Multi-epitope-based peptide vaccine design against SARS-CoV-2 using its spike protein

Debarghya Mitra[1#], Janmejay Pandey[2], Shiv Swaroop[1,*]

1: Department of Biochemistry, School of Life Sciences, Central University of Rajasthan, Ajmer, Rajasthan, India. 305817

2: Department of Biotechnology, School of Life Sciences, Central University of Rajasthan, Ajmer, Rajasthan, India. 305817

*** Corresponding Authors:**

Shiv Swaroop

Department of Biochemistry, School of Life Sciences

Central University of Rajasthan

Bandarsindri, NH-8, Kishangarh, Ajmer (305817)

Rajasthan - INDIA

E-mail: shivswaroop@curaj.ac.in

Phone: +91- 1463-238733 (Off.); +91-9972230316 (Mobile)

Current Address

Department of Biosciences and Bioengineering, Indian Institute of Bombay, Powai, Mumbai, Maharashtra, India. 400076

Abstract:

SARS CoV-2 has particularly been efficient in ensuring that many countries are brought to a standstill. With repercussions ranging from rampant mortality, fear, paranoia and economic recession, the virus has brought together countries in order to look at possible therapeutic countermeasures. With prophylactic interventions possibly months away from being particularly effective, a slew of measures and possibilities concerning the design of vaccines are being worked upon. We attempted a structure-based approach utilizing a combination of epitope prediction servers to develop a multi-epitope-based subunit vaccine that involves the two major domains of the spike glycoprotein of SARS CoV-2 (S1 and S2) coupled with a substantially effective chimeric adjuvant to create stable vaccine constructs through MD simulations. The designed constructs were evaluated based on their docking with Toll Like Receptor (TLR) 4. Our findings provide an epitope-based peptide fragment; which can be a potential candidate for the development of a vaccine against SARS-CoV-2. Recent experimental studies based on determining immunodominant regions across the spike glycoprotein of SARS-CoV-2 indicate the presence of the predicted epitopes included in this study.

Keywords: SARS-CoV-2, vaccine design, Spike protein, Docking, MD simulations, Adjuvant

1. Introduction:

In December 2019, the town of Wuhan in Hubei, China, experienced an outbreak of a highly pathogenic virus prevalent with high transmissibility across humans and potentially responsible for varied symptoms associated with respiratory-based complications. Months after the outbreak, the disease has spread across the globe, infecting 8.7 million people and responsible for the deaths of another 4.6 lacs (at submission). It has all been implicated in a pathogen belonging to the genus *Betacoronavirus* of the Coronaviridae family and has been identified as the severe acute respiratory syndrome coronavirus-2 (SARS CoV 2). The WHO declared it a pandemic on 12th March, 2020 allowing countries across the world to take appropriate measures to fend themselves from the disease and prevent its spread.

The disease transmission has left very few countries free from the repercussions of the disease with surmounting pressure on hospital settings and resources to contain this public health emergency. The scientific community is on an overdrive to look into prophylactic measures as the first line of defense against the disease. The development of rapid diagnostics to determine and isolate disease affected individuals along with repurposing and identifying existing FDA-approved drugs to provide substantial treatment are underway. Still, going by the looks of it and the unprecedented pace of development of therapeutics, it would require several months to cross over several clinical trials and finally be made available to the population.

The study of SARS- CoV-2 has been easier due to the manifestation of a similar homolog of the disease SARS-CoV more than a decade ago and comprises of similar structural counterparts that aid in the process of infection, including a spike glycoprotein, membrane glycoprotein, an envelope protein and a nucleocapsid protein[1, 2]. The genome sequence and its analysis were done promptly and made available [3, 4].Several structures over SARS-Cov-2 S protein in

interaction with human ACE-2 (hACE-2) have been made publicly available over the RCSB PDB[5-8], allowing the community to proceed towards the development of probable therapeutic measures ranging from repurposing drugs[9], design of small molecule inhibitors, identification of unique targets and elucidating molecular mechanisms of the viral protein machinery[10] and how it establishes itself inside the host body. The interaction between the spike glycoprotein and the human ACE-2 receptor is important as it initiates the process of viral entry into human beings through contact with an infected individual. The identification of a probable site upstream of the receptor binding motif (which includes the residues on S1 that interact with hACE-2), which undergoes cleavage and requires priming by TMPRSS2 before stable fusion of the viral membrane complex to the hACE2[11] and involves sufficient nonreversible conformational changes that put into motion the process of viral entry[1, 5].

Vaccines have been pegged as an effective measure against this positive, single-stranded RNA virus and several are already underway since they can be regarded as the only suitable measure in immunizing individuals the world over and stopping the disease from steamrolling through countries. This paper investigates the utilization of the spike glycoprotein of SARS CoV-2 as a probable immunogen for the design of a multi-epitope peptide-based subunit vaccine with a chimeric adjuvant in tow. The spike glycoprotein has been evidenced to be an appropriate immunogen capable of eliciting neutralizing antibodies, which will inadvertently lead to establishment and elicit an immune response to prevent viral entry and act as a prophylactic [6].

The work envisages the prediction of significant epitopes and design of a multi-epitope-based peptide subunit vaccine coupled with a chimeric adjuvant. The immunogen considered involves the two domains of the spike glycoprotein involved in binding with hACE-2 followed by viral and cellular membrane fusion through S1 and S2, respectively. The cleavage site present

between S1 and S2 post-fusion has not been considered due to previous apprehensions on the hypersensitivity of the immune system considering the entire spike glycoprotein of SARS- CoV. The utilization of these epitopes and the vaccine construct across an experimental setting will help provide essential evaluation and aid in the development of a vaccine to generate a robust immunological prophylactic response against SARS-CoV-2.

2. Methods

2.1 Sequence Retrieval for SARS CoV 2 Spike Glycoprotein:

The SARS-CoV-2 spike glycoprotein sequence (PDB ID: 6VSB)[6] was checked for specific domains and the identified S1 domain (residues) comprising the Receptor Binding Domain(RBD) along with the S2 domain(residues) were edited from the cryo-EM structure of the surface glycoprotein using UCSF Chimera[12].

2.2 Prediction of B-cell epitopes

Based on the sequences of S1 and S2 domains, they were separately used to determine linear B cell epitopes through different servers. The Artificial Neural Network (ANN) approach-based server ABCPred[13], Bcepred[14] and the Immune Epitope Database and Analysis Resource (IEDB) based linear B cell epitope prediction tools[15-20] were utilized to predict probable epitope sequences across the query sequences. The latter two servers utilize physicochemical parameters like hydrophilicity, polarity, surface accessibility and flexibility to predict a B cell epitope as has been previously evidenced. The predicted linear B cell sequences from ABCPred were matched with the probable epitopes predicted over the different parameters inclusive of surface accessibility area, hydrophilicity, polarity and flexibility in that order over the IEDB-based Bepipred 2.0 based linear prediction of epitopes and the results from the two servers were

corroborated with these parameters. These same parameters were then again checked with the predicted epitopes by Bcepred, which predicts epitopes on the basis of the same parameters. Consensus sequences which emerged from all three were considered to be the probable epitopes in this regard. Since conformational or discontinuous B cell epitopes could be predicted through ElliPro [21] and Discotope[22], they were utilized for analysis of the consensus sequences arrived upon earlier. Consensus 16-mer epitopes were predicted using these servers.

2.3 Prediction of MHC II based Helper T Lymphocytes ($CD4^+$)

The probable MHC II binding epitopes on S1 and S2 domains were predicted and again a consensus approach was utilized to arrive at the conclusive sequences. The prediction of helper T lymphocytes was made using MHCPred[23], SYFPEITHI[24], NetMHCIIpan 3.2 server [25] and the IEDB server [26, 27]. A consensus selection of epitopes from the four different servers allowed us to improve upon and circumvent the limitations associated with prediction of MHC II binders due to their polymorphic nature, peptide length variation and also determination of the appropriate peptide binding core. Hence, looking into the limitations these servers were determined to be among the best available that can be utilized in this avenue. Each of these prediction servers was compared with experimental datasets to assess their performance. SYFPEITHI predicts nonamer sequences based on the weighted contribution of each amino acid sequence present across a predicted epitope sequence. MHCPred allows prediction of 9-mer epitopes based on multivariate statistical methods and NetMHCIIpan 3.2 allows the determination of 15-mer epitope fragments but with limited allele-specific choices. The IEDB-based MHC II binding epitope sequence prediction server was also utilized because it employs a combination of methods ranging from ANN to SMM and ranks probable epitopes on the basis of a percentile score and an IC_{50} score. The sequences predicted over the different servers were

either 15 mers or 9 mers based on the server in use, but based on the consensus selection across these platforms and overlapping regions of the predicted epitope sequences, uniform 15-mer epitopes were selected.

2.4 Prediction of MHC I based Cytotoxic T Lymphocytes ($CD8^+$)

The prediction of Cytotoxic T Lymphocyte cells (T_{CTL}) through MHC I binding servers involved the utilization of NetMHC 4.0 [28], MHC-NP[29], NetCTL 1.2[30] and the IEDB-based T cell epitope tools[31]. All these servers predict a nonamer epitope sequence and with the use of a default dataset along with probable interacting human leukocyte antigen alleles with SARS CoV identified from literature. Employing a consensus selection of the predictions from the four servers, we were able to list the appropriate T_{CTL} epitopes with relative confidence. The NetMHC 4.0 utilizes an Artificial Neural Network to predict epitopes, NetCTL 1.2 server allows for the identification of epitopes based on improved datasets with a sensitivity of 0.80 and specificity of 0.97 across the filtering threshold employed, MHC-NP employs a Machine Learning approach towards prediction of naturally processed epitopes by the MHC, whereas IEDB-based MHC I based prediction of T cell epitopes sorts epitopes based on the percentile score and low IC_{50} values across a combination of ANN and SMM approaches based on an appropriate peptide library. The MHC I alleles specific for a SARS CoV 2 manifestation were based on the alleles confirmed during the outbreak of SARS CoV at the beginning of the millennium.

2.5 Validation of predicted epitopes:

The consensus sequences thus arrived upon can suitably be considered as being capable of eliciting the necessary immune response. Additionally, each of the sequences were filtered based on their predicted antigenicity over Vaxijen[32], allergenicity over AllerTop[33] and

Algpred[34], toxicity over ToxinPred[35], capable of eliciting Interferon-gamma over IFN-G[36] servers. Also, the epitope sequences were matched with human proteins to prevent cases of antibody response against a self-antigen with the Multiple Peptide Match Tool. Each of the epitope sequences were matched with their corresponding secondary structure over the cryoEM structure (PDB ID: 6VSB)[6] and listed to be either alpha helices, coils, strands or beta sheets. This was done to allow for sequence arrangement to allow for improved molecular modeling. Another thing that we had not considered initially based on the conservation of the surface glycoprotein across the different strains deposited over the GISAID repository (https://www.gisaid.org/about-us/mission/) and the opinion that not much variability has been observed across these strains led us to check the coverage of the epitope sequences we have predicted with the sequences from the repository of surface glycoprotein and also the genome sequence deposited over NCBI from India [37].

2.6 Design of Adjuvant

In a peptide-based subunit vaccine, the predicted B and T cell epitopes are not sufficient to elicit a strong immune response to generate the necessary prophylactic measures. Therefore, a suitable adjuvant must also be added to the vaccine design[38]. In most cases, such purification can be carried out separately as in the purification of the adjuvant and the peptide epitope sequence separately. Recently, several approaches have been made to utilize a chimeric adjuvant comprising two or more separately identified adjuvants [39, 40]. A similar approach was carried out and an in silico design comprising of three different adjuvants with evidence as being used as adjuvants in different vaccines or as agonists of TLR-4 receptors were utilized. A recent review on suitable adjuvants for different TLRs using this approach has been published[41]. It was backed up by experimental evidence of the identified adjuvants being capable of eliciting an

immune response on interaction with toll-like receptors (TLRs) separately and downstream immune signaling generating the necessary prophylactic measures that can be expected from a vaccine against SARS CoV-2[42]. The generation of neutralization antibodies by the peptide sequences of the surface glycoprotein has already been evidenced in previous instances and hence their linking with a triagonist chimeric adjuvant supports their utilization as a suitable prophylactic measure. A suitable rigid linker was utilized to join them with the peptide sequence and their position was rearranged at the N-terminal to design several constructs of the vaccine. A single construct with adjuvants at the N and C-terminal of the vaccine was also considered.

2.7 Design and characterization of Vaccine Constructs:

Appropriate linkers were utilized to join intra B cell epitopes, T-Helper and T-Cytotoxic epitopes and also between them in the vaccine construct[43, 44]. The construct design comprises of the B cell epitopes linked to the adjuvant through a rigid EAAAK linker (helix forming) toward the N-terminal, followed by a GPGPG linker linking the T_{HTL} epitopes and an AAY linker for the T_{CTL} epitopes toward the C-terminal. A DP linker was utilized to link the three adjuvants at the N terminal. The three adjuvants were rearranged each time and the vaccine constructs were analyzed for physicochemical properties using ProtPARAM tool[45], allergenicity[33], antigenicity[32] and scanned for probable stimulants of interferon gamma[36]. The ToxinPred database[35] was used to analyze each of the units of the vaccine construct and hence each of these constructs can be considered nontoxic.

2.8 Molecular Modeling of Vaccine Constructs:

The main challenge associated with the molecular modeling of the vaccine constructs was that the adjuvant-specific region and the epitope-specific region matched with two different

templates. Although in both cases we had sufficient sequence coverage that calls for homology modeling of the vaccine constructs, the multi-template alignment led to a modeled structure comprising mainly of strands, which would have made the modeled constructs unstable. An initial approach included modeling the two parts of the vaccine construct separately using a single template, but linking both of these through loops and energy minimizations of these constructs fell through. The ROBETTA server[46], which allows for comparative modeling, brought the most promising model of the vaccine constructs with sufficient secondary structure coverage. Each of these constructs was then deposited over GALAXY webserver[47] for refinement. Based on select parameters based on which each modeled construct was refined, the best structures were then carried forward for validation through Ramachandran Plot, Z-score over PROCHECK[48] and ProSA-web server[49]. An additional validation was carried out using ERRAT-3D server[50]. Each of the modeled vaccine constructs was also checked for sequence mismatches.

2.9 Stability of the Modeled Vaccine Constructs:

Each of these constructs was then assessed for their stability using the ProtParam server. The various markers for stability made available through the Instability index (II) [51], PEST hypothesis[52], *in vivo* half-life and the isoelectric point[53]. The assessments based on Ramachandran Plot, ERRAT 3D[50], WHATCHECK[54] and the ProSA-web server[49] allowed us to select a single vaccine construct for a prolonged molecular dynamic simulation to determine the stability of the construct and probable use in *in vivo* settings.

2.10 Molecular dynamics simulations

MD simulations to obtain elaborate insights on the dynamic stability of the construct in physiological conditions were performed on Desmond from D E Shaw Research (DESRES)[55]. The protein structures were subjected to minimization in a 10 X 10 X 20 $Å^3$ periodic boundary box with the TIP3P water model using a fast scalable algorithm utilizing a distributed memory parallel program. Counter-ions and salt were added to neutralize the system. Pressure and temperature were maintained using a Nose-Hoover thermostat and a Martyna-Tobias-Klein barostat at 310K and 1 bar, respectively. All the systems were run for 100ns in the NPT ensemble. Trajectories were analyzed with a simulation interaction diagram in Desmond[55] and VMD[56].

2.11 Molecular Docking of Vaccine Construct with TLR-4:

The rationale behind selecting TLR-4 (PDB ID: 3FXI_A)[57] is the fact that in the case of SARS CoV, HIV, Influenzae and other RNA-based viruses, this Toll-like receptor has been experimentally evidenced to be implicated[58] . The binding sites over TLR-4 were determined through available PDB IDs that showed binding of an adjuvant lipopolysaccharide with the Toll like receptor and also through ElliPro[21] and Castp[59]. It was carried out through Patchdock[60] and Firedock[61].

2.12 MM/GBSA based evaluation of Docked Pose:

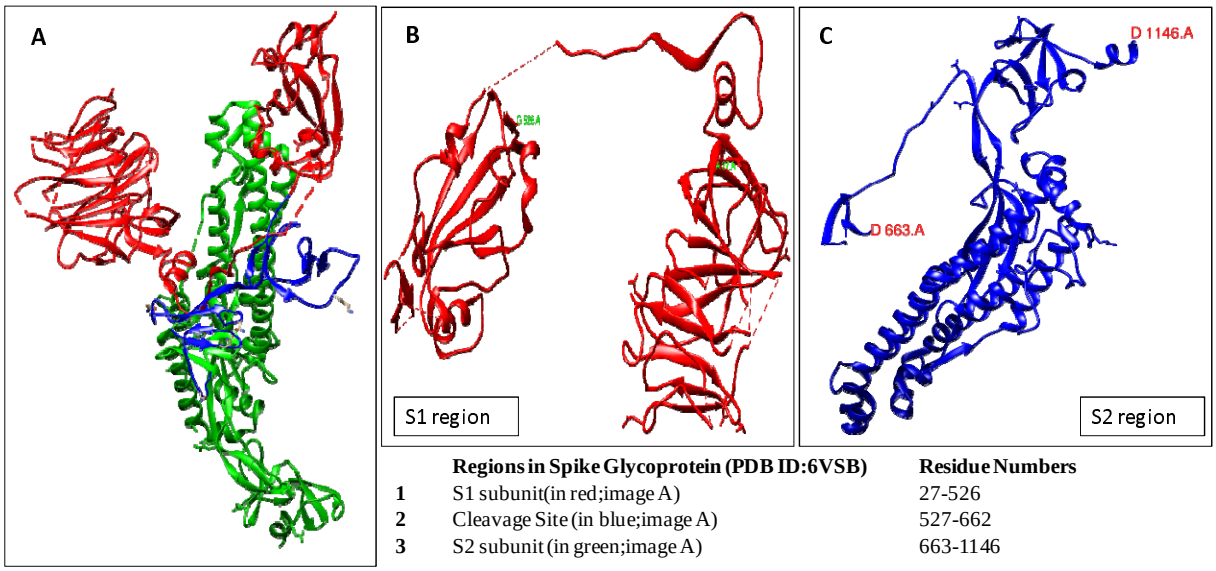
Post docking, the top pose was evaluated through Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) calculations over the HawkDock Server[62]. The server allows the docked pose to be assessed on a per-residue basis across Van der Waal potentials, electrostatic potentials, polar solvation free-energy and solvation free energy through empirical models. The

docked pose is minimized for 5000 steps, including 2000 cycles of steepest descents and 3000 cycles of conjugate gradient minimizations based on an implicit solvent model, the ff02 force field.

3. Results:

3.1 Sequence Retrieval for SARS CoV 2 Spike Glycoprotein:

The sequence of the SARS CoV 2 surface glycoprotein was retrieved from PDB ID: 6VSB in FASTA format. The UCSF Chimera visualization software[12] was utilized to edit the structure and include the S1 and the S2 domains as separate entities. The use of S1 and S2 domains separately without including the cleavage site, which falls in between the two domains **(Figure 1)** for the development of an immunogen is due to previously evidenced cases of increased immune hypersensitivity with concerns over utilizing the entire length of spike glycoprotein as a vaccine immunogen. [63]



	Regions in Spike Glycoprotein (PDB ID:6VSB)	Residue Numbers
1	S1 subunit(in red;image A)	27-526
2	Cleavage Site (in blue;image A)	527-662
3	S2 subunit (in green;image A)	663-1146

Figure 1: The S1 and S2 domains considered in the investigation are reproduced from PDB ID: 6VSB.

3.2 Prediction of B cell, T_{HTL} and T_{CTL} Epitopes:

The selected epitopes were based on multiple servers and since this rationale is not heuristic but based on the role of the different algorithms that lead each server to arrive at the predicted epitope.Based on the stringent filters applied in each case, the epitopes predicted over the different servers were manually curated to arrive upon the consensus sequences in each case. All the epitope sequences specific for S1 and S2, separately predicted by the servers, are listed. In the case of B cell and MHC II epitopes 15 mers epitope sequences were generated, while in the case of MHC I epitope sequences, nonamer fragments were generated.

The predicted sequences were matched with the data generated by [64] through IEDB on significant epitopes present on the surface glycoprotein, which did mention the need for analysis of the predicted sites and hence the utility of the stringent filters over the different servers followed by validation. The utilization of specific alleles which have been identified in the previous manifestation of the SARS CoV has helped us sort the predicted T cell epitopes appropriately on the different servers, which otherwise would have made prediction biased [65-68]. The selected epitopes are listed in **Table 1**.

Table 1(A): Predicted and Selected B cell Epitopes from S1 and S2

B cell Epitopes							
Sequence	Antigen	Allergen	IFN-G	Toxin	Mol Wt	pI	Length
YVGYLQPRTFLLKYNE	positive	Negative	positive	negative	2004.32	8.43	16
EVRQIAPGQTGKIADY	positive	Negative	positive	negative	1745.95	6.17	16
TGKIADYNYKLPDDFT	positive	Negative	positive	negative	1861.04	4.43	16
YFKIYSKHTPINLVRD	positive	Negative	positive	negative	1994.32	9.53	16
TVEKGIYQTSNFRVQP	positive	Negative	positive	negative	1867.09	8.26	16
HVTYVPAQEKNFTTAP	positive	Negative	positive	negative	1803	6.75	16

Table 1(B): Predicted and Selected T_{HTL} cell Epitopes from S1 and S2

T HTL(MHCII, CD4)							
Sequence	Antigen	Allergen	IFN-G	Toxin	Mol Wt	pI	Length
QSLLIVNNATNVVIK	positive	Negative	positive	negative	1625.93	8.75	15
AYYVGYLQPRTFLLK	positive	Negative	positive	negative	1832.18	9.53	15
REGVFVSNGTHWFVT	positive	Negative	positive	negative	1735.92	6.75	15
AIPTNFTISVTTEIL	positive	Negative	positive	negative	1619.88	4	15
GVVFLHVTYVPAQEK	positive	Negative	positive	negative	1686.97	6.75	15

Table 1(C): Predicted and Selected T_{CTL} cell Epitopes from S1 and S2

T CTL (MHC I, CD8)							
Sequence	Antigen	Allergen	IFN-G	Toxin	Mol Wt	pI	Length
VLPFNDGVY	positive	Negative	positive	negative	1023.15	3.8	9
LSETKCTLK	positive	Negative	positive	negative	1022.22	8.2	9
STQDLFLPF	positive	Negative	positive	negative	1067.21	3.8	9
YYVGYLQPRTF	positive	Negative	positive	negative	1406.6	8.5	11
FNATRFASV	positive	Negative	positive	negative	1012.13	9.75	9
SIAIPTNFTISVTTEILP	positive	Negative	positive	negative	1917.23	4	18
STECSNLLL	positive	Negative	positive	negative	979.11	4	9
LPPLLTDEM	positive	Negative	positive	negative	1028.23	3.67	9
AEIRASANL	positive	Negative	positive	negative	944.06	6.05	9

3.3 Validation and Filtering of selected epitopes:

The various parameters against which the selected epitopes and adjuvants were filtered against included antigenicity, allergenicity, toxicity, capability of generating interferon gamma response (**Tables 1A,B and C**) and chances of being identified as self-antigen by the human immune system. Among the epitope sequences hence filtered, 100 percent sequence coverage was detected in all but single residue mutations in 4 of them. These residues were not considered in the vaccine constructs and the aberrations were generally isolated to the S1 domain only (Figure 2B). Hence, sequence coverage based validation of the selected epitopes was also carried out. The list of epitopes considered for design of the vaccines did not include any sequences that did not conform to the 100% sequence coverage. The percent identity matrix utilized to indicate that

the Indian sequence shows percent identity of more than 90% with the sequence retrieved from PDB ID:6VSB and the epitopes utilized show 100% sequence coverage across the sequence submitted from India of the spike glycoproteins[37] **(Figure 2)**.

A

Percent Identity Matrix - created by Clustal2.1

1: INDIA_1	100.00	99.92	99.92	99.92	99.92	99.92	99.92	99.84
2: USA_1	99.92	100.00	100.00	100.00	100.00	100.00	100.00	99.92
3: USA_2	99.92	100.00	100.00	100.00	100.00	100.00	100.00	99.92
4: China_1	99.92	100.00	100.00	100.00	100.00	100.00	100.00	99.92
5: China_2	99.92	100.00	100.00	100.00	100.00	100.00	100.00	99.92
6: China_3	99.92	100.00	100.00	100.00	100.00	100.00	100.00	99.92
7: NEPAL_1	99.92	100.00	100.00	100.00	100.00	100.00	100.00	99.92
8: SWEDEN_1	99.84	99.92	99.92	99.92	99.92	99.92	99.92	100.00

B

28: IND_QIA98583.1	100.00	100.00	100.00	100.00	80.00	87.50	88.89	88.89		
100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		

Figure 2: Percent Identity Matrix (PIM) of SARS CoV-2 spike glycoprotein from GISAID (A) and (B) represents that apart from four of the epitopes selected, all have 100% sequence coverage across the submitted Indian sequence of SARS CoV-2. The overall sequence similarity of the submitted Indian sequence across the 7 other sequences from GISAID indicates the overall coverage sequence of the epitopes selected for vaccine design falls within acceptable parameters.

3.4 Design of Adjuvant:

The triagonist chimeric adjuvant has the propensity to elicit immune responses to allow for its use as a suitable adjuvant against SARS CoV 2. RS09[69], TR-433[70] and a Mtb based Hsp70 partial sequence[71, 72] were utilized in this endeavor. An appropriate rigid polar linker (DP) was utilized in this regard. Among the designed constructs all of the rearrangements were validated based on their antigenicity score predicted by Vaxijen. The adjuvants considered were all verified as TLR-4 agonists. The use of fragments of *Mycobacterial* hsp70 towards the generation of cytokines and natural killer cells has been verified and also for antigen-specific CTL responses [73-75]. The initial constructs and the parameters across which they were

assessed are mentioned in **Supplementary Table 1A and 1B**. The use of lipopolysaccharide based adjuvants has been abrogated in this study due to constraints of utilizing modeling, docking and simulation-based studies from their perspective.

3.5 Design and Characterization of Vaccine Constructs:

The arrangement and validation of antigenicity of the vaccine constructs based on the linkers and the adjuvants led to the terminal selection of 5 different vaccine constructs, which were then subsequently characterized based on the ProtParam parameters, which determined the molecular weight and isoelectric point and predicted each to be a stable construct. Also, the constructs were found to be suitably antigenic, nontoxic and non-allergens. Regions of the vaccine constructs were predicted to have sufficient B cell epitopes and capable of generating interferon gamma response and have 100% sequence coverage **(Table 2 and Figure 3)**.

Table 2: Validation of the five major vaccine constructs designed

Construct	Antigen	Allergen	IFN-G	Toxin	Length	Mol Wt	pI
C1	positive	negative	positive	negative	460	50088.79	9.29
C2	positive	negative	positive	negative	447	49165.7	9.25
C3	positive	negative	positive	negative	468	51662.61	9.24
C4	positive	negative	positive	negative	469	51775.77	9.24
C5	positive	negative	positive	negative	474	52246.29	9.23

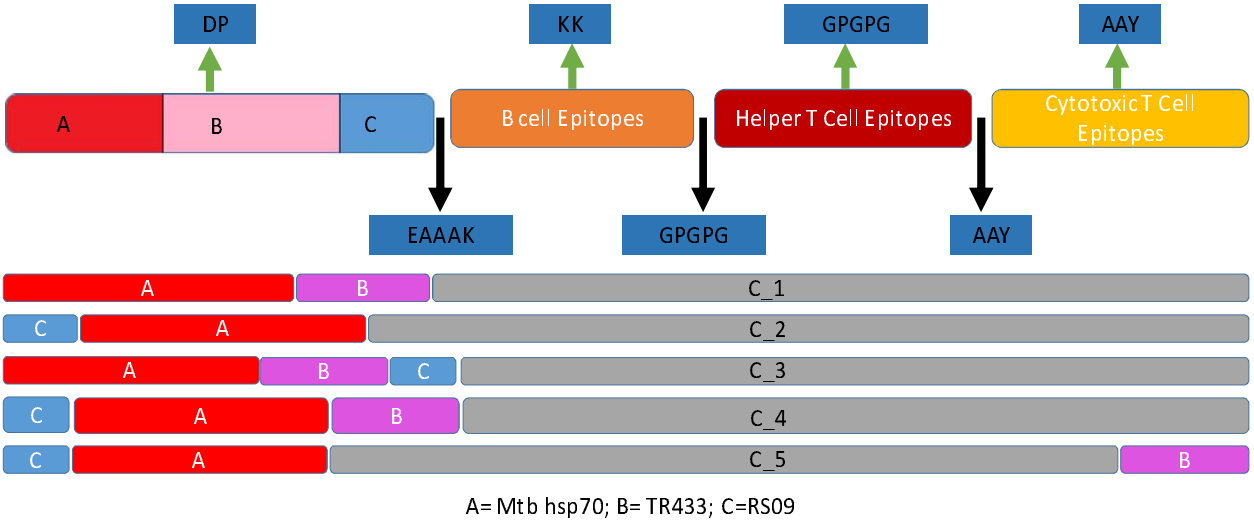


Figure 3: Designed vaccine constructs from C_1 to C_5 with linkers and adjuvants. Vaccine construct arrangement A,B and C refer to the three different adjuvants used. Designed Constructs based on rearrangement and numbered appropriately.

3.6 Molecular Modeling and Validation of Vaccine Constructs:

The vaccine constructs thus designed were made through comparative modeling over ROBETTA and revealed a structure better than the other platforms utilized in this regard **Figure 4**.

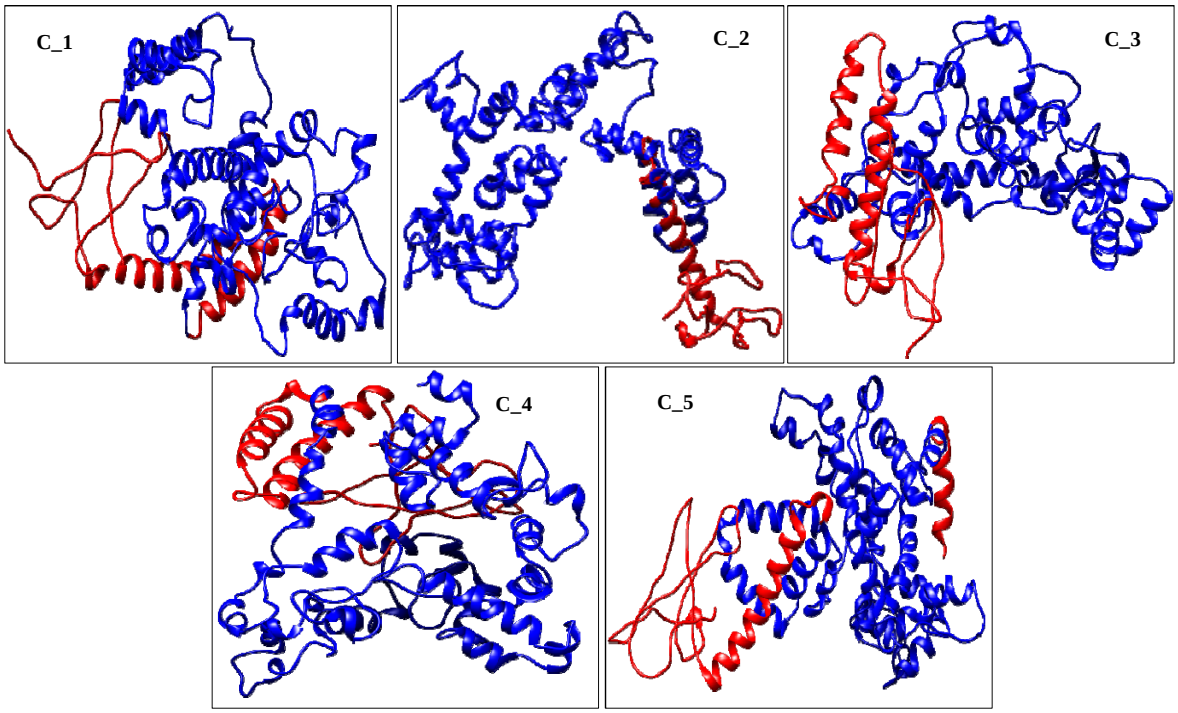


Figure 4: CM based modeling of designed vaccine constructs over ROBETTA server. The blue region is the epitope sequences with the red indicating the adjuvants used and their positions.

Validation and refinement of the modeled structures revealed a stable construct in each case which can be prominently detected by the predominant presence of defined secondary structures across the span of the construct interspersed by coils and helices. The modeled constructs were refined through GALAXY server **(Table 3)** and validated through Ramachandran Plots and Z-score from ProSA-webserver **(Table 4C)**.

Table 3: Refinement based on GALAXY SERVER of modeled constructs

Construct	Rama	Poor Rotamer	RMSD	Mol Probity	RMSD
C_5	96.6	0.5	0.356	1.779	0.36
C_3	95.3	0.5	0.356	1.967	0.37
C_4	96.6	0.5	0.367	1.774	0.36
C_2	96.6	0	0.972	1.83	0.35
C_1	96.1	0.3	0.351	1.792	0.35

3.7 Stability of the Modeled Vaccine Constructs:

Attempts were made to assess the stability of the designed constructs before moving on to docking and simulation studies with TLRs. Since it comprises three different adjuvants, an ensemble of different sequences interspersed across the spike glycoprotein, it becomes imperative that the constructs be evaluated. In this endeavor, we utilized the Expasy ProtParam Tool and based our assumptions on the PEST hypothesis and the Instability Index determinants, which include analysis of the residues that make any biomolecule have lower half-life or increased instability (includes Proline(P), Glutamic acid (E), Serine (S), Threonine (T), Methionine (M) and Glutamine (Q)) in the designed constructs and including Guruprasad's basis[51] that the relative abundance of Asparagine (N), Lysine (K) and Glycine (G) has been found in stable proteins (**Table 4B**). Moreover, basic or neutral isoelectric points also indicate a stable *in vivo* half-life for the modeled constructs. Since we cannot assess it entirely on these assumptions, we investigated the stability of the constructs through a 100 nanoseconds Molecular Dynamics Simulation. But due to constraints of computational time and effort we assessed the stability of Construct_4 since it gave the best estimate of stability based on our assumptions as listed in **Table 4(A,B,C)**. The abovementioned parameters went into this assumption coupled with the Ramachandran Plot, ERRAT-3D and the Z-score.

Table 4(A): The ProtParam based physicochemical based validation of vaccine constructs

Construct	Length	Mol Wt	II	Half life	pI
C1	460	50088.79	27.64	1.4	9.29
C2	447	49165.7	26.89	4.4	9.25
C3	468	51662.61	29.49	1.4	9.24
C4	469	51775.77	28.12	4.4	9.24
C5	474	52246.29	28.25	4.4	9.23

Table 4(B): Assessment of Stability based on PEST hypothesis and Guruprasad's observations.

Construct	P	E	S	T	Q	M	NKG
C1	6.5	5.4	4.8	9.8	3.9	0.4	19.5
C2	7.2	5.4	4.5	10.1	3.8	0.2	19.4
C3	7.1	5.3	4.9	9.6	3.8	0.4	19.2
C4	7	5.3	4.9	9.6	3.8	0.4	19.2
C5	7	5.5	4.9	9.5	3.8	0.4	19.2

Table 4(C): Validation of Modeled vaccine Constructs

Construct	Ramachandran Plot		Z Score	ERRAT 3D	WHATCHECK
	Allowed + Favored regions (in percent)	Disallowed regions(in percent)			
C1	91.5+7.3+0.8	0.5	-4.02	88.1395	Pass
C2	93.2+5.7+0.3	0.8	-4.59	83.698	Pass
C3	88.9+10.9+0.0	0.2	-4.31	86.0045	Pass
C4	92.1+6.9+0.2	0.7	-4.62	90.4444	Pass
C5	93.1+5.1+0.5	1.2	-5.57	87.3085	Pass

3.8 MD of vaccine construct

During the initial phase of MD simulation, major conformational changes in the vaccine construct were observed, as illustrated in **Figure 5A**. During the rest of the simulation time average Root Mean Square Deviation (RMSD) of the C-alpha atoms remained constant around 11.5 Å **(Figure 5A)**. This suggests that after the initial structural rearrangements, the vaccine construct achieved stable three-dimensional structure, as evident in **Supplementary Figure 2**. Followed by the initial structural rearrangements, the vaccine construct achieves a stable three dimensional structure, as evidenced by the superimposed structures obtained at 0 ns and 10 ns in **Figure 5D,** followed by another at 10 ns and 100 ns in **Figure 5E**. Further, to identify the regions of the vaccine construct that contribute to structural rearrangements, Root Mean Square

Fluctuations (RMSF) of C-alpha atoms were calculated at 0-10 ns and another from 10-100 ns in **Figures 5B and 5C,** respectively.

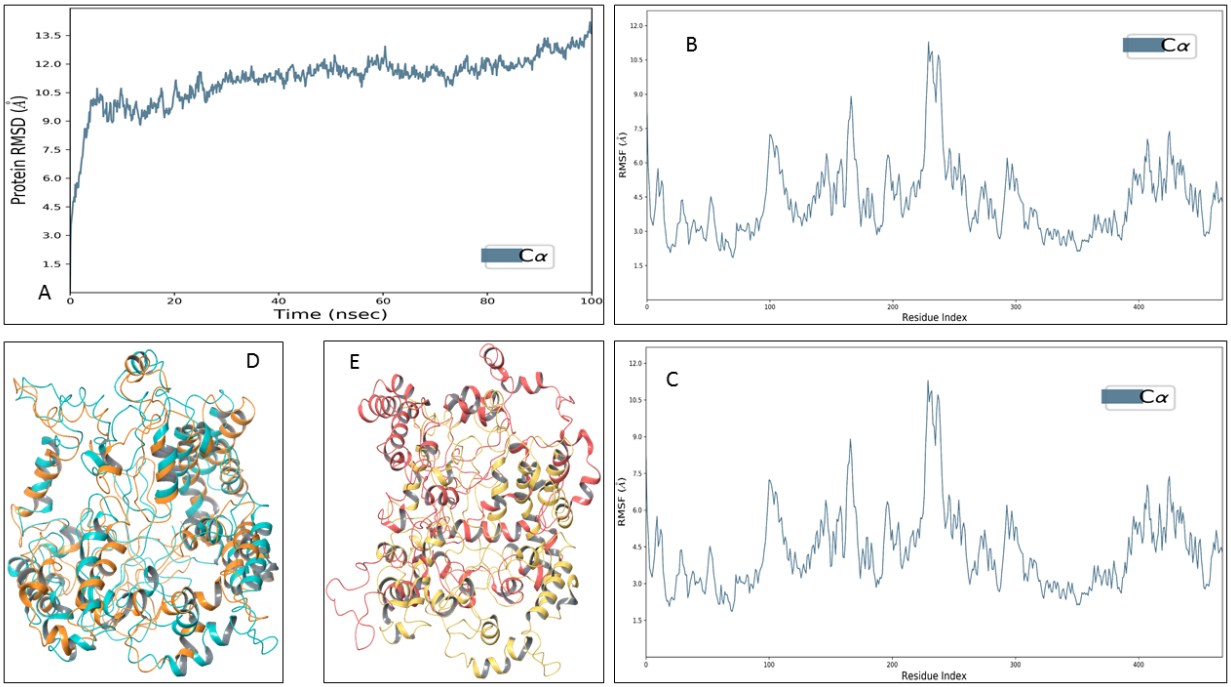


Figure 5: A: C-alpha RMSD of Vaccine Construct_4 for a 100ns MD simulation; **B:** RMSF plot comprising C-alpha atoms based on Vaccine Construct_4 from 0-10 ns; **C:** RMSF plot comprising C-alpha atoms based on Vaccine Construct_4 from 10-100 ns, the peaks represent the regions where loop is in abundance based on residue sequence (across residue index); **D:** Initial rearrangement is depicted through superimposed frames of Vaccine Construct_4 at 0 ns (in orange) and at 10 ns (in cyan); **E:** Rearrangement through superimposed frames at 10 ns (in yellow) and at 100 ns (in red).

The high residue fluctuation is due to the low distribution of stable secondary structure rearrangements in the region between residues 1-70 and 183-246, accompanied by an abundance of loops, as observed from the RMSF plots. The abundance of loops confers its affinitive ability for its complementary receptor. An evaluation of the secondary structure features present in the vaccine construct was evaluated through a Dictionary of Secondary Structure Plot (DSSP)[76] in **Supplementary Figure 1**.

The simulation indicates that Construct_4 conforms to the stability directives that are expected of it and can be reproduced under experimental conditions. Considering the MD, we may definitely extrapolate our results to include Constructs 2 and 5 as suitable vaccine constructs that can be considered in the development of vaccines since they scored on a similar index across the parameters on which we assessed and went ahead with the simulation study of Construct_4.

3.9 Docking of construct with TLR-4

Based on the MD simulation and the observed shift in the modeled vaccine construct 4, molecular docking was carried out between TLR4 receptor and the vaccine construct 4 through Patchdock, which carries out docking of rigid molecules based on a protocol that employs molecular shape representation, surface patch mapping followed by filtering and scoring. The obtained results were refined and ranked using Firedock based on a global energy value (in kJ/mol) that helps determine the binding affinity of the molecules being considered. The results indicate that the post simulation Construct_4 does bind TLR-4 with high binding affinity **(Supplementary Table 2)** with a Global Energy Value (which is an analog of ranking based on binding affinity) of -26.36 kJ/mol. The interface between Construct_4 and TLR4 was analyzed post-docking through UCSF Chimera and DimPlot. These mainly include the non-covalent interactions that are determined through distance constraints between the two docked molecules and involve hydrogen bonds and hydrophobic interactions across the interface. Each of the residues have been specified that contribute to these interactions has been visualized as spheres and depicted on the basis of the corresponding non-covalent interaction across the protein-protein interface in **Figure 6** and in **Supplementary Figure 3**. A similar docking was carried out between the modeled vaccine construct 4 (pre-MD simulation) with TLR-4 with a higher binding

affinity value (-36.05 kJ/mol) and the residues implicated in both the dockings were found to be different as is represented in **Supplementary Figure 4**.This can be attributed to the post simulation transition that is represented in the structural rearrangement. The docking with the simulated vaccine construct indicated the presence of N terminal regions of the vaccine construct to be involved along with several residues like R410,L383,I296,H268,A274,P273,F293 and T294. However, the docking with the modeled vaccine construct indicated residues like K221, K222, T225, A229, F234, A237 and D255 are involved in binding with TLR-4 only. This change in the residue pattern can conclusively indicate that the docking pose had changed significantly post the 100ns simulation run. The docking can be conclusively used to assume that the binding of the vaccine construct to the TLR receptor may elicit signal cascades responsible for the desired immune response towards conferring a suitable prophylactic effect. The utilization of servers that allow simulation of immune response generated based on vaccine dosage were not utilized in this regard owing to the limitations of the predictions across experimental settings.

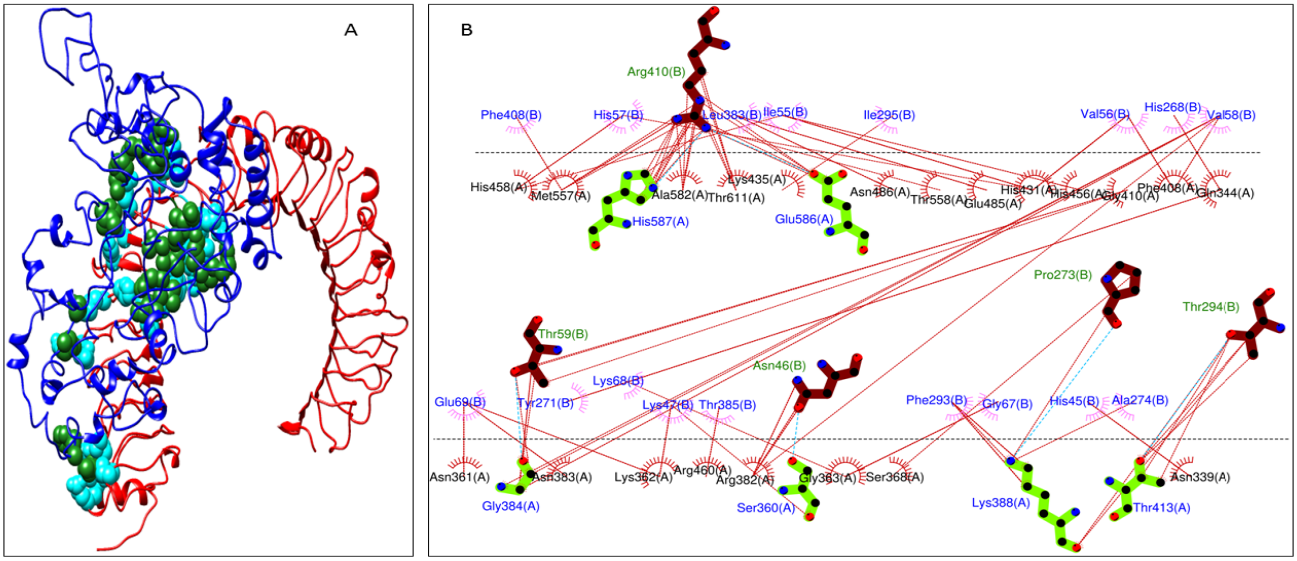


Figure 6: (A) The interface of the Vaccine Construct 4 (Chain B: blue) and TLR4 (Chain A: red) are marked by the residues involved across the protein-protein interface through UCSF Chimera. The residues have been marked by cyan colored spheres (for residues on TLR4 receptor) and dark green colored spheres (for residues on Vaccine Constructs 4) in between; **(B)** Each

residue across the interface has been labeled and the corresponding interactions they take part in across the interface. The interacting residues are labeled in **Supplementary Figures 3 and 4**.

3.10 MM/GBSA based evaluation of Docked Pose:

Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) methods are utilized to determine binding free energies for macromolecules and have been utilized to efficiently in the scoring of protein-protein interactions and suitably circumvent the computationally extensive free-energy calculation methods. Datasets employed to validate their role by comparing their scoring based on experimentally solved protein-protein interactions indicates the superiority of MM/GBSA based calculation models in determining correct docking poses. MM/GBSA has been evaluated to be better than the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) when it comes to evaluation of protein-protein interactions. The HawkDock Server has been used explicitly in this purpose post-docking to evaluate the docked pose as being accurate. The residues that were implicated across the protein-protein interface on LigPlot of the top scoring pose by FireDock have been evaluated by a scoring function based on force-field parameters influenced through MM/GBSA calculations with similar residues being implicated. The overall binding free energy of the TLR-4 and vaccine construct was predicted to be -16.91 kcal/mol (**Supplementary Table 3**).

4. Discussion

One of the most potent options that are being explored to curtail the spread of SARS-CoV-2 includes the design and development of appropriate vaccines. The entire process of vaccine development involves an extensive timeline ranging from experimental to clinical settings. In

recent times, advancement in molecular immunology and development of several epitope mapping methods, several semi-empirical approaches at vaccine design have been introduced wherein, multi-epitope-peptide based subunit vaccines based on predicted B and T cell epitopes utilizing similar bioinformatics tools has made remarkable strides. Despite the cons of low immunogenicity and multiple doses, the trade-off with eliciting neutralizing antibodies, humoral immune response and relative safety when associated with attenuated or inactivated virus vaccines plays in favor of these approaches[77].

Looking into the probable limitations, strategies have been employed to design an epitope based peptide vaccine utilizing the major subunits of the spike glycoprotein of SARS-CoV-2. In selecting the concerned immunogen, S1 and the S2 domains of the surface glycoprotein have been utilized separately, keeping in mind the limitations associated with utilizing the entire surface glycoprotein[63, 78]. The prediction of the three different types of epitopes (B cells, T_{HTL} and T_{CTL}) by utilizing multiple servers in each case based on different algorithms and selection based on a consensus involving all the servers, including experimental basis behind alleles (HLA Class I and Class II), helps add sustenance to the concluded list of epitopes[64-68]. The predicted epitopes were found to be experimentally identified across a recent microarray study of mapped epitopes on the surface glycoprotein [79]. Each of the predicted epitopes was verified based on described parameters and were found to be antigenic, non-allergenic, non-toxic and share high coverage across the spike glycoprotein of SARS-CoV-2 sequences[37]. The epitope sequences were also matched with a recently deposited SARS-CoV-2 sequence from Gujarat (QJC19491.1) and showed complete sequence coverage for all Indian sequences of the spike glycoprotein deposited till date. Based on the necessity of an immunostimulant, the role of an adjuvant becomes important and specifically, a chimeric adjuvant comprising TLR-4 agonists

was designed and validated to have overlapping B cell epitope regions through the IEDB server. The adjuvant sequences have been experimentally evidenced to generate neutralizing antibodies, cytokines and natural killer cells. A rigorous assessment of the implicated toll like receptor in the case of SARS CoV-2 has been carried out before deciding upon the use of TLR4 agonists.

The adjuvant and the predicted epitope sequences were appropriately joined by rigid linkers based on their arrangement. Even the chimeric adjuvant sequences were linked by short rigid linkers $(XP)_n$ to aid in ensuring stability. GPGPG and AAY linkers were used as intra HTL and CTL linkers, respectively. Separation between the vaccine components has been arrived upon by the role of EAAAK as a linker between adjuvant and the peptide epitope sequences[43, 44]. The designed vaccine constructs based on the arrangement of the chimeric adjuvant were validated and concluded to be highly antigenic, non-allergenic and non-toxic. This makes up for the limitations of subunit vaccines as mentioned above.

The stability of the vaccine constructs was assessed based on physico-chemical parameters that make it suitable for purification across experimental settings. The presence of cysteine bonds (without introduction of any engineered mutations) points to the stability of the modeled constructs. The presence of defined secondary structure characteristics across the modeled construct despite the presence of mainly strand and loop regions due to predicted epitopes has led us to verify the stability dynamics through a 100ns MD simulation run. Moreover, the assessment of stability through amino acid composition[51, 52], correlation with in vivo half-life periods and calculated isoelectric points[53] helps us identify three of the five modeled vaccine constructs to exhibit similar stability parameters. Based on the MD simulation of Construct 4, it can be extrapolated to include Construct 2 and Construct 5. The utilization of the ProSA web

server[49], SAVES server for Z-score and Ramchandran plot[48], ERRAT-3D[50] scores, respectively helps validate the modeled constructs.

A rigid based docking of TLR-4 and Construct 4 was carried out after this and although there are no experimental data to insinuate any such interaction but the structural features of the vaccine construct and the functional characteristics of TLR-4, an accepted dock model can be achieved in this scenario. Moreover, the N terminal is composed of TLR-4 agonists which helps compound the interaction between the two proteins. Based on TLR-4 PDB structures[57] and their corresponding binding sites followed by Ellipro[21] and Castp[59] predictions of binding sites over the receptor, docking was carried out with the vaccine construct over PatchDock[60] server and the top pose based on binding affinity values was selected. The interacting residues at the protein-protein interaction were mapped. The interaction occurs over the surfaces of TLR-4 which has a high electrostatic potential. The mapped non-covalent interactions are sufficient to indicate probable binding to the receptor surface through DimPlot and visualization through UCSF Chimera. The post docking MM/GBSA evaluation of the docked pose validates the docked pose and the involved residues across the protein-protein interaction interface.

The binding of B cell epitope sequences in the vaccine construct with the predicted B cell regions over TLR-4 will help in generation of innate immune responses through plasma B cells producing antibodies and memory B cells which confer long-lasting prophylactic response against the virus. Moreover, recent MD studies has indicated as to the effect of trimer of spike protein having much improved stability over the monomer and hence may influence the stability of spike based vaccines.

The MD simulations of the docked pose were not carried out because of computational constraints. Moreover, it becomes imperative that since no experimental data is evidenced to

verify the docking of the TLR4 and the vaccine construct, an actual determination of the protein-protein interaction between them may be carried out by running extensive MD simulations involving both the proteins (tempered binding) and determining their association and dissociation profiles alongside residence time to assess the best docking pose of the two proteins in their energy minimized conditions circumventing the limitations associated with unavailability of experimental evidence to a certain extent. Therefore, the study does not include MD simulation of the docked pose since that would have made our results biased.

5. Conflict of Interests

Authors declare no conflict of interest.

6. Authors' contributions

SS and DM conceived the idea of this study and designed the experiments together. DM and NS performed the experiments. AJ, SS, DM and JP analyzed the data. DM, NS, SS, AJ and JP contributed to drafting the manuscript.

References:

[1] W. Liu, J.S. Morse, T. Lalonde, S. Xu, Learning from the past: possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV, Chembiochem, (2020).
[2] Y. Wan, J. Shang, R. Graham, R.S. Baric, F. Li, Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus, Journal of virology, 94 (2020).
[3] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, A new coronavirus associated with human respiratory disease in China, Nature, 579 (2020) 265-269.

[4] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, The Lancet, 395 (2020) 565-574.
[5] A.C. Walls, Y.-J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, Cell, (2020).
[6] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.-L. Hsieh, O. Abiona, B.S. Graham, J.S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, Science, 367 (2020) 1260-1263.
[7] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, F. Li, Structural basis of receptor recognition by SARS-CoV-2, Nature, (2020) 1-8.
[8] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2, Science, 367 (2020) 1444-1448.
[9] G. Li, E. De Clercq, Therapeutic options for the 2019 novel coronavirus (2019-nCoV), Nature Publishing Group, 2020.
[10] D.E. Gordon, G.M. Jang, M. Bouhaddou, J. Xu, K. Obernier, M.J. O'meara, J.Z. Guo, D.L. Swaney, T.A. Tummino, R. Huttenhain, A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing, BioRxiv, (2020).
[11] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor, Cell, (2020).
[12] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, Journal of computational chemistry, 25 (2004) 1605-1612.
[13] S. Saha, G.P.S. Raghava, Prediction of continuous B-cell epitopes in an antigen using recurrent neural network, Proteins: Structure, Function, and Bioinformatics, 65 (2006) 40-48.
[14] S. Saha, G.P.S. Raghava, BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties, International Conference on Artificial Immune Systems, Springer, 2004, pp. 197-204.
[15] E.A. Emini, J.V. Hughes, D. Perlow, J. Boger, Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide, Journal of virology, 55 (1985) 836-839.
[16] P.Y. Chou, Prediction of the secondary structure of proteins from their amino acid sequence, Advances in enzymology and related areas of molecular biology, 47 (1978) 45-148.
[17] P. Karplus, G. Schulz, Prediction of chain flexibility in proteins, Naturwissenschaften, 72 (1985) 212-213.
[18] A. Kolaskar, P.C. Tongaonkar, A semi-empirical method for prediction of antigenic determinants on protein antigens, FEBS letters, 276 (1990) 172-174.
[19] J. Parker, D. Guo, R. Hodges, New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites, Biochemistry, 25 (1986) 5425-5432.
[20] M.C. Jespersen, B. Peters, M. Nielsen, P. Marcatili, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, Nucleic acids research, 45 (2017) W24-W29.
[21] J. Ponomarenko, H.-H. Bui, W. Li, N. Fusseder, P.E. Bourne, A. Sette, B. Peters, ElliPro: a new structure-based tool for the prediction of antibody epitopes, BMC bioinformatics, 9 (2008) 514.

[22] J.V. Kringelum, C. Lundegaard, O. Lund, M. Nielsen, Reliable B cell epitope predictions: impacts of method development and improved benchmarking, PLoS computational biology, 8 (2012).
[23] C.K. Hattotuwagama, P. Guan, I.A. Doytchinova, C. Zygouri, D.R. Flower, Quantitative online prediction of peptide binding to the major histocompatibility complex, Journal of Molecular Graphics and Modelling, 22 (2004) 195-207.
[24] H.-G. Rammensee, J. Bachmann, N.P.N. Emmerich, O.A. Bachor, S. Stevanović, SYFPEITHI: database for MHC ligands and peptide motifs, Immunogenetics, 50 (1999) 213-219.
[25] K.K. Jensen, M. Andreatta, P. Marcatili, S. Buus, J.A. Greenbaum, Z. Yan, A. Sette, B. Peters, M. Nielsen, Improved methods for predicting peptide binding affinity to MHC class II molecules, Immunology, 154 (2018) 394-406.
[26] P. Wang, J. Sidney, Y. Kim, A. Sette, O. Lund, M. Nielsen, B. Peters, Peptide binding predictions for HLA DR, DP and DQ molecules, BMC bioinformatics, 11 (2010) 568.
[27] S. Paul, J. Sidney, A. Sette, B. Peters, TepiTool: a pipeline for computational prediction of T cell epitope candidates, Current protocols in immunology, 114 (2016) 18.19. 11-18.19. 24.
[28] M. Andreatta, M. Nielsen, Gapped sequence alignment using artificial neural networks: application to the MHC class I system, Bioinformatics, 32 (2016) 511-517.
[29] S. Giguère, A. Drouin, A. Lacoste, M. Marchand, J. Corbeil, F. Laviolette, MHC-NP: predicting peptides naturally processed by the MHC, Journal of immunological methods, 400 (2013) 30-36.
[30] M.V. Larsen, C. Lundegaard, K. Lamberth, S. Buus, O. Lund, M. Nielsen, Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction, BMC bioinformatics, 8 (2007) 424.
[31] M. Moutaftsi, B. Peters, V. Pasquetto, D.C. Tscharke, J. Sidney, H.-H. Bui, H. Grey, A. Sette, A consensus epitope prediction approach identifies the breadth of murine T CD8+-cell responses to vaccinia virus, Nature biotechnology, 24 (2006) 817-819.
[32] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, BMC bioinformatics, 8 (2007) 4.
[33] I. Dimitrov, I. Bangov, D.R. Flower, I. Doytchinova, AllerTOP v. 2—a server for in silico prediction of allergens, Journal of molecular modeling, 20 (2014) 2278.
[34] S. Saha, G. Raghava, AlgPred: prediction of allergenic proteins and mapping of IgE epitopes, Nucleic acids research, 34 (2006) W202-W209.
[35] S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, G.P. Raghava, Peptide toxicity prediction, Computational Peptidology, Springer2015, pp. 143-157.
[36] S.K. Dhanda, P. Vir, G.P. Raghava, Designing of interferon-gamma inducing MHC class-II binders, Biology direct, 8 (2013) 30.
[37] P.D. Yadav, V.A. Potdar, M.L. Choudhary, D.A. Nyayanit, M. Agrawal, S.M. Jadhav, T.D. Majumdar, A. Shete-Aich, A. Basu, P. Abraham, Full-genome sequences of the first two SARS-CoV-2 viruses from India, The Indian journal of medical research, (2020).
[38] S. Shi, H. Zhu, X. Xia, Z. Liang, X. Ma, B. Sun, Vaccine adjuvants: Understanding the structure and mechanism of adjuvanticity, Vaccine, (2019).
[39] T.J. Albin, J.K. Tom, S. Manna, A.P. Gilkes, S.A. Stetkevich, B.B. Katz, M. Supnet, J. Felgner, A. Jain, R. Nakajima, Linked Toll-Like Receptor Triagonists Stimulate Distinct, Combination-Dependent Innate Immune Responses, ACS central science, 5 (2019) 1137-1145.

[40] V. Pavot, N. Rochereau, J. Rességuier, A. Gutjahr, C. Genin, G. Tiraby, E. Perouzel, T. Lioux, F. Vernejoul, B. Verrier, Cutting edge: New chimeric NOD2/TLR2 adjuvant drastically increases vaccine immunogenicity, The Journal of Immunology, 193 (2014) 5781-5785.
[41] N. Gupta, H. Regar, V.K. Verma, D. Prusty, A. Mishra, V.K. Prajapati, Receptor-ligand based molecular interaction to discover adjuvant for immune cell TLRs to develop next-generation vaccine, International Journal of Biological Macromolecules, (2020).
[42] Y. Fu, Y. Cheng, Y. Wu, Understanding SARS-CoV-2-mediated inflammatory responses: from mechanisms to potential therapeutic tools, Virologica Sinica, (2020) 1-6.
[43] X. Chen, J.L. Zaro, W.-C. Shen, Fusion protein linkers: property, design and functionality, Advanced drug delivery reviews, 65 (2013) 1357-1369.
[44] G. Li, Z. Huang, C. Zhang, B.-J. Dong, R.-H. Guo, H.-W. Yue, L.-T. Yan, X.-H. Xing, Construction of a linker library with widely controllable flexibility for fusion protein design, Applied microbiology and biotechnology, 100 (2016) 215-225.
[45] E. Gasteiger, C. Hoogland, A. Gattiker, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, The proteomics protocols handbook, Springer2005, pp. 571-607.
[46] Y. Song, F. DiMaio, R.Y.-R. Wang, D. Kim, C. Miles, T. Brunette, J. Thompson, D. Baker, High-resolution comparative modeling with RosettaCM, Structure, 21 (2013) 1735-1742.
[47] W.-H. Shin, G.R. Lee, L. Heo, H. Lee, C. Seok, Prediction of protein structure and interaction by GALAXY protein modeling programs, Bio Design, 2 (2014) 1-11.
[48] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, Journal of applied crystallography, 26 (1993) 283-291.
[49] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic acids research, 35 (2007) W407-W410.
[50] C. Colovos, T.O. Yeates, Verification of protein structures: patterns of nonbonded atomic interactions, Protein science, 2 (1993) 1511-1519.
[51] K. Guruprasad, B.B. Reddy, M.W. Pandit, Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence, Protein Engineering, Design and Selection, 4 (1990) 155-161.
[52] M. Rechsteiner, S.W. Rogers, PEST sequences and regulation by proteolysis, Trends in biochemical sciences, 21 (1996) 267-271.
[53] J.F. Dice, A.L. Goldberg, Relationship between in vivo degradative rates and isoelectric points of proteins, Proceedings of the National Academy of Sciences, 72 (1975) 3893-3897.
[54] R.W. Hooft, G. Vriend, C. Sander, E.E. Abola, Errors in protein structures, Nature, 381 (1996) 272-272.
[55] K.J. Bowers, D.E. Chow, H. Xu, R.O. Dror, M.P. Eastwood, B.A. Gregersen, J.L. Klepeis, I. Kolossvary, M.A. Moraes, F.D. Sacerdoti, Scalable algorithms for molecular dynamics simulations on commodity clusters, SC'06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, IEEE, 2006, pp. 43-43.
[56] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, Journal of molecular graphics, 14 (1996) 33-38.
[57] B.S. Park, D.H. Song, H.M. Kim, B.-S. Choi, H. Lee, J.-O. Lee, The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex, Nature, 458 (2009) 1191-1195.

[58] S. Akira, S. Uematsu, O. Takeuchi, Pathogen recognition and innate immunity, Cell, 124 (2006) 783-801.
[59] W. Tian, C. Chen, X. Lei, J. Zhao, J. Liang, CASTp 3.0: computed atlas of surface topography of proteins, Nucleic acids research, 46 (2018) W363-W367.
[60] D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H.J. Wolfson, PatchDock and SymmDock: servers for rigid and symmetric docking, Nucleic acids research, 33 (2005) W363-W367.
[61] E. Mashiach, D. Schneidman-Duhovny, N. Andrusier, R. Nussinov, H.J. Wolfson, FireDock: a web server for fast interaction refinement in molecular docking, Nucleic acids research, 36 (2008) W229-W232.
[62] F. Chen, H. Liu, H. Sun, P. Pan, Y. Li, D. Li, T. Hou, Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein–protein binding free energies and re-rank binding poses generated by protein–protein docking, Physical Chemistry Chemical Physics, 18 (2016) 22129-22139.
[63] M. Jaume, M. Yip, Y. Kam, C. Cheung, F. Kien, A. Roberts, P. Li, I. Dutry, N. Escriou, M. Daeron, SARS CoV subunit vaccine: antibodymediated neutralisation and enhancement, (2012).
[64] A. Grifoni, J. Sidney, Y. Zhang, R.H. Scheuermann, B. Peters, A. Sette, A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2, Cell host & microbe, (2020).
[65] D. Weiskopf, M.A. Angelo, E.L. de Azeredo, J. Sidney, J.A. Greenbaum, A.N. Fernando, A. Broadwater, R.V. Kolla, A.D. De Silva, A.M. de Silva, Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells, Proceedings of the National Academy of Sciences, 110 (2013) E2046-E2053.
[66] H. Spínola, HLA loci and respiratory infectious diseases, Journal of Respiratory Research, 2 (2016) 56-66.
[67] M.H. Ng, K.-M. Lau, L. Li, S.-H. Cheng, W.Y. Chan, P.K. Hui, B. Zee, C.-B. Leung, J.J. Sung, Association of human-leukocyte-antigen class I (B* 0703) and class II (DRB1* 0301) genotypes with susceptibility and resistance to the development of severe acute respiratory syndrome, Journal of Infectious Diseases, 190 (2004) 515-518.
[68] J. Greenbaum, J. Sidney, J. Chung, C. Brander, B. Peters, A. Sette, Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes, Immunogenetics, 63 (2011) 325-335.
[69] A. Shanmugam, S. Rajoria, A.L. George, A. Mittelman, R. Suriano, R.K. Tiwari, Synthetic Toll like receptor-4 (TLR-4) agonist peptides as a novel class of adjuvants, PloS one, 7 (2012).
[70] A. Tandon, M. Pathak, M.K. Harioudh, S. Ahmad, M. Sayeed, T. Afshan, M. Siddiqi, K. Mitra, S.M. Bhattacharya, J.K. Ghosh, A TLR4-derived non-cytotoxic, self-assembling peptide functions as a vaccine adjuvant in mice, Journal of Biological Chemistry, 293 (2018) 19874-19885.
[71] M. Dabaghian, A.M. Latifi, M. Tebianian, F. Dabaghian, S.M. Ebrahimi, A truncated C-terminal fragment of Mycobacterium tuberculosis HSP70 enhances cell-mediated immune response and longevity of the total IgG to influenza A virus M2e protein in mice, Antiviral research, 120 (2015) 23-31.
[72] S.G. Reed, F.-C. Hsu, D. Carter, M.T. Orr, The science of vaccine adjuvants: advances in TLR4 ligand adjuvants, Current opinion in immunology, 41 (2016) 85-90.
[73] Y. Wang, T. Whittall, E. McGowan, J. Younson, C. Kelly, L.A. Bergmeier, M. Singh, T. Lehner, Identification of stimulating and inhibitory epitopes within the heat shock protein 70

molecule that modulate cytokine production and maturation of dendritic cells, The Journal of Immunology, 174 (2005) 3306-3316.
[74] G. Multhoff, K. Pfister, M. Gehrmann, M. Hantschel, C. Gross, M. Hafner, W. Hiddemann, A 14-mer Hsp70 peptide stimulates natural killer (NK) cell activity, Cell stress & chaperones, 6 (2001) 337.
[75] P.M. Moyle, Biotechnology approaches to produce potent, self-adjuvanting antigen-adjuvant fusion protein subunit vaccines, Biotechnology advances, 35 (2017) 375-389.
[76] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers: Original Research on Biomolecules, 22 (1983) 2577-2637.
[77] W. Shang, Y. Yang, Y. Rao, X. Rao, The outbreak of SARS-CoV-2 pneumonia calls for viral vaccines, npj Vaccines, 5 (2020) 1-3.
[78] Y.W. Kam, F. Kien, A. Roberts, Y.C. Cheung, E.W. Lamirande, L. Vogel, S.L. Chu, J. Tse, J. Guarner, S.R. Zaki, Antibodies against trimeric S glycoprotein protect hamsters against SARS-CoV challenge despite their capacity to mediate FcγRII-dependent entry into B cells in vitro, Vaccine, 25 (2007) 729-740.
[79] H. Wang, X. Hou, X. Wu, T. Liang, X. Zhang, D. Wang, F. Teng, J. Dai, H. Duan, S. Guo, SARS-CoV-2 proteome microarray for mapping COVID-19 antibody interactions at amino acid resolution, bioRxiv, (2020).

Multi-epitope based peptide vaccine design against SARS-CoV-2 using its spike protein

Supplementary Data

Supplementary Table 1A: Initial adjuvant arrangements considered

Constructs	Arrangement	Sequence
1	A+B	NTTIPTKRSETFTTADDNQPSVQIQVYQGEREIAAHNKFDIDANGIVHVTAKKDKGTGKENTAHAEEDRKRREEADVRNQAKFVKEQREAEGGSKVNLKQMSEFSVFLSLRNLIYL
2	B+A	NLKQMSEFSVFLSLRNLIYLNTTIPTKRSETFTTADDNQPSVQIQVYQGEREIAAHNKFDIDANGIVHVTAKKDKGTGKENTAHAEEDRKRREEADVRNQAKFVKEQREAEGGSKV
3	C+B+A	APPHALSNLKQMSEFSVFLSLRNLIYLNTTIPTKRSETFTTADDNQPSVQIQVYQGEREIAAHNKFDIDANGIVHVTAKKDKGTGKENTAHAEEDRKRREEADVRNQAKFVKEQREAEGGSKV
4	B+C+A	NLKQMSEFSVFLSLRNLIYLAPPHALSNTTIPTKRSETFTTADDNQPSVQIQVYQGEREIAAHNKFDIDANGIVHVTAKKDKGTGKENTAHAEEDRKRREEADVRNQAKFVKEQREAEGGSKV
5	B+A+C	NLKQMSEFSVFLSLRNLIYLNTTIPTKRSETFTTADDNQPSVQIQVYQGEREIAAHNKFDIDANGIVHVTAKKDKGTGKENTAHAEEDRKRREEADVRNQAKFVKEQREAEGGSKVAPPHALS
6	C+A	APPHALSNTTIPTKRSETFTTADDNQPSVQIQVYQGEREIAAHNKFDIDANGIVHVTAKKDKGTGKENTAHAEEDRKRREEADVRNQAKFVKEQREAEGGSKV
7	A+B+C	NTTIPTKRSETFTTADDNQPSVQIQVYQGEREIAAHNKFDIDANGIVHVTAKKDKGTGKENTAHAEEDRKRREEADVRNQAKFVKEQREAEGGSKVNLKQMSEFSVFLSLRNLIYLAPPHALS
8	A+C+B	NTTIPTKRSETFTTADDNQPSVQIQVYQGEREIAAHNKFDIDANGIVHVTAKKDKGTGKENTAHAEEDRKRREEADVRNQAKFVKEQREAEGGSKVAPPHALSNLKQMSEFSVFLSLRNLIYL
9	C+A+B	APPHALSNTTIPTKRSETFTTADDNQPSVQIQVYQGEREIAAHNKFDIDANGIVHVTAKKDKGTGKENTAHAEEDRKRREEADVRNQAKFVKEQREAEGGSKVNLKQMSEFSVFLSLRNLIYL
Color Coding Key:- A: Hsp70, B: TR-433, C: RS09		

Supplementary Table 1B : The 2nd, 6th, 7th and 9th construct were found to be the most antigenic (a stringent Vaxijen score of =>0.9 whereas threshold lies at 0.4) and were utilized to design the vaccine constructs

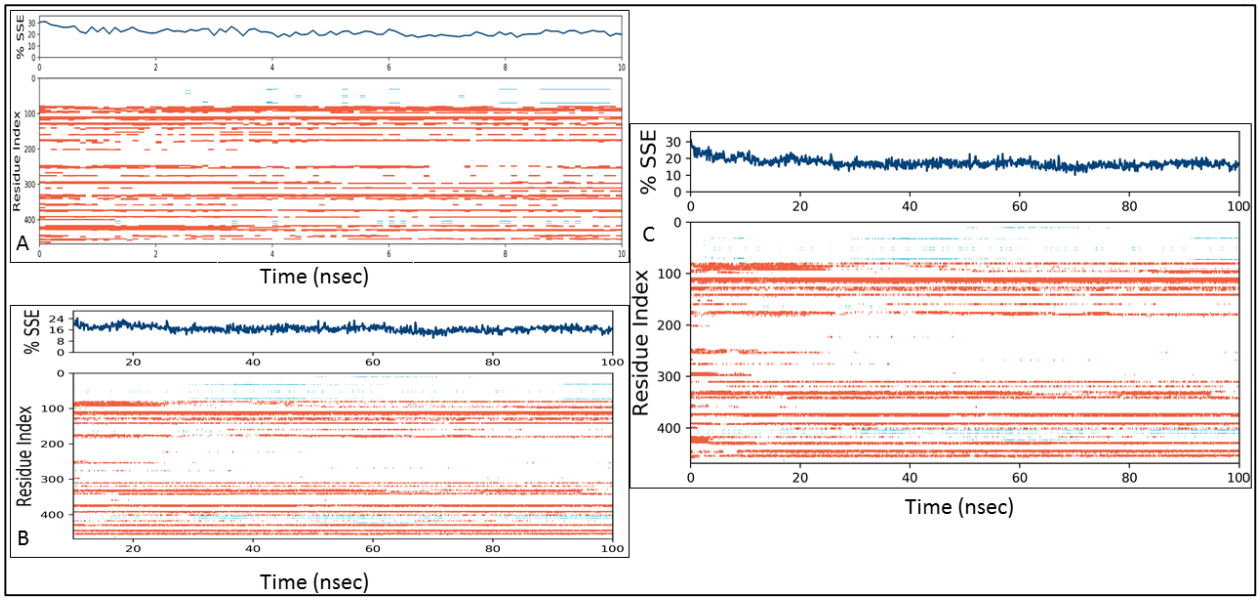
Adjuvant	Arrangement	Allergen	Antigen	Toxin
1	B+A	NA	0.88	No
2	A+B	NA	0.9332	No
3	C+B+A	NA	0.8151	No
4	B+C+A	NA	0.88	No
5	B+A+C	NA	0.85	No
6	C+A	NA	1.0145	No
7	A+B+C	NA	0.931	No
8	A+C+B	NA	0.82	No
9	C+A+B	NA	0.9	No
**NA refers to a non-allergen and all adjuvants were predicted to be non-toxic in nature				

Supplementary Table 2: Patchdock mediated docking between Construct_4 and TLR-4 with Ranking by FireDock based on Global Energy Value.

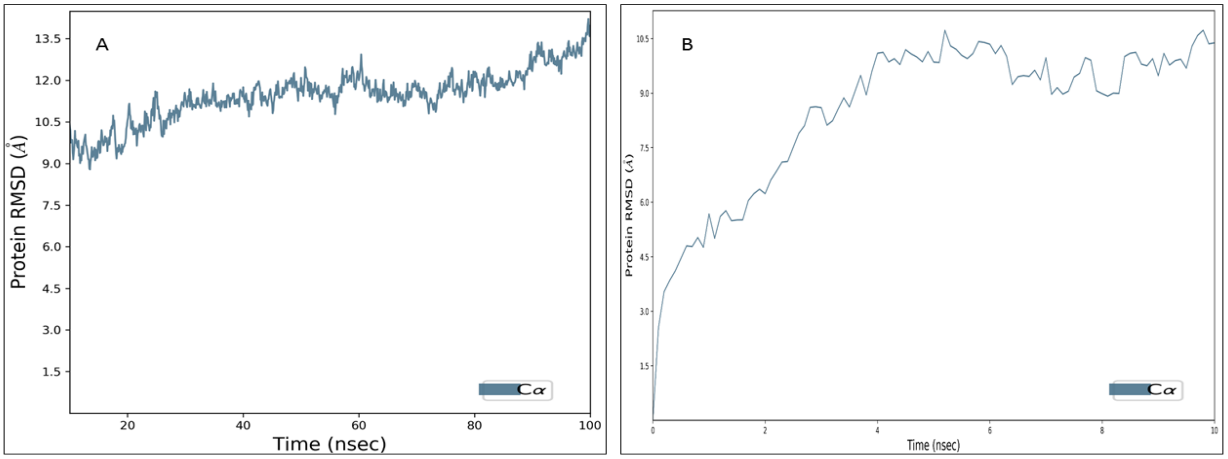
Sl. No.	Docked Poses	Global Energy Value (kJ/mol)	Score	Area	ACE	HB
1	Pose_1	-26.36	14654	2021.40	177.80	-3.21
2	Pose_2	-18.86	14002	2138.60	196.19	-1.89
3	Pose_3	-18.43	12262	1559.00	213.92	-4.19
4	Pose_4	-11.04	10796	1574.50	216.66	-4.57

Supplementary Table 3: The docked pose of the TLR-4 and the vaccine construct has been evaluated across Van der Waal potentials (**VDW**), electrostatic potentials (**ELE**), polar solvation free energies predicted by the Generalized Born model (**GB**) and the non-polar contribution to the solvation free energy calculated through an empirical model (**SA**).

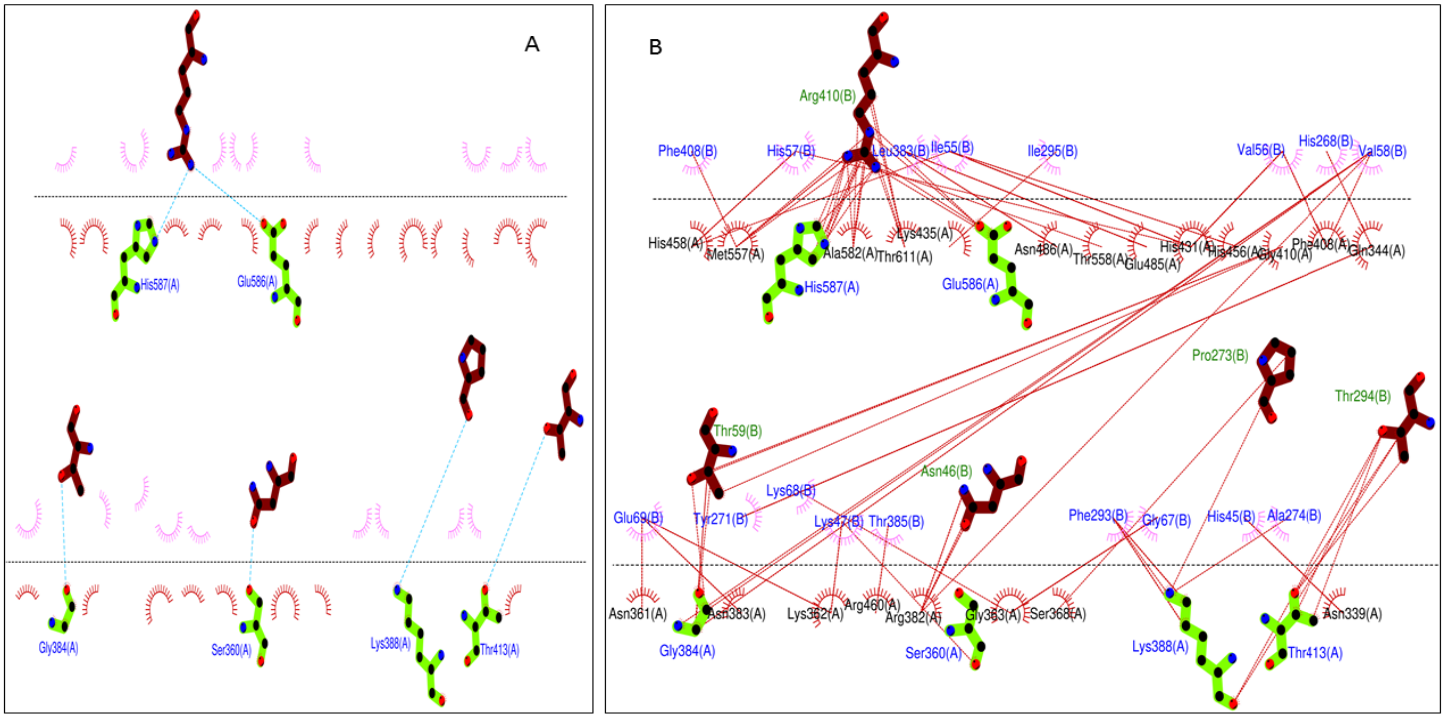
Complex	VDW	ELE	GB	SA	TOTAL
Pose_1	-145.27	-1063.43	1208.14	-16.36	-16.91



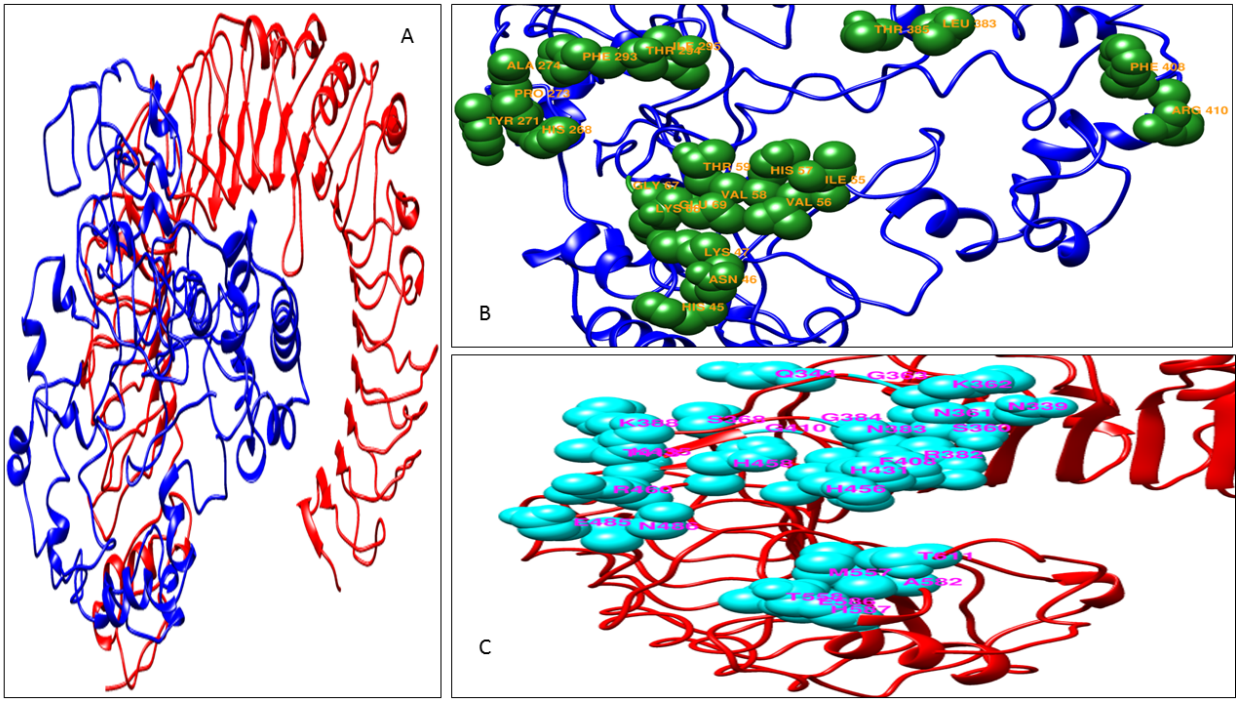
Supplementary Figure 1: The DSSP plot of Vaccine Construct_4 at **A:** 0-10 ns; **B:** 10-100ns and **C:** 0-100ns of the MD simulation.



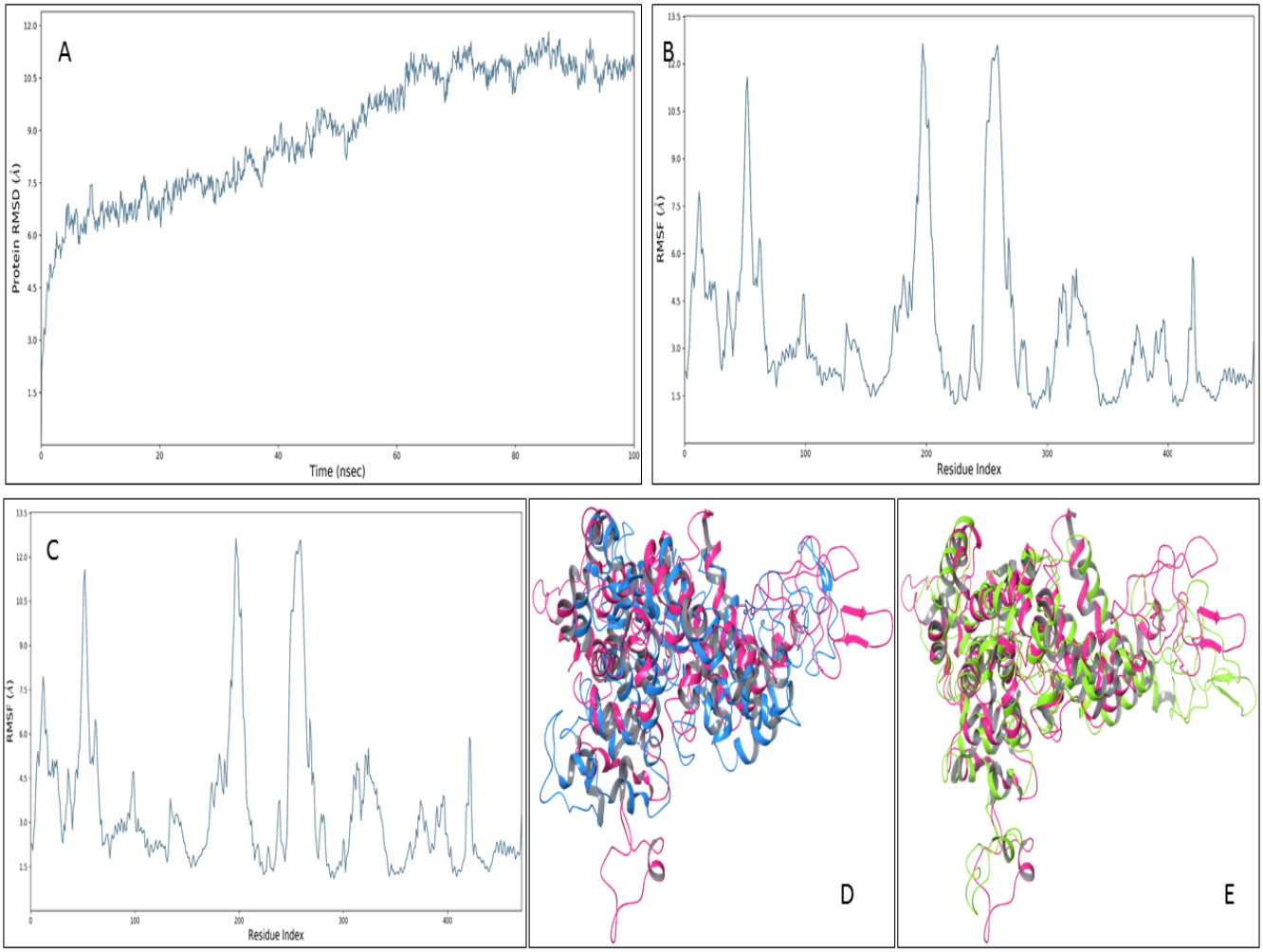
Supplementary Figure 2: The RMSD plot of Vaccine Construct_4 at **A:** 10-100 ns; **B:** 0-10ns.



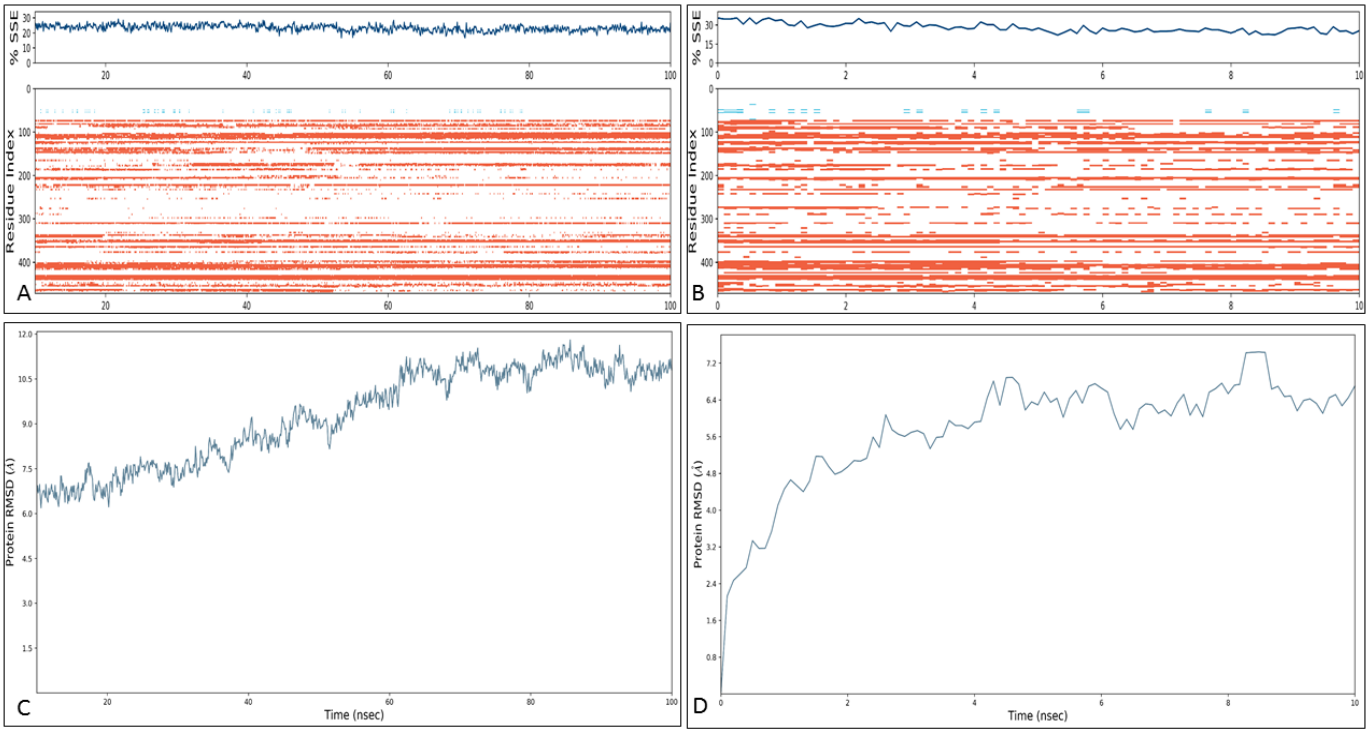
Supplementary Figure 3: The residues which contribute to contacts across the docked protein-protein interface of Construct 4 and TLR4; **A:** The corresponding hydrogen bond contacts and **B:** The corresponding hydrophobic contacts across the interface.



Supplementary Figure 4: The residues which contribute to interactions across the docked protein-protein interface of **A:** Vaccine Construct 4 (Chain B **blue**) and TLR4 (Chain A **red**); **B:** Green spheres upon the Vaccine Construct 4 (Chain B in blue) which contribute to the protein-protein interaction; **C:** Cyan spheres upon the TLR 4 (Chain A in red) which contribute to the protein-protein interaction .



Supplementary Figure 5: (**A**) C-alpha RMSD of Vaccine Construct_5 for a 100ns MD simulation; **(B):** RMSF plot comprising of C-alpha atoms based on Vaccine Construct_5 from 0-10 ns; **(C)** RMSF plot comprising of C-alpha atoms based on Vaccine Construct_5 from 10-100 ns, the peaks represents the regions where loop is in abundance based on residue sequence (across residue index); **(D)** Initial rearrangement is depicted through superimposed frames of Vaccine Construct_5 at 10 ns (in pink) and at 100 ns (in blue); **(E)** Rearrangement through superimposed frames at 0 ns (in green) and at 10 ns (in pink).



Supplementary Figure 6: The DSSP plot of Vaccine Construct_5 at **(A)** 0-10o ns; **(B)** 0-10ns of the MD simulation. The RMSD plot of Vaccine Construct_5 at **(C)** 10-100 ns; **(D)** 0-10ns.