

# Excavating SARS-coronavirus 2 genome for epitope-based subunit vaccine synthesis using immunoinformatics approach

Varun Chauhan<sup>1</sup>  | Tripti Rungta<sup>1</sup> | Manmeet Rawat<sup>2</sup>  | Kapil Goyal<sup>1</sup> |  
 Yash Gupta<sup>3</sup> | Mini P. Singh<sup>1</sup>

<sup>1</sup>Department of Virology, Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh, India

<sup>2</sup>Department of Internal Medicine (GI), Health Science Centre, School of Medicine, University of New Mexico, Albuquerque, New Mexico

<sup>3</sup>Department of Internal Medicine, Loyola University Medical Center, Chicago, Illinois

## Correspondence

Varun Chauhan and Mini P. Singh, Department of Virology, Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh 160012, India.

Email: varunc1784@gmail.com (V. C.) and minipsingh@gmail.com (M. P. S.)

## Abstract

Since the outbreak of severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) in December 2019 in China, there has been an upsurge in the number of deaths and infected individuals throughout the world, thereby leading to the World Health Organization declaration of a pandemic. Since no specific therapy is currently available for the same, the present study was aimed to explore the SARS-CoV-2 genome for the identification of immunogenic regions using immunoinformatics approach. A series of computational tools were applied in a systematic way to identify the epitopes that could be utilized in vaccine development. The screened-out epitopes were passed through several immune filters, such as promiscuity, conservancy, antigenicity, nonallergenicity, population coverage, nonhomologous to human proteins, and affinity with human leukocyte antigen alleles, to screen out the best possible ones. Further, a construct comprising 11 CD4, 12 CD8, 3 B cell, and 3 interferon- $\gamma$  epitopes, along with an adjuvant  $\beta$ -defensin, was designed in silico, resulting in the formation of a multiepitope vaccine. The in silico immune simulation and population coverage analysis of the vaccine sequence showed its capacity to elicit cellular, humoral, and innate immune cells and to cover up a worldwide population of more than 97%. Further, the interaction analysis of the vaccine construct with Toll-like receptor 3 (immune receptor) was carried out by docking and dynamics simulations, revealing high affinity, constancy, and pliability between the two. The overall findings suggest that the vaccine may be highly effective, and is therefore required to be tested in the lab settings to evaluate its efficacy.

## KEY WORDS

dynamics simulations, immunoinformatics, molecular docking, promiscuous, SARS-CoV-2

## 1 | INTRODUCTION

Coronaviruses (CoVs), belonging family—Coronaviridae, subfamily—Orthocoronavirinae, and order Nidovirales, are responsible for causing enzootic infections in mammals and birds. However, since the last

decade, they have shown their capability of infecting humans as well. The lethality of coronaviruses can be demonstrated by the outbreaks caused by severe acute respiratory syndrome-coronavirus (SARS-CoV) and Middle East respiratory syndrome-coronavirus (MERS-CoV) in 2002 and 2012, respectively (Schoeman & Fielding, 2019). The novel

Varun Chauhan and Tripti Rungta are co-first authors as they contributed equally to this study.

flu-like coronavirus which caused the recent outbreak in December 2019 in China was initially named as 2019 novel-coronavirus (2019 n-CoV), and later as SARS-CoV-2 by the International Committee on Taxonomy of Viruses (Zhu et al., 2020). Since then, several cases of SARS-CoV-2 have been reported from different countries due to rapid and easy transmission through droplet route and from humans to humans and fomites (Li et al., 2020). As per a World Health Organization report published on June 23, 2020, about 8,993,659 confirmed SARS-CoV-2 cases with 469,587 deaths (Situation Report 155) have been reported worldwide so far (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>). It is a single-stranded, positive-sense RNA virus belonging to  $\beta$ -coronavirus genera, sharing varied genomic identity with SARS- and MERS-CoV (Chan et al., 2020). The genome of SARS-CoV-2 is about 30 kb (29,891 nucleotides), consisting of 16 nonstructural proteins (NSPs) consisting of two viral cysteine proteases, that is, NSP3 and NSP5 (which codes for papain-like protease and main protease, respectively), NSP12 (RNA-dependent RNA polymerase), NSP13 (helicase), and some others playing roles in replication and transcription of the virus. In addition, the genome consists of four structural proteins, that is, Envelope (E), Membrane (M), Spike (S), and Nucleocapsid (N; Chan et al., 2020). Currently, there is no established therapy available for the same in the form of vaccine or drug. Thus, it is important to find an alternative solution so that the replication and circulation of the virus can be controlled and prevented. Efforts are being made to develop the vaccines on a fast-track mode but still it is not known about the duration required to pass through various phases of clinical trials (Chen WH, Hotez, & Bottazzi, 2020). Moreover, multiple approaches of vaccine development are needed to start simultaneously as it is difficult to predict the failure of any vaccine candidate at any stage. The conventional vaccinology is a technique classically used by various scientists to develop a successful vaccine where pathogen is cultured and is used either in a killed form or in an attenuated form. However, these vaccines may be associated with various side effects due to high titer of antigen load present in vaccine formulations. Also, the chances of reversion from live attenuated strain to wild strain cannot be predicted (Heinson, Woelk, & Newell, 2015). In such a scenario, the vaccines prepared using immunoinformatics approach holds several promises over classical vaccinology approaches as it reduces the unnecessary genomic load, saves time, cost, and is comparatively less labor intensive. However, it is associated with some limitations, for example, the vaccine candidates developed in some studies have been found to be weakly immunogenic. The latter issue can be overcome by using a suitable adjuvant. The T cell lymphocytes-based immunotherapies are important in providing protective and long-lasting immunity against a number of infectious diseases and their exhaustion is often correlated with disease progression (Zhang et al., 2019). An exhaustion in CD4 and CD8 cells was reported in the peripheral blood mononuclear cells of SARS-CoV-2-infected patients (Zheng et al., 2020). Similarly, in another study, an exhaustion in CD4<sup>+</sup>, CD8<sup>+</sup>, B cells, and natural killer cells was reported in patients with SARS-CoV-2 (F. Wang et al., 2020). The following studies indicate the immunomodulation by SARS-CoV-2, causing depletion of different immune cells types in infected patients, suggesting the role of these cells in

development of protective immunity against SARS-CoV-2. Further, the interferon- $\gamma$  (IFN- $\gamma$ ) cells are well known for their immunoregulatory and antiviral properties, and thus are important for vaccine designing (Chauhan, Rungta, Goyal, & Singh, 2019). Recently, several researchers have proposed the immunotherapeutic potential of epitope-based therapeutics (consisting of T-cell, B-cell, and IFN- $\gamma$  epitopes) against the number of viral, bacterial, and parasitic infections, such as hepatitis C virus, Nipah virus, herpes simplex virus, Acinetobacter, Vibrio, malaria, Echinococcus, and Leishmania (Abbas, Zafar, Ahmad, & Azam, 2020; Chauhan & Farooq, 2016; Chauhan, Goyal, & Singh, 2018; Chauhan, Singh, & Ratho, 2018; Damfo, Reche, Gatherer, & Flower, 2017; He et al., 2019; Ravichandran, Venkatesan, & Febin Prabhu Dass, 2018; Solanki & Tiwari, 2018). Even the recently proposed RTS,S/AS01 vaccine for malaria consists of T-cell epitopes (Pance, 2019). Such epitope-based vaccines are composed of peptides that could elicit the activation of different subset of T cells and B cells specific to target protein, and thus have immense potential in vaccine designing. Thus, considering the above-mentioned points and the nonavailability of any established therapy for SARS-CoV-2, the present study was proposed to identify the immunogenic markers in its genome in the form of T cell-, B cell-, and IFN- $\gamma$ -stimulating epitopes. The results of the present study can be validated by wet laboratory experiments and can be tested in clinical trials at a faster pace so as to curtail the threat of SARS-CoV-2.

## 2 | MATERIALS AND METHODS

### 2.1 | SARS-CoV-2 genome and structural analysis

The amino acid sequences of SARS-CoV-2 genome-associated proteins, that is, ORF1ab (which is a complex of the following proteins: 5'-untranslated region, nsp-2, nsp-3, nsp-4, 3C-like protease, nsp-6, nsp-7, nsp-8, nsp-9, nsp-10, RNA-dependent RNA polymerase [nsp-12], helicase [nsp13], 3'-5' exonuclease [nsp-14], endoRNase [nsp-15], and o-ribose methyltransferase [nsp-16]), surface glycoprotein "S," ORF-3a, Envelope protein "E," Membrane glycoprotein "M," ORF-6, ORF-7, ORF-8, Nucleocapsid phosphoprotein "N" and ORF-10, were retrieved from the National Center for Biotechnology Information database. The aim of carrying out phylogenetic analysis was to determine the relatedness of SARS-CoV-2 proteins with the proteins of SARS and MERS coronaviruses. The antigenicity, allergenicity, and other physicochemical properties of the proteins were determined by VaxiJen (Doytchinova & Flower, 2007; <http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>), AlgPred (Saha & Raghava, 2006; <http://webs.iiitd.edu.in/raghava/algpred/submission.html>), and Protparam servers (<https://web.expasy.org/protparam/>), respectively. The proteins were also checked for having any homology at the sequence level with the human proteome using Blastp analysis (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>). The secondary and tertiary structural analysis of the SARS-CoV-2 proteins were determined by SOPMA (Geourjon & Deleage, 1995; [http://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_sopma.html](http://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)) and RaptorX tools (<http://raptord.uchicago.edu/StructPredV2/predict/>; Kallberg et al., 2012), respectively.

## 2.2 | Identification of epitopes

NetCTL1.2 server (<http://www.cbs.dtu.dk/services/NetCTL/>) was utilized for the identification of cytotoxic T-cell (CTL) epitope prediction (Larsen et al., 2007). The commonly found human leukocyte antigen (HLA) Class I alleles in human population worldwide (more than 90%), were targeted for epitope prediction (Chauhan & Singh, 2020). The server predicts the peptide based on the following three parameters: (a) proteasomal mediated cleavage at C terminal, (b) major histocompatibility complex (MHC) Class I binding, and (c) efficiency of transporter associated with antigen processing. All the parameters were taken into consideration during prediction without altering any parameter, at a prediction threshold of 0.75.

NetMHCIIPan 3.2 (<http://www.cbs.dtu.dk/services/NetMHCIIPan/>; Jensen et al., 2018) and the Immune Epitope Database (IEDB) consensus methods (<http://tools.iedb.org/mhcii/>; P. Wang et al., 2010) were employed for Helper T-cell (HTL) epitope prediction. Both the servers are recently updated versions over previously used CD4 epitope prediction servers with improved accuracy, trained on extended datasets. The alleles targeted for epitopes prediction were supposed to cover >95% of the worldwide population (Chauhan & Singh, 2020). The epitopes predicted were divided into three different categories based on the percentile ranks. Based on percentile ranks of 2%, 2–10%, and >10%, the epitopes were designated as strong, intermediate, and weak binders, respectively. The prediction of IFN- $\gamma$  cells was carried out using the IFNepitope server (Dhanda, Vir, & Raghava, 2013) by scanning module using motif and support vector machine (SVM) hybrid approach and the model for prediction was IFN- $\gamma$  versus Non-IFN- $\gamma$ . The BCPred 2.0 server (<http://crdd.osdd.net/raghava/bcpred/>; El-Manzalawy, Dobbs, & Honavar, 2008) and ElliPro server (<http://tools.iedb.org/ellipro/>; Ponomarenko et al., 2008) were utilized for linear/continuous B-cell epitopes and conformational/discontinuous epitopes prediction, respectively. The parameters, such as antigenicity, hydrophilicity, surface accessibility,  $\beta$ -turn, and flexibility of the predicted linear B-cell epitopes, were also taken into consideration during prediction. The predicted B-cell epitopes from different proteins were mapped on their respective three-dimensional (3D) models, to determine their surface localization. Both BCPred 2.0 and ElliPro servers were run at default without altering any prediction parameter.

## 2.3 | Filtering out the predicted epitopes based on following immune filters

### 2.3.1 | Promiscuous epitope prediction

The promiscuous epitopes are important in vaccine designing as they have affinity toward multiple HLA alleles (Chauhan et al., 2019). Moreover, such epitopes have a large population coverage due to their promiscuous nature. Thus, the screened-out epitopes (with high binding affinity scores for HLAs) were further subjected to promiscuity analysis.

### 2.3.2 | Overlapping epitope prediction

The overlapping epitopes have an integral sequence containing both HTL and CTL epitopes, and thus can activate both HTLs and CTLs (Chauhan et al., 2019). Thus, the HTL epitopes which were found to be overlapping with CTL epitopes, were also screened out for further analysis.

### 2.3.3 | Antigenicity, allergenicity, and population coverage analysis of the epitopes

The VaxiJen v2.0 tool and AlgPred tools were used for antigenicity and allergenicity prediction of the epitopes. The amino acid composition-based SVM module was utilized for allergenicity prediction of epitopes, the threshold for which is -0.4. At this value, the sensitivity of the predictive value was reported to be 88.87% and specificity was 81.86%, respectively, at fivefold cross-validation. Among the allergenicity prediction methods of AlgPred, this method has the highest sensitivity and was thus selected for the analysis. The population size covered by the epitopes was analyzed by IEDB population coverage analysis. The Blastp search engine was used against the human proteome to identify any similarity of the predicted epitopes with any human proteins. If found any, was eliminated for further analysis.

### 2.3.4 | Posttranslational modification sites

The epitopes predicted were also checked for their location within any posttranslational modification sites. For this, the amino acid sequences of each protein were subjected to NetOGlyc 4.0 server (<http://www.cbs.dtu.dk/services/NetOGlyc/>) for glycosylation sites prediction (Steentoft et al., 2013). The epitopes which were observed to be within posttranslational modification sites were excluded.

### 2.3.5 | Interaction analysis of peptides and human leukocyte antigens

#### *Molecular docking*

ClusPro (<http://cluspro.bu.edu/>), a protein–protein docking server by Schrodinger, was utilized for determining the interaction patterns of the CD4 and CD8 epitopes with Class II and I HLA alleles, respectively (Kozakov et al., 2017).

## 2.4 | Vaccine and the immune receptor construction and analyzing their different properties

### 2.4.1 | Multiepitope chain construction

The best possible epitopes were screened out based on the above-mentioned immune filters and merged to form a single peptide chain. The T- and B-cell epitopes were linked by GPGPG linkers, HEYGAELERA motifs, and GGGG linkers.  $\beta$ -Defensin was used as an adjuvant and linked

to the vaccine chain, via EAAAK linker. The physicochemical properties and immunogenicity of the vaccine were confirmed using ProtParam and VaxiJen tools, respectively. The allergenicity of the vaccine sequence was determined by two different servers, AlgPred and AllerTOP v2.0 tools (<http://www.ddg-pharmfac.net/AllerTOP/contact.html>).

#### 2.4.2 | Modeling and validation of vaccine and Toll-like receptor 3

The secondary structural confirmations, such as  $\alpha$ -helix,  $\beta$ -sheets, and  $\beta$ -loops were determined using the SOPMA tool. The multiepitope vaccine sequence was shuffled randomly into six different sequences and each sequence was modeled using homology modeling tools, such as RaptorX (<http://raptorg.uchicago.edu/StructPredV2/predict/>; Kallberg et al., 2012), Phyre2 (<http://www.sbg.bio.ic.ac.uk/~phyre2/>; Kelley, Mezulis, Yates, Wass, & Sternberg, 2015), Swiss Model workspace (<http://swissmodel.expasy.org/interactive>; Waterhouse et al., 2018), I-Tasser (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>; Yang et al., 2015), and Modweb (<http://modbase.compbio.ucsf.edu/modweb/>; Pieper et al., 2006). Among the models generated by different homology modeling tools, the one with the best structural configuration was selected for further analysis. The model was further refined using GalaxyRefine server and the reliability of the finalized model was analyzed by Ramachandran Plot ProSA web analysis (<http://prosa.services.came.sbg.ac.at/prosa.php>). Finally, the disulfide engineering of the vaccine was performed to enhance its thermostability, using Design 2 server

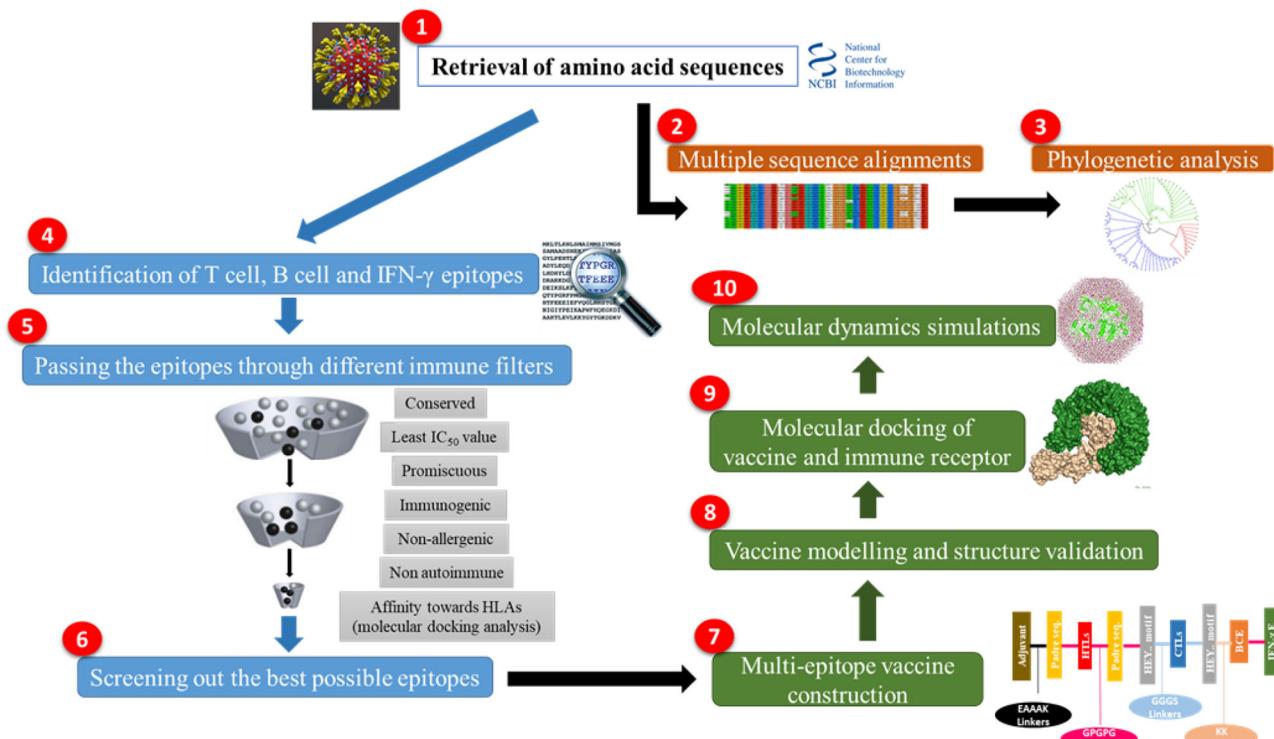
(<http://cptweb.cpt.wayne.edu/DbD2/>; Craig & Dombkowski, 2013). The tertiary structure of Toll-like receptor 3 (TLR-3) was retrieved from the Protein Data Bank with ID: 2AOZ.

#### 2.4.3 | Immune simulation by vaccine sequence

An in silico immune simulation was carried out using C-ImmSim 10.1 server (<http://www.cbs.dtu.dk/services/C-ImmSim-10.1/>), which is based on position-specific scoring matrix, to analyze the activation of different immune markers by the vaccine construct against SARS-CoV-2 (Rapin, Lund, Bernaschi, & Castiglione, 2010). The server is based on stimulating three major compartments, that is, lymph node, thymus, and bone marrow, found in mammals. The simulation was run to analyze the activation of different cellular, humoral, and innate immune components like B cells, T-helper cells (CD4), T-cytotoxic cells (CD8), immunoglobulins, cytokines, dendritic cells, macrophages, and epithelial cells. The random seed, simulation steps, and time set of injections were set as described previously (Chauhan & Singh, 2020).

### 2.5 | Vaccine model and immune receptor interaction analysis

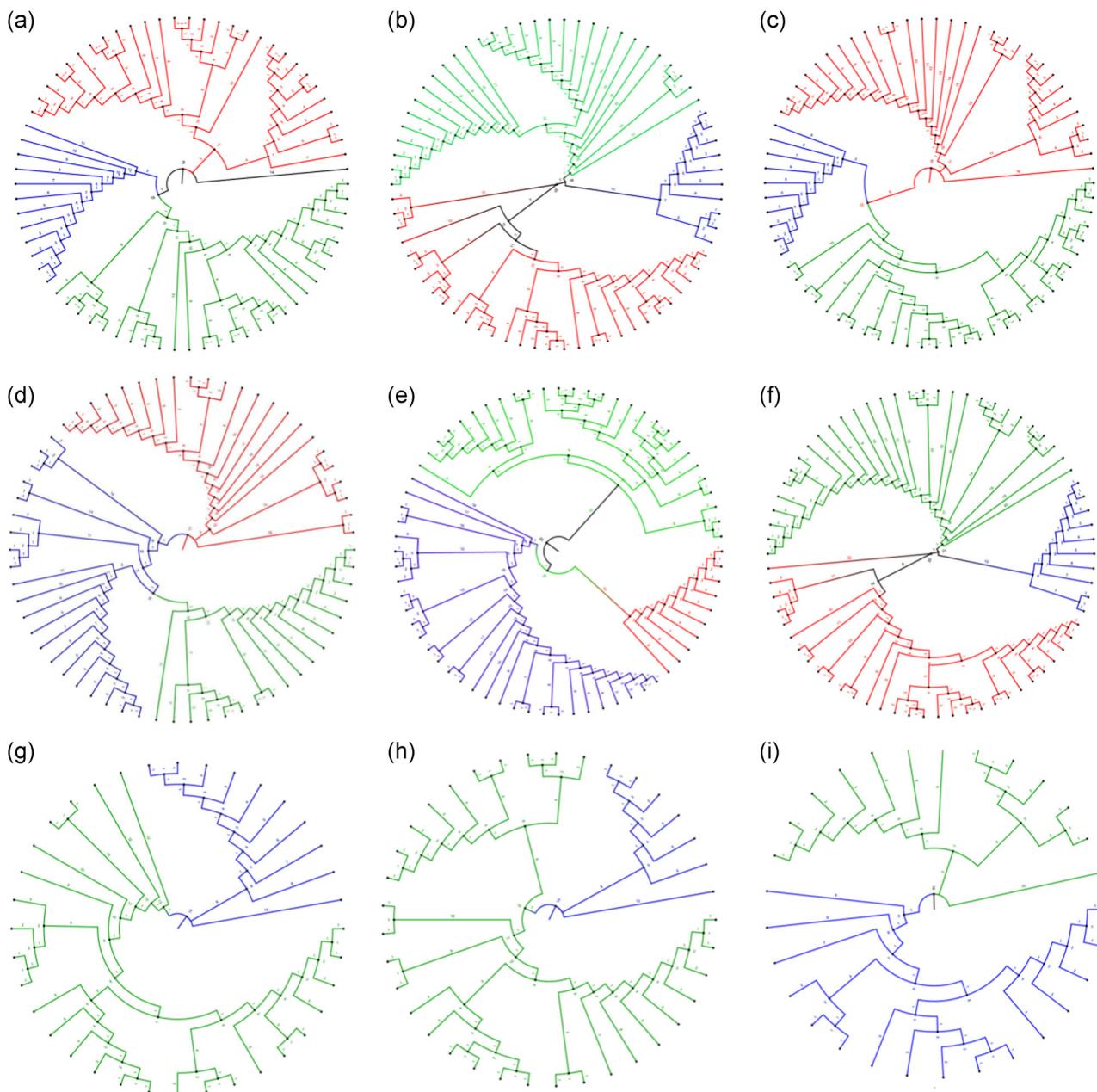
The molecular interaction pattern between the candidate vaccine and immune receptor was determined using ClusPro, a protein-protein docking tool. Among the 29 docked models generated by



**FIGURE 1** Schematic representation of the methodology employed for epitope identification and multiepitope vaccine construction against severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2)

ClusPro, the one with best docked configuration was selected for further analysis. Further, the Desmond tool (Schrodinger) was employed to perform the molecular dynamics (MD) simulations for analyzing the interaction pattern of the docked complex at the microscopic level. The TLR-3 and individual coordinates were optimized in the protein preparation wizard of Schrodinger's Maestro suite (2019-3), where hydrogens were added; hetero molecules were removed, and the complex structure was minimized using the OPLS2005 force field. The complete docked protein complex system was solvated in a cuboid box with TIP3P water molecules and 0.15 mM NaCl (physiological conditions), with a minimum 10<sup>Å</sup>

buffering distance minimized volume in all three orthogonal dimensions using System Builder Wizard within the Desmond module of Schrodinger (2019-3). The protein charges were neutralized by excess "8" chloride ions. The system was then relaxed before actual simulation. The MD simulation steps involved the built setup system heated from 50 to 300K, followed by maintaining the system's temperature of 300K for 20 ns. Once the desired temperature of the system was achieved in all subsequent simulation steps, a pressure of 1.02 bar at isobaric (NPT) ensemble was maintained. The trajectory concatenation and interaction site residue visualization were performed using Maestro interface of Schrodinger (2019-3).



**FIGURE 2** Phylogenetic trees showing genetic relatedness of SARS-CoV-2 proteins with SARS and Middle East respiratory syndrome-coronavirus (MERS-CoVs). The blue, red, and green branches belong to SARS-CoV-2, MERS, and SARS proteins. The phylogenetic trees are represented in the order: (a) membrane, (b) Nucleocapsid, (c) surface, (d) Envelope, (e) ORF1ab, (f) ORF3, (g) ORF6, (h) ORF7, (i) ORF8. SARS-CoV-2, severe acute respiratory syndrome-coronavirus 2

In addition, iMODs server (<http://imods.chaconlab.org/>) was also used to determine the stability of the docked complex (Lopez-Blanco, Aliaga, Quintana-Orti, & Chacon, 2014). The iMODs server utilizes the normal mode analysis for determining the collective motions in internal coordinates. The simulation analysis was expressed in terms of covariance map, elastic network, eigenvalues, B-factor/mobility, and variance.

The protocol employed for screening out the epitopes and multiepitope construction is represented in Figure 1.

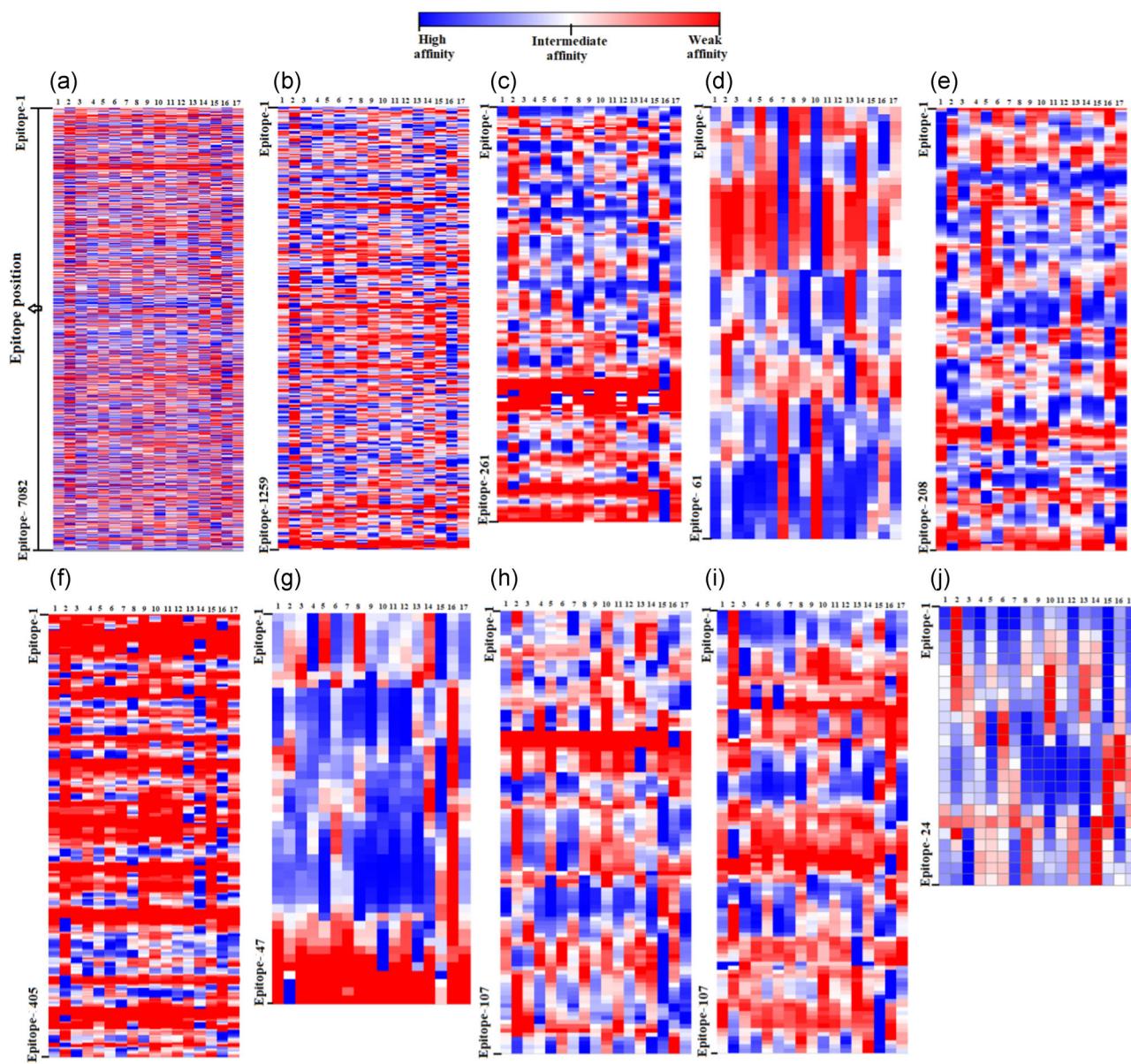
## 2.6 | Cloning of polyepitope vaccine by in silico approach

The amino acid sequence of the vaccine was reverse-transcribed and its various properties, such as codon adaptation index (CAI) and guanine-cytosine (GC) content for efficient cloning was analyzed using the Java Codon Adaptation Tool (JCAT; <http://www.jcat.de/Start.jsp>). The *Escherichia coli* K12 strain was used for expressing the protein of interest by optimizing its codon. As per the tool recommendation, the ideal CAI and

**TABLE 1** SARS-CoV-2 proteins: antigenicity, allergenicity, and secondary structural properties

Proteins	Antigenicity score	Allergenicity score	Molecular weight (Dalton [Da])	Secondary structural properties
ORF1ab	0.46	0.14	794,057.79	α Helix – 38.75% Extended strand – 22.14% β Turn – 9.94% Random coil – 29.17%
Surface glycoprotein	0.46	0.54	141,178.47	α Helix – 28.59% Extended strand – 23.25% β Turn – 3.38% Random coil – 44.78%
ORF3a	0.49	-0.89	31,122.94	α Helix – 26.18% Extended strand – 29.82% β Turn – 10.18% Random coil – 33.82%
Envelope	0.60	-0.87	8,365.04	α Helix – 44.0% Extended strand – 26.67% β Turn – 9.33% Random coil – 20.0%
Membrane glycoprotein	0.51	-1.62	25,146.62	α Helix – 34.68% Extended strand – 21.17% β Turn – 6.76% Random coil – 37.39%
ORF6	0.61	-0.17	7,272.54	α Helix – 70.49% Extended strand – 9.84% β Turn – 8.20% Random coil – 11.48%
ORF7	0.64	-0.45	13,744.17	α Helix – 42.98% Extended strand – 19.01% β Turn – 9.92% Random coil – 28.10%
ORF8	0.65	-0.183	13,831.01	α Helix – 19.83% Extended strand – 35.54% β Turn – 4.96% Random coil – 36.67%
Nucleocapsid	0.50	-0.95	45,625.70	α Helix – 21.24% Extended strand – 16.71% β Turn – 6.92% Random coil – 55.13%
ORF10	0.54	0.0014	4,449.23	α Helix – 28.95% Extended strand – 36.84% β Turn – 5.26% Random coil – 28.95%

Abbreviation: SARS-CoV-2, severe acute respiratory syndrome-coronavirus 2.



**FIGURE 3** Epitopes showing varying affinities (strong, intermediate, and weak) to different human leukocyte antigen (HLA) Class II alleles, represented in the form of heat map. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, and 17 are HLA alleles HLADRB1\* 01:01, HLA-DRB1\*03:01, HLA-DRB1\*04:01, HLA-DRB1\*07:01, HLA-DRB1\*08:01, HLA-DRB1\*09:01, HLA-DRB1\*10:01, HLA-DRB1\*11:01, HLA-DRB1\*12:01, HLA-DRB1\*13:02, HLA-DRB1\*14:01, HLADRB1\* 15:01, HLA-DRB3\*02:02, HLA-DRB5\*01:01, HLA-DPA1\*02:01-DPB1\*01:01, HLA-DQA1\*01:02- DQB1\*06:02, and HLA-DPA1\*02:01-DPB1\*11:401. a–j are n-CoV proteins in the following order: (a) ORF1ab, (b) surface glycoprotein, (c) ORF 3a, (d) Envelope, (e) membrane glycoprotein, (f) N, (g) ORF-6a, (h) ORF- 7a, (i) ORF-8a, and (j) ORF-10a

GC content should range between 0.8 and 1.0 and 30% and 70%, respectively, for efficient cloning. Finally, the optimized sequence was cloned in pET28a(+) expression vector, using SnapGene, an in silico cloning tool.

### 3 | RESULTS AND DISCUSSION

#### 3.1 | Genomic and structural analysis

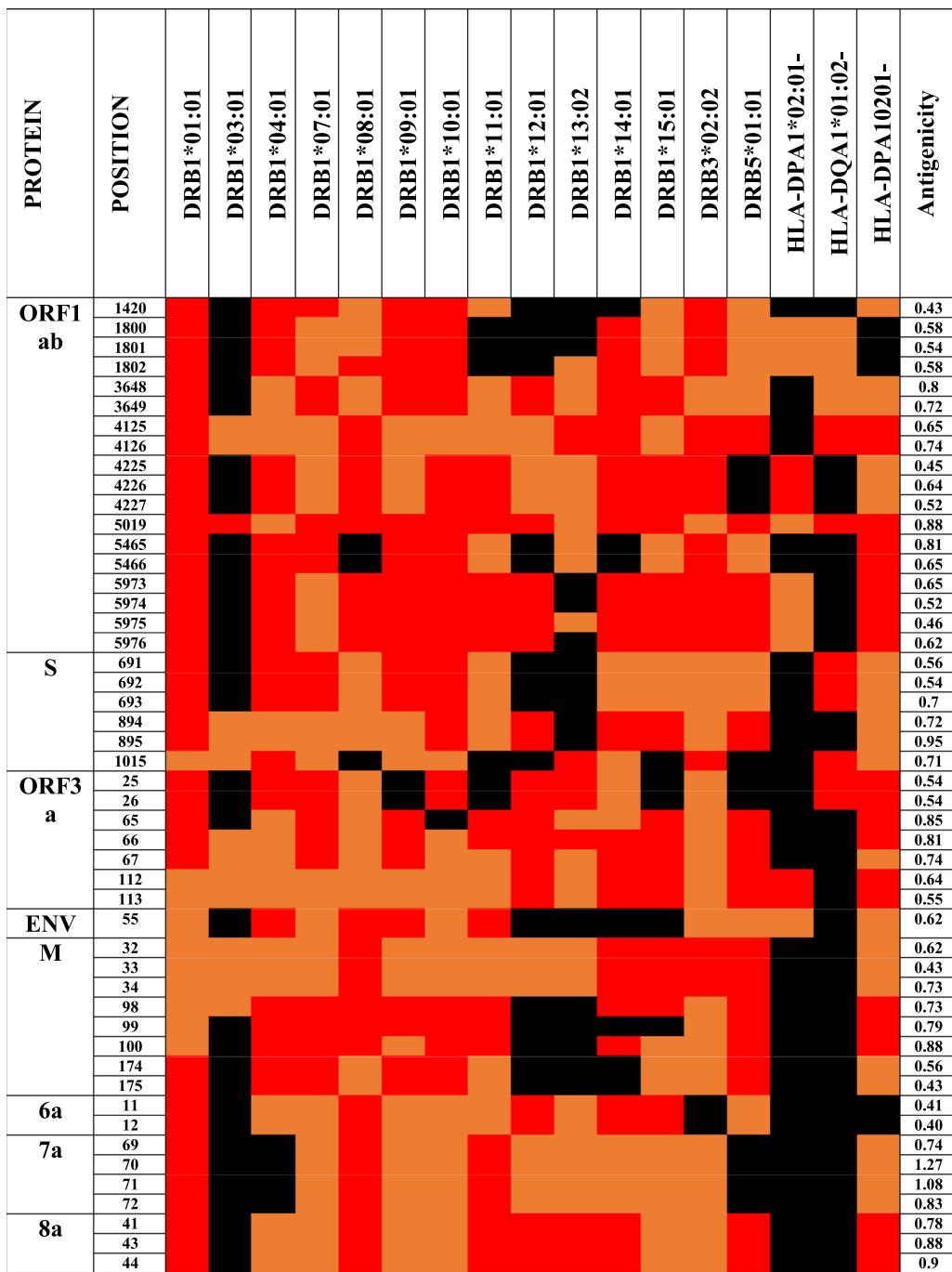
The Blastn analysis revealed that the genome of SARS-CoV-2 had around 88% similarity with SARS-CoV and only 12–15% similarity

with the MERS-CoV genome. The individual proteins of SARS-CoV-2 were also subjected to Blast analysis for analyzing their similarity with other CoV strains. The ORF1ab polyprotein of SARS-CoV-2 showed the highest similarity of about 98.5% with ORF1ab of SARS-CoV and around 50.8% similarity with that of MERS-CoV. Similarly, the surface glycoprotein “S” showed 97.4% similarity to the S protein of SARS-CoV and around 36% similarity with the S protein of MERS-CoV. ORF3a showed around 92% similarity to SARS, but did not find any similarity with that of MERS-CoV. Envelope “E” protein showed 95% similarity to that of SARS and about 38% to that of MERS-CoV. Membrane glycoproteins showed 99% similarity to that of SARS and 50% to

MERS-CoV. The nucleocapsid phosphoprotein showed around 96% similarity with SARS-CoV and around 53% similarity with MERS-CoV. ORF 6 and ORF 7 of SARS-CoV-2 had 93%, 97%, and 95% similarity respectively, with the ORF-6, ORF-7, and ORF-8 proteins of SARS-CoV and did not show any similarity with that of MERS-CoV. ORF-10 did not show any similarity with SARS- and MERS-CoVs. Further, the sequences were subjected to phylogenetic analysis. The analysis was carried out at 1,000 bootstraps replication using the maximum likelihood method (Kumar, Stecher & Tamura, 2017; Figure 2).

The phylogenetic analysis of SARS-CoV-2 proteins was carried out to investigate the relatedness of the individual proteins of SARS-CoV-2 with other CoV strains.

The proteins were also checked for having any homology at the sequence level with the human proteome using Blastp analysis; none of the SARS-CoV-2 proteins showed any homology with that of human proteins. The secondary structural configurations and other physico-chemical properties of the proteins are shown in Table 1. The tertiary structures of the proteins were also generated to explore and map the



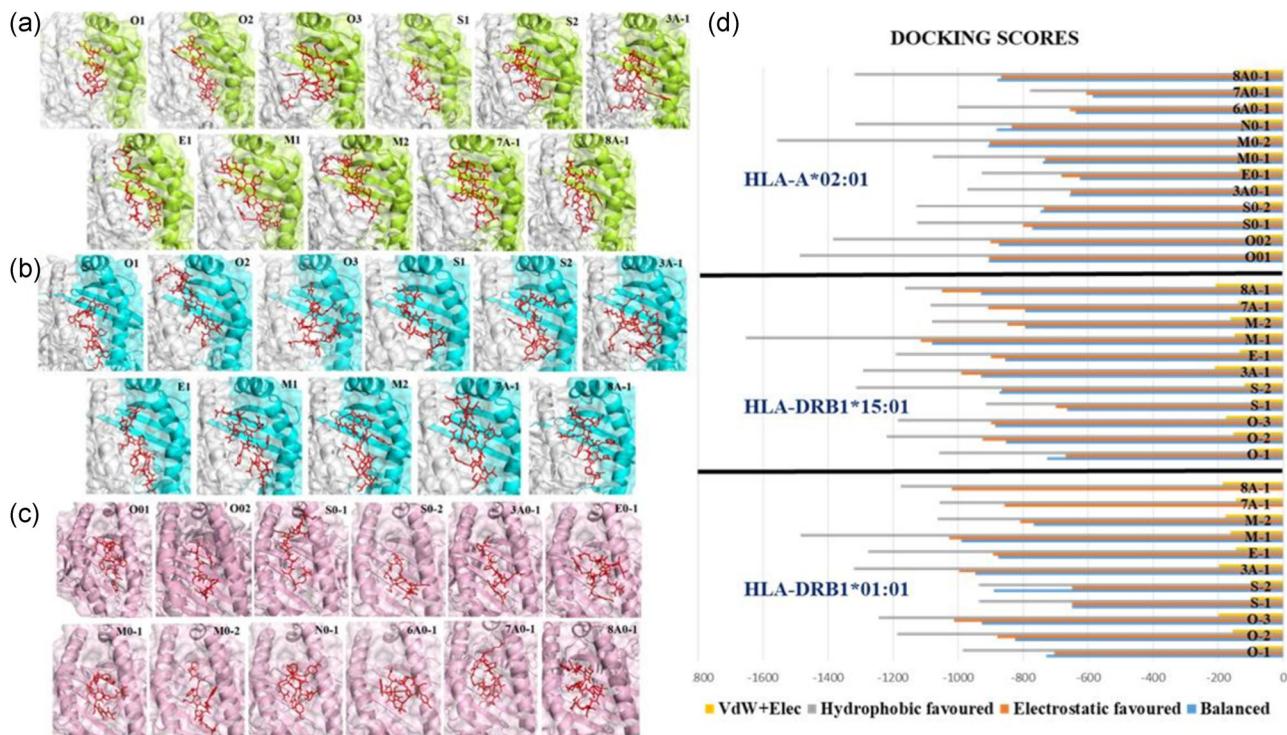
**FIGURE 4** Screened-out HLA Class-II epitopes. The epitopes represented are highly promiscuous, conserved, and antigenic. Red, orange, and black colors represent strong, intermediate, and weak binding affinities with HLAs. HLA, human leukocyte antigen

location of the screened-out T- and B-cell epitopes. The details of the template used for modeling the 3D models of the proteins and their Ramachandran plot analysis are represented in Table S1.

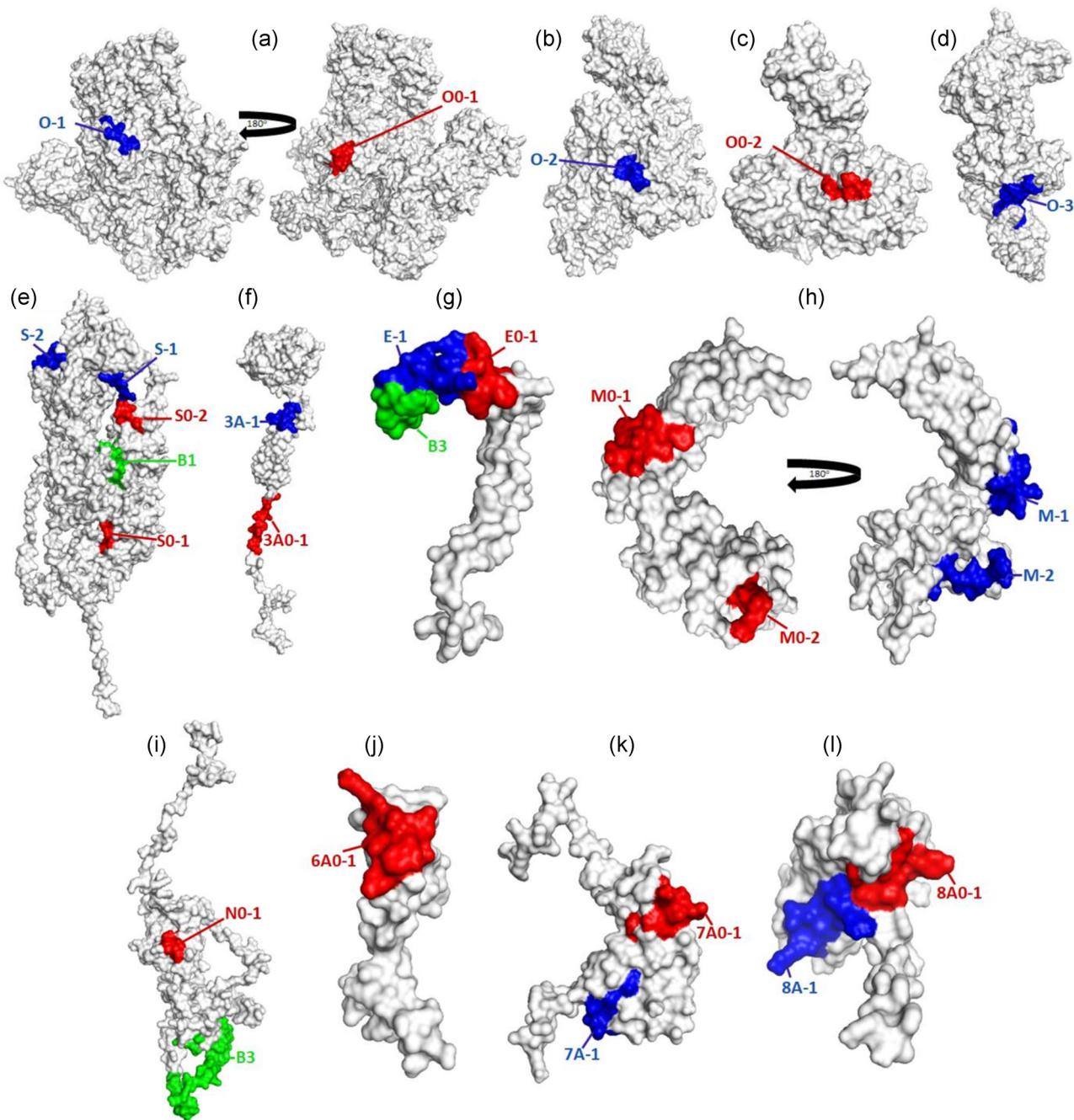
### 3.2 | T-cell, IFN- $\gamma$ , and B-cell epitope recognition

CTLs are specialized immune cells capable of recognizing virus-infected cells and release cytotoxic factors, such as perforin and granzymes in response, thereby averting the survival of invading virus. CTLs further release certain cytokines, IFN- $\gamma$ , and tumor necrosis factor- $\alpha$ , thus enhancing the antiviral response. The T-helper cells (CD4) help in recognizing foreign antigens and secrete certain cytokines that further aid in activating T and B cells through different signaling cascades. In addition, T-helper cells aid B cells to differentiate into memory B cells (Rosendahl Huber, van Beek, de Jonge, Luytjes, & van Baarle, 2014). Therefore, the identification of epitopes capable of stimulating CD4 and CD8 cells is important in vaccine designing. Thus, the CD8 T-cell epitopes were identified by NETCTL1.2 server, at a threshold of 0.75 without altering any parameters (Table S2) and the CD4 T cells were identified by IEDB consensus and NetMHCIIpan 3.2 servers, respectively (Table S3). Based on thresholds of 2%, 10%, and  $>10\%$ , the CD4 epitopes were considered as strong, intermediate, and unbinding epitopes (Figure 3). The worldwide population coverage of  $>90\%$  was likely to be covered by the HLA Class I alleles, targeted for epitope prediction. The epitopes predicted (both CD4 and CD8 T-cell

epitopes) were progressed through numerous immune filters to screen out the most promising ones. For example, they should be antigenic, nonallergenic, promiscuous, should have 100% conservancy, should not be located within posttranslational modification sites, should have high affinity with HLA alleles, and should target larger population (Figure 4). Further, the screened-out T-cell epitopes were docked with HLA Class I and II alleles that are commonly present in human population, to analyze their binding patterns and affinities. The affinities of the docked complexes were analyzed in the form of Van der Waals, hydrophobic, and electrostatic interactions. The epitopes with better binding affinities with HLA alleles (as revealed by docking scores and the number of H-bonds formed) were finally selected for further analysis (Figure 5). The similar kind of strategy has been followed earlier by Chauhan et al. (2019), to screen out epitopes with better affinities. IFN- $\gamma$  cells are components of innate immune response and are well known for its antiviral properties. Thus, the IFN- $\gamma$  cells were predicted using IFNepitope server (Table S4). Only those IFN- $\gamma$  epitopes were screened out which were conserved, antigenic, and nonallergenic. The B cells on activation differentiate into plasma and memory cells and are thus important for providing long-lasting immunity (Chauhan & Singh, 2020). Thus, the linear and conformational B-cell epitopes were predicted by BCPred 2.0 and ElliPro servers (Figure S1), respectively. The mapping of B-cell epitopes on their respective proteins 3D models was also carried out to affirm their surface location (Figure 6). The antigenicity and conservancy of the predicted B-cell and IFN- $\gamma$  epitopes were also determined (Table S5).



**FIGURE 5** Results of docking analysis. (a–c) Docked complex of finally selected epitopes with HLA Class II alleles: HLA-DRB1\*01:01 (Protein Data Bank ID- 2g9hr) and HLA-DRB1\*15:01 (1BX2) and with HLA Class I allele: HLA-A\*02:01 (1QEW), respectively. (d) Docking scores of different epitopes. HLA, human leukocyte antigen

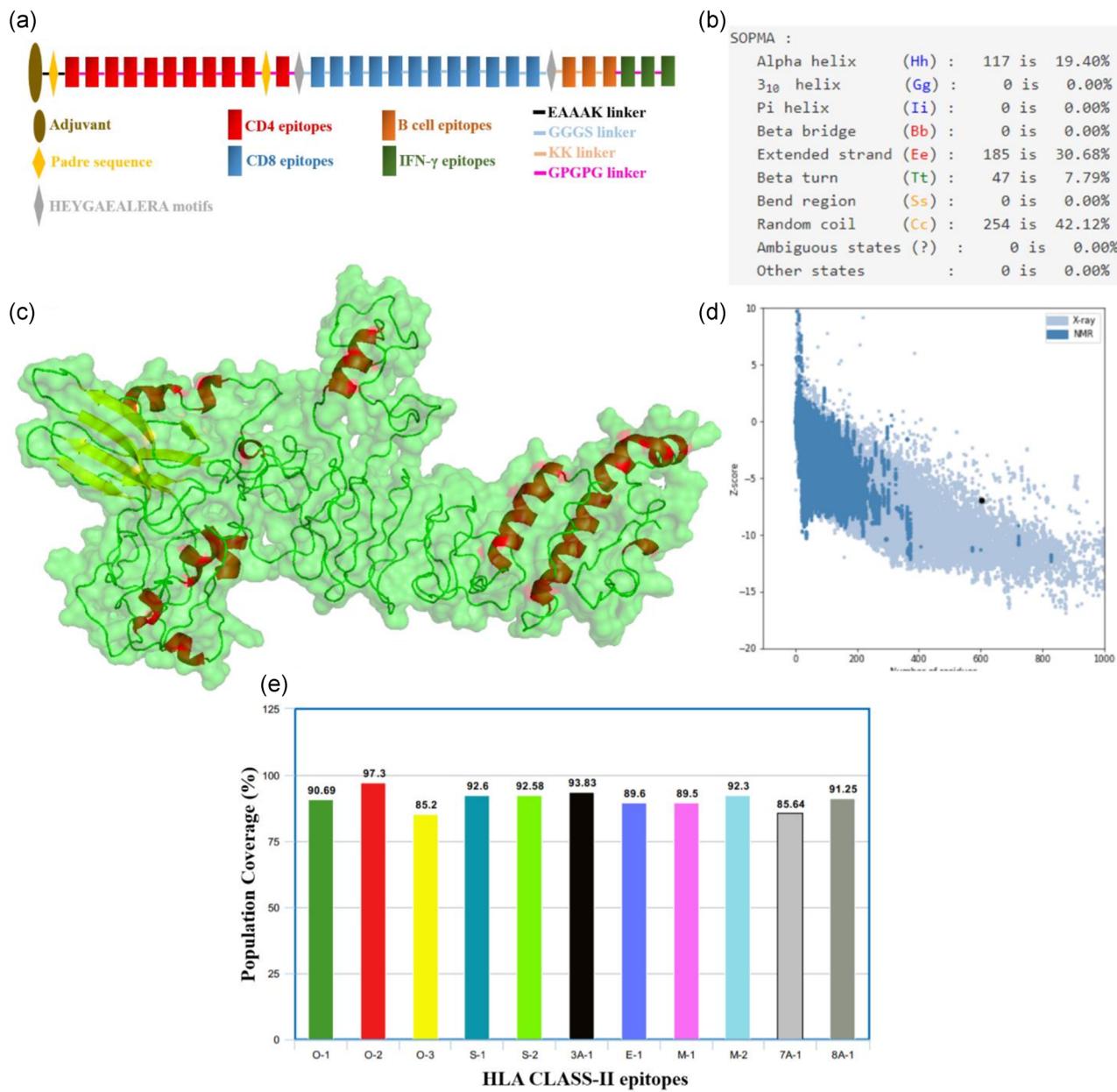


**FIGURE 6** Mapping of epitopes on the three-dimensional (3D) modeled proteins (represented in white surface view). The red, blue, and green colors represent HLA Class I, II, and B cell epitopes. a-l are proteins in order ORF1ab (a) nsp3, (b) RNA-dependent RNA polymerase, (c) helicase, (d) guanine-N7-methyltransferase, (e) surface glycoprotein, (f) ORF3a, (g) envelope protein, (h) membrane glycoprotein, (i) nucleocapsid, (j) ORF6a, (k) ORF7a, and (l) ORF8a. HLA, human leukocyte antigen

### 3.3 | Multiepitope vaccine construction and analysis of its different properties

The multiepitope vaccine is a vaccine containing a chain of several epitopes (CD8 and/or CD4 and/or B cell and/or IFN- $\gamma$  epitopes). The vaccines designed using such strategies holds several advantages over classical methods employed for vaccine designing. For example, autoimmune generation by such vaccines are low, could cover large

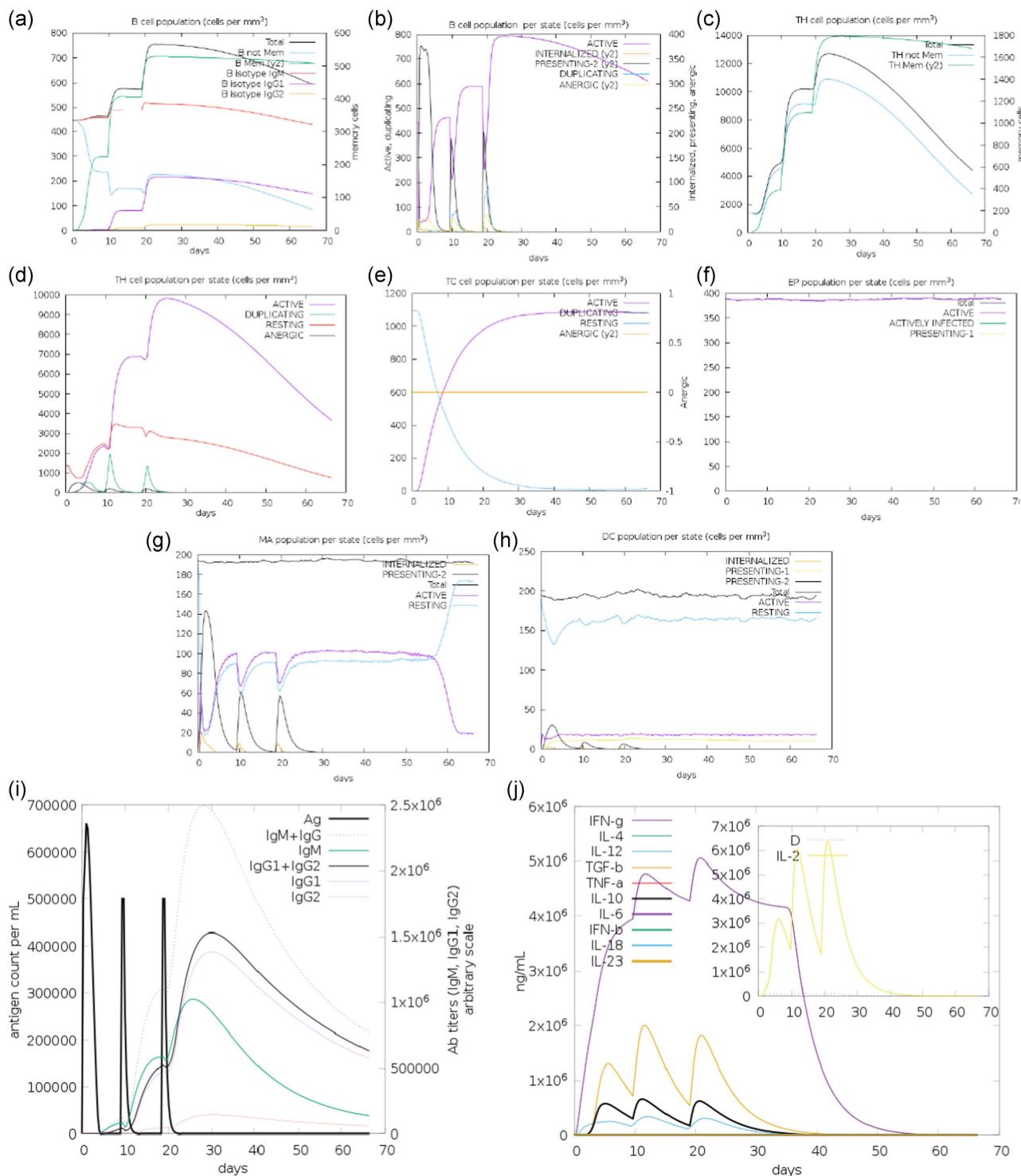
population, could stimulate both cellular and humoral immune responses due to presence of both T-cell and B-cell epitopes, could be effective even if the pathogen is prone to have mutations since it consists of several conserved epitopes belonging to different proteins, and it saves time and cost over classical methods (Chauhan & Singh, 2020). Thus, we aimed to design a vaccine containing T-cell, B-cell, and IFN- $\gamma$  cell epitopes. The epitopes (11 CD4, 12 CD8, and 3 IFN- $\gamma$  and B-cell epitopes each) finally selected for inclusion in the



**FIGURE 7** Vaccine construct. (a) Overall layout of the vaccine construct. (b) Secondary structural configuration. (c) Representation of vaccine as 3D model (in surface and cartoon view). (d) The ProSa web results of the constructed model. (e) Population coverage of each HLA-Class II epitope included in the vaccine. 3D, three-dimensional; HLA, human leukocyte antigen

multipitope vaccine were strictly as per the criteria designed, such as promiscuity, antigenicity, nonallergenicity, affinity and docking scores, conserved, nonhomologous to human proteins, and large population coverage (Table S6). Most of the T- and B-cell epitopes selected had overlapping conformational B-cell epitopes. The finally selected epitopes were attached via GPGPG, GGGS, and AAY linkers, respectively. In addition, a 12 residue peptide—HEYGAEALERA motifs were also used as linkers in between CD8 T-cell epitopes, which aids in enhancing the epitope presentation by providing the specific sites for both proteasomal and lysosomal mediated cleavages, that is, A5-E6, Y3-G4, A7-L8, L8-E9, and R10-A11 (Nezafat, Ghasemi, Javadi, Khoshnoud, & Omidinia, 2014). Further, the PADRE sequence,

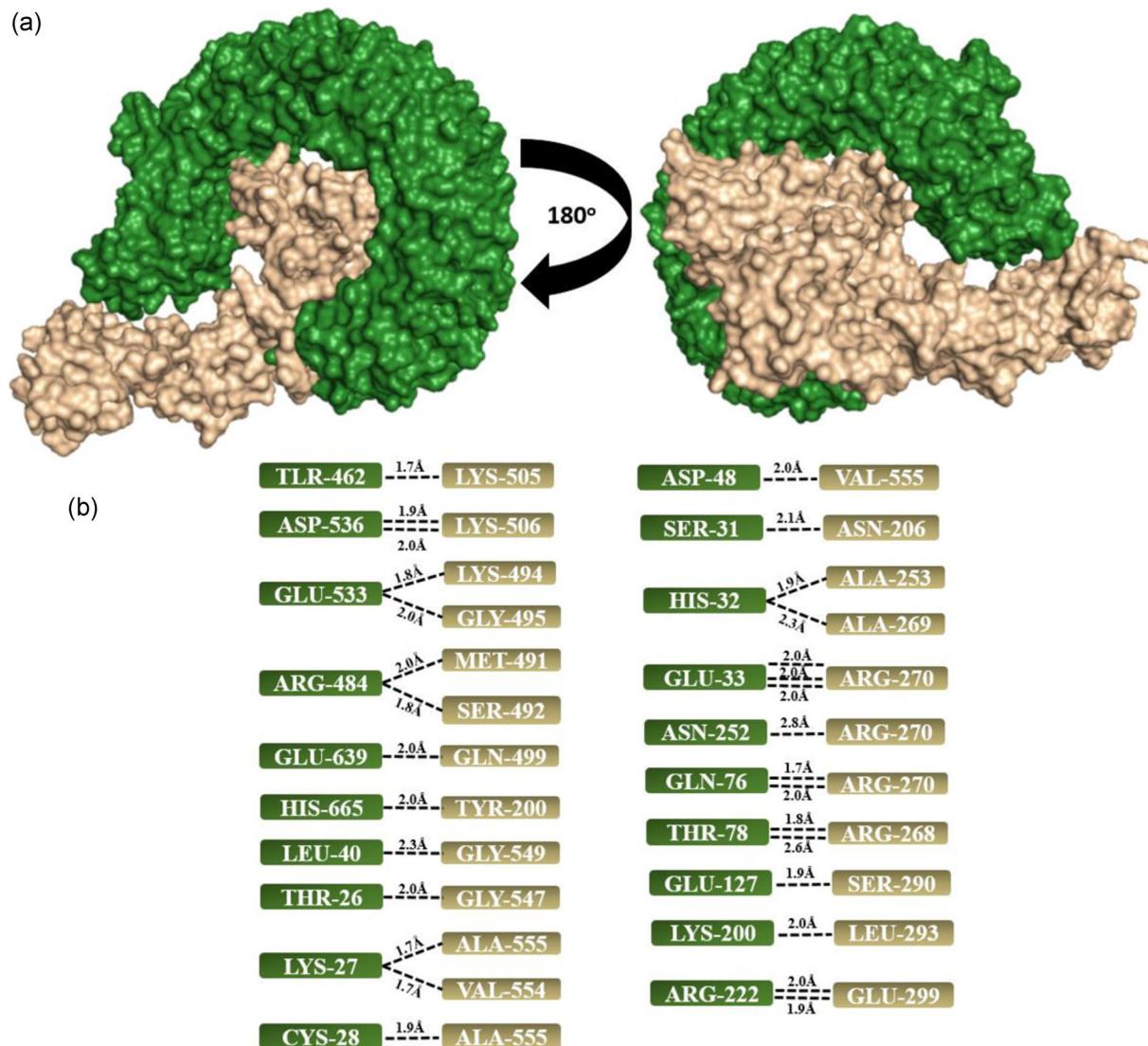
composed of 13 amino acid residues—AKFVAAWTLKAAA—was also considered in the vaccine sequence. The PADRE sequence has affinity toward number of human and mouse HLA Class II alleles and induce the T-helper responses (Athanasou et al., 2017). An adjuvant  $\beta$ -defensin was linked via EAAAK linker to the vaccine sequence to enhance its immunity. The adjuvant on interacting with certain immune receptors (TLRs and CCR6) is well known for activating innate and adaptive immune responses, by recruiting immature dendritic cells and T cells at the site of infection (Ojha, Nandani, & Prajapati, 2019). Finally, the vaccine sequence was constructed containing  $\beta$ -defensin adjuvant, 11 CD4, 12 CD8, 3 IFN- $\gamma$ , and 3 B-cell epitopes and 2 PADRE and HEYGAEALERA motifs (Figure 7a).



**FIGURE 8** The immune simulation results of the vaccine sequence. The B cell simulation results (a,b); the T-helper cells simulation (c,d); the T-cytotoxic cells simulation (e), epithelial cell activation (f); macrophages levels (g); dendritic cells levels (h); immunoglobulins levels with respect to antigen concentration (i); and cytokine levels (j)

The vaccine sequence was composed of 603 amino acids, had molecular weight of 63,898.95 Da, isoelectric point 10.27, and the instability index II was 39.13, classifying it as a stable protein. The grand average of hydropathicity of the vaccine sequence was predicted to be -0.224. All these parameters were determined using ExPASy

ProtParam tool. The vaccine sequence was arranged into six different sequences by random reshuffling of epitopes and were subjected to different homology modeling tools as stated in Section 2.4.2, to achieve the best configuration. The model generated by Raptor-X server for the first sequence (Supporting Information Material S1)

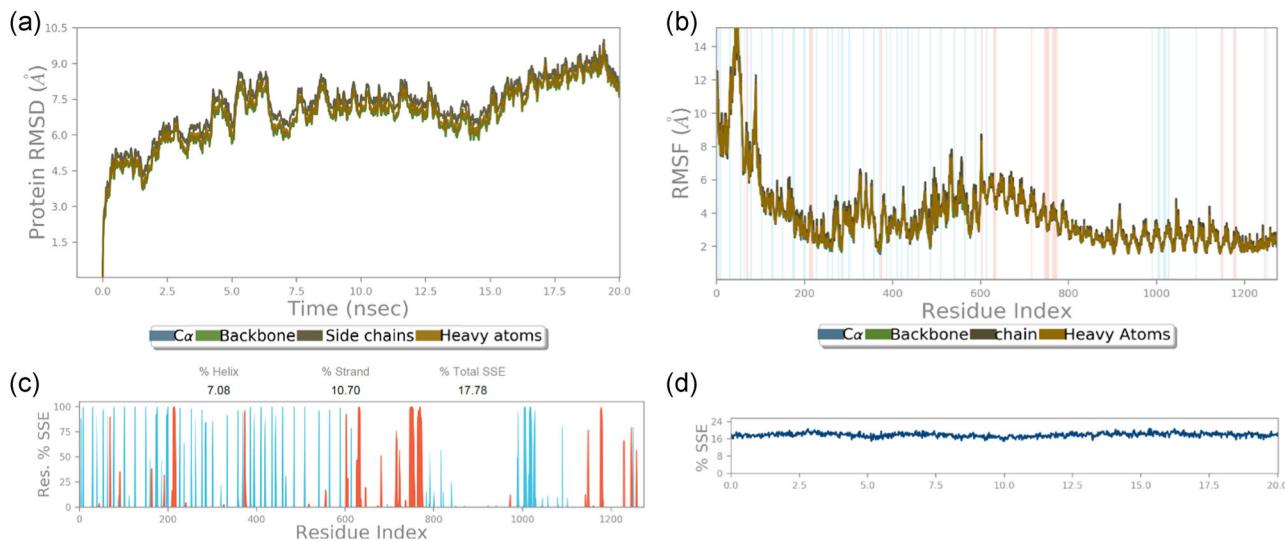


**FIGURE 9** Docking analysis results. (a) The interaction pattern of the vaccine (golden color) and Toll-like receptor 3 (TLR-3; dark green) in surface view. (b) The interacting residues of the vaccine and TLR-3

revealed better structural configuration as compared to others (Figure 7b). The model was finalized based on ProSA web and Ramachandran Plot results. The amino acid sequence of the finalized model was observed to be immunogenic and nonallergenic with scores 0.57 (by VaxiJen server) and -1.406 (by AlgPred server) respectively. In addition, AllerTOP v2.0 server also predicted the sequence to be nonallergenic. Further, the model generated was refined via GalaxyRefine server. It is considered among the best servers for improving the global and local structural configuration of a protein. The ProSA web and Ramachandran plot results revealed an improved structural properties of the finalized 3D model of the multiepitope vaccine (Figure S2). The Ramachandran plot analysis revealed that 98.2% region was in favored and allowed region with only 1.8% lying in outliers (Figure 7c,d). The population coverage analysis of HLA-Class II epitopes included in the vaccine showed an overall high coverage of 97.3% (Figure 7e). Further, the modeled vaccine was subjected to

disulfide engineering to further increase its thermo stability utilizing Design 2 server, which reduces the conformational entropy of the unfolded state of the protein (Figure S2).

The amino acid sequence of the finalized vaccine model was further subjected to in silico immune simulations, revealing that it is competent in generating an effective immune response. Elevated levels of immunoglobulin M in beginning followed by increased levels of immunoglobulin G and its subclasses with reduced antigen concentrations were observed. The ability of the vaccine in generating an effective and long lasting immunity was indicated by elevated levels of activated and memory B cells. Further, the elevated levels of T cells (both Th and Tc), macrophages and dendritic cells, and lowered levels of regulatory T cells were observed during antigen exposure. An effective antiviral response by the vaccine was also indicated by increased levels of IFN- $\gamma$  and interleukin-2 (Figure 8).



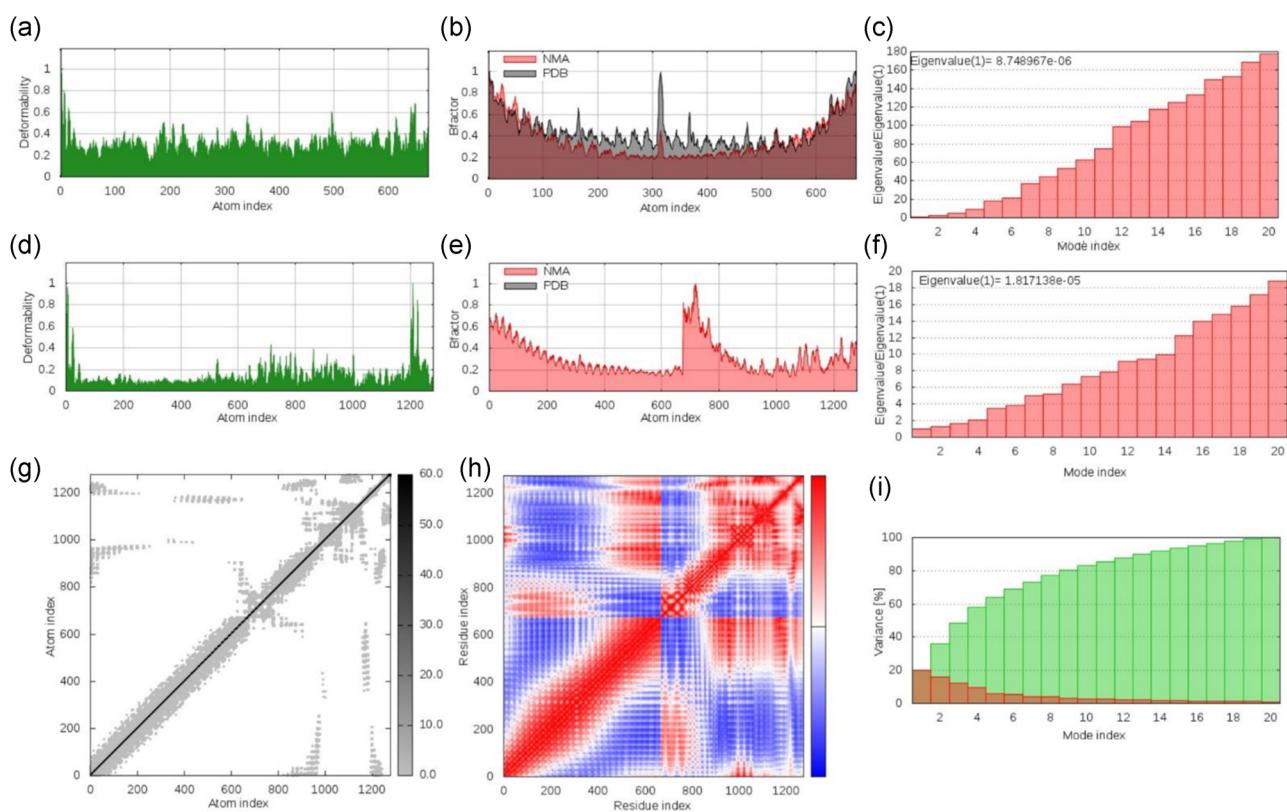
**FIGURE 10** The molecular dynamics simulation results generated by Desmond, carried out at 20 ns. (a) Root-mean-square deviation, (b) root-mean-square fluctuation, and (c,d) secondary structural elements of the vaccine and TLR-3 docked complex, observed during simulations. TLR-3, Toll-like receptor 3

The polyepitope-based vaccines leveraging such mechanisms of activating different immune cell types have recently been put forward for consideration by several researchers, against the number of infective diseases, such as Kaposi sarcoma (Chauhan et al., 2019), Epstein–Barr virus (Ojha et al., 2019), Zika virus (Kumar Pandey, Ojha, Mishra, & Kumar Prajapati, 2018), cytomegalovirus (Chauhan & Singh, 2020), human T-lymphotropic virus-1 (Pandey et al., 2019), Ebola virus (Bazhan et al., 2019), and so forth.

### 3.4 | Interaction analysis

Initially, all the epitopes that were included in vaccine sequence were docked with the HLA alleles for determining the affinity of each epitope (as discussed in Section 3 and Figure 4). A vaccine or drug could be effective if there is prolonged and stable interaction with their receptor. Thus, molecular docking analysis was performed using ClusPro, to analyze the interaction pattern of the finalized modeled vaccine with the immune receptor, TLR-3. Among the 29 docked models generated by ClusPro, the 11th model showed better docking results, with a score of -1260.3, forming 29 H-bonds between the two (Figure 9). Further, the docked complex was subjected to MD simulations to analyze the physical movements. Overall, from docking calculations and MD simulations, we can say that the predicted binding pose is a strong complex forming binding pose. The side chains of the vaccine were deeply buried into the peptide binding extreme loops of TLR, and not with lipophilic pocket at the trunk that binds to lipopolysaccharide, and with polar interactions with the polar residues. Throughout the MD simulation (20 ns), it is possible to observe that the receptor is stable and the binding antigen remains attached without any significant change. During simulation, the root-mean-square fluctuation (RMSF) was used for determining the

flexibility of vaccine, bound to the immune receptor TLR-3, during each confirmatory shift (Figure 10a). The RMSF of each residue within the docked complex showed that there were two very stable interacting regions in both proteins. The observed initial changes in RMSF show total system relaxation in MD simulation environment and not intermolecular interaction changes (Figure 10b). Over the course of simulation, the complex had a significant decay in global energy, suggesting positive complexation without loss in binding integrity of the complex. The stable confirmation of the complex throughout simulation was further indicated by the secondary structural elements, indicating the prolonged binding of vaccine with TLR-3, which is essential for the generation of an effective immune response (Figure 10c,d). Further, the docked complex was also subjected to iMODS server, for determining its stability. The deformability analysis was carried out to determine the flexibility of each residue of the docked complex and was compared to that of TLR-3 in monomeric state (Figure 11a,d). The results revealed that there was significant reduction in the distortions in the docked complex, due to stabilization, as indicated by lower hinges, as compared to that of TLR-3 in monomeric state. The B factor analysis further affirmed the stability of the complex by showing lower atomic distortions as compared to TLR-3 in monomer state (Figure 11b,e). The eigenvalue of docked complex was observed greater than monomeric TLR-3, further indicating the stabilization of the complex formed, as higher eigenvalue is indicative of higher energy required to destabilize the protein (Figure 11c,f). Further, the covariance matrix analysis was carried out to analyze the coupling of docked complex residue pairs, where blue, white, and red colors represent the anticorrelated, uncorrelated, and correlated motions, respectively (Figure 11g). Further, the elastic network analysis and variance analysis were also performed to determine the stiffness and cumulative and individual variances of the complex (Figure 11g,i). The results obtained by MD



**FIGURE 11** Simulation results by iMODs. TLR-3 monomer: deformability, B-factor, and eigenvalue (a–c). Docked complex (vaccine and TLR-3): deformability, B-factor, and eigenvalue (d–f). Elastic network (g); covariance map (h); and variance analysis (i). TLR-3, Toll-like receptor 3

simulations to determine the stability of the protein complex were in accordance to our previously published articles (Chauhan & Singh, 2020; Chauhan et al., 2019).

### 3.5 | In silico cloning

An in silico cloning was performed to analyze the cloning and expression efficiency of the reverse-transcribed sequence of polyepitope vaccine in vector. The sequence optimized using JCAT was composed of 1,809 nucleotides and was cloned in *E. coli* (strain K12). The GC content of vaccine sequence was observed to be 56.75 and the CAI was 1.0, indicating the efficient cloning properties of the vaccine sequence. Finally, the restriction cloning of the vaccine sequence in an expression vector- pET28a (+) was carried out using SnapGene tool (Figure S3). Similar kind of strategy of in silico cloning analysis of the epitope-based vaccine sequence was performed by Chauhan et al. (2019), and the results obtained are in accordance with the study.

## 4 | CONCLUSION

The pandemic situation imposed by SARS-CoV-2 has attracted worldwide attention to develop therapeutics for its prevention and

control on urgent basis. Though some studies have shown the efficacy of using drugs, such as chloroquine, ritonavir, and angiotensin converting enzyme inhibitors against SARS-CoV-2, there is a need of effective vaccine to prevent spread of infection. In the present study, the rigorous immunoinformatics analysis was performed in a careful and sequential manner to screen out the most promising epitopes. The finally screened-out epitopes strictly followed the criteria designed for filtering out the most promising ones like promiscuity, conservancy, nonallergenicity, antigenicity, high population coverage, and affinity with HLA alleles. Further, an attempt was made to design a multiepitope vaccine by assembling the finally screened-out epitopes. The finalized construct was composed of 11 CD4, 12 CD8, 3 B-cell, and 3 IFN- $\gamma$  epitopes. As the vaccine is composed of innate, humoral, and cellular immune markers, generation of an effective immune response by the vaccine could be achieved, as indicated by the results of in silico immune simulations. The docking results revealed the high affinity of the vaccine construct with the immune receptor, TLR-3. Further, the MD simulations revealed the stability of the docked complex. The results generated in the present study will undeniably aid researchers in identifying the immunogenic regions in the n-CoV genome, which could be utilized in vaccine development.

## ACKNOWLEDGMENT

The authors are thankful to PGIMER, Chandigarh, India, for providing the opportunity and workstation for carrying out the following work.

## CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

## AUTHOR CONTRIBUTIONS

V. C. and T. R. carried out the immunoinformatics analysis. V. C. and M. P. S. designed the protocol. M. R. and K. G. assisted in writing the manuscript. Y. G. carried out MD simulations.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the Supporting Information Material of this paper.

## ORCID

Varun Chauhan  <http://orcid.org/0000-0001-6769-8691>

Manmeet Rawat  <http://orcid.org/0000-0002-8325-3390>

## REFERENCES

- Abbas, G., Zafar, I., Ahmad, S., & Azam, S. S. (2020). Immunoinformatics design of a novel multi-epitope peptide vaccine to combat multi-drug resistant infections caused by *Vibrio vulnificus*. *European Journal of Pharmaceutical Sciences*, 142, 105160. <https://doi.org/10.1016/j.ejps.2019.105160>
- Athanasiou, E., Agallou, M., Tatsoglou, S., Kammona, O., Hatzigeorgiou, A., Kiparisides, C., & Karagouni, E. (2017). A poly (lactic-co-glycolic) acid nanovaccine based on chimeric peptides from different Leishmania infantum proteins induces dendritic cells maturation and promotes peptide-specific IFN $\gamma$ -producing CD8(+) T cells essential for the protection against experimental visceral Leishmaniasis. *Frontiers in Immunology*, 8, 684. <https://doi.org/10.3389/fimmu.2017.00684>
- Bazhan, S. I., Antonets, D. V., Karpenko, L. I., Oreshkova, S. F., Kaplina, O. N., Starostina, E. V., ... Ilyichev, A. A. (2019). In silico designed Ebola virus T-cell multi-epitope DNA vaccine constructions are immunogenic in mice. *Vaccines*, 7(2), <https://doi.org/10.3390/vaccines7020034>
- Chan, J. F., Kok, K. H., Zhu, Z., Chu, H., To, K. K., Yuan, S., & Yuen, K. Y. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections*, 9(1), 221–236. <https://doi.org/10.1080/22221751.2020.1719902>
- Chauhan, V., Goyal, K., & Singh, M. P. (2018). Identification of broadly reactive epitopes targeting major glycoproteins of Herpes simplex virus (HSV) 1 and 2—an immunoinformatics analysis. *Infection, Genetics and Evolution*, 61, 24–35. <https://doi.org/10.1016/j.meegid.2018.03.004>
- Chauhan, V., Rungta, T., Goyal, K., & Singh, M. P. (2019). Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach. *Scientific Reports*, 9(1), 2517. <https://doi.org/10.1038/s41598-019-39299-8>
- Chauhan, V., & Singh, M. P. (2020). Immuno-informatics approach to design a multi-epitope vaccine to combat cytomegalovirus infection. *European Journal of Pharmaceutical Sciences*, 147, 105279. <https://doi.org/10.1016/j.ejps.2020.105279>
- Chauhan, V., Singh, M. P., & Ratho, R. K. (2018). Identification of T cell and B cell epitopes against Indian HCV-genotype-3a for vaccine development- An in silico analysis. *Biologicals*, 53, 63–71. <https://doi.org/10.1016/j.biologicals.2018.02.003>
- Chauhan, V., & Farooq, U. (2016). Identification of T cell and B cell epitopes derived from EG95 antigen of *Echinococcus granulosus* using in silico approach for therapeutic vaccine development. *Indo American Journal of Pharmaceutical Research*, 6, 4639–4645.
- Chen WH, S. U., Hotez, P. J., & Bottazzi, M. E. (2020). The SARS-CoV-2 vaccine pipeline: An overview. *Current Tropical Medicine Reports*, 3, 1–4.
- Craig, D. B., & Dombkowski, A. A. (2013). Disulfide by Design 2.0: A web-based tool for disulfide engineering in proteins. *BMC Bioinformatics*, 14(1), 346.
- Damfo, S. A., Reche, P., Gatherer, D., & Flower, D. R. (2017). In silico design of knowledge-based plasmodium falciparum epitope ensemble vaccines. *Journal of Molecular Graphics & Modelling*, 78, 195–205. <https://doi.org/10.1016/j.jmgm.2017.10.004>
- Dhanda, S. K., Vir, P., & Raghava, G. P. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct*, 8, 30. <https://doi.org/10.1186/1745-6150-8-30>
- Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, 8, 4. <https://doi.org/10.1186/1471-2105-8-4>
- El-Manzalawy, Y., Dobbs, D., & Honavar, V. (2008). Predicting linear B-cell epitopes using string kernels. *Journal of Molecular Recognition*, 21(4), 243–255. <https://doi.org/10.1002/jmr.893>
- Geourjon, C., & Deleage, G. (1995). SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Computer Applications in the Biosciences: CABIOS*, 11(6), 681–684. <https://doi.org/10.1093/bioinformatics/11.6.681>
- He, J., Huang, F., Zhang, J., Chen, H., Chen, Q., Zhang, J., ... Chen, J. (2019). DNA prime-protein boost vaccine encoding HLA-A2, HLA-A24 and HLA-DR1 restricted epitopes of CaNA2 against visceral leishmaniasis. *Immunology*, 156(1), 94–108. <https://doi.org/10.1111/imm.13007>
- Heinson, A. I., Woelk, C. H., & Newell, M. L. (2015). The promise of reverse vaccinology. *International Health*, 7(2), 85–89. <https://doi.org/10.1093/inthealth/ihv002>
- Jensen, K. K., Andreatta, M., Marcatili, P., Buus, S., Greenbaum, J. A., Yan, Z., ... Nielsen, M. (2018). Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology*, 154(3), 394–406. <https://doi.org/10.1111/imm.12889>
- Kallberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nature Protocols*, 7(8), 1511–1522. <https://doi.org/10.1038/nprot.2012.085>
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), 845–858. <https://doi.org/10.1038/nprot.2015.053>
- Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., ... Vajda, S. (2017). The ClusPro web server for protein-protein docking. *Nature Protocols*, 12(2), 255–278. <https://doi.org/10.1038/nprot.2016.169>
- Kumar, S., Stecher, G., & Tamura, K. (2017). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874.
- Kumar Pandey, R., Ojha, R., Mishra, A., & Kumar Prajapati, V. (2018). Designing B- and T-cell multi-epitope based subunit vaccine using immunoinformatics approach to control Zika virus infection. *Journal of Cellular Biochemistry*, 119(9), 7631–7642. <https://doi.org/10.1002/jcb.27110>
- Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*, 8, 424. <https://doi.org/10.1186/1471-2105-8-424>
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., ... Feng, Z. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *The New England Journal of Medicine*, 36, 280–287. <https://doi.org/10.1056/NEJMoa2001316>
- Lopez-Blanco, J. R., Aliaga, J. I., Quintana-Orti, E. S., & Chacon, P. (2014). iMODS: Internal coordinates normal mode analysis server. *Nucleic Acids Research*, 42(Web Server issue), W271–W276. <https://doi.org/10.1093/nar/gku339>

- Nezafat, N., Ghasemi, Y., Javadi, G., Khoshnoud, M. J., & Omidinia, E. (2014). A novel multi-epitope peptide vaccine against cancer: An in silico approach. *Journal of Theoretical Biology*, 21, 349–134.
- Ojha, R., Nandani, R., & Prajapati, V. K. (2019). Contriving multiepitope subunit vaccine by exploiting structural and nonstructural viral proteins to prevent Epstein-Barr virus-associated malignancy. *Journal of Cellular Physiology*, 234(5), 6437–6448. <https://doi.org/10.1002/jcp.27380>
- Pance, A. (2019). How elusive can a malaria vaccine be? *Nature Reviews Microbiology*, 17(3), 129. <https://doi.org/10.1038/s41579-018-0148-3>
- 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. *Journal of Cellular Physiology*, 234(6), 8717–8726. <https://doi.org/10.1002/jcp.27531>
- Pieper, U., Eswar, N., Davis, F. P., Braberg, H., Madhusudhan, M. S., Rossi, A., ... Sali, A. (2006). MODBASE: A database of annotated comparative protein structure models and associated resources. *Nucleic Acids Research*, 34(Database issue), D291–D295. <https://doi.org/10.1093/nar/gkj059>
- Ponomarenko, J., Bui, H. H., Li, W., Fusseder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, 9, 514. <https://doi.org/10.1186/1471-2105-9-514>
- Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One*, 5(4), e9862. <https://doi.org/10.1371/journal.pone.0009862>
- Ravichandran, L., Venkatesan, A., & Febin Prabhu Dass, J. (2018). Epitope-based immunoinformatics approach on RNA-dependent RNA polymerase (RdRp) protein complex of Nipah virus (NiV). *Journal of Cellular Biochemistry*, <https://doi.org/10.1002/jcb.27979>
- Rosendahl Huber, S., van Beek, J., de Jonge, J., Luytjes, W., & van Baarle, D. (2014). T cell responses to viral infections - opportunities for Peptide vaccination. *Frontiers in Immunology*, 5, 171. <https://doi.org/10.3389/fimmu.2014.00171>
- Saha, S., & Raghava, G. P. (2006). AlgPred: Prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Research*, 34(Web Server issue), W202–W209. <https://doi.org/10.1093/nar/gkl343>
- Schoeman, D., & Fielding, B. C. (2019). Coronavirus envelope protein: Current knowledge. *Virology Journal*, 16(1), 69. <https://doi.org/10.1186/s12985-019-1182-0>
- Solanki, V., & Tiwari, V. (2018). Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against *Acinetobacter baumannii*. *Scientific Reports*, 8(1), 9044. <https://doi.org/10.1038/s41598-018-26689-7>
- Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T., ... Clausen, H. (2013). Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *The EMBO Journal*, 32(10), 1478–1488. <https://doi.org/10.1038/embj.2013.79>
- Wang, F., Nie, J., Wang, H., Zhao, Q., Xiong, Y., Deng, L., ... Zhang, Y. (2020). Characteristics of peripheral lymphocyte subset alteration in COVID-19 pneumonia. *The Journal of Infectious Diseases*, 221(11), 1762–1769. <https://doi.org/10.1093/infdis/jiaa150>
- Wang, P., Sidney, J., Kim, Y., Sette, A., Lund, O., Nielsen, M., & Peters, B. (2010). Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics*, 11, 568. <https://doi.org/10.1186/1471-2105-11-568>
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gummienny, R., ... Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296–w303. <https://doi.org/10.1093/nar/gky427>
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The iTASSER Suite: Protein structure and function prediction. *Nature Methods*, 12(1), 7–8. <https://doi.org/10.1038/nmeth.3213>
- Zhang, C., Wang, X. M., Li, S. R., Twelkmeyer, T., Wang, W. H., Zhang, S. Y., ... Chen, X. W. (2019). NKG2A is a NK cell exhaustion checkpoint for HCV persistence. *Nature Communications*, 10(1), 1–11. <https://doi.org/10.1038/s41467-019-09212-y>
- Zheng, H. Y., Zhang, M., Yang, C. X., Zhang, N., Wang, X. C., Yang, X. P., ... Zheng, Y. T. (2020). Elevated exhaustion levels and reduced functional diversity of T cells in peripheral blood may predict severe progression in COVID-19 patients. *Cellular & Molecular Immunology*, 17(5), 541–543. <https://doi.org/10.1038/s41423-020-0401-3>
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., ... Tan, W. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *The New England Journal of Medicine*, 382(8), 727–733. <https://doi.org/10.1056/NEJMoa2001017>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Chauhan V, Rungta T, Rawat M, Goyal K, Gupta Y, Singh MP. Excavating SARS-coronavirus 2 genome for epitope-based subunit vaccine synthesis using immunoinformatics approach. *J Cell Physiol*. 2021;236: 1131–1147. <https://doi.org/10.1002/jcp.29923>