Designing A Candidate Multi-Epitope Vaccine Against SARS-CoV-2 Using Reverse Vaccinology Approach

Amir Atapour (✉ amir.atapoor58@yahoo.com )
Shiraz University of Medical Sciences

Ali Golestan
Shiraz University of Medical Sciences

Research Article

Keywords: COVID-19, Vaccine, Reverse Vaccinology, Multi-epitope, Molecular docking, MD Simulation

DOI: https://doi.org/10.21203/rs.3.rs-134740/v1

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

Coronavirus 2019 (COVID-19) infection is a global epidemic that is spreading dramatically from day today. Currently, many efforts have been made against COVID-19 through the designing or developing of specific vaccines or drugs, worldwide. Unfortunately, to date, it has not been successful. Therefore, an effective vaccine against COVID-19 is mandatory. In this study, we used the bioinformatics approach to design an effective multi-epitope vaccine against COVID-19 based on Spike protein. Here, we implemented in silico tools to identify potential T and B cell epitopes that can induce cellular and humoral immunity. Then, the peptide sequence of potential T, B cell epitopes, and flagellin (as an adjuvant molecule) was joined together by suitable linkers to construct of candidate multi-epitope vaccine (MEV). Subsequently, immunological and structural evaluations such as antigenicity, 3D modeling, etc. were performed. In the following, molecular docking of vaccine constructs with Toll-Like Receptors 5 (TLR5), Molecular Dynamics (MD) simulation as well as in silico cloning were carried out. Immunological and structural computational data showed that designed MEV potentially has proper capacity for inducing cellular and humoral immune responses against COVID-19. Based on the preliminary results, in vitro and in vivo experiments are required for validation in the future. Keywords: COVID-19, Vaccine, Reverse Vaccinology, Multi-epitope, Molecular docking, MD Simulation.

Introduction

A novel strain of coronavirus, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), in December 2019, appeared in Wuhan City, Hubei province, China. It was first seen at the seafood Huanan store, which may be due to contact with the animal and then transmitted very quickly from human to human throughout China. The new strain named by the World Health Organization (WHO) as 2019 novel coronavirus (2019-nCoV) or coronavirus disease 2019 (COVID-19) and announced . On January 30th, 2020, due to rapid human-to-human transmission of the 2019-nCoV disease and reporting of laboratory-confirmed infection cases in some countries, WHO declared a “public health emergency of international concern (PHEIC) (1, 2). Some members of the Coronaviridae family result in mild respiratory disease in humans. On the other hand, the number of coronavirus strains such as severe acute respiratory syndrome-associated coronavirus (SARS-CoV) and the Middle East respiratory syndrome-associated coronavirus (MERS-CoV) cause severe respiratory diseases in the human population, SARS disease and MERS disease, respectively (3-5). Coronaviruses (CoVs) genome is single-stranded RNA (∼26–32 kb in length) and positive-sense. Human CoVs belong to the Coronaviridae family and they are enveloped viruses. The structural proteins of mature coronaviruses are the spike, envelope (E), membrane (M) and nucleocapsid (N). The spike protein plays important roles including receptor recognition, virus attachment and entry into human cells by attaching angiotensin-converting enzyme 2 (ACE2) as an entry receptor (6-8). At present, there is no approved vaccine or cure for the treatment of COVID-19 infection. According to the increasing spread of COVID-19 infection, the design and development of an effective vaccine or antiviral drugs against this infection are very crucial. The development of conventional (traditional) vaccines has some disadvantages such as costly, need months to years for development
and in vitro culture of the target virus, maybe there be concerns about biologically safety and instability in diverse storage condition (9-12). In contrast, designing of new vaccine approaches such as multi-epitope vaccines (MEVs) that contain B cell epitopes, cytotoxic T lymphocyte (CTL) and helper T lymphocyte epitopes (HTL) by in silico approaches (bioinformatics tools) overcome the above-stated problems. These multiepitope vaccines use some benefits compared to others, including cost-effectiveness, high specificity and stability in different temperature conditions (13-16). According to recent advances in sequencing methods and access to the sequences of many pathogenic genomes as well as the availability of protein sequence databases, bioinformatics-based approaches have attracted much attention from researchers. In the beginning, the selection of suitable target antigens and the prediction of the immunodominant B-cell and T-cell epitopes is a very significant step in the development of an ideal multi-epitope vaccine (17-20). In this context, the bioinformatics tools can play an important role in the prediction of the epitopes. The predicted immunodominant epitopes should activate an effective response against a targeted infection. Thus, it seems that a suitable designed multi-epitope vaccine can be an ideal approach for the prevention and treatment of infections such as this emerging infection (21-25). Because of the low immunogenicity of multi-epitope vaccines, the incorporation of some adjuvants such as Toll-like receptor (TLR) agonists could be a suitable strategy. TLRs as a class of the pathogen recognition receptors (PRR) are conserved type I transmembrane receptors that are expressed on the immune and non-immune cells. TLRs respond to specific molecules of pathogens known as pathogen-associated molecular patterns (PAMPs) that lead to adjust the induction of more effective immune responses (26-29). Bacterial cell wall flagellin (fliC), as TLR5 agonist, can activate both innate and adaptive immune systems by interacting with the TLR5, which has been used as an effective adjuvant in different candidate vaccines (30-32). On the other hand, various studies showed that spike protein is a valuable target for the development of vaccines and therapeutic approaches against COVID-19 infection (33-37). Therefore, in this study, we aim to use different bioinformatics methods to design a potential multi-epitope vaccine consists of several T-cell, B cell epitopes derived from the S protein of SARS-CoV-2, as the target antigen. Further, fliC was used as a biologic adjuvant in the final construct to increase vaccine immunogenicity.

**Results**

**Protein sequence retrieval.** In this research, the extracellular domain of spike glycoprotein (from SARS-CoV-2) amino acid sequence (without signal sequence) was chosen as virus target antigen and then was subjected to in silico analysis to predict potential B cell, CTL and HTL epitopes. Flagellin (fliC) protein sequence (5-143 amino acid) from Salmonella typhimurium as an agonist of TLR5 was selected as an adjuvant for raising the efficacy of vaccines.

**Linear B-cell epitope prediction.** Linear B-cell epitopes with 14 amino acid length, in S protein of SARS-CoV-2, were selected for the final vaccine construct. Linear B-cell epitopes with the highest score were shown in **Table 1**.
**Helper T Lymphocytes (HTL) epitopes prediction.** Prediction of high scored MHC-II epitopes (as HTL epitopes) for mouse H2 class II alleles were performed using NetMHCII 2.3 web server based on their IC50 scores. Finally, for generating the final MEV construct several 10 high score HTL epitopes were selected (Table 2).

**Cytotoxic T Lymphocyte (CTL) epitopes prediction.** Some 37 CTL (9-mer) epitopes were predicted for S protein using NetCTL 1.2 server set at the default threshold. Of these, eight high-scored CTL epitopes were selected in the final multi-epitope vaccine construct (as listed in Table 3).

**Design and construction of multi-epitope vaccine candidate sequence.** To generate the vaccine candidate sequence, selected high-scoring CTLs, high-affinity HTLs epitopes, as well as B-cell epitopes were linked to each other using suitable linkers (GPGPG, and AYY). Domains D0/D1 of flagellin as an adjuvant was added at the N-terminal of multi-epitope construct sequence using EAAAK linker to improve the immunogenicity of the vaccine construct. UniProt database (http://www.uniprot.org/) was used to achieve the flagellin domain sequences. Also, at the C-terminal of the multi-epitope construct sequence six histidine amino acid residues (as a His-tag) were included for the purification process in the future. Finally, a vaccine construct with 536 amino acid residues was designed. The schematic diagram of the final vaccine construct is indicated (Fig 1).

**Physiochemical parameters and solubility prediction.** The theoretical isoelectric point value (pI) and Molecular weight (Mw) of the final protein were predicted 6.07 and 55.26 kDa, respectively. The half-life was calculated to be 30 hours in mammalian reticulocytes (in vitro) >20 hours in yeast (in vivo) and >10 hours in Escherichia coli (in vivo). GRAVY and aliphatic index were estimated -0.343 and 65.00, respectively. An instability index (II) with a 21.67 score was estimated. The vaccine protein was calculated to be soluble upon expression with a solubility score of 0.557194, in an *E. coli* host.

**IFN-γ inducing epitope prediction.** Identification of the IFN-gamma inducing epitopes from MHC-II binding epitope fragments in the final vaccine construct was performed using the IFNepitope server. The numbers of 5 epitopes were selected as IFN-gamma inducing epitopes (Table 4).

**Antigenicity and allergenicity of the vaccine construct.** The antigenicity of the whole vaccine sequence (including the adjuvant sequence) was estimated by the VaxiJen 2.0 and ANTIGENpro servers to be
0.7751 with a bacteria model at a threshold of 0.5 and 0.931486, respectively. The allergenicity prediction results using both the AllerTOP v.2 and AlgPred servers indicated that the constructed vaccine sequences to be non-allergenic.

**Secondary structure prediction.** Based on PSIPRED data, the prediction of the secondary structure of the final protein vaccine was estimated to contain 47.26% alpha-helix, 5.74% beta-strand and 47% coil (Fig 2). The application of the predicted secondary structure is to the refinement of the tertiary structure of a protein.

**3D structure homology modeling and validation.** Various servers were implemented for 3D structure modelings such as Phyre2, SWISS-MODEL and I-TASSER. Based on primary validation results, c3k1hA_.20 model from Phyre2 server (Fig 3) was selected as the best model to follow the protein construction process. In the selected model, the ProSA z-score was -2.95 (Fig 4A) and ERRAT, the overall quality factor of the selected model was 100% (Fig 4B). The PROCHECK’s Ramachandran plot analysis of the selected model revealed that 94.8%, 5.2% and 0.0% of residues are located in favored, allowed and outlier regions, respectively (Fig 4C).

**Conformational (Discontinuous) B-cell epitopes prediction.** Conformational epitopes play an important role in the humoral response. High-rank residues as conformation epitopes were determined in the 3D model of the final MEV construct. Here, discontinuous peptides with a value of 0.7 or higher were chosen (Table 5). Additionally, the predicted discontinuous epitopes in the 3D structure of the final multiepitope protein are illustrated (Fig 5).

**Molecular docking of designed vaccine with TLR5.** Molecular interaction between the final MEV model and TLR5 was determined using the ClusPro server. The docked structure with the lowest energy score was selected as the best-docked complex. Therefore, based on global free energy, the best possible docked complex and with the highest binding affinity (total free energy -992.4) was chosen. The best-docked model of MEV and TLR5 complex is shown (Fig 6). PyMOL 1.1eval was used for analysis and visualization of the docked complex.

**Molecular dynamics (MD) simulations.** Vaccine flexibility is a key aspect of its performance. Therefore, in this study, we have analysis the flexibility of the designed vaccine by CABS- Flex 2.0 with 50 cycles’
simulation at 1.4 °C temperature. Our complex structure achieved a high level of fluctuations in the residue positions 100, 39 and 150 were recognized to be 2.64 Å, 2.09 Å and 1.99 Å, respectively (Fig 7A).

Also, CABS-Flex 2.0 software provides 10 different models based on parameters such as structural heterogeneity, their optimum free energy, and highly stable configuration, that the first model was selected. Stable protein complex structure after MD simulations using CABS-Flex 2.0 showed (Fig 7B).

In silico cloning and codon optimization of MEV construct. Reverse translation and codon optimization of nucleotide sequence was done using the JCAT tool for efficient protein expression in E. coli. In this study, parameters like codon adaptive index (CAI) and GC content of our optimized nucleotide sequence were 0.93 and 55.59, respectively. These parameters were represented as a good adaptation, which permitted the high rate of expression of the MEV construct in E. coli K12. Finally, the optimized sequence (1608 nucleotides) was cloned into a pET28a expression vector using Xhol and Ncol restriction sites. Also, a poly histidine-tag (6xHis-tag) at the C-terminus of the multiepitope protein was incorporated for purification purposes (Fig 8).

Discussion

Currently, the Wuhan Novel Coronavirus 2019 (2019-nCoV) or (COVID-19) disease is now widespread in all countries and has become a global pandemic that has led to death in infected people (38). Although efforts are ongoing, to date, no specific antiviral drug or vaccine has been introduced to control the COVID-19 disease. Today, researchers are conducting extensive studies to develop multi-subunit vaccines as one of the most effective methods of vaccination because of their benefits (39, 40). Therefore, in this study, we use the methods of bioinformatics designed a potential candidate epitope-based vaccine against the SARS-CoV-2 based on the spike protein, which is one of the major antigenic proteins of SARS-CoV-2 and viral entry into the host cell (3). For this purpose, we first selected surface glycoprotein (S protein) encoded by SARS-CoV-2 as the target antigenic protein for further analysis. T cells (helper T cells and cytotoxic T cells) are known to be the main mediators of cell-mediated immunity. In this way, antigens are identified by cytotoxic T cells (CTLs), whereas helper T cells activate B cells, macrophages, and even cytotoxic T cells. In the case of intracellular pathogens such as viruses, cellular immunity identify and destruct infected cells by secreting antiviral cytokines and creating a life-long immunity (18, 41, 42). Furthermore, in the following, prediction of B-cell and T-cell possible epitopes from S protein was done to construct the potential multi-epitope vaccine (MEV) to induce both cell-mediated and humoral response. Then, to construct the MEV, the predicted epitopes fused using suitable linkers (AAY and GPGPG linkers), as specialized spacer sequences. The AAY linkers play a role in increase epitope presentation and remove junctional epitopes. In the about of GPGPG linkers, these types of linkers can cause stimulate T-helper responses and conformational dependent immunogenicity of helper, as well as antibody epitopes (43-46). Also, to overcome the low immunogenicity challenge of these vaccines, D0/D1 domains of flagellin protein from S. Typhimurium bacteria were used as an adjuvant to enhance
effectively stimulation of immune system responses (27, 32). D0/D1 domains fused at the N-terminal region of the multi-epitope sequence using the EAAAK linker. The EAAAK linker help to decrease interference adjuvant part with other protein segments by effective separation and improve the level of expression and bioactivity of the target fusion protein (47). Finally, a candidate vaccine with a length of 536 amino acids, including some linear B-cell, CTL and HTL epitopes fused to the adjuvant sequence was constructed. Moreover, we predicted three interferon-γ (IFN-γ) epitopes in the final multi-epitope construct. IFN-γ secretion has been shown to stimulate innate immune responses and may directly eliminate viral replication (48-50). Therefore, it seems that IFN-gamma inducing epitopes can be effective in vaccine development. MEV was predicted to be antigenic with the probability of antigenicity 0.900656 and non-allergenic. This means that our MEV has potentially the ability to produce a strong immune response without an allergic reaction, so, make it a potent vaccine. The analysis of the physicochemical properties of the structure was as follows: Molecular weight (MW) of the MEV construct was 55650.26 kDa. The theoretical pI value was 6.07, indicating that MEV is acidic in nature. The amount of light that is absorbed in a certain wavelength can be explained as an extinction coefficient index for a particular compound. The extinction coefficient of the construct was 78620M-1 cm-1. The instability index (II) score computed 21.67 which, indicated the candidate vaccine protein as a stable protein (II of >40 indicates instability) (51). The aliphatic index of a protein is related to the relative volume occupied by aliphatic amino acids (alanine, valine, isoleucine, and leucine) in the protein side chains. It may be regarded as a positive factor for the increase of the thermostability of globular protein. The aliphatic index of the construct was 65.00, which was shown as a thermostable protein. The GRAVY (Grand Average of Hydropathy) value for a peptide or protein is calculated as the sum of hydrophathy values of all the amino acids, divided by the number of residues in the sequence. The positive and negative values represent the hydrophobic and hydrophilic properties of a compound, respectively. Our designed construct had the GRAVY value of -0.343 that predicted construct is hydrophilic protein (51). The half-life is a prediction of the time it takes for half of the amount of protein in a cell to disappear after its synthesis in the cell. ProtParam tool predicted the half-life of our construct in the following; 30 hours (mammalian reticulocytes, in vitro), >20 hours (yeast, in vivo) and >10 hours (Escherichia coli, in vivo). For many biochemical and functional evaluations of recombinant proteins, the solubility overexpression in the *E. coli* host is one of the necessary requirements that help in the efficient purification process, in later stages (51). Here, the solubility upon overexpression of our multi-epitope protein construct predicted (by SOLpro server) with probability 0.557194, which indicates the overexpression of our multi-epitope protein in *E. coli*, insoluble form (52). In this research, analysis of the secondary structure using the PSIPRED method illustrated that our protein construct composed predominantly of alpha-helical (47.26%), 47% coils and 5.74% of the amino acids in strand formation. In this study, the tertiary structure of the vaccine construct was obtained by various servers (SWISS-MODEL, phyre2 and I-TASSER). Validation process, to identify the potential errors and improve the quality of the predicted 3D model was done by ProSA-web, RAMPAGE and ERRAT servers. Validation data showed that the c3k1hA_.20 model generated by the phyre2 server was selected as a final 3D model. Validation output indicated that the selected model (c3k1hA_.20) had a high quality and does not require a refinement process. According to the Ramachandran plot, most of the
residues are located in the favored (94.8%) and 5.2% in allowed regions. In vivo, it expected that interaction between designed MEV constructs with the TLR5 on professional antigen-presenting cells (APCs), inducing a potentially protective immune response against the virus. For this proposal, immune interaction between the designed MEV construct and TLR5 receptor was performed using Cluspro server. Cluspro presented dozens of models which docked models scored based on hydrophobicity, geometry and electrostatic complementarity of protein surfaces. Here, we were selected the best possible docked model between hydrophobicity models. So, the docked structure with the lowest energy score (-992.4) was selected as the best-docked complex. In this research, CABS-Flex 2.0 software was used for MD simulation. CABS-Flex presents the stable arrangement of the TLR5-designed vaccine complex. Based on the root mean square fluctuation (RMSF) values (using CABS Flex 2.0), the fluctuation of the individual amino acid residues as described. The highest RMSF value and the lowest value indicates more fluctuation and low fluctuation of our complex structure during the simulation process, respectively. Fluctuations in the structure of the MEV indicate its high flexibility and validate it as a potential structure of the vaccine (53). The codon optimization process was performed to reach high-level expression and translation efficacy of our multi-epitope protein in E. coli (strain K12). Here, CAI value (0.93) and GC content (55.59%) parameters were obtained which showing a possible higher expression of the protein vaccine within the E. coli K-12 system (54, 55). Finally, the pET-28a vector including the MEV sequence was constructed which should efficiently and effectively encode the MEV protein in the E. coli cells. Based on the results of this study, to develop our candidate vaccine against COVID-19, we recommend that validation assays containing in vitro and in vivo analysis be performed in the future.

Methodology

Retrieving of SARS-COV-2 spike protein and flagellin sequences. In the first stage of the research, the amino acid sequence of Spike protein of SARS-COV-2 and flagellin (flIC) protein from Salmonella enterica serovar Typhimurium (S. Typhimurium) were retrieved from the UniProt database (uniport.org) (P0DTC2, Q66PQ5, respectively).

Prediction of B-cell epitopes. The identification of B-cell epitopes as potential antigens recognized by the surface receptor of B-cell lymphocytes leads to generate a specific immune response. Thus, B-cell epitopes play a key role in vaccine design. The BCPREDS server (56) allows users to choose the method for predicting B-cell epitopes among several developed prediction methods. Herein, the length of 14-mer B-cell epitopes and Specificity of 75% were selected.

Prediction of CTL epitopes. The freely-accessible NetCTL 1.2 server (57) was used for CTL epitopes prediction. This method integrates prediction of peptide MHC class I binding, proteasomal C terminal cleavage and TAP (Transporter Associated with Antigen Processing) transport efficiency. The server
allows for predictions of CTL epitopes restricted to 12 MHC class I supertype. In this study, we have chosen the A1 supertype and threshold value set at the default score (0.75) for the prediction of CTL epitopes.

**HTL epitopes prediction.** The amino acid sequence of S protein subjected to NetMHCII 2.3 Server (58) for screening HTL epitopes with 15-mer length. This server using an artificial neuron network predicts binding of MHC II epitopes for human (HLA-DR, HLA-DQ, HLA-DP) and mouse MHC class II alleles (H-2). Herein, we selected mouse H2 class II alleles (H-2-IAb, H-2-IEd, and H-2-IAd). The prediction values inferred from the IC50 values, and as a percentile-rank to a set of random natural peptides. Strong and weak binding peptides are illustrated in the output. An IC50 values <50 nM, <500 nM and <5000 nM indicates high-affinity, intermediate affinity and low affinity, respectively.

**Multi-epitope vaccine candidate construction.** The selected epitope candidates including high-scoring CTLs, high-affinity HTLs epitopes and B-cell epitopes were joined together using AAY and GPGPG linkers (47) to produce the multi-epitope vaccine (MEV) construct. Flagellin (FliC) as an adjuvant was chosen and was included at the N-terminal region of the MEV construct by an EAAAK linker (47).

**Interferon-gamma (IFN-γ) inducing epitope prediction.** One of the signatures of the innate and adaptive immune systems is interferon-gamma (IFN-γ) which has an antiviral function, immune regulation as well as anti-tumor activity. IFN-γ cytokine, as the major arm of the Th1 response, has a key role in the control of intracellular pathogens such as viruses. IFNepitope server (http://crdd.osdd.net/raghava/ifnepitope/scan.php) (59) developed to predict and design IFN-gamma inducing MHC class II binder peptides from target proteins to design an effective subunit vaccine. The Support Vector Machine (SVM) hybrid algorithms along with Motif was used for the prediction of IFN-γ epitopes. This server is based on a dataset including IFN-γ inducing and non-inducing MHC class-II binder, which has the potential to activate T-helper cells.

**Prediction of antigenicity of the designed vaccine construct.** To analyze antigenicity of multi-epitope vaccine construct (contain the flagellin adjuvant sequence), two servers including Vaxijen 2.0 (Vaxijen 2.0 is freely available online at the URL: http://www.jenner.ac.uk/VaxiJen) (60) and ANTIGENpro (ANTIGENpro is integrated into the SCRATCH suite of predictors available at http://scratch.proteomics.ics.uci.edu) (61) were used. The base of the VaxiJen 2.0 server is on auto- and cross-covariance (ACC) transformation of target protein sequences into uniform vectors of principal amino acid properties. The prediction approach used in VaxiJen v2.0 is alignment-free and is based on various physicochemical properties of the protein.
The threshold value of Vaxijen 2.0 was set at 0.5. ANTIGENpro server for checking protein antigenicity applies protein antigenicity microarray data and illustrate an antigenicity index. It was estimated the accuracy of the ANTIGENpro server to be 76%, based on cross-validation experiments and using the combined dataset.

**Allergenicity prediction of designed vaccine construct.** AllerTOP V2.0 ([http://www.ddg-pharmfac.net/AllerTOP](http://www.ddg-pharmfac.net/AllerTOP)) \(^{(62)}\) and Algpred ([Available at: http://webs.iiitd.edu.in/raghava/algpred](http://webs.iiitd.edu.in/raghava/algpred)) \(^{(63)}\) servers were used to predict the allergenicity nature of the multi-epitope vaccine construct. In AllerTOP V2.0 server various methods are employed for classification of allergens such as amino acid E-descriptors, auto- and cross-covariance transformation, and the k nearest neighbors (kNN) machine learning. In this method, an accuracy of 85.3% has been presented at 5-fold cross-validation. Algpred allows the prediction of allergens based on the similarity of known epitope with any region of the protein. In AlgPred a systematic attempt has been made to integrate various approaches to predict allergenic proteins with high accuracy. In this study, a hybrid approach of the server, using a combined method (SVMc + IgE epitope + ARPs BLAST + MAST), was used to predict allergen.

**Different physicochemical properties and solubility analysis of vaccine construct.** Different physicochemical parameters of MEV such as amino acid composition, theoretical pl, Grand Average Hydropathy, Stability Profiling, Instability Index, Molecular Weight, Half-Life, and Aliphatic Index were determined using ProtParam ([http://web.expasy.org/protparam/](http://web.expasy.org/protparam/)) \(^{(51)}\) online server. SOLpro was used for solubility analysis of MEV. SOLpro is integrated in the SCRATCH suite of predictors available at [http://scratch.proteomics.ics.uci.edu](http://scratch.proteomics.ics.uci.edu) \(^{(52)}\). SOLpro predicts if the protein is soluble or not (SOLUBLE/INSOLUBLE) and gives the corresponding probability (≥ 0.5).

**Secondary structure prediction of vaccine construct.** The PSIPRED server was used for our vaccine protein secondary structure prediction. PSIPRED, a web-based freely-accessible online server ([http://bioinf.cs.ucl.ac.uk/psipred/](http://bioinf.cs.ucl.ac.uk/psipred/)) \(^{(64)}\), is based on primary amino acid sequences input in a precise manner. In this method, a very stringent cross-validation approach use to evaluate the method’s efficiency. PSIPRED 3.2 server combines two feed-forward neural networks that do an analysis of output obtained from PSI-BLAST (Position-Specific Iterated - BLAST).

**Tertiary structure prediction.** Three different servers including I-TASSER, SWISS-MODEL and Phyre2 were used to achieve the best 3D structural models. In the following, Phyre2 software was selected. Phyre2 is available at [http://www.sbg.bio.ic.ac.uk/phyre2](http://www.sbg.bio.ic.ac.uk/phyre2) \(^{(65)}\). This server is one of the most widely used protein
structure prediction servers which uses advanced remote homology detection methods to build 3D models.

**3D structure validation.** Given the importance of model validation and to find the possible errors in primary 3D structure models, we utilized 3 tools for model validation. ProSA-web at https://prosa.services.came.sbg.ac.at/prosa.php, ERRAT server at http://nihserver.mbi.ucla.edu/ERRAtv2/ (66) and PROCHECK's Ramachandran plot analysis at https://servicesn.mbi.ucla.edu/PROCHECK/ (67) were used. The overall quality of a specific input structure calculates using ProSA-web (68) and is presented as a quality score. ProSA-web provides an easy-to-use interface to the program ProSA which is frequently used in protein structure validation. The ProSA-web score is shown in a plot, which including the z-score of experimentally determined structures deposited in PDB. ERRAT program is used for verifying protein structures determined by crystallography. This program is based on the statistics of non-bonded atom-atom interactions in the reported structure. PROCHECK's Ramachandran plot program checks the stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry.

**Discontinuous B-cell epitope prediction in final MEV construct.** ElliPro (http://tools.iedb.org/ellipro/) server (69) was employed for the prediction of conformational B-cell epitopes from the validated 3D structure. ElliPro is a new web-tool for the prediction of antibody epitopes based on the geometrical properties of protein structure. ElliPro considers a score to each output epitope defined as a PI (Protrusion Index) value averaged over epitope residues. In the method, the protein's 3D shape is approximated by several ellipsoids, thus that the ellipsoid with PI = 0.9 would include within 90% of the protein residues with 10% of the protein residues being outside of the ellipsoid.

**Molecular docking of designed vaccine with TLR-5 receptor.** The interaction patterns of the final MEV construct as a ligand with the TLR5 receptor (PDB ID: 3J0A) were analyzed by using ClusPro 2.0 server (https://cluspro.org) (70). ClusPro is a web-based server for the direct docking of two interacting proteins. Three computational steps including (1) rigid-body docking, (2) RMSD based clustering of the 1000 lowest energy structures, and (3) the removal of steric clashes by energy minimization performs using ClusPro software.

**Molecular dynamics (MD) simulations.** CABS-flex webserver was used for fast simulations of TLR5-designed vaccine complex flexibility. CABS-flex 2.0 is freely available at http://biocomp.chem.uw.edu.pl/CABSex2 (53). The only data required as input is a protein structure in
the PDB format (or a protein PDB code). Here, the selected docked TLR5-vaccine complex was used as the input for the MD simulation. In this server, the protein flexibility, contact map and root-mean-square fluctuations (RMSFs) of atoms in a protein complex are presented. CABS-flex server show RMSF simulation of all amino acid residues provide in a particular protein, in nanosecond time.

In silico cloning and codon optimization of MEV construct. The Java Codon Adaptation Tool (JCat) ([http://www.prodoric.de/JCat](http://www.prodoric.de/JCat)) (71) was used for codon optimization and reverse translation of the MEV construct sequence to express the MEV in a suitable expression vector. Here, an E. coli (strain K12) host was selected to express the final MEV sequence and other options such as rho-independent transcription termination, prokaryote ribosome binding site, and restriction enzyme cleavage sites were included. To ensure the high-level protein expression, two JCat output indexes include the codon adaptation index (CAI) and percentage GC content can be used. CAI supply data on codon usage biases; the CAI score is optimal at 1.0, but a score above 0.8 (> 0.8) is also good. Usually, the optimal GC content range between 30–70% is favorable effects on translational and transcriptional efficiencies. Also, to clone the final MEV sequence in E. coli pET-28a (+) vector, two restriction enzyme sites contain XhoI and Ncol were included at N and C-terminals of the target sequence, respectively. To ensure MEV expression, the optimized MEV sequence (with restriction sites) was inserted into the pET-28a (+) vector using the SnapGene tool.

Conclusion

Nowadays, the Wuhan Novel Coronavirus 2019 outbreaks worldwide and has become a global problem. Unfortunately, due to a lack of antiviral drugs and any vaccine against COVID-19, morbidity and mortality are increasing. Therefore, the prevention and control of this infection are very mandatory. In this study, we have tried to make a multi-epitope vaccine (MEV) against the Wuhan Novel Coronavirus 2019 using immunoinformatics methods. Thus, in addition to performing fewer experiments and errors, saving time and costs, in silico methods can be used to design and develop safe and potential MEV vaccines. Here, different computational tools were used for the design of MEV based on immunogenic epitopes (B and T cells) from the antigenic spike (1) protein of COVID-19. In this work, study results revealed that our vaccine construct might confer proper immunogenic responses. Furthermore, in vitro and in vivo studies is necessary to ensure the efficacy and safety of the proposed vaccine.

Declarations

Conflict of Interest

The authors have no conflicts of interest to declare.

Acknowledgments
The authors wish to thank Shiraz University of Medical Sciences for supporting the conduct of this research.

References


46. Saadi M, Karkhah A, Nouri HR. Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. Infection, Genetics and
Evolution. 2017;51:227-34.


### Tables

<table>
<thead>
<tr>
<th>Start-end position</th>
<th>Sequence</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1067</td>
<td>YVPAQEKNFTTAPA</td>
<td>0.988</td>
</tr>
<tr>
<td>596</td>
<td>SVITPGTNSTSNQVA</td>
<td>0.982</td>
</tr>
<tr>
<td>14</td>
<td>QCVNLTTTRTQLPPA</td>
<td>0.963</td>
</tr>
<tr>
<td>114</td>
<td>TQSLLIVNATNVV</td>
<td>0.951</td>
</tr>
<tr>
<td>516</td>
<td>ELLHAPATVCGPKK</td>
<td>0.934</td>
</tr>
</tbody>
</table>

**Table 1.** List of the final B-cell epitopes from S protein of SARS-CoV-2 by BCPREDs server.
<table>
<thead>
<tr>
<th>Allele</th>
<th>HTL epitopes</th>
<th>1-log50k(aff)</th>
<th>affinity(nM)</th>
<th>%Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-2-IAd</td>
<td>QQLIRAAEIRASANL</td>
<td>0.5960</td>
<td>79.2</td>
<td>0.50</td>
</tr>
<tr>
<td>H-2-IAd</td>
<td>TQQLIRAAEIRASAN</td>
<td>0.5958</td>
<td>79.3</td>
<td>0.60</td>
</tr>
<tr>
<td>H-2-IAb</td>
<td>QMAYRFNGIGVTQNV</td>
<td>0.5904</td>
<td>84.1</td>
<td>0.40</td>
</tr>
<tr>
<td>H-2-IAb</td>
<td>QEKNTTAPAIChDG</td>
<td>0.5899</td>
<td>84.6</td>
<td>0.40</td>
</tr>
<tr>
<td>H-2-IAb</td>
<td>MQMAYRFNGIGVTQN</td>
<td>0.5857</td>
<td>88.5</td>
<td>0.40</td>
</tr>
<tr>
<td>H-2-IAb</td>
<td>DSSSGWTAGAAAAYV</td>
<td>0.5849</td>
<td>89.3</td>
<td>0.40</td>
</tr>
<tr>
<td>H-2-IAb</td>
<td>MAYRFNGIGVTQNVL</td>
<td>0.5772</td>
<td>97.0</td>
<td>0.40</td>
</tr>
<tr>
<td>H-2-IAb</td>
<td>SSSGWTAGAAAYVG</td>
<td>0.5771</td>
<td>97.1</td>
<td>0.40</td>
</tr>
<tr>
<td>H-2-IAb</td>
<td>GDSSSGWTAGAAAYY</td>
<td>0.5771</td>
<td>97.1</td>
<td>0.40</td>
</tr>
<tr>
<td>H-2-IAd</td>
<td>YVTQQLIRAAEIRAS</td>
<td>0.5776</td>
<td>96.6</td>
<td>0.70</td>
</tr>
</tbody>
</table>

**Table 2.** List of the high score HTL epitopes from S protein of SARS-CoV-2 by NetMHCII 2.3 Server.
<table>
<thead>
<tr>
<th>Position</th>
<th>CTL epitopes Sequence</th>
<th>Prediction score</th>
<th>MHC binding affinity</th>
</tr>
</thead>
<tbody>
<tr>
<td>851</td>
<td>LTDEMIAQY</td>
<td>3.6616</td>
<td>0.7953</td>
</tr>
<tr>
<td>244</td>
<td>HTSSMRGVY</td>
<td>3.1128</td>
<td>0.6735</td>
</tr>
<tr>
<td>590</td>
<td>TSNQAVAVLY</td>
<td>3.0758</td>
<td>0.6559</td>
</tr>
<tr>
<td>347</td>
<td>CVADYSVLY</td>
<td>2.5759</td>
<td>0.5348</td>
</tr>
<tr>
<td>719</td>
<td>KTSVDCTMY</td>
<td>2.3795</td>
<td>0.4908</td>
</tr>
<tr>
<td>732</td>
<td>STECSNLLL</td>
<td>2.3492</td>
<td>0.5136</td>
</tr>
<tr>
<td>182</td>
<td>NIDGYFKIY</td>
<td>1.9606</td>
<td>0.3921</td>
</tr>
<tr>
<td>146</td>
<td>YSSANNCTF</td>
<td>1.9531</td>
<td>0.3975</td>
</tr>
</tbody>
</table>

**Table 3.** List of the high-scored CTL epitopes from S protein of SARS-CoV-2 by NetCTL 1.2 server.

<table>
<thead>
<tr>
<th>Result</th>
<th>Position</th>
<th>Epitope</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Epitope_1</td>
<td>FPSVYAWERKKISNC</td>
<td>1</td>
</tr>
<tr>
<td>Positive</td>
<td>Epitope_3</td>
<td>YNYKYRYLRHGKLRP</td>
<td>1.0637672</td>
</tr>
<tr>
<td>Positive</td>
<td>Epitope_4</td>
<td>LIRAAEIRASANLAA</td>
<td>0.6887053</td>
</tr>
</tbody>
</table>

**Table 4.** IFN-gamma inducing epitopes predicted by IFNepitope server.
<table>
<thead>
<tr>
<th>No.</th>
<th>Residues</th>
<th>Number of residues</th>
<th>Score</th>
</tr>
</thead>
</table>

**Table 5.** Conformational epitopes of the designed MEV as predicted by the ElliPro server.

**Figures**

**Figure 1**

The schematic diagram of the final vaccine construct
Based on PSIPRED data, the prediction of the secondary structure of the final protein vaccine was estimated to contain 47.26% alpha-helix, 5.74% beta-strand and 47% coil.

Figure 3

c3k1hA_.20 model from Phyre2 server
Figure 4

In the selected model, the ProSA z-score was -2.95 (Fig 4A) and ERRAT, the overall quality factor of the selected model was 100% (Fig 4B). The PROCHECK’s Ramachandran plot analysis of the selected model revealed that 94.8%, 5.2% and 0.0% of residues are located in favored, allowed and outlier regions, respectively (Fig 4C).

Figure 6

The best-docked model of MEV and TLR5 complex