JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
https://doi.org/10.1080/07391102.2020.1844060

Multi-epitope vaccine against SARS-CoV-2 applying immunoinformatics and molecular dynamics simulation approaches

Jyotisha, Samayaditya Singh and Insaf Ahmed Qureshi

Department of Biotechnology and Bioinformatics, School of Life Sciences, University of Hyderabad, Hyderabad, Telangana, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

COVID-19, caused by SARS-CoV-2, is severe respiratory illnesses leading to millions of deaths worldwide in very short span. The high case fatality rate and the lack of medical counter measures emphasize for an urgent quest to develop safe and effective vaccine. Receptor-binding domain (RBD) of spike protein of SARS-CoV-2 binds to the ACE2 receptor on human host cell for the viral attachment and entry, hence considered as a key target to develop vaccines, antibodies and therapeutics. In this study, immunoinformatics approach was employed to design a novel multi-epitope vaccine using RBD of SARS-CoV-2 spike protein. The potential B- and T-cell epitopes were selected from RBD sequence using various bioinformatics tools to design the vaccine construct. The *in silico* designed multi-epitope vaccine encompasses 146 amino acids with an adjuvant (human beta-defensin-2), which was further computationally evaluated for several parameters including antigenicity, allergenicity and stability. Subsequently, three-dimensional structure of vaccine construct was modelled and then docked with various toll-like receptors. Molecular dynamics (MD) study of docked TLR3-vaccine complex delineated it to be highly stable during simulation time and the stabilization of interaction was majorly contributed by electrostatic energy. The docked complex also showed low deformation and increased rigidity in motion of residues during dynamics. Furthermore, *in silico* cloning of the multi-epitope vaccine was carried out to generate the plasmid construct for expression in a bacterial system. Altogether, our study suggests that the designed vaccine candidate containing RBD region could provide the specific humoral and cell-mediated immune responses against SARS-CoV-2.

ARTICLE HISTORY
Received 30 July 2020
Accepted 23 October 2020

KEYWORDS
SARS-CoV-2; receptor-binding domain; immunoinformatic approach; multi-epitope vaccine; molecular dynamics simulation

1. Introduction

Since the end of 2019, a big group of human pneumonia cases was reported by Chinese authorities, from Wuhan city of China and the disease was termed as coronavirus disease 2019 (COVID-19). Whole-genome sequencing of the causative agent identified it as a novel coronavirus which was officially designated as SARS-CoV-2. It causes severe respiratory infections including pneumonia and acute respiratory failure causing loss of life (Li & Ma, 2020; Zhu et al., 2020). SARS-CoV-2 is predominantly transmitted among the individuals through contact routes and respiratory droplets. The situation with COVID-19 is worsening worldwide in quicker manner as the number of verified deaths has reached into the millions (WHO, 2020). Particularly, the old age people with comorbidities have been found more prone to the infection of COVID-19. The treatment depends upon clinical symptoms controlling and oxygen therapy along with usage of mechanized ventilators for patients suffering from respiratory failure. Even though some antiviral drugs such as remdesivir are being tested against the fatal disease (Wang et al., 2020), till date no drug has been precisely approved for COVID-19. The high case fatality rate and the lack of medical counter measures advocate for an urgent need of a safe vaccine that promptly induces effective and long-lasting immune response against this infectious agent.

SARS-CoV-2 is an envelope virus with positive-sense, single-stranded RNA genome, which encodes several non-structural proteins (nsp) as well as four major structural proteins, namely spike (S), nucleocapsid (N), membrane (M), and envelope (E) proteins (Phan, 2020). It binds to surface cellular receptors of the human host and initiate infection; hence recognition of specific receptor is considered as one of the most important event and defines the tissue tropism of a virus. The S protein of SARS-associated coronaviruses is surface exposed and contains two subunits (S1 and S2). It plays the most important roles during entry of viral particles into its host through firstly binding to the S1 subunit followed by integrating the viral and host– cell membranes by S2 subunit (Liu et al., 2004), which makes S protein one of the most promising antigen formulations for vaccine research against the deadly COVID-19 disease. Moreover, homologue proteins have been used for formulating the vaccine against MERS-CoV and SARS-CoV and were found to be effective (Du et al., 2009; Zhou et al., 2018). MERS-CoV identifies dipeptidyl peptidase 4 (DPP4) as its host receptor (Raj et al., 2013), while SARS-CoV and SARS-CoV-2 recognizes angiotensin-converting

CONTACT Insaf Ahmed Qureshi insaf@uohyd.ac.in Department of Biotechnology & Bioinformatics School of Life Sciences, University of Hyderabad, Hyderabad, Telangana 500046, India.

enzyme 2 (ACE2) for binding to viral S protein (Hoffmann et al., 2020).

As the receptor-binding domain (RBD) of S1 subunit of S protein binds to the ACE2 receptor of host cells (Lan et al., 2020), specific antibodies resulting from the immunization of RBD could successfully inhibit the viral invasion by blocking this interaction. Hence, the RBD domain has been exploited for generation of vaccines against MERS-CoV and SARS-CoV (Zhu et al., 2013). The RBDs of the SARS-CoV-2 and SARS-CoV possess identical structural architecture and significant sequence homology with 75% identity and 83% similarity. In spite of possessing high sequence similarity within RBDs of SARS-CoV and SARS-CoV-2, 72 residues of the receptor-binding motif (RBM) in RBD show only 59% similarity, which might result in limited cross-reactivity in the neutralizing antibodies (mAbs) generated against the RBM of SARS-CoV (Chen et al., 2020). Simultaneously, Wang and colleagues (2020) have revealed that substitutions in key residues of RBD of SARS-CoV-2 are responsible for higher affinity with hACE2 than SARS-CoV RBD. It was also observed that antibodies targeting RBD of SARS-CoV did not interact with the S protein of SARS-CoV-2, which suggests remarkable variances in the antigenicity pertaining to SARS-CoV and SARS-CoV-2 that could be further explored for development of therapeutics against the emerging virus.

In present study, we have employed the immunoinformatics based methods to construct an immunogenic multi-epitope vaccine from the RBD sequence of SARS-CoV-2 spike protein. The vaccine encompassed helper T lymphocyte, cytotoxic T lymphocyte, and B-cell epitopes from the RBD domain including RBM region. Simultaneously, human beta defensin-2 (hBD-2) was incorporated into multi-epitope vaccine as an adjuvant to reinforce the immunogenic response. Furthermore, molecular docking and simulation studies were used to validate the binding of designed vaccine candidate to TLR3 receptor of human host to provide a novel and efficient vaccine candidate against deadly COVID-19.

2. Materials and methods

2.1. Protein sequence retrieval

Amino acid sequence of the receptor-binding domain (RBD) of S protein (PDB: 6M17_F) was retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/) in FASTA format and then subjected to BLASTp tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) for evaluation of cross reactivity. Further analysis of exomembrane, physiochemical, antigenic and allergenic properties was done using online servers TMHMM v2.0 (http://www.cbs.dtu.dk/services/TMHMM/), ProtParam (https://web.expasy.org/protparam/), VaxiJenv2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) and AllergenFP v1.0 (http://ddg-pharmfac.net/AllergenFP/), respectively. Subsequently, RBD sequence was used to design a suitable vaccine candidate against SARS-CoV-2 by employing the strategy represented in Figure 1.

2.2. Cytotoxic T lymphocyte (CTL) epitope prediction

The CTL epitopes of SARS-CoV-2 RBD were determined using the server, NetCTL v1.2 (http://www.cbs.dtu.dk/services/NetCTL/) for all the MHC-I supertypes (A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, B62) with 0.75 threshold value having 0.80 sensitivity and 0.97 specificity, while TAP transport efficiency and weight on C-terminal cleavage were kept as default parameters. Subsequently, the antigenicity and immunogenicity of these epitopes were evaluated with VaxiJen v2.0 server and IEDB MHC-I immunogenicity tool (http://tools.iedb.org/immunogenicity/), respectively. Simultaneously, AllergenFP v1.0 server and ToxinPred server (http://crdd.osdd.net/raghava/toxinpred/) were used to analyse the allergic and toxic properties, respectively. Further, MHC-I binding alleles for the selected 9-mer CTL epitopes with antigenic, immunogenic, non-allergenic and non-toxic properties were predicted by a Consensus method of IEDB MHC-I binding tool (http://tools.iedb.org/mhci/) with human as a MHC-I source species and percentile rank < 2.

2.3. Helper T lymphocyte (HTL) epitope prediction

Fifteen amino acids long HTL epitopes for RBD domain were predicted using Immune Epitopes and Analysis Resource (IEDB) server (http://tools.iedb.org/mhcii/). The NN_Align method was applied to calculate IC50 value and percentile rank for peptide interacting with the MHC-II molecules. Human was considered as a source species, while loci, HLA-DP, HLA- DQ and HLA-DR were used for analysis. The IC50 value <20 nM and percentile rank <2 were considered for prediction as these values correspond to higher binding affinity. Finally, Predicted HTL epitopes were screened according to their antigenicity, allergenicity and toxicity.

2.4. Interferon-γ (IFN-γ) inducing epitope prediction

Helper T-cells produce cytokines like IFN-γ (interferon-gamma) that leads to the activation of many immune cells including cytotoxic T-cells, making cytokine-stimulating HTL epitopes a critical factor for vaccine development. IFNepitope server (http://crdd.osdd.net/raghava/ifnepitope/) was applied to identify the HTL epitopes that can produce IFN-γ with SVM and Motif based hybrid method and IFN-γ versus Non-IFN-γ model.

2.5. Linear B lymphocyte (LBL) epitope prediction

Detection of B-cell epitopes is an essential part to develop epitope-based vaccine as they generate antibodies that elicit humoral immune response. ABCpred server (http://crdd.osdd.net/raghava/abcpred/) was employed to predict the epitopes followed by further evaluation using BepiPred server of IEDB (http://tools.iedb.org/bcell/help/#Bepipred). The epitopes that are common for both of the tools were selected for further analysis. As the B-cell epitope is present on the cell surface, exomembrane topology is considered as one of the essential parameter. The allergenicity, antigenicity, toxicity and transmembrane topology of the epitopes were evaluated by AllergenFP v1.0, VaxiJen v2.0,

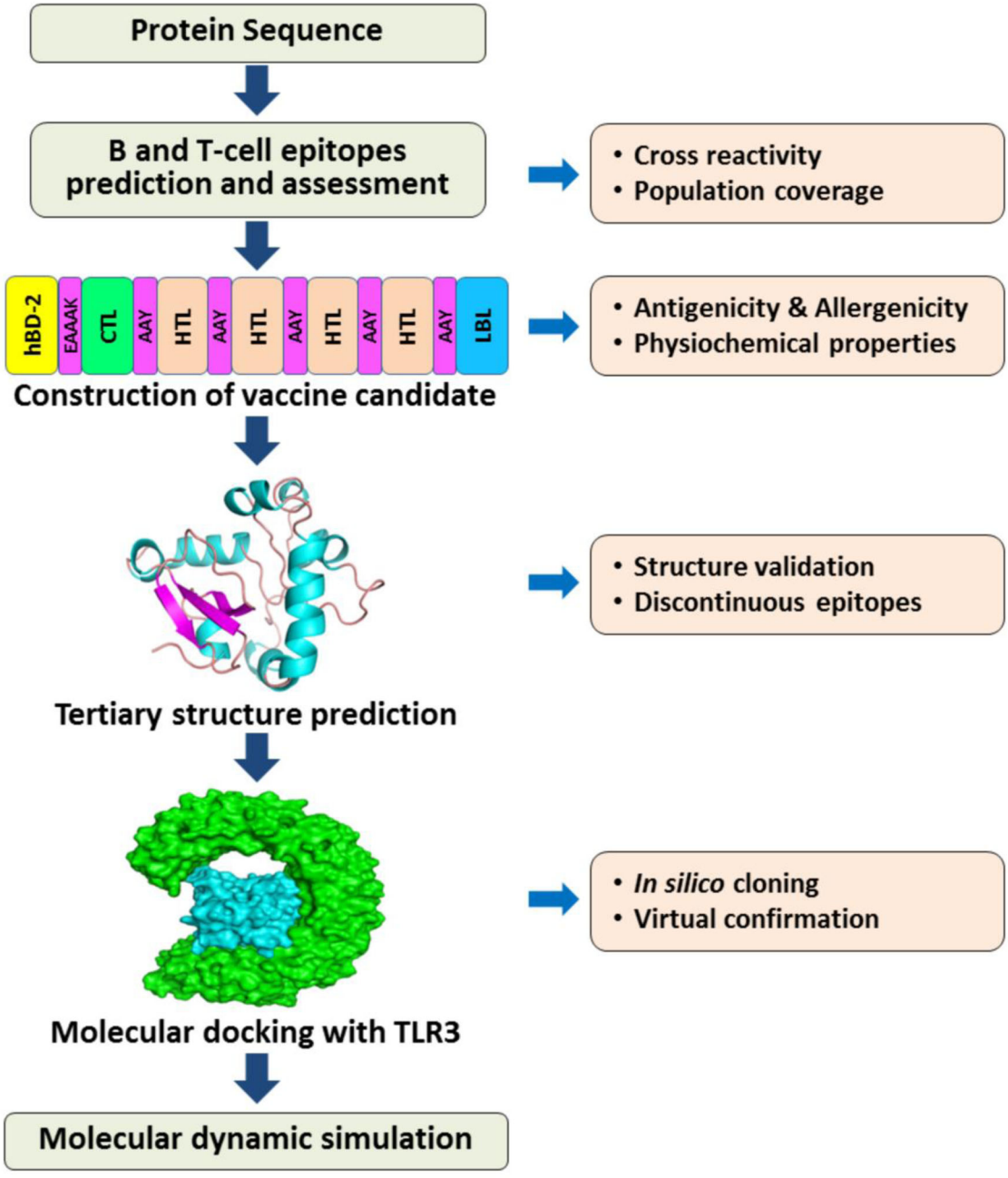

Figure 1. Schematic illustration to design the multi-epitope vaccine candidate with CTL, HTL and B-cell epitopes of SARS-CoV-2 RBD followed by molecular docking, dynamics simulation and *in silico* cloning.

ToxinPred and TMHMM v2.0 server, respectively. B-cell epitopes were also characterized by its surface accessibility, hydrophilicity, flexibility, and beta-turn by using Emini surface accessibility prediction tool, Parker hydrophilicity scale, Karplus & Schulz flexibility scale and Chou & Fasman beta-turn prediction, respectively (http://tools.iedb.org/bcell/).

2.6. Analysis of cross-reactivity and population coverage

To avoid the cross-reactivity, the predicted epitopes were employed to search against the database of human proteome from UniProt using Multiple Peptide Match tool of Protein Information Resource (https://research.bioinformatics.udel.edu/peptidematch/batchpeptidematch.jsp). Simultaneously, the Population Coverage tool of IEDB (http://tools.iedb.org/population/) was used to analyse the potential CTL and HTL epitopes as expression and distribution of its HLA alleles diverge within the ethnicities, regions and countries worldwide that affect the development of epitope-based vaccines. This tool utilizes the dissemination of human MHC binding alleles and provides an estimation of population coverage for epitopes in diverse regions throughout the world.

2.7. Construction of multi-epitope vaccine

The multi-epitope vaccine candidate was designed by amalgamating the selected CTL, HTL, and LBL epitopes with the

suitable linker (AAY) that contribute towards maximal flexibility of the amino acids to fold into appropriate conformations. It has been reported that the expression of human beta-defensin-2 (hBD-2) upregulates at the time of lung infection and increases innate and adaptive immunity (Kim et al., 2018). Hence, hBD-2 (PDB: 1FD3) was chosen as an adjuvant and incorporated at N-terminus to enhance the immune response of the vaccine construct.

2.8. Evaluation of antigenic, allergenic and physiochemical properties

To analyse antigenicity and non-allergenic nature, vaccine sequence was examined using VaxiJen v2.0 and AllergenFP v1.0, respectively. Antigenicity and allergenicity were further analysed with ANTIGENpro (http://scratch.proteomics.ics.uci.edu/) and AllerTOP v2.0 (https://www.ddg-pharmfac.net/AllerTOP/), respectively. Simultaneously, the vaccine protein sequence was subjected to ProtParam server to evaluate different physiochemical characteristics. Furthermore, ProteinSol server (https://protein-sol.manchester.ac.uk/) was employed to estimate the solubility of the vaccine candidate.

2.9. Secondary structure and solvent accessibility analysis

The secondary structural features were analyzed using the PSIPRED 4.0 server (http://bioinf.cs.ucl.ac.uk/psipred/), which employs two feed-forward neural networks for accurate prediction on the basis of position-specific scoring matrix generated by PSI-BLAST. Solvent accessibility of the vaccine candidate was estimated through the RaptorX property server (http://raptorx.uchicago.edu/StructurePropertyPred/predict/) that uses an emergent machine learning model titled DeepCNF (Deep Convolutional Neural Fields) for solvent accessibility prediction.

2.10. Tertiary structure prediction and validation

The prediction of tertiary structure for multi-epitope vaccine was accomplished by iterative threading modelling method on I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/). Initially, the server conducts search using LOMET (Local Meta-Threading Server) for suitable structural templates against the given protein sequence. Afterward, iterative template-based fragment assembly simulation was performed to create a full atomic structure of query. Finally, five full atomic models of the query sequence along with their respective C-score and TM-score were generated. The best model was further refined by a two-step procedure using GalaxyRefine server. Firstly, GalaxyLoop server (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=LOOP) was implemented for the loop modelling of predicted structure and then subjected to the GalaxyRefine tool (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE) for the refinement. GalaxyRefine server rebuild or re-adjust the side chain rotamers of protein structure followed by the mild and aggressive relaxation steps for the removal of clashes and bad contacts by performing the molecular dynamics simulation. The stereo-chemical qualities of models were validated by Ramachandran plot obtained from PROCHECK server (https://servicesn.mbi.ucla.edu/PROCHECK/). The overall quality of the predicted structure was validated through Z-score from ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php). Simultaneously, the statistics of non-bonded atom-atom interactions was evaluated by ERRAT server (https://servicesn.mbi.ucla.edu/ERRAT/).

2.11. Screening for conformational B lymphocyte epitopes

In order to predict the conformational B-cell epitopes, resulted vaccine protein was subjected to IEDB Ellipro tool (http://tools.iedb.org/ellipro/) with default setting. Ellipro uses the structure based method and predicts the antibody epitopes by estimating shape, neighbouring residue and protrusion index (PI) of the protein.

2.12. Molecular docking analysis

Toll-like receptors (TLR 2, 3, and 4) are integral membrane proteins which express on the sentinel cells of innate immunity and generate antiviral response. The structural coordinates of the TLR 2, 3, and 4 were retrieved from the Protein Data Bank (https://www.rcsb.org/) using respective PDB IDs: 2Z7X, 2A0Z and 3FXI. The TLR 2, 3, and 4 were used as receptor molecules for the docking with vaccine protein as ligand using ClusPro 2.0 server (https://cluspro.bu.edu/publications.php) that is based on PIPER, a Fast Fourier Transform (FFT) correlation approach for protein docking with pairwise interaction potential (Kozakov et al., 2017). Each docking output had thirty models which were generated from the three sequential steps including rigid body docking, clustering of lowest energy structure, and structural refinement. The docked structures were visualized through PyMol (http://www.pymol.org) to analyse the interactions between vaccine and TLRs. The complex with better center and lowest energy as well as high number of epitope interacting residues was further carried for molecular dynamics (MD) simulation study.

2.13. Molecular dynamics simulation

The MD simulations were performed for the apo form of vaccine construct and TLR3 as well as the complex of TLR3-vaccine construct by the GROMACS 5.1.4 version (Abraham et al., 2015) using the CHARMM36-AA force field (Best et al., 2012) and TIP3P water model (Jorgensen et al., 1996). Initially, suitable counter ions were added to neutralize the system followed by usage of the steepest descent algorithm to minimize system energy in 50000 steps. After the minimization, 1 ns each NVT (at 300 K) and NPT (at 1 bar) ensemble equilibration was performed for each system. Finally, the equilibrated systems were provided to the production simulation of 100 ns. The production simulation data was further used for the enumeration of root mean square deviation (RMSD), root mean square fluctuations (RMSF), radius of gyration (Rg) and hydrogen bonds for each system. The binding energy of the complex was calculated by the g_MMPBSA

package (Kumari et al., 2014) of GROMACS for every 0.1 ns frame of the last 50 ns MD simulation trajectories of the complex. Simultaneously, the rigidity and deformability in each residue of the complex were determined by the normal mode analysis (NMA) through the iMODS server (López-Blanco et al., 2014), which provides deformability, B-factors, eigenvalues and variance plot.

2.14. Codon optimization and in silico cloning

Java Codon Adaptation Tool (JCat) server (http://www.jcat.de/) was employed for maximum expression of vaccine protein in an extensively used bacterial system, *E. coli* strain K12. The options like avoid the prokaryote ribosome binding site, rho-independent transcription terminator and restriction enzyme cleavage sites were selected during the analysis. In order to find the cleavage site of commercial restriction enzymes, the nucleotide sequence obtained from the JCat was evaluated using NEBcutter (http://nc2.neb.com/NEBcutter2/). Further, the restriction sites of *Nde* I and *Xho* I were incorporated in optimized vaccine sequence at N- and C-terminal sites, respectively. Finally, the optimized vaccine DNA sequence was cloned *in silico* into the multiple cloning site (MCS) of pET-28a(+) plasmid vector followed by virtual confirmation using the SnapGene tool (https://www.snapgene.com/).

3. Results

3.1. SARS-CoV-2 RBD for vaccine construction

The amino acid sequence of the RBD region of spike protein was obtained from the NCBI database in FASTA format. BLASTp analysis revealed that no homolog of RBD is present in humans and mouse, while TMHMM v2.0 server predicted the RBD region to be localized outside the cell membrane suggesting that it is readily accessible to cellular as well as humoral immune systems. The RBD of SARS-CoV-2 contains 223 amino acids with approximately 25 kDa molecular weight. Further analysis showed its antigenic nature with the values of 0.52 calculated by VaxiJen v2.0. It was also found to be non-allergenic, hence chosen as a target for vaccine designing against SARS-CoV-2.

3.2. CTL epitope prediction and assessment

The CTL epitopes were identified from RBD by the NetCTL 1.2 server. A total of 83 CTL epitopes (9-mer) were predicted from all the MHC-I supertypes. After further assessment, 14 epitopes were observed to be antigenic, immunogenic and non-toxic. Among them, only 4 CTL epitopes (VRFPNITNL, YQPYRVVVL, PYRVVVLSF and VLSFELLHA) were non-allergenic, which were analysed for the MHC-I binding alleles (Table 1). In spite of binding with an appropriate number of alleles and showing good immunogenicity value, three (YQPYRVVVL, PYRVVVLSF and VLSFELLHA) out of four epitopes were not considered for further study as they overlap with selected HTL and B-cell epitopes.

3.3. HTL epitope prediction and evaluation

The HTL epitopes were evaluated for HLA-DR locus based on the percentile rank and IC50 with values <2 and <20 nM, respectively. In total, 39 HTL epitopes (15-mer) were obtained with corresponding MHC-II binding molecules. Out of 39 HTL epitopes, 14 were found to be antigenic, non-allergenic and non-toxic. These 14 HTL epitopes were also assessed for HLA-DP and HLA-DR locus. Additionally, the resultant epitopes were analysed for their IFN-γ inducing ability and only 9 HTL epitopes were found positive for induction of IFN-γ cytokine (Table 2). Finally, total 4 HTL epitopes (RFASVYAWNRKRISN, GCVIAWNSNNLDSKV, EGFNCYFPLQSYGFQ and RVVVLSFELLHAPAT) out of 9 were selected for construction of vaccine candidate. The "RVVVLSFELLHAPAT" epitope was chosen from four overlapping HTL epitopes. Similarly, "RFASVYAWNRKRISN" epitope was preferred over three overlapping HTL epitopes to reduce the size of the final vaccine sequence. It has also been observed that selected epitopes provide high population coverage and antigenic value to the vaccine construct in comparison to other overlapping HTL epitopes.

3.4. Linear B-cell epitope prediction and analysis

To construct a vaccine candidate, ABCpred server was employed to obtain the linear B-cell epitopes in the receptor-binding domain of SARS-CoV-2. In total, 23 epitopes showing 0.5 or above score with 16-mer lengths were analysed further with BepiPred server and five epitopes were found common in both of the servers. Successively, these five epitopes were characterized based on antigenicity, allergenicity, toxicity, exomembrane topology, surface accessibility, hydrophilicity, flexibility, and beta turn.

Out of five, 3 epitopes were found to be antigenic, non-allergenic and non-toxic, whereas only two epitopes possess exomembrane property that is essential for epitope recognition by B-cells. The B-cell epitope should be surface accessible, flexible and possess the hydrophilic regions. The average scores for Emini surface accessibility, Karplus and Schulz flexibility and Parker hydrophilicity were found to be 1.000, 0.992 and 1.393, respectively. Simultaneously, beta-turns were predicted through Chou and Fasman algorithm as betaturns are known to be hydrophilic and contribute towards instigating the immune response (Rose et al., 1985). The computed scores for beta-turn were 1.039 (average), 0.694 (minimum), and 1.397 (maximum). The epitopes with greater than the average values of the tools were considered as potently surface accessible, flexible, hydrophilic, and beta-turn carriers (Figure 2(A–D)), which could readily elicit B-cell immune response. Among five, only one epitope (LQSYGFQPTNGVGYQP) was found to fulfil all parameters required for being a potential linear B-cell epitope as well as located in the RBD motif region (Table 3).

3.5. Cross-reactivity and population coverage study

Cross-reaction evaluation against human proteome showed that all the six selected RBD epitopes have no human homolog, hence no cross-reactivity should arise in normal human cells. As SARS-CoV-2 is a global pandemic, the epitopes used in vaccine construction should also cover the maximum allele population. For estimation of population coverage, selected 1 CTL and 4 HTL with

Table 1. Cytotoxic T lymphocyte (CTL) epitopes prediction from RBD of SARS-CoV-2.

Peptide	Position	Antigenicity	Allergenicity	Immunogenicity	Toxicity	No. of alleles bind (Percentile rank < 2)
VLSFELLHA	194–202	Antigen	Non-Allergen	0.1607	Non-toxic	Nil
YQPYRVVVL	187–195	Antigen	Non-Allergen	0.1409	Non-toxic	HLA-B*48:01 HLA-B*39:01 HLA-B*15:01 HLA-A*02:06
PYRVVVLSF	189–197	Antigen	Non-Allergen	0.0313	Non-toxic	HLA-A*23:01 HLA-A*24:02 HLA-C*14:02 HLA-C*07:02
VRFPNITNL	9–17	Antigen	Non-Allergen	0.1748	Non-toxic	HLA-C*06:02 HLA-C*07:01 HLA-C*07:02 HLA-B*27:05 HLA-B*38:01 HLA-B*14:02

Table 2. Evaluation of SARS-CoV-2 RBD helper T lymphocyte (HTL) epitopes.

HTL epitope	Position	Alleles	Antigenicity	Allergenicity	IFN-γ	Toxicity
VLSFELLHAPATVCG	194–208	HLA-DRB1*01:01 HLA-DRB1*10:01	Antigen	Non-allergen	Positive	Non-toxic
RVVVLSFELLHAPAT	191–205	HLA-DRB1*01:01 HLADPA1*03:01/DPB1*04:02 HLADPA1*01:03/DPB1*06:01	Antigen	Non-allergen	Positive	Non-toxic
VVLSFELLHAPATVC	193–207	HLA-DRB1*01:01 HLA-DRB1*10:01	Antigen	Non-allergen	Positive	Non-toxic
VVVLSFELLHAPATV	192–206	HLA-DRB1*01:01 HLA-DRB1*10:01	Antigen	Non-allergen	Positive	Non-toxic
LSFELLHAPATVCGP	195–209	HLA-DRB1*01:01 HLA-DRB1*10:01	Antigen	Non-allergen	Negative	Non-toxic
TRFASVYAWNRKRIS	27–41	HLA-DRB5*01:01 HLA-DRB1*11:01 HLA-DRB1*13:01 HLA-DRB4*01:03	Antigen	Non-allergen	Positive	Non-toxic
RFASVYAWNRKRISN	28–42	HLA-DRB5*01:01 HLA-DRB1*11:01 HLA-DRB1*13:01 HLA-DRB4*01:03	Antigen	Non-allergen	Positive	Non-toxic
NATRFASVYAWNRKR	25–39	HLA-DRB5*01:01	Antigen	Non-allergen	Positive	Non-toxic
TESIVRFPNITNLCP	5–19	HLA-DRB1*15:01	Antigen	Non-allergen	Negative	Non-toxic
KSNLKPFERDISTEI	140–154	HLA-DRB3*01:01	Antigen	Non-allergen	Negative	Non-toxic
GCVIAWNSNNLDSKV	113–127	HLA-DRB3*02:02	Antigen	Non-allergen	Positive	Non-toxic
TGCVIAWNSNNLDSK	112–126	HLA-DRB3*02:02	Antigen	Non-allergen	Negative	Non-toxic
ESIVRFPNITNLCPF	6–20	HLA-DRB1*15:01	Antigen	Non-allergen	Negative	Non-toxic
EGFNCYFPLQSYGFQ	166–180	HLA-DRB1*10:01	Antigen	Non-allergen	Positive	Non-toxic

their corresponding HLA alleles were used and they individually showed 55.88% and 83.01% of world population coverage, respectively. Notably, combination of CTL and HTL epitopes resulted in 92.51% population coverage globally. Since our vaccine protein contains both types of epitopes (MHC-I & MHC-II), population coverage analysis by using the combined method is important. Maximum population coverage was found to be 99.77% for North American countries, while it was observed to be 94.33% for India. As the SARS-CoV-2 was likely originated in China from where it spread all over the world, the population coverage calculation for China is crucial for designing of vaccine. Remarkably, population coverage for China was 84.05% (Figure 3).

3.6. Construction of multi-epitope vaccine

In order to design the multi-epitope vaccine against SARS-CoV-2, a total of six (1 CTL, 4 HTL, 1 LBL) epitopes were joined by using proper linkers. The linkers can prevent junction epitopes formation, which arises as a major concern for constructing a multi-epitope vaccine. To enhance the immunogenecity of the vaccine construct, human β-defensin-2 (hBD-2, 41 amino acids) was utilized as an adjuvant at N-terminus of proposed vaccine. The adjuvant was conjugated to the first CTL epitope with the EAAAK linker to minimize its interaction with different sections of the vaccine (Figure 4(A)). Altogether, the final vaccine construct encompasses 146 amino acids containing an adjuvant and six RBD epitopes connected by the linkers.

3.7. Prediction of antigenicity, allergenicity, and physicochemical parameters

The antigenicity, allergenicity and other parameters of vaccine construct were analysed using various servers. The constructed multi-epitope vaccine has a good antigenic property with a value of 0.70 and 0.84 predicted by VaxiJen v2.0 and ANTIGENpro respectively. In addition, online tools also

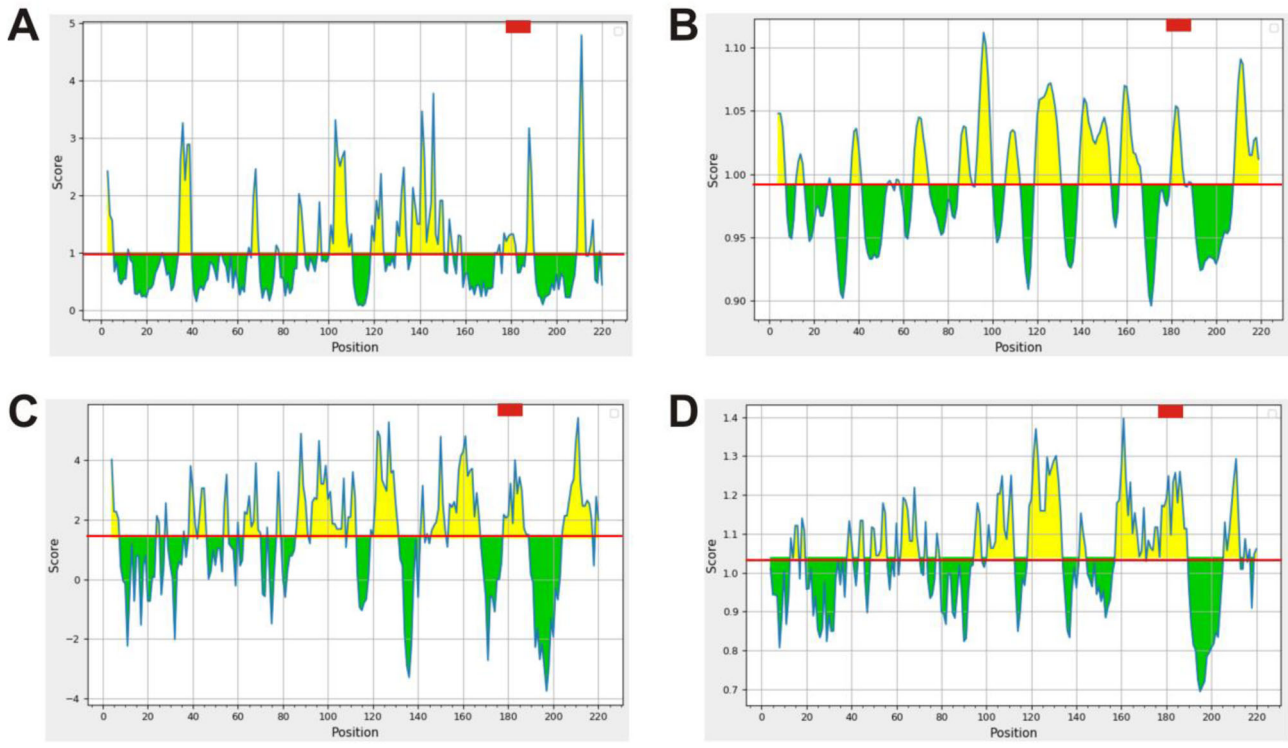

Figure 2. (A) Emini surface accessibility, (B) Karplus and Schulz flexibility, (C) Parker hydrophilicity and (D) Chou and Fasman beta-turn prediction, where X- and Y-axis represent the sequence position and respective scores. The line represents the average values of the respective calculated scores and the yellow regions above the average are considered as good, while the selected peptide region (LQSYGFQPTNGVGYQP) marked as red.

Table 3. Assessment of linear B lymphocyte (LBL) epitopes from SARS-CoV-2 RBD.

B-cell epitope	Position	BepiPred	Antigenicity	Allergenicity	Toxicity	Transmembrane topology
LQSYGFQPTNGVGYQP	174–189	Yes	Antigen	Non-allergen	Non-toxic	Outside
TEIYQAGSTPCNGVEG	152–167	Yes	Non-antigen	Non-allergen	Non-toxic	Inside
EVRQIAPGQTGKIADY	88–103	Yes	Antigen	Non-allergen	Non-toxic	Inside
TGKIADYNYKLPDDFT	97–112	Yes	Antigen	Non-allergen	Non-toxic	Inside
YKLPDDFTGCVIAWNS	105–120	Yes	Non-antigen	Allergen	Non-toxic	Outside

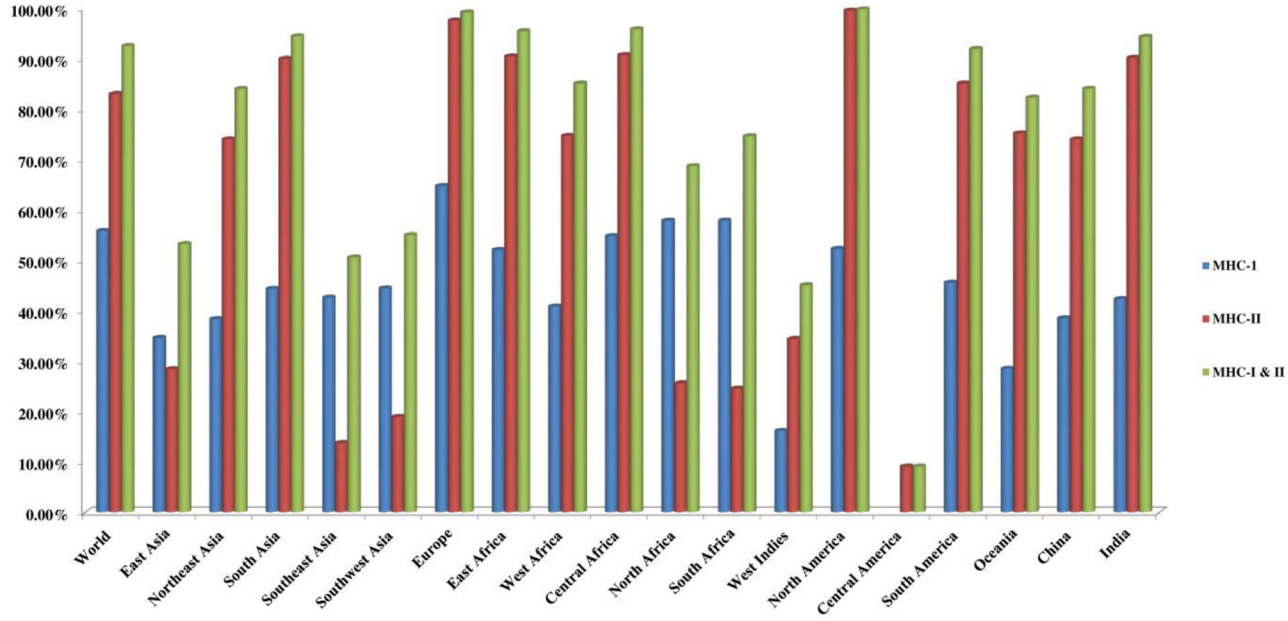

Figure 3. Worldwide population coverage by T-cell epitopes of multi-epitope vaccine on the basis of their respective HLA binding alleles with their MHC-I (blue), MHC-II (red) and combined MHC-I & II (green) coverage rate.

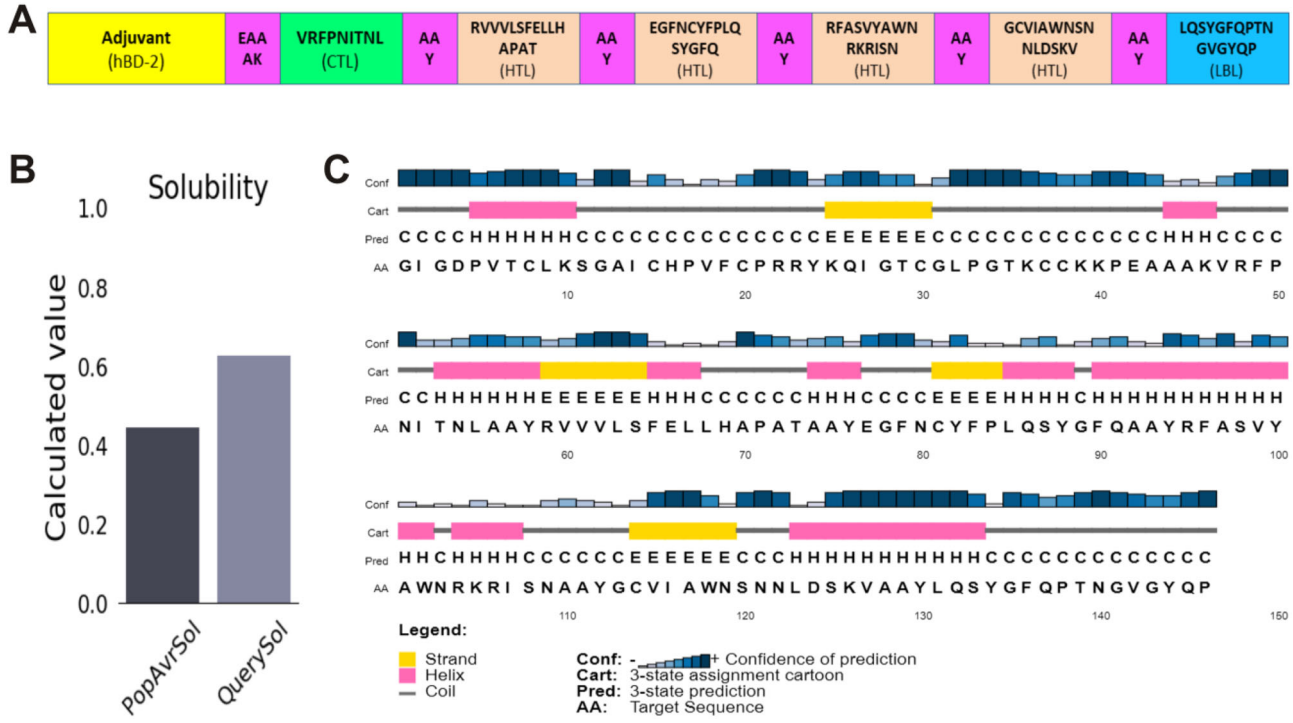

Figure 4. (A) Designing of multi-epitope vaccine from RBD domain of SARS-CoV-2. The CTL, HTL and B-cell epitopes are presented as box with green, beige and blue color, respectively, while adjuvant and linkers are represented in yellow and magenta color, respectively. (B) Solubility and (C) Secondary structure analysis of vaccine construct using Protein-Sol and PSIPRED server, respectively.

suggested that vaccine construct is non-allergenic in nature. Subsequently, various physicochemical characteristics were calculated using ProtParam server. The vaccine construct depicted to possess the molecular weight of 15.99 kDa that suggests its good antigenic nature. The theoretical isoelectric point (pI) was 9.35 indicating towards the basic nature of vaccine construct. The instability index was calculated to be 37.31 advocating for the stable nature of the construct after expression, while the aliphatic index (72.95) represented the vaccine candidate to be highly thermostable. The grand average of hydropathicity (GRAVY) denotes the amphipathic nature of the proteins with positive and negative value representing the hydrophobic and hydrophilic nature of the amino acid side chains, respectively. The GRAVY index of vaccine construct was found to be −0.05 that indicates hydrophilic nature of the vaccine suggesting better interaction with the water molecules. The estimated half-life of the vaccine protein in mammalian reticulocyte (*in vitro*) was observed as 30 h, while it was >20 h and >10 h for yeast and *Escherichia coli* (*in vivo*), respectively. Additionally, the solubility prediction showed a higher solubility rate with a score of 0.63 (Figure 4(B)).

3.8. Secondary structure and solvent accessibility prediction

The secondary structure of the vaccine construct was obtained using PSIPRED v4.0 workbench. The secondary structure analysis of 146 amino acid long vaccine construct demonstrated the presence of 36.3% α-helices, 15.06% β-strands and 48.63% random coils (Figure 4(C)). Solvent accessibility analysis by RaptorX server showed that 44% of residues were exposed, whereas 21% and 34% residues to be medium exposed and buried, respectively.

3.9. Tertiary structure prediction and validation

The sequence of designed vaccine candidate containing 146 amino acids was used to predict the three-dimensional structure using I-TASSER, which provided five structures based on 10 threading templates. Out of 10 templates, 6 templates showed good alignment according to their Z-score that was in the range of 1.12 − 3.27 and the five predicted models were represented by C-score values ranging from −4.44 to −3.62. The C-score typically ranges from −5 to 2 with higher C-score representing higher confidence level for structure prediction. The model showing the C-score −3.78 was selected for further evaluation. The analysis of model via Ramachandran plot indicated that 35.5, 42.7, 13.7 and 8.1% of residues were in favoured, additionally allowed, generously allowed and disallowed regions, respectively. Hence, the selected model was subjected to two step refinement using GalaxyLoop and GalaxyRefine server for model selected from I-TASSER. Out of five different 3D refined models, the best structure (Figure 5(A)) was selected on the basis of quality assessment. The Ramachandran plot suggested that most of the residues were found in the favoured (91.1%) and allowed (8.1%) regions, while only one residue (Val18) was present in the disallowed region of refined modelled structure (Figure 5(B)). Finally, quality-verification and potential error of the 3D vaccine model were evaluated through

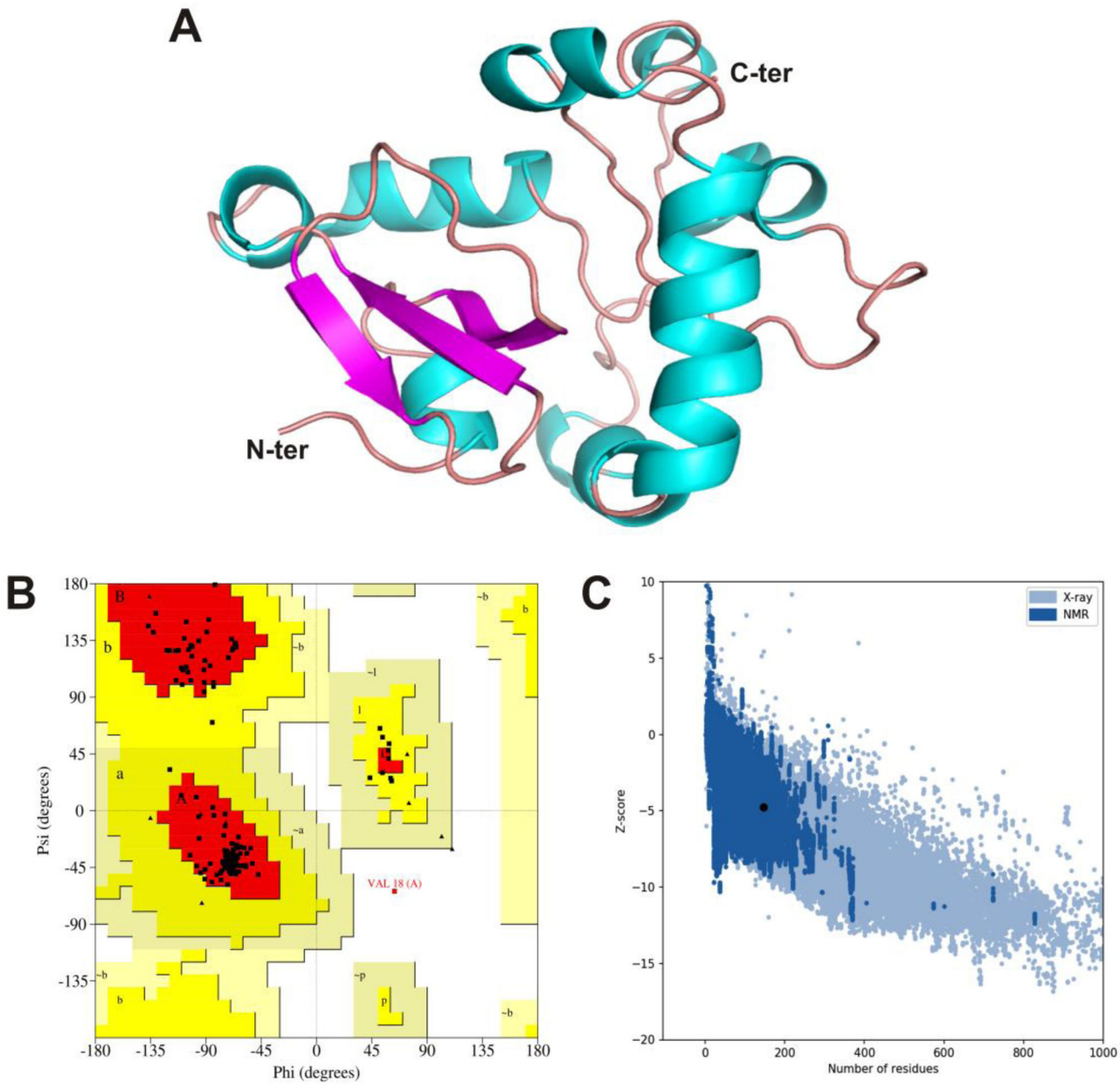

Figure 5. (A) Modelled three-dimensional structure of the multi-epitope vaccine showing α-helices (cyan), β-sheets (magenta) and random coils (red). (B) Ramachandran plot indicating the presence of amino acid residues in favoured, allowed and disallowed region. (B) ProSA-web validation presenting the Z-score of predicted 3D structure.

ProSA-web and ERRAT plot, respectively. The modelled structure after refinement displayed an overall quality factor of 79.71% with ERRAT, whereas ProSA-web depicted a Z-score of −4.78 for the vaccine structure (Figure 5(C)).

3.10. Conformational B-cell epitope prediction

Ellipro tool was used for conformational B-cell epitopes evaluation in multi-epitope vaccine. A total of 86 residues with scores ranging from 0.569 to 0.771 were observed, which were present within four conformational B-cell epitopes with size varying from 12 to 37 residues (Table 4). The top scorer epitope (containing residues Y24, K25, K40, P41, E42, A43, A44, A45, K46, V47, R48, F49, P50, N51, N54) with

Table 4. Conformational B-cell epitopes from vaccine protein using Ellipro server.

Residues	Number of residues	Score
Y24, K25, K40, P41, E42, A43, A44, A45, K46, V47, R48, F49, P50, N51, N54	15	0.771
H69, A70, P71, A72, T73, A74, A75, Y76, E77, F79, N80, C81, Y82, F83, P84, L85, Y88, F90, Q91, A92, Y94, R95	22	0.641
R22, F96, Y100, N103, R104, R106, I107, N109, A110, A111, Y112, N119, S120, N121, N122, L123, D124, S125, K126, A129, Y130, L131, Q132, S133, Y134, G135, F136, Q137, P138, T139, N140, G141, V142, G143, Y144, Q145, P146	37	0.632
G1, I2, G3, D4, T29, C30, G31, L32, P33, G34, Q86, S87	12	0.569

PI value (0.771) has highlighted the presence of 77.1% residues in the ellipsoid region.

3.11. Molecular docking

Since interaction of the vaccine candidate and target immune receptor is important to trigger an effective immune response, ClusPro v2.0 program was used for docking and predicting binding affinity of vaccine candidate with TLRs. A total of thirty docked complexes for each TLR bound vaccine construct were generated through the ClusPro and their respective center and lowest energy scores were calculated. The docking result of the TLR2-vaccine and TLR4-vaccine complexes showed lesser center and lowest energy scores in comparison to the TLR3-vaccine complex. Additionally, more residues from epitopic regions of vaccine candidate have shown interaction with TLR3 as compared to TLR2 and TLR4 receptors (Table 5). The lowest energy score of the finally selected model was −952.1, and the center energy score (the energy between the ligand and the receptor) was −897.0.

The interactions between the TLR3 receptor and vaccine candidate were visualized using Pymol and the docked complex is represented as surface mode in Figure 6. Residues of vaccine construct (Arg48, Glu77, Asn80, Ser87, Arg95, Asn103, Lys105, Arg106, Asn119, Ser120, Asn121, Asn122, Leu123) were found to make polar contacts with residues (Asp36, Ser38, Tyr326, His359, Asn361, Tyr383, Gln483, Arg484, Thr509, Asp512, Glu533, Asp536, Lys589, Ser614, Leu637, Thr638, Glu639) of TLR3 receptor.

3.12. Molecular dynamics simulation

To analyse the structural stability of the docked TLR3-vaccine complex, molecular dynamics simulations of 100 ns were performed for both complex and apo forms. The backbone atom RMSD of the TLR3-vaccine complex showed that initially mean deviation was between 0.2-0.4 nm (approximately up to 49 ns). However, the structural deviations in the complex got stabilized after 50 ns at a RMSD value of nearly 0.35 nm throughout the remaining simulation time with average RMSD of 0.39 nm (Figure 7(A)). The RMSD data analysis inferred that the interaction between the vaccine and TLR3 resulted in the stabilization of the complex structure. Further, the stabilization of vaccine structure after the interaction with TLR3 was analysed by comparing the backbone atom RMSD of the vaccine in the complex as well as in the apo form. It revealed that RMSD of the vaccine structure in the complex attains a lower deviation value (approximately ~0.3 nm) than the vaccine in apo form. The comparison indicated that the selected docked conformation of the vaccine candidate in the complex was structurally more stable after interacting with the TLR3 during the dynamics. Subsequently, root mean square fluctuation (RMSF) was calculated for the backbone atom of the residues in the complex that showed the average fluctuation in the complex residues to be 0.138 nm, with maximum fluctuations at the N and C terminus (Figure 7(B)). The analysis of the RMSF of complex inferred the lower fluctuation in the complex forming residues, which represents the stability of the interface between the vaccine and TLR3 during the dynamics. The RMSF for the backbone atoms of the vaccine residues presented a significant relaxation in the mean fluctuation of the residues side chain of the complex as compared to the apo form of the vaccine.

The compactness in the complex structure was determined by enumerating the Rg data of the complex, which showed that the gyration in the backbone atoms of the complex structure before 49 ns was unstable and varies from 3.10 nm to 3.15 nm. However, after 50 ns, the gyration in the structure started to attain stability at a value of approximately 3.14 nm in the remaining simulation time with average Rg of 3.12 nm (Figure 7(C)). The stabilization of the complex structure Rg at a lower value in comparison to the initial structure of the complex suggested gain in the stability of TLR3-vaccine complex structure compactness during the dynamics. The comparison of backbone atom Rg of the vaccine in complex and in apo form showed that Rg of the vaccine structure in the complex become more compact and stable (approximately at an Rg of ~1.45 nm) after interacting with TLR3 in comparison of the backbone atom Rg of its apo form. The analysis inferred the increase in compactness of the vaccine construct in the complex, which may aid in the stabilization of the complex during the dynamics. Calculation of hydrogen bonds between the TLR3 and vaccine in the complex defines the stability of interface in the TLR3-vaccine complex during simulation. The analysis showed that an average of 13 hydrogen bonds were maintained at the interface up to ~49 ns, which increased up to 17 after the stabilization (approximately at ~50 ns) of the complex (Figure 7(D)). The increase in hydrogen bonds represents the stabilization of the interface between the TLR3 and vaccine in the complex.

Furthermore, the deformability and mobility stiffness of the residues in the complex were analyzed by enumerating normal mode analysis (NMA) of its internal coordinates through iMODS. Analysis showed low distortion in the residues of the complex in the deformability plot (Figure 8(A)), and the B-factor values computed by the NMA were less deviated than the values of complex (Figure 8(B)). Simultaneously, eigenvalue for the complex was 3.842825e-05 and it increased gradually in each mode during the dynamics (Figure 8(C)). The variance plot depicted a gradual decrease of the individual variance in each successive mode (Figure 8(D)). Overall analysis of iMODS shows more stiffness and low deformation in the internal coordinates of the complex during the dynamics, which also advocates towards the stability of the interface between the TLR3-vaccine complex.

The average binding energy of the complex was found to be -2031.27 ± 4.19 kJ mol^{-1} using MMPBSA tool of the GROMACS, whereas the electrostatic free energy (-3168.29 ± 7.18 kJ mol^{-1}) was the major contributing free energy in the binding of vaccine with the TLR3. The average binding energy of the complex during the dynamics was lower (-2031.27 kJ mol^{-1}) as compared to the binding energy of the initial complex structure (-1686.88 kJ mol^{-1}), which

Table 5. Energy scores of the docked complexes of the vaccine candidate with different toll-like immune receptors.

Complex	Center energy score	Lowest energy score	No. of residues from epitopic region
TLR2-Vaccine	−794.6	−847.7	8
TLR3-Vaccine	−897.0	−952.1	13
TLR4-Vaccine	−821.3	−846.8	10

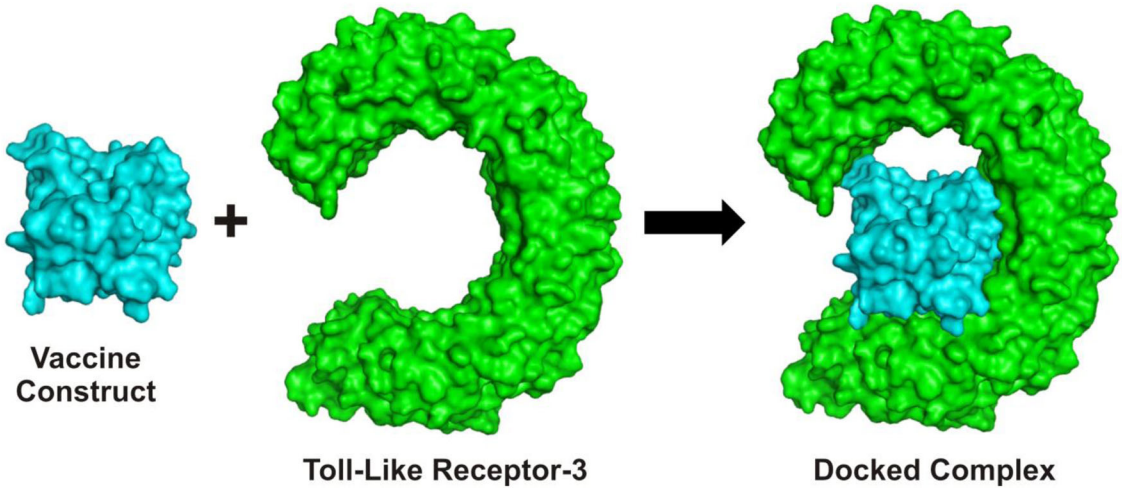

Figure 6. The docked complex of TLR3 and multi-epitope vaccine obtained from ClusPro is presented as surface and colored in green and cyan, respectively.

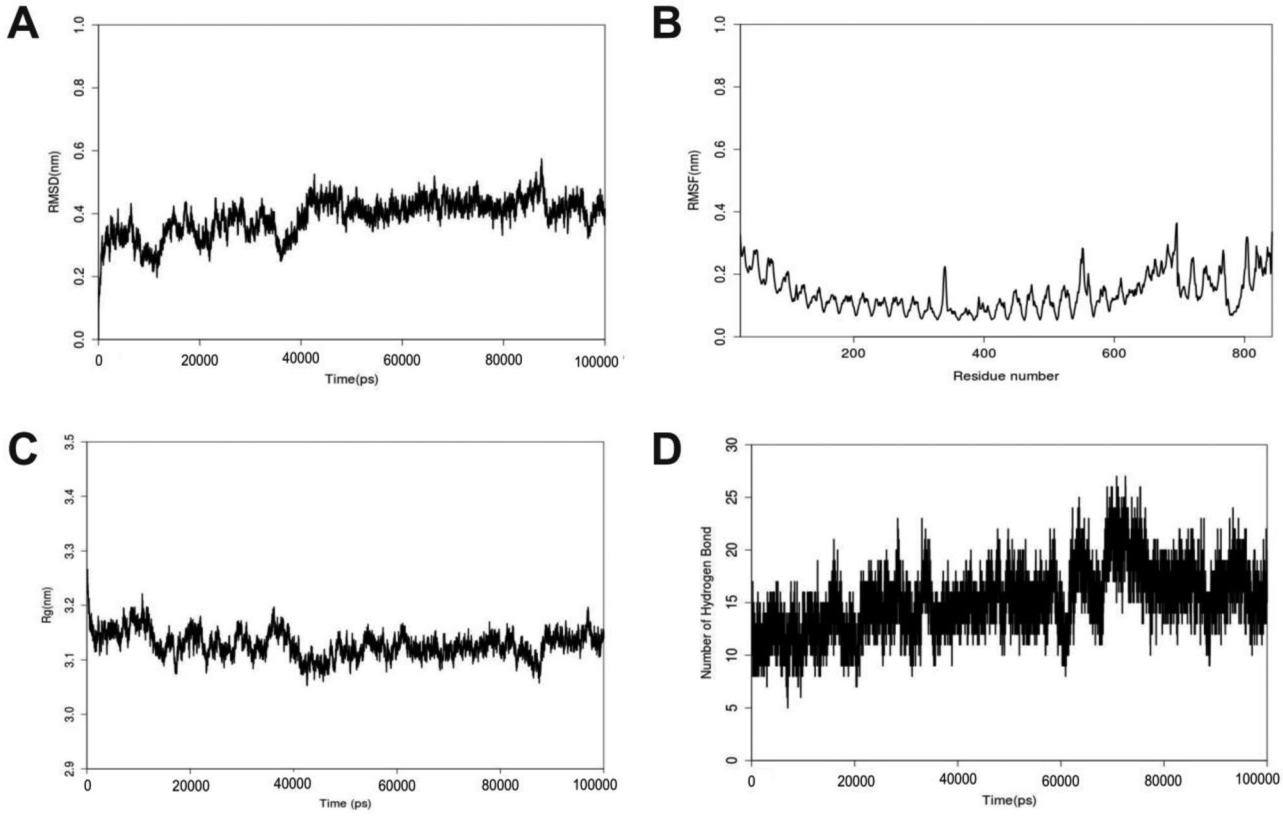

Figure 7. MD simulation analysis. (A) RMSD, (B) RMSF and (C) Rg of the backbone-atoms of the docked TLR3 and multi-epitope vaccine complex. (D) Number of hydrogen bonds formed during MD simulation.

suggests the increase in the binding affinity of the TLR3 and vaccine in the complex structure during the dynamics. Subsequently, the interactions between the TLR3 and the vaccine in the initial and stabilized complex structure (extracted from the trajectories) were calculated by Protein Interaction Calculator (PIC, Tina et al., 2007) and hydrogen bonds using GROMACS (Table 6).

Additionally, number of polar contacts in the initial and stabilized complex were analysed and visualized by Pymol (Figure 9(A–B)). It revealed an increment in the total number

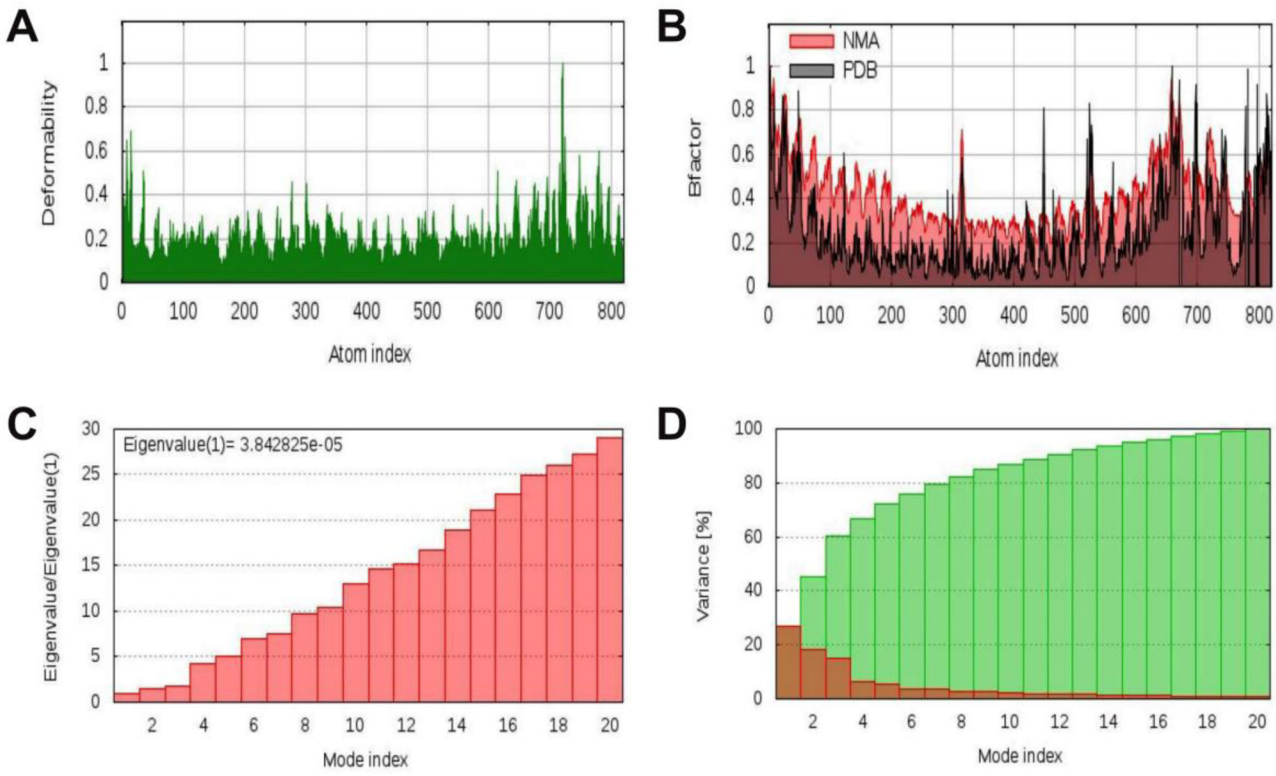

Figure 8. iMODS analysis of the TLR3-vaccine complex showing plots of (A) deformability, (B) B-factor, (C) eigenvalue and (D) NMA variance.

Table 6. Interactions of the vaccine-receptor complex during MDS study.

Interactions name	Number of interactions in initial complex	Number of interactions in stable complex
Hydrophobic	13	16
Hydrogen bond	15	25
Ionic	6	8
Aromatic-Aromatic	1	1
Cation-π	1	3

of interactions in the stable TLR3-vaccine complex than the initial one.

Interaction data of the stable complex was also corroborated by the binding free energy of the system and also shown the favourable increment in each free energy term of the stable complex as compared to the initial complex (Table 7).

3.13. Codon optimization and in silico cloning

The objective of codon optimization and *in silico* cloning is efficient cloning and over expression of vaccine protein in the bacterial system like *E. coli*. For codon optimization of vaccine protein, the Java Codon Adaption tool (JCat) was utilized. The optimized codon sequence was 438 nucleotides long, while Codon Adaptation Index (CAI) and GC content for optimized nucleotide sequence were 1.0 and 52.73%, respectively. These values were found within the optimum range of CAI (0.8–1.0) and GC (30–70%), representing a high probability of good expression of the multi-epitope vaccine candidate in the bacterial host.

The restriction sites *Nde* I and *Xho* I were incorporated at N- and C-terminus of codon-optimized DNA sequence followed by its insertion between *Nde* I and *Xho* I restriction sites within multiple cloning site (MCS) of *E. coli* expression vector, pET-28a(+) (Figure 10). His-tag/thrombin/T7-tag present at N-terminus of the pET-28a(+) plasmid makes it universally used vector for cloning and expression. Furthermore, virtual confirmation of clone using agarose gel simulation tool of SnapGene revealed presence of insert (~0.5 kb) along with vector (~5.5 kb) after digestion with *Nde* I and *Xho* I enzymes (Figure 11). The size of insert was found to be in accordance of the calculated molecular weight of vaccine candidate.

4. Discussion

The employment of immunoinformatics approaches is widely used for developing novel and potential multi-epitope vaccine candidate to combat many deadly diseases. Recently, immunoinformatics methods have been extensively applied to design multi-epitope based vaccine against several virulent pathogens including Dengue virus, Ebola virus etc. (Ahmad et al., 2019; Sabetian et al., 2019). Owing to the limited therapeutic measures and high mortality rate of COVID-19, serious efforts are being made worldwide to develop specific vaccine using full-length and particular components of viral proteins with limited success till now.

Spike proteins are exposed on the viral cell surface and considered as a promising candidate for developing vaccine against SARS associated coronaviruses. Recently, several groups have reported construction of multi-epitope vaccine against SARS-CoV-2 through immunoinformatics approach using various viral proteins. For instance, spike glycoprotein

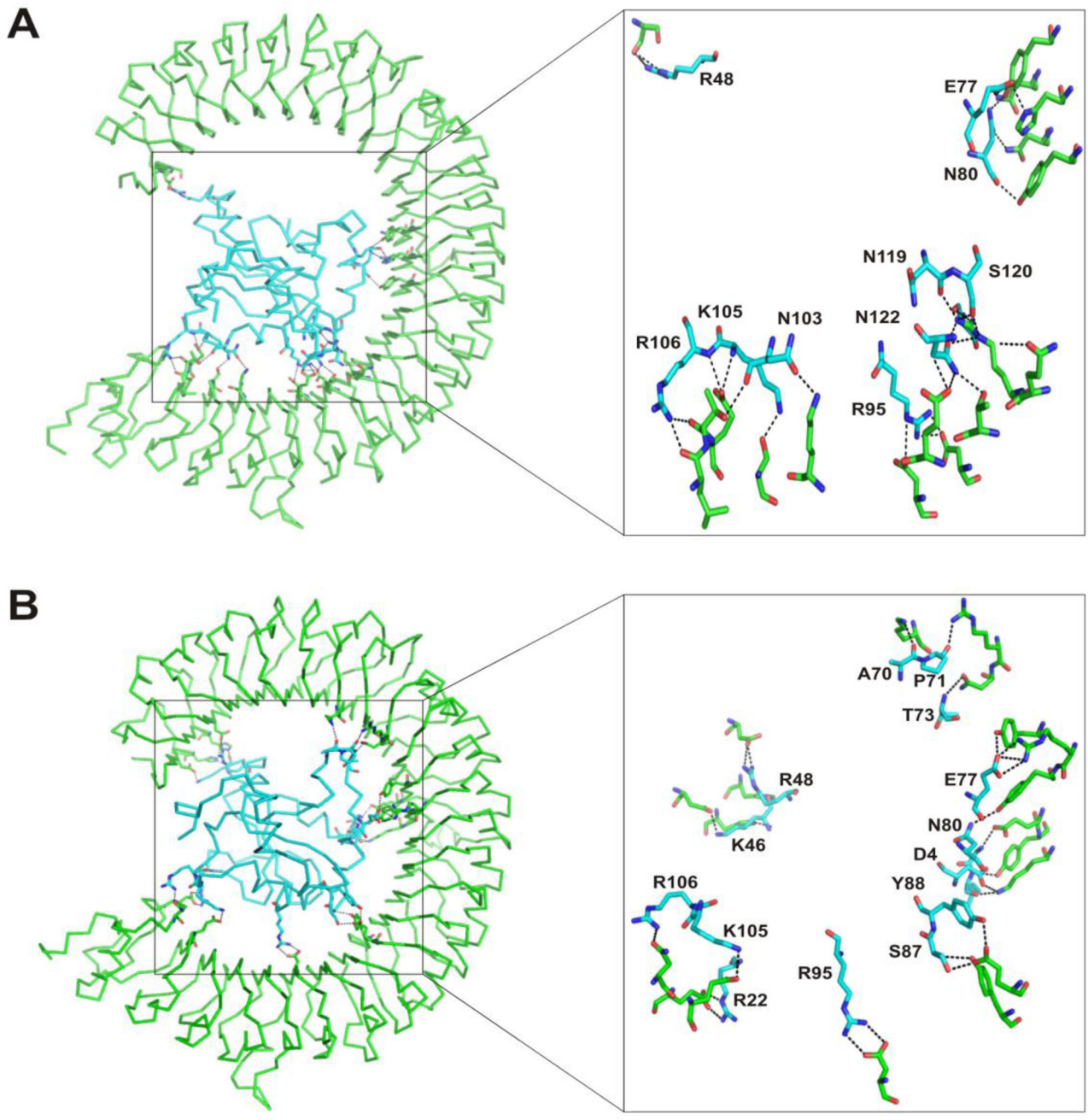

Figure 9. Important interactions of (A) the initial TLR3-vaccine complex structure is presented with (B) the stable TLR3-vaccine complex obtained from MD simulation. The TLR3 and vaccine construct are colored in green and cyan, respectively. The residues depicting interactions between TLR3 and vaccine candidate are shown as stick model, while vaccine construct residues are labelled and polar contacts are represented by black dashed lines.

Table 7. The Electrostatic, van der Waal, Polar solvation, SASA (solvent-accessible surface area) and Binding energy of the vaccine-receptor complex, kJ/mol, calculated by g_MMPBSA method.

	Electrostatic	Van der Waal	Polar solvation	SASA	Binding energy
Initial complex	−2627.95	−590.64	1603.63	−71.91	−1686.88
Stable complex	−3486.91	−488.52	2052.55	−74.86	−1997.75

of SARS-CoV-2 has been evaluated extensively to design different multi-epitope vaccines (Abraham Peele et al., 2020; Kar et al., 2020, Samad et al., 2020). Additionally, spike protein has also been used in combination of epitopes from other viral proteins including nucleocapsid and membrane (Kalita et al., 2020), protein 8 (Nsp8) and 3 C-like proteinase (Ahmad et al., 2020), main protease, Nsp12 (polymerase) and Nsp13 (helicase) (Rehman et al., 2020) for construction of peptide-based vaccines. These vaccine candidates were found to be antigenic, non-allergenic and stable in nature, but presence of additional epitopes from other region of S protein or other viral proteins might result into excessive antigenic load along with increased chances of allergenic response. RBD region of SARS-CoV-2 S protein plays a significant role in interaction with human ACE2 receptor that mediates entry of virus into the host cell and could be a valid target for the vaccine designing. Notably, the RBD region of the spike protein doesn't show any remarkable homology with human proteome, so chances of causing autoimmunity

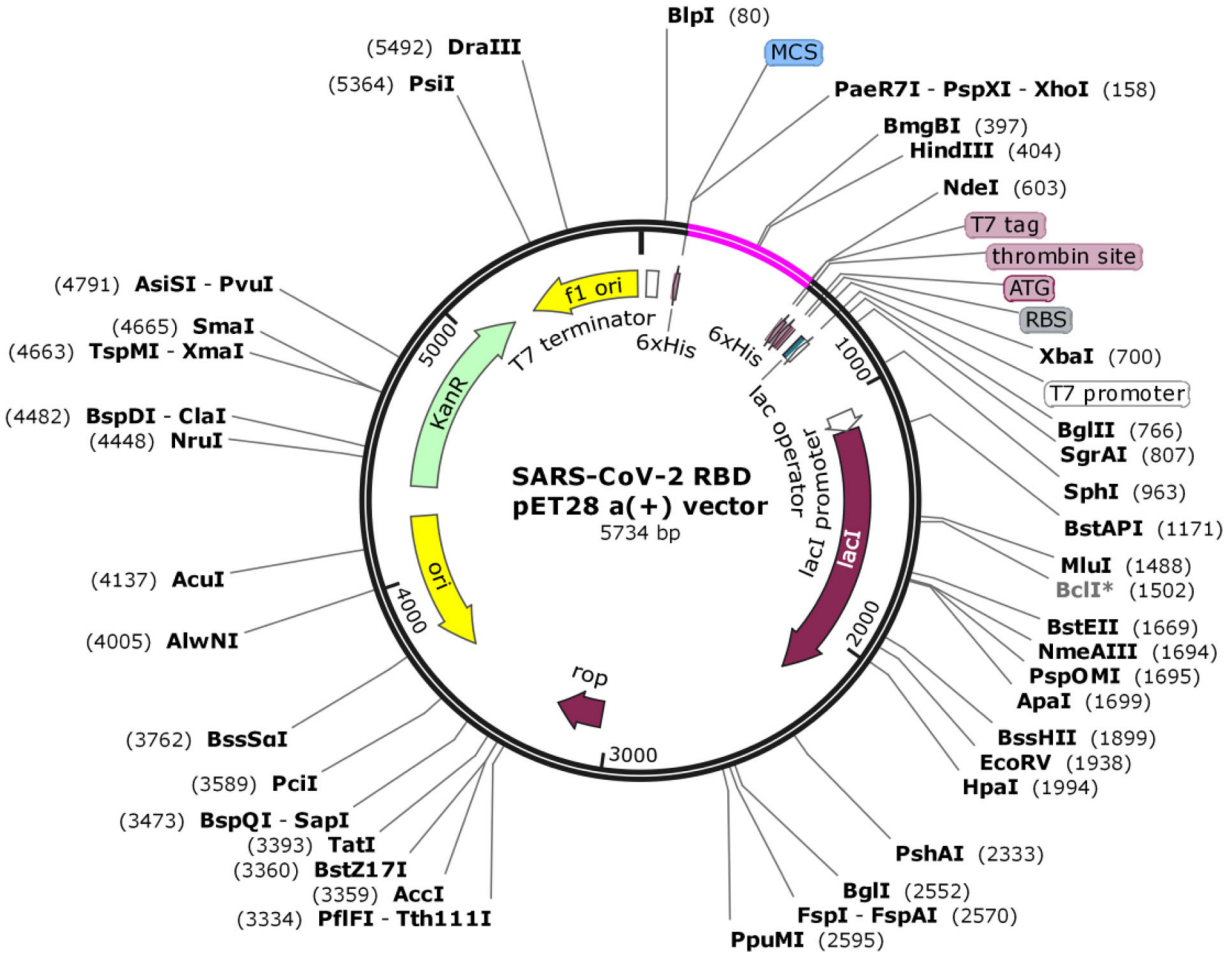

Figure 10. *In silico* cloning of the multi-epitope vaccine optimised codons (magenta color) into pET-28a(+) expression vector between *Nde* I and *Xho* I restriction sites.

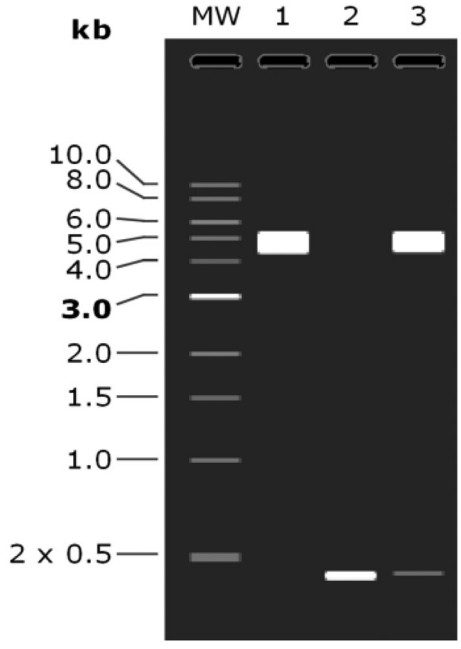

Figure 11. Virtual conformation of RBD vaccine clone by double digestion. Lane 1 represents native pET-28a(+) vector digested with *Nde* I and *Xho* I, Lane 2 indicates insert (Vaccine) and Lane 3 shows double digestion of vaccine construct (Vaccine + Vector).

is very low that is one of the major criteria to be considered during designing of potential vaccine. Antigenicity and non-allergenicity are other parameters that make this candidate suitable for vaccine construction as it represents effective binding with the host protein to stimulate the immune system without causing any allergenic reaction (Solanki et al., 2019). Hence, in the present study, we have utilized RBD including its RBM region of the spike protein to identify the immunogenic B- and T-cell epitopes by rigorous screening.

Various cells of immune cascade recognise the epitopes of viral protein and initiate activation of T-helper cells to release the required cytokines that leads to cytotoxic T-cells and B-cells instigation causing destruction of pathogens, therefore T-cell epitope mapping is essential for vaccine construction (Channappanavar et al., 2014). Altogether, one CTL and four HTL epitopes were selected on the basis of antigenic, immunogenic, non-allergenic, non-toxic and number of MHC-I & II binding allele properties. Additionally, IFN-γ inducing ability of HTL epitopes was evaluated as they recruit the neutrophils and augment chemokine production (McLoughlin et al., 2008). Significantly, one of the HTL epitopes selected by us has also been employed in a recent study for the construction of vaccine candidate using spike protein against SARS-CoV-2 (Abraham Peele et al., 2020). Additionally, B-cell epitope in vaccine can stimulate the cell-

mediated as well as humoral immune response. Hence, based upon several parameters including antigenicity, allergenicity, surface accessibility etc., one linear B-cell epitope was selected that also encompasses the T-cell epitope. Remarkably, several residues of the predicted HTL as well as linear B-cell epitopes participate during interaction with human ACE2 receptor that has been depicted in crystal structure of RBD-hACE2 complex (Wang et al., 2020).

The epitopes present on the vaccine should not show any homology with human proteins to prevent auto-immunity, so cross-reaction assessment is one of the critical parameters and we did not find any homolog of the selected epitopes in humans. Simultaneously, selected epitopes showed maximum population coverage including North American countries, India, China etc., which suggests that our vaccine construct could help to fight against COVID-19 globally. Consequently, six epitopes were selected to design multi-epitope vaccine by excluding overlapping epitopes to reduce additional antigenic load and allowing epitopic variability in vaccine candidate. Additionally, the selected epitopes were combined using suitable linkers that support the natural structure of epitopes and enhance the stability and folding, thus, ultimately increase the recognition of the multi-epitope vaccine (Chen et al., 2013). The human β-defensin-2 (hBD-2) was utilized as an adjuvant in vaccine construct to enhance the immunogenicity. It could also improve the delivery of vaccines by providing protection from degradation as well as delivering it to dendritic cells (Perrie et al., 2008). The vaccine construct showed high antigenicity (0.70) with adjuvant as compared to without adjuvant (0.62), suggesting that the adjuvant is playing significant role for vaccine chimera. In addition, solubility of vaccine is an important physiochemical property that provides basis to many biochemical and functional investigations. Computational analysis of our vaccine candidate depicted high solubility inside the bacterial expression system, which in turn would be useful for its easy purification and large scale production. The final vaccine candidate has shown good antigenic and non-allergenic properties along with stable nature after expression. Remarkably, overall idiosyncratic outputs for various characteristics advocate this multi-epitope vaccine to be a potent candidate against SARS-CoV-2. Mostly residues (44%) of vaccine construct were found to be accessible to solvent, which is in accordance to the earlier report that established a positive correlation between the exposed and solvent accessible vaccine protein residues with proficient cellular and humoral immune response generation (Kar & Srivastava, 2018).

The residues of vaccine construct represent well defined secondary structural elements with presence of more α-helices. Tertiary structure analysis shows all residues to be in proper conformation, while ProSA output is in accordance to the score range typically found for native proteins in the structure database. Altogether, the validation results suggest that the quality of obtained model is good and could be used for further analysis. A conformational B-cell epitope is the cluster of amino acids that are discontinuous in the protein sequence, but present proximally in the three-dimensional structure of the vaccine protein (Zheng et al., 2015). Approximately 60% residues of vaccine candidate are present within the conformational B-cell epitopes that advocates eliciting strong neutralizing antibody response against the virus.

Toll-like receptors (TLRs) are significant components of antiviral innate immune system of host as they identify conserved pathogen-associated molecular patterns (PAMPs) of micro-organisms including viruses (Lester & Li, 2014). Our study highlights several interactions between the vaccine candidate and toll-like receptors with lowest energy score during molecular docking, which could result a prompt immune response and facilitate the process of targeted delivery of vaccine candidate into the host cells. Docking studies showed that vaccine candidate interacts with more affinity to the TLR3 receptor as compared to other toll-like receptors (TLR2 and 4). Previous studies involving immune receptor and S protein vaccine complex (Abraham Peele et al., 2020; Samad et al., 2020) were focused only on energy score of the complex, while our study has emphasized on the epitope dominating interactions along with lowest energy of the complex to achieve better immune response in natural conditions. The TLR3-vaccine candidate complex also shows compactness and stability during MD simulation. It has also been observed earlier that complex of TLR3 receptor with SARS-CoV-2 spike protein was stable during simulation and the average RMSD time was also in accordance with our study (Samad et al., 2020). Simultaneously, the binding study shows that the interactions between the TLR3 and vaccine in the stable complex structure provide more affinity with highly negative binding energy and stability to the structure than the initial complex. Similar observations have been reported in other recently published research showing interactions between viral vaccine candidates and host immune receptors including TLR3, HLA-A*11:01 (Enayatkhani et al., 2020; Tahir Ul Qamar et al., 2020). Moreover, codon optimization was employed to increase transcriptional and translational proficiency to achieve over-expression of vaccine protein in the bacterial system (Makrides, 1996). *In silico* cloning of vaccine candidate and its virtual confirmation support experimental expression of protein in *E. coli* host to enable the speedy and large scale production of the vaccine.

In conclusion, we have designed a multi-epitope vaccine candidate against SARS-CoV-2 from its RBD domain including its motif region applying immunoinformatics approach. Altogether, six (CTL, HTL and LBL) epitopes were obtained through bioinformatics tools that encompass several residues which are significant for binding of viral RBD domain to the human ACE2 receptor. Subsequently, the vaccine candidate was found to be highly antigenic, stable and soluble along with non-allergenic for host cells. The binding study of vaccine candidate with TLR3 immune receptor establishes the stability of the complex, while *in silico* cloning confirms its efficient and stable expression into the bacterial system. Furthermore, this vaccine could stimulate humoral and cell-mediated immunity into the human body due to the presence of CTL, IFN-γ inducing HTL and B-cell epitopes. However, the experimental validation of the proposed vaccine candidate is required to establish its potency and efficacy to fight against dreaded CoVID-19.

Acknowledgements

The authors acknowledge Indian Council of Medical Research (ICMR), Government of India for financial assistance, and Centre for Modelling Simulation & Design (CMSD), University of Hyderabad, Hyderabad for computational resources for MD simulation studies. The authors are thankful to Dr. N.P. Prabhu for his suggestions during revision of the manuscript.

Disclosure statement

No potential conflict of interest is reported by the authors.

References

Abraham Peele, K., Srihansa, T., Krupanidhi, S., Vijaya Sai, A., & Venkateswarulu, T. C. (2020). Design of multi-epitope vaccine candidate against SARS-CoV-2: A *in-silico* study. *Journal of Biomolecular Structure & Dynamics*, 1–9. https://doi.org/10.1080/07391102.2020.1770127

Abraham, M. J., Murtola, T., Schulz, R., Pall, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

Ahmad, B., Ashfaq, U. A., Rahman, M. U., Masoud, M. S., & Yousaf, M. Z. (2019). Conserved B and T cell epitopes prediction of ebola virus glycoprotein for vaccine development: An immuno-informatics approach. *Microbial Pathogenesis*, *132*, 243–253. https://doi.org/10.1016/j.micpath.2019.05.010

Ahmad, S., Navid, A., Farid, R., Abbas, G., Ahmad, F., Zaman, N., Parvaiz, N., & Syed, S. S. (2020). Design of a novel multi epitope-based vaccine for pandemic coronavirus disease (COVID-19) by vaccinomics and probable prevention strategy against avenging zoonotics. *European Journal of Pharmaceutical Sciences: Official Journal of the European Federation for Pharmaceutical Sciences*, *151*, 105387. https://doi.org/10.1016/j.ejps.2020.105387

Best, R. B., Zhu, X., Shim, J., Lopes, P. E., Mittal, J., Feig, M., & Mackerell, A. D. Jr (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. *Journal of Chemical Theory and Computation*, *8*(9), 3257–3273. https://doi.org/10.1021/ct300400x

Channappanavar, R., Zhao, J., & Perlman, S. (2014). T cell-mediated immune response to respiratory coronaviruses. *Immunologic Research*, *59*(1–3), 118–128. https://doi.org/10.1007/s12026-014-8534-z

Chen, W. H., Hotez, P. J., & Bottazzi, M. E. (2020). Potential for developing a SARS-CoV receptor-binding domain (RBD) recombinant protein as a heterologous human vaccine against coronavirus infectious disease (COVID)-19. *Human Vaccines & Immunotherapeutics*, *16*(6), 1234–1239. https://doi.org/10.1080/21645515.2020.1740560

Chen, X., Zaro, J. L., & Shen, W. C. (2013). Fusion protein linkers: Property, design and functionality. *Advanced Drug Delivery Reviews*, *65*(10), 1357–1369. https://doi.org/10.1016/j.addr.2012.09.039

Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J., & Jiang, S. (2009). The spike protein of SARS-CoV-a target for vaccine and therapeutic development. *Nature Reviews. Microbiology*, *7*(3), 226–236. https://doi.org/10.1038/nrmicro2090

Enayatkhani, M., Hasaniazad, M., Faezi, S., Guklani, H., Davoodian, P., Ahmadi, N., Einakian, M. A., Karmostaji, A., & Ahmadi, K. (2020). Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: An *in silico* study. *Journal of Biomolecular Structure & Dynamics*, 1–16. https://doi.org/10.1080/07391102.2020.1756411

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N.-H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell, *181*(2), 271–280.e8. https://doi.org/10.1016/j.cell.2020.02.052

Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, *118*(45), 11225–11236. https://doi.org/10.1021/ja9621760

Kalita, P., Padhi, A. K., Zhang, K., & Tripathi, T. (2020). Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2. *Microbial Pathogenesis*, *145*, 104236. https://doi.org/10.1016/j.micpath.2020.104236

Kar, P. P., & Srivastava, A. (2018). Immuno-informatics analysis to identify novel vaccine candidates and design of a multi-epitope based vaccine candidate against *Theileria* parasites. *Frontiers in Immunology*, *9*, 2213. https://doi.org/10.3389/fimmu.2018.02213

Kar, T., Narsaria, U., Basak, S., Deb, D., Filippo Castiglione, F., David, M., Mueller, D. M., & Srivastava, A. P. (2020). A candidate multi-epitope vaccine against SARS-CoV-2. *Scientific Reports*, *10*(1), 10895.https://doi.org/10.1038/s41598-020-67749-1

Kim, J., Yang, Y. L., Jang, S. H., & Jang, Y. S. (2018). Human β-defensin 2 plays a regulatory role in innate antiviral immunity and is capable of potentiating the induction of antigen-specific immunity. *Virology Journal*, *15*(1), 124. https://doi.org/10.1186/s12985-018-1035-2

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & 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

Kumari, R., Kumar, R., & Lynn, A., Open Source Drug Discovery Consortium. (2014). g_mmpbsa-a GROMACS tool for high-throughput MM-PBSA calculations . J Chem Inf Model, *54*(7), 1951–1962. https://doi.org/10.1021/ci500020m

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature, *581*(7807), 215–220. https://doi.org/10.1038/s41586-020-2180-5

Lester, S. N., & Li, K. (2014). Toll-like receptors in antiviral innate immunity. *Journal of Molecular Biology*, *426*(6), 1246–1264. https://doi.org/10.1016/j.jmb.2013.11.024

Li, X., & Ma, X. (2020). Acute respiratory failure in COVID-19: Is it "typical" ARDS? *Critical Care (London, England)*, *24*(1), 198. https://doi.org/10.1186/s13054-020-02911-9

Liu, S., Xiao, G., Chen, Y., He, Y., Niu, J., Escalante, C. R., Xiong, H., Farmar, J., Debnath, A. K., Tien, P., & Jiang, S. (2004). Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: Implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet (London, England)*, *363*(9413), 938–947. https://doi.org/10.1016/S0140-6736(04)15788-7

López-Blanco, J. R., Aliaga, J. I., Quintana-Ortí, E. S., & Chacón, 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

Makrides, S. C. (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiological Reviews*, *60*(3), 512–538.

McLoughlin, R. M., Lee, J. C., Kasper, D. L., & Tzianabos, A. O. (2008). IFN-gamma regulated chemokine production determines the outcome of *Staphylococcus aureus* infection. *Journal of Immunology (Baltimore, Md. : 1950)*, *181*(2), 1323–1332. https://doi.org/10.4049/jimmunol.181.2.1323

Perrie, Y., Mohammed, A. R., Kirby, D. J., McNeil, S. E., & Bramwell, V. W. (2008). Vaccine adjuvant systems: Enhancing the efficacy of sub-unit protein antigens. *International Journal of Pharmaceutics*, *364*(2), 272–280. https://doi.org/10.1016/j.ijpharm.2008.04.036

Phan, T. (2020). Novel coronavirus: From discovery to clinical diagnostics. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *79*, 104211 https://doi.org/10.1016/j.meegid.2020.104211

Raj, V. S., Mou, H., Smits, S. L., Dekkers, D. H., Müller, M. A., Dijkman, R., Muth, D., Demmers, J. A., Zaki, A., Fouchier, R. A., Thiel, V., Drosten, C., Rottier, P. J., Osterhaus, A. D., Bosch, B. J., & Haagmans, B. L. (2013). Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature, *495*(7440), 251–254. https://doi.org/10.1038/nature12005

Rehman, H. M., Mirza, M. U., Ahmad, M. A., Saleem, M., Froeyen, M., Ahmad, S., Gul, R., Alghamdi, H. A., Aslam, M. S., Sajjad, M., & Bhinder, M. A. (2020). A Putative prophylactic solution for COVID-19: Development of novel multiepitope vaccine candidate against SARS-COV-2 by comprehensive immunoinformatic and molecular modelling approach. Biology, *9*(9), 296. https://doi.org/10.3390/biology9090296

Rose, G. D., Gierasch, L. M., & Smith, J. A. (1985). Turns in peptides and proteins. *Advances in Protein Chemistry*, *37*, 1–109.https://doi.org/10.1016/s0065-3233(08)60063-7

Sabetian, S., Nezafat, N., Dorosti, H., Zarei, M., & Ghasemi, Y. (2019). Exploring dengue proteome to design an effective epitope-based vaccine against dengue virus. *Journal of Biomolecular Structure & Dynamics*, *37*(10), 2546–2563. https://doi.org/10.1080/07391102.2018.1491890

Samad, A., Ahammad, F., Nain, Z., Alam, R., Imon, R. R., Hasan, M., & Rahman, M. S. (2020). Designing a multi-epitope vaccine against SARS-CoV-2: An immunoinformatics approach. *Journal of Biomolecular Structure & Dynamics*, 1–17. https://doi.org/10.1080/07391102.2020.1792347

Solanki, V., Tiwari, M., & Tiwari, V. (2019). Prioritization of potential vaccine targets using comparative proteomics and designing of the chimeric multi-epitope vaccine against *Pseudomonas aeruginosa*. *Scientific Reports*, *9*(1), 5240. https://doi.org/10.1038/s41598-019-41496-4

Tahir Ul Qamar, M., Shokat, Z., Muneer, I., Ashfaq, U. A., Javed, H., Anwar, F., Bari, A., Zahid, B., & Saari, N. (2020). Multiepitope-based subunit vaccine design and evaluation against respiratory syncytial virus using reverse vaccinology approach. Vaccines, *8*(2), 288. https://doi.org/10.3390/vaccines8020288

Tina, K. G., Bhadra, R., & Srinivasan, N. (2007). PIC: Protein interactions calculator. *Nucleic Acids Research*, *35*(Web Server issue), W473–W476. https://doi.org/10.1093/nar/gkm423

Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., Lu, G., Qiao, C., Hu, Y., Yuen, K. Y., Wang, Q., Zhou, H., Yan, J., & Qi, J. (2020). Structural and functional basis of SARS-CoV-2 entry by using human ACE2. Cell, *181*(4), 894–904.e9. https://doi.org/10.1016/j.cell.2020.03.045

Wang, Y., Zhang, D., Du, G., Du, R., Zhao, J., Jin, Y., Fu, S., Gao, L., Cheng, Z., Lu, Q., Hu, Y., Luo, G., Wang, K., Lu, Y., Li, H., Wang, S., Ruan, S., Yang, C., Mei, C., … Wang, C. (2020). Remdesivir in adults with severe COVID-19: A randomised, double-blind, placebo-controlled, multicentre trial. *Lancet (London, England)*, *395*(10236), 1569–1578. https://doi.org/10.1016/S0140-6736(20)31022-9

Zheng, W., Ruan, J., Hu, G., Wang, K., Hanlon, M., & Gao, J. (2015). Analysis of conformational b-cell epitopes in the antibody-antigen complex using the depth function and the convex hull. *PloS One*, *10*(8), e0134835 https://doi.org/10.1371/journal.pone.0134835

Zhou, Y., Jiang, S., & Du, L. (2018). Prospects for a MERS-CoV spike vaccine. *Expert Review of Vaccines*, *17*(8), 677–686. https://doi.org/10.1080/14760584.2018.1506702

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., & Tan, W, China Novel Coronavirus Investigating and Research Team. (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

Zhu, X., Liu, Q., Du, L., Lu, L., & Jiang, S. (2013). Receptor-binding domain as a target for developing SARS vaccines. *Journal of Thoracic Disease*, *5 Suppl 2*(Suppl 2), S142–S148. https://doi.org/10.3978/j.issn.2072-1439.2013.06.06