JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
https://doi.org/10.1080/07391102.2021.1977702

In silico approach to design a multi-epitopic vaccine candidate targeting the non-mutational immunogenic regions in envelope protein and surface glycoprotein of SARS-CoV-2

M. Susithra Priyadarshni[a], S. Isaac Kirubakaran[b] and M. C. Harish[a]

[a]Department of Biotechnology, Thiruvalluvar University, Vellore, Tamil Nadu, India; [b]Division of Pulmonary, Critical Care and Sleep Medicine, Department of Internal Medicine, University of Kansas Medical Center, KS, USA

Communicated by Ramaswamy H. Sarma

ABSTRACT

The novel corona virus (COVID-19) is a causative agent for severe acute respiratory syndrome (SARS-CoV-2) and responsible for the current human pandemic situation which has caused global social and economic commotion. The currently available vaccines use whole viruses whereas there is scope for peptide based vaccines. Thus, the global raise in statistics of this infection at an alarming rate evoked us to determine a novel and effective vaccine candidate against SARS-CoV-2. To find the potential vaccine candidate targets, immunoinformatics approaches were used to analyze the mutations in the envelope protein and surface glycoprotein and determine the conserved region; further specific T-cell epitopes *VSLVKPSFY, SLVKPSFYV, RVKNLNSSR, SEETGTLIV, LVKPSFYVY, LTDEMIAQY, YLQPRTFLL, RLFRKSNLK, SPRRARSVA, AEIRASANL, TLLALHRSY, YSRVKNLNS* and *FELLHAPAT* and B-cells epitopes *TLAILTALRLCAYCCN* and *AGTITSGWTFGAGAAL* were identified. The 3 D structure of epitope was predicted, refined and validated. The molecular docking analysis of multi-epitope vaccine candidates with TLR receptors, predicted effective binding. Overall, using bioinformatics approach this multi-epitopic target facilitates the proof of concept for SARS-CoV-2 conserved epitopic vaccine design.

ARTICLE HISTORY
Received 27 February 2021
Accepted 3 September 2021

KEYWORDS
SARS-CoV-2; *In silico*; envelope protein; surface glycoprotein; epitopic vaccine

Introduction

Severe acute respiratory syndrome corona virus 2 (SARS-CoV-2), a positive sense RNA virus belongs to the family Beta corona virus is the causative agent of ongoing pandemic corona virus disease 2019 (COVID-19). The outbreak of COVID-19 was first identified in Wuhan, China later the virus has spread to 213 countries with 175, 686, 814 cases and 3, 803, 592 deaths as on 14th June, 2021 (WHO, n.d). SARS-CoV-2 is a zoonotic infection similar to SARS and MERS (Middle East Respiratory Syndrome). Bats are found to be the natural reservoirs for this SARS-CoV-2 virus, further the transmission to human is believed to be through intermediate host pangolin, but there is no confirmed evidence for the same (Lam et al., 2020). The SARS-CoV-2 consists of four structural proteins viz. Surface glycoprotein (S), Envelope protein (E), Membrane protein (M) and Nucleocapsid protein (N). The envelope protein is present along with the surface glycoprotein and plays an important role in viral genome assembly (Westerbeck & Machamer, 2019; Zheng et al., 2021). These envelope proteins oligomerize and forms ion channels, thereby responsible for pathogenesis. They are also involved in crucial phases of viral life cycle like envelope formation, pathogenesis, budding and assembly (Ashour et al., 2020). The viral entry occurs through the contact between the surface glycoprotein and the host cell receptor, i.e. human ACE2 receptor (angiotensin converting enzyme 2) (Li et al., 2005). The surface glycoprotein is divided into S1 and S2 region, the S1 region has the receptor binding domain (RBD) which has high affinity for hACE2 receptor (Yan et al., 2020). Interestingly, this RBD region in S1 region of surface glycoprotein is getting mutated to specifically bind to hACE2 receptor (Wan et al., 2020). The rapid genetic evolution of SARS-CoV-2 virus leads to increased spread and possible immune evasion. The surface glycoprotein undergoes rapid mutations; the first widely identified mutation is D614G (Biswas & Majumder, 2020; Isabel et al., 2020). Newer variants with further mutations are spreading swiftly around the world, variant B.1.1.7 in UK, variant B.1.351 in South Africa, variant B.1.1.248 in Brazil and variant B.1.429 in California (Naveca et al., 2021; Rambaut et al., 2020; Tegally et al., 2021; Zhang et al., 2021). Mutations in surface glycoprotein are of major concern because this protein involves in initial entry and attachment of virus and foremost target for neutralizing antibodies and vaccines (Piccoli et al., 2020; Shen et al., 2021). Thus, the envelope protein and surface glycoprotein has become a target site for candidate peptide vaccine development. Multi-epitopic vaccines with T-cell and B-cell epitopes are capable to evoke both cell mediated and humoral immune response without resulting in any immune complications. Moreover, the effectiveness of the vaccine can be regulated by selecting the epitopes that binds with large number of alleles resulting in diverse

CONTACT M. C. Harish mc.harishin@tvu.edu.in Department of Biotechnology, Thiruvalluvar University, Serkkadu, Vellore 632115, India
Supplemental data for this article can be accessed online at https://doi.org/10.1080/07391102.2021.1977702.

immune response (Elfiky, 2021; Foged, 2011; Sadat et al., 2021; Yang et al., 2021).

This study aimed to identify potential T-cell and B-cell epitopes from conserved and immunogenic regions of surface glycoprotein and envelope protein and design a multi-epitopic vaccine target for SARS-CoV-2. The designed multi-epitopic vaccine target was subjected to secondary and tertiary structure prediction and validation. Immunological properties were analyzed, docking and molecular dynamics simulation was performed to evaluate the interaction between the multi-epitopic target and TLR receptors. We investigated the immune response of this multi-epitopic vaccine target by immune simulation analysis.

Methods

Study design

The present study includes the following steps: (1) Data set construction and multiple sequence alignment and genome variation analysis, (2) Epitopes prediction, (3) Construction of multi-epitope vaccine and codon optimization, (4) Multi-epitopic vaccine construct property analysis, (5) Structure prediction and validation, (6) Immunological analysis for multi-epitopic vaccine and disulphide engineering, (7) Molecular docking and molecular dynamics simulation, (8) *In silico* cloning and (9) Immune simulation analysis. The study design is illustrated in Figure 1.

Data set construction and multiple sequence alignment

Genome sequence of SARS-CoV-2 (Ref. Sequence: NC_045512.2) from China and other parts of world were retrieved from the NCBI genome database (Supplementary Table S1). These datasets were constructed by including the NCBI accession numbers and GenBank IDs of envelope protein and surface glycoprotein, the source country and specific regions for all the retrieved sequences. The MSA was carried out using CLUSTAL W (https://www.genome.jp/tools-bin/clustalw) (Thompson et al., 1994) and the aligned sequences are viewed using BioEdit software version 7.0.5 (Hall, 1999).

Genome variation analysis and phylogenetic analysis

The genome variation analysis for envelope protein and surface glycoprotein was carried out by Genome Detective Corona virus Typing Tool (version 1.1.3) (Cleemput et al., 2020), which permit rapid recognition and categorization of corona virus genome. The nucleotide sequences were submitted in FASTA format and this tool helps in identifying the mutation present in each gene. The phylogenetic analysis was carried out to confirm that all the retrieved sequences fall under same clade of SARS-CoV-2.

Prediction of cytotoxic T-Lymphocyte and analysis of binding affinity with MHC I and cluster analysis

NetCTL 1.2 server was employed to predict CTL epitopes of SARS-CoV-2 envelope and surface glycoprotein (http://www.cbs.dtu.dk/services/NetCTL/) (Larsen et al., 2007). Three key parameters for CTL prediction are used by the NetCTL server: MHC-I binding peptide, proteasomal C-terminal cleavage, and transport efficiency TAP (Transporter Associated with Antigen Processing). Artificial neural network (ANN) was employed to predict MHC-I peptide binding and proteasomal C-terminal cleavage while the TAP transport efficiency was achieved by the weight matrix. The scores of all the three methods were integrated and the threshold is converted into sensitivity or specificity values. Threshold value is set as 0.75. NetMHCpan 4.0 server (http://www.cbs.dtu.dk/services/NetMHCpan/) (Jurtz et al., 2017) was used to perform binding affinity analysis. The top two highest scoring epitopes from the envelope protein and surface glycoprotein for each supertype were selected for binding affinity analysis. The default % ranks for strong binders is 0.5 and 2 for weak binders. In this study, we used a tool from the MHCcluster v2.0 (Thomsen et al., 2013) server that provides the functional alliance between MHC variants with pictorial tree-based visualizations and highly instinctive heat maps.

HLA-peptide docking analysis

The 3 D structure of predicted CTL epitopes was modelled using PEPFOLD (https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/) (Maupetit et al., 2009) online server. Molecular docking was performed for epitopes with their specific HLA alleles. HLA A1 (PDB: 1W72), HLA A2 (PDB: 3MRG), HLA A3 (2XPG), HLA B7 (5WMO), HLA B44 (1M6O), HLA B62 (1XR9). The PDB files of receptor and ligand was submitted in the ClusPro2.0 (https://cluspro.bu.edu/login.php) server (Kozakov et al., 2013; 2017; Vajda et al., 2017).

Helper T-lymphocyte (HTL) epitope prediction

HTL epitope prediction is crucial in the development of peptide-based vaccines because binding of peptide with MHC-II proteins is important in eliciting immune response. MHC-II binding alleles were predicted using IEDB (http://tools.iedb.org/mhcii/) (Nielsen & Lund, 2009). The MHC-II binding predictions were executed using the IEDB analysis resource Consensus tool. To evaluate the interaction potential of T-cell epitopes and MHC-II alleles, the Half Maximum Inhibitory Concentration (IC50) was set at less than 50 for strong binding peptides. IEDB recommended method was used to determine promising MHC-II epitopes with low IC50 value.

B cell epitope prediction

B-cells plays a crucial role in eliciting the humoral immune response and antibody production, which neutralizes the antigens during infection, thus B-cell epitopes have an important role in designing the vaccine. The sequential B-cell epitopes were predicted by means of ABCPred server (http://crdd.osdd.net/raghava/abcpred/) (Saha & Raghava, 2006). ABCPred uses ANN using fixed length patterns for prediction of B-cell epitopes with a default threshold of 0.5.

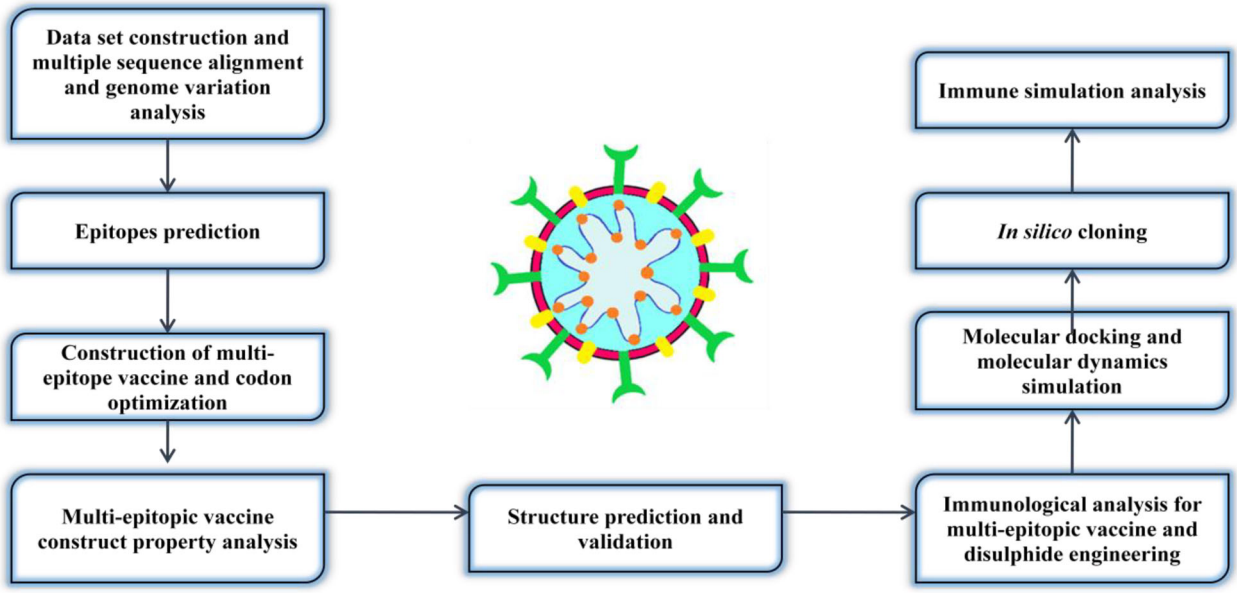

Figure 1. Steps involved in multi-epitopic vaccine design.

Toxicity prediction

The ToxinPred tool (https://webs.iiitd.edu.in/raghava/toxinpred/index.html) uses support vector machine method was employed for the prediction of toxicity of the predicted CTL, HTL and B-cell epitopes (Open Source Drug Discovery Consortium, 2013). The tool permits to categorize highly toxic and nontoxic short epitopic sequences besides with analysis of hydrophobicity, hydropathicity, hydrophilicity, charge, and molecular weight.

Construction of multi epitopic vaccine and codon optimization

Multi-epitopic vaccine was designed using B-Cell, HTL and CTL epitopes from the envelope and surface glycoprotein. To enhance the immunogenicity of the multi-epitope vaccine, β-defensin 2 was selected as an adjuvant. β-defensin 2 induce signal transduction through TLR-4 and regulate adaptive immunity during infection process (Amelin et al., 2002), β-defensin 2 also modifies adaptive immunity by eliminating activated antigen presenting cells (APC) through self-destructing signalling (Biragyn et al., 2008). EAAAK linker was used to link the adjuvant with B-cell epitopes, followed by HTL and CTL epitopes. The intra B-cell and intra HTL epitopes were linked using GPGPG linkers and intra CTL epitopes were linked by AAY linkers respectively. JCAT server (http://www.jcat.de) was used to perform codon optimization according to *Escherichia coli* (*E. coli*) codon usage (Grote et al., 2005). Several parameters that are crucial to effectiveness of gene expression including codon adaptation index (CAI) and GC content adjustment, both were optimized.

Population coverage analysis

The multi-epitopic vaccine construct with the predicted epitopes were submitted in IEDB population coverage analysis tool (http://tools.iedb.org/population/) to analyze population coverage within the global population. Epitopes and MHC restriction datas were submitted in the tool and prediction calculation were carried out for MHC I and MHC II specific epitopes (Bui et al., 2006).

Antigenicity, allergenicity and solubility profile prediction

The antigenicity of the multi-epitope vaccine was predicted using the VaxiJen server (http://www.ddgpharmfac.net/vaxijen/VaxiJen/VaxiJen.html) (Doytchinova & Flower, 2007) and ANTIGENpro (http://scratch.proteomics.ics.uci.edu/) (Magnan et al., 2010) the allergenicity of the vaccine construct was evaluated by AllerTOP v2.0 (http://www.ddg-pharmfac.net/AllerTOP/). In order to sort the allergens, the AllerTOP v2.0 server employs machine learning methods. The correctness of this server is 85.3% at five-fold cross validation (Dimitrov et al., 2013). In addition, we used the SOLpro online method to estimate solubility when the protein structure is over-expressed in *E. coli* with 74% reliable prediction (http://scratch.proteomnoics.ics.uci.edu/) (Cheng et al., 2005).

Determination of physiochemical properties

The ExPASy-ProtParam server, accessible at (http://web.expasy.org/protparam/), has calculated the physio-chemical properties of the vaccine construct. This Web tool calculates multiple parameters like aliphatic index, instability index, half-life, isoelectric point (pI), molecular weight and atomic composition, including grand average hydropathicity (GRAVY). The half-life of the protein represents the time that the molecule disappears after its synthesis in the cell. The index of instability indicates the consistency of a protein molecule *in vitro* (Gasteiger et al., 2003). For positive controls, SARS-CoV-2 structural protein based multi-epitopic vaccine and *in silico* predicted and *in vivo* tested studies (C1) (Di

Natale et al., 2020), (C2) (Kar et al., 2020), (C3) (Tahir Ul Qamar et al., 2020), (C4) (Safavi et al., 2019; 2021), (C5) (Safavi et al., 2019) and (C6) (Safavi et al., 2020) were selected for comparative evaluation of constructed multi-epitopic vaccine.

Secondary and tertiary structure prediction, refinement and validation

The secondary structure of the multi-epitope vaccine construct was calculated using Garnier-Osguthorpe-Robson (GOR IV) online server (https://npsa-prabi.ibcp.fr/cgi-bin/npsaautomat.pl?page=npsa_gor4.html) with mean accuracy of 64.4% (Garnier et al., 1996) and position specific iterated prediction (PSIPRED) analysis on outputs from PSI-BLAST (http://bioinf.cs.ucl.ac.uk/psipred/) (Buchan & Jones, 2019). The tertiary structure was modelled using I- TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Yang et al., 2015). The 3 D structure modelled by I- TASSER were subjected to refinement by GalaxyRefine server (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE) (Heo et al., 2013). The output consist of five refined models, with varying parameter scores, including GDT-HA, RMSD, MolProbity, Clash score, Poor rotamers and Rama favoured (Ko et al., 2012). The refined structure was validated by PROCHECK (https://servicesn.mbi.ucla.edu/PROCHECK/) checks the stereochemical value of a protein structure by investigating residue-by-residue geometry and overall structure geometry (Laskowski et al., 1993).

Immunological analysis for the multi-epitopic vaccine

The multi-epitopic vaccine was analyzed for linear (continuous) and conformational (discontinuous) B-cell epitope. Linear B-cell epitopes were predicted by the BcePred (https://webs.iiitd.edu.in/raghava/bcepred/index.html) (Saha & Raghava, 2004). Conformational B-cell epitopes were predicted by ElliPro, which predicts linear and conformational epitopes based on antigen's 3 D structure (http://tools.iedb.org/ellipro/) (Ponomarenko et al., 2008).

Vaccine protein disulfide engineering

Disulfide engineering is an important biotechnological tool for the design of new disulfide bonds in the target protein in a highly mobile protein region via cysteine residue mutation. Disulfide bonds have important stability and strengthen the geometric conformation of proteins. The online DbD2 server, available at http://cptweb.cpt.wayne.edu/DbD2/index.php, was used to this purpose. If each amino acid residue mutated to cysteine, this web server will detect the pair of residues capable of forming a disulfide bond (Craig & Dombkowski, 2013).

Molecular docking and molecular dynamics simulation

Molecular docking was performed for the constructed multi-epitope vaccine with TLR-2, TLR-3, TLR-4 and TLR-8 receptor (PDB ID: 2Z7X (TLR-2), 2A0Z (TLR-3), 4G8A (TLR-4) and 3W3M (TLR-8). The PDB files of receptor and ligand was submitted in the ClusPro2.0 (https://cluspro.bu.edu/login.php) server (Kozakov et al., 2013; 2017; Vajda et al., 2017) it utilizes the Fourier correlation algorithm and sort out the models with the amalgamation of desolvation and electrostatic energies. The iMODS server was used to perform the internal coordinates analysis based on the protein-protein structural complex (López-Blanco et al., 2014). Normal mode analysis mobility permitted us to explore the large scale mobility and the stability of macromolecules. The server calculates a specific combined motion of large macro-molecule along with the NMA of dihedral co-ordinates of Cα atoms. Furthermore, iMODS estimates B-factor (a dis-order of an atom in a protein), structural deformability, and computes the eigenvalue.

In silico cloning of vaccine construct

Assessment of cloning and expression of the vaccine construct in a suitable expression vector was achieved using *in silico* cloning tool called the SnapGene tool.

Immune simulation

Computational immune simulation was performed using C-IMMSIM (http://150.146.2.1/C-IMMSIM) online server to determine the immune response analysis of vaccine construct. Agent based algorithm was used for the estimation of antigen and foreign particles on immune activity. The simulation was performed with default parameters (Rapin et al., 2010).

Results

Multiple sequences alignment

The multiple sequence alignment of SARS-CoV-2 envelope protein and surface glycoprotein was performed to analyze the conserved region. The envelope protein sequences retrieved from various isolates (strains) were highly conserved (100%) when compared to SARS-CoV-2 reference sequence, whereas the amino acid sequences of surface glycoprotein of various sequences (strains) showed 99.9% similarity with SARS-CoV-2 reference sequence.

Phylogenetic analysis

Assembled SARS-CoV-2 genome sequence in FASTA format from the dataset was used for corona virus typing tool analysis. Evolutionary analysis was made with the reference SARS-CoV-2 envelope protein and surface glycoprotein genome sequences, a cladogram was generated showing that all the sequences comes under SARS-CoV-2 clade.

Genome variation analysis

To understand the variations in the envelope protein and surface glycoprotein of SARS-CoV-2 we compared SARS-CoV genome with the current SARS-CoV-2 genome. The SARS-

CoV-2 genome showed 93.5% similarity in nucleotide level, 94.8% similarity in protein level for envelope protein and 72.9% similarity in nucleotide level, 77% similarity in protein level with SARS-CoV for surface glycoprotein. Furthermore, we also observed mutations in surface glycoprotein from most of the selected sequence when compared with the reference sequence of SARS-CoV-2 (Supplementary Table S2), which denotes the rapid evolution of the virus.

Prediction of cytotoxic T-lymphocyte and analysis of binding affinity with MHC I and cluster analysis

CTL epitopes for SARS-CoV-2 envelope protein and surface glycoprotein were predicted using NetCTL 1.2 server, screening was based on the combined high score which denotes low sensitivity and high specificity for MHC-I super types. The binding affinity analysis was carried out using NetMHCpan 4.0 server, strong binders are defined as having percentage rank < 0.05. For each protein top 2 high scoring epitopes with least percentage rank in binding affinity were selected and used for peptide- MHC I binding analysis. A total of 11 epitopes were predicted, among them 5 epitopes were from envelope protein and 6 epitopes were from surface glycoprotein that was screened based on their binding affinity scores with their respective MHC I super types (Supplementary Table S3). MHCcluster v2.0 provided a cluster of 22 HLA molecules of class I that have been identified to interact with our predicted epitopes. The output was generated on the basis of sequence data available for various HLA-A and HLA-B alleles using the traditional phylogenetic method. The function-based clustering of HLA alleles (heat map) with red areas showing high correlation and the yellow zone showing lesser correlation is illustrated in Figure 2.

HLA- peptide binding analysis

The tertiary structure of predicted CTL epitopes was modelled by PEPFOLD and docking was performed by Cluspro server. The results are listed in Supplementary Table S4. The low weighed scores indicate efficient binding of epitopes with HLA alleles.

HTL epitope prediction

SARS-CoV-2 envelope protein and surface glycoprotein were analyzed for HTL epitopes using IEDB analysis resource using NN-align with half-maximal inhibitory concentration ($IC50 \leq 50$). Top epitopes were selected for both the proteins on the basis of least percentile rank and IC50 value less than 50 nM thereby indicating their higher affinity for receptor molecule (Supplementary Table S5).

B cell epitope prediction

The main objectives in epitope driven vaccine development is identification and prediction of B-cell epitopes in the target antigen. Latest reports suggested that B cell epitope play a crucial role in inducing immune response and inhibiting receptor binding and viral entry into the host cell (Du et al., 2009; Jiang et al., 2020; 2020; Zhou et al., 2019). The epitopes above the threshold score of 0.51 were selected. Among them high scoring epitopes from both the envelope and surface glycoprotein's were selected for vaccine design. All the predicted epitopes falls under conserved region (Supplementary Table S6).

Toxicity prediction

The predicted CTL, HTL and B- cell epitopes were tested for toxicity by ToxinPred online tool and the results revealed that all the selected epitopes for design of multi-epitopic vaccine are non- toxin (Supplementary Tables S3, S5, and S6).

Multi-epitopic vaccine design

The screened B-cell, HTL and CTL epitopes was used to design the multi-epitope vaccine construct. β-defensin 2 serves as an adjuvant which supports the expression of immune stimulators and antiviral compounds (Kim et al., 2018); thus β-defensin 2 sequences was added as an adjuvant at N- terminal of the vaccine construct linked by EAAAK linker followed by B-cell epitopes and linked GPGPG linkers, further the HTL and CTL epitopes were linked by AAY linkers. The final vaccine construct was 270 amino acids long (Figure 3).

Population coverage analysis

The evaluation of MHC I and MHC II based population coverage for the multi-epitope vaccine construct was performed by IEDB for global population (Figure 4, Supplementary Table S7). 91.38% of the world population would be covered with the administration of designed vaccine construct. The population coverage rate exceeded 75% in most of the global regions.

Antigenicity, allergenicity and solubility prediction of multi-epitopic vaccine

The main criteria that has to be ensured while vaccine designing is the antigenicity to persuade a humoral and/or cell-mediated immune response against the target microbe and allergenicity of the constructed vaccine. The antigenic score for multi-epitope vaccine is 0.6809 predicted by VaxiJen v2.0 server. The vaccine was non-allergenic predicted using AllerTOP v2.0. The predicted solubility leading over expression by SOLpro server showed the vaccine construct as soluble with probability 0.7561

Physiochemical characterization of designed vaccine

The physiochemical parameters of the vaccine construct has crucial effects on properties like stability and immunogenicity (Dey et al., 2014). The designed multi-epitope vaccine is

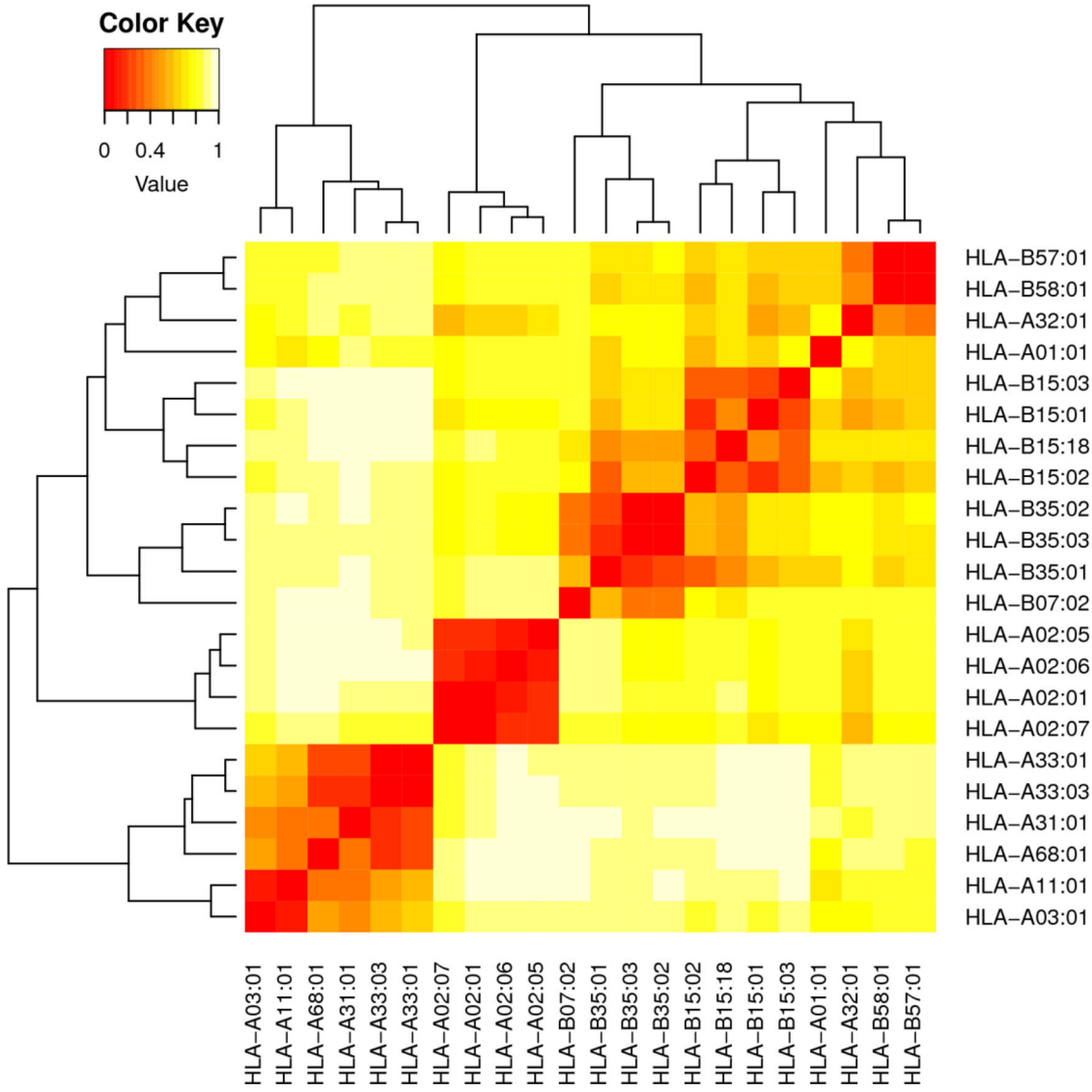

Figure 2. Heat map of MHC-I cluster analysis.

composed of 270 amino acids with a molecular weight of approximately 29.00 kDa. The theoretical pI was 9.54, indicating the vaccine construct is notably basic in nature. The total number of negative residues in the vaccine construct was predicted to be 9 and positive residue is 27. The molecular formula of the vaccine construct is C_{1316} H_{2035} N_{349} O_{365} S_{13} with 4078 atoms. The estimated half-life is 30 h, >20 h, > 10 h in mammalian reticulocytes (*in vitro*), yeast (*in vivo*) and *Escherichia coli* (*in vivo*) respectively. The instability index was 35.42, implying a stable protein. The thermal stability of the vaccine is important and it is evaluated by aliphatic index (Brandau et al., 2003), further solubility is a significant physiochemical parameter for the expression of protein in suitable expression system (Saha & Raghava, 2004). The calculated aliphatic index and grand average of hydropathicity (GRAVY) was 85.89 and 0.169 respectively, signifying that the vaccine is thermostable and hydrophilic. The final multi-epitopic vaccine construct's physiochemical properties were compared with positive controls (Table 1), the positive controls aids in validation of the predicted results (Rawi et al., 2018). The results exhibited high solubility, stability and bioavailability.

Secondary and tertiary structure prediction

SOPMA produced the secondary structure of the vaccine construct with 45.93% Alpha Helix, 17.41% Extended Strand, 9.26% Beta Turn and 27.41% Random Coil. The results were validated by PSIPRED (Supplementary Figures S1 and S2). The tertiary structure of the vaccine was modelled using I-TASSER server. The selected model has a probable TM-score of 0.38 ± 0.13 with a probable RMSD of 13.0 ± 4.2 Å. The TM-score is a measure of structural similarity between two structures. The C-score is −2.96. A TM-score > 0.5 indicates a model of correct topology, and a TM-score < 0.17 means random similarity. These cut-off values are independent of protein length.

Structure validation

The 3 D model was subjected to further refinement by GalaxyRefine server; the output showed 89.9% residues were present in the favoured region for vaccine construct (Figure 5). Further validation of the model suggested that the

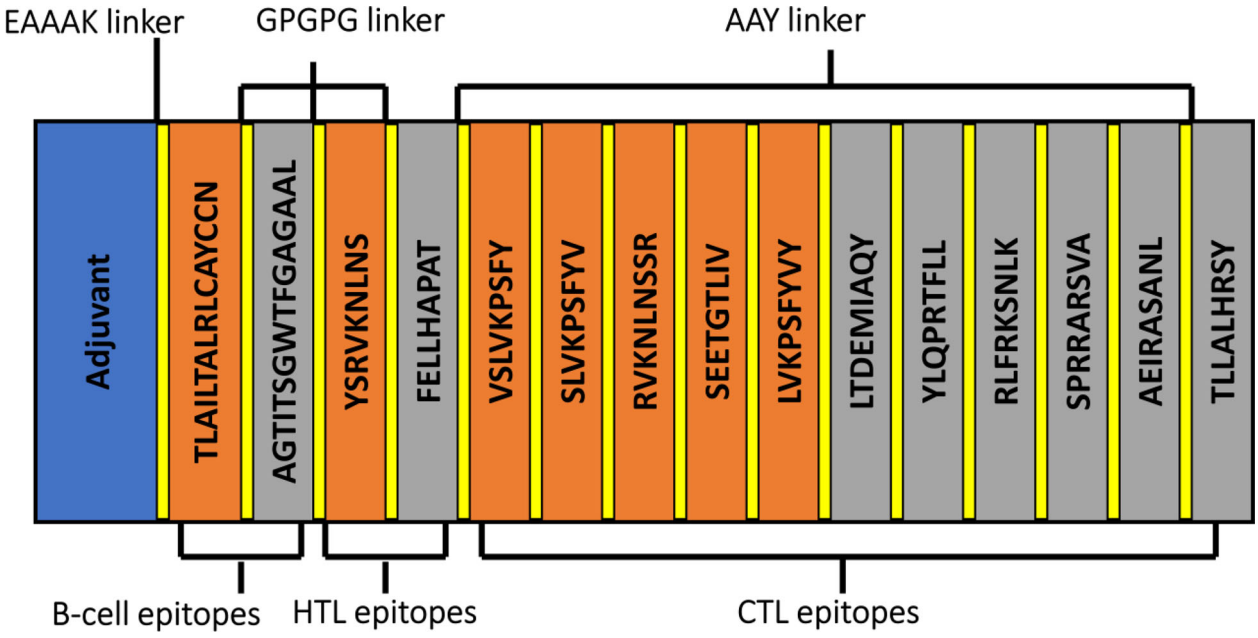

Figure 3. Schematic representation of final multi-epitopic vaccine construct. The 270- amino acid long peptide containing β-defensin adjuvant at N-terminal followed B cell, HTL and CTL epitopes. Depicted in orange are epitopes from envelope protein and grey are epitopes from surface glycoprotein.

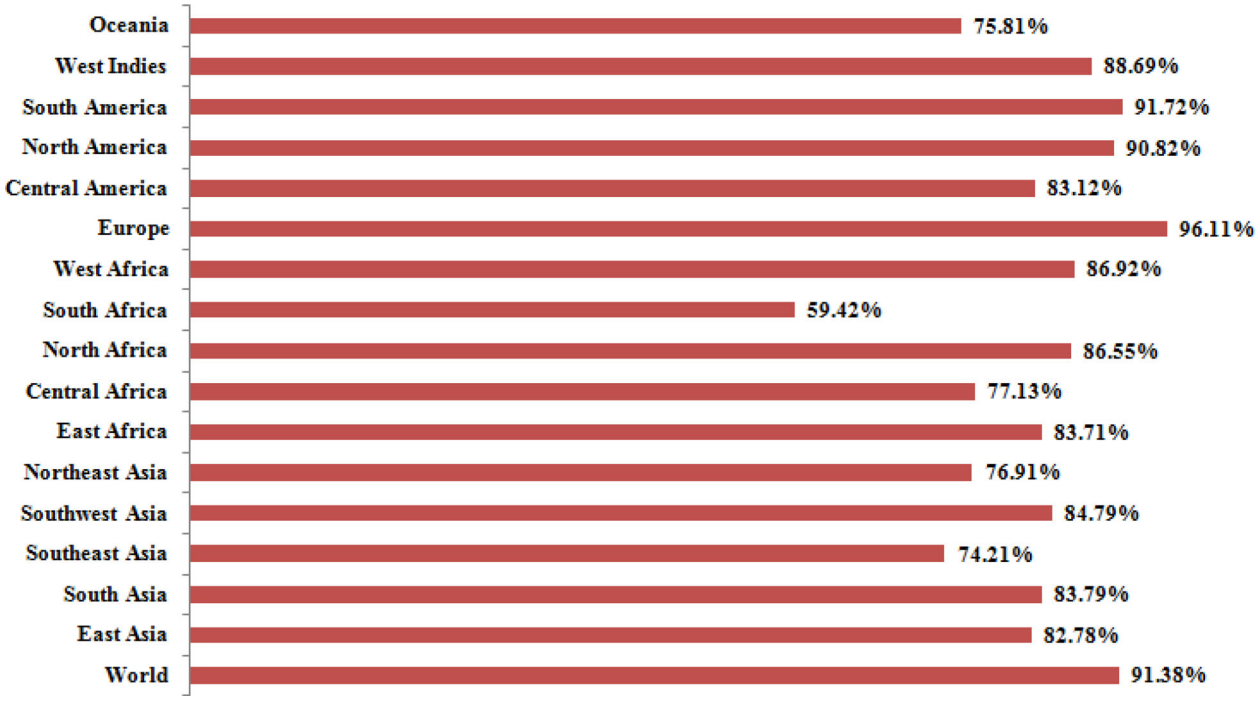

Figure 4. Population coverage analysis of the final multi-epitope vaccine construct across world as predicted by the population coverage analysis tool of the IEDB database.

Table 1. Comparative study of physiochemical properties of vaccine positive controls (C1, C2, C3, C4, C5 and C6) and SARS-CoV-2 multi-epitopic vaccine candidate.

		Value/ score						
Properties	Parameters/tools	C1	C2	C3	C4	C5	C6	Multi-epitopic vaccine construct
Physiochemical properties	Molecular weight	45.13 kDa	44.15 kDa	35.17 kDa	6.38 kDa	39.83 kDa	51.64 kDa	29.0 kDa
	Isoelectric point (pI)	9.01	9.96	10.31	9.61	9.61	10	9.54
	Instability index (II)	23.46	31.34	–	33.38	30.03	27.09	35.42
	Aliphatic index (AI)	–	78.74	–	53.53	84.57	79	85.89
	GRAVY	–	−0.088	0.395	−1.127	−0.215	−0.354	0.169

refined model has high stability and good quality, the refined model was used in docking studies.

Immunological properties assessment

The BcePred server was used for prediction of continuous B-cell epitopes in chimeric multi-epitope vaccine with default parameters. Important properties of epitopes like hydrophilicity, antigenicity, flexibility, accessibility, polarity and exposed surfaces were predicted (Table 2). Ellipro was used to predict the conformational B cell epitopes in multi-epitope vaccine. In total seven conformational epitopes were predicted (Table 3, Figure 6).

Vaccine protein disulfide engineering

A total number of 23 pairs of amino acid residues were predicted having potential to form disulfide bond by DbD2 server. Following residue evaluation by chi3 and B-factor energy parameters, only one residue was found to be suitable for disulfide bond formation 194 SER-199 ALA which was replaced by cysteine residue (Figure 7). Residue screening was done on the basis of $-$ 87 to $+$97 chi3 value and $<$ 2.5 energy value.

Molecular docking analysis and molecular dynamics simulation

The binding energy between the vaccine construct and the TLR-2, TLR- 3, TLR-4 and TLR-8 receptors (PDB ID: 2Z7X (TLR-2), 2A0Z (TLR-3), 4G8A (TLR-4) and 3W3M (TLR-8)) was analyzed using the ClusPro server. The model with lowest binding energy value was preferred for each docked complex (Table 4) (Figure 8). Normal mode analysis (NMA) was executed to evaluate the stability of docked complex and their large scale mobility. According to the normal mode analysis mobility, upon binding, the vaccine construct and the TLR-2, TLR-3, TLR-4 and TLR-8 receptors were notably directed to each other (Figure 9a). The deformability of the complex was associated with distortion of individual residues, representing hinges in the chain (Figure 9b). The B-factor values inferred by NMA is analogous to RMS, B-factor was greatly reduced from its PDB B-factor (Figure 9c). The eigenvalue showed a contrary relationship with the variance of the protein-protein docking complex, and the estimated eigenvalue for vaccine construct and TLR-2, TLR-3, TLR-4 and TLR-8 receptors were $6.482268e^{-05}$, $5.516701e^{-05}$, $1.145536e^{-05}$ and $1.094214e^{-04}$ respectively (Figure 9d). The covariance matrix is illustrated through the graphical representation using white, red and blue colour variations representing the correlated, uncorrelated, and anti-correlated pairs of amino acid residues, respectively (Figure 9e). Springs of atomic contact are plotted as grey dots in the elastic network model, the darker grey represent stiffer springs and vice versa (Figure 9f).

Codon optimization and in silico cloning

Reverse translation and codon optimization was performed using JCAT server. The CAI in the adapted sequence was 1.0 and GC content of the improved sequence was 53.70% wherein the GC-Content of *E. coli* was 50.7. To perform *in silico* cloning the codon optimized vaccine construct sequence were inspected for restriction enzymes sites, HindIII and BamHI enzymes were not found in the vaccine sequence so these enzymes were used for *in silico* cloning purpose. Lastly, a successful clone of 6,160 bp was obtained following insertion of the vaccine construct into pET28a(+) vector (Figure 10).

Immune simulation

Results from the C-ImmSim server showed a significant increase in immune response generation. High levels of IgM and IgG were signs of humoral immune response (Figure 11a). An increase in B memory cells and B cell isotype IgM with a corresponding reduction in antigen concentration was found in the B cell population (Figure 11b). In addition, the TH1 and TC cell populations with corresponding memory growth, a high response was observed (Figure 11c and d). The development of immune memory increased antigen clearance on subsequent exposures. In addition, these findings were compatible with the induced level of IFN-c developed after the proposed vaccine construct was immunized (Figure 8e).

Discussion

Vaccinations are the most effectual way to rapidly improve public health and the best way to control the ongoing pandemic. The epitope identification and prediction using computational biology and bioinformatics started almost three decades back by Rammensee et al., by identifying binding motifs of T-cell antigens (Falk et al., 1991), from then on numerous vaccine designs aiming for various infectious pathogens were predicted using *in silico* approaches. The advantages of this approach aids in rapid detection of antigenic motifs from immune response eliciting proteins of the pathogens, which is current need of the hour.

Current research is focused on development of epitope-based peptide vaccine of envelope and surface glycoprotein from SARS-CoV-2. With rapidly available genomics and proteomics data for SARS-CoV-2 it is now possible to design an epitopic peptide vaccine. This vaccinomics approach has previously supported in defending multiple sclerosis (Bourdette et al., 2005), tumours (Safavi et al., 2019; Kalli et al., 2018), and malaria (López et al., 2001). Many of the prophylactic vaccines against viral diseases targets B-cell immune response to produce neutralizing antibodies. Whereas, the therapeutic vaccines are targeted to evoke cell mediated immune response (Garbuglia et al., 2020). The novelty of the designed vaccine in this study is predicted to have elevated ability to produce neutralizing antibody production against envelope protein and surface glycoprotein of SARS-CoV-2. Addition to that the presence of multiple T-cell epitopes in the vaccine construct can evoke cell-mediated immune response therefore making this vaccine construct suitable for both prophylactic and therapeutic purposes. At present few studies focused on using *in silico* approach for designing multi-epitopic vaccine candidate against SARS-CoV-2,

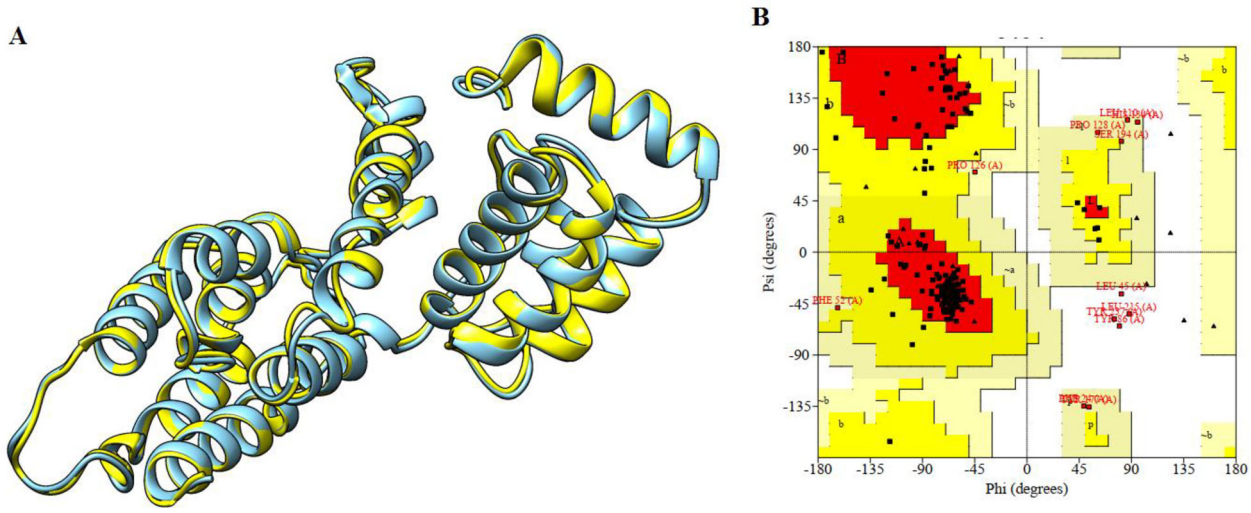

Figure 5. Tertiary structure model and its validation. (A) 3 D structural alignment of the modeled vaccine construct before (yellow) and after (cyan) refinement. (B) Ramachandran plot for the vaccine construct.

Table 2. Prediction of linear B-cell epitopes in chimeric multi-epitope vaccine by different parameters based on BcePred server.

Prediction parameters	Epitope position
Hydrophilicity	126, 179–180,
Flexibility	38–39, 120, 123–124, 168–169, 178, 229, 238–239
Accessibility	32–34, 62–63, 71, 115, 117–121, 123, 166, 168–169, 171, 174, 176–177, 180, 215–217, 227–229, 231–233, 237–243
Exposed surface	232–233
Polarity	31–34, 62–64, 66, 177, 227–229, 231–232, 240–243
Antigenic Propensity	5–15, 35–38, 52, 144–147, 156–159, 192–195

Table 3. Predicted conformational B cell epitopes.

S.No	Residues	Number of residues	Score
1	G113, P114, Y116, S117, R118, V119, K120, N121, L122, N123, S124, G125, P126, G127, P128, G129, F130, E131, L132, L133, H134	21	0.783
2	L184, Y189, A212, Y213, Y214, L215, Q216, P217, R218, T219, F220, L221, L222, A223, A224, Y225, R226, L227, F228, K230, S231, R243, S244, A246, A247, A248, Y249, A250, E251, I252, R253, A254, S255, A256, N257, L258, A259, A260, Y261, T262, L263, L264, A265, L266, H267, S269, Y270	47	0.736
3	T3, S4, L6, L7, T10, G22, N23, F24, L25, T26, G27, L28, G29, H30, R31, S32, D33, N36, S40, Q43, C44, L45, Y46, S47, A48, C49, P50, I51, F52, T53, K54, I55, Q56, G57, T58, R61	36	0.736
4	M1, R2, Y5, N89, G90, P91, G92, P93, G94, A95, G96	11	0.687
5	K65, K68, E69, A71, A72, K73, T74	7	0.641
6	S154, L155, V156, K157, P158	5	0.54
7	R229, N232, K234, A235, A236, Y237	6	0.501

epitopes from spike protein (Kar et al., 2020; Samad et al., 2020), and epitopes from four structural proteins (Tahir Ul Qamar et al., 2020) were used; Recently, Safavi et al. designed a multi-epitopic vaccine using the combination of epitopes from six non-structural proteins and functional region from spike protein (Safavi et al., 2020). The immunogenic epitopes from envelope protein and surface glycoprotein, both are crucial in viral pathogenesis were used to design the multi-epitopic vaccine construct which makes the current study unique.

Firstly, to study the mutations in the selected proteins we selected 50 sequences of envelope and surface glycoproteins from various geographical locations through NCBI database. Genome variation analysis and phylogenetic analysis were carried out by Genome Detective Corona Virus Typing Tool. After the conservation and mutational analysis both the proteins were further used to carry out detailed analysis with the aim to identify antigenic epitopes. B-cells play a crucial role in inducing specific immunogenic response following a pathogen attack. Likewise, T-cell focuses on cellular immune response by performing pathogen clearance and autoimmune reactions. Specific T cell and B cell epitopes were predicted and screened for toxicity from both the proteins. Population coverage analysis for the selected epitopes was analyzed and the final vaccine construct can cover approximately 91.3% of world population. The B-cell, HTL and CTL epitopes screened and selected were linked together with appropriate inter and intra epitope linkers. Immune adjuvants play a critical role in triggering host immune response. The most commonly used adjuvants against SARS-CoV-2 vaccine design is Cholera Toxin B (CTB) and Human β-defensin 2, the later has reported to be more specific and immunogenic thus it was added to the N-terminal of multi-epitope vaccine construct to induce the expression of primary immune stimulatory and antiviral compounds (Kim et al., 2018; Rahmani et al., 2021). Adjuvant helps in induction of effective and persistent immune response, thus essential to reduce the quantity of antigen and number of injections (Guy, 2007). The antigenicity, allergenicity and solubility profile of the vaccine construct was analyzed and the results revealed that multi-epitope vaccine is soluble and can induce vigorous immune response without any allergic reaction. The molecular weight of multi-epitopic vaccine construct estimated through physiochemical analysis indicated the good antigenic nature of the vaccine construct (Goodwin et al., 2012). The theoretical pI indicated the vaccine construct is basic. The calculated aliphatic index denotes the thermostable nature and a low GRAVY score signified that the

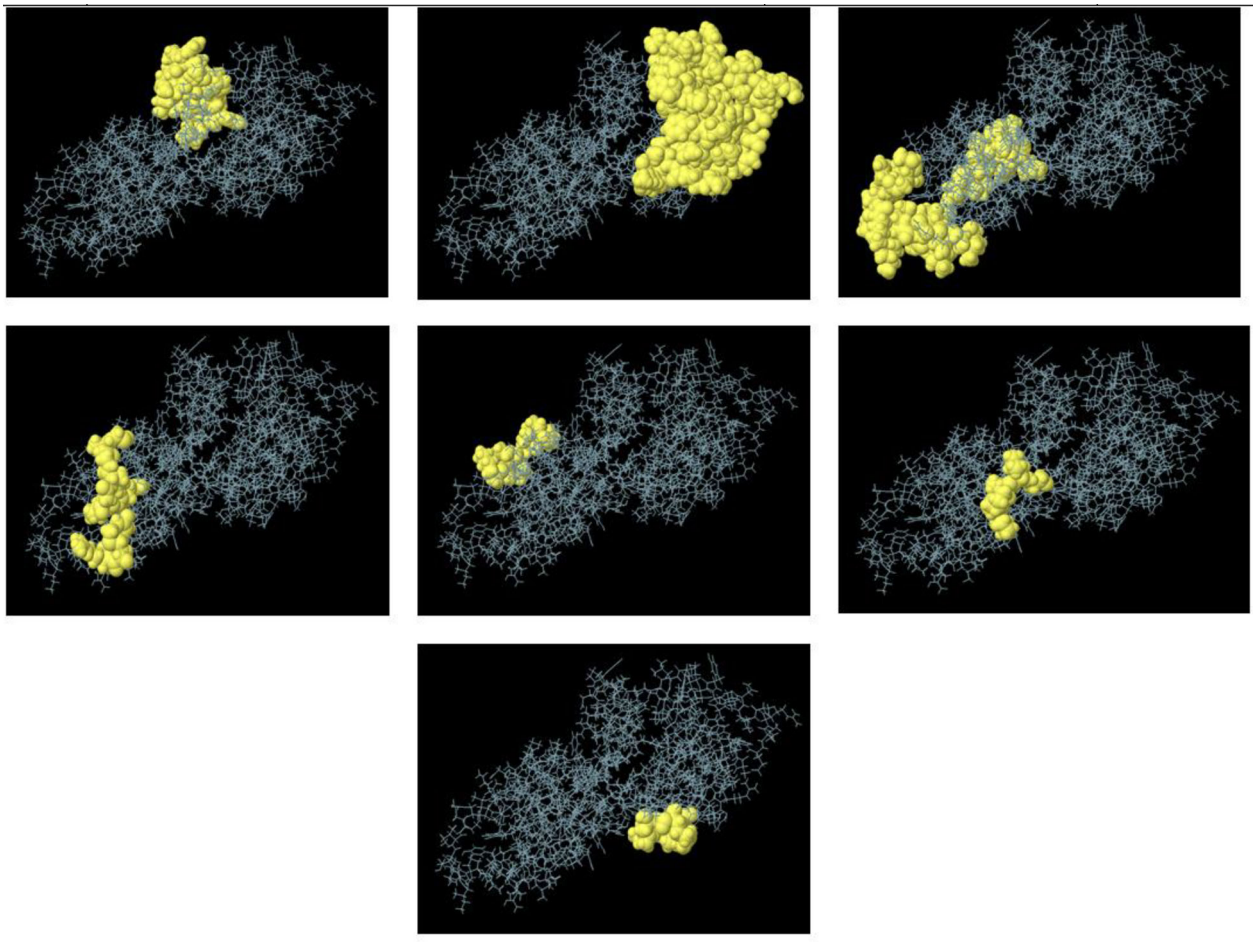

Figure 6. Conformational B cell epitopes in 3 D structure of vaccine construct by ElliPro tool. The immunogenic epitopes are depicted as yellow globules on ball and stick representation of multi-epitopic vaccine construct structure.

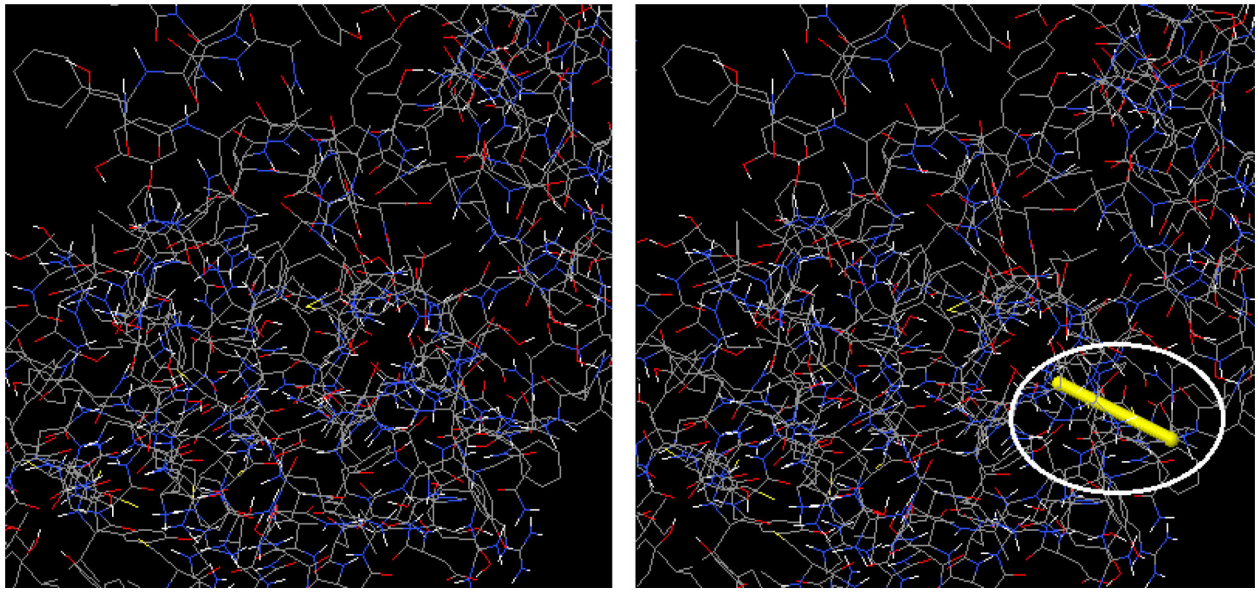

Figure 7. Disulfide engineering of the vaccine construct. (A) Initial model without disulphide bonds, (B) Mutant model the yellow stick, within the circle represents the disulphide bond formation.

vaccine construct is hydrophilic (Validi et al., 2018). The importance of secondary structure prediction of the designed vaccine construct is to determine its stability. The secondary structure of the multi-epitope vaccine construct consists of larger number of amino acids in alpha Helix, followed by extended strand, random coil and beta turn; this shows that the designed vaccine construct is structurally stable. The tertiary structure model was selected based on TM-score, RMSD

and C-score, higher the C-score value better the structure. The tertiary structure was subjected to further refinement; followed by refinement the advantageous feature of multi-epitopic vaccine structure have considerably enhanced. The Ramachandran plot analysis demonstrates majority of the residues were present in the favoured region and additionally allowed region and very few residues in disallowed region indicating the model to be with good overall quality. The hydrophilicity, antigenicity, flexibility, accessibility, polarity, exposed surfaces and conformational B cell epitopes of multi-epitope vaccine construct were predicted. The results disclosed that it can probably interact with antibodies with seven conformational epitopes and are flexible. During this study disulphide bridge was introduced in the vaccine construct to stabilize the vaccine construct, only one residue was found to be suitable for disulfide bond formation which was replaced by cysteine residue. It is necessary to understand the interaction of the vaccine construct with Toll like receptors (TLR) because they are sensors that stimulate the immunity of the natural host through the pathogen-associated molecular pattern (PAMP) and the TLR signalling pathway plays an essential role in host immune response (Steven et al., 2016). This pathway has been recognized as a drug target for many antibacterial or antiviral drugs during the development (Chakraborty et al., 2020). The TLRs (2, 3, 4, and 8) and vaccine construct interaction analysis was performed by ClusPro server, and the best docked model for each complex was selected based on the least binding energy. Molecular dynamics study was carried out to determine the docked complex stability. Structural dynamics were previously studied using atom subsets and covariance analysis and the stability of macromolecules was associated with correlated fluctuations of atoms (Aalten et al., 1997; Caspar, 1995; Clarage et al., 1995). iMODS server was used to determine the stability and probable immune interactions between immune receptors and *in silico* designed vaccines against SARS-CoV-2 (Ashfaq et al., 2021; Bhattacharya et al., 2020). iMODS server was employed to study the essential dynamics of vaccine construct and TLR receptors. The analysis showed slight probability of deformability for all

Table 4. Docking statistics of best docked receptor and vaccine construct.

S. No	Complex	Cluster size	Lowest Binding energy kJ mol^{-1}
1.	VC-TLR2	34	−1212.4
2.	VC-TLR3	28	−1187.7
3.	VC-TLR4	25	−1459.9
4.	VC-TLR8	39	−1488.6

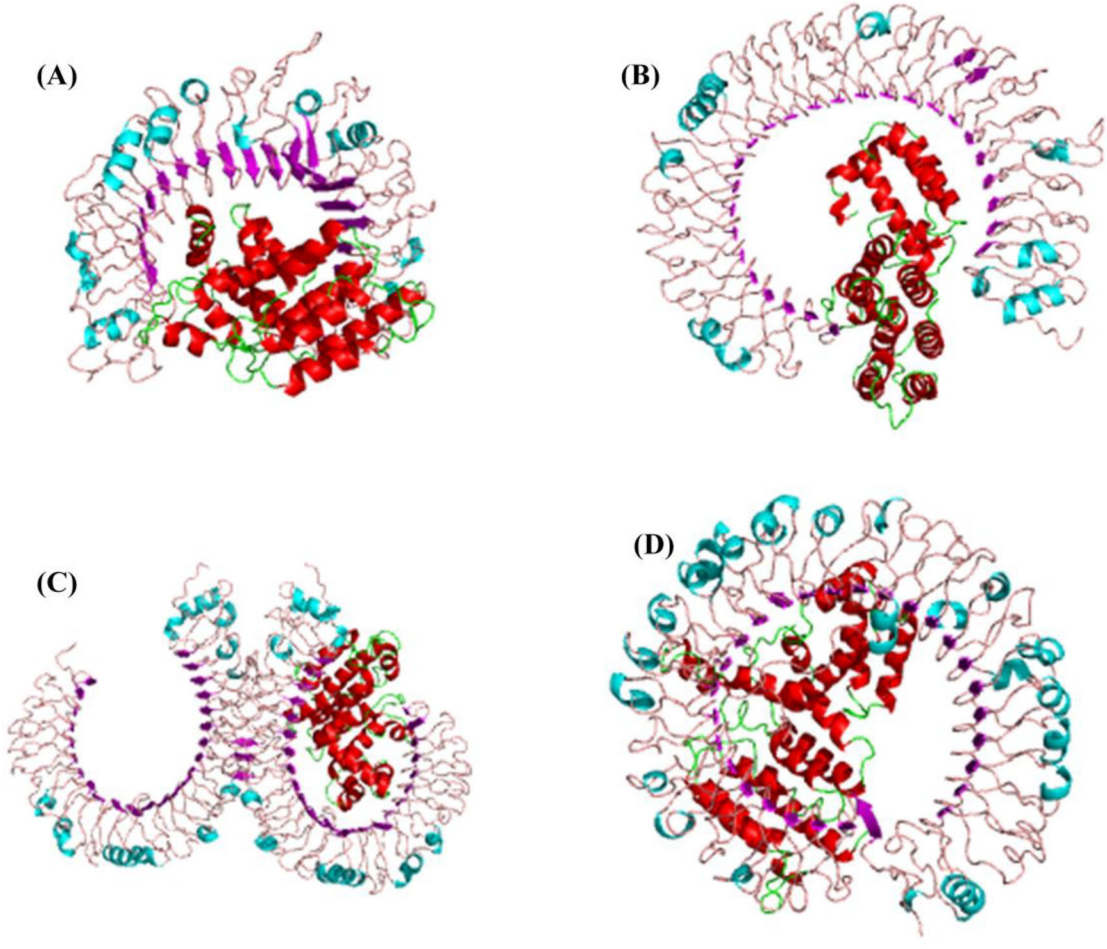

Figure 8. Figure showing (A) TLR-2 and vaccine construct docked complex, (B) TLR-3 and vaccine construct docked complex, (C) TLR-4 and vaccine construct docked complex, (D) TLR-8 and vaccine construct docked complex. Vaccine construct is shown in red colour while TLR receptors are shown in blue colour.

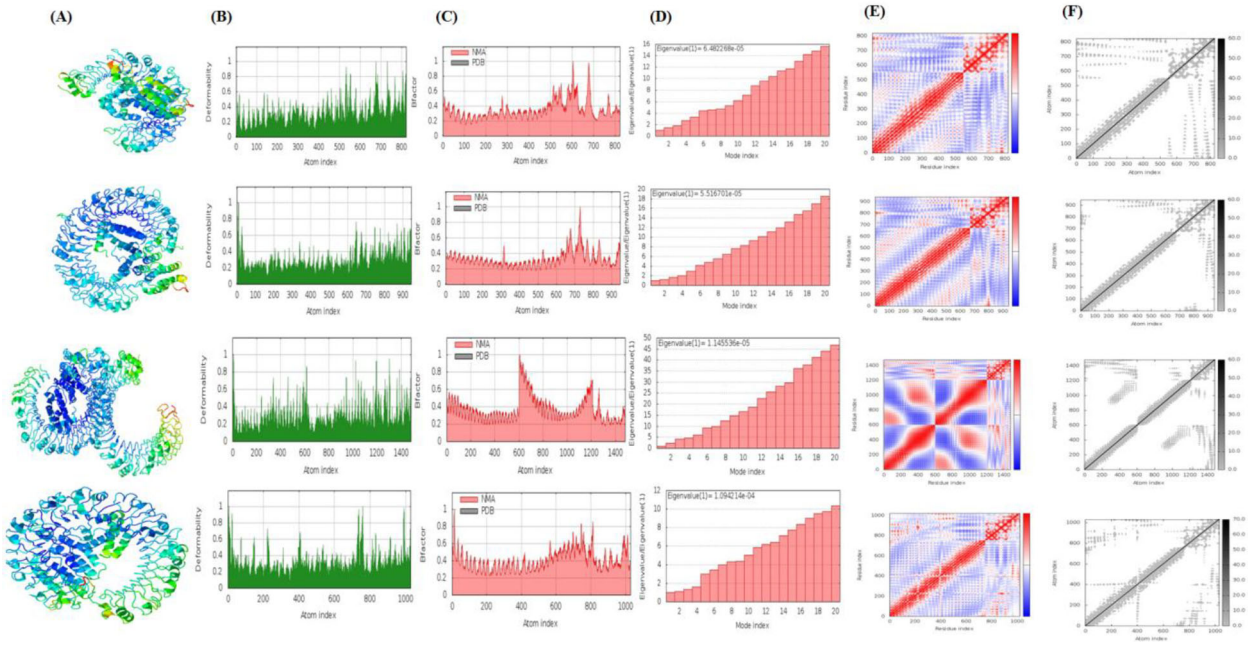

Figure 9. NMA analysis (a) NMA mobility of docked complex (b) deformability plot of atomic fluctuation (c) B-factor plot of PDB and NMA (d) eigenvalue plot (e) covariance matrix plot (f) elastic network plot.

Figure 10. *In silico* restriction cloning of the gene sequence of the vaccine construct into pET28a(+) expression vector.

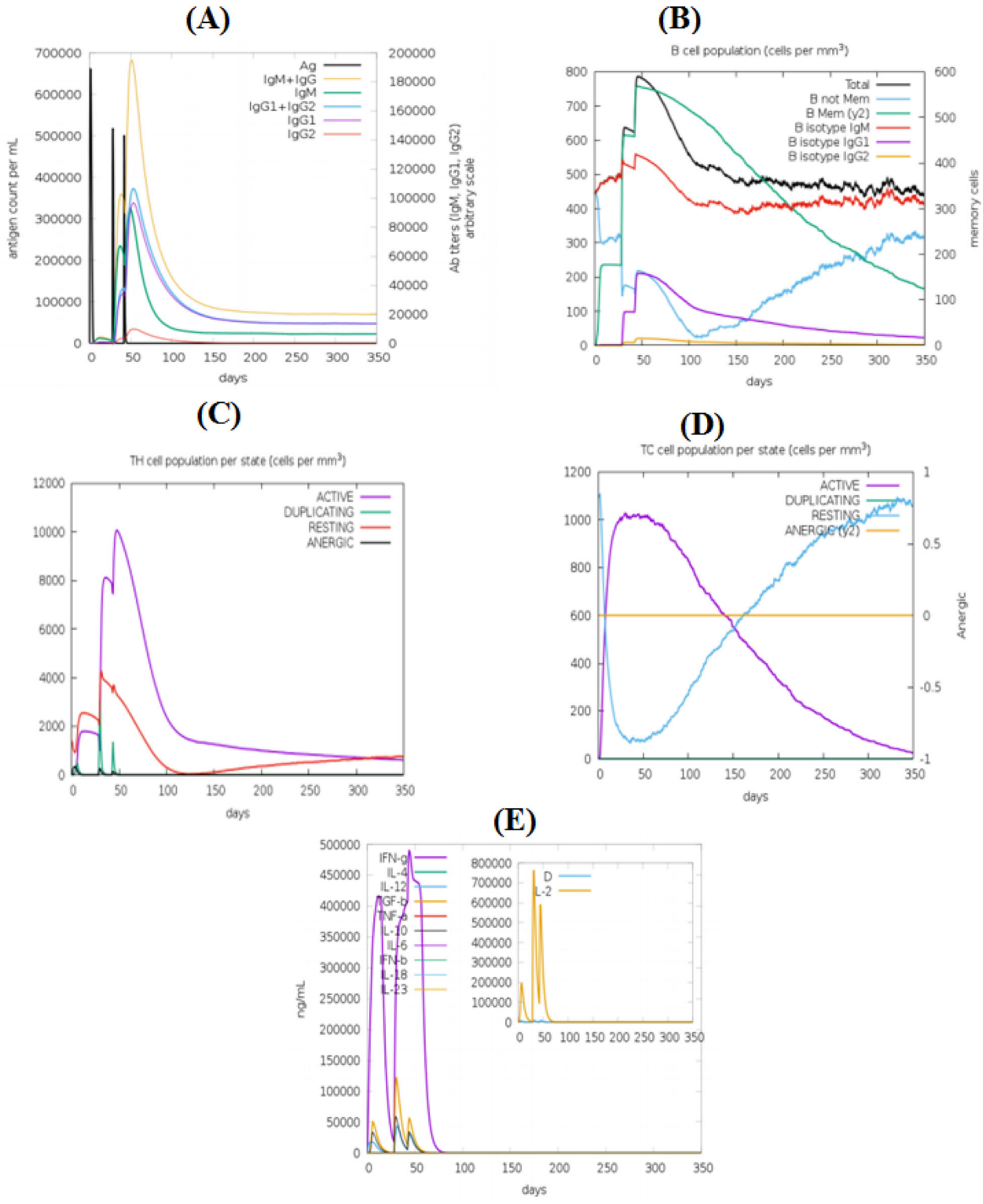

Figure 11. *In silico* immune simulation of vaccine construct. (A) Immunoglobulin production in response to antigen, (B) B-cell populations following exposure to vaccine construct, (C) The progress of T-helper and, (D) T-cytotoxic cell populations per state subsequent to vaccine construct injection. (E) Level of cytokines induced by the vaccine.

individual residues, as location of hinges in the chain was not major and thus validating our prediction. Lastly, the designed vaccine construct was reverse transcribed and codon optimized for *E. coli* strain K12 preceding to insertion within pET28a(+) vector for cloning and expression. The immune simulation results showed the development of immune memory and consequently increased antigen clearance on subsequent exposures. In addition, these findings were compatible with the induced level of IFN-c developed after the proposed vaccine construct was immunized. Our predicted *in silico* results, though, were determined by various computational analysis and using different immune databases, for experimental validation of our proposed vaccine candidates, we suggest further *in vivo* based studies

involving model animals. The designed multi-epitopic vaccine candidate should be coupled with credible delivery system to undergo *in vivo* studies. Antigen delivery systems like lipid particles, nanoparticles and micro particles can be used as carriers that aids in presentation of antigens to immune system in a most favourable mode (O'Hagan et al., 2001; Reed et al., 2009).

Conclusion

Quick identification of antigenic epitopes is of immense significance at the time of emerging pandemic. In the current research, we endeavoured to identify the important CTL epitopes *VSLVKPSFY, SLVKPSFYV, RVKNLNSSR, SEETGTLIV, LVKPSFYVY, LTDEMIAQY, YLQPRTFLL, RLFRKSNLK, SPRRARSVA, AEIRASANL, TLLALHRSY,* HTL epitopes *YSRVKNLNS* and *FELLHAPAT* and B-cells epitopes *TLAILTALRLCAYCCN* and *AGTITSGWTFGAGAAL* from the conserved regions of envelope protein and surface glycoprotein of SARS-CoV-2. The multi-epitopic vaccine candidate with high scoring CTL, HTL and B-cell epitopes of envelope protein and surface glycoprotein could be effective in stimulating both cellular and humoral immunity. Taken all together, according to immunological analysis, structural and physiochemical characterizations the multi-epitope construct can be a potential vaccine candidate.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

M. C. Harish http://orcid.org/0000-0001-5501-1047

References

Aalten, D. M. F., Groot, B. L., Findlay, J. B. C., Berendsen, H. J. C., & Amadei, A. A. (1997). Comparison of techniques for calculating protein essential dynamics. *Journal of Computational Chemistry*, *18*(2), 169–181. https://doi.org/10.1002/(SICI)1096-987X(19970130)18:2<169::AID-JCC3>3.0.CO;2-T

Amelin, Y., Krot, A. N., Hutcheon, I. D., & Ulyanov, A. A. (2002). Lead isotopic ages of chondrules and calcium-aluminum-rich inclusions. *Science (New York, N.Y.)*, *297*(5587), 1678–1683. https://doi.org/10.1126/science.1073950

Ashfaq, U. A., Saleem, S., Masoud, M. S., Ahmad, M., Nahid, N., Bhatti, R., Almatroudi, A., & Khurshid, M. (2021). Rational design of multi epitope-based subunit vaccine by exploring MERS-COV proteome: Reverse vaccinology and molecular docking approach. *PLoS One*, *16*(2), e0245072. https://doi.org/10.1371/journal.pone.0245072

Ashour, H. M., Elkhatib, W. F., Rahman, M. M., & Elshabrawy, H. A. (2020). Insights into the recent 2019 Novel Coronavirus (SARS-CoV-2) in light of past human coronavirus outbreaks. *Pathogens (Basel, Switzerland)*, *9*(3), 186. https://doi.org/10.3390/pathogens9030186

Bhattacharya, M., Sharma, A. R., Patra, P., Ghosh, P., Sharma, G., Patra, B. C., Saha, R. P., Lee, S. S., & Chakraborty, C. (2020). A SARS-CoV-2 vaccine candidate: *In-silico* cloning and validation. *Informatics in Medicine Unlocked*, *20*, 100394. https://doi.org/10.1016/j.imu.2020.100394

Biragyn, A., Coscia, M., Nagashima, K., Sanford, M., Young, H. A., & Olkhanud, P. (2008). Murine beta-defensin 2 promotes TLR-4/MyD88-mediated and NF-kappaB-dependent atypical death of APCs via activation of TNFR2. *Journal of Leukocyte Biology*, *83*(4), 998–1008. https://doi.org/10.1189/jlb.1007700

Biswas, N. K., & Majumder, P. P. (2020). Analysis of RNA sequences of 3636 SARS-CoV-2 collected from 55 countries reveals selective sweep of one virus type. *The Indian Journal of Medical Research*, *151*(5), 450–458. https://doi.org/10.4103/ijmr.IJMR_1125_20

Bourdette, D. N., Edmonds, E., Smith, C., Bowen, J. D., Guttmann, C. R., Nagy, Z. P., Simon, J., Whitham, R., Lovera, J., Yadav, V., Mass, M., Spencer, L., Culbertson, N., Bartholomew, R. M., Theofan, G., Milano, J., Offner, H., & Vandenbark, A. A. (2005). A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, *11*(5), 552–561. https://doi.org/10.1191/1352458505ms1225oa

Brandau, D. T., Jones, L. S., Wiethoff, C. M., Rexroad, J., & Middaugh, C. R. (2003). Thermal stability of vaccines. *Journal of Pharmaceutical Sciences*, *92*(2), 218–231. https://doi.org/10.1002/jps.10296

Buchan, D., & Jones, D. T. (2019). The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Research*, *47*(W1), W402–W407. https://doi.org/10.1093/nar/gkz297

Bui, H. H., Sidney, J., Dinh, K., Southwood, S., Newman, M. J., & Sette, A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*, *7*, 153. https://doi.org/10.1186/1471-2105-7-153

Caspar, D. L. (1995). Problems in simulating macromolecular movements. *Structure (Structure)*, *3*(4), 327–329. https://doi.org/10.1016/S0969-2126(01)00163-0

Chakraborty, C., Sharma, A. R., Sharma, G., Bhattacharya, M., Saha, R. P., & Lee, S. S. (2020). Extensive partnership, collaboration, and teamwork is required to stop the COVID-19 outbreak. *Archives of Medical Research*, *51*(7), 728–730. https://doi.org/10.1016/j.arcmed.2020.05.021

Cheng, J., Randall, A. Z., Sweredoski, M. J., & Baldi, P. (2005). SCRATCH: A protein structure and structural feature prediction server. *Nucleic Acids Research*, *33*(Web Server issue), W72–W76. https://doi.org/10.1093/nar/gki396

Clarage, J. B., Romo, T., Andrews, B. K., Pettitt, B. M., & Phillips, G. N. Jr (1995). A sampling problem in molecular dynamics simulations of macromolecules. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(8), 3288–3292. https://doi.org/10.1073/pnas.92.8.3288

Cleemput, S., Dumon, W., Fonseca, V., Abdool Karim, W., Giovanetti, M., Alcantara, L. C., Deforche, K., & de Oliveira, T. (2020). Genome detective coronavirus typing tool for rapid identification and characterization of novel coronavirus genomes. *Bioinformatics (Oxford, England)*, *36*(11), 3552–3555. https://doi.org/10.1093/bioinformatics/btaa145

Craig, D. B., & Dombkowski, A. A. (2013). Disulfide by Design 2.0: A web-based tool for disulfide engineering in proteins. *BMC Bioinformatics*, *14*, 346 https://doi.org/10.1186/1471-2105-14-346

Dey, A. K., Malyala, P., & Singh, M. (2014). Physicochemical and functional characterization of vaccine antigens and adjuvants. *Expert Review of Vaccines*, *13*(5), 671–685. https://doi.org/10.1586/14760584.2014.907528

Di Natale, C., La Manna, S., De Benedictis, I., Brandi, P., & Marasco, D. (2020). Perspectives in peptide-based vaccination strategies for syndrome coronavirus 2 pandemic. *Frontiers in Pharmacology*, *11*, 578382. https://doi.org/10.3389/fphar.2020.578382

Dimitrov, I., Flower, D. R., & Doytchinova, I. (2013). AllerTOP–A server for in silico prediction of allergens. *BMC Bioinformatics*, *14*(S6), 4. https://doi.org/10.1186/1471-2105-14-S6-S4

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, *8*(1), 4. https://doi.org/10.1186/1471-2105-8-4

Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J., & Jiang, S. (2009). The spike protein of SARS-CoV-a target for vaccine and therapeutic development. *Nature Reviews Microbiology*, *7*(3), 226–236. https://doi.org/10.1038/nrmicro2090

Elfiky, A. A. (2021). SARS-CoV-2 RNA dependent RNA polymerase (RdRp) targeting: An *in silico* perspective. *Journal of Biomolecular Structure & Dynamics*, *39*(9), 3204–3212. https://doi.org/10.1080/07391102.2020.1761882

Falk, K., Rötzschke, O., Stevanović, S., Jung, G., & Rammensee, H. G. (1991). Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature*, *351*(6324), 290–296. https://doi.org/10.1038/351290a0

Foged, C. (2011). Subunit vaccines of the future: The need for safe, customized and optimized particulate delivery systems. *Therapeutic Delivery*, *2*(8), 1057–1077. https://doi.org/10.4155/tde.11.68

Garbuglia, A. R., Lapa, D., Sias, C., Capobianchi, M. R., & Del Porto, P. (2020). The use of both therapeutic and prophylactic vaccines in the therapy of papillomavirus disease. *Frontiers in Immunology*, *11*, 188. https://doi.org/10.3389/fimmu.2020.00188

Garnier, J., Gibrat, J. F., & Robson, B. (1996). GOR method for predicting protein secondary structure from amino acid sequence. *Methods in Enzymology*, *266*, 540–553. https://doi.org/10.1016/s0076-6879(96)66034-0

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, *31*(13), 3784–3788. https://doi.org/10.1093/nar/gkg563

Goodwin, D., Simerska, P., & Toth, I. (2012). Peptides as therapeutics with enhanced bioactivity. *Current Medicinal Chemistry*, *19*(26), 4451–4461. https://doi.org/10.2174/092986712803251548

Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D. C., & Jahn, D. (2005). JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Research*, *33*(Web Server issue), W526–W531. https://doi.org/10.1093/nar/gki376

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Open Source Drug Discovery Consortium., & Raghava, G. P. (2013). In silico approach for predicting toxicity of peptides and proteins. *PLoS One*, *8*(9), e73957 https://doi.org/10.1371/journal.pone.0073957

Guy, B. (2007). The perfect mix: Recent progress in adjuvant research. *Nature Reviews. Microbiology*, *5*(7), 505–517. https://doi.org/10.1038/nrmicro1681

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Heo, L., Park, H., & Seok, C. (2013). GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Research*, *41*(Web Server issue), W384–W388. https://doi.org/10.1093/nar/gkt458

Isabel, S., Graña-Miraglia, L., Gutierrez, J. M., Bundalovic-Torma, C., Groves, H. E., Isabel, M. R., Eshaghi, A., Patel, S. N., Gubbay, J. B., Poutanen, T., Guttman, D. S., & Poutanen, S. M. (2020). Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. *Scientific Reports*, *10*(1), 14031.https://doi.org/10.1038/s41598-020-70827-z

Jiang, S., Du, L., & Shi, Z. (2020). An emerging coronavirus causing pneumonia outbreak in Wuhan, China: Calling for developing therapeutic and prophylactic strategies. *Emerging Microbes & Infections*, *9*(1), 275–277. https://doi.org/10.1080/22221751.2020.1723441

Jiang, S., Hillyer, C., & Du, L. (2020). Neutralizing Antibodies against SARS-CoV-2 and Other Human Coronaviruses. *Trends in Immunology*, *41*(5), 355–359. https://doi.org/10.1016/j.it.2020.03.007

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., & Nielsen, M. (2017). NetMHCpan-4.0: Improved peptide-MHC Class I interaction predictions integrating eluted ligand and peptide binding affinity data. *Journal of Immunology*, *199*(9), 3360–3368. https://doi.org/10.4049/jimmunol.1700893

Kalli, K. R., Block, M. S., Kasi, P. M., Erskine, C. L., Hobday, T. J., Dietz, A., Padley, D., Gustafson, M. P., Shreeder, B., Puglisi-Knutson, D., Visscher, D. W., Mangskau, T. K., Wilson, G., & Knutson, K. L. (2018). Folate receptor alpha peptide vaccine generates immunity in breast and ovarian cancer patients. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *24*(13), 3014–3025. https://doi.org/10.1158/1078-0432.CCR-17-2499

Kar, T., Narsaria, U., Basak, S., Deb, D., Castiglione, F., Mueller, D. M., & Srivastava, A. P. (2020). A candidate multi-epitope vaccine against SARS-CoV-2. *Scientific Reports*, *10*(1), 10895. https://doi.org/10.1038/s41598-020-67749-1

Kim, J., Yang, Y. L., Jang, S. H., & Jang, Y. S. (2018). Human β-defensin 2 plays a regulatory role in innate antiviral immunity and is capable of potentiating the induction of antigen-specific immunity. *Virology Journal*, *15*(1), 124. https://doi.org/10.1186/s12985-018-1035-2

Kim, J., Yang, Y. L., Jang, S. H., & Jang, Y. S. (2018). Human β-defensin 2 plays a regulatory role in innate antiviral immunity and is capable of potentiating the induction of antigen-specific immunity. *Virology Journal*, *15*(1), 124. https://doi.org/10.1186/s12985-018-1035-2

Ko, J., Park, H., Heo, L., & Seok, C. (2012). GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Research*, *40*(Web Server issue), W294–W297. https://doi.org/10.1093/nar/gks493

Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., & Vajda, S. (2013). How good is automated protein docking? *Proteins*, *81*(12), 2159–2166. https://doi.org/10.1002/prot.24403

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S. (2017). The ClusPro web server for protein-protein docking. *Nature Protocols*, *12*(2), 255–278. https://doi.org/10.1038/nprot.2016.169

Lam, T. T.-Y., Jia, N., Zhang, Y.-W., Shum, M. H.-H., Jiang, J.-F., Zhu, H.-C., Tong, Y.-G., Shi, Y.-X., Ni, X.-B., Liao, Y.-S., Li, W.-J., Jiang, B.-G., Wei, W., Yuan, T.-T., Zheng, K., Cui, X.-M., Li, J., Pei, G.-Q., Qiang, X., … Cao, W.-C. (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature*, *583*(7815), 282–285. https://doi.org/10.1038/s41586-020-2169-0

Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*, *8*, 424.https://doi.org/10.1186/1471-2105-8-424

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283–291. https://doi.org/10.1107/S0021889892009944

Li, F., Li, W., Farzan, M., & Harrison, S. C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science (New York, N.Y.)*, *309*(5742), 1864–1868. https://doi.org/10.1126/science.1116480

López, J. A., Weilenman, C., Audran, R., Roggero, M. A., Bonelo, A., Tiercy, J. M., Spertini, F., & Corradin, G. (2001). A synthetic malaria vaccine elicits a potent CD8(+) and CD4(+) T lymphocyte immune response in humans. Implications for vaccination strategies. *European Journal of Immunology*, *31*(7), 1989–1998. https://doi.org/10.1002/1521-4141(200107)31:7<1989::aid-immu1989>3.0.co;2-m

López-Blanco, J. R., Aliaga, J. I., Quintana-Ortí, E. S., & Chacón, P. (2014). iMODS: Internal coordinates normal mode analysis server. *Nucleic Acids Research*, *42*(Web Server issue), W271–W276. https://doi.org/10.1093/nar/gku339

Magnan, C. N., Zeller, M., Kayala, M. A., Vigil, A., Randall, A., Felgner, P. L., & Baldi, P. (2010). High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics (Oxford, England)*, *26*(23), 2936–2943. https://doi.org/10.1093/bioinformatics/btq551

Maupetit, J., Derreumaux, P., & Tuffery, P. (2009). PEP-FOLD: An online resource for de novo peptide structure prediction. *Nucleic Acids Research*, *37*(Web Server issue), W498–W503. https://doi.org/10.1093/nar/gkp323

Naveca, F., Nascimento, V., Souza, V., Corado, A., Nascimento, F., Silva, G., Costa, A., Duarte, D., Pessoa, K., & Gonçalves, L. (2021). Phylogenetic relationship of SARS-CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K and N501Y in the Spike protein. *Virological*. Retrieved February 25, 2021 from https://virological.org/t/phylogenetic-relationship-of-sars-cov-2-sequences-from-amazonas-with-emerging-brazilian-variants-harboring-mutations-e484k-and-n501y-in-the-spike-protein/585.

Nielsen, M., & Lund, O. (2009). NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics*, *10*, 296https://doi.org/https://doi.org/10.1186/1471-2105-10-296

O'Hagan, D. T., MacKichan, M. L., & Singh, M. (2001). Recent developments in adjuvants for vaccines against infectious diseases. *Biomolecular Engineering*, *18*(3), 69–85. https://doi.org/10.1016/S1389-0344(01)00101-0

Piccoli, L., Park, Y.-J., Tortorici, M. A., Czudnochowski, N., Walls, A. C., Beltramello, M., Silacci-Fregni, C., Pinto, D., Rosen, L. E., Bowen, J. E., Acton, O. J., Jaconi, S., Guarino, B., Minola, A., Zatta, F., Sprugasci, N., Bassi, J., Peter, A., De Marco, A., … Veesler, D. (2020). Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike

receptor-binding domain by structure-guided high-resolution serology. *Cell*, *183*(4), 1024–1042.e21. https://doi.org/10.1016/j.cell.2020.09.037

Ponomarenko, J., Bui, H. H., Li, W., Fusseder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, *9*, 514.https://doi.org/10.1186/1471-2105-9-514

Rahmani, A., Baee, M., Saleki, K., Moradi, S., & Nouri, H. R. (2021). Applying high throughput and comprehensive immunoinformatics approaches to design a trivalent subunit vaccine for induction of immune response against emerging human coronaviruses SARS-CoV, MERS-CoV and SARS-CoV-2. *Journal of Biomolecular Structure & Dynamics*, 1–17. Advance online publication. https://doi.org/10.1080/07391102.2021.1876774

Rambaut, A., Loman, N., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., Connor, T., Peacock, T., Robertson, D. L., & Volz, E. (2020). Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. COVID-19 Genomics Consortium UK (CoG-UK).

Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One*, *5*(4), e9862. https://doi.org/10.1371/journal.pone.0009862

Rawi, R., Mall, R., Kunji, K., Shen, C. H., Kwong, P. D., & Chuang, G. Y. (2018). PaRSnIP: Sequence-based protein solubility prediction using gradient boosting machine. *Bioinformatics (Oxford, England)*, *34*(7), 1092–1098. https://doi.org/10.1093/bioinformatics/btx662

Reed, S. G., Bertholet, S., Coler, R. N., & Friede, M. (2009). New horizons in adjuvants for vaccine development. *Trends in Immunology*, *30*(1), 23–32. https://doi.org/10.1016/j.it.2008.09.006

Sadat, S. M., Aghadadeghi, M. R., Yousefi, M., Khodaei, A., Sadat Larijani, M., & Bahramali, G. (2021). Bioinformatics analysis of SARS-cov-2 to approach an effective vaccine candidate against COVID-19. *Molecular Biotechnology*, *63*(5), 389–409. https://doi.org/10.1007/s12033-021-00303-0 Epub 2021 Feb 24.

Safavi, A., Kefayat, A., Abiri, A., Mahdevar, E., Behnia, A. H., & Ghahremani, F. (2019). In silico analysis of transmembrane protein 31 (TMEM31) antigen to design novel multiepitope peptide and DNA cancer vaccines against melanoma. *Molecular Immunology*, *112*, 93–102. https://doi.org/10.1016/j.molimm.2019.04.030

Safavi, A., Kefayat, A., Mahdevar, E., Abiri, A., & Ghahremani, F. (2020). Exploring the out of sight antigens of SARS-CoV-2 to design a candidate multi-epitope vaccine by utilizing immunoinformatics approaches. *Vaccine*, *38*(48), 7612–7628. https://doi.org/10.1016/j.vaccine.2020.10.016

Safavi, A., Kefayat, A., Mahdevar, E., Ghahremani, F., Nezafat, N., & Modarressi, M. H. (2021). Efficacy of co-immunization with the DNA and peptide vaccines containing SYCP1 and ACRBP epitopes in a murine triple-negative breast cancer model. *Human Vaccines & Immunotherapeutics*, *17*(1), 22–34. https://doi.org/10.1080/21645515.2020.1763693

Safavi, A., Kefayat, A., Sotoodehnejadnematalahi, F., Salehi, M., & Modarressi, M. H. (2019). In silico analysis of synaptonemal complex protein 1 (SYCP1) and Acrosin Binding Protein (ACRBP) antigens to design novel multiepitope peptide cancer vaccine against breast cancer. *International Journal of Peptide Research and Therapeutics*, *25*(4), 1343–1359. https://doi.org/10.1007/s10989-018-9780-z

Saha, S., & Raghava, G. P. (2004). S. BcePred:Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In G. Nicosia, V. Cutello, P. J. Bentley, & J. Timis (Eds.), *ICARIS 2004, LNCS*. 3239 (pp. 197–204). Springer.

Saha, S., & Raghava, G. P. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*, *65*(1), 40–48. https://doi.org/10.1002/prot.21078

Samad, A., Ahammad, F., Nain, Z., Alam, R., Imon, R. R., Hasan, M., & Rahman, M. S. (2020). Designing a multi-epitope vaccine against SARS-CoV-2: An immunoinformatics approach. *Journal of Biomolecular Structure & Dynamics*, 1–17. https://doi.org/10.1080/07391102.2020.1792347

Shen, X., Tang, H., McDanal, C., Wagh, K., Fischer, W., Theiler, J., Yoon, H., Li, D., Haynes, B. F., Sanders, K. O., Gnanakaran, S., Hengartner, N., Pajon, R., Smith, G., Glenn, G. M., Korber, B., & Montefiori, D. C. (2021). SARS-CoV-2 variant B.1.1.7 is susceptible to neutralizing antibodies elicited by ancestral spike vaccines. *Cell Host & Microbe*, *29*(4), 529–539.e3. Volume Issue https://doi.org/10.1016/j.chom.2021.03.002

Steven, G. R., Fan-Chi, H., Darrick, C., & Mark, T. O. (2016). The science of vaccine adjuvants: Advances in TLR4 ligand adjuvants. *Current Opinion in Immunology*, *41*, 85–90. https://doi.org/10.1016/j.coi.2016.06.007

Tahir Ul Qamar, M., Shahid, F., Aslam, S., Ashfaq, U. A., Aslam, S., Fatima, I., Fareed, M. M., Zohaib, A., & Chen, L. L. (2020). Reverse vaccinology assisted designing of multiepitope-based subunit vaccine against SARS-CoV-2. *Infectious Diseases of Poverty*, *9*(1), 132.https://doi.org/10.1186/s40249-020-00752-w

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E. J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A. J., Engelbrecht, S., Van Zyl, G., … de Oliveira, T. (2021). Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*, *592*(7854), 438–443. https://doi.org/10.1038/s41586-021-03402-9

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Thomsen, M., Lundegaard, C., Buus, S., Lund, O., & Nielsen, M. (2013). MHCcluster, a method for functional clustering of MHC molecules. *Immunogenetics*, *65*(9), 655–665. https://doi.org/10.1007/s00251-013-0714-9

Vajda, S., Yueh, C., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., & Kozakov, D. (2017). New additions to the ClusPro server motivated by CAPRI. *Proteins*, *85*(3), 435–444. https://doi.org/10.1002/prot.25219

Validi, M., Karkhah, A., Prajapati, V. K., & Nouri, H. R. (2018). Immuno-informatics based approaches to design a novel multi epitope-based vaccine for immune response reinforcement against Leptospirosis. *Molecular Immunology*, *104*, 128–138. https://doi.org/10.1016/j.molimm.2018.11.005

Wan, Y., Shang, J., Graham, R., Baric, R. S., & Li, F. (2020). Receptor recognition by the novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS coronavirus. *Journal of Virology*, *94*(7), e00127–20. https://doi.org/10.1128/JVI.00127-20

Westerbeck, J. W., & Machamer, C. E. (2019). The Infectious bronchitis coronavirus envelope protein alters golgi pH to protect the spike protein and promote the release of infectious virus. *Journal of Virology*, *93*(11), e00015–19. https://doi.org/10.1128/JVI.00015-19

WHO. (n.d.). https://covid19.who.int/

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science (New York, N.Y.)*, *367*(6485), 1444–1448. https://doi.org/10.1126/science.abb2762

Yang, Z., Bogdan, P., & Nazarian, S. (2021). An in silico deep learning approach to multi-epitope vaccine design: A SARS-CoV-2 case study. *Scientific Reports*, *11*(1), 3238. https://doi.org/10.1038/s41598-021-81749-9

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER suite: Protein structure and function prediction. *Nature Methods*, *12*(1), 7–8. https://doi.org/10.1038/nmeth.3213

Zhang, W., Davis, B. D., Chen, S. S., Martinez, J. M. S., Plummer, J. T., & Vail, E. (2021). Emergence of a novel SARS-CoV-2 strain in Southern California, USA. *medRxiv*. https://doi.org/10.1101/2021.01.18.21249786.

Zheng, M., Karki, R., Williams, E. P., Yang, D., Fitzpatrick, E., Vogel, P., Jonsson, C. B., & Kanneganti, T.-D. (2021). TLR2 senses the SARS-CoV-2 envelope protein to produce inflammatory cytokines. *Nature Immunology*, *22*(7), 829–838. https://doi.org/10.1038/s41590-021-00937-x

Zhou, Y., Yang, Y., Huang, J., Jiang, S., & Du, L. (2019). Advances in MERS-CoV vaccines and therapeutics based on the receptor-binding domain. *Viruses*, *11*(1), 60. https://doi.org/10.3390/v11010060