STUDIA UNIVERSITATIS BABEŞ-BOLYAI BIOLOGIA, 69, 1, 2024 (p. 11-32)
doi:10.24193/subbbiol.2024.1.01

Designing a multi-epitope candidate vaccine against SARS-CoV-2 through *in silico* approach for producing in plant systems

Fatemeh Goudarziasl[1], Fatemeh Kheiri[2], Azam Rahbar[3], Reza Mohammadhassan[3]✉, Javad Mohammadi-Asl[4], Arsalan Jalili[5,6], Melika Hajkazemian[7]

[1]Department of Biochemistry, Faculty of Biological Science, University of Mazandaran, Mazandaran, Iran; [2]Department of Biotechnology, Faculty of Biological Sciences, Alzahra University, Tehran, Iran; [3]Amino Techno Gene Private Virtual Lab (NGO), Tehran, Iran; [4]Noorgene Genetic & Clinical Laboratory, Molecular Research Center, Ahvaz, Iran; [5]Department of Stem Cells and Developmental Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACER, Tehran, Iran; [6]Parvaz Research Ideas Supporter Institute, Tehran, Iran; [7]Department of Molecular Biosciences, Wenner-Gren Institute, Stockholm University, Stockholm, Sweden;
✉***Corresponding author, E-mail: rezarmhreza22@gmail.com***

Article history: Received 04 August 2023; Revised 20 January 2024; Accepted 25 January 2024; Available online 30 June 2024.

Abstract. The COVID-19 is considered as a type of severe acute respiratory syndrome (SARS-CoV-2). The current pandemic causes a vital destruction in international social and economic systems. Current available vaccines involve entire viruses; however, peptide-based vaccines could be also beneficial. In the present study, a computationally candidate vaccine was designed against SARS-CoV-2. Surface glycoproteins (E, M, and S proteins) and N protein amino acid sequences were analyzed to predict high score of the B and T cell epitopes as antigenic proteins of the virus. High score epitopes, and the B subunit of *Vibrio cholerae* toxin, as an adjuvant put together by appropriate linkers to construct a multi-epitope candidate vaccine. Bioinformatics tools were used to predict the secondary, tertiary structure and physicochemical properties, such as aliphatic index, theoretical pH, molecular weight, and estimated half-life of the multi-epitope candidate vaccine. The interaction of candidate vaccine with TLR2 and TLR4 was computationally evaluated by molecular docking. Finally, the codon optimization and the secondary structure of mRNA were calculated, and

in silico cloning was performed into plant expression vector by SnapGENE. This designed candidate vaccine along with the computational results requires laboratory evaluations to be confirmed as a candidate vaccine against SARS-COV-2 infection.

Keywords: COVID-19, SARS-CoV-2, in silico, Multi-epitope candidate vaccine, Plant systems.

Introduction

Since late 2019, the prevalence of the COVID-19, as a viral infection caused by a novel coronavirus called SARS-CoV-2, has been a vital concern for human society due to severe acute respiratory syndrome caused by the virus and rapid worldwide outbreak. Elders, pregnant women, and individuals suffering an underlying disease such as diabetes and immune deficiency could be less resistant against COVID-19 (Huang *et al*., 2020; Zhou *et al*., 2020). The virus can be transmitted among people by respiratory droplets and through faecal-oral route (Dibner, 2021). The droplets can be in different sizes from <5 μm to >5-10 μm (Boopathi *et al*., 2020). There is a wide range, from mild to serve, of symptoms for COVID-19. Symptoms can appear 2-14 days after the virus infects cells. The symptoms of COVID-19 include fever, cough, chills, shortness of breath or difficult breathing, fatigue, headache, sore throat, nausea or vomiting, and diarrhea (Nokhostin *et al*., 2020).

Coronavirus is an enveloped non-segmented virus containing a positive-single-strain RNA. The genome of the virus is about 26 to 32 Kbp as one of the most extended viral RNAs (Fehr and Perlman, 2015). There are four genotypes and serotypes of coronavirus to classify as α, β, γ, and δ. Novel coronavirus-19 belongs to the β-coronavirus genus (Lim *et al*., 2016; Rahbaran, 2021). COVID-19 is the third detected viral infection caused by Coronaviridae family. Previously, severe acute respiratory syndrome (SARS) and middle east respiratory syndrome (MERS) diseases have been identified and studied, respectively, as the first and second viral infections caused by species of the β-coronavirus genus (Amanat and Krammer, 2020; Yoshimoto, 2020). However, the outbreak of the viral infection needs to be control by the immunization. There are recently 6 novel vaccines immunizing individuals against SARS-CoV-2, including Moderna, Johnson and Johnson (made in the USA), Pfizer-Biontech (made by cooperation of US and Germany), Astera-Zeneka (made by cooperation of UK and Sweden), Sinopharm (made in China), and Sputnik V (made by Russia) (Zheng *et al*., 2022).

As an *in silico* approach, immunoinformatics contains various reliable and precise tools to analyze the new candidate vaccines. In this strategy, the conserved and effective epitopes, including the smallest part of the antigen's protective activity, were used to design a new vaccine instead of using the whole antigen. The advantages of this approach include low-cost production and a more specific design of the subunit vaccine. This strategy can be recommended to manage many epidemics and even pandemics such as COVID-19 (Miles *et al*., 2019; Shey *et al*., 2019; Mohammadhassan *et al*., 2020). It has been demonstrated that the *in silico* designed vaccines against various infections can positively affect cellular and humoral immunity (Liu and Chen, 2004, Cong *et al*., 2008, Hajissa *et al*., 2019). In the present study, a candidate multi-epitope vaccine against SARS-CoV-2 was designed and investigated by *in silico* tools regarding high score B and T cell epitopes of the structural proteins.

Materials & Methods

Protein selection for preparing multi-epitope protein

The protein database in NCBI (https://www.ncbi.nlm.nih.gov/protein) was used to access the sequences of the E protein (accession number = QHD43418.1), M protein (accession number = QHD43419.1), S protein (accession number = QHD43416.1) and N protein (accession number = QHD43423.2) of SARS-CoV-2, as FASTA format for further analysis.

Predicting linear B-cell epitopes

BpiPred (http://www.cbs.dtu.dk/services/BepiPred), ABCpred (http://crdd.osdd.net/raghava/abcpred), and SVMTrip (http://sysbio.unl.edu/SVMTriP) were employed for linear B-cell epitopes prediction. Moreover, VaxiJen v.2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html), ToxinPred (http://crdd.osdd.net/raghava/toxinpred), AllerTOP v.2.0 (https://www.ddg-pharmfac.net/AllerTOP) and PepCalc (https://pepcalc.com) were respectively used to evaluate antigenicity, toxicity, allergenicity, and solubility of the high score predicted linear B-cell epitopes.

Predicting T-cell epitopes

The Immune Epitope Database server (IEBD) was utilized to predict MHC class I epitopes regarding 9 mer HLA reference alleles. T-cell epitope prediction was performed on S, E, N, and M proteins. High-score epitopes (regarding percentile rank and inhibitory concentration (IC50)) were evaluated in the aspect of antigenicity, toxicity, hydrophobicity, allergenicity, and immunogenicity

by VaxiJen v.2.0, ToxinPred**,** Peptide 2.0, AllerTOP v.2.0, immune epitope (http://tools.immuneepitope.org/mhci/result) servers, respectively. To predict MHC class II epitopes, we also used the IEDB server (which was employed (http://tools.iedb.org/mhcii/) regarding IEDB recommended 2.22 method and 15 mer total HLA reference alleles. Epitopes with low percentile rank and NetMHCIIpan IC50 (nM) were selected, then analyzed by VaxiJen v.2.0, ToxinPred**,** Peptide 2.0 (https://www.peptide2.com), and AllerTOP v.2.0 to identify their antigenicity, toxicity, hydrophobicity, allergenicity, and immunogenicity, respectively.

Gamma interferon production

To identify regions with the potential gamma interferon production stimulation in the candidate multi-epitope vaccine, we used the IFNepitope server (http://crdd.osdd.net/raghava/ifnepitope) by hybrid approaches.

Multi-epitope subunit vaccine design

Nontoxic and non-allergenic high score epitopes with proper antigenicity and solubility were linked together with the aid of appropriate linkers to construct a multi-epitope vaccine structure. The B-cell and the T-cell epitopes were respectively bound GPGPG and the KK linkers. To better stimulate the immune system by final construction, nontoxic cholera subunit B (CTB) was used as an adjuvant and linked by EAAAK linker. The final construction was screened for antigenicity by VaxiJen v 2.0 and ANTIGENpro (http://scratch. proteomics.ics.uci.edu), respectively. AllerTOP v. 2.0 was used for the evaluation of allergenicity. For probable trans-membrane regions and signal peptide, the sequence of the multi-epitopes candidate vaccine was checked by the TMHMM v.2.0 server (http://www.cbs.dtu.dk/services/TMHMM) and TargetP 5.0 (http://www.cbs.dtu.dk/services/SignalP/index.php), respectively.

Physicochemical properties

The physicochemical properties were evaluated by Expasy ProtParam tool (https://web.expasy.org/protparam). This server computes parameters such as amino acid composition, molecular weight, in-vitro and in-vivo half-life, theoretical isoelectric point (pI), aliphatic index, grand average of hydropathicity (GRAVY), and instability index. The aliphatic index explains the thermo-stability of the query protein.

Secondary and tertiary structure prediction

I-TASSER server (https://zhanggroup.org/I-TASSER/) was employed for the prediction of Tertiary structure. This server is an online tool for annotating structure-based function and predicting automated protein structure (Yang and Zhang, 2015).

Refinement and validation of tertiary structure

ModRefiner server (https://zhanggroup.org/ModRefiner/) was used for refining and modeling protein structure prediction, and then Prosa-web (https://prosa.services.came.sbg.ac.at/prosa.php), and Rampage server (https://zlab.umassmed.edu/bu/rama/) were respectively employed for the structural validation.

Molecular docking

The HDock online server (https://cluspro.bu.edu) was used for docking the refined candidate vaccine construction by TLR4. For this purpose, the PDB structures of TLR 4 (PDB ID: 3FXI) were provided from the RCSB PDB server (https://www.rcsb.org/). As a receptor, this PDB structure, along with its ligand, which is our refined constructed protein, were given to the HDock server (http://hdock.phys.hust.edu.cn/).

Molecular dynamic simulation

iMOD server (https://imods.iqfr.csic.es/) was used to explain the stability of the candidate vaccine-TLR2/TLR4 complexes and perform molecular dynamics and RMSD calculations.

Predicting codon optimization and mRNA secondary structure

The synthetic construct was codon optimization for expression in the desired prokaryote (*E. coli*) and eukaryote (plants) host by different online software (IDT DNA, optimizer, cool, Jcat, and Genscript). The amino acid sequence of the candidate vaccine along with the codon usage table was subjected to OPTIMIZER (http://genomes.urv.es/OPTIMIZER/). The optimized DNA was submitted to GENScript and some features such as GC content, Codon Adaption Index (CAI), and Codon Frequency Distribution (CFD) were evaluated. For predicting the mRNA secondary structure, the optimized DNA sequence was submitted to the RNAfold server (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). SnapGENE was also used for the insertion of our construct into the PBI121 vector.

Results

Linear B-cell epitopes prediction

Full-length S and N proteins were subjected to ABCpred, SVMTrip, and BepiPred, separately. Regarding these servers, 110 epitopes were predicted. High score epitopes from each server were analyzed from the aspects of antigenicity, toxicity, allergenicity, and solubility by VaxiJen v.2.0, ToxinPred, AllerTOP v.2.0, and PepCalc, respectively. Finally, among 13 non-allergic and nontoxic epitopes with proper antigenicity and solubility, 4 epitopes with IFNepitope positive were selected to participate in the multi-epitope candidate vaccine. Final selected Linear B-cell epitopes which fulfilled all the criteria for non-allergenicity, antigenicity, non-toxicity, and could also induce the IFN-γ immune response (Tab. 1).

T-cell epitopes prediction

Epitopes with low percentile rank and IC50 have a high affinity with MHCs. Consequently, antigenic, nontoxic, hydrophobic, and non-allergen epitopes with low percentile rank and IC50 were selected and participated in the multi-epitope candidate vaccine (Tab. 2).

Table 1. Linear B-cell epitopes prediction

EPITOPE	Protein name	Start position	Predicted score
QQQQGQTVTKKSAAEASKKP	N protein	239	1.000
RRGPEQTQGNFGDQELIRQG	N protein	276	0.865
HGKEDLKFPRGQGVPI	N protein	59	0.87
TRRIRGGDGKMKDLSP	N protein	91	0.94

Table 2. T-cell epitopes prediction

Epitope	PR name	Start position	MHC binding
VIGFLFLTW	M protein	23	MHC-I
KLIFLWLLW	M protein	50	MHC-I
TLACFVLAA	M protein	61	MHC-I
ACFVLAAVY	M protein	63	MHC-I

Epitope	PR name	Start position	MHC binding
FVLAAVYRI	M protein	65	MHC-I
IAIAMACLV	M protein	80	MHC-I
SELVIGAVI	M protein	136	MHC-I
LVIGAVILR	M protein	138	MHC-I
YYKLGASQR	M protein	178	MHC-I
RYRIGNYKL	M protein	198	MHC-I
LPFNDGVYF	S protein	84	MHC-I
GVYFASTEK	S protein	89	MHC-I
TLDSKTQSL	S protein	109	MHC-I
YYHKNNKSW	S protein	144	MHC-I
FEYVSQPFL	S protein	168	MHC-I
KIYSKHTPI	S protein	202	MHC-I
WTAGAAAYY	S protein	258	MHC-I
YYVGYLQPR	S protein	265	MHC-I
KSNLKPFER	S protein	458	MHC-I
PYRVVVLSF	S protein	507	MHC-I
QLTPTWRVY	S protein	628	MHC-I
SPRRARSVA	S protein	680	MHC-I
LGAENSVAY	S protein	699	MHC-I
IAIPTNFTI	S protein	712	MHC-I
IPTNFTISV	S protein	714	MHC-I
FTISVTTEI	S protein	718	MHC-I
LLFNKVTLA	S protein	821	MHC-I
HWFVTQRNF	S protein	1101	MHC-I
VLKGVKLHY	S protein	1264	MHC-I
QIGYYRRATRRIRGG	N protein	83	MHC-II
IGYYRRATRRIRGGD	N protein	84	MHC-II
DAALALLLLDRLNQL	N protein	216	MHC-II

Epitope	PR name	Start position	MHC binding
AALALLLLDRLNQLE	N protein	217	MHC-II
ALALLLLDRLNQLES	N protein	218	MHC-II
QIAQFAPSASAFFGM	N protein	303	MHC-II
AQFAPSASAFFGMSR	N protein	305	MHC-II

Antigenicity and allergenicity prediction, physicochemical properties

The final multi-epitope candidate vaccine comprises 568 amino acids and consists of three domains, including CTB as adjuvant, linear B-cell epitopes, and T-cell epitopes. Also, for protein purification and identification, a 6xHis tag was added at the C-terminal of the multi-epitope construction (Fig. 1). The antigenicity of the whole multi-epitope structure was calculated to be 0.6633. Furthermore, the results of the AllerTOP server revealed that our protein is non-allergen. Molecular weight and theoretical isoelectric point were 64.48 and 10.39 kDa, respectively. The estimated half-life was 0.8 hours for mammalian reticulocytes, 10 min in yeast, > 10 hours for *E.coli*, and 8 hours for plant. The instability index, aliphatic index, and GRAVY were 31.87, 79.89, and -0.5, respectively. According to these results, our protein is classified as stable, thermostable, and soluble protein. Connecting selected epitopes by appropriate linkers may form a signal peptide or transmembrane regions in the constructed structure. According to this possibility, the amino acid sequence of the multi-epitope candidate vaccine was checked by TMHMM v.2.0 and TargetP 5.0 online servers. Results did not show any signal-peptide or trans-membrane region in the protein construction.

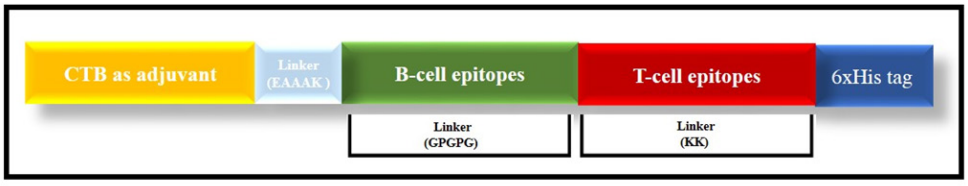

Figure1. The scheme of the final multi-epitope vaccine peptide. The 568-amino acid long peptide sequence containing adjuvant (yellow) at both N terminal was linked with the multi-epitope sequence through an EAAAK linker. HTL epitopes and B-cell epitopes are linked using GPGPG linkers (green) while the CTL epitopes are linked with KK linkers (red).

Secondary and tertiary structure modelling, refinement, and validation

Psipred analyzed secondary structure prediction. The result showed that 44.0, 16.0, and 40.0 of the total 568 amino acids were organized in alpha helix, extended strand, and random coil, respectively.

I-TASSER web server was employed to predict five tertiary 3D structures of the designed vaccine, according to ten threading templates, with Z score values (1.10–2.78) and confidence score (C-score) values (-0.75 to -4.01). Usually, the C score series is from−5 to 2, with high scores representing high sureness. All models refined by ModRefiner. The loop and energy were respectively refined and minimized to achieve the high quality of the predicted structure. The refined structures were exposed to the Ramachandran plot analysis using the RAMPAGE web server.

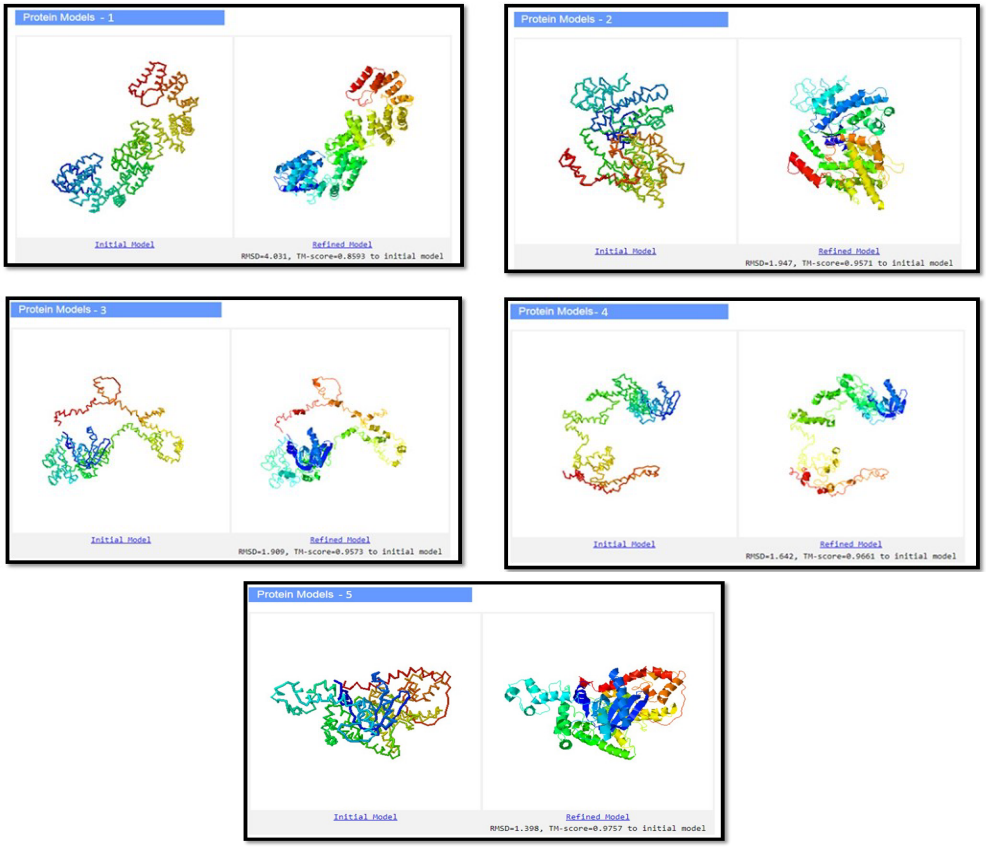

Figure 2. Protein 3D modeling and refining. I-TASSER server was used to provide the 3D model of a multi-epitope vaccine by following homology modeling (left). The protein was refined by the ModRefiner server of a refined 3D structure (right).

Between the 5 models generated by this server, model 5 (Fig. 2) was considered as the best model. For validation of the refined tertiary structures, PROSA-web, VADAR-web, and Rampage servers were utilized. PROSA-web calculated the Z-core of -7.34 for the best tertiary structure (Fig. 3a). This number indicates the tertiary structure of our multi-epitope candidate vaccine is outside the range of the scores that are determined for native proteins of similar size. The quality factor was calculated by the PROSA-web server (Fig. 3b), and the analysis of the Ramachandran plot with the Rampage server indicated that 92.38% of residues were arranged in favored regions, 6.01% of residues in additional allowed regions, 1.6% of residues in disallowed regions (Fig. 3c). This model was selected for an additional study.

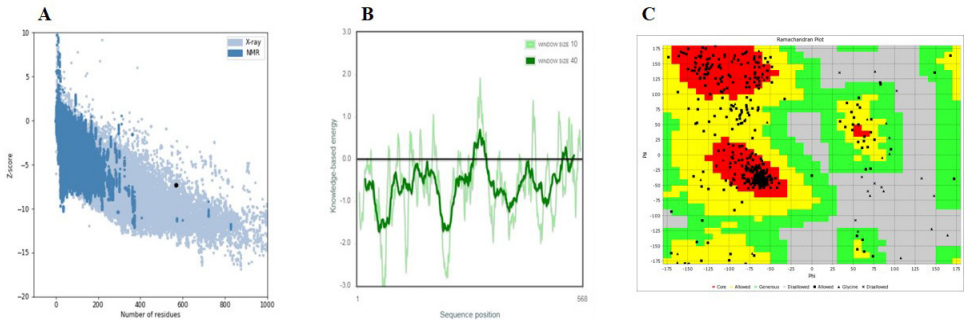

Figure 3. Protein validation. (A) ProSA-web, with a Z score of–7.34. (B) ProSA-web with quality input model graph (positive score is related to areas of the input structure that have low quality. In this model more than 90% of the input structure are in high areas). (C) Ramachandran plot analysis showing 92.38% in favored, 6.0% in allowed, and 1.6% in disallowed regions of protein residues.

Docking

Ten models were promoted for the structure of the designed candidate vaccine with TLR 4. According to the lowest energy scores (-291.56 $kJ.mol^{-1}$), model 5 was selected for the TLR4-candidate vaccine complex as the best-docked complex (Fig. 4).

Molecular dynamic simulation of the candidate vaccine-receptor complex

The deformability of each residue in the candidate vaccine-TLR4 complexes is very low. Therefore, the stability of complex structures is high. Atomic fluctuations of both complexes also have a minor deviation in the B-factor plot. Consequently, the interaction between vaccine and receptors is strong (Fig. 5).

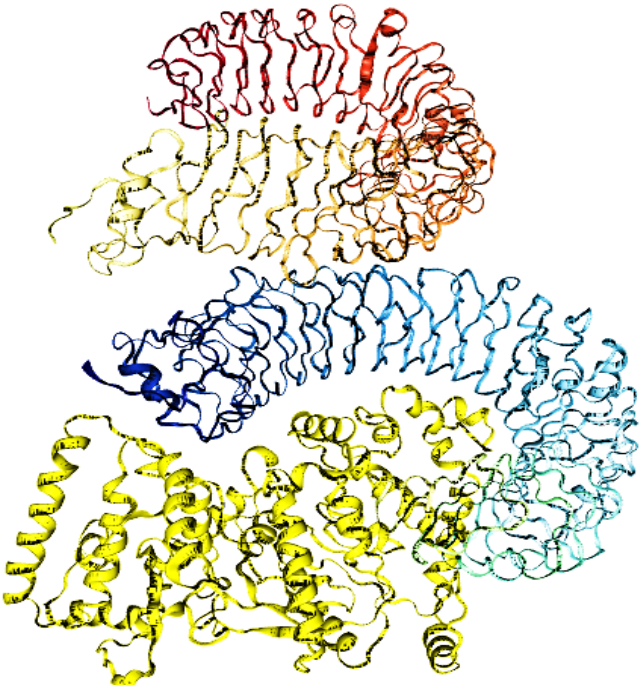

Figure 4. The results of best model molecular docking of vaccine construct (yellow) and TLR-4 (red and blue). The lowest energy score of this complex model is -291.56 kcal.mol^{-1}, indicating good binding affinity.

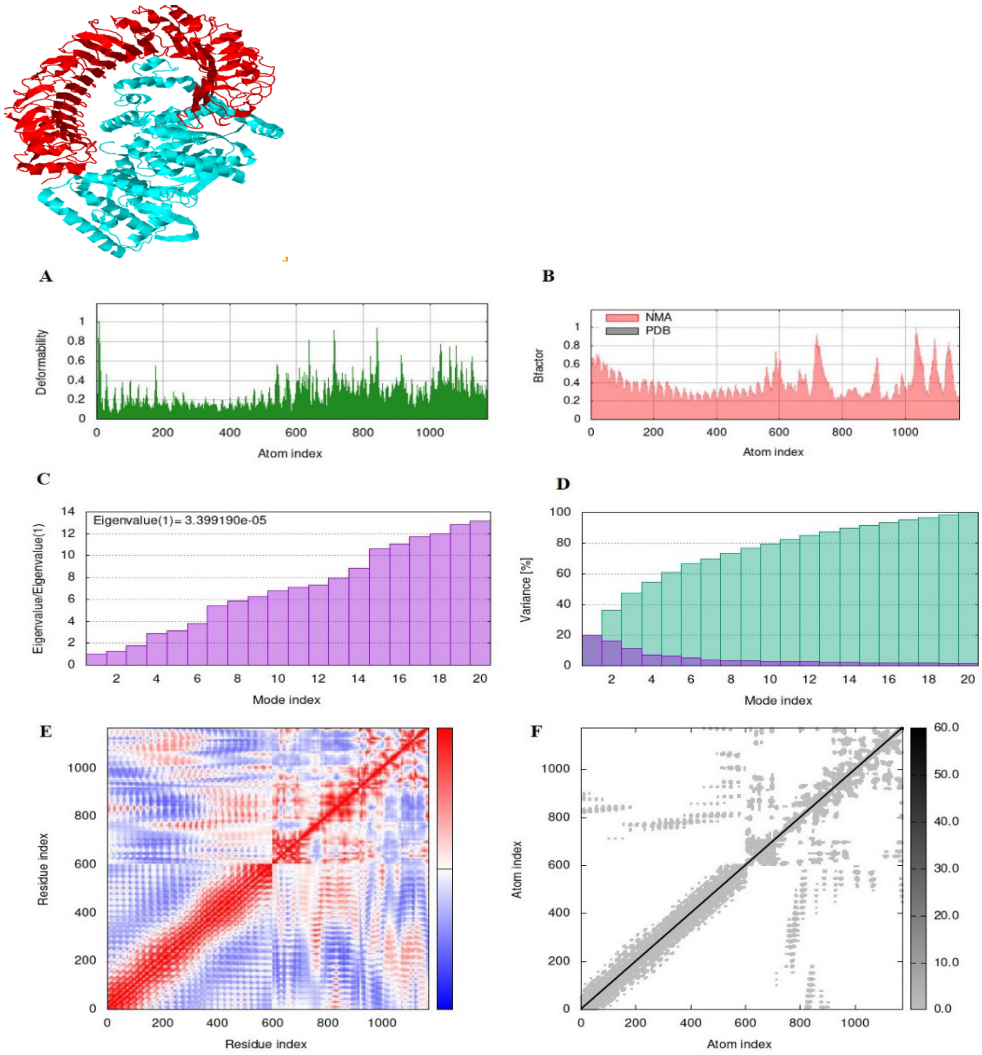

Figure 5. The results of simulating vaccine construct and TLR-4 docked complex via molecular dynamics. (A) Deformability. The deformability simulation of main-chain, the hinges as the regions with high deformability. (B) B-factor values were calculated by normal mode analysis, quantifying the uncertainty of each atom. (C) Eigenvalues; showing the energy required to deform the structure. (D) Variance. The covariance matrix between pairs of residues (white: uncorrelated, red: correlated, blue: anti-correlated). (Red color indicates individual variances and green color indicates cumulative variances), (E) co-variance map (correlated (red), uncorrelated (white) or anti-correlated (blue) motions) and (F) The elastic network model which is suggesting atom-spring connections. The springs with darker gray elastic network are more rigid.

Codon optimization of the vaccine construction and the secondary structure prediction of the mRNA

After optimization, the length of the optimized DNA sequence of the candidate vaccine was 1704 nucleotides for plant-based system. CAI of the optimized nucleotide sequence was 1.00; a number of > 0.8 is considered suitable for expression in a host, and a lower number indicates that your gene may be expressed poorly. The optimal percentage range of GC content was 31.9%. The optimal GC content should be from 30% to 70%. The 20% value of CFD was obtained for our sequence. A value of >30% for CFD may reduce the transcriptional and translational efficiency. The optimal secondary structure of mRNA was predicted with a minimum free energy of -332.97 kcal.mol^{-1}. Finally, the *SacI* and *XbaI* restriction sites were introduced to the N and C-terminals of the sequence, respectively, and this construction was inserted in the PBI121 vector (Fig. 6).

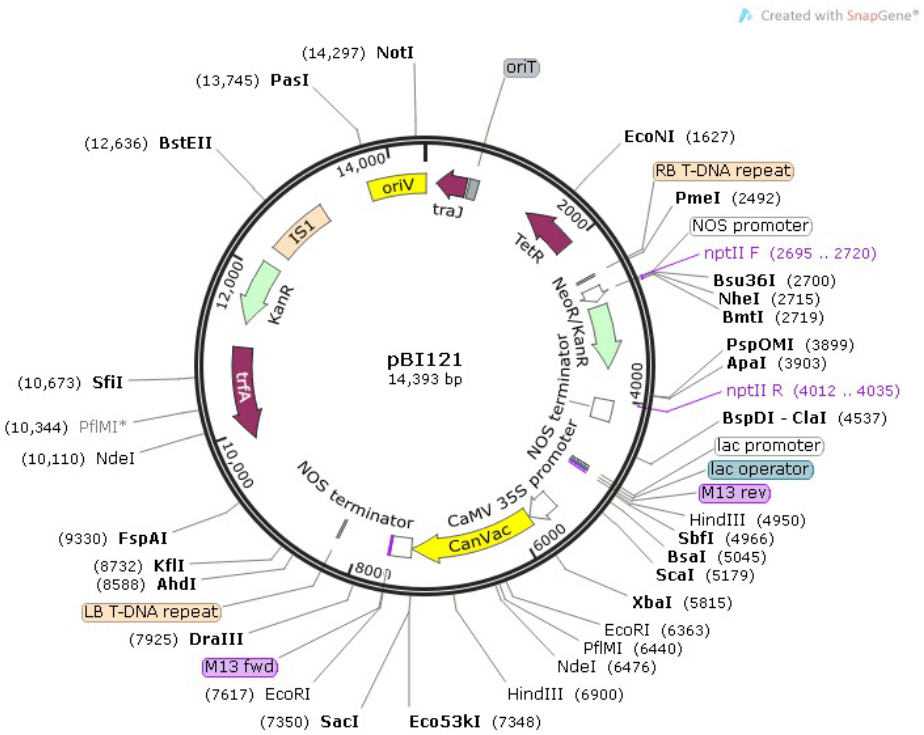

Figure 6. *in silico* cloning of the designed candidate vaccine. The yellow long region represents the codon-optimized candidate vaccine (CanVac) inserted into the pBI121 as an effective plant expression vector.

Discussion

Vaccination could be an appropriate option to manage COVID-19 prevalence (Le *et al*., 2020). Currently, two vaccines against COVID-19, produced by Moderna and Pfizer, are approved by Food and Drug Administration (FDA). Pfizer and Moderna require to store at -80 and -20, respectively (Dyer, 2020). Recent advances in computational biology and bioinformatics tools facilitate accurate identification of the immunogenic components, such as multi-epitope vaccines (Kaur *et al*., 2020). Inactive or weak pathogens are not used in multi-epitope vaccines. They are time-saving, cost-effective, and cross-protective (María *et al*., 2017).

According to the available data, SARS-CoV-2 infects cellular and humoral systems (Grifoni *et al*., 2020). T-cell response against spike protein of the SARS-CoV-2 has been significant. There is a high correlation between IgG and IgA antibody titers in patients (Shrotri *et al*., 2021; Braun *et al*., 2020; van Elslande *et al*., 2020). Besides, antibodies against N and S proteins have been detected in seropositive patients (Apasov & Sitkovsky, 2005). At the start of the infection, the virus binds to the host cell by the interaction of the ACE receptor and S protein; therefore, this protein is a primary target for neutralizing antibodies (Grifoni *et al*., 2020). In this study, several computational tools were used to design the multi-epitope vaccine against SARS-CoV-2. According to the reports, the prediction of linear B-cell epitopes was performed on the sequences of S and N proteins. Some characteristics such as High epitope score, good solubility, antigenic, nontoxic, and non-allergen were considered to select suitable B-linear epitopes. On the other hand, M, E, S, and N protein sequences were used for T-cell epitope prediction. Predicted T-cell epitopes were also filtered according to high epitope score, high hydrophobicity, antigenicity, non-toxicity, and non-allergenicity. T-cells are the main components of the adaptive immune system and have a central role in cell-mediated immunity (Apasov & Sitkovsky, 2005). They control antibody responses, activation of innate immune cells, and lysis of infected cells (Suárez-Fueyo *et al*., 2019). In human and animal models, most of the T-cell epitopes presented by MHC complexes were derived from structural proteins of the coronavirus (Shah *et al*., 2020). Therefore, in this study, the prediction of CTL and HTL epitopes was performed on structural proteins.

To construct a multi-epitope vaccine, selected B-cell epitopes are adjoined together with GPGPG linkers and T-cell epitopes with KK linkers. GPGPG and KK linkers as flexible spacers help the folding of the protein and increase the accessibility of the immune system to B-cell and T-cells epitopes (Dong *et al*., 2020). Cholera toxin subunit B (CTB) was added by EAAAK linker as an adjuvant at the N-terminal of the chimeric construction. Generally, subunit vaccines alone have

weak efficiency in stimulating the immune system. Therefore, adjuvants are used in combination with this type of vaccine (Christensen, 2016). The High potential of CTB in the activation of dendritic cells, helper T-cells, and IFN-γ secretion makes it a proper choice to use as a mucosal adjuvant in vaccines (Antonio-Herrera *et al*., 2018; Wiedinger *et al*., 2017). The Fusion of CTB at the N-terminal of a chimeric vaccine creates an opportunity for CTB to form a pentameric structure, consequently binds better to GM1-gangliosids (Lichtenstein & Höcker, 2018). EAAAK linker is a rigid spacer between the CTB domain from others and also can form an alpha-helix structure (Caparco *et al*., 2022). Analysis by AllerTOP and Vaxijen servers revealed that our design candidate vaccine was non-allergen and antigenic. ProtParam analysis showed the molecular weight of the construction was 64.48 kD and instability index was 31.87, which classify designed candidate vaccine in the stable proteins. If a protein's instability index is less than 40, this protein is predicted as a stable protein (Walker, 2005). According to the results of the ProtParam analysis, the half-life of the candidate vaccine was evaluated 0.8 h (mammalian reticulocytes, *in vitro*), >10 min (yeast, *in vitro*), >10 h (*Escherichia coli*, *in vivo*), and 8 h (Plant, *in vitro*). GRAVY index of the candidate vaccine was -0.5. A negative GRAVY score indicates the polarity nature of a protein and effective interaction with water. Consequently, negative GRAVY score reflects the solubility of a protein. The aliphatic index of our construction (aliphatic index =79.89) reflected the thermostability of the protein (Annunziato, & Costantino, 2020). The structural validation was performed by ProSA-web, ERRAT and PRECHECK online servers. Z-score (-7.34), calculated by ProSA-web server, indicates the overall quality for the tertiary structure of the input protein is outside the range of scores of the native proteins with similar size, which determined their structures by NMR and X-ray crystallography experiments (Kar *et al*., 2020). Ramachandran plot developed by PRECHECK server indicated that 92.38% of residues were arranged in favored regions, 6.01% of residues in additional allowed regions, and 1.6% of residues in disallowed regions. These factors validate the quality of our protein structure.

In the body, strong binding of the vaccine products with immune receptors leads to immunological responses (Sheik Amamuddy *et al*., 2020). TLRs are conserved membrane receptors recognizing pathogen-associated molecular patterns, such as bacterial and fungal patterns, and nucleic acids. Also, TLRs play the central role in the initiation of cellular innate immune responses (Zaheer *et al*., 2020; Eisenbarth *et al*., 2019; Smith *et al*., 2019). TLR2 and TLR4 are present on the surface of the cells and are triggered by viral structural and non-structural proteins (Athari, 2019). COVID-19 and SARS have similar clinical symptoms (Huang *et al*., 2020); therefore, the mechanism of pathogenesis of SARS-CoV-2 may be similar to SARS-CoV (Zheng *et al*., 2021). Some studies on SARS-CoV-2 infection suggested that TLR2 and TLR4 have a significant role in immune

responses (van der Donk *et al*., 2022). For example, Yao *et al*. (2022) indicated that in the infection of SARS-CoV-2, after 24 hours, the expression of TLR4 is upregulated in monocytes. Khan *et al*. (2021) demonstrated that mice with deficient TLR4 are more susceptible to SARS-CoV-2 infection than wild-type mice.

In our study, the binding affinity (ΔG value) of the docked complex was -332.97 kcal.mol^{-1}. A negative sign in ΔG value indicates interaction is thermostatically possible and it can happen in nature (Kakkanas *et al*., 2022).

A T-cell epitope can stimulate the cellular immune system when represented on the cell surface and detected by TCRs of T-cells (Siebenmorgen & Zacharias, 2020). MHC molecules bind to peptide fragments derived from a pathogen and expose them to immune system components (Reynisson *et al*., 2020). Therefore, the strong binding of T-cell epitopes with MHCs can be the key feature in the stimulation of cellular immunity (Peters *et al*., 2020).

Low deviation in deformability and B-factor plots showed vaccine- receptor has acceptable stability. The Eigenvalue indicated the vaccine- TLR4 complex is stable. Finally, the codon adaption and in silico cloning were performed by the OPTIMER server and SnapGENE, respectively. The optimized DNA had a good amount of GC content (31.9%) and CAI value (1.00), which indicates that the DNA sequence has an exact amount of favorable codons; therefore, it is likely to be expressed in a plant.

Mammalian cell culture and microbial fermentation systems are extensively used to produce recombinant proteins commercially (Rahbaran *et al*., 2021). But there can be found many benefits for molecular farming, using plants to produce recombinant proteins (Mohammadhassan and Asadishad, 2023). The costs of pharmaceutical recombinant proteins, for instance, are significantly lower than other systems, because plants can widely cultivated in greenhouses and farm without any specific facilities such as bioreactors and fementors (Chung *et al*., 2022; Bhat *et al*., 2022). In contrast to bacterial systems, plants are capable of producing pharmaceutical medicinal proteins that require post-translational modifications such as glycosylation and aggregation of different subunits, because, plant cells, same as mammalian cells, have intracellular eukaryotic membrane systems, which enables the post-translational modifications (Margolin *et al*., 2020a; Ratre *et al*., 2023). At contarary to animal system which can transmits and causes zoonotic diseases (Fallahi and Mohammadhassan, 2020), there is no common disease between humans and plants (Bhat *et al*., 2022). An an example, Margolin *et al*. (2020b) could produce high level of recombinant SARS-CoV-2 spike protein, as a vaccine, in *Nicotiana benthamiana*.

The most widely used plant expression vector for introducing and expressing recombinant proteins in plants is the pBI121 vector. According to several publications, pBI121 was employed in 40% of the 180 papers on Agrobacterium-mediated transformations (Mohammadhassan *et al*., 2014).

Other factors for pBI121 adoption include its simplicity, the ability to replace the target gene instead of the GUS gene, and the suitable expression of the transferred gene via *CaMV* 35S as an effective promoter (Mohammadhassan *et al*., 2018).

Conclusions

According to the COVID-19 prevalence, the development of other vaccines against the disease is necessary to increase the production capacity of the vaccine and the diversity of vaccines. Computational approach can be employed to create an effective vaccine in lesser time. Therefore in this study, we used bioinformatics tools to design a multi-epitope vaccine against SARS-CoV-2. The designed vaccine consists of HTL, CTL, and linear B-cell epitopes of E, M, N, and S proteins of the virus. The results showed the designed vaccine was antigenic and immunogenic. Molecular docking of the three-dimensional structural model of the vaccine with TLR2/TLR4 indicated the designed vaccine could stimulate the innate immune system. Molecular docking of the HTL and CTL epitopes with their respective MHCs showed the selected epitopes had an acceptable interaction with MHCs. As a result, cellular immunity is more likely to develop. However, bioinformatics results suggest that the designed vaccine may stimulate immunity, but laboratory tests are necessary for final confirmation.

Acknowledgements. The study has been granted by Amin Techno Gene Private Virtual Lab (NGO), Tehran, Iran.

References

Annunziato, G., & Costantino, G. (2020). Antimicrobial peptides (AMPs): a patent review (2015–2020). *Exp Opin Thera Patents, 30*(12), 931-947. Doi: 10.1080/13543776.2020.1851679.

Antonio-Herrera, L., Badillo-Godinez, O., Medina-Contreras, O., Tepale-Segura, A., García-Lozano, A., Gutierrez-Xicotencatl, L. Soldevila, G., Esquivel-Guadarrama, F.R., Idoyaga, J. and Bonifaz, L.C. (2018). The nontoxic cholera B subunit is a potent adjuvant for intradermal DC-targeted vaccination. *Front Immunol, 9*, 2212-2228. Doi: 10.3389/fimmu.2018.02212.

Amanat, F. & Krammer, F. (2020). SARS-CoV-2 vaccines: status report. *Immunity, 52*(4), 583-589. Doi: 10.1016/j.immuni.2020.03.007.

Apasov, S.G., & Sitkovsky, M.V. (2005). T cell-mediated immunity. In: *Principles of Immunopharmacology*, Nijkamp FP, Parnham MJ. (ed.), *Springer*, Singapore, pp. 19-28. Doi: 10.1007/978-3-030-10811-3.

Athari, S.S. (2019). Targeting cell signaling in allergic asthma. *Signal Transduct Target Ther, 4*(1), 1-19. Doi: 10.1038/s41392-019-0079-0.

Bhat, K.A., Tariq, L., Ayaz, A., Manzoor, M., Zargar, S.M., & Shah, A.A. (2022). Molecular farming: sustainable manufacturing of vaccines, antibodies, and other therapeutic substances. In: Metabolic engineering in plants*, Springer Nature*, Singapore, pp. 239-261. Doi: 10.1007/978-981-16-7262-0_10.

Boopathi, S., Poma, A.B. & Kolandaivel, P. (2020). Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. *J Biomol Struct Dyn, 39*(9), 3409-3418. Doi: 10.1080/07391102.2020.1758788.

Braun, J., Loyal, L., Frentsch, M., Wendisch, D., Georg, P., Kurth, F., Hippenstiel, S., Dingeldey, M., Kruse, B., Fauchere, F. & Baysal, E. (2020). SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature, 587*(7833), 270-274. Doi: 10.1038/s41586-020-2598-9.

Caparco, A.A., Dautel, D.R., & Champion, J.A. (2022). Protein mediated enzyme immobilization. *Small, 18*(19), 2106425. Doi: 10.1002/smll.202106425.

Christensen, D. (2016). Vaccine adjuvants: why and how. *Hum Vaccines Immunother*, *12*(10), 2709-2711. Doi: 10.1080%2F21645515.2016.1219003.

Chung, Y.H., Church, D., Koellhoffer, E.C., Osota, E., Shukla, S., Rybicki, E.P., & Steinmetz, N. F. (2022). Integrating plant molecular farming and materials research for next-generation vaccines. *Nat Rev Mater*, *7*(5), 372-388. Doi: 10.1038/s41578-021-00399-5.

Cong, H., Gu, Q.M., Yin, H.E., Wang, J.W., Zhao, Q.L., Zhou, H.Y., Li, Y. & Zhang, J.Q. (2008). Multi-epitope DNA vaccine linked to the A2/B subunit of cholera toxin protect mice against *Toxoplasma gondii*. *Vaccine, 26*(31), 3913-3921. Doi: 10.1016/j.vaccine.2008.04.046.

Dibner, J. (2021). Fecal-oral transmission of COVID-19: could hypochlorhydria play a role? *J Med Virol*, *93*(1), 166-167. Doi: 10.1002%2Fjmv.26265.

Dong, R., Chu, Z., Yu, F. & Zha, Y. (2020). Contriving multi-epitope subunit of vaccine for COVID-19: immunoinformatics approaches. *Front Immunol, 11*, 1784-1802. Doi: 10.3389/fimmu.2020.01784.

Dyer, O. (2020). Covid-19: many poor countries will see almost no vaccine next year, aid groups warn. *BMJ Br Med J, 371*, 1-2. Doi: 10.1136/bmj.m4809.

Eisenbarth, S.C. (2019). Dendritic cell subsets in T cell programming: location dictates function. *Nat Rev Immunol, 19*(2), 89-103. Doi: 10.1038%2Fs41577-018-0088-1.

Fallahi S, & Mohammadhassan R. (2020). A review of pharmaceutical recombinant proteins and gene transformation approaches in transgenic poultry. *J Trop Life Sci, 10*(2), 163-173. Doi: 10.11594/jtls.10.02.09.

Fehr, A.R. & Perlman, S. (2015). Coronaviruses, an overview of their replication and pathogenesis, In: Coronaviruses methods in molecular biology, Maier, H., Bickerton, E. & Britton, P. (ed.), *Springer*, Singapore, pp. 1-23. Doi: 10.1007%2F978-1-4939-2438-7_1.

Grifoni, A., Weiskopf, D., Ramirez, S.I., Mateus, J., Dan, J.M., Moderbacher, C.R., Rawlings, S.A., Sutherland, A., Premkumar, L., Jadi, R.S. & Marrama, D. (2020). Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell, 181*(7), 1489-1501. Doi: 10.1016/j.cell.2020.05.015.

Hajissa, K., Zakaria, R., Suppian, R. & Mohamed, Z. (2019). Epitope-based vaccine as a universal vaccination strategy against *Toxoplasma gonidii* infection: a mini-review. *J Adv Vet Anim Res, 6*(2), 174–182. Doi: 10.5455%2Fjavar.2019.f329.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X. & Cheng, Z. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet, 395*(10223), 497-506. Doi: 10.1016/S0140-6736(20)30183-5.

Kakkanas, A., Karamichali, E., Koufogeorgou, E.I., Kotsakis, S.D., Georgopoulou, U., & Foka, P. (2022). Targeting the YXXΦ motifs of the SARS coronaviruses 1 and 2 ORF3a peptides by in silico analysis to predict novel virus—host interactions. *Biomolecules, 12*(8), 1052. Doi: 10.3390/biom12081052.

Kar, T., Narsaria, U., Basak, S., Deb, D., Castiglione, F., Mueller, D.M. & Srivastava, A.P. (2020). A candidate multi-epitope vaccine against SARS-CoV-2. *Sci Rep, 10*(1), 1-24. Doi: 10.1038/s41598-020-67749-1.

Kaur, R., Arora, N., Jamakhani, M.A., Malik, S., Kumar, P., Anjum, F., Tripathi, S., Mishra, A. & Prasad, A. (2020). Development of multi-epitope chimeric vaccine against *Taenia solium* by exploring its proteome: an in silico approach. *Expert Rev Vaccines, 19*(1), 105-114. Doi: 10.1080/14760584.2019.1711057.

Khan, S., Shafiei, M. S., Longoria, C., Schoggins, J. W., Savani, R. C., & Zaki, H. (2021). SARS-CoV-2 spike protein induces inflammation via TLR2-dependent activation of the NF-κB pathway. *Elife, 10*, e68563. Doi: 10.7554/eLife.68563.

Le, T.T., Andreadakis, Z., Kumar, A., Román, R.G., Tollefsen, S., Saville, M., & Mayhew, S. (2020). The COVID-19 vaccine development landscape. *Nat Rev Drug Discov, 19*(5), 305-306. Doi: 10.1038/d41573-020-00073-5.

Lichtenstein, B. R., & Höcker, B. (2018). Engineering an AB5 protein carrier. *Sci Rep, 8*(1), 1-11. Doi: 10.1038/s41598-018-30910-y.

Lim, Y.X., Ng, Y.L., Tam, J.P. & Liu, D.X. (2016). Human coronaviruses: a review of virus–host interactions. *Diseases, 4*(3), 26-54. Doi: 10.3390/diseases4030026.

Liu, Z. & Chen, Y.H. (2004). Design and construction of a recombinant epitope-peptide gene as a universal epitope-vaccine strategy. *J Immunol Methods, 285*(1), 93-97. Doi: 10.1016/j.jim.2003.10.018.

Margolin, E., Crispin, M., Meyers, A., Chapman, R., & Rybicki, E.P. (2020a). A roadmap for the molecular farming of viral glycoprotein vaccines: engineering glycosylation and glycosylation-directed folding. *Front Plant Sci*, *11*, 609207. Doi: 10.3389/fpls.2020.609207.

Margolin, E., Verbeek, M., Meyers, A., Chapman, R., Williamson, A.L., & Rybicki, E.P. (2020b). Calreticulin co-expression supports high level production of a recombinant SARS-CoV-2 spike mimetic in *Nicotiana benthamiana*. *BioRxiv*, 2020-06. Doi: 10.1101/2020.06.14.150458.

María, R.R., Arturo, C.J., Alicia, J.A., Paulina, M.G. & Gerardo, A.O. (2017). The impact of bioinformatics on vaccine design and development, In: Vaccines, Afrin, F., Hemeg, H. & Ozbak, H. (eds.), *IntechOpen*, London, pp. 3-6. Doi: 10.5772/intechopen.69273.

Miles, S., Portela, M., Cyrklaff, M., Ancarola, M.E., Frischknecht, F., Durán, R., Dematteis, S. & Mourglia-Ettlin, G. (2019). Combining proteomics and bioinformatics to explore novel tegumental antigens as vaccine candidates against *Echinococcus granulosus* infection. *J Cell Biochem, 120*(9), 15320-36. Doi: 10.1002/jcb.28799.

Mohammadhassan, R., & Asadishad, T. (2023). Using in silico tools to analyze the 5' untranslated regions of the alcohol dehydrogenase gene from *Arabidopsis thaliana* and Omega Sequence. *Makara J Sci*, *27*(4), 8. Doi: 10.7454/mss.v27i4.1472.

Mohammadhassan, R., Fallahi, S. & Mohammadalipour, Z. (2020). ADMET and pharmaceutical activity analysis of caffeic acid diversities by in silico tools. *Lett Appl NanoBioSci, 9*(1), 840-8. Doi: 10.33263/LIANBS91.840848.

Mohammadhassan, R., Esfahani, K. & Kashefi, B. (2018). Constructional and functional evaluation of two new plant expression vectors—pBI121 gus-6 and pBI121 5+1. *Banat J Biotechnol*, *9*(17), 60-68. Doi: 10.7904/2068–4738–IX(17)–60.

Mohammadhassan. R., Kashefi, B. & Delcheh, K.S. (2014). *Agrobacterium*-based vectors: a review. *Int J Farm All Sci, 3*(9), 1002-1008. http://ijfas.com/wp-content/uploads/2014/10/1002-1008.pdf.

Nokhostin, F., Dargahi MalAmir, M., Tutunchi, S., & Rezaeeyan, H. (2020). Evaluation of prognostic/diagnostic value of Hematological markers in the detection of inflammation in coronavirus disease: a review study. *JAdv Med Biomed Res*, *28*(128), 171-174. Doi: 10.30699/jambs.28.128.171.

Peters, B., Nielsen, M., & Sette, A. (2020). T cell epitope predictions. *Annu Rev Immunol, 38*, 123-145. Doi: 10.1146/annurev-immunol-082119-124838.

Rahbaran, M., Razeghian, E., Maashi, M.S., Jalil, A.T., Widjaja, G., Thangavelu, L., & Jarahian, M. (2021). Cloning and embryo splitting in mammalians: brief history, methods, and achievements. *Stem Cells Int*, *2021*, 1-11. Doi: 10.1155/2021/2347506.

Rahbaran, M. (2021). Pregnancy and Covid-19: a brief review. 1st National Conference in Health, Social sciences and Humanities focusing on COVID19. *Tbilisi*, Georgia. https://b2n.ir/360005637.

Ratre, Y.K., Mehta, A., Shinde, S., Sinha, V., Soni, V.K., Sonkar, S.C., & Vishvakarma, N.K. (2023). Molecular farming and anticancer vaccine: current opportunities and openings. In*:* Microbial bioreactors for industrial molecules, Singh S.P. & Upadhyay, S.K. (eds), *Wiley,* Hoboken, pp. 355-373. Doi: 10.1002/9781119874096.ch17.

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nuc Acids Res, 48*(W1), W449-W454. Doi: 10.1093%2Fnar%2Fgkaa379.

Shah, V.K., Firmal, P., Alam, A., Ganguly, D. & Chattopadhyay, S. (2020). Overview of immune response during SARS-CoV-2 infection: lessons from the past. *Front Immunol, 11*, 1949-66. Doi: 10.3389/fimmu.2020.01949.

Shrotri, M., van Schalkwyk, M.C., Post, N., Eddy, D., Huntley, C., Leeman, D., Rigby, S., Williams, S.V., Bermingham, W.H., Kellam, P. & Maher, J. (2021). T cell response to SARS-CoV-2 infection in humans: a systematic review. *PloS one, 16*(1), e0245532. Doi: 10.1371/journal.pone.0245532.

Sheik Amamuddy, O., Veldman, W., Manyumwa, C., Khairallah, A., Agajanian, S., Oluyemi, O., Verkhivker, G.M. & Tastan Bishop, Ö. (2020). Integrated computational approaches and tools for allosteric drug discovery. *Int J Mol Sci, 21*(3), 847. Doi: 10.3390/ijms21030847.

Shey, R.A., Ghogomu, S.M., Esoh, K.K., Nebangwa, N.D., Shintouo, C.M., Nongley, N.F., Asa, B.F., Ngale, F.N., Vanhamme, L. & Souopgui, J. (2019). In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Sci Rep, 9*(1), 4409. Doi: 10.1038/s41598-019-40833-x.

Siebenmorgen, T., & Zacharias, M. (2020). Computational prediction of protein–protein binding affinities. *Wiley Interdiscip Rev: Computat Mol Sci, 10*(3), e1448.

Smith, K.G., Kamdar, A.A. & Stark, J.M. (2019). Lung defenses: intrinsic, innate, and adaptive. In: Kendig's disorders of the respiratory tract in children, Wilmott, R.W., Li, A., Sly, P., Bush, A., Deterding, R. & Ratjen, F. (eds.), *Elsevier*, Amsterdam, pp. 120-133. Doi: 10.1016/C2015-0-01292-8.

Suárez-Fueyo, A., Crispín, J.C. & Tsokos, G.C. (2019). T cells. In: Dubois' lupus erythematosus and related syndromes, Wallace, D.J. & Hahn, B.H. (eds.), *Elsevier*, Amsterdam, pp. 116-124. Doi: 10.1016/C2010-0-66018-4.

van der Donk, L.E., Eder, J., van Hamme, J.L., Brouwer, P.J., Brinkkemper, M., van Nuenen, A.C., van Gils, M.J., Sanders, R.W., Kootstra, N.A., Bermejo-Jambrina, M. & Geijtenbeek, T.B. (2022). SARS-CoV-2 infection activates dendritic cells via cytosolic receptors rather than extracellular TLRs. *European J Immunol, 52*(4), 646-655. Doi: 10.1002/eji.202149656.

van Elslande, J., Decru, B., Jonckheere, S., van Wijngaerden, E., Houben, E., Vandecandelaere, P., Indevuyst, C., Depypere, M., Desmet, S., André, E. & van Ranst, M. (2020). Antibody response against SARS-CoV-2 spike protein and nucleoprotein evaluated by four automated immunoassays and three ELISAs. *Clin Microbiol Infect, 26*(11), e1-7. Doi: 10.1016/j.cmi.2020.07.038.

Walker, J.M. (2005). The proteomics protocols handbook, *Springer*, New York. Doi: 10.1385/1592598900.

Wiedinger, K., Pinho, D. & Bitsaktsis, C. (2017). Utilization of cholera toxin B as a mucosal adjuvant elicits antibody-mediated protection against *S. pneumoniae* infection in mice. *Ther Adv Vaccines, 5*(1), 15-24. Doi: 10.1177/2051013617691041.

Yang, J. & Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res, 43*(W1), W174-81. Doi: 10.1093/nar/gkv342.

Yao, Y., Subedi, K., Liu, T., Khalasawi, N., Pretto-Kernahan, C.D., Wotring, J.W., Wang, J., Yin, C., Jiang, A., Fu, C. & Dimitrion, P. (2022). Surface translocation of ACE2 and TMPRSS2 upon TLR4/7/8 activation is required for SARS-CoV-2 infection in circulating monocytes. *Cell Discov, 8*(1), 89. Doi: 10.1038/s41421-022-00453-8.

Yoshimoto, FK. (2020). The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. *Protein J, 39*(3), 198-216. Doi: 10.1007/s10930-020-09901-4.

Zaheer, T., Waseem, M., Waqar, W., Dar, H.A., Shehroz, M., Naz, K., Ishaq, Z., Ahmad, T., Ullah, N., Bakhtiar, S.M. & Muhammad, S.A. (2020). Anti-COVID-19 multi-epitope vaccine designs employing global viral genome sequences. *Peer J, 8*, e9541. Doi: 10.7717/peerj.9541.

Zheng, C., Shao, W., Chen, X., Zhang, B., Wang, G., & Zhang, W. (2022). Real-world effectiveness of COVID-19 vaccines: a literature review and meta-analysis. *Int J Infect Dis, 114*, 252-260. Doi: 10.1016/j.ijid.2021.11.009.

Zheng, M., Karki, R., Williams, E.P., Yang, D., Fitzpatrick, E., Vogel, P., Jonsson, C.B. & Kanneganti, T.D. (2021). TLR2 senses the SARS-CoV-2 envelope protein to produce inflammatory cytokines. *Nat Immunol, 22*(7), 829-838. Doi: 10.1038/s41590-021-00937-x.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L. & Chen, H.D. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature, 579*(7798), 270-273. Doi: 10.1038/s41586-020-2012-7.