



RESEARCH ARTICLE

Reverse vaccinology approach towards the in-silico multiepitope vaccine development against SARS-CoV-2

[version 1; peer review: 2 approved]

Vipul Kumar¹, Sudhakar Kancharla², Prachetha Kolli³, Manoj Jena¹

¹Department of Biotechnology, School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, Punjab, 144411, India

²Devansh Lab Werks, Homewood, AL, 35209, USA

³Microgen Health Inc, Chantilly, VA, 20151, USA

V1 First published: 23 Jan 2021, 10:44
<https://doi.org/10.12688/f1000research.36371.1>
 Latest published: 23 Jan 2021, 10:44
<https://doi.org/10.12688/f1000research.36371.1>

Abstract

Background: The novel severe acute respiratory syndrome related corona virus-2 (SARS-CoV-2) belongs to the "Coronaviridae" family and order "Nidovirales", which has caused the pandemic coronavirus disease 2019 (COVID-19). SARS-CoV-2 has been spread in more than a 100 countries, and more than a million have lost their lives. Vaccination and immunization could be an effective strategy to combat fatal COVID-19.

Methods: For identification of effective vaccine candidate against COVID-19, various immunoinformatics online tools and softwares were used to predict epitopes. Cytotoxic T cell epitopes, helper T cell epitopes, and B cell epitopes from three structural polyproteins (Spike, Membrane, and Nucleocapsid (SMN) based on the binding affinity towards MHC, antigenicity, non-allergenicity, and non-toxicity) were identified for vaccine development. The multiepitope based vaccine was constructed linking two additional adjuvants human beta-defensin-3 and human beta-defensin-2 at N and C terminal, respectively.

Results: The constructed vaccine sequence was found to be a good antigen and non-allergen for the human body. The constructed vaccine was docked with the TLR-3 receptor. The docked complex was further taken for molecular dynamics simulations and RMSD was calculated, which showed stable binding of the complex. The codon adaptation index (CAI) of 0.92 and GC content of 55.5% for *E. coli* (K12 strain) suggested efficient expression of the predicted vaccine.

Conclusion: The current study can be helpful in the reduction of time and cost for further experimental validations and could give a valuable contribution against this pandemic.

Open Peer Review

Reviewer Status  

Invited Reviewers

1

2

version 1


23 Jan 2021




report



report

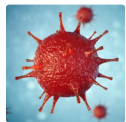
1. **Sivaraman Natarajan** , St. Jude Children's Research Hospital, Memphis, USA

2. **Tirumala Bharani K. Settypalli** , International Atomic Energy Agency, Vienna, Austria

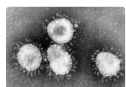
Any reports and responses or comments on the article can be found at the end of the article.

Keywords

SARS-CoV-2, COVID-19, Immunoinformatics, multiepitope, docking, simulations.



This article is included in the [Disease Outbreaks](#) gateway.



This article is included in the [Coronavirus](#) collection.

Corresponding author: Manoj Jena (manoj.20283@lpu.co.in)

Author roles: **Kumar V:** Conceptualization, Methodology, Writing – Original Draft Preparation; **Kancharla S:** Conceptualization, Methodology, Writing – Original Draft Preparation; **Kolli P:** Writing – Review & Editing; **Jena M:** Conceptualization, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2021 Kumar V *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Kumar V, Kancharla S, Kolli P and Jena M. **Reverse vaccinology approach towards the in-silico multiepitope vaccine development against SARS-CoV-2 [version 1; peer review: 2 approved]** F1000Research 2021, 10:44 <https://doi.org/10.12688/f1000research.36371.1>

First published: 23 Jan 2021, 10:44 <https://doi.org/10.12688/f1000research.36371.1>

Introduction

Coronaviruses (CoVs) belong to the family of coronaviridae in the order Nidovirales, and have single-strand positive-sense RNA¹. The size of the RNA of coronavirus is the largest among the viruses (~30 kb)². They have glycoprotein projections on the envelope, which gives the corona appearance. CoVs are pathogens mainly involved in respiratory and gastrointestinal diseases in a wide range of animals and humans^{1,2}. CoVs are divided into four sub-categories, namely alpha, beta, gamma, and delta, out of which alpha and beta coronavirus are known to infect humans^{1,3}. From alpha and beta, four strains are responsible for the common cold, and two strains were found to be responsible for severe acute respiratory syndrome (SARS-CoV) and Middle East respiratory syndrome CoV (MERS-CoV)^{4,5}. Recently in December 2019, a novel coronavirus SARS-CoV-2 was detected from patients of novel coronavirus disease 2019 (COVID-19) in the Wuhan, China⁶⁻⁸. The symptoms of COVID-19 infection include headache, fever, pneumonia, and asthenia^{9,10}. A severe and fatal outbreak of this virus has taken many lives and has created enormous economic loss worldwide. The treatment and prevention from this infection is the need of the hour.

COVID-19 has given a serious and tough challenge to biomedical research scientists and researchers all around the world. Much research is looking at repurposing antiviral drugs, and developing new drugs and vaccines against the SARS-CoV-2¹¹⁻¹⁵. Here, an attempt has been given to construct an in-silico vaccine, which can be further validated through experimental assays and could play a major role in the management of this pandemic. In this study, three structural proteins from SARS-CoV-2, based on the antigenicity has been selected for the construction of the vaccine. The first structural protein is Spike (S) glycoprotein, which has been reported to be a crucial surface protein of SARS-CoV-2, which facilitates the entry of the virus inside the host cell. It has been reported that for the entry of the SARS-CoV-2, S protein first binds with Angiotensin-Converting Enzyme-2 (ACE-2) receptor, and then is primed by the host serine protease (TMPRSS2)^{16,17}. This priming of the S protein lets it to fuse into the host cell membrane and entry inside the cell. The second crucial structural protein, which induces a strong immune response, is Membrane (M) glycoprotein. It plays a crucial role in virus morphogenesis and assembly by interacting with several other viral proteins¹⁸. The third structural protein chosen for vaccine construct is Nucleocapsid (N) phosphoprotein; it links the viral genome to the envelope. It consists of two domains, N terminal and C terminal, and both can bind to RNA. It has been reported that the C terminal domain facilitates the physical interaction of the RNA genome and envelope^{19,20}. All these three structural proteins are predicted to be good antigens and could induce the immune response.

In this pandemic situation, an immunoinformatics approach could be a fast, scientifically sound, and reliable option for quicker vaccine development. These three proteins chosen for the present study were predicted to be good antigens, which gives the opportunity to predict B and T cell epitopes. When naive B cells interact with the antigenic B cell epitopes via its transmembrane bound antibody, they differentiate into two types of cells plasma and memory cells^{21,22}. Plasma cells lack receptors, but they produce a large number of antibodies against the antigen. Memory cells express membrane-bound antibody molecules, but they are functionally inactive unless they encounter the same antigen again²³⁻²⁵. Furthermore, T cell epitopes are recognized by Major Histocompatibility Class (MHC), a glycoprotein present on the variety of the cells, which display the antigen to T cells^{26,27}. Antigen-presenting MHCs are divided into major two classes, MHC class-I are expressed on nucleated cells while MHC class-II are only expressed by antigen-presenting cells. Class II MHC interacts with T helper cells and activates B cells via cytokines, while Class-I MHC interacts with cytotoxic T cells, which kills virus-infected host cells^{28,29}. Further to know the ability of a constructed vaccine for inducing innate as well as antigen-specific acquired immunity, the constructed vaccine must dock with Toll-Like Receptor-3 (TLR-3). TLRs are mainly expressed on various leukocytes such as dendritic cells, natural killer cells, and cells of adaptive immunity such as T cells and B cells^{30,31}. Hence, in this study, an attempt has been made to construct the multiepitope vaccine consists of Helper T cells (HTLs), Cytotoxic T Cells (CTLs), and B cell epitopes, which could interact with TLR-3 and generate the immune response. This constructed multiepitope vaccine may induce both humoral as well as cell-mediated immune responses.

Methods

Retrieval of structural polyproteins of the SARS-CoV-2

The complete sequence of all three structural polyproteins from SARS-CoV-2 reference sequence (NC045512.2) were retrieved from NCBI on the basis of their antigenicity. The spike (S) glycoprotein (YP_009724390.1), Membrane (M) glycoprotein (YP_009724393.1) and Nucleocapsid (N) phosphoprotein (YP_009724397.2) were retrieved in FASTA format. These three proteins together are referred to as SMN (Spike, Membrane, and Nucleocapsid) polyprotein in this study.

Cytotoxic T cell epitopes (CTL) prediction

First, the CTL epitopes for SMN polyproteins were predicted using Netct1 1.2 server³². Prediction of the epitopes depends on three major attributes: (1) binding affinity of MHC-1 class; (2) ability of the proteasome cleavage; and (3) TAP

transport efficiency. The first two are predicted with the artificial neural network algorithm while third one using weight matrix. For the prediction of the epitopes threshold for epitopes, identification was chosen to be 0.75, weight on C terminal cleavage was set on 0.15, while the weight on TAP transport efficiency was set on 0.05. The predicted epitopes were ranked according to the combined score.

Helper T cell (HTL) epitope prediction

For the prediction of HTL epitopes, the [IEDB MHC II server](#) was used³³. The species/locus was selected as Human/HLA-DR, and a 7-allele HLA reference set was selected for the prediction. Further, 15 mer length of the epitopes were retrieved and ranked according to the percentile. The percentile rank is given after comparing the peptides score with five million 15 mers from the SWISSPROT database. The higher percentile value means a lower binding affinity of MHC-II. For further refinements of the HTL epitopes, these selected HTL epitopes were subjected to investigate whether they can induce IFN gamma immune response using the [IFN epitope server](#)³⁴. For IFN gamma inducing epitopes selection, the Motif/SVM hybrid approach was chosen, and the model was set to be IFN gamma versus non-IFN gamma. Finally, the epitopes whose results were positive for the IFN gamma response were chosen for the in-silico vaccine development.

B cell epitope prediction

B cell epitopes were predicted using the [ABCpred server](#)³⁵. This server predicts B cell epitopes using recurrent neural network algorithm. For the identification of the epitopes, the threshold was set on 0.51, while the window length for the prediction was chosen to be 16, keeping overlapping filter on. Top predicted epitopes having scored more than 0.9 was only chosen for the development of the candidate vaccine. Further, after the construction of the vaccine, linear as well as discontinuous conformational B cell epitopes were identified in the vaccine construct using [EliPro](#), an online server³⁶. EliPro predicts the antibody epitopes taking protein antigen tertiary structure as input.

Antigenicity, allergenicity, and toxicity prediction

The important attributes such as the antigenicity, allergenicity and toxicity were predicted for all the predicted epitopes individually as well as after construction of the vaccine. First of all, the antigenicity was investigated using the [VaxiJen 2.0 server](#)³⁷, and only probable antigen epitopes were chosen for the construction of the vaccine. Further, the allergenicity was predicted using the [AlgPred server](#)³⁸ and only non-allergenic epitopes were selected. Finally, all the epitopes were investigated for toxicity using the [ToxinPred server](#)³⁹ and non-toxic epitopes were selected. All the predicted epitopes had to cross all these barriers. The overall construct of the vaccine was also tested for these attributes.

Construction of multiepitope vaccine sequence

The vaccine sequence was constructed using the best identified CTL, HTL, and B cell epitopes. For the construction of the sequence at the N terminal and C terminal, an adjuvant was added using EAAAK linkers. While HTL epitopes were linked using GPGPG, linkers and CTL epitopes were linked using AAY linkers. In the C terminal, HHHHHH was added for the easy purification of the vaccine.

Physicochemical properties of the vaccine sequence

The physicochemical properties such as molecular weight, PI, half-life, aliphatic index, and hydrophobicity were predicted using online tool [ProtParam](#)⁴⁰.

Secondary structure of the vaccine sequence

Protein secondary structure prediction gives further opportunity to predict the tertiary structure as well as gives information about the activity and function of the protein. The secondary structure of the final multiepitope vaccine sequence was predicted by the free online web tool [CFSSP](#)⁴¹.

Tertiary structure prediction

The tertiary structure of the constructed vaccine was predicted using the [Rosetta web tool](#)⁴². Rosetta tool applies a deep neural network algorithm to predict the inter-residue distances as well as orientations. Then these orientations are converted to smooth inter-residue constraints followed by gradient descent energy minimization. Further, the coarse-grained models are generated, and full atom refinement is done. The validation of the model has been done through Ramachandran plot analysis using [VADAR web tool](#)⁴³. Further, the modelled structure was validated through the [ProSA](#) web tool⁴⁴, which gives the quality Z score of the modelled protein based on the already known similar size of the proteins crystal structures.

Minimization and equilibration of the predicted structure

To achieve a more stable structure, the predicted structure was further taken for molecular dynamics (MD) simulations using [Gromacs software](#)⁴⁵. The structure was minimized using the steepest descent algorithm with 50000 steps, followed

by NVT and NPT equilibration for 100 picoseconds, followed by MD simulations of 500 picoseconds. The last frame of the MD trajectory was taken for further analysis.

Molecular docking of constructed vaccine with TLR-3 receptor

For molecular docking, the last frame from the MD simulations of the constructed vaccine was taken, and the TLR-3 structure was retrieved from Protein Data Bank (PDB; ID 1ZIW). The downloaded structure was prepared and processed for docking using dock prep tool UCSF Chimera software. For the docking, the vaccine construct and TLR-3 was uploaded to patchDock server⁴⁶. Further, for refinement of the rigid body molecular docking solutions, FireDock server was used⁴⁷. It gave the best 10 docked confirmation based on global energy and Van der Waal's interactions.

Reverse translation and codon optimization

Finally, for expressing the constructed multipeptide, the vaccine needs to be expressed in the suitable vector inside the prokaryotic system. Hence, reverse translation and codon optimization were analysed using the Java codon adaptation web tool (Jcat)⁴⁸. The codon optimization was performed for *E. coli* strain K12 as a host. Jcat gives the codon adaption index (CAI) and percentage GC content as output. The CAI gives the information of codon usage, generally score between 1 and 0.8, while GC contents should be between 40 % to 70%, values lying outside the given margin is suggested to be inefficient⁴⁹.

Results

Retrieval of the polyproteins and antigenicity

The amino acid sequence of all three (SMN) structural proteins were retrieved from the NCBI database in fasta format. The proteins were investigated for antigenicity by Vaxijen web tool, and it was found that all the three chosen proteins could be good antigens. The default threshold of 0.4 was chosen as the criteria for the antigenicity in the Vaxijen tool. The spike protein showed a score of 0.46; membrane glycoprotein showed a score of 0.51; while nucleocapsid protein showed a score of 0.50. Hence, all three proteins were chosen for further predictions of B cell and T cell epitopes and the construction of the vaccine.

Prediction of CTL epitopes

CTL epitopes were predicted using Netc1 1.2 server for all the three selected proteins. A total of 38 epitopes was predicted from spike glycoprotein; 10 epitopes were predicted from membrane glycoprotein; while 9 were predicted from nucleocapsid protein. Out of all these predicted CTL epitopes, only 8 were selected for the construction of the vaccine, based on a high binding affinity towards MHC-I, antigenicity, non-allergenicity, and non-toxicity predictions, as shown in Table 1.

Prediction of HTL epitopes

HTL epitopes were predicted using the IEDB MHC II server for all the three SMN structural proteins. Finally, 4 HTL epitopes were selected on the basis of binding affinity, antigenicity, non-allergenicity, and non-toxicity, as shown in Table 2. Four human alleles and position of predicted epitopes are HLA-DRB1*07:01 (166-180), HLA-DRB4*01:01 (298-312), HLA-DRB5*01:01 (232-246), HLA-DRB5*01:01 (345-359).

Table 1. List of the final selected CTL epitopes which have fulfilled all the criteria for antigenicity, non-allergenicity, non-toxicity and could bind efficiently to MHC-I.

Peptide sequence	MHC binding affinity	Rescale binding affinity	C-terminal cleavage affinity	Transport efficiency	Prediction score	Polyprotein
WTAGAAAYY	0.6735	2.8596	0.7339	2.8630	3.1128	S protein
WMESEFRVY	0.3902	1.6569	0.7993	2.9290	1.9232	S protein
AGDSGFAAY	0.1341	0.5695	0.9652	2.6730	0.8480	M protein
LVGLMWLSY	0.2694	1.1440	0.7240	2.8970	1.3974	M protein
VATSRTLSTY	0.2752	1.1684	0.9679	3.0130	1.4642	M protein
LSPRWYFY	0.4837	2.0538	0.9746	2.8150	2.3408	N protein
DLSRWYFY	0.2866	1.2167	0.9760	2.7250	1.4994	N protein
NTASWFTAL	0.1772	0.7523	0.9557	1.1280	0.9521	N protein

Table 2. List of the final selected HTL epitopes which fulfilled all the criteria for antigenicity, non-allergenicity, non-toxicity and could also induce the IFN gamma immune response.

Allele	Start	End	Peptide sequence	Percentile score	Polyprotein
HLA-DRB5*01:01	232	246	GINITRFQTLALHR	0.52	S protein
HLA-DRB5*01:01	345	359	TRFASVYAWNKRKIS	0.52	S protein
HLA-DRB1*07:01	166	180	KEITVATSRTLSSYYK	2.1	M protein
HLA-DRB4*01:01	298	312	YKHWPQIAQFAPSAS	11	N protein

Table 3. Predicted linear B cell epitopes, binding score better than 0.9, are only selected for the final vaccine construct.

Peptide sequence	Predicted score	Polyprotein
AGTITSGWTFGAGAAL	0.97	S protein
GVSITPGTNTSNQVA	0.95	S protein
TRRIRGGDGKMKDLSP	0.94	N protein
KSAAEASKKPRQKRTA	0.93	N protein

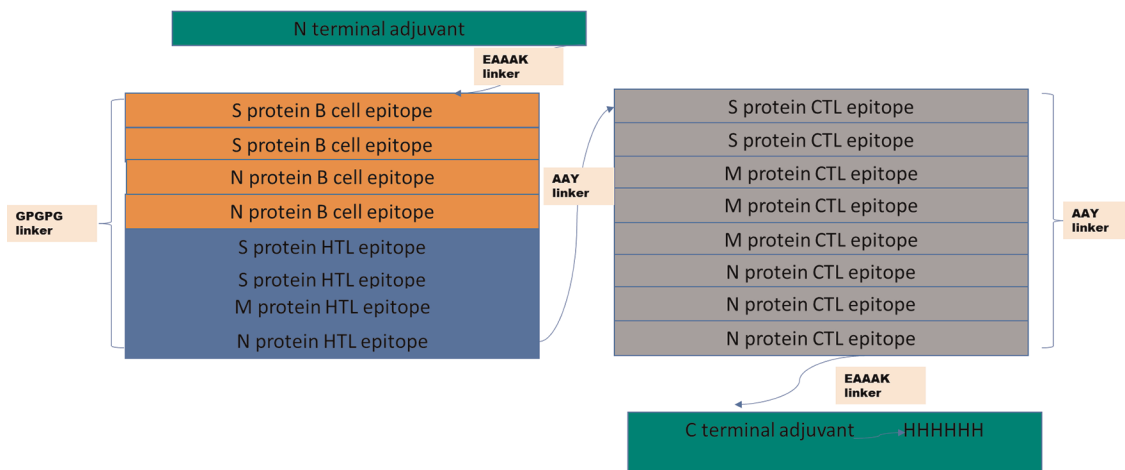


Figure 1. Structure of the final multi-epitope based vaccine. At the C terminal, an adjuvant human B defesin3 has been added, and then it is linked with B cell epitopes using EAAAK linkers. B cell epitopes are linked with HTL with GPGPG linkers, and HTL are linked with CTL with AAY linkers. At N terminal, another adjuvant human B defensin-2 has been linked with six histidine sequences.

B cell epitope prediction

For the prediction of B cell epitopes, ABCpred server was used. Based on the binding score (>0.9), non-allergenicity and non-toxicity, a total of four B cell epitopes were finally selected, as shown in **Table 3**.

Construction of multi-epitope based vaccine

The four B cell epitopes, four HTL epitopes and 8 CTL epitopes were selected for vaccine construction, which fulfilled all the criteria of binding affinity, antigenicity, non-toxicity and non-allergenicity. Besides these epitopes, two adjuvants were also added at the N terminal (human beta defensin-3) and at C terminal (human beta defensin-2) of the vaccine to increase the antigenicity. Adjuvant were linked via EAAAK linkers to the epitopes, HTL epitopes were linked via GPGPG linkers, while CTL epitopes were linked with AAY linkers, as shown in **Figure 1**. The constructed vaccine sequence was again checked for antigenicity, non-allergenicity, non-toxicity and it fulfilled all the criteria.

Prediction of physicochemical parameters of the constructed vaccine sequence

The physicochemical parameters of the vaccine sequence were predicted by the ProtParam server. The molecular weight of the construct was predicted to be 38.8 KDa, and the theoretical PI value was 9.92. The predicted half-life in *E. coli* was

more than 10 hours, and the instability index in the test tube was found to be stable. The aliphatic value of the vaccine sequence was 58.7 and the grand average of hydropathicity (GRAVY) was -0.348.

Secondary structure prediction of the vaccine sequence

Secondary structure prediction was made using the CFSSP web tool. The result showed the presence of helix at 44.5%, sheet at 35.6%, and turns at 14 %.

Tertiary structure prediction of the vaccine sequence

The 3D structure of the multi-epitope predicted vaccine was predicted using the Rosetta web tool. It uses de-novo structure prediction using deep neural network algorithm to predict the inter-residue distances as well as orientations. Then these orientations are converted to smooth inter-residue constraints followed by gradient descent energy minimization. Further, coarse-grained models are generated, and full atom refinement is done. It gave five best-predicted models, and based on the TM score, one model was selected for further investigation, as shown in **Figure 2A**. Further to validate the predicted model, Ramachandran plot analysis was done, and results showed that 96.3% residues were in the favourable region, 2.5% were in the allowed region while ~1% were in the outlier region (**Figure 3**). Additionally, the PROSA web tool was used to predict the quality of the modelled vaccine, which predicted a Z score of -6.34. Ramachandran plot and Z score have suggested that the predicted model of protein was valid and could be taken for further analysis.

Conformational B cell epitope analysis from modelled vaccine

EliPro predicts the antibody epitopes taking protein 3D structure as input. Linear B, as well as discontinuous conformational epitopes, were identified in the vaccine construct using ElliPro, an online server. A total of 8 linear epitopes were predicted, and the sequence of the top 3 epitopes have been reported in **Table 4** and has been shown structurally in **Figure 2B**. Various discontinuous epitope residues were predicted from vaccine sequence length 232-253 (21 epitope residues), between 299-357 (55 epitope residues), between 1-54 (52 epitope residues), between 69-128 (33 epitope residues) and between 168-176 (9 epitope residues) were predicted. The individual score of each of the discontinuous epitopes has been shown in **Figure 4**.

Docking of vaccine with TLR-3 receptor

The modelled structure of the vaccine was taken through energy minimization, equilibration, and MD simulations before docking. The last frame from the simulated trajectory was taken further for docking. The simulated structure has been compared with the crude modelled structure, as shown in **Figure 2A**. The TLR-3 structure was retrieved from PDB (ID 1ZIW). The downloaded structure was prepared and processed for docking using the dock prep tool UCSF Chimera software⁵⁰. The simulation was done using the PatchDock server and further refined using FireDock. The best-docked complex had global energy of -14.91 Kcal/mol, and attractive Van der Waal's energy was -18.1 Kcal/mol, which shows a decent binding affinity of the vaccine towards TLR-3. Further, the best binding pose was investigated for polar interactions using discovery studio visualizer⁵¹ between TLR-3 and vaccine, and it was found that GLN352, SER428, ILE370 of TLR-3 was making the hydrogen bond with TYR260, ARG321, and LYS166 of vaccine respectively (**Figure 5**).

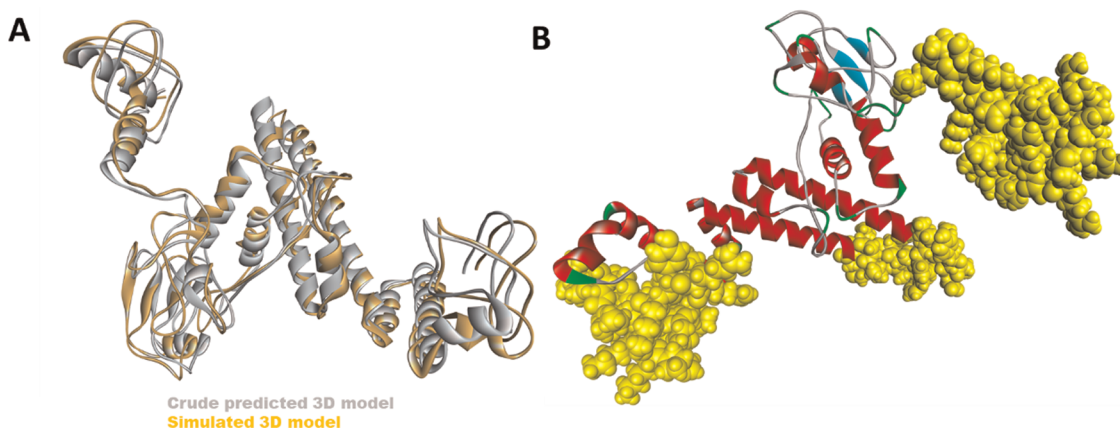


Figure 2. (A) The crude 3D modelled structure of the vaccine (grey) has been superimposed with the simulated model (yellow). (B) The top 3 conformational B cell epitopes predicted in the vaccine has been shown with yellow spheres.

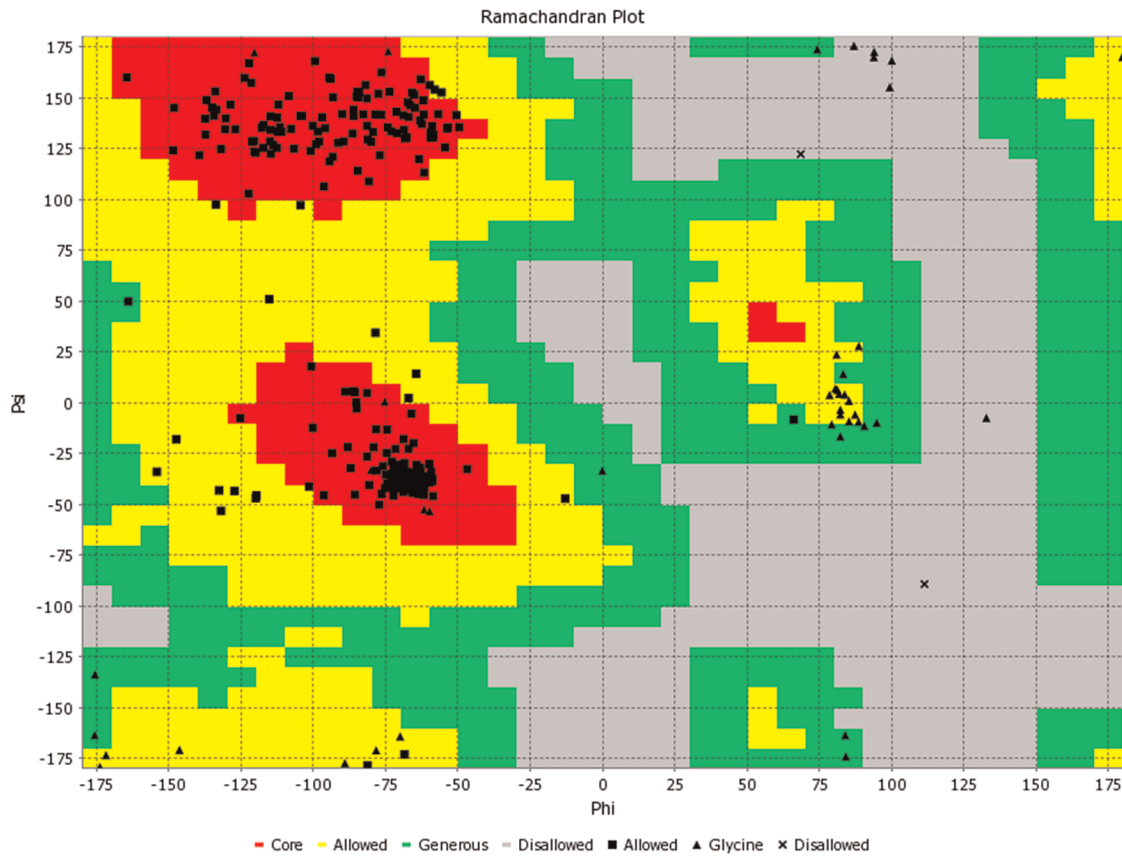


Figure 3. Ramachandran plot of the 3D modelled vaccine construct.

Table 4. Predicted top three conformational B cell epitopes from the 3D modelled vaccine construct.

Start	End	Peptide sequence	score
329	357	FCPRRYKQIGTCGLPGTKCCKKPHHHHHH	0.78
232	253	VYAAYAGDSGFAAYAAYLVGLM	0.74
1	54	GIINTLQKYICRVRGGRAVLSCLPKEEQIGKCTRGRKCCRRKKEAAKAGTI	0.73

Reverse translation and codon optimization

Jcat was used for the optimization of the codon for the proper expression of the protein. *E. coli* strain K12 was chosen as a host, with additional options such as avoid rho-independent transcription terminators, avoid prokaryotic ribosome binding sites, and avoid Cleavage Sites of Restriction Enzymes. The CAI gives the information of codon usage, generally score between 1 and 0.8, while GC contents should be between 40 % to 70%, values lie outside the given margin is suggested to be inefficient. CAI of the optimized nucleotide sequence of the vaccine was found to be 0.92, with a GC content of 55.6%, which indicates the effective expression of the protein in the *E. coli*.

Discussion

The structural polyproteins from SARS-CoV-2 were selected for developing a multiepitope based vaccine. Initially, four structural proteins were chosen, namely spike, membrane, envelope, and nucleocapsid protein, based on their antigenicity prediction, role in facilitating the entry of the virus, and packaging inside the host cells. However, when CTL, HTL, and B cell epitopes were predicted, it was found that epitopes from envelope protein were not able to satisfy the criteria of non-allergenicity and antigenicity simultaneously, and hence envelope protein was not considered further in the study. While epitopes selected from SMN structural polyprotein were satisfying, all the criteria such as antigenicity, non-allergenicity, non-toxicity, and high binding affinity towards MHC and also HTL epitopes were able to induce IFN gamma immune response. The constructed multiepitope vaccine from the selected epitopes from all the three SMN polyproteins was again

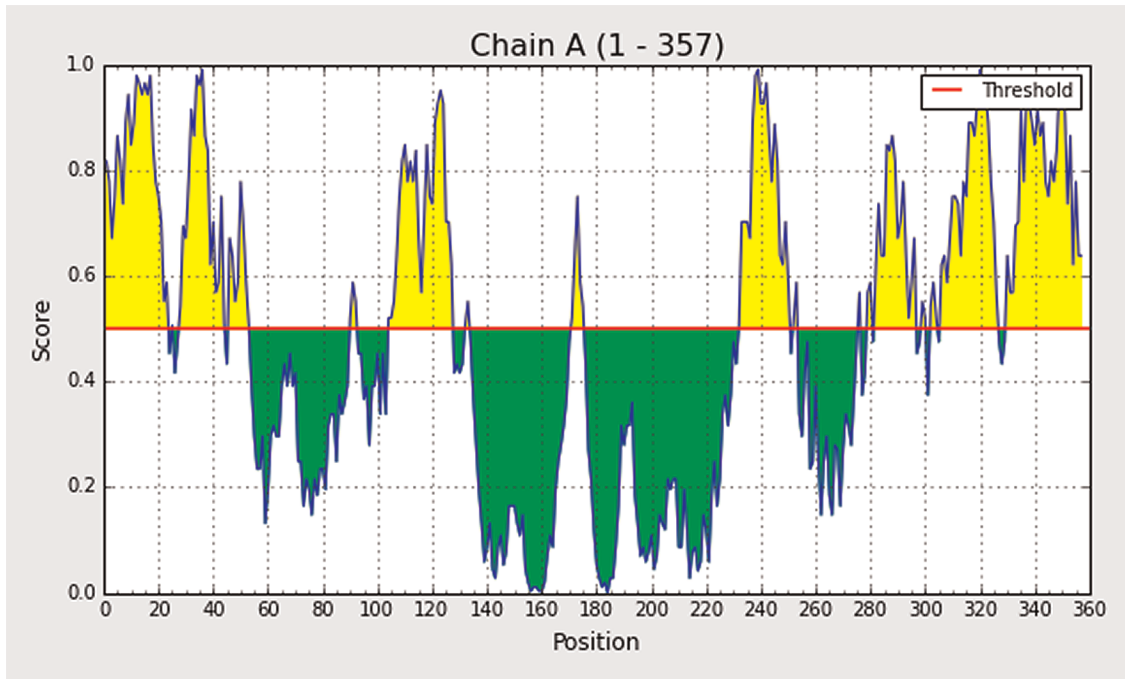


Figure 4. The individual score of discontinuous B cell epitopes predicted in the modelled vaccine

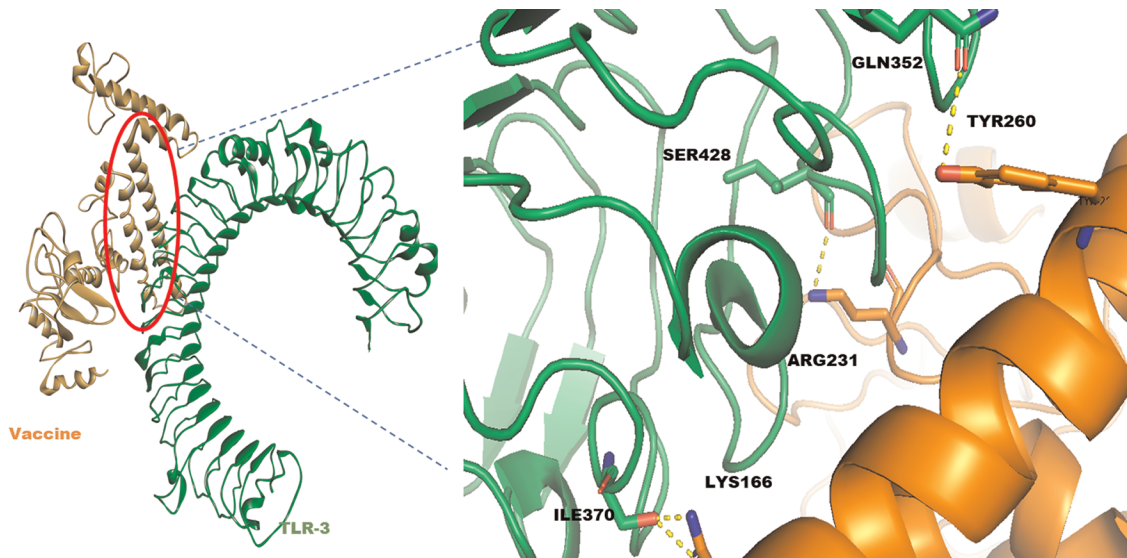


Figure 5. The docked structure of TLR-3 (green)-modelled vaccine (orange) complex illustrating critical residues involved in the interactions.

investigated for antigenicity, and it was found that vaccine construct predicted to be a potent antigen with score 0.60 (predicted by Vaxijen). Further, the vaccine construct was looked for its allergenicity, and it was found that the vaccine was a non-allergen with a score of -0.59 (threshold was set on -0.4, predicted by AlgPred tool). Further physicochemical parameters were analysed for vaccine sequence, and it was predicted to have a molecular weight of 38.8 KDa, PI of 9.92, and half-life inside the *E. coli* >10 hours, which shows that protein can easily express and isolated. The 3D model of the constructed vaccine sequence was predicted from the Rosetta web server. Eight conformational linear B cell epitopes were found in the modelled structure of the vaccine, as predicted by the Elipro web tool. It was evident from the prediction that the constructed vaccine model could easily produce adaptive immune response specific to the SARS-CoV-2 antigens. Further, to investigate the ability of the modelled vaccine to interact with TLR receptors on immune cells,

the TLR-3 receptor was docked with the modelled vaccine. The results showed that the modelled vaccine had a good binding affinity towards TLR-3, and it was found that GLN352, SER428, ILE370 of TLR-3 was making the hydrogen bond with TYR260, ARG321, and LYS166 of vaccine, respectively. This interaction of vaccine with TLR-3 was predicting that vaccine have the potential to generate both innate as well as adaptive humoral and cell-mediated immune responses. For efficient expression of the protein inside the *E. coli*, codon optimization was done to improve the translation and transcription efficiency. The constructed vaccine sequence was reverse translated, and CAI and GC content were assessed, taking *E. coli* (K12) as a host organism. The CAI index of 0.92 and GC content of 55.6 % and half-life was already predicted to be more than 10 hours, suggests the efficient expression of recombinant protein inside the *E. coli*. This immunoinformatics study suggests that the predicted vaccine could generate specific adaptive immunity against SARS-CoV-2 and could provide a valuable contribution to the management of the COVID-19. This predicted vaccine candidate strongly warrant in-vitro and in-vivo study for the practical implications.

Conclusions

In this study, a multiepitope (CTL, HTL, and B cell) vaccine construct has been predicted and modelled through immunoinformatics techniques. The predictions suggest that the constructed vaccine could generate both humoral and cell-based adaptive immunity towards SARS-CoV-2. Further, it was also predicted that it may easily be expressed inside the *E. coli* strain (K12). This immunoinformatics study may reduce the expenditure and time for vaccine research and may give a significant value in the management of COVID-19. This in silico prediction warrants the in-vitro and in-vivo study to test the practical implications of the predicted vaccine.

Data availability

Underlying data

NCBI: Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome, Accession number NC045512.2: https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2/

NCIB: surface glycoprotein [Severe acute respiratory syndrome coronavirus 2], Accession number YP_009724390.1: https://www.ncbi.nlm.nih.gov/protein/YP_009724390.1/

NCBI: membrane glycoprotein [Severe acute respiratory syndrome coronavirus 2], Accession number YP_009724393.1: https://www.ncbi.nlm.nih.gov/protein/YP_009724393.1

NCBI: nucleocapsid phosphoprotein [Severe acute respiratory syndrome coronavirus 2], Accession number YP_009724397.2: https://www.ncbi.nlm.nih.gov/protein/YP_009724397.2

Protein Data Bank: Human Toll-like Receptor 3 extracellular domain structure, Accession number 1ZIW: <http://doi.org/10.2210/pdb1ZIW/pdb>

Acknowledgements

A previous version of this article is available on Research Square: <https://doi.org/10.21203/rs.3.rs-31779/v1>

References

- Denison MR, *et al.*: **Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity.** *RNA Biol* 2011; **8**(2): 270–9. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fehr AR, Perlman S: **Coronaviruses: an overview of their replication and pathogenesis.** *Methods Mol Biol* 2015; **1282**: 1–23. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ye ZW, *et al.*: **Zoonotic origins of human coronaviruses.** *Int J Biol Sci* 2020; **16**(10): 1686–97. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bhattacharya S, *et al.*: **Recent apprise on coronavirus and its terrible insinuations.** *Virusdisease* 1–7, 2020. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vabret A, *et al.*: **Human coronavirus NL63, France.** *Emerg Infect Dis* 2005; **11**(8): 1225–9. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li Q, *et al.*: **Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia.** *N Engl J Med* 2020. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hui DS, *et al.*: **The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China.** *Int J Infect Dis* 2020; **91**: 264–6. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Huang C, *et al.*: **Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.** *Lancet* 2020. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lechien JR, *et al.*: **Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study.** *Eur Arch Otorhinolaryngol* 2020; 1–11. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

10. Nicastrì E, *et al.*: **National institute for the infectious diseases "L. Spallanzani", IRCCS. Recommendations for COVID-19 clinical management.** *Infect Dis Rep* 2020; **12**(1).
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Sindhu T, *et al.*: **Antiviral screening of Clerodol derivatives as COVID-2 main protease inhibitor in Novel Corona Virus Disease: In silico approaches.** *Asian J Pharm Clin Res* 2020; **10**(2): 60–4.
[Publisher Full Text](#)
12. Yadav AR, Mohite SK: **A Novel Approach for Treatment of COVID-19 with Convalescent Plasma.** *RJPDT* 2020; **12**(3): 227–30.
[Publisher Full Text](#)
13. Shankhdhar PK, *et al.*: **Turmeric: Plant Immunobooster against COVID-19.** *Res Rev J Pharmacogn Phytochem* 2020; **12**(3): 174–7.
[Publisher Full Text](#)
14. Goswami S, *et al.*: **A Meticulous Interpretation on a Sanguinary Disease COVID-19.** *Res Rev J Pharmacogn Phytochem* 2020; **12**(3): 231–3.
[Publisher Full Text](#)
15. Naresh B: **A Review of the 2019 Novel Coronavirus (COVID-19) Pandemic.** *Asian J Pharm Sci* 2020; **10**(3): 233–8.
[Publisher Full Text](#)
16. Hoffmann M, *et al.*: **SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor.** *Cell* 2020.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
17. Iwata-Yoshikawa N, *et al.*: **TMPS22 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection.** *J Virol* 2019; **93**(6).
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Prajapat M, *et al.*: **Drug targets for corona virus: A systematic review.** *Indian J Pharmacol* 2020; **52**(1): 56–65.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Brunel J, *et al.*: **Sequence of events in measles virus replication: role of phosphoprotein-nucleocapsid interactions.** *J Virol* 2014; **88**(18): 10851–63.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Surjit M, Lal SK: **The SARS-CoV nucleocapsid protein: a protein with multifarious activities.** *Infect Genet Evol* 2008; **8**(4): 397–405.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Nemazee D: **Mechanisms of central tolerance for B cells.** *Nat Rev Immunol* 2017; **17**(5): 281.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Oleinika K, *et al.*: **B Cell Activation and B Cell Tolerance.** *The Autoimmune Diseases: Elsevier* 2020; 171–87.
[Publisher Full Text](#)
23. Schmitt ME, *et al.*: **The B cell antigen receptor of IgE-switched plasma cells regulates memory IgE responses.** *J Allergy Clin Immunol* 2020.
[PubMed Abstract](#) | [Publisher Full Text](#)
24. Gold MR, Reth MG: **Antigen receptor function in the context of the nanoscale organization of the B cell membrane.** *Annu Rev Immunol* 2019; **37**: 97–123.
[PubMed Abstract](#) | [Publisher Full Text](#)
25. Brink R, Phan TG: **Self-reactive B cells in the germinal center reaction.** *Annu Rev Immunol* 2018; **36**: 339–57.
[PubMed Abstract](#) | [Publisher Full Text](#)
26. Karunakaran KP, *et al.*: **Discordance in the Epithelial Cell-Dendritic Cell Major Histocompatibility Complex Class II Immunoproteome: Implications for Chlamydia Vaccine Development.** *J Infect Dis* 2020; **221**(5): 841–50.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Peters B, *et al.*: **T Cell Epitope Predictions.** *Annu Rev Immunol* 2020; **38**.
[Publisher Full Text](#)
28. Gaudino SJ, Kumar P: **Cross-talk between antigen presenting cells and T cells impacts intestinal homeostasis, bacterial infections, and tumorigenesis.** *Front Immunol* 2019; **10**.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Miller CT: **The Expression Of Connexin-43 By CD11c+ Dendritic Cells Is Required to Maintain CD4+ Foxp3+ Regulatory T Cell Population in Peripheral Lymphoid Organs.** 2019.
[Publisher Full Text](#)
30. Rich RR, Chaplin DD: **The human immune response.** *Clinical Immunology: Elsevier* 2019; **e1**: 3–17.
[Publisher Full Text](#)
31. Toubi E, Vadasz Z: **Innate immune-responses and their role in driving autoimmunity.** *Autoimmun Rev* 2019; **18**(3): 306–11.
[PubMed Abstract](#) | [Publisher Full Text](#)
32. Larsen MV, *et al.*: **Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction.** *BMC Bioinformatics* 2007; **8**: 424.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Zhang Q, *et al.*: **Immune epitope database analysis resource (IEDB-AR).** *Nucleic Acids Res* 2008; **36**(Web Server issue): W513–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Dhanda SK, *et al.*: **Designing of interferon-gamma inducing MHC class-II binders.** *Biol Direct* 2013; **8**: 30.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Saha S, Raghava GP: **Prediction of continuous B-cell epitopes in an antigen using recurrent neural network.** *Proteins* 2006; **65**(1): 40–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
36. Ponomarenko J, *et al.*: **EliPro: a new structure-based tool for the prediction of antibody epitopes.** *BMC Bioinformatics* 2008; **9**: 514.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
37. Doytchinova IA, Flower DR: **Vaxijen: a server for prediction of protective antigens, tumour antigens and subunit vaccines.** *BMC Bioinformatics*. 2007; **8**: 4.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
38. Saha S, Raghava GP: **AlgPred: prediction of allergenic proteins and mapping of IgE epitopes.** *Nucleic Acids Res* 2006; **34**(Web Server): W202–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Gupta S, *et al.*: **In silico approach for predicting toxicity of peptides and proteins.** *PLoS One* 2013; **8**(9): e73957.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
40. Garg VK, *et al.*: **MFPPI - Multi FASTA ProtParam Interface.** *Bioinformatics* 2016; **12**(2): 74–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
41. Kumar TA: **CFSSP: Chou and Fasman secondary structure prediction server.** *Wide Spectrum* 2013; **1**(9): 15–9.
42. Yang J, *et al.*: **Improved protein structure prediction using predicted interresidue orientations.** *Proc Natl Acad Sci U S A* 2020; **117**(3): 1496–503.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
43. Willard L, *et al.*: **VADAR: a web server for quantitative evaluation of protein structure quality.** *Nucleic Acids Res* 2003; **31**(13): 3316–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
44. 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**(Web Server): W407–10.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
45. Hess B, *et al.*: **GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation.** *J Chem Theory Comput* 2008; **4**(3): 435–47.
[PubMed Abstract](#) | [Publisher Full Text](#)
46. Schneidman-Duhovny D, *et al.*: **PatchDock and SymmDock: servers for rigid and symmetric docking.** *Nucleic Acids Res* 2005; **33**(Web Server): W363–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
47. Mashiach E, *et al.*: **FireDock: a web server for fast interaction refinement in molecular docking.** *Nucleic Acids Res* 2008; **36**(Web Server): W229–32.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Grote A, *et al.*: **JCat: a novel tool to adapt codon usage of a target gene to its potential expression host.** *Nucleic Acids Res* 2005; **33**(Web Server): W526–31.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
49. Das S, *et al.*: **Comparative Analysis of Predicted Gene Expression among Crenarchaeal Genomes.** *Genomics Inform* 2017; **15**(1): 38–47.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
50. Goddard TD, *et al.*: **Software extensions to UCSF chimera for interactive visualization of large molecular assemblies.** *Structure* 2005; **13**(3): 473–82.
[PubMed Abstract](#) | [Publisher Full Text](#)
51. BIOVIA DS: **Discovery Studio Modeling Environment.** 2020.

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 17 February 2021

<https://doi.org/10.5256/f1000research.39433.r77969>

© 2021 Settypalli T. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Tirumala Bharani K. Settypalli 

Department of Nuclear Sciences and Applications, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, International Atomic Energy Agency, Vienna, Austria

The current research article by Kumar *et al.*, utilized *in silico* Reverse vaccinology approach to develop multiepitope vaccine against SARS-CoV-2. Overall, the manuscript was well written and is a valuable contribution at this time of the pandemic.

Given the pandemic situation, the *in silico* approach is a reliable and fast option for quicker vaccine development. The authors have utilized various immunoinformatic online tools to predict T-cell and B-cell epitopes from three structural polyproteins and constructed vaccine sequence.

Although the current study lacks the *in vitro* / *in vivo* studies, the immunoinformatic approach is helpful to reduce the cost and time for further experimental validation and can be a valuable contribution during the pandemic.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.**Reviewer Expertise:** Molecular genetics, Immunology**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 09 February 2021

<https://doi.org/10.5256/f1000research.39433.r77965>

© 2021 Natarajan S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Sivaraman Natarajan** 

Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA

Kumar et al. in the manuscript attempted in-silico vaccine development against SARS-CoV-2 using publicly available SARS-CoV2-Sequence. They used three different crucial structural proteins (glycoproteins) from Helper T cell, Cytotoxic T cell, and B cell epitope. The epitope was predicted using publicly available tools. After predicting the epitope, they attempted an in-silico vaccine construction by applying various conditions including adding linked adjuvants to predict the efficiency of the vaccine. This immunoinformatics approach is very interesting given the SARS-CoV-2 evolution of new mutant strains, help to predict effective vaccine *in-silico*.

As the author mentioned in the manuscript it lacks any *invitro* experiment to confirm any of the predictions as proposed in the manuscript. However, the in-silico finding from the study will be immensely beneficial to the researcher who studies the SARS-Co-V2.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, Single-cell Technology, and Cancer Biology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research