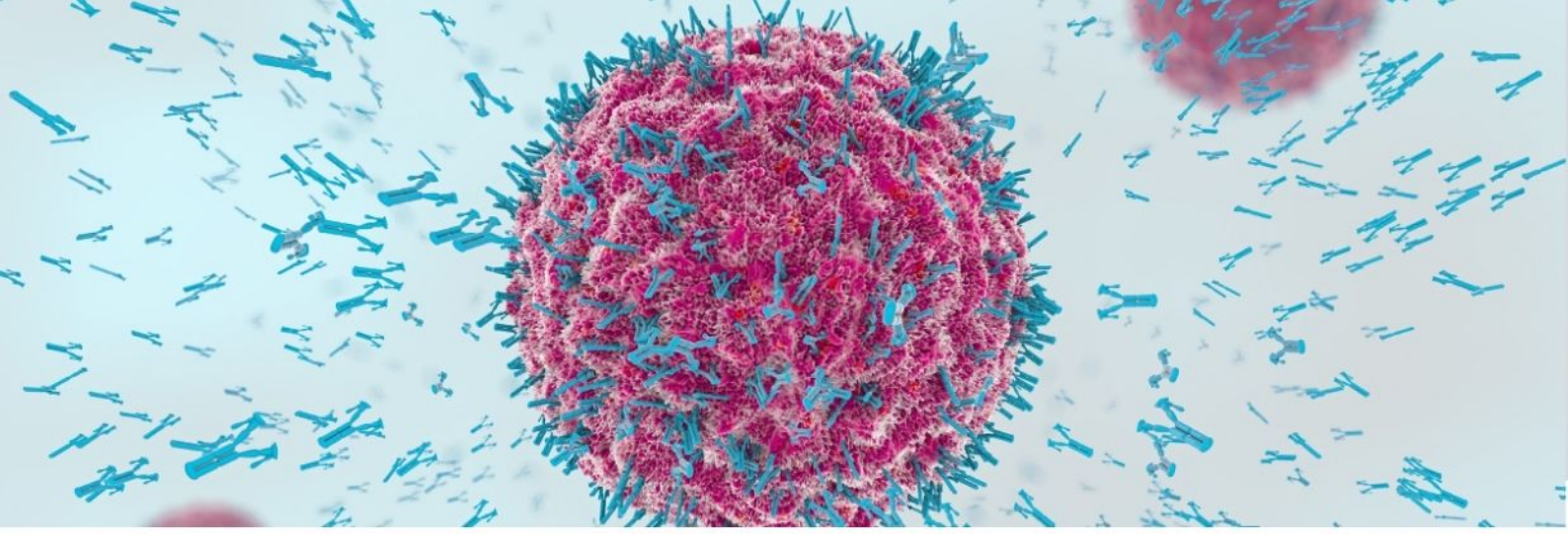

Single Cell Multiomics in Autoimmunity - Advance Your Immunology Studies with Single Cell and Spatial Tools

Free Virtual Seminar | January 26, 2022 | 9 AM PST

There are over 80 different types of autoimmune diseases that affect about 5% of the global population. Recent studies detailing the types of cells involved in autoimmune pathology have provided a critical step in understanding how autoimmune diseases develop, and they continue to inform new therapeutic and diagnostic development avenues. We invite you to attend our webinar, sponsored by 10x Genomics, to learn how integrating these cellular profiles with epigenomic, proteomic, and spatial analysis will help provide even richer insights.

In This Webinar, You Will Learn -
- How scientists are using genomics tools to characterize, diagnose, and treat autoimmune diseases
- Single cell tools that can be used to identify cell populations that may be masked when doing bulk RNA sequencing and/or flow or mass cytometry
- Integrating multiomic profiling into your research to discover deeper insights into autoimmune disease complexity

Register here

10x GENOMICS

CURRENT PROTOCOLS
A Wiley Brand

Received: 26 March 2020 | Accepted: 5 May 2020

DOI: 10.1002/jmv.25993

RESEARCH ARTICLE

JOURNAL OF MEDICAL VIROLOGY WILEY

3CL hydrolase-based multiepitope peptide vaccine against SARS-CoV-2 using immunoinformatics

Renu Jakhar[1] | Samander Kaushik[2] | Surendra K. Gakhar[1]

[1]Centre for Medical Biotechnology, Maharshi Dayanand University, Rohtak, Haryana, India

[2]Centre for Biotechnology, Maharshi Dayanand University, Rohtak, Haryana, India

Correspondence
Renu Jakhar, Centre for Medical Biotechnology, Maharshi Dayanand University, Rohtak, Haryana-124001, India.
Email: renujakhar22@gmail.com

Abstract

The present study provides the first multiepitope vaccine construct using the 3CL hydrolase protein of SARS-CoV-2. The coronavirus 3CL hydrolase (Mpro) enzyme is essential for proteolytic maturation of the virus. This study was based on immunoinformatics and structural vaccinology strategies. The design of the multiepitope vaccine was built using helper T-cell and cytotoxic T-cell epitopes from the 3CL hydrolase protein along with an adjuvant to enhance immune response; these are joined to each other by short peptide linkers. The vaccine also carries potential B-cell linear epitope regions, B-cell discontinuous epitopes, and interferon-γ-inducing epitopes. Epitopes of the constructed multiepitope vaccine were found to be antigenic, nonallergic, nontoxic, and covering large human populations worldwide. The vaccine construct was modeled, validated, and refined by different programs to achieve a high-quality three-dimensional structure. The resulting high-quality model was applied for conformational B-cell epitope selection and docking analyses with toll-like receptor-3 for understanding the capability of the vaccine to elicit an immune response. *In silico* cloning and codon adaptation were also performed with the pET-19b plasmid vector. The designed multiepitope peptide vaccine may prompt the development of a vaccine to control SARS-CoV-2 infection.

KEYWORDS

3CL hydrolase, COVID-19, immunoinformatics, multiepitope, peptide vaccine, SARS-CoV-2

1 | INTRODUCTION

Recently, millions of people in the world have lost their lives due to the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and 3 249 594 (WHO, 30 April 2020) people are affected with this deadly virus. This new virus strain outbreak started from Wuhan city of China in December 2019 and became a pandemic.[1,2] SARS-CoV-2 belongs to the family Coronaviridae and the order Nidovirales which is common among animals.[3] The Coronaviridae family is divided into four genera on the basis of their genetic properties, including alpha, beta, gamma, and delta coronavirus genera.[4] The SARS-CoV-2 is an enveloped positive-sense RNA, beta coronavirus with a genome of 29.9 kb.

There is an urgent need for the development of safe and effective vaccines against SARS-CoV-2 due to the continuous increase in patients. The development of a vaccine by conventional methods is a time-consuming process. Subsequently, a therapeutic approach based on in silico informatics has become supportive and can lend a hand for the prevention of disease. The development of a new vaccine for this new emergent strain by using therapeutic and preventive approaches can be readily applied to save human lives. The use of peptides or epitopes as therapeutics is a good strategy as it has the advantages of design, stability, and delivery.[5] Moreover, there is a growing emphasis on the use of peptides in vaccine design by predicting immunogenic cytotoxic T-cell (CTL), helper T-cell (HTL), and B-cell epitopes from tissue-specific proteins of organisms.

These epitopes can elicit both cellular immunity, and neutralizing antibodies against SARS-CoV-2 necessary for efficient vaccine development. More importantly, T-cell-based cellular immunity is essential for cleaning SARS-CoV-2 infection because it is memory-based.

When SARS-CoV-2 infects a cell, it generates multiple copies of itself by using the molecular machinery of the host to create the long chains of proteins required by the virus. These long viral proteins, however, only become functional when cut into smaller pieces by proteases.[6] Thus, the coronavirus proteases of SARS-CoV-2 play an integral role in propagating the virus.[7] The coronavirus 3CL hydrolase (Mpro) enzyme, also known as the 3C-like main protease, is essential for proteolytic maturation of the virus.[8] Its distinctive heart shape is the result of two identical protein subunits coming together to form a functional protease. 3CL hydrolase is a proven drug discovery target in the case of SARS-CoV.[9] In the present study, an attempt has been made for multiepitope vaccine prediction against SARS-CoV-2 by utilizing immunoinformatic approaches. The study was designed to advance the path of vaccine development. Short peptides as potential epitopes, CTLs, and HTLs were identified in this study. The vaccination aims to induce immunity against specific pathogens.[10] By a combination of these epitopes, the vaccine is capable of inducing a specific humoral or cellular immune response against specific pathogens.[11] We used these epitopes to design a multiepitope vaccine construct. We report here, the predicted multiepitope peptide vaccine from SARS-CoV-2 3CL hydrolase for further consideration in vitro and in vivo analysis.

2 | MATERIALS AND METHODS

2.1 | Retrieval of protein sequence from the database

The 3CL hydrolase (main protease) protein sequence of SARS-CoV-2 (accession no. 6LU7) was retrieved from the NCBI database. A total of 180 3CL hydrolase protein sequences from SARS-CoV-2 were retrieved from the NCBI database until 25 April 2020 (Supporting Information File 1). An NCBI conserved domain search was used to discover the conserved domains of 3CL hydrolase. These 180 sequences retrieved were collected from different parts of the world; the retrieved sequences and their accession numbers are listed in the Supporting Information File 2. Furthermore, the multiple sequence alignment of 3CL hydrolase protein sequences was carried out through clustalW.

2.2 | Prediction of CTL epitope

A NetCTL server[12] was used to predict T-cell epitopes. The threshold score for epitope identification was set as 0.50 and all the supertypes were taken during the submission of a protein sequence. On the basis of the combined score, the selection of the best epitopes was done. The putative epitope candidate's epitopes were further tested with the MHC-I binding tool at the Immune Epitope Database (IEDB).[12] The stabilized matrix-based method was used to calculate the threshold values for strong binding peptides (IC50). For further analysis of the selected epitopes, alleles having an IC50 value less than 200 nm were selected.[13] The immunogenicity prediction tool at IEDB was used to predict the immunogenicity of the epitopes.[14]

2.3 | Prediction of HTL epitope

HTL epitopes were predicted by using the MHC-II binding tool on IEDB.[15] It covers all HLA class II alleles including HLA-DR, HLA-DP, and HLA-DQ.[16] An IC50 below 200 nM shows the maximum interaction potential of the HTL epitope and MHC-II allele.[17] Accordingly, putative epitopes were selected. The predicted HTL epitopes were submitted to the interferon (IFN) epitope server to check whether the MHC-II-binding epitopes had the ability to induce IFN-γ.[18]

2.4 | B-cell epitope prediction

Linear B-cell epitope prediction carried out by using six different prediction tools at the IEDB resource. The protein sequence was subjected to Bepipred linear epitope prediction,[19] Emini[20] surface accessibility, Kolaskar and Tongaonkar[21] antigenicity, Parker[22] hydrophilicity, Chou and Fasman[23] beta turn, and Karplus and Schulz[24] Flexibility Prediction prediction methods in IEDB, that predict the probability of specific regions in the protein binding to the B-cell receptor, being on the surface, being immunogenic, being in a hydrophilic region and being in a beta-turn region, respectively. Structure-based linear and discontinuous B-cell epitopes were generated by using the Ellipro tool from IEDB resource.[25]

2.5 | Epitope antigenicity, allergenicity, and toxicity prediction

The antigenicity of the predicted CTL, HTL, and B-cell epitopes was predicted by the VaxiJen v2.0 server[26] with all the default parameters. The allergenicity was analyzed by AllerTOP.[27] A ToxinPred server was used for toxicity assessment of epitopes.[28]

2.6 | Epitope conservation analysis

The IEDB Conservancy analysis resource was also used for the analysis of the epitope conservancy.[29] The conservancy analysis of epitopes was done among all 3CL hydrolase protein sequences of the SARS-CoV-2 retrieved from the data base. Only 100% conserved epitopes were selected.

2.7 | Population coverage prediction

Human population coverage for selected epitopes was checked by the population coverage tool at IEDB.[30] Every epitope and their HLA alleles were added; also, various ethnic groups and geographical regions across the world were selected.

2.8 | Construction of multiepitope peptide

The HTL and CTL epitopes selected from the 3CL hydrolase protein of SARS-CoV-2 were considered for the final vaccine construct. The vaccine construct consists of adjuvants at the N-terminal, immunogenic CTL and HTL epitopes connected to each other by EAAAK and GGGGS linkers.[10] Algpred and AllergenFP was used to predict allergenicity of multiepitope peptide.[31,32]

2.9 | Secondary structure analysis

Antigenicity depends on the secondary and tertiary structure of the protein. Therefore, ExPASy's secondary structure prediction server ProtParam[33] was used. Physiochemical properties including the amino acid composition, molecular weight, theoretical pI, extinction coefficient, atomic composition, grand average of hydropathicity (GRAVY) values, estimated half-life, instability index, and aliphatic index were investigated. The self-optimized prediction method was used to predict alpha helix, coiled structures, and beta sheets of the vaccine construct.[34] A Vaxijen server was used for antigenicity prediction of the peptide vaccine.

2.10 | Tertiary structure of vaccine constructs

The tertiary structure of the peptide vaccine was predicted by employing Swissmodeller,[35] which utilizes alignment of the target sequence and template structures for homology modeling. Energy minimization was done with Chimera.[36] The final tertiary Structure Refinement of the obtained model was processed by the Galaxyrefine tool[37] for generating loop modeling. The structure validation of selected tertiary structure was carried out with Rampage and Swiss-PdbViewer.[38]

2.11 | In silico codon adaptation and cloning

The template DNA sequence was reverse translated from the peptide sequence. A Java Codon Adaptation Tool (JCAT) was used for codon optimization.[39] The JCAT was used to improve translational efficiency in the mammalian host cell line.[40] Finally, the adapted nucleotide sequence was cloned into the pET-19b plasmid vector by using the SnapGene tool.[41]

2.12 | Protein-protein docking

A three-dimensional structure of toll-like receptor-3 (TLR-3) (PDB id: 2A0Z) was obtained from RCSB. TLR-3 was used as a receptor and the final peptide vaccine model was used as a ligand for docking studies.[10] Analysis of protein-protein docking and binding interactions between TLR-3 and the vaccine construct was assessed using a HEX server.[42] Chimera was used for visualization and determination of binding sites.

3 | RESULTS AND DISCUSSION

3.1 | CTL epitope identification

T-cell based development of vaccines seems to have potential because of antigenic drift as foreign particles can easily engineer the escape from the antibody memory response. In addition, T-cell-mediated immunity tends to be long-lasting. Epitopes having high combinatorial scores were considered as the epitopes with the most potential, as predicted by the NetCTL prediction tool. The predicted total scores of the proteasome score, tap score, MHC score, processing score, and MHC-I binding are summarized as a total score in Table 1. Then, the affinity (IC50, <200) of these potential epitopes with MHC-I alleles was further analyzed by IEDB. The affinity for binding of the epitopes with the MHC-I alleles was inversely proportional to the IC50 values. The peptide that passes several criteria is considered to be a good epitope candidate, such as possessing antigenicity, nonallergen, and immunogenicity. A total number of 180 3CL hydrolase protein sequences of SARS-CoV-2 retrieved from the NCBI database were aligned to witness the conservation of the predicted epitopes. All predicted epitopes were found to be 100% conserved. Also, these epitopes are antigenic, immunogenic, nonallergic, and nontoxic in nature.

Among these selected seven T-cell epitopes, the 9-mer epitope, GSVGFNIDY was found to have the highest combined score and immunogenicity but it interacted with only two MHC-I alleles. However, the TLNDFNLVA epitope found to have a good combined score and showed interaction with four MHC-I alleles. The epitope MLNPNYEDL was found to have the maximum number of allelic interactions with good population coverage rather than the other epitopes. Similarly, SAQTGIAVL has a good allelic interaction and shows good immunogenicity.

3.2 | HTL epitope identification

The epitopes with a higher affinity (IC50 <200) with MHC-II alleles were selected for analysis. The putatively selected nine HTL (9-mer sequences) epitopes were antigenic, nonallergic, nontoxic, and had interactions with numerous HLA-DR alleles (Table 2). The percentage of the peptide conservancy was 100% for all predicted epitopes. The predicted HTL epitopes were submitted to the IFN epitope server to check whether the MHC-II-binding epitopes had

TABLE 1 The immunogenicity, antigenicity, and allergenicity of putative cytotoxic candidates T-cell epitopes of COVID-19 3CL hydrolase protein interacting with MHC-II alleles

Epitopes	Position in sequence	Combined score	Interaction of MHC-I allele with an affinity IC50 < 200	Immunogenicity	Antigenic	Allergen	Toxic
GSVGFNIDY	146	1.6	HLA-C*12:03, HLA-C*03:03	0.28	Yes	No	No
TLNDFNLVA	226	1.1	HLA-C*05:01, HLA-C*12:03 HLA-C*14:02, HLA-A*02:01	0.143	Yes	No	No
MLNPNYEDL	49	0.9	HLA-B*15:02, HLA-C*03:03, HLA-C*05:01, HLA-C*14:02, HLA-C*07:02, HLA-C*12:03, HLA-C*07:01	0.07	Yes	No	No
SAQTGIAVL	254	0.9	HLA-C*03:03, HLA-B*15:02, HLA-C*12:03, HLA-C*15:02, HLA-C*14:02	0.22	Yes	No	No
LAWLYAAVI	205	0.8	HLA-C*15:02, HLA-C*12:03, HLA-C*03:03, HLA-C*14:02	0.15	Yes	No	No
TTLNDFNL	225	0.8	HLA-A*02:06, HLA-C*12:03, HLA-C*15:02, HLA-C*14:02, HLA-C*03:03	0.11	Yes	No	No
DEFTPFDVV	289	0.8	HLA-C*12:03, HLA-B*18:01	0.21	Yes	No	No

the ability to induce IFN-γ, positive results, confirming that these epitopes have the capability to induce IFN-γ. The epitope YKFVRIQPG was found to induce IFN-γ and have the maximum number of allele-binding interactions among all predicted epitopes.

3.3 | Prediction of linear and conformational B-cell epitopes

The potential linear B-cell epitopes should get above threshold scores in Bepipred linear epitope prediction, Emini surface accessibility, Parker hydrophobicity, Karplus & Schulz Flexibility Prediction, and Chou and Fasman beta turn prediction methods in IEDB (Figure S1). Nine epitopes predicted by the Bepipred linear epitope prediction method showed overlapped sequences with the Emini surface accessibility method and these epitopes also satisfied the threshold score of the other five prediction methods (Table S1).

The conformational epitopes were predicted by using a three-dimensional (3D) structure of the 3CL hydrolase protein of SARS-CoV-2. ElliPro gives a score to each output epitope, which is the Protrusion Index (PI) value averaged over each epitope residue.[43] Some ellipsoids approximated the tertiary structure of the protein. The highest probability of a conformational epitope was calculated at 76% (PI score: 0.76). The residues involved in conformational epitopes, their number, location and scores are shown in Table S2. The overlapping regions between the predicted linear B-cell epitopes and

TABLE 2 The antigenicity, nontoxicity, and allergenicity of putative helper candidates T-cell epitopes of SARS-CoV-2 3CL hydrolase protein interacting with MHC-II alleles

Epitope	Position in sequence	Interaction of MHC-II alleles having IC50 < 200 nm	Antigenic	Allergen	Toxic	Induce IFN-γ
YKFVRIQPG	101	HLADRB4*01:01, HLA-DRB1*09:01, HLA-DRB5*01:01, HLA-DRB1*01:01, HLA-DRB1*04:04, HLA-DRB1*07:01	Yes	No	No	Positive
KYKFVRIQP	100	HLA-DRB1*11:01	Yes	No	No	Positive
FVRIQPGQT	103	HLA-DRB4*01:01, HLA-DRB5*01:01, HLA-DRB1*01:01	Yes	No	No	Positive
LAWLYAAVI	207	HLA-DRB1*01:01	Yes	No	No	Positive
PLSAQTGIA	252	HLA-DRB1*01:01	Yes	No	No	Positive
GPLSAQTGI	251	HLA-DRB1*09:01	Yes	No	No	Negative
VSFCYMHHM	157	HLA-DRB1*07:01, HLA-DRB1*01:01	Yes	No	No	Negative
ILGPLSAQT	249	HLA-DRB1*01:01	Yes	No	No	Negative
EDLLIRKSN	55	HLA-DRB1*15:01	Yes	No	No	Negative

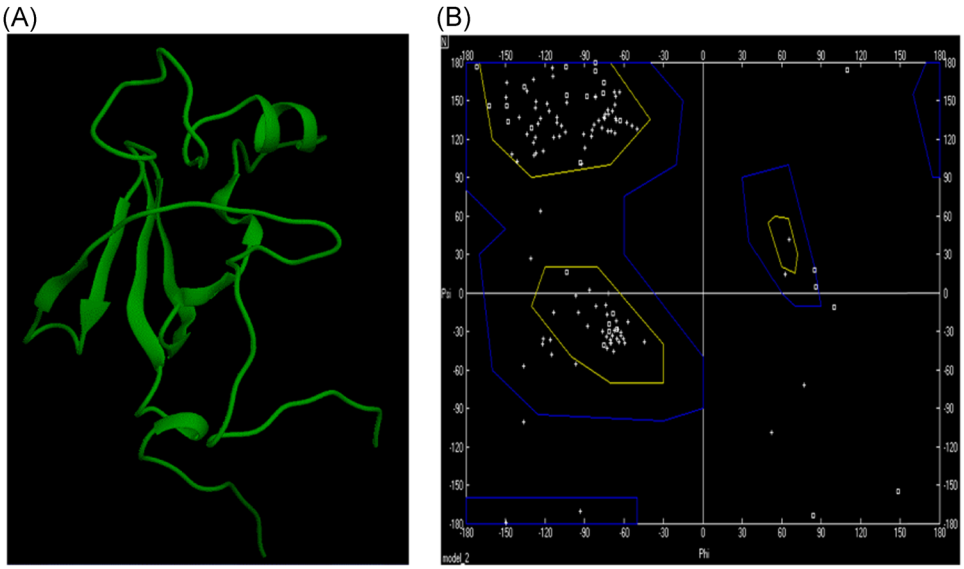

FIGURE 1 A, Structure of multiepitope protein prepared by Swissmodeller. B, Ramachandran plot of multiepitope protein of SARS-CoV-2. Amino acids represented by cross sign except for glycine shown with square

conformational epitopes presented highly immunogenic surface assessable epitopes. The proposed epitope of the B cell is conserved in all retrieved sequences of the SARS-CoV-2 3CL hydrolase protein.

There were several overlapping sequences found between B- and T-cell epitopes. The overlapping sequences between CTL and B-cell epitopes included ^{49}MLNPNYEDL^{57}, ^{146}GSVGFNIDY^{154}, ^{225}TTLNDFNL^{233}, ^{289}DEFTPFDVV^{297}, and the overlapping sequences between HTL and B-cell epitopes included ^{55}EDLLIRKSN^{63}, ^{100}KYKFVRIQP^{108}, ^{101}YKFVRIQPG^{109}, and ^{103}FVRIQPGQT^{111}. CTL and HTL epitopes were also screened with some overlapping amino acid sequences/regions, suggesting the possibility of antigen presentation to immune cells via both MHC class-I and -II pathways, that is, ^{254}SAQTGIAVL^{262} (CTL), ^{251}GPLSAQTGI^{34} (HTL), and ^{252}PLSAQTGIA^{260} (HTL) (Tables 1 and 2).

The predicted CTL (^{49}LNPNYEDL^{57}, ^{146}GSVGFNIDY^{154}, ^{225}TTLNDFNL^{233}, and ^{289}DEFTPFDVV^{297}) and HTL (^{101}YKFVRIQPG^{109}, ^{252}PLSAQTGIA^{260}, ^{55}EDLLIRKSN^{63}, and ^{157}VSFCYMHHM^{165}) epitopes carrying overlapping potential B-cell linear epitope regions, B-cell discontinuous epitopes as well as IFN-γ-inducing epitopes for cellular and humoral immunity, and could be synthesized for further in vivo and in vitro assays. These results are based on an analysis of available data on various immune databases. Also, an NCBI conserved domain search of this protein shows the presence of a conserved domain from amino acid 4 to 300. All short-listed epitopes lie within the conserved domain of 3CL hydrolase. The results of the present study suggest that the predicted epitopes are good candidates for making a peptide vaccine, which may initiate an effective immune response in vivo.

3.4 | Population coverage

Different ethnic populations have a high polymorphism in HLA and its reaction to T-cell epitopes is restricted by HLA proteins. Therefore, to stimulate immune responses in human populations among the world, the HLA specificity of T-cell epitopes has to be measured as the main criteria for the selection of the epitopes. A high population coverage was found in all putative HTL epitopes and CTL epitopes in 16 identified geographic regions of the world. On the basis of the above study, the epitope candidates should bind to the maximum HLA alleles to get a better population coverage.

In this study, the population coverage of the combined use of all short-listed HTL and CTL epitopes for the vaccine construct was found to be 84.7% on an average around the world and reached above-average values in Europe (89%), North America (83%), East Asia (82%), and South Asia (80%).

3.5 | Multiepitope-based vaccine

We used the epitopes predicted from SARS-CoV-2 3CL hydrolase to design a multiepitope vaccine construct. These predicted CTL and HTL epitopes were used for making a peptide vaccine which may initiate an effective immune response in vivo.

Keeping in mind the end goal to effectively activate both innate and adaptive immune response, the multiepitope vaccine must consist of a strong immunostimulatory adjuvant.

3.6 | Multiepitope vaccine construction

The predicted CTL (^{49}MLNPNYEDL^{57}, ^{146}GSVGFNIDY^{154}, ^{225}TTLNDFNL^{233}, and ^{289}DEFTPFDVV^{297}) and HTL (^{101}YKFVRIQPG^{109}, ^{252}PLSAQTGIA^{260}, ^{55}EDLLIRKSN^{63}, and ^{157}VSFCYMHHM^{165}) epitopes were selected from the 3CL hydrolase protein of SARS-CoV-2 and incorporated into the final vaccine construct. Our final vaccine construct consisted of adjuvants at the N-terminal, immunogenic CTL and HTL

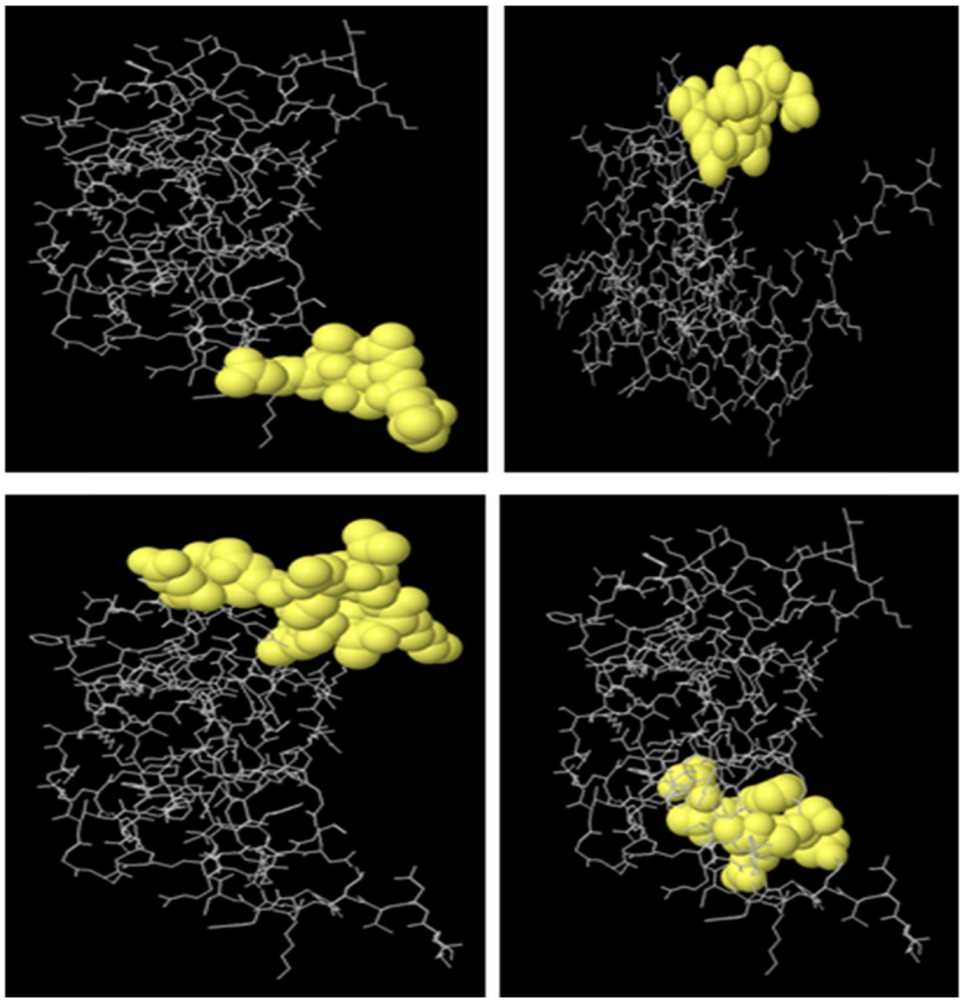

FIGURE 2 Three-dimensional representations of conformational epitopes on the multiepitope protein construct showing the residues composed of antigenic epitopes that will come in direct contact with the immune receptor. The antigenic residues are represented by a yellow surface, and the bulk of the protein is represented in sticks

epitopes connected to each other by linkers.[44] The adjuvant β-defensin-3, hBD-3 (PDB ID: 1KJ6), sequence—GIINTLQKYYCRVRGGRCAV LSCLPKEEQIGKCSTRGRKCCRRKK was used as an adjuvant for the N-terminal, connected to the CTL epitope by the EAAAK linker, and intra CTL and HTL connected to each other by GGGGS.[10] Linkers are essential components of the multiepitope peptide vaccines[45] that have functional and structural roles in the vaccine construct.[46] A final vaccine construct of 156 amino acid residues was designed.

Sequence of vaccine construct—

GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKK **EAAAK**MLNPNYEDL**GGGGS**GSVGFNIDY**GGGGS**TTLNDFNL**GGGGS** DEFTPFDVV**GGGGS**EDLLIRKSN**GGGGS**YKFVRIQPG**GGGGS**VSFCY MHHM**GGGGS**PLSAQTGIA

3.7 | Antigenicity and allergenicity prediction

The antigenicity was predicted by using a vaxijen server. The 0.8% antigenicity probability shows that the vaccine construct's nature is antigenic and able to stimulate T- and B-cell responses. AllerTOP, Algpred, and AllergenFB tools predicted the vaccine protein as nonallergic, Algpred predicted the overall score as 0.54 which is above the threshold value, meaning the protein vaccine does not cause any type of allergic response and inflammation. It is safe for human use.[47]

3.8 | Physiochemical properties assessment

The molecular weight of the vaccine protein was predicted to be approx 15.9 kDa. The theoretical pI was found to be 9.02 showing a slightly basic nature. The total number of positively charged residues was 18 and that of negatively charged residue was 12. The predicted half-life was 30 hours in mammalian reticulocytes, in vitro; while more than 20 hours in yeast, in vivo and more than 10 hours in *Escherichia coli*, in vivo. The extinction coefficient was found to be 8950 $M^{-1}{\cdot}cm^{-1}$, at 280 nm measured in water. The instability index was 56.1. The GRAVY value was found as −0.372, not only

indicating its hydrophilic nature but also indicating the presence of residues mostly on the surface.[39] The value of the aliphatic index was found to be 59.3 and the estimated value of the aliphatic index showed the designed vaccine was thermostable in nature. The secondary structure prediction showed the presence of 14.7% helix, 32% extended strand, 10.9% beta turn, and 42.3% coiled structure.

3.9 | Tertiary structure prediction, refinement, and validation

After our vaccine construction, the primary amino acid sequence of the multiepitope protein was submitted for modeling its 3D structure. The tertiary structure of the protein vaccine construct was built by swiss-modeler (Figure 1A). Furthermore, to achieve a high-quality 3D structure for docking analysis, the 3D protein structure was then applied for the refinement process using GalaxyRefine. Validation of refined tertiary structures was checked by Verify 3D, Ramachandran plot, and Prosa Z score and the best-refined model were selected. This led to an increase in the number of residues in the favored region (Figure 1B). Before refinement, 88% of residues were found in favored region, whereas in the refined model, the residues in the favored region reached 96%, 4% residue was placed in the allowed region, and 0% in outlier regions.[48]

3.10 | B-cell epitope mapping

Out of the 61 residues, five conformational B-cell epitopes were identified by using the final 3D model of the vaccine construct. Epitopes ^{124}RIQPG^{128} and ^{2}IINTLQ^{7}, nearly 5 to 6 amino acids long were found with the highest PI score (0.91). Other epitopes, that is, ^{46}EAAAKMLNPNYED^{58} and ^{89}GGSDEFTP^{96} were also screened with a good PI score (0.77). The obtained Protrusion score confirms that the designed subunit vaccine is assessable to the surface and immunogenic in nature[43] (Figure 2).

Thus, our final vaccine construct consists of both B- and T-cell segments that may have the ability to enhance cell-mediated as well as humoral immunity.

3.11 | Molecular docking of protein vaccine constructs with TLR-3

Analyses of protein-protein interactions between TLR-3 and vaccine construct analyses of protein-protein interactions between TLR-3 and vaccine construct were assessed using a HEX server (Figure 3). A docking study between the TLR-3 receptor and vaccine protein was conducted and the docked complex showed the lowest energy score of −670. The docking of the modeled vaccine construct with TLR-3 was done indicating durable interactions. A strong binding affinity was found between these two structures.[49]

3.12 | Codon adaptation and in silico cloning

Reverse translation and codon optimization of the designed vaccine construct into the nucleotide sequence were performed by a codon usage wrangler server. The codons were adapted as per the mammalian host cell line (human). A codon adaptive index (CAI) > 0.8 is

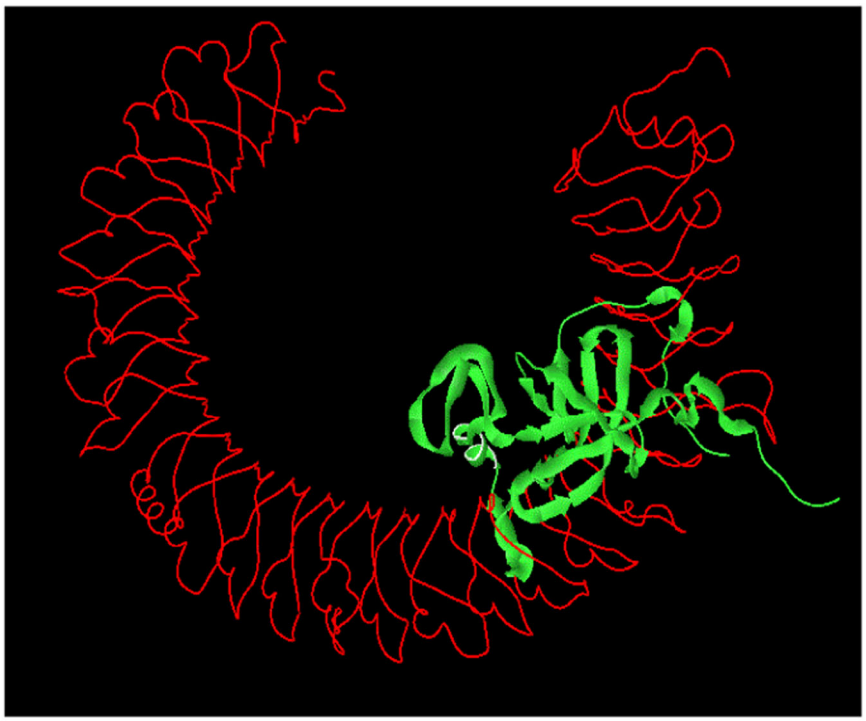

FIGURE 3 Docked complex of vaccine protein and toll-like receptor-3 (TLR-3). Here, the vaccine construct is shown in ribbon form, docked within the TLR-3 receptor

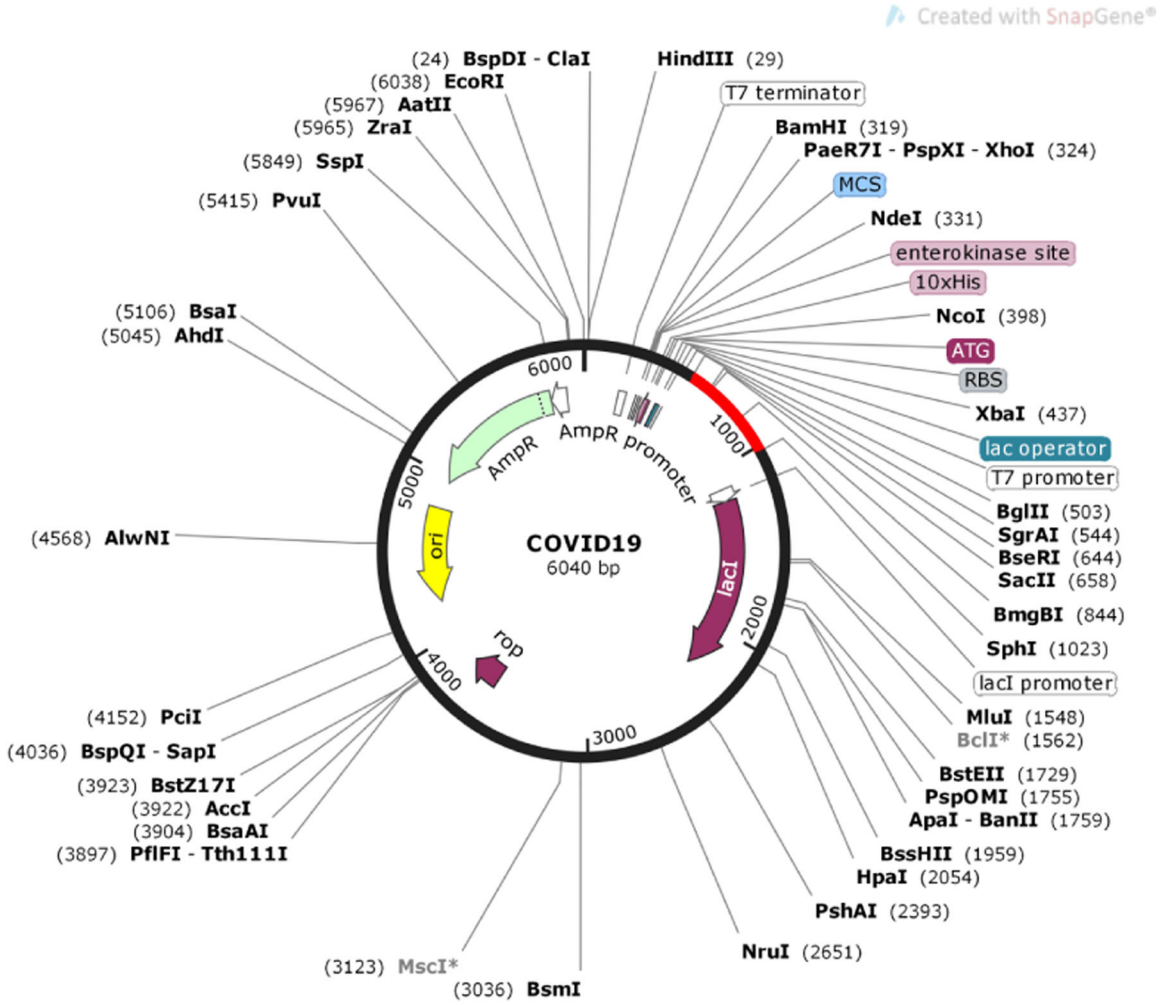

FIGURE 4 In silico cloning for adapted vaccine sequence into pET-19b vector showing the region of choice in red surrounded by SgrAI and SphI restriction sites while the vector is shown in black lines

considered good and optimal guanine-cytosine (GC) content between 30% to 70%, and codon frequency distribution (CFD) < 30% are the important properties of the gene sequence to achieve high expression level in the host.[50] Values within these ranges are suitable for transcription and translation with high efficacy and also for expression. The CAI of our optimized complementary DNA (cDNA) sequence was 0.99, the average GC content was 70.9% and the 0% CFD supported the maximum protein expression in the mammalian host cell line (human). Later, to clone the gene in pET-19b vector plasmid, SgrAI and SphI restriction sites were introduced. The desired sequence was present in between the mentioned restriction sites in the clone and represented in red (Figure 4). The clone was 6040 base pairs in size.

The constructed vaccine consists both B- and T-cell epitopes, has antigenic properties and absence of allergic and toxic properties. The codon based generated cDNA of multiepitope vaccine constructs are shown to have a higher expression in the human (host) cell line and have good binding with the TLR-3 receptor. Moreover, this study needs experimental validation for confirmation.

4 | CONCLUSION

In the present study, we attempted to design a multiepitope vaccine by immunoinformatics driven tools for the selection of vaccine targets to step up the development of a vaccine against SARS-CoV-2. The vaccine peptide consists of HTL, CTL and B-cell epitopes derived from SARS-CoV-2 3CL hydrolase protein as potential vaccine targets. This vaccine hopefully generates cell-mediated and humoral immunity and can act as an effective immune response. In silico prediction and prescreening methods have potential as it offers the scientific community with valuable data for further wet lab trials. Overall, to validate the efficacy of our epitope vaccine, it requires to be cloned, expressed and tested by in vitro and in vivo experiments in the laboratory, along with this in silico study for proficient use as a vaccine against SARS-CoV-2.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

RJ conducted the study, performed in silico analysis, and wrote the manuscript. SKG planned the study and revised the manuscript. SK analyzed the data.

ORCID

Renu Jakhar http://orcid.org/0000-0002-3556-6536

REFERENCES

1. Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance*. 2020; 25(3):2000045
2. Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan China: the mystery and the miracle. *J Med Virol*. 2020;92(4):401–402.
3. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;3:1-5.
4. Chan JF, Yuan S, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet*. 2020; 395(10223):514-523.
5. Sanchez-Trincado JL, Gomez-Perosanz M, Reche PA. Fundamentals and methods for T-and B-cell epitope prediction. *J Immunol Res*. 2017. 2017:2680160.
6. Ghosh AK, Xi K, Ratia K, et al. Design and synthesis of peptidomimetic severe acute respiratory syndrome chymotrypsin-like protease inhibitors. *J Med Chem*. 2005;48(22):6767-6771.
7. Pillaiyar T, Manickam M, Namasivayam V, Hayashi Y, Jung SH. An overview of severe acute respiratory syndrome–coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy. *J Med Chem*. 2016;59(14):6595-6628.
8. Needle D, Lountos GT, Waugh DS. Structures of the Middle East respiratory syndrome coronavirus 3C-like protease reveal insights into substrate specificity. *Acta Crystallogr D*. 2015;71(5):1102-1111.
9. Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science*. 2003;300(5626):1763-1767.
10. Srivastava S, Kamthania M, Singh S, Saxena AK, Sharma N. Structural basis of development of multi-epitope vaccine against middle east respiratory syndrome using in silico approach. *Infect Drug Resist*. 2018;11:2377.
11. Shi J, Zhang J, Li S, Sun J, Teng Y, Wu M. Epitope-based vaccine target screening against highly pathogenic MERS-CoV: an in silico approach applied to emerging infectious diseases. *PLOS One*. 2015;10(12): e0144475.
12. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*. 2007;8(1):424.
13. Vita R, Overton JA, Greenbaum JA, et al. The Immune Epitope Database (IEDB) 3.0. *Nucleic Acids Res*. 2015;43(D1):D405-D412.
14. Calis JJ, Maybeno M, Greenbaum JA, et al. Properties of MHC class I presented peptides that enhance immunogenicity. *PLOS Comput Biol*. 2019;9:10.
15. Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3.0, a common panspecific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. *Immunogenetics*. 2013;65(10):711-724.
16. Wilson JA, Hart MK. Protection from Ebola virus mediated by cytotoxic T lymphocytes specific for the viral nucleoprotein. *J Virol*. 2001; 75(6):2660-2664.
17. Zhang L, Udaka K, Mamitsuka H, Zhu S. Toward more accurate pan-specific MHC-peptide binding prediction: a review of current methods and tools. *Brief Bioinform*. 2012;13(3):350-364.
18. Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct*. 2013;8:30.
19. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res*. 2006;2:2.
20. Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol*. 1985;55:836-839.
21. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett*. 1990;276: 172-174.
22. Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X ray-derived accessible sites. *Biochemistry*. 1986;25: 5425-5432.
23. Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol*. 1978;47:45-148.
24. Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. *Naturwissenschaften*. 1985;72:212-213.
25. Ponomarenko J, Bui HH, Li W, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*. 2008; 9:514.
26. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007;8:4.
27. Dimitrov I, Flower DR, Doytchinova I. AllerTOP: a server for in silico prediction of allergens. *BMC Bioinformatic*. 2013;14(suppl 6):S4.
28. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP. Open Source Drug Discovery Consortium. In silico approach for predicting toxicity of peptides and proteins. *PLOS One*. 2013;8:9.
29. Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2007;8:361.
30. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2006;7:153.
31. Saha S, Raghava GP. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res*. 2006;34(suppl_2): W202-W209.
32. Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics*. 2014; 30(6):846-851.
33. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*. 1999;112:531-552.
34. Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*. 1995;11(6):681-684.
35. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296-W303.
36. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera, a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605-1612.
37. Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res*. 2013;41(W1): W384-W388.
38. Johansson MU, Zoete V, Michielin O, Guex N. Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. *BMC Bioinformatics*. 2012;13(1):173.
39. Khatoon N, Pandey RK, Prajapati VK. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunitvaccine using immunoinformatics approach. *Sci Rep*. 2017;7:8285.

40. Grote A, Hiller K, Scheer M, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res*. 2005;33(suppl_2):W526-W531.
41. Pandey RK, Bhatt TK, Prajapati VK. Novel immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating anopheles salivary protein. *Sci Rep*. 2017;8(1):1125.
42. Macindoe G, Mavridis L, Venkatraman V, Devignes MD, Ritchie DW. HexServer: an FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res*. 2010;38(suppl_2): W445-W449. 5.
43. Jakhar R, Kumar P, Sehrawat N, Gakhar SK. A comprehensive analysis of amino-peptidase N1 protein (APN) from Anopheles culicifacies for epitope design using immuno-informatics models. *Bioinformation*. 2019;15(8):600.
44. Hu W, Li F, Yang X, et al. A flexible peptide linker enhances the immunoreactivity of two copies HBsAg preS1 (21–47) fusion protein. *J Biotech*. 2004;107(1):83-90.
45. Duits LA, Nibbering PH, vanStrijen E, et al. Rhinovirus increases human β-defensin-2 and-3 mRNA expression in cultured bronchial epithelial cells. *FEMS Immunol Med Microbiol*. 2003;38(1):59-64.
46. Chauhan V, Rungta T, Goyal K, Singh MP. Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach. *Sci Rep*. 2019;9(1):1-5.
47. Azim KF, Hasan M, Hossain MN, et al. Immunoinformatics approaches for designing a novel multi epitope peptide vaccine against human norovirus (Norwalk virus). *Infect Genet Evol*. 2019;74: 103936.
48. Nezafat N, Karimi Z, Eslami M, Mohkam M, Zandian S, Ghasemi Y. Designing an efficient multi-epitope peptide vaccine against Vibrio cholerae via combined immunoinformatics and protein interaction based approaches. *Comput Biol Chem*. 2016;62:82-95.
49. Ojha R, Pareek A, Pandey RK, Prusty D, Prajapati VK. Strategic development of a next-generation multi-epitope vaccine to prevent Nipah virus zoonotic infection. *ACS Omega*. 2019;4(8):13069-13079.
50. Solanki V, Tiwari M, Tiwari V. Prioritization of potential vaccine targets using comparative proteomics and designing of the chimeric multi-epitope vaccine against Pseudomonas aeruginosa. *Sci Rep*. 2019; 9(1):1-9.

SUPPORTING INFORMATION

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

How to cite this article: Jakhar R, Kaushik S, Gakhar SK. 3CL hydrolase-based multiepitope peptide vaccine against SARS-CoV-2 using immunoinformatics. *J Med Virol*. 2020;92: 2114–2123. https://doi.org/10.1002/jmv.25993