

## Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study

Maryam Enayatkhani, Mehdi Hasaniazad, Sobhan Faezi, Hamed Guklani, Parivash Davoodian, Nahid Ahmadi, Mohammad Ali Einakian, Afsaneh Karmostaji & Khadijeh Ahmadi

To cite this article: Maryam Enayatkhani, Mehdi Hasaniazad, Sobhan Faezi, Hamed Guklani, Parivash Davoodian, Nahid Ahmadi, Mohammad Ali Einakian, Afsaneh Karmostaji & Khadijeh Ahmadi (2020): Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study, Journal of Biomolecular Structure and Dynamics, DOI: [10.1080/07391102.2020.1756411](https://doi.org/10.1080/07391102.2020.1756411)

To link to this article: <https://doi.org/10.1080/07391102.2020.1756411>



Accepted author version posted online: 15 Apr 2020.



Submit your article to this journal [↗](#)



Article views: 2151



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)

# Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an *in silico* study

Maryam Enayatkhani<sup>a</sup>, Mehdi Hasaniazad<sup>a</sup>, Sobhan Faezi<sup>b</sup>, Hamed Guklani<sup>a</sup>, Parivash Davoodian<sup>a</sup>, Nahid Ahmadi<sup>c</sup>, Mohammad Ali Einakian<sup>d</sup>, Afsaneh Karmostaji<sup>a</sup>, Khadijeh Ahmadi<sup>a,\*</sup>

<sup>a</sup>Infectious and Tropical Diseases Research Center, Hormozgan Health Institute, Hormozgan University of Medical Sciences, Bandar Abbas, Iran

<sup>b</sup>Medical Biotechnology Research Center, School of Paramedicine, Guilan University of Medical Sciences, Rasht, Iran

<sup>c</sup>Department of pharmaceutical chemistry, School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, Iran

<sup>d</sup>Food Health Research Center, Hormozgan University of Medical Sciences, Bandar Abbas, Iran

\*Corresponding author: Dr. Khadijeh Ahmadi, Infectious and Tropical Diseases Research Center, Hormozgan Health Institute, Hormozgan University of Medical Sciences, Bandar Abbas, Iran. Email: khadijeh\_6482@yahoo.com

## Abstract

At present, novel *Coronavirus* (2019-nCoV, the causative agent of COVID-19) has caused worldwide social and economic disruption. The disturbing statistics of this infection promoted us to develop an effective vaccine candidate against the COVID-19. In this study, bioinformatics approaches were employed to design and introduce a novel multi-epitope vaccine against 2019-nCoV that can potentially trigger both CD<sup>4+</sup> and CD<sup>8+</sup> T-cell immune responses and investigated its biological activities by computational tools. Three known antigenic proteins (Nucleocapsid, ORF3a, and Membrane protein, hereafter called NOM) from the virus were selected and analyzed for prediction of the potential immunogenic B and T-cell epitopes and then validated using bioinformatics tools. Based on *in silico* analysis, we

have constructed a multi-epitope vaccine candidate (NOM) with five rich-epitopes domain including highly scored T and B-cell epitopes. After predicting and evaluating of the third structure of the protein candidate, the best 3D predicted model was applied for docking studies with Toll-like receptor 4 (TLR4) and HLA-A\*11:01. In the next step, molecular dynamics (MD) simulation was used to evaluate the stability of the designed fusion protein with TLR4 and HLA-A\*11:01 receptors. MD studies demonstrated that the NOM-TLR4 and NOM-HLA-A\*11:01 docked models were stable during simulation time. *In silico* evaluation showed that the designed chimeric protein could simultaneously elicit humoral and cell-mediated immune responses.

**Keywords:** COVID-19; *Coronavirus*; Vaccine; Epitope; Immunoinformatics

List of abbreviations:

TLR4	Toll-like receptor4
MD	Molecular dynamics
WHO	World Health Organization
SARS	Severe Acute Respiratory Syndrome
MERS	Middle East Respiratory Syndrome
ORF	Open reading frame
MHC	Major histocompatibility complex
CTL	Cytotoxic T-lymphocytes
HTL	Helper T-lymphocytes
RMSD	Root mean square deviation
ns	Nanoseconds

## 1. Introduction

*Coronavirus* disease COVID-19 outbreak began in late December 2019 in Wuhan, the capital of Hubei Province, China (Wang et al., 2020). Scientists from all over the world are attempting to investigate this novel virus, known as 2019-nCoV, which is highly contagious, and to discover effective interventions to control and prevent the disease (Heymann, 2020; Huang et al., 2020). *Coronaviruses* are positive-sense single-stranded RNA viruses (ssRNA<sup>+</sup>) belonging to the *Coronaviridae* family. Human *Coronaviruses* HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1 are observed in almost one-third of the common cold (Lim et al., 2016). However, recently some cases of human *coronavirus* infections have led to fatal endemics, including SARS (Severe Acute Respiratory Syndrome), MERS (Middle East Respiratory Syndrome) and COVID-19 that are common diseases between humans and animals whose belong to the genus Betacoronavirus of the *Coronaviridae* family (Al-Tawfiq

et al., 2014). So far, the novel COVID-19 has caused more than 700,000 illnesses and more than 33,000 deaths worldwide (Organization, 2020). The genome size of this virus is about 30 kb and encodes structural and non-structural proteins like other *coronaviruses*. Structural proteins include S protein (Spike), E protein (Envelope), M protein (Membrane), and N protein (Nucleocapsid) (Ahmed et al., 2020). The increasing rate of COVID-19 disease and the high morbidity necessitate the development of a specific and safe vaccine candidate as soon as possible. There is very little known actually about the pathogenesis of the virus; therefore, an immunoinformatics-based approach to investigate the immunogenic epitopes and vaccine design using data from proteins sequencing of the COVID-19 is required.

N protein is the only structural protein that associates with replicase-transcriptase complexes and binds with genomic RNA in *coronaviruses* (Cong et al., 2017). This protein is multifunctional and one of the most crucial structural components of *coronaviruses*. N protein is a structural and antigenic protein that is involved in packaging, transcription, and replication *coronaviruses* (4). This data showed that N protein is a suitable candidate for targeting drug and vaccine design because this protein is conserved, antigenic and multifunctional (6). Leung and *et al.* concluded that N protein can be a suitable vaccine candidate against SARS-Cov because induce strong antibody and this process may trigger cytokine production (Leung et al., 2004). *Coronaviruses* M protein also has a key role in the assembly of virions. The SARS-CoV M protein can interact with N protein and make a network of interactions with the genomic RNA (He et al., 2004). Ong and *et al.* the COVID-19 antigens such as S, N and M proteins introduced as a vaccine candidate (6). This protein has also been studied as an epitope vaccine candidate against SARS-CoV (7).

Open reading frame 3a (ORF3a) is required for viral replication and virulence of SARS CoV. Severe induction of proinflammatory cytokine is a sign of SARS-CoV and MERS-CoV infections. ORF3a activates both pro-IL-1 $\beta$  gene expression and IL-1 $\beta$  secretion and leads to severe lung injury. (Siu et al., 2019). Also, ORF3a has an important role in SARS-CoV assembly or budding with the participation of M and E proteins (McBride and Fielding, 2012). These proteins are not only involved in the pathogenesis of the COVID-19 virus but also have high antigenicity (Chan et al., 2020; Siu et al., 2019; Xu et al., 2020). In this study E, M, N, ORF10, ORF8, ORF3a and M proteins were evaluated by available bioinformatics tools for designing an efficient multi-epitope vaccine for the stimulation of immune responses against COVID-19 infection.

Since the COVID-19 has been recently discovered, little immunological information is available. Preliminary studies based on phylogenetic analyses of the COVID-19 whole

genome have suggested that this virus is very similar to the SARS-CoV (79.7% Identify)(9, 14). Given the apparent similarity between the two viruses, it could be concluded that previous studies on the protective immune responses against SARS-CoV may be useful for developing a vaccine for COVID-19. Previous studies have suggested that both humoral and cellular immunity play important roles in protective responses against this virus (Deming et al., 2007; Yang et al., 2004). Studies revealed that the formation of antibodies against the N protein of SARS-CoV, an immunogenic protein that is highly expressed during infection, is relatively common in patients infected with this virus (Liu et al., 2004; Ying et al., 2003). Although these antibodies are effective, they have a shorter lifespan in recovering the patients. In addition to the specific humoral immunity, it has been shown that the CD<sup>4+</sup> and CD<sup>8+</sup> responses provide long-lasting protection against COVID-19. These studies showed that besides antibody-mediated immune response, cellular immunity is critical to induce protectivity in these infections (Liu et al., 2017). The concept of a multi-epitope vaccine is to efficiently identify and assemble B and T-cell epitopes that are more capable of stimulating the immune system and therefore can induce more potent and effective both arms of immune responses. Peptides and epitopes have shown to be desirable candidates for vaccine development due to their relatively easy production, chemical stability, and lack of infectious potential (Patronov and Doytchinova, 2013). The experimental design and production of multi-epitope vaccines have improved dramatically in recent years. These vaccines are mainly made up of B-cell, CD<sup>8+</sup> cytolytic T-cell (CTLs) and CD<sup>4+</sup> helper T-cells (HTLs) epitopes (Chiarella et al., 2009). Since the antigenic epitopes of a protein could be predicted and detected, therefore the whole protein is not suitable to stimulate an immune response (Testa and Philip, 2012; Zheng et al., 2017). During the development of a vaccine candidate against COVID-19, complex pathogenic mechanisms and numerous pathogenic factors should be considered in vaccine formulation.

In the present study, we aimed to design a novel multi-epitope fusion protein (Nucleocapsid, ORF3a, and Membrane protein or NOM) containing more efficient antigenic epitopes-rich domains. The biological activity of the engineered fusion protein was assessed by bioinformatics tools using the interaction between the vaccine candidate and the innate immune system receptor (TLR4) and cellular immune system receptor (HLA-A\*11:01). We strongly believe that the outcome of the present report will provide a potential vaccine candidate against 2019-nCoV.

## **2. Materials and methods**

In this study, we designed a suitable vaccine candidate against COVID-19, by exploiting the programs of reverse vaccinology (Fig. 1)

### **2.1. Retrieval of protein sequences**

At first, the amino acid sequences of proteins were retrieved from the National Centre for Biotechnology Information (NCBI) at [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov) in FASTA format and performed for subsequent analysis (Table 1).

### **2.2. Selection of antigenic proteins**

We selected six proteins of COVID-19 virus (Table 1) that have an essential role in virulence and replication of the virus, and previous studies have highlighted the necessity of these proteins in coronaviruses function. After the antigenic analysis of these proteins, three proteins of N, ORF3a, and M were selected for final analysis.

### **2.3. Prediction of T-cell (HLA class I and II) epitopes**

1d sequence-based screening server RANKPEP was used to identify T-cell epitopes (Reche and Reinherz, 2007). This server predicts peptide binders to MHC molecules from protein sequences using the position-specific scoring matrix (PSSM). We have selected all HLA class I alleles from the selection panel of RANKPEP server for prediction of epitopes of HLA class I. To prediction of epitopes of HLA class II, we considered DRB1\*0101, DRB1\*0301, DRB1\*0401, DRB1\*0701, DRB1\*0801, DRB1\*1101, DRB1\*1301, and DRB1\*1501 that cover HLA variability of over 95% of the human population worldwide (Kruiswijk et al., 2020).

### **2.4. B-cell epitopes (linear) identification**

For the prediction of B-cell epitopes, the amino acid sequence was analyzed using BepiPred and Kolaskar & Tongaonkar Antigenicity (<http://www.iedb.org/>) servers (Ponomarenko et al., 2010). BepiPred for linear epitope prediction uses both hidden Markov model and amino acid propensity scales methods. Kolaskar and Tongaonkar evaluate the protein for B cell epitopes using the physicochemical properties of the amino acids and their frequencies of occurrence in recognized B cell epitopes (Kolaskar and Tongaonkar, 1990; Mirza et al., 2016).

### **2.5. Selection of epitope-rich domains and the final sequence**

According to the prediction results of the servers used, B cell epitopes and HLA class I and II epitopes that have had high scores were extracted and combined to generate multi-epitope protein. B cell and T cell epitope-rich domains of N, ORF3a and M proteins were selected and joined to each other with an AAA linker.

## **2.6. Antigenicity and allergenicity evaluation**

Antigenicity of designed recombinant protein predicted using the VaxiJen v2.0 server. The VaxiJen classified antigens based on auto cross-covariance (ACC) transformation of protein sequences into uniform vectors of principal amino acid properties which is a novel alignment-independent method and overcome the limitations of alignment-dependent sequence methods (Doytchinova and Flower, 2007). The prediction of vaccine candidate allergenicity is essential. The allergenicity of the designed protein was computed by AllerTOP (<http://www.ddg-pharmfac.net/AllerTOP/>). AllerTOP method predicts recombinant protein allergenicity on auto cross-covariance ACC that describe residue hydrophobicity, size, abundance, helix- and  $\beta$ -strand forming propensities (Dimitrov et al., 2013). AllerTOP v.2 has the highest accuracy (88.7 %) compared to several servers for allergen prediction (Dimitrov et al., 2013).

## **2.7. The physicochemical parameters**

The analyzed parameters consisted of the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity that were evaluated by ProtParam online server (<http://us.expasy.org/tools/protparam.html>) (Gasteiger et al., 2005).

## **2.8. Secondary and tertiary structure prediction**

GOR was used for the designed protein secondary structure prediction ([https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_gor4.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html)) (Kloczkowski et al., 2002). GOR4 predict protein secondary structure using information theory.

The tertiary structure was built using the Galaxy web. GalaxyWEB server (<http://galaxy.seoklab.org/tbm>) is based on the TBM method. This server detects similar proteins and alignment with the target sequence, then make the model and finally perform model refinement (Shin et al., 2014).

## **2.9. Tertiary structure refinement and validation**

The best-modeled structure was refined using the Galaxy Refine server at (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>). Galaxy refined the model by molecular dynamics simulation. This method showed one of the best performances in improving protein structure quality.

Analysis of the final 3D model was made using Molprobit, ProSA and Ramachandran plot. Ramachandran plot obtained from RAMPAGE calculates torsional angles residue-by-residue in protein and indicates that residues are in allowed, favored or outlier regions (Oberholser, 2010). ProSA web used to recognize the errors in the generated 3D models using atomic coordinates of the model. ProSA web was created Z-score (overall model quality) and a plot of residue energies of proteins (Wiederstein and Sippl, 2007). Clash analysis is very important for the validation of proteins. MolProbit is a structure-validation web service that calculates the clash score, Protein-geometry score, Poor rotamers, Ramachandran plot, and Molprobit score.

#### **2.10. Defining discontinuous B-cell epitopes of designed protein**

The interaction between antigen epitopes and antibodies is very essential to eliminate the infection. Conformational epitopes are the most important and the most prevalent epitopes that are recognized by antibodies. All conformational epitope prediction methods require the 3D structures of proteins. Discontinuous epitopes of recombinant protein predicted using the ElliPro server. This web-based predicts discontinuous epitopes based on protein-antibody interactions. ElliPro server predicts conformational and linear B cell epitopes using Thornton's method and by MODELLER program or BLAST search of PDB predict and visualize of antibody epitopes (Ponomarenko et al., 2008).

#### **2.11. Obtaining and preparing the structures of immune receptors and NOM recombinant protein**

The crystallographic structures of TLR4 (PDB ID: 2Z62 (Kim et al., 2007)) and HLA-A\*11:01 (PDB ID: 5WJL (Culshaw et al., 2017)) were obtained from the PDB database (Berman et al., 2000). The structures were cleaned and crystallographic waters and co-crystallized molecules were deleted and only the monomer forms of each receptor were kept. Then, the structures of the two receptors and the NOM protein were prepared by the Dock prep tool in UCSF Chimera software (Pettersen et al., 2004) where hydrogen atoms and charge were added to the structures to make them ready for the next step. In the next step, to stabilize and relax the structures in an aqueous physiological environment, each of the



monomer form of the receptors and the NOM protein were simulated for 100 ns. The details of the MD simulation will be explained later.

### **2.12. Protein-Protein molecular docking and refinement**

After the MD simulation of each receptor and the NOM recombinant protein, the last frame of the trajectory of each simulation was extracted and used as the input structures for protein-protein molecular docking. The PatchDock webserver (Schneidman-Duhovny et al., 2005) with default parameters was used to predict the best positions and orientations where the two proteins can have the highest number of favorable interactions. Afterward, the top 10 solutions were subjected to Firedock webserver (Andrusier et al., 2007; Mashiach et al., 2008) to refine the interaction of protein-protein complexes resulted from molecular docking. The top two complexes from Coronavirus NOM protein-TLR4 and NOM recombinant protein-HLA-A\*11:01 complexes were chosen based on global energies. These four complexes were then taken to the next step, MD simulation.

### **2.13. Molecular dynamics simulation**

After performing the protein-protein molecular docking and finding the best orientations of the vaccine candidate and the receptor proteins to interact with one another, the two best-scored complexes were subjected to MD simulation. All of the MD simulations were done by GROMACS 2018 package (Abraham et al., 2015) and OPLS-AA force field (Jorgensen et al., 1996). The monomer form of each protein and also the docked complexes were placed in the center of a triclinic box with a distance of 1 nm from all edges and solvated with TIP3P water model (Jorgensen et al., 1983). Then, sodium and chloride ions were added to produce a neutral physiological salt concentration of 150 mM. Each system was energy minimized, using the steepest descent algorithm, until the  $F_{max}$  was found to be smaller than  $10 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-1}$ . All of the covalent bonds were constrained using the Linear Constraint Solver (LINCS) algorithm (Hess et al., 1997) to maintain constant bond lengths. The long-range electrostatic interactions were treated using the Particle Mesh Ewald (PME) method (Darden et al., 1993; Essmann et al., 1995) and the cut off radii for Coulomb and Van der Waals short-range interactions were set to 0.9 nm. Then 100 ps NVT (constant number of particles (N), volume (V), and temperature (T)) and 300 ps NPT (constant number of particles (N), pressure (P), and temperature (T)) equilibrations were performed for each system. After the necessary equilibrations, 100 ns of production run were performed for each of the four complexes. Finally, simulations were carried out under the periodic boundary conditions (PBC), set at XYZ coordinates to ensure that the atoms had stayed inside the simulation box, and the

subsequent analyses were then performed using GROMACS utilities, VMD (Humphrey et al., 1996) and USCF Chimera, and also the plots and the figures were created using Daniel's XL Toolbox (v 7.3.2) add-in (Kraus, 2014) and EzMol (Reynolds et al., 2018). All of the computations were performed on an Ubuntu desktop PC, with a [Intel(R) Xeon(R) CPU E5-2630 v3 + NVIDIA GeForce GTX 1080] configuration.

#### **2.14. MMPBSA binding free energy**

The binding free energies of the protein-protein complexes were calculated by Molecular Mechanics-Poisson-Boltzmann Solvent-Accessible surface area, MMPBSA method (Kollman et al., 2000) using g\_mmpbsa package (Kumari et al., 2014). In this method, the binding free energy is the result of the free energy of the complex minus the sum of the free energies of the ligand and the protein. In our case, the NOM recombinant protein is defined as the ligand and the immune receptors are defined as the receptors. The MMPBSA calculation was done for every ns of each simulation trajectories.

$$\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - (\Delta G_{\text{ligand}} + \Delta G_{\text{Receptor}})$$

### **3. Results and discussion**

#### **3.1. Defining T-cell epitopes**

The results of several studies have shown that strong virus-specific T-cell response is required for the elimination of respiratory virus infections such as SARS-CoV and influenza A and para-influenza. These studies conclude that future vaccine interventions should also consider strategies to enhance T cell response to provide robust long-term memory (Channappanavar et al., 2014; Janice Oh et al., 2012). Studies have shown that high levels of T cell responses against N protein were found 2 years after the patient's recovery (Peng et al., 2006). Antibodies are essential to combat SARS-CoV infection, and the body needs SARS-CoV specific CD4<sup>+</sup> T helper cells to produce these specific antibodies. Also, CD8<sup>+</sup> cytotoxic T cells are important for recognizing and killing infected cells, especially in the lungs of infected individuals. We used the Rankpep server which covers almost all HLA supertypes to predicted different epitops from N, ORF3a and M proteins sequence according to HLA I and HLA II alleles. Antigenic epitopes with high binding affinity score were predicted that are summarized in table 2a-b.

#### **3.2. Defining linear B-cell epitopes**

Successful vaccination against viruses such as measles and rubella reflects the importance of protective antibodies. Protection against virus infection such as *Coronaviruses* depends on the simulation of neutralizing antibodies in addition to the T cell-mediated immunity. While cytotoxic lymphocytes can kill infected cells, antibodies have the potential to eliminate infected cells and prevent the infectious virus from infecting a cell (neutralization). SARS-CoV-specific Neutralizing antibodies block viral entry (Dörner and Radbruch, 2007; Hsueh et al., 2004). In this study, for the linear epitope prediction, the Bepipred server was employed. Bepipred analysis revealed several continuous predicted epitopes of N, ORF3a and M proteins. For cross-checking the predicted epitopes, the sequence of all three proteins was also predicted by Kolaskar & Tongaonkar Antigenicity. The linear B-cell epitopes are represented in Table 3. Given that both cellular and humoral immune responses are essential against coronaviruses infection (Janice Oh et al., 2012), finally, epitopes that were shared between B cell and T-cell were selected.

### **3.3. Selected targets for designing the final vaccine**

The main strategy in the present study was to design and construct a novel multi-epitope protein from COVID-19 based on *in silico* methods to elicit humoral and cellular immune responses. Due to the low immunogenicity of the epitope, we chose epitope-rich domains to generate a more diverse and robust response (Wieser et al., 2010). Based on *in silico* analysis, five epitope-rich domains including highly scored and shared epitopes between T and B-cell epitopes were selected and joined to each other with a three AAA linker (Table 4). The schematic diagram of designed vaccine domains with linker's sites has been shown in Figure 2.

### **3.4. Vaccine features**

#### **3.4.1. Assessment of antigenicity and allergenicity**

Prediction of the vaccine candidate antigenicity represents a numerical criterion for the capability of the vaccine to bind to the B- and T-cell receptors and increase the immune response in the host cell. VaxiJen v2.0 was used to predict the antigenicity of the designed protein. VaxiJen analysis of NOM protein showed potent antigenicity 0.5999 with a 0.4% threshold. The results indicate that the designed protein sequences without adjuvant are antigenic.

Allergen proteins induce an IgE antibody response (Dimitrov et al., 2013). The designed vaccine candidate must not show an allergic reaction to the body. The allergenicity of the sequence was predicted using the Allertop tool. Based on prediction approaches in Allertop, this protein was not recognized as an allergen.

### **3.4.2. The physicochemical parameters and protein secondary structure prediction**

Primary structure analysis NOM designed protein (337 aa) has a molecular weight of 37513.57 D. The total number of positively charged residues (Arg+Lys) and negatively charged residues (Asp+Glu) were 19 and 32 respectively. The theoretical isoelectric point (PI) was calculated at 9.62. The instability index (< 40) indicates that the designed protein has high stability for the initiation of an immunogenic reaction. The instability index (II) was 35.75 which classifies the protein as stable. The aliphatic index of NOM recombinant protein was calculated 97.92 and indicates this protein has stability in several temperatures.

Natively unfolded protein regions and alpha-helix are important forms of structural antigens that can be arranged in their native structure and thus identified by antibodies that are produced in response to infection. (Shey et al., 2019). The composition of the predicted secondary structure of the NOM multi-epitope vaccine candidate was 43.92% (alpha helix), 16.02% (extended strand), and 40.06% (random coil). All this information indicates the designed protein is suitable as a vaccine candidate.

### **3.4.3. Tertiary structure prediction and validation**

Three-dimensional structure was modeled by GalaxyWEB for our designed protein (Fig. 3a). This model was used for evaluation and refinement. For validation of the model, ProSA-web, Ramachandran plot and Molprobability were used that compare and analyze the protein structure. The ProSA Web determined Z-score for the best predicted 3D model (Fig.3b). The Z-score of the NOM predicted model was -4.42, which is within the range of scores typically found for native proteins of similar size. The Ramachandran plot indicates that most of the residues are found in the favored and allowed regions (99.7%) and only 0.3% in the outlier region, this shows that the quality of the designed model is satisfactory. Phe 282, Asn311, Tyr228, Lys194, Asp223, Ser229, Ser4, His197, Tyr68 residues were observed in the outlier regions of Ramachandran plot. (Fig.3c). In MolProbability analysis, the all-atom clash score was 9.37, the Molprobability score was 2.21. All structural images were created using PyMol (The PyMOL Molecular Graphics System, Version 1.1, Schrödinger, LLC).

#### **3.4.4. Defining discontinuous B-cell epitopes**

Ellipro Server was used for predicting conformational B-cell epitopes from the 3D structure of NOM recombinant protein. Discontinuous B-cell epitopes were predicted with scores ranging from 0.679 to 0.945. Amino acid residues, the number of residues, sequence location as well as their scores have been listed in Tables 5. The graphical representation of the discontinuous epitopes has been displayed in Figure 4.

#### **3.4.5. Establishing the stability of the initial structures**

A vaccine can interact with different receptors of the immune system. We have docked designed final constructs (NOM) with TLR4 and HLA-A\*11:01 receptors. The interaction between these receptors and NOM recombinant protein induce different immune responses. To achieve a stable and relaxed state of the NOM recombinant protein and the immune receptors, a MD simulation of 100 ns was performed for each structure. These simulations ensure that the structures are stable enough to be used for protein-protein molecular docking. After 100 ns of the production run, as it is shown in figure 5, the RMSD (Root Mean Square Distance) values of the monomers show that the backbone structure of TLR4 (PDB ID: 2Z62) is very rigid and stable. The structure of the NOM recombinant protein also reached stability after about 25 ns after a considerable structural change. The backbone structure of the HLA-A\*11:01 (PDB ID: 5WJL), however, showed big spikes in the RMSD values but all three structures were visually inspected and were considered stable enough for the next step of the project.

#### **3.4.6. Protein-protein molecular docking**

To find the best orientation for optimal interaction of the NOM protein with the immune receptors we decided to use protein-protein molecular docking. We used PatchDock webservice which concentrates on recognizing and matching patterns of the surfaces of the proteins to put them in the best possible positions. Afterward, the top 10 best solutions were subjected to the FireDock webservice to refine the docked structures. The FireDock algorithms refine the docked complexes by side-chain rearrangement and soft rigid-body optimization. The ranking of the docked complexes is based on short-range, long-range, attractive and repulsive interaction energies between the residues of the two proteins, which are all summed up in the global energy or the binding energy of the complexes. We considered global energy as the main criterion for choosing the best complexes and chose the two top-scored solutions

with the best global energy. As it is shown in Table 6, the solution numbers 5 and 6 from the NOM-TLR4 complex and the solution numbers 2 and 3, from NOM-HLA-A\*11:01 complex, were chosen for further calculations. The total energy plots of the simulations are shown in Fig. 6. The results showed a strong interaction between designed protein amino acids and receptors. The interactions between the binding receptors and the docked NOM protein were visualized using the UCSF Chimera program. Residues of HLA-A\*1101 (GLU<sup>53</sup>, ASN<sup>174</sup>, GLY<sup>56</sup>, ASN<sup>174</sup>, GLU<sup>53</sup>, PRO<sup>57</sup>) participated in the interaction with ARG<sup>49</sup>, SER<sup>97</sup>, GLN<sup>91</sup>, MET<sup>102</sup>, ARG<sup>49</sup> and GLN<sup>91</sup> residues of the candidate vaccine. Also, nine residues of TLR4 (TLR4 LYS<sup>123</sup>, ILE<sup>319</sup>, THR<sup>66</sup>, SER<sup>90</sup>, LYS<sup>124</sup>, LYS<sup>46</sup>, HIS<sup>85</sup>, LYS<sup>123</sup>) participated in the interaction with SER<sup>45</sup>, GLU<sup>199</sup>, GLN<sup>91</sup>, ALA<sup>84</sup>, MET<sup>328</sup>, GLN<sup>65</sup>, SER<sup>90</sup>, GLU<sup>68</sup> residues of the designed candidate vaccine.

### 3.4.7. Molecular dynamics simulation of top scored solutions

After the protein-protein molecular docking, 4 solutions were chosen for MD simulation, solution numbers 2 and 3 from NOM protein-HLA-A\*11:01 complex and solution numbers 5 and 6 from NOM protein-TLR4 complex. Each was simulated for 100 ns. To examine their stability throughout the simulations period, the RMSD values of each protein was analyzed. As it is shown in figure 7, each protein has different RMSD values and in some of them, the backbone atoms have considerable movements during the simulations. The reason is that the interactions can get optimized and rearranged by introducing water molecules and physiological conditions. The RMSD values of TLR4 in NOM protein-TLR4 complexes and the values of HLA-A\*11:01 in NOM protein-HLA-A\*11:01 complexes are very stable and the average values do not reach very high values. However, the RMSD values of the NOM protein in every complex reach high values which means that the structure of the vaccine has more movements to refine the interactions with the immune receptors. However, the interactions between the NOM recombinant protein and the immune receptors are quite strong and MMPBSA binding energy calculations exhibit great binding energies. As it is shown in table 7, the Van der Waals and electrostatic energies of the complexes are strong enough to keep the two proteins in contact with each other. Moreover, the deviation of each value is very small which means that the interactions are very stable and consistent throughout the simulations. In another word, the higher scale of RMSD values with consistent binding energies only show that the structure of the NOM protein is well optimized for interacting with the immune receptors.

In addition to the analysis described above, to analyze the fluctuations of the backbone atoms of structures of the proteins, we decided to perform RMSF (Root Mean Square Fluctuation) analysis. In this analysis, the average value of fluctuation of each residue during the simulation is plotted, figure 8. As it is shown in figure 8a, the RMSF values of NOM recombinant protein in five simulations indicate that the fluctuation of the monomer form of the vaccine in many regions is considerably more than the complex forms. This is a direct indication that the NOM recombinant protein is much more stable when it is in complex with the two immune receptors. Furthermore, the RMSF values of the NOM recombinant protein in complex with HLA-A\*11:01 are lower compared with that of NOM recombinant protein in complex with TLR4. Also, with the evidence that the binding energies are also lower, we can conclude that NOM recombinant protein can bind to the HLA-A\*11:01 better than the other immune receptor.

In Figure 8b and Figure 8c, the RMSF values of the monomer forms of the immune receptors show lower values compared to the complex forms. This behavior is exactly the opposite of the NOM designed protein. This can be an indication that the structures are very stable in their natural function and their structures considerably change when they get into contact with the designed NOM protein. Furthermore, the structural illustration of the monomer forms and also the complex forms of the vaccine and the immune receptors, Figure 9, shows that the vaccine can fill the cavities and bind tightly to them, as it was proved by the binding energies.

#### **4. Conclusion**

COVID-19 pandemic is much more than a health crisis. It leads to a political, social and economic crisis in the world. The development of a safe and effective vaccine could reduce the rate of this infection. Immunoinformatics methods are valuable in reducing time and cost in vaccine design and other fields of life sciences. We have predicted and validated NOM recombinant protein against HLA-A\*11:01 and TLR4 receptors. Our evaluation was based on vaccine candidate structural analysis and molecular docking and MD simulations study. The NOM-TLR4 and NOM-HLA-A\*11:01 complexes were very stable in their natural function with strong molecular interactions in around 100 ns. Higher binding energy even after the MD simulation of 100ns confirmed the stability and specificity NOM-TLR4 and NOM-HLA-A\*11:01 interaction. Our vaccine candidate can stimulate both cellular and humoral

immunity, given that B and T-cell epitopes have been selected in the final construct. Taken all together, according to physicochemical evaluations as well as structural and immunological analyses, NOM recombinant protein could be considered as a possible vaccine candidate against COVID-19.

### **Acknowledgments**

This work was supported by the Hormozgan University of Medical Sciences under Grant [number: HUMS-REC.1398.647], Bandar Abbas, Iran.

### **Competing financial interests**

Authors declare no competing financial interests.

### **References**

- Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2, 19-25.
- Ahmad, T.A., Eweida, A.E., and El-Sayed, L.H. (2016). T-cell epitope mapping for the design of powerful vaccines. *Vaccine Reports* 6, 13-22.
- Ahmed, S.F., Quadeer, A.A., and McKay, M.R. (2020). Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses* 12, 254.
- Al-Tawfiq, J.A., Zumla, A., and Memish, Z.A. (2014). Travel implications of emerging coronaviruses: SARS and MERS-CoV. *Travel medicine and infectious disease* 12, 422-428.
- Andrusier, N., Nussinov, R., and Wolfson, H.J. (2007). FireDock: fast interaction refinement in molecular docking. *Proteins* 69, 139-159.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res* 28, 235-242.
- Chan, J.F.-W., Kok, K.-H., Zhu, Z., Chu, H., To, K.K.-W., Yuan, S., and Yuen, K.-Y. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections* 9, 221-236.
- Channappanavar, R., Zhao, J., and Perlman, S. (2014). T cell-mediated immune response to respiratory coronaviruses. *Immunologic research* 59, 118-128.
- Chiarella, P., Massi, E., De Robertis, M., Fazio, V.M., and Signori, E. (2009). Recent advances in epitope design for immunotherapy of cancer. *Recent patents on anti-cancer drug discovery* 4, 227-240.



Cong, Y., Kriegenburg, F., de Haan, C.A., and Reggiori, F. (2017). Coronavirus nucleocapsid proteins assemble constitutively in high molecular oligomers. *Scientific reports* 7, 1-10.

Culshaw, A., Ladell, K., Gras, S., McLaren, J.E., Miners, K.L., Farenc, C., van den Heuvel, H., Gostick, E., Dejnirattisai, W., Wangteeraprasert, A., *et al.* (2017). Germline bias dictates cross-serotype reactivity in a common dengue-virus-specific CD8<sup>+</sup> T cell response. *Nature Immunology* 18, 1228-1237.

Dar, H., Zaheer, T., Rehman, M.T., Ali, A., Javed, A., Khan, G.A., Babar, M.M., and Waheed, Y. (2016). Prediction of promiscuous T-cell epitopes in the Zika virus polyprotein: An in silico approach. *Asian Pacific journal of tropical medicine* 9, 844-850.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* 98, 10089-10092.

Deming, D., Sheahan, T., Heise, M., Yount, B., Davis, N., Sims, A., Suthar, M., Harkema, J., Whitmore, A., and Pickles, R. (2007). Correction: vaccine efficacy in senescent mice challenged with recombinant SARS-CoV bearing epidemic and zoonotic spike variants. *PLoS medicine* 4, 397-398.

Desai, D.V., and Kulkarni-Kale, U. (2014). T-Cell Epitope Prediction Methods: An Overview. In *Methods in Molecular Biology* (Springer New York), pp. 333-364.

Dimitrov, I., Flower, D.R., and Doytchinova, I. (2013). AllerTOP-a server for in silico prediction of allergens. In *BMC bioinformatics* (BioMed Central), p. S4.

Dörner, T., and Radbruch, A. (2007). Antibodies and B cell memory in viral immunity. *Immunity* 27, 384-392.

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

EL-Manzalawy, Y., Dobbs, D., and Honavar, V. (2008). Predicting linear B-cell epitopes using string kernels. *Journal of Molecular Recognition: An Interdisciplinary Journal* 21, 243-255.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics* 103, 8577-8593.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (Springer), pp. 571-607.

He, R., Leeson, A., Ballantine, M., Andonov, A., Baker, L., Dobie, F., Li, Y., Bastien, N., Feldmann, H., and Strocher, U. (2004). Characterization of protein–protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. *Virus research* 105, 121-125.

Hess, B., Bekker, H., Berendsen, H.J.C., and Fraaije, J.G.E.M. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* 18, 1463-1472.

Heymann, D.L. (2020). Data sharing and outbreaks: best practice exemplified. *The Lancet* 395, 469-470.

Hsueh, P.-R., Huang, L.-M., Chen, P.-J., Kao, C.-L., and Yang, P.-C. (2004). Chronological evolution of IgM, IgA, IgG and neutralisation antibodies after infection with SARS-associated coronavirus. *Clinical microbiology and infection* *10*, 1062-1066.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., and Gu, X. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* *395*, 497-506.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics* *14*, 33-38.

Janice Oh, H.-L., Ken-En Gan, S., Bertolotti, A., and Tan, Y.-J. (2012). Understanding the T cell immune response in SARS coronavirus infection. *Emerging microbes & infections* *1*, 1-6.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* *79*, 926-935.

Jorgensen, W.L., Maxwell, D.S., and Tirado-Rives, J. (1996). Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* *118*, 11225-11236.

Kim, H.M., Park, B.S., Kim, J.-I., Kim, S.E., Lee, J., Oh, S.C., Enkhbayar, P., Matsushima, N., Lee, H., Yoo, O.J., and Lee, J.-O. (2007). Crystal Structure of the TLR4-MD-2 Complex with Bound Endotoxin Antagonist Eritoran. *Cell* *130*, 906-917.

Kloczkowski, A., Ting, K.L., Jernigan, R., and Garnier, J. (2002). Combining the GOR V algorithm with evolutionary information for protein secondary structure prediction from amino acid sequence. *Proteins: Structure, Function, and Bioinformatics* *49*, 154-166.

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

Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., *et al.* (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* *33*, 889-897.

Kraus, D. (2014). Consolidated data analysis and presentation using an open-source add-in for the Microsoft Excel® spreadsheet software. *Medical Writing* *23*, 25-28.

Kruiswijk, C., Richard, G., Salverda, M.L., Hindocha, P., Martin, W.D., De Groot, A.S., and Van Riet, E. (2020). In silico identification and modification of T cell epitopes in pertussis antigens associated with tolerance. *Human Vaccines & Immunotherapeutics*, 1-9.

Kumari, R., Kumar, R., Open Source Drug Discovery, C., and Lynn, A. (2014). g\_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model* *54*, 1951-1962.

Leung, D.T.M., Chi Hang, T.F., Chun Hung, M., Sheung Chan, P.K., Cheung, J.L.K., Niu, H., Tam, J.S.L., and Lim, P.L. (2004). Antibody response of patients with severe acute respiratory syndrome (SARS) targets the viral nucleocapsid. *The Journal of infectious diseases* *190*, 379-386.

Lim, Y.X., Ng, Y.L., Tam, J.P., and Liu, D.X. (2016). Human coronaviruses: a review of virus–host interactions. *Diseases* 4, 26.

Liu, W.J., Zhao, M., Liu, K., Xu, K., Wong, G., Tan, W., and Gao, G.F. (2017). T-cell immunity of SARS-CoV: Implications for vaccine development against MERS-CoV. *Antiviral research* 137, 82-92.

Liu, X., Shi, Y., Li, P., Li, L., Yi, Y., Ma, Q., and Cao, C. (2004). Profile of antibodies to the nucleocapsid protein of the severe acute respiratory syndrome (SARS)-associated coronavirus in probable SARS patients. *Clin. Diagn. Lab. Immunol.* 11, 227-228.

Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R., and Wolfson, H.J. (2008). FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Res* 36, W229-232.

McBride, R., and Fielding, B.C. (2012). The role of severe acute respiratory syndrome (SARS)-coronavirus accessory proteins in virus pathogenesis. *Viruses* 4, 2902-2923.

Mirza, M.U., Rafique, S., Ali, A., Munir, M., Ikram, N., Manan, A., Salo-Ahen, O.M., and Idrees, M. (2016). Towards peptide vaccines against Zika virus: Immunoinformatics combined with molecular dynamics simulations to predict antigenic epitopes of Zika viral proteins. *Scientific reports* 6, 37313.

Oberholser, K. (2010). Proteopedia entry: Ramachandran plots. *Biochemistry and Molecular Biology Education* 38, 430-430.

Organization, W.H. (2020). Coronavirus disease 2019 (COVID-19): situation report, 45.

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

Peng, H., Yang, L.-t., Wang, L.-y., Li, J., Huang, J., Lu, Z.-q., Koup, R.A., Bailer, R.T., and Wu, C.-y. (2006). Long-lived memory T lymphocyte responses against SARS coronavirus nucleocapsid protein in SARS-recovered patients. *Virology* 351, 466-475.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605-1612.

Ponomarenko, J., Bui, H.-H., Li, W., Fusseder, N., Bourne, P.E., Sette, A., and Peters, B. (2008). ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC bioinformatics* 9, 514.

Ponomarenko, J., Papangelopoulos, N., Zajonc, D.M., Peters, B., Sette, A., and Bourne, P.E. (2010). IEDB-3D: structural data within the immune epitope database. *Nucleic acids research* 39, D1164-D1170.

Reche, P.A., and Reinherz, E.L. (2007). Prediction of peptide-MHC binding using profiles. In *Immunoinformatics* (Springer), pp. 185-200.

Reynolds, C.R., Islam, S.A., and Sternberg, M.J.E. (2018). EzMol: A Web Server Wizard for the Rapid Visualization and Image Production of Protein and Nucleic Acid Structures. *Journal of molecular biology* 430, 2244-2248.

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H.J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic acids research* 33, W363-W367.

Shey, R.A., Ghogomu, S.M., Esoh, K.K., Nebangwa, N.D., Shintouo, C.M., Nongley, N.F., Asa, B.F., Ngale, F.N., Vanhamme, L., and Souopgui, J. (2019). In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Scientific reports* 9, 1-18.

Shin, W.-H., Lee, G.R., Heo, L., Lee, H., and Seok, C. (2014). Prediction of protein structure and interaction by GALAXY protein modeling programs. *Bio Design* 2, 1-11.

Siu, K.-L., Yuen, K.-S., Castaño-Rodríguez, C., Ye, Z.-W., Yeung, M.-L., Fung, S.-Y., Yuan, S., Chan, C.-P., Yuen, K.-Y., and Enjuanes, L. (2019). Severe acute respiratory syndrome coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. *The FASEB Journal* 33, 8865-8877.

Skwarczynski, M., and Toth, I. (2016). Peptide-based synthetic vaccines. *Chemical science* 7, 842-854.

Testa, J.S., and Philip, R. (2012). Role of T-cell epitope-based vaccine in prophylactic and therapeutic applications. *Future virology* 7, 1077-1088.

Wang, C., Horby, P.W., Hayden, F.G., and Gao, G.F. (2020). A novel coronavirus outbreak of global health concern. *The Lancet* 395, 470-473.

Wiederstein, M., and Sippl, M.J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research* 35, W407-W410.

Wieser, A., Romann, E., Magistro, G., Hoffmann, C., Nörenberg, D., Weinert, K., and Schubert, S. (2010). A multiepitope subunit vaccine conveys protection against extraintestinal pathogenic *Escherichia coli* in mice. *Infection and immunity* 78, 3432-3442.

Xu, J., Zhao, S., Teng, T., Abdalla, A.E., Zhu, W., Xie, L., Wang, Y., and Guo, X. (2020). Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV. *Viruses* 12, 244.

Yang, Z.-y., Kong, W.-p., Huang, Y., Roberts, A., Murphy, B.R., Subbarao, K., and Nabel, G.J. (2004). A DNA vaccine induces SARS coronavirus neutralization and protective immunity in mice. *Nature* 428, 561-564.

Ying, L., Xu, S., Yang, R.F., Li, Y.X., Ji, Y.Y., He, Y.Y., De Shi, M., Wei, L., Shi, T.L., and Jin, W. (2003). Identification of an epitope of SARS-coronavirus nucleocapsid protein. *Cell research* 13, 141-145.

Zheng, J., Lin, X., Wang, X., Zheng, L., Lan, S., Jin, S., Ou, Z., and Wu, J. (2017). In silico analysis of epitope-based vaccine candidates against hepatitis B virus polymerase protein. *Viruses* 9, 112.

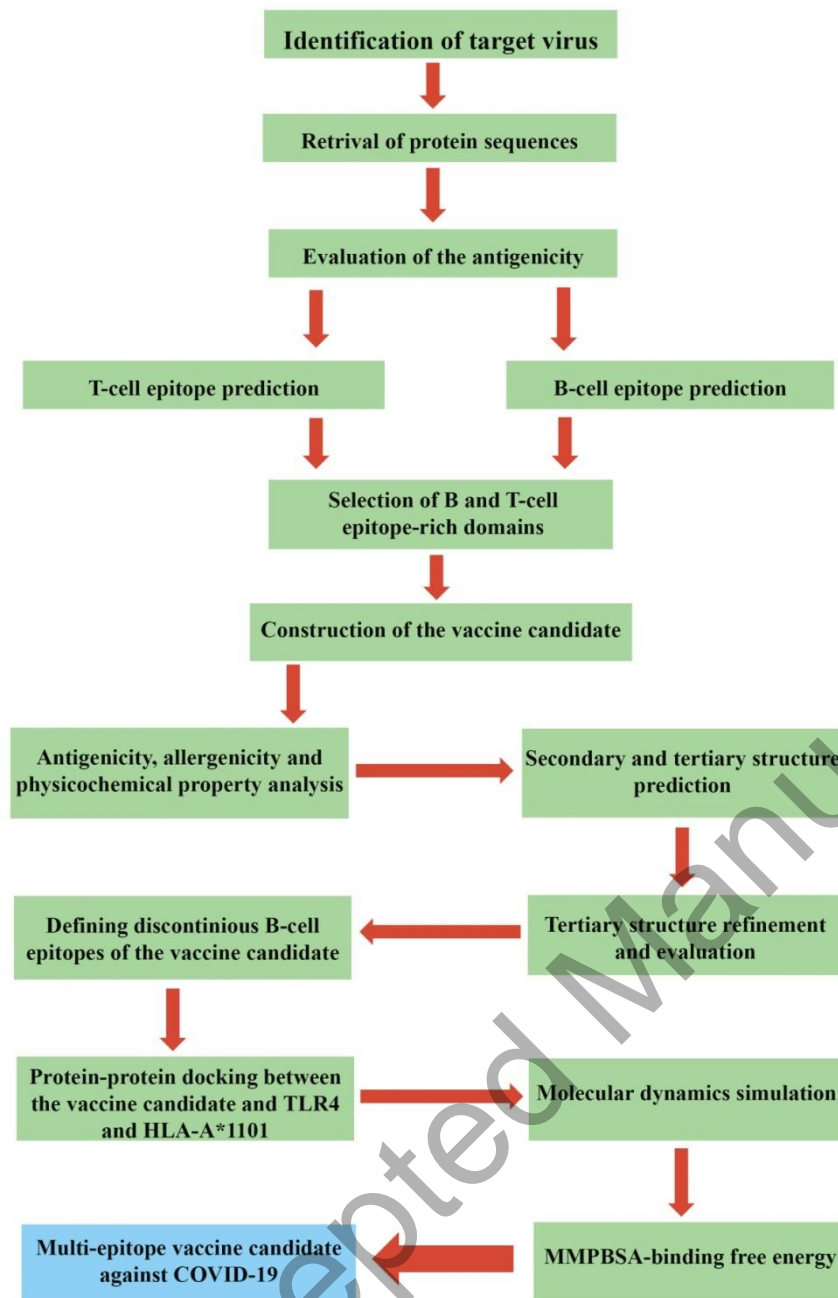


Fig. 1. Strategies employed in the overall study

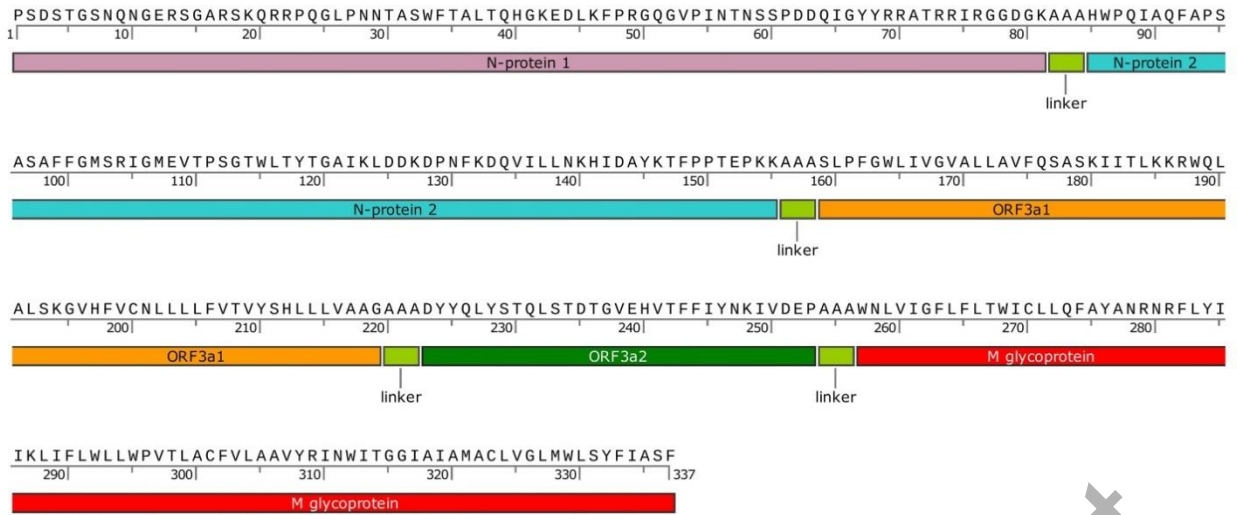


Fig. 2. The schematic diagram of the vaccine candidate construct consists of N, ORF3a and M proteins epitopes of the COVID-19 linked together with AAA linkers.

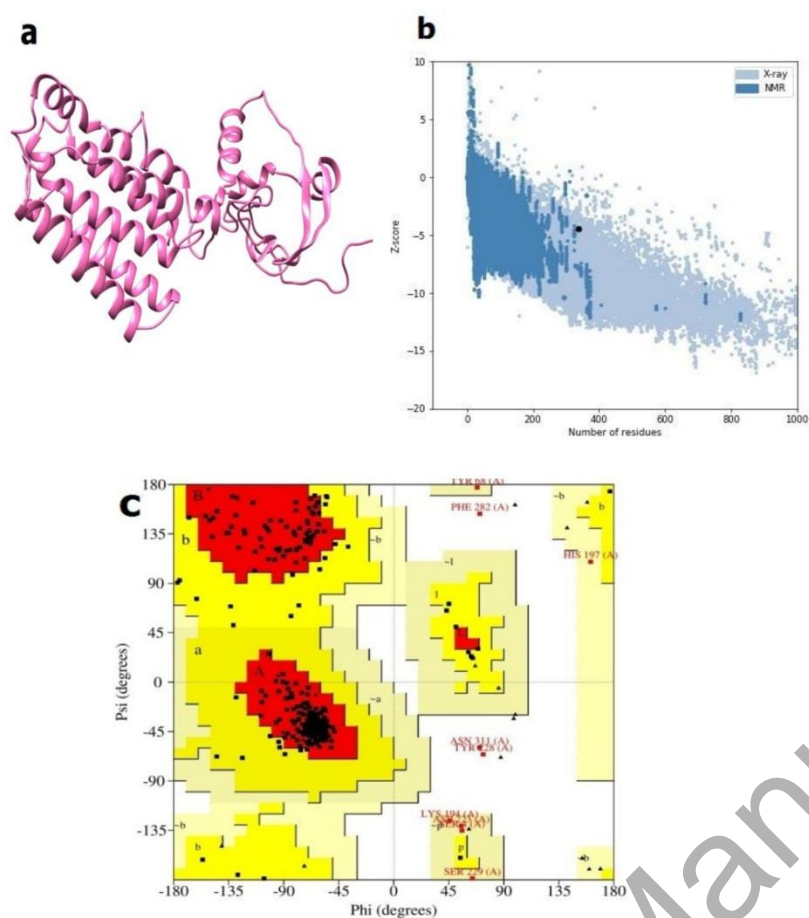


Fig. 3. (a-c) Prediction and validation of tertiary structure of the NOM recombinant protein using (a) Prediction of the tertiary structure of the NOM recombinant protein, (b) ProSA web, (c) Ramachandran plot.

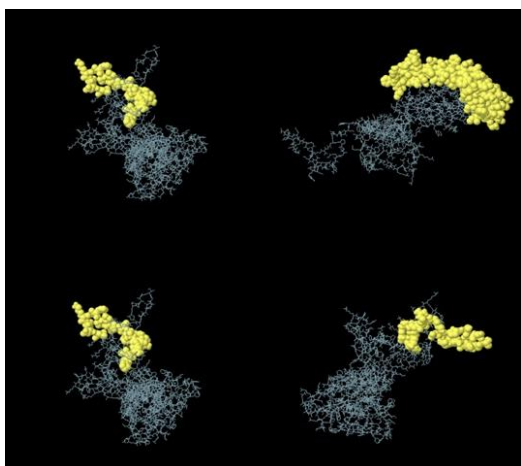


Fig. 4. Three-dimensional representation of discontinuous epitopes of the NOM designed protein. The epitopes are represented by a yellow surface, and the bulk of the polyprotein is represented in grey sticks.

Accepted Manuscript



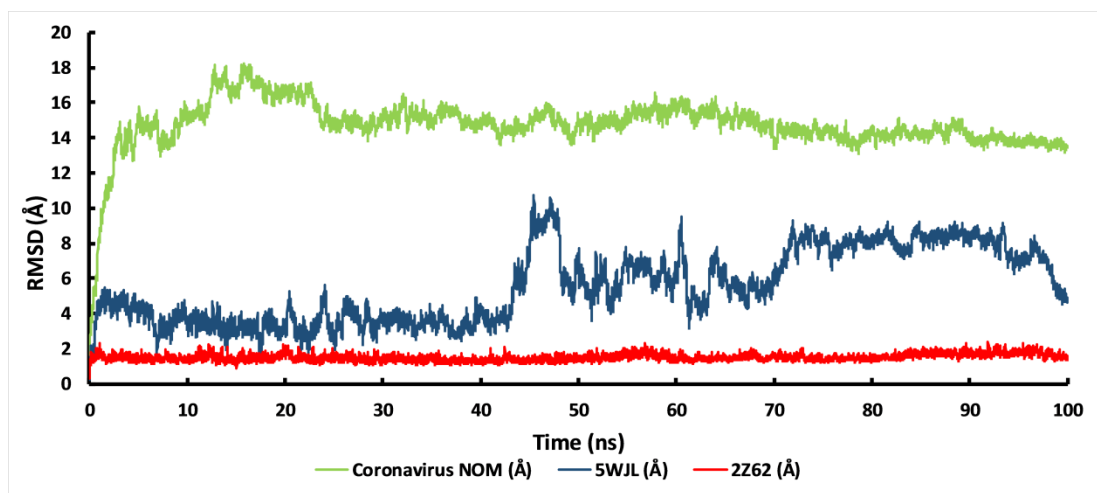


Fig. 5. The RMSD values of the simulated monomer forms of the proteins throughout the 100 ns of production runs.

Accepted Manuscript

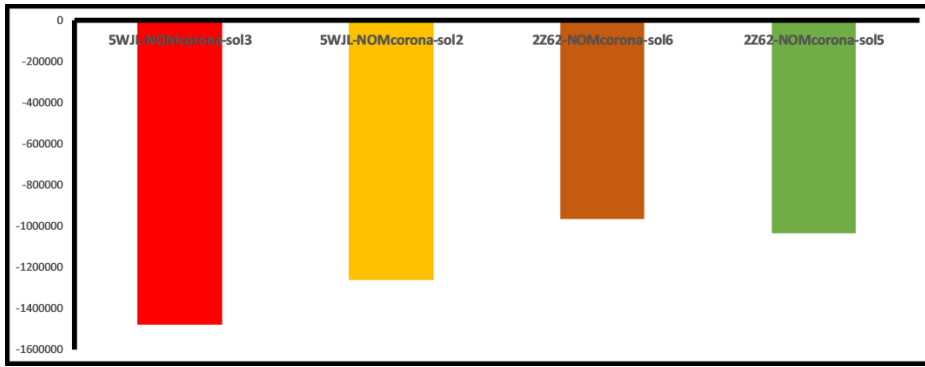


Fig. 6. The total energy plots of the simulations

Accepted Manuscript

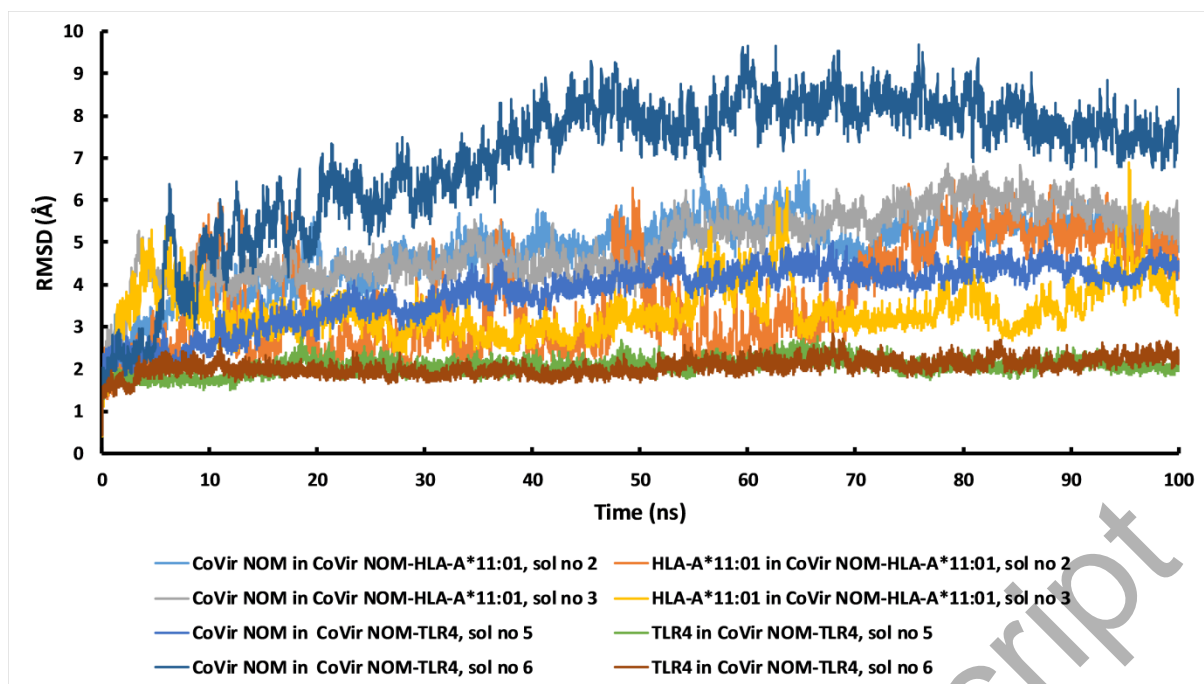


Figure 7.

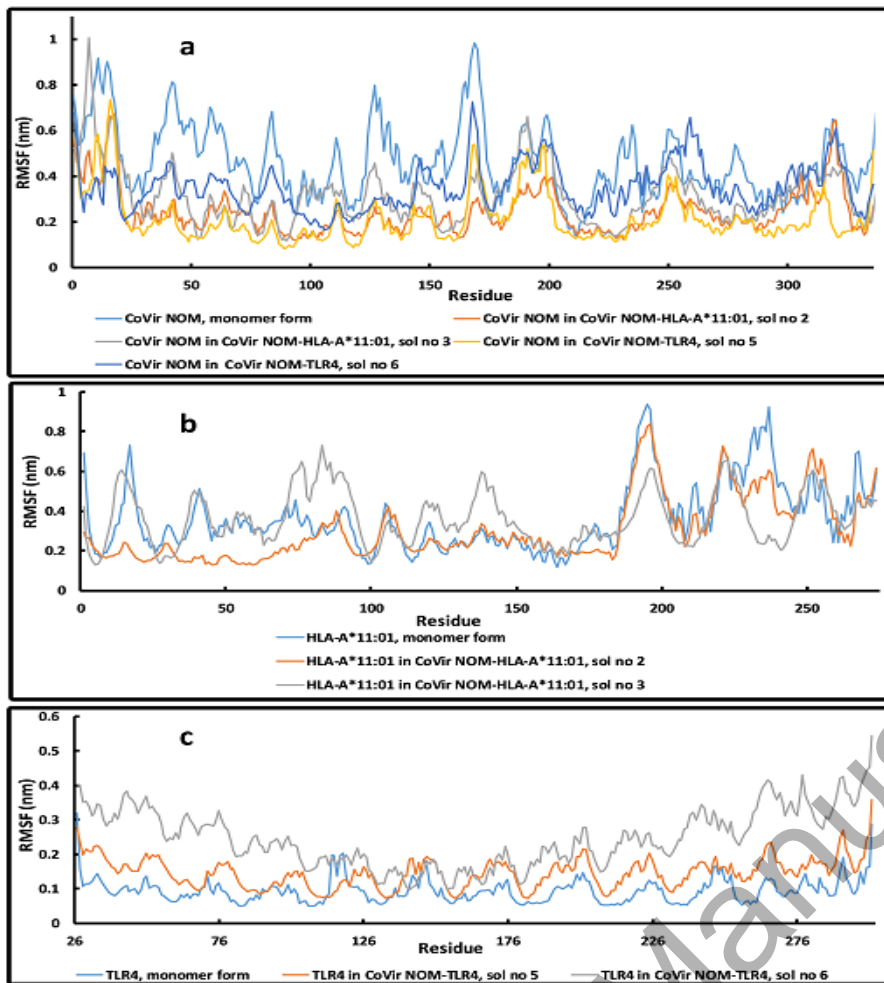


Fig. 8. The RMSF values of each protein in the simulated complexes compared to the simulated monomer forms of the proteins throughout the 100 ns of production runs. a, The comparison of the RMSF values of NOM recombinant protein in the complexes with the monomer form. b, The comparison of the RMSF values of HLA-A\*11:01 in the complexes with the simulated monomer form. c, The comparison of the RMSF values of TLR4 in the complexes with the simulated monomer form.

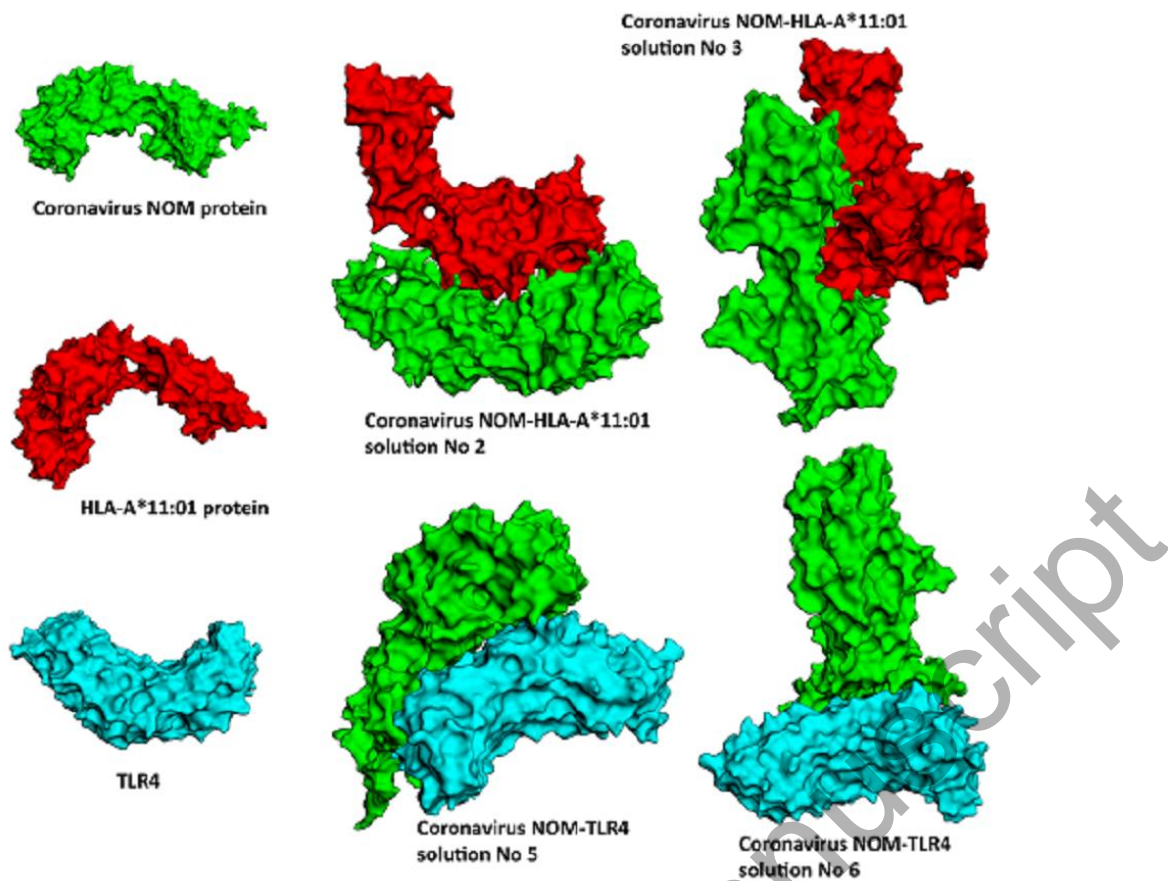


Fig. 9. The graphical illustration of the monomer forms and the complex forms of the NOM recombinant protein and the HLA-A\*11:01 and TLR4 immune receptors.

**Table 1:** Amino acid sequences of proteins were retrieved from NCBI.

Name protein	Accession number	FASTA
Nucleocapsid protein [Severe acute respiratory syndrome coronavirus 2]	<b>QIC53221.1</b>	>QIC53221.1 nucleocapsid protein [Severe acute respiratory syndrome coronavirus 2] MSDNGPQNQRNAPRITFGGPSDDTGSNQNNGERSGARSKQRRPQGLPNNTASWFTALTQHGKEDL KFPRGQGVPIINTSSPDDQIGYYRRATRRIRGGDGKMKDLSRPFYFYLGTGPEAGLPYGANKDGII WVATEGALNTPKDHIGTRNPANNAIIVLQLPQGTTLPKGFYAEGRSGGSQASSRSSRSRNSRNS TPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKQQQQQTVTKKSAAEASKKPRQ KRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEV TPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTFPPEPKDKKKKKADETQALPQRQKKQ QVTLLPAADLDDFSKQLQQSMSSADSTQA
Membrane protein [Severe acute respiratory syndrome coronavirus 2]	<b>QIC53216.1</b>	>QIC53216.1 membrane protein [Severe acute respiratory syndrome coronavirus 2] MADSNGTITVEELKLLLEQWNLVIGFLFLTWICLLQFAYANRNRFYIILKIFLWLLWPVTLACFV LAAVYRINWITGGIAIAMAACLVGLMWLSYFIASFRLFARTRSMWSFNPETNILLNVPPLHGTLTRPL LESELVIGAVILRGLHRIAGHHLGRCDIKDLPKEITVATSRTLSTYKLGASQRVAGDSGFAAYSRY RIGNYKLNTHSSSSDNIALLV
ORF10 protein [Severe acute respiratory syndrome coronavirus 2]	<b>QIC53212.1</b>	>QIC53212.1 ORF10 protein [Severe acute respiratory syndrome coronavirus 2] MGYINVFAPFTIYSLLLCRMNSRNYIAQVDVVFNL
Envelope protein [Severe acute respiratory syndrome coronavirus 2]	<b>QIC53206.1</b>	>QIC53206.1 envelope protein [Severe acute respiratory syndrome coronavirus 2] MYSFVSEETGLIVNSVLLFLAFVFLVTLAILTALRLCAYCCNIVNVSIVKPSFYVYSRVKLNLS SRVPDLLV
ORF8 protein [Severe acute respiratory syndrome coronavirus 2]	<b>QIC53210.1</b>	>QIC53210.1 ORF8 protein [Severe acute respiratory syndrome coronavirus 2] MKFLVFLGIITVAAFHQECSLQSQCTQHQPYPVDDPCPIHFYSKYRIRVGARKSAPLIELCVDEAG SKSPIQYIDIGNYTVSCLPFTINCOEPKLGSLVVRCSFYEDFLEYHDVRRVLDLFI
ORF3a protein [Severe acute respiratory syndrome coronavirus 2]	<b>QIC53205.1</b>	>QIC53205.1 ORF3a protein [Severe acute respiratory syndrome coronavirus 2] MDLFMRIFTIGTVTLKQGEIKDATPSDFVRATATIPIQASLPFGWLIVGVALLAVFQSASKIITLKKR WQLALSKGVHFCNLLLFVTVYSHLLLVAAGLEAPFLYLYALVYFLQSFVRIIMRLWLCWKC RSKNPLLYDANYFLCWHNTCYDYCIPYNSVTSSIVITSGDGTTPISEHDYQIGGYTEKWESGVKD CVVLSYFTSDYYQLYSTQLSTDTGVEHVTFFIYNKIVDEPEEHVQIHTIDVSSGVVNPVMEPIYDE PTTTTSVPL

**Table 2a: HLA I antigenic epitopes predicted using Rankpep.**

Antigen	HLA_A0201	HLA_A0204	HLA_A0206	HLA_B0702	HLA_B51	HLA_B5401	HLA_B5301
Nucleocapsid protein	316-324 GMSRIGMEV	299-307 KHWPQIAQF		66-74 FPRGQGVPI	343-351 DPNFKDQVI 45-53 LPNNTASWF 66-74 FPRGQGVPI	66-74 FPRGQGVPI	45-53 LPNNTASWF
ORF3a	39-47 ASLPFGWLIV 46-55 LIVGVALLAV 64-73 TLKKRWQLAL 79-87 FVCNLLLLFV	39-47 ASLPFGWLI 72-80 ALSKGVHFV 220-228 STDTGVEHV	45-53 WLIVGVALL 72-80 ALSKGVHFV 82-90 NLLLLFVTV	35-43 IPIQASLPF	41-49 LPFGWLIVG 35-43 IPIQASLPF		
Membrane protein	16-24 LLEQWNLVI 15-23 KLLEQWNLV 65-73 FVLAAYYRI	56-64 LLWPVTLAC 96-104 FIASFRLFA	21-29 NLVIGFLFL <b>65-73</b> FVLAAYYRI 89-97 GLMWLSYFI 96-104 FIASFRLFA	58-66 WPVTLACFV	58-66 WPVTLACFV	58-66 WPVTLACFV	

**Table 2b:** HLA II antigenic epitopes predicted using Rankpep.

Antigen	HLADRB1010 1	HLADRB1040 1	HLADRB1040 2	HLADRB1040 2	HLADRB1070 1	HLADRB1080 1	HLADRB1110 1	HLADRB1150 1
Nucleocapsid protein	298-305 YKHWPQIAQ 354-362 NKHIDAYKT 86-94 YYRRATRRRI 305-313 AQFAPSASA 52-60 WFTALTQHG	52-60 WFTALTQHG 86-94 YYRRATRRRI 300-308 HWPQIAQFA 49-57 TASWFTALT			346-354 FKDQVILLN 300-308 HWPQIAQFA	41-49 RPQGLPNNT 348-356 DQVILLNKH	87-95 YRRATRRIR 86-94 YYRRATRRRI 34-42 GARSKORRP 298-306 YKHWPQIAQ 301-309 WPQIAQFAP	50-58 ASWFTALTQ 360-368 YKTFPPTEP
ORF3a	211-219 YYQLYSTQL 212-220 YQLYSTQLS 65-73 LKKRWQLAL	211-219 YYQLYSTQL 45-53 WLIVGVALL	45-53 WLIVGVALL 211-219 YYQLYSTQL	62-70 IITLKKRWQ 85-93 LLFVTVYSH	59-67 ASKIITLKK 87-95 FVTVYSHLL 54-62 AVFQSASKI		62-70 IITLKKRWQ 68-76 RWQLALSKG	77-85 VHFVCNLLL 84-92 LLLFTVYYS
Membrane protein	32-40 ICLLQFAYA 65-73 FVLAAYVRI 76-84 ITGGIALAM 80-88 IAIAMAACLV	28-36 FTWICLLQ 55-63 WLLWPVTLA 65-73 FVLAAYVRI 71-79 YRINWITGG	<b>28-36</b> FTWICLLQ <b>49-57</b> IKLIFLWLL	<b>37-45</b> FAYANRNRF	65-73 FVLAAYVRI 48-56 IIKLIFLWL	<b>44-52</b> RELYIHKLI 55-63 WLLWPVTLA <b>58-66</b> WPVTLACFV	<b>31-39</b> WICLLQFAY 39-47 YANRNRFY 90-98 LMWLSYFIA	



**Table 3:** Predicted epitopes of N, ORF3a and M proteins via Bepipred and Kolaskar & Tongaonkar antigenicity.

Antigen	Server	Amino acid Position	Sequence
Nucleocapsid protein (QIC53221.1)	<b>Bepipred</b>	1-51	MSDNGPQNQRNAPRITFGGSPSDSTGSNQNGERSGARSQRRPQGLPNNTAS
		58-85	QHGKEDLKFPRGQGVPIINTSSPDDQIG
		93-104	RIRGGDGKMKDL
		306-310	QFAPS
		323-321	EVTPSGTWL
		338-347	KLDDKDPNFK
		361-390	KTFPPEPKDKKKKADETQALPQRQKKQQ
	<b>Kolaskar &amp; Tongaonkar</b>	52-59	WFTALTQH
		69-75	GQGVPI
		83-89	QIGYYRR
		299-315	KHWPQIAQFAPSASAFF
		333-339	YTGAIKL
		347-363	KDQVILLNKHIDAYKTF
ORF3a (QIC53205.1)	<b>Bepipred</b>	17-28	QGEIKDATPSDF
		61-71	KIITLKKRWQL
		213-214	QL
		216-225	STQLSTDTGV
	<b>Kolaskar &amp; Tongaonkar</b>	44-58	GWLIVGVALLAVFQS
		74-100	SKGVHFVCNLLLLFVTVYSHLLLVAAG
		212-217	YQLYST
Membrane protein (QIC53207.1)	<b>Bepipred</b>	5-20	NGTTTVEELKKLLEQW
		40-41	AN
		132-137	PLLESE
		180-191	KLASQRVAGDS
	<b>Kolaskar &amp; Tongaonkar</b>	29-38	LTWICLLQFA
		46-71	LYIILIFLWLLWPVTLACFVLAAYY
		83-91	AMACLVGLM
		93-101	LSYFIASFR

**Table 5:** Conformational B-cell epitopes from vaccine protein using Ellipro server

No.	Residues	Number of residues	Score
1	A:P1, A:S2, A:D3, A:S4, A:T5, A:G6, A:S7, A:N8, A:Q9, A:N10, A:G11, A:E12, A:S14, A:G15, A:A16, A:R17, A:S18, A:K19, A:Q20, A:R21, A:R22	21	0.945
2	A:K123, A:L124, A:D125, A:D126, A:K127, A:D128, A:P129, A:N130, A:F131, A:K132, A:D133, A:Q134, A:V135, A:I136	14	0.819
3	A:P23, A:Q24, A:G25, A:L26, A:P27, A:N28, A:N29, A:T30, A:A31, A:S32, A:W33, A:F34, A:T35, A:A36, A:L37, A:T38, A:Q39, A:H40, A:G41, A:K42, A:E43, A:D44, A:L45	23	0.779
4	A:I181, A:I182, A:T183, A:L184, A:K185, A:K186, A:R187, A:W188, A:Q189, A:L190, A:A191, A:L192, A:S193, A:K194, A:G195, A:V196, A:H197, A:F198, A:V199, A:C200, A:N201, A:L203, A:F244, A:I245, A:Y246, A:N247, A:K248, A:I249, A:V250, A:D251, A:E252, A:P253, A:A254, A:A255, A:A256, A:W257, A:N258, A:L259, A:V260, A:I261, A:G262, A:F263, A:L264, A:F265, A:L266, A:T267, A:W268, A:I269, A:C270, A:L271, A:L272, A:Q273, A:F274, A:A275, A:Y276, A:A277, A:N278, A:R279, A:N280, A:R281, A:F282, A:L283, A:I285, A:I286, A:I289, A:L304, A:A305, A:A306, A:Y308, A:R309, A:I310, A:N311, A:W312, A:I313, A:T314, A:G315, A:G316, A:I317, A:I319	79	0.679

**Table 6.** The rankings of the solution of the complexes of NOM protein and the immune receptors sorted by global energy (kJ/mol).

Complex	No of Solution	glob energy	aVdW	rVdW	aElec	rElec	laElec	lrElec
CoVir NOM-TLR4	6	-70.72	-35.8	15.25	-59.29	32.2	-9.04	7.84
CoVir NOM-TLR4	5	-36.56	-38.35	23.35	-35.83	0	-8.39	8.75
CoVir NOM-TLR4	7	-36.39	-24.48	25.24	0	0	0	0
CoVir NOM-TLR4	9	-36.21	-35.91	16.81	-16.15	12.49	-16.07	8.56
CoVir NOM-TLR4	4	-34.54	-20.58	9.45	-11.25	0	-5.55	0
CoVir NOM-TLR4	3	-21.92	-34.54	8.21	-25.98	58.08	-29.15	12.41
CoVir NOM-TLR4	2	-19.49	-23.45	11.3	-82.71	81.52	-17.3	0
CoVir NOM-TLR4	8	-18.15	-17.09	6.82	0	0	0	0
CoVir NOM-TLR4	10	-6.55	-17.77	6.58	-7.49	5.9	-8.21	7.88
CoVir NOM-TLR4	1	26.74	-3.48	49.43	0	0	-2.35	3.58
CoVir NOM-HLA-A*11:01	2	-26.26	-30.64	19.51	-26.6	52.43	-24.6	8.53
CoVir NOM-HLA-A*11:01	3	-21.9	-19.09	7.4	-47.3	14.42	-39.27	20.53
CoVir NOM-HLA-A*11:01	1	-5.38	-5.26	0.43	0	0	-5.89	3.65
CoVir NOM-HLA-A*11:01	8	-0.97	-8.76	3.27	0	0	0	0
CoVir NOM-HLA-A*11:01	5	0.15	-25.97	7.15	-37.47	90.84	-26.96	28.97
CoVir NOM-HLA-A*11:01	6	1.52	-0.8	0.61	0	0	0	0
CoVir NOM-HLA-A*11:01	9	1.74	-28.7	15.11	-90.47	136.21	-32.16	26.54
CoVir NOM-HLA-A*11:01	4	10.29	-10.6	22.49	-40.45	11.12	-24.32	24.47
CoVir NOM-HLA-A*11:01	7	18.39	-21.1	10.69	-35.18	99.31	-30.65	40.48
CoVir NOM-HLA-A*11:01	10	1498.09	-34.31	2561.4	-73.6	10.46	-7.59	18.87

**Table 7.** The Van der Waals, Electrostatic, Polar solvation, SASA and Binding Energy of protein complexes, kJ/mol, calculated by MMPBSA method.

Complex	Van der Waals	Electrostatic	Polar solvation	SASA	Binding Energy
CoVIR NOM-HLA-A*11:01, sol no 2	-556 +/- 8	-2991.1 +/- 28.5	1354.4 +/- 20.5	-79.46 +/- 1.2	-2267.7 +/- 6.5
CoVIR NOM-HLA-A*11:01, sol no 3	-381.9 +/- 8.7	-2705.2 +/- 11.5	729.4 +/- 10.9	-57.1 +/- 1	-2423.3 +/- 24.9
CoVIR NOM-TLR4, sol no 5	-503 +/- 5.7	-1598.9 +/- 10.8	1202 +/- 6.1	-75.4 +/- 0.3	-978.4 +/- 7.6
CoVIR NOM-TLR4, sol no 6	-407.6 +/- 5.9	-1406.8 +/- 23.4	1009.3 +/- 12.3	-64.8 +/- 0.6	-865.7 +/- 10.6

Accepted Manuscript