

The multi-epitope vaccine prediction to combat Pandemic SARS-CoV-2, an immunoinformatic approach

K. M. Kaderi Kibria (✉ km_kibria@yahoo.com)

Department of Biotechnology and Genetic Engineering, Faculty of Life Science, Mawlana Bhashani Science and Technology University

Md. Shaid bin Islam

Department of Biotechnology and Genetic Engineering, Faculty of Life Science, Mawlana Bhashani Science and Technology University

Hedayet Ullah

Department of Biotechnology and Genetic Engineering, Faculty of Life Science, Mawlana Bhashani Science and Technology University

Mojnu Miah

Infectious Diseases Division, International Centre for Diarrheal Diseases Research, Bangladesh

Research Article

Keywords: SARS-CoV-2, COVID-19, Vaccine, Epitope, Immunoinformatics, Envelope protein, Membrane glycoprotein, Spike surface glycoprotein

Posted Date: April 7th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-21853/v1>

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

Abstract

Novel coronavirus (SARS-CoV-2) leads to coronavirus disease 19 (COVID-19) recently declared as a pandemic for its outbreak within almost 190 countries worldwide. No effective drugs and/or vaccines authenticated against this rapidly spreading virus till now. This study aims to establish an efficient multi-epitope vaccine that could elicit both T-cell and B-cell responses sufficient to recognize confirmed surface proteins of the virus. The sequences of the viral surface proteins, e.g. envelope protein (E), membrane glycoprotein (M), and S1 and S2 domain of spike surface glycoprotein (S) collected from the NCBI database. We adopted an immunoinformatic strategy to identify the immunogenic region of the proteins and assessed their affinity with MHC class-I and MHC class-II by various bioinformatics tools. Top epitopes have been selected and assessed for population coverage and conservancy among 180 SARS-CoV-2 genomes. Along with the above analyses, and results of Antigenicity, Allergenicity, and transmembrane location prediction, we selected top epitopes from these four proteins. The epitopes were assembled by the AAY linker to form a multi-epitope vaccine is 70 aa long, can be synthesized commercially. This should be processed by Antigen-presenting cells; consequently, the surface proteins might be recognized by the helper and cytotoxic T-cells as well as by B-cells. We also assessed the structural and various physicochemical properties of the novel chimeric peptide for its suitability as a multi-epitope vaccine. This in-silico study leads to a rationally designed potential vaccine candidate that could be assessed by wet-lab experiments driving towards efficient combat of the novel coronavirus outbreak.

Introduction

An outbreak of 2019-novel coronavirus (SARS-CoV-2) that leads to coronavirus disease 19 (COVID-19; atypical pneumonia) has emerged in China and has spread rapidly from the first known cases in Wuhan in late 2019 to 186 countries and territories around the world (March 21, 2020). SARS-CoV-2 has been declared a pandemic outbreak by the World Health Organization, with more than 303,051 confirmed cases and 12,955 deaths worldwide up to date (March 21, 2020). The epidemic spreads rapidly by human-to-human transmission with incubation period: 2- 14 days (possible outliers: 0-27 days; median: 3.0 days) 1, 2 and from the onset of symptoms the median time for developing pneumonia is 4.0 days (interquartile range, 2.0 to 7.0) (Guan, W. et al 2020). This is worrisome as nCoV-19 can be transmitted from person-to-person even if the infected person is not showing any symptoms. The basic reproduction number (R_0) for the human-to-human (direct) transmission ranged from 1.4 to 2.5 3

There is currently no recommended therapeutic or vaccine exists for the treatment of COVID-19 except for supportive care as needed. Currently, control measures have been depending on the rapid detection and isolation of symptomatic cases as of lack of approved drugs or vaccines 4. During the SARS-CoV outbreak in 2003, several agents used, but the efficacy of these drugs has not been established. In this context, a safe and efficient vaccine urgently required. In response to this global health crisis, researchers are working on developing a potential vaccine and drugs to fight the new coronavirus, called SARS-CoV-2. Researchers are working with different approaches for developing a coronavirus vaccine including, a) inactivated and attenuated whole virus vaccine, b) antibody vaccine, c) nucleic acid vaccine, and d) protein subunit vaccine. Possible phenotypic or genotypic reversion of whole viral vaccine, short-term protection from antibody vaccine makes it difficult to be used as a vaccine against newly emerged SARS-CoV-2. Alternatively, subunit vaccines based on the surface spike and other structural proteins of SARS- CoV-2 could be considered as they do not contain any live pathogens.

SARS-CoV-2, as like other Betacoronavirus SARS-CoV and MERS-CoV, is an enveloped, positive-sense, single-stranded RNA virus with a genome comprising ~30 kilobases 5, 6. SARS- CoV-2 genome encodes four structural proteins, the spike-surface glycoprotein (S), the envelope protein (E), the membrane glycoprotein (M), and the nucleocapsid protein (N), as well as several nonstructural proteins. The nucleocapsid protein holds the RNA genome, and the S, E, and M proteins together create the viral envelope. All these proteins may serve as antigens to stimulate neutralizing antibodies and important targets for cell-mediated immunity (both CD4+/CD8+ T-cell responses) 7, 8.

Preliminary studies suggest that based on the whole genome analysis, the novel SARS-CoV-2 is relatively similar to SARS-CoV 9, 10 and they used similar human cell receptors and cell entry mechanisms 9, 11, 12. Previous knowledge on the understanding of protective immune responses against SARS-CoV unriddles a new way to develop potential vaccine candidates for SARS-CoV- 2 as they are similar. Both humoral and cell-mediated immune responses have a protective role against SARS-CoV. The most exposed proteins of SARS-CoV are surface glycoprotein S, a major antigen that stimulates neutralizing antibody (humoral response) 13 as well as protects from infection in mouse models 14, 15. The spike proteins are also important targets of cytotoxic lymphocytes. The S protein of SARS-CoV-2 comprises two subunits; S1 (685 aa) and S2 (588 aa) 6. S1 subunit mainly contains a receptor-binding domain (RBD), which

mediates virus entry into the epithelial cells through interaction with the cell surface receptor ACE2, angiotensin-converting enzyme 2 16. Administration of RBD of S1 protein can play key parts in the induction of neutralizing antibodies and T-cell responses, as well as long-lasting protective immunity in animal models 17. Among SARS-CoV-2 viruses, the S2 subunit is well conserved and shares 99% identity with the bat SARS-like-coronaviruses 6. The vaccine design based on the conserved domain of the S2 subunit of the surface glycoprotein may provide broad-spectrum protection and is worth testing in animal models. Consequently, the SARS-CoV-2 vaccine can be developed based on the conserved domain, and in particular, the RBD region, following the strategy for the SARS-CoV vaccine development 18 that can be tested further in suitable models before clinical trials 6. The M protein of SARS-CoV induces dominant cellular immunogenicity, as well as strong humoral response, and may serve as potential targets for SARS-CoV-2 vaccine design 19. As the antibodies against the N protein do not provide immunity to the infection this protein is excluded for not being a suitable candidate for vaccine development 20.

In this study, we adopted an immunoinformatics approach to identify the potential T-cell and B-cell epitope(s) to design a multi-epitope vaccine based on the spike, envelope, and membrane proteins. The multi-epitope vaccine could be used as a promising candidate to prevent SARS-CoV-2 virus infection and severe outbreaks.

Methods

The methodology of the entire study to identify and manifest the top epitopes for SARS-CoV-2 has been described in a flow diagram (Fig 1).

Sequence retrieval and T-cell epitope prediction

The envelope protein, membrane glycoprotein, Surface glycoprotein of the SARS-CoV-2 virus retrieved from NCBI (<http://www.ncbi.nlm.nih.gov/protein>) protein database in FASTA format. The Surface glycoprotein was subjected to a CDD domain search where the S1 and S2 domain sequences extracted (Fig S1). All four proteins (E, M, S1, and S2) were analyzed for T-cell epitope prediction by adopting the previously used approaches described elsewhere 21–23. Briefly, NetCTL v1.2 server 24 utilized to predict the most antigenic regions of the proteins by selecting the 9-mer T cell epitopes based on the interaction with commonly occurring HLA Class I supertypes in the human population, i.e. A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, B59, and B62. This tool is well versed to estimate a combined score depending on the algorithms for MHC class-I binding, the transporter of antigenic peptide (TAP) transport efficiency, and proteasomal C-terminal cleavage prediction expressing antigenicity. The default threshold parameters were set in NetCTL.1.2 server, for example, proteasomal C-terminal cleavage, TAP transport efficiency, and epitope identification were set as 0.15, 0.05, and 0.75, respectively.

The affinity of the epitopes with different alleles of MHC class I and MHC class II estimated by tools available in Immune Epitope Database and Analysis Resource (IEDB-AR). The half-maximal inhibitory concentration (IC₅₀) of the selected 9-mer epitope from NetCTL analyses measured by stabilized matrix method (SMM) for the affinity with MHC class I. The affinity of the epitope with the specific HLA-DP, HLA-DQ, and HLA-DR loci of MHC class II was measured by NetMHCpan 2.0. For MHC class II binding analysis, Fifteen-mer epitopes were projected based on the preselected 9-mer epitope as a core peptide. The HLA alleles that have an affinity with epitopes by IC₅₀ < 250nM for the MHC class I and IC₅₀ < 200 nM for MHC class

II alleles, respectively, were selected. The MHC class I binding was crosschecked by the software EPISOPT (<http://bio.med.ucm.es/episopt.html>). We also utilized the MHC class II binding prediction tool, PREDIVAC, to evaluate their affinity with HLA-DRB1 alleles including 01:01, 03:01, 04:01, 07:01, 08:01, 10:01, 11:01, 12:01, 13:02, 14:01, and 15:01. These HLA Class II alleles assessed expected to cover more than 95% of the worldwide population 25. The antigenicity of the peptides was evaluated by Vaxijen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>). The threshold was kept 0.4 for antigenicity prediction.

Population Coverage Analysis

Population coverage evaluated by the IEDB population coverage calculation tool for the most antigenic epitopes 26. We selected area_country_ethnicity for the query, and the combined score for MHC class I and II used to determine the population coverage of the whole world population as well as different regions of the world.

Homology Modelling and Molecular Docking Analysis.

A homology model of envelope, membrane and spike protein built by MODELLER v9 27. The simulated model evaluated by the PROCHECK server 28. The disordered region in the protein sequences measured by DISOPRED v329. The prediction of the transmembrane region performed by TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>). We employed CABSDOCK WEB SERVER [<http://biocomp.chem.uw.edu.pl/CABSdock>] to perform molecular docking studies using the best possible epitope obtained from the above analyses and the PDB structures of HLA alleles retrieved from RCSB Protein Data Bank server (<https://www.rcsb.org/>). The MHC class I molecules with their 9-mer epitope pair and the MHC class II molecules with their 15-mer epitope pair submitted to CABS-dock with 50 simulation cycles. We docked the peptides for envelope protein, CVEnvA1 with the crystal structure of HLA-A*02:03 (3ox8), and CVEnvA2 with the crystal structure HLA-DRB1*01:01 (2FSE) by CABSDOCK WEB SERVER. Similarly, the peptides for membrane glycoprotein CVMemB1 and CVMemB2 were docked with the alleles HLA-A*68:01 (6PBH) and HLA-DRB1*01:01 (2FSE), respectively. The peptides for S1 protein, CVS1A1 and CVS1A2 were docked with alleles HLA-C*07:02 (5VGE) and DQA1/DQB1 (1JK8), respectively. The peptides for S2 protein CVS2B1 and CVS2B2 were docked with the alleles HLA-A*11:01 (6JP3) and HLA-DRB1*01:01 (2FSE), respectively. CABS-dock provides a docking simulation of the binding site permits full flexibility of the peptide and sometimes enables minor fluctuations of the receptor backbone.

Allergenicity investigation and B-Cell Epitope Prediction.

The allergenicity of the proposed epitopes assessed by the AllerTop server (<https://www.ddg-pharmfac.net/AllerTOP/>) where the support vector machine (SVM) algorithm based analyses were utilized 30. The selected T-cell epitope (15-mer) checked for the suitability as the B-cell epitope by IEDB-AR using several sequence-based tools like Emini surface accessibility prediction, Karplus & Schulz Flexibility Prediction, Kolaskar & Tongaonkar Antigenicity, and Parker Hydrophilicity Prediction tools 31–34.

Conservancy analyses

A total of 180 whole-genome sequences of SARS-CoV-2 isolated from the human, environment, and canine were retrieved from the GISAID database (<https://www.gisaid.org>). We excluded sequences that were too short/ or likely had sequencing errors. Retrieved sequences were then translated into amino acid residues by the ExPasy translate tool and retrieved three structural proteins according to the coding sequence positions provided along the reference sequence for SARS-CoV-2 proteins (S, M, and E). Translated proteins aligned with the NCBI reference protein sequence (S: YP_009724390.1; M: YP_009724393.1; E: YP_009724392.1). We have also retrieved similar sequences of SARS and MERS viruses by the NCBI server. All the protein sequences were aligned separately for each protein by multiple sequence alignment program ClustalW using BioEdit tool (version 7.1.3.0) with a number of bootstrap values 1000 (Hall, T. et al. 2011). The conservancy of all the proposed epitopes assessed by the IEDB conservancy analysis tool.

Designing of multi-epitope vaccine and analyses of its properties

We selected the best epitope from the envelope, membrane as well as S1 and S2 domain of spike protein based on the analyses of their affinity with MHC class I and II alleles, antigenicity, population coverage, conservancy, and allergenicity. The epitopes linked by the AAY linker and a Cysteine residue set at the N-terminal of the epitope. The cysteine residue will be required to conjugate the peptide with the carrier protein via a disulfide linkage. The antigenicity of the novel vaccine construct assessed by the VaxiJen v2.0 tool. The threshold for antigenicity prediction used 0.4. The allergenicity of the vaccine assessed by using the AllerTop server. The ProtParam server utilized to evaluate various physiochemical parameters of the multi-epitope vaccine, e.g. molecular weight, theoretical isoelectric point (pI), *in vitro* and *in vivo* half-life, instability and aliphatic index, and grand average of hydropathicity (GRAVY).

Modeling, refinement and validation of Vaccine-structure

The secondary structural properties of the multi-epitope vaccine construct assessed by the SOPMA server. The 3D model of the vaccine constructed by using the homology modeling tool- iTASSER 35. The model refined by using the Galaxy Refine server (<http://galaxy.seoklab.org/>). This server performs the repacking and molecular dynamics simulation to relax the structure, a CASP10 based refinement technique. The Galaxy Refine server is regarded as one of the best performing algorithms to improve the local

structural quality, as per the CASP10 evaluation. The tertiary structures of the vaccines were validated using ProSA-web 36. The server evaluated the overall quality of the model in the form of a z-score. If the z-scores of the predicted model are outside the range of the characteristic for native proteins, it indicates the erroneous structure. Further, the Ramachandran plot analysis of the predicted model performed by using the PROCHECK server (<https://servicesn.mbi.ucla.edu/PROCHECK/>) to determine its overall quality.

Linear and Conformational epitopes prediction.

The prediction of linear and conformational B cell epitopes in the multi-epitope vaccine construct was performed by using the ABCPred server (https://webs.iitd.edu.in/raghava/abcpred/ABC_submission.html) and ElliPro37, respectively, using default parameters.

Results

Sequence retrieval and identification of T-cell epitope

The protein sequences of the Envelope protein (YP_009724392), Membrane glycoprotein (YP_009724393), and surface glycoprotein (YP_009724390) of SARS-corona virus-2 (SARS CoV-2) retrieved from the NCBI server. The sequences assessed by the Vaxijen and found that all these three proteins are antigenic in nature with a score of 0.6025 for Envelope protein, 0.5102 for Membrane glycoprotein, and 0.4646 for Surface glycoprotein. The T-cell epitopes for these three proteins identified by the NetCTLv1.2 server, where the epitope prediction was restricted to 12 MHC class I supertypes. The top 10 epitopes for Envelope protein and top 12 for Membrane glycoprotein and Surface glycoprotein were (Table 1) selected based on the highest combined score, listed for further analysis.

Both the MHC class I and MHC class II-restricted alleles predicted by IEDB analysis resource based on the IC50 value. All the predicted epitopes in Table 1 evaluated for the analyses of MHC interaction. The MHC class I alleles interacted with epitopes of E-protein, M-protein, S1, and S2 protein summarized in Table S1, S3, S5, and S7, respectively. The number of MHC class I alleles interacted with the predicted epitopes for all these four proteins summarized in Table 2. The MHC class II alleles interacted with epitopes of E-protein, M-protein, S1, and S2 protein summarized in Table S2, S4, S6, and S8, respectively. The MHC class II epitopes (15-mer) selected depending on the 9-mer epitope as a core. The number of MHC class II alleles interacted with the predicted epitopes for all these four proteins summarized in Table 3. After MHC class I and MHC Class II analyses, we selected top interacting peptides and denoted each by a name (Table 4), for example, two peptides for Envelope protein (CEnvA2, KPSF YVYSRVKLNLS, and CEnvB2, NIVN VSLVKPSFYVY), Two peptides for Membrane glycoprotein (CMemA2, VGLM WLSYFIASFRL, and CMemB2, VIGAVILRGHLRIAG),

three peptides for S1 protein (CVS1A2, FNA TRFASVYAWNRK; CVS1B2, ADSFVIRGDEVRQIA, and CVS1C2, ISNCVADYSVLYNSA) and two peptides for S2 protein (CVS2A2, IWLGFIAGLIIMMV, and CVS2B2, FLHVITYVPAQKNFT).

The MHC class I interaction has been crosschecked by EPISOPT software, the result shown in Table S9. The result showed that the peptide, VSLVKPSFY is not a suitable MHC class-I epitope. The interaction with MHC class II has been validated by a software PREDIVAC, which predicts on the basis of the specificity-determining residue (SDR) concept. We assessed the epitopes for the interaction with the HLA-DRB1 alleles including 01:01, 03:01, 04:01, 07:01, 08:01, 10:01, 11:01, 12:01, 13:02, 14:01, and 15:01 that are expected to cover more than 95% of the worldwide population (Table S10) 25. The peptides were checked for antigenicity by the Vaxijen software, and it was found that all the peptides are potential antigens except CVS1B2 and CVS1C2 (Table S10).

Population Coverage and Conservancy Analysis.

The prediction of both MHC class I- and MHC class II-based coverage of the selected epitopes performed by IEDB analysis resources for the world population as well as different regions of the world. The world population coverage of CEnvA2 and CEnvB2 found to be 96.77%, and 71.88%, respectively, that enlisted in Table S11. The world population coverage of CMemA2 and CMemB2 found to be 99.82% and 82.11%, respectively (Table S12). The world population coverage of CVS1A2, CVS1B2, and CVS1C2 found to be 94.07%, 79.01%, and 70.77%, respectively, that enlisted in (Table S13). The world population coverage of CVS2A2 and CVS2B2 found to be 87.5% and 57.36%, respectively (Table S14). All these peptides were 96.12–100% conserved among the Sars-CoV-2 isolates, however very poorly conserved in SARS and MERS isolates (Table 4). These analyses give an assumption that CEnvA2, CMemA2, CVS1A2, and CVS2A2 are the top peptides for the vaccine in the whole world population.

Homology Modelling and Model Validation

MODELLER modeled the three-dimensional structure of the Envelope protein, Membrane glycoprotein, and Surface glycoproteins through the best multiple template-based modelling approaches (Fig 2). The Envelope protein modeled using 5X29_B, The Membrane glycoprotein modeled using 4N31_B and 5xpd_b, and The Surface glycoprotein modeled using 6ACC_C. The models validated by the PROCHECK server represented as Ramachandran plot and illustrated in Figure S3. In the case of the Envelope protein, 90%, 10%, and 0.0% residues were in the most favoured region, allowed region and disallowed region, respectively. In the case of the Membrane glycoprotein, 82.6%, 15.9%, and 1.4% residues were in the most favored region, allowed region and disallowed region, respectively. While in the case of the Surface glycoprotein, 82.8%, 14.8%, and 1.7% residues were in the most favoured region, allowed region and disallowed region, respectively. The disorder of the protein sequences was measured by the DISOPRED server to retrieve disorder among the targeted sequences (Fig S4). Both analyses showed that the potential peptide placed in the stable part of the protein. Moreover, the proposed epitopes were shown to be on the surface of the protein, give evidence for their surface accessibility (Fig 2).

Allergenicity assessment and Trans-membrane helix prediction

The allergenicity assessment by the AllerTop server showed that CVEnvB2, CVS1A2, CVS1B2, and CVS1C2 were probable allergen (Table 4). The Trans-membrane region prediction by TMHMM server has been depicted in Fig S2 and summarized in Table 4. The potential peptide, CVMemA2, and CVS2A2 found to be located in the transmembrane region of the protein. So, we then go for the next potential peptides CVMemB2 and CVS2B2.

Molecular Docking Analysis.

We selected HLA alleles of MHC class I and MHC class II found from the IEDB analyses interacted with the respective epitopes (Table 2 and 3). The 3D structures HLA alleles retrieved from RCSB-PDB server and docked by CABS-DOC server. The docking interface visualized with the PyMOL Molecular Graphics System. There are several polar and non-polar interactions identified in the docking simulation analyses. The polar contacts extracted by PyMOL and visualized in the figures. The Docking scheme and the receptor amino acid residues interacted for MHC class I is depicted in Figure 4, and MHC class II is depicted in Figure 5. The amino acid residues of peptides CVEnvA1, CVMemB1, CVS1A1, and CVS2B1 interacted with those of MHC class I alleles were illustrated in Figure S5, S6, S7, and S8, respectively. While the amino acid residues of peptides CVEnvA2, CVMemB2, CVS1A2, and CVS2B2 interacted with those of MHC class II alleles were illustrated in Figure S9, S10, S11, and S12, respectively. The polar contacts were shown by red font in the respective figures. The cluster density, average RMSD, maximum RMSD and elements involved also described in Figure 4 and 5.

B-cell Epitope Prediction.

We used the sequence-based approaches for B-cell epitope prediction of the potential peptides CVEnvA2, CVMemB2, CVS1A2, and CVS2B2 by the Kolaskar and Tongaonkar antigenicity scale to assess the antigenic property of the epitope with a maximum propensity score of 1.152, 1.180, 1.095, and 1.183, respectively. Another important benchmark for being a potential B-cell epitope is peptide surface accessibility that was evaluated by Emini surface accessibility of the predicted peptide and found to be with a maximum propensity score of 2.048, 2.471, 2.910, and 2.507 for CVEnvA2, CVMemB2, CVS1A2, and CVS2B2, respectively. The Parker hydrophilicity prediction utilized to find the hydrophilic regions of CVEnvA2, CVMemB2, CVS1A2, and CVS2B2 with a maximum propensity score of 2.5, 0.371, 2.557, and 3.857, respectively. The Karplus & Schulz Flexibility Prediction also utilized to find the flexibility regions of our proposed epitopes of CVEnvA2, CVMemB2, CVS1A2, and CVS2B2 with a maximum propensity score of 1.021, 0.983, 0.998, and 1.043, respectively. These analyses strengthen the prediction that the proposed epitopes might elicit B-cell response (Fig 6).

Multi-epitope vaccine-construction, structural properties and B-cell epitope prediction

The final peptide candidates from all the analyses concluded that CVEnvA2, CVEnvB2, CVMemB2, CVS1A2, and CVS2B2 were the top peptides that can be utilized as vaccines for recognizing the SARS CoV-2 viruses. However, CVEnvB2 and CVS1A2 found to be a probable allergen. This prompted us to modify our predicted peptides. As up to 100 amino acid can be synthesized commercially, we

combine the top four peptides via AAY linker that are components of Envelope, Membrane, S1, and S2 proteins (denoted as CVMW) suitable for the world population. A cysteine residue added at the N-terminal of the multi-epitope peptide that can be utilized for conjugation with a carrier protein. As the peptide, CVEnvA2 has poor population coverage for South Africa than CVEnvB2 (3.15% vs. 40.9%) (Table S11), so we constructed another multi-epitope vaccine suitable for South Africa using the second one (denoted as CVMS). Finally, two multi-epitope constructs designed, CVMW (Fig 7A (i)) and CVMS (Fig 7B (i)) those are 70 amino acid long and found to be antigenic with a Vaxijen score of 0.6525 and 0.6927, respectively (Fig 7 and Table S15). Both the vaccines found to be non-allergic in nature (Table S15). The secondary structural properties and the theoretical physicochemical properties of the vaccines shown in Table S15. The 3D model of CVMW (Fig 7A (ii)) and CVMS (Fig 7B (ii)) constructed using iTASSER. Furthermore, the models were subjected to refinement by the Galaxy Refine server. The finalized models were subjected to ProSA-web to analyze the model quality (Fig 7A (iii) and B (iii)). The results revealed a z score of -4.75 for the model, CVMW, and a z score of -2.41 for the model, CVMS. The overall quality of the finalized model of the multi-epitope vaccine constructs checked by Ramachandran plot analysis. The results revealed 52.5%, 39.7%, 6.3%, and 1.6% residues of CVMW lying in favored, allowed, generously allowed and disallowed regions, whereas, 82.5%, 12.7%, 1.6%, and 3.2% residues of CVMS lying in favored, allowed, generously allowed and disallowed regions respectively (Fig. 7 A (iv) and B (iv)).

The linear/continuous and conformational/discontinuous B cell epitopes in the multi-epitope vaccine construct predicted by using ABCPred and Ellipro server, respectively, considering the default parameters. The servers predicted 4 linear and 3 conformational B cell epitopes for CVMW; and 5 linear and 3 conformational B cell epitopes for CVMS (Table S16 and S17).

Discussion

The novel coronavirus disease 2019 (COVID-19) outbreak has declared a Public Health Emergency of International Concern and number of infections and deaths increasing day by day. Reinfection of SARS-CoV-2 is possible, apparently because of weakening immunity 38, 39. So, *Vaccination* is an effective way to prevent the pandemic virus infection and severe outcomes.

Multi-epitope based (includes conserved multiple epitopes) vaccines designing is a novel approach that represents inducing specific cellular immunity and highly potent neutralizing antibodies 40-42. These epitope-based vaccines also provide increased safety and have the ability to focus on immune responses as they include conserved multiple epitopes 43. He Y et al. demonstrated that recombinant RBD of the S1 subunit consists of multiple conformational immunogenic epitopes that induce highly potent neutralizing antibodies against SARS-CoV 44. The membrane protein of SARS-CoV holds dominant cellular immunogenicity and a strong humoral response 19. The SARS-CoV-2 is similar to SARS-CoV, so, based on previous SARS-CoV immunological studies, we can consider designing potential vaccine targets for the novel coronavirus (SARS-CoV-2) 45. Combining subunit vaccines comprises S1 protein and/or the RBD element, epitope of envelope and membrane protein with adjuvants may turn into a faster and safer strategy to move through early clinical development for the immediate control of SARS-CoV-2 infections 46. These previous findings guided us to multi-epitope based vaccine designing. This study aimed to initial designing of novel multi-epitope based vaccines, based on the conserved region of surface proteins that could be used as promising vaccines against SARS- CoV-2.

To induce specific humoral or cellular immunity against pathogens, an ideal vaccine should contain both B-cell epitopes and T-cell epitopes 47. Initially, we utilized the immunoinformatic tool to find the most immunogenic T-cell epitopes among the three structural proteins (S, E, and M) of SARS-CoV-2, that are expected to elicit an immune response in most of the regions of the world. Based on several rational analysis, five 15.0-mer peptides, two peptides from Envelop protein (CVEnvA2, and CVEnvB2), one peptide from M protein (CVMemB2), two peptides from Spike protein (S1: CVS1A2; S2: CVS2B2) found to be the most potent and highly interacting human leukocyte antigen (HLA) candidates for both MHC class I and MHC class II molecules (Table 4).

According to IEDB analysis, 9.0-mer epitopes of the finalized five peptides; CVEnvA2, CVEnvB2, CVMemB2, CVS1A2, and CVS2B2 were 100% conserved and their projected 15.0-mer epitopes were also 100% conserved except CVS1A2 showed 99.45% conservancy among all the three structural proteins of SARS-CoV-2, respectively (as of 180 isolates)(Table 4). To design a universal vaccine that would protect against SARS-CoV-2 worldwide, the vaccine candidates must have wider population coverage to get the acceptability. In our analysis, the world population coverage of our proposed epitopes CVEnvA2, CVEnvB2, CVMemB2, CVS1A2, and CVS2B2 was 96.77%, 71.88%, 82.11%, 94.07%, and 57.36%, respectively (Table S11-S14 and Fig 3). The study results suggest that proposed epitopes are possessed high conservancy and broader population coverage.

The proposed epitopes and MHC molecules were subjected to docking analysis to analyze the pattern of interaction among them. Molecular docking revealed strong interactions between both 9.0-mer and 15.0-mer predicted epitopes and the respective HLA alleles. These bindings with lower RMSD values and higher cluster densities shown from docking simulation study strengthen the evidence about the interaction of peptide-MHC class I and II alleles.

Finally, two suitable multi-epitope vaccine constructs have been designed; at each construct, the top four peptides from four proteins combined via AAY linker. In this context, the epitope presentation enhanced by the AAY linker motif. TAP transporter or other chaperons bind to the C-terminal of epitopes which are generated after the cleavage behind the AAY motifs of the combined multi-epitope 48. The multi-epitope vaccine constructs, CVMW comprises CEnvA2, CVMemB2, CVS1A2, and CVS2B2 peptides, and a cysteine residue at the N-terminal of the construct (Fig 7A (i)). CVMW construct is suitable for the world population. The second vaccine constructs, CVMS, comprises CEnvB2, CVMemB2, CVS1A2, and CVS2B2 peptides and a cysteine residue at the N-terminal of the construct (Fig 7B (i)). The CVMS construct was designed for covering the South African population as CEnvA2 (CVMW component) showed poor population coverage for South Africa than CEnvB2 (3.15% vs. 40.9%). Both multi-epitope vaccine constructs were 70 amino acid long and found to be antigenic with a Vaxijen score of 0.6525 and 0.6927, respectively. Both the vaccines found to be non-allergic in nature, according to the FAO/WHO allergenicity evaluation scheme (Table S15). In most cases, multi-epitope vaccines designed by assembling antigenic portions of the proteins, and the construct are long enough to be expressed as well as purified for further experiments. Our construct is less than 100 amino acids that can easily be synthesized for animal experiments, reducing a huge amount of work.

The overall quality of the finalized models of the multi-epitope vaccine constructs was checked by the PROCHECK server, represented as Ramachandran plot (Fig. 7 A (iv) and B (iv)). Moreover, the proposed epitopes were shown to be on the surface of the protein (Fig 2). Based on the previously described approaches, 31–34 each proposed peptide can also elicit B-cell responses. In the case of the multi-epitope vaccine, 4 linear and 3 conformational B cell epitopes for CVMW, and 5 linear and 3 conformational B cell epitopes for CVMS also predicted (Table S16 and S17). Predicted B cell epitopes also expected to elicit strong neutralizing antibody responses. If a strong B-cell response occurred in animal experiments (mice or rabbit), these antibodies can be used in diagnostic purposes, as they should recognize the prominent antigens in the viral surface.

We acknowledge that this is a preliminary study based on the limited sequence data available for SARS-CoV-2. As this RNA virus contains an error-prone RNA polymerase, they have high mutation rates. Though SARS-CoV-2 continues to evolve, such mutations may not affect our analysis as of the inclusion of four peptides from four proteins in a multi-epitope vaccine with the high conservancy. Further, *in vitro* and *in vivo* studies are required to determine the potentiality of the constructed multi-epitope vaccines to induce a positive immune response against SARS-CoV-2.

Conclusion

In this study, to design a potential epitope-based vaccine candidate against coronavirus disease 19, two multi-epitope constructs were defined from a highly conserved domain of SARS-CoV-2 antigens (S1, S2, E, and M). After narrowing down the window of lots of potential epitopes, we designed probable top vaccine, CVMS, suitable for South Africa and CVMW, suitable for the rest of the world. So it can be concluded that the proposed multi-epitope vaccine has the potential to stimulate a high degree of protection against the pandemic SARS-CoV-2.

Declarations

Funding

No funding was obtained for this study.

Acknowledgements

We thank all the authors, the originating and submitting laboratories for their sequence and metadata shared through GISAID, on which this conservancy analyses performed.

Ethical Approval

As samples from human or animals had not been used in this study, ethical clearance is not applicable.

Conflicts of Interest

No potential competing interest was reported by the authors.

Authors' Contributions

MSBI and HU conceived the study and carried out the immunoinformatic analyses. MM performed in conservancy analyses and drafted the manuscript. KMKK conceived, designed, and guided the study; performed immunoinformatic analyses, drafted and critically revised the manuscript; all authors read and approved the final manuscript.

References

1. Guan, W.-j. *et al.* Clinical characteristics of 2019 novel coronavirus infection in *MedRxiv* (2020).
2. Reuters Coronavirus incubation could be as long as 27 days, Chinese provincial government says. Reuters, 22, 2020. Retrieved on 21 March, 2020. (2020).
3. Organization, W.H. Laboratory testing for 2019 novel coronavirus (2019-nCoV) in suspected human cases, World Health Organization (WHO). <https://www.who.int/health-topics/coronavirus/laboratory-diagnostics-for-novel-coronavirus>. (2020).
4. Chen, Y., Liu, & Guo, D. Emerging coronaviruses: genome structure, replication, and pathogenesis. *Journal of medical virology* (2020).
5. Wu, F., Zhao, S. & Yu, B. A new coronavirus associated with human respiratory disease in China.[published on February 03, 2020]. *Nature*.
6. Chan, J.F.-W. *et al.* Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections* **9**, 221-236 (2020).
7. Jiang, S., He, Y. & Liu, SARS vaccine development. *Emerging infectious diseases* **11**, 1016 (2005).
8. Regla-Nava, J.A. *et al.* Severe acute respiratory syndrome coronaviruses with mutations in the E protein are attenuated and promising vaccine candidates. *Journal of virology* **89**, 3870- 3887 (2015).
9. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 1-4 (2020).
10. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* **395**, 565-574 (2020).
11. Letko, M.C. & Munster, V. Functional assessment of cell entry and receptor usage for lineage B β -coronaviruses, including 2019-nCoV. *bioRxiv* (2020).
12. Hoffmann, M. *et al.* The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells. *BioRxiv* (2020).
13. Enjuanes, L. *et al.* Vaccines to prevent severe acute respiratory syndrome coronavirus- induced disease. *Virus research* **133**, 45-62 (2008).
14. Yang, Z.-y. *et al.* A DNA vaccine induces SARS coronavirus neutralization and protective immunity in *Nature* **428**, 561-564 (2004).
15. Deming, D. *et al.* Vaccine efficacy in senescent mice challenged with recombinant SARS- CoV bearing epidemic and zoonotic spike variants. *PLoS medicine* **3** (2006).
16. Wan, Y., Shang, J., Graham, R., Baric, R.S. & Li, Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *Journal of virology* **94** (2020).
17. Song, Z. *et al.* From SARS to MERS, thrusting coronaviruses into the spotlight. *Viruses* **11**, 59 (2019).
18. Babcock, G.J., Eshshaki, D.J., Thomas, W.D. & Ambrosino, D.M. Amino acids 270 to 510 of the severe acute respiratory syndrome coronavirus spike protein are required for interaction with receptor. *Journal of virology* **78**, 4552-4560 (2004).
19. Liu, *et al.* The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes. *The Journal of infectious diseases* **202**, 1171-1180 (2010).

20. Gralinski, L.E. & Menachery, V.D. Return of the Coronavirus: 2019-nCoV. *Viruses* **12**, 135 (2020).
21. Hossain, M.U. *et al.* Design of peptide-based epitope vaccine and further binding site scrutiny led to groundswell in drug discovery against Lassa virus. *3 Biotech* **8**, 81 (2018).
22. Oany, A.R. *et al.* Vaccinomics Approach for Designing Potential Peptide Vaccine by Targeting Shigella spp. Serine Protease Autotransporter Subfamily Protein SigA. *J Immunol Res* **2017**, 6412353 (2017).
23. Chauhan, V., Rungta, T., Goyal, K. & Singh, M.P. Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach. *Sci Rep* **9**, 2517 (2019).
24. Larsen, M.V. *et al.* Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* **8**, 424 (2007).
25. Moise, L. *et al.* In silico-accelerated identification of conserved and immunogenic variola/vaccinia T-cell epitopes. *Vaccine* **27**, 6471-6479 (2009).
26. Bui, H.H. *et al.* Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* **7**, 153 (2006).
27. Sali, , Potterton, L., Yuan, F., van Vlijmen, H. & Karplus, M. Evaluation of comparative protein modeling by MODELLER. *Proteins* **23**, 318-326 (1995).
28. Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R. & Thornton, J.M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8**, 477-486 (1996).
29. Ward, J.J., McGuffin, L.J., Bryson, K., Buxton, B.F. & Jones, D.T. The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **20**, 2138-2139 (2004).
30. Liao, L. & Noble, S. Combining pairwise sequence similarity and support vector machines for detecting remote protein evolutionary and structural relationships. *J Comput Biol* **10**, 857-868 (2003).
31. Kolaskar, A.S. & Tongaonkar, P.C. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* **276**, 172-174 (1990).
32. Emini, E.A., Hughes, J.V., Perlow, D.S. & Boger, J. Induction of hepatitis A virus- neutralizing antibody by a virus-specific synthetic peptide. *J Virol* **55**, 836-839 (1985).
33. Parker, J.M., Guo, D. & Hodges, R.S. New hydrophilicity scale derived from high- performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* **25**, 5425-5432 (1986).
34. Karplus, P.A. & Schulz, G.E. Prediction of chain flexibility in *Naturwissenschaften* **72**, 212-213 (1985).
35. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7-8 (2015).
36. Wiederstein, M. & Sippl, J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* **35**, W407-410 (2007).
37. Ponomarenko, J. *et al.* ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* **9**, 514 (2008).
38. CDC CDC Reports first known employee with COVID-19 Infection. <https://cdc.gov/media/releases/2020/s0316-cdc-employee-covid-19.html>. Retrieved on 16 March 2020. (2020).
39. Forbes Can You Get Coronavirus Twice? How Long Are You Immune After COVID-19? <https://forbes.com/sites/brucelee/2020/03/15/can-you-get-infected-by-coronavirus-twice-how-does-covid-19-immunity-work/#4ffe1df35c0f>. Retrieved on Mar 15, 2020. (2020).
40. Mahmoodi, S., Nezafat, N., Sarmadi, , Zarghami, N. & Ghasemi, Y. Expression and purification of a novel multi-epitope peptide vaccine for breast cancer immunotherapy. *Minerva Biotechnologica* **29**, 1-7 (2017).
41. Dawood, R.M. *et al.* A multiepitope peptide vaccine against HCV stimulates neutralizing humoral and persistent cellular responses in *BMC infectious diseases* **19**, 932 (2019).
42. Vakili, B. *et al.* A new multi-epitope peptide vaccine induces immune responses and protection against *Leishmania infantum* in BALB/c mice. *Medical microbiology and immunology*, 1-11 (2019).
43. Zhou, W.-Y. *et al.* Therapeutic efficacy of a multi-epitope vaccine against *Helicobacter pylori* infection in BALB/c mice *Vaccine* **27**, 5013-5019 (2009).

44. He, Y., Lu, H., Siddiqui, P., Zhou, Y. & Jiang, S. Receptor-binding domain of SARS coronavirus spike protein contains multiple conformational-dependant epitopes that induce highly potent neutralizing antibodies. *Immunol* **174**, 4908-4915 (2005).
45. Ahmed, S.F., Quadeer, A.A. & McKay, M.R. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses* **12**, 254 (2020).
46. Shang, W., Yang, Y., Rao, Y. & Rao, X. The outbreak of SARS-CoV-2 pneumonia calls for viral vaccines. *npj Vaccines* **5**, 1-3 (2020).
47. Purcell, A.W., McCluskey, J. & Rossjohn, More than one reason to rethink the use of peptides in vaccine design. *Nature reviews Drug discovery* **6**, 404-414 (2007).
48. Livingston, B. *et al.* A rational strategy to design multi-epitope immunogens based on multiple Th lymphocyte epitopes. *The Journal of Immunology* **168**, 5499-5506 (2002).

Tables

Table 1: The predicted T-cell epitopes of Envelope protein, Membrane glycoprotein, and S1 and S2 domain of spike surface glycoprotein by the NetCTL server based on the combined score. Here, epitopes for the 10 different HLA supertypes of Envelope protein, 12 HLA supertypes of Membrane glycoprotein, S1 and S2 domain have been represented.

Super types	Predicted epitopes for Envelope protein (start position)	Combined score	Predicted epitopes for Membrane glycoprotein (start position)	Combined score	Predicted epitopes for S1 protein (start position)	Combined score	Predicted epitopes for S2 protein (start position)	Combined score
a1	LTALRLCAY (34)	2.3303	ATSRTLSYY (171)	3.7695	CVADYSVLY (361)	2.3459	LTDEMIAQY (865)	5.2958
a2	SLVKPSFYV (50)	1.4461	GLMWLSYFI (89)	1.5462	KIADYNYKL (417)	1.116	FIAGLIAIV (1220)	1.4022
a3	SFYVYSRVK (55)	1.3954	AVILRGHLR (142)	1.4462	SVYAWNRRK R (349)	1.431	VTYVPAQEK (1065)	1.7543
a24	VFLLVTLAI (25)	1.2004	RYRIGNYKL (198)	1.3445	PYRVVLSF (507)	1.6823	VYDPLQPEL (1137)	1.2944
a26	LTALRLCAY (34)	1.5975	ATSRTLSYY (171)	1.385	RISNCVADY (357)	1.4543	GTITSGWTF (880)	1.4942
b4					GEVFNATRF (339)	1.6183	AEIRASANL (1016)	1.6389
b7	FVSEETGTL (4)	1.1281	FAAYSRYRI (193)	1.2196	APATVCGPK (520)	0.8604	SPRRARSVA (680)	1.8685
b8	YVYSRVKN L (57)	1.1002	ELKKLLEQ W (12)	1.2715	PYRVVLSF (507)	1.6219	QDKNTQEVF (774)	1.4769
b27			SRYRIGNYK (197)	2.8527	TRFASVYAW (345)	0.8532	GRLQSLQTY (999)	2.6259
b39	FVSEETGTL (4)	0.9372	SFNPETNIL (111)	1.5475	IRGDEVQRQI (402)	1.9243	AHFPPREGVF (1087)	2.2944
b44			KEITVATSR (166)	1.1148				
b58	VSLVKPSFY (49)	1.2216	KLIFLWLLW (50)	1.8977				
b59					ASFSTFKCY (372)	1.139	QSAPHGVPF (1054)	1.9203
b62	LVKPSFYVY (51)	1.6543	WLSYFIASF (92)	1.6198	GQTGKIADY (413)	2.962	LGAENSVAY (699)	1.4998

Table 2: The potential CD8+ T-cell epitopes and their respective number of MHC class I alleles for Envelope protein, Membrane glycoprotein, S1 and S2 domain of spike surface glycoprotein (IC50 < 250 nM).

Predicted epitopes for Envelope protein	Number of MHC class 1 alleles	Predicted epitopes for Membrane glycoprotein	Number of MHC class 1 alleles	Predicted epitopes for S1 protein	Number of MHC class 1 alleles	Predicted epitopes for S2 protein	Number of MHC class 1 alleles
LTALRLCAY	3	ATSRTLSYY	8	CVADYSVLY	9	LTDEMIAQY	4
SLVKPSFYV	4	GLMWLSYFI	5	KIADYNYKL	7	FIAGLIAIV	5
SFYVYSRVK	2	AVILRGHLR	6	SVYAWNRRK	7	VTYVPAQEK	7
VFLVTLAI	1	RYRIGNYKL	6	PYRVVLSF	6	VYDPLQPEL	6
LTALRLCAY	3	ATSRTLSYY	8	RISNCVADY	5	GTITSGWTF	4
FVSEETGTL	4	FAAYSRYRI	4	GEVFNATRF	4	AEIRASANL	5
YVYSRVKNL	1	ELKKLLEQW	2	APATVCGPK	3	SPRRARSVA	3
FVSEETGTL	4	SRYRIGNYK	6	PYRVVLSF	6	QDKNTQEVF	3
VSLVKPSFY	1	SFNPETNIL	5	TRFASVYAW	8	GRLQSLQTY	7
LVKPSFYVY	4	KEITVATSR	2	IRGDEVQRQI	5	AHFPREGVF	5
		KLIFLWLLW	4	ASFSTFKCY	7	QSAPHGVVF	7
		WLSYFIASF	5	GQTGKIADY	3	LGAENSVAY	5

Table 3: The potential epitopes for CD4+ T-cells and their interacting MHC class II alleles with IC50 < 200nM. Here core epitopes are shown in bold font.

Predicted epitopes for Envelope protein	Number of MHC class 2 alleles	Predicted epitopes for Membrane glycoprotein	Number of MHC class 2 alleles	Predicted epitopes for S1 protein	Number of MHC class 2 alleles	Predicted epitopes for S2 protein	Number of MHC class 2 alleles
LAILTALR LCAYCCN	6	VIGAVILR GHLRIAG	119	ISNCVADY SVLYNSA	25	IWLGFIAG LIAIMV	38
MYSFVSEE TGTLIVN	15	FAAYSRYR IGNYKLN	52	FASVYAW NRKRISNC	11	FLHVITYVP AQEKNFT	28
KPSFYVYS RVKNLNS	127	GFAAYS YRIGNYKL	64	NRKRISNC VADYSVL	5	YTMSLGA ENSVAYS	2
NIVNVSLV KPSFYVY	98	FAAYSRYR IGNYKLN	1	FNATRFAS VYAWNRRK	43		
		VGLMWLS YFIASFRL	164	ADSFVIRG DEVQRQIA	50		

Table 4: Conservancy, allergenicity and transmembrane location prediction of the selected peptides. All the peptides were denoted by a distinct name.

Name of the Peptides	Peptide sequences	Conservancy among COVID-19 isolates	Conservancy among other SARS and MERS viruses	Allergenicity prediction by AllerTop v. 2.0	Trans-membrane location (T*) Prediction
CVEnvA1	YVYSRVKNL	100% (180/180)	37.66% (29/77)	NON-ALLERGEN	Non-T
CVEnvB1	VSLVKPSFY	100% (180/180)	1.30% (1/77)	NON-ALLERGEN	Non-T
CVMemA1	WLSYFIASF	100% (180/180)	7.29% (7/96)	NON-ALLERGEN	T
CVMemB1	AVILRGHLR	100% (180/180)	3.12% (3/96)	ALLERGEN	Non-T
CVS1A1	TRFASVYAW	100% (180/180)	0.00% (0/97)	ALLERGEN	Non-T
CVS1B1	IRGDEVROI	100% (180/180)	0.00% (0/97)	ALLERGEN	Non-T
CVS1C1	CVADYSVLY	96.12% (173/180)	95.88% (93/97)	ALLERGEN	Non-T
CVS2A1	FIAGLIAIV	99.45% (179/180)	94.85% (92/97)	NON-ALLERGEN	T
CVS2B1	VTYVPAQEK	100% (180/180)	1.03% (1/97)	ALLERGEN	Non-T
CVEnvA2	KPSFYVYSRVKNLNS	100% (180/180)	0.00% (0/77)	NON-ALLERGEN	Non-T
CVEnvB2	NIVNVSLVKPSFYVY	100% (180/180)	0.00% (0/77)	ALLERGEN	Non-T
CVMemA2	VGLMWLSYFIASFRL	100% (180/180)	7.29% (7/96)	NON-ALLERGEN	T
CVMemB2	VIGAVILRGHLRIAG	100% (180/180)	3.12% (3/96)	NON-ALLERGEN	Non-T
CVS1A2	FNATRFASVYAWNRK	99.45% (179/180)	0.00% (0/97)	ALLERGEN	Non-T
CVS1B2	ADSFVIRGDEVROIQA	100% (180/180)	0.00% (0/97)	ALLERGEN	Non-T
CVS1C2	ISNCVADYSVLYNSA	96.12% (173/180)	0.00% (0/97)	ALLERGEN	Non-T
CVS2A2	IWLGFIAGLIAIVMV	99.45% (179/180)	0.00% (0/97)	NON-ALLERGEN	T
CVS2B2	FLHVITYVPAQEKNFT	100% (180/180)	1.03% (1/97)	NON-ALLERGEN	Non-T

*T= Located in transmembrane region, Non-T= Not located in transmembrane region

Figures

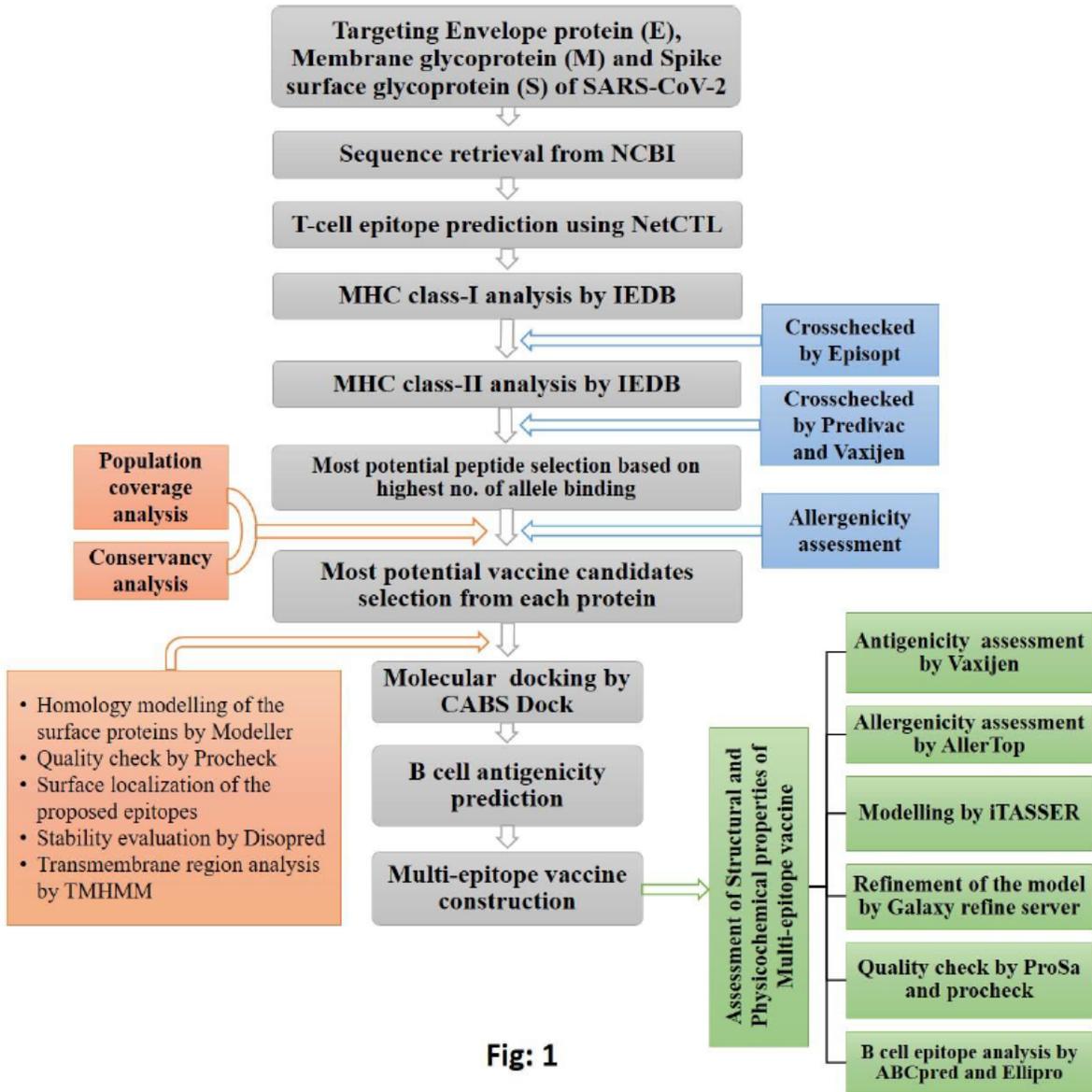


Fig: 1

Figure 1

Flow diagram of the methodology to design multi-epitope vaccine against novel coronavirus (SARS-CoV-2).

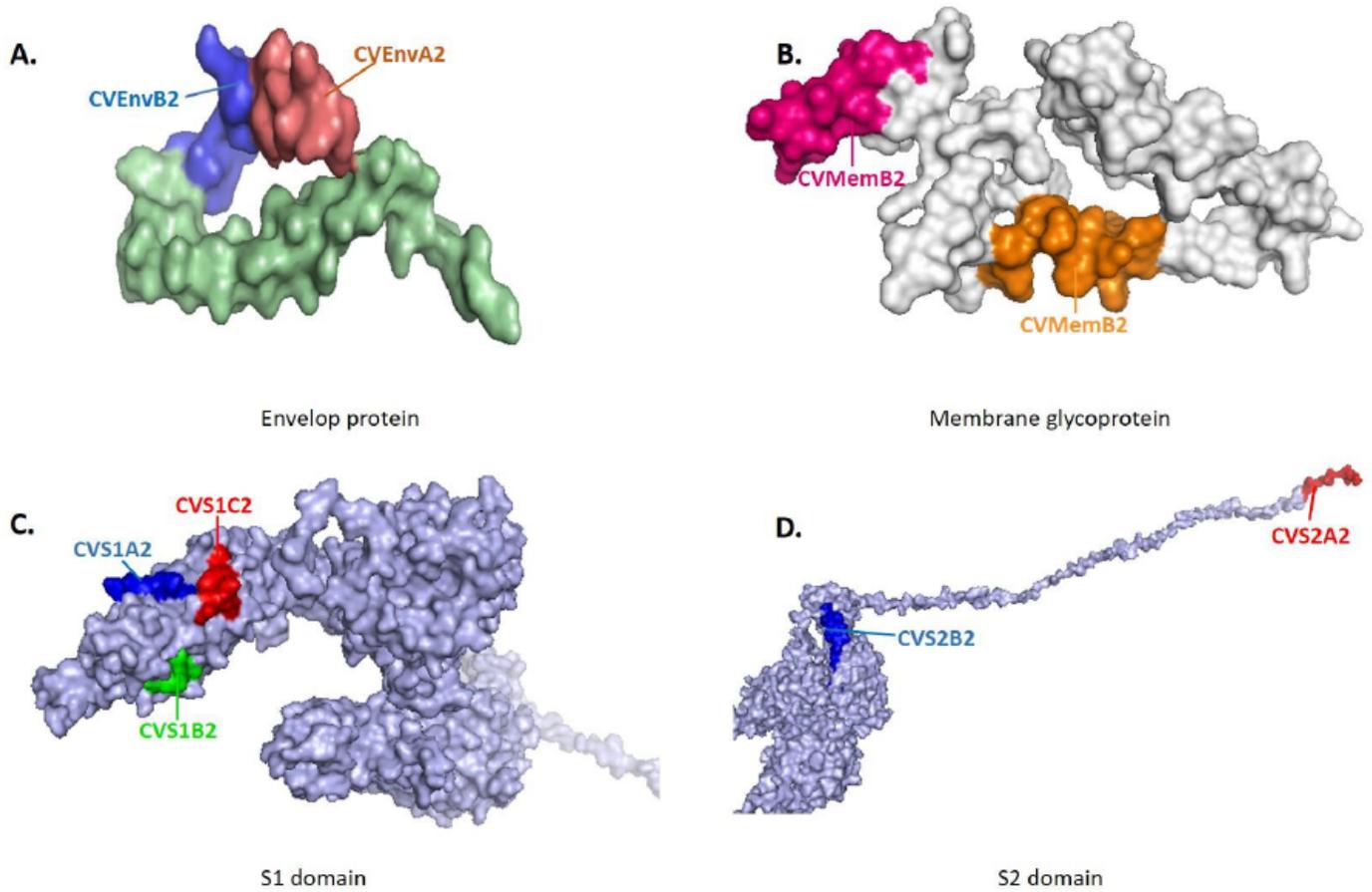


Fig: 2

Figure 2

The 3D models of the Envelope protein (A), Membrane glycoprotein (B) and Surface glycoprotein (C and D). The surface location of the T-cell epitopes on the protein, A. CVEnvA2 (light red) and CVEnvB2 (Light Blue), B. CVMemA2 (Orange) and CVMemB2 (Pink), C. CVS1A2 (Blue), CVS1B2 (Green) and CVS1C2 (Red) and D. CVS2A2 (Red) and CVS2B2 (Blue).

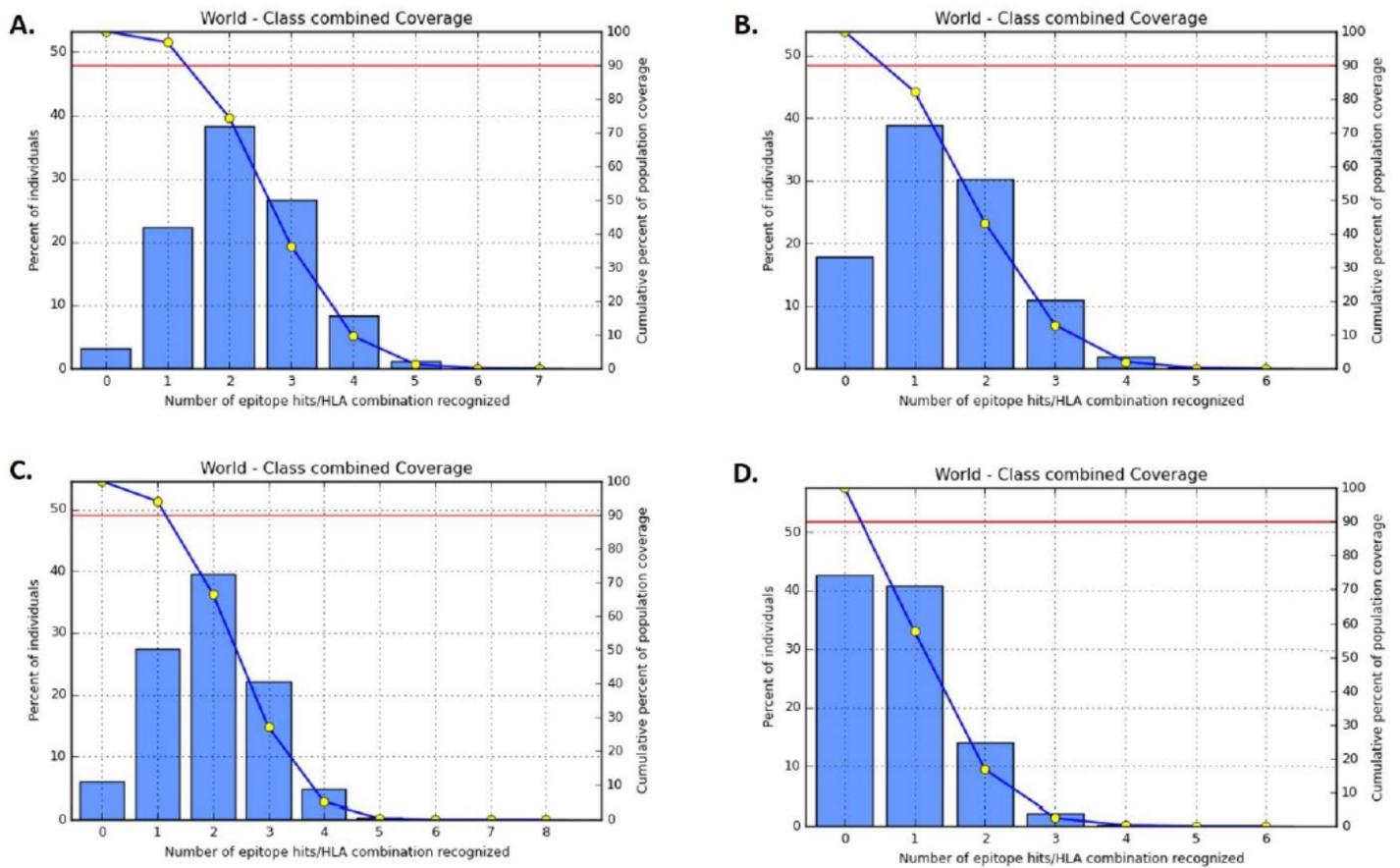


Fig: 3

Figure 3

Population coverage analysis for four prominent epitopes on the basis of HLA interaction in the whole world. (A) The population coverage for epitope CVEnA2, (B) The population coverage for epitope CVMemB2 (C) The population coverage for epitope CVS1A2 (D) The population coverage for epitope CVS2B2. The figures were prepared depending on the combined score of both MHC class I and MHC class II interaction. Notes: the line (-o-) in the graphs, points out the cumulative percentage of population coverage of the epitopes; the bars indicate the population coverage for each epitope.

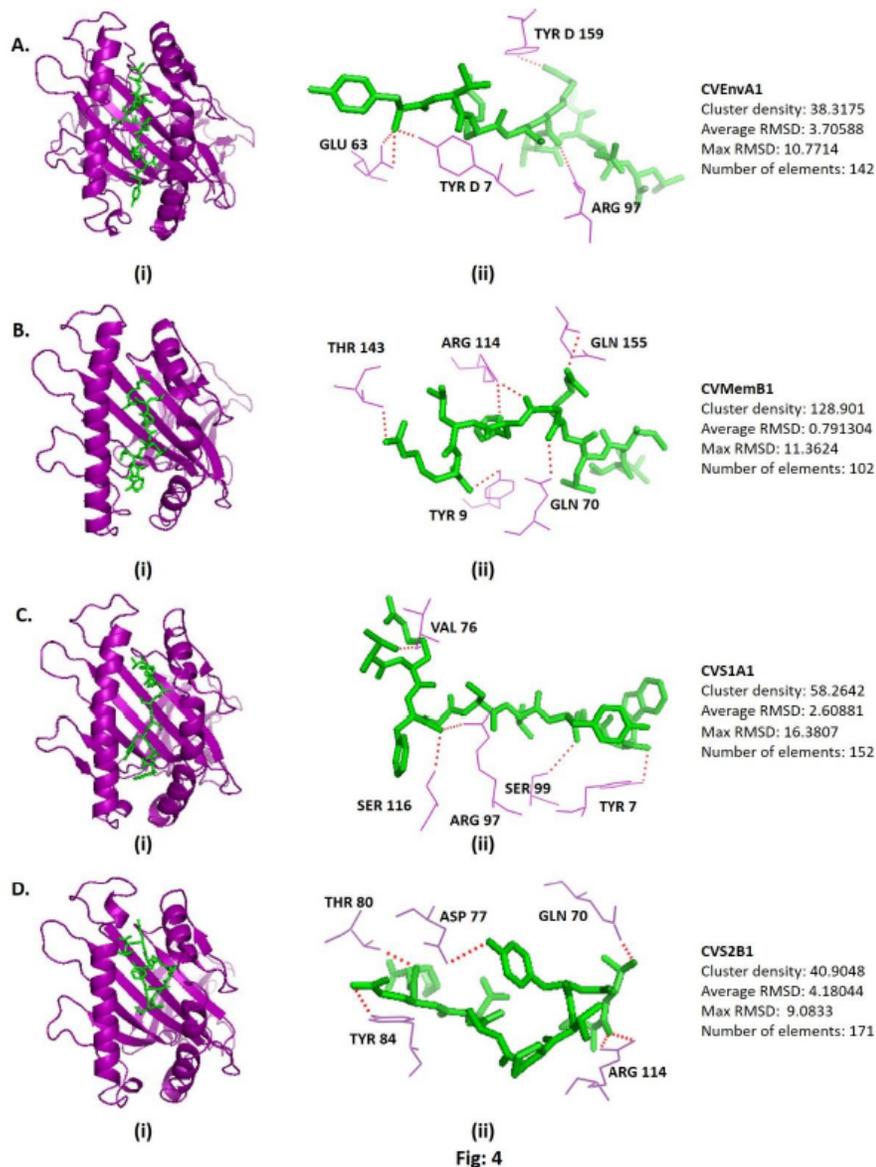


Figure 4

Docking analysis of the proposed 9-mer epitopes with MHC class I molecule. Docking of the epitope for (A) envelope protein, CVEnvA1 with the allele HLA-A*02:03 (3ox8), (B) Membrane glycoprotein, CVMemB1 with the alleles HLA-A*68:01 (6PBH), (C) S1 domain of Spike surface glycoprotein, CVS1A1 with alleles HLA-C*07:02 (5VGE), and (D) S2 domain of Spike surface glycoprotein, CVS2B1 with the alleles HLA-A*11:01 (6JP3). i. Showing the cartoon view. ii representing the interaction between the amino acid residues of HLA and the peptide. The cluster density, average RMSD, maximum RMSD and elements involved also described right side of the respective docked figure.

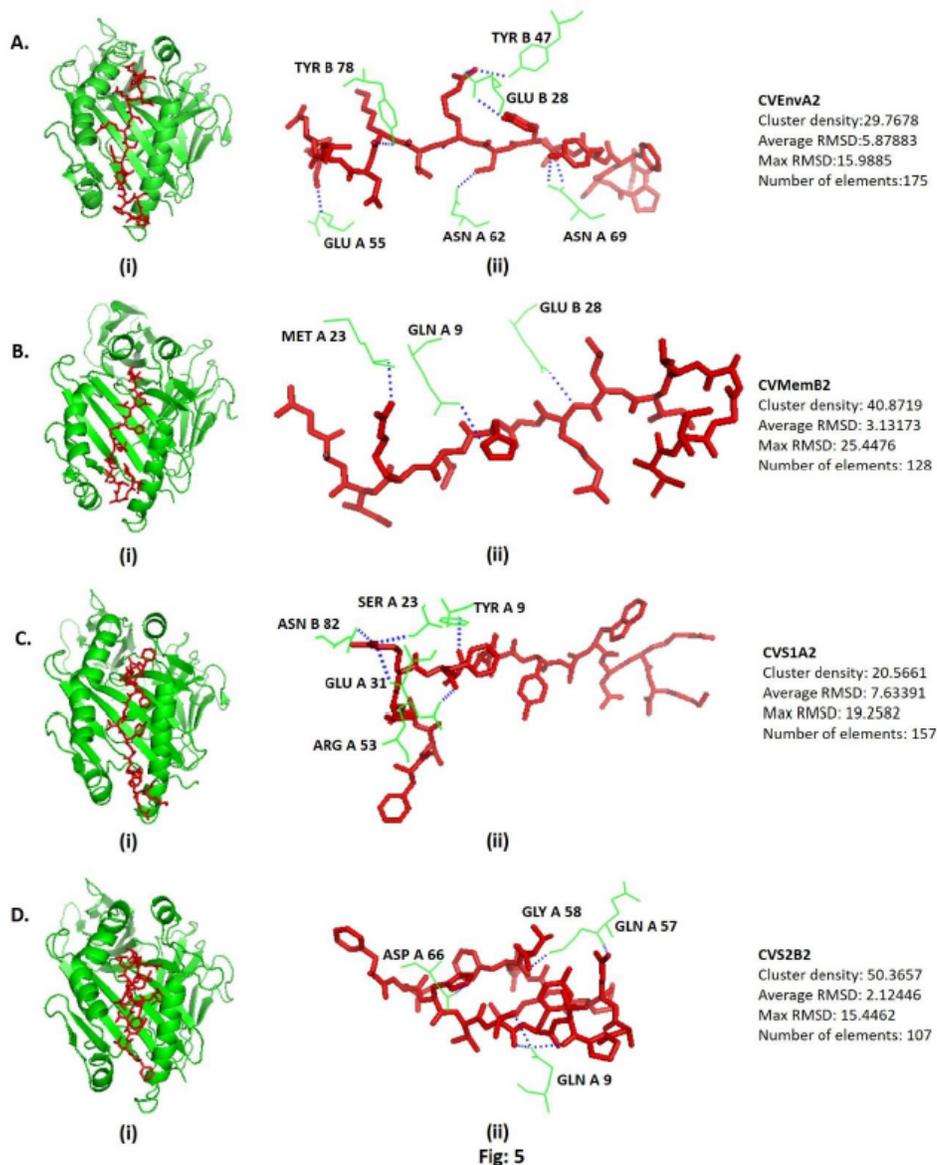


Figure 5

Docking analysis of the proposed 15-mer epitopes with MHC class II molecule. Docking of the epitope for (A) envelope protein, CVEnvA2 with the allele HLA-DRB1*01:01 (2FSE), (B) Membrane glycoprotein, CVMemB2 with the alleles HLA-DRB1*01:01 (2FSE), (C) S1 domain of Spike surface glycoprotein, CVS1A2 with alleles DQA1/DQB1 (1JK8), and (D) S2 domain of Spike surface glycoprotein, CVS2B2 with the alleles HLA-DRB1*01:01 (2FSE). i. Showing the cartoon view. ii representing the interaction between the amino acid residues of HLA and the peptide. The cluster density, average RMSD, maximum RMSD and elements involved also described right side of the respective docked figure.

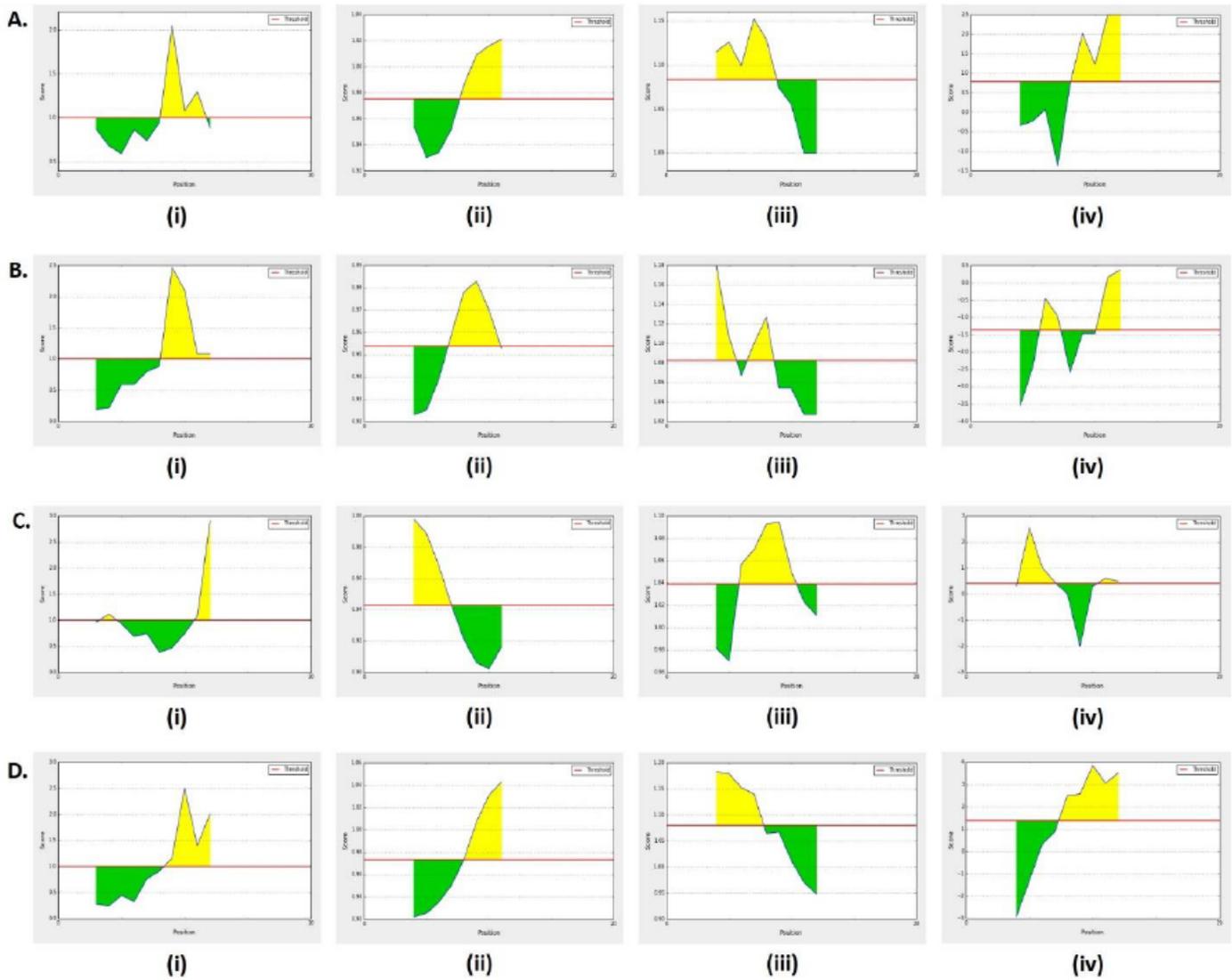


Fig: 6

Figure 6

Prediction of B-cell epitope for epitope (A) CVEnvA2, (B) CVMemB2, (C) CVS1A2 and (D) CVS2B2. The column showed the analyses of the respective peptides with i. Emini surface accessibility prediction, ii. Karplus & Schulz Flexibility Prediction, iii. Kolaskar and Tongaonkar antigenicity prediction and, iv. Parker hydrophilicity prediction. Notes: the x-axis and y-axis of the plot reflects the sequence location and antigenic propensity, respectively. The areas over the threshold line are antigenic that are labeled as yellow.

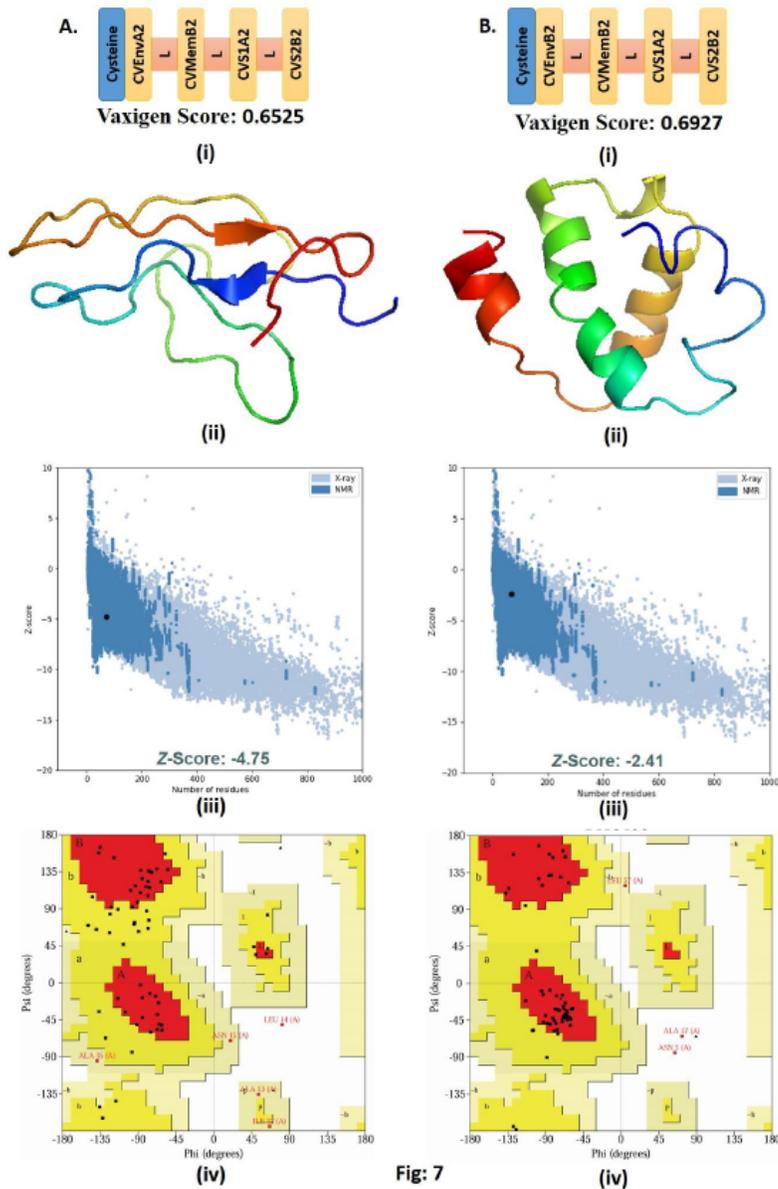


Fig: 7

Figure 7

Schematic diagram of final multi-epitope vaccine construct proposed for A (i) whole world (CVMW); B (i) South Africa (CVMS). 'L' is indicated for the AAY linker (Light red) used for linking the most prominent T-cell epitopes found from the whole analyses. A (ii) and B (ii) represents the 3D model of the multi-epitope vaccine construct CVMW and CVMS, respectively. A (iii) and B (iii) represents the validation of the respective models using ProSa. A (iv) and B (iv) Ramachandran plot analysis of the respective models using PROCHECK server.

Supplementary Files

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

- [Supplementaryinformation.pdf](#)