



ORIGINAL RESEARCH

# Designing an efficient multi-epitope vaccine displaying interactions with diverse HLA molecules for an efficient humoral and cellular immune response to prevent COVID-19 infection

Soumya Ranjan Mahapatra<sup>a</sup>, Susrita Sahoo<sup>a</sup>, Budheswar Dehury<sup>a</sup>, Vishakha Raina<sup>a</sup>, Shubhransu Patro<sup>c</sup>, Namrata Misra<sup>a,b</sup> and Mrutyunjay Suar<sup>a,b</sup>

<sup>a</sup>School of Biotechnology, Kalinga Institute of Industrial Technology (KIIT-DU), Bhubaneswar 751024, India; <sup>b</sup>KIIT-Technology Business Incubator (KIIT-TBI), Kalinga Institute of Industrial Technology (KIIT-DU), Bhubaneswar 751024, India; <sup>c</sup>Kalinga Institute of Medical Sciences (KIMS) Kalinga Institute of Industrial Technology (KIIT-DU), Bhubaneswar 751024, India

## ABSTRACT

**Background:** The novel SARS-CoV-2 coronavirus, the causative agent of the ongoing pandemic COVID-19 disease continues to infect people globally and has infected millions of humans worldwide. However, no effective vaccine against this virus exists.

**Method:** Using Immunoinformatics, epitopic sequences from multiple glycoproteins that play crucial role in pathogenesis were identified. Particularly, epitopes were mapped from conserved receptor-binding domain of spike protein which have been experimentally validated in SARS-CoV-1 as a promising target for vaccine development.

**Results:** A multi-epitopic vaccine construct comprising of B-cell, CTL, HTL epitopes was developed along with fusion of adjuvant and linkers. The epitopes identified herein are reported for the first time and were predicted to be highly antigenic, stable, nonallergen, nontoxic and displayed conservation across several SARS-CoV-2 isolates from different countries. Additionally, the epitopes associated with maximum HLA alleles and population coverage analysis shows the proposed epitopes would be a relevant representative of large proportion of the world population. A reliable three-dimensional structure of the vaccine construct was developed. Consequently, docking and molecular-dynamics simulation ensured the stable interaction between vaccine and innate-immune receptor.

## ARTICLE HISTORY

Received 4 May 2020  
Accepted 13 August 2020

## KEYWORDS

SARS-CoV-2; COVID-19;  
multi-epitopic vaccine;  
immunoinformatics;  
receptor-binding domain

## 1. Introduction

The recent emergence of highly pathogenic SARS-CoV-2 coronaviruses, the causal agent of COVID-19, has created a global concern for public health [1,2]. Since its first emergence in late 2019 in China, the virus has infected more than 7 million individuals and 414,340 deaths, with exponential growth of numbers in many countries. The clinical symptoms associated with this virus infection includes fever, headache trailed by severe respiratory distress and the most severe pathology in individuals above 60 years of age and especially in those with underlying chronic illness like diabetes and hypertension [3]. Moreover, COVID-19 is highly transmissible and the reproductive number ( $R_0$ ) of the infection is estimated to be 2.2, which indicates that one infected individual either symptomatic or asymptomatic can transmit the virus to 2.2 other individuals in a population with a mean incubation period of 5.8 days [3]. Given the severity of the disease, which is above that of seasonal influenza or pandemic H1N1 influenza and the ability of rapid transmission of virus by asymptomatic individual triggering its ability to cause pandemic outbreak, development of a safe and effective vaccine to fight against this novel virus is urgently needed [4,5]. Despite, frequent

periodic outbreaks of human coronavirus infection reported every decade in the twenty-first century like severe acute respiratory syndrome coronavirus (SARS-CoV-1) epidemic in 2002, middle- east respiratory syndrome coronavirus (MERS-CoV) in 2012 and now SARS-CoV-2, till date there are no existing vaccines available in the market to tackle this severe pathogenic virus.

Comparative homology between SARS-CoV-2 and SARS-CoV-1 genomes, both belonging to genus *Betacoronavirus* and exhibiting genome size of approximately 30 Kb [6] showed high sequence and structure conservation, speculating that both coronaviruses shares the same receptor, ACE2 (angiotensin-converting enzyme 2) found in human lungs [7]. On this basis, the lessons learnt through the challenges faced during SARS-CoV-1 vaccine trials like undesired immunopotentiation [8], waning of antibody response after short term [9] should be prioritized while designing SARS-CoV-2 vaccine strategies. Another consideration for effective SARS-CoV-2 vaccine development is the intended target population. Because, the clinical manifestation of COVID-19 disease is more severe in older individuals due to immune senescence [10], it is important to develop vaccine that protects this high-risk segment of the population with reduced immunity. Many leading multinational companies/consortium are presently working relentlessly exploring

**CONTACT** Mrutyunjay Suar msbiotek@yahoo.com; mrutyunjay@kiitincubator.in School of Biotechnology, Kalinga Institute of Industrial Technology (KIIT-DU), Bhubaneswar 751024, India; Fax, (+91) 674 272 5732

Supplemental data for this art can be accessed here.

© 2020 Informa UK Limited, trading as Taylor & Francis Group

different vaccine production platforms and technologies for SARS-CoV-2 like whole virus vaccines (Johnson & Johnson, Codagenix/Serum Institute of India); subunit vaccines (Novavax, Clover biopharmaceuticals, University of Queensland/CEPI); nucleic acid vaccines (Inovio/Beijing Advaccine Biotechnology Co./CEPI, CureVac/CEPI/, Moderna/NIH/CEPI) [11]. However, current vaccine development through conventional techniques might take years and each of the strategy has its distinct advantages and disadvantages leading to serious health consequences as adverse side effects [12]. Considering the present scenario of the COVID-19 pandemic worldwide, it is therefore essential to quickly integrate different vaccine design approaches so as to minimize both time and costs in the process of target vaccine candidate identification, which can then be evaluated for efficacy and safety.

The availability of pathogen whole genomes and a pool of robust vaccinomics driven sequence analysis tools have made it possible to computationally search on a genome scale for promiscuous epitopes as potential vaccine candidates which has the ability to elicit desired immune response against the pathogenic organism [13]. Subunit vaccine formulations containing cocktail of defined T-cell and B-cell multi-epitopes which determines the antigenic part of the pathogens have the potential to stimulate both humoral and cellular immune systems within the human body. Besides nullifying the chances of pathogenicity reversal as in case of live attenuated vaccines, the major advantage of employing subunit vaccine includes its efficacy in a person with reduced immunity. Application of epitopic vaccine construct has been successfully demonstrated in several disease conditions like HIV-1 [14], cancer [15–17], and also allergic and autoimmune diseases [18].

The four important glycoproteins that majorly contribute to the structure of all coronaviruses are the spike protein (S), small envelope protein (E), membrane protein (M), and nucleocapsid (N) protein [9]. Particularly, the spike protein, containing putative S1 and S2 domains, mediates coronavirus entry into the host cells by binding to the ACE2 receptor which triggers a cascade of events leading to fusion of both viral and host membranes [19,20]. A recent report on the structure of SARS-CoV-2 spike showed that specifically a fragment located in the middle region of the S1 subunit defined as ‘receptor-binding domain’ (RBD; residue 317–569) of human novel coronavirus ARS-CoV-2 strain spike protein (Uniprot Id P59594) binds to the peptidase domain of ACE2 [21]. Therefore, it is proposed that the RBD region of the spike protein should be used for the development of SARS-CoV-2 vaccine, because it is not only a functional domain for receptor-binding and viral entry but also a critical neutralization determinant of coronaviruses [22–24]. Furthermore, experimental data suggests that RBD-based vaccines may induce broad protection against both human and animal SARS-CoV variants [22]. Besides spike protein, other structural glycoproteins also have diverse functional role in the viral pathogenesis, like the membrane protein mediates assembly and binding of virus particles [25], nucleocapsid protein packages the viral genome into a helical ribonucleocapsid and hence fundamentally assist in viral self-assembly and pathogenesis [26,27] while the envelope protein play an important role in the assembly of the virion and in virus replication [28].

Recently, few studies have been reported on epitope identification for SARS-CoV-2 vaccine construct using immunoinformatics approach [29–32]. However, these *in silico* analysis are majorly on spike protein and most importantly, regardless of the importance of RBD region of spike protein in the pathogenesis of SARS-CoV-2 virus, none have focused on identification of conserved epitopes spanning the antigenic RBD region. An advantage of subunit vaccine comprising of only the RBD of the SARS-CoV spike protein is shown to minimize host immunopotentiation [8]. Consequently, this study aims to design a multi-epitope SARS-CoV-2 vaccine construct using all antigenic surface structural proteins, including the RBD-spike protein, envelope protein, membrane protein, and nucleocapsid protein. Promiscuous vaccine candidates comprising of most potential T-cell and B-cell epitopes from each of the four antigenic protein sequences have been mapped along with analysis of the epitopes conservancy, potential immunogenicity, allergenicity, toxicity, and conservancy analysis against different isolates to ensure formulation of robust widely protective SARS-CoV-2 vaccine with the ability to generate the humoral as well as cell mediated immunity. Physicochemical characterization and population coverage analysis of the final multi-epitope vaccine construct was assessed followed by homology modeling and docking of validated vaccine model with the immune receptor (TLR-3) present on lymphocytes cells. Consequently, the stability of the interaction of the vaccine with the receptor was determined by molecular dynamics simulation studies.

## 2. Methodology

### 2.1. Selection of SARS-CoV-2 proteins for potential epitope screening

The major structural glycoproteins that majorly contribute to the structure of SARS-CoV2 and play vital roles in viral pathogenesis were selected for the purpose of vaccine development. The sequences of spike protein (UniProt Id: P59594), envelope protein (UniProt Id: P59637), membrane protein (UniProt Id: P59596), and nucleocapsid protein (UniProt Id: P59595) were retrieved from UniProt database (<http://www.uniprot.org/>) in FASTA format. Further, in order to avoid cross-reactivity and generation of autoimmune disorder, sequence similarity search of each of the above selected proteins were done to find out the homology with the human proteome by using NCBI-Protein-Protein Blast ([https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE\\_TYPE=BlastSearch&LINK\\_LOC=blasthome](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)).

### 2.2. Mapping of B-cell epitopes

The humoral immune response specific B-cell epitopes were predicted for the aforementioned SARS-CoV-2 proteins by using the ABCpred server (<http://crdd.osdd.net/raghava/abcpred/>) at the threshold of 0.51 and window length of 10. The server with 65.93% accuracy, utilizes feed forward and recurrent neural network algorithm based on training data set of 700 B-cell epitopes and equal number of non-B-cell epitopes from SwissProt database [33].

## 2.3. Mapping of T-cell epitopes

### 2.3.1. Cytotoxic T-Lymphocyte (CTL)Epitopes Mapping

The NetMHCpan 4.1 Server (<http://www.cbs.dtu.dk/services/NetMHCpan-4.1/>) was used to screen the antigenic CTL epitopes for the selected SARS-CoV-2 coronavirus proteins. NetMHCpan server predicts binding of peptides to any MHC molecule of known sequence using artificial neural networks. The binding affinity data covers major twelve supertypes of human, including five major supertypes of HLA-A locus, A1, A2, A3, A24, and A26, and the HLA-B locus of seven major supertypes, B7, B8, B27, B39, B44, B58, and B62 [34] (Supplementary Table S1). For the present study representative HLA alleles belonging to the major twelve HLA supertypes were selected. In addition to NetMHCpan 4.0 Server, NetCTL 2.0 (<http://www.cbs.dtu.dk/services/NetCTL/>) and IEDB (Immune Epitope Database And Analysis Resource) tool 'Proteasomal cleavage/TAP transport/MHC Class I combined predictor' (<http://tools.immuneepitope.org/mhci/>) servers. The total score generated by NetCTL 2.0 for each CTL epitope gives a combined scoring which includes the proteasomal cleavage of the C-terminal end of the peptide, TAP (N-terminal) transport efficiency, and MHC Class I affinity [35]. IEDB predicts the antigenic CTL epitopes and their binding to MHC Class I alleles based on the IC<sub>50</sub> (nM) score [36]. Epitopes with IC<sub>50</sub> values are predicted as high (<50 nM), intermediate (<500 nM) and of least affinity (<5000 nM) for different epitopes. The top three high scoring CTL epitopes that were consensually predicted by all the tools, NetMHCpan 4.1, NetCTL, and IEDB tools for each protein were selected for further analysis.

### 2.3.2. Helper T-Lymphocyte (HTL) epitopes mapping

The NetMHCIIPan-4.0 server (<http://www.cbs.dtu.dk/services/NetMHCIIPan-4.0/>) was used to predict the immunogenic HTL epitopes [34]. The server predicts peptide binding to any MHC II molecule of known sequence using Artificial Neural Network. It is trained on an extensive dataset of HLA alleles, covering the three human MHC class II isotypes HLA-DR, HLA-DQ, HLA-DP, as well as the mouse molecules (H-2). All HLA alleles belonging to DRB1, DRB3, DRB4, DRB5, DP, and DQ were selected for the present epitope prediction (Supplementary Table S1). The high scoring epitopes predicted by NetMHCIIPan-4.0 server were also validated using the IEDB tool 'MHC II Binding predictions' (<http://tools.iedb.org/mhcii/>) by selecting IEDB recommended alleles, consisting of a reference panel of 27 HLA alleles (Supplementary Table S1). The epitopes that showed strong binding to maximum number of HLA alleles by both the tools were finally selected for further analysis.

## 2.4. Characterization of screened epitopes

### 2.4.1. Epitope conservation analysis

Conservancy analysis of all the shortlisted CTL, HTL and B-cell epitopes was performed by 'Epitope Conservancy Analysis tool of IEDB' (<http://tools.iedb.org/conservancy/>). The tool computes the percentage of epitope sequence conservancy amongst a given set of protein sequences at a given identity level. The higher the percentage of different SARS-CoV-2 strains carrying the screened

epitope in their protein sequences, the greater would be its effectiveness as a widely protective vaccine candidate.

### 2.4.2. Toxicity prediction analysis

The ToxinPred tool (<https://webs.iiitd.edu.in/raghava/toxinpred/index.html>) that employs support vector machine method [37] was used for the prediction of possible toxicity of all the screened CTL, HTL and B-cell epitopes. The tool allows identifying highly toxic or nontoxic short peptide sequences along with analysis of hydrophobicity, hydropathicity, hydrophilicity, charge, and molecular weight of the predicted peptides.

### 2.4.3. Antigenicity analysis

The antigenicity of the screened epitopes was calculated using VaxiJen 2.0 [<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>]. VaxiJen uses an alignment-free approach to identify antigens directly [38]. Instead of concentrating on predicting antigenic regions, the method uses statistical models based on discriminating positive and negative sets of bacterial, viral, and tumor antigen datasets to predict whole-protein antigenicity. VaxiJen has shown high prediction accuracy of up to 89%.

### 2.4.4. Allergenicity analysis

Allergens are proteins or glycoproteins recognized by immunoglobulin E (IgE), which is produced by the immune system in allergic individuals. It is said that a protein is considered an allergen if it has a region or peptides identical to a human IgE epitope [39]. Allergenicity of the epitopes was checked by the AllerTOPV 2.0 tool [<https://www.ddg-pharmfac.net/AllerTOP/index.html>]. The tool has been developed based on the robust k-nearest neighbors (kNN) method and predicts allergenicity/nonallergenicity with 86.7% sensitivity, 90.7% specificity, and 88.7% accuracy [40].

## 2.5. Criteria for selection of final multi-epitopes for vaccine construct

For each of the four proteins viz., spike, membrane, nucleocapsid, and envelope, one HTL epitope, one B-cell epitopes and 3 CTL epitopes (that were consensually predicted as high scoring epitopes binding to maximum number of HLA alleles by the NetMHCpan 4.1, NetCTL as well as IEDB tool) were selected for final vaccine development based on the following criteria (1) highest antigenicity score (2) high score by the predicted algorithm, (3) promiscuous peptides, which means epitopes binding to numerous histocompatibility alleles of the world population, (4) nonallergenicity and nontoxic (5) conservancy above 80% across multiple isolates and strains.

## 2.6. Vaccine development

For devising the multi-epitope subunit vaccine, the final selected B-cell, CTL, and HTL epitopes of corresponding spike, envelope, membrane, and nucleocapsid proteins were combined together in the same sequential manner to obtain the final vaccine construct. The N-terminal of the designed vaccine sequence was adjuvanted with β-defensin, which is

known to exhibit antimicrobial and antiviral activity and the ability to enhance protective immune response [41]. Additional sequences that play a vital role in the chimeric vaccine construct are linkers, acting as spacers that facilitate protein folding and stability to achieve a proper epitope display for binding to MHC Class I and MHC Class II molecules [42]. In this study, the helix forming linker 'EAAAK' was added to join the adjuvant and the first B-cell epitopes, while 'KK' 'AAY,' and 'GPGPG' linkers were added to join together all B-cell epitopes, CTL epitopes and HTL epitopes, respectively. These linkers were appended to attain the efficient separation of individual epitopes *in vivo* [43].

### 2.7. Population coverage analysis of vaccine construct

The predicted epitopes constituting the vaccine construct with the corresponding HLA alleles (Class I and Class II) were submitted to the population coverage analysis tool of the IEDB database (<http://tools.iedb.org/population/>) by keeping the default parameters on (109 countries covering 16 different geographical regions) (Supplementary Table S2). Population coverage analysis tool calculates the fraction of individuals predicted to respond to a given set of epitopes with known MHC restrictions. For individual epitope, the tool computes the following: (i) projected population coverage, (ii) average number of epitope hits/HLA combinations recognized by the populations, and (iii) minimum number of epitope hits/HLA combinations recognized by 90% of the population (PC90). These calculations are made on the basis of HLA genotypic frequencies assuming nonlinkage disequilibrium between HLA loci [44].

### 2.8. Physicochemical features and secondary structure prediction of the vaccine construct

Physicochemical properties of the vaccine construct comprising of amino acid composition, molecular weight, number of positive/negatively charged amino acids, instability index, aliphatic index, GRAVY, hydropathy plot, and solubility was analyzed using the ProtParam tool (<https://web.expasy.org/protparam/>) [45]. Instability index was calculated to estimate the stability of proteins in a test tube. The aliphatic index shows the relative volume occupied by aliphatic side chains and GRAVY is the average of hydropathy values of amino acids in a protein [45]. The secondary structure of the predicted vaccine was carried out by online tool PSIPRED (<http://bioinf.cs.ucl.ac.uk/>) and the percentage of helix, strand, and coils were determined.

### 2.9. Tertiary structure prediction and validation of the vaccine construct

The tertiary structure of vaccine construct was generated using Robetta server [46]. The best 3D model developed was analyzed further using various structure validation tools. PROCHECK checks the geometrical and stereochemical constraints of the protein structure by inspecting the accuracy of the dihedral angles (i.e.  $\alpha$  and  $\psi$ ) in the Ramachandran plot [47]. VERIFY3D analyzes the compatibility of the 3D protein model with its amino acid sequence [48] while ERRAT

examines the nonbonded atomic interactions [49]. A higher score from VERIFY3D and ERRAT indicates a more reliable structure for further analysis.

### 2.10. Analysis of molecular interaction between vaccine candidate and immune cell receptor

Molecular docking between the receptor Toll-like receptor 3 (TLR-3) and vaccine was performed with the help of the ClusPro server (<https://cluspro.org>), which is a widely used tool for protein–protein docking [50]. Subsequently, molecular dynamics (MD) simulations was performed to investigate the dynamical motion and stability of the docked vaccine-receptor complex with the help of GROMACS version2018.4 molecular dynamics package [51,52] using Amber ff99SB-ILDN force field in TIP3P water model. The system net charge neutralization was done by adding required Na<sup>+</sup> and Cl<sup>-</sup> ions. After electro-neutralization, the solvated model was subjected to energy minimization using the steepest descent algorithm until the maximum force of 1000 kJ mol<sup>-1</sup> nm<sup>-1</sup> had been reached. Following energy minimization, using an NVT ensemble at 271 K temperature equilibration was carried out via the V-rescale method and the pressure was equilibrated using an NPT ensemble at 1 bar with the Parinello–Rahman algorithm for 2 ns. After equilibration, final production MD simulation was carried out for 50 ns. The particle mesh Ewald method was used to treat long-range electrostatic interactions. The short-range electrostatic and van der Waals interactions were calculated by specifying a cutoff 12-Å distance. The bonds involving hydrogen atoms were constrained using the linear-constraint-solving algorithm. The periodic boundary conditions were applied during production MD simulation. Trajectory data harvested from production MD was used to compute the dynamics stability of the system. The root mean square deviations (RMSD) for protein backbone and root mean square fluctuation (RMSF) for Ca-atoms were generated using gmx rms and gmx rmsf utility tools. The plots were rendered using grace program (<http://plasma-gate.weizmann.ac.il/Grace/>).

## 3. Results

### 3.1. Selection of B-cell epitopes

Recent findings have demonstrated that neutralizing antibodies nAbs targeting RBD regions of spike protein is potential for prophylaxis and treatment of human coronavirus infection including SARS, MERS and also COVID-19 other by inhibiting receptor binding and viral entry into the host cell [53–56]. Therefore, in this study, B-cell epitopes from the specific RBD region of spike protein as well as from other antigenic proteins, viz., nucleocapsid, envelope, and membrane proteins were predicted using ABCPred server. The high scoring peptides were then checked for toxicity and allergenicity using ToxinPred and AllerTOP V 2.0 tool, respectively. Out of many putative epitopes, only few peptide sequences for each protein were found to be nonallergen and nontoxic to human, and were considered for further analysis (Supplementary Table S3). Finally, the best B-cell epitope from each of the antigenic protein, with highest antigenicity score as predicted by VaxiJen server was used for vaccine construction (Table 1). Only one epitope DISNVPFSPD was predicted to fall within the RBD

**Table 1.** Predicted B-cell epitopes from SARS-CoV-2 proteins and their corresponding immunogenic properties to design multi-epitope vaccine construct.

Uniprot_ID	B-Cell Epitope	Position	Score	Conservancy	Toxicity	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt.
Spike Protein (P59594)	DISNVPFSPD	454	0.74	80.00%	Nontoxin	-0.09	-0.38	0.1	-2	1090.28
Small Envelope Protein (P59637)	VFLVTLAIL	25	0.83	100.00%	Nontoxin	0.46	3.2	-1.54	0	1101.6
Membrane Protein (P59596)	IGFLFLAWIM	23	0.77	80.00%	Nontoxin	0.48	2.46	-1.74	0	1210.7
Nucleocapsid Protein (P59595)	GDGKMKELSP	98	0.72	90.00%	Nontoxin	-0.28	-1.23	0.92	0	1061.36

region (residue 317–569) of the spike protein and the same was also found to possess highest antigenicity, thereby indicating as a potential vaccine candidate. Likewise, epitopes VFLLVTLAIL, IGFLFLAWIM, and GDGKMKELSP were selected from Small Envelope Protein, membrane protein and nucleocapsid protein, respectively. The molecular weight, charge, hydrophobicity, hydropathicity, and hydrophilicity values of the selected epitopes are given in **Table 1**.

### 3.2. Selection of CTL epitopes

CTL epitopes from the four major SARS-CoV-2 antigenic proteins were identified using the NetMHCpan 4.1 server, NetCTL server and the IEDB tool. Nontoxin, nonallergen, ten overlapping high scoring epitopes predicted by both the algorithms were selected for each protein for antigenicity analysis (Supplementary Table S4). Finally based on highest antigenicity score, three epitopes were selected for each protein. Notably CTL epitopes predicted for spike protein were VLSFELLNA, TSTGNYYNYK, and VVLSFELL; envelope protein epitopes were RLCAYCCNI, VVFLLVTLA, LAILTALRL; membrane protein epitopes were NLVIGFLFL, GTITVEELK, RINWVTGGI; and for nucleocapsid protein CTL epitopes were KLDDKDPQF, ELSPRWYFY, and DPQFKDNVI (**Table 2**). The affinity between identified CTL epitopes and their corresponding major histocompatibility complex (MHC) class I alleles alongwith their immunogenic properties is shown in **Table 2**.

### 3.3. Selection of HTL epitopes

The HTL epitopes were predicted using the NetMHCIIpan-4.0 server and the IEDB server as described in the methodology section. Out of many epitopes identified (Supplementary Table S5), based on antigenicity score and binding affinity to maximum number of HLA alleles as predicted by both the tools, the final four HTL epitopes selected each for spike, envelope, membrane, and nucleocapsid proteins were VVLSFELLNAPATVC, LLFLAFVVFLVTLA, IGFLFLAWIMLLQFA, and PRWYFYYLGTGPEAS, respectively (**Table 3**).

### 3.4. Epitope conservation analysis

Genomes of 96 recently sequenced SARS-CoV-2 isolates from different origin were received from NCBI and their corresponding spike, membrane, envelope and nucleocapsid protein sequences (Supplementary Table S6). Consequently, each of the shortlisted CTL, HTL, and B-cell epitopes were analyzed against the corresponding homologous protein sequences using the Conservancy Analysis tool of IEDB. It was observed that all selected epitopes exhibited conservancy of more than 80% in all isolates from different counties (**Tables 1–3**) while

eight antigenic epitopes out of total twenty showed complete 100% conservation.

### 3.5. Designing a multi-epitope chimeric vaccine construct

A major challenge for current vaccine development is the fact that many new subunit vaccines are poorly immunogenic and elicit insufficient immune responses for protective immunity. Adjuvants are therefore needed in vaccine formulations to enhance the immune response and for proper delivery of vaccine antigens to antigen-presenting cells [42]. Hence, β-defensin (TLR-3 agonist) was added at the N-terminal region of the vaccine sequence to induce humoral and cellular immunity [57]. Additional sequences that play an important role in designing of a vaccine construct are linkers which acts as spacers for facilitating folding in order to achieve a proper epitope display [58]. Finally, the designed chimeric vaccine construct was composed of adjuvant, 4 B-cell epitopes, 12 CTL epitopes, 4 HTL epitopes linked together in the same order utilizing EAAAK, KK, AAY, and GPGPG linkers, respectively. The schematic illustration of the multi-epitope vaccine construct is shown in **Figure 1**.

### 3.6. Population coverage analysis of vaccine construct

MHC molecules are extremely polymorphic in their peptide binding regions, and as a result, they exhibit widely varying binding specificity. Therefore, population coverage analysis is crucial for design of applicable epitope-based subunit vaccines that bind to several alleles of HLA supertypes [59]. In order to calculate the population coverage of the predicted putative epitopes, the predicted putative epitopic core sequences along with their binding MHC Class I and MHC Class II alleles were submitted to the population coverage analysis tool housed at the Immuno Epitope Database.

The IEDB results corroborated that with the administration of designed vaccine construct, 98.81% of the world population would be covered, which encompasses 109 countries covering 16 different geographical regions (**Figure 2**; Supplementary Table S2). It was noted that the population coverage rate exceeded 90% in almost all countries across worldwide (**Figure 2**; Supplementary Table S2).

### 3.7. Physicochemical features and secondary structure prediction of the vaccine construct

The physicochemical characteristics, such as the molecular weight, stability, hydrophilicity, charge of the vaccine construct is crucial for their uptake, processing and antigen presentation, and the concomitant induction of immunity [42]. The molecular weight of the final vaccine construct was



Table 2. Predicted CTL epitopes from SARS-CoV-2 proteins to design multi-epitope vaccine construct with their corresponding MHC Class I alleles and their immunogenic properties.

Proteins	Epitope	High affinity HLA Alleles binding to predicted epitopes as analyzed using the NetMHCpan 4.1 server	NetCTL Score	4.1 Score	NetMHCpan Position	Conservancy	Toxicity	Hydrophobicity	Hydrophilicity	Charge	Mol wt.
Spike Protein (P59594)	VLSFELLNA	HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-A*26:01, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05, HLA-B*39 :01, HLA-B*4:01, HLA-B*58:01, HLA-B*15:01	0.8431	0.7462	498–506	88.89%	Nontoxin	0.14	1.08	-0.64	-1 1102.44
TSTGNVNYK		HLA-A*03:01, HLA-A*02:01, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05, HLA-B*39 :01, HLA-B*4:01, HLA-B*58:01, HLA-B*15:01	0.8456	1.0140	431–439	80.00%	Nontoxin	-0.28	-1.74	-0.4	1 1210.4
WVVLSEELL		HLA-A*24:02, HLA-A*02:01, HLA-A*01:01, HLA-A*03:01, HLA-A*26:01, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05, HLA-B*39 :01, HLA-B*4:01, HLA-B*58:01, HLA-B*15:01	0.5299	3.1896	496–504	100.00%	Nontoxin	0.23	1.9	-0.89	-1 1132.52
Small Envelope Protein (P59637)	RLCAYCCNI	HLA-A*02:01, HLA-B*07:01, HLA-A*03:01, HLA-A*24:02, HLA-A*26:01, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05, HLA-B*39 :01, HLA-B*4:01, HLA-B*58:01, HLA-B*15:01	0.8192	6.1136	38–46	100.00%	Nontoxin	-0.02	1.25	-0.77	1 1157.56
WFLLVTLA		HLA-A*02:01, HLA-A*01:01, HLA-A*03:01, HLA-A*24:02, HLA-A*26:01, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05, HLA-B*39 :01, HLA-B*4:01, HLA-B*58:01, HLA-B*15:01	0.3465	3.9037	24–32	100.00%	Nontoxin	0.46	3.24	-1.51	0 1087.57
Membrane Protein (P59596)	LALTALRL	HLA-B*58:01, HLA-A*02:01, HLA-A*01:01, HLA-A*03:01, HLA-A*24:02, HLA-A*26:01, HLA-A*26:01, HLA-B*07:02, HLA-B*08:01, HLA-B*27 :05, HLA-B*39:01, HLA-B*40:01, HLA-B*15:01	0.5001	1.6850	31–39	100.00%	Nontoxin	0.15	2.06	-0.84	1 1086.55
	NLVIGFFFL	HLA-A*02:01, HLA-A*01:01, HLA-A*03:01, HLA-A*24:02, HLA-A*26:01, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05, HLA-B*39 :01, HLA-B*4:01, HLA-B*58:01, HLA-B*15:01	0.7903	1.0027	20–28	100.00%	Nontoxin	0.38	2.36	-1.4	0 1106.53
	GTITVEELK	HLA-A*03:01, HLA-A*02:01, HLA-B*08:01, HLA-B*27:05, HLA-B*39 :01, HLA-B*4:01, HLA-B*58:01, HLA-B*15:01	0.6722	0.7941	5–13	100.00%	Nontoxin	-0.14	-0.37	0.33	-1 1117.42
	RINWWVTGGI	HLA-B*07:02, HLA-A*02:01, HLA-B*08:01, HLA-B*27:05, HLA-B*39 :01, HLA-B*4:01, HLA-B*58:01, HLA-B*15:01	0.5665	4.0169	71–79	88.89%	Nontoxin	0.04	0.46	-0.62	1 1086.41
Nucleocapsid Protein (P59595)	KLDDKDPQF	HLA-A*02:01, HLA-A*01:01, HLA-B*08:01, HLA-B*27:05, HLA-B*39 :01, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05, HLA-B*39 :01, HLA-B*4:01, HLA-B*58:01, HLA-B*15:01	0.5527	0.4409	339–347	88.89%	Non-Toxin	-0.51	-2.07	1.39	0 1233.52
	ELSPRMVYFY	HLA-A*26:01, HLA-A*02:01, HLA-B*08:01, HLA-B*27:05, HLA-B*39 :01, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05, HLA-B*39 :01, HLA-B*4:01, HLA-B*58:01, HLA-B*15:01	1.0943	0.0760	104–112	88.89%	Nontoxin	-0.11	-0.86	-0.83	0 1423.72
	DPQFKDNVI	HLA-B*03:01, HLA-A*02:01, HLA-B*08:01, HLA-B*27:05, HLA-B*39 :01, HLA-B*4:01, HLA-B*58:01, HLA-B*15:01	0.4606	0.6519	344–352	80.00%	Nontoxin	-0.15	-0.42	0.18	-1 1188.49

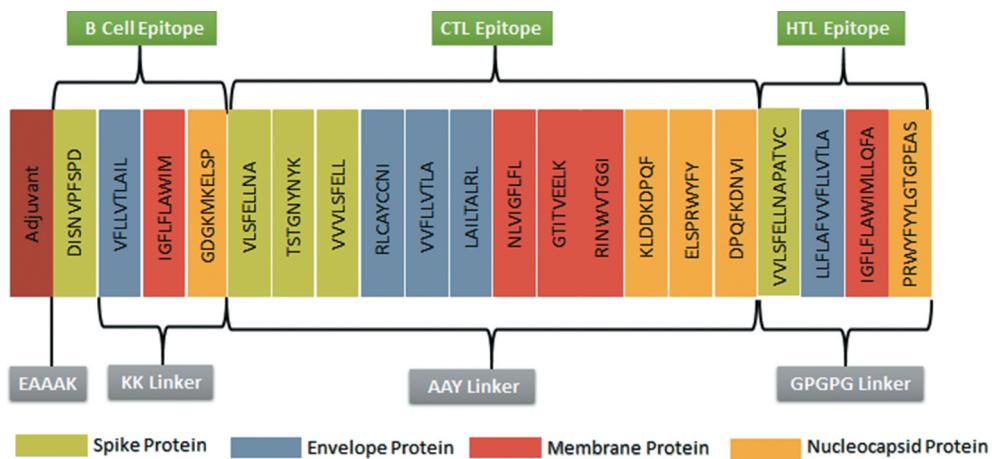
**Table 3.** Predicted HTL epitopes from SARS-CoV-2 proteins to design multi-epitope vaccine construct with their corresponding MHC Class II alleles and their immunogenic properties.

Proteins	MHC II Epitope	High affinity HLA Alleles binding to predicted epitopes as analyzed using the NetMHCIIpan-4.0 server	IC50 value	NetMHCIIpan-4.0 Score	Conservancy	Toxicity	Hydrophobicity	Hydropathicity	Charge	Mol wt.	
Spike Protein (P59594)	WLSFELLNAPATV/C	DRB1*01:12, DRB1*01:15, DRB1*01:17, DRB1*01:24, DRB1*01:25, DRB1*01:27, DRB1*03:31, DRB1*03:46, DRB1*03:48, DRB1*04:27, DRB1*04:41, DRB1*07:01, DRB1*09:02, DRB1*11:35, DRB1*11:65, DRB1*11:78, DRB1*12:03, DRB1*12:18, DRB1*13:10, DRB1*13:12, DRB1*14:52, DRB1*14:53, DRB1*15:26, DRB1*15:44, DRB1*16:14, DRB3*01:08, DRB3*01:14, DRB3*02:04, DRB3*02:23, DRB3*02:25, DRB4*01:04, DRB4*01:07, DRB4*01:08, DRB5*01:02, DRB5*01:03, DRB5*02:05, DPA1*02:04, DRB1*48:01, DPA1*02:02, DRB1*11:80:1, DPA1*02:01DRB1*69:01, DQA1*01:04DQB1*03:07, DQA1*03:03 DQB1*06:35, DQA1*04:04DQB1*05:01, DRB1*01:27, DRB1*01:35, DRB1*03:38, DRB1*03:46, DRB1*03:48, DRB1*04:29, DRB1*04:41, DRB1*07:05, DRB1*09:02, DRB1*11:35, DRB1*11:65, DRB1*11:78, DRB1*12:07, DRB1*12:12, DRB1*13:10, DRB1*13:12, DRB1*14:52, DRB1*14:53, DRB1*15:21, DRB1*15:44, DRB1*16:16, DRB3*01:05, DRB3*01:12, DRB3*02:04, DRB3*02:12, DRB3*02:23, DRB3*02:25, DRB3*03:03, DRB4*01:01, DRB4*01:03, DRB4*01:06, DRB4*01:07, DRB4*01:08, DRB5*01:02, DRB5*01:08 N, DRB5*02:02, DRB5*02:03, DPA1*02:04, DRB1*48:01, DPA1*02:02, DRB1*11:80:1, DPA1*04:01DRB1*69:01, DQA1*01:04DQB1*03:07, DQA1*03:03 DQB1*06:35, DQA1*04:04DQB1*05:01, DQA1*04:04DQB1*05:01, DRB1*01:07, DRB1*01:13, DRB1*01:19, DRB1*01:26, DRB1*01:32, DRB1*01:35, DRB1*03:43, DRB1*03:46, DRB1*03:54, DRB1*04:03, DRB1*04:04, DRB1*04:16, DRB1*04:39, DRB1*04:47, DRB1*07:03, DRB1*08:06, DRB1*08:07, DRB1*08:34, DRB1*09:07, DRB1*10:02, DRB1*11:45, DRB1*11:59, DRB1*11:74, DRB1*12:07, DRB1*12:16, DRB1*13:13, DRB1*13:33, DRB1*14:42, DRB1*14:53, DRB1*15:21, DRB1*15:44, DRB1*16:16, DRB3*01:05, DRB3*01:12, DRB3*02:04, DRB3*02:12, DRB3*02:23, DRB3*02:24, DRB3*03:01, DRB3*03:03, DRB4*01:01, DRB4*01:03, DRB4*01:06, DRB4*01:07, DRB4*01:08, DRB5*01:02, DRB5*01:08 N, DRB5*01:14, DRB5*02:03, DRB5*02:04, DPA1*02:04DQB1*03:07, DPA1*04:04DQB1*06:35, DQA1*01:04DQB1*03:07, DQA1*03:03 DQB1*06:35, DQA1*04:04DQB1*05:01	497-511	4	0.50	93.33%	Nontoxin	0.17	1.52	-0.75	-1 1576.1
Small Envelope Protein (P59637)	LLFLAFLVFLVTLA		18-32	7	0.40	100.00%	Nontoxin	0.46	3.11	-1.61	0 1679.4
Membrane Protein (P59596)	IGFLFLAWMLLQFA		23-37	17	0.50	86.67%	Nontoxin	0.4	2.22	-1.59	0 1783.49

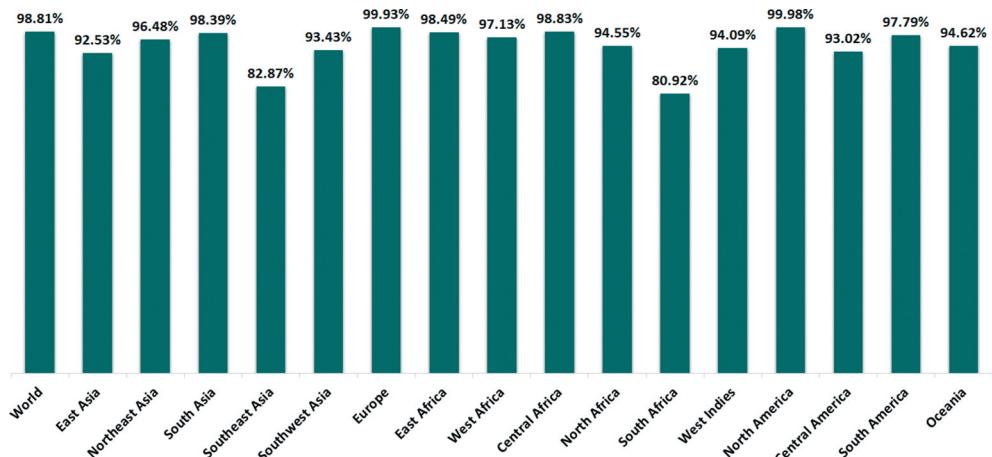
(Continued)

Table 3. (Continued).

Proteins	MHC II Epitope	High affinity HLA Alleles binding to predicted epitopes as analyzed using the NetMHCpan-4.0 server	Pos	IC50 value	NetMHCpan-4.0 Score	Conservancy	Toxicity	Hydrophobicity	Hydropathicity	Charge	Mol wt.	
Nucleocapsid Protein (P59595)	PRWYFYLLGTGPPEAS	DRB1*01:06, DRB1*01:11, DRB1*01:24, DRB1*01:29, DRB1*01:30, DRB1*01:36, DRB1*03:38, DRB1*03:40, DRB1*03:49, DRB1*04:040, DRB1*04:16, DRB1*04:34, DRB1*07:03, DRB1*07:11, DRB1*08:05, DRB1*08:21, DRB1*09:03, DRB1*09:09, DRB1*10:01, DRB1*10:02, DRB1*11:70, DRB1*11:83, DRB1*12:03, DRB1*12:18, DRB1*13:10, DRB1*13:12, DRB1*14:52, DRB1*14:53, DRB1*15:26, DRB1*15:44, DRB1*16:14, DRB3*01:13, DRB3*01:14, DRB3*02:07, DRB3*02:16, DRB3*02:23, DRB3*03:03, DRB4*01:04, DRB4*01:06, DRB4*01:07, DRB5*01:02, DRB5*01:01, DRB5*01:05, DRB5*01:11, DRB5*02:04, DPA1*03:02, DPB1*46:01, DPA1*02:02, DPB1*11:801, DPA1*04:01DPB1*69:01, DQA1*01:04DQB1*03:07, DQA1*03:03 DQB1*06:35, DQA1*04:04DQB1*05:01	107–121	44	0.50	93.33%	Nontoxin	-0.05	-0.66	-0.61	0	1807.2



**Figure 1.** Structural arrangement of the final vaccine construct.



**Figure 2.** Population coverage analysis of the final multi-epitope vaccine construct across world as predicted by the population coverage analysis tool of the IEDB database (<http://tools.iedb.org/population/>).

calculated to be 37.57 kDa that ensures it has good antigenicity. The vaccine candidate was found to be basic in nature due to its 9.26 theoretical pI value. It is composed of 19 negatively charged amino acid residues and 33 positively charged amino acid residues. It is observed that charge affects the physical stability of the formulation and the degree of antigen association. Specifically, cationic particles generally interact more efficiently with antigen presenting cells than neutral and anionic particles due to electrostatic binding to cell-surface heparan sulfate proteoglycans, thereby enhancing antigen acquisition [42]. The Aliphatic index value was found to be 109.33 which defined the high thermal stability of the vaccine in experimental set up. The estimated half-life of designed vaccine construct was predicted to be 30 hours in mammalian reticulocytes, *in vitro* and greater than 20 hours and 10 hours in yeast and *Escherichia coli*, *in vivo*, respectively. The vaccine construct was predicted to be highly stable with Instability index value less than 40 (31.12 Instability Index) (Supplementary Table S7).

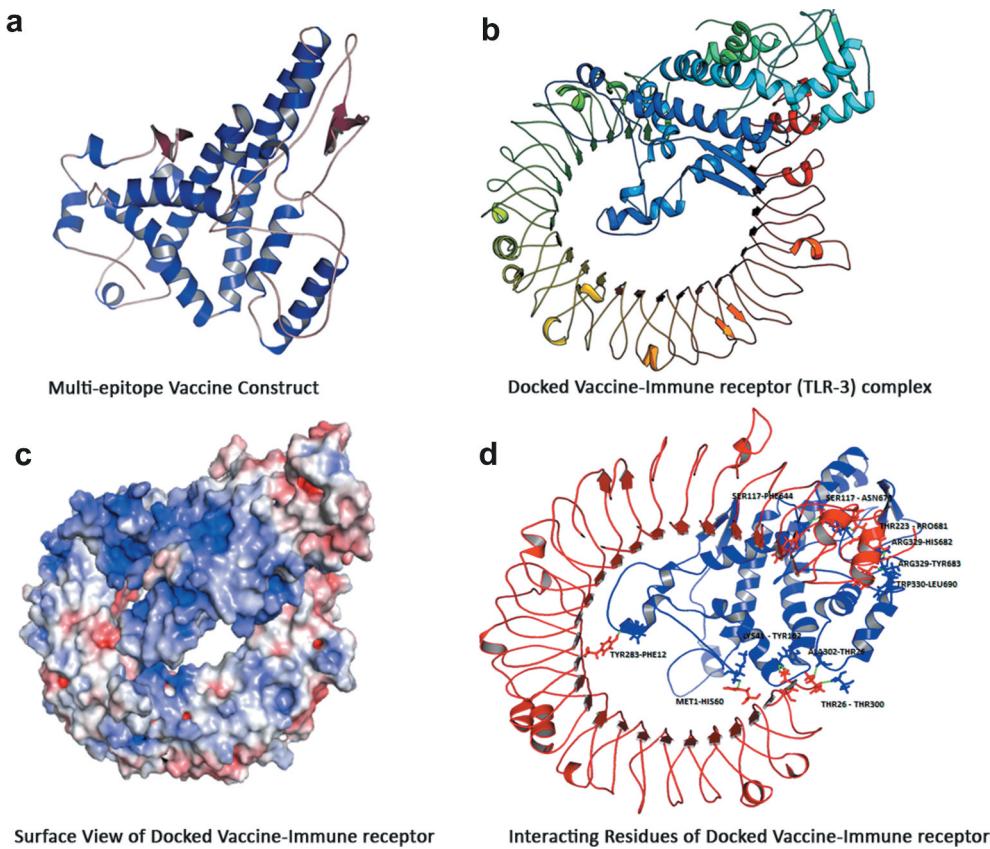
From the PSIPRED program output, it was found that secondary structure of the multi-epitope vaccine constitutes 44.44% alpha helix, 13.45% extended beta sheet, and 42.1% coils as shown in Supplementary Figure S1.

### 3.8. Tertiary structure prediction and validation of the vaccine construct

Three-dimensional structure of the multi-epitope final vaccine was modeled by using Robetta server. The final structure developed has been shown in Figure 3(a). The quality of the predicted 3D model was assessed using various structure validation tools. The calculated Ramachandran plot using PROCHECK revealed that all residues fall in allowed regions, with 92.8%, 7.2% in most favored region and additional allowed region, respectively and no single amino acid present in disallowed region. Further high scoring values, 82.7% in VERIFY3D and 89.12% in ERRAT conclusively demonstrated the accuracy and reliability of the three-dimensional structure of the multi epitopic vaccine construct for further studies (Supplementary Figure S2).

### 3.9. Analysis of molecular interaction between vaccine candidate and immune cell receptor

It has been demonstrated that TLR-3 contributes to a protective innate immune response to severe Acute Respiratory Syndrome



**Figure 3.** Molecular three-dimensional structure of (a) modeled multi-epitope vaccine construct (b) docked vaccine-immune receptor (TLR-3) complex. (c) Surface View of Docked Vaccine-Immune receptor. (d) Interacting Residues of Docked Vaccine-Immune receptor.

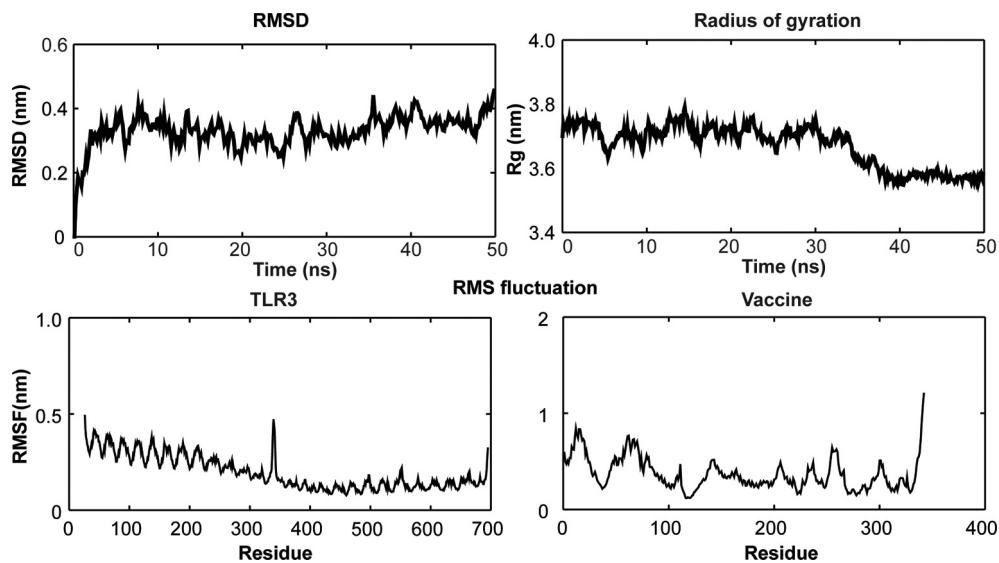
Coronavirus infection [60]. Therefore, the final vaccine construct was docked with TLR-3 (PDB Id: 2A0Z) using Cluspro server. The ribbon representation and surface view of the docked complex with lowest energy score of -919.8 as viewed using Pymol is shown in Figure 3(b,c). High binding affinity was observed between the vaccine construct and the immune receptor as displayed by multiple strong hydrogen bond (H-bond) interactions with less distance between the interacting residues. The residues of docked vaccine-TLR3 complex showing H-bond interactions were MET<sup>1</sup>-HIS<sup>60</sup>, SER<sup>117</sup>-PHE<sup>644</sup>, ARG<sup>329</sup>-HIS<sup>682</sup>, TRP<sup>330</sup>-LEU<sup>690</sup>, PHE<sup>12</sup>-TYR<sup>283</sup>, SER<sup>117</sup>-ASN<sup>678</sup>, THR<sup>223</sup>-PRO<sup>681</sup>, ALA<sup>302</sup>-THR<sup>26</sup>, ARG<sup>329</sup>-TYR<sup>683</sup>, THR<sup>300</sup>-THR<sup>26</sup>, and TYR<sup>162</sup>-LYS<sup>41</sup> with distance of 2.39 Å, 1.69 Å, 2.54 Å, 2.08 Å, 1.79 Å, 2.62 Å, 2.42 Å, 2.49 Å, 2.48 Å, 2.64 Å, and 2.26 Å, respectively (Figure 3(d)). The stability of the docked complex was determined by performing molecular dynamics simulation of 50 ns using GROMACS. The stable RMSD (Figure 4) after 40 ns affirmed the protein-receptor complex attained equilibrium and tend to display a stable trend till 50 ns. While the RMSF plot of the vaccine showed more fluctuations than receptor (TLR3) indicating that the structure of the vaccine made more movements to refine the interaction with receptor in order to elicit higher immune response.

#### 4. Discussion

SARS-CoV-2 continues to infect people globally, even under the current public health emergency with the concomitant urgency to identify appropriate therapeutic targets to prevent

and treat COVID-19 infection and control its spread [61,62]. Although SARS-CoV-1 and MERS-CoV coronaviruses outbreak originated almost a decade before, only a small number of SARS-CoV-1 and MERS-CoV reached to phase I clinical trials before funding dried up and eradication of the virus was possible through nonpharmaceutical interventions and preventive measures [63]. Nevertheless, few of the advances made in developing vaccines and therapeutics for SARS-CoV and MERS-CoV could be exploited for countering current SARS-CoV-2 [11,63]. Furthermore, besides identification of vaccine candidates, it is equally important to understand the key steps needed before bringing a SARS-CoV-2 vaccine into clinical trials. Firstly, the vaccine construct is tested in appropriate animal models to confirm it is protective. In case of SARS-CoV-2, it is seen that the virus does not grow in wild-type mice and only confers mild disease in transgenic animals. One alternative approach to validate the vaccine is by *in vitro* neutralization assays using serum from vaccinated animal. Secondly, toxicology analysis of vaccine in animal models, e.g. in rabbits has to be performed before commencing the clinical development trials [11,63].

Earlier experiments have stated that developing effective neutralizing antibodies generated by B-cell epitopes have greater potency to neutralize coronaviruses infection [53]. Besides the humoral response, the role of cell mediated immune response mediated by T- cells in combating viral infections has been known to be equally important for protecting against re-infection. Whilst neutralizing antibodies can



**Figure 4.** Root mean square deviation (RMSD) and root mean square fluctuation (RMSF) analysis of protein backbone and side chain residues of MD simulated vaccine construct.

prevent viral entry, the body also requires SARS-CoV specific CD4 + T helper (HTL) cells for the development of these specific antibodies. Similarly, CD8+ cytotoxic T cells (CTL) are essential for the recognition and killing of infected cells [64,65]. Experiments on SARS-CoV virus-infected mice models have suggested that inefficient immune activation and a poor virus-specific T cell response underlay severe infection [66,67].

The four major structural glycoproteins namely spike protein, envelope protein, membrane protein and nucleocapsid protein is known to play vital roles in coronavirus pathogenies and greater potency to fight against coronaviruses infection. Amongst the SARS-CoV structural proteins, the spike protein has been found to be primarily responsible for eliciting neutralizing antibodies with its major immunodominant epitopes found specifically in the receptor binding region of the protein which is responsible for interaction with host receptor, ACE2 [68–70]. Besides the spike glycoprotein, persistently high levels of neutralizing antibodies for nucleocapsid protein and T cell responses were also seen in the SARS-recovered individuals after two years of post-infection [71,72]. Similarly, Tsao et al. showed several epitopes in the nucleocapsid protein they could induce specific CTL responses in transgenic mice models. In another study conducted by Li et al. showed that 11% of their SARS subjects gave positive T-cell responses against epitopes present in the nucleocapsid protein [73]. Additionally, animal studies using DNA vaccines suggest that the membrane protein of coronaviruses may also induce T-cell response [74,75]. These comprehensive studies demonstrate that cocktail of immunogenic epitopes from spike, membrane, envelope, and nucleocapsid proteins could be ideal for vaccine development against the novel SARS-CoV-2 coronavirus strain.

The search for potential immunogenic epitopes to be used as vaccine candidates by conventional techniques is time consuming and too expensive. Now a days, immunologists are frequently using in silico approaches to identify the most promiscuous antigenic epitopes which has high affinity to maximum HLA alleles before being subjected to wet-lab confirmatory analysis and thus accelerating the experimental planning in development of

epitope-based vaccines. In the present study, using immunoinformatic approaches, a vaccine construct comprising of a cocktail of potential T-cell and B-cell epitopes from multiple antigenic proteins from SARS-CoV-2 have been predicted which would elicit both cellular and humoral immune responses to combat COVID-19 disease. Schematic representation of the workflow for the development of multi-epitope vaccine is shown in Figure 5. The selected epitopes were found to be highly antigenic, nonallergenic and nontoxic for human application. Further the promiscuous peptide sequences showed high conservation against different isolates to ensure formulation of robust widely protective SARS-CoV-2 vaccine. BLAST screening revealed that all the epitopes were nonself, since they lack any significant similarity with human proteome. The physicochemical characteristics, such as the molecular weight, stability, hydrophilicity, of the vaccine construct was optimum for their uptake and evoking immunity. Additionally, the selected epitopes were observed to promiscuously bind with high affinity to maximum HLA alleles, and indicating that the vaccine construct to exhibit approximately 97.97% of the world population coverage. The epitopic sequences were linked together with adjuvant and linkers to enhance immunogenicity and further the three-dimensional structure of the vaccine construct was developed using Robetta server. Various structure validation tools like Procheck, ERRAT, and VERIFY3D confirmed the reliability and accuracy of the developed structure. Efficacy of this vaccine was explored by molecular docking and dynamic simulation with TLR-3 immune receptor. Molecular docking and simulation analysis confirmed the strong and stable binding between vaccine and receptors with multiple hydrogen bond interactions.

## 5. Conclusion

The present work successfully identified novel peptides of virulent, essential and antigenic proteins, which can act as promising targets for vaccine development against SARS-CoV-2. The methodology employed utilizing immunoinformatic approach can be applied as framework for future characterization of the B-cell and

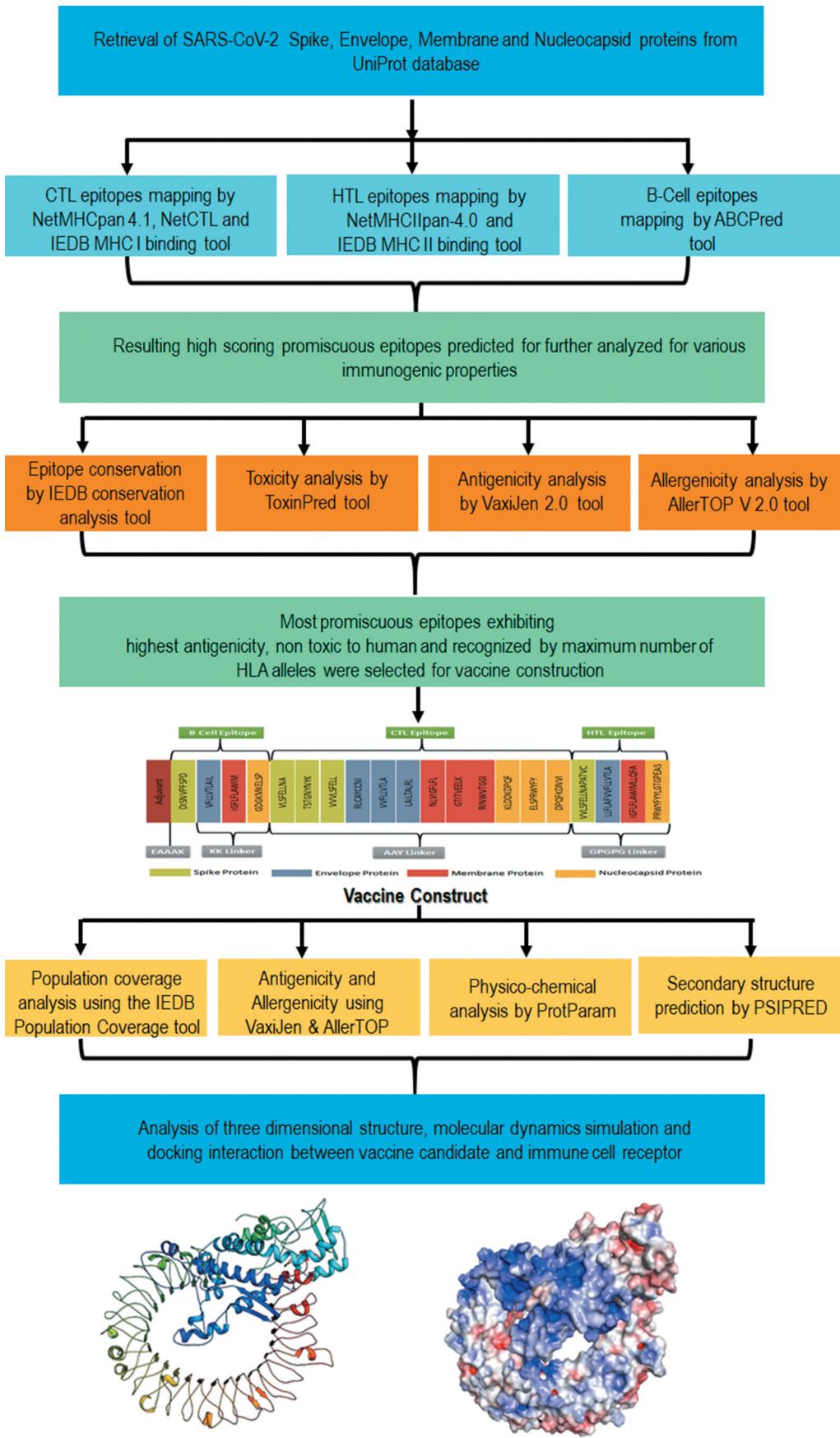


Figure 5. Schematic representation of the workflow for the development of multi-epitope vaccine against SARS-CoV-2 infection.

T-cell epitopes of emerging novel coronavirus strains as well as any other pathogenic microorganisms, and which would be of great help for advancing the peptide-based vaccine approach. However, the immune stimulation potential of the predicted multi-epitopic chimeric vaccine construct is needed to be validated in animal models for their efficacy and safety. In the present scenario of not having any direct antiviral agent and vaccines against SARS-CoV-2, strict implementation of social distancing measures and hygiene safety precautionary measures is of utmost importance to check the further spread and control of this virus.

## 6. Expert Opinion

The promiscuous epitopes capable of inducing humoral and cellular immune response can be considered as potential vaccine candidates against SARS-CoV-2.

## Acknowledgments

The authors acknowledge the School of Biotechnology, Kalinga Institute of Industrial Technology (KIIT), Deemed to be University, Bhubaneswar 751024, India for providing necessary infrastructure to carry out this work.

## Author contributions

Conception and design: MS, NM; Computational work: SM, SS; Data Analysis and Curation: NM, BD, VR; Original Draft Preparation: NM; Writing- Reviewing and Editing: MS, VR, BD, SP. The manuscript has been read and approved by all authors.

## Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

## Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose

## Supporting information

The additional information of the present computational work and details of the analysis are included in supporting information.

## References

- Papers of special note have been highlighted as either of interest (•) or of considerable interest (++) to readers.
- Gorbalenya AE, Baker SC, Baric RS, et al. The species Severe acute respiratory syndrome related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*. **2020**;5:536–544.
  - Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. **2020**;579(7798):265–269.
  - ++ This research article discusses the genome characterization of novel SARS-CoV-2 virus isolated from Wuhan and its evolutionary relationships with previously identified coronaviruses strains of Coronaviridae family.**
  - Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med*. **2020**;382(13):1199–1207.
  - Rothe C, Schunk M, Sothmann P, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *N Engl J Med*. **2020**;382(10):970–971.
  - Chakraborty C, Sharma AR, Bhattacharya M, et al. The 2019 novel coronavirus disease (COVID-19) pandemic: A zoonotic prospective. *Asian Pac J Trop Med*. **2020**;13(6):242.
  - Wu F, Zhao S, Yu B, et al. Complete genome characterisation of a novel coronavirus associated with severe human respiratory disease in Wuhan, China. *bioRxiv*. **2020**:2020.01.24.919183. doi:[10.1101/2020.01.24.919183](https://doi.org/10.1101/2020.01.24.919183).
  - Hoffmann M, Kleine-Weber H, Krüger N, et al. The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells. *bioRxiv*. **2020**. DOI:[10.1101/2020.01.31.929042](https://doi.org/10.1101/2020.01.31.929042). 2020.01.31.929042.
  - Jiang S, ME B, Du L, et al. Roadmap to developing a recombinant coronavirus S protein receptor-binding domain vaccine for severe acute respiratory syndrome. *Expert Rev Vaccines*. **2012**;11(12):1405–1413.
  - To understand the roadmap for development of the RBD-S SARS vaccine against SARS coronavirus.**
  - Li G, Fan Y, Lai Y, et al. Coronavirus infections and immune responses. *J Med Virol*. **2020**;92(4):424–432.
  - Sambhara S, McElhaney JE. Immunosenescence and influenza vaccine efficacy. *Curr Top Microbiol*. **2009**;333:413–429.
  - Chen WH, Strych U, Hotez PJ, et al. The SARS-CoV-2 vaccine pipeline: an overview. *Curr Top Med Rep*. **2020**;1–4.
  - .. this review comprehensively discusses the current ongoing efforts to develop a vaccine against SARS-CoV-2 virus.**
  - Karch CP, Burkhard P. Vaccine technologies: from whole organisms to rationally designed protein assemblies. *Biochem Pharmacol*. **2016**;120:1–14.
  - Vivona S, Gardy JL, Ramachandran S, et al. Computer-aided biotechnology: from immuno-informatics to reverse vaccinology. *Trends Biotechnol*. **2008**;26(4):190–200.
  - Liu Z, Xiao Y, Chen YH. Epitope-vaccine Strategy against HIV-1: today and Tomorrow. *Immunobiology*. **2003**;208(4):423–428.
  - Mocellin S, Pilati P, Nitti D. Peptide-based anticancer vaccines: recent advances and future perspectives. *Curr Med Chem*. **2009**;16(36):4779–4796.
  - Lazoura E, Lodding J, Farrugia W, et al. Enhanced major histocompatibility complex class I binding and immune responses through anchor modification of the non-canonical tumour-associated mucin 1-8 peptide. *Immunology*. **2006**;119(3):306–316.
  - Pietersz GA, Pouniotis DS, Apostolopoulos V. Design of peptide-based vaccines for cancer. *Curr Med Chem*. **2006**;13(14):1591–1607.
  - Ali FR, Larche M. Peptide-based immunotherapy: a novel strategy for allergic disease. *Expert Rev Vaccines*. **2005**;4(6):881–889.
  - Structure LF. Function, and evolution of coronavirus spike proteins. *Annu Rev Virol*. **2016**;3(1):237–261.
  - Chakraborty C, Sharma AR, Sharma G, et al. SARS-CoV-2 causing pneumonia-associated respiratory disorder (COVID-19): diagnostic and proposed therapeutic options. *Eur Rev Med Pharmacol Sci*. **2020**;24(7):4016–4026.
  - Wrapp D, Wang N, Corbett KS, et al., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. **2020**;367(6483):1260–1263.
  - .. Report on a 3.5-angstrom-resolution structure of the SARS-CoV-2 trimeric spike protein, which is a key target for potential therapies and diagnostics.**
  - He Y, Li J, Li W, et al. Cross-neutralization of human and palm civet severe acute respiratory syndrome coronaviruses by antibodies targeting the receptor-binding domain of spike protein. *J Immunol*. **2006**;176(10):6085–6092.
  - He Y, Jiang S. Vaccine design for severe acute respiratory syndrome coronavirus. *Viral Immunol*. **2005**;18(2):327–332.

24. Jiang S, He Y, Liu S. SARS vaccine development. *Emerg Infect Dis*. **2005**;11(7):1016–1020.
25. Marra MA, Jones SJ, Astell CR, et al. The genome sequence of the SARS-associated coronavirus. *Science*. **2003**;300(5624):1399–1404.
26. Chang CK, Hou MH, Chang CF, et al. The SARS coronavirus nucleocapsid protein – forms and functions. *Antiviral Res*. **2014**;104(103):39–50.
27. Wei WY, Li HC, Chen CY, et al. SARS-CoV nucleocapsid protein interacts with cellular pyruvate kinase protein and inhibits its activity. *Arch Virol*. **2012**;157(4):635–645.
28. Jimenez-Guardeno JM, Nieto-Torres JL, DeDiego ML, et al. The PDZ-binding motif of severe acute respiratory syndrome coronavirus envelope protein is a determinant of viral pathogenesis. *PLoS Pathog*. **2014**;10(8):e1004320.
29. Bhattacharya M, Sharma AR, Patra P, et al. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-CoV-2): immunoinformatics approach. *J Med Virol*. **2020**;92(6):618–631.
30. Kumar S. Drug and vaccine design against Novel Coronavirus (2019-nCoV) spike protein through Computational approach. **2020**. Preprints 2020, 2020020071. DOI: 10.20944/preprints202002.0071.v1
31. Feng Y, Qiu M, Zou S, et al. Multi-epitope vaccine design using an immunoinformatics approach for 2019 novel coronavirus in China (SARS-CoV-2). *bioRxiv*. **2020**. DOI:10.1101/2020.03.03.962332
32. Enayatkhani M, Hasaniazad M, Faezi S, et al. Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. *J Biomol Struct Dyn*. **2020**:1–19. doi:10.1080/07391102.2020.1756411.
33. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*. **2006**;65(1):40–48.
34. Reynisson B, Alvarez B, Paul S, et al. NetMHCpan-4.1 and NetMHCIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res*. **2020**;48(W1):W449-W454.
35. Larsen MV, Lundsgaard C, Lamberth K, et al. Large-scale validation of methods for cytotoxic Tlymphocyte epitope prediction. *BMC Bioinformatics*. **2007**;8(1):424.
36. Flerl W, Paul S, Dhanda SK, et al. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front Immunol*. **2017**;8:278.
37. Gupta S, Kapoor P, Chaudhary K, et al. In silico approach for predicting toxicity of peptides and proteins. *PLoS One*. **2013**;8(9):e73957.
38. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. **2007**;8(1):4.
39. Tomar N, De RK. Immunoinformatics: an integrated scenario. *Immunology*. **2010**;131(2):153–168.
40. Dimitrov I, Bangov I, Flower DR, et al. AllerTOP v.2—a server for in silico prediction of allergens. *J Mol Model*. **2014**;20(6):2278.
41. Tani K, Murphy WJ, Chertov O. Defensins act as potent adjuvants that promote cellular and humoral immune responses in mice to a lymphoma idiotype and carrier antigens. *Int Immunol*. **2000**;12(5):691–700.
42. Foged C. Subunit vaccines of the future: the need for safe, customized and optimized particulate delivery systems. *Ther Deliv*. **2011**;2(8):1057–1077.
43. Pandey RK, Ojha R, Aathmanathan VS, et al. Immunoinformatics approaches to design a novel multi-epitope subunit vaccine against HIV infection. *Vaccine*. **2018**;36(17):2262–2272.
44. H H B, Sidney J, Dinh K, et al. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*. **2006**;7:153.
45. Wilkins MR, Gasteiger E, Bairoch A, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*. **1999**;112:531–552.
46. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res*. **2004**;32:W526–531.
47. Laskowski RA, MacArthur MW, Moss DS, et al. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. **1993**;26(2):283–291.
48. Eisenberg D, Luthy R, Bowie JU. Verify 3D: assessment of protein models with three dimensional profiles. *Methods Enzymol*. **1997**;277:396–404.
49. Colovos C, Yeates TO. Verification of protein structure: patterns of non-bonded atomic interaction. *Protein Sci*. **1993**;2(9):1511–1519.
50. Kozakov D, Hall DR, Xia B, et al. The ClusPro web server for protein-protein docking. *Nat Protoc*. **2017**;12(2):255–278.
51. Lindahl E, Hess B, Van Der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Mol Model Ann*. **2001**;7:306–317.
52. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graphics*. **1996**;14(1):33–38.
53. Jiang S, Hillyer C, Du L. Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses. *Trends Immunol*. **2020**;41(5):355–359.
54. Zhou Y, Yang Y, Huang J, et al. Advances in MERS-CoV vaccines and therapeutics based on the receptor-binding domain. *Viruses*. **2019**;11(1):60.
55. Du L, He Y, Zhou Y, et al. The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nat Rev Microbiol*. **2009**;7(3):226–236.
56. Jiang S, Du L, Shi Z. An emerging coronavirus causing pneumonia outbreak in Wuhan, China: calling for developing therapeutic and prophylactic strategies. *Emerg Microbes Infect*. **2020**;9(1):275–277.
57. Oppenheim JJ, Biragyn A, Kwak LW, et al. Roles of antimicrobial peptides such as defensins in innate and adaptive immunity. *Ann Rheum Dis*. **2003**;62(90002):17–21.
58. Soria-Guerra RE, Nieto-Gomez R, Govea-Alonso DO, et al. An overview of bioinformatics tools for epitope prediction: implications on vaccine development. *J Biomed Inform*. **2015**;53:405–414.
59. Sette A, Sidney J. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics*. **1999**;50(3–4):201–212.
60. Totura AL, Whitmore A, Agnihothram S, et al. Toll-like receptor 3 signaling via TRIF contributes to a protective innate immune response to severe acute respiratory syndrome coronavirus infection. *mBio*. **2015**;6(3):e00638–15.
61. Saha A, Sharma AR, Bhattacharya M, et al. Probable molecular mechanism of remdesivir for the treatment of COVID-19: need to know more. *Arch Med Res*. **2020**. DOI:10.1016/j.arcmed.2020.05.001.
62. Chakraborty C, Sharma AR, Bhattacharya M, et al. Consider TLR5 for new therapeutic development against COVID-19. *J Med Virol*. **2020**. DOI:10.1002/jmv.25997.
63. Amanat F, Krammer F. SARS-CoV-2 vaccines: status report. *Immunity*. **2020**;52(4):583–589.
- This review discusses the therapeutic and prophylactic interventions for SARS-CoV-2 and the lessons learned from the challenges faced during previous research efforts for SARS-CoV-1 vaccine development.**
64. Janice Oh HL, Ken-En Gan S, Bertoletti A, et al. Understanding the T cell immune response in SARS coronavirus infection. *Emerg Microbes Infect*. **2012**;1(9):e23.
65. Chen J, Lau YF, Lamirande EW, et al. Cellular immune responses to severe acute respiratory syndrome coronavirus (SARS-CoV) infection in senescent BALB/c mice: CD41 T cells are important in control of SARS-CoV infection. *J Virol*. **2010**;84(3):1289–1301.
66. Zhao J, Perlman S. T cell responses are required for protection from clinical disease and for virus clearance in severe acute respiratory syndrome coronavirus-infected mice. *J Virol*. **2010**;84(18):9318–9325.

67. Zhao J, Van Rooijen N, Perlman S. Evasion by stealth: inefficient immune activation underlies poor T cell response and severe disease in SARS-CoV-infected mice. *PLoS Pathog.* **2009**;5:e1000636.
68. Hsueh PR, Huang LM, Chen PJ, et al. Chronological evolution of IgM/IgA, IgG and neutralisation antibodies after infection with SARS-associated coronavirus. *Clin Microbiol Infect.* **2004**;10(12):1062–1066.
69. Buchholz UJ, Bukreyev A, Yang L, et al. Contributions of the structural proteins of severe acute respiratory syndrome coronavirus to protective immunity. *Proc Natl Acad Sci USA.* **2004**;101(26):9804–9809.
70. Wang YD, Sin WY, Xu GB, et al. T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus spike protein elicit a specific T-cell immune response in patients who recover from SARS. *J Virol.* **2004**;78(11):5612–5618.
71. Peng H, Yang LT, Wang LY, et al. Long-lived memory T lymphocyte responses against SARS coronavirus nucleocapsid protein in SARS-recovered patients. *Virology.* **2006**;351(2):466–475.
72. Leung DT, Tam FC, Ma CH, et al. Antibody response of patients with severe acute respiratory syndrome (SARS) targets the viral nucleocapsid. *J Infect Dis.* **2004**;190(2):379–386.
73. Li CK, Wu H, Yan H, et al. T cell responses to whole SARS coronavirus in humans. *J Immunol.* **2008**;181(8):5490–5500.
74. Jin H, Xiao C, Chen Z, et al. Induction of Th1 type response by DNA vaccinations with N, M, and E genes against SARS-CoV in mice. *Biochem Biophys Res Commun.* **2005**;328(4):979–986.
75. Yang L, Peng H, Zhu Z, et al. Persistent memory CD41 and CD81 T-cell responses in recovered severe acute respiratory syndrome (SARS) patients to SARS coronavirus M antigen. *J Gen Virol.* **2007**;88(10):2740–2748.