

# Applying high throughput and comprehensive immunoinformatics approaches to design a trivalent subunit vaccine for induction of immune response against human emerging coronaviruses SARS-CoV, MERS-CoV and SARS-CoV2

**Abolfazl Rahmani**

Babol University of Medical Science Faculty of Medicine

**Masoud Bae**

Babol University of Medical Science Faculty of Medicine

**Kiarash Saleki**

Babol University of Medical Science Faculty of Medicine

**Saeed Moradi**

Babol University of Medical Science Faculty of Medicine

**Hamid Reza Nouri** (✉ [nourhr851@gmail.com](mailto:nourhr851@gmail.com))

Babol University of Medical Science Faculty of Medicine

---

## Research

**Keywords:** Immunoinformatic, MERS-CoV, SARS-CoV, SARS-CoV-2, Subunit vaccine

**Posted Date:** October 26th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-92515/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

## Background

Coronaviruses (CoV) cause diseases such as severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS) and Coronavirus disease 2019 (COVID-19). Therefore, this study was conducted to contrast a trivalent subunit vaccine against SARS, MERS and COVID-19. The CTL, HTL, MHC I, and IFN- $\gamma$  epitopes were predicted. Moreover, to stimulate strong helper T lymphocytes (HTLs) responses, Pan HLA DR-binding epitope (PADRE) was used. Also, for boosting immune response,  $\beta$ -defensin 2 was added to the construct as an adjuvant. Furthermore, TAT was applied in the vaccine to facilitate the intracellular delivery.

## Results

Based on the predicted epitopes, a trivalent multi-epitope vaccine with a molecular weight of 74.8 kDa as a strong antigen, a non-allergenic, and soluble protein was constructed. Furthermore, analyses validated the stability of the proposed vaccine. The binding affinity of the vaccine construct with the TLR3 was confirmed by molecular docking and, stability of the docked complex was simulated. The predicted epitopes demonstrated strong potential to stimulate T and B-cell mediated immune responses. Furthermore, codon optimization and *in silico* cloning guaranteed increased expression.

## Conclusions

In this work, immunoinformatics investigations demonstrated that this next-generation approach may provide a new horizon for the development of a highly immunogenic vaccine against SARS-CoV, MERS-CoV, and SARS-CoV-2.

## Background

Coronaviruses (CoVs) are positive-strand viruses from the family Coronaviridae that are named according to their crown-like shape [1]. CoVs have caused major outbreaks in recent years [2, 3]. Various human CoVs have spread globally, causing mostly mild illness, while others are highly pathogenic. The first severe acute respiratory syndrome (SARS)-CoV emerged in china in 2001 and resulted in a global outbreak. However, this outbreak was successfully contained. In 2012, another CoV variant named Middle East respiratory syndrome (MERS)-CoV emerged in Jeddah, Saudi Arabia. However, this outbreak never perfectly diminished, and hospitals and communities outbreaks continue to occur annually[4]. Recently, a novel coronavirus, SARS-CoV-2 initiated an outbreak wich turned into a pandemic, raising global consternation. This newly surfaced virus that caused a new disease named coronavirus disease (COVID-19) has a considerably lower mortality rate. However, the rapid transmission of the virus has turned COVID-19 into a fatal pandemic, which indicates a raising need for novel treatments and preventive methods.

Efforts have been made for vaccine and drug design regarding CoVs. As a result, some vaccines are currently being tested in phase I clinical trials for SARS-CoV-2 in the USA (NCT04283461) and China (NCT04313127). Despite the efforts, outcomes of therapeutics and preventive approaches aimed at CoVs, in particular vaccines

have not been adequately successful. Several aspects of CoVs may highlight the importance of preventing. Notably, various subtypes of CoVs, including some subtypes like MERS-CoVs that have caused past outbreaks, have remained relevant [5].

CoV proteins can be classified into two subgroups, structural and non-structural proteins (nsps). The genome of CoVs is surrounded by structural proteins comprising spike (S) protein, nucleocapsid (N), envelope (E), and membrane (M) proteins [6]. S protein exerts various functions and mediates cellular entrance of CoVs. Following attachment of viral envelope-anchored S protein to its receptor, CoVs can enter the cytoplasm. Protease breakdown of S protein is required for virus-cell fusion and the entry of genomic RNA [6–8]. S protein of SARS-CoV-2 is highly similar to SARS-CoV. Interestingly, polyclonal antibodies for SARS-CoV suppress spike-regulated entry of SARS-CoV-2 [9]. HCoV-NL63, a virus that only causes mild symptoms, also has S protein and recruits the angiotensin-converting enzyme 2 (ACE2) for entering the host cell, while showing very low pathogenicity. This indicates that additional elements are responsible for the pathogenicity of CoVs that have caused significant outbreaks. [10]. Furthermore, the N protein has been suggested as a promising candidate for cross-protective vaccine design. This is because of the high similarity of SARS-CoV, SARS-CoV-2, and MERS-CoV N proteins, while being significantly different from CoVs that mainly cause mild illness. [11]. The E and M proteins have been suggested as vaccine candidates in RV assessments. The E and M proteins exert crucial roles in the viral assembly of CoVs. Further, the N protein is essential for viral RNA synthesis. Deletion of E protein has inhibited the virulence of CoVs, and a number of researches have been directed at the efficacy of live attenuated vaccines that comprise recombinant SARS-CoV or MERS-CoV with a mutated E protein [12]. Also, M protein has been utilized for DNA vaccine construction against SARS-CoV [13]. Most vaccines comprise attenuated or inactivated whole virus and structural proteins. However, such vaccines could not confer complete protection and whole vaccines may cause adverse reactions. The non-structural proteins (nsp), a part of ORF1a polyprotein, may confer additional protection. Following S protein, nsp3 is the second-highest pathogenic element of SARS-CoV-2 protein [11].

Taken together, an efficient multipotent preventive solution against various subclasses of the rapidly emerging human CoVs remains to be discovered. This is highly important because of the recent CoV outbreak, which has raised an urgent need for the development of vaccine research. For this purpose, bioinformatics analyses have been utilized to aid the vaccine design process. Hence, we utilized *in silico* methods to design a trivalent novel cross-protective multi-epitope vaccine against SARS-CoV, MERS-CoV, and SARS-CoV-2 via spike (S) protein, nucleocapsid (N), envelope (E), membrane (M) protein, nsp3, and nsp8 antigens.

## Results

### Primary analysis of the retrieved sequences

In this study, spike (S) protein, nucleocapsid (N), envelope (E), membrane (M) protein, nsp3 and nsp8 antigens (totally 6 protein) were used for bioinformatics analysis to find CTL, HTL, IFN- $\gamma$  and MHC I binder's epitopes. The highly conserved regions of  $\beta$ -defensin 2 were employed as an adjuvant because the vaccine requires a strong induction of T cell immune responses. The signal sequences were excluded from all constructed structures. The workflow of the designed work representing the overall procedures of vaccine design is shown in Fig. 1.

## **Cytotoxic T lymphocyte (CTL) prediction**

CTL epitopes were selected according to the highest specificity and lowest sensitivity for their adaptive immune receptor. Top three high score epitopes were chosen from spike S, N, E, M, nsp3 and nsp8 antigens to get a total of 18 final CTL epitopes (Table 1). On the other hand, we utilized the Immune Epitope Database (IEDB) tool to predict MHC-I peptide binders to human alleles HLA-B\*57:01, HLA-A\*26:01, HLA-A\*11:01, HLA-B\*35:01, HLA-A\*03:01, HLA-B\*15:01, HLA-A\*01:01, HLA-A\*11:01, HLA-A\*30:02, HLA-A\*23:01, HLA-B\*38:01, HLA-B\*07:02, HLA-C\*07:01, HLA-A\*03:01, HLA-A\*25:01, HLA-A\*31:01 and HLA-A\*29:02. In the end, CTL epitopes were analysed and IFN- $\gamma$  inducing epitopes were applied in final construct.

## **Helper T-lymphocyte (HTL) epitope prediction**

The HTL epitope prediction was done with the help of the IEDB online server for human HLA-II alleles including HLA-DPA1\*01:03, HLA-DPB1\*02:01, HLA-DQA1\*01:01, HLA-DQB1\*05:01, HLA-DRB1\*09:01, HLA-DRB1\*13:02, HLA-DRB1\*13:21, HLA-DRB1\*04:26, HLA-DRB1\*13:02, HLA-DRB1\*01:01, HLA-DQA1\*05:01, HLA-DQB1\*02:01, HLA-DPB1\*02:01, HLA-DRB1\*09:01, HLA-DRB1\*11:14, HLA-DQA1\*01:02, HLA-DQB1\*06:02, HLA-DPA1\*03:01, HLA-DPB1\*04:02, HLA-DPA1\*03:01, HLA-DPB1\*04:02, HLA-DRB1\*07:03 and HLA-DRB1\*01:01 for antigens. Finally, 18 epitopes were chosen from the predicted HTL epitopes according to their lowest percentile rank (Table 2). The lowest percentile rank reflecting their good binding affinity for the MHC-II receptor.

## **Construction of the subunit vaccine**

To design the final vaccine construct, the top-scored epitopes of CTL and HTL were joined together using HEYGAEALERAG linker. To facilitate cell penetration of the proposed vaccine, the TAT peptide of HIV (GRKKRRQRRPPQ) was added at the N-terminal of the construct. In addition, for increasing the immunogenicity of the vaccine, a 41 amino acid-long sequence of  $\beta$ -defensin 2 was added after TAT peptide as an adjuvant and then followed by the PADRE sequence (AGLFQRHGEGTKATVGEPV). The TAT peptide and  $\beta$ -defensin 2 were linked to the PADRE sequence through GPGPG and EAAAK linkers, respectively. The order of the final construct is summarized in Fig. 2.

## **Population coverage**

The prediction of MHC-I and MHC-II based coverage of the applied epitopes in the constructed vaccine accomplished by IEDB analysis resources for the world population. As SARS-CoV-2 is global pandemic, hence, worldwide analysis has been accomplished. The world population coverage of MHC-I and MHC-II found to be 88.77%, and 97.91%, respectively, that showed in Fig. 3.

## **Optimized codon and mRNA structure of the chimeric vaccine**

We use the JCAT online server to obtain the codon-optimized DNA sequence of the proposed vaccine to verify for cloning and expressing in *E. coli* strain K12 as a host. The optimized codon sequences had 2091 nucleotides long, GC-content was reached to 52.9 % and its codon adaptation index was raised to 0.98. The graphical illustration of the optimized gene was shown in Fig. 4A. In addition, the mfold server was used to get free energy related to the whole mRNA structure and its 5' end. Its details are shown in Table 3. As shown in Fig.

3B, C, and D, the minimum free energy of the predicted secondary RNA structure was  $\Delta G = -595.30$  kcal/mol at 37 °C without hairpin nor pseudoknot at the 5' side.

### **Evaluation of allergenicity, antigenicity, physicochemical parameters and solubility of the vaccine construct**

Unveiling allergenicity of the proposed vaccine was a crucial consideration to avoid allergic reactions. The resulting output of the Allpred online server determined the non-allergenic behavior of the vaccine constructs with a score of -0.59 (Threshold=-0.4). Likewise, the antigenicity of the construct was predicted by the VaxiJen v2.0 online server, and it confirmed that the protein was a probable antigen with an overall prediction score of antigen 0.4996. Furthermore, the physicochemical parameters were evaluated, and the molecular weight of vaccine protein was found to be 74.8 kDa which is good to support the antigenic nature of the vaccine construct. The predicted instability index of the proposed vaccine was 39.09, which is below 40 hence, classifying the protein as stable. The isoelectric point (pI) of the construct was 9.30 and it shows the basic nature of the vaccine.

The estimated half-life (in-vitro) in mammalian reticulocytes was found to be 30 h, while *in vivo* studies in *Yeast* and *E. coli* more than 20 and 10 h, respectively. In addition, the aliphatic index was 80.26, which determines the thermostable nature of the proposed vaccine, and the grand average of hydropathicity obtained to be 0.063, which proves the hydrophilic nature of the constructed vaccine. The value for the solubility tendency of the proposed vaccine upon overexpression in *E. coli* host was 0.55.

### **Secondary and tertiary structure of subunit vaccine**

PSIPRED server was employed to predict the secondary structure of the proposed vaccine (Fig.5). This server showed that the vaccine is made of 43.76% alpha-helix, 16.93% extended strand, and 39.31% random coil. The tertiary structure was predicted as 3 domain structures using the galaxyWEB web server which is a template-based protein structure modeling web server that resulted in a modeled protein structure (Fig. 6). The 1FD3\_C, 5MKE\_A, and 5MKF\_A were used as the best-suited template to model the tertiary structure. All 697 amino acid residues were modeled, while the 70-92, 681-688, and 690-697 amino acids of the residues were unreliable local regions.

### **Refinement and validation of the 3D structure**

The generated 3D structure was submitted for the refinement process by the GalaxyRefine online server to enhance the quality of predicted 3D modeled structure beyond the accuracy. Furthermore, the refined model from GalaxyRefine was analyzed by the ramachandran plot. Validation of 3D structure by ramachandran plot before refinement demonstrated 540(92%), 37 (6.3%), and 9(1.5%) of residues were placed in favored, additionally allowed and disallowed regions, respectively (Fig. 7A). In the refined model, 550 (93.7%), 26 (4.4%), and 7 (1.2%) of residues were placed in favored, additionally allowed and disallowed regions, respectively (Fig. 7B).

### **Prediction of intrinsic protein disorder**

DisEMBL online server identified and annotated disordered regions. Amino acids in the input sequence are considered disordered by remark-465 definition in the following regions: 1-23, 464-478, 689-697. Based on

IUPred results, amino acids were considered disordered when the confidence score is higher than 0.5. Disordered regions were also presented in Fig. 8.

### **Predicted B-cell epitopes**

Ellipro software and the Discotope online server were utilized to predict the linear B-cell and discontinuous B cell epitopes, respectively. The linear B-cell and discontinuous B cell epitopes on the 3D structure are shown in Fig. 9A-G and Fig. 9H-L, respectively. Moreover, epitopes predicted in modeled vaccine according to various parameters were listed in Table 4.

### **Vaccine protein disulfide engineering**

Disulfide engineering leads to an increase in vaccine stability with mutating the residues present in high mobility areas with cysteine. 76 residue pairs were found to have the ability of disulfide bond formation. But, when all these residue pairs were more analyzed on the other parameters like energy on chi3 value and B-factor, only 1 pairs of residues provided the disulfide bond formation. Those residue pairs were GLY360-SER368 and ASP580- GLY601 (Fig. 10). These two residues were mutated to the cysteine. The energy value analyzed for the residue screening was less than 2.2 and the value of chi3 was between -87 to +97.

### **Molecular docking of vaccine construct with the TLR-3**

Protein-protein docking was done with the help of the Hdock online server between the final vaccine construct as a ligand and the TLR-3 as a receptor. We have procured 20 models and the model with the lowest binding energy was selected. The Docking score for the best acceptable model was -326.86, whereas the ligand RMSD (Å) was 205.96. PatchDock server was used to reaffirm the docking between the final vaccine construct and TLR-3. Top 20 models were provided by the server and the model with the top-ranked and lowest binding energy was chosen. The obtained output encompasses the score (17080), Area (2493.80), ACE (408.17), and transformation (-2.83 -0.57 -1.83 -49.29 3.75 160.90). The docked complex achieved from the server has been shown in Fig. 11A and B.

### **Molecular dynamics simulation results**

MD simulation was used to analyze the physical movements of the atoms and molecules of the proposed vaccine. At the initial step of MD simulation, energy minimization was done and several parameters including pressure and temperature were measured. Regular mode investigation allowed the demonstration of the stability of proteins and the large scale mobility. The iMODS server performed such an investigation based on internal coordinates of the receptor-ligand complex (Fig. 12A). In 3D model, the direction of each residue was given through arrows and the length of the line represented the extent of mobility. The proposed vaccine protein and TLR3 receptor were oriented towards each other. The eigenvalue found for the complex was 4.953620e-08 (Fig. 12B). The B-factor values deduced from normal mode investigation was analogous to *root mean square* (Fig. 12C). Hinges in the chain indicated the feasible deformability of the complex calculated through the complication of each residue (Fig. 12D). The variance related to each normal mode was inversely related to the eigenvalue. Covariance matrix described the occlusion between pairs of residues was correlated, uncorrelated or anti-correlated motions were represented with red, white, and blue colors, respectively (Fig. 12 E). The result also provided an elastic network model (Fig. 12F) that identified the pairs of atom connected by

springs. Each dot in the diagram was indicated one spring between the corresponding pair of atoms and colored based on the grade of rigidity.

### **In silico cloning**

To examine the expression of the recommended vaccine into a vector, cloning was done by utilizing the SnapGene server. As the expression system of the human and *E. coli* K12 differ from each other and the codon requires the adaptation as per the host expression system, the codon sequence that was optimized by JCat was reverse transcribed and modified as per the *E. coli* K12. The restriction enzyme sites *Not I* and *Xho I* were introduced at the N and C-terminal, respectively. The adapted codon was entered into the pET21b (+) vector and a recombinant plasmid with the length of 7514 base pair was obtained as displayed in Fig. 13.

### **Immune simulation**

C-ImmSim server results revealed a noticeable increase in the generation of immune responses. The humoral immune response was identified by high levels of IgM along with IgG + IgM (Fig. 14A). Estimation of the B cell population uncovered an increase in B memory cells and B-cell isotype IgM with a corresponding reduction in antigen concentration (Fig. 14B). In addition, the high response was observed in the TH1 and TC cell populations with corresponding memory development (Fig. 10C, D). This profile identified the development of immune memory and consequently enhanced clearance of the antigen upon subsequent exposures. In contrast, regulatory T cells were estimated in low level (Fig. 14E). Moreover, these predictions were consistent with the induced level of IFN- $\gamma$  produced after immunization with the proposed vaccine (Fig. 14F).

## **Discussion**

Design novel vaccines for preventing the ever-rising global burden of disease is an important issue that cannot be ignored. Coronaviruses (CoV) are a wide family of zoonotic viruses that reason illness ranging from the common cold to more severe diseases such as Severe Acute Respiratory Syndrome (SARS-CoV) Middle East Respiratory Syndrome (MERS-CoV) and SARS-CoV-2 [14, 15]. While many coronavirus vaccine studies targeting various structural proteins were conducted, most of these endeavors finally stopped soon after the outbreak of SARS and MERS [11].

In the present study, we purposed to design an epitope-based peptide vaccine with structural and nonstructural proteins against the more severe form of human coronavirus SARS-CoV-2, SARS-CoV, and MERS-CoV via immunoinformatics strategies utilizing various in silico tools. Advancement in genomics has alternated the procedures and concepts to vaccine candidate design and chosen [16]. Reverse vaccinology, a novel method to merge immunogenetics and immunogenomics via immunoinformatic has been used tremendously to introduce novel vaccine [17, 18]. Thus, in this study, we developed a novel trivalent subunit vaccine from immune-dominant peptides of spike (S) protein, nucleocapsid (N), envelope (E), membrane (M) protein, nsp3, and nsp8 antigens. In our suggested construct, the CTL, MHC-I, MHC-II, epitopes along with IFN- $\gamma$  containing epitopes were predicted and used in our designed vaccine. Cytotoxic CD8 T lymphocytes (CTL) limit the spread of infectious pathogens via recognizing and killing infected cells or secreting particular antiviral cytokines [19, 20]. Thus, T-cell epitope-based vaccination is a unique role of eliciting a potent immune response against infectious pathogens such as viruses [21]. Helper T lymphocyte (HTL) responses act an important role in the

induction of both humoral and cellular immune responses. Therefore, HTL epitopes are believably to be a key part of prophylactic and therapeutic vaccines [22]. Several pieces of research introduced IFN- $\gamma$  as the effective cytokine of both the adaptive and natural immune system eliciting antiviral, immune regulatory as well as anti-tumor activities. The secretion of IFN-gamma is the key part of the Th1 response critical for the rein of pathogens such as viruses [23]. Furthermore, once the development of the novel multi-epitope vaccine begins, there will be great concern about its immunogenicity. Thus, adjuvants were established and attached to vaccine construction to enhance its efficacy. There are some complications with the most routine aluminum adjuvants such as anthrax, macrophagic myofasciitis, and post-vaccination phenomena [24]. Recently, the role of TLRs in the relation of innate and adaptive immunity for vaccine design has been highlighted. TLRs are a subgroup of pattern recognition receptors on antigen-presenting cells (APCs), which activating via an appropriate TLR agonist will elicit cellular and humoral reactions as well as innate responses [25, 26]. There are multiple ways to induce TLRs.  $\beta$ -defensin 2 which induce signal transduction via TLR-4 regulates adaptive immunity during infection [27]. Self-destructive signaling to delete activated APCs and thus modifying the adaptive immunity is also possible through utilizing  $\beta$ -defensin 2[28].

In our vaccine construction human  $\beta$ -defensin 2 serves as an adjuvant that promotes the expression of primary immune-inducing and antiviral compounds level [29]. To increase the quality of vaccine delivery, a cell-penetrating peptide, namely TAT, from HIV was engaged to involve T cells via APCs [30]. PADRE T helper epitope can bind to a prominent number of HLA-DR types with an affinity spanning from intermediate to high [31]. Also, the YQVNNLEEI epitope located in a conserved region of EBOV NP added to the C terminal of the vaccine construct. Some studies indicated that this epitope has a protective CTL epitope against COVID-19 [32].

Different parts of the vaccine structure were connected through AAY, GPGPG, HEYGAEALERAG, EAAAK, and GGGG linkers. These sequences exert a critical role in the 3D structure and subsequent exposure of candidates [33]. The AAY linker was used to join HTL epitopes. AAY enhances two subsequent epitopes cleavage [34]. EAAAK linker was used to link PADRE and adjuvant to decrease the interaction with other vaccine domains and cause more effective separation. We used the GGGG linker to connect CTL epitopes. GGGG linker containing sequences are powerful stimulants of the immune response [35, 36]. We utilized the GPGPG linker to connect TAT and adjuvant. Through polypeptide and DNA vaccination, GPGPG can act as a prominent HTL inducer. This spacer, however, as an assistant for epitope presentation by HLA-II, inhibits generation of junction [37, 38]. The HEYGAEALERAG linker provides the specific cleavage target for both the lysosomal and proteasomal degradation system. The twelve-residue peptide consists of all five appropriate cleavage sites, A7-L8, A5-E6, Y3-G4, R10-A11, and L8-E9, that are determined for eukaryotic proteasome complexes [39]. For joining the CTL epitopes to HTL epitopes and HTL epitopes to the PADRE and YQVNNLEEI epitope to the histidine chain, the HEYGAEALERAG linker was used.

Allergenicity is one of the main problems in vaccine development. Recently, most vaccines stimulate the immune system into an allergic reaction [40]. based on the WHO/FAO, if a sequence has a recognize of at least six contiguous amino acids over a window of 80 amino acids (0.35% sequence identity) to identified allergens, it is considered to be potentially allergenic. In this study, AlgPred online server was applied for the allergenicity assessment of the designed vaccine. The result of the allergenicity assessment showed that the vaccine isn't allergen. Thus, the subunit chimeric vaccine was constructed with 697 residues in length. Restriction sites, instability elements, and all the cis-acting sites were omitted from the construct which significantly interfering

with cloning. Moreover, codon optimization was performed to facilitate the high-level expression of the multi-epitope vaccine in *E. coli* and improve translational and transcriptional output. mRNA secondary structure is a serious parameter in the expression of subunit vaccines. The outputs from mRNA prediction by Mfold online server showed that the mRNA had enough stability for impressive translation in the host. Studies show that higher stability result fully leads to higher expression rate. Physico-chemical factors analysis via ProtParam online server showed that an estimated half-life of the subunit vaccine was 30 hours in mammalian reticulocytes, more than 20 hours in yeast, and more than 10 hours in *E. coli*. Furthermore, the instability index was 39.09 that indicates a subunit vaccine is a stable protein. Also, the tendency of vaccine solubility upon overexpression in *E. coli* was computed 0.95% showing an admissible percentage of solubility in an overexpressed state. Outputs of the primary model with the Galaxyweb server showed that the Galaxyweb program model was applied to generate the primary 3D structure of the protein vaccine. Obtained results showed that 43.76% alpha-helix, 16.93% extended strand, and 39.31% random coil structure was in the tertiary model. Besides, the refinement of modeled structure was carried out in GalaxyRefine online server via molecular dynamics simulation. After the refinement, applying via the GalaxyRefine server resulted in modeling showed a high-quality tertiary model. The analysis of all discontinuous and continuous B-cell epitopes showed that the identified epitopes on the protein surface could interact comfortably with antibodies, and they were commonly flexible. DisEMBL online server was utilized to specify unstructured regions and disorders in the subunit vaccine. Obtained results indicated that disorder regions were sited among positions 1–23, 464–478, 689–697. An intrinsically disordered protein (IDP) is classically described as a protein with absenting a fixed or well-structured 3D fold. IDPs include a spectrum of states from wholly unstructured to partially structured and contain (pre-) molten globules, random coils, and large multi-domain proteins connected via flexible linkers. Disordered areas are mostly functional. Many disordered segments fold on binding to their biological targets and some others organize flexible linkers causing an assembly of the macromolecular array [41]. Then, the docking process between TLR3 and the designed vaccine was accomplished and the foremost docked model was selected and investigated via MD simulation to compute the stability of the vaccine–protein complex. Outputs of MD simulation support that the subunit vaccine can interact with TLR3 protein befittingly. The interactions between the two molecules were modified during the MD simulation.

The infectious form of the attenuated vaccine may reverse or inactivation process may fail to repress the virulence property, In the case of conventional vaccines [42, 43]. Furthermore, some recent clinical trial studies enhanced safety concerns while corroborating on conventional method for vaccine development [44]. Some when the cellular lines employed for the generation of the vaccine may be contaminated via microorganisms that do not produce any cytopathic effect while undergoing replication along with the vaccine virus [45, 46]. These contaminations perhaps proved dangerous for the host, when the quality control process failures too. However, the recombinant multi-epitope vaccine does not carry any risk of the emergence of virulence property as it contains the antigenic sections only (not the entire virus itself). Moreover, several instruments have been employed in the current study to accredit the non-allergic and nontoxic behavior of the designed vaccine construct. However, the predicted bioinformatics results were based on the various investigation of sequences and different immune servers. We offer future wet lab-based analyses employed in vivo models for the experimental validation of our subunit vaccine.

## Conclusion

In the current study, we work to develop an efficient trivalent subunit vaccine against CoVs via different *in silico* servers. Immunoinformatics analysis indicated that this vaccine was able to induce both humoral and cellular immune responses effectively due to the presence of several epitopes from a combination of antigens, PADRE, and adjuvant. Thus, this subunit vaccine can conceivably be utilized for prophylactic usages against emerging human pathogenic coronaviruses SARS-CoV, MERS-CoV, and SARS-CoV2.

## Methods

### Retrieval of protein sequences

A thorough survey was done to identify the crucial immunogenic proteins of the SARS-CoV-2, SARS-CoV, and MERS-CoV, respectively. Six proteins were selected and the amino acid sequence of nsp3 (Accession no. YP\_009725299.1, ACZ72252.1:819-2740 and YP\_009047231.1), nsp8 (Accession no. YP\_009725304.1, ACZ72252.1:3920-4117 and YP\_009047236.1), E protein (Accession no. QIH45055.1, ABI96961.1 and ATG84893.1), M protein (Accession no. QII57163.1, P59596.1, and AXN73531.1), S protein (Accession no. QII57161.1, AAQ01597.1, and AXN73514.1) and N protein (Accession no. QIH45060.1, P59595.1 and ATG84895.1) were retrieved from the Universal Protein Resource (UniProt) at <http://www.uniprot.org> and the National Center for Biotechnology Information (NCBI) at <https://www.ncbi.nlm.nih.gov/protein/> saved in FASTA format.

### Cytotoxic T lymphocytes (CTL) and IFN- $\gamma$ inducing epitopes prediction

Choosing CTLs epitopes is one of the most important elements of vaccine development, because CTLs kill virus-infected or damaged cells by recognizing the epitope presented by MHC-I molecule. After all, the initial step to stimulate the immune system is presenting an antigen on MHC-I. CTLPred (<http://crdd.osdd.net/raghava/ctlpred/>) is a freely accessible web server for the prediction of CTL epitopes. This server is based on the artificial neural network (ANN), support vector machine (SVM), the consensus and, the combined prediction methods. In addition, IFN- $\gamma$  plays an essential role in the innate and adaptive immune response by stimulating natural killer cells and macrophages and provides an elevated response to MHC antigens. IFN- $\gamma$  epitope server (<http://crdd.osdd.net/raghava/ifnepitope/scan.php>) was also used to predict interferon-positive sequences in the selected antigens.

### Helper T-lymphocyte (HTL) epitope prediction

IEDB prediction tool (<http://tools.immuneepitope.org/mhcii/>) was employed to predict helper T cell (HTL) epitopes among the selected antigens and each peptide was 15mer in length. The epitopes were predicted for human HLA-II alleles as target species. We used the IEDB recommended method to achieve the best possible result.

### Trivalent subunit multi-epitope vaccine construct and gene optimization

The N-terminal sequence of the subunit vaccine starts with TAT peptide of HIV as cell-penetrating peptides (CPPs) and followed by the adjuvant ( $\beta$ -defensin 2) as adjuvant. In the following, PADRE was utilized as helper T lymphocyte epitopes. Then, HTL and CTL selected epitopes were added sequentially for each SARS-CoV-2, SARS-CoV, and MERS-CoV protein, respectively. The AAY, GPGPG, EAAAK, KK, and GGG amino acids

sequences were used as suitable linkers to attach selected epitopes to achieve the multi-epitope vaccine consists of an antigenic part of a pathogenic organism to elicit an immunological response. Finally, the DNA sequence of the subunit vaccine and its adaptation to *E. coli* codon usage was obtained via the JCAT server at <http://www.jcat.de>.

### **Population coverage**

IEDB population coverage tool (<http://tools.iedb.org/population/>) was used to figure out the scale of the population coverage of the final construct within worldwide population. Epitope/MHC restriction data were loaded on IEDB and, calculations were set to be done for MHC-I and MHC-II separate epitopes.

### **Physico-chemical parameters analysis**

The codon usage was analyzed by Java Codon Adaptation Tool (JCat) carry out the codon optimization. The Mfold server (<http://unafold.ma.albany.edu/>) was used to predict the second structure of RNA. Mfold online server predicts structures and the true positive base pairs with minimum  $\Delta G$  with its thermodynamic methods. In addition, the physical and chemical parameters that are associated with the vaccine were discovered. Some physicochemical characteristics of the designed structure, including amino acid composition, instability index, molecular weight, PI value, estimated half-life *in vitro* and *in vivo*, aliphatic index, extinction coefficient, and grand average of hydropathicity (GRAVY) were evaluated using ProtParam online server at <https://web.expasy.org/protparam/>.

### **Prediction of solubility, antigenicity and allergenicity**

SOLpro online server (<http://scratch.proteomics.ics.uci.edu>) was utilized to predict the propensity of a protein to be soluble upon overexpression in *Escherichia Coli*. This server uses a two-stage SVM architecture method to evaluate the protein solubility based on multiple representations of the primary sequence. In addition, prediction of allergenicity with high precision was discovered by AlgPred at <http://crdd.osdd.net/raghava/algpred/>. Furthermore, antigenicity of the proposed vaccine was assessed using another online server VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>). The accuracy of this server varies from 70% to 89% depends on the organism.

### **Prediction of secondary and tertiary structure**

The secondary structure of the predicted vaccine was carried out by using GOR IV and the PSIPRED online servers at <http://expasy.org/tool/gor4.html> and <http://bioinf.cs.ucl.ac.uk/psipred/>, respectively. The percentage of the helix (H), strand (S), and coils (C) were determined. Homology modeling of the final protein construct was done by GalaxyTBM at <http://galaxy.seoklab.org/index.html>. GalaxyTBM is a template-based tertiary structure prediction online server and it predicts protein structure from the sequence when experimental structures of homologous proteins are available as templates and refines loop or terminus regions by ab initio modeling.

### **Refinement and validation of the 3D modeled structure**

The tertiary structure of the ultimate vaccine construct received from the RaptorX server depends on the degree of similarity between available template structure options and the target protein. Therefore, the GalaxyRefine

server (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) was used to improve the modeled structure beyond the accuracy. The CASP10 assessed the GalaxyRefine server and reported that the GalaxyRefine is one of the best-performing methods to improve the local structure quality. In addition, SAVES v5.0 server at <http://servicesn.mbi.ucla.edu/PROCHECK/> was employed to validate the refined model and the Ramachandran plot of this structure was designed.

### **Prediction of intrinsic protein disorder**

DisEMBL 1.5 (<http://dis.embl.de/>) and IUPred (<http://iupred.enzim.hu/pred.php>) were utilized to predict intrinsic protein disorder and unstructured regions within a protein. DisEMBL is a public online server for predicting disorder in proteins. Also, DISOPRED3 was employed to predict intrinsic protein disorder and unstructured regions in our proposed vaccine.

### **Prediction of conformational and linear B-cell epitopes**

We utilize Discotope 2.0 server at <http://www.cbs.dtu.dk/services/DiscoTope/> to predict conformational epitopes for the modeled vaccine. It's using the calculation of surface accessibility estimated in terms of contact numbers and a new epitope propensity amino acid score to predict discontinuous B cell epitopes from the tertiary structure. The final scores are determined by combining the propensity scores of residues in spatial proximity and contact numbers. In addition, ElliPro online server at <http://tools.iedb.org/ellipro/> was employed to predict linear and discontinuous B-cell epitopes. ElliPro classifies the conformational epitopes according to their protrusion index values and clusters them based on distance  $R$  (Å), while the Discotope method utilizes a calculation of surface accessibility and epitope propensity amino acid score. Finally, conformational epitopes with high-score were selected.

### **Disulfide bond engineering**

Improving the stability of the modeled vaccine was an important step that was completed by utilizing a logical approach to disulfide bond formation. Disulfide bonds provide notable stability to protein structure and well-defined geometric conformation. Disulfide engineering was done by Disulfide by design v2.0 (DbD2) to design novel disulfide bonds in the vaccine protein construct[47].

### **Molecular docking between vaccine construct and TLR-3**

Molecular Docking was done to perceive the interaction between the receptor (TLR-3) and the ligand (final vaccine construct) at their stable form and to check the binding affinity between them. In humans, the TLR-3 receptor stimulates the immune system by activating the NF- $\kappa$ B signaling pathway followed by the production of different cytokines[48]. The protein-protein docking was performed by the Hdock online server at [http://hdock.phys.hust.edu.cn/\\_HDOCK](http://hdock.phys.hust.edu.cn/_HDOCK) is a docking server that efficiently integrates various components: sequence search, template selection, and model building. The predictive power of the HDOCK server can be also improved by ranking first of the template-based model. Moreover, the center of extremely populated clusters and the lowest binding energy of the receptor-ligand complex were predicted. After several rotations of ligand, the docked complex was selected based on the lowest binding energy. Moreover, re-affirmation was performed by the PatchDock online server at <https://bioinfo3d.cs.tau.ac.il/PatchDock/> to evaluate the interaction between the designed vaccine construct and the receptor. This server gives the output based on three

algorithms: Filtering and scoring, molecular shape representation, and surface patch matching. The output composed of rank, global energy, atomic contact energy, transformations, and PDB file complex.

### **Molecular dynamics simulation study**

The molecular dynamic simulation was employed to check the physical basis of the structure and function of the biological macromolecules. The iMODS server was used to explain the collective motion of proteins by analysis of normal modes (NMA) in internal coordinates. Essential dynamics extracts the correlated motions of proteins to understand the motions that are most essential to the activity of the protein and it is a strong tool and alternative to the costly atomistic simulation that can be contrasted to the normal modes of proteins to specify their stability. This server predicted the direction and magnitude of the immanent motions of the complex in terms of eigenvalues, deformability, covariance, and B-factors[49]. The structural dynamics of the ligand-receptor complex was analyzed.

### ***In silico* cloning**

The optimized sequence of proposed vaccine was reversly trascribed and to ensure the restriction cloning, *XhoI* and *Not I* restriction site were added at the N- and C-terminal of sequense, respectively. Then, we used the Snap-Gene restriction cloning module to insert the optimized sequence into pET21b (+) vector between defined restriction sites.

### **Immune simulation**

The C-ImmSim online server (<http://150.146.2.1/C-IMMSIM/index.php>) was used to determine the immune response profile of the constructed vaccine[50]. C-ImmSim simulates three sections that represent three different anatomical regions in mammals including the bone marrow, tertiary lymphatic organ, and the thymus. This is an important step for the simulation of the immune response because it determines immunogenicity. The binding of the epitope, which is the immunogenic part of an invading pathogen, together with activation and cooperation from T helper cells, is required to trigger an immune response in the affected host.

## **Declarations**

### **Ethics approval**

This study was approved by the Ethical Committee of Babol University of Medical Sciences

### **Consent for publication**

All authors agree to publish this study

### **Availability of data and material**

The datasets applied and analyzed during the current study are available from the corresponding author on reasonable request.

### **Conflicts of interest/Competing interests**

# The authors declare that they have no competing interests.

## Funding

This study was supported with the Deputy for Research and Technology of Babol University of Medical Sciences. They had no role in the design, collection, analysis and writing of the manuscript.

## Authors' contributions

Design and supervision of the study: Hamid Reza Nouri. Methodology performed by Abolfazl Rahmani , Masoud Baei, Kiarash Saleki and Saeed Moradi. Writing and correction of the manuscript was accomplished by all authors.

## Acknowledgements

We thank all staffs in Immunoregulation Research Center, Health Research Institute, Babol University of Medical Sciences.

## References

1. Tortorici MA, Velesler D: **Structural insights into coronavirus entry.** *Advances in virus research* 2019, **105**:93-116.
2. Azhar EI, Lanini S, Ippolito G, Zumla A: **The Middle East respiratory syndrome coronavirus—a continuing risk to global health security.** In *Emerging and Re-emerging Viral Infections.* Springer; 2016: 49-60
3. Habibzadeh P, Stoneman EK: **The novel coronavirus: a bird's eye view.** *Int J Occup Environ Med (The IJOEM)* 2020, **11**:1921-1965-1971.
4. Fehr AR, Perlman S: **Coronaviruses: an overview of their replication and pathogenesis.** In *Coronaviruses.* Springer; 2015: 1-23
5. Kleine-Weber H, Elzayat MT, Wang L, Graham BS, Müller MA, Drosten C, Pöhlmann S, Hoffmann M: **Mutations in the spike protein of MERS-CoV transmitted in Korea increase resistance towards antibody-mediated neutralization.** *Journal of Virology* 2018.
6. Rappuoli R, Bottomley MJ, D'Oro U, Finco O, De Gregorio E: **Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design.** *Journal of Experimental Medicine* 2016, **213**:469-481.
7. Dawson P, Malik MR, Parvez F, Morse SS: **What have we learned about Middle East respiratory syndrome coronavirus emergence in humans? A systematic literature review.** *Vector-Borne and Zoonotic Diseases* 2019, **19**:174-192.
8. Zumla A, Hui DS, Perlman S: **Middle East respiratory syndrome.** *The Lancet* 2015, **386**:995-1007.
9. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Velesler D: **Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein.** *Cell* 2020.
10. Hofmann H, Pyrc K, van der Hoek L, Geier M, Berkhout B, Pöhlmann S: **Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry.** *Proceedings of the National Academy of Sciences* 2005, **102**:7988-7993.

11. Ong E, Wong MU, Huffman A, He Y: **COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning.** *BioRxiv* 2020.
12. Shang W, Yang Y, Rao Y, Rao X: **The outbreak of SARS-CoV-2 pneumonia calls for viral vaccines.** *npj Vaccines* 2020, **5**:1-3.
13. Shi S-Q, Peng J-P, Li Y-C, Qin C, Liang G-D, Xu L, Yang Y, Wang J-L, Sun Q-H: **The expression of membrane protein augments the specific responses induced by SARS-CoV nucleocapsid DNA immunization.** *Molecular immunology* 2006, **43**:1791-1798.
14. Lu H, Stratton CW, Tang YW: **Outbreak of Pneumonia of Unknown Etiology in Wuhan China: the Mystery and the Miracle.** *Journal of Medical Virology.*
15. Hui D, IAzharE MT: **The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in Wuhan.** *China* 2020, **91**:264-266.
16. Ahluwalia PK, Pandey RK, Sehajpal PK, Prajapati VK: **Perturbed microRNA expression by Mycobacterium tuberculosis promotes macrophage polarization leading to pro-survival foam cell.** *Frontiers in immunology* 2017, **8**:107.
17. Poland GA, Ovsyannikova IG, Jacobson RM: **Application of pharmacogenomics to vaccines.** 2009.
18. Hasan M, Ghosh PP, Azim KF, Mukta S, Abir RA, Nahar J, Khan MMH: **Reverse vaccinology approach to design a novel multi-epitope subunit vaccine against avian influenza A (H7N9) virus.** *Microbial pathogenesis* 2019, **130**:19-37.
19. Garcia KC, Teyton L, Wilson IA: **Structural basis of T cell recognition.** *Annual review of immunology* 1999, **17**:369-397.
20. Sharmeen R, Hossain MN, Rahman MM, Foysal MJ, Miah MF: **In-vitro antibacterial activity of herbal aqueous extract against multi-drug resistant Klebsiella sp. isolated from human clinical samples.** *International Current Pharmaceutical Journal* 2012, **1**:133-137.
21. Shrestha B, Diamond MS: **Role of CD8+ T cells in control of West Nile virus infection.** *Journal of virology* 2004, **78**:8312-8321.
22. Alexander J, Fikes J, Hoffman S, Franke E, Sacci J, Appella E, Chisari FV, Guidotti LG, Chesnut RW, Livingston B: **The optimization of helper T lymphocyte (HTL) function in vaccine development.** *Immunologic research* 1998, **18**:79-92.
23. Schoenborn JR, Wilson CB: **Regulation of interferon- $\gamma$  during innate and adaptive immune responses.** *Advances in immunology* 2007, **96**:41-101.
24. Gherardi R, Aouizerate J, Cadusseau J, Yara S, Authier F: **Aluminum adjuvants of vaccines injected into the muscle: Normal fate, pathology and associated disease.** *Morphologie* 2016, **100**:85-94.
25. Dowling JK, Mansell A: **Toll-like receptors: the swiss army knife of immunity and vaccine development.** *Clinical & translational immunology* 2016, **5**:e85.
26. Steinhagen F, Kinjo T, Bode C, Klinman DM: **TLR-based immune adjuvants.** *Vaccine* 2011, **29**:3341-3355.
27. Amelin Y, Krot AN, Hutcheon ID, Ulyanov AA: **Lead isotopic ages of chondrules and calcium-aluminum-rich inclusions.** *Science* 2002, **297**:1678-1683.
28. Biragyn A, Coscia M, Nagashima K, Sanford M, Young HA, Olkhanud P: **Murine  $\beta$ -defensin 2 promotes TLR-4/MyD88-mediated and NF- $\kappa$ B-dependent atypical death of APCs via activation of TNFR2.** *Journal of*

*leukocyte biology* 2008, **83**:998-1008.

29. Kim J, Yang YL, Jang S-H, Jang Y-S: **Human  $\beta$ -defensin 2 plays a regulatory role in innate antiviral immunity and is capable of potentiating the induction of antigen-specific immunity.** *Virology journal* 2018, **15**:124.
30. Pouniotis D, Tang C-K, Apostolopoulos V, Pietersz G: **Vaccine delivery by penetratin: mechanism of antigen presentation by dendritic cells.** *Immunologic research* 2016, **64**:887-900.
31. Feng GD XX, Gao ML, Wang XF, Shu Z, Mu N, Gao Y, Wang ZL, Hao Q, Li WN, Li M.: **Therapeutic effects of PADRE-BAFF autovaccine on rat adjuvant arthritis.** *BioMed research international* 2014.
32. Herst CV BS, Sidney J, Sette A, Harris PE, Massey S, Brasel T, Cunha-Neto E, Rosa DS, Chao WC, Carback R.: **An Effective CTL Peptide Vaccine for Ebola Zaire Based on Survivors' CD8+ Targeting of a Particular Nucleocapsid Protein Epitope with Potential Implications for COVID-19 Vaccine Design.** . *Vaccine* 2020.
33. Khatun F, Stephenson RJ, Toth I: **An overview of structural features of antibacterial glycoconjugate vaccines that influence their immunogenicity.** *Chemistry–A European Journal* 2017, **23**:4233-4254.
34. Velders MP, Weijzen S, Eiben GL, Elmishad AG, Kloetzel P-M, Higgins T, Ciccarelli RB, Evans M, Man S, Smith L: **Defined flanking spacers and enhanced proteolysis is essential for eradication of established tumors by an epitope string DNA vaccine.** *The Journal of Immunology* 2001, **166**:5366-5373.
35. Lee M, Bang K, Kwon H, Cho S: **Enhanced antibacterial activity of an attacin-coleopterucin hybrid protein fused with a helical linker.** *Molecular biology reports* 2013, **40**:3953-3960.
36. Yang Y, Sun W, Guo J, Zhao G, Sun S, Yu H, Guo Y, Li J, Jin X, Du L: **In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations.** *Human vaccines & immunotherapeutics* 2015, **11**:795-805.
37. Livingston B, Crimi C, Newman M, Higashimoto Y, Appella E, Sidney J, Sette A: **A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes.** *The Journal of Immunology* 2002, **168**:5499-5506.
38. Nezafat N, Eslami M, Negahdaripour M, Rahbar MR, Ghasemi Y: **Designing an efficient multi-epitope oral vaccine against Helicobacter pylori using immunoinformatics and structural vaccinology approaches.** *Molecular BioSystems* 2017, **13**:699-713.
39. Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T: **Design of the linkers which effectively separate domains of a bifunctional fusion protein.** *Protein engineering* 2001, **14**:529-532.
40. McKeever TM, Lewis SA, Smith C, Hubbard R: **Vaccination and allergic disease: a birth cohort study.** *American journal of public health* 2004, **94**:985-989.
41. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Higgs KW: **Intrinsically disordered protein.** *Journal of molecular graphics and modelling* 2001, **19**:26-59.
42. Hasson SSAA, Al-Busaidi JKZ, Sallam TA: **The past, current and future trends in DNA vaccine immunisations.** *Asian Pacific Journal of Tropical Biomedicine* 2015, **5**:344-353.
43. Kaufmann SH, McElrath MJ, Lewis DJ, Del Giudice G: **Challenges and responses in human vaccine development.** *Current opinion in immunology* 2014, **28**:18-26.
44. Tameris MD, Hatherill M, Landry BS, Scriba TJ, Snowden MA, Lockhart S, Shea JE, McClain JB, Hussey GD, Hanekom WA: **Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial.** *The Lancet* 2013, **381**:1021-1028.

45. Merten O-W: **Virus contaminations of cell cultures—a biotechnological view.** *Cytotechnology* 2002, **39**:91-116.
46. Committee ISR: *Immunization safety review: vaccinations and sudden unexpected death in infancy.* National Academies Press; 2003.
47. Craig DB, Dombkowski AA: **Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins.** *BMC Bioinformatics* 2013, **14**:346.
48. Jiang Z, Mak TW, Sen G, Li X: **Toll-like receptor 3-mediated activation of NF-kappaB and IRF3 diverges at Toll-IL-1 receptor domain-containing adapter inducing IFN-beta.** *Proc Natl Acad Sci U S A* 2004, **101**:3533-3538.
49. Lopez-Blanco JR, Garzon JI, Chacon P: **iMod: multipurpose normal mode analysis in internal coordinates.** *Bioinformatics* 2011, **27**:2843-2850.
50. Rapin N, Lund O, Bernaschi M, Castiglione F: **Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system.** *PLoS One* 2010, **5**:e9862.

## Tables

**Table 1.** Predicted CTL and IFN- $\gamma$  epitopes with MHC I binder's epitopes against of emerging human pathogenic coronaviruses; SARS-CoV, MERS-CoV and SARS-CoV2.

Selected Antigen	Organism	Position	CTL epitope	IFN- $\gamma$ epitope	MHC-I epitope	Allele	Percentile rank
nsp3	SARS-CoV-2	829-840	LNHTKKWKY	+	MSALNHTKKWKY	HLA-B*57:01	0.26
	SARS-CoV	1184-1195	FEVLAVEDT	+	FEVLAVEDTQGM	HLA-A*26:01	0.62
	MERS-CoV	391-402	SFDYLIREA	+	AVSFDYLIREAK	HLA-A*11:01	0.32
nsp8	SARS-CoV-2	12-23	YAAFATAQE	+	YAAFATAQEAYE	HLA-B*35:01	0.21
	SARS-CoV	35-46	KKLLKSLNV	+	LKKLLKSLNVAK	HLA-A*03:01	2.10
	MERS-CoV	39-50	AVNIAKNAY	+	LQKAVNIAKNAY	HLA-B*15:01	0.52
S protein	SARS-CoV-2	27-38	YTNSFTRGV	+	AYTNSFTRGVYY	HLA-A*01:01	0.04
	SARS-CoV	454-465	ISNVPFSPD	+	DISNVPFSPDGK	HLA-A*11:01	1.40
	MERS-CoV	303-314	IQSDRKAWA	+	IQSDRKAWAAFY	HLA-A*01:01	0.11
M protein	SARS-CoV-2	185-196	RVAGDSGFA	+	QRVAGDSGFAAY	HLA-A*30:02	0.06
	SARS-CoV	53-64	LWLLWPVTL	+	LWLLWPVTLACF	HLA-A*23:01	0.32
	MERS-CoV	152-163	HFGACDYDR	+	GMHFGACDYDRL	HLA-B*38:01	0.49
N protein	SARS-CoV-2	45-56	NNTASWFTA	+	LPNNTASWFTAL	HLA-B*07:02	0.40
	SARS-CoV	62-73	EELRFPRGQ	+	KEELRFPRGQGV	HLA-C*07:01	1.40
	MERS-CoV	249-260	HKRTSTKSF	+	KMRHKRTSTKSF	HLA-A*03:01	0.80
E protein	SARS-CoV-2	33-44	TALRLCAYC	+	ILTALRLCAYCC	HLA-A*25:01	1.40
	SARS-CoV	50-61	VKPTVYVYS	+	SLVKPTVYVYSR	HLA-A*31:01	0.87
	MERS-CoV	49-60	VQPALYLYN	+	TLLVQPALYLYN	HLA-A*29:02	1.30

**Table 2.** Predicted MHC II epitopes against emerging human pathogenic coronaviruses SARS-CoV, MERS-CoV, and SARS-CoV2 antigens.

Antigen	Organism	Allele	Position	HTL epitopes	Percentile rank
nsp3	SARS-CoV-2	HLA-DPA1*01:03	1512-1526	AYILFTRFFYVLGLA	0.01
		HLA-DPB1*02:01			
	SARS-CoV	HLA-DQA1*01:01	1517-1531	FISNSWLMWFIISIV	0.01
		HLA-DQB1*05:01			
	MERS-CoV	HLA-DRB1*09:01	1473-1487	DWRSYNYAVSSAFWL	0.01
nsp8	SARS-CoV-2	HLA-DRB1*13:02	184-198	LIVTALRANSAVKLQ	0.09
	SARS-CoV	HLA-DRB1*13:21	86-100	AMQTMLFTMLRKLDN	0.16
	MERS-CoV	HLA-DRB1*04:26	178-192	ENLTWPLVLECTRAS	0.23
S Protein	SARS-CoV-2	HLA-DRB1*13:02	113-127	KTQSELLIVNNATNVV	0.01
	SARS-CoV	HLA-DRB1*01:01	499-513	LSFELLNAPATVCGP	0.01
	MERS-CoV	HLA-DQA1*05:01	239-253	FMYTYNITEDEILEW	0.03
		HLA-DQB1*02:01			
M Protein	SARS-CoV-2	HLA-DPA1*01:03	89-103	GLMWLSYFIASFRLF	0.05
		HLA-DPB1*02:01			
	SARS-CoV	HLA-DRB1*09:01	173-187	RTLSYYKLGASQRVG	0.07
	MERS-CoV	HLA-DRB1*11:14	8-62	FKMFVLWLLWPSSMA	0.06
N Protein	SARS-CoV-2	HLA-DQA1*01:02	150-164	NPANNAAIVLQLPQG	0.03
		HLA-DQB1*06:02			
	SARS-CoV	HLA-DRB1*09:01	306-320	AQFAPSASAFFGMSR	0.01
	MERS-CoV	HLA-DRB1*07:03	328-342	FLRYSGAIKLDPKNP	0.04
E Protein	SARS-CoV-2	HLA-DPA1*03:01	18-32	LLFLAFVVFLLVTLA	0.02
		HLA-DPB1*04:02			
	SARS-CoV	HLA-DPA1*03:01	15-29	NSVLLFLAFVVFLLV	0.06
		HLA-DPB1*04:02			
	MERS-CoV	HLA-DRB1*01:01	46-60	GFNTLLVQPALYLYN	0.03

**Table 3.**Free energy details related to 5' end of recombinant gene mRNA structure predicted by mfold web server.

<b>Structural element</b>	<b>Free energy (kcal/mol)</b>	<b>Base pair</b>
External loop	-8.70	27 ss bases & 8 closing helices.
Stack	-2.90	External closing pair is G <sup>1780</sup> -C <sup>2089</sup>
Stack	-2.90	External closing pair is G <sup>1783</sup> -C <sup>2086</sup>
Stack	-1.80	External closing pair is U <sup>1785</sup> -A <sup>2084</sup>
Helix	-18.60	9 base pairs.
Stack	-2.30	External closing pair is G <sup>1804</sup> -C <sup>2074</sup>
Interior loop	0.80	External closing pair is G <sup>1798</sup> -C <sup>2080</sup>

**Table 4.** B cell epitopes predicted in constructed vaccine based on different parameters.

Prediction parameter	Epitope sequence
Hydrophilicity	AKGRKKRRQRRR, QGPGPGG, ERAGKTQS, DNHEYGA, GGSIQSD, AGGGSNNTAS, TAGGGSHKRTSTKS, PRGQGGGSTA, GACDYDRGGGS, AVEDTGGGSKK, SPDGGGSVQ, YNGGGSRVAGDS, REAGGGSY, ATAQEGGSA
Flexibility	EAAAKGRKKRRRQRRRPP, EALERAGKT, GSIQSDRKAWAGGGSNNT, FTAGGGSHKRTSTKSFGGGS, LRFPRGQGGG, AYCAGGGS, YVYSGGG, CDYDRGGG, LAVEDTGGGSKLK, VPFPDGG, YLYNGGGS, PVTLLGGSS, LIREAGG, ATAQEGGG
Accessibility	AAAKGRKKRRRQRRRPPQGP, FCPRRYKQIGT, TKCCKKPEAAAK, LERAGKTQS, YTYNITEDE, YNAAYRTLSYYKLGASQR, AYDWRSYNYAV, AYNPANNA, IKLDPKNPAA, MLRKLDNHEYGAE, ERAGYTNSFTR, SIQSDRKAWA, GGSHKRTSTKSF, SEELRFPRGQG, SLNHTKKWKYGG, GSKKLLKSLN, AKNAYHEYGAE, YQVNNLEE
Turns	YNPANNA, KLDNHEY, GGSNNTAS, RAGHHHHH
Exposed Surface	AAKGRKKRRRQRRRPPQG, CPRRYKQIG, TKCCKKPE, IKLDPKNP, RKLDNHE, HKRTSTKS, NHTKKWKY, GSKKLLKSLN
Polarity	AAKGRKKRRRQRRRPPQG, FCPRRYKQIG, GTKCCKKPEA, KAAAHEYGAEALERAGKTQ, TEDEILEW, LRKLDNHEYGAEA, GGGSHKRTSTKS, GGSEELRFPRG, SLNHTKKWKY, GSKKLLKSLN, AKNAYHEYGAEA, NNLEEIHEYGAEALERAGHHHHH
Antigenic Propensity	IGDPVTCLKSG, ICHPVFCPRRY, LPGTKCC, TQSLIVN, FNTLLVQP, TLSYYKLG, LTWPLVLECTR, YLSFELL, IVLQLPQG, FVVFLLVTL, YNSVLLFL, FVVFLLVA, YFKMFVLWLLWPS, TRFFYVLGL, WFIISIV, SVKPTVYVYSGG, SLWLLWPVTLG

## Figures

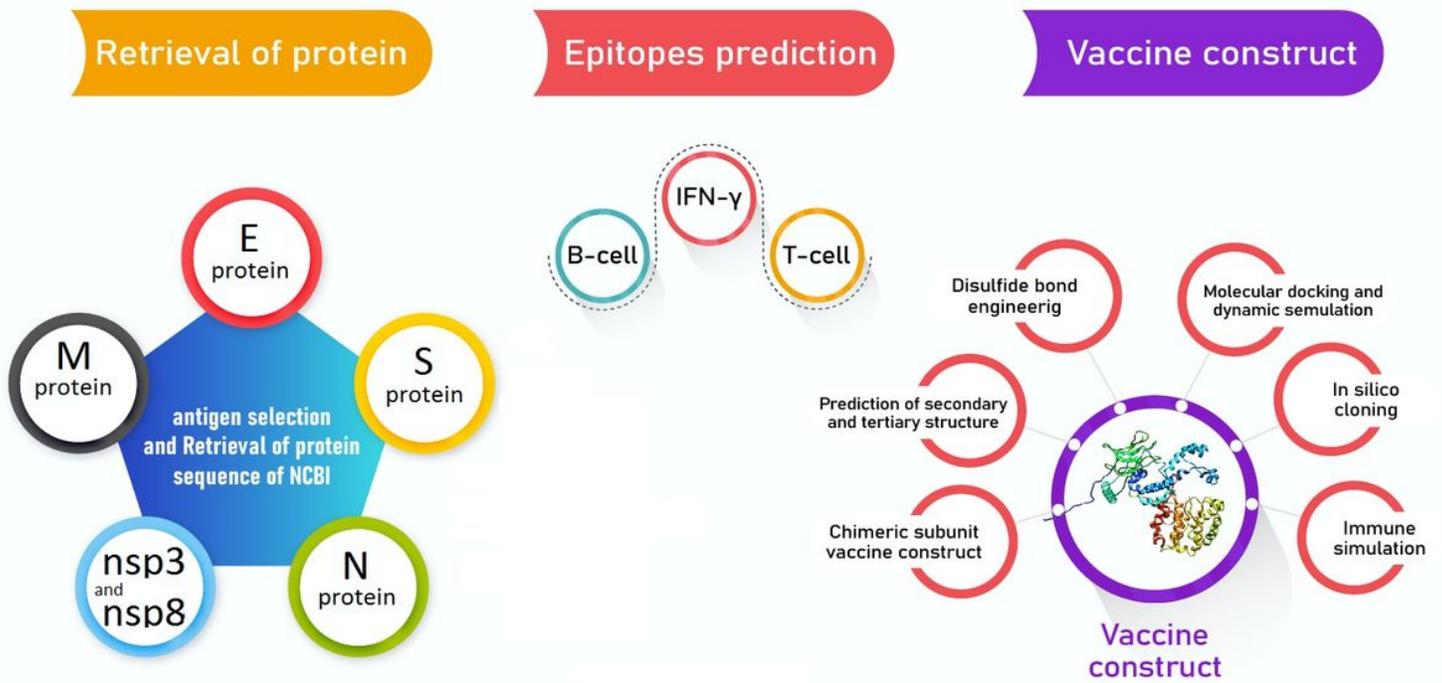


Figure 1

Flowchart summarizing the different steps of a SARS-CoV trivalent subunit vaccinedesign.

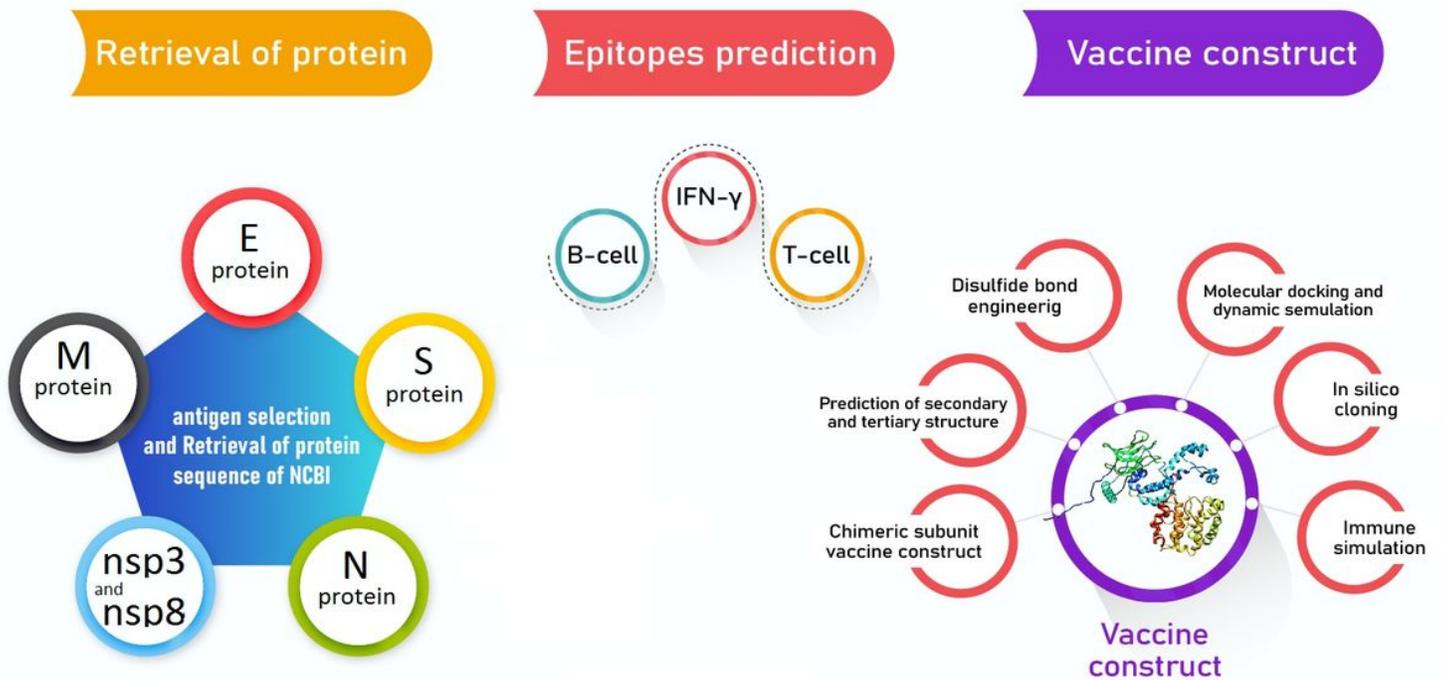


Figure 1

Flowchart summarizing the different steps of a SARS-CoV trivalent subunit vaccinatedesign.



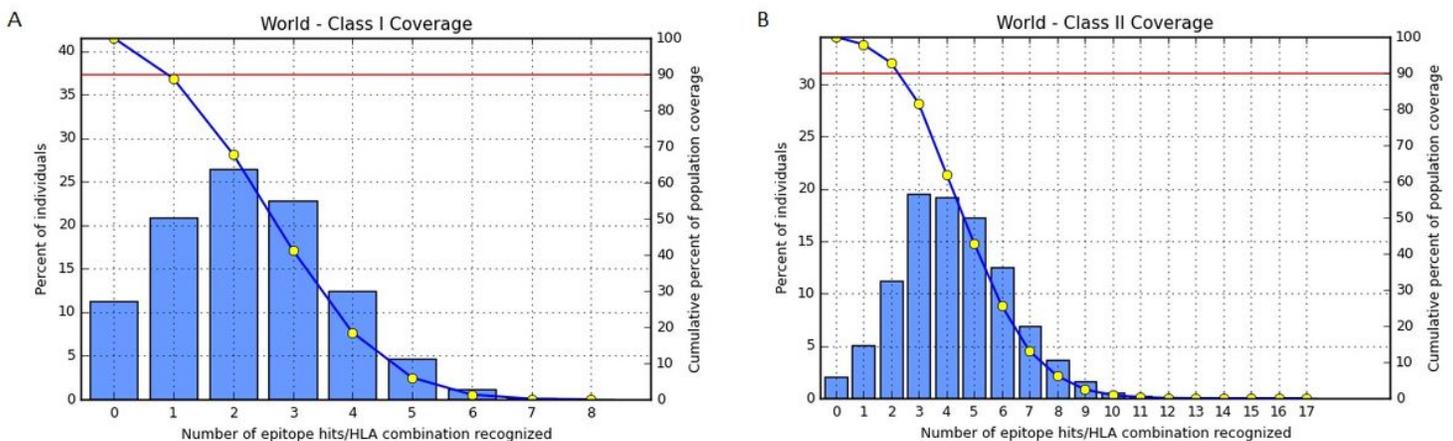
Figure 2

Schematic representation of SARS-CoV trivalent subunit vaccine construct. This vaccine construct covering of TAT sequence,  $\beta$ -defencin and PADRE at N-terminal that linked to HTL and CTL epitopes with appropriate linker and, terminated with His tag in C-terminal.



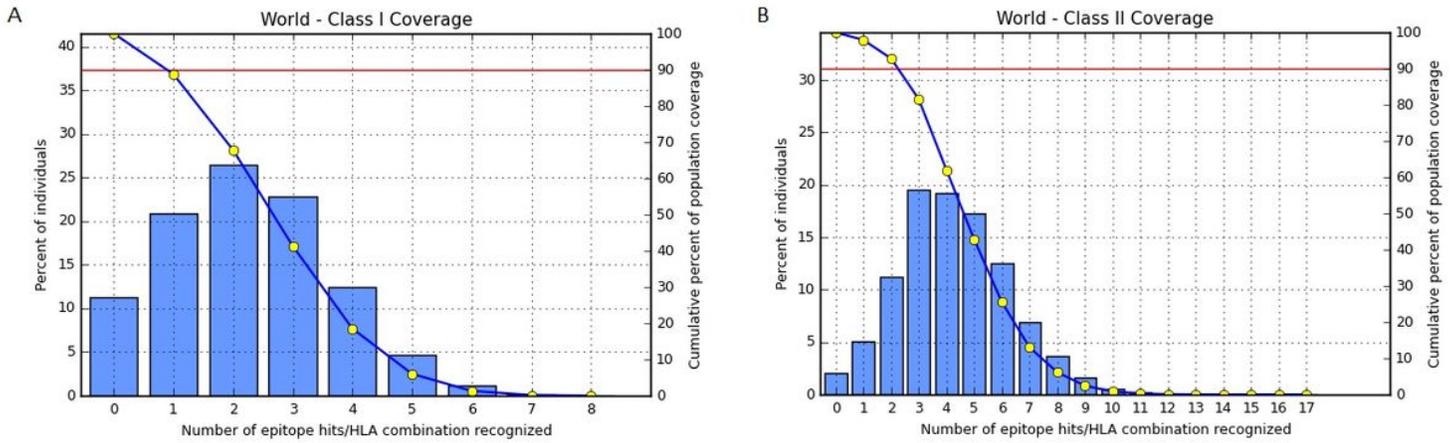
Figure 2

Schematic representation of SARS-CoV trivalent subunit vaccine construct. This vaccine construct covering of TAT sequence,  $\beta$ -defencin and PADRE at N-terminal that linked to HTL and CTL epitopes with appropriate linker and, terminated with His tag in C-terminal.



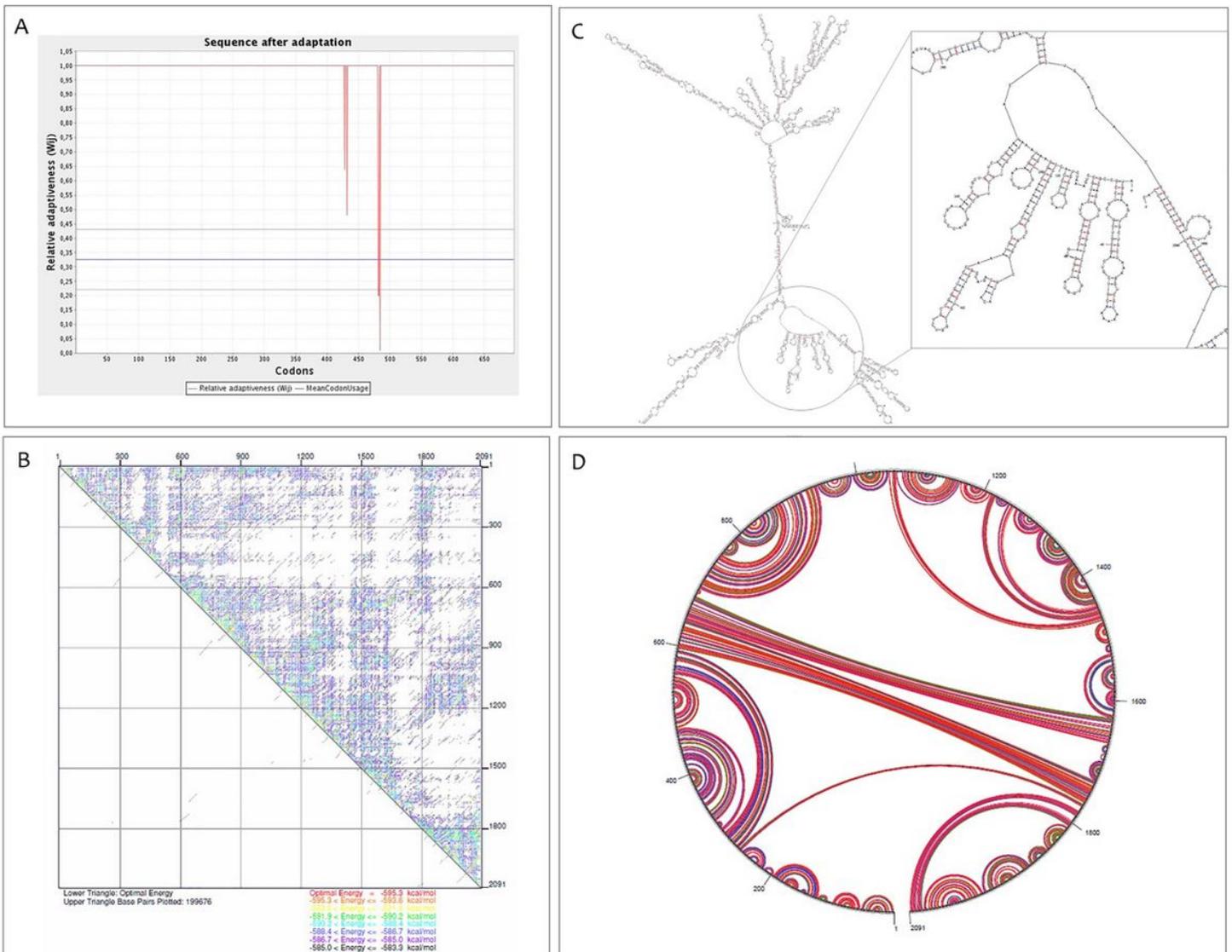
**Figure 3**

Worldwide population coverage of selected epitopes based on (A) MHC-I and (B) MHC-II binding alleles.



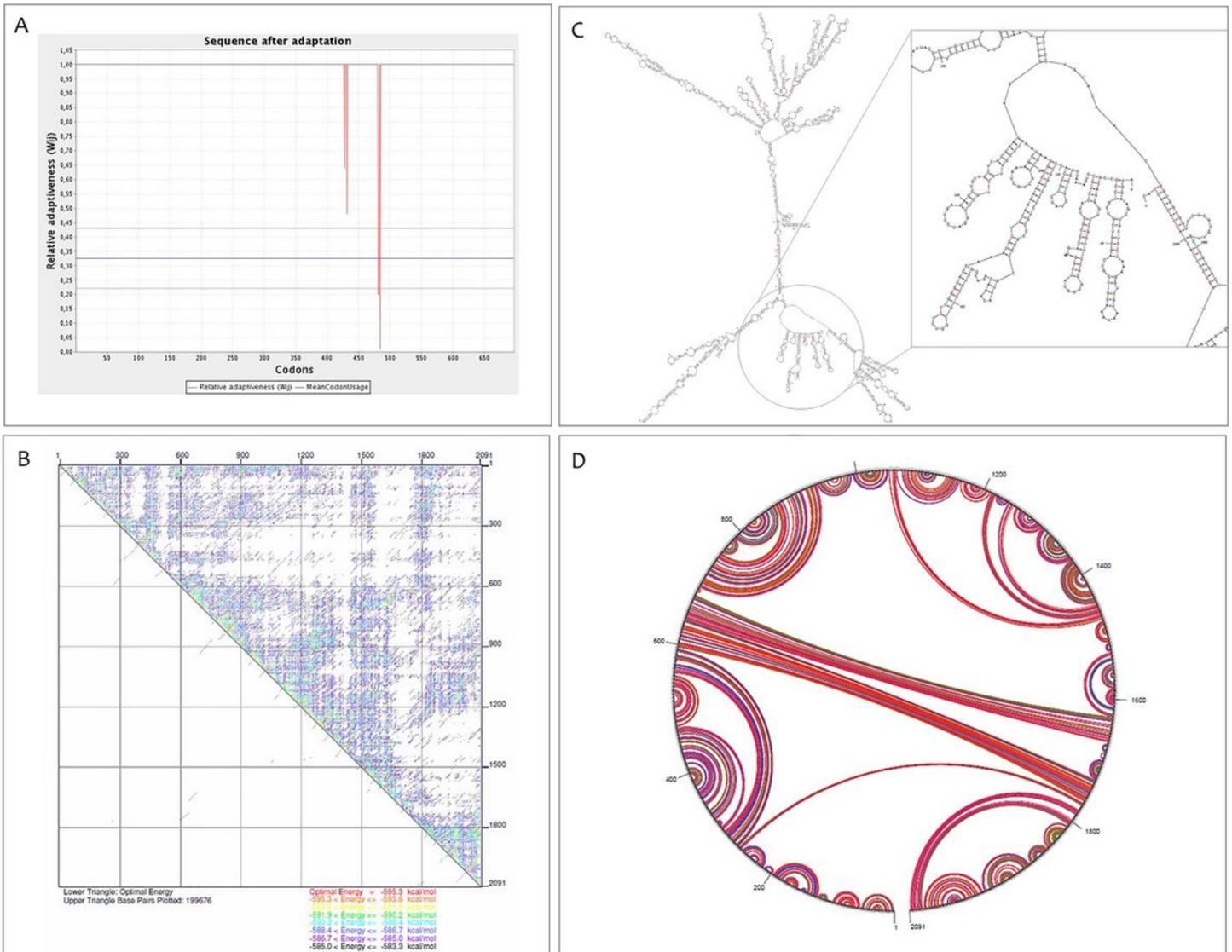
**Figure 3**

Worldwide population coverage of selected epitopes based on (A) MHC-I and (B) MHC-II binding alleles.



**Figure 4**

Worldwide population coverage of selected epitopes based on (A) MHC-I and (B) MHC-II binding alleles.



**Figure 4**

Worldwide population coverage of selected epitopes based on (A) MHC-I and (B) MHC-II binding alleles.



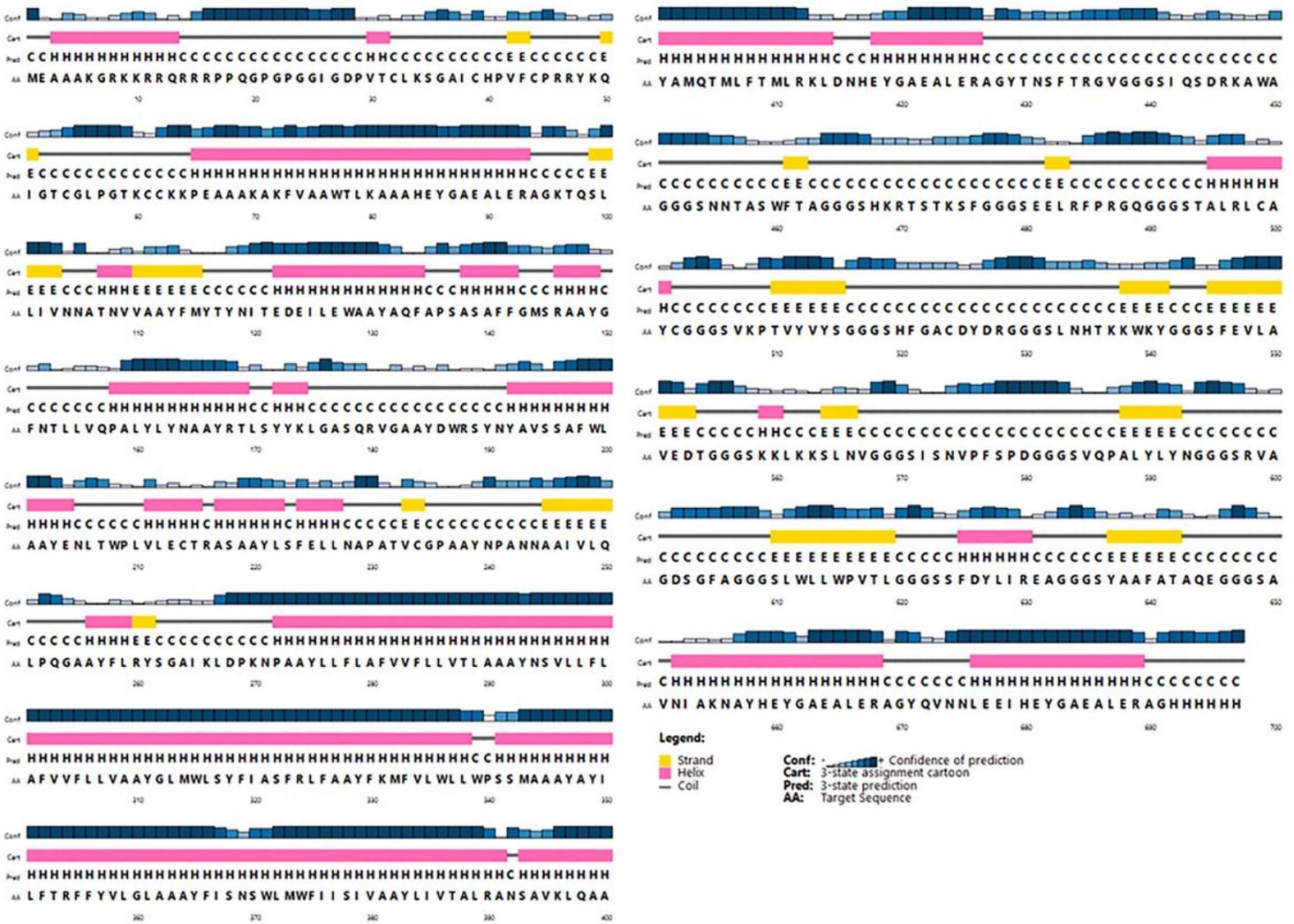
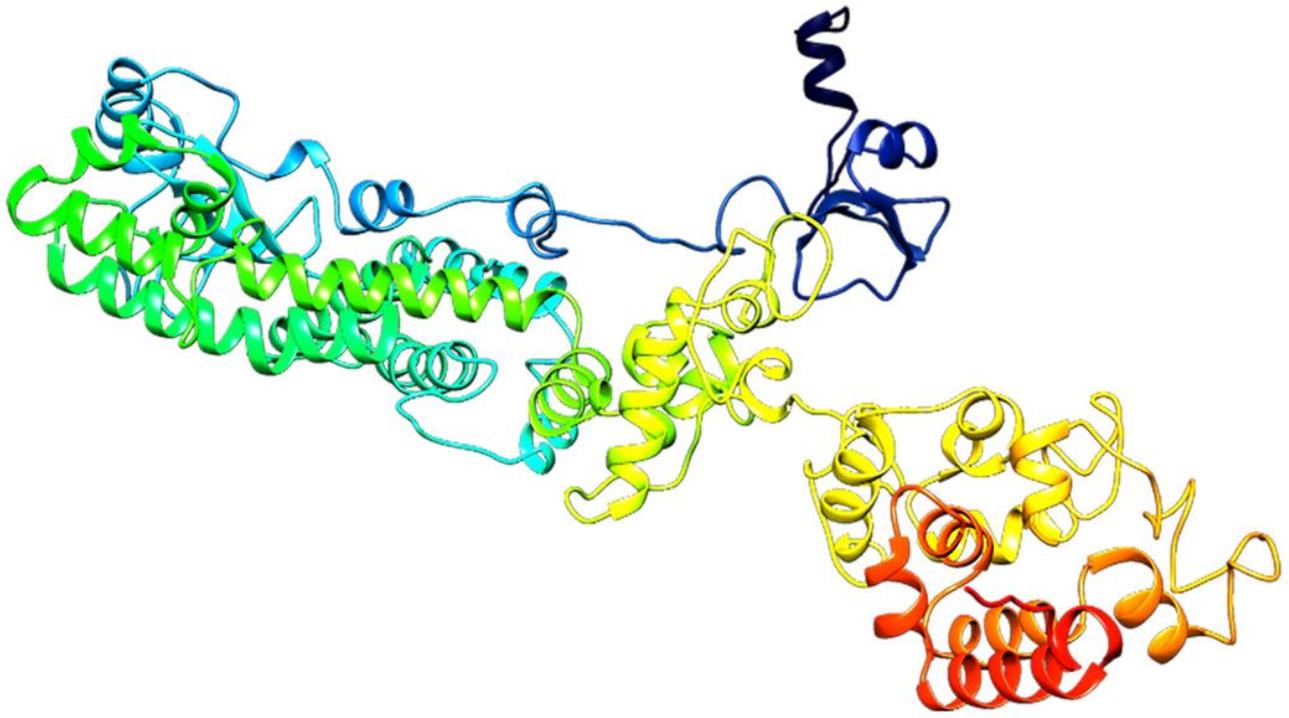


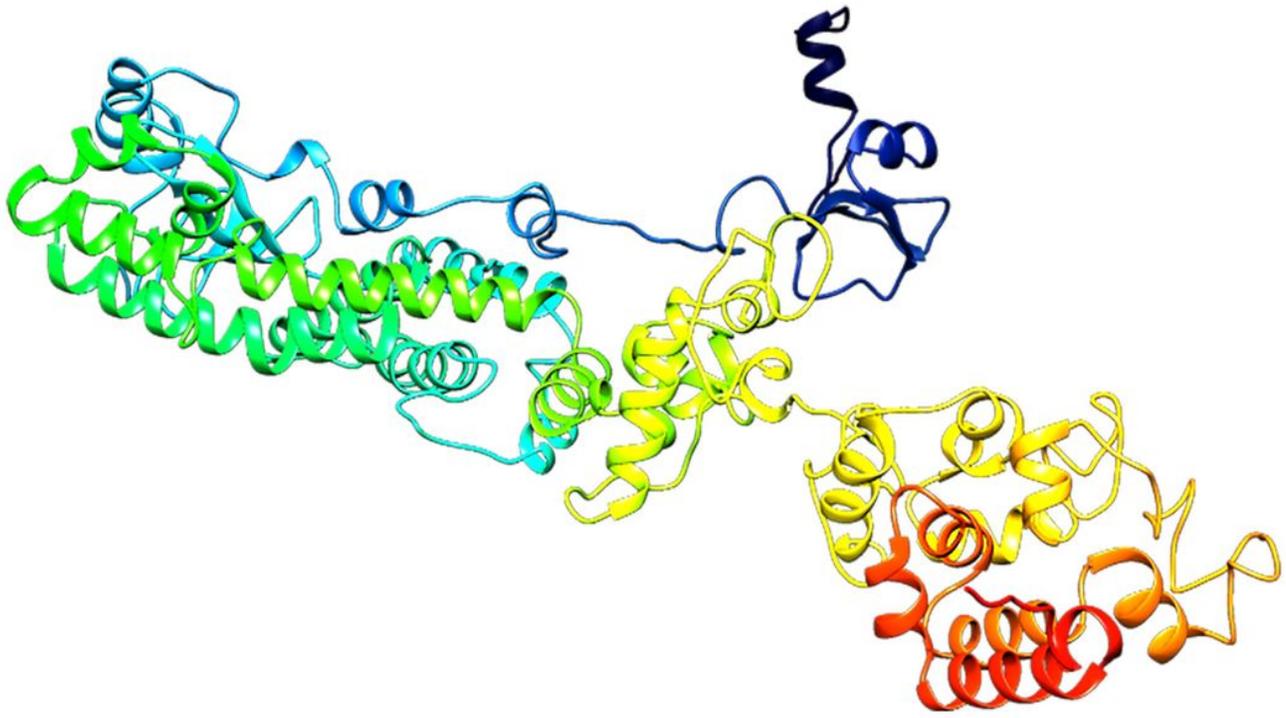
Figure 5

Analysis RNA structures. (A) Graphical view of codon usage in optimized gene. (B) The energy dot plot for predicted mRNA. (C) The predicted of RNA secondary structure has no hairpin and pseudo knot at 5' site of mRNA. (D) A circle graph is one way to display base pairs of RNA structure.



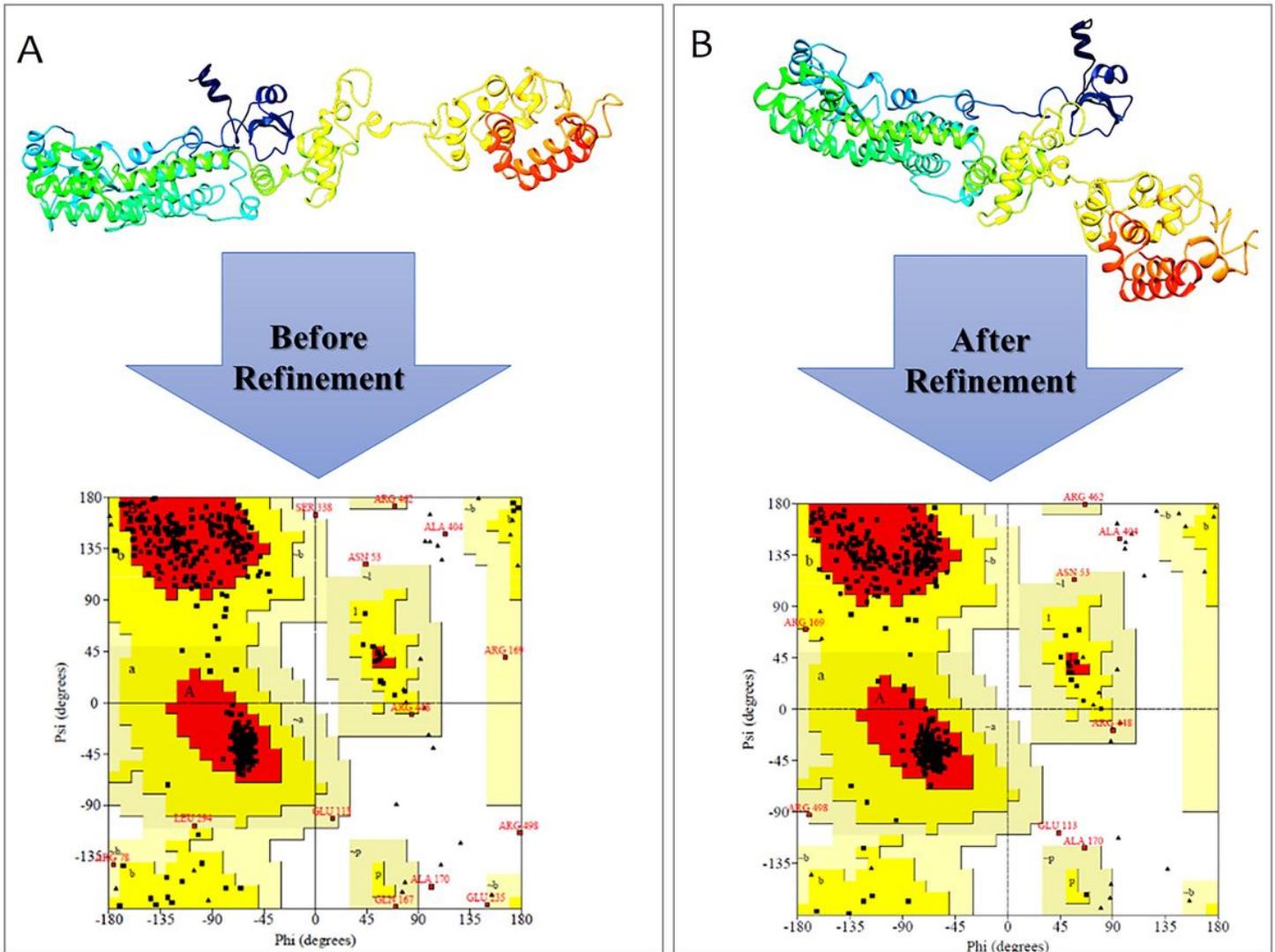
**Figure 6**

Predicted 3D structure of constructed vaccine.



**Figure 6**

Predicted 3D structure of constructed vaccine.



**Figure 7**

Validation 3D model structure. (A) Ramachandran plot of initial model showed 92%, 6.3% and 1.5% of residues were located in favored, allowed and outlier regions, respectively. (B) In the refined model 93.7%, 4.4% and 1.2% of residues were located in favored, allowed and outlier regions, respectively.

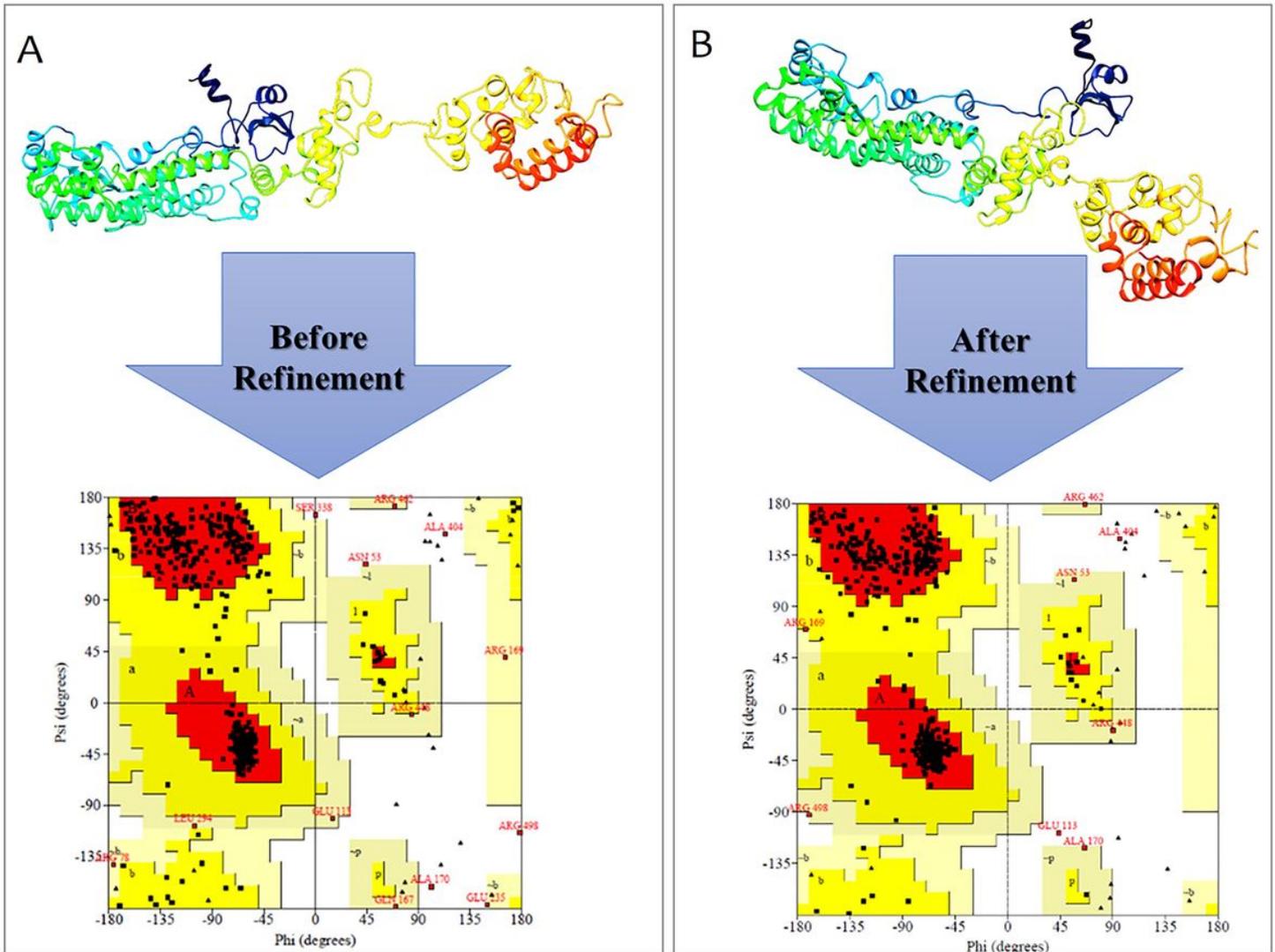


Figure 7

Validation 3D model structure. (A) Ramachandran plot of initial model showed 92%, 6.3% and 1.5% of residues were located in favored, allowed and outlier regions, respectively. (B) In the refined model 93.7%, 4.4% and 1.2% of residues were located in favored, allowed and outlier regions, respectively.

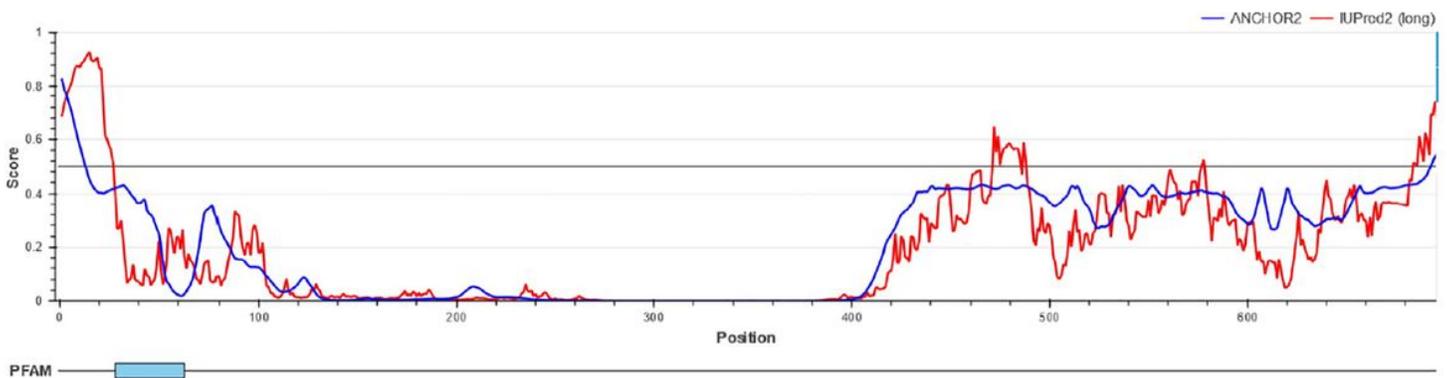
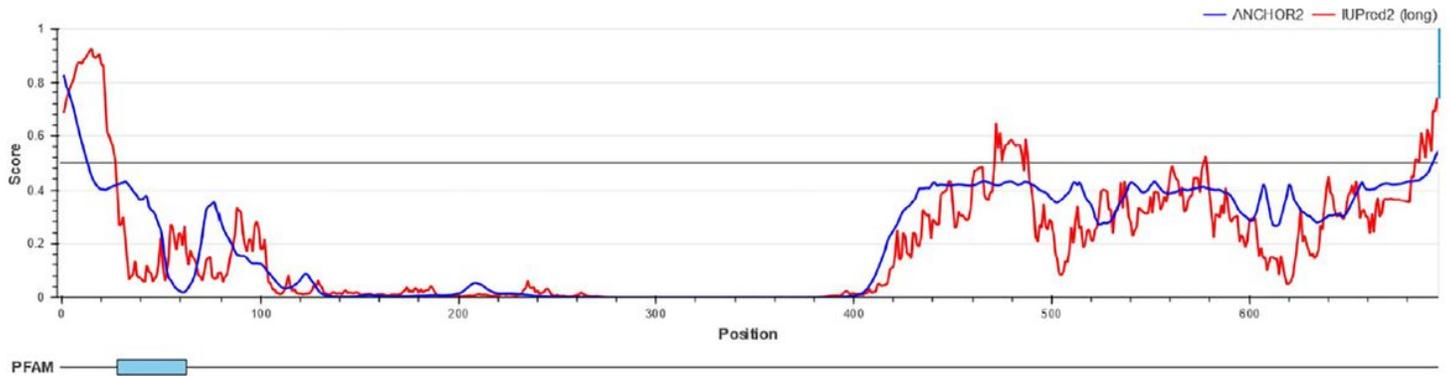


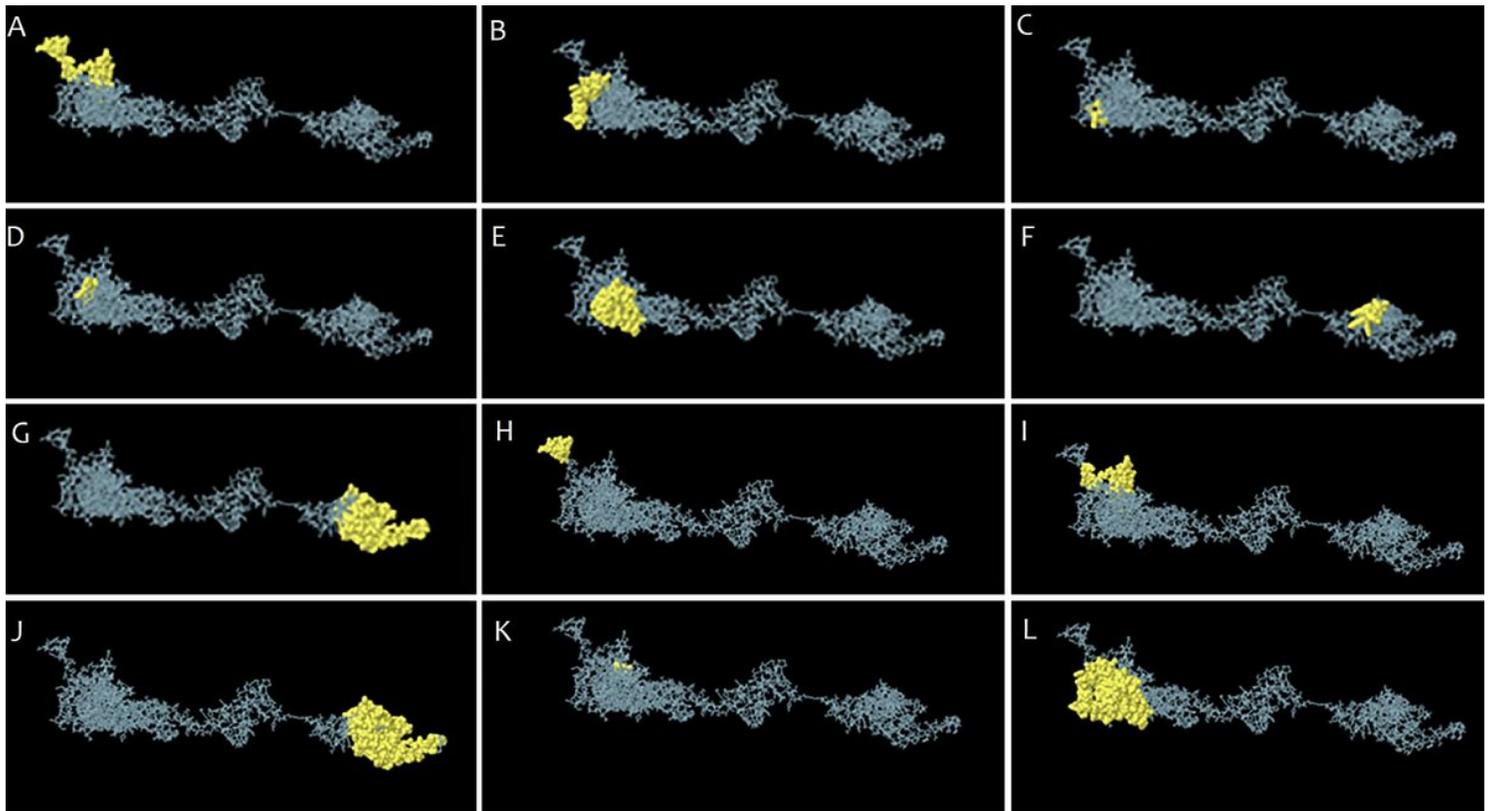
Figure 8

Intrinsically disorder regions. Amino acids in the input sequence were considered disordered when the red line is above the confidence score is higher than 0.5.



**Figure 8**

Intrinsically disorder regions. Amino acids in the input sequence were considered disordered when the red line is above the confidence score is higher than 0.5.



**Figure 9**

The B-cell epitopes. (A-G) Linear and (H-L) discontinuous B cell epitopes on the 3D structure.

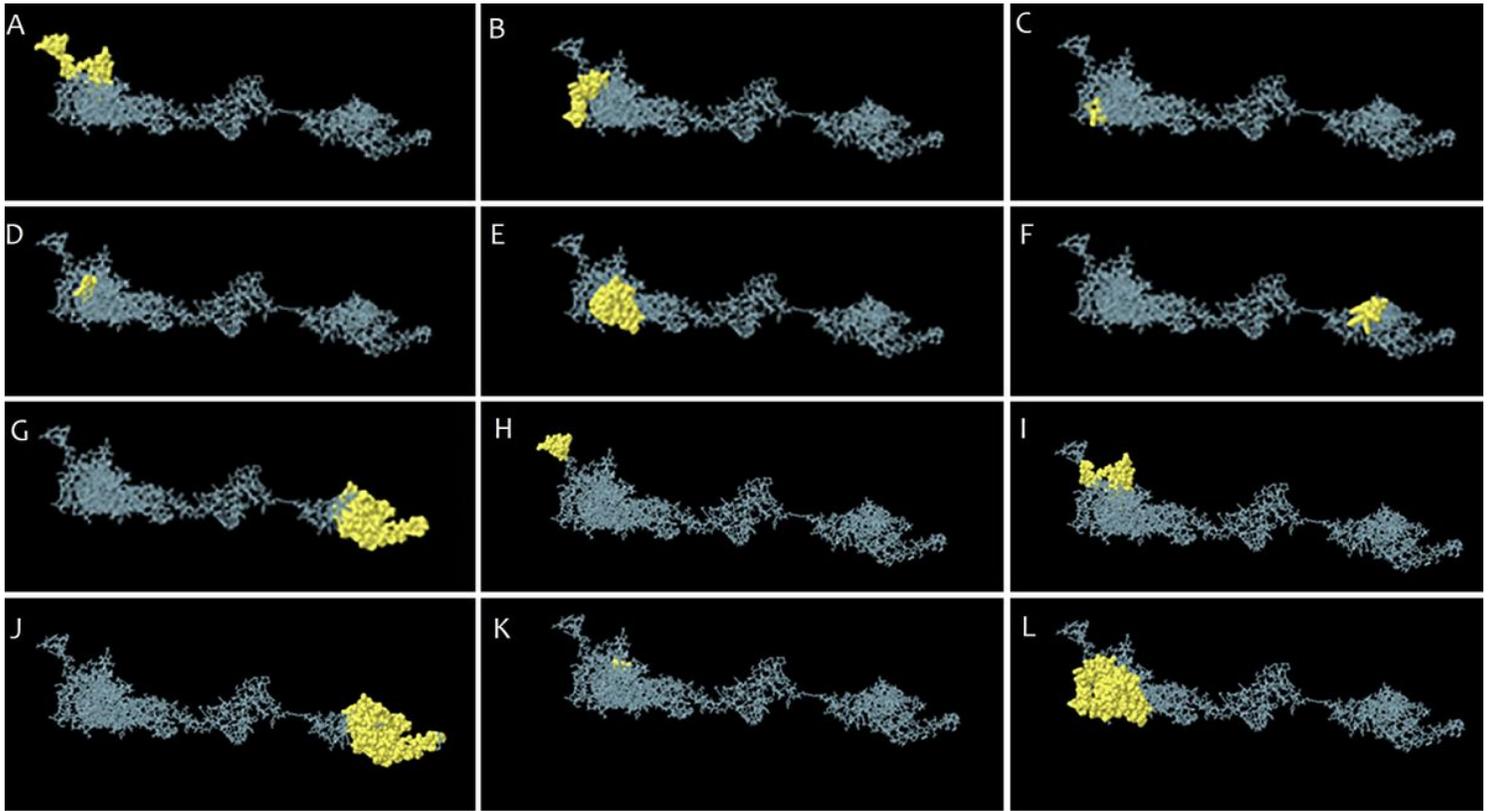


Figure 9

The B-cell epitopes. (A-G) Linear and (H-L) discontinuous B cell epitopes on the 3D structure.

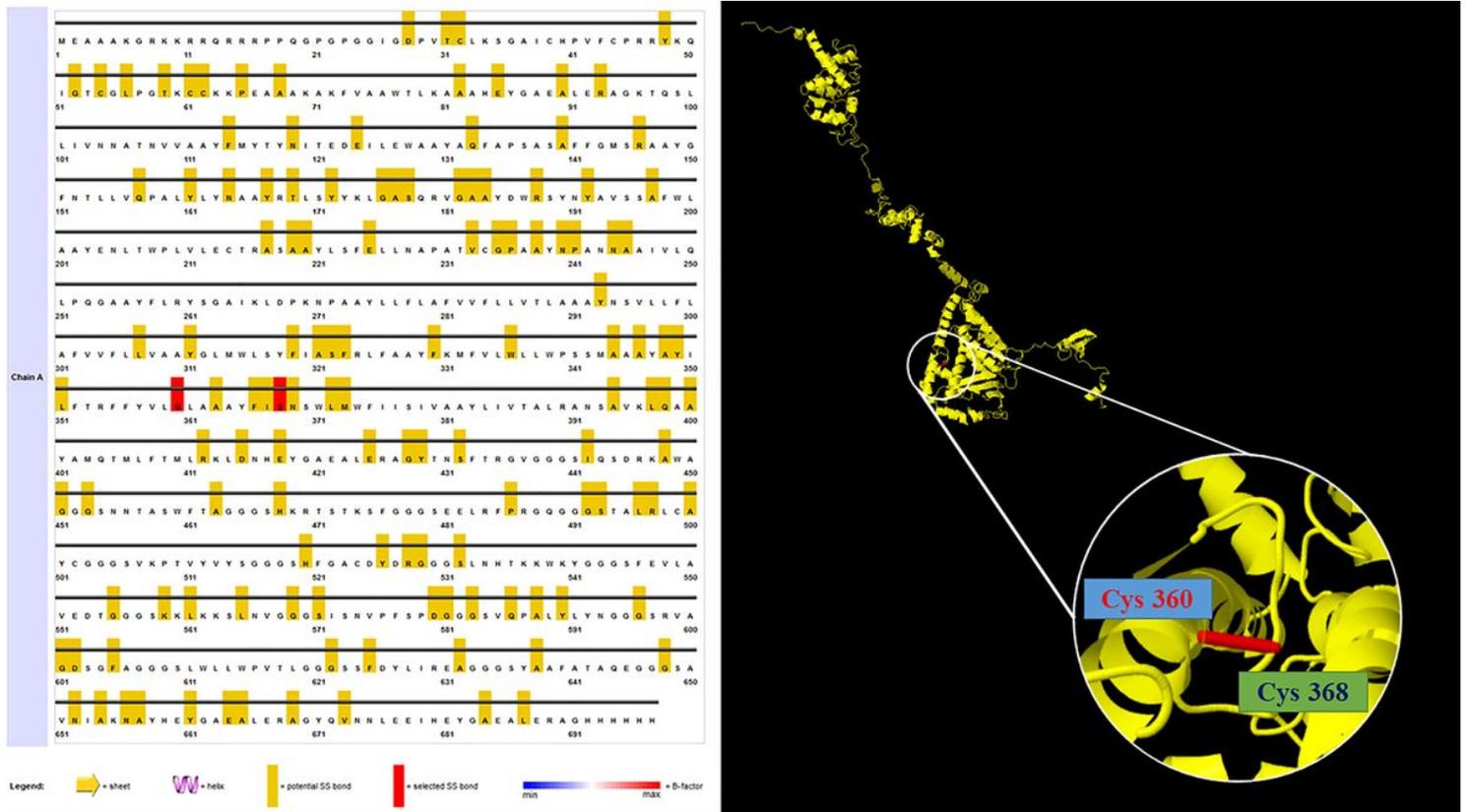


Figure 10

Disulfide engineering to improve protein stability. Two mutated residues pairs were selected based on their energy, chi3 value, and B-factor.

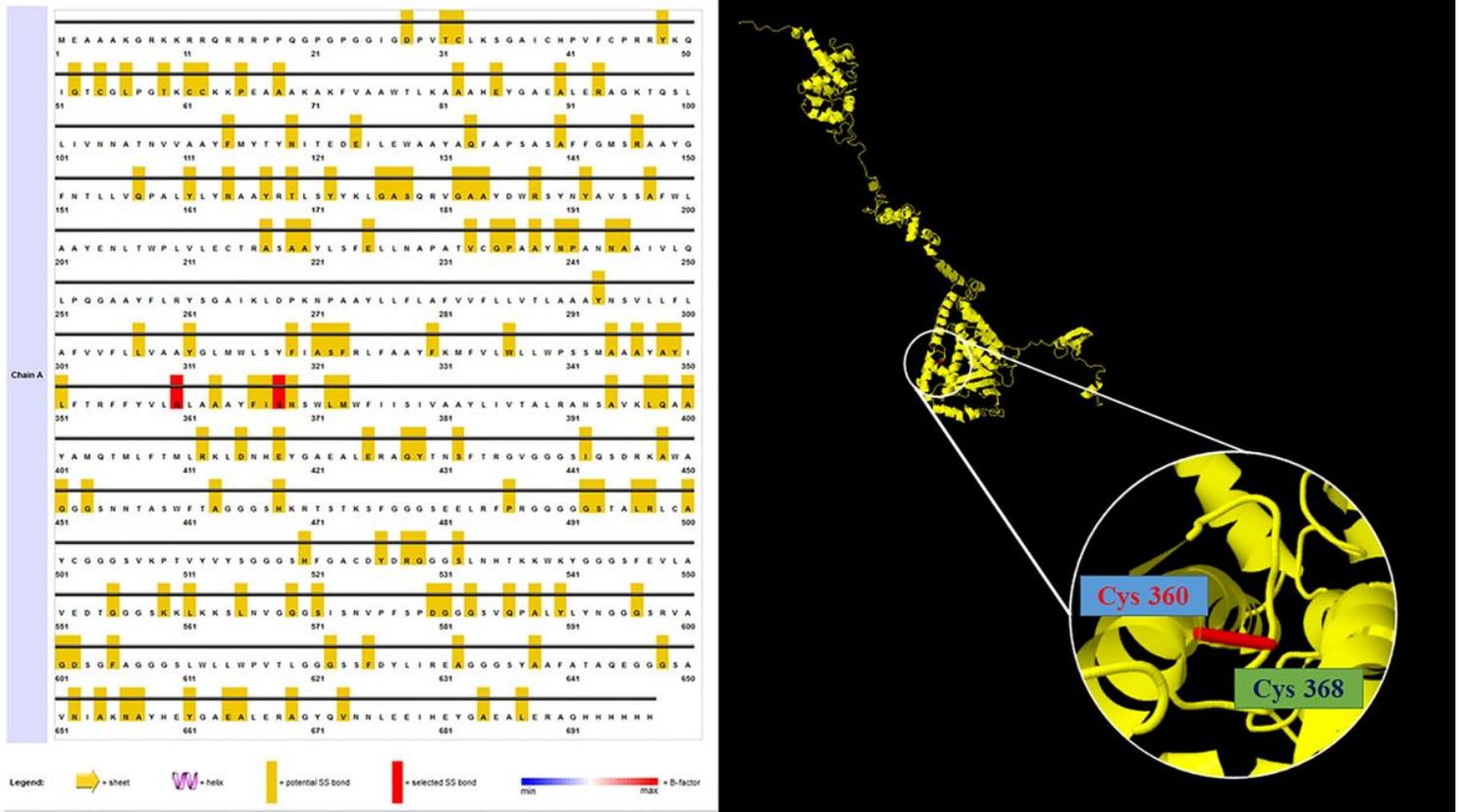


Figure 10

Disulfide engineering to improve protein stability. Two mutated residues pairs were selected based on their energy, chi3 value, and B-factor.

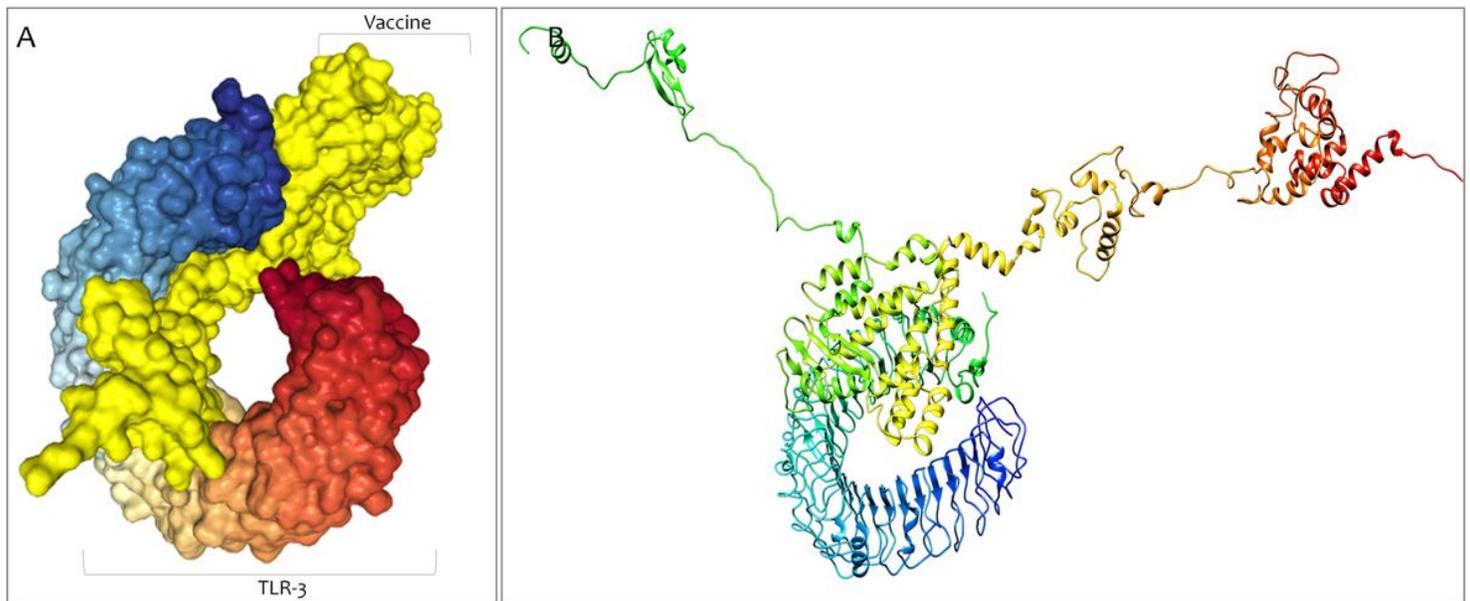
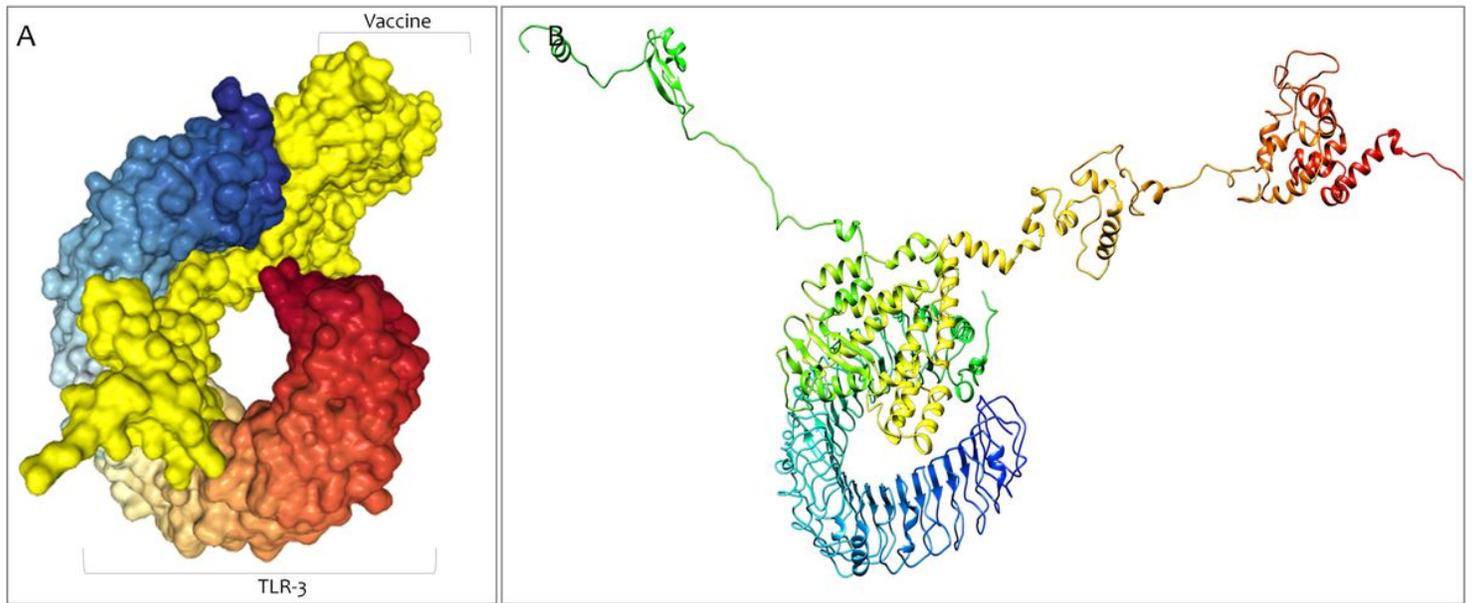


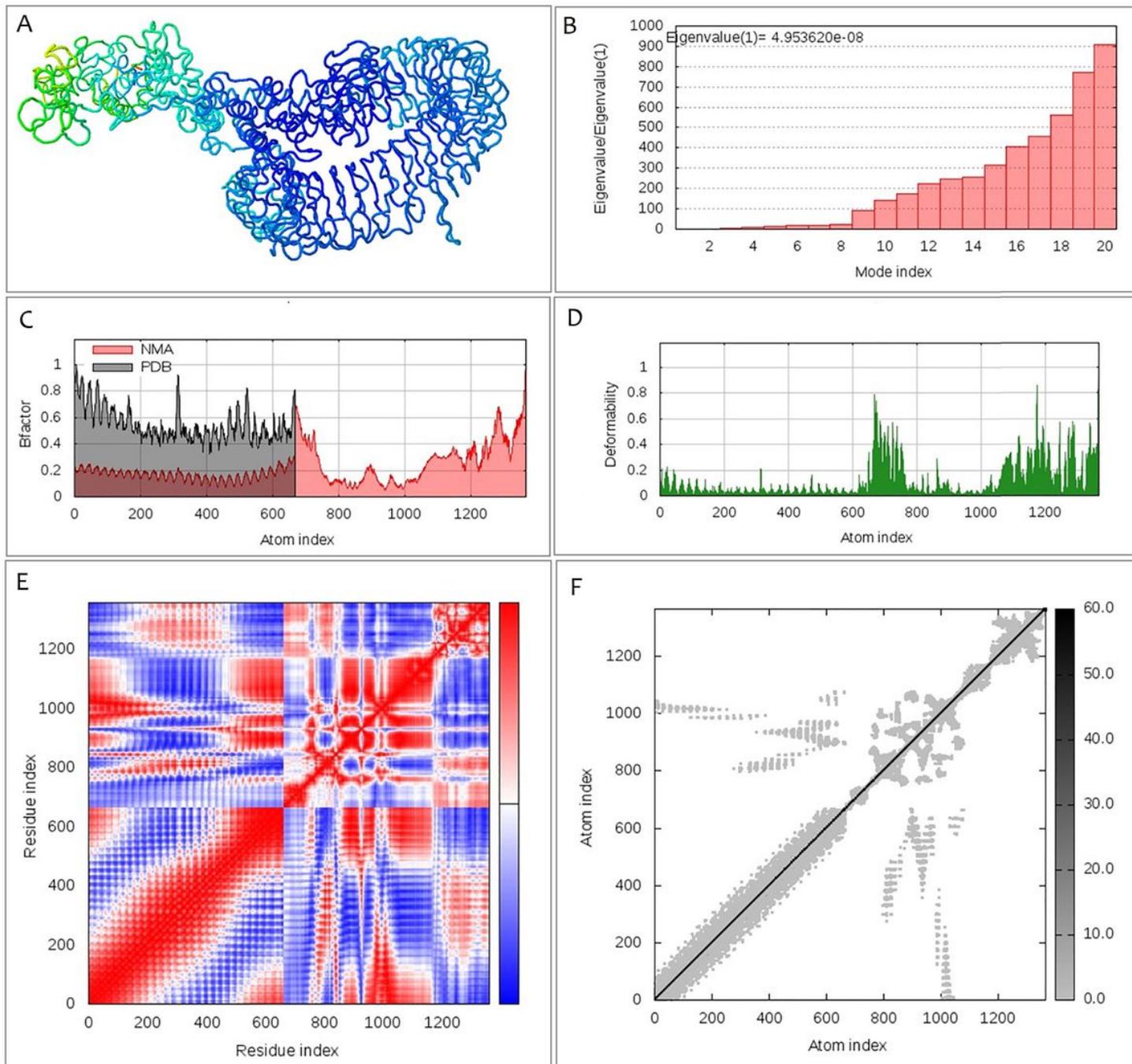
Figure 11

Docking complex of vaccine construct with TLR-3. (A) The result obtained from protein-protein docking with the help of the Hdock. (B) PatchDock server was used to reaffirm the docking between the final vaccine construct and TLR-3.



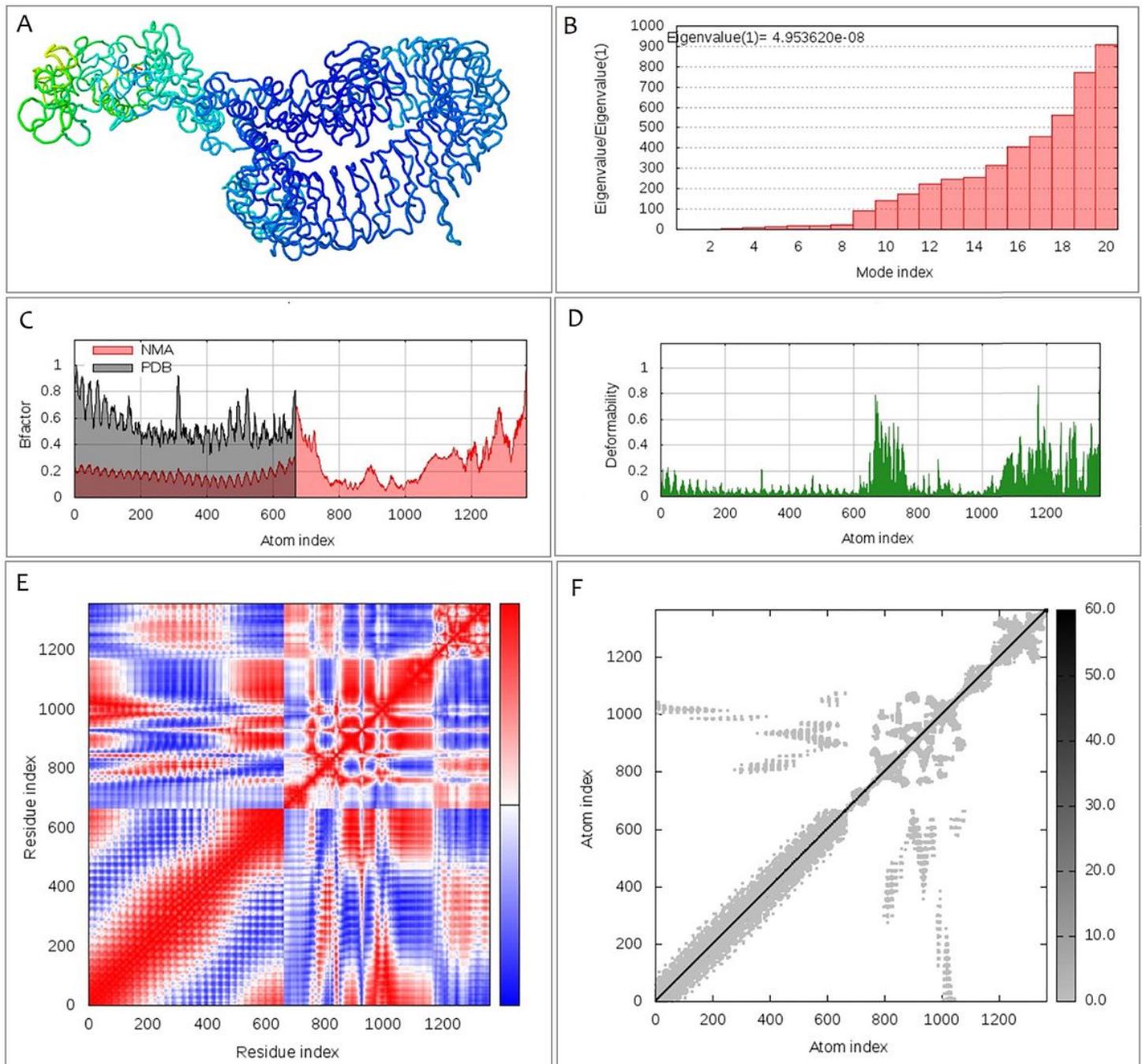
**Figure 11**

Docking complex of vaccine construct with TLR-3. (A) The result obtained from protein-protein docking with the help of the Hdock. (B) PatchDock server was used to reaffirm the docking between the final vaccine construct and TLR-3.



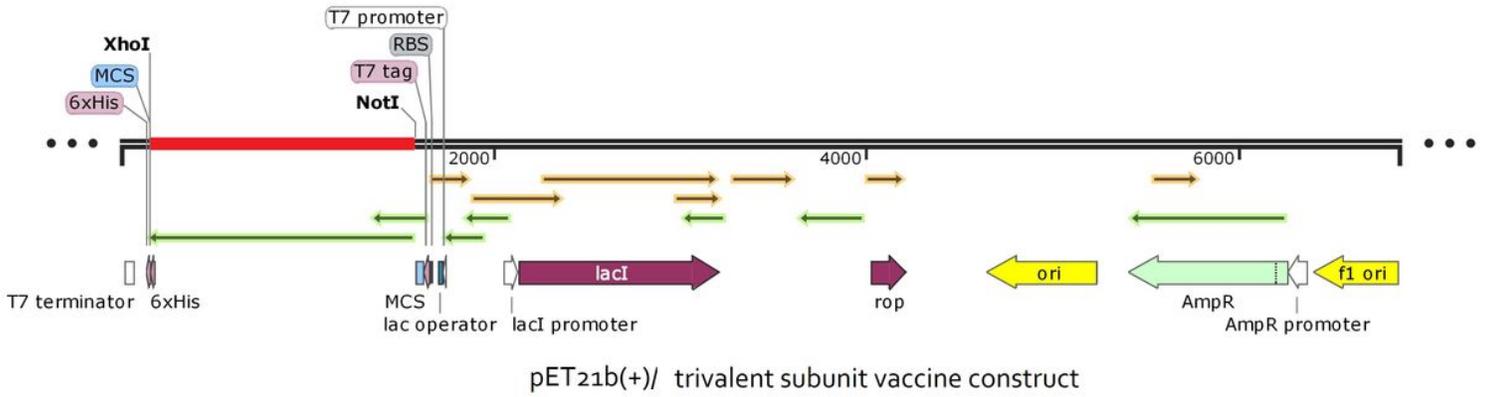
**Figure 12**

Molecular dynamics simulation of vaccine construct and TLR-3. Stability of the protein-protein complex was investigated through (A) mobility, (B) eigenvalue, (C) B factor, (D) deformability, (E) covariance and (F) elastic network analysis.



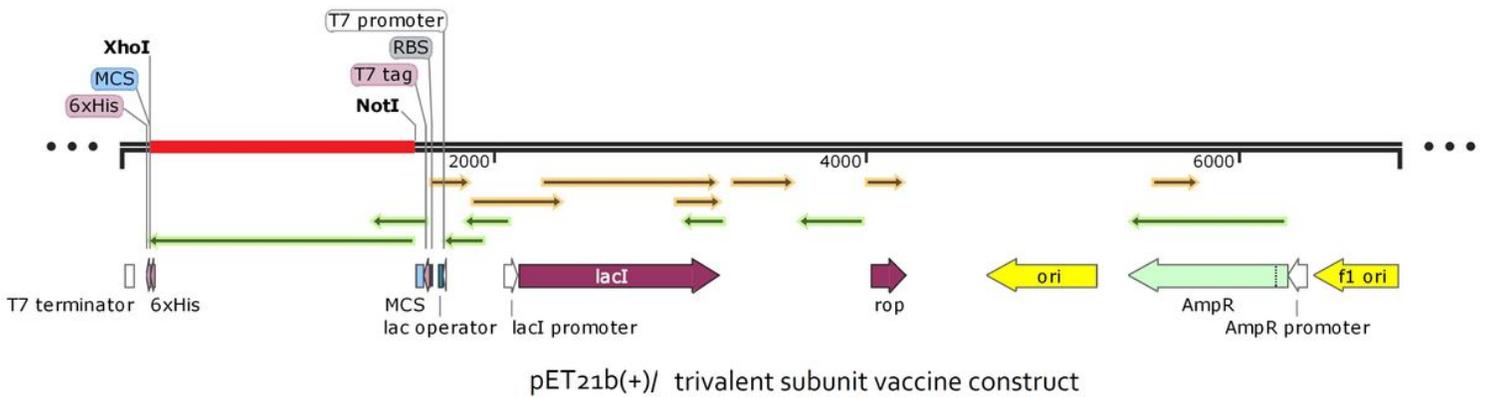
**Figure 12**

Molecular dynamics simulation of vaccine construct and TLR-3. Stability of the protein-protein complex was investigated through (A) mobility, (B) eigenvalue, (C) B factor, (D) deformability, (E) covariance and (F) elastic network analysis.



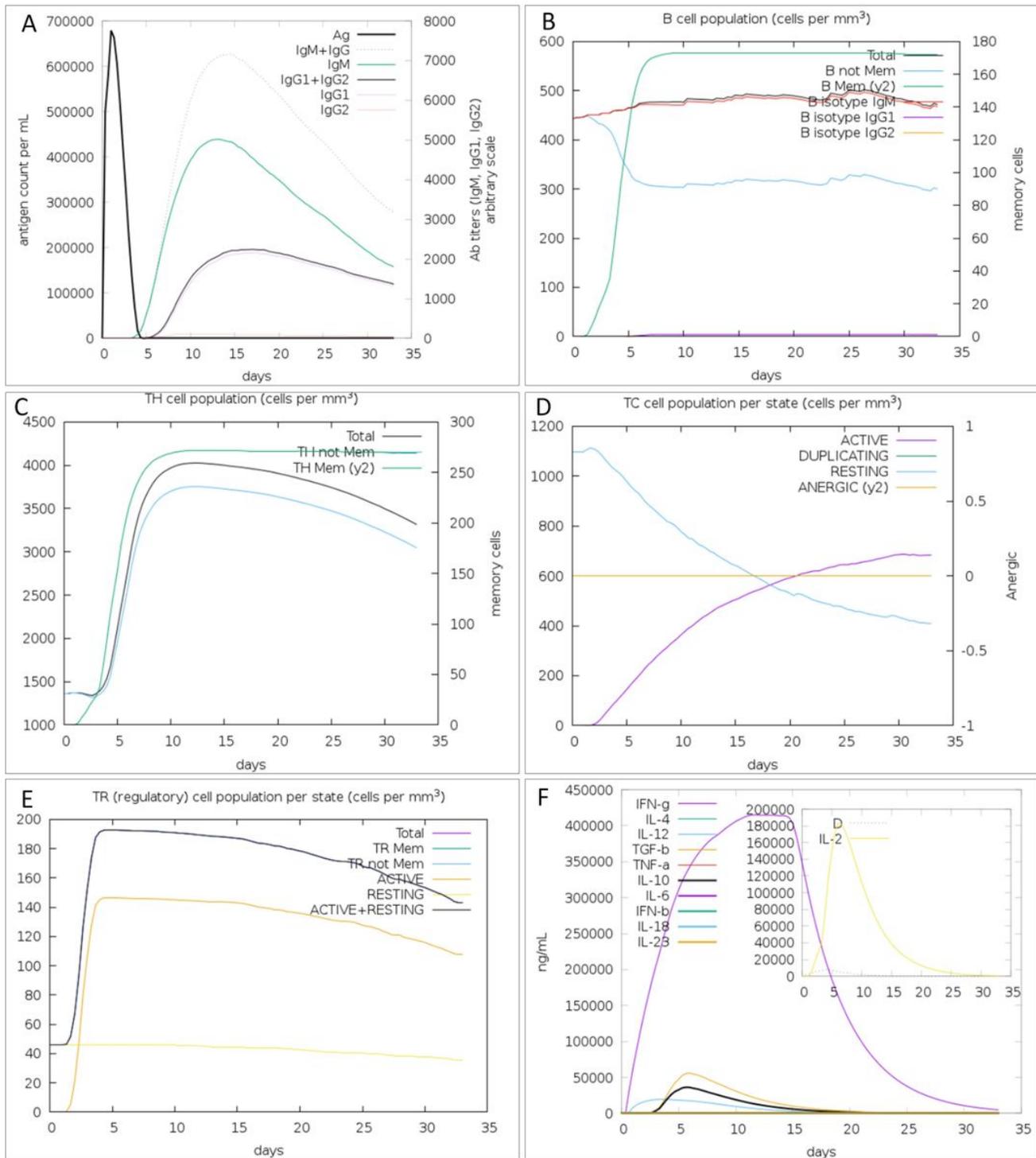
**Figure 13**

In silico cloning to express final vaccine construct in *E. coli*. The vaccine construct gene (red) was inserted into the pET21b (+) vector (black) between NotI and XhoI restriction site to create a recombinant vector.



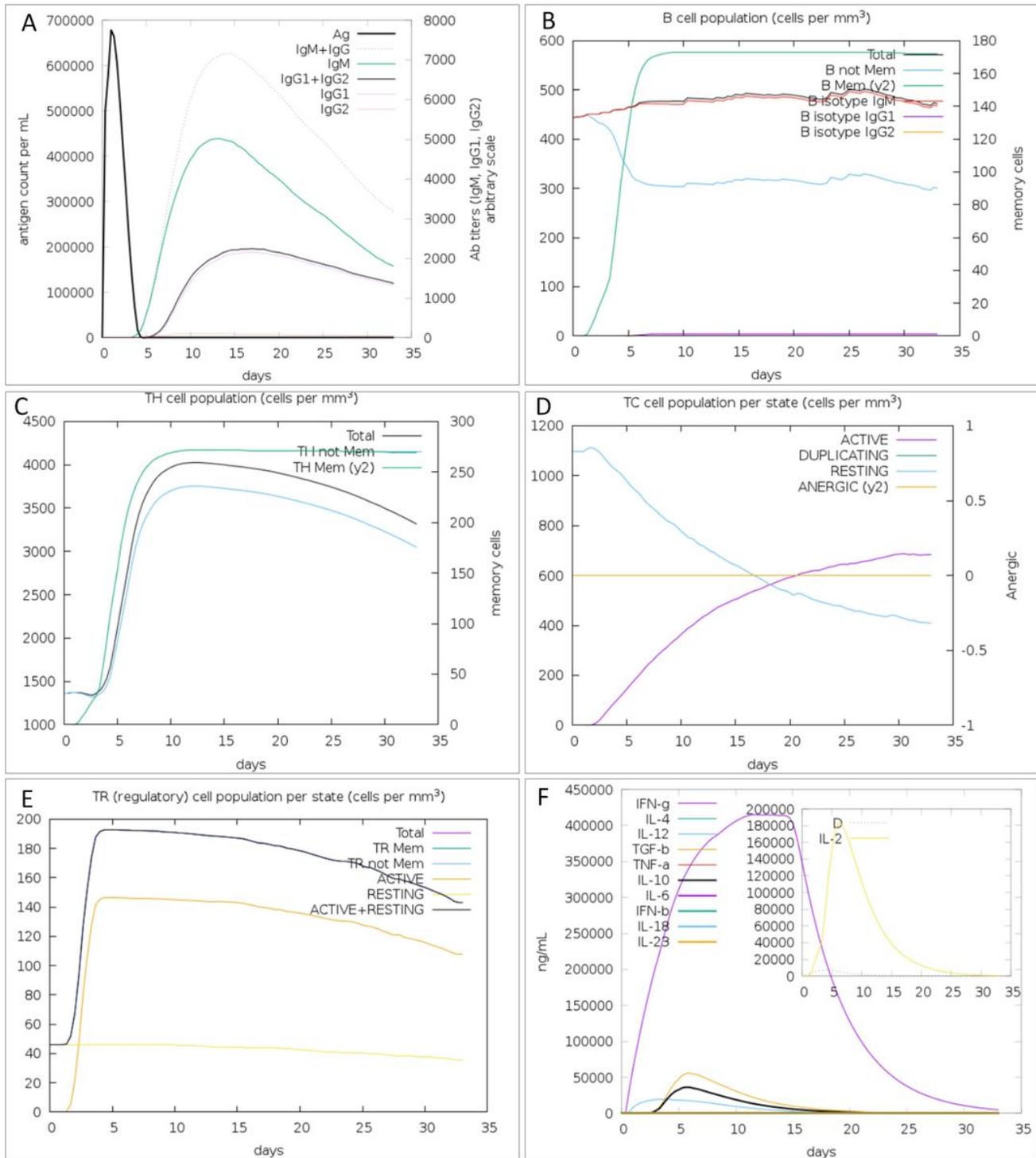
**Figure 13**

In silico cloning to express final vaccine construct in *E. coli*. The vaccine construct gene (red) was inserted into the pET21b (+) vector (black) between NotI and XhoI restriction site to create a recombinant vector.



**Figure 14**

In silico immune simulation with the trivalent subunit vaccine. (A) Immunoglobulin production in response to antigen (B) B-cell populations after exposure to antigen (C) The evolution of T-helper, (D) T-cytotoxic and (E) T-regulatory cell populations per state after antigen injection. The resting state characterized cells not presented with the antigen while the anergic state showed tolerance of the T-cells. (F) Level of cytokines induced by vaccine. The insert plot shows IL-2 level with the Simpson index, D indicated by the dotted line.



**Figure 14**

In silico immune simulation with the trivalent subunit vaccine. (A) Immunoglobulin production in response to antigen (B) B-cell populations after exposure to antigen (C) The evolution of T-helper, (D) T-cytotoxic and (E) T-regulatory cell populations per state after antigen injection. The resting state characterized cells not presented with the antigen while the anergic state showed tolerance of the T-cells. (F) Level of cytokines induced by vaccine. The insert plot shows IL-2 level with the Simpson index, D indicated by the dotted line.