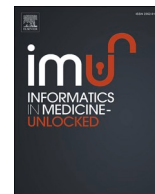




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Immunoinformatics and molecular modeling approach to design universal multi-epitope vaccine for SARS-CoV-2

Md. Tahsin Khan<sup>a,b,\*</sup>, Md. Jahirul Islam<sup>a</sup>, Arpana Parihar<sup>a,c</sup>, Rahatul Islam<sup>b</sup>, Tarhima Jahan Jerin<sup>d</sup>, Rupali Dhote<sup>a,c</sup>, Md. Ackas Ali<sup>a</sup>, Fariha Khan Laura<sup>a</sup>, Mohammad A. Halim<sup>a,e</sup>

<sup>a</sup> Division of Infectious Diseases, The Red-Green Research Centre, BICCB, 16 Tejkunipara, Tejgaon, Dhaka, 1215, Bangladesh

<sup>b</sup> Department of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet, 3114, Bangladesh

<sup>c</sup> Department of Genetics, Barkatullah University, Bhopal, Madhya Pradesh, 462026, India

<sup>d</sup> Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Tangail, 1902, Bangladesh

<sup>e</sup> Department of Physical Sciences, University of Arkansas-Fort Smith, Fort Smith, AR, USA

### ARTICLE INFO

#### Keywords:

Immunoinformatics  
Multi-epitope vaccine  
SARS-CoV-1  
SARS-CoV-2  
Molecular dynamics simulation

### ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a highly transmittable and pathogenic human coronavirus that caused a pandemic situation of acute respiratory syndrome, called COVID-19, which has posed a significant threat to global health security. The aim of the present study is to computationally design an effective peptide-based multi-epitope vaccine (MEV) against SARS-CoV-2. The overall model quality of the vaccine candidate, immunogenicity, allergenicity, and physicochemical analysis have been conducted and validated. Molecular dynamics studies confirmed the stability of the candidate vaccine. The docked complexes during the simulation revealed a strong and stable binding interactions of MEV with human and mice toll-like receptors (TLR), TLR3 and TLR4. Finally, candidate vaccine codons have been optimized for their *in silico* cloning in *E. coli* expression system, to confirm increased expression. The proposed MEV can be a potential candidate against SARS-CoV-2, but experimental validation is needed to ensure its safety and immunogenicity status.

### 1. Introduction

The world health organization (WHO) on March 11, 2020 declared the recent novel coronavirus (SARS-CoV-2) outbreak as a pandemic. This outbreak not only affected millions of people worldwide, and posed a serious health burden to society, but also led to extreme economic imbalance globally. SARS-CoV-2 belongs to the family beta coronaviruses and are closely associated with Severe Acute Respiratory Syndrome Virus (SARS-CoV) and Middle East Respiratory Syndrome Virus (MERS-CoV), which were responsible for the outbreak in 2002 and 2012 respectively [1,2].

SARS-CoV-2 is a small virus containing positive sense single stranded RNA genome of ~30 Kb in length which encodes 9860 amino acids. The genome is translated into mainly three types of proteins: 1) non-structural proteins (ORF1a/b, nsp2, nsp3, nsp4, nsp9), 2) accessory proteins (ORF3a, ORF8, ORF7a/b), and 3) structural proteins which includes spike, envelope, membrane and nucleocapsid protein [3,4]. In

this regard, a recent study on the SARS-CoV-2 human protein interaction map illustrates several SARS-CoV-2 and human protein receptors as putative therapeutic targets [2]. Non-structural proteins include viral enzymes and other accessory proteins that assist genome replication and pathogenesis. Structural proteins play key roles in virus pathogenesis, replication, and spread of infection. Basically, interaction of spike protein with human ACE2 receptors leads to the internalization of viruses inside the host cells, where the viral genome is replicated and interacted with nucleocapsid protein. After replication, the envelope and membrane proteins help with the assembly and packaging of progeny virus particles [3–6].

There is no specific drugs or vaccine established for the treatment of COVID-19 disease. The Pfizer/BioNTech vaccine is 95% effective in preventing COVID-19 infection [7], but its approval status varies worldwide. In countries where the vaccine has not been approved by the relevant regulatory authority, it is an investigational drug, and its safety and efficacy have not been established [8]. The most promising

\* Corresponding author. Division of Infectious Diseases, The Red-Green Research Centre, BICCB, 16 Tejkunipara, Tejgaon, Dhaka, 1215, Bangladesh.  
E-mail address: [tahsinkhan570@gmail.com](mailto:tahsinkhan570@gmail.com) (Md.T. Khan).

<https://doi.org/10.1016/j.imu.2021.100578>

Received 20 December 2020; Received in revised form 12 April 2021; Accepted 13 April 2021

Available online 21 April 2021

2352-9148/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

approach to deal with any pandemic is vaccination; however, the development of vaccines is a challenging process. Lessons and experiences learned from earlier SARS and MERS outbreaks as well as the advancement of technology has accelerated the process of development of vaccines. So far, many vaccine candidates have either entered the clinical trial or are ready to start the trial process by fulfilling preliminary evaluation protocols. For instance, Moderna's mRNA-1273 based vaccine entered phase 1/2 trial, DNA Vaccine named INO-4800 by Inovio pharmaceuticals, microneedle array delivered recombinant coronavirus vaccines etc. have shown promising results [9]. The most important strategies for the development of vaccines against SARS-CoV-2 are inactivated or weakened virus, replicating or non-replicating viral vector-based approaches, DNA, RNA, virus particle like approaches and protein subunit-based approaches, the later includes immunoinformatics approach [10].

Among all, the immunoinformatics and reverse vaccinomics approaches are rapid, accurate, cost-effective and reliable against pathogens [11]. In this regard, pioneering work of Ruth Arnon on an epitope-based vaccine against influenza (H5N1) virus proved the reliability of immunoinformatics approach [11,12]. Several groups also reported the immunoinformatics approaches to design subunit vaccines against SARS-CoV-2, however, only two of them are under preclinical trials [13]. Several studies proposed subunit vaccines through the immunoinformatics approach to accelerate the race over safe vaccine development, but they were limited to either single protein [14,15] or very few proteins of SARS-CoV-2 [16,17].

This study aims to design a multi-epitope vaccine against SARS-CoV-2 that efficiently elicits both innate and adaptive immune responses into the host body and thereby provides maximum protection. Therefore, 13 out of 21 structural and non-structural proteins (spike, nucleocapsid, membrane, envelope, endoRNase, nsp4, nsp9, nsp6, ORF3a, ORF6, ORF7a, ORF8 and ORF10) of SARS-CoV-2 are selected based on their antigenicity score. These proteins are analyzed for prediction of Cytotoxic T cell (CTL), Helper T cell (HTL) and B cell epitope. The toxicity, allergenicity and conservancy of epitopes are checked along with the population coverage study. Combining adjuvants, most promising CTL, HTL and B-Cell epitope along with suitable linker, the vaccine is constructed. Vaccine structure is modeled and docked against TLR3 and TLR4 to ensure its efficacy of inducing an immune response. Furthermore, the docked complex is validated and analyzed using molecular dynamics simulations. Finally, to ensure the expression and translation efficiency of vaccine the *in silico* cloning is performed. The unique feature of our study over other existing reported peptide vaccine is that the selected epitopes could be able to elicit immune response in both human and mice, which would be beneficial in terms of cost reduction in the process of *in vivo* efficacy testing of the vaccine construct.

## 2. Methods

The bioinformatics resources used in the current study are presented in [Supplementary Table S1](#).

### 2.1. Protein selection

The reference proteins of SARS Corona Virus-2 were retrieved from National Center for Biotechnology Information (NCBI). VaxiJen v2 server was utilized to screen out highly antigenic proteins setting the cutoff score of >0.45. We selected 21 protein in FASTA format including nsp2 (YP\_009725298.1), nsp4 (YP\_009725300.1), 3C-like proteinase (YP\_009725301.1), nsp6 (YP\_009725302.1), nsp7 (YP\_009725303.1), nsp8 (YP\_009725304.1), nsp9 (YP\_009725305.1), nsp10 (YP\_009725306.1), RNA-dependent RNA polymerase (YP\_009725307.1), endoRNase (YP\_009725310.1), helicase (YP\_009725308.1), spike (YP\_009724390.1), envelope (YP\_009724392.1), nucleocapsid (YP\_009724397.2), ORF3a (YP\_009724391.1), ORF6 (YP\_009724394.1), ORF7a

(YP\_009724395.1), ORF8 (YP\_009724396.1), leader protein (YP\_009725297.1), 3'-to-5' exonuclease (YP\_009725309.1), and 2'-O-ribose methyltransferase (YP\_009725311.1). Proteins showing antigenic score above the threshold were selected for epitope screening.

### 2.2. MHC class I (CD8<sup>+</sup> T cell) epitope screening

Selected proteins were screened through NetCTL 1.2 server with default parameters (Threshold- 0.75; Weight on C terminal cleavage: 0.15 and Weight on TAP transport efficiency: 0.05) to predict 9-mer CTL epitopes against 12 human super-families of HLA Class-I alleles (A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, B62). The epitope length was selected as 9 mer, because the NetCTL 1.2 server has been trained on a set of 886 known 9 mer MHC class I ligands, whereas the average AUC score (Area under the ROC Curve) of the server is 0.941. Epitopes showing good binding affinity with human supertypes were screened through 1) TMHMM server to predict the topology, 2) VaxiJen v2.0 server to predict antigenic score, 3) AllerTOP v.2.0 server to predict allergenicity, and 4) ToxinPred server to predict toxicity. Epitopes which passed the screening as topologically outside, antigenic, non-allergen and nontoxic were passed through NetMHC 4 server for Mice alleles (H-2-D<sup>b</sup>, H-2-D<sup>d</sup>, H-2-K<sup>b</sup>, H-2-K<sup>d</sup>, H-2-K<sup>k</sup> and H-2-L<sup>d</sup>) binding. Epitopes showing positive results were further cross-checked with IEDB class I immunogenicity server for Human HLA class I binding. Finally, most HLA covering epitopes were selected as final epitopes and subjected to population coverage analysis.

### 2.3. MHC class II (CD4<sup>+</sup> T cell) epitope screening

The IEDB MHC II binding server was used to predict epitopes from selected antigenic proteins. Firstly, 15-mer HTL epitopes were predicted for Mice alleles (H2-IA<sup>b</sup>, H2-IA<sup>d</sup>, H2-IE<sup>d</sup>) with a selecting percentile rank <10. Then, the TMHMM, VaxiJen v2.0, AllerTOP v.2.0 and ToxinPred servers were utilized as in the first filtration such as for CTL screening. Secondly, epitopes were screened through the IFNepitope server to select epitopes with the capability to induce IFN- $\gamma$  production. Finally, epitopes showing positive results were further cross-checked with the IEDB class II immunogenicity server for Human HLA class II binding. As with CTL screening, most HLA class II covering epitopes were selected as final HTL epitopes and then subjected to population coverage analysis.

### 2.4. Linear and discontinuous B cell epitope screening

The available webserver used for B cell epitope prediction was not accurate enough, and B-cell epitope prediction is a semi-low accuracy procedure. Hence, in our study we predicted the B cell epitopes through a multi-method approach including (i) the physicochemical-based algorithms (e.g., Bcepred and IEDB-based linear epitope prediction) and (ii) the machine learning based algorithms (e.g., ABCpred, BepiPred, and LBtope). We initially used the ABCpred webserver as a standard for BCL (B cell lymphoma) prediction (14, 16, 18 and 20-mer), and furthermore, other webserver were used to find overlapped epitopes through cross checking. Finally, overlapped consensus epitope were screened through VaxiJen v2.0, AllerTOP v.2.0 and ToxinPred servers for selecting the final linear BCL epitopes.

We also predicted the discontinuous epitopes through IEDB Ellipro tools in our study. Structural proteins were used for this analysis. Three dimensional (3D) structures of structural proteins were retrieved from RCSB (Research Collaboratory for Structural Bioinformatics) and I-TASSER online server. Finally, shortlisted overlapped consensus epitopes were cross-checked with the IEDB Ellipro webserver to find overlapped discontinuous epitope residues.

### 2.5. Population coverage and conservancy analysis

Shortlisted CTL and HTL epitopes along with respective HLA allele

types were submitted to the IEDB population coverage analysis tool. Calculations were performed separately for MHC class I and MHC class II as well as in the combined condition.

Structural protein sequences from other strains of SARS-CoV-2 were retrieved from NCBI database. Then, the IEDB Epitope Conservancy analysis tool was employed for shortlisted epitopes with a  $\geq 100\%$  sequence identity threshold.

### 2.6. Vaccine construction and structural model

To construct a multi-subunit vaccine, CTL, HTL and BCL were joined in an orderly fashion with appropriate linker. The vaccine construct started with beta defensin-3 adjuvant at the N-terminal end followed by EAAAK linker and PADRE sequence. GGGG linkers were added to join each CTL epitopes, whereas GPGPG and KK linkers were used to join HTL and BCL, respectively. Finally, an additional PADRE sequence was added at C-terminal end. Several parameters such as Physicochemical properties, solubility, antigenicity and allergenicity of designed vaccine construct were analyzed through the ExpASY-ProtParam, Protein-Sol, VaxiJen v.20 and AlgPred servers, respectively.

The secondary structure of the vaccine protein was predicted using the Self-Optimized Prediction Method with Alignment (SOPMA) server [18]. It predicts four conformational states including helix, beta sheets and bridges, turns and coils. The Robetta server was utilized to construct 3D vaccine protein models. The five generated models were evaluated through Rampage Ramachandran plot analysis, while only the best model was optimized in YASARA dynamics suite through 100 ns of MD simulation.

### 2.7. Molecular docking and dynamics simulation of vaccine-TLRs complex

Human TLR-3 (PDB id 1ZIW), human TLR-4 (PDB id: 4G8A), mice TLR3 (3CIG) and mice TLR4 (PDB id 2Z64) were retrieved from the protein database (PDB) and prepared for docking analysis. Protein-protein docking was performed through the ClusPro 2.0 server. Based on the lowest energy score, efficiently docked complexes were selected and downloaded. Interactions between vaccine and TLR complexes were evaluated.

The docked Toll-like Receptor TLRs (human TLR3, 4 and mice TLR3, 4) and constructed multi-epitope vaccine complex MD simulations were computed using the YASARA dynamics suite [19]. The YASARA dynamics suite was also used to estimate conformational change and all possible interactions. AMBER 14 [20] was considered for all MD simulations at constant temperature (298 K) using the NVT ensemble. The 0.9% concentration of NaCl salt was used to neutralize the system where water molecules (0.998 g/cm<sup>3</sup> density) were added. The cell size was 20 Å larger than the complex in all cases. A cuboid cell box and periodic boundary conditions were employed for execution of the simulation. A cut-off radius of 8.0 Å was used for the particle-mesh Ewald (PME) method [21] for long-range electrostatic interaction calculations. For each system, the total simulation time for each case was 100 ns MD simulations with 100 ps time interval.

### 2.8. Codon adaptation, in silico cloning in E. coli expression system and mRNA structure analysis

The Jcat online tool was used to design optimized codons of the vaccine protein for *E. coli* K12 expression system. Prokaryotic ribosome binding site, rho-independent transcription terminators and four restriction sites (*AclI*, *NtoI*, *EagI* and *EgeI*) were avoided in optimized reverse translated sequence. GC content and Codon Adaptation Index (CAI) calculated through the Jcat server were compared and evaluated with the results from GenScript server. Highly efficient cDNA sequence was inserted in pETite N-His SUMO Kan vector (Lucigen, USA) under T7 promoter through SnapGene v5.1 software which would translate the

recombinant protein containing Histidine tags, SUMO protein and vaccine protein. mRNA secondary structure of the fusion protein was predicted through the RNAfold server to check the translation efficiency and thermodynamic stability.

## 3. Results

### 3.1. Selection of significant protein

The NCBI database was used to retrieve amino acid sequences of 21 structural and non-structural proteins. All proteins were evaluated through the VaxiJen server, which produced an overall score for each protein sequence prevailing their antigenicity response. We found an antigenicity score of more than 0.45 for 13 proteins (nsp4, nsp6, nsp9, endoRNase, Spike, ORF3a, Envelop, ORF6, ORF7a, Membrane, ORF8, Nucleocapsid, and ORF10 protein), as shown in [Supplementary Table S2](#). Based on antigenic scores, these proteins were chosen for further design of a multi-epitope vaccine.

### 3.2. Identification and evaluation of CTL (cytotoxic T lymphocytes) epitopes

CTL epitopes of structural and non-structural proteins were initially predicted by a two-stage filtration/screening process. First filtration was conducted after CTL selection by the NetCTL1.2 server, which predicted CTL epitope of 9-mer in length. After identifying CTL epitopes through NetCTL1.2 server, epitopes were filtered by checking antigenicity, allergenicity, non-toxicity and immunogenicity in the first filtration process. Finally, epitopes were evaluated through best binding affinity with mice alleles in the second filtration. About 1582 epitopes were predicted from 13 proteins by NetCTL1.2 server and 109 CTL epitopes were screened out of 1582 in first stage filtration, as shown in [Supplementary File 1](#). Out of 109, 54 CTL epitopes were selected in the second filtration stage, as presented in [Supplementary File 1](#).

### 3.3. Identification and evaluation of HTL (helper T lymphocytes) epitopes

Initially, HTL epitopes for 13 proteins were predicted by the IEDB server for mice MHC-II alleles (H2-IA<sup>b</sup>, H2-IA<sup>d</sup>, H2-IE<sup>b</sup>). Most significant epitopes were selected based on their percentile rank, which should be less than 10. About 703 HTL epitopes were predicted from the IEDB MHC-II server ([Supplementary File 2](#)). Subsequently, these epitopes were screened by following the CTL-first filtration process. However, a total of 120 unique HTL epitopes were predicted for 9 proteins including nsp4, nsp6, nsp9, spike, membrane, envelope, ORF3a, ORF8, and nucleocapsid protein along with their respective mice alleles. Among these 120, after second filtration through IFNepitope server only 59 epitopes were found as IFN- $\gamma$  positive, as provided in [Supplementary File 2](#).

### 3.4. Prediction of B-cell epitopes and discontinuous B-cell epitopes

The most significant linear B-cell epitopes for spike, envelope, membrane, and nucleocapsid proteins were identified by using a multi-method approach, as detailed in the method section. Initially, a pool of linear B cell epitopes was generated from ABCpred server for all four structural proteins. Epitopes with higher antigenicity, non-toxicity and non-allergenicity were shortlisted from the initial pool. In the case of the spike protein, two consensus 14-mer epitopes (S<sub>206-219</sub> and S<sub>374-388</sub>) were finally selected for vaccine construction. IEDB Bepipred 2.0 method predicted epitopes from the residue number 206 to 217 and 369 to 394 susceptible to induce immune responses as linear B-lymphocyte epitopes ([Figure S1a](#)) ([Supplementary File 3](#)). Most residues of these positions were also predicted by other IEDB tools, such as Chou and Fashman B turn prediction ([Figure S1b](#)), Emini surface accessibility prediction method ([Figure S1c](#)), and Karplus and Schulz flexibility



prediction method (Figure S1d). Bcepred and Lbtope webserver also ensured this final two B cell epitopes (S<sub>206-219</sub> and S<sub>374-388</sub>) can act as linear B cell epitope. Furthermore, Ellipro tools in IEDB also confirmed these shortlisted epitopes from ABCpred to be a part of both linear and discontinuous epitopes by using a spike protein crystal structure 6VSB [22], as presented in Supplementary File 3.

In the case of envelope protein, one consensus BCL epitope (E<sub>57-73</sub>) from ABCpred was selected as the final epitope. Initially, the IEDB bepiped prediction method predicted that only one epitope from 57 to 71 residues can induce immune response as linear B-lymphocyte epitope (Supplementary File 3). Additionally, this region was also found most common in other IEDB tools including Chou and Fashman B turn prediction, Emini surface accessibility prediction, and Karplus and Schulz flexibility prediction, as shown in Figure S2. Besides, this residue is common in the Bcepred and Lbtope webserver prediction. Ellipro tools in IEDB also reconfirmed this final epitope from ABCpred to be a part of both linear and discontinuous epitopes by using an envelope protein crystal structure (QHD43418) retrieved from I-TASSER webserver (Supplementary File 3).

Similarly, from membrane protein and N protein two BCL epitopes (M<sub>176-193</sub>, N<sub>91-106</sub>) were selected based on their consensus results in another prediction method as shown in Supplementary File 3, Figure S3 and S4. Thus, we secured final 5 consensus BCL epitopes for vaccine design with higher antigenicity, non-toxicity, and non-allergenicity properties (Supplementary File 3).

### 3.5. Epitope conservancy, final epitopes prediction and population coverage analysis

For effective immunization, epitopes conservancy across other strains around the world is essential. The results of epitope conservancy analysis revealed that CTL, HTL, and B-cell epitopes of the final proteins were 100% conserved, except for one B-cell epitopes of E protein which showed 96% conservancy, as presented in Supplementary File 1, 2 and 3.

After Conservancy analysis, we selected the final CTL and HTL epitopes. We selected total 12 CTL epitopes from each protein by considering their high antigenic and immunogenic score and multiple types of mice allele and human super type coverage (Supplementary File 1). Thereafter, we predicted the MHC-I binding allele of each of the 12 CTL epitopes from IEDB server. The cut off value for the prediction was selected as percentile rank  $\leq 1$ , and according to IEDB recommendations, a percentile rank of  $\leq 1$  was used as the cutoff to predict peptide binders for MHC-I and a percentile rank of  $\leq 10$  was used as the cutoff to predict peptide binders for MHC-II [23].

In the case of CD4<sup>+</sup> T cell epitope prediction, similarly 9 HTL epitopes were selected from each protein by considering their mice allele percentile rank, high antigenic score and maximum number of the HLA-II alleles that are covered by these binders under a percentile rank of 10 (Supplementary File 2).

The population coverage analysis was done for selected 12 CTL and 9 HTL epitopes and their corresponding HLA allele among 13 areas of the world provided in IEDB population coverage tool. We found 87.12% (CTLs) and 90.42% (HTLs) of global population coverage in an individual manner (Supplementary File 4). As CTL and HTL epitopes are contained by a vaccine protein, we focused on their combined coverage which was 98.64% of the global population. The maximum coverage (100%) was found in three countries i.e., Europe, North America, and East Africa. In addition, the average population coverage for each of the CTL epitopes and HTL epitopes can be found in Supplementary File 4.

### 3.6. Vaccine construction and physicochemical analysis

A total 12 CTL, 9 HTL and 5 BCL epitopes based on their antigenic nature, binding energy, and the non-allergenic property were linked together by using GGGG, GPGPG, and KK linkers to construct the final

vaccine. In the N-terminal of vaccine, adjuvant was added to protect it from degradation. Moreover, an EAAK linker was used to join the adjuvant to the CTL epitopes. On the other hand, GGGG, GPGPG, and KK linkers were joined with the CTL, HTL and BCL epitopes, as shown in Fig. 1a. The final vaccine construct consisted of 502 amino acids (Fig. 1b).

The evaluated physicochemical properties of constructed vaccines are represented in the Table S3. The vaccine construct was found to be basic in nature by identifying the Isoelectric point (PI) of 10.06 with 52.29 kDa molecular weight (MW). In addition, half-life was calculated  $>30$ ,  $>20$  and  $>10$  h in mammalian reticulocytes (*in vitro*), in yeast (*in vivo*), and in *E. coli* (*in vivo*) respectively. GRAVY was estimated as 0.316, indicating hydrophobic in nature. Moreover, the vaccine construct was found to be soluble (0.593) in the *E. coli* system during overexpression. Thermo-stability (Aliphatic index) of the vaccine was found to be 89.34, which contain a higher amount of hydrophobic amino acids. Furthermore, the instability index (II) was calculated to be 33.48 classified that protein would be stable in a wet lab. In addition, antigenicity for the vaccine design was 0.6045, as calculated by the Vaxijen server, which indicates that the vaccine is immunogenic and can stimulate sufficient immune response.

### 3.7. Homology modeling, optimization, and structural quality of the vaccine

To analyze the tertiary structure of the vaccine construct, the 502 amino acids peptide sequence was utilized for the prediction of the homology model (Figure S5a). The Robetta server predicted five homology models of the designed vaccine protein based on comparative modeling and *ab initio* method.

The initial homology model of the vaccine construct was optimized by performing 100 ns of MD simulation in physiological environments. The stability of the homology model during the MD simulation was determined by its deviation from the elementary structure in terms of RMSD. The RMSD-C $\alpha$  values of the vaccine protein in the entire MD simulation trajectories was presented in Figure S5b. The homology model of the vaccine protein reached a stable state after 50 ns, while the average RMSD-C $\alpha$  was 8.72 Å.

The MD simulation optimized homology model of the vaccine protein was evaluated through a Ramachandran plot. The results showed that 93% of residues were placed in most favored regions, 6% in additional allowed regions, and 0.3% in disallowed regions respectively, as predicted by RAMPAGE server (Figure S5c). Model quality was also checked by ERRAT and the Verify3D server where the quality of vaccine construct was found 86.62% and 87.45%, respectively.

The secondary structure of the final vaccine construct was predicted by the SOPMA web server, which predicts a secondary structure according to the amino acids sequence of the vaccine protein. Among 502 amino acids, 187 are involved in beta sheet formation and 185 in coil formation, while alpha helix and beta turns are formed by 91 and 39 amino acids, respectively. Overall secondary structure prediction result indicates 37.25% are coiled, 36.85% are beta strands, 18.13% are alpha helix, and 7.77% are beta turn, as shown in Figure S6.

### 3.8. Molecular docking and molecular dynamics of the vaccine with human and mice TLRs

Human TLRs (3,4) and mice TLRs (3,4) were used as receptors for docking analysis with vaccine construct using ClusPro, and the calculated energy scores are tabulated in Table 1. Energy scores attained for human TLR3 and TLR4 were -1400.3 and -1713.7, respectively, and for mice TLR3 and TLR4 were -1436.6 and -1393.4, respectively. These complexes were subjected to MD simulations to analyze their stability.

To evaluate the dynamics of the constructed vaccine complex, MD simulation was performed. The RMSD of alpha-carbon, Rg and RMSF for

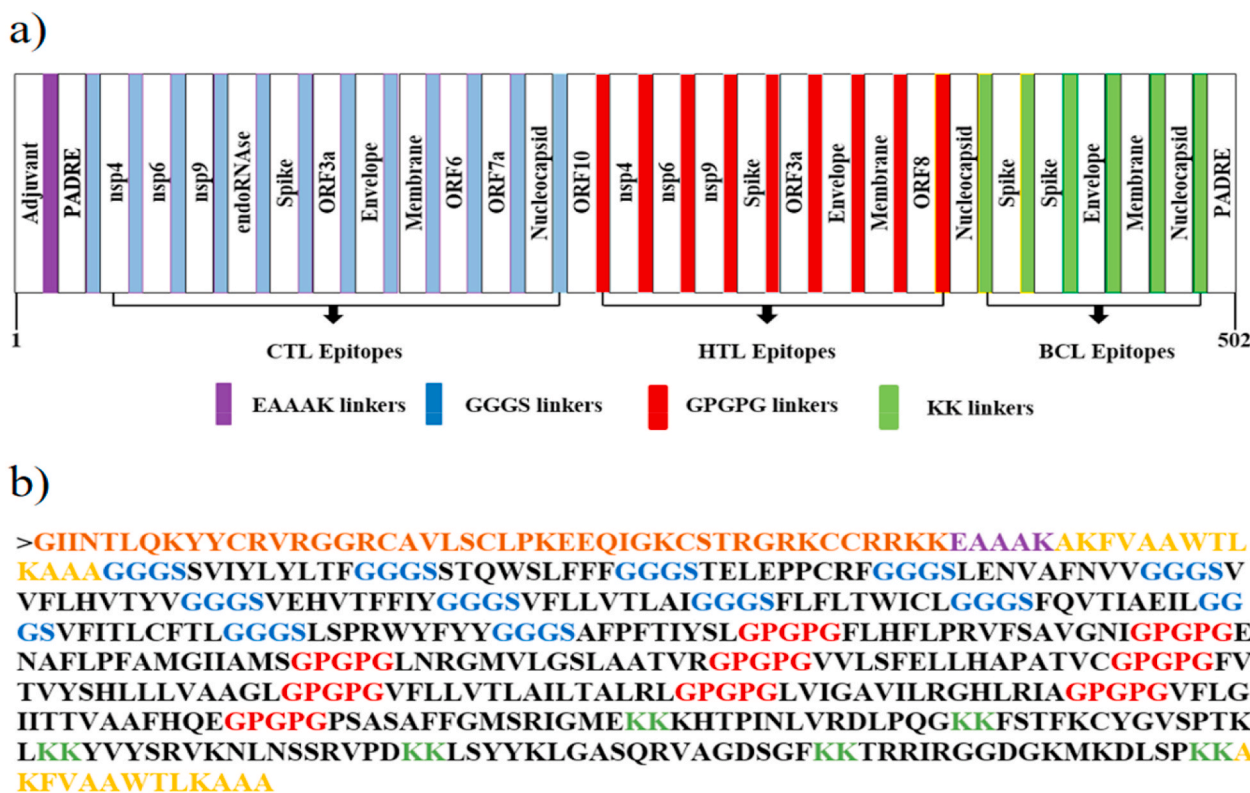


Fig. 1. (a) Schematic view of T and B-cell epitopes in multi-epitope vaccine construct (b) Finally constructed vaccine 502 amino acids long in which residues are attached to Adjuvant, PADRE, EAAAK (linker), GGGG (linker), GPGPG (linker) and KK (linker).

Table 1

Docking score of vaccine construct with TLRs.

TLR name	Lowest Energy
Human TLR3	-1400.3
Human TLR4	-1713.7
Mice TLR3	-1436.6
Mice TLR4	-1393.4

all amino acid residues of the protein were analyzed for 100 ns. In case of human TLR3 complex, RMSD showed an initial deviation of 0.53 Å and gradually increased to 4.11 Å at 50 ns. Between 55 and 85 ns, the fluctuation increased sharply and then stabilized till the end of the simulation, as shown in Fig. 2a. A similar pattern was observed in the case of the mice TLR3 complex; fluctuation was increased sharply within 5–35 ns while the RMSD value range was 0.52–4.40 Å. However, the average deviation was found to be 4.75 Å and 4.61 Å for human and mice TLR3 complexes, respectively. In contrast, there was no sharp deviation observed in human TLR4 and mice TLR4 complexes. After 35 ns to the end the simulation revealed the stable nature of both complexes with mild fluctuation. Investigating the compactness of protein structure during the MD simulation, we analyzed the radius of gyration (Rg). The Rg value showed a wide difference between human TLR4 complex and other three TLR complex structures with an increased trend for human TLR3 and mice TLR (3 and 4) complexes, as shown in Fig. 2b. In case of human TLR4 complex, the lowest Rg value was observed from starting to the end of the simulation, which was ~32.18 Å (average). In contrast, the higher Rg value was found for human TLR3 complex at ~36.23 Å (average), which was followed by mice TLR4 (average ~35.71 Å) then mice TLR3 (average ~35.66 Å). Furthermore, the root-mean-square fluctuation (RMSF) of amino acid residues was analyzed to investigate the stability of the receptor-ligand interaction (Fig. 3). In the RMSF plot, a slight fluctuation was observed in amino acid residues, which may

reflect the incessant interaction between receptor and ligand complex.

As a result, highly fluctuated regions in the plots indicate that the degree of flexibility increased in the receptor-ligand complex.

To gain further insight and understanding of the residue interactions between the docked complexes, we analyzed non-bonding and hydrogen bonding interactions after MD simulations using PDBsum algorithm [24]. The constructed vaccine produced 9 H-bonding interactions with human TLR3 (Fig. 4a and Table S4) and 3 salt bridges and 12 H-bond interactions with mice TLR3 (Fig. 4b and Table S4). On the other hand, 1 salt bridge and 8 H-bond interactions with human TLR4 and 4 salt bridge and 27 H-bond interactions were formed by constructed vaccine, as shown (Fig. 4c, d and Table S4).

### 3.9. Codon optimization for *in silico* cloning in *E. coli* and mRNA secondary structure analysis

Codon Optimization index (CAI) and GC contents of the optimized 1506 bp long nucleotide sequence of vaccine construct was evaluated and compared. To have better expression (transcription and translation) in organisms, the GC content ranging between 30% and 70% is considered optimal, whereas a CAI value must be higher than 0.8 up to 1 [25,26]. GC contents of optimized nucleotide sequence obtained from Jcat and GeneScript servers were 53.25% and 56.79%, respectively, while calculated CAI values were 0.92 and 0.78, respectively. Since Jcat server generated better optimized codons, this nucleotide sequence was therefore used as Gene of Interest (GOI) in SnapGene v 5.1 to construct recombinant pETite N-His SUMO Kan plasmid vector (Lucigen, USA). Stop codon TAA was added at 3' end of GOI followed by *EagI* restriction site (C-terminus) while *AclI* restriction site (N-terminus) was added upstream of GOI followed by the addition of part of SUMO nucleotide sequence (Fig. 5). If expressed in *E. coli* K12, the constructed vector will produce a large fusion protein starting with Methionine (N-terminus) followed by six Histidine tags joined with SUMO protein and vaccine protein (C-terminus). After removal of His tags during downstream

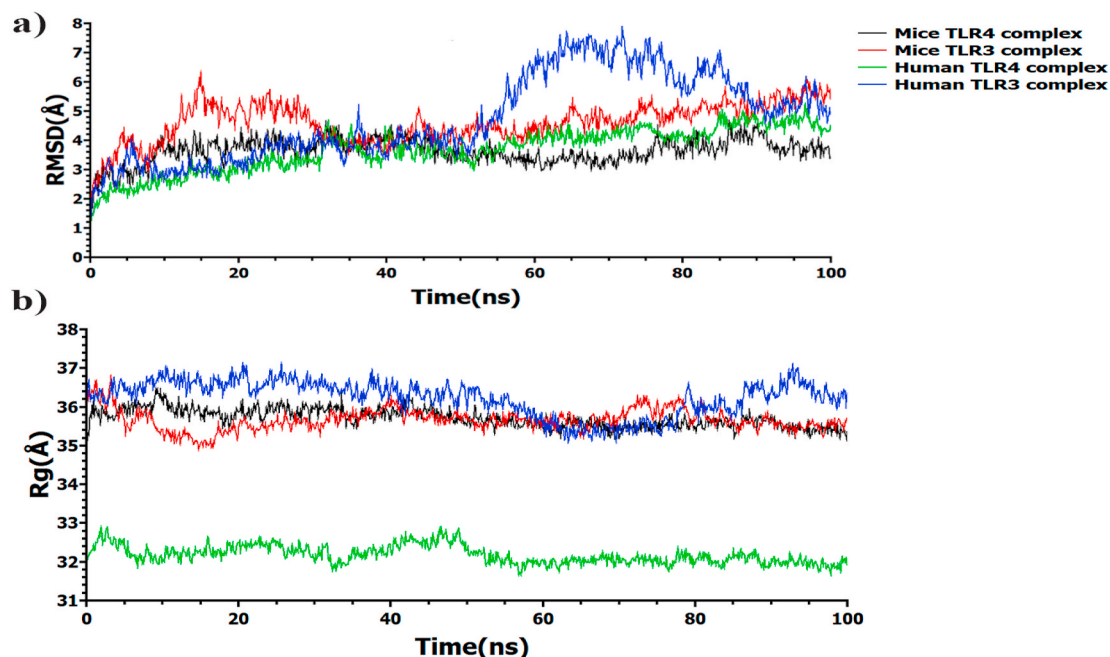


Fig. 2. Molecular dynamics simulation of the ligand-receptor complex (vaccine & TLRs). (a) RMSD-Root Mean Square Deviation shows the stability of docked complexes. (b) Rg plot showing the vaccine construct is stable in its compact form during the simulation time.

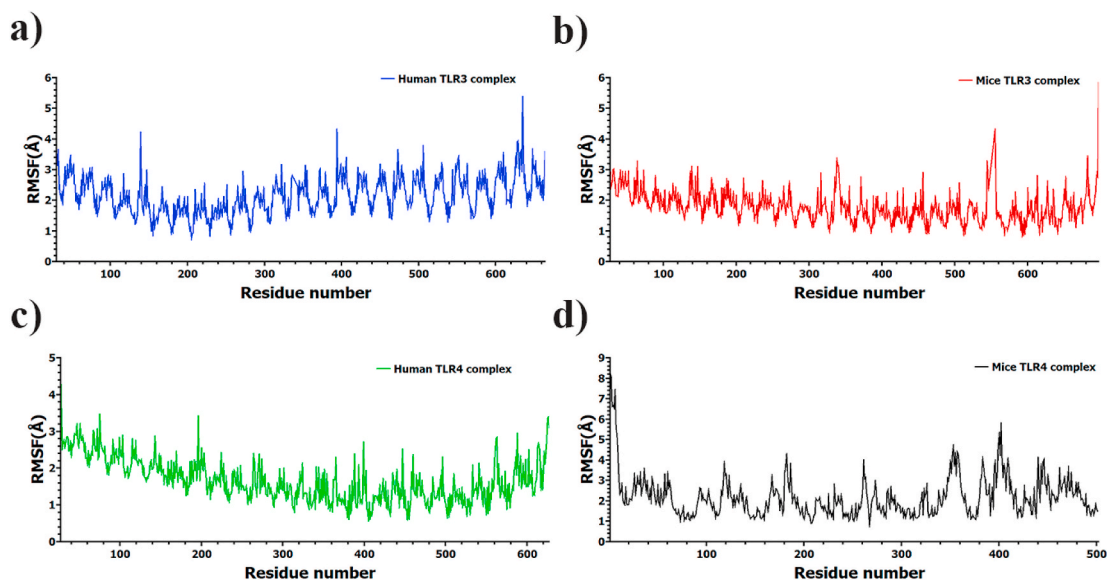


Fig. 3. Root Mean Square Fluctuation (RMSF) plot. Peaks shows the regions with high flexibility.

processes, the native vaccine protein can be obtained by adding SUMO protease which will cleave the SUMO fusion protein and thereby release the exact vaccine protein with native N terminal residue (Glycine of adjuvant) as designed.

Moreover, mRNA structure of expressed genes needs to be stable and the 5' end should not have pseudoknots or long stable hairpin. RNAfold server predicted that the minimum free energy (MFE) of expressed mRNA secondary structure was  $-622.10$  kcal/mol. Lower MFE indicates higher thermodynamic stability of mRNA secondary structure (Figure S7). No pseudoknot or stable hairpin was detected at the 5' end of expressed mRNA, even first 10 nucleotides were free from forming secondary structure. These findings conclude on high level expression of gene construct [27,28] in *E. coli* K12, hence facilitating the production of the vaccine protein.

#### 4. Discussion

Vaccines are considered the most potential immune modalities [29, 30]. The production and development of vaccines is relatively cost effective, yet a laborious and time-consuming process [31,32]. In light of current advances in molecular immunology and computation, researchers are changing conventional vaccinology methods and applying immunoinformatics approaches (reverse vaccinology) to rationally design vaccines, especially for infectious diseases [33]. By using immunoinformatics approach, a vaccine was first designed against serogroup B *Neisseria meningitidis* [34] and successfully produced. Since then, numerous vaccines were developed against such diseases as *Chlamydia pneumoniae*, *Streptococcus pneumoniae* [34], *H. pylori* [35], *Enterotoxigenic Escherichia coli* [36], *Rickettsia prowazekii* [37] based on



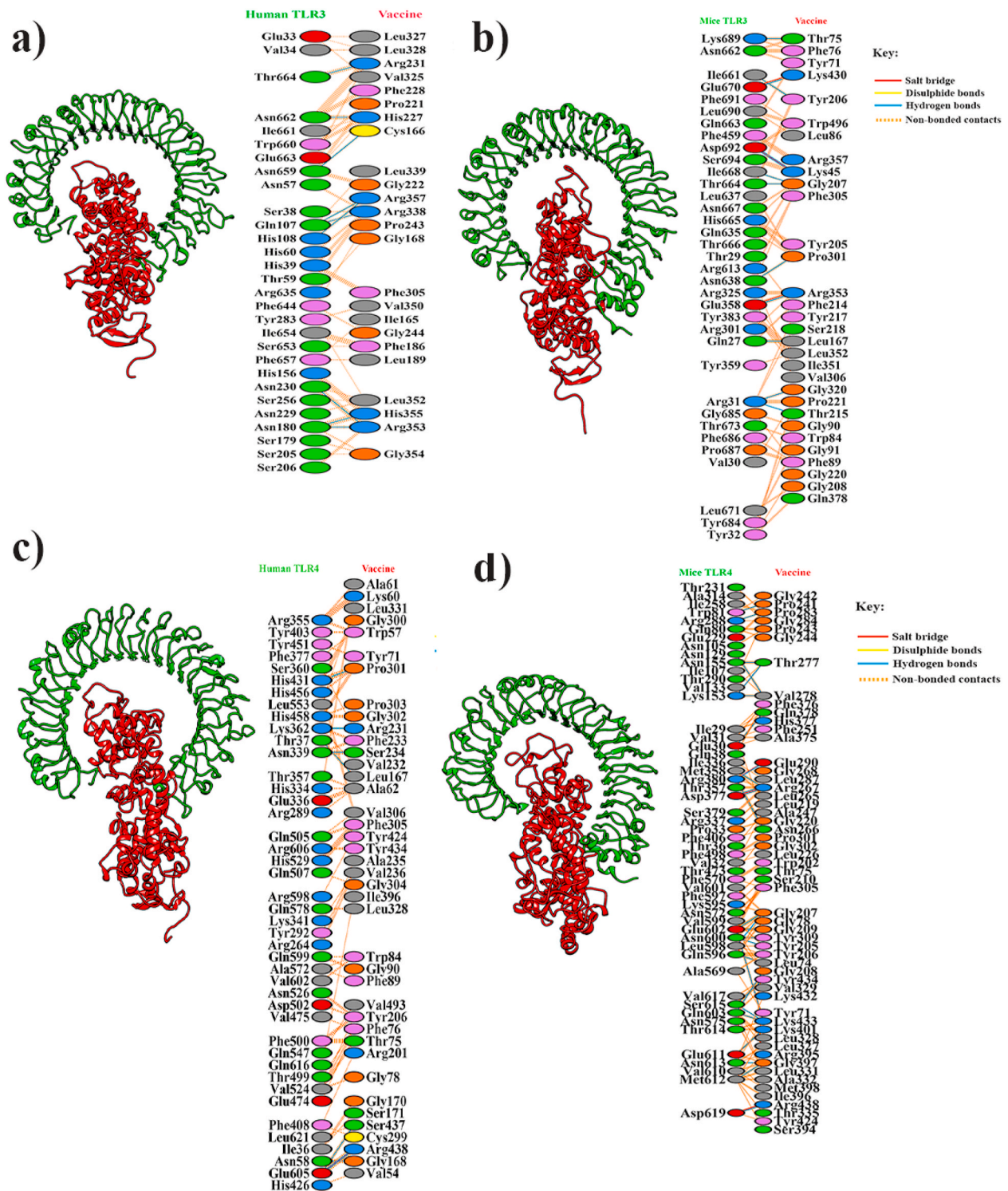


Fig. 4. Vaccine construct-TLRs docked complexes. Data obtained after 100ns molecular dynamics simulation. Interacting residues are illustrated between docked vaccine and TLRs complexes.

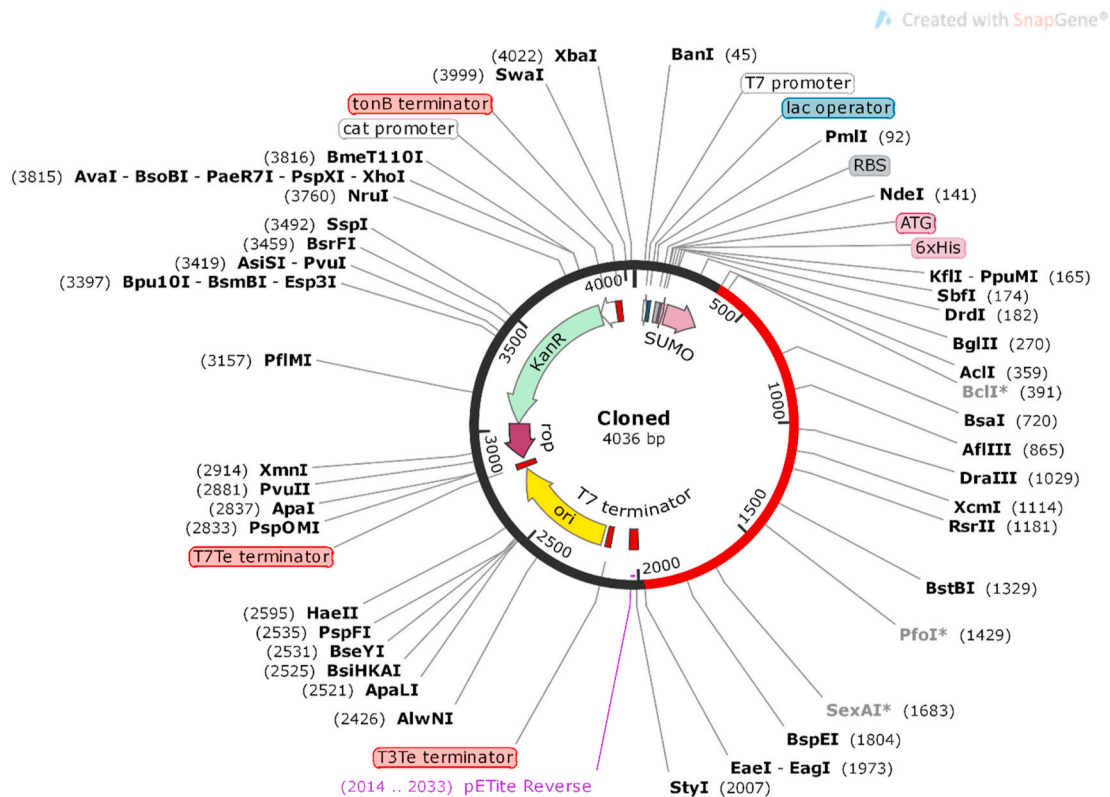
immunoinformatics strategies.

In this study, we designed a multi-epitope vaccine targeting 13 proteins of SARS-CoV-2. Multi-epitope vaccines have several advantages over traditional and single-epitope based vaccines: a) multiple epitopes of MHC Class I and Class II can be recognized by TCRs; b) CTL, HTL, and B cell epitopes can be overlapped, and thus have the capability to induce an immune response of cellular and humoral immunity simultaneously, c) linking of an adjuvant with vaccine epitopes provides long-lasting

immune response, d) the complications of *in vitro* antigen expression can be avoided [38–41].

An ideal multi-epitope vaccine should hold both categories of antigenic epitopes including T-cell and B-cell epitopes to generate particular immune response against predicted antigens [42]. Therefore, T-cell and B-cell epitopes were predicted from 13 proteins of SARS-CoV-2. The selection procedure was maintained based on their immunogenic properties such as antigenicity, toxicity, allergenicity, and cytokine





**Fig. 5.** Optimized codons (1506 base pairs) of vaccine construct inserted in pETite N-His SUMO Kan plasmid vector under T7 promoter for *E. coli* expression. Gene of Interest was inserted between *AclI* and *EagI* at the 5' end.

production. In most cases, we allowed variable epitopes and discarded overlapping epitopes in the vaccine protein. The final multi-epitope vaccine construct comprised of 12 CTL epitopes (nsp4<sub>343-351</sub>, nsp6<sub>28-36</sub>, nsp9<sub>67-75</sub>, nsp15<sub>2-10</sub>, ORF3a<sub>225-232</sub>, ORF6<sub>7-15</sub>, ORF7a<sub>8-16</sub>, ORF10<sub>8-16</sub>, S<sub>1060-1068</sub>, E<sub>25-33</sub>, M<sub>26-34</sub>, and N<sub>105-113</sub>), 9 HTL epitopes (nsp4<sub>118-132</sub>, nsp6<sub>39-53</sub>, nsp9<sub>97-111</sub>, ORF3a<sub>87-101</sub>, ORF8<sub>5-19</sub>, S<sub>511-525</sub>, E<sub>25-39</sub>, M<sub>138-152</sub>, and N<sub>310-324</sub>), and 4 BCL epitopes (S<sub>373-387</sub>, E<sub>57-72</sub>, M<sub>176-193</sub>, and N<sub>91-106</sub>).

The vaccine construct has 502 amino acid residues (52.29 kDa) with adjuvant beta defensin (a 45 mer peptide) [43], PADRE sequence [44] and linkers. The adjuvant sequence is an eminent agonist for TLR which is essential for the activation of different mediators of innate and adaptive immunity, compatible for vaccine development [45,46]. The adjuvant was also joined with EAAAK linker at the N-terminal of the vaccine protein, merged for effective separation of a bi-functional fusion protein [47,48]. The EAAAK linker was joined with PADRE sequence which is one of the helper T-cell epitope, significant for enhancing CTL responses of different antigens [49]. Moreover, GGG [50] and GPGPG [51] linkers were merged with CTL and HTL epitopes for efficient recognition and separation of the multi-epitope vaccine [48]. Finally, the BCL epitopes were linked with a KK linker to preserve their distinct immunogenic activity [52]. However, the effectiveness of the vaccine depends on the broad population coverage in which the vaccination is used [53]. The epitope conservancy analysis was conducted with 51 different SARS-CoV-2 strains. Notably, the construct showed 98.64% of global population coverage in pathogenic invaded province (stated earlier). Therefore, this constructed vaccine would cover majority of the global population.

The final vaccine construct was found to be basic in nature and stable in physiological pH range as predicted by physicochemical analysis. Moreover, the estimated aliphatic index score revealed that the vaccine is more thermostable, while the positive GRAVY indicated its hydrophobicity suggesting strong interactions with amino acid residues. In

addition, the construct was found to be soluble during overexpression in *E. coli* system.

For effective transportation into the body, the vaccine construct should have strong binding affinity with Toll-like receptors (TLRs). Although, TLRs play a significant role in the identification of pathogens and installation of the innate immune response against viruses, TLR recognition of SARS-CoV-1 is not well characterized [54]. In this study, we first assessed docking analysis of mice TLRs (3 and 4) against vaccine construct and compared it with human TLRs (3 and 4) respectively. One study revealed that mice TLR3 and TLR4 were more susceptible to SARS-CoV-1 [54]. Recently, a vaccine clinical trial was done with mice TLR4 against S protein of MERS-CoV. They also developed the standard operating procedures for the rapid development of clinic grade MNA SARS-CoV-2 subunit vaccines [55]. One more rationale for the selection of TLR3 is that it could recognize the RNA or dsRNA genome of coronavirus and TLR4 might recognize the outer component of coronavirus the spike protein [56]. The docking scores revealed that the vaccine construct has high binding interaction with human and mice TLRs, thus suggesting that the vaccine can trigger TLR activation and can enhance immune responses against SARS-CoV-2.

Molecular dynamics studies can provide information on the stability of vaccine-TLR complexes. In the case of the RMSD results, slight fluctuations were observed at 55–85 ns for vaccine-human TLR3 complex, however, all vaccine-receptor complexes were highly stable throughout the 100 ns simulation time. These results are in line with the structural compactness and high degree of vaccine-receptor complex flexibility which were analyzed by Rg and RMSF.

Furthermore, the intermolecular protein-protein interactions between vaccine and TLRs were determined via PDBSum (Fig. 4 and Table S4). Structural analysis showed that vaccine protein formed higher number H-bond and salt bridge interaction with TLR4 than TLR3. In the case of human and mice TLR3, human TLR3 formed 9 and mice TLR3 formed 12 H-bond interactions with vaccine protein, although 3

salt bridge interactions were present in mice TLR3 complex, but this type of interactions was absent in human TLR3 complex. On the other hand, human TLR4 formed 8 and mice TLR4 formed 27 H-bond interactions with the vaccine protein. Although, salt bridge interactions were produced in both complexes, a common salt bridge interaction was found with a residue Arg438 of vaccine construct. Among all non-covalent interactions, salt bridges are the strongest interaction and contribute to a major extent in biomolecular stability [57]. *In silico* cloning and respective mRNA secondary structure analysis revealed that efficient gene expression and translation initiation can be achieved in wet lab trials.

In our study, certain epitopes of SARS-CoV-2 were dominant for T-cell and B-cell responses and these epitopes were almost well conserved in terms of sequence with SARS-CoV-1. Among them, of particular interest is the SARS-CoV-2 S<sub>1060-1068</sub> CTL epitope found to have 100% conservation in SARS-CoV-1 S<sub>1042-1050</sub> that could induce recall response when IFN- $\gamma$  stimulation of blood CD8<sup>+</sup> T-cells, revealing a significant difference between normal healthy donors and SARS-recovered patients [58]. Moreover, SARS-CoV-2 S<sub>373-387</sub> BCL epitope is located in the receptor binding domain (RBD) which has higher binding affinity for angiotensin-converting enzyme 2 (ACE2) of human respiratory epithelial cell [59]. This epitope is >90% conserved in SARS-CoV-1 S<sub>361-374</sub>. Zhi et al. (2004) reported that FSTFKCYGVSATKLN epitope corresponding to SARS-CoV-1 S<sub>361-375</sub> and CYGVSATKLNLCFS epitope corresponding to SARS-CoV-1 S<sub>366-380</sub> was identified as the major positive peptides responsible for specific stimulation of T-cells to produce IFN- $\gamma$ . They also found that CYGVSATKL (SARS-CoV-1 S<sub>366-374</sub>) has strong binding affinity for mice H-2-K<sup>d</sup> [60]. Another group reported the frequencies of IFN- $\gamma$  producing cells elicited by CD8<sup>+</sup> T cell epitope KCYGVSATKL (SARS-CoV-1 S<sub>365-374</sub>) [61]. One region from SARS-CoV-2 E<sub>57-72</sub> (YVYSRVKLNLSRVPD) protein was partially nesting in the region E<sub>55-70</sub> (TVYYSRVKLNLSNSEG), which was recognized as having high neutralizing activity in a SARS-CoV-1 convalescent sera and found to have the capability to elicit marked IgG, IgA, and IgM responses [62].

Moreover, two dominant epitopes correspond to 26–34 and 138–152 from SARS-CoV-2 M protein sharing high homology with SARS-CoV-1 M<sub>25-33</sub>, 137-151. One research group recognized SARS-CoV-1 M<sub>21-44</sub> as immunodominant peptide that could elicit cellular immunity with a predominance of CD8<sup>+</sup> T-cell response. Wang et al. (2003) reported that M<sub>137-158</sub> peptide had the significant affinity for forming peptide-antibody complexes with SARS serum [63]. Three regions were identified from SARS-CoV-2 N protein (91–106, 104–112, and 309–323) which were highly homologous with SARS-CoV-1. Immunization of mice with synthetic peptide (SARS-CoV-1 N<sub>91-105</sub>) containing appropriate helper T-cell epitopes elicited strong immunity *in vivo* [64]. Peng et al. (2006) suggested the major immunodominant epitopes corresponding to the sequences of residues 293–376 in SARS-CoV-1 N protein. They also conducted an experiment on SARS-CoV-1 recovered donors and figured out SARS-CoV-1 N<sub>307-323</sub> peptide could have positive response in production of IFN- $\gamma$  [65]. Besides, one study reported that the isolated CD8<sup>+</sup> T-cells were stimulated with SARS-CoV-1 N<sub>317-325</sub> peptide produced high amounts of IFN- $\gamma$ <sup>+</sup> both in normal healthy and full recovered SARS-CoV-1 infected patients [58].

Currently, there are no effective drugs or vaccines established for the treatment of COVID-19, although vaccine development is a time-consuming process that involves experimental to clinical trial settings, and no vaccine candidate has completed clinical trials yet [7,66–70]. The majority of the vaccine designed against SARS-CoV-2 are mainly focused on a single protein (such as S, E, M, and N) [71–73], but single proteins such as the spike protein-based vaccine design still has many obstacles which may induce harmful immune response, causing liver damage in vaccinated animals [74]. Besides, another study designed a vaccine based on M protein, but their vaccine construct showed lower antigenic response to the virus [73]. For a potential vaccine candidate, structural, non-structural and accessory proteins of the virus are suitable

selections which induce protective immune responses [75–77]. There are many studies that propose epitope and multi-epitope based vaccines against SARS-CoV-2 via the application of various computational and immunoinformatics approaches [78–86]. After scanning the literature, we found most of the published studies concerning epitope prediction on the S protein [71,87,88]. Furthermore, we compared our findings with reported studies manually, and found that none of the predicted epitopes have been projected earlier, which indicates the novelty of our study. In our study, the selected best epitopes have a good immune response to mice alleles besides human alleles, and also vaccine-mice TLRs complexes remain stable during the MD simulation. This type of study has not appeared in any current studies. Asaf et al. and Srivastava et al. recently reported that computational identification of multi-epitopes for CD4<sup>+</sup> and CD8<sup>+</sup> T cell based on multi proteins of SARS-CoV-2. But they only predicted T cell epitopes, a B cell peptide was not predicted [89,90]. Bhattacharya et al. predicted MHC-I and MHC-II epitopes based on S protein, but did not predict the capability of producing IFN- $\gamma$  [15]. A report from Feng et al. outlined the identification of MHC-I epitopes in the main structural proteins of SARS-CoV-2, but did not consider any accessory and non-structural proteins [91]. A comprehensive analysis was conducted by Grifoni and colleagues, in which they identified all SARS-CoV-2 protein CD8<sup>+</sup> T cell epitopes, but their main focus was on the peptides with sequence homologous to known SARS-CoV-1 epitopes [92]. Though a research group has recently conducted a study like us, it never considered accessory proteins for probable vaccine construction. Rather, they have designed a multi-epitope vaccine from four proteins (Spike, Mpro, RdRp, and Nsp13) of SARS-CoV-2 which showed a lower antigenicity score (0.4259) [93] when compared to our study.

## 5. Conclusion

In this study, we designed a unique multi-epitope vaccine against SARS-CoV-2 by using various computational tools. Our vaccine construct showed satisfactory physiochemical properties, antigenicity, non-allergenicity, anti-toxicity, and immunogenicity. Molecular docking analysis followed by 100 ns MD simulation of the vaccine construct with TLR3 and TLR4 was executed, which showed stable interactions of the vaccine candidate with these receptors. However, the vaccine construct showed strong expression, as confirmed by computational runs, and further experimental validation is required to ensure vaccine construct efficacy against COVID-19.

## Author contribution

Conceptualization, M.A.H., M.T.K., A.P.; Protocol designing, M.A.H., M.T.K., M.J.I., R.I.; Immunoinformatics and molecular docking, M.T.K., A.P., R.I., T.J.J., R.D., F.K.L., M.J.I.; MD simulation, M.A.A.; MD simulation data analysis, M.A.A., M.J.I., M.A.H.; Manuscript writing, M.T.K., M.J.I., A.P., Supervised the work, M.A.H.; Revisions, R.I., M.J.I., M.T.K. M.A.H.; All authors approved the final manuscript.

## Declaration of competing interest

The authors declare that they have no known competing interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

We are grateful to our donors for support in building our computational platform (<http://grc-bd.org/donate/>). The authors acknowledge the World Academy of Science (TWAS) for the High-Performance Computer for the molecular dynamics simulations. The authors also acknowledge Dr. Ian Hosein for his careful revision of this manuscript.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.imu.2021.100578>.

## References

- Kuiken T, Fouchier RAM, Schutten M, Rimmelzwaan GF, Van Amerongen G, Van Riel D, et al. Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet* 2003;362:263–70. [https://doi.org/10.1016/S0140-6736\(03\)13967-0](https://doi.org/10.1016/S0140-6736(03)13967-0).
- Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 2020;1–13. <https://doi.org/10.1038/s41586-020-2286-9>.
- Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020;181:281–92. <https://doi.org/10.1016/j.cell.2020.02.058>.
- Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 2020;581:215–20. <https://doi.org/10.1038/s41586-020-2180-5>.
- Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020;395:507–13. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7).
- Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020;395:565–74. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8).
- Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med* 2020;383:2603–15. <https://doi.org/10.1056/NEJMoa2034577>.
- Pfizer-BioNTech COVID-19 vaccine | [covidvaccine.com.n.d](http://covidvaccine.com.n.d). <https://www.covidvaccine.com/> [accessed 15.02.21].
- Tu Y-F, Chien C-S, Yarmishyn AA, Lin Y-Y, Luo Y-H, Lin Y-T, et al. A review of SARS-CoV-2 and the ongoing clinical trials. *Int J Mol Sci* 2020;21:2657. <https://doi.org/10.3390/ijms21072657>.
- Sakharkar KR, Sakharkar MK, Chandra R. Post-genomic approaches in drug and vaccine development. 2015.
- Adar Y, Singer Y, Levi R, Tzehoval E, Perk S, Banet-Noach C, et al. A universal epitope-based influenza vaccine and its efficacy against H5N1. *Vaccine* 2009. <https://doi.org/10.1016/j.vaccine.2009.02.011>.
- Gottlieb T, Ben-Yedidia T. Epitope-based approaches to a universal influenza vaccine. *J Autoimmun* 2014. <https://doi.org/10.1016/j.jaut.2014.07.005>.
- Amanat F, Krammer F. SARS-CoV-2 vaccines: status report. *Immunity* 2020;52:583–9. <https://doi.org/10.1016/j.immuni.2020.03.007>.
- Ismail S, Ahmad S, Azam SS. Immuno-informatics characterization SARS-CoV-2 spike glycoprotein for prioritization of epitope based multivalent peptide vaccine. *BioRxiv* 2020. <https://doi.org/10.1101/2020.04.05.026005>.
- Bhattacharya M, Sharma AR, Patra P, Ghosh P, Sharma G, Patra BC, et al. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): immunoinformatics approach. *J Med Virol* 2020;92:618–31. <https://doi.org/10.1002/jmv.25736>.
- Panda PK, Arul MN, Patel P, Verma SK, Luo H-G W, Rubahn YK, et al. Structure-based drug designing and immunoinformatics approach for SARS-CoV-2. *Sci Adv* 2020;6:1–38. <https://doi.org/10.1126/sciadv.abb8097>.
- Kalita P, Padhi AK, Zhang KYJ, Tripathi T. Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2. *Microb Pathog* 2020;145:104236. <https://doi.org/10.1016/j.micpath.2020.104236>.
- Geourjon C, Deléage G. Soppa: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* 1995;11:681–4. <https://doi.org/10.1093/bioinformatics/11.6.681>.
- Krieger E, Vriend G. New ways to boost molecular dynamics simulations. *J Comput Chem* 2015;36:996–1007. <https://doi.org/10.1002/jcc.23899>.
- Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The Amber biomolecular simulation programs. *J Comput Chem* 2005;26:1668–88. <https://doi.org/10.1002/jcc.20290>.
- Darden T, York D, Pedersen L. Particle mesh Ewald: an  $N \cdot \log(N)$  method for Ewald sums in large systems. *J Chem Phys* 1993;98:10089–92. <https://doi.org/10.1063/1.464397>.
- Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020;367(80):1260–3. <https://doi.org/10.1126/science.aax0902>.
- Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, et al. Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res* 2008;36:W513–8. <https://doi.org/10.1093/nar/gkn254>.
- Laskowski RA. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res* 2001;29:221–2. <https://doi.org/10.1093/nar/29.1.221>.
- Morla S, Makhija A, Kumar S. Synonymous codon usage pattern in glycoprotein gene of rabies virus. *Gene* 2016;584:1–6. <https://doi.org/10.1016/j.gene.2016.02.047>.
- Shey RA, Ghogomu SM, Eshok KK, Nembang ND, Shintouo CM, Nongley NF, et al. In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Sci Rep* 2019;9. <https://doi.org/10.1038/s41598-019-40833-x>.
- Wu X, Jörnvall H, Berndt KD, Oppermann U. Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance. *Biochem Biophys Res Commun* 2004;313:89–96. <https://doi.org/10.1016/j.bbrc.2003.11.091>.
- Gonzalez de Valdivia EI, Isaksson LA. Abortive translation caused by peptidyl-tRNA drop-off at NGG codons in the early coding region of mRNA. *FEBS J* 2005. <https://doi.org/10.1111/j.1742-4658.2005.04926.x>.
- Nain Z, Karim MM, Sen MK, Adhikari UK. Structural basis and designing of peptide vaccine using PE-PGRS family protein of *Mycobacterium ulcerans*—an integrated vaccinomics approach. *Mol Immunol* 2020;120:146–63. <https://doi.org/10.1016/j.molimm.2020.02.009>.
- Nezafat N, Karimi Z, Eslami M, Mohkam M, Zandian S, Ghasemi Y. Designing an efficient multi-epitope peptide vaccine against *Vibrio cholerae* via combined immunoinformatics and protein interaction based approaches. *Comput Biol Chem* 2016;62:82–95. <https://doi.org/10.1016/j.compbiocchem.2016.04.006>.
- Kazi A, Chuah C, Majeed ABA, Leow CH, Lim BH, Leow CY. Current progress of immunoinformatics approach harnessed for cellular- and antibody-dependent vaccine design. *Pathog Glob Health* 2018;112:123–31. <https://doi.org/10.1080/20477724.2018.1446773>.
- Plotkin S, Robinson JM, Cunningham G, Iqbal R, Larsen S. The complexity and cost of vaccine manufacturing – an overview. *Vaccine* 2017;35:4064–71. <https://doi.org/10.1016/j.vaccine.2017.06.003>.
- Six A, Bellier B, Thomas-Vaslin V, Klatzmann D. Systems biology in vaccine design. *Microb Biotechnol* 2012;5:295–304. <https://doi.org/10.1111/j.1751-7915.2011.00321.x>.
- Jeannette A-B, Barbara C, Davide S, Rino R, Mariagrazia P. Two years into reverse vaccinology. *Vaccine* 2003;605–10. [https://doi.org/10.1016/s0264-410x\(02\)00566-2](https://doi.org/10.1016/s0264-410x(02)00566-2).
- Naz A, Awan FM, Obaid A, Muhammad SA, Paracha RZ, Ahmad J, et al. Identification of putative vaccine candidates against *Helicobacter pylori* exploiting exoproteome and secretome: a reverse vaccinology based approach. *Infect Genet Evol* 2015;32:280–91. <https://doi.org/10.1016/j.meegid.2015.03.027>.
- Mehla K, Ramana J. Identification of epitope-based peptide vaccine candidates against enterotoxigenic *Escherichia coli*: a comparative genomics and immunoinformatics approach. *Mol Biosyst* 2016;12:890–901. <https://doi.org/10.1039/c5mb00745c>.
- Caro-Gomez E, Gazi M, Goez Y, Valbuena G. Discovery of novel cross-protective *Rickettsia prowazekii* T-cell antigens using a combined reverse vaccinology and in vivo screening approach. *Vaccine* 2014;32:4968–76. <https://doi.org/10.1016/j.vaccine.2014.06.089>.
- Lu C, Meng S, Jin Y, Zhang W, Li Z, Wang F, et al. A novel multi-epitope vaccine from MMSA-1 and DKK1 for multiple myeloma immunotherapy. *Br J Haematol* 2017;178:413–26. <https://doi.org/10.1111/bjh.14686>.
- He R, Yang X, Liu C, Chen X, Wang L, Xiao M, et al. Efficient control of chronic LCMV infection by a CD4 T cell epitope-based heterologous prime-boost vaccination in a murine model. *Cell Mol Immunol* 2018;15:815–26. <https://doi.org/10.1038/cmi.2017.3>.
- Jiang P, Cai Y, Chen J, Ye X, Mao S, Zhu S, et al. Evaluation of tandem *Chlamydia trachomatis* MOMP multi-epitopes vaccine in BALB/c mice model. *Vaccine* 2017;35:3096–103. <https://doi.org/10.1016/j.vaccine.2017.04.031>.
- Lu IN, Farinelle S, Sausy A, Muller CP. Identification of a CD4 T-cell epitope in the hemagglutinin stalk domain of pandemic H1N1 influenza virus and its antigen-driven TCR usage signature in BALB/c mice. *Cell Mol Immunol* 2017;14:511–20. <https://doi.org/10.1038/cmi.2016.20>.
- Zhang L. Multi-epitope vaccines: a promising strategy against tumors and viral infections. *Cell Mol Immunol* 2018;15:182–4. <https://doi.org/10.1038/cmi.2017.92>.
- Ali Mudassar, Pandey Rajan Kumar, Khatoon Nazia, Narula Aruna, Mishra Amit, Prajapati Vijay Kumar. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci. Rep.* 2017;7(1):1–13. <https://doi.org/10.1038/s41598-017-09199-w>.
- Wu CY, Monie A, Pang X, Hung CF, Wu TC. Improving therapeutic HPV peptide-based vaccine potency by enhancing CD4+ T help and dendritic cell activation. *J Biomed Sci* 2010;17. <https://doi.org/10.1186/1423-0127-17-88>.
- Narula A, Pandey RK, Khatoon N, Mishra A, Prajapati VK. Excavating chikungunya genome to design B and T cell multi-epitope subunit vaccine using comprehensive immunoinformatics approach to control chikungunya infection. *Infect Genet Evol* 2018;61:4–15. <https://doi.org/10.1016/j.meegid.2018.03.007>.
- Ali M, Pandey RK, Khatoon N, Narula A, Mishra A, Prajapati VK. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci Rep* 2017;7. <https://doi.org/10.1038/s41598-017-09199-w>.
- Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev* 2013;65:1357–69. <https://doi.org/10.1016/j.addr.2012.09.039>.
- Hajjigharamani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatbadi SS, Ghasemi Y. Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*. *Infect Genet Evol* 2017;48:83–94. <https://doi.org/10.1016/j.meegid.2016.12.010>.
- Bahrami AA, Payandeh Z, Khalili S, Zakeri A, Bandehpour M. Immunoinformatics: in silico approaches and computational design of a multi-epitope, immunogenic protein. *Int Rev Immunol* 2019. <https://doi.org/10.1080/08830185.2019.1657426>.



- [50] Yang Y, Sun W, Guo J, Zhao G, Sun S, Yu H, et al. In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations. *Hum Vaccines Immunother* 2015;11:795–805. <https://doi.org/10.1080/21645515.2015.1012017>.
- [51] Negahdaripour M, Nezafat N, Eslami M, Ghoshoon MB, Shoolian E, Najafipour S, et al. Structural vaccinology considerations for in silico designing of a multi-epitope vaccine. *Infect Genet Evol* 2018;58:96–109. <https://doi.org/10.1016/j.meegid.2017.12.008>.
- [52] Gu Y, Sun X, Li B, Huang J, Zhan B, Zhu X. Vaccination with a paramyosin-based multi-epitope vaccine elicits significant protective immunity against *Trichinella spiralis* infection in mice. *Front Microbiol* 2017;8. <https://doi.org/10.3389/fmicb.2017.01475>.
- [53] Phillips RO, Phanuz DM, Beissner M, Badziklou K, Luzolo EK, Sarfo FS, et al. Effectiveness of routine BCG vaccination on Buruli ulcer disease: a case-control study in the Democratic Republic of Congo, Ghana and Togo. *PLoS Neglected Trop Dis* 2015;9. <https://doi.org/10.1371/journal.pntd.0003457>.
- [54] Totura AL, Whitmore A, Agnihothram S, Schäfer A, Katze MG, Heise MT, et al. Toll-like receptor 3 signaling via TRIF contributes to a protective innate immune response to severe acute respiratory syndrome coronavirus infection. *mBio* 2015;6: 1–14. <https://doi.org/10.1128/mBio.00638-15>.
- [55] Kim E, Erdos G, Huang S, Kenniston TW, Balmert SC, Carey CD, et al. Microneedle array delivered recombinant coronavirus vaccines: immunogenicity and rapid translational development. *EBioMedicine* 2020;55. <https://doi.org/10.1016/j.ebiom.2020.102743>.
- [56] Astuti I, Ysrafil. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): an overview of viral structure and host response. *Diabetes Metab Syndr Clin Res Rev* 2020;14:407–12. <https://doi.org/10.1016/j.dsx.2020.04.020>.
- [57] Bosshard HR, Marti DN, Jelesarov I. Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *J Mol Recogn* 2004;17:1–16. <https://doi.org/10.1002/jmr.657>.
- [58] Tsao YP, Lin JY, Jan JT, Leng CH, Chu CC, Yang YC, et al. HLA-A\*0201 T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus nucleocapsid and spike proteins. *Biochem Biophys Res Commun* 2006;344:63–71. <https://doi.org/10.1016/j.bbrc.2006.03.152>.
- [59] Loganathan SK, Schleicher K, Malik A, Quevedo R, Langille E, Teng K, et al. Rare driver mutations in head and neck squamous cell carcinomas converge on NOTCH signaling. *Science* 2020;367(80):1264–9. <https://doi.org/10.1126/science.aax0902>.
- [60] Zhi Y, Kobinger GP, Jordan H, Suchma K, Weiss SR, Shen H, et al. Identification of murine CD8 T cell epitopes in codon-optimized SARS-associated coronavirus spike protein. *Virology* 2005;335:34–45. <https://doi.org/10.1016/j.virol.2005.01.050>.
- [61] Huang J, Cao Y, Du J, Bu X, Ma R, Wu C. Priming with SARS CoV S DNA and boosting with SARS CoV S epitopes specific for CD4+ and CD8+ T cells promote cellular immune responses. *Vaccine* 2007;25:6981–91. <https://doi.org/10.1016/j.vaccine.2007.06.047>.
- [62] Guo JP, Petric M, Campbell W, McGeer PL. SARS corona virus peptides recognized by antibodies in the sera of convalescent cases. *Virology* 2004;324:251–6. <https://doi.org/10.1016/j.virol.2004.04.017>.
- [63] Wang J, Wen J, Li J, Yin J, Zhu Q, Wang H, et al. Assessment of immunoreactive synthetic peptides from the structural proteins of severe acute respiratory syndrome coronavirus. *Clin Chem* 2003;49:1989–96. <https://doi.org/10.1373/clinchem.2003.023184>.
- [64] Zhao J, Huang Q, Wang W, Zhang Y, Lv P, Gao X-M. Identification and characterization of dominant helper T-cell epitopes in the nucleocapsid protein of severe acute respiratory syndrome coronavirus. *J Virol* 2007;81:6079–88. <https://doi.org/10.1128/jvi.02568-06>.
- [65] Peng H, Yang Li-tao, Wang Ling-yun, Li J, Huang J, Lu Zhi-qiang, et al. Long-lived memory T lymphocyte responses against SARS coronavirus nucleocapsid protein in SARS-recovered patients. *Virology* 2006;351:466–75. <https://doi.org/10.1016/j.virol.2006.03.036>.
- [66] Zhu FC, Li YH, Guan XH, Hou LH, Wang WJ, Li JX, et al. Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored COVID-19 vaccine: a dose-escalation, open-label, non-randomised, first-in-human trial. *Lancet* 2020; 395:1845–54. [https://doi.org/10.1016/S0140-6736\(20\)31208-3](https://doi.org/10.1016/S0140-6736(20)31208-3).
- [67] Folegatti PM, Ewer KJ, Aley PK, Angus B, Becker S, Belij-Rammerstorfer S, et al. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet* 2020;396:467–78. [https://doi.org/10.1016/S0140-6736\(20\)31604-4](https://doi.org/10.1016/S0140-6736(20)31604-4).
- [68] Walsh EE, Frenck R, Falsey AR, Kitchin N, Absalon J, Gurtman A, et al. RNA-based COVID-19 vaccine BNT162b2 selected for a pivotal efficacy study. *MedRxiv* 2020. <https://doi.org/10.1101/2020.08.17.20176651>.
- [69] Smith TRF, Patel A, Ramos S, Elwood D, Zhu X, Yan J, et al. Immunogenicity of a DNA vaccine candidate for COVID-19. *Nat Commun* 2020;11:1–13. <https://doi.org/10.1038/s41467-020-16505-0>.
- [70] Wang F, Kream RM, Stefano GB. An evidence based perspective on mRNA-SARScov-2 vaccine development. *Med Sci Mon Int Med J Exp Clin Res* 2020;26. <https://doi.org/10.12659/MSM.924700>.
- [71] Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses* 2020;12. <https://doi.org/10.3390/v12030254>.
- [72] Abdelmagede MI, Abdelmoneim AH, Mustafa MI, Elfadol NM, Murshed NS, Shantier SW, et al. Design of a multi-epitope-based peptide vaccine against the e protein of human COVID-19: an immunoinformatics approach. *BioMed Res Int* 2020;2020. <https://doi.org/10.1155/2020/2683286>.
- [73] Ayyagari VS, Venkateswarulu TC, Abraham Peele K, Srirama K. Design of a multi-epitope-based vaccine targeting M-protein of SARS-CoV2: an immunoinformatics approach. *J Biomol Struct Dyn* 2020. <https://doi.org/10.1080/07391102.2020.1850357>.
- [74] Czub M, Weingartl H, Czub S, He R, Cao J. Evaluation of modified vaccinia virus Ankara based recombinant SARS vaccine in ferrets. *Vaccine* 2005;23:2273–9. <https://doi.org/10.1016/j.vaccine.2005.01.033>. Elsevier BV.
- [75] Sedeyn K, Schepens B, Saelens X. Respiratory syncytial virus nonstructural proteins 1 and 2: exceptional disrupters of innate immune responses. *PLoS Pathog* 2019;15: e1007984. <https://doi.org/10.1371/journal.ppat.1007984>.
- [76] Liu W, Lu Y, Chen Y. Bioinformatics analysis of SARS-Cov M protein provides information for vaccine development\*. *Prog Nat Sci* 2003;13:844–7. <https://doi.org/10.1080/10020070312331344530>.
- [77] Kern DM, Sorum B, Hoel CM, Sridharan S, Remis JP, Toso DB, et al. Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs. *BioRxiv* 2020. <https://doi.org/10.1101/2020.06.17.156554>.
- [78] Yazdani Z, Rafiei A, Yazdani M, Valadan R. Design an efficient multi-epitope peptide vaccine candidate against SARS-CoV-2: an in silico analysis. *Infect Drug Resist* 2020;13:3007–22. <https://doi.org/10.2147/IDR.S264573>.
- [79] Nerli S, Sgourakis NG. Structure-based modeling of SARS-CoV-2 peptide/HLA-A02 antigens. *Front Med Technol* 2020;2:553478. <https://doi.org/10.3389/fmed.2020.553478>.
- [80] Liu G, Carter B, Bricken T, Jain S, Viard M, Carrington M, et al. Computationally optimized SARS-CoV-2 MHC class I and II vaccine formulations predicted to target human haplotype distributions. *Cell Syst* 2020;11:131–44. <https://doi.org/10.1016/j.cels.2020.06.009>. e6.
- [81] Lee CH, Koohy H, Wilkinson K, de Palma R, Bonsack M, Riemer A. In silico identification of vaccine targets for 2019-nCoV. *F1000Research* 2020;9:1–10. <https://doi.org/10.12688/f1000research.22507.2>.
- [82] Ismail S, Ahmad S, Azam SS. Immunoinformatics characterization of SARS-CoV-2 spike glycoprotein for prioritization of epitope based multivalent peptide vaccine. *J Mol Liq* 2020;314:113612. <https://doi.org/10.1016/j.molliq.2020.113612>.
- [83] Alam A, Khan A, Imam N, Siddiqui MF, Waseem M, Malik MZ, et al. Design of an epitope-based peptide vaccine against the SARS-CoV-2: a vaccine-informatics approach. *Briefings Bioinf* 2020. <https://doi.org/10.1093/bib/bbaa340>.
- [84] Gupta E, Mishra RK, Niraj R. Identification of potential vaccine candidates against SARS-CoV-2, a step forward to fight novel coronavirus 2019-nCoV: a reverse vaccinology approach. *BioRxiv* 2020. <https://doi.org/10.1101/2020.04.13.039198>.
- [85] Fast E, Altman RB, Chen B. Potential T-cell and B-cell epitopes of 2019-nCoV. *BioRxiv* 2020. <https://doi.org/10.1101/2020.02.19.955484>.
- [86] Baruah V, Bose S. Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. *J Med Virol* 2020;92:495–500. <https://doi.org/10.1002/jmv.25698>.
- [87] Chen HZ, Tang LL, Yu XL, Zhou J, Chang YF, Wu X. Bioinformatics analysis of epitope-based vaccine design against the novel SARS-CoV-2. *Infect Dis Poverty* 2020;9:88. <https://doi.org/10.1186/s40249-020-00713-3>.
- [88] Lin L, Ting S, Yufei H, Wendong L, Yubo F, Jing Z. Epitope-based peptide vaccines predicted against novel coronavirus disease caused by SARS-CoV-2. *Virus Res* 2020;288:198082. <https://doi.org/10.1016/j.virusres.2020.198082>.
- [89] Poran A, Harjanto D, Malloy M, Arieta CM, Rothenberg DA, Lenkala D, et al. Sequence-based prediction of SARS-CoV-2 vaccine targets using a mass spectrometry-based bioinformatics predictor identifies immunogenic T cell epitopes. *Genome Med* 2020;12:70. <https://doi.org/10.1186/s13073-020-00767-w>.
- [90] Srivastava S, Verma S, Kamthania M, Agarwal D, Saxena AK, Kolbe M, et al. Computationally validated SARS-CoV-2 CTL and HTL multi-patch vaccines, designed by reverse epitomics approach, show potential to cover large ethnically distributed human population worldwide. *J Biomol Struct Dyn* 2020:1–20. <https://doi.org/10.1080/07391102.2020.1838329>.
- [91] Feng Y, Qiu M, Liu L, Zou S, Li Y, Luo K, et al. Multi-epitope vaccine design using an immunoinformatics approach for 2019 novel coronavirus (SARS-CoV-2). *BioRxiv* 2020. <https://doi.org/10.1101/2020.03.03.962332>.
- [92] Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe* 2020;27:671–80. <https://doi.org/10.1016/j.chom.2020.03.002>. e2.
- [93] Rehman HM, Mirza MU, Ahmad MA, Saleem M, Froeyen M, Ahmad S, et al. A putative prophylactic solution for COVID-19: development of novel multi-epitope vaccine candidate against SARS-CoV-2 by comprehensive immunoinformatic and molecular modelling approach. *Biology* 2020;9:296. <https://doi.org/10.3390/biology9090296>.