JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
https://doi.org/10.1080/07391102.2020.1850357

Taylor & Francis
Taylor & Francis Group

EXPRESS COMMUNICATION

Design of a multi-epitope-based vaccine targeting M-protein of SARS-CoV2: an immunoinformatics approach

Vijaya Sai Ayyagari, Venkateswarulu T. C., Abraham Peele K. and Krupanidhi Srirama

Department of Biotechnology, Vignan's Foundation for Science, Technology & Research, Vadlamudi, Guntur, Andhra Pradesh, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

In the present study, one of the targets present on the envelopes of coronaviruses, membrane glycoprotein (M) was chosen for the design of a multi-epitope vaccine by Immunoinformatics approach. The B-cell and T-cell epitopes used for the construction of vaccine were antigenic, nonallergic and nontoxic. An adjuvant, β-defensin and PADRE sequence were included at the N-terminal end of the vaccine. All the epitopes were joined by linkers for decreasing the junctional immunogenicity. Various physicochemical parameters of the vaccine were evaluated. Secondary and tertiary structures were predicted for the vaccine construct. The tertiary structure was further refined, and various parameters related to the refinement of the protein structure were validated by using different tools. Humoral immunity induced by B-cells relies upon the identification of antigenic determinants on the surface of the vaccine construct. In this regard, the vaccine construct was found to consist of several B-cell epitopes in its three-dimensional conformation. Molecular docking of the vaccine was carried out with TLR-3 receptor to study their binding and its strength. Further, protein–protein interactions in the docked complex were visualized using LigPlot+. Population coverage analysis had shown that the multi-epitope vaccine covers 94.06% of the global population. The vaccine construct was successfully cloned *in silico* into pET-28a (+). Immune simulation studies showed the induction of primary, secondary and tertiary immune responses marked by the increased levels of antibodies, INF-γ, IL-2, TGF-β, B- cells, CD4+ and CD8+ cells. Finally, the vaccine construct was able to elicit immune response as desired.

ARTICLE HISTORY
Received 4 August 2020
Accepted 23 October 2020

KEYWORDS
COVID-19; SARS-CoV2; membrane glycoprotein; immunoinformatics; multi-epitope vaccine

1. Introduction

The outbreak of Coronavirus disease − 19 (COVID-19) is caused by the Severe Acute Respiratory Syndrome – related Coronavirus 2 (SARS-CoV2) has brought the world to a standstill (Dagur & Dhakar, 2020). The outbreak of SARS-CoV2 had its origin in Wuhan, China (Prasad & Prasad, 2020). It has now spread to over 216 countries. The number of confirmed cases crossed 29 million, with more than 0.9 million confirmed deaths reported worldwide (https://covid19.who.int/ accessed on 16 September 2020). World Health Organization characterized COVID-19 as a pandemic on 11 March 2020 (https://www.who.int/news-room/detail/27-04-2020-who-timeline—covid-19/ accessed on 23 July 2020). COVID-19 outbreak is the result of transmission of the virus to humans from animals (Gorbalenya et al., 2020). There is neither a drug nor a vaccine available for SARS-CoV2 (Prasad & Prasad, 2020).

SARS-CoV2 belongs to the genus – Betacoronavirus, Subfamily – Orthocoronavirinae, Family – Coronaviridae, Suborder – Cornidovirineae and Order – Nidovirales. MERS-CoV and SARS-CoV also belongs to the same genus Betacoronavirus (Gorbalenya et al., 2020). The genome of the SARS-CoV2 is a positive sense single stranded RNA that acts as mRNA immediately after its release into the host cell. The genome of coronavirus is approximately 20 kb–30 kb. Two-third of the genome of Coronavirus towards 5′ - end is made up of two overlapping open reading frames, ORF1a and ORF1b that codes for the synthesis of several nonstructural proteins. ORF1a is translated and results in the synthesis of ORF1a polyprotein. However, due to ribosomal frame shifting, the ribosome continues to translate the ORF1b resulting in the synthesis of ORF1ab polyprotein. Cleavage of the mature peptide of ORF1ab gives rise to the following peptides – nsp2 to nsp4, nsp6 to nsp11, 3C-like proteinase, RNA-dependent RNA polymerase, helicase, 3′- to 5′-exonuclease, endoRNAse and 2′-O-ribose methyltransferase. Remaining one-third of the genome of Coronavirus towards 3′-end is made up of structural proteins consisting of Surface Glycoprotein (S), Envelope (E), Membrane Glycoprotein (M) and Nucleocapsid phosphoprotein (N). Interspersed between these four structural genes are ORF3a, ORF6, ORF7a–ORF7b, ORF8 and ORF10 genes (NCBI Accession MT434760) (Graham et al., 2008; Joshi et al., 2020; Kim et al., 2020; Prasad & Prasad, 2020).

Membrane glycoprotein is the most abundant protein in the Coronaviruses (Rottier, 1995). As the name indicates, the 'M' protein is embedded in the viral lipid bilayer. The N- terminal end is exposed outside and the C- terminal end is present inside the virus with three transmembrane helices embedded in the lipid bilayer (Bianchi et al., 2020; Kumar

CONTACT Venkateswarulu T. C. venki_biotech327@yahoo.com Department of Biotechnology, Vignan's Foundation for Science, Technology & Research (Deemed to be University), Vadlamudi, Guntur, Andhra Pradesh 522213, India; Krupanidhi S. krupanidhi.srirama@gmail.com Department of Biotechnology, Vignan's Foundation for Science, Technology & Research (Deemed to be University), Vadlamudi, Guntur, Andhra Pradesh 522213, India.

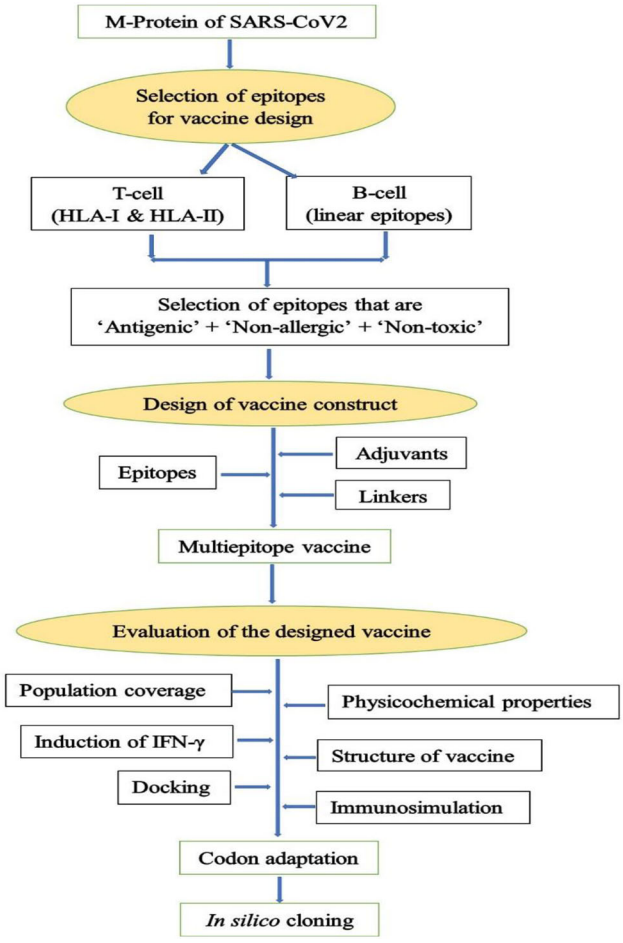

Figure 1. Flowchart summarizing the steps involved in the rational design of vaccine.

et al., 2020). It is an important glycoprotein that alone plays an important role in intracellular budding (Kumar et al., 2020; Rottier, 1995; Tang et al., 2009). It is different from the remaining glycoproteins of Coronavirus in terms of its structure and function (Rottier, 1995). It is flanked on its 5′-end by gene 'E' coding for Envelope protein and towards 3′ – end by gene ORF6 (NCBI Accession MT434760).

Vaccines provide protection from infectious diseases (Thomas & Luxon, 2013). For the successful development of a vaccine candidate till its release into the market takes considerable amount of time as it needs to qualify in different phases in the clinical trials along with the huge investment involved. However, at times, the rate of success of the vaccine is also bleak. In such a scenario, Immunoinformatics approach is helpful in reducing the time taken for the development of vaccines compared with the conventional approaches and there are examples of the successful development of vaccine candidates by adopting *in silico* approaches (Ada et al., 2018; Bhatnager et al., 2020; He et al., 2010; Kim et al., 2019; Oli et al., 2020; Patronov & Doytchinova, 2013; Tordello et al., 2017 and the appropriate references cited therein). The whole genome sequence of the SARS-CoV2 is made available in the public domain by next generation sequencing. This enabled the design of several multi-epitope *in silico* vaccine candidates targeting different structural and nonstructural proteins of the SARS-CoV2 (Bhatnager et al., 2020; Bhattacharya et al., 2020; Devi & Chaitanya, 2020; Dong et al., 2020; Enayatkhani et al., 2020; Kar et al., 2020; Lizbeth et al., 2020; Peele et al., 2020; Samad et al., 2020). In the present study, we aimed to design a novel multi-epitope vaccine targeting membrane glycoprotein by immunoinformatics approach.

2. Methods

A flowchart summarizing the steps involved in the design of the multi-epitope vaccine in the present study is shown in Figure 1.

2.1. Retrieval of nucleotide sequences from NCBI GenBank

Amino acid sequences of the Membrane glycoprotein corresponding to different geographical regions with the following accession numbers viz. MT434760, MT259226, MT066156, MT292577 and MT339041 were retrieved from NCBI GenBank. Multiple sequence alignment of the retrieved sequences was carried out online using Clustal Omega available at https://www.ebi.ac.uk/Tools/msa/clustalo/ to find out the extent of similarity among the sequences. It was observed that the extent of similarity among all the accessions is 100%. Subsequently, 222 amino acid sequences were used for the design of multi-epitope vaccine. Beta- Defensin 3 amino acid sequence was retrieved from UniProt database (Q5U7J2).

2.2. HLA-1 & II molecules utilized in the present study

A total of 27 HLA-I alleles and 27 HLA-II alleles (HLA-A*01:01, HLA-A*02:01, HLA-A*02:03, HLA-A*02:06, HLA-A*03:01, HLA-A*11:01, HLA-A*23:01, HLA-A*24:02, HLA-A*26:01, HLA-A*30:01, HLA-A*30:02, HLA-A*31:01, HLA-A*32:01, HLA-A*33:01, HLA-A*68:01, HLA-A*68:02, HLA-B*07:02, HLA-B*08:01, HLA-B*15:01, HLA-B*35:01, HLA-B*40:01, HLA-B*44:02, HLA-B*44:03, HLA-B*51:01, HLA-B*53:01, HLA-B*57:01, HLA-B*58:01; HLA-DRB1*01:01, HLA-DRB1*03:01, HLA-DRB1*04:01, HLA-DRB1*04:05, HLA-DRB1*07:01, HLA-DRB1*08:02, HLA-DRB1*09:01, HLA-DRB1*11:01, HLA-DRB1*12:01, HLA-DRB1*13:02, HLA-DRB1*15:01, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB4*01:01, HLA-DRB5*01:01, HLA-DQA1*05:01/DQB1*02:01, HLA-DQA1*05:01/DQB1*03:01, HLA-DQA1*03:01/DQB1*03:02, HLA-DQA1*04:01/DQB1*04:02, HLA- DQA1*01:01/DQB1*05:01, HLA-DQA1*01:02/DQB1*06:02, HLA-DPA1*02:01/DPB1*01:01, HLA-DPA1*01:03/DPB1*02:01, HLA-DPA1*01:03/DPB1*04:01, HLA-DPA1*03:01/DPB1*04:02, HLA-DPA1*02:01/DPB1*05:01, HLA-DPA1*02:01/DPB1*14:01) were used in the present study for predicting the affinities of different peptides derived from the membrane glycoprotein of SARS-CoV2 for HLA - I & II. These alleles are considered to occur more frequently than others. (https://help.iedb.org/hc/en-us/articles/114094151851 accessed on 05 July 2020 and the appropriate references cited therein).

2.3. Prediction of HLA-I restricted (CTL) epitopes

MHC class-I Epitopes were identified in the present study using NetCTLpan – 1.1 (https://services.healthtech.dtu.dk/service.php?NetCTLpan-1.1) (Stranzl et al., 2010). It is

considered to outperform any other CTL epitope predictor tools. Apart from this, it is more efficient in predicting new HLA-I/CTL epitopes compared to NetMHCpan and NetCTL tools (Stranzl et al., 2010). The output of the programme gives prediction scores for MHC- I binding affinity, TAP transport efficiency, Proteasomal C terminal cleavage, Ligand combined score and Percentage rank. Threshold for epitope identification was set at its default value of 1.0. The peptides whose %Rank was less than 1.0 were considered qualified for further analyses.

2.4. Prediction of HLA-II restricted epitopes

NetMHCIIpan–4.0 (https://services.healthtech.dtu.dk/service.php?NetMHCIIpan-4.0) was used in the present study to predict the peptides that bind to a HLA-II molecule using Artificial Neural Networks. NetMHCIIpan is superior in terms of prediction of ligand compared to any of the existing T_H cell epitope prediction tools (Reynisson et al., 2020). The output of the programme gives information on the peptide sequence, core peptide region, eluted ligand prediction score, Percentile rank of eluted ligand prediction score and Bind level (binding affinity of the peptide). Threshold for strong binding peptides was set to default value. Only strong binding peptides, whose %Rank was less than the threshold value of 2.0, were included in the present study.

2.5. Population coverage analysis

HLA genotype frequencies vary across different populations in the world. The nature of HLA polymorphism has its influence on the binding of a given peptide derived from the multi-epitope vaccine to HLA-I/II molecules (Bui et al., 2006). Therefore, it is important to know the extent of coverage of world population by the multi-epitope vaccine constructed in the present study. To determine this, IEDB Analysis Resource available at http://tools.iedb.org/population/ (Bui et al., 2006) was used. A comprehensive account on the fundamental principles of the tool is available at http://tools.iedb.org/population/help/for further reading. In the present study, population (World) coverage analysis was carried for HLA class I and class II combined. A total of eight T-cell epitopes as well as their corresponding HLA alleles were given as input.

2.6. Prediction of B-cell epitopes

Linear B-cell epitopes were predicted using BepiPred-2.0 (https://services.healthtech.dtu.dk/service.php?BepiPred-2.0) at the default threshold of 0.5 (Jespersen et al., 2017). Only peptides which are of a minimum of 10 amino acids long are considered for further analyses.

2.7. Prediction of antigenicity, allergenicity and toxicity of HLA – I, II & B-cell epitopes

Initially each of the peptide was analyzed for important parameters viz. antigenicity, allergenicity and toxicity. Antigenicity of the peptides was determined using VaxiJen v2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html). Allergenicity of the peptides was determined using AllerTOP v2.0 (https://www.ddg-pharmfac.net/AllerTOP/) (Dimitrov et al., 2014) and Toxicity of the peptides was determined using ToxinPred (https://webs.iiitd.edu.in/raghava/toxinpred/design.php) (Gupta et al., 2013). Determination of all the above three properties for the individual peptides as well as for the entire vaccine construct was carried out at default settings (unless stated otherwise). However, for B-cell epitopes, the threshold for VaxiJen was set to 0.4 to accommodate for a greater number of B-cell epitopes in the final vaccine construct.

2.8. Construction of multi-epitope vaccine

For the construction of a multi-epitope vaccine, β - defensin was used as an adjuvant. PADRE sequence was also included in the final vaccine construct. The following four linkers viz. EAAAK, GPGPG, AAY and KK were used for joining the epitopes. EAAAK was used for joining adjuvant and PADRE sequence. CTL, HTL and B-cell epitopes were joined by AAY, GPGPG and KK, respectively. A His6 tag was added to the c-terminal end of the vaccine construct to aid in purification.

2.9. Interferon-γ inducing epitope prediction

Interferon-γ (IFN-γ) inducing ability of the vaccine was determined using IFNepitope (https://webs.iiitd.edu.in/raghava/ifnepitope/scan.php) (Dhanda et al., 2013). The online server generates overlapping peptides (15 amino acids long) and predicts the IFN-γ inducing ability of each of the peptide generated. 'Motif and SVM hybrid' method of prediction was selected in the 'Scan' module.

2.10. Evaluation of physicochemical properties of the vaccine

ProtParam (https://web.expasy.org/protparam/) Gasteiger et al. (2005) was used to predict the physicochemical properties of the vaccine viz. molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).

2.11. Prediction of discontinuous B-cell epitopes in the final vaccine construct

Linear and discontinuous B-cell epitopes in the final vaccine construct were predicted using BepiPred − 2.0 (https://services.healthtech.dtu.dk/service.php?BepiPred-2.0) and DiscoTope v2.0 (http://www.cbs.dtu.dk/services/DiscoTope/ & http://tools.iedb.org/discotope/) (Kringelum et al., 2012), respectively. Default threshold value of 0.5 and −3.7 were used for the prediction of epitopes in BepiPred and DiscoTope, respectively.

For the prediction of discontinuous B-cell epitopes using DiscoTope, three dimensional structure of the vaccine construct in '.PDB' format was given as an input. The output of

the programme consists of chain id, residue number, amino acid, contact number, propensity score, DiscoTope score and identified B cell epitope (shown against the residue).

2.12. Prediction of the secondary structure

Secondary structure of the vaccine construct was predicted using PSIPRED 4.0 web server (http://bioinf.cs.ucl.ac.uk/psipred/) (Buchan & Jones, 2019; Jones, 1999).

2.13. Prediction, refinement and validation of tertiary structure

I-TASSER (Iterative Threading ASSEmbly Refinement) available online at https://zhanglab.ccmb.med.umich.edu/I-TASSER/ was used for the prediction of tertiary structure of the protein (Roy et al., 2010). In the first step, it employs LOMETS, a meta-server threading programme for template-based protein structure prediction. It contains 11 threading programs. Each threading programme generates many templates from PDB that are aligned with the query sequence. However, only one template that has shown the highest *z*-score from each of the threading programs is selected. Subsequently, 10 threading programs are ranked. Thus, the output from LOMETS contains the top 10 threading programs along with their respective high score template. Using the templates generated from LOMETS, in the next series of steps, I-TASSER generates top five final models. Each model is given a Confidence score (*C*-score). The quality of each of the predicted models is estimated by means of *C*-score. A higher *C*-score implies higher confidence in the particular model generated. Apart from *C*-score, I-TASSER computes TM-score and RMSD value for all the five models. I-TASSER chooses the top model out of the five models generated which has a strong correlation between *C*-score and TM-score.

The tertiary structure generated by I-TASSER was submitted to 'GalaxyRefine2' (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE2) available at GalaxyWEB server (http://galaxy.seoklab.org/index.html) for the refinement of protein structure obtained from the top model generated by I-TASSER. GalaxyRefine2 is an advanced version of 'GalaxyRefine', utilizes local operators and global operators as well as local error estimation and homology structure information for the improvement of the accuracy of the structure of the input protein (Lee et al., 2019). It generates 10 refined models and provides information related to RMSD, MolProbity, Clash score, Poor rotamers, Rama favored and 'GALAXY energy' for all the ten models in comparison with the user submitted model. The refined model obtained from GalaxyRefine2 was used as a final vaccine candidate for docking and *in silico* cloning.

The refined tertiary structure obtained from GalaxyRefine2 was further validated using ProSA-web, a protein structure analysis tool available at https://prosa.services.came.sbg.ac.at/prosa.php (Sippl, 1993; Wiederstein & Sippl, 2007). The overall quality of the protein is indicated by the *z*-score. The output of the programme is represented in the form of a graph consisting of *z*-scores (*y*-axis) of the structures of native proteins deduced from X-ray and NMR spectroscopy against the number of residues (*x*-axis).

The Ramachandran plot of the protein was obtained using RAMPAGE (accessed at http://mordred.bioc.cam.ac.uk/~rapper/rampage.php). The output of the programme gives information on the number of residues in favored region, allowed region and in the outlier region. Overall quality factor of the unrefined and refined protein was computed using ERRAT web server available at https://servicesn.mbi.ucla.edu/ERRAT/ (Colovos & Yeates, 1993).

2.14. Docking of the vaccine with TLR-3

TLRs (Toll Like Receptors) are immune receptors play an important role in eliciting immune responses against infection by bacteria and viruses. Adjuvants have an ability to induce innate and adaptive immune responses upon their interaction with TLRs. In the present study, β – defensin was included in the vaccine construct as an adjuvant for enhancing the efficacy of the vaccine. It acts as an agonist for TLR3. As a result, docking study was carried out between TLR3 (Receptor) and Vaccine (Ligand) to study the binding affinity of the ligand with the receptor (Bhatnager et al., 2020; Pandey et al., 2018).

Docking of the Vaccine with TLR-3 (PDB: 1ZIW) was performed online using PatchDock available at https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php (Duhovny et al., 2002; Schneidman-Duhovny et al., 2005) with the following default parameters consisted of (i) Clustering RMSD: 4.0, (ii) Complex type: default. Prior to docking, the ligands (NDG, NAG, SO_4 & Glycerol) and water molecules associated with the crystal structure of the TLR-3 was removed using PyMOL. The top-10 binding modes generated from PatchDock were submitted to FireDock (Fast Interaction Refinement in molecular Docking) available online at http://bioinfo3d.cs.tau.ac.il/FireDock/php.php (Andrusier et al., 2007; Mashiach et al., 2008). FireDock yields a refined output consisting of the ranked list of the top 10 models generated by PatchDock.

DIMPLOT under LigPlot + v2.2 was used for studying the interactions between the residues of Chain-A (TLR-3) and Chain-B (Vaccine construct) in the docked complex.

2.15. Codon adaptation and in silico cloning of vaccine candidate

JCat (JAVA Codon Adaptation Tool), an online tool available at http://www.jcat.de/Start.jsp (Grote et al., 2005) was utilized for the Codon-adaptation of the vaccine construct in *Escherichia coli* (Strain K12). Standard genetic code was used for the conversion of input amino acid sequence of the vaccine construct to the DNA sequence. Apart from this, additional options such as (i) avoid rho-independent transcription terminators, (ii) avoid prokaryotic ribosome binding sites and (iii) avoid cleavage sites of restriction enzymes, were selected for the generation of the optimized DNA sequence corresponding to the input amino acid sequence of the vaccine construct.

SnapGene® 5.1.4.1 was utilized for the *in silico* cloning of the DNA sequence (obtained from JCat) into pET-28a (+) vector. Prior to cloning, the DNA sequence of the vaccine

Table 1. Prediction of HLA-I epitopes.

Sr. No.	Position	Allele	Peptide	Antigenicity (Prediction score)	Allergenicity	Toxicity
1	**14**	**HLA-A*02:01, HLA-A*02:06**	**KLLEQWNLV**	**Antigen* (0.8252)**	**Nonallergen**	**Nontoxin**
2	28	HLA-B*57:01, HLA-B*58:01	LTWICLLQF	Nonantigen	Nonallergen	Nontoxin
3	38	HLA-A*01:01, HLA-A*30:02 HLA-B*35:01, HLA-B*53:01	YANRNRFLY	Nonantigen	Allergen	Nontoxin
4	46	HLA-A*23:01 HLA-B*57:01, HLA-B*58:01	YIIKLIFLW	Nonantigen	Nonallergen	Nontoxin
5	49	HLA-A*32:01 HLA-B*57:01, HLA-B*58:01	KLIFLWLLW	Nonantigen	Nonallergen	Nontoxin
6	53	HLA-A*23:01, HLA-A*24:02	LWLLWPVTL	Antigen* (0.7272)	Allergen	Nontoxin
7	**56**	**HLA-A*23:01, HLA-A*24:02**	**LWPVTLACF**	**Antigen* (0.5599)**	**Nonallergen**	**Nontoxin**
8	63	HLA-A*33:01	CFVLAAVYR	Nonantigen	Allergen	Nontoxin
9	64	HLA-A*02:01, HLA-A*02:06	FVLAAVYRI	Nonantigen	Nonallergen	Nontoxin
10	66	HLA-B*57:01, HLA-B*58:01	LAAVYRINW	Antigen* (0.6884)	Allergen	Nontoxin
11	71	HLA-A*32:01	RINWITGGI	Antigen* (0.5347)	Allergen	Nontoxin
12	83	HLA-B*57:01, HLA-B*58:01	MACLVGLMW	Nonantigen	Nonallergen	Nontoxin
13	86	HLA-A*01:01, HLA-A*30:02	LVGLMWLSY	Nonantigen	Nonallergen	Nontoxin
14	88	HLA-A*02:01, HLA-A*02:03, HLA-A*02:06, HLA-A*32:01	GLMWLSYFI	Nonantigen	Nonallergen	Nontoxin
15	91	HLA-A*26:01, HLA-A*32:01 HLA-B*08:01, HLA-B*15:01, HLA-B*35:01,	WLSYFIASF	Nonantigen	Nonallergen	Nontoxin
16	92	HLA-A*11:01, HLA-A*31:01, HLA-A*33:01, HLA-A*68:01	LSYFIASFR	Nonantigen	Nonallergen	Nontoxin
17	93	HLA-A*23:01, HLA-A*24:02	SYFIASFRL	Nonantigen	Nonallergen	Nontoxin
18	94	HLA-A*23:01, HLA-A*24:02	YFIASFRLF	Nonantigen	Allergen	Nontoxin
19	96	HLA-A*31:01, HLA-A*33:01, HLA-A*68:01	IASFRLFAR	Nonantigen	Allergen	Nontoxin
20	**98**	**HLA-A*31:01**	**SFRLFARTR**	**Antigen* (0.5617)**	**Nonallergen**	**Nontoxin**
21	100	HLA-A*30:01, HLA-A*32:01 HLA-B*07:02, HLA-B*08:01, HLA-B*15:01	RLFARTRSM	Antigen* (0.6876)	Allergen	Nontoxin
22	101	HLA-A*23:01, HLA-A*24:02	LFARTRSMW	Nonantigen	Allergen	Nontoxin
23	135	HLA-B*40:01, HLA-B*44:02, HLA-B*44:03	SELVIGAVI	Nonantigen	Nonallergen	Nontoxin
24	137	HLA-A*68:01	LVIGAVILR	Nonantigen	Nonallergen	Nontoxin
25	141	HLA-A*31:01	AVILRGHLR	Nonantigen	Allergen	Nontoxin
26	147	HLA-A*30:01 HLA-B*07:02, HLA-B*08:01, HLA-B*15:01	HLRIAGHHL	Antigen* (0.6265)	Allergen	Nontoxin
27	**149**	**HLA-A*31:01**	**RIAGHHLGR**	**Antigen* (0.5394)**	**Nonallergen**	**Nontoxin**
28	169	HLA-A*01:01, HLA-A*30:02 HLA-B*15:01, HLA-B*35:01, HLA-B*53:01, HLA-B*58:01	VATSRTLSY	Nonantigen	Allergen	Nontoxin
29	**170**	**HLA-A*01:01, HLA-A*03:01, HLA-A*11:01, HLA-A*26:01, HLA-A*30:02 HLA-B*15:01, HLA-B*57:01, HLA-B*58:01**	**ATSRTLSYY**	**Antigen* (0.5067)**	**Nonallergen**	**Nontoxin**
30	171	HLA-A*03:01, HLA-A*11:01, HLA-A*30:01, HLA-A*31:01, HLA-A*68:01	TSRTLSYYK	Nonantigen	Nonallergen	Nontoxin
31	**177**	**HLA-A*33:01**	**YYKLGASQR**	**Antigen* (0.9954)**	**Nonallergen**	**Nontoxin**
32	187	HLA-A*01:01, HLA-A*30:02	AGDSGFAAY	Nonantigen	Nonallergen	Nontoxin
33	190	HLA-A*30:02 HLA-B*15:01	SGFAAYSRY	Nonantigen	Allergen	Nontoxin
34	195	HLA-A*26:01, HLA-A*01:01, HLA-A*30:02, HLA-B*15:01	YSRYRIGNY	Nonantigen	Allergen	Nontoxin
35	197	HLA-A*23:01, HLA-A*24:02, HLA-A*30:01	RYRIGNYKL	Nonantigen	Nonallergen	Nontoxin
36	211	HLA-A*68:02	SSSDNIALL	Antigen* (1.4683)	Allergen	Nontoxin
37	212	HLA-A*01:01	SSDNIALLV	Antigen* (1.0318)	Allergen	Nontoxin

construct was verified for the absence of specific restriction enzyme recognition sites that were to be utilized for cloning into the vector. After confirmation, those restriction enzyme sequences were added on N- terminal and C-terminal regions of the vaccine construct for successful cloning.

2.16. Immune-simulation

In the present study, C-IMMSIM an agent-based model available at http://150.146.2.1/C-IMMSIM/index.php?page=1 (accessed on 25 July 2020) was used to simulate the immune

Table 2. Prediction of HLA-II epitopes.

Sr.No.	Pos.	MHC	Peptide	Antigenicity (Prediction score)	Allergenicity	Toxicity
1	1	HLA-DQA10301-DQB10302, HLA-DQA10401-DQB10402	MADSNGTITVEELKK	Antigen*(0.7854)	Allergen	Nontoxin
2	5	HLA-DPA10201-DPB10101, HLA-DPA10301-DPB10402, HLA-DPA10201-DPB10501	NGTITVEELKKLLEQ	Antigen*(0.6935)	Allergen	Nontoxin
3	**34**	**DRB1_1302**	**LLQFAYANRNRFLYI**	**Antigen*(0.5577)**	**Nonallergen**	**Nontoxin**
4	67	DRB3_0202	LAAVYRINWITGGIA	Nonantigen	Nonallergen	Nontoxin
5	71	HLA-DQA10501-DQB10301	YRINWITGGIAIAMA	Nonantigen	Nonallergen	Nontoxin
6	94	HLA-DPA10201-DPB10501	SYFIASFRLFARTRS	Nonantigen	Nonallergen	Nontoxin
7	97	DRB1_0802, DRB1_1101	IASFRLFARTRSMWS	Nonantigen	Allergen	Nontoxin
8	107	DRB3_0202,	RSMWSFNPETNILLN	Nonantigen	Nonallergen	Nontoxin
9	**109**	**DRB1_0101**	**MWSFNPETNILLNVP**	**Antigen* (0.5349)**	**Nonallergen**	**Nontoxin**
10	114	DRB1_1302, DRB3_0202	PETNILLNVPLHGTI	Nonantigen	Allergen	Nontoxin
11	136	HLA-DQA10102-DQB10602, HLA-DQA10501-DQB10301	SELVIGAVILRGHLR	Nonantigen	Nonallergen	Nontoxin
12	146	DRB1_0802, DRB1_1101	RGHLRIAGHHLGRCD	Antigen*(0.5183)	Allergen	Nontoxin
13	148	HLA-DPA10103-DPB10201	HLRIAGHHLGRCDIK	Antigen*(0.5755)	Allergen	Nontoxin
14	164	DRB1_0701	LPKEITVATSRTLSY	Antigen*(0.5118)	Allergen	Nontoxin
15	165	DRB1_0101, DRB1_0901, DRB1_1201, DRB1_1302, DRB1_1501, DRB5_0101,	PKEITVATSRTLSYY	Antigen*(0.5797)	Allergen	Nontoxin
16	166	DRB1_1201	KEITVATSRTLSYYK	Nonantigen	Allergen	Nontoxin
17	175	DRB5_0101	TLSYYKLGASQRVAG	Nonantigen	Allergen	Nontoxin
18	176	DRB1_0101, DRB1_0701, DRB1_0901, DRB3_0202, DRB5_0101, HLA-DQA10501-DQB10301	LSYYKLGASQRVAGD	Nonantigen	Allergen	Nontoxin
19	177	DQA10501-DQB10301	SYYKLGASQRVAGDS	Antigen*(0.6318)	Allergen	Nontoxin
20	184	DRB1_0301, DRB3_0101	SQRVAGDSGFAAYSR	Nonantigen	Allergen	Nontoxin
21	190	DRB1_1302, HLA-DPA10201-DPB10501	DSGFAAYSRYRIGNY	Nonantigen	Allergen	Nontoxin
22	196	HLA-DPA10103-DPB10201	YSRYRIGNYKLNTDH	Nonantigen	Allergen	Nontoxin
23	201	DRB1_0401, DRB1_0405, DRB1_0802, DRB1_1302, DRB3_0202	IGNYKLNTDHSSSSD	Nonantigen	Nonallergen	Nontoxin
24	202	DRB1_0401, DRB3_0101	GNYKLNTDHSSSSDN	Nonantigen	Allergen	Nontoxin

response of host for the vaccine designed in the present study. A new fast Position Specific Scoring Matrix (PSSM) was utilized in C-IMMSIMM for the prediction of epitopes that binds to HLA molecules. Further, the strength of interaction of a given peptide-HLA complex with a T-cell receptor is scored using Miyazawa–Jernigan residue–residue potential in the tool (Rapin et al., 2010). For the immune-simulation studies, host HLA selection consisted of heterozygous combination of HLA-A (HLA-A*01:01, HLA-A*02:01), HLA-B (HLA-B*15:01, HLA-B*57:01) and HLA-II (DRB1_0101, DRB1_1302). These HLA molecules were randomly chosen from among the HLA alleles, whose interacting peptides were utilized for the construction of multi-epitope vaccine. Three injections of the vaccine (without LPS) were given at four weeks interval with the following time steps (one time step corresponds to 8 h): 1, 84 and 168. (Nain et al., 2020; Shey et al., 2019). Rest of the parameters viz. Random seed, simulation volume, number of antigens to inject were set to their default values except for 'Simulation Steps' that was set to 1050.

3. Results and discussion

3.1. Selection of epitopes for vaccine design

3.1.1. Prediction of HLA-I & II epitopes

A total of 37 and 24 peptides which have binding affinity for chosen HLA-I (HLA-A & HLA-B) (Table 1) and HLA-II alleles (Table 2) were identified by NetCTLpan and NetMHCIIpan, respectively. These peptides were selected from among many peptides predicted by NetCTLpan and NetMHCIIpan on the basis of their percentage rank.

3.1.2. Prediction of linear and discontinuous B-cell epitopes

BepiPred predicted a total of five epitopes. Of these, three epitopes were more than 10 amino acids long. Among the three epitopes, two were found to be nonallergic, nontoxic and antigenic. Hence, these two epitopes, KLGASQRVAGDS and RYRIGNYKLNTDHSSSSDNIA were included in the vaccine construct (Table 3).

Unlike BepiPred, which predicts B-cell epitopes that are continuous in a linear peptide sequence, DiscoTope predicts B-cell residues present on the surface of the three-dimensional structure of a protein. This implies that those residues what are distant in a linear peptide sequence, can come close when the protein assumes three-dimensional conformation to form a B-cell epitope (Kringelum et al., 2012). DiscoTope at a threshold of −3.7 identified 25 B-cell epitope residues (residue numbers 172, 182, 186, 188, 192, 195–196, 202–203, 207–211, 213–223) out of 223 total residues (Figure 2).

The final vaccine construct consists of both linear and discontinuous epitopes as determined from the results obtained using BepiPred and DiscoTope, respectively, which can

Table 3. Linear B-cell epitopes predicted by BepiPred.

Position	Peptide	Antigenicity (Prediction score)	Allergenicity	Toxicity
5-20	NGTITVEELKKLLEQW	Antigen (0.4974)	Allergen	Nontoxin
180-191	**KLGASQRVAGDS**	**Antigen (0.6960)**	**Nonallergen**	**Nontoxin**
198-218	**RYRIGNYKLNTDHSSSSDNIA**	**Antigen (0.4339)**	**Nonallergen**	**Nontoxin**

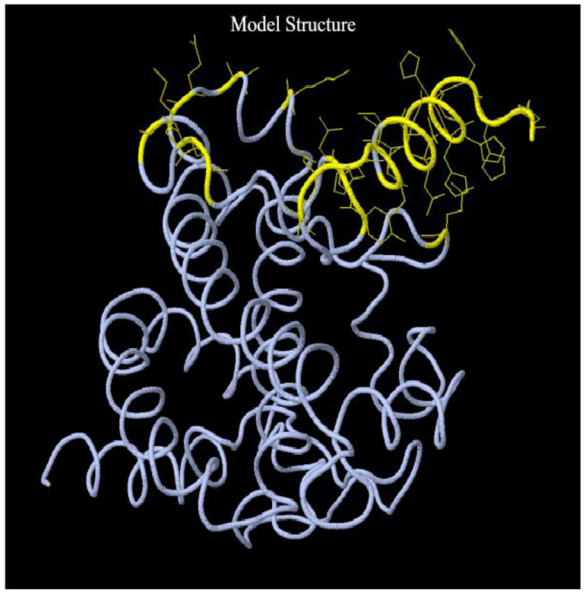

Figure 2. Tertiary structure of the vaccine with discontinuous B-cell epitopes. The side chain of each predicted B-cell epitopes by DiscoTope is highlighted in yellow.

induce humoral immune responses. This was also evident from the results obtained from Immune simulation.

3.1.3. Antigenicity and allergenicity

All the T and B- cell epitopes in the present study were screened for their antigenicity, allergenicity and toxicity. Only those peptides which were antigenic, nonallergic and non-toxic were used in the present study for the design of vaccine construct. The final vaccine construct (223 amino acids) was found to be antigenic in all the three models viz. Tumor (threshold: 0.49), Virus (threshold: 0.4) and Bacteria (threshold: 0.4). It was also found to be nonallergic.

3.2. Design of vaccine construct

The final vaccine construct is 223 amino acids long (Table 4). It consists of a total of six HLA-I, two HLA-II and two B-cell epitopes. Apart from this, at the N-terminal end, β-defensin was included in the vaccine construct for increasing the immunogenicity of the multi-epitope vaccine. β-defensin acts as an adjuvant and elicits antiviral responses (Kim et al., 2018; Lei et al., 2019). β-defensins are anti-microbial peptides involved in eliciting innate immune responses. Apart from this, β-defensin interact with the immune receptors such as TLRs or CCR6 and induces innate and adaptive immune responses (Lei et al., 2019). It is used as an adjuvant in the construction of several vaccines by Immunoinformatics approach (Ali et al., 2017; Bhatnager et al., 2020; Dong et al., 2020; Narula et al., 2018).

Apart from β-defensin, PADRE (pan DR epitope) sequence was also included in the vaccine construct to increase the immunogenicity of the vaccine construct as well as to enhance the T-cell help (Alexander et al., 2000; Sarkar et al., 2020; Wu et al., 2010). PADRE sequence acts as a TLR agonist adjuvant. It is a 13 amino acid long synthetic peptide that has an ability to effectively induce CD4+ T cell help. It has got several advantages, e.g. ability to bind to most of the HLA-DR molecules, clinically safe for humans and more potent than other universally available T_H cell epitopes (Ghaffari-Nazari et al., 2015). In the present study, it is placed at the N-terminal end between β-defensin and HLA-I epitopes and was connected to these two by EAAAK and AAY linkers, respectively.

In the present study, EAAAK linker was attached at the N-terminal end of the vaccine construct (Mittal et al., 2020; Sarkar et al., 2020) and was also used to join the β-defensin sequence with PADRE sequence (Sarkar et al., 2020). EAAAK is a rigid α-helix forming peptide linker, with intramolecular hydrogen bonding having a closed packed backbone. Rigid linkers possess several advantages when compared to the flexible linkers. EAAAK linkers offer efficient separation of the functional domains by keeping a fixed distance with minimal interference between the epitopes thereby maintaining their individual functional properties (Chen et al., 2013). This helps in the effective separation of domains in a bifunctional fusion protein (Arai et al., 2001).

GPGPG linker was used to join the HLA-II epitopes (Livingston et al., 2002; Sarkar et al., 2020) in the present study. GPGPG linker was designed by Livingston et al. (2002) as a universal spacer (Li et al., 2015). GPGPG linkers were demonstrated to be able to induce T_H lymphocyte (HTL) responses which is crucial for a multi-epitope vaccine. Further, GPGPG linker is a valuable tool in breaking the junctional immunogenicity, which leads to the restoration of immunogenicity of the individual epitopes. This was experimentally demonstrated by Livingston et al. (2002) in mice models.

HLA-I epitopes were joined by AAY linker (Bhatnager et al., 2020) in the present study. AAY (Ala-Ala-Tyr) linker is the cleavage site for the proteasomes in mammalian cells. Therefore, epitopes joined using AAY linker gets separated effectively within the cells; thereby reducing the junctional immunogenicity. AAY linker also increases the immunogenicity of the multi-epitope vaccine (Yang et al., 2015).

In the present study, KK linker was used to join the B-cell epitopes (Gu et al., 2017; Nain et al., 2020; Sarkar et al., 2020). The Lysine linker is the target for the Cathepsin B, a lysosomal protease involved in processing of the antigenic

Table 4. Vaccine construct.

EAAAKGIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKKEAAAKAKFVAAWTLKAAAAAYKLLEQWNLVAAYLWPVTLACFAAYSFRLFARTRAAYRIAGHHLGRAAY
ATSRTLSYYAAYYYKLGASQRGPGPGLLQFAYANRNRFLYIGPGPGMWSFNPETNILLNVPKKKLGASQRVAGDSKKRYRIGNYKLNTDHSSSSDNIAHHHHHH

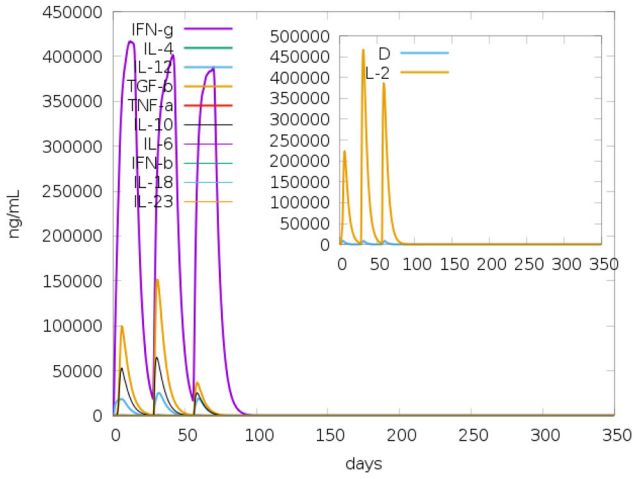

Figure 3. Production of various cytokines in response to the administration of vaccine obtained from c-ImmSim (Inset figure: Simpson Index, D).

peptides for their presentation on the cell surface in an MHC-II restricted antigen presentation. It also plays a crucial role in reducing the junctional immunogenicity by avoiding the induction of antibodies for the peptide sequence that individual epitopes can form when they are joined linearly (Yano et al., 2005). KK linkers also increase the immunogenicity (Li et al., 2015).

Linkers play an important role in minimizing the junctional immunogenicity and also in preserving the identity of each individual epitope during the processing of the vaccine within the cells thereby ensuring the immunogenicity of each epitope (Bhatnager et al., 2020; Meza et al., 2017; Shey et al., 2019). A multi-epitope vaccine without linkers may result in a new protein with unknown properties or may result in the formation of neoepitopes/junctional epitopes (Livingston et al., 2002; Meza et al., 2017).

3.3. Evaluation of the designed vaccine

3.3.1. Population coverage analysis

Population coverage analysis using the T-cell epitopes used for the construction of the vaccine covers 94.06% of world's population (HLA-I & II combined).

3.3.2. IFN-γ inducing epitope prediction

IFN- γ is a cytokine that modulates various immune responses. It is mainly secreted by activated T- cells and Natural Killer cells. It plays an important role in promoting macrophage activation, enhancing the antigen presentation, activation of the innate immune system, promoting antiviral and antibacterial immunity, etc. (Tau & Rothman, 1999). IFNepitope generated a total of 215 epitopes. Out of these, 88 epitopes had positive score and 127 had negative score. The IFN-γ inducing epitopes constitute around 41% of the total number of peptides. This is consistent with the results obtained from the Immune-simulation studies (Figure 3) which had showed an increase in the levels of IFN-γ at all the three stages of the immune response towards the vaccine.

3.3.3. Evaluation of physicochemical properties of the vaccine

Molecular weight and theoretical pI of the protein was 24.9 kDa and 10.16, respectively. Based on the theoretical pI of the vaccine, it is basic in nature. Extinction coefficient of the protein was found to be 45,840 $M^{-1} cm^{-1}$ at 280 nm when all Cys residues are reduced. Total number of negatively charged residues (Asp + Glu) is nine and the total number of positively charged residues (Arg + Lys) is 37. Estimated half-life was 1 h in mammalian reticulocytes (*in vitro*), 30 min in yeast, *in vivo* and more than 10 h in *Escherichia coli*, *in vivo*. The instability index of the protein vaccine was computed to be 36.21; this classifies the vaccine as stable. A protein whose instability index is greater than 40 is considered unstable and vice-versa (https://web.expasy.org/protparam/protparam-doc.html; accessed on 28 July 2020). Aliphatic index (defined as the relative volume occupied by aliphatic side chains) of the protein is 73.77. This indicates that the vaccine is thermo stable. The GRAVY value for the protein is −0.400. This indicates that the protein is hydrophilic

3.3.4. Prediction of the secondary structure

The secondary structure of the final vaccine construct contained 16% beta strand, 39% alpha helix and 44% coil (Figure 4).

3.3.5. Prediction, refinement and validation of tertiary structure

***I-TASSER*:** PDB ID codes of the top threading templates generated in I-TASSER by LOMETS were as follows: 1kj6, 2z36A, 3r9cA, 1gh6B, 6gk5A, 5id6A (*z* score <1). I-TASSER reported five models which correspond to five large structure clusters. Model 1 was chosen as the best template out of the five models generated by I-TASSER. It has the *C*-score of −4.67, Exp.TM-Score: 0.23 ± 0.07 & Exp. RMSA value of 17.2. This model has shown comparatively strong correlation between *C*-score and TM-score. For the remaining models, a weak correlation was observed between *C*-score and TM-score.

***GalaxyRefine2*:** GalaxyRefine2 was used for the refinement of tertiary structure predicted by I-TASSER (Figure 5). Of the 10 models generated, Model 7 was selected for further analyses and consisted of the following parameters viz. RMSD, 2.089; MolProbity, 1.903; Clash score, 5.4; Poor rotamers, 0; Rama favored, 87.3 and GALAXY energy, −4085.36.

***ProSA-web*:** The *z*-value of the protein was −5.77, which fell within the range characteristic for native proteins of similar sizes as can be observed from Figure 6. This is an indication that the predicted tertiary structure probably does not

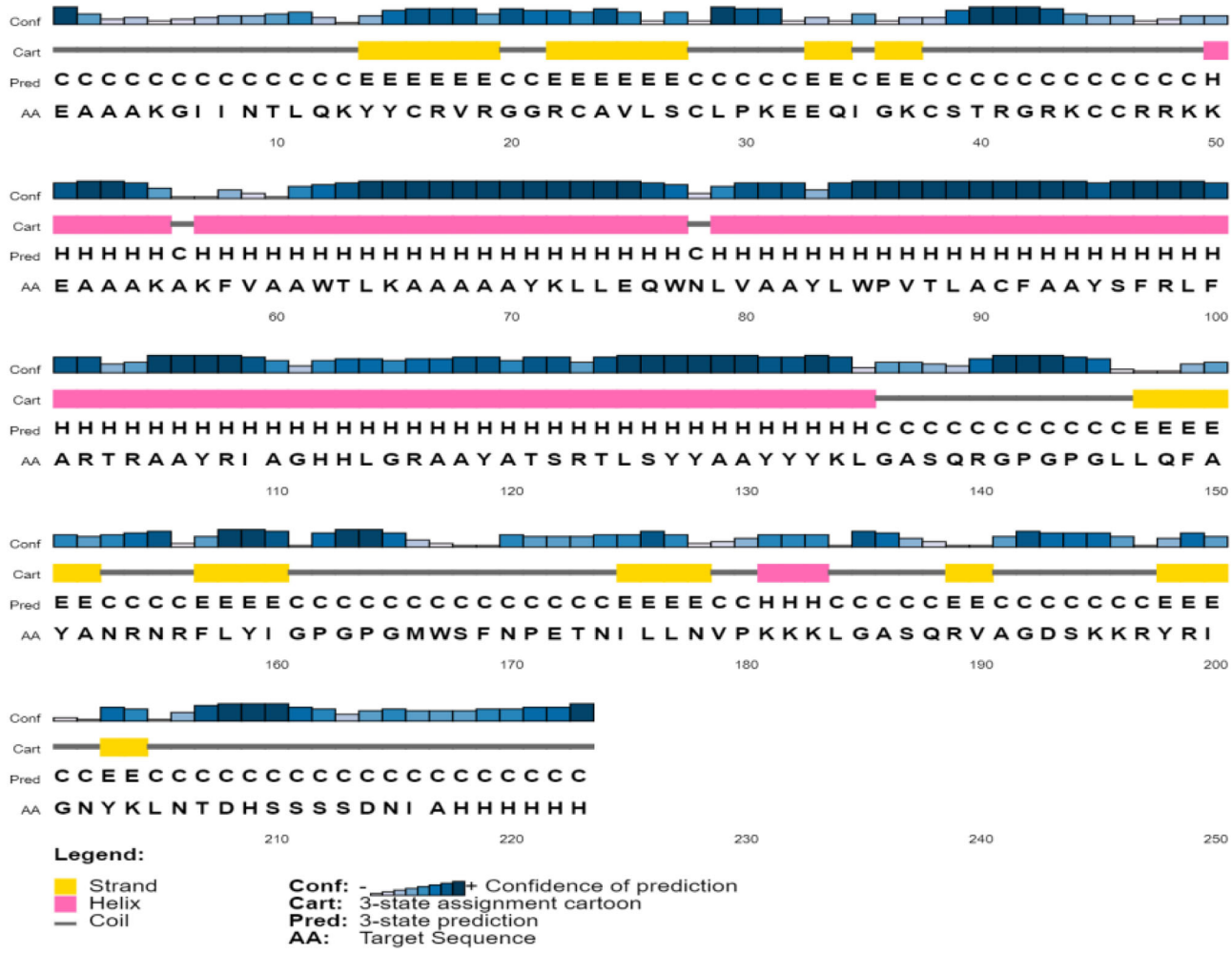

Figure 4. Prediction of secondary structure of the vaccine by PSIPRED.

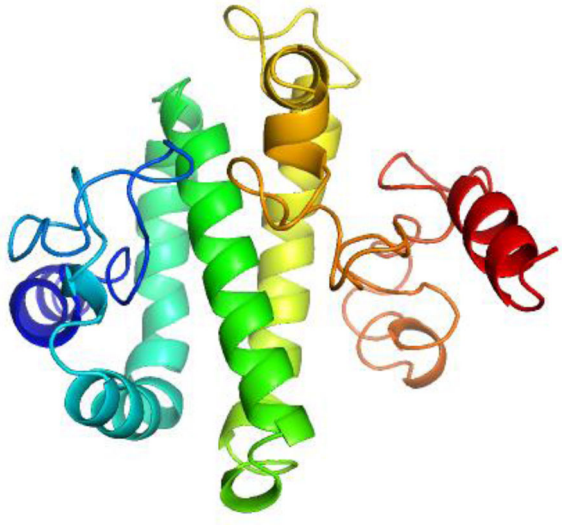

Figure 5. 3D structure of the vaccine refined by GalaxyRefine2.

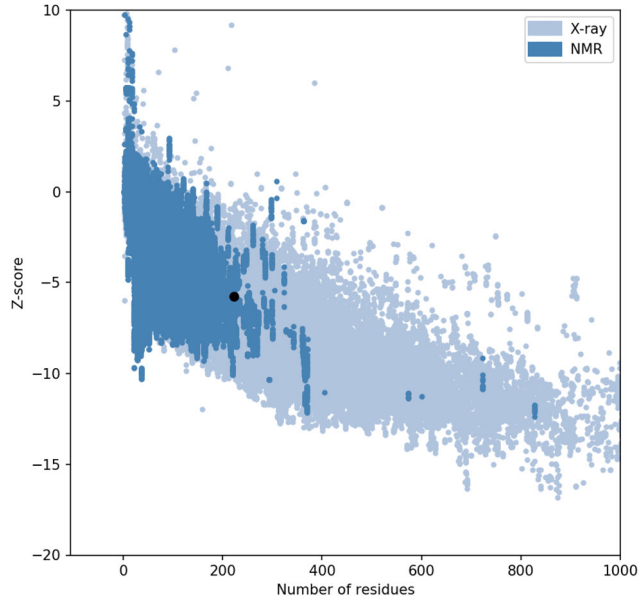

Figure 6. *z*-score of the vaccine determined using ProSA-web.

contain errors. Visual interpretation of *z*-value obtained for the tertiary structure predicted by I-TASSER (image not shown) as well as for the refined models generated using GalaxyRefine (image not shown) and GalaxyRefine2 (Figure 6) indicated that *Z*-values fell within the region normally observed for native proteins of similar size.

***RAMPAGE*:** Analysis of the Ramachandran plot of the refined protein obtained from GalaxyRefine2 showed that 88.7% of residues in favored region, 8.1% of residues in allowed region and 3.2% of residues in outlier region (Figure 7).

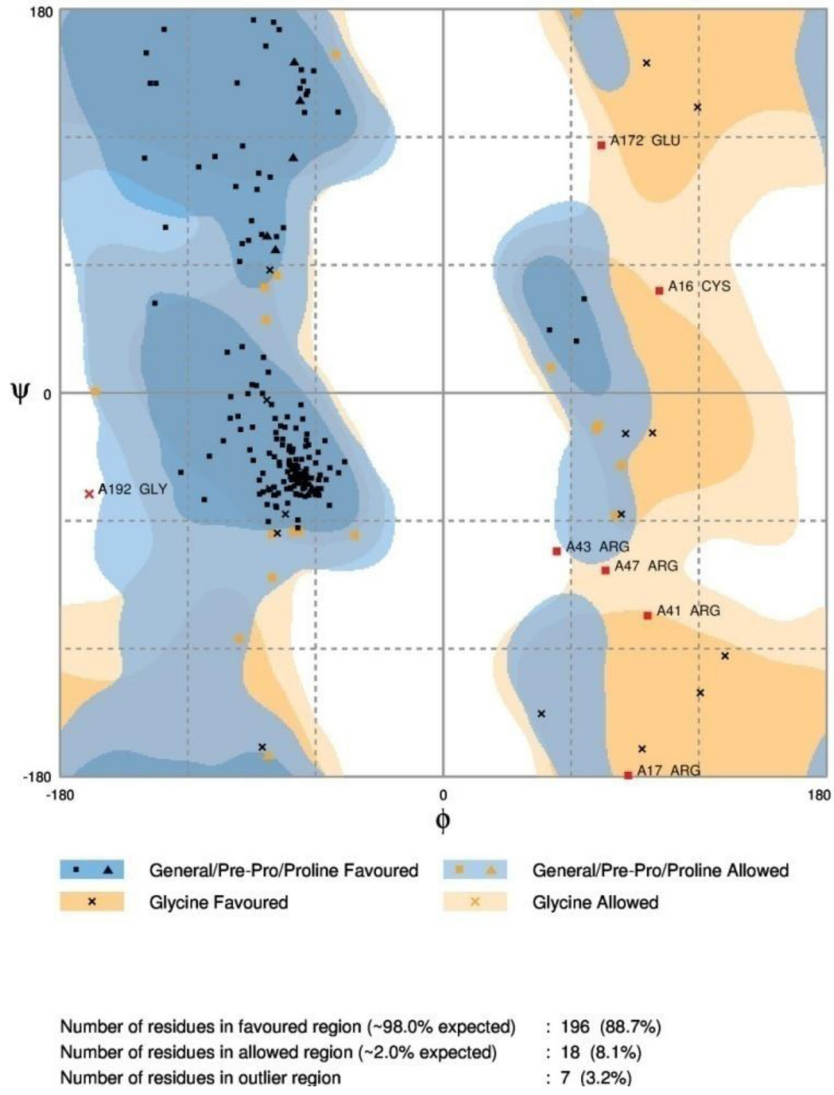

Figure 7. Ramachandran plot analysis of the vaccine.

Figure 8. Docked complex of TLR-3 and the vaccine construct. In the complex, vaccine is depicted in 'Magenta' color, whereas TLR-3 is depicted in 'Cyan' color.

When the parameters of Ramachandran plot obtained for the model generated and refined using I-TASSER (60.6% of residues in favored region, 20.4% residues in allowed region and 19% of residues in outlier region) and GalaxyRefine (81% of the residues are in favored region, 13.1% of the residues in the allowed region and 5.9% of the residues are in outlier region), respectively, were evaluated, the model obtained from GalaxyRefine2 is better when compared with the models obtained from GalaxyRefine and I-TASSER.

The ERRAT score for the refined protein obtained using GalaxyRefine2 was found to be 70% (data not shown); whereas, for the unrefined (I-TASSER) and refined (GalaxyRefine) models, the ERRAT score was found to be 54% and 46%, respectively (data not shown).

3.3.6. Docking of the vaccine with TLR-3

All the top 10 models obtained as a result of docking the vaccine with TLR-3 in PatchDock were refined further and ranked by FireDock. Solution number 6 was the refined docked complex ranked as number 1 by FireDock (Figure 8), with the following parameters: Global energy, 8.71; Attractive VdW, −47.04; Repulsive VdW, 27.04; ACE, 23.02; and HB, −6.25.

The docked complex showed interactions between the ligand and the receptor. The interacting residues between the

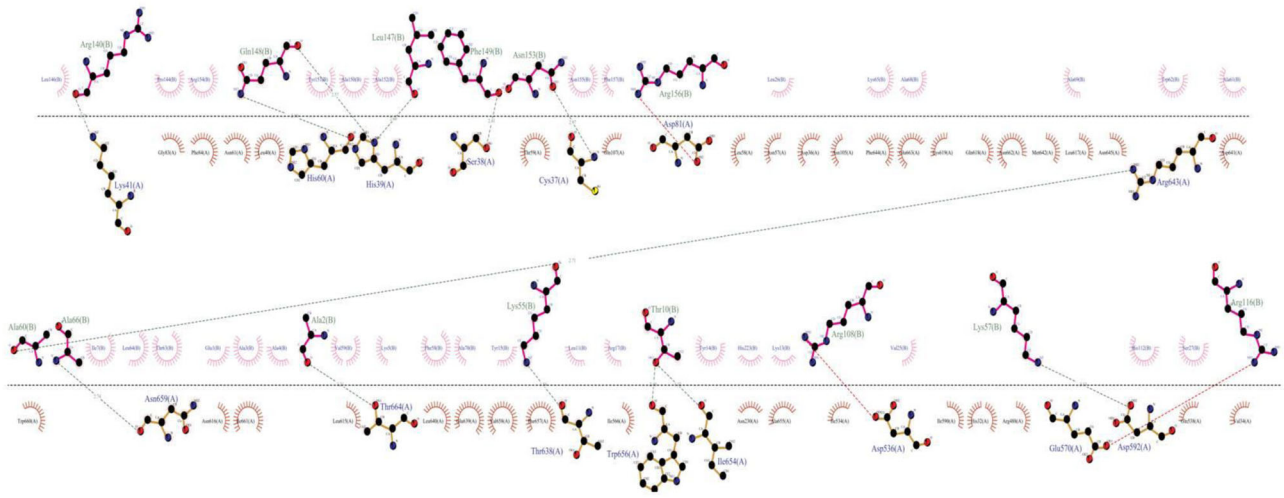

Figure 9. Representation of the molecular interactions between Chain A (TLR-3) and Chain B (vaccine) in the docked complex obtained using DIMPLOT module in LigPlot+. Dashed lines in Pink color indicate salt bridge interactions and dashed lines in blue indicate Hydrogen bonds.

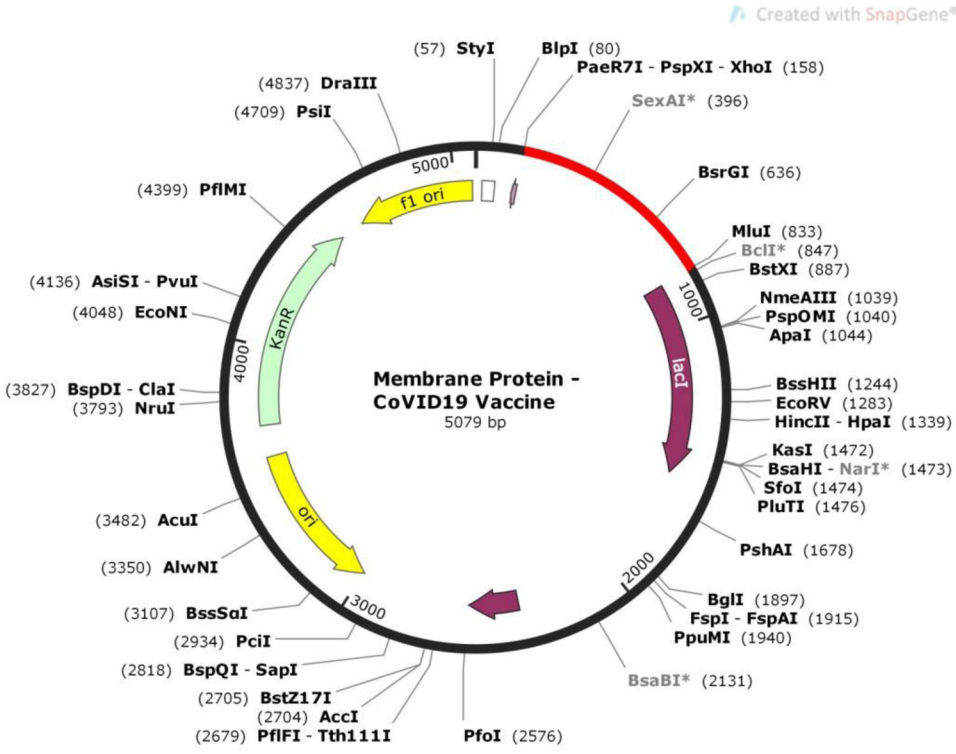

Figure 10. *In silico* restriction cloning of the vaccine into pET-28a (+) vector in between Pae71-PspXI-XhoI and MluI restriction sites. Vaccine constrict is depicted in 'Red' color.

receptor and ligand in the docked complex was visualized using DIMPLOT module in LigPlot + as shown in Figure 9.

3.3.7. Codon adaptation

The multi-epitope vaccine construct consisted of 223 amino acids and the total number of nucleotides present in the DNA sequence generated by JCat after codon-adaptation consisted of 669. CAI value and GC-content of the improved sequence were 0.96 and 52.31, respectively. The CAI value >0.8 is considered as a good value for high-level expression and the range for optimal GC content is 30–70% (Nezafat et al., 2016). Thus, the CAI value and GC-content obtained are highly satisfactory for the expression of the multi-epitope vaccine in the *E coli* (strain K12) (Nezafat et al., 2016; Pandey et al., 2018).

3.3.8. *In silico* cloning of vaccine candidate

DNA sequence was cloned into pET-28a (+) vector in between Pae71-PspXI-XhoI and MluI restriction sites. These restriction sites were not present at both the ends as well as within the DNA sequence of the vaccine construct. As a result, the XhoI and MluI restriction sites were added at the

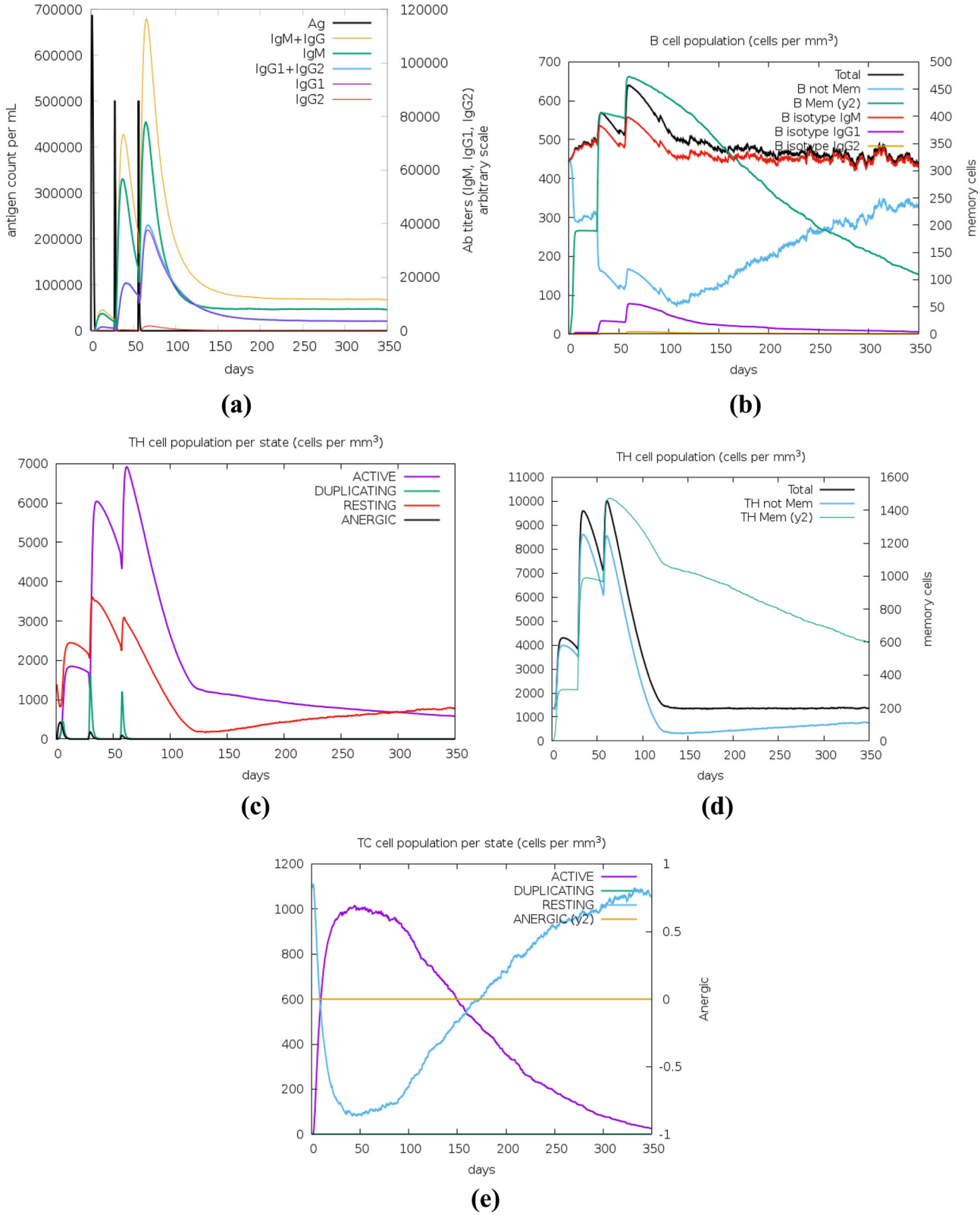

Figure 11. *In silico* immune simulation results obtained using c-ImmSim server after administration of three injections of the vaccine. (a) Production of various types of immunoglobulins, (b) B-cell population, (c) T_H cell population per state, (d) T-helper cell population, (e) T_C lymphocytes count per entity-state.

N- terminal and C- terminal ends, respectively. After this, the insert was treated with these two restriction enzymes *in silico* in SnapGene and the insert DNA was successfully cloned into the vector (Figure 10). The vector along with the insert was 5079 bp in length.

3.3.9. Immune simulation

Immune simulation results obtained from C-ImmSim showed increased levels of IgM + IgG, IgM, IgG1 and IgG1 + IgG2 antibodies in the secondary and tertiary immune responses when compared to primary immune response. IgM and

IgM + IgG antibodies were seen in the primary immune response. However, the peaks depicting the titers of IgM and IgM + IgG are distinctly separated in the secondary and tertiary responses than in primary response. There is a decrease in the levels of antigen at each stage of immune response with the increase in the levels of antibodies (Figure 11(a)). The decrease in antigen count is also attributed to the increase in the total count of B (Figure 11(b)) and T-lymphocytes. Apart from B-lymphocytes, active (Figure 11(c)) and total T_H (helper) (Figure 11(d)) cell populations along with memory T_H cell (Figure 11(d)) populations were increased during the secondary and tertiary immune responses. The T_H cell count (active and total) in the secondary and tertiary immune responses is more when compared to primary immune response. There is also an increase in the active T_C (Cytotoxic) cell population per state (Figure 11(e)).

Immune simulation studies showed an increase in the levels of antibodies, B- lymphocytes, T_H (including memory cells) and T_C cells owing to the administration of the vaccine designed in the present study. This is evident with the development of primary, secondary and tertiary immune responses' and subsequent clearance of antigen from the system.

4. Conclusion

In the present study, a vaccine was designed against membrane glycoprotein of SARS-CoV2 by different *in silico* tools. T-cell and B- cell epitopes were designed using different online servers that are efficient in predicting the epitopes. Only those epitopes which are antigenic, nonallergic and nontoxic were used for the construction of vaccine. Epitopes were joined by different linkers to reduce junctional immunogenicity and also for efficient separation and presentation to T-cell and B- cell receptors. Population coverage analysis (using the epitopes that were used for the construction of the vaccine) had shown that the vaccine can cover 94.06% of the world-wide population. Secondary and tertiary structures were predicted for the protein. Subsequently, the tertiary model generated was further refined and validated. Molecular docking was performed between the vaccine construct and TLR-3 receptor. Various molecular interactions, i.e. hydrogen bonding, hydrophobic interactions, etc. were visualized between the protein and TLR-3 receptors. Condon optimization was carried for cloning and expression of the vaccine in *E. coli*. CAI value of 0.9 was obtained for codon-optimization; indicating the high-level of protein expression in the *E coli* host. The vaccine was successfully cloned into pET-28a (+) vector. The multi-epitope vaccine was efficient in eliciting primary, secondary and tertiary immune responses as evident from Immune simulation studies. The proposed vaccine needs to be further validated for its efficacy by *in vivo* studies.

Disclosure statement

No potential conflict of interest is reported by the authors.

Funding

The authors acknowledge to VFSTR (Deemed to be university) and DST-FIST (LSI- 576/2013) networking facility to carry out this work.

ORCID

Vijaya Sai Ayyagari http://orcid.org/0000-0001-5438-2268
Venkateswarulu T. C. http://orcid.org/0000-0002-7822-8860
Abraham Peele K. http://orcid.org/0000-0001-8173-9672
Krupanidhi Srirama http://orcid.org/0000-0001-7702-8974

References

Ada, K., Chuah, C., Majeed, A. B. A., Leow, C. H., Lim, B. H., & Leow, C. Y. (2018). Current progress of immunoinformatics approach harnessed for cellular- and antibody-dependent vaccine design. *Pathogens and Global Health*, *112*(3), 123–131. https://doi.org/10.1080/20477724.2018.1446773

Alexander, J., del Guercio, M. F., Maewal, A., Qiao, l., Fikes, J., Chesnut, R. W., Paulson, J., Bundle, D. R., DeFrees, S., & Sette, A. (2000). Linear PADRE T helper epitope and carbohydrate B cell epitope conjugates induce specific high titer IgG antibody responses. *Journal of Immunology*, *164*(3), 1625–1633. https://doi.org/10.4049/jimmunol.164.3.1625

Ali, M., Pandey, R. K., Khatoon, N., Narula, A., Mishra, A., & Prajapati, V. K. (2017). Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Scientific Reports*, *7*(1), 9232. https://doi.org/10.1038/s41598-017-09199-w

Andrusier, N., Nussinov, R., & Wolfson, H. J. (2007). FireDock: Fast interaction refinement in molecular docking. Proteins, *69*(1), 139–159. https://doi.org/10.1002/prot.21495

Arai, R., Ueda, H., Kitayama, A., Kamiya, N., & Nagamune, T. (2001). Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Engineering*, *14*(8), 529–532. https://doi.org/10.1093/protein/14.8.529

Bhatnager, R., Bhasin, M., Arora, J., & Dang, A. S. (2020). Epitope based peptide vaccine against SARS-COV2: An immune-informatics approach. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2020.1787227

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

Bianchi, M., Benvenuto, D., Giovanetti, M., Angeletti, S., Ciccozzi, M., & Pascarella, S. (2020). Sars-CoV-2 envelope and membrane proteins: Structural differences linked to virus characteristics? *BioMed Research International*, *2020*, 4389089. https://doi.org/10.1155/2020/4389089

Buchan, D. W. A., & 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*, *17*, 153. https://doi.org/10.1186/1471-2105-7-153

Chen, X., Zaro, J. L., & Shen, W. C. (2013). Fusion protein linkers: Property, design and functionality. *Advanced Drug Delivery Reviews*, *65*(10), 1357–1369. https://doi.org/10.1016/j.addr.2012.09.039

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

Dagur, H. S., & Dhakar, S. S. (2020). Genome organization of Covid-19 and emerging severe acute respiratory syndrome Covid-19 outbreak: A pandemic. *Eurasian Journal of Medicine and Oncology*, *4*(2), 107–115.

Devi, A., & Chaitanya, N. S. N. (2020). In silico designing of multi-epitope vaccine construct against human coronavirus infections. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2020.1804460

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

Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2014). AllerTOP v.2–a server for in silico prediction of allergens. *Journal of Molecular Modeling*, *20*(6), 2278. https://doi.org/10.1007/s00894-014-2278-5

Dong, R., Chu, Z., Yu, F., & Zha, Y. (2020). Contriving multi-epitope subunit of vaccine for COVID-19: Immunoinformatics approaches. *Frontiers in Immunology*, 11, *1784* . https://doi.org/10.3389/fimmu.2020.01784

Duhovny, D., Nussinov, R., & Wolfson, H. J. (2002). Efficient unbound docking of rigid molecules. In Gusfield, D. (Ed.), *Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy, Lecture Notes in Computer Science 2452* (pp. 185–200). Springer Verlag.

Enayatkhani, M., Hasaniazad, M., Faezi, S., Gouklani, H., Davoodian, P., Ahmadi, N., & Ahmad, 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*, 1–19. https://doi.org/10.1080/07391102.2020.1756411

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch A. (2005). Protein identification and analysis tools on the ExPASy Server. In John M. Walker (Ed), *The proteomics protocols handbook* (pp. 571–607). Humana Press.

Ghaffari-Nazari, H., Tavakkol-Afshari, J., Jaafari, M. R., Tahaghoghi-Hajghorbani, S., Masoumi, E., & Jalali, S. A. (2015). Improving multi-epitope long peptide vaccine potency by using a strategy that enhances CD4+ T help in BALB/c mice. *PLoS One*, *10*(11), e0142563. https://doi.org/10.1371/journal.pone.0142563

Gorbalenya, A. E., Baker, S. C., & Baric, R. S. (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

Graham, R. L., Sparks, J. S., Eckerle, L. D., Sims, A. C., & Denison, M. R. (2008). SARS coronavirus replicase proteins in pathogenesis. *Virus Research*, *133*(1), 88–100. https://doi.org/10.1016/j.virusres.2007.02.017

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 Res*, *33*, W526–W531. https://doi.org/10.1093/nar/gki376

Gu, Y., Sun, X., Li, B., Huang, J., Zhan, B., & Zhu, X. (2017). Vaccination with a paramyosin-based multi-epitope vaccine elicits significant protective immunity against *Trichinella spiralis* infection in mice. *Frontiers in Microbiology*, *8*, 1475. https://doi.org/10.3389/fmicb.2017.01475

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

He, Y., Rappuoli, R., De Groot, A. S., & Chen, R. T. (2010). Emerging Vaccine Informatics. *Journal of Biomedicine and Biotechnology*, *2010*, 218590. https://doi.org/10.1155/2010/218590

Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research*, *45*(W1), W24–W29. https://doi.org/10.1093/nar/gkx346

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, *292*(2), 195–202. https://doi.org/10.1006/jmbi.1999.3091

Joshi, A., Joshi, B. C., Mannan, M. A., & Kaushik, V. (2020). Epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach. *Informatics in Medicine Unlocked*, *19*, 100338. https://doi.org/10.1016/j.imu.2020.100338

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, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., & Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. Cell, *181*(4), 914–921. https://doi.org/10.1016/j.cell.2020.04.011

Kim, J., Yang, Y. L., Jang, S. H., & Jang, Y. S. (2018). Human beta-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, Y. S., Yoon, N., Karisa, N., Seo, S., Lee, J., Yoo, S., Yoon, I., Kim, Y., Lee, H., & Ahn, J. (2019). Identification of novel immunogenic proteins against *Streptococcus parauberis* in a zebrafish model by reverse vaccinology. *Microbial Pathogenesis*, *127*, 56–59. https://doi.org/10.1016/j.micpath.2018.11.053

Kringelum, J. V., Lundegaard, C., Lund, O., & Nielsen, M. (2012). Reliable B cell epitope predictions: Impacts of method development and improved benchmarking. *PLOS Computational Biology*, *8*(12), e1002829. https://doi.org/10.1371/journal.pcbi.1002829

Kumar, S., Nyodu, R., Maurya, V. K., & Saxena, S. K. (2020). Morphology, genome organization, replication, and pathogenesis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Coronavirus Disease 2019 (COVID-19)*, *2020*, 23–31. https://doi.org/10.1007/978-981-15-4814-7_3

Lee, G. R., Won, J., Heo, L., & Seok, C. (2019). GalaxyRefine2: Simultaneous refinement of inaccurate local regions and overall protein structure. *Nucleic Acids Research*, *47*(W1), W451–W455. https://doi.org/10.1093/nar/gkz288

Lei, Y., Zhao, F., Shao, J., Li, Y., Li, S., Chang, H., & Zhang, Y. (2019). Application of built-in adjuvants for epitope-based vaccines. PeerJ, *6*, e6185. https://doi.org/10.7717/peerj.6185

Li, X., Guo, L., Kong, M., Su, X., Yang, D., Zou, M., Liu, Y., & Lu, L. (2015). Design and evaluation of a multi-epitope peptide of human metapneumovirus. Intervirology, *58*(6), 403–412. https://doi.org/10.1159/000445059

Livingston, B., Crimi, C., Newman, M., Higashimoto, Y., Appella, E., Sidney, J., & Sette, A. (2002). A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. *The Journal of Immunology*, *168*(11), 5499–5506. https://doi.org/10.4049/jimmunol.168.11.5499

Lizbeth, R. G., Jazmín, G. M., José, C. B., & Marlet, M. A. (2020). Immunoinformatics study to search epitopes of spike glycoprotein from SARS-CoV-2 as potential vaccine. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2020.1780944

Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R., & Wolfson, H. J. (2008). FireDock: A web server for fast interaction refinement in molecular docking. *Nucleic Acids Research*, *36*, W229–232. https://doi.org/10.1093/nar/gkn186

Meza, B., Ascencio, F., Sierra-Beltrán, A. P., Torres, J., & Angulo, C. (2017). A novel design of a multi-antigenic, multistage and multi-epitope vaccine against Helicobacter pylori: An *in silico* approach. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *49*, 309–317. https://doi.org/10.1016/j.meegid.2017.02.007

Mittal, A., Sasidharan, S., Raj, S., Balaji, S. N., & Saudagar, P. (2020). Exploring the Zika genome to design a potential Multiepitope vaccine using an immunoinformatics approach. *International Journal of Peptide Research and Therapeutics*, *26*(4), 2231–2240. https://doi.org/10.1007/s10989-020-10020-y

Nain, Z., Abdulla, F., Rahman, M. M., Karim, M. M., Khan, M. S. A., Sayed, S. B., Mahmud, S., Rahman, S. M. R., Sheam, M. M., Haque, Z., & Adhikari, U. K. (2020). Proteome-wide screening for designing a multi-epitope vaccine against emerging pathogen elizabethkingia anophelis using immunoinformatic approaches. *Journal of Biomolecular Structure and Dynamics*, *38*(16), 4850–4867. https://doi.org/10.1080/07391102.2019.1692072

Narula, A., Pandey, R. K., Khatoon, N., Mishra, A., & Prajapati, V. K. (2018). Excavating chikungunya genome to design B and T cell multi-epitope subunit vaccine using comprehensive immunoinformatics approach to control chikungunya infection. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *61*, 4–15. https://doi.org/10.1016/j.meegid.2018.03.007

Nezafat, N., Karimi, Z., Eslami, M., Mohkam, M., Zandian, S., & Ghasemi, Y. (2016). Designing an efficient multi-epitope peptide vaccine against

Vibrio cholerae via combined immunoinformatics and protein interaction based approaches. *Computational Biology and Chemistry*, *62*, 82–95. https://doi.org/10.1016/j.compbiolchem.2016.04.006

Oli, A. N., Obialor, W. O., Ifeanyichukwu, M. O., Odimegwu, D. C., Okoyeh, J. N., Emechebe, G. O., Adejumo, S. A., & Ibeanu, G. C. (2020). Immunoinformatics and vaccine development: An overview. *ImmunoTargets and Therapy*, *9*, 13–30. https://doi.org/10.2147/ITT.S241064

Pandey, R. K., Ojha, R., Aathmanathan, V. S., Krishnan, M., & Prajapati, V. K. (2018). Immunoinformatics approaches to design a novel multi-epitope subunit vaccine against HIV infection. Vaccine, *36*(17), 2262–2227. https://doi.org/10.1016/j.vaccine.2018.03.042

Patronov, A., & Doytchinova, I. (2013). T-cell epitope vaccine design by immunoinformatics. *Open Biology*, *3*(1), 120139. https://doi.org/10.1098/rsob.120139

Peele, K., A., Srihansa, T., Krupanidhi, S., Ayyagari, V. S., & Venkateswarulu, T. C. (2020). Design of multi-epitope vaccine candidate against SARS-CoV-2: An *in-silico* study. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2020.1770127

Prasad, A., & Prasad, M. (2020). SARS-CoV-2: The emergence of a viral pathogen causing havoc on human existence. *Journal of Genetics*, *99*(1), 37. https://doi.org/10.1007/s12041-020-01205-x

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

Reynisson, B., Barra, C., Kaabinejadian, S., Hildebrand, W. H., Peters, B., & Nielsen, M. (2020). Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data. *Journal of Proteome Research*, *19*(6), 2304–2315. https://doi.org/10.1021/acs.jproteome.9b00874

Rottier, P. J. M. (1995). The coronavirus membrane glycoprotein. In S. G. Siddell (Ed.), *The Coronaviridae. The viruses*. Springer. https://doi.org/10.1007/978-1-4899-1531-3_6

Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, *5*(4), 725–738. https://doi.org/10.1038/nprot.2010.5

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 and Dynamics*. https://doi.org/10.1080/07391102.2020.1792347

Sarkar, B., Ullah, M. A., Johora, F. T., Taniya, M. A., & Araf, Y. (2020). Immunoinformatics-guided designing of epitope-based subunit vaccines against the SARS Coronavirus-2 (SARS-CoV-2). Immunobiology, *225*(3), 151955. https://doi.org/10.1016/j.imbio.2020.151955

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005). PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*, *33*, W363–W367. https://doi.org/10.1093/nar/gki481

Shey, R. A., Ghogomu, S. M., Esoh, K. K., Nebangwa, N. D., Shintouo, C. M., Nongley, N. F., Asa, B. F., Ngale, F. N., Vanhamme, L., & Souopgui, J. (2019). In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Scientific Reports*, *9*(1), 4409. https://doi.org/10.1038/s41598-019-40833-x

Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. Proteins, *17*(4), 355–362. https://doi.org/10.1002/prot.340170404

Stranzl, T., Larsen, M. V., Lundegaard, C., & Nielsen, M. (2010). NetCTLpan: Pan-specific MHC class I pathway epitope predictions. Immunogenetics, *62*(6), 357–368. https://doi.org/10.1007/s00251-010-0441-4

Tang, X., Li, G., Vasilakis, N., Zhang, Y., Shi, Z., Zhong, Y., Wang, L. F., & Zhang, S. (2009). Differential stepwise evolution of SARS coronavirus functional proteins in different host species. *BMC Evolutionary Biology*, *9*(1), 52. https://doi.org/10.1186/1471-2148-9-52

Tau, G., & Rothman, P. (1999). Biologic functions of the IFN-gamma receptors. Allergy, *54*(12), 1233–1251. https://doi.org/10.1034/j.1398-9995.1999.00099.x

Thomas, S., & Luxon, B. A. (2013). Vaccines based on structure-based design provide protection against infectious diseases. *Expert Review of Vaccines*, *12*(11), 1301–1311. https://doi.org/10.1586/14760584.2013.840092

Tordello, E. D., Rappuoli, R., & Delany, I. (2017). Reverse vaccinology: Exploiting genomes for vaccine design. In K. Modjarrad & W.C. Koff (Eds.), *Human vaccines: Emerging technologies in design and development* (pp. 65–86). Academic Press.

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*, W407–W410. https://doi.org/10.1093/nar/gkm290

Wu, C. Y., Monie, A., Pang, X., Hung, C. F., & Wu, T. C. (2010). Improving therapeutic HPV peptide-based vaccine potency by enhancing CD4+ T help and dendritic cell activation. *Journal of Biomedical Science*, *17*(1), 88. https://doi.org/10.1186/1423-0127-17-88

Yang, Y., Sun, W., Guo, J., Zhao, G., Sun, S., Yu, H., Guo, Y., Li, J., Jin, X., Du, L., Jiang, S., Kou, Z., & Zhou, Y. (2015). In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations. *Human Vaccines & Immunotherapeutics*, *11*(3), 795–805. https://doi.org/10.1080/21645515.2015.1012017

Yano, A., Onozuka, A., Asahi-Ozaki, Y., Imai, S., Hanada, N., Miwa, Y., & Nisizawa, T. (2005). An ingenious design for peptide vaccines. Vaccine, *23*(17–18), 2322–2326. https://doi.org/10.1016/j.vaccine.2005.01.031