Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

European Journal of Pharmaceutical Sciences 151 (2020) 105375

Contents lists available at ScienceDirect

European Journal of Pharmaceutical Sciences

journal homepage: www.elsevier.com/locate/ejps

High throughput and comprehensive approach to develop multiepitope vaccine against minacious COVID-19

Rupal Ojha[a,1], Nidhi Gupta[a,1], Biswajit Naik[a], Satyendra Singh[a], Vijay Kumar Verma[b], Dhaneswar Prusty[a], Vijay Kumar Prajapati[a,*]

[a] Department of Biochemistry, School of Life Sciences, Central University of Rajasthan, NH-8, Bandarsindri, Kishangarh, 305817, Ajmer, Rajasthan, India
[b] Department of Microbiology, School of Life Sciences, Central University of Rajasthan, Bandarsindri, Kishangarh, 305817, Ajmer, Rajasthan, India

ARTICLE INFO

Keywords:
Adjuvant
Coronavirus
Multi-epitope
Vaccine
Immune response

ABSTRACT

The ongoing enigmatic COVID-19 outbreak, first reported from Wuhan, China, on last day of the year 2019, which has spread to 213 countries, territories/areas till 28th April 2020, threatens hundreds of thousands human souls. This devastating viral infection has stimulated the urgent development of viable vaccine against COVID-19 across the research institutes around the globe. The World Health Organization (WHO) has also confirmed that the recent pandemic is causing Public Health Emergency of International apprehension. Moreover, the earlier two pathogenic SARS-CoV and MERS-CoV and many others yet to be identified pose a universal menace. Here, in this piece of work, we have utilized an *in silico* structural biology and advanced immunoinformatic strategies to devise a multi-epitope subunit vaccine against ongoing COVID-19 infection. The engineered vaccine sequence is adjuvanted with ß-3 defensin and comprised of B-cell epitopes, HTL epitopes and CTL epitopes. This is very likely that the vaccine will be able to elicit the strong immune response. Further, specific binding of the engineered vaccine and immune cell receptor TLR3 was estimated by molecular interaction studies. Strong interaction in the binding groove as well as good docking scores affirmed the stringency of engineered vaccine. The interaction is stable with minimal deviation in root-mean square deviation and root-mean-square fluctuation was confirmed by the molecular dynamics simulation experiment. The immune-simulation by C-ImmSim server, which mimics the natural immune environment, yielded more potent immune response data of B-cells, Th cells, Tc cells and IgG for vaccine. The encouraging data obtained from the various *in-silico* works indicated this vaccine as an effective therapeutic against COVID-19.

1. Introduction

Coronavirus disease (COVID-19) named by the World Health Organization (WHO) is the causative agent for latest ongoing respiratory infection outbreak, which started in late 2019 in Wuhan, Hubei, China (Huang et al., 2020). Its first case was reported in late November 2019. Coronavirus belongs to Coronaviridae family and is an enveloped, non-segmented virus, which possess positive sense RNA as a genomic material. This particular outbreak is thought to be correlated with a crowded seafood market in Wuhan, China, which has been closed since 23 January 2020. Several investigation teams are working for the identification of origin of the spread. By April 28, 2020, there are 2,954,222 (~ 3 million) confirmed cases of COVID-19 globally out of them 84,341 were from China, mainland and ~4643 deaths have been attributed, reported by WHO. Outside China mainland, there are reports that 213 other countries/territories around the globe are affected with highest cases recorded in European region which has crossed 1.3 million (1,359,380) confirmed cases and 124,525 claimed lives followed by Regions of Americas which has 1,140,520 confirmed cases of COVID-19 and 58,492 claimed lives (WHO Situation Report 98). There are several host where Coronavirus infections have been reported such as avian hosts (Zhu et al., 2020), mammals like dogs, cats, camels, mice masked palm civets and bats. Reports has been shown the identification of novel mammalian coronaviruses now almost on regular basis. For instance, Coronavirus named HKU2 was originated from bat in 2018 and infected the pigs, which further showed the fatal diarrhea syndrome (https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200305-sitrep-45-covid-19.pdf?sfvrsn=ed2ba78b_2). Spread of already known human coronaviruses is not so threatening i.e. they have shown only mild

* Corresponding author.
E-mail address: vkprajapati@curaj.ac.in (V.K. Prajapati).
[1] Both authors have contributed equally for this work.

https://doi.org/10.1016/j.ejps.2020.105375
Received 15 March 2020; Received in revised form 29 April 2020; Accepted 8 May 2020
Available online 14 May 2020
0928-0987/

clinical symptoms (Cavanagh, 2007). While two evident exclusions are Severe Acute Respiratory Syndrome (SARS) coronavirus and Middle East Respiratory Syndrome (MERS) coronavirus in 21st Century means in last two decades. SARS-CoV belongs to genera Betacoronaviruses that originated from Guangdong Province in Southern part of China in year 2002 (Ismail et al., 2003), which unfortunately claimed 774 human lives and >8000 infections in human from 37 countries around the globe in 2002 to 2003 (Zhou et al., 2018). Another outbreak, MERS-CoV was started from Saudi Arab in 2012 (Zhou et al., 2018) and 2494 infection cases of MERS-CoV were reported as laboratory confirmed cases with 858 fatal cases from 2012 to date which includes claim of 38 lives (Su et al., 2016). Past two outbreaks and a recent ongoing outbreak may only be a fraction of what yet to come or can be seen and understood i.e. more severe and possibly new zoonotic infections of coronaviruses incidences yet to be discovered which shows a worldwide warning to public health (Peiris et al., 2004).

Researchers are trying to solve the enigma of the origin of the SARS-CoV-2. As per evolutionary virologist, it is rarely possible the direct jump of the virus from bat to human there should be one connecting link between them. Now, the puzzle is which animal worked as a transient host from bats to humans. Recent announcement but those studies still have not published, revealed an intermediate candidate i.e. Pangolins as transient host.

A high mortality rate of the past incidences shows the imperative need of the successful vaccine candidate, which can ultimately lead to the prevention of such outbreaks. The best way to overcome with the present situation and also preparedness for the upcoming outbreaks is the development of the vaccine against this deadly COVID-19. Particularly, in the absence of any effective drug treatment against these deadly viruses, vaccines are considered to be an obligate requirement. The continuous spread of this devastating viral infection with high mortality rate demands a protected and efficient vaccine. In modern times, multi-epitope subunit vaccine (MESV) formulation is newest way to fight against deadly pathogens. Epitopes (Cytotoxic T Lymphocyte (CTL), Helper T Lymphocytes (HTL) and B-cell Lymphocytes (BCL) i.e. short stretches of amino acid sequences of protein can elicits direct and strong immune response, then response generated by whole protein itself (Kao et al., 2009). Nevertheless, for the development of the MESV, amino acid sequence of the immunogenic proteins of interest should be known. Scientist working in Chinese Center for Disease control and Prevention & National Institute for Viral Disease and Prevention in Beijing, China and their associate collaborators emancipated whole genome sequence of COVID-19 and shared on all public databases, demonstrating fast sequencing of genome in the response of devastating outbreak which is helping in better diagnosing and treatment of further incidences. Upon initial analyses, sequencing data revealed the homology at protein sequence level with SARS-CoV and might latches onto human angiotensin-converting enzyme 2 (ACE2) receptor (Chan-Yeung and Yu, 2003).

Developing vaccine against COVID-19 is very challenging, since research to develop a vaccine against earlier coronaviruses outbreaks like SARS and MERS, generated only limited responses (Zumla et al., 2019). The past research work was more attentive on the viral Spike (S) glycoprotein and receptor binding domain (RBD) of this glycoprotein (Shi et al., 2015). But those strategies are now giving us the jumping off points for the development of vaccine against deadly novel COVID-19.

Hypothetically, almost all proteins of the miniature organism (viruses) are potential vaccine targets (Shi et al., 2015). In an incredibly fast piece of study, it has been shown that once this virus installed itself into the human body, it encodes its “tool proteins” which helps in the replication. These “tool proteins” i.e. non-structural proteins are the key (toolkit) to synthesize the other structural proteins (Spike, Membrane, Envelope and Nucleocapsid) and factors which requires to grow its progeny (Lewis et al., 2020). In our strategy of vaccine development, we have engineered the toolkit i.e. non-structural proteins which includes 3-C like proteinase (YP_009725301.1), RNA dependent RNA polymerase (YP_009725307.1), Helicase (YP_009725308.1), 3′−5′ exonuclease (YP_009725309.1), 2′-O-ribo methyl transferase (YP_009725311.1) & endoRNase (YP_009725310.1). All these proteins have also used as a drug target for commercially available antiviral drugs to fight against sudden outbreak of COVID-19, since there are no specific antivirals available (Beck et al., 2020). Further, several non-structural proteins make the replicase-transcriptase complex, which ultimately helps in the replication and transcription of virus genomic RNA (Ziebuhr, 2005). The 3-C like proteinase is a protease present in all coronaviruses, which helps in proteolysis of the non-structural protein 5 (nsp5) during replication process of virus (Tomar et al., 2015). This makes it a high value vaccine candidate against the coronavirus. Further, a key enzyme of coronavirus replication and transcription system is RNA-dependent RNA polymerase (RdRp), which is encoded by nsp12, also appears as an attractive target for therapeutics (Gao et al., 2020). This enzyme catalyses the viral RNA synthesis and therefore plays an important role in the replication/transcription process during COVID-19 infection. Helicase of coronavirus encodes by nsp13 helps in the parting of dsDNA and RNA in a 5′ – 3′ direction (Adedeji and Lazarus, 2016). Endoribonuclease is an enzyme encoded by nsp15, which cleaves specially at uridine residues and another replicative enzyme that is exoribonuclease encoded by nsp14 possess the N7-methyltransferase activity and also involved in checking fidelity of the replication whereas nsp16 encodes for 2′-O-ribo methyl transferases. Remarkably, nsp14 and nsp15, both ribonucleases are distinct to Nidovirales order which are envisaged genetic markers for the viruses belongs to this order (Bhardwaj et al., 2004; Snijder et al., 2003). However, a recent *insilico* study has anticipated that several non-structural proteins work as adhesins, which plays role in the adhesion and invasion into the host cell (Ong et al., 2020). Hence, these tool proteins, which makes the replication/transcription machinery complex, and also function as adhesins could be good candidate for vaccine development. To develop a strong and potential vaccine, we have used an immunoinformatic approach and also shown its binding with the human immune receptor TLR-3 by molecular interaction studies. The strategy we have used helped in identifying the antigenic epitopes for all proteins used in the study and vaccine construct was analysed for several other parameters like antigenicity, allergenicity, physicochemical properties which is suited as a candidate vaccine. The strategy used in the study has shown in the Fig. 1.

2. Methodology

2.1. Protein sequences selection from SARS-Cov-2 proteome for vaccine designing

In the current study, we have selected the non-structural proteins of SARS-CoV-2 which is the responsible for the pandemic, COVID-19. The sequences of the proteins were retrieved from the NCBI (National centre for Biotechnology Information) (https://www.ncbi.nlm.nih.gov/) in the FASTA format. The details of proteins along with their accession numbers and functions has been given in introduction section. After the sequence retrieval, all the sequences were individually to subjected to pBLAST server (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) to check the homology of the selected viral proteins with human proteome. Next step was to check the antigenicity of each protein and this was done with the help of ANTIGENpro protein antigenicity prediction server which is a part of Scratch protein predicter (http://scratch.proteomics.ics.uci.edu/).

2.2. Prediction of Helper T Lymphocytes (HTL) epitopes for vaccine designing

After the selection of protein sequences, we have predicted the HTL epitopes from each protein with the help of Immune epitope database and analysis resource, IEDB (https://www.iedb.org) server. This freely

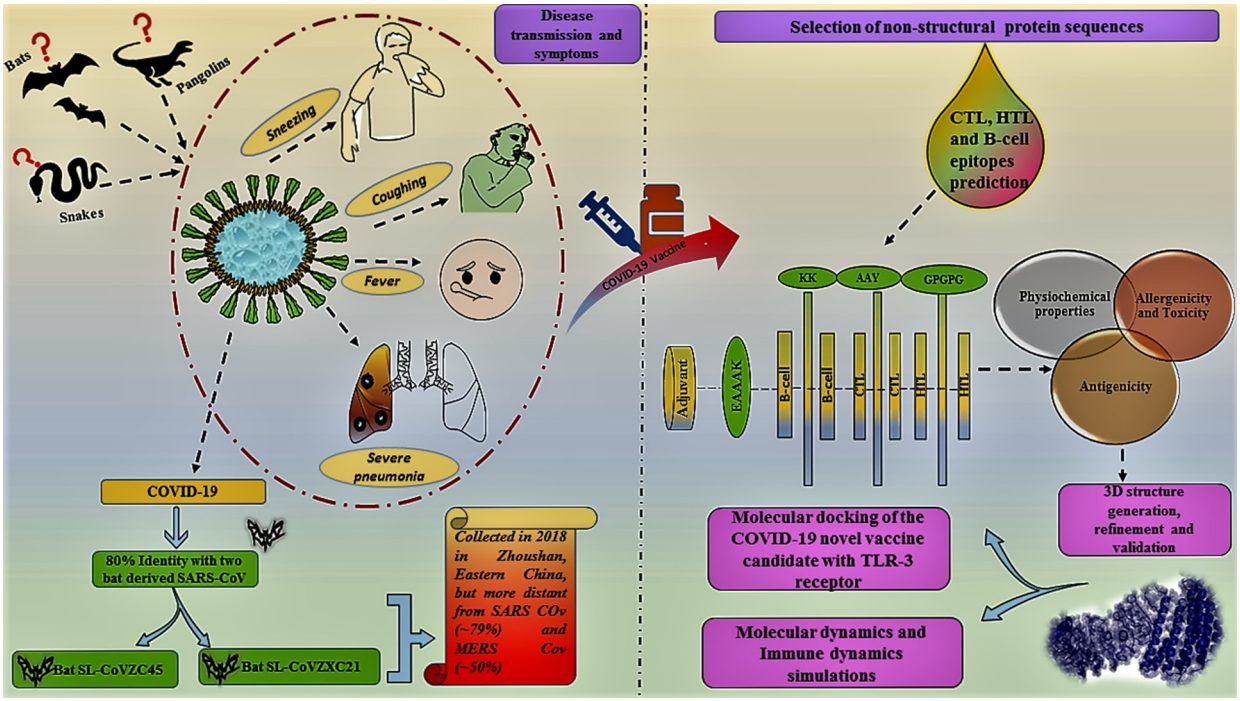


Fig. 1. Schematic flow of the transmission of recent outbreak COVID-19 and work flow used in the designing of vaccine candidate against menacious coronavirus (COVID-19).

available online server encompasses the experimentally validated data of all the species (animals and humans) on B-cell and T-cell epitopes perspective to infectious diseases (Vita et al., 2015). For the HTL epitope prediction, the selected protein sequences were given as input in FASTA format. The selection of allele was done for the human and the full HLA reference set was selected in order to obtain the epitopes prevalent to any geographical region.

2.3. Prediction of Cytotoxic T Lymphocytes (CTL) epitopes for vaccine designing

The CTL epitopes for vaccine designing were predicted by using NetCTL 1.2 server (http://www.cbs.dtu.dk/services/NetCTL/). This server comprises of two datasets, one is of HIV (human immunodeficiency virus) CTL epitopes and other one is their source proteins. The proteins present in the datasets have been cleaved into all probable 9 mers. The exception was only for the annotated epitopes and the remaining ones are considered as non-epitopes. The protein sequences in FASTA has been required as input for the prediction of potential epitopes. By using neural networks and weight matrix-based system, the server predicted the MHC I binding efficiency of epitopes, proteasomal cleavage sites (C terminal) and TAP transport efficiency. Next, the identification of epitopes was constrained to 12 supertypes, which are known to enumerate the variance in binding specificities with MHC I class molecules (Larsen et al., 2007).

2.4. Prediction of B-Lymphocytes epitopes for vaccine designing

B-cell epitopes in this study have been predicted via ABCpred server (http://crdd.osdd.net/raghava/abcpred/). This highly cited and reliable server utilizes machine learning algorithm and artificial neural network (ANN) for the epitope prediction from antigenic sequence. The datasets present in this server contains around 700 B-cell as well as 700 non-B-cell epitopes, having maximum length of 20 amino acids, but it also has the option of selecting the length for predicted epitope (10, 12, 14, 16 and 20 amino acids length) (Saha and Raghava, 2007). As per recurrent neural network the server has the accuracy of 65.93%. Here, we have given the protein sequence as input and set the other parameters threshold (0.51) and overlapping filter (ON) as default except for amino acid window length which was set as 10.

2.5. Merging of epitopes for the generation of potential vaccine candidate

All the selected epitopes were combined together with the help of linkers or spacer sequence of amino acids. Different linkers have been used to link the B-cell, HTL and CTL epitopes for instance KK linker was used for making B-cell epitope combinations while AAY and GPGPG were used for CTL and HTL epitope joining. Next, the adjuvant at the starting of sequence (N-terminal) was added and then to link the adjuvant epitopes one more linker was used i.e. EAAAK linker. After joining all the adjuvant, linkers and epitopes in a sequential manner, a single potential vaccine candidate was generated.

2.6. Physiochemical properties, antigenicity, allergenicity and toxicity prediction

The physiochemical properties of the designed COVID-19 vaccine construct was predicted with the help of ProtParam- an ExPASy server (https://web.expasy.org/protparam/). This server computes the various parameters of protein sequence which includes theoretical pI, molecular weight, estimated half-life, extension coefficient, instability index, aliphatic and GRAVY index.

For the prediction of antigenicity of the designed construct, ANTIGENpro server (http://scratch.proteomics.ics.uci.edu), share of Scratch protein predictor, was utilized. This is the first server which predicted antigenicity of the entire protein by using experimental proven microarray analysis reactivity data meant for five pathogens. This server predicts the antigenicity of protein on the basis of alignment free and manner. It is a sequence based, pathogen independent server whose predictions are made by five machine learning algorithms and further the antigenic or non-antigenic nature of the protein get finalized by SVM (Support Vector Machine) classifier.

Further, the allergenicity of the vaccine candidate was predicted by an online server namely Algpred (http://crdd.osdd.net/raghava/

algpred/) and AllerTop (https://www.ddg-pharmfac.net/AllerTOP/) servers. The AlagPred server predicts the allergenicity on the basis of matching of vaccine sequence against the known epitope sequences (Saha and Raghava, 2006) whereas the AllerTop server utilizes the machine learning techniques which are inclusive of DT (Decision tree), kNN (k nearest neighbours), LR (Logistic regression), MLP (Multilayer perceptron) NB (naïve Bayes), RF (Random forest) but among them the five-fold cross validation was only showed by the kNN method with the maximum precision of 85.3% (Dimitrov et al., 2013). Along with this, ToxinPred tool (http://crdd.osdd.net/raghava/toxinpred/) was used to predict the toxicity of the designed vaccine candidate. This tool predicts the toxicity of peptides as well as entire protein sequences and accepts the input in fast format and then work on score of position specific basis of Quantitative Matrix, which further identifies the toxicity of the submitted sequence (Gupta et al., 2015).

2.7. Tertiary (3D) structure prediction

Tertiary structure was modelled with the help of RaptorX server (http://raptorx.uchicago.edu/StructPredV2/predict/). The server compares the input sequence with other nonredundant homologs in order to evaluate the quality of model and then on this basis generates the 3D structure accordingly. RaptorX server graded second and it works by utilizing CNF (conditional neural fields), CRF (conditional random field) and MTT (multiple template threading) which assimilate various biological signals and convert them into a plausible nonlinear scoring function (Källberg et al., 2014).

2.8. Tertiary structure (3D) refinement, model quality assessment and validation

The refinement of the obtained tertiary structure was done with the help of 3D protein structure refinement server (http://sysbio.rnet.missouri.edu/3Drefine/). This CSAP (Critical Assessment of protein Structure Prediction) evaluated server is freely available and demands either the initial structure or protein sequence as input in order to perform the statistical and graphical analysis. The server utilizes complete physics and intelligence-based force fields to perform the frequentative optimization of hydrogen bonds and energy minimization (Bhattacharya et al., 2016).

After refinement of the 3D structure, the quality of model was again verified with the help of QMEAN (Qualitative Model Energy Analysis) server (https://swissmodel.expasy.org/qmean/). The QMEAN server evaluates the global and local quality of the structure. As input the server requires a 3D model in PDB format or sequence of the protein. Before proceeding to the submit option, user can choose scoring function as per their requirement, there are three QMEAN scoring functions- 1. QMEAN 2. QMEANDisco 3. QMEANBrane. The complex QMEAN scoring function includes two subtypes QMEAN4 and QMEAN6 which estimates the quality of complete structure and also for each amino acid by using single PDB model. The second one is QMEANDISCO, this also perform the same function as QMEAN does, except calculating the pairwise distance constraints (DisCo) between each amino acid residues. The last one is QMEANBrane, this only evaluates the local quality of models, especially for membrane proteins and also show the membrane, soluble, interface and transmembrane part of the input PDB model (Benkert et al., 2009).

At last, after the structure modeling, refinement and quantitative energy analysis its validation is a very decisive part. For the validation purpose, we have utilized RAMPAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php). As an input the server requires the PDB structure on protein and then utilizes the quantum mechanics to predict the favoured, allowed and disallowed regions. The prediction was done on the basis of phi vs psi backbone angles, steric clashes, hydrogen bond energy, planarity of peptide bonds (Lovell et al., 2003).

2.9. Molecular docking of vaccine candidate and TLR-3 receptor

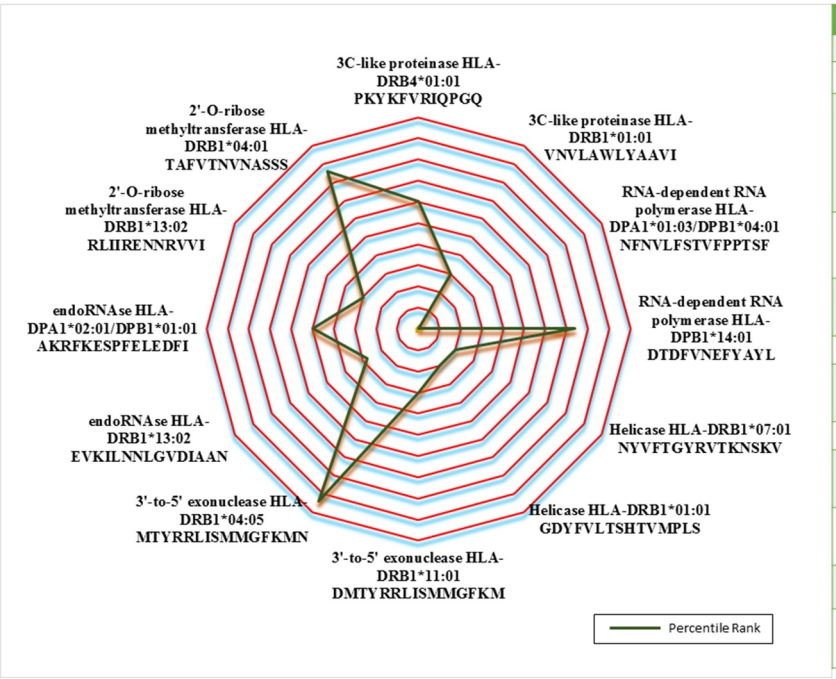
Protein-protein interaction between the TLR-3 receptor and vaccine candidate was done with the help of High Ambiguity Driven protein-protein DOCKing (HADDOCK) server (http://milou.science.uu.nl/services/HADDOCK2.2/haddockserver-easy.html). This server utilized both biochemical bioinformatics and biophysical methods information to improve sampling and scoring of docking results. As an input the server requires the active and passive residues from both the receptor and ligand as well as their energy minimized PDB models. These interactive residues firstly, predicted with the help of CPROT server and then used in HADDOCK. The server provides the output on the basis of van der Waals energy, Z score, desolvation energy, RMSD score, etc., in which the Z-score represents the estimation of standard deviation (De Vries et al., 2010).

2.10. Molecular dynamics simulation of designed vaccine candidate and TLR-3 receptor

For the dynamical simulation of designed vaccine candidate and TLR-3 receptor, in this study, we have utilized GROMACS (GROningen MAchine for Chemical Simulations) v5.0 program with the force field GROMOS96 43A1. This works on the command lines and after the run of program, trajectory files were generated which is a time taken process but defines the motion of all the atoms over the time (Abraham et al., 2015; Ojha et al., 2020; Van Der Spoel et al., 2005). Before starting the first step in GROMACS all the HOH atoms were get removed via PYMOL software and then input the PDB file to start the first module of GROMACS as pdb2gmx. After this the topology file of the input protein was generated which have all the specifics about bonded and non-bonded interaction or features. After this the next step was of solvation, with the help of commands the protein get centered in a box of 1.0 nm, which was then filled with the solvent (can be water or any other solvent). Further, the ions (Na or Cl ions) were added in the solvation box, having the charged protein, in order to exchange the HOH molecules with it. Then, prior to MD run, energy minimization of the protein was done to achieve the lowest energy of the protein. Subsequently, the equilibration of the solvent was done in association with pressure (NPT) and temperature (NVT). After the stabilization of pressure and temperature ensembles, the final MD simulation ran for 10 ns to positively produce the trajectory files, which defines the microscopic movement of the system.

2.11. Immune dynamics simulation of the designed COVID-19 vaccine candidate

In this study, we have performed an *in silico* experiment which generates the immune-response for vaccine candidate. There is an online dynamic immune simulation server developed by Rapin et al., namely C-ImmSim server (https://www.iac.cnr.it/~filippo/projects/c-immsim-online.html) (Rapin et al., 2010) which uses the sequence of vaccine candidate as an input. The advanced C-ImmSim server, which is consistent with earlier version uses the amino acid strings in place of Bit strings. As per the literature, dose interval of 1 month is prescribed for immunization of vaccine, though there are some vaccines where the difference can be 2, 3 or 6 months between first immunization and first booster dose. Thus, we had chosen the one-month interval and identified the dose dependent immunological response of our vaccine candidate. Besides, the likeliness of the immune response generated by the vaccine (in response to pathogen invasion) was analyzed through repeated exposure of X following injection doses. In each instance, one dose was made up of X vaccine molecules where the simulation was conducted for almost X days. Finally, the immune response generated was calculated in the form of Simpson Index (D).



Alleles	Region
HLA-DRB4*01:01	North America
HLA-DRB1*01:01	South East Asia, Europe
HLA-DPA1*01:03/DPB1*04:01	Europe, North America, North East Asia, Oceania, South & Central America, South Asia, South East Asia, Sub-Saharan Africa
HLA- DPB1*14:01	Europe, North Africa, North East Asia, Oceania, South East Asia
HLA-DRB1*07:01	South East Asia, Western Asia, North America
HLA-DRB1*01:01	South East Asia, Europe
HLA-DRB1*11:01	South East Asia, Western Asia, Oceania
HLA-DRB1*04:05	Oceania, South East Asia
HLA-DRB1*13:02	South East Asia, Western Asia, North America
HLA-DPA1*02:01/DPB1*01:01	South & Central America, North East Asia
HLA-DRB1*13:02	South East Asia, Western Asia
HLA-DRB1*04:01	North East Asia, South East Asia, Europe

Fig. 2. Graphical representation of predicted HTL epitopes along with their percentile rank and HLA distribution.

3. Result and discussion

3.1. Screening of SARS-CoV-2 replication machinery antigenic protein for the sequence retrieval

In general, proteome of the SARS-CoV-2 has both structural and non-structural proteins and after literature survey we came across this point that the non-structural proteins of virus are mainly responsible for their replication in host (Ojha et al., 2019a). Here, in this study we have targeted "toolkit" of the virus i.e. the non-structural proteins for vaccine designing. Recently, researchers have reported the study about multi-epitope vaccine against the novel coronavirus and in that, they have targeted the structural protein of coronavirus whereas in our current study we have selected the non-structural proteins as they have also been reported as key or potential antiviral drug targets. So, targeting the already reported drug targets for vaccine designing purpose will be more effective. Further, as per recent survey report, it was found that SARS-CoV-2 and SARS-CoV both belongs to the same beta coronavirus group with the 70% similarity between their genomes (Cheng and Shan, 2020). So, targeting the replication machinery proteins of this deadly coronavirus for vaccine designing could be a better way to block the expansion of disease either its SARS or COVID-19.

For the development of immunogenic vaccine, we have proceeded by retrieving all the six key proteins from NCBI database. Apart from this, by using pBLAST server we have checked the homology of all the viral protein with human proteome. With this process we came to conclusion that none of the viral protein was showing any kind of sequence identity/similarity with the human proteins expect, helicase protein. The helicase protein was showing 22.4% of identity and 34% query coverage with the ZGRF1 isoforms proteins of homo sapiens, but as per Pearson (2013), for the purpose of vaccine designing the sequence similarity between the human and pathogen should not be more than 40% in order to overcome the probabilities of cross reactivity. Here, with this data we can conclude that the non-structural proteins used for vaccine designing will show no cross-reactivity reactions when administered in host. Also, we have checked the antigenicity of all the selected non-structural proteins with the help of ANTIGENpro server. The result indicates that the all the proteins were highly antigenic in nature and their sequences can be utilized for the purpose of vaccine designing (Supplementary Table 1).

3.2. Antigenic B-cell epitopes prediction from the COVID-19 proteome

B-cell epitope prediction was done via ABCPred server and the obtained epitopes were then finalized on the basis of their ranking and highest score. The epitopes with highest score have the highest binding affinity. A total six B-cell epitopes were selected (one from each protein sequence) with the length of 10 mer amino acid each (Supplementary Table 1).

3.3. Antigenic MHC class II epitopes prediction from the COVID-19 proteome

This study has utilized the most reliable and validated immunoinformatic servers for the prediction of antigenic epitopes (B-cell, HTL and CTL) which are probably conserved across the diverse coronavirus species and expected to recognize the human T-cell and B-cell receptors in order to generate a potential immune response against the COVID-19. The immunoinformatic servers provides an easiest way, which curtails the experimental effort for the epitope identification. In the field of vaccinology, many researchers with the help of reverse vaccinology approach have proved the antigenicity of peptide vaccines in contradiction of various infectious diseases (Dikhit et al., 2017; Muruato et al., 2017; Ojha et al., 2018; Pandey et al., 2019).

For the identification of antigenic HTL epitopes from all the protein sequences, IEDB server was utilized. The epitopes were sorted on the basis on percentile rank which was less than or equal to 0.9 and IC_{50} value, less than or equal to 50. The lowest percentile rank and IC50 value denotes good binding affinity of epitopes with the HLA (Human leucocyte antigen). The next criteria, selection of allele, was done on the basis of distribution of population in the COVID-19 endemic regions. In the current study, we have tried to cover almost all regions in which COVID-19 confirmed cases were reported and total 12 epitopes were selected i.e. two epitopes from each protein sequence Alleles according to their frequency and geographical region has been mentioned in Fig. 2. The IC_{50} value and the percentile rank score of each epitope has been given separately in supplementary file as Table 2.

Table 1
Secondary structure characteristics of the COVID-19 vaccine candidate.

Characteristic of Vaccine	Assessment
Number of amino acids	561
Molecular weight (kDa)	61.05
Theoretical pI	9.54
Total no. of negatively charged residues (Asp + Glu)	31
Total no. of positively charged residues (Arg + Lys)	61
Formula	$C_{2808}H_{4313}N_{713}O_{756}S_{28}$
Extinction Coefficients (M^{-1} cm^{-1})	85,205
Estimated half-life	30 h (mammalian reticulocytes)
Instability Index	26.97
Aliphatic Index	81.44
Grand average of hydropathicity (GRAVY)	0.060

3.4. Antigenic MHC class I epitopes prediction from the COVID-19 proteome

The CTL epitopes were predicted via NetCTL server and at the time of prediction all the parameters were set to default for instance, C-terminal cleavage (0.15), threshold value (0.75), TAP transport efficiency (0.05), except for supertype selection. The HLA supertype identification is important for the effective and specified recognition of T-cell epitopes from many viral diseases for instance HIV, MMR, MERS, SARS etc. The supertype selection in this study was done according to the percentage of worldwide population coverage. With the help of literature survey, we came across that the supertypes A2, A3 and B7 together covers around 88% population globally (Sette and Sidney, 1999; Shen et al., 2018). A total of 18 epitopes were selected; three from each protein sequence (Supplementary Table 3).

3.5. Designing of COVID-19 multi-epitope antigenic vaccine candidate

After the prediction of all antigenic vaccine candidates, they were joined with each other with the help of linkers in order to generate a single construct. However, in multi-epitope vaccine each epitope should have the capability to induce the potential immune response and for that linkers are the main factor which are involved in locating the epitopes as well as responsible for the immunogenicity of the entire construct. They are also use to discrete the domains, in a single protein sequence without distressing the function of protein and helps to maintain a stable and flexible protein complex. Here, in this study we have used different types of linkers to connect the epitopes for instance, for linking B-cell epitopes “KK” (Yano et al., 2005) (augmentation of proteasome processing) linkers were used while for linking CTL and HTL epitopes “AAY” (Ojha et al., 2019b) (proteasomal cleavage site and avoid the genesis of new epitope) and “GPGPG (Livingston et al., 2002) (for improving the flexibility of protein)” linkers were availed, respectively. Along with these linkers one more linker was used i.e. EAAAK, which has been reported as a helix forming linker in the designing of next-generation vaccine. Next, to acquire more immunogenic construct, at the N-terminal of the designed vaccine an adjuvant namely, human beta defensin 3 (45 amino acids) was attached. This TLR-3 agonist human beta defensin 3 has been well known to augment the strong T-cell and B-cell immunity (Gupta et al., 2020; Mohan et al., 2013).

After linking of all the epitopes, a single immunogenic vaccine candidate of 561 amino acids was created. Here, we have designed the vaccine candidate in this sequential manner- Adjuvant-B-cell-HTL-CTL. In the vaccine sequence, the former 45 amino acids were occupied by an adjuvant, next 6 amino acids were of EAAAK linker and then from number 51 to 120 amino acids were covered by B-cell epitopes whereas 121–345 and 346–561 were obscured by HTL and CTL epitopes, respectively (Supplementary Fig. 1).

3.6. Assessment of physiochemical properties, antigenicity, allergenicity and toxicity of designed COVID-19 vaccine candidate

The physiochemical properties were computed on the basis of certain parameters such as instability index which was 26.97, (should be less than 40) hence, showing the stable nature of the vaccine candidate. Next, on the basis of N-terminal residue (N-end rule), half-life of the protein was estimated (on models like yeast and *E. coli in vivo*, mammalian reticulocytes in vitro), which denotes that after the synthesis of protein in cell, at what time partial amount of the protein will get dissipate? As per result, we have found that the estimated half-life for the designed vaccine candidate was >30 h and hence, inversely demonstrating that designed vaccine candidate was stale in nature. The calculated GRAVY (Grand average of hydropathicity) score was 0.060 and this positive score specifies the highly hydrophobic nature of protein. Next, the aliphatic index of the protein epitomizes the space reserved by aliphatic side chains (AVIL) and the calculated score was 81.44, the higher aliphatic index denotes highly stable side chains of the protein. Other calculated parameters have been shown in Table 1.

Apart from this, the average antigenic propensity of the designed vaccine candidate was 1.0406, which represents the highly antigenic nature of the vaccine. According to the AntigenPro server, high score signifies the more antigenic nature of the protein. For the antigenicity prediction AlgPred server was used and the result obtained from this server was based on the 6 approaches (IgE mapping, MEME/Mast motif, SVM modules based on both amino acid and dipeptide composition, BLAST search on ARPs and combined approach or hybrid approach which involves all the parameters). Among these 6 approaches, SVM modules based on both amino acid and dipeptide composition and Hybrid approach identified the vaccine construct as allergenic whereas MEME/Mast motif, BLAST search on ARPs, IgE mapping approaches identified the construct as non-allergenic. So, further to cross check this mystery we again checked the allergenicity with AllerTop server, the result from this server exposed the designed vaccine construct as probable non-allergen. Hence, by taking both the results from different servers into consideration, we concluded that designed vaccine candidate was probably non-allergen and will not lead to any allergy during pre and post vaccination. Subsequently, with the help of ToxinPred server we have checked the toxicity of the epitopes, as this server accepts the protein sequences of >500 amino acids and here, in this study the designed vaccine candidate was of 561 amino acid. So, after doing this it was found that all the epitopes involved in the designing of vaccine candidate was nontoxic and will not release any toxic components after administration (Supplementary Tables 5A, B and 5C). This above data suggested that the designed vaccine candidate was highly stable, antigenic, probable non-allergic and nontoxic in nature.

3.7. Structure prediction and model quality assessment of the designed COVID-19 vaccine candidate

Tertiary structure of the designed vaccine candidate was achieved via RaptorX server. The tertiary structure of the vaccine candidate was predicted as 2 domains- domain 1 and domain 2 and the best template with id-5cwmA was generated and the calculated p-value for the model was 4.31*10ˆ−04. The p value represents the quality of the predicted model or structure and value less than 10ˆ−3 is a signal of good quality of model. It was also found that all the residues (100%) of the sequence were modelled and the percentage of disordered region was zero. Next calculated parameters are of percentage composition of coil, helix and beta strands which was 47%, 16% and 35%, respectively (Fig. 3).

3.8. Tertiary structure refinement and validation

The tertiary structure refinement was achieved with the help of 3D refine server. As an output the server generated five refined models and the selection of best model was done on the basis of various parameters.

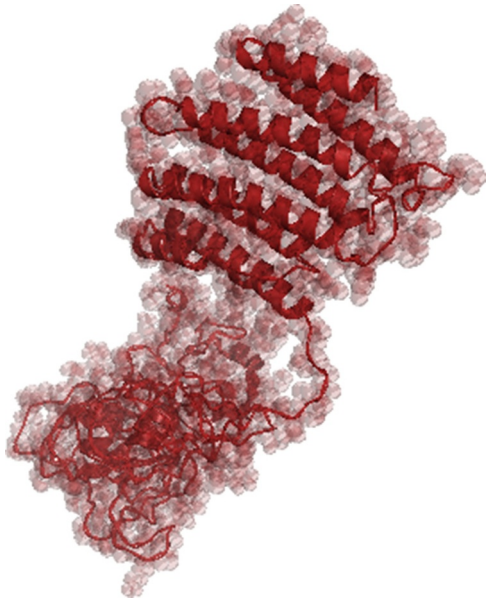


Fig. 3. Tertiary structure prediction of designed vaccine candidate with the help of RaptorX server.

The model number fifth was selected as its 3D refine score was 40,460.5 which was less than all the other four models. Additionally, the GDT-TS score of the selected model was 0.9960 whereas the GDT-HA score was 0.9501, these higher scores denotes the conservative refinement. Next, the RMSD (root mean square deviation) was 0.450, this higher score indicates the aggressive refinement in comparison to initial non-refined one. The MolProbity score was 3.790 and the RWPlus (potential energy) score was −108,736.3 both lower scores indicate the best quality of model.

After refinement, the quantitative model energy assessment was done with help of QMEAN server. For the estimation of local and global quality of model we have assessed the same by using all the three scoring function options available in QMEAN server, as described in methodology. The refined 3D model was given as input and the first estimation was done for QMEAN4 scoring function which evaluates the global and local model quality as well as global IDDT score which is the local distance difference test between per amino acid residue. The score obtained by QMEAN4 server was −7.25 whereas the global IDDT score of the model was 0.9182 (this score should be ranges between 0 and 1). Next, with the help of QMEANDisCo scoring function we have predicted the global quality of model, the obtained global score was 0.35 ± 0.05. This data suggested that the score (global, local, IDDT) obtained for our model structure was higher and in between the range, hence, we can conclude that designed model has best scoring and was of exceptional quality (Fig. 4).

Subsequently, the validation of the refined model was done via RAMPAGE server which predicted the φ (phi) and ψ (psi) torsion angles of the residues. The Ramachandran Plot assessment provides us the information about the backbone, sidechain and geometry of C alpha conformations. Firstly, we have done the analysis of the non-refined vaccine 3D structure and we have found that the favoured region covered 431 residues (77.1%) while the number of residues in the allowed region was 82 (14.7%) and for the disallowed region was 46 (8.2%) (Fig. 5A)Secondly, we have generated the Ramachandran Plot for the refined 3D vaccine structure (Fig. 3) and the graphical representation has showed that the around 461 residues were in the favoured region (82.5%), whereas 64 (11.4%) and 34 (6.1%) amino acid residues were in allowed and disallowed area of the protein sequence, respectively (Fig. 5). This data suggested that the refined vaccine candidate has more stability and flexibility in comparison to non-refined one. However, the best quality model in RAMPAGE should have ~98% residues (expected) in the most favoured region so, here we can conclude that the score obtained for the refined 3D model was good then the non-refined one but not the best.

3.9. Molecular docking or protein-protein interaction of the COVID-19 and TLR-3 receptor

In this study, we have selected TLR-3 receptor for carrying out docking with the designed vaccine candidate (COVID-19). The role of TLR-3 has already been known in the previous studies of SARS coronavirus infection. During the SARS outbreak in 2015 Tortura et al. has reported that mice lacking TLR-3 and TLR-4 are more vulnerable to SARS-CoV in comparison to mice having TLR-3 and TLR-4 receptors; exceptionally weight loss was observed when exposed to infection, but no mortality was detected (Totura et al., 2015). One more rational for the selection of TLR3 was that it mainly recognizes the viruses which have RNA as genetic material. Subsequently, after the selection of TLR3 receptor we have proceeded for the docking and the process was done with the help of HADDOCK server in order to illustrate the functioning of ligand molecules which binds to the active site of the receptor as well as to explicate the essential biochemical activities. Before initializing docking, the active and passive residues (interactive residues) for both the vaccine candidate as well TLR3 receptor has been predicted with the help of CPROT server, a part of HADDOCK. After performing molecular docking, we have obtained 8 clusters of the docked complexes and each cluster was composed of 4 best docked complexes. According to HAADOCK server calculation the attained top 10 clusters were highly reliable. Further, the docked complex was finalized on the basis of Z score and RMSD value (Table 2). Here, we have selected the first best complex of cluster number 1 because the Z score (denotes standard deviations) of the selected complex was −1.0 which was lowest among all obtained clusters, the lowest score always designates the highest binding capacity (Fig. 6). Here, are the details of score which we have obtained after performing docking.

3.10. Molecular dynamics simulation of the docked complex

For the accomplishment of highly stable vaccine candidate we have performed the in silico simulation of the docked complex (vaccine candidate and TLR-3 receptor) with the help of GROMACS. Various steps like energy minimization, pressure assessment, temperature, and potential energy calculations were analyzed after the MD simulation study. The temperature calculation plot reveal that the system remains persistent at 300 K (Fig. 7A) at around 100 ps time interval, whereas in case of pressure the system demonstrates pressure fluctuation with an average value of 1 bar at the same time interval (Fig. 7B). The complex root mean square deviation (RMSD) plot demonstrated the structural stability and flexibility of the docked complex. RMSD vs. time plot reveals that the primary fluctuation started with 0.15 nm and completed at 0.55 nm after 6 ns of a time span and the minor fluctuations were also observed in the docked complex, after this the complex reached to its stable position at 6 ns (Fig. 7C). Next, the root medium square fluctuation (RMSF) was obtained for amino acid side chains in order to analyze the fluctuations of amino acid in the docked complex. The RMSF plot reveals minor shifts in amino acid side chains, indicating the continuing interaction of the vaccine candidate and the TLR-3 receptor. In the RMSF plot the very fluctuated regions indicates the extremely flexible complex and here, the higher peaks (100–200 residues and 600–700 residue) in the RMSF plot indicate approximately 0.4 nm suggests that the complex comprises extremely stable extents (Fig. 7D).

3.11. Generation of immune response by COVID-19 vaccine candidate

The immune simulation experiment was performed using C-ImmSim server which, gives a real immune response result. In our results, we have observed the proliferation in secondary and tertian immune

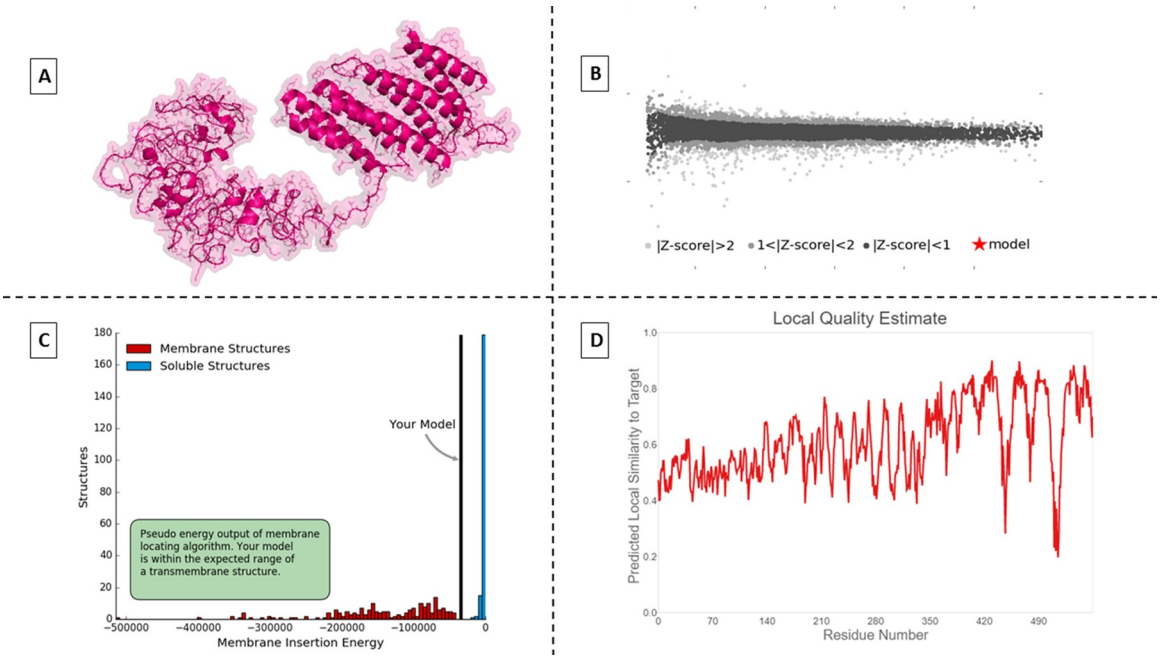


Fig. 4. (A) Refined tertiary structure, refined with the help of 3D refine server. (B and D) Showing the Z score and model quality (C) Showing the membrane insertion energy of the refined protein structure.

response which is identified by IgG1 + IgG2 and IgM. Also, while decreasing the antigen count IgG + IgM also showed the proliferated response in Fig. 8A. Result revealed the immune response development upon immunization. B-cell population level was also stimulated and seems to be very high upon immunization (Fig. 8B). Likewise, in secondary and tertian immune response, cell mediated immunity i.e. the cytotoxic and helper T cells (Tc & Th) levels were high (Fig. 8C & D). Level of IFN- γ was retained during the exposure time which has been shown in Fig. 8E. These results are significant for the immune response against SARS-CoV-2. Hence, the designed vaccine candidate can compete to control the devastating coronavirus outbreak.

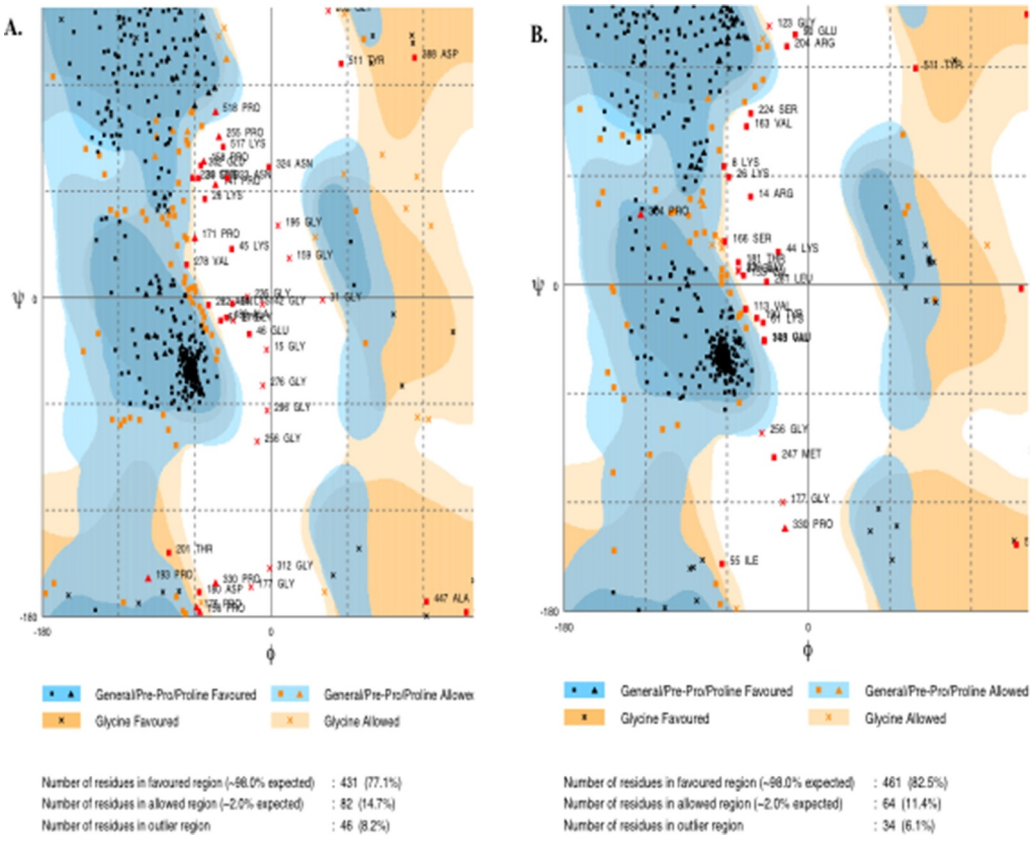


Fig. 5. Ramachandran plot shoeing the differences between the (A) non-refined and (B) refined protein tertiary structures.

Table 2
Table showing the scores obtained from the molecular interaction of vaccine with immune receptor TLR-3 using HADDOCK server.

HADDOCK score	149.3 +/- 10.7
Cluster size	8
RMSD from the overall lowest-energy structure	6.7 +/- 1.1
Van der Waals energy	−86.0 +/- 16.5
Electrostatic energy	−280.8 +/- 51.5
Desolvation energy	−14.4 +/- 13.7
Restraints violation energy	3057.9 +/- 205.90
Buried Surface Area	2551.2 +/- 320.4
Z-Score	−1.0

4. Conclusion

Too much to be known about the COVID-19, which is spreading through China and claiming thousands of lives. Featuring the effect of this debilitating outbreak is not feasible because of its rapid spread. The recent three outbreaks of coronaviruses in past two decades are different from their “common cold” causing friends and can spark a fire in individual's body. The newly dubbed enemy, in the form of virus was initially named as COVID-19 by World Health Organization (WHO) and now formally named as SARS-CoV-2. The virus primarily attacks on respiratory system, which involves three major steps: replication of the virus, hyper-reactivity of the immune system and pulmonary obstruction. As per the WHO, the virus makes the perforated holes in the lungs, which upon infection give them honey-comb appearance. To inhibit the viral infections in human the vaccination is one of the most desirable way. Herein, we have devised a multi-epitope subunit vaccine to elicit the immune response against COVID-19 which comprises of an adjuvant i.e. beta-defensin 3 at N terminal end followed by B-cell epitopes, HTL epitopes and CTL epitopes connected with appropriate linker sequences. The physicochemical parameters of vaccine candidate which includes allergenicity, solubility, pI, hydrophobicity, molecular weight, GRAVY, half-life indicated its effectiveness. The binding

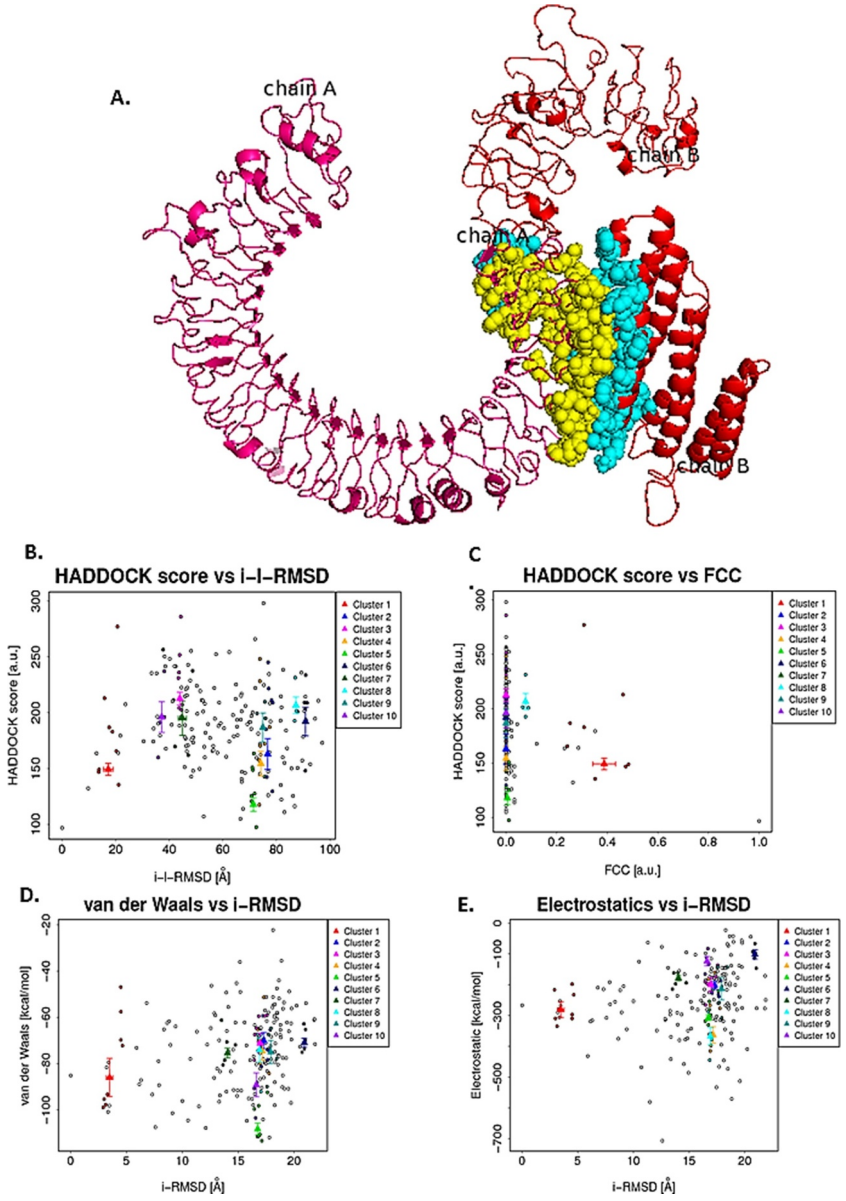


Fig. 6. Molecular interaction studies of vaccine candidate with immune receptor TLR3 using High Ambiguity Driven (HADDOCK) server: (A) docked complex of vaccine and TLR3, (B) – (E) graphical representation of different parameters of docking of 10 clusters, (B) HADDOCK score versus interface-RMSD, (C) HADDOCK score versus fraction of frequent contacts (FCC), (D) Vanderwaals energy versus interface-RMSD, (E) Electrostatic energy of interacted complex versus interface-RMSD.

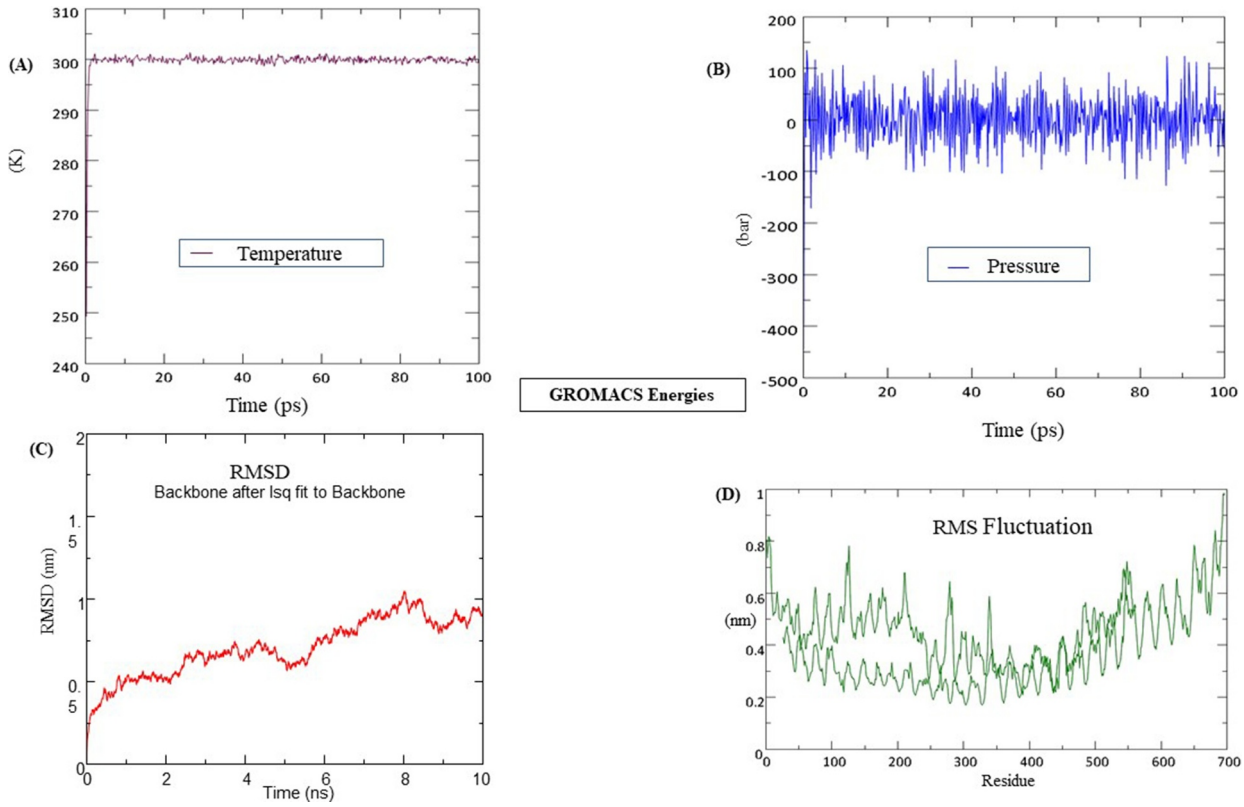


Fig. 7. The figure showing the results of molecular dynamics simulation of vaccine and immune receptor TLR3 using GROMACS. A and B showing the equilibration phase ensembles-temperature (constant at 300k for 100 ps) and pressure (displaying fluctuations at 1 bar value for 100 ps), respectively. (C) RMSD (root mean square deviation) plots reflects the stability between the vaccine candidate and TLR-3 receptor whereas (D) RMSF (root mean square fluctuation) reflecting the flexibility and fluctuation of the amino-acids residues in the side chain of docked complex.

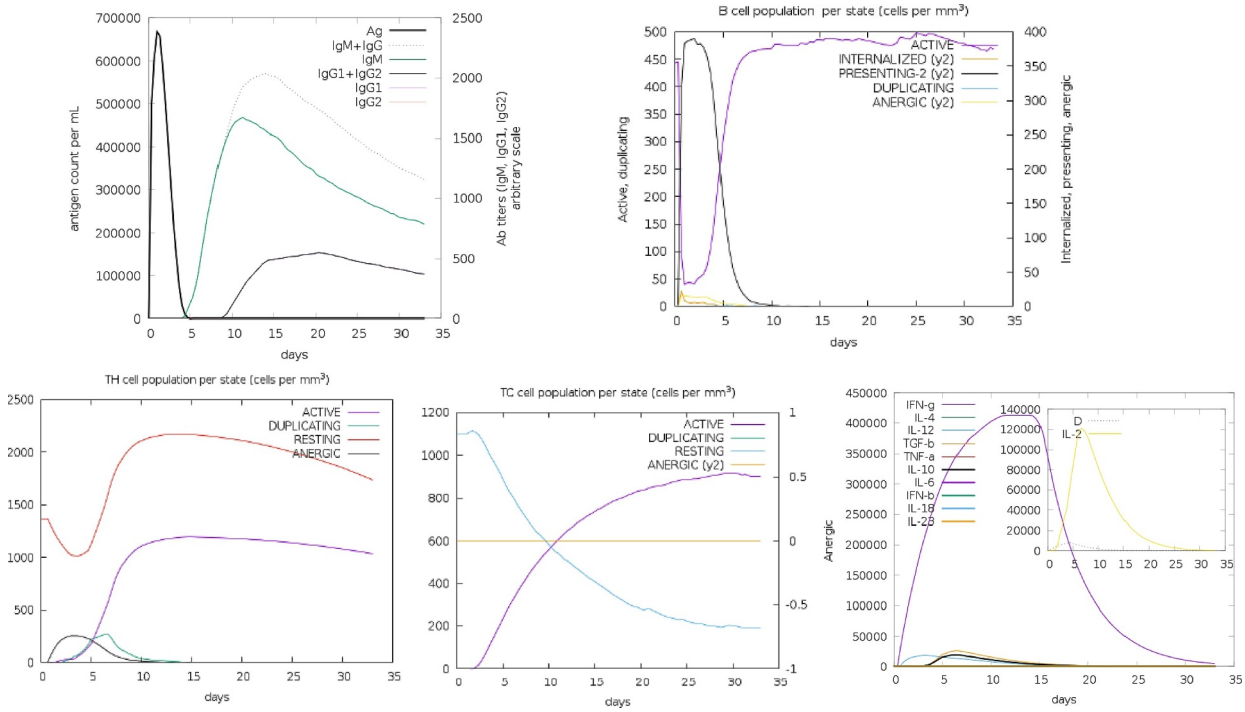


Fig. 8. Immune Simulation results by C-ImmSim: an in silico experiment of the engineered vaccine candidate. (A) represents the antigen injection and proliferation of immune response by immunoglobulin production. Various sub-classes of immunoglobulin are represented as colored peaks. (B) The active B-cell population is observed on the administration of vaccine. (C) The illustration of the generation of Helper-T cells, and (D) The illustration of the generation of cytototoxic-T cells after the vaccine administration. In the graphs, RESTING indicates to the cells, which were not shown to the antigens while ANERGIC indicates the tolerance level of antigen. (E) Graph showing the induced cytokine level upon administration of vaccine. The inset graph indicating the Simpson Index, D of IL-2. Simpson Index, D was inferred as the measurement of diversity.

specificity of the vaccine candidate with human immune cell receptor i.e. TLR3 was assessed by molecular docking interaction. Lowest energy score obtained showed the strong interaction and specificity between vaccine and binding groove of TLR3 receptor. Our devised vaccine candidate showed the stable and prolonged binding in molecular dynamics simulation with negligible deviation in RMSF and RMSD values. Additionally, immune simulation that mimics the natural environment in computational immunology experiment exhibited the strong immunological response data from the major players of the immunity i.e. cell mediated (T-cell) and humoral (B-cell) at different doses at different intervals for COVID-19 multi-epitope subunit vaccine. This work provides an innovative vaccine construct against the recent COVID-19 outbreak and encourages the development of the vaccine against other newly emerging infectious disease.

Author statement

Rupal Ojha: Conceptualization, Data curation, Software analysis and Validation, Writing original draft.

Nidhi Gupta: Conceptualization, Data curation, Software analysis and Validation, Writing original draft.

Biswajit Naik: Data curation and analysis.

Satyendra Singh: Data curation and analysis.

Vijay Kumar Verma: Data curation and analysis.

Dhaneswar Prusty: Data curation and analysis.

Vijay Kumar Prajapati: Conceptualization, Supervision, Writing original draft and finalizing draft.

Declaration of Competing Interest

The authors have declared no competing interest.

Acknowledgement

RO is thankful to UGC for providing university fellowship. VKP is thankful to the Central University of Rajasthan for providing computational facility

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejps.2020.105375.

References

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E.J.S., 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1, 19–25.

Adedeji, A.O., Lazarus, H.J.m., 2016. Biochemical characterization of middle east respiratory syndrome coronavirus helicase. mSphere. 1 e00235-00216.

Beck, B.R., Shin, B., Choi, Y., Park, S., Kang, K.J.b., 2020. Predicting commercially available antiviral drugs that may act on the novel coronavirus (2019-nCoV), Wuhan, China through a drug-target interaction deep learning model. BioRxiv.

Benkert, P., Künzli, M., Schwede, T.J.N.a.r., 2009. QMEAN server for protein model quality estimation. Nucl. Acids Res. 37, W510–W514.

Bhardwaj, K., Guarino, L., Kao, C.C.J.J.o.v., 2004. The severe acute respiratory syndrome coronavirus Nsp15 protein is an endoribonuclease that prefers manganese as a cofactor. J. Virol. 78, 12218–12224.

Bhattacharya, D., Nowotny, J., Cao, R., Cheng, J., 2016. 3Drefine: an interactive web server for efficient protein structure refinement. Nucl. Acids Res. 44, W406–W409.

Cavanagh, D.J.V.r., 2007. Coronavirus avian infectious bronchitis virus. Vet. Res. 38, 281–297.

Chan-Yeung, M., Yu, W.J.B., 2003. Outbreak of severe acute respiratory syndrome in Hong Kong special administrative region: case report. BMJ 326, 850.

Cheng, Z.J., Shan, J., 2020. 2019 novel coronavirus: where we are and what we know. Infection. 155–163.

De Vries, S.J., Van Dijk, M., Bonvin, A.M.J.N.p., 2010. The HADDOCK web server for data-driven biomolecular docking. J. Mol. Biol. 5, 883.

Dikhit, M.R., Kumar, A., Das, S., Dehury, B., Rout, A.K., Jamal, F., Sahoo, G.C., Topno, R.K., Pandey, K., Das, V.N.R., Bimal, S., Das, P., 2017. Identification of potential MHC Class-II-Restricted Epitopes Derived from Leishmania donovani antigens by reverse vaccinology and evaluation of their CD4+ T-Cell Responsiveness against Visceral Leishmaniasis. Front. Immunol. 8 1763-1763.

Dimitrov, I., Flower, D.R., Doytchinova, I., 2013. AllerTOP-a server for in silico prediction of allergens. BMC Bioinform. BioMed. Cent. S4.

Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., Wang, T., Sun, Q., Ming, Z., Zhang, L.J.S., 2020. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science eabb7498.

Gupta, N., Regar, H., Verma, V.K., Prusty, D., Mishra, A., Prajapati, V.K.J.I.J.o.B.M., 2020. Receptor-ligand based molecular interaction to discover adjuvant for immune cell TLRs to develop next-generation vaccine. Int. J. Biol. Macromol. 152, 535–545.

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Raghava, G.P., 2015. Peptide toxicity prediction. Computational Peptidology. Springer, pp. 143–157.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X.J.T.L., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395, 497–506.

Ismail, M., Tang, Y., Saif, Y.J.A.d., 2003. Pathogenicity of turkey coronavirus in turkeys and chickens. Avian Dis. 47, 515–522.

Källberg, M., Margaryan, G., Wang, S., Ma, J., Xu, J., 2014. RaptorX server: a resource for template-based protein structure modeling. Methods Mol. Biol. 1137, 17–27.

Kao, D.J., Hodges, R.S.J.C.b., design, d., 2009. Advantages of a synthetic peptide immunogen over a protein immunogen in the development of an anti-pilus vaccine for Pseudomonas aeruginosa. Chem. Biol. Drug Des. 74, 33–42.

Larsen, M.V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., Nielsen, M.J.B.b., 2007. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinform. 8, 424.

Lewis, R., Yap, A., Easter, S., Mitchell, B., Radek, J., Rogers, R.J.D.S.B., 2020. COVID-19 vaccine will close in on the spikes. DNA Sci. Blog.

Livingston, B., Crimi, C., Newman, M., Higashimoto, Y., Appella, E., Sidney, J., Sette, A., 2002. A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. J. Immunol. 168, 5499–5506.

Lovell, S.C., Davis, I.W., Arendall 3rd, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C., 2003. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins 50, 437–450.

Mohan, T., Sharma, C., Bhat, A.A., Rao, D.J.V., 2013. Modulation of HIV peptide antigen specific cellular immune response by synthetic α-and β-defensin peptides. Vaccine 31, 1707–1716.

Muruato, L.A., Tapia, D., Hatcher, C.L., Kalita, M., Brett, P.J., Gregory, A.E., Samuel, J.E., Titball, R.W., Torres, A.G., 2017. Use of reverse vaccinology in the design and construction of nanoglycoconjugate vaccines against. Clin. Vaccine Immunol. 24, e00206–e00217.

Ojha, R., Khatoon, N., Prajapati, V.K., 2018. Conglomeration of novel Culex quinquefasciatus salivary proteins to contrive multi-epitope subunit vaccine against infections caused by blood imbibing transmitter. Int. J. Biol. Macromol. 118, 834–843.

Ojha, R., Nandani, R., Prajapati, V.K., 2019a. Contriving multiepitope subunit vaccine by exploiting structural and nonstructural viral proteins to prevent Epstein-Barr virus-associated malignancy. J. Cell Physiol. 234, 6437–6448.

Ojha, R., Pandey, R.K., Prajapati, V.K., 2020. Vaccinomics strategy to concoct a promising subunit vaccine for visceral leishmaniasis targeting sandfly and leishmania antigens. Int. J. Biol. Macromol. 156, 548–557.

Ojha, R., Pareek, A., Pandey, R.K., Prusty, D., Prajapati, V.K., 2019b. Strategic development of a next-generation multi-epitope vaccine to prevent nipah virus zoonotic infection. ACS Omega 4, 13069–13079.

Ong, E., Wong, M.U., Huffman, A., He, Y.J.B., 2020. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. BioRxiv.

Pandey, R.K., Ojha, R., Chatterjee, N., Upadhyay, N., Mishra, A., Prajapati, V.K., 2019. Combinatorial screening algorithm to engineer multiepitope subunit vaccine targeting human T-lymphotropic virus-1 infection. J. Cell Physiol. 234, 8717–8726.

Pearson, W.R., 2013. An introduction to sequence similarity ("homology") searching. Current protocols in bioinformatics Chapter 3, Unit3.1.

Peiris, J., Guan, Y., Yuen, K.J.N.m., 2004. Severe acute respiratory syndrome. Proc. Natl. Acad. Sci. 10, S88–S97.

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 e9862-e9862.

Saha, S., Raghava, G.J.N.a.r., 2006. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucl. Acids Res. 34, W202–W209.

Saha, S., Raghava, G.P., 2007. Prediction methods for B-cell epitopes. Immunoinformatics. Springer, pp. 387–394.

Sette, A., Sidney, J., 1999. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics 50, 201–212.

Shen, W.-J., Zhang, X., Zhang, S., Liu, C., Cui, W., 2018. The utility of supertype clustering in prediction for class II MHC-peptide binding. Molecules 23, 3034.

Shi, J., Zhang, J., Li, S., Sun, J., Teng, Y., Wu, M., Li, J., Li, Y., Hu, N., Wang, H.J.P.o., 2015. Epitope-based vaccine target screening against highly pathogenic MERS-CoV: an in silico approach applied to emerging infectious diseases. PLoS One. 10.

Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J., Gorbalenya, A.E.J.J.o.m.b., 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J. Mol. Biol. 331, 991–1004.

Su, S., Wong, G., Shi, W., Liu, J., Lai, A.C., Zhou, J., Liu, W., Bi, Y., Gao, G.F.J.T.i.m., 2016. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. Trends Microbiol. 24, 490–502.

Tomar, S., Johnston, M.L., John, S.E.S., Osswald, H.L., Nyalapatla, P.R., Paul, L.N., Ghosh, A.K., Denison, M.R., Mesecar, A.D.J.J.o.B.C., 2015. Ligand-induced dimerization of Middle East Respiratory Syndrome (MERS) Coronavirus nsp5 Protease (3CLpro) implications for nsp5 regulation and the development of antivirals. J. Biol.

Chem. 290, 19403–19422.
Totura, A.L., Whitmore, A., Agnihothram, S., Schäfer, A., Katze, M.G., Heise, M.T., Baric, R.S.J.M., 2015. Toll-like receptor 3 signaling via TRIF contributes to a protective innate immune response to severe acute respiratory syndrome coronavirus infection. MBio 6 e00638-00615.
Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J., 2005. GROMACS: fast, flexible, and free. J. Comput. Chem. 26, 1701–1718.
Vita, R., Overton, J.A., Greenbaum, J.A., Ponomarenko, J., Clark, J.D., Cantrell, J.R., Wheeler, D.K., Gabbard, J.L., Hix, D., Sette, A., Peters, B., 2015. The immune epitope database (IEDB) 3.0. Nucl. Acids Res. 43, D405–D412.
Yano, A., Onozuka, A., Asahi-Ozaki, Y., Imai, S., Hanada, N., Miwa, Y., Nisizawa, T., 2005. An ingenious design for peptide vaccines. Vaccine 23, 2322–2326.
Zhou, P., Fan, H., Lan, T., Yang, X.-.L., Shi, W.-.F., Zhang, W., Zhu, Y., Zhang, Y.-.W., Xie, Q.-.M., Mani, S.J.N., 2018. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. Nature 556, 255–258.
Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R.J., 2020. A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. 382 (8), 727–733.
Ziebuhr, J., 2005. The coronavirus replicase. Coronavirus Replication and Reverse Genetics. Springer, pp. 57–94.
Zumla, A., Memish, Z.A., Hui, D.S., Perlman, S.J.T.L.I.D., 2019. Vaccine against Middle East respiratory syndrome coronavirus. Lancet Infect. Dis. 19, 1054–1055.