Original Article

Computer-Assisted Multi-Epitopes T-Cell Subunit Covid-19 Vaccine Design

Abdulwahid Abaukaka Yahaya[1], Salihu Sanusi[2], Fatima Umar Malo[3]

Departments of [1]Medical Biotechnology and [2]Toxicology and Pharmacology, Tehran University of Medical Sciences, Tehran, Iran, [3]Department of Chemistry Nasarawa State University, Keffi, Nigeria

Abstract

Background: The world is currently facing the coronavirus disease-2019 pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Researchers from different parts of the world have employed diverse approaches to create a safe and effective vaccine as it saves millions of lives. Vaccines are created from the viral particle to train the body for a natural defense against invading pathogens. It is important to understand the concept of the vaccine design, especially the multi-epitope T-cells subunit vaccine. **Methods:** In this regard, we employed bioinformatics and immunoinformatic tools to illustrate the concept of the computer-based vaccine design. The computational methods consist of evaluation and selection of SARS-CoV-2 structural proteins, prediction of cytotoxic T-lymphocyte (CTL) epitopes, prediction of helper T-cell (HTL) epitope, multi-epitope vaccine candidate construct, antigenicity and allergenicity prediction of the designed candidate vaccine, physiochemical properties and solubility evaluation, secondary/tertiary structure prediction, refinement and validation of model vaccine tertiary structure, molecular docking of fusion proteins and Toll-like receptor 9 protein, and *in silico* cloning of the vaccine. **Results:** A total of 454 amino acid sequences were generated from CTL and HTL epitopes. The query solubility value (QuerySol) of the vaccine construct was 0.419, including the human β-defensin-2 adjuvant and peptide linkers. A circular clone of vaccine and pEX-C-His plasmid was achieved after *in silico* ligation using the annealed primer. **Conclusion:** Here, we provide essential information on computer-assisted multi-epitopes T-cell subunit vaccine design.

Keywords: Computer-assisted, coronavirus disease-2019, multi-epitopes, severe acute respiratory syndrome coronavirus 2, vaccine

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a positive-sense single-stranded RNA virus from genus *Beta coronavirus* (β-CoV) known to cause coronavirus disease-2019 (COVID-19) as the World Health Organization (WHO) declared the outbreak a pandemic.[1] The phylogenetic analysis of the multiple sequence alignment of SARS-CoV-2 shows that the genome is highly identical to bat coronavirus[2] and pangolin coronavirus (Pangolin-CoV) that are 91.02% and 90.55% genetically similar to SARS-CoV-2 and BatCoV RaTG13, respectively, from the complete genomic sequence.[3] By March, 2020; WHO announce person to person transmission through infectious droplet. The report was based on investigation in 66 infected cases. However, droplets are not airborne because of the weight but eventually land on surfaces which people get infected from by touching their eyes, nose, or mouth after they might have touched the surfaces. However, current evidence shows that smaller nuclei particle from the droplets is transmissible within a distance of 1 m from the aerosol in clinical settings.[4] The shape of the SARS-CoV-2 virion is spherical and enveloped with the spike protein; however,[5] on the size of the viral particles, the result from different studies varies. According to Neuman *et al*., the particle diameter ranges from 50 to 150 nm, without the spikes protein, with a mean diameter of 82–94 nm,[6] Cascella *et al*. described the size of the particle to be approximately 60–140 nm.[7] On the contrary, the report of Kannan *et al*. indicates the particle size to be 150–160 nm.[8] The most cited opinion on the description of the viral particle size was the study of

Access this article online

Quick Response Code:

Website: www.bmbtrj.org

DOI: 10.4103/bbrj.bbrj_5_21

Address for correspondence: Dr. Abdulwahid Abaukaka Yahaya, Department of Medical Biotechnology, Tehran University of Medical Sciences, Tehran, Iran.
E-mail: y-abaukaka@razi.tums.ac.ir
ORCID: https://orcid.org/0000-0001-7435-2162

How to cite this article: Yahaya AA, Sanusi S, Malo FU. Computer-assisted multi-epitopes T-cell subunit Covid-19 vaccine design. Biomed Biotechnol Res J 2021;5:27-34.

Submitted: 14-Jan-2021; **Revised:** 20-Jan-2021
Accepted: 21-Feb-2021; **Published:** 13-Mar-2021

Chen *et al*., who described the size of the enveloped virion to be approximately 50–200 nm in diameter.[9]

Like other human β-CoV, SARS-CoV-2 has five major genes with corresponding coded proteins. This includes the orf1ab gene for orf1ab polyprotein replicase, the S gene for surface glycoprotein, E gene for envelope protein, M gene for membrane protein, and N gene for nucleocapsid phosphoprotein.[10] Other genes include the ORF3a, ORF7a, ORF6, ORF8, and ORF10 with corresponding coding proteins. The surface glycoprotein is also known as spike protein that facilitates the entry of SARS-CoV-2 into the host cell via a functional receptor, the angiotensin-converting enzyme 2 through priming of the protein by the host transmembrane protease, serine 2.[11,12] The expression of SARS-CoV membrane protein (M protein) suppresses nuclear factor kappa B (NF-kappaB) activities and facilitates the gene expression of cyclooxygenase-2, indicating a potential role in the pathogenesis of SARS disease;[13] however, there has not been any reported study on the role of the M protein in the COVID-19 pandemic. β-CoVs envelope protein (E protein) are small integral protein that facilitates virus replication by the assembly, budding, enveloping, and disease progression; besides, it serves as ion-channeling viroporin and also interacts with host proteins.[14]

The major priority since the outbreak of the COVID-19 pandemic was the development of a novel vaccine that will curtail the widespread of SARS-CoV-2; however, efforts are in progress to establish vaccine products for clinical usage. Currently, the authorized/approved vaccines include Comirnaty (BNT162b2), mRNA-1273, CoronaVac, COVID-19 Vaccine AstraZeneca (AZD1222), Sputnik V, BBIBP-CorV, EpiVacCorona, and Covaxin, while those in the development include Convidicea (Ad5-nCoV), JNJ-78436735 (formerly Ad26.COV2.S), INO-4800, VIR-7831, CVnCoV, and ZyCoV-D. To understand the concept of immunogenicity and safety of these vaccines, we employed series of immunoinformatic techniques to design a multi-epitopes T-cell subunit candidate vaccine from SARS-CoV-2 structural proteins by fusion of the immunogenic short peptide (epitopes) using protein linkers. Multi-epitopes vaccine aimed at inducing multi-antigenic immunity against large complex pathogens with different strain variants.[15] To acquire immunologic relevance vaccine against COVID-19 pandemic, poly-epitope vaccine candidate that will deliver a series of antibody from T-cells epitopes is essential.

Methods

The immune cells recognized antigenic SARS-CoV-2 structural proteins through different cytoplasmic and membrane spanning receptor, thereby induced immune protective immune response against COVID-19 infection through major histocompatibility complex class (MHC) human leukocyte antigen (HLA). This study was reviewed and granted a waiver of individual informed consent by the Ethical Review Committee of the Tehran medical University.

Evaluation and selection of severe acute respiratory syndrome coronavirus 2 protein for multi-epitope vaccine design

A total of 19 SARS-CoV-2 surface glycoprotein sequences were collected from GeneBank (https://www.ncbi.nlm.nih.gov/genbank/SARS-COV-2-seqs/) based on the accession number, region, date, and country/province from the complete genomic sequence.[16] The countries include Nepal, Brazil, Taiwan, China-Hangzhou, Finland, China-Wuhan, USA-California, USA-Washington, Japan, Sweden, South Korea, Australia, Pakistan-Gilgit, and Spain-Valencia. Protein sequences were aligned in MEGAX software (https://www.megasoftware.net/) using ClustalW to determine residue-to-residue correspondence homology and evolutionary relationship of the common ancestor. The evolutionary history was inferred using the unweighted pair group method with arithmetic mean method followed by assembling of the phylogenetic tree.[17] After evaluating the sequence homology from various gene entries, a candidate gene was selected from the results for immunogenic epitope prediction of the four SARS-CoV-2.

Prediction of cytotoxic T-lymphocyte epitopes

Identification of potential epitopes is an important step in subunit vaccine design to achieve immunologic relevance multi-epitope vaccine candidate. In this study, cytotoxic T-lymphocyte (CTL) epitopes were predicted by NetCTL 1.2 server (http://www.cbs.dtu. dk/services/NetCTL/) with prediction thresholds of weight on C terminal cleavage as 0.15, weight on TAP transport efficiency as 0.05, and the threshold for epitope identification as 0.75. The method integrates prediction of peptide MHC-I (A1 supertype) binding, proteasomal C terminal cleavage, and TAP transport efficiency.[18]

Prediction of helper T-cell epitope

T-helper cells expressed class II molecules on antigen-presenting cells. The 15-mer T-helper cell epitopes were predicted by the NetMHCII 2.2 server (http://www.cbs.dtu.dk/services/NetMHCII-2.2/). The NetMHCII 2.3 server predicts binding of peptides to HLA-DR based on binding energy affinity (1-log50k) and minimum inhibitory concentration (IC_{50}) values (nM). The NetMHCII 2.3 server automatically categorized the epitopes into strong and weak affinity from the rank scored of the IC_{50} and binding energy value.[19]

Multi-epitope vaccine candidate designing

Final multi-epitope vaccine candidate was designed by joining epitopes with high-binding affinity and lower IC_{50} values using suitable linkers. The T lymphocytes (TLC) and helper T-cell (HTL) epitopes of the surface glycoprotein, envelope protein, membrane glycoprotein, and nucleocapsid phosphoprotein were linked together with AAY and GPGPG linker. A potential adjuvant with known antiviral immunomodulatory activities, human beta-defensin 2,[20] PDB

accession number 1FD3, was linked to the N-terminal of CTL epitopes through EAAAK linker. The choice of the adjuvant was based on the antiviral activity, immunoregulatory function, and potent activation of macrophages.

Antigenicity and allergenicity prediction of the designed candidate vaccine

The ability of an antigen to bind to the T-cell receptor and elicit immune response and formation of memory cells is determined by the antigenicity and antigenic propensity of the ligand protein. ANTIGENpro (http://scratch.proteomics.ics.uci.edu/) and VaxiJen v2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) were used to predict the antigenicity of the designed vaccine candidate. ANTIGENpro is a nonhomology-based method that predicts protein antigenicity based on protein microarray data. The accuracy of this tool is based on a combined dataset estimated at 76% by cross-validation experiments.[21] VaxiJen prediction is based on auto–cross-covariance (ACC) alteration of protein sequences into a constant vector of primary amino acid properties.[22] The allergic potential or allergenicity of the designed multi-epitope vaccine was predicted by AllerTOP V2.0 (https://www.ddg-pharmfac.net/AllerTOP/), AllergenFP V1.0 (http://ddg-pharmfac.net/AllergenFP/), and AlgPred (https://webs.iiitd.edu.in/raghava/algpred/submission.html) to exclude the ability of the protein causing sensitization and allergic reactions related with the IgE antibody response.

Physiochemical properties and solubility evaluation

It is known that the best designed vaccine is the one that ensure an appropriate immune response after vaccination. Therefore, the computation of various physical and chemical parameters for the multi-epitope vaccine candidate was calculated through a web server tool, ProtParam (https://web.expasy.org/protparam/). An Expert Protein Analysis System tool computed parameters including the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index (II), aliphatic index, and grand average of hydropathicity (GRAVY).[23] The solubility of the vaccine candidate was determined by Protein-Sol (https://protein-sol.manchester.ac.uk/), a web server software that takes a single amino acid sequence and returns the result of a set of solubility prediction calculations, compared to a solubility database.[24]

Secondary and tertiary structure prediction

The final vaccine candidate designed was subjected to secondary structure prediction to determine appropriate homology modeling of the construct. PSIPRED Server V4.0 (http://bioinf.cs.ucl.ac.uk/psipred/) was used to predict the secondary structure (2D) based on Position-Specific Iterated BLAST (Psi-Blast) from the protein sequence.[25] In addition to the psipred 2D prediction, RaptorX property server (http://raptorx.uchicago.edu/StructurePropertyPred/predict/) was employed to predict a 2D structure properties including solvent accessibility (ACC) and disorder region (DISO). RaptorX property server employs deep convolutional neural fields model to predict complex sequence–structure relationship and interdependency between adjacent property labels.[26] Iterative Threading ASSEmbly Refinement (ITASSER, https://zhanglab.ccmb.med.umich.edu/I-TASSER/) was used to predict the tertiary (three-dimensional [3D]) structure. I-TASSER is a graded approach to protein structure and function prediction through the identification of structural templates from the PDB by multiple threading approach LOMETS, with complete atomic models constructed simulations.[27]

Refinement and validation of model vaccine tertiary structure

ModRefiner (https://zhanglab.ccmb.med.umich.edu/ModRefiner/) and GalaxyRefine serve (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type = REFINE) were used to refine the 3D vaccine model in two-step process. ModRefiner refines protein structures from two steps: C-alpha trace and atomic-level energy minimization, with improved global and local structures, hydrogen-bonding networks, side-chain positions, side-chain positions, and less atomic overlaps.[28] GalaxyRefine is a GalaxyWEB tool based on refinement method driven by side-chain repacking that has been tested with CASP10 assessment and considered to be accurate in performance in both global and local structure quality.[29] ProSA-web tool (https://prosa.services.came.sbg.ac.at/prosa.php) was used to validate the tertiary structure of the vaccine designed. ProSA-web analyzes the protein structure statistically by determining the protein Z-score that strongly predicts erroneous segment of the structure.[30]

Molecular docking of fusion proteins and Toll-like receptor 9 protein

Molecular docking is a computer-assisted method that predicts ligand (chimeric protein) and protein (receptor) predominant

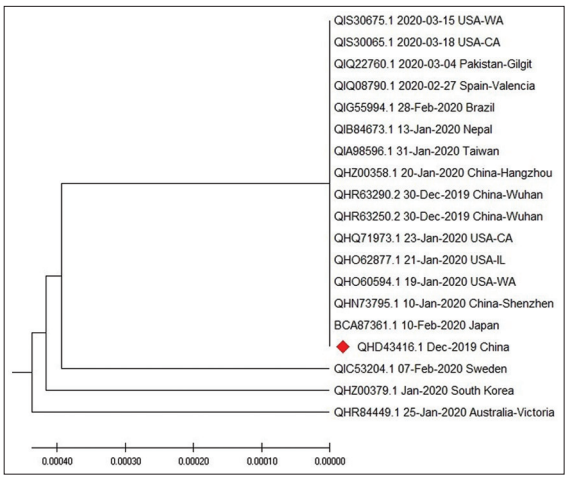

Figure 1: Phylogenetic tree of 19 severe acute respiratory syndrome coronavirus 2 surface glycoprotein from 11 different countries for sequence homology evaluation and evolutionary relationship. The evolutionary relationship was inferred by unweighted pair group method with arithmetic mean method with Jones Taylor Thornton matrix-based substitutional model

binding modes in 3D structure for virtual screening of the candidate vaccine design.[31] The binding of the designed vaccine Toll-like receptor 9 receptor was performed by a Computed Atlas of Surface Topography (CASTp) Server V3.0 (http://sts.bioe.uic.edu/castp/calculation.html). CASTp 3.0 provides accurate identifications and quantifications of protein topography and other ligand–protein assessment.[32] Molecular docking was performed by HDOCK web server. HDOCK is a free hybrid docking server whose algorithm is based on template modeling that runs for 10–20 min.[33] To determine the functional pattern of the designed multi-epitope vaccine on the receptor, COACH-D server[34] and 3DLigandSite[35] were used to predict the receptor–ligand binding site.

In silico cloning of the vaccine design

In subunit peptide vaccine, one of the major goals is the expression of the fusion protein in a host organism using an appropriate vector clone system. Java Codon Adaptation Tool JCat (http://www.jcat.de/) was used to define highly expressed gene by reverse transcription of the designed vaccine protein sequence to the optimized coding sequence. To avoid undesired cleavage sites for restriction enzymes and Rho-independent transcription terminators, additional options include avoidance of rho-independent transcription terminators, prokaryotic ribosome binding site, and cleavage sites of restriction enzymes.[36] The optimized DNA sequence of the vaccine design from JCat reverse transcription result was cloned into pEX-C-His Tagged Cloning Vector, a precision shuttle bacterial vector for inducible expression of a protein with a cleavable C-terminal 6xHis tag. The vector can be expressed in *Escherichia coli* strain BL21/DE3 as tagged protein with a C-terminal His tag. The detection and purification of the clone transgene are best done using an anti-His antibody or a nickel affinity column. *In silico* cloning was performed with molecular cloning designer simulator (MCDS). MCDS (https://github.com/johnnpeacock/-Molecular-Cloning-Designer-Simulator) is a user-friendly software platform that simulates and manages cloning for genetic engineering. The interface has an interactive flow-chart, user-defined workflow, multiple genetic recombineering module and guide to wet-lab work.[37]

Results and Discussion

Protein sequence collection

Before the collection of the structural protein sequence, we determine the homology and evolutionary relationship, of the

Table 1: Cytotoxic T-lymphocyte epitopes predicted against surface glycoprotein, envelope protein, membrane glycoprotein, and nucleocapsid phosphoprotein

Serial number	Position	Epitopes	MHC binding affinity	Combination score (Tap+Cle)
S protein				
1	865	LTDEMIAQY	0.7953	3.6616
2	258	WTAGAAAYY	0.6735	3.1128
3	604	TSNQVAVLY	0.6559	3.0758
4	361	CVADYSVLY	0.5348	2.5759
5	733	KTSVDCTMY	0.4908	2.3795
E protein				
6	34	LTALRLCAY	0.5594	2.6158
7	49	VSLVKPSFY	0.3533	1.7149
M protein				
8	213	SSDNIALLV	0.6531	2.9325
9	171	ATSRTLSYY	0.5463	2.6146
10	196	YSRYRIGNY	0.3214	1.6623
N protein				
11	104	LSPRWYFYY	0.4837	2.3408
12	164	GTTLPKGFY	0.3315	1.6848

Cle: C terminal cleavage affinity, Tap: Transport efficiency, MHC: Major histocompatibility complex

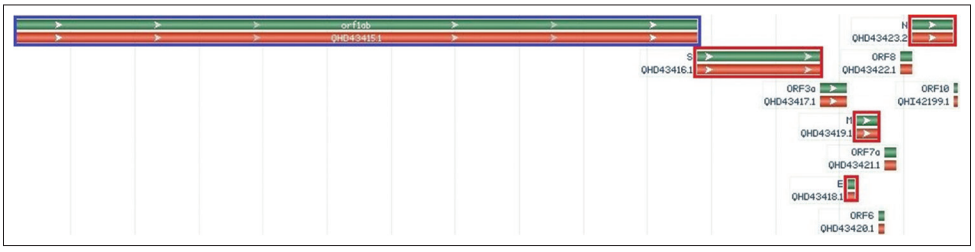

Figure 2: The graphical presentation of the complete genomic sequence of severe acute respiratory syndrome coronavirus 2 (MN908947.3). The orf1ab gene for orf1ab polyprotein replicase in a blue box; the S, E, M, and N gene in the red boxes. Other genes include the ORF3a, ORF7a, ORF6, ORF8, and ORF10 with the corresponding coding protein

Table 2: Helper T-cell epitope predicted against surface glycoprotein, envelope protein, membrane glycoprotein, and nucleocapsid phosphoprotein

Serial number	Allele	Peptide	1-log50k	IC50 (nM)
S protein				
1	DRB1_0101	VLSFELLHAPATVCG	0.9082	2.7
2	DRB1_0101	LQTYVTQQLIRAAEI	0.8745	3.9
3	DRB1_0101	GWTFGAGAALQIPFA	0.8445	5.4
4	DRB1_0404	RAAEIRASANLAATK	0.7886	9.8
5	DRB1_0701	IAIPTNFTISVTTEI	0.8280	6.4
6	DRB1_0701	SQSIIAYTMSLGAEN	0.8086	7.9
E protein				
7	DRB1_0101	LLVTLAILTALRLCA	0.8031	8.4
8	DRB1_0401	KPSFYVYSRVKNLNS	0.6527	1.7
M protein				
9	DRB1_0101	TLSYYKLGASQRVAG	0.9092	2.7
10	DRB1_1101	ASFRLFARTRSMWSF	0.8590	4.6
N protein				
11	DRB1_0701	TPSGTWLTYTGAIKL	0.8671	4.2
12	DRB1_0901	PQIAQFAPSASAFFG	0.8564	4.7
13	DRB1_1101	QIGYYRRATRRIRGG	0.8842	3.5

1-log50k: Log-transformed binding affinity, IC50: Half-maximal inhibitory concentration

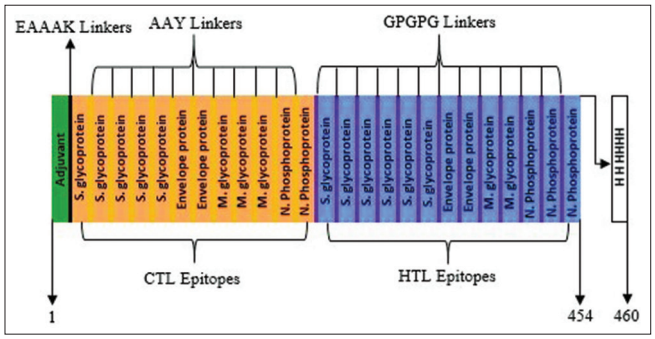

Figure 3: Schematic presentation of the multi-epitope vaccine design. The 460 amino acid vaccine construct was designed by fusing the 41 amino acid adjuvant (green) to the cytotoxic T-lymphocyte epitopes via the EAAAK linker (black) at the amino terminus. The cytotoxic T-lymphocyte peptides were fused with AAY linker (yellow) and the helper T-cell epitopes by GPGPG linker (purple). A 6xHis-tag was joined to the carboxy-terminus of the nucleocapsid phosphoprotein helper T-cell epitope

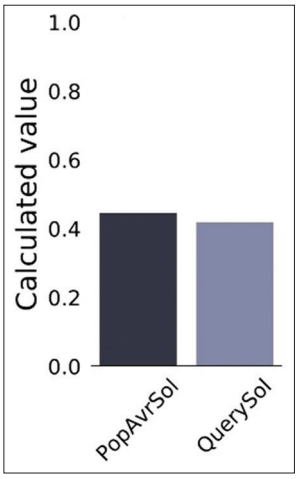

Figure 4: The query solubility value (QuerySol) of the vaccine candidate is 0.419 compared to the population average for the experimental dataset (PopAvrSol) 0.45

SARS-CoV-2, using the spike glycoprotein. The tree was drawn to scale and the branches were used to infer the phylogenetic tree. From the 19 proteins analyzed, three taxa tree were generated from MEGA X as shown in Figure 1. QHD43416.1_Dec-2019_China was used to retrieve other structural protein by mapping out the complete genome (MN908947.3) as shown in Figure 2. Other structural proteins include membrane glycoprotein or M protein (QHD43419.1), envelope protein or E protein (QHD43418.1), and nucleocapsid phosphoprotein or N protein (QHD43423.2).

Predicted cytotoxic T-lymphocyte epitopes

The CTL epitopes of the four structural proteins were predicted with the NetCTL 1.2 server. A total of 20 9-mers immunogenic epitopes were selected from the pool of the predicted epitopes based on the MHC binding affinity, tap, and cleavage score (combined). The 20 selected epitopes were further screened into 12 for the candidate vaccine protein fusion. The selected structural proteins include 5 S protein, 2 E protein, 3 M protein, and 2 N protein as shown in Table 1.

Predicted helper T-cell epitopes

A total of 21 HTL epitopes were selected from the predicted 15-mer peptides based on the allele-peptide binding energy (1-log50k) and IC_{50}. Immunogenic epitopes were tagged strong binding (SB) from the results of the NetMHCII 2.2 server prediction. Thirteen epitopes were selected for the four structural protein for vaccine candidate design. Table 2 shows the predicted HTL epitopes surface glycoprotein, envelope protein, membrane glycoprotein, and nucleocapsid phosphoprotein.

Designing the final multi-epitope vaccine

The final fusion protein for the multi-epitope vaccine candidate was contracted using 12 CTL and 13 HTL epitopes. The vaccine constructs consist of 460 amino acids residue and 4 domains, which include CTL epitopes, HTL epitopes, human Beta-defensin 2 adjuvant, and Hexa histidine-tag (6 × His-tag). The 12 predicted CTL epitopes were fused by AAY linkers, while the 13 HTL epitopes were fused by GPGPG linkers. Fusion epitopes were joined to the adjuvant at the N-terminal by EAAAK linkers. 6 × His-tag was added to the C-terminal of the epitopes/adjuvant complex for purification and identification of the chimeric protein after cloning as shown in Figure 3.

Assessment of antigenicity and allergenicity of the designed candidate vaccine

The final vaccine candidate designed was evaluated for antigenicity using ANTIGENpro and VaxiJen v2.0. The result of the prediction from ANTIGENpro was 0.721263 and VaxiJen was 0.5323. Both the results indicate the antigenic properties

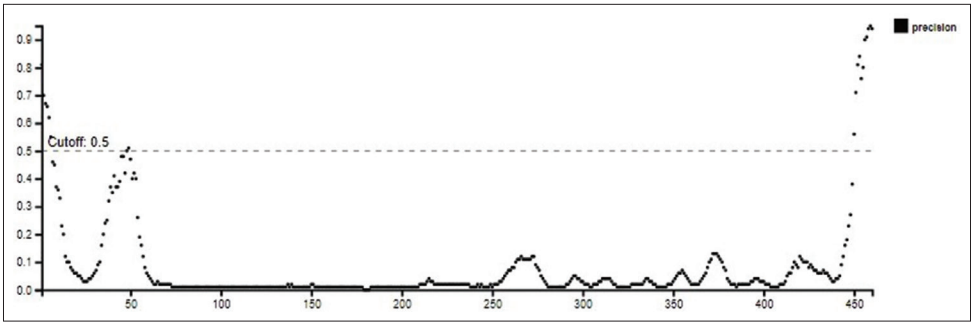

Figure 5: Graphical presentation showing the disordered position in the secondary structure of the multi-epitope vaccine candidate. 26 (5%) amino acids were predicted to be in disordered position (above 0.5 cutoff value) as shown

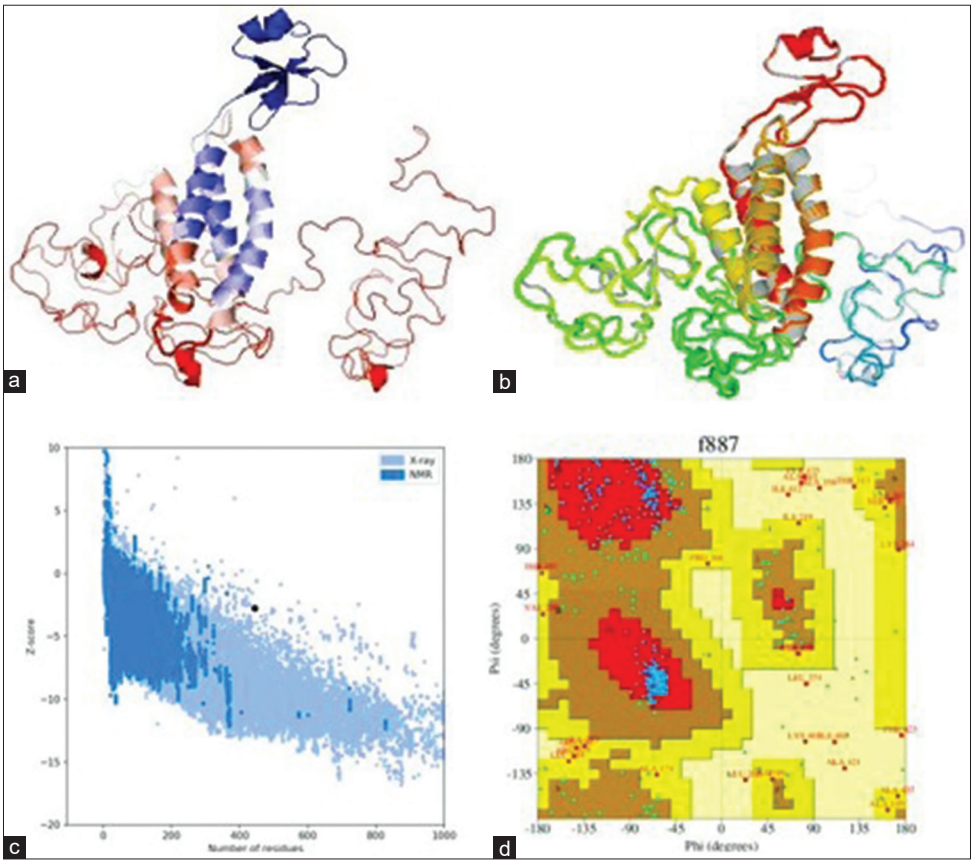

Figure 6: Graphical representation of the tertiary structure of the vaccine construct, the refined model, and validation chart. (a) Three-dimensional structure of the vaccine designed with a global model quality score of 0.2878 and estimated $P = 2.488E\text{-}1$ predicted by IntFOLD server. (b) A better quality of the three-dimensional structure refined with Galaxy web server. The GDT-HA value of the structure was 0.9555, RMSD 0.404, MolProbity 2.398, Clash score 18.7, Poor rotamers 0.6, and Rama favored 86.7. (c and d) Validation of the tertiary structure with ProSA-web server and PDB sum, respectively. The Z-score for the ProSA-web is 2.85 and most favored regions and G-Factors of the PDB sum were 71.6% and −0.38, respectively

based on interpretation of chimeric peptide. To assess the safety of the vaccine, we determine the ability of the fused protein to induce an allergic reaction using three allergenic prediction tools. The AllerTOP V2.0, AllergenFP V1.0, and AlgPred predicted the vaccine candidate as nonallergenic.

Physiochemical properties and solubility of the vaccine candidate

The physicochemical properties of the final vaccine designed were evaluated by ProtParam server tool and nine parameters were assessed. This includes the molecular weight of 47.1 kDa and theoretical pI of 9.52; total negative and positive charge residue were 14 and 37, respectively. Estimated half-life is 30 h in mammalian reticulocytes, *in vitro*; >20 h in yeast, *in vivo*; and >10 h in *E. coli*, *in vivo*. The II of the vaccine candidate is computed to be 35.32, which classifies the protein as stable, while the aliphatic index and GRAVY are 74.38 and 0.070, respectively. The solubility of the vaccine candidate was queried using the Protein-Sol server. QuerySol value was 0.419 compared to the population average for the experimental dataset (PopAvrSol) which is 0.45 as shown in Figure 4.

Secondary and tertiary structure of the vaccine construct. Refinement and validation of three-dimensional structure

The secondary structure of the final vaccine construct contains 27% alpha-helix, 19% beta-sheet, and 53% coil. The solvent accessibility predicted was 46% expose, 19% medium, and 34% buried. Using both the PSIPRED sever V 4.0 and and RaptorX property server predicated disposition of 26(5%) of amino acids [Figure5]. Five tertiary structures were predicted from I-TASSER server based on the templates of the highest significance in the threading alignments that measures the difference between the raw and average scores in the unit of standard deviation, the Z-score. We selected a model with higher C-score value as it signifies a model with higher confidence [Figure 7]. The TM-score and root mean square deviation (RMSD) estimated for the mode were 0.55 ± 0.15 and 10.2 ± 4.6Å, respectively, which indicates the model to be of better quality as shown in Figure 6a. The tertiary structure was refined with Galaxy web server in Figure 6b and subjected to validation using ProSA-web server and PDB sum tool in Figure 6c and d, respectively. Using PSIPRED server V4.0 Secondary structure of the final vaccine showing different color - as following; strand (yellow), helix (pink), and coli (ash). They were 27% alpha-helix, 19% beta-sheet (strand), and 53% coil.

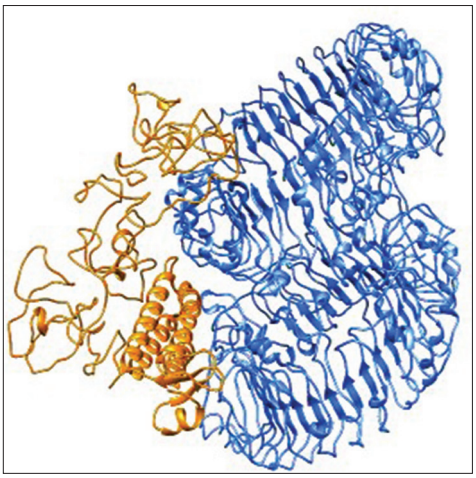

Figure 7: Chimeric structure of the docking complex between the vaccine construct (colored orange) and Toll-like receptor 9 receptor (blue)

In silico cloning of the vaccine design

The vaccine construct was cloned into pEX-C-His Tagged after digesting the plasmid with restriction enzymes, BgIII and HindIII through restriction analysis. Following the digestion of the plasmid, vaccine fragment was constructed from the nucleotide sequence of the multi-epitopes design. PCR was performed on the vaccine design to anneal the designed primer for final cloning. Vaccine fragment was cloned into the pEX-C-His plasmid by ligation, and circular clones were selected as shown in Figure 8.

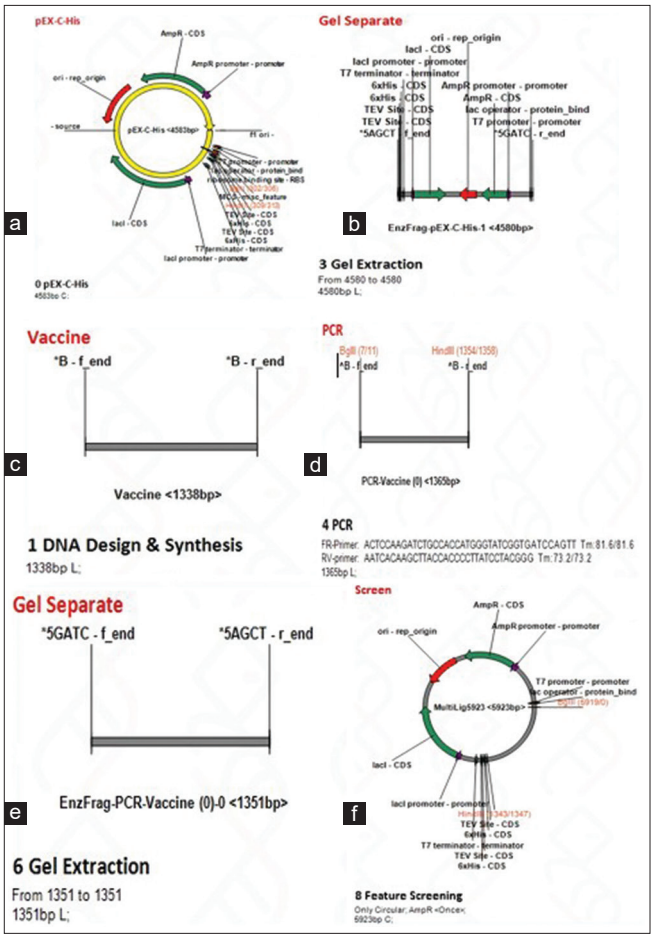

Figure 8: Cloning of vaccine construct into the pEX-C-His plasmid. (a) Restriction analysis of the plasmid to identified key restriction enzymes. (b) Gel purification after restriction digestion. (c) Vaccine fragment designed from the multi-epitope design. (d) PCR on vaccine fragment to aneal primer, cozac sequence, and restriction enzymes. (e) Gel purification of the vaccine PCR product and (f) a circular clone of the vaccine into pEX-C-His plasmid

Conclusion

We presented a comprehensive guideline for computer-aided multi-epitopes T-cell subunit vaccine design from virus epitopes selection to molecular cloning bioinformatic tools. This methodology and tools can be used to design different peptide vaccine before *in vitro* and *in vivo* validation of vaccine efficacy and safety.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents 2020;55:105924.
2. Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, *et al*. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – An update on the status. Mil Med Res 2020;7:11.
3. Zhang T, Wu Q, Zhang Z. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. Curr Biol 2020;30:1346-1351.
4. World Health Organization. Coronavirus Disease 2019 (COVID-19): Situation Report, 66. World Health Organization; 2020.
5. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, *et al*. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020;382:727-33.
6. Neuman BW, Adair BD, Yoshioka C, Quispe JD, Orca G, Kuhn P, *et al*. Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. J Virol 2006;80:7918-28.
7. Cascella M, Rajnik M, Cuomo A, Dulebohn SC, Di Napoli R. Features, Evaluation and Treatment Coronavirus (COVID-19). In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020.
8. Kannan S, ALI PS, Sheeza A, Hemalatha K. COVID-19 (novel coronavirus 2019) – Recent trends. Eur Rev Med Pharmacol Sci 2020;24:2006-11.
9. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, *et al*. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. Lancet 2020;395:507-13.
10. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov 2020;6:14.
11. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, *et al*. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 2020;181:271-80.e8.
12. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, *et al*. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun 2020;11:1620.
13. Fang X, Gao J, Zheng H, Li B, Kong L, Zhang Y, *et al*. The membrane protein of SARS-CoV suppresses NF-kappaB activation. J Med Virol 2007;79:1431-9.
14. Schoeman D, Fielding BC. Coronavirus envelope protein: Current knowledge. Virol J 2019;16:69.
15. Suhrbier A. Multi-epitope DNA vaccines. Immunol Cell Biol 1997;75:402-8.
16. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2016;44:D67-72.
17. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 2018;35:1547-9.
18. 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.
19. Nielsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinformatics 2009;10:296.
20. Kim J, Yang YL, Jang SH, Jang YS. Human β-defensin 2 plays a regulatory role in innate antiviral immunity and is capable of potentiating the induction of antigen-specific immunity. Virol J 2018;15:124.
21. Magnan CN, Zeller M, Kayala MA, Vigil A, Randall A, Felgner PL, *et al*. High-throughput prediction of protein antigenicity using protein microarray data. Bioinformatics 2010;26:2936-43.
22. Doytchinova IA, Flower DR. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 2007;8:4.
23. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. In: The Proteomics Protocols Handbook. Humana Press. https://doi.org/10.1385/1-59259-890-0:571: Springer; 2005. p. 571-607.
24. Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. Protein-Sol: A web tool for predicting protein solubility from sequence. Bioinformatics 2017;33:3098-100.
25. Buchan DWA, Jones DT. The PSIPRED protein analysis workbench: 20 years on. Nucleic Acids Res 2019;47:W402-7.
26. Wang S, Li W, Liu S, Xu J. RaptorX-Property: A web server for protein structure property prediction. Nucleic Acids Res 2016;44:W430-5.
27. Yang J, Zhang Y. I-TASSER server: New development for protein structure and function predictions. Nucleic Acids Res 2015;43:W174-81.
28. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys J 2011;101:2525-34.
29. Heo L, Park H, Seok C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. Nucleic Acids Res 2013;41:W384-8.
30. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007;35:W407-10.
31. Morris GM, Lim-Wilby M. Molecular docking. Methods Mol Biol 2008;443:365-82.
32. Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: Computed atlas of surface topography of proteins. Nucleic Acids Res 2018;46:W363-7.
33. Yan Y, Zhang D, Zhou P, Li B, Huang SY. HDOCK: A web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. Nucleic Acids Res 2017;45:W365-73.
34. Wu Q, Peng Z, Zhang Y, Yang J. COACH-D: Improved protein-ligand binding sites prediction with refined ligand-binding poses through molecular docking. Nucleic Acids Res 2018;46:W438-42.
35. Wass MN, Kelley LA, Sternberg MJ. 3DLigandSite: Predicting ligand-binding sites using similar structures. Nucleic Acids Res 2010;38:W469-73.
36. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, *et al*. JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 2005;33:W526-31.
37. Shi Z, Vickers CE. Molecular Cloning Designer Simulator (MCDS): All-in-one molecular cloning and genetic engineering design, simulation and management software for complex synthetic biology and metabolic engineering projects. Metab Eng Commun 2016;3:173-86.