Designing of a next generation multiepitope based vaccine (MEV) against SARS-COV-2: Immunoinformatics and *in silico* approaches

Muhammad Tahir ul Qamar[a†], Abdur Rehman[b†], Usman Ali Ashfaq[b*], Muhammad Qasim Awan[b], Israr Fatima[b], Farah Shahid[b], Ling-Ling Chen[a*]

[a]*College of Life Science and Technology, Guangxi University, Nanning, P. R. China*

[b]*Department of Bioinformatics and Biotechnology, Government College University Faisalabad, Pakistan*

[†]**These authors contributed equally in this study**

***Correspondence authors**:

Prof. Ling-Ling Chen; llchen@mail.hzau.edu.cn;

Dr. Usman Ali Ashfaq: usmancemb@gmail.com

Abstract

Coronavirus disease 2019 (COVID-19) associated pneumonia caused by severe acute respiratory coronavirus 2 (SARS-COV-2) was first reported in Wuhan, China in December 2019. Till date, no vaccine or completely effective drug is available for the cure of COVID-19. Therefore, an effective vaccine against SARS-COV-2 is needed to be design. This study was conducted to design an effective multi-epitope vaccine (MEV) against SARS-COV-2. Seven antigenic proteins were taken as a target and epitopes (B cell, IFN- γ and T cell) were predicted. Highly antigenic and overlapping epitopes were shortlisted. Selected T cell epitopes indicated significant interactions with the HLA-binding alleles and 99.29% coverage of the world's population. Finally, 505 amino acids long MEV was designed by connecting sixteen MHC class I and twelve MHC class II epitopes with suitable linkers and adjuvant. Linkers and adjuvant were added to enhance the immunogenicity response of the vaccine. The allergenicity, physiochemical properties, antigenicity and structural details of MEV were analyzed in order to ensure safety and immunogenicity. MEV construct was non-allergenic and antigenic. Molecular docking demonstrated a stable and strong binding affinity of MEV with TLR3 and TLR8. Codon optimization and *in silico* cloning ensured increased expression in the *Escherichia coli* K-12 system. However, to ensure its safety and immunogenic profile, the proposed vaccine needs to be experimentally validated.

KEY WORDS: SARS-COV-2; COVID-19; Pneumonia; Epitopes; Vaccine; Linkers; Adjuvant

1. Introduction

Viruses are dangerous and can cause irreparable losses to human. The world hardly learns to deal with a virus when new emerges and threatens the future of humanity. A similar situation arose when a new strain of coronavirus not previously identified in humans was reported last year [1]. Positive-sense RNA viruses called corona viruses belong to the Coronaviridae family that are distributed broadly among human and mammals. In the last two decades there have been more than 10,000 reported infections of two types of coronaviruses such as severe acute respiratory coronavirus (SARS-COV) or Middle East Coronavirus (MERS-COV) [2]. The type of pneumonia caused by COVID-19 is a highly infectious disease and an ongoing epidemic has been declared by WHO as a global health emergency. COVID-19 pneumonia was first reported in Wuhan, Hubei province, China, in December 2019 and subsequently occurred in Hubei province and other parts of the country [3-5]. A novel coronavirus was indicated root cause of COVID-19 through deep sequencing analysis from lower respiratory tract samples, and later termed as SARS-COV-2 [6]. SARS-COV-2 strain, according to World Health Organization, is almost 70% like the SARS-COV strain, and 40% similar to the MERS-COV strain [7]. Symptoms of SARS-COV-2 may occur within 2 days or upto 14 days after exposure. Symptoms such as fever, diarrhea and respiratory disorder are present in infected people [4]. According to the recent research SARS-CoV-2 has an identical genomic organization as of beta-coronaviruses, 5'-untranslated region (UTR) includes orf1ab (replicas complex), nsps (encoding non-structural proteins), S (spike protein) and 3'-UTR includes E- protein (envelope protein), M-protein (membrane protein), Oraf6, orf7a, orf8, N-protein (nucleocapsid protein), orf10, and a lot of unknown non-structural open reading frames [3,8].

There is currently no vaccine or approved treatment for humans, but the treatment options for COVID-19 could be Chinese traditional medicine such as ShuFengJieDu Capsules and Lianhuaqingwen Capsules. Nonetheless, no clinical trials support the safety and efficacy of these medicinal products [9,10]. Many other options could be used for the treatment of COVID-19 such as peptides, vaccines, small molecule drugs and monoclonal-oligonucleotides based therapies but these preventive methods need month to years for development [11]. There is no clinical trial-based vaccine present. To prevent viral diseases vaccine is the most effective method. The availability of genomic, software algorithms and immunological data has facilitated scientists to identify the effective epitopes that can be used to develop active subunit vaccines [12,13]. The subunit vaccine contains the fragments of antigenic proteins that can mimic the presence of the natural pathogen and cause an immune response to the target pathogen [14,15]. The design of the vaccine candidate against chikungunya, MERS virus, Ebola virus and Zika has produced promising results [16-19]. The *in silico* methods reduce the number of *in vitro* experiments and save time, overcome cost obstacles and increase the potential for successful vaccine design [20,21]. A lot of peptides involved in a multi epitope vaccine that induces the activation of adaptive-immune-response are the best technique for the treatment of viral-infections and SARS-CoV-2 [22-25].

In the recent study, SARS-CoV-2 proteome was explored to determine the antigenic proteins and various T-cell and B-cell epitopes were predicted with their MHC (major histocompatibility complex) alleles. Antigenicity of these epitopes was checked. Finally, multi-epitope vaccine (MEV) was designed using the most potent epitopes with suitable adjuvant and linkers. Online tools analyze the antigenicity, allergenicity, structural detail and physiochemical properties of the MEV. Molecular docking analyzed the binding interaction and stability of the

vaccine-receptor complex. At the end, the vaccine codon was optimized for *E. coli* system and i*n silico* cloning was performed (Figure 1).

2. Material and methods

2.1. Sequence retrieval

In the first step, the sequence of target proteins was retrieved from GENBANK [26]. Then all the sequences were stored as FASTA format for further analysis.

2.2. Antigenicity and physiochemical properties evaluation

The Expassy Protparam tool was used to determine the physical and chemical properties of proteins [27]. To check protein antigenicity, the Vaxijen 2.0 software was also used [28]. The threshold value was held at 0.5, and the secondary structure of proteins was predicted by using SOPMA (Alignment self-optimized prediction method) tool [29].

2.3. Tertiary structure prediction of target proteins

Different online tools such as Swiss model, Phyre 2, and Raptor X were used for the tertiary structure prediction of SARS-COV-2 proteins [30-32]. Models retrieved were then refined by galaxy refine server and validated by Ramachandran plot analysis.

2.4. Epitope prediction

Inducing epitope (MHC I and II and IFN-γ) is shown in the sequence of amino acids of all proteins.

2.4.1. B-cell epitope prediction

A surface receptor of B-cell recognizes B-cell epitopes, resulting the generation of antigen-specific immunoglobulins [33]. In immune system the B-Cell epitope helps to detect viral infection and activities. An online database of ABCPred was used to predict 14-mer B-cell epitopes [34,35]. Conformational epitopes were predicted by Ellipro server [36].

2.4.2. T-cell Epitope prediction

The most important step to develop an epitope-based vaccine is selection of T-cell epitopes with strong binding affinity to appropriate HLA molecules [4]. Current challenge in the immunological prediction of T-cell epitope sequences is the accurate prediction of interacting molecules. Most popular method considered, for epitope prediction is binding affinity prediction for a range of MHC molecules [5]. It is necessary for recognition by cytotoxic T-cells through molecular binding between antigenic peptides and MHC molecules. Therefore, it is crucial to identify MHC-binding peptides by any T-cell epitope prediction algorithm. For a functional T-cell response, peptide binding to MHC and the interaction of this complex with a specific T-cell receptor are required [6]. The IEDB consensus method was used to predict 12 mer MHC classes I and 15 mer MHC II epitopes. Due to a large number of HLA alleles used in this method, results are very important. Sequence was given in a FASTA format, selecting all alleles for prediction. To be considered a good binder, epitopes with a consensus score of less than 2 was used, and selected for further research [7]

2.5. Immunogenicity prediction of peptides

To evaluate the antigenicity of T-cell and B cell epitope, Vaxijen v2.0 was used. 0.5 was used as default value. Non-antigenic epitopes were removed, and antigenic epitopes were selected.

2.6. Conservation of peptides analysis

Immune Epitope Database (IEDB) Conservancy Analysis tool was used, to monitor the degree of conservation in the protein sequence of B cell and T cell epitopes [8]. Epitopes showing 100 percent conservation were selected for further analysis.

2.7. Interferon-Y epitope prediction

IFN-γ is acknowledged to elicit intrinsic safe responses and can directly detain viral duplication. Besides, they can validly actuate the versatile immune reaction by preparing cytotoxic T cells and T helper cells. IFN epitope server calculates IFN-γ epitopes of selected proteins of SARS-CoV-2 using SVM hybrid algorithms along with Motif [37].

2.8. Population coverage analysis of selected epitopes

Predicting a T-cell epitope is not sufficient to become a good candidate for a vaccine. The predicted peptide(s) should effectively cover people in major areas. For population coverage, the separate epitopes of selected T-cells (class I and II) with the related HLA alleles were submitted to the IEDB population coverage analysis tool by maintaining the default analysis parameters [38]. The population coverage analysis tool forecast and calculates each epitope of various regions of the world based on the distribution of human alleles that bind to MHC.

2.9. Construction of multi epitope vaccine

To construct a vaccine sequence, a high scoring CTLs, high-affinity HTLs epitopes and simultaneously B cell epitope predicted by ABCPRED were selected. It was observed some sequence may act as both B and T cell epitope. The chosen epitopes should be preserved overlapping and immunogenic. The adjuvant and first CTL epitope were combined with the support of the EAAAK while different epitopes were connected using AAY and GPGPG linkers.

2.10. Structural analysis of vaccine construct

To determine the physiochemical properties of MEV the ProtParam tool was used. It analyzes various physical and chemical attributes that depended on the pK of amino acids involved. The theoretical pI, Grand Average Hydropathy, Stability Profiling, Instability Index, Half-Life, and Aliphatic Index were checked. AllerTOP V2.0 server was used to evaluate the allergic and non-allergic nature of the vaccine [39]. To check the antigenicity of vaccine construct the vaxijen2.0 server was used. This identification depends on various protein physicochemical characteristics and is alignment-free in VaxiJen. Secondary structure was predicted by GOR4 and PSIPRED [40,41].

2.11. Tertiary structure and validation of MEV

The swiss model server was used to develop final MEV structure. The protein homology model was developed based on the quality of the QMEAN model. It gives results as QMean value and GMQE value. GMQE (Global Model Quality Estimation) reflects the certainty of a model by taking into account the coverage, template, and arrangement of the target. It also offers a

calculation of quality by combining target-template alignment properties with the template search method. The higher the GMQE value, the better the model's quality. It is usually calculated in 0 and 1 range. Globally, QMean Z-score provides an intuitive level of basic highlights present in a model. If the score is around zero, the apprehension among the virtually identical size model structures and test structures ought to be remarkable. However, if this value is below or -4.0 the model is considered a low-quality model [32]. The Galaxy Refine server has been used for strong and delicate methods for refining the structure. It produces several models with structure deviations from the given structure. Ramachandran plot was created to approve the tertiary structure of the vaccine developed by the online server RAMPAGE[42]. Consequently, ProSA-web has been used as a quality score for the whole structure of the modified vaccine protein [43]. The quality score outside the usual range of native proteins indicates possible errors in the predicted protein structure. To evaluate the statistics of Non-bonded interactions ERRAT server was used [44].

2.12. Molecular docking of immune receptor

The interaction of vaccine molecule with the immune receptor was analyzed by the molecular docking. HDOCK carried out the molecular docking of the MEV with TLR3 and TLR8, to confirm the immune reaction. Protein-protein docking and protein-DNA/RNA docking was performed in HDOCK [45].

2.13. In silico cloning and codon optimization

Codon use differs from species to species in the organism and therefore unadapted codon can lea d to a lower host rate of expression. It should, therefore, be optimized to improve the gene expression by the host translation machinery. Codon optimization and reverse translation are

examined by the tool of java codon adaptation to allow for proper vector translation expression and cloning efficiency. *Escherichia coli* (K12 strain) was identified as a host organism for the expression of MEV. Data were collected in graphical representations and the CAI values [46].

3. Results

3.1. Pre-vaccine design analysis

The amino acid sequence of target protein (ORF1 [QHD43415.1], S [QHD43416.1], ORF3a [QHD43417.1], E [QHD43418.1], M [QHD43419.1], ORF6 [QHD43420.1], ORF7a [QHD443421.1], ORF8 [QHD43422.1], N [QHD43423.2] and ORF10 [QHI42199.1]) of SARS-CoV-2 were retrieved from Genbank in FASTA format. Vaxijen was used to check the antigenicity of these target proteins. Total 7 highly antigenic proteins were detected. The most antigenic ORF10 protein has been found followed by the E, M, ORF6, ORF7a, ORF8, and N having antigenic values of 0.7185, 0.6502, 0.6441, 0.6131, 0.6025, 0.5102 and 0.5059 respectively. ORF1, surface glycoprotein, and ORF3 proteins had antigenic values less than 0.5, so they were excluded. In addition, Blast p analysis was performed against Homo sapiens with predetermined parameters to identify non-homologous proteins. Proteins with less than 37% identity were considered non-homologous proteins. None of the protein showed significant similarity with the human proteins.

Five non-structural proteins (M, N, ORF6, ORF7a, and ORF10) and two structural proteins (E, ORF8) were selected as targets for further analysis. Other physicochemical characters like Theoretical pi, molecular weight, half-life, stability profile, aliphatic index, etc were predicted from protparam (Table 1) and secondary structure was predicted by SOPMA (Table 2).

3.2. Tertiary structure and validation

The 3D models obtained using Phyre 2, Raptor X and Swiss model homology approaches were evaluated through Ramachandran plot analysis. The model specifications and reliability were assessed using different homology modeling tools. The structures retrieved from Swiss model were of better quality than those from other homology modeling tools like the Raptor X and Phyre-2. There was no suitable template found for ORF10 because of small number of residues so its structure was predicted by PEPFOLD [47]. The target protein pattern alignment was also observed with precise statistics. Besides, Galaxy refine server was used to optimize the models (Table 3).

3.3. Prediction of B-cell epitope

ABCpred was used to predict the target protein Linear B cell epitopes. Vaxijen was used to check the antigenicity. Screened out B cell epitope were 100% conserved in protein sequence and are antigenic. All the target proteins were predicted to have a total 55 linear epitopes (E-4, M-12, ORF6-1, ORF7a- 6, ORF8-9, N- 22, and ORF10-1) (Table 4). For determination of conformational epitopes, the Ellipro server was used. A total of 24 (E-4, M-2, ORF6-3, ORF7a-4, ORF8-4, NC- 4, and ORF10-3) conformational epitopes were forecast of all proteins (Table 5).

3.4. T-cell prediction

T-cell epitopes of the target proteins were predicted by using the IEDB consensus method. 100% conserved in protein sequence and antigenic epitopes were selected. Peptides that can bind to multiple alleles because of their strong defense capabilities are considered the most appropriate peptide. Total 31 MHC class I (E-9, M-4, ORF6-2, ORF7a-3, ORF8-7, N-4, ORF10-2) (Table 6) and 40 MHC class II (E-4, M-5, ORF6-4, ORF7a-7, ORF8-12, N-4, ORF10-4) epitopes were chosen for further study (Table 7).

3.5. Population coverage

The distribution and expression of HLA alleles vary by ethnic groups and regions of the world. Therefore, it affects the successful development of an epitope-based vaccine. The IEDB population tool was used to check the population coverage of the designed vaccine. Selected epitopes showed the 99.29% world population coverage. The highest coverage of population found within a country Sweden 99.79% however, the least population coverage 5.72% was predicted in the United Arab Emirates. SARS-CoV-2 showed the population coverage of 84.51% in China where it was first identified (Figure 2).

3.6. MEV construction

A total 28 T cell epitopes were selected to form a vaccine construct.16 epitopes of MHCI (E-1, M-3,ORF6-1,ORF7a-3,ORF8-5,N-2,ORF10-1) and 12 epitopes of MHCII (E-1, M-13, ORF6-1, ORF7a-1, ORF8-2, N-3, ORF10-1) were merge with the AAY and GPGPG linkers respectively. GPGPG prevents the generation of junctional epitopes, which is a major concern in the design of epitope vaccines; On the other hand, it facilitates the immunization and presentation of HTL epitopes [48,49]. To join the CTL epitope, the AAY motif was used as a linker. The final length of the vaccine was 460 after long merging of adjuvant. Besides, the EAAAK connector was used as

an adjuvant to the N-end of the vaccine to attach 45 amino acid long β-defensin. The vaccine's immunogenicity was increased with an adjuvant. An adjuvant and a CTL epitope were connected to the N-terminal of the structure of the vaccine by using the EAAAK linker to reduce interaction with other protein regions with efficient separation [50,51]. After the addition of linkers and adjuvant final construct was 505 amino acid long (Figure 3).

3.7. Linear and conformational B epitopes and other physiochemical properties of MEV

ABCPred 2.0 and Ellipro servers have been used to predict linear/continuous and conformational/discontinuous B cell epitopes of the vaccine model without changing the prediction parameters. 92 continuous/ linear and 3 conformational/discontinuous B cell epitopes were predicted by servers (Table 8 & 9). The physiochemical properties of SARS-CoV-2 vaccine construct computed via Protparam so it contains 505 amino acid with molecular weight 55426.35 KDa which reflects good antigenic nature. Theoretical isoelectric point (PI) of MEV was 9.12 which indicates negative in nature. The isoelectric point less than 7 showed negatively charged protein. Instability index computed by the protparam is 33.41, this categories protein as a stable. Aliphatic index 82.75 which devotes a thought of proportional volume hold by aliphatic side chain and GRAVY value for protein sequence is 0.105 which indicates the hydrophobic nature of MEV. Half-life of protein depicted as the total time taken for its vanishing after it has been synthesized in cell, which was computed as 30 h for mammalian-reticulocytes, > 20 h for yeast, > 10 h for *Escherichia coli*. Total number of Carbon (C), Oxygen (O), Nitrogen (N), Hydrogen (H) and Sulfur (S) were entitled by formula $C_{2549}H_{3850}N_{666}O_{669}S_{28}$.

3.8. Prediction of allergenicity and antigenicity

AllerTOP was used to check the allergenicity of the vaccine construct which describe the non-allergenic behavior of vaccine. The vaccine antigenicity was 0.6741 at 0.5% threshold, indicating the antigenic nature of the vaccine, according to the VaxiJen.

3.9. Prediction of Secondary structure

Secondary structure of MEV was predicted by PSIPRED and GOR4. Among the 505 amino acids, the formation of α-helix is comprised of 113 amino acids representing 22.38%, 0 in β-strands and 271 amino acids forms the coils which are 53.66% of the whole MEV construct.

3.10. Tertiary structure prediction and refinement

To determine the tertiary structure of vaccine Swiss model was used. GMQE was 0.04 and Q - 2.88, which indicates the high quality of the vaccine. Structure was improved by Galaxy refine. After examining the modifications of 5 models, Model 2 was found to be the best model and was selected for further analysis (Figure 4). The improved model exhibits 95% favored region in RAMPAGE and qRMSD as 0.428, poor rotamers as 0%, MolProbity as 1.889, clash score as 13.6 and RAMPAGE server analyzes and validates the tertiary structure by producing Ramachandran plot. In the more favorite region, 96.3% of residues are generated, 3.7% of amino acids are resident in the permitted region, and 0.0% are present in outer regions according to the vaccine construct Ramachandran plot. ProSA-web also gives−2.25 Z, which is within an acceptable range of values. In addition, the refined model showed 0 errors with PROCHECK. The refined model was 85.7143 in total quality (ERRAT score). These results show that the refined model is of good quality.

3.11. Molecular docking with TLR3 and TLR8

An appropriate association between immune receptor molecules and the antigen molecule is necessary to activate an immune responsiveness. HDOCK server has thus been used to perform the docking of the MEV with TLR3 and TLR8. The blue and red color shows the MEV in the docked complex, whereas TLR3 and TLR8 are depicted by a rainbow color respectively. The docking score and RMSD value of TLR3 and TLR8 with MEV were -244.88, -204.07 and 75.80, 80.46 respectively (Figure 5).

3.12. In silico cloning

The codon adaptation tool, an online server, was used for codon optimization analysis evaluating a cDNA sequence followed by a codon adaptation index (IAC) and GC data. The GC content of the structure is in the ideal range 55.51% (30-70%), and CAI 1.0 (0.8-1.0), indicating the effective expression of the protein reliability.

4. Discussion

Vaccination has many useful effects for improving people health in a low-cost manner and best aid to inhibit transmission of diseases around the world. Today researchers are searching methods for the development of subunit vaccines for complete genome. Vaccines are basically consisting of different immunogenic ingredients of disease (pathogens), comparatively the whole pathogenic agent [52]. Epitope prediction for antibodies becomes more significant with the advancement of the computational tools for designing a vaccine [53]. In the field of bioinformatics Immuno-informatics is a sub-branch that includes a lot of tools & databases. Immunological datasets prediction and *in silico* analysis are done with the help of those tools. With the advancement of tools and a variety of data availability like genomic, proteomic, and different

algorithms made it more effective for scientists to predict epitopes that are much effective in the development of the sub-unit vaccines [12,13,54].

An outbreak of Wuhan SARS-COV-2 in late December 2019 resulted in hundreds of deaths in China [6]. After the outbreak of Wuhan, remarkable progress has been made towards the identification, genome, and structure of virus and the development of effective drugs against SARS-COV-2, yet its injection mechanism is still unclear. Till date, no vaccine or effective drug is available for the cure of COVID-19. Coronavirus vaccines need to be designed so that they can be eradicated form the globe entirely.

The recent study was conducted to design a multi-epitope vaccine (MEV) against SARS-COV-2 by using immunoinformatics approach. MEV is advantageous compared to monovalent vaccine because it can elicit cellular and humoral immunity [55]. Two subunit vaccines against SARS-COV-2 was reported where they have used one and three proteins respectively for vaccine development, but this study included more number of proteins and epitopes [10,37]. The current vaccine subunit is therefore more effective and immunogenic against COVID-19. Amino acid sequences of ten proteins of SARS-COV-2 were taken from Genbank and their antigenicity was checked. Highly antigenic proteins were selected for further analysis. After the complete physiochemical analysis of antigenic proteins, B cell and T cell epitopes were predicted. Predicting the B-cell and T-cell epitopes is an important step in vaccine development that's why both types of epitopes were predicted form protein sequences and their antigenicity was evaluated [56]. After that antigenic T-cell epitopes were screened to overlap with B-cells and IFN-γ epitopes. Vaccine was designed with the help of linkers and adjuvant. An adjuvant was added to the N-end of the vaccine and epitopes were linked with the help of AAY and GPGPG linkers. Adjuvant was added to increase the immunogenicity of the vaccine. Linkers were added to help

maintain the function of each epitope so that after being imported into the human body they can function independently [31].

Finally, 505 amino acids long vaccine was designed including adjuvant. The vaccine protein was 55426.35 kDa in molecular weight. MEV was basic, according to the theoretical pI value, which can ensure stable physiological pH interaction. The calculated aliphatic index and instability index scores showed that the vaccine protein may be stable and thermostable. Positive score of grand average of hydropathy suggests its hydrophobic nature. MEV was found to be highly antigenic, immunogenic and non-allergenic indicating an epitope vaccine's potential to cause robust immune responses without any allergic reactions. The effectiveness of MEV depends on the population in which the vaccine is used. T cell epitopes included in the MEV showed 99.29% of the world's population, indicating that the vaccine developed would be effective for most of the world population. The strong binding affinity of the vaccine with immune receptor (TLR3 & TLR8) is necessary to effectively transport vaccine into the body. For this purpose, docking between the TLR3, TLR8 and MEV was performed. The translation efficiency of foreign genes within the host system differs due to the incompatibility of mRNA codons, which require codon optimization for higher expression [31]. CAI value obtained was 1.0 and GC content 55.51% was also within the optimum limit indicating possible higher expression within *E. coli* K-12 system. Thus, MEV designed with caution using such a methodology could become an integral asset in the fight against tumors and viral contaminations. The results of recent study suggest that the vaccine being designed may undergo *in vivo* and *in vitro* experimental analysis to develop a potential vaccine against COVID-19.

5. Conclusion

A serious global problem is morbidity and mortality, nowadays COVID-19 has, unfortunately, affected several precious lives in various regions of the world due to a lack of SARS-COV-2 vaccination. Several antiviral medicines have been tested but none have shown effective results against the infection. In this study, an attempt was made to design an MEV against SARS-COV-2 that targets its antigenic proteins. Immunoinformatics and *in silico* approached were used to develop potential and safe MEV that can trigger three types of the immune responses: humoral, innate and cellular. However, current research is the result of an integrated vaccinomics approach. Therefore, experimental validation is necessary to demonstrate the efficacy and safety of the designed vaccine.

Acknowledgements

This work was supported by the Starting Research Grant for High-level Talents from Guangxi University and Postdoctoral research platform grant of Guangxi University.

Authors' contributions

MTQ, LLC and UAA conceived and designed this study; MTQ and AR performed the experiments; MQA, IF and FS analyse the results; MTQ and AR wrote the manuscript; UAA and LLC improved and revised the manuscript, and all the authors approved the final version.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

1 Tripp, R. A. & Tompkins, S. M. *Roles of Host Gene and Non-coding RNA Expression in Virus Infection*. Vol. 419 (Springer, 2018).

2 Xu, X. *et al.* Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Science China Life Sciences*, 1-4 (2020).

3 Zhu, N. *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine* (2020).

4 Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* (2020).

5 Tahir ul Qamar, M., Alqahtani, S. M., Alamri, M. A. & Chen, L.-L. Structural Basis of SARS-CoV-2 $3CL^{pro}$ and Anti-COVID-19 Drug Discovery from Medicinal Plants. *Preprints* **2020020193**, doi:doi: 10.20944/preprints202002.0193.v1 (2020).

6 Hui, D. S. *et al.* The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China. *International Journal of Infectious Diseases* **91**, 264-266 (2020).

7 Alamri, M. A., Tahir ul Qamar, M. & Alqahtani, S. M. Pharmacoinformatics and Molecular Dynamic Simulation Studies Reveal Potential Inhibitors of SARS-CoV-2 Main Protease 3CLpro. *Preprints* **2020020308**, doi:10.20944/preprints202002.0308.v1 (2020).

8 Ahmed, T. *et al.* Coronavirus Disease 2019 Assosiated Pneumonia in China: Current Status and Future Prospects. *Preprints* **2020020358**, doi:10.20944/preprints202002.0358.v3 (2020).

9 Lu, H. Drug treatment options for the 2019-new coronavirus (2019-nCoV). *BioScience Trends* (2020).

10 Abdelmageed, M. I. *et al.* Design of multi epitope-based peptide vaccine against E protein of human 2019-nCoV: An immunoinformatics approach. *BioRxiv* (2020).

11 Li, G. & De Clercq, E. (Nature Publishing Group, 2020).

12 De Gregorio, E. & Rappuoli, R. Vaccines for the future: learning from human immunology. *Microbial biotechnology* **5**, 149-155 (2012).

13 Patronov, A. & Doytchinova, I. T-cell epitope vaccine design by immunoinformatics. *Open biology* **3**, 120139 (2013).

14 Saadi, M., Karkhah, A. & Nouri, H. R. Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. *Infection, Genetics and Evolution* **51**, 227-234 (2017).

15 Nain, Z., Karim, M. M., Sen, M. K. & Adhikari, U. K. Structural Basis and Designing of Peptide Vaccine using PE-PGRS Family Protein of Mycobacterium ulcerans–An Integrated Vaccinomics Approach. *bioRxiv*, 795146 (2019).

16 Tahir ul Qamar, M. *et al.* Peptide vaccine against chikungunya virus: immuno-informatics combined with molecular docking approach. *Journal of translational medicine* **16**, 298 (2018).

17 Tahir ul Qamar, M. *et al.* Epitope□based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. *Journal of Translational Medicine* **17**, 362, doi:10.1186/s12967-019-2116-8 (2019).

18 Ahmad, B., Ashfaq, U. A., Rahman, M.-u., Masoud, M. S. & Yousaf, M. Z. Conserved B and T cell epitopes prediction of ebola virus glycoprotein for vaccine development: an immuno-informatics approach. *Microbial pathogenesis* **132**, 243-253 (2019).

19 Shahid, F., Ashfaq, U. A., Javaid, A. & Khalid, H. Immunoinformatics guided rational design of a next generation multi epitope based peptide (MEBP) vaccine by exploring Zika virus proteome. *Infection, Genetics and Evolution* **80**, 104199 (2020).

20 Mirza, M. U. *et al.* Towards peptide vaccines against Zika virus: Immunoinformatics combined with molecular dynamics simulations to predict antigenic epitopes of Zika viral proteins. *Scientific reports* **6**, 37313 (2016).

21 Khan, A. *et al.* Computational identification, characterization and validation of potential antigenic peptide vaccines from hrHPVs E6 proteins using immunoinformatics and computational systems biology approaches. *PloS one* **13** (2018).

22 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 (2017).

23 Jensen, K. K. *et al.* Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* **154**, 394-406 (2018).

24 Lu, C. *et al.* A novel multi□epitope vaccine from MMSA□1 and DKK 1 for multiple myeloma immunotherapy. *British journal of haematology* **178**, 413-426 (2017).

25 Chauhan, V., Rungta, T., Goyal, K. & Singh, M. P. Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach. *Scientific reports* **9**, 1-15 (2019).

26 Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic acids research* **37**, D26-D31 (2008).

27 Walker, J. M. *The proteomics protocols handbook*. (Springer, 2005).

28 Doytchinova, I. A. & Flower, D. R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC bioinformatics* **8**, 4 (2007).

29 Deléage, G. ALIGNSEC: viewing protein secondary structure predictions within large multiple sequence alignments. *Bioinformatics* (2017).

30 Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nature Protocols* **7**, 1511-1522, doi:10.1038/nprot.2012.085 (2012).

31 Pandey, R. K., Bhatt, T. K. & Prajapati, V. K. Novel immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating anopheles salivary protein. *Scientific reports* **8**, 1-11 (2018).

32 Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* **46**, W296-W303 (2018).

33 Cooper, M. D. The early history of B cells. *Nature Reviews Immunology* **15**, 191-197 (2015).

34 Zheng, J. (2017).

35 Kolaskar, A. & Tongaonkar, P. C. A semi□empirical method for prediction of antigenic determinants on protein antigens. *FEBS letters* **276**, 172-174 (1990).

36 Ponomarenko, J. *et al.* ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC bioinformatics* **9**, 514 (2008).

37 Tahir ul Qamar, M., Shahid, F., Ali, U., Fareed, A. Z. & Chen, L.-L. Structural modeling and conserved epitopes prediction against SARS-COV-2 structural proteins for vaccine development. *Research Square*, doi:10.21203/rs.2.23973/v1 (2020).

38 Bui, H.-H., Sidney, J., Li, W., Fusseder, N. & Sette, A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC bioinformatics* **8**, 361 (2007).

39 Dimitrov, I., Flower, D. R. & Doytchinova, I. in *BMC bioinformatics.* S4 (BioMed Central).

40 McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404-405 (2000).

41 Sen, T. Z., Jernigan, R. L., Garnier, J. & Kloczkowski, A. GOR V server for protein secondary structure prediction. *Bioinformatics* **21**, 2787-2788 (2005).

42 Lovell, S. C. *et al.* Structure validation by Cα geometry: □, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics* **50**, 437-450 (2003).

43 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 (2007).

44 Lengths, M. & Angles, M. Limitations of structure evaluation tools errat. *Quick Guideline Comput Drug Des* **16**, 75 (2018).

45 Yan, Y., Zhang, D., Zhou, P., Li, B. & Huang, S.-Y. HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic acids research* **45**, W365-W373 (2017).

46 Grote, A. *et al.* JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic acids research* **33**, W526-W531 (2005).

47 Lamiable, A. *et al.* PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic acids research* **44**, W449-W454 (2016).

48 Nezafat, N. *et al.* Designing an efficient multi-epitope peptide vaccine against Vibrio cholerae via combined immunoinformatics and protein interaction based approaches. *Computational biology and chemistry* **62**, 82-95 (2016).

49 Livingston, B. *et al.* A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. *The Journal of Immunology* **168**, 5499-5506 (2002).

50 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* **14**, 529-532 (2001).

51 Pandey, R. K. *et al.* Exploring dual inhibitory role of febrifugine analogues against Plasmodium utilizing structure-based virtual screening and molecular dynamic simulation. *Journal of Biomolecular Structure and Dynamics* **35**, 791-804 (2017).

52 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. *Sci Rep* **7**, 8285, doi:10.1038/s41598-017-08842-w (2017).

53 Dubey, K. K. *et al.* Vaccine and antibody production in plants: developments and computational tools. *Briefings in functional genomics* **17**, 295-307 (2018).

54 Yang, X. & Yu, X. An introduction to epitope prediction methods and software. *Reviews in medical virology* **19**, 77-96 (2009).

55 Amanna, I. J. & Slifka, M. K. Contributions of humoral and cellular immunity to vaccine-induced protection in humans. *Virology* **411**, 206-215 (2011).

56 Van Regenmortel, M. H. Mapping epitope structure and activity: from one-dimensional prediction to four-dimensional description of antigenic specificity. *Methods* **9**, 465-472 (1996).

Figures

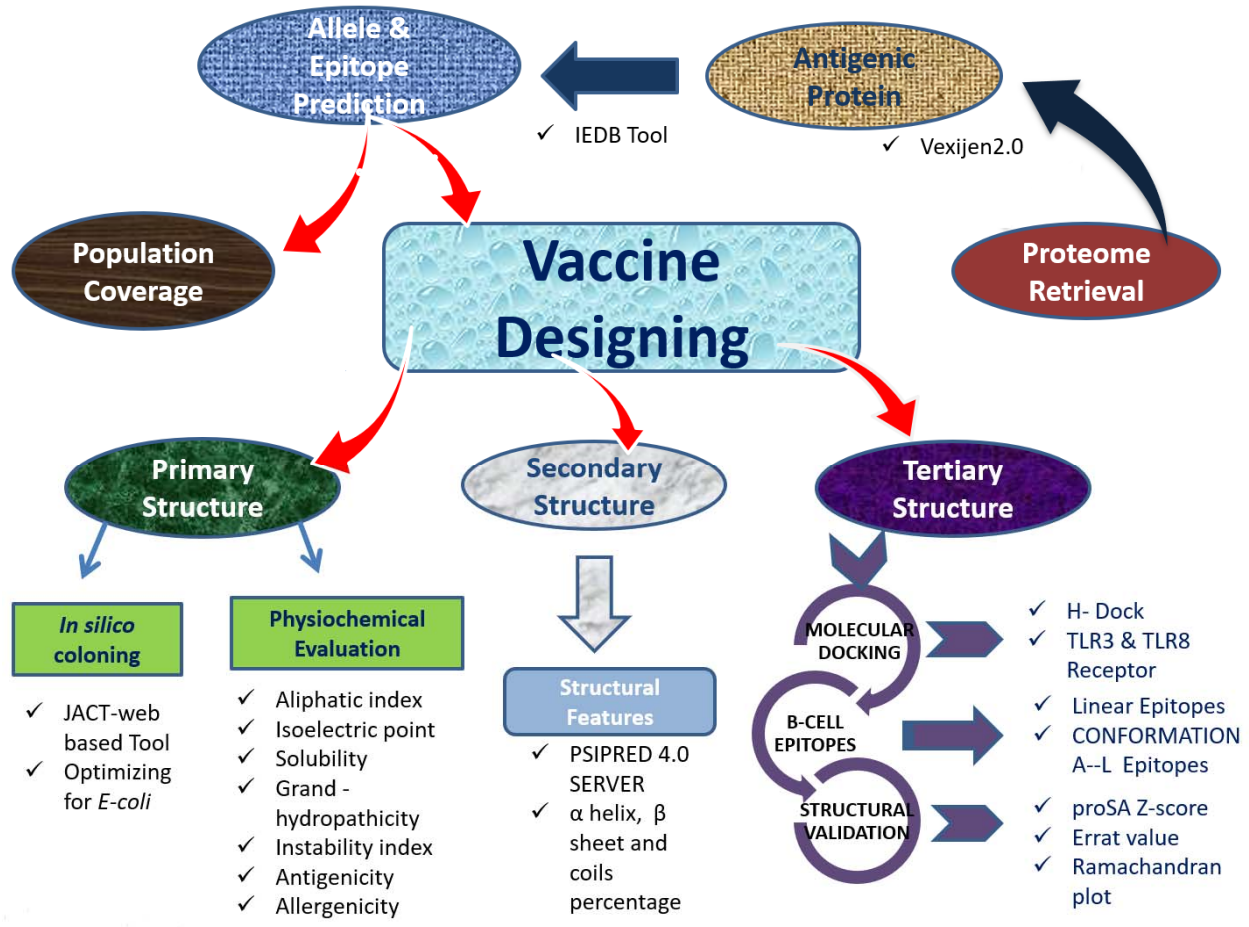


Figure 1. The overall experimental workflow used to develop MEV against SARS-COV-2.

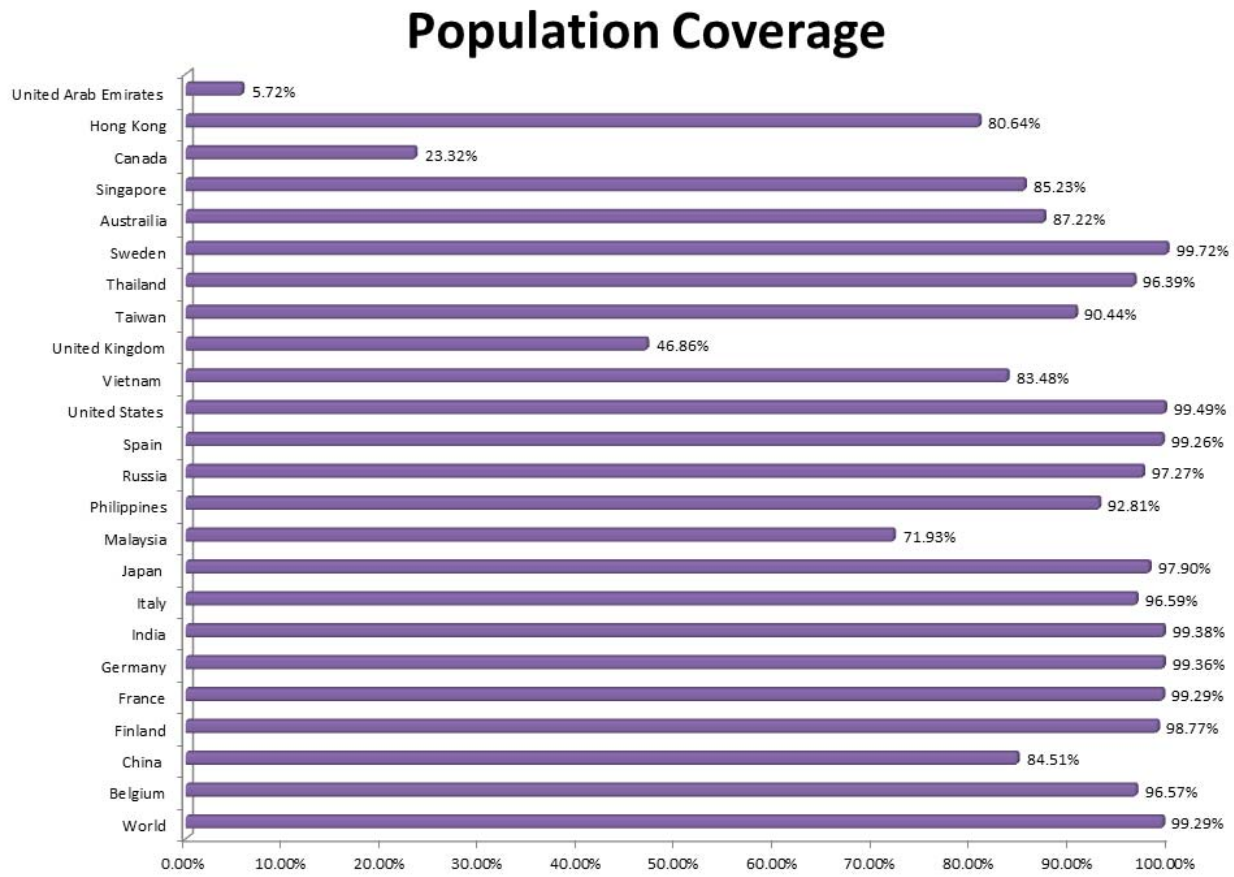


Figure 2. Worldwide population coverage by T-cell epitopes based on their respective HLA binding alleles.

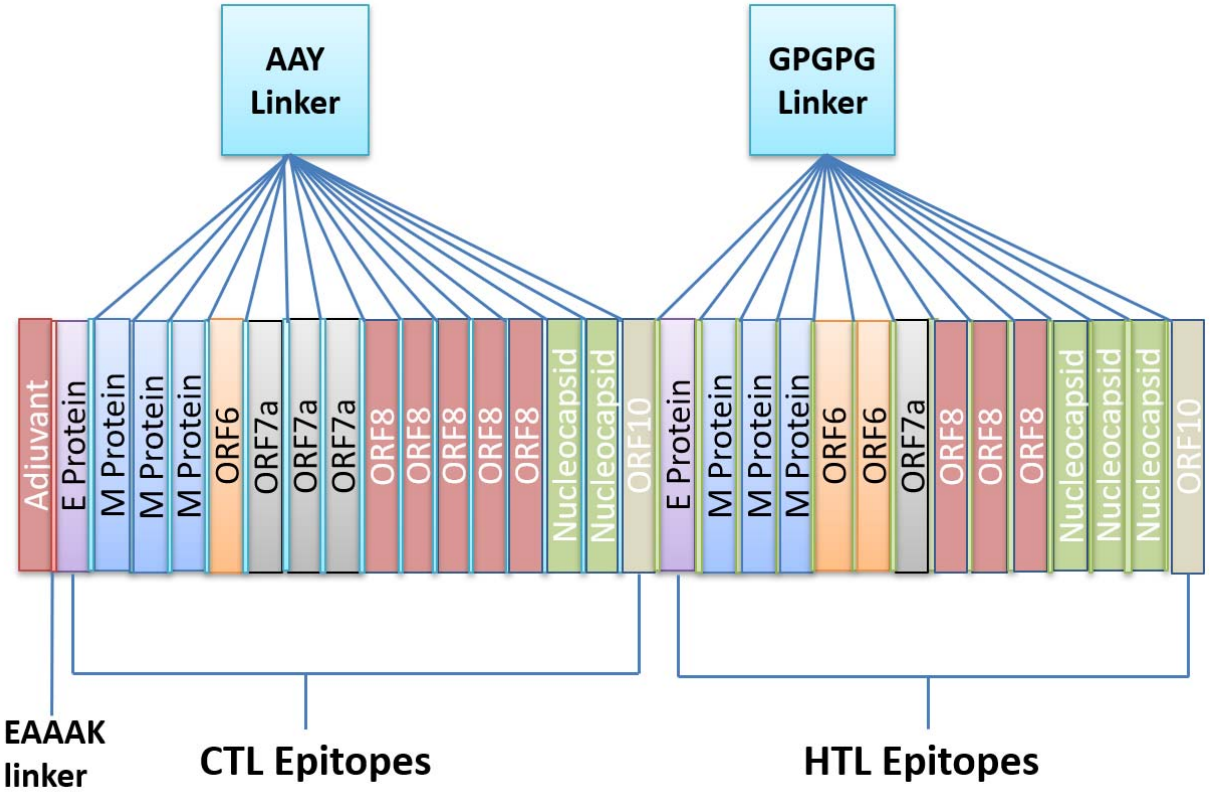


Figure 3. Schematic diagram of MEV construct: A 505 amino acid long MEV sequence consisting an adjuvant (Maroon) linked at N-terminal linked with a MEV sequence with the help of EAAAK linker (Pink). CTL epitopes are joined by AAY linkers (Blue) while HTL epitopes are joined by GPGPG linkers (Green).

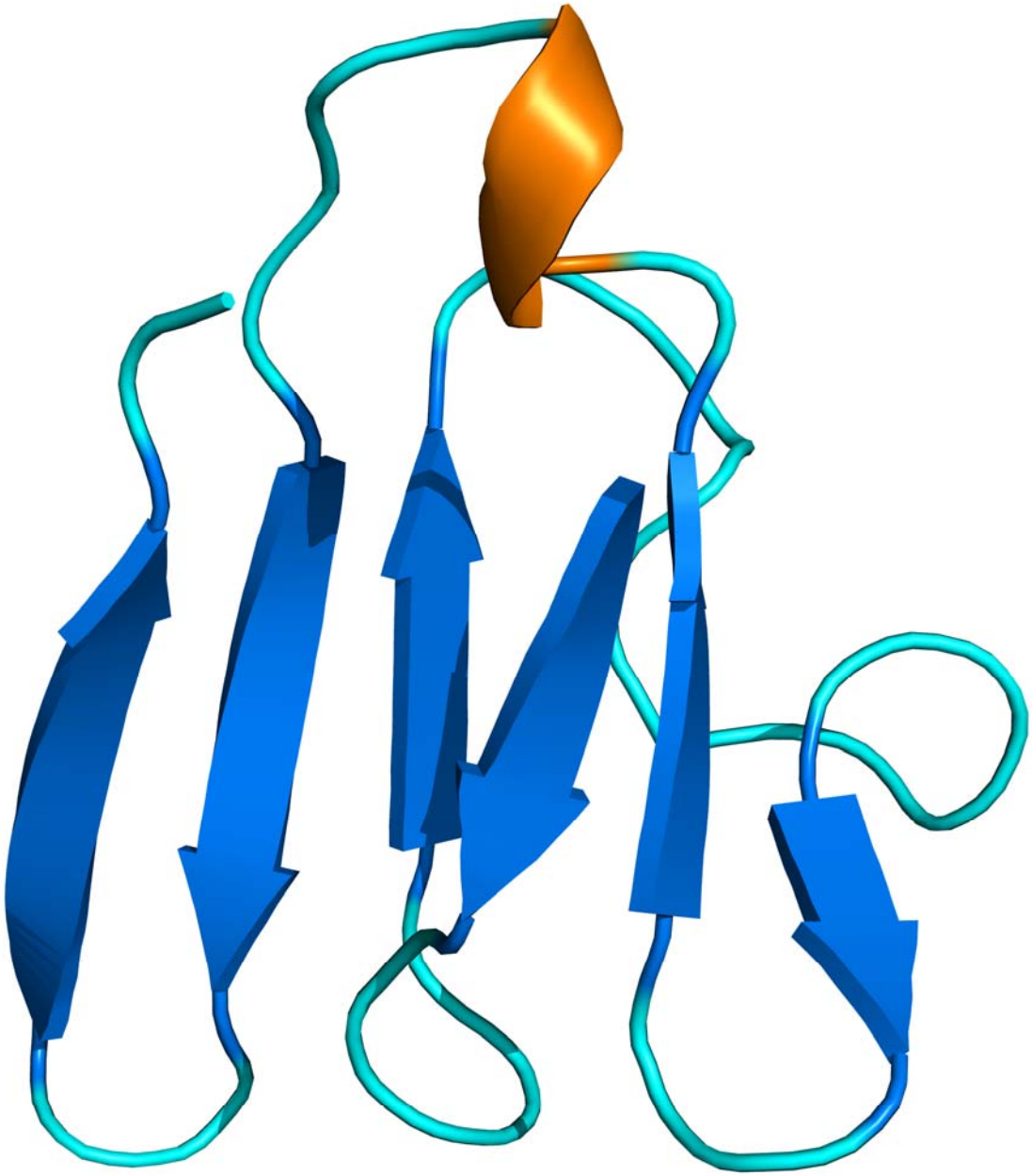


Figure 4. The three-dimensional structure of the MEV vaccine construct (Alpha helixes: Bown; Beta sheets: Blue; Loops: Cyan).

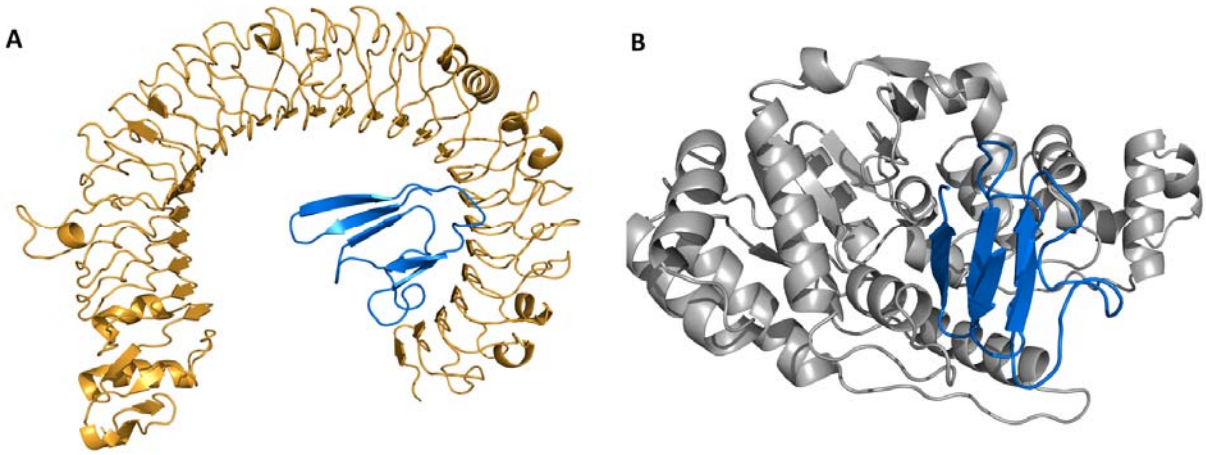


Figure 5. (A) MEV Construct-TLR3 docked complex. TLR-3 displayed with brown color and MEV vaccine construct displayed with blue color. (B) MEV Construct-TLR8 docked complex. TLR-8 displayed with grey color and MEV vaccine construct displayed with blue color.

Tables

Table 1. Physiochemical properties of the SARS-COV-2 proteins analyzed by Protparam Tool

Proteins	Molecular Weight	Theoretical pI	Instability index	Half□life	Stability Profiling	Aliphatic Index	Grand Average of Hydropathy
Envelope protein	8365.04	8.57	38.68	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).	stable	144.00	1.128
Membrane glycoprotein	25146.62	9.51	39.14	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).	stable	120.86	0.446
ORF6 protein	7272.54	4.60	31.16	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).	stable	130.98	0.233

ORF7a	13744.17	8.23	48.66	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo)	unstable	100.74	0.318
ORF8 protein	13831.01	5.42	45.79	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo) >10 hours (Escherichia coli, in vivo).	unstable	97.36	0.219
Nucleocapsid phosphoprotein	45625.70	10.07	55.09	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).	unstable	52.53	-0.971
ORF10 protein	4449.23	7.93	16.06	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).	stable	107.63	0.637

1

2

Table 2. Secondary structure of the SARS-COV-2 proteins predicted via SOPMA

Proteins	Sequence Length	α-helix	β-Turn	Random Coils
Envelope	75	44.00%	9.33%	20.00%
M glycoprotein	222	34.68%	6.76%	37.39%
ORF6 protein	61	70.49%	8.20%	11.48%
ORF7a	121	42.98%	9.92%	28.10%
ORF8 protein	121	19.83%	4.96%	39.67%
Nucleocapsid	419	21.24%	6.92%	55.13%
ORF10	38	28.95%	5.26%	28.95%

Table 3. Structural details of the SARS-COV-2 proteins model

Proteins	Tool utilized for Modeling	Template	Ramachandran plot			Errat Results
			Favored Region	Allowed Region	Disallowed Region	Quality Factor
E	Swiss model	5x29.1	92.9%	3.9%	3.2%	61.22
M	Swiss model	6ck1.1	98.0%	2.0%	0.0%	100
N	Swiss model	1ssk.1	68.2%	18.6%	13.2%	23.9316
Orf8	Swiss model	1xak.1	86.1%	9.7%	4.2%	24.61
Orf7a	Swiss model	1yo4.1	80.2%	16.0%	3.7%	77.4648
Orf6	Swiss model	5vyj.1	97.1%	2.9%	0.0%	100
Orf10	Pep-fold		82.9%	11.4%	5.7%	100

1 **Table 4.** Linear B cell epitopes predicted through ABCPred 2.0 server.

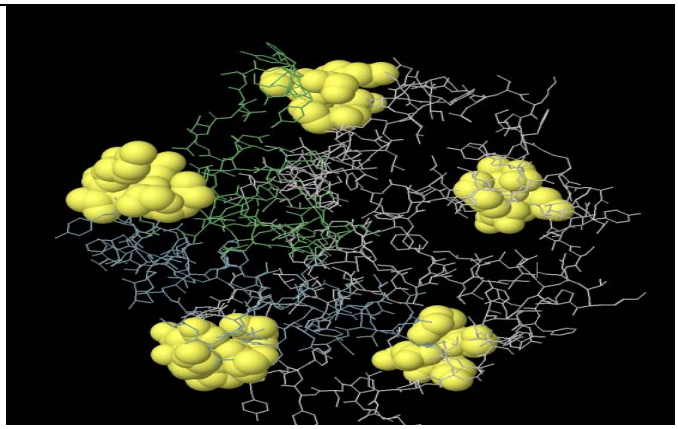
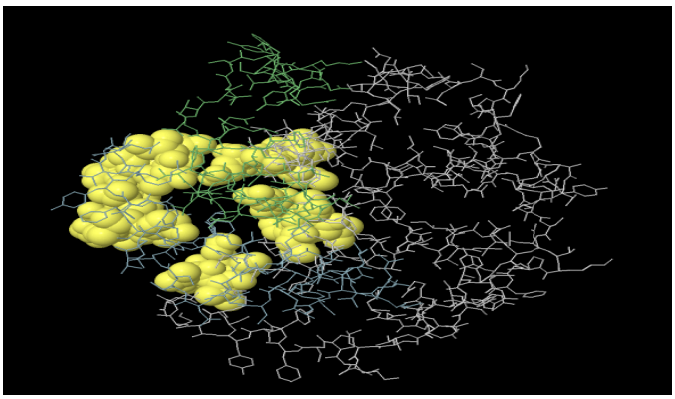
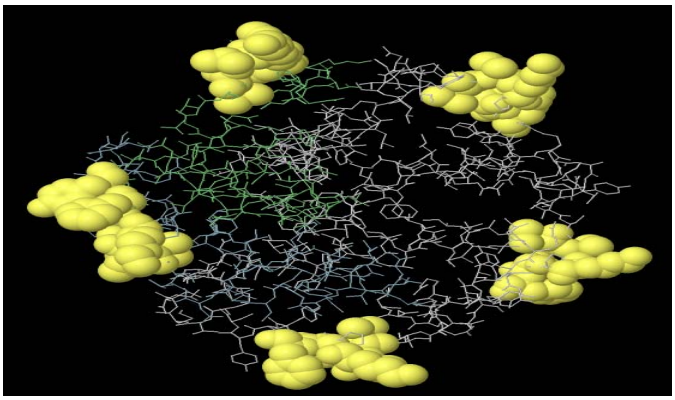
Protein	Peptide (Position)	Antigenicity
E protein	SLVKPSFYVYSRVK (50)	0.6
	LCAYCCNIVNVSLV (39)	1.1
	ILTALRLCAYCCNI (33)	0.7
	FLLVTLAILTALRL (26)	0.8
Membrane protein	SELVIGAVILRGHL (136)	0.54
	TRPLLESELVIGAV (130)	0.58
	DIKDLPKEITVATS (160)	O.67
	PVTLACFVLAAVYR(59)	O.97
	FLTWICLLQFAYAN (28)	0.60
	AAVYRINWITGGIA (68)	0.88
	ATSRTLSYYKLGAS (171)	0.64
	GGIAIAMACLVGLM (78)	0.87
	LEQWNLVIGFLFLT (17)	0.94
	MADSNGTITVEELK(1)	0.62
	VIGFLFLTWICLLQ (23)	0.93
	FRLFARTRSMWSFN (100)	0.71
ORF6	FHLVDFQVTIAEIL(2)	1.2
ORF7a	SSGTYEGNSPFHPL(36)	0.5
	FALTCFSTQFAFAC (54)	1.6
	VYQLRARSVSPKLF (74)	0.6
	PFHPLADNKFALTC (45)	1.2
	FSTQFAFACPDGVK (59)	0.8
	DGVKHVYQLRARSV (69)	0.6
ORF8	EAGSKSPIQYIDIG(64)	1.2
	GNYTVSCLPFTINC (77)	1.6
	PIHFYSKWYIRVGA (38)	1.09
	SKWYIRVGARKSAP (43)	0.75
	FLVFLGIITTVAAF (3)	0.50

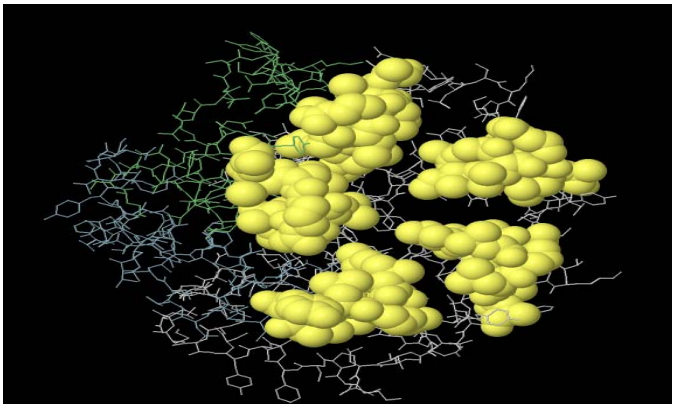
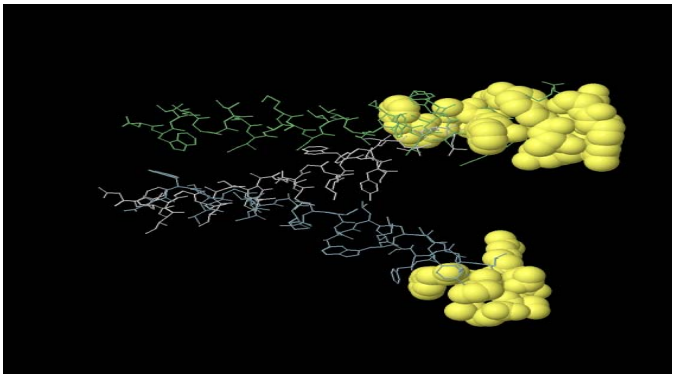
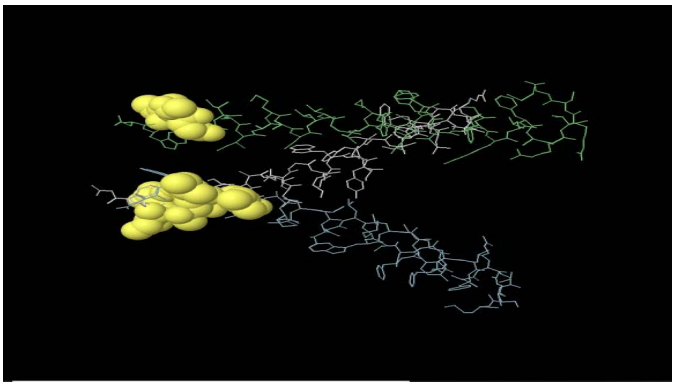
	TQHQPYVVDDPCPI (26)	0.69
	VVDDPCPIHFYSKW (32)	0.50
	LPFTINCQEPKLGS (84)	1.34
	KLGSLVVRCSFYED (94)	0.79
Nucleocapsid	TNSSPDDQIGYYRR (76)	0.68
	GSRGGSQASSRSSS (175)	0.89
	DGKMKDLSPRWYFY (98)	1.13
	PQNQRNAPRITFGG (6)	0.73
	GNFGDQELIRQGTD (284)	0.55
	GQTVTKKSAAEASK (243)	0.53
	FGMSRIGMEVTPSG (315)	1.01
	SQASSRSSSRSRNS (180)	1.07
	KAYNVTQAFGRRGP (266)	0.62
	LLLLDRLNQLESKM (221)	0.55
	LSPRWYFYYLGTGP (104)	1.37
	NGERSGARSKQRRP (29)	0.52
	FTALTQHGKEDLKF (53)	1.24
	KDPNFKDQVILLNK (342)	1.16
	EVTPSGTWLTYTGA (323)	0.69
	GQQQQGQTVTKKSA (238)	0.59
	NSTPGSSRGTSPAR (196)	0.58
	GPEQTQGNFGDQEL (278)	1.19
	TGAIKLDDKDPNFK (334)	1.80
	NQLESKMSGKGQQQ(228)	1.09
	GTDYKHWPQIAQFA (295)	0.65
	SSSRSRNSSRNSTP (186)	0.80
ORF10	CRMNSRNYIAQVDV (19)	0.67

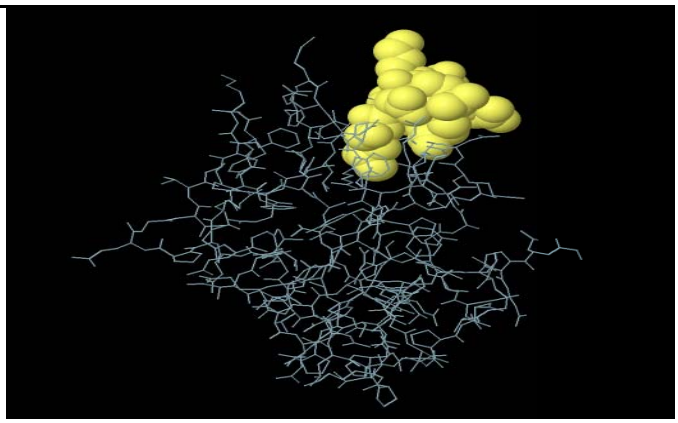
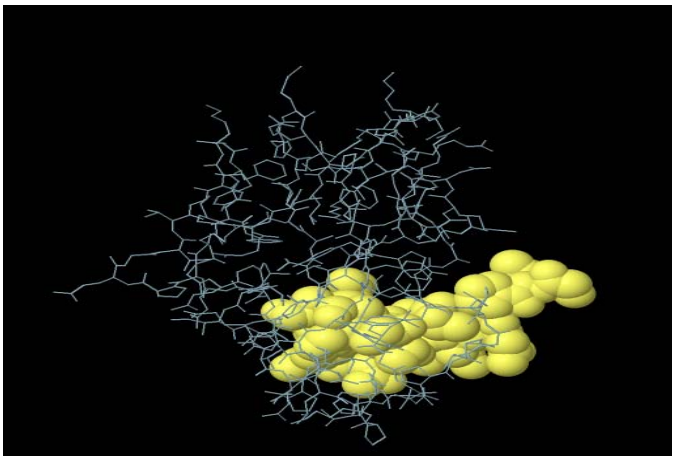
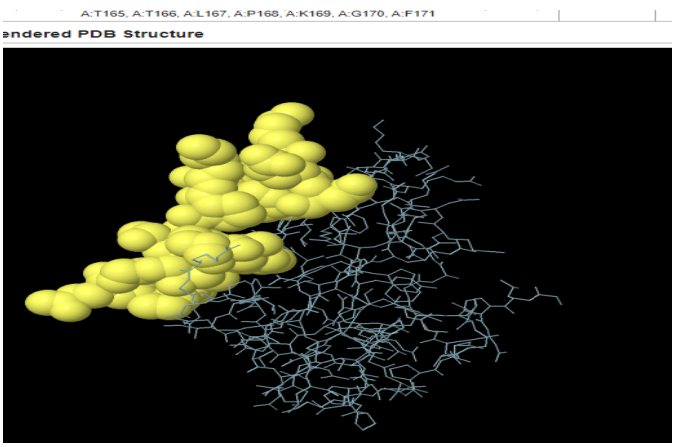
1

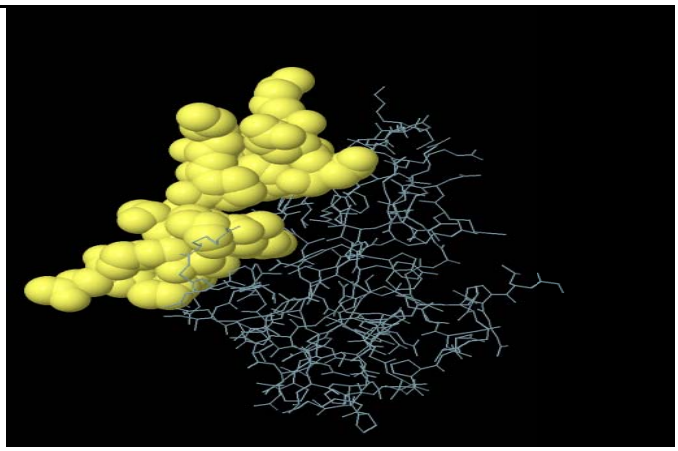
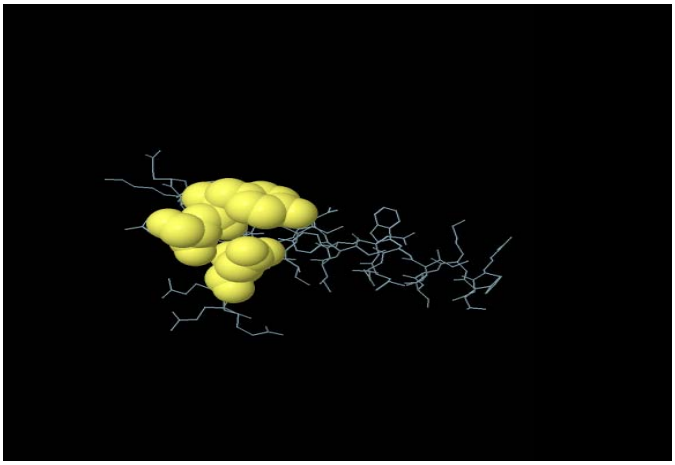
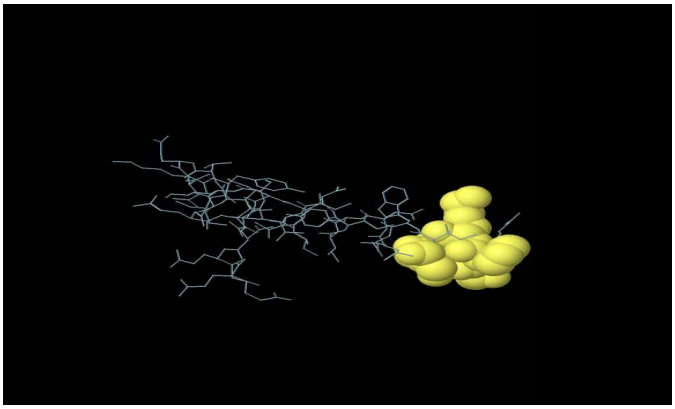
1

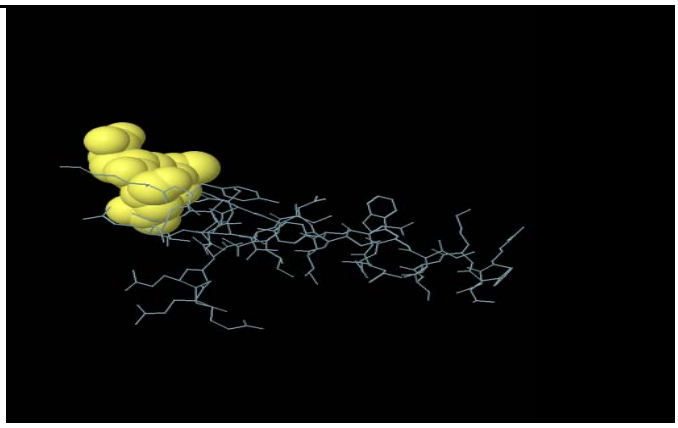
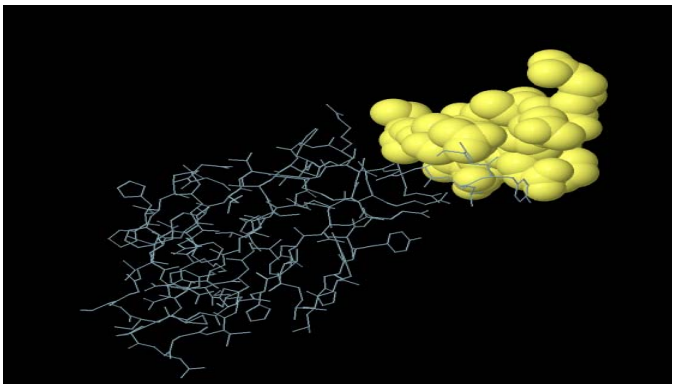
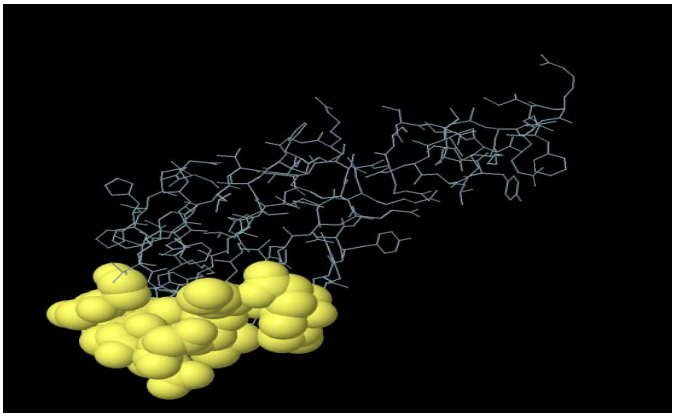
2 **Table 5.** Conformational B cell epitopes of SARS-COV-2 proteins predicted by Ellipro Server

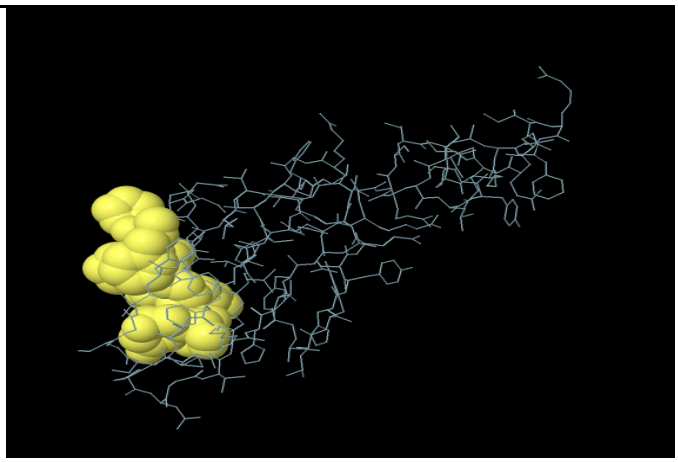
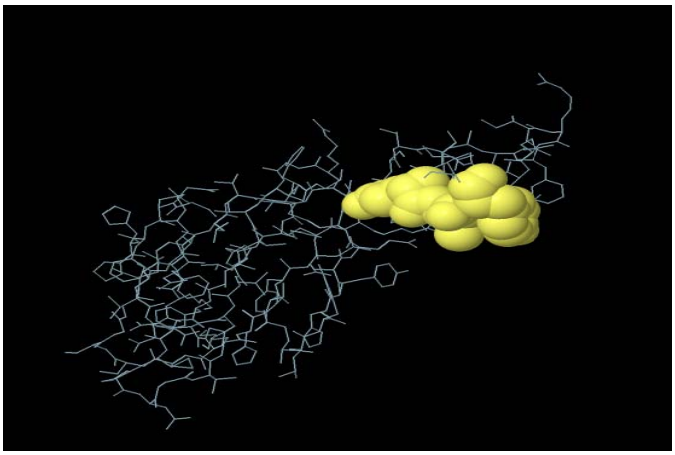
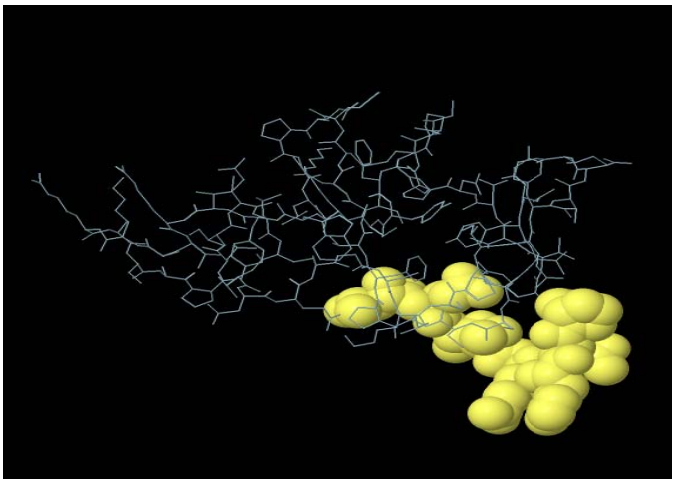
PROTEIN	CONFORMATIONAL B-CELL EPITOPES	3D-STRUCTURE
E-PROTEIN	A:S60, A:R61, A:V62, A:K63, A:N64, A:L65	
	A:E8, A:T9, A:G10, A:T11, A:L12, A:I13, A:V14, A:S16	
	A:L51, A:V52, A:K53, A:S55, A:F56, A:Y59	

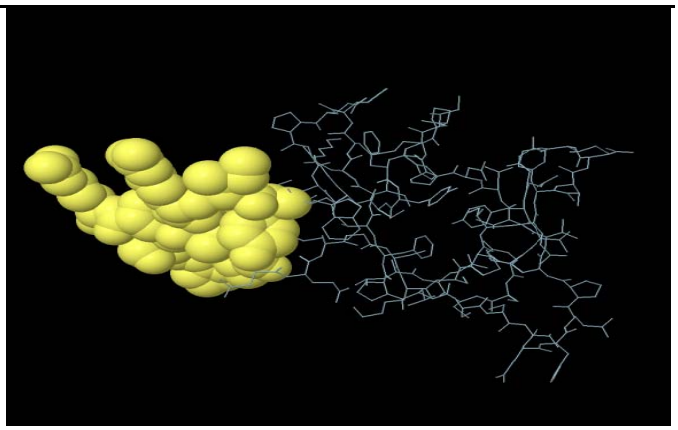
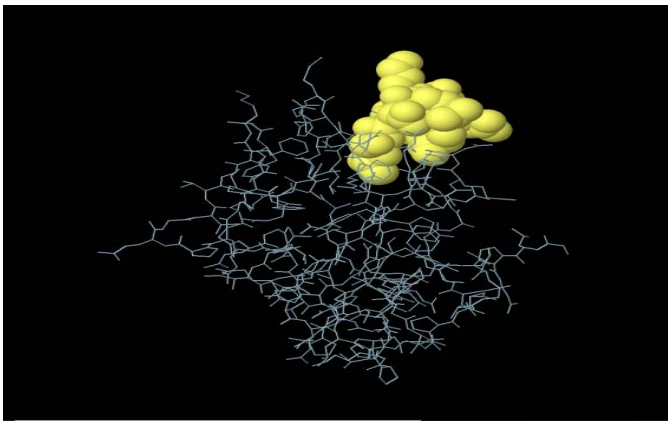
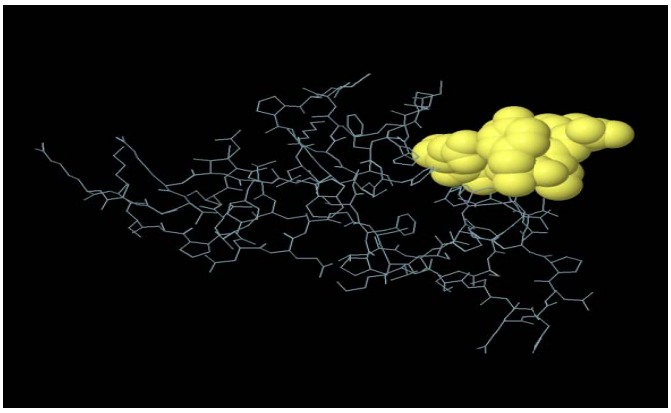
	A:A32, A:I33, A:T35, A:A36, A:L37, A:R38, A:L39, A:C40, A:A41, A:Y42	
M-PROTEIN	A:F100, A:F103, A:A104, A:R105, A:T106, A:R107, A:S108, A:M109	
	A:I76, A:T77, A:G78, A:I80	

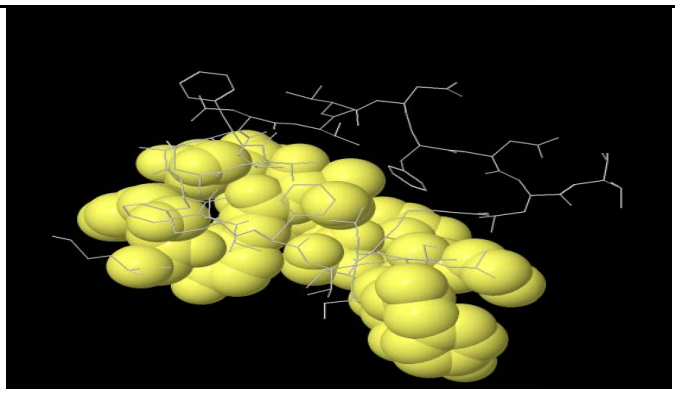
N-Protein	A:R93, A:R95, A:G96, A:G97, A:D98, A:G99, A:K100, A:M101, A:D103, A:L104	
	A:Q58, A:H59, A:G60, A:K61, A:E62, A:D63, A:F66, A:P67, A:L161, A:P162, A:Q163, A:G164, A:T165, A:T166, A:L167, A:P168, A:K169, A:G170, A:F171	
	A:G69, A:Q70, A:T76, A:N77, A:S78, A:S79, A:P80, A:D81, A:D82, A:G114, A:T115, A:G116, A:P117, A:E118, A:A119, A:G120, A:L121, A:P122, A:G124, A:A125, A:N126, A:K127, A:D128, A:G129, A:T135, A:E136, A:G137, A:A138, A:L139, A:N140, A:T141, A:P142, A:K143, A:D144, A:H145, A:I146	

	A:G44, A:L45, A:P46, A:N47, A:N48, A:T49, A:W52, A:G147, A:T148, A:R149, A:N150, A:P151, A:A152, A:N153, A:N154, A:A155, A:A156	
ORF6	A:Y49, A:S50, A:D53	
	A:T21, A:F22, A:K23, A:V24, A:S25	

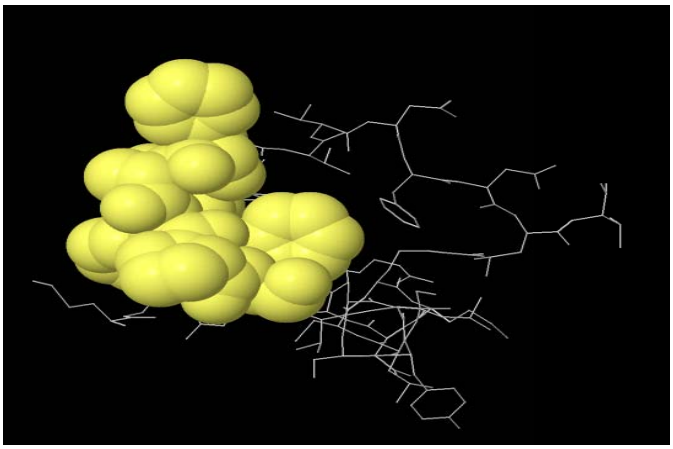
	A:T45, A:E46, A:N47	
ORF7a	A:S81, A:V82, A:S83, A:P84, A:K85, A:L86, A:F87, A:I88, A:R89, A:Q90, A:E91, A:E92, A:V93, A:Q94	
	A:E16, A:L17, A:C35, A:S36, A:S37, A:G38, A:C67, A:P68, A:D69, A:G70, A:V71, A:K72	

	A:E33, A:F46, A:H47, A:P48, A:L49, A:A50, A:D51, A:N52, A:K53	
	A:E95, A:L96, A:Y97	
Orf8	A:L22, A:Q23, A:S24, A:C25, A:T26, A:Q27, A:H28, A:Q29, A:P30	

	A:Y46, A:I47, A:R48, A:V49, A:G50, A:A51, A:R52, A:S54, A:A55, A:P56, A:L57, A:I58, A:V62	
	A:C90, A:E92, A:P93, A:K94	
	A:D34, A:D35, A:Y73, A:I74, A:D75, A:I76, A:G77, A:N78	
ORF10		
	_:G2, _:Y3, _:Y14, _:L18, _:M21, _:N22, _:S23, _:R24, _:N25, _:Y26, _:I27, _:A28, _:Q29	



_:N5, _:F7, _:A8, _:F9, _:P10, _:F11



_:V6, _:V32, _:V33, _:N34, _:F35, _:N36, _:L37, _:T38

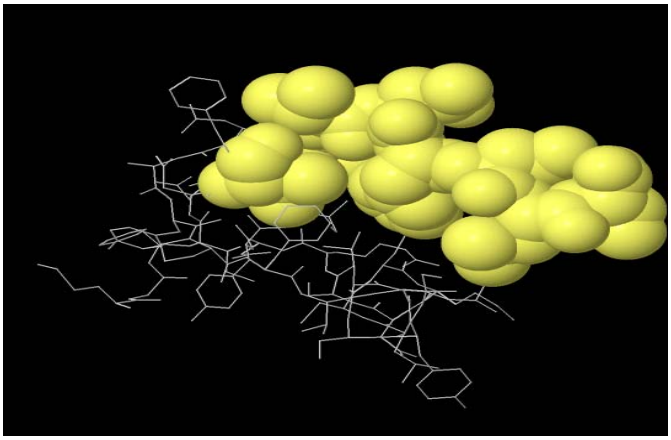


Table 6. MHC Class I Epitopes. The epitopes listed in the table showed 100% conservancy as predicted by IEDB conservancy analysis tool) among the protein sequences included in the present study

Proteins	Peptide	position	allele	antigenicity
Envelope protein	FLLVTLAILTAL	26-37	HLA-A*02:01	0.6
			HLA-A*02:06	
	LLFLAFVVFLLV	18-29	HLA-A*02:01	0.6
			HLA-B*51:01	
	FLAFVVFLLVTL	20-31	HLA-A*02:01	0.7
			HLA-B*51:01	
	NVSLVKPSFYVY	48-59	HLA-A*30:02	0.6
			HLA-A*29:02	
	AILTALRLCAYC	32-43	HLA-A*25:01	0.7
			HLA-A*01:01	
	LFLAFVVFLLVT	19-30	HLA-A*02:01	0.6
			HLA-B*51:01	
	VYSRVKNLNSSR	58-69	HLA-A*31:01	0.5
	VSLVKPSFYVYS	49-60	HLA-A*30:02	0.5
	PSFYVYSRVKNL	54-65	HLA-C*14:02	0.8
			HLA-C*07:01	
			HLA-C*06:02	
Membrane protein	FRLFARTRSMWS	100-111	HLA-C*07:01	
			HLA-B*27:05	
			HLA-B*14:02	
			HLA-A*32:01	
	RLFARTRSMWSF	101-112	HLA-A*32:01	0.5
			HLA-B*08:01	
			HLA-B*46:01	
			HLA-B*27:05	

			HLA-B*15:01	
			HLA-B*57:01	
			HLA-A*24:02	
			HLA-A*23:01	
	LFARTRSMWSFN	102-113	HLA-B*46:01	0.9
			HLA-B*08:01	
			HLA-A*24:02	
	ITVATSRTLSYY	168-179	HLA-A*01:01	0.7
			HLA-A*30:02	
			HLA-A*26:01	
			HLA-A*29:02	
			HLA-B*57:01	
Orf6	FKVSIWNLDYII	22-33	HLA-A*32:01	0.5
	FHLVDFQVTIAE	2-13	HLA-B*38:01	1.5
Orf7a	HPLADNKFALTC	58-12	HLA-B*35:03	1.3
			HLA-B*39:01	
			HLA-B*07:02	
	GTYEGNSPFHPL	38-49	HLA-B*40:01	0.6
			HLA-B*40:02	
			HLA-B*15:02	
	STQFAFACPDGV	61-71	HLA-A*68:02	0.9
Orf8	HQPYVVDDPCPI	28-39	HLA-B*51:01	0.5
	GNYTVSCLPFTI	77-88	HLA-A*24:02	1.7
	DDPCPIHFYSKW	34-45	HLA-B*53:01	0.8
	LPFTINCQEPKL	84-95	HLA-B*35:03	1.1
			HLA-B*51:01	
	DFLEYHDVRVVL	107-118	HLA-B*40:01	0.6
			HLA-B*18:01	
			HLA-C*07:02	
	PIHFYSKWYIRV	38-49	HLA-A*31:01	0.7

	LEYHDVRVVLDF	109-120	HLA-B*18:01	1.0
			HLA-B*44:02	
Nucleocapsid	TATKAYNVTQAF	263-274	HLA-B*46:01	0.5
protein	KMKDLSPRWYFY	100-111	HLA-A*29:02	1.4
			HLA-E*01:01	
			HLA-A*32:01	
			HLA-A*01:01	
			HLA-A*03:01	
			HLA-A*31:01	
			HLA-B*15:01	
	MEVTPSGTWLTY	322-333	HLA-B*35:01	0.7
			HLA-B*53:01	
			HLA-B*44:03	
			HLA-B*18:01	
			HLA-A*01:01	
			HLA-B*44:02	
			HLA-A*26:01	
			HLA-A*29:02	
	DPNFKDQVILLN	343-354	HLA-B*35:03	1.3
Orf10	CRMNSRNYIAQV	19-30	HLA-C*06:02	0.6
			HLA-B*27:05	
	FAFPFTIYSLLL	7-18	HLA-C*12:03	0.7
			HLA-C*03:03	
			HLA-B*46:01	
			HLA-B*51:01	
			HLA-B*53:01	
			HLA-C*06:02	
			HLA-C*07:01	
			HLA-A*68:02	
			HLA-B*38:01	

Table 7. MHC Class II Epitopes. The epitopes listed in the table showed 100% conservancy as predicted by IEDB conservancy analysis tool) among the protein sequences included in the present study

PROTEIN	PEPTIDE (POSITION)	ALLELES	ANTIGENICITY
E PROTEIN	FLLVTLAILTALRL C(26-40)	HLA-DRB1*11:04 HLA-DRB1*11:06 HLA-DRB1*13:11 HLA-DRB1*01:01 HLA-DRB1*01:02 HLA-DPA1*03:01/ DPB1*04:02 HLA-DRB1*07:01 HLA-DRB1*11:28 HLA-DRB1*13:05 HLA-DRB1*13:07	0.6311
	LLFLAFVVFLLVTL A(18-32)	HLA-DPA1*03:01/ DPB1*04:02 HLA-DPA1*01:03/ DPB1*02:01 HLA-DPA1*01/ DPB1*04:01 HLA-DPA1*02:01/ DPB1*01:01 HLA-DRB1*15:02 HLA-DRB1*04:23 HLA-DRB1*04:04 HLA-DRB1*04:08 HLA-DRB1*04:10 HLA-DQA1*05:01/ DQB1*02:01 HLA-DRB1*08:13 HLA-DRB1*07:03 HLA-DRB1*01:02 HLA-DRB1*04:05	0.8122
	VLLFLAFVVFLLVT L(17-31)	HLA-DPA1*03:01/ DPB1*04:02 HLA-DPA1*01:03/ DPB1*02:01 HLA-DPA1*01/ DPB1*04:01	0.6386

		HLA-DPA1*02:01/ DPB1*01:01 HLA-DRB1*15:02 HLA-DQA1*05:01/ DQB1*02:01 HLA-DRB1*15:01 HLA-DRB1*15:06 HLA-DRB1*07:03	
	LFLAFVVFLLVTLA I(19-33)	HLA-DPA1*03:01/ DPB1*04:02 HLA-DPA1*01:03/ DPB1*02:01 HLA-DPA1*01/ DPB1*04:01 HLA-DPA1*02:01/ DPB1*01:01 HLA-DRB1*15:02 HLA-DRB1*04:23 HLA-DRB1*11:04 HLA-DRB1*04:08 HLA-DRB1*13:11 HLA-DRB1*04:08 HLA-DRB1*04:10 HLA-DRB1*11:28 HLA-DRB1*13:05 HLA-DRB1*04:21 HLA-DRB1*08:13 HLA-DRB1*04:26 HLA-DRB1*07:03 HLA-DRB1*01:02 HLA-DRB1*04:04 HLA-DRB1*04:05	0.7471
M PROTEIN	ESELVIGAVILRGH L(135-149)	HLA-DRB1*03:09 HLA-DRB1*11:07 HLA-DRB1*03:06 HLA-DRB1*03:07 HLA-DRB1*03:08 HLA-DQA1*01:02/ DQB1*06:02	**0.5735**
	PVTLACFVLAAVY RI(59-73)	HLA-DRB1*07:03 HLA-DRB1*11:20 HLA-DRB1*01:02 HLA-DRB1*07:01 HLA-DRB1*11:14	**0.8548**

		HLA-DRB1*13:23	
		HLA-DRB1*03:09	
		HLA-DRB1*13:07	
		HLA-DRB1*11:28	
		HLA-DRB1*13:05	
		HLA-DRB1*03:05	
		HLA-DRB1*04:08	
	LEQWNLVIGFLFLTW(17-31)	HLA-DPA1*01:03/ DPB1*02:01	**1.0231**
		HLA-DPA1*01/ DPB1*04:01	
		HLA-DRB5*01:05	
	RNRFLYIIKLIFLWL (42-56)	HLA-DRB1*11:28	
		HLA-DRB1*13:05	
		HLA-DRB1*13:21	
		HLA-DRB1*11:01	
		HLA-DPA1*02:01/ DPB1*01:01	
		HLA-DRB1*08:13	
		HLA-DRB1*08:01	
		HLA-DRB1*07:03	
		HLA-DPA1*01/ DPB1*04:01	
		HLA-DRB1*11:14	
		HLA-DRB1*13:23	
		HLA-DRB1*03:09	
		HLA-DRB4*01:01	
		HLA-DPA1*01:03/ DPB1*02:01	
		HLA-DRB1*08:17	
		HLA-DRB1*13:07	
		HLA-DRB1*15:01	
		HLA-DRB1*15:06	
	SFRLFARTRSMWSFN(99-113)	HLA-DRB1*08:13	**0.7955**
		HLA-DRB1*11:14	
		HLA-DRB1*13:23	
		HLA-DRB1*15:02	
		HLA-DRB1*11:20	
		HLA-DRB1*11:01	
		HLA-DRB1*13:07	
		HLA-DRB1*15:06	
		HLA-DRB1*11:28	
		HLA-DRB1*13:05	
		HLA-DRB1*04:01	
		HLA-DRB1*04:26	
		HLA-DRB1*11:02	

		HLA-DRB1*11:21 HLA-DRB1*13:22 HLA-DRB1*03:05	
Orf6	MFHLVDFQVTIAEI L(1-15)	HLA-DRB1*07:01 HLA-DRB1*07:03 HLA-DQA1*04:01 DQB1*04:02 HLA-DPA1*03:01 DPB1*04:02	1.0366
	FHLVDFQVTIAEIL L(2-16)	HLA-DPA1*03:01/ DPB1*04:02 HLA-DRB1*07:03 HLA-DQA1*04:01/ DQB1*04:02	1.1567
	DFQVTIAEILLIIMR (6-20)	HLA-DRB1*07:01 HLA-DRB1*07:03 HLA-DPA1*03:01/ DPB1*04:02	0.8778
	VDFQVTIAEILLIIM (5-19)	HLA-DRB1*07:01 HLA-DPA1*03:01/ DPB1*04:02 HLA-DRB1*07:03	0.9754
		HLA-DRB1*07:01 HLA-DPA1*02:01/ DPB1*01:01	
Orf7A	DGVKHVYQLRARS VS(69-83)	HLA-DRB1*08:01 HLA-DRB1*08:13 HLA-DRB1*01:01 HLA-DRB1*08:04	0.7457
	HVYQLRARSVSPK LF(73-87)	HLA-DRB1*08:13 HLA-DRB1*08:01 HLA-DRB1*08:04	0.5654
	GVKHVYQLRARSV SP(70-84)	HLA-DRB1*08:13 HLA-DRB1*08:01 HLA-DRB1*01:01 HLA-DRB1*08:04	1.2761
	KHVYQLRARSVSP KL(72-86)	HLA-DRB1*08:13 HLA-DRB1*08:01 HLA-DRB1*01:01 HLA-DRB1*08:04	0.8381
	VKHVYQLRARSVS PK(71-85)	HLA-DRB1*08:13 HLA-DRB1*08:01 HLA-DRB1*01:01 HLA-DRB1*08:04	1.0865
	YQLRARSVSPKLFI R(75-89)	HLA-DRB1*08:13 HLA-DRB1*08:01	0.7805

		HLA-DRB1*08:04	
	AAIVFITLCFTLKR K(105-119)	HLA-DRB1*07:03 HLA-DPA1*02:01 DPB1*05:01	1.8170
Orf8	CTQHQPYVVDDPC PI(25-39)	HLA-DRB3*01:01 HLA-DRB1*03:09 HLA-DRB1*03:05 HLA-DRB1*04:21	0.5165
	HQPYVVDDPCPIHF Y(28-42)	HLA-DRB3*01:01 HLA-DRB1*03:09 HLA-DRB1*03:05 HLA-DRB1*04:21 HLA-DRB1*03:01	0.5587
	QHQPYVVDDPCPI HF(27-41)	HLA-DRB3*01:01 HLA-DRB1*03:09 HLA-DRB1*03:05 HLA-DRB1*04:21 HLA-DRB1*03:01	0.8637
	TQHQPYVVDDPCPI H(26-40)	HLA-DRB3*01:01 HLA-DRB1*03:09 HLA-DRB1*03:05 HLA-DRB1*04:21	0.6706
	PYVVDDPCPIHFYS K(30-44)	HLA-DRB1*03:09 HLA-DRB1*03:05 HLA-DRB3*01:01 HLA-DRB1*04:21	0.6766
	FLGIITTVAAFHQE C(6-20)	HLA-DRB1*04:08 HLA-DRB1*04:23 HLA-DRB1*04:10 HLA-DRB1*04:02 HLA-DRB1*04:26	0.6904
	VVDDPCPIHFYSK W (3-17)	HLA-DRB1*04:08 HLA-DRB1*15:02 HLA-DRB1*04:23 HLA-DRB1*04:10 HLA-DRB1*11:04 HLA-DRB1*11:06 HLA-DRB1*13:11 HLA-DRB1*04:05 HLA-DRB1*01:01 HLA-DRB1*04:02 HLA-DRB1*04:01 HLA-DRB1*04:26	0.5754
	LVFLGIITTVAAFH Q(4-18)	HLA-DRB1*04:08 HLA-DRB1*04:23 HLA-DRB1*04:10	0.6791

		HLA-DRB1*11:04	
		HLA-DRB1*11:06	
		HLA-DRB1*13:11	
		HLA-DRB1*08:02	
		HLA-DRB1*04:05	
		HLA-DRB1*04:02	
		HLA-DRB1*04:04	
		HLA-DRB1*04:01	
		HLA-DRB1*04:26	
	MKFLVFLGIITTVA	HLA-DRB1*04:08	0.5366
	A(1-15)	HLA-DRB1*15:02	
		HLA-DRB1*04:23	
		HLA-DRB1*04:05	
		HLA-DRB1*04:04	
		HLA-DRB1*01:01	
		HLA-DRB1*04:10	
		HLA-DRB1*11:04	
		HLA-DRB1*11:06	
		HLA-DRB1*13:11	
		HLA-DRB1*04:02	
		HLA-DRB1*04:01	
		HLA-DRB1*04:26	
	LGIITTVAAFHQEC	HLA-DRB1*04:23	0.7908
	S(7-21)	HLA-DRB1*04:08	
		HLA-DRB1*04:10	
		HLA-DRB1*04:02	
		HLA-DRB1*04:26	
	VFLGIITTVAAFHQ	HLA-DRB1*04:08	0.6382
	E(5-19)	HLA-DRB1*04:23	
		HLA-DRB1*04:10	
		HLA-DRB1*04:23	
		HLA-DRB1*04:08	
		HLA-DRB1*04:26	
		HLA-DRB1*04:01	
	PKLGSLVVRCSFYE	HLA-DRB1*08:06	0.7902
	D(93-107)	HLA-DRB1*11:02	
		HLA-DRB1*11:21	
		HLA-DRB1*13:22	
		HLA-DRB1*13:01	
		HLA-DRB1*13:27	
		HLA-DRB1*13:28	
		HLA-DRB1*08:01	
		HLA-DRB1*13:04	
NC PROTEIN	FFGMSRIGMEVTPS	HLA-DRB1*11:28	**0.9397**
	G(314-328)	HLA-DRB1*13:05	
		HLA-DRB1*13:21	

	FFGMSRIGMEVTPSG(314-328)	HLA-DRB1*11:28 HLA-DRB1*13:05 HLA-DRB1*13:21	**0.9397**
	KAYNVTQAFGRRGPE(266-280)	HLA-DRB5*01:05 HLA-DRB5*01:01	**0.6104**
	ALLLLDRLNQLESKM(220-234)	HLA-DRB1*11:04 HLA-DRB1*11:06 HLA-DRB1*13:11 HLA-DRB1*13:21 HLA-DRB1*13:07 HLA-DRB1*11:02 HLA-DRB1*11:21 HLA-DRB1*13:22 HLA-DRB1*13:04 HLA-DRB1*08:17 HLA-DPA1*03:01/DPB1*04:02 HLA-DRB1*08:06 HLA-DRB1*11:28 HLA-DRB1*13:05 HLA-DRB1*08:04 HLA-DRB1*11:14 HLA-DRB1*13:23	**0.5669**
ORF10	CRMNSRNYIAQVDV(8-22)	HLA-DRB1*15:02	**0.6757**
	FAFPFTIYSLLLCRM(7-21)	HLA-DRB1*15:02 HLA-DPA1*01/DPB1*04:01 HLA-DPA1*02:01/DPB1*01:01 HLA-DPA1*03:01/DPB1*04:02 HLA-DRB1*08:17 HLA-DPA1*01:03/DPB1*02:01 HLA-DRB1*08:13	**0.5718**
	FPFTIYSLLLCRMNS(9-23)	HLA-DRB1*15:02 HLA-DRB1*08:17 HLA-DPA1*01/DPB1*04:01 HLA-DRB1*08:13	**0.6560**
	PFTIYSLLLCRMNSR(10-24)	HLA-DRB1*15:02 HLA-DRB1*08:17 HLA-DRB1*08:13	**0.9445**

1

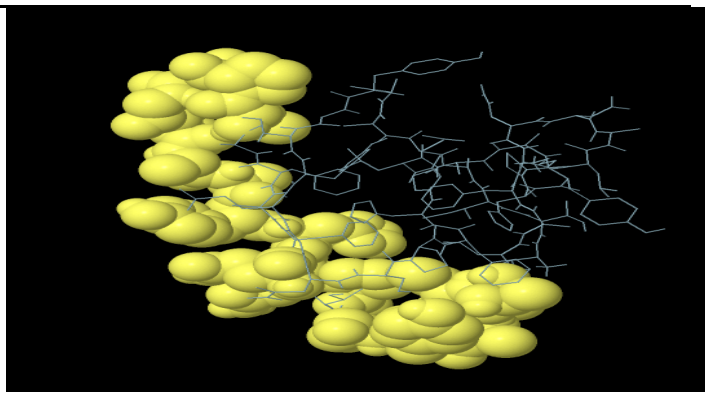
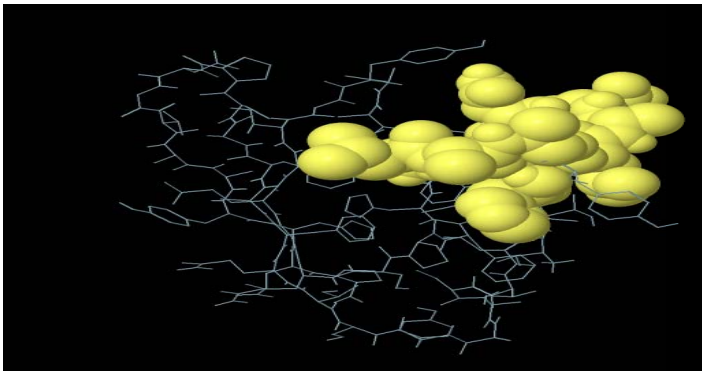
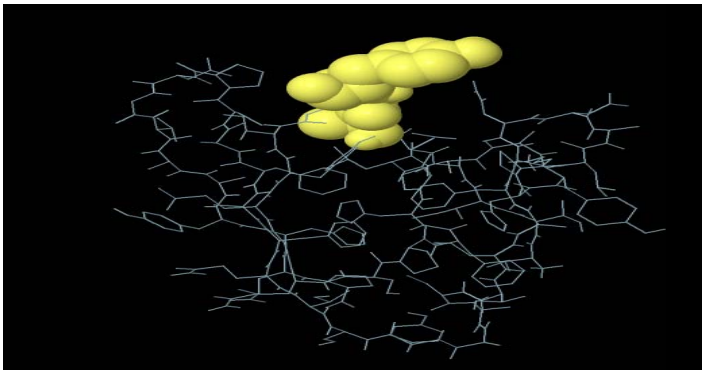
1 **Table 8.** Linear B cell epitopes in the final MEV construct

B cell Epitopes	Position	Antigenicity
LVDFQVTIAEAAYG	113	1.0
GPGPGMFHLVDFQV	368	0.8
PIHFYSKWGPGPGP	419	0.6
AAYFHLVDFQVTIA	108	1.3
LACFVLAAVYRIGP	336	1.0
PGPVTLACFVLAAV	331	1.1
GPGMFHLVDFQVTI	370	1.2
IGKCSTRGRKCCRR	30	1.2
TRGRKCCRRKKEAA	35	1.0
LPKEEQIGKCSTRG	24	0.7
MWSFNAAYFHLVDF	103	1.1
GPGPGESELVIGAV	348	0.5
VGPGPGFLLVTLAI	287	0.6
YIAQVGPGPGFLLV	283	0.5
PYVVDDPCPIAAYD	173	0.5
PLADNKFALTCAAY	142	1.2
GPGPKLGSLVVRCS	429	0.7
ELVIGAVILRGHLG	355	0.8
SKWAAYPIHFYSKW	195	0.8
WSFNAAYFHLVDFQ	104	1.5
WGPGPGPKLGSLVV	426	0.5
VIGAVILRGHLGPG	357	1.0
NLVIGFLFLTWGPG	317	1.1
AYDPNFKDQVILLN	259	1.1
PCPIHFYSKWAAYP	188	0.9
NAAYFHLVDFQVTI	107	1.3
GMFHLVDFQVTIAE	327	1.1
GESELVIGAVILRG	352	0.5
AVYRIGPGPGESEL	343	0.5
AQVGPGPGFLLVTL	285	0.8
DDPCPIHFYSKWGP	415	1.1
GPKLGSLVVRCSFY	431	1.0
FHPLAAYHPLADNK	134	0.6
SELVIGAVILRGHL	354	0.5
GLEQWNLVIGFLFL	312	1.0
YDDPCPIHFYSKWA	185	0.9

FNAAYFHLVDFQVT	106	1.4
LRGHLGPGPGMFHL	363	0.9
SKWGPGPGPKLGSL	424	0.6
GPGVVDDPCPIHFY	410	0.5
HFYSKWAAYPIHFY	192	0.9
DDPCPIAAYDDPCP	177	0.8
LAAYHPLADNKFAL	137	0.5
CPIHFYSKWGPGPG	418	0.6
DDPCPIHFYSKWAA	186	0.8
AYDDPCPIHFYSKW	184	0.8
YGNYTVSCLPFTIA	215	1.6
AYGTYEGNSPFHPL	124	0.7
AEAAYGTYEGNSPF	121	0.6
YSKWGPGPGPKLGS	423	0.5
ESELVIGAVILRGH	353	0.6
FVLAAVYRIGPGPG	339	0.6
PGPGPVTLACFVLA	329	0.9
PKEEQIGKCSTRGR	25	1.3
PGPGPKLGSLVVRC	428	0.7
PIHFYSKWAAYPIH	190	1.0
DPCPIAAYDDPCPI	178	1.0
VDDPCPIAAYDDPC	176	0.5
YHPLADNKFALTCA	140	1.1
AAYHPLADNKFALT	138	0.6
VDDPCPIHFYSKWG	414	0.8
LSPRWYFYAAYDPN	250	1.1
YTVSCLPFTIAAYL	218	1.3
FAFACPDGVAAYHQ	159	0.7
LAAVYRIGPGPGES	341	05
IAAYDDPCPIHFYS	182	0.6
RGHLGPGPGMFHLV	364	1.1
AAYDPNFKDQVILL	258	1.2
NFKDQVILLNAAYC	263	0.8
QPYVVDDPCPIAAY	172	0.5
VTLACFVLAAVYRI	334	0.9
EQWNLVIGFLFLTW	314	1.2

LEQWNLVIGFLFLT	313	0.9
AYLPFTINCQEPKL	229	1.0
AAYLPFTINCQEPK	228	1.0
PCPIHFYSKWGPGP	417	0.9
GPGESELVIGAVIL	350	0.5
ACFVLAAVYRIGPG	337	0.9
KWGPGPGPKLGSLV	425	0.6
LPFTIAAYLPFTIN	223	1.2
AVLSCLPKEEQIGK	19	0.5
GVVDDPCPIHFYSK	412	0.6
TIAAYLPFTINCQE	226	1.1
DPCPIHFYSKWAAY	187	0.9
EAAYGTYEGNSPFH	122	0.6
PGPKLGSLVVRCSF	430	0.8
IGPGPGESELVIGA	347	0.9
LVIGFLFLTWGPGP	318	1.0
VVDDPCPIHFYSKW	413	0.5
IGAVILRGHLGPGP	358	1.2
RWYFYAAYDPNFKD	253	1.2
VSCLPFTIAAYLPF	220	1.3

1 **Table 9.** Conformational B cell epitopes in the final MEV construct

Conformational B cell Epitopes	3D Structure
A:T151, A:C152, A:A153, A:A154, A:Y155, A:S156, A:Q158, A:C163, A:P164, A:D165, A:G166, A:V167, A:A168, A:A169, A:Y170, A:H171, A:Q172, A:P173	
A:L137, A:A138, A:A139, A:H141, A:P142, A:L143, A:A144, A:D145, A:N146	
A:A123, A:A124, A:Y125	

2

3

4