Immunoinformatics design of Bivalent Vaccine targeting S1-NTD and HA2 to simultaneously protect against SARS-CoV-2 and Influenza infections

Marzieh Rezaei (m.rezaeibio@gmail.com)
University of Isfahan
Somaieh Sabzali
University of Isfahan
Mohammad Satari
Malayer University

Research Article

Keywords: SARS-COV2, Influenza A (H1N1), Bivalent vaccine, Fusion peptide, HA 2 protein, spike glycoprotein S1

Posted Date: April 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1532015/v1

Abstract

Introduction: Two of the most challenging viruses for vaccine development are SARS-COV-2 causing the current COVID-19 pandemic and influenza virus (H1N1) which spread annually causing seasonal epidemics or increase the pandemic risk. In this study, we analyzed the immonodominant 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 IFN gene by two (G4S)3 linker.

Method: The 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 were accessed and the docking or binding affinity of designed vaccine to the Toll-like receptor 3 (TLR3) were analyzed. The efficiency of the constructed vaccine confirmed by appropriate expression of 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 k12.

Result and discussion: In conclusion, we suggest that this fusion peptide would be an attractive design strategy toward developing bivalent vaccine against both COVID-19 and Influenza as promising vaccine candidate without need to reformulation or vaccination each year.

1. Introduction

It was probably never thought that the illness of the five patients who were hospitalized in Wuhan, China, would cause a terrible pandemic of the century. The disease was caused by a virus from the family of coronaviruses that had previously caused SARS and MERS epidemics, but this time the infectivity of the virus was much higher than the previous two viruses. The new coronavirus, so called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Symptoms in people with Covid-19 have a wide range variety, including 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 (Zhang et al. 2020).The S glycoprotein of SARS-COV-2 virus is most important surface protein which includes the receptor binding S1-subunit and the membrane fusion S2-subunit. The N-terminal domain (NTD) of the S1 subunit is known as undoubtedly immunogenic peptide and as a receptor binding domain which 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 (Dai and Gao 2021).

On the other hand, the influenza virus (H1N1), which was first identified in 2009 and then in 2015, led to a pandemic. Between 250,000 and 500,000 deaths occur worldwide each year from the flu virus. The genome of this virus is made up of 8 pieces of RNA. 18 serotypes of hemagglutinin and 11 serotypes of neuraminidase have been identified, the most famous of which is H1N1(Dou et al. 2018) .According to the continuous antigenic drift and high variable region of H1N1, the rate of Influenza infection is being increased yearly. Emerging new strains due to high mutation rates emphasize the need to develop broadly protective vaccine against high spreading H1N1 each year (Dou et al. 2018). Antigenic variation mechanism of influenza virus to evade the human immune system because of continuous re-assortment in its segmented genome can lead to recurrent infections in humans with high rates of morbidity and mortality and increases the potential pandemic risk. Annual influenza infections and epidemics caused by H1N1 are due to high rates of mutations in the hemagglutinin and neuraminidase (Krammer and Palese 2019). Hemagglutinin (HA) is one of the two important surface glycoproteins of the influenza virus (Krammer and Palese 2019). Because of different mutation changes considered as antigenic drift in the HA1 domain of the hemagglutinin protein which causing to emergence of new virus strains(Wang and Palese 2011), the conserved regions of hemagglutinin need to be targeted to design a vaccine based on the HA2 domain of the influenza virus (Du et al. 2020).

In previous studies, a set of different platform was conducted to develop vaccine against influenza based on the variable domain of hemagglutinin protein (Song et al. 2015; Angeletti et al. 2017). So, one of the purposes of this paper was to introduce the strategy for vaccine designing based on the constant domain of the major surface glycoprotein (HA2) to confer protection against H1N1 without need to reformulation of influenza vaccines due to high variable new emerging strains in each year (Gerdil 2003). Recently, IFN-γ has licensed for human vaccines as an adjuvant and as a proper antiviral cytokine in this construct to highly improve the effective immune response .Novel approaches in vaccine development has been designed to fuse immunogenic peptides with immunostimulatory cytokines as a self-adjuvant (Tripathi and Shrivastava 2018). So, we selected the IFN-γ as an adjuvant and also antiviral cytokine fused to S1-NTD and HA2 peptides to propose an attractive candidate vaccine to simultaneously protect against SARS-CoV-2 and influenza viruses. We introduced the fusion Peptide to include the S1-NTD of SARS-COV2, HA2 of H1N1 and IFN as an adjuvant along with appropriate linkers. Glycine rich linker, such as (G4S)3 was preferred to link the peptides as it improves overall solubility and flexibility between fused peptides.

2. Material And Methods

The comprehensive bioinformatics analysis was conducted using prediction software to predict (*i*) T- and B-cell epitopes, (*ii*) the secondary and tertiary structures, and (*iii*) the antigenicity of fusion protein.

2.1. The protein sequence retrieval and designing the construct (S1-NTD-HA2-IFN)

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 IFN gene (NM_000619) by two (G4S)3 linker. The 3 end of the fusion gene includes a six-histidine tag for easier purification.

2.2. B-cell and T-cell epitope prediction and Confirmation of Antigenicity

A variety of freely online accessible servers like ABCpred, BepiPred, BCPred, and IEDB (Immune-Epitope-Database And Analysis-Resource) were used to predict which a lot of B cell epitopes of the fusion protein. The online prediction tools IEDB, SYFPEITH, PropredI, Propred, Net MHC, Net Cytotoxic T lymphocyte (CTL), CTL Pred were assessed for their ability to predict the T-cell epitopes. All predicted epitopes were detected as probable antigen with Vaxijen v2.0 value (threshold 0.4%).

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

2.3. Prediction of Secondary structure of the fusion peptides, the transmembrane helix and signal peptide

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

In addition, to predict the surface exposed epitopes, the TMHMM (Transmembrane Helices Hidden Markov Models Server v.2.0 was used. The transmembrane helix (TMH) of the construct was predicted using the online TMHMM server *v*2.0, (Wu and Zhang 2007) and the potential signal peptide (SigP) cleavage site was identified by The SignalP *v*4.1 (D-cutoff: 0.45) online tool.

2.4. Physicochemical properties and efficiency of the vaccine construct

The post-injection behavior of the designed vaccine into the body is the main 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. In this web-server, various parameters were computed; including (i) molecular weight (kDa), (ii) theoretical isoelectric point (pI), (iii) in vitro and in vivo based estimated half-life, (iv) stability index, (v) aliphatic index, (vi) extinction coefficient, and (vii) grand average of hydropathicity (GRAVY)

2.5. Three dimensional structure modelling

The 3-dimensional modeled structure of final vaccine construct generated by I-TASSER software (https://zhanglab.ccmb.med.Umich.Edu/I-TASSER/). The modelled structure was refined with online server Galaxy refine tool (Ko et al. 2012).The overall model quality was checked using ProSA web-server.

The stereochemistry quality in the vaccine model was analyzed based on the Ramachandran plot. To validate the 3D model, Rampage (Lovell et al. 2003) server was used based on the allowed and disallowed regions of the protein structural model.

2.6. SiteSeer search

The SiteSeer search was managed to find our designed structure itself, to scan the target structure against a prepared database of templates and to match functionally important sites (Laskowski et al. 2005).This raises the possibility to report possible matches after scanning 400 auto-generated templates from the query structure against representative structures in the PDB.

2.7. 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 (Consortium 2007).

2.8. 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 (Pal et al. 2020). Simple Nests are either RL or LR. Larger nests (> 2 residues long) may be RLR, LRL, RLRL, etc., which are 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(Pal et al. 2020). ProFunc server (http://www.ebi.ac.uk/thornton-srv/databases/ProFunc) was used for predicting probable protein function from 3D structure.

2.9. Docking and Binding Affinity Analysis of Vaccine candidate with TLR-3

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 (Xue et al. 2016). The CASTp web server was used to determine the active binding pockets of refined vaccine construct for TLR-3 receptor (Sharma et al. 2021).

The use of new molecular methods beside a clinical trial can lead to get a better result. Due to the genome organization of 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 third 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 designed vaccine and human Toll-like receptor (TLR3) protein was performed from protein data bank web site (http://www.rcsb.org/pdb). Molecular docking investigations were performed by using Molegro Virtual Docker 7 to analyze the interaction probability of this vaccine against TLR3.

2.10. Reverse Translation, Codon Adaptation Index (CAI) and in silico expression of vaccine candidate

The codon adaptation plays an important 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 by applying an algorithm from Carbone et al.(Carbone et al. 2003). An in-silico cloning and expression of vaccine candidate in *E. coli* pET–28(+) vector was performed using the SnapGene 4.2 tool to verify the maximum expression of vaccine in expression vector at *XhoI* and *NotI* restriction sites.

2.11. The conservation level and cross-protection of designed vaccine

Due to emergence of new strain of SARS-COV-2 and current challenging to control new epidemiological situation 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 that deposited in NCBI.

3. Result

3.1. Sequence retrieval

The protein sequence retrieval and designing the construct (S1-NTD-HA2-IFN) 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 Human IFN gene by two (G4S)3 linker.

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

A lot of MHC I and MHC II epitopes were predicted as probable by Net MHC, SYFPEITH, Propred and PropredI online servers. T cell and B cell epitopes prediction description Link is shown in Table S1 and the final result were shown in Table 1.

CTLpred and NetCTL 1.2 servers (Larsen et al. 2007) were used for reliable predicting human CTL epitopes toward rational vaccine design. The prediction result by CTLpred and NetCTL 1.2 servers are shown in Table S2 and Table S3, respectively. In addition, Enzymatic degradation of 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 epitopes in Table 2.

Finally, two identified epitopes were confirmed by VaxiJen v2.0 server as probable antigens (keeping score threshold of 0.4). The first probable antigenic epitope was amino acid residues from 383–392 (KRIENLNKK) and another epitope was predicted from 385–394 residues (IENLNKKVD) (Table 1.) Final antigenicity of the fusion peptides was confirmed by VaxiJen v2.0 server [18] with a score of 0.5107. A score > 0.4 is considered to be antigenic and indicates it is a probable antigen to activate immunity against SARS-COV-2 and Influenza. The position of final predicted epitope with high score in vaccine designed and verified by VaxiJen v2.0 server was shown in Fig. 1.

3.3. Analysis of Physicochemical properties and efficiency of the vaccine candidate

To understand the vaccine response and its stability, the physiochemical properties of the vaccine construct were analyzed with the use of online tool ProtParam (Gasteiger et al. 2005). The designed construct of 697 amino acid residues with the molecular weight of 78.23 KDa, also found well fit within the defined range (40–50 KDa) of average molecular weight of vaccine based on recombinant protein (Kumar Pandey et al. 2018). Some important properties of the protein can affect its antigenic nature. So, we assessed some physiochemical properties (e.g. the theoretical pI, the GRAVY index, the instability index etc). The theoretical pI of the vaccine was found to be 8.04. The aliphatic index of vaccine was 73.69, which suggest that the vaccine would be thermostable nature. The higher aliphatic index of a protein can be indicated as greater thermostability (Walker 2005). 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 *Escherichia coli, in vivo.* The Grand average hydropathicity (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 aqueous environment. The instability index of vaccine candidate was 36.68, which indicates it as stable protein. Generally a protein whose instability index is < 40 is classified as a stable protein(Walker 2005).

3.4. Analyzing the Secondary structure

Analyzing the secondary structure by SOPMA method and TMHMM servers showed that the random coils was detected 32.14% (Fig. 2.). The higher presence of predicted random coil shows the higher prefer binding to an antibody. Increased number of some secondary structures e.g. extended strands and random coils in the protein indicating more antigenicity of proposed epitopes (Shaddel et al. 2018) and illustrating predicted relative surface accessibility. To predict the surface exposed epitopes, the TMHMM Server v.2.0 was used and the graphical output of TMHMM showed that the amino acid residues from 1-494 are outside and would be probability considered as surface exposed sequences.

3.5. Structure-based assessments of the vaccine construct

Structure-based assessments of the vaccine construct obtained after molecular modelling. The 3-dimensional modeled structure for the fusion peptides was generated by I-TASSER software (Yang and Zhang 2015). Five models of the 3-dimensional structure were predicted. Model 1 from the output of the I-TASSER server was selected (Fig. 3A).The information about modelled structure refinement after analysis with online server Galaxy refine tool was shown in Fig. 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 (Fig. 3C), which indicates that it is near to experimentally determined structures of similar sizes and also confirms the near-native configuration [25]. The analysis of 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 (Fig. 3D). So, it would be suggested that the structural model has high quality.

3.6 Binding affinity of vaccine candidate and Kd prediction

The result of PRODIGY web server (Xue et al. 2016) showed the negative ΔG value of our designed vaccine candidate (-6.2 ΔG kcal mol^{-1}).The negative ΔG value reveals that the molecular association between our designed candidate and TLR3 structure would be thermodynamically possible (Kar et al. 2020). In addition, the dissociation constant (Kd) value was found to be $2.7E^{-05}$ and estimated at 25 ℃.

3.6. Reverse Translation, Codon Adaptation Index (CAI) of Vaccine Candidate.

The Codon Usage of the designed vaccine candidate to the *E. coli* K12 strain, as a sequenced prokaryotic organism was obtained after undergoing reverse translation followed by codon adaptation tool. It was revealed that the CAI-Value of the improved sequence was 0.833873. As this value is greater than 0.8, it was found to be convenient (Morla et al. 2016). The GC-Content of the improved sequence was computed and found to be 53.3333. GC-Content of *Escherichia coli* (strain K12) is 50.7340. As we know that GC contents should be within 30–70%, optimal results were achieved (Ali et al. 2017). Finally, in silico cloning of vaccine construct gene in pET–28(+) expression vector was performed at *XhoI* and *NotI* restriction sites and succeeded by its virtual confirmation (Fig. 4.).

3.7. SiteSeer search

The SiteSeer search predicted the possibly 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)(Laskowski et al. 2005). These top 20 hits are the ones most likely to have been preserved and hence give the highest local similarity scores. 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 (Laskowski et al. 2005). 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".

3.8. Results of BLAST search vs UniProt

Results of BLAST search vs UniProt (Consortium 2007) showed the best alignment between the query sequence and the found sequences (hits). The BLAST High score and e-value to zero indicated the higher quality of alignment and would be helpful in best identifying, characterizing and comparative analyzing of our designed fusion protein.

3.9. Nest analysis results

Seven nests were identified and are shown in the Table S4. 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(Laskowski et al. 1996) .One of the important factors determining how proteins interact with other molecules is the size of clefts in the protein's surface (Laskowski et al. 1996). The depth in clefts indicates how they are functionally important. The large clefts are shown in the 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.

3.10. Docking result

The human Toll-like receptor (TLR3) protein has a critical role in the control and regulation of host immune response. The TLR3 protein is very sensitive against viral infection and other inflammatory responses. Figure 5. shows the schematic structure of interaction of chain A of human Toll-like receptor (TLR3) protein (red) with model protein (green). Using the Pymol software package was found that human Toll-like receptor (TLR3) protein and model protein interact through van der Waals interactions and hydrogen bonding. The RMSD from the overall lowest-energy structure is equal to 32.8 +/- 0.4. Total Van der Waals energy is equal to -88.8 +/- 8.4 and total electrostatic energy is equal to -233.4 +/- 25.5. The desolation energy is equal to -65.5 +/- 4.5. The restraints violation energy is equal to 2121.5 +/- 87.6. The buried surface area is equal to 3047.8 +/- 244.5. The

z-score for docking of human Toll-like receptor protein and model protein is equal to -1.9. As a result, a human Toll-like receptor protein could be considered as the potential binding with designed vaccine candidate due to the lowest receptor-average Z-score (Kim et al. 2019). 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 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.

3.11. 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 were shown in Table S5 and Table S6, respectively. These data reflect the fact that this designed vaccine candidate is conserved in almost all of the isolated of SARS-COV-2 and Influenza A (H1N1) viruses. This investigation also confirmed the cross- protection between Coronaviruses and Influenza viruses due to high conservation level and provided some information in terms of the vaccine coverage.

4. Discussion

Emergence of new viral strains in resource poor countries represents a huge global disease burden. There is significant interest in developing novel peptide-based infectious disease vaccines for many pathogens. COVID-19 and influenza are both highly contagious respiratory diseases cause great threat to 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 (Ao et al. 2021).

There are different strategies for developing vaccine against infectious diseases. Applying any strategy is depended on the knowledge of the mechanisms of infection of the target pathogen and evaluating the host 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 (Berry et al. 2020). One of the next generation platforms of vaccines against infectious disease is Peptide-based vaccines. Developing new vaccine strategy based on the peptide and recombinant proteins have more advantages over other vaccine developing approaches (e.g. the nucleic acid or viral vector vaccines) (Hotez and Bottazzi 2020; Pollet et al. 2021) because of lower manufacture cost, more simple and non-complicated procedure and lower complications (Tripathi and Shrivastava 2018; McMillan et al. 2021). Recently, experimentally analysis of all pathogen gene products is time and cost consuming procedure. So, applying the desirable, low cost and fast approaches are promising for designing to predict viral or bacterial antigens. It should be noticed that the experimentally determined epitopes will never be equal to predicted ones through immunoinformatics and the variable results would be anticipated (Roider et al. 2014). The antigenicity and analysis of two dimensional structure of our designed vaccine candidate showed its good accessibility and its effectiveness to be 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 to structurally favorable. Since TLR 3 has proven recognition capability in both SARS-COV (Gralinski et al. 2017) and Influenza (Harris et al. 2021), the docking analysis and CASTp server results (showing active binding package of the vaccine structure for the TLR3 receptor) can suggest that the vaccine construct has significant affinity to TLR 3 to act as an important toll like receptor for recognizing molecular patterns of pathogen and activating immune response.

5. 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 the whole viruses or whole surface proteins. The prediction methods can reduce the cost for vaccine development. The final comprehensive bioinformatic analyses have proposed that the bivalent construct (NTDS1-HA2- IFN-γ) would be a computational design strategy toward developing vaccine to overcome both influenza and COVID19 disease.

Abbreviations

SARS-COV2
Sever Acute Respiratory Syndrome Coronavirus2
S
Spike
S1-NTD
the N-terminal Domain
HA
Hemagglutinin
NCBI
National Centre for Biotechnology Information
CTL
Cytotoxic T Lymphocyte

PDB
Protein Data Bank
PI
Isoelectric Point
MHC
Major Histocompatibility Complex
TLR-3
Toll-like Receptor-3
CAI
Codon Adaptation Index

Declarations

Conflict of interest

The authors declare that there is no conflict of interest in this study.

Authors' contributions

MR conceived and designed this study; MR, SS and MS performed the experiments; MR and SS analyzed the results; MR and SS wrote the manuscript; MR, SS and MS improved and revised the manuscript. All authors read and approved the final manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Ali M, Pandey RK, Khatoon N, Narula A, Mishra A, Prajapati VK (2017) Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. Scientific reports 7:1–13
2. Angeletti D et al. (2017) Defining B cell immunodominance to viruses. Nature immunology 18:456–463
3. Ao Z et al. (2021) Development and Characterization of Recombinant Vesicular Stomatitis Virus (rVSV)-based Bivalent Vaccine Against COVID-19 Delta Variant and Influenza Virus. bioRxiv
4. Berry DA et al. (2020) A cost/benefit analysis of clinical trial designs for COVID-19 vaccine candidates. PloS one 15:e0244418
5. Carbone A, Zinovyev A, Képes F (2003) Codon adaptation index as a measure of dominating codon bias. Bioinformatics 19:2005–2015
6. Consortium U (2007) The universal protein resource (UniProt). Nucleic acids research 36:D190-D195
7. Dai L, Gao GF (2021) Viral targets for vaccines against COVID-19. Nature Reviews Immunology 21:73–82
8. Dou D, Revol R, Östbye H, Wang H, Daniels R (2018) Influenza A virus cell entry, replication, virion assembly and movement. Frontiers in immunology 9:1581
9. Du R, Cui Q, Rong L (2020) Flu Universal Vaccines: New Tricks on an Old Virus. Virologica Sinica:1–12
10. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook:571–607
11. Gerdil C (2003) The annual production cycle for influenza vaccine. Vaccine 21:1776–1779
12. Gralinski LE et al. (2017) Allelic variation in the toll-like receptor adaptor protein ticam2 contributes to SARS-coronavirus pathogenesis in mice. G3: Genes, Genomes, Genetics 7:1653–1663
13. Harris R, Yang J, Pagan K, Cho SJ, Stout-Delgado H (2021) Antiviral Gene Expression in Young and Aged Murine Lung during H1N1 and H3N2. International journal of molecular sciences 22:12097
14. Hotez PJ, Bottazzi ME (2020) Developing a low-cost and accessible COVID-19 vaccine for global health. PLoS neglected tropical diseases 14:e0008548
15. Kar T et al. (2020) A candidate multi-epitope vaccine against SARS-CoV-2. Scientific reports 10:1–24
16. Kim SS, Aprahamian ML, Lindert S (2019) Improving inverse docking target identification with Z-score selection. Chemical biology & drug design 93:1105–1116
17. Ko J, Park H, Heo L, Seok C (2012) GalaxyWEB server for protein structure prediction and refinement. Nucleic acids research 40:W294-W297
18. Krammer F, Palese P (2019) Universal influenza virus vaccines that target the conserved hemagglutinin stalk and conserved sites in the head domain. The Journal of infectious diseases 219:S62-S67

19. Kumar Pandey R, Ojha R, Mishra A, Kumar Prajapati V (2018) Designing B-and T-cell multi-epitope based subunit vaccine using immunoinformatics approach to control Zika virus infection. Journal of cellular biochemistry 119:7631–7642
20. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC bioinformatics 8:1–12
21. Laskowski RA, Luscombe NM, Swindells MB, Thornton JM (1996) Protein clefts in molecular recognition and function. Protein science: a publication of the Protein Society 5:2438
22. Laskowski RA, Watson JD, Thornton JM (2005) Protein function prediction using local 3D templates. Journal of molecular biology 351:614–626
23. Lovell S, Davis I, Arendall W (2003) 3RD, De Bakker, PI, Word, JM, Prisant, MG, Richardson, JS & Richardson, DC 2003. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins 50:437–450
24. McMillan CL, Young PR, Watterson D, Chappell KJ (2021) The next generation of influenza vaccines: Towards a universal solution. Vaccines 9:26
25. Morla S, Makhija A, Kumar S (2016) Synonymous codon usage pattern in glycoprotein gene of rabies virus. Gene 584:1–6
26. Pal D, Sahu S, Banerjee R (2020) New facets of larger Nest motifs in proteins. Proteins: Structure, Function, and Bioinformatics 88:1413–1422
27. Pollet J, Chen W-H, Strych U (2021) Recombinant protein vaccines, a proven approach against coronavirus pandemics. Advanced Drug Delivery Reviews
28. Roider J et al. (2014) Comparison of experimental fine-mapping to in silico prediction results of HIV-1 epitopes reveals ongoing need for mapping experiments. Immunology 143:193–201
29. Shaddel M, Ebrahimi M, Tabandeh MR (2018) 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. Journal of Parasitic Diseases 42:269–276
30. Sharma R, Rajput VS, Jamal S, Grover A, Grover S (2021) An immunoinformatics approach to design a multi-epitope vaccine against Mycobacterium tuberculosis exploiting secreted exosome proteins. Scientific Reports 11:1–12
31. Song L et al. (2015) An avian influenza A (H7N9) virus vaccine candidate based on the fusion protein of hemagglutinin globular head and Salmonella typhimurium flagellin. BMC biotechnology 15:1–8
32. Tripathi NK, Shrivastava A (2018) Recent developments in recombinant protein–based dengue vaccines. Frontiers in immunology 9:1919
33. Walker JM (2005) The proteomics protocols handbook. Springer
34. Wang TT, Palese P (2011) Catching a moving target. Science 333:834–835
35. Wu S, Zhang Y (2007) LOMETS: a local meta-threading-server for protein structure prediction. Nucleic acids research 35:3375–3382
36. Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A (2016) PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. Bioinformatics 32:3676–3678
37. Yang J, Zhang Y (2015) I-TASSER server: new development for protein structure and function predictions. Nucleic acids research 43:W174-W181
38. Zhang W et al. (2020) 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. Clinical immunology 214:108393

Tables

Table 1. Final identified B-Cell and T –Cell epitopes.

Final identified B –Cell epitopes		
Position	Sequences	vaxiJen score
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		
Position	Sequences	vaxiJen score
383-392	KRIENLNKK	Probable ANTIGEN
385-394	IENLNKKVD	Probable ANTIGEN

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

Protein digest analysis of final B- cell epitopes.				
p Position Position	epitopes	Mass (Da)	pI	Undigested enzyme
246-262	GWTAGAAAYYVGYLQP	1687.87	5.52	Trypsin, Clostripain, Cyanogen_Bromide, Proline_Endopept, Staph_Protease, Trypsin_K, Trypsin_R, AspN
509-517	VSLGAISFW	979.14	5.49	Trypsin, Clostripain, Cyanogen_Bromide, IodosoBenzoate, Proline_Endopept, Staph_Protease, Trypsin_K, Trypsin_R, AspN,
5-23	VNLTTRTQLPPAYTNSFTR	2180.45	10.83	Cyanogen_Bromide, IodosoBenzoate, Staph_Protease, Trypsin_K, AspN
247-256	WTAGAAAYY	973.05	5.52	Trypsin, Clostripain, Cyanogen_Bromide, Proline_Endopept, Staph_Protease, Trypsin_K, Trypsin_R, AspN,

Protein digest analysis of final T- cell epitopes.				
Position	Epitopes	Mass (Da)	pI	Undigested enzyme
383-392	KRIENLNKK	1142.37	10.29	Chymotrypsin, Cyanogen_Bromide, IodosoBenzoate, Proline_Endopept, AspN,
385-394	IENLNKKVD	1072.23	6.19	Chymotrypsin, Clostripain, Cyanogen_Bromide, IodosoBenzoate, Proline_Endopept, Trypsin_R,

Table 3: The SiteSeer search predicted result

Hit no.	E-value	Simil-arity score	Neighbours ident/simil [equiv]	Template id	Matched PDB entry	Longest fitted segment	Seq lengths query/target	Over-lap	%-tage seq id	Structural similarity
1	0.00E+00	870.00	47/0 [47]	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 [46]	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 [45]	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 [44]	TMP00395	6acc×34: Trypsin-cleaved and low ph-treated sars-cov spike glycoprotein and ace2 complex, ace2-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 [41]	TMP00395	5x4s×76: Structure of the n-terminal domain (ntd)of sars-cov spike protein	150/256	697/269	293	54.28%	99.6%
6	0.00E+00	725.94	36/7 [43]	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 [36]	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%

Supplementary Tables

Supplementary Tables S1-S6 are not available with this version.

Figures

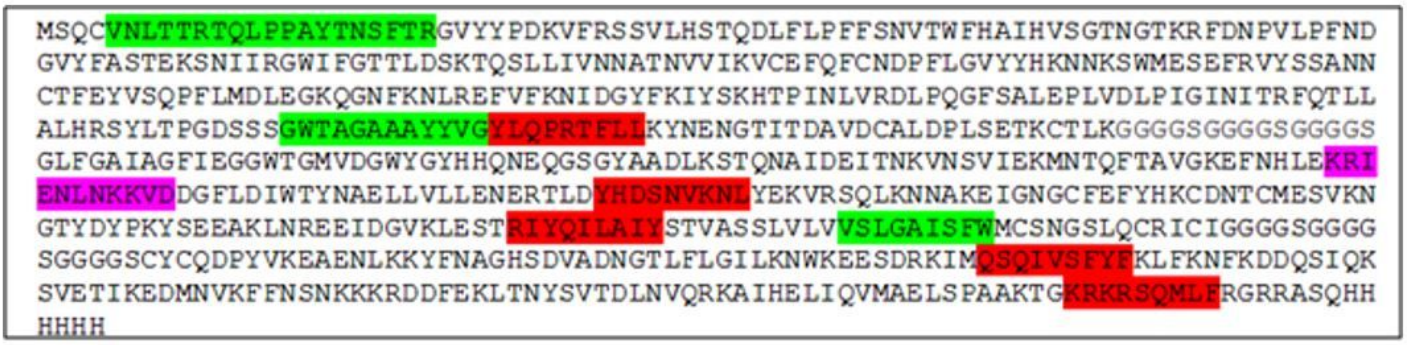

Figure 1

The position of final predicted epitope with high score by VaxiJen v2.0.

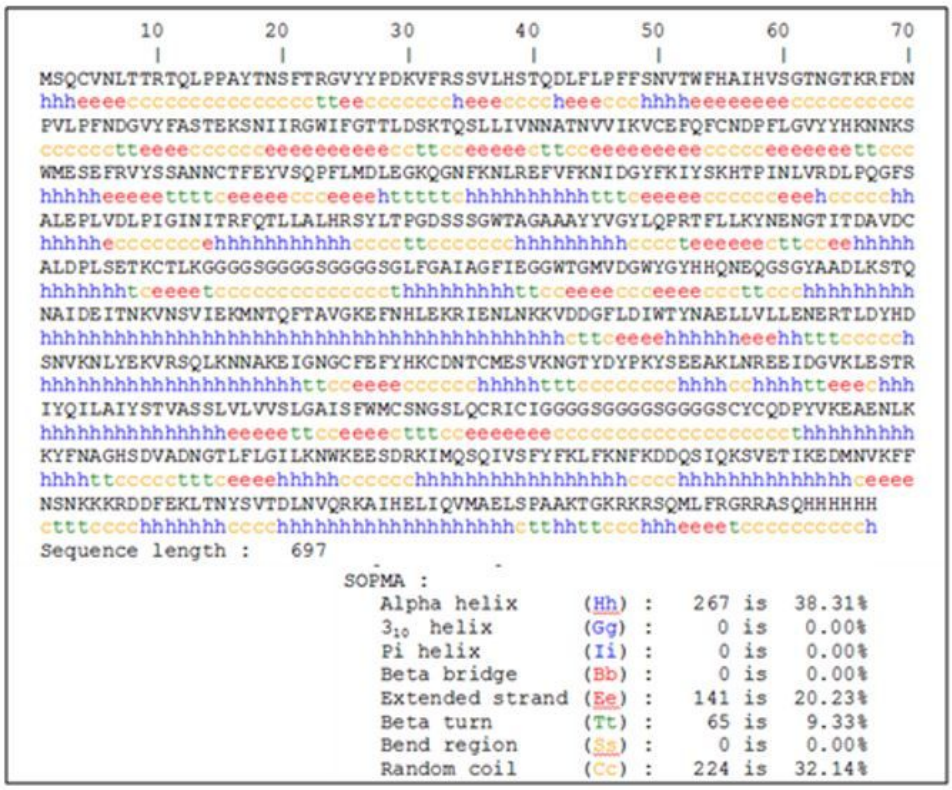

Figure 2

Analysis of secondary structure of the fusion peptides by the SOPMA method:

Four conformational states are presented in different color lines: Blue: α helix; Green: β turn; Red: extended strand and Purple: random coil.

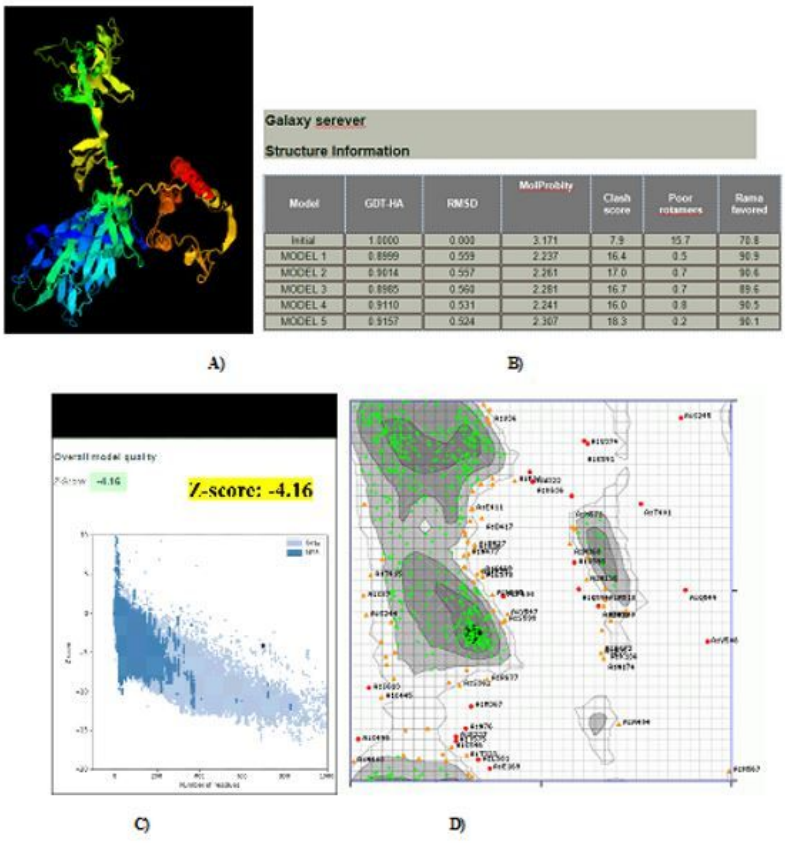

Figure 3

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

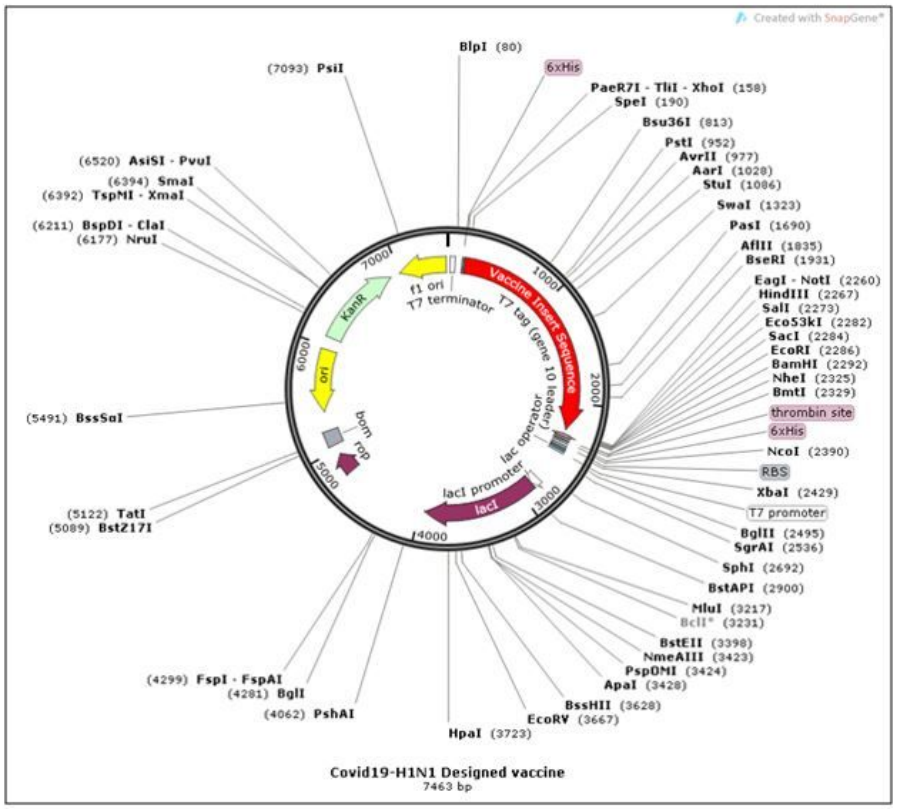

Figure 4

In silico cloning of vaccine. The segment represented in red is the designed vaccine insert in pET-28(+) expression vector.

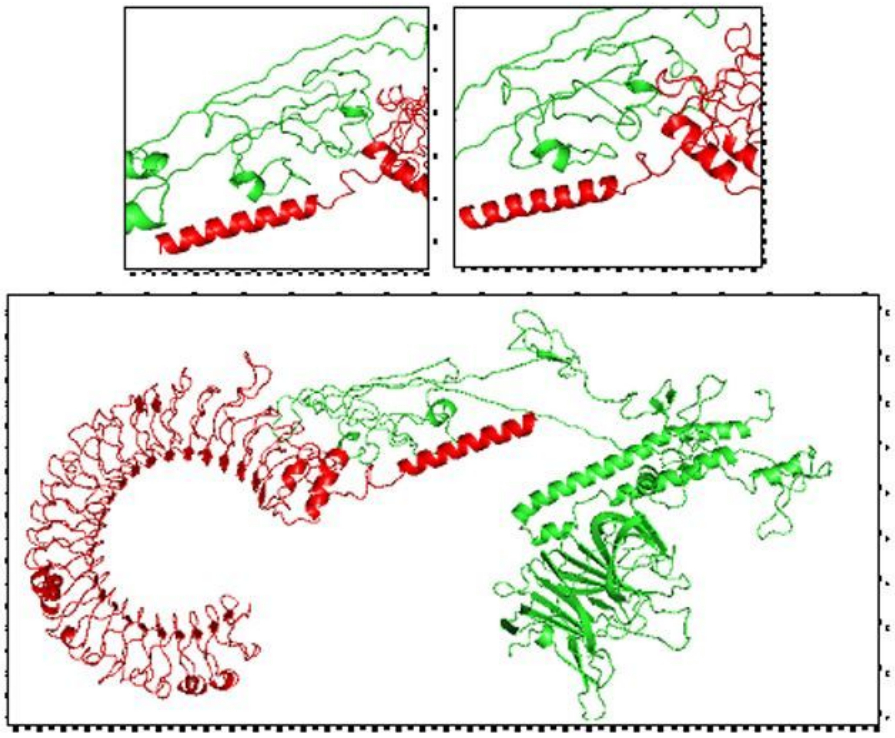

Figure 5

Interaction of human Toll-like receptor (TLR3) protein (red) and model protein (green).