Designing a new multi epitope-based vaccine against COVID-19 disease: an immunoinformatic study based on reverse vaccinology approach

Afshin Samimi Nemati
University of Mazandaran Faculty of Basic Science

Majid Tafrihi (✉ m.tafrihi@umz.ac.ir)
University of Mazandaran Faculty of Basic Science  https://orcid.org/0000-0001-7200-4252

Fatemeh Sheikhi
University of Mazandaran Faculty of Basic Science

Abolfazl Rostamian Tabari
University of Mazandaran Faculty of Basic Science

Amirhossein Haditabar
University of Mazandaran Faculty of Basic Science

Research Article

Keywords: COVID-19, Reverse vaccinology, Epitope-based vaccine, Immunoinformatics

DOI: https://doi.org/10.21203/rs.3.rs-206270/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has currently caused a significant pandemic among worldwide populations. The transmission speed and the high rate of mortality caused by the disease necessitate studies for the rapid designing and effective vaccine production. The purpose of this study is to predict and design a novel multi-epitope vaccine against the SARS-CoV-2 virus using bioinformatics approaches. Coronavirus envelope proteins, ORF7b, ORF8, ORF10, and NSP9 were selected as targets for epitope mapping using IEDB and BepiPred 2.0 Servers. Also, molecular docking studies were performed to determine the candidate vaccine's affinity to TLR3, TLR4, MHC I, and MHC II molecules. Thirteen epitopes were selected to construct the multi-epitope vaccine. We found that the constructed peptide has valuable antigenicity, stability, and appropriate half-life. The Ramachandran plot 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. This study confirmed that the designed vaccine has a good antigenicity and stability and could be a proper vaccine candidate against the 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 a 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 COVID-19 infection is rapidly transmissible among people worldwide [4]. The SARS-CoV-2 genome is a non-segmented positive-sense RNA of 29,903 nucleotides [5] that has ~80% similarity to the genome of the Severe Acute Respiratory Syndrome (SARS-CoV-1) and Middle East Respiratory Syndrome (MERS-CoV) viruses, that 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].

During this short period of the COVID-19 pandemic, many efforts have been made to introduce therapeutic methods and new drugs to treat Covid-19 infection [14]. To this end, anti-viral drugs, interferon-α nebulization, and broad-spectrum antibiotics have shown promising results in some therapeutic cases [15–17]. Despite these, there is no approved drug available in the medical centers yet [18].

Reverse vaccinology (RV), an approach that utilizes the expressed genomic sequences to find new potential vaccines [19], was firstly applied for Neisseria meningitidis group B and now is accepted as a successful method for producing new vaccines [20]. 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) [21]. For this purpose, to acquire an appropriate peptide vaccine, the
biochemical and physicochemical properties of the constructed peptide considered. According to the current advances in this field, several servers were established [22].

In the present study, we designed a new multi-epitope peptide vaccine against SARS-CoV-19 infection. 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 Toll-like receptors (TLR3 and TLR4) and major histocompatibility complex (MHC I and MHC II).

**Materials And Methods**

We took all steps of this study according to Fig. 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 structure of TLR 3, TLR 4, MHC I, and MHC II proteins were obtained from the protein databank (available at https://www.rcsb.org/), in PDB format, with PDB entry of 3ULS, 4R7N, 4UQ3A, and 4GBX, respectively.

**Selecting the protein antigen**

Among protein sequences, the five top score proteins including; 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) selected as candidates that had the maximum of antigenicity and were non-allergens. Antigenicity and the allergenicity of the selected proteins 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 [22, 23].

**B-cell and T-cell epitope prediction**

In the following, IEDB server (available at https://www.iedb.org/) and BepiPred 2.0 accessible at (http://www.cbs.dtu.dk/) have used to predict the linear B-cell epitopes from the chosen proteins, [24, 25]. For the prediction of the T-cell related epitopes that contain the MHC I and MHC II, the IEDB server was utilized. Antigenicity and allergenicity analysis for the selected epitopes using Vaxijen 2.0 and AllerTOP v. 2.0 servers was repeated. The ToxinPred server (available at https://webs.iiitd.edu.in/raghava/toxinpred/algo.php) used to predict the toxic region in the selected epitopes.

**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 the epitope selection could lead to reducing the length of the final amino acid sequence. Lysine-lysine (KK) cross-linkers connected 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 Proline-rich (PAPAP) rigid linker.

**Analyzing the physicochemical properties**

In this step, the physicochemical properties of the proposed vaccine performed. To this end, amino acid composition, molecular weight, iso-electric 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 [26]. Finally, using the Iupred 2a server (https://iupred2a.elte.hu/) predicted the protein stability [27].

**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). In the next step, the 3D structure of the protein has been predicted by using the I-TASSER server (https://zhanglab.ccb.med.umich.edu) [28, 29]. 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 performed using the GalaxyWEB server (available at galaxy.seoklab.org/) to reduce possible structural faults [30, 31]. Moreover, the geometric features of the proposed vaccine validated based on the Ramachandran plot by using the PROCHECK server (available at https://services.mbi.ucla.edu/PROCHECK/) [32]. The molecular docking process was conducted on 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 [33-36].

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

The Snap-Gene 3.2.1 software performed the back-translation of 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 [37]. To have the correct translation, the open reading frame (ORF) has
checked by using Snap-Gene 3.2.1 software. Finally, in silico restriction cloning was performed using Snap-Gene 3.2.1 software. 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, chosen proteins compared with each other in terms of antigenicity, allergenicity, and toxicity. Finally, five chosen proteins had high levels of antigenicity and were non-allergen (see 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 of the B-cell epitopes have 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, thirteen epitopes were selected and then linked together using lysine-lysine (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 [38]. The amino acid sequence of cholera toxin B subunit (CTXB) that is 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 (Fig. 2).

Physicochemical properties of constructed peptide

Physicochemical properties and amino acid composition of the constructed peptide evaluated using Pepcalc and Protparam servers. The results showed that the proposed vaccine was stable, water-soluble, with a molecular weight of 39.4 kDa. The calculated pl value was 9.6, 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 (Fig. 3A). Also, the analysis of protein sequence stability was performed through predicting the protein disorder regions (by using Iupred 2A server), and the results confirmed the stability of the designed peptide (Figs 3B and 3C). The Iupred 2A 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 selected for investigation. Protein disorders of the proposed vaccine were predicted by Iupred 2.0 designed graph. Due to the designed graph that showed in Fig. 3C, the sequence of the proposed vaccine has not a great chance to be as 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 as protein disorders and protein orders, respectively.
Secondary and tertiary structure of the constructed peptide

The prediction of the secondary structure of the proposed vaccine 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 Fig. 4 shows, the secondary structure of the proposed vaccine contains alpha-helix (28.57%), extended strand (19.76%), and random coil (51.67 %). Using the I-TASSER server predicted the tertiary structure of the peptide. This server suggests various models for the input sequence and the quality of prediction models reflected on the form of c-scores (-5 to 2). The higher values of the C-score relate to the higher confidence levels for the predicted model (Fig. 5).

Model refinement and molecular docking of the vaccine candidate peptide

The refinement process has performed for the predicted model using the GalaxyWeb server. In this process, the server refined the secondary structure elements like; loop regions and side chains based on several factors containing similarity score (GDT-HA), clash score, RMSD, and MolProbity. The global distance test (GDT_TS) is a measure of similarity between 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. The root-mean-square deviation (RMSD) is the average distance between backbone atoms in the protein structure. The MolProbity 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. This Table presented the GalaxyWeb server as a result of the refinement process for the 3D predicted model of the proposed vaccine. Table 4 shows the five suggested refined models. The first presented model with GDT-HA and RMSD of 0.8951 and 0.542 selected for further considerations. However, the clash score for the chosen model was 33.6, reported score for the initial model was about 24.1. The Rama favored scores was the other score that indicates the percentage of the residues in the most favored regions of the Ramachandran plot. This score changed from 64.2 to 84.4, simultaneously for the refined model in comparison to the initial model.

Moreover, the geometric quality of the refined model has evaluated using the Ramachandran plot by the PROCHECK server. The quality of the predicted model was investigated before refinement and after the refinement process. Fig. 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 Fig. 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 78.7% that confirmed the refining process.

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 -434.3, -515.1, -602.1, and -567.3 Kcal.mol\(^{-1}\) for MHC I, MHC II, TLR3, and TLR4, respectively (Table 5). Also, this server predicted a 3D structure for the docked molecules models (Fig. 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 sequence of peptide vaccine converted to nucleotide sequence using Snap-Gene 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 added to the optimized nucleotide sequence. Also, the ORF frames for considering correct protein expression performed using the Snapp-Gene 3.2.1 software. Finally, the cloning of the nucleotide sequence into the PET 21 expression vector has simulated by using Snapp-Gene 3.2.1 offline software (Fig. 8).

Discussion

The SARS-CoV-2 virus and its associated disease COVID-19 have become a crucial world health concern resulted in a significant mortality. On the other hand, there are no approved vaccines against COVID-19, yet, and these conditions resist until a specific medicine or a vaccine gets approved. Therefore, researchers started to discover drugs and vaccines by using bioinformatics methods. Some of the studies reported new vaccines and medical drugs for the COVID-19 infectious disease that some of them were in the clinical trial phases [39].

There are several methods to produce new vaccines like; live-attenuated form, subunit vaccines, multi-epitope vaccines, and DNA vaccines [40, 41]. The reverse vaccinology approach is a relatively new method to find a successful and safe vaccine in a relatively short time [42]. Main immunoinformatics approaches have improved this method by creating several databases and algorithms for epitope prediction, which increases the speed and accuracy of the vaccine development [40]. Nosrati and his colleagues have identified the first epitope-based vaccine against Crimean-Congo hemorrhagic fever virus based on the reverse vaccinology approach. Their designed peptide vaccine showed good stability, water-solubility, non-allergenicity, and highly antigenic activity [43]. Ahmad et al. (2020) suggested a novel vaccine for the COVID-19 virus using molecular docking and molecular dynamics of interaction with TLR3 and TLR 4 molecules [44].

In this study, we used immunoinformatics to design an appropriate vaccine for the newfound coronavirus. We chose both immunogen B and T-cell related epitopes from different proteins of the SARS-CoV-2 virus using the epitope prediction methods of the IEDB server. In the same way, Bhattacharya et al. (2020) 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 that were connected by the EAAAK linker [45].

As is shown in Fig. 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 structure, stability, and half-life of the constructed peptide was considered using several servers. Also, the prediction of protein stability and protein disorders has performed using the lupred 2.0 server [46]. Protein disorders are functional domains or protein sequence segments that have not had a highly populated stable secondary and tertiary structure under physiological conditions. These sequences do not have a stable three-dimensional structure, so they may lead to protein instability [47, 48]. 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 [49-52]. 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 Bipred2.0 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 used and the most frequently identified epitopes that interact with most of MHC I and MHC II alleles have selected (Table 3) [25, 53].

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 5).

Sanami and colleagues have suggested a multi-epitope vaccine against the human coronavirus 2. In that study, the spike protein of the virus has targeted, and HTL, CTL, and B-cell epitopes were selected using the ProPred-1 and ABCpred servers. They have 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 [54]. According to Table 4 and Fig. 8, the results of our molecular docking process confirmed the immune reaction between the antigen and selected antibodies. As Table 4 shows, related receptors interacted with the proposed vaccine. In this regard, using the Cluspro 2.0 server, TLR 3 molecule showed the best interaction comparing to other receptors with the lowest energy of -602.1 kcal.mol-1 in antibody-mode of Cluspro2.0 server [33]. In this server, the antibody-mode was different from the balanced-mode. Brenke et al. (2012) 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 [33].

**Conclusion**

The process of development of a vaccine is a time-consuming and costly process that involves bioinformatics and experimental approaches. Therefore, an epitope-based vaccine that is designed by immunoinformatics methods can be considered as 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.

Declarations

Authors' contributions

A.S.N., designed research, performed research, analyzed data, and drafted, M.T., designed research, analyzed data, wrote and revised the draft, F. Sh., performed research, and analyzed data, A. R. T., performed research, and analyzed data, and A. H., performed research, and analyzed data.

Compliance with Ethical statement

This article does not contain any studies involving animals or human participants performed by any of the authors.

Conflict of Interest Statement

The authors have no relevant financial or non-financial interests to disclose.

Informed consent

This study does not contain any human samples, so informed consent does not apply to this study.

Ethical approval

This is a computational study, and the University of Mazandaran Research Ethics Committee has confirmed that no ethical approval is required.

Funding Sources

The authors did not receive support from any organization for the submitted work.

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;13:242-6.
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;14:254-64.
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;395:1054-62.
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.
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;52:986-1001.

6. Mohd HA, Al-Tawfiq JA, Memish ZA. Middle East respiratory syndrome coronavirus (MERS-CoV) origin and animal reservoir. Virol J 2020;13:87.

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;63:119-24.

8. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579:265-9.

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.

10. Rothan H, Byrareddy S. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 2020;109.

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;115:766-73.

12. Rothan H, Byrareddy S. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 2020;109.

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;92:688-92.

14. Shah B, Modi P, R.Sagar S. In silico studies on therapeutic agents for COVID-19: Drug repurposing approach. Life Sci 2020;252:117652.

15. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First Case of 2019 Novel Coronavirus in the United States. New Engl J Med 2020;382:929-36.

16. Nile SH, Nile A, Qiu J, Li L, Jia X, Kai G. COVID-19: Pathogenesis, cytokine storm and therapeutic potential of interferons. Cytokine Growth Factor Rev 2020;53:66-70.

17. Wang BX, Fish EN. Global virus outbreaks: Interferons as 1st responders. Semin Immunol 2019;43:101300.

18. Lu H. Drug treatment options for the 2019-new coronavirus (2019-nCoV). Biosci Trends 2020;14:69-71.

19. Rappuoli R. Reverse vaccinology Curr opin Microbiol 2000;3:445-50.

20. Masignani V, Rappuoli R, Pizza M. Reverse vaccinology: a genome-based approach for vaccine development. Expert opinion on biological therapy. Expert Opin Biol Ther 2003;2::895-905.

21. Sette A, Fikes J. Epitope-based vaccines: an update on epitope identification, vaccine design and delivery. Curr Opin Immunol 2003;15:461-70.

22. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007;8.
23. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v. 2—a server for in silico prediction of allergens. J Mol Model 2014;20:2278.

24. 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:W502-W6.

25. 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:W24-W9.

26. 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;30:271-7.

27. Erdős G, Dosztányi Z. Analyzing Protein Disorder with IUPred2A. Curr Protoc Bioinformatics 2020;70.

28. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res 2015;43:W174-W81.

29. 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:W291-W9.

30. Ko J, Park H, Heo L, Seok C. GalaxyWEB server for protein structure prediction and refinement. Nucleic acids research. Nucleic Acids Res 2012;40:W294–W7.

31. Shin W-H, Lee GR, Heo L, Lee H, Seok C. Prediction of protein structure and interaction by GALAXY protein modeling programs. Bio Design. 2014;2:1-11.

32. 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.

33. Brenke R, Hall DR, Chuang G-Y, Comeau SR, Bohnuu T, Beglov D, et al. Application of asymmetric statistical potentials to antibody-protein docking. Bioinformatics. 2012;28:2608–14.

34. Kozakov D, Beglov D, Bohnuu T, Mottarella SE, Xia B, Hall DR, et al. How good is automated protein docking? Proteins. 2013;81:2159-66.

35. Kozakov D, Hall DR, Xia B, Porter KA, Padhomy D, Yueh C, et al. The ClusPro web server for protein-protein docking. Nat Protoc 2018;12:255-78.

36. Vajda S, Yueh C, Beglov D, Bohnuu T, Mottarella SE, Xia B, et al. New additions to the Clus Pro server motivated by CAPRI. Proteins. 2017;85:435-44.

37. 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:W526–W31.

38. Perrie Y, Mohammed AR, Kirby DJ, McNeil SE, Bramwell VW. Vaccine adjuvant systems: enhancing the efficacy of sub-unit protein antigens. Int J Pharm 2008;364:272-80.

39. Mahase E. Oxford researchers halt vaccine trial while adverse reaction is investigated. BMJ. 2020;370:m3525.

40. Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. Open Biol 2013;3:120139.
41. Soltani S, Farahani A, Dastranj M, Momenifar N, Mohajeri P, Emamie AD. DNA vaccine: Methods and mechanisms. Adv Hum Biol 2018;2018:132-9.

42. Dellagostin OA, Grassmann AA, Rizzi C, Schuch RA, Jorge S, Oliveira TL, et al. Reverse vaccinology: an approach for identifying leptospiral vaccine candidates. Int J Mol Sci 2017;18:158.

43. Nosrati M, Behbahani M, Mohabatkar H. Towards the first multi-epitope recombinant vaccine against Crimean-Congo hemorrhagic fever virus: A computer-aided vaccine design approach. J Biomed Inform 2019;93:103160.

44. Ahmad S, Navid A, Farid R, Abbas G, Ahmad F, Zaman N, et al. Design of a Novel Multi Epitope-Based Vaccine for Pandemic Coronavirus Disease (COVID-19) by Vaccinomics and Probable Prevention Strategy against Avenging Zoonotics. Eur J Pharm Sci 2020;151:105387.

45. 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;92:618-31.

46. 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;347:827-39.

47. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Soh J, et al. Intrinsically disordered protein. J Mol Graph Model 2001;19:26-59.

48. 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:e1259708.

49. Bowden TA, Jones EY, Stuart DI. Cells under siege: viral glycoprotein interactions at the cell surface. J Struct Biol 2011;175:120-6.

50. 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.

51. Lee JE, Saphire EO. Neutralizing ebolavirus: structural insights into the envelope glycoprotein and antibodies targeted against it. Curr Opin Struct Biol 2009;19:408-17.

52. 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;82:11217-27.

53. 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:W513-W8.

54. Sanami S, Zandi M, Pourhossein B, Mobini G-R, 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.

Tables
Table 1: The list of protein candidates. The sequences were retrieved from the NCBI database and then their antigenicity, allergenicity, and toxicity were analyzed. Selected proteins have been marked with *

| Protein                                      | NCBI Reference Sequence | Antigenicity (Vaxijen2.0) | Allergenicity (Aller-Top) | 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            |

Table 2: Prediction of linear B-cell epitope. Similar epitopes were selected by servers

| Epitope                        | Antigenicity | Allergenicity |
|--------------------------------|--------------|---------------|
| ELQDHNE                        | 0.9331       | Non-allergen  |
| DEAGSKSPIQYIDIGN               | 0.9110       | Non-allergen  |
| AGTTQTACTDDNALAYNTTKGG         | 0.5741       | Non-allergen  |
Table 3 T-cell epitopes binding to MHC I and MHC II. 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.

| Epitope                     | Antigenicity | Allergenicity |
|-----------------------------|--------------|---------------|
| FWFSLELQDHNETCH             | 1.1195       | Non-allergen  |
| IYSLLLCRMNSRNYI             | 0.6011       | Non-allergen  |
| VRCSFYEDFLEYHDV             | 0.7525       | Non-allergen  |
| TTQTACTD                   | 0.6570       | Non-allergen  |
| FPFTIYSLL                 | 0.8834       | Non-allergen  |
| IELSLIDFY                  | 1.7424       | Non-allergen  |
| LYFIKGLNNLNRGMV           | 0.5997       | Non-allergen  |
| DNAALAYY                   | 0.631        | Non-allergen  |
| NVFAFPFTI                | 0.8365       | Non-allergen  |
| VLGSLAATV             | 0.6548       | Non-allergen  |

Table 4 Results of the model refinement

| Model                  | GDT-HA | RMSD  | MolProbity | Clash score | Poor rotamers | Rama favored |
|------------------------|--------|-------|------------|-------------|---------------|--------------|
| Initial model          | 1.000  | 0.000 | 3.712      | 24.1        | 17.9          | 64.2         |
| Refined model 1        | 0.8951 | 0.542 | 2.678      | 33.6        | 0.3           | 84.4         |
| Refined model 2        | 0.9043 | 0.550 | 2.792      | 33.8        | 1.4           | 84.1         |
| Refined model 3        | 0.8982 | 0.539 | 2.795      | 33.2        | 1.4           | 83.5         |
| Refined model 4        | 0.8906 | 0.551 | 2.763      | 31.9        | 1.4           | 84.4         |
| Refined model 5        | 0.8951 | 0.549 | 2.674      | 32.8        | 0.7           | 84.1         |

Table 5 Coefficient Weights formula and docking score in Clus pro server for selected models using Antibody-mode. E=0.50Erep+-0.20Eatt+600Eelec+0.25EDARS
| Selected model | Representative | Weighted Score Kcal.mol\(^{-1}\) |
|----------------|----------------|----------------------------------|
| MHC I          | Center         | -392.7                           |
|                | The Lowest Energy | -434.3                      |
| MHC II         | Center         | -494.2                           |
|                | The Lowest Energy | -515.1                      |
| TLR 3          | Center         | -479.0                           |
|                | The Lowest Energy | -602.1                      |
| TLR 4          | Center         | -402.4                           |
|                | The Lowest Energy | -567.3                      |

**Figures**

1. Protein selection (NCBI)
2. Selected protein screening
   Antigenicity (vaxijen2.0)
   Allergenicity (Aller-Top)
   Toxicity (Toxin-pred)
3. Epitope selection
   B-cell (IEDB/BepiPred2.0)
   T-cell (IEDB-MHC I and MHCII)
4. Selected epitope screening
   Antigenicity (vaxijen2.0)
   Allergenicity (Aller-Top)
   Toxicity (Toxin-pred)
5. Peptide vaccine designing
   Linkers
   Adjuvant
6. Biochemical studies of designed peptide
   Stability
   physicochemical properties
7. Secondary and 3D structure of constructed vaccine
8. Molecular docking (Clus pro)
9. In silico cloning (Snapp-gene)

**Figure 1**
Schematic representation of steps for designing of the peptide vaccine against SARS-CoV-2

Figure 2

Schematic representation of the designed peptide vaccine. The designed sequence consisted of four components including CTXB as an adjuvant, PAPAP, and KK as linkers, B-cell, and T-cell related epitopes.

Figure 3

Estimated half-life:
The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>30 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).
A) Physicochemical properties of the constructed peptide by the PepCalc server. Analyses showed that the net charge of the peptide is 17.3, and the pI value is about 9.62. B) Analysis of the amino acid composition and estimation of the half-life of the protein using the ProtParam database. C) The stability of protein was analyzed using Iupred 2a by studying the protein disorders of the peptide sequence. In this graph, the blue line indicates the ANCHOR2, and the red line indicates the IUPred2. According to the red line on the presented graph, the results confirmed the stability of the protein.

Figure 4

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%).
Figure 5

The predicted tertiary structure of the constructed vaccine by the I-TASSER server
Figure 6

Ramachandran plot for constructed peptide after refinement. A) Ramachandran plot for the initial model that showed 59.1% of residues located in most favored regions and 31.2% were 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 showed 78.7% of residues located in the most favored regions. In this regard, the designated additional allowed, generously allowed, and disallowed regions residues were 15.9%, 2.3%, and 3.0%, respectively.
Figure 7

Docking results for the predicted 3D structure. A) 3D structure after docked the peptide vaccine and MHC I, B) MHC II, C) TLR3, and D) TLR4
Figure 8

The map of PET 21a expression vector. The inserted fragment has shown in orange color and labeled as the COVID vaccine.