JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
https://doi.org/10.1080/07391102.2021.1924265

Engineering a multi epitope vaccine against SARS-CoV-2 by exploiting its non structural and structural proteins

Vikrant Singh Rajput[a]*, Ritika Sharma[a]*, Anchala Kumari[a,b]*, Nidhi Vyas[a], Vijay Prajapati[c] and Abhinav Grover[a]

[a]School of Biotechnology, Jawaharlal Nehru University (JNU), Delhi, India; [b]Department of Biotechnology, Teri School of Advanced Studies, New Delhi, India; [c]Department of Biochemistry, School of Life Sciences, Central University of Rajasthan, Ajmer, Rajasthan, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

SARS-CoV-2, the causative agent behind the ongoing pandemic exhibits an enhanced potential for infection when compared to its related family members- the SARS-CoV and MERS-CoV; which have caused similar disease outbreaks in the past. The severity of the global health burden, increasing mortality rate and the emergent economic crisis urgently demands the development of next generation vaccines. Amongst such emergent next generation vaccines are the multi-epitope subunit vaccines, which hold promise in combating deadly pathogens. In this study we have exploited immunoinformatics applications to delineate a vaccine candidate possessing multiple B and T cells epitopes by utilizing the SARS-CoV-2 non structural and structural proteins. The antigenicity potential, safety, structural stability and the production feasibility of the designed construct was evaluated computationally. Furthermore, due to the known role of human TLR-3 immune receptor in viral sensing, which facilitates host cells activation for an immune response, the vaccine construct was examined for its binding efficiency using molecular docking and molecular dynamics simulation studies, which resulted in strong and stable interactions. Finally, the immune simulation studies suggested an effective immune response on vaccine administration. Overall, the immunoinformatics analysis advocates that the proposed vaccine candidate is safe and immunogenic and therefore can be pushed as a lead for *in vitro* and *in vivo* investigations.

ARTICLE HISTORY
Received 5 December 2020
Accepted 24 April 2021

KEYWORDS
SARS-COV-2; immunoinformatics; multi-epitope; vaccine; docking; MD simulation

1. Introduction

Coronavirus disease (COVID-19) effectuated by the severe acute respiratory syndrome coronavirus 2 virus (SARS-CoV-2), is a globally asserted pandemic by the World Health Organization (WHO) (Bianchi et al., 2020; Perlman & Netland, 2009).The outbreak was first documented from Wuhan, China, in 2019 (Riou & Althaus, 2020). The highly infectious and deadly virus has been successful in proliferating to 218 countries and territories around the globe till date. According to World Health Organization (WHO), by 18th Mar, 2021, there are 120,915,219 known cases of Covid-19 worldwide, out of which 53,340,393 cases alone are attributed to America, while Europe has 41,811,235 cases, leading to a death toll of over 2,674,078 worldwide. Coronaviruses belongs to the family Coronaviridae (that encompasses two subfamilies of Coronavirinae and Torovirinae), of the order Nidovirales (Cavanagh, 2005; Emmanuel, 2020). These are large, enveloped, positive-stranded RNA viruses, with a genome size typically ranging from 27 to 32 kb, largest amongst all RNA viruses. The nucleocapsid protein (N) forms a helical capsid structure that captures the viral genome inside it, this capsid is enclosed by an envelope comprising the membrane (M) protein, envelope (E) protein, and the spike (S) protein (Nieto-Torres et al., 2011; Venkatagopalan et al., 2015).

In the past vaccine development against the outbreaks of SARS and MERS have accomplished limited success, which indicates that developing the vaccine against COVID-19 could be a formidable task to achieve (Hilgenfeld & Peiris, 2013; Ojha et al., 2020). Recently, among other approaches, the multi epitope subunit vaccine (MESV) has proved to be an efficacious approach against the virus. Various *in silico* tools could rationally provide a competent vaccine construct that can be pursued further. A successful vaccine candidate, demands attributes such as high immunogenicity, antigenicity and acceptable physicochemical properties along with no allergenicity and toxicity. The designing of a vaccine using the MESV approach could exactly address these issues (Urrutia-Baca et al., 2019). To date numerous vaccines have been created using *in silico* methods against various viruses such as the Hepatitis C virus, Human immunodeficiency

CONTACT Abhinav Grover agrover@jnu.ac.in School of Biotechnology, Jawaharlal Nehru University (JNU), Delhi 110067, India
*These authors contributed equally to the study.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/07391102.2021.1924265.

virus-1, and many more (Nezafat et al., 2016; Yang et al., 2015).Short stretches of amino acid residues from antigenic proteins called as epitopes, are recognisable by B-cell Lymphocytes (BCL), Helper T Lymphocytes (HTL) and Cytotoxic T Lymphocytes (CTL) which may evoke a direct and vigorous immune reaction as compared to the response produced by a whole protein alone (Kao & Hodges, 2009). Many innovative vaccination strategies centred on rational design of B- and T-cell epitopes from one and/or several immunogenic stretches of different pathogens are now advancing to the clinical trials phase (Farhani, 2019; Li et al., 2014; Oyarzún & Kobe, 2016; Skwarczynski & Toth, 2016). However, the malaria vaccine Mosquirix™ also known as RTS, S is the only licensed multi epitope based vaccine to be available currently (Laurens, 2019; Oyarzún & Kobe, 2016). In late 2019, this vaccine was authorized for use in three African countries on children in high risk groups, by the WHO (Adepoju, 2019; Schuerman, 2019). Since the emergence of the novel coronavirus or SARS CoV-2 and the realisation of its catastrophic proficiency, researchers all across the world have been trying to advance both prophylactic and therapeutic treatments against it. The availability of the whole genome sequence of SARS-Cov-2 has provided a better understanding, for the development of vaccines, diagnostics and therapeutics against the same (Hilgenfeld & Peiris, 2013; Ratnadeep Saha, 2020). Immunoinformatics and reverse vaccinology has the advantage of providing the scientific community with proficient leads in a short span of time.

The engineered vaccine construct proposed in the current study utilizes non-toxic antigenic epitopes against B (BCL) and T cells (HTL/T_H and CTL/T_C) by utilizing non-structural proteins viz. non-structural protein 2 (NSP2), non-structural protein 8 (NSP8), and non-structural protein13 (NSP13)/helicase, which are referred as the tool proteins executing the viral replication (Ojha et al., 2020) and structural proteins, namely nucleoprotein (N) and receptor binding domain (RBD) of the spike (S) protein. NSP2 is known to be conserved in both SARS-CoV and SARS-CoV-2; it binds to host proteins prohibitin 1 and prohibitin 2 that are vital for cell cycle passage, cellular differentiation, apoptosis, cell migration and mitochondrial biogenesis. NSP2 binding is critical in disruption of the host cellular environment (Yoshimoto, 2020). Further, recent evidences have shown that the highly contagious nature of SARS-Cov-2 might be because of a stabilizing mutation in the endosome associated protein like domain of the NSP2 protein (Angeletti et al., 2020); thereby making it an important target for vaccine development. NSP8, a peptide cofactor forms a heterodimer with another peptide cofactor, NSP7 and this complex in turn binds NSP12, which eventually organizes the RNA polymerase complex (Yoshimoto, 2020). This makes it a valuable protein for vaccine search. The nucleoprotein is a major SARS-CoV structural protein which is known to bind RNA and stabilize it (Yoshimoto, 2020). The SARS-CoV-2 nucleoprotein displays 94% homology to the nucleoprotein of the SARS-CoV; and likewise, it also suppresses RNAi by sequestrating ds RNA; and is thus directly involved in viral RNA replication. Because of this suppression, the N protein represents an important factor in host immune evasion by SARS-Cov-2 and therefore, contributes to its pathogenicity; implying that it can also play a crucial role in developing an effective vaccine (Mu et al., 2020). The spike protein (S), facilitates the viral attachment and, thus, the virus entry into the host cells (Shang et al., 2020). The RBD of the spike protein specifically recognizes the human Ace2 receptor. This receptor recognition is crucial for the pathogenesis and host range along with viral infectivity, making it another vital target for vaccine development (Shang et al., 2020).

The proposed vaccine construct has been evaluated for its immunogenic potential, through interaction studies with the human TLR3 immune receptor molecule. TLR3 is involved in viral sensing through the recognition of viral RNA in endosomes, and facilitating host cells activation for an immune response (Iwasaki & Yang, 2020).Molecular docking and molecular dynamics (MD) simulation analyses were carried out to deduce plausible interplay between the vaccine construct and TLR3. This was followed by conducting immune-simulation studies that provided an overview of the possible immune response that could be elicited on administration of the purported vaccine.

2. Materials and methods

2.1. Servers and tools

All servers and tools utilized in the study are mentioned in their corresponding sections of usage and their links are provided in Supplementary Table 8.

2.2. Selection of proteins

The available literature for the SARS-CoV-2 was surveyed and based on it we decided to incorporate structural as well as non-structural proteins of its proteome into creation of a multi-epitope subunit vaccine candidate. The complete peptide sequence for each of the structural and non-structural proteins encoded by this virus was downloaded from the NCBI database, in the form of FASTA files. All the structural and non-structural proteins were analysed for their immunogenic potential via ANTIGENpro tool of the Scratch protein properties interpreter server (Ojha et al., 2020). Those proteins that exhibited a higher ANTIGENpro score ($>$0.65) were taken forward into the investigation. To ensure that the selected proteins exhibit no similarity to any human protein, a BLASTp search was conducted for each individual protein against the human protein dataset.

2.3. Identification of B cell epitopes

The filtered antigenic proteins of SARS-CoV-2, bearing insignificant homology with the human proteome, obtained from the previous step, were individually subjected to analysis by ABCpred server. The ABCpred server is routinely employed in reverse vaccinology for machine learning and artificial neural network (ANN) based computation of amino acid stretches within antigens that can mediate a humoral immune

response on interactions with B-lymphocytes under physiological conditions. The default threshold value and window size of 10 were used in identification of the epitopes (Ojha et al., 2020). The default threshold value of 0.51 is based on the evaluation of the sensitivity, specificity, accuracy, and positive prediction values and results in optimal performance by the server.

2.4. Identification of T_H cell epitopes

For predicting Helper T lymphocytes- MHCII recognised epitopes, the IEDB database was utilized. The selection of predicted binders can be based on the percentile rank or MHC binding affinity. IEDB currently recommends making selections based on a consensus percentile rank and the same strategy was employed in the current study. Individually, each antigenic protein sequence was provided as input to T cell epitope- MHCII binding tool of IEDB, while selecting the complete human HLA reference set in the parameter window. The rationale behind selecting the entire HLA reference set was to yield epitopes covering all the globally present MHC II alleles (Ojha et al., 2020).

2.5. Identification of T_C cell epitopes

Long lasting cellular immunity is provided by the development of cytotoxic cells on being presented with MHC-I bound antigenic peptides or epitopes. Cytotoxic cells have the ability to eliminate the circulating virus and the virus infected cells. The prediction of Cytotoxic T Lymphocytes-MHCI binding epitopes was accomplished by employing the Net-CTL 1.2 server. It utilizes ANN, and yields the immunogenic epitopes, by envisaging functions such as- MHCI binding affinity, proficiency of proteasomal cleavage sites and transporter associated with antigen processing (TAP). The amino acid sequence corresponding to each antigenic protein was provided as an input to the server individually. The epitopes were predicted by selecting default parameters (to give an optimal performance of epitope prediction and specificity as described on the server website) for the three MHCI supertypes- A2, A3 and B7, as these three collectively cover >88% of the global population (Ojha et al., 2020; Sette & Sidney, 1999; Shen et al., 2018).

2.6. Assessment of toxicity, prediction of antigenic potency and merging of epitopes to generate the vaccine construct and its efficacy determination

Toxicity of all the generated epitopes was assessed by employing the ToxinPred server which utilizes models formulated on the technique of machine learning and quantitative matrix, by applying various peptide properties, for predicting toxicity of the input peptide/protein sequence (Nosrati et al., 2019; Ojha et al., 2020). Before combining the different non toxic epitopes for fabricating the vaccine construct, the antigenic potential of each of these predicted B, T_H and T_C cell epitopes were determined by using VaxiJen (version 2.0) server that is available online. VaxiJen, replaces conventional alignment-based strategies with the auto-cross covariance (ACC) methodology to yield reliable results, when predicting antigenicity. A higher VaxiJen score is indicative of an enhanced antigenicity potential.

The predicted epitopes that yielded lower VaxiJen scores (below the threshold value of 0.4) were left out of the vaccine design study. The remaining epitopes were joined together with the aid of well-known linker/spacer amino acids sequences. The merging of epitopes was performed in the following order –starting at the amino (N) terminal, the peptide sequence corresponding to the beta defensin-3 adjuvant was incorporated, which was linked to the collection of different cell epitopes using the linker sequence 'EAAK'. The B cell epitopes were fused with each other using the 'KK' amino acids, the 'GPGPG' linker was used for combining the different T_H cell epitopes and the T_C cell epitopes were linked in series by means of the 'AAY' linker. The different linkers utilized, have been chosen based on previous reports and are known to perform specific functions, such as increased proteasome processing (KK linker) (Yano et al., 2005), improved protein flexibility (GPGPG linker) (Livingston et al., 2002), providing a site for proteasomal cleavage (AAY linker) (Ojha et al., 2019) and promoting α helix secondary structure formations (EAAAK linker) (Ojha et al., 2020) in the designed protein (vaccine molecule). The human beta defensin-3 is a well-known TLR-3 agonist, and increases the immunogenic potential of a vaccine by enhancing the effectivity of the humoral and cellular immune response, which defines its application in the current study (Gupta et al., 2020; Mohan et al., 2013). The VaxiJen server and ANTIGENpro tool of SCRATCH suite were utilized to check the efficacy of the fabricated construct. In order to ensure the use of an effective adjuvant, L7/L12 ribosomal protein and HABA protein, which are other routine adjuvants utilized in formulation of vaccines, was also attempted (Sarkar et al., 2020).

2.7. Physiochemical properties and safety assessment of the designed vaccine construct

The ExPASy portal's ProtParam tool, and the PepCalc tools were utilized for physiochemical evaluation of the designed vaccine construct (Nosrati et al., 2019; Ojha et al., 2020). Additionally, the possible solubility of this protein upon overexpression in *E. coli*, was computed by the SOLpro tool of the Scratch suite. For assessing the safety and allergenicity of the proposed vaccine construct, Algpred, AllerTop and AllergenFP webservers were utilized, which are freely available online (Dimitrov et al., 2014; Naik et al., 2020). These servers use sequence matching and machine learning algorithms, for predicting the allergenicity of a given peptide or protein sequence.

2.8. Prediction of trans-membrane helices

The membrane localized proteins bearing single or multiple trans-membrane domains, may not be explored by B cells for an immune response. Therefore, the vaccine construct was

evaluated for the presence of a possible trans-membrane helix conformation using the TMHMM, TMPred and the TOPCONS servers. The TOPCONS server also evaluates the presence of a possible signal peptide (SigP) cleavage site in the construct (Pourseif et al., 2019).

2.9. MHC alleles cluster analysis

The human MHC genomic region (HLA) is known to contain several thousand alleles, and is therefore extremely polymorphic (Thomsen et al., 2013). Thus, it is important to identify MHC allele molecules that might possess similar binding specificities for the vaccine construct. This was accomplished through the MHCcluster server. During the execution of this analysis, the default values for the parameters (sum of peptides to be involved, the number of bootstrap calculations, fraction of peptides to include in correlation), were selected while utilizing the A2, A3 and B7 supertypes for MHC class-I and the corresponding HTL epitope allele (supplementary table 3) for MHC class-II. The server produces a MHC specificity tree and a MHC specificity heat map.

2.10. Population coverage and conservancy analysis

The total population coverage was evaluated using the IEDB population coverage analysis tool (https://tools.iedb.org/population/) against MHC class-I and MHC class-II binding alleles. The tool investigates the total population covered by the epitopes of the proposed vaccine. Since, COVID-19 is a global pandemic, the entire world was selected for analysis, while keeping other parameters as default (Bui et al., 2006; Kar et al., 2020).

Conservancy status of the different epitopes used in fabricating the vaccine construct was evaluated using the conservation across the antigen tool (http://tools.iedb.org/conservancy/) of IEDB (Bui et al., 2007). The amino acid sequence corresponding to the individual ORF's for each of the selected proteins (NSP2, NSP8, NSP13, RBD and N) was retrieved in FASTA format, from all the SARS-CoV-2 reference genomes/CDS (92 in total) which are freely available at NCBI. Next, the epitopes utilized in creating the vaccine construct (as explained in the previous sections) from each of the above-mentioned proteins, were separately examined for their conservancy against their respective protein sequences, downloaded as described above (for all the 92 different strains/reference genomes of SARS-CoV-2 sequenced worldwide). The epitope linear sequence conservancy was evaluated by keeping the sequence identity threshold to $>=$50%.

2.11. Tertiary structure prediction

The three-dimensional structure of the designed multi-epitope vaccine was computed by the 3Dpro tool of the Scratch suite. This tool combines the use of predicted secondary structural features, a fragment library and energy terms derived from the PDB statistics to generate the tertiary structure of a given amino acid sequence (Cheng et al., 2005).

2.12. Modelled structure's refinement and validation

The predicted 3 D models for a given protein sequence are likely to have biophysical and topological errors such as unusual bond angles and lengths, irregular contacts or hydrogen bonds, and non-physical atomic clashes, which restrains their usage for further studies. Thus, it is essential to alleviate the errors and improve their stereochemistry through structural relaxation. As a single-long model-sampling protocol is usually found trapped in local energy minima (Garg et al., 2016; Runthala, 2012; Runthala & Chowdhury, 2019), the energetic landscape of the constructed model was traversed through Modeller 9.24 (Webb & Sali, 2016) to select the lowest normalized energy model with the lowest Molprobity score (V. B. Chen et al., 2010). The sampling is iteratively implemented until convergence (Runthala & Chowdhury, 2014). Modeller 9.24 was thus utilized to generate an energy refined structure for the designed vaccine construct. The refined structure was subsequently analysed for structural stability through online available applications like ProSA-web, PROCHECK and ERRAT.

2.13. Molecular interaction study with TLR-3

The possibility of an interaction with the immune receptor molecule -TLR3, for the designed vaccine construct was predicted via performing a molecular docking study. The human TLR-3 protein structure (PDB ID: 2A0Z), was downloaded from the RCSB PDB database and the refined vaccine structure obtained as described previously, served as the ligand for this receptor. .The likely interaction between the two protein molecules was checked using the High Ambiguity Driven protein-protein docking (HADDOCK) server. HADDOCK is different from other docking methods as it integrates a wide variety of experimental and bioinformatics data to guide the modelling process (Van Zundert et al., 2016). The active and passive amino acid residues of the receptor as well as the ligand, and their energy minimized PDB structures, are pre-requisites for performing docking by HADDOCK. To obtain these interactive residues, initially the CPROT tool was utilized which also demands PDB structures of the protein molecules as input. Docking was executed by the HADDOCK server after successful submission of all the input requirements.

2.14. Molecular dynamics simulations of TLR-3 and premeditated vaccine candidate

In this study, the TLR-3 and designed vaccine candidate docked complex was simulated with the help of GROMACS (GROningen Machine for Chemical Simulation) v5.0 program, utilizing the GROMOS96 54a7 force field, having water model SPC216 along with time step 1 fs for 20 ns (Abraham et al., 2015). It is a command line based application where various trajectory files are generated through all time frames and the motion of every atomic coordinate is defined. The simulation box for the MD run was produced, having a size of $10.075 \times 8.901 \times 7.011$ nm, which was later filled with about

78659 water molecules of SPC model for TLR-3-vaccine candidate complex simulation system. The total charge on TLR-3-vaccine candidate complex was found to be −6.000e, therefore to neutralize charge +6.000e was appended into the simulation system by compensating the water molecules in arbitrary locations inside the simulation box. The NPT ensembles, along with periodic boundary conditions, were utilized to carry out MD simulations. A cut-off of 12 Å was used in order to achieve the Vander Waals forces. The Particle Mesh Ewald model having a cut-off of 14 Å was further utilized to compute the electrostatic interactions (Darden et al., 1993). The TLR-3-vaccine candidate was solvated through a slab of about 10 Å in all directions. The neighbour list was rationalised to a frequency of 10 ps.

The simulation system consisted of four key steps. The first phase involved the energy minimization of the whole system employing the integrator of steepest descent in perpetuation with subsequent integrator of conjugate gradients algorithms. The second phase comprised of minimization and molecular dynamics of NVT and NPT ensembles for 500 ps and 1000 ps respectively, permitting the solvents and ions to evolve (here the starting configuration for the structures was kept similar). The third phase involved heating of the systems, having a lower temperature coupling ($\tau = 0.1$ ps) along with pressure coupling ($\tau = 0.5$ ps) to achieve equilibrium at 300 K and 1 atm of temperature and pressure. In the equilibration phase, the thermostat and barostat were estimated through the Berendsen algorithm (Berendsen et al., 1984). The hydrogen-containing bond lengths were repressed with the help of the LINCS algorithm (Hess et al., 1997). The last and fourth phase, termed as the production step is where the MD simulations for 20 ns at 300 K temperature having 2 fs of time step were completed, and the final structures were attained. The Maxwell Boltzmann distribution was employed in order to reassign the velocities at each step. Nose Hoover thermostat and Parrinello Rahman barostat were the respective thermostat and barostat for the final MD or production run (Berendsen et al., 1984).

Various interpretations were achieved with the assistance of inbuilt investigation commands of GROMACS. The root mean square deviation (RMSD) is a magnitude of the dimensional inequality between the two stagnant structures, and RMSD calculation is attained depending upon the native structure and every consecutive trajectory frame in the simulation. Furthermore, the root mean square fluctuation (RMSF) profile measures the displacement of every amino acid residue relative to the fluctuation about an average position within all MD simulations (Knapp et al., 2011). Therefore, RMSD and RMSF of TLR-3-vaccine candidate complex were determined to scrutinize stability and residual fluctuations. Further, the radius of gyration (R_g) was examined to estimate the compactness of the simulation system. Also, the hydrogen bond study was accomplished to check the neighboring interactions within the simulation system, including the hydrophobic interactions with the help of the DimPlot tool for TLR-3-vaccine complex before and after simulation.

Additionally, solvent accessibility surface area (SASA) was also computed to scrutinize the solvent attributable areas of TLR-3-vaccine candidate complex. Cluster exploration with a cut-off value of 0.25 nm liable upon the RMSD profile was employed to identify the most prominent conformations, found intermittently throughout the trajectory. Here, all the structures having RMSD values of below 0.25 nm for all components within a cluster are incorporated to the initial cluster. It is rare that a molecule displaying a RMSD value higher than 0.25 nm from other cluster, prospectively be treated as a structure. The secondary structure interpretation was also achieved by the DSSP program (Martin et al., 2005). The visualization of protein nature throughout the entire simulation was accomplished through Visual Molecular Dynamics (VMD) (Humphrey et al., 1996) and UCSF Chimera (Pettersen et al., 2004).

2.15. Immune simulation

The *in silico* immune reaction produced by the proposed vaccine construct was determined using C-ImmSim server (Rapin et al., 2010). It is an online available, dynamic immune simulation tool that produces well constructed and compatible results, by utilizing basic principles of immunology. The multilevel model expresses the immune reaction at the mesoscopic level, while considering the perception operations between immune system mediators, based on prediction tools for epitope identification. The default parameters were selected, which included 100 simulation steps (1 simulation step ≈ 8 h) with 10 µL of simulation volume of the antigen. The amino acid sequence of the proposed vaccine candidate was given as an input for executing the study. Since, at this juncture, the clinical data regarding the HLA susceptibility is missing, the HLA selection was also taken as default.

2.16. Codon optimization and design of expression vector

For over-expressing the proposed vaccine candidate in an appropriate *E. coli* expression vector, the protein sequence of the construct was reverse translated into the corresponding coding DNA sequence by utilizing the Gene infinity server, assembled on the codon usage of the *E.coli* K12 strain. Thereafter, using the GenScript tool, the optimized DNA sequence was examined based on parameters that are crucial for successful recombinant protein expression. These parameters include GC content, codon adaptation index (CAI), and codon frequency distribution (CFD) (Nosrati et al., 2019). Finally, the expression vector design was executed through the WebDSV ver. 2.0 server using the sequence of the expression vector, pET28a (+) which was retrieved from the 'addgene' vector database (Bhattacharya et al., 2020).

3. Results

3.1. Antigenic proteins selection, their retrieval and homology with human proteome

There exist two kinds of proteins in SARS-Cov-2 proteome, the non structural and the structural proteins. The non

structural proteins principally participate in the process of viral replication, whereas the structural proteins function in the progression of viral assembly and infection within the host (Mirza & Froeyen, 2020). Therefore, both classes of proteins should bear equal consideration while either developing a vaccine or a drug. In this study too, we considered a combination of both non structural and structural proteins for fabricating a vaccine. All the proteins encoded by the SARS CoV 2 proteome were analysed for their immunogenic potential via ANTIGENprotool. The NSP2, NSP5, NSP8, NSP9, NSP10, NSP14, NSP15, NSP16, NSP13/Helicase (H), Nucleoprotein (N), Spike glycoprotein (S) and separately only the RBD region of S protein exhibited high antigenicity with score greater than 0.65 (Supplementary Table 1). Since the previous study by Ojha et al. (2020) had utilized NSP5, NSP14, NSP13, NSP15 and NSP16 for creating a vaccine construct against SARS CoV-2, we decided to exclude all these proteins from our study except for the helicase. Utilizing RBD as opposed to the complete S protein is more advantageous as it eliminates the possibility of immunopotentiation, which is observed as a result of the presence of certain epitopes in the complete S protein (Amanat & Krammer, 2020; W. H. Chen et al., 2020; Jiang et al., 2012; Wang et al., 2016). Another observation was that RBD exhibited a higher antigenicity score as compared to S, thus RBD was preferred over S in this study. The BLASTp similarity check revealed viral helicase to display 22.4% of sequence identity to the human ZGRF1 isoform X10 protein, with a query coverage of just 34%. Since, the probability of cross reactivity is considered as very low, if the similarity between human and pathogen proteins is less than 40% (Pearson, 2013) we selected helicase for our further studies. Similarly, the nucleoprotein exhibited 55.56% identity with the human immunoglobin heavy chain junction region which appears high, but considering the exhibited query coverage value of just 4% and a high e-value of 5.4 in the BLASTp results, it was deemed insignificant. On this basis nucleocapsid protein was also taken forward into the investigation. All the other proteins that qualified antigenicity filter did not show any similarity against the human proteome.

3.2. B cell epitopes prediction

For eliciting an ideal humoral immune response, recognition of B cell epitopes by the B-cell receptors is essential. The designed vaccine therefore must present the B cell epitopes so that an efficient immune response can be triggered. The ABCPred server predicted the B cell epitopes against each selected protein. Based on their ranking and the highest score, the best epitope of 10 mer length for each protein was chosen for further investigations (Supplementary Table 2).

3.3. T_H cell epitopes prediction

The HTL epitopes play a vital part in producing both humoral and cellular immune responses. These epitopes yield helper cell (T_H) responses, that are essential for the development of both memory cytotoxic cells (T_C) and stimulation of B-cells for antibody production. The IEDB database was utilized to predict Helper T lymphocytes- MHCII recognised epitopes for each protein. The percentile rank and IC_{50} values were the main criterias utilized for selection of the predicted epitopes (specifically a cut off in percentile rank of less than or equal to 0.9 and an IC_{50}value of below or equal to 50). Lower the percentile rank and the IC_{50} value, better is the binding efficiency of the epitope (Ojha et al., 2020). In order to cover maximum population possible, the complete human HLA reference set was stipulated in the parameter window. The best two epitopes for each protein were selected. The preferred T_H cell epitopes with their IC_{50} value and the percentile rank score are presented in supplementary file in Table 3.

3.4. T_C cell epitopes prediction

The Net-CTL server was applied to identify the T_C cell epitopes. Barring the supertype selection, all other parameters such as threshold value (0.75), C-terminal cleavage (0.15), and TAP transport efficiency (0.05) were set to default values. The HLA supertype selection determines the specific and efficient perception of epitopes for T cell which are aimed at several viral diseases, including SARS, MERS, HIV and MMR (Ojha et al., 2020). Based on the fact that A2, A3 and B7 HLA supertypes together cover more than 88% of the global population, epitopes were predicted by selecting these supertypes in the parameter window. The study resulted in identification of 3 epitopes, each for every protein (Supplementary Table 4).

3.5. Prediction of antigenicity of the epitopes, assessment of toxicity and construction of the multi epitope vaccine construct and its efficacy determination

All the epitopes identified for the proteins NSP 2, NSP 8, NSP 9, NSP 10, NSP13/H, N and RBD were evaluated for their individual antigenicity using the VaxiJen server (Supplementary Tables 2–4). Those epitopes which exhibited VaxiJen scores below the threshold value of 0.4, were left out from the study. It was observed that certain epitopes predicted for NSP9 and NSP10 did not have a desired antigenicity value as assessed by VaxiJen server. Therefore, it was decided that the epitopes from these proteins be omitted from further analysis and inclusion in the final vaccine construct. The plausible toxicity of all the selected B and T cell epitopes was examined using the ToxinPred server and all of them were predicted to be safe and non toxic (Supplementary Tables 5–7).

The selected epitopes (antigenic peptides) were coupled together, utilizing the linkers to yield the vaccine candidate. These linkers, not only assist in maintenance of construct immunogenicity, but also in making the resulting protein complex more flexible and stable (Ojha et al., 2020). The B cell epitopes were linked using linkers 'KK' while the T_H cell and the T_C cell epitopes were connected through 'GPGPG'

Table 1. Predicted physicochemical properties of the vaccine construct.

Feature of the vaccine construct	Assessment using ProtParam	Assessment using PepCalc A
Number of amino acids	310	310
Molecular weight (g/mol)	33197.27	33196.88
Theoretical pI	9.79	10.16
Extinction coefficient ($M^{-1}\,cm^{-1}$)	46800	42670
Estimated half-life	30 h (mammalian reticulocytes, in vitro). >20 h (yeast, in vivo). >10 h (*Escherichia coli*, in vivo)	–
Instability index	20.26 (stable)	–
Aliphatic index	77.84	–
Grand average of hydropathicity (GRAVY)	−0.273	–
Estimated solubility	–	Good water solubility

and 'AAY' linkers, respectively. Finally, the 'EAAAK' linker joined the adjuvant – human beta defensin-3 (which is known to augment a sturdy B and T cell response (Gupta et al., 2020; Mohan et al., 2013)) to the assembled epitopes (at the N terminal) for generating the complete vaccine construct (Supplementary Figure 1). Thus, the final vaccine construct comprised of beta defensin adjuvant coupled with highly antigenic B and T cell epitopes from the viral NSP2, NSP8, NSP13 (helicase), Nucleoprotein and RBD of spike glycoprotein. On evaluating the generated vaccine construct for its antigenicity, score values of 0.60 and 0.84 were obtained by VaxiJen and ANTIGENpro respectively, confirming the strong immunogenic potential of the proposed multi subunit vaccine. The antigenicity of the merged epitopes was also tested by using either L7/L12 ribosomal protein and the HABA protein, individually as adjuvants in place of human beta defensin-3 and the antigenicity scores of 0.54 and 0.56 respectively, were obtained using the VaxiJen server. However, using ANTIGENpro server the antigenic scores did not alter significantly while utilizing each adjuvant individually with the merged epitopes. Based on these observations, finally we decided to proceed with human beta defensin-3 adjuvanted vaccine construct as our lead.

3.6. Physiochemical properties and safety assessment of the designed vaccine construct

Various physicochemical properties were determined for the vaccine construct using the ProtParam and PepCalc A tools. Both the tools yielded acceptable physicochemical properties for the vaccine construct. Briefly, the instability index was found to be 20.26 (should be below 40) and the half-life of the protein was predicted to be approximately- 30 h in mammalian reticulocytes, >20 h in yeast and >10 h in *E. coli*. The calculated GRAVY (Grand average of hydropathicity) score was −0.273, the value of aliphatic index of the protein was 77.4, the instability index (II) was determined to be 20.26, and the construct was predicted to display good water solubility. Additionally, the solubility upon recombinant expression was examined using the SOLpro tool, where the protein was anticipated to be soluble with a probability score of 0.793223. Table 1 depicts all these properties, evaluated by the mentioned tools.

The AlgPred server predicted the possible allergenicity for the vaccine construct. It applies six methods (IgE mapping, MEME/Mast motif, amino acid as well as dipeptide composition based SVM modules, BLAST inspection on ARPs and a hybrid approach that encompasses all these parameters) for determination of protein allergenicity. Among these methods, the SVM modules that take amino acid and dipeptide composition into consideration predicted the protein as allergenic; whereas the hybrid approach, MEME/Mast motif, BLAST search on ARPs and IgE surveying predictors classified the vaccine construct as non-allergenic. In addition, evaluation of the construct using the AllerTop and AllergenFP servers also identified the protein to be non-allergenic. Considering these results, the proposed vaccine candidate can be indexed as a probable non allergen.

3.7. Topology prediction of the vaccine construct

The trans-membrane helix forming regions, or the membrane spanning peptide portion of proteins are immunologically non-accessible. The vaccine construct has no noteworthy alpha-helix trans-membrane (TM) topology as concluded by the TMHMM probability plot (Supplementary Figure 2A). Using the TMPred web-server, where scores greater than 500 are considered as crucial for TM helix prediction, the proposed vaccine construct exhibited three fragments, 89–110 (score 633) in inside-to-outside (i-o) orientation, 130–149 (score 562) in outside-to-inside (o-i) orientation and 283–301 (score 1064) inside-to-outside (i-o) orientation when computed through the strongly preferred model (Supplementary Figure 2B). Whereas the alternative model predicted two TM helices, viz. 92–110 (score 1136) in outside-to-inside (o-i) orientation and 283–301 (score 1064) in inside-to-outside (i-o) orientation (Supplementary Figure 2B). The TOPCONS server which utilizes six algorithms to detect TM helix was also applied. Neither the TM helix nor a signal sequence was predicted in the vaccine construct by this server (Supplementary Figure 2C).

3.8. MHC cluster analysis

The potential MHC I and MHC II allele interactions with the epitopes in the vaccine candidate were examined by the MHCcluster server. The relationship of the cluster of alleles with the epitopes is represented in the form of a specificity heat map and tree (Supplementary Figure 3). The red and the yellow zone symbolize strong and weaker interactions respectively.

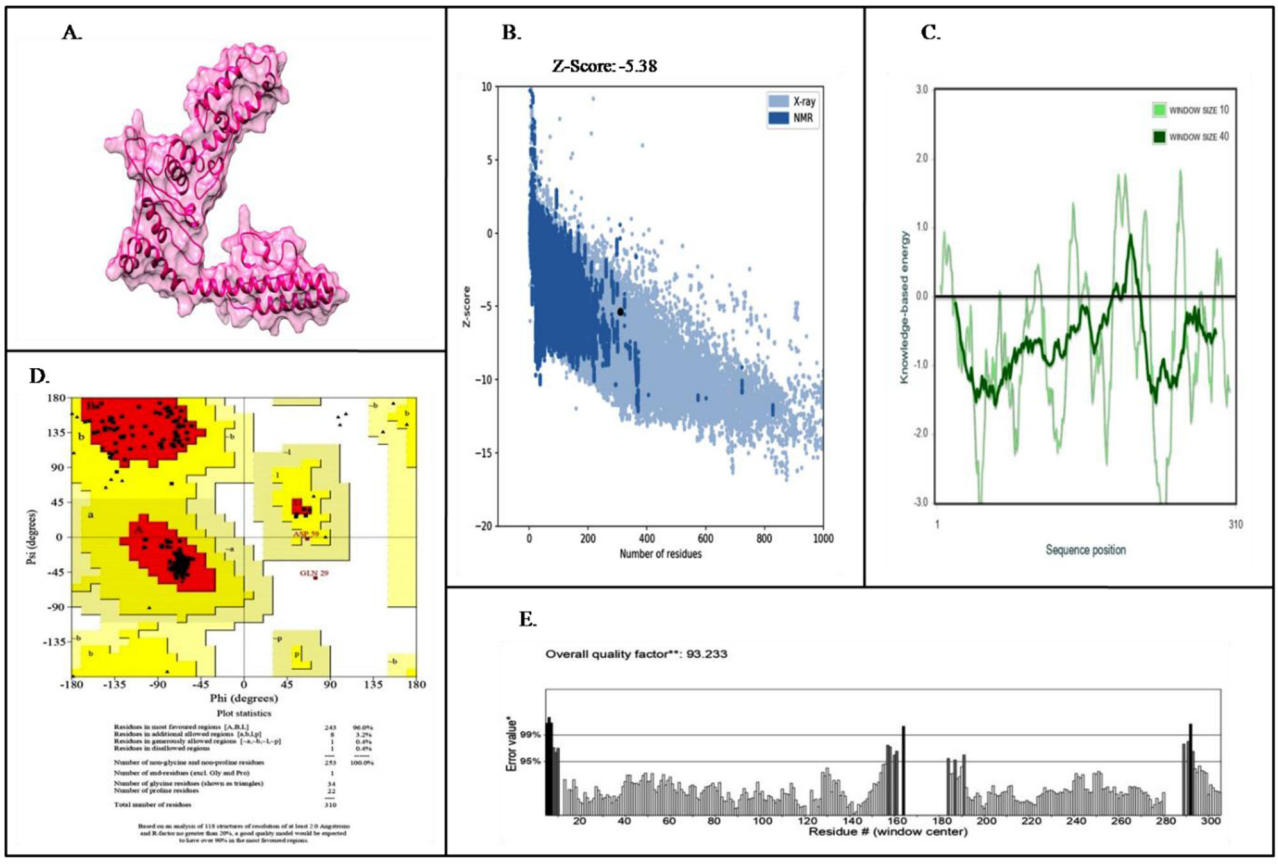

Figure 1. Structural validation of the refined modelled vaccine construct. (A) Refined tertiary structure of the vaccine protein. (B) ProSA Z-score (Overall model quality). (C) ProSA graphical plot (Local model quality). (D) Ramachandran plot produced using PROCHECK. The zones displayed using different colors viz. red, yellow and light yellow depicts the most favoured regions (96%), additional allowed regions (3.2%), and disallowed regions (0.4%), respectively. (E) The ERRAT plot.

3.9. Populations covered and conservation across SARS-CoV-2 isolates

For an efficient vaccine outcome, the evaluation of HLA allele distribution across the globe is important (Kar et al., 2020). Therefore, the efficiency of the proposed vaccine construct against the world population was investigated. The MHC class I and MHC class II epitopes selected for the vaccine construct displayed world population coverage of 80.35% (Supplementary Table 9).

All the epitopes utilized in the vaccine construct were evaluated for their conservancy against all the available SARS-CoV-2 reference sequences. All of the epitopes were observed to be conserved across the SARS-CoV-2 isolates reported till date (Supplementary Tables 5–7), suggesting that the proposed vaccine construct would be efficient globally, and that the mutation rates are extremely low for these epitopes.

3.10. Tertiary structure prediction and quality assurance of the designed candidate vaccine

The 3Dpro module of Scratch protein predictor, web-server estimated the likely 3 D configuration of the proposed vaccine candidate. On analysis by structure visualization tools such as PyMol and UCSF Chimera, it was observed that the designed multi-epitope vaccine majorly comprised of alpha helices, with very little contribution from a beta sheet structure. Structural refinement was performed by Modeller 9.24 and the refined structure (Figure 1A) was validated using ProSA, PROCHECK and ERRAT2. ProSA estimates a global quality score for the structure provided as input. The structure is considered free of errors if the ProSA -z score falls within a limit that is representative of native proteins having similar number of residues. The ProSA score for the fabricated vaccine construct's structure was observed to be −5.38 which falls well within the range, and thus was confirmed for its quality (Figure 1B and C). The PROCHECK generated Ramachandran plot validated the quality of the 3 D structure, as 96% of the residues were positioned in the most favoured regions and only 0.4% of them were localized in the disallowed regions (Figure 1D). The analysis by ERRAT2 predicted the overall quality factor as 93.233 (signified as proportion of the protein for which the computed rate of structural inaccuracy lies within the 95% elimination limit) (Figure 1E).

3.11. Molecular docking of the vaccine construct with TLR-3

For activation of an immune response, a suitable interaction amidst the immune receptor and the antigen molecules is essential. Therefore, to assure that the designed vaccine candidate is capable of associating with TLR-3, we performed a docking study using HADDOCK web-server. HADDOCK

Table 2. The scores procured from the molecular interaction of vaccine construct with immune receptor TLR-3 utilizing the HADDOCK server.

HADDOCK score	120.3 ± 24.7
Cluster size	15
RMSD from the overall lowest-energy structure	27.5 ± 0.2
Van der Waals energy	−90.7 ± 17.4
Electrostatic energy	−305.1 ± 24.8
Desolvation energy	−38.4 ± 5.4
Restraints violation energy	3103.9 ± 173.2
Buried Surface Area	3327.0 ± 246.9
Z-Score	−1.9

clustered 40 structures of the docked complex in a total of 6 clusters, the cluster with the best statistics (RMSD values; Van der Waals, Electrostatic and Desolvation energy values) was found to have a Z score of −1.9 (lowest of all the clusters). The more negative, the Z score, the better is the interplay between the receptor and the ligand. The details of the topmost docked cluster are given in the Table 2 and the binding of the vaccine candidate to TLR-3 is represented in Figure 2.

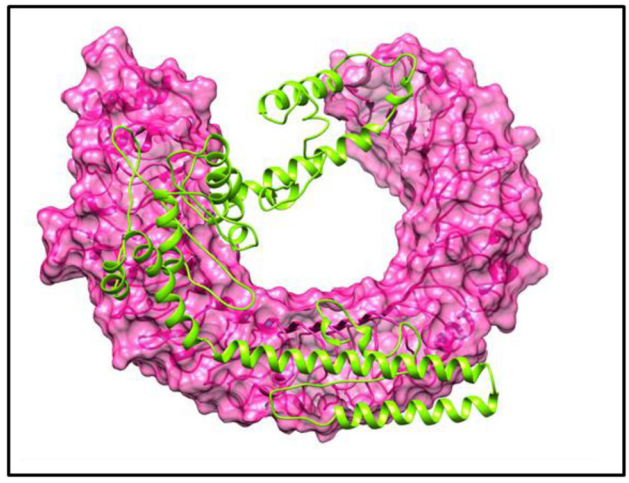

Figure 2. Molecular interaction study of vaccine candidate with immune receptor TLR3 using High Ambiguity Driven protein-protein docking (HADDOCK) server.

3.12. MD Simulation

MD simulation for the docked complex of TLR-3 and the designed vaccine was evaluated through trajectory study. For the progression of 20 ns MD run, the stable trajectory was perceived and the representative structures were acquired. Different steps for energy minimization, pressure and density equilibrium, and temperature persistence were analysed (Supplementary Figure 4). The deviation of the backbone atoms for simulated structures, comparative to the starting structures, was utilized as a reference and was assessed through RMSD (Figure 3A). A fluctuating RMSD profile during the course of the simulation run suggests a flexible and unstable interaction, or the variation in the force field. The assessment of RMSD in Figure 3A reveals that the RMSD fluctuation stabilizes around 10 ns into the MD simulation and thus, the total time for simulation was adequate. In the time frame 10–20 ns, the RMSD for TLR-3-vaccine candidate complex has an approximate value of about 0.7–0.8 nm.

The fundamental magnitude of this principle is attained by evaluating the disparities arising from modifications of each of the protein residues that majorly feature the most flexible chain frames. Hence, we validated the residual fluctuations by calculating the mean fluctuations for the stable trajectory of the simulation. The RMSF evaluation of all protein residues was attained in order to check the residues that may have inclined to an enhancement in the RMSD results (Figure 3B). Convincing fluctuations existed in the terminal residues, few β-sheet regions along with the loops linking the alternative β-sheets of TLR-3 (residues 330–350, and 490–550) to about 0.32 nm, and few α-helical sections of vaccine candidate (residues 130–210) to about 0.73 nm. Also, we noticed that residues 299–355 of TLR-3 which form the binding segment for the corresponding vaccine candidate residues (175–272) showed comparatively reduced variation in RMSF values after MD simulation. Interestingly, the binding region for TLR-3 and the designed vaccine before and after MD simulation was found to be similar, suggesting a strong binding interaction of the vaccine designed withTLR-3, and provides stability to the docked complex (Figure 4A). The residues in the interface and the stability of the interactions along the length of the simulations has been identified and shown in the supplementary data (Supplementary Tables 10 and 11).

The radius of gyration investigation was achieved to determine the modification in compactness of TLR-3 and the designed vaccine throughout the MD run. The R_g plot shows decrease in radius of gyration values and hence increase in compactness of the protein complex, which persists after MD simulation (Figure 3C). The compactness of the TLR-3 protein complex is due to the strong binding interaction of the designed vaccine. Similar observations were determined through SASA analysis representing the solvent defined protein surface and its orientation through folding, making the alterations in the exposed and buried regions of the surface area of proteins. The SASA values for the simulation was about 485 nm/S^2/N after 15 ns (Figure 3D). Also, it remains similar till 20 ns suggesting that TLR-3-vaccine solvation profile shows a convincing SASA value indicating a stable structure and strong binding interaction with the vaccine candidate.

Also, the hydrogen bond landscape with respect to the simulation time was evaluated, which exposed the dynamic equilibration of the complex trajectory with a high number of hydrogen bonds, as shown in Figure 5. The consistent high numbers of hydrogen bonds were perceived which contributed significantly to the proximal binding of the designed vaccine with the TLR-3 receptor. Further, these results are strengthened by the vital contribution ofthe complex binding energies throughout the simulation run. These calculations along with consistent high binding energies and large hydrogen bonds involvement demonstrate the stable binding of candidate vaccine with TLR-3.

Further, cluster analysis having a RMSD based cut-off value of 0.25 nm confirmed the development of 17 distinctive clusters for TLR-3 vaccine complex system. The most dominant cluster attained after 10 ns of MD simulation is

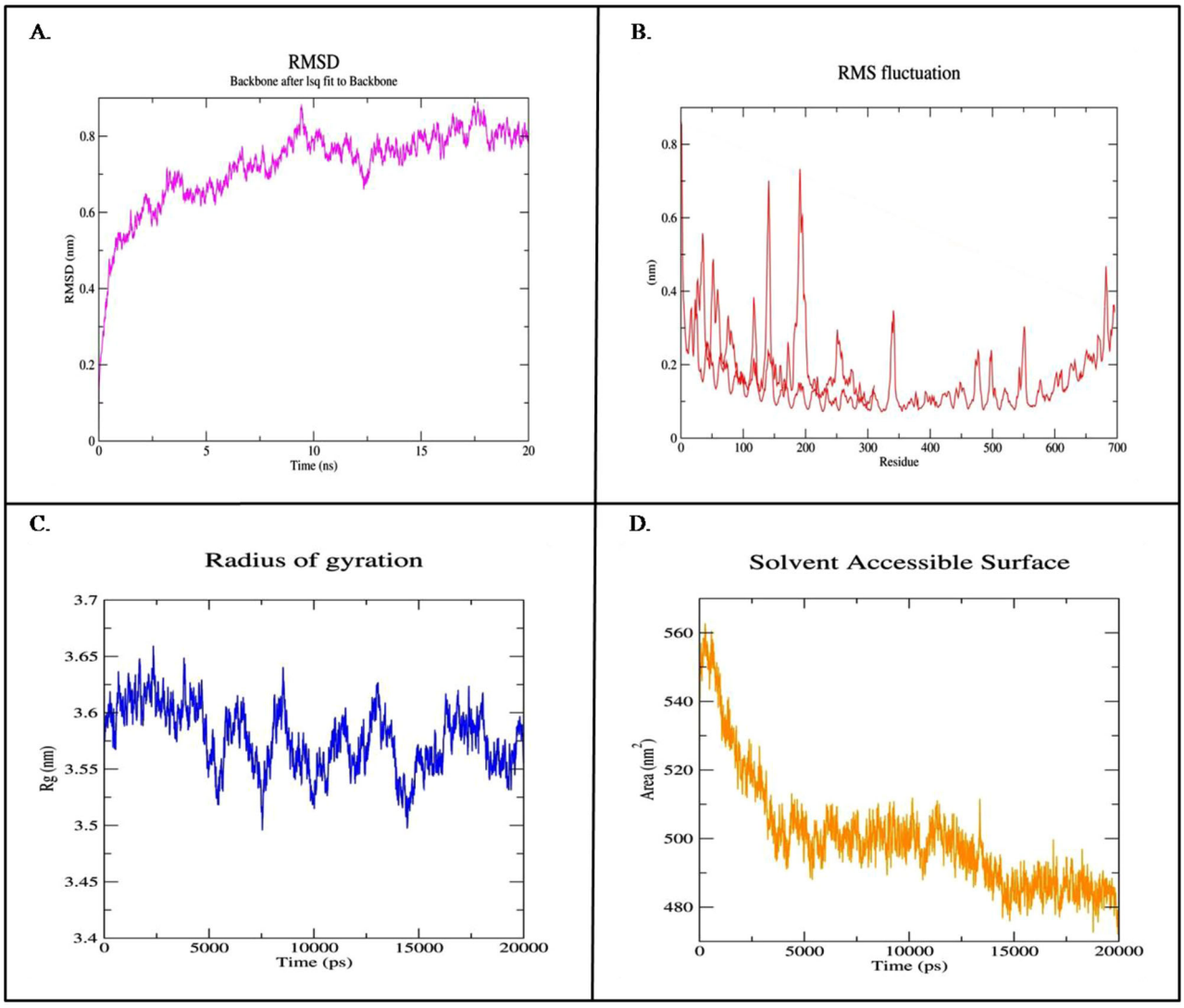

Figure 3. The stability of TLR-3-vaccine candidate as predicted by molecular dynamics simulations. (A) Root mean square deviation (RMSD) profile. (B) Root mean square fluctuation (RMSF) profile. (C) Radius of gyration (R_g) profile and (D) Solvent accessible surface area (SASA) profile. The comprehensive computational strategy was utilized to attain insights towards the designed vaccine candidate antigenicity against TLR-3.

depicted in Supplementary Figure 5. Also, a secondary structure investigation of the stable trajectory was implemented by the DSSP tool of GROMACS. The TLR-3-vaccine complex was formed mainly of continuous coils and conserved β-sheet regions as secondary structure elements infused with various small segments of bend, α-helix, turn, and β-bridge (Figure 4B). Both cluster analysis and secondary structure interpretation reveals the conformational modifications before and after simulations for TLR-3 vaccine complex structure.

3.13. Immune simulation

The C-ImmSim server was utilized for immune simulation study, which gives a picture of the possible immune response on vaccine administration. The immune response was generated for a single administration dose of the vaccine, for a total simulation time of ≈33 days as per the default settings of the server. The peaks for IgG1 + IgG2 and IgM show a secondary and tertiary immune reaction along with the presence of IgG + IgM, all of which seems to peak between 10-15 days with high titre (Figure 6A). The active B cell population was stimulated, which peaks in around 8 days and remains stagnant for many days (Figure 6B). Similarly, the cell mediated immune response which is led by T_C and T_H cells was also found to be high (Figure 6C and D). The IFN- γ level was also found to be inflated during vaccine administration (Figure 6E). These results indicate a high probabilityof generating a competent immune response which can potentially control the SARS-CoV-2 pandemic.

3.14. Codon optimization and design of expression vector

The vaccine construct was reverse translated into the corresponding nucleotide sequence and codon optimization was executed using the GenScript tool, which also predicts crucial parameters to forecast the feasibility of recombinant protein production in the host. The CAI value and GC content of the sequence was 1 and 62.37% respectively, along with 0% CFD

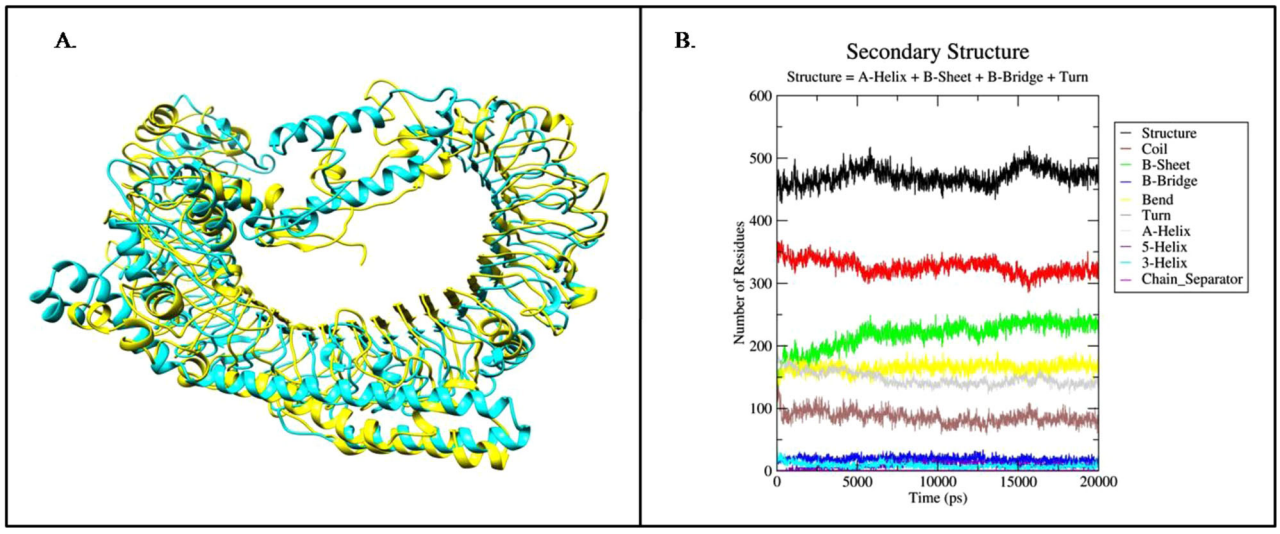

Figure 4. (A) Superimposed structure for the docked complex of TLR-3-designed vaccine candidate before simulation (cyan color) and after simulation (yellow color), the simulated structure attained compact conformation due to the strong interaction of vaccine candidate into the binding pocket of TLR-3 receptor. (B) The secondary structure analysis of the stable trajectory for TLR-3 receptor-vaccine candidate was performed by DSSP tool of GROMACS. Maximum of coil conformation is attained (red color) and β-sheets (green).

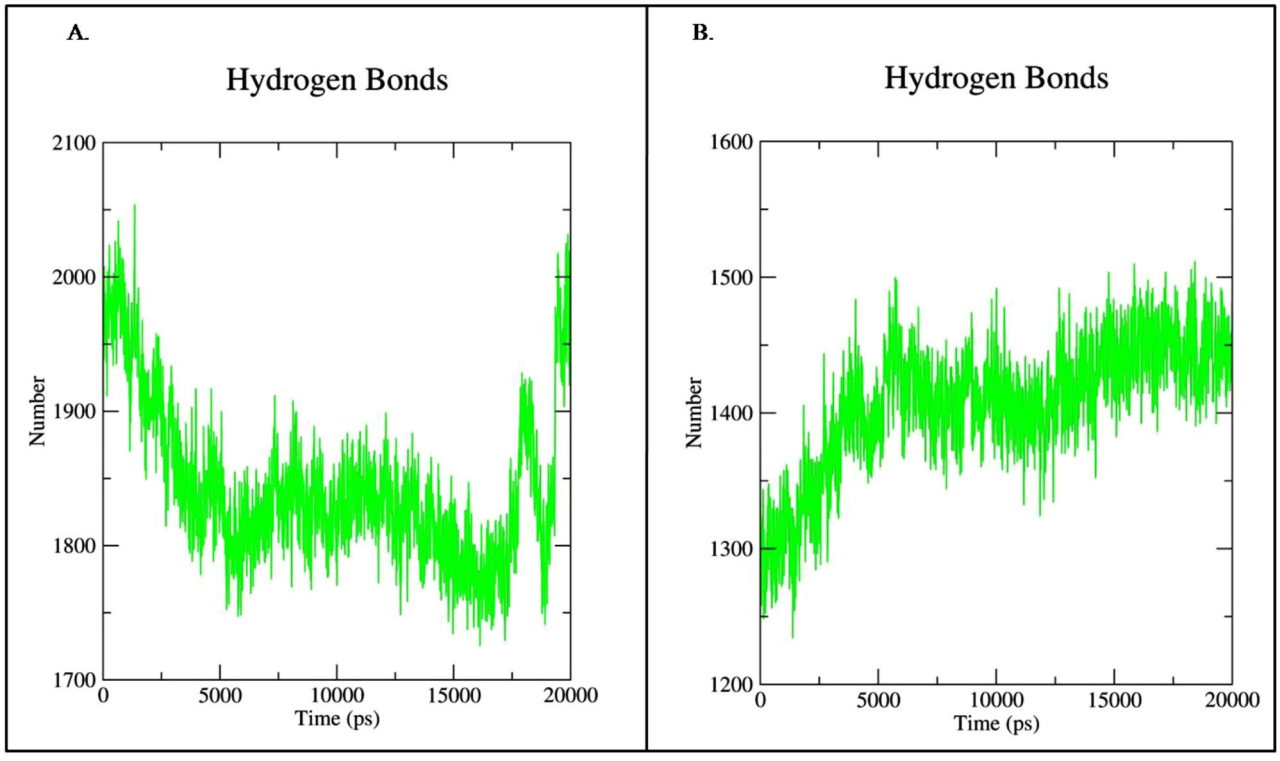

Figure 5. (A) Depiction of the time evolution of the total number of hydrogen bonds between TLR-3 receptor and the surrounding solvent molecules throughout the simulation time. (B) The statics of consistent high number of hydrogen bonds involved in strong binding of designed vaccine candidate to TLR-3 receptor with respect to the simulation time.

of the gene in *E. coli* (Supplementary Figure 6). These outcomes imply that the optimized nucleotide sequence is suitable for cloning and expression in *E. coli* (Mohammadi et al., 2019). The expression vector carrying the vaccine construct was designed using the WebDSV 2.0 server, utilizing the pET28a (+) expression vector's HindIII and BamHI restriction sites. The vector design is presented in the Figure 7.

4. Discussion

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has led to alarming numbers of fatalities across the world causing a global health emergency. According to the Milken Institute, COVID-19 Treatment and Vaccine Tracker, as of 18th Mar, 2021, a total of 251 vaccines against SARS CoV-2

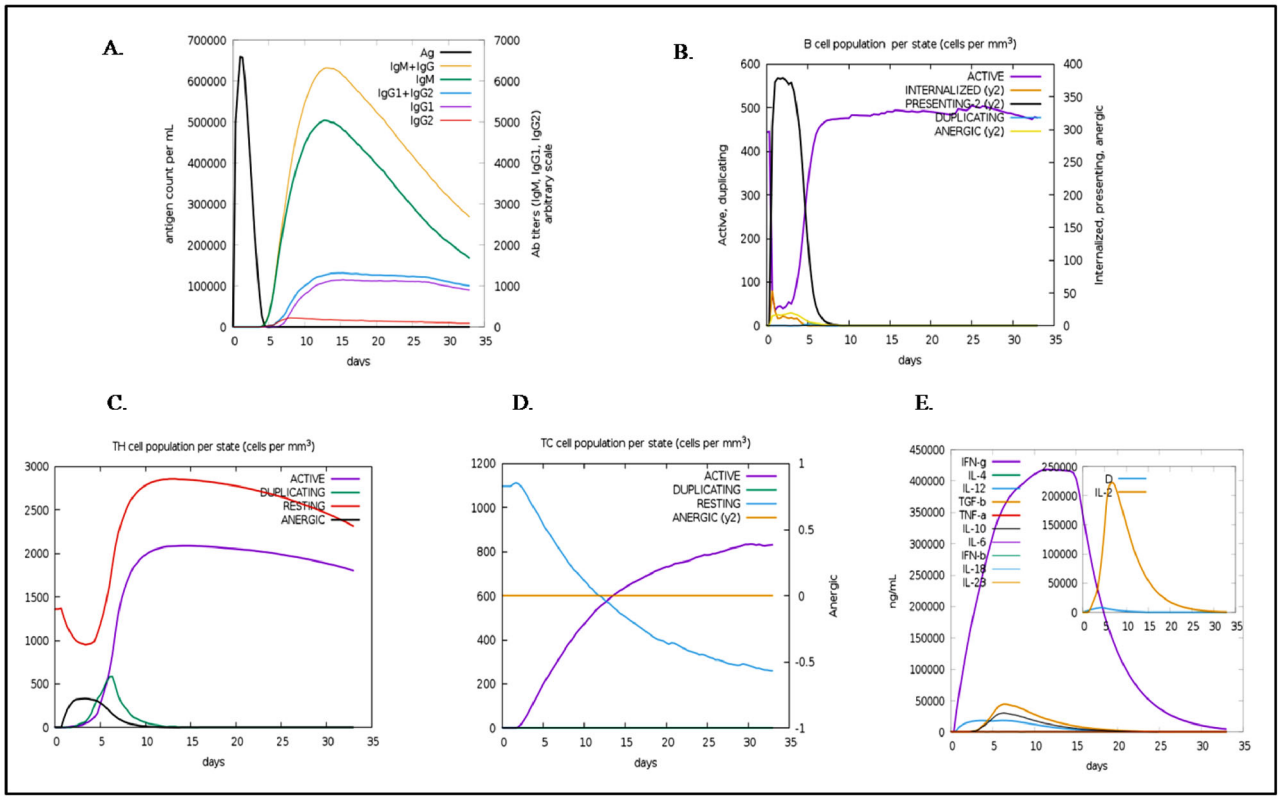

Figure 6. *In silico* immune simulation results of the vaccine construct using C-ImmSim. (A) Immunoglobulin generation response on antigen injection. Various subclasses of immunoglobulin are depicted as colored peaks. (B) Active B-cell population observed after administrating the vaccine construct. (C) The demonstration of the emergence of Helper-T cells. (D) The illustration of the production of cyototoxic-T cells on vaccine administration. In the graphs, RESTING implies to the cells, which were not shown to the antigens while ANERGIC indicates the tolerance level of antigen. (E) Graph displaying the induced cytokine level after vaccine administration. The inset graph illustrates the Simpson Index, D of IL-2. Simpson Index, D was interpreted as the measurement of diversity.

are underway in different stages of development. While the World Health Organization, Draft landscape of COVID-19 candidate vaccines – of 16th Mar, 2021, a total of 82 candidate vaccines are in the clinical phase evaluation while 182 candidate vaccines are in pre-clinical phase evaluation (Milken Institute, 2020; World Health Organization, 2020). To the best of our knowledge there are 4 vaccines in this cohort that are utilizing the epitope-based approach for generation of immunity against SARS CoV-2, and all of them are at the initial stages of development (Milken Institute, 2020; World Health Organization, 2020). ImmunoPrecise Antibodies and LiteVax BV have come together and developed a vaccine candidate using the epitopes screened from the spike surface glycoprotein (Milken Institute, 2020; Release, 2020; World Health Organization, 2020). The St. Petersburg Research Institute of Vaccines and Sera vaccine candidate belongs to the Protein Subunit, Recombinant Protein, Nanoparticles category and utilizes S-protein and other epitopes. This next generation candidate vaccine has been included in the list of promising vaccines by the WHO (Milken Institute, 2020; Release, 2020; World Health Organization, 2020). The University of Bristol and its associated company Imophoron have established several leads based on Imophoron'sADDomer© platform which is a synthetic, self-assembling, virus-like particle (VLP) for delivery of multiple immunogenic epitopes from SARS CoV-2 (Milken Institute, 2020; World Health Organization, 2020). The COVAXX's vaccine candidate is the first and only truly multi epitope peptide vaccine (MEPV) designed to provide protection against Covid-19. The COVAXX vaccine targets a crucial antigen from the RBD of the spike glycoprotein, along with other viral epitopes (from additional structural proteins) to promote both humoral and cell-mediated memory responses. This vaccine has progressed to the phase I trials in Sep, 2020 (Milken Institute, 2020; Solution & Status, 2021; World Health Organization, 2020).

The development of an efficient and safe vaccine against any particular pathogen, by utilizing the traditional methods of vaccine generation, is a time-consuming process. Hence, *in silico* approaches for vaccine design are gaining interest, as they can potentially expedite the vaccine development process, by providing lead constructs that can be further verified for desired characteristics in future experimental studies. Computational immunology or immunoinformatics, utilizes computational strategies to investigate different immunological events, along with the development of algorithms for identifying potent B and T cell epitopes. The application of immunoinformatics for designing multi-epitope subunit vaccines, is a cost effective modus operandi against infectious pathogens, displaying high degrees of specificity, safety and stability values (De Groot et al., 2002; Nosrati et al., 2017; Patronov & Doytchinova, 2013; Suhrbier, 1997). Computational vaccinology technique has proven its value against various infectious diseases. Utilizing the immunoinformatics strategies, many effective vaccines against pathogens like *Chlamydia pneumoniae*, *Streptococcus pneumonia*

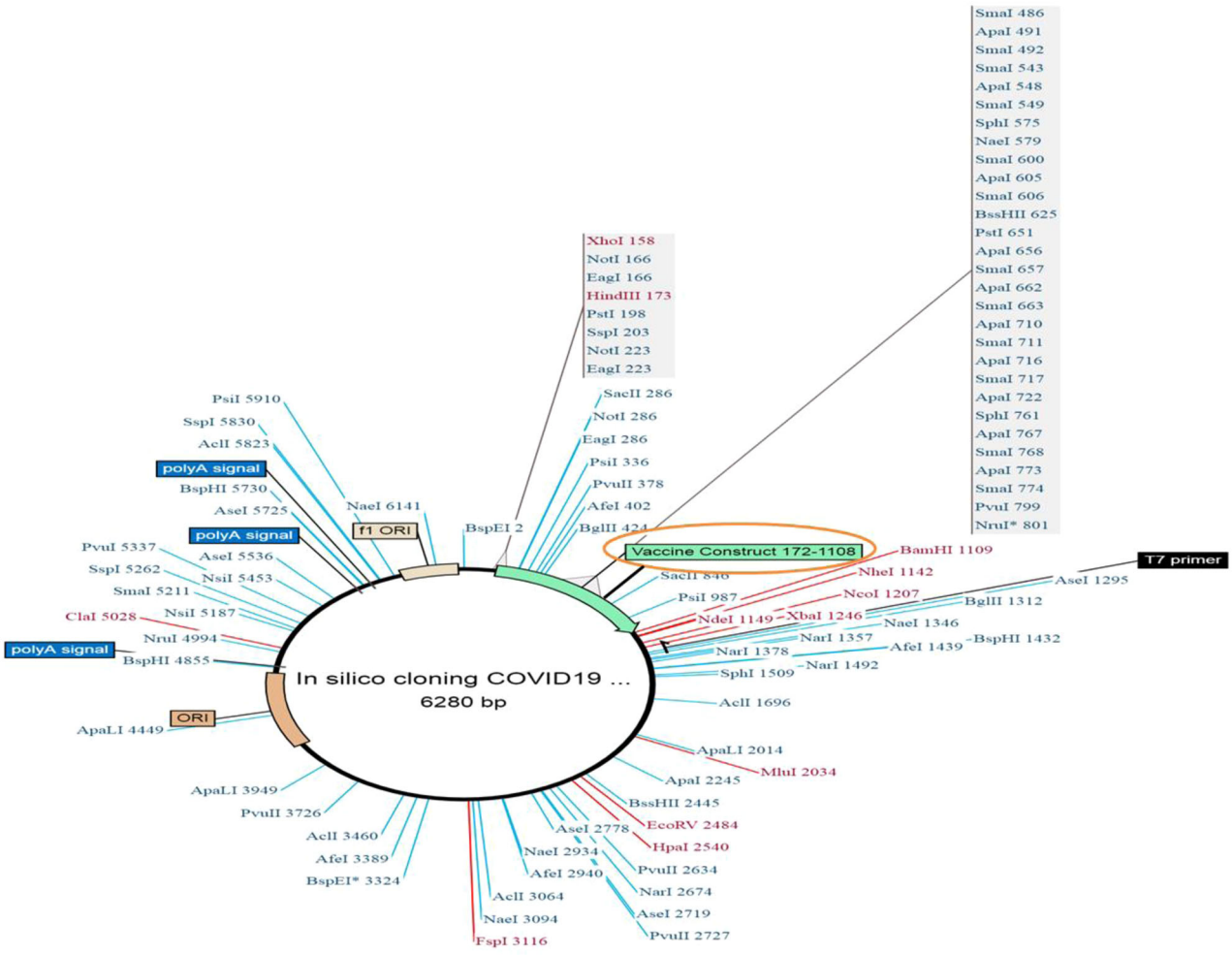

Figure 7. Recombinant pET28a(+) expression vector carrying the vaccine construct.

(Pourhajibagher & Bahador, 2016), *Rickettsia prowazekii* (Caro-Gomez et al., 2014), enterotoxigenic *E. coli* (Mehla & Ramana, 2016) and *Staphylococcus aureus* (Delfani et al., 2015) have emerged. Several groups have utilized this approach in developing candidate vaccines to combat Covid-19. Most of these immunoinformatics studies have focussed on prediction of epitopes from a single protein, majorly the surface spike glycoprotein (Abraham Peele et al., 2020; Bhatnager et al., 2020; Bhattacharya et al., 2020; Kar et al., 2020; Naz et al., 2020; Rahman et al., 2020; Samad et al., 2020; Sanami et al., 2020) or a single category of proteins, either structural or non-structural (Kalita et al., 2020; Ojha et al., 2020; Sarkar & Ullah, 2020); relatively few are focused on a candidate that utilizes both these classes of proteins (Ahmad et al., 2020; Chauhan et al., 2021; Enayatkhani et al., 2021; Srivastava et al., 2020; Tahir et al., 2020).Some others have employed only CTL epitopes leaving the significant B-cell or HTL epitopes (Mishra, 2020). In the present study we have proposed a non allergenic and non toxic vaccine construct with acceptable physiological properties, which is composed of highly immunogenic B and T cell epitopes from two structural and three non-structural proteins of SARS-CoV-2. In comparison to our study with other similar reports where both classes of proteins have been utilized (Ahmad et al., 2020; Chauhan et al., 2021; Enayatkhani et al., 2021; Srivastava et al., 2020), we found that our proposed candidate exhibits enhanced immunogenic potential on account of its predicted antigenicity score. Moreover, the combinations of proteins/epitopes that we have utilized have not been explored by any other immunoinformatics study so far.

In the present study, we have exploited immunoinformatics to delineate a novel multi-epitope subunit vaccine against the deadly SARS-Cov-2. For this purpose, five vital proteins displaying significant immunogenicity and no homology to humans were selected. Although helicase displayed 22.4% sequence identity to the human protein ZGRF1 isoform, but since the chances of cross reactivity are negligible when the sequence similarity is below 40% between the pathogen and human (Ojha et al., 2020), helicase was considered for the study. Similarly, the nucleoprotein displayed 55.56% identity with the human immunoglobin heavy chain junction region, but with a meagre 4% query coverage value and a high E value of 5.4, this similarity was considered insignificant. In order to generate an effective humoral and cellular immune response by a vaccine candidate, the epitopes ought to be accessible to B cells as well as MHC I and MHC II molecules (Patra et al., 2020). Therefore, B and T cell epitopes were mapped using online web servers. In order to cover

largest achievable population coverage, the full HLA reference set was selected for HTL epitopes identification and supertypes A2, A3 and B7, which jointly enclose about 88% population worldwide were chosen for CTL determination (Ojha et al., 2020; Sette & Sidney, 1999; Shen et al., 2018). Based on either ranking or highest score (B cell), lowest percentile rank and IC_{50} value (HTL) and high score (CTL), the best possible epitopes were selected. These epitopes were once again scrutinized for antigenicity individually, and only those epitopes which displayed significant antigenicity potential were carried forward in the investigation. These antigenic epitopes were further evaluated for toxicity. . Fortunately, none of the selected epitope displayed toxicity potential. . These epitopes were then used to generate the vaccine construct using appropriate linkers. The vaccine construct was also evaluated for its antigenicity along with their physicochemical properties and allergenicity. The proposed vaccine construct displayed significant antigenicity using two online web servers (VaxiJen and ANTIGENpro), and did not exhibit any unacceptable characteristics. The evaluation of trans-membrane forming regions in the vaccine construct, which represents an inaccessible portion for B cells, showed absence of any such domains. The MHC class-I alleles and MHC class-II alleles cluster analysis was performed to functionally cluster them in accordance with their predicted binding specificity and to infer their association with each other. Moreover, the population coverage and conservancy analysis suggested that the proposed vaccine construct would be efficient for the entire world population infected with different SARS-CoV-2 isolates.

The structure corresponding to the protein sequence of the vaccine candidate was envisaged by 3Dpro tool, which showed it to have majorly an alpha helical conformation. The generated structure was refined to remove biophysical and topological errors via energy minimization through Modeller 9.24. The refined structure so generated was observed to be convincing for further computational experiments. The ProSA Z score was noticed to be within the distinctive range for a native protein. The Ramachandran plot generated through PROCHECK validated the quality of the structure. The modelled vaccine candidate displayed only 0.4% of the residues in disallowed regions, thereby indicating minimum steric hindrance created among side chain and main chain atoms. Further, the ERRAT server analysis, which identifies the arrangement of non-bonded atomic interactions, was carried out. The ERRAT score (93.233), which is an indicator of overall quality suggests that a high quality model is produced.

Previous reports suggest that mice in which TLR-3 and TLR-4 are absent, are more prone to SARS-CoV contagion when compared to the mice possessing them (Totura et al., 2015). The TLR-3 trigger could suppress the production of neurotropic MHV-A59, MHV-JHM, and MHV-3 strains of Mouse Hepatitis Virus (MHV), the model of betacoronaviruses, conventionally used to study human diseases. TLR 3 is also known to play crucial role in providing protection against several viruses such as hepatitis B, human simplex virus and lethal influenza virus (Mazaleuskaya et al., 2012). Additionally, it is observed that TLR-3 recognises viral RNA that is produced during the replication cycle of most viruses and this activation of TLR-3 further triggers appropriate host antiviral immune responses (Christopher & Wong, 2012; Iwasaki & Yang, 2020). The role of TLR3 is also particularly important because, unlike the other TLR's it does not utilize a common signalling pathway for triggering anti-viral activities, and thus can provide a broader functional innate immune response (Christopher & Wong, 2012). Based on these observations, the vaccine construct was evaluated for the likelihood of its interaction with TLR-3 using the HADDOCK server. On performing docking, 6 clusters of the docked complexes were obtained. Cluster 1 with the best statistics was observed to have a Z score of -1.9, the lowest value attained among all the clusters, indicating a high efficiency binding. In sum, the designed vaccine construct/protein was successful in binding with the human TLR-3 immune receptor with a good docking score. Furthermore, the docking complex of TLR-3 and designed vaccine got stabilized within $\sim$10ns into the simulation, and the 20 ns simulation trajectory conformation displayed negligible topological variations in the range of 0.7–0.8 nm. The residue network of the interacting substructure of TLR-3 and the designed vaccine was found quite similar during and after the run, and it reliably confirms the topological stability. The increased compactness of the complex indicates a strong binding of TLR-3 and the designed vaccine. The difference between the exposed and buried surface area and a strong hydrogen-bonding network further affirms the significant potential of the designed vaccine.

By carrying out the immune simulation experiment, which mimics the innate framework of an immune system, the immune response from the major stakeholders of the immune system namely, B cells and T cells which are involved in humoral and cellular responses, respectively, at a single dose against the SARS-Cov-2 multi-epitope subunit vaccine was affirmed (Ojha et al., 2020). The outcome from this study indicates a powerful immune reaction by the proposed vaccine construct. Finally, in order to push the proposed vaccine construct towards laboratory evaluation, it is crucial that the optimum-level of protein expression in hosts such as *E. coli* could be achieved. Therefore, CAI, CFD, and GC content of the gene are some of the variables that need to be optimized. Our results very well support the proposed vaccine construct gene in this regard (Nosrati et al., 2019).

5. Conclusions

In sum, the study indicates that the proposed vaccine candidate bears competent structural, acceptable physiochemical and desirable immunological attributes that can elicit a powerful humoral and cellular immune reaction and therefore should be pushed as a prospective lead candidate for further *in vitro* and *in vivo* investigations against SARS-CoV-2.

Acknowledgements

VSR is thankful to Dept. of Health Research (Govt. of India) for the young scientist fellowship (No. 12014/13/2018-HR/E-Office: 3151268). RS and AK are contended to Indian Council of Medical Research (ICMR), India for the SRF position. AG is thankful to University Grant Commission, India for the faculty Recharge position. NV is thankful to DST Inspire fellowship for research grant.

Disclosure statement

The authors declare no conflicting interests.

Funding

This research did not receive specific grant form funding agencies in the public, commercial, or not-for-profit sectors.

ORCID

Vijay Prajapati http://orcid.org/0000-0001-6510-0596
Abhinav Grover http://orcid.org/0000-0002-3296-7860

Data availability statement

All datasets generated for this report are included in the article/Supplementary Material.

Contributions

VSR conceptualized and visualized the whole study. VSR and RS designed and conducted all the studies pertaining to epitope identification and selection. AK carried out the structure generation, *in silico* docking and MD studies. NV assisted in data curation and manuscript writing. VP guided the study and reviewed the manuscript. AG managed resources and reviewed the manuscript. VSR revised the manuscript after peer review. All authors read and approved the manuscript.

References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindah, 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

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 and Dynamics*, 1–9. https://doi.org/10.1080/07391102.2020.1770127

Adepoju, P. (2019). RTS,S malaria vaccine pilots in three African countries. *The Lancet)*, *393*(10182), 1685. https://doi.org/10.1016/S0140-6736(19)30937-7

Ahmad, S., Navid, A., Farid, R., Abbas, G., Ahmad, F., Zaman, N., Parvaiz, N., & Azam, 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*, *151*(March), 105387. https://doi.org/10.1016/j.ejps.2020.105387

Amanat, F., & Krammer, F. (2020). SARS-CoV-2 vaccines: Status report. *Immunity*, *52*(4), 583–589. https://doi.org/10.1016/j.immuni.2020.03.007

Angeletti, S., Benvenuto, D., Bianchi, M., Giovanetti, M., &, Pascarella, S., & Ciccozzi, M. (2020). COVID-2019 the role of the nsp2 and nsp3 in its pathogenesis. *Journal of Medical Virology*, *92*, 584–588.

Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, *81*(8), 3684–3690. https://doi.org/10.1063/1.448118

Bhatnager, R., Bhasin, M., Arora, J., & Dang, A. S. (2020). Epitope based peptide vaccine against SARS-COV2: An immune-informatics approach. *Journal of Biomolecular Structure and Dynamics*, 1–16. https://doi.org/10.1080/07391102.2020.1787227

Bhattacharya, M., Sharma, A. R., Patra, P., Ghosh, P., Sharma, G., Patra, B. C., Lee, S. S., & Chakraborty, C. (2020). Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. *Journal of Medical Virology*, *92*(6), 618–631. https://doi.org/10.1002/jmv.25736

Bhattacharya, M., Sharma, A. R., Sharma, G., Patra, P., Mondal, N., Patra, B. C., Lee, S. S., & Chakraborty, C. (2020). Computer aided novel antigenic epitopes selection from the outer membrane protein sequences of *Aeromonas hydrophila* and its analyses. *Infection, Genetics and Evolution*, *82*, 104320. https://doi.org/10.1016/j.meegid.2020.104320

Bianchi, M., Benvenuto, D., Giovanetti, M., Angeletti, S., & Pascarella, S. (2020). Sars-CoV-2 Envelope and membrane proteins: Differences from closely related proteins linked to cross-species transmission? *BioMed Research International*, *2020*, 1–14. https://doi.org/10.20944/preprints202004.0089.v1

Bui, H. H., Sidney, J., Dinh, K., Southwood, S., Newman, M. J., & Sette, A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*, *7*(1), 153. https://doi.org/10.1186/1471-2105-7-153

Bui, H. H., Sidney, J., Li, W., Fusseder, N., & Sette, A. (2007). Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics*, *8*, 361–366. https://doi.org/10.1186/1471-2105-8-361

Caro-Gomez, E., Gazi, M., Goez, Y., & Valbuena, G. (2014). Discovery of novel cross-protective *Rickettsia prowazekii* T-cell antigens using a combined reverse vaccinology and in vivo screening approach. *Vaccine*, *32*(39), 4968–4976. https://doi.org/10.1016/j.vaccine.2014.06.089

Cavanagh, D. (2005). Coronaviridae: A review of coronaviruses and toroviruses. In A. Schmidt, O. Weber, & M. H. Wolff (Eds.), *Coronaviruses with special emphasis on first insights concerning SARS* (pp. 1–54). Birkhäuser Basel. https://doi.org/10.1007/3-7643-7339-3_1

Chauhan, V., Rungta, T., Rawat, M., Goyal, K., Gupta, Y., & Singh, M. P. (2021). Excavating SARS-coronavirus 2 genome for epitope-based subunit vaccine synthesis using immunoinformatics approach. *Journal of Cellular Physiology*, *236*(2), 1131–1147. https://doi.org/10.1002/jcp.29923

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 1), 12–21. https://doi.org/10.1107/S0907444909042073

Chen, W. H., Strych, U., Hotez, P. J., & Bottazzi, M. E. (2020). The SARS-CoV-2 vaccine pipeline: An overview. *Current Tropical Medicine Reports*, *7*(2), 61–64. https://doi.org/10.1007/s40475-020-00201-6

Cheng, J., Randall, A. Z., Sweredoski, M. J., & Baldi, P. (2005). SCRATCH: A protein structure and structural feature prediction server. *Nucleic Acids Research*, *33*(Web Server), W72–W76. https://doi.org/10.1093/nar/gki396

Christopher, M. E., & Wong, J. P. (2012). Use of toll-like receptor 3 agonists against respiratory viral infections. *Anti-Inflammatory & Anti-Allergy Agents in Medicinal Chemistry*, *10*(5), 327–338. https://doi.org/10.2174/1871523011109050327

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*(12), 10089–10092. https://doi.org/10.1063/1.464397

De Groot, A. S., Sbai, H., Aubin, C. S., McMurry, J., & Martin, W. (2002). Immuno-informatics: Mining genomes for vaccine components. *Immunology and Cell Biology*, *80*(3), 255–269. https://doi.org/10.1046/j.1440-1711.2002.01092.x

Delfani, S., Imani Fooladi, A. A., Mobarez, A. M., Emaneini, M., Amani, J., & Sedighian, H. (2015). In silico analysis for identifying potential vaccine candidates against *Staphylococcus aureus*. *Clinical and Experimental Vaccine Research*, *4*(1), 99–106. https://doi.org/10.7774/cevr.2015.4.1.99

Dimitrov, I., Naneva, L., Doytchinova, I., & Bangov, I. (2014). AllergenFP: Allergenicity prediction by descriptor fingerprints. *Bioinformatics*, *30*(6), 846–851. https://doi.org/10.1093/bioinformatics/btt619

Emmanuel, C. (2020). Coronaviruses with SARS-CoV-2 (Causative agent for COVID-19) as a case study, 2(March). https://doi.org/10.13140/RG.2.2.31080.42242

Enayatkhani, M., Hasaniazad, M., Faezi, S., Gouklani, H., Davoodian, P., Ahmadi, N., Einakian, M. A., Karmostaji, A., & Ahmadi, K. (2021). Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: An in silico study. *Journal of Biomolecular Structure and Dynamics*, *39*(8), 2857–2816. https://doi.org/10.1080/07391102.2020.1756411

Farhani, I. (2019). Multi-epitope vaccines (MEVs), as a novel strategy against infectious diseases. *Current Proteomics*, *17*(5), 354–364. https://doi.org/10.2174/1570164617666190919120140

Garg, S., Kakkar, S., & Runthala, A. (2016). Improved protein model ranking through topological assessment. In K-.C. Wong (Ed.), *Computational Biology and Bioinformatics* (pp. 406–424). CRC Press. https://doi.org/10.1201/b20026-24

Gupta, N., Regar, H., Verma, V. K., Prusty, D., Mishra, A., & Prajapati, V. K. (2020). Receptor-ligand based molecular interaction to discover adjuvant for immune cell TLRs to develop next-generation vaccine. *International Journal of Biological Macromolecules*, *152*, 535–545. https://doi.org/10.1016/j.ijbiomac.2020.02.297

Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, *18*(12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hilgenfeld, R., & Peiris, M. (2013). From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. *Antiviral Research*, *100*(1), 286–295. https://doi.org/10.1016/j.antiviral.2013.08.015

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38. https://doi.org/10.1016/j.carbon.2017.07.012

Iwasaki, A., & Yang, Y. (2020). The potential danger of suboptimal antibody responses in COVID-19. *Nature Reviews. Immunology*, *20*(6), 339–341. https://doi.org/10.1038/s41577-020-0321-6

Jiang, S., Bottazzi, M. E., Du, L., Lustigman, S., Tseng, C. T. K., Curti, E., Jones, K., Zhan, B., & Hotez, P. J. (2012). Roadmap to developing a recombinant coronavirus S protein receptor-binding domain vaccine for severe acute respiratory syndrome. *Expert Review of Vaccines*, *11*(12), 1405–1413. https://doi.org/10.1586/erv.12.126

Kalita, P., Padhi, A. K., Zhang, K. Y. J., & Tripathi, T. (2020). Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information. January.

Kao, D. J., & Hodges, R. S. (2009). Advantages of a synthetic peptide immunogen over a protein immunogen in the development of an anti-pilus vaccine for Pseudomonas aeruginosa. *Chemical Biology & Drug Design*, *74*(1), 33–42. https://doi.org/10.1111/j.1747-0285.2009.00825.x

Kar, T., Narsaria, U., Basak, S., Deb, D., Castiglione, F., Mueller, D. M., & Srivastava, A. P. (2020). A candidate multi-epitope vaccine against SARS-CoV-2. *Scientific Reports*, *10*(1), 1–24. https://doi.org/10.1038/s41598-020-67749-1

Knapp, B., Frantal, S., Cibena, M., Schreiner, W., & Bauer, P. (2011). Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible? *Journal of Computational Biology*, *18*(8), 997–1005. https://doi.org/10.1089/cmb.2010.0237

Laurens, M. B. (2019). RTS,S/AS01 vaccine (Mosquirix™): An overview. *Human Vaccines and Immunotherapeutics*, *16*(3), 480–489. https://doi.org/10.1080/21645515.2019.1669415

Li, W., Joshi, M. D., Singhania, S., Ramsey, K. H., & Murthy, A. K. (2014). Peptide vaccine: Progress and challenges. *Vaccines*, *2*(3), 515–536. https://doi.org/10.3390/vaccines2030515

Livingston, B., Crimi, C., Newman, M., Appella, E., Sidney, J., Livingston, B., Crimi, C., Newman, M., Higashimoto, Y., Appella, E., Sidney, J., & Sette, A. (2002). A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. *Journal of Immunology*, *168*(11), 5499–5506. https://doi.org/10.4049/jimmunol.168.11.5499

Martin, J., Letellier, G., Marin, A., Taly, J. F., De Brevern, A. G., & Gibrat, J. F. (2005). Protein secondary structure assignment revisited: A detailed analysis of different assignment methods. *BMC Structural Biology*, *5*, 17. https://doi.org/10.1186/1472-6807-5-17

Mazaleuskaya, L., Veltrop, R., Ikpeze, N., Martin-Garcia, J., & Navas-Martin, S. (2012). Protective role of toll-like receptor 3-induced type I interferon in murine coronavirus infection of macrophages. *Viruses*, *4*(5), 901–923. https://doi.org/10.3390/v4050901

Mehla, K., & Ramana, J. (2016). Identification of epitope-based peptide vaccine candidates against enterotoxigenic *Escherichia coli*: A comparative genomics and immunoinformatics approach. *Molecular BioSystems*, *12*(3), 890–901. https://doi.org/10.1039/c5mb00745c

Milken Institute. (2020). *COVID-19 treatment and vaccine tracker*, 1–45. https://milkeninstitute.org/sites/default/files/2020-03/Covid19TrackerNEW3-24-20-REVISED.pdf

Mirza, M. U., & Froeyen, M. (2020). Structural elucidation of SARS-CoV-2 vital proteins: Computational methods reveal potential drug candidates against main protease, Nsp12 polymerase and Nsp13 helicase. *Journal of Pharmaceutical Analysis*, *10*(4), 320–328. https://doi.org/10.1016/j.jpha.2020.04.008

Mishra, D. (2020). *T cell epitope-based vaccine design for pandemic novel coronavirus 2019-nCoV*. https://doi.org/10.26434/chemrxiv.12029523

Mohammadi, S., Mostafavi-Pour, Z., Ghasemi, Y., Barazesh, M., Kavousi Pour, S., Atapour, A., Mokarram, P., & Morowvat, M. H. (2019). In silico analysis of different signal peptides for the excretory production of recombinant NS3-GP96 fusion protein in *Escherichia coli*. *International Journal of Peptide Research and Therapeutics*, *25*(4), 1279–1290. https://doi.org/10.1007/s10989-018-9775-9

Mohan, T., Sharma, C., Bhat, A. A., & Rao, D. N. (2013). Modulation of HIV peptide antigen specific cellular immune response by synthetic α- and β-defensin peptides. *Vaccine*, *31*(13), 1707–1716. https://doi.org/10.1016/j.vaccine.2013.01.041

Mu, J., Xu, J., Zhang, L., Shu, T., Wu, D., Huang, M., Ren, Y., Li, X., Geng, Q., Xu, Y., Qiu, Y., & Zhou, X. (2020). SARS-CoV-2-encoded nucleocapsid protein acts as a viral suppressor of RNA interference in cells. *Science China Life Sciences*, *63*(9), 1413–1416. https://doi.org/10.1007/s11427-020-1692-1

Naik, B., Gupta, N., Ojha, R., Singh, S., Prajapati, V. K., & Prusty, D. (2020). High throughput virtual screening reveals SARS-CoV-2 multi-target binding natural compounds to lead instant therapy for COVID-19 treatment. *International Journal of Biological Macromolecules*, *160*, 1–17. https://doi.org/10.1016/j.ijbiomac.2020.05.184

Naz, A., Shahid, F., Butt, T. T., Awan, F. M., Ali, A., & Malik, A. (2020). Designing multi-epitope vaccines to combat emerging coronavirus disease 2019 (COVID-19) by employing immuno-informatics approach. *Frontiers in Immunology*, *11*, 1663–1613. https://doi.org/10.3389/fimmu.2020.01663

Nezafat, N., Karimi, Z., Eslami, M., Mohkam, M., Zandian, S., & Ghasemi, Y. (2016). Designing an efficient multi-epitope peptide vaccine against Vibrio cholerae via combined immunoinformatics and protein interaction based approaches. *Computational Biology and Chemistry*, *62*, 82–95. https://doi.org/10.1016/j.compbiolchem.2016.04.006

Nieto-Torres, J. L., DeDiego, M. L., Álvarez, E., Jiménez-Guardeño, J. M., Regla-Nava, J. A., Llorente, M., Kremer, L., Shuo, S., & Enjuanes, L. (2011). Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. *Virology*, *415*(2), 69–82. https://doi.org/10.1016/j.virol.2011.03.029

Nosrati, M., Hajizade, A., Nazarian, S., Amani, J., Namvar Vansofla, A., & Tarverdizadeh, Y. (2019). Designing a multi-epitope vaccine for cross-protection against Shigella spp: An immunoinformatics and structural

vaccinology study. *Molecular Immunology*, *116*, 106–116. https://doi.org/10.1016/j.molimm.2019.09.018

Nosrati, M., Mohabatkar, H., & Behbahani, M. (2017). A novel multi-epitope vaccine for cross protection against hepatitis C virus (HCV): An immunoinformatics approach. *Research in Molecular Medicine*, *5*(1), 17–26. https://doi.org/10.29252/rmm.5.1.17

Ojha, R., Gupta, N., Naik, B., Singh, S., Verma, V. K., Prusty, D., & Prajapati, V. K. (2020). High throughput and comprehensive approach to develop multiepitope vaccine against minacious COVID-19. *European Journal of Pharmaceutical Sciences*, *151*, 105375. https://doi.org/10.1016/j.ejps.2020.105375

Ojha, R., Nandani, R., & Prajapati, V. K. (2019). Contriving multiepitope subunit vaccine by exploiting structural and nonstructural viral proteins to prevent Epstein-Barr virus-associated malignancy. *Journal of Cellular Physiology*, *234*(5), 6437–6448. https://doi.org/10.1002/jcp.27380

Oyarzún, P., & Kobe, B. (2016). Recombinant and epitope-based vaccines on the road to the market and implications for vaccine design and production. *Human Vaccines & Immunotherapeutics*, *12*(3), 763–767. https://doi.org/10.1080/21645515.2015.1094595

Patra, P., Mondal, N., Patra, B. C., & Bhattacharya, M. (2020). Epitope-based vaccine designing of nocardia asteroides targeting the virulence factor Mce-family protein by immunoinformatics approach. *International Journal of Peptide Research and Therapeutics*, *26*(2), 1165–1176. https://doi.org/10.1007/s10989-019-09921-4

Patronov, A., & Doytchinova, I. (2013). T-cell epitope vaccine design by immunoinformatics. *Open Biology*, *3*(1), 120139. https://doi.org/10.1098/rsob.120139

Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. *Current Protocols in Bioinformatics*, *42*(1), 1–8. https://doi.org/10.1002/0471250953.bi0301s42

Perlman, S., & Netland, J. (2009). Coronaviruses post-SARS: Update on replication and pathogenesis. *Nature Reviews. Microbiology*, *7*(6), 439–450. https://doi.org/10.1038/nrmicro2147

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera – A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Pourhajibagher, M., & Bahador, A. (2016). Designing and in silico analysis of PorB protein from *Chlamydia trachomatis* for developing a vaccine candidate. *Drug Research*, *66*(9), 479–483. https://doi.org/10.1055/s-0042-110319

Pourseif, M. M., Yousefpour, M., Aminianfar, M., Moghaddam, G., & Nematollahi, A. (2019). A multi-method and structure-based in silico vaccine designing against Echinococcus granulosus through investigating enolase protein. *BioImpacts*, *9*(3), 131–144. https://doi.org/10.15171/bi.2019.18

Rahman, N., Ali, F., Basharat, Z., Shehroz, M., Khan, M. K., Jeandet, P., Nepovimova, E., Kuca, K., & Khan, H. (2020). Vaccine design from the ensemble of surface glycoprotein epitopes of SARS-CoV-2: An immunoinformatics approach. *Vaccines*, *8*(3), 423–417. https://doi.org/10.3390/vaccines8030423

Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One*, *5*(4), e9862. https://doi.org/10.1371/journal.pone.0009862

Ratnadeep Saha, B. V. L. S. P. (2020). In silico approach for designing of a multi-epitope based vaccine against novel Coronavirus (SARS-COV-2). *BioRxiv*, *9*(5–6), 10–16.

Release, P. (2020). Collaborative COVID-19 vaccine project between immunoprecise antibodies Europe and LiteVax BV, funded through TRANSVAC2, 1–5.

Riou, J., & Althaus, C. L. (2020). Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance*, *25*(4), 1–5. https://doi.org/10.2807/1560-7917.ES.2020.25.4.2000058

Runthala, A. (2012). Short review: Protein structure prediction: Challenging targets for CASP10. *Journal of Biomolecular Structure & Dynamics*, *30*(5), 607–615. https://doi.org/10.1080/07391102.2012.687526

Runthala, A., & Chowdhury, S. (2014). Iterative optimal TM-score and Z-score guided sampling significantly improves model topology. *Lecture Notes in Engineering and Computer Science*, *2209*(January), 123–128.

Runthala, A., & Chowdhury, S. (2019). Refined template selection and combination algorithm significantly improves template-based modeling accuracy. *Journal of Bioinformatics and Computational Biology*, *17*(2), 1950006–1950021. https://doi.org/10.1142/S0219720019500069

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 and Dynamics*, 1–17. https://doi.org/10.1080/07391102.2020.1792347

Sanami, S., Zandi, M., Pourhossein, B., Mobini, G., Safaei, M., & Abed, A. (2020). Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 Resource Centre is Hosted on Elsevier Connect, the Company' s Public News and Information, January.

Sarkar, B., Ullah, A., Tuz, F., Afrin, M., & Araf, Y. (2020). Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19. The COVID-19 Resource Centre is Hosted on Elsevier Connect, the Company's Public News and Information, January.

Sarkar, B., & Ullah, M. A. (2020). Designing novel subunit vaccines against herpes simplex virus-1 using reverse vaccinology approach. BioRxiv, 2020.01.10.901678. https://doi.org/10.1101/2020.01.10.901678

Schuerman, L. (2019). RTS,S malaria vaccine could provide major public health benefits. *The Lancet*, *394*(10200), 735–736. https://doi.org/10.1016/S0140-6736(19)31567-3

Sette, A., & Sidney, J. (1999). Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics*, *50*(3-4), 201–212. https://doi.org/10.1007/s002510050594

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., & Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, *581*(7807), 221–224. https://doi.org/10.1038/s41586-020-2179-y

Shen, W. J., Zhang, X., Zhang, S., Liu, C., & Cui, W. (2018). The utility of supertype clustering in prediction for class II MHC-peptide binding. *Molecules*, *23*(11), 3034–3018. https://doi.org/10.3390/molecules23113034

Skwarczynski, M., & Toth, I. (2016). Peptide-based synthetic vaccines. *Chemical Science*, *7*(2), 842–854. https://doi.org/10.1039/c5sc03892h

Solution, T., & Status, D. (2021). *COVAXX vaccine technology – Fact sheet*.

Srivastava, S., Verma, S., Kamthania, M., Kaur, R., Badyal, R. K., Saxena, A. K., Shin, H.-J., Kolbe, M., & Pandey, K. C. (2020). Structural basis for designing multiepitope vaccines against COVID-19 infection: In silico vaccine design and validation. *JMIR Bioinformatics and Biotechnology*, *1*(1), e19371. https://doi.org/10.2196/19371

Suhrbier, A. (1997). Multi-epitope DNA vaccines. *Immunology and Cell Biology*, *75*(4), 402–408. https://doi.org/10.1038/icb.1997.63

Tahir, M., Rehman, A., Tusleem, K., Ashfaq, A., Qasim, M., Zhu, X., Fatima, I., Shahid, F., & Chen, L. (2020). Designing of a next generation multiepitope based vaccine (MEV) against SARS-COV-2: Immunoinformatics and in silico approaches. *PLoS One*, *2019*, 1–25. https://doi.org/10.1371/journal.pone.0244176

Thomsen, M., Lundegaard, C., Buus, S., Lund, O., & Nielsen, M. (2013). MHCcluster, a method for functional clustering of MHC molecules. *Immunogenetics*, *65*(9), 655–665. https://doi.org/10.1007/s00251-013-0714-9

Totura, A. L., Whitmore, A., Agnihothram, S., Schäfer, A., Katze, M. G., Heise, M. T., & Baric, R. S. (2015). Toll-like receptor 3 signaling via TRIF contributes to a protective innate immune response to severe acute respiratory syndrome coronavirus infection. *MBio*, *6*(3), 1–14. https://doi.org/10.1128/mBio.00638-15

Urrutia-Baca, V. H., Gomez-Flores, R., De La Garza-Ramos, M. A., Tamez-Guerra, P., Lucio-Sauceda, D. G., & Rodríguez-Padilla, M. C. (2019). Immunoinformatics approach to design a novel epitope-based oral

vaccine against *Helicobacter pylori*. *Journal of Computational Biology*, *26*(10), 1177–1190. https://doi.org/10.1089/cmb.2019.0062

Van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., Melquiond, A. S. J., Van Dijk, M., De Vries, S. J., & Bonvin, A. M. J. J. (2016). The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. *Journal of Molecular Biology*, *428*(4), 720–725. https://doi.org/10.1016/j.jmb.2015.09.014

Venkatagopalan, P., Daskalova, S. M., Lopez, L. A., Dolezal, K. A., & Hogue, B. G. (2015). Coronavirus envelope (E) protein remains at the site of assembly. *Virology*, *478*, 75–85. https://doi.org/10.1016/j.virol.2015.02.005

Wang, Q., Zhang, L., Kuwahara, K., Li, L., Liu, Z., Li, T., Zhu, H., Liu, J., Xu, Y., Xie, J., Morioka, H., Sakaguchi, N., Qin, C., & Liu, G. (2016). Immunodominant SARS coronavirus epitopes in humans elicited both enhancing and neutralizing effects on infection in non-human primates. *ACS Infectious Diseases*, *2*(5), 361–376. https://doi.org/10.1021/acsinfecdis.6b00006

Webb, B., & Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics*, *54*(1), 5.6.1–5.6.37. https://doi.org/10.1002/cpbi.3

World Health Organization. (2020). Draft landscape of COVID-19 candidate vaccines – 15 May 2020. *WHO*, *March*, 3. https://www.who.int/who-documents-detail/draft-landscape-of-covid-19-candidate-vaccines

Yang, Y., Sun, W., Guo, J., Zhao, G., Sun, S., Yu, H., Guo, Y., Li, J., Jin, X., Du, L., Jiang, S., Kou, Z., & Zhou, Y. (2015). In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations. *Human Vaccines & Immunotherapeutics*, *11*(3), 795–805. https://doi.org/10.1080/21645515.2015.1012017

Yano, A., Onozuka, A., Asahi-Ozaki, Y., Imai, S., Hanada, N., Miwa, Y., & Nisizawa, T. (2005). An ingenious design for peptide vaccines. *Vaccine*, *23*(17-18), 2322–2326. https://doi.org/10.1016/j.vaccine.2005.01.031

Yoshimoto, F. K. (2020). The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. *The Protein Journal*, *39*(3), 198–216. https://doi.org/10.1007/s10930-020-09901-4