

Design an efficient multi-epitope peptide vaccine candidate against SARS-CoV-2:

An in silico analysis

Zahra Yazdani¹, Alireza Rafiei^{*1}, Mohammadreza Yazdani², Reza Valadan¹

1- Department of Immunology, Molecular and Cell Biology Research Center, School of Medicine, Mazandaran University of Medical Sciences, Sari, Iran

2- Department of Chemistry, Isfahan University of Technology, Isfahan, 84156-83111, Iran

Corresponding author:

Alireza Rafiei,

Department of Immunology, Molecular and Cell Biology Research Center,

School of Medicine,

Mazandaran University of Medical Sciences,

KM 18 Khazarabad road, Khazar Sq, Sari, Iran,

E-mail: rafiei1710@gmail.com,

Orcid Id:0000-0002-1766-6605, Tel: +981133543614

Abstract

Background: To date, no specific vaccine or drug has been proven to be effective for SARS-CoV-2 infection. Therefore, we implemented immunoinformatics approach to design an efficient multi-epitopes vaccine against SARS-CoV-2.

Results: The designed vaccine construct has several immunodominant epitopes from structural proteins of Spike, Nucleocapsid, Membrane and Envelope. These peptides promote cellular and humoral immunity and Interferon gamma responses. In addition, these epitopes have antigenicity ability and no allergenicity probability. To enhance the vaccine immunogenicity, we used three potent adjuvants; Flagellin, a driven peptide from high mobility group box 1 as HP-91 and human beta defensin 3 protein. The physicochemical and immunological properties of the vaccine structure were evaluated. Tertiary structure of the vaccine protein was predicted and refined by I-Tasser and galaxi refine and validated using Rampage and ERRAT. Results of Ellipro showed 242 residues from vaccine might be conformational B cell epitopes. Docking of vaccine with Toll-Like Receptors 3, 5 and 8 proved an appropriate interaction between the vaccine and receptor proteins. In silico cloning demonstrated that the vaccine can be efficiently expressed in *Escherichia coli*.

Conclusions: The designed multi epitope vaccine is potentially antigenic in nature and has the ability to induce humoral and cellular immune responses against SARS-CoV-2. This vaccine can interact appropriately with the TLR3, 5, and 8. Also, this vaccine has high quality structure and suitable characteristics such as high stability and potential for expression in *Escherichia coli*.

Key words: SARS-CoV-2, multi-epitope vaccine, structural proteins, humoral immunity, cellular immunity

1. Introduction:

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which the cause of COVID-19 disease, was first reported as a pneumonia epidemic in the Chinese city of Wuhan (Hubei province) in December 31, 2019 belongs to the Beta coronavirus genus[1-3]. Coronaviruses are positive-sense single-stranded RNA viruses. They belong to the order of Nidovirales and superfamily of orthocoronaviridae. This superfamily has four genus including: alpha, beta, gamma and delta coronaviruses. Gamma and delta coronaviruses infect birds, while alpha and beta coronaviruses generally infect mammals such as human. In immunocompetent individuals, they generally cause mild respiratory infections, such as common cold, while in some humans coronavirus infections cause serious disease, such as SARS (Severe Acute Respiratory Syndrome) and MERS (Middle East Respiratory Syndrome) epidemics.

The SARS-CoV-2, the viral cause of COVID-19, has a large mRNA genome, 26 to 32 kb in length, meaning a 5' cap cover structure and a 3' polyadenylated. This encodes several structural and nonstructural proteins. Among structural proteins, spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins are important in inducing immune responses [3-5]. Protein S is the main tool for virus entry into the cell, which interacts with host cell receptor, angiotensin-converting enzyme 2 (ACE2). ACE2 is a metallopeptidase expressed in variant tissues, including alveolar epithelial cells, fibroblasts, endothelial cells, and enterocytes. [6-10]. The N protein is an essential RNA-binding protein for transcription and replication of the viral RNA. It has the significant roles in forming of the helical ribonucleoproteins during packaging the RNA genome, regulating of viral RNA synthesis during replication, transcription and modulating of metabolism in the infected cells [11-14]. T cell responses against the S and N proteins of SARS-CoV virus are most dominant and long-lived than other structural proteins[15]. protein E has an important role in the assembly of the viral genome[16, 17]. The M protein plays a pivotal role in virus assembly, budding and replication of virus particles in the host cells[18].

SARS-CoV-2 is transmitted quickly and causes considerable fatality rate, so that World Health Organization (WHO) reported over 2 114 269 confirmed cases globally and at least 145 144 deaths because of this disease until April 17, 2020[19]. There is currently no vaccine or approved treatment for

this disease. One of the major challenges for scientists in dealing with the new coronavirus pandemic will be to get a useful and effective vaccine. For this reason, trying to develop an effective vaccine to control the virus has been the subject of research by many scientists around the world [4, 20-22]. However, achieving such a vaccine with conventional methods is very time consuming and expensive. As such, *in silico* prediction of the vaccine targets is very important because the targets can be selected with a more open eye in a limited time and resources. Bioinformatic approaches are very helpful to identify the effective epitopes and developing vaccine. Several vaccines for virus diseases were designed based on these approaches that include effective vaccines on Human papilloma viridea (HPV) [23],Ebola [24], Zika[25]and MERS[26, 27]viruses. There are also few reports of a vaccine candidate for COVID19[28, 29]. These vaccines contained multiple cytotoxic T lymphocytes (CTL) and B cell epitopes against several proteins of SARS-CoV-2 virus. However, those vaccines didn't protect against all structural immunodominant proteins of virus. Therefore, to overwhelm these limitations, we designed a multi-epitope peptide vaccine consisted of S, M, N, and E viral proteins. The vaccine has appropriate physicochemical properties such as stability at room temperature, antigenic capability, and no possibility of allergy. This multi-epitope vaccine consists of three epitopes from S protein and one epitope from each of M, N, and E proteins, respectively. Each of the epitopes has a high ability to stimulate humoral and cellular immune responses, especially the proper production of Interferon gamma (IFN- γ). In addition, we used appropriate adjuvants in the vaccine structure to potentiate the immunogenicity of the antigens.

Therefore, we incorporated the potent adjuvants of N and C terminus of Flagellin as a toll like receptor (TLR) 5, a driven peptide from high mobility group box1 (HMGB1) as HP-91 and human beta defensin 3 (HBD-3) in the construct of the vaccine. The vaccine segments were connected to each other by appropriate linkers. Physicochemical properties, structural stability, and immunological characterizations of the vaccine were evaluated. Modeling, refinement, and validation were performed to access a high quality three dimensional (3D) structure of the vaccine protein. Docking evaluation showed an appropriate interaction between the vaccine and TLRs 3, 5 and 8. In silico cloning showed that the

vaccine could be effectively expressed in *E.coli*. Totally, a potential vaccine candidate with proper immunological and stable physicochemical properties against SARS-CoV-2 was designed. It is expected the vaccine could be capable to protect humans from COVID19 disease.

2. Methodology

2.1 Protein sequence retrieval

The protein sequences of spike glycoprotein (accession number of QIC53213.1), nucleocapsid phosphoprotein (QHU79211.1), membrane glycoprotein (QIC53207.1) and envelope protein (QIC53215.1) were driven using NCBI (<https://www.ncbi.nlm.nih.gov/>) databases and saved in FASTA format for subsequent analysis.

2.2 Screening of potential epitopes

2.2.1 Screening of Major histocompatibility complex class I (MHC-I) epitopes

MHC-I humans alleles with 9 mer length were selected using IEDB (www.iedb.org) database through IEDB recommended 2.22 method and Net-MHC 4.0 online server (<http://www.cbs.dtu.dk/services/NetMHC/>). IEDB (Instructor/Evaluator Database) is a repository of web-based tools which predicts and analysis of immune epitopes. It uses a consensus method based on artificial neural network (ANN), stabilized matrix method (SMM), and Combinatorial Peptide Libraries (CombLib), if any corresponding predictor is available for the peptide sequence [30-48]. NetMHC 4.0 software is a free server for the prediction of peptide-MHCI binding affinity by gapped sequence alignment. Prediction methods of this server are based on alignments that include insertions and deletions and have significantly higher performance than methods based on peptides of single lengths methods[32].

2.2.2 Screening of Cytotoxic T lymphocytes (CTL) epitopes

Prediction of CTL epitopes was done by an online server CTL Pred. This method is based on quantitative matrix (QM) and machine learning techniques for example, artificial neural network (ANN) and support vector machine (SVM). The server also uses the consensus and combined prediction approaches. The consensus and combined prediction approaches are more specific and sensitive, respectively, in comparison with individual approaches such as ANN and SVM[49].

2.2.3 Screening of Human leukocyte antigen (HLA)-II epitopes

IEDB database (IEDB recommended 2.22 method) [40, 50-53] (www.iedb.org) and RANKPEP online server (<http://imed.med.ucm.es/Tools/rankpep.html>) were employed for screening of HLA-II epitope. The database of IEDB was described previously. The RANKPEP server predicts MHC-II binding epitope by position specific scoring matrices (PSSMs) that are structurally consistent with the binding mode of MHC-II ligands [54].

2.2.4 Screening of Linear B cell epitopes

Prediction of linear B cell epitope was done by IEDB[55] (Emini surface accessibility method) and BepiPred software. BepiPred 2.0 predicts the location of linear B-cell epitopes by a combination of a hidden Markov model and a propensity scale method. The epitope Threshold of this server was selected 0.5[56].

2.3 Selection of the epitope segments

The results of all the above predictions were pooled and compared together, and the regions with the highest overlaps were determined. These immunodominant regions were employed for future analyses to finally select the most appropriate epitope domains for vaccine construct.

2.4 In silico analyzing of IFN- γ inducing epitopes

IFN epitope server (<http://crdd.osdd.net/raghava/ifnepitope>) was employed for determining the induce IFN- γ production ability in the selected epitopes. This web server classifies MHC binder epitopes into

IFN- γ inducing (positive numbers) and non-inducing IFN- γ (negative numbers) using several methods such as; motifs-based search, machine learning technique and a hybrid approach. Accuracy of Best prediction based on hybrid approach in this software is 82.10%[57].

2.5 Evaluation of antigenicity

Antigenicity of epitopes was assessed by ANTIGENpro (<http://scratch.proteomics.ics.uci.edu/>) and VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) servers. Web server of ANTIGENpro is the first sequence-based, alignment-free, and pathogen independent predictor, using protein antigenicity microarray data for predicting the protein antigenicity[58]. VaxiJen is also the first server for prediction of antigens. This server applies a new alignment-independent approach that is based on auto-cross covariance (ACC) transformation of protein sequences into uniform vectors of principal amino acid properties. Depending on the target organism (bacteria, virus or tumor) accuracy of this server varies from 70% to 89%. The Threshold of this server was selected 0.5[59, 60].

2.6 Allergenicity assessment of predicted epitopes

For predicting protein allergenicity with a high accuracy, the software of AllergenFP v.1.0 (<http://ddg-pharmfac.net/AllergenFP/>) was used. This is an online server that identifies allergens based on amino acid principle properties such as hydrophobicity, helix and β -strand forming with accuracy about 88% [40].

2.7 Vaccine engineering, evaluation of physicochemical properties, antigenicity and allergenicity

According to the results of the previous steps, four epitopes from S protein, two epitope from N protein and one epitope from E protein were selected to be incorporated in the vaccine construction. The whole construction was designed using joining these epitopes to two adjuvant sequences of N and C terminals of Flagellin and the synthesis peptide of HP-91. Various segments of the designed-vaccine connected to each other by suitable linkers.

After designing the vaccine, Several physicochemical parameters including molecular weight, theoretical isoelectric point (pI), total number of positive and negative residues, extinction coefficient, Instability index, half-life, aliphatic index, and grand average hydropathy (GRAVY) were computed using Expasy's ProtParam at <http://us.expasy.org/tools/protparam.html> [61]. Antigenicity of the final vaccine construct was evaluated using vaxijen v2.0 and ANTIGENpro and allergenicity was assayed by AllergenFP v.1.0 server.

2.8 Tertiary structure development

The 3-D modeling of designed construct was predicted using the I-TASSER software (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) [61]. This server showed five models as a result of the prediction. This model was accepted as a vaccine structure with the highest reliability score (c-score) and was refined using Galaxy Refine web services. It successfully tested in CASP10 (Critical assessment of techniques for protein structure prediction) experiments. This server refines the whole protein with gentle and aggressive relaxation methods. This first reconstructs all side-chain conformations and repeatedly relaxes the structure with short molecular dynamics simulations after side-chain repacking perturbations [62]. Finally, Galaxy Refine showed five structures as refined models of vaccine structure. The RAMPAGE (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>), and ERRAT (<http://services.mbi.ucla.edu/ERRAT/>) servers were used to validate the refined 3D structures obtained from Galaxy Refine web service. [63, 64]. Finally, the refined and high-quality 3D structure of the vaccine was observed by PyMOL software v2.1.1. PyMOL is an open-source that widely used for bimolecular function[65].

2.9 Prediction of discontinuous B cell epitope

Discontinuous B-cell epitopes were predicted from the 3D vaccine structure using ElliPro in IEDB database (<http://tools.immuneepitope.org/tools/ElliPro>). ElliPro is a useful research tool for identifying immunoepitopes in protein antigens which implements Thornton's method and with a residue clustering

algorithm. The MODELLER program and the Jmol viewer allow the prediction and visualization of immunopeptides in a given protein sequence or structure. In comparison with six other structure-based methods that are used for epitope prediction, ElliPro performs the best prediction and gave an AUC value of 0.732, when the most significant prediction is considered for each protein. [66].

2.10 Molecular docking with TLR- 3, 5 and 8

The first, tertiary structure of the human TLR-3, 5, and 8 was obtained from PDB database (www.rcsb.org) with codes of 3J0A, 5GS0 and 3w3g, respectively. Next, protein-protein docking of the vaccine structure (as ligand) and each TLR (as receptor) was performed by CLUSPRO 2.0 online server (cluspro.bu.edu/login.php). CLUSPRO 2.0 is a fully automated and fast rigid-body protein-protein docking server. This server evaluates docking of two interacting proteins based on three techniques: first, the Fast Fourier Transform (FFT) correlation approach, second, clustering of the best energy conformations, third, refining the obtained model using short Monte Carlo simulations and the medium-range optimization method SDU [67-69].

2.11 Codon optimization and *in silico* cloning

The reverse translation of the designed-vaccine gene sequence was performed by Sequence Manipulation Suite (https://www.bioinformatics.org/sms2/rev_trans.html) to prepare a suitable vaccine sequence for cloning and expression in an appropriate expression system. The properties of sequence gene such as Codon Adaptation Index (CAI), GC content, and Codon Frequency and Distribution (CFD) have the key roles in attaining a high-level of protein expression in the host were evaluated using GenScript online server (<https://www.genscript.com/tools/rare-codon-analysis>)[24]. Finally, the restriction sites of XhoI and EcoRV were respectively added to the N- and C-terminus of the vaccine DNA sequence using CLC Sequence viewer v8.0 (<http://www.cacbio.com>) to facilitate the cloning in *E. coli* expression system.

3. Results:

3.1. Immunoinformatics analysis

Protein S, M, N, and E were used to predict HLA-I binding epitopes (HLA-A, B, and C) using IEDB and NetMHC 4. CTL Pred server was used to predict CTL epitopes of SARS-CoV-2 structural proteins. HLA-II binding epitope (DP, DQ, and DR) from these proteins were predicted using RANKEP and IEDB servers. Prediction of linear B cell epitopes was performed by BepiPred and IEDB. Finally, the obtained epitopes from the comparison of above analyses were applied to future selection (Table.1).

3.2. Antigenic protein screening and avoid allergenicity

All selected peptides were submitted to the IFNepitope server for evaluation their ability to induce IFN- γ . Allergenicity probability of epitopes evaluated by AllergenFPE valuation of epitope antigenicity was performed by ANTIGEN pro and VaxiJen v2.0. The six epitopes (three from S and one from each of E, M, N proteins) that had the highest ability in IFN- γ induction, low allergenicity and potent antigenic ability were chosen to be used in the vaccine structure (Table 1).

3.3 Designing a multi-epitopes vaccine construct

Six epitopes with high score which selected to be incorporated in the final vaccine construct were the sequences of 20-49, 418-446, and 505-534 from S, 51-71 from E, 360-388 from N and 198-214 from M protein. In addition, three immunopotent adjuvants Flagellin, HP-91, and HBD-3 were also added to the vaccine structure. The vaccine pieces were linked to each other by using a short linker sequence of LE dipeptide repeats. The final multi epitope peptide vaccine was 473 amino acid residues (Fig.1).

4.3 Immunological and physicochemical properties of the vaccine structure

To evaluate antigenicity of the whole protein vaccine, the “virus” option was chosen as a target organism. The antigenicity of 0.5319 (probable antigen) was estimated at 0.5% threshold for the virus model. Assessment of antigenicity by Antigen pro showed this vaccine is high antigenic with antigenicity of 0.941411. AllergenFP showed that the T cell epitomes in the vaccine protein are non-allergen. The molecular weight (MW) and theoretical isoelectric point (pI) of the vaccine protein were computed as 51.916kDa and 7.45, respectively. The predicted half-lives were calculated as 30 hour (h) (mammalian

reticulocytes, *in vitro*), 20 h (*yeast*, *in vivo*), and 10 h (*E. coli*, *in vivo*). The instability index was 35.77, indicating that the protein vaccine was stable enough (Table.2)

5.3 Tertiary structure modeling, refinement, and validation

The primary 3D model of the protein vaccine construct was predicted using I-TASSER. The server suggested the top five models for the protein vaccine with C scores ranging from -5 to +2. The C score of the selected model was 2.97, which was the highest score of C. Then, selected I-TASSER model was refined by GalaxyRefine software. The quality of the designed-vaccine was evaluated using the Ramachandran plot in the RAMPAGE server and characteristic atomic interaction in the ERRAT server. The RAMPAGE results of the final model showed that the majority of residues (89.6%) are located in the favored region and 6.4% are in allowed, and only 4% of residues are the outlier. The ERRAT result showed that the refined model has the ERRAT score 87.955 (Fig.2). The outputs of the Ramachandran plot and ERRAT indicated that the refined 3D structure is good and therefore, can be utilized as a reliable model for further evaluations.

6.3 Conformational B-cell epitope Identification

Tertiary structure of the designed vaccine was used as an input for conformational (discontinuous) B cell epitope prediction via Ellipro in IEDB. From 472 amino acid residues, 247 were defined as discontinues B cell epitope.

7.3 Identifying binding sites and protein–protein docking.

ClusPro server has been used to perform the docking of the vaccine with TLR3, 5, and 8 molecules. Cluspro generated thirty models for each of interaction. The docked model was selected based on most atoms interacted in the vaccine and each of receptors (Figures 3 to 5).

8.3 *In silico* codon optimisation of the vaccine construct

The reverse translation of the protein vaccine into a nucleotide sequence was performed simultaneously using the Seipence Manipulation Suite server to express high-level protein in *E. coli*. Codon Adaptation Index (CAI), GC content and Codon Frequency and Distribution (CFD) were evaluated using GenScript online server. A CAI of the vaccine optimized nucleotide sequence was 1. The GC content of the structure was in the ideal range of 55.78% (30-70%), and CFD100 indicating the effective expression of the protein in the host. To clone the gene into *E. coli* vectors, XhoI and EcoRV restriction sites were added into the N- and C-terminals of the sequence using CLC Sequence viewer v.8. Finally, the vaccine gene was inserted into the PET-28 vector (Fig.6).

4. Discussion:

To date, there remains an utmost need to design efficient vaccines in order to halt the inexorable spread of COVID19. In light of recent advances in bioinformatics approaches, we can identify immunodominant T and B cell epitopes and design a potential vaccine to prevent the disease. We used an in silico analysis to design a potent multi-epitope peptide vaccine against SARS-CoV-2. The vaccine contains 472 amino acids which constructed from three specific epitopes from S (receptor-binding domain) and one epitope from each of the structural proteins of E, M, and N. These proteins have essential roles in infection of host cells and host immune modulation. Therefore, it is expected that the vaccine will have a high ability to induce the production of neutralizing antibodies by B lymphocytes and the production of IFN- γ by (T helper) Th and CTL cells.

Adjuvants are essential components of multi-epitope vaccines, because they increase immunologic properties of the vaccine structures. We used Flagellin, a TLR5 agonist, in two N and C terminous of the vaccine construct. Flagellin induces innate immune effectors such as cytokines and nitric oxide[70, 71], activate of TLR5-positive DCs, neutrophils and epithelial cells[72-75] and stimulate the activation of adaptive immune responses mainly Th2-type and IgA production [72, 73, 76-79]. Intranasal administration of Flagellin stimulates the signaling of TLR5 in lung epithelial cells and pneumonocytes [80]. It was used broadly as adjuvant in vaccine structure designed against viral infections such as

influenza [79, 81, 82] and HPV[23, 83]. In addition, we incorporated HP-91 and HBD-3 in the final vaccine construct. The peptide of HP-91 derived from B-box domain amino acids of 89-108 from HMGB1 protein. This peptide induces high levels of IL-2 and IL-15, increases secretion of IL-12 and IFN- α in human DCs and augmented T cell activation[84]. Also, It activates (Dendritic cells) DCs independent of TLR2, 4, and 9, and MyD88 pathways[85]. This is a potent immune adjuvant for inducing cellular and humoral immune responses and used in vaccines against viral infection such as HIV and HPV [84-87]. HBD-3 is the third adjuvant in our vaccine, used as adjuvant to viruses such as influenza and MERS-CoV [26, 88]. This peptide blocks viral fusion using creating a protective barricade of immobilized surface proteins[88]. It activates APCs via TLR1 and TLR2[89] and induces IL-22[90], TGF- α [91, 92] and IFN- γ [93, 94]. It has been implicated in the chemotaxis of immature DCs and T cells via its interaction with chemokine receptor 6 (CCR6) and the chemotaxis of monocytes via interaction with CCR2[95], As well as this peptide activates NK cells, promotes and activates myeloid DCs directly and dependent Natural killer (NK) activity[89, 94]. For linking the different pieces of the vaccine, a repeated of a dipeptide linker of LE was used. This linker improves the expression of vaccine protein via decreasing the pI [96, 97].

An efficient vaccine should not only have stimulating ability immune response but also possess good physicochemical properties during production, formulation, storage and consumption. According to the results of bioinformatics predictions, the designed vaccine was stable with a stability index of 35.77 and had pI of 7.45. Conformational B cell epitopes have a central role in induction humoral responses. The accessible of a significant number of B cell epitopes in the vaccine molecule indicates the high capability of this structure in the stimulation of B lymphocytes. The binding affinity of the vaccine with immune receptor (TLR3, 5 and 8) is necessary to effectively transport vaccine protein into antigen presenting cells. The results of docking analysis showed the vaccine protein properly interact with TLR3, 5, and 8.

5. Conclusion

In this study, using a variety of bioinformatics methods, we developed a multi-epitope subunit vaccine against the SARS-Cov-2 virus. The results of this study showed that it could be possible to predict vaccine candidates against new emerging viral diseases such as COVID-19 with the help of reverse vaccinology. The present study, with its effective vaccine design against SARS-CoV-2, showed that bioinformatics approaches could help to develop effective treatments for other emerging infectious diseases in a short time and at low cost. However, the *in silico* results of this study need to be verified using laboratory and animal models.

6. List of abbreviations

abbreviation	Full name
ACE2	angiotensin-converting enzyme 2
ANN	artificial neural network
CAI	Codon Adaptation Index
CASP10	Critical assessment of techniques for protein structure prediction
CCR	C-C Motif Chemokine Receptor
CFD	Codon Frequency and Distribution
CombLib	Combinatorial Peptide Libraries
CTL	Cytotoxic T lymphocytes
DC	Dendritic cells
E	envelope

E.coli	Escherichia coli
h	hour
HBD-3	human beta defensin 3
HMGB1	high mobility group box1
HLA	Human leukocyte antigen
KDa	Kilo Dalton
INF	Interferon
M	membrane
MHC	Major histocompatibility complex class
MW	molecular weight
N	nucleocapsid
NK	Natural killer
pI	isoelectric point
PSSMs	position specific scoring matrices
QM	Quantitative matrix
S	spike
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SMM	stabilized matrix method
SVM	support vector machine

Th	T helper
TLR	Toll-Like Receptor
3D	Three dimensional

8. Declarations

7.1. Funding

This study received a small grant (Grant no: 7252) from Research and Technology deputy of Mazandaran University of Medical Sciences.

7.2 Authors' contributions

A.R conceived and designed this study; supervised the project and collected the data. Z.Y and M.Y contributed to performing the analysis and writing the draft. A.R and R.V edited the manuscript.

7.3 Ethics approval and consent to participate

The protocol of this study approved by Research Ethics committee of Mazandaran University of Medical Sciences IR.MAZUMS.REC.1398.1430.

7.4 Consent for publication

Not applicable.

7.5 Declarations of interest

The authors declared no conflict of interest.

8. References:

- [1] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395(10223):497-506.

- [2] Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet* 2020;395(10223):470-3.
- [3] Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020;395(10224):565-74.
- [4] Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses* 2020;12(3):254.
- [5] Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020:1-4.
- [6] Stein RA. The 2019 Coronavirus: Learning Curves, Lessons, and the Weakest Link. *Int J Clin Pract* 2020:e13488.
- [7] Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *J. Virol* 2020.
- [8] Li F. Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol* 2016;3:237-61.
- [9] Bosch BJ, van der Zee R, de Haan CA, Rottier PJ. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol* 2003;77(16):8801-11.
- [10] Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Sci* 2020;367(6483):1260-3.
- [11] Nelson GW, Stohlman SA, Tahara SM. High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA. *J. Gen. Virol* 2000;81(1):181-8.
- [12] Stohlman S, Baric R, Nelson G, Soe L, Welter L, Deans R. Specific interaction between coronavirus leader RNA and nucleocapsid protein. *J. Virol* 1988;62(11):4288-95.
- [13] Cong Y, Ulasli M, Schepers H, Mauthe M, V'kovski P, Kriegenburg F, et al. Nucleocapsid protein recruitment to replication-transcription complexes plays a crucial role in coronaviral life cycle. *J. Virol* 2020;94(4).
- [14] Kang S, Yang M, Hong Z, Zhang L, Huang Z, Chen X, et al. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *bioRxiv* 2020.
- [15] Li CK-f, Wu H, Yan H, Ma S, Wang L, Zhang M, et al. T cell responses to whole SARS coronavirus in humans. *J. Immunol* 2008;181(8):5490-500.
- [16] Westerbeck JW, Machamer CE. The infectious bronchitis coronavirus envelope protein alters Golgi pH to protect the spike protein and promote the release of infectious virus. *J. virol* 2019;93(11):e00015-19.
- [17] Gupta MK, Vemula S, Donde R, Gouda G, Behera L, Vadde R. In-silico approaches to detect inhibitors of the human severe acute respiratory syndrome coronavirus envelope protein ion channel. *J. Biomol. Struct. Dyn* 2020(just-accepted):1-17.
- [18] Song Z, Yang Y, Wang L, Wang K, Ran L, Xie Y, et al. EIF4A2 interacts with the membrane protein of transmissible gastroenteritis coronavirus and plays a role in virus replication. *Res. Vet. Sci* 2019;123:39-46.
- [19] Control ECfDPa. COVID-19, <http://mdanderson.libanswers.com/faq/26219>; 2020 [accessed April 17, 2020]
- [20] Pang J, Wang MX, Ang IYH, Tan SHX, Lewis RF, Chen JI-P, et al. Potential rapid diagnostics, vaccine and therapeutics for 2019 novel coronavirus (2019-nCoV): a systematic review. *J. Clin. Med* 2020;9(3):623.

- [21] ul Qamar MT, Alqahtani SM, Alamri MA, Chen L-L. Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. *J. Pharm. Anal* 2020; 26.
- [22] Robson B. Computers and viral diseases. Preliminary bioinformatics studies on the design of a synthetic vaccine and a preventative peptidomimetic antagonist against the SARS-CoV-2 (2019-nCoV, COVID-19) coronavirus. *Comput. Biol. Chem* 2020:103670.
- [23] Yazdani Z, Rafiei A, Valadan R, Ashrafi H, Pasandi M, Kardan M. Designing a potent L1 protein-based HPV peptide vaccine; a bioinformatics approach. *Computational Biology and Chemistry* 2020:107209.
- [24] Ahmad B, Ashfaq UA, Rahman M-u, Masoud MS, Yousaf MZ. Conserved B and T cell epitopes prediction of ebola virus glycoprotein for vaccine development: an immuno-informatics approach. *Microb. Pathog* 2019;132:243-53.
- [25] Shahid F, Ashfaq UA, Javaid A, Khalid H. Immunoinformatics guided rational design of a next generation multi epitope based peptide (MEBP) vaccine by exploring Zika virus proteome. *Infect. Genet. Evol* 2020;80:104199.
- [26] Srivastava S, Kamthania M, Singh S, Saxena AK, Sharma N. Structural basis of development of multi-epitope vaccine against middle east respiratory syndrome using in silico approach. *Infect. Drug Resist* 2018;11:2377.
- [27] Ul Qamar MT, Saleem S, Ashfaq UA, Bari A, Anwar F, Alqahtani S. Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. *J Transl Med* 2019;17(1):362.
- [28] ul Qamar MT, Rehman A, Ashfaq UA, Awan MQ, Fatima I, Shahid F, et al. Designing of a next generation multi-epitope based vaccine (MEV) against SARS-COV-2: Immunoinformatics and in silico approaches. *BioRxiv* 2020.
- [29] Behbahani M. In silico Design of novel Multi-epitope recombinant Vaccine based on Coronavirus surface glycoprotein. *bioRxiv* 2020.
- [30] Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, et al. Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res* 2008;36(suppl_2):W513-W8.
- [31] Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res* 2019;47(D1):D339-D43.
- [32] Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* 2016;32(4):511-7.
- [33] Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res* 2008;36(suppl_2):W509-W12.
- [34] Lundegaard C, Nielsen M, Lund O. The validity of predicted T-cell epitopes. *Trends Biotechnol* 2006;24(12):537-8.
- [35] Lundegaard C, Lund O, Nielsen M. Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. *Bioinformatics* 2008;24(11):1397-8.
- [36] Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* 2003;12(5):1007-17.
- [37] Buus S, Lauemøller S, Worning P, Kesmir C, Frimurer T, Corbet S, et al. Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. *Tissue antigens* 2003;62(5):378-84.
- [38] Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC bioinform* 2005;6(1):132.

- [39] Kim Y, Sidney J, Pinilla C, Sette A, Peters B. Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior. *BMC bioinform* 2009;10(1):394.
- [40] Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, et al. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunome Res* 2008;4(1):2.
- [41] Moutaftsi M, Peters B, Pasquetto V, Tschärke DC, Sidney J, Bui H-H, et al. A consensus epitope prediction approach identifies the breadth of murine T CD8⁺-cell responses to vaccinia virus. *Nat. Biotechnol* 2006;24(7):817-9.
- [42] Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol* 2017;199(9):3360-8.
- [43] Nielsen M, Andreatta M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med* 2016;8(1):33.
- [44] Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* 2009;61(1):1.
- [45] Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and-B locus protein of known sequence. *PloS one* 2007;2(8).
- [46] Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics* 2012;64(3):177-86.
- [47] Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. *Bioinformatics* 2009;25(10):1293-9.
- [48] Rasmussen M, Fenoy E, Harndahl M, Kristensen AB, Nielsen IK, Nielsen M, et al. Pan-specific prediction of peptide-MHC class I complex stability, a correlate of T cell immunogenicity. *J. Immunol* 2016;197(4):1517-24.
- [49] Bhasin M, Raghava G. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine* 2004;22(23-24):3195-204.
- [50] Nielsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC bioinform* 2009;10(1):296.
- [51] Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC bioinform* 2007;8(1):238.
- [52] Sturniolo T, Bono E, Ding J, Radrizzani L, Tuereci O, Sahin U, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat. Biotechnol* 1999;17(6):555-61.
- [53] Andreatta M, Karosiene E, Rasmussen M, Stryhn A, Buus S, Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics* 2015;67(11-12):641-50.
- [54] Reche PA, Glutting J-P, Zhang H, Reinherz EL. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. *Immunogenetics* 2004;56(6):405-19.
- [55] Emini EA, Hughes JV, Perlow D, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J. Virol* 1985;55(3):836-9.
- [56] Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 2017;45(W1):W24-W9.
- [57] Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. *Biology direct* 2013;8(1):30.

- [58] Magnan CN, Baldi P. SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. *Bioinformatics* 2014;30(18):2592-7.
- [59] Agostini F, Cirillo D, Livi CM, Ponti RD, Tartaglia GG. ccSOL omics: a webserver for large-scale prediction of endogenous and heterologous solubility in *E. coli*. *Bioinformatics* 2014;30(20):2975-7.
- [60] Doytchinova IA, Flower DR. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine* 2007;25(5):856-66.
- [61] Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. *The proteomics protocols handbook*. Springer; 2005, p. 571-607.
- [62] Ko J, Park H, Heo L, Seok C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res* 2012;40(W1):W294-W7.
- [63] Lovell SC, Davis IW, Arendall III WB, De Bakker PI, Word JM, Prisant MG, et al. Structure validation by α geometry: ϕ , ψ and $C\beta$ deviation. *Proteins* 2003;50(3):437-50.
- [64] Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein sci* 1993;2(9):1511-9.
- [65] Lill MA, Danielson ML. Computer-aided drug design platform using PyMOL. *J. Comput. Aided Mol. Des* 2011;25(1):13-9.
- [66] Ponomarenko J, Bui H-H, Li W, Fussedder N, Bourne PE, Sette A, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC bioinform* 2008;9(1):514.
- [67] Vajda S, Yueh C, Beglov D, Bohnuud T, Mottarella SE, Xia B, et al. New additions to the ClusPro server motivated by CAPRI. *Proteins* 2017;85(3):435-44.
- [68] Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. *Nat. Protoc* 2017;12(2):255.
- [69] Kozakov D, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, et al. How good is automated protein docking? *Proteins* 2013;81(12):2159-66.
- [70] Mizel SB, Honko AN, Moors MA, Smith PS, West AP. Induction of macrophage nitric oxide production by Gram-negative flagellin involves signaling via heteromeric Toll-like receptor 5/Toll-like receptor 4 complexes. *J. Immunol* 2003;170(12):6217-23.
- [71] Moors MA, Li L, Mizel SB. Activation of interleukin-1 receptor-associated kinase by gram-negative flagellin. *Infect. Immun* 2001;69(7):4424-9.
- [72] Didierlaurent A, Ferrero I, Otten LA, Dubois B, Reinhardt M, Carlsen H, et al. Flagellin promotes myeloid differentiation factor 88-dependent development of Th2-type response. *J. Immunol* 2004;172(11):6922-30.
- [73] Lee SE, Kim SY, Jeong BC, Kim YR, Bae SJ, Ahn OS, et al. A bacterial flagellin, *Vibrio vulnificus* FlaB, has a strong mucosal adjuvant activity to induce protective immunity. *Infect. Immun* 2006;74(1):694-702.
- [74] Han S, Yang K, Ozen Z, Peng W, Marinova E, Kelsoe G, et al. Enhanced differentiation of splenic plasma cells but diminished long-lived high-affinity bone marrow plasma cells in aged mice. *J. Immunol* 2003;170(3):1267-73.
- [75] Pino O, Martin M, Michalek SM. Cellular mechanisms of the adjuvant activity of the flagellin component FljB of *Salmonella enterica* Serovar Typhimurium to potentiate mucosal and systemic responses. *Infect. Immun* 2005;73(10):6763-70.
- [76] Honko AN, Mizel SB. Mucosal administration of flagellin induces innate immunity in the mouse lung. *Infect. Immun* 2004;72(11):6676-9.

- [77] Cunningham AF, Khan M, Ball J, Toellner KM, Serre K, Mohr E, et al. Responses to the soluble flagellar protein FliC are Th2, while those to FliC on Salmonella are Th1. *Eur. J. Immunol* 2004;34(11):2986-95.
- [78] McSorley SJ, Ehst BD, Yu Y, Gewirtz AT. Bacterial flagellin is an effective adjuvant for CD4+ T cells in vivo. *J. Immunol* 2002;169(7):3914-9.
- [79] Wang B-Z, Quan F-S, Kang S-M, Bozja J, Skountzou I, Compans RW. Incorporation of membrane-anchored flagellin into influenza virus-like particles enhances the breadth of immune responses. *J. Virol* 2008;82(23):11813-23.
- [80] López-Boado YS, Cobb LM, Deora R. Bordetella bronchiseptica flagellin is a proinflammatory determinant for airway epithelial cells. *Infect. Immun* 2005;73(11):7525-34.
- [81] Huleatt JW, Nakaar V, Desai P, Huang Y, Hewitt D, Jacobs A, et al. Potent immunogenicity and efficacy of a universal influenza vaccine candidate comprising a recombinant fusion protein linking influenza M2e to the TLR5 ligand flagellin. *Vaccine* 2008;26(2):201-14.
- [82] Skountzou I, del Pilar Martin M, Wang B, Ye L, Koutsonanos D, Weldon W, et al. Salmonella flagellins are potent adjuvants for intranasally administered whole inactivated influenza vaccine. *Vaccine* 2010;28(24):4103-12.
- [83] Negahdaripour M, Eslami M, Nezafat N, Hajighahramani N, Ghoshoon MB, Shoolian E, et al. A novel HPV prophylactic peptide vaccine, designed by immunoinformatics and structural vaccinology approaches. *Infect. Genet. Evol* 2017;54:402-16.
- [84] Telusma G, Datta S, Mihajlov I, Ma W, Li J, Yang H, et al. Dendritic cell activating peptides induce distinct cytokine profiles. *Int. Immunol* 2006;18(11):1563-73.
- [85] Saenz R, Futralan D, Leutenez L, Eekhout F, Fecteau JF, Sundelius S, et al. TLR4-dependent activation of dendritic cells by an HMGB1-derived peptide adjuvant. *J Transl Med* 2014;12(1):211.
- [86] Talebi S, Bolhassani A, Sadat SM, Vahabpour R, Agi E, Shahbazi S. Hp91 immunoadjuvant: An HMGB1-derived peptide for development of therapeutic HPV vaccines. *Biomed. Pharmacother* 2017;85:148-54.
- [87] Milani A, Bolhassani A, Shahbazi S, Motevalli F, Sadat SM, Soleymani S. Small heat shock protein 27: An effective adjuvant for enhancement of HIV-1 Nef antigen-specific immunity. *Immunol. Lett* 2017;191:16-22.
- [88] Leikina E, Delanoe-Ayari H, Melikov K, Cho M-S, Chen A, Waring AJ, et al. Carbohydrate-binding molecules inhibit viral fusion and entry by crosslinking membrane glycoproteins. *Nat. Immunol* 2005;6(10):995-1001.
- [89] Funderburg N, Lederman MM, Feng Z, Drage MG, Jadowsky J, Harding CV, et al. Human β -defensin-3 activates professional antigen-presenting cells via Toll-like receptors 1 and 2. *Proc. Natl. Acad. Sci* 2007;104(47):18631-5.
- [90] Wolk K, Kunz S, Witte E, Friedrich M, Asadullah K, Sabat R. IL-22 increases the innate immunity of tissues. *Immunity* 2004;21(2):241-54.
- [91] Sørensen OE, Cowland JB, Theilgaard-Mönch K, Liu L, Ganz T, Borregaard N. Wound healing and expression of antimicrobial peptides/polypeptides in human keratinocytes, a consequence of common growth factors. *J. Immunol* 2003;170(11):5583-9.
- [92] Ferris LK, Mburu YK, Mathers AR, Fluharty ER, Larregina AT, Ferris RL, et al. Human beta-defensin 3 induces maturation of human Langerhans cell-like dendritic cells: an antimicrobial peptide that functions as an endogenous adjuvant. *J. Invest. Dermatol* 2013;133(2):460-8.
- [93] Joly S, Organ CC, Johnson GK, McCray Jr PB, Guthmiller JM. Correlation between β -defensin expression and induction profiles in gingival keratinocytes. *Mol. Immunol* 2005;42(9):1073-84.

- [94] Judge CJ, Reyes-Aviles E, Conry SJ, Sieg SS, Feng Z, Weinberg A, et al. HBD-3 induces NK cell activation, IFN- γ secretion and mDC dependent cytolytic function. *Cell. Immunol* 2015;297(2):61-8.
- [95] Röhrl J, Yang D, Oppenheim JJ, Hehlhans T. Human β -defensin 2 and 3 and their mouse orthologs induce chemotaxis through interaction with CCR2. *J. Immunol* 2010;184(12):6688-94.
- [96] Chen X, Zaro JL, Shen W-C. Fusion protein linkers: property, design and functionality. *Adv. Drug Deliv. Rev* 2013;65(10):1357-69.
- [97] Chen X, Lee H-F, Zaro JL, Shen W-C. Effects of receptor binding on plasma half-life of bifunctional transferrin fusion proteins. *Mol. Pharm* 2011;8(2):457-65.

Figure legend:

Figure 1, Schematic representation of the designed multi-epitope peptide based-vaccine. The vaccine consist of ten parts: Epitopes from structural proteins S, E, N, and M, adjuvants: Flagellin (in N-and C-terminus), HP-91 and HBD-3 that join to each other by linkers of repeat sequence of LE (A figure). Tertiary structure of the modeled multi-epitope vaccine construct (figure B). The 3D structure of the designed vaccine was predicted via homology modeling by I-Tasser, then the best predicted model was refined by Galaxy Refine and visualized using Pymol software. N-and C-terminus of Flagellinis shown in green, HP-91 in black, S epitopes in blue, E epitopes in purple, N epitope in orange, M epitope in yellow, HBD-3 in green, and linkers are shown in light pink.

FIG.2, Validation results of the refined 3D structure of multi-epitope vaccine structure by ERRAT software. ERRAT plot showed the overall quality score of the refined structure as 87.955.

Fig. 3. Docking model (cartoon representation) of human TLR3 in complex with the vaccine obtained using Cluspro server. TLR3 protein is shown in limon, Flagellin in green, HBD-3 in tv-green, S epitopes in white, M epitope in orange, N epitope in purple, E epitope in red, HP-91in cyan and linkers are shown in light pink. As it shows some parts of two S, M, N, and E epitopes, HBD-3, and HP-91 interacted with TLR3. To more visualized interaction points, some of the interacting residues of the vaccine and TLR3 are magnified in 20 Angstrom. Docked model was visualized via Pymol software.

Fig. 4. Docking model (cartoon representation) of human TLR5 in complex with the vaccine obtained using Cluspro server. TLR5 protein is shown in limon, Flagellin in green, HBD-3 in tv-green, S epitopes

in white, M epitope in orange, N epitope in purple, E epitope in red, HP-91 in cyan and the linkers are shown in light pink. As the figure shows some parts of Flagellin, S-epitope and HP-91 interacted with TLR5. In order to more visualized interaction points, some of the interacting residues of the vaccine and TLR5 are magnified in 20 Angstrom. Docked model was visualized by Pymol software.

Fig. 5. Docking model (cartoon representation) of human TLR8 in complex with the vaccine obtained using Cluspro server. TLR8 protein is shown in limon, Flagellin in green, HBD-3 in tv-green, S epitopes in white, M epitope in orange, N epitope in purple, E epitope in red, HP-91 in cyan, and linkers are shown in light pink. As the figure shows some parts of S, M, N, and E epitopes, HBD-3 and HP-91 interacted with TLR8. In order to more visualized interaction points, some of the interacting residues of the vaccine and TLR8 are magnified in 20 Angstrom. Docked model was visualized using Pymol software.

Fig.6. Evaluation of the three important parameters of the codon-optimized gene to express high level protein in E.coli. A) CAI of the gene sequence was 1. It is noted that a CAI of > 0.8 will be considered as good for expression in the selected host. B) Average GC content of the sequence was 55.78%. C) Codon frequency distribution (CFD) value of the gene sequence is 100. A CFD equal to 100 supports maximum protein expression in the desired host. D) Insertion of the vaccine gene in the PET28 vector by restriction enzymes of XhoI and EcoRV.

Tables:

Table.1 The final immunodominant epitopes were selected from S, M, N, and E proteins of SARS-CoV-2

Table.2 Analysis of the physicochemical and immunological properties of the SARS-CoV-2 vaccine

Table.3: Conformational B-cell epitopes identified in the refined tertiary structure of the multi-epitope vaccine using Ellipro tool.

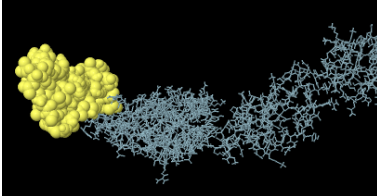
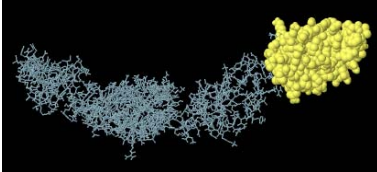
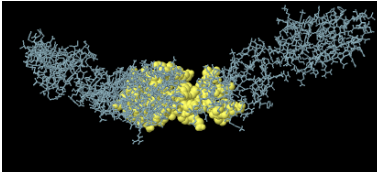
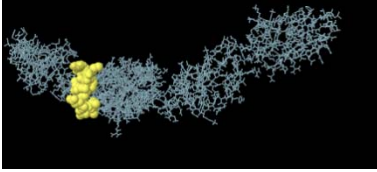
Table1:

Protein	Sequence	Number of aminoacids	HLA-I (Netmhc4-IEDB)	HLA-II (RANKP EP/IED)	CTL (CTLPRE D)	Linear B cell (BEPIPRED/IEDB)	Antigenicity	IFNstimulation	Allergenicity
Spike glycoprotein [SARS-CoV-2] (QIC53213.1)	QCVNLTTRTQLPPAY TNSFTRGVYYPDKVF	20-49	✓	✓	✓	✓	✓ 0.6331	+1	Probable allergen non-
	KIADYNYKLPDDFTG CVIAWNSNNLDSKVG	418-446	✓	✓	✓	✓	0.7009	+5.0777656	Probable allergen non-
	YQPYRVVVLSEFLLH APATVCGPKKSTNLV	505-534	✓	✓	✓	✓	✓0.5147	+4.9606125	Probable allergen non-
Envelope protein [SARS-CoV-2] (QIC53215.1)	LVKPSFYV YSRVKNLNSSRVP	51-71	✓	✓	✓	✓	✓ 0.5468	+1.3230951	Probable allergen non-
Nucleocapsid phosphoprotein [SARS-CoV-2] (QHU79211.1)	YKTFPPTEPKKDKKK KADET QALPQRQKK	360-388	✓	✓	✓	✓	✓0.5844	+3.3604363	Probable allergen non-
Membrane glycoprotein [SARS-CoV-2] (QIC53207.1)	RYRIGNYKLNTDHSS SS	198-214	✓	✓	✓	✓	✓1.2640	0.53643062	Probable allergen non-

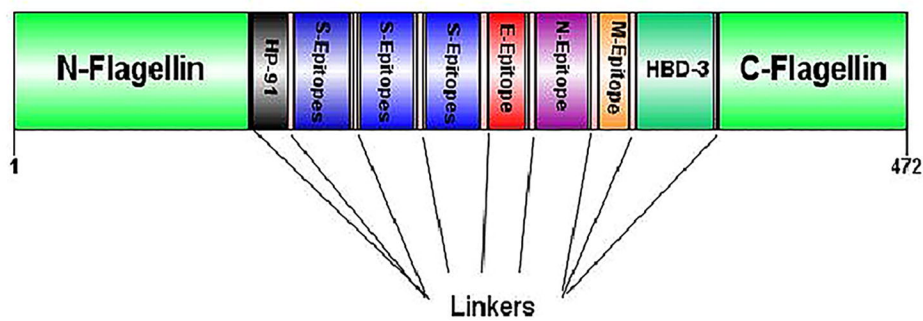
Table 2:

Physicochemical properties	Value
Molecular Weight (Da)	51916.54
Instability index	35.77
Gravy	-0.458
Aliphatic index	87.03
Theoretical pI	7.45
No. amino acids	472
Total no. of negatively charged residues (Asp+Glu)	53
Total no. of positively charged residues (Arg+Lys)	54
No. of atoms	7308
Antigenicity	
Antigenicity / ANTIGENpro	0.941411
Antigenicity /vaxijen	0.5319 (Probable antigen)
Allergenicity / AllergenFP v.1.0	Probable non-allergen

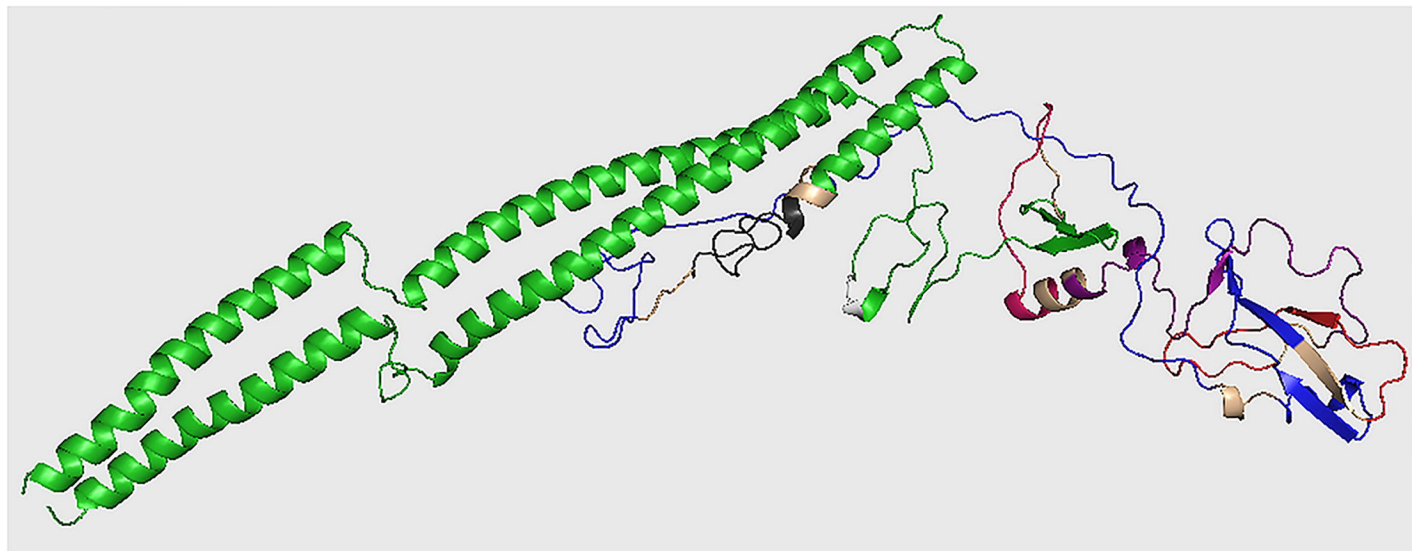
Table 3:

Residues	Number of residues	Score	3D structure
1 A:M1, A:A2, A:Q3, A:V4, A:I5, A:N6, A:T7, A:N8, A:S9, A:L10, A:S11, A:L12, A:L13, A:T14, A:Q15, A:N16, A:N17, A:L18, A:N19, A:K20, A:S21, A:Q22, A:S23, A:S24, A:L25, A:S26, A:S27, A:A28, A:I29, A:E30, A:R31, A:L32, A:S33, A:S34, A:G35, A:L36, A:D435, A:A436, A:D437, A:Y438, A:A439, A:T440, A:E441, A:V442, A:S443, A:N444, A:M445, A:S446, A:K447, A:A448, A:Q449, A:I450, A:L451, A:Q452, A:Q453, A:A454, A:G455, A:T456, A:S457, A:V458, A:L459, A:A460, A:Q461, A:A462, A:N463, A:Q464, A:V465, A:P466, A:Q467, A:N468, A:V469, A:L470, A:S471, A:L472, A:L473	75	0.801	
2 A:N207, A:L208, A:D209, A:S210, A:K211, A:V212, A:G213, A:L214, A:E215, A:L216, A:E217, A:Y218, A:Q219, A:P220, A:Y221, A:R222, A:V223, A:V224, A:V225, A:L226, A:S227, A:F228, A:E229, A:L230, A:L231, A:H232, A:A233, A:P234, A:A235, A:T236, A:V237, A:C238, A:G239, A:P240, A:K241, A:K242, A:S243, A:T244, A:N245, A:L246, A:V247, A:L248, A:E249, A:L250, A:E251, A:L252, A:V253, A:K254, A:P255, A:S256, A:F257, A:Y258, A:V259, A:Y260, A:S261, A:R262, A:V263, A:K264, A:N265, A:L266, A:N267, A:S268, A:S269, A:R270, A:V271, A:P272, A:L273, A:E274, A:L275, A:E276, A:Y277, A:K278, A:T279, A:F280, A:P281, A:P282, A:T283, A:E284, A:P285, A:K286, A:K287, A:D288, A:K289, A:K291, A:K292, A:A293, A:D294, A:E295, A:T296, A:Q297	90	0.767	
3 A:Q90, A:V92, A:R93, A:E94, A:L95, A:S96, A:V97, A:Q98, A:A99, A:T100, A:N101, A:G102, A:T103, A:N104, A:S105, A:D106, A:S107, A:D108, A:L109, A:K110, A:S111, A:I112, A:E115, A:E181, A:L182, A:K184, A:A186, A:D187, A:Y188, A:N189, A:Y190, A:K191, A:L192, A:P193, A:D194, A:D195, A:F196, A:T197, A:L318, A:N319, A:T320, A:D321, A:H322, A:S323, A:S324, A:S325, A:S326, A:G333, A:D334, A:P335, A:A379, A:A380, A:K381, A:K382, A:S383, A:T384, A:A385, A:N386, A:P387, A:L388, A:A389, A:S390, A:I391, A:D392, A:S393, A:A394, A:L395, A:S396, A:K397, A:D399, A:A400, A:S4	72	0.619	
4 A:R37, A:I38, A:N39, A:S40, A:A41, A:K42, A:D43, A:D44, A:A45, A:A46	10	0.575	

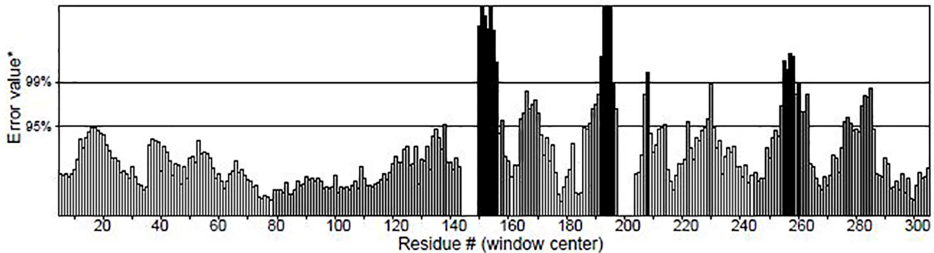
A)



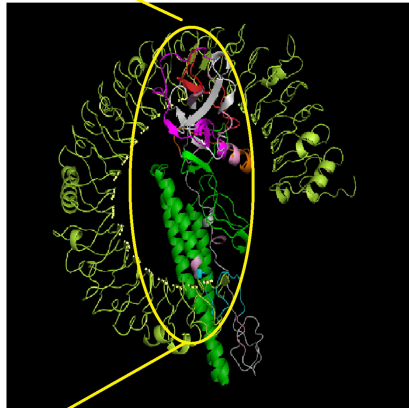
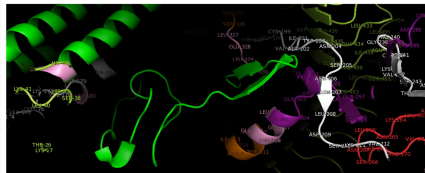
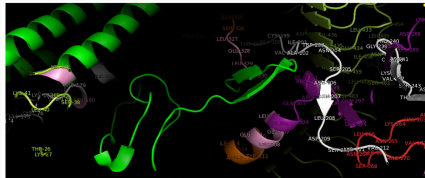
B)



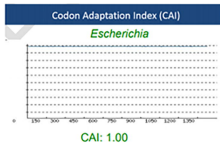
Overall quality factor**: 87.955



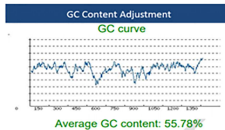
*On the error axis, two lines are drawn to indicate the confidence with



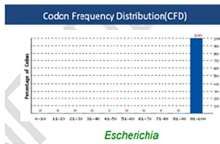
A)



B)



C)



D)

