDSK_{113}$, $_{634}RVYST_{638}$, $_{1054}QSAPH_{1058}$, $_{1086}KAHFP_{1090}$, and $_{1137}VYDPLQPELDSF_{1148}$). Among these epitopes only one epitope $_{1054}QSAPH_{1058}$ 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 MHC I epitopes based on the high antigenicity score and great linkage with MHCI alleles class A, B and C. These epitopes were ($_{898}FAMQMAYRF_{906}$, $_{258}WTAGAAAYY_{266}$ *and* $_{2}FVFLVLLPL_{10}$, $_{202}KIYSKHTPI_{210}$, $_{718}FTISVTTEI_{726}$, $_{712}IAIPTNFTI_{720}$ *and* $_{1060}VVFLHVTYV_{1068}$) (Table 6).

In MHCII prediction methods, many core sequences were predicted to interact with huge numbers of alleles as well as high antigenicity score. The core $_{898}FAMQMAYRF_{906}$, that predicted in MHCI methods was interacted with 101 MHCII alleles. $_{888}FGAGAALQI_{896}$ and $_{342}FNATRFASV_{350}$ 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 antigenicity, allergenicity and toxicity of predicted epitopes respectively. The predicted MHCI and II epitopes were antigenic, but $_{1060}VVFLHVTYV_{1068}$ and $_{898}FAMQMAYRF_{906}$, 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 "$_{1209}YIKWPWYIW_{1217}$" shared between MHC I 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 physiochemical properties. All predicted epitopes were stable except $_{718}FTISVTTEI_{726}$ (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 only78.09 population coverage against whole population (Table 9).

3.11 Molecular Docking:

The epitope $_{1060}VVFLHVTYV_{1068}$ interacted strongly with TLR8 (global energy -84.58) followed by $_{2}FVFLVLLPL_{10}$ (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 $_{2}FVFLVLLPL_{10}$ (global energy 78.81) followed by $_{1060}VVFLHVTYV_{1068}$ (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 physiochemical 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, Emini and Kolaskar and Tongaonkar antigenicity measurement tools sequentially. Only one epitope ($_{1054}QSAPH_{1058}$) 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 epitopes 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 ($_{898}FAMQMAYRF_{906,\ 258}WTAGAAAYY_{266}$, $_{2}FVFLVLLPL_{10}$, $_{202}KIYSKHTPI_{210}$, $_{712}IAIPTNFTI_{720}$ and $_{1060}VVFLHVTYV_{1068}$). These epitopes showed high antigenicity and safety.

MHCII prediction methods, predicted many epitopes such as ($_{898}FAMQMAYRF_{906,\ 888}FGAGAALQI_{896}$ and $_{342}FNATRFASV_{350}$) as they interacted with great numbers of HLA alleles as well as high antigenicity and safety. However, $_{898}FAMQMAYRF_{906}$ epitope that predicted in MHCI was also expected to interact with huge number of MHCII 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 $_{718}FTISVTTEI_{726}$. 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, $_{202}KIYSKHTPI_{210}$ showed negative GRAVY (hydrophilic) [62]. The theoretical pI values of epitopes were also varies in range of 9.70 to 4.00. In a vaccine designed for injection, pI 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 identifies 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. $_{1060}VVFLHVTYV_{1068}$ epitope interacted strongly with TLR8 that indicated by the lower global energy − 84.58 followed by $_{2}FVFLVLLPL_{10}$ (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, $_{2}FVFLVLLPL_{10}$ produced lower global energy 78.81 which indicates the strong binding affinity in comparison with other epitopes followed by $_{1060}VVFLHVTYV_{1068}$ (global energy − 63.20).

Furthermore, the proposed epitopes for MHCI revealed high coverage (95.74%) against whole population whereas the MHCII epitopes showed only78.09% 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 Coronary 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), Tall 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), l neural network-based alignment (NN-align).

Declarations

Acknowledgements

Not applicable.

Authors' contributions

All authors participating indesigning 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 Gaafar

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

Zahra Abdelmagid

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

Sumaia AwadElkariem Ali

Department of Veterinary Medicine and Surgery, College of Veterinary Medicine, Sudan University of Science and Technology, Khartoum, Sudan.

Corresponding author

Correspondence to Eman, Ali Awadelkareem and Sumaia AwadElkariem Ali

References

1. Yuen K-S, Ye Z-W, Fung S-Y, Chan C-P, Jin D-Y. SARS-CoV-2 and COVID-19: The most important research questions. Cell bioscience. 2020;10(1):1–5.
2. Organization WH. Coronavirus disease 2019 (COVID-19): situation report, 70. 2020.
3. Chan JF-W, Yuan S, Kok K-H, To KK-W, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. The Lancet. 2020;395(10223):514–23.
4. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nature Medicine. 2020:1–5.
5. Zheng J. SARS-CoV-2: an emerging coronavirus that causes a global threat. Int J Biol Sci. 2020;16(10):1678.
6. Deng C-X. The global battle against SARS-CoV-2 and COVID-19. International Journal of Biological Sciences. 2020;16(10):1676.
7. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nature communications. 2020;11(1):1–12.
8. Ibrahim IM, Abdelmalek DH, Elshahat ME, Elfiky AA. COVID-19 spike-host cell receptor GRP78 binding site prediction. Journal of Infection. 2020.
9. Aronson JK, Ferner RE. Drugs and the renin-angiotensin system in covid-19. British Medical Journal Publishing Group; 2020.
10. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell. 2020.
11. Ng O-W, Chia A, Tan AT, Jadi RS, Leong HN, Bertoletti A, et al. Memory T cell responses targeting the SARS coronavirus persist up to 11 years post-infection. Vaccine. 2016;34(17):2008–14.
12. Abdelmageed MI, Abdelmoneim AH, Mustafa MI, Elfadol NM, Murshed NS, Shantier SW, Makhawi AM. Design of a Multiepitope-Based Peptide Vaccine against the E Protein of Human COVID-19: An Immunoinformatics Approach. BioMed Research International. 2020.
13. Zhao K, Wang H, Wu C. The immune responses of HLA-A* 0201 restricted SARS-CoV S peptide-specific CD8 + T cells are augmented in varying degrees by CpG ODN, PolyI: C and R848. Vaccine. 2011;29(38):6670–8.
14. Vanderpuye V, Elhassan MMA, Simonds H. Preparedness for COVID-19 in the oncology community in Africa. The Lancet Oncology. 2020.
15. Gilbert M, Pullano G, Pinotti F, Valdano E, Poletto C, Boëlle P-Y, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. The Lancet. 2020;395(10227):871–7.
16. Hall T. BioEdit: an important software for molecular biology. GERF Bull Biosci. 2011;2(1):60–1.
17. Bui H-H, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinform. 2007;8(1):361.
18. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular biology evolution. 2016;33(7):1870–4.
19. Gasteiger EGA, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31:3784–8.
20. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. Sep. 1997;1(17):3389–402. 25.
21. Eddy SR. Profile hidden Markov models. Bioinformatics (Oxford, England). 1998 Jan 1;14 (9):755 – 63.
22. Bateman ABE, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL. The Pfam protein families database. Nucleic Acids Res. 2002;30(1):276–80.
23. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction: Springer Science & Business Media; 2009.
24. Hitchcock AL, Auld K, Gygi SP, Silver PA.A subset of membrane-associated proteins is ubiquitinated in response to mutations in the endoplasmic reticulum degradation machinery. Proceedings of the National Academy of Sciences. 2003;100(22):12735-40.
25. Peng J, Schwartz D, Elias JE, Thoreen CC, Cheng D, Marsischky G, et al. A proteomics approach to understanding protein ubiquitination. Nature biotechnology. 2003;21(8):921–6.
26. Gomi M, Sonoyama M, Mitaku S. High performance system for signal peptide prediction: SOSUIsignal. Chem-Bio Inform J. 2004;4(4):142–7.
27. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 1997;25(17):3389–402.
28. Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schäffer AA, et al. Protein database searches using compositionally adjusted substitution matrices. FEBS J. 2005;272(20):5101–9.
29. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR. etal. The immune epitope database (IEDB) 3.0. Nucleic acids research. 2014;43(D1):D405-D12.

30. Yao B, Zheng D, Liang S, Zhang C. Conformational B-cell epitope prediction on antigen protein structures: a review of current algorithms and comparison with common binding site prediction methods. PloS one. 2013; 8(4).
31. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome research. 2006;2(1):2.
32. Emini EA, Hughes JV, Perlow D, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. Journal of virology. 1985;55(3):836–9.
33. Kolaskar A, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 1990;276(1–2):172–4.
34. Sun P, Ju H, Liu Z, Ning Q, Zhang J, Zhao X, et al. Bioinformatics resources and tools for conformational B-cell epitope prediction. Computational and mathematical methods in medicine. 2013; 2013.
35. Chan WM, Rogers SE, Nash SM, Buning PG, Meakin R. User's manual for Chimera grid tools, version 1.8. NASA Ames Research Center, URL: http://people nas nasa gov/~ rogers/cgt/doc/man html [cited 19 July 2006]. 2003.
36. Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. Open biology. 2013 Jan 8; 3 (1):120139.
37. Nielsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinform. 2009;10(1):296.
38. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinform. 2007;8(1):4.
39. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v. 2—a server for in silico prediction of allergens. J Mol Model. 2014;20(6):2278.
40. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP, et al. In silico approach for predicting toxicity of peptides and proteins. PloS one. 2013; 8(9).
41. Bui H-H, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinform. 2006;7(1):153.
42. Zhu J, Wang S, Bu D, Xu J. Protein threading using residue co-variation and deep learning. Bioinformatics. 2018 Jul 1;34 (13):i263-73.
43. Xu J. Distance-based protein folding powered by deep learning. arXiv preprint arXiv:1811.03481. 2018 Nov 8.
44. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. Template-based protein structure modeling using the RaptorX web server. Nature protocols. 2012 Aug;7(8):1511.
45. Ma J, Peng J, Wang S, Xu J. A conditional neural fields model for protein threading. Bioinformatics. 2012 Jun 15;28 (12):i59-66.
46. Ma J, Wang S, Zhao F, Xu J. Protein threading using context-specific alignment potential. Bioinformatics. 2013 Jun 19;29 (13):i257-65.
47. Beaufays J, Lins L, Thomas A, Brasseur R. In silico predictions of 3D structures of linear and cyclic peptides with natural and non-proteinogenic residues. J Pept Sci. 2012;18(1):17–24.
48. Shen Y, Maupetit J, Derreumaux P, Tufféry P. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. J Chem Theory Comput. 2014;10(10):4745–58.
49. Choe J, Kelker MS, Wilson IA. Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. Science. 2005;309(5734):581–5.
50. Tanji H, Ohto U, Shibata T, Miyake K, Shimizu T. Structural reorganization of the Toll-like receptor 8 dimer induced by agonistic ligands. Science. 2013;339(6126):1426–9.
51. Duhovny D, Nussinov R, Wolfson HJ, editors. Efficient unbound docking of rigid molecules. International workshop on algorithms in bioinformatics; 2002: Springer.
52. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic acids research. 2005;33(suppl_2):W363-W7.
53. Andrusier N, Nussinov R, Wolfson HJ. FireDock: fast interaction refinement in molecular docking. Proteins: Struct Funct Bioinf. 2007;69(1):139–59.
54. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Bioinformatics. 1992;8(3):275–82.
55. Binns MM, Boursnell ME, Cavanagh D, Pappin DJ, Brown TDK. Cloning and sequencing of the gene encoding the spike protein of the coronavirus IBV. J Gen Virol. 1985;66(4):719–26.
56. Prabakaran P, Gan J, Feng Y, Zhu Z, Choudhry V, Xiao X, et al. Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. J Biol Chem. 2006;281(23):15829–36.
57. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020;367(6483):1260–3.
58. Wang Y, Rattray JB, Thomas SA, Gurney J, Brown SP. In silico bacteria evolve robust cooperation via complex quorum-sensing strategies. BioRxiv. 2019:598508.
59. 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. Journal of translational medicine. 2019 Dec 1; 17(1):362.
60. Haste Andersen P, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Sci. 2006;15(11):2558–67.
61. Baruah V, Bose S. Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. Journal of Medical Virology. 2020.
62. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic acids research. 2003;31(13):3784–8.

63. Negahdaripour M, Nezafat N, Eslami M, Ghoshoon MB, Shoolian E, Najafipour S, et al. Structural vaccinology considerations for in silico designing of a multi-epitope vaccine. Infection Genetics Evolution. 2018;58:96–109.
64. Lester SN, Li K. Toll-like receptors in antiviral innate immunity. Journal of molecular biology. 2014;426(6):1246–64.
65. Mukherjee S, Karmakar S, Babu SPS. TLR2 and TLR4 mediated host immune responses in major infectious diseases: a review. Brazilian Journal of Infectious Diseases. 2016;20(2):193–204.

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.

No.	Accession No.	Country	Year
1	YP_009724390.1*	China	2020
2	QHR63290.2	China	2020
3	QHR63280.2	China	2020
4	QHR63270.2	China	2020
5	QHR63260.2	China	2020
6	QHR63250.2	China	2020
7	QIC53213.1	China	2020
8	QIA20044.1	China	2020

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

No	N-terminal	Transmembrane region	C-terminal	Type	Length
1	1	MFVFLVLLPLVSSQCVNLTTRT	22	Secondary	22
2	1223	GLIAIVMVTIMLCCMTSCCSCLK	1245	Primary	23

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.

NCBI Protein ID	Protein Name	E- value	Identity
YP_009724390.1	Severe acute respiratory syndrome-related coronavirus	0.0	75.96%
YP_003858584.1	Bat coronavirus BM48-31/BGR/2008	0.0	71.96%
YP_009273005.1	Rousettus bat coronavirus	0.0	35.86%
YP_009047204.1	Middle East respiratory syndrome-related coronavirus	1e-176	35.10%
YP_009555241.1	Human coronavirus OC43	2e-145	37.63%
YP_209233.1	Murine hepatitis virus strain JHM	4e-142	36.65%
YP_009194639.1	Camel alphacoronavirus	7e-109	31.54%
YP_003767.1	Human coronavirus NL63	4e-103	30.78%
YP_001941166.1	Turkey coronavirus	4e-103	36.92%
NP_040831.1	Infectious bronchitis virus	2e-101	35.91%
YP_004070194.1	Feline infectious peritonitis virus	9e-99	31.95%
NP_058424.1	Transmissible gastroenteritis virus	1e-98	31.98%
YP_009199242.1	Swine enteric coronavirus	2e-93	30.60%
NP_598310.1	Porcine epidemic diarrhea virus	4e-91	30.44%

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.

*Epitope sequence	length	Start	End	+Emini	#Kolaskar	Allergenicity	Toxinicity
LDSK	4	110	113	1.497	1.014	Allergen	Non Toxin
RVYST	5	634	638	1.426	1.068	Allergen	Non Toxin
**QSAPH	5	1054	1058	1.597	1.052	Non-Allergen	Non Toxin
KAHFP	5	1086	1090	1.191	1.051	Allergen	Non Toxin
VYDPLQPELDSF	12	1137	1148	1.279	1.073	Allergen	Non Toxin

Table 5: Discontinuous epitopes of SARS-CoV-2 spike S protein *predicted through DISCOTOPE 2.0 Server. Parameter was set at ≥ 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
810	SER	9	-0.669	-1.627
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	19	-1.693	-3.684
1118	ASP	4	-3.016	-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.

Epitopes	Start	End	Antigenicity	Alleles
FAMQMAYRF	898	906	1.0278	HLA-A*02:06; HLA-A*23:01; HLA-A*24:02; HLA-A*29:02; HLA-B*08:01; HLA-B*15:01; HLA-B*35:01; HLA-B*53:01; HLA-B*58:01; HLA-C*03:03; HLA-C*05:01; HLA-C*12:03
WTAGAAAYY	258	266	0.6306	HLA-A*01:01; HLA-A*26:01; HLA-A*29:02;HLA-A*30:02; HLA-A*68:01; HLA-A*68:02;HLA-B*15:01; HLA-B*35:01; HLA-B*58:01
FVFLVLLPL	2	10	0.8601	HLA-A*02:01; HLA-A*02:06; HLA-A*68:02;HLA-B*35:01; HLA-B*39:01; HLA-C*03:03;HLA-C*12:03; HLA-C*14:02
KIYSKHTPI	202	210	0.7455	HLA-A*02:01; HLA-A*02:06; HLA-A*30:01;HLA-A*32:01; HLA-C*03:03; HLA-C*14:02;HLA-C*15:02
FTISVTTEI	718	726	0.8535	HLA-A*02:01; HLA-A*02:06; HLA-A*68:02; HLA-B*58:01; HLA-C*03:03; HLA-C*12:03; HLA-C*15:02
IAIPTNFTI	712	720	0.7052	HLA-A*02:06; HLA-A*23:01; HLA-B*53:01; HLA-B*58:01; HLA-C*03:03; HLA-C*12:03
*VVFLHVTYV	1060	1068	1.5122	HLA-A*02:01; HLA-A*02:06; HLA-A*68:02; HLA-C*06:02; HLA-C*07:01; HLA-C*12:03

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

Epitopes	MW*	Theoretica pI	Estimated half-life	Formula	Ext. coefficient*	II*
*FAMQMAYRF	1164.41	8.75	1.1 hours (mammalian reticulocytes, in vitro)	$C_{54}H_{77}N_{13}O_{12}S$	1490	stable
WTAGAAAYY*	973.05	5.52	2.8 hours (mammalian reticulocytes, in vitro)	$C_{47}H_{60}N_{10}O_{13}$	8480.	stable
FVFLVLLPL*	1060.39	5.52	1.1 hours (mammalian reticulocytes, in vitro)	$C_{46}H_{69}N_{13}O_{13}$	should not be visible by UV spectrophotometry.	stable
KIYSKHTPI	1086.30	9.70	1.3 hours (mammalian reticulocytes, in vitro)	$C_{51}H_{83}N_{13}O_{13}$	1490	stable
FTISVTTEI	1010.15	4.00	1.1 hours (mammalian reticulocytes, in vitro)	$C_{46}H_{75}N_9O_{16}$	should not be visible by UV spectrophotometry.	unstable
IAIPTNFTI	989.18	5.52	20 hours (mammalian reticulocytes, in vitro)	$C_{47}H_{76}N_{10}O_{13}$	should not be visible by UV spectrophotometry.	stable
VVFLHVTYV	1076.30	6.71	100 hours (mammalian reticulocytes, in vitro)	$C_{54}H_{81}N_{11}O_{12}$	1490	stable

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

Epitopes	MW*	Theoretical pI	Estimated half-life	Formula	*Ext. coefficient	II*	*GRAVY
FAMQMAYRF	1164.41	8.75	(hours (mammalian reticulocytes, in vitro 1.1	$C_{54}H_{77}N_{13}O_{12}S$	1490	stable	0.411
FGAGAALQI	846.98	5.52	(hours (mammalian reticulocytes, in vitro 1.1	$C_{39}H_{62}N_{10}O_{11}$	should not be visible by UV spectrophotometry.	stable	1.356
FNATRFASV	1012.13	9.75	(hours (mammalian reticulocytes, in vitro 1.1	$C_{46}H_{69}N_{13}O_{13}$	should not be visible by UV spectrophotometry.	stable	0.433

Table 9: Population Coverage for Proposed MHCI and II Epitopes.

Epitopes	MHCI Coverage	Epitopes	MHCII Coverage
FAMQMAYRF			
FNATRFASV		FAMQMAYRF	
FGAGAALQI	95.74	WTAGAAAYY	78.09
KIYSKHTPI		FVFLVLLPL	
IAIPTNFTI			
VVFLHVTYV			

Table 10: Docking of MHCI epitopes with TLR8. Molecular docking was performed using Patch Dock online autodock tools by submitting MHCI predicted epitopes and 3D structures of TLR8 (PDB: 3W3G). VVFLHVTYV produced low global energy which indicates strong binding affinity.

No	MHCI epitopes	Global Energy	Attractive VdW
1	FAMQMAYRF	-49.69	-29.90
2	WTAGAAAYY	-37.73	-27.38
3	FVFLVLLPL	-64.23	-24.16
4	KIYSKHTPI	-35.81	-25.05
5	IAIPTNFTI	-56.95	-27.05
6	VVFLHVTYV	-84.58	-36.05

Table 11: Docking of MHCI epitopes with HLA-B7. Molecular docking was achieved using Patch Dock online autodock tools by submitting MHCI predicted epitopes and 3D structures of HLA-B7 (BDP: 3VCL). FVFLVLLPL showed low global energy which indicates strong binding affinity.

No	MHCI epitopes	Global Energy	Attractive VdW
1	FAMQMAYRF	-52.00	-26.51
2	WTAGAAAYY	-51.14	-23.90
3	FVFLVLLPL	-78.81	-33.08
4	KIYSKHTPI	-46.40	-23.32
5	IAIPTNFTI	-52.94	-26.75
6	VVFLHVTYV	-63.20	-33.01

Figures

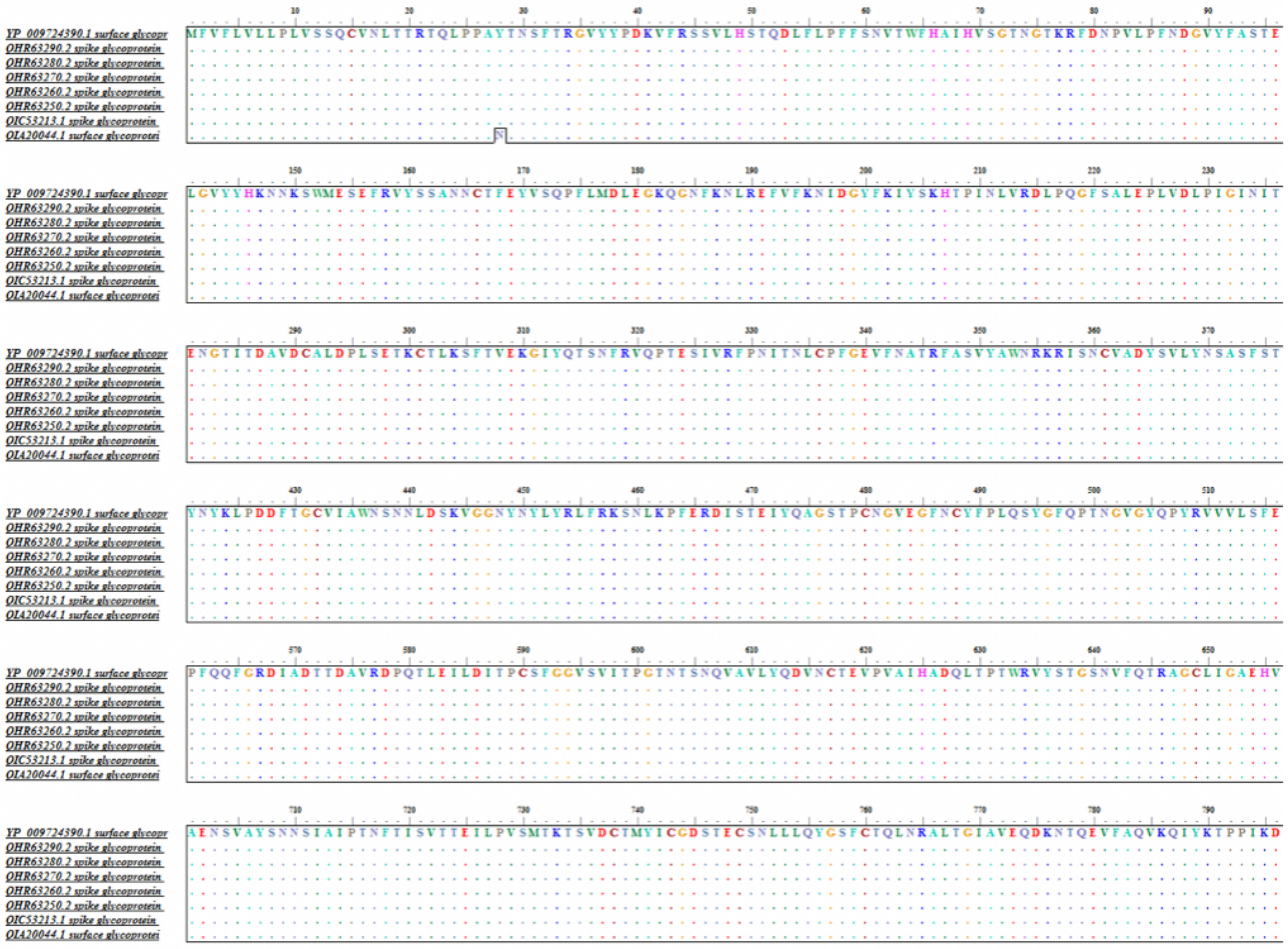

Figure 1

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.

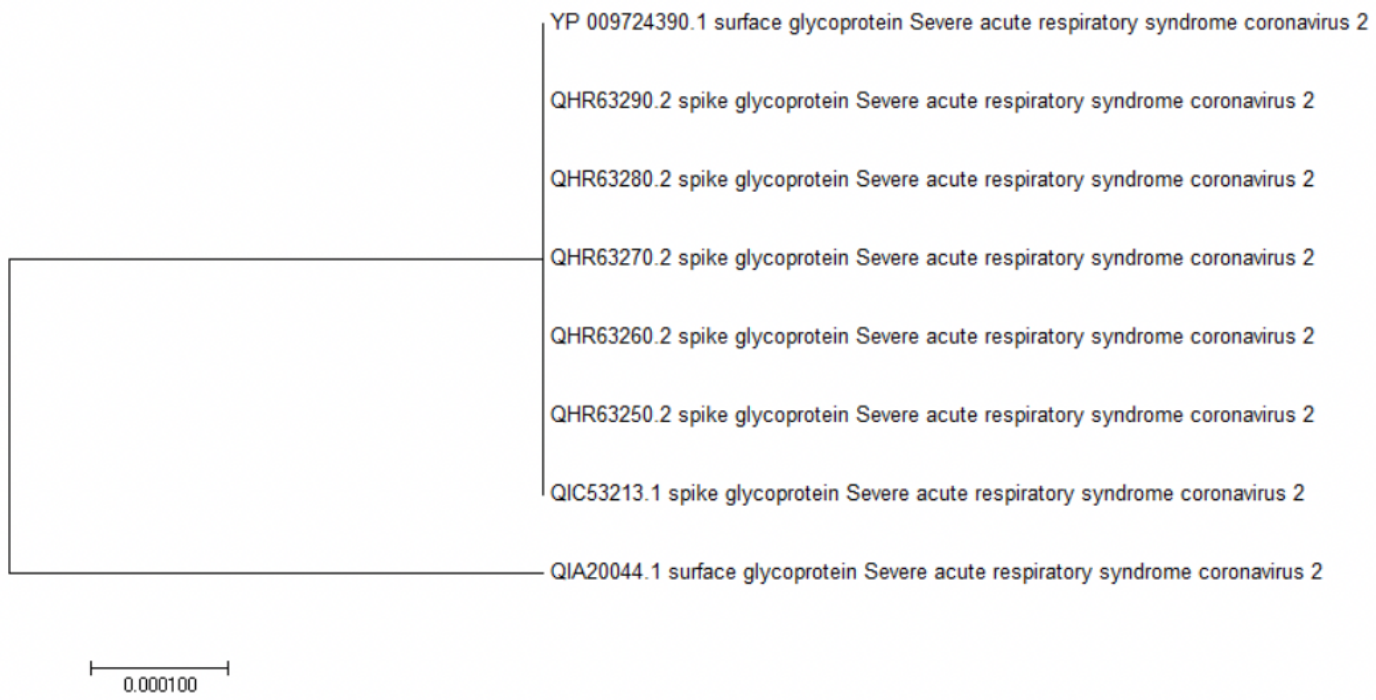

Figure 2

Phylogenetic tree of SARS-CoV-2 spike S protein using MEGA7 software using maximum likelihood parameter.

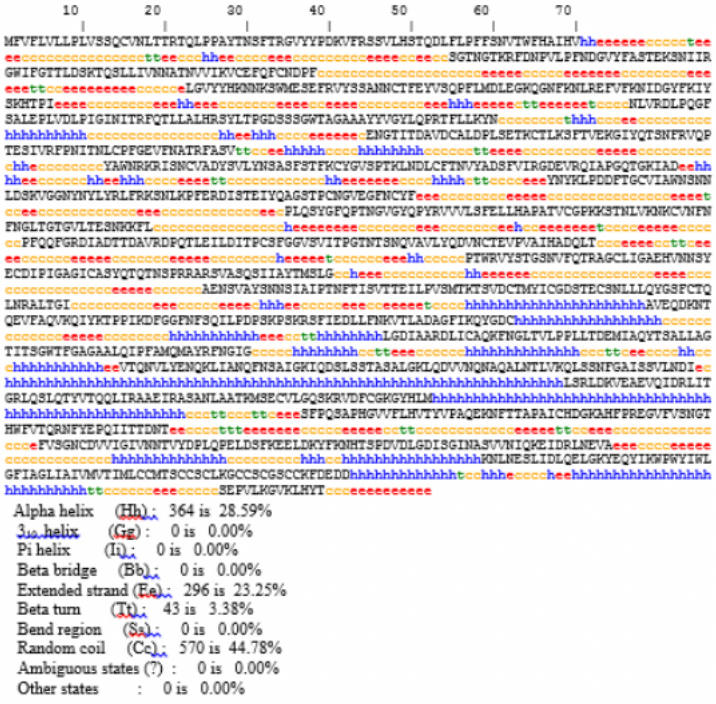

Figure 3

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

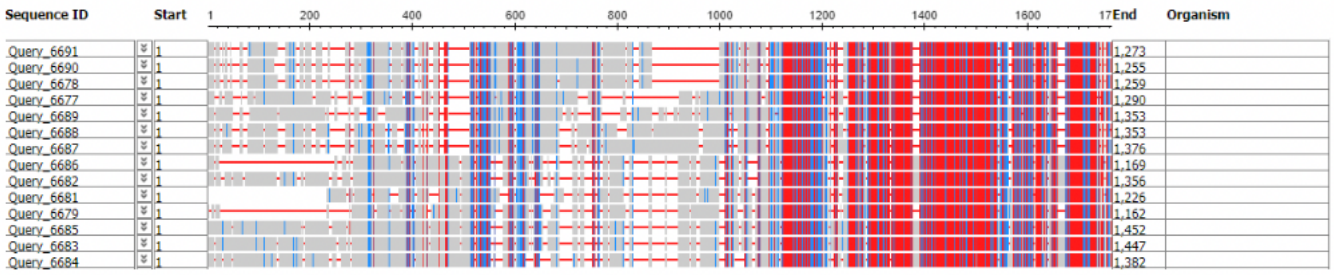

Figure 5

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.

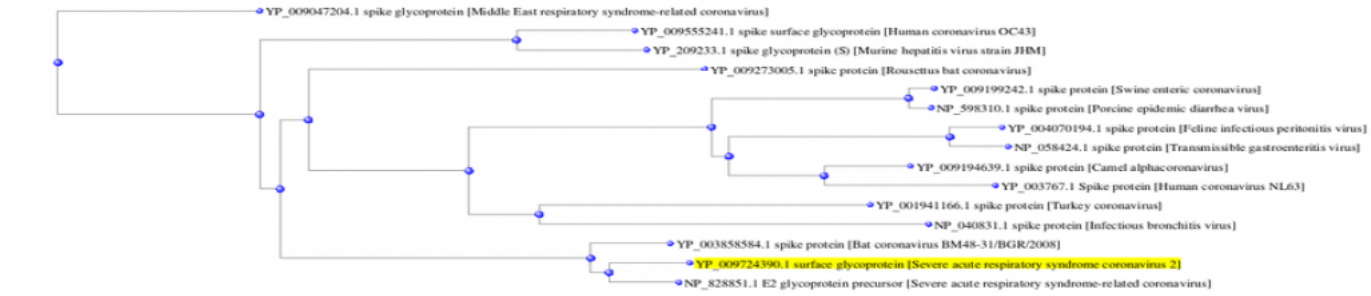

Figure 6

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).

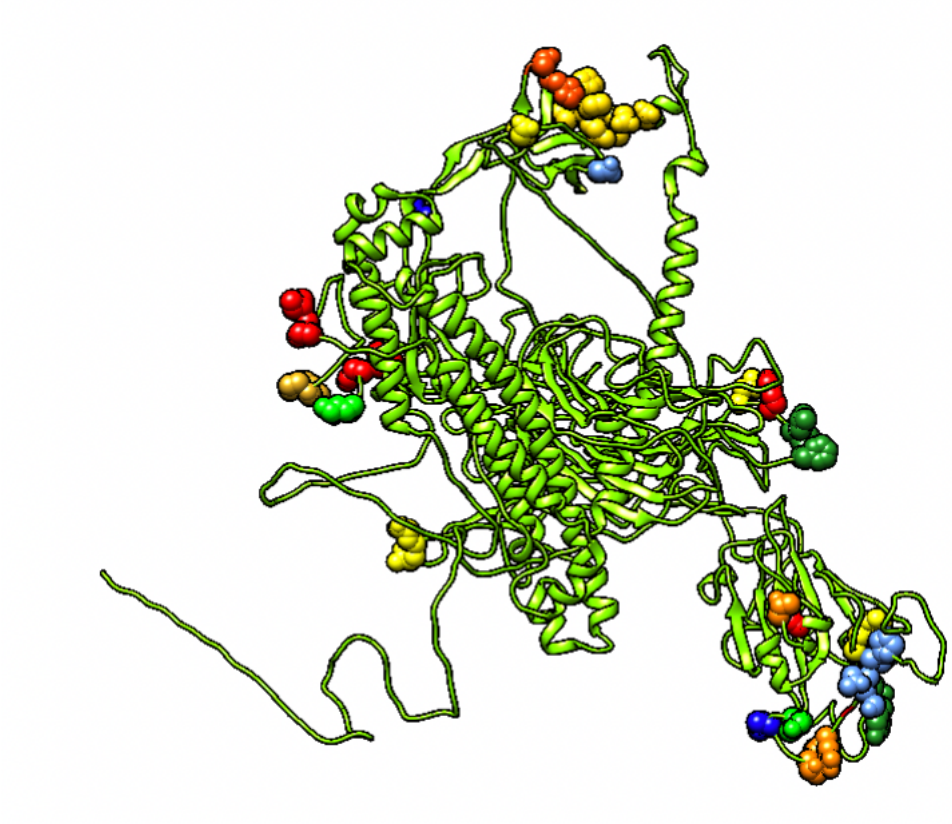

Figure 7

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.

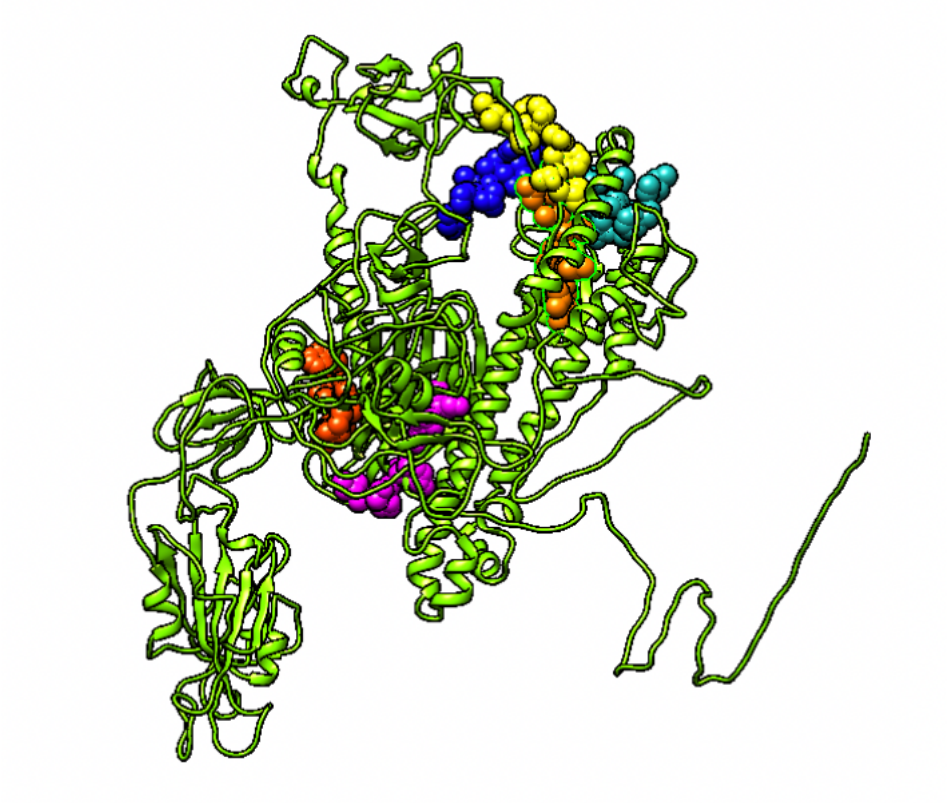

Figure 8

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

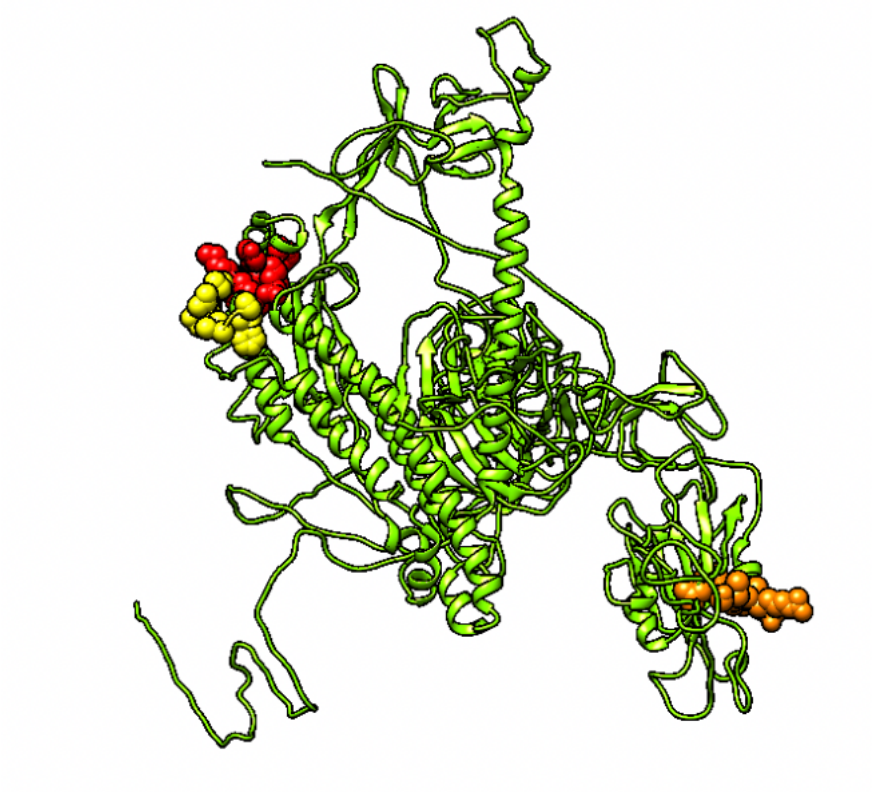

Figure 9

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

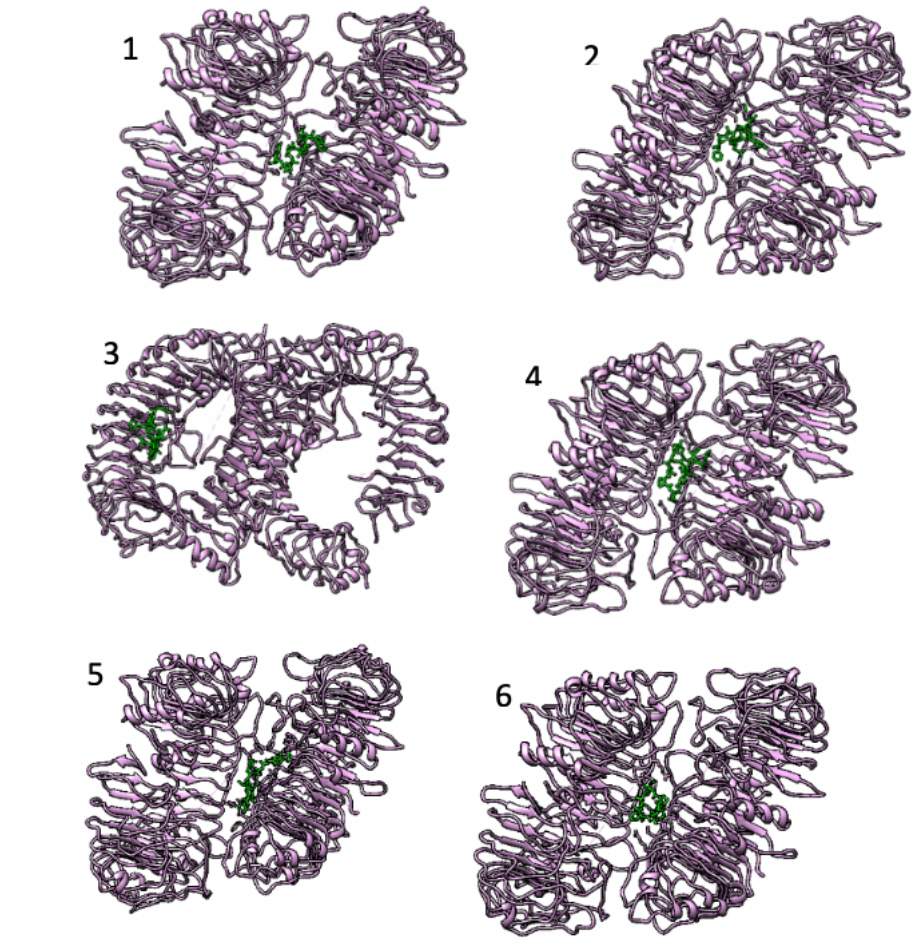

Figure 10

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.

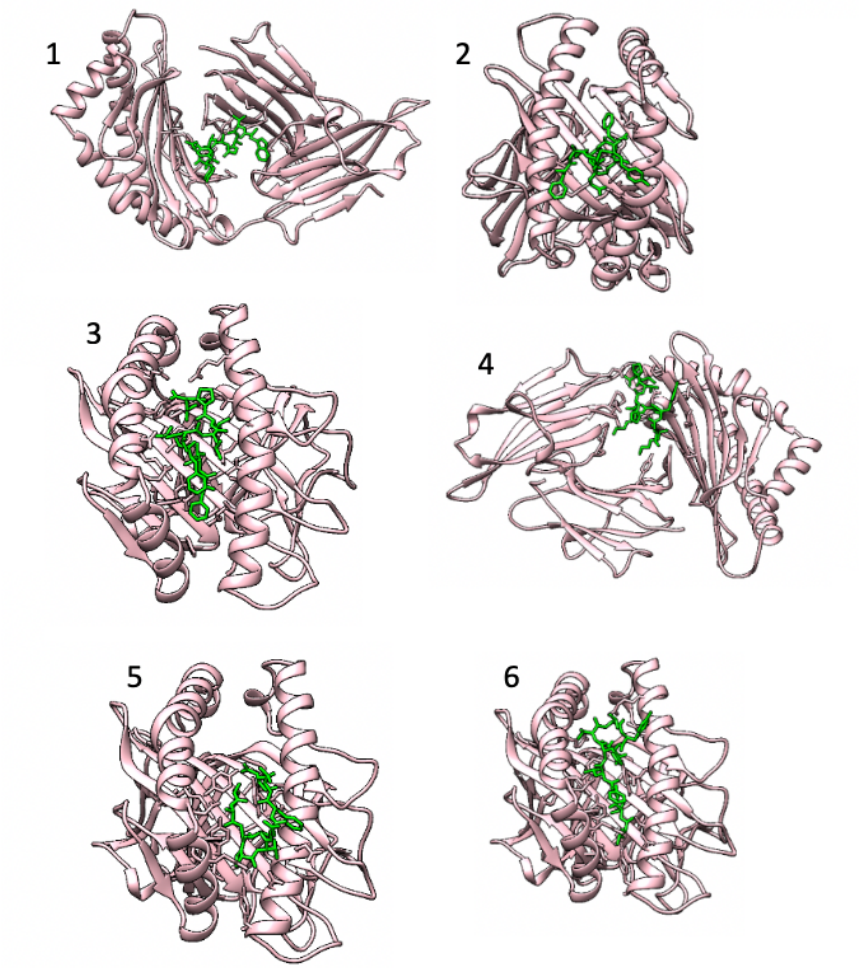

Figure 11

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