Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Computers in Biology and Medicine 136 (2021) 104703

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

An in-depth *in silico* and immunoinformatics approach for designing a potential multi-epitope construct for the effective development of vaccine to combat against SARS-CoV-2 encompassing variants of concern and interest

Abdullah Al Saba [a], Maisha Adiba [a], Piyal Saha [a], Md. Ismail Hosen [a], Sajib Chakraborty [b], A.H. M. Nurun Nabi [a,*]

[a] Laboratory of Population Genetics, Department of Biochemistry and Molecular Biology, University of Dhaka, Bangladesh
[b] Molecular Systems Biology Laboratory, Department of Biochemistry and Molecular Biology, University of Dhaka, Bangladesh

ARTICLE INFO

Keywords:
COVID-19
SARS-CoV-2
T cell epitope
B cell epitope
Multi-epitope vaccine
Immunoinformatics

ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the latest of the several viral pathogens that have acted as a threat to human health around the world. Thus, to prevent COVID-19 and control the outbreak, the development of vaccines against SARS-CoV-2 is one of the most important strategies at present. The study aimed to design a multi-epitope vaccine (MEV) against SARS-CoV-2. For the development of a more effective vaccine, 1549 nucleotide sequences were taken into consideration, including the variants of concern (B.1.1.7, B.1.351, P.1 and, B.1.617.2) and variants of interest (B.1.427, B.1.429, B.1.526, B.1.617.1 and P.2). A total of 11 SARS-CoV-2 proteins (S, N, E, M, ORF1ab polyprotein, ORF3a, ORF6, ORF7a, ORF7*b*, ORF8, ORF10) were targeted for T-cell epitope prediction and S protein was targeted for B-cell epitope prediction. MEV was constructed using linkers and adjuvant beta-defensin. The vaccine construct was verified, based on its antigenicity, physicochemical properties, and its binding potential, with toll-like receptors (TLR2, TLR4), ACE2 receptor and B cell receptor. The selected vaccine construct showed considerable binding with all the receptors and a significant immune response, including elevated antibody titer and B cell population along with augmented activity of T_H cells, Tc cells and NK cells. Thus, immunoinformatics and *in silico*-based approaches were used for constructing MEV which is capable of eliciting both innate and adaptive immunity. In conclusion, the vaccine construct developed in this study has all the potential for the development of a next-generation vaccine which may in turn effectively combat the new variants of SARS-CoV-2 identified so far. However, *in vitro* and animal studies are warranted to justify our findings for its utility as probable preventive measure.

1. Introduction

The human race has encountered several bacterial and viral pathogens that have acted as a threat to human health around the world. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the latest of them. The coronavirus disease (COVID-19) has emerged as the third human zoonosis of this century and has become significantly harmful to human health [1,2]. Previously, in 2002, coronavirus (CoV) belonging to the *Betacoronavirus* genus infected humans to cause a severe acute respiratory syndrome (SARS-CoV). The Middle East respiratory syndrome (MERS-CoV) occurred in 2012, also belonged to the same genus. But neither of those has caused the vast irreparable loss to humankind as of the pandemic of SARS-CoV-2 [3,4]. This deadly virus contains an unsegmented, 3′ polyadenylated and 5′ capped positive-sense, single-stranded, and ~30 kilobases long RNA genome which is 79.0% and 50.0% identical to the genomes of SARS-CoV and MERS-CoV, respectively [5]. This genome in a sequential manner, results in non-structural and structural proteins such as - ORF1ab, surface glycoprotein (S), ORF3a, envelop protein (E), membrane glycoprotein (M), ORF6, ORF7a, ORF7b, ORF8, nucleocapsid phosphoprotein (N), and ORF10 proteins [6].

Thus, to prevent COVID-19 and control the outbreak, the

* Corresponding author. Laboratory of Population Genetics, Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka, 1000, Bangladesh.
E-mail address: nabi@du.ac.bd (A.H.M.N. Nabi).

https://doi.org/10.1016/j.compbiomed.2021.104703
Received 2 May 2021; Received in revised form 25 July 2021; Accepted 26 July 2021
Available online 30 July 2021

development of vaccines against SARS-CoV-2 is one of the most important strategies. Till now, more than 180 different types of vaccines (e.g., mRNA, DNA, subunit etc vaccines) are under development. Among these, the clinical trials of 42 vaccine candidates have been started while 10 are in phase III trials [7]. CoVaxin from Bharat Biotech (inactivated virus vaccine) [8], Gam-Covid-Vac or Sputnik V from Russia (viral vector vaccine) [9], Moderna (mRNA1273) [10] and Pfizer (BNT162b2) [11] (RNA and DNA vaccines) are some of the vaccines that are either approved or in trial stages. Among them, Pfizer and Moderna vaccine are stated to be 95% and 94.1% effective according to the clinical trials in people with no evidence of previous COVID-19 infection [12]. However, for Moderna the success rate dropped to 86.4% for 65 or more than 65 years old people, and Pfizer was suspected to be less effective against the South African (B.1.351) variant [12]. Also, according to Bharat Biotech, the efficacy in those without prior infection of CoVaxin is 81% in neutralizing the UK variant and other heterologous strains [13]. On the other hand, Sputnik V, a recombinant adenovirus-based vaccine, was reported to be 91.6% effective in lessening the disease's severity [14]. However, it is still suspected to be effective against a fast-spreading variant [15].

Thus, although it is too early to predict these vaccines' ability to provide long-term protection, at present it is affirmed that they can protect human for 6–8 months after a booster vaccination by developing cellular immune response [16]. However, the complex genetic makeup and high mutation rate of the SARS-CoV-2 make the outcome of the vaccine response uncertain. Thus, a strategic development of a vaccine by targeting all the proteins is required [17].

The high mutation frequency of SARS-CoV-2 is the major concern at the present time. The Centre for Disease Control and Prevention (CDC) has designated B.1.1.7, B.1.351, P.1 and B.1.617.2 variants as "variants of concern" [18]. B.1.1.7, B.1.351, P.1, and B.1.617.2 are the Pango lineage nomenclatures [19] and the corresponding Nextstain nomenclatures are 20I/501Y.V1, 20H/501.V2, 20J/501Y.V3 and 20A/S:478K, respectively. B.1.1.7, B.1.351, P.1 and B.1.617.2 variants were first detected in United Kingdom, South Africa, Japan/Brazil and India, respectively. Variants B.1.427, B.1.429, B.1.525, B.1.526, B.1.617.1, B.1.617.3 and P.2 were designated as "variants of interest" by CDC [18]. WHO labeled the variants B.1.1.7, B.1.351, P.1, P.2, B.1.525, B.1.526, B.1.617.1 and B.1.617.2 as alpha, beta, gamma, zeta, eta, iota, kappa and delta, respectively. Both variants B.1.427 and B.1.429 had been designated as epsilon by WHO [18]. All the variants of concern and interest harbor the *D614G mutation.* This mutation in the spike protein increases the infectivity of SARS-CoV-2 virus [20]. Serum neutralization titers of B.1.351 were lower compared to P.1 and B.1.1.7, whereas for P.1 and B.1.1.7 the serum neutralization titers were almost equal [21]. P.1 had the greatest fold reduction in susceptibility when Bamlanivimab and Etesivimab were used together compared to B.1.1.7, B.1.351, P.1, B.1.427, B.1.429 and B.1.526 [22]. Variants with L452R and E484K mutations in spike protein had markedly reduced susceptibility to Bamlanivimab and, for Etesevimab and Casirivimab, the susceptibly might be also reduced [22–24]. Variants harboring receptor binding domain mutations K417 N/T, E484K, and N501Y had high resistance to neutralization [25]. B.1.526, B.1.427, and B.1.429 lineages of SARS-CoV-2 contain the L452R mutation, whereas B.1.525, P.2, P.1, and B.1.351 contain the E484K mutation (some B.1.526 and B.1.1.7 also contain the E484K mutation) [23]. The B.1.351 lineage had greater resistance to neutralization compared to the variant B.1.429 by Moderna and Novavax, which elicited a neutralizing antibody [26]. Although many variants contain the E484Q and L452R separately, recently both of these mutations were found together in India. They were designated as a "double mutant" which is of B.1.617 lineage [27].

While there are many approaches for vaccine development like killed vaccine, live-attenuated vaccine, DNA/RNA vaccines, etc., epitope-based chimeric or subunit vaccines have advantages over them. This is because an ideal multi-epitope vaccine (MEV) can enhance immune response and eventually scale down the risk of re-infection by magnifying the host immunogenicity [28]. Moreover, they are safer as they do not require an entire pathogen. Also, highly promiscuous epitopes can bind multiple alleles simultaneously which can ensure the desired immune response among a heterogeneous human population. Further, usage of multiple epitopes from different virus antigens can expand the spectra of targeted viruses of high mutation frequency [29]. Because they contain multiple MHC restricted epitopes, they can be recognized by TCRs of various T cell subsets. Along with inducing strong cellular and humoral immune responses with the help of the combination of multiple kinds of epitopes, they also can enhance long-lasting immune responses with the help of their adjuvant [30].

Our approach is to design epitope-based chimeric vaccines by screening all existing proteins in SARS-CoV-2 so that the most immunogenic peptides can be used. In this study, the analysis was done using the conserved regions of proteins of different SARS-CoV-2 strains obtained from NCBI virus portal. This ensures the predicted MEVs will target all the SARS-CoV-2 strains including the concerned lineages.

2. Methods

2.1. Retrieval of nucleotide and protein sequences of SARS-CoV-2

Nucleotide and protein sequences of SARS-CoV-2 were retrieved from NCBI virus portal (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/, queried on 22/04/21). The Virus, Nucleotide Completeness and Host fields were set to "Severe acute respiratory syndrome coronavirus 2 (SARS –CoV-2), taxid: 2697049", "complete" and "Homo sapiens (human), taxid: 9606" respectively in the NCBI Virus portal for retrieving the sequences analyzed in the present study. The Collection Date field was set from 1st December 2019 to 22nd April 2021. A meta data (containing accession number, Release Date, Geographical Location, Collection Date) of all the nucleotide sequences of SARS-CoV-2 fulfilling the above-mentioned criteria was also obtained from the same database. Sequences which did not mention the specimen collection month were not included. One percent (1%) of accession numbers of SARS-CoV-2 nucleotide sequences from each month were randomly selected using the sample_n () function of dplyr package in R [31]. Finally, randomly selected nucleotide sequences of SARS-CoV-2 were retrieved. The Pango Lineages of selected sequences were assigned using Pangolin webserver (https://pangolin.cog-uk.io/). The rationale for selecting sequences on the basis of collection date was that the sequences collected would closely reflect the viral strain circulating in that particular month.

In NCBI Virus portal, there are a maximum of 12 SARS-CoV-2 (S, N, E, M, ORF1ab polyprotein, ORF1a polyprotein, ORF3a, ORF6, ORF7a, ORF7*b*, ORF8, ORF10) proteins with respect to a particular SARS-COV-2 nucleotide accession number. As ORF1a polyprotein is a part of ORF1ab, 11 proteins were therefore selected in this study. Finally, the protein sequences of these 11 proteins corresponding to a particular nucleotide sequence were obtained from the virus portal.

2.2. Multiple sequence alignment and phylogenetic tree construction

Multiple sequence alignment (MSA) of all the selected nucleotide sequences of SARS-CoV-2 was performed using online MAFFT platform (https://mafft.cbrc.jp/alignment/server/). Default parameters were used. The phylogenetic tree was constructed using IQ-TREE employing the parameter rich–model called GTG + I + G model, as suggested by Abadi et al. [32].

Protein sequences corresponding to the selected nucleotide sequences of SARS-CoV-2 for a particular protein was also aligned with the online MAFFT platform (https://mafft.cbrc.jp/alignment/server/).

Selection of conserved regions from MSA of protein sequences and prediction of antigenicity The MSA files were visualized using MEGA X (Version 10.1.7) [33] and 100% conserved regions of each of the 11 proteins were manually extracted. Conserved regions shorter

than 9 residues were not included. The density of conservancy was calculated for each protein by using the following formula:

Density of conservancy = (Number of conserved regions of length ≥ 9 residues for a protein) *(Average length of conserved regions for a protein)/(Length of the protein).

The formula for density of conservancy accounted for the length of the protein and the length of the conserved regions in that particular protein. The density of conservancy serves as a measure of conservancy of a protein. The higher the density of conservancy, the higher is the conservancy.

The antigenicity of each of the proteins sequences was determined using VaxiJen server 2.0 (http://www.ddg-pharmfac.net/vaxijen/). which is a web-based server for the alignment-independent prediction of antigenicity [34]. This server predicts each submitted sequence as an antigen or non-antigen along with a probability score. The mean antigenicity score was then calculated by averaging the antigenicity scores of all the protein sequences of a particular protein. The process was repeated for all the 11 SARS-CoV-2 proteins.

2.3. B cell epitope prediction

The conserved amino acid sequences of spike glycoprotein were submitted with a 0.5 cut-off value to find their chance of being antigenic or non-antigenic. Along with this, for transmembrane topology prediction, TMHMM v0.2 server (http://www.cbs.dtu.dk/services/TMHMM/) was used. This server is based on the hidden Markov model and has a high degree of accuracy in predicting protein topology. Each of the conserved sequences was subjected to this server to discriminate intracellular and surface protein regions [35].

Conserved sequences that were predicted as antigen along with fulfilling exomembrane characteristics, were analyzed for linear B cell epitope prediction using BepiPred-2.0 (http://www.cbs.dtu.dk/services/BepiPred/) with the default threshold value [36]. The same analysis was also performed using another web-based tool ABCpred (https://webs.iiitd.edu.in/raghava/abcpred/ABC_submission.html) which has a default threshold value of 0.51 with an epitope length of 16 mer [37]. After obtaining the outputs, peptides that were found to be fully overlapping in both tools were selected to be potential epitopes.

To determine the surface accessibility, hydrophilicity, flexibility, and secondary structure properties of these epitopes, IEDB analysis resource (http://tools.immuneepitope.org/bcell/) was used. Using Emini surface accessibility prediction, surface exposure probabilities of the amino acids were predicted with default threshold 1.0, using surface accessibility scale [38]. Along with this, for the analysis of hydrophilicity of the amino acid residues, Parker Hydrophilicity prediction was used. This tool uses a hydrophilic scale which is dependent on peptide retention times of 20 model synthetic peptides during high-performance liquid chromatography (HPLC) on a reversed-phase column [35]. Karplus' and Schulz's flexibility prediction was used with default parameters for determining flexibility of the residues [39]. The presence of beta-turn, an extremely important factor for an epitope, was predicted using beta-turn prediction algorithm developed by Chou and Fasman [40]. Finally, all these prediction results were analyzed and their antigenicity, allergenicity and toxicity were assessed using VaxiJen v2.0, AllergenFP and ToxinPred respectively. The most common, conserved, non-allergen and non-toxic peptides were considered to be the most potential B cell epitopes.

2.4. T cell MHC class I and class II epitope prediction

Epitopes for Cytotoxic T cell lymphocytes (CTL) were predicted for all the 11 proteins, using the conserved regions of each protein through artificial neural network algorithm-based online server NetCTL 1.2. (http://www.cbs.dtu.dk/services/NetCTL/). Default parameters were used. Epitopes were predicted against all the 12 MHC class I supertypes. Epitopes, identified as MHC ligands, were further selected for subsequent analysis.

The epitopes were subjected to MHC I binding prediction, using the peptide binding to MHC class I molecules tool [41] found in Immune Epitope Database or IEDB (http://tools.iedb.org/mhci/). NetMHCpan EL 4.1 was used as the prediction method as recommended by IEDB using specific MHC Class I alleles (Supplementary File 1, Sheet 1) [42]. Nonamer was selected as the peptide length. A percentile rank with a threshold of 1% for MHC class I binding epitopes was used to filter out peptide-allele combinations with weak binding affinity. The lower the percentile ranks, the higher was the interaction shown between peptide and MHC molecules.

The peptide binding to MHC class II molecules tool of IEDB was used to predict T-helper cell (HTL) epitopes for 11 proteins (http://tools.iedb.org/mhcii/) [41,43,44]. In this case, conserved regions of each protein were used for the prediction. IEDB recommended 2.22 be used as the prediction method using specific MHC class II alleles (Supplementary File 1, Sheet 2) [45]. Peptide length was selected as 15 mer. A percentile rank with a threshold of 10% was used for MHC class II binding epitopes to filter out peptide-allele with weak binding affinity. The selected allele files for HLA class I and II follow the most common specificities in the general population, based on data available from DbMHC and allelefrequencies.net, which represents commonly shared binding specificities (i.e., supertypes) [46]. Peptides with the number of binding alleles ≥5 for MHC class I binding epitopes and ≥8 for MHC class II binding epitopes were selected for further analyses.

The selected CTL and HTL epitopes were analyzed using Vaxijen server 2.0 (http://www.ddg-pharmfac.net/vaxijen/) to find their antigenic properties [34]. The threshold value was set at 0.6. The epitopes that passed the antigenicity test were analyzed for allergenicity using AllergenFP (http://ddg-pharmfac.net/AllergenFP/) and toxicity using ToxinPred (http://crdd.osdd.net/raghava/toxinpred/). Class II epitopes that passed antigenicity, toxicity, and allergenicity tests were further analyzed for Interferon-γ production using IFNepitope (http://crdd.osdd.net/raghava/ifnepitope/), for IL-4 production using IL4pred (http://crdd.osdd.net/raghava/il4pred/) and for IL-10 production using IL-10Pred (http://crdd.osdd.net/raghava/IL-10pred/).

2.5. Epitope conservancy and population coverage of selected epitopes

During epitope prediction, fully conserved regions from the proteins were used. Moreover, as COVID-19 caused by SARS-CoV-2 is a pandemic, so the constructed MEV should be effective in different populations around the globe to combat this crisis. Thus, population coverage is a very important area to be taken into consideration for the predicted epitopes, as we can observe their probable effective utility for the potential vaccine development in different populations. Population coverage was performed for whole world and specially for the countries mostly affected by death and infection (as mentioned in John Hopkins Covid tracker) by SARS-CoV-2 using the IEDB population coverage analysis tool (http://tools.iedb.org/population/) [47].

2.6. Molecular interaction analysis of selected CTL and HTL epitopes with HLA alleles

For docking of the CTL and HTL epitopes with representative HLA alleles, firstly the three-dimensional structure of the epitopes was predicted using PEP-FOLD3 (https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/) server [48]. As the representative HLA alleles, complex structures of HLA-A*02:01 (PDB ID: 1DUZ) and DRB1*01:01 (PDB ID: 1AQD) were retrieved from Protein Data Bank which served as MHC class I and II receptors respectively. The peptides which were already bounded to the receptor PDB structures ("LLFGYPVYV" with PDB ID: IDUZ and "VGSDWRFLRGYHQYA" with PDB ID: 1AQD) were considered as control peptides. These control peptides were removed using PyMOL and energy minimization was carried out. Afterward, the selected CTL and HTL epitopes were docked with the processed MHC

class I and II receptors respectively. For docking, ClusPro 2.0 protein-protein docking server (https://cluspro.bu.edu/home.php) was used [49]. This server performs blind docking by sampling billions of conformations using CAPRI experiment, determines the most likely models based on RMSD clustering, and refines the selected docked structures using energy minimization [50]. The best receptor-epitope complexes were selected based upon the lowest interaction energy and the visual similarity with the control peptide-receptor complex. Finally, the molecular interactions of the complexes were evaluated using PDBsum Protein-Protein interaction (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html).

2.7. Construction of multi-epitope vaccine

The pre-selected B and T cell epitopes were used for the multi-epitope vaccine (MEV) construction by adapting the previously-described protocols [51]. Two linkers GPGPG, and AAY were used to link the epitopes together. The GPGPG linker was used to connect the B cell epitopes and MHC Class II epitope, and the AAY linker was used to connect the MHC Class I epitopes. These linkers helped to protect their individualistic immunogenic properties after their inter-interaction compatibility authentication. Moreover, at the N terminal, Beta-defensin 3 (Q5U7J2_HUMAN) (https://www.uniprot.org/uniprot/Q5U7J2#sequences) was linked using EAAAK linker as an adjuvant for boosting of the immunogenicity of the vaccine. Finally, to enable protein purification and identification, 6xHis tag was added at the C-terminal end.

2.8. Evaluation of physicochemical properties of the vaccines

The vaccine that was finally developed was subjected to antigenicity and allergenicity prediction using VaxiJen v2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) and AllergenFP V.1.0 (http://ddg-pharmfac.net/AllergenFP/), respectively [34,52]. Also, the construct sequences were submitted to Expasy ProtParam tool (https://web.expasy.org/protparam/) to calculate different parameters such as the number of amino acids, molecular weight, theoretical iso-electric point (pI), half-life, instability index, aliphatic index, GRAVY score, etc [53]. Based on the physicochemical properties of the construct, the quality of the vaccine was assessed.

2.9. Secondary structure prediction

Protein secondary structure plays an important role in protein folding properties and its physicochemical characteristics [54]. Thus, for the prediction of the constructed MEV secondary structure, and to assess its secondary structural properties, RaptorX property (http://raptorx.uchicago.edu/StructurePropertyPred/predict/) was used. RaptorX property performs template free secondary structure prediction by using a machine-learning model called DeepCNF (Deep Convolutional Neural Fields). Along with the secondary structure of the MEV, solvent accessibility and disorder regions (DISO) were also speculated, based on the output of 66% and 89% accuracy respectively [55].

2.10. Tertiary structure prediction, molecular refinement, and structure validation of MEV

The three-dimensional structure of the constructed MEV was predicted using the 3Dpro tool (http://scratch.proteomics.ics.uci.edu/) [56]. Because of the lack of any suitable template for homology modeling, this ab initio approach was used. Thus, for the refinement of the predicted protein structure, the GalaxyRefine web server (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE) was employed [57,58]. This server follows a refinement method that rebuilds the side-chains along with side-chain repacking. It also causes structural relaxation using molecular dynamics simulation, improving the structural quality significantly.

Following the refinement, the models were subjected to quality checking or assessment using the ERRAT (https://servicesn.mbi.ucla.edu/ERRAT/) [59], ProSA (https://prosa.services.came.sbg.ac.at/prosa.php) [60], and the Ramachandran plot analysis using SWISS-MODEL Structure Assessment platform (https://swissmodel.expasy.org/assess) [61]. Based on these results, the qualities of the refined models were validated and the most suitable model was selected for further steps.

2.11. Conformational B cell epitopes prediction of the constructed MEV

To predict conformational B cell epitopes on the MEV construct DiscoTope-2.0 (http://www.cbs.dtu.dk/services/DiscoTope/), ElliPro (http://tools.iedb.org/ellipro/), and SEPPA v3.0 (http://www.badd-cao.net/seppa3/index.html) were used. The PDB structure of the constructed MEV was submitted to each server with default parameters (DiscoTope threshold = −3.700, ElliPro threshold = 0.5, and SEPPA threshold = 0.064). The residues predicted as conformational B cell epitope by all of the three tools were selected and visualized using PyMOL.

2.12. Molecular docking and binding affinity analysis

To assess the binding affinity of selected potential vaccine candidates with TLR2 (PDB ID: 2Z7X), TLR4 (PDB ID: 4G8A), ACE2 receptor (PDB ID: 3SCI), and B Cell Receptor (PDB ID: 3KG5), molecular docking was performed using the ClusPro 2.0 protein-protein docking server (https://cluspro.bu.edu/home.php) [49]. From the outcome, the vaccine receptor complex that appeared to be docked with maximum residue interaction and lowest interaction energy was selected, because the lower energy score corresponds to the better binding affinity [62]. Further, the complexes were analyzed using PDBsum Protein-Protein interaction (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html) to observe the interactions between MEV and the receptors [63].

2.13. Molecular simulation study of the vaccine construct and TLR4

The binding stability and flexibility of the vaccine and TLR4 complex was analyzed using iMODS server (http://imods.Chaconlab.org/) which performs Normal Mode Analysis (NMA) in internal (dihedral) coordinates using an elastic network model (ENM) [64]. The server provides B-factor and deformability plots, covariance map, mode variance plot, eigenvalues and elastic network.

2.14. Immune simulation

The immunogenicity and immune response profile of the selected vaccine were simulated through the C-ImmSim server (https://150.146.2.1/C-IMMSIM/?page=0). This server uses a position-specific scoring matrix (PSSM) and machine learning techniques for the prediction of the immune response. The simulation volume was set to 10, random seed 12345, number of steps 1095, adjuvant 100, simulation time steps 1, 84, and 168 along with homozygous host haplotypes HLA-A0321, HLA-A0110, HLA-B5801, HLA-B3501, HLA-DRB1-0801, and HLA-DRB1_1501 [65].

2.15. Codon optimization and in silico cloning

Because of the codon bias, codon adaptation is a vital step in reverse vaccinology which can promote recombinant protein expression significantly. For this purpose, the Java codon adaptation index (JCAT) (http://www.jcat.de/) was used to get the optimized DNA sequence of the final vaccine to construct along with the codon adaptation index (CAI) value and GC content. Here, the selected target organism was

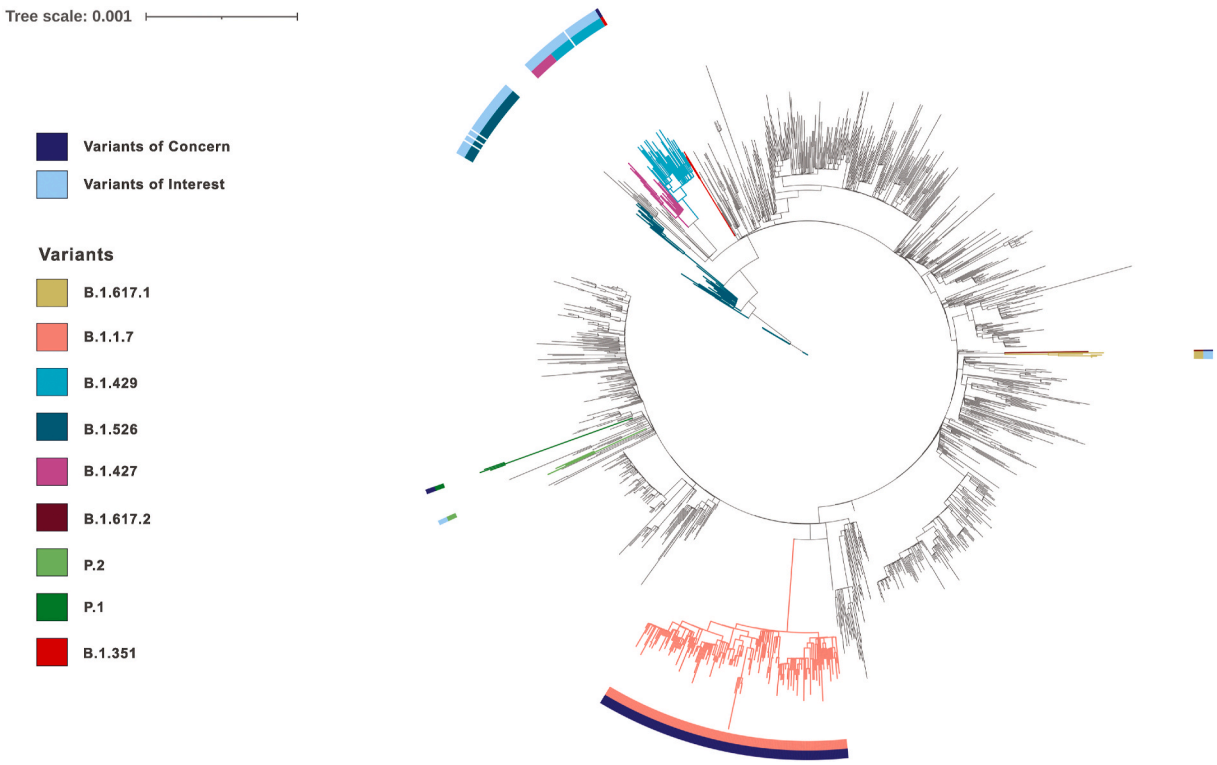

Fig. 1. Phylogenetic tree of the nucleotide sequences of different variants of SARS-CoV-2 to represent the relatedness as well as diversity of different variants of coronavirus. A total of 1549 nucleotide sequences were applied to construct the tree. The colored strip charts represent different variants. The color codes illustration of each variant have been demonstrated at the left side of the figure. The branches of the tree are also colored according to the variants. Deep blue represents the variants of concern while sky blue represents the variants of interest. The nucleotide sequences of variants B.1.351, B.1.427 and B.1.429 are more closely related as they had clustered together in the tree. The same is true for P.1 and P.2; B.1.617.1 and B.1.617.2. Tree scale is 0.01.

Escherichia coli (strain K12), and *EcoR*I and *BamH*I as the restriction enzyme cleavage sites [66]. SnapGene software (version 5.2.3) (https://www.snapgene.com/) was used to integrate the adapted DNA sequence to pET-28a (+) vector, between the *EcoR*I and *BamH*I restriction sites. This vector enables enhanced protein recovery and purification due to its N-terminally 6 × His-tagged proteins.

3. Results

3.1. Retrieval of nucleotide and protein sequences of SARS-CoV-2

The meta data list retrieved consisted of 154805 sequences from 1st December 2019 to 22nd April 2021 (based on collection date, accessed on 22/04/21). A total of 501 sequences were excluded because the specimen collection month was not mentioned. Thus, the final list consisted of 154,304 sequences. A total of 1283 accession numbers of SARS-CoV-2 sequences were obtained from the final list after randomly selecting 1% of the sequences from each month (except March 2021). In the case of March 2021, 1% of sequences were randomly selected differently to consider the new B.1.617.1 and B.1.617.2 variants. Initially, accession numbers of five B.1.617.1 variants and the only B.1.617.2 variants were included; then, accession numbers for 260 SARS-CoV-2 sequences were retrieved randomly from March 2021. A total of 1549 (1283 + 266) SARS-CoV-2 nucleotide sequences were retrieved from NCBI virus portal using the selected accession numbers. The number of sequences taken from each month are given in Supplementary File 2 (Sheet 1). Most sequences are taken from March 2021 (266).

The number of protein sequences retrieved for a particular protein corresponding to 1549 SARS-CoV-2 nucleotide sequences are given in Supplementary File 2 (Sheet 2). There should be 1549 sequences for each protein corresponding to 1549 SARS-CoV-2 sequences. However, some SARS-CoV-2 nucleotide sequences do not have protein sequences as they were part of genome assembly projects (533).The accession numbers of these sequences (without any proteins) are given in Supplementary File 2 (Sheet 3). Of the 1016 (1549–533) SARS-CoV-2 nucleotide sequences, 158 sequences did not have protein sequences for Orf 8 (Supplementary File 2, Sheet 4). MW553306 and MW629363 did not have sequences for Orf 3a and Orf 7a, respectively.

The Pango lineage distribution of all the 1549 sequences are given in Supplementary File 2 (Sheet 5). The number of sequences taken from each Pango Lineage is given in Supplementary File 2 (Sheet 6). The variants B.1.1.7, B.1.351, P.1, B.1.427, B.1.429, B.1.526 and P.2 had 162, 2, 3, 17, 38, 50 and 3 SARS-CoV-2 nucleotide sequences, respectively.

3.2. Multiple sequence alignment and phylogenetic tree construction

IQ-TREE constructs a phylogenetic tree by maximum likelihood. GTG-I-G was used as the substitution model, as previously described [32]. Distinct clusters of B.1.1.7, B.1.351, P.1, B.1.427, B.1.429, B.1.526, B.1.617.1, B.1.617.2and P.2 variants were observed in the constructed phylogenetic tree. The percentages of sequences harboring the mutations D614G, N501Y, E484K, L452R, and P681H in S protein are shown in Fig. 2A.

The insertions and deletions found in different proteins after inspecting the MSA of 11 proteins are given in Supplementary File 2 (Sheet 7). Deletion of three amino acids (SGF) was found in Orf 1 ab protein of 215 SARS-CoV-2 sequences. Most of them were of B.1.1.7 variants (155) followed by B.1.526 (50). Both B.1.351 and all the three P.1 variants had this deletion too (Supplementary File 2, Sheet 7).

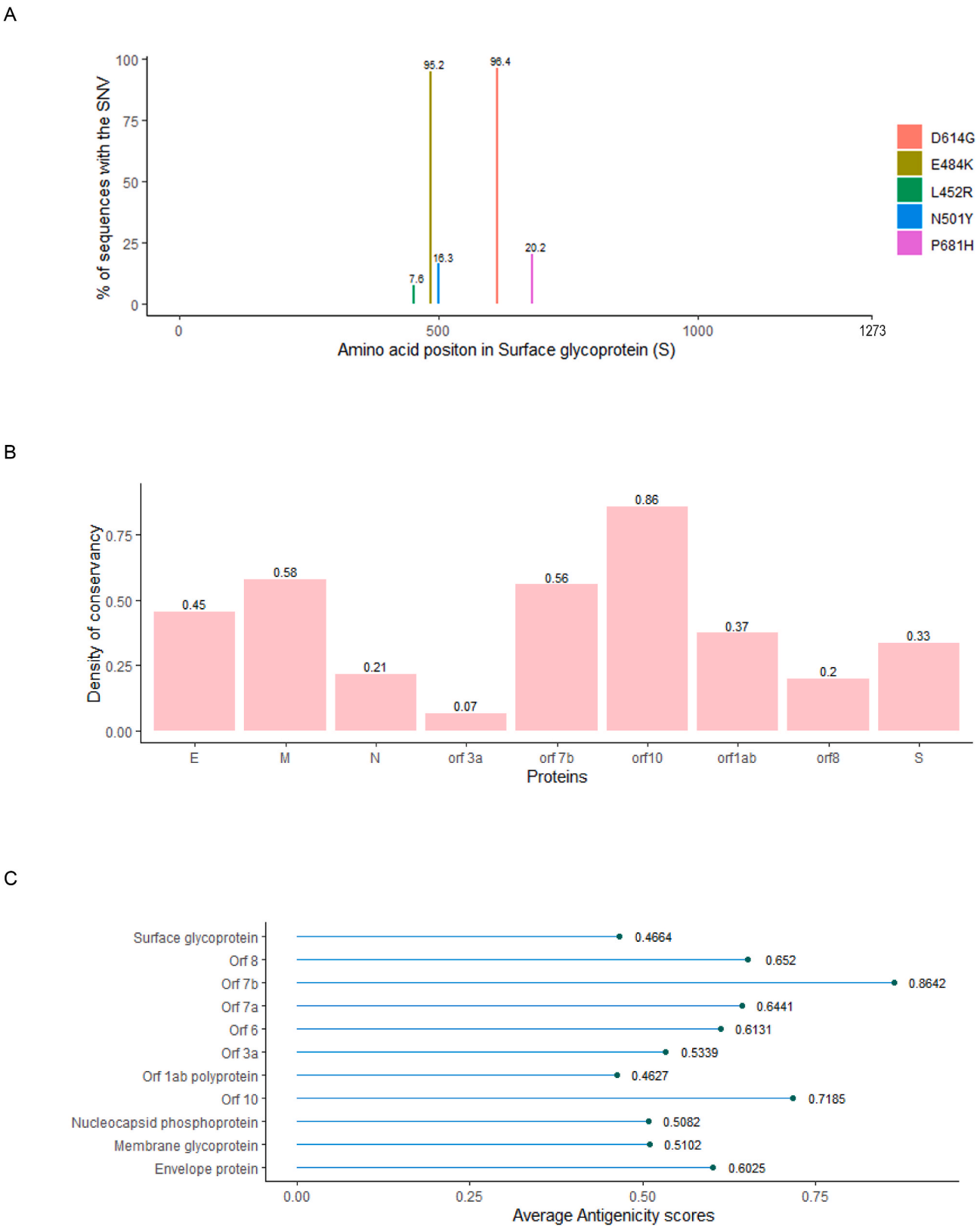

Fig. 2. A) A bar plot showing the % of sequences containing a particular mutation. The y–axis denotes the % of sequences with the single nucleotide variation (SNV) and the x-axis showing the position of the mutation in the Surface glycoprotein. The bars are colored according to the mutation type and the color code is given on the right of figure. B) Bar plot showing the density of conservancy of each protein constructed using information of the density of conservancy (y-axis) and the proteins of coronavirus (x-axis) used in this study. The values on the top of each bar represents the value of density of conservancy. C) A lollipop plot showing the average antigenicity scores of each protein. The average antigenicity scores have been shown at the right sides of the lollipops.

Table 1
Selected potential B cell, MHC class I and class II epitopes.

Epitope	Overlapped Epitope	VaxiJen 2.0	AllergenFP v.1.0	ToxinPred	Protein	Type	Identifier
RVVFNGVSF		0.7252	Non Allergen	Non Toxic	ORF1ab	MHC Class I	**E1**
YPSLETIQI		0.8151	Non Allergen	Non Toxic	ORF1ab	MHC Class I	**E2**
YTEISFMLW		1.2159	Non Allergen	Non Toxic	ORF1ab	MHC Class I	**E3**
TVKPGNFNK		1.3778	Non Allergen	Non Toxic	ORF1ab	MHC Class I	**E4**
VVSTGYHFR		1.4741	Non Allergen	Non Toxic	ORF1ab	MHC Class I	**E5**
EIDPKLDNY		1.6159	Non Allergen	Non Toxic	ORF1ab	MHC Class I	**E6**
GVVFLHVTY		1.4104	Non Allergen	Non Toxic	Surface Glycoprotein	MHC Class I	**E7**
LVIGAVILR		1.1027	Non Allergen	Non Toxic	Membrane Glycoprotein	MHC Class I	**E8**
DGYFKIYSKHTPINL	DGYFKIYSKHTPINLV	0.7526	Non Allergen	Non Toxic	Surface Glycoprotein	MHC Class II	E9
GYFKIYSKHTPINLV		0.9278	Non Allergen	Non Toxic			
FNMVYMPASWVMRIM	AYFNMVYMPASWVMRIM	0.6579	Non Allergen	Non Toxic	ORF1ab	MHC Class II	E10
AYFNMVYMPASWVMR		0.8153	Non Allergen	Non Toxic			
FLLVTLAILTALRLC	VVFLLVTLAILTALRLC	0.6311	Non Allergen	Non Toxic	Envelope protein	MHC Class II	E11
VVFLLVTLAILTALR		0.7559	Non Allergen	Non Toxic			
LVSTQEFRYMNSQGL		0.6617	Non Allergen	Non Toxic	ORF1ab	MHC Class II	E12
KVKYLYFIKGLNNLN		0.9061	Non Allergen	Non Toxic	ORF1ab	MHC Class II	**E13**
IADYNYKLP		1.1012	Non Allergen	Non Toxic	Surface Glycoprotein	B cell	**E14**
WFVTQRNFY		0.7376	Non Allergen	Non Toxic	Surface Glycoprotein	B cell	**E15**
NSYECDIPIGAGIC		0.6147	Non Allergen	Non Toxic	Surface Glycoprotein	B cell	**E16**

*Bolded epitopes are used to construct multi-epitope vaccine.

3.3. Conserved region selection from MSA of protein sequences and prediction of antigenicity

The conserved regions extracted from the protein sequences of a particular protein are presented in Supplementary File 2 (Sheet 8). Fig. 2B represents the density of conservancy for each protein. ORF10 of SARS-CoV-2 was the most conserved among the SARS-CoV-2 proteins, with a density of conservancy of 0.86. Following this, antigenicity of the targeted proteins was observed. The average antigenicity score of ORF7b was found to be the highest compared to other proteins. The average antigenicity scores of different proteins have been demonstrated in Fig. 2C.

3.4. B cell epitopes

Among the total of 32 conserved sequences of spike glycoprotein, 16 sequences satisfied the VaxiJen v2.0 threshold score and were predicted to contain antigenicity properties. All these sequences belonged to the outside region of the protein according to the TMHMM v0.2 output. BepiPred-2.0 and ABCpred predicted 34 and 134 peptides, respectively from the 16 sequences from which 5 overlapping peptides were identified.

On the basis of the IEDB analysis resource prediction of surface accessibility, hydrophilicity, flexibility, secondary structure properties of these peptides, 3 of them (“^{427}IADYNYKLP^{435}”, “^{667}NSYECDIPIGAGIC^{680}”, “^{1111}WFVTQRNFY^{1119}”) appeared to possess all these desired properties by crossing the prediction threshold values (Supplementary File 4). Thus, these epitopes were surface accessible, hydrophilic, flexible and contained a beta turn in the secondary structure. After further assessment of allergenicity, toxicity and antigenicity, these 3 epitopes were selected as the potential B cell epitopes as these were predicted to be non-allergen, non-toxic and antigenic (Table 1).

3.5. Selection of potent CTL and HTL epitopes for vaccine construct

A total of 363 CTL epitopes (ORF 1 ab = 280, S = 40, M = 16, E = 9, N = 9, ORF7b = 7 and ORF10 = 2) were predicted from NetCTL 1.2 (http://www.cbs.dtu.dk/services/NetCTL/). Of them, 289 epitopes (ORF 1 ab = 227,S = 32,M = 13,E = 9,N = 7 and ORF 7b = 1) had a percentile rank of ≤1%, as predicted with MHC class I molecules tool of IEDB (http://tools.iedb.org/mhci/). Of these 289 epitopes, 65 epitopes (ORF 1 ab = 49, S = 11, M = 2, E = 2 and N = 1) had bound with ≥5 MHC Class I alleles.

Out of 65, 8 CTL epitopes (ORF 1 ab = 6, S = 1 and M = 1), satisfied the antigenicity, allergenicity and toxicity tests. Finally, these epitopes were selected for vaccine construction. Epitope “^{1891}EIDPKLDNY^{1899}” of ORF1ab had the highest antigenic score of 1.6159 and epitope “^{3164}RVVFNGVSF^{3172}” bound with the highest number of MHC Class I alleles (9) (Supplementary File 3).

A total of 323 HTL epitopes (ORF 1 ab = 252, S = 43, M = 14, E = 6, N = 6, ORF7b = 1 and ORF 10 = 1) were predicted by MHC class II molecules tool of IEDB (http://tools.iedb.org/mhci/) having percentile

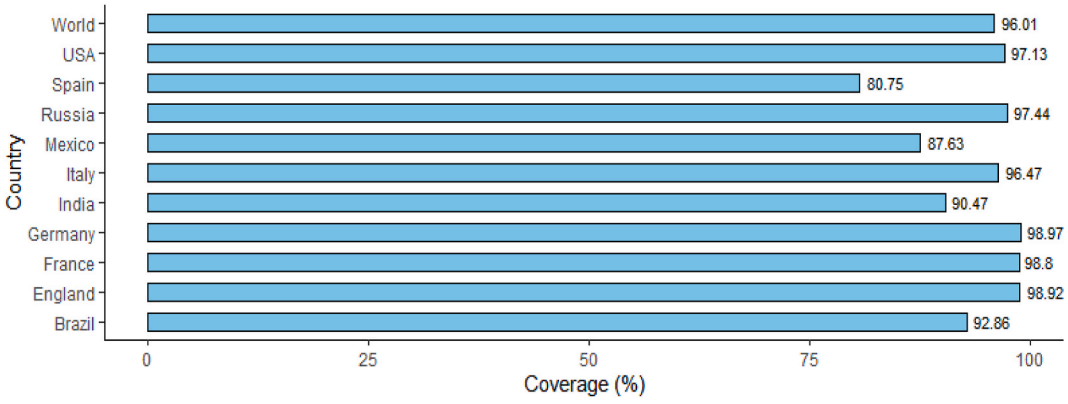

Fig. 3. Population coverage of the selected epitopes for vaccine construct in different countries world-wide. The y-axis shows the countries and the x axis shows the % coverage. The percentage of population coverage is shown right to the bars.

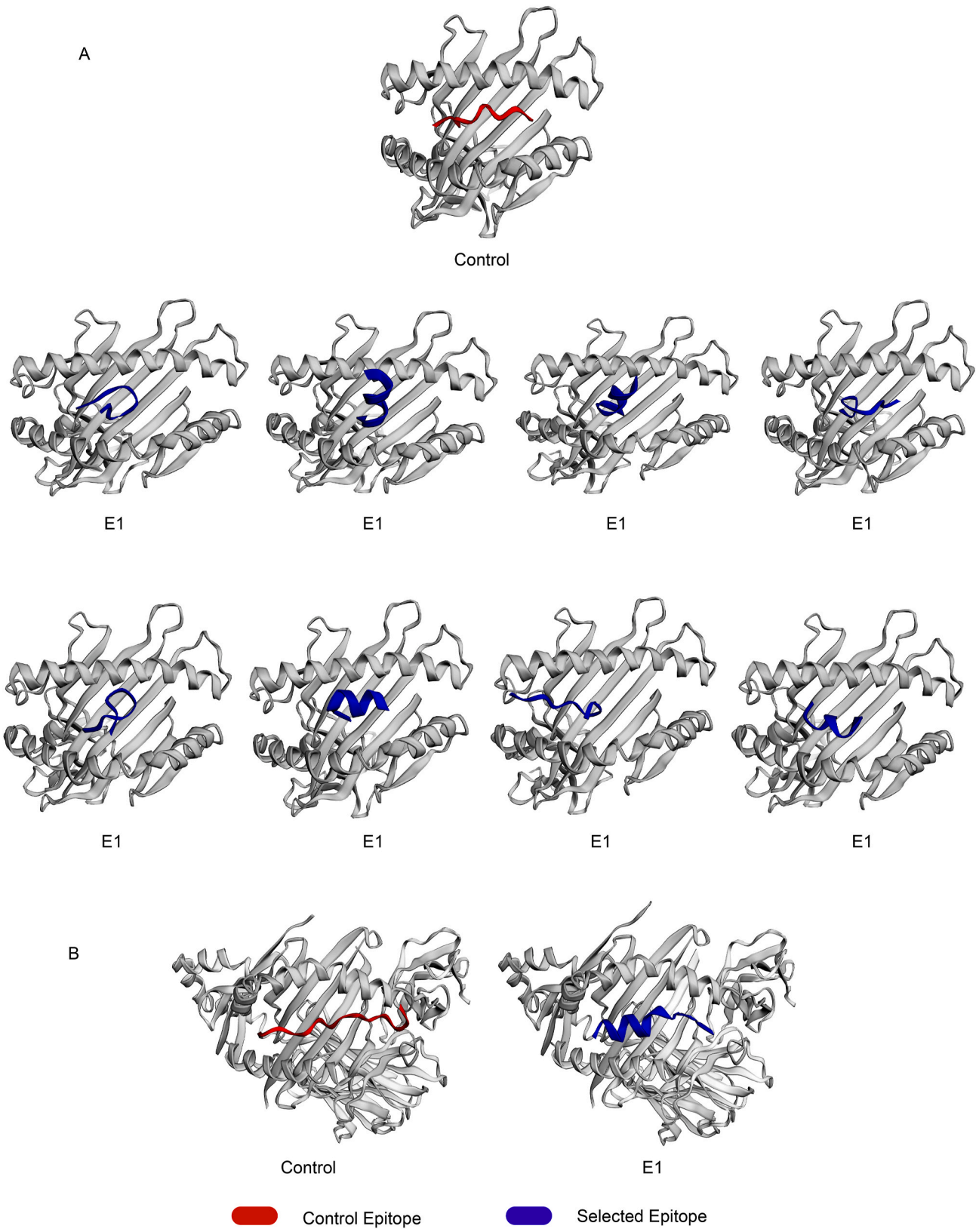

Fig. 4. The view of- A) HLA-A*02:01 presenting control and CTL epitopes (E1-E8); B) DRB1*01:01 presenting control and HTL epitope (E13). HLA alleles, control peptides and selected viral epitopes are shown in grey, red and blue respectively.

Table 2
The interactions between the selected epitopes and MHC receptors of the docked complexes.

Epitope	Receptor	Lowest interaction energy	No of H bonds	No of Non-bonded contacts
Control	MHC I	−1163.6	10	155
E1	MHC I	−732.1	5	104
E2	MHC I	−609.8	5	63
E3	MHC I	−736.1	5	103
E4	MHC I	−462.6	11	107
E5	MHC I	−756.7	8	143
E6	MHC I	−532.1	4	89
E7	MHC I	−855.5	1	150
E8	MHC I	−763.1	5	118
Control	MHC II	−1629.8	14	199
E13	MHC II	−895	10	261

rank ≤ 10%. Of these 323 epitopes, 42 (ORF 1 ab = 33, S = 4, E = 4 and ORF7b = 1) epitopes bound with ≥8 MHC Class II alleles.

Out of the 42 epitopes, 8 HTL epitopes (ORF 1 ab = 4, S = 2 and E = 2) passed the antigenicity, allergenicity, and toxicity tests. But only epitope, "^{424}KVKYLYFIKGLNNLN^{438}" of Orf 1 ab was found to be IFN-**γ** producer, IL-10 and IL-4 non-producers. "^{424}KVKYLYFIKGLNNLN^{438}" had an antigenicity score of 0.9061 and bound with 17 HLA Class II alleles. Therefore, it was selected as the final HTL epitope for vaccine construction (Table 1) (Supplementary File 3).

3.6. Population coverage of selected epitopes

The population coverage of the selected epitopes for different countries and the whole world is shown in Fig. 3. The selected epitopes had the highest population coverage of 98.97% in Germany. The result for England was nearly equal (98.92%). The epitopes showed a high population coverage in all the countries which are greatly affected by SARS-CoV-2. This shows that our vaccine constructed from these selected epitopes will be able to efficiently combat this pandemic.

3.7. Interaction of selected epitopes with HLA alleles

As the outcome of molecular docking simulations of the CTL and HTL epitopes with MHC class I (HLA-A*02:01) and MHC Class II (DRB1*01:01) receptors, it appeared that the epitopes interacted to the receptor grooves in a similar manner to the control peptides (Fig. 4). For the CTL and MHC class I receptor interaction, the control peptide formed 10H bonds and 155 non-bonded contacts with the receptor with the lowest interaction energy score −1163.6. Whereas, the selected CTLs interacted with the receptors with similar energy scores. Among them, E4 showed the highest no of H bonds and E7 showed the highest non-bonded interactions with the receptor which were closest to the control complex (Table 2). For the HTL-MHC class II complex, the non-bonded contacts were much higher for E13-MHC class II complex but the H bond number was lesser than the control complex (Table 2). Thus, the selected epitopes showed strong potential ability to complex with MHC receptors in comparison with the control peptides.

3.8. Multi epitope vaccine construction

A total of 12 epitopes- 3 B cell, 1 MHC Class II, and 8 MHC Class I epitopes were used to design the MEV. The predicted epitope sequences were joined using GPGPG and AAY linkers. The adjuvant Beta-defensin was attached to the N terminal with the help of EAAAK linker. Also, to facilitate protein purification, a 6xHis tag was added to the C terminal. Thus, the final construct generated was as follows (Bold letters indicate the linkers): "**EAAAK**GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCST RGRKCCRRKK**EAAAK**IADYNYKLP**GPGPG**WFVTQRNFY**GPGPG**NSYEC DIPIGAGIC**GPGPG**KVKYLYFIKGLNNLN**AAY**RVVFNGVSF**AAY**YPSLETI-QIA**AYY**TEISFMLW**AAY**TVKPGNFNK**AAY**VVSTGYHFR**AAY**EIDPKLDN YA**AY**GVVFLHVTY**AAY**LVIGAVILRHHHHHH" (Fig. 5A).

3.9. Evaluation of physicochemical parameters

The Expasy ProtParam tool, VaxiJen v2.0, and AllergenFP v.1.0 predicted the physicochemical properties of the designed MEV construct. The vaccine construct was composed of 219 amino acids and possessed 24.326 kDa molecular weight. As the molecular weight is lower than 110 kDa, it can be considered as a good vaccine candidate [67]. The vaccine was predicted to be slightly basic in nature because the estimated pI was 9.4. The half-life was predicted to be 1 h in mammalian reticulocytes in vitro. The instability index of the vaccine was 23.69, which is less than 40 and thus indicating the construct to be stable. The aliphatic index of the construct was 79.82, which indicates the construct was thermostable [68]. The hydrophilic nature of the MEV construct was proved by its GRAVY score −0.124, which is a negative value [69]. Moreover, the VaxiJen score of the construct was 0.6199, indicating a probable antigen; also, it was predicted to be non-allergen.

3.10. Secondary structure assessment

RaptorX property was used to evaluate the secondary structure of the MEV construct. The secondary structure of the vaccine construct (Fig. 5B) predicts the vaccine to be composed of 16% alpha helix, 36% beta strand and 47% coil. From the solvent accessibility prediction by RaptorX property, it was observed that 41% of the vaccine residues were in an exposed region, 22% in moderately exposed region and 36% in a buried region (Fig. 5C). It was also predicted that only 12 residues (5%) were in a disordered region (Fig. 5D).

3.11. Tertiary structure evaluation

Using 3Dpro, the preliminary 3D structure of the MEV was predicted. Here, due to the lack of any proper PDB template this *de novo* method was used. Thus, to improve the quality, the predicted structure was refined using GalaxyRefine server. This server refined the structure and five refined models were obtained. These models were further evaluated based on their ERRAT score, Ramachandran plot, and ProSA Z-score. On the basis of these results, the best-resulting model was selected for further analyses (Fig. 6A). Here, the final selected structure was found to have a ProSA Z-score of −2.97. This value confirms its near-native quality as it is placed close to the experimentally resolved structures of similar sizes (Fig. 6B). From the Ramachandran plot analysis, it was observed that 96.2% residues of the finally selected refined model were in the favorable and allowed region (Fig. 6C), whereas, before refinement, the favorable region residues were 89.4%. Finally, analysis of ERRAT server predicted the quality score of 79.0576, which indicated the presence of very few inaccurate regions in the MEV 3D structure.

3.12. Discontinuous B-cell epitopes on MEV

Out of 219 total residues, DiscoTope-2.0 predicted 20 residues as conformational B cell epitope. Whereas, ElliPro and SEPPA v3.0 predicted 97 and 106 residues to possess discontinuous epitope quality. Among them, "^{17}R", "^{61}YKLP^{64}", "^{81}GPGNSYECD^{89}", "^{160}PG^{161}", residues were common in all the three predictions (Fig. 6D). The ElliPro score of these residues ranged between 0.584 and 0.995.

3.13. Interaction of vaccines with potential receptors

To evaluate the interaction between the MEV construct and TLR2, TLR4, ACE2 receptors, and B Cell Receptor (BCR), blind molecular docking was performed using ClusPro 2.0 server. Among all the generated docking models, the ones with the higher cluster size and lower interaction energy scores were considered as the best-docked complex. Because of these properties, it could be estimated that the MEV construct

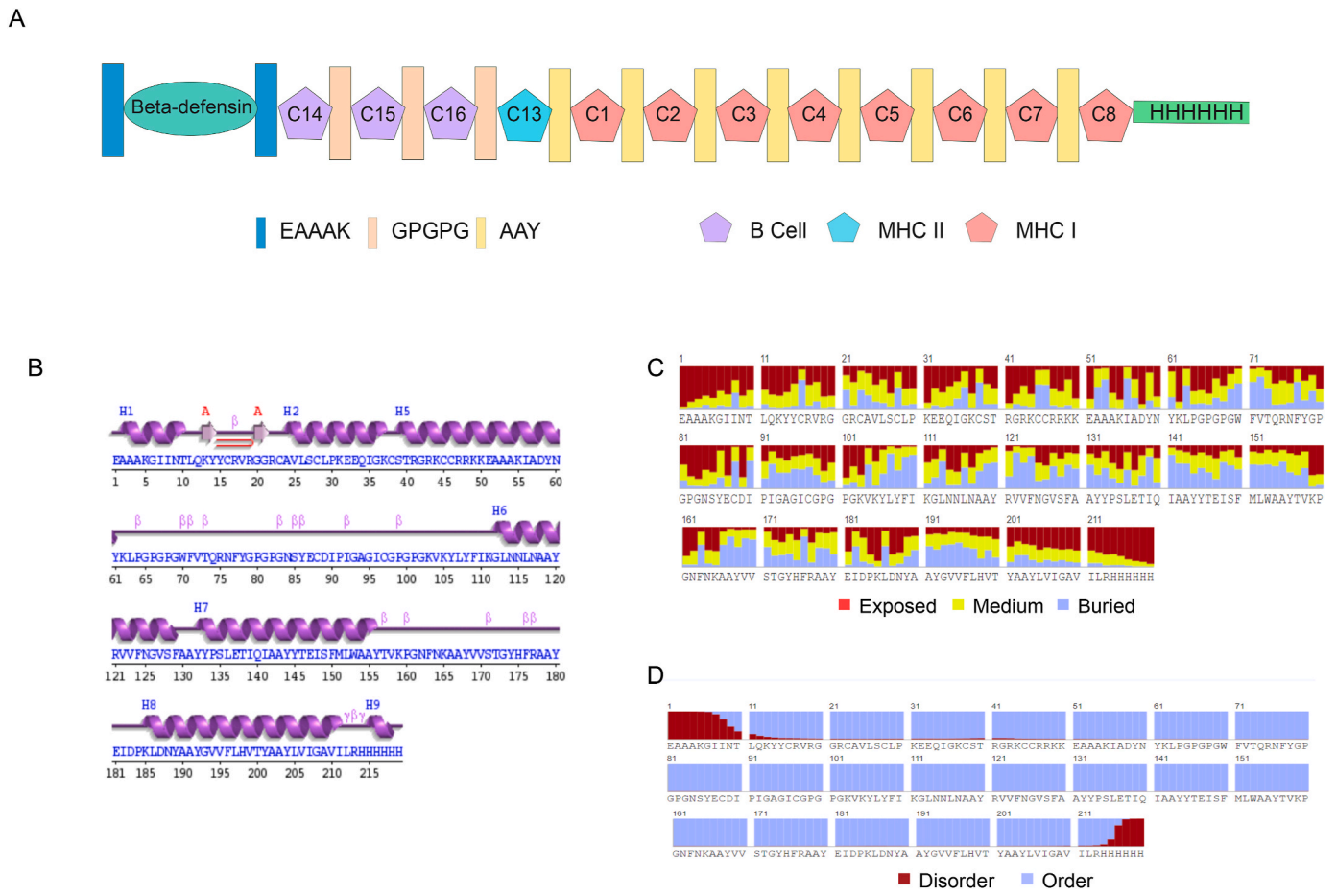

Fig. 5. A) Combination of linkers, adjuvant, and epitopes and their assembly for multi epitope vaccine (MEV) construction, B) Predicted secondary structure of the MEV construct along with amino-acid sequence and positions, C) Solvent accessibility prediction result indicating the exposed, medium, and buried regions of the MEV protein, D) Disordered region prediction result indicating the disordered and ordered regions of the MEV protein.

in the complexes occupies the receptor properly along with satisfactory binding affinity. Here, the cluster size of the selected MEV complexes with TLR2, TLR4, ACE2 receptor, and BCR are 46, 36, 106, and 114; and the lowest interaction energy score was −1004.1, −1256.9, −1411.8, and −1162.4 respectively (Supplementary File 5).

Following the selection of the best-docked complexes, molecular interactions that formed between the MEV and the receptors were visualized with the help of PDBsum. For each complex, the number of interacting interface residues and the quantification of the interface area were predicted for the MEV and receptor. Also, the number of salt bridges, hydrogen bonds, and non-bonded contacts formed between them were predicted. According to the results, the MEV formed a number of different bonds with each receptor (Table 2). However, unlike the others, for MEV-BCR it was observed that the MEV forms bond with both chains of the BCR homodimer (Fig. 7 and Table 3).

3.14. Molecular simulation study of the vaccine construct and TLR4

The affine –model based arrow is mainly directed towards function-based motions [64]. The affine–model based arrows showed the vaccine construct and TLR4 were directed towards each other. This inferred functional role and also indicated strong binding (Fig. 8A). Deformability depicts the flexibility of the protein whereas B-factor is a measure of the mobility of the protein [70]. The relative amplitude of the displacements of atoms about an equilibrium position is indicated by B-factors [64]. The peaks in the deformability curve demonstrated flexible regions. There were few such peaks in vaccine –TLR 4 complex which illustrates stable binding (Fig. 8B). Few fluctuations of atomic displacements were observed for the vaccine–TLR4 complex (Fig. 8C). The eigenvalue was $7.077913e^{-06}$ (Fig. 8D), and there is an inverse relationship between normal mode variance and the eigenvalues. The red bars indicate variance of individual modes while the green bars indicate cumulative variance (Fig. 8E). In vaccine-TLR4 complex, approximately 80% of the variance was justified by the first 8 normal modes (Fig. 8E). The coupling between pairs of residues is illustrated by the covariance map where red, white and blue color corresponds to correlated, uncorrelated and anti-correlated motions (Fig. 8F). The elastic network shows the pairs of atoms connected by a spring. The darker the grey color, the stiffer the spring (i.e. more rigid positions in the complex) (Fig. 8G).

3.15. Immune simulation through in silico simulation

The immune simulation was performed in an agent-based representation of immune cells. The secondary and tertiary immune responses were better than the primary response as they resulted in reduced antigen count per mL. The primary response resulted in a slight increase in IgM antibody which is normal as IgM are the first antibody types to be produced. The antibody levels (IgG1 + IgG2, IgM, IgG + IgM) increased in secondary and tertiary responses which had led to decreased antigen level (Fig. 9A). There was a marked increase in the IgM + IgG level in the secondary response compared to the primary response. The IgM + IgG further increased in the tertiary response. There was a significant increase in memory B cells during secondary response which depicts active B –cell proliferation. The memory B-cell rose further during tertiary response (Fig. 9B). The elevation of different B-cell types indicated class

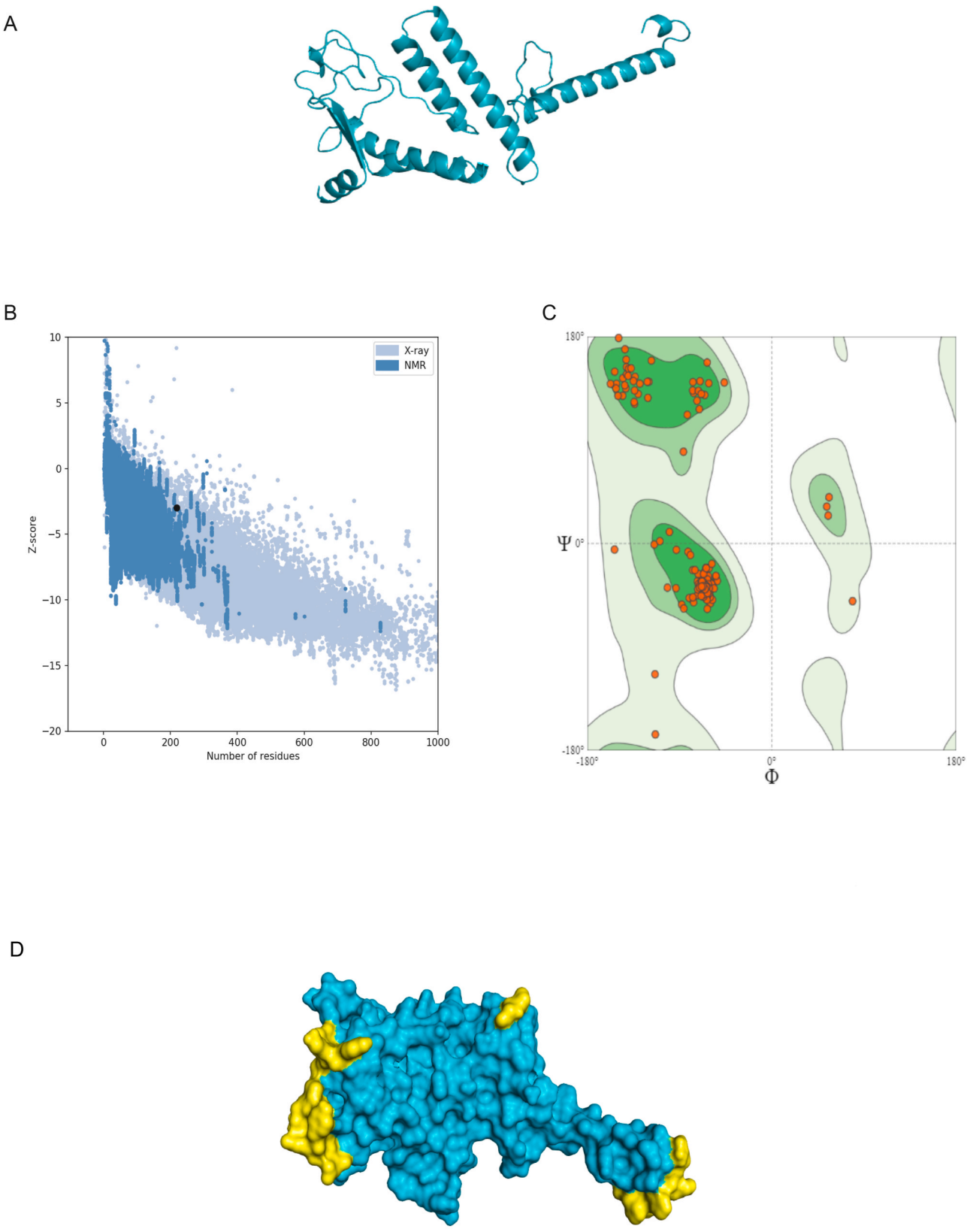

Fig. 6. A) The tertiary structure of the MEV construct finally selected in the study, B) Validation of the three dimensional structure of MEV using ProSA. The Z-score of the refined model is −2.97 which is lying among the score range, C) Ramachandran plot analysis showing 96.2% MEV residues in favorable allowed region. D) The surface view of constructed MEV structure. The yellow indicated regions represent the discontinuous B cell epitopes.

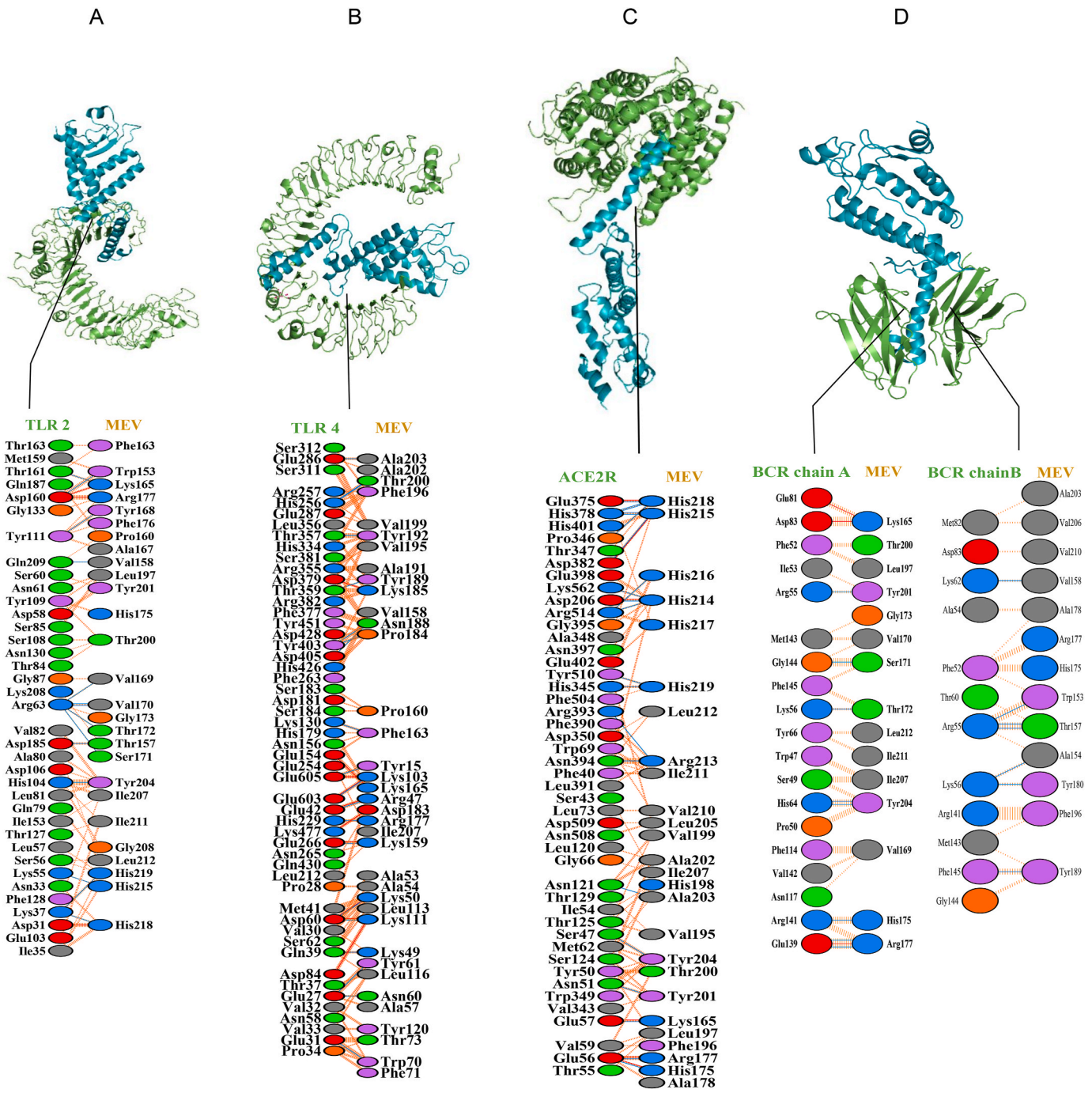

Fig. 7. The docked complexes of MEV-TLR2 (A), MEV-TLR4 (B), MEV-ACE2 receptor (C), and MEV-BCR (D) are presented (Blue = MEV, Green = Receptor) along with the diagram of interactions between MEV and receptor chains. Interacting chains are joined by colored lines, each representing a different type of interaction (Red = salt bridges, blue = H-bonds, striped line = non-bonded contacts). Here, the number of salt bridge and H-bond lines indicates the number of potential hydrogen bonds between the residues. As non-bonded contacts can be plentiful, the width of the striped line is proportional to the number of atomic contacts.

Table 3
The interactions between MEV and receptors of the docked complexes.

Complex	No. of interface residue		No. of interface area (Å^2)		No. of salt bridges	No. of H-bonds	No. of non-bonded contacts
	MEV	Receptor	MEV	Receptor			
MEV-TLR2	27	37	1606	1523	3	14	233
MEV-TLR4	37	52	2141	2008	10	20	288
MEV-ACE2 receptor	26	45	1852	1691	6	21	279
MEV- BCR (chain A)	15	19	890	808	3	8	96
MEV- BCR (chain B)	13	12	816	835	None	7	73

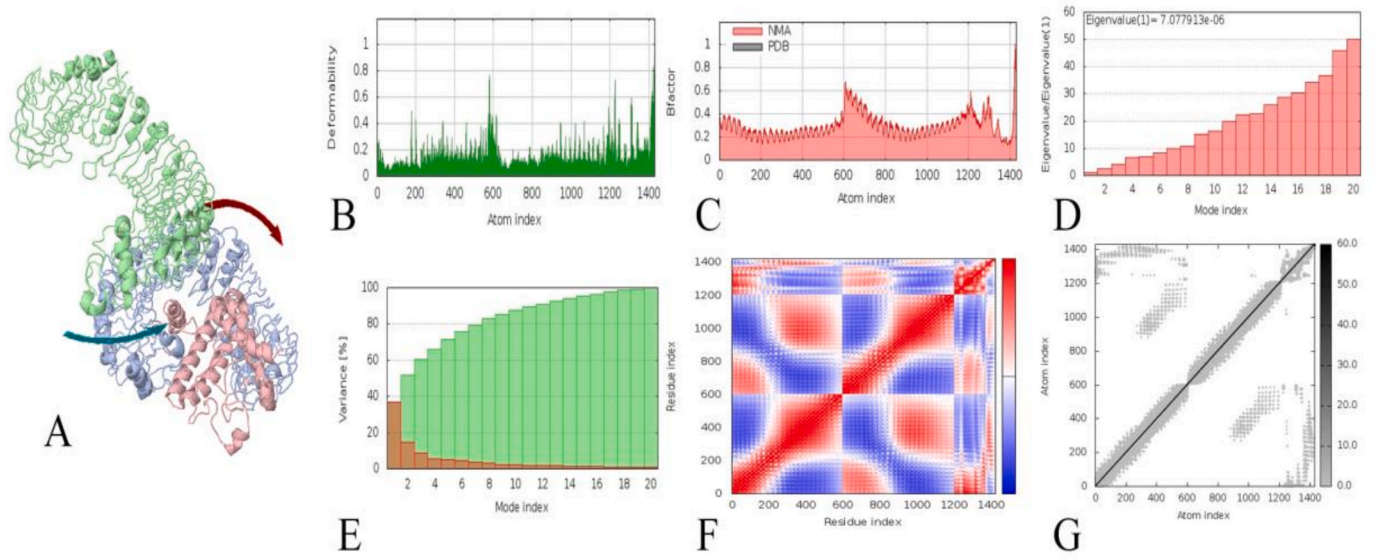

Fig. 8. Output of normal mode analysis (NMA) study by iMODS. A) NMA mobility of the vaccine-TLR complex with affine-model arrows, B) Deformability plot, C) B-factor plot, D) Eigenvalue plot, E) Normal mode variance plot. The red bars indicate variance of individual modes while the green bars indicate cumulative variance F) Covariance map. Red, white and blue color corresponds to correlated, uncorrelated and anti-correlated motions G) Elastic network. The darker the grey color, the stiffer the spring.

switching potential. The responses were long-lasting (Fig. 9B). The plasma B cells also produced increased IgM and IgG response in secondary and tertiary responses, which also indicated the presence of memory B cells and active B-cell proliferation (Fig. 9C). An elevated level of T_H cells and T_c cells was also observed along with the formation of memory T_H and T_c cells (Fig. 9 D, E and F). There was a substantial increase in T_H memory cell after the second dose and the level was almost equal in the third dose. The memory cells were long-lasting (Fig. 9D). The level of active state T_H cell was also increased sharply during the secondary dose and was almost equal for tertiary response (Fig. 9E). The T_c cells elevated to a maximum of greater than 1150 cell per mm^3 but then the level fluctuated (Fig. 9F). The NK cell numbers were also increased but the level fluctuated. The average level was 350 cells per mm^3 (Fig. 9G). The number of active macrophages increased with subsequent doses and then decreased to 20–25 cells per mm^3 after 44 days from the third dose (Fig. 9H). The IFN- γ production by the constructed vaccine was high which plays a vital role in immunity against virus (Fig. 9I). The increased IFN-γ production justifies the selection of IFN- γ producing MHC-Class II epitope, "^{424}KVKYLYFIKGLNNLN^{438}". The IL-2 production was also high, which demo nstrated that the constructed MEV will promote T-cell differentiation and proliferation.

3.16. Codon adaptation and cloning

The amino acid sequence of vaccine was translated back to a cDNA nucleotide sequence to inspect the expression and to clone the constructed vaccine inside a suitable vector. After codon optimization, the V1 vaccine construct showed 50.98935% GC content. The normal estimated GC content range is 40–60%, thus the GC content of the V1 vaccine is acceptable. Also, the codon optimization index (CAI) value was 0.983051, which is more than 0.5, thus is an admissible value and indicates high expression in the vector (Fig. 10).

4. Discussion

SARS-CoV provoked a large-scale epidemic and the pathogenic mechanisms behind this have been partially decrypted. Recent therapies have failed to meet the initial enthusiastic expectations [71], and now new waves of SARS-CoV-2, through the evolution of its new strains, are imminent in several parts of the world. Under the current condition, more therapeutic and preventive approaches are warranted, which might ameliorate the current pandemic by combating SARS-CoV-2.

Orf 10 was found to be the most conserved protein in this study as depicted by the highest density of conservancy value of 0.86 (Fig. 2B). Within RNA viruses, RNA-dependent RNA polymerase is thought to be a conserved protein [72]. RdRp consists of 932 amino acids (from Serine 4393 to Glutamine 5324 of the Orf 1 ab). Inspection of the RdRp portion of Orf1ab revealed that 729 sites out of 932 were conserved in all the sequences (there was no mutation at these sites), Two amino acid residues at positions 4619 and 4715 had high variations. At 4619, 7.09% of the sequences had leucine instead of proline and at 4715, 4.13% of the sequences had proline instead of leucine. The density of conservancy for RdRp was calculated to be 0.68. However, the whole Orf1ab protein had a density of conservancy of 0.37 and hence is less conserved. The high conservancy of Orf10 and RdRp shows that these proteins may play vital roles in SARS-CoV-2 life cycle. Targeting these proteins might be a great weapon for preventing viral infection and replication. The MEV constructed consisted of two epitopes "^{4733}VVSTGYHFR^{4741}" and "^{4801}TVKPGNFNK^{4809}" of ORF1ab that belong to the RdRp region and hence, the construct may play a crucial role in blocking the action of this essential protein.

The finally selected epitopes for vaccine construct were matched with the epitopes deposited in the IEDB database to identify any experimentally determined immune epitopes. Predicted T cell epitopes "^{4733}VVSTGYHFR^{4741}", "^{4801}TVKPGNFNK^{4809}" and "^{1891}EIDPKLDNY^{1899}" of ORF1ab, "^{1059}GVVFLHVTY^{1067}" of S and "^{138}LVIGAVILR^{146}" of M exactly matched the epitopes with IEDB IDs 71837, 67051, 1315808, 23200 and 132149, respectively. Epitope with IEDB ID 71837 ("^{4733}VVSTGYHFR^{4741}") bound with HLA-A*31:01, HLA-A*68:01, HLA-A*03:01,HLA-A*11:01 and HLA-A*33:01 with IC_{50} values of 1.69 nM,7.4 nM,31.5 nM,8.55 nM and 41.2 nM, respectively in MHC ligand binding assays. Epitope (^{4801}TVKPGNFNK^{4809}) labeled as 67051 in IEDB had been found to bind with HLA-A*31:01, HLA-A*68:01, HLA-A*03:01 and HLA-A*11:01 and HLA-A*33:01 with IC_{50} values of 41.6 nM, 122 nM, 81.5 nM and 33 nM, respectively. The epitopes (labeled as 71837 and 67051 in IEDB) were found in Replicase polyprotein 1 ab of SARS coronavirus Tor2. In T cell receptor (TCR) dependent Activation Induced

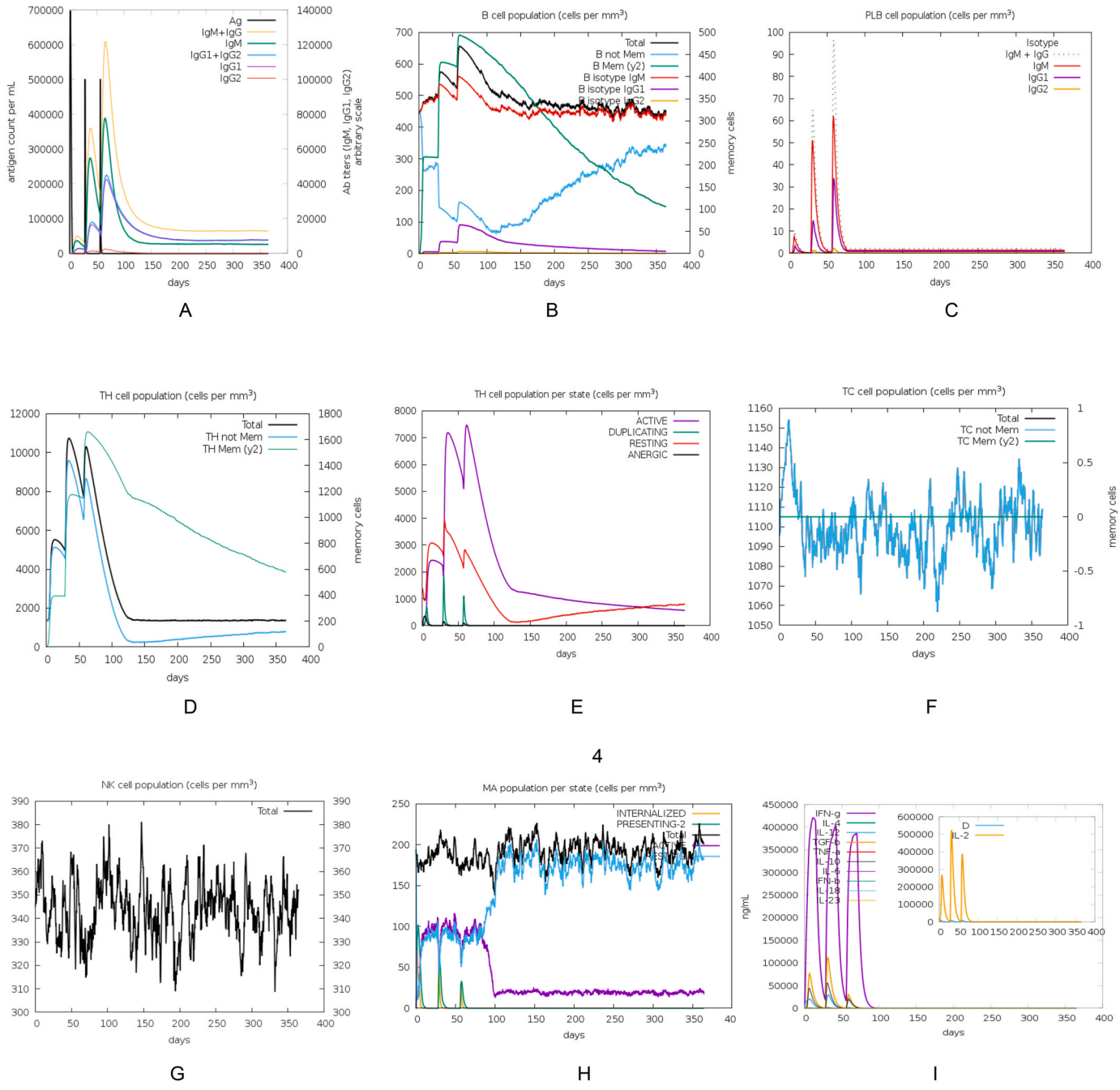

Fig. 9. Outcome of the immune simulation analysis of A) The Ab titers level increasing with each successive injection along with decreasing antigen count, B) B cell population (indicates elevation of different B-cell types and their class switching potential), C) PLB (Plasma B cell) population (indicates the presence of memory B cells and active B-cell proliferation), D) T_H cell population (indicates a substantial increase in T_H memory cell), E) T_H cell population per state (indicates the increase of active state T_H cell), F) T_c cell population (indicates the fluctuation of T_c cell population with time), G) NK cell population (indicates the fluctuation of the NK cell population with time), H) Macrophage population (indicates the fluctuation of the Macrophage population with time), I) Concentration of cytokines and interleukins (indicates the increased IFN- γ and IL-2 production). Inset plot shows danger signal together with leukocyte growth factor IL-2.

Marker (AIM) assays (which measured the combination of markers $CD69^{+}CD137+$),"^{4733}VVSTGYHFR^{4741}" of ORF1ab polyprotein, "^{1059}GVVFLHVTY^{1067}" of S and "^{138}LVIGAVILR^{146}" of M showed positive response in 100%, 50% and 50% of the subjects, respectively and the MHC restriction were HLA-A*11:01, HLA-A*32:01 and HLA-A*68:01, respectively [73]. Activation Induced Marker (AIM) assays measures T-cell response. In the same experiment, "^{1891}EIDPKLDNY^{1899}" of ORF1ab gave positive response in 50% of the subjects for both MHC restriction HLA-A*01:01 and HLA-A*26:01 by TCR dependent AIM assays [73]. Moreover, the HTL epitope "^{424}KVKYLYFIKGLNNLN^{438}" of ORF1ab and the B-cell epitopes "^{427}IADYNYKLP^{435}", "^{667}NSYECDIPIGAGIC^{680}" and "^{1111}WFVTQRNFY^{1119}" had overlapped with 5, 7, 9 and 25 immune epitopes available in IEDB database, respectively. Thus, the predicted T and B-cell epitopes are expected to elicit good cellular and humoral immune response.

B cell epitopes are entities that can be specifically recognized and can react with particular paratopes. Upon interaction with the epitopes followed by stimulation of the humoral immune response, B cell secretes antibodies which bind to the virus-surface structure protein at the time of viral entry [74]. In case of SARS-CoV-2, the S, M, E and N proteins are

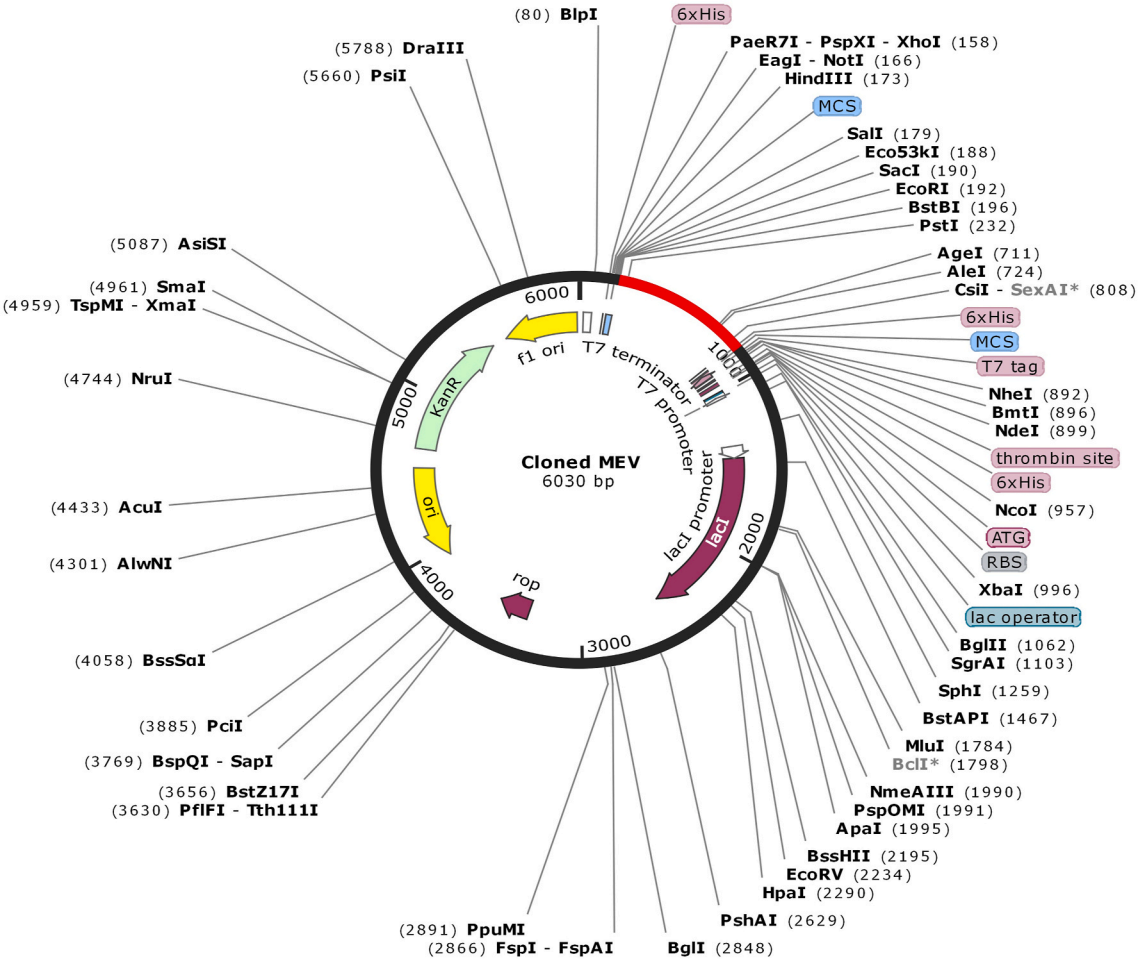

Fig. 10. In silico cloning of MEV gene in a restriction cloning vector pET28a (+) in *E.coli* host. Here, the red areas indicate the MEV, and the black areas represent the expression vector, pET28a (+).

involved in forming the outer layer of the virus. Among them, S protein forms the outermost structure which facilitates virus entry through binding to the ACE2 receptor [75]. Thus, for a more effective and rapid antibody-dependent immune response upon viral insertion, S protein is the most suitable target candidate for inhibiting host-virus interaction at the very beginning before other structural proteins come into play [76]. For the prediction of the linear or continuous epitope, two tools were used for a more specific outcome as each bioinformatics-based web server has its own algorithm and sensitivity. Thus, rather than random prediction, sequences that overlapped fully in both tools were considered for analysis to obtain a more accurate and reliable result [77]. However, a significant portion of the predicted epitopes (“^{427}IADYNYKLP^{435}”, “^{667}NSYECDIPIGAGIC^{680}”, “^{1111}WFVTQRNFY^{1119}”) passed all the featured criteria of being surface accessible, hydrophilic, flexible, and beta-turn containing, along with being non-allergen and non-toxic. These qualities are significant as beta-turns and hydrophilic residues are generally surface-exposed and play a significant role in immune response initiation [78].

For vaccine construction, a multi-epitope based vaccine was preferred compared to a single epitope-based vaccine because (i) it can induce both T_H and T_c response as it contains both MHC Class I and II epitopes (ii) it can lead to antibody production as it contains B cell epitopes, and (iii) it contains adjuvant beta–defensin which activate the primary innate antiviral immune response [79]. Hence, both humoral and cell-mediated immune response will be elicited by the vaccine construct, as suggested in the present study. For T cell MHC class II epitope prediction, the epitopes were analyzed for interferon-γ, IL-4 and IL-10 production. IL-10 is an immunosuppressive cytokine. Thus, IL-10 inducing epitopes can lead to immune suppression. IL-10 inhibits IFN-γ secretion from T_H1 cells [80]. IL-4 is responsible for allergic reactions. This is because IL-4 directs T_H2 development and thus leads to IgE production [81]. IFN- γ plays an important role in clearing viral infection [82].

To evaluate the interaction of the selected epitopes with HLA alleles, HLA-A*02:01and DRB1*01:01 were chosen as the representative HLA alleles as these are two of the most frequent alleles in world population [17]. The resulted interaction of the epitopes with the receptors indicated their potential role in eliciting an immune response.

After epitope selection, different linkers and adjuvant were used to develop a potent vaccine construct. Immunogenic adjuvant beta-defensin was used to activate the primary innate antiviral immune response along with other immunomodulatory activities including modulation of the adaptive immune response, and also has previously been appeared as a potent adjuvant when linked to MERS-CoV antigens [79]. It was attached to the N-terminal of the vaccine sequence with an EAAAK linker. This rigid spacing linker helped to increase the immunogenic properties by its helix formation properties. To complement its rigidity, to link the epitopes linkers GPGPG and AAY was used to decrease junctional immunogenicity and facilitate the rational design construction of a potent MEV [69,83,84].

The vaccine constructs were evaluated based on their physicochemical parameters, as these parameters have a major impact on the success of immunization [85]. For instance, the molecular weight and isoelectric point (pI) are important parameters for the solubility and the

nature of the vaccine construct. Molecular weight of the construct (24.326 kDa) developed in this study could be a concern with respect to immune response. Vaccine candidates with higher molecular weight can be difficult to develop and have the possibility of interfering with the normal function of the complement system. Also, it has been reported that the MEVs with molecular weight less than 110 kDa could be a good vaccine candidate [86,87]. Although our vaccine construct is 219 amino acids long, 53.4% of this construct is concentrated with the epitopes chosen through robust filtering process. Additionally, beta-defensin adjuvant will also activate innate antiviral immune response. The high percentage of residues of the vaccine construct represents epitopes consisting of the best characteristics (one of them is high antigenicity) that supports its potential immunogenicity. Moreover, Singh et al. (2007) identified a few low molecular weight (as low as 18 kDa) proteins in the Omps of *Salmonella* which showed potent T cell mediated immune response in patients with salmonella-induced reactive arthritis [86]. Thus, the constructed MEV with 24.326 kDa molecular weight could be a potent candidate for strong immune response. Additionally, the use of adjuvant in the vaccine construct helps to overcome any obstacle that could arise due to the lower molecular weight by increasing vaccine immunogenicity [88]. The construct possessed pI > 7 (9.4), indicating its alkaline nature. The half-life (1 h in mammalian reticulocytes in vitro) indicates vaccines required time for disappearance of half of the initial amount, which is related to the stability of the protein. The stability of the vaccine construct is indicated by the stability index. Here, the stability index was 23.69, which is < 40 and indicates the vaccine construct to be stable in standard condition [89]. Another parameter was the grand average of hydropathicity - GRAVY score −0.124. The mean value of hydropathy of proteins amino acids ranges between −2 and +2. The negative value indicates hydrophilicity and the hydrophilic nature indicates improved solubility [90]. Thus, the constructed MEV appeared to be suitable when all the physicochemical aspects were investigated/evaluated. We then opted to search for regions in the designed MEV that may serve as potential discontinous (conformational) B-cell epitopes and it was predicted that specially two discontinuous epitopes "^{61}YKLP^{64}"and "^{81}GPGNSYECD^{89}" overlapping the linear B cell epitopes present in the vaccine construct. These conformational epitopes may help to evoke the humoral immune response [51]. The designed MEV was docked with TLR2 and TLR4 which indicated the possibility of inducing innate immune response along with playing a vital role in maintaining the balance of T_H1 and T_H2 response [91]. TLR4, which is an agonist to beta-defensin, interacts with the adjuvant to enhance the immune response [87]. Also, along with TLR2, it plays role in the activation of the MyD88-dependant pathway which leads to early-phase NF-κB activation, and type 1 IFN production [92]. Moreover, due to the presence of TLR2 and TLR4 on innate immune cells like dendritic cells, macrophage, monocytes, binding of MEV with the receptors can lead to production of pro-inflammatory cytokines and chemokines [93]. Based on this MEV-TLR2/4 recognitions, activation of dendritic cells can link innate and adaptive immunity by efficiently processing and presenting antigens to T cells [94].

As COVID-19 uses angiotensin-converting enzyme 2 (ACE2) as its potential receptor, docking analysis of MEV was also performed with this and favorable interaction was observed. Thus, the constructed MEV has the potential to interrupt the interaction of viral surface glycoprotein with ACE2 receptor and thus can be considered as a possible therapeutic target for COVID-19. Finally, the interaction with BCR indicated the ability of the MEV to elicit cellular or hormonal immune response [95]. The docking interface and the bonds between the residues, along with the binding affinity indicated the MEVs possibility to exert its immunogenic function. Following the docking analysis, the immune simulation was done to assess the competency of the vaccine construct to the adaptive immune system mechanisms. The antibody titers showed elevation after the second dose that indicated enhanced primary immune response. Also, as response of the MEV, an increasing amount of T_H cells along with an elevated concentration of IFN-γ was observed. The increased T_H cells support B cell clonal expansion and antibody synthesis [96]. IFN-γ plays an important role in both innate and adaptive immunity by containing antiviral factors [97]. It promotes macrophage activation, T_H1 cell generation along with suppressing the T_H2 response [98]. In this way, T_H1/T_H2 balance is maintained so that excess T_H2 response cannot counteract the T_H1 mediated antiviral actions [99].

The iMODS server was used for the prediction of the binding stability and flexibility of the vaccine-TLR4 complex. The large eigenvalue of $7.077913e^{-06}$ (Fig. 8D) showed that high energy is required to deform the complex. Very few peaks were observed in the deformability plot (Fig. 8B) that illustrated high stability of the complex with an extremely low chance of deformability. The complex had a high number of correlated residues as shown by the red color in the covariance map (Fig. 8F). A higher number of stiffer regions was present in the vaccine-TLR4 complex, as illustrated by the grey dots in the elastic network (Fig. 8G). The results from iMODs demonstrated stable binding of the vaccine-TLR4 complex.

The MEV construct was optimized by codon adaptation approach that helps to increase protein expression by up to >1000-fold [100]. Thus, it facilitates and maximizes protein expression by controlling expression limitations related to codon usage [101]. Through this process, the Codon Adaptation Index (CAI) and GC content were obtained. The CAI index varies from 0 to 1, where a value equal to 1 means the maximum codon affinity. As the CAI value of the MEV construct was 0.983051, it can be stated that the MEV obtained nearly maximum codon affinity. The GC content 40–60% indicates a good comprehensive stability of mRNA from the synthetic gene [102], thus the 50.98935% GC content of the MEV fulfilled the criterion. However, apart from all these analyses and findings, additional experimental validation is needed to confirm the safety and effectiveness of the designed MEV.

Most of the nucleotide sequences deposited in NCBI are from USA. The variant B.1.1.7 is highly prevalent in USA and 162 sequences included in the present study also belonged to B.1.1.7. However, the low prevalence of P.1 and B.351.1 led to retrieval of few sequences from these variants (B.1.351 = 2 and P.1 = 3). Though the sequences retrieved were mostly of USA origin, the MEV constructed had high population coverage for countries having high morbidity and infection.

Supplementary File 6 shows the viral strains and proteins considered in different epitope-based vaccine construction studies against SARS-CoV-2. The presence of immunodominant $CD8^+$ T-cells against internal proteins of SARS-CoV-2 in convalescent individuals highlights the importance of those proteins in vaccine construction [103]. Most of the vaccine construction studies against SARS-CoV-2 had focused only on S or the structural proteins (S, M, E and N) for both B and T-cell epitope prediction [17,62,79,87,89,104–108]. In the present study, we predicted T cell epitopes from the conserved regions of all the 11 proteins of SARS-CoV-2 (conserved regions presented in Supplementary file 2, sheet 8). Due to stringent and robust selection criteria, 8 CTL epitopes from three proteins (ORF 1 ab = 6, S = 1, and M = 1) and 1 HTL epitope from Orf 1 ab were selected for designing the MEV construct together with three B-cell epitopes. The epitopes from other proteins did not pass the filtering process and hence, were not included in the chimeric vaccine construct. Moreover, the epitopes were designed using conserved region of protein sequences of SARS-CoV-2 which included the variants of concern (B.1.1.7, B.1.351, P.1,and B.1.617.2) and the variants of interest (B.1.427, B.1.429, B.1.526, B.1.617.1and P.2). Although some studies considered the alpha (B.1.1.7) variant of concern in vaccine construction, none included as many variants of concern and interest as we did in this present study [106,108]. This surely justifies the robustness of our vaccine construct in combating the new variants. Moreover, the resemblance of B and T cell epitopes with immune epitopes in IEDB explains the possible elicitation of cellular and humoral response.

5. Conclusion

This global SARS-CoV-2 pandemic has taken many precious lives.

This contiguous virus has crippled the whole world economy. The only effective way to contain the spread of this devastating virus is the successful vaccination program. Though several countries have initiated mass vaccination, successful vaccination against all the strains of coronavirus is still far from attainable. In this study, we have designed a multi-epitope-based vaccine to tackle this vicious virus. *In silico* based immunoinformatics approaches were used for the construction of multi-epitope vaccines. The predicted multi-epitope vaccine is immunogenic and appears safe to use. The constructed multi-epitope vaccine is capable of eliciting both innate and adaptive immunity (humoral and cellular). However, in vitro and animal studies are warranted to demonstrate the efficacy of the designed multi-epitope vaccine for its possible utility as a potent preventive measure (Fig. 1).

Author contributions

AHMNN conceived the idea and supervised the research work. AAS and MA retrieved the data and performed data analyses for B cell and T cell epitope prediction. PS and MIH performed the tertiary structure prediction, molecular refinement, and structure validation of MEV work. SC performed the docking analysis. AHMNN, AAS and MA prepared the first draft and then revised the draft carefully and critically. All authors approved the manuscript.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgement

Part of this research work was supported by the establishment developed from the special grants (Code#1280101-120008431-3631107) provided by Information and Communication Technology Division under the Ministry of Posts, Telecommunications and Information Technology, Government of the People's Republic of Bangladesh (2019–2020).

Appendix A. Supplementary data

Supplementary File 1. Prediction of allele binding to MHC Class I and MHC II molecules. Sheet 1: MHC Class I binding prediction; Sheet 2: MHC Class II binding prediction.**Supplementary file 2**. Sequence retrieval and conserved sites. Sheet 1: Number of sequences taken from each month; Sheet 2: Number of protein sequences retrieved for each protein; Sheet 3: Accession numbers of sequences for which there were no protein sequences for any protein.; Sheet 4: Accession numbers of sequences for which there were proteins sequences for Orf 8 protein.; Sheet 5: Pango lineage distribution of all the 1549 sequences; Sheet 6: Number of sequences belonging to each Pango lineage; Sheet 7: Insertions and deletions in different protein sequences. Sheet 8: Conserved sites from each protein.**Supplementary File 3**. Final list of epitopes with interacting alleles.**Supplementary File 4**. IEDB analysis resource prediction- Sheet 1) Emini surface accessibility prediction: list of surface accessible peptides. Sheet 2) Parker Hydrophilicity prediction. Sheet 3) Karplus and Schulz flexibility prediction. Sheet 4) Chou and Fasman Beta Turn Prediction.**Supplementary File 5**. ClusPro Balanced Model Scores- Sheet 1) MEV-TLR2 complex. Sheet 2) MEV- TLR4 complex. Sheet 3) MEV-ACE2 receptor complex. Sheet 4) MEV- BCR complex. **Supplementary File 6**. Proteins for T-cell epitope prediction and SARS-CoV-2 strains considered in different vaccine-construction against SARS-CoV-2 studies.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2021.104703.

References

[1] J.M. Abramo, A. Reynolds, G.T. Crisp, et al., Individuality in music performance, Assess Eval. High Educ. 37 (2012) 435.
[2] X. Ma, Ph D, D. Wang, et al., A Novel Coronavirus from Patients with Pneumonia in China, 2019, pp. 727–733, 2020.
[3] A. Almofti Y, K. Ali Abd-elrahman, S. Abd Elgadir Gassmallah, et al., Multi epitopes vaccine prediction against severe acute respiratory syndrome (SARS) coronavirus using immunoinformatics approaches, Am. J. Microbiol. Res. 6 (2018) 94–114.
[4] M.M. Badawi, M.A. Salaheldin, M.M. Suliman, In Silico Prediction of a Novel Universal Multi-Epitope Peptide Vaccine in the in Silico Prediction of a Novel Universal Multi-Epitope Peptide Vaccine in the Whole Spike Glycoprotein of MERS CoV, 2016.
[5] M.S. Rahman, M.N. Hoque, M.R. Islam, et al., Epitope-based chimeric peptide vaccine design against S, M and e proteins of SARS-CoV-2, the etiologic agent of COVID-19 pandemic: an in silico approach, PeerJ 8 (2020).
[6] J.F. Chan, K. Kok, Correction to: Genomic Characterization of the 2019 Novel Human-Pathogenic Coronavirus Isolated from a Patient with Atypical Pneumonia after Visiting Wuhan, Emerging Microbes & Infections, 2020, 9, 1, (221-236), 10.1080/22221751.2020.1719902). Emerg. Microbes Infect. 2020; 9:540.
[7] F. Krammer, SARS-CoV-2 vaccines in development, Nature 586 (2020) 516–527.
[8] G.N. Sapkal, P.D. Yadav, R. Ella, et al., Neutralization of UK-variant VUI-202012/01 with COVAXIN vaccinated human serum, bioRxiv 2021 (2021), 01.26.426986.
[9] D.Y. Logunov, I.V. Dolzhikova, D.V. Shcheblyakov, et al., Safety and efficacy of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine: an interim analysis of a randomised controlled phase 3 trial in Russia, Lancet 397 (2021) 671–681.
[10] L.A. Jackson, E.J. Anderson, N.G. Rouphael, et al., An mRNA vaccine against SARS-CoV-2 — preliminary report, N. Engl. J. Med. 383 (2020) 1920–1931.
[11] F.P. Polack, S.J. Thomas, N. Kitchin, et al., Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine, N. Engl. J. Med. 383 (2020) 2603–2615.
[12] B.Y. Kathy, K. April, Comparing the COVID-19 Vaccines : How Are They Different ? the Three Vaccines Authorized by the FDA Thumbnail Illustration of a Needle on a Blue Background Pfizer-BioNTech, 2021.
[13] G.N. Sapkal, P.D. Yadav, R. Ella, et al., Inactivated COVID-19 vaccine BBV152/COVAXIN effectively neutralizes recently emerged B 1.1.7 variant of SARS-CoV-2, J. Trav. Med. 28 (4) (2021) taab051.
[14] I. Jones, P. Roy, V. Sputnik, COVID-19 vaccine candidate appears safe and effective, Lancet 397 (2021) 642–643.
[15] S. Ikegame, M.N.A. Siddiquey, C.-T. Hung, et al., Qualitatively distinct modes of Sputnik V vaccine-neutralization escape by SARS-CoV-2 Spike variants, medRxiv 2021 (2021), 03.31.21254660.
[16] B.R. Singh, Pros and Cons of Covid-19 Vaccines and Vaccination, 2021.
[17] A. Singh, M. Thakur, L.K. Sharma, et al., Designing a multi-epitope peptide based vaccine against SARS-CoV-2, Sci. Rep. 10 (2020) 1–12.
[18] SARS-CoV-2 Variant Classifications and Definitions.
[19] A. Rambaut, E.C. Holmes, Á. O'Toole, et al., A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology, Nat. Microbiol. 5 (2020) 1403–1407.
[20] B. Korber, W.M. Fischer, S. Gnanakaran, et al., Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus, Cell 182 (2020) 812–827, e19.
[21] Y. Liu, J. Liu, H. Xia, et al., Neutralizing activity of BNT162b2-elicited serum, N. Engl. J. Med. 384 (2021) 1466–1468.
[22] Moruf, Alicia. FACT SHEET FOR HEALTH CARE PROVIDERS EMERGENCY USE AUTHORIZATION (EUA) OF BAMLANIVIMAB AND ETESEVIMAB AUTHORIZED USE.
[23] SARS-CoV-2 Variants of Concern | CDC.
[24] Mani, Nina. FACT SHEET for HEALTH CARE PROVIDERS EMERGENCY USE AUTHORIZATION (EUA) of REGEN-COV TM (Casirivimab with Imdevimab).
[25] W.F. Garcia-Beltran, E.C. Lam, St Denis K, et al., Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity, Cell 184 (9) (2021) 2372–2383.
[26] X. Shen, H. Tang, R. Pajon, et al., Neutralization of SARS-CoV-2 variants B.1.429 and B.1.351, N. Engl. J. Med. 384 (24) (2021) 2352–2354. NEJMc2103740.
[27] Explained: B.1.617 Variant and the Covid-19 Surge in India | Explained News,The Indian Express.
[28] A. Patronov, I. Doytchinova, T-cell epitope vaccine design by immunoinformaticsPatronov, A., & Doytchinova, I., T-cell epitope vaccine design by immunoinformatics. Open Biology, 3(1), 120139, Open Biol. 2013 3 (2013) 120139.
[29] L. Zhang, Multi-epitope vaccines: a promising strategy against tumors and viral infections. Cell, Mol. Immunol. 15 (2018) 182–184.
[30] X. Lin, S. Chen, X. Xue, et al., Chimerically fused antigen rich of overlapped epitopes from latent membrane protein 2 (LMP2) of Epstein–Barr virus as a potential vaccine and diagnostic agent, Cell. Mol. Immunol. 13 (2016) 492–501.
[31] H. Wickham, R. Francois, L. Henry, et al., Dplyr: A Grammar of Data Manipulation, 2020.
[32] S. Abadi, D. Azouri, T. Pupko, et al., Model selection may not be a mandatory step for phylogeny reconstruction, Nat. Commun. 10 (2019) 1–11.
[33] S. Kumar, G. Stecher, M. Li, et al., Molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549.
[34] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, BMC Bioinf. 8 (2007) 1–7.

[35] A. Krogh, B. Larsson, G. Von Heijne, et al., Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J. Mol. Biol. 305 (2001) 567–580.
[36] M.C. Jespersen, B. Peters, M. Nielsen, et al., BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, Nucleic Acids Res. 45 (2017). W24–W29.
[37] G. Sanchez, Las instituciones de ciencia y tecnología en los procesos de aprendizaje de la producción agroalimentaria en Argentina. El Sist. argentino innovación Inst. Empres. y redes, El desafío la creación y apropiación Conoc 48 (2013) 40–48.
[38] E.A. Emini, J.V. Hughes, D.S. Perlow, et al., Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide, J. Virol. 55 (1985) 836–839.
[39] P.A. Karplus, G.E. Schulz, Prediction of chain flexibility in proteins - a tool for the selection of peptide antigens, Naturwissenschaften 72 (1985) 212–213.
[40] P.Y. Chou, G.D. Fasman, Prediction of the secondary structure of proteins from their amino acid sequence, Adv. Enzymol. Relat. Area Mol. Biol. 47 (1978) 45–148.
[41] B. Reynisson, B. Alvarez, S. Paul, et al., NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data, Nucleic Acids Res. 48 (2020) W449–W454.
[42] D. Weiskopf, M.A. Angelo, E.L. De Azeredo, et al., Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells, Proc. Natl. Acad. Sci. U.S.A. 110 (2013).
[43] P. Wang, J. Sidney, C. Dow, et al., A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach, PLoS Comput. Biol. 4 (2008), e1000048.
[44] P. Wang, J. Sidney, Y. Kim, et al., Peptide binding predictions for HLA DR, DP and DQ molecules, BMC Bioinf. 11 (2010) 568.
[45] J. Greenbaum, J. Sidney, J. Chung, et al., Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes, Immunogenetics 63 (2011) 325–335.
[46] Tang, et al., 基因的改变NIH public access, Bone 2008 23 (2005) 1–7.
[47] COVID-19 map - johns Hopkins coronavirus resource center.
[48] J. Maupetit, P. Derreumaux, P. Tufféry, A fast method for large-scale de novo peptide and miniprotein structure prediction, J. Comput. Chem. 31 (2010) 726–738.
[49] B. Xia, S. Vajda, D. Kozakov, Accounting for pairwise distance restraints in FFT-based protein-protein docking, Bioinformatics 32 (2016) 3342–3344.
[50] S.R. Comeau, D.W. Gatchell, S. Vajda, et al., ClusPro: a fully automated algorithm for protein-protein docking, Nucleic Acids Res. 32 (2004). W96–W99.
[51] R.A. Shey, S.M. Ghogomu, K.K. Esoh, et al., In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases, Sci. Rep. 9 (2019) 4409.
[52] I. Dimitrov, L. Naneva, I. Doytchinova, et al., AllergenFP: allergenicity prediction by descriptor fingerprints, Bioinformatics 30 (2014) 846–851.
[53] Gasteiger E, Hoogland C, Gattiker A, et al., The Proteomics Protocols Handbook. Proteomics Protoc. Handb, 2005, pp. 571–608.
[54] E. Saghapour, M. Sehhati, Physicochemical position-dependent properties in the protein secondary structures, Iran. Biomed. J. 23 (2019) 253–261.
[55] S. Wang, W. Li, S. Liu, et al., RaptorX-Property: a web server for protein structure property prediction, Nucleic Acids Res. 44 (2016) W430–W435.
[56] J. Cheng, A.Z. Randall, M.J. Sweredoski, et al., SCRATCH: a protein structure and structural feature prediction server, Nucleic Acids Res. 33 (2005) W72–W76.
[57] L. Heo, H. Park, C. Seok, GalaxyRefine: protein structure refinement driven by side-chain repacking, Nucleic Acids Res. 41 (2013) 384–388.
[58] J. Ko, H. Park, L. Heo, et al., GalaxyWEB server for protein structure prediction and refinement, Nucleic Acids Res. 40 (2012) W294–W297.
[59] C. Colovos, T.O. Yeates, Verification of protein structures: patterns of nonbonded atomic interactions, Protein Sci. 2 (1993) 1511–1519.
[60] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res. 35 (2007).
[61] V. Mariani, M. Biasini, A. Barbato, et al., lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests, Bioinformatics 29 (2013) 2722–2728.
[62] B. Sarkar, M.A. Ullah, Y. Araf, et al., Engineering a novel subunit vaccine against SARS-CoV-2 by exploring immunoinformatics approach, Informatics Med. Unlocked 21 (2020) 100478.
[63] R.A. Laskowski, PDBsum new things, Nucleic Acids Res. 37 (2009) D355–D359.
[64] J.R. López-Blanco, J.I. Aliaga, E.S. Quintana-Ortí, et al., IMODS: internal coordinates normal mode analysis server, Nucleic Acids Res. 42 (2014) 271–276.
[65] N. Rapin, O. Lund, M. Bernaschi, et al., Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system, PloS One 5 (2010).
[66] A. Grote, K. Hiller, M. Scheer, et al., JCat: a novel tool to adapt codon usage of a target gene to its potential expression host, Nucleic Acids Res. 33 (2005) 526–531.
[67] S. Baseer, S. Ahmad, K.E. Ranaghan, et al., Towards a peptide-based vaccine against Shigella sonnei: a subtractive reverse vaccinology based approach, Biologicals 50 (2017) 87–99.
[68] A. Ikai, Thermostability and aliphatic index of globular proteins, J. Biochem. 88 (1980) 1895–1898.
[69] M. Ali, R.K. Pandey, N. Khatoon, et al., Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection, Sci. Rep. 7 (2017) 9232.
[70] J.A. Kovacs, P. Chacón, R. Abagyan, Predictions of protein flexibility: first-order measures, Proteins Struct. Funct. Genet. 56 (2004) 661–668.
[71] S. Iacob, D.G. Iacob, SARS-CoV-2 treatment approaches: numerous options, No certainty for a versatile virus, Front. Pharmacol. 11 (2020) 1–15.
[72] S.O. Aftab, M.Z. Ghouri, M.U. Masood, et al., Analysis of SARS-CoV-2 RNA-dependent RNA polymerase as a potential therapeutic drug target using a computational approach, J. Transl. Med. 18 (2020) 275.
[73] A. Tarke, J. Sidney, C.K. Kidd, et al., Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases, Cell Reports Med 2 (2021).
[74] J.R. Lon, Y. Bai, B. Zhong, et al., Prediction and evolution of B cell epitopes of surface protein in SARS-CoV-2, Virol. J. 17 (2020) 165.
[75] Z. Noorimotlagh, C. Karami, S.A. Mirzaee, et al., Immune and bioinformatics identification of T cell and B cell epitopes in the protein structure of SARS-CoV-2: a systematic review, Int. Immunopharm. 86 (2020) 106738.
[76] M.H.V. Van Regenmortel, Immunoinformatics may lead to a reappraisal of the nature of B cell epitopes and of the feasibility of synthetic peptide vaccines, J. Mol. Recogn. 19 (2006) 183–187.
[77] T. Yasmin, A.H.M.N. Nabi, B and T Cell epitope-based peptides predicted from evolutionarily conserved and whole protein sequences of ebola virus as vaccine targets, Scand. J. Immunol. 83 (2016) 321–337.
[78] B. Ahmad, U.A. Ashfaq, M.-U. Rahman, et al., Conserved B and T cell epitopes prediction of ebola virus glycoprotein for vaccine development: an immuno-informatics approach, Microb. Pathog. 132 (2019) 243–253.
[79] E. Behmard, B. Soleymani, A. Najafi, et al., Immunoinformatic design of a COVID-19 subunit vaccine using entire structural immunogenic epitopes of SARS-CoV-2, Sci. Rep. 10 (2020) 1–12.
[80] Y. Yanagawa, K. Iwabuchi, K. Onoé, Co-operative action of interleukin-10 and interferon-γ to regulate dendritic cell functions, Immunology 127 (2009) 345–353.
[81] J.J. Ryan, Updates on cells and cytokines, J. Allergy Clin. Immunol. 99 (1997) 1–5.
[82] D.A. Chesler, C.S. Reiss, The role of IFN-γ in immune responses to viral infections of the central nervous system, Cytokine Growth Factor Rev. 13 (2002) 441–454.
[83] B. Meza, F. Ascencio, A.P. Sierra-Beltrán, et al., A novel design of a multi-antigenic, multistage and multi-epitope vaccine against Helicobacter pylori: an in silico approach, Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 49 (2017) 309–317.
[84] N. Khatoon, R.K. Pandey, V.K. Prajapati, Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach, Sci. Rep. 7 (2017) 8285.
[85] E. Only, Guidelines on Stability Evaluation, 2006, pp. 23–27.
[86] A.J. Obaidullah, M.M. Alanazi, N.A. Alsaif, et al., Immunoinformatics-guided design of a multi-epitope vaccine based on the structural proteins of severe acute respiratory syndrome coronavirus 2, RSC Adv. 11 (2021) 18103–18121.
[87] R. Saha, P. Ghosh, V.L.S.P. Burra, Designing a next generation multi-epitope based peptide vaccine candidate against SARS-CoV-2 using computational approaches, 3 Biotech 11 (2021) 1–14.
[88] M. TopuzoGullari, T. Acar, P. Pelİt Arayici, et al., An insight into the epitope-based peptide vaccine design strategy and studies against COVID-19, Turkish J. Biol. = Turk Biyol. Derg. 44 (2020) 215–227.
[89] T. Kar, U. Narsaria, S. Basak, et al., A candidate multi-epitope vaccine against SARS-CoV-2, Sci. Rep. 10 (2020) 1–24.
[90] A. Kaur, P.K. Pati, A.M. Pati, et al., Physico-chemical characterization and topological analysis of pathogenesis-related proteins from Arabidopsis thaliana and Oryza sativa using in-silico approaches, PloS One 15 (2020), e0239836.
[91] S. Mukherjee, S. Karmakar, S.P.S. Babu, TLR2 and TLR4 mediated host immune responses in major infectious diseases: a review, Brazilian J. Infect. Dis. an Off. Publ. Brazilian Soc. Infect. Dis. 20 (2016) 193–204.
[92] J.K. Dowling, A. Mansell, Toll-like receptors: the swiss army knife of immunity and vaccine development, Clin. Transl. Immunol. 5 (2016) e85–e85.
[93] L. Oliveira-Nascimento, P. Massari, L.M. Wetzler, The role of TLR2 in infection and immunity, Front. Immunol. 3 (2012) 79.
[94] R. Noubade, S. Majri-Morrison, K.V. Tarbell, Beyond cDC1: emerging roles of DC crosstalk in cancer immunity, Front. Immunol. 10 (2019) 1014.
[95] A. Kazi, C. Chuah, A.B.A. Majeed, et al., Current progress of immunoinformatics approach harnessed for cellular- and antibody-dependent vaccine design, Pathog. Glob. Health 112 (2018) 123–131.
[96] Smith Km, Pottage L, Thomas Er, et al., Th1 and Th2 CD4+ T cells provide help for B cell clonal expansion and antibody synthesis in a similar manner, In Vivo. J. Immunol 165 (6) (2000) 3136–3144.
[97] N. Kopitar-Jerala, The role of interferons in inflammation and inflammasome activation, Front. Immunol. 8 (2017) 873.
[98] P. Rothman, Biologic Functions of the IFN- c Receptors, 1999, 1233–1251.
[99] A. Berger, Th1 and Th2 responses: what are they? BMJ 321 (2000) 424.
[100] C. Gustafsson, J. Minshull, S. Govindarajan, et al., Engineering genes for predictable protein expression, Protein Expr. Purif. 83 (2012) 37–46.
[101] V.P. Mauro, Codon optimization in the production of recombinant biotherapeutics: potential risks and considerations, BioDrugs 32 (2018) 69–81.
[102] H. Dana, G.M. Chalbatani, E. Gharagouzloo, et al., In silico analysis, molecular docking, molecular dynamic, cloning, expression and purification of chimeric protein in colorectal cancer treatment, Drug Des. Dev. Ther. 14 (2020) 309–329.

[103] Y. Peng, A.J. Mentzer, G. Liu, et al., Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19, Nat. Immunol. 21 (2020) 1336–1345.
[104] J. Kumar, R. Qureshi, S.R. Sagurthi, et al., Designing of nucleocapsid protein based novel multi-epitope vaccine against SARS-COV-2 using immunoinformatics approach, Int. J. Pept. Res. Therapeut. 27 (2020) 941–956.
[105] A. Alam, A. Khan, N. Imam, et al., Design of an epitope-based peptide vaccine against the SARS-CoV-2: a vaccine-informatics approach, Briefings Bioinf. 22 (2021) 1309–1323.
[106] A. Khan, S. Khan, S. Saleem, et al., Immunogenomics guided design of immunomodulatory multi-epitope subunit vaccine against the SARS-CoV-2 new variants, and its validation through in silico cloning and immune simulation, Comput. Biol. Med. 133 (2021) 104420.
[107] S.R. Mahapatra, S. Sahoo, B. Dehury, et al., Designing an efficient multi-epitope vaccine displaying interactions with diverse HLA molecules for an efficient humoral and cellular immune response to prevent COVID-19 infection, Expert Rev. Vaccines 19 (2020) 871–885.
[108] Y.A. Almofti, K.A. Abd-elrahman, E.E.M. Eltilib, Vaccinomic approach for novel multi epitopes vaccine against severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), BMC Immunol. 22 (2021) 1–20.