EXPERT REVIEW OF VACCINES
https://doi.org/10.1080/14760584.2021.1874925

Taylor & Francis
Taylor & Francis Group

ORIGINAL RESEARCH

Check for updates

Immunoinformatics-guided designing and in silico analysis of epitope-based polyvalent vaccines against multiple strains of human coronavirus (HCoV)

Bishajit Sarkar [a], Md. Asad Ullah [a], Yusha Araf [b], Nafisa Nawal Islam [a] and Umme Salma Zohora [a]

[a]Department of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Jahangirnagar University, Savar, Dhaka, Bangladesh; [b]Department of Genetic Engineering and Biotechnology, School of Life Sciences, Shahjalal University of Science and Technology, Sylhet, Bangladesh

ABSTRACT

Objectives: The group of human coronaviruses (HCoVs) consists of some highly pathogenic viruses that have caused several outbreaks in the past. The newly emerged strain of HCoV, the SARS-CoV-2 is responsible for the recent global pandemic that has already caused the death of hundreds of thousands of people due to the lack of effective therapeutic options.

Methods: In this study, immunoinformatics methods were used to design epitope-based polyvalent vaccines which are expected to be effective against four different pathogenic strains of HCoV i.e., HCoV-OC43, HCoV-SARS, HCoV-MERS, and SARS-CoV-2.

Results: The constructed vaccines consist of highly antigenic, non-allergenic, nontoxic, conserved, and non-homologous T-cell and B-cell epitopes from all the four viral strains. Therefore, they should be able to provide strong protection against all these strains. Protein-protein docking was performed to predict the best vaccine construct. Later, the MD simulation and immune simulation of the best vaccine construct also predicted satisfactory results. Finally, in silico cloning was performed to develop a mass production strategy of the vaccine.

Conclusion: If satisfactory results are achieved in further in vivo and in vitro studies, then the vaccines designed in this study might be effective as preventative measures against the selected HCoV strains.

ARTICLE HISTORY
Received 21 April 2020
Accepted 8 January 2021

KEYWORDS
Coronaviruses; sars-CoV; mers-CoV; sars-CoV-2; vaccines; immunoinformatics

1. Introduction

Coronaviruses are a group of pathogenic viruses that mainly infect mammals and birds. These viruses cause diseases in the respiratory tract of humans, ranging from the common cold in otherwise healthy individuals to more serious and lethal conditions and even death [1,2]. The coronavirus family, also known as *Coronaviridae*, is the largest family of *Nidovirales* order. The coronavirus family consists of two subfamilies, *Letovirinae* and *Orthocoronavirinae*. Among these two subfamilies, the *Orthocoronavirinae* contains four genera, *Alphacoronavirus, Betacoronavirus, Gammacoronavirus*, and *Deltacoronavirus*. And among these four genera, the *Alphacoronavirus* and *Betacoronavirus* are known to cause diseases in humans [3]. Coronaviruses are enveloped viruses containing a positive-sense single stranded RNA genome. The genome of coronaviruses ranges from approximately 25 to 34 kb. The viral envelope comprises a lipid bilayer where the membrane (M) and spike (S) structural proteins are anchored [4,5].

Till now, seven strains of CoV have been identified which can cause diseases in humans. Among them, HCoV-OC43, HCoV-229E, HCoV-NL63, and HCoV-HKU1 strains are responsible for relatively milder respiratory illnesses. Among these four strains, the HCoV-OC43 strain is the most prevalent and according to a study, during the times of viral peak activity, this strain can be responsible for up to 35% of upper respiratory tract infection cases [6]. However, the other three strains, severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV) and the recently emerged strain, severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) are highly pathogenic and lethal than any other types of coronaviruses [1,3]. The HCoVs are transmitted through the respiratory droplets produced from sneezes or cough of an infected individual. The SARS-CoV outbreak occurred between November 2002 and July 2003, which caused the infection of total 8096 individuals in 27 countries including 774 deaths [7]. However, random cases of SARS-CoV infections are still being reported to this date. The MERS-CoV was first identified in 2012 and as of November 2019, it has caused 2494 infected cases and 858 deaths [1]. The recent coronavirus disease 2019 (COVID-19) outbreak by SARS-CoV-2 is by far the most lethal and dangerous among all the previous outbreaks of coronaviruses and the SARS-CoV-2 outbreak has been announced as a 'pandemic'. According to the reports of Centers for Disease Control and Prevention (CDC) (https://www.worldometers.info/coronavirus/) [8], the SARS-CoV-2 viral transmission

CONTACT Umme Salma Zohora zohoraus@bgeju.edu.bd Department of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Jahangirnagar University, Dhaka, Bangladesh.

Supplemental data for this article can be accessed here.

© 2021 Informa UK Limited, trading as Taylor & Francis Group

started in Wuhan, China in December 2019, and as of 20 January 2021, about 9.6 million people from all over the world got infected with the mortality rate of about 2.5% [9].

Currently, neither any satisfactory antiviral therapy nor any vaccine is available to fight the infections of HCoVs including SARS-CoV, MERS-CoV, and SARS-CoV-2. Different types of antiviral therapies have been tried and experimented over the years but none of these treatments has gained success to the expected threshold. Antiviral preparations like Ribavirin, Lopinavir, and Ritonavir, in combination with various interferons, had been used to alleviate the MERS and SARS viral infections in recent years [7]. However, the results of these studies are inconsistent and due to lack of any definitive proof from the trials, these treatments could not be declared 100% effective to fight the coronaviruses [7,10]. Although much emphasis has already been put forward to develop potential treatments for COVID-19, still no approved preventing option is available. Several antiviral therapies are being tried by scientists, but these therapies offer varying therapeutic results from patient to patient of COVID-19 [11]. Much progress was reported in vaccine development against the MERS-CoV. However, due to numerous safety and efficacy concerns in the vaccinated individuals, these vaccines have failed to provide any definitive immunity to the viral infections [7,10].

In our study, the methods of immunoinformatics and reverse vaccinology have been utilized to design potential epitope-based polyvalent vaccines targeting the spike glycoproteins of multiple strains of HCoV i.e., HCoV-OC43, SARS-CoV, MERS-CoV, and SARS-CoV-2. Reverse vaccinology and immunoinformatics are quick, easy, and cost-effective approaches of vaccine designing and development where the novel antigens of a pathogen are identified by analyzing its genetic makeup with the aid of different tools of in silico biology [,]. A flowchart of the methods employed in the vaccine designing in this experiment has been illustrated in Figure 1. The spike glycoprotein (SG) sequences (also known as S protein) of CoVs are among the most important proteins that mediate the virus entry and pathogenesis, determine the host range of the virus and elicit the effects of the neutralizing antibodies [15–17]. The SG proteins of HCoVs have two subunits i.e., S1 and S2 subunits. The S1 subunit functions in binding with the host cell receptor and the S2 subunit functions in the fusion of viral and cellular membranes. The S1 subunit has two domains i.e., N-terminal domain (S1-NTD) and the C-terminal domain (S1-CTD). During infection, the S1-CTD subunit recognizes the protein receptors of the host cells like the ACE2, APN, and DPP4 for binding. Then the S2 subunit causes the fusion of the viral and host cell membrane, thus the viral genome is entered into the host cell. Therefore, SG proteins on the surface of HCoVs play very important roles in mediating their infection [18,]. The SG proteins also give the HCoVs their characteristic crown-like appearance because these proteins stud and protrude from the viral envelope. For these reasons, targeting the SGs for designing vaccines against HCoVs presents an effective strategy [20].

2. Materials and Methods

2.1. Strain Identification and Protein Sequence Retrieval

The strains of HCoVs and the target proteins were identified and retrieved from the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/) as well as the UniProt database (https://www.uniprot.org/).

2.2. Prediction of Antigenicity and Physicochemical Properties of the Proteins

The antigenicity of the target protein sequences was predicted by the online antigenicity prediction tool, VaxiJen v2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.htm), keeping the prediction accuracy parameter threshold at 0.4. The accuracy, sensitivity, and specificity of a prediction depend on the accuracy prediction threshold and the 0.4 threshold improves the prediction accuracy of the server [21,22]. Thereafter, the physicochemical properties i.e., the number of amino acids, theoretical pI, instability indexes, extinction coefficients, half-lives, grand average of hydropathicity (GRAVY) etc. of the target proteins were predicted by ExPASy's online tool ProtParam (https://web.expasy.org/protparam/) [].

2.3. T-cell and B-cell Epitope Prediction

The T-cell and B-cell epitopes of the protein sequences were predicted for the vaccine construction. The cytotoxic T-cells are important for specific antigen recognition and the helper T-cells are the most important component of adaptive immunity, which function in activating the B-cell, macrophages, and even cytotoxic T-cells [24,25]. As a result, both cytotoxic and helper T-cell epitopes are essential for a successful vaccine designing study.

The epitopes were predicted using the Immune Epitope Database or IEDB (https://www.iedb.org/), which contains a vast collection of experimental data on T-cell epitopes and antibodies [26]. The MHC class-I or CD8+ cytotoxic T-lymphocytic (CTL) epitopes were predicted using the IEDB recommended 2020.04 (NetMHCpan EL 4.0) method . The top MHC class-I epitopes were selected for further analysis. Again, MHC class-II or CD4+ helper T-lymphocytic (HTL) epitopes were predicted using the IEDB Recommended Method 2.2 which gives predictions in percentile scores. The top predicted MHC class-II epitopes were considered for further analysis. The determined MHC class-I epitopes were 9-mers and the MHC class-II epitopes were 15-mers. The HLA alleles for which these predicted epitopes were found, were also noted and listed. B-cell epitopes in a vaccine trigger the antigen-specific immunoglobulin production which are crucial components of adaptive immunity []. The B-cell epitopes of the proteins were predicted by BepiPred linear epitope prediction method [] and the epitopes with more than

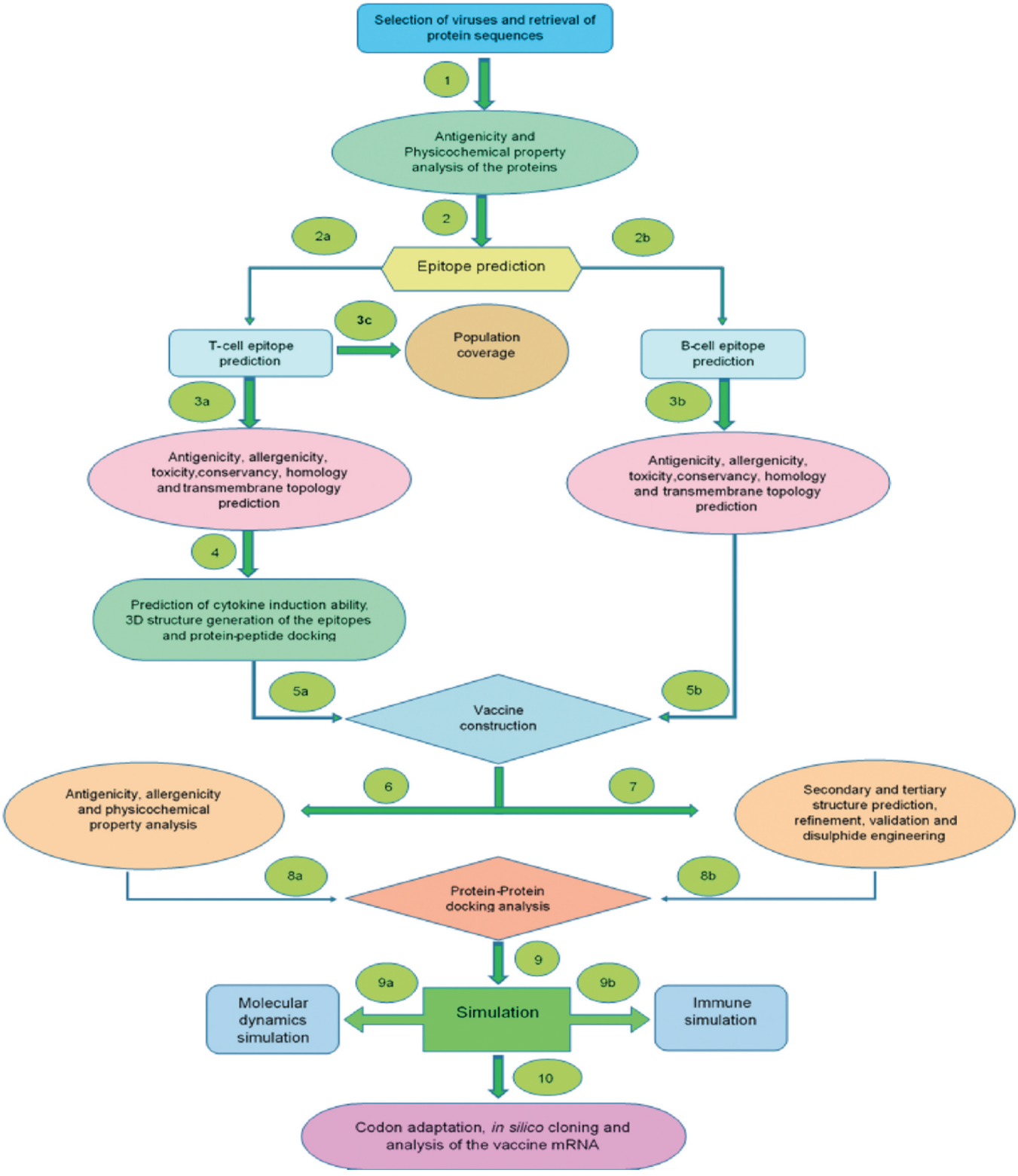


Figure 1. Stepwise procedures adapted in this study to design polyvalent vaccine.

ten amino acids were considered as potential candidates for vaccine construction. Using a combination of a hidden Markov model and a propensity scale method, BepiPred predicts the location of linear B-cell epitopes [].

2.4. Antigenicity, Allergenicity, Toxicity, Conservancy, and Human Homology Prediction of the Epitopes

In this step, the predicted T-cell and B-cell epitopes from the previous step were screened with several tools for predicting

their antigenicity, allergenicity, toxicity, conservancy, and human homology. Vaccine components should be highly antigenic and at the same time non-allergenic. Moreover, the vaccine components should also devoid of any toxic reaction. The antigenicity determining tool, VaxiJen v2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.htm) was again used in this step where the threshold was kept at 0.4 [21,22]. Two different tools i.e., AllerTOP v2.0 (https://www.ddgpharmfac.net/AllerTOP/), as well as AllergenFP v1.0 (http://ddg-pharmfac.net/AllergenFP/) were used for allergenicity prediction of the epitopes. The AllerTOP v2.0 server has better prediction accuracy of 88.7% than AllergenFP v1.0 server (87.9%) [29,30]. After that, ToxinPred (http://crdd.osdd.net/raghava/toxinpred/) server was used for toxicity prediction of all the epitopes. The default support vector machine (SVM) method was used for prediction keeping all the parameters default [31]. For conservancy analysis of the epitopes, multiple sequence alignment (MSA) was performed using the online tool Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) as well as the alignment module of UniProt database (https://www.uniprot.org/). In the MSA, SGs sequences from total 80 different isolates of HCoVOC43, SARS-CoV, MERS-CoV, and SARS-CoV-2 (20 isolates from each of the virus strain) were used (**Supplementary Table S1**). The homology of the epitopes to the human proteome was also determined in this step. The protein BLAST module (blastP) of BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) tool was used in the human homology determination, where *Homo sapiens* (taxid: 9606) was used for comparison keeping all other parameters default. An e-value cutoff of 0.05 was set and epitopes that showed no hits below e-value inclusion threshold were selected as non-homologous pathogen peptides [32]. The epitopes that were found to be highly antigenic, non-allergenic, nontoxic, 100% conserved among all the corresponding isolates and non-homologous to the human proteome, were considered as the 'best selected epitopes' and used for vaccine construction in the later stages.

2.5. IFN-gamma, IL-4, and IL-10 Induction Capacity and Transmembrane Topology Prediction

The helper T-cells are known to produce different types of cytokines like INF-gamma, IL-4 (interleukin-4), and IL-10 (interleukin-10). These cytokines later aid in the activation of different immune cells i.e., cytotoxic T-cells, macrophages, etc [33]. Therefore, the cytokine inducing capability of the HTL or MHC class-II epitopes were determined in this study. The interferon-gamma (IFN-gamma) induction capability of the predicted HTL epitopes was carried out using IFNepitope (http://crdd.osdd.net/raghava/ifnepitope/) server, using the Hybrid (Motif + SVM) prediction approach. The Hybrid prediction approach is a highly accurate approach for IFN-gamma inducing epitope prediction [34]. Moreover, IL-4 and IL-10 inducing properties of the HTL epitopes were predicted using IL4pred and IL10pred servers, respectively [,36]. Both the IL4pred and IL10pred predictions were conducted based on the SVM method where the default threshold values were set at 0.2 and −0.3, respectively. Thereafter, the transmembrane topology experiment of all the epitopes predicted in sub-section 2.3 was carried out using the TMHMM v2.0 server (http://www.cbs.dtu.dk/services/TMHMM/) [37].

2.6. Population Coverage Analysis

The distribution of specific HLA alleles among different ethnicities and population around the world is an essential criterion for designing a successful multi-epitope vaccine. The expression of different HLA alleles varies from one ethnicity to another around the world. The IEDB population coverage tool (http://tools.iedb.org/population/) was used to determine the population coverage of the best-selected epitopes across multiple HLA alleles among different nations around the world [38,39].

2.7. 3D Structure Generation and Molecular Docking of the Epitopes

Molecular docking study is one of the necessary steps in reverse vaccinology since it predicts the binding of epitopes with the MHC receptors or sometimes, antibodies [14]. The 3D structures of the best-selected epitopes (epitopes that followed the previously mentioned criteria of high antigenicity, non-allergenicity, nontoxicity, 100% conservancy, and non-homology to human proteome) were taken into consideration for peptide-protein docking. The 3D structure generation process was carried out by PEP-FOLD3 (http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/) [40–42].

Thereafter, the best-selected MHC class-I epitopes were docked with HLA allele HLA-A*11-01 (PDB ID: 5WJL). And the MHC class-II alleles were docked with HLA DRB1*04-01 allele (PDB ID: 5JLZ). The successful docking of the selected MHC epitopes with the HLA alleles ensured that the epitopes might interact with these MHC alleles during an immune response. PatchDock (https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php) online molecular docking tool was used for the peptide-protein docking. After the docking by the PatchDock tool, the results were refined and rescored by the FireDock server (http://bioinfo3d.cs.tau.ac.il/FireDock/php.php) which provides output in the context of global energies and the lowest global energy represents the best prediction [43–46]. The docked complexes from the docking experiment were then visualized and analyzed by Discovery Studio Visualizer [47].

2.8. Vaccine Construction

The best-selected epitopes were conjugated with each other to construct three possible vaccines. From these three vaccines, one best possible vaccine would be selected based on the molecular docking analysis. The CTL, HTL and BCL epitopes were conjugated by GGGGS, GPGPG, and KK linkers. Three vaccine constructs contained three different adjuvant sequences: L7/L12 ribosomal protein and HBHA protein (*M. tuberculosis*, accession number: AGV15514.1) and beta-defensin-3 (UniProt accession number: Q5U7J2), that were linked to the epitopes by EAAAK linkers. Adjuvants are known to enhance the antigenicity, immunogenicity, stability, and longevity of the constructed vaccines [48,49]. The vaccine constructs also contained the pan HLA-DR epitope (PADRE) sequence, attached with the adjuvant and epitopes. The EAAAK linkers provideeffective separation of domains of the

bifunctional fusion proteins [50]. Again, the GPGPG linkers have the capability to prevent the generation of junctional epitopes and facilitate the immune processing and presentation [51]. Furthermore, the bi-lysine (KK) linkers preserve the independent immunological activities of the epitopes of a vaccine [52]. And the GGGGS is a flexible linker which was proved to be effective in conferring resistance to proteases [53].

Significant elevation of toll-like receptor-2 (TLR-2) and TLR-4 was reported in different studies after the HCoV infections in human body [,54]. L7/L12 ribosomal protein and HABA protein can stimulate the activity of TLR-4 and beta-defensin has the capability to activate TLR-1, 2, 4, etc. Moreover, studies have also shown that the PADRE sequence improves the immune response enhancing ability of the CTL epitopes. It also increases the potency of vaccines with minimal toxicity. For this reason, the adjuvants and PADRE sequence might confer potential immunity against the HCoV infections [55–62]. Figure 2 represents a schematic diagram of the three vaccines in their appropriate orientation that were constructed using different adjuvants and linkers. The three vaccines differ from each other only in their adjuvant sequences.

2.9. Antigenicity, Allergenicity and Physicochemical Property Analysis

The antigenicity of the three vaccine constructs was predicted by the VaxiJen v2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.htm) keeping the threshold at 0.4 [21,22]. The predicted results of the Vaxijen v2.0 server were further cross-checked by the ANTIGENpro module of the SCRATCH protein predictor (http://scratch.proteomics.ics.uci.edu/) [63]. The constructed vaccines should be highly antigenic to stimulate a better immune response. The allergenicity of the vaccine

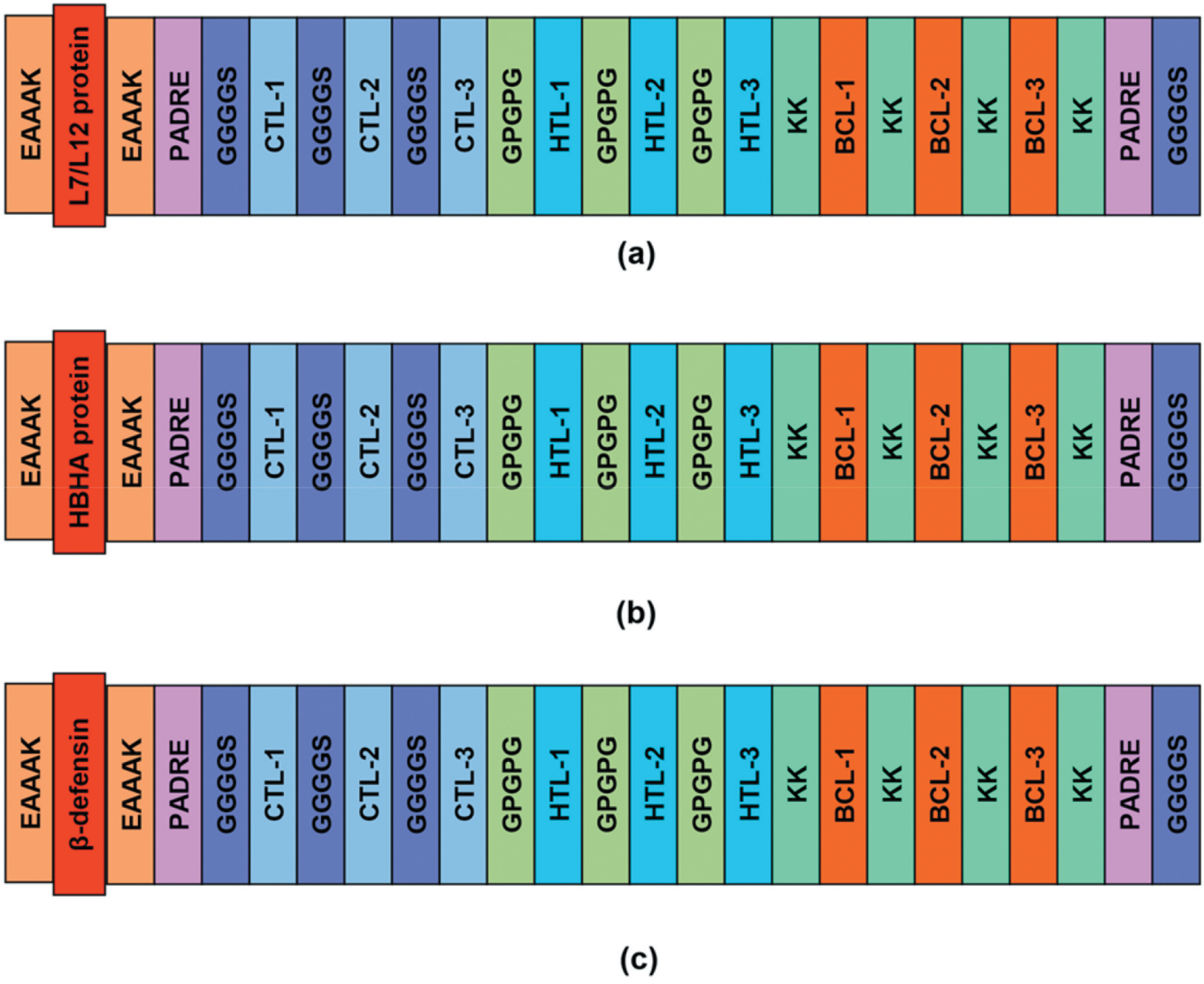


Figure 2. A schematic representation of the three possible vaccine constructs with linkers (EAAAK, GGGS, GPGPG, KK), PADRE sequence, adjuvants (L7/L12 protein, HBHA protein and β-defensin-3) and epitopes (CTL, HTL, BCL) in sequential and appropriate manner. (a) is the first vaccine constructed using the L7/L12 protein adjuvant, (b) is the second vaccine constructed using HABA adjuvant protein and (c) is the third vaccine constructed using β-defensin-3 protein as an adjuvant. CTL; cytotoxic T lymphocytic epitope, HTL; helper T lymphocytic epitope, BCL; B cell lymphocytic epitope. The three vaccine constructs differ from each other only in their adjuvant sequences.

constructs was predicted by three different online tools i.e., AlgPred (http://crdd.osdd.net/raghava/algpred/), AllerTop v2.0 (https://www.ddgpharmfac.net/AllerTOP/), and AllergenFP v1.0 (http://ddg-pharmfac.net/AllergenFP/) for improving the prediction accuracy. The AlgPred (http://crdd.osdd.net/raghava/algpred/) server attempts to integrate different approaches of allergenicity determination to predict potential allergenic proteins with high accuracy [64]. Thereafter, the physicochemical properties of the constructed vaccines were predicted again by the online tool ProtParam (https://web.expasy.org/protparam/) []. Along with the physicochemical property analysis, the solubility of vaccine constructs upon over-expression in *E. coli* host was predicted by the SOLpro module of the SCRATCH protein predictor (http://scratch.proteomics.ics.uci.edu/) and later further clarified by the Protein-sol server (https://protein-sol.manchester.ac.uk/) [,66].

2.10. Secondary and Tertiary Structure Prediction of the Vaccine Constructs

The secondary structures of the vaccine constructs were predicted by several online tools for improving prediction accuracy i.e., PRISPRED (http://bioinf.cs.ucl.ac.uk/psipred/) (using PRISPRED 4.0 prediction method), GOR IV (https://npsa-prabi.ibcp.fr/cgibin/npsa_automat.pl?page=/NPSA/npsa_gor4.html), SOPMA (https://npsa-prabi.ibcp.fr/cgibin/npsa_automat.pl?page=/NPSA/npsa_sopma.html), and SIMPA96 (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_simpa96.html). These tools are easy, accurate and efficient online servers to predict the amounts or percentage of amino acids in αhelix, β-sheet and coil structure formations [67–72]. All the parameters were kept default while predicting the secondary structures of the vaccine constructs. On the other hand, RaptorX (http://raptorx.uchicago.edu/) tool was used for predicting the tertiary or 3D structures of the three vaccine constructs. The server predicts the tertiary structure of a query protein or peptide by template-based method [73–75].

2.11. Tertiary Structure Refinement and Validation

When protein 3D structures are predicted by computational methods, they may lack their true, native structures. Therefore, the 3D structure refinement was conducted to convert the low resolution predicted model to models that closely resemble the native protein structure. The generated 3D structures of the vaccines were refined by GalaxyRefine module of the GalaxyWEB server (http://galaxy.seoklab.org/). The tool uses CASP10 tested refinement method and dynamics simulation to provide better-refined structures [76–78].

After structure refinement, the vaccine constructs were validated by analyzing the Ramachandran plots, generated by PROCHECK (https://servicesn.mbi.ucla.edu/PROCHECK/) server [79,80]. Moreover, another online tool, ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php) was also used for protein validation. It is a tool for checking the 3D models of protein structures for potential errors. The server generates z-score which is used to express the quality of a query protein structure. A z-score within the range of the z-scores of all the experimentally determined protein chains in the current PDB database represents better quality of a query protein [81].

2.12. Vaccine Protein Disulfide Engineering

Since disulfide bonds are necessary for the conformational stability of folded proteins, the disulfide engineering of the vaccine constructs was carried out in this experiment. The disulfide bond prediction was conducted by the Disulfide byDesign 2 v12.2 (http://cptweb.cpt.wayne.edu/DbD2/) server [82]. During the experiment, the χ3 angle was kept −87° or +97° ±5 to discard numerous putative disulfides which were generated using the default angles of +97° ±30° and −87° ±30°. The mentioned χ3 angle was used for identifying the fewer amount of disulfide bonds with desired characteristics. Again, the Cα-Cβ-Sγ angle was set at its default value (114.6° ±10) because studies have estimated that the Cα-Cβ-Sγ angle reaches a peak near 115° and covers a range from 105° to 125° in known disulfide bonds. Finally, the residue pairs with energy less than 2.2 Kcal/mol were selected and mutated to cysteine residue to allow disulfide bridge formation [83]. The energy value 2.2 Kcal/mol was selected as threshold because 90% of native disulfide bonds are generally found to have energy value less than 2.2 Kcal/mol [82].

2.13. Protein-Protein Docking

In protein-protein docking, the constructed vaccines were docked against multiple MHC alleles and TLRs and one best vaccine was selected based on their docking scores. The vaccines should have a good binding affinity with the MHC protein segments encoded by different MHC alleles. This is important because the MHC molecules produce potential immune responses after recognizing the vaccines which mimic the original viral infections. Thus vaccines help to produce immunity toward a particular virus or viruses [84]. In this study, the vaccine constructs were docked against the MHC alleles i.e., HLA-A*01:01 (PDB ID: 4NQX), HLA-A*02:01 (PDB ID: 4U6X), HLA-A*03:01 (PDB ID: 2XPG), and HLA-A*11:01 (PDB ID: 5WJL). Moreover, TLR-8 is responsible for generating immune responses against RNA viruses and TLR-3 aids in immune response generation against the DNA viruses [85,86]. Since HCoVs are positive-sense RNA viruses, the vaccines were also docked against the TLR-8 (PDB ID: 3W3M) []. To further assess the efficacy of the vaccines, they were docked with two more TLRs i.e., TLR-2 (PDB ID: 6NIG), and TLR-4 (PDB ID: 4G8A) because TLRs are some of the main players of the immune system.

In order to improve the prediction accuracy, the protein-protein docking was carried out by three different online tools. At first, the docking was carried out by ClusPro 2.0 (https://cluspro.bu.edu/login.php) where the lower energy score corresponds to the better binding affinity [88–90]. The ClusPro server calculates the energy score based on the following equation:

$$E = 0.40E_{rep} + (-0.40E_{att}) + 600E_{elec} + 1.00E_{DARS}(89,\ 90)$$

Thereafter,thedockingwasagaincarriedoutbythePatchDock (https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php) server and the results were then refined by FireDock server (http://bioinfo3d.cs.tau.ac.il/FireDock/php.php) [43,44,46]. Finally, another round of docking was performed by the HawkDock server (http://cadd.zju.edu.cn/hawkdock/) along with the Molecular Mechanics/Generalized Born Surface Area (MM-GBSA) study. In these servers, the lower output scores correspond to better binding affinity and vice versa [91–94]. Visualization of the best vaccine-TLR-8 complex was carried out by Discovery Studio Visualizer [47]. The vaccine showing the best performance in the docking study was considered as the best-selected/predicted vaccine construct.

2.14. Screening for Conformational B-lymphocytic Epitopes

Antibody-mediated humoral immunity initiates within the body when the B-cells interact with their epitopic counterparts. Therefore, the vaccines should have effective conformational B-cell epitopes to provide better immunity. The conformational B-cell epitopes of the best-predicted vaccine protein from the 2.13 sub-section were determined by IEDB ElliPro tool (http://tools.iedb.org/ellipro/), using the default parameters of a minimum score of 0.5 and a maximum distance of 6 angstrom [95].

2.15. Molecular Dynamics Simulation Analysis

The molecular dynamics simulation study was conducted by the online simulation tool, iMODS (http://imods.chaconlab.org/). It is a fast, easy and user-friendly tool for dynamics simulation study. The dynamics simulation was conducted only for the complex containing the best-selected vaccine (selected based on the docking study) and TLR-8. Different dynamic parameters, for example, deformability, B-factor (mobility profiles), eigenvalues, variance, co-variance map, and elastic network of the protein complex are predicted quite efficiently by this server. The deformability of a protein can be defined as the ability of the amino acids to deform at a specific position within the protein backbone. Moreover, the deformability is also depended on the eigenvalue of a protein. Lower eigenvalue represents easy deformability of the protein complex. The dynamics simulation study was conducted to predict the stability of the vaccine construct [96–100].

2.16. Immune Simulation

The immune simulation study was conducted for the best-predicted vaccine to predict its immunogenicity and immune response profile. The C-ImmSim server (http://150.146.2.1/CIMMSIM/index.php) was used for the immune simulation study which predicts the real-life-like immune interactions using position-specific scoring matrix (PSSM) and machine learning techniques [101]. During the experiment, all the parameters were kept default, except the time steps which were set at 1, 84, and 170 (time step 1 is injection at time = 0 and each time step is 8 hours) . So, three injections are expected to be required at four weeks interval because the recommended interval between two doses of most of the commercial vaccines is four weeks [102]. The Simpson's Diversity index, D was calculated from the figures.

2.17. Codon Adaptation, In Silico Cloning, Prediction of the mRNA Secondary Structure, and Expression of the best Vaccine Protein with SUMO-fusion

Codon adaptation and *in silico* cloning are two important steps in reverse vaccinology. An amino acid can be encoded by more than one codon in different organisms, a phenomenon known as codon bias. The cellular machinery of an organism may be completely different from another organism, so the same amino acid can be encoded by completely different codons in different organisms. For this reason, codon adaptation was conducted in this study to predict the suitable codon that might encode a specific amino acid efficiently in a particular organism. In codon adaptation study, the best-predicted vaccine was reverse translated to the possible DNA sequence that is expected to encode the vaccine protein [103,104]. The Java Codon Adaptation Tool or JCat server (http://www.jcat.de/) was used for the codon adaptation study which ensures maximal expression of a protein in a target organism [105]. Prokaryotic *E. coli* strain K12 was selected as the target organism and rho-independent transcription terminators, prokaryotic ribosome binding sites and EaeI and StyI cleavage sites of restriction enzymes, were avoided at the server. The vaccine protein sequence was reverse translated to the optimized DNA sequence by the server. Then the optimized DNA sequence was collected and EaeI and StyI restriction sites were attached to the N-terminal and C-terminal sites, respectively. Finally, the SnapGene restriction cloning software was used for inserting the newly adapted DNA sequence between the EaeI and StyI restriction sites of the pETite vector (Lucigen, USA) [106]. The pETite vector plasmid contains a small ubiquitin-like modifier or SUMO-tag and 6X-His tag which facilitates the solubilization and effective affinity purification of the recombinant protein [107].

The mRNA secondary structure prediction was conducted by two online servers i.e., Mfold (http://unafold.rna.albany.edu/?q=mfold) and RNAfold (http://rna.tbi.univie.ac.at/cgibin/RNAWebSuite/RNAfold.cgi). Both of these servers predict the mRNA secondary structures thermodynamically and provide minimum free energy (ΔG Kcal/mol) for each of the generated structures. The lower the minimum free energy, the more stable the folded mRNA is and vice versa [108–111]. To predict the mRNA folding and secondary structure of the best-selected vaccine, at first the optimized DNA sequence from JCat server was taken and converted to possible RNA sequence by DNA<->RNA>Protein tool (http://biomodel.uah.es/en/lab/cybertory/analysis/trans.htm). Then the RNA sequence was collected from the tool and pasted in the Mfold and RNAfold servers for prediction using the default settings for all the parameters.

3. Results

3.1. Strain Identification and Protein Sequence Retrieval

The SG sequences of HCoV-OC43, SARS-CoV, MERS-CoV, and SARS-CoV-2 were identified as potential targets for the current study by reviewing the literatures from the NCBI database. The

Table 1. List of the spike glycoproteins (SGs) of different HCoVs with their accession numbers used in the vaccine designing study.

No	Name of the virus	Accession number of the protein	Protein sequence length (aa)
01	Human coronavirus OC43 (HCoV-OC43)	UniProt accession no: P36334	1353
02	Human SARS coronavirus (SARS-CoV)	UniProt accession no: P59594	1255
03	MERS coronavirus (MERS-CoV)	UniProt accession no: K9N5Q8	1353
04	SARS-CoV-2	GenBank accession no: QHR63290.2	1273

protein sequences of HCoV-OC43, SARS-CoV, and MERS-CoV were retrieved from UniProt and the protein sequence of SARS-CoV-2 was retrieved from NCBI database (Table 1).

3.2. Prediction of Antigenicity and Physicochemical Properties of the Proteins

All the SG sequences from the four viral strains were found to be antigenic. All of them had a predicted half-life of 30 h in mammalian reticulocytes and all of them were predicted to be stable. The SG from HCoV-OC43 had the highest predicted aliphatic index of 85.94 as well as the highest predicted extinction co-efficient of 196,595 M^{-1} cm^{-1} and the SG of SARS-CoV-2 was found to have the lowest GRAVY value of −0.079. Except the SARS-CoV-2, all the other proteins from the selected viruses had almost similar theoretical pI (**Supplementary Table S2**).

3.3. T-cell and B-cell Epitope Prediction, IFN-gamma, IL-4, IL-10 Induction Capacity and Transmembrane Topology Analysis

The T-cell and B-cell epitopes of the SG sequences of the four viral strains were predicted for vaccine construction. These epitopes were expected to stimulate potential T-cell and B-cell immune responses. Based on their rankings, the top MHC class-I and MHC class-II epitopes as well as B-cell epitopes with length over ten amino acids, were taken into consideration for all the viral strains. From these epitopes, the highly antigenic, non-allergenic, nontoxic, 100% conserved and non-homologous epitopes were considered as the best selected epitopes and selected for final vaccine construction. IFN-gamma, IL-4, and IL-10 inducing capacity prediction of the HTL epitopes had showed that most of the selected HTL epitopes had at least one of these cytokine producing capabilities. Moreover, about 30% of all the epitopes were found to reside outside of the cell membrane according to the transmembrane topology test. **Supplementary Table S3, Supplementary Table S4, Supplementary Table S5** and **Supplementary Table S6** list the epitopes predicted for the four viral strains.

3.4. Antigenicity, Allergenicity, Toxicity, Conservancy and Human Homology Prediction of the Epitopes

The antigenicity and allergenicity, toxicity, conservancy, and human homology predictions were carried out to find out the best epitopes that followed the previously mentioned selection criteria which could be used effectively in vaccine construction. Most of the epitopes were found to be 100% conserved among the SG sequences of the selected isolates and almost all of them were also found to be non-homologous to the human proteome. The best-selected epitopes of the selected virus strains are listed in Table 2.

3.5. Population Coverage Analysis of the Epitopes and their MHC Alleles

The population coverage analysis showed that the best-selected MHC class-I and MHC class-II epitopes covered about 86.11% and 84.94% of the world population, respectively and the epitopes covered 91.23% of the world population in combination. The highest prevalence of the MHC class-I epitopes was found in China (90.49%) and the lowest prevalence of the epitopes were found in Central Africa (54.29%). India was predicted to have 91.40% of the population covered by the MHC class-II epitopes and only 43.67% of the West African population was predicted to have the best-selected MHC class-II coverage. The highest combining coverage of MHC class-I and ClassII epitopes was predicted to be in Saudi Arabia (95.43%) and lowest in West Africa (51.90%). The results of the population coverage analysis among some HCoV infected countries are visualized in **Supplementary Figure S1.**

3.6. 3D Structure Generation and Molecular Docking of the T-cell Epitopes

All the best-selected T-cell epitopes were subjected to the PEPFOLD3 server for 3D structure generation. Thereafter, the 3D structures were used in the peptide-protein docking experiment to determine, whether all of the best-selected epitopes had the ability

Table 2. List of the best-selected epitopes from the spike glycoproteins (SGs) of all the four viral strains that followed the selection criteria and used in vaccine construction. Selection criteria: high antigenicity, non-allergenicity, nontoxic, 100% conservancy and non-homolohous to human proteome.

MHC class	HCoV-OC43	SARS-CoV	MERS-CoV	SARS-CoV-2
MHC class-I epitopes	VVYAQHCFK	GVIADYNYK	AALSAQLAK	GVYFASTEK
	STCAVNYTK	EVMPVSMAK	ASIGDIIQR	GTHWFVTQR
	-	TLADAGFMK	YSVSSFEAK	TLADAGFIK
	-	TTTSTALGK	SVIYDKETK	ASANLAATK
	-	SVYAWERKK	-	-
MHC class-II epitopes	YRIDTTATSCQLYYN	NTLVKQLSSNFGAIS	SVRNLFASVKSSQSS	DLFLPFFSNVTWFHA
	NYRIDTTATSCQLYY	REGVFVFNGTSWFIT	IYPAFMLGSSVGNFS	LQYGSFCTQLNRALT
	FNYRIDTTATSCQLY	IFLLFLTLTSGSDLD	-	QDLFLPFFSNVTWFH
B-cell epitopes	RTINSTQDGDNK	DDVRQIAPGQTGVI	KTWPRPIDVSKA	LTPGDSSSGWTAG
	CVGSGPGKNNGIGTCPAGT	YDPLQPELDSF	TVWEDGDYYRK	VRQIAPGQTGKIAD

to bind with the MHC class-I and MHC class-II molecule or not. The selected epitopes were docked against the HLA- A*11-01 allele and HLA-DRB1*04-01. Among the MHC class-I epitopes of SG proteins from HCoVOC43, SARS-CoV, MERS-CoV, and SARS-CoV-2, VVYAQHCFK, GVIADYNYK, YSVSSFEAK, and GTHWFVTQR generated the best results (global energies: −47.40, −44.54, 21.70, and −33.41, respectively). On the other hand, among MHC class-II epitopes of HCoV-OC 43, SARS-CoV, MERS-CoV, and SARS-CoV-2, FNYRIDTTATSCQL, NTLVKQLSSNFGAIS, IYPAFMLGSSVGNFS, and LQYGSFCTQLNRALT generated the best global energies of −60.26, −37.22, −22.41, and −10.11, respectively (**Supplementary Table S7**). **Supplementary Figure S2 & S3** illustrate the docked epitopes within the binding pocket of their respective targets. Since all the best-selected epitopes were docked successfully with their receptors, it can be concluded that all of the receptors have good affinities toward the MHC class-I and MHC class-II epitopes.

3.7. Vaccine Construction

After successful docking, the best-selected epitopes were joined with one another by different types of linkers to construct the vaccine. GGGGS linkers were used to conjugate the CTL epitopes, GPGPG linkers were used to join HTL epitopes and KK linkers were used to attach the BCL epitopes. The PADRE sequence and the adjuvants were linked by EAAAK linkers. The three vaccines were designated as: V1, V2, and V3 (Table 3). The three vaccines differ from each other only in their adjuvant sequences.

3.8. Antigenicity, Allergenicity and Physicochemical Property Analysis

All the three vaccine constructs were found to be potentially antigenic and non-allergenic. Moreover, all of them had quite similar theoretical pI of more than 9.00, so all of them were predicted to be basic in nature. However, V2 had the highest predicted extinction co-efficient of 91,260 M^{-1} cm^{-1} and it also had the lowest GRAVY value of −0.403. Furthermore, all the three vaccine constructs were found to be stable and had similar predicted half-lives and all of them were found to be soluble upon over-expression in *E. coli*. The results of the physicochemical property analysis are listed in Table 4.

3.9. Secondary and Tertiary Structure Prediction of the Vaccine Constructs

The secondary structure prediction of the three vaccine constructs were carried out by several online tools for improving the accuracy of the prediction. The results of the secondary structure prediction are listed in **Supplementary Table S8**. All the proteins were found to have quite similar predictions by all the servers. The secondary structure prediction showed that the adjuvants had caused potential differences among the secondary structures of the vaccine proteins. **Supplementary Figure S4** illustrates the results of secondary structure prediction carried out by the PRISPRED server. The tertiary structures of the three vaccine constructs were conducted by the RaptorX server. **Supplementary Figure S5** illustrates the 3D structures generated by the RaptorX server.

3.10. Tertiary Structure Refinement and Validation

The tertiary structures of the vaccine constructs were refined by GalaxyRefine module of the GalaxyWEB server and later validated by analyzing the Ramachandran plots and z-scores. The protein structure validation of the three vaccine constructs showed that all the three vaccines were much more improved than their crude, unrefined structures. The V1, V2, and V3 had z-scores of 6.87, −5.67, and −5.73, respectively, which reflected that all of them had scores well within the range of experimentally proven X-ray crystal structures [81]. Again, all of them had almost similar percentages of amino acids in the Rama favored region (**Supplementary Table S9**).

Table 3. Protein sequences of the three vaccine constructs. The bold letters represent the linker sequences.

Name of the vaccines	Vaccine constructs
V1	**EAAAK**MAKLSTDELLDAFKEMTLLELSDFVKKFEETFEVTAAAPVAVAAAGAAPAGAAVEAAEEQSEFDVILEAAGDKKIGVIKVVREIVSGLGLKEAKDLVDGAPKPL LEKVAKEAADEAKAKLEAAGATVTVK**EAAAK**AKFVAAWTLKAAA**GGGGS**VVYAQHCFK**GGGGS**STCAVNYTK**GGGGS**GVIADYNYK**GGGGS**EVMPVSMAK **GGGGS**TLADAGFMK**GGGGS**TTTSTALGK**GGGGS**SVYAWERKK**GGGGS**AALSAQLAK**GGGGS**ASIGDIIQR**GGGGS**YSVSSFEAK**GGGGS**SVIYDKETK**GGG GS**GVYFASTEK**GGGGS**GTHWFVTQR**GGGGS**TLADAGFIK**GGGGS**ASANLAATK**GPGPG**YRIDTTATSCQLYYN**GPGPG**NYRIDTTATSCQLYY**GPGPG**FNY RIDTTATSCQLY**GPGPG**NTLVKQLSSNFGAIS**GPGPG**REGVFVFNGTSWFIT**GPGPG**IFLLFLTLTSGSDL**GPGPG**SVRNLFASVKSSQSS**GPGPG**IYPAFMLGSS VGNFS**GPGPG**DLFLPFFSNVTWFHA**GPGPG**LQYGSFCTQLNRALT**GPGPG**QDLFLPFFSNVTWFH**KK**RTINSTQDGDNK**KK**CVGSGPGKNNGIGTCPAGT**KK** DDVRQIAPGQTGVI**KK**YDPLQPELDSF**KK**KTWPRPIDVSKA**KK**TVWEDGDYYRK**KK**LTPGDSSSGWTAG**KK**VRQIAPGQTGKIAD**KK**AKFVAAWTLKAAA**GGGGS**
V2	**EAAAK**MAENPNIDDLPAPLLAALGAADLALATVNDLIANLRERAEETRAETRTRVEERRARLTKFQEDLPEQFIELRDKFTTEELRKAAEGYLEAATNRYNELVERGEA ALQRLRSQTAFEDASARAEGYVDQAVELTQEALGTVASQTRAVGERAAKLVGIEL**EAAAK**AKFVAAWTLKAAA**GGGGS**VVYAQHCFK**GGGGS**STCAVNYTK**G GGGS**GVIADYNYK**GGGGS**EVMPVSMAK**GGGGS**TLADAGFMK**GGGGS**TTTSTALGK**GGGGS**SVYAWERKK**GGGGS**AALSAQLAK**GGGGS**ASIGDIIQR**GGG GS**YSVSSFEAK**GGGGS**SVIYDKETK**GGGGS**GVYFASTEK**GGGGS**GTHWFVTQR**GGGGS**TLADAGFIK**GGGGS**ASANLAATK**GPGPG**YRIDTTATSCQLYYN**GP GPG**NYRIDTTATSCQLYY**GPGPG**FNYRIDTTATSCQLY**GPGPG**NTLVKQLSSNFGAIS**GPGPG**REGVFVFNGTSWFIT**GPGPG**IFLLFLTLTSGSDL**GPGPG**SVRNL FASVKSSQSS**GPGPG**IYPAFMLGSSVGNFS**GPGPG**DLFLPFFSNVTWFHA**GPGPG**LQYGSFCTQLNRALT**GPGPG**QDLFLPFFSNVTWFH**KK**RTINSTQDGDNK **KK**CVGSGPGKNNGIGTCPAGT**KK**DDVRQIAPGQTGVI**KK**YDPLQPELDSF**KK**KTWPRPIDVSKA**KK**TVWEDGDYYRK**KK**LTPGDSSSGWTAG**KK**VRQIAPGQT GKIAD**KK**AKFVAAWTLKAAA**GGGGS**
V3	**EAAAK**GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKK**EAAAK**AKFVAAWTLKAAA**GGGGS**VVYAQHCFK**GGGGS**STCAVNYTK**GGGGS** GVIADYNYK**GGGGS**EVMPVSMAK**GGGGS**TLADAGFMK**GGGGS**TTTSTALGK**GGGGS**SVYAWERKK**GGGGS**AALSAQLAK**GGGGS**ASIGDIIQR**GGGGS** YSVSSFEAK**GGGGS**SVIYDKETK**GGGGS**GVYFASTEK**GGGGS**GTHWFVTQR**GGGGS**TLADAGFIK**GGGGS**ASANLAATK**GPGPG**YRIDTTATSCQLYYN**GP GPG**NYRIDTTATSCQLYY**GPGPG**FNYRIDTTATSCQLY**GPGPG**NTLVKQLSSNFGAIS**GPGPG**REGVFVFNGTSWFIT**GPGPG**IFLLFLTLTSGSDL**GPGPG** SVRNLFASVKSSQSS**GPGPG**IYPAFMLGSSVGNFS**GPGPG**DLFLPFFSNVTWFHA**GPGPG**LQYGSFCTQLNRALT**GPGPG**QDLFLPFFSNVTWFH**KK** RTINSTQDGDNK**KK**CVGSGPGKNNGIGTCPAGT**KK**DDVRQIAPGQTGVI**KK**YDPLQPELDSF**KK**KTWPRPIDVSKA**KK** TVWEDGDYYRK**KK**LTPGDSSSGWTAG**KK**VRQIAPGQTGKIAD**KK**AKFVAAWTLKAAA**GGGGS**

Table 4. The results of physicochemical property analysis of the three vaccine constructs with probabilities. AN; Antigenicity, AG; allergenicity, II; instability index, AI; aliphatic index; GRAVY; Grand average of hydropathicity, SI; solubility.

Name of the vaccine	AN (VaxiJen 2.0 server, threshold 0.4)	AN (AntigenPro server)	AG	Total number of positively charged amino acid	Total number of Negatively charged amino acid	pI	Extinction coefficients (in M^{-1} cm^{-1})	Estimated half-life	II	AI	GRAVY	SI
V1	Antigen	Antigen	Non-allergen	77	57	9.30	86790	1 hours (mammalian reticulocyte), >10 hours (*E.coli*)	Stable (24.59)	63.70	-0.238	Soluble (SolPro: 0.898, ProteinSol: 0.519)
V2	Antigen	Antigen	Non-allergen	82	63	9.31	91260	1 hours (mammalian reticulocyte), >10 hours (*E.coli*)	Stable (29.44)	62.56	-0.403	Soluble (SolPro: 0.907, ProteinSol: 0.545)
V3	Antigen	Antigen	Non-allergen	74	34	9.74	90145	1 hours (mammalian reticulocyte), >10 hours (*E.coli*)	Stable (26.90)	55.73	-0.358	Soluble (SolPro: 0.957, ProteinSol: 0.581)

Supplementary Figure S6 depicts the Ramachandran plot of each of the three vaccine constructs (generated by the PROCHECK server) and z-score or model quality graphs generated by the ProSA-web server.

3.11. Vaccine Protein Disulfide Engineering

The V1, V2, and V3 vaccine constructs generated 23, 29, and 21 possible pairs of amino acids, respectively, capable of undergoing disulfide bond formation. However, only 2 pairs from V1: 73 Glu-105 Ala and 161 Ala-196 Ser; 8 pairs from V2: 297 Thr-320 Thr, 410 Ala-436 Tyr, 454 Asn-463 Ser, 490 Pro-512 Ser, 489 Ala-511 Gln, 546 Ile-566 Ser, 312 Arg-472 Asp, and 691 Ser-694 Arg and 2 pairs from V3: 182 Thr-202 Arg, 389 Arg-477 Phe, were selected for disulfide bond formation by mutation because they had bond energy of less than 2.2 kcal/mol. **Supplementary Figure S7** gives the visual representation of the selected amino acid pairs in their original form and mutated form for disulfide bond formation. With 8 pairs of possible disulfide bonds, V2 was found to be most stable among the three vaccine constructs.

3.12. Molecular Docking Analysis

Much emphasis was given on molecular docking analysis of the three vaccine constructs with several MHC alleles and TLR-8 because one best vaccine construct was predicted based on the docking analysis. Therefore, to improve the prediction accuracy, the docking was carried out by three different online tools. From the docking analysis listed in Table 5, it can be declared that V3 vaccine construct was the best vaccine construct. It generated the lowest and so the best binding free energies as well as ClusPro global energy scores with all the target proteins in the MM-GBSA and ClusPro analyses, respectively. Again, V3 also generated the best scores with all the targets except the HLA-A*01:01 allele, when docked by HawkDock server. Therefore, this server also pointed toward V3 as the best vaccine construct.

Moreover, when analyzed by PatchDock and FireDock servers, V3 was predicted to have best scores with HLA-A*01:01, HLA-A*11:01, TLR-2, TLR-4, and TLR-8. And most importantly, V3 showed the satisfactory performances when docked against the TLR-2, TLR-4, and TLR-8 by all the servers. None of the V1 and V2 vaccine constructs generated such good results like V3. For this reason, V3 vaccine construct was considered to be the best vaccine construct and the later analyses were conducted only for the V3 construct. Figure 3 illustrates the interaction of the best vaccine V3 and the receptor protein TLR-8 and their interacting amino acids.

3.13. Screening for Conformational B-lymphocytic Epitopes

The conformational B-cell epitopes were predicted using ElliPro server which predicts conformational epitopes from tertiary structures. Total 357 residues were found with scores varying from 0.516 to 0.876. The length of the epitopes were ranged from 3 to 158 amino acids and predicted to be located

Table 5. Results of the molecular docking between the three vaccine constructs and their selected receptors.

Name of the vaccines	Targets	PDB IDs of the targets	ClusPro energy score	Global energy	HawkDock score (the lowest score)	MM-GBSA (binding free energy, in kcal mol^{-1})
V1	HLA-A*01:01	4NQX	−989.8	−16.04	−6348.92	−77.18
	HLA-A*02:01	4U6X	−939.4	−22.91	−5892.05	−48.30
	HLA-A*03:01	2XPG	−1112.1	−5.39	−6892.94	−108.69
	HLA-A*11:01	5JLZ	−1075.5	3.45	−6120.40	−101.54
	TLR-2	6NIG	−1031.1	−1.55	−6940.39	−77.38
	TLR-4	4G8A	−957.6	−0.29	−6853.22	−77.19
	TLR8	3W3M	−1134.6	−10.45	−7211.00	−86.04
V2	HLA-A*01:01	4NQX	−933.6	−14.20	−5509.57	−56.52
	HLA-A*02:01	4U6X	−870.4	−6.78	−6091.20	−73.14
	HLA-A*03:01	2XPG	−967.1	−8.22	−5492.71	−112.95
	HLA-A*11:01	5JLZ	−813.2	−6.30	−5436.69	−86.34
	TLR-2	6NIG	−941.4	−10.71	−6161.79	−91.15
	TLR-4	4G8A	−917.8	−11.13	−6198.50	−76.00
	TLR8	3W3M	−899.3	−7.52	−5981.45	−60.39
V3	HLA-A*01:01	4NQX	−1087.2	−28.73	−6274.76	−98.25
	HLA-A*02:01	4U6X	−972.5	−8.43	−6120.33	−80.67
	HLA-A*03:01	2XPG	−1260.0	−1.56	−7287.23	−121.11
	HLA-A*11:01	5JLZ	−1178.5	−12.36	−6602.48	−132.01
	TLR-2	6NIG	−1202.5	−13.10	−7295.12	−101.56
	TLR-4	4G8A	−1004.5	−13.68	−7546.51	−117.33
	TLR8	3W3M	−1345.2	−18.40	−7324.56	−93.76
Remarks	NA	NA	Best vaccine construct: V3 (considering the docking scores against all the targets)	Best vaccine construct: V3 (considering the docking scores against targets: HLA-A*01:01, HLA-A*11:01, TLR-2, TLR-4 and TLR-8)	Best vaccine construct: V3 (considering the docking scores against all the targets except HLA-A*01:01)	Best vaccine construct: V3 (considering the docking scores against all the targets)

within five conformational B-cell epitopes (**Supplementary Table S10** and **Supplementary Figure S8**).

3.14. Molecular Dynamics Simulation

Figure 4 illustrates the results of molecular dynamics simulation of V3-TLR-8 docked complex. The molecular dynamics simulation study is performed in revere vaccinology to determine the stability and physical movements of atoms and molecules of a vaccine construct [112]. Therefore, the simulation study was conducted in this study to determine the movements of molecules in the V3 vaccine construct. The deformability graph of the complex illustrates the peaks in the graphs which represent the regions of the protein with deformability (Figure 4b). The B-factor graph of the docked complex provides an easy understanding and visualization of the comparison between the NMA and PDB field of the complex (Figure 4c). The graph of eigenvalue of the complex is illustrated in Figure 4d which shows that V3 and TLR-8 docked complex generated an eigenvalue of 1.278806e-07. The variance graph represents the individual variance by red colored bars and cumulative variance by green colored bars (Figure 4e). Figure 4f illustrates the co-variance map of the complex where the correlated motion between a pair of residues is depicted by red color, uncorrelated motion is represented by white color, and anticorrelated motion is indicated by blue color. Moreover, the elastic map of the complex represents the connection between the atoms and darker gray regions indicate stiffer regions (Figure 4g) [98–100].

3.15. Immune Simulation

The immune simulation of the best-selected vaccine V3 was performed by the C-ImmSimm server which predicts the generation of adaptive immunity as well as epitopes and immune interactions [101]. The simulation study had revealed that after each of the three vaccine injections, the primary immune response against the antigenic fragments was predicted to increase

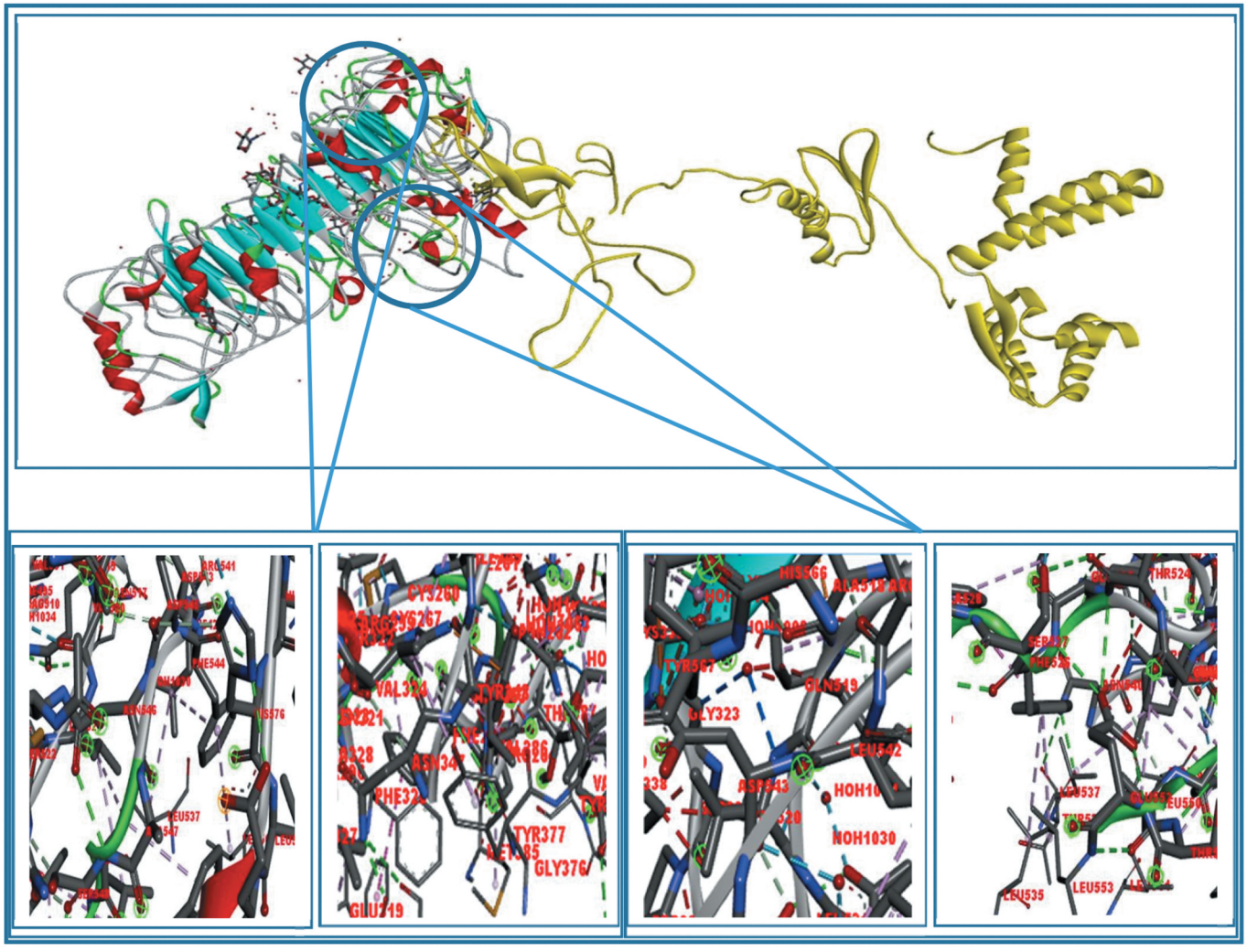


Figure 3. Interaction of the V3 vaccine construct (ligand in yellow color) with TLR-8 (receptor in variable color). The amino acids took part in the interaction: Leu 535 (receptor)-Phe 526(ligand), Leu 553 (receptor)-Phe 526 (ligand), Tyr 563 (receptor)-Asp 543 (ligand), Lys 333 (receptor)-Asp 543 (ligand), Arg 259 (receptor)-Tyr 322 (ligand), Tyr 377 (receptor)-Val 324(ligand), Ala 518 (receptor)-Lys 340 (ligand), Phe 501 (receptor)-Ile 496 (ligand), Leu 561 (receptor)-Phe 544 (ligand), Tyr 563 (receptor)- Pro 414(ligand).

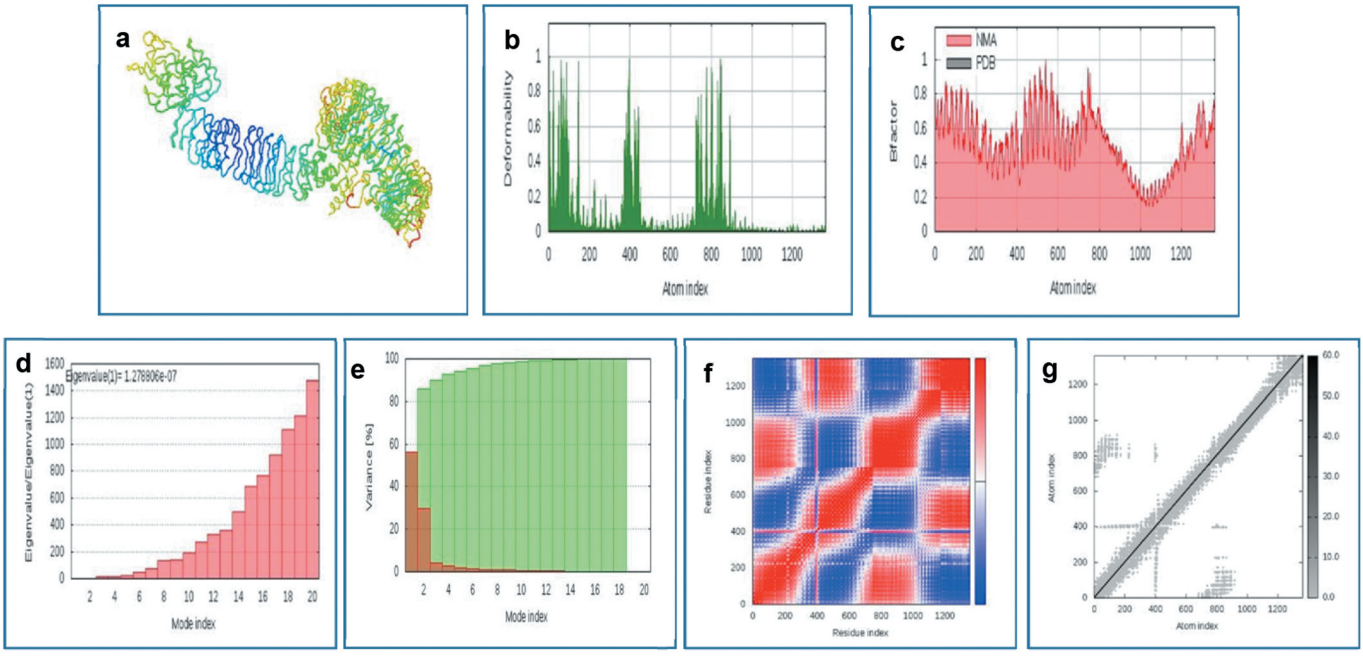


Figure 4. Figure depicting the results of molecular dynamics simulation study of V3 and TLR-8 docked complex. Here, (a) NMA mobility, (b) deformability, (c) B-factor, (d) eigenvalues, (e) variance (red color indicates individual variances and green color indicates cumulative variances), co-variance map (correlated (red), uncorrelated (white) or anti-correlated (blue) motions) and elastic network (darker gray regions represent more stiffer regions).

sharply as indicated by the gradual elevation in concentrations of different immunoglobulins (figure 5a). Again, in response to the primary immune stimulation, the secondary immune responses were also found to be increased. The gradual increase in the concentrations of active Bcell (figure 5b and figure 5c), plasma B-cell (figure 5d), helper T-cell (figure 5e and figure 5f), regulatory T-cell, and cytotoxic T-cell (figure 5g, figure 5h and figure 5i) was found, which indicated a strong secondary immune response, very good immune memory generation and the increased clearance of antigen after exposure. Moreover, the increase in the concentrations of dendritic cells and macrophages pointed towards very good antigen presentation by these antigen presenting cells (APCs) (figure 5j and figure 5k). The vaccine was also able to generate good amount of various types of cytokines like the IFN-gamma, IL-23, IL-10, and IFN-beta which are the most significant cytokines for generating immune response against viruses (Figure 5l). Overall, the immune simulation study had showed that with the predicted capability of generating high levels of immunoglobulins, active B-cells and T-cell, cytokines and APCs, the polyvalent vaccine V3 might be able to provide good immunogenic protection against the targeted four HCoVs.

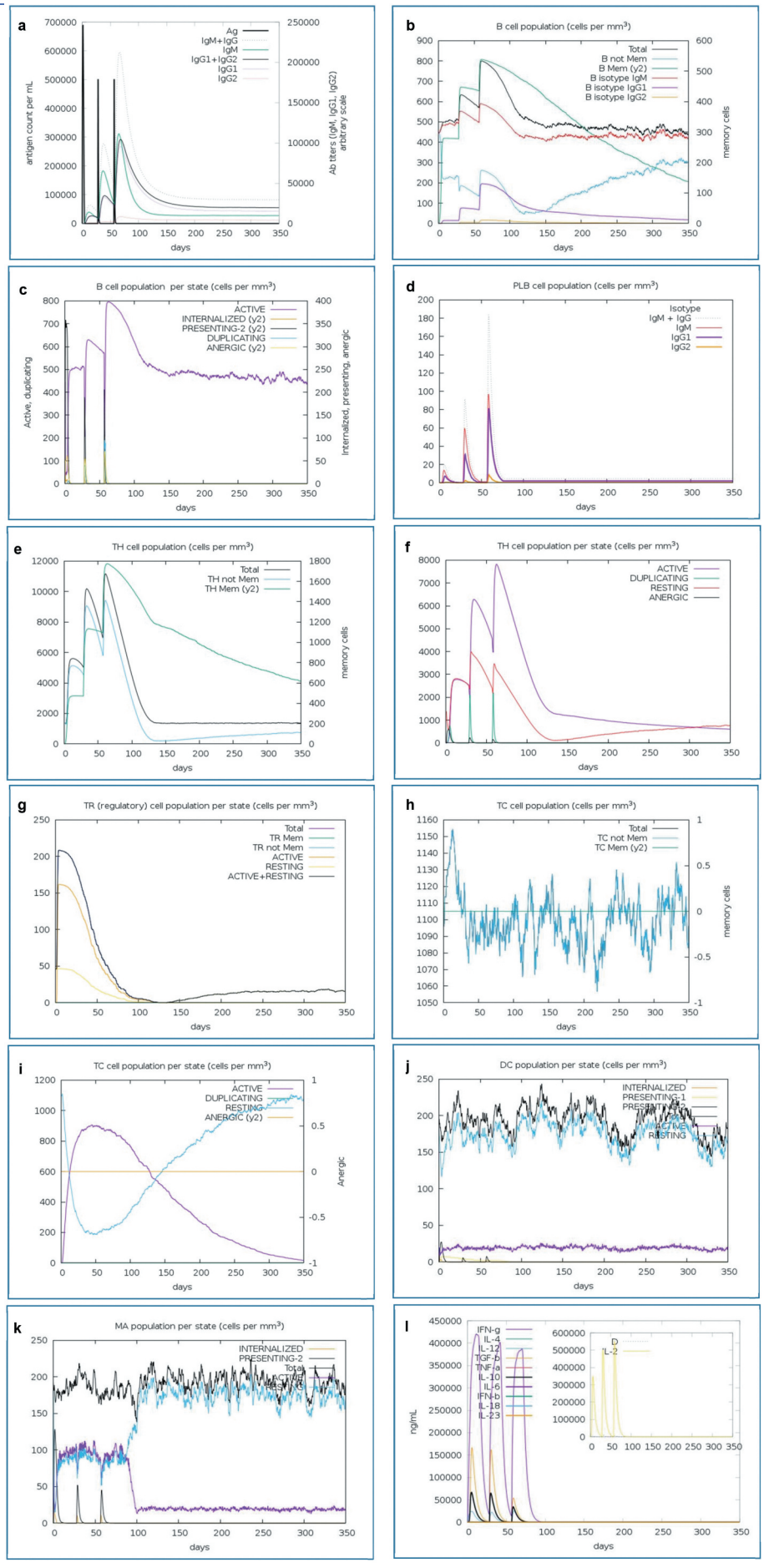
3.16 Codon Adaptation, In Silico *Cloning, Prediction of the mRNA Secondary Structure and Expression of the best Vaccine Protein with SUMO-fusion*

The codon adaptation experiment of best vaccine construct V3 revealed that the adapted sequence had Codon Adaptation Index of 0.94 and GC content of 52.4%. The newly adapted DNA sequence was then inserted into EaeI and StyI restriction sites of the pETite vector plasmid. Upon translation in the *E. coli* host, the vaccine protein was expected to be expressed in fusion with SUMO protein and 6X His tag, which should help effective purification and solubilization of the protein [107] (**Supplementary Figure 9** and Figure 6). The newly constructed plasmid with the V3 sequence was designated as 'CSMV_3' plasmid.

When the lowest minimum free energy of the V3 vaccine was determined by the Mfold server, the best predicted structure (among 46 generated structures by the server) for the optimized construct showed ΔG value of −656.30 kcal/mol. The outcomes of the Mfold server were in agreement with the data generated by RNAfold server, where the ΔG value of the mRNA structure was predicted to be −617.70 kcal/mol. **Supplementary Figure S10** illustrates the predicted secondary structure of the V3 vaccine.

4. Discussion

Vaccines are widely produced and used worldwide to control and prevent the infections caused by many types of pathogens. Conventional approaches are mainly used for vaccine development and production, although they are costly and time-consuming [14]. In contrast to such conventional methods of vaccine development, today's cutting-edge research and technology as well as the availability of information about the genome and proteome of almost all the viruses and organisms has made it possible to design and develop novel peptide-based 'subunit vaccines' comprised of the antigenic protein portions from a target pathogen. A major advantage of subunit vaccines is that toxic and immunogenic parts of an antigen can be eliminated during a vaccine designing study so that the vaccine would be safe to use in humans [113–115]. This had led to the development of techniques of bioinformatics and immunoinformatics which can be explored to design novel subunit vaccines in a safe, effective, efficient and inexpensive way [116,117]. In this study, these methods of


a
Ag
IgM+IgG
IgM
IgG1+IgG2
IgG1
IgG2
antigen count per mL
Ab titers (IgM, IgG1, IgG2) arbitrary scale
days
b
B cell population (cells per mm3)
Total
B not Mem
B Mem (y2)
B isotype IgM
B isotype IgG1
B isotype IgG2
memory cells
days
c
B cell population per state (cells per mm3)
ACTIVE
INTERNALIZED (y2)
PRESENTING-2 (y2)
DUPLICATING
ANERGIC (y2)
Active, duplicating
Internalized, presenting, anergic
days
d
PLB cell population (cells per mm3)
Isotype
IgM + IgG
IgM
IgG1
IgG2
days
e
TH cell population (cells per mm3)
Total
TH not Mem
TH Mem (y2)
memory cells
days
f
TH cell population per state (cells per mm3)
ACTIVE
DUPLICATING
RESTING
ANERGIC
days
g
TR (regulatory) cell population per state (cells per mm3)
Total
TR Mem
TR not Mem
ACTIVE
RESTING
ACTIVE+RESTING
days
h
TC cell population (cells per mm3)
Total
TC not Mem
TC Mem (y2)
memory cells
days
i
TC cell population per state (cells per mm3)
ACTIVE
DUPLICATING
RESTING
ANERGIC (y2)
Anergic
days
j
DC population per state (cells per mm3)
INTERNALIZED
PRESENTING-1
PRESENTING-2
ACTIVE
RESTING
days
k
MA population per state (cells per mm3)
INTERNALIZED
PRESENTING-2
Total
ACTIVE
RESTING
days
l
IFN-g
IL-4
IL-12
TGF-b
TNF-a
IL-10
IL-6
IFN-b
IL-18
IL-23
D
IL-2
ng/mL
days

immunoinformatics were exploited to design possible polyvalent vaccines against multiple strains of the HCoVs i.e., HCoV-OC43, SARS-CoV, MERS-CoV, and SARS-CoV-2.

The target protein sequences, spike glycoproteins (SGs) of the four HCoV strains, were identified and retrieved from the NCBI database. Thereafter, their antigenicity and physicochemical properties i.e., theoretical pI, extinction co-efficient, estimated half-life, instability index, aliphatic index, grand average of hydropathicity (GRAVY), and molecular weight were determined. All the proteins were found to be antigenic. The theoretical pI describes the pH at which aprotein should have no net charge. The extinction co-efficient of acompound depicts the amount of light that is absorbed by that compound at acertain wavelength [118,119]. The instability index of acompound describes the probability of that particular compound to be stable and acompound with instability an index over 40 is considered to be unstable [120]. The aliphatic index of aprotein refers to the relative volume of the amino acids in its side chains occupied by the aliphatic amino acids i.e., alanine, valine, etc [121]. The GRAVY value of aprotein is represented as the sum of hydropathy values of all the amino acids of that protein which is then divided by the total number of residues in the protein sequence. The negative GRAVY value represents hydrophilic characteristic and the positive GRAVY value represents hydrophobic characteristic of acompound [122,123]. All the SG proteins from the four viruses were found to be potentially antigenic. The highest extinction co-efficient of SG from HCoV-OC43 of 196,595 $M^{-1}cm^{-1}$ describes that it was predicted to absorb the highest amount of light at acertain wavelength. Again, since all of the proteins had instability index of less than 40, all of them were predicted to be stable and have the same amount of aliphatic amino acids in their side chains, according to the aliphatic index prediction. Moreover, all of the proteins had theoretical half-life of 30 h and their negative GRAVY values represented their hydrophilic characteristics.

An effective multi-epitope subunit vaccine should contain the CTL, HTL, and B-cell epitopes so that during the immunogenic response, the vaccine will stimulate the cytotoxic T-cells, helper Tcells, and B-cell. These cells are the most important cells that provide immunity to the body [115]. The B-cells mediate the humoral immune response by producing antibodies and keeping the memory of a previous infection. However, the humoral immune response provided by the B-cells may get weaker overtime and pathogens or antigens can easily overcome the humoral immune response [124]. In such cases, the cell-mediated immune response can provide much broader, often life-long immunity by secreting antiviral cytokines and specifically identifying and destroying the infected cells. For this reason, the T-cell and B-cell epitopes were predicted for the vaccine construction.

The possible T-cell and B-cell epitopes of the proteins were predicted for vaccine construction so that the vaccines would be able to provoke potential immune responses. However, since a lot of possible epitopes were generated by the IEDB server, some special criteria were set for final selection of the best epitopes for vaccine construction. The epitopes must be highly antigenic because if they are not antigenic, then they would not be able to induce strong immune response. Again, the epitopes have to be non-allergenic and nontoxic so that they would not be able to produce any harmful, toxic and allergenic reaction within the body. Again, since polyvalent vaccines were designed in this study, so the epitopes that were found to be 100% conserved among the selected SG proteins of different isolates, were considered as potential epitopes for the polyvalent vaccines. The 100% conservancy had ensured their efficacy and potency among different isolates of the HCoV viruses. Furthermore, the epitopes must be non-homologous to the human proteome because if they are homologous, then they won't be recognized as foreign antigenic sequences or particles. The epitopes that followed these criteria were considered as the best-selected epitopes and used for vaccine construction. Again, since cytokines like the IFNgamma, IL-10, and IL-4 are required for the activation of many immune cells [33], the cytokine production ability of the HTL epitopes was also predicted. Most of the HTL epitopes were found to be at least one cytokine inducer (among IFN-gamma, IL-10, and IL-4) and all the HTL epitopes selected for vaccine construction were also found to possess at least on cytokine inducing ability. This ability would impact greatly on the immunogenic activities of the vaccine. In the next step, the population coverage analysis was conducted. The population coverage analysis of the MHC epitopes and alleles showed that significant portions of the population around different countries of the world possess the epitopes and respective alleles within their genome. The prevalence of the selected alleles and epitopes was relatively higher among the Asian population. Then the molecular docking study of the best-selected epitopes was carried out. Since all the epitopes showed good results in the docking study, it can be concluded that all of the best-selected epitopes had the capability to bind with their respective MHC class-I and MHC class-II molecules. After successful docking, the best-selected epitopes were conjugated with each other using the appropriate linkers (EAAAK, GGGGS, GPGPG, KK) to construct three vaccines, V1, V2, and V3. All these vaccines were predicted to be highly

Figure 5. C-IMMSIMM representation of the immune simulation of the best predicted vaccine, V3. a. The immunoglobulin and immunocomplex response to the antigen (V3 vaccine) inoculations (black colored lines) and specific subclasses are indicted by colored lines, b. Increase in the B-cell population over the course of the three injections, c. Elevation of the B-cell population per state over the course of vaccination, d. Increase in the plasma B-cell population over the course of vaccination, e. Increase in the helper T-cell population over the course of the injections, f. Elevation of the helper T-cell population per state over the course of three injections, g. Increase in the regulatory T lymphocyte over the course of the vaccinations, h. Rise in the cytotoxic T lymphocyte population over the course of the vaccinations, i. Increase in the active cytotoxic T lymphocyte population per state over the course of the injections, j. Augmentation of the active dendritic cell population per state over the course of the vaccinations, k. Increase in the macrophage population per state over the course of three injections, l. Rise in the concentrations of different types of cytokines over the course of vaccination.

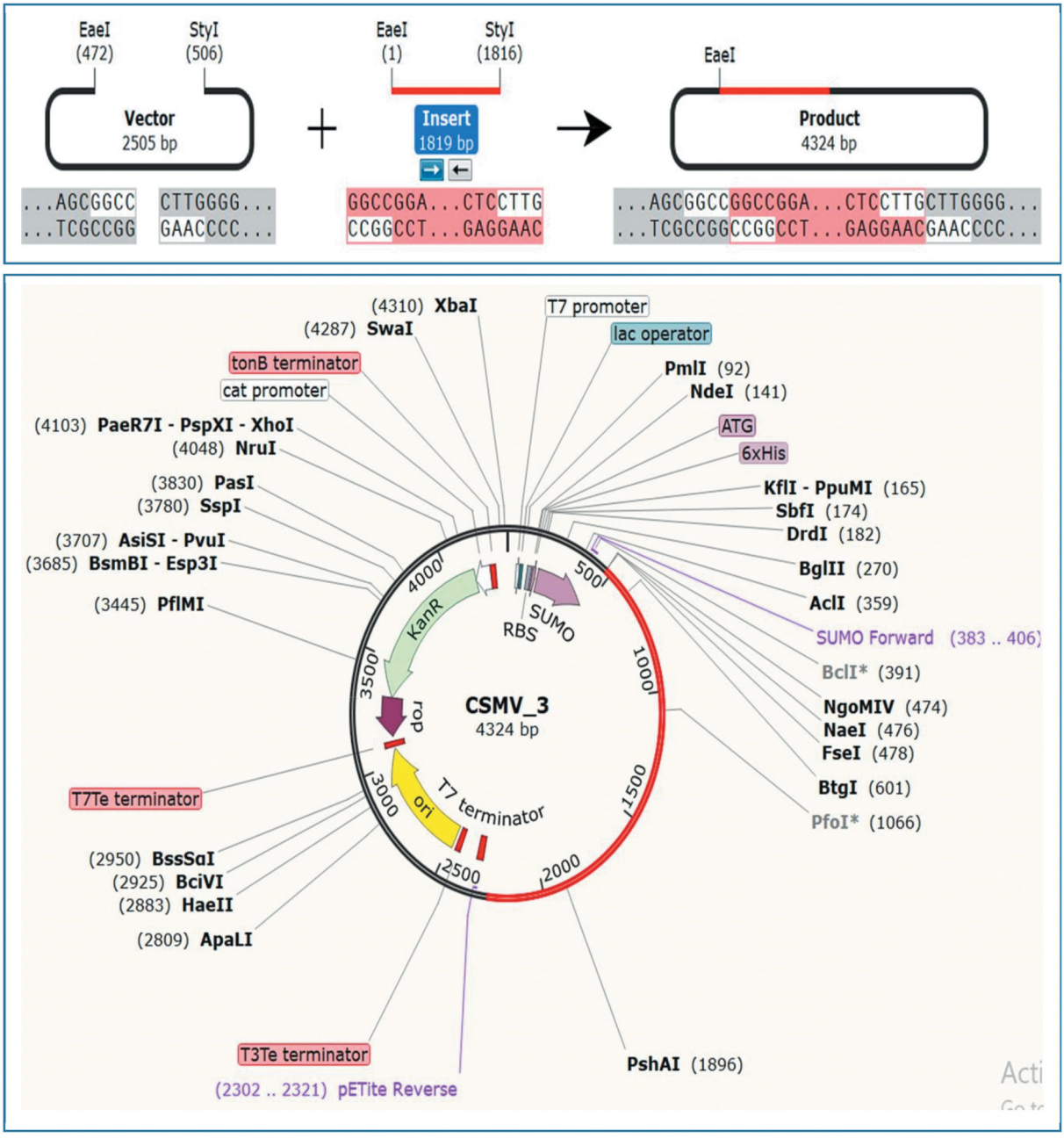


Figure 6. The results of the *in silico* cloning study of the V3 vaccine construct. The adapted DNA sequence of the V3 vaccine was inserted into the pETite plasmid between EaeI and StyI restriction sites which is indicated by the red color in the plasmid. The plasmid with the sequence had been designated as 'CSMV_3' plasmid.

antigenic as well as nonallergenic. For this reason, they might provoke high immune responses and at the same time, they would not cause any unwanted allergenic reactions within the body.

Thereafter, the physicochemical properties of the three vaccine constructs were determined. All the proteins were found to be basic with theoretical pI of more than 9.00 and quite stable with very good instability indexes (less than 40). Again, the aliphatic index represents the protein's thermal stability and the higher the aliphatic index of a protein, the more thermostable it is. Since all the vaccine constructs were predicted to have quite high aliphatic indexes, all of them were considered to be quite thermostable. Furthermore, the negative GRAVY value of the vaccine constructs revealed that all of them might be hydrophilic in nature. Moreover, all of the vaccine constructs were predicted to have more than 10 h of half-life in *E. coli*, which might not be a problem during the mass production of the vaccines. Solubility is one of the major factors for the post-production studies. The more soluble a protein on overexpression, the easier the purification of that protein is []. Since all the proteins were predicted to be soluble upon over-expression in *E. coli* by both servers (SolPro and Protein-Sol), therefore, their purification steps should be much easier. Considering all these aspects, the predicted vaccine constructs might be suitable for a potential vaccine candidate.

The secondary and tertiary structure prediction of the three vaccine constructs showed that the adjuvant sequences had caused some significant differences among the three vaccines. All the tools predicted almost similar percentages of amino acids in the α-helix, β-strand, and coil structure formation for each of the vaccines. After secondary structure determination, the tertiary structures of the three vaccine constructs were predicted. The tertiary structure refinement and validation study revealed that the quality of all the three vaccine constructs were significantly improved after refinement in the context of GDT-HA, MolProbity, Rama favored amino acid percentage and zscores. All the three refined structures of the vaccine constructs resembled the crystallographic structures and all of them had quite satisfactory Rama favored amino acid percentage with very few amino acids in the outlier regions. After that, the refined structures were used for vaccine protein disulfide engineering. Since disulfide bonds confer structural stability to the proteins, V2 construct with its eight disulfide bonds was predicted to be more stable than the other two constructs.

The docking of the three vaccine constructs with different MHC alleles as well as TLR-2, TLR-4, and TLR-8 was conducted to find out the best vaccine construct. The docking step is very important for two reasons: it would help to predict the interaction ability of the vaccines with different MHC alleles and the docking would also aid in prediction of the best vaccine from these three constructs. So, the docking was conducted using several online tools for better prediction accuracy. When analyzing the docking results, all the three vaccine constructs showed good capability of binding to all the MHC alleles and also the TLRs. However, V3 was considered as the best vaccine construct since the MM-GBSA study and docking studies of V3 by all the servers predicted the best scores against almost all the targets. Although, V1 and V2 also showed quite good outcomes in some aspects, the best results were generated by V3 construct with most of the receptor proteins. And V3 also showed the best docking results by all the servers when docked with the selected TLRs. Therefore, V3 was considered as the best vaccine construct and selected for further analysis.

For providing a strong humoral immunity, a vaccine construct should have conformational B-cell epitopes because these epitopes would aid to activate and stimulate the B-cells when these cells encounter them. Therefore, the discontinuous conformational B-cell epitopes on the surface of the best vaccine V3 were predicted that can mimic the original infection and thus stimulate the antibody production [95]. Total 357 amino acids in five different B-cell epitopic regions were identified with scores ranging from 0.516 to 0.876. The molecular dynamics simulation study was performed for docked TLR-8 and V3 complex using the online tool iMODS. The study showed that the complex had less chance of deformability with quite high eigenvalue of 1.278806e-07. The deformability graph (Figure 4b) had confirmed that the location of the hinges in the complex were not quite significant and therefore, the complex might have good stability with a lower degree of deforming for each individual amino acid residue. Moreover, the complex also had a good number of correlated amino acids and a large number of stiffer regions. Therefore, the V3-TLR-8 complex showed good results in the molecular dynamics simulation study.

The immune simulation study of the best-selected vaccine V3 showed that the immune response of the vaccine within the host might be consistent with the typical immune response. After each of the vaccine injections, the primary immune response was found to be triggered, which later activated the secondary immune response. An increase in the concentrations of the memory Bcells, plasma B-cell, cytotoxic T-cells, and helper T-cells as well as different antibodies indicated that a good humoral and cell-mediated immune response might have been built in the body after each of the vaccinations and the memory B-cell was predicted to last for several months after the vaccinations. Again, the stimulated helper T-cells will also aid in increasing the growth and proliferation of B-cells, thus elevating the adaptive immunity. Moreover, the increase in the concentrations of macrophages and dendritic cells ensured very good antigen presentation and the increase in the cytokine profile after each of the vaccinations might also contribute to the immunity provided by the vaccine. On the other hand, the negligible Simpson index (D) suggests a diverse immune response [101]. Since the vaccine construct contained multiple numbers of B and T-cell epitopes, so it would be able to generate a diverse immune response. As a result, it can be declared that the vaccine V3 might be able to generate good immune response in the body.

Finally, codon adaptation and in silico cloning studies were carried out to identify the possible codons for expressing the vaccine V3 in *E. coli* strain K12. The in silico cloning was conducted for *E. coli* as the host organism because *E. coli* expression is the recommended system for the production of recombinant proteins. During codon adaptation, CAI value 0.94 and GC content of 52.4% was generated. The optimal range and limit of the CAI value were measured to be close to 1.0 but the any value greater than 0.80 can be considered as good score. Again, the optimal range for GC content of an optimized DNA sequence should be 30% to 70% [103. Therefore, the CAI value of 0.94 and GC content of 52.4% can be declared as very good scores. Thereafter, the optimized sequence was inserted into the EaeI and StyI restriction sites of pETite plasmid vector so that the efficient expression of the vaccine protein would occur. Moreover, the expressed protein might have SUMO and 6X His tag fused with it since the pETite plasmid offers these tags as fusion partners. Fusion of these tags would help the downstream processing of the vaccine. The newly constructed plasmid with the vaccine V3 sequence was designated as 'CSMV_3' plasmid. When the stability of the mRNA secondary structure of the vaccine protein was determined, the both Mfold and RNAfold servers generated negative and much lower minimal free energies of −656.30 and −617.70 kcal/mol, respectively. Since, the lower minimal free energy always represents better mRNA stability, so it can be concluded that the predicted vaccine protein might be quite stable upon transcription.

The genome based technology will continue to dominate in the field of vaccine development. The current COVID-19 pandemic will require much more attention from the immunoinformatics field to design and develop potential vaccine candidates

which will be effective to combat the SARS-CoV-2. Scientists from around the world have already started working on designing vaccine using this novel approach. Yazdani et al., Joshi et al., Lizbeth et al., Chauhan et al., and other authors have already proposed probable blue-prints effective against the SARS-CoV -2. The authors have used different proteins and different epitopes in numerous combinations in their works and all these designed vaccines showed quite promising results in their experiments. However, in our work, novel polyvalent vaccines were designed that might be effective against the four most dangerous strains of HCoVs i.e., HCoV-OC43, HCoV-SARS, HCoV-MERS, and SARS-CoV-2, targeting their highly accessible spike proteins. As a result, such vaccines will be effective against all the selected viral strains simultaneously and no separate vaccine will be needed for each of the strains. Such works on polyvalent vaccines are not very common but our findings should definitely open new avenues to design epitope based polyvalent vaccines to fight against these lethal viruses.

Overall, this study recommends V3 vaccine as the best vaccine construct based on the strategies employed in the study to be an effective countermeasure against the four mentioned HCoVs. However, further research are suggested to finally determine and validate the immunogenicity, stability, safety, efficacy, and various physicochemical characteristics of the suggested vaccines of this study.

5. Expert opinion

The *Coronaviridae* family of viruses is responsible for many outbreaks over the past two decades and the current pandemic caused by the SARS-CoV-2 has already taken the lives of millions of people. Scientists are working hard and soul to develop effective vaccines against all these strains, although none of them generated sound and satisfactory outcomes. In our study, effective polyvalent vaccines were designed by mining the spike glycoproteins of the four most prevalent strains of the *Coronaviridae* family i.e., HCoV-OC43, SARS-CoV, MERS-CoV, and SARS-CoV-2. If positive results are achieved in further research, then our designed vaccines might have the capability to fight off these deadly viruses.

6. Conclusion

Human Coronaviruses (HCoVs) are a group of lethal viruses that has caused many outbreaks in the recent decades and at present the SARS-CoV-2 is the new strain of HCoV and its outbreak has become a global pandemic. The scientific community around world is racing for a satisfactory prevention measure to control the transmission of coronaviruses, especially the spread of SARS-CoV-2. In this study, epitope-based polyvalent vaccines were developed that might confer immunogenic protection against the four strains of HCoV i.e., HCoV-OC43, HCoV-MERS, HCoV-SARS, and SARS-CoV-2. Since the vaccines contained multiple T-cell as well as B-cell epitopes from all of these four viruses, they are expected to provoke both humoral and cell-mediated immunogenic responses within the body. Results of different experiments that were conducted in the study indicated that these polyvalent vaccines should be quite safe, effective, and responsive to use. However, since all these predictions were done based on the computational methods, more wet lab-based research is needed to finally confirm the outcomes of this study. With high cost requirements and multiple limitations for developing the live, attenuated or inactivated vaccine preparation for contagious agents like the HCoV, these peptide-based vaccine candidates might be relatively cheap and effective options to reach the entire world to combat the HCoVs.

Acknowledgments

Authors are thankful to Swift Integrity Computational Lab, Dhaka, Bangladesh, a virtual platform of young researchers, for providing the support and tools.

Declaration of interest

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

Reviewer disclosures

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

ORCID

Bishajit Sarkar http://orcid.org/0000-0001-8439-6994
Md. Asad Ullah http://orcid.org/0000-0002-6615-6433
Yusha Araf http://orcid.org/0000-0002-0144-5875

References

1. Paules CI, Marston HD, Fauci AS. Coronavirus infections—more than just the common cold. Jama. 2020;323(8):pp.707–708.
2. Pyrc K, Berkhout B, Van der hoek L. Identification of new human coronaviruses. Expert Rev Anti Infect Ther. 2007;5(2):pp.245–253.
3. Banerjee A, Kulcsar K, Misra V, et al. Bats and coronaviruses. Viruses. 2019;11(1):p.41.
4. Sexton NR, Smith EC, Blanc H, et al. Homology-based identification of a mutation in the coronavirus RNA-dependent RNA polymerase that confers resistance to multiple mutagens. J Virol. 2016;90(16): pp.7415–7428.
5. Lai MM, Cavanagh D. The molecular biology of coronaviruses. Adv Virus Res. 1997;48: pp.1–100. Academic Press.
6. Jean A, Quach C, Yung A, et al. Severity and outcome associated with human coronavirus OC43 infections among children. Pediatr Infect Dis J. 2013Apr1;32(4):325–329.
7. de wit E, van doremalen N, Falzarano D, et al. SARS and MERS: recent insights into emerging coronaviruses. Nature Rev Microbiol. 2016;14(8):p.523.
8. https://www.worldometers.info/coronavirus/. Accessed on: 2020Apr04.
9. Bernheim A, Mei X, Huang M, et al. Chest ct findings in coronavirus disease-19 (covid-19): relationship to duration of infection. Radiology. 2020;295(3):p.200463.
10. Yin Y, Wunderink RG. MERS, SARS and other coronaviruses as causes of pneumonia. Respirology. 2018;23(2):pp.130–137.
11. Li G, Fan Y, Lai Y, et al. Coronavirus infections and immune responses. J Med Virol. 2020;92(4):pp.424–432.

12. Qian Z, Dominguez SR, Holmes KV. Role of the spike glycoprotein of human middle east respiratory syndrome coronavirus (mers-cov) in virus entry and syncytia formation. PloS One. 2013;8:10.
13. Li W, Hulswit RJ, Widjaja I, et al. Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. Proc Nat Acad Sci. 2017;114(40):pp.E8508–E8517.
14. Gallagher TM, Buchmeier MJ. Coronavirus spike proteins in viral entry and pathogenesis. Virology. 2001;279(2):pp.371–374.
15. Li F. Structure, function, and evolution of coronavirus spike proteins. Annu Rev Virol. 2016;3:pp.237–261.
16. Bhattacharya M, Sharma AR, Patra P, et al. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): immunoinformatics approach. J Med Virol. 2020;92(6):pp.618631.
17. Doytchinova IA, Flower DR. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine. 2007;25(5):pp.856866.
18. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007;8(1):p.4.
19. Chaudhri G, Quah BJ, Wang Y, et al. T cell receptor sharing by cytotoxic T lymphocytes facilitates efficient virus control. Proc Nat Acad Sci. 2009;106(35):pp.14984–14989.
20. Zhu J, Paul WE. Cd4 t cells: fates, functions, and faults. blood. J Am Soc Hematol. 2008;112(5):pp.1557–1569.
21. Vita R, Mahajan S, Overton JA, et al. The immune epitope database (IEDB): 2018 update. Nucleic Acids Res. 2019;47(D1):pp.D339–D343.
22. Cooper MD. The early history of B cells. Nat Rev Immunol. 2015;15 (3):pp.191–197.
23. Dimitrov I, Naneva L, Doytchinova I, et al. AllergenFP: allergenicity prediction by descriptor fingerprints. Bioinformatics. 2014;30(6): pp.846–851.
24. Dimitrov I, Flower DR, Doytchinova I. AllerTOP-a server for in silico prediction of allergens. In BMC bioinformatics. BMC. 2013 April; Vol.14(No. 6):S4.
25. Gupta S, Kapoor P, Chaudhary K, et al. Consortium, osdd; raghava, gps in silico approach for predicting toxicity of peptides and proteins. PLoS One. 2013;8:p.e73957.
26. Mehla K, Ramana J. Identification of epitope-based peptide vaccine candidates against enterotoxigenic escherichia coli: a comparative genomics and immunoinformatics approach. Mol Biosyst. 2016;12 (3):pp.890–901.
27. Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. Biol Direct. 2013;8(1):p.30.
28. Nagpal G, Usmani SS, Dhanda SK, et al. Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. Sci Rep. 2017;7(p):42851.
29. Möller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics. 2001;17 (7):pp.646–653.
30. Adhikari UK, Rahman MM. Overlapping cd8+ and cd4+ t-cell epitopes identification for the progression of epitope-based peptide vaccine from nucleocapsid and glycoprotein of emerging rift valley fever virus using immunoinformatics approach. Infect Genet Evol. 2017;56:pp.75–91.
31. Bui HH, Sidney J, Dinh K, et al. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics. 2006;7(1):p.153.
32. Lamiable A, Thévenet P, Rey J, et al. PEPFOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res. 2016;44(W1):pp.W449–W454.
33. Shen Y, Maupetit J, Derreumaux P, et al. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. J Chem Theory Comput. 2014;10(10):pp.4745–4758.
34. Thévenet P, Shen Y, Maupetit J, et al. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. Nucleic Acids Res. 2012;40(W1):pp.W288–W293.
35. Biovia DS, Berman HM, Westbrook J, et al. Dassault systèmes biovia, discovery studio visualizer, v. 17.2, san diego: dassault systèmes, 2016. J Chem Phys. 2000;10:pp.0021–9991.
36. Meza B, Ascencio F, Sierra-beltrán AP, et al. A novel design of a multi-antigenic, multistage and multi-epitope vaccine against helicobacter pylori: an in silico approach. Infect Genet Evol. 2017;49: pp.309–317.
37. Lee S, Nguyen MT. Recent advances of vaccine adjuvants for infectious diseases. Immune Netw. 2015;15(2):pp.51–57.
38. Arai R, Ueda H, Kitayama A, et al. Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Eng. 2001;14(8):pp.529–532.
39. Saadi M, Karkhah A, Nouri HR. Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. Infection. Gene Evolut. 2017;51:pp.227–234.
40. Gu Y, Sun X, Li B, et al. Vaccination with a paramyosinbased multi-epitope vaccine elicits significant protective immunity against trichinella spiralis infection in mice. Front Microbiol. 2017;8:p.1475.
41. Wen D, Foley SF, Hronowski XL, et al. Discovery and investigation of O-xylosylation in engineered proteins containing a (GGGGS) n linker. Anal Chem. 2013;85(9):pp.4805–4812.
42. Law HK, Cheung CY, Sia SF, et al. Toll-like receptors, chemokine receptors and death receptor ligands responses in SARS coronavirus infected human monocyte derived dendritic cells. BMC Immunol. 2009;10(1):p.35.
43. Rana A, Akhter Y. A multi-subunit based, thermodynamically stable model vaccine using combined immunoinformatics and protein structure based approach. Immunobiology. 2016;221(4): pp.544–557.
44. Hancock RE, Nijnik A, Philpott DJ. Modulating immunity as a therapy for bacterial infections. Nature Rev Microbiol. 2012;10 (4):pp.243–254.
45. Toussi DN, Massari P. Immune adjuvant effect of molecularly-defined toll-like receptor ligands. Vaccines (Basel). 2014;2(2):pp.323–353.
46. Funderburg N, Lederman MM, Feng Z, et al. Human β-defensin -3 activates professional antigen-presenting cells via toll-like receptors 1 and 2. Proc Nat Acad Sci. 2007;104(47): pp.1863118635.
47. Lee SJ, Shin SJ, Lee MH, et al. A potential protein adjuvant derived from mycobacterium tuberculosis rv0652 enhances dendritic cells-based tumor immunotherapy. PloS One. 2014;9:8.
48. Hajighahramani N, Nezafat N, Eslami M, et al. Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against staphylococcus aureus. infection. Gene Evolut. 2017;48:pp.83–94.
49. Pandey RK, Sundar S, Prajapati VK. Differential expression of miRNA regulates T cell differentiation and plasticity during visceral leishmaniasis infection. Front Microbiol. 2016;7:p.206.
50. Wu CY, Monie A, Pang X, et al. Improving therapeutic HPV peptide-based vaccine potency by enhancing CD4+ T help and dendritic cell activation. J Biomed Sci. 2010;17(1):p.88.
51. Magnan CN, Zeller M, Kayala MA, et al. High-throughput prediction of protein antigenicity using protein microarray data. Bioinformatics. 2010;26(23):pp.2936–2943.
52. Saha S, Raghava GP. S: algPred: prediction of allergenic proteins and mapping of ige epitopes. Nucl Acids Res. 2006;34:pp.202–209.
53. Magnan CN, Randall A, Baldi P. SOLpro: accurate sequence-based prediction of protein solubility. Bioinformatics. 2009;25(17): pp.2200–2207.
54. Hebditch M, Carballo-amador MA, Charonis S, et al. Protein– sol: a web tool for predicting protein solubility from sequence. Bioinformatics. 2017;33(19):pp.30983100.
55. Buchan DW, Jones DT. The PSIPRED protein analysis workbench: 20 years on. Nucleic Acids Res. 2019;47(W1):pp.W402–W407.
56. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999;292(2): pp.195–202.
57. Garnier J, Gibrat JF, Robson B. [32] GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol. 1996;266: pp.540–553. Academic Press.

58. Montgomerie S, Sundararaj S, Gallin WJ, et al. Improving the accuracy of protein secondary structure prediction using structural alignment. BMC Bioinformatics. 2006;7(1):p.301.
59. Källberg M, Wang H, Wang S, et al. Template-based protein structure modeling using the raptorx web server. Nat Protoc. 2012;7(8): p.1511.
60. Ma J, Wang S, Zhao F, et al. Protein threading using context-specific alignment potential. Bioinformatics. 2013;29(13): pp.i257–i265.
61. Peng J, Xu J. RaptorX: exploiting structure information for protein alignment by statistical inference. proteins. Struct Funct Bioinf. 2011;79(S10):pp.161–171.
62. Nugent T, Cozzetto D, Jones DT. Evaluation of predictions in the CASP10 model refinement category. Proteins Struct Funct Bioinf. 2014;82:pp.98–111.
63. Pandey RK, Bhatt TK, Prajapati VK. Novel immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating anopheles salivary protein. Sci Rep. 2018;8(1):pp.1–11.
64. Ko J, Park H, Heo L, et al. GalaxyWEB server for protein structure prediction and refinement. Nucleic Acids Res. 2012;40(W1):pp. W294–W297.
65. Morris AL, MacArthur MW, Hutchinson EG, et al. Stereochemical quality of protein structure coordinates. proteins: structure. Funct Bioinf. 1992;12(4):pp.345–364.
66. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007;35(suppl_2):pp.W407–W410.
67. Craig DB, Dombkowski AA. Disulfide by design 2.0: a web-based tool for disulfide engineering in proteins. BMC Bioinformatics. 2013;14(1):p.346.
68. Petersen MTN, Jonson PH, Petersen SB. Amino acid neighbours and detailed conformational analysis of cysteines in proteins. Protein Eng. 1999;12(7):pp.535–548.
69. Stern LJ, Calvo-calle JM. HLA-DR: molecular insights and vaccine design. Curr Pharm Des. 2009;15(28):pp.3249–3261.
70. Thompson JM, Iwasaki A. Toll-like receptors regulation of viral infection and disease. Adv Drug Deliv Rev. 2008;60(7):pp.786–794.
71. Lester SN, Li K. Toll-like receptors in antiviral innate immunity. J Mol Biol. 2014;426(6):pp.1246–1264.
72. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. in coronaviruses. Humana Press: New York, NY; 2015. pp.1–23.
73. Vajda S, Yueh C, Beglov D, et al. New additions to the c lus p ro server motivated by capri. proteins: structure. Funct Bioinf. 2017;85 (3):pp.435–444.
74. Kozakov D, Hall DR, Xia B, et al. The cluspro web server for protein–protein docking. Nat Protoc. 2017;12(2):p.255.
75. Kozakov D, Beglov D, Bohnuud T, et al. How good is automated protein docking? Proteins: structure. Funct Bioinf. 2013;81(12): pp.2159–2166.
76. Weng G, Wang E, Wang Z, et al. HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA. Nucleic Acids Res. 2019;47(W1):pp.W322–W330.
77. Hou T, Wang J, Li Y, et al. Assessing the performance of the mm/pbsa and mm/gbsa methods. 1. the accuracy of binding free energy calculations based on molecular dynamics simulations. J Chem Inf Model. 2011;51(1):pp.69–82.
78. Feng T, Chen F, Kang Y, et al. HawkRank: a new scoring function for protein–protein docking based on weighted energy terms. J Cheminform. 2017;9(1):p.66.
79. Sun H, Li Y, Tian S, et al. Assessing the performance of mm/pbsa and mm/gbsa methods. 4. accuracies of mm/pbsa and mm/gbsa methodologies evaluated by various simulation protocols using pdbbind data set. Phys Chem Chem Phys. 2014;16(31): pp.16719–16729.
80. Ponomarenko J, Bui HH, Li W, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics. 2008;9(1):p.514.
81. Awan FM, Obaid A, Ikram A, et al. Mutation-structure-function relationship based integrated strategy reveals the potential impact of deleterious missense mutations in autophagy related proteins on hepatocellular carcinoma (HCC): A comprehensive informatics approach. Int J Mol Sci. 2017;18(1):p.139.
82. Prabhakar PK, Srivastava A, Rao KK, et al. Monomerization alters the dynamics of the lid region in campylobacter jejuni cstii: an md simulation study. J Biomol Struct Dyn. 2016;34(4):pp.778–791.
83. López-blanco JR, Aliaga JI, Quintana-ortí ES, et al. iMODS: internal coordinates normal mode analysis server. Nucleic Acids Res. 2014;42(W1):pp.W271–W276.
84. Lopéz-blanco JR, Garzón JI, Chacón P. iMod: multipurpose normal mode analysis in internal coordinates. Bioinformatics. 2011;27(20): pp.2843–2850.
85. Kovacs JA, Chacón P, Abagyan R. Predictions of protein flexibility: first-order measures. proteins: structure. Funct Bioinf. 2004;56(4): pp.661–668.
86. Rapin N, Lund O, Bernaschi M, et al. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. PLoS One. 2010;5:4.
87. Khatoon N, Pandey RK, Prajapati VK. Exploring leishmania secretory proteins to design b and t cell multi-epitope subunit vaccine using immunoinformatics approach. Sci Rep. 2017;7(1):pp.1–12.
88. Angov E. Codon usage: nature's roadmap to expression and folding of proteins. Biotechnol J. 2011;6(6):pp.650–659.
89. Grote A, Hiller K, Scheer M, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res. 2005;33(suppl_2):pp.W526–W531.
90. Solanki V, Tiwari V. Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against acinetobacter baumannii. Sci Rep. 2018;8(1): pp.1–19.
91. Biswal JK, Bisht P, Mohapatra JK, et al. Application of a recombinant capsid polyprotein (P1) expressed in a prokaryotic system to detect antibodies against foot-and-mouth disease virus serotype O. J Virol Methods. 2015;215:pp.45–51.
92. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31(13):pp.3406–3415.
93. Mathews DH, Sabina J, Zuker M, et al. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol. 1999;288(5):pp.911–940.
94. Mathews DH, Turner DH, Zuker M. RNA secondary structure prediction. Curr Protoc Nucleic Acid Chem. 2007;28(1):pp.11–2.
95. Gruber AR, Lorenz R, Bernhart SH, et al. The vienna RNA websuite. Nucleic Acids Res. 2008;36(suppl_2):pp.W70–W74.
96. Chauhan V, Rungta T, Goyal K, et al. Designing a multi-epitope based vaccine to combat kaposi sarcoma utilizing immunoinformatics approach. Sci Rep. 2019;9(1):pp.1–15.
97. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. Nat Rev Drug Discov. 2007;6(5):pp.404–414.
98. De groot AS, Sbai H, Aubin CS, et al. Immuno-informatics: mining genomes for vaccine components. Immunol Cell Biol. 2002;80(3): pp.255–269.
99. Zhang L. Multi-epitope vaccines: a promising strategy against tumors and viral infections. Cell Mol Immunol. 2018;15(2): pp.182–184.
100. Rappuoli R, Bottomley MJ, D'Oro U, et al. Reverse vaccinology 2.0: human immunology instructs vaccine antigen design. J Exp Med. 2016;213(4):pp.469–481.
101. Rappuoli R. Reverse vaccinology. Curr Opin Microbiol. 2000;3(5): pp.445–450.
102. Gill SC, Von hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem. 1989;182(2): pp.319–326.
103. Pace CN, Vajdos F, Fee L, et al. How to measure and predict the molar absorption coefficient of a protein. Protein Sci. 1995;4(11): pp.2411–2423.

104. Guruprasad K, Reddy BB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Engineering. Des Sel. 1990;4(2): pp.155–161.
105. Ikai A. Thermostability and aliphatic index of globular proteins. J Biochem. 1980;88(6):pp.1895–1898.
106. Kyte J, Doolittle. RF. Mol. Biol. 1982;157:pp.105–132.
107. Chang KY, Yang JR. Analysis and prediction of highly effective antiviral peptides based on random forests. PloS One. 2013;8:8.
108. Cooper NR, Nemerow GR. The role of antibody and complement in the control of viral infections. J Invest Dermatol. 1984;83(s 1): pp.121–127.
109. Bacchetta R, Gregori S, Roncarolo MG. CD4+ regulatory T cells: mechanisms of induction and effector function. Autoimmun Rev. 2005;4(8):pp.491–496.
110. Cano RLE, Lopera HDE. Introduction to T and B lymphocytes. In Autoimmunity: From Bench to Bedside [Internet]. El Rosario University Press; 2013.
111. Garcia KC, Teyton L, Wilson IA. Structural basis of T cell recognition. Annu Rev Immunol. 1999;17(1):pp.369–397.
112. Panda S, Chandra G. Physicochemical characterization and functional analysis of some snake venom toxin proteins and related non-toxin proteins of other chordates. Bioinformation. 2012;8(18): p.891.
113. Almofti YA, Abd-elrahman KA, Gassmallah SAE, et al. Multi epitopes vaccine prediction against severe acute respiratory syndrome (sars) coronavirus using immunoinformatics approaches. Am J Microbiol Res. 2018;6(3):pp.94–114.
114. Carvalho LH, Sano GI, Hafalla JC, et al. IL-4-secreting CD4+ T cells are crucial to the development of CD8+ T-cell responses against malaria liver stages. Nat Med. 2002;8(2):pp.166–170.
115. Shey RA, Ghogomu SM, Esoh KK, et al. In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. Sci Rep. 2019;9(1):pp.1–18.
116. Hoque MN, Istiaq A, Clement RA, et al. Metagenomic deep sequencing reveals association of microbiome signature with functional biases in bovine mastitis. Sci Rep. 2019;9(1):pp.1–14.
117. Kambayashi T, Laufer TM. Atypical MHC class II-expressing antigen-presenting cells: can anything replace a dendritic cell? Nat Rev Immunol. 2014;14(11):pp.719–730.
118. Pei H, Liu J, Cheng Y, et al. Expression of sars-coronavirus nucleocapsid protein in escherichia coli and lactococcus lactis for serodiagnosis and mucosal vaccination. Appl Microbiol Biotechnol. 2005;68(2):pp.220–227.
119. Morla S, Makhija A, Kumar S. Synonymous codon usage pattern in glycoprotein gene of rabies virus. Gene. 2016;584(1):pp.1–6.
120. Hamasaki-katagiri N, Lin BC, Simon J, et al. The importance of mRNA structure in determining the pathogenicity of synonymous and non-synonymous mutations in haemophilia. Haemophilia. 2017;23(1):pp.e8–e17.
121. Yazdani Z, Rafiei A, Yazdani M, et al. Design an efficient multi-epitope peptide vaccine candidate against sars-cov-2: an in silico analysis. Infect Drug Resist. 2020 Aug: ;3007–3022. doi: 10.2147/IDR.S264573
122. Joshi A, Joshi BC, Mannan MA, et al. Epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach. Inf Med Unlocked. 2020 Apr;29;100338
123. Lizbeth RS, Jazmín GM, José CB, et al. Immunoinformatics study to search epitopes of spike glycoprotein from sars-cov-2 as potential vaccine. J Biomol Struct Dyn. 2020 Jun;23;1–5
124. Chauhan V, Rungta T, Rawat M, et al. Excavating SARS-coronavirus 2 genome for epitope-based subunit vaccine synthesis using immunoinformatics approach. J Cell Physiol. [2020 Jul9]. doi: 10.1002/jcp.29923