Structural modeling and conserved epitopes prediction against SARS-COV-2 structural proteins for vaccine development

Muhammad Tahir ul Qamar

College of Life Science and Technology, Guangxi University, Nanning, P. R. China <https://orcid.org/0000-0003-4832-4250>

Farah Shahid

Department of Bioinformatics and Biotechnology, Government College University Faisalabad (GCUF), Faisalabad, Pakistan

Usman Ali Ashfaq

Department of Bioinformatics and Biotechnology, Government College University Faisalabad (GCUF), Faisalabad, Pakistan

Sidra Aslam

Department of Bioinformatics and Biotechnology, Government College University Faisalabad (GCUF), Faisalabad, Pakistan

Israr Fatima

Department of Bioinformatics and Biotechnology, Government College University Faisalabad (GCUF), Faisalabad, Pakistan

Muhammad Mazhar Fareed

Department of Bioinformatics and Biotechnology, Government College University Faisalabad (GCUF), Faisalabad, Pakistan

Ali Zohaib

Department of Healthcare Biotechnology, Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad, Pakistan

Ling-Ling Chen (✉ llchen@mail.hzau.edu.cn)

College of Life Science and Technology, Guangxi University, Nanning, P. R. China

Research

Keywords: SARS-CoV-19, COVID-19, Structural Proteins, Vaccine, Immuno-informatics, Epitopes

Posted Date: February 19th, 2020

DOI: <https://doi.org/10.21203/rs.2.23973/v1>

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

Abstract

Background: Coronavirus disease 2019 (COVID-19) caused by Severe Acute Respiratory Syndrome Corona virus 2 (SARS-CoV-2) was first diagnosed in December 2019, Wuhan, China. Little is known about this new virus and it has the potential to cause severe illness and pneumonia in some people, therefore the development of an effective vaccine is highly desired.

Methods: Immunoinformatics and statistical approaches were used in this study to forecast B- and T- cell epitopes for the SARS-CoV-2 structural proteins (Surface glycoprotein, Envelope protein, and Membrane glycoprotein) that may play a key role in eliciting immune response against COVID-19. Different types of B cell epitopes (linear as well as discontinuous) and T cell (MHC class I and MHC class II) were determined. Moreover, their antigenicity and allergenicity were also estimated.

Results: The antigenic B-cell epitopes exposed to the outer surface were screened out and 23 linear B cell epitopes were selected. "SPTKLNLCFTNVY" had the highest antigenicity score among B cell epitopes. The T-cell epitopes bound to multiple alleles, antigenic, non-allergen, non-toxic, and conserved in the protein sequence were shortlisted. In total, 16 epitopes (9 from MHC class I and 7 from MHC class II) were selected. Among the T-cell epitopes, MHC class I (IPFAMQMAYRFN) and MHC class II (VTLCFVLAAYRIN) were classified as strongly antigenic. Digestion analysis verified the safety and stability of the peptides predicted during this study. Furthermore, docking analyses of predicted peptides showed significant interactions with the HLA-B7 allele.

Conclusion: The putative antigen epitopes identified in this study may serve as vaccine candidates and can help to eliminate/control growing health threat of COVID-19.

Background

Viruses have the potential to become dangerous life threat and cause irreparable loss to human beings. Hardly the world learns to cope with one strain of virus when another emerges and poses a threat to the future of humanity. A similar situation has emerged when a new strain of novel coronavirus (CoV) that has not been previously identified in humans is reported last year late December [1]. Coronaviruses are the largest among RNA viruses belonging to *Coronaviridae*, *Roniviridae* and *Arteriviridae* families. Coronaviridae are unsegmented, 3' polyadenylated and 5' capped positive sense single-stranded RNA viruses cause various respiratory diseases in humans [2]. CoV is classified into 4 classes; alpha, beta, delta, and gamma. Amongst them, beta and alpha CoVs have been reported for infecting humans [3]. CoV continuously changes its strains making it a herculean task to develop a vaccine. Recent CoV strain has received tremendous attention from researchers, as it causes severe respiratory complications similar to its other family members. [4]. According to the WHO reports, China was intimated about the cases of pneumonia having unknown etiology in Wuhan on December 31, 2019 [1, 5]. Later, SARS-CoV-2 was identified as the causative agent of this outbreak by the CCDCP (Chinese Center for Disease Control and Prevention) [6]. Currently, there is a paucity of research regarding its transmission among people. Initial diagnostic procedures indicate that the SARS-CoV-2 is primarily spread through respiratory droplets from sneezing/coughing, body contact and to some extent through fecal contact [7]. The SARS-CoV-2 can exhibit symptoms in less than 2 days or up to 14 days after exposure. Symptoms of patients infected with COVID-19 include fever, runny nose, cough, and dyspnea [8]. Although the entire genome sequence of the virus has been published, the origin and proliferation mechanism of the new coronavirus is still ambiguous as stated by the WHO [9]. The study of genome sequences has cast a shadow that SARS-CoV-2 is closely related to the SARS-CoV which is the causative agent of the SARS outbreak in humans in 2002/03 [10].

The SARS-CoV-2 outbreak initially was related to the Huanan seafood market in the South China region. However, major gaps in knowledge of the origin, human transmission period, clinical spectrum, and epidemiology of COVID-19 need fulfillment by future studies. Some patients diagnosed with disease revealed that they had searched or wandered the market [11]. From shellfish to snakes, birds and other small mammals including marmots, the market is said to be the selling point of a variety of animals that are purported to be the cause of dissemination of the virus. WHO reported that environmental samples of the market were positive for the SARS-CoV-2. However, no specific animal relationships were sorted out yet [12]. Initial reports claimed that snakes could be a possible source due to codon usage [13], but the claim is under debate and needs substantial research to prove it [6]. Researchers are currently working to sort out the SARS-CoV-2 source, including possible intermediate animal vectors.

The samples taken from a respiratory system-throat swab or lung fluid- are helpful in diagnosing its infection in patients. Unfortunately, there is no vaccine or therapy available to cure this disease [14]. Therefore, an effective remedy to cope with this disease is direly needed to contain this rapidly spreading menace. For this purpose, immunoinformatics approaches can be applied to examine viral antigens, prediction of its epitopes and assessment of its immunogenicity [15]. Moreover, this approach could be both time and cost-effective [2, 16, 17]. Excessive respiratory infection can also resolve with T cell reactions and antibodies [18]. Furthermore, rapid identification, isolation, disease prevention, and control measures are required to hinder its spread of SARS-CoV-2 at homes, communities and healthcare units [19, 20]. In various studies, therapeutic approaches against the Ebola virus, Zika virus and MERS-CoV were developed using in silico peptide prediction [2, 17, 21]. The purpose of this study was to pinpoint the potential T-cell and B-cell epitopes from SARS-CoV-2 structural proteins which would help in the future to develop its vaccine.

Methods

Flow chart of methodology used in present study is graphically represented in Additional file 1: Figure S1. Complete details about software/resources and their availability URLs enlisted in Additional file 2. Table S1.

Data retrieval

Three proteins, Surface glycoprotein (S), Envelope protein (E) and Membrane glycoprotein (M) of SARS-COV-2 were taken as targets. Their amino acid sequences were collected in fasta format from Genbank [22].

Structural analysis

The ExPASy ProtParam tool was used to evaluate the physical and chemical properties of target proteins [23]. Protein's secondary structure was analyzed through PSIPRED. The transmembrane topology of the proteins was checked through the TMHMM tool. DIANNA v1.1, an online tool, checked the presence of disulphide-bonds in proteins [24]. It makes predictions based on the neural trained system. Allergenicity and antigenicity of proteins were checked through AllerTOP v2.0 and VaxiJen v2.0 respectively [25, 26].

3D structure prediction and validation

3D structures of SARS-COV-2 proteins (S, E, and M) were predicted using homology modeling approach. PDB database was used to find appropriate templates. PDB homologous proteins were scanned by PSI-blast [27]. The initial alignment between template and target was created using the ALIGN2D tool. The 3D structures were built in MODELLER v.9.12 using a restrained-based approach [28]. These models were then visualized by Chimera [29]. Galaxy refines server and ModRefiner was used to refine the retrieved models [30, 31]. Besides, the refined structure needs to be validated based on experimentally validated 3D structure of proteins. Refined structures were therefore applied in the PROSA web providing a quality score for a given structure [32]. The quality score beyond the usual range of native proteins indicates a possible error in protein structure. Ramachandran plot was created by rampage server where the principle of PROCHECK is applied to validate the protein structure [33].

Prediction of B cell epitopes

The epitopes of B cells help to detect viral infections in the immune system. ABCpred was used to forecast 14-mer B cell epitopes for target proteins at 0.51 threshold [34]. Epitopes evident on the outer surface were picked, and other intracellular epitopes were removed. The vaxijen server tested the antigenicity of the selected epitopes at a threshold of 0.5. B cell epitope identification was based upon antigenicity, flexibility, linear epitope predictions, hydrophilicity, and surface accessibility [35]. Parker hydrophilicity prediction algorithms, Emini surface accessibility prediction method, Kolaskar and Tongaonkar antigenicity scale, and Karplus and Schulz flexibility prediction tool were used to perform hydrophilicity, accessibility of surface, antigenicity and flexibility analysis respectively [36]. As discontinuous epitopes become more evident and have higher dominant properties than linear epitopes, DiscoTop 2.0 server was used to forecast discontinuous epitopes from 3D structures of surface glycoprotein, membrane protein and envelope protein [37]. The position of epitopes on 3D structures of proteins was visualized by Pymol [38].

Prediction of T cell epitopes

In vaccine designation, T cell epitopes play a crucial role. More specifically, it reduces the cost and time compared with laboratory experiments [39]. IEDB consensus method was used to predict 12 mer MHC class I and 15 mer MHC class II epitopes. The results of this method are very important due to a large number of HLA alleles used in the calculation. The sequence was given in a FASTA format and all the alleles were selected for prediction. Epitopes with less than 2 consensus score believed to be good binders and chosen for further research.

Epitopes conservation analysis

The degree of conservation of predicted T cell and B cell epitopes within the protein sequence was analyzed by IEDB conservancy analysis tool. Epitopes having 100% conservancy were shortlisted [40].

Population coverage analysis

The expression and distribution of HLA alleles vary depending on the world's ethnicities and regions, thereby impacting the effective production of an epitope-based vaccine [41]. The population coverage was calculated using the IEDB population coverage tool and for this purpose selected MHC class I and MHC class II epitopes and corresponding HLA-binding alleles were considered. This tool estimates population coverage for each epitope for various regions of the world based on the distribution of human alleles that bind to MHC [42]. However, this study will highlight areas of particular importance concerning our pathogen.

Properties evaluation of peptides

Antigenicity and allergenicity of the peptides were checked by Vaxijen v 2.0 and Allergen Fp 1.0 respectively [43]. The threshold of VaxiJen was set at 0.5. Physiochemical properties of the peptides like Theoretical pI, the molecular weight was evaluated by ExPASy ProtParam tool. Protein digesting enzyme server was used to predict peptide digester enzymes. ToxinPred is used for Non-toxic/toxic peptide prediction. Non-toxic peptides were selected for further analysis [44].

Modeling of peptides and docking with HLA allele

The 3D structures of nine MHC class I epitopes were predicted using PEPFOLD. Five models for each peptide were created and the best model for each peptide was picked [45]. The crystallographic X-ray structure of the commonly found HLA allele in the human population i.e HLA-B7 allele was obtained from PDB (PDB ID: 3VCL). Molecular docking was performed using similar protocol of our already published studies [1, 2, 16, 46, 47].

Results

Sequence retrieval and structural analysis of the target proteins

Amino acid sequences of target proteins (S [Genbank: QHD43416.1], E [Genbank: QHD43418.1] and M [QHD43419.1]) of SARS-COV-2 were collected in fasta format from Genbank. Their allergenicity and antigenicity were checked by AllerTop and vaxijen respectively. All proteins were found to be non-allergenic and antigenic. E protein was considered to be the most antigenic followed by M and S protein with 0.60, 0.51 and 0.46 antigenic values respectively. Besides, 40 S protein disulphide bonds, 3 E protein disulphide bonds and 6 M protein disulphide bonds calculations were made in DIANNA v1.1 (Additional File 3: Table S2). Physiochemical properties evaluated by protparam showed that all proteins are stable. S protein was found to be acidic and hydrophilic, while E and M protein was found to be basic and hydrophobic (Table 1). TMHMM carried out a prediction of the transmembrane topology of the target proteins displaying one transmembrane helix in S and E protein while three transmembrane helices in M protein. Residues of the S protein from 1-1213 were on the surface while residues from 1214-1236 were within the transmembrane region and residues from 1237-1273 were buried within the core region of S protein. Similarly, in E protein residues from 1-11 were exposed on the surface and residues from 12-34 were inside the transmembrane region and residues from 35-75 were buried inside the core region of E protein. In M protein residues from 1-19 and 74-77 were exposed on the surface, while residues from 20-39, 51-73, and 78-100 were inside the transmembrane region and residues from 40-50 and 101-222 were buried inside the core region. Secondary structure details of the target proteins predicted by PSIPRED are mentioned in Additional File 4: Table S3.

Homology modeling and validation

To predict structure-based epitopes of SARS-COV-2 proteins (S, E, and M), homology modeling of the proteins was performed. Chain A, Spike glycoprotein of SARS-COV (PDB ID: 6ACC) and Chain A, Envelope small membrane protein of SARS-COV (PDB ID: 5X29) were found to be the best template based on E-value, percent identity, and query coverage for S and E protein respectively. S protein was found to have a 75% identity with a spike protein of SARS-COV (6ACC). E protein had 88.71% identity with Chain A, Envelope small membrane protein of SARS-COV. No suitable template was found for M protein, so its structure was predicted by Raptor X [48]. The amino acid sequence was provided as an input that yielded to have two domains and the best template used was model 5yckA. The P-value for the modeled structure was calculated at 9.14×10^{-5} . All 222 (100%) amino acids were modeled and 27 (12%) positions were predicted as disordered in the model. Visualization of the models was done by Chimera (Additional file 1: Figure S2). Structures were refined by Galaxy refine server and ModRefiner. The quality factor (z-score) and Ramachandran plot values of refined models are mentioned in Additional File 5: Table S4 (Additional file 1: Figure S3-S5).

B cell epitopes prediction

ABCpred was used to predict the linear B cell epitopes of the target proteins and the TMHMM server was used to check the surface availability. Antigenicity was checked by vaxijen. Epitopes that were exposed on the surface, antigenic, and 100% conserved in the protein were selected from all the predicted epitopes. Total 23 epitopes (S-19, E-1, and M-3) were selected based on these criteria. Among the chosen epitopes, 'SPTKLNDLCFTNVY' of S protein showed the highest antigenicity (1.69) and predicted score (Table 2). The position of epitopes on their respective protein structure was visualized by Pymol (Figure 1).

Besides, potential B cell surface accessibility needs to be examined. By analyzing the physiochemical properties of amino acids and their concentration in previously identified B cell epitopes, the Kolaskar and Tongaonkar antigenicity tools evaluated target proteins for predicting B cell epitopes. The estimation threshold was set at 1.045, while the window size was maintained 7. It predicted the protein antigenic propensity value of S protein as 1.041 (Average), 0.866 (Minimum) and 1.261 (Maximum) (Additional file 1: Figure S6-A), of E protein as 1.119 (Average), 0.947 (Minimum) and 1.262 (Maximum) (Additional file 1: Figure S7-A) and of M protein as 1.053 (average), 0.904 (minimum), and 1.235 (maximum) (Additional file 1: Figure S8-A). Chou and Fasman beta-turn analyzing algorithm was used to predict beta-turn in target proteins since beta-turn in nature is exposed to the surface and hydrophilic and plays a vital role in beginning the defensive response. The threshold value was adjusted at 1.009, it calculated values of 0.097 (average), 0.541 (minimum) and 1.484 (maximum) in S protein (Additional file 1: Figure S6-B), 0.883 (average), 0.554 (minimum) and 1.264 (maximum) in E protein (Additional file 1: Figure S7-B), and 0.915 (average), 0.600 (minimum) and 1.384 (maximum) in M protein (Additional file 1: Figure S8-B). The findings show that area from 251 to 257 amino acids in S protein, area from 63-69 amino acids and 67-73 amino acids in E protein and area from 209-216 amino acids in M protein are more likely to reassure beta turns in peptide structure. Experimental experience shows that the parts of the epitope bound with antibodies or alleles are essentially elastic. Hydrophilic protein regions are usually exposed on the surface and play a key role in eliciting an immune response. ABCpred score and the antigenic value calculated by vaxijen certainly show that all expected peptides are part of the transmembrane-protein extracellular region and capable of optimizing a defensive response within the host during COVID-19 infection. Therefore, to determine the hydrophilicity and the surface abundance of possible B cell epitopes, the parker surface accessibility prediction method having a threshold of 1.279 and Emini surface accessibility prediction methods having a threshold of 1.00 were used. Values calculated by parker-hydrophilicity was 1.238 (average), -7.629 (minimum) and 7.743 (maximum) in S protein (Additional file 1: Figure S6-C), -0.911 (average), -6.843 (minimum) and 4.929 (maximum) in E protein, and -0.499 (average) (Additional file 1: Figure S7-C), -9.257 (minimum) and 6.871 (maximum) in M protein (Additional file 1: Figure S8-C). Emini surface accessibility analyzing results are given in the Additional File 6: Table S5 and Additional file 1: Figure S6-D, S7-D, S8-D. Flexibility analysis by Karplus and schulz tool showed that area from 251-257 amino acids in S protein (Additional file 1: Figure S6-E), 65-71 amino acids in E protein (Additional file 1: Figure S7-E) and area from 210-216 amino acids in M protein (Additional file 1: Figure S8-E) are highly versatile.

To further improve the specificity and variety of B-cell epitopes, Discotop 2.0 server was used to calculate surface abundance concerning residual contact number and use the novel amino acid score to forecast discontinuous epitopes. 3D structures of the target proteins were used to predict discontinuous epitopes; 90% specificity, - 3.700 thresholds and 22.000 Angstroms propensity score radius. 138 discontinuous epitopes of S protein, 1 epitope of the E protein and 22 epitopes of M protein were calculated (Table 3). The position of epitopes on their respective protein structure was visualized by Pymol (Figure 2).

T cell epitopes prediction and properties evaluation

T cell (MHC class I & MHC class II) epitopes of target proteins were predicted by the IEDB consensus method. Peptides that can bind to multiple alleles are considered the most appropriate peptides due to their strong defensive ability. Their antigenicity and allergenicity were checked by Vaxijen and Allergen FP 1.0. Their conservation in protein sequences was analyzed by the IEDB conservancy analysis tool. Epitopes that are bound to multiple alleles, highly antigenic, non-allergenic and 100% conserved were screened out. Based on these criteria, 9 MHC class I (S-3, E-3, and M-3) and 7 MHC class II (S-1, E-3 and M-3) were shortlisted. Between MHC class I epitopes 'IPFAMQMAYRFN' of S protein suggested higher antigenicity score 1.3 binding with multiple alleles including HLA-B*35:01, HLA-B*35:03, HLA-A*24:02, HLA-B*51:01, HLA-B*53:01. The peptide 'VTLCFVLAAYRIN' of M protein was considered most antigenic for its higher antigenicity score 1.0 between MHC class II epitopes and it was bound to multiple alleles including HLA-DRB1*07:03, HLA-DRB1*11:20, HLA-DRB1*01:02, HLA-DRB1*07:01, HLA-DRB1*11:14, HLA-DRB1*13:23, HLA-DRB1*03:09, HLA-DRB1*13:07, HLA-DRB1*11:28, HLA-DRB1*13:05, HLA-DRB1*03:05, HLA-DRB1*04:08 (Table 4). Protein digesting enzyme server was used to estimate peptides digesting enzymes. Enzymes that do not digest peptide into fragments is considered non-digestive enzyme. Peptides digestible with many enzymes are not stable. Less enzyme digested peptides, on the other hand, are very stable and favored vaccine candidates (Table 5).

World population coverage

Population coverage was calculated with finally selected nine MHC class I and seven MHC class II epitopes and related HLA alleles. Selected MHC class I and MHC class II epitopes represented 88.23% and 66.66% of the world's population. Highest coverage of MHC class I epitopes found in the population of two countries Finland (96.16%) and Italy (96.38%). Similarly, the highest coverage of MHC class II epitopes was found within France (79.44%). In China where SARS-CoV-2 was first identified, MHC class I epitopes showed 63.26% population coverage and MHC class II showed 41.17% coverage. Also, the population coverage was higher in regions where SARS-CoV-2 cases have been reported (Figure 3).

Molecular docking analysis

The 3D structures of selected nine MHC class I epitopes were predicted using PEPFOLD (Additional file 1: Figure S9). Molecular docking is important to understand the patterns of protein-peptides interaction. To analyze the interaction of HLA allele with selected MHC class I epitopes, molecular docking was performed. Docking results revealed all the selected peptides binds inside the receptor binding site of HLA allele with strong interactions (Figure 4).

Discussion

CoVs have long been considered as insignificant pathogens causing "colds" in humans. In the 21st century, two extremely pathogenic CoVs named SARS-CoV and MERS-CoV emerged from the livestock reservoirs and cause deadly outbreaks. A new strain of CoV officially named as SARS-CoV-2 was identified in the Chinese city of Wuhan on December 31, 2019. The final dimension and impact of this outbreak are currently uncertain due to the rapidly changing situation [3]. After the recombination of various virus genomes particles, the novel virus infects the host cells rapidly. No reliable medication is currently available for the said infection. The emergence of COVID-19 results in a significant global disease burden, for which preventative measures are urgently needed [49]. Recent advancements in the immunological bioinformatics area has resulted in a variety of tools and servers that can lessen the time and cost of traditional vaccine advancement. Due to the problems in the selection of suitable antigen candidates and immunodominant epitopes, the development of effective multiple epitope vaccines remains toilsome. Thus, the prediction of appropriate antigenic epitopes of a targeted protein by the immunoinformatics approaches is very essential for designing a multiple epitope vaccine [50].

T-cell epitopes reside on the antigen-presenting cell (APC) surface, where they bind to major histocompatibility molecules (MHC) to induce the immune response [51]. Class I MHC molecules generally comprise the peptides of 8-11 amino acids while MHC class II binding peptides are typically 12-25 amino acids long. Adaptive immune response trigger by the cells is unique to the pathogens if appropriate epitopes are present. In specific types of cells, Class II MHCs are expressed like CPA's such as B cells, macrophages and dendritics, while class I MHCs are present in all nuclear cells of the body [52]. B-cell epitopes are pathogen surface antigen markers that interact with B cell receptors. Hydrophobic B cell receptor binding site has six hypervariable cycles of different length and amino acid composition [53]. B-cell epitopes are classified as linear/continuous and conformational/non-continuous. The experiments are mainly based on linear epitopes [54].

Here, we explored the development of epitope-based vaccines targeting the structural proteins of the SARS-CoV-2. T- and B-cell epitopes of the target proteins were predicted to support the host's immune response. Research was performed at primary, secondary and tertiary structural levels of proteins. IEDB analysis resource and ABCPred predicted B-cell conserved epitopes. Certain IEDB techniques were utilized to evaluate antigenicity, solvent accessibility, disulphide bonds, and flexibility. The 'SPTKLNDLCFTNVY' produced a higher immunogenicity score (1.6982) and could represent a potential B cell epitope and candidate for the vaccine. In addition, antigenic T cell determinants with potential to bind MHC class I and MHC class II have been predicted by the IEDB consensus approach. MHC-I (IPFAMQMAYRFN) and MHC-II (VTLCFVLAAYRIN) epitopes associate with many HLA alleles and are strongly antigenic. The position of epitopes on 3D structures of proteins was visualized by Pymol. DiscoTop server was used to predict discontinuous epitopes. To further improve specificity and selectivity, allergenicity, toxicity and physiochemical properties of predicted epitopes were checked. Digestion analysis verified that the peptides predicted during the analysis were stable and safe to use. For future studies the predicted epitopes should be tested for therapeutic potential.

The rapid development of structural and genomic databases combined with computational tools, helps in the design and discovery of new vaccine candidates. COVID-19 infection is a severe problem of morbidity and mortality worldwide. Unfortunately, the unavailability of the vaccinations against COVID-19 has impacted several precious lives, in different regions of the world. To successfully eradicate the disease, researchers have been trying to collect data associated with CoVs to understand its transmission, pathophysiology and biology. Our analysis will help to develop potential peptide vaccines to eradicate the COVID-19 and help to combat against SARS-CoV-2.

Conclusions

A reverse vaccinology technique for classifying surface-exposed antigens was used in this research, rather than focusing on the whole pathogen, a less successful approach. T and B cell epitopes were identified through sequence, structure and conservational analysis. Due to their prediction methods, yields, speeds, and low costs, B and T cell epitopes have become the focal point of immunoinformatics studies. A preliminary sequence of epitopes for future SARS-CoV-2 vaccine development could be seen in this study, which can help manage this increased health hazard. The predicted epitopes should be tested for therapeutic potency in future studies.

Abbreviations

SARS-CoV-2: Severe Acute Respiratory Syndrome Corona virus 2; COVID-19: Coronavirus disease 2019; S: Spike; M: Membrane; E: Envelope; PDB: Protein Data Bank; PI: Isoelectric Point; MHC: Major Histocompatibility Complex

Declarations

Ethics approval and consent to participate

Not Applicable

Consent for publication Availability of data and material

Not Applicable

Competing interests

All authors have no competing interests.

Funding

This work was supported by the Starting Research Grant for High-level Talents from Guangxi University and Postdoctoral research platform grant of Guangxi University.

Authors' contributions

MTQ, LLC and UAA conceived and designed this study; MTQ and FS performed the experiments; SA, IF, MMF and AZ analyse the results; MTQ and FS wrote the manuscript; LLC and UAA improved and revised the manuscript, and all the authors approved the final version.

Acknowledgements

Authors would like to acknowledge Guangxi University and Government College University Faisalabad for providing facilities for this study.

References

1. Tahir ul Qamar M, Alqahtani SM, Alamri MA, Chen L-L: **Structural Basis of SARS-CoV-2 3CL^{Pro} and Anti-COVID-19 Drug Discovery from Medicinal Plants.** *Preprints* 2020, **2020020193**.
2. Tahir ul Qamar M, Saleem S, Ashfaq UA, Bari A, Anwar F, Alqahtani S: **Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study.** *Journal of Translational Medicine* 2019, **17**:362.
3. de Wilde AH, Snijder EJ, Kikkert M, van Hemert MJ: **Host factors in coronavirus replication.** In *Roles of Host Gene and Non-coding RNA Expression in Virus Infection*. Springer; 2017: 1-42
4. Wang C, Horby PW, Hayden FG, Gao GF: **A novel coronavirus outbreak of global health concern.** *The Lancet* 2020.
5. Wu P, Hao X, Lau EH, Wong JY, Leung KS, Wu JT, Cowling BJ, Leung GM: **Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020.** *Eurosurveillance* 2020, **25**:2000044.
6. Gralinski LE, Menachery VD: **Return of the Coronavirus: 2019-nCoV.** *Viruses* 2020, **12**:135.
7. Cotten M, Watson SJ, Zumla AI, Makhdoom HQ, Palser AL, Ong SH, Al Rabeeah AA, Alhakeem RF, Assiri A, Al-Tawfiq JA: **Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus.** *MBio* 2014, **5**:e01062-01013.
8. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X: **Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.** *The Lancet* 2020.
9. Chen Y, Liu Q, Guo D: **Coronaviruses: genome structure, replication, and pathogenesis.** *Journal of Medical Virology*.
10. Peiris JS, Yuen KY, Osterhaus AD, Stöhr K: **The severe acute respiratory syndrome.** *New England Journal of Medicine* 2003, **349**:2431-2441.
11. Wu P, Hao X, Lau EH, Wong JY, Leung KS, Wu JT, Cowling BJ, Leung GM: **Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020.**
12. Nishiura H, Jung S-m, Linton NM, Kinoshita R, Yang Y, Hayashi K, Kobayashi T, Yuan B, Akhmetzhanov AR: **The Extent of Transmission of Novel Coronavirus in Wuhan, China, 2020.** Multidisciplinary Digital Publishing Institute; 2020.

13. Ji W, Wang W, Zhao X, Zai J, Li X: **Homologous recombination within the spike glycoprotein of the newly identified coronavirus may boost cross-species transmission from snake to human.** *Journal of Medical Virology* 2020.
14. Shammah A, Budoor H, Suad M, Richard K, Maha S, Asokan G: **Middle East Respiratory Syndrome Corona Virus (MERS-CoV): Levels of Knowledge and Awareness in Bahrain.** *KnE Life Sciences* 2018:98–114-198–114.
15. 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*, n/a.
16. Tahir ul Qamar M, Bari A, Adeel MM, Maryam A, Ashfaq UA, Du X, Muneer I, Ahmad HI, Wang J: **Peptide vaccine against chikungunya virus: immunoinformatics combined with molecular docking approach.** *Journal of translational medicine* 2018, **16**:298.
17. Ahmad B, Ashfaq UA, Rahman M-u, Masoud MS, Yousaf MZ: **Conserved B and T cell epitopes prediction of ebola virus glycoprotein for vaccine development: an immuno-informatics approach.** *Microbial pathogenesis* 2019, **132**:243-253.
18. Cockrell AS, Johnson JC, Moore IN, Liu DX, Bock KW, Douglas MG, Graham RL, Solomon J, Torzewski L, Bartos CJSr: **A spike-modified Middle East respiratory syndrome coronavirus (MERS-CoV) infectious clone elicits mild respiratory disease in infected rhesus macaques.** 2018, **8**:10727.
19. Amer H, Alqahtani AS, Alaklobi F, Altayeb J, Memish ZA: **Healthcare worker exposure to Middle East respiratory syndrome coronavirus (MERS-CoV): revision of screening strategies urgently needed.** *International Journal of Infectious Diseases* 2018, **71**:113-116.
20. Hui DS, Azhar EI, Kim Y-J, Memish ZA, Oh M-d, Zumla A: **Middle East respiratory syndrome coronavirus: risk factors and determinants of primary, household, and nosocomial transmission.** *The Lancet Infectious Diseases* 2018, **18**:e217-e227.
21. Ashfaq UA, Ahmed B: **De novo structural modeling and conserved epitopes prediction of Zika virus envelop protein for vaccine development.** *Viral immunology* 2016, **29**:436-443.
22. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW: **GenBank.** *Nucleic acids research* 2008, **37**:D26-D31.
23. 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**:3784-3788.
24. Ferré F, Clote P: **DIANNA 1.1: an extension of the DIANNA web server for ternary cysteine classification.** *Nucleic acids research* 2006, **34**:W182-W185.
25. Doytchinova IA, Flower DR: **VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines.** *BMC bioinformatics* 2007, **8**:4.
26. Dimitrov I, Flower DR, Doytchinova I: **AllerTOP-a server for in silico prediction of allergens.** In *BMC bioinformatics*. BioMed Central; 2013: S4.
27. 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* 1997, **25**:3389-3402.
28. Webb B, Sali A: **Protein structure modeling with MODELLER.** In *Protein structure prediction*. Springer; 2014: 1-15
29. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: **UCSF Chimera—a visualization system for exploratory research and analysis.** *Journal of computational chemistry* 2004, **25**:1605-1612.
30. Shin W-H, Lee GR, Heo L, Lee H, Seok C: **Prediction of protein structure and interaction by GALAXY protein modeling programs.** *Bio Design* 2014, **2**:1-11.
31. Xu D, Zhang Y: **Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization.** *Biophysical journal* 2011, **101**:2525-2534.
32. Wiederstein M, Sippl MJ: **ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins.** *Nucleic acids research* 2007, **35**:W407-W410.
33. Lovell SC, Davis IW, Arendall III WB, De Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC: **Structure validation by Ca geometry: ϕ , ψ and C β deviation.** *Proteins: Structure, Function, and Bioinformatics* 2003, **50**:437-450.
34. Saha S, Raghava GPS: **Prediction of continuous B-cell epitopes in an antigen using recurrent neural network.** *Proteins: Structure, Function, and Bioinformatics* 2006, **65**:40-48.
35. Fieser TM, Tainer JA, Geysen HM, Houghten RA, Lerner RA: **Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein alpha-helix.** *Proceedings of the National Academy of Sciences* 1987, **84**:8568-8572.
36. Jespersen MC, Peters B, Nielsen M, Marcatili P: **BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes.** *Nucleic acids research* 2017, **45**:W24-W29.
37. Sun P, Ju H, Liu Z, Ning Q, Zhang J, Zhao X, Huang Y, Ma Z, Li Y: **Bioinformatics resources and tools for conformational B-cell epitope prediction.** *Computational and mathematical methods in medicine* 2013, **2013**.
38. DeLano WL: **Pymol: An open-source molecular graphics tool.** *CCP4 Newsletter on protein crystallography* 2002, **40**:82-92.
39. Zhang M, Ishii K, Hisaeda H, Murata S, Chiba T, Tanaka K, Li Y, Obata C, Furue M, Himeno K: **Ubiquitin-fusion degradation pathway plays an indispensable role in naked DNA vaccination with a chimeric gene encoding a syngeneic cytotoxic T lymphocyte epitope of melanocyte and green fluorescent protein.** *Immunology* 2004, **112**:567-574.
40. 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 bioinformatics* 2007, **8**:361.
41. Adhikari UK, Tayebi M, Rahman MM: **Immunoinformatics approach for epitope-based peptide vaccine design and active site prediction against polyprotein of emerging oropouche virus.** *Journal of immunology research* 2018, **2018**.
42. 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 bioinformatics* 2006, **7**:153.
43. Dimitrov I, Naneva L, Doytchinova I, Bangov I: **AllergenFP: allergenicity prediction by descriptor fingerprints.** *Bioinformatics* 2014, **30**:846-851.

44. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP, Consortium OSDD: **In silico approach for predicting toxicity of peptides and proteins.** *PloS one* 2013, **8**.
45. Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P: **PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex.** *Nucleic acids research* 2016, **44**:W449-W454.
46. Tahir ul Qamar M, Maryam A, Muneer I, Xing F, Ashfaq UA, Khan FA, Anwar F, Geesi MH, Khalid RR, Rauf SA, Siddiqi AR: **Computational screening of medicinal plant phytochemicals to discover potent pan-serotype inhibitors against dengue virus.** *Scientific Reports* 2019, **9**:1433.
47. Durdagi S, Tahir ul Qamar M, Salmas RE, Tariq Q, Anwar F, Ashfaq UA: **Investigating the molecular mechanism of staphylococcal DNA gyrase inhibitors: A combined ligand-based and structure-based resources pipeline.** *Journal of Molecular Graphics and Modelling* 2018, **85**:122-129.
48. 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, **7**:1511-1522.
49. Douglas MG, Kocher JF, Scobey T, Baric RS, Cockrell AS: **Adaptive evolution influences the infectious dose of MERS-CoV necessary to achieve severe respiratory disease.** *Virology* 2018, **517**:98-107.
50. Nain Z, Karim MM, Sen MK, Adhikari UK: **Structural Basis and Designing of Peptide Vaccine using PE-PGRS Family Protein of Mycobacterium ulcerans—An Integrated Vaccinomics Approach.** *bioRxiv* 2019:795146.
51. Madden DR: **The three-dimensional structure of peptide-MHC complexes.** *Annual review of immunology* 1995, **13**:587-622.
52. Janeway CA, Capra JD, Travers P, Walport M: *Immunobiology: the immune system in health and disease.* 1999.
53. Greenbaum JA, Andersen PH, Blythe M, Bui HH, Cachau RE, Crowe J, Davies M, Kolaskar A, Lund O, Morrison S: **Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools.** *Journal of Molecular Recognition: An Interdisciplinary Journal* 2007, **20**:75-82.
54. Saha S, Bhasin M, Raghava GP: **Bcipep: a database of B-cell epitopes.** *BMC genomics* 2005, **6**:79.

Tables

Table 1. Physiochemical properties of the SARS-COV-2 structural proteins analyzed through ProtParam tool

Proteins	Molecular Weight	Theoretical pI	Instability index	Half-life	Stability Profiling	Aliphatic Index	Grand Average of Hydropathy	Amino Acid Composition
S	141178.47	6.24	33.01	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (<i>Escherichia coli</i> , in vivo).	Stable	84.67	-0.079	C-6336 H-9770 N-1656 O-1894 S-54
E	8365.04	8.57	38.68	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (<i>Escherichia coli</i> , in vivo).	Stable	144.00	1.128	C-390 H-625 N-91 O-103 S-4
M	25146.62	9.51	39.14	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (<i>Escherichia coli</i> , in vivo).	Stable	120.86	0.446	C-1165 H-1823 N-303 O-301 S-8

Table 2. Linear B cell epitopes predicted through ABCPred 2.0 server (NT: nontoxic)

Protein	B cell epitopes (Position)	Score	Antigenicity	Toxicity
S	FSTFKCYGVSPTKL (374)	0.90	0.8	NT
	ILPVSMTKTSVDCT (726)	0.89	1.6	NT
	AGCLIGAEHVNNSY (647)	0.87	0.8	NT
	LSSTASALGKLQDV (938)	0.85	0.8	NT
	DLPIGINITRFQTL (228)	0.84	1.1	NT
	LTGTGVLTESNKKF (546)	0.83	0.7	NT
	SIAYTMSLGAENS (691)	0.81	0.7	NT
	EILDITPCSFSGGVS (583)	0.81	1.6	NT
	DPQTLLEILDITPCS (578)	0.81	1.2	NT
	VNFNFNGLTGTGVL (539)	0.80	1.2	NT
	QPYRVVVLSFELLH (506)	0.78	0.9	NT
	GVVFLHVTYVPAQE (1059)	0.78	1.0	NT
	LPLVSSQCVNLTTR (8)	0.76	1.3	NT
	ISVTTEILPVSMTK (720)	0.75	1.2	NT
	AIPTNFTISVTTEI (713)	0.73	0.8	NT
	LQYGSFCTQLNRAL (754)	0.69	1.0	NT
	SPTKLNDLCFTNVY (383)	0.67	1.6	NT
	TPCSFSGGVSVITPG (588)	0.56	0.9	NT
	MSLGAENSVAYSNN (697)	0.51	0.8	NT
E	FLLVTLAILTALRL (26)	0.60	0.8	NT
M	PVTLACFVLAAYR (59)	0.75	0.9	NT
	GGIAIAMACLVGLM (78)	0.68	0.8	NT
	LEQWNLVIGFLFLT (17)	0.67	0.9	NT

Table 3. Discontinuous epitopes predicted through DiscoTop 2.0 server

Protein	Residues Position	Residues Names	Number of contacts	Propensity Score	DicsoTope score
S Protein	20	THR	0	-2.399	-2.123
	69	HIS	9	-2.720	-3.442
	70	VAL	16	-0.522	-2.302
	71	SER	0	-0.579	-0.513
	72	GLY	6	-0.453	-1.091
	73	THR	12	-1.188	-2.431
	75	GLY	9	-2.669	-3.397
	97	LYS	7	-2.338	-2.874
	98	SER	4	-2.876	-3.005
	145	TYR	16	-1.112	-2.824
	146	HIS	5	-0.164	-0.430
	147	LYS	8	0.591	-0.397
	148	ASN	15	0.699	-1.107
	149	ASN	2	3.237	2.634
	150	LYS	11	2.037	0.538
	151	SER	13	2.690	0.885
	152	TRP	16	1.959	-0.106
	153	MET	8	1.653	0.543
	154	GLU	25	1.182	-1.829
	155	SER	16	1.161	-0.813
	156	GLU	12	2.411	0.754
	157	PHE	9	-0.155	-1.172
	158	ARG	12	-0.632	-1.939
	159	VAL	18	-0.854	-2.826
	160	TYR	3	-1.197	-1.404
	162	SER	0	-3.205	-2.837
	163	ALA	11	-2.706	-3.660
	169	GLU	0	-3.602	-3.188
	171	VAL	5	-3.120	-3.336
	173	GLN	14	-1.647	-3.068
	174	PRO	13	-1.869	-3.149
	177	MET	0	-2.514	-2.225
	178	ASP	1	-1.776	-1.687
	179	LEU	14	-1.375	-2.826
	180	GLU	14	-0.767	-2.288
	181	GLY	18	0.278	-1.824
	182	LYS	9	2.192	0.905
	183	GLN	16	1.250	-0.734
	184	GLY	9	1.802	0.560
	185	ASN	16	3.439	1.203
	186	PHE	18	2.213	-0.112
	187	LYS	13	3.203	1.340
	188	ASN	20	2.040	-0.495
	189	LEU	9	1.785	0.545
	190	ARG	6	1.541	0.674
	191	GLU	5	0.670	0.018
	192	PHE	6	-0.452	-1.090
	193	VAL	18	0.177	-1.913
	194	PHE	5	1.153	0.445
	195	LYS	10	2.424	0.996
	196	ASN	3	1.869	1.309
	197	ILE	4	1.887	1.210
	198	ASP	8	1.944	0.801
	199	GLY	11	2.928	1.327
	200	TYR	5	0.998	0.308
	201	PHE	12	1.620	0.054
	202	LYS	8	-0.112	-1.019
	203	ILE	16	-1.238	-2.935
	205	SER	2	-2.369	-2.327
	320	VAL	10	-0.331	-1.443
	321	GLN	14	-1.119	-2.039

322	PRO	23	1.247	-1.541
323	THR	12	-0.745	-2.039
324	GLU	9	-0.992	-1.913
325	SER	8	-0.364	-1.242
326	ILE	11	-1.503	-2.595
327	VAL	5	-3.080	-3.301
414	GLN	13	-0.959	-2.344
415	THR	10	-0.831	-1.885
416	GLY	11	-2.662	-3.621
437	ASN	2	-3.160	-3.026
438	SER	14	-2.155	-3.517
439	ASN	0	-1.470	-1.301
440	ASN	0	-0.203	-0.180
441	LEU	13	-0.388	-1.839
442	ASP	12	0.102	-1.290
443	SER	1	-0.194	-0.287
444	LYS	6	-0.582	-1.205
445	VAL	17	0.131	-1.839
446	GLY	5	1.035	0.341
447	GLY	12	2.009	0.397
448	ASN	8	1.181	0.125
449	TYR	21	-0.9602	-3.266
450	ASN	9	-1.220	-2.115
451	TYR	17	-1.650	-3.416
460	ASN	2	-3.837	-3.626
466	ARG	16	-2.081	3.682
467	ASP	9	-2.559	-3.299
468	ILE	17	0.988	-2.830
469	SER	4	-1.198	-1.520
470	THR	21	-0.363	-2.736
471	GLU	4	1.297	1.245
472	ILE	25	1.276	-1.746
473	TYR	14	1.332	-0.431
478	THR	3	-2.903	-2.914
480	CYS	2	-3.542	-3.365
527	PRO	1	-3.443	-3.162
529	LYS	3	-3.381	-3.337
532	ASN	0	-2.618	-2.317
558	LYS	2	-1.413	-1.491
560	LEU	13	-1.977	-3.244
561	PRO	0	-0.721	-0.638
562	PHE	4	-1.484	-1.773
564	GLN	21	-3.118	-3.679
678	THR	4	0.881	0.319
679	ASN	10	0.189	0.983
680	SER	14	-1.574	-3.003
681	PRO	1	0.047	-0.073
682	ARG	9	1.473	0.268
683	ARG	0	2.197	1.945
684	ALA	3	1.850	1.292
685	ARG	0	2.284	2.021
686	SER	7	1.867	0.848
687	VAL	6	1.784	0.889
688	ALA	18	-0.621	-2.620
793	PRO	2	-1.171	-1.266
795	LYA	6	-2.174	-2.614
796	ASP	1	-3.607	-3.307
807	PRO	4	0.332	-0.166
808	ASP	13	-1.341	-2.682
809	PRO	3	0.610	0.195
810	SER	13	-0.402	-1.851
811	LYS	3	-1.662	-1.816
812	PRO	4	1.515	0.881

	813	SER	3	2.083	1.499
	814	LYS	15	0.158	-1.585
	815	ARG	20	-0.253	-2.524
	936	ASP	2	-3.731	-3.532
	1104	VAL	21	-0.766	-3.093
	1105	THR	6	1.205	0.376
	1106	GLN	15	3.212	1.118
	1107	ARG	16	3.213	1.118
	1108	ASN	19	2.938	0.415
	1109	PHE	20	-0.099	-2.388
	1110	TYR	13	0.162	-1.352
	1191	LYS	8	-3.002	-3.576
	1192	ASN	1	-3.503	-3.215
	1195	GLU	0	-4.126	-3.651
E Protein	63	LYS	4	-3.653	-3.693
M Protein	1	MET	9	-0.899	-1.831
	2	ALA	6	-2.467	-2.873
	3	ASP	10	-2.746	-3.580
	4	SER	0	-2.407	-2.131
	202	GLY	2	-3.269	-3.123
	203	ASN	2	-2.220	-2.195
	204	TYR	17	-1.043	-2.878
	205	LYS	5	0.198	-0.400
	206	LEU	4	-0.429	-0.840
	207	ASN	9	0.348	-0.727
	208	THR	6	1.343	0.498
	209	ASP	0	1.026	0.908
	210	HIS	3	0.399	0.008
	211	SER	3	0.359	-0.027
	212	SER	6	-0.919	-1.503
	213	SER	2	-0.541	-0.709
	214	SER	3	-0.862	-1.108
	215	ASP	0	-1.638	-1.449
	216	ASN	3	-2.148	-2.246
	217	ILE	3	-2.798	-2.821
	218	ALA	3	-3.375	-3.332
	222	GLN	3	-3.606	-3.537

Table 4. MHC class-I allele and MHC class-II binding peptides with their antigenicity scores

Protein	Peptides (position)	Alleles	Antigenicity
MHC Class I			
S	VRFPNITNLCPF (327-338)	HLA-B*35:03, HLA-B*53:01, HLA-A*24:02, HLA-C*07:01,, HLA-B*35:01, HLA-C*06:02, HLA-A*23:01, HLA-B*51:01, HLA-C*14:02,	1.2
	ALQIPFAMQMAY (893-904)	HLA-B*35:01, HLA-B*15:01, HLA-A*29:02, HLA-A*03:01, HLA-A*30:02, HLA-B*18:01, HLA-A*25:01	0.9
	IPFAMQMAYRFN (896-907)	HLA-B*35:01, HLA-B*35:03, HLA-A*24:02, HLA-B*51:01, HLA-B*53:01	1.3
E	PSFYVYSRVKNL (54-65)	HLA-C*14:02, HLA-C*07:01, HLA-C*06:02	0.8
	SFYVYSRVKNLN (55-66)	HLA-C*14:02, HLA-C*07:01, HLA-C*06:02	0.9
	FYVYSRVKNLNS (56-67)	HLA-C*14:02, HLA-C*07:01, HLA-C*06:02	0.6
M	YRINWITGGIAI (71-82)	HLA-B*27:05, HLA-A*32:01, HLA-C*06:02, HLA-B*39:01, HLA-C*07:01	1.2
	SFRLFARTRSMW (99-110)	HLA-C*06:02, HLA-C*07:01, HLA-B*14:02, HLA-A*32:01, HLA-B*57:01	0.6
	ITVATSRTLSYY (168-179)	HLA-A*01:01, HLA-A*30:02, HLA-A*26:01, HLA-A*29:02, HLA-B*57:01	0.7
MHC Class II			
S	FVFLVLLPLVSSQCV (2-16)	HLA-DRB1*13:21, HLA-DRB1*01:01, HLA-DRB1*15:02, HLA-DRB1*11:28, HLA-DRB1*13:05, HLA-DPA1*03:01/DPB1*04:02, HLA-DRB1*13:07, HLA-DRB1*11:01, HLA-DRB1*11:02, HLA-DRB1*11:21, HLA-DRB1*13:22, HLA-DRB1*11:04, HLA-DRB1*11:06, HLA-DRB1*13:11, HLA-DRB1*08:17, HLA-DRB1*13:01, HLA-DRB1*13:27, HLA-DRB1*13:28, HLA-DRB1*11:14, HLA-DRB1*13:23, HLA-DRB1*07:03, HLA-DRB1*04:08	0.7
E	LLFLAFVVFLVTLA (18-32)	HLA-DPA1*03:01/DPB1*04:02, HLA-DPA1*01:03/DPB1*02:01, HLA-DPA1*01/DPB1*04:01, HLA-DPA1*02:01/DPB1*01:01, HLA-DRB1*15:02, HLA-DRB1*04:23, HLA-DRB1*04:04, HLA-DRB1*04:08, HLA-DRB1*04:10, HLA-DQA1*05:01/DQB1*02:01, HLA-DRB1*08:13, HLA-DRB1*07:03, HLA-DRB1*01:02, HLA-DRB1*04:05	0.8
	AFVVFLVTLAILTA (22-36)	HLA-DPA1*03:01/DPB1*04:02, HLA-DRB1*15:02, HLA-DRB1*04:23, HLA-DRB1*11:04, HLA-DRB1*11:06, HLA-DRB1*13:11, HLA-DRB1*04:08, HLA-DRB1*04:10, HLA-DRB1*11:28, HLA-DRB1*13:05, HLA-DPA1*01/DPB1*04:01, HLA-DRB1*04:21, HLA-DRB1*08:13, HLA-DRB1*04:01, HLA-DRB1*04:26, HLA-DRB1*07:03, HLA-DRB1*01:01, HLA-DPA1*02:01/DPB1*01:01, HLA-DRB1*04:05, HLA-DRB1*13:07,	0.6
	FVVFLVTLAILTAL	HLA-DPA1*03:01/DPB1*04:02, HLA-DRB1*15:02, HLA-DRB1*04:23, HLA-DRB1*11:04, HLA-DRB1*11:06, DRB1*13:11, HLA-DRB1*04:08, HLA-DRB1*04:10, HLA-DRB1*11:28, HLA-DRB1*13:05, HLA-DPA1*01/DPB1*04:01, HLA-DRB1*01:01, HLA-DRB1*04:01, HLA-DRB1*04:26, HLA-DRB1*07:03, HLA-DPA1*02:01/DPB1*01:01, HLA-DRB1*01:02, HLA-DRB1*04:05, HLA-DRB1*13:07	0.5
M	ASFRLFARTRSMWSF (98-112)	HLA-DRB1*08:13, HLA-DRB1*11:14, HLA-DRB1*13:23, HLA-DRB1*15:02, HLA-DRB1*11:20, HLA-DRB1*11:01 HLA-DRB1*13:07, HLA-DRB1*15:06 HLA-DRB1*11:28, HLA-DRB1*13:05 HLA-DRB1*04:01, HLA-DRB1*04:26 HLA-DRB1*07:01, HLA-DRB1*11:02 HLA-DRB1*11:21, HLA-DRB1*13:22 HLA-DRB1*03:05,	0.7
	FRLFARTRSMWSFNP (100-114)	HLA-DRB1*08:13, HLA-DRB1*11:14, HLA-DRB1*13:23, HLA-DRB1*15:02, HLA-DRB1*11:20, HLA-DRB1*13:07, HLA-DRB1*11:01, HLA-DRB1*15:06, HLA-DRB1*11:28, HLA-DRB1*04:01, HLA-DRB1*04:26, HLA-DRB1*11:02, HLA-DRB1*11:21, HLA-DRB1*13:22, HLA-DRB1*03:05	0.8
	VTLACFVLAAYRIN (60-74)	HLA-DRB1*07:03, HLA-DRB1*11:20, HLA-DRB1*01:02, HLA-DRB1*07:01, HLA-DRB1*11:14, HLA-DRB1*13:23, HLA-DRB1*03:09, HLA-DRB1*13:07, HLA-DRB1*11:28, HLA-DRB1*13:05, HLA-DRB1*03:05, HLA-DRB1*04:08	1.0

Table 5. Digestion, Allergenicity, Toxicity and Physiochemical profiling of Selected Peptides (NA: not allergic; NT: nontoxic)

Peptides	Non-Digesting Enzyme	Allergenicity	Toxicity	Hydrophilicity	Hydrophobicity	Charge	pI	M
MHC Class I								
VRFPNITNLCPF	Chymotrypsin,Cyanogen_Bromide, IodosoBenzoate,Staph_Protease, Trypsin_K, AspN,	NA	NT	-0.67	-0.02	1.00	8.60	14
ALQIPFAMQMAY	Trypsin, Clostripain, IodosoBenzoate, Staph_Protease, Trypsin_K, Trypsin_R, AspN,	NA	NT	-1.01	0.14	0.00	5.88	13
IPFAMQMAYRFN	IodosoBenzoate, Staph_Protease, Trypsin_K, AspN	NA	NT	-0.78	-0.01	1.00	9.10	14
PSFYVYSRVKNL	Cyanogen_Bromide, IodosoBenzoate, Staph_Protease, AspN	NA	NT	-0.42	-0.15	2.00	9.72	14
SFYVYSRVKNLN	AspN, Staph_Protease, Proline_Endopept, IodosoBenzoate, Cyanogen_Bromide	NA	NT	-0.41	-0.20	2.00	9.72	14
FYVYSRVKNLNS	Cyanogen_Bromide, IodosoBenzoate, Proline_Endopept, Staph_Protease, AspN	NA	NT	-0.41	-0.20	2.00	9.70	14
YRINWITGGIAI	Cyanogen_Bromide, Proline_Endopept, Staph_Protease, Trypsin_K, AspN	NA	NT	-0.88	0.11	1.00	9.10	13
SFRLFARTRSMW	IodosoBenzoate, Proline_Endopept, Staph_Protease, Trypsin_K, AspN	NA	NT	-0.23	-0.28	3.00	12.31	15
ITVATSRTLSTYY	Cyanogen_Bromide, IodosoBenzoate, Proline_Endopept, Staph_Protease, Trypsin_K, AspN	NA	NT	-0.65	-0.06	1.00	8.93	13
MHC Class II								
FVFLVLLPLVSSQCV	Trypsin, Clostripain, Cyanogen_Bromide, IodosoBenzoate, Staph_Protease, Trypsin_K, Trypsin_R, AspN	NA	NT	-1.23	0.28	0.00	5.85	16
LLFLAFVVFLVTLA	Trypsin, Clostripain, Cyanogen_Bromide, IodosoBenzoate, Proline_Endopept, Staph_Protease, Trypsin_K, Trypsin_R, AspN	NA	NT	-1.61	0.46	0.00	5.88	16
AFVVFLVTLAILTA	AspN, Trypsin_R, Trypsin_K, Staph_Protease, Proline_Endopept, IodosoBenzoate, Cyanogen_Bromide, Clostripain, Trypsin	NA	NT	-1.39	0.41	0.00	5.88	15
FVVFLVTLAILTAL	Trypsin, Clostripain, Cyanogen_Bromide, IodosoBenzoate, Proline_Endopept, Staph_Protease, Trypsin_K, Trypsin_R, AspN	NA	NT	-1.47	0.42	0.00	5.88	16
ASFRLFARTRSMWSF	Proline_Endopept, Staph_Protease, Trypsin_K, AspN	NA	NT	-0.18	-0.37	3.00	12.31	18
FRLFARTRSMWSFNP	Proline_Endopept, Staph_Protease, Trypsin_K, AspN	NA	NT	-0.34	-0.23	3.00	12.31	19
VTLACFVLAAYRIN	AspN, Trypsin_K, Staph_Protease, Proline_Endopept, IodosoBenzoate, Cyanogen_Bromide	NA	NT	-0.96	0.15	1.00	8.57	16

Figures

Figure 1

Specific sites of B cells predicted linear epitopes on the 3D structure of SARS-COV-2 proteins: (A) S protein, (B) E protein and (C) M protein.

Figure 2

Specific sites of B cells discontinues epitopes predicted through DISCOTOPE 2.0 server on the 3D structure of SARS-COV-2 proteins: (A) S protein, (B) E protein and (C) M protein.

Figure 3

Population Coverage of MHC class I and MHC class II peptides around the globe predicted by IEDB population Coverage Tool.

Figure 4

The 3D binding pattern of the selected 9 peptides (magenta) docked with HLA-B7 allele (lime green).

Supplementary Files

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

- [AdditionalFile6.docx](#)
- [Additionalfile1.docx](#)
- [AdditionalFile4.docx](#)
- [AdditionalFile5.docx](#)
- [Additionalfile3.docx](#)
- [Additionalfile2.docx](#)