Designing a New Multiepitope-based Vaccine Against COVID-19

Afshin Samimi Nemati[1] (iD), **Majid Tafrihi**[1*] (iD), **Fatemeh Sheikhi**[1] (iD), **Abolfazl Rostamian Tabari**[1] (iD), **Amirhossein Haditabar**[1] (iD)

1. Department of Molecular and Cell Biology, Faculty of Basic Sciences, University of Mazandaran, Babolsar, Iran.

ABSTRACT

Background: Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has currently caused a significant pandemic among worldwide populations. The high transmission and mortality rates of the disease necessitate studies for rapid designing and effective vaccine production. This study aims to predict and design a novel multi-epitope vaccine against the SARS-CoV-2 virus using bioinformatics approaches.

Materials and Methods: Coronavirus envelope proteins, Open Reading Frame 7b (ORF7b), Open Reading Frame 8 (ORF8), Open Reading Frame 10 (ORF10), and Nonstructural protein 9 (Nsp9) were selected as targets for epitope mapping using Immune Epitope Data Bank (IEDB) and BepiPred 2.0 servers. Also, molecular docking studies were performed to determine the candidate vaccine's affinity to Toll-Like Receptor (TLR3, TLR4) and Major Histocompatibility Complex (MHC I and MHC II) molecules. Thirteen epitopes were selected to construct the multi-epitope vaccine.

Results: We found that the constructed peptide has valuable antigenicity, stability, and appropriate half-life. The Ramachandran and ERRAT plots approved the quality of the predicted model after the refinement process. Molecular docking investigations revealed that antibody-mode in the ClusPro 2.0 server showed the lowest binding energy for MHC I, MHC II, TLR3, and TLR4.

Conclusion: The designed vaccine has a good antigenicity and stability and could be a proper vaccine candidate against the Coronavirus Disease 2019 (COVID-19) infectious disease though, in vitro and in vivo experiments are necessary to complete and confirm our results.

Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is the zoonotic and causative agent behind the COVID-19 pandemic [1, 2]. This outbreak was first observed in December 2019 in Wuhan, China, and has since caused the death of millions of people worldwide [3]. The Coronavirus 2019 (COVID-19) infection is rapidly transmissible among people worldwide [4]. The SARS-CoV-2 genome is a non-segmented positive-sense RNA of 29903 nucleotides [5] with ~80% similarity to the genome of the Severe Acute Respiratory Syndrome (SARS-CoV-1) and

* Corresponding Author:
Majid Tafrihi, PhD.
Address: Department of Molecular and Cell Biology, Faculty of Basic Sciences, University of Mazandaran, Babolsar, Iran.
Phone: +98 (11) 35302450
E-mail: m.tafrihi@umz.ac.ir

Middle East Respiratory Syndrome (MERS-CoV) viruses. Those viruses caused global pandemics in 2002 and 2011, respectively [6-9]. The COVID-19 infection shows different symptoms, including fever, tiredness, difficulty breathing, dry cough, chest pain, and pneumonia [10-12]. The largest SARS-CoV-2 gene is *orf1ab*, which encodes two large polyproteins pp1a and pp1ab. The pp1ab is the largest protein in coronaviruses replicating and transcribing the viral genome [13]. The Nonstructural protein 9 (Nsp9) is one of the proteins produced from the processing of pp1a and pp1ab polyproteins. The Nsp proteins are vital for the forming of the Replication-Transcription Complex (RTC) that mediates the synthesis of the nested subgenomic minus-strands of RNA [(–)sgRNA] during the transcription process [14]. It has been shown that the Open Reading Frame 8 (*ORF8*) gene in the SARS-CoV-2 virus is a unique gene. Also, immune evasion could occur during down-regulating the Major HistoCompatibility complex (MHC) I molecule by the ORF8 protein [15].

There are several methods to produce new vaccines, such as live-attenuated form, subunit vaccines, multi-epitope vaccines, and DNA vaccines [16-18]. Recently, most developed or under developing vaccines have focused on virus surface proteins such as spike and nucleocapsid protein. These proteins show good immunity, but some studies have indicated the high immunogenicity of the ORF8 protein [19]. Also, the ORF7b and ORF10 proteins are unique, and they lacked any sequence homology with other viral proteins [20]. Therefore, the ORF7b, ORF8, and ORF10 could be proper candidates to study their immunogenicity.

Toll-Like Receptors (TLRs) have a significant role in innate immunity against viral infections. Pro-inflammatory cytokines, such as Interleukin-1 (IL-1) and IL-6, have been found in innate immunity in COVID-19 patients, resulting from activation of TLRs pathways. Various studies have shown that TLR2, TLR3, TLR4, TLR6, TLR7, TLR8, and TLR9 are potentially crucial in COVID-19 infection [21].

Reverse Vaccinology (RV) is an approach that utilizes the expressed genomic sequences to find new potential vaccines [22]. This method significantly reduces the time required for identifying candidate vaccines through using various servers and databases to identify candidate proteins and efficient epitope(s) [23]. Main immunoinformatics approaches have improved this method by creating several databases and algorithms for epitope prediction, which increases the speed and accuracy of vaccine development [17]. For this purpose, to acquire an appropriate peptide vaccine, the biochemical and physicochemical properties of the constructed peptide are considered. According to the current advances in this field, several servers were established [24].

In the present study, we designed a new multi-epitope peptide vaccine against SARS-CoV-19 infection. Because of rare mutations observed in *ORF7b*, *ORF8*, *ORF10*, and *Nsp9* encoding genes, they were selected for epitope prediction. The envelope protein of the virus was selected to increase the possible chance for humoral immunity. For this purpose, various epitopes have been selected and engineered based on their biochemical and physicochemical properties. Finally, in silico studies were carried out on the bioactivity of the constructed peptide by docking the vaccine with TLRs (TLR3 and TLR4) and Major Histocompatibility Complex (MHC I and MHC II).

Materials and Methods

We took all steps of this study according to Figure 1.

Data collection

To find proteins associated with SARS-CoV-19 infection, we used the NCBI database (available at https://www.ncbi.nlm.nih.gov/protein). Then, the amino acid sequence of the Cholera Toxin Subunit B (CTXB) (GenBank: BBG62270.1) has obtained in the FASTA format from the protein database of the NCBI. Accordingly, the three-dimensional (3D) structure of TLR 3 (2mka), TLR 4 (2z62), MHC I (1akj), and MHC II (4i5b) proteins were obtained from the protein databank (available at https://www.rcsb.org/), in PDB format.

Selecting the protein antigen

Among protein sequences, the five top score proteins of the envelope protein (YP_009724392.1), ORF7b protein (YP_009725318.1), ORF8 protein (YP_009724396.1), ORF10 protein (YP_009725255.1), and Nsp9 protein (YP_009725305.1) were selected as candidates that had the maximum of antigenicity and were non-allergens. Antigenicity and the allergenicity of the selected proteins were analyzed by using the VaxiJen 2.0 server (available at http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) and AllerTOP v. 2.0 (available at https://www.ddg-pharmfac.net/AllerTOP), respectively [24, 25].

B-cell and T-cell epitope prediction

Afterward, the Immune Epitope Data Bank (IEDB) server (available at https://www.iedb.org/) and BepiPred 2.0 (accessible at http://www.cbs.dtu.dk/) were used to
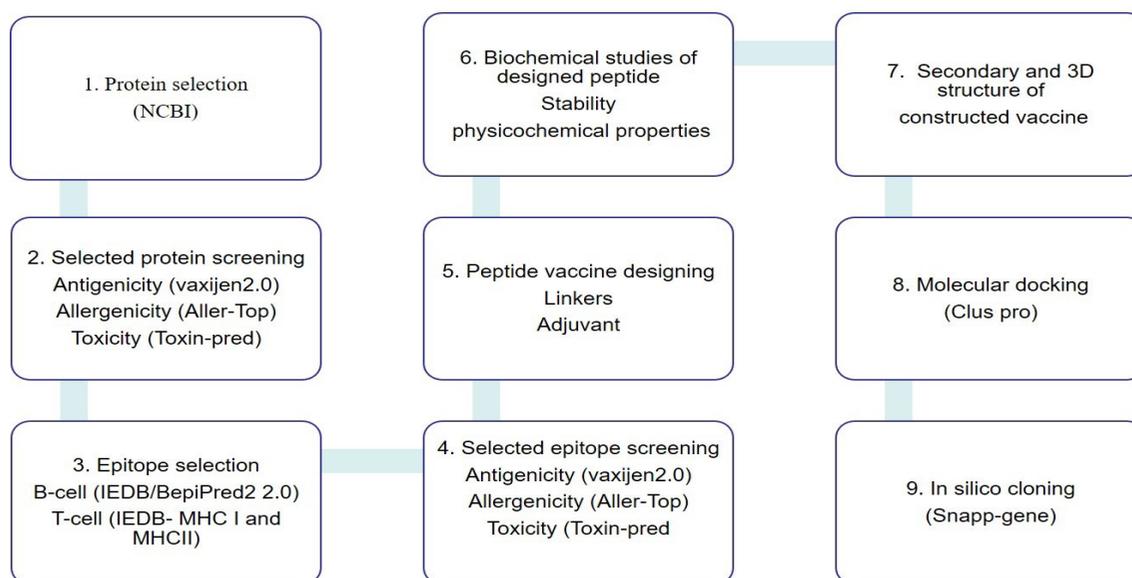
Figure 1. Schematic representation of steps for designing peptide vaccine against SARS-CoV-2

predict the linear B-cell epitopes from the chosen proteins [26, 27]. For the prediction of the T-cell-related epitopes that contain the MHC I and MHC II, the IEDB server was utilized [28]. Antigenicity and allergenicity analyses for the selected epitopes were repeated using VaxiJen 2.0 and AllerTOP v. 2.0 servers. The ToxinPred server (available at https://webs.iiitd.edu.in/raghava/toxinpred/algo.php) was used to predict the toxic regions in the selected epitopes [29].

Peptide designing and adjuvating

For this purpose, by engineering the amino acid sequences of the selected epitopes, a peptide sequence has been constructed. Some of the selected epitope sequences contain both the B- cell and T-cell epitopes that are named B-cell-derived T-cell epitopes. This strategy for epitope selection could reduce the length of the final amino acid sequence. Lysine-lysine (KK) cross-linkers connected the selected epitopes. Usually, protein adjuvants were used to increase the probability of exposing the peptide vaccines. Therefore, the CTXB amino acid sequence was added to the epitope sequences using a proline-rich (PAPAP) rigid linker. Also, the final amino acid sequence of the designed vaccine was used to predict the surface accessibility of the vaccine containing epitopes by the IEDB server. To this aim, the Emini surface accessibility prediction algorithm was used [30]. Finally, discontinuous B cell epitopes of the proposed vaccine were predicted by using the ElliPro (The antibody epitope prediction tool of the IEDB server) [31].

Analyzing the physicochemical properties

In this step, the physicochemical properties of the proposed vaccine are performed. To this end, amino acid composition, molecular weight, isoelectric point, net charge at pH=7, water-solubility, protein half-life in mammalian and bacterial cells were evaluated by using ProtParam (available at http://web.expasy.org/cgi-bin/protparam/protparam), and PepCalc (available at http://pepcalc.com/) databases [32]. Finally, using the IUPred2A server (https://iupred2a.elte.hu/) predicted the protein stability [33].

Structural analysis

The secondary structure of the designed vaccine has been predicted using the GOR4 secondary structure prediction method of the Prabi server (available at https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html). Also, the protein's disulfide bonds were predicted by using DiANNA 1.1 web server, which is accessible at (http://clavius.bc.edu/~clotelab/DiANNA/) [34]. In the next step, the 3D structure of the protein has been predicted by using the C-I-TASSER server (https://zhanglab.ccmb.med.umich.edu) [35, 36]. Finally, the predicted 3D structure of the designed vaccine sequence is visualized by Discovery Studio Client offline software 2017 [37]. The prediction of the 3D structure of the proposed vaccine is required to validate the affinity of binding to the immune system elements in docking studies.

Refinement and molecular docking of the candidate peptide

The refinement process of the predicted model has been performed using the GalaxyWEB server (available at galaxy.seoklab.org/) to reduce possible structural faults [38, 39]. Moreover, the geometric features of the proposed vaccine were validated based on the Ramachandran plot using the PROCHECK (available at https://servicesn.mbi.ucla.edu/PROCHECK/) and ERRAT servers. These servers calculate the overall scores for the predicted models [33, 40]. The molecular docking process was conducted using the ClusPro 2.0 protein-protein docking server (https://cluspro.org/help.php) to ensure the interaction between receptor molecules TLR3, TLR4, MHC I, and MHC II and the refined model of candidate protein as a ligand. The antibody mode of the ClusPro 2.0 server was used for the docking process [41-44].

Back translation, codon optimization, and in silico cloning of the candidate protein

The SnapGene 3.2.1 software was used to back translate the designed peptide into the nucleotide sequence. Also, to optimize the rate of protein expression in Escherichia coli, the JCat online server (http://www.jcat.de/) was utilized [45]. To have the correct translation, the Open Reading Frame (ORF) has been checked using SnapGene 3.2.1 software. Finally, in silico restriction cloning was performed with SnapGene 3.2.1. To this aim, the poly-histidine-tag sequence was fused to the proposed vaccine sequence and then inserted into the multiple cloning site of the PET21 expression vector.

Results

Protein antigen selection

Firstly, protein sequences have aligned with conserved and selected regions. Then, the chosen proteins are compared with each other in terms of antigenicity, allergenicity, and toxicity. Finally, five chosen proteins had high levels of antigenicity and were non-allergen (Table 1).

Linear B-cell and T-cell epitope prediction

To predict the linear B-cell and T-cell epitopes IEDB, and BepiPred2.0 servers have used, and similar epitopes were selected (Table 2). Briefly, MHC I- and MHC II-restricted epitopes were predicted and ranked according to the IEDB scores (Table 3). Some B-cell epitopes have been derived from T-cell epitopes to reduce the final length of the sequence. Finally, epitopes with the highest

antigenicity score and lack of allergenicity and toxicity were selected to construct the peptide vaccine.

Peptide designing and adjuvating

For both B-cell and T-cell, 13 epitopes were selected and then linked together using KK linkers. To increase the immunogenicity of the designed vaccine, an adjuvant sequence was added to the peptide sequence. Various peptide adjuvants were used in previous studies [44]. The amino acid sequence of CTXB, the non-toxic portion of cholera toxin, was added into the initial part of the peptide and connected to the epitopes by PAPAP rigid linker (Figure 2-A). The antigenicity of the proposed vaccine was evaluated as 0.57 using VaxiJen2.0, and the results of the AllerTOP 2.0 server confirmed that the vaccine is not an allergen. The result of the Emini surface accessibility and ElliPro algorithm, used to reveal the surface accessibility and discontinuous B cell epitopes of the final sequence of the designed vaccine, has been shown in Figure 2, parts B, C, D, and E, respectively.

Physicochemical properties of constructed peptide

Physicochemical properties and amino acid composition of the constructed peptide were evaluated using Pep-Calc and ProtParam servers. The results showed that the proposed vaccine was stable, water-soluble, with a molecular weight of 3745.98 g/mol. The calculated pI value was 9.62, the net charge was 17.3, and the estimated half-life of the protein in mammalian, yeast, and *E. coli* cells were 30, 20, and 10 hours, respectively (Table 4). Also, the analysis of protein sequence stability was performed by predicting the protein disorder regions (using the IUPred2A server). The results confirmed the stability of the designed peptide (Figure 3B). The IUPred2A server presents three types of analysis; IUPred2 long disorder, IUPred2 short disorder, and IUPred2 structured domains. In this study, the IUPred2 long disorder mode has been selected for investigation. Protein disorders of the proposed vaccine were predicted by IUPred 2.0 designed graph. According to the designed graph (Figure 3B), the sequence of the proposed vaccine has not had a great chance to be an anchor for binding to the registered structures in the IUPred2A server. In the presented graph, residues with a score above and below 0.5 are considered protein disorders and protein orders, respectively.
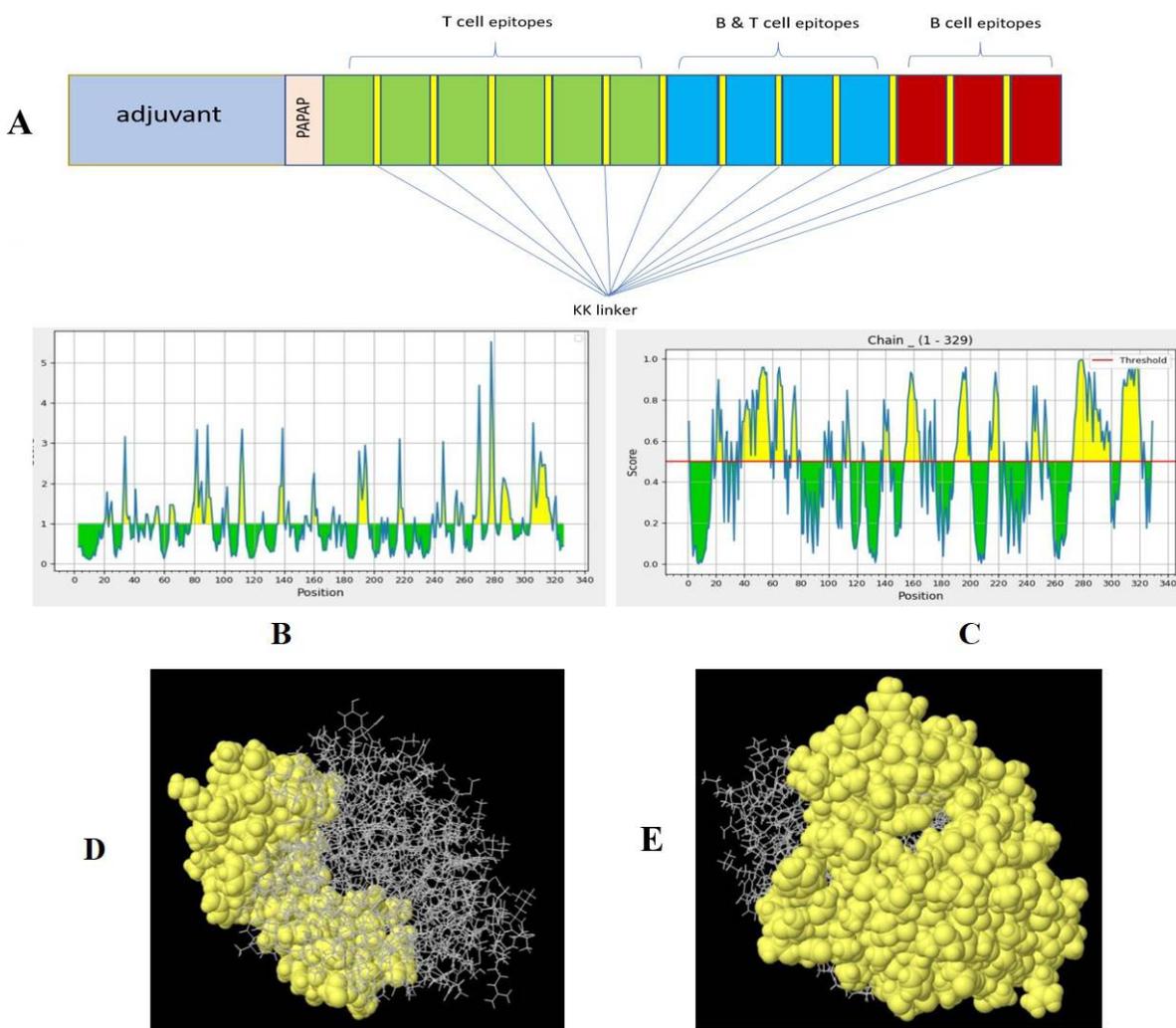
Figure 2. Schematic representation of the designed peptide vaccine

A: The designed sequence consisted of four components, including Cholera Toxin Subunit B (CTXB) as an adjuvant, proline-rich (PAPAP), and lysine-lysine (KK) as linkers, B-cell, and T-cell related epitopes; B: The graph of the emini surface accessibility prediction from IEDB Server, the default designated Threshold of 1.00 and results confirmed the accessibility of the vaccine containing components; C: The visual graph with default threshold of 0.5 of ElliPro: antibody epitope prediction tool of discontinuous B-cell epitopes of the proposed vaccine; D and E: The prediction results of ElliPro: Epitope 3D structures for designed vaccine.

The secondary and tertiary structure of the constructed peptide

The prediction of the secondary structure of the proposed vaccine was performed using the GOR 4 method of the Prabi server (available at https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html). As Figure 4 shows, the secondary structure of the proposed vaccine contains alpha-helix (28.57%), extended strand (19.76%), and random coil (51.67 %). Free cysteine amino acids in the protein structure were predicted, and results showed that only cysteine at position 243 was non-bonding. Results of the prediction of the disulfide bonds are shown in Figure 4-B. Using the C-I-TASSER

server, the tertiary structure of the peptide is predicted. This server suggests various models for the input sequence and the quality of prediction models reflected in C-Scores (-5 to 2). The higher values of the C-Score indicate the higher confidence levels for the predicted model. The predicted model was viewed using the Discovery Studio 2020 Client offline software (Figure 5-A).

Model refinement and molecular docking of the vaccine candidate peptide

The refinement process has been performed for the predicted model using the GalaxyWEB server. In this process, the server refined the secondary structure elements

Table 1. List of protein candidates

Protein	NCBI Reference Sequence	Antigenicity (VaxiJen2.0)	Allergenicity (AllerTOP)	Toxicity (ToxinPred)
Envelope protein (SARS-CoV-2)	YP_009724392.1	0.6025*	Non-allergen	Non-toxin
ORF6 protein (SARS-CoV-2)	YP_009724394.1	0.6131	Non-allergen	Non-toxin
ORF7a protein (SARS-CoV-2)	YP_009724395.1	0.6441	Non-allergen	Non-toxin
ORF8 protein (SARS-CoV-2)	YP_009724396.1	0.6502*	Non-allergen	Non-toxin
Nucleocapsid phosphoprotein (SARS-CoV-2)	YP_009724397.2	0.5059	Non-allergen	Non-toxin
ORF10 protein (SARS-CoV-2)	YP_009725255.1	0.7185*	Non-allergen	Non-toxin
nsp6 (SARS-CoV-2)	YP_009725302.1	0.5813	Non-allergen	Non-toxin
nsp9 (SARS-CoV-2)	YP_009742616.1	0.6476	Non-allergen	Non-toxin
nsp6 (SARS-CoV-2)	YP_009742613.1	0.5813	Non-allergen	Non-toxin
ORF7b (SARS-CoV-2)	YP_009725318.1	0.8462*	Non-allergen	Non-toxin
endoRNAse (SARS-CoV-2)	YP_009725310.1	0.5554	Non-allergen	Non-toxin
nsp9 (SARS-CoV-2)	YP_009725305.1	0.6476*	Non-allergen	Non-toxin

The sequences were retrieved from the NCBI database, and then their antigenicity, allergenicity, and toxicity were analyzed.

*Marks the selected proteins.

like loop regions and side chains based on several factors containing similarity score (GDT-HA), clash score, Root-Mean-Square Deviation (RMSD), and MolProbity. The Global Distance Test (GDT_TS) compares two protein structures with known amino acid correspondences with a different tertiary structure. The GDT-HA is a high-accuracy version of GDT_TS, which selects smaller cut-off distances that were half of the size of GDT_TS and thus is more rigorous. RMSD is the average distance between backbone atoms in the protein structure. The Mol-Probity score reflects the crystallographic resolution. A structure with a numerically lower MolProbity score than its actual crystallographic resolution is, quality-wise, better than the average structure at that resolution.

Five suggested refined models resulted from the refinement process for the 3D predicted model of the proposed vaccine have been shown in Table 5.

The second presented model with GDT-HA and RMSD of 0.9878 and 0.282 was selected for further considerations. However, the clashscore for the chosen model was 19.3, and the reported score for the initial model was about 22.4. The Rama favored scores were the other score that indicates the percentage of the residues in the most favored regions of the Ramachandran plot. This score changed from 64.9% to 89.6%, simultaneously for the refined model compared to the initial model. The 3D

Table 2. Prediction of linear B-cell epitope

Epitope	Protein	Antigenicity	Allergenicity
ELQDHNE	ORF10	0.9331	Non-allergen
DEAGSKSPIQYIDIGN	ORF8	0.9110	Non-allergen
AGTTQTACTDDNALAYYNTTKGG	Nsp9	0.5741	Non-allergen
IYSLLLCRMNSRNYI	Envelope protein	0.6011	Non-allergen
VRCSFYEDFLEYHDV	ORF8	0.7225	Non-allergen

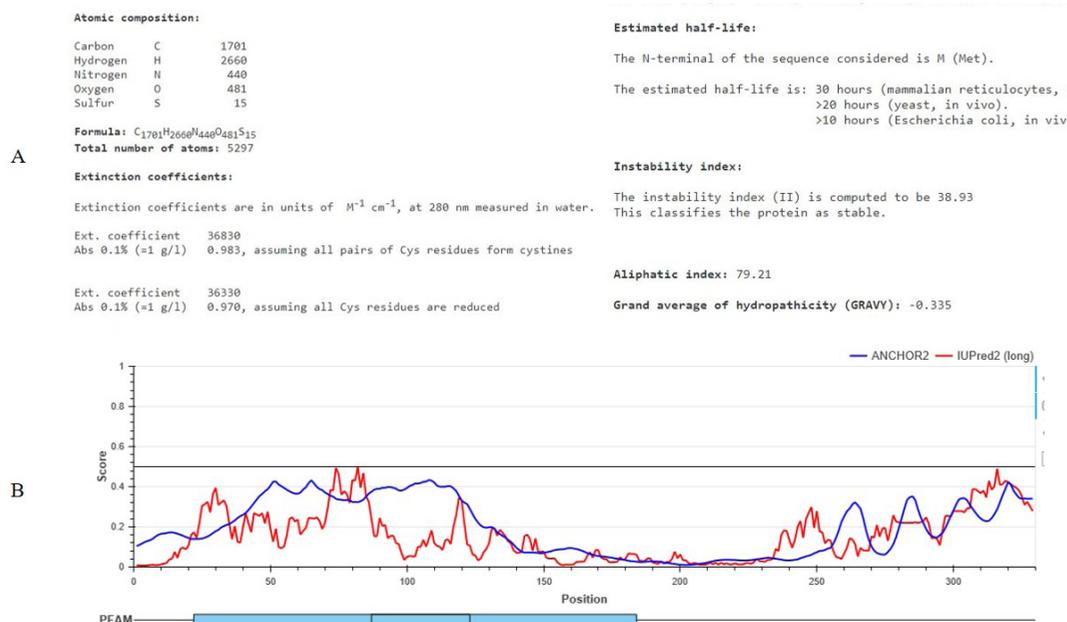Similar epitopes were selected by servers.

Figure 3. Physiochemical analyses of designed peptide vaccine

A: Physicochemical properties of the constructed peptide by the ProtParam Server, analysis of the amino acid composition and estimation of the half-life of the protein using the ProtParam database, data showing the estimated half-life of 30 hours in mammalian reticulocytes, the Instability Index (II) was measured as 38.93 indicating the protein stability; B: The stability of protein analyzed using IUPred 2a by studying the protein disorders of the peptide sequence, in this graph, the blue line indicating the ANCHOR2, and the red line the IUPred2, according to the red line on the presented graph, the results confirmed the stability of the protein
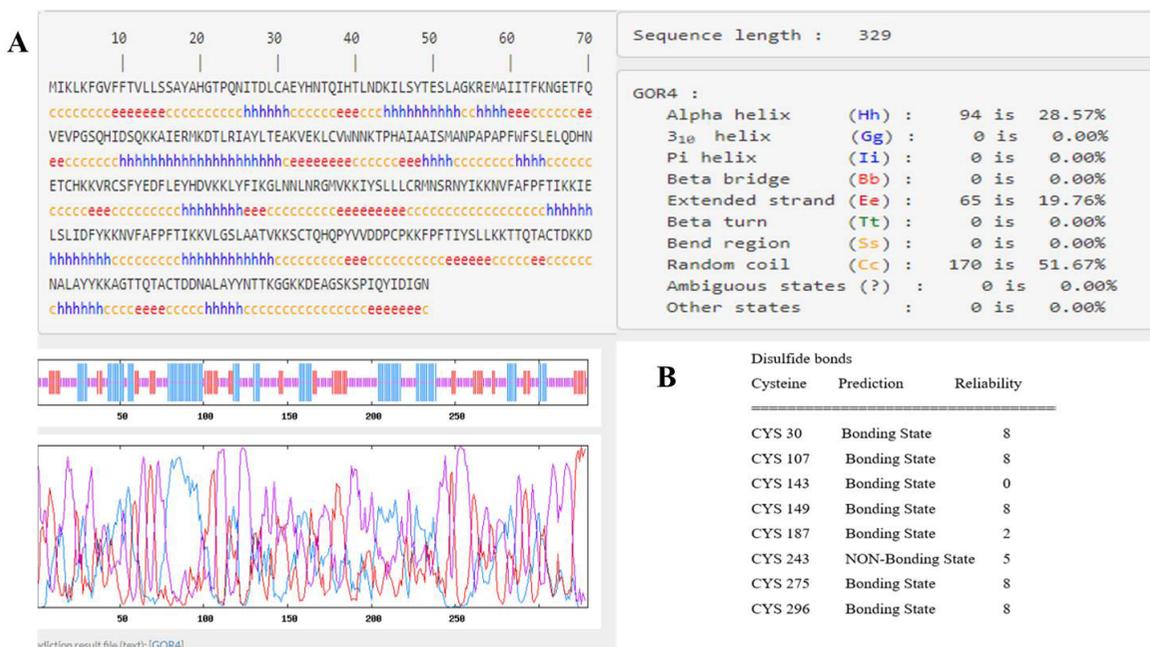

Figure 4. Structural analyses of the designed peptide vaccine

A: Secondary structure of the peptide vaccine predicted by the Prabi Server and the GOR IV method, the length of the peptide is 326 amino acids, the secondary structure includes alpha-helix (28.57%), extended strand (19.76%), and random Coil (51.67%); B: The results of prediction of disulfide bonds.
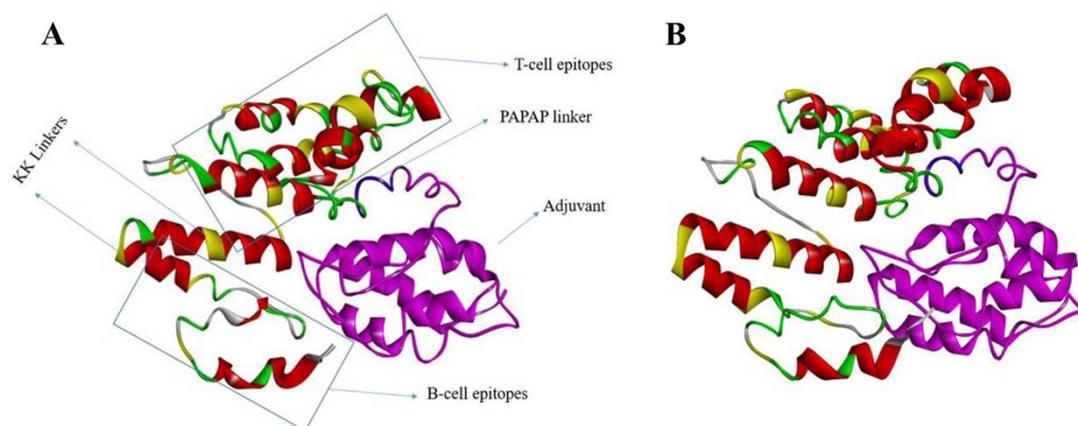
Figure 5. Analysis of 3D structure of the designed peptide vaccine

A: The predicted 3D structure of the constructed vaccine by the C-I-TASSER Server visualized by using discovery studio visualizer 2.0, the purple color indicating the adjuvant sequence, the blue and yellow colors in the scheme related to proline-pich (PAPAP), and lysine-lysine (KK) linkers, Respectively, the B cell and T cell epitopes defined in other colors; B: The predicted 3D structure of the constructed vaccine after refinement by GalaxyWEB Server.

structure of the predicted model after the refinement process has been shown in Figure 5-B.

Moreover, the geometric quality of the refined model has been evaluated using the Ramachandran plot by the PROCHECK server. The quality of the predicted model was investigated before refinement and after the refinement process. Figure 6-A shows the main Ramachandran plot for the 3D model before the refining process. Also,

the Ramachandran plot for the refined model has shown in Figure 6-B. The Ramachandran plot results of the initial structure of the designed vaccine included 59.1% in most favored regions, while these proportions in the refined model were 83.7% that confirmed the refining process. Also, the overall quality score of the model obtained from the ERRAT servers was 70.09 % (Figure 6-C).

Table 3. T-Cell epitopes binding to Major Histocompatibility Complex (MHC) I and MHC II

Epitope	T-Cell Epitope	Protein	Antigenicity	Allergenicity
FWFSLELQDHNETCH	MHC I, MHC II	ORF7b	1.1195	Non-allergen
IYSLLLCRMNSRNYI	MHC I	Envelope protein	0.6011	Non-allergen
VRCSFYEDFLEYHDV	MHC II	ORF8	0.7525	Non-allergen
TTQTACTD	MHC II	Nsp 9	0.6570	Non-allergen
FPFTIYSLL	MHC I, MHC II	ORF8	0.8834	Non-allergen
IELSLIDFY	MHC I	Envelope protein	1.7424	Non-allergen
LYFIKGLNNLNRGMV	MHC I	Nsp 9	0.5997	Non-allergen
DNALAYY	MHC I,MHCII	ORF 10	0.631	Non-allergen
NVFAFPFTI	MHC I	Envelope protein	0.8365	Non-allergen
VLGSLAATV	MHC I	NSp 9	0.6548	Non-allergen

Antigenicity and allergenicity of the most repeated epitopes were analyzed using the VaxiJen 2.0 and the AllerTOP servers, respectively. The non-allergen epitopes with the most antigenicity score were selected.
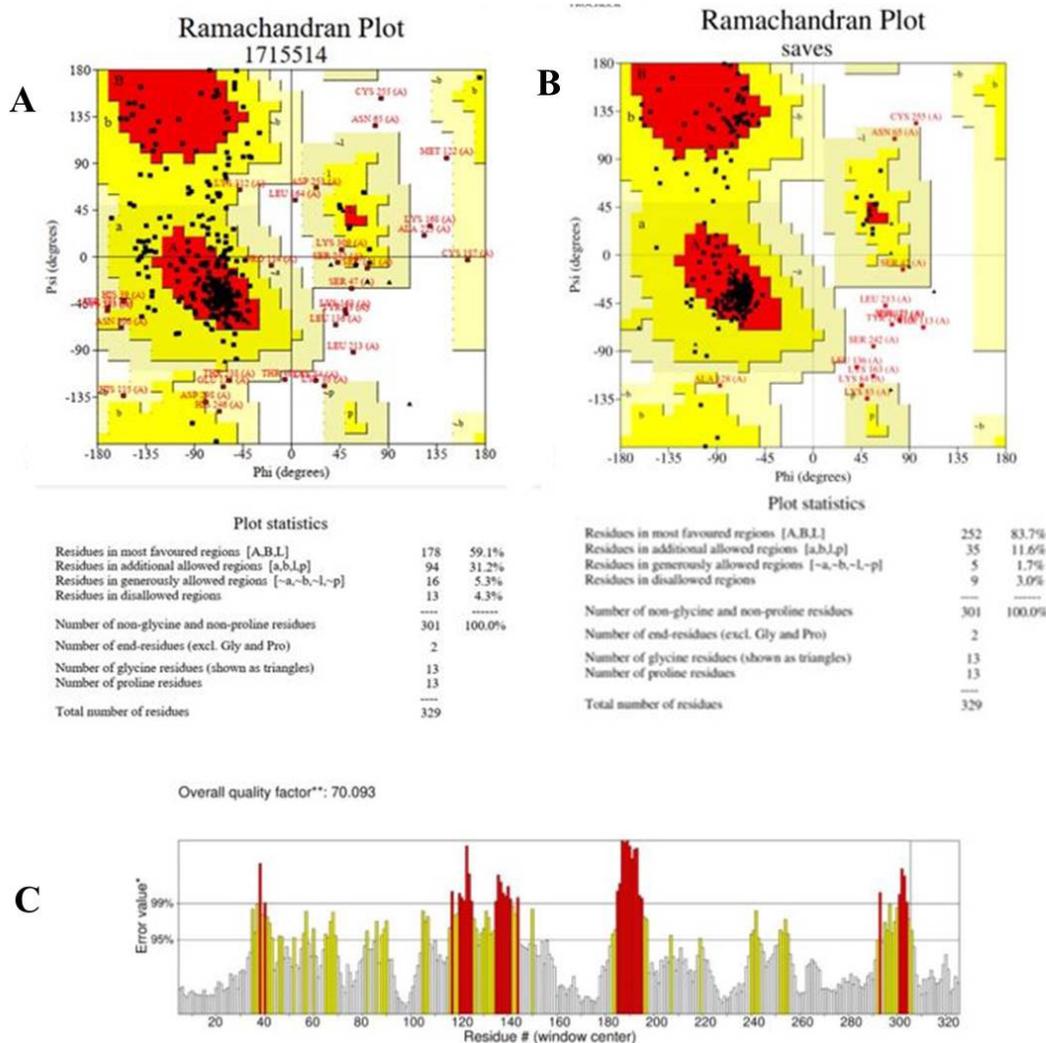
Figure 6. Analysis of the geometrics quality and overall score for constructed peptide before and after refinement

A: Ramachandran plot for the initial model showing 59.1% of residues located in most favored regions and 31.2% located in additional allowed regions, also, 301 residues from a total of 329 residues were non-glycine and non-proline; B: The ramachandran plot for the refined model showing 83.7% of residues located in the most favored regions, the designated additional allowed, generously allowed, and disallowed regions residues were 11.6%, 1.7%, and 3.0%, respectively; C: The ERRAT plot of the vaccine 3D model

Furthermore, to validate the binding affinity and interaction of the peptide vaccine with TLR3, TLR4, MHC I, and MHC II molecules, the molecular docking process has been performed in the antibody mode of the ClusPro 2.0 server. Results of docking scores in the antibody-mode included the lowest energy of -404.7, -485.7, -567.3, and -602.1 kcal/mol for MHC I, MHC II, TLR3, and TLR4, respectively (Table 6). Also, this server predicted a 3D structure for the docked molecules models (Figure 7).

Back translation, codon optimization, and in silico cloning the candidate protein

While the bioinformatic and biochemical analysis was done on the constructed peptide vaccine, the amino acid sequence should be back-translated into nucleotide sequence and then inserted into an expression vector for expression in the bacterial system or other expression systems. For this goal, at first, the final amino acid se-

Table 4. Physiochemical properties of the candidate vaccine

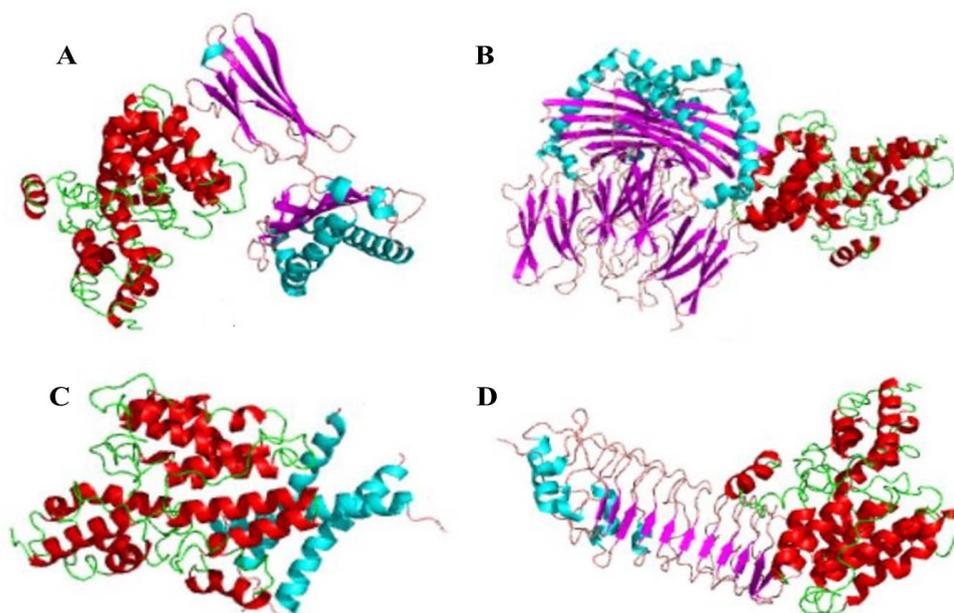| Number of Residues	Molecular Weight	Isoelectric pH	Solubility	Net Charge
329	3745.98 g/mol	9.62	Water soluble	17.3

Figure 7. Docking results for the predicted 3D structure

A: 3D structure after docking the peptide vaccine and Major Histocompatibility Complex (MHC) I; B: MHC II; C: Toll-Like Receptor (TLR)3; and D: TLR4.
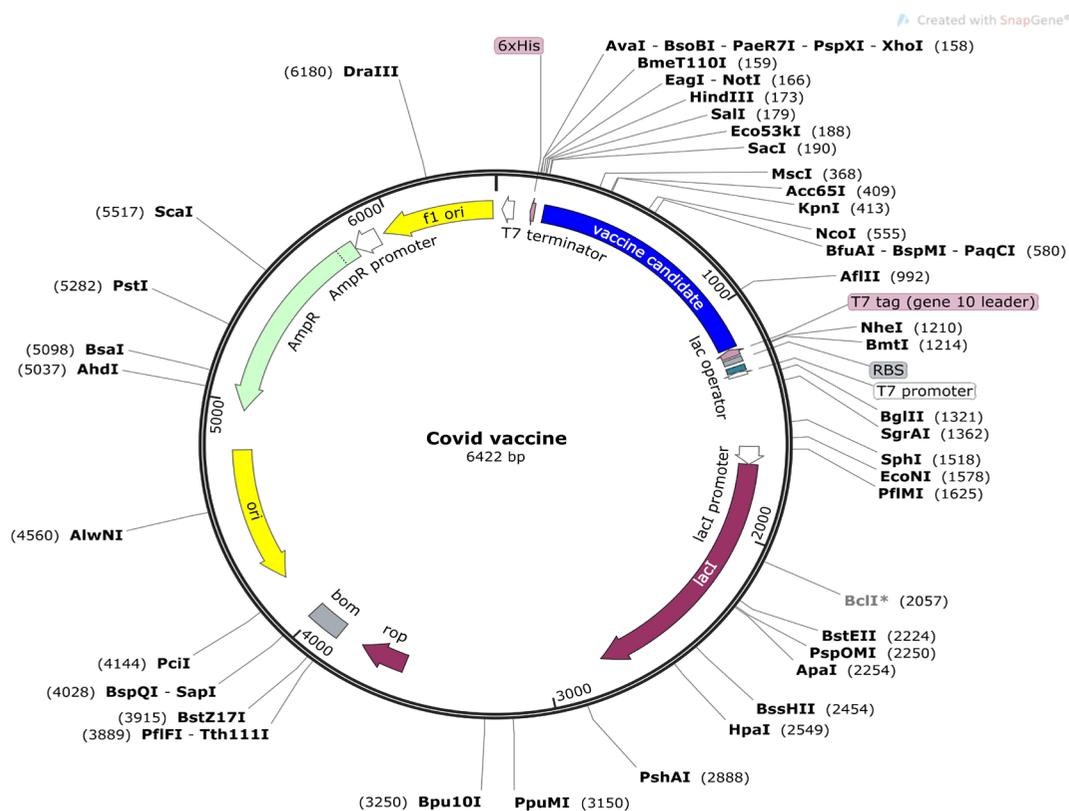

Figure 8. The map of PET 21a expression vector, the inserted fragment are shown in blue color and labeled as the COVID-19 vaccine

Table 5. Results of the model refinement

Models	GDT-HA	RMSD	MolProbity	Clashscore	Poor Rotamers	Rama Favored
Initial	1.0000	0.000	2.478	23.4	1.0	64.9
MODEL 1	0.9932	0.269	2.398	20.6	0.7	88.4
MODEL 2	0.9878	0.282	2.352	19.3	1.0	89.6
MODEL 3	0.9848	0.302	2.408	21.9	0.3	89.0
MODEL 4	0.9894	0.282	2.413	21.1	1.0	88.7
MODEL 5	0.9886	0.281	2.419	23.4	0.3	89.6

RMSD: Root-Mean-Square Deviation.

quence of peptide vaccine converted to nucleotide sequence using SnapGene 3.2.1 offline software. The nucleotide sequence was then optimized for codon usage in *E. coli* using the JCat server. For the next step, the restriction enzyme recognition site and polyhistidine tag were added to the optimized nucleotide sequence. Also, the ORF frames for considering correct protein expression were performed using the SnapGene 3.2.1 software. Finally, the cloning of the nucleotide sequence into the pET-21 expression vector has been simulated by SnapGene 3.2.1 offline software (Figure 8).

Discussion

In this study, we used immunoinformatics to design an appropriate vaccine for the newfound coronavirus. We chose both immunogen B-cell and T-cell related epitopes from different proteins of the SARS-CoV-2 virus using the epitope prediction methods of the IEDB server. Likewise, Bhattacharya et al. developed an epitope-based vaccine against the novel coronavirus (SARS-COV-2) using the immunoinformatics approach. They identified potential B and T cell epitopes of the spike glycoprotein of the coronavirus. The final vaccine construction contains thirteen MHC I and three MHC II related epitopes connected by the EAAAK linker [46].

As shown in Figure 2, our designed vaccine contains CTXB, as a natural adjuvant, and various B- and T-cell epitopes that were connected by linkers. The biochemical and structural properties of the proposed vaccine include amino acid composition, secondary and tertiary structures, stability, and half-life of the constructed peptide was considered using several servers. Also, the prediction of protein stability and protein disorders has been performed using the IUPred 2.0 server [47]. Protein disorders are functional domains or protein sequence segments that have not highly populated stable secondary and tertiary structures under physiological conditions. These sequences do not have a stable 3D structure, so

Table 6. Coefficient weights formula and docking score in ClusPro Server for selected models using antibody mode (E=0.50E rep+−0.20Eatt+600Eelec+0.25EDARS)

Selected Model	Members	Representative	Weighted Score kcal/mol
MHC I	127	Center	-334.5
		The lowest energy	-404.7
MHC II	140	Center	-390.1
		The lowest energy	-485.7
TLR3	55	Center	-479.0
		The lowest energy	-567.3
TLR4	227	Center	-402.4
		The lowest energy	-602.1

MHC: Major Histocompatibility Complex; TLR: Toll-Like Receptors.

that they may lead to protein instability [48, 49]. Our analyses confirmed the stability of the designed chimeric protein by predicting the half-life for the sequence.

Generally, viral surface proteins like glycoproteins play a crucial role in virus fusion to the host cells. On the other hand, these proteins are usually more mutable [50-53]. Therefore, in this study, we selected conserved proteins with the highest antigenicity levels regardless of their functions. The prediction of T-cell and B-cell epitopes is a crucial step to achieve an immunodominant vaccine. Consequently, in this study, IEDB and BepiPred2.0 servers were used to predict linear B-cell epitopes and thus similarly resulted in selected epitopes. Also, to predict appropriate T-cell epitopes (with high antigenicity and non-allergenicity), the IEDB server was used and the most frequently identified epitopes that interact with most of MHC I and MHC II alleles have been selected (Table 3) [27, 54].

The results of the molecular docking process confirmed the immune reaction between the antigen and selected antibodies. All receptors can get an immune interaction with the proposed vaccine (Table 6).

Using computational and immunoinformatics methods has led to an improvement in the vaccine development process. In this way, the reverse vaccinology and structural vaccinology approaches have helped researchers to find the proper epitopes and develop a competent and effective protein-based vaccine. Structural vaccinology was applied to take information about the 3D structure of the designed peptide and model assessment. This approach also developed various web servers for model quality validation and docking studies [55]. Currently, bioinformatics approaches have accelerated the vaccine design and encouraged various studies in this field. For example, Grifoni et al. used the IEDB tool to predict B-cell and T-cell epitopes for SARS-CoV-1 and SARS-CoV-2 viruses based on their genomic similarities [56]. In another study, the envelope protein of the COVID-19 virus has targeted to predict the potential T-cell epitopes for a peptide vaccine design by using immunoinformatics and comparative genomic methods. The authors have reported 10 MHC class I and II binding peptides [57].

Edison Ong et al. performed the same study. They studied predicting the COVID-19 vaccine candidates using Vaxign and the newly developed machine learning-based Vaxign-ML reverse vaccinology tools against the nonstructural Nsp3 protein [58]. Sanami and colleagues suggested a multi-epitope vaccine against the human coronavirus 2. In that study, the spike protein of the virus has been targeted, and HTL, CTL, and B-cell epitopes were selected using the ProPred-1 and ABCpred servers. They used ClusPro 2.0 server for the docking studies of the designed vaccine and HLA-I and HLA-II molecules, which resulted in the lowest energy weighted score of -1165.4 and -1279.1 kcal/mol, respectively [59]. According to Table 6 and Figure 8,

the results of our molecular docking process confirmed the immune reaction between the antigen and selected antibodies. As Table 6 shows, the related receptors interacted with the proposed vaccine. In this regard, using the ClusPro 2.0 server, TLR 4 molecule showed the best interaction compared to other receptors with the lowest energy of -602.1 kcal/mol in antibody-mode of ClusPro2.0 server [39]. In this server, the antibody mode was different from the balanced mode. Brenke et al. removed the usual assumption of symmetry using asymmetric pairwise potential. Thus, each atom on the antibody molecule is different from the corresponding atom on the antigen protein [41].

Conclusion

The development of a vaccine is a time-consuming and costly process that involves bioinformatics and experimental approaches. Therefore, an epitope-based vaccine designed by immunoinformatics methods can be considered a practical approach for designing and producing a novel vaccine against this virus. The results of this in silico study confirmed that the designed vaccine that had high antigenicity and stability could be a proper vaccine candidate against the COVID-19 disease. Consequently, to analyze the immune-dominant features, the proposed vaccine should be considered for further in vitro and in vivo studies to investigate the potential immunogenicity.

Ethical Considerations

Compliance with ethical guidelines

All ethical principles were considered in this article.

Authors' contributions

Conceptualization and supervision: Afshin Samimi Nemati and Majid Tafrihi; Methodology: Fatemeh Sheikhi, Abolfazl Rostamian Tabari, Amirhossein Haditabar; Investigation, writing – original draft: All authors; Writing – review & editing: Majid Tafrihi; Data collection: Afshin Samimi Nemati, Fatemeh Sheikhi, Abolfazl Rostamian Tabari, Amirhossein Haditabar; Data analysis: Afshin Samimi Nemati and Majid Tafrihi.

Conflict of interest

The authors declared no competing interests.

References

[1] Chakraborty C, Sharma AR, Bhattacharya M, Sharma G, Lee SS. The 2019 novel coronavirus disease (COVID-19) pandemic: A zoonotic prospective. Asian Pac J Trop Med. 2020; 13(6):242-6. [DOI:10.4103/1995-7645.281613]

[2] Contini C, Nuzzo MD, Barp N, Bonazza A, Giorgio RD, Tognon M, et al. The novel zoonotic COVID-19 pandemic: An expected global health concern. J Infect Dev Ctries. 2020; 14(3):254-64. [DOI:10.3855/jidc.12671] [PMID]

[3] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. Lancet 2020; 395(10229):1054-62. [DOI:10.1016/S0140-6736(20)30566-3]

[4] Shereen MA, Khana S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. J Adv Res. 2020; 24:91-8. [DOI:10.1016/j.jare.2020.03.005] [PMID] [PMCID]

[5] Fernandes JD, Hinrichs AS, Clawson H, Gonzalez JN, Lee BT, Nassar LR, et al. The UCSC SARS-CoV-2 Genome Browser. Nat Genet. 2020; 52(10):991-8. [DOI:10.1038/s41588-020-0700-8] [PMID] [PMCID]

[6] Mohd HA, Al-Tawfiq JA, Memish ZA. Middle East respiratory syndrome coronavirus (MERS-CoV) origin and animal reservoir. Virol J. 2020; 13:87. [DOI:10.1186/s12985-016-0544-0] [PMID] [PMCID]

[7] Park SE. Epidemiology, virology, and clinical features of severe acute respiratory syndrome-coronavirus-2 (SARS-COV-2; Coronavirus Disease-19). Clin Exp Pediatr. 2020; 63(4):119-24. [DOI:10.3345/cep.2020.00493] [PMID] [PMCID]

[8] Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020; 579(7798):265-9. [DOI:10.1038/s41586-020-2008-3] [PMID] [PMCID]

[9] Yang Y, Peng F, Wang R, Guan K, Jiang T, Xu G, et al. The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. J Autoimmun. 2020; 109:102434. [DOI:10.1016/j.jaut.2020.102434] [PMID] [PMCID]

[10] Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun. 2020; 109:102433. [DOI:10.1016/j.jaut.2020.102433] [PMID] [PMCID]

[11] Pan L, Mu M, Yang P, Sun Y, Wang R, Yan J, et al. Clinical characteristics of COVID-19 patients with digestive symptoms in Hubei, China: A descriptive, cross-sectional, multicenter study. Am J Gastroenterol 2020; 115(5):766-73. [DOI:10.14309/ajg.0000000000000620] [PMID] [PMCID]

[12] Odayar J, Myer L, Malaba TR. The epidemiology and pathogenesis of SARS-CoV-2 infection in pregnancy: More questions than answers. EClinicalMedicine. 2020; 26:100534. [DOI:10.1016/j.eclinm.2020.100534] [PMID] [PMCID]

[13] Cárdenas-Conejo Y, Liñan-Rico A, García-Rodríguez DA, Centeno-Leija S, Serrano-Posada H. An exclusive 42 amino acid signature in pp1ab protein provides insights into the evolutive history of the 2019 novel human-pathogenic coronavirus (SARS-CoV-2). J Med Virol. 2020; 92(6):688-92. [DOI:10.1002/jmv.25758] [PMID] [PMCID]

[14] Díaz J. SARS-CoV-2 molecular network structure. Front Physiol. 202; 11:870. [DOI:10.3389/fphys.2020.00870] [PMID] [PMCID]

[15] Michel CJ, Mayer M, Poch O, Thompson JD. Characterization of accessory genes incoronavirus genomes. Virol J. 2020; 17(1):131. [DOI:10.1186/s12985-020-01402-1] [PMID] [PMCID]

[16] Mahase E. Covid-19: Oxford researchers halt vaccine trial while adverse reaction is investigated. BMJ. 2020; 370:m3525. [DOI:10.1136/bmj.m3525] [PMID]

[17] Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. Open Biol. 2013; 3(1):120139. [DOI:10.1098/rsob.120139] [PMID] [PMCID]

[18] Soltani S, Farahani A, Dastranj M, Momenifar N, Mohajeri P, Emamie AD. DNA vaccine: Methods and mechanisms. Adv Hum Biol. 2018; 2018:132-9. [DOI:10.4103/AIHB.AIHB_74_17]

[19] Wang X, Lam JY, Chen L, Wing-Ngor Au S, To KKW, Yuen KY, et al. Mining of linear B cell epitopes of SARS-CoV-2 ORF8 protein from COVID-19 patients. Emerg Microbes Infect. 2021; 10(1):1016-23. [DOI:10.1080/22221751.2021.1931465] [PMID] [PMCID]

[20] Hassan SS, Choudhury PP, Roy B. Rare mutations in the accessory proteins ORF6, ORF7b, and ORF10 of the SARS-CoV-2 genomes. Meta Gene. 2021; 28:100873. [DOI:10.1016/j.mgene.2021.100873] [PMID] [PMCID]

[21] Khanmohammadi S, Rezaei N. Role of toll-like receptors in the pathogenesis of COVID-19. J Med Virol. 2021; 93(5):2735-9. [DOI:10.1002/jmv.26826] [PMID] [PMCID]

[22] Rappuoli R. Reverse vaccinology. Curr Opin Microbiol. 2000; 3(4):445-50. [DOI:10.1016/S1369-5274(00)00119-3]

[23] Sette A, Fikes J. Epitope-based vaccines: An update on epitope identification, vaccine design and delivery. Curr Opin Immunol. 2003; 15(4):461-70. [DOI:10.1016/S0952-7915(03)00083-9] [PMID]

[24] Doytchinova IA, Flower DR. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007; 8:4. [DOI:10.1186/1471-2105-8-4] [PMID] [PMCID]

[25] Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v. 2-a server for in silico prediction of allergens. J Mol Model. 2014; 20(6):2278. [DOI:10.1007/s00894-014-2278-5] [PMID]

[26] Dhanda SK, Mahajan S, Paul S, Yan Z, Kim H, Jespersen MC, et al. IEDB-AR: Immune epitope database-analysis resource in 2019. Nucleic Acids Res. 2019; 47(W1):W502-6. [DOI:10.1093/nar/gkz452] [PMID] [PMCID]

[27] 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):W24-9. [DOI:10.1093/nar/gkx346] [PMID] [PMCID]

[28] Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, et al. The design and implementation of the immune epitope database and analysis resource. Immunogenetics. 2005; 57(5):326-36. [DOI:10.1007/s00251-005-0803-5] [PMID] [PMCID]

[29] Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci. 1993; 2(9):1511-9. [DOI:10.1002/pro.5560020916] [PMID] [PMCID]

[30] Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis a virus-neutralizing antibody by a virus specific synthetic peptide. J Virol. 1985; 55(3):836-9. [DOI:10.1128/jvi.55.3.836-839.1985] [PMID] [PMCID]

[31] Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, et al. ElliPro: A new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics. 2008; 9:514. [DOI:10.1186/1471-2105-9-514] [PMID] [PMCID]

[32] Lear S, Cobb SL. Pep-Calc. com: A set of web utilities for the calculation of peptide and peptoid properties and automatic mass spectral peak assignment. J Comput Aided Mol Des. 2016; 30(3):271-7. [DOI:10.1007/s10822-016-9902-7] [PMID] [PMCID]

[33] Erdős G, Dosztányi Z. Analyzing protein disorder with IUPred2A. Curr Protoc Bioinformatics. 2020; 70(1):e99. [DOI:10.1002/cpbi.99] [PMID]

[34] Ferrè F, Clote P. DiANNA 1.1: An extension of the DiANNA web server for ternary cysteine classification. Nucleic Acids Res. 2006; 34(Web Server issue):W182-5. [DOI:10.1093/nar/gkl189] [PMID] [PMCID]

[35] Yang J, Zhang Y. I-TASSER server: New development for protein structure and function predictions. Nucleic Acids Res. 2015; 43(W1):W174-81. [DOI:10.1093/nar/gkv342] [PMID] [PMCID]

[36] Zhang C, Freddolino PL, Zhang Y. COFACTOR: Improved protein function prediction by combining structure, sequence and protein-protein interaction information. Nucleic Acids Res. 2017; 45(W1):W291-9. [DOI:10.1093/nar/gkx366] [PMID] [PMCID]

[37] Biovia DS. Discovery studio visualizer. San Diego: Dassault Systèmes; 2017. https://scholar.google.com/scholar?q=Biovia+DS+Discovery+studio+visualizer+2017+CA,+USA+San+Diego+

[38] Ko J, Park H, Heo L, Seok C. GalaxyWEB server for protein structure prediction and refinement. Nucleic acids research. Nucleic Acids Res. 2012; 40(Web Server issue):W294-7. [DOI:10.1093/nar/gks493] [PMID] [PMCID]

[39] Shin WH, Lee GR, Heo L, Lee H, Seok C. Prediction of protein structure and interaction by GALAXY protein modeling programs. Bio Design. 2014; 40(Web Server issue):W294–7. https://www.readcube.com/articles/10.1093/nar/gks493

[40] Laskowski RA, Macarthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Crystallogr. 1993; 26:283-91. [DOI:10.1107/S0021889892009944]

[41] Brenke R, Hall DR, Chuang GY, Comeau SR, Bohnuud T, Beglov D, et al. Application of asymmetric statistical potentials to antibody-protein docking. Bioinformatics. 2012; 28(20):2608-14. [DOI:10.1093/bioinformatics/bts493] [PMID] [PMCID]

[42] 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. [DOI:10.1002/prot.24403] [PMID] [PMCID]

[43] Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. Nat Protoc. 2018; 12(2):255-78. [DOI:10.1038/nprot.2016.169] [PMID] [PMCID]

[44] Vajda S, Yueh C, Beglov D, Bohnuud T, Mottarella SE, Xia B, et al. New additions to the Clus Pro server motivated by CAPRI. Proteins. 2017; 85(3):435-44. [DOI:10.1002/prot.25219] [PMID] [PMCID]

[45] Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res. 2005; 33(Web Server issue):W526-31. [DOI:10.1093/nar/gki376] [PMID] [PMCID]

[46] Bhattacharya M, Sharma AR, Patra P, Ghosh P, Sharma G, Patra BC, et al. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. J Med Virol. 2020; 92(6):618-31. [DOI:10.1002/jmv.25736] [PMID] [PMCID]

[47] Dosztányi Z, Csizmók V, PéterTompa, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol. 2005; 347(4):827-39. [DOI:10.1016/j.jmb.2005.01.071] [PMID]

[48] Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Soh J, et al. Intrinsically disordered protein. J Mol Graph Model. 2001; 19(1):26-59. [DOI:10.1016/S1093-3263(00)00138-8]

[49] Lieutaud P, Ferron F, Uversky AV, Kurgan L, Uversky VN, Longhi S. How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. Intrinsically Disord Proteins. 2016; 4(1):e1259708. [DOI:10.1080/21690707.2016.1259708] [PMID] [PMCID]

[50] Bowden TA, Jones EY, Stuart DI. Cells under siege: viral glycoprotein interactions at the cell surface. J Struct Biol. 2011; 175(2):120-6. [DOI:10.1016/j.jsb.2011.03.016] [PMID] [PMCID]

[51] Dash R, Das R, Junaid M, Akash MFC, Islam A, Hosen SZ. In silico-based vaccine design against Ebola virus glycoprotein. Adv Appl Bioinforma Chem. 2017; 10:11-28. [DOI:10.2147/AABC.S115859] [PMID] [PMCID]

[52] Lee JE, Saphire EO. Neutralizing ebolavirus: Structural insights into the envelope glycoprotein and antibodies targeted against it. Curr Opin Struct Biol. 2009; 19(4):408-17. [DOI:10.1016/j.sbi.2009.05.004] [PMID] [PMCID]

[53] Young KA, Herbert AP, Barlow PN, Holers VM, Hannan JP. Molecular basis of the interaction between complement receptor type 2 (CR2/CD21) and Epstein-Barr virus glycoprotein gp350. J Virol. 2008; 82(22):11217-27. [DOI:10.1128/JVI.01673-08] [PMID] [PMCID]

[54] Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, et al. Immune epitope database analysis re-

source (IEDB-AR). Nucleic Acids Res. 2008; 36(Web Server issue):W513-8. [DOI:10.1093/nar/gkn254] [PMID] [PMCID]

[55] Ishack S, Lipner SR. Bioinformatics and immunoinformatics to support COVID-19 vaccine development. J Med Virol. 2021; 3(9):5209-11. [DOI:10.1002/jmv.27017] [PMID] [PMCID]

[56] Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. Cell Host Microbe. 2020; 27(4):671-80. [DOI:10.1016/j.chom.2020.03.002] [PMID] [PMCID]

[57] Abdelmageed MI, Abdelmoneim AH, Mustafa MI, Elfadol NM, Murshed NS, Shantier SW, et al. Design of a multiepitope-based peptide vaccine against the E protein of human COVID-19: an immunoinformatics approach. Biomed Res Int. 2020; 2020:2683286. [DOI:10.1101/2020.02.04.934232] [PMID] [PMCID]

[58] Ong E, Wong MU, Huffman A, He Y. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. Front Immunol. 2020; 11:1581. [DOI:10.3389/fimmu.2020.01581] [PMID] [PMCID]

[59] Sanami S, Zandi M, Pourhossein B, Mobini GR, Safaei M, Abed A, et al. Design of a multi-epitope vaccine against SARS-CoV-2 using immunoinformatics approach. Int J Biol Macromol. 2020; 164:871-83. [DOI:10.1016/j.ijbiomac.2020.07.117] [PMID] [PMCID]

This Page Intentionally Left Blank