



Computational design of a broad-spectrum multi-epitope vaccine candidate against seven strains of human coronaviruses

Avinash Kumar¹ · Ekta Rathi¹ · Suvarna Ganesh Kini^{1,2}

Received: 26 August 2021 / Accepted: 29 July 2022 / Published online: 23 August 2022
© The Author(s) 2022

Abstract

Spike (S) proteins are an attractive target as it mediates the binding of the SARS-CoV-2 to the host through ACE-2 receptors. We hypothesize that the screening of the S protein sequences of all the seven known HCoV-229E would result in the identification of potential multi-epitope vaccine candidates capable of conferring immunity against various HCoV-229E. In the present study, several machine learning-based *in-silico* tools were employed to design a broad-spectrum multi-epitope vaccine candidate targeting the S protein of seven known strains of human coronaviruses. Herein, multiple B-cell epitopes and T-cell epitopes (CTL and HTL) were predicted from the S protein sequences of all seven known HCoV-229E. Post-prediction they were linked together with an adjuvant to construct a potential broad-spectrum vaccine candidate. Secondary and tertiary structures were predicted and validated, and the refined 3D-model was docked with an immune receptor. The vaccine candidate was evaluated for antigenicity, allergenicity, solubility, and its ability to achieve high-level expression in bacterial hosts. Finally, the immune simulation was carried out to evaluate the immune response after three vaccine doses. The designed vaccine is antigenic (with or without the adjuvant), non-allergenic, binds well with TLR-3 receptor and might elicit a diverse and strong immune response.

Keywords Covid-19 · Coronaviruses · Cheminformatics · Immunoinformatic · SARS-CoV-2 · Vaccine

Abbreviations

ACE-2	Angiotensin-converting enzyme 2
CTL	Cytotoxic T lymphocytes
HCoV-229E	Human coronavirus 229E
HTL	Helper T lymphocyte
RBD	Receptor binding domain
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
TLR-3	Toll-like receptor

Introduction

A recent study suggests that coronaviruses (CoVs) and bats have co-evolved for millions of years but CoVs seldom jumped across to infect humans (Joffrin et al. 2020). They were considered a comparatively harmless pathogen causing mild respiratory illness in humans (Song et al. 2019). But in the last two decades after the incidences of SARS infection in 2003 and MERS infection in 2013, the understanding regarding CoVs has changed. SARS-CoV-2 has been reported to target pneumocytes which cause severe respiratory distress characterized by bronchiolitis, bronchitis and pneumonia (Peele et al. 2020)

SARS-CoV-2 like SARS and MERS belongs to the *Coronaviridae* family. The presence of spikes on the outer surface of CoVs has a resemblance to the crown and hence the name corona (in Latin corona means crown) (Pandey et al. 2020). SARS-CoV-2 is an enveloped, positive-sense, single-stranded RNA beta-coronavirus with 26–32 thousand base pairs (de Wilde et al. 2018). Its genome encodes both structural like E (envelop) protein, N (nucleocapsid) protein, M (membrane) protein, S (spike) protein and non-structural proteins (nsp) like proteases (Mousavizadeh and Ghasemi

✉ Suvarna Ganesh Kini
suvarna.gk@manipal.edu

¹ Department of Pharmaceutical Chemistry, Manipal College of Pharmaceutical Sciences, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

² Manipal Mc Gill Centre for Infectious Diseases, Prasanna School of Public Health, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

2020). Out of all the four structural proteins, the S protein plays a prominent role in the binding of the SARS-CoV-2 to the host target receptors, especially through the ACE-2 (angiotensin-converting enzyme 2) receptor (Wrapp et al. 2020; Zhou et al. 2020). Cleavage of S protein by furin or other proteases has been reported in many of the avian and mammalian CoVs. The cleavage of S results in S1 which bears the receptor-binding domain (RBD) and S2 responsible for fusion (Wrapp et al. 2020).

Epitope-based vaccine approach has been employed to combat several diseases like a solid tumours (Knutson et al. 2001), malaria (Wang et al. 2004) and multiple sclerosis (Bourdette et al. 2005). Recently, many authors have reported epitope-based vaccine candidates for the mitigation of Covid-19 (Oany et al. 2014; Abdelmageed et al. 2020; Enayatkhani et al. 2020; Mukherjee et al. 2020; Naz et al. 2020; Abraham et al. 2020). Yang et al. have combined immunoinformatic methods with the deep learning algorithm to directly predict potential vaccine subunits from the S protein sequence of SARS-CoV-2 and constructed a multi-epitope vaccine for the SARS-CoV-2 virus (Yang et al. 2020b). Similarly, Singh et al. have designed multi-epitope vaccine peptides by targeting four different structural proteins (Spike, Envelop, Membrane and Nucleocapsid) of SARS-CoV-2 (Singh et al. 2020). Lin et al. have also employed several immunoinformatic tools to design a multi-epitope vaccine by targeting three different structural proteins of SARS-CoV-2 (Lin et al. 2020).

Employing modern computational tools, several compounds inhibiting different targets of SARS-CoV-2 have also been reported. Singh et al. have reported some 1,2,3-triazole scaffolds as potential main-protease (M^{pro}) inhibitors by using steered MD simulation studies (Singh et al. 2022). Employing *in-silico* studies, dicaffeoylquinic acid has also been reported as a potential S-protein and human ACE2

fusion inhibitor (Singh et al. 2021). Although several compounds have been identified and biologically evaluated for potential anti-SARS-CoV-2 activity, still it will take a long time for a new anti-COVID drug to come into the market. Computational studies help to reduce the initial timeline of drug discovery i.e., during discovery and pre-clinical stages but the clinical trials are still a lengthy stage. But in a short span of 2–3 years, several anti-COVID vaccines have been developed using various platforms and they have proved to be quite successful in containing the spread and severity associated with SARS-CoV-2. Moreover, the lesson learnt to develop a vaccine against this strain of the virus might come in handy in developing vaccines for different strains or different viruses in future.

In a short span of 2 years from the onset of Covid-19, several vaccines have been developed and used worldwide. Till now, ten vaccines have been approved for emergency usage by WHO (world health organization). For the first time novel vaccine platforms like RNA-based vaccines have been developed against a virus. A list of all the approved vaccines, their manufacturer and platform employed has been summarised in Table 1.

In the present study, we have targeted the S protein of SARS-CoV-2 along with all the seven known strains of human coronaviruses (HCoVs) i.e., both alpha (HCoV-NL63 and HCoV-229E) and beta coronaviruses (HCoV-OC43, HCoV-HKU1, MERS, SARS-CoV and SARS-CoV-2). As S protein is a common structural protein among the HCoVs, hence, we have taken up the S protein sequence to screen for potential epitopes. The objective of this study was to develop a broad-spectrum vaccine candidate capable of conferring immunity against all the seven known strains of HCoVs. Our hypothesis of identifying an epitope capable of providing immunity against various CoVs is also supported by the recent work of Mateus et al., where they have identified

Table 1 Vaccines that have been approved by WHO for emergency use

Sr. no	Vaccine	Type of vaccine/vaccine platform used	Manufacturer/developer
1	CoronaVac	Inactivated virus	SinoVac
2	Covilo	Inactivated virus	Sinopharm
3	Covaxin	Inactivated virus	Bharat Biotech
4	Covishield (Oxford/ Astra-Zeneca formulation)	Non-replicating viral vector	Serum Institute of India
5	Vaxzevria	Non-replicating viral vector	Oxford/AstraZeneca
6	Ad26.COV2.S	Non-replicating viral vector	Janssen (Johnson & Johnson)
7	Comirnaty	RNA	Pfizer/BioNTech
8	Spikevax	RNA	Moderna
9	COVOVAX (Novavax formulation)	Protein subunit	Serum Institute of India
10	Nuvaxovid	Protein subunit	Novavax

cross-reactive T-cell epitopes of SARS-CoV-2 in unexposed humans. They have concluded that pre-existing immunity could be derived from exposure to other HCoV strains like HCoV-OC43, HCoV-229E, HCoV-NL63, or HCoV-HKU1, which cause the common cold (Mateus et al. 2020). Yang et al. have reported that a vaccine candidate targeted towards the RBD of the S protein of SARS-CoV-2 led to the induction of protective immunity (Yang et al. 2020a). A flowchart summarizing the methodology and tools employed for the current work has been depicted in Fig. S2.

Materials and methods

Computational configuration of the *in-silico* studies

All the immunoinformatic exercises were carried out using free web-based tools. For cheminformatics experiments like molecular docking and molecular dynamics simulations were carried out by using Maestro (Release 2018–3) which is a commercial small-molecule drug discovery suite from Schrodinger Inc. (USA) on an HP computer with Linux Ubuntu 18.04.1 LTS operating system.

Retrieval of S protein sequences

For the present study, we have taken the sequences of S protein from seven strains of coronaviruses (CoVs): (i) NL63 (ii) 229E (iii) OC43 (iv) HKU1 (v) MERS (vi) SARS-CoV and (vii) SARS-CoV-2. The sequences were retrieved in the FASTA format from the UniProtKB database (Apweiler et al.).

Prediction of T-cell epitopes

Herein, we have predicted 9-mer CTL epitopes for each of the HCoV strains employing the NetMHCpan4.1 web server which is also recommended by IEDB (Immune Epitope Database) (Reynisson et al. 2020). The NetMHCpan-4.1 server employs artificial neural networks (ANNs) to predict the binding of epitopes to any MHC (major histocompatibility complex) molecule of known sequence. Its prediction method has been developed using a combined training set of > 850,000 quantitative BA (Binding Affinity) and Mass-Spectrometry EL (Eluted Ligands) peptides (Reynisson et al. 2020). In the present study, the HLA-A*24:02 supertype of the human leukocyte antigen (HLA) allele has been employed for all the predictions. The rationale behind selecting this supertype allele was the availability of a high-resolution X-ray crystal structure (PDB ID 3I6L) of HLA-A*24:02 in a complex with an epitope N1 derived from SARS-CoV N protein. The other reason behind the selection of this supertype was that according to Lu et al.,

one of the most common HLA-A alleles found in the Asian population is HLA-A*24:02 (Lu et al. 2005).

The best two 9-mer peptides from each HCoV strain were selected based on their binding affinity (nM) and percentile rank. Thus, a total of 14 CTL epitopes were selected for further study. But we wanted to refine our selection of CTL epitopes further and hence these epitopes were put for epitope conservancy analysis by employing the IEDB tool for the epitope conservancy analysis (Bui et al. 2007). For the present work, linear epitope conservancy analysis was done with a sequence identity threshold of 100%. The epitopes which were conserved in all the strains (reflected by their minimum sequence identity) were taken up for further analysis. The higher the value of minimum percentage identity, the higher the chances of those epitopes being conserved among all the strains. Further, these 14 epitopes were put for molecular docking studies using peptide docking under the Biologics suite of Maestro (Release 2018–3). The X-ray crystal structure of HLA-A*24:02 protein (PDB ID 3I6L) was retrieved from the protein data bank (PDB) (Berman et al. 2000; Liu et al. 2010). The retrieved protein structure was prepared using the protein preparation wizard of Maestro and the optimised protein structure was energetically minimized by employing the OPLS3e force field (Sastry et al. 2013; Kumar et al. 2019). The epitope sequences were imported into the workspace and using peptide builder peptides were generated. The generated peptides were energetically minimized and docked against the receptor grid generated at the binding site of epitope N1 derived from SARS-CoV N protein on the HLA-A*24:02 protein. The number of poses to return for each docking run was fixed at 10 as peptides are very flexible hence, several docking runs were used for each peptide. Each docking run employed a different conformation of the epitope to enhance the sampling (Tubert-Brohman et al. 2013). The MM-GBSA (molecular mechanics generalized Born model and solvent accessibility) was used as the scoring method which is more accurate than the Glide score method but less fast (Li et al. 2011).

The epitopes were ranked based on the docking score and the highest-ranked CTL epitope was further put for molecular dynamics (MD) simulations analysis by using the Desmond module of the Maestro suite (Bowers et al. 2006). The docked complex of the selected epitopes with the HLA-A*24:02 protein (PDB ID 3I6L) was first put for the system builder where the Na⁺ and Cl⁻ were added to neutralize the charges. A total of 9 Na⁺ ions were added. SPC (simple point charge) was used as the solvent system in an orthorhombic boundary box shape. The solvated system of epitope and the HLA-A*24:02 protein was put for minimization by using the SD (steepest descent) method. The convergence threshold was fixed at 1 kcal/mol/Å with a maximum iteration of 2000. A slow relaxation protocol

was followed for equilibration at a temperature of 300 K and pressure of 1.01325 bar. Nose–Hoover thermostat was used to maintain the temperature while Martina–Tobias–Klein was the barostat employed. The minimized protein-epitope complex was simulated for 50 ns to analyse the stability of the epitope-protein complex under biological conditions. Based on the above methodologies i.e., percentile rank, binding affinity, minimum percentage conservancy and docking score, a total of ten CTL epitopes were selected for the design of vaccine candidate.

Prediction of linear B-cell epitope and helper T-cell (HTL) epitopes

B-cell epitopes are a specific part of the antigen recognized by the B lymphocytes of our immune system. For the present work, the BepiPred-2.0 webserver has been employed to predict linear B-cell epitopes of 14 sequence lengths for all seven HCoV strains. BepiPred-2.0 is superior to other sequence-based linear epitope prediction tools and employs RF (random forest) algorithm trained on peptides annotated from antigen–antibody protein complexes (Jespersen et al. 2017). HTL epitopes of 15 sequence lengths were predicted using the NetMHCII 2.2 Server for all seven strains (Reynisson et al. 2020). The NetMHCII 2.2 server employs artificial neuron networks to predict the binding of epitopes to HLA-DQ, HLA-DP, and HLA-DR alleles (Nielsen and Lund 2009).

Multi-epitope vaccine candidate construction and its conservancy analysis

The multi-epitope vaccine candidate was designed by linking different epitopes and adjuvants. CTLs epitopes were linked through an AAY linker while HTL and B-cell epitope was linked using a GPGPG linker. At the N-terminal, a 50 S ribosomal protein L7/L12 (Locus RL7_MYCTU) was added as an adjuvant (Accession no. P9WHE3) through an EAAAK linker. The adjuvant sequence was retrieved from the UniProt database (Bateman et al. 2017). At the carboxy-terminal end, a 6x-His tail was also added to generate a final vaccine construct. The final designed vaccine peptide was put for conservancy analysis without the adjuvant across the seven strains of the selected S-protein sequences of the HCoVs (Bui et al. 2007).

Prediction of IFN-gamma-inducing epitope

Cytokines like Interferon-gamma (IFN- γ) play a crucial role in the immune responses by stimulating NK (natural killer) cells and macrophages and providing an enhanced response to MHC antigen. Herein we have employed the IFNepitope server to predict 15-mer IFN- γ epitopes for the

designed vaccine candidate. IFN- γ epitopes were predicted separately for the adjuvant and the main vaccine peptide due to the limitation on the number of sequences that can be used as input in the server. The server employs a support vector machine (SVM) hybrid approach for the prediction of IFN- γ epitopes (Dhanda et al. 2013).

Antigenicity, allergenicity and assessment of physicochemical properties of the vaccine construct

Antigenicity prediction for the vaccine construct was carried out using ANTIGENpro (Magnan et al. 2010) and the VaxiJen2.0 server. Like ANTIGENpro the antigenicity prediction by VaxiJen v2.0 is alignment-free and employs the physicochemical parameters of the protein (Doytchinova and Flower 2007). Allergenicity of the designed vaccine construct was predicted using AllerCatPro and AllergenFP servers. AllerCatPro predicts the allergenicity of the query proteins by comparing their 3D structure along with their amino acid sequence with a dataset of 4180 unique protein sequences which have been reported to be allergenic (Maurer-Stroh et al. 2019). To predict the allergenicity of proteins, AllergenFP employs a descriptor-based fingerprint approach and thus is an alignment-free tool (Dimitrov et al. 2014). The physicochemical parameters like theoretical pI, in vitro and in vivo half-life, amino acid composition, instability index, molecular weight, aliphatic index and GRAVY (grand average of hydropathicity) of the vaccine construct were analysed by the ProtParam web server (Gasteiger et al. 2005). The solubility of the vaccine construct was analysed by the Protein-Sol web server (Hebditch et al. 2017).

Secondary and tertiary structure prediction of the vaccine construct

The secondary structure of the designed vaccine construct was predicted by employing PSIPRED and RaptorX Property web servers. PSIPRED 4.0 performed Position-specific iterated BLAST (Psi-Blast), for the identification of sequences that showed considerable homology to the vaccine peptide (McGuffin et al. 2000). These sequences were selected to build a position-specific scoring matrix. The RaptorX Property web server employs a template-free approach to predict the secondary structure properties of the query protein. Its algorithm is based on DeepCNF (Deep Convolutional Neural Fields) to simultaneously predict secondary structure (SS), disorder regions (DISO) and solvent accessibility (ACC) (Wang et al. 2016). The tertiary structure of the designed multi-epitope vaccine candidate was predicted by the I-TASSER server webserver (Yang and Zhang 2015). The I-TASSER (Iterative Threading Assembly Refinement) server is an integrated platform to predict 3D structure from

a peptide sequence. It generates 3D atomic models from multiple threading alignments and iterative structural assembly simulations and has been recognized as the best server for the prediction of protein structure (Roy et al. 2010). The best 3D-model for the vaccine construct obtained from the I-TASSER webserver was refined first by the ModRefiner and then by using the GalaxyRefine server. The refinement of protein structures by the ModRefiner server is based on a two-step, atomic-level energy minimization which leads to improvements in both local and global structures (Xu and Zhang 2011). GalaxyRefine method first rebuilds side chains and then carries out side-chain repacking and finally employs MD simulation for overall structural relaxation (Heo et al. 2013). Finally, the refined 3D model of the vaccine construct was validated using ProSA-web (Wiederstein and Sippl 2007) and ERRAT server (Colovos and Yeates 1993). Ramachandran plot was obtained from the protein structure quality tool of the Prime module in Maestro1.4.

Protein–protein docking studies

Molecular docking was carried out to understand the binding of the vaccine candidate with the TLR-3 receptor as the immune response depends upon the interaction between an antigenic molecule (here the designed vaccine candidate) and an immune receptor (in this case TLR-3). The first step in this milieu was the prediction of the binding sites of the proteins which was carried out using the CASTp server (Binkowski et al. 2003). Molecular docking of the vaccine candidate with the TLR3 (PDB ID: 1ZIW) (Choe et al. 2005) receptor was carried out using the HADDOCK 2.4 webserver (Van Zundert et al. 2016) and GRAMM-X Simulation webserver (Tovchigrechko and Vakser 2006). The interactions were visualized using LIGPLOT v.4.5.3 (Wallace et al. 1995). Finally, the binding energy of the top-ranked docking pose of the vaccine candidate-TLR4 complex was predicted using the PRODIGY webserver (Xue et al. 2016).

Prediction of discontinuous B-cell epitopes

It has been reported that more than 90% of B-cell epitopes are discontinuous, i.e. they comprise distantly separated segments in the pathogen protein sequence which are brought close by the folding of the protein (Barlow et al. 1986). Herein, we employed ElliPro for the prediction of discontinuous B-cell epitopes of the validated 3D structure of the vaccine candidate (Ponomarenko et al. 2008).

In-silico cloning and codon optimization of the vaccine construct

Java Codon Adaptation Tool (JCat) server was employed for reverse translation and codon optimization so that the

designed multi-epitope vaccine candidate could be expressed in a selected expression vector (Grote et al. 2005). Codon optimization was carried out to express the final vaccine candidate in the *E. coli* (strain K12) host. The output of the JCat server comprises the codon adaptation index (CAI) and percentage GC content. CAI score reflects codon usage biases and ideally, the value of CAI should be 1.0 but a score above 0.8 is a good score (Grote et al. 2005). The gene sequences of the designed vaccine candidate were optimized, and two restriction sites (*Nde* I and *Xho*) were introduced at the C and N-terminals of the sequence. In the last step, the optimized gene sequence of the vaccine construct along with the inserted restriction sites was inserted into the pET-28a (+) vector using the SnapGene programme.

Immune simulation

In-silico immune simulations were carried out for further characterization of the immunogenicity and immune response profile of the designed multi-epitope vaccine candidate. Herein, we employed the C-ImmSim server for immune simulation studies which is an agent-based model that employs a position-specific scoring matrix (PSSM) (Rapin et al. 2010). Three injections were given at four weeks intervals by keeping all simulation parameters at the default setting with time steps set at 1, 84, and 168 (each time step is 8 h and time step 1 is injection at time = 0).

Results and discussion

Prediction of T-cell epitopes

The goal of T-cell epitope prediction is to identify peptides with the shortest sequence length within an antigen that is capable of stimulating either CD4 or CD8 T-cells (Ahmed and Maeurer 2009). Antigen-presenting cells (APCs) present T-cell epitopes on the surface where they bind to either MHC class I protein or MHC-II proteins. It has been reported that MHC-I protein presents T-cell epitopes of 8–11 amino acid residue length while MHC-II presents longer peptides in the range of 13–17 amino acid residues (Steers et al. 2014). T-cell epitopes bound to MHC I protein are recognized by CD8 T-cells which become CTL (cytotoxic T lymphocytes), while those presented by MHC-II are recognized by CD4 T-cells and become helper T-cells (Sanchez-Trincado et al. 2017). Many authors have reported these tools for CTL epitope prediction not only against SARS-CoV-2 but also against SARS-CoV (Oany et al. 2014; Enayatkhani et al. 2020; Naz et al. 2020; Abraham et al. 2020; Panda et al. 2020). But this is the first such analysis involving seven strains of the HCoV. Our present hypothesis to screen the S protein sequences of all the seven reported HCoVs for

identifying CTL epitopes as a potential vaccine candidate against SARS-CoV-2 is also supported by the recent findings of Barun et al., who have demonstrated the presence of S-cross-reactive T cells in unexposed healthy individuals probably because of previous exposure to endemic coronaviruses like 229E or OC43 (Braun et al. 2020).

A total of 142 CTL (9-mer) epitopes were predicted to be strong binders for the seven selected S proteins of different HCoV strains using the NetMHCpan4.1 server with default settings. We selected the best two epitopes based on their percentile rank, binding affinity (nM) and binding level for each of the HCoV strains. Thus, a total of fourteen 9-mer CTL epitopes were selected for further study. The conservancy analysis of these fourteen CTL epitopes showed that the QYIKWPWYI epitope from the SARS-CoV-2 strain had a minimum conservancy of 66.67% across all the strains. Another epitope with sequence YYNKWPWYI from the MERS strain showed a minimum conservancy of 55.56% across all the strains. Hence, these two CTL epitopes were selected for the final vaccine design. Peptide docking analysis of these epitopes with HLA-A*24:02 protein showed that these peptides showed favourable binding with the protein. But we selected only one CTL epitope (NYNYLYRLF) which showed the highest docking score of -12.759 kcal/mol. MD simulations study for this epitope showed that the epitope-protein complex was stable throughout the simulations period as reflected by their RMSD (root mean square deviation) plot in Fig. 1. A histogram representing the interaction of the epitopes with the amino acid residues of the protein was also plotted as shown in Fig. 2. The non-bonding interactions were similar to the ones observed during peptide

docking studies, but water bridges were formed with LYS66, HIS70, ASP74, TYR84, THR143 and GLN155 residues. A 2D interaction diagram (Fig. 3) suggested that apart from intermolecular H-bond interactions, there were quite several intramolecular H-bond interactions that could make the epitope-protein complex more stable. The importance of intramolecular H-bonding for the epitopes binding to HLA-A*24:02 protein has also been highlighted by Liu et al. (Liu et al. 2010). Several other authors like Abraham et al. and Enayatkhani et al. have also employed MD simulation studies to analyse the stability of the designed epitopes and protein complex (Enayatkhani et al. 2020; Abraham et al. 2020). Hence, for the final vaccine design, a total of ten CTL epitopes (Table 2) were selected out of which seven were the best-ranked epitope from each strain, and two were the best-conserved epitope across all the strains and one was the best-docked epitope with HLA-A*24:02 protein.

Prediction of linear B-cell epitope and Helper T-cell (HTL) epitopes

A total of seven 14-mer B-cell epitopes (Table 2) were selected based on the epitope threshold and relative surface accessibility as predicted by the BepiPred2.0 server. A total of seven 15-mer high-binding MHC-II epitopes (Table 2) for human alleles HLA-DR, HLA-DQ and HLA-DP as predicted by the NetMHCII 2.2 webserver was selected. Some of the B-cell epitopes and HTL epitopes overlapped.

Fig. 1 RMSD plot of CTL-epitope 2 in complex with HLA-A*24:02 protein, where blue colour shows protein RMSD while red colour shows ligand RMSD

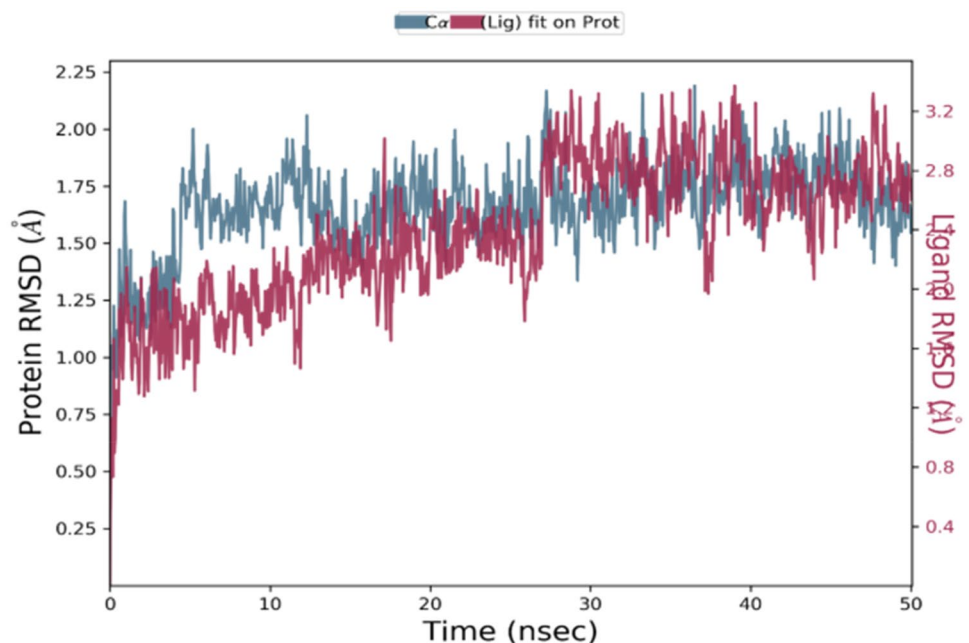


Table 2 List of the selected CTL, HTL and linear B-cell epitopes from S protein sequences of different HCoV strains for constructing the multi-epitope vaccine candidate

HCoV strains	CTL epitopes (9-mer)	B-cell epitopes (14-mer)	HTL epitopes (15-mer)
229E	FYINGYRYF	PDGFYSTSPIQPVE	VQVEYLQITSTPIVV
NL63	FYINGFKYF	ITGVYPVSGIREF	VSTFVGLPPTVREI
OC43	TYNSWQNL	KNRRSRAITTYGR	QSYKGKIVLPLLSE
HKU1	VYLNTLLF	LGINEEKCGTQLNH,	QSFNGKIVLPPILSE
MERS	LYGGNMFQF YYNKWPWYI*	IADPGYMQGYDDCM	TIKYYSIIPHSIRSI
SARS	VYSTGVNVF	NVSKGIYQTSNFRV	IVAYTMSLGAENSIA
SARS-CoV-2	VYSTGSNVF QYIKWPWYI* NINYLYRLF#	APATVCGPKKSTNL	IVAYTMSLGAENSIA

*CTL epitopes selected based on the epitope conservancy analysis

#CTL epitope selected based on the docking score

Multi-epitope vaccine candidate construction and its conservancy analysis

The final multi-epitope vaccine constructs comprised 10 CTL epitopes, 7 linear B-cell epitopes, and 7 HTL epitopes. The predicted CTL epitopes were joined by using an AAY linker while B-cell epitope and TL epitopes were fused with GPGPG linkers. The 50S ribosomal L7/L12 (Locus RL7_MYCTU) was selected as the adjuvant to accentuate antigen-specific immune responses and fused at the N-terminal by using the EAAAK linker. Finally, to help in protein identification and purification a 6xHis tail was inserted at the carboxy-terminal end of the vaccine peptide. Thus, the final multi-epitope vaccine candidate was constructed with a total of 532 amino acid residues derived from 24 merged multi-epitopes. The final designed vaccine peptide was put for conservancy analysis without the adjuvant across the seven strains of the selected S-protein sequences of the HCoV strains. The sequences of the vaccine candidate showed a minimum and maximum conservancy of 73.97% across all the strains of the HCoVs. A higher degree of conservation across all the HCoV strains might confer broader protection across multiple strains by recognition of the conserved epitopes by our immune system. Hence, we hypothesize that this vaccine construct which has more than 73% of sequences conserved across the seven HCoV strains might be a potential broad-spectrum vaccine candidate against HCoVs.

Prediction of IFN-gamma-inducing epitope

Due to the restriction on the number of residues that can be used as input in the IFNepitope server, IFN- γ inducing epitopes were predicted separately for the adjuvant and main vaccine peptide. For the adjuvant, a total of 127 potential IFN- γ inducing epitopes (15-mer) were predicted. For the main vaccine peptide, 389 potential 15-mer epitopes were predicted out of which 82 had a positive prediction score.

Antigenicity, allergenicity and assessment of physicochemical properties of the vaccine construct

As per the predictions of the VaxiJen 2.0 server, the antigenicity of the vaccine constructs along with the adjuvant sequence was found to be 0.7106 in a bacteria model at a threshold of 0.5 while the main vaccine sequence without the adjuvant showed a score of 0.8135. As per the predictions of the ANTIGENpro server the probable antigenicity of the vaccine candidate with the adjuvant was 0.829721 while without the adjuvant was 0.869507. This result indicates that the designed vaccine candidates are antigenic (with or without adjuvant) in nature. The main vaccine sequence seems to be more antigenic without the adjuvants as per the predicted scores. The allergenicity of the vaccine candidate was predicted by the AllerCatPro and AllergenFP servers and both predicted it to be non-allergenic. The molecular weight (MW) of the final vaccine candidate was predicted to be 56.5 kDa. The theoretical isoelectric point (pI) value was predicted to be 7.11. The half-life was predicted to be 30 h in mammalian reticulocytes (in vitro), > 10 h in *E. coli* and > 20 h in yeast (in vivo). The instability index (II) was computed to be 22.81 which classifies the vaccine candidate to be stable. It had an aliphatic index of 79.29 which suggests it to be thermostable (Ikai 1980). The GRAVY score for the vaccine candidate was computed to be -0.407 which indicates that it could be hydrophilic (Ali et al. 2017). The predicted scaled solubility value was computed to be 0.432 which indicates that the vaccine construct would have an acceptable solubility profile.

Secondary and tertiary structure prediction of the vaccine construct

The secondary structure of the final multi-epitope vaccine peptide was computed, and it was predicted to have 20% alpha-helix, 20% beta-sheet and 59% coil (Fig. 4). The

solvent accessibility of amino acid residues as predicted by the RaptorX property server showed that 49% of the total residues were exposed, 24% were medium exposed and 26% was buried. A total of 42 residues i.e., 7% of residues were computed to be in the disordered region. Five tertiary structure models of the final vaccine construct were predicted by the I-TASSER server based on 10 threading templates, of which PDB hits 1dd4A, 5tsjN, 1rquA, 2nbiA, 1dd3A, 2ocwA and 2ftc were the best hits. All of the 10 selected templates showed good alignment which was evident from their *Z-score* values which ranged from 1.02 to 5.35. The first model (Fig. 5a) was selected for further refining which had an estimated C-score of 0.77, TM score of 0.62 ± 0.14 and RMSD value of 9.2 ± 4.6 Å. The C-score range is between -5 and 2 , where higher values towards

2 indicate higher confidence. The TM-score is an indication of the structural similarity between two structures and unlike RMSD, it is not sensitive to local error (Zhang and Skolnick 2004). From the TM value of the selected model, we can infer that the model has the correct topology. This selected model was further put for refining first by ModRefiner and then by GalaxyRefine server which generated five models (Table 3). According to the model quality scores (Table 2) for all the five refined models, model 3 (Fig. 5b) was found to be the best model. It had a GDT-HA score of 0.9408, RMSD value of 0.443, MolProbity score of 2.072, clash score of 11.3 and a poor rotamers score was 0.2. The Ramachandran plot score for model 3 was predicted to be 91.5% which was less when it was put for validation by the Ramachandran plot analysis. The Ramachandran plot of the

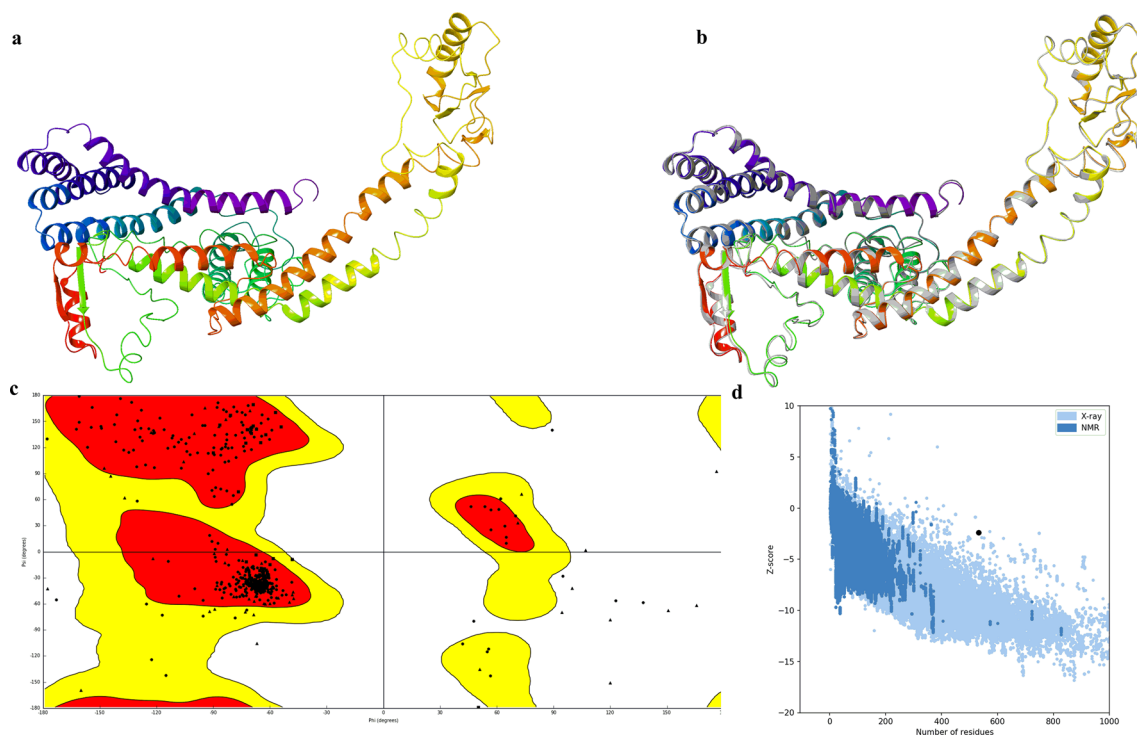


Fig. 5 Tertiary structure modelling, refining and validation. **a** The final homology 3D model of the vaccine construct as computed by the I-TASSER. **b** Refinement of the selected I-TASSER 3D model by GalaxyRefine where the refined 3D model (coloured) has been super-

imposed on the initial crude model (grey). **c** Ramachandran plot of the refined 3D model of the vaccine candidate **d** Plot by ProSA-web showing the Z-score

Table 3 Detailed result of the five models refined by the GalaxyRefine server

Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favoured
Initial	1.0000	0.000	3.083	61.4	2.0	88.1
MODEL 1	0.9314	0.467	2.126	12.6	0.5	91.1
MODEL 2	0.9305	0.476	2.111	11.6	0.5	90.6
MODEL 3	0.9408	0.443	2.072	11.3	0.2	91.5
MODEL 4	0.9375	0.465	2.180	14.2	0.5	90.9
MODEL 5	0.9337	0.457	2.122	12.1	0.5	90.8

selected model 3 (Fig. 5c) suggests that it's 94.528% of the residues are in the favourable region with 2.64% of residues in the allowed region. The overall quality factor was computed by the ERRAT server and for the selected model 3, it was predicted to be 86.235. The ProSA-web computed the Z-score for model 3 and it was found to be -2.41 (Fig. 5d).

Protein–protein docking studies

The CASTp 3.0 server was employed to compute the protein binding and hydrophobic interaction sites on the surface of the TLR-3 receptor and the final refined 3D model of the vaccine construct. A total of 105 binding pockets were predicted for the vaccine candidate with different molecular surface areas and volumes. Similarly, for the TLR-3 protein (PDB ID 1ZIW), a total of 77 binding pockets were predicted. For the present work, a pocket (Fig. 6a) with a molecular surface area of 2125.000 \AA^2 and volume of $13,438.169 \text{ \AA}^3$ was selected for the TLR-3 receptor. For the vaccine candidate, the selected pocket (Fig. 6b) had a molecular surface area of 917.628 \AA^2 and a volume of 2414.204 \AA^3 . Both the proteins were put for molecular docking simulations first using the GRAMM-X server which led to the generation of ten binding poses. The poses were visualised using the DIMPLOT of LIGPLUS version 4.5.3 and the best-docked pose of the vaccine-TLR-3 complex in 3D has been depicted in Fig. 7a and b. The non-bonding interactions between the docked protein complex were also visualized in 2D and have been shown in Fig. 8. It was observed that ARG306, TYR526, ANS521 and GLY46 residues of the vaccine construct showed hydrogen bonding with TRP273, ANS247, LEU243 and HIS432 residues of the TLR-3

protein respectively. The 2D interaction plot also showed a possible salt bridge between GLU307 and GLY46 of the vaccine construct with the LYS272 and SER481 of the TLR-3 protein. Further, the best-docked pose was put for the prediction of binding affinity (ΔG) and dissociation constant (K_d). The binding energy (ΔG) for the vaccine-TLR-3 complex was computed to be -20.8 kcal/mol with a K_d value of $5.9 \times 10^{16} \text{ M}$ at $25.0 \text{ }^\circ\text{C}$. The Prodigy server also calculated the number of interfacial contacts (ICs) per property within the threshold distance of 5.5 \AA and its ICs charged-charged, ICs polar-polar, and ICs apolar-apolar were computed to be 14, 7 and 64 respectively. It suggests that most of the interfacial contacts were hydrophobic.

The HADDOCK2.4 server clustered 103 structures in 14 clusters, which represent 51% of the water-refined model generated by the HADDOCK. The clusters were ranked on several parameters and out of the 14 clusters, cluster 3 (cluster size 10) was reported to be the best cluster. It had a HADDOCK score of -79.4 ± 7.1 , RMSD from the overall lowest-energy structure was 43.8 ± 0.4 , Van der Waals energy $-100.6 \pm 15.7 \text{ kcal/mol}$, Electrostatic energy $-415.7 \pm 78.8 \text{ kcal/mol}$, Desolvation energy $-27.2 \pm 4.0 \text{ kcal/mol}$ and Z-Score -1.5 . The detailed result for all the clusters has been depicted in Figs. 9 and 10. The best pose from cluster 3 was also put for the prediction of binding affinity (ΔG) and dissociation constant (K_d). The binding energy (ΔG) was computed to be -20.3 kcal/mol with a K_d value of $1.3 \times 10^{15} \text{ M}$ at $25.0 \text{ }^\circ\text{C}$ which was similar to that of the best docked pose generated by the GRAMM-X server. The ICs charged-charged, ICs polar-polar, and ICs apolar-apolar were computed to be 16, 7 and 25 respectively.

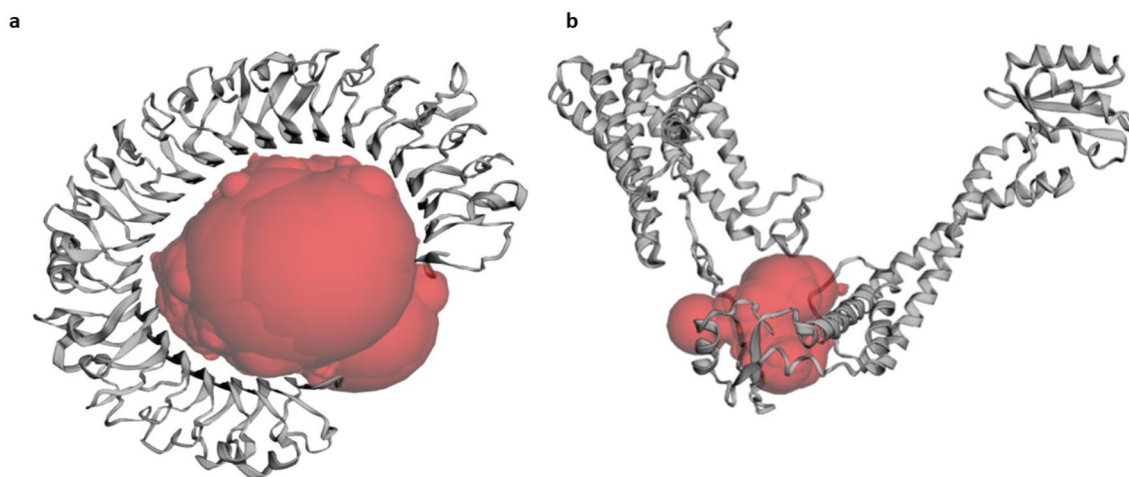


Fig. 6 The selected binding pocket of the proteins as predicted by the CASTp3.0 server. **a** The predicted binding pocket of the TLR-3 protein. **b** The predicted binding pocket of the vaccine candidate

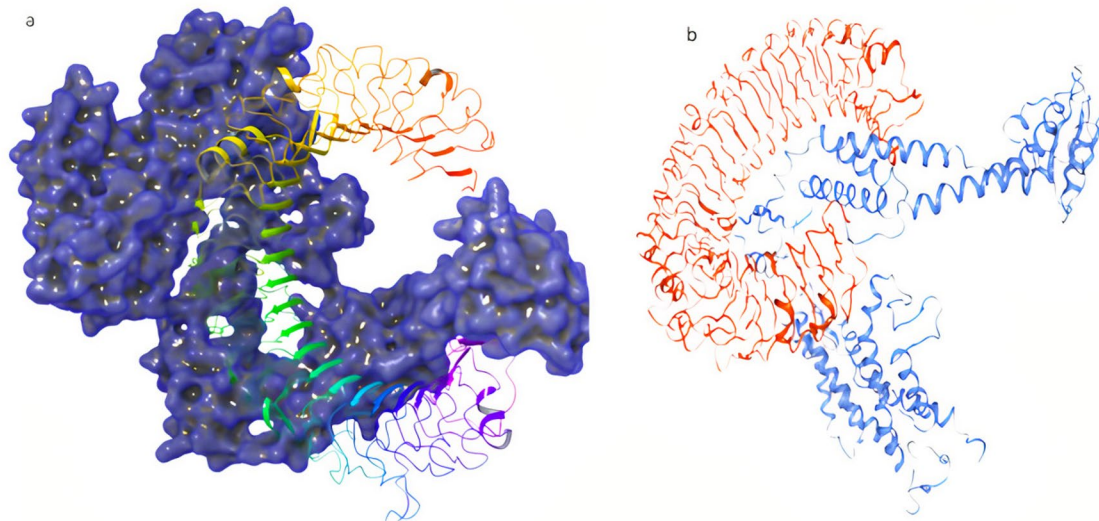


Fig. 7 Visualization of the best-docked poses for the vaccine-TLR-3 receptor complex. **a** 3D-interaction diagram of the protein-protein complex generated by the GRAMM-X server where blue mesh structure represents the vaccine candidate while coloured ribbon shows

the TLR-3 receptor. **b** 3D-interaction diagram of the protein-protein complex generated by the HADDOCK2.4 server where the blue ribbon represents the vaccine candidate while the orange ribbon shows the TLR-3 receptor

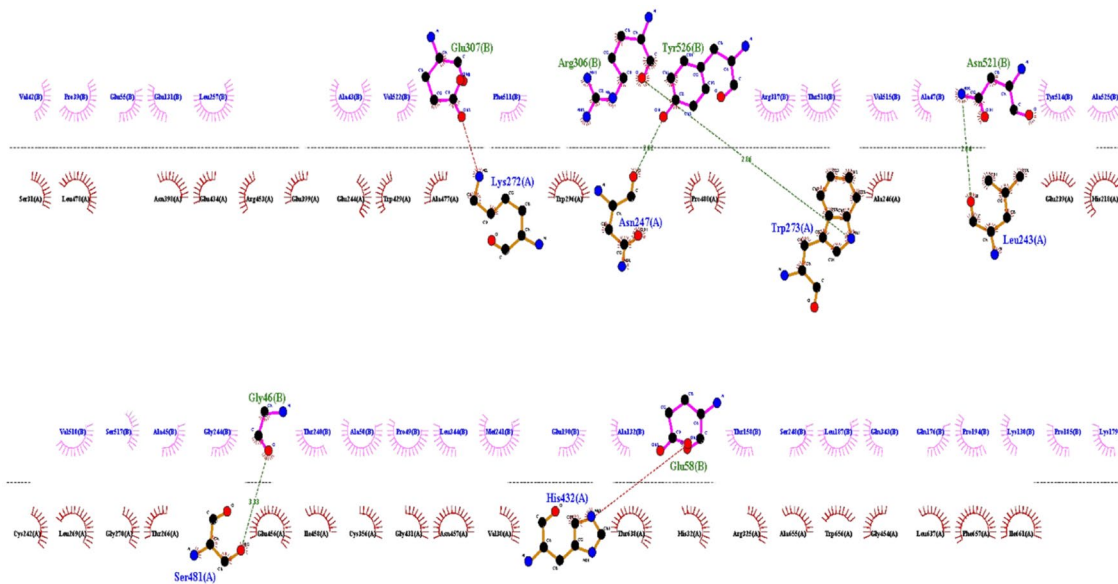


Fig. 8 2D-interaction diagram of the protein-protein complex generated by the GRAMM-X server where chain A is from the TLR-3 receptor and chain B is the vaccine construct

Prediction of discontinuous B-cell epitopes

The ElliPro server predicted a total of 7 discontinuous B-cell epitopes of the validated 3D structure of the vaccine candidate comprising 270 residues. The conformational B-cell epitopes had scores ranging from 0.598 to 0.875 where the best epitope had only three residues while the last ranked epitope comprised of 8 residues. This study suggests that the designed

multi-epitope vaccine might lead to heightened immune response through the conformational B-cell epitopes also.

In-silico cloning and codon optimization of the vaccine construct

Immunoreactivity through serological assays is needed to validate a designed vaccine candidate and the first step for it

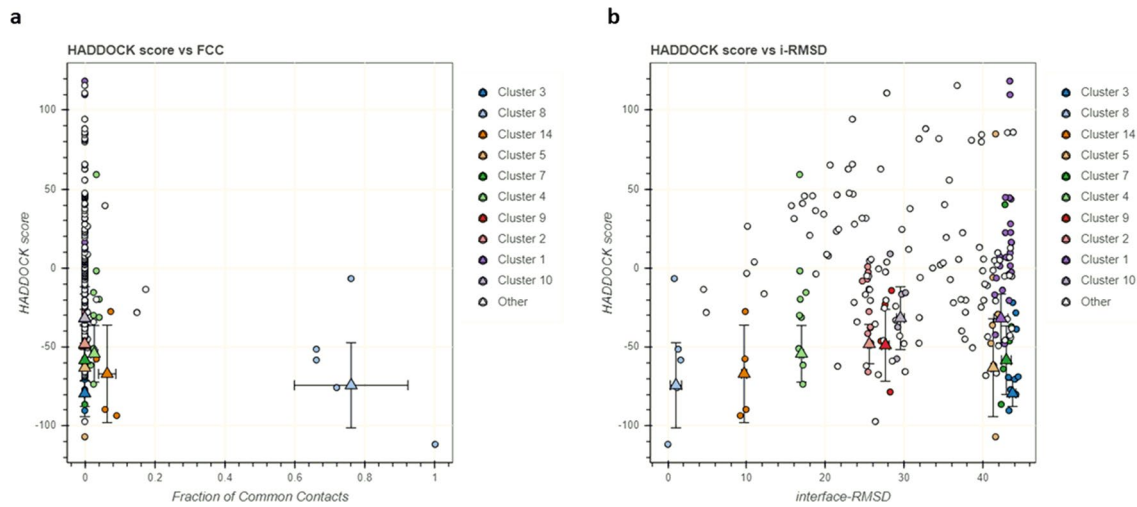


Fig. 9 Bokeh plot showing the result of various parameters for all the clusters by the HADDOCK2.4 server. **a** Plot between HADDOCK score and FCC (fraction of common contacts). **b** Plot depicting HADDOCK score vs interface-RMSD

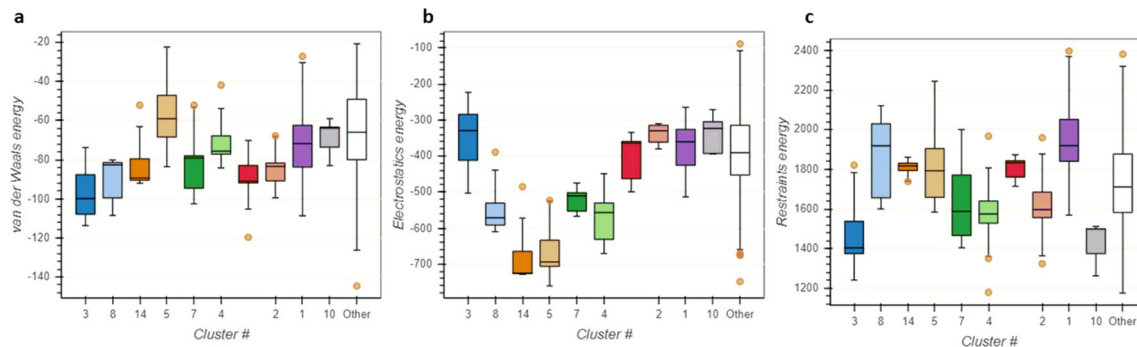


Fig. 10 Bokeh plot depicting the analysis of various energy parameters for all the clusters as predicted by the HADDOCK2.4 server. **a** Van der Waals energy **b** Electrostatics energy **c** Restraints energy

is the expression of the multi-epitope vaccine candidate in a suitable host (Gori et al. 2013). *Escherichia coli* expression systems have been widely used to produce recombinant proteins and hence to achieve maximal protein expression, we selected the *E. coli* (strain K12) system for the codon optimization by the Java Codon Adaptation Tool (JCat) (Chen 2012; Rosano and Ceccarelli 2014). The optimized codon sequence for the final vaccine candidate was 1596 with a CAI value of 0.9237 and 53.57% of GC content. Ideally, the GC content should be in the range of 30% to 70% and hence, our vaccine candidate might show high-level expression in the bacterial host system. Finally, the optimized gene sequences of the designed vaccine candidate with two restriction sites (*Nde* I and *Xho*) at the C and N-terminals of the sequence were inserted into the pET-28a (+) vector using SnapGene software to generate the sequences of the recombinant plasmid (6892 bp) as shown in the supplementary file (Fig. S1).

Immune simulation

The immune stimulation by the C-ImmSim server showed results that are consistent with actual immune responses. The primary response after the first dose of the vaccine was characterized by increased levels of IgM as shown in Fig. 11a. After the second and third doses of the vaccine, there was a remarkable increase in the levels of IgG1, IgG1 + IgG2, IgM, and IgG + IgM antibodies (Fig. 11a). The level of IgM + IgG peaked after the third dose to around 2,30,000 and IL-2 level (after 2nd dose) was around 9,00,000 which suggests a strong humoral immune response which is better than reported by Abraham et al. (around 2,00,000 for IgM + IgG after 3rd dose and IL-2 level was around 7,50,000 ng/ml after 2nd dose) for multi-epitope vaccine design against SARS-CoV-2 (Abraham et al. 2020). Similarly, when compared to Shey et al. for multi-epitope vaccine candidates (the level of IgM + IgG was around

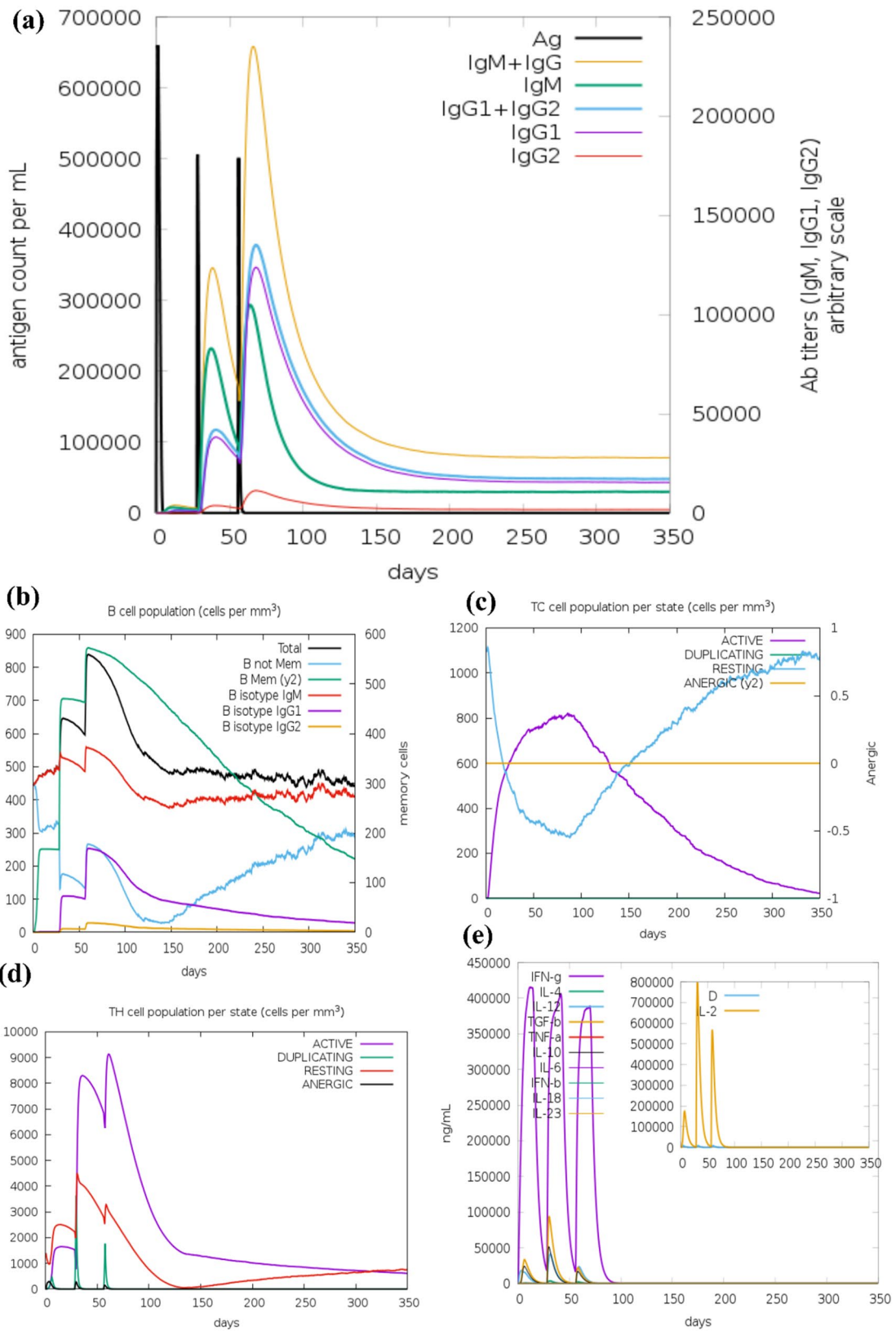


Fig. 11 Results of immune simulation by the C-ImmSim server. **a** Immunoglobulin and immunocomplexes production in response to antigen injections at different time intervals (black vertical lines); specific subclasses have been shown as coloured peaks. **b** The increase in the populations of B-cell post three doses of the vaccine. **c** The evolution of T-cytotoxic cell population per state post three vaccine doses. **d** The evolution of T-helper cell population per state post three vaccine doses. **e** The main plot depicts the levels of cytokine levels post three vaccine doses. The insert plot depicts the IL-2 level with the Simpson index, D which is indicated by the green dotted line

6,50,000) against onchocerciasis our designed vaccine-elicited stronger immune response (Shey et al. 2019). It was also observed that after the second exposure the population of the antigen was also falling which indicates increased clearance by the immune system. B-cell population (Fig. 11b) and especially the B-memory cell population (> 800 cells per mm^3) also increased and peaked after the third dose. Similarly, a greater response was observed in the cytotoxic T-cells (Fig. 11c) and helper-T cell populations (Fig. 11d) with corresponding memory development which lasted for many months. It was also observed that the levels of IFN- γ and IL-2 (Fig. 11e) increased significantly after the first dose of the vaccine and maintained their peak levels after multiple exposures to the antigen. The Simpson index, D which indicates a measure of diversity, predicted this vaccine candidate to have a diverse immune response. This diverse immune response might be possible as the vaccine comprises different epitopes.

Conclusions

In the present study, several machine learning-based *in-silico* tools were used to design a potential broad-spectrum multi-epitope vaccine candidate against spike protein of human coronaviruses. To the best of our knowledge, it is one of the first studies, where multiple B-cell epitopes and T-cell epitopes (CTL and HTL) were predicted from the spike protein sequences of all the seven known human coronaviruses. The final vaccine candidate was found to have a minimum conservancy of 73.97% across all the seven HCoV strains and it might protect from SARS-CoV-2 as well as other HCoVs. The designed vaccine is predicted to be antigenic and non-allergenic. Its physicochemical properties are also in the acceptable range. Molecular docking of the refined tertiary structure of the vaccine candidate with the TLR-3 protein also indicated its favourable binding with the TLR-3 receptor. *In-silico* cloning study indicated that the designed vaccine candidate might show high-level expression in the bacterial host system. The immune simulation analysis suggests that the vaccine might elicit a strong immune response and when compared to other similar publications which have reported immune simulation for the designed vaccines, our

designed vaccine candidate induced a higher level of immunoglobulins (IgM + IgG), immunocomplexes and interleukins (IL-2). The Simpson index, D which indicates a measure of diversity, predicted this vaccine candidate to have a diverse immune response. The major limitation of this study is that it's a computational study. MD simulation could have been done for 100 ns for more epitopes rather than one epitope that we selected based on the docking analysis. A population-wise study can also be done to understand the immune response elicited by the vaccine construct among different races. The predictions made by different *in-silico* tools need to be validated through various immunological assays.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13205-022-03286-0>.

Acknowledgements We are grateful to Manipal-Schrödinger Centre for Molecular Simulations, Manipal Academy of Higher Education and Manipal College of Pharmaceutical Sciences for providing the necessary support and facilities to carry out the present research work.

Funding Open access funding provided by Manipal Academy of Higher Education, Manipal.

Data availability Available from the corresponding author on request.

Code availability Not applicable.

Declarations

Conflict of interest Authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdelmageed MI, Abdelmoneim AH, Mustafa MI, et al (2020) Design of a Multi-epitope-Based Peptide Vaccine against the e Protein of Human COVID-19: An Immunoinformatics Approach. *Biomed Res Int* 2020. <https://doi.org/10.1155/2020/2683286>
- Abraham PK, Srihansa T, Krupanidhi S et al (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>

- Ahmed RKS, Maeurer MJ (2009) T-cell epitope mapping. *Methods Mol Biol* 524:427–438. https://doi.org/10.1007/978-1-59745-450-6_31
- Ali M, Pandey RK, Khatoun N et al (2017) Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci Rep* 7:1–13. <https://doi.org/10.1038/s41598-017-09199-w>
- Apweiler R, Bairoch A, Wu CH et al (2004) UniProt: the universal protein knowledgebase. *Nucleic Acid Res* 32:D115–D119. <https://doi.org/10.1093/nar/gkh131>
- Barlow DJ, Edwards MS, Thornton JM (1986) Continuous and discontinuous protein antigenic determinants. *Nature* 322:747–748. <https://doi.org/10.1038/322747a0>
- Bateman A, Martin MJ, O'Donovan C et al (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 45:D158–D169. <https://doi.org/10.1093/NAR/GKW1099>
- Berman HM, Westbrook J, Feng Z et al (2000) The protein data bank. *Nucleic Acids Res* 28(1):235–242. <https://doi.org/10.1093/nar/28.1.235>
- Binkowski TA, Naghibzadeh S, Liang J (2003) CASTp: computed atlas of surface topography of proteins. *Nucleic Acids Res* 31:3352–3355. <https://doi.org/10.1093/nar/gkg512>
- Bourdette DN, Edmonds E, Smith C et al (2005) A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. *Mult Scler* 11:552–561. <https://doi.org/10.1191/1352458505ms1225oa>
- Bowers KJ, Chow DE, Xu H, et al (2006) Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In: *ACM/IEEE SC 2006 Conference (SC'06)*. <https://doi.org/10.1145/1188455.1188544>
- Braun J, Loyal L, Frensch M et al (2020) SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature* 587:270–274. <https://doi.org/10.1038/s41586-020-2598-9>
- Bui HH, Sidney J, Li W et al (2007) Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics* 8:361. <https://doi.org/10.1186/1471-2105-8-361>
- Chen R (2012) Bacterial expression systems for recombinant protein production: *E. coli* and beyond. *Biotechnol Adv* 30:1102–1107. <https://doi.org/10.1016/j.biotechadv.2011.09.013>
- Choe J, Kelker MS, Wilson IA (2005) Structural biology: crystal structure of human toll-like receptor 3 (TLR3) ectodomain. *Science* 309:581–585. <https://doi.org/10.1126/science.1115253>
- Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 2:1511–1519. <https://doi.org/10.1002/PRO.5560020916>
- de Wilde AH, Snijder EJ, Kikkert M, van Hemert MJ (2017) Host factors in coronavirus replication. *Curr. Top. Microbiol. Immunol.* Springer Verlag, 1–42. https://doi.org/10.1007/82_2017_25
- Dhanda SK, Vir P, Raghava GPS (2013) Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct* 8:30. <https://doi.org/10.1186/1745-6150-8-30>
- Dimitrov I, Naneva L, Doytchinova I, Bangov I (2014) AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* 30:846–851. <https://doi.org/10.1093/bioinformatics/btt619>
- Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 8:4. <https://doi.org/10.1186/1471-2105-8-4>
- Enayatkhani M, Hasaniazad M, Faezi S et al (2022) 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>
- Gasteiger E, Hoogland C, Gattiker A, et al (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: *The Proteomics Protocols Handbook*. Humana Press, 571–607. <https://doi.org/10.1385/1-59259-890-0:571>
- Gori A, Longhi R, Peri C, Colombo G (2013) Peptides for immunological purposes: Design, strategies and applications. *Amino Acids* 45:257–268
- Grote A, Hiller K, Scheer M et al (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* 33:W526. <https://doi.org/10.1093/nar/gki376>
- Hebdtich M, Carballo-Amador MA, Charonis S et al (2017) Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* 33:3098–3100. <https://doi.org/10.1093/bioinformatics/btx345>
- Heo L, Park H, Seok C (2013) GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res* 41:W384–W388. <https://doi.org/10.1093/nar/gkt458>
- Ikai A (1980) Thermostability and aliphatic index of globular proteins. *J Biochem* 88:1895–1898
- Jespersen MC, Peters B, Nielsen M, Marcotili P (2017) BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 45:W24–W29. <https://doi.org/10.1093/nar/gkx346>
- Joffrin L, Goodman SM, Wilkinson DA et al (2020) Bat coronavirus phylogeography in the Western Indian Ocean. *Sci Rep* 10:1–11. <https://doi.org/10.1038/s41598-020-63799-7>
- Knutson KL, Schiffman K, Disis ML (2001) Immunization with a HER-2/neu helper peptide vaccine generates HER-2/neu CD8 T-cell immunity in cancer patients. *J Clin Invest* 107:477–484. <https://doi.org/10.1172/JCI11752>
- Kumar A, Rathi E, Kini SG (2019) Identification of E6 inhibitors employing energetically optimized structure-based pharmacophore modelling, ligand docking and molecular dynamics simulations studies. *ChemistrySelect* 4:10701–10708. <https://doi.org/10.1002/slct.201902105>
- Li J, Abel R, Zhu K et al (2011) The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling. *Proteins Struct Funct Bioinforma* 79:2794–2812. <https://doi.org/10.1002/prot.23106>
- Lin L, Ting S, Yufei H et al (2020) Epitope-based peptide vaccines identified against novel coronavirus disease caused by SARS-CoV-2. *Virus Res* 288:198082. <https://doi.org/10.1016/j.virusres.2020.198082>
- Liu J, Wu P, Gao F et al (2010) Novel immunodominant peptide presentation strategy: a featured HLA-A*2402-restricted cytotoxic T-lymphocyte epitope stabilized by intrachain hydrogen bonds from severe acute respiratory syndrome coronavirus nucleocapsid protein. *J Virol* 84:11849–11857. <https://doi.org/10.1128/jvi.01464-10>
- Lu X, Wu X, Liang Z et al (2005) Construction and functional test of HLA-A*2402-peptide tetramer. *Cell Mol Immunol* 2:145–149
- Magnan CN, Zeller M, Kayala MA et al (2010) High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics* 26:2936. <https://doi.org/10.1093/BIOINFORMATICS/BTQ551>
- Mateus J, Grifoni A, Tarke A et al (2020) Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science* 370(6512):89–94. <https://doi.org/10.1126/science.abd3871>
- Maurer-Stroh S, Krutz NL, Kern PS et al (2019) AllerCatPro-prediction of protein allergenicity potential from the protein sequence. *Bioinformatics* 35:3020–3027. <https://doi.org/10.1093/bioinformatics/btz029>
- McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16:404–405. <https://doi.org/10.1093/BIOINFORMATICS/16.4.404>
- Mousavizadeh L, Ghasemi S (2021) Genotype and phenotype of COVID-19: their roles in pathogenesis. *J Microbiol Immunol*

- Infect 54(2):159–163. <https://doi.org/10.1016/j.jmii.2020.03.022>
- Mukherjee S, Tworowski D, Detroja R et al (2020) Immunoinformatics and structural analysis for identification of immunodominant epitopes in SARS-CoV-2 as potential vaccine targets. *Vaccines* 8:1–17. <https://doi.org/10.3390/vaccines8020290>
- Naz A, Shahid F, Butt TT et al (2020) Designing multi-epitope vaccines to combat emerging coronavirus disease 2019 (COVID-19) by employing immuno-informatics approach. *Front Immunol* 11:1663. <https://doi.org/10.3389/fimmu.2020.01663>
- Nielsen M, Lund O (2009) NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* 10:296. <https://doi.org/10.1186/1471-2105-10-296>
- Oany AR, Al EA, Jyoti TP (2014) Design of an epitope-based peptide vaccine against spike protein of human coronavirus: an in silico approach. *Drug Des Devel Ther* 8:1139–1149. <https://doi.org/10.2147/DDDT.S67861>
- Panda PK, Arul MN, Patel P et al (2020) Structure-based drug designing and immunoinformatics approach for SARS-CoV-2. *Sci Adv* 6(28):1–14. <https://doi.org/10.1126/sciadv.abb8097>
- Pandey A, Nikam AN, Shreya AB et al (2020) Potential therapeutic targets for combating SARS-CoV-2: Drug repurposing, clinical trials and recent advancements. *Life Sci* 256:117883. <https://doi.org/10.1016/j.lfs.2020.117883>
- Ponomarenko J, Bui HH, Li W et al (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* 9:514. <https://doi.org/10.1186/1471-2105-9-514>
- Rapin N, Lund O, Bernaschi M, Castiglione F (2010) Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS ONE* 5:9862. <https://doi.org/10.1371/journal.pone.0009862>
- Reynisson B, Alvarez B, Paul S et al (2020) NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res* 48:W449–W454. <https://doi.org/10.1093/NAR/GKAA379>
- Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol* 5:172. <https://doi.org/10.3389/fmicb.2014.00172>
- Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5:725–738. <https://doi.org/10.1038/nprot.2010.5>
- Sanchez-Trincado JL, Gomez-Perosanz M, Reche PA (2017) Review Article Fundamentals and Methods for T- and B-Cell Epitope Prediction. <https://doi.org/10.1155/2017/2680160>
- Sarkar B, Ullah MA, Johora FT et al (2020) Immunoinformatics-guided designing of epitope-based subunit vaccines against the SARS Coronavirus-2 (SARS-CoV-2). *Immunobiology* 225(3):151955. <https://doi.org/10.1016/j.imbio.2020.151955>
- Sastry GM, Adzhigirey M, Day T et al (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* 27:221–234. <https://doi.org/10.1007/s10822-013-9644-8>
- Shey RA, Ghogomu SM, Esoh KK et al (2019) In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Sci Rep* 9:4409. <https://doi.org/10.1038/s41598-019-40833-x>
- Singh A, Thakur M, Sharma LK, Chandra K (2020) Designing a multi-epitope peptide based vaccine against SARS-CoV-2. *Sci Rep* 10:16219. <https://doi.org/10.1038/s41598-020-73371-y>
- Singh R, Bhardwaj VK, Das P, et al (2022) Benchmarking the ability of novel compounds to inhibit SARS-CoV-2 main protease using steered molecular dynamics simulations. *Comput Biol Med* 146: <https://doi.org/10.1016/J.COMPBIOMED.2022.105572>
- Singh R, Bhardwaj VK, Sharma J et al (2021) Identification of potential plant bioactive as SARS-CoV-2 Spike protein and human ACE2 fusion inhibitors. *Comput Biol Med* 136:104631. <https://doi.org/10.1016/J.COMPBIOMED.2021.104631>
- Song Z, Xu Y, Bao L et al (2019) From SARS to MERS, thrusting coronaviruses into the spotlight. *Viruses* 11(1):59. <https://doi.org/10.3390/v11010059>
- Steers NJ, Currier JR, Jobe O et al (2014) Designing the epitope flanking regions for optimal generation of CTL epitopes. *Vaccine* 32:3509–3516. <https://doi.org/10.1016/j.vaccine.2014.04.039>
- Tovchigrechko A, Vakser IA (2006) GRAMM-X public web server for protein–protein docking. *Nucleic Acids Res* 34:W310–W314. <https://doi.org/10.1093/nar/gkl206>
- Tubert-Brohman I, Sherman W, Repasky M, Beuming T (2013) Improved docking of polypeptides with glide. *J Chem Inf Model* 53:1689–1699. <https://doi.org/10.1021/ci400128m>
- Van Zundert GCP, Rodrigues JPGLM, Trellet M et al (2016) The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J Mol Biol* 428:720–725. <https://doi.org/10.1016/j.jmb.2015.09.014>
- Wallace AC, Laskowski RA, Thornton JM (1995) Ligplot: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng Des Sel* 8:127–134. <https://doi.org/10.1093/protein/8.2.127>
- Wang R, Epstein J, Charoenvit Y et al (2004) Induction in humans of CD8+ and CD4+ T cell and antibody responses by sequential immunization with malaria DNA and recombinant protein. *J Immunol* 172:5561–5569. <https://doi.org/10.4049/jimmunol.172.9.5561>
- Wang S, Peng J, Ma J, Xu J (2016) Protein secondary structure prediction using deep convolutional neural networks. *Sci Rep* 6:1–11. <https://doi.org/10.1038/srep18962>
- Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35:W407. <https://doi.org/10.1093/NAR/GKM290>
- Wrapp D, Wang N, Corbett KS et al (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367:1260–1263. <https://doi.org/10.1126/science.aax0902>
- 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* 101:2525–2534. <https://doi.org/10.1016/j.bpj.2011.10.024>
- Xue LC, Rodrigues JP, Kastiris PL et al (2016) PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* 32:3676–3678. <https://doi.org/10.1093/bioinformatics/btw514>
- Yang J, Wang W, Chen Z et al (2020) A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity. *Nature* 586:572–577. <https://doi.org/10.1038/s41586-020-2599-8>
- Yang J, Zhang Y (2015) Protein Structure and Function Prediction Using I-TASSER. *Curr Protoc Bioinformatics* 52(1):5–8. <https://doi.org/10.1002/0471250953.BI0508S52>
- 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* 11:3238. <https://doi.org/10.1038/s41598-021-81749-9>
- Zhang Y, Skolnick J (2004) Scoring function for automated assessment of protein structure template quality. *Proteins Struct Funct Genet* 57:702–710. <https://doi.org/10.1002/prot.20264>
- Zhou P, Lou YX, Wang XG et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270–273. <https://doi.org/10.1038/s41586-020-2012-7>