

In-silico vaccine design against novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS- COV2) using immunoinformatics approach.

Joy Dip Barua¹, Niloy Das¹, Najia Absar Munia¹ and Suraj Raju².

¹Department of Pharmacy, BGC Trust University Bangladesh, Chandanish, Chattogram-4381, Bangladesh.

²Nitza Bioventure, Hyderabad, India.

suraj.balraj@uqconnect.edu.au

Abstract

The novel Coronavirus, also known as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is a contagion for the coronavirus disease commonly known as COVID-19. Globally, the COVID-19 pandemic is a public health crisis responsible for 18.2 million cases with 692,000 fatalities worldwide as of 04th August 2020. The aim of this study was to computationally design a multi-epitope vaccine (MEV) using emerging immunoinformatics approach to effectively prevent SARS-CoV-2 infection. Promising Cytotoxic T Lymphocyte (CTL), Helper T Lymphocyte (HTL) and Linear B-Cell specific epitopes were screened from highly antigenic viral proteins to generate a chimeric vaccine construct. The CTL epitopes were found to form a stable complex with HLA molecules. The multi-epitope vaccine candidate was designed by combining identified epitopes of diverse specificity using linker peptides. A model of the vaccine also included Toll like receptor (TLR-2) agonist as an adjuvant. The vaccine structure was refined and analyzed for its suitability of use. The utility of the vaccine was assessed by conducting molecular docking with TLR-2 followed by Molecular Dynamics simulation for determining the complex stability. To measure the potency, a simulation of immune system upon vaccine candidate administration was executed. The multi-epitope vaccine was predicted to stimulate both humoral and cell-mediated immune response after multi-stage *in-silico* study. The viral protein based vaccine sequence optimized and cloned in a bacterial expression system using pET-28a(+) vector. A Poly-His (6xHis) tag was added to facilitate protein purification.

Key words: SARS-CoV-2, Multi-epitope vaccine, COVID-19, Adjuvant, TLR-2

1. INTRODUCTION

Coronaviruses (CoV) are a group of zoonotic viruses responsible for symptoms ranging from a mild cold to severe respiratory diseases such as Middle East Respiratory Syndrome (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS-CoV)^[1]. Six virulent strains of Coronavirus (CoV) have been identified in the last decade^[1]. However, a novel strain namely Severe Acute Respiratory Syndrome Coronavirus 2(SARS-CoV-2) was found to be associated with the recent global pandemic^[2, 3]. The severe respiratory illness caused by the SARS-CoV-2 strain is commonly described as Coronavirus Disease 2019 (COVID-19). COVID-19 outbreak was confirmed by the World Health Organization's (WHO) committee as a global health emergency on 30 January 2020^[4, 5] and on 11 March 2020, it was declared a global pandemic^[6].

SARS-CoV-2 is an un-segmented, enveloped, and positive-sense RNA virus that is widely distributed in mammalian species and belongs to the *Coronaviridae* family^[7, 8]. SARS-CoV-2 is a novel β -coronavirus which is approximately 65-125nm in diameter, containing single RNA strands, nucleocapsids and crown-like spikes on the surface. The spike protein facilitates infection of the host epithelial cell leading to pulmonary failure and potentially fatal inflammation of the lower respiratory tract^[9]. It bears a genome size ranging from 29.8kb to 29.9kb^[10] that comprises of four main structural proteins: spike (S) glycoprotein, small envelope (E) glycoprotein, membrane (M) glycoprotein, nucleocapsid (N) and associated accessory proteins^[11]. The spike protein is a transmembrane glycoprotein that forms homotrimers on the viral surface and enables host-viral interaction by attaching to

angiotensin-converting enzyme 2 (ACE2) receptor expressed in lower respiratory tract cells^[12, 13]. The Nucleocapsid called N protein is the structural portion of CoV and assists in viral budding at the Golgi-Endoplasmic reticulum region^[9]. It is highly phosphorylated and plays an important role in packaging the viral RNA^[12]. The Membrane or M protein is another structural component that forms envelope shape. Envelope (E) protein is involved in the development and maturation of the virus^[14].

As of 4th August 2020, a total of 18.2 million cases were confirmed with 692,000 fatalities. Currently, there are no commercially available vaccines approved while licensed therapies such as remdesivir and favipiravir do not guarantee complete recovery in severe cases caused by an increased viral-load. Some traditional Chinese medicines such as Lianhuaqingwen and ShuFengJieDu based therapy may serve as potential alternative COVID-19 treatments, but no clinical reports confirm the effectiveness and safety of these medicines for Covid-19^[15].

Vaccination has proven to induce quick immune response for the prevention of numerous deadly infectious diseases^[16-18]. In traditional vaccine development approach, a whole pathogen is attenuated to serve as a vaccine candidate. The key limitation is the risk associated with toxicity and allergenicity of the whole pathogenic accessory proteins used. In-complete attenuation can lead to additional safety concerns. Moreover, *in-vitro* and *in-vivo* cell-based toxicity assays are required to meet additional regulatory requirements which can be time-consuming and expensive^[19, 20]. The *In-silico* peptide-based vaccine design offers a promising solution to mitigate risk and develop a potent vaccine capable of stimulating the immune system.

The selectivity of peptide-based vaccines allows precise activation of the immune response against specific antigens that do not require whole agent thereby rendering them safe^[21, 22]. The peptide-based vaccine is constructed by incorporating T-cell and B-cell epitopes which an Antigen Presenting Cell (APC) recognizes through MHC-I and MHC-II molecules and then elicits cytokine and Interferon-gamma (IFN- γ) production for inducing specific cellular and humoral immune response. Antigenic T-cell epitopes can be linked to antigenic epitopes of diverse specificities to create an immunogenic vaccine protein^[23, 24]. The aim of this research is to design a multi-epitope chimeric vaccine candidate (MEV) by applying various computational tools of the emerging immunoinformatics approach. It is achieved by predicting T-cell, B-cell and cytokine targeted epitopes from antigenic SARS-CoV-2 proteins then validating the effectiveness of epitopes to be used as a potential constituent of the chimeric vaccine candidate. The potency of the vaccine is confirmed by a series of simulations including docking, molecular dynamics and immune-simulation.

2. METHODS AND MATERIAL:

2.1 Retrieval of protein sequence and antigenicity prediction :

The protein sequences of SARS-CoV-2 (9 sequences) were retrieved from National Center for Biotechnology Information (NCBI) in FASTA format and the antigenicity of all protein sequences was assessed using the VaxiJen 2.0 server where the threshold value was set to 0.4^[25]. It predicts antigenicity based on the transformation method of auto cross-covariance (ACC) to maintain predictive accuracy of 70-89%^[25]. For further analysis, the proteins with the highest antigenic scores were chosen.

2.2 Prediction of Epitopes of Cytotoxic T-lymphocyte (CTL) and Binding Alleles of MHC-I:

The CTL epitopes of nine amino acid length for 12 supertypes (A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, and B62) were identified from the selected protein sequences through the NetCTL 1.2 server. It uses three criteria, namely C-terminal cleavage, MHC-I binding peptides and Tap transport efficiency where MHC-I binding and C-terminal cleavage are obtained by using Artificial Neural Network (ANN)^[26]. The threshold was set in this analysis at 0.5. Similarly, the consensus method of the IEDB analysis tool was used to predict the MHC-I binding alleles of the CTL epitope. Percentile rank was adjusted to ≤ 5 for counting binding alleles because lower score indicates higher affinity and source species was selected as Human^[27].

2.3 Prediction of Antigenicity, Allergenicity and Toxicity of CTL Epitopes:

The antigenicity of selected CTL epitopes was analysed to ensure their capacity to induce immune response with the help of VaxiJen 2.0 server where threshold value was calibrated to 0.4^[25]. Immunogenicity prediction is important for the vaccine, thus it was tested with the MHC-I immunogenicity tool of the IEDB server^[28]. Both Allpred (<https://www.bionet.sccc.ru/psd/cgi-bin/programs/Allpred/allpred.cgi>) and AllerTOP 2.0 server were used to predict allergenicity^[29]. Toxicity of the CTL

epitopes were predicted with the ToxinPred server to isolate non-toxic CTL epitopes^[30].

2.4 Prediction of Helper T-lymphocyte (HTL) Epitopes and MHC-II Alleles:

Helper T-lymphocyte (HTL) responses play multiple functions that is essential to developing the immune capacity to fight pathogens. Therefore, the HTL epitopes (15-mer) were identified using the IEDB MHC-II binding tool by applying the NN-align method^[31]. A different percentile rank was generated for each epitope by comparing randomly selected 15-mer peptide's scores from the SWISSPROT database. The percentile rank indicates the binding affinity, but the smaller the percentile rank, the better the binding affinity. So in this analysis, the percentile rank ≤ 5 was considered^[32].

2.5 Identification of Cytokine-Inducing HTL Epitopes:

The helper T-cell secretes a variety of cytokines such as Interferon-gamma (IFN- γ), Interleukin-4 (IL-4), interleukin-10 (IL-10) which activate the B-lymphocytes, cytotoxic T-cells and other immune cells that are responsible for the immune response^[33]. HTL targeting epitopes were identified by using INFepitope server^[34]. Furthermore, IL4pred and IL10pred server was used to predict IL-4 and IL-10 stimulating properties of HTL epitopes^[34, 35].

2.6 Identification of Linear B-lymphocyte (LBL) Epitopes:

B-cell epitopes are antigenic which recognize and adhere to the B-lymphocytes leading to the secretion of antibodies^[36]. ABCpred server was used to identify the B-cell targeting epitopes which were developed based on the recurrent neural networks^[37]. The isolated LBL epitopes were then analyzed with VaxiJen 2.0, AllerTOP 2.0 and ToxinPred servers to predict their antigenicity, allergenicity, and toxicity.

2.7 Discontinuous B-cell Epitope Prediction:

Most B-cell epitopes are sequentially discontinuous, indicating that they are composed of amino acid residues that may be far apart in sequence and interact directly with the receptor of the immune system^[38]. We used the ElliPro tool of the IEDB server for discontinuous B-cell epitope prediction.

2.8 Population Coverage Analysis:

In diverse ethnicities different Human Leukocyte Antigen (HLA) alleles, as well as their expression, are sensationally distributed at different frequencies^[39]. Hence, HLA-alleles distribution pattern varies among various ethnic groups and geographic areas around the world. IEDB population coverage tool was used to calculate population coverage.

2.9 Conservancy Analysis:

Conservancy analysis was analyzed using the IEDB's epitope conservancy analysis tool. This tool calculates the degree of an epitope's conservation within a given protein sequence set at a given identity level^[40].

2.10 Molecular Docking Analysis between CTL Epitopes and HLA-Alleles:

Molecular docking is an important step to predict the interaction between two molecules. The MDockPeP server^[41] was used for predicting interactions between selected CTL epitopes and human HLA-A*11:01 (PDB ID:

4UQ2). The binding affinity (kcal/mol) and Dissociation constant (K_D) for the epitope-HLA-A*11:01 complex was predicted employing the PRODIGY server^[80].

2.11 Construction of Multi-epitope vaccine construct:

Multi-epitope vaccine was constructed by connecting all isolated epitopes with appropriate linkers. For this purpose, chosen CTL, HTL, and LBL epitopes were combined together with AAY, GPGPG and KK linkers respectively^[42]. Linkers (AAY, GPGPG, KK) play a very important role for protein folding, flexibility and protein domain separation which makes the protein more stable^[43]. Using adjuvants potentially increases immunogenicity^[19] hence five different vaccine candidates were designed, using five different adjuvants including human beta-defensin, human beta-defensin 2, L7/L12 ribosomal protein, HABA protein (M. tuberculosis, accession number: AGV15514.1), OmpA protein (GenBank: AFS89615.1). These 5 adjuvants have been used to construct the vaccine in different previous articles^[44, 45]. Therefore, different

vaccine sequences were constructed using these 5 adjuvants to select the highest antigenic vaccine sequence. A Linker peptide is necessary to join two epitopes for the proper functioning of each epitope^[43]. Adjuvants were linked with the first CTL epitopes with the help of EAAAK linker^[46].

2.12 Antigenic Properties prediction of Constructed Vaccine Sequence:

The five constructed vaccine sequences assessed for their antigenicity through the VaxiJen 2.0 server for immune response. For further analysis, the most antigenic sequence among the five probable vaccine candidates were isolated.

2.13 Physicochemical and Allergenicity Evaluation of Selected Vaccine Sequence:

AllerTOP 2.0 server was used to predict allergenicity. Physicochemical properties of the constructed vaccine sequence were predicted using the ProtParam web-server to determine stability, half-life and other parameters^[47]. Additionally, solubility of the vaccine construct was predicted by the SOLpro tool in the SCRATCH suite^[48].

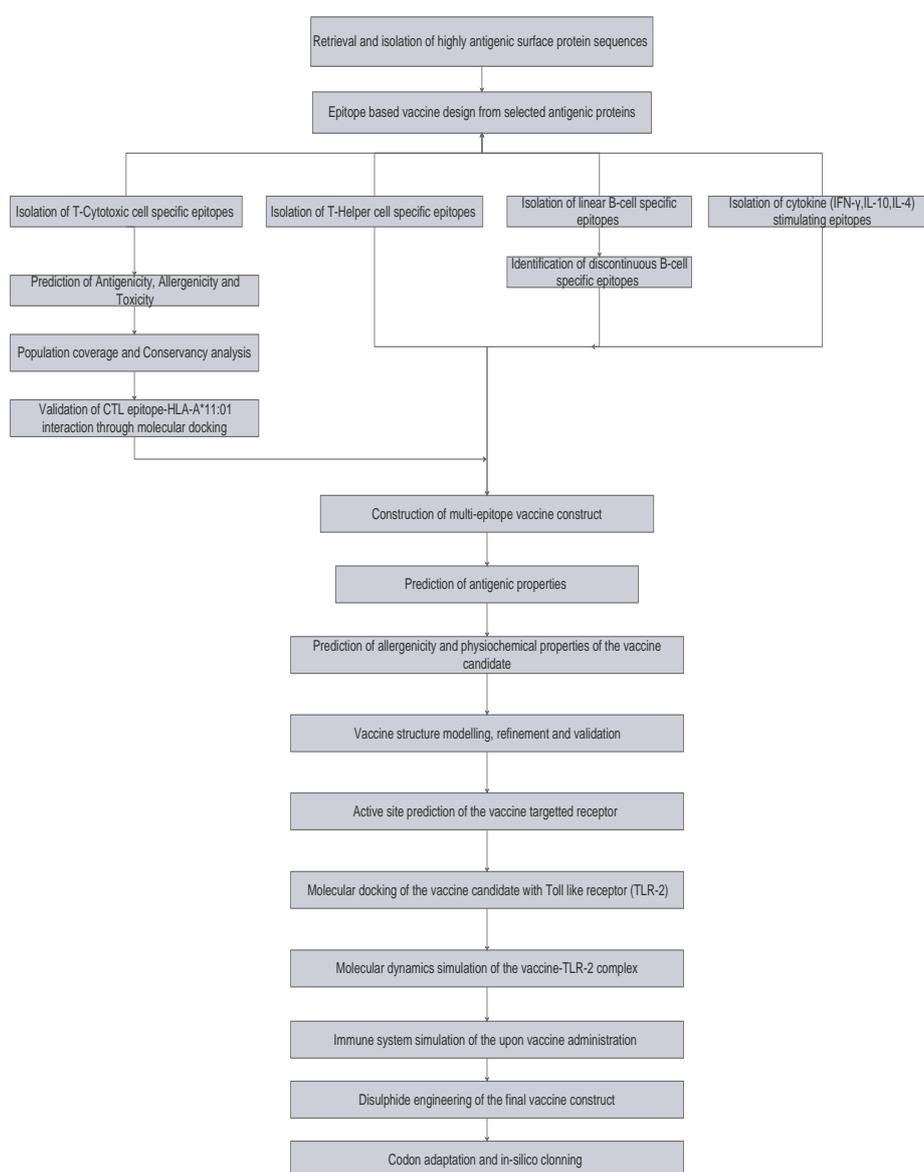


Figure 1: Schematics of the methodology implemented in chimeric vaccine development.

2.14 Vaccine Structure Modeling, Refinement, and Validation:

The secondary structure of the vaccine sequence was generated through PSIPRED 4.0 server^[49]. vaccine sequence was then converted into 3D structure using homology modelling based Phyre2 web server^[50]. Although computational homology modeling methods perform very well to generate an accurate model but most models do not reach experimental accuracy and therefore are not always useful for the application. Hence protein structure refinement is required to bring the model closer to the native state. GalaxyRefine web server was used to perform structure refinement^[51]. Subsequently, structural validation and model quality assessment was performed using Pros-SA server^[49]. It analyzes the protein structure and compares with the X-Ray and NMR based experimental structures^[52]. To analyze specific non-bonded atomic interactions ERRAT server was used^[53]. Finally, the Ramachandran plot was generated using the PROCHECK server. It calculates possible phi and psi angles that account for the amino acid residues^[54].

2.15 Active Site Prediction of Receptor:

Toll-like Receptor-2 (TLR2) plays a critical role recognizing viral proteins on innate immunity-based cells such as macrophages (MA), dendritic cells (DC), B-cell, T-cell particularly T-regulatory cells^[81]. SCFBio server (<http://www.scfbio-iitd.res.in/dock/ActiveSite.jsp>) was used for prediction of the ligand binding site on TLR-2.

2.16 Protein-protein Docking between Vaccine and TLR-2:

Molecular docking is a study of how two molecular structures are perfectly fit to match together and also evaluate binding affinity^[55, 56]. TLR-2 agonist namely OmpA protein was used as an adjuvant^[57]. It's binding to TLR-2 (PDB ID: 3A7B) was validated employing the vaccine construct as a ligand for docking analysis. Molecular docking was executed in the HDOCK server^[58]. Binding affinity (Kcal/mol) and K_D were predicted with the help of PRODIGY server^[80]. Subsequently, the binding energy of the complex was calculated by molecular mechanics generalized Born surface area (MM/GBSA) method using Hawkdock server^[82]. The protein-protein interaction hotspots were identified using Spot-on server^[85]. Further, hot-regions of interaction were isolated with the help of Hot-region server^[86]. Contact map representing inter-residue interaction was derived from COCOMAPS server^[87].

2.17 Molecular Dynamics simulation:

Stability of protein-protein can be determined by comparing essential dynamics with their normal modes^[59, 60]. iMODS server was used to interpret the collective motion through normal mode analysis (NMA)^[19]. The server calculated the position and magnitude of the complex's immanent motions in terms of deformability, B-factors, eigenvalues, and covariance. Similarly, atomic mobility and inter-residue network was predicted by Dynamics server^[83].

2.18 In-silico evaluation of Immune Response: To assess the immunogenic potential of the final vaccine construct, *in-silico* immune response after vaccine administration was measured using C-Immsim server^[60]. The server uses a position-specific scoring matrix (PSSM) and machine learning techniques respectively for predicting epitopes and their immune interactions^[61]. All parameters were set at default with the time steps being set at 1, 84, and 168 wherein each time step represented 8 hours. Booster doses were administered on day 30 and day 60 respectively. The levels of antibodies, cytokines, B-cell, T-cell, DCs and macrophages were measured.

2.19 Disulfide Engineering of Final Vaccine Construct:

Disulfide bonds are covalent bonds that play a major role in the stabilization of protein structure. Disulfide engineering of the vaccine construct was executed with the help of the Disulfide by Design 2.12 web tool^[62].

2.20 Codon Adaptation and In-silico Cloning:

Codon adaptation is important to accelerate the rate of expression of codon use in the prokaryotic organisms. Codon adaptation was undertaken using Java Codon Adaptation Tool (Jcat) server wherein *Escherichia coli* strain K12 was selected as a host^[63]. Codon adaptation index (CAI) and GC content of the adapted sequence were obtained^[64]. Furthermore, *E. coli* pET28a(+) was used as a vector for *in-silico* cloning utilizing the SnapGene 4.2 software.

3. RESULT:

3.1 Retrieval of Protein Sequence and Antigenicity Prediction:

Nine potential protein sequences of SARS-CoV-2 were retrieved from NCBI. Among the nine protein sequences, the envelope (E) protein and membrane (M) protein showed the highest antigenicity [Table 1]. These protein sequences are chosen for deriving antigenic epitopes.

Table 1: Retrieved protein sequences and their antigenicity

Name protein	Accession number	Antigenicity
membrane glycoprotein	QIV15191.1	0.5102
spike glycoprotein	QIC53213.1	0.4646
leader protein	YP_009742608.1	0.4064
envelope protein	YP_009724392.1	0.6025
ORF3a protein	QIC53205.1	0.5094
nucleocapsid protein	QIC53221.1	0.5025
ORF1a polyprotein	QIS60287.1	0.4786
nsp2	YP_009742609.1	0.4021
nsp4	YP_009742611.1	0.4699

3.2 Cytotoxic T-lymphocyte (CTL) Epitopes and MHC-I Binding Alleles Prediction:

The 126 CTL epitopes from M protein and 51 CTL epitopes from E protein were obtained from NetCTL II server. Subsequently, 80 unique epitopes (M protein) and 27 unique epitopes (E protein) were isolated. Those unique epitopes were further analyzed with different parameters like antigenicity, allergenicity, toxicity and immunogenicity. Further, 20 epitopes from M protein and 12 epitopes from E proteins were found to be suitable. Based on maximum MHC-I binding alleles, four CTL epitopes from each protein were derived [Table 2].

3.3 Prediction of Helper T-lymphocyte (HTL) Epitopes Along with MHC-II Binding Alleles:

A total of 208 HTL epitopes were obtained from M protein of which 17 epitopes were found to induce cytokine secretion. Based on maximum MHC-II binding alleles three HTL epitopes were selected [Table 3]. Similarly, from 61 epitopes of E protein, four potential HTL epitopes were capable of inducing cytokine production and three epitopes are selected for further analysis based on maximum MHC-II binding potential [Table 3].

3.4 Prediction of Linear B-lymphocyte (LBL) and discontinuous epitopes: A total of 20 potential LBL epitopes were derived from M protein while seven LBL epitopes were obtained from E protein. Two LBL epitopes

from each protein were isolated based on low toxicity and allergenicity with high antigenicity [Table 4]. The ElliPro of IEDB tool was used for discontinuous B-cell epitope prediction wherein seven discontinuous epitopes were derived. The PI score for discontinuous B-cell epitopes ranged between 0.559 and 0.787.

3.5 Molecular Docking Analysis between CTL Epitopes and HLA-Alleles: Molecular docking CTL-HLA-A*11:01 (PDB ID: 4UQ2) results are shown in [Table 5]. The result showed stable interaction between 'LTWICLLQF' epitope of M protein and HLA-A*11:01 with a binding affinity (ΔG) of -12.5 Kcal/mol with K_D of $6.4E-10$ at $25^\circ C$. While the envelope protein-based CTL 'LLFLAFVVF' epitope showed binding affinity of -8.1 Kcal/mol with K_D of $1.2E-6$ at $25^\circ C$. The CTL epitope 'NSVLLFLAF' demonstrated highest binding affinity among envelope-based proteins which is -10 Kcal/mol with K_D of $4.9E-8$ at $25^\circ C$. The docking poses of both LTWICLLQF- HLA- A*11:01 and LLFLAFVVF- HLA- A*11:01 complexes are represented in [Figure 2A and 2B]

3.6 Population Coverage Analysis:

In this study, the selected epitopes and their MHC-I and MHC-II restricted alleles were shown to cover an average 99.24% population across the world [Figure 3]. The population coverage by country is given in [Table 6].

Table 2: Isolated T-Cytotoxic cell specific epitopes for multi-epitope vaccine development

Protein	MHC-I Super-type	Isolated Peptide	Antigenicity	Binding alleles	Immunogenicity	Allergenicity	Toxicity
Membrane protein	A2	FVLAAYVYRI	0.5136	25	0.13985	Non-Allergen	Non-Toxin
	A2	FLFLTWICL	1.4835	21	0.35397	Non-Allergen	Non-Toxin
	A2	NLVIGFLFL	1.2917	16	0.34956	Non-Allergen	Non-Toxin
	B58	LTWICLLQF	1.1393	16	0.06584	Non-Allergen	Non-Toxin
Envelope protein	A2	FLLVTLAIL	0.5508	21	0.17608	Non-Allergen	Non-Toxin
	A2	SVLLFLAFV	0.5155	16	0.19022	Non-Allergen	Non-Toxin
	B8	LLFLAFVVF	0.6301	16	0.2341	Non-Allergen	Non-Toxin
	B62	NSVLLFLAF	0.5256	21	0.11514	Non-Allergen	Non-Toxin

Table 3: Isolated T-Helper cell specific epitopes for multi-epitope vaccine development

Protein	Isolated peptide	Antigenicity	Immunogenicity	Binding Alleles	INF- γ induction	IL-10 induction	IL-4 induction
Membrane Protein	FLYIIKLIFLWLLWP	0.42	94.6742	26	Positive	Positive	Positive
	RFLYIIKLIFLWLLW	0.4833	93.4588	26	Positive	Positive	Positive
	YIIKLIFLWLLWPVT	0.4298	93.1993	17	Positive	Positive	Positive
Envelope Protein	LLFLAFVVFLLVTLA	0.8122	89.6735	34	Positive	Positive	Positive
	VNSVLLFLAFVVLL	0.4339	86.7887	17	Positive	Positive	Positive
	GTLIVNSVLLFLAFV	0.426	96.8955	18	Positive	Positive	Positive

Table 4: Isolated Linear B-cell specific epitopes for multi-epitope vaccine development

Protein	Peptide	Probability	Antigenicity	Allergenicity	Toxicity
Membrane Protein	LLESELVIGAVI	0.74	0.5385	Non-Allergen	Non-Toxin
	IAIAMACLVLGM	0.73	0.9515	Non-Allergen	Non-Toxin
Envelope Protein	AILTALRLCAYC	0.65	0.786	Non-Allergen	Non-Toxin
	LLVTLAILTALR	0.59	0.6581	Non-Allergen	Non-Toxin

Table 5: Docking results of identified CTL epitopes and HLA-A*11:01 complex

	Isolated CTL Epitope	ΔG (kcal/mol)	K_D (M)
Envelope protein	LLFLAFVVF	-8.1	1.20xE-06
	FLLVTLAIL	-9.4	1.40xE-07
	NSVLLFLAF	-10.0	4.90xE-08
	SVLLFLAFV	-9.0	2.70xE-07
Membrane protein	LTWICLLQF	-12.5	6.40xE-10
	NLVIGFLFL	-9.4	1.20xE-07
	FLFLTWICL	-9.5	1.10xE-07
	FVLAAYVRI	-10.7	1.40xE-08

Table 6: Population coverage of MHC-I and MHC-II targeted epitopes used in the chimeric vaccine development

Country	Coverage
Austria	99.92%
Belgium	98.28%
Portugal	99.72%
China	94.28%
Denmark	90.64%
Germany	99.89%
India	97.64%
Italy	99.59%
Japan	98.93%
Korea; South	99.17%
Russia	98.87%
Singapore	96.62%
Spain	99.28%

3.7 Conservancy Analysis:

Epitope conservancy analysis of the IEDB tool used for this analysis and the results shown that all the epitopes (B-cell and T-cell) were 100% conserved for both M and E protein. It is expected that the use of conserved epitopes can provide broader protection across several strains or even species. The conservancy analysis result is shown in [Table 7].

Table 7: Epitope conservancy analysis

Specific ity	Membrane protein		Envelope protein	
	Epitope	Identy	Epitope	Identy
MHC-I	FVLAAYVRI	100%	FLLVTLAIL	100%
	FLFLTWICL	100%	SVLLFLAFV	100%
	NLVIGFLFL	100%	LLFLAFVVF	100%
	LTWICLLQF	100%	NSVLLFLAF	100%
	FLYIIKLIFLWL LWP	100%	LLFLAFVVFL VTLA	100%
MHC-II	RFLYIIKLIFLW LLW	100%	VNSVLLFLAFV VFLL	100%
	YIIKLIFLWLL WPVT	100%	GTLIVNSVLLFL AFV	100%
	LLESELVIGAV I	100%	AILTALRLCAY C	100%
B-cell	IAIAMAQLVGL M	100%	LLVTLAILTAL R	100%

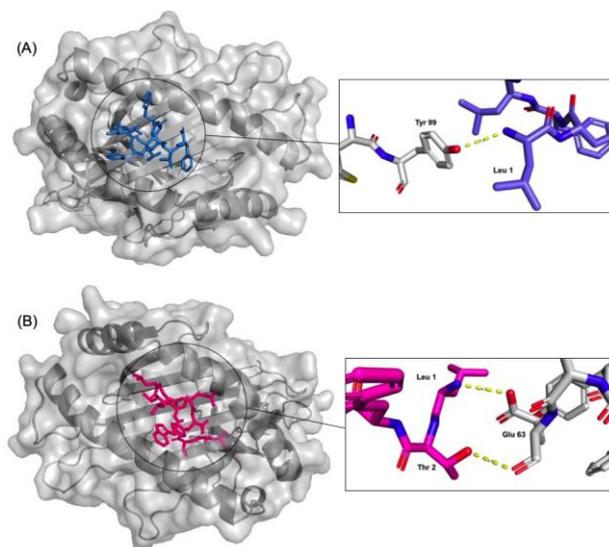


Figure 2: (A) Simulated interaction between the T-Cytotoxic cell specific epitope LLFLAFVVF (blue) and HLA-A*11:01 (grey). (B) Simulated interaction between the T-Cytotoxic cell specific epitope LTWICLLQF (pink) and HLA-A*11:01 (grey).

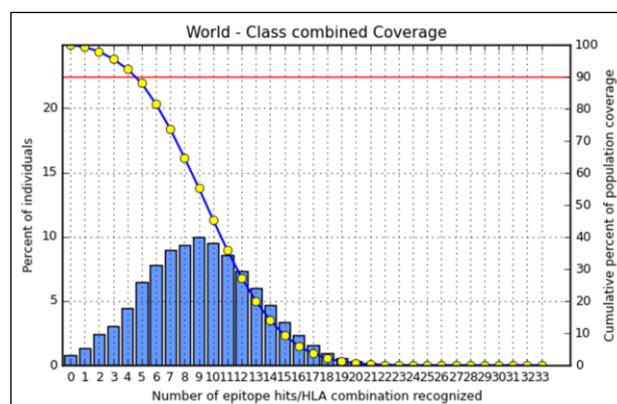


Figure 3: Population coverage of isolated MHC-I and MHC-II epitopes used in the vaccine candidate.

3.8 Multi-epitope chimeric vaccine candidates and Antigenicity Prediction:

From M and E protein, eight CTL, six HTL, and four LBL epitopes were selected for final vaccine construct. CTL, HTL, and LBL epitopes were connected by AAY, GPGPG, and KK linkers respectively. To boost immunogenicity, TLR-2 agonist adjuvant sequence was added to the vaccine construct by means of EAAAK linker. In this study, five different adjuvants-based vaccine candidates were assessed for their immunogenicity (V1, V2, V3, V4, V5). PADRE sequence also added to increase efficacy and potency of the vaccine construct. Antigenicity of all potential candidates were evaluated and the vaccine sequence with the highest antigenicity was used for further analysis. The vaccine candidate 'V5' was chosen due to its antigenicity value of 0.6606. It bears OmpA adjuvant protein in the N-terminal along with PADRE sequence. The outline of multi-epitope chimeric vaccine construct V5 is represented in [Figure 4].

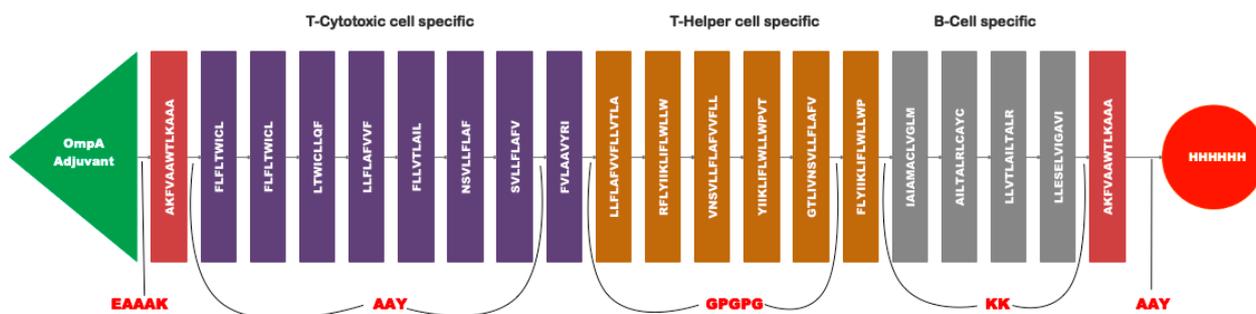


Figure 4: Outline of the designed Multi-epitope chimeric vaccine construct.

3.9 Physicochemical properties and secondary structure of the Vaccine Construct:

The vaccine construct was found to have a theoretical isoelectric point (pI) of 9.54 suggesting the protein to be basic in nature [Table 8]. The molecular weight was calculated to be around 71.43 kilodaltons. The vaccine construct is highly stable with an instability index (II) of 19.97 (>40 implies instability). The predicted half-life of the protein was 30 hours in mammalian reticulocytes (*in-vitro*), >20 hours (yeast, *in-vivo*) and >10 hours (*Escherichia coli*, *in-vivo*). The Grand average of hydropathicity (GRAVY) was estimated to be 0.590 indicating hydrophobic in nature [Table 9]. This is due the presence of multiple hydrophobic pockets and respective aminoacids that facilitate protein-protein interactions through non-covalent bonding. The vaccine protein is soluble when over-expressed with a probability of 0.96609. The aliphatic index of the vaccine construct was 116.56 implying high thermo-stability. The secondary structure by PSIPRED revealed, the structure to constitute 50% helix, 16.97% strand, and 33.03% coil [Figure 5]. Only 16% of the sequence was found to be dis-ordered.

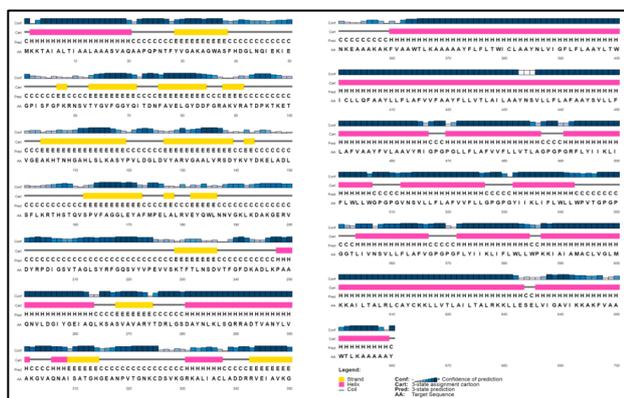


Figure 5: Secondary structure of the chimeric vaccine construct.

Table 8: Physicochemical Properties of selected vaccine construct

Formula	C ₃₃₇₃ H ₅₂₄₅ N ₈₂₃ O ₈₅₉ S ₁₁
Mol. Weight (Dalton)	71423.43
Amino acids	660
Theoretical pI	9.54
Half-life (hours)	30
Instability index	19.97
Aliphatic index	116.56
Negatively charged residues (Asp + Glu)	40
Positively charged residues (Arg + Lys)	66
Solubility	Soluble with 0.96609 probability

3.10 Tertiary Structure Prediction and Refinement: The homology modelling method of the Phyre2 server was used for modelling the 3D structure of vaccine protein candidate (V5). Protein structure refinement was executed with the GalaxyRefine server. Five potential refined models were obtained. The result of the best-refined model showed 92.1% of Rama favored region, GDT-HA score 0.9163, RMSD value 0.498, MolProbity score 2.208, and poor rotamers 0.8. Model-2 was selected for further analysis [Figure 6C] [Table 9].

3.11 Validation Of refining vaccine protein: The Ramachandran plot showed 87.1% residues in the most favored region, 9.8% residues in the additional allowed region, 0.7% residues in the generously allowed region and 2.4% in the disallowed region [Figure 6A]. The z-score was calculated to be -2.82 in the ProSA-web server for vaccine protein [Figure 6B] and the construct was near experimentally derived structures. The ERRAT score for the vaccine construct was 83.6127% (quality factor >80% is acceptable). The overall quality of the protein structure was satisfactory for further analysis.

Table 9: 3D structure refinement results

Model	GDT-HA	RMSD	MolProbity score	Poor rotamers	Rama favored
MODEL 1	0.9174	0.504	2.296	1.2	90.4
MODEL 2	0.9163	0.498	2.208	0.8	92.1
MODEL 3	0.9129	0.507	2.218	0.8	90.9
MODEL 4	0.9163	0.499	2.224	1	91.8
MODEL 5	0.9117	0.5	2.216	1	91.5

3.12 Active Site Prediction of TLR2 Receptor:

SCFBio server predicted the geometrical shape of the active site as 0.7686 in x-center, -21.1211 in y-center and 30.8294 in z-center. The residue of the active site is shown in [Supplementary Figure 2].

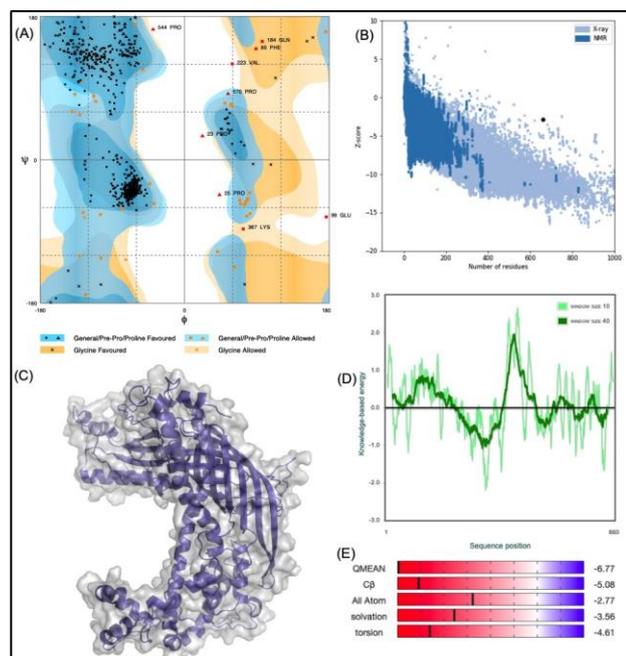


Figure 6: (A) Ramachandran plot for multi-epitope vaccine construct (B) Z-plot demonstrating the proximity of vaccine construct to the experimental models derived through X-ray and NMR analysis (C) Homology modelling based refined vaccine structure (D) Knowledge based energy per residue (E) Predicted QMEAN4 scores for the vaccine construct.

3.13 Analysis of Protein-protein Docking Results:

Molecular docking was executed in the HDock server where the vaccine (V5) was considered as a ligand and TLR2 as a receptor protein. The server provided ten potential docking positions based on their docking score. The first docking position out of ten docking poses showed the lowest docking score (-322.47 kcal/mol) with reduced RMSD (109.52 Å). Hence, the first docking position was selected as the ideal docking position shown in [Figure 7C]. The binding affinity (ΔG) was found to be -16.4 Kcal/mol and K_D of 8.7×10^{-13} at 25°C which implies a stable interaction. Further, the binding affinity of the complex using MM/GBSA method was calculated to be -13.21 Kcal/mol. The total interface area (Å^2) between the receptor and ligand was 2423.85. There were 142 residues at the area of interface [Supplementary Table 4]. Ten hydrogen bonding interactions were observed during molecular docking [Supplementary Table 3]. Interestingly, 327 Ser of the vaccine construct which was found to participate in hydrogen bonding was also identified as an interaction hot-spot thereby validating the molecular docking [Figure 7B] [Supplementary table 1]. The residues 411 Leu, 413 Leu and 415 Phe of the vaccine construct were determined to be Hot-regions of interaction [Supplementary table 2].

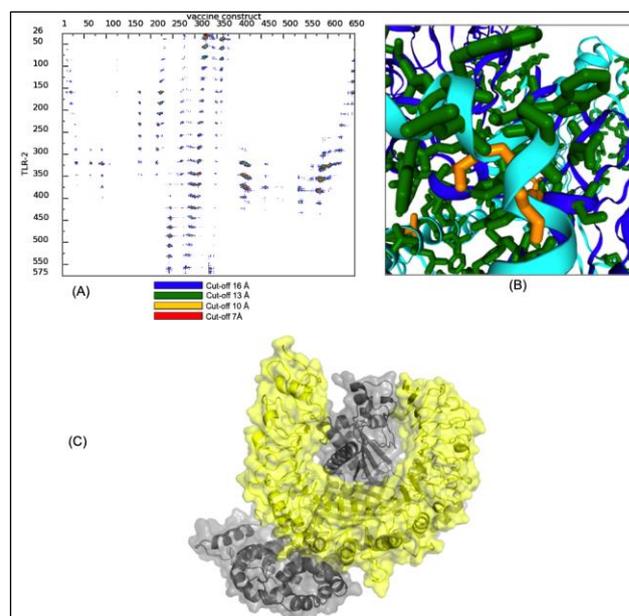


Figure 7: (A) Contact map showing the inter-molecular distance between vaccine and TLR-2 residues (B) Interaction hot-spots (orange) between TLR-2 (cyan) and vaccine construct (blue) (C) Simulated interaction involving the vaccine construct (grey) and TLR-2 (yellow).

3.14 Molecular Dynamics Simulation:

Molecular dynamics revealed the mobility around Vaccine-TLR2 interface was restricted while some degree of mobility was observed at the ends [Figure 8G]. The Eigenvalue found for the complex was 3.988793×10^{-5} which is significantly high, implying a stable complex formation [Figure 8C]. The values of the B-factor inferred via NMA were RMS-equivalent [Figure 8A] and [Figure 8B]. The variance [Figure 8D] associated with each normal mode has been inversely related to the eigenvalue^[65, 66]. The deformability of the complex depends on the individual distortion of each residue which is represented by hinges in the chain [Figure 8A]. Coupling between residue pairs was shown by the covariance matrix where various pairs displayed correlated, anti-correlated, and uncorrelated motions respectively defined by red, blue, and white colors [Figure 8E]. An elastic network generated level of stiffness indicates the association of atoms along the complex as springs [Figure 8F]. The atomic network is relatively elastic shown by the light grey dots on the residue map [Figure 8F]. The interacting residues of vaccine-TLR-2 are shown in [Figure 8H].

3.15 In-silico Evaluation of Immune Response:

The C-ImmSim immune simulator server considers the efficient immune responses of the cell state and models the immune system. The results of immune simulation on the C-ImmSim server confirmed compatibility with real immune responses. The primary response was observed when IgM concentration increased upon vaccine administration. In the response to the booster doses, secondary and tertiary immune reaction showed a significant increase in the concentration of IgG1+IgG2, IgM, and IgG+IgM antibodies [Figure 9A]. A consistent

increase in the B-cell and T-Helper cell populations were noted at day 30 and day 60 of which majority of them were in active phase [Figure 9B and 9C]. An adequate amount of cytokine production was detected [Figure 9D]. Interestingly, a strong IFN- γ response was visible implying an anti-viral potential of the vaccine construct. The levels peaked with the administration of booster doses on day 30 and day 60. Similarly, a modest level of other cytokines including Interleukin-10 were observed. The dendritic cell and macrophage activity were also visible [Figure 9E and 9F]. Both humoral and cell-mediated immune system was activated upon vaccination [supplementary report 1 and supplementary report 2].

3.16 Disulfide Engineering:

Disulfide engineering of the final modeled structure was performed for the purpose of stabilization where highly unstable protein sequence regions were mutated with cysteine. A total of 60 amino acid residue pairs were identified by the DbD2 server to enable disulfide engineering. However, five residue pairs [Table 12] were chosen for mutation after evaluation on parameters such as energy, chi3, and B-factor i.e. the energy value was less than 2.2 and chi3 (χ_3) should be between -87 to +97 [84]. Therefore, LEU8-ALA14, GLY69-GLU79, GLY190-ARG203, TYR258-LYS302, PHE565-GLY598 [Figure 10] residues were mutated to cysteine residues.

3.17 Codon Adaptation and In-silico Cloning:

The aim of codon optimization was to maximize expression of viral based vaccine protein into the E. coli K12 strain host system by removing existing codon bias. A 1980 base pairs long optimized nucleotide sequence was obtained from Java codon Adaptation Tool (Jcat). The GC-content of optimized codon sequence was found to be 51.8687% which is within the optimal range of 30% to 70%. Codon Adaptation Index (CAI) value was calculated to be 0.93666 (the closer the CAI value is to 1, the better the adaptation) which implies good expression levels. Finally, the optimized codon was inserted between XhoI and BamHI restriction sites at multiple cloning site (MCS) of the E. coli plasmid vector

pET28a(+) where the target area is shown in red color [Figure 11]. Additionally, a poly-histidine tag (6xHis) was adjoined at the C-terminal to facilitate protein purification [Figure 4] and [Figure 11].

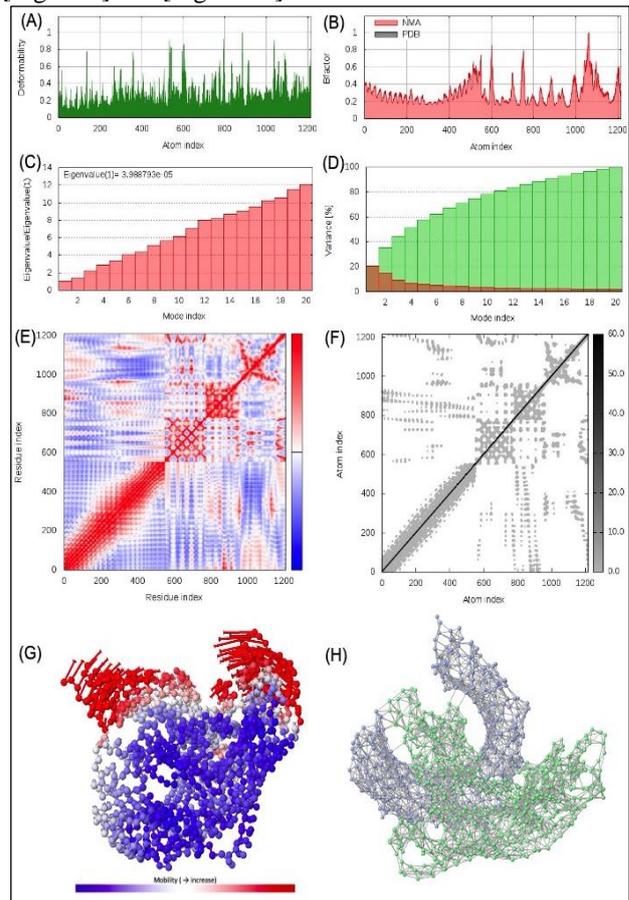


Figure 8: Molecular dynamics simulation analysis of vaccine protein-TLR2 complex. Protein-protein complex stability was investigated through (A) B-factor (B) eigenvalue (C) variance (D) deformability (E) covariance (F) elastic network analysis (G) Atomic Mobility and (H) Inter-residue interaction network.

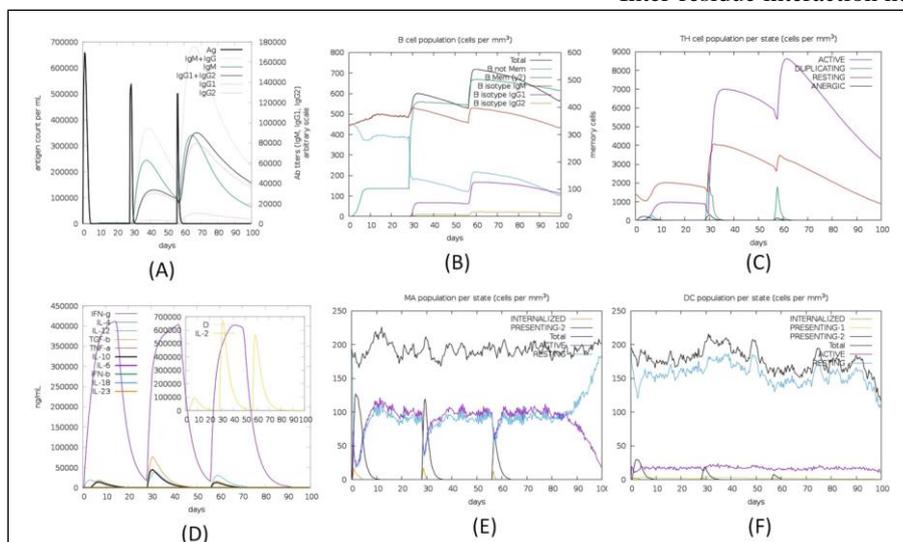


Figure 9: Simulated immune response after administration of the vaccine candidate (A) Antigen and immunoglobulins per millilitre (B) B-lymphocytes cell population (C) CD4+ helper T-cells population per state (D) Induced levels of the cytokine and Simpson index D (E) Macrophages population per state (F) Dendritic cell population per state.

Table 10: Residues identified for Disulphide engineering

Residue 1		Residue 2		Bond		
Sequence number	Aminoacid	Sequence number	Aminoacid	χ_3	kcal/mol	Σ B-factor
8	LEU	14	ALA	91.99	1.85	0
69	GLY	79	GLU	-86.05	1.57	0
190	GLY	203	ARG	-83.24	1.58	0
258	TYR	302	LYS	95.66	0.67	0
565	PHE	598	GLY	-74.18	1.47	0

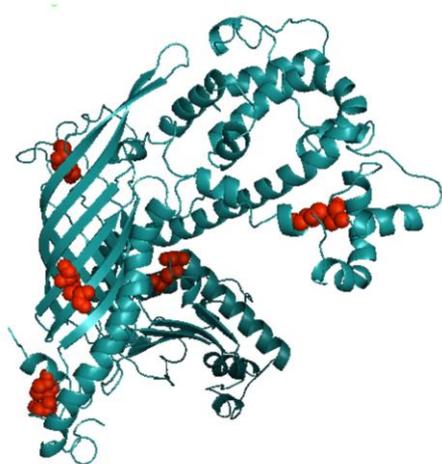


Figure 10: Disulphide engineered vaccine construct wherein mutated pairs are represented as red spheres.

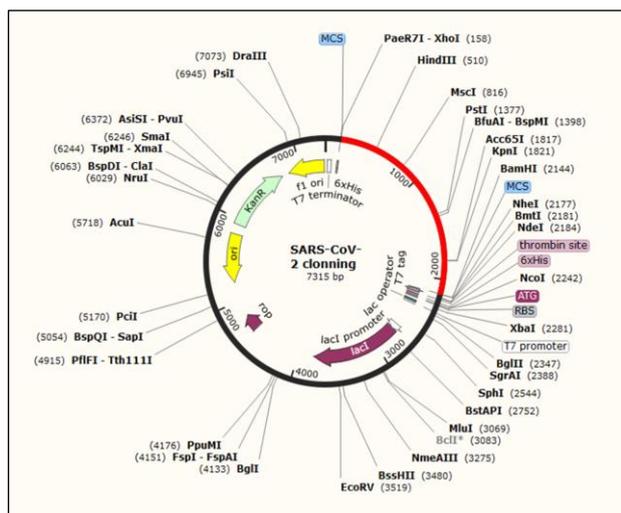


Figure 11: *In-silico* restriction cloning of the vaccine gene sequence into pET-28a(+) expression vector shown as red region encompassed between XhoI (173) and BamHI (489). Figure 11: *In-silico* restriction cloning of the vaccine gene sequence into pET-28a(+) expression vector shown as red region encompassed between XhoI (173) and BamHI (489).

4.0 DISCUSSION:

SARS-CoV-2 is a potent contagion known for a global pandemic infecting many populations across the world, thus developing an effective vaccine is one of the best ways to

contain transmission. Designing a novel vaccine candidate in short period of time would be time-consuming and expensive. There is additional risk associated with live or attenuation based vaccines which contain unwanted allergens^[67] Multi-epitope subunit vaccines eliminate the presence of allergens thus reducing adverse effects^[68]. Cost-effectiveness, safety, improved potency and immunogenicity of conserved epitopes are the pillars of epitope based chimeric vaccine design ^[3].With the availability of protein sequence database and help of advanced *in-silico* based analysis, it is now possible to design a peptide vaccine based on immunoinformatics through different computational tools. Particularly, this approach showed potency against the Oropouche virus^[69], Rift valley fever virus^[39], Dengue virus^[70], etc. Acknowledging the benefits of multi-epitope vaccines, a novel multi-epitope chimeric vaccine was intended to be designed against COVID-19.

Similar attempts were made to target the spike (S) glycoprotein through immune informatics ^[15, 71]. However, the importance of M and E protein for viral entry, replication and infection were avoided^[72, 73]. The membrane based structural glycoproteins offer an ideal target ^[74-76]. They showed good antigenicity [Table 1] among retrieved proteome sequences thus we selected M and E protein to design a vaccine model in this study.

Vaccines trigger T-cell and B-cell based responses simultaneously. A vaccine stimulates plasma cells that are responsible for antibody production and pathogenic recognition^[77]. However, memory B-cells play a crucial role in recognizing the pathogenic epitope profile to prevent disease. The CD8+ and CD4+ T-cells impart necessary antiviral immunity^[78]. It is important to derive CTL, HTL, and LBL epitopes which are needed to elicit T-cell and B-cell response. Therefore, a potent multi-epitope vaccine should contain CTL, HTL, and LBL epitopes to provoke an innate immune response against specific incoming pathogen^[68].

Based on antigenicity, allergenicity, toxicity, cytokine production, and highest binding affinity of HLA molecules we selected eight CTL and six HTL epitopes from both M and E proteins using various computational tools. Four LBL epitopes from both M and E protein were also considered after assessing their antigenicity, allergenicity, and probability score from both protein sequences. Roy, Tonmoy et al. (2020) isolated antigenic epitopes from S, M, and E proteins. However, it lacked the design of chimeric vaccine construct and its interaction with TLR^[79]. Another study targeted E protein for predicting epitopes with the aim of constructing a vaccine, but no B-cell epitopes were predicted ^[1]. The current study targeted both M and E

protein for predicted epitopes and identified 8 CTL, 6 HTL and 4 LBL epitopes which are highly antigenic, non-allergenic, and non-toxic. This study also confirmed that the predicted epitopes have not been reported yet. Validation of CTL-HLA-A*11:01 interaction was performed to demonstrate efficacy. The epitope 'LTWICLLQF' showed the highest binding affinity of -12.5 Kcal/mol and K_D of $6.4E-10$ at 25°C. Similarly, promising LBL and HTL epitopes were also isolated. Five different vaccine candidates with five different adjuvants along with the PEDRE sequence were assembled. Studies show, the presence of PADRE sequence in a vaccine construct demonstrated better CTL response^[80]. Each epitope was joined by suitable linkers to ensure a better protective response. Adjuvants are needed to boost the innate immune response. Among the five constructed vaccine sequences, vaccine candidate V5 was chosen based on antigenicity. The vaccine candidate was found to be non-allergen and non-toxic. The molecular weight of the V5 construct is 71.4 KDa, the physicochemical studies have also shown that the vaccine protein is basic in nature, thermo-stable and soluble. The secondary structure of the vaccine construct was analyzed before generating a 3D model. Refinement of the model was performed and the best model was selected based on RMSD, MolProbity, and Rama favored value. The active site of TLR-2 was predicted and molecular docking was performed. Stable interaction between the vaccine construct and TLR-2 was confirmed by the binding affinity (ΔG) of -16.4 Kcal/mol, K_D of $8.7xE-13$ at 25°C and MMGBSA based binding affinity of -13.21 Kcal/mol. Further, normal mode analysis based molecular dynamics showed a high eigenvalue of $3.988793e^{-05}$ validating the executed docking simulation. Immune simulation showed the induction of both humoral and cell-mediated based response. Disulfide bridging was undertaken to improve stability. Finally, the vaccine sequence was optimized for E. coli strain K12 before insertion within the pET28a(+) vector for gene cloning and expression. A poly-histidine (6xHis) tag was added at the C-terminal to facilitate protein purification through the broadly available Nickel-Nitrilotriacetic acid (Ni-NTA) sepharose column. The designed vaccine candidate holds promise but to continue the vaccine development process *in-vitro* and *in-vivo* serological assays are highly recommended.

5.0 CONCLUSION:

The control of COVID-19 pandemic requires strengthening many areas of public health including diagnostics, epidemiology and drug development. Preventive strategies such as vaccines and mathematical modelling can decrease infection rates and their associated fatalities. The *in-silico* epitope-based vaccine design complements this effort. The multi-epitope subunit vaccine against SARS-CoV-2 successfully elicited both humoral and cell-mediated response during immune simulation. The recommended next steps would include *in-vivo* and *in-vitro* serological assays to further the vaccine development process.

Conflict of interest:

The authors declare no existing conflict of interest.

Acknowledgement:

The authors like to acknowledge Nouman Ali from University of Central Punjab and also Zulkar Nain from Islamic University for their technical support and useful comments.

Author Contributions:

Joy Dip Barua (JDB), Niloy Das (ND), Najia Absar Munia (NAM) conceived and executed the idea. JDB derived epitopes using immunoinformatics. ND executed docking and molecular dynamics simulation. Suraj Raju (SR) and NAM performed immune-simulation analysis. SR and JDB compiled the data, illustrated the figures and tables. SR, JDB and ND wrote the manuscript. SR supervised the whole work. All authors approved the final manuscript.

Funding Information:

This research received no particular grant from public, commercial or not-for-profit funding agencies.

Ethical approval: Not applicable.

Code availability: Not applicable.

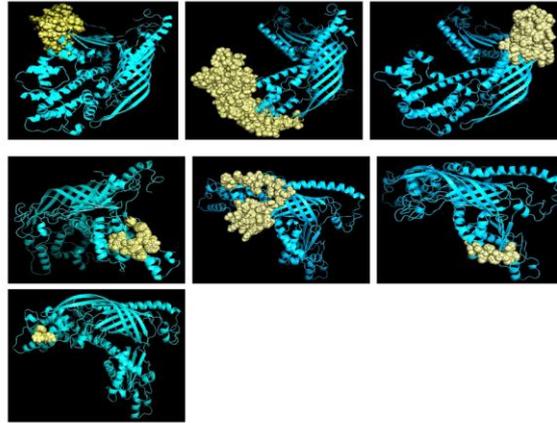
Data availability: Provided as supplementary material.

REFERENCES

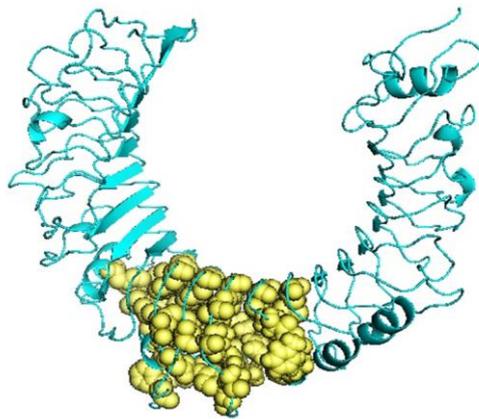
1. Abdelmageed, M.I., et al., *Design of a Multi-epitope-Based Peptide Vaccine against the E Protein of Human COVID-19: An Immunoinformatics Approach*. BioMed Research International, 2020. 2020.
2. Hui, D.S., et al., *The continuing epidemic threat of novel coronaviruses to global health-the latest novel coronavirus outbreak in Wuhan, China*. International Journal of Infectious Diseases, 2020.
3. Shey, R.A., et al., *In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases*. Scientific reports, 2019. 9(1): p. 1-18.
4. Gallegos, A., *WHO Declares Public Health Emergency for Novel Coronavirus*. Medscape Medical News, 2020.
5. Wee, S.-L., D.G.M. Jr., and J.C. Hernández, *W.H.O. Declares Global Emergency as Wuhan Coronavirus Spreads*. The New York Times, 2020.
6. Times, T.N.Y., *Coronavirus Live Updates: W.H.O. Declares Pandemic as Number of Infected Countries Grows*. The New York Times, 2020.
7. Huang, C., et al., *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China*. The lancet, 2020. 395(10223): p. 497-506.
8. M.T., u.Q., et al., *Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants*. J Pharm Anal, 2020.
9. Astuti, I., *Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response*. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 2020.
10. Khailany, R.A., M. Safdar, and M. Ozaslan, *Genomic characterization of a novel SARS-CoV-2*. Gene Reports, 2020: p. 100682.
11. Jiang, S., C. Hillyer, and L. Du, *Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses*. Trends in immunology, 2020.
12. Fehr, A.R. and S. Perlman, *Coronaviruses: an overview of their replication and pathogenesis*, in *Coronaviruses*. 2015, Springer. p. 1-23.
13. Walls, A.C., et al., *Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein*. Cell, 2020.
14. Schoeman, D. and B.C. Fielding, *Coronavirus envelope protein: current knowledge*. Virology journal, 2019. 16(1): p. 69.
15. Peele, K.A., et al., *Design of multi-epitope vaccine candidate against SARS-CoV-2: a in-silico study*. Journal of Biomolecular Structure & Dynamics, 2020: p. 1.

16. Bachler, B.C., et al., *Novel biopanning strategy to identify epitopes associated with vaccine protection*. Journal of virology, 2013. **87**(8): p. 4403-4416.
17. Black, M., et al., *Advances in the design and delivery of peptide subunit vaccines with a focus on toll-like receptor agonists*. Expert review of vaccines, 2010. **9**(2): p. 157-173.
18. Perrie, Y., et al., *Recent developments in particulate-based vaccines*. Recent patents on drug delivery & formulation, 2007. **1**(2): p. 117-129.
19. López-Blanco, J.R., et al., *iMODS: internal coordinates normal mode analysis server*. Nucleic acids research, 2014. **42**(W1): p. W271-W276.
20. Lo, Y.-T., et al., *Prediction of conformational epitopes with the use of a knowledge-based energy function and geometrically related neighboring residue characteristics*. BMC bioinformatics, 2013. **14**(S4): p. S3.
21. L Dudek, N., et al., *Epitope discovery and their use in peptide based vaccines*. Current pharmaceutical design, 2010. **16**(28): p. 3149-3157.
22. Purcell, A.W., J. McCluskey, and J. Rossjohn, *More than one reason to rethink the use of peptides in vaccine design*. Nature reviews Drug discovery, 2007. **6**(5): p. 404-414.
23. Dermime, S., et al., *Vaccine and antibody-directed T cell tumour immunotherapy*. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 2004. **1704**(1): p. 11-35.
24. Meloen, R.H., et al., *Synthetic peptide vaccines: unexpected fulfillment of discarded hope?* Biologicals, 2001. **29**(3-4): p. 233-236.
25. Doytchinova, I.A. and D.R. Flower, *VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines*. BMC bioinformatics, 2007. **8**(1): p. 4.
26. Larsen, M.V., et al., *Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction*. BMC bioinformatics, 2007. **8**(1): p. 424.
27. Moutafsi, M., et al., *A consensus epitope prediction approach identifies the breadth of murine T CD8+ -cell responses to vaccinia virus*. Nature biotechnology, 2006. **24**(7): p. 817-819.
28. Calis, J.J., et al., *Properties of MHC class I presented peptides that enhance immunogenicity*. PLoS computational biology, 2013. **9**(10).
29. Dimitrov, I., et al., *AllerTOP v. 2—a server for in silico prediction of allergens*. Journal of molecular modeling, 2014. **20**(6): p. 2278.
30. Gupta, S., et al., *In silico approach for predicting toxicity of peptides and proteins*. PloS one, 2013. **8**(9).
31. Nielsen, M. and O. Lund, *NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction*. BMC bioinformatics, 2009. **10**(1): p. 296.
32. Paul, S., et al., *TepiTool: a pipeline for computational prediction of T cell epitope candidates*. Current protocols in immunology, 2016. **114**(1): p. 18.19. 1-18.19. 24.
33. Luckheeram, R.V., et al., *CD4+ T cells: differentiation and functions*. Clinical and developmental immunology, 2012. **2012**.
34. Dhanda, S.K., P. Vir, and G.P. Raghava, *Designing of interferon-gamma inducing MHC class-II binders*. Biology direct, 2013. **8**(1): p. 30.
35. Nagpal, G., et al., *Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential*. Scientific reports, 2017. **7**: p. 42851.
36. EL-Manzalawy, Y., D. Dobbs, and V. Honavar, *Predicting linear B-cell epitopes using string kernels*. Journal of Molecular Recognition: An Interdisciplinary Journal, 2008. **21**(4): p. 243-255.
37. Saha, S. and G.P.S. Raghava, *Prediction of continuous B-cell epitopes in an antigen using recurrent neural network*. Proteins: Structure, Function, and Bioinformatics, 2006. **65**(1): p. 40-48.
38. Kringelum, J.V., et al., *Structural analysis of B-cell epitopes in antibody: protein complexes*. Molecular immunology, 2013. **53**(1-2): p. 24-34.
39. Adhikari, U.K. and M.M. Rahman, *Overlapping CD8+ and CD4+ T-cell epitopes identification for the progression of epitope-based peptide vaccine from nucleocapsid and glycoprotein of emerging Rift Valley fever virus using immunoinformatics approach*. Infection, Genetics and Evolution, 2017. **56**: p. 75-91.
40. Bui, H.-H., et al., *Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines*. BMC bioinformatics, 2007. **8**(1): p. 361.
41. Xu, X., C. Yan, and X. Zou, *MDockPeP: An ab-initio protein-peptide docking server*. Journal of computational chemistry, 2018. **39**(28): p. 2409-2413.
42. Gu, Y., et al., *Vaccination with a paramyosin-based multi-epitope vaccine elicits significant protective immunity against Trichinella spiralis infection in mice*. Frontiers in microbiology, 2017. **8**: p. 1475.
43. Nezafat, N., et al., *A novel multi-epitope peptide vaccine against cancer: an in silico approach*. Journal of theoretical biology, 2014. **349**: p. 121-134.
44. Sayed, S.B., et al., *Exploring Lassa Virus Proteome to Design a Multi-epitope Vaccine Through Immunoinformatics and Immune Simulation Analyses*. International Journal of Peptide Research and Therapeutics, 2020: p. 1-19.
45. Hasan, M., et al., *Reverse vaccinology approach to design a novel multi-epitope subunit vaccine against avian influenza A (H7N9) virus*. Microbial pathogenesis, 2019. **130**: p. 19-37.
46. Arai, R., et al., *Design of the linkers which effectively separate domains of a bifunctional fusion protein*. Protein engineering, 2001. **14**(8): p. 529-532.
47. Wilkins, M., et al., *RD and Hochstrasser, DF (1999) Protein Identification and Analysis Tools in the ExPASy Server. In 2-D proteome analysis protocols*. Humana Press. Totowa, NJ.
48. Magnan, C.N., A. Randall, and P. Baldi, *SOLpro: accurate sequence-based prediction of protein solubility*. Bioinformatics, 2009. **25**(17): p. 2200-2207.
49. Buchan, D.W., et al., *Scalable web services for the PSIPRED Protein Analysis Workbench*. Nucleic acids research, 2013. **41**(W1): p. W349-W357.
50. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis*. Nature protocols, 2015. **10**(6): p. 845.
51. Heo, L., H. Park, and C. Seok, *GalaxyRefine: protein structure refinement driven by side-chain repacking*. Nucleic acids research, 2013. **41**(W1): p. W384-W388.
52. Wiederstein, M. and M.J. Sippl, *ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins*. Nucleic acids research, 2007. **35**(suppl_2): p. W407-W410.
53. Colovos, C. and T.O. Yeates, *Verification of protein structures: patterns of nonbonded atomic interactions*. Protein science, 1993. **2**(9): p. 1511-1519.
54. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structures*. Journal of applied crystallography, 1993. **26**(2): p. 283-291.
55. Gane, P.J. and P.M. Dean, *Recent advances in structure-based rational drug design*. Current opinion in structural biology, 2000. **10**(4): p. 401-404.
56. Solanki, V. and V. Tiwari, *Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against Acinetobacter baumannii*. Scientific reports, 2018. **8**(1): p. 1-19.
57. Park, J.S., et al., *Mechanism of anchoring of OmpA protein to the cell wall peptidoglycan of the gram-negative bacterial outer membrane*. The FASEB Journal, 2012. **26**(1): p. 219-228.
58. Yan, Y., et al., *HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy*. Nucleic acids research, 2017. **45**(W1): p. W365-W373.
59. Van Aalten, D.M., et al., *A comparison of techniques for calculating protein essential dynamics*. Journal of Computational Chemistry, 1997. **18**(2): p. 169-181.
60. Wüthrich, K., et al., *Correlations between internal mobility and stability of globular proteins*. Biophysical journal, 1980. **32**(1): p. 549-560.
61. Rapin, N., et al., *Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system*. PLoS One, 2010. **5**(4).
62. Craig, D.B. and A.A. Dombkowski, *Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins*. BMC bioinformatics, 2013. **14**(1): p. 346.
63. Grote, A., et al., *JCat: a novel tool to adapt codon usage of a target gene to its potential expression host*. Nucleic acids research, 2005. **33**(suppl_2): p. W526-W531.
64. Sharp, P. and W. Li, *The codon adaptation index—a measure of directional synonymous codon usage bias, and its possible applications*. Nucleic Acids Res, 1987. **15**: p. 1281-1295.

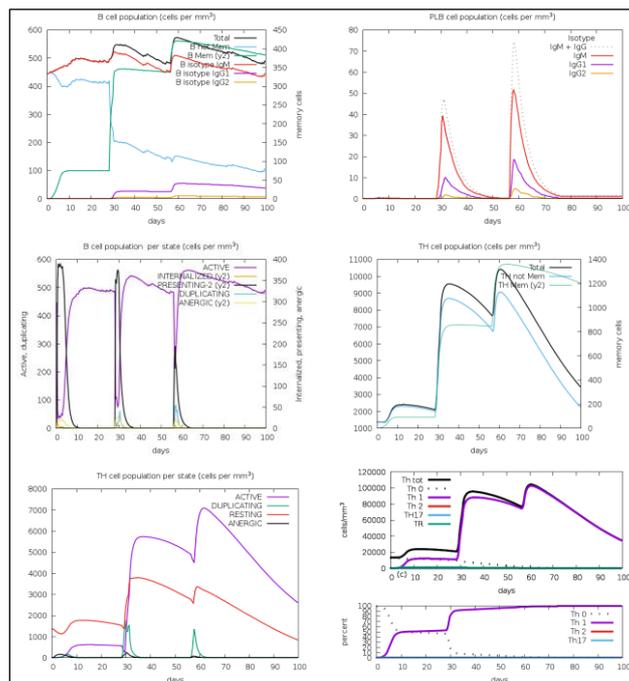
65. Kovacs, J.A., P. Chacón, and R. Abagyan, *Predictions of protein flexibility: first-order measures*. Proteins: Structure, Function, and Bioinformatics, 2004. **56**(4): p. 661-668.
66. Lopéz-Blanco, J.R., J.I. Garzón, and P. Chacón, *iMod: multipurpose normal mode analysis in internal coordinates*. Bioinformatics, 2011. **27**(20): p. 2843-2850.
67. Li, W., et al., *Peptide vaccine: progress and challenges*. Vaccines, 2014. **2**(3): p. 515-536.
68. Zhang, L., *Multi-epitope vaccines: a promising strategy against tumors and viral infections*. Cellular & molecular immunology, 2018. **15**(2): p. 182-184.
69. Adhikari, U.K., M. Tayebi, and M.M. Rahman, *Immunoinformatics approach for epitope-based peptide vaccine design and active site prediction against polyprotein of emerging oropouche virus*. Journal of immunology research, 2018. **2018**.
70. Ali, M., et al., *Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection*. Scientific reports, 2017. **7**(1): p. 1-13.
71. Bhattacharya, M., et al., *Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach*. Journal of medical virology, 2020. **92**(6): p. 618-631.
72. Alsaadi, E. and I. Jones, *Membrane binding proteins of coronaviruses*. Futur Med [Internet]. 2019 Apr 1 [cited 2020 Jun 10]; **14** (4): 275–86.
73. Schoeman, D. and B.C. Fielding, *Coronavirus envelope protein: current knowledge*. Virology journal, 2019. **16**(1): p. 1-22.
74. Al-Amri, S.S., et al., *Immunogenicity of candidate MERS-CoV DNA vaccines based on the spike protein*. Scientific reports, 2017. **7**: p. 44875.
75. Shi, S.-Q., et al., *The expression of membrane protein augments the specific responses induced by SARS-CoV nucleocapsid DNA immunization*. Molecular immunology, 2006. **43**(11): p. 1791-1798.
76. Zhao, P., et al., *Immune responses against SARS-coronavirus nucleocapsid protein induced by DNA vaccine*. Virology, 2005. **331**(1): p. 128-135.
77. Cooper, N.R. and G.R. Nemerow, *The role of antibody and complement in the control of viral infections*. Journal of investigative dermatology, 1984. **83**(s 1): p. 121-127.
78. Sesardic, D., *Synthetic peptide vaccines*. Journal of medical microbiology, 1993. **39**(4): p. 241-242.
79. Roy, A.S., et al., *Multi-epitope Based Peptide Vaccine Design Using Three Structural Proteins (S, E, and M) of SARS-CoV-2: An In Silico Approach*. bioRxiv, 2020.
80. Wu, C.-Y., et al., *Improving therapeutic HPV peptide-based vaccine potency by enhancing CD4+ T help and dendritic cell activation*. Journal of biomedical science, 2010. **17**(88): p. 1-10.



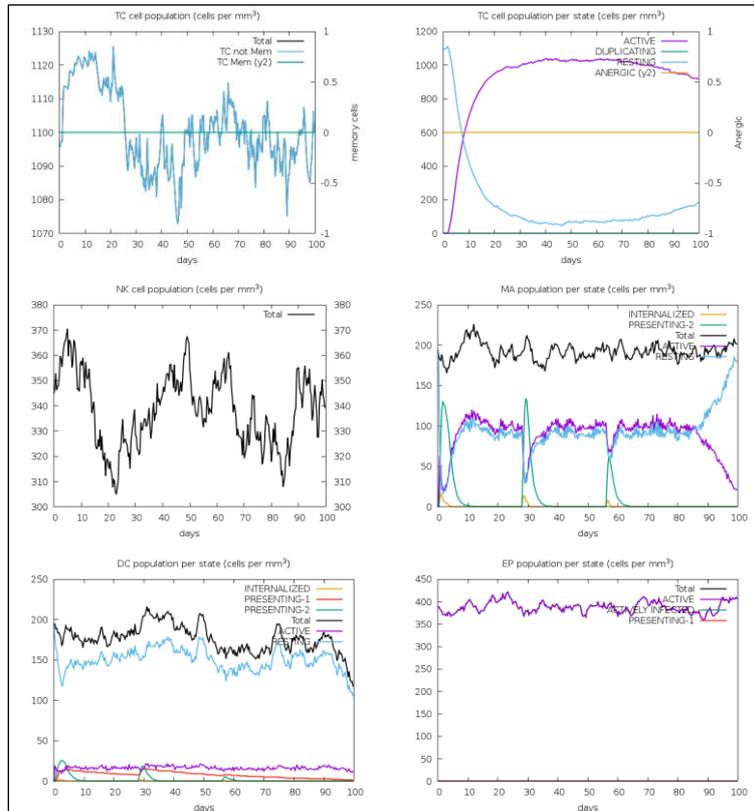
Supplementary figure 1: Vaccine structure showing the discontinuous B-cell epitopes as pale-yellow spheres.



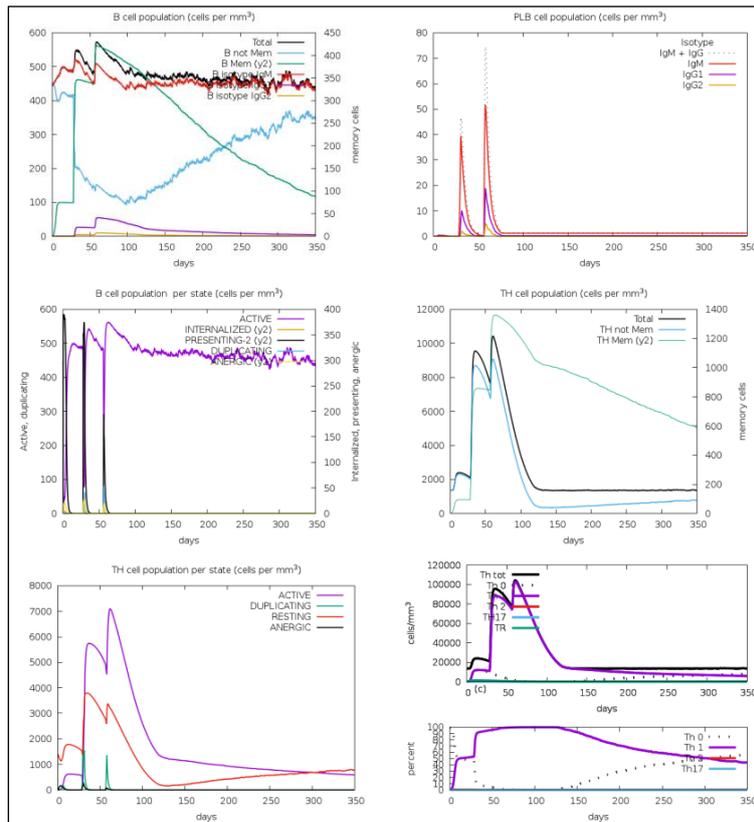
Supplementary figure 2: Receptor structure wherein pale-yellow spheres represent the active site residues.



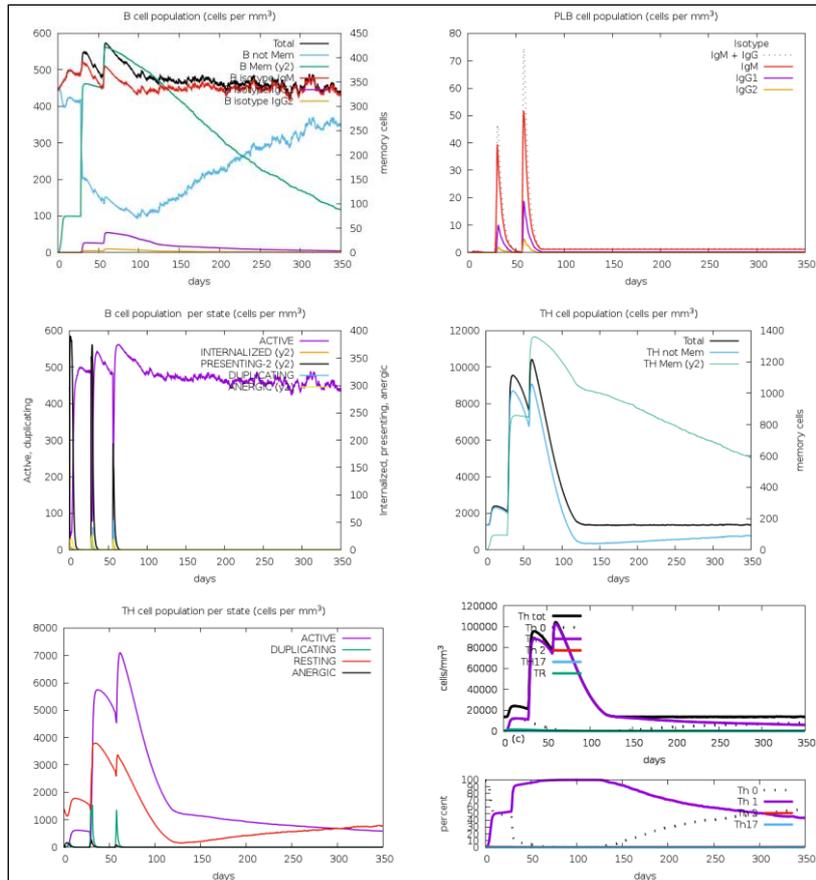
Supplementary figure 3: Cell counts upon vaccine administration up to 100 days, Mem is memory.



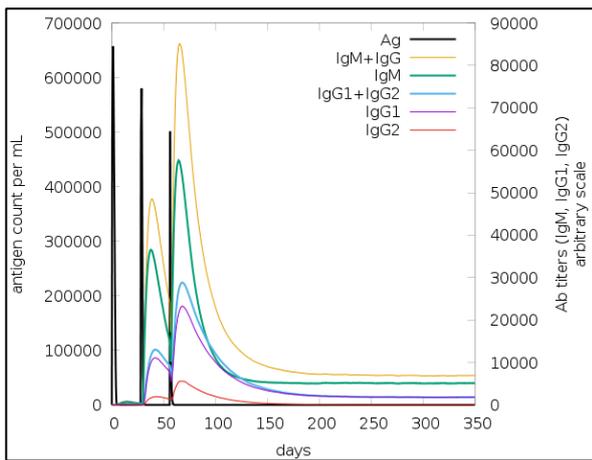
Supplementary figure 4: Cell counts upon vaccine administration up to 100 days, Mem is memory.



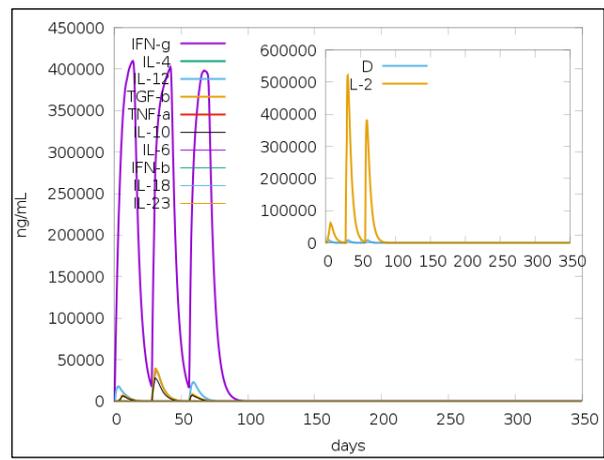
Supplementary figure 5: Cell counts upon vaccine administration up to 350 days, Mem is memory.



Supplementary figure 6: Cell counts upon vaccine administration up to 350 days, Mem is memory.



Supplementary figure 7: Antibodies and antibody-complexes produced in response to coronavirus antigen.



Supplementary figure 8: Interleukins and cytokines secreted in response to vaccine administration.

Supplementary Table 1: Interaction Hot-spots identified by Spot-on server (vaccine construct is designated as chain B)

Residue Index	Residue Name	Residue Chain
588	LYS	B
584	LEU	B
414	ALA	B
327	SER	B
309	ILE	B

Supplementary Table 2: Interacting hot-regions identified using Hot-region server (TLR-2 is designated as Chain A while vaccine construct is designated as Chain B)

Interface Name	Residue Number	Residue Type	Chain	Relative Complex ASA	Relative Monomer ASA	Pair Potential	Hotspot Status	Hot-region Status	Complex ASA	Monomer ASA
Final	31	ASP	A	26.09	54.88	8.9	NH		36.63	77.05
Final	32	ALA	A	83.16	108.33	6.35	NH		89.77	116.94
Final	33	SER	A	39.47	66.82	5.09	NH		45.98	77.85
Final	35	VAL	A	10.63	47.77	19.53	H	-	16.1	72.34
Final	39	ARG	A	13.32	46.75	12.99	NH		31.8	111.63
Final	40	SER	A	37.99	68.58	15.13	NH		44.26	79.89
Final	61	PHE	A	17.22	50.75	19.8	H	-	34.36	101.23
Final	137	GLN	A	41.35	66.72	13.81	NH		73.81	119.1
Final	160	GLU	A	40.13	81.14	4.9	NH		69.13	139.76
Final	161	THR	A	13.21	34.2	11.34	NH		18.4	47.63
Final	316	ARG	A	5.55	34.7	20.6	H	-	13.25	82.86
Final	318	HIS	A	26.92	53.64	5.78	NH		49.23	98.09
Final	321	GLN	A	29.7	65.44	22.46	NH		53.01	116.81
Final	322	PHE	A	64.03	71.92	0	NH		127.73	143.47
Final	323	TYR	A	32.89	81.92	18.52	NH		69.97	174.3
Final	324	LEU	A	64.66	89.43	11.68	NH		115.5	159.75
Final	325	PHE	A	47.26	87.81	6.11	NH		94.27	175.17
Final	329	SER	A	32.06	56.17	8.52	NH		37.35	65.44
Final	332	TYR	A	11.36	16.01	20.85	H	-	24.18	34.06
Final	333	SER	A	17.12	26.62	7	NH		19.95	31.01
Final	336	GLU	A	18.34	46.98	15.8	NH		31.59	80.93
Final	345	ASN	A	4.16	40.63	16.27	NH		5.99	58.48
Final	347	LYS	A	36.13	77.24	8.01	NH		72.55	155.1
Final	349	PHE	A	16.74	53.91	16.41	NH		33.4	107.54
Final	350	LEU	A	16.91	32.17	24.78	H	-	30.2	57.46
Final	357	GLN	A	9.74	34.45	25.66	H	-	17.38	61.5
Final	358	HIS	A	7.09	39.66	13.89	NH		12.97	72.53
Final	360	LYS	A	13.89	63.06	9.98	NH		27.9	126.63
Final	371	LEU	A	9.52	37.64	22.39	H	-	17.01	67.23
Final	375	GLU	A	64.35	79.98	7.72	NH		110.84	137.77
Final	376	TYR	A	6.54	42.91	25.96	H	-	13.92	91.29
Final	379	ASN	A	18.63	49.72	16.5	NH		26.81	71.57
Final	383	LYS	A	38.72	90.22	10.14	NH		77.75	181.17
Final	384	GLY	A	2.23	41.06	10.24	NH		1.79	32.89
Final	396	GLN	A	27.37	42	14.87	NH		48.86	74.97
Final	422	ARG	A	28.74	43.73	22.76	NH		68.61	104.41
Final	466	ASN	A	28.46	36.49	11.49	NH		40.96	52.53
Final	486	ARG	A	15.94	56.37	24.54	H	-	38.07	134.59
Final	508	SER	A	15.97	33.03	11.95	NH		18.6	38.48
Final	532	THR	A	32.15	59.34	12.5	NH		44.78	82.64
Final	561	LYS	A	30.45	49.76	5.66	NH		61.14	99.92
Final	89	LYS	B	16.85	24.42	17.02	NH		33.83	49.04
Final	90	VAL	B	6.39	24.27	26.22	H	-	9.67	36.75
Final	223	VAL	B	29.27	42.07	11.35	NH		44.33	63.71
Final	224	PRO	B	33.79	64.64	4.96	NH		46	88
Final	241	PHE	B	69.14	105.19	6.55	NH		137.92	209.83
Final	242	ASP	B	25.33	55.77	8.67	NH		35.56	78.3
Final	244	ALA	B	13.43	34.15	14.62	NH		14.5	36.87
Final	245	ASP	B	22.2	65.77	8.76	NH		31.17	92.34
Final	297	ASN	B	18.1	69.9	14.02	NH		26.05	100.61
Final	306	GLN	B	25.72	72.44	15.21	NH		45.91	129.31
Final	307	ASN	B	23.65	107.12	6.67	NH		34.04	154.19
Final	309	ILE	B	2.03	19.07	27.26	H	-	3.55	33.4
Final	310	SER	B	15.81	48.68	9.95	NH		18.42	56.71
Final	323	ASN	B	48.47	84.17	7.65	NH		69.77	121.16

Final	324	LYS	B	35.66	65.82	11.52	NH		71.61	132.17
Final	327	SER	B	6.04	62.05	11.6	NH		7.04	72.29
Final	328	VAL	B	85.13	104.99	10.09	NH		128.92	158.99
Final	359	LYS	B	49.19	64.54	9.31	NH		98.77	129.6
Final	360	PHE	B	27.82	45.45	16.91	NH		55.49	90.67
Final	361	VAL	B	39.92	67.24	16.26	NH		60.45	101.83
Final	363	ALA	B	13.14	75.73	13.66	NH		14.18	81.75
Final	410	LEU	B	22.39	50.56	18.04	NH		40	90.31
Final	411	LEU	B	18.59	76.34	30.36	H	+	33.21	136.37
Final	413	LEU	B	12.05	37.84	22.08	H	+	21.52	67.6
Final	414	ALA	B	1.27	58.95	13.41	NH		1.37	63.64
Final	415	PHE	B	14.56	31.82	26.74	H	+	29.05	63.47
Final	421	TYR	B	36.42	52.69	13.07	NH		77.48	112.1
Final	577	LYS	B	37.9	61.43	12.09	NH		76.11	123.35
Final	578	LEU	B	38.64	49.33	12.66	NH		69.03	88.11
Final	581	LEU	B	38.22	84.57	10.78	NH		68.27	151.06
Final	582	TRP	B	66.89	66.9	7.08	NH		166.79	166.82
Final	584	LEU	B	3.96	43.79	30.89	H	-	7.07	78.22
Final	585	TRP	B	49.66	75.24	6.96	NH		123.82	187.62
Final	588	LYS	B	6.82	63.81	13.34	NH		13.7	128.13
Final	604	ILE	B	31.68	70.33	17.07	NH		55.48	123.16
Final	609	ARG	B	21.62	52.68	18.01	NH		51.61	125.79
Final	654	LYS	B	46.34	67.09	9.22	NH		93.05	134.73
Final	660	TYR	B	54.48	92.4	14.43	NH		115.91	196.58

Supplementary Table 3: Hydrogen bonding between vaccine construct (designated chain B) and TLR-2 (designated chain A)

Acceptor				Donor				Hydrogen bonding
Residue 1	Position	Atom1	Chain	Residue 2	Position	Atom 2	Chain	Distance (Å)
GLN	137	NE2	A	TYR	660	OH	B	2.71
ASN	345	ND2	A	GLN	306	O	B	3.25
LYS	360	NZ	A	LEU	581	O	B	2.87
ASN	379	ND2	A	TYR	421	OH	B	3.17
ARG	486	NH2	A	ASP	245	OD1	B	1.53
SER	508	OG	A	ASP	242	OD1	B	2.47
SER	327	OG	B	ASP	31	OD2	A	2.53
LYS	89	NZ	B	TYR	323	OH	A	3.21
LYS	588	NZ	B	SER	329	O	A	2.79
TRP	585	NE1	B	GLU	336	OE2	A	3.10

Supplementary Table 4: Interaction data in the Vaccine-TLR-2 complex (TLR-2 is designated as MOL1 and Vaccine is designated as MOL2)

Title	Value
Buried area upon the complex formation (Å ²)	4847.7
Buried area upon the complex formation (%)	7.82
Interface area (Å ²)	2423.85
Interface area MOL1 (%)	9.51
Interface area MOL2 (%)	6.64
POLAR Buried area upon the complex formation (Å ²)	2857.3
POLAR Interface (%)	58.94
POLAR Interface area (Å ²)	1428.65
NON-POLAR Buried area upon the complex formation (Å ²)	1990.3
NON-POLAR Interface (%)	41.06
NON-POLAR Interface area (Å ²)	995.15
Residues at the interface TOTAL (n)	142
Residues at the interface_MOL1	75
Residues at the interface_MOL2	67