Immunoinformatics Vaccine Targeting S1-NTD and HA2 Against SARS-CoV-2 and Influenza

Marzieh Rezaei[1*] ![ORCID], **Somaieh Sabzali**[2] ![ORCID], **Mohammad Satari**[3] ![ORCID], **Maryam Parhamfar**[4] ![ORCID]

1. Department of Cell and Molecular Biology & Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran.
2. Department of Biology, Faculty of Science, Lorestan University, Khorramabad, Iran.
3. Department of Biology, Faculty of Science, Malayer University, Malayer, Iran.
4. Molecular and Medicine Research Center, Arak University of Medical Science, Arak, Iran.

ABSTRACT

Background: Two of the most challenging viruses for vaccine development are SARS-CoV-2, causing the current COVID-19 pandemic, and influenza virus (H1N1), which spreads annually, causing seasonal epidemics or increasing the pandemic risk.

Materials and Methods: In this study, we analyzed the immunodominant epitope regions in fusion peptides consisting of the Spike_S1_N-terminal domain from SARS-CoV-2 in frame to hemagglutinin H2 (*HA2*) gene from influenza A virus (H1N1) and also human *IFNγ* gene by two (G4S)3 linkers. A comprehensive analysis based on immunoinformatic has been conducted on prediction servers to predict T- and B-cell epitopes. In silico cloning and expression in pET-28(+) expression vector and vaccine optimization were assessed. The overall model quality was accessed, and the docking or binding affinity of the designed vaccine to the Toll-like receptor 3 was analyzed.

Results: The efficiency of the constructed vaccine confirmed by appropriate expression of the designed vaccine candidate tested by in silico cloning in pET–28(+) vector and codon optimization might increase the production of vaccine candidate into *Escherichia coli* strain K-12.

Conclusion: In conclusion, we suggest that this fusion peptide would be an attractive design strategy for developing a bivalent vaccine against COVID-19 and influenza as promising vaccine candidates without the need to reformulate or vaccinate each year.

* Corresponding Author:
Marzieh Rezaei, PhD.
Address: *Department of Cell and Molecular Biology & Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran.*
Phone: *+98 (913) 0925435*
E-mail: *m.rezaeibio@gmail.com*

Introduction

It was probably never thought that the illness of the five patients hospitalized in Wuhan, China, would cause a terrible pandemic of the century. The disease was caused by a virus from the family of coronavirus that had previously caused SARS and MERS epidemics, but this time, the infectivity of the virus was much higher. The new coronavirus was called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Symptoms in people with COVID-19 may present with fever, fatigue, weakness, myalgia, arthralgia, sweating, whooping cough, dry cough, nausea, and diarrhea. Studies on the mechanism of pathogenesis and virus spread show that the most common way of transferring the virus is through human-to-human respiratory droplets. These droplets enter the other person's body through the ACE-2 receptors on the surface of cells and begin to multiply and cause infection [1]. The S glycoprotein of the SARS-CoV-2 virus is the most important surface protein, and it includes the receptor binding S1-subunit and the membrane fusion S2-subunit. The N-terminal domain (NTD) of the S1 subunit is known as an immunogenic peptide and a receptor binding domain that can be involved in virus attachment to host cells by recognizing specific sugar molecules, and its main function is the transition of the S protein [2].

On the other hand, the influenza virus (H1N1), first identified in 2009 and then in 2015, led to a pandemic. Between 250000 and 500000 deaths occur worldwide each year from the flu virus. The genome of this virus is made up of 8 pieces of RNA. Eighteen serotypes of HA and 11 serotypes of neuraminidase have been identified, the most famous of which is H1N1 [3]. According to the continuous antigenic drift and the high variable region of H1N1, the influenza infection rate is increasing yearly. Emerging new strains due to high mutation rates emphasize the need to develop a broadly protective vaccine against high spreading H1N1 each year [3]. The antigenic variation mechanism of the H1N1 to evade the human immune system because of continuous re-assortment in its segmented genome can lead to recurrent infections in humans with high morbidity and mortality rates and increase the potential pandemic risk. Annual influenza infections and epidemics caused by H1N1 are due to high rates of mutations in the hemagglutinin (HA) and neuraminidase [4]. HA is one of the two essential surface glycoproteins of the H1N1 [4]. Because of different mutation changes considered antigenic drift in the HA1 domain of the HA protein, which led to the emergence of new virus strains [5], the conserved regions of HA need to be targeted to design a vaccine based on the HA2 domain of the H1N1 [6].

In previous studies, a set of different platforms was conducted to develop a vaccine against influenza based on the variable domain of HA protein [7, 8]. So, one of the purposes of this paper was to introduce a strategy for vaccine designing based on the constant domain of the major surface glycoprotein (HA2) to confer protection against H1N1 without the need for the reformulation of influenza vaccines due to high variable new emerging strains in each year [9]. Recently, IFN-γ has been licensed for human vaccines as an adjuvant and as a proper antiviral cytokine in this construct to improve the effective immune response. Novel approaches in vaccine development have been designed to fuse immunogenic peptides with immunostimulatory cytokines as a self-adjuvant [10]. So, we selected IFN-γ as an adjuvant and antiviral cytokine fused to S1-NTD and HA2 peptides to propose an attractive candidate vaccine to protect against SARS-CoV-2 and influenza viruses simultaneously. We introduced the fusion peptide to include the S1-NTD of SARS-CoV2, HA2 of H1N1, and IFNγ as an adjuvant, along with appropriate linkers. Glycine-rich linkers, such as (G4S)3, are preferred to link the peptides as they improve overall solubility and flexibility between fused peptides.

Materials and Methods

A comprehensive bioinformatics analysis was conducted using prediction software to predict T- and B-cell epitopes, the secondary and tertiary structures, and the antigenicity of fusion protein.

The protein sequence retrieval and designing the construct (S1-NTD-HA2-IFNy)

The full-length open reading frame of the Spike_S1_NTD gene (YP_009724390) from SARS-CoV-2 without its signal peptide sequence and the TAA stop codon was fused in frame to *HA2* gene (YP_009118626.1) from influenza A virus/California/07/2009 (H1N1) and also human *IFNγ* gene (NM_000619) by two (G4S)3 linkers. The 3 end of the fusion gene includes a six-histidine tag for easier purification.

B-cell and T-cell epitope prediction and confirmation of antigenicity

Various freely accessible online servers like ABCpred [11], BepiPred [12], BCPred [13], and IEDB (Immune Epitope Database and Analysis Resource) [14] were used to predict many B-cell epitopes of the fusion protein. The online prediction tools IEDB [15], SYFPEITHI [16], PropredI [17], Propred [18], Net MHC [19], Net Cytotoxic T lymphocyte (CTL) [20], and CTLPred [21]

were assessed for their ability to predict the T-cell epitopes. All predicted epitopes were detected as probable antigens with VaxiJen software, version 2.0 [22] value (threshold 0.4%).

The finally predicted epitopes should be revealed by Immunoproteasome analysis for some dominant enzymes of host cells to prevent peptide degradation during antigen processing. So, the Protein Digest server [23] was used to predict enzymatic degradation sites.

Prediction of the secondary structure of the fusion peptides, the transmembrane helix, and the signal peptide

The secondary structure of the fusion peptides was analyzed using the improved self-optimized prediction method with alignment (SOPMA) software [24]. This analysis was performed based on the conformational states of the fusion peptides (alpha-helices, beta sheets turns, and coils).

In addition, the TMHMM (Transmembrane Helices Hidden Markov Models Server v.2.0) was also used to predict the surface exposed epitopes. The transmembrane helix (TMH) of the construct was predicted using the online TMHMM server v2.0 [11], and the potential signal peptide (SigP) cleavage site was identified by The SignalP software, version 4.1 (D-cutoff: 0.45) online tool [25, 26].

Physicochemical properties and efficiency of the vaccine construct

The post-injection behavior of the designed vaccine into the body is the primary goal of vaccination. Therefore, the physicochemical features of the formulated vaccine candidate should be analyzed. Therefore, we used the ProtParam tool of ExPasy web server 33 [27]. In this web server, various parameters were computed: Molecular weight (kDa), theoretical isoelectric point (pI), in vitro and in vivo based estimated half-life, stability index, aliphatic index, extinction coefficient, and grand average of hydropathicity (GRAVY).

Three-dimensional structure modelling

The 3-dimensional modeled structure of the final vaccine construct generated by I-TASSER software. The modeled structure was refined with the online server Galaxy refine tool [28]. The overall model quality was checked using the ProSA web server.

The stereochemistry quality in the vaccine model was analyzed based on the Ramachandran plot. To validate the 3D model, the Rampage [13] server was used based on the allowed and disallowed regions of the protein structural model.

SiteSeer search

The SiteSeer search was employed to find our designed structure, scan the target structure against a prepared templates database, and match functionally important sites [29, 30]. This process raises the possibility of reporting possible matches after scanning 400 auto-generated templates from the query structure against representative structures in the PDB.

The BLAST search of our protein sequence vs UniProt

To find a sequence similarity between our protein sequence and the found sequence (hit), the BLAST search (Basic Local Alignment Search Tool) vs UniProt was done [31].

Nest analysis

Nests are structural motifs that are often found in functionally important regions of protein structure formed by consecutive enantiomeric left-handed (L) and right-handed (R) helical conformation of the backbone [32]. Simple nests are either RL or LR. Larger nests (>2 residues long) may be RLR, LRL, RLRL, etc. composed of simple overlapping nests that have not been studied despite their extensive involvement in protein function. The most abundant doublets and triplets in nests have a propensity for particular secondary structures, suggesting a strong sequence-structure relationship in the larger nest [33]. ProFunc server was used for predicting probable protein function from 3D structure.

Docking and binding affinity analysis of vaccine candidate with TLR3

To predict the binding affinity of our vaccine candidate and dissociation constant (Kd), the Gibbs free energy (ΔG) as a critical thermodynamic parameter was analyzed by using the PRODIGY web server [34]. The CASTp web server determined the active binding pockets of refined vaccine construct for TLR3 receptor [35].

Using new molecular methods aside from clinical trials can improve results. Due to the genome organization of the SARS-CoV-2 virus has been identified, the important enzymatic and structural proteins of this virus have been recognized. Virtual

methods such as molecular docking can be used to identify effective viral therapies. Almost two-thirds of the viral genome was shown to be translated into pp1a and pp1ab polyproteins that cleaved and processed into nonstructural proteins (16 proteins). The crystal structure of the designed vaccine and human TLR3 protein was performed from the Protein Data Bank website. Molecular docking investigations were performed using Molegro Virtual Docker 7 to analyze the probability of interaction with this vaccine against TLR3.

Reverse translation, codon adaptation index (CAI), and in silico expression of vaccine candidate

The codon adaptation plays an essential role in the expression of the desired foreign gene in a different host and is used to adapt the codon usage to most sequenced prokaryotic organisms and selected eukaryotic organisms. The CAI values were calculated using an algorithm from Carbone et al. [35]. An in silico cloning and expression of the vaccine candidate in *Escherichia coli* pET–28(+) vector was performed using the SnapGene 4.2 tool to verify the maximum expression of the vaccine in expression vector at XhoI and NotI restriction sites.

The conservation level and cross-protection of designed vaccine

Due to the emergence of a new strain of SARS-CoV-2 and the current challenge to control new epidemiological situations derived from its new strains, we analyzed the value of similarity and per identifies of our designed vaccine and surface glycoprotein of SARS-CoV-2 deposited in NCBI.

Results

Sequence retrieval

The construct's protein sequence retrieval and designing (S1-NTD-HA2-IFNγ) was done. The ExPASy translation result confirmed that the full-length open reading frame of the *Spike_S1_NTD* gene from SARS-CoV-2 was fused in frame to *HA2* gene from influenza A virus/California/07/2009 (H1N1) and also the Human *IFNγ* gene by two (G4S)3 linker.

B- and T-cell epitopes prediction result and final confirmation of the antigenicity

Many MHC I and MHC II epitopes were predicted to be probable by Net MHC, SYFPEITHI, and ProPred and ProPred-I online servers. The T-cell and B-cell epitopes prediction description Link is shown in Table S1, and the final results are shown in Table 1.

CTLPred and NetCTL 1.2 servers [36] were used to predict human CTL epitopes and rationalize vaccine design reliably. The prediction results by CTLPred and NetCTL 1.2 servers are shown in Table S2 and Table S3, respectively. In addition, the enzymatic degradation of the final B- and T-cell predicted epitope was analyzed. The proposed epitopes in this study have no proteasomal digestion sites for most cell dominant enzymes (Table 2). Protein digest analysis of final B- and T-cell epitopes was reported as undigested enzymes with mass (Da) and pI of each epitope in Table 2.

Finally, two identified epitopes were confirmed by the VaxiJen v2.0 server as probable antigens (keeping a score threshold of 0.4). The first probable antigenic epitope was amino acid residues from 383-392 (KRIEN-LNKK), and another epitope was predicted from 385-394 residues (IENLNKKVD) (Table 1). The VaxiJen v2.0 server confirmed the final antigenicity of the fusion peptides with a score of 0.5107. A score >0.4 is considered antigenic and indicates it is a probable antigen to activate immunity against SARS-CoV-2 and influenza. The position of the final predicted epitope with a high score in the vaccine designed and verified by the VaxiJen v2.0 server is shown in Figure 1.

Analysis of physicochemical properties and efficiency of the vaccine candidate

To understand the vaccine response and stability, the vaccine construct's physiochemical properties were analyzed using the online tool ProtParam. The designed construct of 697 amino acid residues with a molecular weight of 78.23 kDa was also found to fit within the defined range (40–50 kDa) of average molecular weight of vaccine based on recombinant protein [37]. Some essential properties of the protein can affect its antigenic nature. So, we assessed some physiochemical properties (e.g. the theoretical pI, the grand average hydropathicity [GRAVY] index, the instability index, etc.). The theoretical pI of the vaccine was found to be 8.04. The aliphatic index of the vaccine was 73.69, which suggests that the vaccine would be thermostable. The higher aliphatic index of a protein can be indicated as greater thermostability [36, 37]. The determined half-life of the vaccine, as predicted by ExPASy, is 30 h in mammalian reticulocytes, in vitro; >20 h in yeast, in vivo; and >10 h in *E. coli*, in vivo. The GRAVY score is -0.421 (the lower the GRAVY score, the greater the solubility), which indicates the hydrophilic nature of the vaccine candidate, meaning the probably appropriate interaction with the aqueous environment. The instability index of the vaccine candidate was 36.68, which indicates it is a stable protein. Generally, a protein whose instability index is <40 is classified as a stable protein [38].
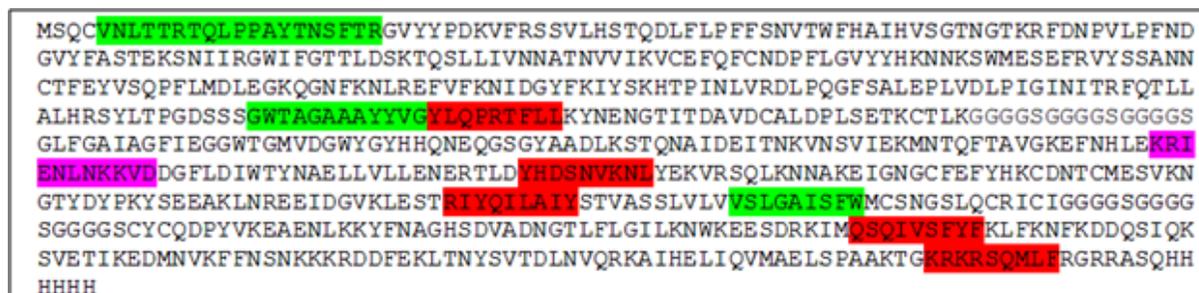
MSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFND
GVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANN
CTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLL
ALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKGGGGSGGGGSGGGGS
GLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSGYAADLKSTQNAIDEITNKVNSVIEKMNTQFTAVGKEFNHLEKRI
ENLNKKVDDGFLDIWTYNAELLVLLENERTLDYHDSNVKNLYEKVRSQLKNNAKEIGNGCFEFYHKCDNTCMESVKN
GTYDYPKYSEEAKLNREEIDGVKLESTRIYQILAIYSTVASSLVLVVSLGAISFWMCSNGSLQCRICIGGGGSGGGG
SGGGGSCYCQDPYVKEAENLKKYFNAGHSDVADNGTLFLGILKNWKEESDRKIMQSQIVSFYFKLFKNFKDDQSIQK
SVETIKEDMNVKFFNSNKKKRDDFEKLTNYSVTDLNVQRKAIHELIQVMAELSPAAKTGKRKRSQMLFRGRRASQHH
HHHH

Figure 1. The position of the final predicted epitope with high score by VaxiJen software, version 2.0

Analyzing the secondary structure

Analyzing the secondary structure by the SOPMA method and TMHMM servers showed that the random coils were detected at 32.14% (Figure 2). The higher presence of predicted random coil shows a higher preference for binding to an antibody. An increased number of secondary structures, e.g. extended strands and random coils in the protein indicate more antigenicity of proposed epitopes [39] and illustrate predicted relative surface accessibility. To predict the surface exposed epitopes, the TMHMM server software, version 2.0 was used, and the graphical output of TMHMM showed that the amino acid residues from 1 to 494 are outside and would be probability considered as surface exposed sequences.

Structure-based assessments of the vaccine construct

Structure-based assessments of the vaccine construct were obtained after molecular modeling. The 3-dimensional modeled structure for the fusion peptides was generated by I-TASSER software [40]. Five models of the 3-dimensional structure were predicted. Model 1 from the output of the I-TASSER server was selected (Figure 3A). The information about modeled structure refinement after analysis with the online server Galaxy refine tool was shown in Figure 3B. Model 1 from the output of the I-TASSER server was visualized by Swiss-Pdb Viewer software. The ProSA Z-score of the model was -4.16 (Figure 3C), indicating that it is near to experimentally determined structures of similar sizes and confirms
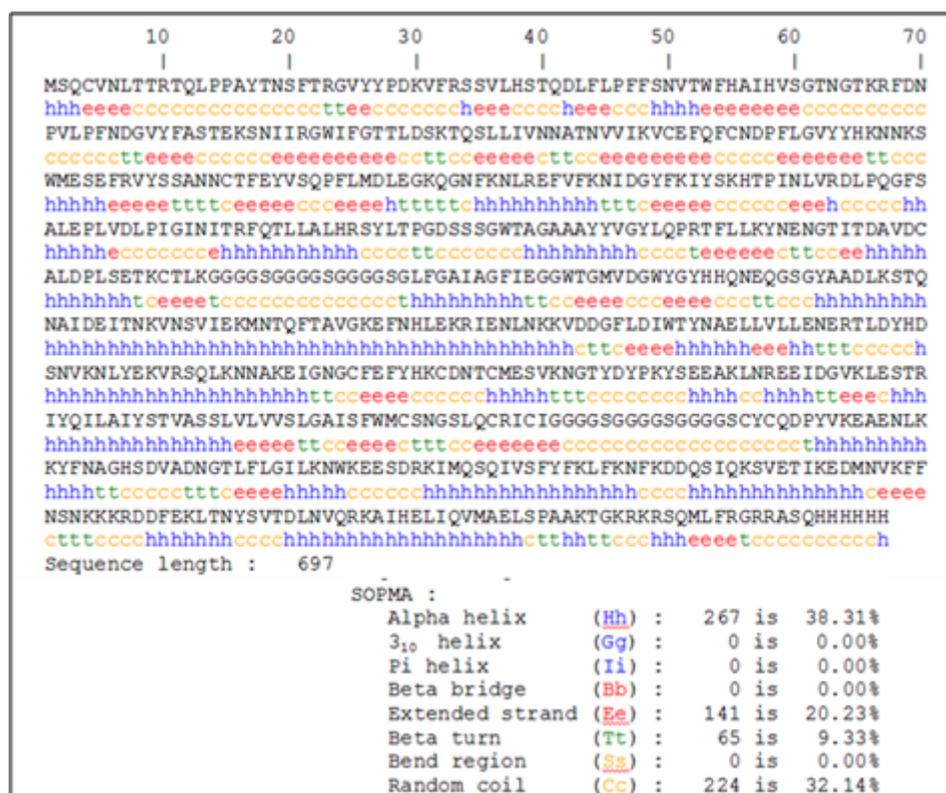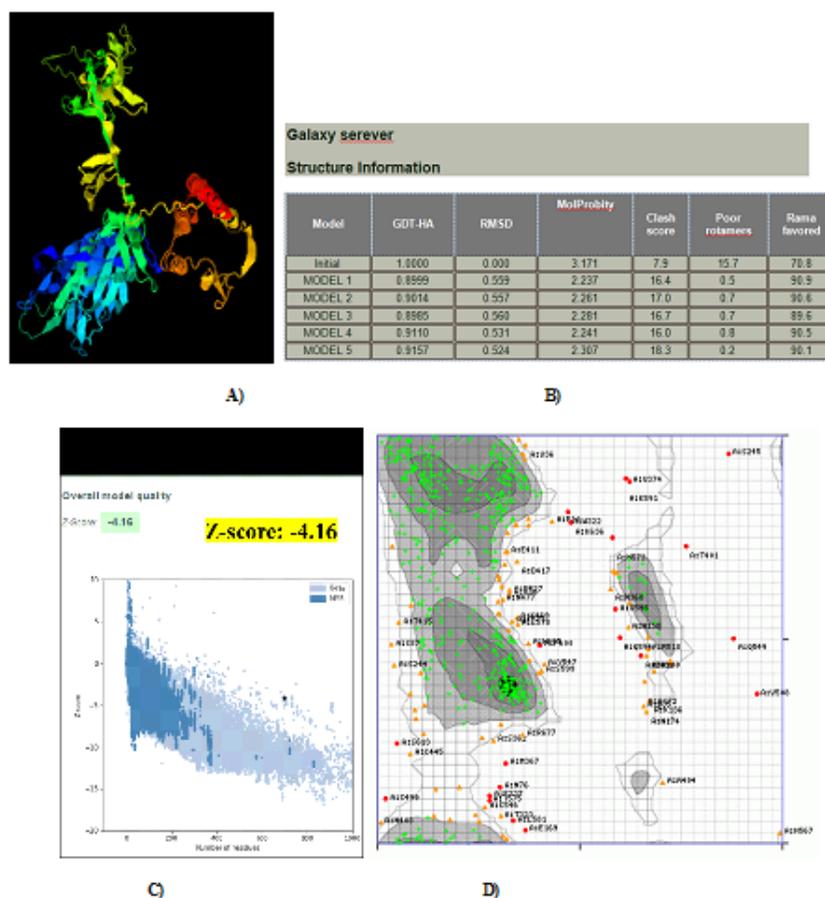

Figure 2. Analysis of the secondary structure of the fusion peptides by the SOPMA

Figure 3. The 3-dimensional modeled structure for the fusion peptides generated by I-TASSER software

the near-native configuration. The analysis of the Ramachandran plot predicted 89.8% of residues present in the most favorable region, 12.56% in the permissible region, and only 3.4% of amino acids are in the far region (Figure 3D). So, it would be suggested that the structural model has high quality.

Binding affinity of vaccine candidate and dissociation constant prediction

The result of the PRODIGY web server [34] showed the negative ΔG value of our designed vaccine candidate (-6.2 ΔG kcal/mol). The negative ΔG value reveals that the molecular association between our designed candidate and TLR3 structure would be thermodynamically possible. In addition, the dissociation constant (Kd) value was found to be $2.7E^{-05}$ and estimated at 25 °C.

Table 1. Final identified B-cell and T-cell epitopes

Final Identified B –Cell Epitopes	Position	Sequence	VaxiJen Score
Final identified B-cell epitopes	246-262	GWTAGAAAYYVGYLQP	Probable ANTIGEN
	509-517	VSLGAISFW	Probable ANTIGEN
	5-23	VNLTTRTQLPPAYTNSFTR	Probable ANTIGEN
	247-256	WTAGAAAYY	Probable ANTIGEN
Final identified T-cell epitopes	383-392	KRIENLNKK	Probable ANTIGEN
	385-394	IENLNKKVD	Probable ANTIGEN
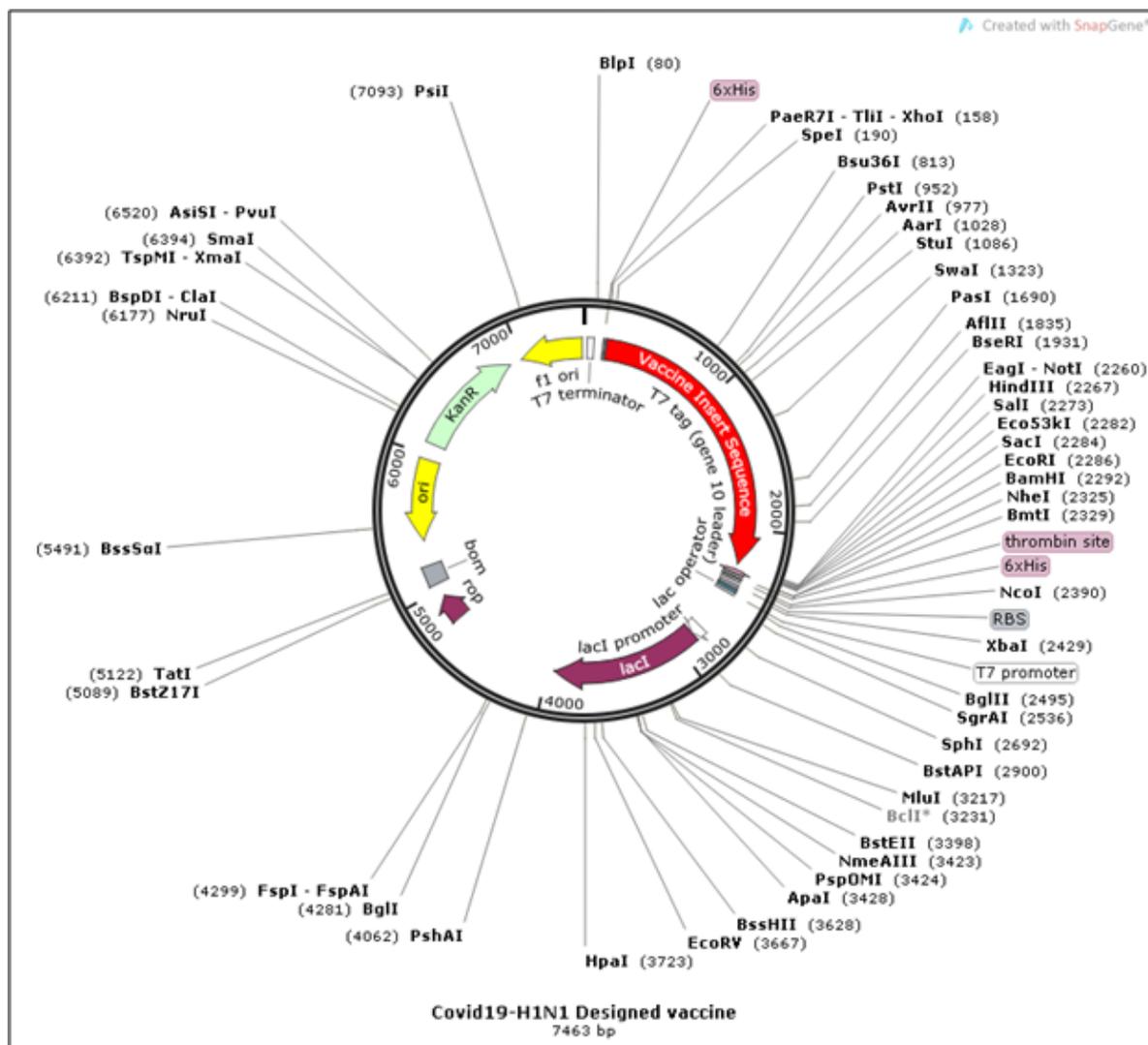
Figure 4. In silico cloning of vaccine

Note: The segment represented in red is the designed vaccine insert in the pET-28(+) expression vector.

Reverse translation, CAI of vaccine candidate

The codon usage of the designed vaccine candidate to the *E. coli* K-12 strain as a sequenced prokaryotic organism was obtained after reverse translation, followed by a codon adaptation tool. It was revealed that the CAI value of the improved sequence was 0.833873. As this value is >0.8, it was found to be convenient [41]. The GC content of the improved sequence was computed and found to be 53.3333. GC-content of *E. coli* (strain K-12) is 50.7340. As we know, GC contents should be within 30%-70%, and optimal results were achieved [42]. Finally, in silico cloning of the vaccine construct gene in pET–28(+) expression vector was performed at XhoI and NotI restriction sites and succeeded by its virtual confirmation (Figure 4).

SiteSeer search

The SiteSeer search predicted the possible functional relationship between our new structure and existing PDB entries, as presented in Table 3. The top 20 hits are listed as certain matches (E value<1.00E^{-06}) [31]. These top 20 hits are most likely to have been preserved and hence give the highest local similarity scores. The similarity score gives the similarity between the neighborhood around the matched side chains in the query structure and the neighborhood around the template side chains in their parent structure. The scores being between 250 and 480 place the hits in the possible matches [31]. The PDB code is hyperlinked to its PDB sum entry. Where the same structure has been hit more than once, only the top-scoring hit is shown here. The number of hits, n, is shown as "×n."
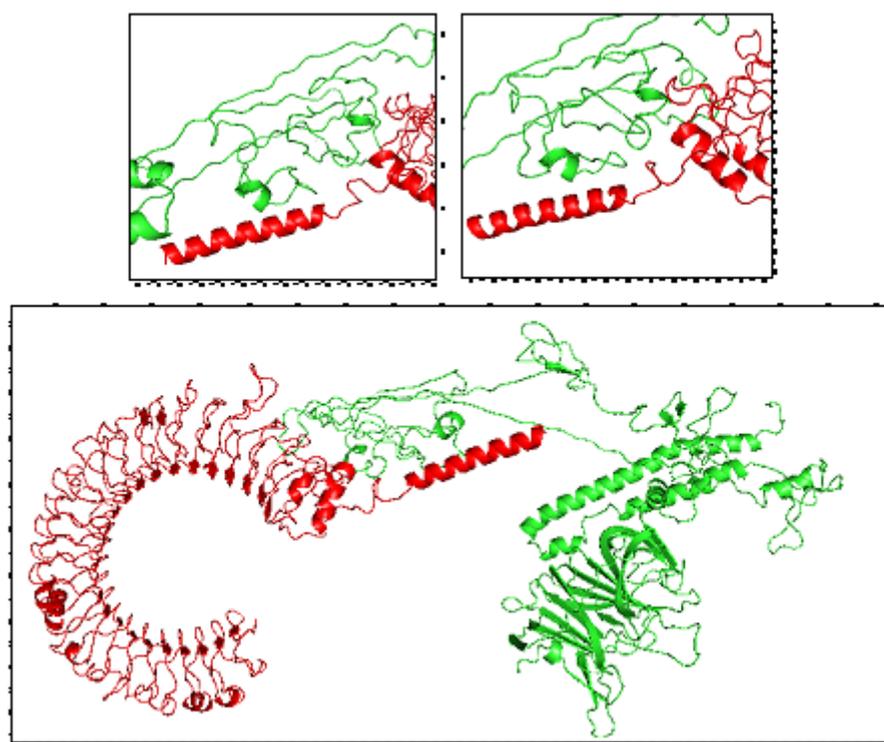
Figure 5. Interaction of human Toll-like receptor (TLR3) protein (red) and model protein (green)

Results of BLAST search vs UniProt

BLAST search vs UniProt results showed the best alignment between the query and found sequences (hits). The BLAST high score and E value of zero indicated the higher quality of alignment and would be helpful in best identifying, characterizing, and comparative analysis of our designed fusion protein.

Nest analysis results

Seven nests were identified. The nest score indicates how functionally significant the nest is likely to be. A score above 2.0 is suggestive of the nest being a functionally significant one [43]. One of the important factors determining how proteins interact with other molecules is the size of clefts in the protein's surface [43]. The depth in clefts indicates how they are functionally important. The

Table 2. Protein digest analysis of final T-cell and B- cell epitopes

Protein Digest Analysis	Position	Epitope	Mass (Da)	pI	Undigested Enzyme
Protein digest analysis of final B-cell epitopes	246-262	GWTAGAAAYYVGYLQP	1687.87	5.52	Trypsin, clostripain, cyanogen_bromide, proline endopeptidase, staph_protease, trypsin_K, trypsin_r, AspN
	509-517	VSLGAISFW	979.14	5.49	Trypsin, clostripain, cyanogen_bromide, iodosobenzoate, proline endopeptidase, staph_protease, trypsin_K, trypsin_r, AspN,
	5-23	VNLTTRTQLPPAYTNS-FTR	2180.45	10.83	Cyanogen_bromide, iodosobenzoate, staph_protease, trypsin_K, AspN
	247-256	WTAGAAAYY	973.05	5.52	Trypsin, clostripain, cyanogen_bromide, proline endopeptidase, staph_protease, trypsin_K, trypsin_R, AspN,
Protein digest analysis of final T-cell epitopes	383-392	KRIENLNKK	1142.37	10.29	Chymotrypsin, cyanogen_bromide, iodosobenzoate, proline endopeptidase, AspN,
	385-394	IENLNKKVD	1072.23	6.19	Chymotrypsin, clostripain, cyanogen_bromide, iodosobenzoate, proline endopeptidase, trypsin_R,

Table 3. The SiteSeer search predicted result

Hit No.	E Value	Similarity Score	Neighbors Identification/Similarity [Equivalent]"	Template ID	Matched PDB Entry	Longest Fitted Segment	Seq Lengths Query/Target	Overlap	%-tag -seq Id	Structural Similarity (%)
1	0.00E+00	870.00	47/0 [46]	TMP00092	6vxx×38: Structure of the SARS-CoV-2 spike glycoprotein (closed state)	241/341	697/972	960	42.47	86.3
2	0.00E+00	850.00	46/0 [45]	TMP00395	6x29×72: SARS-CoV-2 rS2d down state spike protein trimer	241/328	697/972	951	41.46	86.6
3	0.00E+00	817.00	45/0 [44]	TMP00395	6vsb×36: Prefusion 2019-nCoV spike glycoprotein with a single receptor-binding domain up	238/393	697/973	973	42.18	86.7
4	0.00E+00	734.31	36/8 [43]	TMP00395	6acc×34: Trypsin-cleaved and low pH-treated SARS-CoV spike glycoprotein and ACE2 complex, ACE-free conformation with three RBD in down conformation	153/355	697/1065	1027	32.14	87.4
5	0.00E+00	730.47	35/8 [40]	TMP00395	5x4s×76: Structure of the NTD of SARS-CoV spike protein	150/256	697/269	293	54.28	99.6
6	0.00E+00	725.94	36/7 [42]	TMP00395	6crv×70: SARS spike glycoprotein, stabilized variant, c3 symmetry	150/454	697/881	938	28.26	91.7
7	0.00E+00	613.12	32/4 [19]	TMP00092	6nb6×22: SARS-CoV complex with human neutralizing s230 antibody Fab fragment (state 1)	254/374	697/1026	1065	32.71	84.2
8	1.30E-08	409.50	17/10 [27]	TMP00132	6q04×8: MERS-CoV s structure in complex with 5-n-acetyl neuraminic acid	61/173	697/1159	1152	20.23	75.8
9	3.49E-08	396.88	16/11 [27]	TMP00132	6j11×6: MERS-CoV spike N-terminal domain and 7D10 scFv complex	56/206	697/336	356	18.75	94.3
10	3.49E-08	396.88	16/11 [27]	TMP00132	5x59×7: Prefusion structure of MERS-CoV spike glycoprotein, three-fold symmetry	55/285	697/1141	1106	22.67	80.0
11	1.27E-07	372.41	16/8 [24]	TMP00342	5w9i×14: MERS s ectodomain trimer in complex with variable domain of neutralizing antibody g4	58/213	697/513	773	21.44	74.4
12	2.80E-07	365.62	17/8 [24]	TMP00332	6nzk×24: Structural basis for human coronavirus attachment to sialic acid receptors	109/283	697/1175	1191	22.53	80.0

large clefts are shown in Table S4. In addition, residue conservation shows the conservation score for each nest residue, as determined from a multiple sequence alignment of the protein's sequence against BLAST hits from the UniProt sequence database. The conservation score ranges from 0.0, signifying that the residue is not at all conserved, to 1.0, which indicates it is perfectly conserved.

Docking result

The human TLR3 protein has a critical role in controlling and regulating host immune response. The TLR3 protein is susceptible to viral infection and other inflammatory responses. Figure 5 shows the schematic structure of the interaction of chain A of human TLR3 protein (red) with model protein (green). The PyMOL software package found that human TLR3 and model

proteins interact through van der Waals interactions and hydrogen bonding. The root mean square deviation (RMSD) from the overall lowest-energy structure is equal to (RMSD=32.8±0.4 Å). Total van der Waals energy equals -88.8±8.4 kcal/mol, and total electrostatic energy equals -233.4±25.5 kcal/mol. The desolation energy is equal to -65.5±4.5 kcal/mol. The restraints violation energy is equal to 2121.5±87.6 kcal/mol. The buried surface area is equal to 3047.8±244.5 Å². The Z-score for docking human TLR3 protein and model protein equals -1.9. As a result, a human TLR3 protein could be considered the potential binding with the designed vaccine candidate due to the lowest receptor-average Z-score. In addition, active pockets and ligand binding sites of the TLR3 receptor and refined vaccine construct were determined using the CastP web server. The CASTp web server shows an active binding pocket of the vaccine construct for the TLR3 receptor. So, it would be concluded that the determined active binding packet of the vaccine construct for the TLR3 receptor confirmed the result by docking analysis.

Assessing the conservation level and cross-protection between Coronaviruses

The similarity and per identified between our vaccine and NCBI deposited sequence of SARS-CoV-2 and influenza A (H1N1) viruses respectively. These data reflect that this designed vaccine candidate is conserved in almost all isolated SARS-CoV-2 and influenza (H1N1) viruses. This investigation also confirmed the cross-protection between coronaviruses and influenza viruses due to the high conservation level and provided some information in terms of vaccine coverage.

Discussion

The emergence of new viral strains in resource-poor countries represents a huge global disease burden [44]. There is significant interest in developing novel peptide-based infectious disease vaccines for many pathogens [45]. COVID-19 and influenza are highly contagious respiratory diseases that threaten global public health. Co-infection with SARS-CoV-2 and influenza viruses is one of the most important epidemiological challenges worldwide, and there is a serious need to develop a divalent vaccine that can control these two infectious epidemics [46].

There are different strategies for developing vaccines against infectious diseases [47, 48]. Applying any strategy depends on knowing the mechanisms of infection and the target pathogen and evaluating the host's immune protection. In addition, main factors such as cost, benefits, risk, safety, and the balance between these factors should certainly be analyzed before introducing the desired vaccine candidate to the clinical phases [28]. Peptide-based vaccines are one of the next-generation platforms of vaccines against infectious diseases [49]. Developing a new vaccine strategy based on the peptide and recombinant proteins has more advantages over other vaccine-developing approaches (e.g. the nucleic acid or viral vector vaccines) [50, 51] because of lower manufacture costs, simpler and non-complicated procedures, and fewer complications [10, 52]. Recently, experimentally analyzing all pathogen gene products has been time- and cost-consuming. So, applying the desirable, low-cost, and fast approaches is promising for designing and predicting viral or bacterial antigens. It should be noted that the experimentally determined epitopes will never be equal to predicted ones through immunoinformatics, and variable results will be anticipated [53]. The antigenicity and analysis of the two-dimensional structure of our designed vaccine candidate showed its good accessibility and effectiveness in being recognized by the B- and T-cells. The zero and negative values in ProSA plots were related to the stabilized models. Therefore, the 3D predicted models were considered structurally favorable. Since TLR3 has proven recognition capability in both SARS-CoV [54] and influenza [55], the docking analysis and CASTp server results (showing an active binding package of the vaccine structure for the TLR3 receptor) can suggest that the vaccine construct has significant affinity to TLR3 for recognizing molecular patterns of pathogen and activating immune response [31].

Conclusion

This vaccine based on fusion peptide can be designed to include multiple determinants from two SARS-CoV2 and influenza viruses. This fusion peptide was designed as a reverse vaccinology approach to target most immunogenic peptides rather than whole viruses or whole surface proteins. The prediction methods can reduce the cost of vaccine development. The final comprehensive bioinformatic analyses have proposed that the bivalent construct (NTDS1-HA2- IFN-γ) would be a computational design strategy for developing a vaccine to overcome both influenza and COVID-19 disease.

Ethical Considerations

Compliance with ethical guidelines

There were no ethical considerations to be considered in this research.

References

[1] Zhang W, Zhao Y, Zhang F, Wang Q, Li T, Liu Z, et al. The use of anti-inflammatory drugs in the treatment of people with severe coronavirus disease 2019 (COVID-19): The perspectives of clinical immunologists from China. Clin Immunol. 2020; 214:108393. [DOI:10.1016/j.clim.2020.108393] [PMID] [PMCID]

[2] Dai L, Gao GF. Viral targets for vaccines against COVID-19. Nat Rev Immunol. 2021; 21(2):73-82. [DOI:10.1038/s41577-020-00480-0] [PMID] [PMCID]

[3] Dou D, Revol R, Östbye H, Wang H, Daniels R. Influenza a virus cell entry, replication, virion assembly and movement. Front Immunol. 2018; 9:1581. [DOI:10.3389/fimmu.2018.01581] [PMID] [PMCID]

[4] Krammer F, Palese P. Universal influenza virus vaccines that target the conserved hemagglutinin stalk and conserved sites in the head domain. J Infect Dis. 2019; 219(Suppl_1):S62-7. [DOI:10.1093/infdis/jiy711] [PMID] [PMCID]

[5] Wang TT, Palese P. Biochemistry. Catching a moving target. Science. 2011; 333(6044):834-5. [DOI:10.1126/science.1210724] [PMID]

[6] Du R, Cui Q, Rong L. Flu universal vaccines: New tricks on an old virus. Virol Sin. 2021; 36(1):13-24. [DOI:10.1007/s12250-020-00283-6] [PMID] [PMCID]

[7] Song L, Xiong D, Kang X, Yang Y, Wang J, Guo Y, et al. An avian influenza A (H7N9) virus vaccine candidate based on the fusion protein of hemagglutinin globular head and Salmonella typhimurium flagellin. BMC Biotechnol. 2015; 15:79. [DOI:10.1186/s12896-015-0195-z] [PMID] [PMCID]

[8] Angeletti D, Gibbs JS, Angel M, Kosik I, Hickman HD, Frank GM, et al. Defining B cell immunodominance to viruses. Nat Immunol. 2017; 18(4):456-63. [DOI:10.1038/ni.3680] [PMID] [PMCID]

[9] Gerdil C. The annual production cycle for influenza vaccine. Vaccine. 2003; 21(16):1776-9. [DOI:10.1016/S0264-410X(03)00071-9] [PMID]

[10] Tripathi NK, Shrivastava A. Recent developments in recombinant protein-based dengue vaccines. Front Immunol. 2018; 9:1919. [DOI:10.3389/fimmu.2018.01919] [PMID] [PMCID]

[11] Miller ME. The interactive tester (Version 4.0) [Computer software]. Psytek Services; 1993. [Link]

[12] St. James J, Schneider W, Eschman A. PsychMate student guide (Version 2.0) [Computer software]. 2003. [Link]

[13] Saha S, Raghava GP. Prediction methods for B-cell epitopes. In: Flower DR, editor. Immunoinformatics. Methods in Molecular Biology™, vol 409. New Jersey: Humana Press. [Link]

[14] Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 2017. [Link]

[15] Greenbaum J. Next-Generation IEDB Tools Website and API - Release 1.1. Immune Epitope Database and Analysis Resource (IEDB); 2024. [Link]

[16] No Author. SYFPETTHI [Internet]. 2025 [Updated 5 April 2025]. Available from: [Link]

[17] Singh H, Raghava GP. ProPred1: Prediction of promiscuous MHC Class-I binding sites. Bioinformatics. 2003; 19(8):1009-14. [PMID]

[18] Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: Application to the MHC class I system. Bioinformatics. 2016; 32(4):511-7. [DOI:10.1093/bioinformatics/btv639] [PMID]

[19] Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics. 2007; 8:424. [DOI:10.1186/1471-2105-8-424] [PMID] [PMCID]

[20] Bhasin M, Raghava GP. Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine. 2004; 22(23-24):3195-204. [DOI:0.1016/j.vaccine.2004.02.005] [PMID]

[21] Doytchinova IA, Flower DR. Bioinformatic approach for identifying parasite and fungal candidate subunit vaccines. Open Vaccines J. 2008; 1:22-6. [Link]

[22] No Author. Protein Digest server. 2025. [Link]

[23] No Author. SOPMA. 2025. [Link]

[24] Hallgren J, Tsirigos KD, Pedersen MD, Armenteros JJA, Marcatili P, Nielsen H, et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv. 2022. [Link]

[25] Nielsen H. Protein function prediction. In: Kihara D, editor. Protein function prediction methods and protocols. Berlin: Springer; 2017. [Link]

[26] Wu S, Zhang Y. LOMETS: A local meta-threading-server for protein structure prediction. Nucleic Acids Res. 2007; 35(10):3375-82. [DOI:10.1093/nar/gkm251] [PMID] [PMCID]

[27] Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. The Proteomics Protocols Handbook. Humana Press; 2005. p. 571-607.

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

[29] Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, et al. Structure validation by Calpha geometry: Phi,psi and Cbeta deviation. Proteins. 2003; 50(3):437-50. [DOI:10.1002/prot.10286] [PMID]

[30] Laskowski RA, Watson JD, Thornton JM. Protein function prediction using local 3D templates. J Mol Biol. 2005; 351(3):614-26. [DOI:10.1016/j.jmb.2005.05.067] [PMID]

[31] UniProt Consortium. The universal protein resource (UniProt) in 2010. Nucleic Acids Res. 2010; 38(Database issue):D142-8. [DOI:10.1093/nar/gkp846] [PMID] [PMCID]

[32] Pal D, Sahu S, Banerjee R. New facets of larger Nest motifs in proteins. Proteins. 2020; 88(11):1413-22. [DOI:10.1002/prot.25961] [PMID]

[33] Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A. PRODIGY: A web server for predicting the binding affinity of protein-protein complexes. Bioinformatics. 2016; 32(23):3676-8. [DOI:10.1093/bioinformatics/btw514] [PMID]

[34] Al Tbeishat H. Novel in silico mRNA vaccine design exploiting proteins of M. tuberculosis that modulates host immune responses by inducing epigenetic modifications. Sci Rep. 2022; 12(1):4645. [DOI:10.1038/s41598-022-08506-4] [PMID] [PMCID]

[35] Carbone A, Zinovyev A, Képès F. Codon adaptation index as a measure of dominating codon bias. Bioinformatics. 2003; 19(16):2005-15. [DOI:10.1093/bioinformatics/btg272] [PMID]

[36] Kumar Pandey R, Ojha R, Mishra A, Kumar Prajapati V. Designing B- and T-cell multi-epitope based subunit vaccine using immunoinformatics approach to control Zika virus infection. J Cell Biochem. 2018; 119(9):7631-42. [DOI:10.1002/jcb.27110] [PMID]

[37] Walker JM. The proteomics protocols handbook. Berlin: Springer; 2005. [DOI:10.1385/1592598900]

[38] Shaddel M, Ebrahimi M, Tabandeh MR. Bioinformatics analysis of single and multi-hybrid epitopes of GRA-1, GRA-4, GRA-6 and GRA-7 proteins to improve DNA vaccine design against Toxoplasma gondii. J Parasit Dis. 2018; 42:269-76. [Link]

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

[40] Morla S, Makhija A, Kumar S. Synonymous codon usage pattern in glycoprotein gene of rabies virus. Gene. 2016; 584(1):1-6. [DOI:10.1016/j.gene.2016.02.047] [PMID]

[41] Ali M, Pandey RK, Khatoon N, Narula A, Mishra A, Prajapati VK. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. Sci Rep. 2017; 7(1):9232. [DOI:10.1038/s41598-017-09199-w] [PMID] [PMCID]

[42] Laskowski RA, Luscombe NM, Swindells MB, Thornton JM. Protein clefts in molecular recognition and function. Protein Sci. 1996; 5(12):2438-52. [DOI:10.1002/pro.5560051206] [PMID]

[43] Keni R, Alexander A, Nayak PG, Mudgal J, Nandakumar K. COVID-19: Emergence, spread, possible treatments, and global burden. Front Public Health. 2020; 8:216. [DOI:10.3389/fpubh.2020.00216] [PMID] [PMCID]

[44] Malonis RJ, Lai JR, Vergnolle O. Peptide-based vaccines: Current progress and future challenges. Chem Rev. 2020; 120(6):3210-29. [DOI:10.1021/acs.chemrev.9b00472] [PMID] [PMCID]

[45] Ao Z, Jing Ouyang M, Olukitibi TA, Warner B, Vendramelli R, Truong T, et al. Development and characterization of recombinant vesicular stomatitis virus (rVSV)-based bivalent vaccine against COVID-19 delta variant and influenza virus. bioRxiv. 2021 [Unpublished]. [DOI:10.1101/2021.12.14.472657]

[46] Bai Y, Wang Q, Liu M, Bian L, Liu J, Gao F, et al. The next major emergent infectious disease: reflections on vaccine emergency development strategies. Expert Rev Vaccines. 2022; 21(4):471-81. [DOI:10.1080/14760584.2022.2027240] [PMID]

[47] Berry DA, Berry S, Hale P, Isakov L, Lo AW, Siah KW, et al. A cost/benefit analysis of clinical trial designs for COVID-19 vaccine candidates. Plos One. 2020; 15(12):e0244418. [DOI:10.1371/journal.pone.0244418] [PMID] [PMCID]

[48] Orosco F. Recent advances in peptide-based nanovaccines for re-emerging and emerging infectious diseases. J Adv Biotechnol Exp Ther. 2024; 7(1):106-17. [DOI:10.5455/jabet.2024.d10]

[49] Hotez PJ, Bottazzi ME. Developing a low-cost and accessible COVID-19 vaccine for global health. PLoS Negl Trop Dis. 2020; 14(7):e0008548. [DOI:10.1371/journal.pntd.0008548] [PMID] [PMCID]

[50] Pollet J, Chen WH, Strych U. Recombinant protein vaccines, a proven approach against coronavirus pandemics. Adv Drug Deliv Rev. 2021; 170:71-82. [DOI:10.1016/j.addr.2021.01.001] [PMID] [PMCID]

[51] McMillan CLD, Young PR, Watterson D, Chappell KJ. The next generation of influenza vaccines: Towards a universal solution. vaccines. 2021; 9(1):26. [DOI:10.3390/vaccines9010026] [PMID] [PMCID]

[52] Roider J, Meissner T, Kraut F, Vollbrecht T, Stirner R, Bogner JR, et al. Comparison of experimental fine-mapping to in silico prediction results of HIV-1 epitopes reveals ongoing need for mapping experiments. Immunology. 2014; 143(2):193-201. [DOI:10.1111/imm.12301] [PMID] [PMCID]

[53] Gralinski LE, Menachery VD, Morgan AP, Totura AL, Beall A, Kocher J, et al. Allelic variation in the toll-like receptor adaptor protein ticam2 contributes to SARS-coronavirus pathogenesis in mice. G3. 2017; 7(6):1653-63. [DOI:10.1534/g3.117.041434] [PMID] [PMCID]

[54] Harris R, Yang J, Pagan K, Cho SJ, Stout-Delgado H. Antiviral gene expression in young and aged murine lung during H1N1 and H3N2. Int J Mol Sci. 2021; 22(22):12097. [DOI:10.3390/ijms222212097] [PMID] [PMCID]

[55] Devi A, Chaitanya NSN. In silico designing of multi-epitope vaccine construct against human coronavirus infections. J Biomol Struct Dyn. 2021; 39(18):6903-17. [DOI:10.1080/07391102.2020.1804460] [PMID] [PMCID]

Table S1. T Cell and B cell epitopes prediction description link

Name		Description Link
IEDB	SVM- and ANN-based method for prediction	http://tools.immuneepitope.org/mhci
SYFPEITH	A database of MHC ligands and peptide motifs; predictive server for MHC binding peptide	http://www.syfpeithi.de
NetMHC	ANN-based method for prediction of HLA	http://www.cbs.dtu.dk/services/NetMHC
NetCTL	Prediction of cytotoxic T lymphocyte (CTL) epitope	http://www.cbs.dtu.dk/services/NetCTL
CTLpred	Prediction of T-cell epitope	http://www.imtech.res.in/raghava/ctlpred/
PropredI	Predict MHC class I binding peptides	http://www.imtech.res.in/raghava/propred1
Propred	Predict MHC class II binding peptides	http://www.imtech.res.in/raghava/propred
	B cell epitopes prediction	
IEDB	Physiochemical properties of amino acids based predictive server for linear B cell epitope	http://tools.immuneepitope.org/tools/bcell/iedb_input
Bcepred	Physio-chemical properties of amino acids based predictive server for linear B cell epitope	http://www.imtech.res.in/raghava/bcepred
ABCpred	ANN-based predictive server	http://www.imtech.res.in/raghava/abcpred
BepiPred	Predictor of linear B cell epitopes using a combination of a hidden Markov model and a propensity scale method	http://www.cbs.dtu.dk/services/NetCTL

Abbreviations: ANNs: Artificial neural networks; SVM: Support vector machine; MHC: Major histocompatibility complexes; HLA: Human leukocyte antigen.

Table S2. Final identified cytotoxic T lymphocyte (CTL) epitopes by CTLpred

Position	Sequence	Vaxijen Score
383-392	KRIENLNKK	Probable ANTIGEN
385-394	IENLNKKVD	Probable ANTIGEN

Table S3. Final identified CTL epitopes by NetCTL

Position		Sequences	Vaxijen Score
MHC supertype A2, MHC supertype B8	258	YLQPRTFLL	Probable ANTIGEN
MHC supertype A3	490	RIYQILAIY	Probable ANTIGEN
MHC supertype B8	258	YLQPRTFLL	Probable ANTIGEN
MHC supertype B27	676	KRKRSQMLF	Probable ANTIGEN
MHC supertype B39	418	YHDSNVKNL	Probable ANTIGEN
MHC supertype B58	594	QSQIVSFYF	Probable ANTIGEN

Table S4. Nest analysis result

Nest	Score	Residue Range	Residue	Ramachandran Region	Cleft	Depth in Cleft	Residue Conservation
1.	3.80	Leu292(A)-Gly294(A)	Leu292(A)	RIGHT		15.22	1.00
			Lys293(A)	LEFT	-	-	1.00
			Gly294(A)	-	-	-	0.40
2.	3.00	Ile530(A)-Gly532(A)	Ile530(A)	RIGHT		15.45	0.00
			Gly531(A)	LEFT		19.27	0.00
			Gly532(A)	-		14.89	0.00
3.	2.00	Asn88(A)-Ile90(A)	Asn88(A)	LEFT	8	9.87	1.00
			Ile89(A)	RIGHT	8	16.00	1.00
			Ile90(A)	-	-	-	1.00
4.	2.00	Asn271(A)-Thr273(A)	Asn271(A)	RIGHT	-	-	1.00
			Gly272(A)	LEFT	-	-	1.00
			Thr273(A)	-	10	4.51	1.00
5.	0.00	His380(A)-Glu382(A)	His380(A)	LEFT	-	-	0.00
			Leu381(A)	RIGHT	-	-	0.00
			Glu382(A)	-	-	-	0.00
6.	0.00	Arg431(A)-Gln433(A)	Arg431(A)	RIGHT	-	-	0.00
			Ser432(A)	LEFT	-	-	0.00
			Gln433(A)	-	-	-	0.00
7.	0.00	Val553(A)-Glu555(A)	Val553(A)	LEFT	-	-	0.00
			Lys554(A)	RIGHT	-	-	0.00
			Glu555(A)	RIGHT	-	-	0.00

This Page Intentionally Left Blank