Journal of Biomolecular Structure and Dynamics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tbsd20

Epitope based peptide vaccine against SARS-COV2: an immune-informatics approach

Richa Bhatnager, Maheshwar Bhasin, Jyoti Arora & Amita S. Dang

To cite this article: Richa Bhatnager, Maheshwar Bhasin, Jyoti Arora & Amita S. Dang (2021) Epitope based peptide vaccine against SARS-COV2: an immune-informatics approach, Journal of Biomolecular Structure and Dynamics, 39:15, 5690-5705, DOI: 10.1080/07391102.2020.1787227

To link to this article: https://doi.org/10.1080/07391102.2020.1787227

Published online: 03 Jul 2020.

Submit your article to this journal

Article views: 3246

View related articles

View Crossmark data

Citing articles: 13 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tbsd20

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
2021, VOL. 39, NO. 15, 5690–5705
https://doi.org/10.1080/07391102.2020.1787227

Taylor & Francis
Taylor & Francis Group

Check for updates

Epitope based peptide vaccine against SARS-COV2: an immune-informatics approach

Richa Bhatnager[a], Maheshwar Bhasin[b], Jyoti Arora[a] and Amita S. Dang[a]

[a]Centre for Medical Biotechnology, M.D. University, Rohtak, Haryana, India; [b]Department of Neonatology, Lady Hardinge Medical College and associated hospital, New Delhi, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

World is witnessing exponential growth of SARS-CoV2 and fatal outcomes of COVID 19 has proved its pandemic potential already by claiming more than 3 lakhs deaths globally. If not controlled, this ongoing pandemic can cause irreparable socio-economic and psychological impact worldwide. Therefore a safe and effective vaccine against COVID 19 is exigent. Recent advances in immunoinformatics approaches could potentially decline the attrition rate and accelerate the process of vaccine development in these unprecedented times. In the present study, a multivalent subunit vaccine targeting S2 subunit of the SARS-CoV2 S glycoprotein has been designed using open source, immunoinformatics tools. Designed construct comprises of epitopes capable of inducing T cell, B cell (Linear and discontinuous) and Interferon γ. physiologically, vaccine construct is predicted to be thermostable, antigenic, immunogenic, non allergen and non toxic in nature. According to population coverage analysis, designed multiepitope vaccine covers 99.26% population globally. 3D structure of vaccine construct was designed, validated and refined to obtain high quality structure. Refined structure was docked against Toll like receptors to confirm the interactions between them. Vaccine peptide sequence was reverse transcribed, codon optimized and cloned in pET vector. Our *in-silico* study suggests that proposed vaccine against fusion domain of virus has the potential to elicit an innate as well as humoral immune response in human and restrict the entry of virus inside the cell. Results of the study offer a framework for in-vivo analysis that may hasten the process of development of therapeutic tools against COVID 19.

ARTICLE HISTORY
Received 11 June 2020
Accepted 21 June 2020

KEYWORDS
SARS-COV2; COVID 19; epitope mapping; population coverage; insilico cloning

Introduction

COVID 19, Corona virus disease 2019, has brought the world to a staggering halt. The overwhelming human transmission of SARS-CoV2 is unprecedented. The disease was first reported in Wuhan district, China during late December 2019 and has claimed lives in over 210^{+} countries and territories worldwide. Due to the virus knack for exponential growth and being highly contagious to humans, it was regarded as a Public Health Emergency of International Concern (PHEIC) by WHO on 30th January 2020 and had been declared a pandemic on 11th March 2020 (Grasselli et al., 2020; Wachyuni & Kusumaningrum, 2020; World Health Organization, 2020a).

As of May 16th 2020, more than 300,000 lives are lost globally due to ongoing pandemic and 4 million confirmed cases are battling with the disease across the globe (World Health Organization, 2020a). The daunting task in front of the scientific community is to find measures to treat/prevent the outbreak of disease. The novelty of the virus SARS-CoV2 augments to the uncertainties in pathophysiology, clinical management, prevention and treatment of the underlying COVID 19 disease.

Accrediting to the advances in next generation sequencing techniques and availability of various informatics tools at our disposal, the entire genome of SARS-CoV2 was sequenced and made available on January 12th, 2020 (GENEBANK MN908947.1) (Wu et al., 2020). Phylogenetically, SARS-CoV2 is a member of genus Betacornavirus and has a positive sense single strand RNA genome of ∼ 30000 bases in length. The RNA genome codes for 4 structural proteins i.e. namely Spike (S), Envelope (E), Membrane (M) and Nucleocapsid (N) (Andersen et al., 2020).

SARS-CoV 2 is phylogenetically classified as a strain of severe acute respiratory syndrome related coronaviruses by International Committee on Taxonomy of Viruses (Gorbalenya et al., 2020). The striking similarities between the genetic makeup of SARS-CoV2 with its predecessor human corona viruses SARS-CoV (79.7%) and MERS-CoV (50%) could potentially provide insights into the pathogenesis of COVID 19 (Lu et al., 2020).

Attempts made to gain insights into pathophysiology of COVID 19 based on SARS-CoV have shown that both the viruses attach to ACE 2 (Angiotensin converting enzyme 2) receptors with an almost identical spike (S) protein as a mode of cell entry. However, the affinity of SARS-CoV2 S

CONTACT Amita S. Dang suneja_a@yahoo.co.in Centre for Medical Biotechnology, M. D. University, Rohtak, Haryana124001, India
© 2020 Informa UK Limited, trading as Taylor & Francis Group

protein to ACE 2 is 10x to 20x more than SARs-CoV, this might explain the contagious nature of SARS CoV2 (Tai et al., 2020; Wrapp et al., 2020).

According to Asian Development Bank report, an economic loss of $8.8 trillion dollars is projected due to COVID 19 and together with the deteriorating impact on human life it makes it axiomatic to dedicate our current resources towards combating COVID 19 (Park et al., 2020). Vaccines are a proven cost effective measure and to prevent the outbreak we should dwell upon developing safe and effective vaccines against highly contagious SARS-CoV2.

WHO has launched a solidarity trail to identify favourable attributes of safe and effective vaccines (World Health Organization, 2020a; von Tigerstrom & Wilson, 2020). The compressed timeline, safety of vulnerable population and large-scale manufacturing of vaccine remains a few imminent challenges. The projected goals of vaccine development against COVID 19 are unparalleled.

A study on clinical development success rate for various drugs concludes a likelihood of acceptance (LOA) of 16.2% for vaccine candidates from phase 1 to phase 3. Another study projects an average timeline for development of vaccine candidate to be 10 .71 years and a market entry rate of 6% (Le et al., 2020; Pronker et al., 2013). The rapid development, urgency of the situation and mutation rate of SARS-CoV2 might add up to attrition rate of the potential vaccine candidates.

Exploiting immunoinformatics tools offer a spark of hope in reducing the attrition rate and selection of putative valid vaccine target. The reverse vaccinology approach has been successful in development of FDA licensed vaccine bexsero (Folaranmi et al., 2015). The infusion of bioinformatics analysis of vaccine candidates in the decision-making pipeline could potentially help save billions of dollars, and guide to researchers to make an informed choice on selection of candidates.

Studies to identify potential vaccine candidates using genome wide approach against SARS-CoV2 have suggested targeting S glycoprotein as a potential for vaccine development (Ong et al., 2020). S glycoprotein is a major structural protein of corona viruses, crucial for virulence, adhesion mediated cell entry and tissue tropism. S glycoprotein is also known for triggering the production neutralizing antibody against the viruses (Walls et al., 2020). S glycoprotein (Uniprot: PODTC2) consists of subunits S1 and S2 that mediates cell entry by S1 binding to ACE 2 receptor, and S2 acts as a class 1 fusion protein (UniProt Consortium, 2007). Unlike other hCoVs, SARS-CoV2 possess a 4 furin ring at the cleavage site between S1/S2 domain, suggesting fusion protein might play a pivotal role in contagious nature of SARS-CoV2 (Xia et al., 2020).

Viral fusion proteins are potential therapeutic and vaccine targets. However, conserved antibody binding sites in S2 subunit, (Walls et al., 2020) and its potential to elicit a neutralizing antibody response (Zhang et al., 2004) could generate an antibody mediated immune response only for a short period. A tailored multiepitope vaccine has a potential to generate more potent response by activating not just B cell but also cytotoxic T and Helper T cells. Moreover, their added advantages like feasibility of manufacturing peptides and epitopes, chemical stability and lack of the infectious component of the virus presents them as a promising candidate to prevent the emergent situation of COVID 19 (Enayatkhani et al., 2020).

In the present study, a multivalent subunit vaccine targeting S2 subunit of the SARS-CoV2 S glycoprotein pathogenic SARS-CoV2 has been designed using open source, immunoinformatics tools. Epitopes predicted for T cell, B cell, MHC 1 and MHC II were used to design vaccine construct that was further analysed for physiochemical properties, antigenicity, allergenicity, and cross reactivity to human and mouse proteome. The structure was validated and analysed for prediction of B cell epitope and IFN-γ epitope. Results of the study offer a framework for in-vivo analysis that may hasten the process of development of therapeutic tools against COVID 19.

Materials and methods

Antigen sequence retrieval

For the current analysis, one of the important the structural protein of viral genome the Spike glycoprotein S2 domain was selected as candidate antigen for the study. S2 domain plays a prominent role in virus fusion with the host cell and mediates the entry process. The Spike glycoprotein S2 (SPIKE_SARS2) protein sequence of SARS-CoV-2 (accession no. P0DTC2) was retrieved from NCBI in FASTA format (https://www.ncbi.nlm.nih.gov). Reference protein sequence of 'S' subunit of all seven variant of human coronavirus i.e. 229E (alpha coronavirus, Accession no: NP_073551.1), NL63 (alpha coronavirus, Accession no: ANU06084.1), OC43 (beta coronavirus, Accession no: YP_009555241.1), HKU1 (beta coronavirus, Accession no: AGW27881.1) , MERS-CoV (the beta coronavirus that causes Middle East Respiratory Syndrome, or MERS, Accession no: YP_009047204.1) , SARS-CoV (the beta coronavirus that causes severe acute respiratory syndrome, or SARS, Accession no: ABB29898.2), SARS-CoV-2 (the novel coronavirus that causes coronavirus disease 2019, or COVID-19, Accession no: >QJX59884.1) were retrieved from NCBI and subjected to multiple sequence alignment through clustal W. Toll-like receptor 3 (TLR3) agonist β-defensin was retrieved in FASTA format from NCBI.

Predictions of T- and B-cell for designing the vaccine construct

SPIKE_SARS2 protein sequence was first analyzed to predict probable T- and B-cell epitopes followed by the fusion of adjuvant and linkers with the selected epitopes and was subsequently used for further analysis

Epitope prediction for MHC-I

For the prediction of MHC 1 compatible epitopes, two web based servers IEDB and RANKPEP servers were utilized to predict 9-mer epitopes in the sequence that might bind to

Human MHC-I alleles. Both the server requires the sequence in FASTA format as retrieved in step 2.1. IEDB (http://tools.iedb.org/mhci) operates on different methods for epitope prediction, such as a consensus method based on artificial neural network (ANN), stabilized matrix method (SMM), and combinatorial peptide libraries (CombLib). The prediction was done based on the IEDB recommended prediction method (Kim et al., 2012). FASTA sequence of S2 domain was used as an input along with MHC source species 'Human' and human specific MHC-I alleles. RANKPEP server (http://imed.med.ucm.es/Tools/rankpep. html) utilizes position specific scoring matrix (PSSM) for prediction. This server was employed to identify MHC-I binding epitopes of S2 domain with immunodominance threshold 59.4% sensitivity and 69.4% specificity (Reche & Reinherz, 2007).

Prediction of CTL binding epitopes and TCR-peptide/peptide-MHC interfaces

To predict, CTL epitopes, we utilized a web based server CTLPred (http://crdd.osdd.net/raghava/ctlpred/). The method employed by the server is based on elegant machine learning techniques like an Artificial Neural network and support vector machine. The server also utilizes consensus and combinatorial approach to predict CTL specific epitopes (Bhasin & Raghava, 2004). Single letter amino acid FASTA sequence of Spike glycoprotein S2 was submitted for 9-mer CTL epitopes to obtain sequences of MHC restricted epitopes.

To investigate the interfaces between peptide-MHC and peptide-TCR for the identification ofr homologous peptide antigens from a query, PAComplex (http://pacomplex.life.nctu.edu.tw) was employed. It is an online server that gives the detailed atomic interactions, homologous peptide antigen, the binding models, and joint Z-value of each sequence (Liu et al., 2011).

MHC-II binding epitope prediction

RANKPEP and IEDB servers were used again to predict epitopes for human MHC-II alleles (Bhasin & Raghava, 2004). In the RANKPEP (http://imed.med.ucm.es/Tools/rankpep.html), similar to MHC I epitopes, the prediction is carried out using PSSMs. The top score of MHC-II restricted epitopes with the length of nine amino acids was predicted at 2% binding threshold. The IEDB (http://tools.iedb.org/mhcii/) was used to predict 15-mer MHC-II binding epitopes in a way similar to MHC-I binding epitopes approach.

Designing of vaccine construct

Overlapping fragments that are consensus by the prediction of all servers were selected. All the overlapped epitopes of HTL and CTL were linked together with the help of GPGPG and AAY linkers, respectively. The β-defensin amino acid sequence was added as an adjuvant which was linked at the N-terminal of construct with the help of EAAAK linker.

Evaluation of the designed vaccine properties

Physicochemical properties

Physiochemical properties of designed vaccines were computed by ProtParam (http://web.expasy.org/protparam). This tool predict the various properties like theoretical isoelectric point (pI), molecular weight, amino acid and atomic compositions, aliphatic index, extinction coefficient, grand average of hydropathicity (GRAVY) half-life and instability index (Gasteiger et al., 2005).

Antigenicity

Two servers were employed to predict the antigenicity of vaccine. VaxiJen v2.0 (http://www.ddgpharmfac.net/vaxijen/VaxiJen/VaxiJen.html), was used with the selection of virus and threshold = 0.4 threshold to make the alignment-independent prediction of protective antigens. Accuracy of this server ranged from 70% to 89% depending on the target organism (Doytchinova & Flower, 2007). ANTIGENpro (http://scratch.proteomics. ics.uci.edu/) server that is mainly sequence dependent uses a two stage process to predict the antigenicity. Its prediction also based on the machine learning algorithms and multiple representations of primary sequence (Magnan & Baldi, 2014).

Allergenicity

AllergenFP v.1.0 (http://ddg-pharmfac.net/ AllergenFP/), and AlgPred (http://crdd.osdd.net/raghava/algpred/index.html) were utilized for the allergenicity prediction (Dimitrov et al., 2014; Saha & Raghava, 2006). AllergenFP utilizes the descriptor fingerprint approach that is alignment with accuracy of about 88%. On the other hand, AlgPred predicts allergens based on similarity of the known epitope with any region of protein by integrating different approaches.

Prediction of protein solubility

Protein solubility on over expression in E. coli was predicted SOLpro (http://scratch.proteomics.ics.uci.edu) (Magnan et al., 2009). This server utilizes a two-stage SVM architecture method that is having an accuracy of more than74%.

Analysis of cross-reactivity with mouse and human proteomes

To measure the sequence similarity of the designed epitope-based vaccine with mouse and human proteomes, BLASTP search was done on the NCBI database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) according to the default parameters. None of antigenic sequence was found to have more than 10% similarity with mouse and human proteome.

Secondary and tertiary structure analyses

Secondary structure

Secondary structure of vaccine construct was analyzed by PSIPRED. Prediction of server is based on the incorporation of two feed-forward neural networks which perform an

analysis on output obtained from PSI-BLAST (Position Specific Iterated - BLAST)

Homology modeling. Homology modeling of the vaccine structure was done using Swiss Modeller which utilizes alignment of the target sequence and template structures for homology modeling (Waterhouse et al., 2018). Energy minimization was done with Chimera (Pettersen et al., 2004). Pymol 1.7 was used to visualize the 3D structures after modeling.

Refinement and validation of the 3D model. GalaxyRefine server (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE) was used to refine the best 3D structure on the basis of Z score. Refinement done by server was based on the repeated structure perturbation followed by overall relaxation of structure (Heo et al., 2013).

Validation of native and refined 3D model was carried out by the Ramachandran plot (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php), ProSA-web server (https://prosa.services.came.sbg.ac.at/prosa.php), and ERRAT server (http://services.mbi.ucla.edu/ERRAT/).

ProSA prediction is based on atomic coordinates by which it recognizes errors in the 3D models of protein structures through estimating an overall quality score (Z score) If the estimated Z score is outside the range of native proteins, the 3D structure is likely to contain errors (Wiederstein & Sippl, 2007).

The Ramachandran plot is based on ϕ and ψ dihedral angles that verifies backbone conformation of the protein structures by analyzing for each residue and finally classifies the residues into favorable, allowed, and outlier regions (Lovell et al., 2003).

ERRAT is a server that verifies the predicted 3D structures by calculating error values based on the statistics of nonbonded atomic interactions in the structures (Colovos & Yeates, 1993).

Linear and discontinuous B-cell epitope prediction

After designing the final vaccine construct, linear and discontinuous B-cell epitopes was predicted by BCPRED and Discotope servers (Kim et al., 2012). The BCPRED (http://ailab.ist.psu.edu/bcpred/predict.html) prediction is based on kernel method for prediction. BCPred default method that is 75% specific upto the epitope length of 20mer was employed for the current analysis.

Discotope (http://www.cbs.dtu.dk/ services/DiscoTope/) was employed to identify discontinuous B-cell epitopes on the designed 3D structure, the based on the default threshold (−3.7). The server uses calculation of surface accessibility and epitope propensity amino acid score with the sensitivity and specificity of 0.47 and 0.75, respectively.

IFN-γ epitope prediction

Vaccine construct was also screened for the epitopes that can induce the IFN-γ release from CD4+ T cells. IFNepitope (http://crdd.osdd.net/raghava/ifnepitope/) was employed for the this. This server uses a dataset consisting of IFN-γ inducing and non IFN-γ inducing MHC class II binders and predict the antigens from protein sequences having the potential to induce IFN-γ secretion from T_H cells. Three different approaches i.e. motif-based, SVM-based, and hybrid approaches are available to predict the antigenicity, however hybrid approach shows a maximum accuracy of 81.39% in 0.1 cut-off (Dhanda et al., 2013).

Population coverage and conservation across antigen

Because of polymorphic nature of MHC molecules, it is important to check the whether the vaccine construct is able to induce immunogenicity in ethnically biased population. For this purpose, Population coverage analysis was carried out by using Population coverage tool of IEDB server (Kim et al., 2012). this tool was employed to calculate the fraction of individuals supposed respond to a given epitope set on the basis of HLA genotypic frequencies and MHC binding and/or T cell restriction data. Conserved epitopes offered broader protection across different strain of an pathogen, therefore precise information indicating the to the degree of conservancy of the considered epitope is crucial. For conservation across antigen of selected epitopes, IEDB tool 'conservation across antigen' was deployed for both linear and discontinuous epitopes.

Interaction between vaccine construct and TLR3 and 8

Protein–protein docking was performed by Cluspro 2.0 (http://cluspro.bu.edu/login.php) server. It evaluates a variety of protein complexes and finally selects the top-ranked complexes according to their clustering properties (Kozakov et al., 2017). Chimera was used for visualization and determination of binding sites.

Codon optimization and in silico cloning

To clone the nucleotide sequence of the vaccine construct in a suitable expression vector, template DNA sequence was reverse translated from the peptide sequence. Java Codon Adaptation Tool (JCAT) was used for codon optimization (Grote et al., 2005). The JCAT was used to improve translational efficiency in mammalian host cell line. To predict the success of protein expression in the host, different features of the sequence including Codon Adaptation Index (CAI), guanine cytosine (GC) content, and Codon Frequency Distribution (CFD) were evaluated. Therefore, the optimized DNA sequence was submitted to the GenScript Rare Codon Analysis Tool at (https://www.genscript.com/tools/rarecodon-analysis). The adapted nucleotide sequence was cloned into the plasmid vector by using the SnapGene tool. Restriction and insertion cloning method was used for the *insilico* cloning.

Results

Identification and selection of MHC I binding epitope

S2 domain of SPIKE_SARS 2 protein sequence was screened for cytotoxic T cell (Tc cells) and Helper T cell (TH cells) specific epitopes. Tc cells are MHC I restricted, therefore probable epitopes for MHC I were identified by IEDB and RANKPEP servers. IEDB server predicted peptide fragment of 9mer and 10 mer with a combined score. Top 10 peptides with a higher score were selected for further analysis (Table 1). Peptide sequence 865LTDEMIAQY873 was found to have maximum combined score, indicating the maximum efficacy to be antigenic and immunogenic. In addition to maximum score, 865LTDEMIAQY873 is restricted to most of the MHC class I (HLA-A, HLA-B, HLA-C) as suggested in Table1. Another nine sequences listed in table also restricted to majority of class I MHC, suggesting their potential to be antigenic and immunogenic. RANKPEP predicted 26 potential epitopes, out of which topmost 8 peptide sequences were selected for further analysis (Table 1).

Identification and selection of CTL binding epitopes and TCR-peptide/peptide-MHC interfaces

After prediction of MHCI restricted epitopes, Prediction of CTL binding epitopes and TCR-peptide/peptide-MHC interfaces was done. CTLPred and PAcomplex server used to predict CTL binding epitopes. CTLPred server predicted five potential epitopic region based on ANN and SVM score along with MHC restriction of the epitopes (Table 2). PAComplex results predicted only the antigenic peptide having scores more than 4.0. S2 domain protein sequence was submitted for its prediction and it predicted 16 antigenic peptide having score more than 4 which are restricted to HLA-A*0201,HLA-B*0801,HLA-B35*08, H2Kb, H2Kbm3,H2Kbm8, H2Ld mainly.

Identification and selection of MHC-II binding epitopes

MHC-II binding epitope prediction was also done by IEDB and RANKPEP server. In IEDB server combined score along with SMM alignment. IC50 score was considered to filter out most antigenic epitopes. Epitopes having IC50 score less than 200 were considered for further analysis (Table 3). Peptide sequence ^{865}LTDEMIAQYTSALLA^{869} was found to have maximum combined score via IEDB recommended method. According to both IEDB and RANKPEP server, a number of peptide sequences were found to be overlapping, antigenic, immunogenic and non toxic in nature.

Population coverage analysis of selected epitopes

Population coverage analysis was done on the selected epitopes and it was found that these epitopes are conserved in 94.62% in Asian population, 91.65% in Black population, 82.69% in European Caucasian, 84.54% in North-American Caucasian population and 99.26% worldwide.

Designing of vaccine construct

All the overlapped epitopes of HTL and CTL were linked together with the help of GPGPG and AAY linkers, respectively. The β-defensin amino acid sequence was added as an adjuvant which was linked at the N-terminal of construct with the help of EAAAK linker.

Among all the overlapping sequences, sequences are selected on the basis of overall score and shown here 687-695: 9AA(CTL), 864-873:10AA(CTL), 896-904: 9AA(CTL), 976-984:9 AA(CTL), 1201-1209:9AA(CTL), 1006-1020: 15AA(HTL), 1124-1138:15AA(HTL) and 1212-1225(15AA). These sequences were considered for putative vaccine construct.

After ligation vaccine construct obtained is GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKKEAAAKVASQSIIAYAAYLLTDEMIAQYAAYIPFAMQMAYAAYVLNDILSRLGGGGSTYVTQQLIRAAEIRAGPGPGGNCDVVIGIVNNTVYGGGGSQELGKYEQYGPGPGKWPWYIWLGFIAGLI.

Physiochemical properties analyses of vaccine construct

Physiochemical properties of synthetic vaccine construct such as molecular weight, half-life, theoretical isoelectric point (pI), amino acid and atomic compositions, extinction coefficient, instability, aliphatic index, and grand average of was determined by Protparam as detailed in Table 3. According to the results of Protparam server, construct is composed of 170AA with a molecular weight of 18.35Kda unit and pI 9.27. Construct is having 10 negatively charged residues (Asp + Glu) and 19 positively charged residues (Arg + Lys). Molar Ext. coefficient is 36245 M^{-1} cm^{-1}, at 280 nm. Half life of vaccine construct was estimated in three reference cells and predicted to be 30 h in mammalian cells (mammalian reticulocytes, in vitro), >20 h in yeast (*in vivo*) and >10 h in Escherichia coli (in vivo). Instability index of predicted protein is stable with an aliphatic index of 88.47 and Grand average of hydropathicity (GRAVY) index is −0.002.

Allergenicity and antigenicity prediction of vaccine construct

As synthetic vaccine construct was found to be stable and can survive in human body, next step of its testing was carried out. This step was to check whether the construct is allergenic or not. To achieve this, AllergenFP v.1.0 and AlgPred were deployed and it was found that vaccine is non-alergenic in nature. AlgPred score is 0.73 which is above threshold value.

Antigenicity was predicted by VaxiJen v2.0 and ANTIGENpro. Vaxijen predicted construct as antigenic as having higher score than the threshold (0.4). ANTIGENpro server predicted the score of 0.658 suggesting the protein is antigenic. In addition to this, vaccine solubility upon over-expression in *E. coli* host was assessed by SOlPro server which predicts it as insoluble with probability of 70.7%.

Secondary and tertiary structure prediction and validation of vaccine construct

After confirming the antigenicity and non-allergenic nature of synthetic vaccine construct, Secondary structure of vaccine construct was predicted by PSIPRED (Figure 1(a)). According

TABLE 1. Prediction of MHC-I binding Epitope based on IEDB and RANKPEP server.

Allele	Start	End	Peptide length	Peptide Seq	Score
HLA-A*01:01,HLA-A*30:02,HLA-B*35:01, HLA-B*58:01,HLA-A*26:01,HLA-A*68:01, HLA-B*57:01, HLA-B*53:01, HLA-B*15:01, HLA-A*11:01, HLA-A*32:01, HLA-A*03:01, HLA-A*02:06,HLA-A*68:02, HLA-A*30:01, HLA-B*51:01,HLA-A*31:01, HLA-A*33:01, HLA-A*23:01, HLA-A*02:03, HLA-A*24:02, HLA-A*02:01,HLA-B*44:03,HLA-B*44:02, HLA-B*08:01,HLA-B*40:01,HLA-B*07:02	865	873	9	LTDEMIAQY	0.998741
HLA-A*01:01,HLA-A*30:02,HLA-B*15:01, HLA-A*26:01,HLA-B*58:01,HLA-B*35:01, HLA-A*68:01,HLA-A*02:03,HLA-B*57:01, HLA-A*11:01,HLA-A*03:01, HLA-A*32:01, HLA-B*53:01, HLA-A*02:06, HLA-A*02:01, HLA-A*31:01, HLA-A*23:01, HLA-A*24:02, HLA-A*30:01, HLA-A*68:02, HLA-A*33:01, HLA-B*51:01,HLA-B*08:01,HLA-B*44:02, HLA-B*44:03, HLA-B*40:01 ,HLA-B*07:02	864	873	10	LLTDEMIAQY	0.990087
HLA-B*35:01, HLA-B*53:01, HLA-B*51:01, HLA-B*07:02,HLA-A*26:01,HLA-A*30:02, HLA-A*68:01,HLA-A*01:01,HLA-B*58:01, HLA-B*15:01,HLA-A*32:01,HLA-B*57:01, HLA-B*08:01,HLA-A*03:01,HLA-A*33:01, HLA-B*44:03,HLA-B*44:02,HLA-A*11:01, HLA-A*68:02,HLA-A*23:01,HLA-A*30:01, HLA-B*40:01,HLA-A*31:01,HLA-A*24:02, HLA-A*02:06,HLA-A*02:01	896	904	9	IPFAMQMAY	0.987213
HLA-A*02:03,HLA-A*02:01,HLA-A*02:06, HLA-A*32:01,HLA-A*68:02,HLA-B*15:01, HLA-B*08:01,HLA-B*58:01,HLA-A*23:01, HLA-A*30:02,HLA-A*24:02,HLA-A*26:01, HLA-A*01:01,HLA-B*57:01,HLA-A*30:01, HLA-A*31:01,HLA-B*51:01,HLA-B*53:01, HLA-A*03:01,HLA-B*07:02,HLA-B*40:01, HLA-A*33:01,HLA-A*68:01,HLA-B*35:01, HLA-A*11:01,HLA-B*44:02,HLA-B*44:03	976	984	9	VLNDILSRL	0.98603
HLA-B*40:01,HLA-B*44:03HLA-B*44:02, HLA-B*15:01,HLA-A*02:06,HLA-B*08:01, HLA-B*53:01,HLA-A*26:01,HLA-A*68:02, HLA-B*35:01,HLA-A*30:02,HLA-B*58:01, HLA-A*02:03,HLA-A*30:01,HLA-B*51:01, HLA-A*23:01HLA-B*57:01,HLA-B*07:02, HLA-A*32:01,HLA-A*24:02,HLA-A*01:01, HLA-A*02:01,HLA-A*68:01,HLA-A*31:01, HLA-A*03:01,HLA-A*33:01,HLA-A*11:01	1016	1024	9	AEIRASANL	0.974826
HLA-A*24:02,HLA-A*23:01,HLA-B*35:01, HLA-A*30:02,HLA-A*32:01,HLA-B*15:01, HLA-B*58:01,HLA-B*53:01,HLA-B*57:01, HLA-B*08:01,HLA-B*51:01,HLA-A*01:01, HLA-A*30:01,HLA-A*26:01,HLA-B*07:02, HLA-A*33:01,HLA-A*31:01,HLA-B*44:02, HLA-B*44:03,HLA-B*40:01,HLA-A*68:02, HLA-A*68:01,HLA-A*03:01,HLA-A*02:03, HLA-A*02:06,HLA-A*11:01,HLA-A*02:01	635	643	9	VYSTGSNVF	0.969649
HLA-B*44:03,HLA-B*44:02,HLA-A*30:02, HLA-B*40:01,HLA-B*15:01,HLA-A*01:01, HLA-B*35:01,HLA-A*26:01,HLA-B*53:01, HLA-B*58:01,HLA-B*57:01, HLA-A*68:01,HLA-B*08:01,HLA-A*32:01, HLA-A*23:01,HLA-A*33:01, HLA-A*03:01,HLA-A*02:06,HLA-A*24:02, HLA-A*31:01,HLA-B*51:01,HLA-A*11:01, HLA-A*30:01,HLA-A*68:02,HLA-A*02:01, HLA-A*02:03	1201	1209	9	QELGKYEQY	0.968666
HLA-A*32:01,HLA-A*33:01,HLA-A*31:01, HLA-A*68:02,HLA-A*30:02,HLA-A*02:06, HLA-B*51:01,HLA-A*30:01,HLA-B*58:01, HLA-B*53:01,HLA-A*02:01,HLA-A*02:03, HLA-A*68:01,HLA-A*26:01,HLA-A*01:01, HLA-B*57:01,HLA-B*08:01,HLA-A*11:01, HLA-B*40:01,HLA-B*35:01,HLA-B*44:02, HLA-B*44:03,HLA-A*03:01,HLA-B*07:02, HLA-B*15:01	1208	1216	9	QYIKWPWYI	0.963837
	687	695	9	VASQSIIAY	0.96344

(*continued*)

TABLE 1. Continued.

Allele	Start	End	Peptide length	Peptide Seq	Score
HLA-B*35:01,HLA-B*15:01,HLA-B*58:01, HLA-A*30:02,HLA-A*01:01,HLA-B*53:01, HLA-B*57:01,HLA-A*26:01,HLA-A*68:01, HLA-A*11:01,HLA-A*32:01,HLA-A*03:01, HLA-A*30:01,HLA-B*51:01,HLA-A*31:01, HLA-B*07:02,HLA-B*08:01,HLA-B*44:03, HLA-B*44:02,HLA-A*02:06,HLA-A*33:01, HLA-A*68:02,HLA-A*23:01,HLA-A*24:02, HLA-B*40:01,HLA-A*02:03,HLA-A*02:01 HLA-A*23:01,HLA-A*24:02,HLA-A*32:01, HLA-A*33:01,HLA-A*31:01,HLA-A*68:02, HLA-A*30:02,HLA-A*02:06,HLA-B*51:01, HLA-A*30:01,HLA-B*58:01,HLA-B*53:01, HLA-A*02:01,HLA-A*02:03,HLA-A*68:01, HLA-A*26:01,HLA-A*01:01,HLA-B*57:01, HLA-B*08:01,HLA-A*11:01,HLA-B*40:01, HLA-B*35:01,HLA-B*44:02,HLA-B*44:03, HLA-A*03:01,HLA-B*07:02,HLA-B*15:01	1208	1216	9	QYIKWPWYI	0.96134

RANKPEP

Position	C terminal	Peptide seq	N terminal	Mol Weight	Opt %
1121	DNT	FVSGNCDVV	IGI	921.03	50.27
797	IKD	FGGFNFSQI	LPD	998.11	49.17
1042	RVD	FCGKGYHLM	SFP	1037.26	40.96
951	LQD	VVNQNAQAL	NTL	938.04	40.55
915	TQN	VLYENQKLI	ANQ	1101.31	37.78
1122	NTF	VSGNCDVVI	GIV	887.01	31.81
1055	FPQ	SAPHGVVFL	HVT	908.07	31.07
1236	MLC	CMTSCCSCL	KGC	932.17	29.72

TABLE 2. Prediction of CTL binding epitopes and TCR-peptide/peptide-MHC interfaces.

CTLPred

Start Position	Sequence	Score(ANN/SVM)	Prediction	MHC Restriction
1192	NLNESLIDL	0.92/1.5879094	Epitope	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb ,H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
236	TRFQTLLAL	0.99/1.4877641	Epitope	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb ,H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01, HLA-35
983	RLDKVEAEV	0.71/1.6892505	Epitope	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb ,H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
679	NSPRRARSV	0.87/1.0916413	Epitope	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb ,H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
292	ALDPLSETK	0.92/1.0117189	Epitope	HLA-A2, HLA-A*0201, HLA-A*0202, HLA-A24, HLA-B*070,HLA-Cw*0401, H2-Db, H2-Dd,H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01

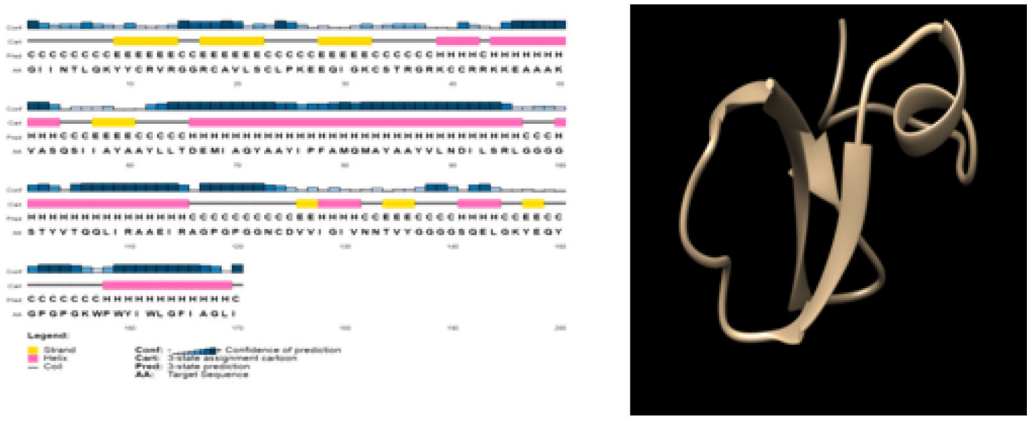
PAComplex

PEPTIDE	Position	Z-Score	Allele
SIIAYTMSL	691	4.51	HLAA0201
SLQTYVTQQ	1003	4.41	
QYIKWPWYI	1208	4.22	
DFGGFNFSQ	796	4.02	
IVNNTVYDP	1132	4	**HLAB0801**
LQDVVNQNAQALN	948	4.12	HLAB3508
TQRNFYEP	1105	4.9	H2Kb
NKVTLADA	824	4.42	
KNHTSPDV	1157	4.21	
QNVLYENQ	913	4.02	
TQRNFYEP	1105	5.38	H2Kbm3
QVKQIYKT	784	4.47	
EMIAQYTS	868	4.01	
DKYFKNHT	1153	4.39	H2Kbm8
QRNFYEPQ	1106	4.28	
ELDSFKEEL	1144	4.32	H2Ld

to prediction majority of amino acids are involved in helix and coil formation. Amino acid residues ranging from 40–53, 65–95 and 100–115 are the major peptide involved in helix formation. Swiss modeler was employed to predict its three dimensional structure. Two 3D structures were predicted for the vaccine sequence, among which the best structure was chosen with the highest Z-score (-2.02). 3D structure of vaccine model is shown in Figure1b. Energy of selected model was minimized by UCSF chimera. Quality of Selected model was checked by PROSA and Ramachandran plot. After all refinements, the model reached Z-score of −5.76. Ramachandran plot analysis showed that 90.4%, 6.6%, and3% of residues were located in the favored, allowed, and outlier regions (Figure 2), respectively, in the original as compared with 98%, 2%, and 0% of residues in the refined model. Furthermore, the results from the ERRAT plot analysis indicated that the overall quality factors in the original and refined models were 66.66 and 82.3, respectively.

Table 3. Prediction of MHC II binding epitope.

MHC restricted alleles	Position	Peptide Sequence	Comb. score
HLA-DRB1*07:01,HLA-DRB1*15:01,HLA-DRB4*01:01,HLA-DRB5*01:01,HLA-DRB1*03:01,HLA-DRB3*01:01,HLA-DRB3*02:02	687-701	VASQSIIAYTMSLGA	0.92
HLA-DRB1*15:01,HLA-DRB1*07:01,HLA-DRB4*01:01,HLA-DRB5*01:01,HLA-DRB1*03:01,HLA-DRB3*01:01,HLA-DRB3*02:02	865-879	LTDEMIAQYTSALLA	0.93
HLA-DRB5*01:01,HLA-DRB1*15:01,HLA-DRB3*01:01,HLA-DRB1*07:01,HLA-DRB4*01:01,HLA-DRB1*03:01,HLA-DRB3*02:02	896-910	IPFAMQMAYRFNGIG	0.70
HLA-DRB1*03:01,HLA-DRB4*01:01,HLA-DRB3*01:01,HLA-DRB1*15:01,HLA-DRB1*07:01,HLA-DRB5*01:01,HLA-DRB3*02:02	976-990	VLNDILSRLDKVEAE	0.68
HLA-DRB4*01:01,HLA-DRB1*07:01,HLA-DRB1*15:01,HLA-DRB5*01:01,HLA-DRB1*03:01,HLA-DRB3*01:01,HLA-DRB3*02:02	1006-1022	TYVTQQLIRAAEIRA	0.69
HLA-DRB1*07:01,HLA-DRB5*01:01,HLA-DRB1*03:01,HLA-DRB1*15:01,HLA-DRB4*01:01,HLA-DRB3*01:01,HLA-DRB3*02:02	1016-1030	AEIRASANLAATKMS	0.84
HLA-DRB1*15:01,HLA-DRB1*07:01,HLA-DRB5*01:01,HLA-DRB4*01:01,HLA-DRB1*03:01,HLA-DRB3*01:01,HLA-DRB3*02:02	1124-1138	GNCDVVIGIVNNTVY	0.92
HLA-DRB1*15:01,HLA-DRB5*01:01,HLA-DRB4*01:01,HLA-DRB3*01:01,HLA-DRB1*07:01,HLA-DRB1*03:01,HLA-DRB3*02:02	1201-1215	QELGKYEQYIKWPWY	0.87
HLA-DRB1*15:01,HLA-DRB1*07:01,HLA-DRB5*01:01,HLA-DRB4*01:01,HLA-DRB3*01:01,HLA-DRB1*03:01,HLA-DRB3*02:02	1212-1226	WPWYIWLGFIAGLIA	0.58



1a 1b

Figure 1. (a) PSIPRED secondary structure prediction of vaccine construct; (b) 3D structure of multiepitope vaccine construct (Swiss-Modeller) and energy minimized by Chimera.

Predictions of the linear and discontinuous B cell epitopes in vaccine construct

Linear as well as discontinuous B-cell epitopes in the final vaccine structure were predicted by Antibody epitope prediction tool and Discotope tool provided by IEDB server. Bepipred Linear Epitope Prediction tool predicted 5 potential conformational B cell epitope in our construct ^{26}KEEQIGKCSTR^{36}, ^{45}KEAAAKV^{51}, ^{97}GGGGSTY^{103}, ^{114}IRAGPGPGGNC^{124}, ^{135}VYGGGGSQELGKYEQYGPGPGKW^{157} (Figure 3(A–-E)) Emini Surface Accessibility Prediction suggested accessibility of three epitopes ^{5}TLQKYYC^{11}, ^{40}CCRRKKEA^{47}, ^{143}ELGKYEQYGPGPGKWP^{158} to the surface. Karplus & Schulz method predicted the flexible nature of the construct. In addition to this, Kolaskar & Tongaonkar suggest its antigenic nature by predicting antigenic peptide that are also same with the other prediction methods (^{8}KYYCRV^{13}, ^{16}GRCAVLSCLPK^{26}, ^{48}AAKVASQSIIAYAAYLL^{64}, ^{71}QYAAYIPF^{78}, ^{84}YAAYVLND^{91}, ^{124}CDVVIGI^{130}).

According to Ellipro server three different linear epitopes were found on the 3D structure of Vaccine construct ^{20}VLSCLPKEE^{28}, ^{1}GIINTLQKYYCRVR^{14}, ^{33}CSTRGR^{38} (Figure 4(A–-C)) and two discontinuous epitopes (V20, L21, S22, C23, L24, P25, K26, E27, C33, S34, T35, G37, R38 and G1, I2, I3, N4, T5, L6, Q7, K8, Y9, R12, G16) (Figure 4(D–E)). Both continuous and discontinuous epitopes having good pI score ranging from (0.94 to 0.87) suggesting their potential immunogenic nature. According to population coverage and conservation across antigen all the selected epitopes are 100% conserved in nature. IFN epitope server predicted six epitopes that can induce the interferons. All these parameters indicated that our construct is antigenic, immunogenic and accessible to the surface.

Predictions of binding of vaccine construct with toll like receptor 3 and 8

To confirm efficacy of the construct further, molecular docking was done by Cluspro where construct was docked against TLR3, and 8. Docked complex of construct against

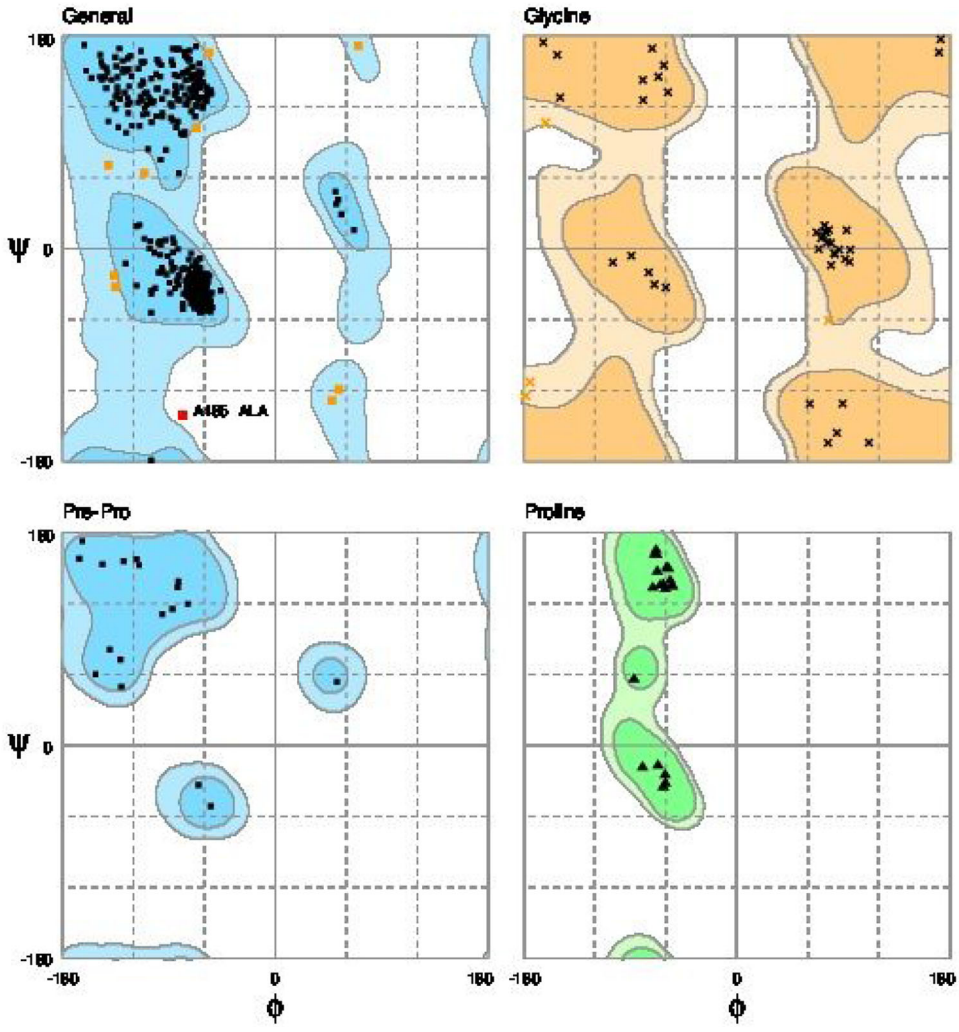


Figure 2. Ramachandran Plot for the energy minimized Vaccine construct that shows 90.4%, 6.6%, and3% of residues were located in the favored, allowed, and outlier regions.

TLR3 and 8 shows minimum energy of −1001, −865.5 J, respectively (Figure 5(a, b)). Protein–protein interaction of the vaccine construct and TLRs is shown in Figure 5(a′, b).

Codon optimization and insilico cloning

Insilico vaccine construction was done through codon optimization and *in-silico* cloning. Vaccine peptide sequence was reverse transcribed and converted in the nucleotide sequences by JCAT server followed by codon optimization. Codon Adaptive Index (CAI) $>$ 0.8 considered good, the optimal GC content between 30% and 70% and Codon Frequency Distribution (CFD) $<$ 30% are the important properties of gene sequence in order to achieve high expression level in the host. Optimized sequence was found to have 69% GC content and CAI score 0.95. Following codon optimization, vaccine optimized nucleotide sequences were subjected *in-isilico* cloning by Snapgene. Snapgene provides a list of vector to be used for cloning. pETvector was used for vaccine construction, nucleotide sequence of vaccine construct was added between the restriction site PpuMI and BstZ17I. The desired sequence was present in between mentioned restriction sites in the clone and represented by red color (Figure 6) with the final vaccine construct of 6274 bp.

Discussion

SARS-CoV2 is an emerging human CoV, and appears highly contagious as compared to previous SARS and MERS outbreaks. COVID 19 that has been declared as a pandemic and possess a great threat for global public health (World Health Organization, 2020b). After the outbreak of COVID-19 in China, SARS-CoV2 has spread to over 210+ countries. Worldwide, over 4 million people have been infected within four months. Socio-economic and psychological loses world has been suffering are devastating. Till date, no effective treatment or sustainable preventative measure for COVID 19 is available. Present analysis is a step to make an effective formulation to treat viral infection.

Despite of different vaccination strategies and trials that are being tried and tested none of them has reached to a level to stave off this pandemic. However, continuous efforts are being made to design an effective vaccine formulation against SARS-CoV. Spike protein RBD domain, Nucleocapsid, ORF3a, and Membrane protein have been

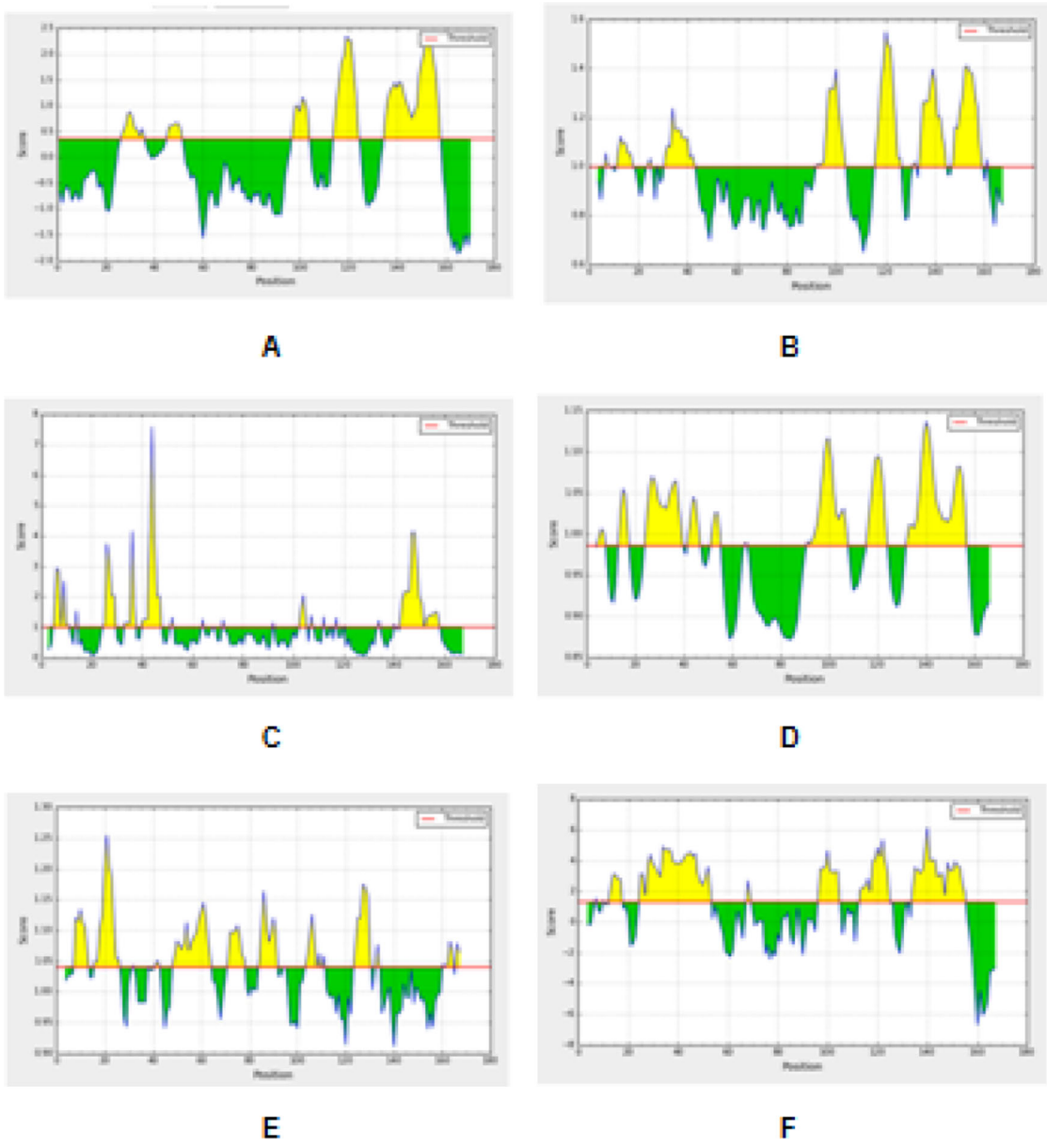


Figure 3. linear epitope properties prediction where (A) represents the Bepipred Linear epitope where yellow peaks correspond to the linear epitopes; (B) represents the Chou & Fasman Beta-Turn Prediction Results; (C) represents the Emini Surface Accessibility Prediction Results where yellow peaks corresponds to surface accessible residues; (D) represents the Karplus & Schulz Flexibility Prediction where yellow peaks nominates the flexible epitopes; (E) is Kolaskar & Tongaonkar Antigenicity Results and figure F is Parker Hydrophilicity Prediction Results showing approximately equal hydrophilic and hydrophobic residues.

studied for exploring their potential for vaccine development by using conventional and reverse vaccinology techniques (Peele et al., 2020; Enayatkhani et al., 2020; Srivastava et al., 2019). However, to the best of our knowledge, there is no prediction and vaccine construct available for S2 domain. To the best of our knowledge, there is no prediction and vaccine construct available for S2 domain. To know the conservation of S2 domain in different SARS species, Multiple sequence alignment was carried out in all reported seven SARS variants and it was found that approximately 97% sequence is conserved in different SARS groups. However, no NCBI conserved domain prediction is available for S2 domain. After thoroughly screening the viral genome and conservation analysis, Spike protein S2 was selected as a putative candidate for further analysis. This domain mediates fusion of the virion and cellular membranes by acting as a class I viral fusion protein (Tai et al., 2020; Wrapp et al., 2020). As per the proposed structure model on UniProt Server the protein has at least three conformational states: pre-fusion native state, pre-hairpin intermediate state, and post-fusion hairpin state. During viral and target cell membrane fusion, the coiled coil regions (heptad repeats) assume a trimer-of-hairpins structure, positioning the fusion peptide in close proximity to the C-terminal region of the ectodomain. The formation of this structure appears to drive apposition and subsequent fusion of viral and target cell membranes. Hence, it can be inferred that S2 domain plays a significant role in mediating cell entry (Tai et al., 2020). If the conformational changes are hampered, the infection can be constrained or we can suggest that if we disrupt the S2 domain, virus will not be able to fuse with the host cell membrane and

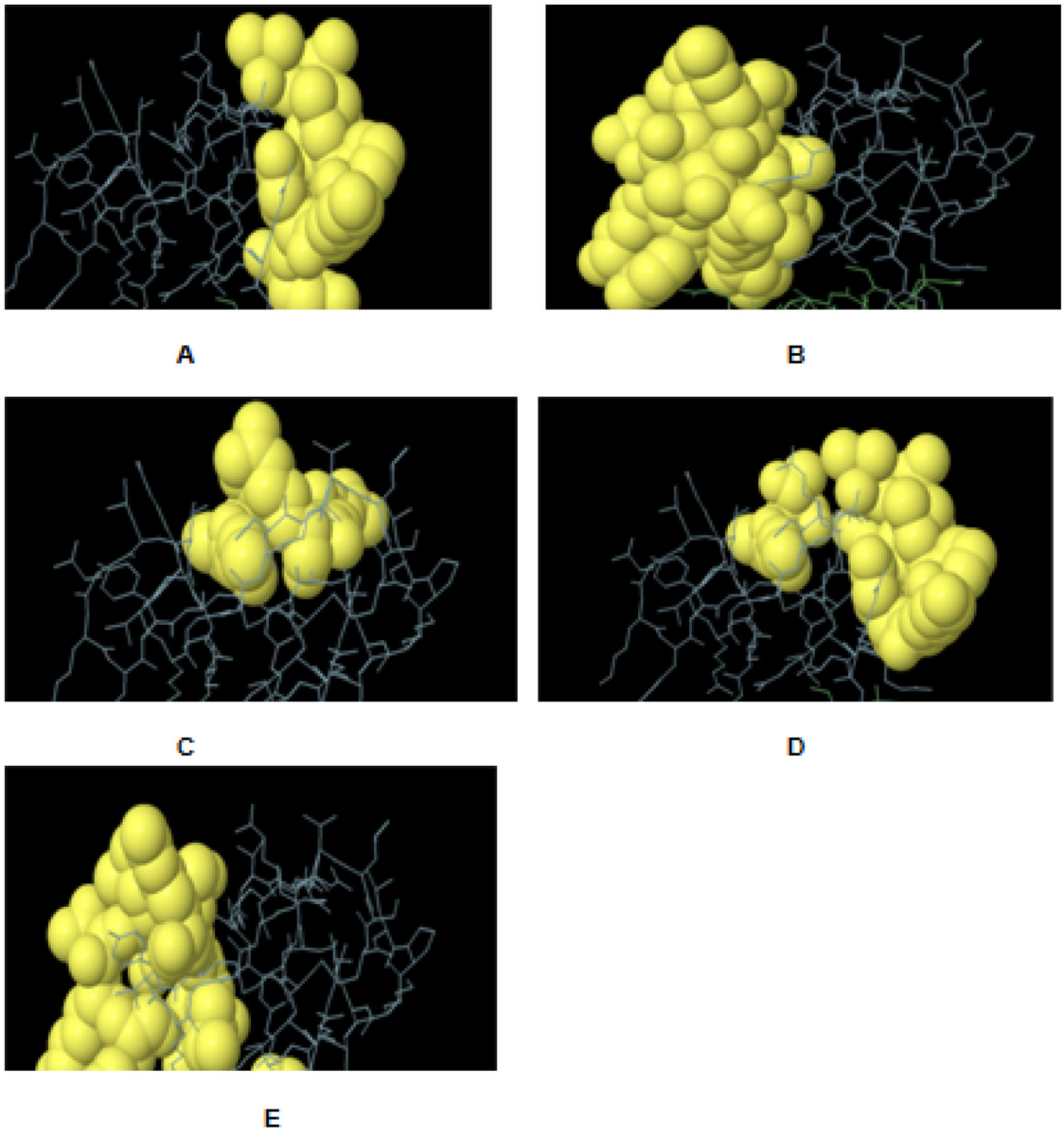


Figure 4. (A–E) Prediction of epitopes on 3D structure of vaccine construct where A corresponds to ^{20}VLSCLPKEE^{28}, B represents ^{1}GIINTLQKYYCRVR^{14},figure C represents the ^{33}CSTRGR^{38} whereas figure D& E are discountinuous epitopes that will be in direct contact with immune receptor where yellow surface is epitopes and sticks are the residual protein.

can be eliminated, therefore can prove to be a potential candidate of vaccine development.

Present vaccine construct was made up of a combination of Tcell and B cell epitopes to induce both cell mediated and humoral response. T cells (Helper T cell and Cytotoxic T cell) are the important mediator of cell mediated immunity. This is well known those helper T cells (T_H) are MHC-II restricted whereas cytotoxic T cells (T_C) are MHC-I restricted. Activation of humoral or cell mediated response depends on the nature of antigen. If an antigen is a short free peptide, is taken up into the B cell through receptor-mediated endocytosis, degraded, and presented to T cells as peptide pieces in complex with MHC-II molecules on the cell membrane and activate the T_H cells and ultimately humoral immune response In case of an intracellular pathogen, pathogen is degraded and displayed as a complex with MHC-I molecule on their cell surface. This Antigen-MHC-I complex induces the activation of T_C cells; cell mediated immune response. Therefore to induce a potential immune response it is necessary to identify those epitopes that are capable enough to activate both MHC-I ans MHC-II restricted cells. Probable epitopes for MHC I were identified by IEDB and RANKPEP servers. Top ten epitopes predicted by IEDB and top eight epitopes predicted by RANKPEP server were considered for further analysis. In addition to this CTL cell epitopes are also identified by CTLPred and PAComplex server to make more precise prediction of most antigenic antigens. The peptide sequences that pass most of the criteria were considered to be a good epitope candidate such as possessing antigenicity, non-allergen, and immunogenicity. Table 1 comprises the selected MHC-I restricted alleles showing all these epitopes are restricted to a number of Class I MHC allele i.e. HLA-A,B,C alleles. After prediction of MHC Class-I restricted epitopes, MHC-II restricted epitopes are also identified by IEDB and

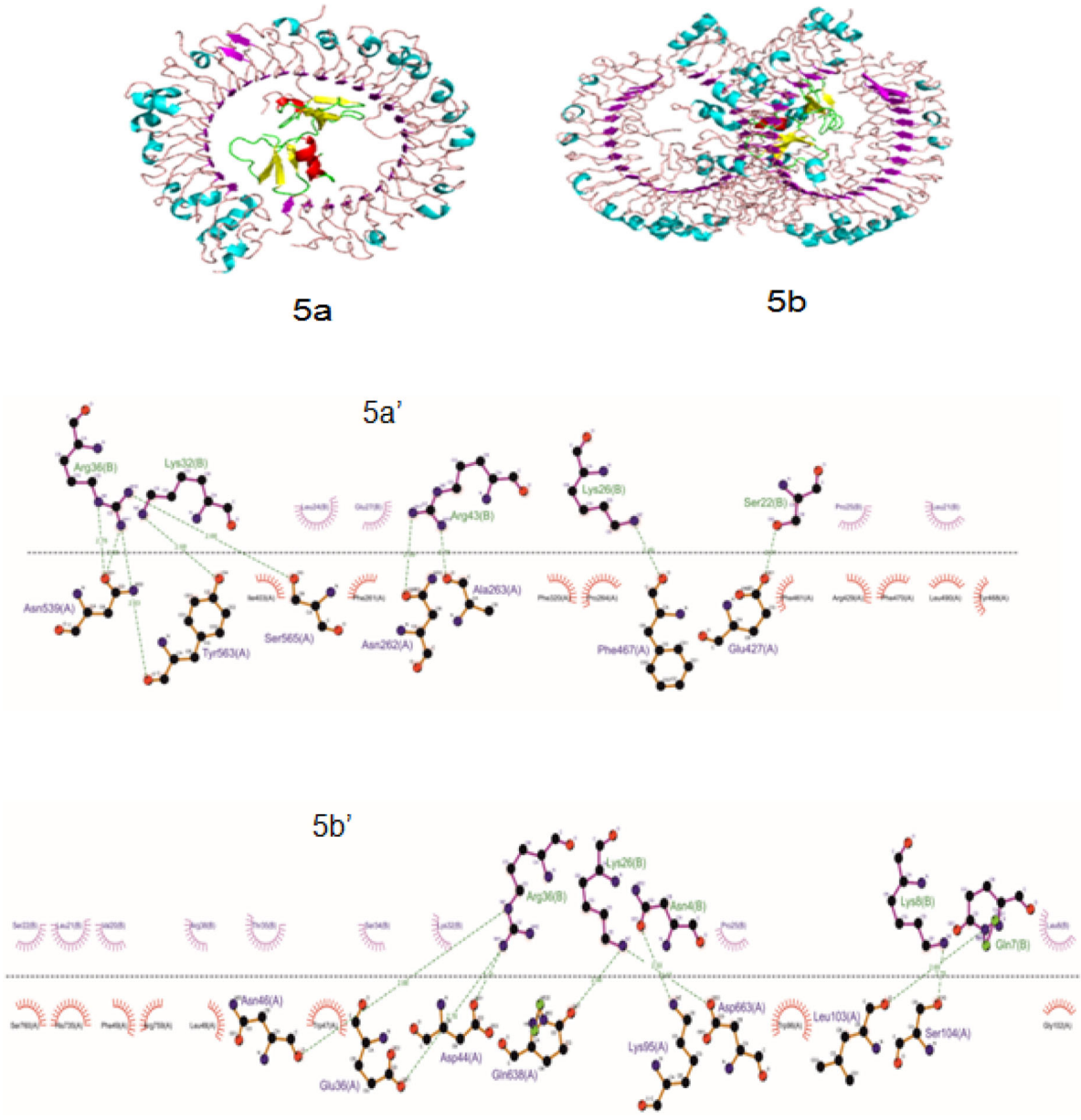


Figure 5. (a) represent the docked model for vaccine construct and TLR3 and (b) represent the docked model for vaccine construct and TLR8 where vaccine sequences are present in yellow, red and green color. (a' and b') showing the protein–protein interaction between vaccine construct TLR3 and TLR8, respectively.

RANKPEP server again. HLA allele are highly polymorphic i.e. can differ from population to population. Complete amino acid sequence of S subunit was screened for potential B cell epitopes by Bepipred prediction of B cell epitopes. According to the analysis, major antigenic epitopes are found in S2 domain, however ACE2 binding region was also found antigenic in S1 subunit. Epitopes related ACE2 binding region was not considered for further analysis, as there are already vaccine construct against ACE2 binding region (Tignanelli et al., 2020). All different HTL and CTL epitopes, those were predicted by various immuno-informatics approaches were analyzed for antigenicity, immunogenicity and scanned for an overlapping sequence with linear B-cell epitopes.

According to denominated MHC restriction, a given epitope will elicit a response only in individuals that express an MHC molecule capable of binding that particular epitope. The human hosts are genetically heterogeneous and express a different set of HLA. Due to this heterogeneity, HLA polymorphism influences the epitope-binding specificities with diverse T-cell epitope repertoires and therefore HLA cross-presentation ability of epitope play a crucial role to cope with this diversity. Therefore, it is important to have such antigen that have more than 80% HLA molecules coverage. To overcome that, population coverage analysis was done on the selected epitopes which predicted selected epitopes can cover 99.26% population worldwide for both MHC-I and II class. This data suggest that these epitopes are present in all the major population and ethnic groups and can elicit immune response in any population/ethnic group.

Selected epitopes are joined by linkers and adjuvant (N terminal). Adjuvant is an important player in generating immunity and increased the half life of vaccine construct. Toll like receptor 3and 8 specific adjuvant used in present

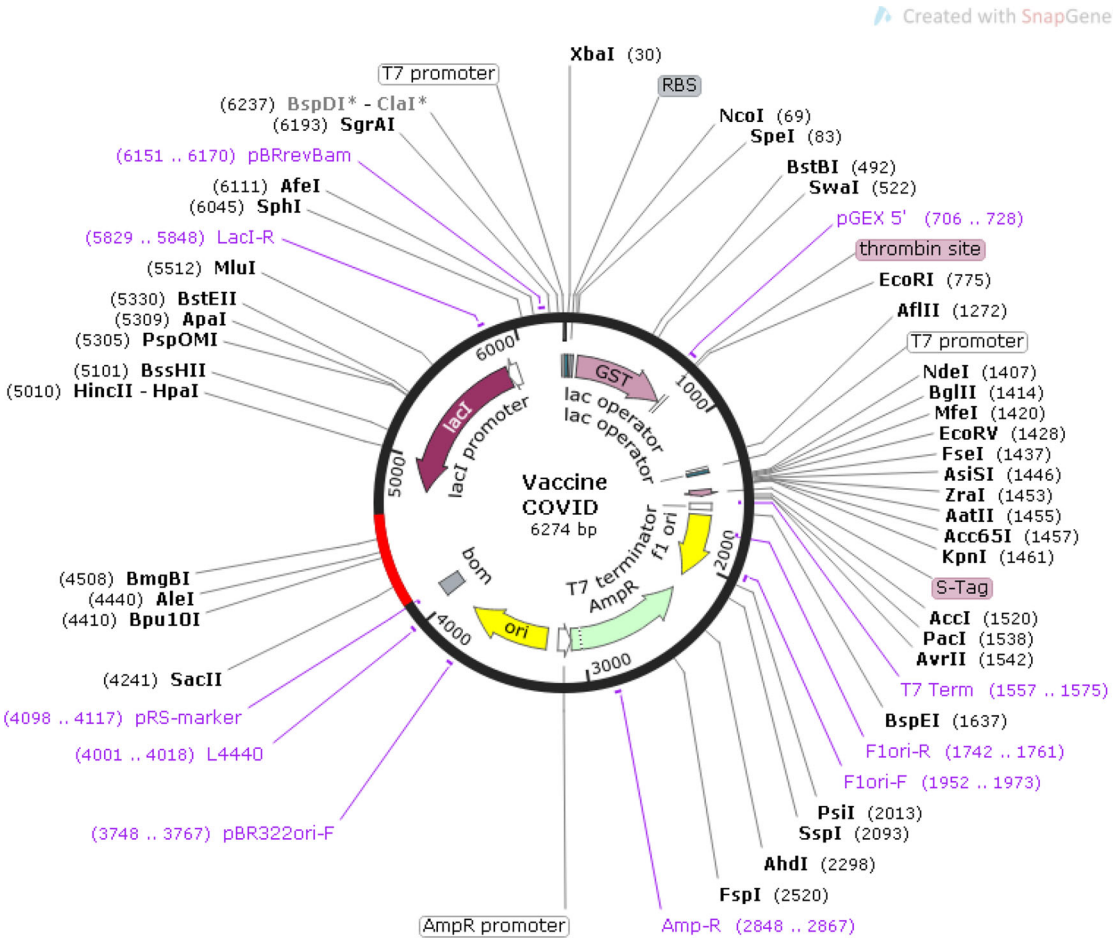


Figure 6. Cloned vaccine construct: optimized nucleotide sequence was cloned in pET vector by ligating the vaccine sequence between PpuMI and BstZ17I restriction sites.

vaccine construct which behaves like agoinst for these TLRs. In addition to adjuvant, linkers are the important part of vaccine construct as they have additional useful applications in fusion proteins, such as enhancing biological activities, improving expression yield, and obtaining desirable pharmacokinetic profiles. In designing our vaccine construct, EAAAK was used between the adjuvant (β-defensin) and epitope fragments to maintain a more effective separation and also their independent function. EAAAK linkers belongs to the rigid linker category and thermodynamically stable due to the formation of α-helical structures. Literature suggest that Naïve T-cell and immature dendritic cells are recruited by β-defensin at the site of infection through the CCR6 receptor, and provide an adaptive immune response, Insertion of linkers between two epitopes provides efficient separation which is required for the effective functioning of each epitope (Hajighahramani et al., 2019). Therefore in present analysis, we selected such adjuvant and linker that can enhance the antigenicity and immunogenicity of our construct.

After ligation of all the epitopes with adjuvant and linkers, physiochemical properties of construct were also analyzed. Our vaccine construct is comprised of 170 amino acids and having a slightly basic isoelectric pH 9.27. Analysis data suggests that synthetic construct is alkaline, thermostable (as instability index is less than 40), having mixed composition of hydrophilic and hydrophobic amino acids that make it survive in the complex environment of human body. Allergenicity prediction indicated that vaccine construct is non-allergen, therefore we can suggest that the protein vaccine does not cause any type of allergic response and inflammation and might be safe for the human use. Antigenicity prediction indicates the construct is highly antigenic and have the potential to generate immune response. By all the analysis and predictions we can suggest that present multiepitope vaccine construct is highly immunogenic, antigenic, non allergen and thermodynamically stable in human physiological environment and can elicit both humoral and cell mediated immunity.

Functionality of protein is highly dependent on its three dimensional structure. This is also evident that antigenicity also depend on the complexity of antigen, primary structure of antigen are least antigenic when compared to its three dimensional arrangement. Amino acid sequence of vaccine construct was deployed for its 3D structure determination and best selected 3D model was energy minimized and validated by various tools. Three dimensional structure also comprised some region that are specifically bind with other residues. So for an antigen to induce immunogenicity, its binding with the receptor that can induce immune response is crucial. These receptor are basically Toll like receptors (TLRs). To study the binding with TLRs, vaccine construct was docked against TLR 3 and 8. Docked complex showed

interaction of construct with TLR and probable binding sites as shown in Figure 5. Docked scores suggest the potential binding between the construct and TLRs and its ability to activate innate immunity.

3D model of vaccine construct was analyzed for the presence of linear and discontinuous B cell epitopes. Linear epitopes are target for both T cell (primarily) and B cell (least). Majority of B cell epitopes are conformational in nature i.e. they are not present in primary structure however after folding some residues comes in close proximity and contribute as epitopes (Mukonyora, 2015). Discontinuous B cell epitope prediction reported the presence of two conformational epitopes that are specific for B cell and can probably elicit humoral response. Parker Hydrophilicity analysis predicted epitopes are hydrophilic in nature. In addition to that, Karplus & Schulz analysis predicted the such epitopes that are more flexible than the other residues. According to a study, B-cell epitopes tend to be more exposed to solvent than their surrounding surface-exposed residues. This high surface-exposure of antigenic regions makes the epitopes highly flexible (Novotný et al., 1986). Flexibility of epitopes is crucial to accommodate the conformational changes that are the resultant of epitopes binding with antibodies (Mukonyora, 2015). Moreover, flexibility is the primary requirement for an antigenic epitope when talking about the plasticity of the CDRs (complementarity determination regions). By all the predictions carried out for B cell epitope prediction we can suggest that present construct might be capable enough to behave like a potent B cell stimulator and could generate humoral response as well.

For determining the potential use of vaccine *invitro, insilico* cloning of vaccine construct was prepared for the expression of vaccine in host cell lines. Amino acid sequence was reverse transcribed into nucleotide and optimized. Optimized product is having high CAI (codon adaptive index). Higher CAI value, higher gene expression therefore this data suggests the maximum efficacy of gene expression in the host human cell lines.

In this investigation, through various immunoinformatics tools, a multivalent subunit vaccine against pathogenic SARS COVID19 was designed. Chimeric vaccine constructed is predicted to be capable of stimulating both cellular and humoral immune responses effectively due to the presence of both B cell and T cell epitopes along with adjuvant. Thus, our designed vaccine is suggested as a potential candidate for COVID19 infections. Future in vitro synthesis of the vaccine construct and in vivo immunological tests are needed to verify the efficacy of the designed vaccine.

Conclusion

Unparalleled losses to human life and world economy witnessed by mankind makes it axiomatic to dedicate our current resources with full acceleration towards combating COVID 19. It is evident that virus that was initially zoonotic has evolved and is now horizontally transmitting in human population at exponential rate thereby causing a huge threat to mankind. Developing a safe and effective novel vaccine expeditiously against this newly evolved and highly contagious SARS-CoV2 is of utmost importance to control the losses. Our *in-silico* study suggests the that a vaccine against fusion domain of virus has potential to activate not just B cells but also T cells and IFN γ by which we can and control/limit the entry of virus inside the host human cell. Our construct was found to very effective in generating innate as well as humoral immune response so might be considered for future studies. However, *in-vivo* immunological studies are required to prove the effectiveness of proposed vaccine construct.

Acknowledgements

The authors acknowledge Centre for Medical Biotechnology, Maharshi Dayanand University, Rohtak for providing research work platform.

Funding

Not applicable.

Disclosure statement

The authors report no declarations of interest.

Ethical approval

Not applicable.

Informed consent

Not applicable.

ORCID

Amita S. Dang http://orcid.org/0000-0002-5015-3502

References

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature Medicine*, *26*(4), 450–452. https://doi.org/10.1038/s41591-020-0820-9

Peele, K. A., Srihansa, T., Krupanidhi, S., Sai, A. V., & Venkateswarulu, T. C. (2020). Design of multi-epitope vaccine candidate against SARS-CoV-2: A In-Silico study. *Journal of Biomolecular Structure & Dynamics*, 1–10. https://doi.org/10.1080/07391102.2020.1770127

Bhasin, M., & Raghava, G. P. S. (2004). Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine*, *22*(23-24), 3195–3204. https://doi.org/10.1016/j.vaccine.2004.02.005

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science: A Publication of the Protein Society*, *2*(9), 1511–1519. https://doi.org/10.1002/pro.5560020916

Dhanda, S. K., Vir, P., & Raghava, G. P. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct*, *8*(1), 30. https://doi.org/10.1186/1745-6150-8-30

Dimitrov, I., Naneva, L., Doytchinova, I., & Bangov, I. (2014). AllergenFP: Allergenicity prediction by descriptor fingerprints. *Bioinformatics (Oxford, England))*, *30*(6), 846–851. https://doi.org/10.1093/bioinformatics/btt619

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

Enayatkhani, M., Hasaniazad, M., Faezi, S., Guklani, H., Davoodian, P., Ahmadi, N., Karmostaji, A., Ahmadi, K. (2020). Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: An in silico study. *Journal of Biomolecular Structure and Dynamics*, 1–16. https://doi.org/10.1080/07391102.2020.1756411

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571–607). Humana Press. https://doi.org/10.1385/1-59259-890-0:571

Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., & Gulyaeva, A. A. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature* Microbiology, *5*, 536–544. https://doi.org/10.1038/s41564-020-0695-z

Grasselli, G., Zangrillo, A., Zanella, A., Antonelli, M., Cabrini, L., Castelli, A., Cereda, D., Coluccello, A., Foti, G., Fumagalli, R., Iotti, G., Latronico, N., Lorini, L., Merler, S., Natalini, G., Piatti, A., Ranieri, M. V., Scandroglio, A. M., Storti, E., Cecconi, M., & Pesenti, A., for the COVID-19 Lombardy ICU Network (2020). Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region. *JAMA*, *323*(16), 1574–1581. https://doi.org/10.1001/jama.2020.5394

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

Hajighahramani, N., Eslami, M., Negahdaripour, M., Ghoshoon, M. B., Dehshahri, A., Erfani, N., Heidari, R., Gholami, A., Nezafat, N., & Ghasemi, Y. (2019). Computational design of a chimeric epitope-based vaccine to protect against Staphylococcus aureus infections. *Molecular and Cellular Probes*, *46*, 101414. https://doi.org/10.1016/j.mcp.2019.06.004

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

Kim, Y., Ponomarenko, J., Zhu, Z., Tamang, D., Wang, P., Greenbaum, J., Lundegaard, C., Sette, A., Lund, O., Bourne, P. E., Nielsen, M., & Peters, B. (2012). Immune epitope database analysis resource. *Nucleic Acids Research*, *40*(Web Server issue), W525–W530. https://doi.org/10.1093/nar/gks438

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

Le, T. T., Andreadakis, Z., Kumar, A., Roman, R. G., Tollefsen, S., Saville, M., & Mayhew, S. (2020). The COVID-19 vaccine development landscape. *Nature Reviews. Drug Discovery*, *19*(5), 305–306. https://doi.org/10.1038/d41573-020-00073-5

Liu, I. H., Lo, Y. S., & Yang, J. M. (2011). PAComplex: A web server to infer peptide antigen families and binding models from TCR-pMHC complexes. *Nucleic Acids Research*, *39*(Web Server issue), W254–W260. https://doi.org/10.1093/nar/gkr434

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., & Richardson, D. C. (2003). Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, *50*(3), 437–450. https://doi.org/10.1002/prot.10286

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., … Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *The Lancet*, *395*(10224), 565–574. https://doi.org/10.1016/S0140-6736(20)30251-8

Magnan, C. N., & Baldi, P. (2014). SSpro/ACCpro 5: Almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. *Bioinformatics (Oxford, England))*, *30*(18), 2592–2597. https://doi.org/10.1093/bioinformatics/btu352

Magnan, C. N., Randall, A., & Baldi, P. (2009). SOLpro: Accurate sequence-based prediction of protein solubility. *Bioinformatics (Oxford, England))*, *25*(17), 2200–2207. https://doi.org/10.1093/bioinformatics/btp386

Mukonyora, M. (2015). A review of important discontinuous B-cell epitope prediction tools. *Journal of Clinical and Cellular Immunology*, *6*, 358–362. https://doi.org/10.4172/2155-9899.1000358

Novotný, J., Handschumacher, M., Haber, E., Bruccoleri, R. E., Carlson, W. B., Fanning, D. W., Smith, J. A., & Rose, G. D. (1986). Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains). *Proceedings of the National Academy of Sciences of the United States of America*, *83*(2), 226–230. https://doi.org/10.1073/pnas.83.2.226

Ong, E., Wong, M. U., Huffman, A., & He, Y. (2020). COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. BioRxiv. https://doi.org/10.1101/2020.03.20.000141

Park, C. Y., Villafuerte, J., Abiad, A. (2020). Updated assessment of the potential economic impact of COVID-19 Asian Development Bank; https://www.adb.org/publications/updated-assessment-economic-impact-covid-19.. https://doi.org/10.22617/BRF200144-2

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera-a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Pronker, E. S., Weenen, T. C., Commandeur, H., Claassen, E. H., & Osterhaus, A. D. (2013). Risk in vaccine research and development quantified. *PloS One*, *8*(3), e57755. https://doi.org/10.1371/journal.pone.0057755

Reche, P. A., & Reinherz, E. L. (2007). Prediction of peptide-MHC binding using profiles. In *Immunoinformatics* (pp. 185–200). Humana Press. https://doi.org/10.1007/978-1-60327-118-9_13

Saha, S., & Raghava, G. P. S. (2006). AlgPred: Prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Research*, *34*(Web Server issue), W202–W209. https://doi.org/10.1093/nar/gkl343

Srivastava, S., Kamthania, M., Kumar Pandey, R., Kumar Saxena, A., Saxena, V., Kumar Singh, S., Kumar Sharma, R., & Sharma, N. (2019). Design of novel multi-epitope vaccines against severe acute respiratory syndrome validated through multistage molecular interaction and dynamics. *Journal of Biomolecular Structure & Dynamics*, *37*(16), 4345–4360. https://doi.org/10.1080/07391102.2018.1548977

Tai, W., He, L., Zhang, X., Pu, J., Voronin, D., Jiang, S., Zhou, Y., & Du, L. (2020). Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: Implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cellular & Molecular Immunology*, *17*(6), 613–620. https://doi.org/10.1038/s41423-020-0400-4

Tignanelli, C. J., Ingraham, N. E., Sparks, M. A., Reilkoff, R., Bezdicek, T., Benson, B., Schacker, T., Chipman, J. G., & Puskarich, M. A. (2020). Antihypertensive drugs and risk of COVID-19. *The Lancet. Respiratory Medicine*, *8*(5), e30–e31. https://doi.org/10.1016/S2213-2600(20)30153-3

UniProt Consortium. (2007). The universal protein resource (UniProt). *Nucleic Acids Research*, *36*(suppl_1), D190–D195. https://doi.org/10.1093/nar/gkm895 https://doi.org/10.1016/S2213-2600(20)30153-3

von Tigerstrom, B., & Wilson, K. (2020). COVID-19 travel restrictions and the International Health Regulations (2005). *BMJ Global Health*, *5*(5), e002629. https://doi.org/10.1136/ bmjgh-2020-002629 https://doi.org/10.1136/bmjgh-2020-002629

Wachyuni, S. S., & Kusumaningrum, D. A. (2020). The Effect of COVID-19 Pandemic: How are the Future Tourist Behavior? *Journal of Education, Society and Behavioural Science*, *33*(4), 67–76. https://doi.org/10.9734/jesbs/2020/v33i430219

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, *181*(2), 281–292.e6. https://doi.org/10.1016/j.cell.2020.02.058

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L.,

Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(Web Server issue), W407–W410. https://doi.org/10.1093/nar/gkm290

World Health Organization. (2020a). Solidarity" clinical trial for COVID-19 treatments. *World Health Organization (WHO). Situation reports. Geneva: WHO*. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/solidarity-clinical-trial-for-covid-19-treatments.

World Health Organization. (2020b). *Protocol for assessment of potential risk factors for coronavirus disease 2019 (COVID-19) among health workers in a health care setting,* (23 March 2020). (No. WHO/2019-nCoV/HCW_risk_factors_protocol/2020.3).

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., Graham, B. S., & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (New York, N.Y.).)*, *367*(6483), 1260–1263. https://doi.org/10.1126/science.abb2507

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome. *Nature*, *579*(7798), 265–269. https://doi.org/10.1038/s41586-020-2008-3

Xia, S., Liu, M., Wang, C., Xu, W., Lan, Q., Feng, S., Qi, F., Bao, L., Du, L., Liu, S., Qin, C., Sun, F., Shi, Z., Zhu, Y., Jiang, S., & Lu, L. (2020). Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Research*, *30*(4), 343–355. https://doi.org/10.1038/s41422-020-0305-x

Zhang, H., Wang, G., Li, J., Nie, Y., Shi, X., Lian, G., Wang, W., Yin, X., Zhao, Y., Qu, X., Ding, M., & Deng, H. (2004). Identification of an antigenic determinant on the S2 domain of the severe acute respiratory syndrome coronavirus spike glycoprotein capable of inducing neutralizing antibodies. *Journal of Virology*, *78*(13), 6938–6945. https://doi.org/10.1128/JVI.78.13.6938-6945.2004