In silico Design and Characterization of Multi-epitopes Vaccine for SARS-CoV2 from Its Spike Protein

Gunderao H. Kathwate[1]

Abstract

COVID 19 is a disease caused by a novel coronavirus, SARS-CoV2 originated in China most probably of Bat origin. Multiepitopes vaccine would be useful in eliminating SARS-CoV2 infections as asymptomatic patients are in large numbers. In response to this, we utilized bioinformatic tools to develop an efficient vaccine candidate against SARS-CoV2. The designed vaccine has effective BCR and TCR epitopes screened from the sequence of S-protein of SARS-CoV2. Predicted BCR and TCR epitopes found antigenic, non-toxic and probably non-allergen. Modeled and the refined tertiary structure predicted as valid for further use. Protein–Protein interaction prediction of TLR2/4 and designed vaccine indicates promising binding. The designed multiepitope vaccine has induced cell-mediated and humoral immunity along with increased interferon-gamma response. Macrophages and dendritic cells were also found to increase upon the vaccine exposure. In silico codon optimization and cloning in expression vector indicates that the vaccine can be efficiently expressed in *E. coli*. In conclusion, the predicted vaccine is a good antigen, probable no allergen, and has the potential to induce cellular and humoral immunity.

Keywords COVID19 · SARS-CoV2 · B cell epitopes · T cell epitopes · Vaccine · Immune-simulation

Introduction

In December 2019, a group of patients from Wuhan city of China was found to have pneumonia-like symptoms and diagnosed with the infection of beta coronavirus. Soon the infection spread across China and later spread all over the globe. These infections were named COVID 19 disease and the virus as SARS-CoV2 by WHO (Sohrabi et al. 2020). On 30 January 2020 due to its spread in more than 120 countries, WHO declared the COVID 19 pandemic as a public health emergency of international concern. The Novelty of SARS-CoV2 is its rapid spread may be due to asymptomatic patients (Gandhi et al. 2020; Bi et al. 2020) and highly sophisticated, time-consuming diagnostic methods (Chan et al. 2020; Yelin et al. 2020, Li and Xia 2020). In the report published by WHO on the spread of COVID 19 infection, 4 million new cases were emerged with a cumulative number of 244 million cases worldwide and 4.6 million deaths since from COVID 19 pandemic declaration (WHO 2020).

In India, due to lockdown positive cases are in control and spread is also marginal. However, still after two years from the pandemic declaration, SAR CoV 2 succumbing cases are increasing due emergence of new more transmissible variants of SARS-CoV2. Till September 16, 2021, a total of 33.35 million cumulative cases with 443,928 deaths were reported by the ministry of health and family welfare, Govt. of India (https://www.mohfw.gov.in/). To date, no promising antiviral drug is available to combat SARS-CoV2 infections. There are few drug candidates in clinical trials, but the process of drug development is slow and time-consuming (Mitjà and Clotet 2020). In South Korean clinical practices, Lopinavir/Ritonavir combination was found significantly effective in lowering the viral load to no detectible or little SARCoV2 titer (Wen et al. 2020). But another research group showed that the same combination was ineffective beyond standard care (Cao et al. 2020). Another drug pair, hydroxychloroquine, an antimalarial drug, and azithromycin reported having an association in the reduction of viral load (Gautret et al. 2020a, b). Nevertheless, QT interval prolongation may cause life-threatening arrhythmia (Chorin et al. 2020). In a large scale study both drugs and drugs alone compared with neither drug was associated with mortality (Rosenberg et al. 2020; Molina et al. 2020). In a drug repurposing study,

✉ Gunderao H. Kathwate
  santoshkathwate@gmail.com

[1] Department of Biotechnology, Savitribai Phule Pune University, Pune, Maharashtra, India

remdesivir, lopinavir, emetine, and homoharringtonine have significantly inhibited the replication of SAR-CoV2 (Choy et al. 2020). But this was in vitro study, and preliminary clinical data suggest no relation with mortality. Moreover, a multicentric study conducted for remdesvir showed no role in lowering the viral load and failed to provide clinical benefits in the patients with Severe COVID (Wang et al. 2020). Prevention is a better option than treatments to break the chain of SARS-CoV2 infections. Several shreds of evidence suggest and support the importance of vaccines to eliminate the COVID19 pandemic (Chen et al. 2020; Bai et al. 2020; Prem et al. 2020). Various epidemiological surveys proved that naturally acquired immunity could decrease the SARS-CoV2 titer (Shi et al. 2020; Prompetchara et al. 2020). More than 80% of patients develop mild and 14% develop severe symptoms of COVID19 with zero case fatality rate (Spychalski et al. 2020). Data published on efficacy of various vaccines released in the market have shown promising hope in the controlling spread of SAR-CoV2 infections (Polack et al. 2020; Olliaro et al. 2021). However, the evolution of new SAR-CoV2 variants like delta and omicron can evade the immunity and lower effectiveness of the vaccines (Lopez Bernal et al. 2021; Planas et al. 2021). On these circumstances, it is indispensable to develop a vaccine that will be effective against various variants of SAR-CoV2..

A highly effective vaccine must have common and more than one immunogenic region from the strains that will raise the various immunological attacks against different variants of a virus. The bioinformatics approach can accelerate the process by predicting common potential candidate peptides. Such peptides can be tailored together with adjuvant protein to enhance the immunological response. Vaccines against MERS, Ebola, and human papillomavirus were designed and successfully developed using bioinformatics approaches (Shey et al. 2019).

SARS-CoV2 belongs to the beta-coronoviridae family including four endemic viral pathogens viz HCoV-HKU1, HCoV-OC43, HCoV-229E, HCoV-NL63, two epidemic viruses like the Middle East Respiratory Syndrome virus (MERS), and SARS (Osman et al. 2020). This virus is a non-segmented positive-strand RNA virus with an envelope. Envelop proteins are categorized into structural, non-structural, and accessory proteins. Structural proteins are involved in protection and bind to the host. Several proteins of SAR-CoV2 are essential for virulence and pathogenesis. For example, Nucleocapsid (N) protein N, is essential for RNA binding and its replication and transcription (Snijder et al. 2016). Envelope (E) and membrane (M) proteins are instrumental in virus assembly and virulence promotion (Ruch and Machamer 2012; Alsaadi et al. 2019). In addition, E and M proteins effectively induce an immune response (Ahmed et al. 2020). Spike (S) protein, a structural protein is responsible for binding to a receptor on the host cell called as Angiotensin-converting enzyme 2 (ACE2) (Hoffmann et al. 2020). S-Protein undergoes proteolytic cleavage by TMPRSS2 that allows subsequent entry of the virus into cells through endocytosis (Glowacka et al. 2011). S- protein was also found to be involved in the activation of T cell response (Li et al. 2020; Grifoni et al. 2020). Entire S protein expressed in chimpanzee adeno 38 (ChAd)-vector showed protection against SARS-CoV2 in mice and *rhesus* macaques (van Doremalen et al. 2020). Single-dose of ChAdOx1 nCoV-19 vaccine is sufficient to elicit a humoral and cell-mediated immune response in both the animals. Viral load also was found to be reduced compared to control animals and symptoms of pneumonia were absent. Here, in this communication a multi-epitopes vaccine is designed from S Protein of SAR-CoV2. This vaccine has ideal properties like good stability at room temperature, immunogenic, antigenic and non-allergen in nature. All the epitopes are effective in the stimulation of humeral as well and cell-mediated immunity.

Methods

Selection of Protein for Epitope Prediction

Complete Spike protein sequence (P0DTC2) of SARS-CoV2 was downloaded in fasta format from UniProt protein database. This sequence was used for further analysis and obtaining potential epitopes for B and T cell receptors.

T Cell Epitope Prediction

Various tools were used to predict the epitopes presented to T cell receptors. Potential epitopes of various bacteria and viruses were predicted by using such online epitopes predicting tools. The following tools were used for the prediction of TCR (T cell Receptor) epitopes IEDB MHC-I processing predictions, MHC-NP, netCTLpan1.1, RANK-PEP, and netMHCpan4.0. All these tools were used to screen the potential peptides as epitopes for TCR. These tools have various methods to predict epitopes. IEDB (Instructor/Evaluator Database) is a web-based epitope analysis resource that includes tools for T cell epitope prediction, B cell epitope prediction, and other analysis tools like epitope conservancy, etc. (Fleri et al. 2017; Dhanda et al. 2019). This resource uses methods based on artificial neural network (ANN), stabilized matrix method (SMM), and Combinatorial Peptide Libraries (CombLib), predict the peptides' way that processed naturally and presented by MHC I (Calis et al. 2013).

Analysis of Immunological Properties

The properties like immunogenicity, toxicity, allergen, and antigenicity were analyzed to predicted the TCR epitopes. IEDB MHC class I immunogenicity server and conservancy tool were used for the determination of immunogenicity and conservancy. The immunogenicity of a peptide MHC complex (http://tools.immuneepitope.org/immunogenicity/) was assessed keeping all the parameters at default. Protein sequence variants used setting sequence identity 100% and other parameters default (Calis et al. 2013). ToxiPred (http://www.imtech.res.in/raghava/toxinpred/index.html) a online tool predicted the toxicity considering the physicochemical properties of selected peptides (Mishra et al. 2014). Online server AllergenFP v.1.0 (http://ddgpharmfac.net/AllergenFP/) was used to predict peptides as allergens (Dimitrov et al. 2014a, b). The epitopes' antigenicity was analyzed by online server VaxiJen v2.0 (http://www.ddg-harmfac.net/vaxijen/VaxiJen/VaxiJen.html)(Doytchinova and Flower 2007b). The threshold value was set to 0.5. It is alignment- independent predictor based on auto-cross covariance (ACC) transformation epitopes sequences into uniform vectors of principle amino acid properties. The accuracy of this server varies between 70% and 89% depending on the targeted organism.

Linear B Cell Receptor Epitope Prediction

Six different methods were used for the prediction of B cell receptor (BCR) epitopes. All these methods generate fragments of the protein. To predict the B cell epitopes, it is necessary to find a linear sequence of B cell epitopes in the protein sequence. BepiPred linear epitope prediction server (http://www.cbs.dtu.dk/services/BepiPred/) uses a hidden Markov model and propensity scale method (Jespersen et al. 2017). Similarly, other properties were also being considered to predict good B cell epitopes. Different methods calculate various properties at the IEDB server (http://tools.iedb.org/bcell/). Kolaskar–Tongaonkar antigenicity scale provides physiology of the amino acid residues (Kolaskar and Tongaonkar 1990). Emini Surface accessible score for the accessible surface of the epitope (Emini et al. 1985). Secondary structure of epitopes also has a role in antigenicity. Karplus-Schulz flexibility score (Karplus and Schulz 1985) and Chou-Fasman β turn methods were used for flexibility and β turns prediction (Garnier 1978) respectively. Parker hydrophilicity prediction method was used for the determination of in silico hydrophilicity of peptides (Parker et al. 1986).

Engineering of Multi-epitope Vaccine Sequence

High-scored and common peptides predicted by various tools were selected for the deriving sequence of the potential vaccine candidate. GPGPG and AAY linked together with selected epitopes for T and B cell receptors. For enhanced immunogenicity, OmpA (GenBank: AFS89615.1) protein was chosen as an adjuvant and was linked through EAAAK at the N-terminal site of the vaccine.

Prediction of Immunogenic Properties of Designed Vaccine

The antigenicity of the chimeric vaccine was predicted by VaxiJen v2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) and ANTIGENpro online tools. VaxiJen 2.0 is an alignment-free antigenicity prediction tool that utilizes auto cross-covariance transformation of protein sequence into uniform vector vectors of principal amino acid properties (Doytchinova and Flower 2007b). ANTIGENpro (http://scratch.proteomics.ics.uci.edu/) is also an online tool that utilizes a protein microarray dataset to predict the antigenicity (Cheng et al. 2005). Based on the cross-validation experiment accuracy of this server is estimated to be around 76% using a combined dataset. AllerTop v2.0 and AllergenFP were two online tools utilized for the prediction of allergenicity of chimeric protein. Amino acid E-descriptors, auto- and cross-covariance transformation, and the k nearest neighbors (kNN) machine learning methods are the basis of AllerTop v2.0 (http://www.ddg-pharmfac.net/AllerTOP) for allergenicity prediction (Dimitrov et al. 2014a). AllergenFP is a descriptor-based fingerprint, alignment-free tool for allergenicity prediction (Dimitrov et al. 2014a, b). The tool uses a four-step algorithm. In the first step, proteins were analyzed based on their properties like hydrophobicity, size, secondary structures formation, and relative abundance. In the second step, generated strings were converted into vectors of equal length by ACC. Then, in the third step, vectors were converted into binary fingerprints and compared in terms of the Tanimoto coefficient. In last step, applying this approach to known allergen and non-allergen can identify 88% of allergen/non-allergen with Mathew's Correlation coefficient of 0.759.

Prediction of Solubility and Physiochemical Properties

For Solubility prediction of multi-epitopes chimeric vaccine, PROSO II server (http://mbiljj45.bio.med.uni-muenchen.de:8888/prosoII/prosoII.seam) was utilized. PROSO II server works on an approach of the classifier which utilizes the difference between TargetDB and PDB and insoluble proteins of TargetDB (Smialowski et al. 2012). The accuracy of this server is 71% at the default threshold of 0.6. Protoparam, the online webserver, was exploited for the evaluation of

physicochemical properties. The properties like amino acid composition, theoretical pI, instability index, in vitro and in vivo half-life, aliphatic index, molecular weight, and grand average of hydropathicity (GRAVY) were evaluated.

Secondary and Tertiary Structure Prediction

PSIPRED and RaptorX Property online servers were used to determine the secondary structure of the predicted vaccine. PSIPRED is a publicly available web server, includes two feed-forward neural networks works on output obtained from PSI-BLAST (McGuffin et al. 2000). PSIPRED 3.2 attains an average Q3 score of 81.6% obtained using very stringent cross approval strategies to assess its performance. RaptorX property is another web-based server prediction secondary structure of protein without template (Wang, et al. 2016a, b). This server utilizes DeepCNF (Deep Convolutional Neural Fields), a new machine learning model that simultaneously predicts secondary structure and solvent accessibility and disorder regions (Wang, et al. 2016a, b). It accomplishes Q3 score of approximately 84 for 3 state secondary structures and approximately 66% for three state solvent accessibility.

The tertiary structure of multi-epitopes chimeric vaccine candidate was built using I-TASSER online server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/). The I-TASSER (Iterative Threading Assembly Refinement) webserver utilizes the sequence-to-structure-to-function paradigm to build protein structure (Roy et al. 2010). It is a top-ranked 3D protein structure web server in community-wide CASP experiments (Zheng et al. 2019).

Tertiary Structure Refinement and Validation

Two web bases servers were used to refine the 3D structure of the multiepitopes chimeric protein. Initially, the Modrefiner server (https://zhanglab.ccmb.med.umich.edu/ModRefiner/) and finally GalaxyRefine server (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE) was used. Modrefiner server is based on an algorithm for atomic-level structure refinement that utilizes c alpha trace, main chain, or atomic model (Xu and Zhang 2011). The output structure was refined in terms of the accurate position of the side chains hydrogen bond network and fewer atomic overlaps. On the other hand, the Galaxy server rebuilds the 3D structure, performs repacking, and uses molecular dynamic simulations to accomplish overall protein structure relaxation. Structure refined by the Galaxy server is of the best quality as per community-wide CASP10 experiments (Ko et al. 2012).

ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php), The ERRAT server (http://services.mbi.ucla.edu/ERRAT/) and RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) web servers were utilized for 3D structure validation obtained after galaxy server refinement. ProSA web server calculates the quality score as Z score that should fall in a characteristic range (Wiederstein and Sippl 2007). Z score obtained for a specific input protein is in context with protein structure available in the public domain. ERRAT server analyses non-bonded atom–atom interaction with the refined 3D structure of protein compared to high resolved crystallographic protein structures (Colovos and Yeates 1993). Ramachandran plot displays energetically allowed and disallowed dihedral psi and phi angles of amino acids. The plot is calculated based on van der Waal radius of the protein side chain. RAMPAGE server determines the Ramachandran plot for a protein that includes the percent of residues in allowed and disallowed regions (Lovell et al. 2003).

Discontinuous B Cell Epitope Prediction

More than 90% B cell epitopes are discontinuous that is they are present in small segments on linear protein and brought to proximity during protein folding. Discontinues B cell epitopes for the designed vaccine were predicted by Ellipro online tool (http://tools.immuneepitope.org/tools/ElliPro) at IEDB. In this tool, three algorithms were implemented to determine the discontinuous B cell epitopes. 3D structure of the input protein is approximated as the number of ellipsoid shapes, calculate protrusion index (PI) and clusters neighboring residues. Ellipro defines the PI score of each residue based on the center of mass of residue residing outside the largest possible ellipsoid (Ponomarenko et al. 2008). considering other epitope predicting tools Ellipro gave an AUC value of 0.732, best of all (Ponomarenko et al. 2007).

Protein–Protein Docking of Designed Vaccine with TLR2 and TLR4

Molecular docking server HADDOCK (https://bianca.science.uu.nl/haddock2.4/) was used to see the interaction of the designed vaccine and TLR4. HADDOCK (High Ambiguity Driven protein–protein DOCKing) is an information-derived flexible docking server (Van Zundert et al. 2016). Galaxyrefine 3D model of multi-epitopes chimeric proteins, adjuvant TLR2 (PDB ID: 6NIG), and TLR4 (PDB ID: 4G8A) were uploaded for docking at HADDOCK server, keeping all the parameter default. Finally, the top five models were downloaded and by PRODIGY (PROtein binDIng enerGY prediction) webserver (https://nestor.science.uu.nl/prodigy/) was utilized for prediction of binding affinities (Xue et al. 2016).

Immune Simulation

In silico cloning of the vaccine was predicted by the C-ImmSim server (http://150.146.2.1/C-IMMSIM/index.

php). C-immSim Server is a freely available web-based server that utilizes a position-specific scoring matrix (PSSM) for the prediction of immune epitopes and machine learning techniques for immune interaction (Rapin et al. 2010; Rapin et al. 2011).

JCat (Java Codon Adaptation Tool) server was utilized for reverse translation and codon optimization. Codon optimization is carried out to express the construct in the *E. coli* host. JCat (http://www.prodoric.de/JCat) output display sequence of nucleotide, other important properties of sequence that includes codon adaptation Index (CAI), percent GC content essential for to assess the protein expression in the host (Grote et al. 2005). Finally, the vaccine construct was cloned in the pET30a (+) plasmid vector by adding XhoI and XbaI restriction sites at C and N terminus, respectively. The Snapgene tool was used to clone the construct to ensure cloning and expression (Goldberg et al. 2018).

Results

Prediction of B and T Cell Receptor Epitopes

T Cell Receptor Epitopes

IEDB recommended MHC-NP, netCTLpan1.1, RANKPEP, and netMHCpan3.0 servers predicted the potential candidate epitopes from the spike glycoprotein sequence of SARS-CoV2. Epitopes commonly predicated by at least four prediction tools were selected for further analysis parameters. Four such epitopes with a high score were predicted by four different tools (Table 1A). All the four epitopes were immunogenic and antigenic, conserved 100%, and predicted to be non-allergen (Table 1B). Immunogenicity scores for the predicted epitopes range from 0.3858 to 0.0961. All the epitopes were with positive scores hence considered for

further analysis. All the peptides were predicted to be non-toxic by toxipred online toxicity prediction tools.

B Cell Receptor Linear Epitopes

After provision of spike protein sequence to the BepiPred server showed an average score of 0.407, with 0.695 maximum and 0.163 minimum scores (Fig. 1). Other properties like the physiology of amino acid residues, accessible surface, hydrophilicity, Flexibility, and β turns were also determined (Table 2 and Fig. 1). Peptides obtained in six different methods were arranged according to higher to lower scores and 2% of high score peptides were selected for overlap screening. Regions commonly overlap in at least four prediction methods selected as potential B epitopes. Regions 250 to 261 and 1246 to 1267 were overlapped regions as per the condition applied. Region 250 to 260 includes BepiPred region 250–260, 251 to 257 predicted by Karplus-Schulz flexibility and Chau-Fasman β turn prediction methods, and 250 to 257 regions predicted by parker hydrophilicity prediction method. Similarly, region 1246 to 1267 covers region predicted by BepiPred 2.0 (1252-1267), Emini surface accessible (1257-1262), Kolaskar, and Tongoankar method (1247-1254), Chau-Fasman β turn prediction method (1246-1252) and Parker hydrophilicity method (1256-1262).

Engineering of Multi-epitope Vaccine Sequence

Three peptides with high scores predicted for T cell receptors, and two peptides for B cell receptors were ligated together by GPGPG or AAY. Additionally, OmpA (GenBank: AFS89615.1) was linked at the N-terminal end of the designed polypeptides by EAAAK linker as an adjuvant protein. For purification purposes, six histidine residues were added. The final amino acid residues of the designed vaccine

Table 1 **A** Common peptides predicted by at least four TCR epitopes prediction tools; **B** Immunogenic properties of that peptides. In the case of immunogenic properties, peptide number 3 has a negative score for antigenicity hence neglected for further analysis

| (A)							
Epitopes	Sequence	Amino acid number	Net CTLpan	RankPep	IEDB	MHC NP	Net MHC 4.0
TCR1	WYIWLGFIAGLIAI	1214–1227	0.8	42.06	0.32	0.2815	0.3
TCR2	FVSNGTHWFV	1095–1105	0.3	–	0.07	0.344184	0.17
TCR3	KLPDDFTGCV	423–432	0.4	39.12	−0.68	0.591582	0.3
TCR4	KLNDLCFTNV	385–393	0.4	34.15	0.13	0.710803	0.9
(B)							
Epitopes	Sequence	Immunogenicity	Conservancy	Toxicity	Antigenicity	Allergen	
TCR1	WYIWLGFIAGLIAI	0.38584	100	Non-toxic	0.5798	Probable Non-allergen	
TCR2	FVSNGTHWFV	0.18796	100	Non-toxic	0.5982	Probable Non-allergen	
TCR3	KLPDDFTGCV	0.1001	100	Non-toxic	−0.5308	Probable Non-allergen	
TCR4	KLNDLCFTNV	0.0961	100	Non-toxic	2.6927	Probable Non-allergen	
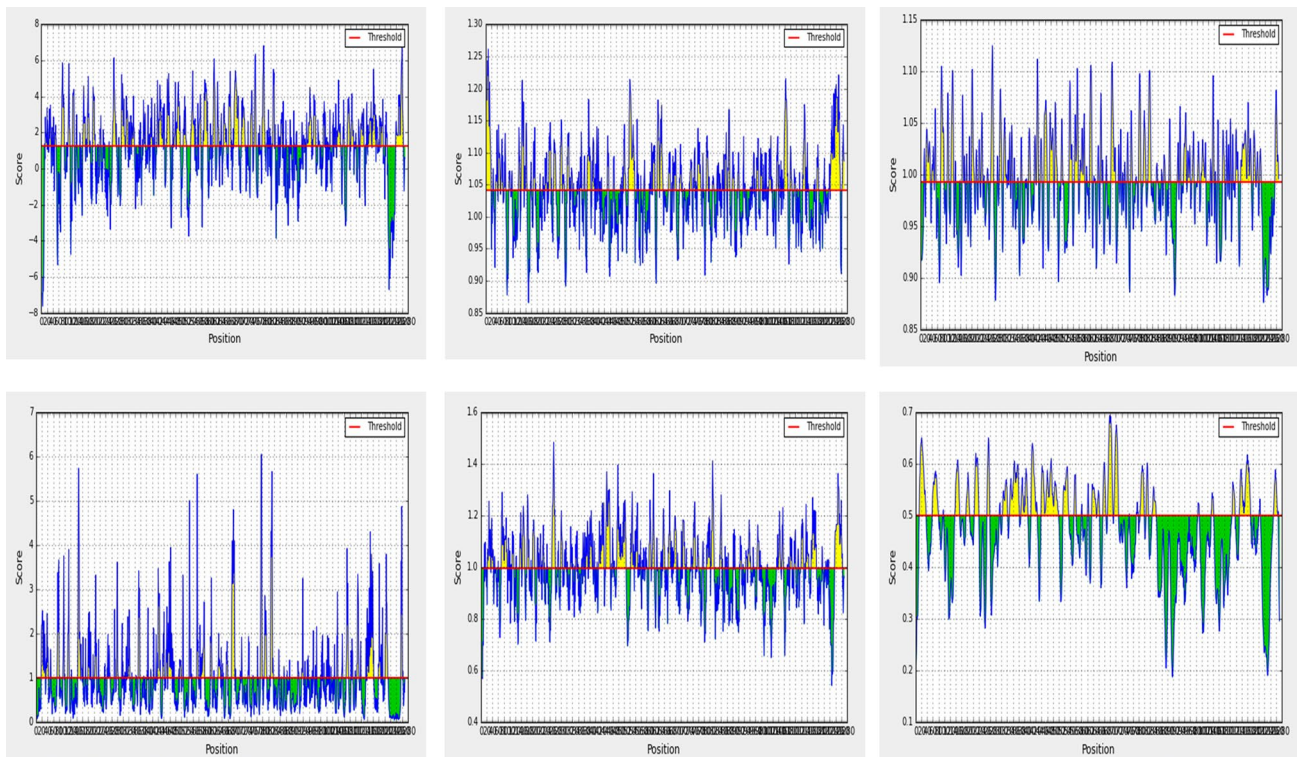
Fig. 1 Graphical representation of prediction of B cell epitopes Immunogenic peptide predicted by various tools probable epitopes are scored above the threshold, showed here as yellow. **A** Parker hydrophilicity prediction, **B** Chau-Fasman β turn prediction **C** Karplus-Schulz flexibility, **D** Emini Surface accessible, **E** Kolaskar and Tongoankar, F: Bepipred Linear Epitope Prediction 2.0

Table 2 B cell receptor epitope prediction, **A** Common B cell receptor epitopes predicted by various methods **B** Score obtained B cell receptor epitopes for Spike glycoprotein of SARS-CoV2

(A)

Method	Average score	Maximum score	Minimum score
BepiPred 2.0	0.470	0.183	0.695
Emini Surface accessible	1.0	0.042	6.051
Chua-Fasman β turn	0.997	0.541	1.484
Karplus-Schulz flexibility	0.993	0.876	1.125
Kolaskar–Tongoankar antigenicity	1.041	0.866	1.261
Parker hydrophilicity	1.238	−7.629	7.743

(B)

BCR epitopes	Region	Sequence
BCR1	250–260 (11 aa)	TPGDSSSGWTA
BCR2	1246–1267 (22 aa)	GCCSCGSCCK-FDEDDSEPVLKG

were 454 when all the five peptides, linkers, and adjuvant were ligated (Fig. 2).

Prediction of Immunogenic Properties of the Designed Vaccine

The antigenicity of the designed multi-epitopes vaccine predicted by the VaxiJen v2.0 server was 0.7048 with probable antigen annotation for the virus as a target organism at 0.5 thresholds. Antigenicity predicted for the same sequence
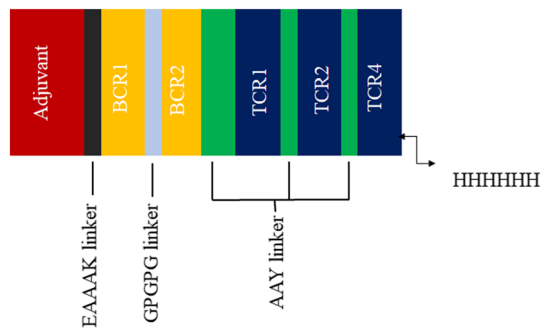
Fig. 2 Schematic presentation of the final multi-epitope vaccine peptide. The 599-amino acid long peptide sequence containing an adjuvant (Red) at the amino-terminal end linked with the multi-epitope sequence through an EAAAK linker (black). BCR epitopes (yellow) are linked using GPGPG linkers (cyan) while the TCR epitopes (blue) are linked with the help of AAY linkers (green). A 6x-His tag is added at the Carboxy terminus for purification and identification purposes (Color figure online)

Table 3 Analysis of the physicochemical and immunological properties of the designed vaccine for SARS-CoV2

Characters	score
Antigenicity (AntigenPro and VaxiJen v2.0)	0.765946 and 0.7048 (Probable antigen)
Allergen (AllergenFP and AllerTop)	0.84 and 0.77 (probable non-allergen)
Possible solubility	0.9557226
Number of amino acids	454
Molecular weight	48.49368
pI	7.94
The instability index (II)	24.58
Aliphatic index	79.02
Grand average of hydropathicity (GRAVY)	−0.172

of vaccines by the AntigenPro server was 0.765946. At the same time 0.953508 was predicted as probable solubility upon overexpression. The antigenicity of peptides without adjuvant sequence was 0.671028 with 0.861451 probable solubilities of peptides upon overexpression. The results obtained indicate that with and without adjuvant predicted vaccine sequence is potentially antigenic. AllergenFP and AllerTop servers predicted that both vaccines with adjuvant and without adjuvant are probable non-allergen with 0.84 and 0.77 Tanimoto similarity indexes, respectively.

Prediction Physiochemical Properties

The computed molecular weight of the designed vaccine was 48.49368 KDa and the theoretical pI was 7.96. Predicted pI indicates the slight alkaline nature of the vaccine (Table 3). The estimated half-lives were 30 h in mammalian reticulocytes (in vitro), >20 h in yeast (in vivo), and >10 h in *Escherichia coli* (in vivo). The vaccine was highly stable with an instability index of 24.58 classifying the protein as stable. The estimated aliphatic index was 79.02 indicating high thermo-stability of the protein (Doytchinova and Flower 2007a). The predicted GRAVY score was -0.175 (Table 3). The negative GRAVY score indicates that protein is hydrophilic and will interact with water molecules (Dimitrov et al. 2014b).

Secondary Structure Prediction

Secondary structure predicted by online tool RaptorX property includes 16% alpha-helix region, 42% beta-sheet region, and 40% coil region. Furthermore, solvent accessibility of amino acid residues was predicted to be 42% exposed, 27% medium exposed and 24% buried. A total of 56 residues (12%) were predicted as residues in the disordered region. Pictorial representation of secondary structure predicted of final protein by PSIPRED is shown in Fig. 3.

Tertiary Structure Prediction Refinement and Validation

I-TASSER predicted the top five 3D structure models of designed vaccines utilizing 10 threading templates. Z scores for these 10 templates were in the range of 0.95 to 5.50, indicating good alignment with the sequence submitted. C score, critical in the quality of the built model which was quantitatively measured. C score was in the range of −5 to 2 signifies the best model quality. C score of the top five models ranging from −4.14 to −3.41. A model with a high C score i.e. −3.41, estimated TM score 0.34 ± 0.11 and estimated RMSD 15.6 ± 3.3 was chosen for further analysis (Fig. 4A) (Smialowski et al. 2012).

For refinement purposes, the tertiary structure predicted by I-TASSER is initially submitted to ModRefiner and finally to GalaxyRefine. Among the top five refined models, model 5 was found to be the best model based on various parameters like GDT-HA=0.8398, RMSD=0.696, MolProbit=3.526, and Rama favored=77.1(Fig. 4B).

The Ramachandran plot at RAMPAGE webserver validated the quality of refined model 5. 79.5% of modeled vaccine residues were in the favored region, 13.1% in the allowed region and 7.3% were in the outlier region (Fig. 4D). Error in the model is analyzed by the ProSA web and ERRAT web server. Z score was −1.64 (Fig. 4E) and the

Fig. 3 Graphical representation of secondary structure characters; 42% beta-sheet region, 40% coil region and 16% alpha-helix region
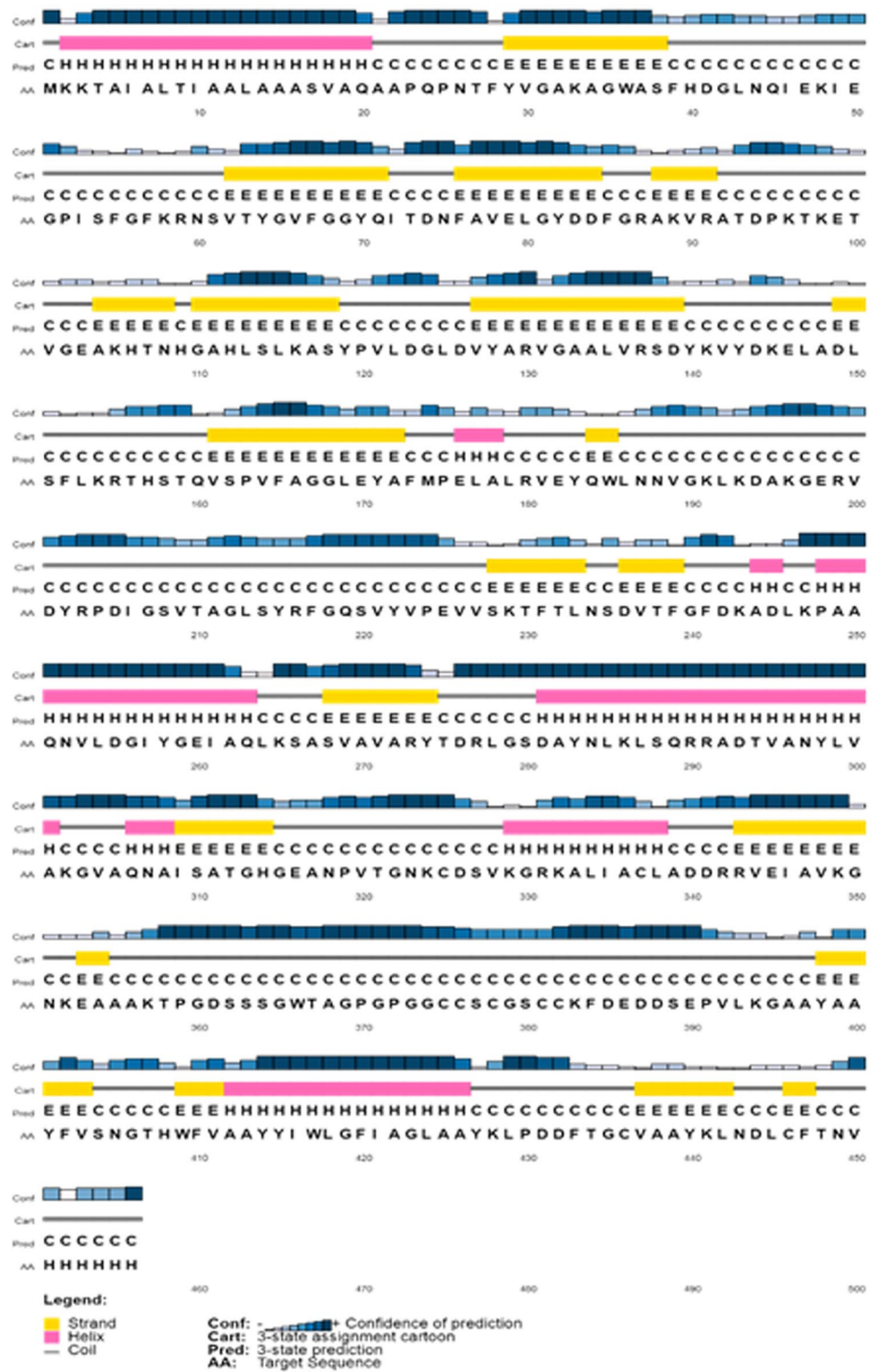
ERRAT score index was 74.387 (Fig. 4C). Ramachandran Plot Z score and ERRAT score showed that refined model 5 is of good quality and can be used further.

Dis-continuous Epitopes for B Cell

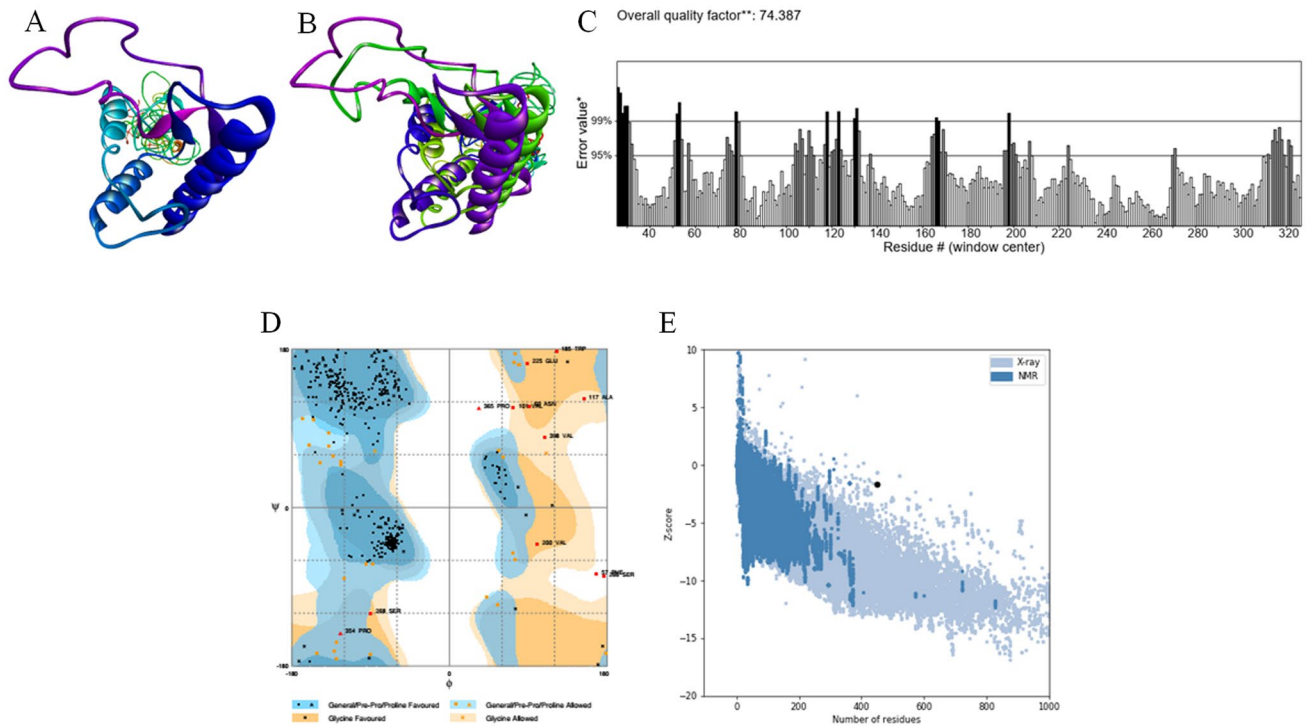Four discontinuous epitopes, including 265 residues, were predicted from the sequence of Model 5. The score of the predicted epitopes ranges from 0.58 to 0.749. The shortest

Fig. 4 Vaccine tertiary structure modeling, refinement, and validation **A** Tertiary structure of ITASSER modeled multi-epitopes Vaccine along with OmpA as adjuvant; **B** Refined 3D model (green) after GalaxyRefine superimposed on ITASSER crude model (violet) **C** Errata score for a refined model with graphical representation, **D** Ramachandran plot for a refined 3D of vaccine, **E** Z-score − 1.64 by ProSA webtool (Color figure online)

and longest discontinuous B cell epitopes are 14 and 97 residues long respectively.

Molecular Docking

HADDOCK web bases server was utilized to dock the designed vaccine to TLR2 and TLR 4. TLR2/4 eliciting immune response toward the designed vaccine was analyzed by conformational change concerning adjuvant TLR2/4 complexes. Top models from each group with the lowest HADDOCK score were selected. HADDOCK score for TLR2/Adjuvant and TLR2/Vaccine were −89.3 and −95.9 kcal/mol, respectively. Also, relative binding energies were −9.8 and −12.8 for the TLR2/adjuvant and TLR2/vaccine, respectively. Similarly, HADDOCK scores for the TLR4/adjuvant complex and TLR4/vaccine complex were −75.6 and −139.9 kcal/mol, respectively. The relative binding energy of the TLR4/adjuvant complex was −10.7 kcal/mol and −8.2 for the TLR4 vaccine complex. The difference in scores and binding energies of adjuvants and vaccines indicates the change in conformation that may stimulate TLR2/4 receptors. Even there was a difference in the number of interacting residues at the juncture. TLR2/Adjuvant complex have charged-charged 2, charged

polar 6, charged-apolar 24, polar-polar 2, polar-apolar 22 and apolar-apolar21. While TLR2/Vaccine has charged-charged 3, charged polar 7, charged-apolar 15, polar-polar 1, polar-apolar 13, and apolar-apolar 119 (Fig. 5A, B). Similarly, TLR4/Adjuvant complex have charged-charged 3, charged polar 10, charged-apolar 21, polar-polar 3, polar-apolar 15 and apolar-apolar20. While TLR4/vaccine complex has charged-charged 1, charged-polar 9, charged-apolar 10, polar-polar 9, polar-apolar 15 and apolar-apolar 8 interface contacts were observed (Fig. 5C, D).

Codon Optimization

Enhance expression of nucleotide construct in *Escherichia coli*, is reverse translated in Jcat online web tool. The total number of nucleotides in the construct was 1350. After codon optimization, CAI was 1.0 and the average GC content of the optimized nucleotide construct was 50.7340 indicating good probable expression of the vaccine in *E. coli* K12. To clone the gene XhoI and XbaI restriction sites are added at N and C terminus of the construct respectively and was cloned in pET30a + plasmid, by using SnapGene software (Fig. 6).

Fig. 5 Haddock docking cartoon models showing interaction **A** TLR2-Vaccine; TLR2 in red and Vaccine in blue; **B** TLR2-Adjuvant; TLR2 in red and Adjuvant in blue; **C** TLR4-Vaccine; TLR4 in red and Vaccine in red; **D** TLR4-Adjuvant; TLR4 in red and Adjuvant in blue
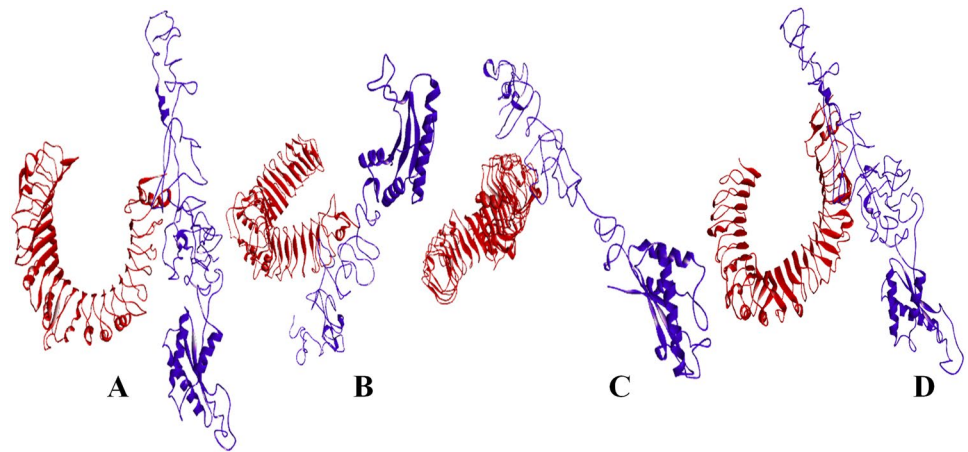
Immune Simulation

C-ImmSim Immune simulator webserver was used for determining the ability of a vaccine to induce immunity. Results obtained are consistent with the experimental results published elsewhere. Within the first week of vaccine administration, the primary response observed was the high-level secretion of IgM. The secondary and tertiary immune response seen with an increase in the B cell number, level of IgM, IgG1 + IgG2, and IgM + IgG with decreasing concentration of the antigen (Fig. 7A). Different B cell isotype populations were also found increasing indicating isotype switching and memory formation (Fig. 7B). B cell activities were also found high, especially B Isotype IgM and IgG1,

which were observed with prominent memory cell formation (Fig. 7B). Similarly, the cell populations of Th and Tc cells are found high along with memory development (Fig. 7C, D). Also, active macrophages were found to be increased consistently after each antigen shot and declined upon antigen clearance (Fig. 7E). Another cell from cell-mediated immunity, dendritic cells were also found increased (Fig. 7F). IFNγ and IL2 expression were found high with low Simpson Index indicating sufficient immunoglobulin production, suggesting good humoral immune response (Fig. 7G). Simpson Index (D) is a measure of diversity. An increase in D over time indicates the emergence of different epitope-specific dominant clones of T-cells. The smaller the D value, the lower the diversity.
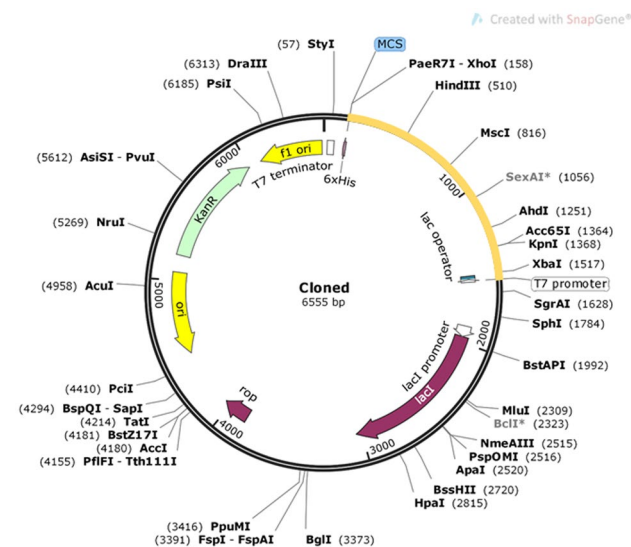

Fig. 6 In silico cloning of designed vaccine construct in pET30a + expression vector where the yellow arc segment represents cloned construct and the remaining segment represents vector backbone

Discussion

To date there is no sure cure available for COVID19, suggesting the urgent requirement of drug or vaccine to control the spread of SAR-CoV2 infections. Advancements in bioinformatics tools, the vaccine development process can be facilitated by identifying potential epitopes for T and B cells that will lead to a vaccine for the prevention of COVID19. Several reports and animal trials suggest that S-protein can be a potential target for vaccine development (Prompetchara et al. 2020; Tai et al. 2020; Du et al. 2009; Pogrebnyak et al. 2005; Jiang et al. 2005; Hotez et al. 2020). Several reports are suggesting the importance of Spike protein in the activation of cells of immunity and vaccine development (Tufan et al. 2020; Prompetchara et al. 2020; Zhang et al. 2020a, b). HLA-A2 restricted epitopes from S protein of SARS-CoV2, elicit T cell-specific immune response in SARS-CoV2 recovered patients (Wang et al. 2004). Similarly, serum samples of SARS-CoV2 recovered patients were sufficient to neutralize epitope-rich regions on the spike S2 protein from SARS-CoV2 (Zhong et al. 2005). Programing
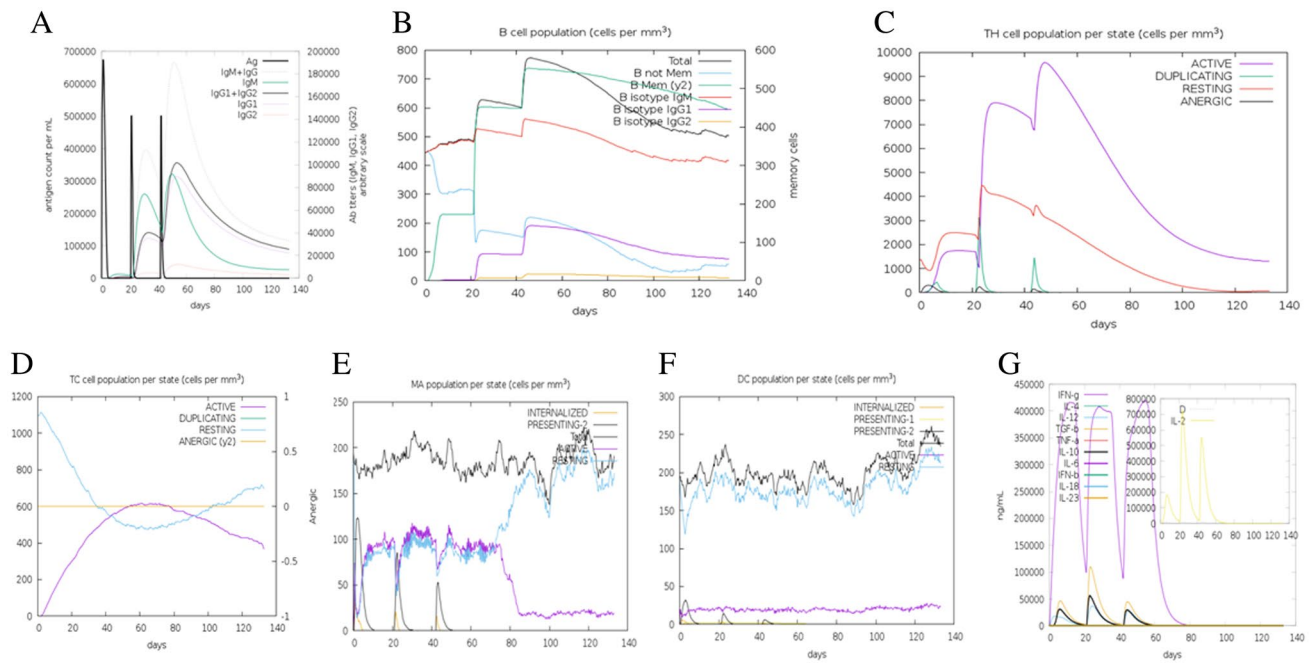
Fig. 7 Immune response simulation by a designed vaccine, **A** Immune response upon antigen exposure; **B** B cell population per state (cells/mm3); C: TH cell population state (cells/mm3); **D** TC cell population state (cells/mm3); **E** Activity of macrophage population in three subsequent immune responses; **F** Dendritic cell population state (cells/mm3) upon antigen exposure; G:Cytokine-induced by three injections of vaccine after a one-week interval, The main plot shows cytokine levels after the injections. The insert plot shows IL-2 level with the Simpson index, D indicated by the dotted line

of a live attenuated form of the pathogen is a commonly used strategy for vaccine development. Despite the efficacy of such vaccines, the reversion of virulence remains a challenge and is not utilized in weak immune patients (Plitnick and Herzyk 2013; Plitnick 2013). Here in case SARS-CoV2 due to its high communal transmission vaccine research requires highly skilled workers, sophisticated instrumentation, and biosafety level III facility. This may likely slow down the process of vaccine development, enhancing cost that leads to a restriction on the availability of vaccines for the mass population. Advancements in the field of bioinformatics and molecular biology techniques have provided an opportunity for the development of a vaccine with high efficacy that reduces vaccine development time, and cost-effective products that may reduce the cost of vaccine in time for the large population. Various research groups across the world are searching for highly effective vaccine to stop the spread of virus. Research communications published till now has one common thing and that is spike protein as the protein of interest for the development of vaccine. Covisheid of Oxford, JNJ-78436735 of Janssen Pharmaceuticals and Gam-COVID-Vac of Sputnik are actually contained whole spike protein as active ingredient of SARS-CoV2 origin. On the other hand, Moderna's mRNA vaccine codes spike protein to provide the immunity against SARS-CoV2. As already described above the advantages of bioinformatic

tools in vaccines development process, there are some research papers in the public domain focusing on the multiepitopic vaccine candidate against SARS-CoV2. Spike protein is at the center for most of the studies and used various strategies. Various structural proteins like nucleocapsid ORF3a, envelop and membrane protein were evaluated for epitope screening including spike proteins (Kumar et al. 2020; Rahman et al. 2020; Enayatkhani et al. 2021). Long stretches (more than 20 aa), short stretches (less than 20 aa) along various linker proteins were the primary constituent of most of the multiepitopic SARS-CoV2 vaccine (Kumar et al. 2020; Abraham Peele et al. 2021). Recently deep learning approach was used to search the epitopes out of spike protein sequence for the prediction of vaccine subunit. DeepVacPred a AI-based frame work predicted 26 vaccine subunits within a second, skipping almost 95% unnecessary prediction and manual curing (Yang et al. 2021).

In this study, we designed a multi-epitopes vaccine construct from the S-protein of SARS-CoV2. Several online tools were utilized for the designing of epitopes and thereby a vaccine. From various epitopes predicted by the online server based on common sequence and high score, three TCR and two BCR epitopes were selected as part of COVID19 vaccine. All the five TCR and BCR epitopes were linked by GPGPG and AAY linker peptides (Fig. 2). To enhance the antigenicity of TCR/BCR linked epitopes

peptide, adjuvant OmpA were linked by EAAAK linker sequence for a complete vaccine design which was found to restore antigenicity and non-allergen status. (Table 2). For a successful vaccine candidate, secondary and tertiary structures characters play an essential role. The secondary structure of the designed vaccine contains 16% alpha helixes, 42% beta-sheets and 40% coils. Upon refinement tertiary structure of the designed vaccine showed improvement to a desirable level as indicated by the Ramachandran plot. The designed vaccine showed a high antigenicity score, no probable allergen.

TLR are essential receptor proteins in the activation of the innate immune response. That recognizes and responds by induction of immune reactions to PAMPs (pathogen-associated molecular patterns). Various TLRs have been shown to activate the immune response to the virus through their interaction with nucleic acids and envelopes proteins. TLR3 and TLR7/8 are involved in nucleic acid detection, while TLR2 and TLR4 recognize envelope glycoproteins (Xagorari and Chlichlia 2008). Interaction of S protein with TLR2 increases the production of IL8 in human monocyte macrophages (Dosch et al. 2009). TLR 2 is also involved in eliciting innate immune response upon recognition of several other virus components (Finberg and Kurt-Jones 2004; Luo et al. 2020; Aguilar-Briseño et al. 2020; Lin et al. 2020). In a comparative binding efficacy of S-proteins to TLRs, TLR 4 showed the strongest protein–protein interaction with hydrogen and hydrophobic bonds on the extracellular domain (Choudhury and Suprabhat 2020). In molecular docking analysis, TLR2 and TLR4 showed stable protein–protein interaction with the designed vaccine compared to adjuvant protein. TLR 2 showed a more stable interaction than TLR4, which is consistent with the previous report (Choudhury and Suprabhat 2020). Stable interaction of designed vaccine indicates the efficiency of vaccine for activation of TLRs, involved in dendritic cell activation, thereby subsequent antigen processing, and presentation on the surface to T cells (Englmeier 2020; Li et al. 2020). The immune stimulation showed typical natural immune response patterns upon multiple exposures to the same antigen. The server predicted elevated levels of B cell and T cell for a longer time on repeated exposure of antigen. Increased levels of antiviral cytokine IFNγ and IL2 indicate the potential of subsequent activation of T-helper cells thereby the high level of Ig production, supporting humoral immune response (Majid and Andleeb 2019).

To validate the designed vaccine in the screening of immune reactivity by serological test, it is necessary to express the construct in a preferred host like *E. coli* K12 for recombinant proteins. Codon optimization was performed to achieve a high expression designed vaccine in *E coli* 12. Codon adaptability index (CAI = 1.0) and GC content (51.66%) of the vaccine construct were optimum for high expression of recombinant protein in the host. The immediate step, hereafter is to express the protein in a preferred host and validate results obtained in this report by performing several immunological assays.

Different variants of SARS-CoV2 were emerged in mean time evading neutralizing antibody response (Edara et al. 2021; Planas et al. 2021). Interestingly, the predicted epitopes do not fall in the region reported as a unique part of different variants of SARS-CoV2 including delta variant (P681R) (UniprotKB. 2021). A vaccine made of such epitopes may work against variants of SARS-CoV2.

Conclusion

With an inadequate number of drugs and the time-consuming process of drug development, the multiplication chain of SARS-CoV2 infection can be controlled only by the potential vaccine. We utilized tools and techniques of immune informatics to design a potential vaccine candidate. This vaccine has epitopes from the S protein of the SARS-CoV2 virus for T and B cell receptors. One more time, This scientific report showed that the bioinformatics approach could effectively be used to develop vaccines in a short time and with incurred less cost. Although in silico results point out the effectiveness of the vaccine, efficacy needs to be analyzed by performing laboratory experiments and animal model studies.

Declarations

Conflict of Interest G H Kathwate declares no conflict of interest.

Animal Rights This article does not contain any studies with human participants performed by the author.

References

Abraham Peele K, Srihansa T, Krupanidhi S, Ayyagari VS, Venkateswarulu TC (2021) Design of multi-epitope vaccine candidate against SARS-CoV-2: a in-silico study. J Biomol Struct Dyn 39(10):3793–3801. https://doi.org/10.1080/07391102.2020.1770127

Aguilar-Briseño José A, Upasani Vinit, Ter Ellen Bram M, Moser Jill, Pauzuolis Mindaugas, Ruiz-Silva Mariana, Heng Sothy, Laurent Denis, Choeung Rithy, Philippe J (2020) TLR2 on blood monocytes senses dengue virus infection and its expression correlates with disease pathogenesis. Nat Commun Dussart 11(1):1–14

Ahmed SF, Quadeer AA, McKay MR (2020) Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses 12(3):254

Alsaadi J, Entedar A, Jones IM (2019) Membrane binding proteins of coronaviruses. Futur Virol 14(4):275–286

Bai Y, Yao L, Wei T, Tian F, Jin D-Y, Chen L, Wang M (2020) Presumed asymptomatic carrier transmission of COVID-19. JAMA 323(14):1406–1407

Calis Jorg JA, Maybeno Matt, Greenbaum Jason A, Weiskopf Daniela, De Silva Aruna D, Sette Alessandro, Keşmir Can, Peters Bjoern (2013) Properties of MHC class I presented peptides that enhance immunogenicity. PLoS Comput Biol 9(10):e1003266

Bin C, Wang Y, Wen D, Liu W, Wang J, Fan G, Ruan L, Song B, Cai Y, Wei M (2020) A trial of lopinavir–ritonavir in adults hospitalized with severe Covid-19. New Eng J Med.

Chan JF-W, Yip CC-Y, To KK-W, Tang TH-C, Wong SC-Y, Leung K-H, Fung AY-F, Ngen AC-K, Zou Z, Tsoi H-W (2020) Improved molecular diagnosis of COVID-19 by the novel, highly sensitive and specific COVID-19-RdRp/Hel real-time reverse transcription-PCR assay validated in vitro and with clinical specimens. J Clin Microb. https://doi.org/10.1128/JCM.00310-20

Chen Tao, Di Wu, Chen Huilong, Yan Weiming, Yang Danlei, Chen Guang, Ma Ke, Dong Xu, Haijing Yu, Wang Hongwu (2020) Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. BMJ 368:m1091

Cheng J, Randall AZ, Sweredoski MJ, Baldi P (2005) SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res. https://doi.org/10.1093/nar/gki396

Chorin E, Wadhwani L, Magnani S, Dai M, Shulman E, Nadeau-Routhier C, Knotts R, Bar-Cohen R, Kogan E, Barbhaiya C (2020) QT interval prolongation and torsade de pointes in patients with COVID-19 treated with hydroxychloroquine/azithromycin. Heart Rhythm.

Choudhury A, Mukherjee S (2020) In silico studies on the comparative characterization of the interactions of SARS-CoV-2 spike glycoprotein with ACE-2 receptor homologs and human TLRs. J Med Virol.

Choy K-T, Wong AY-L, Kaewpreedee P, Sia S-F, Chen D, Hui KPY, Chu DKW, Chan MCW, Cheung PP-H, Huang X (2020) Remdesivir, lopinavir, emetine, and homoharringtonine inhibit SARS-CoV-2 replication in vitro. Antiviral Res:104786.

Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2(9):1511–1519

Dhanda SK, Mahajan S, Paul S, Yan Z, Kim H, Jespersen MC, Jurtz V, Andreatta M, Greenbaum JA, Marcatili P, Sette A, Nielsen M, Peters B (2019) IEDB-AR: immune epitope database—analysis resource in 2019. Nucleic Acids Res 47(W1):W502–W506. https://doi.org/10.1093/nar/gkz452%JNucleicAcidsResearch

Dimitrov I, Bangov I, Flower DR, Irini J (2014) AllerTOP v. 2—a server for in silico prediction of allergens. J Mol Model 20(6):2278

Dimitrov I, Naneva L, Doytchinova I, Bangov I (2014b) AllergenFP: allergenicity prediction by descriptor fingerprints. Bioinformatics 30(6):846–851

Dosch SF, Mahajan SD, Collins AR (2009) SARS coronavirus spike protein-induced innate immune response occurs via activation of the NF-κB pathway in human monocyte macrophages in vitro. Virus Res 142(1–2):19–27

Doytchinova IA, Flower DR (2007a) Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine 25(5):856–866

Doytchinova IA, Flower DR (2007b) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 8(1):4

Du L, He Y, Zhou Y, Liu S, Zheng B-J, Jiang S (2009) The spike protein of SARS-CoV—a target for vaccine and therapeutic development. Nat Rev Microbiol 7(3):226–236

Edara VV, Hudson WH, Xie X, Ahmed R, Suthar MS (2021) Neutralizing Antibodies Against SARS-CoV-2 Variants After Infection and Vaccination. JAMA 325(18):1896–1898. https://doi.org/10.1001/jama.2021.4388%JJAMA

Emini EA, Hughes JOSEPHV, Perlow DoS, Boger J (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol 55(3):836–839

Enayatkhani M, Hasaniazad M, Faezi S, Gouklani H, Davoodian P, Ahmadi N, Einakian MA, Karmostaji A, Ahmadi K (2021) Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. J Biomol Struct Dyn 39(8):2857–2872. https://doi.org/10.1080/07391102.2020.1756411

Englmeier L (2020) A theory on SARS-CoV-2 susceptibility: reduced tlr7-activity as a mechanistic link between men, obese and elderly. J Biol Regulator Homeostat Agents.

Finberg RW, Kurt-Jones EA (2004) Viruses and Toll-like receptors. Microbes Infect 6(15):1356–1360

Fleri Ward, Paul Sinu, Dhanda Sandeep Kumar, Mahajan Swapnil, Xiaojun Xu, Peters Bjoern, Sette Alessandro (2017) The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. Front Microbiol. https://doi.org/10.3389/fimmu.2017.00278

Gandhi M, Yokoe DS, Havlir DV (2020) Asymptomatic transmission, the Achilles' heel of current strategies to control COVID-19. Mass Medical Soc.

Garnier J (1978) Analysis of the accuracy and implication of simple methods for predicting the secondary structure of globular proteins. J Mol Biol 120:97–120

Gautret P, Lagier J-C, Parola P, Meddeb L, Sevestre J, Mailhe M, Doudier B, Aubry C, Amrane S, Seng P (2020) Clinical and microbiological effect of a combination of hydroxychloroquine and azithromycin in 80 COVID-19 patients with at least a six-day follow up: a pilot observational study. Travel Med Infect Dis:101663.

Gautret P, Lagier J-C, Parola P, Meddeb L, Mailhe M, Doudier B, Courjon J, Giordanengo V, Vieira VE, Dupont HT (2020) Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. Int J Antimicrobial Agents:105949.

Glowacka I, Bertram S, Müller MA, Allen P, Soilleux E, Pfefferle S, Steffen I, Tsegaye TS, He Y, Gnirss K (2011) Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. J Virol 85(9):4122–4134

Goldberg Michael F, Roeske Elizabeth K, Ward Lauren N, Pengo Thomas, Dileepan Thamotharampillai, Kotov Dmitri I, Jenkins Marc K (2018) Salmonella persist in activated macrophages in T cell-sparse granulomas but are contained by surrounding CXCR3 ligand-positioned Th1 cells. Immunity 49(6):1090–11027

Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Moderbacher CR, Rawlings SA, Sutherland A, Premkumar L, Jadi RS (2020) Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. Cell.

Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, Jahn D (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 33 (suppl_2):W526–W531.

Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Muller MA, Drosten C, Pohlmann S (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically

proven protease inhibitor. Cell 181(2):271-288 e8. https://doi.org/10.1016/j.cell.2020.02.052

Hotez PJ, Corry DB, Bottazzi ME (2020) COVID-19 vaccine design: the Janus face of immune enhancement. Nat Rev Immunol 20(6):347–348

Jespersen MC, Peters B, Nielsen M, Marcatili P (2017) BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res 45(W1):W24–W29

Jiang S, He Y, Liu S (2005) SARS vaccine development. Emerg Infect Dis 11(7):1016

Karplus PA, Schulz GE (1985) Prediction of chain flexibility in proteins. Naturwissenschaften 72(4):212–213

Ko J, Park H, Heo L, Seok C (2012) GalaxyWEB server for protein structure prediction and refinement. Nucleic Acids Res 40(W1):W294–W297

Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett 276(1–2):172–174

Kumar A, Kumar P, Saumya KU, Kapuganti SK, Bhardwaj T, Giri R (2020) Exploring the SARS-CoV-2 structural proteins for multiepitope vaccine development: an in-silico approach. Expert Rev Vaccines 19(9):887–898. https://doi.org/10.1080/14760584.2020.1813576

Li G, Fan Y, Lai Y, Han T, Li Z, Zhou P, Pan P, Wang W, Dingwen Hu, Liu X (2020) Coronavirus infections and immune responses. J Med Virol 92(4):424–432

Li Y, Xia L (2020) Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management. Am J Roentgenol:1–7.

Lin Y-T, Chen Y-P, Fang C-H, Huang P-Y, Liang S-M (2020) Capsid proteins of foot-and-mouth disease virus interact with TLR2 and CD14 to induce cytokine production. Immunol Lett .

Lopez Bernal J, Andrews N, Gower C, Gallagher E, Simmons R, Thelwall S, Stowe J, Tessier E, Groves N, Dabrera G, Myers R, Campbell CNJ, Amirthalingam G, Edmunds M, ZambonM, Brown KE, Hopkins S, Chand M, Ramsay M (2021) Effectiveness of Covid-19 vaccines against the B.1.617.2 (delta) variant. 385 (7):585–594. https://doi.org/10.1056/NEJMoa2108891.

Lovell SC, Davis IW, Bryan Arendall III W, de Bakker PIW, Michael Word J, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins 50(3):437–450. https://doi.org/10.1002/prot.10286

Luo X, Lu M, Baba HA, Gerken G, Wedemeyer H, Broering R (2020) Hippo signaling counter regulates early innate immunity in hepatocytes exposed to Hepatitis B virus. Zeitschrift für Gastroenterologie 58(01):519

Majid M, Andleeb S (2019) Designing a multi-epitopic vaccine against the enterotoxigenic Bacteroides fragilis based on immunoinformatics approach. Sci Rep 9(1):1–15

McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16(4):404–405. https://doi.org/10.1093/bioinformatics/16.4.404%JBioinformatics

Mishra NK, Singla D, Agarwal S, Raghava GPS (2014) ToxiPred: a server for prediction of aqueous toxicity of small chemical molecules in T. Pyriformis. J Trans Toxicol 1(1):21–27. https://doi.org/10.1166/jtt.2014.1005

Mitjà O, Clotet B (2020) Use of antiviral drugs to reduce COVID-19 transmission. Lancet Glob Health 8(5):e639–e640

Molina JM, Delaugerre C, Le Goff J, Mela-Lima B, Ponscarme D, Goldwirt L, de Castro N (2020) No evidence of rapid antiviral clearance or clinical benefit with the combination of hydroxychloroquine and azithromycin in patients with severe COVID-19 infection. Med Mal Infect 50 (384):10.1016.

Olliaro P, Torreele E, Vaillant M (2021) COVID-19 vaccine efficacy and effectiveness—the elephant (not) in the room. Lancet Microbe 2(7):e279–e280. https://doi.org/10.1016/S2666-5247(21)00069-0

Osman EEA, Toogood PL, Neamati N (2020) COVID-19: Living through Another Pandemic. ACS Infect Dis.

Parker JMR, Guo D, Hodges RS (1986) New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry 25(19):5425–5432

Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, Planchais C, Porrot F, Robillard N, Puech J, Prot M, Gallais F, Gantner P, Velay A, Le Guen J, Kassis-Chikhani N, Edriss D, Belec L, Seve A, Courtellemont L, Péré H, Hocqueloux L, Fafi-Kremer S, Prazuck T, Mouquet H, Bruel T, Simon-Lorière E, Rey FA, Schwartz O (2021) Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature 596(7871):276–280. https://doi.org/10.1038/s41586-021-03777-9

Plitnick L, Herzyk D (2013) Nonclinical development of novel biologics, biosimilars, vaccines and specialty biologics. Academic Press.

Plitnick LM (2013) Global regulatory guidelines for vaccines. In Nonclinical development of novel biologics, biosimilars, vaccines and specialty biologics, 225–241. Elsevier.

Pogrebnyak N, Golovkin M, Andrianov V, Spitsin S, Smirnov Y, Egolf R, Koprowski H (2005) Severe acute respiratory syndrome (SARS) S protein production in plants: development of recombinant vaccine. Proc Natl Acad Sci 102(25):9062–9067

Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, Perez JL, Marc GP, Moreira ED, Zerbini C, Bailey R, Swanson KA, Roychoudhury S, Koury K, Li P, Kalina WV, Cooper D, Frenck RW, Hammitt LL, Türeci O, Nell H, Schaefer A, Ünal S, Tresnan DB, Mather S, Dormitzer PR, Şahin U, Jansen KU, Gruber WC (2020) Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. 383 (27):2603–2615. https://doi.org/10.1056/NEJMoa2034577.

Ponomarenko JV, Philip E (2007) Antibody-protein interactions: benchmark datasets and prediction tools evaluation. BMC Struct Biol 7(1):64

Ponomarenko J, Bui H-H, Li W, Fusseder N, Bourne PE, Sette A, Peters B (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics 9(1):514

Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Flasche S, Clifford S, Pearson CAB, Munday JD (2020) The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. Lancet Public Health.

Prompetchara E, Ketloy C, Palaga T (2020) Immune responses in COVID-19 and potential vaccines: lessons learned from SARS and MERS epidemic. Asian Pac J Allergy Immunol 38(1):1–9

Qifang B, Wu Y, Mei S, Ye C, Zou X, Zhang Z, Liu X, Wei L, Truelove SA, Zhang T (2020) Epidemiology and transmission of covid-19 in Shenzhen China: analysis of 391 cases and 1,286 of their close contacts. MedRxiv.

Rahman MS, Hoque MN, Islam MR, Akter S, Rubayet Ul Alam ASM, Siddique MA, Saha O, Rahaman MM, Sultana M, Crandall KA, Hossain MA (2020) Epitope-based chimeric peptide vaccine design against S, M and E proteins of SARS-CoV-2, the etiologic agent of COVID-19 pandemic: an in silico approach. PeerJ 8:e9572. https://doi.org/10.7717/peerj.9572

Rapin Nicolas, Lund Ole, Bernaschi Massimo, Castiglione Filippo (2010) Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. PLoS One 5(4):e9862

Rapin N, Lund O, Castiglione F (2011) Immune system simulation online. Bioinformatics 27(14):2013–2014. https://doi.org/10.1093/bioinformatics/btr335

Rosenberg ES, Dufort EM, Udo T, Wilberschied LA, Kumar J, Tesoriero J, Weinberg P, Kirkwood J, Muse A, DeHovitz J (2020) Association of treatment with hydroxychloroquine or azithromycin with in-hospital mortality in patients with COVID-19 in New York state. JAMA.

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5(4):725–738

Ruch TR, Machamer CE (2012) The coronavirus E protein: assembly and beyond. Viruses 4(3):363–382

Shey RA, Ghogomu SM, Esoh KK, Nebangwa ND, Shintouo CM, Nongley NF, Asa BF, Ngale FN, Vanhamme L, Souopgui J (2019) In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. Sci Rep 9(1):1–18

Shi Y, Wang Y, Shao C, Huang J, Gan J, Huang X, Bucci E, Piacentini M, Ippolito G, Melino G (2020) COVID-19 infection: the perspectives on immune responses. Nature Publishing Group.

Smialowski P, Doose G, Torkler P, Kaufmann S, Frishman D (2012) PROSO II–a new method for protein solubility prediction. FEBS J 279(12):2192–2200

Snijder EJ, Decroly E, Ziebuhr J (2016) The nonstructural proteins directing coronavirus RNA synthesis and processing. In Advances in virus research, 59–126. Elsevier.

Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, Iosifidis C, Agha R (2020) World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg 76:71–76. https://doi.org/10.1016/j.ijsu.2020.02.034

Spychalski P, Błażyńska-Spychalska A, Kobiela J (2020) Estimating case fatality rates of COVID-19. Lancet Infect Dis.

Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, Zhou Y, Du L (2020) Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cell Mol Immunol:1–8.

Tufan A, Güler AA, Matucci-Cerinic MJ (2020) COVID-19, immune system response, hyperinflammation and repurposing antirheumatic drugs. Turkish J Med Sci 50 (SI-1):620–632.

UniprotKB (2021) UniProtKB - P0DTC2 (SPIKE_SARS2), Pathology & Biotech-Mutagenesis. https://www.uniprot.org/uniprot/P0DTC2.

van Doremalen N, Lambe T, Spencer A, Belij-Rammerstorfer S, Purushotham JN, Port JR, Avanzato V, Bushmaker T, Flaxman A, Marta %J bioRxiv Ulaszewska. (2020) ChAdOx1 nCoV-19 vaccine prevents SARS-CoV-2 pneumonia in rhesus macaques. Nature 586:578–582

Van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond ASJ, Marc van Dijk SJ, de Vries AMJJ (2016) The HADDOCK2. 2 web server: user-friendly integrative modeling of biomolecular complexes. J Mol Biol 428(4):720–725

Wang B, Chen H, Jiang X, Zhang M, Wan T, Li N, Zhou X, Yanfeng Wu, Yang F, Yizhi Yu (2004) Identification of an HLA-A* 0201–restricted CD8+ T-cell epitope SSp-1 of SARS-CoV spike protein. Blood 104(1):200–206

Wang S, Li W, Liu S, Jinbo Xu (2016a) RaptorX-Property: a web server for protein structure property prediction. Nucleic Acids Res 44(W1):W430–W435

Wang S, Peng J, Ma J, Jinbo Xu (2016b) Protein secondary structure prediction using deep convolutional neural fields. Sci Rep 6(1):1–11

Wang Y, Zhang D, Guanhua Du, Ronghui Du, Zhao J, Jin Y, Shouzhi Fu, Gao L, Cheng Z, Qiaofa Lu, Yi Hu, Luo G, Wang Ke, Yang Lu, Li H, Wang S, Ruan S, Yang C, Mei C, Wang Yi, Ding D, Feng Wu, Tang X, Ye X, Ye Y, Liu B, Yang J, Yin W, Wang A, Fan G, Zhou F, Liu Z, Xiaoying Gu, Jiuyang Xu, Shang L, Zhang Yi, Cao L, Guo T, Wan Y, Qin H, Jiang Y, Jaki T, Hayden FG, Horby PW, Cao B, Wang C (2020) Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. The Lancet 395(10236):1569–1578. https://doi.org/10.1016/S0140-6736(20)31022-9

Wen CY, Xie ZW, Li YP, Deng XL, Chen XT, Cao Y, Ou X, Lin WY, Li F, Cai WP, Li LH (2020) Real-world efficacy and safety of lopinavir/ritonavir and arbidol in treating with COVID-19: an observational cohort study. Zhonghua Nei Ke Za Zhi 59:E012. https://doi.org/10.3760/cma.j.cn112138-20200227-00147

WHO. 2020. Situation report - 184:coronavirus disease 2019 (COVID-19). https://www.who.int/publications/m/item/situation-report---184.

Wiederstein, Markus, and Manfred J Sippl. 2007. "ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins." *Nucleic acids research* 35 (suppl_2):W407-W410.

Xagorari A, Chlichlia K (2008) Toll-like receptors and viruses: induction of innate antiviral immune responses. The Open Microbiology Journal 2:49

Xu D, Zhang Y (2011) Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys J 101(10):2525–2534

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(23):3676–3678

Yang Z, Bogdan P, Nazarian S (2021) An in silico deep learning approach to multi-epitope vaccine design: a SARS-CoV-2 case study. Sci Rep 11(1):3238. https://doi.org/10.1038/s41598-021-81749-9

Yelin I, Aharony N, Shaer-Tamar E, Argoetti A, Messer E, Berenbaum D, Shafran E, Kuzli A, Gandali N, Hashimshony T (2020) Evaluation of COVID-19 RT-qPCR test in multi-sample pools. MedRxiv.

Zhang, Yaguang, Jin Xu, Ran Jia, Chunyan Yi, Wangpeng Gu, Pengcheng Liu, Xinran Dong, Hao Zhou, Bo Shang, Shipeng Cheng (2020) Protective humoral immunity in SARS-CoV-2 infected pediatric patients. Cell Mol Immunol:1–3.

Zhang B-Z, Hu Y-F, Chen L-I, Yau T, Tong Y-G, Hu J-C, Cai J-P, Chan K-H, Dou Y, Deng J (2020) Mining of epitopes on spike protein of SARS-CoV-2 from COVID-19 patients. Cell Res:1–3.

Zheng W, Zhang C, Bell EW, Zhang Y (2019) I-TASSER gateway: a protein structure and function prediction server powered by XSEDE. Futur Gener Comput Syst 99:73–85

Zhong X, Yang H, Guo Z-F, Sin W-Y, Chen W, Junjie Xu, Ling Fu, Jie Wu, Mak C-K, Cheng C-S (2005) B-cell responses in patients who have recovered from severe acute respiratory syndrome target a dominant site in the S2 domain of the surface spike glycoprotein. J Virol 79(6):3401–3408