Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Vaccine 38 (2020) 7612–7628

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Exploring the out of sight antigens of SARS-CoV-2 to design a candidate multi-epitope vaccine by utilizing immunoinformatics approaches

Ashkan Safavi [a], Amirhosein Kefayat [b], Elham Mahdevar [c], Ardavan Abiri [d], Fatemeh Ghahremani [e,*]

[a] Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran
[b] Department of Oncology, Cancer Prevention Research Center, Isfahan University of Medical Sciences, Isfahan, Iran
[c] Department of Biology, Faculty of Science and Engineering, Science and Arts University, Yazd, Iran
[d] Department of Medicinal Chemistry, Faculty of Pharmacy, Kerman University of Medical Sciences, Kerman, Iran
[e] Department of Medical Physics and Radiotherapy, Arak School of Paramedicine, Arak University of Medical Sciences, Arak, Iran

ARTICLE INFO

Article history:
Received 5 May 2020
Received in revised form 25 August 2020
Accepted 6 October 2020
Available online 9 October 2020

Keywords:
SARS-CoV-2
Immunoinformatics
Vaccine
COVID-19
In silico

ABSTRACT

SARS-CoV-2 causes a severe respiratory disease called COVID-19. Currently, global health is facing its devastating outbreak. However, there is no vaccine available against this virus up to now. In this study, a novel multi-epitope vaccine against SARS-CoV-2 was designed to provoke both innate and adaptive immune responses. The immunodominant regions of six non-structural proteins (nsp7, nsp8, nsp9, nsp10, nsp12 and nsp14) of SARS-CoV-2 were selected by multiple immunoinformatic tools to provoke T cell immune response. Also, immunodominant fragment of the functional region of SARS-CoV-2 spike (400–510 residues) protein was selected for inducing neutralizing antibodies production. The selected regions' sequences were connected to each other by furin-sensitive linker (RVRR). Moreover, the functional region of β-defensin as a well-known agonist for the TLR-4/MD complex was added at the N-terminus of the vaccine using (EAAAK)3 linker. Also, a CD4 + T-helper epitope, PADRE, was used at the C-terminal of the vaccine by GPGPG and A(EAAAK)2A linkers to form the final vaccine construct. The physicochemical properties, allergenicity, antigenicity, functionality and population coverage of the final vaccine construct were analyzed. The final vaccine construct was an immunogenic, non-allergen and unfunctional protein which contained multiple CD8 + and CD4 + overlapping epitopes, IFN-γ inducing epitopes, linear and conformational B cell epitopes. It could form stable and significant interactions with TLR-4/MD according to molecular docking and dynamics simulations. Global population coverage of the vaccine for HLA-I and II were estimated 96.2% and 97.1%, respectively. At last, the final vaccine construct was reverse translated to design the DNA vaccine. Although the designed vaccine exhibited high efficacy *in silico*, further experimental validation is necessary.

1. Introduction

SARS-CoV-2 is a newly identified member of the coronavirus family. It causes a severe respiratory disease called COVID-19 and global health currently is facing its devastating outbreak [1]. SARS-CoV-2 is an enveloped positive-strand RNA virus. Its genome encodes 28 proteins which are divided into three categories including non-structural (nsp), structural, and accessory proteins. The non-structural proteins (nsp1-nsp16) are only produced during the virus RNA translation at the infected host cells. The envelop (E), membrane (M), nucleoprotein (N) and spike (S) are the structural proteins which assemble the virus particle. Also, the accessory proteins (n = 8) play key roles in the virus assembly, virulence and pathogenesis [2]. No specific treatment is approved for COVID-19, until now. Vaccines are one of the most effective tools to control infectious diseases. Recently, immunoinformatics approaches have gained lots of attention for designing vaccines. In the past, vaccine development was completely dependent to immunological experiments which are relatively expensive and time-consuming. However, recent advances in the field of immunological bioinformatics have provided feasible tools which can significantly decrease the time and cost required for vaccine development [3,4].

Adaptive immune response consists of cellular and humoral arms. It has determinative role against viral infections. CD8 + T cells account for about 80% of total infiltrative inflammatory cells in the lungs interstitium of the COVID-19 patients. These cytotoxic cells attack the virus-infected cells. On the other hand, humoral immune produces neutralizing antibodies which play a protective

* Corresponding author.
E-mail address: F.ghahremani@arakmu.ac.ir (F. Ghahremani).

https://doi.org/10.1016/j.vaccine.2020.10.016
0264-410X/

role by limiting infection. Most of the immune responses in the coronavirus infected patients were found to be against the structural proteins, particularly S glycoprotein [5]. The S glycoproteins protrude from the SARS-CoV-2 envelope and play a determinative role in the virus binding and invasion to the target cells through angiotensin-converting enzyme 2 receptor (ACE2) [6]. Many studies have used this glycoprotein to develop COVID-19 vaccine due to its pivotal role and high exposure on the virus surface [7–10]. However, immunogenic parts of these surface antigens are occluded with a dense coat of host-derived glycans. In addition, the surface proteins' glycosylation sites are highly variable during the virus evolution. Taking together, this glycan shield and its potential diversity of modification can lead to immune evasion [11]. To conquer this challenge, we selected the interacting region of the S protein with ACE2 for incorporation to the designed vaccine structure. This region has a high-specific and conserved glycosylation pattern which is suitable as an antibody target. Attaching of neutralizing antibodies to this site of S protein blocks virus binding and entrance to the host cells [12].

On the other hand, T cells' role in the immune response for controlling coronaviruses infections is significantly more vital than B cells. More than 70% of the T cell immune response targets the structural proteins of coronaviruses [5]. Naive T cells are stimulated by antigen-presenting cells (APC). As APCs phagocyte the virus, only the structural proteins of the coronavirus enter the major histocompatibility complex (MHC) processing due to lack of nsps at virus particles. However, two third of the coronavirus genomes encodes nsps and they are highly expressed at the infected cells. The nsps are presented by the infected cells' MHC-I from the first day of infection [13]. Therefore, we incorporated the immunodominant regions of nsps in the designed vaccine to activate the immune system against these antigens which may stay out of the immune system sight.

In this study, we tried to designed a multi-epitope vaccine with high efficacy to boost both T and B cells immune responses against SARS-CoV-2. According to the best of our knowledge, this is the first time to select nsps for designing vaccine against SARS-CoV-2. Also, an immunogenic region of the S protein interacting domain with ACE2 receptor was incorporated into the vaccine construct to cause production of neutralizing antibodies with ability to block virus binding and invasion to the host cells.

2. Methods and materials

2.1. Study design

The present study has eight main steps: (1) Antigen screening; (2) Selection of the immunodominant regions of the selected antigens; (3) Designing the final vaccine construct; (4) Tertiary structure prediction and validation; (5) Molecular docking with TLR-4/MD; (6) Designing the DNA vaccine; (7) Data validation; and (8) Immune response simulation. The workflow of this study is shown as Fig. 1.

2.2. Collection of proteins

The sequences of the structural and non-structural proteins of SARS-CoV-2 were obtained from NCBI database (https://www.ncbi.nlm.nih.gov/protein/). Two SARS-CoV-2 proteins including S glycoprotein and Orf1ab polyprotein (Orf1ab polyprotein is cleaved into nsp1-nsp16 [14]) were selected for designing the candidate vaccine. Orf1ab had significantly higher number of high-binding affinity (<500 nM) MHC-I epitopes (across 9-mer peptides) in comparison with all of the structural proteins of SARS-CoV-2 according to predictions of the consensus method of the TepiTool server (http://tools.iedb.org/tepitool/.) [15].

2.3. MHC-I binding epitopes prediction

MHC-I binding epitopes were predicted for the nsp7, nsp8, nsp9, nsp10, nsp12, and nsp14 proteins. Three servers including NetMHCpan 4.0 (http://www.cbs.dtu.dk/services/NetMHCpan/) [16], Artificial neural network (ANN) and Stabilized matrix method (SMM) of the "IEDB" server (http://tools.immuneepitope.org/analyze/html/mhc_binding.html) [17], and the MHCpred server (http://www.ddg-pharmfac.net/mhcpred/MHCPred/) [18] were used for this purpose.

2.4. MHC-II binding epitopes prediction

MHC-II binding epitopes of the mentioned proteins were predicted by NetMHCIIpan 3.2 (http://www.cbs.dtu.dk/services/NetMHCIIpan/) [19], NN-align and SMM-align methods in "IEDB" server (http://tools.iedb.org/mhcii/) [20,21], and the MHCpred (http://www.ddg-pharmfac.net/mhcpred/MHCPred/) [22].

2.5. CTL epitopes prediction

The CTL epitopes were predicted by NetCTLpan 1.1 (http://www.cbs.dtu.dk/services/NetCTLpan/) [23,24], NetCTL 1.2 (http://www.cbs.dtu.dk/services/NetCTL/) [25,26] using all the server-provided HLA supertypes (threshold for epitope identification >0.75). Also, the PAComplex (http://pacomplex.l ife.nctu.edu.tw./) [27] was used for evaluation of both peptide/MHC and peptide/T-cell receptors (TCR) interfaces. Epitopes with Joint *Z*-value >3 were selected in this server.

2.6. Predicting HLA CD4 + immunogenicity

CD4 immunogenicity of the selected proteins were predicted by IEDB Analysis Resource (http://tools.iedb.org/CD4episcore/). The CD4episcore predicts the immunogenicity of antigens for the CD4 + T cell at the population level without need of HLA typing. Its efficacy in different context including disparate techniques for epitope identification, different antigen sources and ethnicities has been validated [28].

2.7. Identification of linear B-cell epitopes of the S protein

The immunodominant linear B cell epitopes of the interacting region of S protein with ACE2 (350–570 residues of S protein), were predicted by several methods and servers including ABCpred Prediction Server (https://webs.iiitd.edu.in/raghava/abcpred/) [29], Bcepred Prediction Server (https://webs.iiitd.edu.in/raghava/bcepred/) [30], ElliPro (http://tools.iedb.org/ellipro/) [31], Antibody Epitope Prediction in IEDB Analysis Resource (http://tools.iedb.org/bcell/) by different methods including Bepipred Linear Epitope Prediction 2.0, Chou & Fasman Beta-Turn Prediction, Emini Surface Accessibility Prediction, Karplus & Schulz Flexibility Prediction, Kolaskar & Tongaonkar Antigenicity, and Parker Hydrophilicity Prediction.

2.8. Identification of the overlapping T cell epitopes

The epitopes with high affinity for multiple alleles of both HLA-I and HLA-II are called overlapping epitopes which are important for designing vaccines. These epitopes have the ability to activate both cytotoxic T cells and helper T cells. Therefore, the consensus method of the TepiTool (http://tools.iedb.org/tepitool/.) server

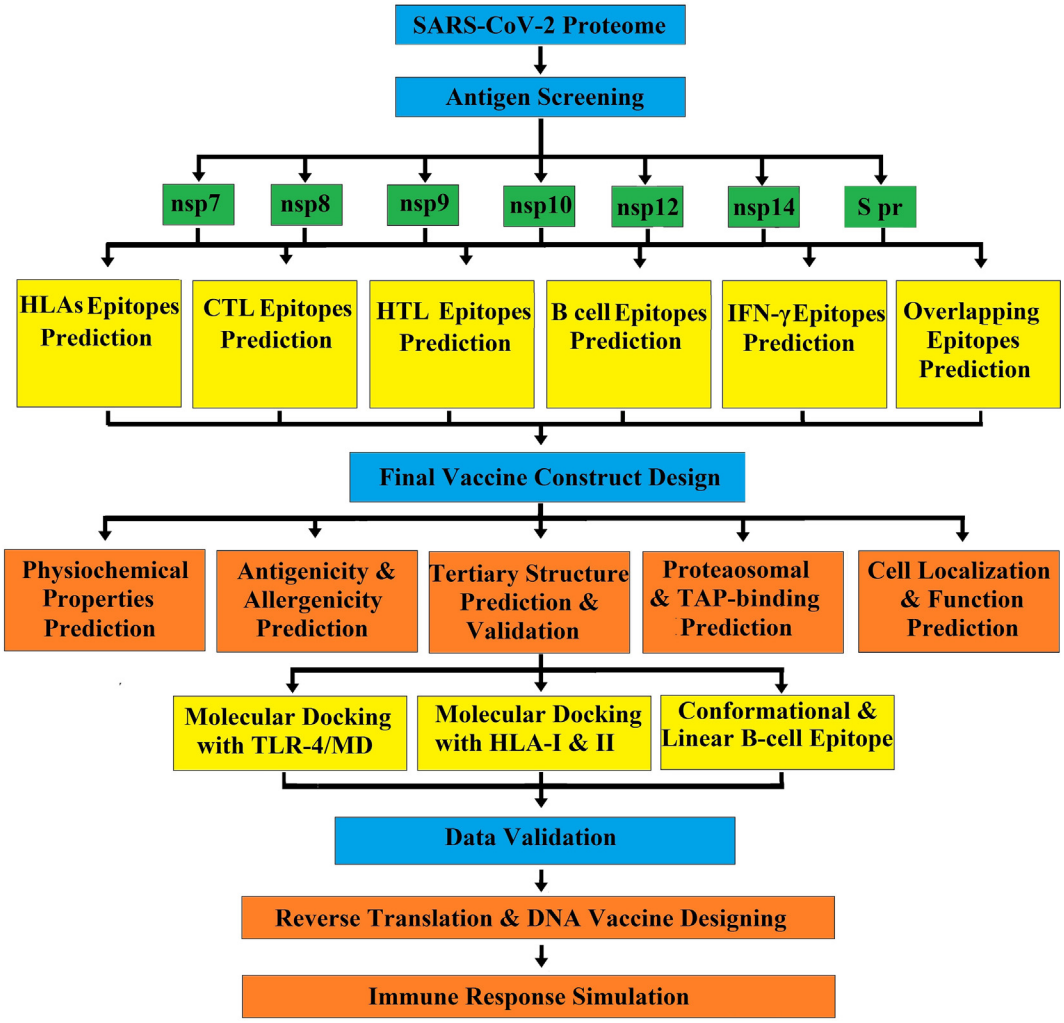

Fig. 1. The flow chart of this study. Designing this vaccine is consisted of six main steps including (1) Antigen screening; (2) Selection of the immunodominant regions of the selected antigens; (3) Designing the final vaccine construct; (4) Tertiary structure prediction and validation; (5) Molecular docking with TLR-4/MD; (6) Designing the DNA vaccine.

was employed for this purpose. In case of consensus prediction, the peptides which were predicted as epitope by all methods (ANN, SMM, NetMHCpan, CombLib for MHC-I and NetMHCIIpan, NN_align, SMM_align, Sturniolo and Combinatorial library for MHC-II) were considered as epitope to enhance the prediction specificity [15].

2.9. Prediction of IFN-γ inducing epitopes

IFN-γ has a pivotal role in both innate and adaptive immune responses against viral pathogens and predicting the epitopes with the capacity to induce IFN-γ is very important for designing vaccine. Therefore, the IFN-γ inducing epitopes of the selected proteins were predicted by the IFNepitope (http://crdd.osdd.net/raghava/ifnepitope/) server [32].

2.10. Assemblage of the multi-epitope vaccine candidate sequence

The candidate region of each protein which contained the immunodominant epitopes were selected to form the multi-epitope vaccine. The selected regions' sequences were connected to each other by furin-sensitive linker (RVRR). Human β-defensin was added at the N-terminus of the vaccine using (EAAAK)3 linker as an internal adjuvant to enhance immunogenicity. Also, CD4 + T-helper epitope, PADRE (Pan human leukocyte antigen-DR reactive epitope), was used at the C-terminal of the vaccine by GPGPG and A(EAAAK)2A linkers. Finally, a methionine was added to the N-terminal and His-tag (six histamines) was added to the C-terminal to form the final vaccine construct.

2.11. Vaccine features

2.11.1. Prediction of physicochemical parameters

The ProtParam online server (http://web.expasy.org/protparam/) was employed to investigate the physicochemical parameters of the vaccine including amino acid composition, grand average of hydropathicity (GRAVY), instability index, theoretical pI (isoelectric point), molecular weight, *in vitro* and *in vivo* half-life, and aliphatic index [33,34]. Also, the self-assembling protein nanoparticles (SAPN) from *Plasmodium falciparum* FMP014 (C1) [35], *Staphylococcus aureus* fusion protein (C2) [36], and a multiepitope vaccine consisted of immunodominant epitopes of SYCP1 and ACRBP antigens (C3) [37] were selected as positive controls for comparative evaluation of the candidate vaccine's properties.

2.11.2. Proteasomal cleavage and TAP-binding peptide prediction

The MAPPP (http://www.mpiib-berlin.mpg.de/MAPPP/cleavage.html) [38] and NetChop (http://tools.iedb.org/netchop/) were used

for predicting the proteasomal processing [39]. In addition, the binding affinity of the multi-epitope vaccine with the TAP (transporter associated with antigen processing) was investigated by the TAPPred (http://crdd.osdd.net/raghava/tappred/index.html) server [40].

2.11.3. Prediction of antigenicity, allergenicity, cell localization and protein function

The VaxiJen v2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) [41], Secret-AAR (http://microbiomics.ibt.unam.mx/tools/aar/) [42], and ANTIGENPro (http://scratch.proteomics.ics.uci.edu/) [43,44] servers were employed to predict the antigenicity of the candidate vaccine, C1, C2, and C3. The AllerCatPro (https://allercatpro.bii.a-star.edu.sg/) [45], AllergenFP v.1.0 (http://ddg-pharmfac.net/AllergenFP/) [46], and all the six approaches of the AlgPred (http://www. imtech.res.in/raghava/algpred/) [47] server were used to predict allergenicity of the candidate vaccine. Subcellular localization of the vaccine was assigned by employing the BaCelLo (http://gpcr.biocomp.unibo.it/bacello/) [48] and SherLoc2 (https://abi-services.informatik.uni-tuebingen.de/sherloc2/webloc.cgi) servers [49]. Presence of the PEST motifs at the vaccine construct was evaluated by the epestfind server (https://emboss.bioinformatics.nl/cgi-bin/emboss/epestfind) [50]. Functional class of the vaccine was predicted by the SVMProt server (http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi) based on Support Vector Machine classification. This server uses the primary sequence of a protein for predicting its functional family classification [51].

2.11.4. Homology modeling

The homology modeling of the vaccine construct was done by employing four different servers including Robetta (http://robetta.bakerlab.org/) [52], I-Tasser (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) [53], RaptorX (http://raptorx.uchicago.edu/) [54], Phyre2 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id = index) [55]. Subsequently, the quality of the obtained homology models were assessed by MolProbity (http://molprobity.biochem.duke.edu/) [56] and SAVES 5 servers (https://servicesn.mbi.ucla.edu/SAVES/). The most efficient structure which was obtained by the Robetta server, was used as the template for further optimization and predicting the refined 3D structure. Robetta is a product of Rosetta Commons, which can use both comparative protein modeling or *de novo* structure prediction and even a combination of both to construct the full structure [57].

2.11.5. Molecular dynamics

To have a valid and stable structure, the preliminary structure (Robetta structure) was subjected to a molecular dynamics' simulation using GROMACS 2018.3 version. Two 20 ns simulation were applied on the primary structure and the final structural frame of the first simulation was placed as the input of the second simulation. The Chemistry at Harvard Macromolecular Mechanics (CHARMM) force-field and simple water charge (SPC) water model, were used for the simulation of the protein in a periodic boundary condition (PBC). The first energy minimization was performed using a maximum number of 5000 nanosteps steepest descent minimization to yield a maximum force of below 1000 Kj/mol/nm. In the next step, NVT ensemble equilibrium was set, using a 50,000 nanosteps leap-frog integrator, which is equivalent to a 100 ps simulation, at 300 K temperature. Then, an NPT ensemble equilibrium was achieved by virtue of a similar integrator and the same nanosteps as in the NVT ensemble. The pressure for the NPT ensemble was 1.0 bar, using the Parrinello-Rahman pressure coupling method, and isotropic coupling style. The final molecular dynamics (MD) simulation was carried out with again leap-frog integrator, 10,000,000 steps (20 ns), and a Verlet cutoff-scheme. Particle Mesh Ewald method was utilized for generation of long-range electrostatics forces. The Fourier spacing grid was 0.16 for FFT (Fast-Fourier Transform). Isothermal compressibility of water was set at 4.5e-5 bar^{-1}. To constraint the geometry of all-bonds (even heavy atom-H bonds), the LINear Constraint Solver (LINCS) algorithm was employed. The backbone values were exploited for calculation of root-mean square deviation (RMSD), root-mean square fluctuation (RMSF), and radius of gyration.

2.11.6. Obtaining the final structure of the vaccine

The best structure of last MD optimization was chosen based on the sampling method each 5 ns and assessing their quality *via* the MolProbity and SAVES 5 severs. The best structure of the last procedure was selected and was subjected to further optimization. To reach a more accurate tertiary structure and optimizing the 3D structure of the vaccine protein, several softwares and servers were used. The swiss-pdb viewer (SPDBV) software [58,59] generated the best results among our tested ones according to the MolProbity and SAVES 5 servers. The optimization of the structure using SPDBV was accomplished in the user-friendly environment of the software, by GROMOS96 43B1 set of parameters, without a reaction field, *in vacuo*.

2.12. Prediction of linear and conformational B cell epitopes of the final vaccine construct

The mentioned serves at 2.7. section was utilized for linear B cell epitopes prediction of the final vaccine construct. In addition, conformational discontinuous B cell epitopes were predicted by the ElliPro (http://tools.iedb.org/ellipro/) [60] and DiscoTope (http://www.cbs.dtu.dk/services/DiscoTope/) [61] servers. The sensitivity and specificity of the DiscoTope serve at threshold −3.7 (default) were 0.47 and 0.75, respectively.

2.13. Post-translational modifications

The post-translational modifications (PTMs) of DNA vaccine product including N-glycosylation (http://www.cbs.dtu.dk/services/NetNGlyc/) [62], O-glycosylation (http://www.cbs.dtu.dk/services/NetOGlyc/) [63], ubiquitination (http:// www.ubpred.org/) [64], and phosphorylation (http://www.cbs.dtu.dk/services/NetPhos/) [65] were predicted.

2.14. Molecular docking

The ClusPro is a web server which is extensively utilized for evaluation of the protein–protein binding interactions [66–68]. This server assesses 70,000 rotational gyrations for the ligand structure (the vaccine) and the ligand is revolved along the ×, y, and z axes of cartesian coordinates, with respect to the receptor (TLR-4 structure) on a grid. Top favorable rotations with best scores (the first 1,000), were assigned for a greedy clustering with a 9 Å Cα radius and then, unfavorable energies were calculated (based on steric clashes) and unresolved structures were ruled out. The best docked structures were compared with previous studies to choose the most reliable model. The interaction energies between these two proteins in the ClusPro are computed using the summation of electrostatic, repulsive, attractive, hydrophobic, and also the pairwise-structure based potential which is established by the Decoys As The Reference State (DARS) method [69]. The vaccine interaction with TLR4 were analyzed and visualized in the Maestro v11.4 environment (by the means of protein interaction analysis module). Also, carbohydrate recognition and neck domains of surfactant protein A (PDB ID: 1R13) was used as control (C4) for docking with TLR-4 receptor [34,70].

According to immunoinformatic analyzes of the final vaccine construct, one of the most immunodominant overlapping epitope "IPLMYKGLPWNVVRI" was recognized which could bind to multiple alleles of HLA-I and II. Then, the epitope-HLA complex formation was validated by the molecular docking. The sequences of "LMYKGLPWNV" and "IPLMYKGLPWNVVRI" were generated by the peptide builder of the Maestro environment for the docking process with HLA class I and II, respectively. The structure of HLA class I and II were retrieved from the PDB (https://www.rcsb.org/) IDs 2GIT and 6ATF. The epitopes structures were then prepared and minimized afterwards using protein preparation tool. The peptide docking module of the Biologics suit of Schrodinger suite was used for the peptide docking process. The two alpha-helical bundles in each structure were defined as center of the grid generation. The grid size was adjusted according to the size of each epitope. Glide score was utilized for the prediction of the correct binding orientation and free energies [71–73]. For each independent running procedure, ten poses were returned for final assessment. The other details of the docking procedure are the same as Pourshojaei *et al.* study [74].

2.15. Population coverage

As COVID-19 pandemic is a global crisis, the population coverage of the designed vaccine was evaluated by Population Coverage Calculation tool of IEDB (http://tools.immuneepitope.org /tools/population/iedb_input) [23]. Calculations were set to be performed for HLA-I and HLA-II epitopes, separately.

2.16. Data validation

The immunodominant regions of the selected antigens were identified based on predictability of specific *in silico* tools. So, the experimentally validated epitopes of SARS-CoV-2 were used as control to compare our predictions' results with previous experimental studies. For this purpose, the IEDB resource which contains an extensive collection of experimentally identified immune epitopes was employed. This resource contains more than 260,000 epitopes and over one million T cell, B cell, MHC binding, and MHC ligand elution assays [75]. In addition, the final vaccine construct was evaluated by the SYFPEITHI server (http://www.syfpeithi.de/) which contains more than 7000 peptide sequences of human and other organisms, such as chicken, mouse and apes, which are known to bind MHC-I and MHC-II molecules. The final vaccine construct was reinvestigated by the same servers for data validation. Several MHCs, CTL and CD4 + T-helper epitopes were predicted within the final vaccine construct (data not shown).

2.17. The DNA vaccine construct

The peptide vaccine sequence was reverse translate to the template DNA sequence. This DNA sequence was codon optimized by using the JCat (http://www.jcat.de/) [76] server. Also, GC content and codon frequency distribution (CFD) were evaluated by employing GenScript (https://www.genscript.com/tools/rare-codon-analysis). The CFD can affect expression efficiency. To enhance expression, a Kozak sequence was added to the 5 of nucleotide sequence. At last, the DNA construct with suitable restriction enzyme sites (Nhe1, Nco1, and Xho1) at the 5′ and 3′ ends was designed.

2.18. Immune simulation

To simulate the immune system response against the final vaccine construct, the C-ImmSim (http://150.146.2.1/C-IMMSIM/index.php) [77] server was used. This server employs a position specific scoring matrix (PSSM) to estimate effects of an antigen on the immune system. This server can define responses of both humoral and cellular arms of the mammalians' immune system against a specific antigen. According to previous studies, the injection profile of prophylactic vaccines which means three sequential injections with 4-week intervals was set [78,79]. 4-week period is the minimum suggested time between immunization times for most of the currently in use vaccines [80]. All the default simulation parameters were used. The time steps of injection were specified at 1, 84 and 168. The simulation volume of vaccine (containing no LPS) injection was set at 1000. Also, for comparative assessment of the final vaccine construct, a SAPN made of two covalently linked coiled-coil domains which was designed to incorporate the membrane proximal external region (SAPN-MPER) of HIV-1 gp41 (C5) was selected as positive control [81,82].

3. Results

3.1. Antigen screening and selection of their most immunodominant regions

Cellular immunity plays a determinative role against viral infections and vaccines against these pathogens should activate CD8 + T cells [83]. HLA-I binding is a critical step for the process of antigen presentation to CD8 + T cells and activation of these cells [84,85]. According to the TepiTool server predictions, the Orf1ab exhibited considerably higher number of high-binding affinity MHC-I epitopes in comparison with all of the SARS-CoV-2 structural proteins (Spike glycoprotein, Envelope, Membrane glycoprotein, and Nucleocapsid) (Table 1). Therefore, the immunodominant regions of Orf1ab polyprotein were used to design the candidate vaccine. The 3888–3917, 4018–4045, 4222–4252, 4322–4373, 4939–4986, and 6074–6150 residues of Orf1ab polyprotein were selected as the most immunodominant regions according to immunoinformatics analyzes. Orf1ab polyprotein is cleaved into a definite number of non-structural proteins as its subunits (nsp1-nsp16) [14]. These above-mentioned regions of Orf1ab were located at 28–58 residue of nsp7, 76–103 residue of nsp8, 82–113 residue of nsp9, 69–120 residue of nsp10, 547–593 residue of nsp12, and 146–180 residue of nsp14, respectively. Also, an immunogenic region of S protein from its interacting domain with ACE2 receptor (the 400–510 residue of S protein) was selected to be incorporated in the candidate vaccine construct. These regions were not homologous to any human protein. They contained multiple epitopes with high affinity for binding to several alleles of both HLA-I and HLA-II (Tables S1-S4 and Tables S5-S8 for HLA-I and HLA-II, respectively). Also, these regions contained multiple CTL and CD4 + epitopes which are listed in Tables S9-S11 and Table S12, respectively.

Table 1
Comparison of the SARS-CoV-2 proteins according to number of HLA-I high-binding-affinity epitopes (across 9-mer peptides) with representative alleles from different HLA supertypes based on the TepiTool server predictions.

SARS-CoV-2 proteins	Reference sequences	Number of 9-mer overlapping peptides	Number of HLA-I binding epitopes
Orf1ab polyprotein	YP_009724389.1	6280	1303
Spike glycoprotein	YP_009724390.1	1221	208
Membrane glycoprotein	YP_009724393.1	190	72
Envelope protein	YP_009724392.1	62	23
Nucleocapsid phosphoprotein	YP_009724397.2	371	35

The used HLA-I alleles: A*01:01, A*02:01, A*03:01, A*24:02, A*26:01, B*07:02, B*08:01, B*27:05, B*39:01, B*40:0, B*58:01, B*15:01.

The overlapping epitopes have the capacity to activate both CTL and helper T cells. These epitopes were obtained from results of the TepiTool server (Table S13 and S14) to enhance the prediction specificity [15]. The overlapping epitopes of the selected regions were briefly illustrated in Table 2. In addition, the selected regions exhibited 96.2% and 97.1%, world population coverage for HLA-I and HLA-II, respectively (Table 3). Moreover, multiple IFN-γ inducing epitopes were identified at the selected regions (Table S15).

3.2. Final construct of the multi-epitope vaccine

Schematic diagram of the final vaccine construct with 462 amino acids, is illustrated in Fig. 2A. The final construct is consisted of three main domains including β-defensin (TLR-4 agonist), the B cell response inducing domain, and the T cell response inducing domain. In addition, a methionine was added at the N-terminal to form the AUG start code and histidine tag was added to the C-terminal.

3.3. The multi-epitope vaccine features

3.3.1. Proteolysis and TAP transport

The number of possible cleavage sites were about 140 according to the NetChop and MAPPP servers' predictions (Figure S1). Also, the TAPPred server was employed for prediction of the TAP binding affinity of the multi-epitope vaccine and more than 230 segments with >3 binding affinity to the TAP protein were identified.

3.3.2. Physicochemical properties of the final vaccine construct

The ProtParam server was used to calculate physicochemical properties of the final vaccine construct. Its molecular weight was calculated 51643.19 g/mol and the pI was 10. Stability is critical to the overall success of vaccine [86]. The instability index provides an estimate of a protein stability in a test tube. The proteins with instability index below 40 are stable. The candidate vaccine was a stable protein complex according to its instability index (II) 27.09. A significant number of individuals dies due to vaccine-preventable diseases each year. A significant portion of this problem results from the thermal instability of many of the currently used vaccines. Therefore, thermal stability is very important to be estimated for the designed vaccine [87]. The final vaccine construct was a thermostable protein based on its aliphatic index 79. The estimated half-life of the final construct which shows the time taken by the protein to reach half of its concentration after its synthesis, was determined to be 30 h, >20 h and >10 h in mammalian reticulocytes (*in vitro*), yeast (*in vivo*) and *Escherichia coli* (*in vivo*), respectively. The GRAVY score was − 0.354, which shows that the protein is hydrophilic in nature and can appropriately interact with its surrounding water molecules. Solubility is a very important physiochemical property for expression and subsequently, manufacturability of a protein. Appropriate hydrophilicity of proteins can significantly decrease their expression efficacy and manufacturability [88,89]. The final vaccine construct was compared with the positive controls regarding physicochemical properties (Table 4). It exhibited approximately similar properties in comparison with positive controls. The final vaccine construct exhibited appropriate physicochemical properties like high solubility and stability which can increase bioavailability, immunogenicity and decrease the probable side effects [90,91].

3.3.3. Antigenicity, allergenicity, cell localization and protein function properties

The antigenicity scores of the candidate vaccine was predicted 0.5936, 0.7425, and 39.8 by the Vaxijen 2.0, ANTIGENpro, and Secret-AAR servers, respectively (Table 4). Comparative investigation of the antigenicity scores of the candidate vaccine and positive controls (C1, C2, C3) is illustrated in Table 4. The final construct was nonallergen based on the predictions of the AllerCatPro, AllergenFP v.1.0, and all approaches of the AlgPred server. Also, subcellular localization of the DNA vaccine product was predicted to be cytoplasm. There was no significant PEST site within the vaccine sequence. Also, the SMVPRot server predicted no functionality for the final vaccine construct.

3.3.4. The 3D structure refinement and validation

The 3D structure is necessary for determining the conformational B cell epitopes and molecular docking. Therefore, refined 3D structure of the final vaccine construct was obtained after multiple steps (Fig. 2B). Molecular dynamics root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) plots are helpful in terms of the relative stability of the protein with regard to its prior state in a series of simulation, and their values highly depend on the amino acid composition of the protein structure and can be quite unique in a certain protein structure. As Fig. 3 illustrates, RMSD (Fig. 3A) and RMSF (Fig. 3B) qualities were considerably satisfactory in terms of low alteration of protein residues over the time of simulation. The highest instability of the structure was observed in the end part of the protein structure (as RMSF plot reveals) which isn't a concern. In general, the RMSD total variation is less than 3 Å, signifying a stable protein construct over the time of simulation. In addition, the Ramachandran plot was performed for the refined 3D structure by the PROCHECK server and 94.4% of the residues were in the favorite regions (Fig. 4A). All related data of the refined model is presented in Table S16. The refined 3D structure was evaluated by ProSA-web and ERRAT to clarify the quality and potential errors. Its Z-score was – 6.57 (Fig. 4B). Also, overall quality factor for the refined 3D structure was 90.37% (Fig. 4C). Its Verify 3D score was 84.20% which was confirmed by 3D-1D value (Fig. 4D).

3.4. Determining the B cell epitopes

The linear B cell epitopes of the S protein's interacting region with ACE2 receptor are listed in Tables S17-S19. Also, the linear and conformation B cell epitopes of the final vaccine construct are listed at Tables S20-S23 and Tables S24-S25, respectively. The final vaccine construct contained several conformational and linear B cell epitopes (Fig. 5A-F). Although presence of multiple linear and conformation B cell epitopes enhances the uptake and presentation of the peptide vaccine by B cells, the most important B cell epitopes for producing neutralizing antibodies are the common ones between the final vaccine construct and the interacting region of S protein (Table 5 and Fig. 5F-H). Neutralizing antibodies not only bind to a virus, but also bind in a manner that might block its interactions with the target receptor [92]. It should be mention some of these common linear B cell epitopes were located at the surficial regions of the final vaccine construct and S protein interacting region which increase their accessibility for the immune response [93]. The neutralizing antibodies are necessary for protection against viral infections. However, it usually takes a long period of time for the immune system to produce highly effective neutralizing antibodies [94]. The designed multi-epitope vaccine can facilitate this process by presenting appropriate B epitopes to the host immune system.

3.5. Protein-protein docking

Molecular docking and dynamics simulations showed formation of stable and significant interactions between the vaccine with the monomer (Fig. 6A) and heterodimer TLR-4/MD (Fig. 6B). Several key residues throughout the entire structure of the final vaccine construct participated in the stability of the TLR-4/MD-

Table 2
The obtained overlapping epitopes from the TepiTool server results.

#	Peptide	IC50	Alleles
1	AMQTMLFTM	44.80	A0201, A0206, B1525, A0301, A1101, A3101, A6801, A0203, A3202
	AMQTMLFTMLRKLDN	24.04	DRB11101, DPA10201DPB10501, DPA10201DPB14001, DRB10701, DRB10801, DPA10301/DPB10402,
2	VLGSLAATV	50.70	A0201, A0202, A0206, A0203, A0301, A1101, A0204, A0207, A6802, B5701
	RGM**VLGSLAATV**RLQ	14.88	DQA10501/DQB10301, DRB10101, DRB10701, DQA10102DQB10602, DQA10103DQB10302, DQA10106DQB10301, DRB10901, DRB11302, DRB50101
3	MLSDTLKNL	54.10	A0201, A0202, A0203, A0204, A0205, A0206, A0207, A1101
	RIKIVQ**MLSDTLKNL**	16.38	DRB10101, DRB10401, DRB10404, DRB40101, DRB40103, DRB40104, DRB40106, DRB40107
4	MLFTMLRKL	69.50	A0201, A0202, A0203, A0206, , A0301, A1101, A6801
	QT**MLFTMLRKL**DNDA	18.73	DRB11101, DRB10801, DPA10201DPB10501
5	LMYKGLPWNV	7.20	A0201, A0203, A0204, A0206
	IP**LMYKGLPWNV**VRI	13.42	DRB10101, DRB10901, DRB10103, DRB10109, DRB10110, DRB10114, DRB10115, DRB50101
6	MVLGSLAATV	8.30	A0201, A0206, A6802
	RG**MVLGSLAATV**RLQ	14.88	DRB10101, DRB10701, DRB10901, DRB11302, DRB50101, DQA10102/DQB10602, DQA10103/DQB10302, DQA10106/DQB10301, DQA10501/DQB10301
7	AMQTMLFTML	78.80	A0201
	AMQTMLFTMLRKLDN	24.04	DRB11101, DRB10701, DRB10801, DRB50101, DPA10301/DPB10402, DPA10201/DPB10501, DPA10201/DPB14001
8	HLIPLMYKGL	85.20	A0201
	HLIPLMYKGLPWNVV	25.95	DRB10101, DRB10901, DRB11501
9	YLYFIKGLNNL	63.50	A0201, A0202, A0203
	YLYFIKGLNNLNRGM	9.69	DRB10101, DRB10401, DRB50101, DRB10104, DRB10105, DRB10107, DRB10108, DRB10109, DRB10110, DRB10111, DRB10112, DRB10113, DRB10114, DRB10115, DRB30201, DRB30204, DRB30202, DRB50101, DRB10405DRB11101, DRB11501,
10	LMYKGLPWNVV	83.90	A0201
	IP**LMYKGLPWNVV**RI	13.42	DRB10101, DRB10901, DRB10103, DRB10110, DRB10114, DRB10115, DRB50101, DRB11501, DRB10901
11	HLIPLMYKGLPWNV	59.20	A0201
	HLIPLMYKGLPWNVV	25.95	DRB10101, DRB10901
12	MQTMLFTML	29.90	A0206, B3901, A0203, A1101
	MQTMLFTMLRKLDND	21.82	DRB11101, DPA10201DPB10501, DPA10201DPB14001, DRB10801, DRB50101, DPA10301/DPB10402
13	MVLGSLAAT	37.70	A0206, A1101, A6802
	RG**MVLGSLAAT**VRLQ	14.88	DRB10101, DRB10901, DRB11302, DRB50101, DQA10501/DQB10301, DRB10701, DQA10102/DQB10602, DQA10103/DQB10302, DQA10106/DQB10301
14	TMLFTMLRK	9.50	A0203, A0301, A1101
	Q**TMLFTMLRK**LDNDA	18.73	DRB11101, DRB10801, DPA10201/DPB10501
15	RQFHQKLLK	17.00	A0301, A0302, A0304, A0305, A0306, A0307, A1101, A1102, A1103, A1104, A1105, A1106, A1107, A3008, A3102, A3103, A3104, A3105, A3106, , A3001, A3101, A6801,
	RQFHQKLLKSIAATR	16.68	DRB10101, DRB11101, DRB10102, DRB50104, DRB50101
	KVKYLYFIK	7.30	A0206, A0301, A1101, A1104, A3001, A3008, A3101
16	P**KVKYLYFIK**GLNNL	16.00	DRB10101, DRB11101, DRB11501, DRB40101, DPA10103/DPB10201, DPA10201/DPB10501, DPA10201/DPB14001
17	MLFTMLRK	64.00	A0301
	QT**MLFTMLRK**LDNDA	18.73	DRB11101, DRB10801, DPA10201/DPB10501
18	HLIPLMYK	94.20	A0301, A0302, A0304, A0305, A0306, A0307, A1103, A1104
	HLIPLMYKGLPWNVV	25.95	DRB10101
19	QTMLFTMLR	3.30	A0302, A0307, A1101, A1102, A1103, A1104, A1105, A1106, A1107, A3101, , A3102, A3103, A3104, A3105, A3301, A3303, A3106, A6801
	QTMLFTMLRKLDNDA	18.73	DRB11101, DPA10201DPB10501
20	ATVVIGTSK	26.60	A0301, A0302, A0304, A0305, A0306, A0307, A1101, A1102, A1103, A1104, A1105, A1106, A1107, A3001, A3008,
	IAATRG**ATVVIGTSK**	38.37	DQA10501/DQB10301, DQA10102/DQB10602, DQA10106DQB10301

Table 2 (*continued*)

#	Peptide	IC50	Alleles
21	MQTMLFTMLR	3.80	A0301, A1101, A3301, A3303, A6801, A3101
	MQTMLFTMLRKLDND	21.82	DRB11101, DPA10201DPB10501, DPA10201DPB14001, DRB50101, DPA10301/DPB10402, DRB10801
22	AMQTMLFTMLR	39.60	A1101, A3101, A6801
	AMQTMLFTMLRKLDN	24.04	DRB10701, DRB10801, DRB11101, DRB50101, DPA10201/DPB10501, DPA10201/DPB14001, DPA10301/DPB10402
23	IVQMLSDTLK	45.70	A1101, A6801
	RIK**IVQMLSDTLK**NL	16.38	DRB10101, DRB40101, DRB40103, DRB40104, DRB40106, DRB40107
24	ATRGATVVI	22.10	A0203, A1101, A6802, A3001, A3008, A3201, A3202, A3206, B0704, B1504
	LLKSIA**ATRGATVVI**	10.97	DRB10101, DQA10501/DQB10301, DRB10701, DRB10901, DRB11302, DRB50101, DRB10103, DRB10104, DRB10105, DRB10107, DRB10108, DRB10110, DRB10111, DRB10112, DRB10113, DRB10115, DRB50104, DRB10901
25	RIKIVQMLS	22.50	A0206, A0301, A1101, A3001, A3008
	RIKIVQMLSDTLKNL	16.38	DRB10101, DRB40101, DRB40104, DRB40106, DRB40107, DRB40101
26	KSIAATRGA	24.60	A3001, A1101, A3008
	QKLL**KSIAATRGA**TV	9.09	DRB10102, DRB10103, DRB10104, DRB10105, DRB10106, DRB10107, DRB10108, DRB10109, DRB10110, DRB10112, DRB10404, DRB10901, DRB11101, DRB11302, DRB30201, DRB30204, DRB50101, DRB50102, DRB50103, DRB50104, DRB50105, DRB11501, DRB50101, DQA10501/DQB10301
27	MYKGLPWNVVR	37.50	A3301, A3101
	IPL**MYKGLPWNVVR**I	13.42	DRB10101, DRB10103, DRB10109, DRB10110, DRB10114, DRB10115, DRB10901, DRB11501, DRB50101
28	LLKSIAATR	46.90	A3101, A6801, A0203, A0206, A3102, A3103, A3104, A3105, A3106
	QK**LLKSIAATR**GATV	9.09	DRB10101, DRB10102, DRB10103, DRB10104, DRB10105, DRB10106, DRB10107, DRB10108, DRB10109, DRB10110, DRB10111, DRB10112, DRB10113, DRB11101, DRB11302, DRB11501, DRB30201, DRB30204DRB50101, DRB50102, DRB50103, DRB50104, DRB50105, DQA10501/DQB10301
29	EAFEKMVSL	33.50	A0202, A0203, A1101, A2501, A2502, A2503, A2504, A2505, A2506, A2507, A2601, A2602, A2603, A2604, A2605, A2606, A2608, A2609, A6801, A6802, B0710, B0801, B0802, B0803, B0804, B0805, B0807, B1401, B1402, B1403, B1404, B1405, B1406, B3501, B3502, B3503, B3504, B3505, B3506, B3507, B3509, B3903, B3904, B3909, B3910, B4007, B4008, B4201, B4202, B4204, B4205, C0102, C0103, C0303, C0304, C1203
	DTT**EAFEKMVSL**LSV	19.68	DRB10101, DRB10401, DRB10405, DRB11101
30	FEKMVSLLSV	36.00	B4002
	DTTEA**FEKMVSLLSV**	19.68	DRB10101, DRB10401, DRB10405, DRB11101
31	TEAFEKMVSL	57.60	B0804, B1802, B1803, B3701, B3704, B3705, B3706, B3803, B3902, B3908, B4001, B4002, B4003, B4004, B4005, B4006, B4007, B4008, B4009, B4010, B4101, B4102, B4103, B4104
	DT**TEAFEKMVSL**LSV	19.68	DRB10101, DRB10401, DRB10405, DRB11101
32	IPLMYKGLPW	35.00	B0702, B5301
	IPLMYKGLPWNVVRI	13.42	DRB10101, DRB10103, DRB10109, DRB10110, DRB10114, DRB10115, DRB11501, DRB50101
33	IAATRGATV	18.80	A0203, A0301, A6801, B0801, B0802, B0803, B0805, B0807, B4202, C0102, C0103, C0303, C0304, C1601, C1203
	QKLLKS**IAATRGATV**	9.09	DRB10102, DRB10103, DRB10104, DRB10105, DRB10106, DRB10107, DRB10108, DRB10109, DRB10110, DRB10111, DRB10112, DRB10113, DRB11101, DRB11302, DRB11501, DRB30201, DRB30204, DRB50101, DRB50102, DRB50103, DRB50104, DRB50105, DQA10501/DQB10301, DRB10101, DRB10701, DRB10901

vaccine interactions (Fig. 6C). Among them, Lys5, Lys14, Arg18, Arg20, Lys110, Asp190, Glu202, Arg214, and Arg235 contributed the most part of these interactions, each creating at least two hydrogen bonds with the TLR-4 residues. The positive nature of the vaccine protein and its remarkable interactions with TLR-4 exhibit that the TLR-4 needs a relatively structure with high negative charge for its proper binding. Among the most important residues of the first 60 residues (the β-defensin domain), Lys14 created hydrogen bonds with Asp500 and Asn524, displaying a SASA of 97.5%, being one of the most crucial amino acids in the interactions with TLR-4. Arg18 formed two hydrogen bonds with His424 and Asp426, with an additional salt bridge with Asp426. Arg20, another critical residue, provided two hydrogen bonds with Glu423. All the interactions between the resides of the docked vaccine and TLR-4/MD complex are listed in Table S26. Also, the C4 against TLR-4/MD docking studies were done as positive control (Fig. 6D). The main interacting residues of TLR4/MD in both TLR4/MD-C4 and TLR-4/MD-candidate vaccine complexes were A:151:Lys, A:32:Pro, A:227:His, A:228:Lys, A:254:His, A:255:Arg, A:284:Glu, A:285:Glu, A:287:Arg, A:332:His, A:334:Glu, A:353:Arg, A:374:Glu, A:375:Phe, A:400:Lys, A:401:Tyr, A:423:Glu, A:424:His, A:449:Tyr, A:472:Glu, A:473:Val, A:475:Lys, A:497:Thr, A:498:Phe, A:521:Gln, A:522:Val, A:545:Gln, A:546:Val, A:570:Ala, A:571:Phe, A:597:Gln, A:598:Leu, C:57:Gln, C:59:Tyr, C:75:Lys, C:82:Ser, C:127:Glu. The docking energy of TLR4/MD-C4 complex (−727.7 Kcal/mol) was higher than the TLR-4/MD-candidate vaccine complex (-1315.6 Kcal/mol) which indicates forming stronger immunological complexes by the candidate vaccine in comparison with the control.

The dockings result of one of the most immunodominant overlapping epitopes of the vaccine "LMYKGLPWNV" (digested by the proteasomal complex) with MHC-I revealed considerable binding

Table 3
Percentage of population coverage.

Area	HLA-I coverage (%)	HLA-II coverage (%)
East Asia	96.5	96.4
Europe	97.1	97.9
North Africa	91.8	91.8
North America	97.4	98.1
World	96.2	97.1
Average	95.8	96.26
Standard deviation	2.28582589	2.583215051

energies between the peptide and the HLA-I (Fig. 7A) and HLA-II (Fig. 7B). The docking glide gscore was −8.135 for 2GIT, and −8.894 for 6ATF structure. This amount of low negative values vindicates the potential ability of these digested fragments to maintain at the surface of the MHC molecules and act as immunogens. As Fig. 7 illustrates, each HLA protein molecule contain two binding sites for holding antigenic peptides and one of them is filled with the fragments. The successful maintenance of the fragments between the two α-helices indicate that these peptides are capable of inducing remarkable immune response.

Asn9 and Val10 each displayed two hydrogen bonds and were among the most important residues in the MHC-I docking interactions. Asn9 of the digested fragment created these bonds with Thr73 and Trp147. In addition, Val10 of the peptide created these interactions with Asp77 and Lys146. By having a buried SASA (solvent-accessible surface area, an indicator of good fitness of a ligand and receptor in a water-soaked environment) ratio of 92.6% for Val10 and 84.3% for Asn9, these interactions seemed significantly critical for the antigen presentation by the MHC-I. Lys4 and Met2 also represented hydrogen bonds with Arg65 and Gln155. For MHC-II structure, Arg14 contributed the most in the stability of the peptide-MHC complex. With a buried SASA of 96.7% and three hydrogen bonds with Glu11, Glu28, and Arg71 this residue appeared the most crucial in staying put in the MHC binding pocket. Lys6 created two hydrogen bonds with Tyr60. This lysine residue of the peptide also contributed a hydrogen bond and a salt bridge with Asp66.

3.6. Post translation modifications

PTMs can significantly affect the DNA vaccine product stability and MHC processing [95]. All the predicted ubiquitination, glycosylation, and phosphorylation sites of the vaccine sequence are listed in Table S27.

3.7. Data validation

According to IEBD database, each of the selected immunodominant regions contained multiple experimentally validated SARS-CoV-2 epitopes (Table 6). These experimentally validated epitopes which are present at the final vaccine construct are listed at Table 6. Therefore, the results of previous experimental studies can support our predictions. Also, the SYFPEITHI server predicted several epitopes for HLA-I and HLA-II at the final vaccine construct (Table S28). Merging of the selected immunodominant regions and other modifications didn't interrupt epitope mapping and antigenicity of the final vaccine construct according to reassessments by the same servers (data not shown).

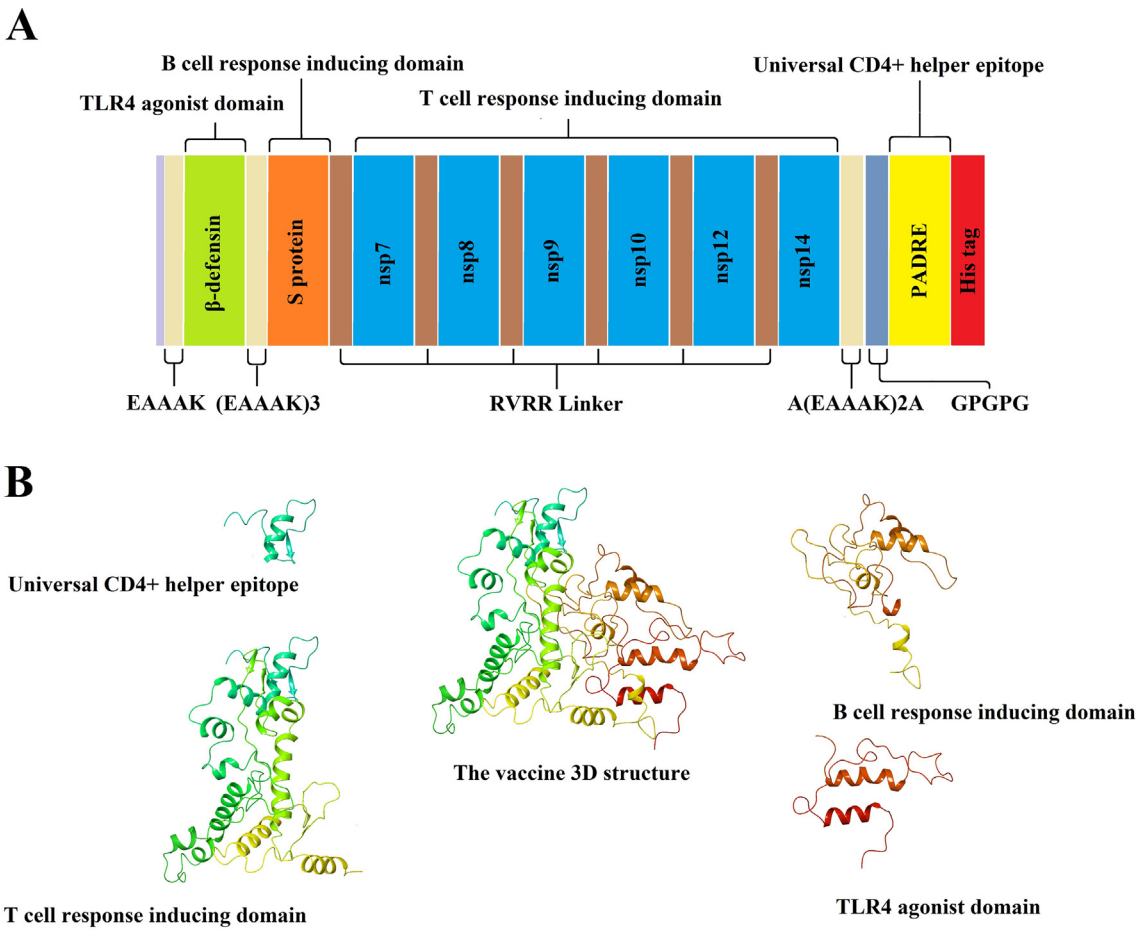

Fig. 2. (A) Schematic diagram of the final vaccine construct. (B) 3D model of the final vaccine construct and its main consisting domains.

Table 4
Comparative investigation of the physicochemical and antigenicity properties of the positive vaccine controls (C1, C2, C3) and SARS-CoV-2 candidate vaccine.

Properties	Parameters/Tools	Value/Score/Probability			
		C1	C2	C3	Candidate vaccine
Physicochemical	Molecular weight	2.44 kDa	5.05 kDa	6.38 kDa	51.64 kDa
	Isoelectric point (pI)	6.24	8.67	9.61	10
	Instability index (II)	28	22.78	33.38	27.09
	GRAVY	−0.88	−0.32	−1.127	−0.354
	Aliphatic index	65.94	–	53.53	79
Antigenicity	ANTIGENpro	0.67	0.94	0.62	0.74
	Secret-AAR	42.6	27.59	35.4	39.8
	VaxiJen	0.65	0.67	0.61	0.59

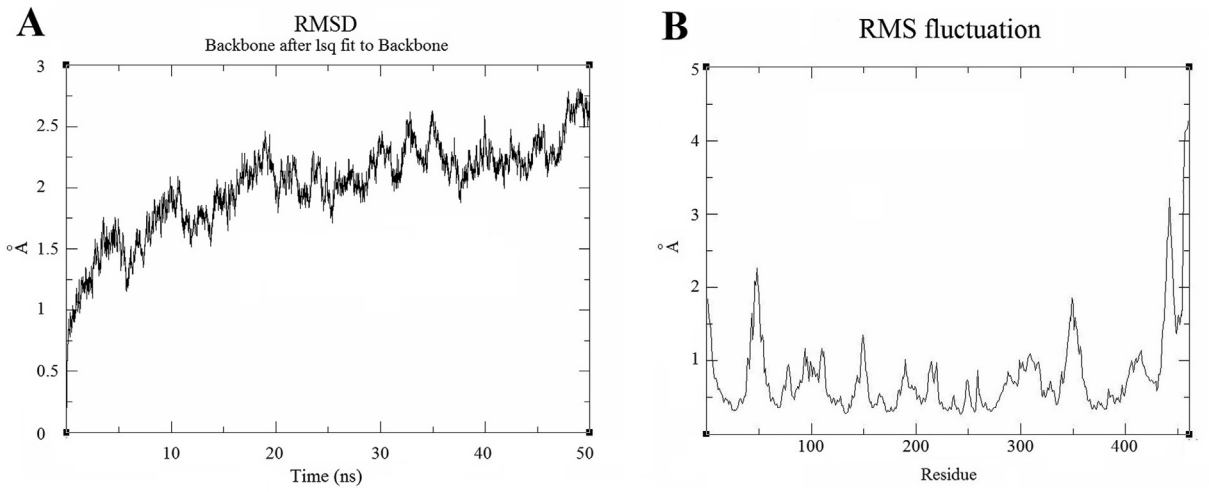

Fig. 3. (A) The root-mean-square deviation (RMSD) and (B) root-mean-square fluctuation (RMSF) of atomic positions in the "last" molecular dynamics simulation of the vaccine construct, depicting stable conformations during the simulation.

3.8. Designing the multi-epitope DNA vaccine

Codon adaptation is necessary to optimize the foreign genes' expression in a special host [96]. Therefore, the final vaccine construct was reverse translated and codon optimized for human host. The optimized codon adaptation index (CAI) for mammalian and overall GC content were 0.99 and 67.90%, respectively (Fig. 8A and B). In addition, suitable restriction enzymes sites and the Kozak sequence was incorporated in the final DNA construct (Fig. 8C). These restriction sites ease the cloning into pcDNA3.1 to form a multi-epitope DNA vaccine or pET-28a for expression in a prokaryote host. In our previous experimental studies, we used the same motifs (Kozak sequence, start codon, stop codon, and the same restriction enzyme sites) in the designed final DNA construct. It was successfully cloned into the pcDNA3.1 and pET-28a. Also, high expression of the recombinant protein was detected in the both transfected prokaryotic and eukaryotic hosts [37,97]. In this study, the designed vaccine construct is about 51 kDa. The pcDNA3.1 has been exhibited high efficacy for *in vivo* expression of different recombinant proteins with different sizes from 14 to 73 kDa [98–100]. Also, high expression of recombinant proteins with sizes ranged from 14 to 51 kDa have been reported for pET-28a vector [101,102]. Therefore, these vectors are appropriate for expression of the final vaccine construct.

3.9. Immune simulation

Most of the currently licensed vaccines against human infectious diseases have been developed for generating neutralizing antibodies [81,103]. But antibodies are insufficient for fighting intracellular infections like SARS-CoV-2. Therefore, both humoral and cellular immunity responses plus activating the innate immunity are necessary for an effective immune response against this virus [104]. Therefore, the immune response for the final vaccine construct was simulated by the C-ImmSim server (Fig. 9A). Furthermore, the C5 was used as the positive control for comparative evaluation of the candidate vaccine's immune simulation (Fig. 9B). High similarity was observed between the simulation results of the candidate vaccine and C5. Significant decrease in antigen count during time was observed for both the candidate vaccine and control. The antibodies titers were considerably raised after second and third dose of injection for the candidate vaccine (Fig. 9Ai) and control (Fig. 9Bi). At the primary response, a rise in IgM antibody (Fig. 9Ai) and the B cell isotype IgM population (Fig. 9Aii) was observed after the candidate vaccine injection. Also, a significant increase in the B cell isotype IgG1 and IgG2 populations with superiority of IgG1 isotype (Fig. 9Aii) was observed after the second time of candidate vaccine injection (Fig. 9Aii). At the tertiary response, IgG1 + IgG2 antibodies titer was higher than IgM (Fig. 9Ai). This exhibits Ig heavy chain class switching which is necessary for an effective vaccination [105]. The total B cell population increased after each injection which was characterized by increase of immunoglobulins' concentration including IgG + IgM, IgG1 + IgG2, IgG1 and IgM antibodies. This increase in the antibodies titer resulted in a decrease in the the antigen concentration (Fig. 9Ai). Total B cell population was found approximately the same for the candidate (Fig. 9 Aii) and control vaccines (Fig. 9Bii). According to the cytokines' simulation plot, high levels of IFN-γ and IL-2 were observed which can exhibit the candidate vaccine (Fig. 9Aiii) ability to generate an appropriate immune

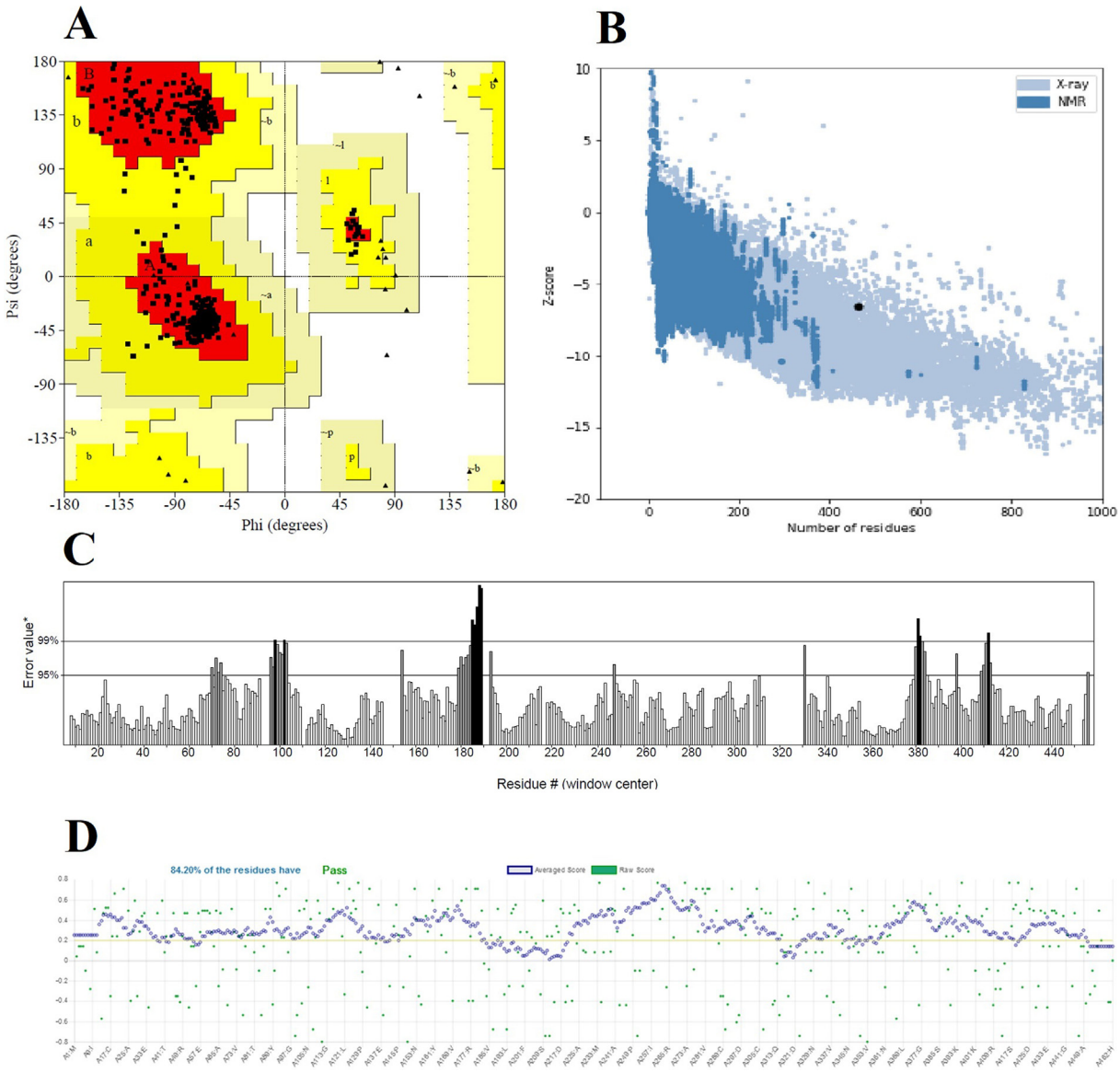

Fig. 4. (A) Ramachandran plot of the refined 3D structure. The Ramachandran plot revealed that 94.4%, 5.6%, and 0% and 0% of the refined model residues are located in the most-favored (red), favored (yellow), generously allowed (pale yellow), and disallowed regions (white), respectively. The vaccine residues are shown as black squares. (B) ProSA-web server analysis of the refined model. The Z-score of the refined models was − 6.57. The plot contains Z-scores of all protein chains in the protein data bank (PDB) determined by X-ray crystallography (light blue) and NMR spectroscopy (dark blue). The Z-score of the refined model is denoted by a large black dot. (C) The ERRAT plot of the refined model. The regions of the refined model that could be rejected at the levels of 95% and 99% are indicated in gray and black lines, respectively. In general, an appropriate high-resolution structure produces a value of ≥ 95%. (D) Verify 3D analyses plot. 84.20% of the residues of the refined model have average 3D-1D score ≥ 0.2 which is pass. At least 80% of the amino acids should score ≥ 0.2 in the 3D-1D profile to get passed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

response [106]. Also, the observed IL-2 concentration level for the candidate vaccine was found remarkably higher than control vaccine. However, the concentration level of IFN-γ was slightly lower for the candidate vaccine (Fig. 9Aiii) in comparison with the control (Fig. 9Biii). A steady increase in the Th (helper) cell population along with developing memory cells was indicated for both candidate vaccine (Fig. 9Aiv) and control (Fig. 9Biv). In addition, developing memory cell at the Tc (cytotoxic) cell population describes that the Tc cell population was highly responsive for both candidate (Fig. 9Av) and C5 (Fig. 9Bv). However, the active Tc cell population per state (cells per mm3) was considerably higher for the candidate vaccine (Fig. 9Avi) in comparison with the control (Fig. 9Bvi). Overall, these simulations exhibit that the designed final vaccine construct can activate the the immune response to provoke an appropriate anti-SARS-CoV-2 immunity.

4. Discussion

Vaccines are the most effective and least expensive way to prevent and control devastating viral outbreaks. An effective vaccine should induce potent humoral and cellular immune responses against the target virus and infected cells. The vaccine efficacy deeply depends on the antigens selection [107]. Most of the previously published *in silico* studies used the S protein of SARS-CoV-2 virus to design their candidate vaccines [108,109]. This glycoprotein rapidly became the main target of vaccine design. Several companies and research institutes have started developing different platforms like recombinant protein vaccine, DNA vaccine and RNA vaccines for SARS-CoV-2 a vaccine that has the receptor binding domain of SARS-CoV-2 protein S as its target.

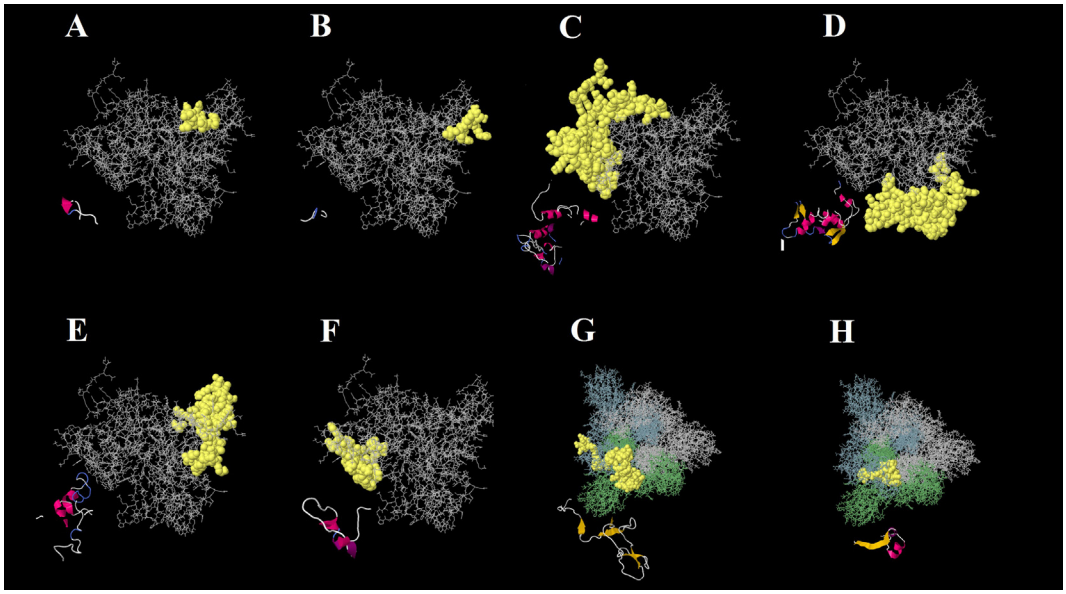

Fig. 5. The predicted linear and conformational B cell epitopes of the final vaccine construct and functional region of S protein by the ElliPro server in the ball and stick model. Yellow balls represent the predicted epitopes' residues. Non-epitope and core residues are shown in white sticks. Also, ribbon representation of the epitopes' residues was separately displayed. (A-E) The conformational B cell epitope of the vaccine construct. (F) The linear B cell epitope of the vaccine construct. (G-H) The linear epitopes of the functional region of S protein. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
The common linear B cell epitopes between the final vaccine construct and the S protein of SARS-CoV-2.

Peptide	Start position	Score
YQAGSTPCNGVE	473	0.72
PFERDISTEIYQ	463	0.61
GKIADYNYKLPDDF	416	0.90
QAGSTPCNGVEGFN	474	0.87
YRLFRKSNLKPFER	453	0.77
GSTPCNGVEGFNCYFP	476	0.91
TEIYQAGSTPCNGVEG	470	0.89
FERDISTEIYQAGSTP	464	0.86
TGKIADYNYKLPDDFT	415	0.84
EGFNCYFPLQSYGFQP	484	0.73
TGCVIAWNSNNLDSKV	430	0.71
YKLPDDFTGCVIAWNS	423	0.71
GCVIAWNSNNLDSKVGGNYN	431	0.81

In this study, the most immunogenic regions of the Orf1ab polyprotein were incorporated into the vaccine construct for T cell response. On the other hand, the most immunodominant fragment of the S protein functional region (400–510 residues) was incorporated in the vaccine construct for inducing neutralizing antibodies production by B cells. The nsps are obtained from cleavage of Orf1ab. The nsps and particularly the replication-transcription complexes (RTCs) related nsps are more conserved in comparison with S and other structural proteins among SARS-CoVs. [110]. Besides, more than 60% of the coronavirus genomes belongs to the nsps. They are highly expressed at the infected cells and presented by their MHC-I from the first days of infection [13]. The nsps have significantly lower glycosylation density in comparison with the structural proteins. Therefore, they may be better choices for designing the vaccine as many studies have reported the negative effects of dense glycosylation of epitopes for their recognition by T cells [110–112]. The selected immunodominant regions of Orf1ab were located at the nsp7, nsp8, nsp9, nsp10, nsp12, and nsp14 which all are RTCs-related nsps [113,114]. According to our *in silico* predictions, Orf1ab polyprotein has the highest number of HLA-I binding epitopes in comparison with all of the structural proteins of SARS-CoV-2 (Table 1). These predictions were consistent with previous bioinformatic studies which reported that Orf1ab contains the highest number of epitopes for binding to different alleles of HLA-I [115]. Besides, Gangaev *et al.* [116] observed remarkable CD8 + T cell responses towards the SARS-CoV-2 Orf1ab in COVID-19 patients.

Most of the prophylactic vaccines against viral diseases activate the B cell immune response for producing neutralizing antibodies. On the other side, the therapeutic vaccines are designed to generate cell-mediated immunity [117]. In this study, the designed vaccine was predicted to have high ability to induce neutralizing antibodies production against binding domain of the SARS-CoV-2 S protein which makes the vaccine appropriate to be used as a prophylactic vaccine. In addition, containing multiple T cells' epitopes

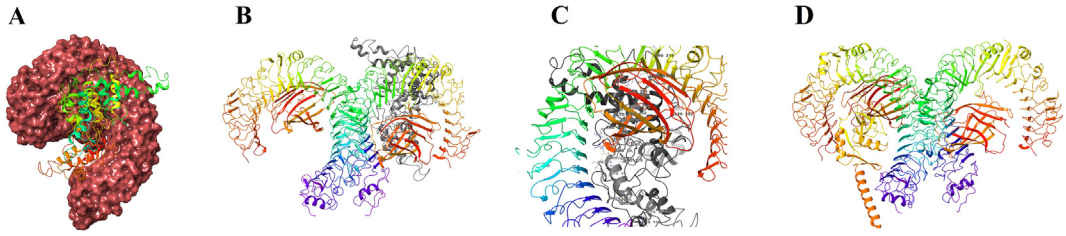

Fig. 6. Molecular docking of the subunit vaccine with TLR-4. (A) Interaction of the subunit vaccine (ribbon representation) with TLR-4/MD monomeric complex (molecular surface representation). (B) Ribbon representation of the interaction pattern of the subunit vaccine (black) with heterodimer TLR-4/MD complex (colored). (C) Illustration of the interacting residues between docked vaccine (black) and TLR-4 (colored) complex. (D) Interaction pattern of the C4 with heterodimer TLR-4/MD (ribbon representation).

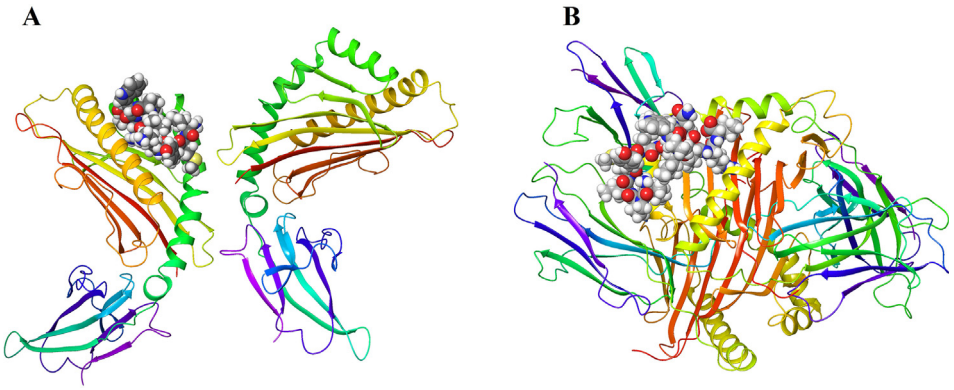

Fig. 7. The docked structures of one of the most immunodominant overlapping epitopes of the vaccine in the binding cleft of (A) MHC class I and (B) MHC class II.

Table 6
Comparative evaluation of the candidate vaccine's sequence with the experimentally validated epitopes of SARS-CoV-2 (nsp7, nsp8, nsp9, nsp10, nsp 12, and nsp14 are subunits of Orf1ab polyprotein).

#	SARS-CoV-2 proteins	Experimentally validated epitopes	Location in Orf1ab (residue)	Location in the vaccine (residue)	IEDB epitope ID
1	S glycoprotein	RQIAPGQTGKIADYNYKL	408–425	74–91	1,075,039
2	YP_009724390.1	YNYKLPDDFTGCVIA	421–435	87–101	1,074,214
3		KLPDDFTGCV	424–435	90–100	1,074,952
4		LPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERD	425–467	91–133	1,074,979
5		KLPDDFTGCV	424–433	90–99	1,074,952
6		DDFTGCVIAWNSNNL	427–441	94–107	1,069,293
7		NLDSKVGGNY	440–449	106–115	1,075,002
8		VGGNYNYLYRLFRKS	445–459	111–125	1,073,698
9		NYNYLYRLFRK	448–458	114–124	1,075,012
10		RLFRKSNLK	454–462	120–128	1,075,031
11		YLYRLFRKSNLKPFE	451–465	117–131	1,074,201
12		RKSNLKPFERDISTE	457–471	123–137	1,072,366
13		KPFERDISTEIY	462–473	128–139	1,074,954
14		AGSTPCNGVEGFNCY	475–486	141–155	1,069,064
15		NGVEGFNCY	481–489	147–155	1,075,001
16		NGVEGFNCYFPLQSY	481–498	147–161	1,071,575
17		YFPLQSYGF	489–497	155–163	1,075,121
18		NCYFPLQSYGFQPTN	487–501	153–167	1,071,518
19	Non-structural protein 7 YP_009725303.1	FEKMVSLLSV	3908–3918	201–210	15,578
20	Non-structural protein 8	AMQTMLFTM	4028–4036	225–233	3179
21	YP_009725304.1	MQTMLFTMLR	4029–4038	226–234	42,417
22		QTMLFTMLR	4030–4038	227–234	52,573
23		TMLFTMLRK	4031–4039	228–236	65,222
24	Non-structural protein 9	KVKYLYFIK	4224–4232	250–258	34,083
25	YP_009725305.1	LYFIKGLNNL	4228–4237	254–263	40,753
26		LNNLNRGMVLGSLAA	4234–4248	260–274	38,283
27		NRGMVLGSLAATVRL	4238–4252	264–278	45,682
28		RGMVLGSLAATVRLQ	4239–4253	265–279	53,926
29	Non-structural protein 10	SFGGASCCLY	4320–4327	282–291	57,776
30	YP_009725305.1	TVCTVCGMWK	4368–4377	330–339	66,958
31	RNA-dependent RNA polymerase (Non-	SICSTMTNR	4953–4961	254–362	58,467
32	structural protein 12)	MTNRQFHQK	4958–4966	359–367	42,843
33	YP_009725307.1	RQFHQKLLK	4961–4969	362–370	55,413
34		NRQFHQKLLKSIAAT	4960–4973	361–375	45,723
35		QKLLKSIAATRGATV	4965–4979	366–380	51,217
36		ATVVIGTSK	4973–4981	378–386	5209
37	3′-to-5′ exonuclease	IPLMYKGLPWNVVRI	6075–6089	396–400	27,933
38	(Non-structural protein 14)	LMYKGLPWNV	6077–6086	398–407	38,165
39	YP_009725309.1	IVQMLSDTLK	6091–6100	412–421	70,610
40		VQMLSDTLK	6092–6100	413–421	29,419

of non-structural proteins in the vaccine construct can cause cell-mediated immunity response against infected cells which makes this vaccine suitable for therapeutic purposes. Therefore, the candidate vaccine can be used for both prophylactic and therapeutic purposes.

Many scientists and even the ordinary people believe vaccine would be the key to solve the COVID-19 pandemic problem. However, developing an effective vaccine through conventional methods is time and cost consuming [118]. Therefore, many studies have focused on using immunoinformatics approaches for designing novel candidate vaccines against SARS-CoV-2 [3,106,119–122]. These *in silico* studies have focused on the structural proteins, but many recent experimental studies have emphasized on the potential of non-structural proteins for SARS-CoV-2 vaccine design. Gangaev *et al.* [116] observed that a substantial fraction of the observed CD8 + T cell responses in COVID-19 patients were direc-

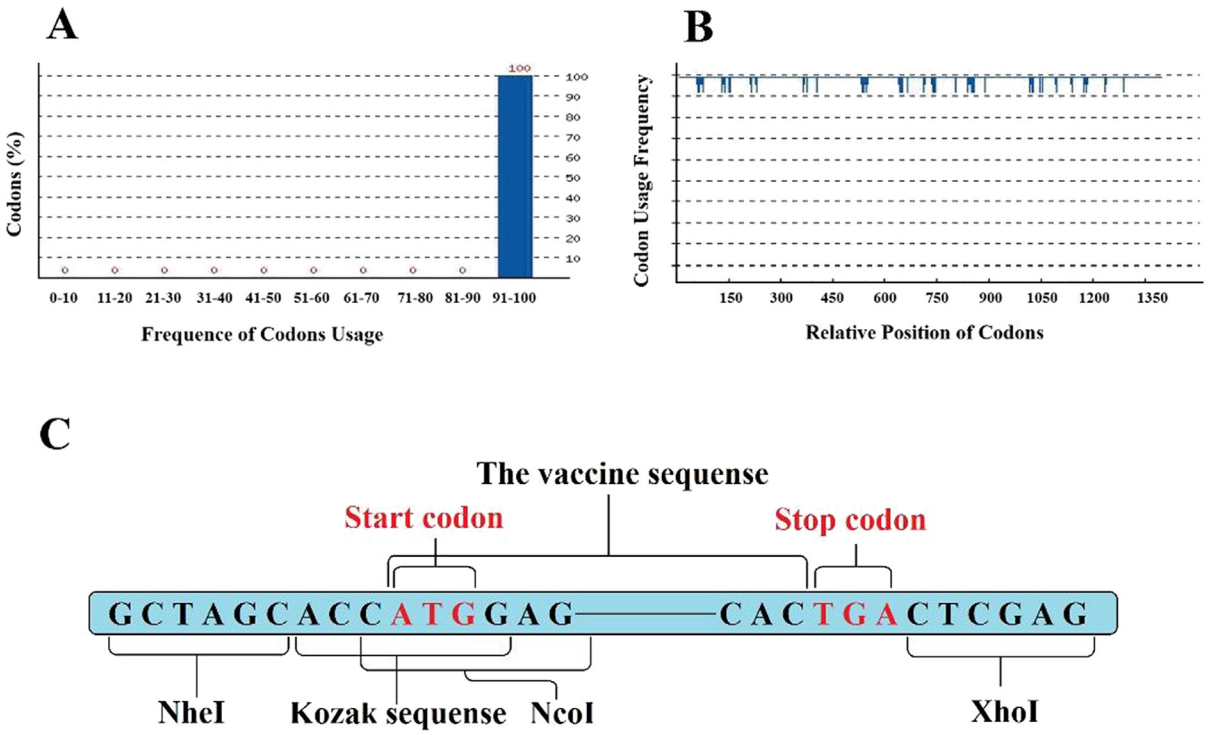

Fig. 8. (A) The codon adaptation index (CAI) and (B) codon frequency distribution of the DNA vaccine sequence. The CAI value is 0.99 for the mammalian host. Also, percentage of codon having a frequency distribution of 91–100 in mammalian host gene is 100%. (C) The schematic diagram of the DNA vaccine sequence. It contains NheI and NcoI restriction enzymes sequences at 5́ end and an XhoI sequence at the 3́ end. Also, it contains Kozak sequence at its 5́ end.

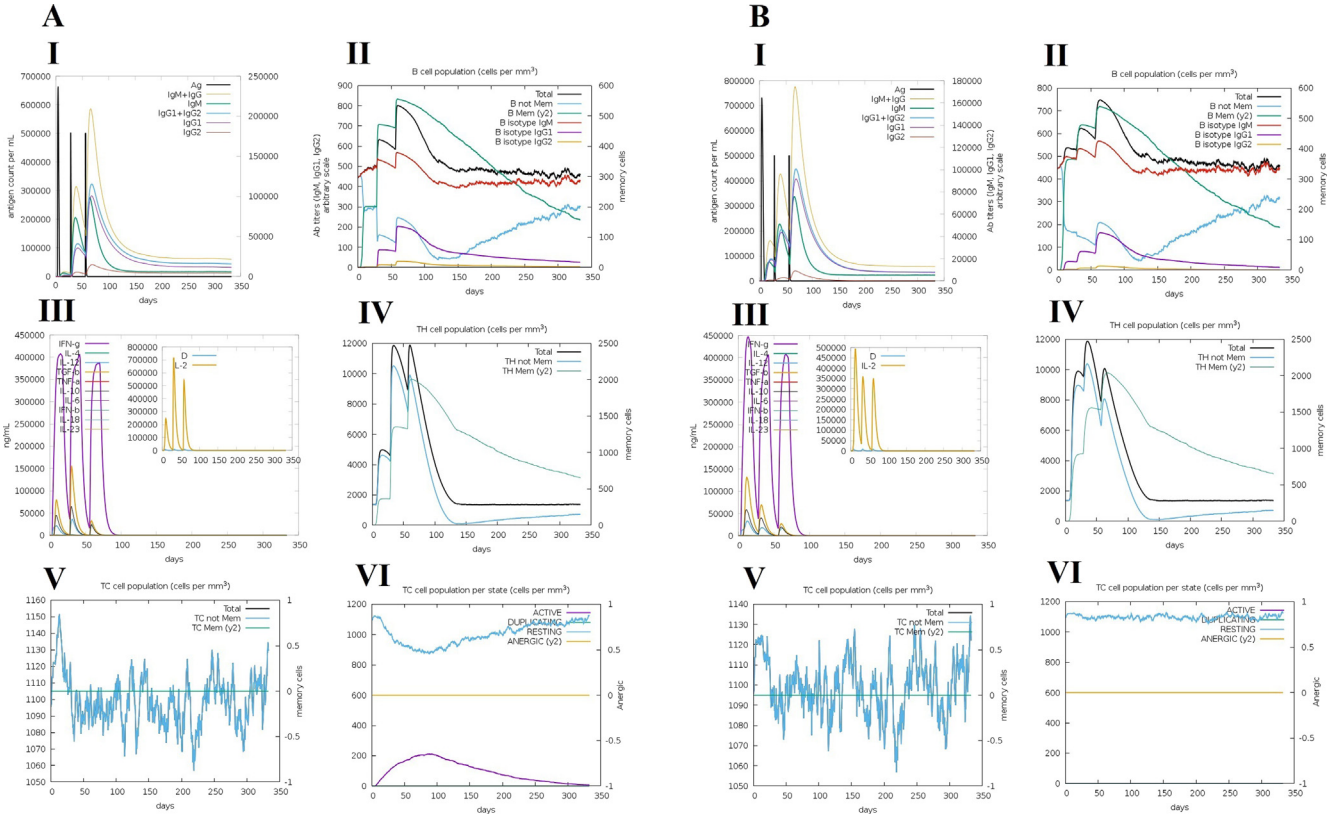

Fig. 9. Simulated immune responses against (A) the final vaccine construct and (B) the C5 positive control (SAPN-MPER of HIV-1 gp41) by the C-ImmSim server. (i) Antigen (Ag) count along with antibody titers with specific subclasses, (ii) B cells population, (iii) Cytokines responses, (iv) Th (helper) cells population, (v) Tc (cytotoxic) cells population, (vi) Tc cells population per state.

ted towards the Orf1ab polyprotein. Also, they mentioned ***"The fact that a major part of the SARS-CoV-2 specific CD8 T cell response is directed against a part of the viral genome that is not included in the majority of vaccine candidates currently in development may potentially influence their clinical activity and toxicity***". According to best of our knowledge this is the first study to design a multiepitope candidate vaccine by using the immunodominant regions of both non-structural and structural proteins of SARS-CoV-2 which makes this vaccine unique. In addition, the final vaccine construct was validated by comparing its epitopes with experimentally identified epitopes of SARS-CoV-2. Zhang *et al.* [123] mapped the immunodominant sites of S protein using sera samples collected from COVID-19 patients. They identified nine linear B cell epitopes at the S protein according to epitope mapping which four of them including 330–349, 375–394, 450–469, and 480–499 residues were located in the receptor-binding domain of S protein. Our final vaccine construct contains two of these residues including 450–469 and 480–499 residues. Also, they selected the 370–395 and 435–479 residues for mice immunization and found out that generation of high levels of specific neutralizing antibodies against SARS-CoV-2. On the other hand, strong T cells response against 480–499 residue was observed in the immunized mice and human sera. Therefore, 435–479 residue of S protein is an epitope for both T cells and B cells responses. This residue is located in our final vaccine construct. In addition, Berry *et al.* [124] nominated 460–492 reside of S protein as "Achilles heel" of SARS-CoV-2 as this residue is exposed regardless of the glycosylation condition and has a critical role in the virus attachment to the host ACE2 receptor. They observed significant production of high efficacy neutralizing antibodies against this region in their experiments. Therefore, multiple epitopes which were identified at experimental studies for humoral and cellular responses against SARS-CoV-2 were present at our designed candidate vaccine. Taking together, this vaccine can be an appropriate platform for further experiments by vaccinologist.

5. Conclusion

Currently, world is facing COVID-19 pandemic and a global competition is ongoing to discover the vaccine. This study presents *in silico* designing of a candidate multi-epitope vaccine to provoke both innate and adaptive immunity against SARS-CoV-2 according immunoinformatic approaches. Different domains were located in the final vaccine construct to enhance its efficacy. In spite of high efficacy of the designed candidate vaccine according to *in silico* analyses, further experimental studies are necessary for the validation of the designed vaccine in different aspects.

6. Contributions

A. Safavi, A. Kefayat, E. Mahdevar, and F. Ghahremani analyzed the antigen and designed the vaccines. The molecular docking was carried out by A. Abiri. The manuscript writing and revisions were done by A. Safavi, A. Kefayat, E. Mahdevar, and F. Ghahremani.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by Arak University of Medical Sciences (grant number: 3626).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vaccine.2020.10.016.

References

[1] Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 2020;109:102433. https://doi.org/10.1016/j.jaut.2020.102433.
[2] Tang X et al. Differential stepwise evolution of SARS coronavirus functional proteins in different host species. BMC Evol Biol 2009;9(1):52.
[3] Sunita et al. Computational tools for modern vaccine development. Hum Vaccin Immunother 2020;16(3):723–35.
[4] Kazi A et al. Current progress of immunoinformatics approach harnessed for cellular-and antibody-dependent vaccine design. Pathog Glob Health 2018;112(3):123–31. https://doi.org/10.1080/20477724.2018.1446773.
[5] Shin H-S et al. Immune responses to Middle East respiratory syndrome coronavirus during the acute and convalescent phases of human infection. Clin Infect Dis 2019;68(6):984–92. https://doi.org/10.1093/cid/ciy595.
[6] Ortega JT et al. Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: An in silico analysis. EXCLI J 2020;19:410–7. https://doi.org/10.17179/excli2020-1167.
[7] Oany AR et al. Design of an epitope-based peptide vaccine against spike protein of human coronavirus: an in silico approach. Drug Des Devel Ther 2014;21(8):1139–49.
[8] Kumar, S.J., Drug and vaccine design against Novel Coronavirus (2019-nCoV) spike protein through Computational approach. 2020.
[9] Ul Qamar MT et al. Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. J Transl Med2019 2019;17(1):362.
[10] Ahmed SF, Quadeer AA, McKay MRJV. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses 2020;12(3):254. https://doi.org/10.3390/v12030254.
[11] Watanabe, Y., et al., Exploitation of glycosylation in enveloped virus pathobiology. BBA-GEN SUBJECTS 2019.1863(10),1480-1497.
[12] Zhang Y et al. Site-specific N-glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins using High-Resolution Mass. Spectrometry. 2020.
[13] Johnson CR, Yu W, Murtaugh MPJJoGV. Cross-reactive antibody responses to nsp1 and nsp2 of Porcine reproductive and respiratory syndrome virus. J Gen Virol 2007(Apr;88(Pt 4):):1184–95. https://doi.org/10.1099/vir.0.82587-0.
[14] Khailany, R.A., M. Safdar, and M.J.G.r. Ozaslan, Genomic characterization of a novel SARS-CoV-2. Gene Rep. 2020 Jun;19:100682. doi: 10.1016/j.genrep.2020.100682. Epub 2020 Apr 16.
[15] Paul S et al. TepiTool: a pipeline for computational prediction of T cell epitope candidates. Curr Protoc Immunol 2016 Aug;1114:18.19.1–18.19.24. https://doi.org/10.1002/cpim.12.
[16] Vita R et al. The immune epitope database 2.0. Nucleic Acids Res 2009;38(suppl_1):D854–62.
[17] Nezafat N et al. A novel multi-epitope peptide vaccine against cancer: an in silico approach. J Theor Biol 2014;349:121–34.
[18] Hattotuwagama CK et al. Quantitative online prediction of peptide binding to the major histocompatibility complex.. J MOL GRAPH MODEL 2004;22(3):195–207.
[19] Chakraborty I et al. Massive electrical conductivity enhancement of multilayer graphene/polystyrene composites using a nonconductive filler. ACS Appl. Mater. Interfaces 2014;6(19):16472–5. https://doi.org/10.1021/am5044592.
[20] Nielsen M, Lund OJBb. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinformatics 2009;1810:296. https://doi.org/10.1186/1471-2105-10-296.
[21] Nielsen M, Lundegaard C, Lund OJBb. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinformatics 2007;8(1):238.
[22] Jensen KK et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology 2018;154(3):394–406. https://doi.org/10.1111/imm.12889Epub 2018 Feb 6.
[23] Khalili S et al. In silico analyses of Wilms tumor protein to designing a novel multi-epitope DNA vaccine against cancer. J Theor Biol 2015;379:66–78.
[24] Stranzl T et al. NetCTLpan: pan-specific MHC class I pathway epitope predictions. Immunogenetics 2010;62(6):357–68.
[25] Pandey RK, Bhatt TK, Prajapati VK. Novel Immunoinformatics Approaches to Design Multi-epitope Subunit Vaccine for Malaria by Investigating Anopheles Salivary Protein. Sci Rep 2018;8(1):1125.
[26] Larsen MV et al. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinf 2007;8(1):424.
[27] Lo, Y.-S., C.-Y. Lin, and J.-M.J.N.a.r. Yang, PCFamily: a web server for searching homologous protein complexes. Nucleic Acids Res. 2010 Jul;38(Web Server issue):W516-22. doi: 10.1093/nar/gkq464. Epub 2010 May 28.
[28] Dhanda SK et al. Predicting HLA CD4 immunogenicity in human populations. Front Immunol. 2018;9:1369.

[29] Saha S, Raghava GPSJPS. Function,, and Bioinformatics, Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 2006 Oct;165(1):40–8. https://doi.org/10.1002/prot.21078.
[30] Saha S, Raghava GPS. BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. International Conference on Artificial Immune Systems. Springer; 2004.
[31] Ponomarenko J et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics. 2008;9(1):514. https://doi.org/10.1186/1471-2105-9-514.
[32] Ikram A et al. Exploring NS3/4A, NS5A and NS5B proteins to design conserved subunit multi-epitope vaccine against HCV utilizing immunoinformatics approaches. Sci. Rep 2018;8(1):1–14.
[33] Gasteiger E et al. Protein identification and analysis tools on the ExPASy server. Methods Mol Biol 1999;112:531531–52. https://doi.org/10.1385/1-59259-584-7.
[34] Pritam, M., et al., A cutting-edge immunoinformatics approach for design of multi-epitope oral vaccine against dreadful human malaria. Int J Biol Macromol 2020. 1;158:159-179. https://doi.org/10.1016/j.ijbiomac.2020.04.191.
[35] Kaba SA et al. Self-assembling protein nanoparticles with built-in flagellin domains increases protective efficacy of a Plasmodium falciparum based vaccine. Vaccine. 2018;36(6):906–14.
[36] Ahmadi K et al. Epitope-based immunoinformatics study of a novel Hla-MntC-SACOL0723 fusion protein from Staphylococcus aureus: Induction of multi-pattern immune responses. Mol Immunol . 2019;114:88–99.
[37] Safavi, A., et al., Efficacy of co-immunization with the DNA and peptide vaccines containing SYCP1 and ACRBP epitopes in a murine triple-negative breast cancer model. Hum Vaccin Immunother. 2020 Jun 4;1-13. https://doi.org/10.1080/21645515.2020.1763693.
[38] Xie J et al. The VHSE-based prediction of proteasomal cleavage sites. Plos One 2013;8:0074506.
[39] Nielsen M et al. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. Immunogenetics 2005;57(1–2):33–41.
[40] Sankar S et al. Short peptide epitope design from hantaviruses causing HFRS. Bioinformation 2017;13(7):231.
[41] Doytchinova IA, Flower D. Bioinformatic approach for identifying parasite and fungal candidate subunit vaccines. The Open Vaccine Journal 2008;1(1):22–6.
[42] Cornejo-Granados F et al. Secret-AAR: a web server to assess the antigenic density of proteins and homology search against bacterial and parasite secretome proteins. Genomics 2019;111(6):1514–6.
[43] Magnan CN et al. High-throughput prediction of protein antigenicity using protein microarray data. Bioinformatics . 2010;26(23):2936–43.
[44] Singh SP, Roopendra K, Mishra BNJJotm. Genome-wide prediction of vaccine candidates for Leishmania major: an integrated approach. Journal of Tropical Medicine 2015:709216.
[45] Maurer-Stroh S et al. AllerCatPro—prediction of protein allergenicity potential from the protein sequence. Bioinformatics 2019;35(17):3020–7.
[46] Adhikari UK, Tayebi M, Rahman MMJJoir. Immunoinformatics approach for epitope-based peptide vaccine design and active site prediction against polyprotein of emerging oropouche virus. Journal of Immunologt Research 2018;6718083.
[47] Schrodinger LJV. The PyMOL molecular graphics system. 2010;1(5).
[48] BaCelLo: a balanced subcellular localization predictor. Bioinformatics 2006;22(14):e408–16.
[49] Briesemeister S et al. SherLoc2: a high-accuracy hybrid method for predicting subcellular localization of proteins. J. Proteome Res. 2009;8(11):5363–6. https://doi.org/10.1021/pr900665y.
[50] Tornavaca O et al. KAP degradation by calpain is associated with CK2 phosphorylation and provides a novel mechanism for cyclosporine A-induced proximal tubule injury. PLoS One. 2011;6(9): e25746. https://doi.org/10.1371/journal.pone.0025746Epub 2011 Sep 28.
[51] Cai C et al. SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence. Nucleic Acids Res . 2003;31(13):3692–7.
[52] Song Y et al. High-resolution comparative modeling with RosettaCM. Structure 2013;21(10):1735–42.
[53] Roy A, Kucukural A, Zhang YJNp. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 2010;5(4):725–38.
[54] Källberg M et al. Template-based protein structure modeling using the RaptorX web server. Nat Protoc 2012;7(8):1511–22.
[55] Kelley LA et al. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 2015;10(6):845–58. https://doi.org/10.1038/nprot.2015.053.
[56] Chen VB et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010;66(1):12–21.
[57] Raman S et al. Structure prediction for CASP8 with all-atom refinement using Rosetta. Proteins 2009;77(S9):89–99. https://doi.org/10.1002/prot.22540.
[58] Johansson MU et al. Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. BMC Bioinformatics. 2012;13(173). https://doi.org/10.1186/1471-2105-13-173.
[59] Guex N, Peitsch MC, Schwede TJE. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective 2009;30(S1):S162–73.
[60] Ponomarenko J et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinf 2008;9(1):514.
[61] Kringelum JV et al. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. PLoS Comput Biol 2012;8(12):e1002829.
[62] Blom N et al. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics 2004;4(6):1633–49.
[63] Steentoft C et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. The EMBO journal 2013;32(10):1478–88.
[64] Radivojac P et al. Identification, analysis, and prediction of protein ubiquitination sites. Proteins Struct Funct Bioinf 2010;78(2):365–80.
[65] Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites1. J Mol Biol 1999;294(5):1351–62.
[66] Kozakov D et al. How good is automated protein docking?. Proteins 2013;81(12):2159–66.
[67] Comeau SR et al. ClusPro: a fully automated algorithm for protein–protein docking. Nucleic Acids Res. 2004;32(suppl_2):W96–9.
[68] Kozakov D et al. The ClusPro web server for protein–protein docking. Nat Protoc . 2017;12(2):255–78.
[69] Kozakov D et al. an FFT-based protein docking program with pairwise potentials. Proteins 2006;65(2):392–406.
[70] Yamada C et al. Surfactant protein a directly interacts with TLR4 and MD-2 and regulates inflammatory cellular response importance of supratrimeric oligomerization. J Biol Chem 2006;281(31):21771–80.
[71] Friesner RA et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem. 2004;47(7):1739–49.
[72] Halgren TA et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem 2004;47(7):1750–9.
[73] Friesner RA et al. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein− ligand complexes. J Med Chem 2006;49(21):6177–96.
[74] Pourshojaei Y et al. phenoxyethyl piperidine/Morpholine Derivatives as pAS and cAS inhibitors of cholinesterases: insights for future Drug Design. Sci. Rep 2019;9(1):1–19.
[75] Fleri W et al. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. Front Immunol 2017;8(278):278. https://doi.org/10.3389/fimmu.2017.00278.eCollection 2017.
[76] Grote A et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 2005;33(suppl_2):W526–31.
[77] Rapin N et al. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. PLoS One 2010;5(4) e9862. https://doi.org/10.1371/journal.pone.0009862.
[78] Peele KA et al. Design of multi-epitope vaccine candidate against SARS-CoV-2: a in-silico study. J Biomol Struct Dyn 2020:1–9.
[79] Majid M, Andleeb SJSR. Designing a multi-epitopic vaccine against the enterotoxigenic Bacteroides fragilis based on immunoinformatics approach. Sci. Rep 2019;9(1):1–15.
[80] Castiglione, F., et al., How the interval between prime and boost injection affects the immune response in a computational model of the immune system. Comput Math Methods Med. 2012;2012:842329. doi: 10.1155/2012/842329. Epub 2012 Sep 11.
[81] Singh G et al. Designing of precise vaccine construct against visceral leishmaniasis through predicted epitope ensemble: a contemporary approach. Comput Biol Chem 2020:107259.
[82] Wahome N et al. Conformation-specific display of 4E10 and 2F5 epitopes on self-assembling protein nanoparticles as a potential HIV vaccine. Chem Biol Drug Des 2012;80(3):349–57. https://doi.org/10.1111/j.1747-0285.2012.01423.x.
[83] Seder RA, Hill AVJN. Vaccines against intracellular infections requiring cellular immunity. Nature 2000;406(6797):793–8. https://doi.org/10.1038/35021239.
[84] Gfeller D, Bassani-Sternberg MJFii. Predicting antigen presentation—What could we learn from a million peptides?. Front Immunol 2018;9:1716.
[85] Singh SP, Mishra BNJHi. Major histocompatibility complex linked databases and prediction tools for designing vaccines. Hum Immunol 2016;77(3):295–306.
[86] Dey AK, Malyala P, Singh MJErov. Physicochemical and functional characterization of vaccine antigens and adjuvants. Expert Rev Vaccines 2014;13(5):671–85.
[87] Brandau DT et al. Thermal stability of vaccines. J Pharm Sci 2003;92(2):218–31. https://doi.org/10.1002/jps.10296.
[88] Rawi R et al. PaRSnIP: sequence-based protein solubility prediction using gradient boosting machine. Bioinformatics 2018;34(7):1092–8. https://doi.org/10.1093/bioinformatics/btx662.
[89] Beebe M et al. Formulation and characterization of a ten-peptide single-vial vaccine EP-2101, designed to induce cytotoxic T-lymphocyte responses for cancer immunotherapy. Hum Vaccin 2008;4(3):210–8. https://doi.org/10.4161/hv.4.3.5291.
[90] Prasad S, Khadatare PB, Roy I. Effect of chemical chaperones in improving the solubility of recombinant proteins in Escherichia coli. Appl Environ Microbiol 2011;77(13):4603–9.

[91] Vazquez E, Corchero JL, Villaverde A. Post-production protein stability: trouble beyond the cell factory. Microb Cell Fact 2011;10(1):60.
[92] Klasse PJAib. Neutralization of virus infectivity by antibodies: old problems in new perspectives. Advances in Biology 2014:157895.
[93] Zobayer N, Hossain AA, Rahman MAJB. A combined view of B-cell epitope features in antigens. Bioinformation 2019;15(7):530–4. https://doi.org/10.6026/97320630015530.
[94] Shi J et al. Epitope-based vaccine target screening against highly pathogenic MERS-CoV: an in silico approach applied to emerging infectious diseases. PLoS One 2015;10(12) e0144475. https://doi.org/10.1371/journal.pone.0144475.
[95] Safavi A et al. In silico analysis of synaptonemal complex protein 1 (SYCP1) and acrosin binding protein (ACRBP) antigens to design novel multiepitope peptide cancer vaccine against. Int. J. Pept. Res. Ther. 2019;25(4):1343–59.
[96] Puigbo P et al. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res 2007;35(suppl_2):W126–31. https://doi.org/10.1093/nar/gkm219.
[97] Safavi A et al. Production, purification, and in vivo evaluation of a novel multiepitope peptide vaccine consisted of immunodominant epitopes of SYCP1 and ACRBP antigens as a prophylactic melanoma vaccine. Int Immunopharmacol 2019;76 105872. https://doi.org/10.1016/j.intimp.2019.105872.
[98] Luo D et al. Protective immunity elicited by a divalent DNA vaccine encoding both the L7/L12 and Omp16 genes of Brucella abortus in BALB/c mice. Infect Immun 2006;74(5):2734–41.
[99] Salehi-Sangani G et al. Immunization against Leishmania major infection in BALB/c mice using a subunit-based DNA vaccine derived from TSA, LmSTI1, KMP11, and LACK predominant antigens. Iran J Basic Med Sci. 2019;22(12):1493–501.
[100] Gupta J et al. Immunogenicity and protective efficacy of Brugia malayi heavy chain myosin as homologous DNA, protein and heterologous DNA/protein prime boost vaccine in rodent model. PLoS One 2015;10(11):e0142548.
[101] Faezi S et al. High yield overexpression, refolding, purification and characterization of Pseudomonas aeruginosa type B-flagellin: an improved method without sonication. Int J Mol Cell Med 2016;5(1):37–48.
[102] Fang L et al. Efficacy of CpG-ODN and Freund's immune adjuvants on antibody responses induced by chicken infectious anemia virus VP1, VP2, and VP3 subunit proteins. Poult. Sci. J. 2019;98(3):1121–6.
[103] Singh G et al. Genome based screening of epitope ensemble vaccine candidates against dreadful visceral leishmaniasis using immunoinformatics approach. Microb Pathog 2019;136 103704. https://doi.org/10.1016/j.micpath.2019.103704.
[104] Ni L et al. Detection of SARS-CoV-2-specific humoral and cellular immunity in COVID-19 convalescent individuals. Immunity 2020;52(6):971–7.
[105] Stavnezer J, Schrader CEJTJoI. IgH chain class switch recombination: mechanism and regulation.J. Immunol 2014;193(11):5370–8.
[106] Ka T et al. A Candidate multi-epitope vaccine against SARS-CoV-2. Int J Biol Macromol 2020;164:871–83.
[107] Nabel GJJNEJoM. Designing tomorrow's vaccines. N Engl J Med 2013;368(6):551–60.
[108] Enayatkhani M et al. Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. J Biomol Struct Dyn 2020;just-accepted:1–19.
[109] Du L et al. The spike protein of SARS-CoV—a target for vaccine and therapeutic development. Nat Rev Microbiol 2009;7(3):226–36.
[110] Hagemeijer MC et al. Dynamics of coronavirus replication-transcription complexes. J Virol 2010;84(4):2134–49.
[111] Li H et al. Proximal glycans outside of the epitopes regulate the presentation of HIV-1 envelope gp120 helper epitopes. J Immunol 2009;182(10):6369–78.
[112] Wolfert MA, Boons G-JJNcb. Adaptive immune activation: glycosylation does matter. Nat Chem Biol 2013;9(12):776–84.
[113] Sexton NR et al. Homology-based identification of a mutation in the coronavirus RNA-dependent RNA polymerase that confers resistance to multiple mutagens. J Virol 2016;90(16):7415–28. https://doi.org/10.1128/JVI.00080-16.
[114] Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. In: Coronaviruses. Springer; 2015. p. 1–23.
[115] Campbell, K.M., et al., Prediction of SARS-CoV-2 epitopes across 9360 HLA class I alleles. 2020.
[116] Gangaev, A., et al., Profound CD8 T cell responses towards the SARS-CoV-2 ORF1ab in COVID-19 patients. 2020.
[117] Garbuglia AR et al. The Use of Both Therapeutic and Prophylactic Vaccines in the Therapy of Papillomavirus Disease. Front Immunol 2020;11(188). https://doi.org/10.3389/fimmu.2020.00188.
[118] Soria-Guerra RE et al. An overview of bioinformatics tools for epitope prediction: implications on vaccine development. J Biomed Inform 2015;53:405–14. https://doi.org/10.1016/j.jbi.2014.11.003.
[119] Sarkar, B., et al., Immunoinformatics-guided designing of epitope-based subunit vaccine against the SARS Coronavirus-2 (SARS-CoV-2). Immunobiology 2020: 225 (3). 151955.
[120] Bhattacharya M et al. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2). Immunoinformatics approach 2020;92(6):618–31.
[121] Joshi A et al. Epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach. Inform. Med. Unlocked 2020;19:100338.
[122] Abdelmageed MI 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:. https://doi.org/10.1155/2020/2683286.
[123] Huang, J.-D., B.-z. Zhang, and Y.-f.J.b. Hu, Mapping the Immunodominance Landscape of SARS-CoV-2 Spike Protein for the Design of Vaccines against COVID-19. 2020.
[124] Berry JD et al. Neutralizing epitopes of the SARS-CoV S-protein cluster independent of repertoire, antigen structure or mAb technology. MAbs. Taylor & Francis; 2010.