

An in-silico approach to develop of a multi-epitope vaccine candidate against SARS-CoV-2 envelope (E) protein

Fatemeh Ghafouri

Shahid Beheshti University

Reza Ahangari Cohan

Pasteur Institute of Iran

Farshid Noorbakhsh

Tehran University of Medical Sciences

Hilda Samimi

Tehran University of Medical Sciences

Vahid Haghpanah (✉ vhaghpanah@tums.ac.ir)

Tehran University of Medical Sciences

Research Article

Keywords: SARS-CoV-2, Envelope (E) protein, COVID-19 vaccine, In-silico

Posted Date: May 20th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-30374/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Since the first appearance of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in China on December 2019, the world has now witnessed the emergence of the SARS-CoV-2 outbreak. Therefore, due to the high transmissibility rate of virus, there is an urgent need to design and develop vaccines against SARS-CoV-2 to prevent more cases affected by the virus. In this study, a computational approach is proposed for vaccine design against the envelope (E) protein of SARS-CoV-2, which contains a conserved sequence feature. First, we sought to gain potential B-cell and T-cell epitopes for vaccine designing against SARS-CoV-2. Second, we attempted to develop a multi-epitope vaccine. Immune targeting of such epitopes could theoretically provide defense against SARS-CoV-2. Finally, we evaluated the affinity of the vaccine to major histocompatibility complex (MHC) molecules to stimulate the immune system response to this vaccine. We also identified a collection of B-cell and T-cell epitopes derived from E proteins that correspond identically to SARS-CoV-2 E proteins. The in-silico design of our potential vaccine against E protein of SARS-CoV-2 demonstrated a high affinity to MHC molecules, and it can be a candidate to make a protection against this pandemic event.

Introduction

The recent outbreak of the new virus in Wuhan City, China, originated possibly from the seafood industry has contributed to the discovery of a new strain named coronavirus and labeled the SARS-CoV-2 of the Coronaviridae family. This virus has caused severe damage and anxiety, leading to the loss of myriad individuals, impacting more than 1,000,000 people (<https://who.sprinklr.com/>). Symptoms such as the flu-like illness, acute respiratory distress syndrome, and clinical or radiological evidence of pneumonia in individuals needing hospitalization are considered as COVID-19¹. Patients diagnosed with COVID-19 are reported to have high levels of interleukin 1 beta (IL1B), interferon gamma (IFN γ), interferon-inducible protein 10 (IP10), and monocyte chemoattractant protein 1 (MCP1), likely leading to activated T-helper-1 (TH1) cell responses. In comparison, patients needing ICU admission had higher concentrations of granulocyte-colony stimulating factor (GCSF), IP10, MCP1, MIP1A, and tumor necrosis factor α (TNF α) than those not needing, suggesting a possible correlation of cytokine storm and disease intensity. Nonetheless, SARS-CoV-2 infection also resulted in the enhanced production of T-helper-2 (TH2) cytokines, such as interleukin 4 (IL4) and interleukin 10 (IL10) that inhibit inflammation that varies from that of SARS-CoV infection². The persistent rise in patients and the high contagious rate of SARS-CoV-2 infection illustrate the immediate need to develop a safe and effective vaccine.

Vaccines were mostly made up of whole pathogens, either destroyed or attenuated. It may be beneficial to use peptide vaccines that are capable of generating an immune response against a specific pathogen³. Epitope-based vaccines (EVs) utilize immunogenic peptides (epitopes) to cause an immune response. The performance of the EV is calculated by the number of epitopes to be used as the foundation. Nevertheless, the experimental identification of the nominee epitopes is costly in terms of time and

money. Besides, different immunological requirements need to be considered for the final choice of epitopes ⁴.

The electron microscope determines the properties of CoVs. They are enveloped viruses with single-stranded positive-sense RNA. The coronavirus genome size varies from 26–32 kb ⁵. Like all coronaviruses, SARS-CoV-2 comprises four viral proteins, namely spike (S) protein, a form of glycoprotein, membrane (M) protein, covering the membrane, envelope (E) protein, a strongly hydrophobic protein that covers the entire coronavirus structure, and nucleocapsid (N) protein, a structural protein that suppresses the RNA interference (RNAi) in order to overcome the host defense ^{6,7} (Fig. 1). Such accessory proteins are not only essential for virion assembly but could also play an extra function in disrupting the host immune responses to promote viral replication⁸. Structural proteins spike and membrane are shown to have substantial mutational modifications, while envelope and nucleocapsid proteins (Fig. 1) are highly conserved indicating differential selection pressures imposed on SARS-CoV-2 during evolution ⁹. Envelope (E) protein is a small intrinsic membrane protein actively engaged in several stages of the life cycle of the virus, such as assembling, propagation, enveloping, and pathogenesis ¹⁰. This protein also slows the transport of proteins through the secretory pathway by adjusting the concentrations of Ca²⁺ and H⁺ in the Golgi and endoplasmic reticulum (ER) compartments and has been suggested to be a mechanism for immune avoidance ¹¹. The envelope (E) protein sequence was collected from a protein database and analyzed with various bioinformatics tools to identify protective epitopes. The toxicity of whole protein was analyzed, and nine epitopes including TALRLCAYCC, ALRLCAYCCN, LRLCAYCCNI, RLCAYCCNIV, LCAYCCNIVN, CAYCCNIVNV, AYCCNIVNVS, YCCNIVNVSL, and CCNIVNVSLV were identified as toxic epitopes. The predicted B-cell and T-cell epitopes were checked in terms of not coinciding with these regions. The toxic epitopes of envelope (E) protein were less than the spike protein. This finding was verified in the literature in which envelope (E) protein was explored in Severe Acute Respiratory Syndrome (SARS) in 2003 and, more recently, in the Middle East Respiratory Syndrome (MERS), the retention of this protein against the seven strains was investigated and confirmed using BioEdit Package tool ^{10,12-17}. Several experiments have explored the ability of CoVs with mutated envelope (E) protein or without mutated envelope (E) protein, concentrating mainly on SARS and MERS-CoV, as live attenuated vaccine candidates with some impressive results ¹⁸⁻²¹. Therefore, we decided to design an effective vaccine against this protein.

Results

Selection of protein sequences

The amino acid sequence of envelope (E) protein of SARS-CoV-2 was collected from the NCBI virus with an accession number of QHD43418 released on 2020-01-13. The FASTA sequence was used to construct a multi-epitope vaccine against SARS-CoV-2.

The sequence of “cholera toxin B (CTB) subunit [Vibrio mimicus]” was collected from the NCBI database and was used as an adjuvant.

B-cell epitopes analysis

ABCpred database has studied the full-length E protein sequence for the analysis of linear B-cell epitopes. All results passed three more filters (antigenicity, allergenicity, and toxicity) and two epitopes (NVSLVKPSFYVYSRVK – YVYSRVKNLNSSRVPD) were chosen as protective epitopes (Tab 1).

T-cell epitopes analysis

The first group of T-cell epitope prediction was performed using vaxign server (Tab 1). The next two groups including binding epitopes to MHC class I and class II molecules were predicted individually by the IEDB database. The three filters aforementioned were applied to identify the protective antigens. Based on the number of alleles, the predominant human leukocyte antigen (HLA) alleles were HLA-A*01:01, HLA-A*24:02, HLA-A*11:01, HLA-A*03:01, and HLA-B*07:02 between MHC class I alleles and HLA-DRA-DBR1*01:01 and HLA-DPB1*01:02 in MHC class II alleles.

Antigenicity of potential epitopes

Antigenicity of both B-cell and T-cell epitopes was foreseen by VaxiJen 2.0, with a threshold of 0.4. Except two epitopes, the antigenicity score of other predicted epitopes were upper the threshold and were considered as “antigen” epitopes. The two more screenings (including allergenicity and toxicity) were not carried out on “non-antigen” epitopes (Tab 1).

Antigenicity of potential epitopes

Allergenicity of both B-cell and T-cell epitopes was estimated by AllerTOP v. 2.0 (Tab 1).

Toxicity of potential epitopes

Toxicity of both B-cell and T-cell epitopes was anticipated by ToxinPred, with a peptide fragment length of 10 (Tab 1).

Construction of chimeric peptide

The tick marked epitopes in Table 1 that remained after screening were chosen for the design of chimeric peptide as a multi-epitope vaccine. Finally, we selected 12 epitopes (including 2 B-cell epitopes, 7 binding epitopes to MHC class I proteins, and 5 binding epitopes to MHC class II proteins) and they were all “antigen”, “non-allergen”, and “non-toxin”. In this part, to make a contiguous sequence in final construction, the overlapping of B-cell and T-cell epitopes were merged. Predicted linear B-cell epitopes and T-cell epitopes were connected utilizing KK linkers as flexible connectors. The “cholera toxin B subunit [Vibrio mimicus]” with GenBank accession AJP16764.1 was selected as an adjuvant and

attached to the amino terminals of the multi-epitope vaccine through PAPAP linkers as rigid linkers to improve antigen-specific immune responses²² (Fig. 1).

Antigenicity, allergenicity and toxicity estimation of the nominee multi-epitope vaccine

The final peptide chimera (Fig. 1) antigenicity (along with the adjuvant sequence) was estimated by the VaxiJen 2.0 server to be 0.5841 with a threshold of 0.4. The primary multi-epitope vaccine sequence (without adjuvant) reported a score of 0.6906. ANTIGENpro was another platform that was utilized to estimate the antigenicity of final peptide. Based on this server the whole protein (Fig. 1) is antigen with a probability of 0.809160.

AllerTOP v. 2.0 server estimated the allergenicity of the final peptide. The result for both states was "non-allergen". By utilizing ToxinPred server, the final peptide was predicted as "non-toxon".

Amino acid composition and physicochemical properties and solubility prediction

Based on the ProtParam database, the final peptide chimera comprised 338 amino acids (Fig. 1) with a molecular weight of **37.7** kDa. The isoelectric point (PI) value was expected to be 9.62. The half-life was estimated to be 30 hours, with adjuvant and 100 hours, without adjuvant in mammalian reticulocytes *in vitro* and more than 20 hours in yeast and over 10 hours in Escherichia coli (E.coli) *in vivo*. An instability index II was predicted to be 41.34, classifying the protein as unstable (II >40 indicates instability), but the amount of this index is really near the border. The estimated Grand Hydropathic Average (GRAVY) was -0.120. The negative attribute indicates that the protein is hydrophilic and can react with water molecules²². Furthermore, based on the PepCalc server, the solubility was predicted to be "good" in water. Based on SOLpro from ANTIGENpro, Peptide chimera was expected to be SOLUBLE with a possibility of 0.684306.

By using the Iupred2a server, the disorder regions of the final peptide, which make it unstable, were identified. The disorders are regarded in the adjuvant areas (Fig. 2A).

Secondary structure prediction

The secondary structure of the final protein (Fig. 1) was analyzed by the Prabi server, the final chimeric peptide was estimated to include 36.09% alpha-helix, 22.49% extended strand, and 41.42% random coil. There is no compactness in alpha-helix locations, which demonstrates there will be less difficulty in future synthesis steps (Fig. 2B). The other server we used to predict the secondary structure of the peptide chimera was PSIPRED 4.0 server. The results and the details of residues and their configurations are given in Figure 2C.

Tertiary structure homology modeling

The final protein tertiary structure was built by I-TASSER server. I-TASSER provided the top 10 proteins from PDB library that had the closest structural similarity to the predicted vaccine model. The average TM-score of these ten proteins was calculated to be 0.41. We selected the best predicted model according

to C-score; a confidence score to estimating the quality of predicted models by I-TASSER. It had the highest C-score of -3.96, and the RMSD was estimated to be $16.4 \pm 3.0 \text{ \AA}$.

Tertiary structure refinement

The GalaxyWEB server was used to refine the predicted model. After refinement, an evident improvement was observed in the percentage of residues in Ramachandran favored regions relative to the initial predicted model.

Tertiary structure validation

The plot B in Figure 3 reveals local quality model by displaying the energy as a function of the amino acid sequence location. This plot is made by ProSA-web server. In general, the negative values refer to the problematic or erroneous areas of the input model, which is observed in some parts in adjuvant regions. About our model, these negative values were observed in middle parts including epitope but the lowest value was relevant to the C-terminus connected adjuvant, not the multi-epitope section.

Based on the ProSA-web, the Z-score of peptide chimera was predicted to be -4.3 (Fig. 3C). This value is in the range of native conformations²³. The diagram of the predicted local similarity to target and the Z-score value of the homology model is shown in Figure 3B and C.

Based on the Procheck server, The Ramachandran plot analysis of the modeled protein reported that 72.0% of residues in the protein are in the favored regions. 22.2% of the residues were predicted to be in additional allowed regions and 2.3% were in generously allowed regions, with 9 residues (3.5%) in the disallowed regions (Fig. 3D).

Molecular docking of subunit vaccine with MHC molecules

Crystal structure of HLA-A*01:01, HLA-A*24:02, HLA-A*11:01, HLA-A*03:01, HLA-B*07:02, HLA-DRA-DBR1*01:01 and HLA-DPB1*01:02 were retrieved from the PDB RCSB database (PDB ID: 4NQX, 5HGH, 6ID4, 6O9B, 5E00, 1AQD, 3LQZ, respectively). The PDB files were edited and cleaned from heteroatoms. PEP-FOLD 2.0 from the RPBS Web Portal server was used to predict the tertiary structure of 6 epitopes of vaccine construction individually. A molecular docking study was carried out on epitopes and the whole vaccine construction with MHC alleles by the ClusPro 2.0 online server. PyMOL software was used to perform a detailed analysis of the interface of protein-protein interaction (PPI) (Fig. 4). The weighted score of the lowest energy docked complexes are reported in Table 2. The best way to rank the model is the cluster size (number of members)^{24,25}. The most populated clusters were found in SFVSEETGT and HLA-A*24:02, TLAILTALR and HLA-A*24:02, VTLAILTAL and HLA-A*24:02, and TLAILTALR and HLA-A*01:01 with 997, 997, 990, and 769 cluster members, respectively.

The docking study on the whole vaccine construction with the MHC molecules depicted one of the most probable position and orientation of the peptide within the MHC binding grooves (Fig. 4H). The active residues of Fig. 4I belongs to [TLAILTALR](#) epitope which interacts with chain E of HLA-DRB1*01:01.

Immune response simulation

We predicted the IL4, IL10 and INF γ inducing peptides from the 6 epitopes in the final vaccine construction via IL4pred server, IL10pred server and INFepitope server, respectively. The result predicted the SFVSEETGT epitope as an IL4 inducer and the NVSLVKPSFYVYSRVK as an IL4, IL10, and INF γ inducer and the SFYVYSRVKLNSSRVPD as an IL4 and IL10 inducer peptide.

The primary and secondary immune responses were stimulated by the C-IMMSIM server. This server simulated the immune response to vaccine candidate (without adjuvants) for three times of injection in the time steps of 1, 84, and 100. Each time step is equal to 8 hours. To make a relative comparison, we made a shuffled sequence of the vaccine candidate as a control protein, and we analyzed the results of the immune response simulation to injection of it. This shuffled sequence was employed to evaluate the significance of the vaccine sequence results, because in immune response simulation by this server, the sequence composition (the final epitopes connected via KK linkers) is an important consideration. Finally, we found out that the results of the vaccine injection varied from those of the controls. The results of the immune response simulations are given in Figures 5 and 6.

Discussion

Based on the World Health Organization (WHO) website (<https://who.sprinklr.com/>) as of April 13, 2020, the number of COVID-19 confirmed cases is estimated at 1,776,867 people and 111,828 deaths. Therefore, there is an immediate need to develop vaccines against this transmissible disease. There is currently no vaccination or licensed medication for humans against SARS-CoV-2. Nonetheless, further clinical trials are still needed to confirm their efficacy and safety ²⁶.

Epitope-based vaccines offer a new strategy for the prophylactic and therapeutic use of pathogen-specific immunity ²⁷. A multi-epitope vaccine consisting of a series of or overlapping peptides seems to be an appropriate solution to the prevention and treatment of viral infections ¹²⁻¹⁷. The perfect multi-epitope vaccine should be engineered to include epitopes which can activate cytotoxic T lymphocyte (CTL), T-cells and B-cells and trigger successful responses to specific viruses ¹².

In this study, we present an in-silico design of a potential multi-epitope vaccine against the E protein of SARS-CoV-2, which is made of both B-cell and T-cell epitopes which can stimulate the immune system responses impressively. Envelope (E) protein is conserved in all CoVs and covers the entire surface of SARS-CoV-2 (Fig. 1). It has less toxic regions, rather the spike protein. Several studies have examined the potential of CoVs with mutated envelope (E) protein or without mutated envelope (E) protein, focusing specifically on SARS-and MERS-CoV, as live attenuated vaccine candidates associated with hopeful results ^{10,18,19,28-30}. First of all, we obtained the FASTA sequence of E protein of SARS-CoV-2 from the NCBI database. B-cell and T-cell epitopes of this protein were predicted by different servers. The epitopes were screened based on three filters of antigenicity, allergenicity, and toxicity. Therefore, we selected the protective epitopes. We merged the overlaps of B-cell and T-cell epitopes and fused them by appropriate

flexible linkers. Previous studies reported that KK linkers preserve independent immune responses when they are inserted between epitopes³¹. Then, we linked CTB adjuvants at the terminus of epitopes by PAPAP linkers as rigid linkers to enhance the biological activities³² (Fig. 1). Another study suggested that CTB, though an important adjuvant through the nasal and oral routes of administration, can also be considered to improve the immune response in intramuscular dosing vaccine regimens³³. Immunological adjuvants are agents that improve the intensity, activation, or longevity of antigen-specific immune responses if used in conjunction with particular vaccine antigens³⁴.

The absence of allergenic properties of the proposed peptide chimera further increases its potential as a vaccine candidate²². Finally, the whole peptide chimera was analyzed for antigenicity, allergenicity, and toxicity, and it was predicted as Antigen³⁵, Non-Allergen³⁶, and Non-Toxin³⁷. PI was calculated to be 9.62, which shows that the final protein is alkaline. It was predicted as "soluble" upon expression in the E.coli host. The instability index II was about 1 unit over the threshold of 40, and it resulted in considering this protein as "unstable". However, more analysis demonstrated that the residues which are responsible for such disorders to make it unstable are in adjuvant regions not in the multi-epitope area (Fig. 2A). Secondary structure analysis predicted that the final protein is consisted of 36.09% alpha-helix, 22.49% extended strand, and 41.42% random coil. Essential types of "structural antigens" have been identified as natively unfolded protein regions and alpha-helical coils peptides. These two structural types, when examined in synthetic peptides, have the capacity to fold into their native structure and are therefore recognized by antibodies naturally triggered in response to infection^{22,38}. Protein three dimensional (3D) structures offer useful insights into their molecular activity and provide a wide variety of applications in bioscience³⁹. The I-TASSER server modeled the tertiary structure of the final protein. Based on the Ramachandran plot, 96% of the residues of refined predicted model were found in favored and allowed regions with 3.5% outliers. Successful refining might improve the applicability of template-based models by offering more reliable structures for functional analysis, molecular design or experimental structure determination⁴⁰. In the context of structural vaccinology, a molecular docking study was needed to forecast the binding affinity of epitopes to the crystallized fragment (FC) of antibodies or MHC molecules^{41,42}. To analyze the affinity of the final multi-epitope vaccine to MHC molecules, we did 26 docking studies. They were carried out on the 6 epitopes of the final vaccine and the whole vaccine construction with MHC class I and class II receptors. The results of docking studies were notable and demonstrated the high affinity of the multi-epitope vaccine and its individual epitopes to MHC molecules. Then the interface of protein-protein interactions was reconsidered by a visualizations tool. At the next step of designing a multi-epitope vaccine, following an approach of systems vaccinology is beneficial in assessing the human complex immune response at different stages of biological structures⁴³. Finally, we utilized an immune simulator server to predict the primary and secondary response of the immune system to three times of injection of the candidate vaccine. From cytokines simulation plot, we noted an increase in amounts of IL-4 and INF γ , similar to that Huang et al.² observed in clinical features of COVID-19 patients (Fig. 5J). Appropriate activation in antigen-presenting cells (APC) cells, the high production of memory cells due to the extensive activation of B-cells and T-cells, control and clearance of antigens due

to the creation of cytokines by the participation of TH memory cells and the evident long-term memory persistence after three times of injection, could confirm the efficiency of our candidate vaccine ⁴⁴.

Conclusion

The goal of this research was to suggest a computational method for predicting protective B-cell and T-cell epitopes of the Envelope (E) protein of SARS-CoV-2 to construct a chimeric peptide candidate against this pandemic disease. The results of the present study demonstrated a high affinity of this chimeric peptide to MHC molecules of the immune system, and the outputs of immune response simulation to the injection of this novel vaccine confirmed our findings. To conclude, the multi-epitope vaccine designed against E protein of SARS-CoV-2 by utilizing immunoinformatics methods may be considered as a new, safe, and efficient approach against SARS-CoV-2.

Methods

Retrieval protein sequence

Based on vaxquery database (<http://www.violinet.org/vaxquery/>) envelope (E) protein of SARS-CoV-2 can be a target to design vaccines because there is a vaccine in research status working on this protein. The amino acid sequence of the envelope (E) protein and spike protein of SARS-CoV-2 were collected from the NCBI virus (<https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/>) with the accession number of QHD43418 and QHR63280, respectively.

B-cell epitopes prediction

B-cell epitopes are pieces of proteins or other molecules in which antibodies (produced by B-cells) bind. Prediction methods are both time saving and cost-effective, and reliable approaches for predicting linear B-cell epitopes thus would be the first move in leading the genome-wide quest for B-cell antigens in pathogenic organism ⁴⁵ (Tab. 1). ABCpred database (<http://crdd.osdd.net/raghava/abcpred/>) has analyzed the full-length envelope (E) sequence for the prediction of linear B-cell epitopes. ABCpred uses a machine learning methodology that requires fixed-length patterns for training or research, while B-cell epitopes range from 5 to 30 residues. To overcome this issue, the server sought to create data sets of fixed-length patterns from B-cell epitopes by removing or linking residues to terminals. (<http://crdd.osdd.net/raghava/abcpred/>).

T-cell epitopes prediction

T-cell epitopes are a group of peptides that can be detected by T-cell receptors after a given antigen have been processed intracellularly and attached to at least one MHC molecule and expressed on the surface of the APC as an MHC-peptide complex. Entities that have at least one MHC molecule that can most eagerly attach to allergenic amino acid sequences from an allergen and at the same time have the correct T-cell clone that can detect this MHC-peptide complex are known to be genetically susceptible to allergic

reactions to this allergen. This concept can be investigated in in-silico by employing advanced statistical methods focused on sophisticated mathematics and statistics⁴⁶. The first group of T-cell epitope prediction was performed using vaxign server (<http://www.violinet.org/vaxign/>) (Tab. 1). Vaxign is the first web-based vaccine design program to predict vaccine targets relying on genome sequences using the reverse vaccine strategy. Foreseen features of the Vaxign pipeline provide protein subcellular location, transmembrane helices, adhesion probability, human or mouse protein retention, sequencing exclusion from the genome(s) of non-pathogenic strain(s), and epitope binding to MHC molecules⁴⁷. The next two groups helper T lymphocyte (HTL) epitopes and cytotoxic T lymphocytes (CTL) epitopes were predicted individually by the IEDB database (<http://tools.iedb.org/mhcii/> and <http://tools.iedb.org/mhci/>) (Tab 1). There is a tool available on the IEDB database to predict binding epitopes to MHC class I molecules. This device can take a series of amino acids, or sequences, and assess the capacity of each subsequence to bind to a different MHC class I molecule. (<http://tools.iedb.org/main/tcell/>). The other tool that is available on the IEDB database estimates peptide attachment to MHC class II molecules. This tool uses a variety of methods to predict MHC class II epitopes, including a consensus approach combining NN-align, SMM-align, and combinatorial library methods (<http://tools.iedb.org/main/tcell/>).

Antigenicity, allergenicity and toxicity prediction

VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) with a threshold of 0.4 was used to predict the antigenicity of both B-cell and T-cell epitopes (Tab. 1). A threshold of 0.4 was used to predict the antigenicity of both B-cell and T-cell epitopes. VaxiJen is the first alignment-independent antigen predictor server. It is created to make the categorization of antigens solely based on the physicochemical properties of proteins without recourse to sequence alignment. The system can be used either on its own or in conjunction with alignment-based prediction methods⁴⁸. VaxiJen v2.0 was used to estimate the antigenicity of whole peptide chimera. Based on this server, the antigenicity score of final protein was 0.5841 (Probable ANTIGEN) with 0.4 thresholds. Likewise, ANTIGENpro (<http://scratch.proteomics.ics.uci.edu/>) was utilized to predict the antigenicity of peptide chimera. ANTIGENpro is an alignment-free, sequence-based, and pathogen-independent protein antigenicity predictor. The forecasts are a two-stage model based on multiple versions of the primary sequence with five machine learning algorithms. The final SVM classifier analyses the corresponding predictions and determines whether or not the protein is probable to be antigenic or not, as well as the relevant probability. ANTIGENpro is the first indicator of all protein antigenicity trained to employ reactivity data from the protein microarray analysis of five pathogens.

(<http://scratch.proteomics.ics.uci.edu/explanation.html#ANTIGENpro>) AllerTOP v2.0 (<http://www.ddg-pharmfac.net/AllerTOP/feedback.py>) was used to predict the Allergenicity of both B-cell and T-cell epitopes (Tab. 1). Protein sequences are sent to this server in simple text. The results page shows the identity of an allergen: "Probable Allergen" or "Probable Non-allergen". The whole protein chimera was predicted as a "probable non-allergen" using this tool⁴⁹.

ToxinPred (<https://webs.iitd.edu.in/raghava/toxinpred/protein.php>) with a peptide fragment length of 10 was used to predict toxicity of both B- and T-cell epitopes (Tab. 1). ToxinPred is a computational tool built

to anticipate and design toxic/non-toxic peptides. The primary dataset used for this approach is comprised of 1805 toxic peptides (≤ 35 residues) (<http://crdd.osdd.net/raghava/toxinpred/>). This server also was used to predict the toxicity of whole protein chimera and no fragment was predicted as toxin.

Construction of chimeric peptide

Selected B-cell and T-cell epitopes (Tab. 1) were fused to construct the protein chimera as a multi-epitope vaccine. Overlaps of B-cell and T-cell epitopes were merged. KK linkers (flexible linkers) were used to connect the epitopes. The bi-lysine (KK) linker was implanted between separate epitopes to maintain their independent immunological functions (Fig. 1). KK is the target sequence of cathepsin B, which is one of the essential antigen processing proteases in the sense of the MHC class II antigen presentation³¹. The CTB subunit was chosen as an adjuvant and applied to the amino terminals of the multi-epitope peptide. There are several clear advantages to the use of adjuvants: 1) stabilizing the formulation of the vaccine, since physical structure of immunostimulators makes them highly unstable in aqueous solutions; 2) co-location of the vaccine antigen and adjuvant, assuring the activation of the same cells that have faced the antigen⁵⁰. PAPAP linkers as rigid linkers were used to link adjuvants to epitopes (Fig. 1) in order to improve biological activity³².

Amino acid composition and physicochemical properties and solubility prediction

ProtParam database (<https://web.expasy.org/protparam/>) was used to calculate and predict the molecular weight, isoelectric point value (PI), in vivo, and in vitro half-life, instability index II and grand average of hydropathicity (GRAVY). ProtParam from the ExPASy server is a reliable algorithm to compute Physico-chemical properties. However, it uses a single sequence per analysis through the interface.

The IUPred2a server (<https://iupred2a.elte.hu/>) was used to analyzing disorders which make the final protein unstable (Fig. 2A). The structural states of proteins involve organized globular domains as well as fundamentally distorted protein areas that function as extremely variable conformational ensembles in isolation. IUPred2A is an integrated web interface that produces energy estimation dependent on IUPred2 order and disordered residue predictions and abnormal binding regions by ANCHOR2. The application produces visual and text outputs⁵¹.

SOLpro from ANTIGENpro (<http://scratch.proteomics.ics.uci.edu/explanation.html#SOLpro>) was used to predict the solubility of peptide chimera upon overexpression. SOLpro predicts the tendency of the protein to be soluble when overexpressed in E.coli using a two-stage SVM model based on multiple representations of the primary sequence. Each first layer classifier uses a separate set of features describing the sequence as input. The final SVM classifier sums up the resulting estimates and predicts whether the protein is soluble or not, as well as the relevant probability. (<http://scratch.proteomics.ics.uci.edu/explanation.html#SOLpro>).

PepCalc server (<https://pepcalc.com/>) is another server to predict the solubility of the final protein. It is only a very rough estimation of water solubility (<https://pepcalc.com/>).

Secondary structure prediction

Prabi server (<https://prabi.ibcp.fr/htm/site/web/services/secondaryStructurePrediction>) was used to predict the secondary structure of the final sequence of peptide chimera (Fig. 2B). All PRABI components provide services in their various areas of expertise (e.g., molecular, phylogeny, genomics, transcriptomics, proteomics, protein structure, and medical biostatistics (<http://www.prabi.fr/spip.php?page=services>)).

PSIPRED 4.0 (<http://bioinf.cs.ucl.ac.uk/psipred/>) is another servers to predict secondary structure employed to achieve the details of residue's configurations. It is a very simple system of secondary prediction based on a simple neural network evaluation of PSI-BLAST-generated profiles and is capable of generating findings that place the process at the very top of the current forecasting system crop ⁵².

Tertiary structure homology modelling

I-TASSER server (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) was used to carry out the homology modeling of the final protein (Fig. 3A). I-TASSER is a hierarchical method for protein structure and function forecasting. First, it identifies structural PDB templates, with full-length atomic models built through iterative template-based fragment assembly simulations. After the structural integration simulation, I-TASSER uses the TM-align structural alignment software to match the first I-TASSER configuration to all the structures in the PDB database. Then it reports the top 10 proteins from the PDB that have the closest structural similarity ⁵³⁻⁵⁵. UCSF Chimera was used to create a high quality image to present the predicted model. UCSF Chimera is a visualization tool to analysis the molecular structures ⁵⁶.

Tertiary structure refinement

GalaxyWEB server (<http://galaxy.seoklab.org/index.html>) was used to refine the tertiary structure predicted model. In this refinement process, the server first reconstructs side chains and performs side-chain repacking following eventual overall structure relaxation through molecular dynamic simulation ^{40,57}.

Validation of the tertiary structure

Model validation is an essential step in the model construction process as it identifies possible defects in the 3D structures expected ⁵⁸. Ramachandran plot analysis was performed using the Procheck server (<https://servicesn.mbi.ucla.edu/PROCHECK/>) (Fig. 3D); as a program to control the stereochemical consistency of the protein structure (<https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>). The Ramachandran plot is particularly useful as a test for geometric validation since ϕ and ψ are not part of the target function for refining. The percentage of residues found in the most favored ϕ and ψ regions is strongly correlated with resolution and is now reported as standard in protein structure papers, whereas specific "outlier" residues incorrect structures are used to denote either possible errors or possibly significant strained conformations⁵⁹. ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php>) was used to analyze the Z-score of the 3D model predicted (Fig. 3C). The Protein Structure Analysis (ProSA)

program is a well-established method with a large user base. It is widely used in the optimization and validation of experimental protein structures and structural prediction and simulation⁶⁰. The Z-score shows the overall quality of the model and calculates the divergence of the total energy of the structure from the distribution of energy from spontaneous conformations⁶¹.

Molecular docking of subunit vaccine with MHC molecules

PEP-FOLD 2.0 from the RPBS Web Portal server (<https://mobyale.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::PEP-FOLD>) performed the prediction of the tertiary structure of vaccine construct epitopes. PEP-FOLD is an online tool that is designed to model 3D peptide conformation structures in aqueous solutions for 9-25 amino acid length peptides (*de novo* modeling). PEP-FOLD conducts a series of 50 simulations, beginning with an amino acid sequence and returns the most critical energy and population-related conformations to be found⁶².

The ClusPro 2.0 server rotates each Final epitopes ligand and whole vaccine peptide with 70,000 rotations. It translated the ligand rotations relative to the MHC receptor alleles in 3 axes (x,y,z) on a grid. Then, it chose the scores of top 1000 lowest energy docked structures 70,000 rotations. These 1000 lowest energy docked structures would be processed subsequently. This set might have the potential to consist of at least some models which are close to the native structure of the complex. Then, the server clustered the 1000 rotations by finding the structure with the most “neighbors” within 9 Å IRMSD radii as the distance measure. Then, it considered this ligand and its neighbors as the “cluster center” and the “members” of the cluster, respectively. This process was repeated under remainder of the ligands to find the next clusters. Finally, the server scores the models and reported the top score ones based on the cluster size (10 most populated clusters)^{24,25}.

PyMOL software was used to analyze docking results. PyMOL is mostly utilized for molecular visualization by crystallographic, molecular dynamic simulation, and protein modeling software packages⁶³.

Immune response simulation

IL4, IL10 and INF γ inducing peptides from the 6 epitopes in the final vaccine construction were predicted via IL4pred server (<https://webs.iiitd.edu.in/raghava/il4pred/design.php>), IL-10Pred server (<https://webs.iiitd.edu.in/raghava/il10pred/predict3.php>) and INFepitope server (<https://webs.iiitd.edu.in/raghava/ifnepitope/predict.php>), respectively.

Immune response to vaccine injection was simulated by the C-IMMSIM server (<http://150.146.2.1/C-IMMSIM/index.php>) (Fig. 5). C-ImmSim is an agent-based computational immune response simulator that utilizes position-specific score matrix (PSSM) and machine learning methods for predicting epitope and immune interactions, respectively⁶⁴. We regulate the parameters based on the predominant HLA alleles of predictions. The host HLA selection parameter for MHC class I was set on A1010, A1101, and B0702 and for DR MHC class II was set on DBR1_0101 and based on literatures the time step to injection

parameter was set on 1, 84 and 100 (maximum allowed value), respectively. We randomly shuffled the vaccine protein sequence (without adjuvants) by using the Stothard P (2000), the Sequence Manipulation Suite server (https://www.bioinformatics.org/sms2/shuffle_protein.html)⁶⁵ to create a control group.

The immune system simulation server mentioned here provides an opportunity to study the overall immunogenicity of the generic protein sequence in the context of its amino acid sequence⁶⁶. The total simulation is focused on three events: 1) B-cell epitopes binding, 2) HLA class I and II epitopes binding, and 3) TCR binding, which HLA-peptide complex interaction should be presented. Such processes are independently conducted by cells described by agents and occupy a specified simulated biological amount⁶⁶.

Declarations

Acknowledgments

The authors are grateful to Mrs. Hajipour for her English language comments on the initial draft of the manuscript. Also the authors are thankful to Dr. Mahmood Naderi for his helpful suggestions and advice. Molecular graphics and analyses performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311. The figure No. 6 was created with [BioRender.com](https://www.bioinformatics.org/sms2/shuffle_protein.html).

Author contributions

F.G. and V.H. created the idea; All authors discussed and designed the study; F.G performed all parts of the study. R.A.C. supervised the bioinformatics section; F.N. supervised the immunology section; H.S. contributed in acquisition and data analysis; V.H. supervised the whole project; F.G. wrote the first draft and all authors critically reviewed and approved the final version of the manuscript.

Conflict of interest

The authors declare no competing interests.

Data availability

All data generated or analysed during this study are included in this published article.

References

1. Razai, M. S., Doerholt, K., Ladhani, S. & Oakeshott, P. Coronavirus disease 2019 (covid-19): a guide for UK GPs. *BMJ* **368**, m800, doi:10.1136/bmj.m800 (2020).
2. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* **395**, 497-506, doi:[https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5) (2020).

3. Nardin, E. H. *et al.* A Totally Synthetic Polyoxime Malaria Vaccine Containing &em>Plasmodium falciparum&/em> B Cell and Universal T Cell Epitopes Elicits Immune Responses in Volunteers of Diverse HLA Types. *The Journal of Immunology* **166**, 481, doi:10.4049/jimmunol.166.1.481 (2001).
4. Bing, Z., Sakharkar, K. R. & Sakharkar, M. K. In Silico Design of Epitope-based Vaccines. *Encyclopedia of Systems Biology*, doi:10.1007/978-1-4419-9863-7_90 (2013).
5. Wu, A. *et al.* Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell host & microbe* **27**, 325-328, doi:10.1016/j.chom.2020.02.001 (2020).
6. Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y. & Jung, S.-H. An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *Journal of Medicinal Chemistry* **59**, 6595-6628, doi:10.1021/acs.jmedchem.5b01461 (2016).
7. Cui, L. *et al.* The Nucleocapsid Protein of Coronaviruses Acts as a Viral Suppressor of RNA Silencing in Mammalian Cells. *Journal of virology* **89**, 9029-9043, doi:10.1128/JVI.01331-15 (2015).
8. Pillaiyar, T., Meenakshisundaram, S. & Manickam, M. Recent discovery and development of inhibitors targeting coronaviruses. *Drug Discovery Today*, doi:https://doi.org/10.1016/j.drudis.2020.01.015 (2020).
9. Ramaiah, A. & Arumugaswami, V. Insights into Cross-species Evolution of Novel Human Coronavirus 2019-nCoV and Defining Immune Determinants for Vaccine Development. *bioRxiv*, 2020.2001.2029.925867, doi:10.1101/2020.01.29.925867 (2020).
10. Schoeman, D. & Fielding, B. C. Coronavirus envelope protein: current knowledge. *Virology Journal* **16**, 69, doi:10.1186/s12985-019-1182-0 (2019).
11. de Jong, A. S. *et al.* The Coxsackievirus 2B Protein Increases Efflux of Ions from the Endoplasmic Reticulum and Golgi, thereby Inhibiting Protein Trafficking through the Golgi. *Journal of Biological Chemistry* **281**, 14144-14150 (2006).
12. Zhang, L. Multi-epitope vaccines: a promising strategy against tumors and viral infections. *Cellular & Molecular Immunology* **15**, 182-184, doi:10.1038/cmi.2017.92 (2018).
13. Buonaguro, L. & Consortium, H. Developments in cancer vaccines for hepatocellular carcinoma. *Cancer Immunology, Immunotherapy* **65**, 93-99, doi:10.1007/s00262-015-1728-y (2016).
14. Brennick, C. A., George, M. M., Corwin, W. L., Srivastava, P. K. & Ebrahimi-Nik, H. Neoepitopes as cancer immunotherapy targets: key challenges and opportunities. *Immunotherapy* **9**, 361-371, doi:10.2217/imt-2016-0146 (2017).
15. Kuo, , Wang, C., Badakhshan, T., Chilukuri, S. & BenMohamed, L. The challenges and opportunities for the development of a T-cell epitope-based herpes simplex vaccine. *Vaccine* **32**, 6733-6745, doi:https://doi.org/10.1016/j.vaccine.2014.10.002 (2014).
16. He, R. *et al.* Efficient control of chronic LCMV infection by a CD4 T cell epitope-based heterologous prime-boost vaccination in a murine model. *Cellular & Molecular Immunology* **15**, 815-826, doi:10.1038/cmi.2017.3 (2018).

17. Lu, I. N., Farinelle, S., Sausy, A. & Muller, C. P. 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. *Cellular & Molecular Immunology* **14**, 511-520, doi:10.1038/cmi.2016.20 (2017).
18. Regla-Nava, J. A. *et al.* Severe Acute Respiratory Syndrome Coronaviruses with Mutations in the E Protein Are Attenuated and Promising Vaccine Candidates. *Journal of Virology* **89**, 3870, doi:10.1128/JVI.03566-14 (2015).
19. Almazán, F. *et al.* Engineering a Replication-Competent, Propagation-Defective Middle East Respiratory Syndrome Coronavirus as a Vaccine Candidate. *mBio* **4**, e00650-00613, doi:10.1128/mBio.00650-13 (2013).
20. Gretebeck, L. M. & Subbarao, K. Animal models for SARS and MERS coronaviruses. *Curr Opin Virol* **13**, 123-129, doi:10.1016/j.coviro.2015.06.009 (2015).
21. Fett, C., DeDiego, M. L., Regla-Nava, J. A., Enjuanes, L. & Perlman, S. Complete protection against severe acute respiratory syndrome coronavirus-mediated lethal respiratory disease in aged mice by immunization with a mouse-adapted virus lacking E protein. *Journal of virology* **87**, 6551-6559, doi:10.1128/JVI.00087-13 (2013).
22. Shey, R. A. *et al.* In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Scientific Reports* **9**, 4409, doi:10.1038/s41598-019-40833-x (2019).
23. Wiederstein, M. & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* **35**, W407-410, doi:10.1093/nar/gkm290 (2007).
24. Kozakov, D. *et al.* How good is automated protein docking? *Proteins* **81**, 2159-2166, doi:10.1002/prot.24403 (2013).
25. Kozakov, D. *et al.* The ClusPro web server for protein-protein docking. *Nat Protoc* **12**, 255-278, doi:10.1038/nprot.2016.169 (2017).
26. Lu, H. Drug treatment options for the 2019-new coronavirus (2019-nCoV). *BioScience Trends* **advpub**, doi:10.5582/bst.2020.01020 (2020).
27. Khan, A. M. *et al.* A systematic bioinformatics approach for selection of epitope-based vaccine *Cellular Immunology* **244**, 141-147, doi:https://doi.org/10.1016/j.cellimm.2007.02.005 (2006).
28. Netland, J. *et al.* Immunization with an attenuated severe acute respiratory syndrome coronavirus deleted in E protein protects against lethal respiratory disease. *Virology* **399**, 120-128, doi:https://doi.org/10.1016/j.virol.2010.01.004 (2010).
29. Lamirande, E. W. *et al.* A Live Attenuated Severe Acute Respiratory Syndrome Coronavirus Is Immunogenic and Efficacious in Golden Syrian Hamsters. *Journal of Virology* **82**, 7721, doi:10.1128/JVI.00304-08 (2008).
30. Fett, C., DeDiego, M. L., Regla-Nava, J. A., Enjuanes, L. & Perlman, S. Complete Protection against Severe Acute Respiratory Syndrome Coronavirus-Mediated Lethal Respiratory Disease in Aged Mice

- by Immunization with a Mouse-Adapted Virus Lacking E Protein. *Journal of Virology* **87**, 6551, doi:10.1128/JVI.00087-13 (2013).
31. Sarobe, P. *et al.* Enhancement of peptide immunogenicity by insertion of a cathepsin B cleavage site between determinants recognized by B and T cells. *Research in Immunology* **144**, 257-262, doi:https://doi.org/10.1016/0923-2494(93)80102-5 (1993).
 32. Chen, X., Zaro, J. L. & Shen, W.-C. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev* **65**, 1357-1369, doi:10.1016/j.addr.2012.09.039 (2013).
 33. Krakauer, T. Enterotoxins: Microbial Proteins and Host Cell Dysregulation. *Toxins* **8**, 17, doi:10.3390/toxins8010017 (2016).
 34. Crowe, J. E. Prevention of Fetal and Early Life Infections Through Maternal–Neonatal Immunization. *Infectious Diseases of the Fetus and Newborn*, 1212–1230, doi:10.1016/b978-1-4160-6400-8.00038-9 (2011).
 35. Doytchinova, A. & Flower, D. R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* **8**, 4, doi:10.1186/1471-2105-8-4 (2007).
 36. Dimitrov, I., Flower, D. R. & Doytchinova, I. AllerTOP—a server for in silico prediction of allergens. *BMC bioinformatics* **14 Suppl 6**, S4-S4, doi:10.1186/1471-2105-14-S6-S4 (2013).
 37. Gupta, S. *et al.* In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLOS ONE* **8**, e73957, doi:10.1371/journal.pone.0073957 (2013).
 38. Bennuru, S. *et al.* Stage-Specific Transcriptome and Proteome Analyses of the Filarial Parasite *Onchocerca volvulus* and Its *Wolbachia* Endosymbiont. *mBio* **7**, e02028-02016, doi:10.1128/mBio.02028-16 (2016).
 39. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**, W296-W303, doi:10.1093/nar/gky427 (2018).
 40. MacCallum, J. L. *et al.* Assessment of the protein-structure refinement category in CASP8. *Proteins* **77 Suppl 9**, 66-80, doi:10.1002/prot.22538 (2009).
 41. Agostino, M., Mancera, R. L., Ramsland, P. A. & Fernández-Recio, J. Optimization of protein-protein docking for predicting Fc-protein interactions. *J Mol Recognit* **29**, 555-568, doi:10.1002/jmr.2555 (2016).
 42. María, R. A. R., Arturo, C. V. J., Alicia, J. A., Paulina, M. L. G. & Gerardo, A. O. The Impact of Bioinformatics on Vaccine Design and Development. *Vaccines*, doi:10.5772/intechopen.69273, (2017).
 43. Raeven, R., Riet, E., Meiring, H., Metz, B. & Kersten, G. Systems vaccinology and big data in the vaccine development chain. *Immunology* **156**, doi:10.1111/imm.13012 (2018).
 44. Six, A., Bellier, B., Thomas-Vaslin, V. & Klatzmann, D. Systems biology in vaccine design. *Microb Biotechnol* **5**, 295-304, doi:10.1111/j.1751-7915.2011.00321.x (2012).
 45. Larsen, J. E. P., Lund, O. & Nielsen, M. Improved method for predicting linear B-cell epitopes. *Immunome Research* **2**, 2, doi:10.1186/1745-7580-2-2 (2006).

46. Konstantinou, G. N. T-Cell Epitope Prediction. *Methods in molecular biology (Clifton, N.J.)* **1592**, 211-222, doi:10.1007/978-1-4939-6925-8_17 (2017).
47. He, Y., Xiang, Z. & Mobley, H. Vaxign: The First Web-Based Vaccine Design Program for Reverse Vaccinology and Applications for Vaccine Development. *Journal of biomedicine & biotechnology* **2010**, 297505, doi:10.1155/2010/297505 (2010).
48. Doytchinova, A. & Flower, D. R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* **8**, 4, doi:10.1186/1471-2105-8-4 (2007).
49. Dimitrov, I., Bangov, I., Flower, D. R. & Doytchinova, I. AllerTOP v.2—a server for in silico prediction of allergens. *Journal of Molecular Modeling* **20**, 2278, doi:10.1007/s00894-014-2278-5 (2014).
50. Christensen, D. Vaccine adjuvants: Why and how. *Human vaccines & immunotherapeutics* **12**, 2709-2711, doi:10.1080/21645515.2016.1219003 (2016).
51. Meszaros, B., Erdos, G. & Dosztanyi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res* **46**, W329-w337, doi:10.1093/nar/gky384 (2018).
52. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices¹¹ Edited by G. Von Heijne. *Journal of Molecular Biology* **292**, 195-202, doi:https://doi.org/10.1006/jmbi.1999.3091 (1999).
53. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**, 725-738, doi:10.1038/nprot.2010.5 (2010).
54. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nature methods* **12**, 7-8, doi:10.1038/nmeth.3213 (2015).
55. Yang, J. & Zhang, Y. I-TASSER server: new development for protein structure and function *Nucleic acids research* **43**, W174-W181, doi:10.1093/nar/gkv342 (2015).
56. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
57. Heo, L., Park, H. & Seok, C. GalaxyRefine: Protein structure refinement driven by side-chain *Nucleic Acids Res* **41**, W384-388, doi:10.1093/nar/gkt458 (2013).
58. Khatoon, N., Pandey, R. K. & Prajapati, V. K. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Scientific Reports* **7**, 8285, doi:10.1038/s41598-017-08842-w (2017).
59. Richardson, J. S., Arendall, W. B. & Richardson, D. C. New Tools and Data for Improving Structures, Using All-Atom Contacts. in *Methods in Enzymology* 374, 385-412 (Elsevier, 2003).
60. Wiederstein, M. & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research* **35**, W407-W410, doi:10.1093/nar/gkm290 (2007).
61. Sippl, M. J. Knowledge-based potentials for proteins. *Current Opinion in Structural Biology* **5**, 229-235, doi:https://doi.org/10.1016/0959-440X(95)80081-6 (1995).

62. Maupetit, J., Derreumaux, P. & Tuffery, P. PEP-FOLD: An online resource for de novo peptide structure prediction. *Nucleic acids research* **37**, W498-503, doi:10.1093/nar/gkp323 (2009).
63. Yuan, S., Chan, H. C. S. & Hu, Z. Using PyMOL as a platform for computational drug design. *WIREs Computational Molecular Science* **7**, e1298, doi:10.1002/wcms.1298 (2017).
64. Nain, Z. *et al.* Immunoinformatic and dynamic simulation-based designing of a multi-epitope vaccine against emerging pathogen Elizabethkingia anophelis. *bioRxiv*, 758219, doi:10.1101/758219 (2019).
65. Stothard, P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* (2000). Available at: <https://ncbi.nlm.nih.gov/pubmed/10868275>. (Accessed: 28th March 2020)
66. Rapin, N., Lund, O. & Castiglione, F. Immune system simulation online. *Bioinformatics* **27**, 2013-2014, doi:10.1093/bioinformatics/btr335 (2011).

Tables

Table 1: T-cell and B-cell protective epitopes

Epitope	B-cell	MHC I	MHC II	Antigenicity	Allergenisty	Toxicity
ABCpred						
o TLAILTALRLCAYCCN	+	+	**	Antigen	Non-Allergen	Toxin
√ NVSLVKPSFYVYSRVK	+	-	-	Antigen	Non-Allergen	Non-Toxin
√ YVYSRVKLNLSRVPD	+	+	+	Antigen	Non-Allergen	Non-Toxin
o LCAYCCNIVNVSLVKP	+	-	-	Antigen	Non-Allergen	Toxin
o FVSEETGTLIVNSVLL	+	-	-	Non-Antigen	Discontinued	Discontinued
IEDB database						
√ SRVKNLNSSR	-	+	-	Antigen	Non-Allergen	Non-Toxin
√ SFVSEETGT	-	+	-	Antigen	Non-Allergen	Non-Toxin
√ SRVKNLNSS	-	+	-	Antigen	Non-Allergen	Non-Toxin
√ VYSRVKNLNS	-	+	-	Antigen	Non-Allergen	Non-Toxin
√ VYSRVKNLN	-	+	-	Antigen	Non-Allergen	Non-Toxin
√ RVKNLNSSR	-	+	-	Antigen	Non-Allergen	Non-Toxin
√ SFYVYSRVKLNLSR	-	-	+	Antigen	Non-Allergen	Non-Toxin
√ FYVYSRVKLNLSRV	-	-	+	Antigen	Non-Allergen	Non-Toxin
√ YVYSRVKLNLSRVP	-	-	+	Antigen	Non-Allergen	Non-Toxin
o IVNSVLLFLAFVVFL	-	-	+	Antigen	Allergen	Discontinued
√ FYVYSRVKLNLSRV	-	-	+	Antigen	Non-Allergen	Non-Toxin
√ SFYVYSRVKLNLSR	-	-	+	Antigen	Non-Allergen	Non-Toxin
Vaxign server						
o LVKPSFYVY	-	+	+	Antigen	Allergen	Toxin
o SLVKPSFYV	-	+	+	Antigen	Allergen	Toxin
√ RVKNLNSSR	-	+	-	Antigen	Non-Allergen	Non-Toxin
o SLVKPSFYVY	-	+	-	Non-Antigen	Discontinued	Discontinued
√ YVYSRVKNL	-	+	-	Antigen	Non-Allergen	Non-Toxin
√ VTLAILTAL	-	+	-	Antigen	Non-Allergen	Non-Toxin
√ TLAILTALR	-	+	-	Antigen	Non-Allergen	Non-Toxin

√	VTLAILTA	-	-	+	Antigen	Non-Allergen	Non-Toxin
o	YSRVKNLNS	-	-	+	Antigen	Allergen	Toxin

T-cell epitopes were identified as best epitopes based on number of alleles and antigenicity score. * Related

** Unrelated

√ Selected epitopes for vaccine construction

o Unselected epitopes for vaccine construction.

Table 2: Docking results

Vaccine Epitopes	Weighted Score of docked complex with						
	HLA-A*01:01	HLA-A*24:02	HLA-B*07:02	HLA-A*11:01	HLA-A*03:01	HLA-DPA1*01:03	HLA-DRB1*01:01
VTLAILTAL	-732.3	-765.2	-662.7	-644.7	-682	-----	-----
TLAILTALR	-700	-626.2	-587.3	-601.5	-635.8	-----	-----
SFVSEETGT	-650.2	-626.4	-523.8	-548.9	-503.3	-----	-----
LVTLAILTA	-777.3	-777.3	-641.3	-660.8	-682	-----	-----
NVSLVKPSFYVYSRVK	-----	-----	-----	-----	-----	-859.9	-972.9
SFYVYSRVKNLNSSRVPD	-----	-----	-----	-----	-----	-959.7	-998.1

The weighted scores of the lowest energy docked structures were based on the cluster size of the most populated cluster.

Figures

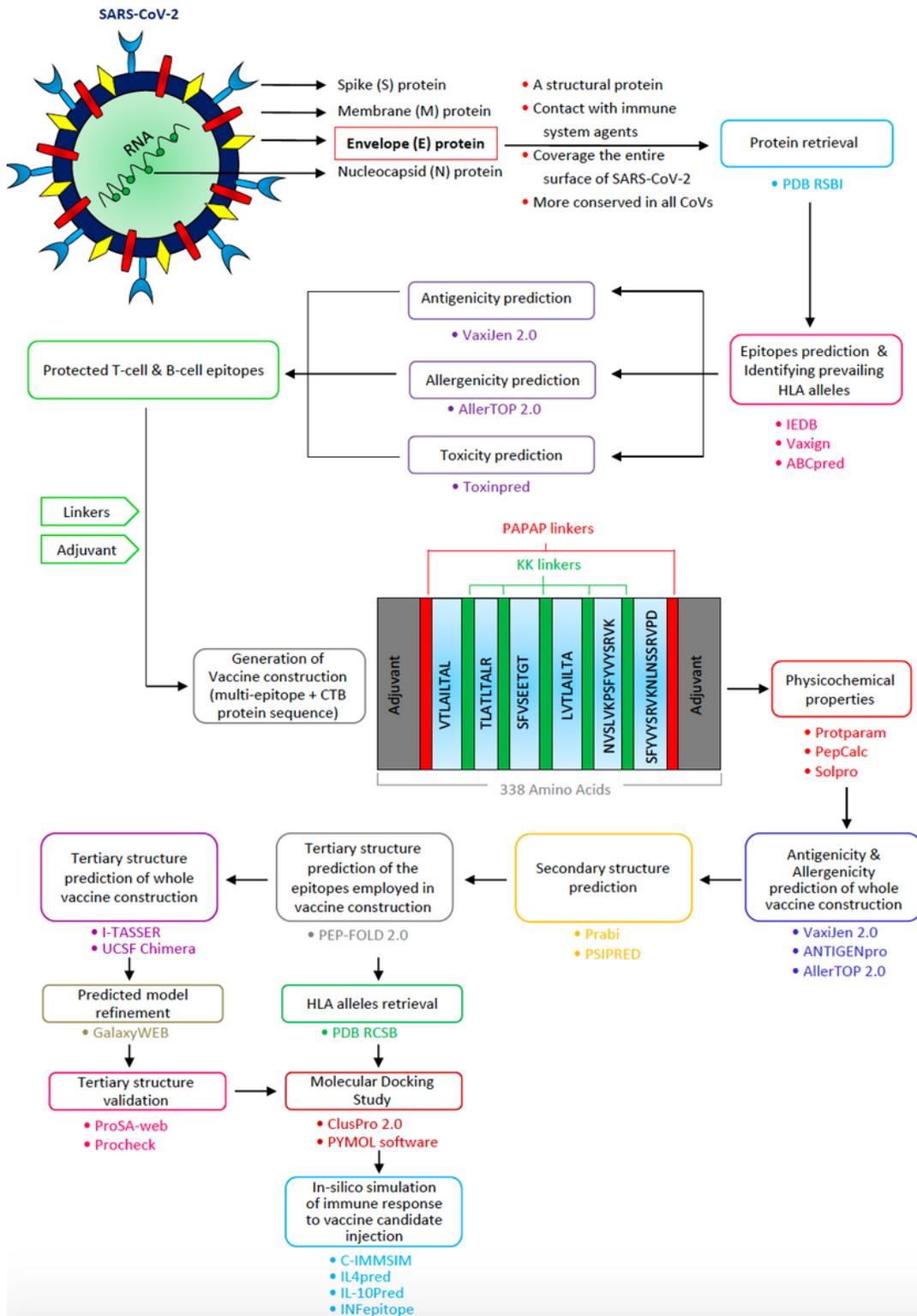


Figure 1

From Envelope (E) protein of SARS-CoV-2 to multi-epitope vaccine Schematic description and the study workflow of in-silico design a multi-epitope vaccine against envelope (E) protein of SARS-CoV-2

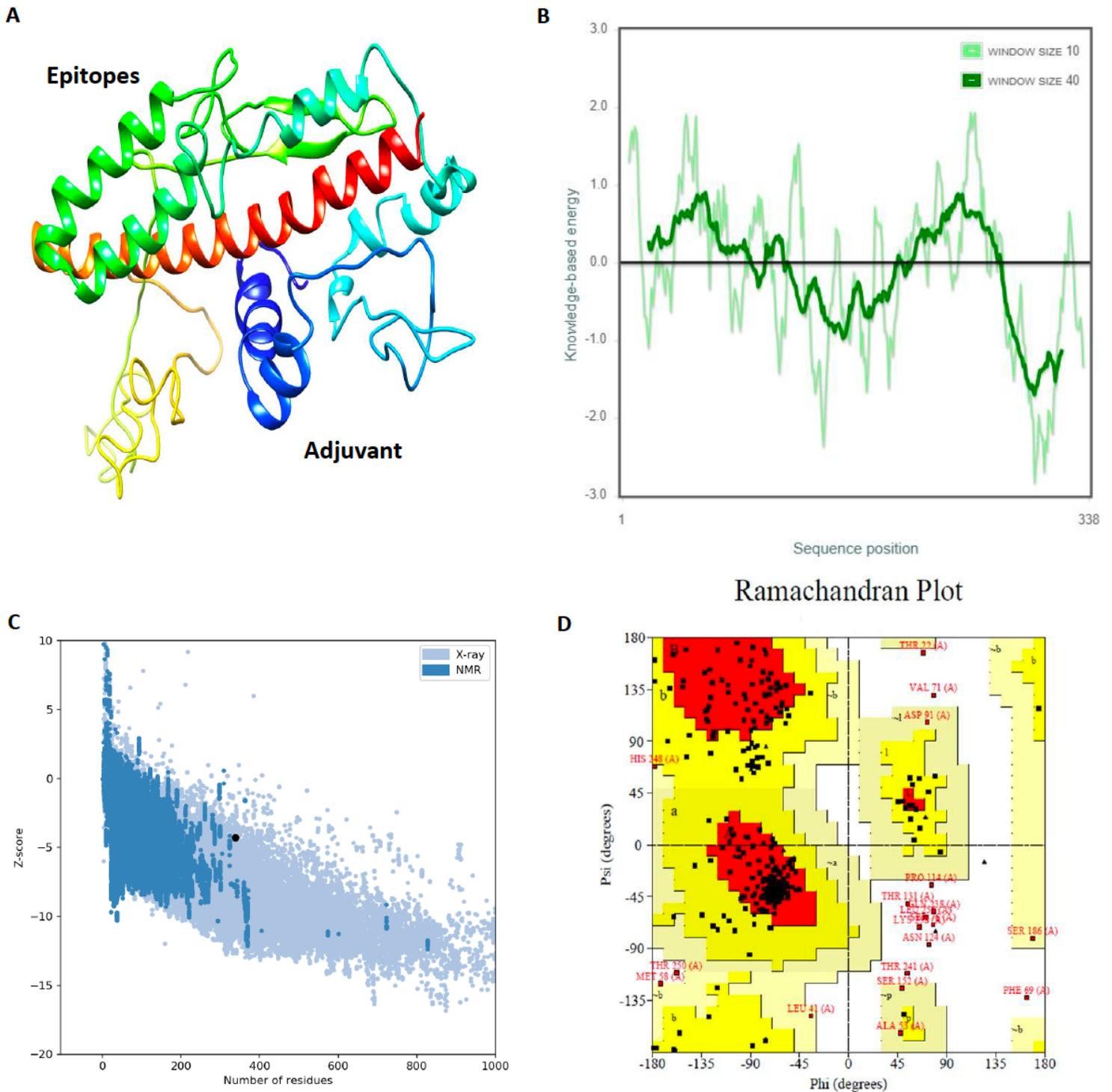


Figure 3

Homology modeling and validation of 3D structure (A) The final protein tertiary structure was built by the I-TASSER server. This presentation of the predicted model was performed by Chimera visualization tool. (B) The local quality model of predicted model. (C) Based on the ProSA-web, the Z-score of predicted model is -4.3 (D) The Ramachandran plot of predicted model

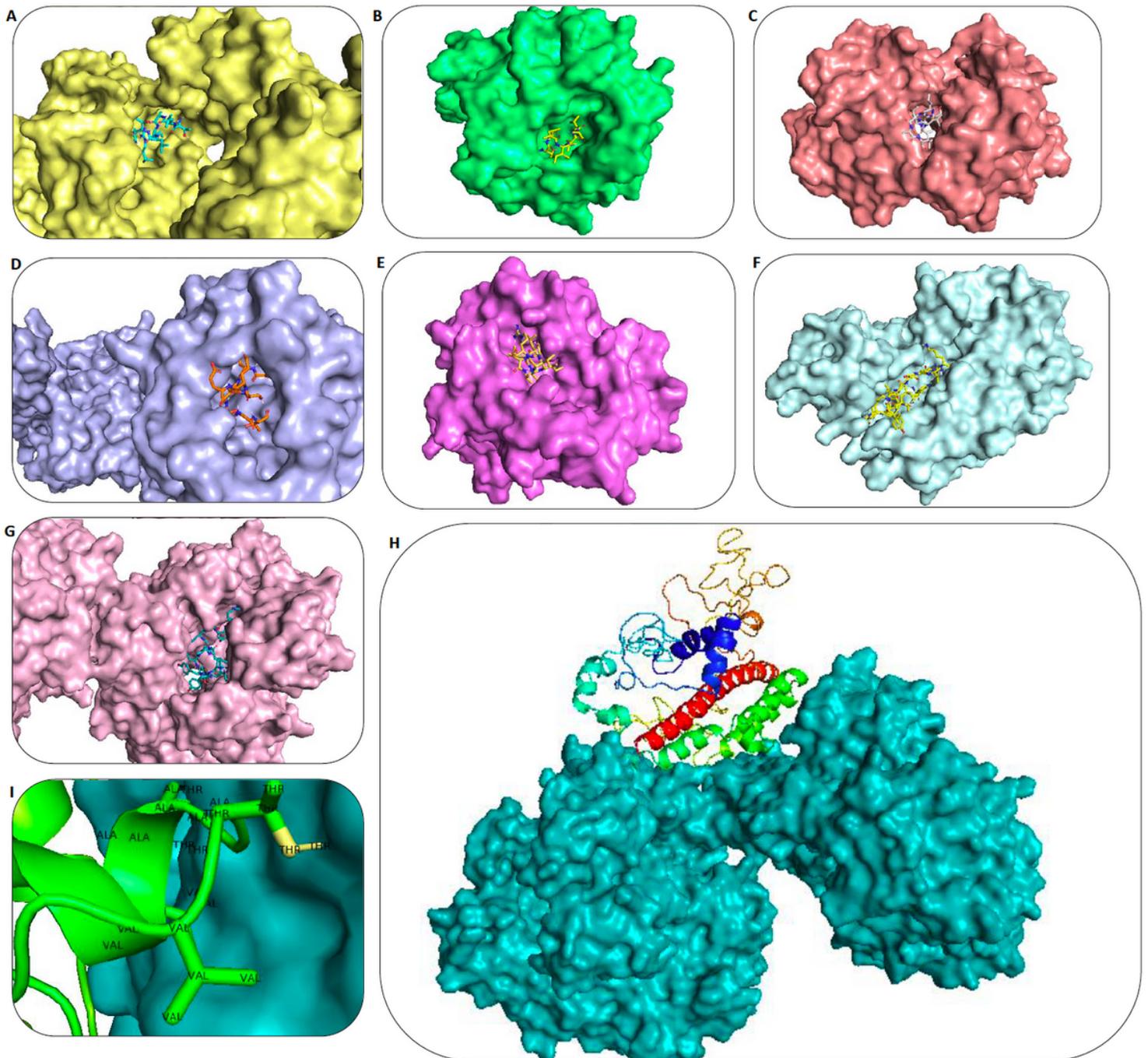


Figure 4

Molecular docking studies (A) TLAILTALR epitope (blue) and HLA-A*01:01 protein (cream). (B) LVTLAILTA epitope (yellow) and HLA-A*24:02 protein (green). (C) VTLAILTAL epitope (gray) and HLA-B*07:02 protein (pink). (D) SFVSEETGT epitope (orange) and HLA-A*11:01 protein (blue). (E) TLAILTALR epitope (yellow) and HLA-A*03:01 protein (purple). (F) NVSLVKPSFYVYSRVK epitope (yellow) and HLA-DPA1*01:03 protein (light blue). (G) SFYVYSRVKLNSSRVDPD epitope (blue) and HLA-DRB1*01:01 protein (pink) (H) Whole vaccine peptide (colorful) and HLA-DRB1*01:01 (teal blue) (I) The active residues (green) of the vaccine peptide in docked complex with HLA-DRB1*01:01 (teal blue).

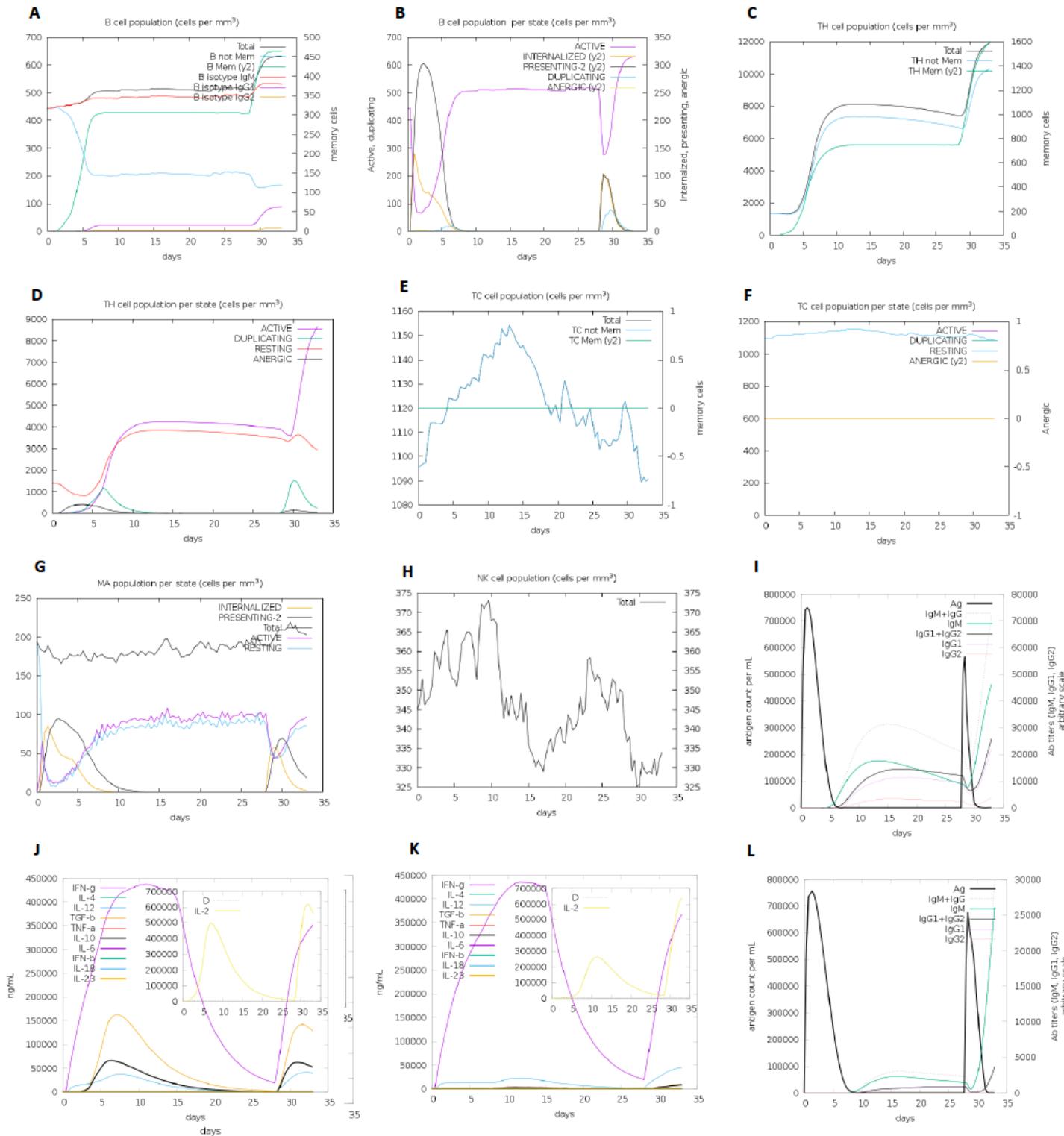


Figure 5

In-silico immune response simulation to the injection of candidate vaccine (without adjuvants) and control protein by C-IMMSIM server. The simulation is performed for three times of injection in the time steps of 1, 84, and 100. Each time step is equal to 8 hours. (A) B-cell population. (B) B-cell population per state. (C) T helper (TH) cell population. (D) TH cell population per state. (E) T cytotoxic (TC) cell population. (F) TC cell population per state. (G) Macrophage (MA) cell population. (H) Natural killer (NK)

cell population. (I) Immunoglobulins. (J) Cytokines. (K) Cytokines after the protein control injection. (L) Immunoglobulins following the protein control injection

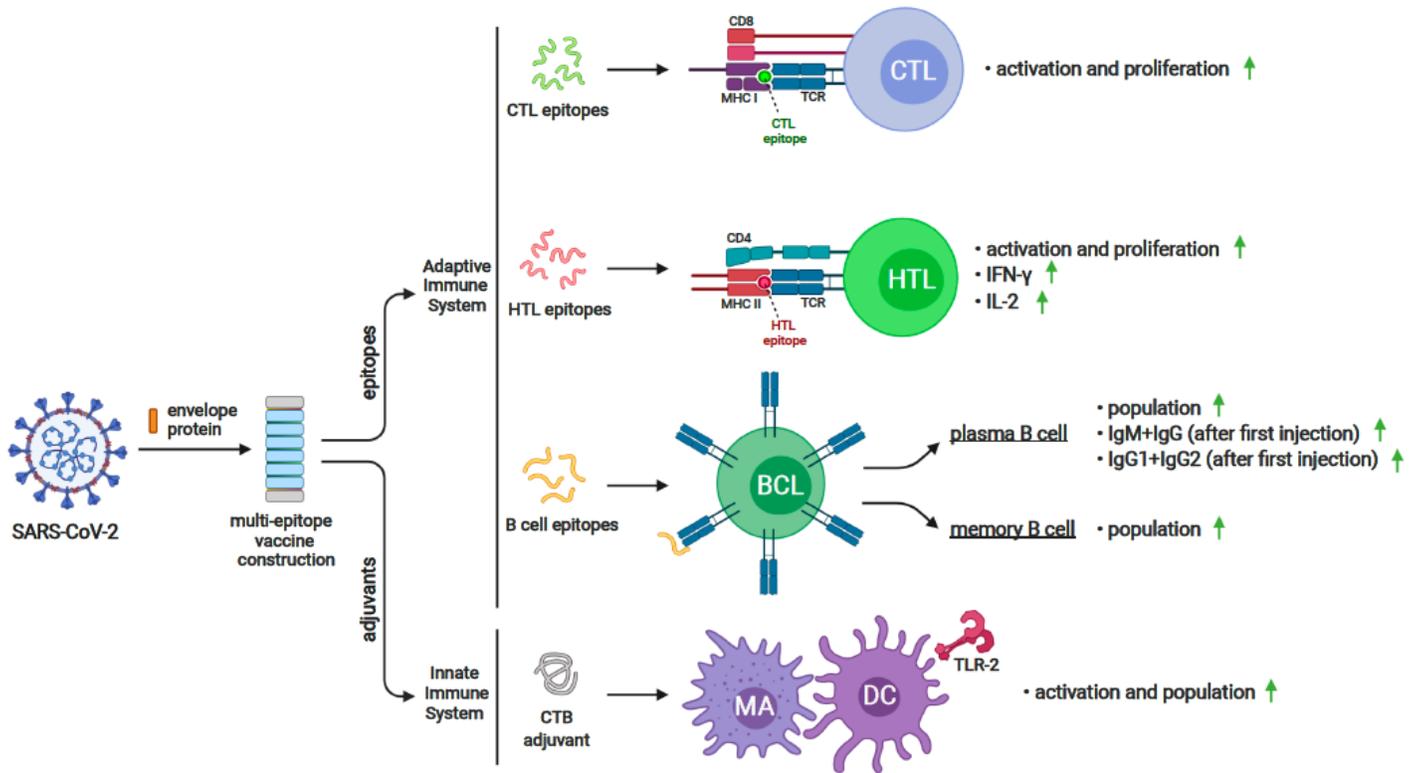


Figure 6

Graphical feature of immune response simulation to the injection of vaccine candidate against E protein of SARS-CoV-2 CTL: cytotoxic T lymphocyte cell , HTL: helper T lymphocyte, BCL: B-cell lymphoma, MA: Macrophage cell, DC: Dendritic cell, TLR-2: Toll Like Receptor 2, IgM: Immunoglobulin M, IgG: Immunoglobulin G, INF- γ : Interferon gamma, IL-2: Interleukin-2, CTB: cholera toxin B subunit, TCR: T-cell receptor, MHC I: major histocompatibility complex class I , MHC II: major histocompatibility complex class II, CD4: cluster of differentiation 4, CD8: cluster of differentiation 8 The image was created by using www.biorender.com illustrator tool.